SCOPE 3 EMISSIONS FROM PURCHASED GOODS: USING FOOD PURCHASING AS A CASE STUDY FOR BEST PRACTICES

A REPORT SUBMITTED TO THE DEPARTMENT OF ENERGY RESOURCES ENGINEERING

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

REBECCA GREKIN

06/03/2022

I certify that I have read this report and that in my opinion
it is fully adequate, in scope and in quality, as partial fulfillments of
the degree of Master of Science in Energy Resources Engineering.

Abstract

This work answers questions relevant to the determination of Scope 3 emissions from purchased goods: 1) how does the use of weight-based versus spend-based emissions factors (EF) affect the magnitude of emissions calculated, 2) how does using different sources for emissions factors within a same type of EF affect overall estimated emissions, and 3) what is needed to standardize and ease calculation of Scope 3 emissions across institutions, countries, and backgrounds such that the process can be automated and ensure replicable results. To answer these questions, a collaboration with seven universities across the United States was formed to share best practices and food purchasing data. This collaboration enabled the development of a python-based categorization tool to ease and standardize calculation of Scope 3 emissions from purchased foods and helped validate and expand the applicability of the results and methodologies. To understand the impact of different emissions factors on calculated emissions outcomes, the Stanford 2019 food purchasing data was used as a case study and both weight-based (kg $CO_{2\text{-eq}}$/kg item purchased) and spend-based (kg $CO_{2\text{-eq}}$/\$ item purchased) EFs were applied to this data set. It was found that for the 2019 Stanford food purchasing case study, the largest estimated emissions were about 2.5times larger than the smallest estimated emissions, and although the range of estimated emissions for the weight-based factors is larger than the range for the spend-based factors, the average is remarkably similar at 10,543 and 11,282 metric tons of CO_2-eq respectively.

Acknowledgements

First, I would like to thank my wonderful advisor, Professor Sally Benson for her guidance, support, understanding, and encouragement, both professionally and personally. I have learned so much and grown tremendously under her guidance. Her curiosity and drive inspire me to be a better researcher daily.

I would also like to thank my collaborators at the Stanford Office of Sustainability (Moira Zbella, Melissa Maigler, Fahmida Bangert, and Annabelle Bardenheier), Stanford Residential and Dining Enterprise (Jackie Bertoldo, Anuradha Chowdhary, and Sophie Egan), SIMAP NWG (Allison Leach, Stacey Baumgarn, Mary Ellen Mallia, Elizabeth Dukes, Sarah Messenger, Jonathan Lantz-Trissel, Sibeal McCourt, and Jim Galloway), and at ExxonMobil (Rosendo Cruz, Nagore Sabio, and Vijay Swarup). Without these individuals providing data, input, discussion, and support, none of this work would have been possible.

To my 5E sibs who have helped me in innumerable ways, Chris, Dom, Eileen, Erika, Helen, Houssam, Jake, Leon, Luke, Radha – thank you for the laughs, memories, and support. Thank you as well to Sunny, Natalie, and Edwin for keeping me sane and always being there for me. Thank you to Mateus for all the coding help and thank you to Sarah and Henry for encouraging me, checking in on me, and always knowing the right thing to say. To River, Bear, Bubba, and Pumpkin thanks for all the hugs and love.

Finally, I would like to thank my family, Tania and Sophia, for everything you have done for me and all the love and support you have given along the way. I love you so much.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background

New effects of climate change are being felt around the world every day. Thus, the urgency with which mitigation efforts must be implemented only grows. In the last decade, there has been significant research conducted related to the quantification of some types of emissions, such as those from power generation and from the use of internal combustion engine vehicles. However, there are large sectors of the economy for which calculating and tracking emissions have not historically been prioritized. Emissions must be quantified in order to be mitigated, thus no large source of emissions can be neglected if climate change targets are to be met. These historically neglected emissions have only grown in cultural and scientific relevance and the pressure to develop methodologies associated with these emissions has also grown as climate change and sustainability has steadily increased in importance in the public's perception [1].

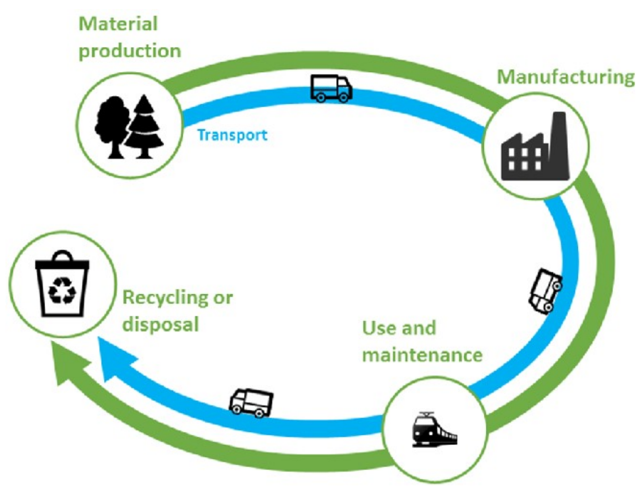


Figure 1. Life-cycle phases for a produced good [2].

A majority of the emissions associated with a purchased good are the greenhouse gas and other pollutants that are emitted during the life cycle of that good. These emissions can be determined through a process called a life cycle analysis (LCA). LCA analyzes emissions from making the product, using it, and eventually recycling or disposing of it, as can be seen in Figure 1. Each step in the life cycle can have emissions associated with it, however, not all products have emissions at each phase and products will have different distributions of impact at each stage. For example, a product such as gasoline will have a disproportionate amount of emissions during the use phase since this product is combusted, while a product that is not combusted, such as a table, will likely have a larger portion of its life cycle emissions attributable to the manufacturing phase as there are no use-phase emissions for a table.

Individuals are often more accustomed to considering the emissions associated with products that have a disproportionate share of emissions during the use phase, as this is the phase most visible to consumers. However, it is important to bring to light that there are emissions associated with all the phases in order to quantify and eliminate them. There have been some improvements in understanding and acceptance by the general population of this concept, for example with the rise of vegetarianism and veganism as mechanisms to reduce food-based emissions. However, there is still much work to be done to bring this understanding to needed levels across all demographics and economy sectors.

Figure 2. Emission scopes as defined by the Greenhouse Gas Protocol [3].

There are three categories of emissions based on their source as illustrated in Figure 2. Scope 1 emissions are related to all emissions from onsite emissions from burning of fuels and other onsite reactions within the boundaries of an entity (e.g., any natural gas water heaters, vehicles). An entity within the context of this work is any unit that is calculating its emissions (for example, a company, university, person, country, etc.). Scope 2 emissions include offsite emissions such as from the electricity that is purchased from the grid, but also includes emissions from other shared large-scale systems, such as steam, if available for purchase from a district steam heating system. Scope 3 emissions are also offsite emissions related to all other upstream and downstream sources from the entity. These include 15 categories which are listed in Figure 2, as defined by the Greenhouse Gas Protocol, the developers of the Scope 1, 2, 3 terminology. Not all 15 categories are relevant for every entity. For example, the processing of sold products would not apply to a university's Scope 3 emissions because a university does not produce sold products. The types of emissions that the public, as well as most companies and universities have focused on in terms of reporting and reductions to date are Scope 1 and 2 emissions. However, there is growing interest in quantifying and setting targets for reducing Scope 3 emissions.

Purchased items can have wildly different emissions associated with their production due to varying levels of complexity in the products themselves, the wide range of types and amounts of primary materials needed, and types of use. By studying an entity's purchased goods and the emissions associated with their production, transportation, and end-of-life, that entity can then understand what categories of purchases contribute the most to the total emissions from purchased goods and take steps to mitigate these emissions. This can take many forms, including finding lower carbon intensive vendors for product categories that contribute significantly, determining if purchasing the historical quantities of these items are necessary, or even implementing refurbishment or recycling programs for items if applicable. Although Scope 3 emissions are considered indirect, they often make up a majority of emissions for many entities, and until every entity producing goods is actively keeping track of and attempting to mitigate its Scope 1 and Scope 2 emissions, it will be important for large entities with purchasing power and leverage to put pressure on the markets when trying to minimize their Scope 3 emissions in order to attain the needed level of reporting and responsibility for emissions across the board.

It is also important to remember that Scope 3 emissions are indirect emissions, and therefore they are another entity's Scope 1 and Scope 2 emissions. Broad adoption of Scope 1 and 2 tracking for all entities will be necessary to minimize and mitigate climate change impacts in the future. A labeling system similar to the nutritional labels that exist for food products would ideally be created and standardized to easily pass on the emissions information from upstream for each and every product. This future is still distant, and until this is a reality, large entities that care about emissions and have made targets will need to put pressure on their supply chains to start accounting for and reducing their Scope 1 and 2 emissions such that the entities can claim reductions in Scope 3 emissions. By creating these pressures in the market and by making the emissions data readily available, entities and consumers will be able to make more informed decisions and the entire economy will move to be more environmentally conscious and emissions reduction will be attained at large scale.

1.2 Methods for Quantifying Scope 3 Emissions

One current methodology for determining the embodied emissions associated with any item is to take a representative quantity, such as weight/volume (weight-based), or cost (spend-based), and multiply that value by an appropriate emissions factor. These emissions factors are often determined through LCAs for items or product categories which analyze each process or phase that occurs during the life of that item and determine the emissions associated with each process or phase. By adding up each process or phase emissions, the life-cycle emissions can be determined.

However, currently there is uncertainty when using these emissions factors and a gap in understanding the accuracy, precision, and specificity of emissions factors. It is not possible to get an item-specific emission factor in most cases without conducting a time intensive analysis of exactly how that one product was produced. Instead, studies have been conducted to determine the average or representative emissions factor for a broader category to obtain a representative value. Thus, to determine the emissions associated with purchased goods, purchases must be grouped into categories for which emissions factors exist. Then, appropriate emissions factors must be used to determine the overall emissions from those purchases. Depending on the category being studied, there can be a variety of emissions factors which include or exclude certain phases of the life cycle, which are spend- or weight-based, which represent the emissions associated with an item produced in a specific region of the world, etc.

There are a number of options for where to obtain emissions factors from. Many research publications exist which have conducted an LCA for a specific item under a strict set of assumptions, and other publications exist which have done literature reviews which attempt to aggregate many LCAs [4,5]. Some tools exist which are developed by companies to help others calculate emissions, and there are also many governmental agencies which provide emissions factors for certain categories of products (e.g. EPA, DEFRA). Relevant examples from these categories are further detailed in Chapter 2: Literature Review. These tools and sources still require purchases to be grouped into the categories for which each of these have emissions factors for, and each of these has a different set of categories with varying amounts of specificity. Using these multiple tools and sources introduces uncertainty with regards to the emissions calculation results in large part due to different LCA assumptions and the different group categorizations needed for each source, in addition to differences in data vintages between tools.

This work aims to develop tools and methodologies to ease and create a standardized practice for Scope 3 emissions calculations for entities, as well as to quantify uncertainties inherent in using different emissions factors. To achieve this goal, food purchasing data was used as a case study to develop tools and answer questions that will be relevant for the quantification of emissions from all categories of purchased goods. Food data was selected as the case study for this research because of the ease of the availability of data both in terms of total weight purchased as well as total spend.

This work will be relevant across many industries as many entities, including large companies and universities, have recently started to set targets for Scope 1, 2, and 3 emissions. In order to mitigate these emissions, they must be quantified, and this work is an important step towards that goal. With the rise in popularity and customer demands for companies to set emissions targets, work to develop standardized methodologies to estimate Scope 3 emissions will only continue to grow in importance.

1.3 Stanford Food Purchasing Case Study

In order to answer these questions related to Scope 3 methodologies and to develop the necessary tools, a real-world data set was needed. Stanford purchasing was selected to be the source of the data set due in part to ease of data availability for the work, but also because universities are on the forefront of entities creating Scope 3 commitments. Food purchasing was selected as a subset of purchasing data because it provides the ability to quantify the difference in estimates when using weight-based versus spend-based emissions factors because it is commonplace to consider both aspects of food. For example, people are used to thinking about food both in terms of 6 pounds of beef but also $30 of beef, whereas it is not common to think about 50 pounds of chair. This characteristic of food has allowed for systems which make keeping track of both spend and weight of food easier.

This work is generally relevant not only due to the recent increase in public interest in climate change and greenhouse gas emissions, but also because it is particularly pertinent within the context of scholarship occurring at Stanford due to the announcement made by the Stanford Faculty Senate in November 2020. This announcement takes the first step towards prioritizing Scope 3 emissions calculations by approving a resolution which set a target for net-zero greenhouse gas emissions from Scope 3 emissions by 2040 [6]. To meet this goal, Stanford will need to quantify all categories in its Scope 3 emissions, and this work provides the first step in doing so for purchased goods. The Faculty Senate committed to net-zero Scope 3 emissions and not a percent reduction, thus the true magnitude of the emissions is important, and therefore, careful development of the methodologies that Stanford employs in determining its Scope 3 emissions is crucial.

Within the categories present in purchased goods, food can make up a large portion of a university's Scope 3 emissions due to the sheer amount of food that must be purchased to keep a campus fed. In order to validate the tools and methodologies developed, a partnership was formed with the Sustainability Indicator Management & Analysis Platform (SIMAP) Nitrogen Working Group [7], through which the food purchasing data sets of seven other universities was acquired and the tools and methodologies were applied to these data sets and modified for improved performance as needed in order to validate the tool. All entities, regardless of their sector, purchase goods, thus any lessons learned or new methodological processes that apply for purchased goods within a university's context will be relevant for a wide range of entities when they begin the Scope 3 emissions quantification process.

1.4 Motivation

Preliminary estimates indicate that Scope 3 emissions are expected to represent a majority of an entity's emissions [8], and thus must be quantified such that they can be incorporated into decision making efforts to mitigate the effects of climate change. Many companies have recently made sweeping Scope 3 emissions commitments, including Microsoft [9] and Amazon [10]. As more entities join in determining their Scope 3 emissions, it will become more important for there to be standard methodologies and tools to ease these calculations. As this field is still in relative infancy, there is much research and development of these methodologies and tools that must be done.

In order to determine Scope 3 emissions from purchased food, the thousands of different food items that are purchased yearly by an entity must be placed into categories for which emissions factors exist. This process can be extremely time-intensive to do by hand, yet that is how many entities have approached this task previously. Thus, creating an automated, standardized way to categorize purchases can significantly decrease the activation energy associated with analyzing these purchased food data sets and be a gateway for widespread adoption of quantification of Scope 3 emissions from purchased food by a many entities. Starting with food purchasing can also be seen as a manageable starting point to quantify an entity's Scope 3 emissions. Since entities up to now were free to categorize their food purchases as they saw most appropriate, the same item purchased at different entities may have been categorized differently between them. Thus, using a common, standardized, categorization tool enables entities to conduct fair comparisons between the calculated results.

In addition to the discussion on the development of methodologies and tools to ease Scope 3 calculations from purchased goods, this work also explores the impacts and attempts to quantify the uncertainty associated with using weight-based versus spend-based emissions factors. Emissions factors are a required input when calculating emissions associated with any item if a full LCA cannot be conducted, thus a determination of the impacts associated with using different emissions factors have applications far beyond purchased goods emissions calculations. The automated categorization tool is easily adaptable to have applications beyond the sole categorization of food items.

1.5 Objectives

This work aims to answer the following questions relevant to the determination of Scope 3 emissions from purchased goods:

1) How does the use of weight-based versus spend-based emissions factors (EF) affect the magnitude of emissions calculated for a data set?
2) How does using different sources for emissions factors within a same type of EF affect overall estimated emissions?
3) What is needed to standardize and ease calculation of Scope 3 emissions across countries and backgrounds such that the process can be automated and replicable results can be ensured?

To answer these questions, food purchases from Stanford in 2019 were used as a case study to develop methodologies for categorization of purchased items as well as to quantify uncertainties in calculated emissions when considering the same data but using different emissions factors. Emissions factors were collected from a variety of commercial and public tools and databases for emissions calculations and include both weight-based and spend-based factors to determine the possible range of quantified emissions from the data sets studied. In addition, this work aims to develop and leverage partnerships with other entities to validate and standardize methodologies and tools to ease the process for others to quantify their Scope 3 emissions in the future.

Chapter 2: Literature Review

Although research has previously been conducted with regards to portions of this work, there is not much identified research that combines spend- and weight-based analyses, comparisons of sets of emissions factors, and the analysis of a real-world data set. There has also not been significant research conducted with regards to categorization methodology for determining Scope 3 emissions from purchased goods and analyzing how different categorizations may impact the final emissions calculation. Many of the identified studies are more academic in nature and do not deal with the practical concerns associated with categorization of large numbers of purchases. Below are summarized the relevant literature for this research.

2.1 Life-Cycle Analysis (LCA)

The fundamental basis of most emissions factors is life-cycle analysis. As discussed in Chapter 1.1 Background, the assumptions and boundaries which are considered whenever an LCA is conducted are very important and impact significantly the total emissions calculated for an item. It is also important to be aware of the boundaries of LCAs when attempting to compare the results from using emissions factors derived from various LCAs. In the Poore et al. [5] paper, which will be further discussed in this section, significant efforts were made to harmonize assumptions before compiling aggregated data from many LCAs. The boundaries for all the LCAs used in this work vary, with some being cradle-to-grave, others being cradle-to-gate, and some including land use emissions or indirect supply chain emissions.

2.2 Weight-Based Emissions Factors

2.2.1 SIMAP Emissions Basis: Heller et al. 2014 [4]

This study conducted a meta-analysis of various emissions factors from the literature to develop representative emissions factors for different food types, and it focuses mostly on estimating the emissions of the average American diet and the impacts of food loss in the United States. In order to achieve this goal, the emissions for food groups were determined by calculating the average emission factors for each food category from the literature. These emissions factors are not intended to provide an exact result for each food type – instead, they are supposed to provide a reasonable range for expected emissions values. This study includes the emissions associated with the production of food that is wasted at both the retail and consumer levels and considers the possible impacts of shifts in dietary habits, but does not include the emissions associated with the disposal of the wasted food. These losses are estimated to be approximately 1.3 billion metric tons per year of food in 2014, which represents almost 1/3rd of food produced for human consumption. Plate waste by consumers in included in the analysis but the paper does not include impacts of waste from retailers due to appearance or issues such as pest infestations, mold, and other spoilage factors. The emissions factors were compiled from a literature review and the average of the reported values were used in the calculations presented here. In addition, the data were collected from around the world because US-specific data were limited. This paper determines that the losses associated with food contribute an average of 1.4 kg CO_2-eq per capita per day in the US, which is approximately equivalent to emissions from 33 million passenger vehicles per year. By using the minimum and maximum emissions factors for each food category, the paper also found that the possible range of emissions associated with food consumption per capita in the US is between 2.5 and 9.2 kg CO_2-eq per day.

In addition to calculating the emissions associated with food loss and consumption, this study also noted several other interesting findings. For example, it was found that while beef makes up approximately 4% of the retail food supply by weight, it makes up 36% of the diet-related emissions. Direct emissions from agriculture were estimated to represent between 10 and 12 % of global emissions, but this range increased to 17-32% when indirect emissions, such as fertilizers, chemical production, fuel use, and agriculturally induced land-use, were included in the calculations. In addition, when compared to processing and other phases, agricultural production dominates the LCA emissions associated with the food commodity.

This study also highlights several important topics that are relevant to the rest of this work. For example, the analysis found that Economic Input Output models often produce larger emissions estimates by virtue of including the impact of indirect emissions in their calculations—emissions which are usually excluded in process level LCA studies, such as emissions associated with producing capital equipment. Finally, it is mentioned that comparisons between LCAs are difficult due to differences in system boundaries, allocation procedures, and geographic idiosyncrasies that include electricity grid mix and typical production methods. Finally, one other comparison was done with the UK-specific Hoolohan et al. (2013) emissions factors which found that the total food emissions was 17% greater than the emissions calculated using the average Heller et al. emissions factors.

2.2.1.1 Heller et al. 2014 Edits for Use in SIMAP [4]

SIMAP is a tool developed by the University of New Hampshire that aims to help universities calculate their Scope 1 and 2 emissions. The tool includes food purchasing in its metrics for analysis due to SIMAP's focus on nitrogen foot-printing. SIMAP has developed an emissions calculator that measures these emissions that is used by many universities across the US. The emissions factors used by SIMAP for food purchasing are based on the Heller et al. 2014 emissions factors. These factors have been available for a longer period and existed when SIMAP added food purchasing emissions to its calculator; however, some edits were put in place in order to make the tool more flexible and capable of handling more specific data sets. For example, users can tag purchases as organic and local, and the Heller et. al 2014 emissions factors are edited to reflect these impacts. If a food purchase is tagged as local, the Heller et al. values are modified to reduce the transportation emissions. It is assumed that transportation from producer to retail accounts for approximately 4% of greenhouse gas emissions from food [11]. This means that for local foods, the Heller et al. 2014 values are reduced by 3% to account for a 75% reduction in transported miles.

2.2.2 US-Based Food Impacts Literature Review- Heller et al. 2018

The same authors as the Heller et al. 2014 paper conducted a new study in 2018 to understand the environmental impact of individual self-selected diets in the US by performing a 1-day dietary recall of 16,800 Americans. To determine the emissions of these diets, Heller et al. updated their food emissions factors from the limited 2014 values by conducting a more exhaustive literature review and including updated data. This was in part possible due to a larger number of food LCAs conducted between 2014 and 2018 due to increased interest in the field. However, it is important to note that there are still many gaps in the food LCA literature, including analyses related to nuts, legumes, and meat substitutes, and that the current literature is biased towards European studies. In addition, this study used a more refined food type characterization as compared to the 2014 study: the new emissions factors were calculated by considering 13 different categories, and an average of the literature review for each of these categories was used. The categories were made up of 841 different LCA data points that were organized into 172 foods. These data were combined to create the database of Food Impacts on the Environment for Linking to Diets (dataFIELD).

To be included in this literature review, papers had to be publicly available, in English, published between 2005 and 2016, and LCAs for one or more food products. The functional unit considered was one kilogram of food, except for meat and fish/seafood, for which the functional unit was adjusted to represent one kilogram of edible boneless weight. LCAs for heated greenhouse vegetable production and for beef from dairy herds were not included in dataFIELD because information on the market share of these production methods was either not available or unreliable. In addition, for juices, vinegar, and maple syrup, additional sources were used to develop valid emissions estimates that possibly were not in English or primary source documents. The LCA boundaries considered for the emissions factors in this study are cradle-to-farm gate. Because most of these commodity foods are used as ingredients and are further processed before being eaten, considering the phases after farm gate would not reflect the actual impacts of the food consumed. However, flours, refined sugars, vegetable oils, and other foods that require processing before becoming ingredients had the emissions associated with these processing steps included in their LCA boundaries. More information on this methodology is available in the supplementary information for the Heller et al. paper. The food losses included in the LCA were calculated using the USDA's loss-adjusted food availability and included losses from retail sites such as supermarkets and restaurants as well as consumer losses from cooking losses and uneaten food. Importantly, this study included an emissions factor for beverages which up to this point were not often identified as a separate food group but ended up being the third largest food group in terms of emissions in this work. This is particularly important because the packaging and use phase portions of the LCA phases are not included in the emissions factors, both of which are high for beverages due to, for example, packaging for soda, and brewing coffee or heating water for other drinks.

The study, unlike previous studies, looked at the impacts of the variability across the LCA studies on the mean self-selected diet emissions and found it caused a $\pm$19% range to this value. This was done by calculating upper and lower bounds for the impacts of the food types and carrying these through to the diet-level impact calculations. The variability in the LCA studies is believed to be due to differences in food production locations (and thus climatic conditions), production practices, and LCA allocation methods for co-products, all characteristics which cannot currently be evaluated for individual diet decisions. Variability is also introduced because many foods referenced in the diet surveys are complex recipes, such as lasagna, or are items which have not yet been studied in life cycle assessment literature, such as blackberries. Thus, appropriate proxies for emissions factors must be used for these items.

In addition, the study concluded that the meat categories contributed 57% of the average dietary greenhouse gas emissions. Within the meats categories, for the total population, 80.6% of the emissions were due to beef consumption, 9.5% from poultry consumption, and 8.5% from pork consumption. Surprisingly, this study found that the 20% of diets that have the largest carbon footprint account for 45.5% of the total diet related emissions in the United States. The greenhouse gas emissions of the 5th quintile were almost 8 times the emissions of the 1st quintile and were 3 times that of the middle quintile. This is in part due to the fact that the 5th quintile consumes on average 2.25 times the number of kilocalories as the 1st quintile. Thus, if the 5th quintile shifted such that their emissions were equivalent to the mean emissions impact of diets, this would be equivalent to reducing emissions in the US by 0.27 million metric tons of CO_2-eq per day or the equivalent of eliminating 661 million average passenger vehicle miles. Thus, behavior change offers an important opportunity for emissions reductions in the US. This study also shows that foods for which LCAs do not currently exist do not contribute significantly to the overall emissions from food consumption in the US as long as appropriate proxies are selected from available literature.

Finally, in order to estimate the emissions associated with the LCA boundaries not considered in the main study, the US EPA EEIO model was used and the estimated contribution of the disregarded phases was 15% of the total cradle-to-processor gate emissions. More information on this methodology can be found in the supporting information for this study. However, these EEIO models consider the food and agricultural sectors in aggregate, thus, they cannot be applied evenly across different food types or for specific diets and are only applicable to the mean.

It is important to consider that these diet-level estimates were conducted using self-reported diets, which usually understate actual food consumption, and that since the boundaries for the study are cradle-to-farm gate, these emissions values should be considered underestimates of the total emissions associated with food consumed in the US.

2.2.3 Global Food Production Impacts Literature Review - Poore et al.[5]

This study evaluates the impacts of food production on five different environmental factors (land use, freshwater withdrawals, greenhouse gas emissions, acidification, and eutrophication) using data from 38,700 farms, and 1,600 processors, packing types, and retailers. It found that the total impact can vary up to 50 times among producers of the same food product. This range of emissions is due to the fact that there are millions of producers of food and a large range of farming techniques, all of which have different impacts associated with them. Different methodologies were accounted for between the LCAs included in the analysis, and were harmonized and reconciled among the global data that was gathered. The data was gathered from 1,530 studies, and 11 criteria were used to standardize the methodology between the LCAs. This led to the incorporation of 570 of these articles, which had a median reference year of 2010. The data from the studies were available because of a recent rapid expansion in literature associated with LCAs, which are done in part by surveying producers around the world. The data included in the analysis represented more than 38,000 farms in 119 countries and 40 products, which make up about 90% of global protein and calorie consumption, and were used to produce emissions factors for 59 food categories. They estimate that the food supply chain emits 13.7 billion metric tons of CO_2-eq per year, which makes up 26% of anthropogenic emissions. The breakdown of the emissions per food category as well as the contributions from each life cycle phase to each food's total emissions can be found in Appendix I: Poore emissions breakdown. The life cycle analysis did not include emissions associated with consumer losses because of a lack of available data.

Major conclusions from this study include the following. Due to the large variability in emissions between farms producing the same products, which can range from 2 to 130 times larger between the 90th percentile to the 10th percentile emissions, there is large potential for reduction in emissions if the largest emissions methods are able to switch to lower emissions methods. For example, across all the food categories, 25% of the products contribute on average 53% of the emissions from that food group. There are also relevant conclusions made within food groups. Methane emissions from flooded rice, enteric methane from ruminants, and concentrated feed for pigs and poultry make up 30% of emissions from food globally. Within this study, nuts are considered to have the potential for negative emissions due to the fact that they can temporarily sequester carbon when they replace croplands or pastures by sequestering carbon in the trees themselves. For food categories such as beef, the losses associated with distribution and retail can make up 12 to 15% of the total emissions for the category. Overall, the largest LCA phase contributor was the farm stage, which was found to account for 61% of the life cycle emissions for food. In addition, contrary to commonly held popular belief, the emissions associated with transportation were not a major contributor to overall life cycle emissions as only 6% of the emissions came from transportation on average.

In order to mitigate emissions, several recommendations were made: to develop systems for producers to monitor their own impacts; to meet environmental targets by choosing less carbon intensive farming methods; and to communicate impacts to consumers in order to encourage consumers to change dietary habits. Another option for lowering emissions is to switch from planting monocultures to diversified crops, which also improve degraded pastures and require less land use. Finally, in the US, the average meat consumption per capita is as much as three times the global average, thus, changes to dietary habits have a large potential for emissions reductions. By providing processors, retailers, and consumers with emissions values for the products, providers could encourage waste reduction and dietary changes where it matters most.

2.2.3.1 Poore et al. Edits for Use in World Resources Institute Cool Food Pledge [12,13]

Edits were made to the Poore et al. emissions factors in order to develop a US-specific and Europe-specific set of emissions factors to be used in the Cool Food Pledge emissions calculator developed by the World Resources Institute. The Cool Food Pledge aims to help entities reduce their climate impacts from food by encouraging shifts to plant-rich diets, and by providing support and methodologies to commit to and reach these science-based targets. To align with the IPCC goals, entities participating in the Cool Food Pledge must commit to reducing food emissions by 25% by 2030 as compared to food emissions in 2015. Entities are required to report animal-based food purchases (ruminant meats, other meats, dairy, fish and seafood, legumes, grains/cereals (except rice), and plant-based milk substitutes), which were found to represent more than 80% of the total emissions from food, including supply chain emissions and carbon opportunity costs. Entities are not required but are encouraged to report purchases of fruits, vegetables, roots and tubers, sugars and sweeteners, vegetable oils, alcohol, and stimulants.

The Cool Food Pledge considers agricultural supply chain emissions, land use, annualized carbon opportunity costs, and total calories in their calculations and commitments. The life cycle stages considered are cradle-to-point of purchase as well as carbon opportunity costs (COC), which are defined as the amount of carbon that could be stored if production of that food type was reduced and the land used by that food was restored to its native vegetation. Most LCAs do not translate the agricultural land use into COCs, and the data for COC were extracted from Searchinger et al. [14] The COCs are a way to interpret the pressure that entities' food purchases cause on forests and other natural ecosystems and their related climate impacts.

The emissions factors were developed based on the Poore et al. emissions factors, but the local (Europe and US) values were determined by using a weighted mean, by calculating the share of the national agricultural production that each analyzed LCA data point represented and the share each country represented for the global values. The losses during harvesting, transportation, processing, and packaging were included, but the analysis did not include retail-level emissions or losses, or post-retail stages such as consumer waste. In addition, the Cool Food Pledge calculator converts purchases of meats and fish into boneless equivalent weights for which the emissions factors were developed. It was found that the annualized COCs are larger than supply chain emissions for almost all the food categories. However, because the COCs have some elements associated with avoided emissions, they are not included in the standard Scope 3 GHG Inventory and must be reported separately from the supply chain emissions.

Although the Cool Food Pledge documentation does include a comparison of the cradle-to-farm gate emissions factors for several food categories from several papers [5,15,16], there is no discussion as to how these differences in emissions factors impact the analysis of a real-world food purchasing data set. This variation in emissions factors is explained by differences in production systems in different regions. These production differences include decisions related to production strategies, such as how to boost yields, reducing direct agricultural production emissions by improving input technologies, improving management, and how carbon is sequestered in soils in different regions. Because of the inclusion of the COCs, the Cool Food Pledge analysis shows that the benefits associated with shifting to more plant-based diets are larger than previously reported.

2.2.4 Land Use Change Emissions

There are various different approaches to accounting for land use change emissions within the LCA community. However, it is estimated that in 2010, land use change, including vegetation clearing and soil plowing, accounted for approximately 10% of anthropogenic GHG emissions, most of which was caused by agricultural expansion into forests and other natural ecosystems [13]. It is also estimated that deforestation and other land-use changes have contributed between 25-33% of the total anthropogenic emissions since 1750 [17]. Typical LCAs often only consider land use in terms of hectares without converting to carbon emissions, and those that do often only consider land use carbon costs if the food produced directly cleared new land, or only consider costs for meat or milk because these crops are expanding into new lands [14]. It is also important to consider that different production practices can significantly impact the land-use change emissions of a particular crop in a particular area. For example, because organically grown foods have yields that can be between 19-25% lower than non-organic production methods, the land-use carbon opportunity costs can be 23-33% higher [13], and the total land needed to grow the same amount of an organic crop would need to be larger than for the non-organic same crop.

2.3 Spend-Based Emissions Factors

2.3.1 Introduction to Economic Input-Output Tables [18]

Economic Input-Output (EIO) models are matrix representations of the monetary flows between the different industrial sectors and thus show what goods and services are consumed by other industries. Each row and column in the matrix representation is a single industrial sector. The intersections of the matrix show the monetary amount that is output by one industry (row) as an input for another (column). By utilizing linear algebra techniques, EIO models can be used to determine the direct effects (those that occur directly from the sector of interest to all other sectors), indirect effects (those that occur between all other sectors in order to meet the demand in those sectors needed to fulfill the demand of the sector of interest), and total effects of changes to the economy. Many nations create EIO models for their own economies, and each model has varying degrees of detail and data update frequency. The data to produce the EIO models comes from surveys sent to a sample of all operating facilities.

These traditional EIO models can be expanded upon to determine not only the monetary flows between industries but also other flows, including energy, electricity, greenhouse gas emissions, and other environmental impacts. This idea of combining LCAs and EIOs into an Economic Input-Output Life Cycle Assessment (EIO-LCA) was developed by Wassily Leontief in the 1930s. The EIO-LCA is one mechanism of performing an LCA, and it is used to estimate the total emissions throughout a supply chain for a specific country. The EIO-LCA is effectively a traditional EIO matrix with an additional column representing the environment. The value of each row within the column represents the environmental impact of one industrial sector to the environment. Because both direct and indirect interactions between the industries in an economy are included, the boundary of the LCA is considered to be broad and inclusive. In addition, EIO models include monetary transactions within a sector, thus any circularity effects of recycling or other circularity influences are included in these LCAs. EIO models are particularly useful when considering economies or large systems because the traditional LCA methods for single products are not scalable or practical for analysis of the entire economy.

Several matrices are needed for an EIO model, and the data held within these matrices are usually in units of dollars (or other currency) such that all the items in an economy are in comparable units. A Matrix, often called A, represents the direct effects where the rows of A represent the amount of output from industry i which is needed to produce one dollar's worth of output from industry j. If a Y vector represents the final demand of goods from each sector in an economy is defined, to determine the direct requirements from all sectors to meet the demand defined in Y, a new vector X can be defined as $X = (1 - A)^{-1}$ where X includes both direct and indirect transactions between sectors. Now, another vector can be defined as $R_i = \frac{total\ external\ output}{X_i}$ where R_i is he impact in sector i and X_i is the total dollar output for sector i. This is very similar conceptually to developing a carbon intensity of sector i. Finally, in order to determine the environmental impacts from meeting the demand in Y, a new vector of total external outputs can be defined as B_i by multiplying the total economic output at each stage by the respective impact.

There are several caveats that should be made clear for consideration when using EIO-LCA models. These models do not include the use phase and the end-of-life phases, but there are some additional analyses that can be done using the EIO-LCA model to help understand the emissions of these phases as well. The data for the environmental impacts of sectors are often gathered using the North American industry classification system (NAICS) or other generic categorization systems (for example, the USDA uses crop types to categorize farms), but these categories do not exactly line up with the economic sectors defined within the EIO models. Thus, harmonization by using weighted averages or external sources must be conducted between the environmental impact categories and the EIO categories. In addition, these EIO models do not explicitly account for imports and exports, which can make up a significant amount of any economy's transactions. Imports are handled by assuming that they have the same production characteristics as the comparable product made in the EIO model country.

The data used in the EIO models also have several important characteristics which should be considered when using these models. The data can be several years old, and because the amount that these data change over time varies widely depending on the industry sector, care should be taken when using a model to replicate the current conditions. In particular, environmental impacts can change over time due to process improvements, such as increases in efficiency, or regulatory improvements, such as setting maximum levels of pollutants. In addition, all the data that is used in the EIO-LCA models are compiled from surveys which are submitted by players within the industries to their respective governments for national statistical purposes. Thus, there is an inherent uncertainty in the sampling, response rate, and missing or incomplete data which contribute to uncertainty in the EIO-LCA model results. Finally, the economic data that forms the basis of the EIO-LCA models represent the producer prices. These are the prices that a producer needs for its goods and services which include taxes and any applicable subsidies. In other words, it represents the costs of buying all the needed materials, running the facilities, paying workers, etc. The purchaser price, on the other hand, includes the producer price and any other costs associated with transporting the product to the point of sale and any wholesale or retail trade margins. Thus, for many physical goods, there can be a large disparity between the producer and purchaser prices.

2.3.2 Environmental Protection Agency (EPA) US Environmental Economic Input Output (EEIO) [19]

The US EEIO tool was developed by the EPA to provide an economic based LCA tool that was transparent, reproducible, open, and up to date. Thus, all files associated with the US EEIO are publicly available, unlike most other EEIO tools. The tool was developed in part to help organizations identify environmental impact hotspots in their supply chains. The US EEIO models use economic input output tables as discussed in the previous section. These EIO tables are updated at the detail level every 5 years and include information for almost 400 industry sectors [18]. The data used for these economic input output tables is the same as are provided by the US Bureau of Economic Analysis (BEA) to determine the yearly US Gross Domestic Product. US EEIO v1.1 combines data on monetary flows between 389 industry sectors and environmental data for the same sectors to produce a life cycle model of 385 US goods and services with data from 2007. The data from the BEA is in the form of Make tables and Use tables, which are a slight variant of the standard IO tables developed by Stone et al in the 1960s in order to account for secondary products.

Cradle-to-gate environmental impacts considered include land, water, energy, and mineral usage; emissions of greenhouse gases and criteria air pollutants; nutrient releases to water; and toxic releases to water, air and soil. The greenhouse gas emissions and sinks data come from the 2013 US GHG Inventory conducted by the US EPA and were collected into satellite tables. Assumptions, primarily that a single impact data point applies equally for all technologies in the sector, had to be made for some impact factors including greenhouse gas emissions for sectors that have mixed technologies. One difficulty associated with developing the satellite tables is that there are a variety of data sources, including both surveys and model simulations, that have varying degrees of quality and uncertainty associated with them. GHG emissions are considered to have average reliability within the model. Most of the data are gathered from individual facility reports which are compiled in the National Emissions Inventory, Toxic Release Inventory, and Discharge Monitoring Report. All the data are either discrete sources which report across the US or are based on average US conditions. The model makes two important assumptions with regard to how secondary products are produced: first is that all commodities which are produced by an industry have the same input structure, and second is that a given commodity has the same input structure regardless of where it is produced.

There are several matrices from the USEEIO v1.1 model which are available for use in analyses of environmental impacts:

- A Matrix (technology matrix, or direct requirements matrix): commodity inputs are the rows of the matrix, and the commodity outputs are the columns of the matrix. The data were derived from the 2007 Input-Output tables developed by the US Bureau of Economic Analysis. This matrix is calculated by the IO model builder by multiplying the commodity-by-industry direct requirements gathered from the Use table by the market share industry-by-commodity matrix gathered from the Make table after the scrap adjustment procedure.

- B Matrix (Environmental matrix): provides data on emissions or resource by commodity. Direct emissions and resource use needed to produce one 2013$ of a commodity are the rows of the matrix and the commodities are the columns of the matrix.

- C Matrix (Characterization Factors Matrix): Indicators make up the rows of the matrix for a given emission or resource, which are the columns of the matrix.

- D Matrix (Direct Environmental Impacts): indicators are the rows of the matrix and commodities are the columns of the matrix. This matrix provides data on the direct environmental impacts associated with producing one 2013$ worth of a commodity. This matrix is calculated by multiplying the C and B matrices.
- L Matrix (Leontif Inverse of the A Matrix): the total direct and indirect dollars of a commodity needed to produce one dollar of a commodity are the rows of the matrix and the commodities are the columns of the matrix. The matrix is calculated by calculating $(I-A)^{-1}$.
- LCI Matrix (Life Cycle Inventory Results): This matrix has the same rows and columns as the B Matrix, but the data represent both the direct and indirect resources or emissions associated with producing one 2013$ worth of a commodity. This matrix is calculated using the following formula: $B*(1-A)^{-1}$.
- U Matrix (Life Cycle Impact Assessment Results): This matrix has the same rows and columns as the D Matrix, but the data represent both direct and indirect environmental impacts associated with producing one 2013$ worth of a commodity. This matrix is calculated using the following formula: $CB(1-A)^{-1}$.

Yang et al.[20] compared, as much as possible, US EEIO v1.1 to CEDA v4.6 and found that the sources for greenhouse gas emissions between the two were largely the same. A true quantitative comparison of results was not conducted because the two tools are based on different reference years. Yang et al. found that at the time of development, US EEIO included more current data, had more extensive consideration of impacts and resources, was more detailed in its interpretation, and included formal data quality evaluations when compared to other EEIO tools available at the time. Yang et al. also compared openIO 1.4 (another EEIO tool) and CEDA v4.6 and found that it is not possible to determine which model is more accurate. Even though the fundamental environmental data sources for greenhouse gas emissions used to create the satellite tables were the same, there were significant differences in the results. Yang hypothesizes that this could be due to differences in allocation methodologies of emissions to different sectors or to differences in the use of primary or secondary sources. Due to a lack of documentation, it was not possible to ascertain exactly where the differences were in the creation of the satellite tables for both tools. Yang et al. also provided a framework for conducting a hotspot analysis for goods and services by using the US EEIO tool and conducted an analysis that showed the structure of the US economy is relatively slow to change. Because of this characteristic, EEIO models can be appropriate for use even when the economic and environmental data years do not necessarily match the year of the data being analyzed, as long as rapidly changing sectors are not the focus of analysis, and the prices are adjusted to the EIO year.

At the time of the release of US EEIO v1.1, it was not common to evaluate uncertainty within EEIO models because there were no established methods for doing so. Since the development, it has become more common, but a formal structure still does not exist. Future iterations of the US EEIO model hope to include use and end-of-life emissions for commodities used by US households as well as regional-specific versions of the model for more specific applications than the broad US economy.

2.3.3 Supply Chain Greenhouse gas Emissions Factors (SEF) [21]

The Supply Chain Greenhouse gas Emissions Factors (SEFs) were developed to provide a comprehensive set of supply chain emissions factors that cover all goods and services within the US economy. Emissions factors were developed for supply chain without margins (cradle-to-factory gate);

supply chain margin emissions (factory gate-to-shelf) which include emissions from transportation, storage and selling (both wholesale and retail) adjustments for price markups; and supply chain emissions with margins (cradle-to-shelf). None of the factors include emissions from the use or end-of-life phases and they do not include biogenic CO_2 emissions nor emissions derived from biomass. The emissions factors are in terms of kilograms of greenhouse gas emissions per 2018$ purchaser price and procedures which are available in the original appendix were done in order to align the direct emissions factors with the dollar year of the IO table data.

SEFs use the US EEIO v2.0 as a basis for the economic and environmental impact flows. US EEIO v2.0 uses 2012 IO tables from the BEA, the most recently available data set. These EIO tables were extended by researchers from the Yale University Center for Industrial Ecology and then further refined for the development of the SEFs. The two main differences between the SEFs and the original US EEIO v1.1 discussed above is that the emissions associated with the supply chain and marginal emissions are broken out and that the commodity prices in the SEFs represent purchaser price whereas US EEIO v1.1 represent purchaser prices. Purchaser prices match spending data more closely and thus can be more accurate for emissions from spending analysis. Additionally, using purchaser prices reduces the uncertainty related to the margin emissions. In addition, the SEFs are available with greenhouse gas emissions specific data from 2010-2016; however, all of these years use the 2012 detail IO data. US EEIO v1.1 uses 2007 detail IO data and emissions factors from 2013 [22]. Like other EEIO models, the SEFs assume that imported commodities are produced in the same way they are in the US thus have the same emissions factors. In other words, international transportation emissions are not included in the margin emissions factors. The 2012 Personal Consumption Expenditures and Private Investment in Equipment bridge tables were used to develop the margin data for both the emissions factors as well as the price adjustments. These tables introduce some difficulties because these are also only updated every five years and their sector categories do not exactly match the sectors from the other data sources.

Supply chain emissions factors can be commodity-based or industry-based which introduce more uncertainty with regards to the total magnitude of emissions estimate depending on which of these two modes is used. Commodity-based and industry-based factors are useful for determining emissions associated with purchases of a specific commodity or from a given industry respectively. Differences arise between these two methodologies when the same commodity is produced by many industries (for example when a commodity is the primary product of one industry and the secondary product of another industry), and each of the industries have varying supply chains. Industry-based and commodity-based SEFs are usually very similar for most sectors except for the utility section which had the largest difference between the two emissions factors, where the industry-based value was significantly larger.

In addition, some comparisons were made when analyzing emissions at different levels of granularity available for the IO tables – at the detail level and the summary level. These refer to the number of sectors considered in the IO tables available from government agencies. The summary level is updated yearly but only includes 73 sectors whereas the detail level includes almost 400 sectors and is only updated every five years. The summary level factors have a larger aggregation error than the detail level factors because of how transactions are aggregated in the IO tables, but the detail level factors require the use of older economic data. The analysis was conducted by using the US EEIO models with the 2016 greenhouse gas emissions data and were converted to 2018$ producer price. The summary level methodology produces total commodity emissions equal to those reported by the National GHG Industry Attribution model, but the detail SEFs had a small difference between the sum and the overall emissions of less than -6% for both CO_2 and N_2O emissions in 2012. The total difference ranged from -12 to 7% when comparing the SEF data to the annual national reported values for the years 2010-2016, which shows there in inherent uncertainty with regard to the results estimated using these tales. These differences arise due to the use of the 2012 IO tables which are the most recently available data for economic transactions between the sectors/commodities. Overall, for CO_2 supply chain emissions factors, the summary level values were found to be smaller than the median detail level values. Specifically for food, the detail level factors for CO_2 emissions had a range of 0.17-0.77 kg/$ with a median of 0.48 kg/$ whereas the summary level factor was 0.42 kg/$. Both the summary level and detail level methodologies provide an average of the commodity or industry performances in the US. Thus, differences in production technologies and practices, environmental controls, scales of production, or locations in the US are not captured in these SEFs.

In conclusion, for most commodities and industries, the indirect emissions are larger than the direct emissions, thus it is important to include the supply chain emissions in emissions calculations. EEIO models are the best equipped to calculate these supply chain emissions when compared to process based LCAs because they are more comprehensive in terms of inclusion of contributors to the supply chain. It was determined that the differences in SEFs across the different sectors are much larger than the variations within a sector between years, thus if determining the order of priority for emissions from purchased goods, using any of the available years should provide the same general result of important categories. However, the magnitude of total emissions calculated using factors from the range of years can change significantly as the relative change within a commodity or industry SEF from one year to the next was found to be up to 50% smaller, highlighting the importance of regularly updating emissions factors.

2.3.4 Comprehensive Environmental Data Archive (CEDA) 5 VitalMetrics – Private Spend Based Emissions Tool

CEDA is a private and proprietary EEIO tool managed by VitalMetrics. It considers cradle-to-gate emissions and includes factors for greenhouse gas emissions as well as other common pollutants and climate impact factors [23]. The first version of CEDA was developed in 2000 and has been used since by the private sector for greenhouse gas accounting and life cycle assessments [23] and is updated annually. With CEDA, it is possible to quantify the supply chain footprint of a company by using spend data in order to identify key contributors [23]. CEDA combines input-output tables, emissions, resource use statistics, and characterization factors from LCAs in order to determine environmental impacts [24]. The 2005 CEDA 3.0 used the US Bureau of Economic Analysis 1998 input output tables [24], and a more recent description of specific sources used for updated versions was not found. The aggregation of environmental data is not a straightforward one as many assumptions are needed as well as harmonization in order to collect often incompatible information from multiple sources [24]. CEDA was developed for multiple intended possible uses including policy modeling, material flow analyses, substance flow analyses, LCAs, consumption and related environmental impact analyses, and selection of alternative materials for environmental design.

The version of CEDA considered in this study is CEDA 5 which uses data available from 2014 and is focused on US specific IO tables and emissions [25], thus is intended for national or at most continental level analyses [24]. In 2017, CEDA was one of the three most used EEIO models [19]. Although some documentation exists, it is not always clear what year is used for some data sources for the most recent versions of CEDA [19]. CEDA 5 includes 389 industrial sectors [26] and includes data from 2700+ environmental exchanges [23].

General conclusions found from the use of CEDA include that in order for progress in LCA methodology to continue, there needs to be significant collaboration between economic and environmental statistics in order to collect appropriate data at a frequency high enough to develop meaningful emissions estimates [24]. In addition, due to a large amount of data necessary for the development of CEDA, it was found to be difficult to quantitatively determine uncertainty with regards to specific data points [24]. Similarly to other IO tables discussed, in CEDA 5, imports are assumed to have the same environmental impacts as products produced domestically, thus there is room for improvement with regards to the handling of imports. Although imports represent a small portion of the US economy in terms of marginal monetary terms, it is possible that the emissions associated with the imports are larger than monetary fraction of the economy [24]. Finally, key sectors which contribute significantly to the emissions of the US economy would benefit from increased disaggregation for better understanding of the sector components that contribute significantly to emissions [24].

Chapter 3: Methodology

3.1 Introduction

In 2019, Stanford budgeted to spend $336.5 million dollars on materials and supplies of which most are for use in laboratories and research settings [27]. These items range broadly from broths for bacterial growth to magnetic resonance imaging machines to desks to ketchup. Emissions from purchased goods are thought to comprise a large portion of Stanford's, and any other research university's Scope 3 emissions. Food is estimated to be the 31st largest spend category among all Stanford purchased goods and services and was used to develop the tools and methodologies needed to ease and standardize calculations of Scope 3 emissions from purchased goods. In addition to the food purchasing data available from Stanford, data from other universities across the United States were gathered and used as a case study. These data were collected by reaching out to university dining hall coordinators, procurement offices, and through a partnership with the SIMAP Nitrogen Working Group (NWG).

The methodology was split into various steps which are presented in Figure 3 and further elaborated in this chapter. First, food purchasing data was collected which included both weight and spend data, and appropriate weights were assigned to these purchases. A categorization tool was developed to group the purchases into appropriate categories for which emissions factors exist. This tool was validated using comparisons to previously hand-categorized data from partner universities from the SIMAP NWG. Once total weights and spends were assigned to categories, different emissions factors were used to calculate estimates for total emissions from the food purchases to determine differences and quantify uncertainty present when using spend-based and weight-based emissions factors.

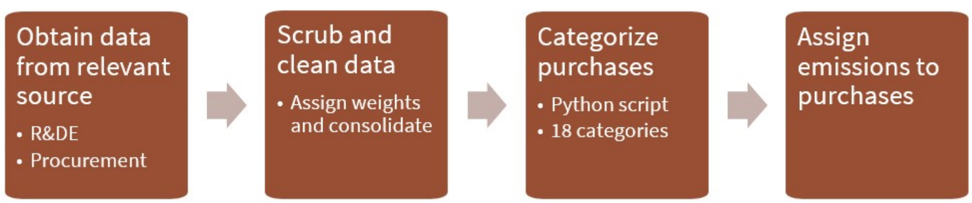


Figure 3. High level view of methodology for determination of emissions from food purchases using spend-based and weight-based emissions factors.

3.2 Data Collection

A partnership with the Stanford Procurement Office, the Stanford Office of Sustainability, and Stanford Residential & Dining Enterprises (R&DE) allowed for the collection of purchasing data from three sources for 2019: Amazon for Business, SmartMart, and Dining Hall data. The data provided are in the form of Microsoft Excel sheets with each individually purchased item's data recorded on a line of the excel file. A partnership with the SIMAP NWG facilitated the data gathering from other universities of six already hand-categorized data sets. This work considers $10,731,524 spent on 7,066 unique food items purchased for Stanford in 2019, including food purchased for all the dining halls on the Stanford campus as well as a subset of the food purchased through two purchasing platforms: 71% of the Amazon-for-business and 92% of the SmartMart food purchases. A subset of the food purchased through these avenues was used because it was not possible to determine both weight and spend data for all food items purchased through these avenues. The dining hall data set was provided by Stanford R&DE, which is the department which manages all the dining halls on campus. The Amazon-for-business and SmartMart data sets were provided by the Stanford Procurement office.

3.2.1 Dining Hall

Stanford R&DE is made up of Stanford Dining, Stanford Housing, Stanford Hospitality & Auxiliaries, and Stanford Conferences [28]. The data provided for this project only encompasses data from Stanford Dining, which represents data from 13 different dining locations [29]. Meals are available to undergraduate and graduate students as well as post-docs and Stanford Dining serves approximately 12,000 meals every day [30]. The dining halls have an internal centralized system for keeping track of food ordering and purchasing. Due to participating in other sustainability initiatives the dining halls already had weights assigned to much of the protein-based foods.

3.2.2 Amazon for Business

The Amazon for Business account is an account that is available to all departments at Stanford through the department's staff members and that can be used to purchase any item available through Amazon. It is the preferred and recommended purchasing method for any item at Stanford [31]. All purchases made through an Amazon for Business account are accessible by the Stanford Procurement department, and thus spending and purchasing reports are available and these are the reports that were used to determine food purchases in 2019 for this work. However, the data from these reports is not a full representation of all the purchases through Amazon at a university. It is also possible to purchase items for a department through a personal Amazon account and be reimbursed, and, depending on the department, the frequency of using a personal Amazon account as compared to the Amazon for Business accounts varies widely. Food purchased through the Amazon for Business account is usually made up of snacks for department break rooms or soda for department events.

3.2.3 SmartMart

SmartMart is an online catalog ordering system which is used by departments at Stanford to purchase items. Particularly, the vendors available through SmartMart are ones who have negotiated special pricing for Stanford [32]. SmartMart is also used for purchases of goods and services which require a Stanford signature. SmartMart is often used to purchase lab and medical products because it has a wider availability of items, but it also includes items for purchase such as software, food, office supplies, and more depending on vendor. Food purchased through the SmartMart account is usually snacks for department break rooms or soda for department events.

3.2.4 Categorized Data from Other Universities

A partnership was formed with the Sustainability Indicator Management & Analysis Platform (SIMAP) Nitrogen Working Group which is a group of universities focused on their nitrogen footprint. At universities, over 50% of the nitrogen footprint can come from food purchases, thus this group of universities had previously done significant work to obtain and process their food purchases. SIMAP is a tool developed by the University of New Hampshire which aims to help universities calculate their Scope 1 and 2 emissions and it includes food purchasing due to their focus on Nitrogen foot-printing. SIMAP has developed a tool to calculate these emissions which they call their emissions calculator, which is used by many universities across the US.

Through the partnership with the SIMAP Nitrogen Working Group (NWG), previously hand categorized data sets were requested and obtained. The methodology implemented to obtain the food data sets for the respective universities is outside the scope of this work but represent the best efforts of each of the universities to obtain relevant food data for input into the SIMAP emissions calculator. The universities who have calculated their food emissions through the SIMAP tool and are a part of the NWG are a self-selected group who are particularly interested in quantifying and mitigating these emissions and thus the data provided from these institutions are considered representative of the university's food purchasing and as accurate as is possible.

3.3 Data Cleaning

The Amazon for Business and SmartMart purchasing reports include all items that were purchased in the calendar year 2019, not only the food purchases. Thus, data cleaning was necessary to extract the food data from these data sets. In addition, to obtain both weight and spend data for these data sets, differing amounts of additional effort were needed depending on the purchasing platform data set. These efforts are detailed in the following sections with specifics necessary for each purchasing platform. The Amazon for Business and SmartMart data sets do not inherently include weight data as a column in the reports. Due to R&DE sustainability related targets, weights for most of the protein rich foods purchased for the dining halls were included, such as Beef, Chicken, and Pork. However, additional analysis was conducted to obtain weights for other categories to obtain a more complete representation of all food purchasing for the dining halls, not just protein rich foods.

One component of both the Amazon for Business and the SmartMart data set that was very helpful in extracting the food portions of the data sets was the inclusion of information on the purchased items United Nations Standard Products and Services Codes (UNSPSC). A UNSPSC is a code which can be up to eight digits long and is standardized and used worldwide for accurate classification of products and services [33]. Each set of two digits provides progressively more information about the item purchased as is illustrated in Figure 4. The first two digits of a UNSPSC code are called the segment and represent the general category of the item purchased. The following two digits are called the family name and give slightly more detail about the item category. The next two digits are called the class name and are even more specific, and finally the last two digits define the purchased item to a large amount of detail.

Segment	Segment name	Family	Family name	Class	Class name	Commodity	Commodity name
44000000	Office Equipment and Accessories and Supplies	44100000	Office machines and their supplies and accessories	44103100	Printer and facsimile and photocopier supplies	44103112	Printer ribbon

Figure 4. United Nations Standard Products and Services Code (UNSPSC) example [34].

Amazon and SmartMart both require UNSPSC codes to be present on the items available for purchase on their platforms, but they do not require a certain level of specificity, and there are frequently errors in the UNSPSC codes that are input by vendors on both platforms. In the case of food, the segment code is 50, and there are situations where the data available from both data sets only include details up to the segment code, such as 50000000, whereas there are other items where the code is specific all the way down to the commodity level, such as 50202306 whose commodity name is soft drinks and represents a purchase of a case of Coca-Cola cans.

3.3.1 Dining Hall Data

Stanford participates in the Association for the Advancement of Sustainability in Higher Education Sustainability Tracking, Assessment & Rating System (AASHE STARS) program, which is a reporting framework for both colleges and universities that measures and tracks their sustainability performances using a variety of metrics [35]. For analysis related to food sustainability, STARS focuses particularly on the impacts and sustainability efforts involving animal products, which it defines as "meat, poultry, fish/seafood, eggs, and dairy products" [36]. Because STARS only focuses on animal products, there were inconsistencies in the formatting of the animal product and non-animal product data, which had to be mitigated. The steps to assign weights to the data are explained in detail in Appendix II: Dining Hall Weight Assignments.

The file provided included 145,991 line items which were delivered to dining halls during the calendar year 2019 and represent $11,333,616 in spend. The dining hall data set had mostly been previously scrubbed to remove non-food items, but some remained in the data set. To remove any lingering non-food data, items which were in the categories of KitchenSupply, CleaningSupply, and Disposables we removed from the data set. After creating a pivot table to consolidate items with the same name, there were 3,135 unique food items purchased for the dining halls in 2019. From those, 2,731 items were able to be assigned both weight and spend which represent $10,364,968 in spend.

3.3.2 Amazon for Business

The original data set provided for this work included 135,788 purchases made during the 2019 calendar year. First, items whose "Order Status" was "Cancelled" were removed from consideration and items whose "Item Subtotal" was "Blank" were removed. This left 134,938 items purchased representing $8,639,308 in spend through Amazon for Business in 2019. To extract the food data, the items with the "Product Category" of "Pantry", "Grocery", and "Amazon Yum" were considered which brought the number of items down to 16,426 and represents $451,071 in spending. Thus, food makes up 12% of the items purchased through Amazon for Business and 5.2% of total spend in 2019.

Then UNSPSC codes were used to filter out items that were not food but had been included in one of the "Product Category" listed above. Items in these product categories that were not food included items such as notebooks, lighters, soap, and hand creams. Next, items with an UNSPSC code that did not begin with 50 were removed as this represents the food category at the segment level. This decreased the data set to 15,699 items and $434,546 in spend. Then a pivot table was created to consolidate the items with the same name which resulted in 5,589 unique food items purchased through Amazon for Business.

Once the food items were extracted from the larger purchasing data set, the items for which weights could be assigned had to be determined. Thus, items whose "Title" did not include any of the following key words which indicated that weight or volume data could be extracted from the purchased item name were removed: "lb", "pound", "fl oz", "gallon", "liter", "gram", "ml", "gal", "1.5L", "2L", "kg", "g", "oz", "ounce", "tea", "k-cups", and "k-cup". By using only the items whose "Title" included one of these key words, both weight and spend based information could be extracted to achieve the goal of comparison between the two approaches for calculating emissions.

Weights were assigned to the items left in the data set based on the weight or volume-based indicator that was present in the item "Title" along with some searching for the items on Amazon in order to assign an appropriate weight when the item "Title" description was confusing. In addition, the number of packs of items purchased was also taken into consideration (for example, 24 pack of 12 fl oz cans was considered a 288 fl oz purchase). Food items purchased through Amazon for Business that were included in the following analyses are 4,058 unique items purchased representing $308,597 in spend, which is equivalent to 71% of the original items tagged as food. After the data cleaning outlined above, the Amazon food data was ready to be categorized.

3.3.3 SmartMart

The entire SmartMart data set represents 241,636 purchases at a total spend of $59,304,426. First the data set was trimmed to only include the items whose "Line Status" was "Approved". The "Invoice Line Extended Price" was used as it represents the total spend based on the total quantity of an item and its associated individual price. Then UNSPSC codes were used to filter out items that were not food, and an additional analysis was done to remove all Mouse food items as those are not intended for human consumption but were still categorized under the 50 UNSPSC segment code. Mouse food items appear in the SmartMart purchasing data set because SmartMart is used in large part for laboratory supplies, including food for lab mice. Items with an UNSPSC code that did not begin with 50 were removed which left 2,726 food purchases which represent $62,893 in spend. Thus, food represents 1.1% of items purchased through SmartMart and 0.01% of spend in SmartMart in calendar year 2019.

Then a pivot table was created to consolidate the items with the same name which resulted in 325 unique food items purchased through SmartMart. Next, items whose "Product Name" did not include any of the following key words were removed: "lb", "fl. oz", "oz", "gallon", "1.5L", "tea", and "k-cups". This left only items for which both weight and spend-based information could be extracted in order to achieve the goal of comparison between the two approaches for calculating emissions.

Weights were assigned to the items left in the data set based on the weight or volume-based indicator that was present in the item "Product Name". In addition, the number of boxes of items purchased was also taken into consideration (for example, 2Oz. Bags Box of 24 would be considered a 48 oz purchase). Food items purchased through SmartMart that were included in the following analyses are 277 unique items purchased representing $57,958 in spend, representing 92% of the items recognized as food. After the data cleaning outlined above, the SmartMart food data was ready to be categorized.

3.4 Categorization Script

To determine emissions from purchased goods, it is important to be able to sort purchased items accurately and consistently into the categories for which emissions factors exist. For example, at the current state of research in the field, it would not be feasible to have a different specific emissions factor for each food item purchased at a university because a detailed LCA would need to be conducted for each of these items. It is much more feasible to have an average emissions factor for the fruits category which considers many LCA studies that are conducted over many different types of fruits produced all over the world and can be applied as an average to all fruits purchased by the university. This same approach can be done for many different categories of foods or purchased goods. Preferably, there would be different emissions factors for each specific type of food, (or even more ideally, different types of foods produced in different regions of the world) but at the current stage of the field, the emissions factors that are available are at the level of detail which represents an average for a category unless a specific LCA is conducted for each item.

These emissions factors can be used to determine the order of magnitude of emissions associated with purchased items but should not be considered an accurate methodology to determine the exact magnitude of emissions for those items. These emissions factors are beneficial because they do not require as much effort to develop because a smaller number of LCAs can be used to represent categories as a whole, but they also introduce some complexity when it comes to assigning purchased foods into these categories.

In the last couple of years, some universities have started quantifying their Scope 3 emissions, and the most popular categories to begin with are employee commuting, university sponsored travel, and food purchasing emissions because these are often the easiest to obtain data for. Many of the universities who have started embarking on this journey have resorted to hand categorizing their food purchasing data into the relevant categories. This means a person, usually someone from an office of sustainability or a member of the management team of a dining hall, must manually categorize the thousands of food items that a university purchases along the course of a year. This inherently takes many hours, makes comparisons between university results difficult because different universities can take different approaches to categorizing foods, and leaves significant room for human error.

Many of the universities who have started looking at their Scope 3 emissions use SIMAP as a tool to calculate their food emissions. SIMAP is used by almost 150 institutions [37] and considers 18 food categories which are defined by Heller et al. [4] and for which they have selected emissions factors for from the same paper. However, SIMAP requires the input data to already be pre-categorized with up to three categories per item to account for multi-ingredient items. To facilitate the categorization process and to standardize and ease the emissions from food purchasing process, a python-based categorization script was developed and is further detailed below.

Figure 5. Data flow through categorization script. Green components are script functions, blue components are food data flows, and gray components are external inputs into the script which aid in the categorization.

Figure 5 represents a schematic of how the categorization script functions. It requires a user input with the food data which will be categorized and includes the optional user input of a list of local vendors. In addition, there are two required inputs which are provided: Keywords and Edge Cases. These two inputs have been developed over the course of this work and represent the needed keywords and edge cases which are used to categorize these food purchases. The development process to create these documents are further explained in Section 3.5 Development of Keywords and Edge Cases Documents. Keywords are words that, if present in the food purchase name, indicate that that purchase should be assigned to one of the 18 categories which are detailed in Table 1. For example, if the word *apple* appears in the food name, then that line should be assigned to Fruits, and if the word *hamburger* appears, then it should be assigned to Beef. This assignment process occurs in the function called categorize_purchases shown in Figure 5, however, the categorize_purchases function only attempts to assign categories up to the dotted line present in Table 1. The rows after the dotted line are considered in the categorize_spices function. Each food item can be assigned to a maximum of three categories. If the output of this script is run through SIMAP, or if the category breakdown scripts (see 3.6 Category breakdown scripts) are run, the weight or spend for the line item will be evenly attributed to each of the assigned categories. For example, for a line item representing nine pounds of *lasagna* would be assigned to Beef, Vegetables, and Grains, and three pounds would be attributed to each of these categories. This follows the convention that SIMAP has created for multi-ingredient items.

The Edge cases input document includes information to deal with situations where combinations of keywords (or combinations of words that are not included in the keywords document) interact such that the assignments from the keywords when considered separately becomes incorrect. For example, when the words *yellow* (not a keyword) and *pepper* (keyword for Spices) are present together in a food item name, the assignment to Spices should be removed and should be replaced with an assignment to Vegetables. Another example of this is when *hamburger* (keyword for Beef) and *bun* (keyword for Grains) are both present, and in this case, only the assignment to Beef should be removed. Dealing with these edge cases occurs in the do_edge_cases function in Figure 5. The edge cases function is also able to handle situations in which the order of the key words matters. For example, if the keyword *chocolate* and then the keyword *milk* appear in that order, the do_edge_cases function removes the assignment to Coffee and Tea, which is the correct categorization for chocolate, however, if the keyword *milk* appears and then the keyword *chocolate*, then the do_edge_cases function removes the assignment to Milk.

Table 1. Extract from Keywords document including the first three keywords for each food category.

Food Categories	Foods in the category
Beef	Beef Steak Hamburger
Pork	Sausage Bacon Pepperoni
Chicken	Chicken Turkey Poultry
Cheese	Cheddar Parmesan Cheese
Eggs	Egg Omelette Mousse
Milk	Milk Yogurt Cream
Fish	Fish Lobster Shrimp
Liquids	Juice Soda Broth
Grains	Wheat Bagel Bread
Fruits	Apple Orange Lemon
Nuts	Cashew Almond Walnut
Oils	Oil Canola Mayonnaise
Beans	Soybean Tofu Bean
Potatoes	Potatoes Potato Fries
Coffee and Tea	Coffee Tea Chocolate
Sugars	Sugar Sweetener Honey
Vegetables	Tomatoes Lettuce Cauliflower
Vegetables	Basil Chive Garlic
Spices	Pepper Clove Basil

As the extract from the keywords document in Table 1 shows, there are two separate vegetables keywords groupings. The first vegetables keywords grouping refers to vegetables which cannot also be spices. In other words, these are vegetables that cannot be dried and considered spices, for example, *lettuce,* and *carrots* versus *basil*, *chive*, and *garlic*. When vegetables have been dried to be turned into a spice, they have a different emissions factor than when they are eaten fresh because the drying process requires energy input which in turn produces emissions. The spices category also includes keywords for things that cannot be vegetables such as *pepper* and *salt*. The categorization process considers the vegetables which, when fresh, are vegetables and, when dried, are spices separately because of two factors. First, spices should not be assigned to items where they do not make up a significant portion of the weight. For example, with the food item *garlic bread* it does make sense to assign it to Grains because of *bread* but it would not be accurate to assign this item to Spices because of the word *garlic* because garlic represents such a small portion of the total weight of any garlic bread. Therefore, the categorize_spices function in Figure 5 only considers lines that have not yet been assigned a category after passing through the first two functions. If the line name includes a keyword that is present in the second vegetables category, and it does not include the keywords *flake(s), dried, powder, granulated, dry, spice, leaf, crushed,* or *herb,* then it is tagged as a Vegetable. Finally, if the line includes any of the keywords for spices and has not been assigned to any category, it is assigned to spices.

After the categorize_spices function has completed, most of the line items in the original food data should be categorized with the exception of some items whose names do not include any words that are keywords and items that are misspelled. To minimize the number of lines that are not categorized, the fix_spelling function only looks at lines where no categories have been assigned and leverages the pyspellchecker library [38], which uses a Levenshtein Distance algorithm to find insertion, deletion, replacement, and transposition permutations for words in a given dictionary. The dictionary used in this categorization script is created based on the words present in the keywords and edge cases document to ensure that words are only spellchecked to words that would be useful for further categorization (food related words). The script keeps track of what lines were updated and adds the spellchecked word to the end of the line item for ease in visual verification by the user at the end of the process. The default setting for fix_spelling is to only consider the most likely autocorrection as defined by the pyspellchecker library, but the user can select the option to check all possible autocorrections by changing the setting to Advanced Autocorrect. This is more likely to find the correct autocorrection if a word is badly misspelled but is also more likely to incorrectly add autocorrections if a word is only slightly misspelled and is close to several keywords. Once the misspelled lines have been autocorrected, only those lines go through the loop again of categorize_purchases, do_edge_cases, and categorize_spices.

Next, the function tag_organic looks at the item names again, and if the item includes keywords which would indicate that that item is organic, such as *organic, org, orgnc* then that item is tagged in organic in the appropriate column. Finally, if the user provided any vendors which should be considered local, the tag_local function will look at the vendor for each line and if it is included in the list of vendors which are local, that line will be tagged as local in the appropriate column. The file that is created at the end of the script is in the correct format for direct input into the SIMAP emissions calculation tool. The file includes information on all the items that the script could not categorize, the autocorrected lines, and, if selected, the breakdown of spend and weight by category which is further explained in Chapter 3.6 Category breakdown scripts.

3.5 Development of Keywords and Edge Cases Documents

In partnership with the SIMAP Nitrogen Working Group [39], food purchasing data from the universities listed in Table 2 was used to test, improve, and validate the Categorization Model. In aggregate, this data represents over 13,600 metric tons of food and more than 37,000 lines of food purchases. By utilizing these data sets, keywords and edge cases were added to their respective documents to accurately categorize a larger proportion of these data sets. Each university's data was run through the script, and the uncategorized lines were analyzed to determine what additional keywords would be necessary to categorize these lines. After all reasonable keywords were added, that university's data was re-run through the program in an iterative fashion until a minimum level of uncategorized data was achieved. In some cases, this represented zero items that the script could not categorize, but in other cases, items could not be categorized due to lack of clarity on what the item was.

Some universities provided hand categorizations that had been conducted for previous years food purchases, and a comparison script was developed to determine which line items had been categorized differently between the hand categorization and the script categorization. The lines which had been categorized differently were analyzed and relevant edge cases were added to the edge cases document when it was determined that the script was incorrectly categorizing an item. These steps were conducted for each university data set and repeated iteratively to ensure that the addition of a keyword or edge case from a different university data set did not negatively affect the categorization of another university's data set.

Table 2. University data sets used to develop edge cases and key words document as well as to determine accuracy of categorization script.

University	Data Time Spanning	Previous Hand Categorization Available
A	October 2018 and April 2019	Yes
B	July 2017 – June 2018	Yes
C	2017-2018 Academic Year	No
D	January – December 2018	Yes
E	January – December 2018	Yes
F	January – December 2019	No
G	May 2015 – April 2016	Yes
H	July 2018 – June 2019	Yes

3.6 Category breakdown scripts

To calculate the total emissions from a data set after categorization, it is helpful to be able to determine the total spend or weight assigned to each of the food categories for which emissions factors exist. To facilitate this determination, an additional functionality was added to the categorization script which calculates the breakdown per category. Users can determine if they would like the total spend, the total weight, or both to be calculated per category.

Figure 6 represents a schematic of how the categorization breakdown script functions. The output of the categorization script is taken in as an input, and a user selects the breakdown of interest, either spend, weight, or both. The do_counts function determines the number of categories that are assigned for each line item as the weight or spend data will be evenly divided between all the categories for which each line is assigned to. The do_conversions function converts all the weights to kilograms using the unit conversions as defined by SIMAP [40], which are available in the Appendix III: SIMAP Unit Conversions, in order to accurately provide one weight value per category. The calc_category_breakdowns function goes through each item in the data set and cumulatively adds the weight or spend divided by the number of categories assigned to the item for the appropriate categories for that item. When the toggle to calculate the total spend or weight per category is selected, the final excel file includes a tab for the total weight and/or a tab for the total spend per category. These values can then be multiplied by the emissions factors for each category as described in Chapter 3.8 Calculating Total Food Emissions.

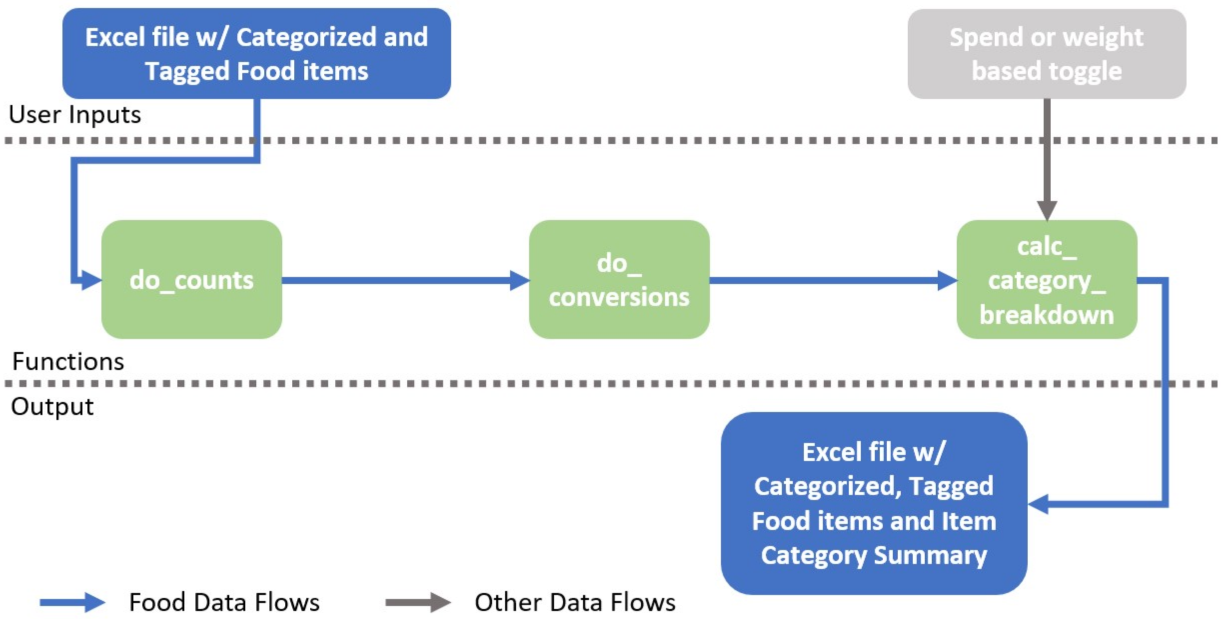


Figure 6. Data flow through category breakdown script. Green components are script functions, blue components are food data flows, and gray components are external inputs into the script which determine which breakdown is calculated.

3.7 Website

To facilitate wide use of the categorization tool developed, a website was created in which users can upload their own food data sets using templates provided on the website and run the categorization script to directly download the categorized data. The website is currently in beta mode and can be found by going to https://food-emissions-categorization.wl.r.appspot.com/. The website is supported by Google cloud platform and is slightly slower than using the python script locally. A copy of the python script can be obtained by reaching out to the author and through the author's GitHub. A visual representation of the website is available in Figure 7. All the functionality described above is available through use of the website, except for comparisons between hand-categorized and script-categorized data sets. This website has begun to be shared with various groups, including the Ivy Plus Sustainability Consortium [41], to determine their emissions associated with food purchasing. Discussions are underway to recommend the tool for use with several other groups.

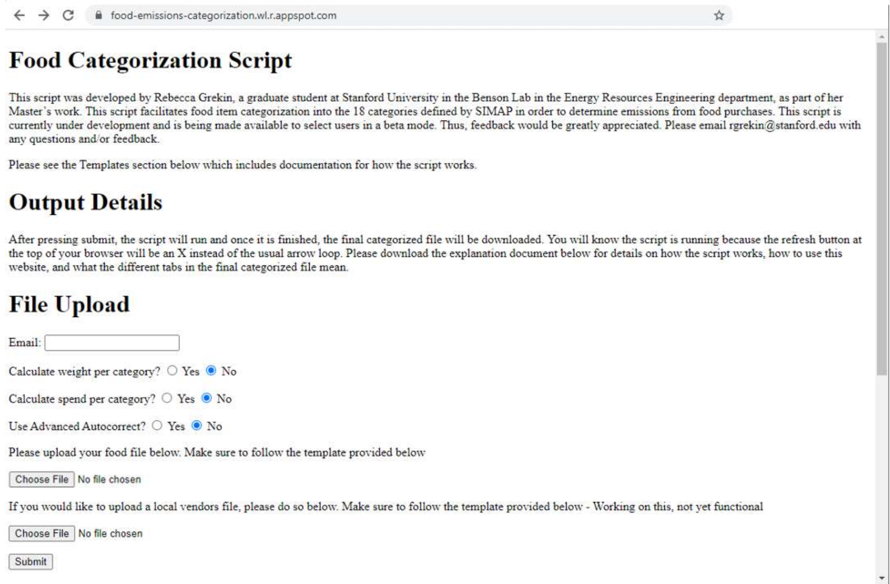


Figure 7. Screenshot of website created to facilitate access to categorization script.

3.8 Calculating Total Food Emissions

Emissions factors are the values which are multiplied by the unit quantity of a relevant item to determine the emissions associated with that item as illustrated by Equation 1 and Equation 2. In terms of environmental impact, CO_2 equivalent emissions are the focus within the context of this work, although there are many other emissions and environmental impacts that can be considered including nitrogen, acidification, eutrophication, land use, etc. There are many sources available for emissions factors and they are integral to the process of translating purchasing or consumption data into emissions from those items.

Some emissions factors are easier to determine, for example, the emissions factor associated with the burning of a fuel (Scope 1) is directly related to the chemical composition of the fuel and the chemical reactions that occur when they are burned, which are relatively straightforward to determine. However, for items which are not burned, the majority of emissions associated with them are instead the Scope 3 emissions, thus the life cycle emissions are dominated by the phases other than the use-phase such as their production and transportation. As can be seen in Figure 8, there are many types of emissions factors, and different levels of transparency associated with each source. In this work, the private emissions factor sources are defined as sources or tools which are perceived by the user as a black box, and where the emissions factors themselves are not readily available. In these cases, the raw data is provided to the source or tool, and a result is returned which contains the quantity of emissions for the data in aggregate, or occasionally, some category breakdown is available. These private sources or tools are not particularly transparent with regards to the assumptions made nor the methodology used to determine the emissions factors themselves. All the private emissions factors considered in this work were spend-based.

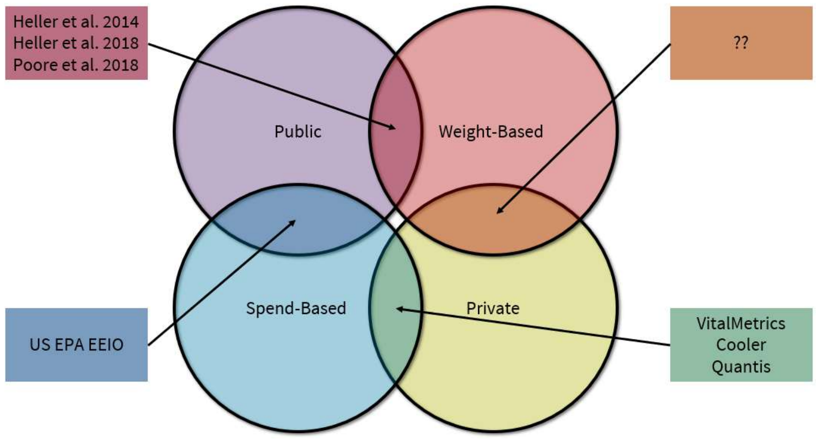


Figure 8. Sources and intersections for different types of emissions factors.

The public emissions factors are defined as ones where the emissions per category are easily accessible and information about methodology and assumptions are readily available. Most of the public emissions factors that were found for food were weight-based except for the US Environmental Protection Agency EEIO Model which is spend-based. Once the emissions factors are gathered and the total weight or spend for each group for which an emissions factor is available is known, Equation 1 and Equation 2 can be used to determine the CO_2 associated with that item or category.

$$CO_{2_{item\ purchased}} = kg_{item\ purchased} * emissions\ factor \left[\frac{kg\ CO_2}{kg_{item\ purchased}}\right]$$

Equation 1. Calculation for determining the emissions when weight-based emissions factors are used.

$$CO_2 = \$_{item\ purchased} * emissions\ factor \left[\frac{kg\ CO_2}{\$_{item\ purchased}}\right]$$

Equation 2. Calculation for determining the emissions when spend-based emissions factors are used.

Weight- and spend-based emissions factors each have their respective pros and cons, mostly in the form of sources of uncertainty, which are summarized in Table 3. Weight-based emissions factors are often determined through a bottoms-up analysis, whereas spend-based emissions factors are more often determined through a top-down analysis. Part of the appeal of spend-based emissions factors comes from the fact that the economy is used to tracking dollars spent but has been less historically interested in keeping track of weights of items as they pass through an economy during an item's lifespan. For example, when considering a purchase of chairs, an entity might reasonably keep track of the $500 that was spent on the chairs, and/or keep track that five chairs were purchased, however, it is very difficult to image that an entity would keep track that 100 kg of chairs were purchased. With the rise in e-commerce, and the subsequent need to ship individual items to consumers, the data to keep track of the weights of items has begun to become more readily available, but there is no evidence that this data has been leveraged for emissions calculations.

Table 3. Summary of pros and cons for weight-based and spend-based emissions factors.

	Weight-based	Spend-Based
Pros	Often public Often bottom-up approach	Often easier to obtain spend-based data
Cons (Sources of Uncertainty)	Obtaining weight data Double Counting co-products Loss of accuracy due to aggregation into categories	Accounting for sale, wholesale, normal pricing Accounting for regional price differences Often top-down approach Loss of accuracy due to aggregation into categories

Because this weight-based data is not currently easily kept track of, using weight-based emissions factors introduce a certain level of uncertainty as is discussed in section 3.3 Data Cleaning as well as 3.9 Determining Outliers. Weight data is not often directly available, thus creative solutions must be implemented to obtain this data which currently require human input, and thus, human error. Methods can be implemented to decrease this human error, for example, by creating automated categorization scripts as described in section 3.4 Categorization Script. In both the weight-based and spend-based emissions, there is inherent uncertainty due to the need to put the purchases into categories in the first place. By aggregating the data into categories, detail is lost, and error is introduced.

3.9 Determining Outliers in Data Set

As previously mentioned, there is inherent uncertainty when using weight-based emissions factors due to the additional analysis required to obtain weights of items purchased since this data is not readily available. One way to mitigate this uncertainty from weight-based calculations is to use the spend based data to identify outliers since there is basically 100% certainty with regards to the spend data for each item purchased.

After the categorization step has been completed, an analysis can be conducted for single ingredient items to determine if there are any outliers in this data set in terms of dollars spent per kg of item for each purchase. Because the items are categorized, statistically relevant indicators within each category of food can be analyzed including the median, mode, minimum and maximum values, and inter quartile range (IQR) which is defined by Equation 3,.

$$IQR_{for\ a\ food\ category} = \frac{\$}{kg}_{Q3\ for\ a\ food\ category} - \frac{\$}{kg}_{Q1\ for\ a\ food\ category}$$

Equation 3. Interquartile Range definition.

The spend per kilogram of item can be calculated for each line item, and Equation 4 and Equation 5 can be used to determine if each item can be considered an outlier either because the spend per kilogram is too high or because the spend per kilogram is too low.

$$\frac{\$}{kg}_{outlier} > \frac{\$}{kg}_{Q3\ for\ a\ food\ category} + \left(1.5 * IQR_{for\ a\ food\ category}\right)$$

Equation 4. Outlier definition for spend per kilogram which is too large.

$$\frac{\$}{kg}_{outlier} < \frac{\$}{kg}_{Q1\ for\ a\ food\ category} - \left(1.5 * IQR_{for\ a\ food\ category}\right)$$

Equation 5. Outlier definition for spend per kilogram which is too small.

When the spend per kilogram is too large compared to the values for a particular category of food, this is a strong indication that the weight assigned to that item is too small, likely because it is not possible to determine from the item name that the purchase includes some kind of pack or multiple items within a purchase. When the spend per kilogram is too small, this is a strong indication that the weight assigned to that item is too large. This can happen when the volume or weight in the title of an item is inaccurate, or if there was a difference in unit counting within the data.

The base case for this work is that all the weights and spend data are considered to be accurate. However, there are two options that are possible in order to mitigate the impacts of these outliers if they are in fact incorrect. The first of which is called the Removal Scenario which involves simply not considering any items which are determined to be outliers based on the methodology described previously. The second option which is called the New Outlier Weight Scenario is to allocate a new weight to the items which are considered outliers by assuming that the spend data is correct and attributing that spend an appropriate weight for that food category based on the median as illustrated by Equation 6.

$$New\ allocated\ weight = \frac{original\ spend\ for\ item}{\frac{\$}{kg}_{median\ for\ a\ food\ category}}$$

Equation 6. Formula for allocating the new weight for each line item that was previously tagged as an outlier.

The results from the outlier analysis for both these scenarios as well as the base case are shown in Chapter 4.4 Outlier Analysis.

3.10 Methodology Take-aways

After validating the categorization script and categorizing the data, emissions factors determined using different methods can be compared to quantify the uncertainty present in the data set. Figure 8 illustrates the types of emissions factors used in this analysis. Public emissions factors are considered such because the methodology and the emissions factors themselves are readily accessible in published literature or government websites. The private emissions factors are ones which can be considered "black boxes" as the data is provided by a user to the emissions calculation tool and the aggregate emissions are returned. These private emissions factors are embedded in these tools, and the emissions factors themselves are unknown for specific categories.

The weight-based emissions factors used in this analysis were obtained from journal articles [4,5,15]. The US EPA publishes an economic environmental input output model as detailed in Chapter 2: Literature Review, which includes emissions factors for many US economy sectors, is spend-based, and is also publicly accessible. There are several private emissions calculating tools which are managed by companies which are spend-based. As of this study, no private weight-based emissions factors had been encountered.

Chapter 4: Results and Discussion

In this section, the results of the validation of the categorization script will be introduced which compare the script categorization output to the hand categorization when they were available. In addition, the results are presented of the comparison of the total emissions estimated for the Stanford case study using different emissions factors found in literature and from private tools. In this section, many of the results will be compared to the SIMAP results because it is the most widely used method for university carbon footprint applications. This is not meant to imply that the SIMAP results are correct in comparison to other methods.

4.1 Categorization Script Validation

Several universities provided hand-categorized food purchasing data sets that had been analyzed previously for input into SIMAP. These hand-categorized data sets provided the opportunity to conduct a validation of the categorization script by comparing the categorization script output to the hand categorized output for the same data. The summary of the data provided by the universities and the outcomes from the script analyzed are detailed in Table 4. Total uncategorized items by the script represent 0.99% of the lines in the original provided data sets, but only 0.39% of the total weight of all the data sets. In the case of the universities where a much larger percent of the weight was not categorized as compared to the average among the universities, a few items make up a large amount of the uncategorized weight. For example, in the University E data set only three items: *21 oz 1-ls 24 core spar, 5 gallon 1-ls core spar,* and *2.5 gallon 1-ls core spar* make up 61.7% of the uncategorized weight.

Table 4. Summary of University Data Sets Analyzed Using Categorization Script. Deleted lines are items which the script was not able to categorize thus were not considered in the final analysis for emissions. Among all the data sets the total percent of weight that was not categorized by the data set is only 0.39% however this value ranges from 0.00 to 2.48% depending on the data set.

University Dataset	Weight (kg)	Total Lines	Number of Autocorrected Lines	Number of Deleted Lines	Deleted Weight (kg)	Percent Deleted Weight (%)
A	442,912.53	770	3	3	200	0.05
B	1,700,312.48	747	3	7	1,629.79	0.10
C	1,158,565.22	6,011	0	21	702.69	0.06
D	3,045,939.53	22,057	13	203	28,242.83	0.93
E	243,532.33	1,185	7	53	6,033.18	2.48
F	2,830,969.24	6,948	0	19	77.95	0.00
G	303,325.71	1,486	13	17	393.82	0.13
H	27,887.36	2,467	61	93	521.49	1.87
Totals	9,753,444.39	41,671	100	416	37,755.65	0.39

The results for runtime required for each university data set is presented in Figure 9. The runtime shows a very linear relationship (with an R^2 of 0.9944) between the number of lines in the data set to be categorized and the number of seconds required for categorization. The relationship is not exactly linear for each data set because of the additional loop through the categorize_purchases, do_edge_cases, and categorize_spices functions that data that was autocorrected must go through as shown in Figure 5. Thus, data sets that appear below the linear best fit line had a number of lines that were autocorrected that is less than the average among all the data sets, and those that are above the line had a larger number of lines that were autocorrected.

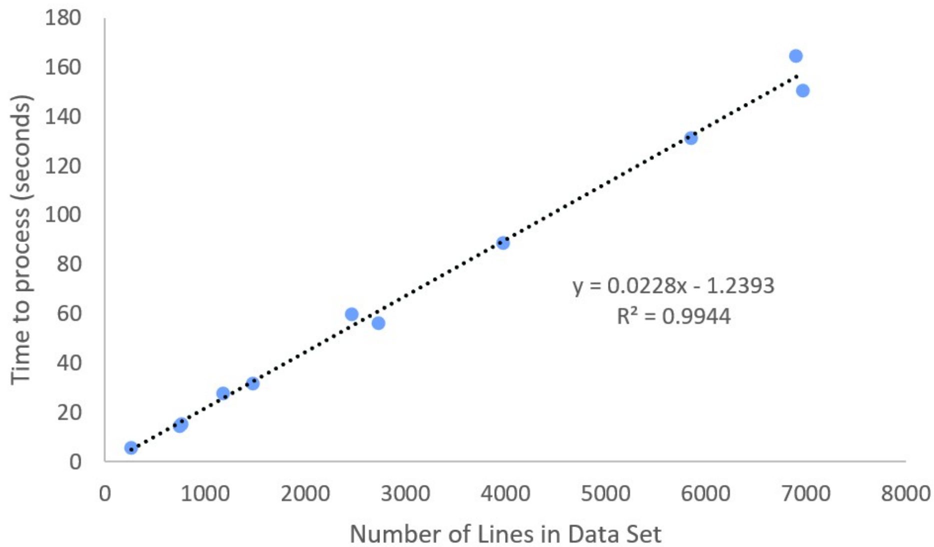


Figure 9. Required time to categorize each data set based on running the script locally.

A total calculated emissions comparison was then conducted using the assignments from the hand categorization and from the script categorization to determine the impact on estimated emissions from hand categorizations and from the script categorization. The emissions comparison was conducted using the emissions factors available in SIMAP in 2021 [37]. Because the university data was categorized for use in SIMAP, the categories were the same for the hand categorization and the script categorization. The difference in emissions arises from items being categorized into different categories for which the emissions factors are different. For each university, the percent difference between the calculated emissions was determined using Equation 7 for the entire university's data set as well as the difference in weight assigned to each of the 18 food categories and those results are illustrated in Table 5 and Table 6 respectively.

$$\%\ difference\ emissions = \frac{emissions_{hand\ categorization} - emissions_{script\ categorization}}{emissions_{hand\ categorization}} * 100$$

Equation 7. Definition of percent difference for emissions.

The goal of the validation is not necessarily to obtain percent differences as close as possible to zero because the hand categorized data sets are not perfectly categorized. For example, from analyzing the items that were categorized differently by the script as compared to the hand categorization it was determined that there is a reasonable amount of human error in these data sets. For example, occasionally items were simply categorized in the wrong category. Examples of this include: misreading *meat less* and categorizing as *meat*, as well as occasions where identical items were categorized differently because they were hundreds of lines apart in the data set and it would be practically impossible for someone doing a hand categorization to ensure that all items with the same or similar name are categorized exactly the same way throughout the data set, particularly when there are thousands of items in a data set.

In addition, there are inherent differences between the emissions calculated for the hand categorization and for the script categorization because the script is not always able to categorize all the items in a data set, but the hand categorization that was provided includes categorizations for every item in the data set. Thus, the total calculated emissions will be different if there is some data that was not categorized in the script because it will be missing from the total weights in that data set. In this case, it would be expected that the percent difference would be positive, which helps explain in part some of the differences between the two methods of categorization. However, referring to Table 4, the maximum amount of weight that was not considered by the script categorization because it was not able to categorize it was in the data set for University E at 2.48%, which only had a percent difference in total calculated emissions of 0.58%, which is by no means the largest magnitude difference in emissions from these data set. Thus, the omission of parts of the data because the script was not able to categorize the items does not explain 100% of the difference in aggregate emissions for each data set.

Table 5. Percent Difference between total calculated carbon footprint between script and hand categorization for SIMAP NWG partner universities.

University	A	B	D	E	G	H
% difference kg CO2eq	4.19	0.46	-0.09	0.58	4.57	-4.79

It is interesting to note that most of the percent differences between the methods are positive, and as can be seen in Equation 7, because of the definition being used for percent difference, a positive percent difference indicates that many of the universities had previously categorized their food data in a way that indicated a higher amount of emissions than would be calculated based on the categorization from the script. Although the magnitude emissions difference between most of the university's data sets is less than 1 percent, in some cases there is almost a 5 percent difference. However, as can be noted in Table 6, the differences in weights assigned to the categories can have a much larger magnitude difference as defined by Equation 8.

$$\% \ difference \ emissions = \frac{weight_{hand \ categorization} - weight_{script \ categorization}}{weight_{hand \ categorization}} * 100$$

Equation 8. Definition of percent difference for weight.

The blue cells in Table 6 represent larger differences in weight assigned to the categories when comparing the hand categorization to the script categorization, both for positive and negative percent differences. The blue cells represent values of percent difference in weight closer to the 10th and 90th percentile for each university while the yellow cells represent values closer to the median of the values for each university. As can be noted visually from the table, the blue cells are concentrated in the lower portions of the table, for which the food categories on the left are ordered in terms of emissions factors, where the highest emissions factors are at the top and the lowest emissions factors are at the bottom. Thus, very large differences in weight assigned to the categories at the bottom do not translate to large changes in calculated emissions for the data set as a whole. More details on what contributed to these differences in emissions can be found in Appendix IV: Script and Hand Categorization Differences.

Table 6. Percent difference in weight assigned to each food category for each university. Blue cells indicate data points that are further from the median values for the percent differences assigned to each food category for each university based on the hand and script categorizations.

University	A	B	D	E	G	H
Beef	5.04	2.83	-4.24	-7.58	10.3	-16.3
Pork	-2.66	1.91	2.63	-6.45	-1.56	-2.83
Chicken	6.66	1.55	4.03	0.10	-7.13	5.80
Fish	8.08	10.7	8.80	0.24	13.9	7.09
Milk	11.5	5.11	11.3	10.9	0.33	0.22
Cheese	11.7	-6.83	0.45	-2.78	2.38	-0.40
Eggs	1.72	0.00	1.22	9.70	17.85	4.56
Grains	3.50	5.24	-8.46	7.94	-10.11	9.92
Vegetables	-9.55	8.82	3.39	16.5	8.72	9.37
Fruits	10.0	-1.81	13.6	-7.77	45.9	-8.59
Potatoes	-6.41	0.09	4.64	-0.16	-58.4	4.90
Beans	3.53	0.49	-0.15	-64.4	-3.72	-28.8
Nuts	54.6	10.0	14.0	-27.0	9.01	-84.8
Liquids	-4.11	-8.66	-55.1	28.4	-1492.8	-54.0
Coffee and tea	-39.4	-172.7	-62.9	-22.4	-70.1	-351.0
Oils	-62.2	-9.86	6.54	-23.9	26.1	1.51
Sugars	-39.7	-60.7	-46.7	25.4	50.7	7.71
Spices	66.4	46.5	48.6	8.50	42.3	59.8

Percentile: 10th 50th 90th

Overall, the categorization script was found to provide an accurate characterization of both the food purchases as well as the emissions when compared to previously hand-categorized data sets. Although the categorization script omits some of data because it cannot categorize it, this number of items is relatively small and represents in aggregate only 0.39% of the total weight analyzed among these universities. Most of the differences in categorization between the script and hand-categorization are in the less emissions intensive categories which translates to small impacts in overall emissions calculations.

4.2 Stanford Food Emissions Case Study: Analysis of results using different emissions factors

Three sources for consumption at Stanford were considered to analyze the difference in emissions for food purchased using different emissions factors: Amazon for Business, SmartMart, and the dining halls run by R&DE. After the validation of the categorization script was conducted, the script was used to categorize the data from these three data sets. The summary of the categorization outcome for these data sets is presented in Table 7.

From the Amazon data set, one item was autocorrected from *juce* to *juice* and 60 lines were not categorized, most of which are not edible food items including hydrogen peroxide and soap, which the cleaning described in Chapter 3.3 Data Cleaning did not identify as non-food. Thus, allowing the script the flexibility of not forcing it to assign categories to all items provides a final check for removal of non-food items. Some of the uncategorized items are food items, but the item names do not include enough information to categorize them, even by hand. Additional analysis into what the item is would be necessary to categorize them. A good example of an item where the item name does not contain enough information for categorization is the deleted line from the SmartMart data set which was *assorted party mix*. Without more information it is impossible to categorize this item correctly.

Table 7. Summary of Stanford Food Purchasing Categorization Results.

Dataset	Weight (kg)	Spend ($)	Number of Autocorrected Lines	Number of Deleted Lines
Amazon for Business	21,107	$307,008	1	60
SmartMart	7,457	$57,943	0	1
Dining Halls	2,802,250	$10,364,968	0	0

Figure 10 shows the distribution of the food weight purchased through the three platforms for each food category based on the categorization script output. As is expected, the variety of foods purchased for the dining halls is much greater than the variety purchased through Amazon for Business and SmartMart due to the differences in purpose the food purchased through these avenues have. For the dining halls, the food provides full balanced meals for thousands of students on campus, but the food purchased through Amazon and SmartMart are for department snacks or events, which is why a majority of the foods purchased fall under the liquids category. A deeper analysis into the liquids purchasing through Amazon and SmartMart is available in Appendix V: Liquids Purchasing Analysis. For the dining halls, more than 50% by weight of the foods purchased are Grains, Vegetables, Fruits, Potatoes, Beans and Nuts, whereas meats including Beef, Pork, Chicken, and Fish make up less than 20% of the weight of foods purchased.

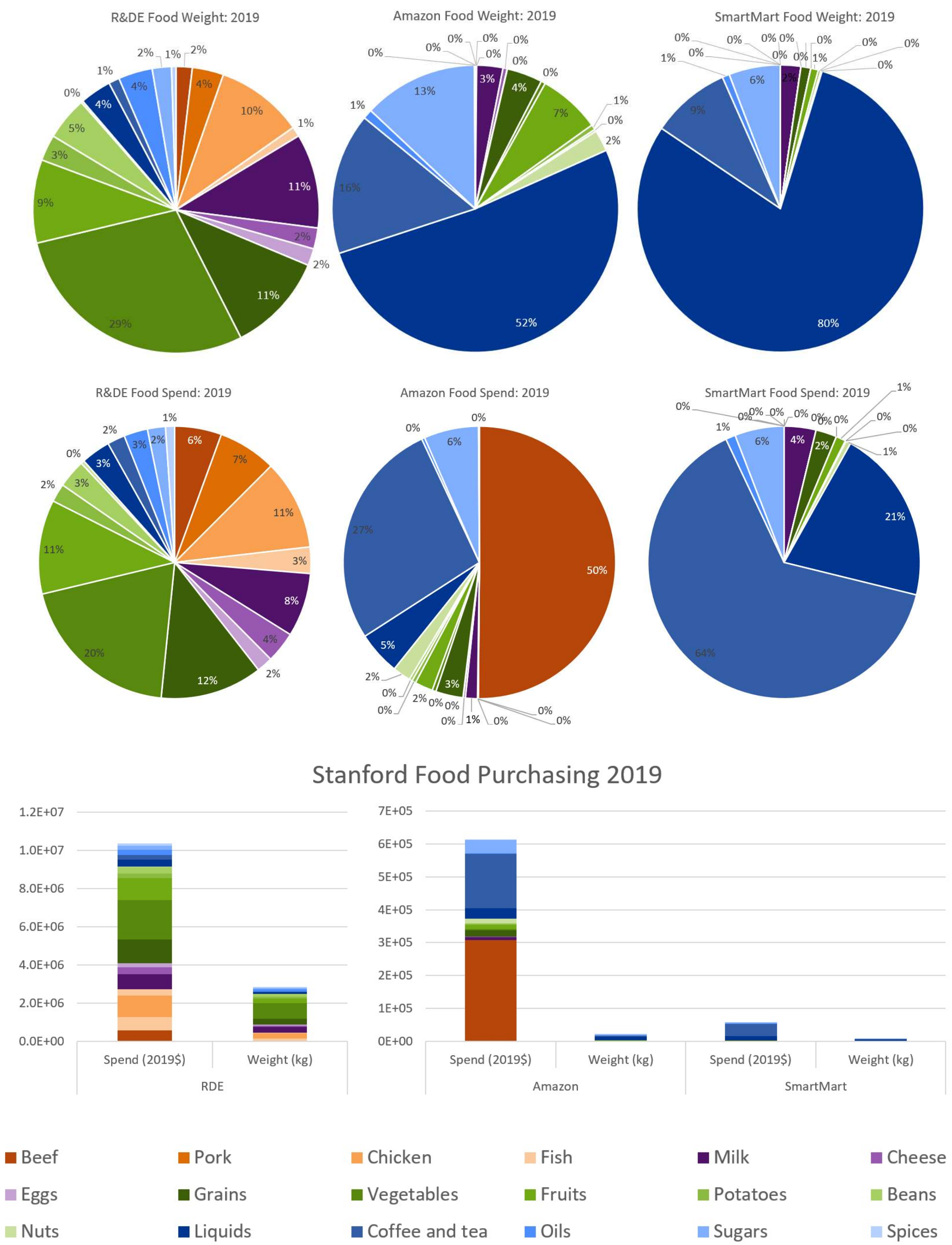


Figure 10. Stanford food weight and spend purchasing distributions.

4.2.1 Weight-Based Emissions Factors

Once weights were assigned to the categories, different emissions factors can be applied to the data set to determine the overall calculated emissions with each emissions factor.

4.2.1.1 Calculating Emissions Using the SIMAP Emissions Factors

First, the emissions factors used by SIMAP were used to determine the total amount of emissions from the three food purchasing sources. SIMAP uses emissions factors from Heller et. al 2014 [4] which uses average emissions for items within a category.

From Figure 11, it is clear to see that meats make up more than 50% of the emissions associated with the food purchases for the dining halls when using the SIMAP emissions factors, even though they make up a minority of the weight purchased whereas the Grains, Vegetables, Fruits, Potatoes, Beans, and Nuts that make up a majority of the weight purchased and represent less than 20% of the emissions.

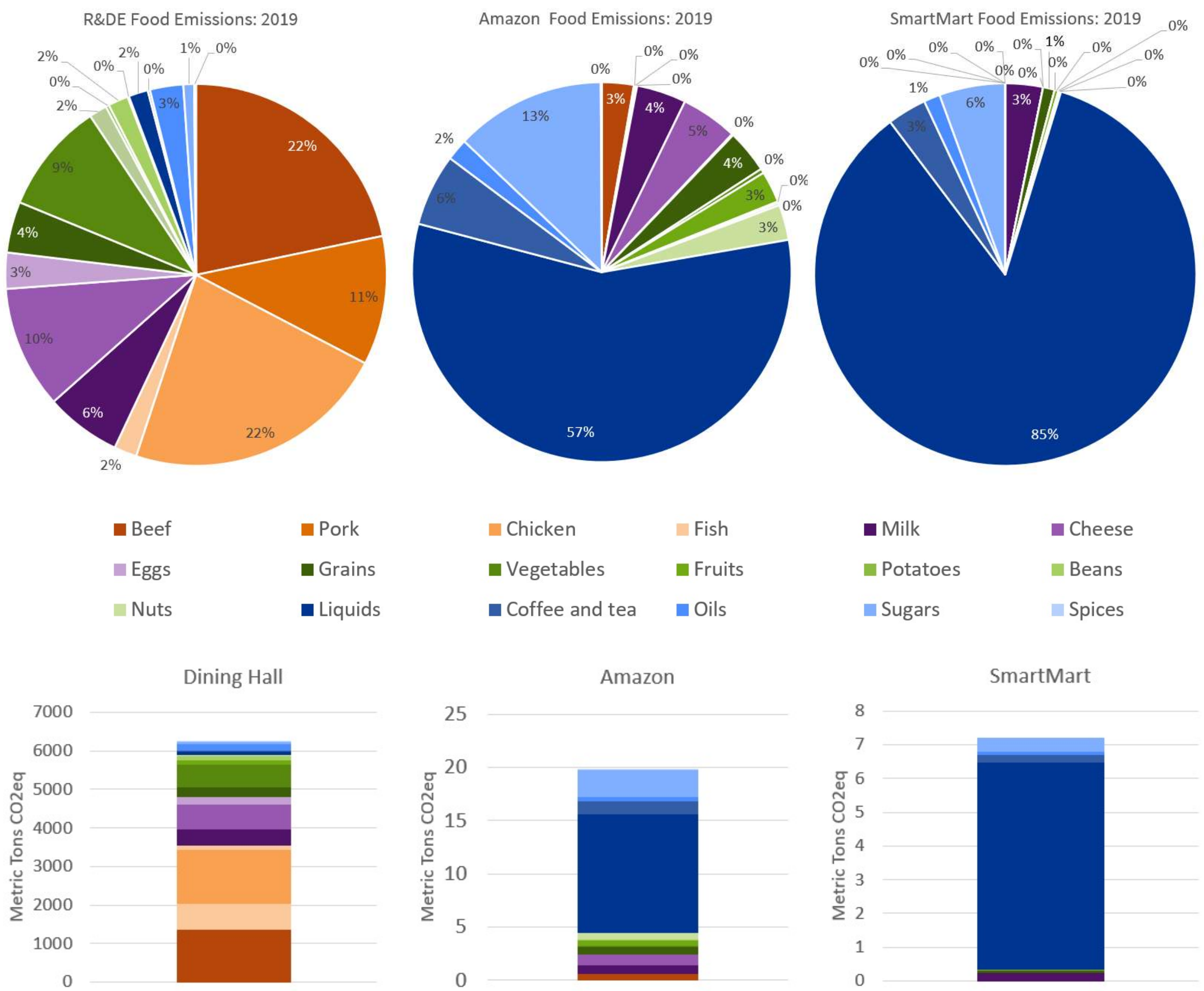


Figure 11. Food Emissions using SIMAP emissions factors.

The emissions associated with the dining hall food purchasing is more than 2 orders of magnitude larger than the emissions associated with the Amazon food or the SmartMart food. Due to a diversified set of foods purchased and the larger number of items purchased, the remaining analysis in this section focuses on the dining hall data and emissions.

4.2.1.2 Calculating Emissions Using Data from Heller et. al 2018

The same author that published the emissions factors that SIMAP uses published an update in 2018 which is further discussed in Chapter 2: Literature Review [15]. This updated paper includes 13 total categories of food items which are slightly different from the SIMAP food categories. See Appendix VI: SIMAP to Heller 2018 Emissions Factors Mapping for mapping from Heller 2018 to SIMAP categories for information related to how the categories from Heller 2018 were mapped to Heller 2014. This set of emissions factors considers cradle-to-farm gate life cycle phases.

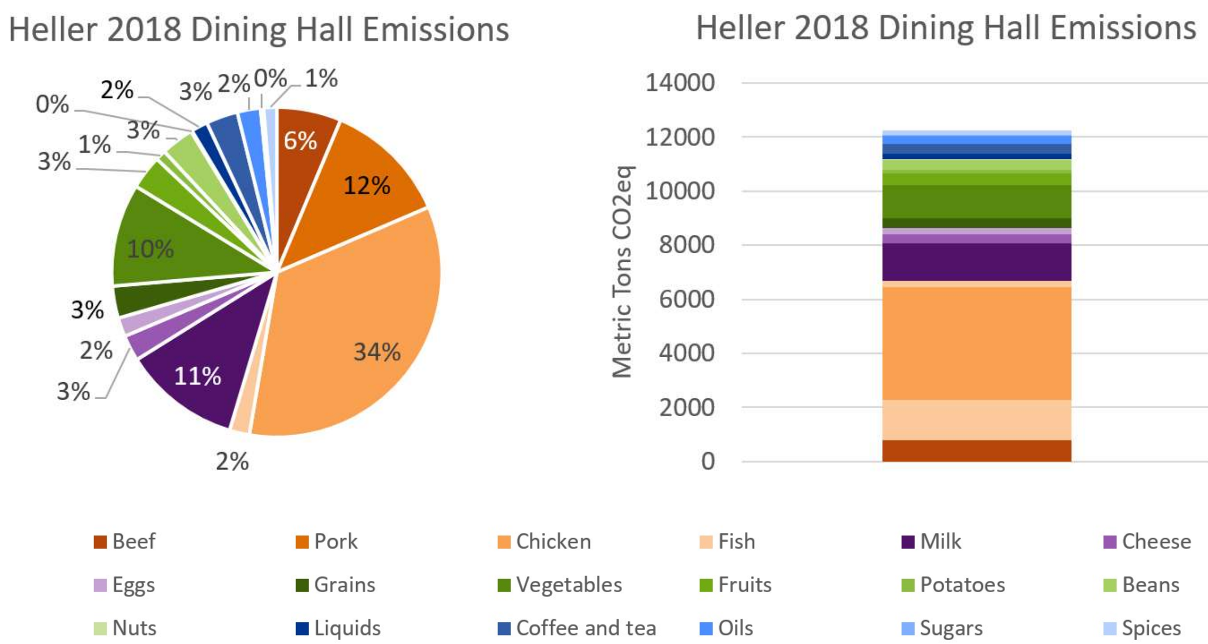


Figure 12. Dining Hall Emissions calculated up updated emissions factors from Heller 2018 [15].

Using only 13 categories instead of 18 means that there is less granularity in the emissions factors, however using updated data can allow advances in technology and farming methods to be accounted for. Using these updated emissions factors increased the total emissions calculated for the dining hall food purchases by 96% and attributes a smaller fraction of those emissions to Beef, but a larger fraction to Chicken and Milk when compared to the SIMAP emissions totals. The SIMAP emissions factors used a less exhaustive literature review in order to develop its emissions factors and the data used was older, thus likely contributing to this significant difference in emissions calculated. When comparing these values to the overarching conclusions of the Heller et al. 2018 paper, meats made up 57% of the average US diet and this data shows that meats make up 54% of the overall diet, showing remarkable alignment. In addition, from the Heller et al. 2018 analysis, in the lowest quintile of dietary emissions poultry was the largest contributor to the meats category at 55% [15], in the data from the dining halls, poultry makes up 63% of the emissions and beef makes up only 11%.

4.2.1.3 Calculating Emissions Using Data from Poore et al. (2018).

Poore et al. includes 59 different categories of food, developed from 2,519 different data points and is further detailed in Chapter 2: Literature Review**Error! Reference source not found.**. The functional units for the emissions factors are 1 kg except for the liquids (beer, wine, milk, and soymilk) which have a functional unit of 1 liter [42]. These liquids were mapped to the SIMAP Liquids category, thus the total weight for the liquids category was converted to liters using the conversion factors available in Appendix III: SIMAP Unit Conversions, which are the same ones used by SIMAP. The mapping used to convert the Poore categories into the SIMAP ones are available in Appendix VII: SIMAP to Poore Emissions Factors

Mapping. The SIMAP Liquids category emissions factor is heavily influenced by fruit juices, but there is no category for fruit juices for Poore et al. (2018), thus it is expected that there will be a difference in the emissions associated with the liquids between the two methods. In addition, there is also no category for Spices in the Poore data set, thus the emissions factor for Vegetables was used for Spices. This is expected to be a under estimation for the emissions for spices because the fresh versions of spices are categorized as Vegetables, and in order to become a spice, a drying process is required which produces emissions.

The individual data from all 2,519 data points is available through the supplementary information from Poore [43], thus there are several methods that can be employed to use all the data points to create appropriate emissions factors. In this section, categories are defined as the 18 categories for which SIMAP has emissions factors, whereas sub-categories are the 59 categories for which Poore defines emissions factors. In the equations below, $\tilde{x}$ refers to taking the median of the data and $\bar{x}$ refers to taking the mean of the data.

For each data point, the emissions associated with each of the following life cycle stages is provided: Land Use Change (LUC) burn, LUC stock, Feed, Farm, Processing, Trans & Str, Packaging, and Retail. There are many ways the data can be aggregated, particularly because Poore and SIMAP have differently defined categories. The total emissions for each data point can be calculated using Equation 9. A data point here is defined as each piece of data that is available for each food item from each paper included in the analysis.

$$Emissions_{data\ point} = LUC\ burn + LUC\ stock + Feed + Farm + Processing + Trans\ \&\ Str + Packaging + Retail$$

Equation 9. Emissions for each data point.

Method 1 for determining the emissions associated with a category leverages Equation 10 to calculate the emissions for each category by determining the median of all the data points within that category.

$$Emissions_{category} = \tilde{x}(LUC\ burn_i) + \tilde{x}(LUC\ stock_i) + \tilde{x}(Feed_i) + \tilde{x}(Farm_i) + \tilde{x}(Processing_i) + \tilde{x}(Trans\ \&\ Str_i) + \tilde{x}(Packaging_i) + \tilde{x}(Retail_i)$$

$$for\ i = all\ data\ points\ within\ a\ category$$

Equation 10. Method 1 for calculating emissions associated with food categories from Poore data set.

The issue that arises with using Equation 10 is that the total emissions for a category will not be representative of any one item within the category since the median of each of the life cycle stages will not likely come from the same food item being analyzed.

Method 2 for determining the emissions associated with a category first requires the calculation of the total life cycle emissions for each data point using Equation 9, and then the median value for the life cycle emissions among all the data points within a category is selected to represent the category.

$$Emissions_{category} = \tilde{x}(Emissions_{data\ point_i})$$

$$for\ i = all\ data\ points\ within\ a\ category$$

Equation 11. Method 2 for calculating emissions associated with food categories from Poore data set.

The issue that arises with using Equation 11 is that it assumes that there is an inherent correlation between the number of LCA studies conducted for a particular food item within a category and the frequency or proportion of that food category that that food item represents. For example, if there are more papers analyzing the LCA of apples than there are LCAs analyzing pears, the data from the apples LCAs will have a bigger influence in the emissions determination for the Fruits category than the data from the pears.

Method 2 is the preferred method because it in fact is how the Poore paper developed the emissions factors for the sub-categories as some of the sub-categories considered in the Poore paper include several types of food such as the Lettuce, Chicory, Endive, and Artichoke sub-category. The median is selected as the most appropriate representation of the data because of the skewed nature of the data set as described in the Poore paper and in Chapter 2: Literature Review.

Method 3 for determining emissions associated with a category involves many steps. First, the median for each life cycle step is determined for each of the sub-categories defined by Poore as illustrated in Equation 12. Then, the average of the subcategories is determined and set as the emissions factor for that category as defined by Equation 13.

$$\begin{aligned} Emissions_{sub-categor} &= \tilde{x}(LUC\ burn_i) + \tilde{x}(LUC\ stock_i) + \tilde{x}(Feed_i) + \tilde{x}(Farm_i) + \tilde{x}(Processing_i) \\ &\quad + \tilde{x}(Trans\ \&\ Str_i) + \tilde{x}(Packaging_i) + \tilde{x}(Retail_i) \end{aligned}$$

$$for\ i = all\ data\ points\ within\ a\ sub - category$$

Equation 12. Calculation of emissions per sub-category based on emissions for each data point.

$$Emissions_{Category} = \bar{x}(Emissions_{sub-category_j})$$

$$for\ j = all\ sub - categories\ within\ a\ category$$

Equation 13. Calculation of emissions per category based on emissions for each sub-category.

The issues with Method 3 are similar to the issues with Method 1 which arise because the median of each of the life cycle stages is determined, thus the final emissions factor does not necessarily represent a real data point or possibility. In addition, because the average among the sub-categories is being used, it is being assumed that all the sub-categories equally contribute to the category, which may not be the case. For example, for the Beef category, Bovine meat from beef herd, Bovine meat from a dairy herd, and Mutton & goat meat, are all contributing equally to the beef emissions factor, although practically, Mutton & goat meat is purchased less often than bovine meat of either type.

Method 4 for determining the emissions associated with a category first requires the determination of the median value for the total life cycle emissions for each sub-category as defined in Equation 14. Then, the average value among the sub-categories is calculated and assigned as the emissions factor for that category as defined in Equation 15.

$$Emissions_{sub-\ category} = \tilde{x}(Emissions_{data\ point_i})$$

$$for\ i = all\ data\ points\ within\ a\ sub - category$$

Equation 14. Method 4 for calculating emissions. First must calculate the relevant values for the sub-categories.

$$Emissions_{Category} = \bar{x}(Emissions_{sub-category_j})$$

$$for\ j = all\ sub-categories\ within\ a\ category$$

Equation 15. Calculation of emissions per category based on emissions for each sub-category.

The issue arises with Method 4 is similar to one of the issues with Method 3 in that it assumes that all the sub-categories are equally representative of the category.

It is also possible to conduct an analysis of emissions for each category and sub-category using the same equations but replacing the median determination with a calculation of the average. It is known that the data available within this analysis is heavily skewed [5], thus the median was determined to be the best representation of the emissions factors for the food categories. However, a discussion of the impacts of using the average emission factors instead of the median emissions factors is available in Section 4.3 Quantifying Uncertainty.

The baseline method to compare the Poore emissions factors was selected as Method 2 which considers all the individual data points collected from all the papers analyzed and uses the median values. The emissions results using Method 2 are presented in Figure 13.

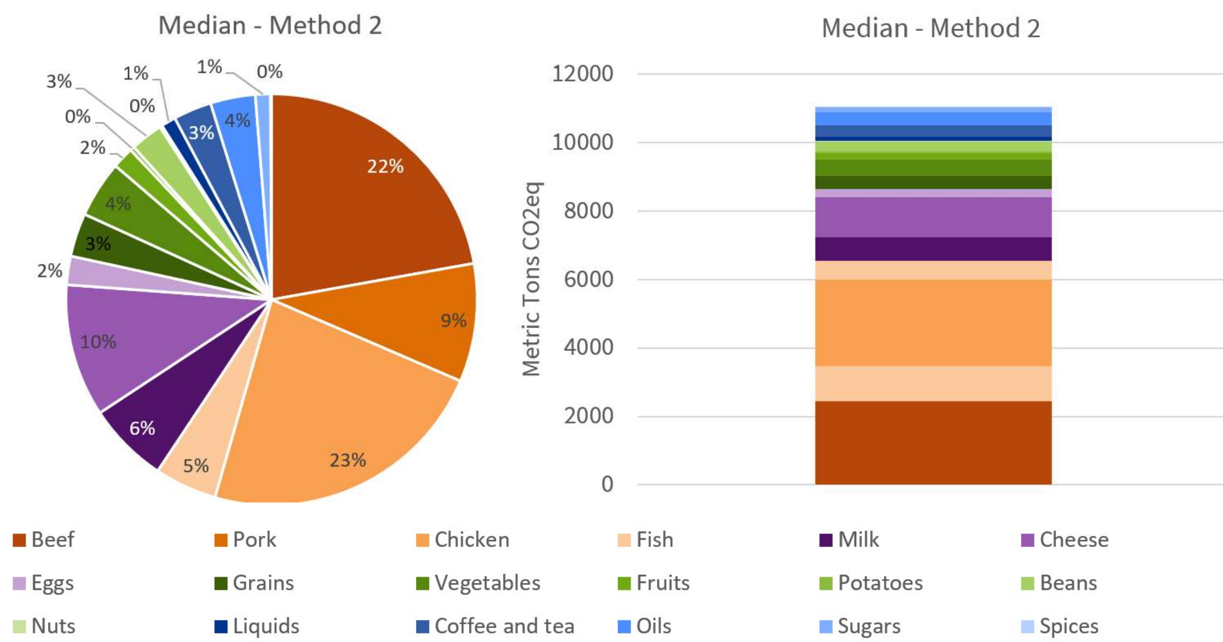


Figure 13. Poore Method 2 Emissions Breakdown.

The emissions calculated using Method 2 are larger than the emissions calculated using the SIMAP emissions factors, but slightly less than the Heller 2018 emissions. However, the percentage of emissions attributed to each category is much more similar in the case of SIMAP and Poore than between Heller 2018 and Poore. In addition, Poore et al. considers cradle-to-grave life cycle phases, so it is important to highlight that if an entity is processing the food on site (i.e. cooking food) and is reporting the emissions associated with processing the food through another avenue, such as the natural gas or electricity and the associated emissions within the kitchens, they must be careful to not double count these emissions.

4.2.1.4 Poore et. al without Land Use Change (LUC) Emissions

As discussed in Section 2.2.4 Land Use Change Emissions, there is still much uncertainty with regards to how to accurately account for LUC emissions as there are many complicated interactions and systems which influence LUC emissions. Thus, to determine the impact of not considering LUC emissions, Method 2 as described in the previous section was used with a slight modification. This modification affects Equation 9 which defines the Emissions for each data point, as was modified as detailed by Equation 16 to exclude all types of LUC Emissions.

$$Emissions_{data\ point\ no\ LUC} = Feed + Farm + Processing + Trans\ \&\ Str + Packaging + Retail$$

Equation 16. No Land Use Change Emissions for each data point.

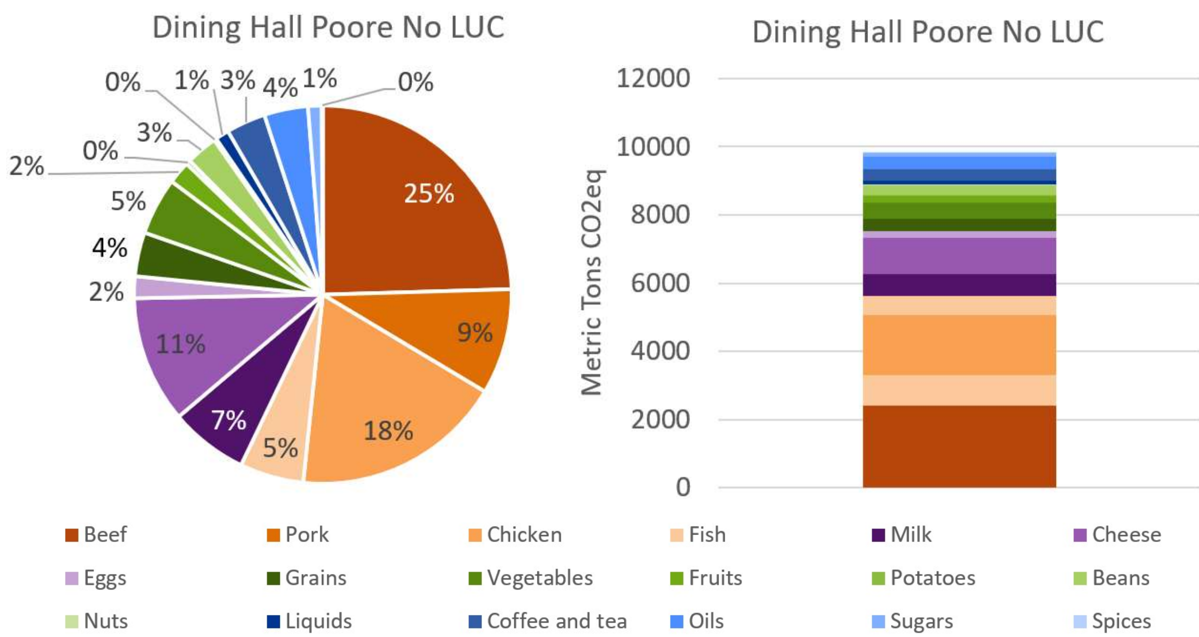


Figure 14. Dining Hall Emissions Calculation using Poore Median Method 2 Emissions without including LUC.

Removing the LUC emissions reduced the total calculated emissions by 10.9% as compared to the calculation including LUC emissions. It most greatly affected the breakdown of emissions associated with the Beef and Chicken categories, but overall, most of the categories remained at the same percentage breakdown as when the LUC emissions are considered. Thus, considering LUC emissions affects the magnitude, but does not necessarily affect the prioritization order for most impactful categories, and mostly affects all the food categories equivalently.

4.2.1.5 Department for Environment, Food, and Rural Affairs (DEFRA) Emissions Factors

The United Kingdom (UK) Department for Environment, Food, and Rural Affairs (DEFRA) released emissions factors for 2019 on 07/31/2020 for the 2019 calendar year. These emissions factors are released such that companies can report their emissions to the government. This source uses the IPCC Fourth Assessment Report over a 100-year period for its emissions factors. These emissions factors are developed specifically for use by UK companies for business conducted within the UK. Thus, it is expected that there will be an inherent difference in the magnitude of emissions as compared to the emissions factors developed specifically for the US and for more global estimates.

DEFRA is a little different from other emissions factors discussed in this research because it defines a single emissions factor for all food and drink. The 2019 emissions factor for food and drink is 4,060 kg CO_{2eq} per metric ton of food and drink purchased [44]. The emissions factors cover extraction, primary processing, manufacturing and transportation to the point of sale, thus it is considered to include cradle-to-gate emissions.

Multiplying the emissions factor by the total amount of weight purchased by the dining halls lead to 11,377 metric tons of CO_{2eq}. This represents an 79.2% increase in estimated emissions as compared to the SIMAP emissions estimate, but is similar to the estimate using Heller 2018 [15] (6.89% decrease) and Median Method 2 Poore [5] (2.15% increase). It is interesting to note that this singular emissions factor is still very much within the range of the emissions estimates of several other weight-based emissions factors which have much more categories. However, this is likely due to the fact that this one emissions factor for all food and drink incorporates an assumed distribution of the types of foods purchased. Thus, it is likely that there would be a much larger difference in total emissions calculated if a less distributed data set was used, such as the Amazon for Business or SmartMart food purchasing data sets.

4.2.3 Using Various Spend-Based Emission Factors

Not all entities keep track of weights of foods purchased, thus it is often useful or necessary to calculate emissions using spend-based emissions factors. Using the tools developed in this work, there are two different ways that the spend based emissions factors can be used for the data sets for the food purchased. First, the categorization script including the keywords and edge cases can be adapted to categorize into the 37 food categories which exist for the US EPA EEIO spend based emissions factors, which are also the categories that exist for the other spend-based emissions factors. The other option is to do a mapping which takes each US EPA EEIO category and maps them to the most appropriate SIMAP category.

Each of these methods has its pros and cons. By adapting the script, one is able to use the more detailed categories which are not available through the 18 SIMAP categories, for example, the US EPA EEIO includes a category for frozen foods which SIMAP does not. However, because no previously categorized data sets for the US EPA EEIO categories were able to be obtained, it was not possible to do an extensive validation of the US EPA EEIO categorization script in the same manner that was conducted for the SIMAP categories. By mapping the categories from SIMAP to US EPA EEIO, one loses granularity, but can be sure of the original SIMAP category assignment with relatively high certainty. The results using these two methods are detailed in the following sections.

One major difference with the spend-based emissions factors as compared to the weight-based emissions factors is that each item is only attributed to one category. In addition, it is important to highlight that there are many prices at which an item can be purchased, and it is important to keep track of if the emissions factors are based on producer price or consumer price and which of these is most appropriate for the current application.

4.2.3.1 Adjustments to Categorization Script

For the spend-based emissions factors, only one category was assigned per item. Thus, there were no edge cases that are included in this categorization. Instead, the order through which the script considers the categories was determined such that the more important or relevant categories were assigned first. For example, the script first searches for keywords associated with the frozen category because they trump other categories, and spices are last, which is only considered for items which were not categorized to any of the other categories. The order of keyword search is available in Appendix VIII: Order of EEIO Categorization.

4.2.3.2 US Environmental Protection Agency (EPA) Environmentally-Extended Input-Output (EEIO) v1.1

In this section, the results discussed are the ones that were obtained using the US EPA EEIO adjusted keywords and edge cases for the categorization script. The results using the mapped categories, both adjusted and not adjusted are discussed in 4.3 Quantifying Uncertainty. The US EPA EEIO is US specific thus introduces an additional difference when comparing this method to the weight-based emissions factors which are more global in their development of emissions factors.

The U matrix described in Section 2.3.2 from the US EPA EEIO v1.1 tables were used as the emissions factor, which are the life cycle impact assessment results and include both the direct and indirect environmental impacts for producing one dollar's worth of a commodity in 2013 dollars. The emissions factors were converted to 2019 dollars accounting for inflation by using an online inflation calculator which considers economy wide adjustments [45]. Then it is multiplied by the total spend per category to determine the total emissions. These prices are representative of producer price, thus they are the prices for producing items in each of these categories. The US EPA EEIO v1.1 factors use 2007 detail Input Output data and the 2013 greenhouse gas emissions for the US. The US EPA EEIO v1.1 tables are widely used and recognized thus provide a good comparison metric across the board.

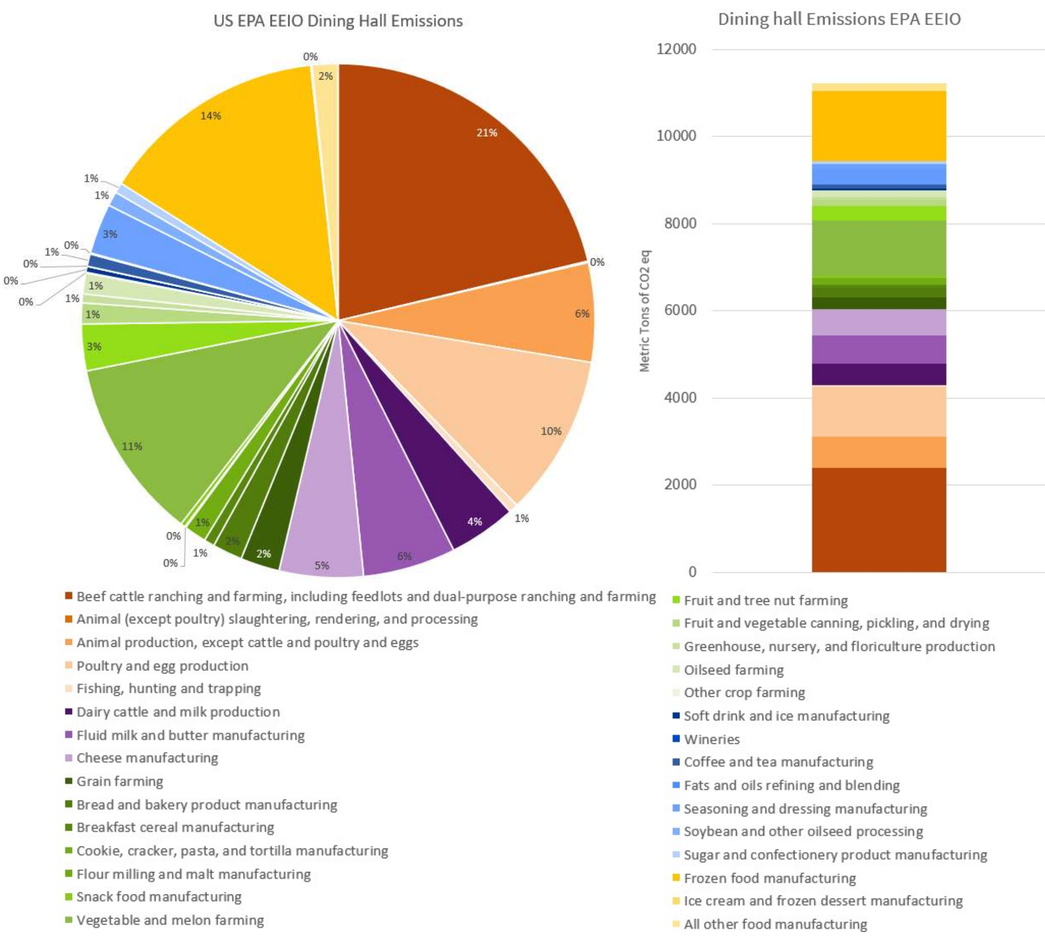


Figure 15. US EPA EEIO Dining Hall Emissions.

The colors for each category were kept as similar as possible to the weight-based emissions graphs, but because of additional categories, more shades were required within some categories, and new categories which did not exist in the previous analysis were added, mainly the yellow categories. These emissions calculations are very similar in magnitude to the emissions calculated with weight-based emissions factors except for the SIMAP emissions factors. In fact, they are only 1.9% less in emissions estimate than Median Method 2 from Poore et al. However, the breakdown of emissions attributed to each category has changed quite significantly depending on the category. For example, Chicken in the SIMAP analysis makes up 22% of the emissions whereas in the US EPA EEIO analysis it makes up only 10%. This is in large part due to the fact that there is a significant amount of emissions that are attributed to the frozen food manufacturing category, which in turn affects the breakdown for all the categories.

Another additional difference inherent in these emissions factors when compared to the weight-based emissions factors includes its consideration of land use change emissions. For example, the US EPA EEIO v1.1 model includes the emissions associated with land use, land use change, and forestry, but considers them as biogenic sequestration using the United Nations Framework Convention on Climate Change (UNFCCC) national GHG inventory conventions [46, 47]. Thus, their impacts to the net carbon emissions calculation are to decrease the total value reported. Finally, these methods do not include use-phase or end-of-life emissions, unlike some of the weight-based factors.

4.2.3.3 Supply Chain Emissions Factors (SEF)

For this analysis, the US EPA EEIO adjusted keywords and edge cases categorization script was also used, as the categories are the same for both sets of emissions factors. The EPA has also more recently published work with supply chain emissions factors [48]. These were also developed using USEEIO models, and thus are US specific. It includes both direct and indirect GHG emissions associated with cradle-to-point of sale in 2018 dollars. These prices are representative of purchaser price, thus is expected to represent the food data sets more appropriately since the universities are purchasing the food to prepare, not producing the food itself. The emissions factors are in kg CO_{2eq} and the conversions to CO_{2eq} are based on the IPCC AR4 global warming potentials for methane and nitrous oxide which are listed in Appendix IX: Carbon Equivalence Conversions. The factors were developed using 2012 Input output data and 2016 greenhouse gas emissions for the US.

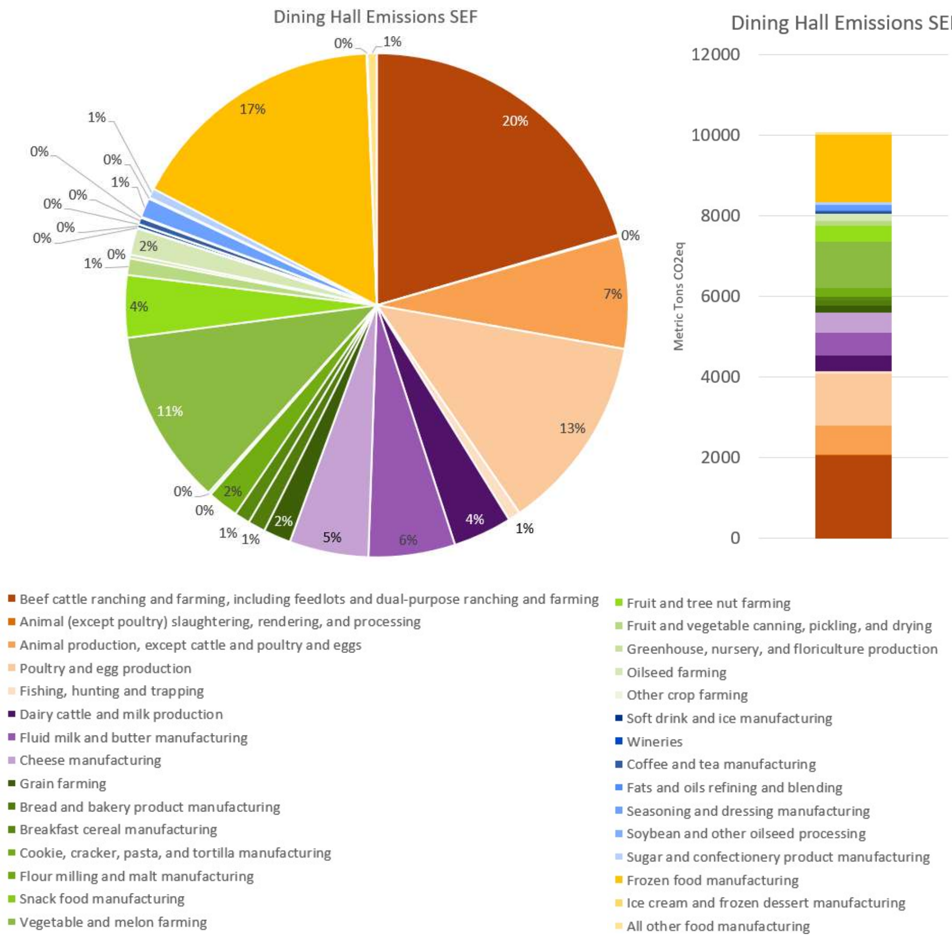


Figure 16. Dining Hall Emissions using Supply Chain Emissions Factors.

The total magnitude emissions calculated are 61.7% higher as compared to the SIMAP weight-based emissions calculation, 17.5% lower than the Heller 2018 emissions calculations, and 9.5% lower than the Median Method 2 Poore calculation. In addition, they are 10.2% lower than the US EPA EEIO v1.1 total emissions calculated. Again, the magnitude of the total emissions has changed as well as per category, but the prioritization, or largest contributing emissions categories, have largely remained the same between the US EPA EEIO v1.1 and the SEF emissions factors. However, because the SEF are much more recent, they are not as widely used nor as robustly checked as the US EPA EEIO v1.1 simply due to the fact that they are newer.

A major difference between the SEF emissions factors and the US EPA EEIO v1.1 model is how they handle biogenic sequestration. In the US EPA EEIO v1.1 model, forest sequestration was assigned to the forestry sector, however the locations in which the sequestration is occurring in the US are not actively harvested and used within the forestry sector [46]. Thus, in the SEF emissions factors biogenic sequestration was removed to avoid a misbalance of carbon accounting [46]. However, this should not impact the food emissions factors significantly as the largest portion of this change directly affected the forestry sector.

4.2.3.4 VitalMetrics

VitalMetrics is the first of the private tools to be discussed in this analysis. By this, it is meant that the tool is a black box where the emissions factors are not directly available. Instead, the categorized data was provided to VitalMetrics, and a summary report was made available which detailed the total emissions calculated for the entire provided data set along with the percentage of emissions for the top 9 categories.

For analysis of emissions using the VitalMetrics tool, the US EPA EEIO adjusted keywords and edge cases categorization script was also used because it uses the same categories which VitalMetrics uses. The category names have slight differences, but they represent the same categories. For this analysis, the country was set to the US. For this analysis, VitalMetrics used CEDA 5 as their emissions factors source.

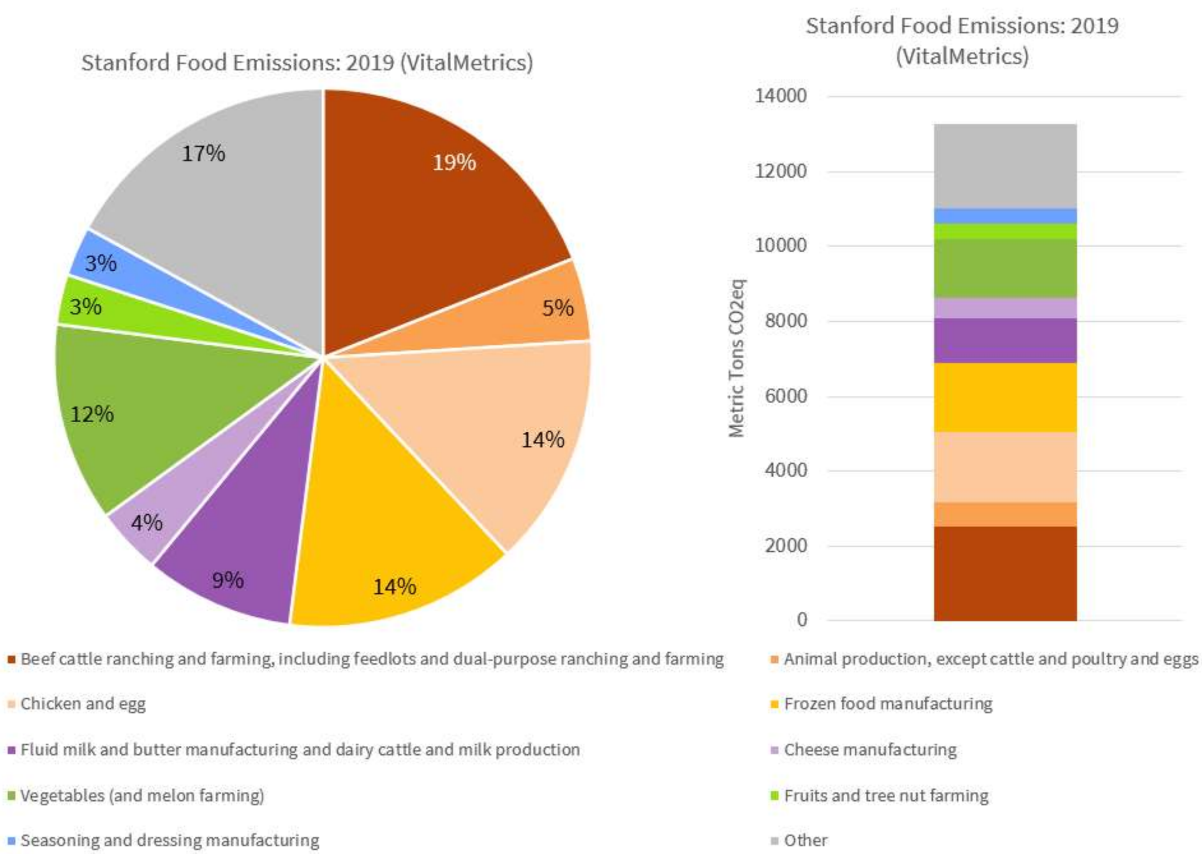


Figure 17. Total Stanford Emissions Breakdown from VitalMetrics Analysis. Includes Foods purchased through dining halls, Amazon for business and SmartMart.

The emissions above include the aggregate emissions from the dining halls, Amazon for Business and SmartMart, however, it was detailed that 98% of the total emissions came from the dining hall data. Thus, the estimate total emissions for the dining hall using VitalMetrics is 13,004 metric tons of CO_{2eq}. This is a

77% increase in emissions when compared to the total emissions calculated using SIMAP emissions factors, a 6.4% increase as compared to Heller 2018, and a 17% increase as compared to Median Method 2 Poore. The percentage of emissions attributed to each category remains mostly the same across the board, with an increase in magnitude of emissions. The VitalMetrics emissions were calculated using the GHG Protocol: Corporate Value Chain (Scope 3) Accounting and Reporting Standard. The emissions do not include biogenic CO_2 emissions nor trades. The CEDA database was modified in order to represent conditions present in 2019 to harmonize with the data year.

4.3 Quantifying Uncertainty

As can be noted from the section above the total estimated magnitude of emissions can have a variance of up to 2x depending on source and method. Although the range can be considered relatively large, both the minimum and maximum estimated emissions are the same order of magnitude which is encouraging. In addition, it is interesting to note that both the highest and lowest estimate are achieved using weight-based emissions factors. Among the spend-based emissions factors, the estimated total emissions range from 10,083 to 13,004 metric tons of CO_{2eq}, a much more narrow range which shows relatively good agreement between the different spend-based emissions factors. The average total estimated magnitude of the spend-based emissions factors is 11,440 metric tons of CO_{2eq}, whereas the same average for the weight-based emissions factors is 10,216 metric tons of CO_{2eq}, again showing remarkable agreement between the methodologies. However, there is still some uncertainty as to what the true magnitude of emissions is and what the uncertainty is with regards to these values which will be further discussed in this section. The estimates for the total emissions are so close, but it is important to remember that these sources consider different life cycle stages in their calculations, different geographies and different data vintages.

4.3.1 Calculation of Emissions per Dollar per Food Category

One way to quantify some uncertainty with regards to using spend-based and weight-based emissions is to back-calculate one using the emissions from the other. Several of the data sets provided from the partner universities through the SIMAP NWG included both weight and spend data, with the relevant year for the data set presented in Table 8. All the spend data was converted to 2019 dollars by adjusting for inflation using an online inflation calculator [45].

Table 8. Data set years for various provided data sets.

University	A	B	C	D	E	F	G	H
Data Set Year	2019	2018	2017	2018	2018	2019	2015	2019

By using the SIMAP emissions values and determining the total emissions per category for each of these data sets, and then dividing these total emissions by the total amount of weight of food purchased in that category for each university, a back-calculated spend-based emissions factor could be determined for each food category for each university. These results are shown in green Figure 18. The blue squares in the same figure represent the different spend-based emissions factors found in literature. Some of the food categories have a very small spread between all the back-calculated and spend-based emissions factors, such as nuts, liquids, and sugars. However, other food categories have a wide spread in emissions such as beef, cheese, and eggs. This can be due in part to the differences in qualities of the foods purchased, which affect the price per kilogram of the food purchased. For example, the lowest back-calculated emissions per dollar spent for beef is associated with the beef purchased from Amazon for Business and was made up of jerkies (beef jerky, turkey jerky, and bacon jerky). These beef products are lower quality and thus are cheaper than regular beef, and the difference in weight is not so significant, thus the emissions per kilogram are much lower for these purchases of beef as compared to regular beef purchases. This illustrates that the more diversified a food data set that is being analyzed, the less likely it is to encounter large differences in emissions using weight- vs spend-based emissions factors. This is further corroborated by a deeper analysis of the Amazon for Business food purchases.

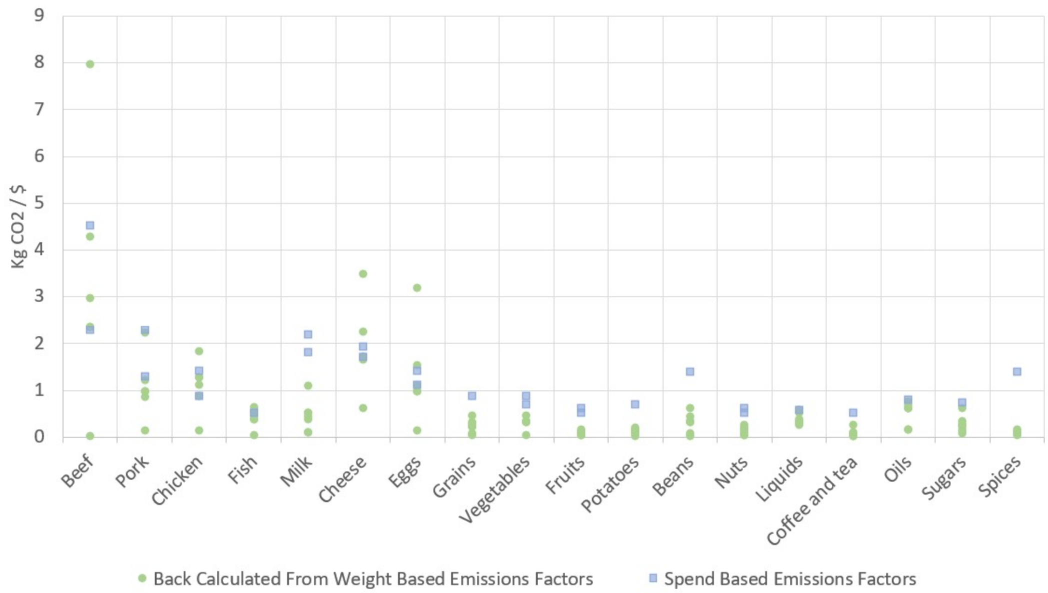


Figure 18. Back-calculated kg of CO2 per 2019 dollar spent using data from various universities for which spend data was available. Blue squares represent spend-based emissions factors found in literature whereas the green points represent the back-calculated spend-based emissions factors for the university data sets where both spend and weight data were available.

4.3.2 Analysis of Impacts of Grouping Poore Categories into SIMAP Categories

As discussed in depth in Chapter 4.2.1.3, there are many ways to aggregate the more numerous Poore categories into the SIMAP categories because the detailed information is available for every data point gathered for the analysis. Previous results using the Poore et al. emissions factors used the median method 2 methodology. Using both the median values and the average values and Equation 9 through Equation 15, a total of eight methods can be developed to use the emissions factors. The impacts of using these methods are detailed in Figure 19 and Figure 20.

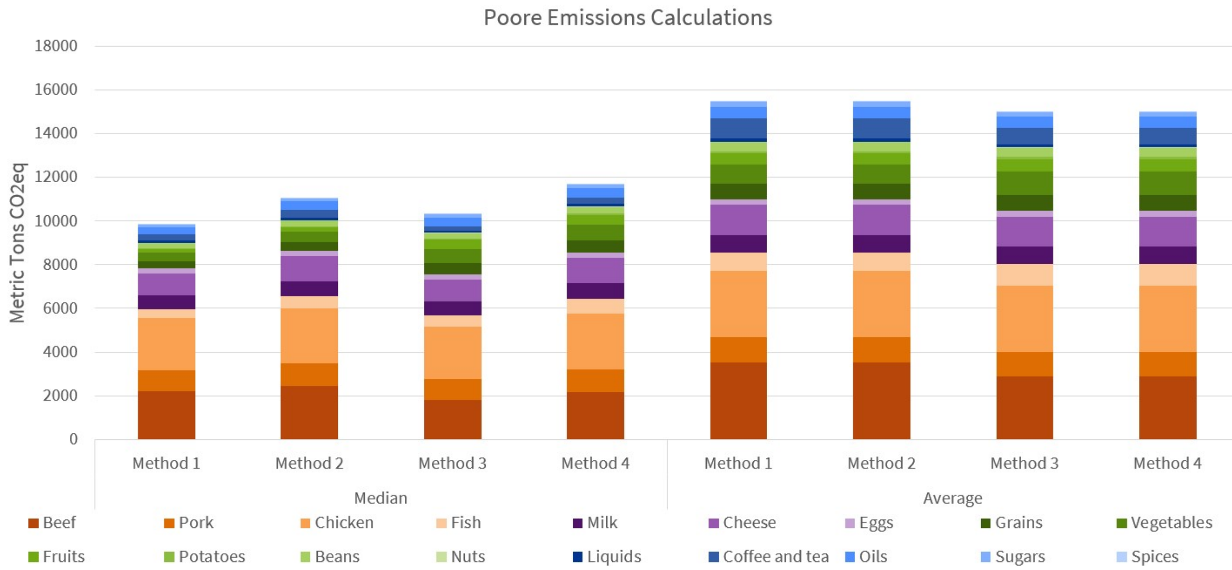


Figure 19. Comparison of Emissions calculations from Poore et al. values.

Due to the skewed nature of the emissions factors, the average emissions factors are much higher for each food category when compared to the median values as shown by Figure 19. This causes the minimum emissions calculated to go from 9,943 metric tons of CO_{2eq} using Method 1 for median values to 15,104 metric tons of CO_{2eq} using Method 3 and 4 with the average values. This represents a 51.9% increase in the calculated emissions. Among the Median Methods, there is some spread, but the highest estimate represents an increase of 18.5% in emissions as compared to the lowest estimate. Among the Average Methods, the highest estimate of emissions is only 3.2% higher than the lowest. Many of the other weight-based emissions factors use the average of the LCA emissions for the items in each category, which according to this analysis can estimate the emissions at a higher level than when compared to when the Median values are used.

The percentage breakdown for each method as illustrated in Figure 20, shows some difference between the percentage of responsibility attributable to each food category between the methods, but not as large a difference as the magnitude of emissions. This indicates that if the purpose of examining emissions is to determine the largest contributors to emissions or the ranking of contributors, any of these methods for developing emissions factor should be able to generally accurately provide a response. However, if the question being asked is what the magnitude of emissions is, then the emissions factor used can make a large difference.

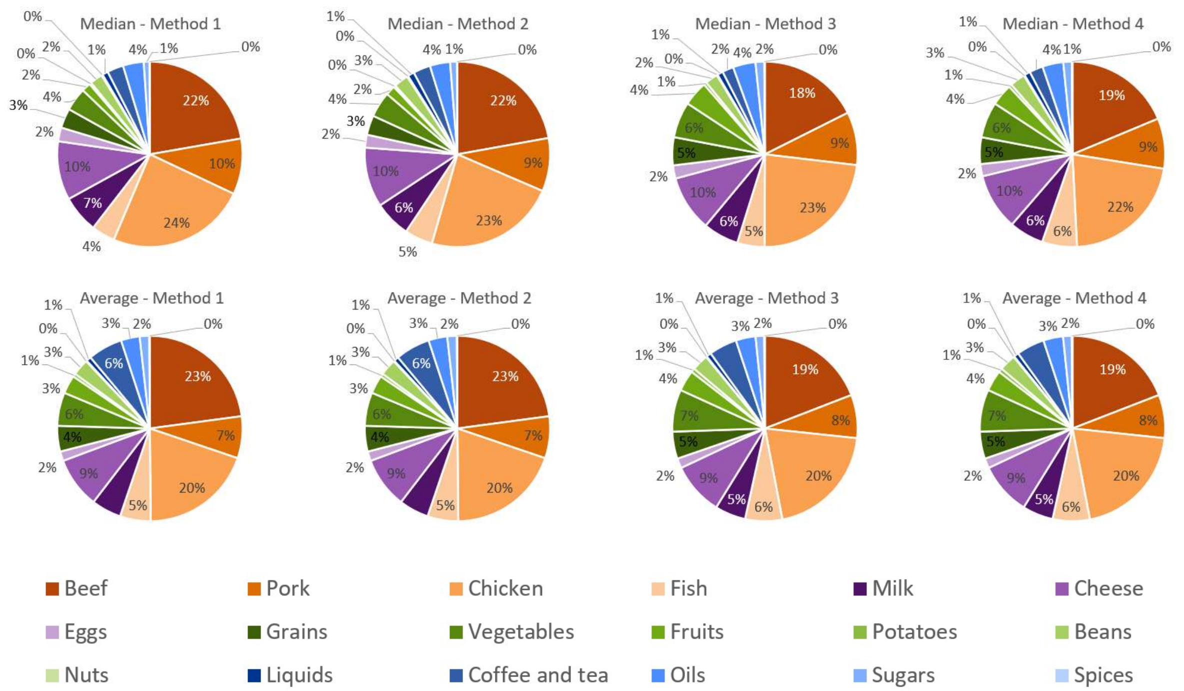


Figure 20. Emissions Breakdown with various methods using Poore et al. data.

4.3.3 Analysis of Impacts of Adjusting Heller 2018 Emissions Factors

For the Heller 2018 emissions factors, each data point used to determine the emissions for each of the categories considered was also available, much like in the Poore data set. However, the total number of items contributing to each category was much smaller in the Heller 2018 data set, thus each individual point has a larger impact on the emissions for that category. Of particular significance is the Other food category, which includes data from 23 different data points. For the previous analysis discussed in this work, the emissions for the dining halls were calculated using the average of all 23 data points for the Other category. However, three points within this category have a significantly higher emissions factor associated with them, and likely do not make up such a significant portion of the items purchased in the Other category. Thus, Cocoa, powder; essential oil, lemon; and essential oil, orange were removed in order to conduct a secondary analysis using the Heller 2018 emissions factors because including these three data points in the average increases the average for the Other category by 6.5 times. The Heller et al. 2018 categories are less granular than any of the other emissions factors sets that were considered for this analysis, thus there are many more occasions of one emissions factor from Heller et al. 2018 being assigned to multiple SIMAP categories, including the Other category. Thus, as seen in Figure 21, excluding the three data points that significantly increase the emissions associated with the Other category has an impact on the total emissions calculated and reduced the total magnitude of emissions by 3.7% as compared to using the average of all the points for all the categories.

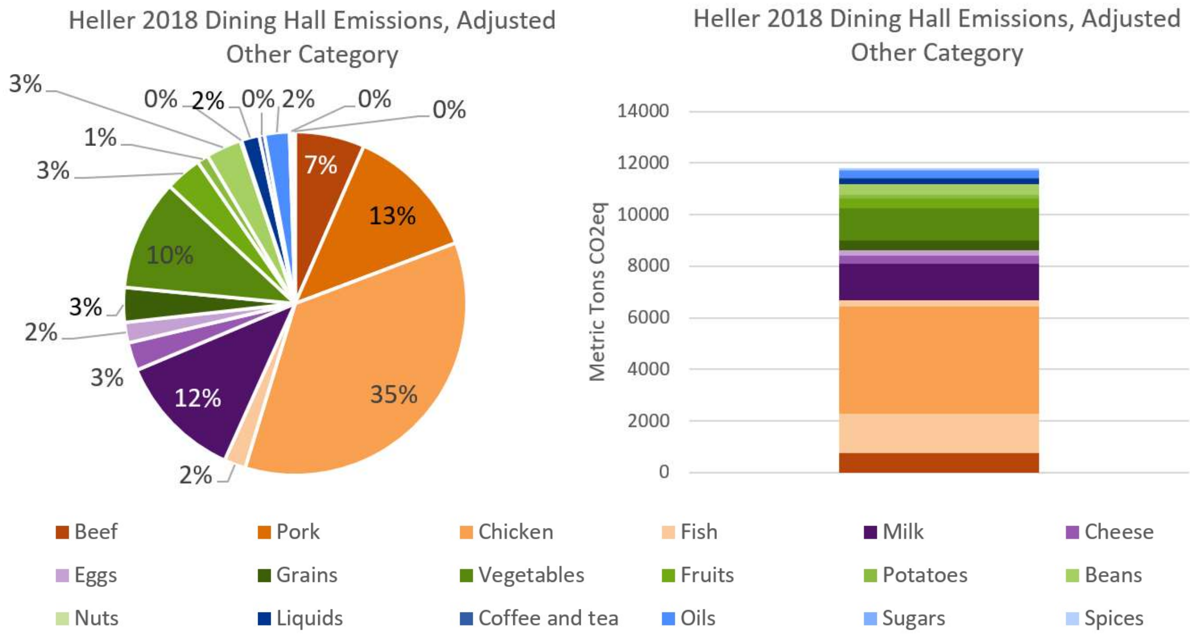


Figure 21. Heller 2018 Dining Hall Emissions with Adjusted Other Category.

4.3.4 Analysis of Using EPA EEIO Categorization Script versus Mapping EPA EEIO categories to SIMAP categories

As discussed at the beginning of Section 4.2.3, the categorization for the spend-based emissions factors can be done in two ways. The categorization script can be adapted to categorize food purchases into the 37 food categories which exist for the US EPA EEIO spend based emissions factors, which is what was used for the analyses up to this point. The other option is to do a mapping which takes each US EPA EEIO category and maps them to the most appropriate SIMAP category. Because no previously categorized data sets for the US EPA EEIO categories were able to be obtained, it was not possible to do an extensive validation of the US EPA EEIO categorization script in the same manner that was conducted for the SIMAP categories. By mapping the categories from SIMAP to US EPA EEIO, one loses granularity, but can be sure of the original SIMAP category assignment with relatively high certainty. The results from mapping the categories is presented in Figure 22.

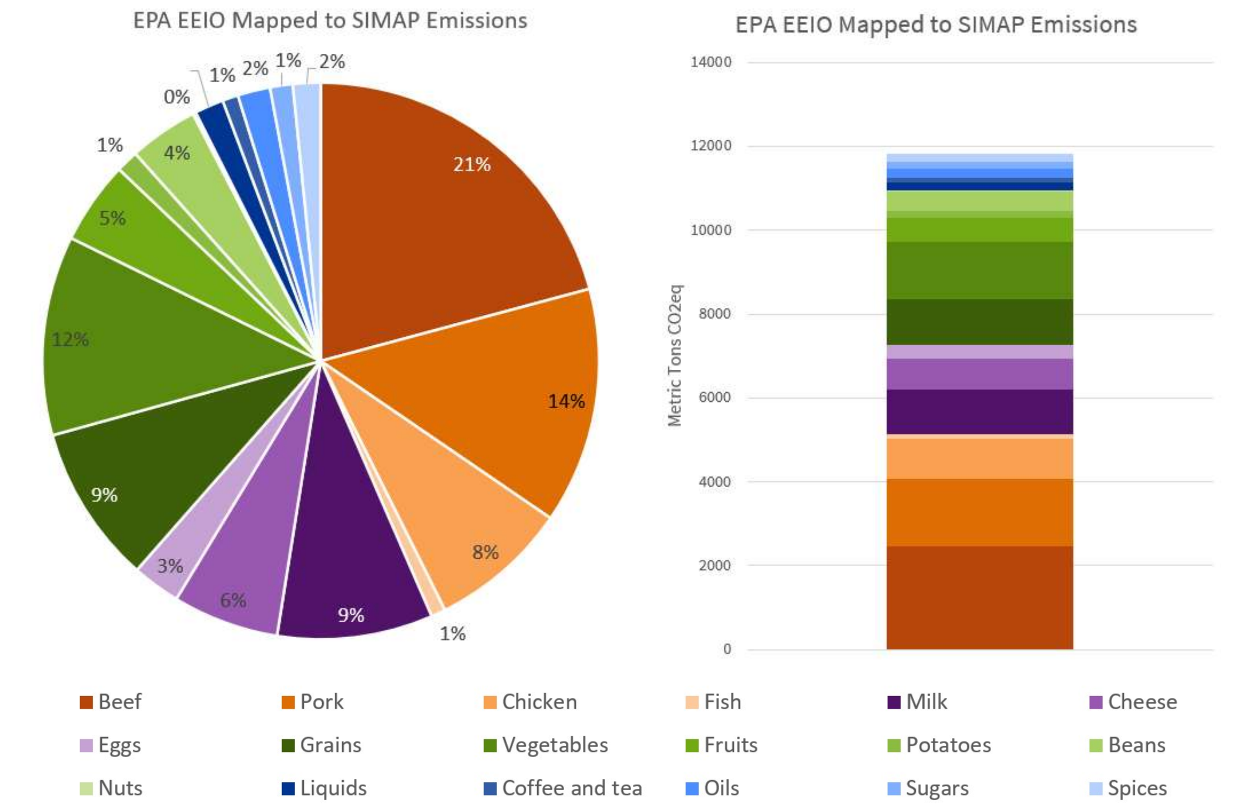


Figure 22. Spend-based emissions factor using mapping from EPA EEIO Categories to SIMAP Categories.

Using the mapped categorization to of EEIO emissions factors to the SIMAP categories reduced the total emissions estimate for the data set to 11,813 kg $CO_{2\text{-eq}}$ or a 5.2% increase from using the EEIO emissions factors and categories directly. Most categories have stayed relatively similar, but Pork has a larger proportion of emissions when compared to weight-based emissions while Chicken has a smaller contribution to the overall emissions. Pork is likely larger because the category for emissions in EEIO is for animal (except poultry) slaughtering, rendering, and processing, which is much more general than a Pork specific emissions factor available through other sources. Milk has grown from 5% in the EEIO categorization to 9% in this mapping, likely due to the fact that ice cream and butter are two separate categories in the EEIO categorization.

4.3.5 Analyzing Amazon for Business Food Emissions in More Depth

As mentioned previously, the types of food purchased through the Amazon for Business are not as varied as what is purchased for the dining halls because it mostly represents purchases done for snacks for meeting and events. Thus, any assumptions made on general food purchasing for these emissions factors are less appropriate for any kind of skewed food purchasing such as the Amazon for Business data set. Table 9 shows the emissions calculated for the Amazon for Business data set using three weight-based emissions factors (SIMAP, Heller 2018, and DEFRA) as well as three spend-based emissions factors (EPA EEIO 1, EPA EEIO 2, and VitalMetrics).

Table 9. Amazon for Business Emissions Calculation Using Various Emissions Factors.

	SIMAP	Heller 2018	DEFRA	EPA EEIO 1	EPA EEIO 2	VitalMetrics
Total Estimated Emissions (metric tons CO2eq)	19.61	70.31	85.22	188.24	213.66	265.40

Differences between these methods for this data set are an order of magnitude different, thus the impact of using different emissions factors in this case is much more significant than the previous differences noted in the dining hall data set. This indicates that the larger the data set and the more varied the types of food purchased the less the method or emissions factor matters if determining the order of magnitude of emissions is the goal. Once one attempts to analyze smaller or less varied data sets, the differences become more apparent. The Heller 2018 emissions factor used is the average of all the items in each category. The EPA EEIO 1 emissions leverage the mapping of the EPA EEIO categories to SIMAP categories whereas the EPA EEIO 2 emissions use the less validated categorization directly into the spend categories for the EPA EEIO.

4.4 Outlier Analysis

An outlier analysis was conducted to attempt to quantify the possible errors present with regards to the weight assignment for the different items purchased. Three scenarios were considered for options of what to do for the line items deemed to be an outlier for each food category.

Scenario 1 is the base case where all the weights and spend data are considered to be accurate. Up to this point in this work, this is the scenario that was utilized. Scenario 2 is where, through the use of the determination of outliers as explained in Section 3.9 Determining Outliers, any item that is considered an outlier is removed from the data set. Scenario 3 is where all the items that were considered outliers are allocated a new weight according to Equation 6 and the line items with the new weight are included in the data set. Scenario 3 assumes that the spend for these line items are correct and that the issue is the weight assignment. Although it is possible the spend is incorrect, this is less likely than the weight being incorrect due to the prevalence of keeping track of spending for budgetary reasons.

Because it is hard to analyze outliers for multi-ingredient items, only single ingredient items were considered for this analysis. The total spend on one ingredient items was \$8,802,915 out of a total spend of \$10,364,968. Therefore, the percent of spend on one ingredient items is 85%. Thus, there is still the possibility of incorrect weights assigned to multi-ingredient items and still uncertainty associated with the range of emissions calculated with the outlier analysis.

Table 10. Summary statistics for each food category in R&DE Dining Hall data set.

	Average	Median	Stdev	Q1	Q2	Q3
Beef	18.17	13.99	12.54	8.60	13.99	24.27
Pork	21.87	8.48	94.25	6.51	8.48	12.25
Chicken	11.74	6.17	26.48	4.23	6.17	11.27
Fish	21.22	15.71	27.37	10.16	15.71	24.14
Milk	5.36	3.63	5.49	2.44	3.63	6.63
Cheese	9.64	7.55	6.32	5.47	7.55	10.57
Eggs	23.36	3.51	71.85	2.35	3.51	5.23
Grains	12.23	5.39	75.73	3.14	5.39	10.11
Vegetables	8.96	5.00	22.90	2.46	5.00	6.96
Fruits	9.18	6.39	12.23	3.28	6.39	9.92
Potatoes	8.44	5.57	6.34	4.46	5.57	12.68
Beans	17.87	3.79	88.59	1.88	3.79	9.03
Nuts	13.90	11.11	10.68	6.17	11.11	19.41
Liquids	3.43	2.24	4.26	1.43	2.24	4.07
Coffee and tea	57.60	20.87	72.79	8.84	20.87	68.27
Oils	8.40	4.73	11.70	2.98	4.73	10.14
Sugars	6.48	4.65	4.95	2.39	4.65	8.81
Spices	36.50	19.34	106.23	12.54	19.34	27.69

A high standard deviation value indicates that there is a lot of variability in the prices per kilogram of the items in that category. In some cases, this is in part due to a variety of items which are in the category, such as coffee and tea also includes chocolate, but in other cases this is likely due to items where the weights were not accurately considered. For example, the five highest standard deviations are for the categories: Spices, Pork, Beans, Grains, and Coffee and tea. In the case of Pork, there is a single line which has a dollars per kilogram of $837/kg of pork purchased, whereas all the other outliers are below $90/kg and most are below $50/kg. More information on the lines that were deemed to be outliers is available in Appendix X: Items Determined to be Outliers from R&DE Data Set. A statistical summary of the items considered to be outliers is available in Table 11.

By using the standard deviation and the first and third quartile designations for each food category, outliers were determined using Equation 4 and Equation 5. There were no outliers found that were too small in terms of spend per kilogram. Table 11 shows the number of outliers for each category as well as the percentage of the spend that is attributable to the items deemed to be outliers (only for single ingredient items).

Table 11. Summary of Outlier information per food category.

	Number of outliers	% spend on outliers	new estimated weight (metric tons)	original weight (metric tons)
Beef	1	0.497342	0.200735	0.043545
Pork	7	1.522776	1.262325	0.231863
Chicken	3	0.024249	0.042052	0.009636
Fish	3	0.520507	0.105960	0.031904
Milk	5	0.223960	1.034575	2.742439
Cheese	11	1.278472	0.516550	0.166768
Eggs	1	0.116926	0.127068	0.001701
Grains	51	3.697458	6.765711	1.384195
Vegetables	38	1.429539	5.005542	1.401134
Fruits	11	0.517573	0.818970	0.204394
Potatoes	2	1.114948	0.216424	0.037195
Beans	5	1.550676	1.195102	0.159664
Nuts	1	0.487408	0.013560	0.002722
Liquids	5	1.205994	1.103145	0.161792
Coffee and tea	4	3.082048	0.208789	0.017577
Oils	4	0.124231	0.059245	0.008603
Sugars	2	1.173385	0.118226	0.026792
Spices	17	9.378843	0.672609	0.103462

Scenario 2 reduced the total emissions for the food purchased through R&DE from 6,348 metric tons of CO_{2eq} using the SIMAP emissions factors to 6,337 metric tons of CO_{2eq}. This represents a -0.17% reduction in the estimated emissions. This is in part a small value because there were no outliers where the $/kg were found to be too small, thus all the outliers were cases where a proportionally small amount of weight was attributed to a large amount of spend. This means that the emissions associated with these outliers were being underestimated, and by simply removing these line items the reduction in emissions is small because the emissions were being estimated using a weight-based emissions factor and the weights were small.

In Scenario 3, the emissions determined were larger than for either of the other two scenarios at 6,371 metric tons of CO_{2eq}, or a 0.37% increase in emissions. Although this value was a small percent of the total emissions, as can be seen in Table 11, there were almost 8 tons of food purchased that were considered outliers, and when adjusted, this weight became almost 20 tons of food. This analysis could lead to a larger impact if the errors in weight were concentrated in the food categories with higher emissions. Thus, it is still important to determine the outliers to ensure that the data set does not have many outliers in these more impactful categories.

Chapter 5: Conclusions and Future Work

5.1 Conclusions

Overall, the differences between the total emissions calculated using emissions factors for large data sets was much smaller than expected as illustrated in Table 12 and Figure 23. However, it was realized that the distribution of the types of food items purchased will affect the magnitude of difference that will be present between weight- and spend-based emissions calculations. The more varied the types of foods purchased, the smaller the difference will be between the methods as is discussed in Section 4.3.2.

Table 12. Dining Hall Emissions Summary.

Emissions Factor Source	Method	CO2eq (Metric Tons)	Comparison
Heller et al 2014	Weight-based	6,348	0.6x
Poore et al 2018 with LUC (Median Method 2)	Weight-based	11,138	x
Poor et all 2018 no LUC	Weight-based	9,884	0.9x
Heller et al 2018 (not modified)	Weight-based	12,219	1.1x
DEFRA	Weight-based	11,377	1.0x
VitalMetrics	Spend-based	13,004	1.2x
US EPA EEIO (categorizing to EEIO categories)	Spend-based	11,232	1.0x
US EPA EEIO (mapping categories from EEIO to SIMAP)	Spend-based	11,813	1.0x
SEF	Spend-based	10,084	0.9x

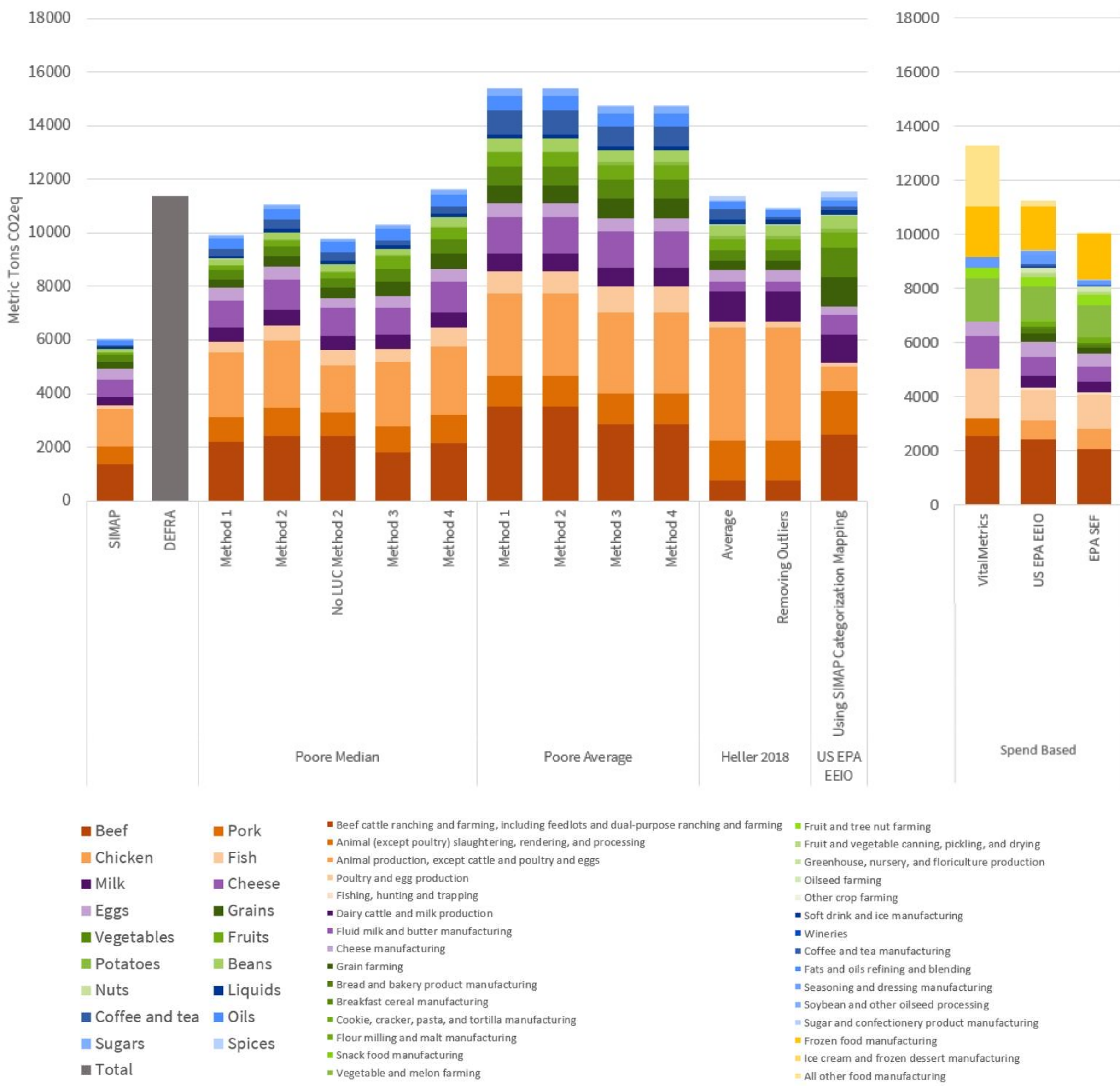


Figure 23. Comparison of all emissions estimation methods. The DEFRA Emissions methodology does not allow a breakdown per food category type which is why it is a single bar that represents the total emissions.

There is also generally not a large difference between the methods for the order of categories with the largest contributions except for the Heller 2018 methods which both estimate a larger portion of emissions associated with Chicken and Milk and less for Beef as can be seen in Figure 24.

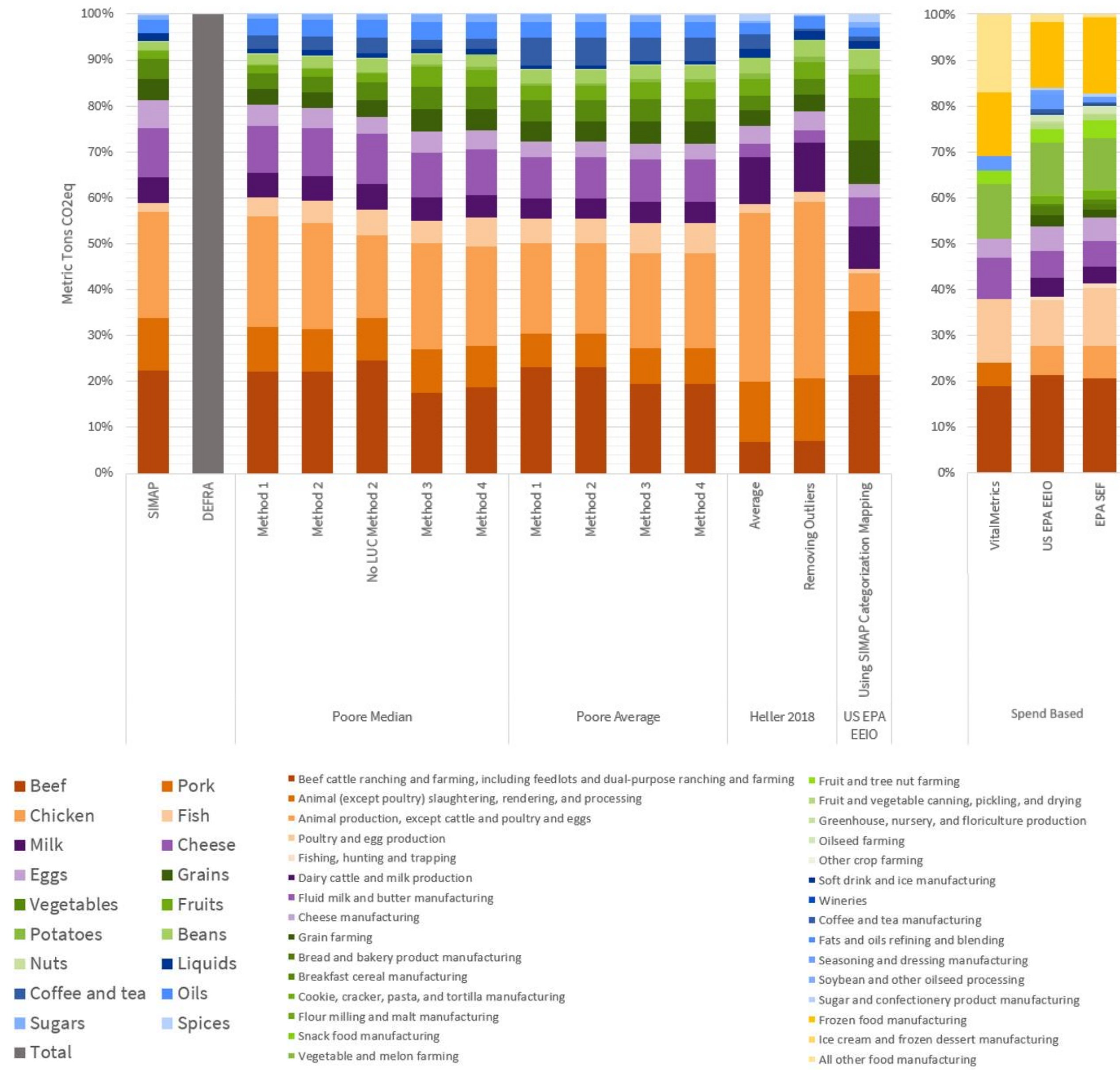


Figure 24. Percentage breakdown for all emissions estimation methods. The DEFRA Emissions methodology does not allow a breakdown per food category type which is why it is a single bar that represents the total emissions.

Although the differences in magnitude between the methods are not so large, there are differences in the LCA assumptions that are being made between these methods, such as how to deal with land use change emissions if they are being dealt with at all, how detailed or not the categories for emissions factors are, geographic applicability, number of ingredients considered, among other factors. Thus, it is very interesting that it seems that these differences in assumptions between emissions factors seem to mostly cancel out.

Most importantly, the goal of quantifying the emissions from food purchases is to determine best courses of action to reduce the emissions from the category of purchases. The breakdown of which food categories contribute the most to the total emissions does not change significantly between the different methodologies and emissions factors. Thus, determining the categories that contribute the most is relatively easy, and is usually animal based foods. This is not surprising but highlights the exciting opportunity for emissions reductions from replacing traditional animal-based products with new foods such as plant-based meat substitutes and generally reducing the overall animal-based consumption across the board.

This work also highlights the importance of creating a standardized labeling system where emissions from each portion of the supply chain are easily kept track of and passed directly from each phase of the life cycle. Developing this system would ease the calculation of Scope 3 emissions and allow users to utilize their purchasing power to force companies and products to make more environmentally friendly decisions in terms of procurement and materials. In the meantime, it is also important to standardize the use of UNSPSC codes and require a detailed UNSPSC code down to the commodity level in order to help with categorizations for the rest of purchased goods. The categorization step is one of the main barriers for entities to begin to quantify their Scope 3 emissions from food purchases and will likely be the main barrier for quantifying Scope 3 emissions from other purchased goods categories as well. By obtaining specific information from the UNSPSC codes, the mapping of UNSPSC codes to EEIO categories can be made much more easily and generally for use by many in a standardized way.

5.2 Limitations of Study

There are many limitations of the study, some of which are inherent to the available data sets and others which arise from the methodologies employed. The data sets analyzed in this work are all in English and are from universities located in the United States (US) and Canada, which possibly restrict the confidence in results to applications of the tools and methodologies to entities in the US and Canada. In addition, some of the emissions factors data sets are US-specific, while others have varying levels of details for the entire world, thus limiting the applicability of these results of uncertainty and variance on emissions bounds to US data sets.

The food data sets that were included in the Stanford case study were not comprehensive of all food purchased on campus and only included the line items for which both weight and spend data could be determined. This introduces to possible limitations. First, because some campus food purchases were not considered including food purchased by on campus housing where students prepare their own food, food purchased for resale at on-campus food trucks, on-campus food franchises, and other food vendors including concessions at sporting events, the distributions of food purchasing analyzed may not be representative of the true distribution across the whole campus, or even the distributions of other entities. Second, most of the food items for which both spend and weight data could not be determined were fruits and vegetables, which introduces a bias where a larger fraction of the true purchases for animal products is considered than for the fruit and vegetable products. In a related fashion, food items that the script was not able to categorize are omitted from the analysis, which can also introduce a bias to the results. This bias in the discussed cases is likely minimal because of the high ability to assign categories to the Stanford data sets, but this could be a problem for other university data sets if a large portion of their data is not able to be categorized.

5.3 Future work

Many avenues exist to expand this work further, including improving the categorization script such that less food items are omitted due to inability to categorize, divulgation of the tool such that more entities begin quantifying, tracking, and mitigating Scope 3 Emissions from food purchasing, and using more emissions factors to better understand the uncertainty with regards to the total calculated emissions from purchased foods. In addition, it is important to understand how this work can inform similar work for other purchased goods.

5.3.1 Categorization Script

In order to minimize uncategorized items, after the fix_spelling function illustrated in Figure 5, an additional function could be added which could estimate the category based on the vendor for items that have not been categorized up to that point. For example, the three items with the largest weights that were uncategorized within the University E data set in Table 4 were from Coca Cola, thus it could be inferred that they were soda and should be categorized as Liquids.

The processing time could also be reduced by changing the structure of the code such that instead of the key words document contains single words, it contains multi-word key words so that most of the edge cases where two words should be assigned a new category could be removed. This would be particularly impactful to the run time because the edge cases function is the portion of the code that takes the longest to run. The processing time could also be significantly reduced by having the script create a pivot table of the provided data and consolidating all items with the same name, instead of assuming the user has provided a data set where this step has already been performed.

In addition, the script could be improved such that it facilitates correct interpretation of the results by the user. This could be achieved by, for example, including warnings of line items that might be outliers such that those can be excluded from the overall data set if appropriate. This would be possible by comparing $/kg of food in each category after categorization and comparing to the statistics from the metadata for all data analyzed by the script previously.

As more universities and entities use the categorization script, it will iteratively improve as more keywords and edge cases are added to accommodate more data points and as mistakes are noticed by users. As more entities use the script there will also be the opportunity to look at trends over time of food purchasing breakdowns. Finally, it is important to continue to advertise the tool and incentivize more entities to use the tool such that it can continue to be improved.

5.3.2 Uncertainty Quantification

The emissions factors used in this work are from governmental agencies or published papers, but there is a lack of emissions factors from LCA software, thus in order to better quantify the emissions calculation uncertainty, it would be interesting and important to include emissions factors from software such as OpenLCA, SimaPro, and Thinkstep by GaBi.

It would also be interesting to quantify the uncertainty that exists between the different ways to obtain a full year's worth of food purchasing at different entities. Some are only able to obtain data for one- or two-months' worth of purchasing and then scale this information up to a whole year, while other entities are able to obtain all the food purchases for the whole year directly. Because there is less diversity of purchasing that occurs in a month than in a whole year, it is likely that there would be a greater difference in emissions in this case for weight- versus spend- based emissions calculations than in the case where the whole years data is used.

In addition, a simplification was used for this analysis where the information of whether the food was organic or local was not taken into consideration when calculating the total data set emissions. Incorporating these data into an analysis may show a larger difference between spend- and weight- based emissions calculations because organic foods are usually more expensive, while local foods might be less expensive. It would also be interesting to consider other environmental impact factors and how these are affected by the use of these multiple emissions factor sources.

There is also an opportunity to include an integration with the US EPA Food Commodities Intake Database which has standardized recipes for foods in agricultural commodity form. It would be interesting to analyze what the impacts of true ingredient analysis of multi-ingredient items would have instead of the simplified maximum of three ingredient process that is being used in this work. In addition, instead of using inflation factors for the entire economy to harmonize cost data, the statistical abstract of the US provides historical price indexes for major commodity groups including food. This has the potential to influence the spend-based emissions factors if the cost of food grew significantly faster or slower in a particular year than the overall US economy.

Finally, as data collection continues from large institutions as they continue to use the tool, there will be exciting new opportunities for insights to be gained from the meta data, which could lead to a better understanding of the uncertainty within these calculated emissions.

5.3.3 Applying Learnings to Other Purchased Good Categories

Food is only one of many different categories of purchased goods, and it is important to understand how these learnings can be translated to other categories as well as what similar analyses can be conducted for other categories to determine the importance and uncertainty related to different methodologies and estimation methods.

For example, a very similar analysis can be conducted for laptops, computers, and some electronics like printers because the exact quantities of these items purchased is known as well as the spend on these items. Many computer companies have done more specific LCAs for individual laptops, computers, and some other electronics which take into consideration specific configurations. Thus, it is possible to determine the difference in total emissions estimated with these two methods to determine the uncertainty for this category. Much of this work has been done thus far in collaboration with the Material Science Laboratory at MIT using the Product Attribute to Impact Algorithm (PAIA) Research Approach [49].

It is also possible to do a similar analysis with chemicals because they are often purchased in very specific quantities which are kept track of and the spend on these chemicals is also known. Both chemical and the subset of electronics discussed here are thought to make up a large portion of a research university's emissions from Purchased Goods, in fact, at Stanford in 2019 according to data sourced from the Procurement Department, Laboratory Material Supplies is the largest category of spend for purchased goods and computers and computing devices is the 6th largest spend category.

General takeaways from this work which are applicable to other purchased goods categories include the learning that it would be helpful for entities to keep track of weights of other items purchased in order for comparisons to be made with the different methodologies. Today, it is possible to obtain weight data of purchased items through shipping information. It would also be beneficial to demand more accurate and specific UNSPSC codes for purchased goods, and for there to be centralized ways to pull purchasing information across an entity.

Bibliography

1. Doherty, C., Kiley, J. & Asheer, N. *As Economic Concerns Recede, Environmental Protection Rises on the Public's Policy Agenda*. vol. 13 www.pewresearch.org. (2020).

2. Hegedić, M., Štefanić, N. & Nikšić, M. Life cycle assessment: Assessing the environmental impact in the railway maintenance. in *MATEC Web of Conferences* vol. 180 (EDP Sciences, 2018).

3. Callahan, W. *et al. Corporate Value Chain (Scope 3) Accounting and Reporting Standard Supplement to the GHG Protocol Corporate Accounting and Reporting Standard GHG Protocol Team*.

4. Heller, M. C. & Keoleian, G. A. Greenhouse Gas Emission Estimates of U.S. Dietary Choices and Food Loss. *J. Ind. Ecol.* **19**, 391–401 (2015).

5. Poore, J. & Nemecek, T. Reducing food ' s environmental impacts through producers and consumers. **992**, 987–992 (2019).

6. Chesley, K. Faculty Senate urges acceleration of Stanford sustainability goals | Stanford News. https://news.stanford.edu/2020/11/05/faculty-senate-urges-action-sustainability/ (2020).

7. Home | SIMAP. https://unhsimap.org/home.

8. Scope 3 Inventory Guidance | US EPA. https://www.epa.gov/climateleadership/scope-3-inventory-guidance.

9. Smith, B. Microsoft will be carbon negative by 2030 - The Official Microsoft Blog. https://blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/ (2020).

10. Carbon Footprint. https://sustainability.aboutamazon.com/environment/sustainable-operations/carbon-footprint.

11. Weber, C. L. & Matthews, H. S. Food-miles and the relative climate impacts of food choices in the United States. *Environ. Sci. Technol.* **42**, 3508–3513 (2008).

12. Cool Food | World Resources Institute. https://www.wri.org/initiatives/cool-food-pledge.

13. Waite, R., Vennard, D. & Pozzi, G. TRACKING PROGRESS TOWARD THE COOL FOOD PLEDGE: SETTING CLIMATE TARGETS, TRACKING METRICS, USING THE COOL FOOD CALCULATOR, AND RELATED GUIDANCE FOR PLEDGE SIGNATORIES.

14. Searchinger, T. D., Wirsenius, S., Beringer, T. & Dumas, P. Assessing the efficiency of changes in land use for mitigating climate change. *Nat. 2018 5647735* **564**, 249–253 (2018).

15. Heller, M., Willits-Smith, A., Meyer, R., Keoleian, G. & Rose, D. LETTER • OPEN ACCESS Greenhouse gas emissions and energy use associated with production of individual self-selected US diets The influence of life cycle inventory approaches on the choice of structural systems to reduce the embodied greenhouse gas emissio. *Environ. Res. Lett.* (2018) doi:10.1088/1748-9326/aab0ac.

16. Clune, S., Crossin, E. & Verghese, K. Systematic review of greenhouse gas emissions for different fresh food categories. *J. Clean. Prod.* **140**, 766–783 (2017).

17. Le Quéré, C. *et al.* Global Carbon Budget 2016. *Earth Syst. Sci. Data* **8**, 605–649 (2016).

18. General Information about EIO-LCA - Economic Input-Output Life Cycle Assessment - Carnegie Mellon University. http://www.eiolca.net/Method/index.html.

19. Yang, Y., Ingwersen, W. W., Hawkins, T. R., Srocka, M. & Meyer, D. E. USEEIO: a New and Transparent United States Environmentally-Extended Input-Output Model. *J. Clean. Prod.* **158**, 308–318 (2017).

20. Yang, Y., Ingwersen, W. W., Hawkins, T. R., Srocka, M. & Meyer, D. E. USEEIO: A new and transparent United States environmentally-extended input-output model. *J. Clean. Prod.* **158**, 308–318 (2017).

21. Ingwersen, W. & Ceser, E. P. A. O. R. D. USEEIO National Models and Applications. (2021).

22. Supply Chain Greenhouse Gas Emission Factors for US Industries and Commodities. https://edg.epa.gov/metadata/catalog/search/resource/details.page?uuid=https://doi.org/10.23719/1517796.

23. Environmental Databases | CEDA — VitalMetrics — VitalMetrics. https://vitalmetricsnew.squarespace.com/environmental-databases.

24. Suh, S. Developing a sectoral environmental database for input-output analysis: The comprehensive environmental data archive of the US. *Econ. Syst. Res.* **17**, 449–469 (2005).

25. CEDA – Comprehensive Environmental Data Archive | Greenhouse Gas Protocol. https://ghgprotocol.org/Third-Party-Databases/CEDA.

26. *VitalMetrics CEDA Uses Brochure*.

27. Trustees, S. U. B. of. *Budget Plan*. *Financial* (2011).

28. About R&DE | Stanford R&DE. https://rde.stanford.edu/about-rde.

29. Dining Locations & Hours | Stanford R&DE. https://rde.stanford.edu/dining/dining-locations-hours.

30. Culinary Excellence | Stanford R&DE. https://rde.stanford.edu/dining/culinary-excellence.

31. Amazon Business in iProcurement. https://fingate.stanford.edu/system/amazon-business-iprocurement.

32. SmartMart Catalog Suppliers. https://fingate.stanford.edu/system/smartmart-catalog-suppliers.

33. UNSPSC > Home. https://www.unspsc.org/.

34. UNSPSC Codes Search - UNSPSC Codes | United States Datasets 💿. https://usa.databasesets.com/unspsc/search?keywords=44103112.

35. Emissions Inventory and Disclosure - The Sustainability Tracking, Assessment & Rating System. https://stars.aashe.org/resources-support/help-center/operations/air-emissions-and-disclosure/.

36. Stanford University | Scorecard | Institutions | STARS Reports. https://reports.aashe.org/institutions/stanford-university-ca/report/2019-02-22/.

37. Emissions Reports | SIMAP Preview. https://cmap-pre.sr.unh.edu/public/emissions.

38. pyspellchecker — pyspellchecker 0.6.3 documentation. https://pyspellchecker.readthedocs.io/en/latest/.

39. SIMAP Nitrogen Working Group | SIMAP. https://unhsimap.org/cmap/resources/NitrogenWorkingGroup.

40. Unit Conversions | SIMAP. https://unhsimap.org/cmap/conversions.

41. Ivy Plus Sustainability Consortium | Sustainability | Brown University. https://sustainability.brown.edu/ivy-plus-sustainability-consortium.

42. Poore, J. & Nemecek, T. Supplementary Materials for Reducing food's environmental impacts through producers and consumers. *Science (80-.).* **360**, 987 (2018).

43. LCA+Meta-Analysis+of+Food+Products+-+Full+Model+v0+(1).

44. conversion-factors-2019-full-set-v01-02.

45. $100 in 2011 → 2022 | Inflation Calculator. https://www.in2013dollars.com/us/inflation/2011?amount=100.

46. Ingwersen, W. W., Yang, Y., Gilkey, K. & Li, M. *USEEIO v1.1 – Description of Satellite Table Updates*. https://pasteur.epa.gov/uploads/10.23719/1365565/documents/USEEIOv1.1-DescriptionofSatelliteTableUpdates.pdf (2017).

47. Gan, Y. & Matthews, H. S. A Comparison of Methods and Results from the 2007 Benchmark USEEIO model and the 2002 EIO-LCA Model. (2018).

48. Supply Chain Greenhouse Gas Emission Factors for US Industries and Commodities | Science Inventory | US EPA. https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=349324&Lab=CESER.

49. Product Attribute to Impact Algorithm - PAIA | Materials Systems Laboratory. https://msl.mit.edu/projects/paia/main.html.

50. Environmental Impacts of Food Production - Our World in Data. https://ourworldindata.org/environmental-impacts-of-food.

Supplementary Information

Appendix I: Poore emissions breakdown [50]

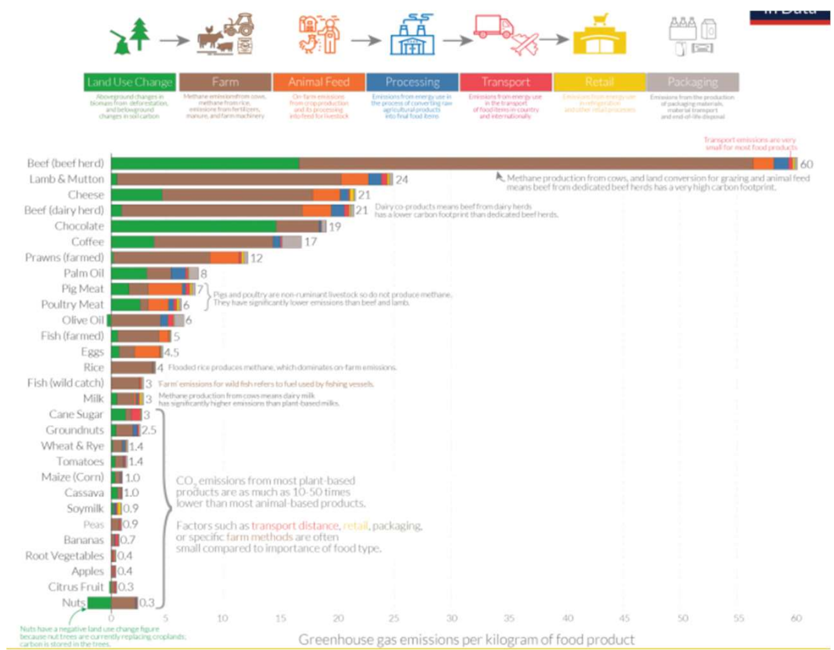


Figure 25. Breakdown of emissions per life cycle phase for Poore et al. food categories.

Appendix II: Dining Hall Weight Assignments

Table 13. Dining Hall Data Column Headers.

Column Header	Column Description
Item	Includes the item name and often some details or description about the item weight or quantity
Vendor	Useful for considerations of locally purchased items
Cost	Dollar amount for purchase per line item
Remark	Relevant purchase comments, sometimes includes details or descript about the item weight or quantity
Quantity (qty)	Quantity of purchase for each line item
Purchase Unit (PU)	The unit for the value in the column Quantity (qty) can be one of: BG, CS, DZ, EA, GL, JR, KG, LB, LF, LG, LT, OZ, PK, SH, SK, TR, TU
Conversion	The value that Quantity (qty) must be multiplied by to obtain the appropriate value to represent the weight of an item purchased within that line item.

$$New\ Quantity = qty * conversion$$

For this analysis, a *new quantity* value had already been calculated and the *conversion* had been set to 1 and the *quantity* in the provided spreadsheet was set to the *new quantity*. This affected approximately 10% of the meat purchases and thus, in the cases where the *conversion* was set to 1, the *new quantity* was set to the *quantity* and the unit was set to the *purchase unit.* For the rest of the data the following steps were conducted.

Any item name that included the following combinations of letters were analyzed: GL, KG, LB, and OZ. These key words were selected because they are indicative of a volume or weight associated with the item purchased. Within the item name, these key words could appear in two formats which are illustrated in Figure 26: a value and a unit as seen in the first line, or a number of packages purchased with a specific value and unit, as seen in the second line.

Bakery Mix WaffleCGM GF 15lb
Bakery Muffin Batter ChocChocChip Frozen 6x3lb

Figure 26. Example of LB found in item names.

The appropriate value was determined for each line that included each of the four key words listed above and put in the conversion column. After this step, all remaining lines for which the original purchase unit was GL, KG, LB, or OZ were considered to be accurate. Next, the remarks included with each of the lines not yet assigned a weight were analyzed to determine if they included weight information. If they did, the appropriate value was added to the conversion column, and the correct unit attributed to that line. Finally, for any remaining uncategorized lines, the item names were analyzed to determine if the exact same name from the same vendor had been attributed a weight or volume through any of the previously mentioned steps, and if so, that weight or volume was assigned. After each line item had a weight or volume associated with it, line items with the same item name, purchasing unit, and vendor were consolidated.

In some occasions items with the same name had different Purchasing units, and this was assumed to be attributable to human error when inputting the data at the food's arrival. In these cases, the most appropriate purchasing unit from the ones present as occurrences for one item name was determined by considering the number of items purchased that was attributed to each purchasing unit. Took original data, did a pivot table to determine which items had the same name but different purchasing units. Not converted are the lines that did not have a new quantity after the analysis done with oz, lt, gl, and lb in name. The provided data set had different purchase order, if the purchase order was not gl, kg, lb, or oz, then there was no way to determine what the purchasing quantity was.

Appendix III: SIMAP Unit Conversions

	Kilogram
1 US Gallon	3.8
1 Liter	1.0038536
1 Pound	0.453592

Appendix IV: Script and Hand Categorization Differences

University A: 188 Example of a mistaken categorization: *chicken sub breast meatless 3.8 oz vegan* in hand categorization went to Chicken and in script categorization went to beans. *Grape tomatoes* went to Fruits in hand categorization and vegetables in script categorization. Generally, not many patterns of differences other than the script assigns more categories to things than the hand categorization did, like *enchilada* for hand categorization was Cheese but for script categorization was cheese and grains.

University B: has 162 items which are categorized differently. Biggest differences come from things like cookie dough, cake, and brownies are being categorized as both grains and sugars in script whereas in hand categorization they were just grains. Same thing with creamer, where script categorizes it as milk, oils, and sugars but hand is only as milk

University D: 657 items which are categorized differently. Chex mix got hand categorized to vegetables and grains when for script it was categorized as grains and cheese when there was a key word for cheese. Clif bars were not categorized as Grains in the script but were in the hand categorization. Cookie dough was categorized as Grains for hand categorization and Grains + sugars for script categorization. Hi chew got categorized as Sugars for hand categorization and Fruits and Sugars for script categorization.

University E: 317 items which are categorized differently. Several items that have keywords which indicate one thing, but outside knowledge may indicate it is something else. Like things that say *2.5 gallon 1-ls fruit veg* which the hand categorization says is liquids and fruits, but the script categorization assigns it to Fruits and Vegetables since there is no explicit indication that this is a juice which would cause it to be assigned to Liquids. In addition, most brownies, cakes, and cookies in the hand categorization are just being assigned to grains but for the script categorization are assigned to Grains and Sugars. Peas are labeled as vegetables in the hand categorization and beans in the script categorization. Many sauces in hand categorization are labeled as both vegetables and oils but in script are just oils.

University G: 506 items which are categorized differently. Largest differences come from all juices and vinegars being categorized as Fruits in hand categorization but liquids in the script categorization. Sauces being categorized as vegetables or beans instead of oils.

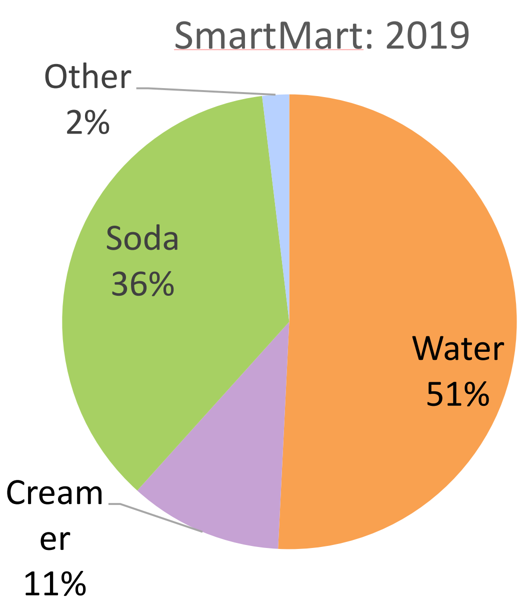
University H: doesn't have digitalized food purchasing information, so must take data from printed out sheets and manually enter them into a computer which leads to a large number of misspellings which the autocorrect function cannot always handle. 371 items which are categorized differently. Many hamburger related items that the hand categorization says is Grains but the script says Beef because the item names are things like: *bbb 4 pl hamburger 3* or *4" wheat hamburger dozen*. Indicates that the person categorizing by hand either had knowledge which is not present in the item name or that these were incorrect categorizations. Also has many items like *Italian parsley bag* which are hand categorized as spices but if they are fresh should be considered vegetables. *Sauce mnara* is supposed to be marinara sauce but is misspelled in such a way that the autocorrect doesn't pick it up so it gets script categorized as oils because of sauce but the hand categorization says vegetables.

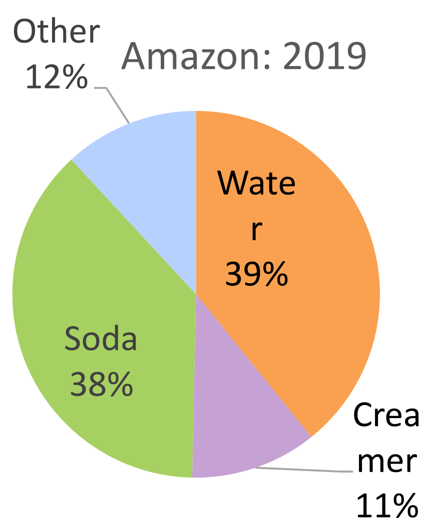
Importance of getting as detailed names as possible. Maybe with University H they eased up a little on the item names since they were doing it by hand, but if going to use the script important to include as much information as possible.

Appendix V: Liquids Purchasing Analysis

	SmartMart	Amazon
Water	896	1321
Creamer	192	375
Soda	642	1273
Other	34	400

A total of 598 items purchased through Amazon for Business had at least one of the categorizations as liquid.





Appendix VI: SIMAP to Heller 2018 Emissions Factors Mapping

Table 14. SIMAP to Heller 2018 Emissions Factors Mapping

SIMAP Food Categories	Heller 2018
Beef	Meat
Pork	Meat
Chicken	Meat
Cheese	Dairy
Eggs	Eggs
Milk	Dairy
Fish	Fish and Seafood
Liquids	Beverages
Grains	Cereals and Grains
Fruits	Fruit
Nuts	Legumes and Nuts
Oils	Oils and Fats
Beans	Meat Substitutes
Potatoes	Vegetables
Coffee and Tea	Other
Sugars	Sweeteners
Vegetables	Vegetables
Spices	Other

Appendix VII: SIMAP to Poore Emissions Factors Mapping

Table 15. SIMAP to Poore Emissions Category Mapping

SIMAP Food Categories	Poore Categories
Beef	Bovine Meat (beef herd), Bovine Meat (Dairy Herd), Mutton & Goat Meat
Pork	Pig Meat
Chicken	Poultry Meat
Cheese	Cheese
Eggs	Eggs
Milk	Milk, Milk
Fish	Fish (farmed), Crustaceans (farmed)
Liquids	Beer, Soybeans (Milk), Wine Grapes (Wine)
Grains	Bread, Wheat/Rye, Maize/Meal, Barley, Oats (Oatmeal), Rice
Fruits	Palm Fruit, Olives, Tomatoes, Citrus Fruits, Bananas, Apples, Berries, Pears (Pyruss ssp.), Stone Fruit (Prunus ssp.), Avocado, Pineapple, Kiwi, Melon and Watermelon
Nuts	Nuts, Groundnuts, Sunflower Seeds, Seeds
Oils	Soybeans (Oil), Palm (Oil), Sunflower (Oil), Rapeseed (Oil), Olive Oil
Beans	Beans & Pulses, Peas, Soybeans (Tofu) , Soybeans
Potatoes	Potatoes, Cassava and Other Roots
Coffee and Tea	Coffee, Chocolate, Cocoa
Sugars	Sugar, Sugar Cane, Sugar Beet Sugar
Vegetables	Sugar Beet Vegetable, Onions and Leeks, Root Vegetables, Cabbages and Other Brassicas, Gourds and Cucumber, Lettuce, Chicory, Endive and Artichoke, Green Beans and Peas
Spices	Sugar Beet Vegetable, Onions and Leeks, Root Vegetables, Cabbages and Other Brassicas, Gourds and Cucumber, Lettuce, Chicory, Endive and Artichoke, Green Beans and Peas

Appendix VIII: Order of EEIO Categorization

Table 16. Order of categorization for EEIO single ingredient analysis.

Frozen food manufacturing
Breakfast cereal manufacturing
Coffee and tea manufacturing
Cookie, cracker, pasta, and tortilla manufacturing
All other food manufacturing
Soft drink and ice manufacturing
Snack food manufacturing
Bread and bakery product manufacturing
Sugar and confectionery product manufacturing
Fruit and vegetable canning, pickling, and drying
Soybean and other oilseed processing
Poultry and egg production
Cheese manufacturing
Fluid milk and butter manufacturing
Seasoning and dressing manufacturing
Oilseed farming
Vegetable and melon farming
Fruit and tree nut farming
Other crop farming
Beef cattle ranching and farming, including feedlots and dual-purpose ranching and farming
Ice cream and frozen dessert manufacturing
Dairy cattle and milk production
Animal production, except cattle and poultry and eggs
Grain farming
Fishing, hunting and trapping
Flour milling and malt manufacturing
Wet corn milling
Fats and oils refining and blending
Dry, condensed, and evaporated dairy product manufacturing
Animal (except poultry) slaughtering, rendering, and processing
Poultry processing
Seafood product preparation and packaging
Flavoring syrup and concentrate manufacturing
Breweries
Wineries
Distilleries
Greenhouse, nursery, and floriculture production

Appendix IX: Carbon Equivalence Conversions

Gas	CO_{2eq} – IPCC AR4
Methane	25
Nitrous Oxide	298

Appendix X: Items Determined to be Outliers from R&DE Data Set

Beef

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
654	2019-01-01	2019-01-01	beef patty 2 1 raw fresh	43.544832	kilogram	No	No	Beef	NaN	NaN	NaN	2808.0	1.0	64.485264

Pork:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
1700	2019-01-01	2019-01-01	pork bacon aws cooked 2x150ea	166.971751	kilogram	No	No	Pork	NaN	NaN	NaN	4956.17	1.0	29.682686
1702	2019-01-01	2019-01-01	pork bacon coppa sliced cwh 12x3oz	6.123492	kilogram	No	No	Pork	NaN	NaN	NaN	300.66	1.0	49.099435
1708	2019-01-01	2019-01-01	pork bacon pancetta sliced cwh 12x2oz	2.721552	kilogram	No	No	Pork	NaN	NaN	NaN	240.40	1.0	88.331952
1709	2019-01-01	2019-01-01	pork bacon pancetta smkd lb	14.043208	kilogram	No	No	Pork	NaN	NaN	NaN	436.43	1.0	31.077656
1721	2019-01-01	2019-01-01	pork chop cc nimanranch lb	10.251179	kilogram	No	No	Pork	NaN	NaN	NaN	247.47	1.0	24.140637
1729	2019-01-01	2019-01-01	pork ham prosciutto sliced cwh 12x1lb	27.215520	kilogram	No	No	Pork	NaN	NaN	NaN	721.40	1.0	26.506934
1738	2019-01-01	2019-01-01	pork pig roasted ea	4.535920	kilogram	No	No	Pork	NaN	NaN	NaN	3800.00	1.0	837.757280

Chicken:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
1093	2019-01-01	2019-01-01	duck breast bnls frzn 7 8oz 32ea	6.944494	kilogram	No	No	Chicken	NaN	NaN	NaN	164.12	1.0	23.633113
1094	2019-01-01	2019-01-01	duck breast whole smkd frz birite 6lb	2.540115	kilogram	No	No	Chicken	NaN	NaN	NaN	67.98	1.0	26.762566
1765	2019-01-01	2019-01-01	poultry turkey patty mediterranean whitemeat r...	0.151103	kilogram	No	No	Chicken	NaN	NaN	NaN	27.48	1.0	181.862902

Fish:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
2078	2019-01-01	2019-01-01	seafood caviar goldpearl salmon 2oz	0.113398	kilogram	No	No	Fish	NaN	NaN	NaN	24.30	1.0	214.289494
2080	2019-01-01	2019-01-01	seafood caviartobiko black 1.1lb	4.575383	kilogram	No	No	Fish	NaN	NaN	NaN	233.19	1.0	50.966231
2117	2019-01-01	2019-01-01	seafood scallop raw60-80ct iqf dm lb	27.215520	kilogram	No	No	Fish	NaN	NaN	NaN	1407.00	1.0	51.698443

Milk:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
49	2019-01-01	2019-01-01	dairy gelato caramelseasalt villadolce cw 0.8l	3.212332	kilogram	No	No	Milk	NaN	NaN	NaN	130.20	1.0	40.531309
1049	2019-01-01	2019-01-01	dairy imitationcheese classicblend vegan 12x8oz	46.266384	kilogram	No	No	Milk	NaN	NaN	NaN	732.09	1.0	15.823368
1050	2019-01-01	2019-01-01	dairy imitationcheese mozzarella vegan 12x8oz	16.329312	kilogram	No	No	Milk	NaN	NaN	NaN	258.38	1.0	15.823079
1079	2019-01-01	2019-01-01	dairyfree butter vegancultured miyokos birite ...	1.360776	kilogram	No	No	Milk	NaN	NaN	NaN	23.96	1.0	17.607600
1432	2019-01-01	2019-01-01	icecream bar vanillamilkchocalmond haagendaz 1...	1.020582	kilogram	No	No	Milk	NaN	NaN	NaN	17.50	1.0	17.147079

Cheese:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
816	2019-01-01	2019-01-01	cheese brie triplecream petite birite 6x4oz	4.762716	kilogram	No	No	Cheese	NaN	NaN	NaN	137.22	1.0	28.811292
817	2019-01-01	2019-01-01	cheese camembert petite birite 6x4oz	4.762716	kilogram	No	No	Cheese	NaN	NaN	NaN	137.22	1.0	28.811292
821	2019-01-01	2019-01-01	cheese nicasiosquare taleggio org birite 4.5lb	1.859727	kilogram	Yes	No	Cheese	NaN	NaN	NaN	46.95	1.0	25.245638
822	2019-01-01	2019-01-01	cheese parmigiano meggiano wheel cwh lb	72.878627	kilogram	No	No	Cheese	NaN	NaN	NaN	1891.09	1.0	25.948486
958	2019-01-01	2019-01-01	dairy cheese burrata 12x4oz	9.525432	kilogram	No	No	Cheese	NaN	NaN	NaN	187.48	1.0	19.682047
960	2019-01-01	2019-01-01	dairy cheese burrata distefano 12x4oz	2.721552	kilogram	No	No	Cheese	NaN	NaN	NaN	54.70	1.0	20.098826
988	2019-01-01	2019-01-01	dairy cheese gournaygarlicherb boursin 12x5.2oz	8.845044	kilogram	No	No	Cheese	NaN	NaN	NaN	206.52	1.0	23.348668
989	2019-01-01	2019-01-01	dairy cheese gruyere loaf lbs	17.472363	kilogram	No	No	Cheese	NaN	NaN	NaN	408.57	1.0	23.383786
1011	2019-01-01	2019-01-01	dairy cheese pepperjack shrded vgn 8x7.8oz	1.769009	kilogram	No	No	Cheese	NaN	NaN	NaN	43.06	1.0	24.341315
1990	2019-01-01	2019-01-01	readymade chilesrelleno montereyjack cheddar 3...	40.823280	kilogram	No	No	Cheese	NaN	NaN	NaN	762.60	1.0	18.680518
2730	2019-01-01	2019-01-01	dairy cheese blue wheel birite lb	1.347634	kilogram	No	No	Cheese	NaN	NaN	NaN	27.05	1.0	20.072210

Eggs:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
296	2019-01-01	2019-01-01	appetizer omelette plain cooked frozen 3oz	1.70097	kilogram	No	No	Eggs	NaN	NaN	NaN	446.4	1.0	262.438491

Grains:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
327	2019-01-01	2019-01-01	baguette ancientgrain spi ea	1.240291	kilogram	No	No	Grains	NaN	NaN	NaN	26.50	1.0	21.365960
339	2019-01-01	2019-01-01	bakery bagel plain mini fresh 10ea	2.083688	kilogram	No	No	Grains	NaN	NaN	NaN	171.01	1.0	82.070818
343	2019-01-01	2019-01-01	bakery biscuit buttermilkplain spi small ea	0.839871	kilogram	No	No	Grains	NaN	NaN	NaN	51.00	1.0	60.723615
344	2019-01-01	2019-01-01	bakery biscuit cheddarchives medium spi ea	0.041994	kilogram	No	No	Grains	NaN	NaN	NaN	3.75	1.0	89.299434
434	2019-01-01	2019-01-01	bakery danish bearclaw medium	2.296310	kilogram	No	No	Grains	NaN	NaN	NaN	67.50	1.0	29.394992
435	2019-01-01	2019-01-01	bakery danish bearclaw mini spi ea	85.473743	kilogram	No	No	Grains	NaN	NaN	NaN	1778.40	1.0	20.806390
436	2019-01-01	2019-01-01	bakery danish bearclaw small	1.063106	kilogram	No	No	Grains	NaN	NaN	NaN	26.25	1.0	24.691794
439	2019-01-01	2019-01-01	bakery danish stick asst spi ea	0.850485	kilogram	No	No	Grains	NaN	NaN	NaN	22.00	1.0	25.867593
468	2019-01-01	2019-01-01	bakery macaron french assorted	1.417475	kilogram	No	No	Grains	NaN	NaN	NaN	75.00	1.0	52.910986
470	2019-01-01	2019-01-01	bakery minipannacotta greentea ea	0.850485	kilogram	No	No	Grains	NaN	NaN	NaN	23.85	1.0	28.042823
477	2019-01-01	2019-01-01	bakery mochi assorted ea	31.461141	kilogram	No	No	Grains	NaN	NaN	NaN	1560.00	1.0	49.584978
478	2019-01-01	2019-01-01	bakery muffin applecinnamon mini spi ea	22.821348	kilogram	No	No	Grains	NaN	NaN	NaN	531.58	1.0	23.293103
479	2019-01-01	2019-01-01	bakery muffin assortment mini ea	29.908722	kilogram	No	No	Grains	NaN	NaN	NaN	708.47	1.0	23.687739
480	2019-01-01	2019-01-01	bakery muffin banana allergy-friendlyvegan	33.339012	kilogram	No	No	Grains	NaN	NaN	NaN	699.72	1.0	20.988024
481	2019-01-01	2019-01-01	bakery muffin banana nonut mini ea	6.520385	kilogram	No	No	Grains	NaN	NaN	NaN	142.60	1.0	21.869874
482	2019-01-01	2019-01-01	bakery muffin bananawalnut mini spi ea	3.401940	kilogram	No	No	Grains	NaN	NaN	NaN	74.40	1.0	21.869874
487	2019-01-01	2019-01-01	bakery muffin blueberry mini spi ea	20.383291	kilogram	No	No	Grains	NaN	NaN	NaN	474.34	1.0	23.271022
488	2019-01-01	2019-01-01	bakery muffin carrotcake allergy-friendlyvegan	4.082328	kilogram	No	No	Grains	NaN	NaN	NaN	85.68	1.0	20.988024
492	2019-01-01	2019-01-01	bakery muffin cranberry mini spi ea	18.427175	kilogram	No	No	Grains	NaN	NaN	NaN	425.40	1.0	23.085470
496	2019-01-01	2019-01-01	bakery muffin greenteacranberry allergy-friend...	10.886208	kilogram	No	No	Grains	NaN	NaN	NaN	228.48	1.0	20.988024
498	2019-01-01	2019-01-01	bakery muffin nonut asst mini ea	235.074054	kilogram	No	No	Grains	NaN	NaN	NaN	5734.48	1.0	24.394355
499	2019-01-01	2019-01-01	bakery muffin oatbran mini spi ea	10.205820	kilogram	No	No	Grains	NaN	NaN	NaN	245.60	1.0	24.064700
500	2019-01-01	2019-01-01	bakery muffin pumpkin creamcheese mini ea	0.708738	kilogram	No	No	Grains	NaN	NaN	NaN	21.25	1.0	29.982892
501	2019-01-01	2019-01-01	bakery muffin pumpkincranberry allergy-friendl...	17.009700	kilogram	No	No	Grains	NaN	NaN	NaN	357.00	1.0	20.988024
503	2019-01-01	2019-01-01	bakery pannacotta passionfruit mini ea	0.850485	kilogram	No	No	Grains	NaN	NaN	NaN	23.85	1.0	28.042823
507	2019-01-01	2019-01-01	bakery petitfours assorted spi ea	1.700970	kilogram	No	No	Grains	NaN	NaN	NaN	58.00	1.0	34.098191
520	2019-01-01	2019-01-01	bakery pie lemonmeringue 1.5in sp ea	7.654365	kilogram	No	No	Grains	NaN	NaN	NaN	229.75	1.0	30.015553
559	2019-01-01	2019-01-01	bakery shell phyllo minicup baked frozen	0.013998	kilogram	No	No	Grains	NaN	NaN	NaN	23.04	1.0	1645.967163
569	2019-01-01	2019-01-01	bakery tart lemoncurd 2.5in	1.530873	kilogram	No	No	Grains	NaN	NaN	NaN	77.40	1.0	50.559387
582	2019-01-01	2019-01-01	batard ancient grain 24in spi ea	191.996989	kilogram	No	No	Grains	NaN	NaN	NaN	5059.50	1.0	26.351976
583	2019-01-01	2019-01-01	batard ancient grain crustybase spi ea	4.989512	kilogram	No	No	Grains	NaN	NaN	NaN	176.00	1.0	35.273991
681	2019-01-01	2019-01-01	bread injera ea	59.399914	kilogram	No	No	Grains	NaN	NaN	NaN	1575.00	1.0	26.515190
682	2019-01-01	2019-01-01	bread loaf cranberrywalnut spi ea	19.050864	kilogram	No	No	Grains	NaN	NaN	NaN	430.50	1.0	22.597400
692	2019-01-01	2019-01-01	breadflat 12x12in frz	0.340194	kilogram	No	No	Grains	NaN	NaN	NaN	38.46	1.0	113.053140
775	2019-01-01	2019-01-01	cereal assorted wellness cup 60x1.45oz	22.197658	kilogram	No	No	Grains	NaN	NaN	NaN	493.87	1.0	22.248743
779	2019-01-01	2019-01-01	cereal cheerios wholegrain ss gf 70x.62oz	6.151841	kilogram	No	No	Grains	NaN	NaN	NaN	152.55	1.0	24.797453
1119	2019-01-01	2019-01-01	focaccia spi hs	13.607760	kilogram	No	No	Grains	NaN	NaN	NaN	296.00	1.0	21.752294
1274	2019-01-01	2019-01-01	grocery cereal assorted kashi ss 36ea	26.943365	kilogram	No	No	Grains	NaN	NaN	NaN	608.00	1.0	22.565853
1322	2019-01-01	2019-01-01	grocery matzoball gf frz 12x10oz	6.803880	kilogram	No	No	Grains	NaN	NaN	NaN	159.00	1.0	23.369019
1453	2019-01-01	2019-01-01	loaf bananawalnut spi ea	16.329312	kilogram	No	No	Grains	NaN	NaN	NaN	385.00	1.0	23.577233
1455	2019-01-01	2019-01-01	loaf breakfast applecinnamonraisin 12c spi ea	78.925008	kilogram	No	No	Grains	NaN	NaN	NaN	1784.00	1.0	22.603735
1458	2019-01-01	2019-01-01	loaf breakfast lemonpoppyseed 12c spi ea	8.164656	kilogram	No	No	Grains	NaN	NaN	NaN	175.50	1.0	21.495088
1459	2019-01-01	2019-01-01	loaf breakfast lemonpoppyseednonut 12c spi ea	78.471416	kilogram	No	No	Grains	NaN	NaN	NaN	1704.25	1.0	21.718099
1530	2019-01-01	2019-01-01	muffin carrotpineapple nonut mini spi ea	7.937860	kilogram	No	No	Grains	NaN	NaN	NaN	210.00	1.0	26.455493
1533	2019-01-01	2019-01-01	muffin lemonpoppyseed nonut mini spi ea	10.772810	kilogram	No	No	Grains	NaN	NaN	NaN	235.60	1.0	21.869874
1534	2019-01-01	2019-01-01	naan garlic calavash frz atoria 4x5.5oz	40.539785	kilogram	No	No	Grains	NaN	NaN	NaN	1798.47	1.0	44.363087
1537	2019-01-01	2019-01-01	naan tandoori garlic 12x5.5oz	89.811216	kilogram	No	No	Grains	NaN	NaN	NaN	2052.53	1.0	22.853827
2002	2019-01-01	2019-01-01	readymade pancake scallion cooked 20x5ea	142.427888	kilogram	No	No	Grains	NaN	NaN	NaN	4980.00	1.0	34.965062
2162	2019-01-01	2019-01-01	snack popcorn seasalt vgn 24x0.6oz	0.816466	kilogram	No	No	Grains	NaN	NaN	NaN	37.70	1.0	46.174634
2170	2019-01-01	2019-01-01	snackbar fruitleather stretchisland unfi 30x0.5oz	3.827183	kilogram	No	No	Grains	NaN	NaN	NaN	109.44	1.0	28.595448
2178	2019-01-01	2019-01-01	snackbar sunbuttercrunch enjoylife 6x5oz	2.551455	kilogram	No	No	Grains	NaN	NaN	NaN	56.26	1.0	22.050164

Vegetables:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
252	2019-01-01	2019-01-01	condiment black pepper ss 6000x0.1gm	1.200000	kilogram	No	No	Vegetables	NaN	NaN	NaN	57.12	1.0	47.600000
253	2019-01-01	2019-01-01	condiment crushed red pepper ss 500x1gm	1.000000	kilogram	No	No	Vegetables	NaN	NaN	NaN	32.32	1.0	32.320000
908	2019-01-01	2019-01-01	condiment pickle ginger pink oz	24.493968	kilogram	No	No	Vegetables	NaN	NaN	NaN	486.00	1.0	19.841620
909	2019-01-01	2019-01-01	condiment pickle radish yellow 14x3.5oz	2.778251	kilogram	No	No	Vegetables	NaN	NaN	NaN	168.00	1.0	60.469698
1095	2019-01-01	2019-01-01	edible flower chefblend micro 4oz	0.113398	kilogram	No	No	Vegetables	NaN	NaN	NaN	12.26	1.0	108.114782
1096	2019-01-01	2019-01-01	edible flower confettiblend 4oz	0.113398	kilogram	No	No	Vegetables	NaN	NaN	NaN	14.15	1.0	124.781742
1097	2019-01-01	2019-01-01	edible leaf shiso red perilla bn	1.020582	kilogram	No	No	Vegetables	NaN	NaN	NaN	77.20	1.0	75.643113
1272	2019-01-01	2019-01-01	grocery caper kosher 32oz	0.907184	kilogram	No	No	Vegetables	NaN	NaN	NaN	24.80	1.0	27.337343
1327	2019-01-01	2019-01-01	grocery mushroom chanterelle dried 1lb	0.907184	kilogram	No	No	Vegetables	NaN	NaN	NaN	94.44	1.0	104.102365
1329	2019-01-01	2019-01-01	grocery mushroom morrel dried 1lb	0.453592	kilogram	No	No	Vegetables	NaN	NaN	NaN	164.22	1.0	362.043422
1331	2019-01-01	2019-01-01	grocery mushroom porcini dried lb	0.907184	kilogram	No	No	Vegetables	NaN	NaN	NaN	78.44	1.0	86.465370
1332	2019-01-01	2019-01-01	grocery mushroom shiitake dried 5lb	99.790240	kilogram	No	No	Vegetables	NaN	NaN	NaN	1772.80	1.0	17.765264
1337	2019-01-01	2019-01-01	grocery onion shallot fried 12x8oz	100.697424	kilogram	No	No	Vegetables	NaN	NaN	NaN	2022.00	1.0	20.079958
1381	2019-01-01	2019-01-01	herb arugula micro 4oz	13.154168	kilogram	No	No	Vegetables	NaN	NaN	NaN	1473.69	1.0	112.032171
1388	2019-01-01	2019-01-01	herb celery micro 4oz	2.267960	kilogram	No	No	Vegetables	NaN	NaN	NaN	234.92	1.0	103.582074
1405	2019-01-01	2019-01-01	herb micro greens intensity 4oz	6.236890	kilogram	No	No	Vegetables	NaN	NaN	NaN	663.54	1.0	106.389563
1406	2019-01-01	2019-01-01	herb micro greens intensity org 4oz	0.453592	kilogram	Yes	No	Vegetables	NaN	NaN	NaN	49.30	1.0	108.687984
1542	2019-01-01	2019-01-01	nonedible leaves ti red	2.267960	kilogram	No	No	Vegetables	NaN	NaN	NaN	32.50	1.0	14.330059
1789	2019-01-01	2019-01-01	precut brussel sprout cleaned halved 5lb	27.215520	kilogram	No	No	Vegetables	NaN	NaN	NaN	450.24	1.0	16.543502
1854	2019-01-01	2019-01-01	precut leeks diced .5in 5lb	20.411640	kilogram	No	No	Vegetables	NaN	NaN	NaN	337.45	1.0	16.532234
1866	2019-01-01	2019-01-01	precut mushroom shiitake sliced .25in 3lb	877.700520	kilogram	No	No	Vegetables	NaN	NaN	NaN	12078.71	1.0	13.761767
1986	2019-01-01	2019-01-01	precut zucchini sliced 0.25in 5lb	47.627160	kilogram	No	No	Vegetables	NaN	NaN	NaN	954.26	1.0	20.036047
1997	2019-01-01	2019-01-01	readymade grilled steak moncuisine veg 6x10oz	8.504850	kilogram	No	No	Vegetables	NaN	NaN	NaN	180.00	1.0	21.164394
1998	2019-01-01	2019-01-01	readymade moroccan chicken moncuisine veg 6x10oz	6.803880	kilogram	No	No	Vegetables	NaN	NaN	NaN	144.00	1.0	21.164394
2007	2019-01-01	2019-01-01	readymade stuffed cabbage moncuisinevegkosher ...	5.102910	kilogram	No	No	Vegetables	NaN	NaN	NaN	108.00	1.0	21.164394
2300	2019-01-01	2019-01-01	spice pepper red flake crushed oz	9.865626	kilogram	No	No	Vegetables	NaN	NaN	NaN	187.34	1.0	18.989165
2456	2019-01-01	2019-01-01	vegetable broccoli crown baby orgdi 5lb	2.267960	kilogram	No	No	Vegetables	NaN	NaN	NaN	31.80	1.0	14.021411
2485	2019-01-01	2019-01-01	vegetable chili jalapeno red lb	6.803880	kilogram	No	No	Vegetables	NaN	NaN	NaN	134.95	1.0	19.834271
2512	2019-01-01	2019-01-01	vegetable favabean fa oya 4lb	9.071840	kilogram	No	No	Vegetables	NaN	NaN	NaN	149.00	1.0	16.424452
2513	2019-01-01	2019-01-01	vegetable favabean leaves org di3 lb	29.937072	kilogram	Yes	No	Vegetables	NaN	NaN	NaN	611.60	1.0	20.429520
2514	2019-01-01	2019-01-01	vegetable favabean org di 3lb	2.721552	kilogram	Yes	No	Vegetables	NaN	NaN	NaN	47.80	1.0	17.563508
2533	2019-01-01	2019-01-01	vegetable lettuce baby lollarossa orgdi 2lb	1.814368	kilogram	No	No	Vegetables	NaN	NaN	NaN	27.60	1.0	15.211908
2549	2019-01-01	2019-01-01	vegetable mushroom chanterelle 5lb	7.257472	kilogram	No	No	Vegetables	NaN	NaN	NaN	384.24	1.0	52.944055
2550	2019-01-01	2019-01-01	vegetable mushroom cordyceps golden 5lb	9.071840	kilogram	No	No	Vegetables	NaN	NaN	NaN	330.36	1.0	36.415986
2552	2019-01-01	2019-01-01	vegetable mushroom enoki 12x3.5oz	59.533950	kilogram	No	No	Vegetables	NaN	NaN	NaN	1012.55	1.0	17.007943
2553	2019-01-01	2019-01-01	vegetable mushroom hedgehog 1lb	4.535920	kilogram	No	No	Vegetables	NaN	NaN	NaN	202.20	1.0	44.577506
2592	2019-01-01	2019-01-01	vegetable pepper chile thai red lb	4.535920	kilogram	No	No	Vegetables	NaN	NaN	NaN	70.00	1.0	15.432371
2645	2019-01-01	2019-01-01	vegetable radish ruby micro 4oz	1.587572	kilogram	No	No	Vegetables	NaN	NaN	NaN	150.42	1.0	94.748459

Fruits:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
1134	2019-01-01	2019-01-01	fruit apricot halvesdried birite 5lb	79.378600	kilogram	No	No	Fruits	NaN	NaN	NaN	1656.90	1.0	20.873384
1145	2019-01-01	2019-01-01	fruit cherry dried 5lb	92.986360	kilogram	No	No	Fruits	NaN	NaN	NaN	1939.80	1.0	20.861124
1149	2019-01-01	2019-01-01	fruit cranberry ea 12oz	0.595340	kilogram	No	No	Fruits	NaN	NaN	NaN	42.00	1.0	70.547981
1173	2019-01-01	2019-01-01	fruit mango dried 5lb	4.535920	kilogram	No	No	Fruits	NaN	NaN	NaN	96.00	1.0	21.164394
1216	2019-01-01	2019-01-01	fruit strawberry dried 5lb	4.535920	kilogram	No	No	Fruits	NaN	NaN	NaN	208.00	1.0	45.856188
1228	2019-01-01	2019-01-01	grab go thatsit bars apple cherry 12x1.2oz	0.816466	kilogram	No	No	Fruits	NaN	NaN	NaN	28.22	1.0	34.563612
1229	2019-01-01	2019-01-01	grab go thatsit bars apple pear dry 12x1.2oz	1.224698	kilogram	No	No	Fruits	NaN	NaN	NaN	42.33	1.0	34.563612
1279	2019-01-01	2019-01-01	grocery coconut flake lb	6.803880	kilogram	No	No	Fruits	NaN	NaN	NaN	225.00	1.0	33.069366
1396	2019-01-01	2019-01-01	herb citrus micro 4oz	6.577084	kilogram	No	No	Fruits	NaN	NaN	NaN	745.82	1.0	113.396758
2182	2019-01-01	2019-01-01	snakbar fruit applemango thatsit12x1.2oz	3.674095	kilogram	No	No	Fruits	NaN	NaN	NaN	134.07	1.0	36.490617
2183	2019-01-01	2019-01-01	snakbar fruit applestrawberry thatsit12x1.2oz	3.265862	kilogram	No	No	Fruits	NaN	NaN	NaN	119.17	1.0	36.489596

Potatoes:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
1400	2019-01-01	2019-01-01	herb galangal root whole lb	0.907184	kilogram	No	No	Potatoes	NaN	NaN	NaN	27.0	1.0	29.762430
1980	2019-01-01	2019-01-01	precut yams halves skinon 5lb	36.287360	kilogram	No	No	Potatoes	NaN	NaN	NaN	1178.0	1.0	32.463095

Beans:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
1496	2019-01-01	2019-01-01	meatless bulk impossibleburger dm 4x5lb	99.790240	kilogram	No	No	Beans	NaN	NaN	NaN	2344.30	1.0	23.492277
1497	2019-01-01	2019-01-01	meatless burger patty2.0 impossiblefood cwh 4x5lb	18.143680	kilogram	No	No	Beans	NaN	NaN	NaN	540.66	1.0	29.798806
1503	2019-01-01	2019-01-01	meatless patty soyprotein impossibleburger 40x4oz	36.287360	kilogram	No	No	Beans	NaN	NaN	NaN	922.22	1.0	25.414359
2214	2019-01-01	2019-01-01	spice bean vanilla 8oz	0.907184	kilogram	No	No	Beans	NaN	NaN	NaN	637.16	1.0	702.349248
2222	2019-01-01	2019-01-01	spice chili cascabel dreid 5lb	4.535920	kilogram	No	No	Beans	NaN	NaN	NaN	90.50	1.0	19.951851

Nuts:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
1595	2019-01-01	2019-01-01	nut pinenut raw 3x2lb	2.721552	kilogram	No	No	Nuts	NaN	NaN	NaN	150.72	1.0	55.380165

Liquids:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
120	2019-01-01	2019-01-01	grocery vinegar wine red passover 12x12.7oz	4.524375	kilogram	No	No	Liquids	NaN	NaN	NaN	37.50	1.0	8.288438
126	2019-01-01	2019-01-01	juicebase cranberrydrink 15 4.7 1 frz mm 4x90oz	138.937500	kilogram	No	No	Liquids	NaN	NaN	NaN	1937.26	1.0	13.943392
223	2019-01-01	2019-01-01	juice bluemachine naked 15.2oz	9.927500	kilogram	No	No	Liquids	NaN	NaN	NaN	366.08	1.0	36.875346
1510	2019-01-01	2019-01-01	milksub coconutblury sodelicious 12x5.3oz	3.606056	kilogram	No	No	Liquids	NaN	NaN	NaN	39.52	1.0	10.959341
2202	2019-01-01	2019-01-01	soupmix consomme osem kosher 12x14.1oz	4.796735	kilogram	No	No	Liquids	NaN	NaN	NaN	89.50	1.0	18.658523

Coffee and tea:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
884	2019-01-01	2019-01-01	coffee instant nescafe 14oz	1.587572	kilogram	No	No	Coffee and tea	NaN	NaN	NaN	334.24	1.0	210.535333
1439	2019-01-01	2019-01-01	icedtea herbal caffeinated 3glyield teavana 48ea	8.618248	kilogram	No	No	Coffee and tea	NaN	NaN	NaN	1821.15	1.0	211.313251
1440	2019-01-01	2019-01-01	icedtea herbal caffeinated 3glyield teavana 54ea	6.803880	kilogram	No	No	Coffee and tea	NaN	NaN	NaN	2094.60	1.0	307.853754
1539	2019-01-01	2019-01-01	nonedible leaf tea red bunch ea	0.566990	kilogram	No	No	Coffee and tea	NaN	NaN	NaN	107.80	1.0	190.126810

Oils:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
146	2019-01-01	2019-01-01	oil evoo kosher halutza 17oz	0.504687	kilogram	No	No	Oils	NaN	NaN	NaN	48.00	1.0	95.108359
167	2019-01-01	2019-01-01	sauce soy sweet 12x21floz	3.562500	kilogram	No	No	Oils	NaN	NaN	NaN	108.00	1.0	30.315789
1611	2019-01-01	2019-01-01	oil spray conditioner grid ea	3.175144	kilogram	No	No	Oils	NaN	NaN	NaN	89.00	1.0	28.030225
2333	2019-01-01	2019-01-01	spice sriracha blend tampico 1lb	1.360776	kilogram	No	No	Oils	NaN	NaN	NaN	35.37	1.0	25.992522

Sugars:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
194	2019-01-01	2019-01-01	syrup maple gradea 6x.5gl	22.80000	kilogram	No	No	Sugars	NaN	NaN	NaN	454.84	1.0	19.949123
1371	2019-01-01	2019-01-01	grocery sugar substitute steviaextr ss	3.99161	kilogram	No	No	Sugars	NaN	NaN	NaN	95.24	1.0	23.860049

Spices:

	Start date	End date	Label	Weight	Unit	Organic	Local	Category 1	Category 2	Category 3	Vendor	Dollars	Counts	$/kg
1104	2019-01-01	2019-01-01	flavoring extract vanilla org nonalch 4oz	0.907184	kilogram	Yes	No	Spices	NaN	NaN	NaN	581.44	1.0	640.928411
1383	2019-01-01	2019-01-01	herb basil micro 4oz	6.690482	kilogram	No	No	Spices	NaN	NaN	NaN	711.97	1.0	106.415352
1387	2019-01-01	2019-01-01	herb bullsblood micro 4oz	23.813580	kilogram	No	No	Spices	NaN	NaN	NaN	2505.71	1.0	105.221894
1389	2019-01-01	2019-01-01	herb chefblend green micro 4oz	14.061352	kilogram	No	No	Spices	NaN	NaN	NaN	1686.68	1.0	119.951481
1394	2019-01-01	2019-01-01	herb cilantro micro 4oz	14.968536	kilogram	No	No	Spices	NaN	NaN	NaN	1725.21	1.0	115.255760
1404	2019-01-01	2019-01-01	herb micro green heartsonfire 4oz	7.030676	kilogram	No	No	Spices	NaN	NaN	NaN	772.33	1.0	109.851457
1425	2019-01-01	2019-01-01	herb wasabi micro 4oz	1.360776	kilogram	No	No	Spices	NaN	NaN	NaN	169.89	1.0	124.847881
2217	2019-01-01	2019-01-01	spice cardamomseed ground 15oz	8.079607	kilogram	No	No	Spices	NaN	NaN	NaN	888.44	1.0	109.960787
2218	2019-01-01	2019-01-01	spice cardamomseed whole 8oz	0.226796	kilogram	No	No	Spices	NaN	NaN	NaN	41.38	1.0	182.454717
2248	2019-01-01	2019-01-01	spice dillweed dry whole 5.5oz	6.548735	kilogram	No	No	Spices	NaN	NaN	NaN	386.40	1.0	59.003766
2258	2019-01-01	2019-01-01	spice herbdeprovence	14.033003	kilogram	No	No	Spices	NaN	NaN	NaN	758.70	1.0	54.065408
2261	2019-01-01	2019-01-01	spice marjoramleaf dry whole 3.5oz	0.893009	kilogram	No	No	Spices	NaN	NaN	NaN	64.44	1.0	72.160507
2304	2019-01-01	2019-01-01	spice peppercorn pink whole 9oz	0.765436	kilogram	No	No	Spices	NaN	NaN	NaN	115.08	1.0	150.345587
2310	2019-01-01	2019-01-01	spice saffron 1oz	1.956115	kilogram	No	No	Spices	NaN	NaN	NaN	2455.34	1.0	1255.212179
2323	2019-01-01	2019-01-01	spice seasoning redfishblackenedtff 24oz	0.680388	kilogram	No	No	Spices	NaN	NaN	NaN	34.49	1.0	50.691664
2329	2019-01-01	2019-01-01	spice seasoning ymyaonoriko nori powder 10x3.5oz	0.992232	kilogram	No	No	Spices	NaN	NaN	NaN	72.00	1.0	72.563638
2335	2019-01-01	2019-01-01	spice tarragonleaf dry whole 4oz	0.453592	kilogram	No	No	Spices	NaN	NaN	NaN	38.80	1.0	85.539428

Appendix XI: Reframing food purchasing and emissions

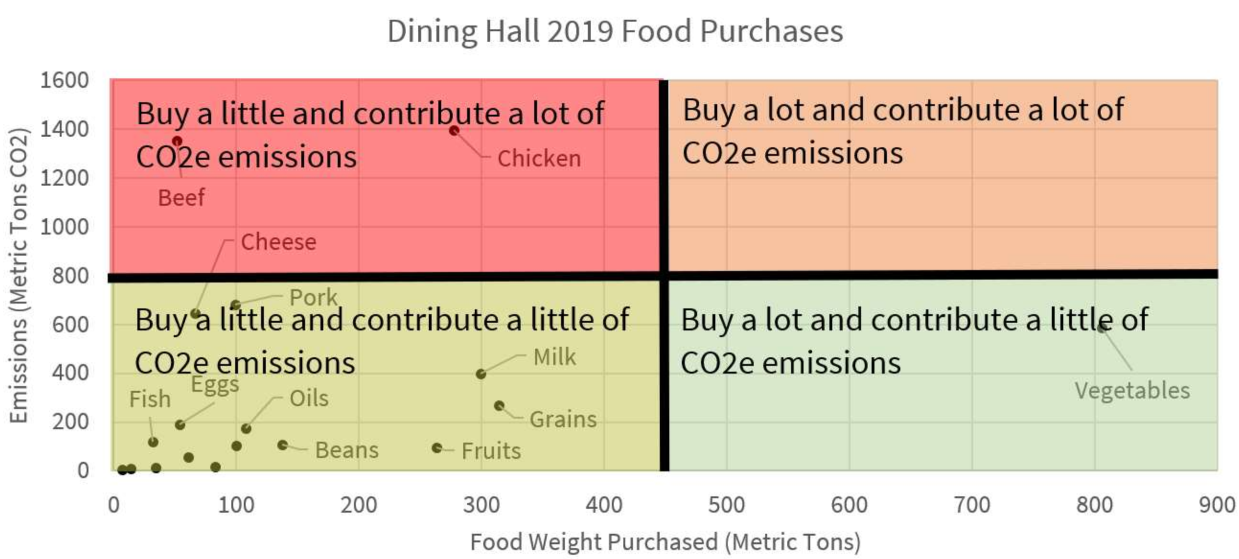


Figure 27. Thinking about food purchasing in quadrants.

Reframing of the way to think about the relationship between the food categories and their emissions. Want to move from the red colored quadrants to the green colored quadrants. This can be a good visual way to coordinate with food purchasing personnel to help them understand the impacts the food purchasing decisions they make have. The quadrants will self-adjust as the purchasing decisions change because the quadrants split the space for which there is data into four sections, instead of splitting a defined range into four sections.

Appendix XII: SIMAP to US EPA EEIO Category Mapping

Table 17. SIMAP to US EPA EEIO Category Mapping.

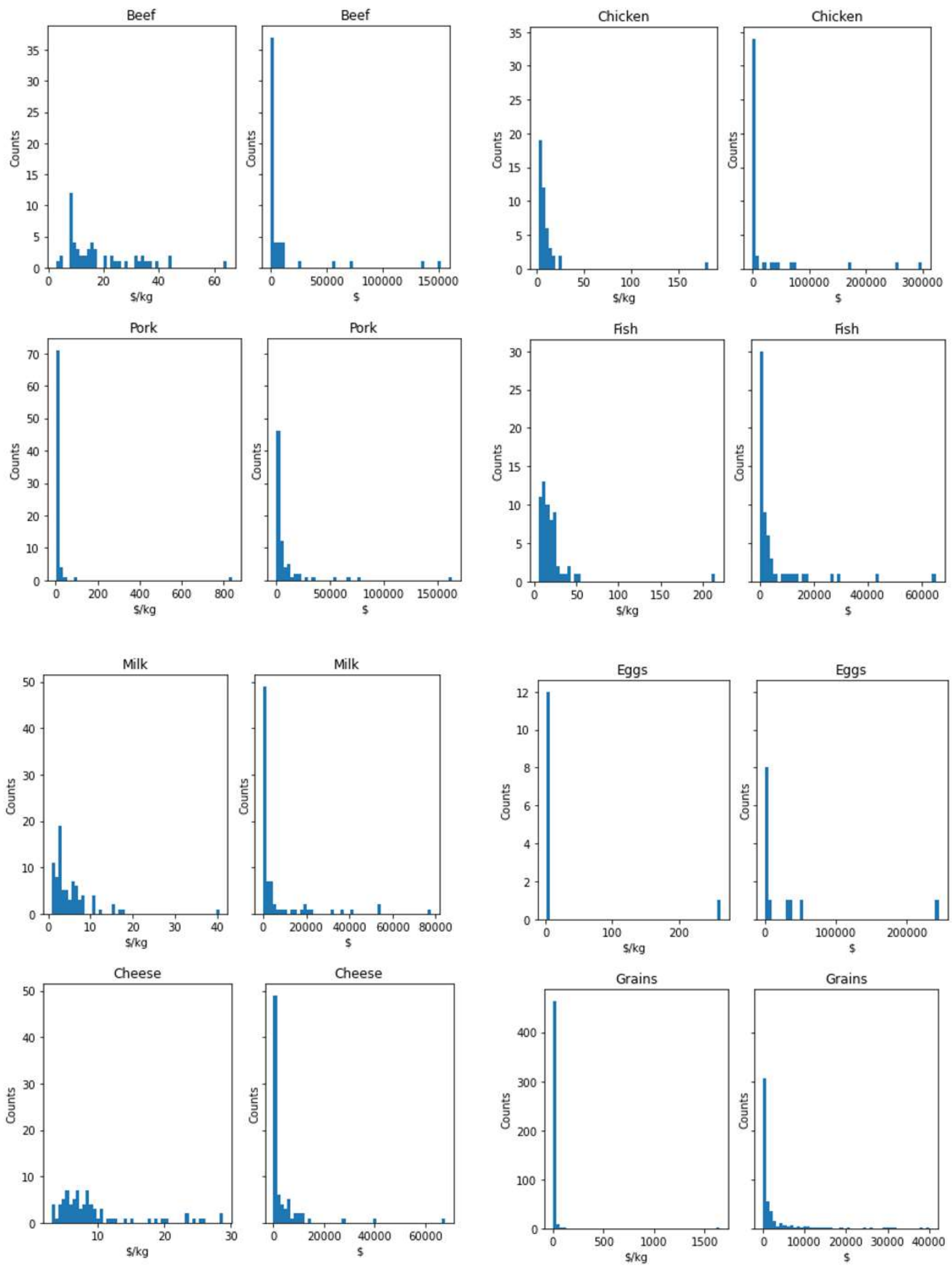
SIMAP Food Categories	EPA US EEIO Food Categories
Beef	Packaged Meat (except poultry)
Pork	Packaged Meat (except poultry)
Chicken	Packaged poultry
Cheese	Cheese
Eggs	Poultry farms
Milk	Fluid milk and butter
Fish	Seafood
Liquids	Soft drink, bottled water, and ice
Grains	Breakfast cereals, cookies, crackers, pastas, and tortillas, bread and other baked goods, flours and malts
Fruits	Fresh fruits and tree nuts
Nuts	Fresh fruits and tree nuts
Oils	Refined vegetable, olive, and seed oils, and seasonings and dressings
Beans	Tobacco, cotton, sugarcane, peanuts, sugar beets, herbs and spices, and other crops
Potatoes	Fresh vegetables, melons and potatoes
Coffee and Tea	Coffee and tea
Sugars	Sugar, candy, and chocolate
Vegetables	Fresh vegetables, melons, and potatoes
Spices	Tobacco, cotton, sugarcane, peanuts, sugar beets, herbs and spices

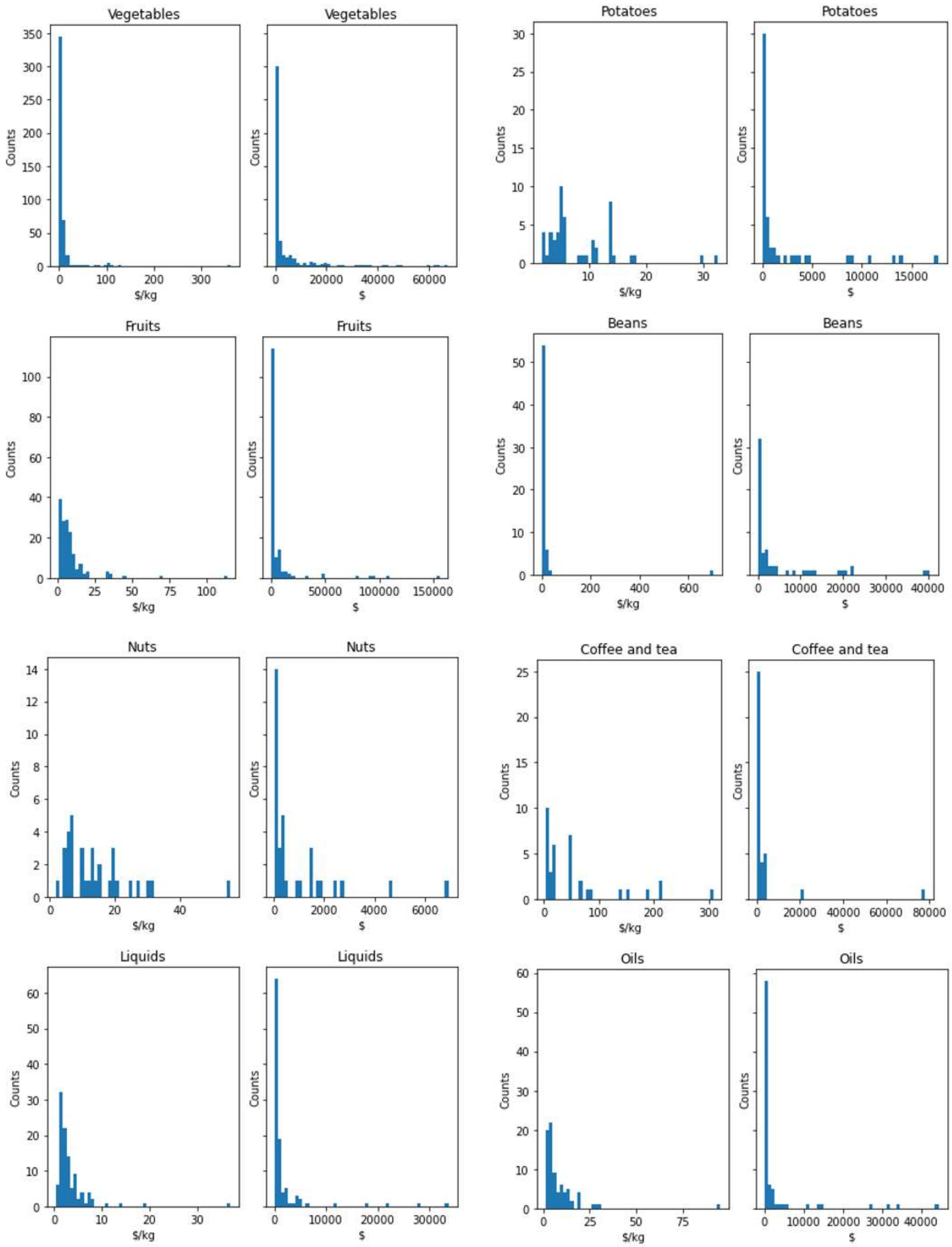
Appendix XIII: OBI to US EPA EEIO Category Mapping

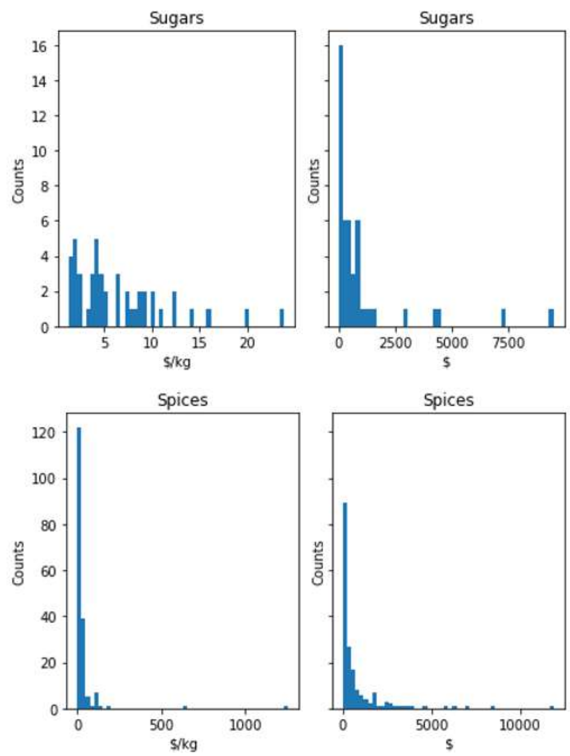
Table 18. Mapping of OBI Categories to US EPA EEIO Categories for preliminary determination of emissions form the top purchases in the OBI data set

OBI Category	US EPA EEIO Category
55210-SUPPLIES MATERIALS LABORATORY	N/A
53115-SU CAP SCIENTIFIC TECH EQUIP	334516/ANALYTICAL LABORATORY INSTRUMENTS/US
55110-SUPPLIES MTL NON-CAP EQ OFFICE	N/A
55120-SUPPLIES MTL NON-OFFIC NON-LAB	339940/OFFICE SUPPLIES (NOT PAPER)/US
55410-COMP SOFTWARE LICENSES	511200/SOFTWARE/US
53511-EQUIPMENT REPAIR AND MAINTENANCE	811200/ELECTRONIC EQUIPMENT REPAIR AND MAINTENANCE/US
55116-COMPUTERS & COMPUTING DEVICES	334111/COMPUTERS/US
53135-SU CAP MODULAR FURNITURE	33721A/OFFICE FURNITURE AND CUSTOM ARCHITECTURAL WOODWORK AND MILLWORK/US
52455-GROUND TRANSPORTATION	485000/PASSENGER GROUND TRANSPORT/US
52355-FOOD	311990/ALL OTHER FOODS/US
53120-SU CAP COMPUTER EQUIP	334111/COMPUTERS/US
55155-MISC FURNISHINGS NONCAPTL	337127/INSTITUTIONAL FURNITURE/US
55230-CHEMICALS COMPRESSED GAS	325120/COMPRESSED GASES/US
54510-PRINTING PUBLICATIONS	323120/PRINTING SUPPORT/US
55240-LAB ANIMAL PURCHASES	N/A
55215-LAB/SCI/TECH NON CAP EQUIPMENT	334516/ANALYTICAL LABORATORY INSTRUMENTS/US
55540-RDE PAPER SUPPLIES	322120/PAPER/US
55935-COGS PRODUCE	AVERAGE OF (111200/FRESH VEGETABLES, MELONS, AND POTATOES/US AND 111300/FRESH FRUITS AND TREE NUTS/US)
55925-COGS MEAT	31161A/PACKAGED MEAT (EXCEPT POULTRY)/US
55930-COGS GROCERIES	311990/ALL OTHER FOODS/US
55940-COGS DAIRY PRODUCTS	31151A/FLUID MILK AND BUTTER/US
55945-COGS BAKERY GOODS	311810/BREAD AND OTHER BAKED GOODS/US

Appendix XIV: R&DE Purchasing Histograms


Beef
Counts
$/kg
Beef
Counts
$
Chicken
Counts
$/kg
Chicken
Counts
$
Pork
Counts
$/kg
Pork
Counts
$
Fish
Counts
$/kg
Fish
Counts
$
Milk
Counts
$/kg
Milk
Counts
$
Eggs
Counts
$/kg
Eggs
Counts
$
Cheese
Counts
$/kg
Cheese
Counts
$
Grains
Counts
$/kg
Grains
Counts
$


Vegetables
Counts
$/kg
Vegetables
Counts
$
Potatoes
Counts
$/kg
Potatoes
Counts
$
Fruits
Counts
$/kg
Fruits
Counts
$
Beans
Counts
$/kg
Beans
Counts
$
Nuts
Counts
$/kg
Nuts
Counts
$
Coffee and tea
Counts
$/kg
Coffee and tea
Counts
$
Liquids
Counts
$/kg
Liquids
Counts
$
Oils
Counts
$/kg
Oils
Counts
$



Appendix XV: SIMAP Details

SIMAP Food Categories	Foods in Category
Beef	Beef Steak Hamburger Meatball Lasagna Goat Sheep Ruminant Jerky Lamb Deer Veal Pastrami All-beef Stk Burger Cheeseburger Rabbit
Pork	Sausage Bologna Bacon Pepperoni Pork Ham Salami CornDog Pig Dumpling Prosciutto Mortadella Capicola Bratwurst HotDog Porkloin
Chicken	Chicken Turkey Poultry Duck Gravy Chkn Duckling
Cheese	Cheddar Parmesan Cheese Feta Gouda Alfredo Pizza Lasagna Queso Mozzarella MacCheese Chss Brie Ricotta Ravioli Asiago Spanakopita Muenster Caprese
Eggs	Egg Eggnog Omelette Mousse Quiche
Milk	Milk Yogurt Cream Butter Condensed Pudding Dairy Shake Eggnog WhipLight Ranch Éclair Icecream Whipped IceCreamBar Éclair CremeBrulee Cannoli Cheesecake Crm Lactaid Buttermilk Tzatziki Mousse Crema Buttercream Chowder Creamer Yog Yoghurt Crème Creme Milkshake Parfait
Fish	Fish Lobster Shrimp Cod Anchovy Salmon Tuna Seafood Clam Mussel Oyster Catfish Crabmeat Crab Scallop Squid Tilapia Halibut Pollock Petrale Swai Seaweed Bonito Swordfish Haddock Littleneck Schrod Anchovies Trout Bluefish Pangasius Perch Crawfish Monkfish Chowder Rockfish Mahi Bass Seabass Redfish

Liquids	Juice Soda Broth Beverage Drink Soup Vinegar Wine Kombucha Water Cola Coca-Cola Dasani Coke Pepsi Gatorade Sprite Ale Dew Lemonade Smartwater Beer Vitaminwater Liquid MilkSub Monster RedBull MilkSubstitute SoupMix Sorbet JuiceBase Nondairy Sparkling SoyMilk Coke-Zero Izze Gravy Giardiniera Cider Bev Powerade Dip Gelatin Non-dairy OJ Drk Smoothie Stew Liq Puree Sobe
Grains	Wheat Bagel Bread Rice Biscuit Cereal Pita Cake Cracker Crepe Cookie Muffin Pastries Pancake Noodle Pasta Pizza Tortilla Chip Barley Rye Oats Millet Sorghum Grain Cornmeal Pie Chex Biscotti Pretzel Ramen Granola Oatmeal Kind Dough Oreo Waffle Churro Brioche Hoagie Croissant CornBread Loaf WholeWheat Scone Danish Bun Naan Roll Focaccia Couscous Pancake Flour BreadCrumb Crouton SandwichThin Tart Cupcake Baguette Lavash Taco CinnamonTwist ConeIceCream Baguette Brownie Cheesecake SourDough PizzaCrust Cobbler CornDog Baklava BreadFlat Phyllo Eclair Macaron Tostada Mochi RiceCracker Muffulletta BreadBowl BreadStick Flan CupCake MiniPannaCotta PannaCotta IceCreamBar Wonton PetitFour Granola Cheez-It Madeleines Alfajore Cheeto Matzo Breaded Samosa Batter Potsticker SpringRoll Donut Bakery Matzoball Tapioca SnackBar Pocky RiceKrispies Cannoli Dolma MacCheese Stuffing Toast ChexMix Pastry Yeast SnackMix Strudal Stirfry Cutlet Eggroll Empanada Blint Quinoa Ravioli Mac Turnover Ziti Spanakopita Quiche Rugelach Farro Blintz Strudel Flatbread Cannoli Macaroni Penne Pierogi Sandwich Breading Oat Wafer Polenta Ritz Nodl Hostess Spaghetti Cheerios Shell Belvita Nacho Tortellini Fettucini Graham Burrito Enchilada Chia Keebler Rice-a-roni Panko Dumpling Florentine Ciabatta Honeywheat Btrd Wrap Panini Multigrain Orzo Popcorn
Fruits	Apple Orange Lemon Grapefruit Citrus Banana Blueberries Strawberries Plantain Pineapple Date Grape Avocadoes Melon Fruit Preserve Lime Blueberry Cantaloupe Raspberry Apricot Strawberry Blackberries Watermelon Peach Kiwi Nectarine Pomegranate Cranberry Cranberries Mandarin Pluot Plum Persimmon Pluott Produce Mango Honeydew Pear Aprium Cherries Raisin Applesauce Coconut Fig Mangoes Cobbler Cherry Acai Prune Berry Fig Cranberries Guava Avocado Papaya Grape Guacamole Crnbry Cran Clementines Raspberries Olive Jackfruit Blackberry Starfruit Lychee Kumquat Peache Tangerine
Nuts	Cashew Almond Walnut Pistachio Peanut Tahini Nut Hazelnut Nutella Seed Torrone Pecan Baklava Chestnut Reese's Planters Pnut Peant Sunflower Pbj Reese Pb
Oils	Oil Canola Soybean Mayonnaise Margarine Lard HoneyMustard Ranch Balsamic Dressing Mayo Coffee-mate Creamer Mustard Sriracha Hollandaise Sauce Shortening Pesto Nutella
Beans	Soybean Tofu Bean Kidney Cannelli Pinto Chickpea Lentil Hummus Fava Pulse Tempeh Gardein Meatless SoyMilk Daiya Papad Falafel Miso Chili Pea Edamame Humus Soy Garbanzo
Potatoes	Potatoes Potato Fries Hash Yam Cassava Root Rutabaga Sunchoke Disiree Frito-Lay Pringles Frito Samosa Jicama Hashbrown Popchips Yucca

Coffee and tea	Coffee Tea Chocolate Chocolove Honest Matcha Chai Ferrero Ghirardelli Fudge Hershey Hershey's K-Cup Teabag Cocoa Truffle K-Cup Lavazza Nespresso Nescafe Ande's Andes Decaf Duplo Torrone Maxwell Café Starbucks Cafe Capsule Coffeemate IcedTea CoffeeBean ChocolateChip ChocChocChip Cacao Reese's Choc Tiramisu Rugelach Cappuccino Reese Chc M&M
Sugars	Sugar Sweetener Honey Candy Glaze Sprinkle Marshmallow Syrup Candies Gum Skittles Butterfinger Smarties Haribo Fudge Twix Snickers LifeSaver Licorice Tootsie Gummi Pop Ande's Andes Peppermint Cadbury Babies Mazapan Gummy Ike Wint-o-Green Sweet'n Lollipop Splenda SweetN Equal Stevia M&M Kat Starburst Werther's Caramel Oreo Cupcake Baklava Sorbet Flan CupCake SweetNLow Creamer Molasses Alfajore Airhead RiceKrispies Jelly Jam Strudal Cake Choc Sweetened Icing Custard Preserve Pudding Brownie Mousse Cheesecake Cookie Wrigley Whoppers Twizzlers Trolli Trident Sweetarts Reeses Orbit Lifesaver Hostess Altoids Airhead Chc Chocolate Halls Dentyne Nutella Ricola Hershey Mentos Cakes
Vegetables	Tomatoes Lettuce Cauliflower Carrot Corn Vegetable Beet Broccoli Chard Kale Leek Cabage Cabbage Fennel Kalette Arugula Sprout Mushroom Brussel Spinach Asparagus Bellpepper Eggplant Mushroom Cucumber Celery Bok Choy Pumpkin Romaine Ketchup Parsnip Kohlrabi Shallot Daikon Zucchini Romain Endive Kolhrabi Stinging Nettle Romanesco Frisee Tatsoi Frissee Radicchio Mizuna Broccolini Coleslaw Flower Produce Salad Cucumber Tomato Pepper Collard Rapini Eggplant Escarole Veggie Popcorners Seaweed Caper Ketchup Lasagna Salsa Snapea Dumpling Artichoke Pickle Horseradish Beet Daiya Pepperoncini Dolma Giardiniera Chili Kimchi Harissa Veg Stirfry Guacamole Taboule Gourd Lotus Sauerkraut Butternut Tomatillo Relish Radish Radishes Turnip Jalapeno Spanakopita Squash Marinara Watercress Okra Shiso Leaves Achiote Orchid Uplandcress Calendula Borage Dandelion Upland Cress Lemongrass Rhubarb Hominy Pesto Caprese
Vegetables	Basil Chive Sesame Garlic Ginger Scallion Cilantro Dill Parsely Rosemary Sage Parsley Marjoram Oregano Thyme Herb Mint Onion Tarragon Coriander
Spices	Pepper Pimento Clove Mustard Spice Seasoning Salt Cinnamon Curry Nutmeg Extract Tumeric Vanilla Sesame Garlic Flower Shiso Leaves Achiote Orchid Scallion Basil Cilantro Dill Parsely Rosemary Sage Parsley Chives Marjoram Oregano Thyme Herb Uplandcress Calendula Borage Dandelion Upland Cress Onion Tarragon Cumin Coriander Anise Paprika

Appendix XVI: VitalMetrics Details

Vital Metrics Food Categories	Description of Category

Frozen food manufacturing	manufacturing frozen fruits; frozen vegetables; and frozen fruit juices, ades, drinks, cocktail mixes and concentrates; manufacturing frozen specialty foods (except seafood), such as frozen dinners, entrees, and side dishes; frozen pizza; frozen whipped topping; and frozen waffles, pancakes, and french toast.
Breakfast cereal manufacturing	manufacturing breakfast cereal foods.
Coffee and tea manufacturing	(1) roasting coffee; (2) manufacturing coffee and tea concentrates (including instant and freeze-dried); (3) blending tea; (4) manufacturing herbal tea; and (5) manufacturing coffee extracts, flavorings, and syrups.
Cookie, cracker, pasta, and tortilla manufacturing	manufacturing cookies, crackers, and other products, such as ice cream cones; (1) manufacturing dry pasta and/or (2) manufacturing prepared flour mixes or dough from flour ground elsewhere.
All other food manufacturing	manufacturing perishable prepared foods, such as salads, sandwiches, prepared meals, fresh pizza, fresh pasta, and peeled or cut vegetables; manufacturing food (except animal food; grain and oilseed milling; sugar and confectionery products; preserved fruits, vegetables, and specialties; dairy products; meat products; seafood products; bakeries and tortillas; snack foods; coffee and tea; flavoring syrups and concentrates; seasonings and dressings; and perishable prepared food). Included in this industry are establishments primarily engaged in mixing purchased dried and/or dehydrated ingredients including those mixing purchased dried and/or dehydrated ingredients for soup mixes and bouillon. Illustrative Examples: Baking powder manufacturing Cake frosting, prepared, manufacturing, Dessert puddings manufacturing, Sweetening syrups (except pure maple) manufacturing, Egg substitutes manufacturing, Gelatin dessert preparations manufacturing, Honey processing, Powdered drink mixes (except chocolate, coffee, tea, or milk based) manufacturing, Popcorn (except popped) manufacturing, Yeast manufacturing
Soft drink and ice manufacturing	manufacturing soft drinks and artificially carbonated waters; purifying and bottling water (including naturally carbonated); manufacturing ice.
Snack food manufacturing	(1) salting, roasting, drying, cooking, or canning nuts; (2) processing grains or seeds into snacks; and (3) manufacturing peanut butter; manufacturing snack foods (except roasted nuts and peanut butter). Illustrative Examples: Corn chips and related corn snacks manufacturing, Popped popcorn (except candy-covered) manufacturing, Pork rinds manufacturing, Potato chips manufacturing, Pretzels (except soft) manufacturing, Tortilla chips manufacturing.
Bread and bakery product manufacturing	retailing bread and other bakery products not for immediate consumption made on the premises from flour, not from prepared dough; manufacturing fresh and frozen bread and bread-type rolls and other fresh bakery (except cookies and crackers) products; manufacturing frozen bakery products (except bread), such as cakes, pies, and doughnuts.

Sugar and confectionery product manufacturing	manufacturing refined beet sugar from sugar beets; (1) processing sugarcane and/or (2) refining cane sugar from raw cane sugar; (1) processing sugarcane and/or (2) refining cane sugar from raw cane sugar; manufacturing nonchocolate confectioneries. retailing nonchocolate confectionery products not for immediate consumption made on the premises; shelling, roasting, and grinding cacao beans and making chocolate cacao products and chocolate confectioneries; manufacturing chocolate confectioneries from chocolate produced elsewhere.
Fruit and vegetable canning, pickling, and drying	manufacturing canned, pickled, and brined fruits and vegetables
Soybean and other oilseed processing	crushing oilseeds and tree nuts, such as soybeans, cottonseeds, linseeds, peanuts, and sunflower seeds.
Poultry and egg production	raising chickens for egg production. The eggs produced may be for use as table eggs or hatching eggs; raising broilers, fryers, roasters, and other meat type chickens; raising turkeys for meat or egg production; hatching poultry of any kind; raising poultry
Cheese manufacturing	(1) manufacturing cheese products (except cottage cheese) from raw milk and/or processed milk products and/or (2) manufacturing cheese substitutes from soybean and other nondairy substances.
Fluid milk and butter manufacturing	(1) manufacturing processed milk products, such as pasteurized milk or cream and sour cream and/or (2) manufacturing fluid milk dairy substitutes from soybeans and other nondairy substances; manufacturing creamery butter from milk and/or processed milk products.
Seasoning and dressing manufacturing	manufacturing mayonnaise, salad dressing, vinegar, mustard, horseradish, soy sauce, tarter sauce, Worcestershire sauce, and other prepared sauces (except tomato-based and gravy); (1) manufacturing spices, table salt, seasonings, flavoring extracts (except coffee and meat), and natural food colorings and/or (2) manufacturing dry mix food preparations, such as salad dressing mixes, gravy and sauce mixes, frosting mixes, and other dry mix preparations.
Oilseed farming	growing soybeans and/or producing soybean seeds; growing fibrous oilseed producing plants and/or producing oilseed seeds, such as sunflower, safflower, flax, rape, canola, and sesame; growing dry peas, beans, and/or lentils.
Vegetable and melon farming	growing potatoes and/or producing seed potatoes; in one or more of the following: (1) growing melons and/or vegetables (except potatoes; dry peas; dry beans; field, silage, or seed corn; and sugar beets); (2) producing vegetable and/or melon seeds; and (3) growing vegetable and/or melon bedding plants
Fruit and tree nut farming	growing apples; growing oranges; growing citrus fruits. growing tree nuts
Other crop farming	growing tobacco; growing cotton; growing sugarcane; growing hay, alfalfa, clover, and/or mixed hay; growing sugar beets; growing peanuts

Beef cattle ranching and farming, including feedlots and dual-purpose ranching and farming	raising cattle (including cattle for dairy herd replacements); feeding cattle for fattening.
Ice cream and frozen dessert manufacturing	ice cream, frozen yogurts, frozen ices, sherbets, frozen tofu, and other frozen desserts (except bakery products).
Dairy cattle and milk production	This industry comprises establishments primarily engaged in milking dairy cattle.
Animal production, except cattle and poultry and eggs	raising hogs and pigs. These establishments may include farming activities, such as breeding, farrowing, and the raising of weanling pigs, feeder pigs, or market size hogs; raising sheep and lambs, or feeding lambs for fattening. The sheep or lambs may be raised for sale or wool production; raising goats; (1) farm raising finfish (e.g., catfish, trout, goldfish, tropical fish, minnows) and/or (2) hatching fish of any kind; farm raising shellfish (e.g., crayfish, shrimp, oysters, clams, mollusks)
Grain farming	growing wheat and/or producing wheat seeds; growing and/or producing corn seeds; growing rice (except wild rice) and/or producing rice seeds; growing a combination of oilseed(s) and grain(s) with no one oilseed (or family of oilseeds) or grain (or family of grains) accounting for one-half of the establishment's agricultural production (value of crops for market). These establishments may produce oilseed(s) and grain(s) seeds and/or grow oilseed(s) and grain(s); growing grains and/or producing grain(s) seeds (except wheat, corn, rice, and oilseed(s) and grain(s) combinations).
Fishing, hunting and trapping	the commercial catching or taking of finfish (e.g., bluefish, salmon, trout, tuna) from their natural habitat; the commercial catching or taking of shellfish (e.g., clams, crabs, lobsters, mussels, oysters, sea urchins, shrimp) from their natural habitat; the commercial catching or taking of marine animals
Flour milling and malt manufacturing	(1) milling flour or meal from grains (except rice) or vegetables and/or (2) milling flour and preparing flour mixes or doughs.; one of the following: (1) milling rice; (2) cleaning and polishing rice; or (3) milling, cleaning, and polishing rice. The establishments in this industry may package the rice they mill with other ingredients; manufacturing malt from barley, rye, or other grains.
Wet corn milling	wet milling corn and other vegetables (except to make ethyl alcohol). Examples of products made in these establishments are corn sweeteners, such as glucose, dextrose, and fructose; corn oil; and starches (except laundry).
Fats and oils refining and blending	(1) manufacturing shortening and margarine from purchased fats and oils; (2) refining and/or blending vegetable, oilseed, and tree nut oils from purchased oils; and (3) blending purchased animal fats with purchased vegetable fats.
Dry, condensed, and evaporated dairy product manufacturing	manufacturing dry, condensed, and evaporated milk and dairy substitute products.

Animal (except poultry) slaughtering, rendering, and processing	slaughtering animals (except poultry and small game). Establishments that slaughter and prepare meats are included in this industry; processing or preserving meat and meat byproducts (except poultry and small game) from purchased meats. This industry includes establishments primarily engaged in assembly cutting and packing of meats (i.e., boxed meats) from purchased meats; rendering animal fat, bones, and meat scraps.
Poultry processing	(1) slaughtering poultry and small game and/or (2) preparing processed poultry and small game meat and meat byproducts.
Seafood product preparation and packaging	1) canning seafood (including soup); (2) smoking, salting, and drying seafood; (3) eviscerating fresh fish by removing heads, fins, scales, bones, and entrails; (4) shucking and packing fresh shellfish; (5) processing marine fats and oils; and (6) freezing seafood. Establishments known as "floating factory ships" that are engaged in the gathering and processing of seafood into canned seafood products are included in this industry; one or more of the following: (1) canning seafood (including soup); (2) smoking, salting, and drying seafood; (3) eviscerating fresh fish by removing heads, fins, scales, bones, and entrails; (4) shucking and packing fresh shellfish; (5) processing marine fats and oils; and (6) freezing seafood.
Flavoring syrup and concentrate manufacturing	manufacturing flavoring syrup drink concentrates and related products for soda fountain use or for the manufacture of soft drinks.
Breweries	brewing beer, ale, lager, malt liquors, and nonalcoholic beer.
Wineries	(1) growing grapes and manufacturing wines and brandies; (2) manufacturing wines and brandies from grapes and other fruits grown elsewhere; and (3) blending wines and brandies.
Distilleries	(1) distilling potable liquors (except brandies); (2) distilling and blending liquors; and (3) blending and mixing liquors and other ingredients.
Greenhouse, nursery, and floriculture production	growing mushrooms under cover in mines underground, or in other controlled environments; growing food crops (except mushrooms) under glass or protective cover. growing and/or producing floriculture products (e.g., cut flowers and roses, cut cultivated greens, potted flowering and foliage plants, and flower seeds) under cover and in open fields.

Vital Metrics Food Categories	**Foods in Category**
Frozen food manufacturing	frozen corndog dumplings pizza fries hash browns frz
Breakfast cereal manufacturing	cereal cereals

Coffee and tea manufacturing	coffee tea teas matcha chai decaf max K-cup teabags teabag k-cups lavazza nespresso nescafe capsule capsules Icedtea coffeebean flavia starbucks
Cookie, cracker, pasta, and tortilla manufacturing	cookie cookies cracker crackers pasta tortilla pita crepe noodles tortillas ramen biscotti ConeIceCream oreo alfajores ricekrispies couscous noodle ricecracker matzo cheez-it crouton pocky
All other food manufacturing	salad sandwich frosting syrup gelatin honey popcorn lasagna eggnog pudding coleslaw broth soupmix samosa guacamole torrone baklava daiya papad gardein falafel meatless dolma popcorners syrups haribo gum syrup gummi gummy wonton Maccheese flan hoagie potsticker springroll yeast cobbler taco muffulletta minipannacotta pannacotta stuffing veggie vegetarian veg
Soft drink and ice manufacturing	ice water soda beverage drinks beverages sprite dew smartwater cola coca-cola dasani pepsi gatorade vitaminwater monster redbull Sparkling coke-zero izze fanta coke croix juice
Snack food manufacturing	chip pretzel chips pretzels Frito-Lay pringles seed seeds frito ChexMix Chex Granola cheetos snackbar kind snackmix bar bars snacks kar's
Bread and bakery product manufacturing	bread cake pie doughnut éclair eclair cannoli bagel biscuit muffins pastries pancakes biscuits muffin naan roll bun pie baguette dough waffle brownie papad bakery baguette brioche croissant cornbread loaf wholewheat scone danish focaccia tostada sourdough pizzacrust cupcake donut cupcakes pastry toast pancake breadcrumb tart cupcakes madeleines macaron breadbowl churro breadstick brownie cinnamontwist breadflat phyllo breaded batter sandwichthin lavash matzoball
Sugar and confectionery product manufacturing	sugar chocolate chocolove ferrero ghirardelli fudge hershey hershey's cocoa truffles nutella hersheys chocolatechip chocchocchip cacao chocolates reese's ande's andes duplo torrone candy glaze sprinkles mints marshmallow candies skittles butterfinger smarties fudge twix snickers marshmallows lifesavers licorice tootsie tootsies pops ande's andes peppermint cadbury babies mazapan ike wint-o-green lollipops pop M&Ms Kat Starburst werther's caramel airheads candies molasses sprinkle Sweeteners Sweetener Sweetner Sweet'n Splenda SweetN Equal Stevia SweetNLow petitfours
Fruit and vegetable canning, pickling, and drying	pickled canned jam jelly ketchup salsa chili soup juicebase lemonade preserves applesauce pickle giardiniera pickle caper marinara
Soybean and other oilseed processing	oil oils tempeh tofu
Poultry and egg production	egg turkey duck quail pheasant chicken poultry eggs omelette
Cheese manufacturing	cheese cheddar parmesan feta gouda cheeses queso mozzarella maccheese

Fluid milk and butter manufacturing	cream butter yogurt half whiplight whipped milksub substitute soymilk nondairy creamer almondmilk oatmilk milksubstitute
Seasoning and dressing manufacturing	mayonnaise vinegar mustard horseradish tarter worcestershire salt extract gravy alfredo ranch sauce relish giardiniera pimento cloves mustard spices seasoning dressings honeymustard dressing italian mayo cilantro dill parsely rosemary sage parsley chives marjoram oregano thyme herbs herb uplandcress calendula borage dandelion upland cress salt cinnamon spice sesame vanilla garlic tumeric turmeric extract nutmeg ginger flower shiso leaves achiote curry tahini hummus miso kimchi harissa mizuna pepper heinz tartar
Oilseed farming	Soybean sunflower safflower flax canola sesame peas beans lentils lentil soybeans kidney cannelli pinto chickpeas fava pulses pea bean chickpea snapeas
Vegetable and melon farming	potato potatoes melon carrot squash tomato watermelon cantaloupe casaba honeydew pepper melons cantaloupes tomatoes onions lettuce carrots greens peas cauliflower corn vegetable beets broccoli chard kale leeks cabage cabbage vegetables honeydew jalapeno yams cassava radish radishes roots turnips rutabagas sunchokes rutabaga turnip yam jicama arugula brussel spinach asparagus bellpepper eggplant celery pumpkin romaine zucchini romain broccolini cucumber onion eggplants vegetable salads sunchokes salad produce peppers artichoke beet horseradish fennel kalette sprouts cucumbers bok choy parsnips kohlrabi shallots daikon endive kolhrabi stinging nettle romanesco frisee tatsoi frissee radicchio collards rapini escarole pepperoncini shallot parsnip
Fruit and tree nut farming	apple orange grapefruit lemon tangerine grape raisin strawberry berry cranberry blackberry currant blueberry raspberry almond pistachio macadamia walnut pecan apricot fig banana cherry peach pineapple date prune apples oranges lemons grapefruit citrus bananas blueberries strawberries plantains pineapples dates grapes avocadoes fruit fruits almonds pistachios limes apricots blackberries kiwi peaches nectarines pomegranates mandarins mandarin pluot plum pluots persimmons pluotts produce mango pears nectarine pomegranate apriums cherries raisins pineapple figs mangoes coconut acai prunes cranberries guava plantain avocado papaya grape cashews walnuts nuts almond hazelnuts nut cashew pecans chestnut chestnuts planters
Other crop farming	sugarcane alfalfa peanut peanuts mint
Beef cattle ranching and farming, including feedlots and dual-purpose ranching and farming	beef steak hamburger meatball meat ruminants
Ice cream and frozen dessert manufacturing	sherbet icecream icecreambar eclair cremebrulee shake sorbet mochi gelato IceCream

Dairy cattle and milk production	milk dairy
Animal production, except cattle and poultry and eggs	pig lamb sheep goat catfish trout goldfish crayfish shrimp oyster clam mollusk algae seaweed honey deer bacon pork ham seaweed
Grain farming	wheat rice sorghum oat rye barley grains oats grain millet oatmeal tapioca
Fishing, hunting and trapping	salmon trout tuna clam crab lobster oyster shrimp fish lobsters cod anchovy seafood clams mussels oysters crabmeat crab scallop squid tilapia halibut pollock petrale swai bonito
Flour milling and malt manufacturing	flour malt
Wet corn milling	glucose dextrose fructose cornmeal
Fats and oils refining and blending	margarine vegetable oil gravy canola olive soybean lard
Dry, condensed, and evaporated dairy product manufacturing	condensed
Animal (except poultry) slaughtering, rendering, and processing	jerky sausage bologna pepperoni salami
Poultry processing	
Seafood product preparation and packaging	
Flavoring syrup and concentrate manufacturing	
Breweries	beer ale lager malt
Wineries	wine brandy
Distilleries	liquor kombucha
Greenhouse, nursery, and floriculture production	Mushroom Mushrooms Flowers flower