

Importing data into R

Dhafer Malouche

<http://dhafermalouche.net>

Aim of this chapter

Show how can we import data into R from several sources:

- Local files (Rdata, csv, other Statistical Software)
- Web Scraping
- pdf documents
- Social Networks
 - *Twitter*
 - *Facebook*

Local files

R environnement, .RData file

```
> load(file="dir_location\\savedfile") # Windows only  
> load(file="dir_location/savedfile") # other OS
```

Reading text files

```
> # Windows only  
> ds = read.table("dir_location\\file.txt", header=TRUE)  
> # all OS (including Windows)  
> ds = read.table("dir_location/file.txt", header=TRUE)
```

Example:

```
> file.exists("banques.txt")
```

```
## [1] TRUE
```

```
> d1 = read.table("banques.txt", header=TRUE)
```

Example:

```
> head(d1)
```

```
##   Classe_PV Age_PV Cadre Non_Cadre Diplome Age_Moy Anc_Moy Surface_  
## 1         3    29   14        10      10  41,71   5,54   12,75  
## 2         3    21    6        14      6   34,2    4,7    11,8  
## 3         5    13    3         5      3  31,13   6,63   12,25  
## 4         3    18    5         7      4  36,58   5,33   16,42  
## 5         4    17    2         5      2  44,86   4,14   27,86  
## 6         4    16    3         3      2  40,33    5     40  
##   Type_concep Nbr_reclam Age_RPV Anc_RPV Anc_RPV_PV Qualt_client  
## 1           A         16     38     21         20         0,37  
## 2           A          7     41     28          9         0,53  
## 3           A          7     46     66         24         0,49  
## 4           A         12     38     28          9         0,69  
## 5           A          2     47      7          7         0,44  
## 6           A          8     37     66          7         0,4
```

Other ASCII files

```
> file.exists("banques.txt")
```

```
## [1] TRUE
```

```
> d1 = readLines("banques.txt")  
> d1[1:3]
```

```
## [1] "Classe_PV\tAge_PV\tCadre\tNon_Cadre\tDiplome\tAge_Moy\tAnc_Moy\tSurface_\tType_concep\tNbr_reclam\tAge_RPV\tAnc_RPV\tAnc_RPV_I  
## [2] "3\t29\t14\t10\t10\t41,71\t5,54\t12,75\tA\t16\t38\t21\t20\t0,37"  
## [3] "3\t21\t6\t14\t6\t34,2\t4,7\t11,8\tA\t7\t41\t28\t9\t0,53"
```


Comma-separated value (CSV)

```
> ds = read.csv("dir_location/file.csv")
```

Other formats

```
> library(foreign)
> ds = read.dbf("filename.dbf") # DBase
> ds = read.epiinfo("filename.epiinfo") # Epi Info
> ds = read.mtp("filename.mtp") # Minitab worksheet
> ds = read.octave("filename.octave") # Octave
> ds = read.ssd("filename.ssd") # SAS version 6
> ds = read.xport("filename.xport") # SAS XPORT file
> ds = read.spss("filename.sav") # SPSS
> ds = read.dta("filename.dta") # Stata
> ds = read.systat("filename.sys") # Systat
```

Using RStudio (1/3)

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains an R script with various data import functions:


```

123 library(foreign)
124 ds = read.dbf("filename.dbf") # DBase
125 ds = read.epiinfo("filename.epiinfo") # Epi Info
126 ds = read.mtp("filename.mtp") # Minitab worksheet
127 ds = read.octave("filename.octave") # Octave
128 ds = read.ssd("filename.ssd") # SAS version 6
129 ds = read.xport("filename.xport") # SAS XPORT file
130 ds = read.spss("filename.sav") # SPSS
131 ds = read.dta("filename.dta") # Stata
132 ds = read.systat("filename.sys") # Systat
133
134
135
136 # Using RStudio
137
138
139

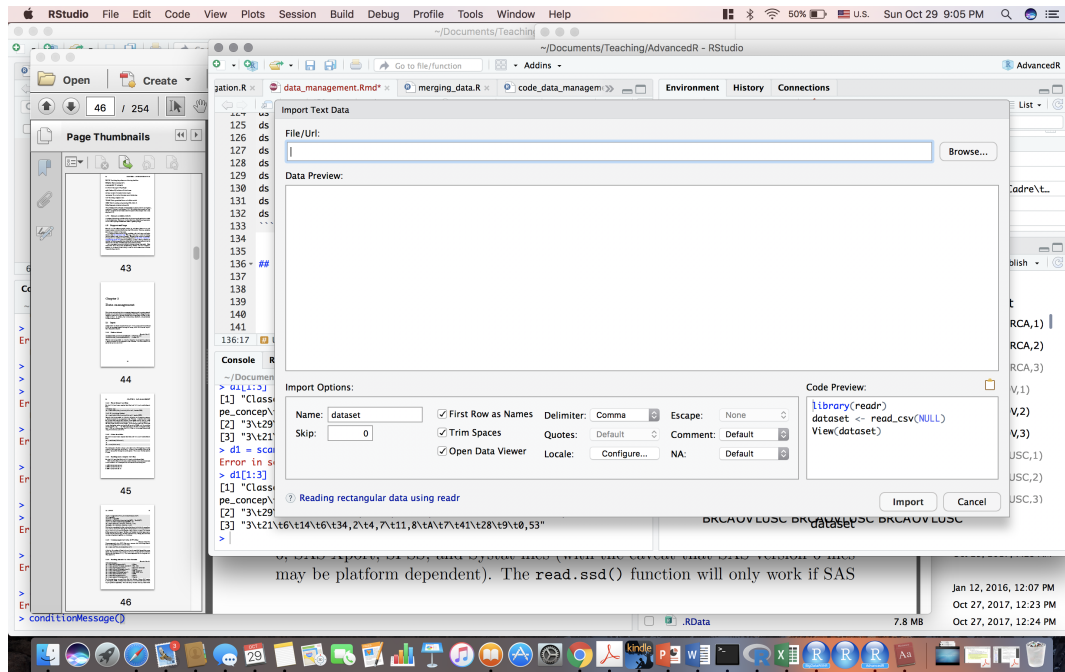
```
- Console:** Shows the execution of `scan("banques.txt")` resulting in an error:


```

Error in scan("banques.txt") : scan() expected 'a real', got 'Classe_PV'
> d1[1:3]
[1] "Classe_PV\tAge_PV\tCadre\tNon_Cadre\tDiplome\tAge_Moy\tAnc_Moy\tSurface\tType_concept\tNbr_reclam\tAge_RPV\tAnc_RPV\tAnc_RPV\tQual_client"
[2] "3\t129\t14\t10\t10\t41,71\t5,54\t12,75\t16\t16\t38\t21\t120\t0,37"
[3] "3\t121\t6\t14\t6\t34,2\t4,7\t11,8\t14\t7\t41\t28\t19\t0,53"

```
- Environment:** A dropdown menu is open, showing options for importing data from various sources like Text, Excel, SPSS, etc.
- Files:** A file browser window is open, showing a directory structure with files like `example_NZ_elections_missing_values.R`, `example_visu_genom_data.R`, and `visu_genom_data.Rmd`.
- Terminal:** Shows a message: "may be platform dependent). The read.ssd() function will only work if SAS".

Using RStudio (2/3)



Using RStudio (3/3)

Import Text Data

File/Url: ~/Downloads/pres_polls.csv

Data Preview:

Day (double)	Len (integer)	State (character)	EV (integer)	Dem (integer)	GOP (integer)	Ind (character)	Date (character)	X9 (character)	X10 (character)
335.0	1	Alabama	9	35	63	02	Dec 01	NA	NA
307.0	7	Alabama	9	36	54	06	Nov 06	NA	NA
300.0	7	Alabama	9	36	54	06	Oct 30	NA	NA
295.0	9	Alabama	9	36	52	08	Oct 26	NA	NA
293.0	7	Alabama	9	37	51	08	Oct 23	NA	NA
232.5	24	Alabama	9	31	53	08	Sep 01	NA	NA
136.0	31	Alabama	9	33	52	NA	May 31	NA	NA
46.0	29	Alabama	9	31	53	NA	Feb 29	NA	NA

Import Options:

Name: pres_polls First Row as Names Delimiter: Comma Escape: None

Skip: 0 Trim Spaces Quotes: Default Comment: Default

Open Data Viewer Locale: Configure... NA: Default

Code Preview:

```
library(readr)
pres_polls <- read_csv("~/Downloads/pres_polls.csv")
View(pres_polls)
```

Console:

```
Error in View: object 'dataset' not found
```

may be platform dependent). The `read.ssd()` function will only work if SAS

Jan 12, 2016, 12:07 PM
Oct 27, 2017, 12:23 PM
Oct 27, 2017, 12:24 PM

Using Rcmdr package (1/2)

```
> install.packages("Rcmdr")  
> library(Rcmdr)
```

Using Rcmdr package (2/2)

The screenshot shows the R Commander application window. The 'Data' menu is open, highlighting 'Import data'. The console shows the following output:

```

Rcmdr Version 2.4-1
Attaching package: 'Rcmdr'

The following object is masked from 'package:shiny':

  radIoButtons
  
```

The package list on the right shows the following installed packages:

Package	Description	Version
ggplot2	Create Elegant Data Visualisations Using the Grammar of Graphics	2.2.1
ggplotgui	Create Ggplots via a Graphical User Interface	1.0.0
cowplot	Streamlined Plot Theme and Plot Annotations for 'ggplot2'	0.8.0
ggbiplot	A ggplot2 based biplot	0.55
ggeffects	Create Tidy Data Frames of Marginal Effects for 'ggplot' from Model Outputs	0.2.2.9000
ggmap	Spatial Visualization with ggplot2	2.6.1
ggpubr	'ggplot2' Based Publication Ready Plots	0.1.5.999
ggrepel	Repulsive Text and Label Geoms for 'ggplot2'	0.7.0
ggsoci	Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'	2.8
ggsignif	Significance Brackets for 'ggplot2'	0.4.0
ggthemes	Extra Themes, Scales and Geoms for 'ggplot2'	3.4.0
survminer	Drawing Survival Curves using 'ggplot2'	0.4.0
ggplot2	Create Elegant Data Visualisations Using the Grammar of Graphics	2.2.1
GGally	Extension to 'ggplot2'	1.3.2
ggdendro	Create Dendrograms and Tree Diagrams Using 'ggplot2'	0.1-20

The console also shows a message about the `readLines()` function:

```

The readLines() function reads arbitrary text, while read.table() can be used to read a file with cases corresponding to lines and variables to fields in the file (the header option sets variable names to entries in the first line)
  
```

Using GrapheR (1/2)

```
> library(GrapheR)
> run.GrapheR()
```


Using Grapher (2/2)

The screenshot shows the XQuartz window with the Grapher application. The main window has a toolbar with icons for data, window, DRAW, and various drawing tools. Below the toolbar, there are sections for 'Loading the dataset', 'External file', and 'Existing R objects'. The 'External file' section includes dropdown menus for Extension (txt), Column separator (space / tabulation), and Decimal separator (point), along with a text field for Code for missing values (NA) and a checkbox for Variables name indicated in the file. The 'Existing R objects' section has an empty list box and a 'Load' button. To the right, there is a 'Connections' panel showing 'Environment is empty' and a 'Help Viewer' panel displaying a list of packages with their descriptions and versions.

The Help Viewer panel shows the following table:

Description	Version
A Multi-Platform GUI for Drawing Customizable Graphs in R	1.9-86
Alluvial Diagrams	0.1-2
Extending 'Dendrogram' Functionality in R	1.5.2
A Grammar of Data Manipulation	0.7.4
Interface to 'Dygraphs' Interactive Time Series Charting Library	1.1.1.4
Functions for Medical Statistics Book with some Demographic Data	0.6.1
Utilities for Graphical Rendering	0.1.6
Create Elegant Data Visualisations Using the Grammar of Graphics	2.2.1
Create Ggplots via a Graphical User Interface	1.0.0
Interactive Grammar of Graphics	0.4.3
Miscellaneous Functions for "Grid" Graphics	2.3
Java Graphics Device	0.6-1
D3 JavaScript Network Graphs from R	0.4
Programmatic Conversion of PDF Tables	0.1
Programme for International Student Assessment (PISA)	1.0
Create Interactive Web Graphics via 'plotly.js'	4.7.1
Propensity Score Analysis Graphics	2.1.1

The terminal window shows the following commands and output:

```

> library("Grapher", lib.loc=~/.Library/R/3.4/Library")
Loading required package: tcltk
*** Grapher v 1.9-86 ***
Available languages: English, French, German, Spanish.
Use run.Grapher(C) to launch or re-launch the interface.
> run.Grapher(C)
<Tcl>
>

```

Using JGR and deducer (1/2)

```
> Sys.setenv(JAVA_HOME = 'JAVA_HOME=/Library/Java/JavaVirtualMachines/1.6.0.jdk/Contents/Home')
> dyn.load("/Library/Java/JavaVirtualMachines/jdk1.8.0_121.jdk/Contents/Home/jre/lib/server/libjvm.dylib")
> library(rJava)
> library(JGR)
> library(Deducer)
> JGR()
```

Using JGR and deducer (2/2)

The screenshot displays the JGR application window. The main area is a 'Data Viewer' showing a table of flight data. The console on the left shows the following commands and output:

```

You are welcome to redistribute it under c
Type 'license()' or 'licence()' for distri
Natural language support but running in
R is a collaborative project with many con
Type 'contributors()' for more informatio
'citation()' on how to cite R or R package
Type 'demo()' for some demos, 'help()' for
'help.start()' for an HTML browser interfa
Type 'q()' to quit R.

[Previously saved workspace restored]
>
Loading required package: JGR
Loading required package: rJava
Loading required package: Jvarkit
starting httpd help server ... done
> JGR::package.manager()
> Variable_description <- read.table("/Us
> library(DeducerExtras)
Loading required package: Deducer
Loading required package: ggplot2
Loading required package: car
Loading required package: MASS
Loading required package: irr
Loading required package: lpsolve
> Variable_description1 <- read.table("/Us
> library(DeducerExtras)
> java.lang.NullPointerException

```

The 'Data Viewer' window shows a table with the following columns: YEAR, MONTH, DAY_OF_M., DAY_OF_W., FL_DATE, UNIQUE_C., AIRLINE_ID, TAIL_NUM, FL_NUM, and ORIGIN. The data is sorted by YEAR (2015) and MONTH (9).

YEAR	MONTH	DAY_OF_M.	DAY_OF_W.	FL_DATE	UNIQUE_C.	AIRLINE_ID	TAIL_NUM	FL_NUM	ORIGIN
2015	9	21	1	2015-09-21	DL	N967AT	2845	12892	
2015	9	21	1	2015-09-21	DL	N967AT	2845	14831	
2015	9	21	1	2015-09-21	DL	N353NB	2846	13495	
2015	9	21	1	2015-09-21	DL	N607AT	2847	12892	
2015	9	21	1	2015-09-21	DL	N607AT	2848	12892	
2015	9	21	1	2015-09-21	DL	N607AT	2848	14831	
2015	9	21	1	2015-09-21	DL	N370NB	2850	10397	
2015	9	21	1	2015-09-21	DL	N370NB	2850	11540	
2015	9	21	1	2015-09-21	DL	N319NB	2851	14908	
2015	9	21	1	2015-09-21	DL	N333NB	2852	14869	
2015	9	21	1	2015-09-21	DL	N928AT	2853	10397	
2015	9	21	1	2015-09-21	DL	N928AT	2853	15370	
2015	9	22	2	2015-09-22	DL	N909DE	7	10397	
2015	9	22	2	2015-09-22	DL	N928AT	2853	10397	
2015	9	22	2	2015-09-22	DL	N519US	8	13303	
2015	9	22	2	2015-09-22	DL	N340NW	11	13487	
2015	9	22	2	2015-09-22	DL	N675DL	14	14869	
2015	9	22	2	2015-09-22	DL	N684DA	15	10397	
2015	9	22	2	2015-09-22	DL	N841DN	16	12892	
2015	9	22	2	2015-09-22	DL	N839DN	17	10397	
2015	9	22	2	2015-09-22	DL	N370NW	18	11697	
2015	9	22	2	2015-09-22	DL	N962DN	21	11433	
2015	9	22	2	2015-09-22	DL	N948DL	28	13244	
2015	9	22	2	2015-09-22	DL	N957DL	29	10397	
2015	9	22	2	2015-09-22	DL	N935DL	30	11298	
2015	9	22	2	2015-09-22	DL	N932DL	31	10397	
2015	9	22	2	2015-09-22	DL	N983AT	42	13303	
2015	9	22	2	2015-09-22	DL	N944DN	52	11433	
2015	9	22	2	2015-09-22	DL	N579NW	53	13487	
2015	9	22	2	2015-09-22	DL	N345NB	54	12266	
2015	9	22	2	2015-09-22	DL	N340NB	55	10397	
2015	9	22	2	2015-09-22	DL	N808DN	61	14771	
2015	9	22	2	2015-09-22	DL	N998DL	62	10721	
2015	9	22	2	2015-09-22	DL	N811DZ	64	14771	
2015	9	22	2	2015-09-22	DL	N590NW	65	10397	
2015	9	22	2	2015-09-22	DL	N393DA	72	12889	
2015	9	22	2	2015-09-22	DL	N904DE	74	13303	
2015	9	22	2	2015-09-22	DL	N140LL	80	12892	
2015	9	22	2	2015-09-22	DL	N1613B	81	10397	
2015	9	22	2	2015-09-22	DL	N603AT	82	13244	
2015	9	22	2	2015-09-22	DL	N904DA	84	15304	
2015	9	22	2	2015-09-22	DL	N581NW	86	14679	
2015	9	22	2	2015-09-22	DL	N582NW	87	11433	
2015	9	22	2	2015-09-22	DL	N535US	89	14679	
2015	9	22	2	2015-09-22	DL	N366NB	94	10397	

Web Scraping

From Wikipedia pages (1/2)

https://en.wikipedia.org/wiki/Upper_Peninsula_of_Michigan

WIKIPEDIA

Coordinates: 46°14′00″N 86°21′00″W

Upper Peninsula of Michigan



The **Upper Peninsula** (**the UP**), also known as **Upper Michigan**, is the northern of the two major peninsulas that make up the U.S. state of Michigan. The peninsula is bounded on the north by Lake Superior, on the east by the St. Marys River, on the southeast by Lake Michigan and Lake Huron, and on the southwest by Wisconsin.

The Upper Peninsula contains 29% of the land area of Michigan but just 3% of its total population. Residents are frequently called Yoopers (derived from "U.P.-ers") and have a strong regional identity. Large numbers of French Canadian, Finnish, Swedish, Cornish, and Italian immigrants came to the Upper Peninsula, especially the Keweenaw Peninsula, to work in the area's mines and lumber industry. The peninsula includes the only counties in the United States where a plurality of residents claim Finnish ancestry.^[1]

Ordered by size, the peninsula's largest cities are Marquette, Sault Ste. Marie, Escanaba, Menominee, Houghton, and Iron Mountain. The land and climate are not very suitable for agriculture because of the long harsh winters. The economy has been based on logging, mining, and tourism. Most mines have closed since the "golden age" from 1890 to 1920. The land is heavily forested and logging remains a major industry.

Contents

- 1 **History**
- 2 **Geography**
 - 2.1 Wildlife
 - 2.2 Climate
 - 2.3 Time zones
- 3 **Government**
 - 3.1 Politics
 - 3.2 Proposed statehood
- 4 **Demographics**
- 5 **Economy**
 - 5.1 Industries
 - 5.2 Notable attractions

Upper Peninsula Michigan	
	
The Lake of the Clouds in the Porcupine Mountains of the Upper Peninsula of Michigan	
Nickname: <i>The UP, The 906</i>	
Country	United States
State	Michigan
Highest point	
- location	Mount Arvon
- elevation	1,979 ft (603 m)
Area	16,377 sq mi (42,416 km ²)
Population	311,361 (2010)
Density	19/sq mi (7/km ²)
Area code	906
	

From Wikipedia pages (2/2)

```
> install.packages("htmltab")
```

```
> library(htmltab)  
> htmltab("http://en.wikipedia.org/wiki/Upper_Peninsula_of_Michigan", 3)
```

```
## Warning in strptime(x, fmt, tz = "GMT"): unknown timezone 'zone/tz/2017c.  
## 1.0/zoneinfo/Africa/Tunis'
```

##	Year	REP	DEM	Others
## 2	2016	56.40%82,018	37.77% 54,923	5.83% 8,476
## 3	2012	50.80%73,529	47.49% 68,747	1.71% 2,477
## 4	2008	46.12% 69,647	51.82%78,257	2.06% 3,108
## 5	2004	51.52%78,276	47.31% 71,888	1.17% 1,781
## 6	2000	50.61%70,256	45.95% 63,791	3.43% 4,768

Using XML and httr package (1/2)

```
> library(httr)
> library(XML)
> url <- "https://en.wikipedia.org/wiki/Upper_Peninsula_of_Michigan"
> r <- GET(url)
> r
```

```
## Response [https://en.wikipedia.org/wiki/Upper_Peninsula_of_Michigan]
##   Date: 2017-11-15 21:11
##   Status: 200
##   Content-Type: text/html; charset=UTF-8
##   Size: 416 kB
## <!DOCTYPE html>
## <html class="client-nojs" lang="en" dir="ltr">
## <head>
## <meta charset="UTF-8"/>
## <title>Upper Peninsula of Michigan - Wikipedia</title>
## <script>document.documentElement.className = document.documentElement.cl...
## <script>(window.RLQ=window.RLQ||[]).push(function(){mw.config.set({"wgCa...
## mw.user.tokens.set({"editToken":"+\\","patrolToken":"+\\","watchToken":"...
##
## });mw.loader.load(["ext.cite.ally","site","mediawiki.page.startup","medi...
## ...
```

Using XML and httr package (2/2)

```
> doc <- readHTMLTable(
+   doc=content(r, "text"))
> doc[6]
```

```
## $`Upper Peninsula land area and population density by county[6]`
##      County Population Land area (sq mi) Land area (km2)
## 1      Alger      9,601           915      2,370
## 2      Baraga     8,860           898      2,330
## 3      Chippewa   38,520          1,558     4,040
## 4      Delta     37,069          1,171     3,030
## 5      Dickinson 26,168           761      1,970
## 6      Gogebic   16,427          1,101     2,850
## 7      Houghton 36,628          1,009     2,610
## 8      Iron      11,817          1,166     3,020
## 9      Keweenaw   2,156           540      1,400
## 10     Luce       6,631           899      2,330
## 11     Mackinac  11,113          1,021     2,640
## 12     Marquette 67,077          1,808     4,680
## 13     Menominee 24,029          1,044     2,700
## 14     Ontonagon 6,780           1,311     3,400
## 15     Schoolcraft 8,485          1,171     3,030
## 16     Total    311,361        16,377    42,420
##      Population density\n(per sq mi) Population density\n(per km2)
## 1              10.5              4.1
## 2              9.8              3.8
## 3             24.7              9.5
## 4             31.6             12.2
## 5             34.4             13.3
## 6             14.9              5.8
## 7             36.3             14.0
## 8             10.1              3.9
## 9              4.0              1.5
## 10             7.3              2.8
## 11            10.8              4.2
## 12            37.1             14.3
## 13            23.0              8.9
## 14              5.1              2.0
## 15              7.2              2.8
## 16            19.0              7.3
```


Data available in the web?

- from R packages
- Available in web pages

From R packages

Example: WDI package

Search, extract and format data from the World Bank's World Development Indicators

<https://data.worldbank.org/data-catalog/world-development-indicators>

or

<https://www.indexmundi.com/facts/indicators>

Exemple: Create a data about R&D in the world (1/5)

```
> library(WDI)
> # Collecting the available indicators
> ind_data=rbind.data.frame(WDIsearch("research"),
+                           WDIsearch("technology"),
+                           WDIsearch("technical")
+                           ,WDIsearch("scientific"))
> # Collecting the data from 1960 to 2016.
> all_data=WDI(indicator = ind_data$indicator,start = 1960,end = 2016)
```

Exemple: Create a data about R&D in the world (2/5)

The indicator data

Show entriesSearch:

indicator	name
GB.XPD.RSDV.GD.ZS	Research and development expenditure (% of GDP)
SP.POP.SCIE.RD.P6	Researchers in RD (per million people)
IC.FRM.TECH.ZS	Firms using technology licensed from foreign companies (% of firms)
IE.ICT.PCAP.CD	Information and communication technology expenditure per capita (current US\$)
IE.ICT.TOTL.CD	Information and communication technology expenditure (current US\$)

Showing 1 to 5 of 24 entries

[Previous](#)[1](#)[2](#)[3](#)[4](#)[5](#)[Next](#)

Exemple: Create a data about R&D in the world (3/5)

The research data (after reshaping it and from 2010 to 2014)

```
> library(reshape)
> i=which(all_data$year>=2010 & all_data$year<=2014)
> x=melt(all_data[i,],id.vars = colnames(all_data)[1:3])
> colnames(x)[4]="indicator"
> datatable(x,options = list(pageLength = 5), rownames = FALSE)
```

Exemple: Create a data about R&D in the world (4/5)

Show entriesSearch:

iso2c	country	year	indicator	value
AD	Andorra	2010	IPJRN.ARTC.SC	1.2
AD	Andorra	2011	IPJRN.ARTC.SC	1
AD	Andorra	2012	IPJRN.ARTC.SC	3.9
AD	Andorra	2013	IPJRN.ARTC.SC	5.9
AD	Andorra	2014	IPJRN.ARTC.SC	

Showing 1 to 5 of 25,872 entries

Previous

[2](#)[3](#)[4](#)[5](#)[...](#)[5175](#)

Next

Exemple: Create a data about R&D in the world (5/5)

I have also created a Shiny App to better explore this data

https://dhafer.shinyapps.io/Research_Indicators/

Please Wait



pisa package

Data from Programme of International Student Assessment (PISA).

Installation of `pisa` package

```
> require(devtools)
> install_github('pisa', 'jbryer', force=T)
```

Some available data (1/3)

School results from the 2009 Programme of International Student Assessment (PISA) as provided by the Organization for Economic Co-operation and Development (OECD).

See <http://www.pisa.oecd.org/> for more information including the code book.

Some available data (2/3)

```
> library(pisa)
> data(pisa.school)
> head(pisa.school)
```

```
##      CNT COUNTRY      OECD SUBNATIO SCHOOLID SC01Q01 SC01Q02 SC01Q03
## 1 Albania Albania Non-OECD Albania 00001      No      No      No
## 2 Albania Albania Non-OECD Albania 00002      Yes     Yes     Yes
## 3 Albania Albania Non-OECD Albania 00003      Yes     Yes     Yes
## 4 Albania Albania Non-OECD Albania 00004      No      No      No
## 5 Albania Albania Non-OECD Albania 00005      Yes     Yes     Yes
## 6 Albania Albania Non-OECD Albania 00006      Yes     Yes     Yes
##      SC01Q04 SC01Q05 SC01Q06 SC01Q07 SC01Q08 SC01Q09 SC01Q10 SC01Q11 SC01Q12
## 1      No      No      No      No      No      No      Yes     Yes     Yes
## 2      Yes     Yes     Yes     Yes     Yes     Yes     No      No      No
## 3      Yes     Yes     Yes     Yes     Yes     Yes     Yes     Yes     Yes
## 4      No      No      No      No      No      No      Yes     Yes     Yes
## 5      Yes     Yes     Yes     Yes     Yes     Yes     No      No      No
## 6      Yes     Yes     Yes     Yes     Yes     Yes     No      No      No
##      SC01Q13 SC01Q14 SC02Q01 SC03Q01 SC03Q02 SC03Q03 SC03Q04 SC04Q01
## 1      <NA>    <NA>    Public    60     40     0     0     Town
## 2      <NA>    <NA>    Public    90     10     0     0     Village
## 3      <NA>    <NA>    Public   100     0     0     0     Town
## 4      <NA>    <NA>    Public    90     10     0     0     Village
## 5      <NA>    <NA>    Public    95     5     0     0     Town
## 6      <NA>    <NA>    Public    99     0     0     1     City
##      SC05Q01 SC06Q01 SC06Q02 SC07Q01 SC07Q02 SC08Q01 SC09Q11
## 1 Two or More Schools 285 316 <NA> <NA> None NA
## 2 Two or More Schools 93 118 <NA> <NA> None 10
## 3 Two or More Schools 475 581 <NA> <NA> None 47
## 4 One Other 121 163 <NA> <NA> None 22
## 5 Two or More Schools 219 187 <NA> <NA> None 26
## 6 Two or More Schools 530 472 <NA> <NA> None 44
##      SC09Q12 SC09Q21 SC09Q22 SC09Q31 SC09Q32 SC10Q01 SC10Q02 SC10Q03
## 1      2      NA      0      NA      0      212     30      0
## 2      0      10     0      10     0      25      0      0
## 3      2      47     2      41     2      79     15     1
## 4      0      22     0      22     0      87     15     1
## 5      0      26     0      21     0      56      8     1
## 6      5      44     5      36     5     123     18     18
##      SC11Q01 SC11Q02 SC11Q03 SC11Q04 SC11Q05
## 1 Not at all Not at all Not at all Very little A lot
## 2 Very little Very little Not at all Very little Very little
## 3 Not at all Not at all Not at all Not at all Very little
## 4 Not at all Not at all Not at all Very little Not at all
## 5 Not at all Very little Not at all Not at all Not at all
## 6 Not at all Not at all Not at all Very little Not at all
```

```

##          SC11Q06      SC11Q07      SC11Q08      SC11Q09      SC11Q10
## 1 To some extent To some extent Very little      Very little      A lot
## 2      Not at all      A lot Very little      A lot      A lot
## 3      Not at all      Very little Not at all      Very little      A lot
## 4      Very little      A lot Very little To some extent Very little
## 5      Not at all To some extent Very little To some extent      A lot
## 6      Very little To some extent Not at all To some extent Not at all
##          SC11Q11      SC11Q12      SC11Q13      SC12Q01
## 1 Not at all      A lot      Not at all Not for any subject
## 2      A lot      Very little      A lot      For all subjects
## 3 Not at all      Very little To some extent Not for any subject
## 4 Very little To some extent      Very little      For some subjects
## 5      A lot      A lot      A lot      For some subjects
## 6 Very little To some extent To some extent Not for any subject
##          SC12Q02 SC13Q01 SC13Q02 SC13Q03 SC13Q04 SC13Q05 SC13Q06
## 1 Not for any subject      Yes      Yes      No      Yes      No      No
## 2      For all subjects      No      No      No      No      No      No
## 3      For some subjects      No      Yes      No      Yes      Yes      Yes
## 4 Not for any subject      Yes      Yes      Yes      Yes      Yes      Yes
## 5 Not for any subject      Yes      No      Yes      No      Yes      No
## 6      For some subjects      Yes      No      No      Yes      No      No
##          SC13Q07 SC13Q08 SC13Q09 SC13Q10 SC13Q11 SC13Q12 SC13Q13 SC13Q14 SC14Q01
## 1      Yes      Yes      Yes      Yes      No      No      No      Yes      <NA>
## 2      Yes      No      No      Yes      No      No      No      Yes      <NA>
## 3      Yes      Yes      Yes      Yes      No      Yes      No      Yes      <NA>
## 4      Yes      Yes      Yes      Yes      Yes      No      No      Yes      <NA>
## 5      Yes      No      Yes      Yes      Yes      Yes      No      Yes      <NA>
## 6      Yes      No      No      Yes      Yes      Yes      No      Yes      <NA>
##          SC14Q02 SC14Q03 SC14Q04 SC14Q05      SC15Q01      SC15Q02
## 1      <NA>      <NA>      <NA>      <NA> 1-2 times a year      Monthly
## 2      <NA>      <NA>      <NA>      <NA> 1-2 times a year      3-5 times a year
## 3      <NA>      <NA>      <NA>      <NA> 1-2 times a year More than once a month
## 4      <NA>      <NA>      <NA>      <NA> 1-2 times a year      3-5 times a year
## 5      <NA>      <NA>      <NA>      <NA> 3-5 times a year      Monthly
## 6      <NA>      <NA>      <NA>      <NA> 1-2 times a year More than once a month
##          SC15Q03      SC15Q04      SC15Q05 SC16Q01 SC16Q02
## 1      Never      1-2 times a year More than once a month      Yes      Yes
## 2      Never      1-2 times a year More than once a month      Yes      Yes
## 3      Never      Monthly More than once a month      Yes      Yes
## 4      Never      Monthly More than once a month      Yes      No
## 5      Never More than once a month More than once a month      Yes      Yes
## 6      Never      Never      3-5 times a year      Yes      Yes
##          SC16Q03 SC16Q04 SC16Q05 SC16Q06 SC16Q07 SC16Q08      SC17Q01
## 1      No      Yes      Yes      Yes      Yes      Yes To some extent
## 2      Yes      Yes      Yes      Yes      Yes      Yes      Not at all
## 3      Yes      No      Yes      No      Yes      No      Not at all
## 4      Yes      Yes      Yes      Yes      Yes      No      Not at all
## 5      Yes      Yes      Yes      No      Yes      Yes      Very little
## 6      Yes      No      Yes      Yes      Yes      No      Very little
##          SC17Q02      SC17Q03      SC17Q04      SC17Q05      SC17Q06
## 1 To some extent Very little Very little Not at all Very little
## 2      Very little Not at all Not at all Not at all Very little
## 3      Not at all Not at all Not at all Not at all Not at all

```

```

## 4 Very little Very little Very little Very little Not at all
## 5 Very little Not at all Very little Very little Very little
## 6 Very little Very little Very little Very little Very little
## SC17Q07 SC17Q08 SC17Q09 SC17Q10 SC17Q11 SC17Q12
## 1 To some extent Not at all Not at all Not at all Not at all Not at all
## 2 Not at all Not at all Not at all Not at all Not at all Not at all
## 3 Very little Not at all Not at all Not at all Not at all Not at all
## 4 Not at all Very little Not at all Not at all Very little Very little
## 5 Very little Very little Not at all Not at all Not at all Very little
## 6 Very little Not at all Not at all Not at all Not at all Not at all
## SC17Q13 SC18Q01 SC19Q01 SC19Q02 SC19Q03 SC19Q04
## 1 Very little Minority of Parents Always Sometimes Sometimes Never
## 2 Not at all Minority of Parents Always Always Always Never
## 3 Very little Minority of Parents Sometimes Always Always Always
## 4 Very little Minority of Parents Always Always Always Never
## 5 To some extent Minority of Parents Never Never Always Always
## 6 Very little Minority of Parents Sometimes Sometimes Sometimes Always
## SC19Q05 SC19Q06 SC19Q07 SC20Q01 SC20Q02 SC20Q03
## 1 Never Sometimes Sometimes Likely Not likely Likely
## 2 Never Always Always Very likely Very likely Not likely
## 3 Sometimes Sometimes <NA> Likely Not likely Likely
## 4 Always Sometimes Sometimes Not likely Likely Not likely
## 5 Sometimes Always Sometimes Very likely Not likely Very likely
## 6 Always Sometimes Sometimes Not likely Not likely Likely
## SC20Q04 SC20Q05 SC20Q06 SC21Q01 SC21Q02 SC21Q03 SC22Q01
## 1 Likely Likely Likely Yes Yes Yes No
## 2 Very likely Very likely Very likely Yes Yes Yes No
## 3 Likely Likely <NA> Yes No No No
## 4 Likely Likely Likely <NA> <NA> <NA> Yes
## 5 Very likely Likely Likely Likely Yes No No No
## 6 Likely Very likely Likely Likely Yes Yes No No
## SC22Q02 SC22Q03 SC22Q04 SC22Q05 SC23Q01 SC23Q02 SC23Q03 SC23Q04 SC24Qa1
## 1 Yes Yes Yes Yes Yes Yes Yes Yes No No Tick
## 2 Yes Yes Yes Yes Yes Yes Yes Yes Yes No Tick
## 3 Yes Yes Yes No Yes Yes Yes Yes Yes No Tick
## 4 Yes Yes No Yes Yes Yes Yes Yes Yes No Tick
## 5 Yes Yes Yes Yes Yes Yes Yes Yes Yes No Tick
## 6 Yes Yes Yes No Yes Yes Yes Yes Yes No Tick
## SC24Qa2 SC24Qa3 SC24Qa4 SC24Qa5 SC24Qb1 SC24Qb2 SC24Qb3 SC24Qb4 SC24Qb5
## 1 No Tick No Tick Tick No Tick No Tick No Tick No Tick Tick No Tick
## 2 No Tick No Tick Tick No Tick No Tick No Tick No Tick Tick No Tick
## 3 No Tick No Tick Tick No Tick No Tick No Tick No Tick Tick No Tick
## 4 No Tick No Tick Tick No Tick No Tick No Tick No Tick Tick No Tick
## 5 No Tick No Tick Tick No Tick No Tick No Tick No Tick Tick No Tick
## 6 No Tick No Tick Tick No Tick No Tick No Tick No Tick Tick No Tick
## SC24Qc1 SC24Qc2 SC24Qc3 SC24Qc4 SC24Qc5 SC24Qd1 SC24Qd2 SC24Qd3 SC24Qd4
## 1 No Tick No Tick No Tick No Tick Tick No Tick No Tick No Tick No Tick
## 2 No Tick No Tick No Tick No Tick Tick No Tick No Tick No Tick No Tick
## 3 No Tick No Tick No Tick No Tick Tick No Tick No Tick No Tick No Tick
## 4 No Tick No Tick No Tick No Tick Tick No Tick No Tick No Tick No Tick
## 5 No Tick No Tick No Tick No Tick Tick No Tick No Tick No Tick No Tick
## 6 No Tick No Tick No Tick No Tick Tick No Tick No Tick No Tick No Tick
## SC24Qd5 SC24Qe1 SC24Qe2 SC24Qe3 SC24Qe4 SC24Qe5 SC24Qf1 SC24Qf2 SC24Qf3

```

```

## 1   Tick No Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick   Tick
## 2   Tick No Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick   Tick
## 3   No Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick No Tick   Tick
## 4   Tick   Tick No Tick   Tick   Tick No Tick   Tick   Tick   Tick
## 5   Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick No Tick   Tick
## 6   Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick No Tick   Tick
##   SC24Qf4 SC24Qf5 SC24Qg1 SC24Qg2 SC24Qg3 SC24Qg4 SC24Qg5 SC24Qh1 SC24Qh2
## 1 No Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick   Tick
## 2 No Tick No Tick   Tick No Tick No Tick No Tick No Tick   Tick No Tick
## 3 No Tick No Tick   Tick No Tick No Tick No Tick No Tick No Tick No Tick   Tick
## 4 No Tick No Tick   Tick   Tick   Tick   Tick   Tick No Tick   Tick
## 5 No Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick
## 6 No Tick No Tick No Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick
##   SC24Qh3 SC24Qh4 SC24Qh5 SC24Qi1 SC24Qi2 SC24Qi3 SC24Qi4 SC24Qi5 SC24Qj1
## 1 No Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick No Tick
## 2 No Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick No Tick
## 3 No Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick No Tick
## 4 No Tick No Tick   Tick   Tick No Tick No Tick   Tick No Tick No Tick
## 5 No Tick No Tick   Tick   Tick No Tick No Tick No Tick No Tick No Tick
## 6 No Tick No Tick   Tick   Tick No Tick No Tick No Tick No Tick No Tick
##   SC24Qj2 SC24Qj3 SC24Qj4 SC24Qj5 SC24Qk1 SC24Qk2 SC24Qk3 SC24Qk4 SC24Qk5
## 1 No Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick   Tick
## 2   Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick   Tick No Tick
## 3   Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick
## 4   Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick   Tick
## 5   Tick No Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick No Tick
## 6   Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick   Tick
##   SC24Ql1 SC24Ql2 SC24Ql3 SC24Ql4 SC24Ql5 SC25Qa1 SC25Qa2 SC25Qa3 SC25Qa4
## 1 No Tick No Tick No Tick No Tick No Tick   Tick   Tick   Tick   Tick No Tick
## 2 No Tick No Tick No Tick   Tick No Tick No Tick No Tick   Tick No Tick
## 3   Tick No Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick No Tick
## 4 No Tick No Tick No Tick No Tick No Tick   Tick   Tick No Tick No Tick   Tick
## 5   Tick No Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick No Tick
## 6 No Tick No Tick No Tick No Tick No Tick   Tick   Tick No Tick   Tick   Tick
##   SC25Qb1 SC25Qb2 SC25Qb3 SC25Qb4 SC25Qc1 SC25Qc2 SC25Qc3 SC25Qc4 SC25Qd1
## 1 No Tick   Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick
## 2 No Tick   Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick
## 3 No Tick   Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick No Tick
## 4   Tick   Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick   Tick
## 5 No Tick   Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick
## 6 No Tick   Tick No Tick No Tick No Tick No Tick No Tick No Tick   Tick No Tick
##   SC25Qd2 SC25Qd3 SC25Qd4 SC25Qe1 SC25Qe2 SC25Qe3 SC25Qe4 SC25Qf1 SC25Qf2
## 1 No Tick No Tick   Tick No Tick No Tick No Tick No Tick No Tick No Tick
## 2 No Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick
## 3 No Tick   Tick No Tick No Tick   Tick No Tick No Tick No Tick No Tick
## 4 No Tick No Tick No Tick   Tick   Tick No Tick No Tick No Tick No Tick
## 5 No Tick   Tick No Tick No Tick No Tick No Tick No Tick No Tick No Tick
## 6 No Tick   Tick   Tick No Tick No Tick No Tick No Tick   Tick No Tick No Tick
##   SC25Qf3 SC25Qf4   SC26Q01   SC26Q02   SC26Q03   SC26Q04
## 1 No Tick   Tick Very often Very often Quite often Very often
## 2 No Tick No Tick Very often Very often Quite often Very often
## 3 No Tick   Tick Very often Very often Very often Very often
## 4   Tick   Tick Quite often Very often Very often Very often

```

```

## 5 No Tick No Tick Quite often Quite often Quite often Very often
## 6 No Tick Tick Very often Quite often Quite often Quite often
## SC26Q05 SC26Q06 SC26Q07 SC26Q08 SC26Q09 SC26Q10
## 1 Very often Very often Very often Very often Quite often Very often
## 2 Very often Very often Quite often Quite often Very often Very often
## 3 Quite often Very often Seldom Quite often Very often Quite often
## 4 Quite often Quite often Very often Quite often Very often Quite often
## 5 Quite often Very often Quite often Seldom Quite often Very often
## 6 Very often Quite often Very often Very often Very often Seldom
## SC26Q11 SC26Q12 SC26Q13 SC26Q14 SC27Q01
## 1 Very often Very often Very often Seldom Male
## 2 Very often Very often Very often Very often Male
## 3 Very often Very often Quite often Seldom Female
## 4 Quite often Very often Very often Quite often Female
## 5 Very often Quite often Very often Seldom Male
## 6 Very often Very often Quite often Quite often Male
## ABGROU COMPWEB IRATCOMP PCGIRLS PROPCERT PROPQUAL SCHSIZE
## 1 Not for any subjects 0.000 0.142 52.5790 NA NA 601
## 2 For all subjects NA 0.000 55.9242 1 1.000 211
## 3 For some subjects 0.067 0.190 55.0189 1 0.875 1056
## 4 For some subjects 0.067 0.172 57.3944 1 1.000 284
## 5 For some subjects 0.125 0.143 46.0591 1 0.808 406
## 6 For some subjects 1.000 0.146 47.1058 1 0.828 1002
## SCTYPE SELSCH STRATIO
## 1 Public At least one sometimes but neither always considered NA
## 2 Public At least one always considered 21.100
## 3 Public At least one always considered 22.000
## 4 Public At least one always considered 12.909
## 5 Public At least one always considered 15.615
## 6 Public At least one sometimes but neither always considered 21.548
## EXCURACT LDRSHP RESPCURR RESPRES SCMATEDU STUDBEHA TCHPARTI TCSHORT
## 1 0.1527 1.6750 -1.1970 -0.8256 -0.5602 0.3454 -1.2768 -0.2606
## 2 -1.2758 1.9462 -0.9125 -0.8256 -1.7503 1.6729 -1.2768 0.4283
## 3 0.7130 0.8741 1.3635 -0.8096 -0.2728 1.6729 -0.7877 -1.0222
## 4 1.4227 1.0501 -1.0548 -0.6658 -0.7064 0.6491 0.0910 -0.2606
## 5 0.4233 0.3866 -0.3435 -0.8096 -1.7503 0.3454 -0.7877 -0.2606
## 6 -0.1104 0.8741 -1.1970 -0.8096 -0.4165 0.9549 -1.2768 -0.2606
## TEACBEHA W_FSCHWT STRATUM VER_SCH
## 1 0.2091 1.3226 ALB: North/Urban/General P2009_07DEC10
## 2 1.4029 25.9638 ALB: South/Rural/General P2009_07DEC10
## 3 1.4029 4.1397 ALB: South/Urban/General P2009_07DEC10
## 4 0.4524 3.2925 ALB: South/Rural/General P2009_07DEC10
## 5 0.2091 7.5364 ALB: Center/Urban/General P2009_07DEC10
## 6 0.2091 1.7948 ALB: South/Urban/General P2009_07DEC10

```


Some available data (3/3)

Parent survey results from the 2009 Programme of International Student Assessment (PISA)

```
> data(pisa.parent)
> head(pisa.parent)
```

```
##      CNT  COUNTRY  OECD  SUBNATIO  SCHOOLID  STIDSTD  PA01Q01  PA01Q02  PA01Q03
## 1  CHL      152      1      15200      00001  00001      2      2      2
## 2  CHL      152      1      15200      00001  00002      2      2      1
## 3  CHL      152      1      15200      00001  00003      1      2      2
## 4  CHL      152      1      15200      00001  00004      2      2      1
## 5  CHL      152      1      15200      00001  00005      1      1      2
## 6  CHL      152      1      15200      00002  00006      1      2      2
##      PA02Q01  PA03Q01  PA03Q02  PA03Q03  PA03Q04  PA03Q05  PA03Q06  PA03Q07  PA03Q08
## 1          2          3          2          3          4          4          2          4          4
## 2          2          4          3          4          4          3          4          4          4
## 3          2          4          2          1          2          1          1          1          1
## 4          2          2          1          3          3          4          4          3          2
## 5          2          4          2          4          1          4          4          1          3
## 6          2          1          1          1          1          4          1          1          3
##      PA03Q09  PA04Q01  PA05Q01  PA06Q01  PA06Q02  PA06Q03  PA06Q04  PA07Q01  PA07Q02
## 1          4          1          4          2          2          4          2          2          2
## 2          4          1          4          2          2          4          2          2          2
## 3          NA          1          4          2          2          3          4          2          2
## 4          2          1          4          2          2          4          1          2          2
## 5          4          1          3          2          2          4          2          2          2
## 6          4          1          3          2          1          4          2          2          2
##      PA07Q03  PA07Q04  PA07Q05  PA07Q06  PA08Q01  PA08Q02  PA08Q03  PA08Q04  PA08Q05
## 1          2          2          2          1          1          2          3          4          3
## 2          2          2          2          1          1          3          3          4          4
## 3          2          2          2          1          3          1          3          4          4
## 4          2          2          1          1          4          4          4          4          2
## 5          2          2          2          1          1          2          1          2          3
## 6          2          2          2          1          1          2          3          3          4
##      PA08Q06  PA08Q07  PA08Q08  PA09Q01  PA09Q02  PA09Q03  PA09Q04  PA10Q01  PA10Q02
## 1          2          4          2          2          2          2          2          2          2
## 2          2          2          4          2          2          2          2          2          2
## 3          1          2          1          2          2          2          1          2          2
## 4          1          3          4          2          2          2          2          2          2
## 5          1          2          1          2          2          2          2          2          2
## 6          1          3          1          2          2          2          1          2          2
##      PA10Q03  PA10Q04  PA11Q01  PA12Q01  PA13Q01  PA14Q01  PA14Q02  PA14Q03  PA14Q04
## 1          2          2          1          3          2          1          1          2          2
## 2          2          1          1          2          3          1          1          1          1
## 3          2          1          3          2          3          1          1          1          1
## 4          1          1          1          2          4          2          2          1          1
```

```

## 5      2      2      1      3      2      2      2      2      2
## 6      2      1      2      2      2      2      2      2      2
##      PA14Q05 PA14Q06 PA14Q07 PA15Q01 PA15Q02 PA15Q03 PA15Q04 PA15Q05 PA15Q06
## 1      2      2      2      1      1      1      1      1      1
## 2      1      1      1      1      1      2      2      2      2
## 3      1      1      1      1      1      1      1      2      2
## 4      1      2      1      2      2      2      2      2      2
## 5      2      2      2      2      2      2      2      2      2
## 6      1      2      2      2      2      2      2      2      2
##      PA15Q07 PA15Q08 PA16Q01 PA17Q01 PA17Q02 PA17Q03 PA17Q04 PA17Q05 PA17Q06
## 1      1      1      3      3      4      2      2      2      2
## 2      2      2      3      1      4      4      4      1      1
## 3      1      2      3      3      3      3      3      NA      2
## 4      2      2      3      1      3      3      4      1      3
## 5      2      2      3      3      3      3      3      1      1
## 6      2      2      3      4      4      2      3      3      1
##      PA17Q07 PA17Q08 PA17Q09 PA17Q10 PA17Q11 PQMISCED PQFISCED PQHISCED
## 1      4      3      3      3      4      0      0      0
## 2      2      3      4      4      4      1      0      1
## 3      3      4      4      4      4      1      1      1
## 4      3      3      3      2      4      2      0      2
## 5      3      3      3      3      3      0      0      0
## 6      3      1      4      3      4      1      1      1
##      CURSUPP MOTREAD PARINVOL PQSCHOOL PRESUPP READRES STRATUM      VER_PAR
## 1      NA      NA      NA      NA      NA      NA      NA      15297 P2009_07DEC10
## 2      NA      NA      NA      NA      NA      NA      NA      15297 P2009_07DEC10
## 3      NA      NA      NA      NA      NA      NA      NA      15297 P2009_07DEC10
## 4      NA      NA      NA      NA      NA      NA      NA      15297 P2009_07DEC10
## 5      NA      NA      NA      NA      NA      NA      NA      15297 P2009_07DEC10
## 6      NA      NA      NA      NA      NA      NA      NA      15297 P2009_07DEC10

```

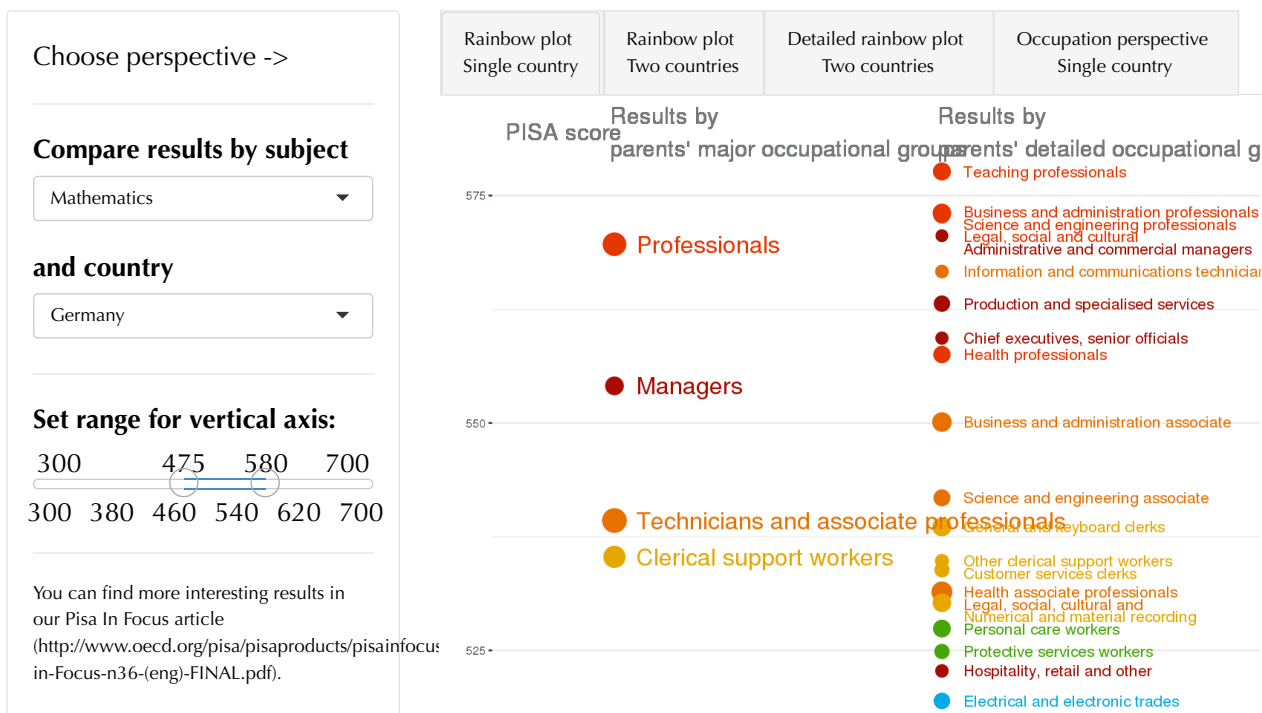
Shiny App to Explore PISA data

<http://mi2.mini.pw.edu.pl:8080/SmarterPoland/PISAoccupations2012/>

Occupations@PISA2012

How much can we infer about a student's performance in school by looking at what his or her parents do for a living? To find out, PISA 2012 asked participating students about their parents' occupations.

Occupations@PISA2012 is a web-based application that allows you to explore the relationship between parents' occupations and their children's performance in mathematics, reading and science - in your own country and in other countries.



From pdf documents

tabulizer package (I)

Installation of the package

```
> library("ghit")  
> ghit::install_github(c("ropensci/tabulizerjars", "ropensci/tabulizer"))
```

Case study: Size of the population and the Schools in Tunis (2013 to 2015)

Our aim is

- to obtain a data on the delegations of Tunis about the population, schools and the numbers de classrooms
- to display an interactive graph

Source of the data

- We download a pdf document in the General Commission of the Regional Development.
- We are interested in Tunis report that can be downloaded from the following link

<http://www.cgdr.nat.tn/upload/files/gouvchiffres/gech2015/Tunis.pdf>

- Put the file “Tunis.pdf” in your working directory.

Extracting the data on population from the pdf reports.

- We extract the table of the page 8.

```
> library(tabulizer)
> tabl <- extract_tables("Tunis.pdf", pages = 8)
> tabl
```

```
## [[1]]
##      [,1]      [,2]      [,3]      [,4]      [,5]
## [1,] ""      "DELEGATION" "2015"  "2014"  "2013"
## [2,] "Carthage" ""      "24906" "24216" "24100"
## [3,] "Tunis Medina" ""      "22009" "21400" "21298"
## [4,] "Bab Bhar" ""      "37241" "36210" "36037"
## [5,] "Bab Souika" ""      "30016" "29185" "29046"
## [6,] "El Omran" ""      "43410" "42208" "42006"
## [7,] "Omrane Superieur" ""      "57094" "55513" "55248"
## [8,] "Tahrir" ""      "22327" "21709" "21605"
## [9,] "Menzah" ""      "43021" "41830" "41630"
## [10,] "Cité Khadra" ""      "36175" "35173" "35005"
## [11,] "Le Bardo" ""      "74010" "71961" "71617"
## [12,] "Sejoumi" ""      "34835" "33870" "33708"
## [13,] "Ezzouhour" ""      "41888" "40728" "40534"
## [14,] "El Hrairia" ""      "113322" "110184" "109658"
## [15,] "Sidi Hessine" ""      "112795" "109672" "109148"
## [16,] "El Ouardia" ""      "33063" "32147" "31993"
## [17,] "El Kabaria" ""      "88474" "86024" "85613"
## [18,] "Sidi El Bechir" ""      "28539" "27749" "27616"
## [19,] "Jebel Jelloud" ""      "24311" "23638" "23525"
## [20,] "La Goulette" ""      "47013" "45711" "45493"
## [21,] "El Kram" ""      "76243" "74132" "73778"
## [22,] "La Marsa" ""      "95635" "92987" "92543"
## [23,] ""      "Total"  "1086327" "1056247" "1051203"
##      [,6]      [,7]
## [1,] "المعمدية" ""
## [2,] ""      "قرطاج"
## [3,] ""      "المدينة"
## [4,] ""      "باب بحر"
## [5,] ""      "باب سويقة"
## [6,] ""      "العمران"
## [7,] ""      "العمران الأعلى"
## [8,] ""      "التحرير"
## [9,] ""      "المنزه"
## [10,] ""      "حي الخضراء"
## [11,] ""      "باردو"
## [12,] ""      "السيجومي"
## [13,] ""      "الزهور"
```



```
## [14,] ""      "الحرائرية"  
## [15,] ""      "سيدي حسين"  
## [16,] ""      "الوردية"  
## [17,] ""      "الكبارية"  
## [18,] ""      "سيدي البشير"  
## [19,] ""      "جبل الجلود"  
## [20,] ""      "حلق الوادي"  
## [21,] ""      "الكرم"  
## [22,] ""      "المرسى"  
## [23,] "الولاية" ""
```

Cleaning the data into R

- We transform `tab1` to a `data.frame` object

```
> dt=tab1[[1]]
> colnames(dt)[c(1,3,4,5,7)]=tab1[[1]][1,2:6]
> dt=dt[1:22,c(1,3,4,5,7)]
> dt=dt[-1,]
> dt[,2]=as.numeric(as.character(dt[,2]))
> dt[,3]=as.numeric(as.character(dt[,3]))
> dt[,4]=as.numeric(as.character(dt[,4]))
> dt=as.data.frame(dt)
```

Reshaping the data

```
> library(reshape2)
```

```
##  
## Attaching package: 'reshape2'
```

```
## The following objects are masked from 'package:reshape':  
##  
##   colsplit, melt, recast
```

```
> dtA=melt(dt,id.vars = colnames(dt)[c(1,5)])  
> colnames(dtA)[3:4]=c("Year","Population")
```

Displaying the data on population sizes in each delegation

```
> DT::datatable(dtA)
```

Show entries

Search:

	DELEGATION	المعتمدية	Year	Population
1	Carthage	قرطاج	2015	24906
2	Tunis Medina	المدينة	2015	22009
3	Bab Bhar	باب بحر	2015	37241
4	Bab Souika	باب سوقية	2015	30016
5	El Omran	العمران	2015	43410
6	Omrane Superieur	العمران الأعلى	2015	57094
7	Tahrir	التحرير	2015	22327
8	Menzah	المنزه	2015	43021
9	Cité Khadra	حي الخضراء	2015	36175
10	Le Bardo	باردو	2015	74010

Showing 1 to 10 of 63 entries

Previous 2 3 4 5 6 7 Next

Extracting now the data on Schools and classrooms

We can extract a second table: Number of schools and Classrooms (page 14)

```
> tab2 <- extract_tables("Tunis.pdf", pages = 14)
> tab2
```

```
## [[1]]
##      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] ""      "salles d'études" ""      "قاعات" "écoles" ""
## [2,] "DELEGATION" ""      ""      ""      ""      ""
## [3,] ""      "2015"      "2014" "2013" "2015" "2014"
## [4,] "CARTHAGE" "76"      "73"    "59"    "8"     "8"
## [5,] "TUNIS MEDINA" "67"     "65"    "67"    "6"     "6"
## [6,] "BAB BHAR" "97"     "97"    "96"    "8"     "8"
## [7,] "BAB SOUIKA" "46"     "50"    "49"    "4"     "4"
## [8,] "EL OMRAN" "114"    "125"   "111"   "10"    "10"
## [9,] "OMRANE SUPERIEUR" "108"   "109"   "109"   "9"     "9"
## [10,] "TAHRIR" "45"     "45"    "46"    "4"     "4"
## [11,] "MENZAH" "58"     "56"    "52"    "6"     "6"
## [12,] "CITE KHADRA" "64"     "63"    "61"    "5"     "5"
## [13,] "LE BARDO" "138"    "136"   "136"   "13"    "13"
## [14,] "SEJOUIMI" "34"     "34"    "37"    "3"     "3"
## [15,] "EZZOUHOUR" "90"     "91"    "89"    "9"     "9"
## [16,] "EL HRAIRIA" "204"    "200"   "211"   "19"    "19"
## [17,] "SIDI HESSINE" "205"    "203"   "195"   "18"    "18"
## [18,] "EL OUARDIA" "72"     "72"    "71"    "7"     "7"
## [19,] "EL KABARIA" "164"    "154"   "163"   "18"    "18"
## [20,] "SIDI EL BECHIR" "58"     "57"    "58"    "6"     "6"
## [21,] "JEBEL JELLOUD" "50"     "50"    "49"    "6"     "6"
## [22,] "LA GOULETTE" "62"     "61"    "60"    "6"     "6"
## [23,] "EL KRAM" "72"     "73"    "71"    "8"     "8"
## [24,] "LA MARSA" "131"    "128"   "123"   "12"    "12"
## [25,] "TOTAL" "1955"   "1942"  "1913"  "185"   "185"
##      [,7]      [,8]
## [1,] "مدارس" ""
## [2,] ""      "المعتمدية"
## [3,] "2013" ""
## [4,] "7"      "قرطاج"
## [5,] "6"      "تونس المدينة"
## [6,] "8"      "باب بحر"
## [7,] "4"      "باب سويقة"
## [8,] "10"     "العمران"
## [9,] "9"      "العمران الأعلى"
## [10,] "4"     "التحرير"
## [11,] "5"     "المنزه"
## [12,] "5"     "حي الحضراء"
```

```
## [13,] "13" "باردو"  
## [14,] "3" "السيجومي"  
## [15,] "9" "الزهور"  
## [16,] "19" "الحريرية"  
## [17,] "18" "سيدي حسين"  
## [18,] "7" "الوردية"  
## [19,] "18" "الكيارية"  
## [20,] "6" "سيدي البشير"  
## [21,] "6" "جبل الجلود"  
## [22,] "6" "حلق الوادي"  
## [23,] "8" "الكرم"  
## [24,] "11" "المرسى"  
## [25,] "182" "المجموع"
```

Cleaning the data (I)

```
> tab2[[1]][1,]
```

```
## [1] "" "salles d'études" "" "قاعات"
## [5] "écoles" "" "مدارس" ""
```

```
> tab2[[1]][2,]
```

```
## [1] "DELEGATION" "" "" "" ""
## [6] "" "" "المعمدية" ""
```

```
> tab2[[1]][3,]
```

```
## [1] "" "2015" "2014" "2013" "2015" "2014" "2013" ""
```

```
> ds1=tab2[[1]][,c(1,2,3,4,8)] # data classrooms
> ds2=tab2[[1]][,c(1,5,6,7,8)] # data school
> colnames(ds1)=colnames(ds2)=c(ds1[2,1],ds1[3,2:4],ds1[2,5])
> ds1=ds1[-c(1:3,25),]
> ds2=ds2[-c(1:3,25),]
```

Cleaning the data (2)

```
> ds1[,2]=as.numeric(as.character(ds1[,2]))
> ds1[,3]=as.numeric(as.character(ds1[,3]))
> ds1[,4]=as.numeric(as.character(ds1[,4]))
> ds1=as.data.frame(ds1)
```

and

```
> ds2[,2]=as.numeric(as.character(ds2[,2]))
> ds2[,3]=as.numeric(as.character(ds2[,3]))
> ds2[,4]=as.numeric(as.character(ds2[,4]))
> ds2=as.data.frame(ds2)
```


Cleaning the data (3)

```
> ds1A=melt(ds1,id.vars = colnames(ds1)[c(1,5)])  
> colnames(ds1A)[3:4]=c("Year","Classrooms")  
> ds2A=melt(ds2,id.vars = colnames(ds2)[c(1,5)])  
> colnames(ds2A)[3:4]=c("Year","Schools")
```

One data for all variables

```
> data_all=cbind.data.frame(dtA,ds1A$Classrooms,ds2A$Schools)  
> colnames(data_all)[c(5,6)]=c("Classrooms","Schools")
```

Displaying the whole data

```
> DT::datatable(data_all)
```

Show entries

Search:

	DELEGATION	المعتدية	Year	Population	Classrooms	Schools
1	Carthage	قرطاج	2015	24906	76	8
2	Tunis Medina	المدينة	2015	22009	67	6
3	Bab Bhar	باب بحر	2015	37241	97	8
4	Bab Souika	باب سويقة	2015	30016	46	4
5	El Omran	العمران	2015	43410	114	10
6	Omrane Superieur	العمران الأعلى	2015	57094	108	9
7	Tahrir	التحرير	2015	22327	45	4
8	Menzah	المنزه	2015	43021	58	6
9	Cité Khadra	حي الخضراء	2015	36175	64	5
10	Le Bardo	باردو	2015	74010	138	13

Showing 1 to 10 of 63 entries

Previous

1

2

3

4

5

6

7

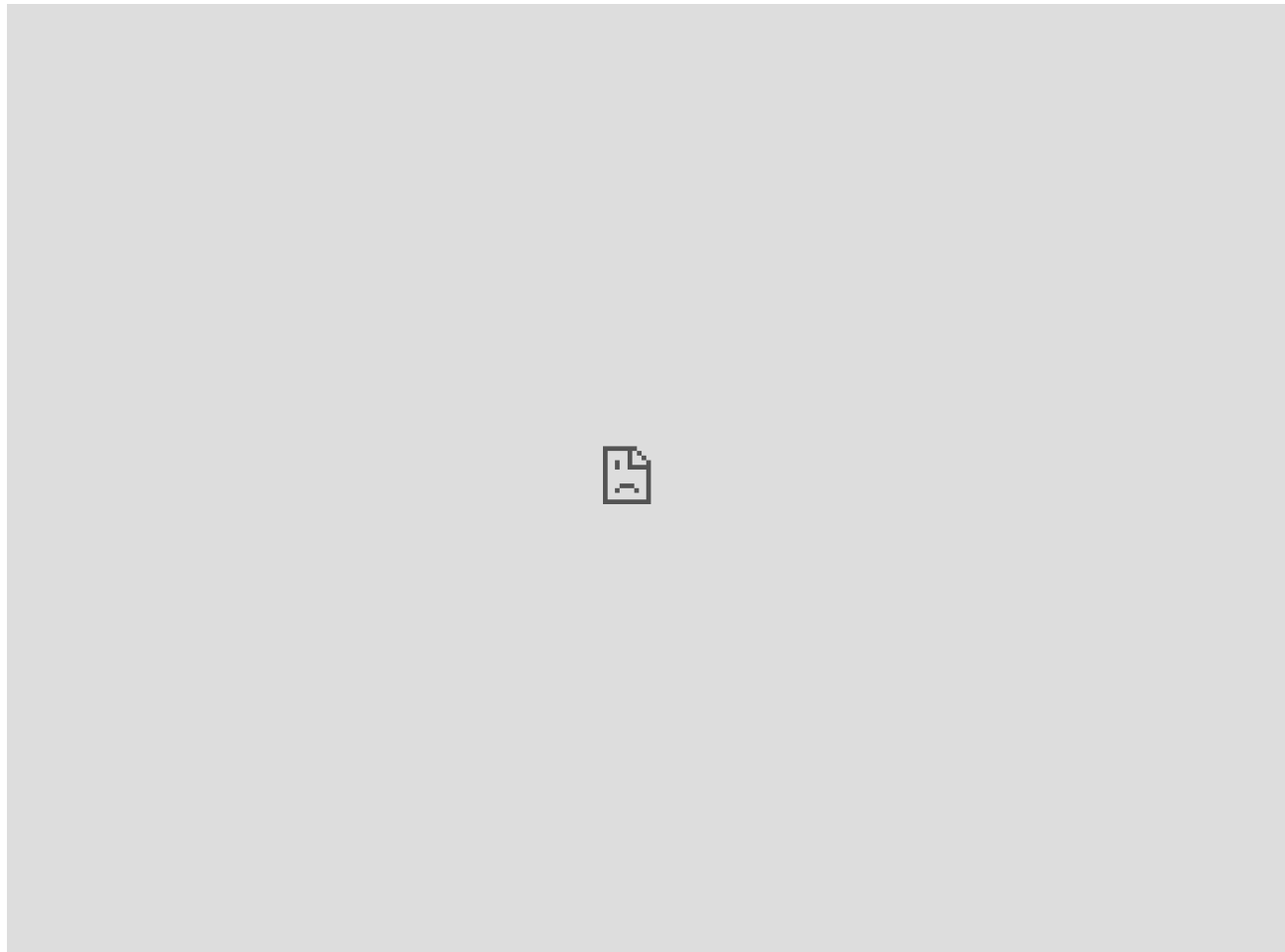
Next

Bubble chart with googleVis

```
> apply(data_all,2,class)
## DELEGATION      المعمدية      Year Population Classrooms Schools
## "character" "character" "character" "character" "character" "character"
> data_all$Year=as.numeric(as.character(data_all$Year))
> data_all$Population=as.numeric(as.character(data_all$Population))
> data_all$Classrooms=as.numeric(as.character(data_all$Classrooms))
```

```
> data_all$Schools=as.numeric(as.character(data_all$Schools))
> op <- options(gvis.plot.tag="chart")
> Motion=gvisMotionChart(data_all[,-2],
+                         idvar="DELEGATION",
+                         timevar="Year")
> print(Motion, file="Motion.html")
> cat(Motion$html$chart, file = "Motion.html")
```

Bubble chart with googleVis

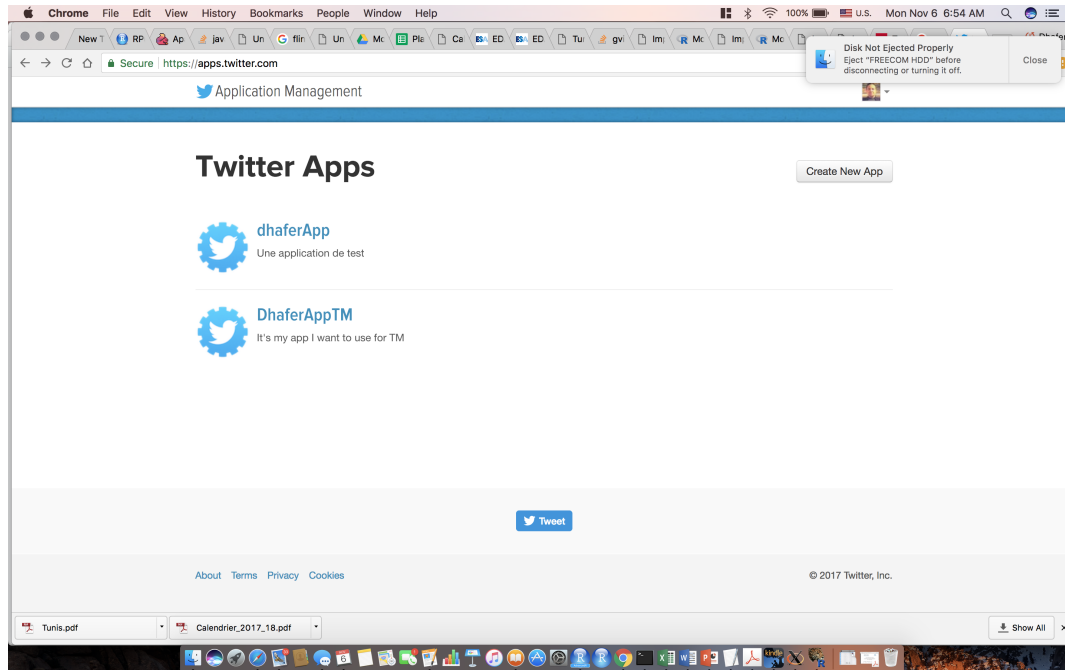


Social Network

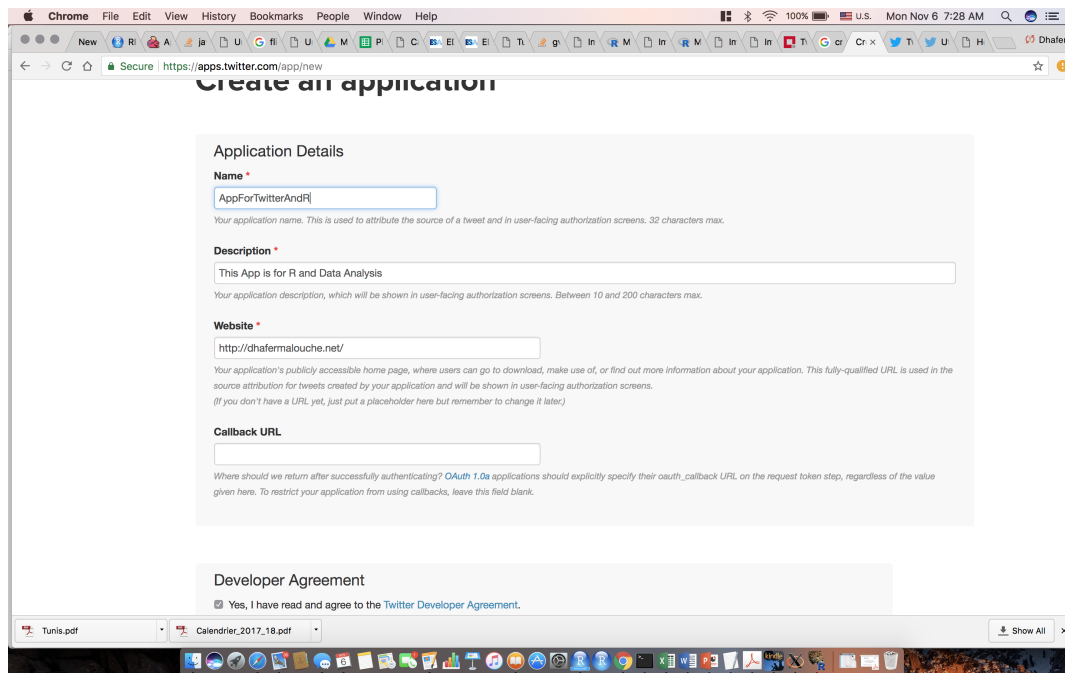
Data from twitter

An App on twitter

Step I: create an app on your account twitter: goto <https://apps.twitter.com/app/>



Setting up the App.



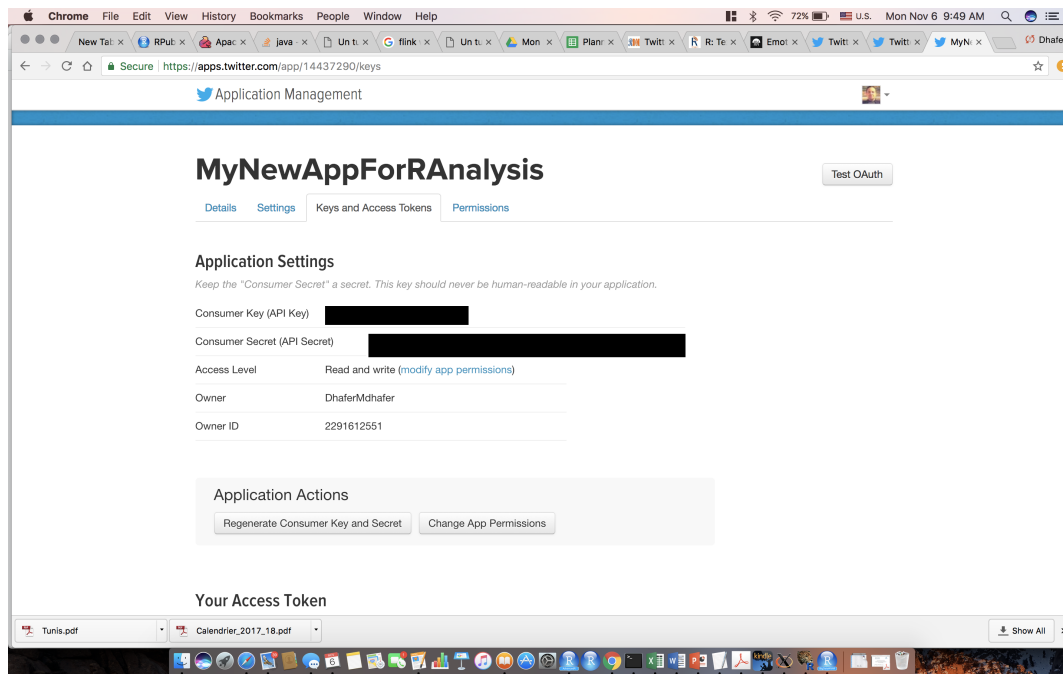
The screenshot shows a web browser window at the URL `https://apps.twitter.com/app/new`. The page title is "Create an application". The main content area is titled "Application Details" and contains the following fields:

- Name ***: A text input field containing "AppForTwitterAndR". Below it is a small note: "Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max."
- Description ***: A text input field containing "This App is for R and Data Analysis". Below it is a small note: "Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max."
- Website ***: A text input field containing "http://dhafermalouche.net/". Below it is a small note: "Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)"
- Callback URL**: An empty text input field. Below it is a small note: "Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank."

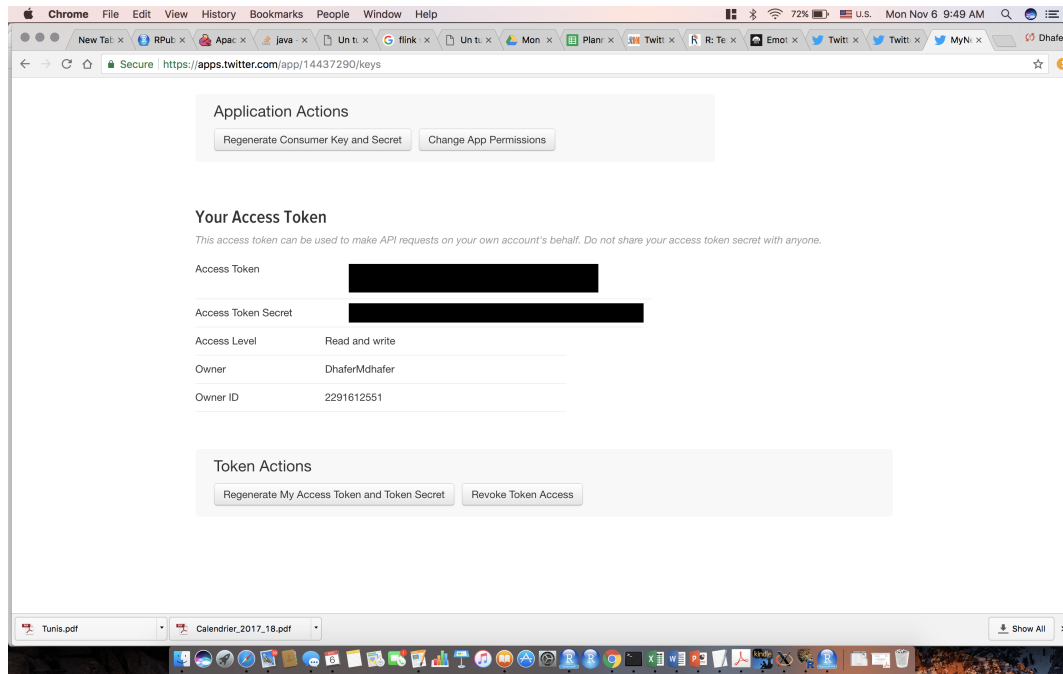
Below the "Application Details" section is a "Developer Agreement" section with a checkbox that is checked and the text "Yes, I have read and agree to the [Twitter Developer Agreement](#)."

The browser's address bar shows "Secure | https://apps.twitter.com/app/new". The top of the browser window shows the Chrome menu, tabs, and system tray information (100% battery, U.S., Mon Nov 6 7:28 AM). The bottom of the browser window shows a taskbar with various application icons and a "Show All" button.

API key secret



Access token



Installing R packages

```
> devtools::install_github("jrowen/twitterR", ref = "oauth_httr_1_0", force=TRUE)
> install.packages('base64enc')
> install.packages('ROAuth')
> library(twitterR)
> library(ROAuth)
> library(base64enc)
```

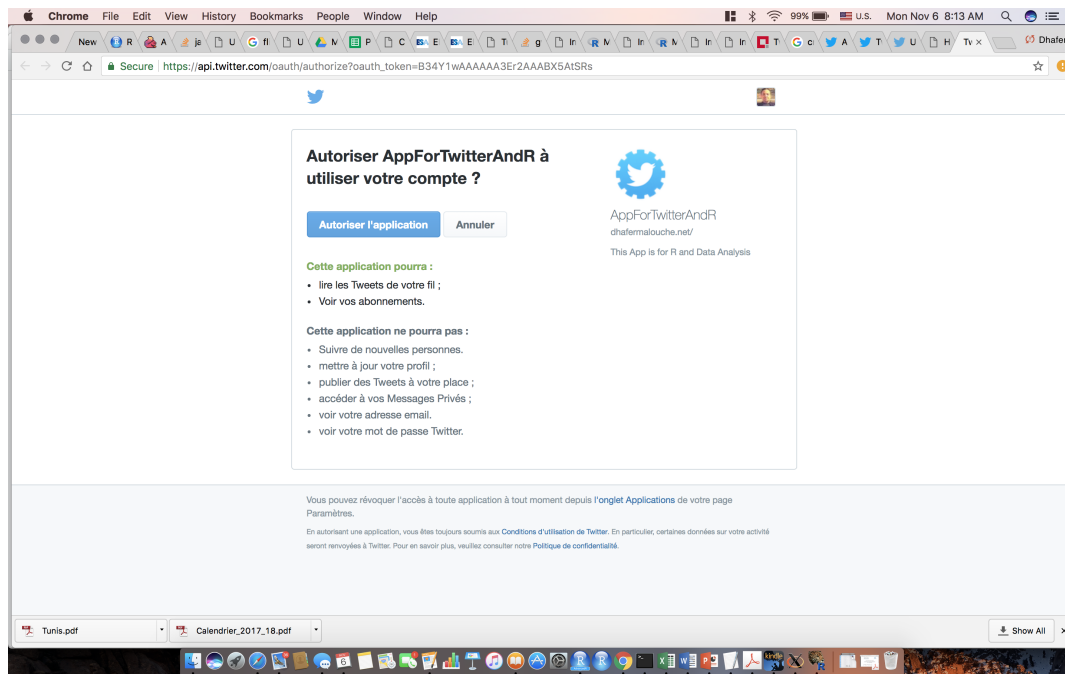
Importing twitter setting into R

```
> api_key = "xxxxxxx" # your api_key
> api_secret = "xxxxxxx" # your api_secret
> access_token = "xxxxx" # your access_token
> access_token_secret = "xxxxxxx" # your access_token_sceret
> credential<-OAuthFactory$new(
+   consumerKey=api_key,
+   consumerSecret=api_secret,requestURL="https://api.twitter.com/oauth/request_token",accessURL="https://api.twitter.com/oauth/access_token",
+   authURL="https://api.twitter.com/oauth/authorize")
```

Getting the authorization from R

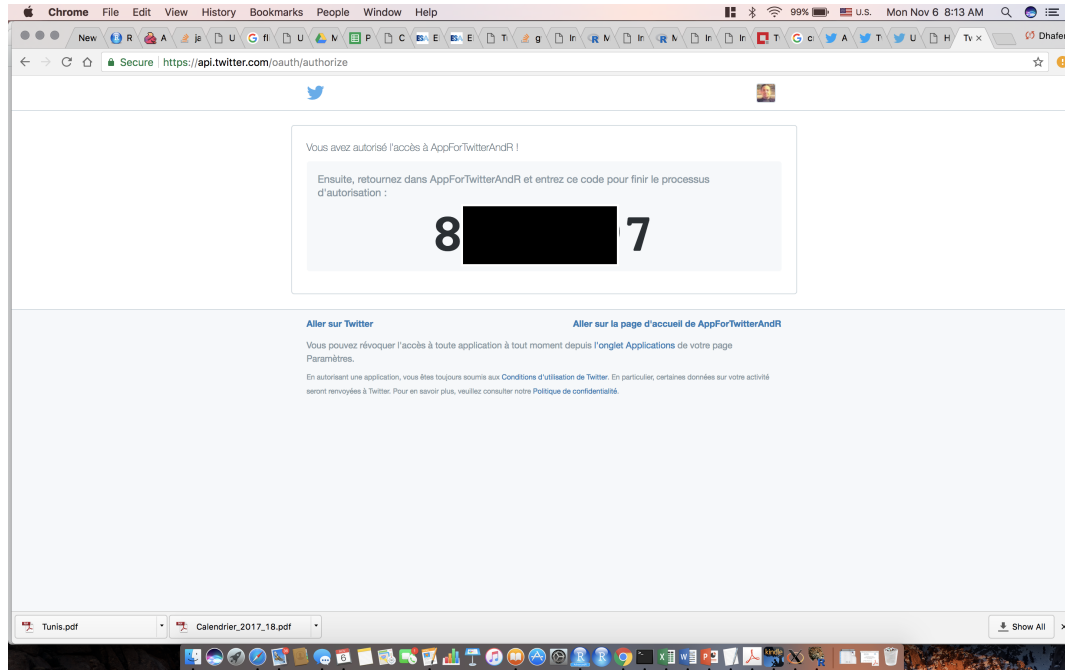
```
> credential$handshake()  
To enable the connection, please direct your web browser to:  
https://api.twitter.com/oauth/authorize?oauth\_token=XXXXXXXX  
When complete, record the PIN given to you and provide it here:
```

Getting the authorization from R



Getting the authorization from R

Copy this pin number and paste it in R



Setting the authorization

```
> setup_twitter_oauth(consumer_key = api_key, consumer_secret = api_secret,  
+                    access_token = access_token, access_secret = access_token_secret)  
[1] "Using direct authentication"
```


Extracting data on tweets.

```
> metoo=searchTwitter("metoo", n=2000, lang="en", since='2017-10-01')  
> df_metoo <- do.call("rbind", lapply(metoo, as.data.frame))
```

```
> colnames(df_metoo)
```

```
## [1] "text"          "favorited"    "favoriteCount" "replyToSN"  
## [5] "created"      "truncated"    "replyToSID"    "id"  
## [9] "replyToUID"   "statusSource" "screenName"    "retweetCount"  
## [13] "isRetweet"    "retweeted"    "longitude"     "latitude"
```

Importing data from Facebook

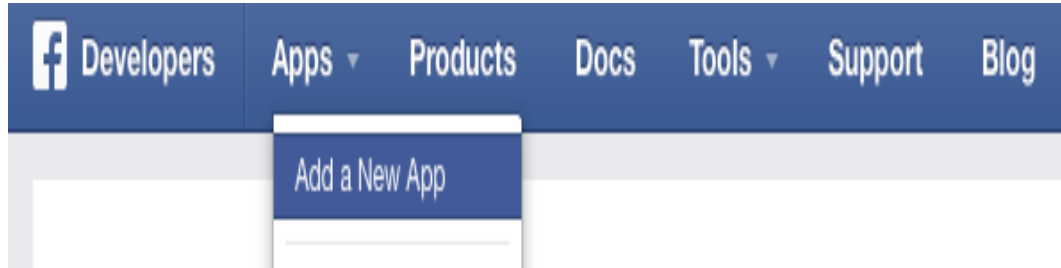
Installing R packages

Using `Rfacebook` package

```
> install.packages("Rfacebook")  
> install.packages("httpuv")  
> library(Rfacebook)  
> library(httpuv)
```

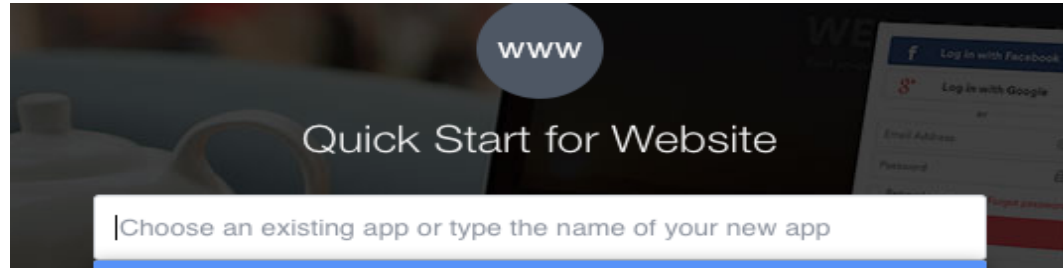
Creating the App on Facebook

Step I: Go to the link <https://developers.facebook.com>



Creating the App on Facebook

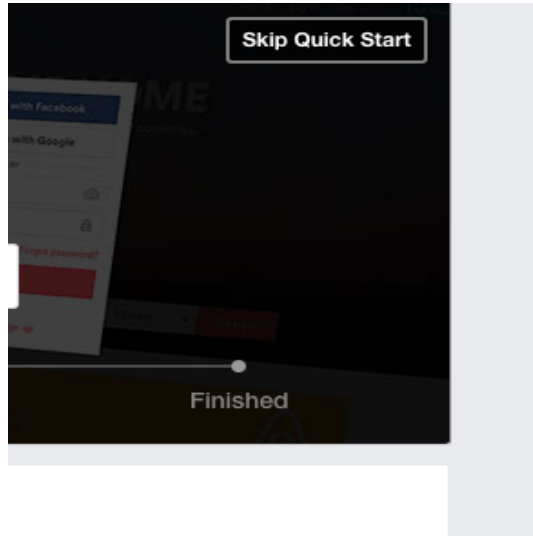
Step2: Give a name to your new App.



Click on “Create a New App ID” and choose a category for your app.

Creating the App on Facebook

Step3: Skip Quick Start and go to the setting of the app.



Setting up the App

The screenshot shows the Facebook Developer Dashboard for an application named "myRapp". The browser is Firefox, and the URL is <https://developers.facebook.com/apps/456791051149784/dashboard/>. The page title is "myRapp" and the app ID is "456791051149784".

The dashboard is in French and features a sidebar on the left with navigation options: "Tableau de bord", "Paramètres", "Rôles", "Alertes", "Examen des apps", and "PRODUITS".

The main content area is titled "Tableau de bord" and includes the following sections:

- myRapp**: A section showing the app's status. It includes a text description: "Cette app est en mode développement et seuls les administrateurs, développeurs et testeurs peuvent l'utiliser. (?)". It displays the "Version de l'API" as "v2.3" and the "ID de l'app" as a redacted value. A "Clé secrète" (secret key) is also shown as redacted, with a "Réinitialiser" (reset) button next to it.
- Démarrer avec le SDK Facebook**: A section with the text "Utilisez nos guides de démarrage rapide pour configurer le SDK Facebook pour votre application, jeu canevase ou site Internet iOS ou Android." and a "Choisir une plate-forme" (choose a platform) button.
- Facebook Analytics pour apps**: A section titled "Set up Analytics" with the text "Analytics for Apps helps you grow your business and learn about the actions people take in your app. It only takes 5 minutes to set up." and buttons for "Try Demo" and "View Quickstart Guide".

The footer of the page reads "facebook for developers".

Getting the authorization

```
> app_id="XXXX"  
> app_secret="XXXX"  
> fb_oauth=fbOAuth(app_id=app_id, app_secret=app_secret, extended_permissions = TRUE)  
> save(fb_oauth, file="fb_oauth")
```


Getting data from Donal Trump official page

<https://www.facebook.com/DonaldTrump/>

```
> fb_page <- getPage(page="DonaldTrump", since='2017-11-01', token=fb_oauth)
```

```
> colnames(fb_page)
```

```
## [1] "from_id"      "from_name"    "message"      "created_time"  
## [5] "type"        "link"         "id"           "story"  
## [9] "likes_count" "comments_count" "shares_count"
```

Getting the data on the posts in DT's page

```
> post <- getPost(post=fb_page$id[1], token=fb_oauth)
```

```
names(post)
```

```
## [1] "post"      "likes"     "comments"
```

Using SocialMediaLab package

```
> if (!"SocialMediaLab" %in% installed.packages()) {  
+   devtools::install_github("voson-lab/SocialMediaLab/SocialMediaLab")  
> library(SocialMediaLab)  
> require(magrittr)
```

Getting the data

```
> fb_page2<-Authenticate("Facebook",appID = app_id, appSecret = app_secret) %>%SaveCredential("FBcredential.RDS") %>%Collect(pageName="DonaldTrump", rangeFrom="2017-  
+ rangeTo="2017-11-5", writeToFile=TRUE)
```

Getting the data

- The data is writing in a csv file

```
> file.exists("2017-11-4_to_2017-11-5_DonaldTrump_FacebookData.csv")
```

```
## [1] TRUE
```

- Importing it again into R

```
> dt<- read.csv("~/Documents/Teaching/AdvancedR/2017-11-4_to_2017-11-5_DonaldTrump_FacebookData.csv", comment.char="#")  
> colnames(dt)
```

```
## [1] "X"           "from_username" "from"  
## [4] "to"         "edgeType"     "postType"  
## [7] "postLink"   "postTimestamp" "commentText"  
## [10] "commentTimestamp"
```