INSTITUTE
OF LAW,
POLITICS AND
DEVELOPMENT

Sant'Anna
School of Advanced Studies – Pisa

Intro to R

Getting set up

Prof. Kevin Koehler
kevin.koehler@santannapisa.it

The class

Aims:

1. Introduce you to R as a programming language, statistical package, and reporting tool (using the RStudio user interface)

The class

Aims:

1. Introduce you to R as a programming language, statistical package, and reporting tool (using the RStudio user interface)
2. Implement basic data analysis tasks in R, including data management, visualization, and regression analysis

The class

Aims:

1. Introduce you to R as a programming language, statistical package, and reporting tool (using the RStudio user interface)
2. Implement basic data analysis tasks in R, including data management, visualization, and regression analysis
3. Glimpse into Web Scraping and Text-as-data

Schedule

1. Getting set up, 8 April, 10-12h
2. Visualization and data management, TBD
3. Data exploration and reporting with R markdown, TBD
4. Replication I: U.S. Soft Power (OLS), 29 April, 10-13h
5. Replication II: Does Counterbalancing Prevent Military Coups? (Logistic), 5 May, 10-13h
6. Basic Web-Scraping with R, 12 May, 10-13h
7. Text as Data, 15 May, 10-13h

Today's class

1. RStudio

Today's class

1. RStudio
2. Basic notions in R: Objects, functions, packages

Today's class

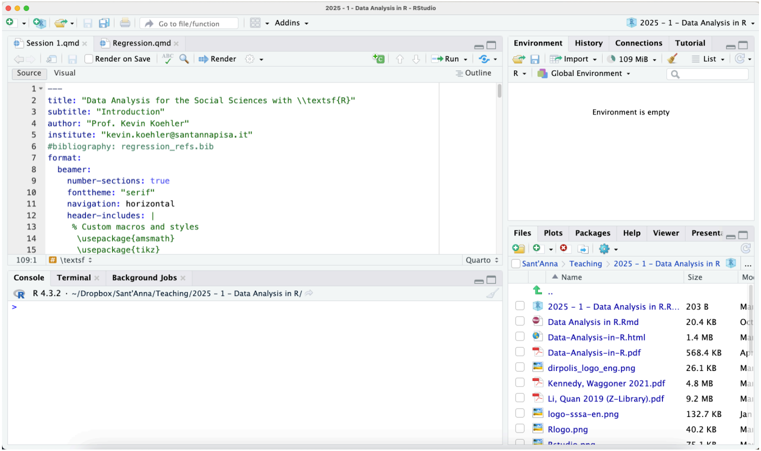
1. RStudio
2. Basic notions in R: Objects, functions, packages
3. Tidy data and data management

Today's class

1. RStudio
2. Basic notions in R: Objects, functions, packages
3. Tidy data and data management
4. Exercises

RStudio

RStudio



```
---
title: "Data Analysis for the Social Sciences with \\textsf{R}"
subtitle: "Introduction"
author: "Prof. Kevin Koehler"
institute: "kevin.koehler@santannapisa.it"
#bibliography: regression_refs.bib
format:
  beamer:
    number-sections: true
    fonttheme: "serif"
    navigation: horizontal
    header-includes: |
     % Custom macros and styles
      \usepackage{amsmath}
      \usepackage{tikz}
```

```
R 4.3.2 · ~/Dropbox/Sant'Anna/Teaching/2025 – 1 – Data Analysis in R/
>
```

Name	Size	Mo
..		
2025 – 1 – Data Analysis in R.R…	203 B	Ma
Data Analysis in R.Rmd	20.4 KB	Oc
Data-Analysis-in-R.html	1.4 MB	Ma
Data-Analysis-in-R.pdf	568.4 KB	Ap
dirpolis_logo_eng.png	26.1 KB	Ma
Kennedy, Waggoner 2021.pdf	4.8 MB	Ma
Li, Quan 2019 (Z-Library).pdf	9.2 MB	Ma
logo-sssa-en.png	132.7 KB	Ja
Rlogo.png	40.2 KB	Ma
Rstudio.png	75.1 KB	M

RStudio

First, create a **project**. Projects are useful for keeping all files in the same place.

1. Create a directory on your computer where you want to save all files related to this class
2. Go to `> File > New Project` in the R menu and then select "Existing directory"
3. Navigate to the folder and name and create the project

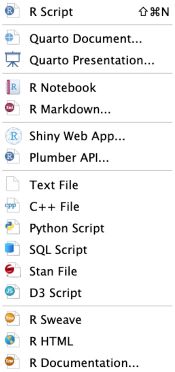
RStudio

The folder you created is now also your **working directory**. You can see the directory at the top of the console. You can also type:

```
getwd()
```

Basic notions in R

Types of R files



R can do many different things, not just statistical analysis.

Consequently, there are many different file types in RStudio:

- R scripts for coding
- Quarto documents and presentations
- R Notebook and R Markdown
- Shiny Apps
- Plumber API
- Files in other languages (C++, Python, SQL...)

R scripts

- Go ahead and open a new R script.
- Type `print("Hello world")`
- With your cursor in the line with the command, press `Ctrl + Enter` (Windows) or `Command + Enter` (Mac)
- The code is executed and the results printed in the Console

objects

Vectors

Vectors are the most basic data structure. They hold a series of numeric or character values and are created with the `c()` function (the c in the function stands for concatenate):

```
c(1,2,3,4,5)
```

```
[1] 1 2 3 4 5
```

```
c("a","b","c","d","e")
```

```
[1] "a" "b" "c" "d" "e"
```

Matrices and data frames

```r
matrix(1:6, nrow = 2, ncol = 3)
```

```
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6
```

```r
data.frame(name = c("A","B"),
           age = c(24,56))
```

```
  name age
1    A  24
2    B  56
```

The environment

The **Environment** is where R stores objects for the duration of a session. You can assign an object to the environment by typing:

```
data <- data.frame(name = c("A","B"),
                   age = c(24,56))
```

(you could replace the `<-` with `=`, but I recommend getting used to `<-`)

After running this code, you should have an `object` called “data” in your environment. This object contains two columns (name and age) with two rows each. We call the columns **variables** and the rows **observations**.

You can click on the object to see what it contains.

Working with data frames

```r
data
```

```
  name age
1    A  24
2    B  56
```

```r
data$name
```

```
[1] "A" "B"
```

```r
data$age[1]
```

```
[1] 24
```

```r
data$age[data$name=="A"]
```

```
[1] 24
```

Functions

R works with `functions`. A function takes an `object` (or multiple `objects`) as input and does something with it.

Examples include:

- `print("Hello world")` prints "Hello world" to the console
- `c(1,2,3,4)` creates a numerical vector with 1,2,3,4 as elements
- `getwd()` returns the active working directory
- `help(print)` returns the help file for the `print()` function
- `lm(x~y)` performs a linear regression of x on y

Functions in R are words followed by brackets. You always need to close the brackets, otherwise your code will not run.

User-defined functions

You can write your own functions in R. Here is a function which takes a number as an argument and tells you whether the number is greater than 5:

```r
greater5 <- function(x) {
  if (!is.numeric(x)) {
    stop(paste0("Argument must be numeric.\n",
                "You provided an object of class: ",
                class(x)[1],
                ". You moron."))
  } # check if input is numeric, return error if not
  result <- ifelse(x > 5,
                   paste(x, "is greater than 5"),
                   paste(x, "is not greater than 5"))
  return(result)  # Return results
}
```

Packages and CRAN

Functions are part of packages. Your version of R comes with base R, but there are many other packages.

We will use the `tidyverse` family of packages. You can install packages by typing `install.packages("tidyverse")` in the Console. This will download the package and save it on your machine. You need to do this only once.

To use specific packages, you need to load them in the beginning of your R session. It is good practice to include all packages needed to run your code in the beginning of your R script. Packages are loaded typing `library(tidyverse)`.

Some statistical notions

Recap on variable types

1. **Nominal Scale**:
   - Categories without a specific order (e.g., gender, color).
2. **Ordinal Scale**:
   - Categories with a defined order but unequal intervals (e.g., rankings, satisfaction ratings).
3. **Interval Scale**:
   - Numeric scales with equal intervals but no true zero (e.g., temperature in Celsius).
4. **Ratio Scale**:
   - Numeric scales with equal intervals and a true zero (e.g., height, weight, age).

How can we describe the typical value for each of these scales?

Measures of Central Tendency

- **Mean**: The average of a data set, calculated by summing all values and dividing by the number of values.

Measures of Central Tendency

- **Mean**: The average of a data set, calculated by summing all values and dividing by the number of values.
    - Mean $= \frac{\sum_{i=1}^{n} x_i}{n}$, where x_i represents each data point, and n is the number of data points.

Measures of Central Tendency

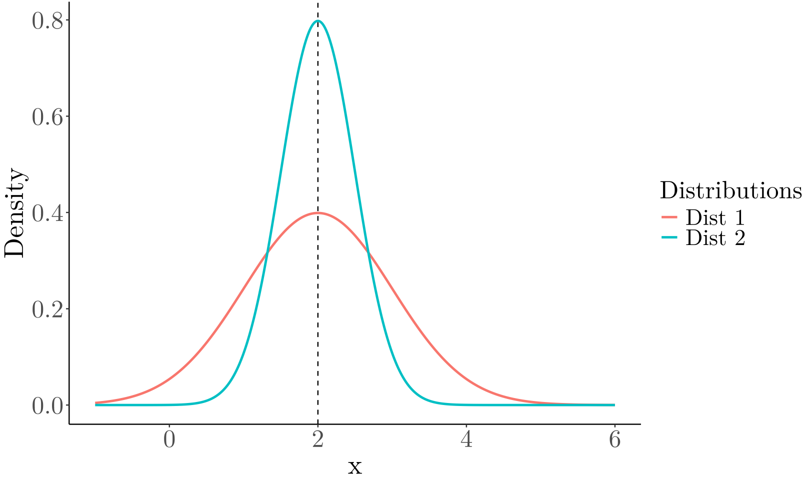
- **Mean**: The average of a data set, calculated by summing all values and dividing by the number of values.
  - Mean $= \frac{\sum_{i=1}^{n} x_i}{n}$, where x_i represents each data point, and n is the number of data points.
- **Median**: The middle value when data is ordered from lowest to highest; 50% of values fall below it.

Measures of Central Tendency

- **Mean**: The average of a data set, calculated by summing all values and dividing by the number of values.
  - Mean $= \frac{\sum_{i=1}^{n} x_i}{n}$, where x_i represents each data point, and n is the number of data points.
- **Median**: The middle value when data is ordered from lowest to highest; 50% of values fall below it.
- **Mode**: The value that occurs most frequently in a data set.

How can we describe variation around a central value?

How can we describe variation around a central value?



Measures of Dispersion

- **Range**: $\text{Range} = \text{Max}(x) - \text{Min}(x)$. The difference between the maximum and minimum values in the data set.

Measures of Dispersion

- **Range**: $\text{Range} = \text{Max}(x) - \text{Min}(x)$. The difference between the maximum and minimum values in the data set.
- **Variance**: $\text{Variance} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$, where x_i represents each data point, μ is the mean, and n is the number of data points.

Measures of Dispersion

- **Range**: $\text{Range} = \text{Max}(x) - \text{Min}(x)$. The difference between the maximum and minimum values in the data set.
- **Variance**: $\text{Variance} = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$, where x_i represents each data point, μ is the mean, and n is the number of data points.
- **Standard Deviation**: $\text{SD} = \sqrt{\text{Variance}}$. The square root of the variance, which gives the spread of data in the same units as the original data.

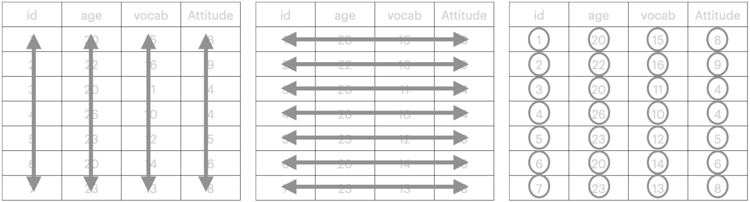
Measures of Dispersion

- **Range**: $\text{Range} = \text{Max}(x) - \text{Min}(x)$. The difference between the maximum and minimum values in the data set.
- **Variance**: $\text{Variance} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$, where x_i represents each data point, μ is the mean, and n is the number of data points.
- **Standard Deviation**: $\text{SD} = \sqrt{\text{Variance}}$. The square root of the variance, which gives the spread of data in the same units as the original data.
- **Interquartile Range (IQR)**: $\text{IQR} = Q3 - Q1$, where $Q3$ is the third quartile and $Q1$ is the first quartile, representing the middle 50% of the data.

What is tidy data

Tidy data

Observations in rows



id	age	vocab	Attitude
1	20	15	8
2	22	16	9
3	20	11	4
4	26	10	4
5	23	12	5
6	20	14	6
7	23	13	8

Variables in columns

Values in cells

Messy data

Party	2014 Vote %	2019 Vote %
Nidaa Tounes	37.56%	1.51%
Ennahda	27.79%	19.63%

Table 1: Vote Percentages for Ennahda and Nidaa Tounes (2014 vs. 2019)

Tidy data

party	**year**	**vote**
Nidaa Tounes	2014	37.56
Nidaa Tounes	2019	1.51
Ennahda	2014	27.79
Ennahda	2019	19.63

Table 2: Vote Percentages for Ennahda and Nidaa Tounes (2014 vs. 2019)

The `tidyverse`

The `tidyverse` is a family of functions for data management. They follow a specific structure:

```
new_data <- data %>%
  function()  %>%
  function()
```

The `%>%` is called a *pipe*. It allows you to stack functions on top of each other.

Data management

Data management functions

The most important data management functions are:

1. Adding or removing variables or observations

```
mutate() # to create new variables
select() # to select specific variables
filter() # to filter for values
```

2. Logical operations

```
ifelse() # ifelse(test,yes,no)
case_when() # series of ifelse statements
```

3. Grouping and summarizing

```
group_by() # to group a data set
summarize() # to summarize data
```

Data management functions

For example, returning to the Tunisia survey, look at

```
tun22 <- read_csv("tunisia_survey.csv")
table(tun22$pres2019_2)
```

```
  1   2  97  98  99 
434  32   8   9  25 
```

From the codebook, we know that 1 stands for Kais Saied and 2 for Nabil Karoui, the two candidates in the run-off round. 97 stands for blank or invalid votes, 98 for DK/don't remember, and 99 for declined to answer.

Wouldn't it be nice to see names instead of numbers?

Data management functions

```r
tun22 <- tun22 %>%
  mutate(pres2019_2_new = case_when(
    pres2019_2==1~"Kais Saied",
    pres2019_2==2~"Nabil Karoui",
    pres2019_2==97~"blank/invalid",
    pres2019_2==98~"don't know",
    pres2019_2==99~"declined to answer"
  ))

tun22 %>% tabyl(pres2019_2_new)
```

New variable with names instead of numbers

```
     pres2019_2_new   n percent valid_percent
         Kais Saied 434   0.434    0.85433071
       Nabil Karoui  32   0.032    0.0629921
      blank/invalid   8   0.008    0.0157480
 declined to answer  25   0.025    0.04921260
         don't know   9   0.009    0.01771654
               <NA> 492   0.492            NA
```

Output with janitor package

Data management functions

Next, assume we want to know the average value of Kais Saied voters on question:

To what extent do you agree with the following statement: “Members of Parliament very quickly lose touch with ordinary people after they assume office.”

From the codebook, we know that this variable is called `mps`.

Data management functions

1. Filtering

```
tun22 %>%
  select(pres2019_2_new, mps) %>%
  filter(pres2019_2_new=="Kais Saied") %>%
  summarize(mps=mean(mps, na.rm = T))
```

```
# A tibble: 1 x 1
    mps
  <dbl>
1  2.38
```

2. Grouping

```
tun22 %>%
  select(pres2019_2_new, mps) %>%
  group_by(pres2019_2_new) %>%
  summarize(mps=mean(mps, na.rm = T))
```

```
# A tibble: 6 x 2
  pres2019_2_new        mps
  <chr>               <dbl>
1 Kais Saied           2.38
2 Nabil Karoui         2.44
3 blank/invalid        1.62
4 declined to answer   2.04
5 don't know           1.89
6 <NA>                 2.77
```

Sant'Anna
School of Advanced Studies – Pisa

Data management functions

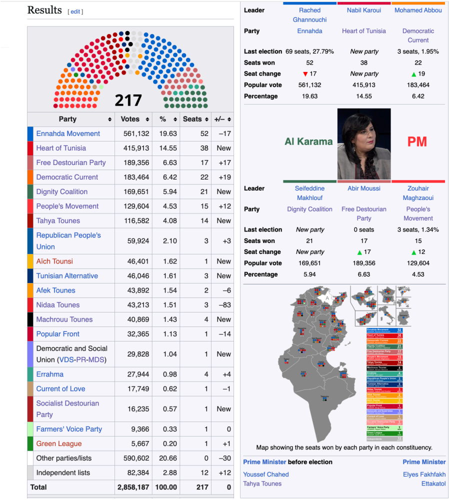
4. Joining data sets

```
inner_join() # joining data sets keeping common obs
left_join() # joining data sets keeping all left-hand obs
right_join() # joining data sets keeping all right-hand obs
full_join() # joining data sets keeping all obs
```

5. Reshaping data sets

```
pivot_longer() # reshape data into long format
pivot_wider() # reshape data into wide format
```

Data management functions



Results [edit]

217

Party	Votes	%	Seats	+/–
Ennahda Movement	561,132	19.63	52	−17
Heart of Tunisia	415,913	14.55	38	New
Free Destourian Party	189,356	6.63	17	+17
Democratic Current	183,464	6.42	22	+19
Dignity Coalition	169,651	5.94	21	New
People's Movement	129,604	4.53	15	+12
Tahya Tounes	116,582	4.08	14	New
Republican People's Union	59,924	2.10	3	+3
Aïch Tounsi	46,401	1.62	1	New
Tunisian Alternative	46,046	1.61	3	New
Afek Tounes	43,892	1.54	2	−6
Nidaa Tounes	43,213	1.51	3	−83
Machrouu Tounes	40,869	1.43	4	New
Popular Front	32,365	1.13	1	−14
Democratic and Social Union (VDS-PR-MDS)	29,828	1.04	1	New
Errahma	27,944	0.98	4	+4
Current of Love	17,749	0.62	1	−1
Socialist Destourian Party	16,235	0.57	1	New
Farmers' Voice Party	9,366	0.33	1	0
Green League	5,667	0.20	1	+1
Other parties/lists	590,602	20.66	0	−30
Independent lists	82,384	2.88	12	+12
Total	2,858,187	100.00	217	0

Leader	Rached Ghannouchi	Nabil Karoui	Mohamed Abbou
Party	Ennahda	Heart of Tunisia	Democratic Current
Last election	69 seats, 27.79%	*New party*	3 seats, 1.95%
Seats won	52	38	22
Seat change	▼ 17	*New party*	▲ 19
Popular vote	561,132	415,913	183,464
Percentage	19.63	14.55	6.42

Al Karama

PM

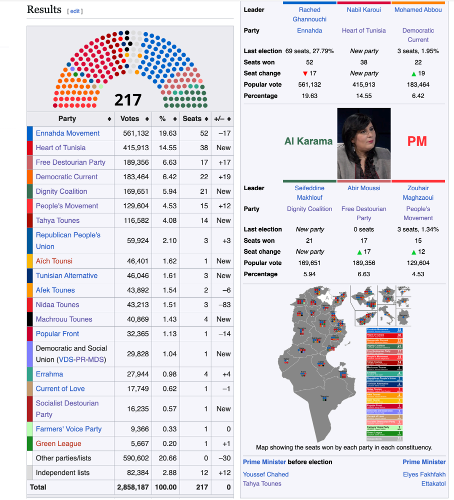
Leader	Seifeddine Makhlouf	Abir Moussi	Zouhair Maghzaoui
Party	Dignity Coalition	Free Destourian Party	People's Movement
Last election	*New party*	0 seats	3 seats, 1.34%
Seats won	21	17	15
Seat change	*New party*	▲ 17	▲ 12
Popular vote	169,651	189,356	129,604
Percentage	5.94	6.63	4.53

Map showing the seats won by each party in each constituency.

Prime Minister before election	Prime Minister
Youssef Chahed Tahya Tounes	Elyes Fakhfakh Ettakatol

How do we get the data from Wikipedia into R?

Data management functions



Results [edit]

Party	Votes	%	Seats	+/–
Ennahda Movement	561,132	19.63	52	–17
Heart of Tunisia	415,913	14.55	38	New
Free Destourian Party	189,356	6.63	17	+17
Democratic Current	183,464	6.42	22	+19
Dignity Coalition	169,651	5.94	21	New
People's Movement	129,604	4.53	15	+12
Tahya Tounes	116,582	4.08	14	New
Republican People's Union	59,924	2.10	3	+3
Aïch Tounsi	46,401	1.62	1	New
Tunisian Alternative	46,046	1.61	3	New
Afek Tounes	43,892	1.54	2	–6
Nidaa Tounes	43,213	1.51	3	–83
Machrouu Tounes	40,869	1.43	4	New
Popular Front	32,365	1.13	1	–14
Democratic and Social Union (VDS-PR-MDS)	29,828	1.04	1	New
Errahma	27,944	0.98	4	+4
Current of Love	17,749	0.62	1	–1
Socialist Destourian Party	16,235	0.57	1	New
Farmers' Voice Party	9,366	0.33	1	0
Green League	5,667	0.20	1	+1
Other parties/lists	590,602	20.66	0	–30
Independent lists	82,384	2.88	12	+12
Total	**2,858,187**	**100.00**	**217**	**0**

Leader	Rached Ghannouchi	Nabil Karoui	Mohamed Abbou
Party	Ennahda	Heart of Tunisia	Democratic Current
Last election	69 seats, 27.79%	*New party*	3 seats, 1.95%
Seats won	52	38	22
Seat change	▼ 17	*New party*	▲ 19
Popular vote	561,132	415,913	183,464
Percentage	19.63	14.55	6.42

	Al Karama		PM
Leader	Seifeddine Makhlouf	Abir Moussi	Zouhair Maghzaoui
Party	Dignity Coalition	Free Destourian Party	People's Movement
Last election	*New party*	0 seats	3 seats, 1.34%
Seats won	21	17	15
Seat change	*New party*	▲ 17	▲ 12
Popular vote	169,651	189,356	129,604
Percentage	5.94	6.63	4.53

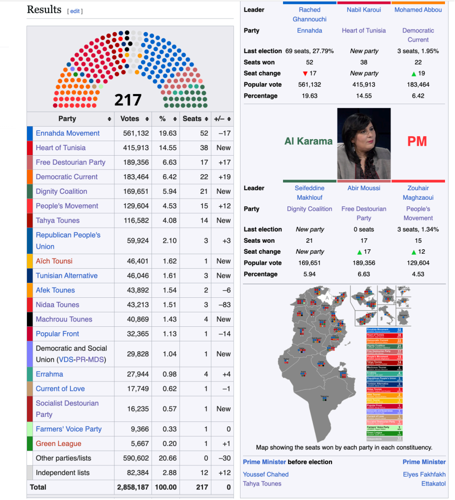
Map showing the seats won by each party in each constituency.

Prime Minister before election	**Prime Minister**
Youssef Chahed Tahya Tounes	Elyes Fakhfakh Ettakatol

We could:

1. manually copy the numbers

Data management functions



Results [edit]

217

Party	Votes	%	Seats	+/–
Ennahda Movement	561,132	19.63	52	–17
Heart of Tunisia	415,913	14.55	38	New
Free Destourian Party	189,356	6.63	17	+17
Democratic Current	183,464	6.42	22	+19
Dignity Coalition	169,651	5.94	21	New
People's Movement	129,604	4.53	15	+12
Tahya Tounes	116,582	4.08	14	New
Republican People's Union	59,924	2.10	3	+3
Aïch Tounsi	46,401	1.62	1	New
Tunisian Alternative	46,046	1.61	3	New
Afek Tounes	43,892	1.54	2	–6
Nidaa Tounes	43,213	1.51	3	–83
Machrouu Tounes	40,869	1.43	4	New
Popular Front	32,365	1.13	1	–14
Democratic and Social Union (VDS-PR-MDS)	29,828	1.04	1	New
Errahma	27,944	0.98	4	+4
Current of Love	17,749	0.62	1	–1
Socialist Destourian Party	16,235	0.57	1	New
Farmers' Voice Party	9,366	0.33	1	0
Green League	5,667	0.20	1	+1
Other parties/lists	590,602	20.66	0	–30
Independent lists	82,384	2.88	12	+12
Total	**2,858,187**	**100.00**	**217**	**0**

Leader	Rached Ghannouchi	Nabil Karoui	Mohamed Abbou
Party	Ennahda	Heart of Tunisia	Democratic Current
Last election	69 seats, 27.79%	*New party*	3 seats, 1.95%
Seats won	52	38	22
Seat change	▼ 17	*New party*	▲ 19
Popular vote	561,132	415,913	183,464
Percentage	19.63	14.55	6.42

	Al Karama		PM
Leader	Seifeddine Makhlouf	Abir Moussi	Zouhair Maghzaoui
Party	Dignity Coalition	Free Destourian Party	People's Movement
Last election	*New party*	0 seats	3 seats, 1.34%
Seats won	21	17	15
Seat change	*New party*	▲ 17	▲ 12
Popular vote	169,651	189,356	129,604
Percentage	5.94	6.63	4.53

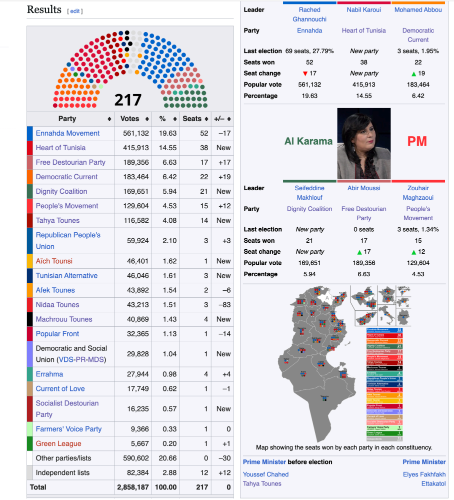
Map showing the seats won by each party in each constituency.

Prime Minister before election	**Prime Minister**
Youssef Chahed Tahya Tounes	Elyes Fakhfakh Ettakatol

We could:

1. manually copy the numbers
2. copy paste into Excel, save the file, read it into R

Data management functions



Results [edit]

Party	Votes	%	Seats	+/–
Ennahda Movement	561,132	19.63	52	–17
Heart of Tunisia	415,913	14.55	38	New
Free Destourian Party	189,356	6.63	17	+17
Democratic Current	183,464	6.42	22	+19
Dignity Coalition	169,651	5.94	21	New
People's Movement	129,604	4.53	15	+12
Tahya Tounes	116,582	4.08	14	New
Republican People's Union	59,924	2.10	3	+3
Aïch Tounsi	46,401	1.62	1	New
Tunisian Alternative	46,046	1.61	3	New
Afek Tounes	43,892	1.54	2	–6
Nidaa Tounes	43,213	1.51	3	–83
Machrouu Tounes	40,869	1.43	4	New
Popular Front	32,365	1.13	1	–14
Democratic and Social Union (VDS-PR-MDS)	29,828	1.04	1	New
Errahma	27,944	0.98	4	+4
Current of Love	17,749	0.62	1	–1
Socialist Destourian Party	16,235	0.57	1	New
Farmers' Voice Party	9,366	0.33	1	0
Green League	5,667	0.20	1	+1
Other parties/lists	590,602	20.66	0	–30
Independent lists	82,384	2.88	12	+12
Total	**2,858,187**	**100.00**	**217**	**0**

Leader	Rached Ghannouchi	Nabil Karoui	Mohamed Abbou
Party	Ennahda	Heart of Tunisia	Democratic Current
Last election	69 seats, 27.79%	*New party*	3 seats, 1.95%
Seats won	52	38	22
Seat change	▼ 17	*New party*	▲ 19
Popular vote	561,132	415,913	183,464
Percentage	19.63	14.55	6.42

Al Karama

PM

Leader	Seifeddine Makhlouf	Abir Moussi	Zouhair Maghzaoui
Party	Dignity Coalition	Free Destourian Party	People's Movement
Last election	*New party*	0 seats	3 seats, 1.34%
Seats won	21	17	15
Seat change	*New party*	▲ 17	▲ 12
Popular vote	169,651	189,356	129,604
Percentage	5.94	6.63	4.53

Map showing the seats won by each party in each constituency.

Prime Minister before election	**Prime Minister**
Youssef Chahed Tahya Tounes	Elyes Fakhfakh Ettakatol

We could:

1. manually copy the numbers
2. copy paste into Excel, save the file, read it into R
3. scrape the tables from Wikipedia using the `rvest` package in R (code is on the GitHub page)

Data management functions

1. Load the result files, inspect them, and remove unnecessary parts

```
res2014 <- read_csv("res2014.csv")
res2014 <- res2014[1:42,]
res2019 <- read_csv("res2019.csv")
res2019 <- res2019[1:22,] %>%
  mutate(Party=ifelse(Party=="Other parties/lists",
                      "Other parties",
                      Party))
```

2. Merge the two result files and select only vote percentages

```
results <- full_join(res2014,res2019,by="Party")
results <- results %>%
  select(Party, Percentage2014, Percentage2019)
```

Is the `results` data set tidy?

Data management functions

3. Reshape the data set to make it tidy

```r
results <- results %>%
  pivot_longer(
    cols = starts_with("Percentage"),
    names_to = "Year",
    values_to = "Percentage",
    names_prefix = "Percentage"
  )
head(results, 4)
```

```
# A tibble: 4 x 3
  Party             Year  Percentage
  <chr>             <chr> <chr>
1 Nidaa Tounes      2014  37.56
2 Nidaa Tounes      2019  1.51
3 Ennahda Movement  2014  27.80
4 Ennahda Movement  2019  19.63
```

Nice tables with stargazer

```
library(stargazer)

tab2014 <- results %>%
  filter(Year==2014 & !is.na(Percentage))

stargazer(tab2014,
          summary = F,
          type = "html",
          out = "table.html") # or type="html" or type="lat
```

Exercises

GitHub repository

There is a GitHub repository for this class where I will share materials with you. You can reach it here:

https://github.com/KevinKoehlerSSSA/Intro-to-R

Exercises

1. Read the data into R, save it in your environment as `tun22`.
2. Consult the codebook to understand what you are looking at
3. Write code to calculate:
    3.1 The typical age of all respondents
    3.2 The typical age of respondents who have voted for Kais Saied in the first round of the presidential elections
4. Describe how much Saied voters differ from each other in terms of the levels of education. Which measure would you use? Why?
    4.1 Write the appropriate code

Additional exercises

1. What is the vote percentage of Kais Saied in the first and second round?
2. Are first round Saied voters significantly younger or older than respondents overall?
3. Are male respondents significantly more likely to have voted Saied in the first round than female respondents?