

Data manipulation & visualization with R

Before starting, (if not yet done) download and install R and RStudio in the following order:

1. R : <https://cran.r-project.org/bin/windows/base/>
2. RStudio : <https://rstudio.com/products/rstudio/download/#download>.

Note: do not install on administrator mode.

1 Dataset description

In the folder `data` you will find the following data :

- the file `results_pres_elections_dept_2017_round_1.csv` contains information describing the results of the first round of the 2017 elections per department in France. It provides the following variables:
 - `region_code` : the code of the region where the voting took place
 - `region_name` : the name of the region
 - `dept_code` : the code of the department where the voting took place
 - `dept_name` : the name of the department
 - `registered_voters` : the number of registered voters in that place
 - `absent_voters` : the number of absent voters
 - `present_voters` : the number of voters that actually voted
 - `blank_ballot` : the number of blank votes
 - `null_ballot` : the number of null votes
 - `votes_cast` : the number of valid votes
 - a variable per candidate (the variable is the name of the candidate) : the number of votes they received
- the file `results_pres_elections_dept_2017_round_2.csv` contains information describing the results of the second tour of the 2017 elections per department in France. The variables in this file are the same ones from the above described file, except for the variables representing the candidates, which were replaced by the two candidates of the second round.
- the file `regions.csv` contains information describing the different regions of France through the following variables:
 - `id`: number of regions
 - `code`: the code of the region as we find in the previous files
 - `name`: the name of the region
 - `slug`: the normalized name of the region (lower case and without accents)
- the file `departments.csv` contains information describing the different departments of France through the following variables:

- **id**: number of departments
- **region_code**: the code of the region to which this department belongs
- **code**: the code of the department as we find in the previous files
- **name**: the name of the department
- **slug**: the normalized name of the department (lower case and without accents)
- the file **coordinates_regions_2016.csv** contains the geographical coordinates of the different regions of France:
 - **insee_reg** : the region’s code as we find in the previous files
 - **latitude** : the region’s latitude
 - **longitude** : the region’s longitude
- the folder **shapefile** contains the files describing the geographical features of the different regions of France:
 - **contours-geographiques-des-regions-2019.shp**: contains all information regarding the geometry of the regions, i.e. the polygons
 - **contours-geographiques-des-regions-2019.shx**: a positional index of the feature geometry to allow seeking forwards and backwards quickly
 - **contours-geographiques-des-regions-2019.dbf**: contains attributes regarding the geometric features of the shapefile (file .shp)
 - **contours-geographiques-des-regions-2019.prj**: information about the coordinates system

2 Loading the data and preparing the working environment

1. Install the following packages (you can use the R console for this):
 - **tidyverse** to manipulate data and create static visualizations : `install.packages("tidyverse")`
 - **sf** to manipulate geographic data : `install.packages("sf")`
 - **leaflet** to create interactive maps : `install.packages("leaflet")`
2. Open a new R script (via the “+” button in the top-left corner of RStudio) and load the required packages using the `library()` function.
3. Load the presidential election results for round 1 from the file **results_pres_elections_dept_2017_round_1.csv** and store them in a variable named **round_1**.
4. Load the presidential election results for round 2 from the file **results_pres_elections_dept_2017_round_2.csv** into a variable named **round_2**.
5. Load the regional coordinates from **coordinates_regions_2016.csv** into a variable named **geo_data**.

3 Exercises

3.1 Data processing

1. Extend the tibbles **round_1** et **round_2** by calculating the following attributes:
 - the abstention rate per department: defined as the ratio between the number of absent and registered voters

- the voting rate per department: defined as the ratio between the number of present and registered voters
 - the blank ballot rate: defined as the ratio between the number of blank votes and votes cast
 - the null ballot rate: defined as the ratio between the number of null votes and votes cast
 - the votes cast rate: defined as the ratio between the number of votes cast and voters
 - the voting rate: defined as the ratio between the number of voters and registered voters
2. Calculate the following attributes from the **tibbles** `round_1` and `round_2`, and store them into new **tibbles** (i.e. create new variables to contain each attribute)
 - the sum of votes cast per region
 - the average number of voters per region
 3. Enrich `round_1` and `round_2` tibbles by joining the regional coordinates (see `dplyr::left_join()`)
 4. In preparation to create a single tibble containing all the data, create a new variable (e.g. named `round`) into each tibble (i.e. `round_1` and `round_2`) with values 1 and 2, respectively. Thus, indicating the corresponding round of the election.
 5. Combine `round_1` and `round_2` in a single **tibble** named `results`. Notice that the previously created variable allows to distinguish both rounds in this new tibble.
 6. Re-structure the tibble to follow a **tidy** format (see `tidyr::gather()`), regarding the presentation of candidates.

3.2 Simple data visualization with R

Should you arrive at this section prior to the professor's introduction to data visualization, please begin the exercise by consulting the corresponding slides on Moodle.

1. Create a bar chart to visualize the number of valid votes in the first round per candidate and region
2. Create a chart that shows the number of valid votes in the first and second round side-by-side per candidate and region
3. Create a static map that shows the number of valid votes for each candidate of the second round
4. Create an interactive map that shows the blank ballot rate per region