Assignment 1

John Hulsey

How R Works

R is a powerful tool for data analysis and visualization. It is a programming language that is designed for data analysis. Researchers and users contribute tools and functions.

- *R* is a software program that you will need to install on your computer.
- *RStudio* is an integrated development environment (IDE) that makes it easier to use R. You will rarely work with R without using RStudio.
- When you work in RStudio, you will write a code script that issues commands to R, which will read in data, perform the analysis and the produce outputs. In this class we'll use .R scripts and Quarto (.qmd) documents. Quarto (.qmd) documents allow you to embed code and output within a text document so that you can produce reports, presentations and websites seamlessly. The document you are reading now is generated from a quarto document.

Working in R Studio

For almost everyting you do in R, you will create variables or objects, which are then passed to functions in order to perform the actions that you want, whether that is reading in data, changings the data (usually in a data frame), or producing output (usually in the form of a figure or a table.)

While it is possible to work in R by using the Console directly (ie. without using a script), you should always work in a script because scripts make your work reproducible. You can save your script and rerun it at any time to reproduce your results. If you find a problem in you output, you can go back and change the script, then re-run it quickly. You can also re-use scripts with minor changes to incorporate new data or new analyses. In this class, most assignments will require you to submit your script (.R or .qmd) along with an output.

Literate Programming using Quarto and .qmd files

Literate programming is a way of mixing code and text in a single document. This allows you to explain your code and your results in the same document. Quarto is a package that allows you to create literate programming documents in R. You can use Quarto to create reports, presentations, and websites that include code, output, and text. This document is an example of a Quarto document.

The body of the document that contains the text is written in Markdown, which is a simple way of formatting text. Markdown is a very simple formatting code that lets you create headings, lists and links as well as italicize and bold text. Here is a sheet with the basics of Markdown formatting.

The analysis is done by R code in *chunks* The code chunks are written in R, and they are surrounded by three backticks. The code chunks are executed by R, and the output is displayed in the document.

```
2+2
```

```
[1] 4
```

You can set options within each chunk that determine whether and how it is shown in the document. For example, the chunks that contain code that prepare the data are run in R but are not shown in the final document, but the chunks that show the results and figures are shown. In this document, I am showing everything.

You can (and should!) also add comments to your code. Comments are lines that start with a #. Comments are not executed by R, but they are useful for explaining what your code is doing. You can use this to explain your code to me, but most important is that it allows present you to communicate with future you about what you were doing.

```
mynumber <- 2+2 # This line creates a variable called mynumber that is the sum of 2+2
```

It is a good practice to have a section at the beginning of your script that loads all the packages that you will use in the script.

Our first Analysis

For this assignment, we're going to work with Municipal-level Data from from the Kantonal elections in Bosnia in 2018.

Download the file munkantonsper2018.csv from here https://hulseyjw.github.io/Izbori2018/munkantonsper2018. and place it in a new folder within the class folder you created for this class. So, something like "Documents\POSC644\Week1\"

Then save this .qmd file Week1Assignment.qmd from Canvas and save it into that folder as well. It is a good practice to have your script file (.qmd) and your data file(s) in a seperate folder for each project or assignment.

Some best practices

Sometimes, R will carry over some loaded packages or data files from previous instances. So, it is a good idea to restart R and clear your environment before running your code from the beginning. To do so, go the Session Menu in RStudio and choose Clear Environment and Restart R.

Packages and Libraries

R comes with many functions built-in like *plot* and *read.csv* but most functions aren't built-in and have to be added as part of a package. There are thousands of packages, which can extend R to perform a wide range of functions. The first time that you want to use a package, you have to install the package. This usually only has to be done once. However, every time you restart R, you have to load the packages that you want to use using the library() command.

We're going to start by installing the tidyverse set of packages, which includes many packages we'll use.

To install, copy *install.packages("tidyverse")* into the Console and press Enter.

Once you've installed the package, you can load it for use by using the library() command. Normally, you'll have a list of packages that you'll use at the top of your R script. Add the library command for tidyverse to your script and run it.

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

In order to render .qmd documents into pdf, you'll also need to install the `tinytex` package. You can install the package using the same method as above. Go ahead and use *install.packages* to install the `tinytex` package.

Load the data

The following code chunk:

-reads in the file you've just downloaded using read_csv.

-figuring out the path can be a challenge. This video may help.

```
kper18 <- read.csv("~/Dropbox/Teaching/EUPSAnalysisVis/GitRepo/Week1/munkantonsper2018.csv")
```

The command creates a data frame called kper18 that contains the data 79 observations of 85 variables. You can see that the data is loaded by looking at the Environment tab in RStudio.

Looking at your data

If you have succeeded in loading the data, we can use several commands to look at the data.

```
class(kper18) # this command tells you what type of object kper18 is. It should be a data fra
```

```
[1] "data.frame"
```

```
dim(kper18) # this command tells you the dimensions of the data frame. It should be 79 obser
```

```
[1] 79 85
```

```
head(kper18) # this command shows you the first 6 rows of the data frame.
```

	sifra	Municipality	Municipality2	District	Ballots	Empty.Ballots	
1	79	ŽIVINICE	Å½IVINICE (079)	KANTON 3	27915	856	
2	116	VISOKO	VISOKO (116)	KANTON 4	17593	801	
3	114	FOJNICA	FOJNICA (114)	KANTON 6	5719	284	
4	80	KALESIJA	KALESIJA (080)	KANTON 3	14114	592	
5	117	BREZA	BREZA (117)	KANTON 4	6016	388	
6	115	KISELJAK	KISELJAK (115)	KANTON 6	8533	338	

	Other.Invalid.Ballots	Numvoters	p00090	p00008	p02321	p00001	p02328	p00004
1	1189	53921	35.78	14.90	11.87	7.03	6.63	5.55
2	953	33885	29.67	13.07	1.02	6.14	2.70	1.50
3	325	10132	37.14	17.57	0.10	0.00	0.12	2.33
4	826	32576	28.95	24.14	16.95	5.44	7.19	1.38
5	422	11357	29.27	12.97	0.96	1.96	1.26	2.73
6	331	16920	24.35	7.57	0.09	0.00	0.05	0.91

	p01698	p01182	p02385	p00877	p02402	p02373	p00036	p00513	p02399	p01276	p02401
1	4.95	4.65	2.46	1.41	1.16	1.12	0.82	0.66	0.52	0.09	0.09
2	10.72	11.35	0.00	4.63	0.22	6.66	0.58	3.17	0.00	0.15	0.00
3	4.48	8.53	0.00	0.00	0.19	1.00	0.12	0.56	0.00	0.07	0.00
4	4.44	4.02	0.05	0.54	0.98	3.37	0.72	1.13	0.42	0.06	0.01
5	7.60	22.21	0.00	1.56	0.93	5.10	0.42	4.92	0.00	0.33	0.00
6	3.82	3.27	0.00	0.00	0.25	0.94	0.14	0.54	0.00	0.12	0.00

	p02325	p02323	p01728	p02368	p00521	p01718	p01178	p01705	pNA	p02403	p02317
1	0.08	0.07	0.06	0.03	0.03	0.02	0.01	0.01	0	0.00	0.00
2	0.20	0.00	0.08	0.07	0.00	0.04	0.00	0.00	0	6.84	0.58
3	0.16	0.00	0.05	0.02	0.00	0.02	0.00	0.00	0	0.00	0.00
4	0.05	0.06	0.08	0.01	0.01	0.01	0.01	0.01	0	0.00	0.00
5	0.20	0.00	0.10	0.08	0.00	0.03	0.00	0.00	0	5.50	1.25
6	0.13	0.00	0.02	0.12	0.00	0.01	0.00	0.00	0	0.00	0.00

	p02382	p02391	p02395	p02314	p01699	p01701	p00028	p02375	p00502	p00769	p02383
1	0.00	0.00	0.00	0.00	0.00	0	0	0	0	0	0
2	0.48	0.13	0.00	0.00	0.00	0	0	0	0	0	0
3	20.13	0.00	4.60	2.73	0.09	0	0	0	0	0	0
4	0.00	0.00	0.00	0.00	0.00	0	0	0	0	0	0
5	0.47	0.15	0.00	0.00	0.00	0	0	0	0	0	0
6	50.94	0.00	5.64	1.05	0.04	0	0	0	0	0	0

	p02311	p00017	p01200	p02384	p01703	p02322	p00734	p01714	p02353	p02358	p02388
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0

	p01270	p02389	p02398	p01188	p01290	p02400	p02396	p02355	p02360	p02361	p02072

```
1      0      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      0      0      0      0      0
5      0      0      0      0      0      0      0      0      0      0      0
6      0      0      0      0      0      0      0      0      0      0      0
  p02354 p02390 p02392 p02376 p00882 p02387 p01989 p01723 p02347 p02397 p02291
1      0      0      0      0      0      0      0      0      0      0      0
2      0      0      0      0      0      0      0      0      0      0      0
3      0      0      0      0      0      0      0      0      0      0      0
4      0      0      0      0      0      0      0      0      0      0      0
5      0      0      0      0      0      0      0      0      0      0      0
6      0      0      0      0      0      0      0      0      0      0      0
  p02359 p01706 p02349 Entity Election
1      0      0      0    Fed  Kantons
2      0      0      0    Fed  Kantons
3      0      0      0    Fed  Kantons
4      0      0      0    Fed  Kantons
5      0      0      0    Fed  Kantons
6      0      0      0    Fed  Kantons
```

```
names(kper18) # this command shows you the names of the variables in the data frame.
```

```
 [1] "sifra"                "Municipality"         "Municipality2"
 [4] "District"             "Ballots"              "Empty.Ballots"
 [7] "Other.Invalid.Ballots" "Numvoters"           "p00090"
[10] "p00008"               "p02321"               "p00001"
[13] "p02328"               "p00004"               "p01698"
[16] "p01182"               "p02385"               "p00877"
[19] "p02402"               "p02373"               "p00036"
[22] "p00513"               "p02399"               "p01276"
[25] "p02401"               "p02325"               "p02323"
[28] "p01728"               "p02368"               "p00521"
[31] "p01718"               "p01178"               "p01705"
[34] "pNA"                  "p02403"               "p02317"
[37] "p02382"               "p02391"               "p02395"
[40] "p02314"               "p01699"               "p01701"
[43] "p00028"               "p02375"               "p00502"
[46] "p00769"               "p02383"               "p02311"
[49] "p00017"               "p01200"               "p02384"
[52] "p01703"               "p02322"               "p00734"
[55] "p01714"               "p02353"               "p02358"
```

```
[58] "p02388"                "p01270"                "p02389"
[61] "p02398"                "p01188"                "p01290"
[64] "p02400"                "p02396"                "p02355"
[67] "p02360"                "p02361"                "p02072"
[70] "p02354"                "p02390"                "p02392"
[73] "p02376"                "p00882"                "p02387"
[76] "p01989"                "p01723"                "p02347"
[79] "p02397"                "p02291"                "p02359"
[82] "p01706"                "p02349"                "Entity"
[85] "Election"
```

One of the most useful functions is summary(), which gives you basic summary information on the object. For a data.frame like kper18, it will give you summary information for each variable.

```
summary(kper18) # this command gives you a summary of the data frame.
```

```
     sifra       Municipality       Municipality2        District
 Min.   :  1.0   Length:79          Length:79          Length:79
 1st Qu.: 53.5   Class :character   Class :character   Class :character
 Median : 98.0   Mode  :character   Mode  :character   Mode  :character
 Mean   : 95.8
 3rd Qu.:132.0
 Max.   :199.0
    Ballots       Empty.Ballots    Other.Invalid.Ballots   Numvoters
 Min.   :  422   Min.   :   4.0   Min.   :  11.0         Min.   :   600
 1st Qu.: 4805   1st Qu.: 129.0   1st Qu.: 184.5         1st Qu.:  9722
 Median : 9102   Median : 290.0   Median : 375.0         Median : 20006
 Mean   :12196   Mean   : 386.7   Mean   : 549.7         Mean   : 25407
 3rd Qu.:15118   3rd Qu.: 544.0   3rd Qu.: 782.0         3rd Qu.: 33871
 Max.   :58225   Max.   :1655.0   Max.   :2911.0         Max.   :112479
     p00090          p00008           p02321           p00001
 Min.   : 0.00   Min.   : 0.000   Min.   : 0.000   Min.   :0.000
 1st Qu.:14.98   1st Qu.: 3.965   1st Qu.: 0.000   1st Qu.:0.000
 Median :25.71   Median : 9.580   Median : 0.000   Median :0.210
 Mean   :23.44   Mean   :12.187   Mean   : 3.968   Mean   :1.085
 3rd Qu.:31.85   3rd Qu.:16.140   3rd Qu.: 0.990   3rd Qu.:1.740
 Max.   :62.24   Max.   :53.040   Max.   :57.700   Max.   :7.030
     p02328           p00004           p01698           p01182
 Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
 1st Qu.: 0.000   1st Qu.: 0.025   1st Qu.: 1.645   1st Qu.: 1.605
 Median : 0.100   Median : 0.850   Median : 5.250   Median : 5.280
```

```
Mean   : 1.456   Mean   : 2.334   Mean   : 4.722   Mean   : 5.771
3rd Qu.: 1.240   3rd Qu.: 2.445   3rd Qu.: 6.840   3rd Qu.: 8.270
Max.   :19.250   Max.   :20.050   Max.   :15.830   Max.   :24.150
    p02385           p00877           p02402           p02373
Min.   :0.000   Min.   : 0.000   Min.   :0.0000   Min.   : 0.000
1st Qu.:0.000   1st Qu.: 0.000   1st Qu.:0.0000   1st Qu.: 0.000
Median :0.000   Median : 0.780   Median :0.0200   Median : 0.000
Mean   :0.301   Mean   : 2.435   Mean   :0.1554   Mean   : 1.423
3rd Qu.:0.000   3rd Qu.: 2.725   3rd Qu.:0.1800   3rd Qu.: 1.590
Max.   :9.460   Max.   :21.290   Max.   :1.1600   Max.   :18.960
    p00036           p00513           p02399           p01276
Min.   :0.0000   Min.   :0.0000   Min.   : 0.0000   Min.   :0.00000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.:0.01000
Median :0.1800   Median :0.4100   Median : 0.0000   Median :0.07000
Mean   :0.5825   Mean   :0.6467   Mean   : 0.4201   Mean   :0.08165
3rd Qu.:0.4900   3rd Qu.:0.7700   3rd Qu.: 0.0000   3rd Qu.:0.12000
Max.   :6.0000   Max.   :4.9200   Max.   :13.1000   Max.   :0.33000
    p02401           p02325           p02323           p01728
Min.   :0.000000   Min.   :0.00   Min.   :0.00000   Min.   : 0.000
1st Qu.:0.000000   1st Qu.:0.02   1st Qu.:0.00000   1st Qu.: 0.020
Median :0.000000   Median :0.06   Median :0.00000   Median : 0.050
Mean   :0.007468   Mean   :0.07   Mean   :0.05456   Mean   : 0.302
3rd Qu.:0.000000   3rd Qu.:0.11   3rd Qu.:0.06000   3rd Qu.: 0.065
Max.   :0.280000   Max.   :0.23   Max.   :0.49000   Max.   :13.820
    p02368           p00521           p01718           p01178
Min.   : 0.000   Min.   :0.000000   Min.   :0.00000   Min.   :0.000000
1st Qu.: 0.010   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.000000
Median : 0.080   Median :0.000000   Median :0.01000   Median :0.000000
Mean   : 2.211   Mean   :0.002785   Mean   :0.02203   Mean   :0.001266
3rd Qu.: 0.280   3rd Qu.:0.000000   3rd Qu.:0.03000   3rd Qu.:0.000000
Max.   :59.060   Max.   :0.070000   Max.   :0.20000   Max.   :0.020000
    p01705           pNA           p02403           p02317
Min.   : 0.0000   Min.   :0   Min.   : 0.0000   Min.   : 0.000
1st Qu.: 0.0000   1st Qu.:0   1st Qu.: 0.0000   1st Qu.: 0.000
Median : 0.0000   Median :0   Median : 0.0000   Median : 0.000
Mean   : 0.7986   Mean   :0   Mean   : 0.8803   Mean   : 1.844
3rd Qu.: 0.0350   3rd Qu.:0   3rd Qu.: 0.0000   3rd Qu.: 0.485
Max.   :17.0800   Max.   :0   Max.   :29.5600   Max.   :19.000
    p02382           p02391           p02395           p02314
Min.   : 0.00   Min.   :0.0000   Min.   :0.0000   Min.   : 0.000
1st Qu.: 0.00   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 0.000
Median : 0.23   Median :0.0000   Median :0.0000   Median : 0.020
Mean   :12.89   Mean   :0.1365   Mean   :0.6954   Mean   : 1.266
```

```
3rd Qu.:13.47   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.: 1.530
Max.   :89.73   Max.   :5.8600   Max.   :9.6500   Max.   :17.870
    p01699            p01701            p00028           p02375
Min.   :0.00000   Min.   : 0.0000   Min.   : 0.000   Min.   : 0.000
1st Qu.:0.00000   1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.000
Median :0.00000   Median : 0.0000   Median : 0.000   Median : 0.000
Mean   :0.05519   Mean   : 0.6365   Mean   : 0.223   Mean   : 1.941
3rd Qu.:0.02500   3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.: 0.000
Max.   :1.47000   Max.   :46.5600   Max.   :11.880   Max.   :46.490
    p00502            p00769            p02383           p02311
Min.   : 0.0000   Min.   : 0.0000   Min.   :0.00000   Min.   :0.00000
1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.:0.00000   1st Qu.:0.00000
Median : 0.0000   Median : 0.0000   Median :0.00000   Median :0.00000
Mean   : 0.3967   Mean   : 0.7362   Mean   :0.02506   Mean   :0.02253
3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.:0.00000   3rd Qu.:0.00000
Max.   :10.6700   Max.   :23.9900   Max.   :1.50000   Max.   :0.72000
    p00017           p01200            p02384           p01703
Min.   :0.0000   Min.   : 0.0000   Min.   : 0.000   Min.   : 0.0000
1st Qu.:0.0000   1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.0000
Median :0.0000   Median : 0.0000   Median : 0.000   Median : 0.0000
Mean   :0.2146   Mean   : 0.9079   Mean   : 2.039   Mean   : 0.6866
3rd Qu.:0.0000   3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.: 0.0000
Max.   :9.4300   Max.   :47.7100   Max.   :55.820   Max.   :23.1500
    p02322           p00734           p01714           p02353
Min.   :0.0000   Min.   :0.0000   Min.   :0.000000   Min.   :0.000000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.:0.000000
Median :0.0000   Median :0.0000   Median :0.000000   Median :0.000000
Mean   :0.1867   Mean   :0.1753   Mean   :0.002405   Mean   :0.001266
3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.000000   3rd Qu.:0.000000
Max.   :9.1400   Max.   :4.9900   Max.   :0.090000   Max.   :0.050000
    p02358            p02388           p01270           p02389
Min.   :0.000000   Min.   :0.00000   Min.   :0.00000   Min.   : 0.000
1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.: 0.000
Median :0.000000   Median :0.00000   Median :0.00000   Median : 0.000
Mean   :0.001266   Mean   :0.08506   Mean   :0.01013   Mean   : 1.886
3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.: 0.000
Max.   :0.060000   Max.   :3.92000   Max.   :0.16000   Max.   :61.050
    p02398           p01188           p01290           p02400
Min.   : 0.0000   Min.   : 0.0000   Min.   :0.0000   Min.   : 0.0000
1st Qu.: 0.0000   1st Qu.: 0.0000   1st Qu.:0.0000   1st Qu.: 0.0000
Median : 0.0000   Median : 0.0000   Median :0.0000   Median : 0.0000
Mean   : 0.6743   Mean   : 0.3594   Mean   :0.1091   Mean   : 0.2565
3rd Qu.: 0.0000   3rd Qu.: 0.0000   3rd Qu.:0.0000   3rd Qu.: 0.0000
```

```
Max.   :30.9900   Max.   :17.7600   Max.   :3.2100   Max.   :10.7800
    p02396            p02355            p02360            p02361
Min.   : 0.000   Min.   :0.000000   Min.   :0.000000   Min.   :0.0000000
1st Qu.: 0.000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.0000000
Median : 0.000   Median :0.000000   Median :0.000000   Median :0.0000000
Mean   : 1.338   Mean   :0.004684   Mean   :0.008481   Mean   :0.0008861
3rd Qu.: 0.000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.0000000
Max.   :38.750   Max.   :0.110000   Max.   :0.200000   Max.   :0.0300000
    p02072            p02354            p02390            p02392
Min.   :0.0000   Min.   :0.000000   Min.   : 0.000   Min.   : 0.0000
1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.: 0.000   1st Qu.: 0.0000
Median :0.0000   Median :0.000000   Median : 0.000   Median : 0.0000
Mean   :0.2166   Mean   :0.003924   Mean   : 3.359   Mean   : 0.5384
3rd Qu.:0.0000   3rd Qu.:0.000000   3rd Qu.: 0.000   3rd Qu.: 0.0000
Max.   :2.4300   Max.   :0.070000   Max.   :72.410   Max.   :16.3900
    p02376            p00882            p02387            p01989
Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
Median :0.0000   Median :0.0000   Median :0.0000   Median :0.0000
Mean   :0.2452   Mean   :0.1362   Mean   :0.1148   Mean   :0.1159
3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:0.0000
Max.   :9.3600   Max.   :1.6200   Max.   :2.7700   Max.   :2.2000
    p01723            p02347            p02397            p02291
Min.   :0.00000   Min.   :0.000000   Min.   :0.000000   Min.   :0.0000
1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.0000
Median :0.00000   Median :0.000000   Median :0.000000   Median :0.0000
Mean   :0.04367   Mean   :0.007342   Mean   :0.008481   Mean   :0.0219
3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.0000
Max.   :1.35000   Max.   :0.090000   Max.   :0.220000   Max.   :0.5100
    p02359            p01706            p02349            Entity
Min.   :0.00000   Min.   :0.000000   Min.   :0.000000   Length:79
1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000000   Class :character
Median :0.00000   Median :0.000000   Median :0.000000   Mode  :character
Mean   :0.00962   Mean   :0.002658   Mean   :0.002405
3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.000000
Max.   :0.17000   Max.   :0.060000   Max.   :0.040000
  Election
Length:79
Class :character
Mode  :character
```

For the numeric variables, it shows the mean, median, range and quartiles of the variable. For the text variables, it just the type class "character". Later, we'll work on turning these character variables into factor variables, which are more useful. You can also click on the data.frame in the Environment tab and it will show it like a spreadsheet.

This is a dataset of the results for the Kantonal elections in Bosnia and Herzegovina in 2018 aggregated at the municipal level. We can see from the output that there are 79 rows and 85 columns. There are 79 rows because there are 79 municipalities in the Federation half of Bosnia, which is where the cantons are located. There are 85 columns because there are 9 columns with information about the municipality like its numeric code (sifra), name(Municipality), Kanton (District), total number of Ballots, and the percentage of votes for each of the parties. There are 76 parties receiving votes in one of the 10 Kantonal elections. It's a lot of parties. This is too many for us to think about right now, so we're going to use some Tidyverse functions to trim things down.

The following lines take a few of the parties and information about the municipality and puts them in a new dataset called kper18_trimmed. The second one replaces all of the column names and replaces the party codes with acronyms for our three parties.

```
kper18_trimmed <- kper18 %>%
  select(sifra, Municipality, District, Ballots, p00090, p00008, p01182, p00877, p01698) # tl

names(kper18_trimmed) <- c("sifra", "Municipality", "District", "Ballots", "SDA", "SDP", "SBI
```

So, now you can see how to select a few columns and rename columns. Note that when you use the names function, you have to give a name to all of the columns, even if you aren't changing all of them. The c("sifra", "Municipality) form is of a list of names and the list of names must be equal to the list of columns in your dataset. You shouldn't have spaces in your column names.

Now, we can look at the data using the same functions as before and see the changes.

```
dim(kper18_trimmed)
```

```
[1] 79  9
```

```
head(kper18_trimmed)
```

```
  sifra Municipality District Ballots   SDA   SDP   SBB   NS    DF
1    79     ŽIVINICE KANTON 3   27915 35.78 14.90  4.65 1.41  4.95
2   116       VISOKO KANTON 4   17593 29.67 13.07 11.35 4.63 10.72
3   114      FOJNICA KANTON 6    5719 37.14 17.57  8.53 0.00  4.48
```

```
4    80     KALESIJA KANTON 3   14114 28.95 24.14  4.02 0.54  4.44
5   117        BREZA KANTON 4    6016 29.27 12.97 22.21 1.56  7.60
6   115     KISELJAK KANTON 6    8533 24.35  7.57  3.27 0.00  3.82
```

Now we have the "sifra" a code for each municipality, the name of the municipality, the district (Kanton) that the municipality is in, the total number of ballots, and the percentage of votes for the SDA, SDP, SBB, NS, and DF parties.

Subsetting by values

Using select() above, we learned how to choose columns. Another useful tool is to be able to subset the dataset by values of particular rows. For example, we might want to see just the municipalities from Kanton 1. To do that, we'll use the filter command.

```
kper18_trimmed %>% filter(District == "KANTON 1")
```

```
  sifra      Municipality District Ballots   SDA   SDP   SBB   NS    DF
1     1    VELIKA KLADUŠA KANTON 1   13927 18.45  3.02  7.35 1.20  2.00
2     2             CAZIN KANTON 1   22173 30.26  5.35  1.31 1.72  8.19
3     3             BIHAĆ KANTON 1   21027 26.27 12.60  8.96 7.45  7.34
4     4    BOSANSKA KRUPA KANTON 1    9194 30.12 19.15  7.25 7.02 15.83
5     5             BUŽIM KANTON 1    6619 26.53  3.91 14.32 0.82  2.60
6    30 BOSANSKI PETROVAC KANTON 1    2210 34.21  7.96  7.24 3.89  5.93
7    32       SANSKI MOST KANTON 1    9604 35.68 17.33 14.71 5.04  7.94
8    59             KLJUČ KANTON 1    4762 39.71 31.90  7.64 1.22  6.89
```

Note that we have to use == when we are doing a logical test and we have to use quotation marks when the value is characters as opposed to a number.

We could also use filter() to find only those municipalities where DF received more than 10 percent of the votes.

I've been using the "tidyverse" way of doing things, which uses the pipe operator %>% to chain functions together. This is a way of making your code more readable. The pipe operator takes the output of the function on the left and uses it as the input for the function on the right.

```
kper18_trimmed %>% filter(DF > 10) #tidy command using the pipe operator which always puts tl
```

```
  sifra   Municipality District Ballots   SDA   SDP   SBB    NS    DF
1   116         VISOKO KANTON 4   17593 29.67 13.07 11.35  4.63 10.72
2     4 BOSANSKA KRUPA KANTON 1    9194 30.12 19.15  7.25  7.02 15.83
3   135        VOGOŠĆA KANTON 9   14083 30.79 10.81  8.52  6.30 10.84
4    50          TUZLA KANTON 3   49076 11.95 29.14  8.31 10.22 10.77
5    96          OLOVO KANTON 4    4764 47.69 15.87  7.33  0.78 13.08
```

```
filter(kper18_trimmed, DF > 10) # the same command without the pipe operator but putting the
```

```
  sifra   Municipality District Ballots   SDA   SDP   SBB    NS    DF
1   116         VISOKO KANTON 4   17593 29.67 13.07 11.35  4.63 10.72
2     4 BOSANSKA KRUPA KANTON 1    9194 30.12 19.15  7.25  7.02 15.83
3   135        VOGOŠĆA KANTON 9   14083 30.79 10.81  8.52  6.30 10.84
4    50          TUZLA KANTON 3   49076 11.95 29.14  8.31 10.22 10.77
5    96          OLOVO KANTON 4    4764 47.69 15.87  7.33  0.78 13.08
```

-Add a chunk to your qmd file to see in how many municipalities SDA received more than 50 percent. Add the answer as markdown text after your chunk.

```
kper18_trimmed %>% filter(SDA > 50) #tidy command using the pipe operator which always puts t
```

```
  sifra  Municipality District Ballots   SDA  SDP  SBB   NS   DF
1    55        TEOČAK KANTON 3    2924 62.21 5.64 1.68 0.48 1.92
2   141 TRNOVO (FBIH) KANTON 9    1695 62.24 4.66 4.07 1.30 1.65
```

We can also chain these functions together using %>%. So, we can look for municipalities in Sarajevo Kanton (Kanton 9) where Nasa Stranka (NS) received more than ten percent of the vote and only include the results for NS.

```
kper18_trimmed %>% filter(NS > 10) %>% filter(District =="KANTON 9") %>% select(Municipality
```

```
         Municipality    NS
1  NOVI GRAD SARAJEVO 12.69
2     CENTAR SARAJEVO 21.29
3 STARI GRAD SARAJEVO 11.96
4       NOVO SARAJEVO 20.95
```

```
kper18_trimmed %>% # for more complex chains, we can put each step on a new line. This line c
  filter(NS > 10) %>% # this line filters the data frame to only include municipalities where
  filter(District =="KANTON 9") %>% # this line filters the data frame to only include munici
  select(Municipality, NS) # this line selects the columns Municipality and NS.
```

```
         Municipality    NS
1  NOVI GRAD SARAJEVO 12.69
2     CENTAR SARAJEVO 21.29
3 STARI GRAD SARAJEVO 11.96
4       NOVO SARAJEVO 20.95
```

Here you can see the results are the same using both ways of chaining the functions. There are four municipalities in Kanton 9 where NS received more than 10 percent of the vote.

-Add a chunk to your qmd file to see in how many municipalities SDA received more than 10 percent of the vote in Kanton 3 (Tuzla Canton). Add the answer as markdown text after your chunk.

```
kper18_trimmed %>% # for more complex chains, we can put each step on a new line. This line c
  filter(SDA > 10) %>% # this line filters the data frame to only include municipalities wher
  filter(District =="KANTON 3") %>% # this line filters the data frame to only include munici
  select(Municipality, SDA) # this line selects the columns Municipality and NS.
```

```
    Municipality   SDA
1       ŽIVINICE 35.78
2       KALESIJA 28.95
3          SAPNA 15.92
4        KLADANJ 18.79
5       GRADAČAC 21.81
6  DOBOJ - ISTOK 35.13
7      GRAČANICA 22.10
8        LUKAVAC 20.71
9      SREBRENIK 24.04
10         TUZLA 11.95
11         ČELIĆ 45.81
12        TEOČAK 62.21
13      BANOVIĆI 21.21
```

Summarizing by Category

We often want to summarize information by a categorical variable. For example, we might want to get the total number of ballots cast in each Kanton and the number of municipalities in each Kanton. To do so, we'd group by District, then create two variables that show the sum of the Ballots and the Number (n) of municipalities in each one.

```
kper18_trimmed %>% #starts with our dataset
  group_by(District) %>% #this defines District as the group we care about.
    summarise(SDPvote = mean(SDP), Number = n()) #this command creates two new variables. SDI
```

```
# A tibble: 10 x 3
   District  SDPvote Number
   <chr>       <dbl>  <int>
 1 KANTON 1   12.7        8
 2 KANTON 10   2.85       6
 3 KANTON 2    4.65       3
 4 KANTON 3   19.3       13
 5 KANTON 4   19.5       12
 6 KANTON 5    8.10       3
 7 KANTON 6   13.0       12
 8 KANTON 7    8.45       9
 9 KANTON 8    0.438      4
10 KANTON 9    9.66       9
```

```
kper18_trimmed %>% #starts with our dataset
    summarise(SDPvote = mean(SDP), Number = n()) #this command creates two new variables.
```

```
   SDPvote Number
1 12.18734     79
```

- *Create your own summary that gives the mean() result for SDP and DF in each Kanton. Note that this doesn't give us the accurate number percentage for the whole Kanton, since it would just take the average of the municipalities and not take into account the number of ballots. As a comment say which Kanton has the highest average result for each party.*

```
kper18_trimmed %>% #starts with our dataset
  group_by(District) %>% #this defines District as the group we care about.
    summarise(SDPvote = mean(SDP), DF = mean(DF)) #this command creates two new variables. SI
```

```
# A tibble: 10 x 3
   District  SDPvote    DF
   <chr>       <dbl> <dbl>
 1 KANTON 1   12.7    7.09
 2 KANTON 10   2.85   1.52
 3 KANTON 2    4.65   2.45
 4 KANTON 3   19.3    5.54
```

```
 5 KANTON 4   19.5    7.01
 6 KANTON 5    8.10   3.94
 7 KANTON 6   13.0    4.78
 8 KANTON 7    8.45   2.42
 9 KANTON 8    0.438  0
10 KANTON 9    9.66   5.87
```

- *Bonus question: Write a comment where you describe a strategy for getting the actual Kanton level percents using the information in this table. You don't need to implement it.*

Rendering and Submitting

Once we've created our full document, including text and analysis, we can use the Render button to output it to .pdf. Make sure you save your .qmd file before rendering.

Submit the .qmd file that you've edited as well as the .pdf file that you've rendered. Submit them both to Canvas to complete the assignment.