01 Intro – Crash Course in Statistics (Summer 2025)

Neuroscience Center Zurich – University of Zurich

Dr. Zofia Baranczuk

2025-08-25

0. R Console vs. R Markdown

- Use the **console** for quick, temporary commands and checking results.
- Use an **R Markdown file (.Rmd)** to write code and text together for reproducible analysis.
- In `.Rmd`, combine:
  - **Text chunks** (like this explanation)
  - **Code chunks** (enclosed in `{r} ...`)

1. Loading and Exploring the Dataset

Load the data about depression per country in years 2019-2021 from the link: https://user.math.uzh.ch/baranczuk/znz/Data/depression-rates-by-country-2025.csv Have a look at the data.

```
depression <-
  read.csv("https://user.math.uzh.ch/baranczuk/znz/Data/depression-rates-by-country-2025.csv")

#View(depression) - if you would like to view the data excel style.
head(depression) #the first few lines
```

```
##   flagCode   country RatePer100k_2021 Percentage_2021 Cases_2021
## 1       GL Greenland          7979.49            8.51    4476.69
## 2       GR    Greece          7297.08            7.59  742471.24
## 3       TN   Tunisia          7264.74            7.57  860332.31
## 4       PT  Portugal          6867.34            7.14  728476.73
## 5       LT Lithuania          6726.37            6.96  183515.69
## 6       LS   Lesotho          6630.94            6.89  124289.37
##   RatePer100k_2020 Percentage_2020 Cases_2020 RatePer100k_2019 Percentage_2019
## 1          7830.61            8.35    4389.61          6450.26            6.88
## 2          6938.31            7.23  711199.30          6539.65            6.82
## 3          6603.45            6.89  142445.01          5816.02            6.08
## 4          6720.60            6.99  715280.21          5831.86            6.07
## 5          6099.90            6.30  168557.49          5503.10            5.68
## 6          5658.28            5.88  106016.60          5116.69            5.32
##   Cases_2019
## 1    3611.28
## 2  674672.63
## 3  680630.61
## 4  622292.77
## 5  153764.06
## 6   95631.79
```

```r
str(depression)
```

```
## 'data.frame':    204 obs. of  11 variables:
##  $ flagCode        : chr  "GL" "GR" "TN" "PT" ...
##  $ country         : chr  "Greenland" "Greece" "Tunisia" "Portugal" ...
##  $ RatePer100k_2021: num  7979 7297 7265 6867 6726 ...
##  $ Percentage_2021 : num  8.51 7.59 7.57 7.14 6.96 6.89 7.06 6.91 6.65 6.59 ...
##  $ Cases_2021      : num  4477 742471 860332 728477 183516 ...
##  $ RatePer100k_2020: num  7831 6938 6603 6721 6100 ...
##  $ Percentage_2020 : num  8.35 7.23 6.89 6.99 6.3 5.88 6.89 6.03 6.85 5.95 ...
##  $ Cases_2020      : num  4390 711199 142445 715280 168557 ...
##  $ RatePer100k_2019: num  6450 6540 5816 5832 5503 ...
##  $ Percentage_2019 : num  6.88 6.82 6.08 6.07 5.68 5.32 5.9 5.3 6.37 5.32 ...
##  $ Cases_2019      : num  3611 674673 680631 622293 153764 ...
```

2. Working with Data Frames

Which country had the highest depression rate in 2021? Which 10 countries had the lowest depression rate in 2021? How many cases where there in the country with the highest rate? Which countries have less than 4% of depression in the population? What are the depression rates and number of cases in Switzerland?

```r
ind_max <- which.max(depression$RatePer100k_2021)
depression$country[ind_max]
```

```
## [1] "Greenland"
```

```r
# or
m <- max(depression$RatePer100k_2021)
ind_max <- which(depression$RatePer100k_2021 == m)
depression$country[ind_max]
```

```
## [1] "Greenland"
```

```r
cat('max depression rate = ', m, '\n')
```

```
## max depression rate =  7979.49
```

```r
cat('Number of cases in the country with the highest depression rates:',
    depression$Cases_2021[ind_max], "\n")
```

```
## Number of cases in the country with the highest depression rates: 4476.69
```

```r
ind_min <- which.min(depression$RatePer100k_2021)
depression$country[ind_min]
```

```
## [1] "Brunei"
```

```r
inds_low <-which(depression$Percentage_2021<4)
depression$country[inds_low]
```

```
##  [1] "Bahamas"                 "Antigua and Barbuda"
##  [3] "Burundi"                 "Jamaica"
##  [5] "Cook Islands"            "Somalia"
##  [7] "Puerto Rico"             "Venezuela"
##  [9] "Pakistan"                "Togo"
## [11] "Liberia"                 "China"
## [13] "Iceland"                 "South Sudan"
## [15] "Niue"                    "Azerbaijan"
```

```
## [17] "Nicaragua"                "Taiwan"
## [19] "Cameroon"                 "Sri Lanka"
## [21] "Haiti"                    "Zambia"
## [23] "Belize"                   "Maldives"
## [25] "Panama"                   "Uzbekistan"
## [27] "Senegal"                  "Malawi"
## [29] "Albania"                  "Guinea-Bissau"
## [31] "Northern Mariana Islands" "Turkmenistan"
## [33] "Honduras"                 "Sierra Leone"
## [35] "Seychelles"               "Guinea"
## [37] "Tokelau"                  "Sao Tome and Principe"
## [39] "Argentina"                "Chad"
## [41] "Benin"                    "North Korea"
## [43] "Poland"                   "Cambodia"
## [45] "Philippines"              "Tuvalu"
## [47] "Burkina Faso"             "Fiji"
## [49] "Ivory Coast"              "Mauritania"
## [51] "Zimbabwe"                 "Japan"
## [53] "Micronesia"               "Indonesia"
## [55] "Tajikistan"               "Nauru"
## [57] "Marshall Islands"         "Vietnam"
## [59] "Laos"                     "South Korea"
## [61] "American Samoa"           "Kiribati"
## [63] "Nigeria"                  "Vanuatu"
## [65] "Peru"                     "Papua New Guinea"
## [67] "Solomon Islands"          "Colombia"
## [69] "Niger"                    "Timor-Leste"
## [71] "Myanmar"                  "Samoa"
## [73] "Mali"                     "Tonga"
## [75] "Singapore"                "Brunei"
```

```
depression[depression$country=="Switzerland",]
```

```
##    flagCode     country RatePer100k_2021 Percentage_2021 Cases_2021
## 31       CH Switzerland          5502.36            5.77   490965.3
##    RatePer100k_2020 Percentage_2020 Cases_2020 RatePer100k_2019 Percentage_2019
## 31          5048.96             5.3   446898.1          4672.31             4.9
##    Cases_2019
## 31   410071.7
```

3. Accessing Elements of a Data Frame

```
depression[1,] # the first row
```

```
##   flagCode   country RatePer100k_2021 Percentage_2021 Cases_2021
## 1       GL Greenland          7979.49            8.51    4476.69
##   RatePer100k_2020 Percentage_2020 Cases_2020 RatePer100k_2019 Percentage_2019
## 1          7830.61            8.35    4389.61          6450.26            6.88
##   Cases_2019
## 1    3611.28
```

```
depression[,1] # the first column
```

```
##   [1] "GL" "GR" "TN" "PT" "LT" "LS" "PS" "LB" "ES" "BY" "UA" "MA" "LV" "SR" "IR"
##  [16] "BH" "GY" "SM" "EE" "GA" "GB" "CV" "MC" "US" "FI" "CG" "NP" "SE" "IE" "LY"
```

```
##  [31] "CH" "CU" "AE" "ZA" "QA" "IT" "BD" "AD" "MU" "TT" "FR" "GQ" "SA" "UG" "CF"
##  [46] "SZ" "KN" "AU" "TR" "KW" "BW" "IL" "MD" "GE" "SY" "CL" "NL" "AO" "OM" "NO"
##  [61] "BR" "SI" "DE" "BE" "DK" "CD" "RU" "HR" "RW" "DZ" "MX" "AF" "GM" "LC" "BM"
##  [76] "BG" "VI" "JO" "HU" "SD" "LU" "DJ" "IN" "BB" "CZ" "AM" "KM" "BA" "YE" "CA"
##  [91] "BT" "RS" "VC" "KE" "CR" "EG" "DO" NA   "ER" "SK" "ME" "BO" "CY" "IQ" "MT"
## [106] "UY" "AT" "RO" "MG" "MN" "KZ" "MK" "MZ" "GD" "NZ" "EC" "GT" "TH" "PW" "SV"
## [121] "ET" "GH" "TZ" "DM" "GU" "PY" "BS" "AG" "BI" "JM" "MY" "CK" "SO" "PR" "KG"
## [136] "VE" "PK" "TG" "LR" "CN" "IS" "SS" "NU" "AZ" "NI" "TW" "CM" "LK" "HT" "ZM"
## [151] "BZ" "MV" "PA" "UZ" "SN" "MW" "AL" "GW" "MP" "TM" "HN" "SL" "SC" "GN" "TK"
## [166] "ST" "AR" "TD" "BJ" "KP" "PL" "KH" "PH" "TV" "BF" "FJ" "CI" "MR" "ZW" "JP"
## [181] "FM" "ID" "TJ" "NR" "MH" "VN" "LA" "KR" "AS" "KI" "NG" "VU" "PE" "PG" "SB"
## [196] "CO" "NE" "TL" "MM" "WS" "ML" "TO" "SG" "BN"
```

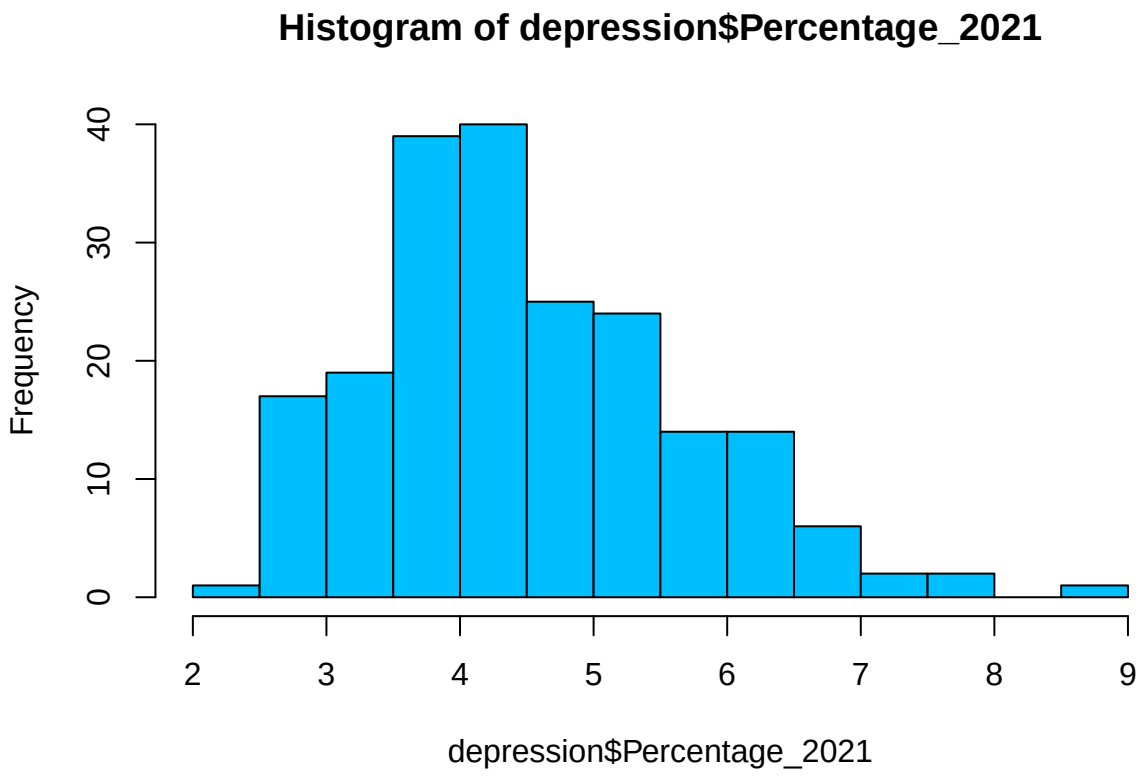
```
depression[c(1:5), c(2,4,5)] # the first 5 rows, columns 2, 4, 5.
```

```
##     country Percentage_2021 Cases_2021
## 1 Greenland            8.51    4476.69
## 2    Greece            7.59  742471.24
## 3   Tunisia            7.57  860332.31
## 4  Portugal            7.14  728476.73
## 5 Lithuania            6.96  183515.69
```

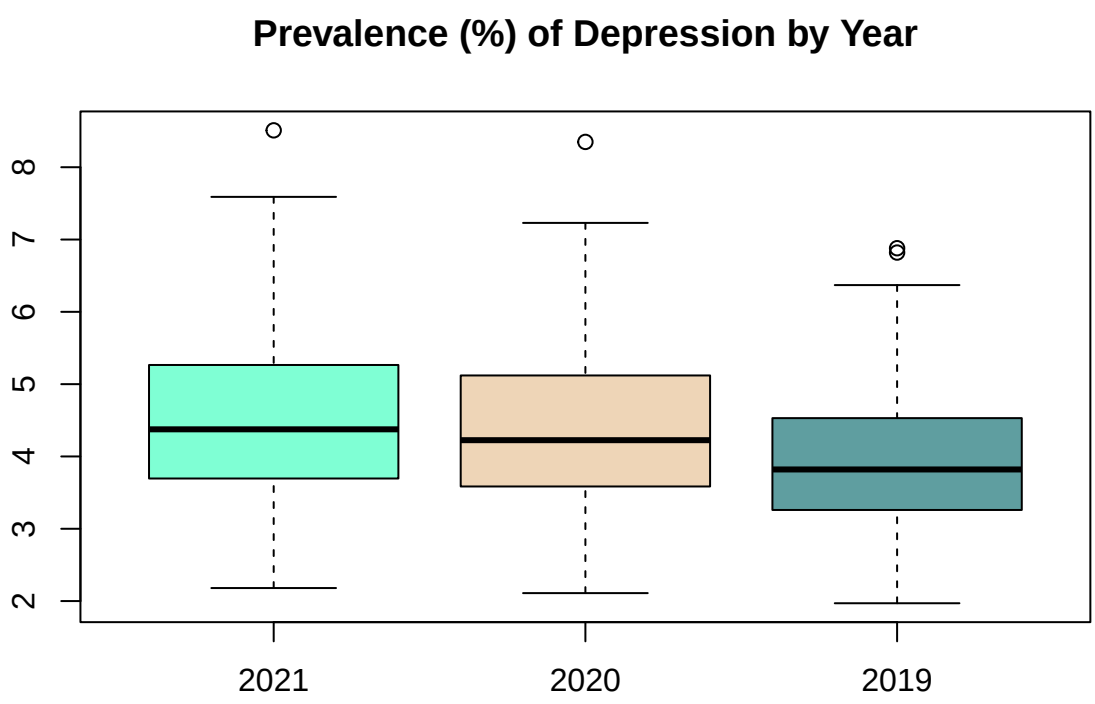
4. Simple plots

Let's have a look at a histogram of depression percentages, boxplots of depression percentages and a scatter plot of depression percentages in 2019 and 2021.

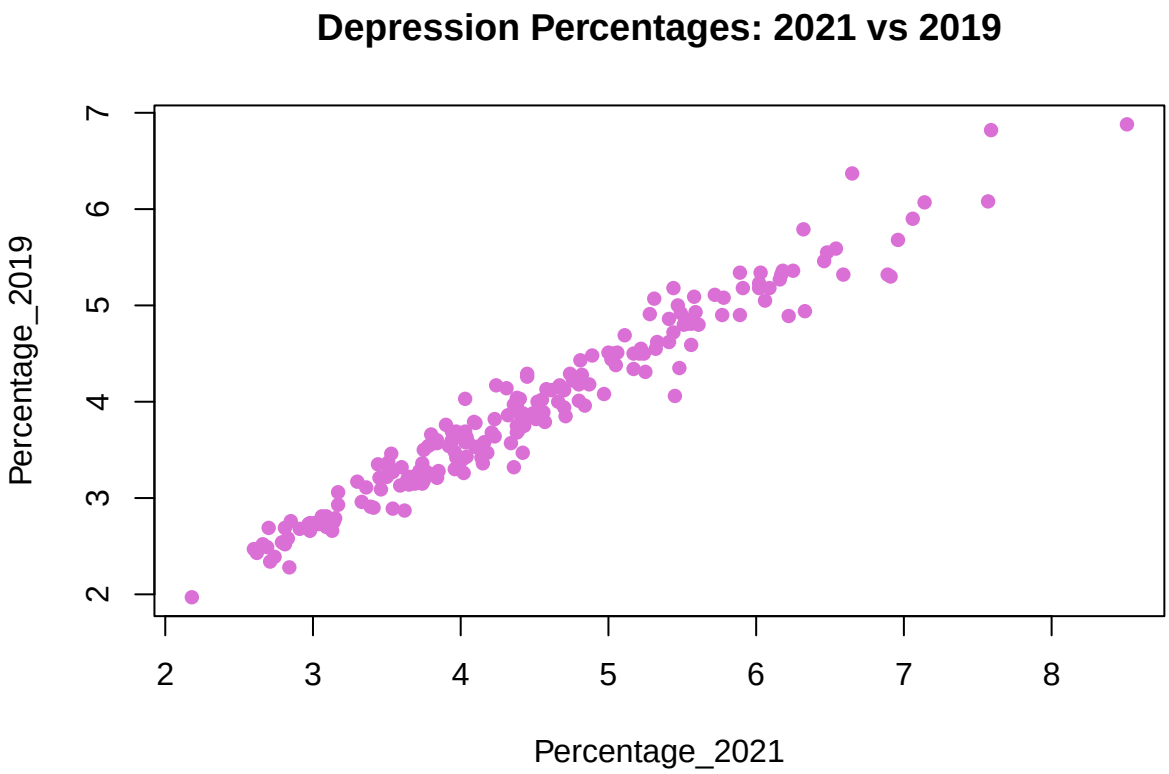
```
hist(depression$Percentage_2021,breaks = 10, col = "deepskyblue" )
```




```r
boxplot( depression$Percentage_2021,  depression$Percentage_2020,
         depression$Percentage_2019 , col = c("aquamarine", "bisque2",
                                              "cadetblue"), xaxt = 'n',
         main = "Prevalence (%) of Depression by Year")
axis(side = 1, at = c(1,2,3), labels = c("2021", "2020", "2019"))
```




```
plot(Percentage_2019 ~ Percentage_2021,
     data = depression,
     col = "orchid",
     pch = 16,
     main = "Depression Percentages: 2021 vs 2019")
```


Depression Percentages: 2021 vs 2019
Percentage_2019
Percentage_2021
2
3
4
5
6
7
8