Some R used in

the workshop

|  |  |
| --- | --- |
| a<- 2+3 | assignment, a will be a variable with the value of 2 + 3 |
| paste("qwer", "ty") | concatenate,  result is "qwer ty" |
| paste0("qwer", "ty") | concatenate,  result is "qwerty" |
| read.csv()  read.csv2()  read\_csv()  read\_csv2()  read\_xlsx() | read data from comma separated values, or ; separated values, or excels |
| install.packages("tidyverse") | install the package "tidyverse" |
| library(tidyverse) | use the package tidyverse for this environmet |
| file.path("datafolder", "filename.csv") | the path that leads to datafolder/filename.csv |
| %>% | pipe operator, the pre thing becomes the first argument of the post thing |
| kable() | nicer display of dataframe between the () |
| gather() | transform dataframe from wide to long |
| spread() | transform dataframe from long to wide |
| select() | choose columns from dataframe |
| filter() | choose rows from dataframe |
| mutate() | make or change columns in dataframe |
| arrange() | order rows in dataframe |
| summarise() | summarise rows in dataframe |
| group\_by | make subgroups from rows, usually combined with 1 of the above |
| n() | amount of rows |
| %in% | "is in the collection" |
| round(1.2345, 1) | 1.2 |
| min() | the minimum |
| max() | the maximum |
| as.integer("3") | 3 as a integer |
| sprintf() | format with a lot of options |
| ggplot() | start of the function that makes nice plots |

**Code sheet 10**

p <- 50

s <- "The number of participants is"

paste(s,p)

**Code sheet 22**

# read file IncomingInternationalDegreeStudents.csv

location\_of\_data <- file.path("data",

"IncomingInternationalDegreeStudents.csv")

dStudents <- read\_csv2(location\_of\_data)

**Code sheet 36**

dStudents %>%

filter(Year %in% c(min\_year, max\_year)) %>%

group\_by(Year) %>%

mutate(Total\_per\_year = sum(n),

part = n / Total\_per\_year,

part = paste0(round(100\*part,0), "%")) %>%

select(Program, Year, part) %>%

spread(Year, part) %>%

# caption in kable means title

kable(caption = "Choice of Program, per year",

align = "lrr")

**Code sheet 41**

# The first year in the data

min\_year <- min(dStudents$Year)

# the most recent year in the data

max\_year <- max(dStudents$Year)

diff\_years <- max\_year - min\_year

# totals per year, only for the min and the max year

dTotalPerYear <- dStudents %>%

group\_by(Year) %>%

summarise(Total = sum(n)) %>%

filter(Year %in% c(min\_year, max\_year))

# base R for calculating the total in max\_year

total\_max\_year <- dTotalPerYear$Total[dTotalPerYear$Year == max\_year] %>%

# the counts were numeric, we want an integer

as.integer()

# another way for the min\_year

total\_min\_year <- dTotalPerYear %>%

filter(Year == min\_year) %>%

select(Total) %>%

# now still a dataset, we want only the value

as.integer()

# the growth

growth\_between <- (total\_max\_year - total\_min\_year) / total\_min\_year

# and now formatted as percentage

growth\_between <- sprintf("%.0f%%", 100\* growth\_between)

**Code sheet 47**

dStudents %>%

ggplot(aes(x = Year, y = n, color = Program)) +

geom\_line(size = 1) +

# better values on the x axis:

scale\_x\_continuous(breaks = seq(from = min\_year, to = max\_year, by = 2)) +

# add titles and better labels

labs(y = "students",

title = "Incoming international students",

# caption in ggplot is used for source & credits information

caption = "(Data: Nuffic | Adaptation; Author)") +

# another theme

theme\_minimal() +

# caption on the left, label x axis on the right

# and set some font sizes

theme(plot.caption = element\_text(hjust = 0, size = 8),

plot.title = element\_text(size = 10),

axis.title = element\_text(size = 9),

axis.title.x = element\_text(hjust = 1),

legend.title = element\_text(size = 9),

legend.text = element\_text(size = 8))