Introduction to R

Day 2

Sereina Herzog

Institute for Medical Informatics, Statistics and Documentation
Medical University of Graz

10.11.2023

Saving plots in R

Example Iris

Example Iris

```
lab_species <- c("Iris setosa", "Iris virginica", "Iris versicolor")

plot_iris_f <-
  ggplot(data = iris, aes(x = Sepal.Length, y = Sepal.Width, fill = Species)) +
      geom_point(aes(shape = Species), size = 3, alpha = 0.7) +

      scale_fill_manual(values = c("darkred", "blue", "orange"),
                        labels = lab_species) +
      scale_shape_manual(values = c(21, 22, 24),
                         labels = lab_species) +
      labs(title = "Dataset Iris") +
      xlab("Sepal length [cm]") +
      ylab("Sepal width [cm]") +
      coord_cartesian(xlim = c(4, 8), ylim = c(2, 5)) +
      theme_bw() +
      theme(plot.title = element_text(face = "bold", size = 25),
            axis.title = element_text(size = 20),
            axis.text = element_text(size = 14),
            legend.title = element_text(size = 16),
            legend.text = element_text(size = 14))

plot_iris_f
```

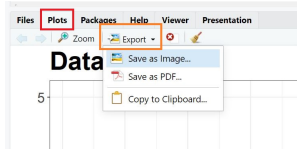
Saving plots

Several possibilities

- plot image in RStudio and work with ‘Plots’ panel (lower right corner)
- save your image in a specific format (e.g., jpeg(), png(), svg(), pdf())
- for ggplot images use ggsave()

Saving plots - ‘Plots’ panel

‘Plots’ panel (lower right corner) → Export → Save as Image or Save as PDF



Saving plots - jpeg()

```
jpeg(filename = "plot_iris_v20231109.jpg",
     width = 12, height = 7, units = "cm", res = 300)

  plot_iris_f

dev.off()
```

- Look at documentation for all options

Saving ggplots

```
ggsave(filename = "plot_iris_ggsave_v20231109.png",
       plot = plot_iris_f,
       units = "cm", width = 12, height = 7)
```

Saving plot - exercise

- Switch to RStudio
- Open Rmd file: *day2_ex1_saveggplot_v20231109.Rmd*
  - is on GitHub in folder 'Course Introduction R 2023/Day2'
- Work through 'Day 2 - Exercise 1'

Rmarkdown - Chunk options

Chunk options in Rmarkdown

- There are several options regarding code chunks, e.g.,
  - **echo** display code in output document (default TRUE)
  - **eval** rund code in chunk (default TRUE)
  - **fig.width** width of plot dimensions in inches (default 7)
  - ...
- The options can be set globally and/or separately for each code chunk
  - globally: use within a code chunk, e.g., *knitr::opts_chunk$set(echo = TRUE)*
  - for one code chunk: within the curly brackets, e.g., ` ``{r, echo = FALSE}`
- See cheat sheet within RStudio
  - Help → Cheat sheets → Rmarkdown Cheat Sheet

Chunk options - exercise

- Make copy of your ‘Day 2 - Exercise 1’ Rmarkdown file and save it as *day2_ex2_chunkoptions_v20231109.Rmd* and try to change the options
  - fig.height
  - fig.width
  - eval
  - echo

Data cleaning with tidyverse

Data table

- each **unit** (e.g. patient, mouse, cell) equals a row
- for each unit the measured **variables** (e.g. age, blood pressure, size) equal columns

id	gender	age	weight	height	smoking
1	1	35	70.5	185	0
2	2	36	65.3	170	0
3	2		90.1	164	1
4	1	21	72.0	177	0
5	1	66	89.4	175	0

Repeated measurements

wide format

id	gender	syst0	syst1
1	1	120	125
2	2	118	125
3	2		110

long format

id	gender	syst	time
1	1	120	0
1	1	125	61
2	2	118	0
2	2	125	60
3	2		
3	2	110	59

What is tidyverse

- tidyverse is a collection of R packages designed for data science
  - they share an underlying design philosophy, grammar, and data structure
    - *ggplot2* for data visualization
    - *readr* for data importation from various file sources
    - *tidyr* and *dplyr* useful for data cleaning
    - ...
  - all core packages can be loaded at once: *library(tidyverse)*
  - 'R for Data Science' (see slide with links)

Useful functions for data cleaning

- **select()** extracts columns and returns a tibble
- **arrange()** changes the ordering of the rows
- **filter()** picks cases based on their values
- **mutate()** adds new variables that are functions of existing variables

What is %>% in Tidyverse?

%>% is used to emphasis a sequence of actions, rather than the object that the actions are being performed on

```
dt_example %>%
  mutate(bmi = weight/(height^2)) %>%
  select(pat_id, sex, bmi)
```

What will we cover

- We will look at
  - importing data (example: .xlsx)
  - useful function for data cleaning
  - save R environment (.Rdata)
- We will work with .Rdata in a Rmarkdown file

Data cleaning - exercise

- Example Glucose:
  - Glucose tolerance was tested by administering 100g glucose drink
  - Glucose was tested before and 1 hour after administering
  - source: R package medicaldata
- Download from GitHub (Course Introduction R 2023/Day2')
  - messy_glucose.xlsx
  - day2_ex2_datacleaning_v20231109.R
- Open R file
- Work through R file (together)

Links

Links

- Introduction to R
  - R for Data Science (https://r4ds.hadley.nz/)
- Plots using ggplot
  - Overview with further links to course material: https://ggplot2.tidyverse.org/
- Display tables using flextable
  - flextable bool https://ardata-fr.github.io/flextable-book/
  - Function references https://davidgohel.github.io/flextable/reference/index.html
- Download R
  - CRAN (https://cran.r-project.org/)
- Download RStudio
  - RStudio Desktop (https://posit.co/download/rstudio-desktop/)

Useful

Useful

- There are links to cheat sheets within RStudio to several topics, among others, about Rmarkdown:
  - Help $\rightarrow$ Cheat sheets $\rightarrow$ Rmarkdown Cheat Sheet