

Hands-on Biological Data Science with R

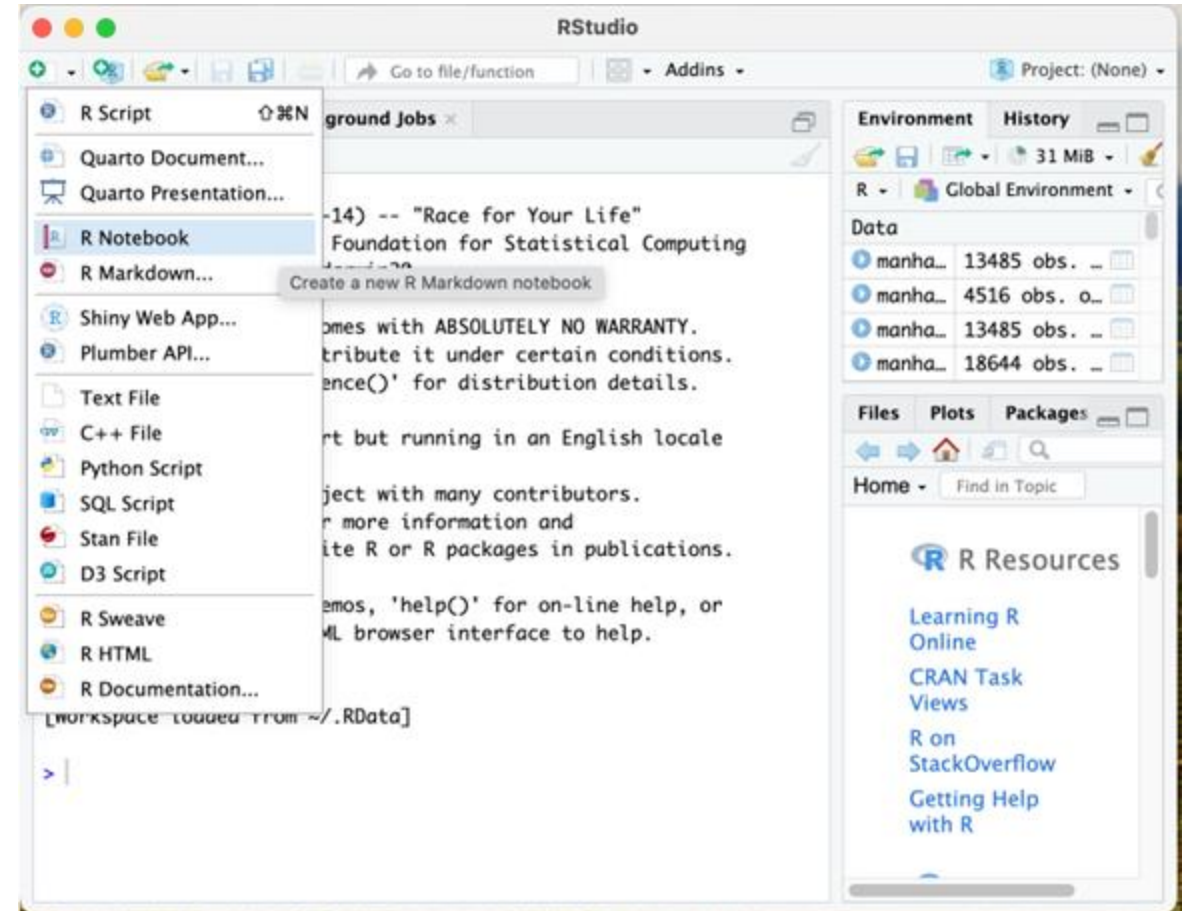
Exercise 1 - First operations in R



Exercise 1

Exercise Upload Format

- Open a new R Notebook in R Studio.
 - The default output is html notebook, don't change that. This means your 'R Notebook' will be saved in two formats, one is Rmd, the format you'll be opening to make edits etc.; and an html format.
- After you are done with it, upload your html file to moodle.
 - Name your html file in the following format **'exercise1_SURNAME_FIRSTNAME'**
 - Make sure it can be opened before you upload.
 - An html file can be opened by any web browser, so either double click and view or, right click and select your browser to open (Chrome, Safari etc.)
- Deadline: **07.11.2025 Fri, 23.59**



Exercise 1

Data

- For our first exercise we use the data from the US Center for Disease Control's Behavioral Risk Factor Surveillance System (BRFSS) annual survey. We are using a small subset of this data, including a random sample of 10000 observations from the years 1990 and 2010.
 - You can check out their [web page](#) for more information.
- Download the required dataset from our med.moodle platform. You can access the dataset using the following link:
<https://www.med.moodle.elearning.lmu.de/mod/resource/view.php?id=155865&forceview=1>

Exercise 1

Part 1 - Loading and Exploring the Data

- Start by loading the dataset and taking a look at its structure.
 - Use a reading function (`read.csv`, `read_csv` etc.) to import the data, and save it in a variable, for example `'brfss'`.
 - To locate the data `'BRFSS-subset.csv'` (which should typically be in your Downloads folder), you can use the `file.choose()` function or if you know your file path, you can use your reading function directly with full path.
 - ```
fname <- file.choose() ## BRFSS-subset.csv
```
    - ```
stopifnot(file.exists(fname))
```
 - ```
brfss <- read.csv(fname)
```
- Explore the data using `head()`, `tail()`, `summary()`, `names()`, `str()`, `class()` etc.

## Exercise 1

### Part 2 - Plotting

- Before you start, make sure to load the necessary libraries in your R environment. (e.g. ggplot2, tidyverse)
- Plot 1:
  - Create a **scatterplot** showing the relationship between the square root of weight ("`sqrt(Weight)`") and height, using `ggplot()`. Color the female and male points differently. In cases of overplotting, making the points slightly transparent can help.
- Plot 2:
  - Plot male and female in different panels using `facet_grid()`.
- Plot 3:
  - Create a subset of the data containing only observations from 2010 and make density curves for male and female groups. Use the fill aesthetic to indicate that each sex is to be calculated separately, and `geom_density()` for the density plot.

## Exercise 1

## References

- 20th Edition of CSAMA Summer School: Biological Data Science. (06/2024, retrieved 10/03/2024, from <https://csama2024.bioconductor.eu/>).

# Thanks for your attention!

## Contacts:

**Ekin Yaman**  
**Kim-Hellmuth**

✉: [Ekin.Yaman@med.uni-muenchen.de](mailto:Ekin.Yaman@med.uni-muenchen.de)

**Dr. med. Paula Rothämel**

✉: [Paula.Rothaemel@med.uni-muenchen.de](mailto:Paula.Rothaemel@med.uni-muenchen.de)

**Dr. med. Sarah**

<https://www.ccrh-hauner.de/kim-hellmuth-labor>