LMU KLINIKUM
Gemeinsam. Fürsorglich. Wegweisend.

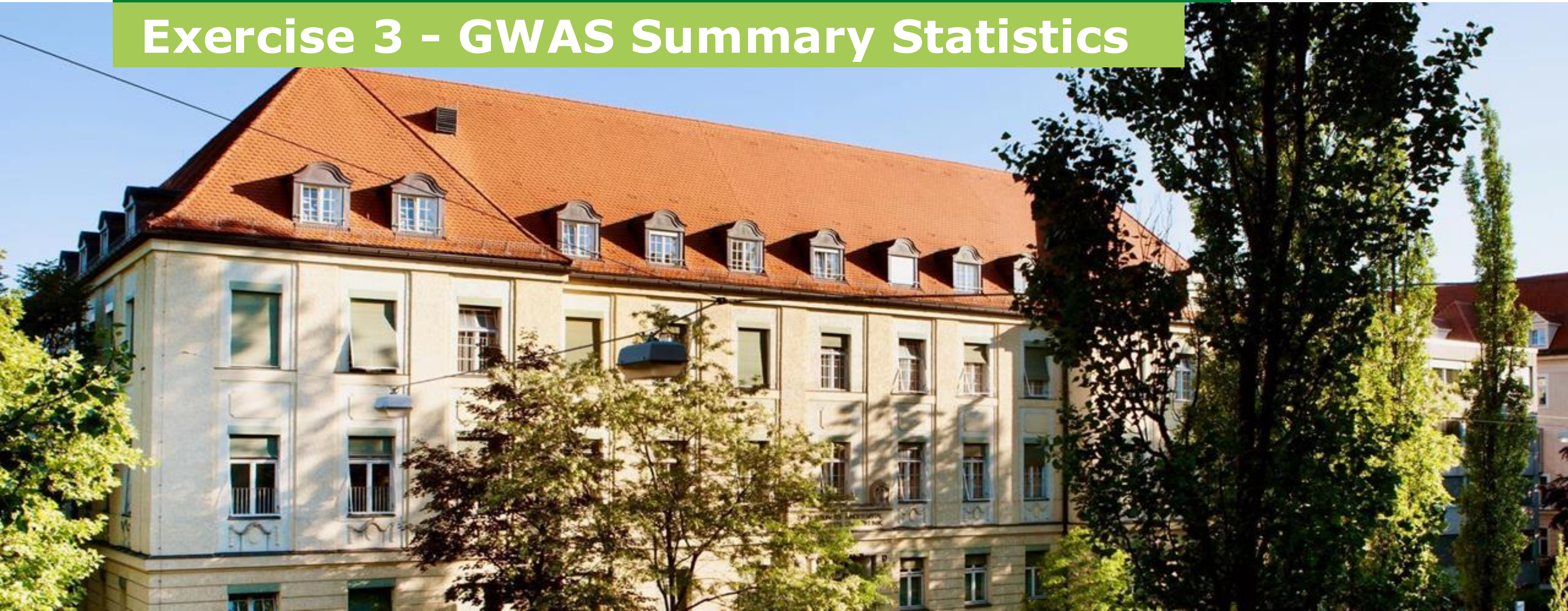
HELMHOLTZ MUNICH

KimHellmuth Lab

DR.VON HAUNERSCHES KINDERSPITAL
Helfen Forschen Heilen. Seit 1846

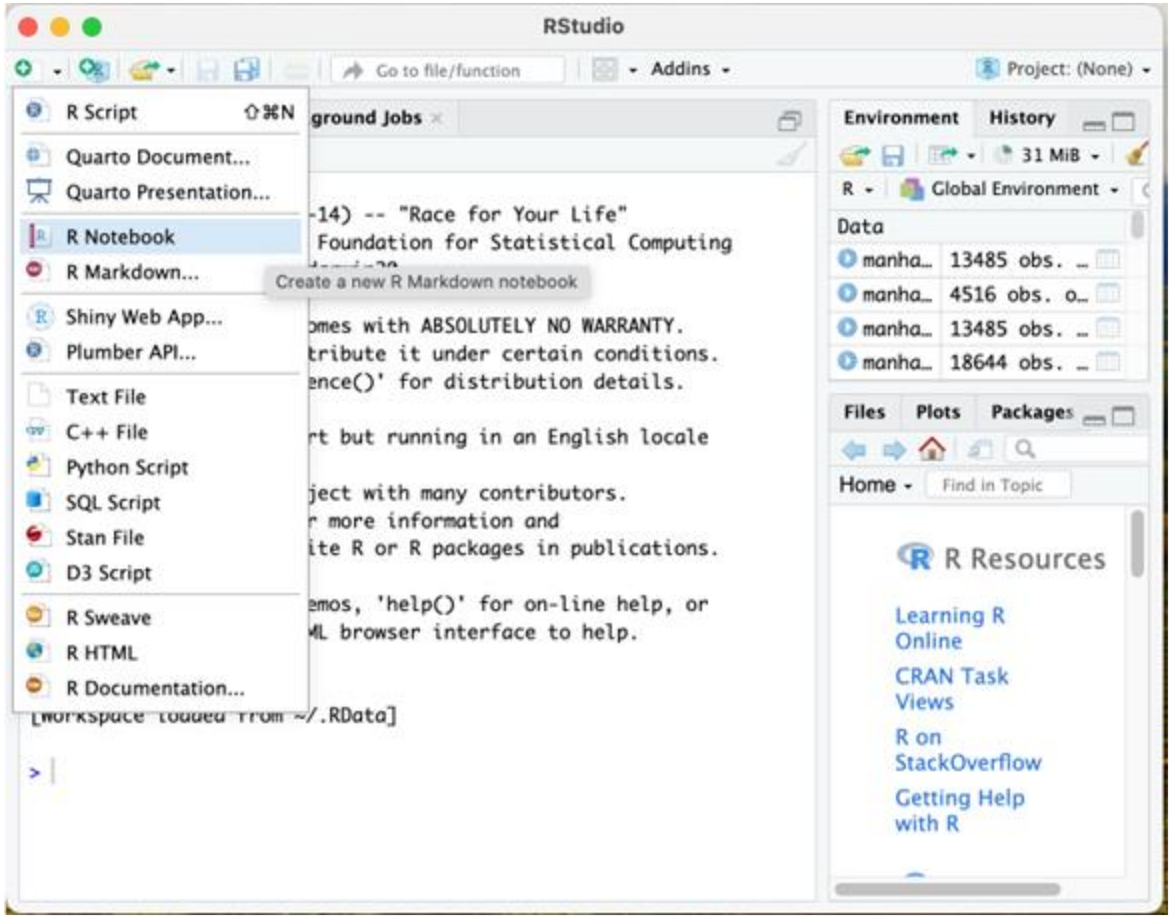
Hands-on Biological Data Science with R

Exercise 3 - GWAS Summary Statistics



Exercise 3
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 **‘exercise3_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.)
- <u>Deadline:</u> **02.01.2026 Fri, 23.59**



```
-14) -- "Race for Your Life"
 Foundation for Statistical Computing

omes with ABSOLUTELY NO WARRANTY.
tribute it under certain conditions.
ence()' for distribution details.

rt but running in an English locale

ject with many contributors.
r more information and
ite R or R packages in publications.

emos, 'help()' for on-line help, or
ML browser interface to help.

[Workspace loaded from ~/.RData]

> 
```

Data	
manha…	13485 obs. …
manha…	4516 obs. o…
manha…	13485 obs. …
manha…	18644 obs. …

Exercise 3
Data

- In this exercise you will use a GWAS summary statistics data computed for height as the phenotype. You can access the data downloading it directly from its source, which is the Uk Biobank's Height GWAS:
  - https://pheweb.org/UKB-Neale/pheno/50
    - We recommend that you explore the webpage after your download as well.
- Due to data size being large, to be able to process the data more easily, you can apply a p-value filtering to the data before do your operations.

Exercise 3

Part 1 - Data Manipulation

- **Format the Data**
  - Explore the data you've read into your notebook (fread is faster with large datasets compared to read_csv) in the usual way
  - In this specific data, you have allele count (ac) column instead of a allele frequency column. Create an allele frequency column. (You can infer the necessary information from the data source)
- **The top hit and the top locus**
  - Look at the top hit of the study (lowest p-value).
    - Print the top SNV's position, rsid, p-value, beta, and allele frequency.
  - Extract the top SNV and every other SNV positionally near the top SNV in a +/-250kb range.

Exercise 3
Part 2 - Visualization

- **Manhattan Plot**
  - Create a manhattan plot and highlight your top SNV and every other SNV near your top SNV’s 250kb vicinity **(Plot 1)**
    - You can use (‘qqman’) package or any other method (ggplot, baseR) for this.
  - In a second manhattan plot zoom in to the +/- 1mb region of your top SNV and show the nearest gene’s (to your top SNV) position in the same plot. **(Plot 2)**
- **QQ Plot**
  - Create a qqplot. **(Plot 3)**
- **Locus Zoom**
  - Lastly, create a locus zoom plot of the selected snv +/- 250kb region. **(Plot 4)**
    - You can create the plot in r using ‘locuszoomr’ package.
    - Or you can export your dataset that includes the snvs from the selected region and upload them to the Locuszoom website.
  - If you choose the second (web) version, don’t forget to import the resulting image to your Rnotebook and upload it alongside with your html file at the end.

Exercise 3
Sources

- Here are some sources that you may use to create your plots.
  - https://cran.r-project.org/web/packages/qqman/vignettes/qqman.html
  - https://r-graph-gallery.com/101_Manhattan_plot.html
  - https://bernatgel.github.io/karyoploter_tutorial/Tutorial/PlotManhattan/PlotManhattan.html

LMU KLINIKUM
Gemeinsam. Fürsorglich. Wegweisend.

HELMHOLTZ MUNICH

KimHellmuth Lab

DR. VON HAUNERSCHES KINDERSPITAL
Helfen Forschen Heilen. Seit 1846

Thanks for your attention!

Contacts:

Ekin Yaman
✉: Ekin.Yaman@med.uni-muenchen.de

Dr. med. Paula Rothämel
✉: Paula.Rothaemel@med.uni-muenchen.de

Dr. med. Sarah Kim-Hellmuth
https://www.ccrc-hauner.de/kim-hellmuth-labor