## **Statistics II – Week 1: What is exploratory data analysis?**

Date:

Group A 6 September 2022 10:45 – 12:30

Group B 6 September 2022 15:15 – 17:00

Group C 5 September 2022 10:45 - 12:30

**Preparation – Lecturer and TAs**

To do before start of term:

* Ensure all team members and students can access to the materials. For instance, check and help all students to set up R environment at their computers.
* Discuss who will lead which class and who will support.
* Ensure all team members and students have received invites for classes.
* Agree ahead of time how staff will communicate with each other during live sessions.
* Enable room for all work sessions (contingency for internet issues)
* Adjust course settings in the Brightspace as needed.

To do 1-2 days before session:

* Review this document if necessary.
* Check discussion forums for questions and pick out recurrent themes on Brightspace to address them in the live session.
* Check other communication channels (emails from students) for questions related to live session and weekly contents.
* Set up any polls in Mentimenter.
* Nominate TA as host to manage the logistics, Mentimeter, updated and publications on Brightspace, R Markdown, support office hours, assignments deadlines whoever is supporting teaching sessions should be assigned to specific students groups (A, B, and C).
* Upload PDF of slides before session.

Notes

* If wanting to use Mac/Windows to work in R environment during the session, ensure this is considering during set it up individual computers.

**Outline of the session**

*Preamble/Housekeeping*

**Actions:**

* Lecturer/TAs managing the session.
* Start live session by following instructions and verifying computers, internet, and video are properly working.
* Note number of students attending.

Lecturer/TA introduce themselves – see slides for script: "Preparing data for analysis using R - Week 1"

*Course overview*

Lecturer outlines Statistics II Syllabus, assessments and explains ‘house rules’ for sessions and discussion - expectation setting.

* Poll [TA please set up beforehand]
  + Have you had a chance to read through the Syllabus?
    - Yes, No, Partly, Not sure
  + Have you had a chance to install R in your computer?
    - Yes, No, Partly, Not sure
  + Have you had any R experience before this course?
    - From zero (less) to ten (more experience).
  + What your thoughts about statistics as a field apply to your live or future work? (Open question)
* Show poll results.
* Give links to instructions, videos for those who haven’t installed R yet – they can catch up during class if needed. Here are some useful links for this session:

### 

### Download and Install R: <https://cran.r-project.org./>

# How to install R: <https://r-coder.com/install-r/>

# Install R and RStudio – A Step-by-Step Guide for Beginners: <https://www.datacamp.com/tutorial/installing-R-windows-mac-ubuntu>

Mentimeter link for this session: **(update this)**

Or go to www.menti.com and use the joining code **(update this)**

*Material*

Open code titled: "Preparing data for analysis using R - Week 1"

Datasets: “gapminder” and “childhood inequality”.

**Actions**

* Students will now do 45 minutes of importing data in text or csv format following the instruction in the scrip titled: "Preparing data for analysis using R - Week 1".
* Check on student progress, give suggestions and answer any questions they might have imported data and exploring data structure.
* Keep an eye on the main room in case any students join late; explain the task and help them to reach out the rest of the class: installing R and getting data into a usable format.
* Make note of the questions the students are mapping and the themes they have picked out, to aid in tying the session together during lab, office hours or next class feedback.
* **To remind to students at 35 mins they have 10 mins left.**

**If there’s time, Lecturer present a final example preparing datasets, visualizing data, and explore basic statistics in the data and brief catchup with teaching team.**

**Actions:**

* Add link Course resources (TA) and share them through Brightspace.
* TA check participation and absents.
* Lecture has a brief catchup with teaching team.

These online textbooks cover a wide range of topics in R data processing using the tidyverse:

* Grolemung G & Wickham H, R for Data Science (https://r4ds.had.co.nz/) (2017)
* Wickham H, ggplot2: Elegant Graphics for Data Analysis (https://ggplot2-book.org) (2020)
* Chang W, R Graphics Cookbook (https://r-graphics.org) (2019)
* Ismay C & Kim AY, Statistical Inference via Data Science (https://moderndive.com) (2019)
* Sanchez G, Handling Strings with R (https://www.gastonsanchez.com/r4strings/) (2016)

Other web resources:

* The RStudio cheat sheets (https://rstudio.com/resources/cheatsheets/) are great for a quick overview of the tidyverse functions.
* The tidyverse overview (https://tidyverse.tidyverse.org) provides links to further information on each individual package.
* The RStudio team also provide some suggested learning paths (https://education.rstudio.com/learn/) for beginner, intermediate and expert R users.

Acknowledgements

All data used in this course are taken from <https://www.gapminder.org/data/>, U.S. Nutrition and Food Survey, and National Aging Survey MHSP Colombia.