Final exercise instructions

(For the advanced group)

1

Build an exercise with these specs

- This should be an exercise that takes 15-20 minutes
  - Aimed at a beginner-intermediate user
- Using a small dataset, contained in a csv file (or multiple files). Anonymized data.
- The exercise should combine 2-3 elements of the following:
  - Generating plots with ggplot2
  - Using tidyverse (packages dplyr, tidyr, e.g.: mutate_*, summarize_*, group_by, *_join)
  - Building custom functions
  - Non standard evaluation (tidyverse style functions)
  - Unsupervised learning (e.g., clustering)
  - Supervised learning (e.g., regression, forests, etc.)
  - Purrr
- Also include one question that shows “understanding” (and not just coding)

Example topics

- The exercise should **combine 2-3 elements** of the following:
  - Plots with ggplot2
  - Using tidyverse (packages dplyr, tidyr, e.g.: mutate_*, summarize_*, group_by, *_join, gather/spread)
  - Building custom functions
  - Non standard evaluation (tidyverse style functions)
  - factors or strings or dates (forcats, stringr, lubridate)
  - Unsupervised learning (e.g., clustering)
  - Supervised learning (e.g., regression, forests, etc.)
  - Purrr
- Also include one question that the learner needs to show “understanding” (and not just coding)

How to build the exercise

- The exercise should be compiled in groups of 2-3
- An exercise should contain 3-6 sub-questions
- In sub-questions where code is needed, provide “faded examples” (partial code with “complete the blanks”)
- Provide the questions in an Rmarkdown file or a script file
- In a separate file provide the answers
- Provide an example anonymized data file
- Add documentation in the file indicating the group participants
- Files should be zipped together (all three files as a single zip file)