Final Exam

Marketing Analytics Intro R

Submission Instructions:

- → Submission is due by **24 October 2021**, 23:59 by e-mail to winkler@wiwi.uni-frankfurt.de
- → Hand in **1** annotated and structured R script file (= text file with file extension .R) named **LastName_FirstName.R**
- → Delete or comment out "annoying" lines of code
  - E.g., `#setwd(...)`, `#install.packages()`, `#View()`
- → Only include the name of the data file (not your file path) when importing the data
  - E.g., `d <- read.csv2("Nameofthedataset.csv")`
- → Add comments (with #) to your code to answer questions if necessary
- → The code must run in R without producing an error message!

Task 1: Creating Certificate Grades

Import the file `certificates.txt` into R and check if all data have been read correctly. The file contains the (randomly generated) scores for examinations of 10 students who have taken three modules from the Marketing Analytics Master (COIN, REMA, ACVM, maximum points 90 points each).

1. Create a new variable that contains the mean of the score of the three modules.
2. Calculate the 20%, 40%, 60% and 80% quantile for the newly created variable. (Hint: read the help page for the `quantile()` function)
3. Now assign the grades A-F for each student on the basis of the average score according to the following scheme. Save the grades as a new variable.
   - → better than 80% of students: A
   - → better than 60% of students: B
   - → better than 40% of students: C
   - → better than 20% of students: D
   - → is one of the 20% worst students: F
4. Now sort the dataset permanently by last names and first names of the students (in alphabetical order).

Task 2: Flirting

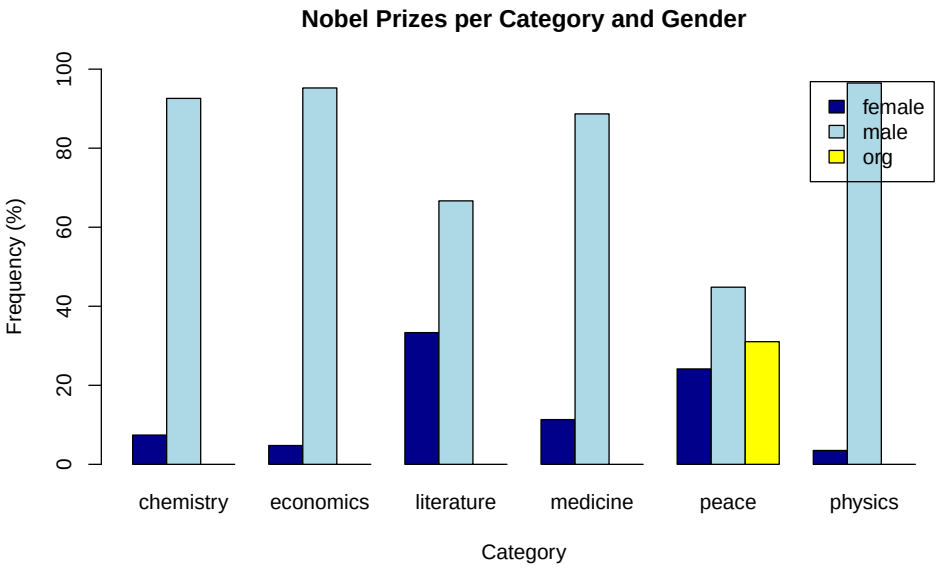
The following is an (imaginary) dataset that contains data on chat-up lines (flirting) during partying (see Field, 2012, p. 347). The chat-up lines of 535 revelers were recorded (example: * I may not be Fred Flintstone, but I could make your bed rock! *). These chat-up lines were then rated on three dimensions by independent raters. The data for this is in the dataset `Chat-Up-Lines_Person.txt`. It was also observed if the chat-up line was successful. The data for this (along with an indication of whether the person being addressed was male or female) is in the `Chat-Up-Lines_Success.txt` record.

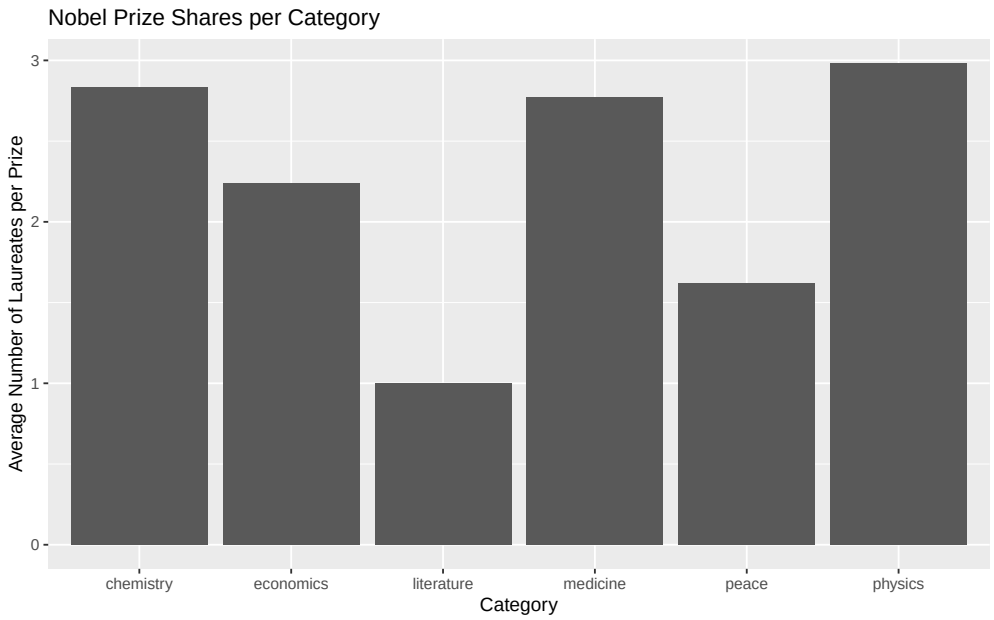
1. Merge the two datasets `Chat-Up-Lines_Person.txt` and`Chat-Up-Lines_Success.txt` using the key variable `id`. The dataset should then contain the following variables (also note that factors are encoded as such in R):
   - → id (person variable)
   - → Funny (Rating of the wittiness of the chat-up line, scale 0 = not funny to 10 = the funniest I've ever heard)
   - → Sex (rating of suggestiveness of the chat-up line, scale 0 = not suggestive to 10 = very suggestive)
   - → Gender: Gender of the person who was "chatted up" (0 = female, 1 = male)
   - → Success: Success of the chat-up line (0 = no response, 2 = phone number)
2. Answer the following questions:
   - → Who is flirted with more (in percentage) - men or women?
   - → Who is addressed with funnier chat-up lines on average - men or women?
   - → Who is approached with more suggestive chat-up lines on average - men or women?
   - → Create a contingency table with the gender as rows and the success of the chat-up lines as columns. Get the (appropriate) relative frequencies (row or column percentages) to answer the question which gender the chat-up lines were more successful at (i.e., the number of phone numbers received more frequently).

Task 3: Nobel Prizes

Import the data of the Nobel Prize committee on all Nobel Prize laureates into R by using the following command: `nobel <- read.csv("http://api.nobelprize.org/v1/laureate.csv")` (Attention: Internet connection is required). Then select only winners from 2000 onwards.

1. Make a (suitable) plot which indicates the distribution of Nobel Prizes per gender (male, female, organization) in each Nobel Prize category (absolute or relative frequencies). The plot could look something like the first picture on the next page.
2. Make a (suitable) plot using the `ggplot2` package which shows how many people share the Nobel Prize on average in each Nobel Prize category (variable `share`). The plot could look something like the second picture on the next page.


Nobel Prizes per Category and Gender
Frequency (%)
0
20
40
60
80
100
female
male
org
chemistry
economics
literature
medicine
peace
physics
Category


Nobel Prize Shares per Category
Average Number of Laureates per Prize
0
1
2
3
chemistry
economics
literature
medicine
peace
physics
Category