Homework 2
of
STAT 3355 Data Analysis for Statisticians & Actuaries

Due: 11:30 am

February 21 (Monday), 2022

Problem 1 (2 points)

Rewrite each code block to comply with the "Homework and Project Code Style Guide"

(a)
```
mat <- matrix( c( 34, 23, 53, 6, 78, 93, 12, 41, 99 ) ,nrow
    = 3)
df <- as.data.frame (mat)
names( df ) <- c("score_given_to_car_on_driving_test",
 "score.given.to.van.on.driving.test",
 "score-given-to-truck-on-driving-test")
```

(b)
```
library( ggplot2 )
  head(mpg)
second_version_of_mpg <- mpg[ mpg$cyl = 6,]
second_version_of_mpg$class <- as.character(second_version_
   of_mpg$class$class)
```

Problem 2 (5 points)

Download the "Teaching Assistant Evaluation Data Set" dataset on the UCI Machine Learning Repository. The link is archive.ics.uci.edu/ml/datasets/Teaching+Assistant+Evaluation. Read the data in its original format (.data) by using the function `read.table()` or `read.csv()` in an appropriate way and rename each variable according to the web page.

In this dataset, each of the 151 observation corresponds to a unique teaching assistant (TA), so create a variable of TA identification (ID) number and assign an ID number from 1 to 151 to all TAs sequentially. In addition, for simplicity (although it may not be true in this case), assume each class can only have one TA at a time. Therefore, each of the 151 observation corresponds to a unique course at a time. If you see multiple observations share the same instructor ID and course ID, that probably means that the courses occurred in different year or semester.

(a) Turn the first variable (whether of not the TA is a native English speaker) into a logical variable, where `TRUE` corresponds to a native English speaker, and `FALSE` otherwise

(b) Turn the fourth variable (summer or regular semester) into a logical variable, where `TRUE` corresponds to regular, and `FALSE` corresponds to summer

(c) Turn the last variable (class attribute or evaluation score) into an ordered factor variable with levels labeled as 'low', 'medium' and 'high'

Hint: You should get the following response (other than variable names) after applying the `str()` function on the cleaned dataset

```
'data.frame':  151 obs. of  6 variables:
 $ eng_speaker  : logi  TRUE FALSE TRUE TRUE FALSE FALSE
   ...
 $ instructor_id: int  23 15 23 5 7 23 9 10 22 15 ...
 $ course_id    : int  3 3 3 2 11 3 5 3 3 3 ...
 $ regular      : logi  FALSE FALSE TRUE TRUE TRUE FALSE
   ...
 $ size         : int  19 17 49 33 55 20 19 27 58 20 ...
 $ score        : Factor w/ 3 levels "low","medium",..: 3 3
   3 3 3 3 3 3 3 3 ...
 $ ta_id        : int  1 2 3 4 5 6 7 8 9 10 ...
```

(d) What is the average and median class size in regular semester? What are those two numbers in summer semester? Round your numeric answer to 2 decimal places.

(e) How many native English speaker TAs are there in regular and summer semester, respectively? What are those two numbers for non native English speaker TAs?

Problem 3 (2 points)

Read "Coping with Hitchhikers and Couch Potatoes on Teams," and write 3 – 4 sentence on how you feel the paper relates to your previous experience in group projects and what you will do to ensure you don't have the same problems with the team projects in this class.

Problem 4 (1 point)

Fill out the team info sheet as below:

(a) Team name:

(b) Team member info

(c) When is the next meeting?

	Name	Major	Class standing
Member 1			
Member 2			
Member 3			