Homework2

2024-02-14

Question 1

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)
data(mpg)
head(mpg)
```

```
## # A tibble: 6 x 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl    class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr> <chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p     compa~
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p     compa~
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p     compa~
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p     compa~
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p     compa~
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p     compa~
```

```
second_version_of_mpg <- mpg[mpg$cyl == 6, ]
second_version_of_mpg_class <- as.character(mpg$class)
```

Question 2

```
setwd("/Users/adi/Desktop/dataverse_files")
senate <- read.csv("1976-2020-senate.csv", header = TRUE, sep = ",")
```

a)

```
senate$year <- factor(senate$year)
senate$state <- factor(senate$state)
senate$party_simplified <- factor(senate$party_simplified)
```

b)

```
texas_data <- senate[senate$state == "TEXAS", c("year",
                                                "state",
                                                "candidatevotes",
                                                "totalvotes",
                                                "party_simplified")]
```

c)

```
avgdem <- round(mean(
                texas_data$candidatevotes[texas_data$party_simplified == "DEMOCRAT"]))
avgrep <- round(mean(
                texas_data$candidatevotes[texas_data$party_simplified == "REPUBLICAN"]))
avglib <- round(mean(
                texas_data$candidatevotes[texas_data$party_simplified == "LIBERTARIAN"]))
avgoth <- round(mean(
                texas_data$candidatevotes[texas_data$party_simplified == "OTHER"]))

meddem <- round(median(
                texas_data$candidatevotes[texas_data$party_simplified == "DEMOCRAT"]))
medrep <- round(median(
                texas_data$candidatevotes[texas_data$party_simplified == "REPUBLICAN"]))
medlib <- round(median(
                texas_data$candidatevotes[texas_data$party_simplified == "LIBERTARIAN"]))
medoth <- round(median(
                texas_data$candidatevotes[texas_data$party_simplified == "OTHER"]))
```

d)

```
dem_win <- texas_data$year[
           texas_data$partysimplified == "DEMOCRAT" & texas_data$candidatevotes ==
           max(texas_data$candidatevotes)]
```

Question 3

```
setwd("/Users/adi/Desktop/teaching+assistant+evaluation")
tae <- read.csv("tae.data", header = TRUE, sep = ",")
names(tae) <- c("English",
                "Instructor",
                "Course",
                "Semester",
                "Size",
                "Attribute")
tae$ta_id <- c(1:150)
```

a)

```
english_speak <- which(tae$English == "1")
tae$English[english_speak] <- TRUE
english_nospeak <- which(tae$English == "2")
tae$English[english_nospeak] <- FALSE
tae$English <- as.logical(tae$English)
```

b)

```
regular_sem <- which(tae$Semester == "2")
summer_sem <- which(tae$Semester == "1")
tae$Semester[regular_sem] <- TRUE
tae$Semester[summer_sem] <- FALSE
tae$Semester <- as.logical(tae$Semester)
```

c)

```
low_class <- which(tae$Attribute== "1")
tae$Attribute[low_class] <- "low"
med_class <- which(tae$Attribute == "2")
tae$Attribute[med_class] <- "medium"
high_class <- which(tae$Attribute == "3")
tae$Attribute[high_class] <- "high"
tae$Attribute <- factor(tae$Attribute, levels = c("low","medium","high"))
```

d)

```
avg_reg_sem <- round(
               mean(
               tae$Size[tae$Semester == TRUE]), 2)
avg_summer_sem <- round(
                  mean(
                  tae$Size[tae$Semester == FALSE]), 2)
median_reg_sem <- round(
                  median(
                  tae$Size[tae$Semester == TRUE]),2)
median_summer_sem <- round(
                     median(
                     tae$Size[tae$Semester == FALSE]),2)
```

e)

```
native_eng_reg <- length(tae$English[tae$English == TRUE & tae$Semester == TRUE])
native_eng_sum <- length(tae$English[tae$English == TRUE & tae$Semester == FALSE])
non_native_reg <- length(tae$English[tae$English == FALSE & tae$Semester == TRUE])
non_native_sum <- length(tae$English[tae$English == FALSE & tae$Semester == FALSE])
```

f)

```
total_eng <- sum(tae$English == TRUE)
high_eng <- sum(tae$English == TRUE & tae$Attribute == "high")
proportion_eng <- round((high_eng / total_eng), 2)
total_no_eng <- sum(tae$English == FALSE)
high_no_eng <- sum(tae$English == FALSE & tae$Attribute == "high")
proportion_no_eng <- round((high_no_eng / total_no_eng), 2)
```

Question 4

After reading "Coping with Hitchhikers and Couch Potatoes on Teams," I've come to realize that in the past I've allowed myself to be taken advantage of by hitchhikers and couch potatoes. When faced with one of those people, I would often take it upon myself to do their work for them, to avoid getting a bad grade. I've never gone to a professor or supervisor with these problems because I felt that I could do the project myself. Reading this has made me realize that I don't have to do that in this class, and I can go to the professor if I feel that I'm being left with all the work.