STAT 3355 - HW 2 - Yebom Kim

2024-02-14

Problem 1

Problem 1 - (a)

```r
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
```

Problem 1 - (b)

```r
##library(ggplot2)
##head(mpg)
##second_version_of_mpg <- mpg[mpg$cyl == 6,]
##second_version_of_mpg$class <- as.character(mpg2$class)
```

Problem 2

```r
setwd("/Users/springkim/Downloads/dataverse_files")
senate_data <- read.csv("1976-2020-senate.csv")
```

Problem 2 - (a)

```r
senate_data$year <- factor(senate_data$year)
senate_data$state <- factor(senate_data$state)
senate_data$party_simplified <- factor(senate_data$party_simplified)
```

Problem 2 - (b)

```r
texas_data <- subset(senate_data, state == "TEXAS", select = c(year, state, candidatevotes, totalvotes,
```

Problem 2 - (c)

```r
#Democratic
dem <- subset(texas_data, party_simplified == "DEMOCRAT")
avgDem <- round(mean(dem[, "candidatevotes"]), 0)
medianDem <- round(median(dem[, "candidatevotes"]), 0)
cat("The average of democratic is: ", avgDem, "\n")
```

```
## The average of democratic is:  2416258
```

```r
cat("The median of democratic is: ", medianDem, "\n")
```

```
## The median of democratic is:  2112490
```

```r
#Republican
rep <- subset(texas_data, party_simplified == "REPUBLICAN")
avgRep <- round(mean(rep[, "candidatevotes"]), 0)
medianRep <- round(median(rep[, "candidatevotes"]), 0)
cat("The average of republican is: ", avgRep, "\n")
```

```
## The average of republican is:  3019937
```

```r
cat("The median of republican is: ", medianRep, "\n")
```

```
## The median of republican is:  2761660
```

```r
#Libertarian
lib <- subset(texas_data, party_simplified == "LIBERTARIAN")
avgLib <- round(mean(lib[, "candidatevotes"]), 0)
medianLib <- round(median(lib[, "candidatevotes"]), 0)
cat("The average of libertarian is: ", avgLib, "\n")
```

```
## The average of libertarian is:  92815
```

```r
cat("The median of libertarian is: ", medianLib, "\n")
```

```
## The median of libertarian is:  72657
```

```r
#Other Candidates
othercan <- subset(texas_data, party_simplified == "OTHER")
avgothercan <- round(mean(othercan$candidatevotes), 0)
medianothercan <- round(median(othercan$candidatevotes), 0)
cat("The average of other candidates is: ", avgothercan, "\n")
```

```
## The average of other candidates is:  21533
```

```r
cat("The median of other candidates is: ", medianothercan, "\n")
```

```
## The median of other candidates is:  4564
```

Problem 2 - (d)

```r
democratic_wins <- as.character((texas_data[[c("year")]]))
cat("Years Democrat Texan Won: ", (unique(as.numeric(democratic_wins))))
```

```
## Years Democrat Texan Won:  1976 1978 1982 1984 1988 1990 1994 1996 2000 2002 2006 2008 2012 2014 2018
```

Problem 3

```r
data <- read.csv("/Users/springkim/Downloads//teaching+assistant+evaluation/tae.data", header = FALSE,
```

Problem 3 - (a)

```r
#data$TA_ID <- 1:nrow(data)
#data$English_speaker <- data$English_speaker == 1
#data$Semester <- data$Semester == 2
#data$Class_attribute <- factor(data$Class_attribute,
# levels = c("low", "medium", "high"), ordered = TRUE)
```

Problem 3 - (b)

```r
data[, "Semester"] <- data[, "Semester"] == 2
```

Problem 3 - (c)

```r
data[, "Rating"] <- factor(data[, "Rating"], levels = 1:3,
labels = c("low", "medium", "high"), ordered = TRUE)
```

Problem 3 - (d)

```r
regular_sem <- subset(data, Semester == TRUE)
summer_sem <- subset(data, Semester == FALSE)
cat("Regular semester average: ", round(mean(regular_sem[, "Size"]), digits = 2), "\n")
```

```
## Regular semester average:  29.34
```

```r
cat("Regular semester median: ", round(median(regular_sem[, "Size"]), digits = 2), "\n")
```

```
## Regular semester median:  29
```

```r
cat("Summer semester average: ", round(mean(summer_sem[, "Size"]), digits = 2), "\n")
```

Summer semester average: 19.7

```r
cat("Summer semester median: ", round(median(summer_sem[, "Size"]), digits = 2), "\n")
```

Summer semester median: 20

Problem 3 - (e)

```r
cat("English TA during regular semester:",
    nrow(subset(regular_sem, English == TRUE)), "\n")
```

English TA during regular semester: 20

```r
cat("English TA during summer semester:",
    nrow(subset(summer_sem, English == TRUE)), "\n")
```

English TA during summer semester: 9

```r
cat("Non-native English TA during regular semester:",
    nrow(subset(regular_sem, English == FALSE)), "\n")
```

Non-native English TA during regular semester: 0

```r
cat("Non-native English TA during summer semester:",
    nrow(subset(summer_sem, English == FALSE)), "\n")
```

Non-native English TA during summer semester: 0

Problem 3 - (f)

```r
native_english <- subset(data, English == TRUE)
cat("Total Native English speaker: ",
nrow(native_english), "\n")
```

Total Native English speaker: 29

```r
cat("High Score Native Proportion: ",
round(mean(native_english[, "Rating"] == "high", na.rm = TRUE),
digits = 2), "\n")
```

High Score Native Proportion: 0.62

```
nonnative_english <- subset(data, English == FALSE)
cat("Total Non-native English speaker: ",
nrow(nonnative_english), "\n")
```

Total Non-native English speaker: 0

```
cat("High Score Non-native Proportion: ", round(mean(nonnative_english[, "Rating"] == "high", na.rm = T
digits = 2), "\n")
```

High Score Non-native Proportion: NaN

Problem 4

```
# About my previous team project experiences, thankfully all my teammates
#I have met are really passionate and participatory so that I have never met
#couch potatoes or hitchhikers. After reading this article and if
#I met couch potatoes or hitchhikers during the team project,
#I'm not absorb them and I'll do the mirroring. I also thought
#I shouldn't be that kind of person.
```