

# mKhan\_Homework2

2023-09-16

#Problem 1a

```
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")
```

#Problem 1b

```
library(ggplot2)
data(mpg)
#second_version_of_mpg <- mpg[mpg$cy2 == 6 ,]
```

#Problem 2

```
#Part a
senate_data <- read.csv("1976-2020-senate.csv", colClasses = c("year" = "factor", "state" = "factor", "party" = "factor", "candidatevotes" = "numeric", "totalvotes" = "numeric"))

#Part b
# Create a subset from main dataset extracting information from Texas with specified columns
texas_subset <- subset(senate_data, state == 'TEXAS', select = c('year', 'state', 'candidatevotes', 'totalvotes'))

#Part c
dem <- subset(texas_subset, party_simplified == 'DEMOCRAT')
round(mean(dem$candidatevotes), digits = 0)

## [1] 2416258
median(dem$candidatevotes)

## [1] 2112490
rep <- subset(texas_subset, party_simplified == 'REPUBLICAN')
round(mean(rep$candidatevotes), digits = 0)

## [1] 3019937
median(rep$candidatevotes)

## [1] 2761660
lib <- subset(texas_subset, party_simplified == 'LIBERTARIAN')
round(mean(lib$candidatevotes), digits = 0)

## [1] 92815
median(lib$candidatevotes)

## [1] 72657
```

```
other <- subset(texas_subset, party_simplified == 'OTHER')
round(mean(other$candidatevotes), digits = 0)
```

```
## [1] 21533
```

```
median(other$candidatevotes)
```

```
## [1] 4564
```

```
#Part d
```

```
dem$year[which(dem$candidatevotes > dem$totalvotes)]
```

```
## factor()
```

```
## 24 Levels: 1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 ... 2021
```