

Assignment 01

```
library(tinytex)
indata <- read.csv("D:/RStudio/File/STATS413/vwjetta_gas.csv")
indata$mileage[indata$mileage == 999999] <- NA
indata$vehage[indata$vehage == 999] <- NA
indata$mpg[indata$mpg == 99] <- NA

indata$congrade <- factor(indata$congrade)
levels(indata$congrade) <- c("rough", "average", "clean", "excellent")
```

Question a

```
options(width = 300)
summary(indata)
```

##	resratio	mileage	vehage	mpg	congrade	modelyr
##	Min. :0.0388	Min. : 1266	Min. : 1.00	Min. :22.00	rough :578	Min. :1997
##	1st Qu.:0.1982	1st Qu.: 30578	1st Qu.: 32.00	1st Qu.:23.00	average :378	1st Qu.:2003
##	Median :0.3425	Median : 63744	Median : 64.00	Median :25.00	clean :209	Median :2005
##	Mean :0.4224	Mean : 64398	Mean : 65.88	Mean :24.06	excellent: 31	Mean :2005
##	3rd Qu.:0.6378	3rd Qu.: 95131	3rd Qu.: 97.00	3rd Qu.:25.00		3rd Qu.:2008
##	Max. :1.0535	Max. :149769	Max. :164.00	Max. :27.00		Max. :2011
##		NA's :5	NA's :8	NA's :6		

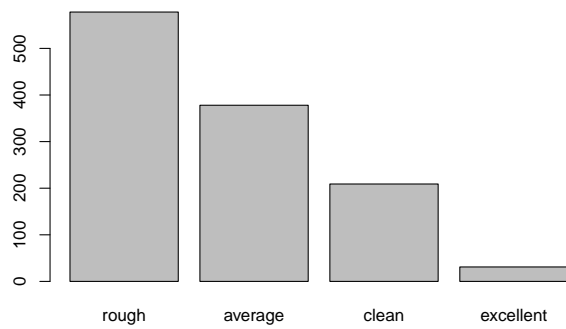
- 5 values are missing in **mileage**
- 8 values are missing in **vehage**
- 6 values are missing in **mpg**

Question b

```
table(indata$congrade)
```

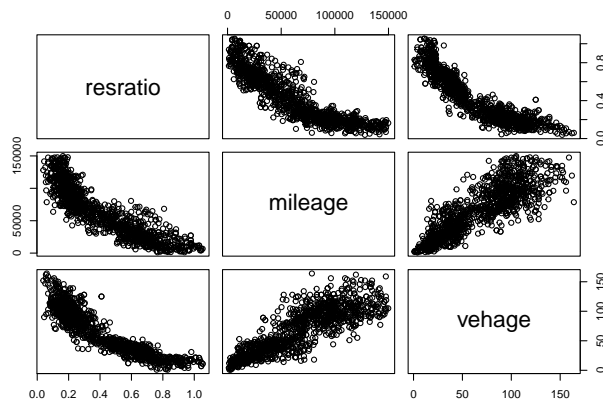
```
##
##    rough    average    clean excellent
##      578      378      209        31
```

```
plot(indata$congrade)
```



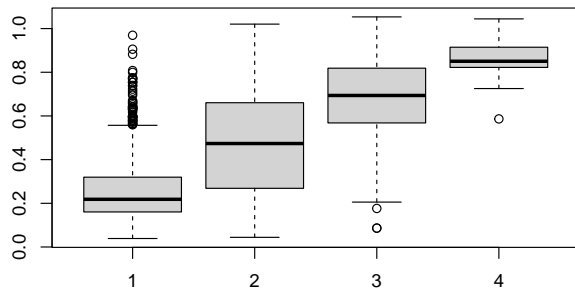
Question c

```
knitr::opts_chunk$set(fig.width = 6, fig.height = 4)
pairs(indata[,c("resratio", "mileage", "vehage")])
```



Question d

```
options(width = 300)
rough_vehicles <- indata[indata$congrade == "rough",]
average_vehicles <- indata[indata$congrade == "average",]
clean_vehicles <- indata[indata$congrade == "clean",]
excellent_vehicles <- indata[indata$congrade == "excellent",]
knitr::opts_chunk$set(fig.width = 6, fig.height = 4)
boxplot(rough_vehicles$resratio, average_vehicles$resratio, clean_vehicles$resratio, excellent_vehicles$resratio)
```



```
mean_resratio_rough <- mean(rough_vehicles$resratio)
mean_resratio_average <- mean(average_vehicles$resratio)
mean_resratio_clean <- mean(clean_vehicles$resratio)
mean_resratio_excellent <- mean(excellent_vehicles$resratio)
cat("mean resratio of rough:", mean_resratio_rough)
```

```
## mean resratio of rough: 0.2722137
```

```
cat("mean resratio of average:", mean_resratio_average)
```

```
## mean resratio of average: 0.4774244
```

```
cat("mean resratio of clean:", mean_resratio_clean)
```

```
## mean resratio of clean: 0.6732804
```

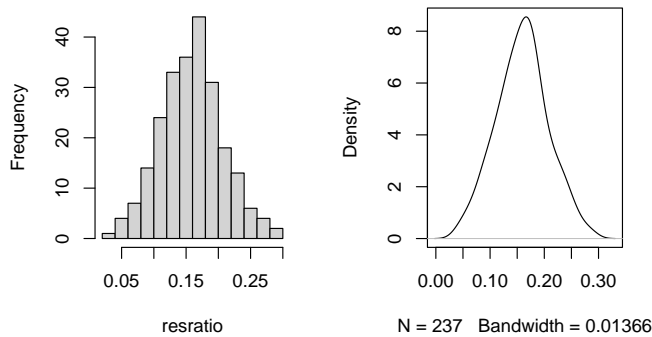
```
cat("mean resratio of excellent:", mean_resratio_excellent)
```

```
## mean resratio of excellent: 0.8596422
```

Question e

```
options(width = 300)
mileage_limit <- indata[indata$mileage > 100000,]
resratio_mileage <- mileage_limit$resratio
par(mfrow = c(1,2))
knitr::opts_chunk$set(fig.width = 6, fig.height = 4)
hist(resratio_mileage, main = "resratio for vehicles with mileage greater than 100,000", xlab = "resratio")
plot(density(resratio_mileage, na.rm = TRUE), main = "kernel density plot of resratio for vehicles with mil")
```

for vehicles with mileage greater of resratio for vehicles with milea



```
print("The distribution is close to normal distribution")
```

```
## [1] "The distribution is close to normal distribution"
```

```
mean_resratio <- mean(resratio_mileage, na.rm = TRUE)
cat("mean of resratio for vehicles with mileage greater than 100,000 is:", mean_resratio)
```

```
## mean of resratio for vehicles with mileage greater than 100,000 is: 0.1593309
```

```
std_resratio <- sd(resratio_mileage, na.rm = TRUE)
cat("standard deviation of resratio for vehicles with mileage greater than 100,000 is:", std_resratio)
```

```
## standard deviation of resratio for vehicles with mileage greater than 100,000 is: 0.04841774
```

Question f

```
options(width = 300)
indata$mile15 <- indata$mileage/150000
ave_vehicle <- indata[indata$congrade == "average",]
mean_average_resratio_edit <- mean(ave_vehicle$resratio)
sd_average_resratio_edit <- sd(ave_vehicle$resratio)
cat("the mean for resratio is:", mean_average_resratio_edit)
```

```
## the mean for resratio is: 0.4774244
```

```
cat("the standard deviation for resratio is:", sd_average_resratio_edit)
```

```
## the standard deviation for resratio is: 0.2266965
```

```
mean_average_mile15_edit <- mean(ave_vehicle$mile15, na.rm = TRUE)
sd_average_mile15_edit <- sd(ave_vehicle$mile15, na.rm = TRUE)
cat("the mean for resratio is:", mean_average_mile15_edit)
```

```
## the mean for resratio is: 0.3651413
```

```
cat("the standard deviation for resratio is:", sd_average_mile15_edit)
```

```
## the standard deviation for resratio is: 0.2220556
```

```
mean_average_vehage_edit <- mean(ave_vehicle$vehage, na.rm = TRUE)
sd_average_vehage_edit <- sd(ave_vehicle$vehage, na.rm = TRUE)
cat("the mean for resratio is:", mean_average_vehage_edit)
```

```
## the mean for resratio is: 58.17772
```

```
cat("the standard deviation for resratio is:", sd_average_vehage_edit)
```

```
## the standard deviation for resratio is: 34.79075
```