

Homework 1

Jing Leng

September 11, 2014

preprocessing:

```
read.csv("autodat.csv") -> autodat
autodat$mpg[autodat$mpg == 99] <- NA
autodat$horsepower[autodat$horsepower == 999] <- NA
autodat$acceleration[autodat$acceleration == 99] <- NA
autodat$origin <- factor(autodat$origin)
levels(autodat$origin) <- c("American", "European", "Japanese")
```

a). #

```
summary(autodat)
```

```
##      mpg      cylinders  displacement  horsepower
##  Min.   : 9.0    Min.   :3.00   Min.   : 68   Min.   : 46.0
##  1st Qu.:17.5    1st Qu.:4.00   1st Qu.:104   1st Qu.: 75.0
##  Median :23.0    Median :4.00   Median :146   Median : 92.5
##  Mean   :23.5    Mean   :5.46   Mean   :194   Mean   :104.5
##  3rd Qu.:29.0    3rd Qu.:8.00   3rd Qu.:262   3rd Qu.:129.0
##  Max.   :46.6    Max.   :8.00   Max.   :455   Max.   :230.0
##  NA's   :5
##      weight      acceleration      year      origin
##  Min.   :1613    Min.   : 8.0   Min.   :70   American:248
##  1st Qu.:2223    1st Qu.:13.8   1st Qu.:73   European: 70
##  Median :2800    Median :15.5   Median :76   Japanese: 79
##  Mean   :2970    Mean   :15.6   Mean   :76
##  3rd Qu.:3609    3rd Qu.:17.1   3rd Qu.:79
##  Max.   :5140    Max.   :24.8   Max.   :82
##      NA's   :8
##      name
##  ford pinto : 6
##  amc matador : 5
##  ford maverick : 5
##  toyota corolla: 5
##  amc gremlin : 4
##  amc hornet : 4
##  (Other) :368
```

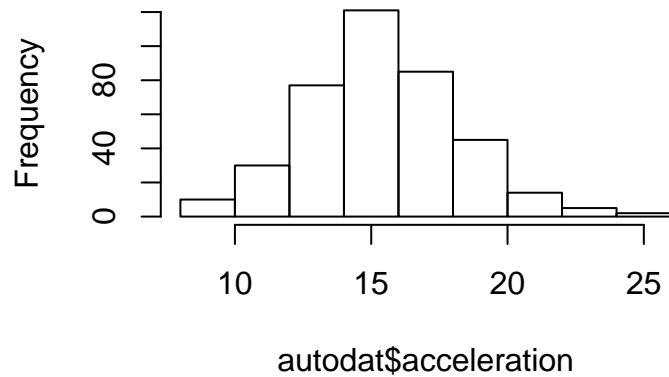
Summary of the data shows there are 5 values missing on **mpg**, 9 on **horsepower**, and 8 on **acceleration**.

b). 248 American, 70 European and 79 Japanese cars.

c).

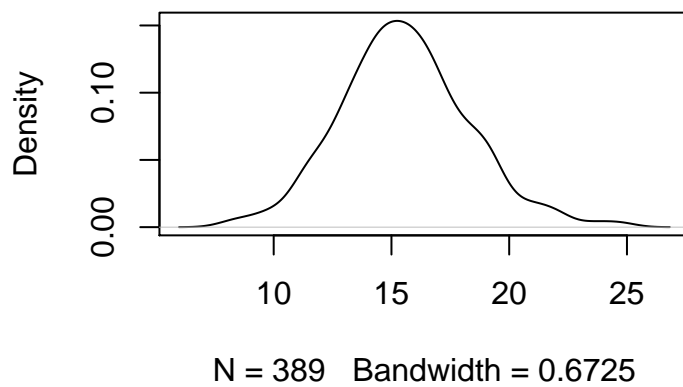
```
hist(autodat$acceleration)
```

Histogram of autodat\$acceleration



```
plot(density(autodat$acceleration, na.rm = T))
```

density.default(x = autodat\$acceleration, na.rm = T)



```
mean(autodat$acceleration, na.rm = T)
```

```
## [1] 15.55
```

```
sd(autodat$acceleration, na.rm = T)
```

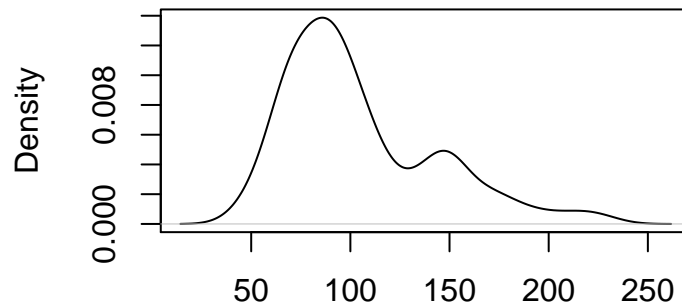
```
## [1] 2.749
```

mean: 15.55 standard deviation: 2.74

d).

```
plot(density(autodat$horsepower, na.rm = T), main = "kernel density of horsepower")
```

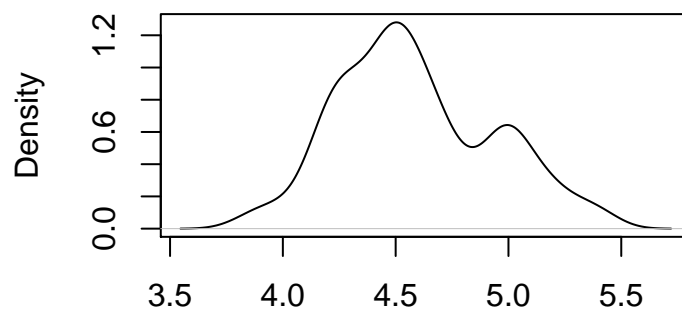
kernel density of horsepower



N = 388 Bandwidth = 10.57

```
plot(density(log(autodat$horsepower), na.rm = T), main = "kernel density of log(horsepower)")
```

kernel density of log(horsepower



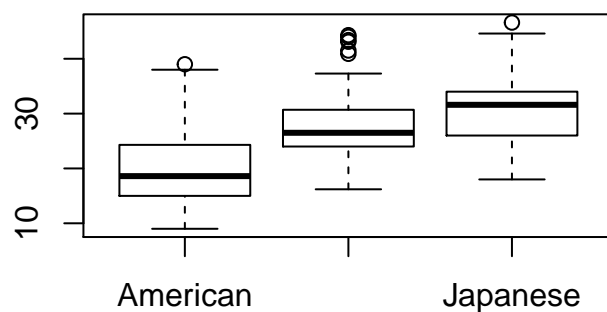
N = 388 Bandwidth = 0.09426

The density plot of horsepower is right-skewed. The density plot of log of horsepower is more symmetric. Both of them have more than one peaks.

e).

```
boxplot(autodat$mpg ~ autodat$origin, main = "boxplot of mpg")
```

boxplot of mpg

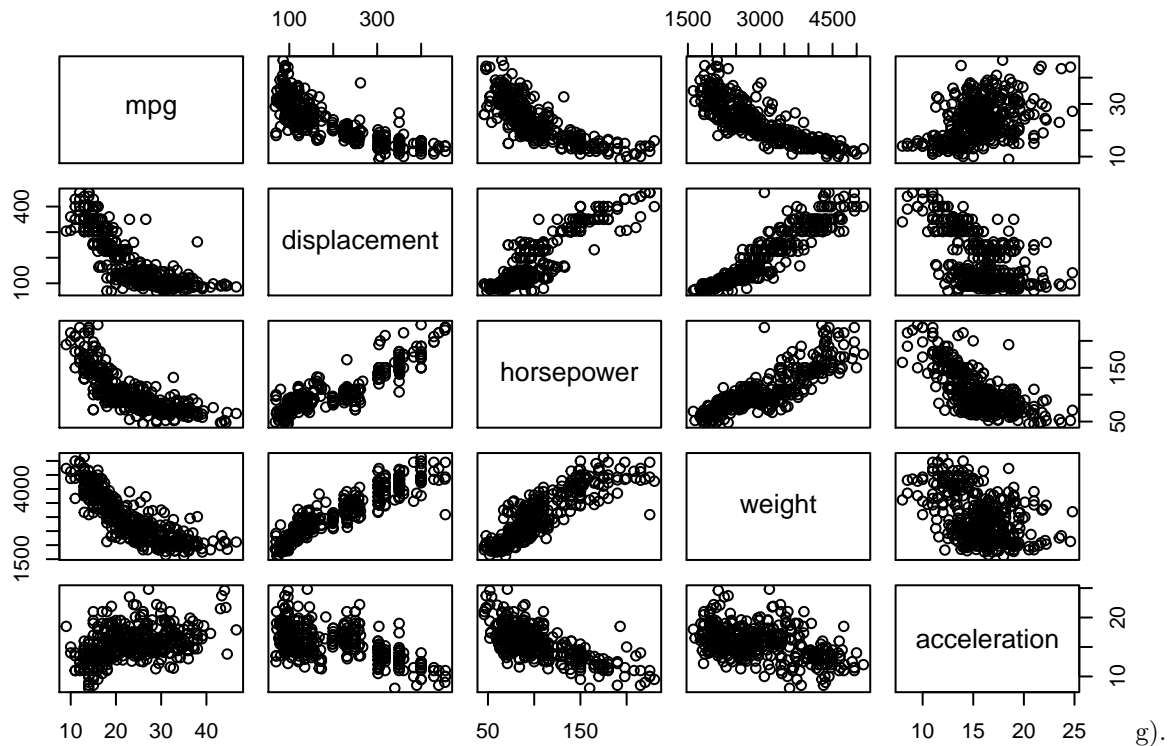


```
tapply(autodat$mpg, autodat$origin, function(x) mean(x, na.rm = T))
```

```
## American European Japanese
##      20.12      27.89      30.40
```

f).

```
pairs(autodat[, c("mpg", "displacement", "horsepower", "weight", "acceleration")])
```



```
apply(autodat[autodat$origin == "American", c("mpg", "displacement", "horsepower", "weight",
"acceleration")], 2, function(x) mean(x, na.rm = T))
```

```
##      mpg displacement horsepower weight acceleration
##      20.12      246.28      119.30  3363.25      15.02
```