MATH 324 Homework 2

Mehmet Arslan

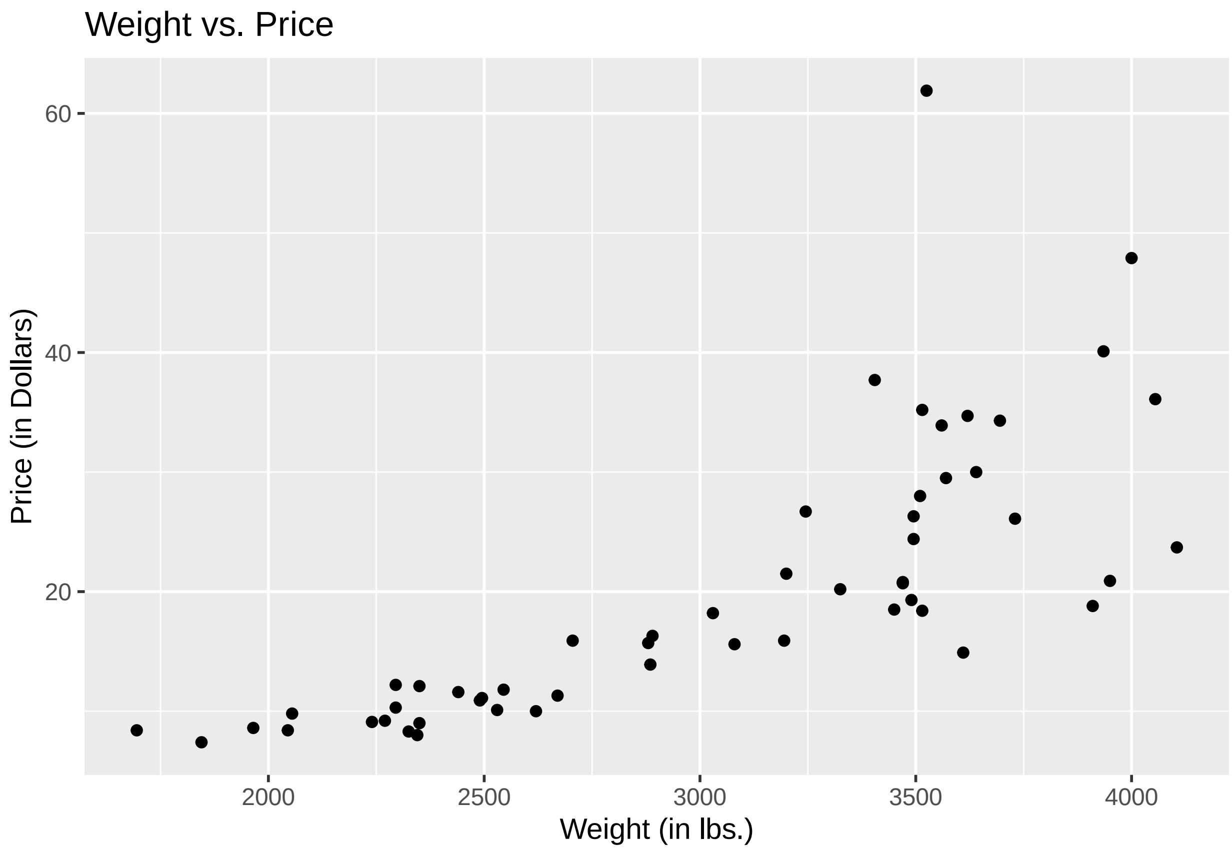
2/3/2021

```r
library(openintro)
library(ggplot2)
library(dplyr)
```

Problem 1

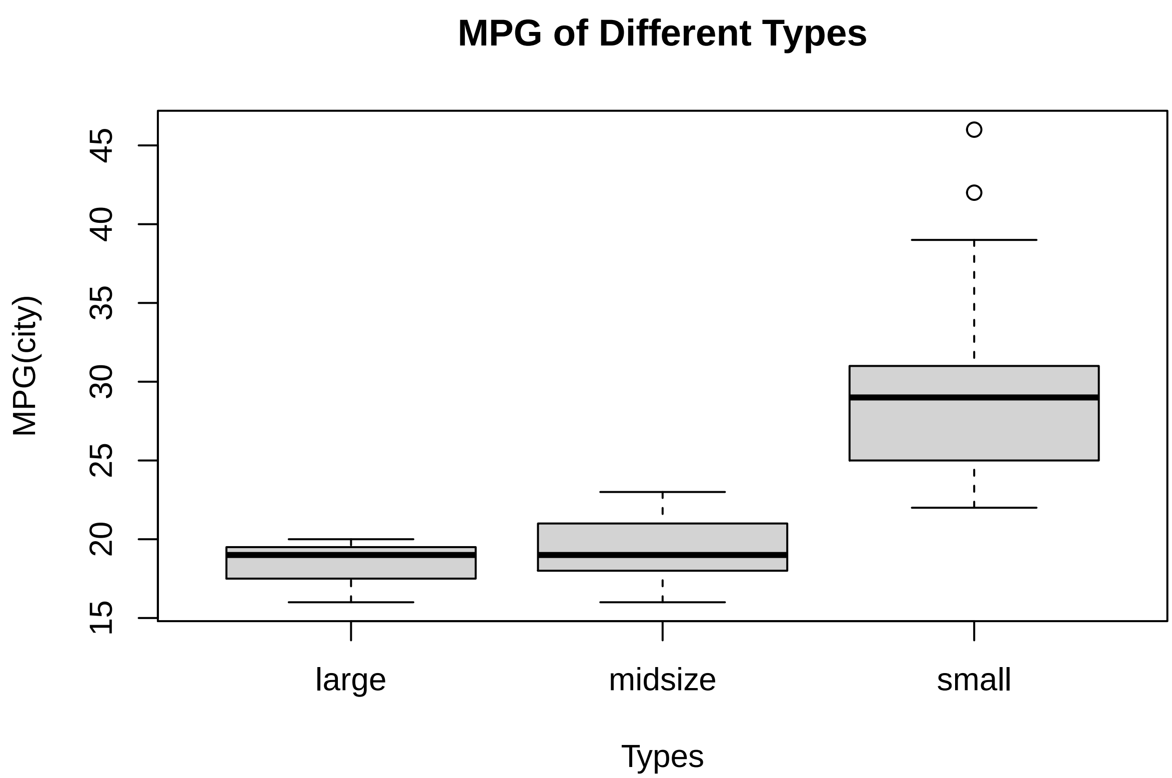
```r
data("cars93")
attach(cars93)


cars93 %>% ggplot(aes(weight, price)) +
  geom_point() +
  labs(x = "Weight (in lbs.)",
       y = "Price (in Dollars)",
       title = "Weight vs. Price")
```



Problem 2

```r
boxplot(mpg_city ~type, xlab = "Types", ylab = "MPG(city)", main = "MPG of Different Types")
```




```r
cars93 %>% group_by(type) %>% summarize(min = min(mpg_city),
                                        q1 = quantile(mpg_city, 0.25),
                                        med = median(mpg_city),
                                        q3 = quantile(mpg_city, 0.75),
                                        max = max(mpg_city),
                                        iqr = IQR(mpg_city))
```

```
## # A tibble: 3 x 7
##   type      min    q1   med    q3   max   iqr
## * <fct>   <int> <dbl> <dbl> <dbl> <int> <dbl>
## 1 large      16  17.5    19  19.5    20     2
## 2 midsize    16  18      19  21      23     3
## 3 small      22  25      29  31      46     6
```

```r
outlier_small = (1.5*6) + 31.0
outlier_midsize = (1.5*3) + 21.0
outlier_large = (1.5*2) + 19.5

outlier_small
```

```
## [1] 40
```

```r
outlier_midsize
```

```
## [1] 25.5
```

```
outlier_large
```

```
## [1] 22.5
```

I calculated the outliers manually using the standard formula: Q3 + (1.5 * IQR). These are my results:

1. 40
2. 25.5
3. 22.5

The boxplots shows that the large and midsize cars are nearly perfect however, the small cars have 2 distinct outliers. I looked back at the actual dataset itself, and then sorted by the mpg variable. I found that there were 2 small cars that both got 42 mpg and 46 mpg respectively, these two cars were 2 outliers.

Problem 3

```r
attach(cars93)

tapply(mpg_city, type, function(x) {c(mean(x), sd(x))})
```

```
## $large
## [1] 18.363636  1.501514
##
## $midsize
## [1] 19.54545  1.89554
##
## $small
## [1] 29.857143  6.109711
```

Small cars: mean = 29.86, standard deviation= 6.11

Midsize cars: mean = 19.54, standard deviation = 1.90

Large cars: mean = 18.36, standard deviation = 1.50

Problem 4 for Small Cars

```
cars_93_small = cars93 %>% select(type, mpg_city) %>% filter(type == "small")
t.test(cars_93_small$mpg_city, conf.level = 0.90)
```

```
##
##  One Sample t-test
##
## data:  cars_93_small$mpg_city
## t = 22.394, df = 20, p-value = 1.234e-15
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  27.55767 32.15662
## sample estimates:
## mean of x
##  29.85714
```

```
32.15662 - 29.85714
```

```
## [1] 2.29948
```

Problem 4 for Midsize Cars

```
cars_93_midsize = cars93 %>% select(type, mpg_city) %>% filter(type == "midsize")
t.test(cars_93_midsize$mpg_city, conf.level = 0.90)
```

```
##
##  One Sample t-test
##
## data:  cars_93_midsize$mpg_city
## t = 48.364, df = 21, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  18.85005 20.24086
## sample estimates:
## mean of x
##  19.54545
```

```
20.24086 - 19.54545
```

```
## [1] 0.69541
```

Problem 4 for Large Cars

```
cars_93_large = cars93 %>% select(type, mpg_city) %>% filter(type == "large")
t.test(cars_93_large$mpg_city, conf.level = 0.90)
```

```
##
##  One Sample t-test
##
## data:  cars_93_large$mpg_city
## t = 40.563, df = 10, p-value = 1.985e-12
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  17.54309 19.18418
## sample estimates:
## mean of x
##  18.36364
```

```r
19.18418 - 18.36363
```

```
## [1] 0.82055
```

1. 90% confidence interval for small cars: 29.86 ± 2.30 or (27.56, 32.16). The t-value was t = 22.394.
2. 90% confidence interval for midsize cars: 19.54 ± 0.70 or (18.85, 20.24). The t-value was t = 48.364.
3. 90% confidence interval for large cars: 17.54 ± 0.82 or (17.54, 19.18). The t-value was t = 40.563.

Problem 5

```r
pnorm(2.15)
```

```
## [1] 0.9842224
```

```r
1 - pnorm(1.97) #Symmetry for a bell curve with max value of 1.
```

```
## [1] 0.02441919
```

1. P(Z <= 2.15) = 0.98 or 98%
2. P(Z >= 1.97) = 0.02 or 2%

Problem 6

```r
pnorm(120, mean = 185, sd = 39)
```

```
## [1] 0.04779035
```

```r
1 - pnorm(150, mean = 185, sd = 39)
```

```
## [1] 0.8152568
```

```r
pnorm(220, mean = 185, sd = 39) - pnorm(190, mean = 185, sd = 39)
```

```
## [1] 0.2642501
```

1. $P(X < 120) = 0.047$ or 4.7%
2. $P(X > 150) = 0.815$ or 81.5%
3. $P(190 < X < 220) = 0.264$ or 26.4%

Problem 7

```
qnorm(c(0.05,0.10,0.30,0.50,0.70,0.90), mean = 185, sd = 39)
```

```
## [1] 120.8507 135.0195 164.5484 185.0000 205.4516 234.9805
```

I rounded all the percentiles to two decimal places.

1. 5th percentile = 120.85
2. 10th percentile = 135.02
3. 30th percentile = 164.55
4. 50th percentile = 185.00
5. 70th percentile = 205.45
6. 90th percentile = 234.98