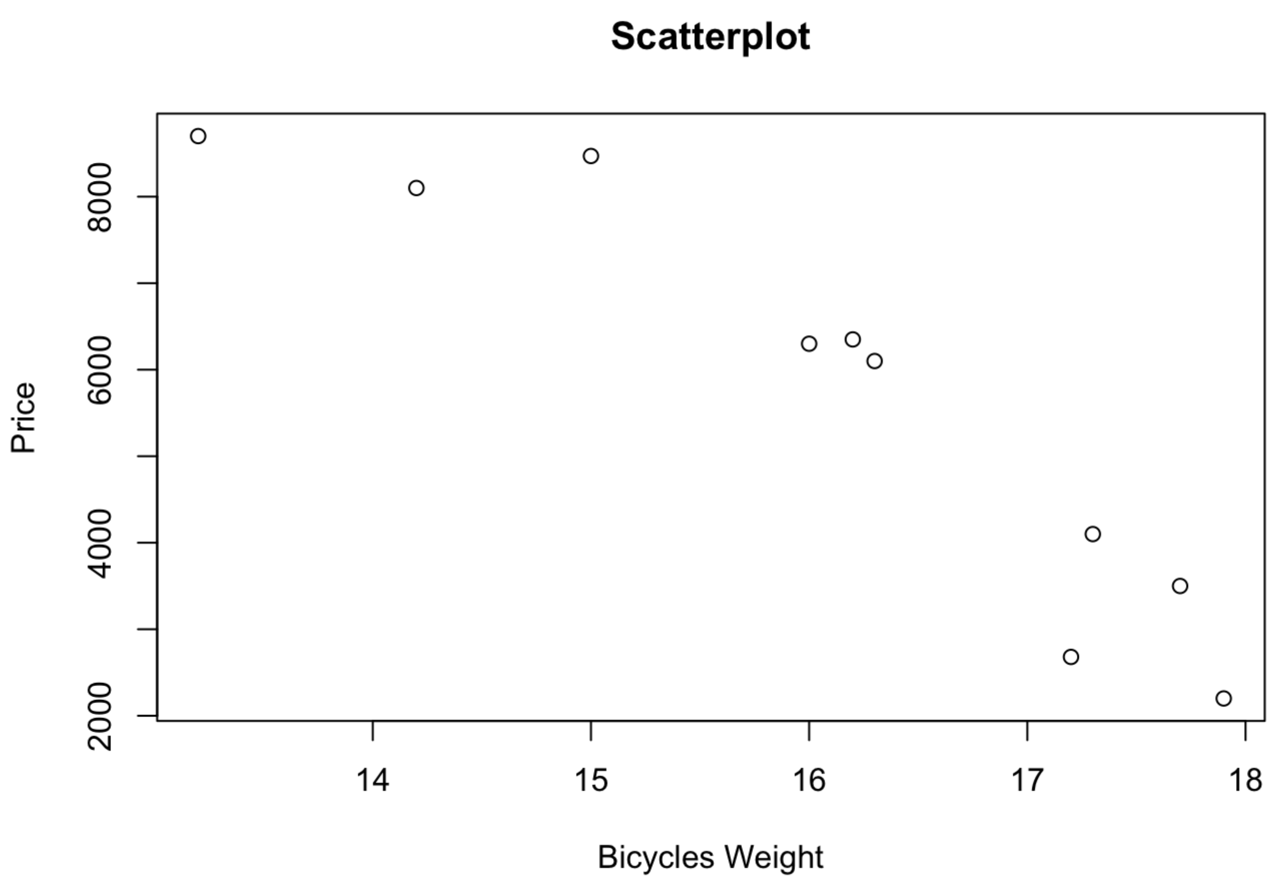
Hw4

Problem 1: Bicycling World Problem a.

```
Model = c('Fierro 7B', 'HX 5000', 'Durbin Ultralight', 'Schmidt', 'WSilton Advanced', 'b
icyclette velo', 'Supremo Team', 'XYC Racer', 'DOnofrio Pro', 'Americana #6')
Weight = c(17.9, 16.2, 15.0, 16.0, 17.3, 13.2, 16.3, 17.2, 17.7, 14.2)
Price = c(2200, 6350, 8470, 6300, 4100, 8700, 6100, 2680, 3500, 8100)
df <- data.frame(Model, Weight, Price)
plot(Weight, Price, main="Scatterplot", xlab="Bicycles Weight", ylab="Price")
```



b.

```
model1 = lm(df$Price ~ df$Weight)
model1$coefficients
```

```
## (Intercept)   df$Weight
##   28818.004   -1439.006
```

y = b0 + b1*x* = *28818 - 1439*x

c.

```
summary(model1)
```

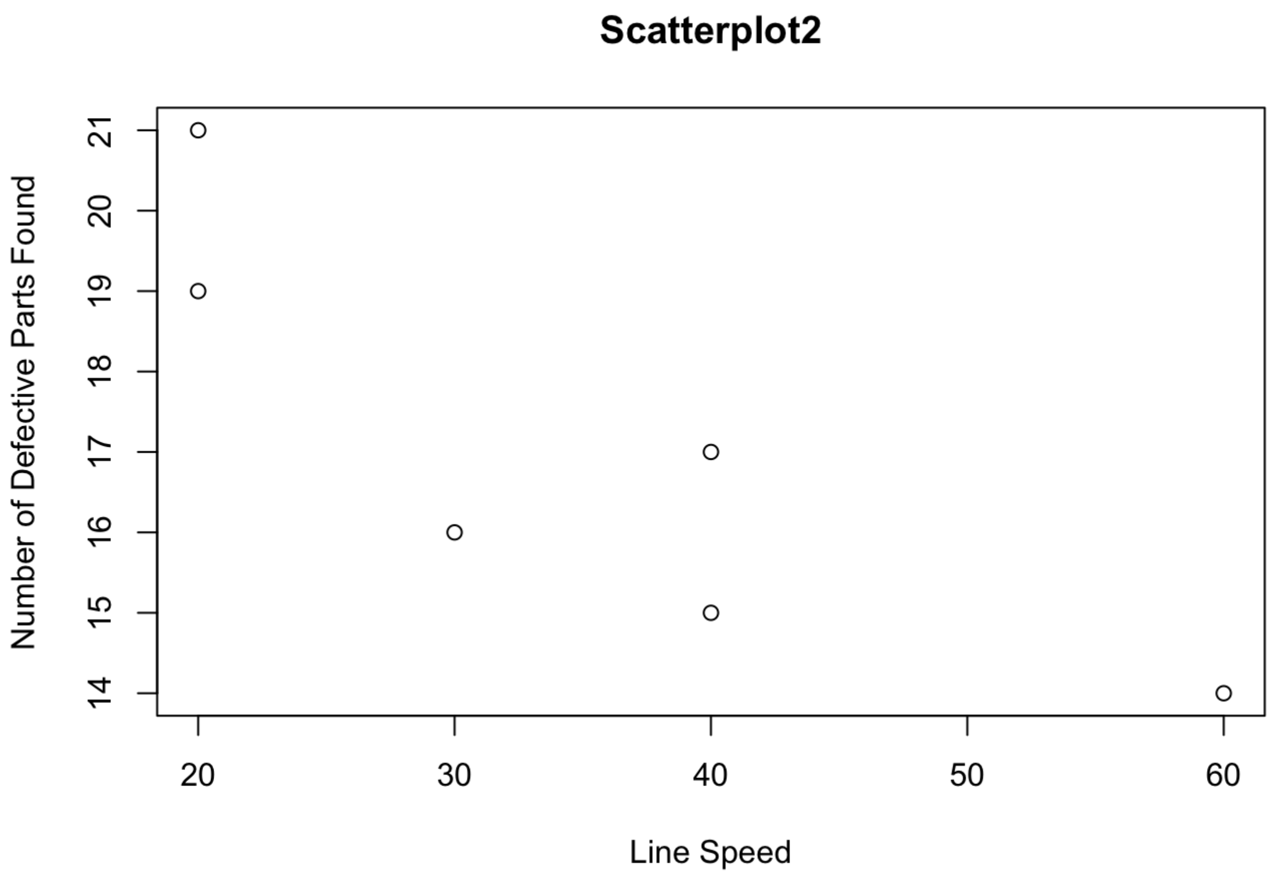
```
##
## Call:
## lm(formula = df$Price ~ df$Weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1387.1  -715.9   164.6   679.9  1237.1
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  28818.0     3267.3   8.820 2.15e-05 ***
## df$Weight    -1439.0      202.1  -7.121 9.99e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 942.3 on 8 degrees of freedom
## Multiple R-squared:  0.8637, Adjusted R-squared:  0.8467
## F-statistic:  50.7 on 1 and 8 DF,  p-value: 9.994e-05
```

H0: beta0 = 0, Ha: beta0 != 0, p-value = 9.99e-05, p-value < 0.05, so slope is significant to the model H0: beta1 = 0, Ha: beta1 != 0, p-value = 2.15e-05, p-value < 0.05, so intercept is significant to the model

d. R-squared = 0.8637, so 86.37% of the variation in the prices of the bicycles is accounted for the weight of bicycles.

Problem 2: Assembly Line Problem a.

```
LineSpeed = c(20,20,40,30,60,40)
numberofdefectivepart = c(21,19,15,16,14,17)
df2 = data.frame(LineSpeed, numberofdefectivepart)
plot(LineSpeed, numberofdefectivepart, main="Scatterplot2", xlab="Line Speed", ylab="Num
ber of Defective Parts Found")
```



The number of defective parts found is decrease as the line speed increase.

b.

```
model2 = lm(df2$numberofdefectivepart ~ df2$LineSpeed)
summary(model2)
```

```
##
## Call:
## lm(formula = df2$numberofdefectivepart ~ df2$LineSpeed)
##
## Residuals:
##       1       2       3       4       5       6
##  1.7826 -0.2174 -1.2609 -1.7391  0.6957  0.7391
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   22.17391    1.65275  13.416 0.000179 ***
## df2$LineSpeed -0.14783    0.04391  -3.367 0.028135 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.489 on 4 degrees of freedom
## Multiple R-squared:  0.7391, Adjusted R-squared:  0.6739
## F-statistic: 11.33 on 1 and 4 DF,  p-value: 0.02813
```

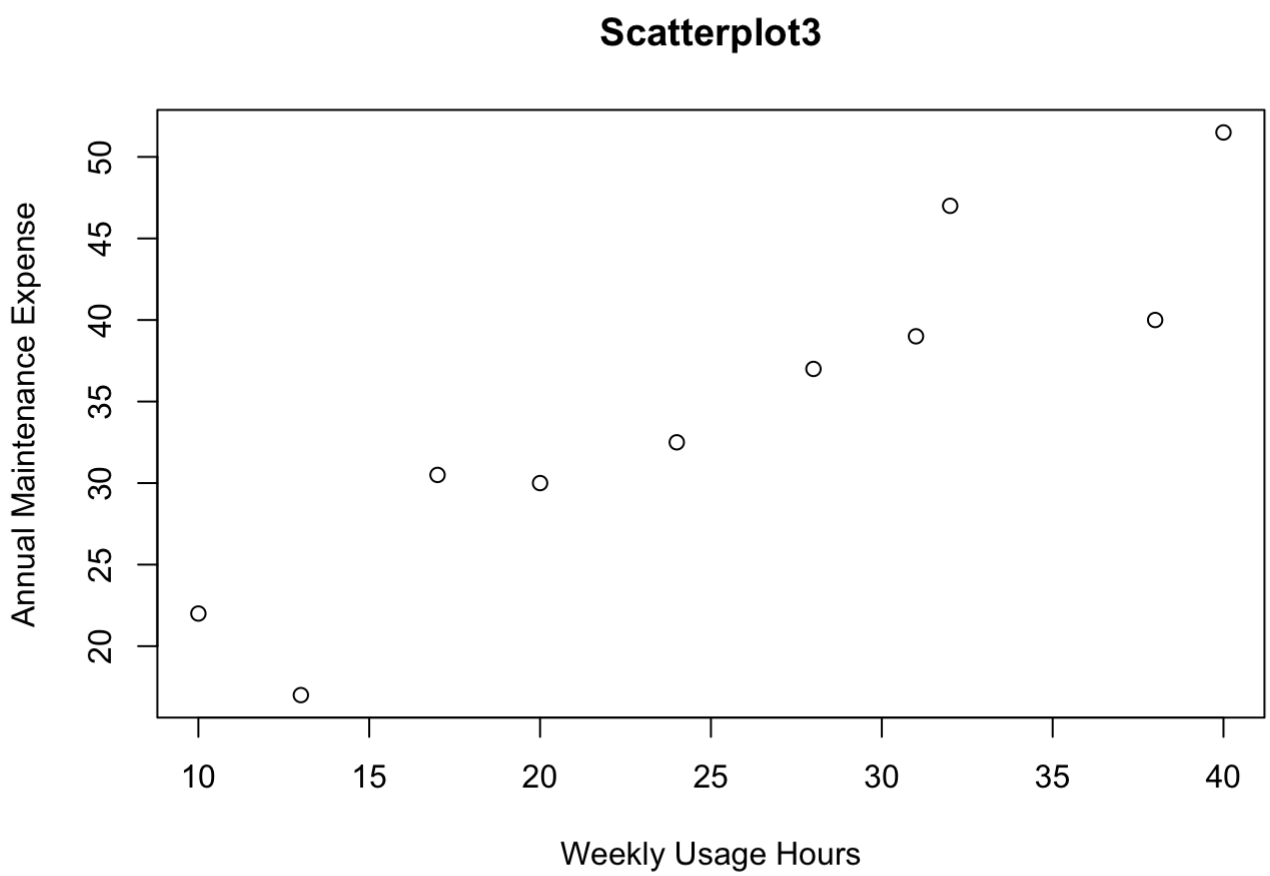
y = b0 + b1*x* = *22.1739 - 0.1478*x

c. H0: beta0 = 0, Ha: beta0 != 0, p-value = 0.000179, p-value < 0.01, so slope is significant to the model H0: beta1 = 0, Ha: beta1 != 0, p-value = 0.028135, p-value < 0.01, so intercept is insignificant to the model

d. R-squared = 0.7391, so 73.91% of the variation in the number of defective part is accounted for the line speed.

Problem 3: Jensen Tire & Auto Problem a.

```
WeeklyUsage = c(13,10,20,28,32,17,24,31,40,38)
AMexpense = c(17,22,30,37,47,30.5,32.5,39,51.5,40)
df3 = data.frame(WeeklyUsage, AMexpense)
plot(WeeklyUsage, AMexpense, main="Scatterplot3", xlab="Weekly Usage Hours", ylab="Annua
l Maintenance Expense")
```



Annual Maintenance Expense is positively related to Weekly Usage Hours

b.

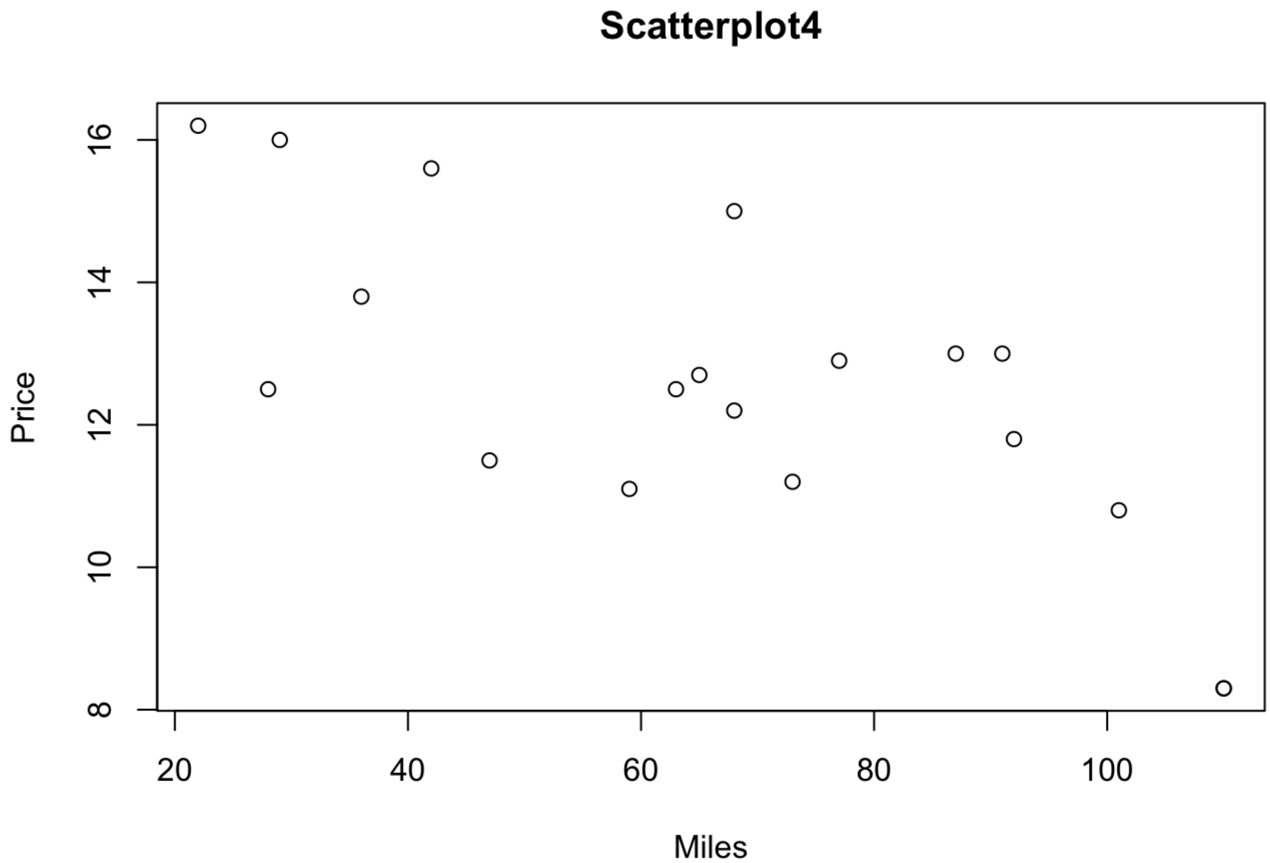
```
model3 = lm(df3$AMexpense ~ df3$WeeklyUsage)
summary(model3)
```

```
##
## Call:
## lm(formula = df3$AMexpense ~ df3$WeeklyUsage)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7587 -1.0411  0.0895  2.6102  5.9619
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      10.5280     3.7449   2.811 0.022797 *
## df3$WeeklyUsage   0.9534     0.1382   6.901 0.000124 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.25 on 8 degrees of freedom
## Multiple R-squared:  0.8562, Adjusted R-squared:  0.8382
## F-statistic: 47.62 on 1 and 8 DF,  p-value: 0.0001244
```

y = b0 + b1*x* = *10.5280* + *0.9534*x

4.Toyota Problem a.

```
Miles = c(22,29,36,47,63,77,73,87,92,101,110,28,59,68,68,91,42,65,110)
Price = c(16.2,16,13.8,11.5,12.5,12.9,11.2,13,11.8,10.8,8.3,12.5,11.1,15,12.2,13,15.6,1
2.7,8.3)
df4 = data.frame(Miles, Price)
plot(Miles, Price, main="Scatterplot4", xlab="Miles", ylab="Price")
```



b.

```
model4 = lm(df4$Price ~ df4$Miles)
summary(model4)
```

```
##
## Call:
## lm(formula = df4$Price ~ df4$Miles)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.32408 -1.34194  0.05055  1.12898  2.52687
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 16.46976    0.94876  17.359 2.99e-12 ***
## df4$Miles   -0.05877    0.01319  -4.455 0.000348 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.541 on 17 degrees of freedom
## Multiple R-squared:  0.5387, Adjusted R-squared:  0.5115
## F-statistic: 19.85 on 1 and 17 DF,  p-value: 0.0003475
```

y = b0 + b1*x* = *16.46976 – 0.05878*x

c. H0: beta0 = 0, Ha: beta0 != 0, p-value = 0.000348, p-value < 0.01, so slope is significant to the model H0: beta1 = 0, Ha: beta1 != 0, p-value = 2.99e-12, p-value < 0.01, so intercept is significant to the model

d. R-squared = 0.5387, so 53.87% of the variation in the price of the car is accounted for the miles of the car.

e.

```
predict(model4)
```

```
##        1        2        3        4        5        6        7        8
## 15.17673 14.76531 14.35389 13.70738 12.76700 11.94416 12.17926 11.35642
##        9       10       11       12       13       14       15       16
## 11.06255 10.53359 10.00462 14.82408 13.00209 12.47313 12.47313 11.12133
##       17       18       19
## 14.00125 12.64945 10.00462
```

```
head(sort(residuals(model4)),2)
```

```
##        12         4
## -2.324085 -2.207380
```

f.

```
16.4698-0.0587*60
```

```
## [1] 12.9478
```

5. Dodger Stadium Attendance Q1:

a.

```
url = "https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/dodge
rs.csv"
Dodger = read.csv(url)
df5 = as.data.frame(Dodger)
table(df5$cap)
```

```
##
##  NO YES
##  79   2
```

```
table(df5$shirt)
```

```
##
##  NO YES
##  78   3
```

```
table(df5$bobblehead)
```

```
##
##  NO YES
##  70  11
```

```
table(df5$fireworks)
```

```
##
##  NO YES
##  67  14
```

b.

```
by(data = df5$attend, INDICES = df5$cap, FUN = mean)
```

```
## df5$cap: NO
## [1] 41112.24
## --------------------------------------------------------
## df5$cap: YES
## [1] 38189.5
```

```
by(data = df5$attend, INDICES = df5$shirt, FUN = mean)
```

```
## df5$shirt: NO
## [1] 40824.55
## --------------------------------------------------------
## df5$shirt: YES
## [1] 46643.67
```

```
by(data = df5$attend, INDICES = df5$bobblehead, FUN = mean)
```

```
## df5$bobblehead: NO
## [1] 39137.93
## --------------------------------------------------------
## df5$bobblehead: YES
## [1] 53144.64
```

```
by(data = df5$attend, INDICES = df5$fireworks, FUN = mean)
```

```
## df5$fireworks: NO
## [1] 41032.18
## --------------------------------------------------------
## df5$fireworks: YES
## [1] 41077.86
```

c.

```
by(data = df5$attend, INDICES = df5$day_of_week, FUN = mean)
```

```
## df5$day_of_week: Friday
## [1] 40116.92
## --------------------------------------------------------
## df5$day_of_week: Monday
## [1] 34965.67
## --------------------------------------------------------
## df5$day_of_week: Saturday
## [1] 43072.92
## --------------------------------------------------------
## df5$day_of_week: Sunday
## [1] 42268.85
## --------------------------------------------------------
## df5$day_of_week: Thursday
## [1] 40407.4
## --------------------------------------------------------
## df5$day_of_week: Tuesday
## [1] 47741.23
## --------------------------------------------------------
## df5$day_of_week: Wednesday
## [1] 37585.17
```

```
by(data = df5$attend, INDICES = df5$month, FUN = mean)
```

```
## df5$month: APR
## [1] 39591.92
## --------------------------------------------------------
## df5$month: AUG
## [1] 42751.53
## --------------------------------------------------------
## df5$month: JUL
## [1] 43884.25
## --------------------------------------------------------
## df5$month: JUN
## [1] 47940.44
## --------------------------------------------------------
## df5$month: MAY
## [1] 37345.72
## --------------------------------------------------------
## df5$month: OCT
## [1] 36703.67
## --------------------------------------------------------
## df5$month: SEP
## [1] 38955.08
```

```
by(data = df5$attend, INDICES = df5$skies, FUN = mean)
```

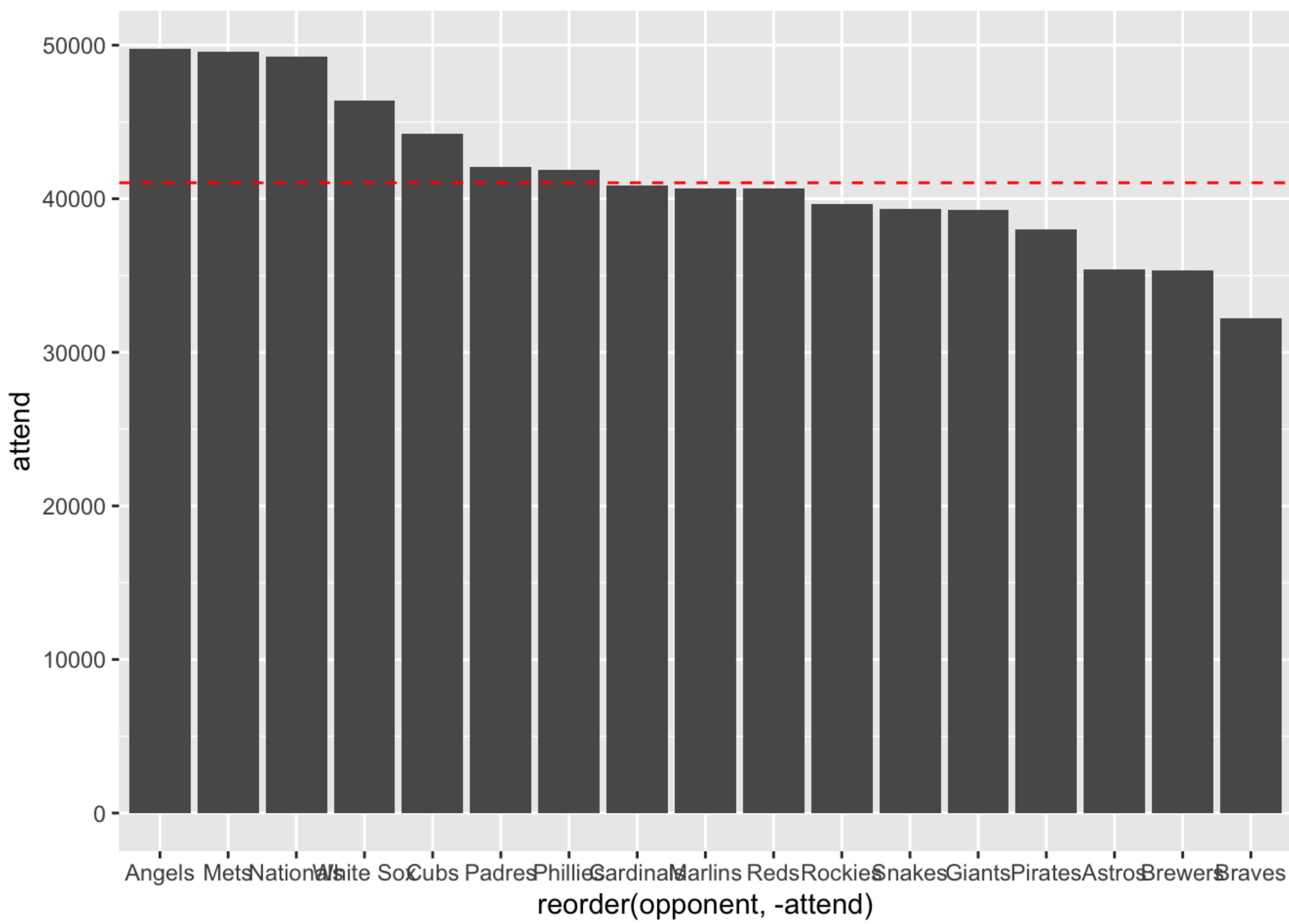
```
## df5$skies: Clear
## [1] 41729.21
## --------------------------------------------------------
## df5$skies: Cloudy
## [1] 38791.32
```

```
by(data = df5$attend, INDICES = df5$day_night, FUN = mean)
```

```
## df5$day_night: Day
## [1] 41793.27
## --------------------------------------------------------
## df5$day_night: Night
## [1] 40868.89
```

d.

```
library(ggplot2)
library(dplyr)
p = ggplot(df5, aes(x=reorder(opponent, -attend), y=attend)) + stat_summary(fun.y="mean"
, geom="bar")
p + geom_hline(yintercept=mean(df5$attend), linetype="dashed", color = "red")
```



Q2: a.

```
model5 = lm(df5$attend ~ df5$bobblehead + df5$fireworks + df5$shirt + df5$cap)
summary(model5)
```

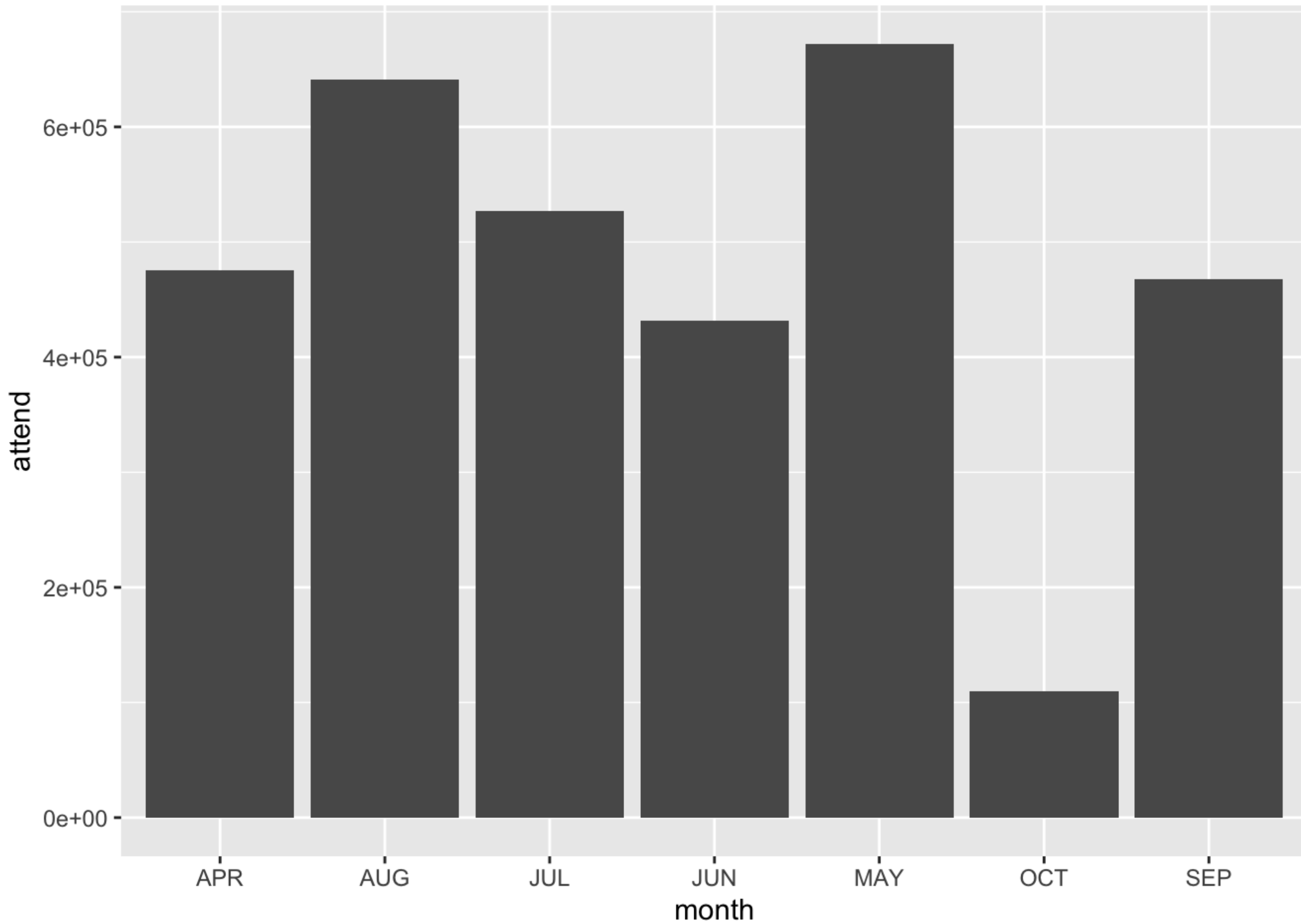
```
##
## Call:
## lm(formula = df5$attend ~ df5$bobblehead + df5$fireworks + df5$shirt +
##     df5$cap)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13889.1  -4466.1   -185.1   3729.1  17798.9
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       38201.08     933.09  40.940  < 2e-16 ***
## df5$bobbleheadYES 14943.56    2215.26   6.746 2.64e-09 ***
## df5$fireworksYES   2876.78    2010.56   1.431   0.1566
## df5$shirtYES       8442.59    3958.77   2.133   0.0362 *
## df5$capYES          -11.58    4803.39  -0.002   0.9981
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6664 on 76 degrees of freedom
## Multiple R-squared:  0.3873, Adjusted R-squared:  0.3551
## F-statistic: 12.01 on 4 and 76 DF,  p-value: 1.293e-07
```

Bobblehead promotions will increase attendance.

b. Bobblehead promotions increase the most attendance and it has the lowest p-value, so I think bobblehead promotions is better than all other promotions.

c.

```
p1 = ggplot(df5, aes(x=month, y=attend)) + stat_summary(fun.y="sum", geom="bar")
p1
```




```
sum(df5[((df5$month == 'JUN'|df5$month == 'JUL'|df5$month == 'AUG') & df5$bobblehead ==
'YES'),]$attend)
```

```
## [1] 432135
```

We should order 432135 booblehead for the summer time.