Final Project

April 10, 2018

1 Final Project

1.0.1 Importing data

```
In [1]: Computer_Stats <- read.csv(file = '/Users/loftina/Downloads/Computer_Stats.csv', heade
```

1.0.2 Splitting it by the monthly and daily data

```
In [2]: Monthly_2014_to_2016 <- Computer_Stats[,1:2]
        Daily_2017 <- Computer_Stats[,3:4]
```

1.0.3 Removing dollar signs

```
In [3]: Monthly_2014_to_2016$AC.ybar = as.numeric(gsub("\\$", "", Monthly_2014_to_2016$AC.ybar
        Daily_2017$ACy = as.numeric(gsub("\\$", "", Daily_2017$ACy))
```

```
In [4]: cat('Daily_2017')
        head(Daily_2017)
```

```
Daily_2017
```

ACy	PPx
345.35	36
346.25	38
351.65	39
341.75	35
347.00	35
288.95	36

```
In [5]: print('Monthly_2014_to_2016')
        head(Monthly_2014_to_2016)
```

```
[1] "Monthly_2014_to_2016"
```

image.png

AC.ybar	PPxbar
347.52	36.3
347.11	37.1
353.43	40.4
342.10	34.3
336.01	35.1
340.10	37.6

1.1 1 Perform following on AC and PP for 2017 stats

1.1.1 a. Find mean, median, mode, and standard deviation

In [61]:
```
Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}
```

In [62]:
```
cat('Daily 2017\n')
cat('AC Mean:', mean(Daily_2017$ACy), '\n')
cat('AC Median:', median(Daily_2017$ACy), '\n')
cat('AC Mode:', Mode(Daily_2017$ACy), '\n')
cat('AC Standard Deviation', sd(Daily_2017$ACy), '\n')
```

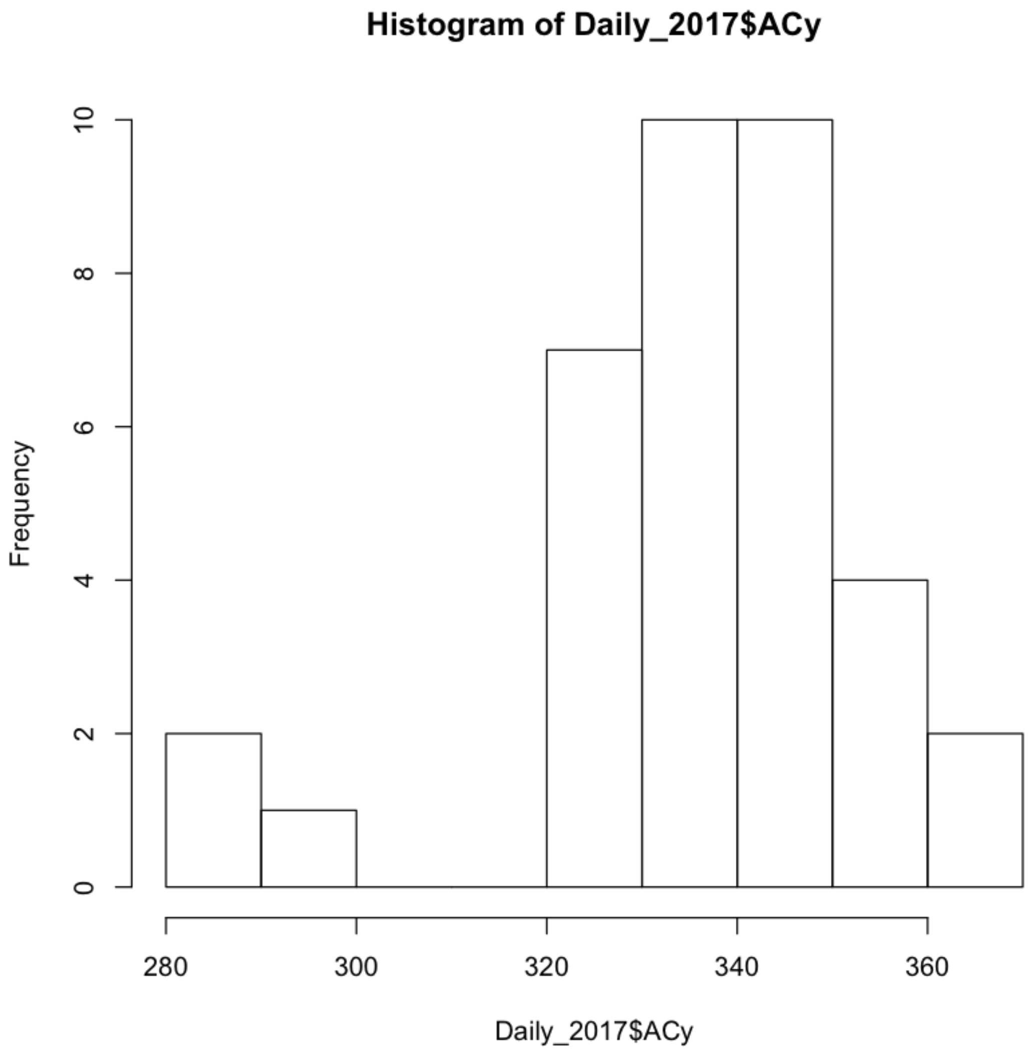
```
Daily 2017
AC Mean: 335.625
AC Median: 337.825
AC Mode: 341.25
AC Standard Deviation 16.69877
```

In [63]:
```
cat('Daily 2017\n')
cat('PP Mean:', mean(Daily_2017$PPx), '\n')
cat('PP Median:', median(Daily_2017$PPx), '\n')
cat('PP Mode:', Mode(Daily_2017$PPx), '\n')
cat('PP Standard Deviation', sd(Daily_2017$PPx), '\n')
```

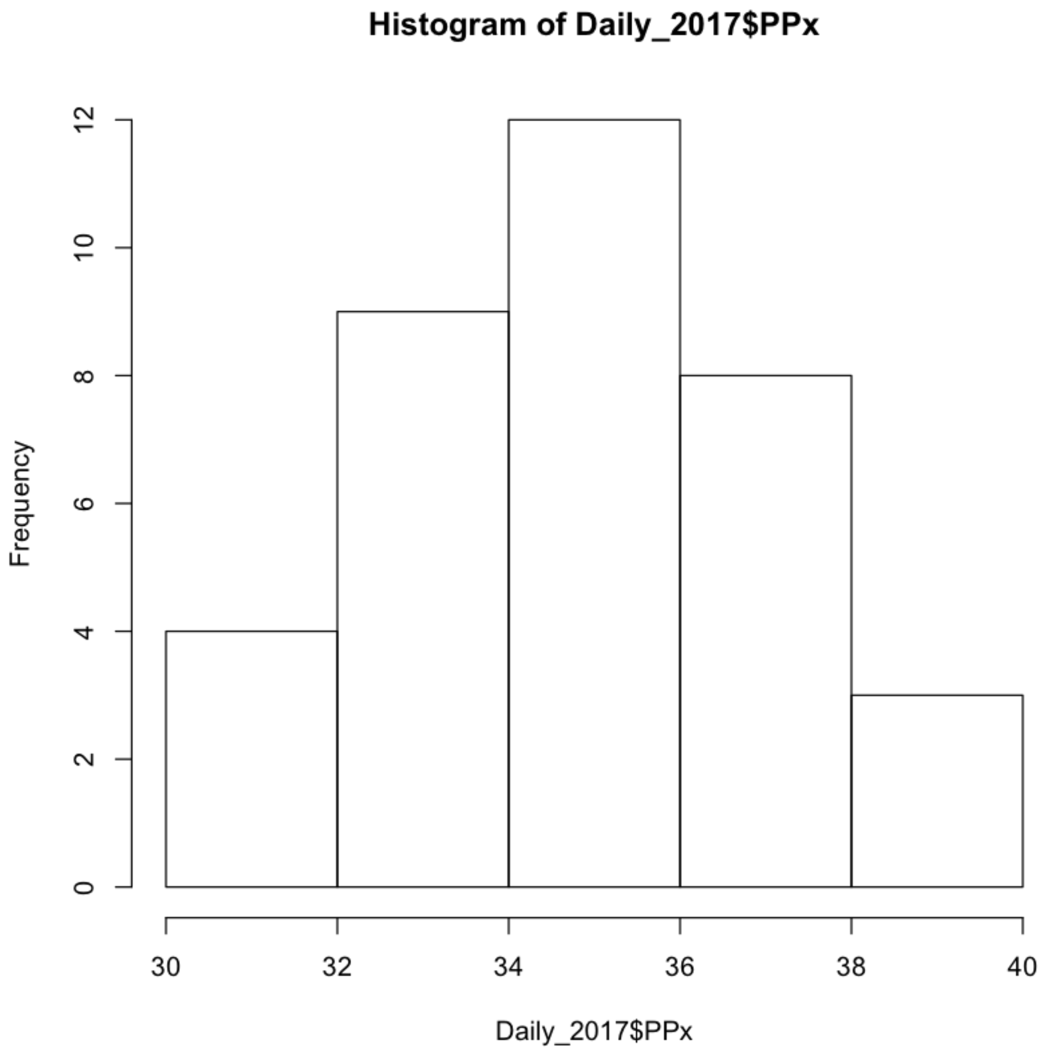
```
Daily 2017
PP Mean: 35.44444
PP Median: 36
PP Mode: 36
PP Standard Deviation 2.477838
```

1.1.2 b. Create a histogram using your frequency table

In [8]:
```
hist(Daily_2017$ACy)
```

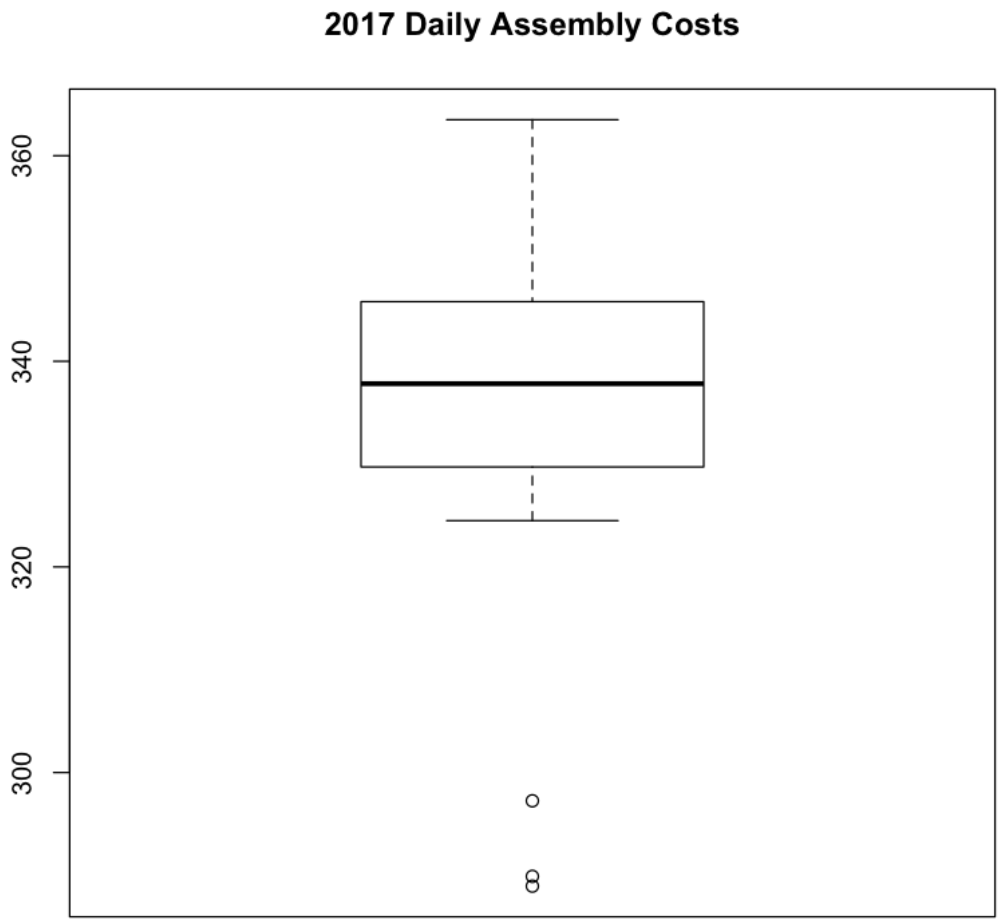



```
In [9]: hist(Daily_2017$PPx)
```



1.1.3 c. Draw a boxplot of the assembly costs

```
In [74]: boxplot(Daily_2017$ACy, main = '2017 Daily Assembly Costs')
```



1.1.4 d. Find precentage of days with AC within +/- 1 standard deviation from mean

```
In [11]: x_bar <- mean(Daily_2017$ACy)
         s <- sd(Daily_2017$ACy)
         lower_bound <- x_bar-s
         upper_bound <- x_bar+s
         counter <- 0
         for (i in 1:length(Daily_2017$ACy)) {
             if (Daily_2017$ACy[i] >= lower_bound && Daily_2017$ACy[i] <= upper_bound) {
                 counter <- counter + 1
             }
         }
         cat('Percentage daily AC within 1 sd:', counter/length(Daily_2017$ACy)*100)
```

```
Percentage daily AC within 1 sd: 80.55556
```

1.1.5 e. What percentage of days have PP > 35.4%

```
In [12]: counter <- 0
         for (i in 1:length(Daily_2017$PPx)) {
             if (Daily_2017$PPx[i] >= 35.4) {
                 counter <- counter + 1
             }
         }
         cat('Percentage of days with PP > 35.4:', counter/length(Daily_2017$ACy)*100)
```

```
Percentage of days with PP > 35.4: 58.33333
```

1.1.6 f. What percentage of days have PP > 35.4% and AC < 330

```
In [13]: counter <- 0
         for (i in 1:length(Daily_2017$PPx)) {
             if (Daily_2017$PPx[i] >= 35.4 && Daily_2017$ACy[i] < 330) {
                 counter <- counter + 1
             }
         }
         cat('Percentage of days with PP > 35.4 with AC < 330:', counter/length(Daily_2017$ACy
```

```
Percentage of days with PP > 35.4 with AC < 330: 11.11111
```

1.2 2 Assume mu of AC for first 36 months is normal

1.2.1 a. Find mu and sigma and what proportion of AC was between 320 and 350

```
In [14]: cat('Monthly_2014_to_2016\n')
         cat('AC Mean:', mean(Monthly_2014_to_2016$AC.ybar), '\n')
         cat('AC Standard Deviation', sd(Monthly_2014_to_2016$AC.ybar), '\n')
```

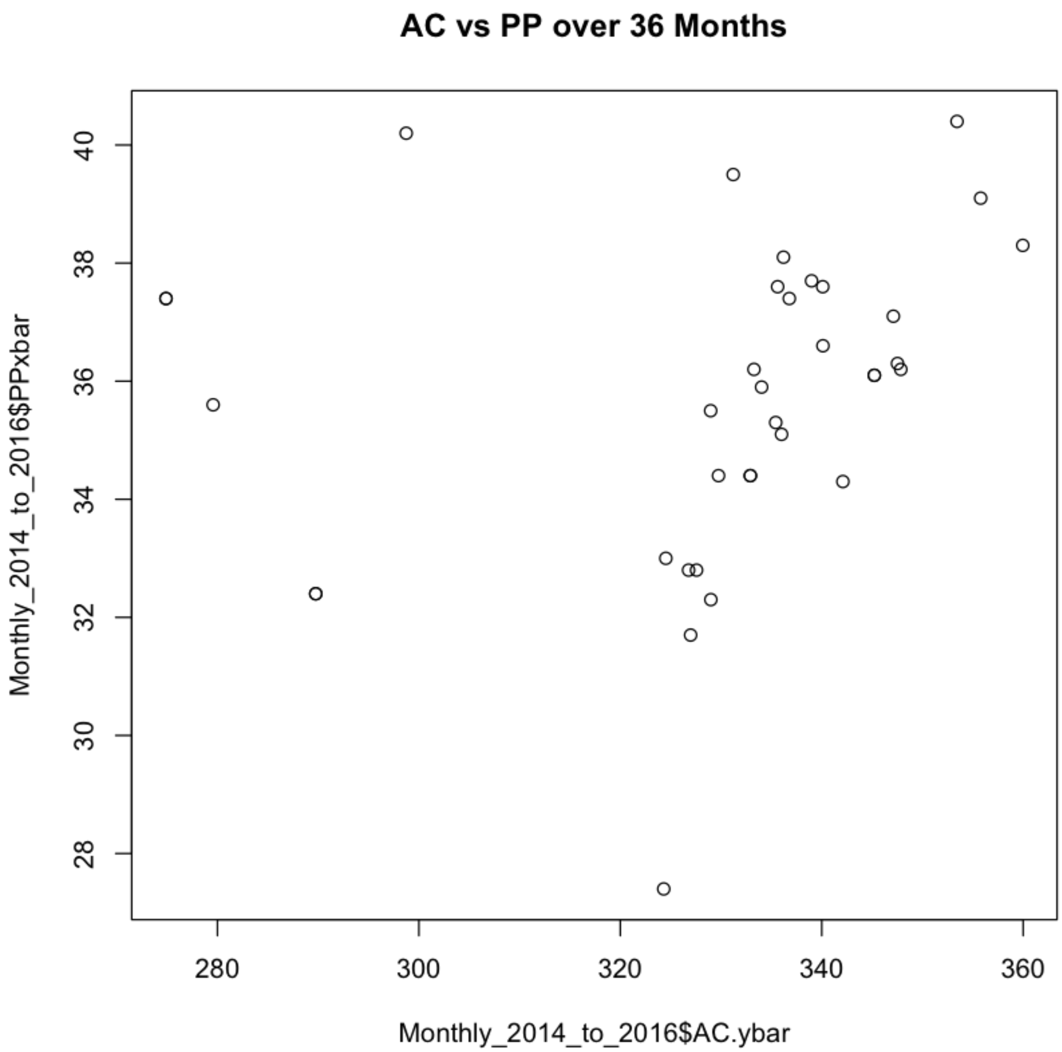
```
Monthly_2014_to_2016
AC Mean: 328.7014
AC Standard Deviation 22.04782
```

```
In [15]: counter <- 0
         for (i in 1:length(Monthly_2014_to_2016$AC.ybar)) {
             if (Monthly_2014_to_2016$AC.ybar[i] >= 320 && Monthly_2014_to_2016$AC.ybar[i] <= 3
                 counter <- counter + 1
             }
         }
         cat('Proportion of AC between 320 and 350:', counter/length(Monthly_2014_to_2016$AC.yb
```

```
Proportion of AC between 320 and 350: 0.75
```

1.2.2 b. Plot AC on x and PP on y for 36 month data

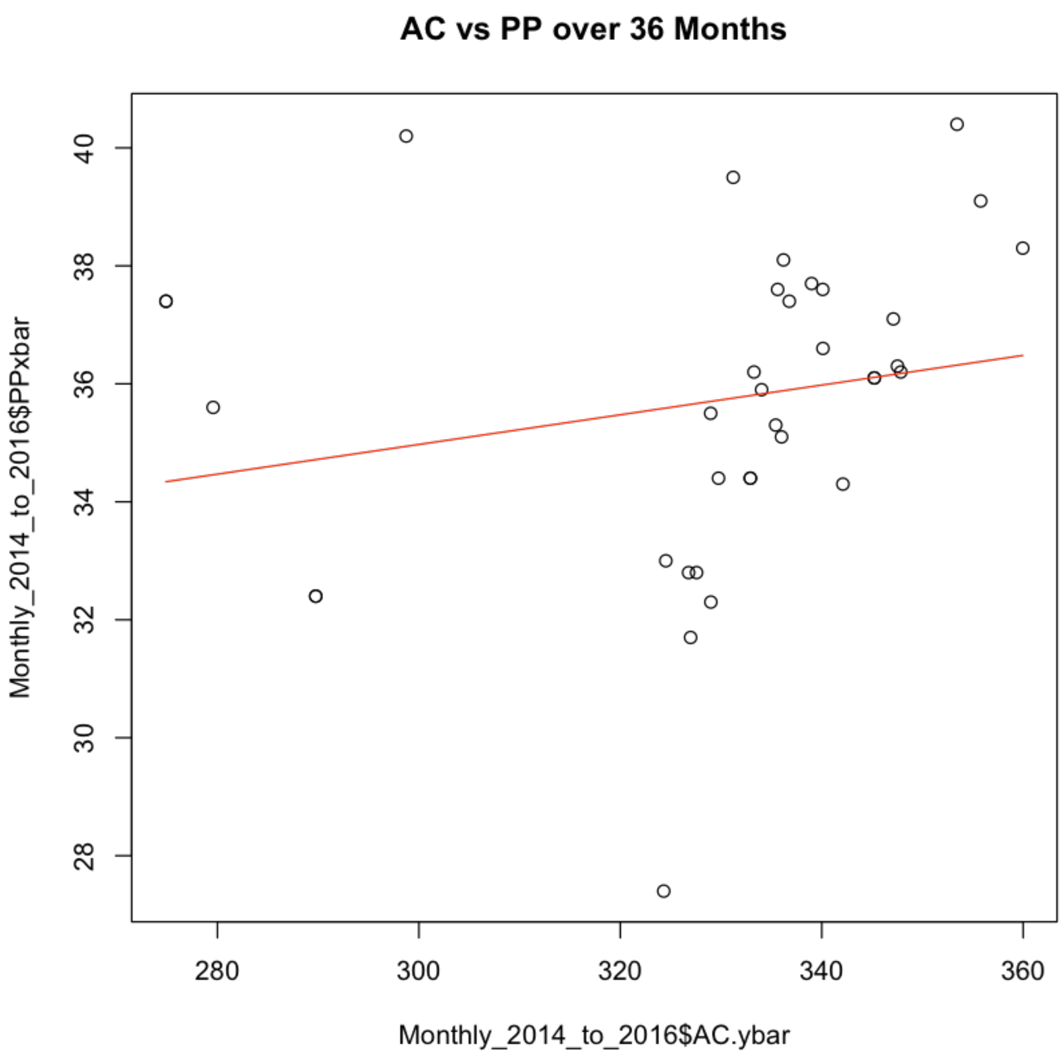
```
In [16]: plot(x = Monthly_2014_to_2016$AC.ybar, y = Monthly_2014_to_2016$PPxbar, main = 'AC vs
```



1.2.3 c. Draw a trend line for the above graph and calculate the equation for the line of regression

```
In [17]: AC_PP_Trend <- lm(Monthly_2014_to_2016$PPxbar ~ Monthly_2014_to_2016$AC.ybar)
```

```
In [18]: plot(x = Monthly_2014_to_2016$AC.ybar, y = Monthly_2014_to_2016$PPxbar, main = 'AC vs
         lines(x = Monthly_2014_to_2016$AC.ybar, y = predict(AC_PP_Trend), col = 'red')
```




```
In [19]: summary(AC_PP_Trend)
```

```
Call:
lm(formula = Monthly_2014_to_2016$PPxbar ~ Monthly_2014_to_2016$AC.ybar)

Residuals:
    Min      1Q  Median      3Q     Max
-8.1839 -1.4832  0.0476  1.7353  5.2595

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                  27.42308    6.76427   4.054 0.000277 ***
```

```
Monthly_2014_to_2016$AC.ybar  0.02516    0.02053   1.225 0.228819
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.678 on 34 degrees of freedom
Multiple R-squared:  0.0423,Adjusted R-squared:  0.01413
F-statistic: 1.502 on 1 and 34 DF,  p-value: 0.2288
```

The linear model predicts a line with the following formula: y = 0.02516x + 27.42308

1.2.4 d. Approximate the PP that could result from AC of 340 and AC that would result from PP of 34.7

```
In [20]: cat('PP from AC of 340:', round(0.02516*340+27.42308), '%')
```

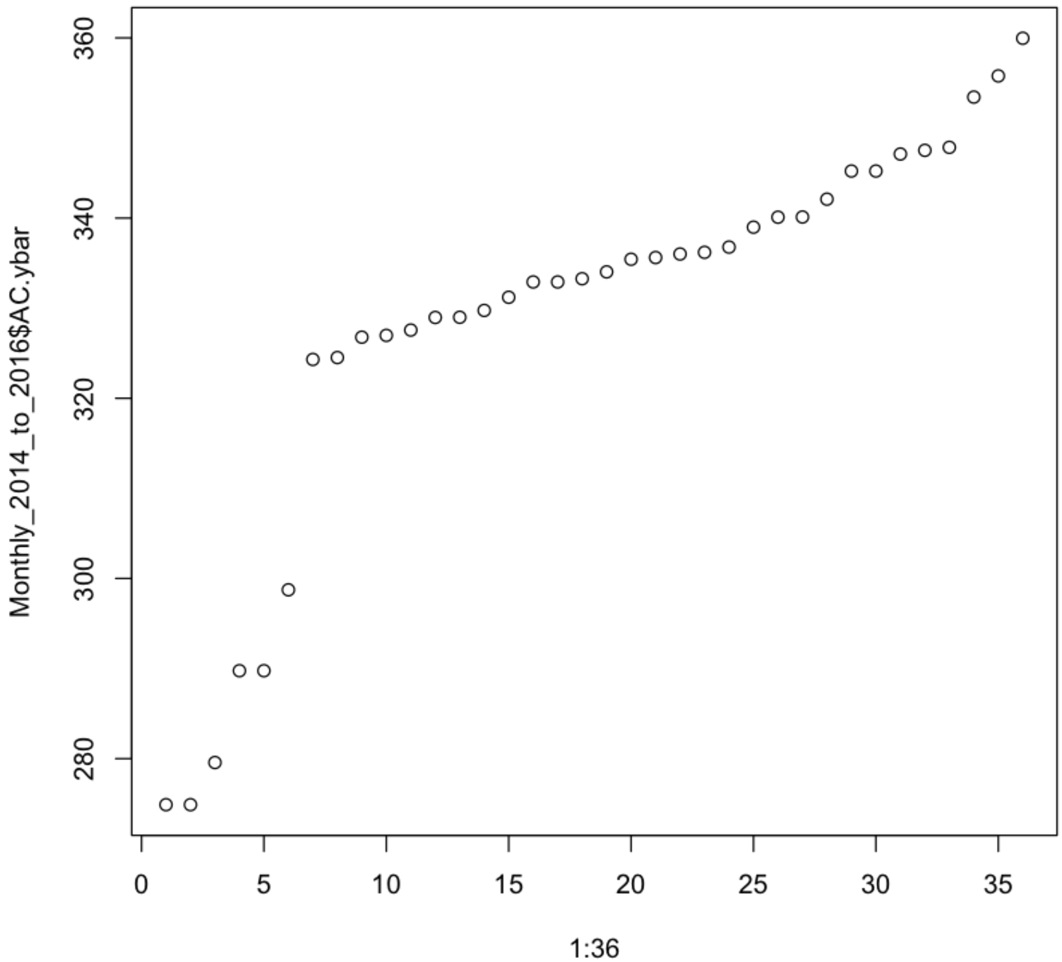
```
PP from AC of 340: 36 %
```

```
In [21]: cat('AC from PP of 34.7: $', round((34.7-27.42308)/0.02516, 2))
```

```
AC from PP of 34.7: $ 289.23
```

1.2.5 e. Show that the data is normally distributed

```
In [83]: qqplot(x = 1:36, Monthly_2014_to_2016$AC.ybar)
```



1.3 3 Assuming Daily AC follows a normal distribution estimate the mean using margin of error and C.I. of 95%

```
In [78]: n <- length(Daily_2017$ACy)
         x_bar <- mean(Daily_2017$ACy)
         s <- sd(Daily_2017$ACy)
         lower_interval <- x_bar - 1.96*s/sqrt(n)
         upper_interval <- x_bar + 1.96*s/sqrt(n)
         cat('Margin of error:', 1.96*s/sqrt(n), '\n')
         cat('95% CI: (',lower_interval,',',upper_interval,')')
```

```
Margin of error: 5.454932
95% CI: ( 330.1701 , 341.0799 )
```

1.4 4 Assume monthly AC data mean and sd as Ho

In [50]:
```
Ho_mu <- mean(Monthly_2014_to_2016$AC.ybar)
Ho_sd <- sd(Monthly_2014_to_2016$AC.ybar)
cat('Ho mean:', Ho_mu, '\n')
cat('Ho sd:', Ho_sd)
```

```
Ho mean: 328.7014
Ho sd: 22.04782
```

1.4.1 a. Has this changed in daily 2017 data?

In [80]:
```
Ha_mu <- mean(Daily_2017$ACy)
Ha_sd <- sd(Daily_2017$ACy)
cat('Ha mean:', Ha_mu, '\n')
cat('Ha sd:', Ha_sd, '\n')
cat('The mean has increased.')
```

```
Ha mean: 335.625
Ha sd: 16.69877
The mean has increased.
```

1.4.2 b. State your alternative hypothesis

In [52]:
```
cat('The population mean is greater than', Ho_mu)
```

```
The population mean is greater than 328.7014
```

1.4.3 c. Use alpha = .05 to find Z-test and P-value to accept or reject the Null

In [82]:
```
z <- (Ha_mu - Ho_mu)/(Ho_sd/sqrt(n))
cat('z test statistic:', z, '\n')
p_value <- pnorm(q = z, lower.tail = F)
cat('p-value:', p_value)
```

```
z test statistic: 1.884162
p-value: 0.02977155
```

p-value is less than alpha, so we reject the null hypothesis.

1.4.4 d. Has there been a statistical change in the mean AC since 2014?

yes, it has increased

1.5 Methods Used

- measures of center and spread
- plotting methods (boxplot, hist, scatter, line)
- linear regression
- confidence intervals
- hypothesis testing