Lab 1

Saul Grimaldo, Melwin Poovakottu, Anirudh Mittal

1/18/2017

```
library(ggplot2)
setwd("../data/")
load("ceo_w203.RData")
```

Exploration of education variables

The college variable is a binary variable. Within this sample, we see that more than 96% of CEO's have a college education

```
# Calculates percent of CEOs in our data set with a college education
number_of_college_CEO = sum(CEO$college)
data_set_size = nrow(CEO)
number_of_college_CEO / data_set_size
```

```
## [1] 0.9621622
```

Only 7 CEOs in our data set did not report having a college education.

```
data_set_size - number_of_college_CEO
```

```
## [1] 7
```

Next, we look at graduate school completion. Within our sample, we see that a total of 55 % of CEO's reported having a graduate level education

```
sum(CEO$grad) / nrow(CEO)
```

```
## [1] 0.5513514
```

Interestingly, 2 CEOs reported having a graduate level education but not a college level education. For our future analysis, we will assume that CEO's who reported having a graduate level education would have also received a college education and will thus be updated to reflect this. This new data will be stored in an adjusted_college variable.

```
# counts the number of CEOs who reported having a graduate level education but no college education
nrow(subset(CEO, grad == 1 & college ==0))
```

```
## [1] 2
```

```
# updates college education for CEOs who reported having a graduate level degree.
CEO$adjusted_college = ifelse(CEO$grad == 1, 1,CEO$college)
CEO$grad = as.factor(CEO$grad)
```

Because of the updates to the data, we now have only 5 CEO's who did not report having a college education or better. Because of the minimal amount of data in our data set for CEOs with no education, we will focus education related questions on graduate level education.

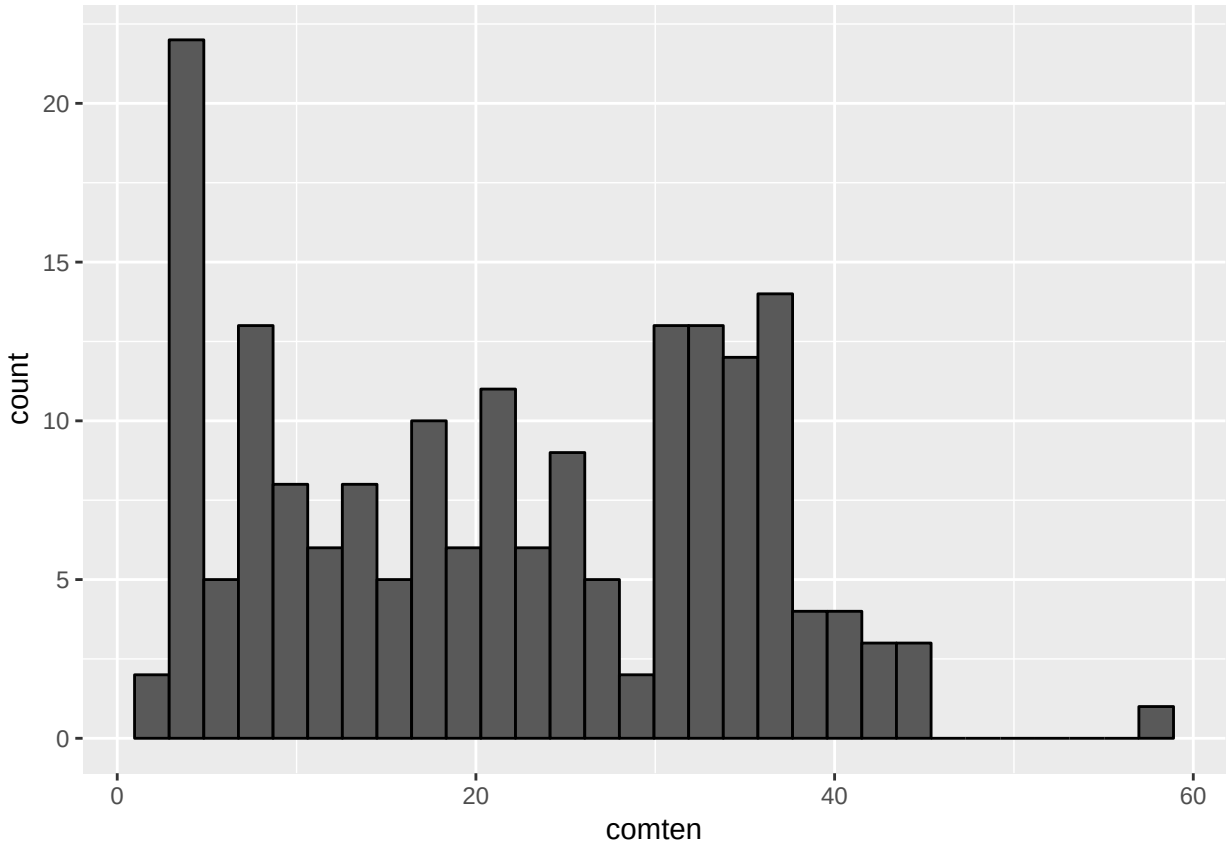
```
nrow(subset(CEO, adjusted_college == 0))
```

```
## [1] 5
```

Exploration of company tenure

Next, we'll explore the company tenure variable.

```
ggplot(CEO, aes(x = comten)) + geom_histogram(color = as.factor(1), bins = 30)
```



We can see a large spike in the number of CEOs with low amounts of company tenure. This likely corresponds to CEOs who were hired from out of the company.

There is also a large bump in the number of CEOs who had 30 - 36 years of company tenure.

Finally, we can see the number of CEOs with more than 36 years of experience dropping off dramatically. This likely represents CEOs who entered retirement.

Lastly, we also see a single data point further down the curve, at around 58. This corresponds to someone who is 80 years old! This 80 year old has been around his/ber company since he/she was 22! Someone getting their first job out of college at 22 is not rare, so this data point is not bad data.

```
subset(CEO, comten == 58)
```

```
##     salary age college grad comten ceoten profits mktval adjusted_college
## 122    396  80       1    0     58     28      53    963                1
```

Below, we can see the range in CEO company tenure going from 2 to 58, with the mean and median being relatively close at 21.66 and 21 respectively.

```
summary(CEO$comten)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00    9.00   21.00   21.66   33.00   58.00
```

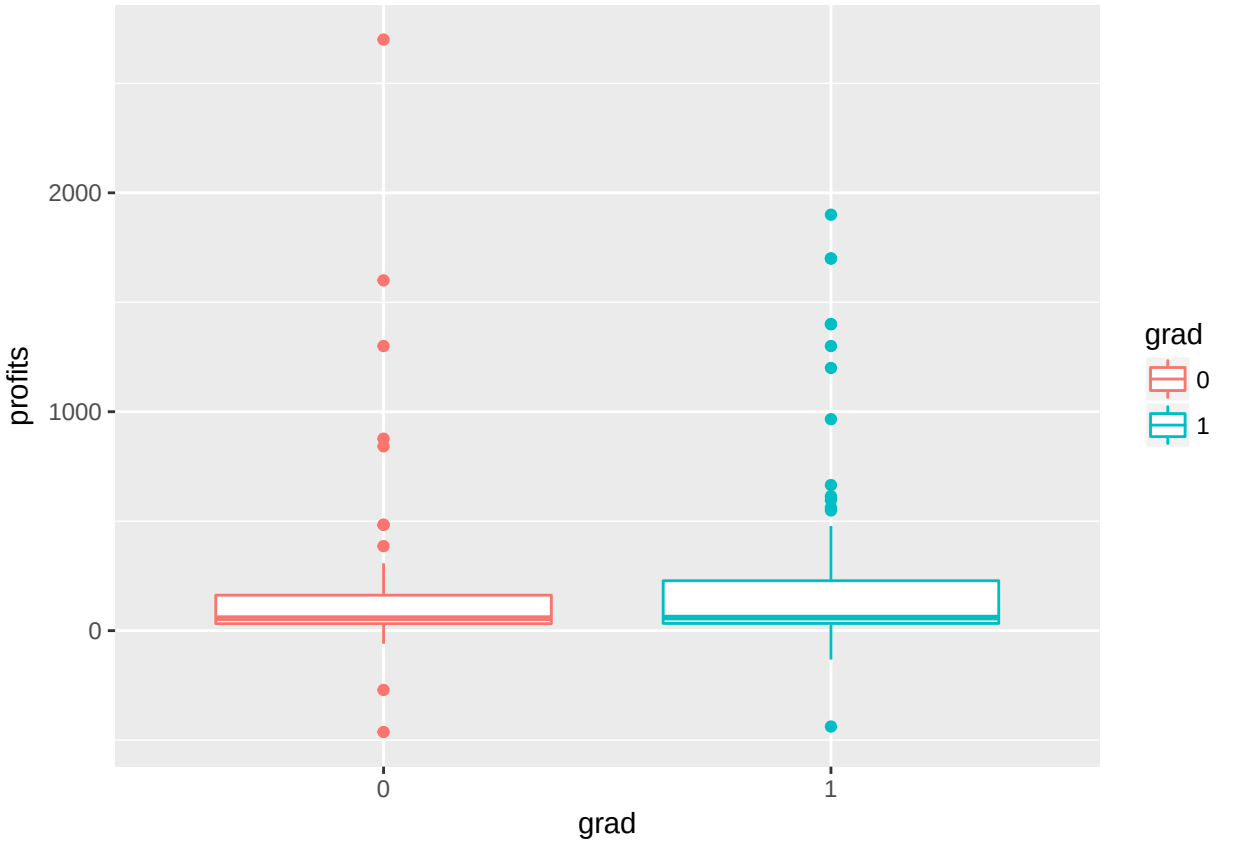
We can also see some interesting trends when we compare the company tenure for CEOs with a graduate-level education with those who don't have one.

Bivariate Relationships

Education vs Profit

First we'll examine the profits for companies run by those with a graduate level education versus those without it.

```
ggplot(CEO, aes(x = grad, y = profits, color = grad)) + geom_boxplot()
```



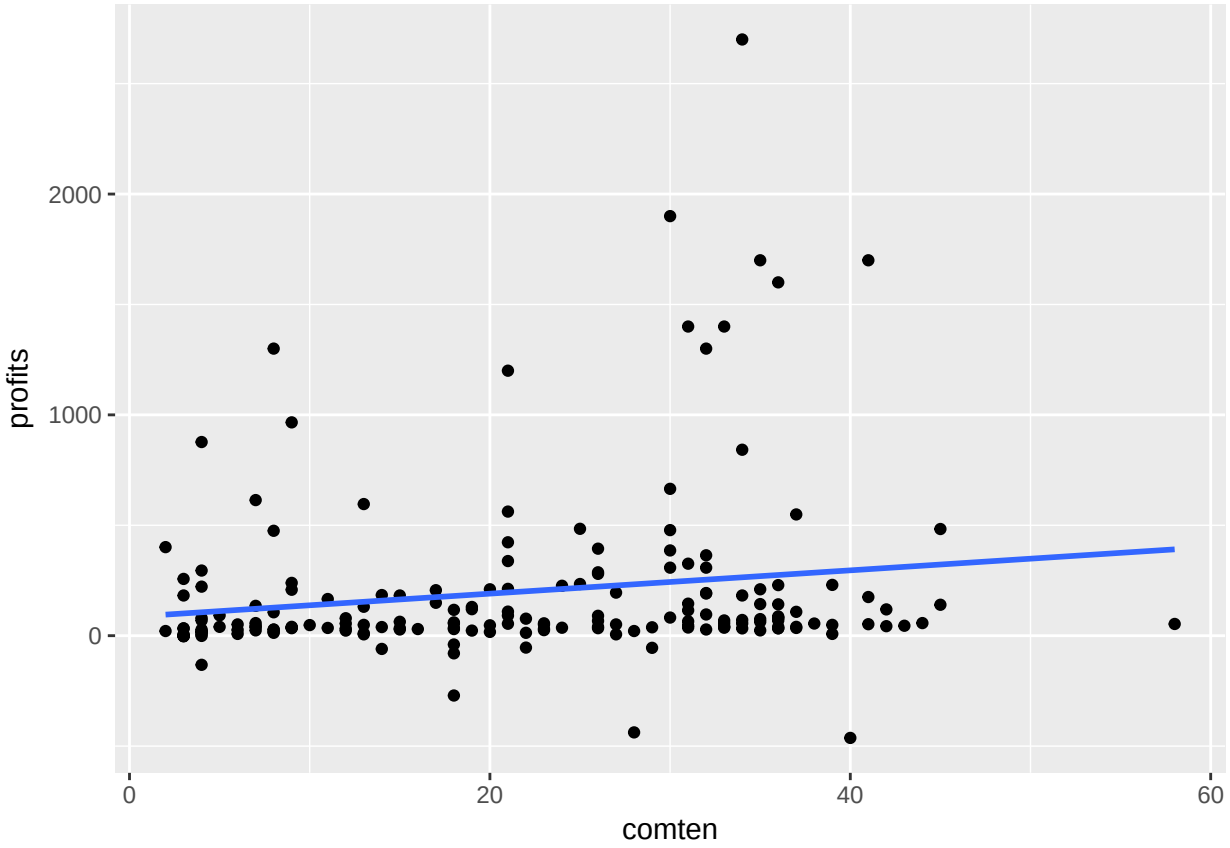
Looking at the relationship between education and profits, we can immediately see that both the 25th and 75th percentile for CEOs with graduate degrees is higher than that of the CEOs without a graduate degree.

Typically, employees with more education will get paid higher than those without education, so perhaps the level of education might have more of an impact on company success than we might expect.

Company Tenure vs Profits

Next, we will examine the correlation between company tenure and profits

```
ggplot(CEO, aes(x = comten, y = profits)) + geom_point() + stat_smooth(method = "lm", se = FALSE)
```



While most data points are clustered closer to the 0 - 1000 range for all company tenure levels, we see that there are more high profits for companies with CEOs with high company tenure, though the same is true for negative profits.

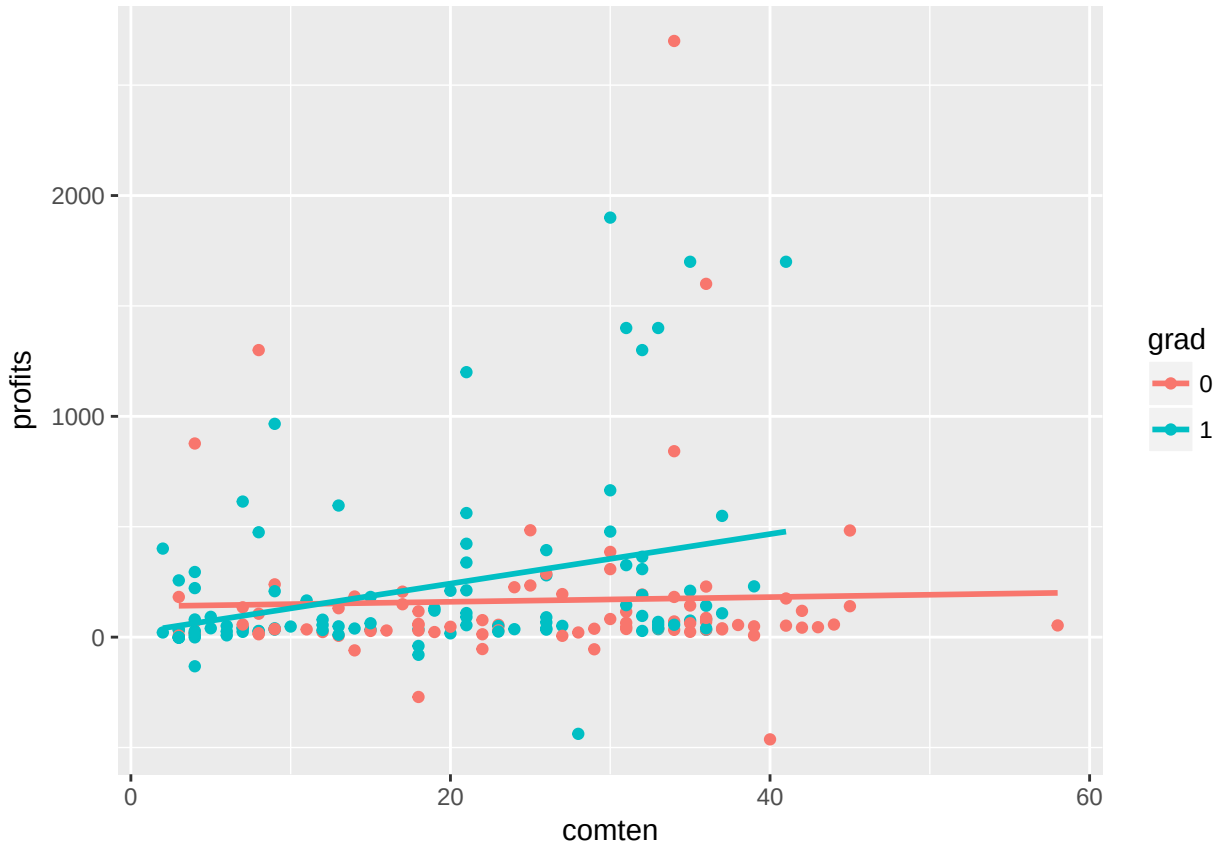
It seems that there might potentially be a weak correlation between company tenure and profits. In fact, the correlation is .1679, confirming that a very weak correlation exists.

```
cor(CEO$comten, CEO$profits)
```

```
## [1] 0.1679311
```

We now explore the how the correlation of company tenure on profits changes based on whether or not the CEO has a graduate degree.

```
ggplot(CEO, aes(x = comten, y = profits, color = grad)) + geom_point() + stat_smooth(method = "lm", se =
```



We can see a very interesting trend here. The group of companies with a CEO who received a graduate-level education had a higher correlation between company tenure and profits.