

Lab 1

Saul Grimaldo, Melwin Poovakottu, Anirudh Mittal

1/18/2017

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

Introduction

We were tasked to determine whether company performance is related to a CEO's salary. To answer this question, we will explore a CEO data set containing the salaries of 185 CEOs.

We operationalize the concept of performance through the use of profits as a proxy for performance.

We will examine whether we can say that a higher salary is related to better company performance.

Data Coding For this analysis, we assumed that if a CEO reported completing a graduate degree, they must have also completed college. This affected 2 data points in the CEO data set.

```
# 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$unaltered_college = CEO$college
CEO$college = ifelse(CEO$grad == 1, 1, CEO$unaltered_college)
```

Furthermore, we encountered cases where market value and profits are set to -1. In the case of profits, this is acceptable, but only 5 data points showed a market value of less than 0. We believe that missing values were encoded as -1, so if profits and market value are both -1 at the same time, we updated both the values to be missing.

```
nrow(subset(CEO, mktval == -1))
```

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

```
CEO$unaltered_mktval = CEO$mktval
CEO$unaltered_profits = CEO$profits
CEO$mktval = ifelse(CEO$unaltered_mktval == -1, NA, CEO$unaltered_mktval)
CEO$profits = ifelse(CEO$unaltered_mktval == -1 & CEO$profits == -1, NA, CEO$profits)
```

Finally, we also noticed that 1 observation exists where ceoten > comten. The data point also belonged to someone who reported being 21 years old, completing both an undergraduate and graduate degree, and being a CEO since he/she was 16. This data point is dubious, so we eliminated it from the data set.

```
print(subset(CEO, ceoten > comten))
```

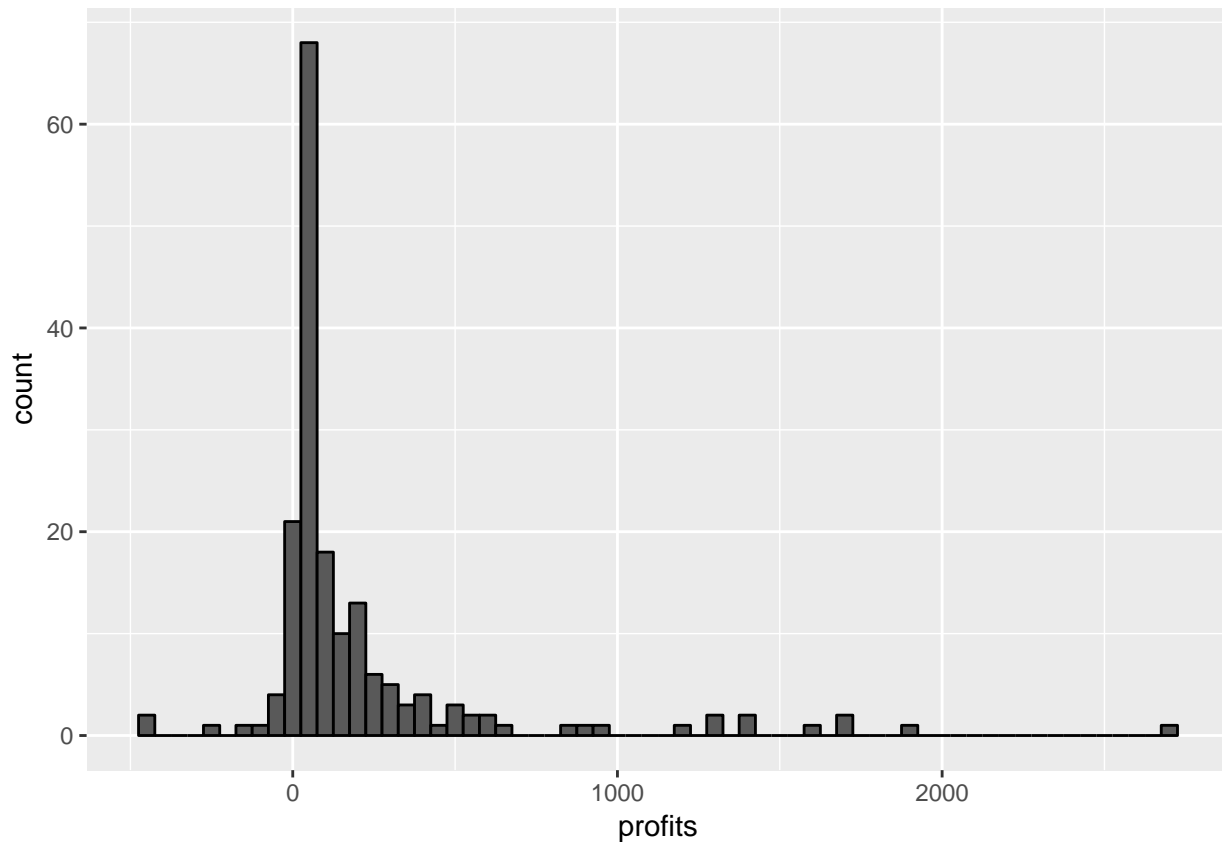
```
##      salary age college grad comten ceoten profits mktval unaltered_college
## 183     877  21      1    1      3      5      -3    303                1
##      unaltered_mktval unaltered_profits
## 183                303                -3
```

```
CEO = subset(CEO, comten >= ceoten)
```

Exploration of profits

We begin our analysis by looking at profits, the variable of interest.

```
ggplot(CEO, aes(x = profits)) +  
  geom_histogram(color=as.factor(1), binwidth = 50, na.rm = T)
```

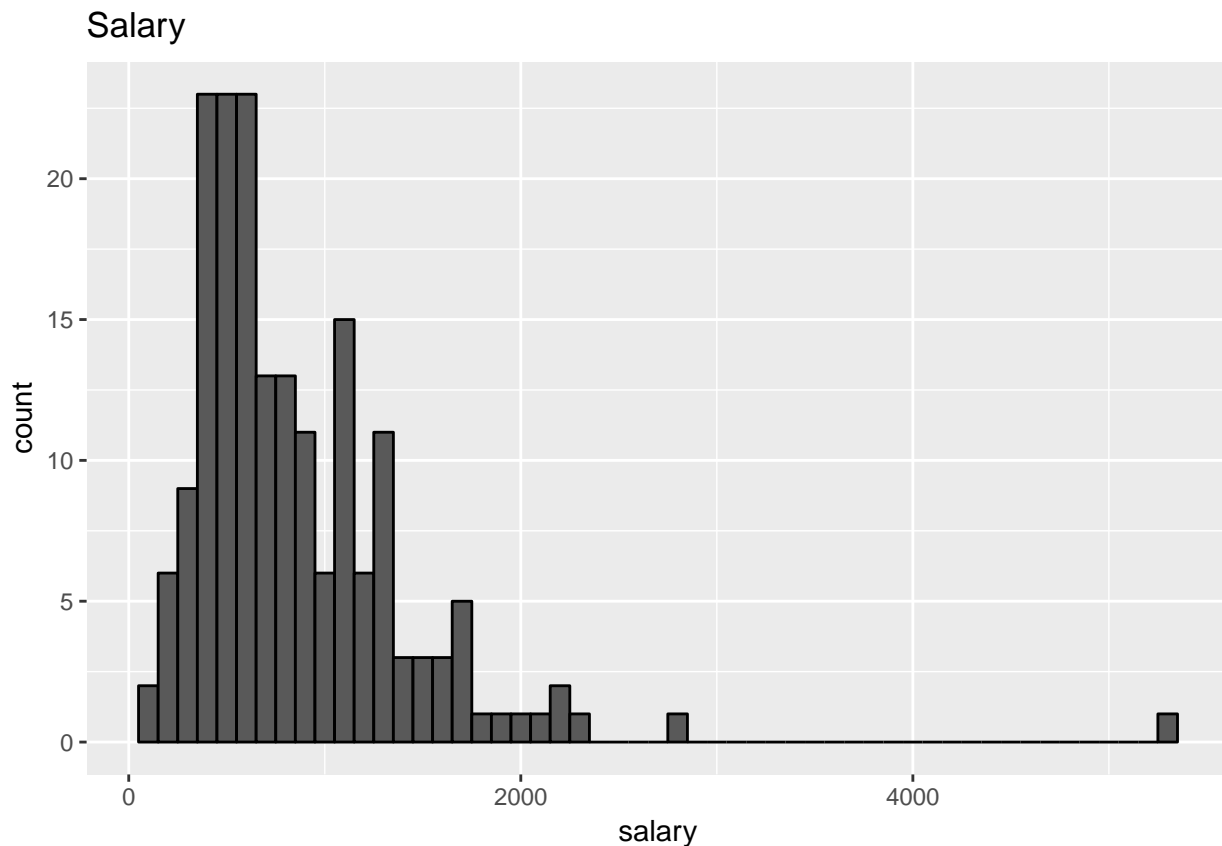


Profits, are highly skewed to the right with the majority of CEO's working at companies making between \$50 million and \$100 million in profits.

Exploration of Salary

We next explore salary.

```
ggplot(CEO,aes(x = salary))+geom_histogram(color=as.factor(1), binwidth = 100) + ggtitle("Salary")
```



We can see that salary, like profits, has a strong right skew. However, data is more dispersed.

```
summary(CEO$salary)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  100.0   466.5   695.5   852.7  1102.0  5299.0
```

We can also see that the salary varies fairly dramatically with the first quartile being as low as \$467,000 and the third quartile being at \$1,101,00

Exploration of Education

Another interesting variable to explore is the education of the CEO.

The college variable is a binary variable. Within this sample, we see that more than 97% 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] 0.9728261
```

Only 5 CEOs in our data set did not report having a college education.

```
data_set_size - number_of_college_CEO
```

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

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)
```

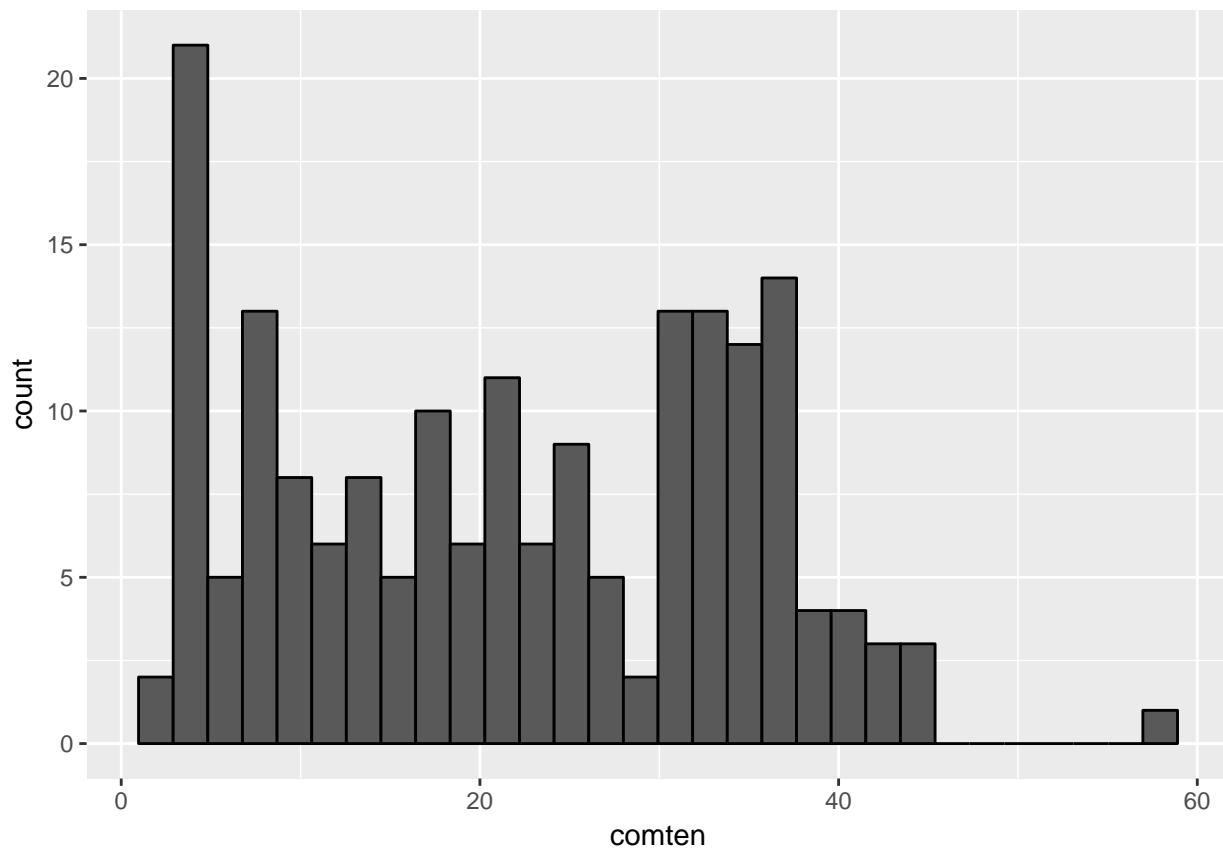
```
## [1] 0.548913
```

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.

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.

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

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

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)
```

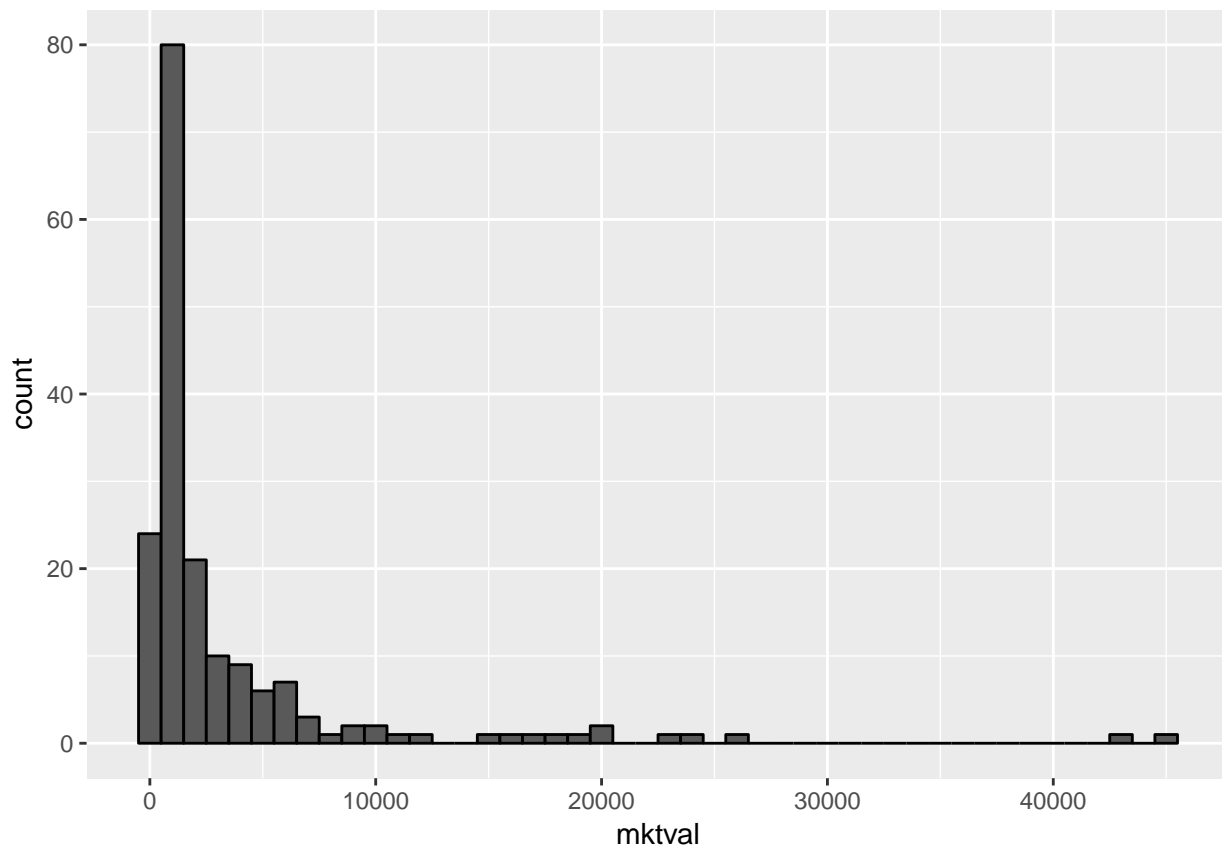
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.00   9.00   21.50   21.77   33.00   58.00
```

Market Value Exploration

We now explore market value.

```
ggplot(CEO,aes(x = mktval))+ geom_histogram(binwidth = 1000, color = as.factor(1))
```

```
## Warning: Removed 5 rows containing non-finite values (stat_bin).
```



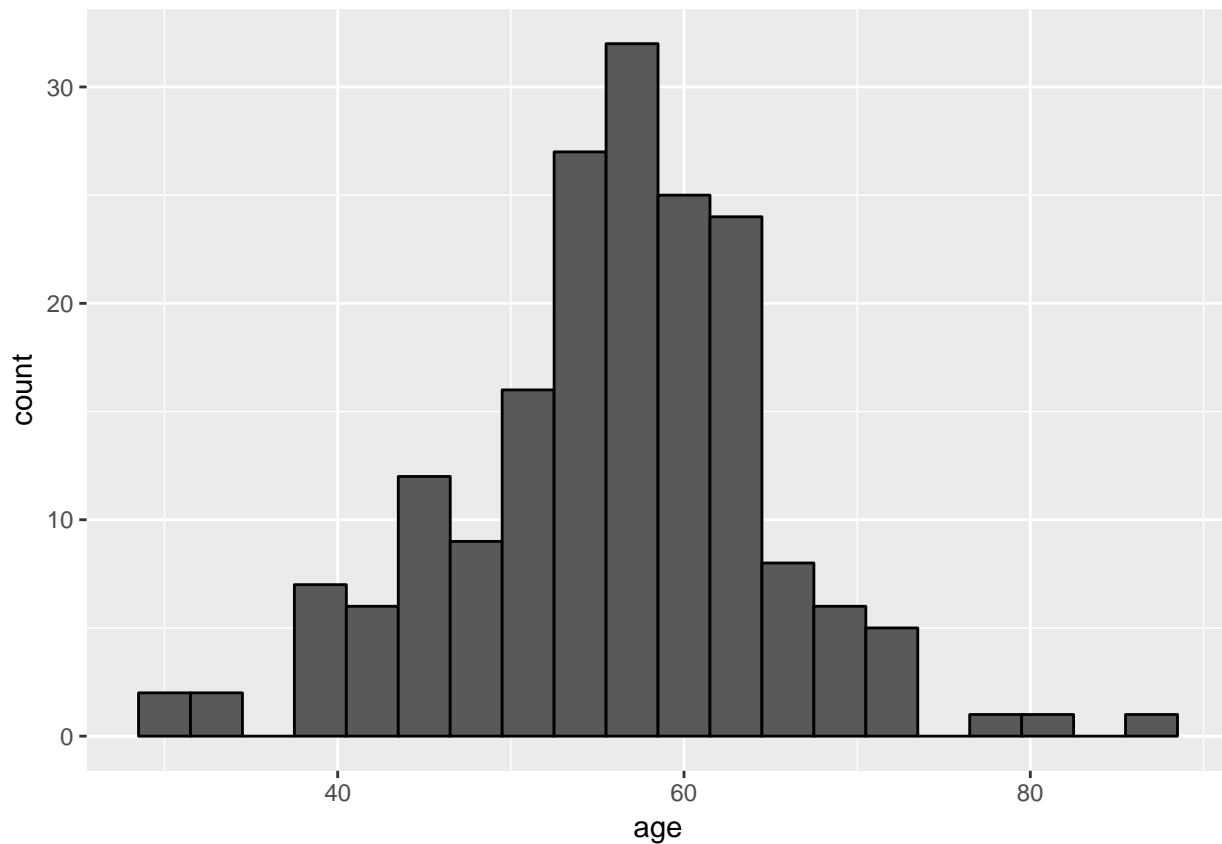
Like salary and profits, we see a strong rightward skew in our data for market value.

In this case, market value, similar to profits, has a large number of data points in the \$1 billion - \$2 billion market value range.

Exploration of Age

Looking at the histogram of age, we can see that the variable is close to a normally distributed curve with the mean(55) and median(57) with a drop after the age of 65, the age of retirement in the US.

```
ggplot(CEO,aes(x = age))+ geom_histogram(binwidth = 3, color = as.factor(1))
```



Bivariate Relationships

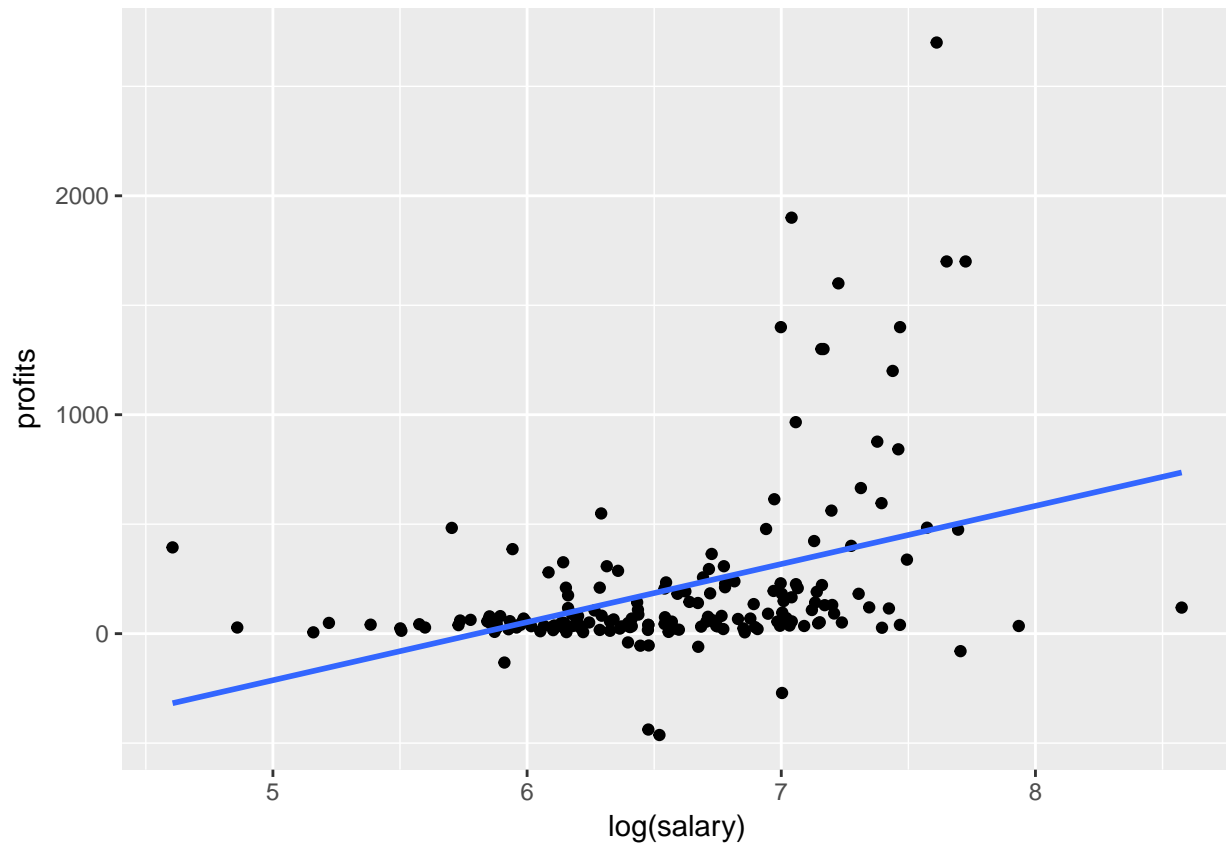
Salary vs Profits

First, we examine the correlation between profits and salary. We transform salary to account for the high level of dispersion in salary values and since it never has a negative value

```
ggplot(CEO,aes(x = log(salary), y = profits)) +  
  geom_point() +stat_smooth(method = "lm", se = F)
```

```
## Warning: Removed 5 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 5 rows containing missing values (geom_point).
```



```
cor(log(CEO$salary), CEO$profits, use = "pairwise.complete.obs")
```

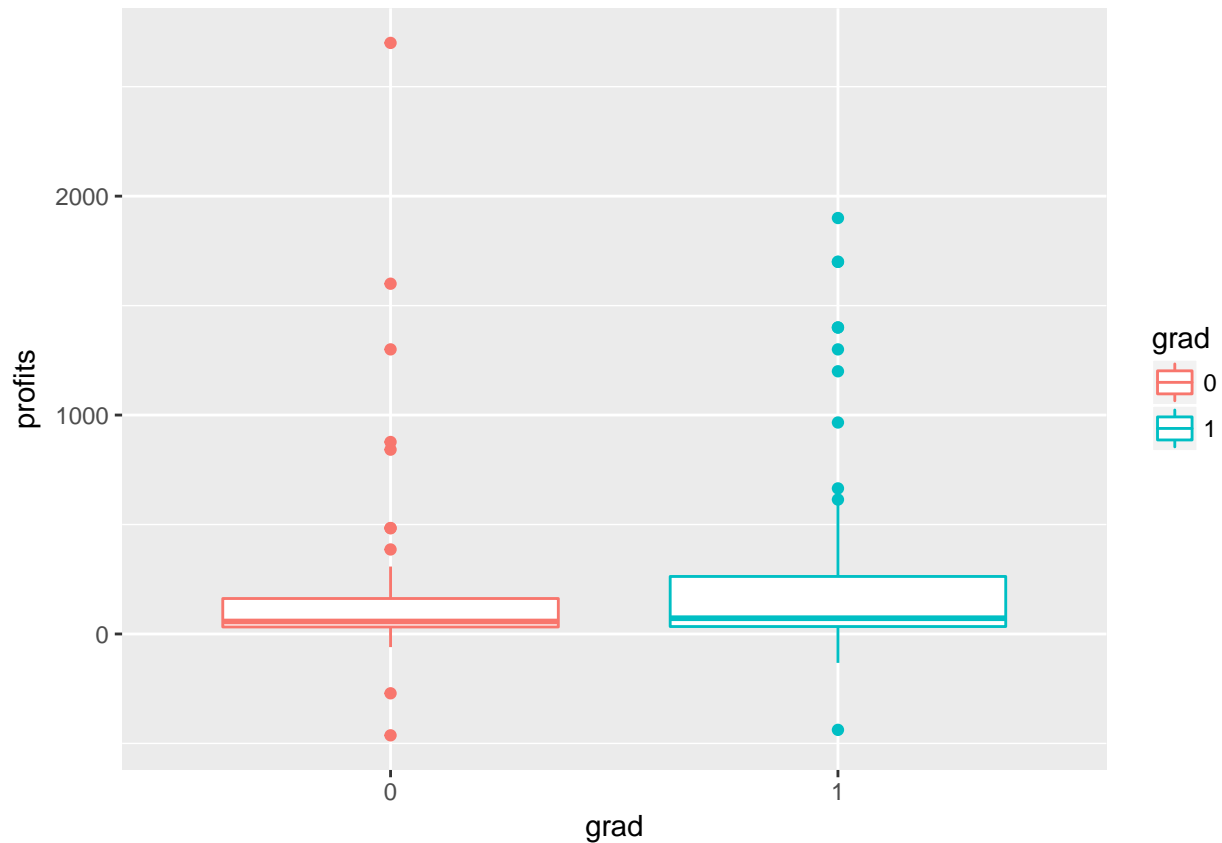
```
## [1] 0.3986899
```

We see that, indeed, salary and profits are relatively well correlated with a correlation coefficient of .3979.

Education vs Profit

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

```
CEO$grad = as.factor(CEO$grad)
ggplot(CEO, aes(x = grad, y = profits, color = grad)) +
  geom_boxplot(na.rm = T)
```

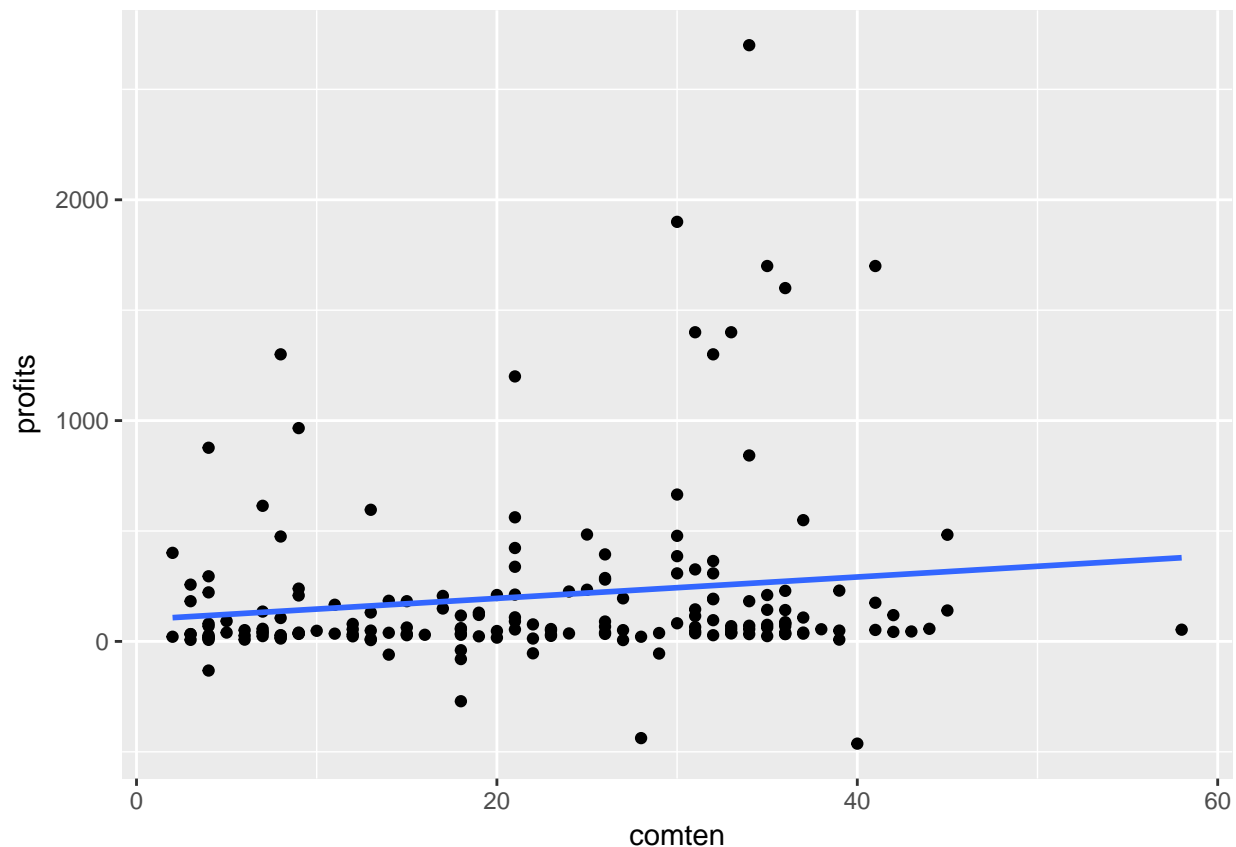


Looking at the relationship between education and profits, we can immediately see that both the 25th and 75th percentile for CEOs with graduate degrees is higher than that of the CEOs without a graduate degree.

Company Tenure vs Profits

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

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

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.

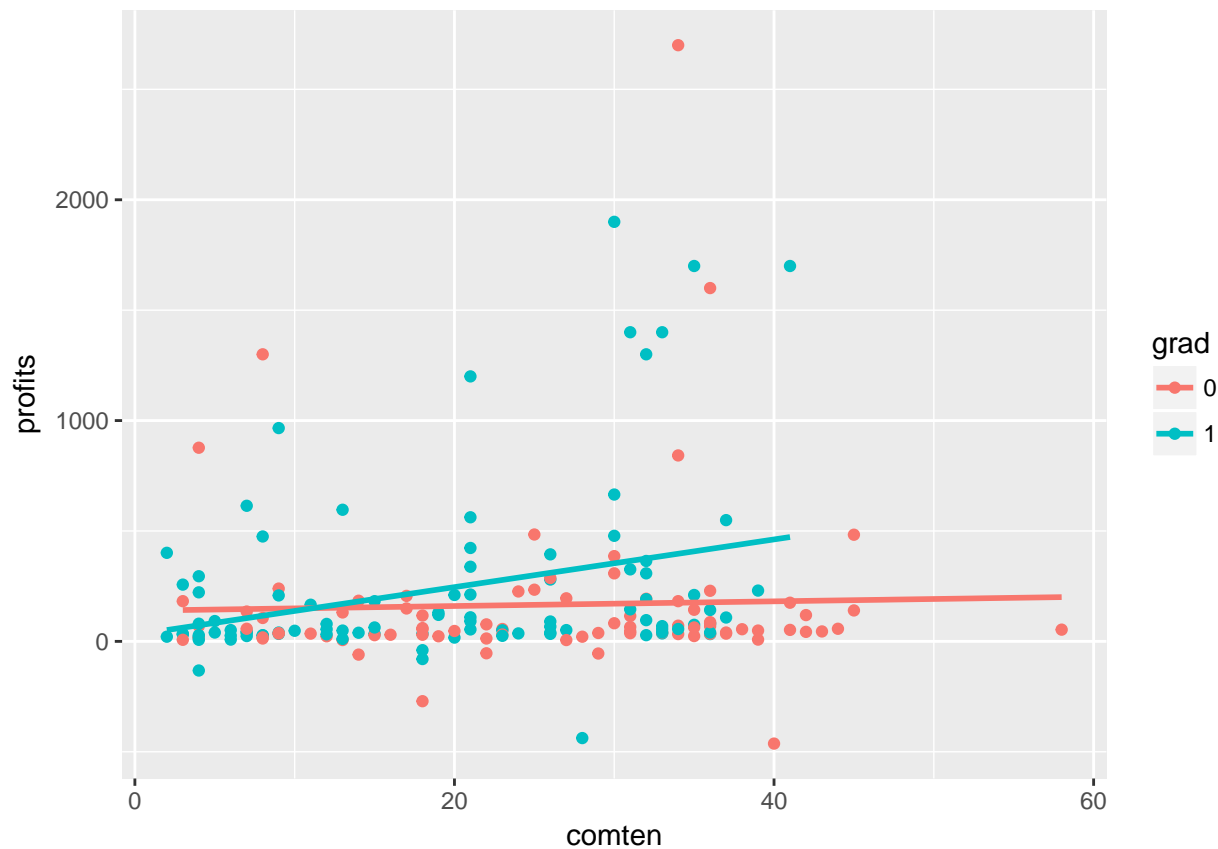
It seems that there might potentially be a weak correlation between company tenure and profits. In fact, the correlation is .1491, confirming that a very weak correlation exists.

```
cor(CEO$comten, CEO$profits, use = "pairwise.complete.obs")
```

```
## [1] 0.1491889
```

We now explore 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(na.rm = T) +  
  stat_smooth(method = "lm", se = FALSE, na.rm = T)
```

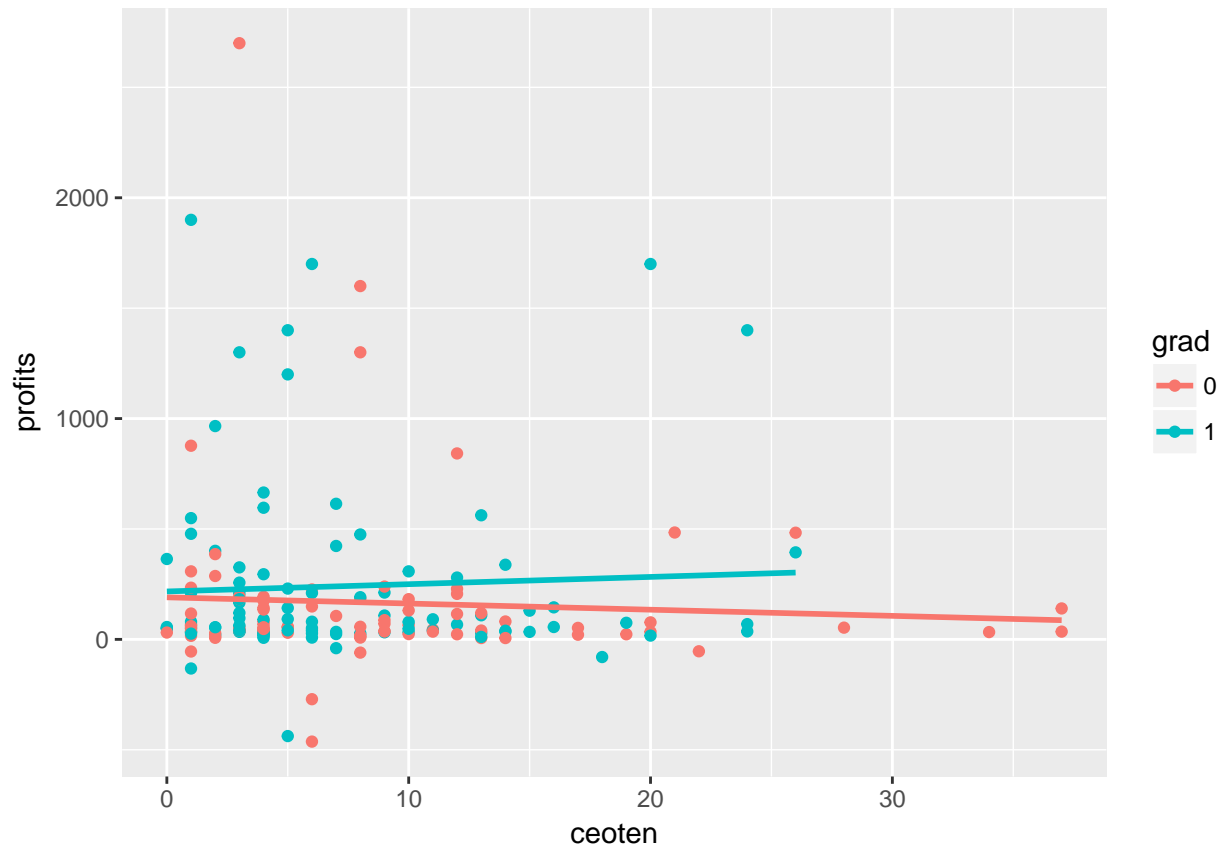


We can see a very interesting trend here. The group of companies with a CEO who received a graduate-level education had a higher correlation between company tenure and profits.

Tenure as CEO vs profits

Next, we see the relationship between tenure as CEO vs profits.

```
ggplot(CEO, aes(x = ceoten, y = profits, color = grad)) +
  geom_point(na.rm = T) + stat_smooth(method = "lm", se = F, na.rm = T)
```

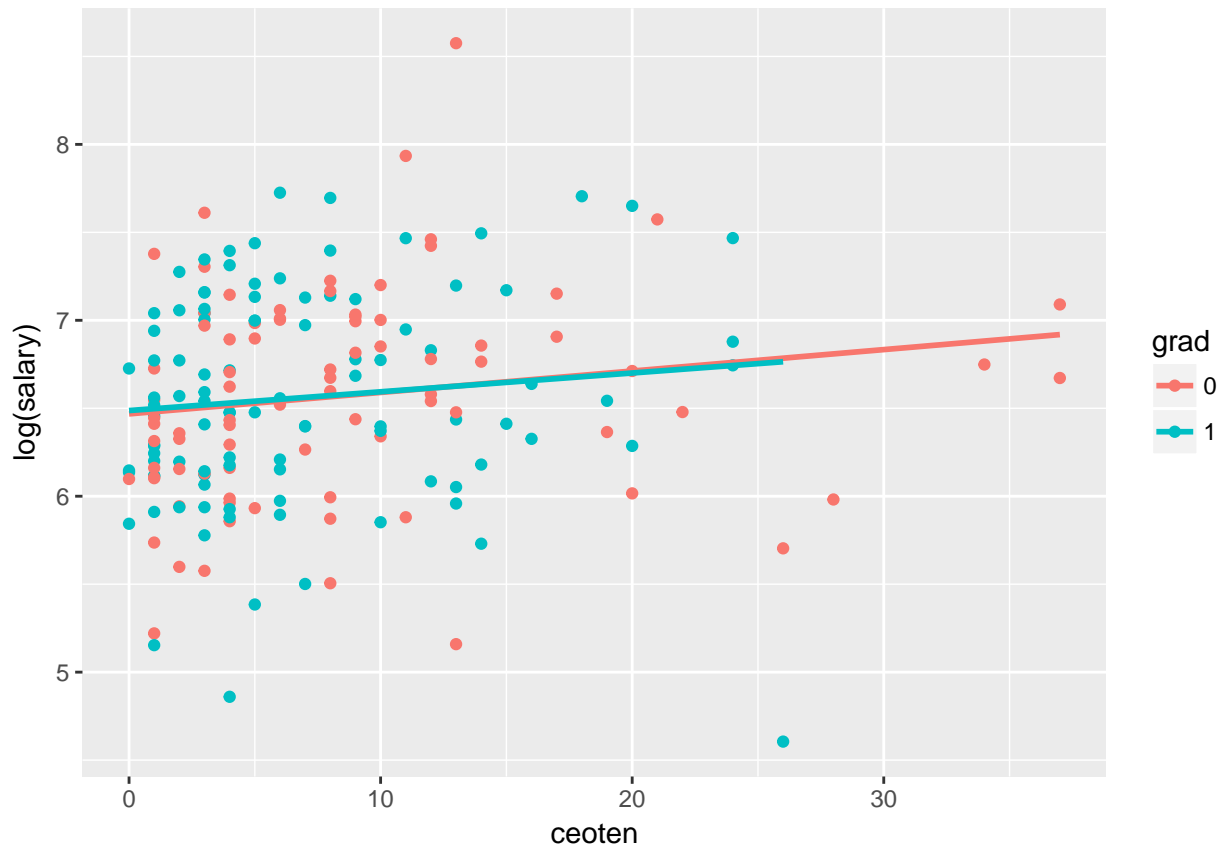


Interestingly enough, we see that CEO tenure and profits are more strongly correlated when the CEO has a graduate degree than when they don't. In fact we also see that there is a negative correlation between tenure as CEO vs profits.

Tenure as CEO vs Salary

Next we also explore the relationship between CEO and salary. Because of the level of dispersion between tenure as CEO and salary, we perform a log transformation of salary.

```
ggplot(CEO, aes(x = ceoten, y = log(salary), color = grad)) +
  geom_point() + stat_smooth(method = "lm", se = F)
```

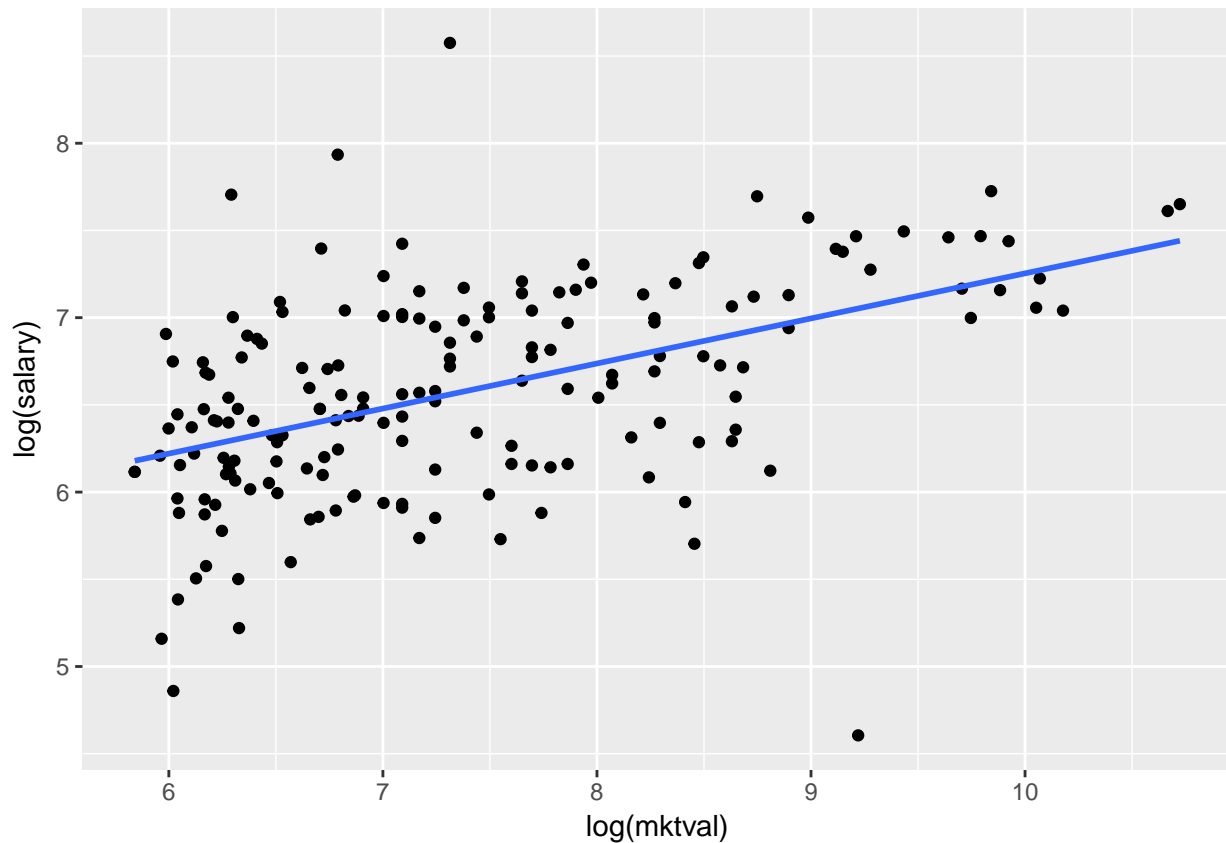


We see that there is a fairly strong linear relationship between tenure as CEO and salary, and that the correlation is roughly similar regardless of education level.

Market Value vs Salary

As we showed earlier, both market value and salary have similar distributions that are strongly right skewed. Both are also always positive, so we will perform log transformations to market value and salary.

```
ggplot(CEO, aes(x = log(mktval), y = log(salary))) +
  geom_point(na.rm = T) + stat_smooth(na.rm = T, method = "lm", se = F)
```



We see a fairly strong correlation between the log of market value and the log of salary, and in fact, with a correlation coefficient of .48, we see that indeed there is a correlation between the two variables.

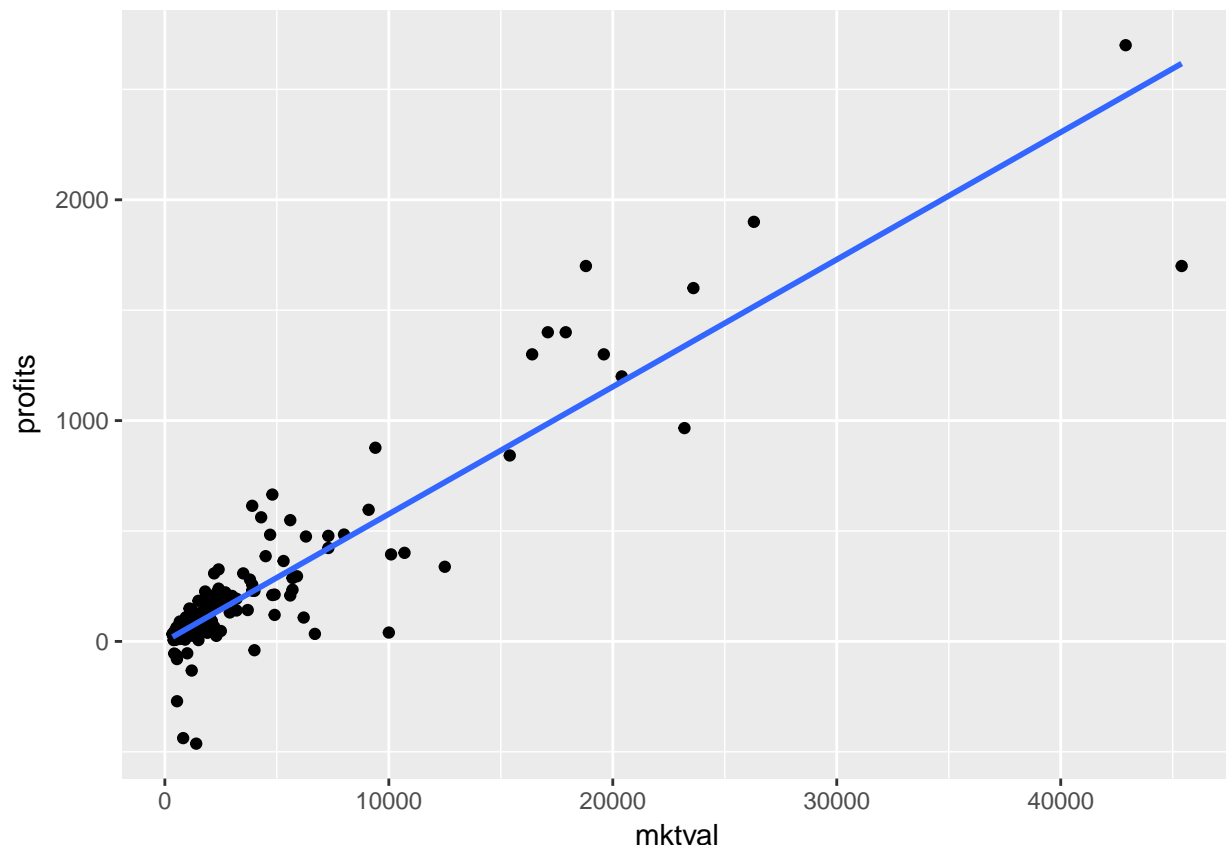
```
cor(log(CEO$mktval), log(CEO$salary), use = "pairwise.complete.obs")
```

```
## [1] 0.4866164
```

Market Value vs Profits

Finally, we explore the correlation between market value and profits. Since profits can be negative, no transformation is performed on it, but it is performed on market value.

```
ggplot(CEO, aes(x = mktval, y = profits)) +  
  geom_point(na.rm = T) + stat_smooth(na.rm = T, method = "lm", se = F)
```



We see a strong linear relationship between the market value of profits, and indeed with a correlation coefficient of .91, this is the strongest correlation we've seen in this data.

```
cor(CEO$mktval, CEO$profits, use = "pairwise.complete.obs")
```

```
## [1] 0.918301
```

Confounding Effects

Correlation is often confused for causation, so it is important to correctly assess what the data is saying. For instance, we saw that salary is correlated with years of experience, market value of the company, and in deed, profits.

However, we also saw that profits is correlated with company tenure, education, market value, and profits.

Since multiple correlations exist, it is very possible that for instance, higher education and experience combined allow a CEO to learn how businesses run effectively and thus lead to higher profits, and because the educated and experienced CEO has a history of showing good results, they would demand a higher salary.

Alternatively, it is also possible that simply having a higher salary would motivate a CEO to perform better.

Another alternative could be that companies with high profits tend to have the budget to hire a CEO with a good performance record and pay him/her a very competetive salary.

To understand the true nature of these correlations, it would be necessary to use more advanced modeling techniques.

Conculsion

As explained in the previous section of this report, without a deeper analysis of the data set, it is difficult to conclusively say what the true nature of the relationship between salary and profits other than that they are correlated with each other.