

# EDA - Corporate Governance and Company Performance

Conor Reid

28/5/2017

Set the working directory and read in data as “spx” (not shown) and load in the required packages.

```
library(reshape)
library(stringr)
library(rpart)
library(lattice)
library(ggplot2)
vars = c("Ticker", "Tobin.s.Q", "AZS", "P.E", "EPS", "P.B", "P.EBITDA",
        "Board.Size", "CEO.Duality", "X..Feml.Execs",
        "X..Feml.Execs.1", "Bd.Avg.Age", "Board.Mtg.Att..",
        "Asset", "Interest", "Tax", "ROE", "ROC",
        "Indep.Directors", "X..Indep.Dir.on.Aud.Cmte", "X..Indep.Dir.on.Aud.Cmte.1", "Sz.Aud.Cmte")
```

## S&P 500

The S&P 500 is an American stock market index including market capitalization of 500 companies, listed on the NYSE or the NASDAQ. It is distinct from other indexes such as the Dow Jones etc due to its diverse constituency and weighting methodology. *Would it be useful to look at other index's?*

Read in full data as below, and carry out some sub-setting. Set NA values.

```
spx_EDA = spx[vars]
spx_EDA[spx_EDA == "#N/A Field Not Applicable" ] <- NA
```

Create categorical variable for Q and Z score, as per MM.

```
m_TQ = median(subset(spx_EDA$Tobin.s.Q, !is.na(spx_EDA$Tobin.s.Q)))
m_AZ = median(subset(spx_EDA$AZS, !is.na(spx_EDA$AZS)))
spx_EDA$Tobin.s.Q.class = with(spx_EDA, ifelse(Tobin.s.Q >= m_TQ, 1, 2))
spx_EDA$AZS.class = with(spx_EDA, ifelse(AZS >= m_AZ, 1, 2))
```

View data types. Convert to numeric columns that need to be numeric. Parse out the actual ticker string (output hidden for clarity).

```
sapply(spx_EDA,class)
cols.num <- c("P.B","EPS","P.E","P.EBITDA")
spx_EDA[cols.num] <- sapply(spx_EDA[cols.num],as.numeric)
sapply(spx_EDA, class)
spx_EDA$TickerID = str_split_fixed(spx_EDA$Ticker, " ", 2)[,1] #parse the ticker itself
```

## Missing Data

Cases with missing values. Very high (254 / 500 = 50.8% loss if we discount them).

```
sum(!complete.cases(spx_EDA))
```

```
## [1] 254
```

81 NA's for Altman Z score, (81 / 500 = 16.2%).

```
summary(spx_EDA)
```

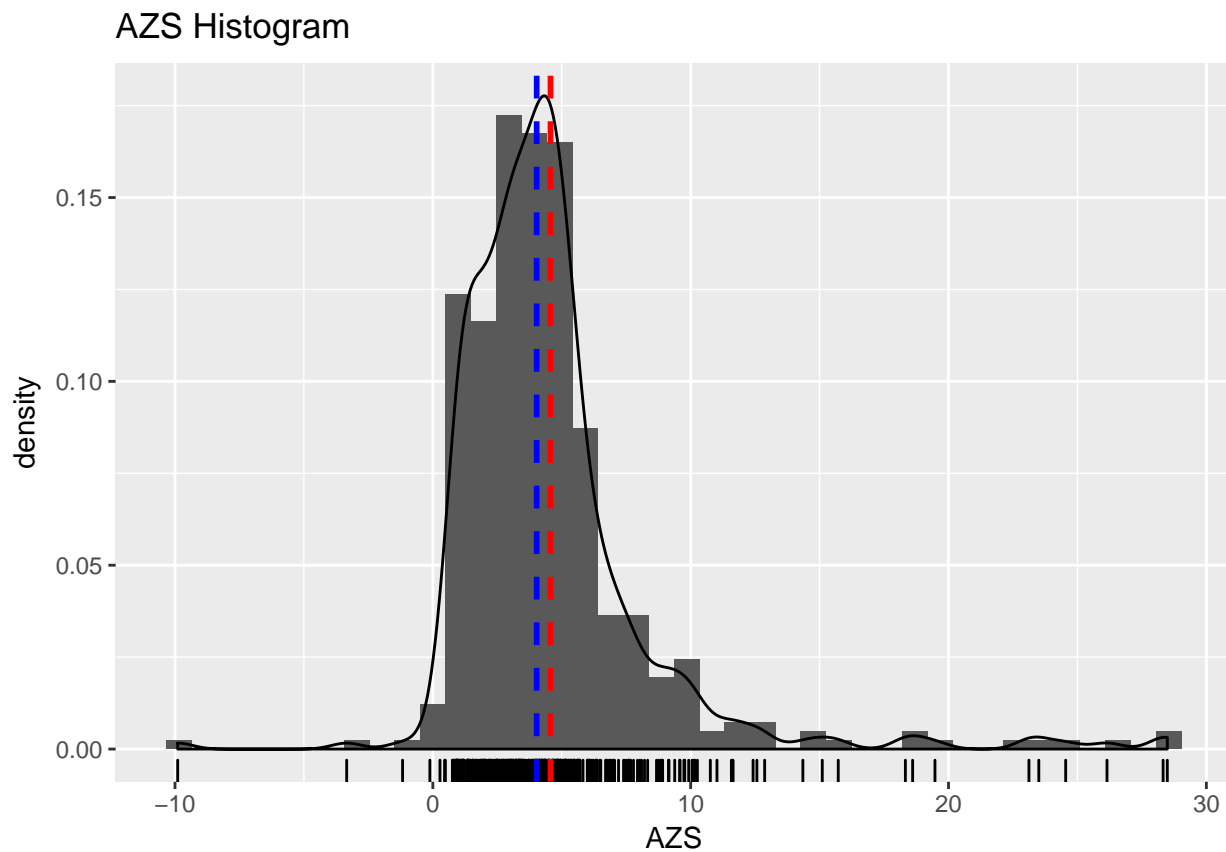
```
##           Ticker      Tobin.s.Q      AZS
## A      UN Equity: 1  Min.   :0.8268  Min.   :-9.896
## AA     UN Equity: 1  1st Qu.:1.2672  1st Qu.: 2.450
## AAPL   UW Equity: 1  Median  :1.7495  Median  : 4.026
## ABBV   UN Equity: 1  Mean    :2.1761  Mean    : 4.562
## ABC    UN Equity: 1  3rd Qu.:2.5184  3rd Qu.: 5.372
## ABT    UN Equity: 1  Max.    :9.6363  Max.    :28.484
## (Other)      :494  NA's    :4      NA's    :81
##      P.E      EPS      P.B      P.EBITDA
## Min.   : 2.0   Min.   : 1.00   Min.   : 2.0   Min.   : 2.0
## 1st Qu.:124.2   1st Qu.: 98.75   1st Qu.:123.8   1st Qu.:115.5
## Median :246.5   Median :181.50   Median :245.5   Median :229.0
## Mean   :246.5   Mean   :183.82   Mean   :245.5   Mean   :229.0
## 3rd Qu.:368.8   3rd Qu.:266.25   3rd Qu.:367.2   3rd Qu.:342.5
## Max.   :491.0   Max.   :374.00   Max.   :489.0   Max.   :456.0
## NA's   :10     NA's   :4      NA's   :12     NA's   :45
##      Board.Size CEO.Duality X..Feml.Execs X..Feml.Execs.1
## Min.   : 5.00   N      :225   Min.   : 0.00   Min.   :0.00
## 1st Qu.:10.00   Y      :272   1st Qu.: 0.00   1st Qu.:0.00
## Median :11.00   NA's:   3     Median :12.50   Median :1.00
## Mean   :10.98           Mean   :13.53   Mean   :1.33
## 3rd Qu.:12.00           3rd Qu.:20.00   3rd Qu.:2.00
## Max.   :29.00           Max.   :62.50   Max.   :8.00
## NA's   :3           NA's   :3      NA's   :3
##      Bd.Avg.Age Board.Mtg.Att.. Asset Interest
## Min.   :40.33   Min.   : 72.15   Min.   :0.03398   Min.   : -52.96
## 1st Qu.:60.85   1st Qu.: 75.00   1st Qu.:0.34901   1st Qu.: 84.63
## Median :63.00   Median : 75.00   Median :0.61024   Median : 92.21
## Mean   :62.80   Mean   : 79.88   Mean   :0.79602   Mean   : 99.11
## 3rd Qu.:64.91   3rd Qu.: 80.00   3rd Qu.:0.94005   3rd Qu.: 98.89
## Max.   :73.21   Max.   :100.00   Max.   :5.91183   Max.   :3706.66
## NA's   :3      NA's   :14     NA's   :3      NA's   :73
##      Tax      ROE      ROC      Indep.Directors
## Min.   : -498.04   Min.   : -42.764   Min.   : -17.959   Min.   : 4.000
## 1st Qu.: 63.23   1st Qu.: 9.661   1st Qu.: 6.425   1st Qu.: 8.000
## Median : 69.02   Median : 15.716   Median : 10.343   Median : 9.000
## Mean   : 86.91   Mean   : 25.190   Mean   : 12.103   Mean   : 9.302
## 3rd Qu.: 77.82   3rd Qu.: 24.001   3rd Qu.: 14.813   3rd Qu.:11.000
## Max.   :3127.97   Max.   :1741.641   Max.   : 86.278   Max.   :23.000
## NA's   :2      NA's   :12     NA's   :167     NA's   :4
##      X..Indep.Dir.on.Aud.Cmte X..Indep.Dir.on.Aud.Cmte.1 Sz.Aud.Cmte
## Min.   :2.000           Min.   : 66.67           Min.   :3.000
## 1st Qu.:4.000           1st Qu.:100.00          1st Qu.:4.000
## Median :4.000           Median :100.00          Median :4.000
```

##	Mean	:4.354	Mean	: 99.88	Mean	:4.359
##	3rd Qu.	:5.000	3rd Qu.	:100.00	3rd Qu.	:5.000
##	Max.	:9.000	Max.	:100.00	Max.	:9.000
##	NA's	:3	NA's	:4	NA's	:4
##	Tobin.s.Q.class	AZS.class	TickerID			
##	Min.	:1.0	Min.	:1.000	Length:500	
##	1st Qu.	:1.0	1st Qu.	:1.000	Class :character	
##	Median	:1.5	Median	:1.000	Mode :character	
##	Mean	:1.5	Mean	:1.499		
##	3rd Qu.	:2.0	3rd Qu.	:2.000		
##	Max.	:2.0	Max.	:2.000		
##	NA's	:4	NA's	:81		

## Altman Z Score

Histogram of the Altman Z Score over the data. Looks to be a pretty normal distribution, albeit with some extreme outliers on either tail.

```
ggplot(data=spx_EDA) +  
  geom_histogram( aes(AZS, ..density..), bins=40 ) +  
  geom_density( aes(AZS, ..density..) ) +  
  geom_rug( aes(AZS) ) +  
  geom_vline(aes(xintercept=mean(AZS, na.rm=T)), # red line is the mean  
             color="red", linetype="dashed", size=1) +  
  geom_vline(aes(xintercept=median(AZS, na.rm=T)), # blue line is the median  
             color="blue", linetype="dashed", size=1) +  
  ggtitle("AZS Histogram")
```



## Independent Directors and Altman Z score

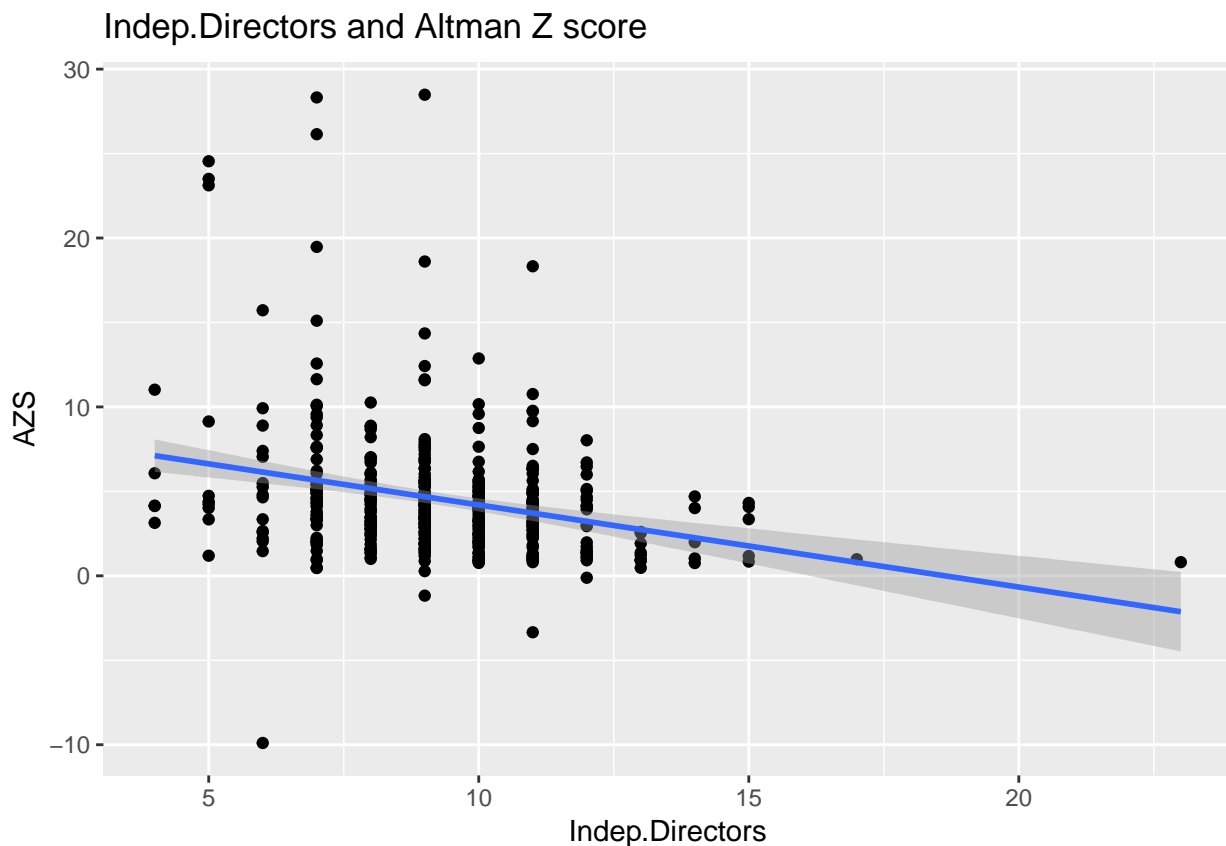
According to M&M;

...while an independent lead director and a financial leverage higher than 2.5 generates a higher risk of bankruptcy.

Risk of bankruptcy is measured by the Z score.

Plot below seems to support this, is it strong enough though? R squared value likely to be pretty small when a proper regression is done.

```
ggplot(data=subset(spx_EDA, !is.na(Indep.Directors)),
  aes(Indep.Directors, AZS )) +
  geom_point() +
  ggtitle("Indep.Directors and Altman Z score") +
  geom_smooth(method = "lm")
```



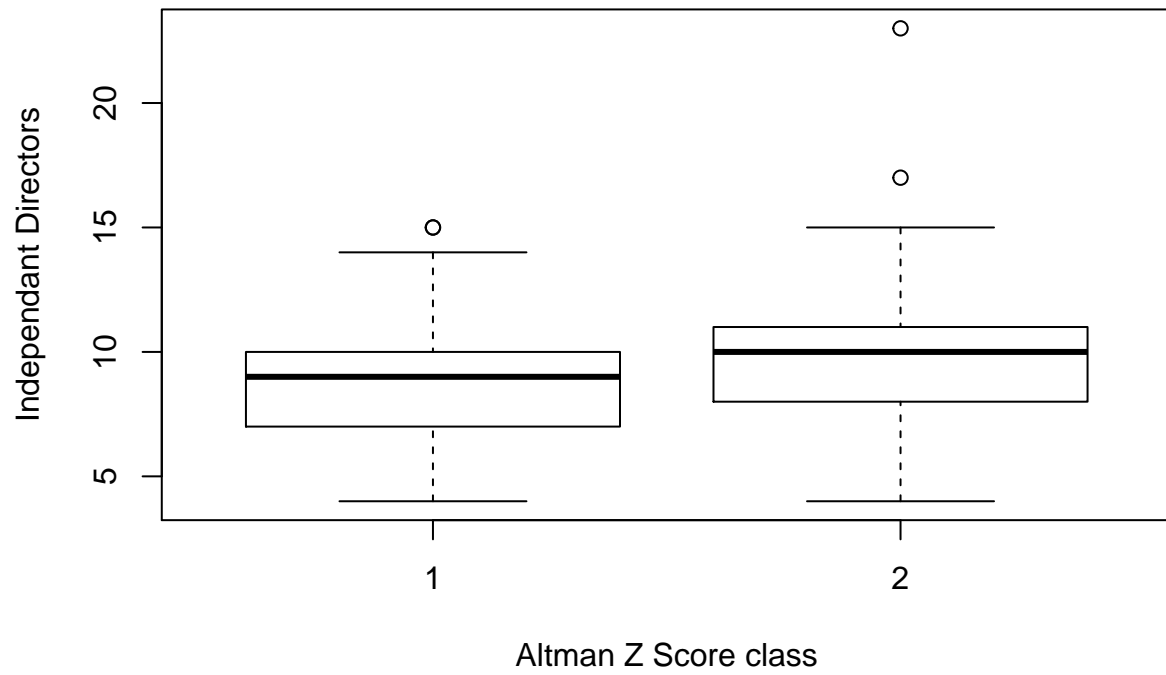
```
table(spx_EDA$AZS.class)
```

```
##
##    1    2
## 210 209
```

Using the class variable for Z score. Not a significant relationship.

```
boxplot(Indep.Directors~AZS.class,data=spx_EDA, main="Altman Z and Number of Indep Directors",
  xlab="Altman Z Score class", ylab="Indepandant Directors")
```

### Altman Z and Number of Indep Directors



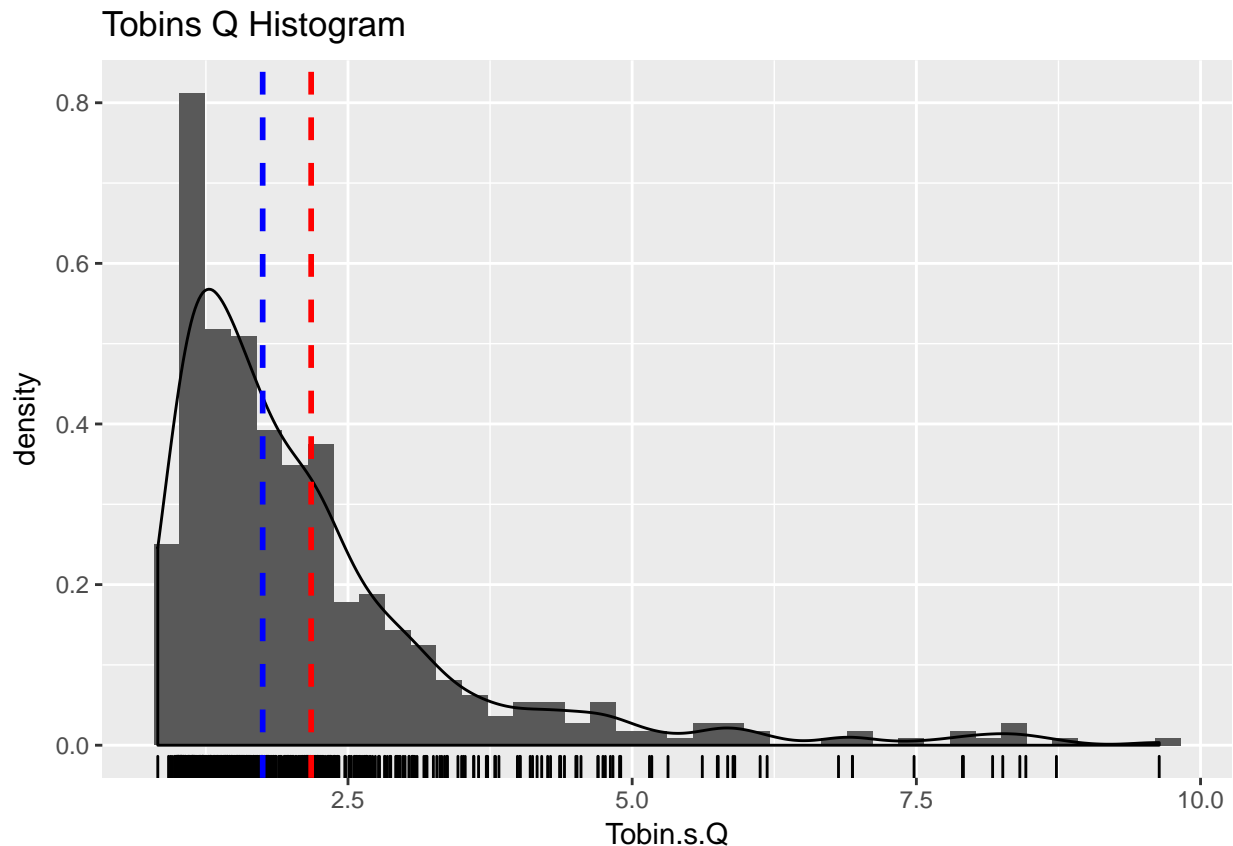
## Tobins Q

Histogram of Tobins Q over the data. Heavily skewed right, most companies have low scores. It is going to be difficult to learn rules for high performing companies? From M&M:

As suggested by Creamer [14], we discretized Tobin's Q ratio in order to obtain two classes, dividing each dataset according to its median value. In this way, a company that lies in the upper side of the median will be looked positively by the machine learning algorithms.

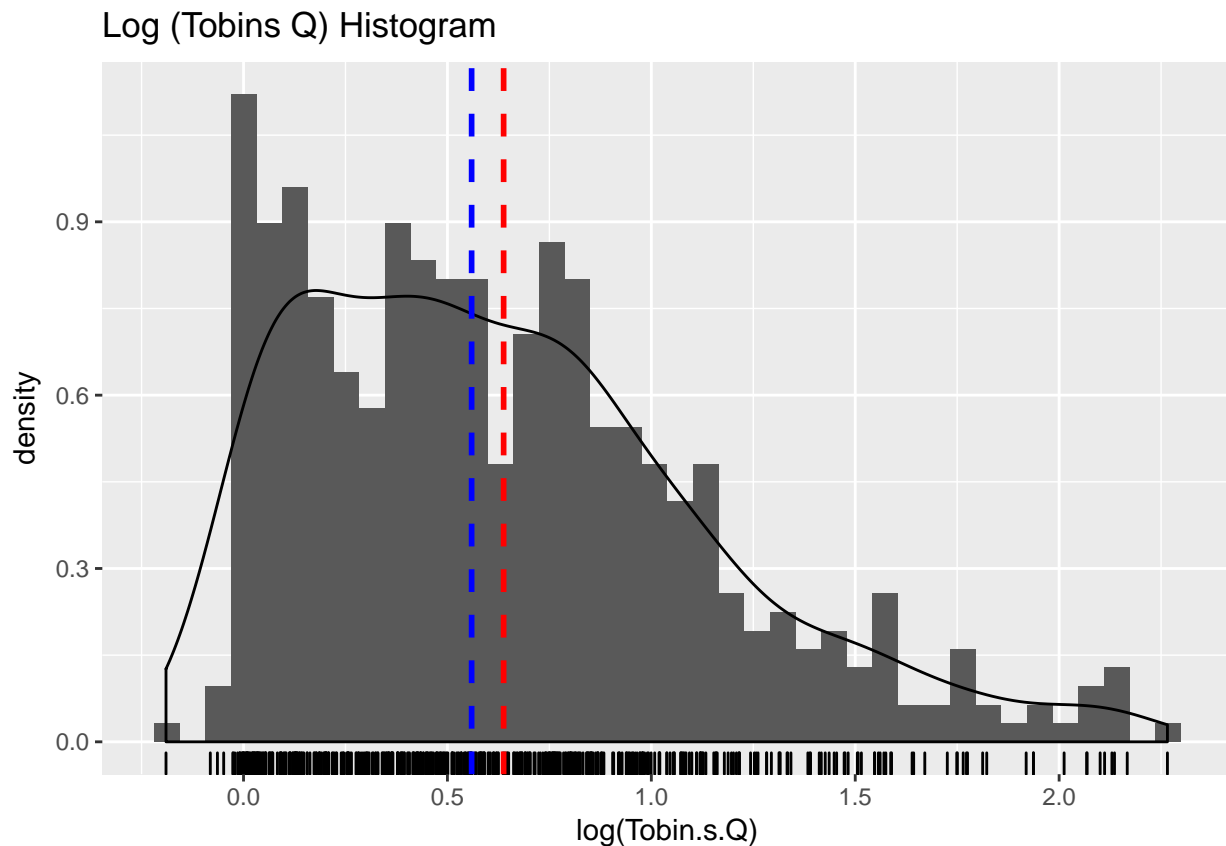
Median (blue) and mean (red) of Q score is shown below. Regression may not be suitable for data that has such a skewness.

```
ggplot(data=spx_EDA) +  
  geom_histogram( aes(Tobin.s.Q, ..density..), bins=40 ) +  
  geom_density( aes(Tobin.s.Q, ..density..) ) +  
  geom_rug( aes(Tobin.s.Q) ) +  
  geom_vline(aes(xintercept=mean(Tobin.s.Q, na.rm=T)), # red line is the mean  
             color="red", linetype="dashed", size=1) +  
  geom_vline(aes(xintercept=median(Tobin.s.Q, na.rm=T)), # blue line is the median  
             color="blue", linetype="dashed", size=1) +  
  ggtitle("Tobins Q Histogram")
```



Try taking the log of Tobins Q instead, histogram shown below. The code is emitted since its the same as previous page but used `log(Tobin.s.Q)` instead. Much less skewed, making regression more applicable? Is it OK to process the dependent variable like this?

```
ggplot(data=spx_EDA) +
  geom_histogram( aes(log(Tobin.s.Q), ..density..), bins=40 ) +
  geom_density( aes(log(Tobin.s.Q), ..density..) ) +
  geom_rug( aes(log(Tobin.s.Q)) ) +
  geom_vline(aes(xintercept=mean(log(Tobin.s.Q), na.rm=T)), # red line is the mean
    color="red", linetype="dashed", size=1) +
  geom_vline(aes(xintercept=median(log(Tobin.s.Q), na.rm=T)), # blue line is the median
    color="blue", linetype="dashed", size=1) +
  ggtitle("Log (Tobins Q) Histogram")
```



Poisson regression could be a good alternative to having to process the Q score directly, obviously assumes the dependent variable has a Poisson distribution.

See [https://en.wikipedia.org/wiki/Poisson\\_distribution](https://en.wikipedia.org/wiki/Poisson_distribution)

I don't think this it actually does, but Poisson regression may be more applicable here that OLS for example.

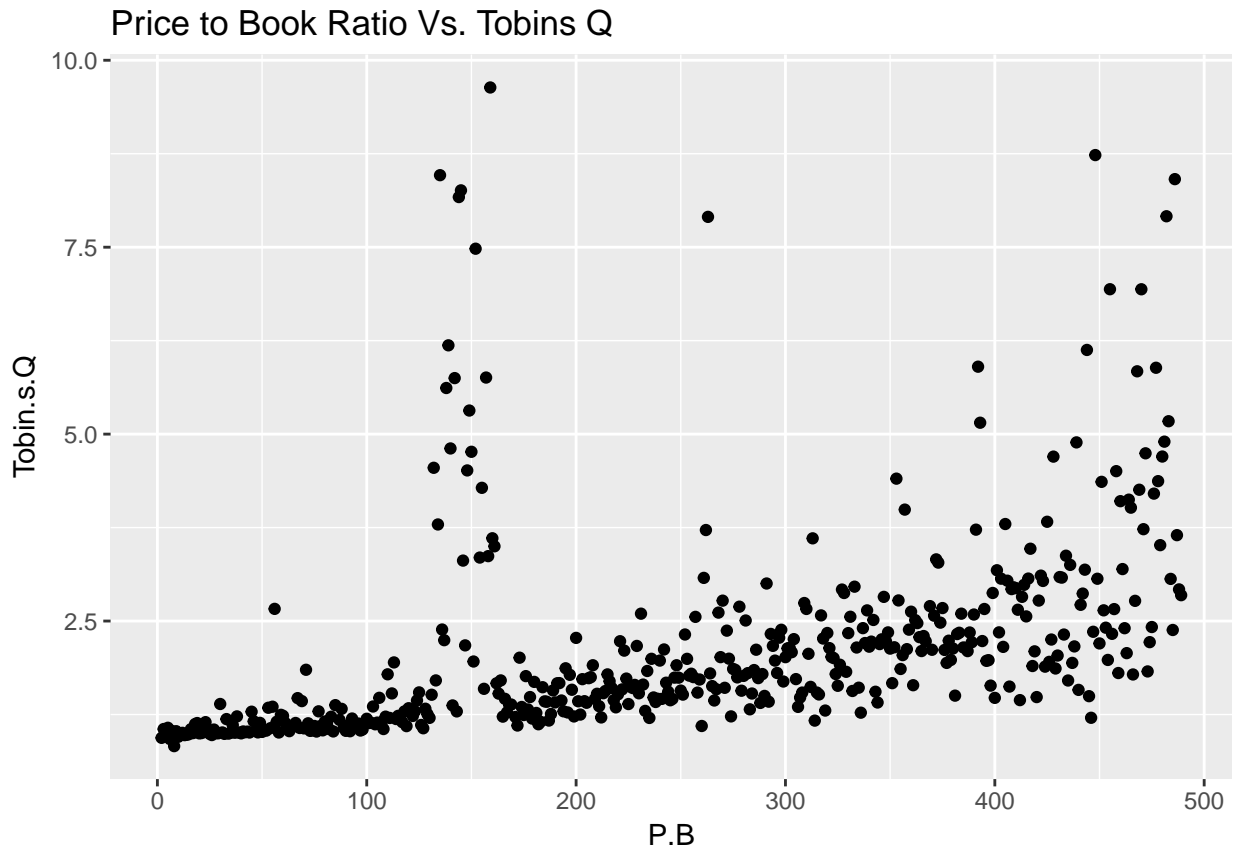


## P.B (price to book ratio) and Q Score

A high P.B means that investors see the company as growing, low means they think the company's book value is exaggerated (<http://www.investopedia.com/terms/p/price-to-bookratio.asp>).

Interesting spike in Q score around the 150 P.B mark. Maybe worth further investigation?

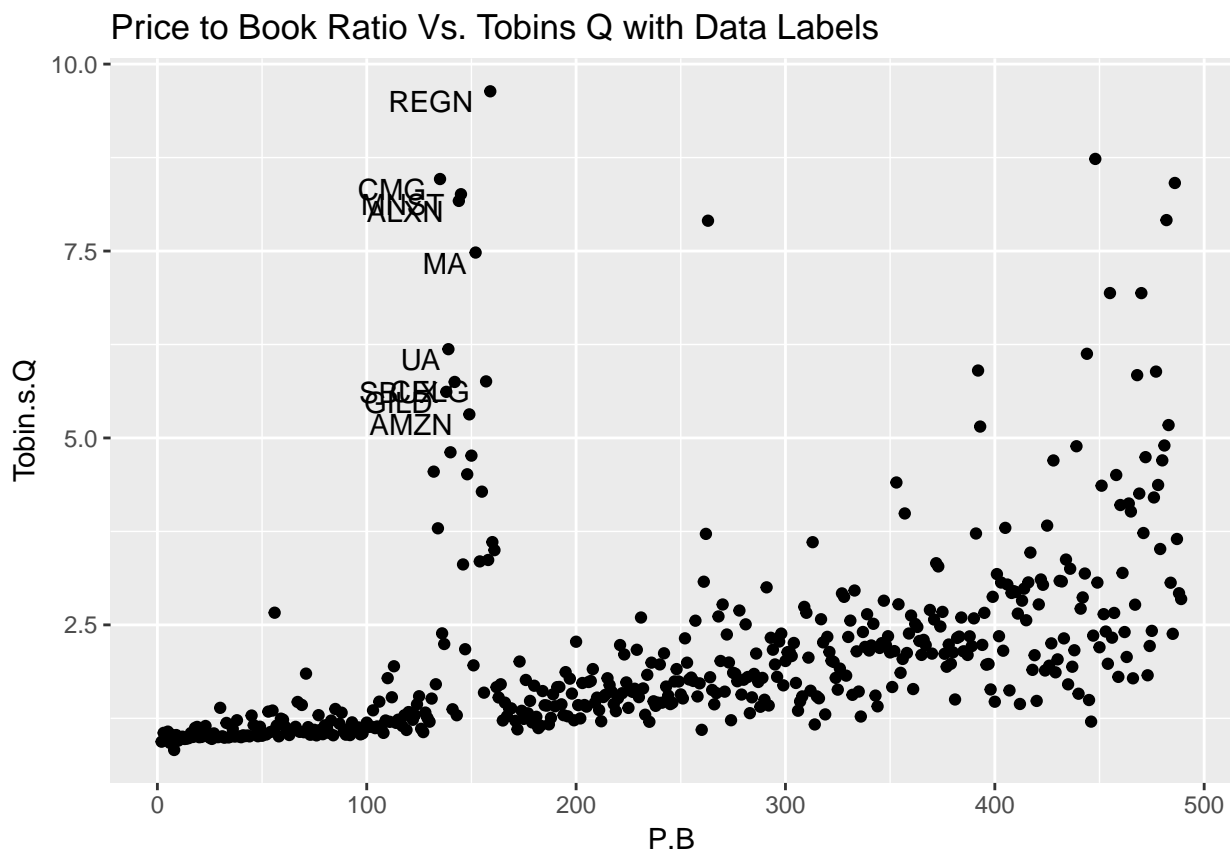
```
##P.B  
ggplot(spx_EDA, aes(P.B, Tobin.s.Q) ) +  
  geom_point() +  
  ggtitle("Price to Book Ratio Vs. Tobins Q")
```



Same as previous page, but showing the tickers for each data point. VRTX is a particular outlier (Vertex Pharmaceuticals Incorporated).

*#with labels*

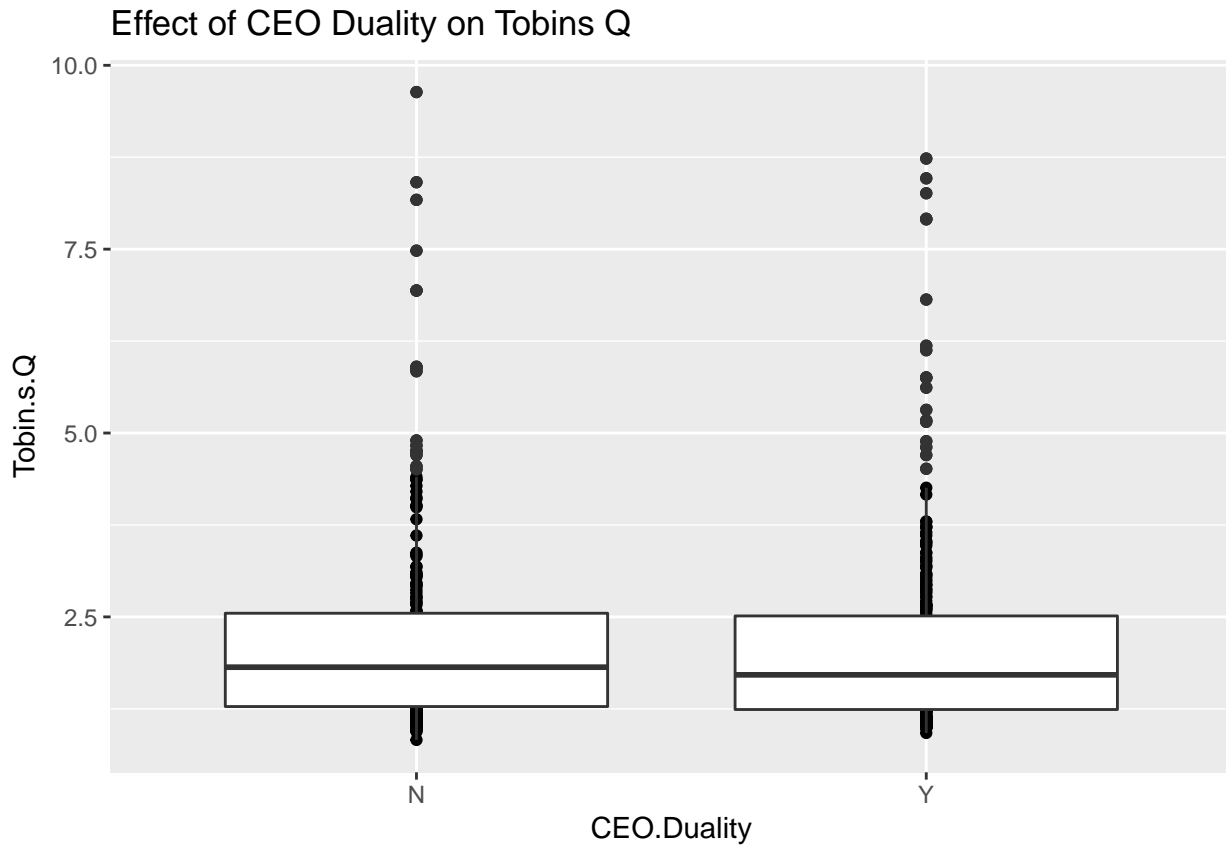
```
ggplot(spx_EDA, aes(P.B, Tobin.s.Q) ) +
  geom_point() +
  geom_text(aes(label=ifelse(P.B>100 & P.B<200 &
                             Tobin.s.Q>5,as.character(TickerID),'')),
            hjust=1.2, vjust=1)+
  ggtitle("Price to Book Ratio Vs. Tobins Q with Data Labels")
```



## CEO Duality and Q Score

Research suggests the CEO Duality can be either a positive or negative influence on the success of a company, looks from the below that perhaps its not that big of an influence at all. No deviation in Q score based on duality.

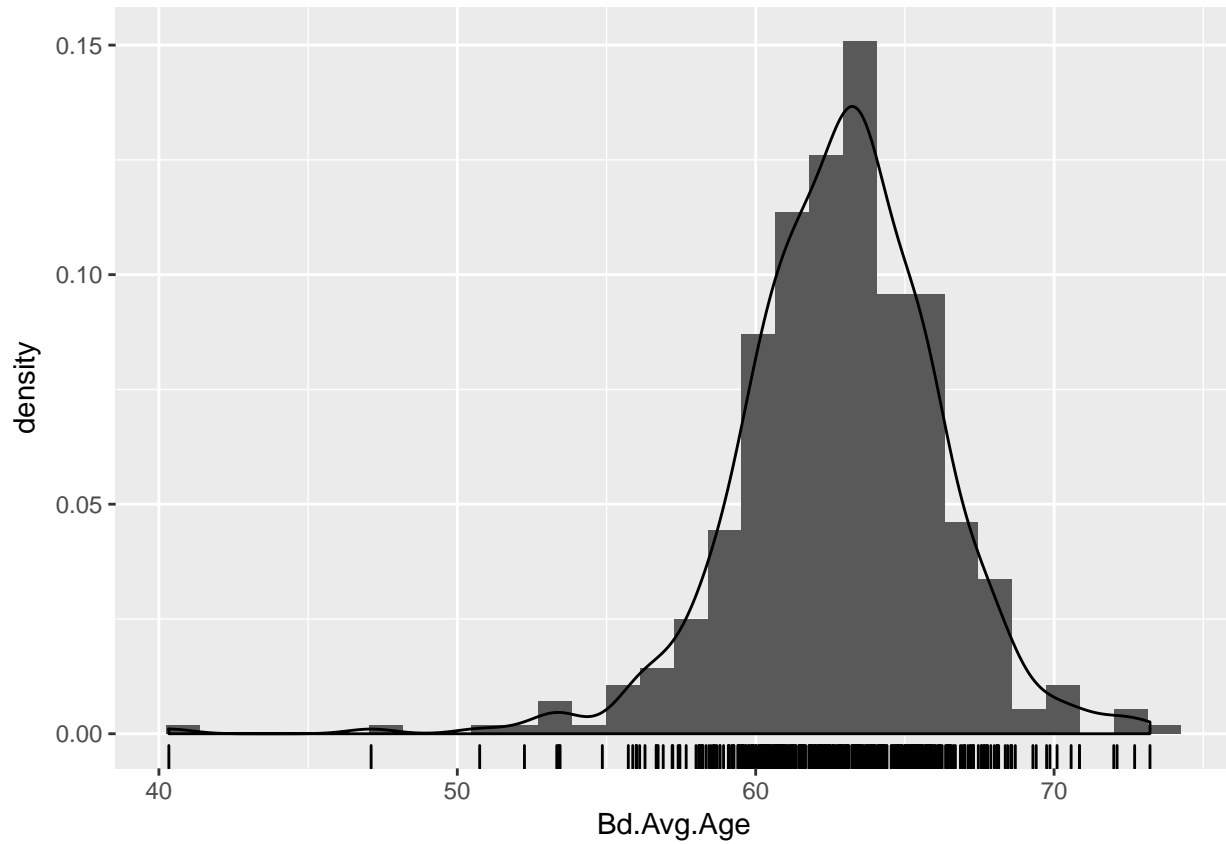
```
#no difference  
ggplot(data=subset(spx_EDA, !is.na(CEO.Duality)),  
  aes(CEO.Duality, Tobin.s.Q )) +  
  geom_point() + geom_boxplot()+  
  ggtitle("Effect of CEO Duality on Tobins Q")
```



## Board Average Age and Q Score

Seems to be pretty normally distributed, skewed left due to some very young boards.

```
#board average age  
ggplot(data=spx_EDA) +  
  geom_histogram( aes(Bd.Avg.Age, ..density..) ) +  
  geom_density( aes(Bd.Avg.Age, ..density..) ) +  
  geom_rug( aes(Bd.Avg.Age) )
```



## Number of Female Executives and Q Score

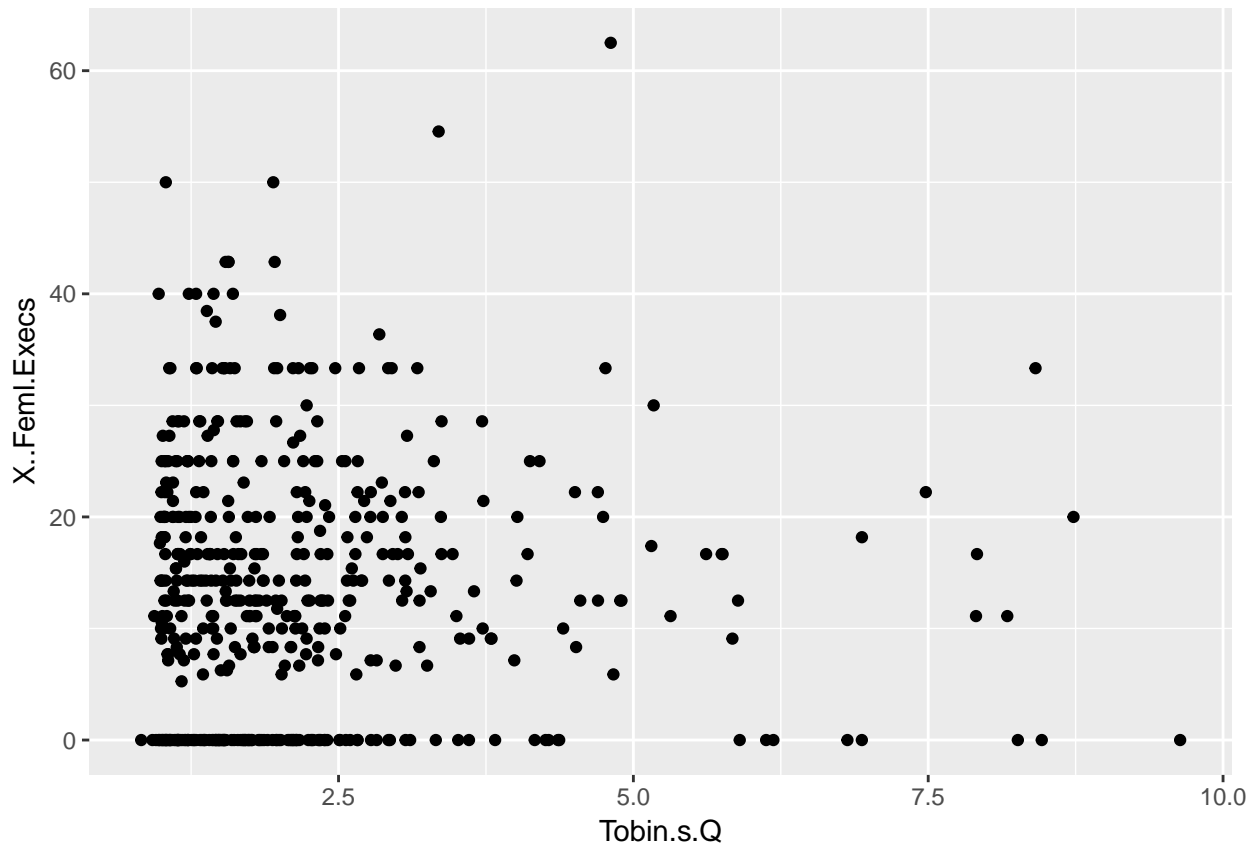
Tobins Q and the number of female executives. Seems to cluster around the bottom left hand corner at low numbers and low performance, but cant see much to the upper right. Doesn't look like a strong relationship here anyway.

According to M&M:

Tobin's Q ratio is positively influenced by the percentage women in the board

May need to look into this further.

```
#female presence on board and success?  
ggplot(spx_EDA, aes(Tobin.s.Q, X..Feml.Execs) ) +  
  geom_point()
```

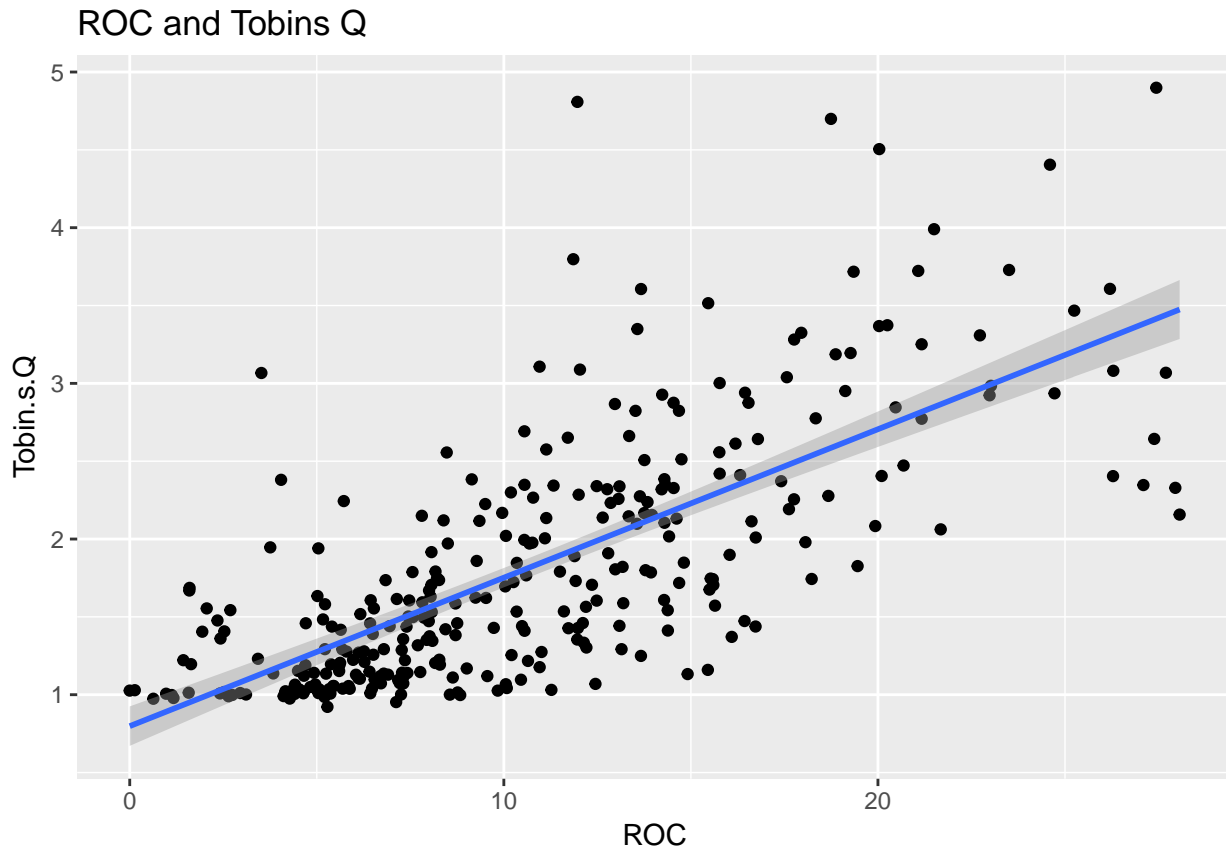


## ROC (Return on Capital) and Q Score

Below shows Q score against ROC (return on capital), subsetting to remove extreme values in both variables. Thus, below is a representation of the main cluster in the data.

Looks to be a positive relationship between the two. Interesting, since ROC (may be) considered a bad strategy in the long term for companies (requires further investigation).

```
ggplot(data=subset(spx_EDA, !is.na(ROC) & ROC<30 & ROC>0 & Tobin.s.Q < 5),  
  aes(ROC, Tobin.s.Q )) +  
  geom_point() +  
  ggtitle("ROC and Tobins Q") +  
  geom_smooth(method = "lm")
```



## STOXX® Europe 600

```
sxxp_EDA = sxxp[vars]
sxxp_EDA[ sxxp_EDA == "#N/A Field Not Applicable" ] <- NA
summary(sxxp_EDA)
```

```
##          Ticker      Tobin.s.Q      AZS
## AAL      LN Equity: 1  Min.    : 0.7437  Min.    :-0.6806
## AALB     NA Equity: 1  1st Qu.: 1.0627  1st Qu.: 2.2148
## ABBN     VX Equity: 1  Median   : 1.4300  Median   : 3.3419
## ABE      SQ Equity: 1  Mean     : 1.9247  Mean     : 4.3981
## ABF      LN Equity: 1  3rd Qu.: 2.0809  3rd Qu.: 5.0353
## ABG/P    SQ Equity: 1  Max.     :72.2901  Max.     :46.5149
## (Other)  :594  NA's    :4      NA's    :127
##          P.E      EPS      P.B      P.EBITDA
## 1.289647253: 1  0.71    : 5  0.290244253: 1  0.874325837: 1
## 1.39824875 : 1  2.6     : 4  0.3408942 : 1  1.080417894: 1
## 10.07537419: 1  0.15    : 3  0.396541621: 1  1.190429818: 1
## 10.09797277: 1  0.27    : 3  0.429090936: 1  1.539923955: 1
## 10.13564468: 1  0.3     : 3  0.472640007: 1  1.580920306: 1
## (Other)    :534  (Other):573  (Other)    :578  (Other)    :466
## NA's       : 61  NA's     : 9  NA's       : 17  NA's       :129
##      Board.Size  CEO.Duality  X..Feml.Execs  X..Feml.Execs.1
## Min.    : 3.00  N      :522  Min.    : 0.000  Min.    :0.0000
## 1st Qu.: 9.00  Y      : 72  1st Qu.: 0.000  1st Qu.:0.0000
## Median :10.50  NA's: 6  Median   : 0.000  Median :0.0000
## Mean    :11.17              Mean     : 8.299  Mean    :0.7036
## 3rd Qu.:13.00              3rd Qu.: 14.290  3rd Qu.:1.0000
## Max.    :23.00              Max.     :100.000  Max.    :6.0000
## NA's    :6              NA's     :15  NA's     :13
##      Bd.Avg.Age  Board.Mtg.Att..  Asset      Interest
## Min.    :10.00  Min.    : 50.00  Min.    :0.0000  Min.    :-16040.00
## 1st Qu.:56.00  1st Qu.: 92.00  1st Qu.:0.2872  1st Qu.: 66.89
## Median :58.22  Median : 96.00  Median :0.6308  Median : 87.41
## Mean    :58.22  Mean    : 94.61  Mean    :0.7179  Mean    : 22.14
## 3rd Qu.:60.75  3rd Qu.: 98.70  3rd Qu.:0.9755  3rd Qu.: 99.49
## Max.    :72.10  Max.    :100.00  Max.    :4.1537  Max.    : 1684.38
## NA's    :207  NA's    :124  NA's    :15  NA's    :120
##      Tax      ROE      ROC      Indep.Directors
## Min.    :-2775.41  Min.    :-33.652  Min.    : -7.242  Min.    : 0.0
## 1st Qu.: 58.50  1st Qu.: 6.956  1st Qu.: 4.676  1st Qu.: 5.0
## Median : 73.10  Median :12.776  Median : 9.313  Median : 6.0
## Mean    : 78.60  Mean    :16.290  Mean    :15.401  Mean    : 6.7
## 3rd Qu.: 81.34  3rd Qu.:19.755  3rd Qu.:13.863  3rd Qu.: 8.0
## Max.    :8242.86  Max.    :370.768  Max.    :1647.036  Max.    :21.0
## NA's    :6      NA's    :18  NA's    :219  NA's    :100
##      X..Indep.Dir.on.Aud.Cmte  X..Indep.Dir.on.Aud.Cmte.1  Sz.Aud.Cmte
## Min.    :1.000  Min.    : 16.67  Min.    :0.000
## 1st Qu.:3.000  1st Qu.: 75.00  1st Qu.:3.000
## Median :3.000  Median :100.00  Median :4.000
## Mean    :3.468  Mean    : 87.52  Mean    :3.947
## 3rd Qu.:4.000  3rd Qu.:100.00  3rd Qu.:5.000
## Max.    :9.000  Max.    :100.00  Max.    :9.000
```

## NA's :124

NA's :125

NA's :32

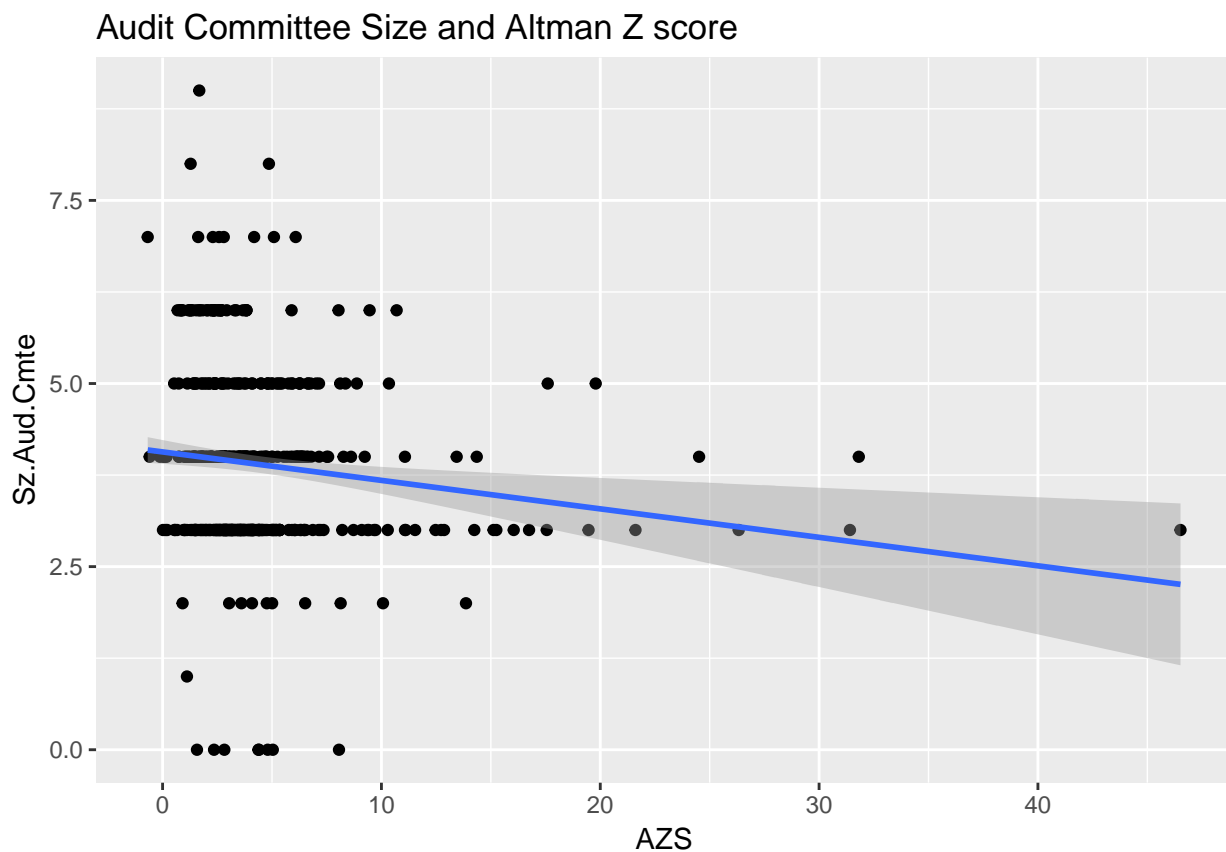
## Audit committee and Altman Z Score

According to M&M:

... while in Western Europe that it is better to employ a larger audit committee in order to lower the bankruptcy risk

Again, result below doesnt look all that convincing but further analysis required.

```
ggplot(data=sxxp_EDA,  
  aes(AZS, Sz.Aud.Cmte )) +  
  geom_point() +  
  ggtitle("Audit Committee Size and Altman Z score") +  
  geom_smooth(method = "lm")
```





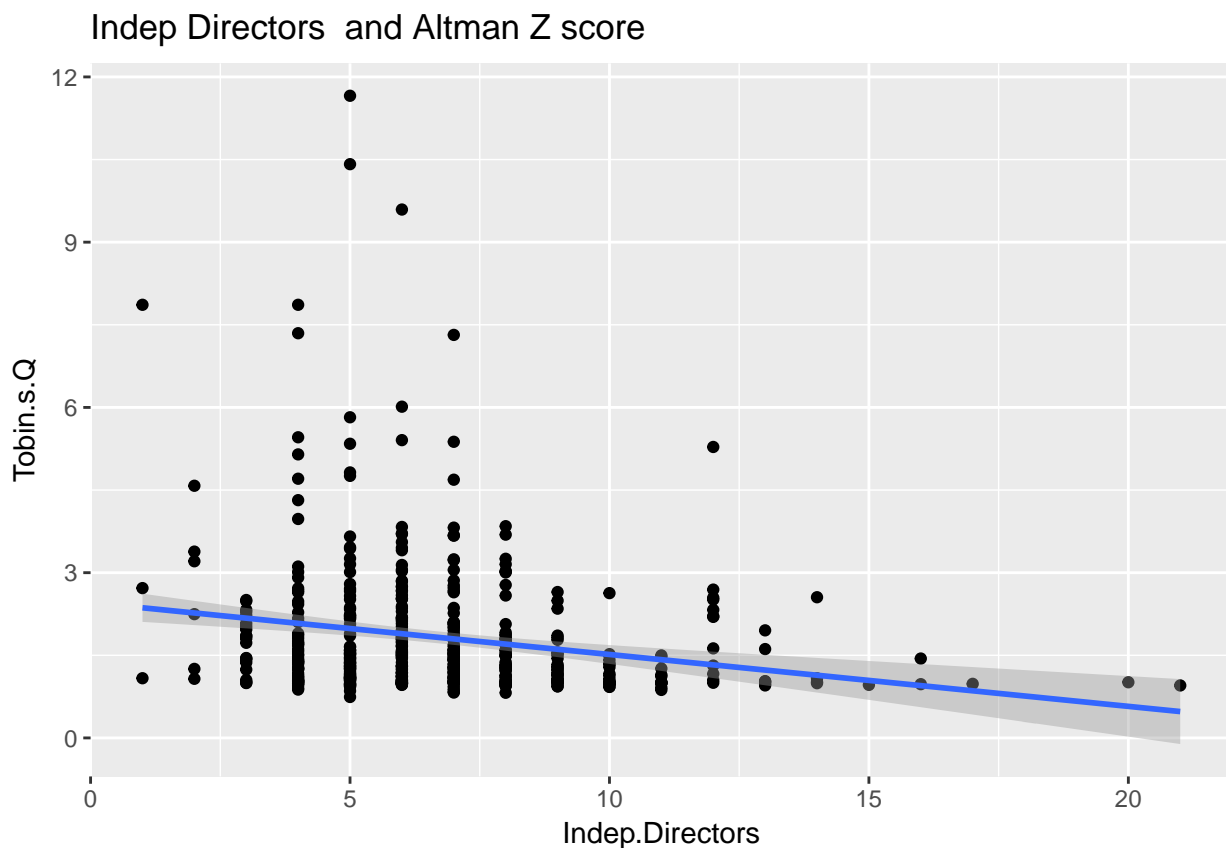
## Indep.Director / Former CEO and Tobin Q

According to M&M;

For the Western European companies, the presence of an independent lead director or a former CEO in the board could be a sign of weaker performances, being negatively correlated with Tobin's Q

Results below (take out one outlier, whose Q score is huge) seem to support this. Again, r score likely to be pretty low based on how distributed the data points are.

```
ggplot(data=subset(sxxp_EDA, Tobin.s.Q<30 ),  
  aes(Indep.Directors, Tobin.s.Q )) +  
  geom_point() +  
  ggtitle("Indep Directors and Altman Z score") +  
  geom_smooth(method = "lm")
```



## STOXX Eastern Europe 300