Group 8 Maths and Stats Project

1. Introduction

Business environment in China is unique. Multinational companies (MNC) often find it difficult to compete with local companies.

China is a country where the government has a relatively strong interception and interference on its economy. Thus, state-owned companies in China may have a stronger relationship with the government and be affected more by the government compared with other kinds of companies.

A stronger connection with the government may benefit the company with easy access to important information, such as big news or change of policies. It could benefit the company with prior gain of resources as well, such as controlled materials. Moreover, one state-owned company could operate more easily with relatively low risk of bankruptcy due to the help offered by the government.

Based on the above information and analysis, it may be interpreted that the relationship between the company-related factors and the government-related factors of one state-owned company should be stronger than that of other kinds of companies.

This report analyse replication dataset provided by Zhu (2016) that consisted of information about different companies in 30 provinces in China. The relationship between variables that affect company revenue will be investigated using correlation test and other different statistical testing. Then, we will examine the significant variables using linear regression to calculate their impact to the log of company revenue.

2. Theory

The theory we have based our analysis on is a paper published in 2012 about political connections and how this affects a firm's performance.

Companies with a strong relationship to the government benefit from easier access to government-related resources and bureaucratic processes, however, research conducted on this relationship with companies' financial performance is varied. In China, 76 out of 98 companies in 2014 Fortune 500 list, are state-owned (Cendrowski, 2015) and it shows the importance of government of China in the development of these companies.

One of the theories on this topic is from Pfeiffer and Salancik (1978), they developed a theory stating that political connections will increase a firms value. Another theory, published by Jensen and Meckling (1976), argue that a firm which have a CEO with strong government connections might utilize this benefit in his own interests rather than the business' interest. We found these theories quite interesting and wanted to see if we could find any support for them in our data.

3. Methods

The analysis is conducted using several statistical tools in R, i.e. statistical test and linear modelling. Some libraries such as `ggplot2`, `scales` and `stargazer` are loaded to the console to visualise the data.

```
library(ggplot2)
library(scales)
library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2015). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2. http://CRAN.R-project.org/package=stargazer
library(plyr)
```

3.1 Description

```
#load("./ForeighPresenceChina -- finally chosen/ForeignPresence&ETCs.RData")
load("C:/Users/Kaixuan/Documents/MSF_Team_8/ForeighPresenceChina -- finally chosen/ForeignPresence&ETCs
mnc <- as.data.frame(x)
ls.str(mnc)
summary(mnc)
```

The dataset contains 40 variables and 11360 observations. The data covered 11360 different companies observed across 30 provinces in China out of a total of 276,474 companies recorded in 2004 by National Bureau of Statistics of China (2014). The variables are divided into three groups, i.e. firm-level variables, industry-level variables and province-level variables. Firm-level variables consist of the properties of each firm, which include the type of ownership, revenue, number of employees and several variables that highlight the relationship between a firm and the government of China in 2004. Industry-level variables illustrate the scale of the industry and the market share of different groups of companies, including foreign firms and the industry leaders. They also try to highlight the amount of government influence by measuring the linear combination of the amount of help received by government officials, as well as the median of tax rate incurred by each firm in the industry. Lastly, province-level variables indicate the macroeconomic information regarding a province, such as its GDP per capita and the level of MNC activities in the region (e.g. foreign direct investments and trades). Below are the details of each variable analysed in this report.

3.1.1 Firm-Level Variables

Variable	Description
`idstd`	Company ID
`lectcs`	Log of firm expenditures on entertainment as a percentage of annual revenue.
`soe`	Whether a firm is state-owned or not. (1 = Yes, 0 = No)
`collective`	Whether a firm is collectively-owned. (1 = Yes, 0 = No)
`private`	Whether a firm is privately-owned. (1 = Yes, 0 = No)
`foreign_HMT`	Whether a firm is owned by foreign owner from Hong Kong, Macao or Taiwan.
`foreign_NonHMT`	Whether a firm is owned by foreign owner other not from Hong Kong, Macao or Taiwan.
`mixedown`	Other firms not identified in `soe`, `collective`, `private`, `foreign_HMT` and `foreign_NonHMT` are coded 1.
`revenue`	Total revenue in 2004 in log form
`lemp`	Total number of employee in 2004 in log form
`sales_otherprov`	Whether a firm sells to other provinces. (1 = Yes, 0 = No)
`govtsales`	Proportion of revenue from transactions with the government
`soesales`	Proportion of revenue from transactions with state-owned enterprises
`relationship`	Years of relationships between the firm and its main clients
`licenses`	Number of licenses required to operate
`lceopay`	CEO annual salary in log form
`interaction`	Number of days of interactions with the government

Variable	Description
gm_govt	Whether the general manager is appointed by the government. (1 = Yes, 0 = No)

3.1.2 Industry-Level Variables

Variable	Description
foreign_output	The share of output by foreign-owned firms
HMT_output	Percentage output by firms owned by shareholders from Hong Kong, Macao or Taiwan
NonHMT_output	Percentage output by firms owned by other foreign nationals
w_con_pdc4	Market share of four biggest firms in the industry
w_con_pdc8	Market share of eight biggest firms in the industry
marketsize	The market size of the industry
growthrate	Output growth 2001-2003
last_intensity	the ratio of fixed asset per number of employees in log form
lscale	The mean of the size of firms that has 50% of industry market share in log form
mgovthelp	A linear combination of government helpfullness based on the contribution of government officials
mtaxrate	Median of firms tax paid divided by total revenue
prov_ind	Industry ID by Province

3.1.3 Province-Level Variables

Variable	Description
MNC	Principal factor analysis of FDI inflow and MNC trades in a province.
lwdist	Geographic distance
lgdpper2003	Log of GDP per capita of the province in 2003
gdp2003	GDP in 2003
lpop2003	Log of the province's population in 2003
province	The name of the province
provinceID	Province ID

3.2. Strength and Limitations

This dataset represent the ties between firms and the government of China. Given the unique feature of Chinese industry that most of big firms in China are state-owned, as of 98 companies in 2014 Fortune 500, 76 of them are state-owned (Cendrowski, 2015), this dataset would highlight how strong the relationship with the goverment affect the CEO salary of a company or the company revenue.

However, the dataset possess some limitations. Although industry ID numbers are provided, the sector was not clearly specified in the dataset or in previous publications. Moreover, the same industry in different provinces were coded differently, thus, the impact of a specific industry could not be measured.

4. Analysis

4.1 Statistics Description of the Dataset

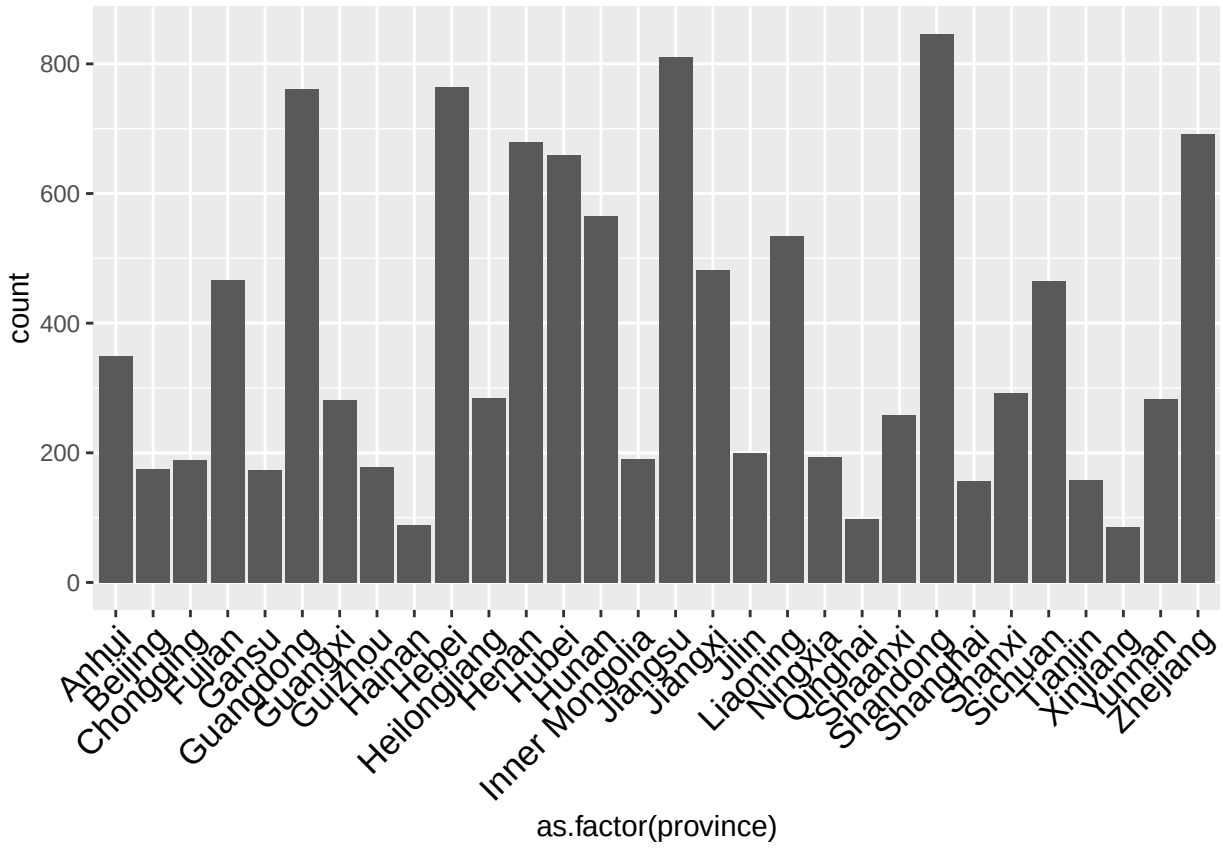
4.1.1 Exploring Province-level Variables

Provinces variable were separated from the datasets. Before the data was plotted, `province` variable needed to be converted to factor. The summary showed the sample of each province ranges between 85 (Xinjiang) and 846 (Shangdong).

```
library(ggplot2)
mnc$province <- as.factor(mnc$province)
summary(as.factor(mnc$province))
```

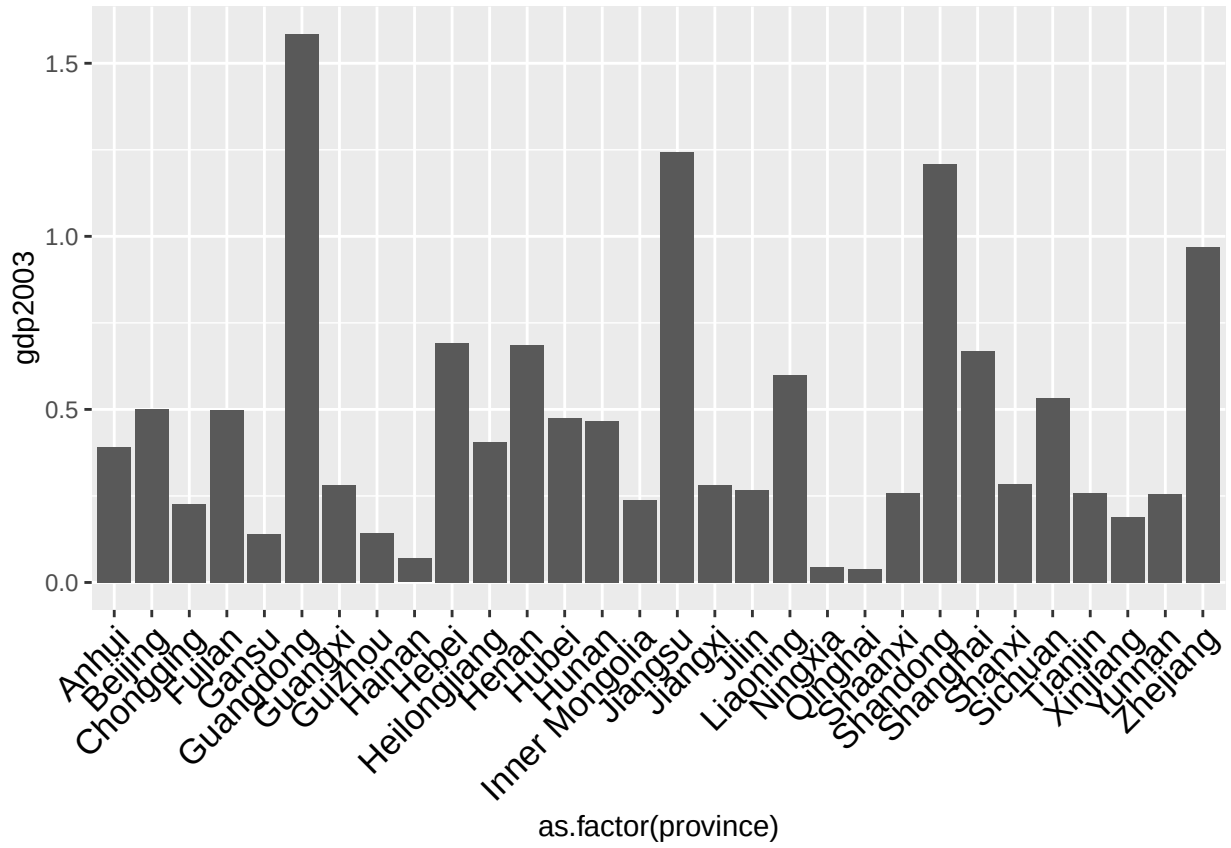
```
##          Anhui        Beijing      Chongqing         Fujian          Gansu
##            350            175            189            467            174
##      Guangdong        Guangxi        Guizhou         Hainan          Hebei
##            761            281            178             88            765
##   Heilongjiang          Henan          Hubei          Hunan Inner Mongolia
##            284            680            660            566            190
##        Jiangsu        Jiangxi          Jilin       Liaoning        Ningxia
##            810            482            199            534            194
##        Qinghai        Shaanxi       Shandong       Shanghai         Shanxi
##             98            258            846            156            292
##        Sichuan        Tianjin       Xinjiang         Yunnan       Zhejiang
##            465            158             85            283            692
```

```
# Plot the numbers of samples of different provinces
ggplot(mnc) + geom_bar(aes(x = as.factor(province))) + theme(axis.text.x = element_text(angle = 45,
    colour = "black", size = 13, hjust = 1))
```



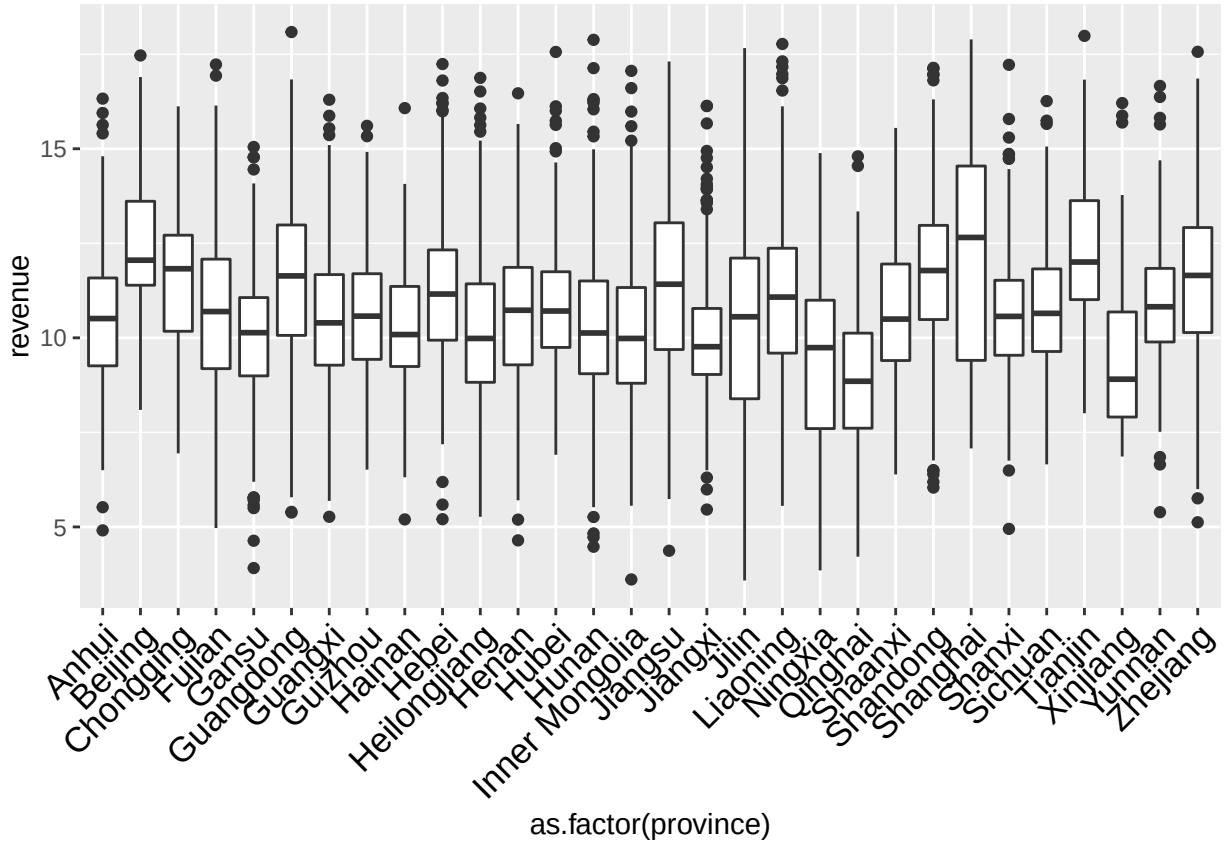

```
# Get the data for province
mnc.prov <- as.data.frame(mnc[!duplicated(x[, "province"]), ])

# Plot the GDP of different provinces
ggplot(mnc.prov) + geom_bar(aes(x = as.factor(province), y = gdp2003), stat = "identity",
    position = "dodge") + theme(axis.text.x = element_text(angle = 45, colour = "black",
    size = 13, hjust = 1))
```



The GDP in 2003 were plotted in a barchart. The first highest three GDP provinces are Guangdong, Jiangsu and Shandong. The lowest three GDP provinces are Qinghai, Ningxia and Hainan. Then, we visualised the relationship between provinces' number of population (log) and their log of GDP per capita.

```
ggplot(data=mnc) + geom_boxplot(aes(as.factor(province), revenue)) + theme(axis.text.x = element_text(a
```



It can be concluded that the Shanghai firms' revenues have the largest interval and Jiangxi has the smallest revenues interval. The median revenues in Shanghai is highest and that in Qinghai is the lowest.

From the above analysis, we may derive the observation that companies located in provinces with relatively higher GDP may gain more revenue than others.

4.1.2 Exploring GDP-related Variables

```
cor.test(mnc.prov[,"lpop2003"], mnc.prov[,"lgdpper2003"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc.prov[, "lpop2003"] and mnc.prov[, "lgdpper2003"]
## t = -0.84312, df = 28, p-value = 0.4063
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.4898505  0.2151136
## sample estimates:
##       cor
## -0.1573501
```

There is no strong relationship between the population and capital. From the correlation test result, we found that the p-value is 0.4043, which is much larger than 0.05, so we cannot reject the Null Hypothesis at a 95% confindence level. Therefore, there is no strong correlation between these two variables. The next part explore the link between provinces' size and their GDP per Capita.

```r
cor.test(mnc.prov[,"lwdist"], mnc.prov[,"lgdpper2003"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc.prov[, "lwdist"] and mnc.prov[, "lgdpper2003"]
## t = 0.99205, df = 28, p-value = 0.3297
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1885141  0.5106395
## sample estimates:
##     cor
## 0.18427
```

The weighted geographic distance and GDP per Capita do not show any significant relationship. The p value is 0.3297, which is much larger than 0.05, thus, the Null Hypothesis cannot be rejected. Although there is no strong correlation, these two variables are positive correlated, which means large province enjoyed slightly more GDP per Capita.

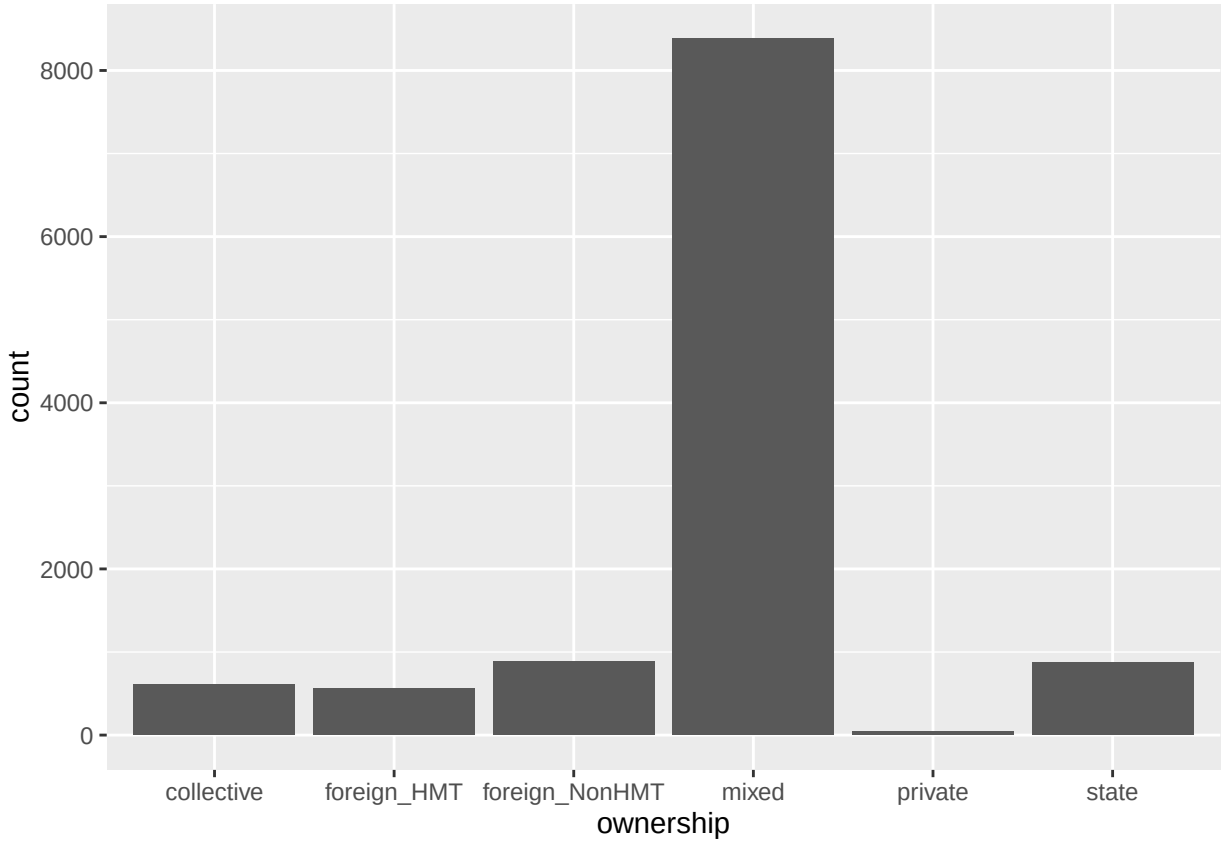
4.1.3 T-test of Company Revenue for Beijing and Shanghai

```r
t.test(mnc[mnc[,"province"] == "Beijing","revenue"],mnc[mnc[,"province"] == "Shanghai","revenue"], alte
```

```
##
##  Welch Two Sample t-test
##
## data:  mnc[mnc[, "province"] == "Beijing", "revenue"] and mnc[mnc[, "province"] == "Shanghai", "reve
## t = 1.4738, df = 237.38, p-value = 0.1419
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1234829  0.8569424
## sample estimates:
## mean of x mean of y
##  12.51317  12.14644
```

We conducted T-test of revenue for Beijing and Shanghai, two biggest cities in China. The p value is 0.1419, which is larger than 0.05, so we can not reject the Null Hypothesis on a 95 % confidence level. We can not find a significant difference in the two means.

4.1.4. Description of the Firms Ownership



Average Revenue of Each Ownership

```
mean(mnc[mnc[,"ownership"]=="state", "revenue"])
```

```
## [1] 11.1182
```

```
mean(mnc[mnc[,"ownership"]=="collective", "revenue"])
```

```
## [1] 9.882829
```

```
mean(mnc[mnc[,"ownership"]=="private", "revenue"])
```

```
## [1] 9.863199
```

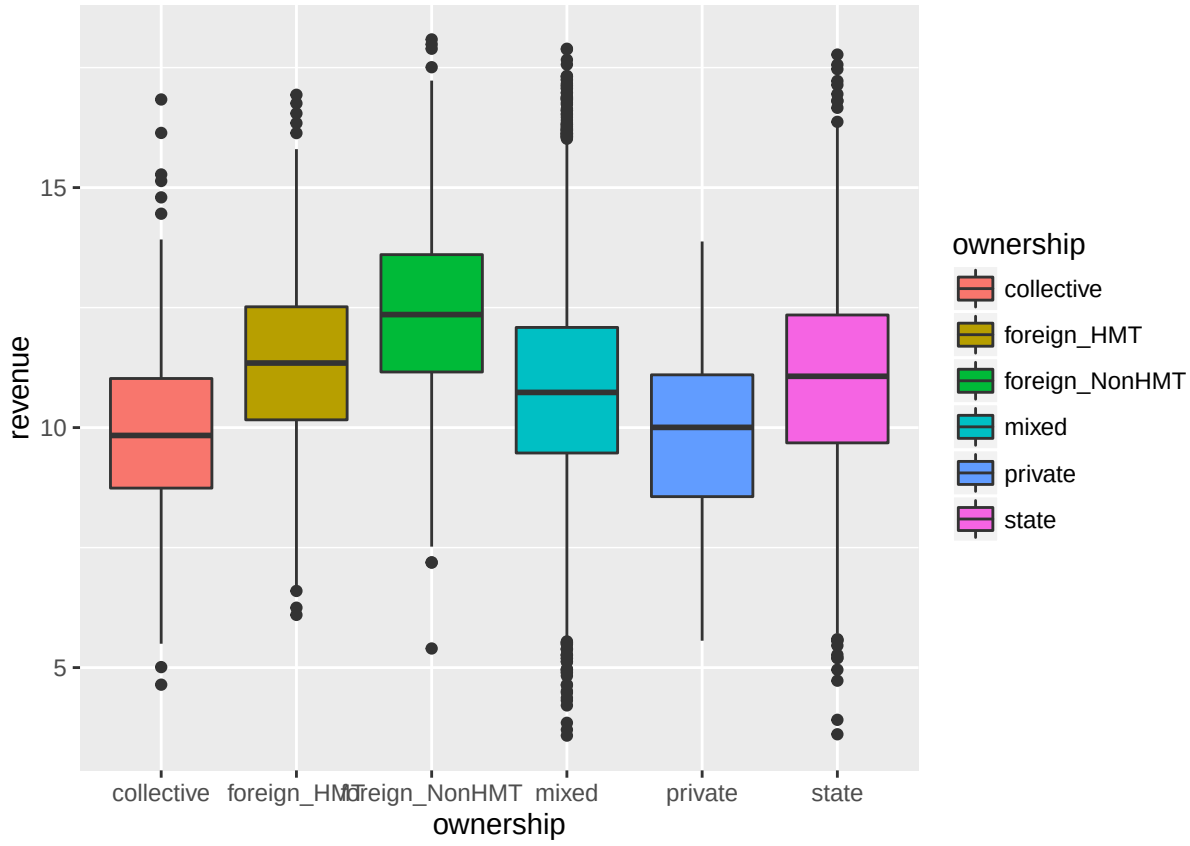
```
mean(mnc[mnc[,"ownership"]=="foreign_HMT", "revenue"])
```

```
## [1] 11.38784
```

```
mean(mnc[mnc[,"ownership"]=="foreign_NonHMT", "revenue"])
```

```
## [1] 12.37844
```

```
ggplot(data=mnc) + geom_boxplot(aes(ownership, revenue, fill = ownership))
```

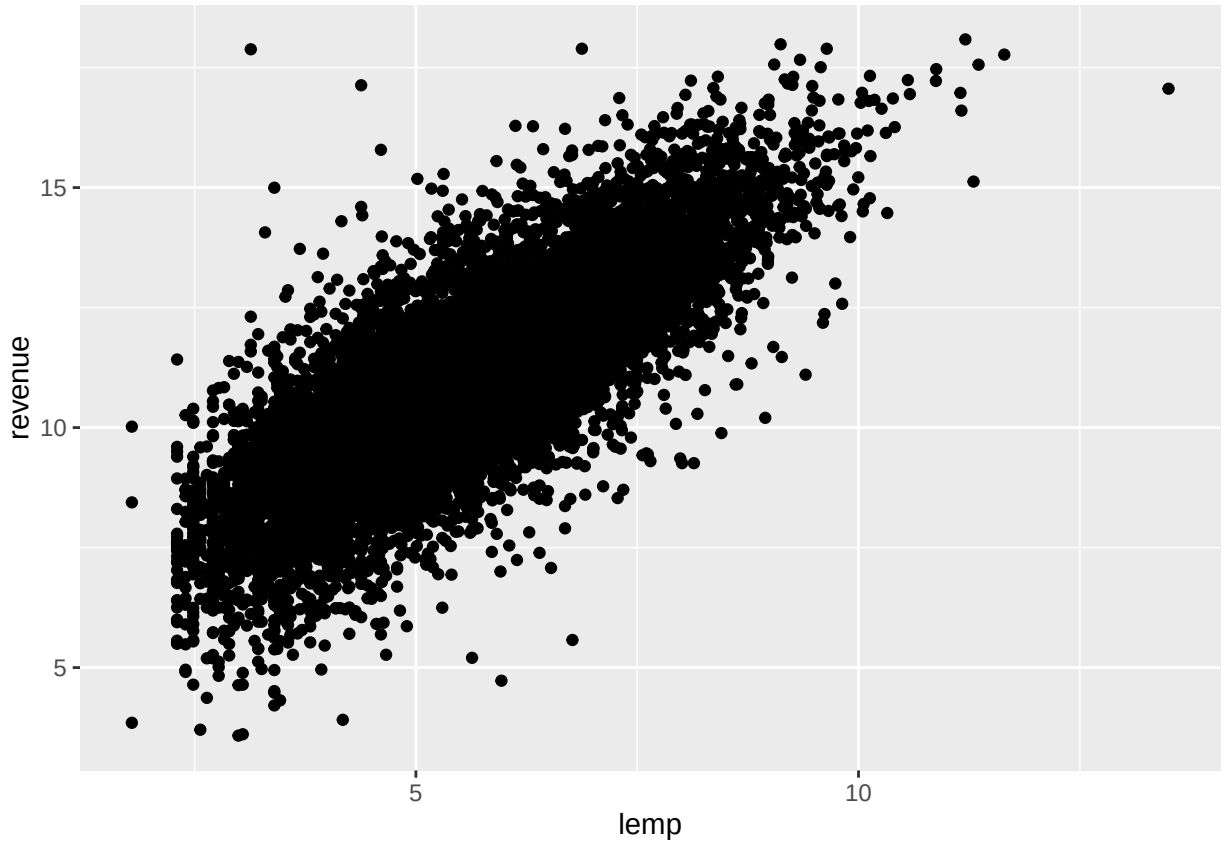


As we can see of the barplot there is a lot of companies in the bar with "mixed" ownership compared to the other categories. Therefore it might be difficult to do any valuable analysis on this data. If we look at the boxplot we can see that it looks like the distribution of revenue might be different between the different kind of ownership.

4.2 Correlation Analysis

4.2.1 Correlation Test among Different Variables with `lemp` (log of total employees in 2004)

```
ggplot(mnc) + geom_point(aes(lemp,revenue))
```




```
cor.test(mnc[,"lemp"],mnc[,"revenue"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "lemp"] and mnc[, "revenue"]
## t = 138.69, df = 11358, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7860050 0.7996623
## sample estimates:
##       cor
## 0.7929332
```

```
cor.test(mnc[,"lemp"],mnc[,"lceopay"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "lemp"] and mnc[, "lceopay"]
## t = 39.122, df = 11023, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.3326759 0.3654589
## sample estimates:
##       cor
## 0.3491743
```

```r
cor.test(mnc[,"lemp"],mnc[,"letcs"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "lemp"] and mnc[, "letcs"]
## t = -10.975, df = 11358, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.12060080 -0.08420784
## sample estimates:
##        cor
## -0.1024386
```

From the scatter diagram, we can find that the with the increasement of the firm's total employees, the total income and CEO payment will increase. But the log of firm expenditure on entertainment and travel cost as % of total revenue will seem to decrease when firm's total employees increases. From the correlation test, we get that the correlation between lemp and revenue, lceopay and letcs are 0.79, 0.34, and -0.10.

4.2.2 Correlation Test among Different Variables with `Revenue` (log of total business income)

4.2.2.1 Correlation Test with Government Relative Factors

According to the dataset provided, we would like to examine the correlation of all factors relative to the government with the revenue of the company. In this case, we analysis the factors:

Variable	Reason of Analyzing it
`mgovthelp`	based on the condition in China, with more help from the government, one company may perform better and gain more revenue
`govtsales`	sales to Chinese government meaning stable sales with only low risk
`interaction`	with more time working on the government assignments and communications, it would be easier for the company to know news and policies from the goverment and act accordingly and timely

```r
cor.test(mnc[, "revenue"], mnc[, "mgovthelp"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "revenue"] and mnc[, "mgovthelp"]
## t = 2.0832, df = 11357, p-value = 0.03726
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.001153993 0.037920347
## sample estimates:
##        cor
## 0.01954378
```

```r
cor.test(mnc[, "revenue"], mnc[, "govtsales"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "revenue"] and mnc[, "govtsales"]
```

```
## t = -0.34543, df = 11358, p-value = 0.7298
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.02162934  0.01514905
## sample estimates:
##          cor
## -0.003241241
```

```
cor.test(mnc[, "revenue"], mnc[, "interaction"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "revenue"] and mnc[, "interaction"]
## t = 5.0441, df = 11358, p-value = 4.627e-07
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.02891254 0.06560914
## sample estimates:
##        cor
## 0.04727679
```

Based on the correlation tests, it could be observed that the total business income of one company has a quite weak connection with the factors relative to the government, containing even negative correlation with some of the factors. In this case, the theory mentioned previously could not be proved since we hardly observe any strong positive or negative correlation among governmental factors with the performance of a company. Furthermore, the factors tested above may not be used as the independent variables for the linear regression model with the revenue as the dependent variables.

4.2.2.2 Correlation Test with Non-government Relative Factors

Based on the above tests, government relative factors have very weak impacts on the revenue of the company. In this case, many other non-government relative factors should be examined for better interpretation of the dataset and smothier linear regression test. The factors are to be analyzed are:

Variable	Reason of Analyzing it
`lgdpper2003`	GDP of one country may have big impact on the performance of companies in that country
`marketsize`	the revenue of one company may fluctuate according to the market size
`growthrate`	the change of the output growth rate may affect the revenue as there exists some direct relationship between them
`relationship`	a good and long existing relationship with the clients and suppliers could help with the increase of the revenue
`licenses`	more licenses one company holds, it would be possible for that company to generate more revenue
`lceopay`	the performance of the company's CEO largely affect the revenue that company earns
`letcs`	the expenditure on entertainment and travel cost directly reflect on the revenue of the company

```
cor.test(mnc[, "revenue"], mnc[, "lgdpper2003"])
```

```
##
##  Pearson's product-moment correlation
```

```
##
## data:  mnc[, "revenue"] and mnc[, "lgdpper2003"]
## t = 26.052, df = 11358, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2200298 0.2547354
## sample estimates:
##       cor
## 0.2374584
```

```
cor.test(mnc[, "revenue"], mnc[, "marketsize"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "revenue"] and mnc[, "marketsize"]
## t = 34.244, df = 11358, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2891464 0.3224844
## sample estimates:
##       cor
## 0.3059092
```

```
cor.test(mnc[, "revenue"], mnc[, "growthrate"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "revenue"] and mnc[, "growthrate"]
## t = -5.7057, df = 11358, p-value = 1.187e-08
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.07177998 -0.03510629
## sample estimates:
##         cor
## -0.05346116
```

```
cor.test(mnc[, "revenue"], mnc[, "relationship"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "revenue"] and mnc[, "relationship"]
## t = 30.468, df = 11358, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2577914 0.2917921
## sample estimates:
##       cor
## 0.2748777
```

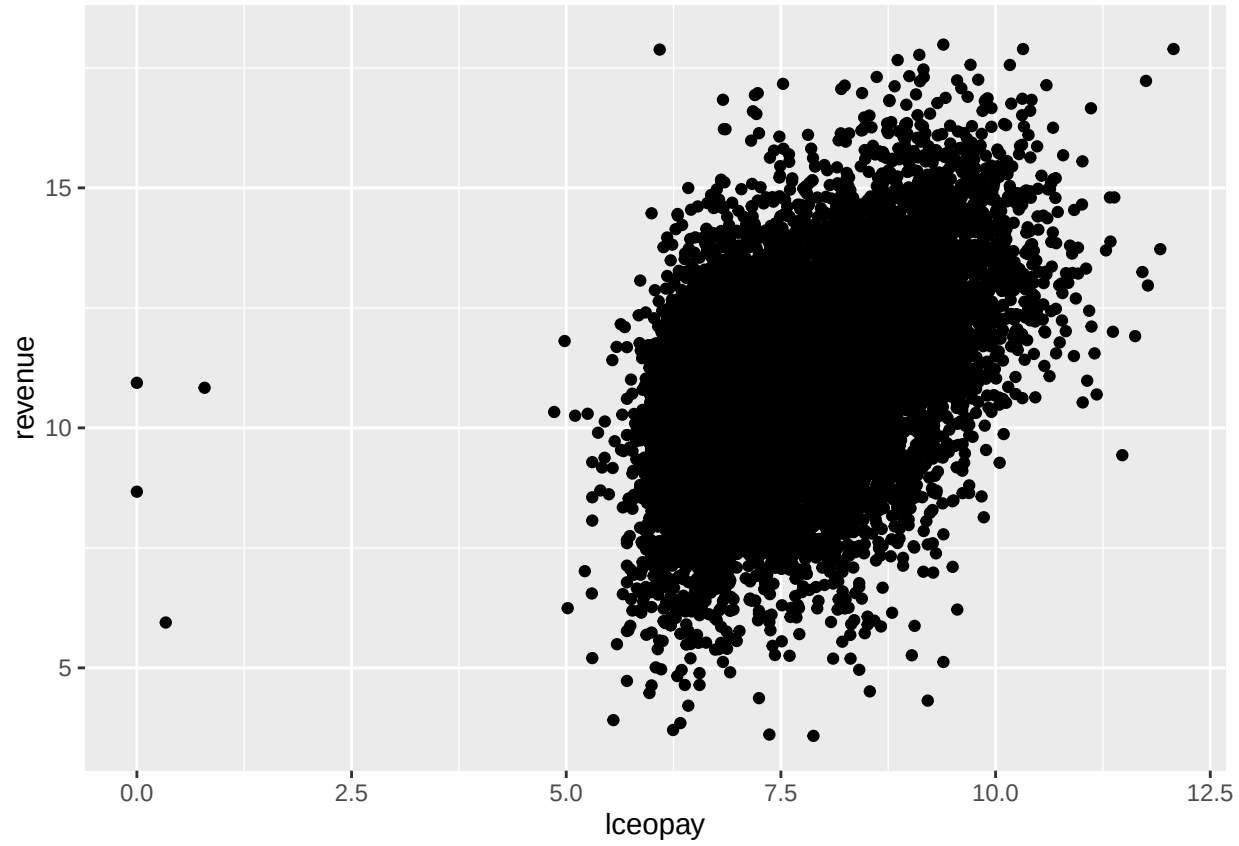
```
cor.test(mnc[, "revenue"], mnc[, "licenses"])
```

```
##
##  Pearson's product-moment correlation
##
```

```
## data:  mnc[, "revenue"] and mnc[, "licenses"]
## t = 13.721, df = 11358, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1095590 0.1457383
## sample estimates:
##       cor
## 0.1276911
```

```
ggplot(data = mnc, aes(x = lceopay, y = revenue)) + geom_point()
```

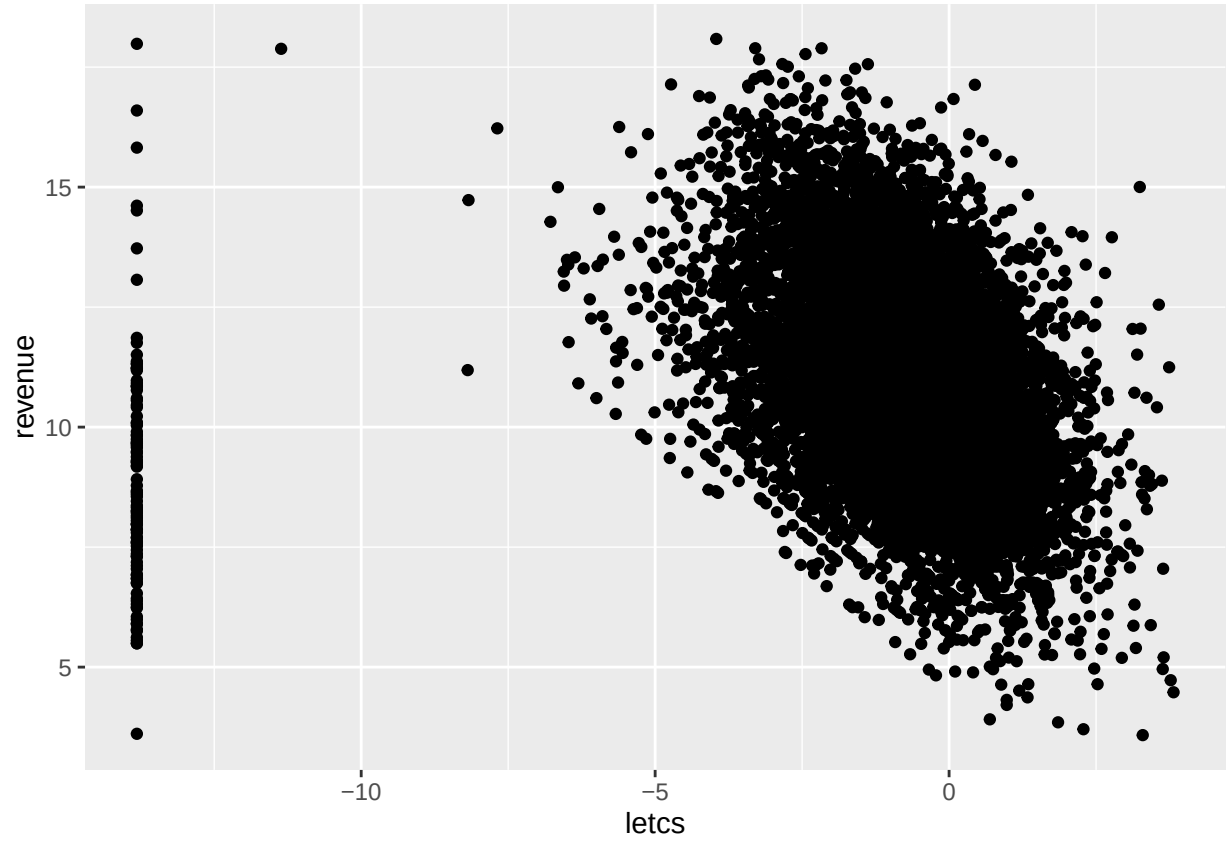
```
## Warning: Removed 335 rows containing missing values (geom_point).
```




```
cor.test(mnc[, "revenue"], mnc[, "lceopay"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "revenue"] and mnc[, "lceopay"]
## t = 59.457, df = 11023, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4785118 0.5067820
## sample estimates:
##       cor
## 0.4927769
```

```
ggplot(data = mnc, aes(x = letcs, y = revenue)) + geom_point()
```




```
cor.test(mnc[, "revenue"], mnc[, "letcs"])
```

```
##
##  Pearson's product-moment correlation
##
## data:  mnc[, "revenue"] and mnc[, "letcs"]
## t = -26.011, df = 11358, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2543821 -0.2196703
## sample estimates:
##        cor
## -0.2371019
```

Based on the correlation tests, it could be observed that the total business income of one company has a positive connection with all the factors mentioned above except the output growth rate (with -0.053 correlation with revenue). The correlation with revenue in descending order is: CEO annual income (0.493), market size (0.306), relationship with clients and suppliers (0.275), GDP per capita in 2003 (0.237) and licenses (0.128). blablabla

4.3 Linear Regression

4.3.2. Single Variable Linear Regression Test with `Revenue` as the Dependent Variable

Based on our correlation analysis we want to look deeper into the variables which had the highest correlation with revenue.

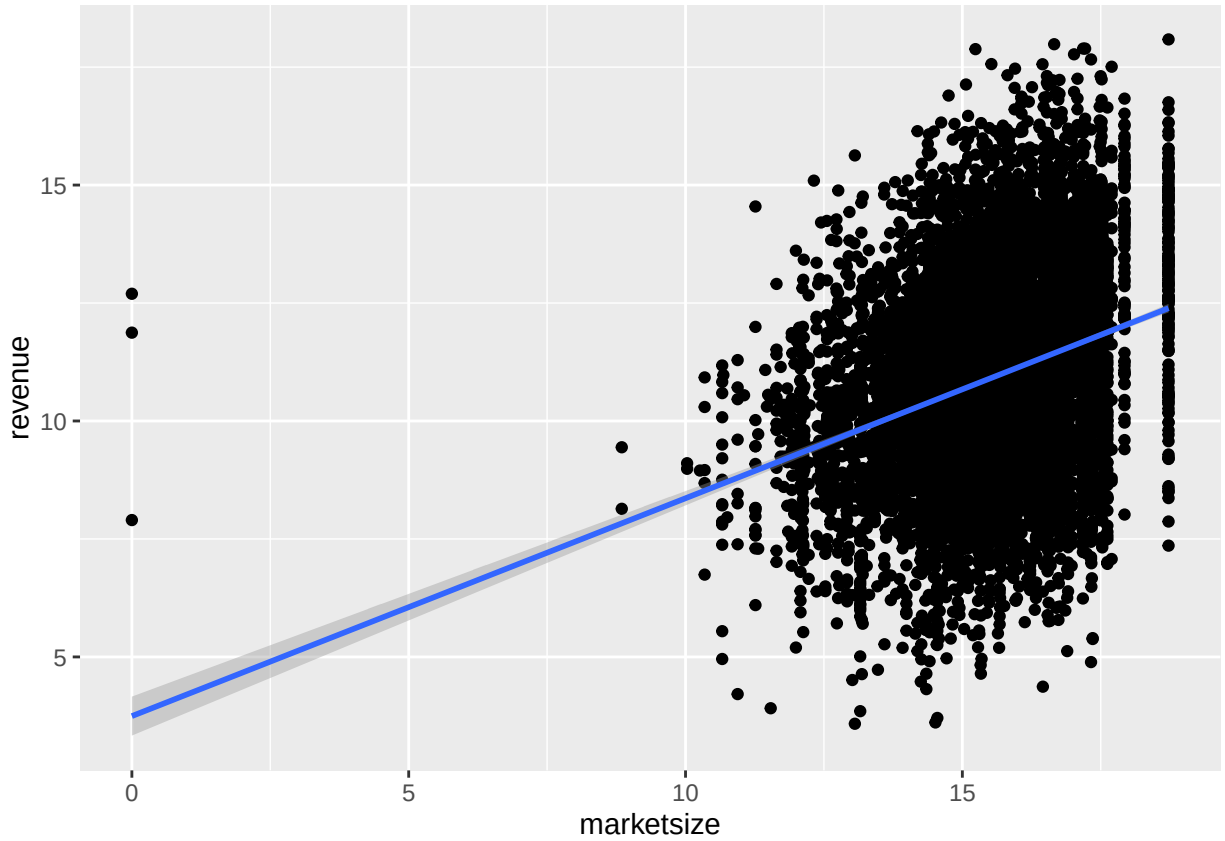
4.3.2.1 `lgdpper2003` (GDP per capita in 2003 (log)) as the Independent Variable

```
##
## Call:
## lm(formula = revenue ~ lgdpper2003, data = mnc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.2752 -1.3110 -0.0171  1.2865  7.2910
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.39905    0.36659   3.816 0.000136 ***
## lgdpper2003  1.02872    0.03949  26.052  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.974 on 11358 degrees of freedom
## Multiple R-squared:  0.05639,    Adjusted R-squared:  0.0563
## F-statistic: 678.7 on 1 and 11358 DF,  p-value: < 2.2e-16
```

The log of GDP per capita in 2003 (lgdpper2003) has a positive inpact on revenue. By increasing the GDP per capita with 1 %, revenue will increase by approximately 1 %. The p-value is less than 0.05 and is significant on a 5% level, and the revenue variable is described by 5.6 % in this model.

4.3.2.2. `marketsize` (Market size) as the Independent Variable

```
##
## Call:
## lm(formula = revenue ~ marketsize, data = mnc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9732 -1.2906 -0.0359  1.2425  8.9507
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.74573    0.21080   17.77   <2e-16 ***
## marketsize   0.46169    0.01348   34.24   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.935 on 11358 degrees of freedom
## Multiple R-squared:  0.09358,    Adjusted R-squared:  0.0935
## F-statistic:  1173 on 1 and 11358 DF,  p-value: < 2.2e-16
```



The market size variable has a positive inpact on revenue. The estimate of market size tells us that 1 % change in market size will result in 0.46 % change in revenue. The p-value is less than 0.05 and is significant on a 5 % level, and the revenue variable is described by 9.4 % in this model.

4.3.2.3 `relationship` (Relationship with clients and suppliers) as the Independent Variable

```
##
## Call:
## lm(formula = revenue ~ relationship, data = mnc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1528 -1.3308 -0.0673  1.2644  7.3152
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.005445   0.066003  136.44   <2e-16 ***
## relationship 0.195060   0.006402   30.47   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.954 on 11358 degrees of freedom
## Multiple R-squared:  0.07556,    Adjusted R-squared:  0.07548
## F-statistic: 928.3 on 1 and 11358 DF,  p-value: < 2.2e-16
```

Relationship has a positive inpact on revenue. By increasing the number of relationships by one, revenue will increase by 19.5 %, holding all other variables constant. The p-value is less than 0.05 and it is significant on

a 5 % level, and the revenue variable is described by 7.5 % in this model.

4.3.2.4 `licenses` (Licenses) as the Independent Variable

```
rev_lic <- lm(revenue ~ licenses, data = mnc)
summary(rev_lic)
```

```
##
## Call:
## lm(formula = revenue ~ licenses, data = mnc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3272 -1.3843 -0.0743  1.3062  7.1159
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.35143    0.04670  221.65   <2e-16 ***
## licenses     0.36451    0.02657   13.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.015 on 11358 degrees of freedom
## Multiple R-squared:  0.01631,    Adjusted R-squared:  0.01622
## F-statistic: 188.3 on 1 and 11358 DF,  p-value: < 2.2e-16
```

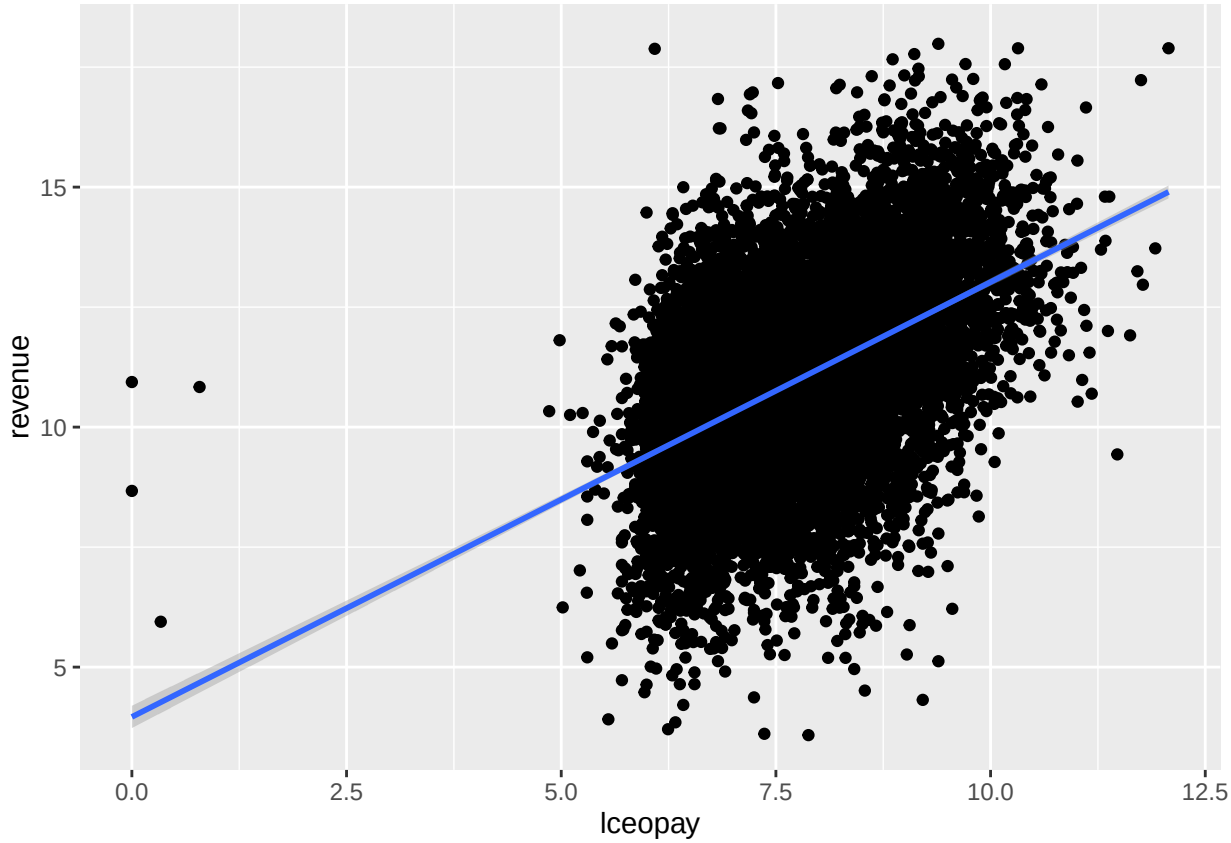
Licenses has a positive inpact on revenue. The estimate of licenses tells us that 1 % change in licenses will result in 0.36 % change in revenue. The p-value is less than 0.05 and it is significant on a 5 % level. However, the revenue variable is described by only 1.6 % in this model. In this case, we would not include this factor when doing the multiple linear regression test.

4.3.2.5 `lceopay` (Log of the CEO pay) as the Independent Variable

```
rev_ceo <- lm(revenue ~ lceopay, data = mnc)
summary(rev_ceo)
```

```
##
## Call:
## lm(formula = revenue ~ lceopay, data = mnc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.9888 -1.1141 -0.0019  1.1664  8.4026
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.96232    0.11840   33.47   <2e-16 ***
## lceopay      0.90593    0.01524   59.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.757 on 11023 degrees of freedom
##   (335 observations deleted due to missingness)
## Multiple R-squared:  0.2428, Adjusted R-squared:  0.2428
## F-statistic:  3535 on 1 and 11023 DF,  p-value: < 2.2e-16
```

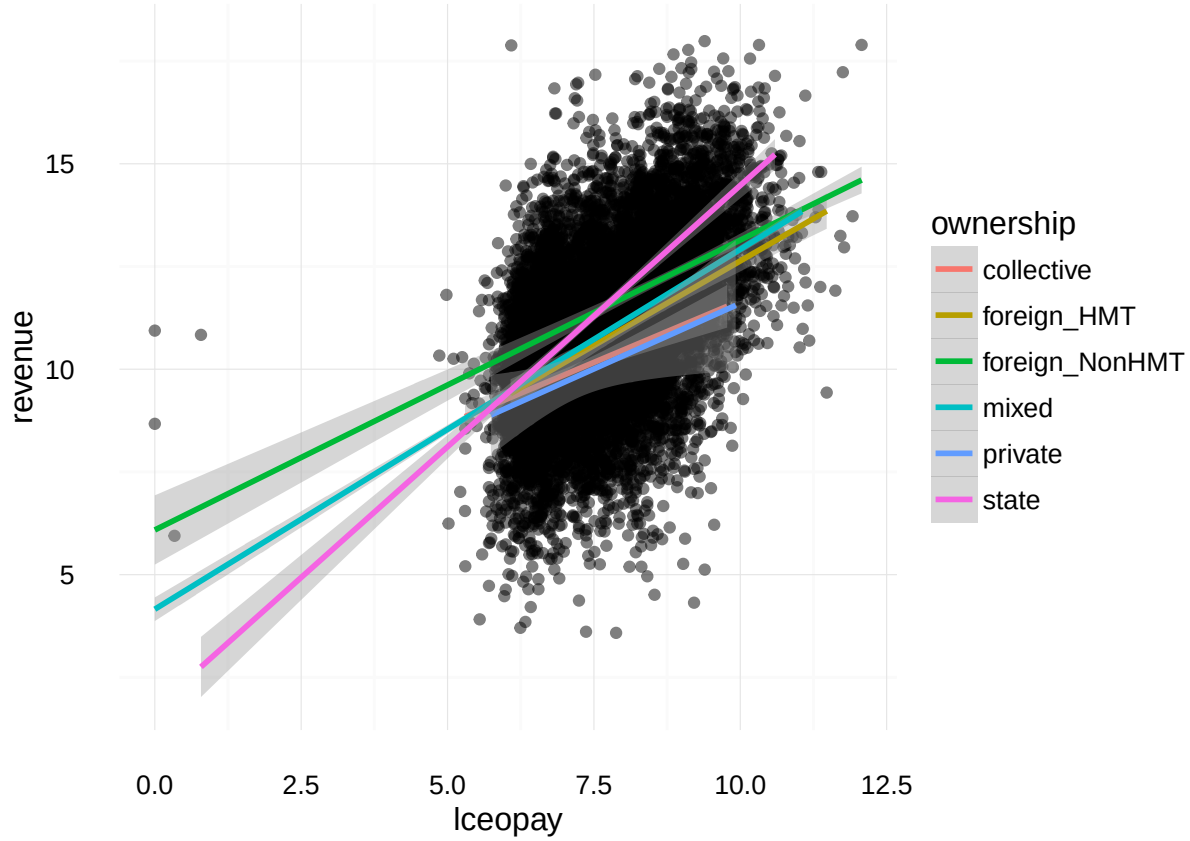
```
ggplot(data = mnc, aes(x = lceopay, y = revenue)) + geom_point(na.rm = TRUE) + geom_smooth(method='lm',
```



4.3.2.6 `lceopay` (Log of the CEO pay) as the Independent Variable with Different Ownership Structure

```
ggplot(data = mnc, aes(x = lceopay, y = revenue)) + geom_point(alpha = 1/2, na.rm = TRUE) + theme_minima
```

```
## Warning: Removed 335 rows containing non-finite values (stat_smooth).
```

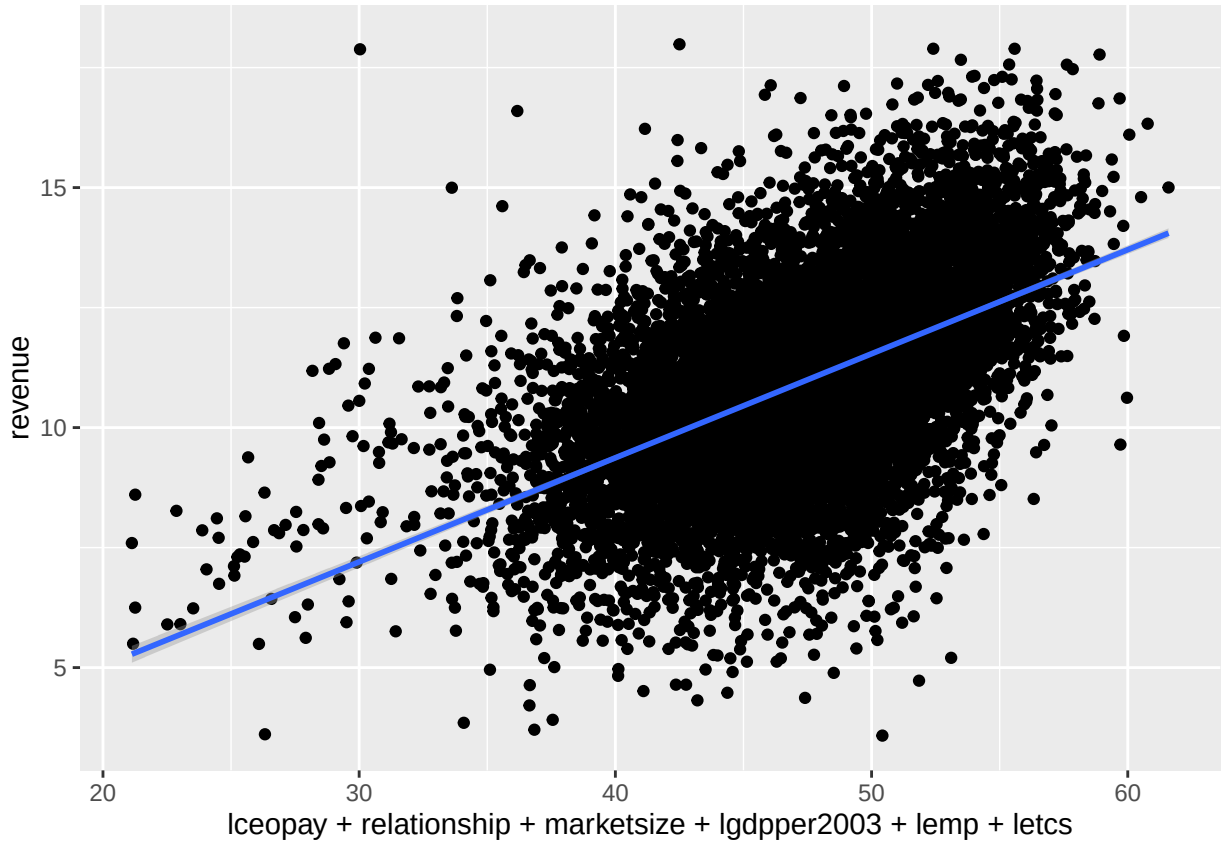


The CEO pay (lceopay) has a positive inpact on revenue. The estimate of lceopay tells us that 1 % change in the CEO pay will result in 0.9 % change in revenue. The p-value is less than 0.05 and is significant on a 5 % level, and the revenue variable is described by 24.3 % in this model.

4.3.3 Multiple Linear Regression Test with `Revenue` as the Dependent Variable

As a result of our correlation analysis and single regression models we would like to look further into the revenue variable by adding relationship, marketsize, CEO pay, number of employees, firm expenditure on entertainment and GDP per capita in 2003 as independent variables.

```
rev_reg1 <- lm(revenue ~ lceopay + relationship + marketsize + lgdpper2003 +
    lemp + letcs, data = mnc)
summary(rev_reg1)
ggplot(data = mnc, aes(x = lceopay + relationship + marketsize + lgdpper2003 +
    lemp + letcs, y = revenue)) + geom_point(na.rm = TRUE) + geom_smooth(method = "lm",
    formula = y ~ x, na.rm = TRUE)
```




```
stargazer(list(rev_ceo, rev_reg1), digits = 2, no.space = TRUE, type = "html",
    title = "Regression Analysis")
```

Regression Analysis

Dependent variable:

revenue

(1)

(2)

lceopay

0.91***

0.40***

(0.02)

(0.01)

relationship

0.02***

(0.004)

marketsize

0.12***

(0.01)

lgdpper2003

0.13***

(0.03)

lemp

0.92***

(0.01)

letcs

-0.17***

(0.01)

Constant

3.96***

-0.75***

(0.12)

(0.21)

Observations

11,025

11,025

R2

0.24

0.72

Adjusted R2

0.24

0.72

Residual Std. Error

1.76 (df = 11023)

1.08 (df = 11018)

F Statistic

3,535.14*** (df = 1; 11023)

4,624.05*** (df = 6; 11018)

Note:

p<0.1; ***p<0.05;*** p<0.01

The linear regression model function:

$$revenue = -0.75 + 0.40 * lceopay + 0.02 * relationship + 0.12 * marketsize + 0.13 * lgdpper2003 + 0.92 * lemp - 0.17 * letcs$$

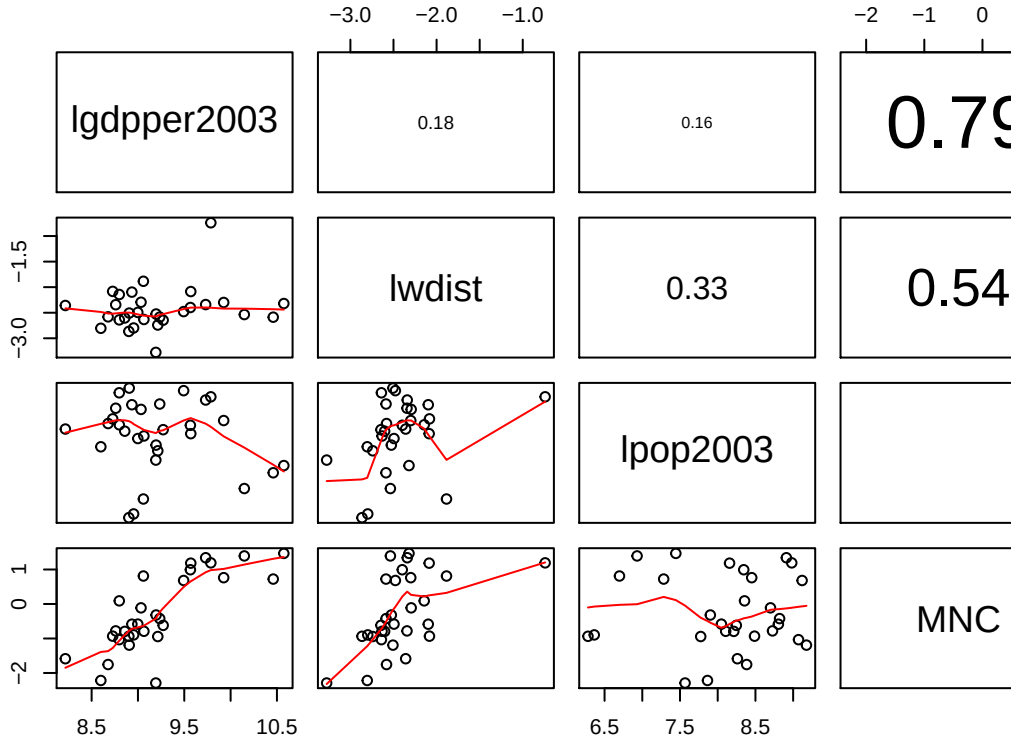
From this regression model we get an adjusted R-squared equal to 0.72, which means that 72% of the revenue variable is described in this model. All the independent variables are significant at the 1% level. The intercept

of -0.75 tells us what the log of revenue will be if all our independent variables are zero. This number is negative and does not give us any valuable information without combining it with more variables.

By looking at the lceopay estimate we can see that 1 % change in the CEO pay will result in 0.4 % change in revenue. If we look at the relationship variable we can see that by increasing the number of relationships by one, revenue will increase by 2.1 %. The estimate of marketsize tells us that 1 % change in marketsize will result in 0.12 % change in revenue. The estimate of lgdpper2003 tells us that an increase in GDP per capita of 1 % will increase revenue by 0.13 %. By looking at the lemp estimate we see that 1 % increase in employees results in 0.92 % change in revenue. The negative relation between letcs and revenue is described by 1 % increase in firms expenditure on entertainment will result in 0.17 % decrease in revenue.

4.3.1. Multiple Linear Regression with `lgdpper2003` (GDP per capita in 2003 (log)) as Dependent Variable

For province-level variables, let's make a regression model to see what kind of variables influence the GDP per



capita and how they influence it.

```
##
## Call:
## lm(formula = lgdpper2003 ~ lwdist + MNC, data = mnc.prov)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.47573 -0.20507 -0.03318  0.11877  0.67695
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.26819    0.38660  21.387  < 2e-16 ***
```

```
## lwdist      -0.44858    0.16155  -2.777  0.00986 **
## MNC          0.49039    0.06258   7.836    2e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3092 on 27 degrees of freedom
## Multiple R-squared:  0.705,  Adjusted R-squared:  0.6831
## F-statistic: 32.26 on 2 and 27 DF,  p-value: 6.969e-08
```

We can find out that the Adjusted R-squared is 0.6831, and we got the linear regression model function that:

$$lgdpper2003 = 8.27 - 0.44 * lwdist + 0.49 * MNC$$

For the linear regression model result, $R^2 = 0.6831$ means 68.31% of variance in Log of GDP per capita can be explained by geographic distance and MNC trades in a province. When holding MNC fixed, every extra increase in $lwdist$ can lead to 44% percentage increasement in GDP per capita. When holding Geographic distance fixed, the change in MNC trades in a province can lead to 49% percentage changes in GDP per capita in the province.

5. Discussion and Conclusion

5.1 Discussion

Based on past reports, the performance of a company is believed to have strong relationship with its connection to the government. Since the Chinese government has a strong interception on its country's economy, and many companies in China are state-owned, we wanted to check whether the revenue of one company is correlated with the governmental factors.

However, in our correlation analysis about the dataset, we could not find any strong correlation between the company's revenue and governmental factors. Therefore, we went further to search for other factors that might have influence on the performance of the company. Beyond this, we found some interesting variables in the dataset that have strong correlation with the company's revenue.

Factors like the payment to the CEO, the number of relationship with clients and suppliers, the market size, the GDP per capita, the number of employee and the expenses on the entertainment have an obvious impact on the performance of the company with a high explanation (72%) of the varibility of the revenue change. We also found that different types of ownership resulted in a mixed impact on the revenue change. Being owned by the state did not immdediately give a company significant advantage.

Moreover, there was a negative correlation between entertainment expenditures and revenue. The variable that Zhu (2015) believed to be an indicator of political relationship and potential corrupt practices does not necessarily improve financial performance. The limitation could be, high entertainment expenditures in a particular year is a process of relationship building. The implication could take time. Moreover, there is no records of entertainment expenditures from previous years in this dataset, which could possibly illustrate the connection better.

Other things we discovered in this analysis were:

The MNC activity and GDP per capita had a high positive correlation, which indicates that FDI (foreign direct investment) could stimulate the economy of China.

Around 73.8 per cent of the companies in the sample is owned by a mix of foreign investors, private and government. None of them held more than 50 per cent of the shares. The number of private owned companies are surprisingly very small and their average revenue is lower than the other firms.

References

Hlavac, M. (2015). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2. http://CRAN.R-project.org/package=stargazer

You, J. & Du, G. (2012). Are Political Connections a Blessing or a Curse? Evidence from CEO Turnover in China, *Corporate Governance: An International Review*, 20(2). Accessed through: EBSCOHost.

Zhu, B. (2016). *Replication Data for: MNCs, Rents, and Corruption: Evidence from China*, doi:10.7910/DVN/VT4KJA, Harvard Dataverse, V1 [UNF:6:LTJM6CHylp/JU5AfkjBK8Q==]