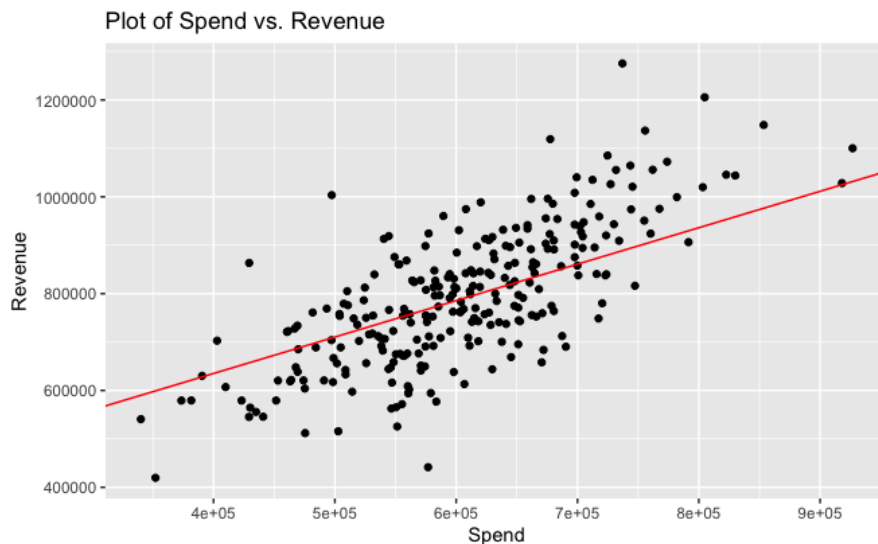


Advertising Report

We are looking at financial marketing data on 267 start-up companies working in the tech sector. The data gives us Spend, the amount of capital resources that were spent initially on the first year of each company and Revenue, the amount of revenue generated during the first year of the company.

Looking at the data, there is one outlier that stands out. It says that the company spent \$1,255,897 and reported \$0 revenue. I'm assuming this is some sort of error. Maybe the user did not report revenue? For this reason, I removed the data point.

1) The relationship between spend and revenue is positive. Seen in the plot below. Generally, more money spent leads to more revenue.



2) The below table predicts the Profit for Spend at five different values. For example, a company that spends \$500,000, we can predict with 95% confidence that their profit will fall between (-9653, 369,660). We can see that profit slightly increases with the more money spent.

Spend	Lower Bound Profit	Average Profit	Upper Bound Profit
500,000	(9,653)	180,003	369,660
550,000	(4,806)	184,548	373,901
600,000	(149)	189,092	378,333
650,000	4,317	193,637	382,956
700,000	8,592	198,181	387,770

Math 536 Homework 2

Shamir Naran

Due February 20, 2020

In this dataset, you will find financial marketing data on 267 start-up companies working in the tech sector.

```
hw2 <- read.csv("hw2.csv", header=T, na.strings="?")  
# We read in the data using the read.csv function.  
# The header = T tells R that the first line of the file contains the variable names.  
# na. strings tells R that when it sees a particular set of characters,  
# it should be treated as a missing element.
```

The variables within this dataset are spend and revenue. Spend contains the amount of capital resources that were spent initially on over the first year of each company. Revenue quantifies the amount of revenue generated during the first year of the company.

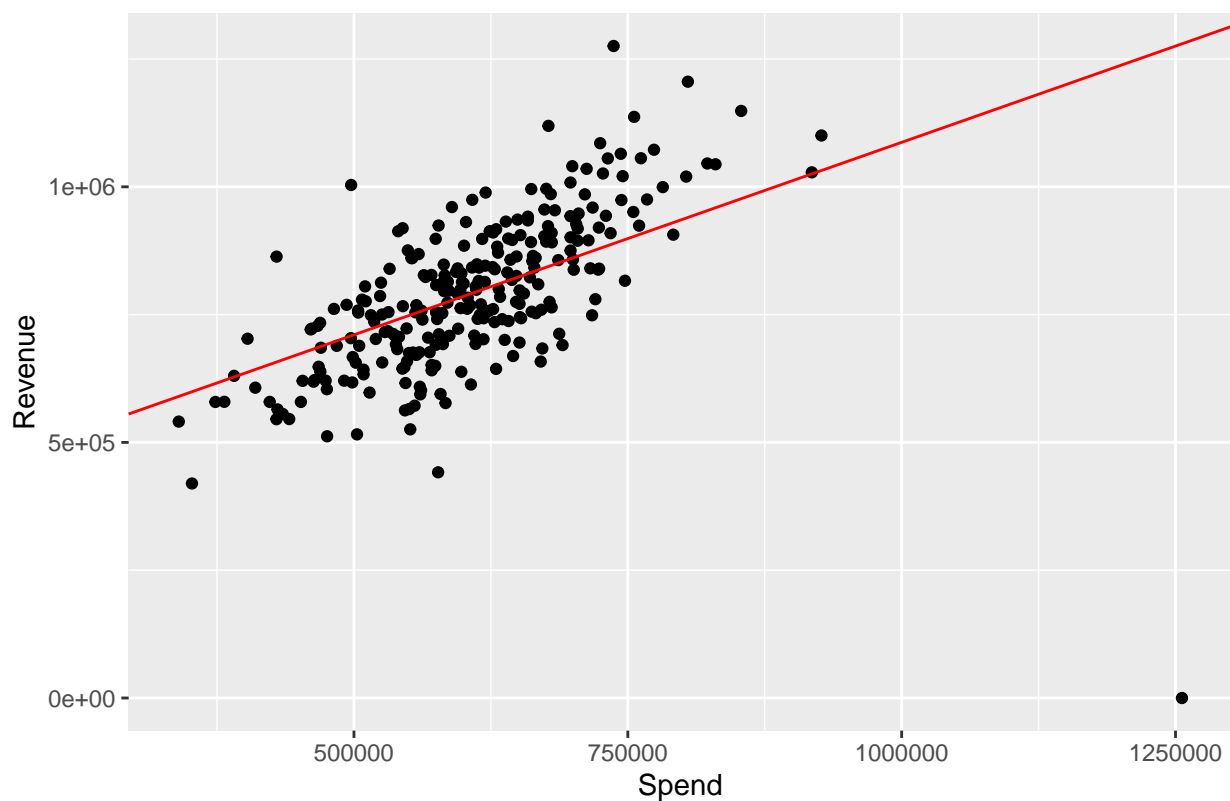
1. What is the relationship between advertising spend and revenue within the first year of a start-up?

```
library(tidyverse) # loads readr, dplyr, ggplot2 and other useful packages
```

```
lm <- lm(Revenue~Spend, data = hw2) # fit linear model  
coef <- coef(lm) #extract coefficients of the linear model
```

```
ggplot(data = hw2, mapping = aes(x = Spend, y = Revenue)) +  
  geom_point() +  
  geom_abline(intercept = coef[1], slope = coef[2], col = "red") +  
  labs(title = "Plot of Spend vs. Revenue")
```

Plot of Spend vs. Revenue



```
cor(hw2$Spend, hw2$Revenue)
```

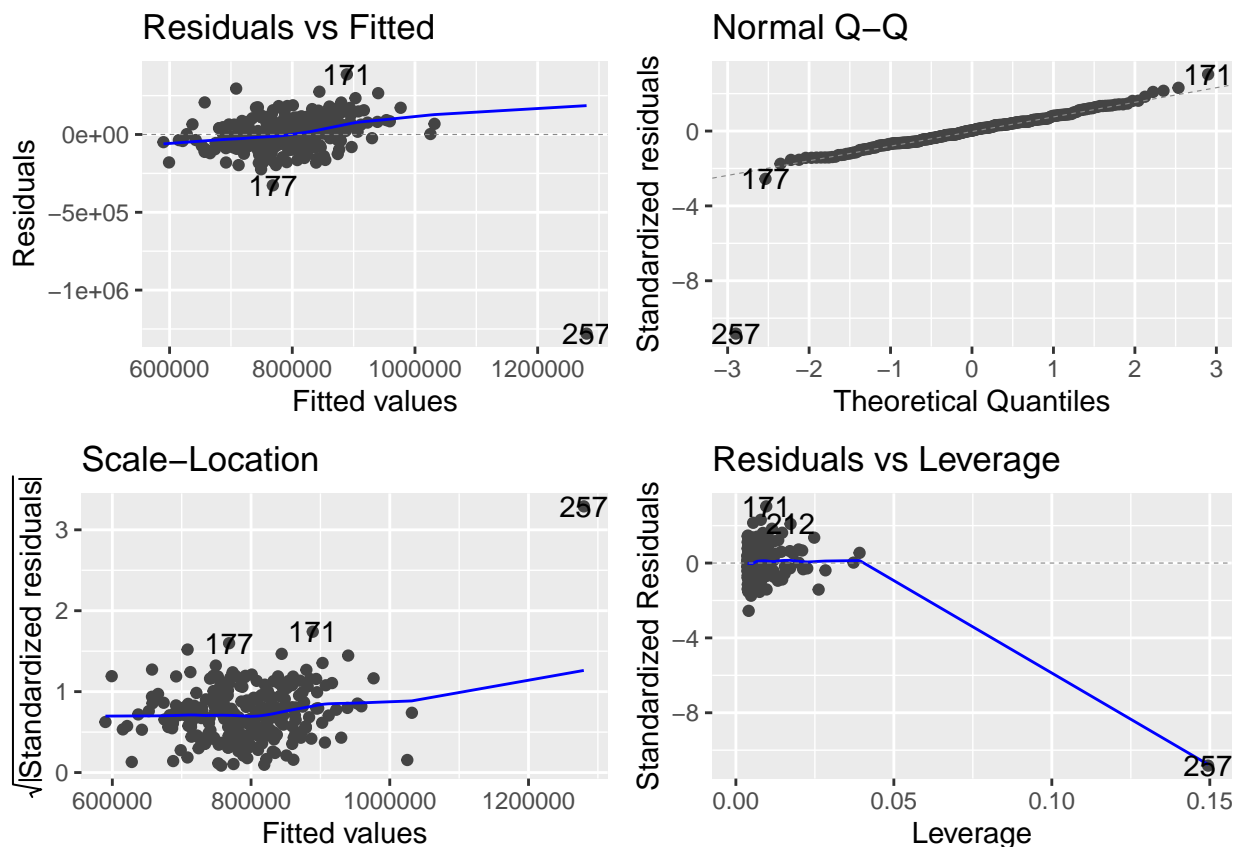
```
## [1] 0.522981
```

There is a positive relationship between Spend and Revenue. The correlation is 0.522981. This suggests that the more money spent leads to higher revenue.

Let's check our 4 assumptions statistical inference.

```
library(ggfortify) # for diagnostic plots
```

```
autoplot(lm)
```

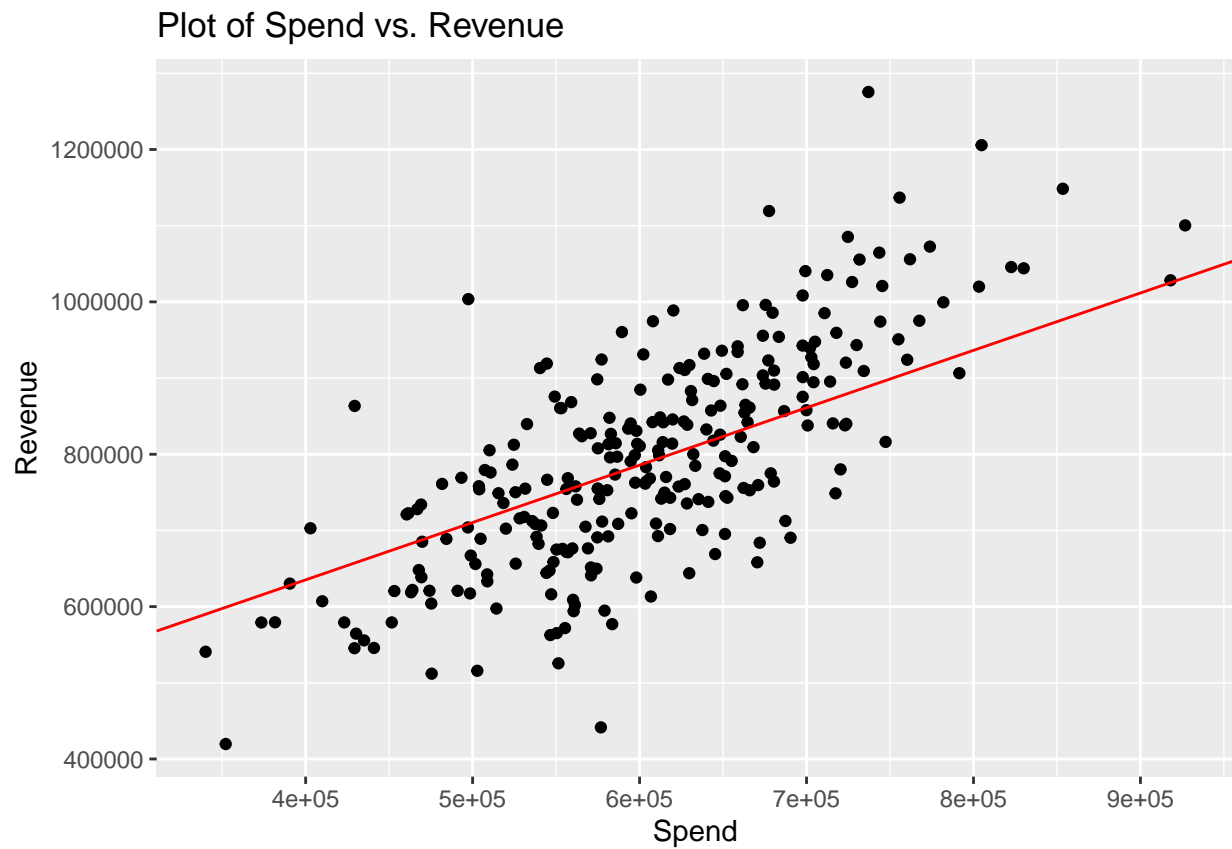


Our outlier is having a drastic impact on the plots. The outlier says that \$1,255,897 was spent with no corresponding revenue. This seems strange. Maybe it's a human error and the person forgot to enter revenue? The data point is so unusual, I decide to remove it.

```
hw2.new <- hw2[-257, ] # remove line 257
```

```
lm.new <- lm(Revenue~Spend, data = hw2.new) # fit linear model
coef.new <- coef(lm) #extract coefficients of the linear model
```

```
ggplot(data = hw2.new, mapping = aes(x = Spend, y = Revenue)) +
  geom_point() +
  geom_abline(intercept = coef[1], slope = coef[2], col = "red") +
  labs(title = "Plot of Spend vs. Revenue")
```

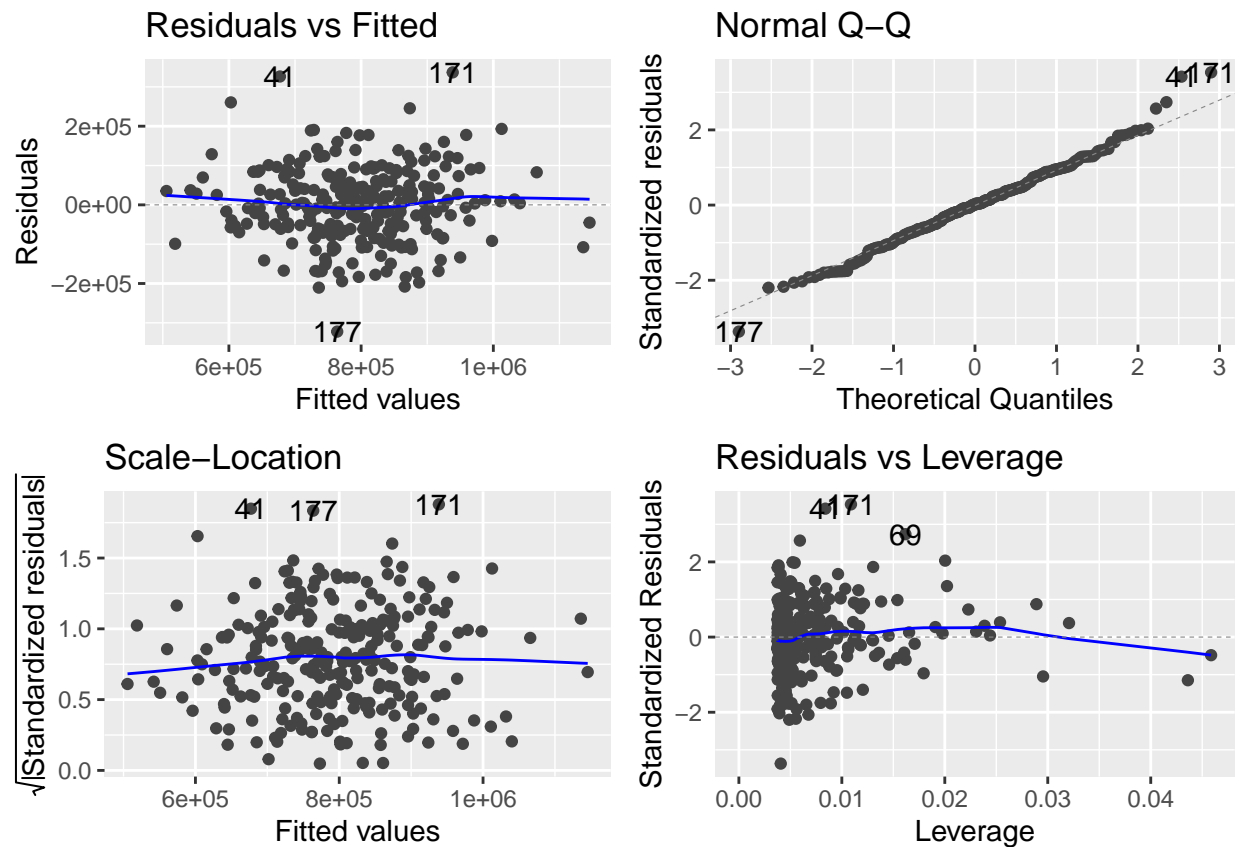


```
cor(hw2.new$Spend, hw2.new$Revenue)
```

```
## [1] 0.7397931
```

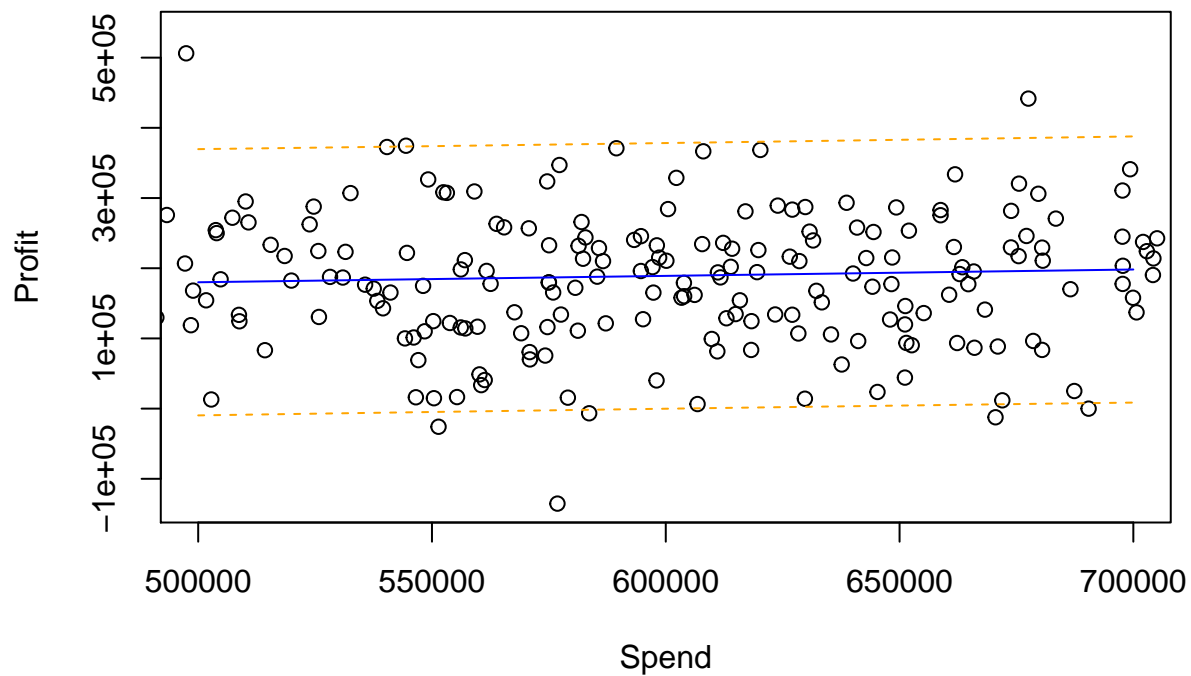
Again, there is a positive relationship between Spend and Revenue. Removing our outlier has increased the correlation to 0.7397931.

```
autoplot(lm.new)
```



Our plots look a lot better after removing the outlier.

```
# Predict Profit based on Spend
newdata <- data.frame(Spend = seq(500000,700000, by = 100))
predictrev <- predict(lm.new,newdata, interval="predict")
plot(hw2.new$Spend, hw2.new$Revenue-hw2.new$Spend, xlim = c(500000,700000),
     xlab = "Spend", ylab = "Profit")
lines(newdata[,1], predictrev[,1]-newdata[,1], col="blue")
lines(newdata[,1], predictrev[,2]-newdata[,1], col="orange", lty=2)
lines(newdata[,1], predictrev[,3]-newdata[,1], col="orange", lty=2)
```



```
# Data frame to show Profit corresponding to Spend
data2 <- data.frame(cbind(newdata[,1],
                          predictrev[,2]-newdata[,1],
                          predictrev[,1]-newdata[,1],
                          predictrev[,3]-newdata[,1]))
```

```
# Print specific values
```

```
report <- filter(data2, X1 == 500000 |
                  X1 == 550000 |
                  X1 == 600000 |
                  X1 == 650000 |
                  X1 == 700000)

report
```

```
##      X1      X2      X3      X4
## 1 500000 -9652.8613 180003.4 369659.6
## 2 550000 -4805.5941 184547.8 373901.1
## 3 600000 -148.9078 189092.2 378333.3
## 4 650000  4316.8586 193636.6 382956.3
## 5 700000  8591.9428 198181.0 387770.0
```