Exercise 1

Jaeho Jang

1. Data Visualization: flights at ABIA

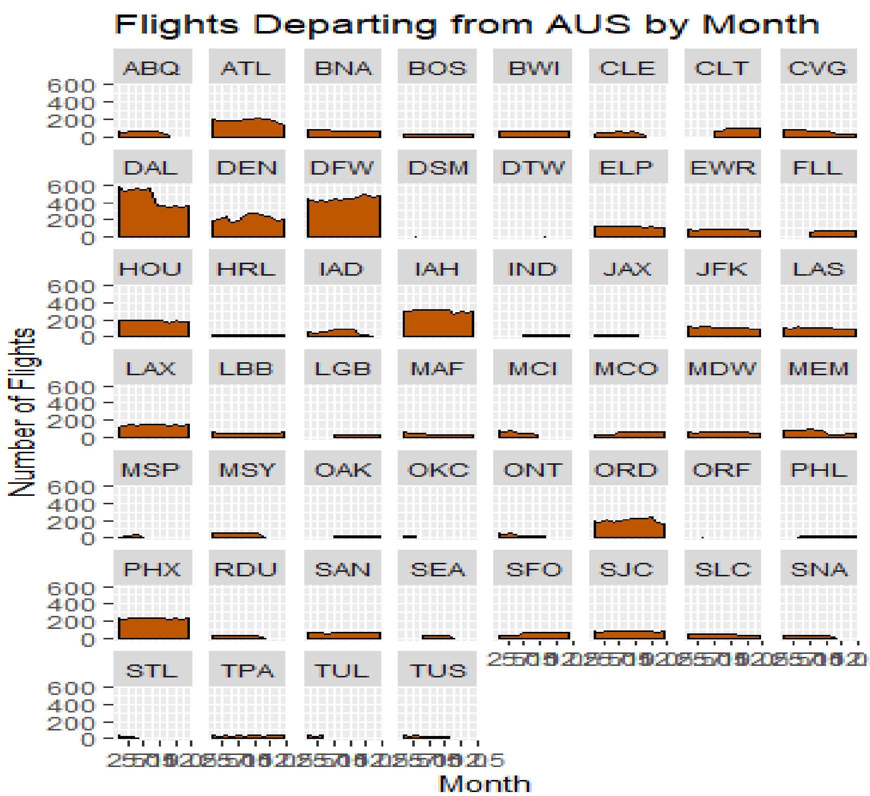
1) Overview

For this analysis, I want to see if:

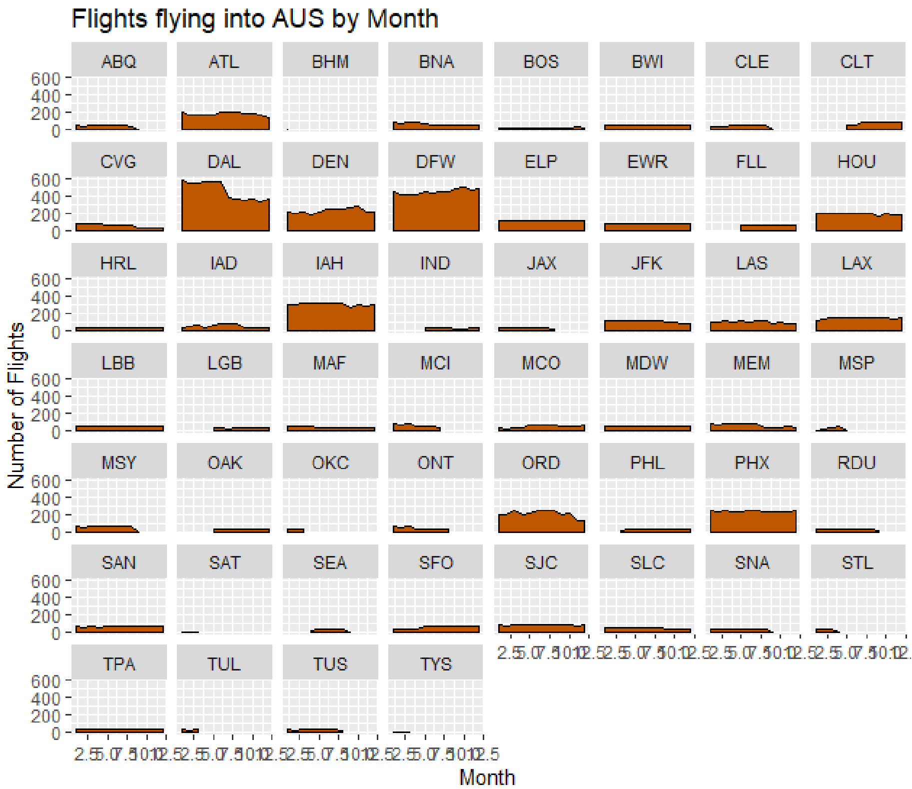
- Flights where actually departing in time and,
- To where people where leaving the most from Austin and,
- From where people where coming to Austin the most

2) Modeling

First, I created a regression by modeling two variables: flights departing from Austin and the month. When they were put into a regression, the distribution looked as in the graph below.



Furthermore, I modeled from where people were coming the most to Austin using each month with two variables: flights flying into Austin and the month. When put into distribution, it looked like the model below.



3) Conclusion

According to our first distribution, it seemed like people were leaving the most to Dallas and Denver. This can be accounted to the geographic closeness of these locations. The second model tells us that similar to the former model, people flight from Dallas and Denver the most to Austin. Accordingly, it could be concluded that people who use the airport are usually from the same area; there is not a variety of people coming from different places to the Austin airport.

2. Regression Practice

1) Overview

In the regression analysis, I wanted to see:

- What creatinine clearance rate should we expect, on average, for a 55-year-old?
- How does creatinine clearance rate change with age?
- Whose creatinine clearance rate is healthier for their age: a 40-year-old with rate of 135, or a 60-year-old with a rate of 112?

2) Modeling

The following regression was conducted to obtain the coefficient from looking at the equation (relationship) between variable creatclear and age.

```
#lm_creat=lm(creatclear~age,data=creatinine)

#coef(lm_creat)

#rate=-0.6198*age+147.8129
```

The regression returns us a coefficient of -0.6198; from this we can conclude that our equation for the two variables is rate = -0.6198*age + 147.8129. Lastly, I ran the following regression to get the residual of each instance; a 40-year-old with a rate of 135 and 60-year-old with rate of 112.

```
#resid_creat=resid(lm_creat)

#resid_creat %>% which.min

#age_data=data.frame(age=c(40,60))

#predict(lm_creat,age_data)

#resid135=135-123.0203

#resid112=112-110.6240
```

3) Conclusion

The regression returns us a coefficient of -0.6198; from this we can conclude that our equation for the two variables is rate = -0.6198*age + 147.8129. To find the creatinine clearance rate for age of 55, we simply plug in the value 55 into the equation; thus, on average, we expect creatinine clearance rate at age of 55 to be **113.7239**. From the coefficient, we can also conclude that the creatinine clearance rate change **decreases by .6198 mL/minutes for each year increase**. Furthermore, The residual of 40-year-old came out to be 123.0203 and that of 60-year-old came out to be 11.06240. Accordingly, we can conclude that the **60-year-old has a healthier rate of 1.376**.

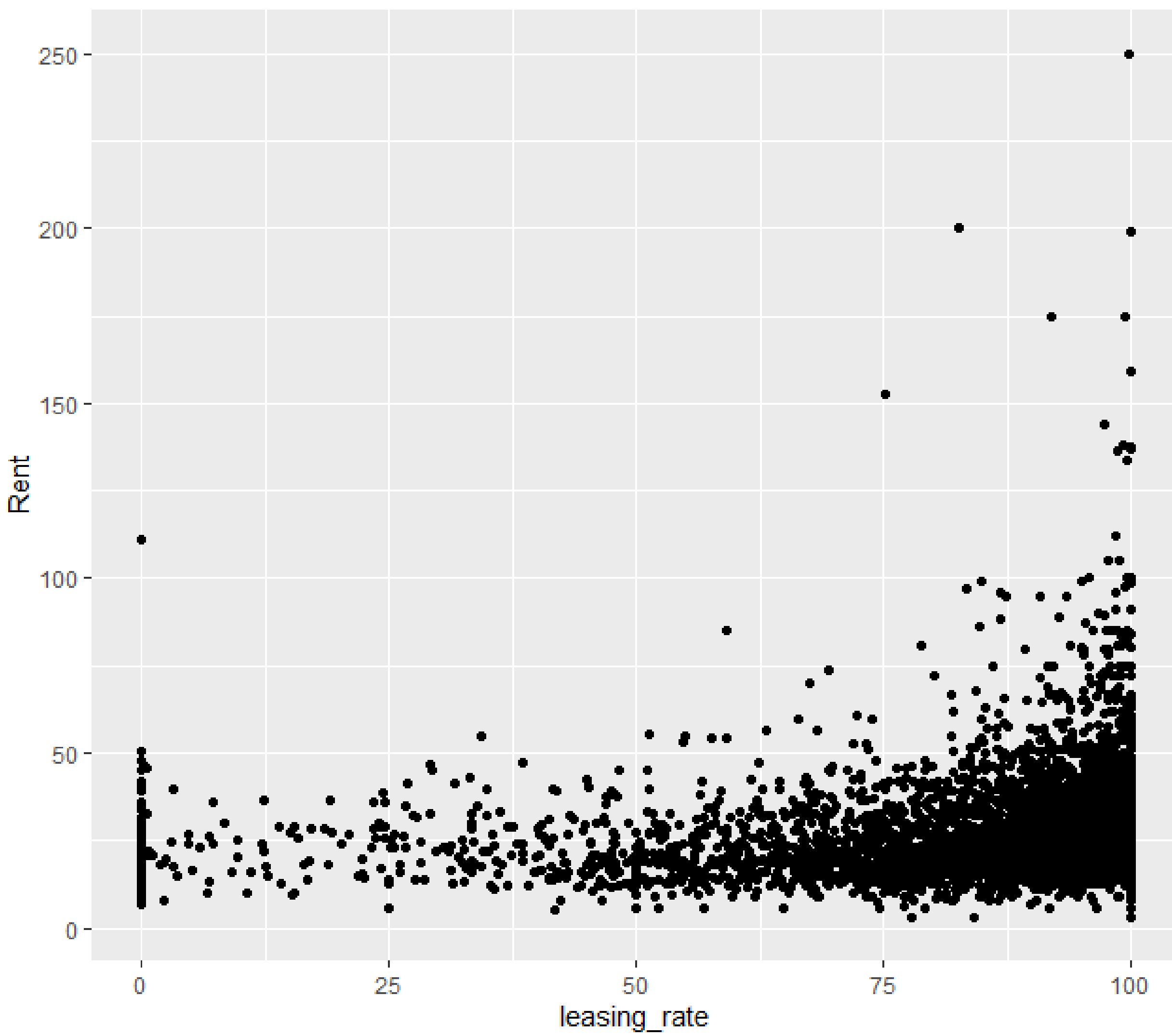
3. Green Building

1) Overview

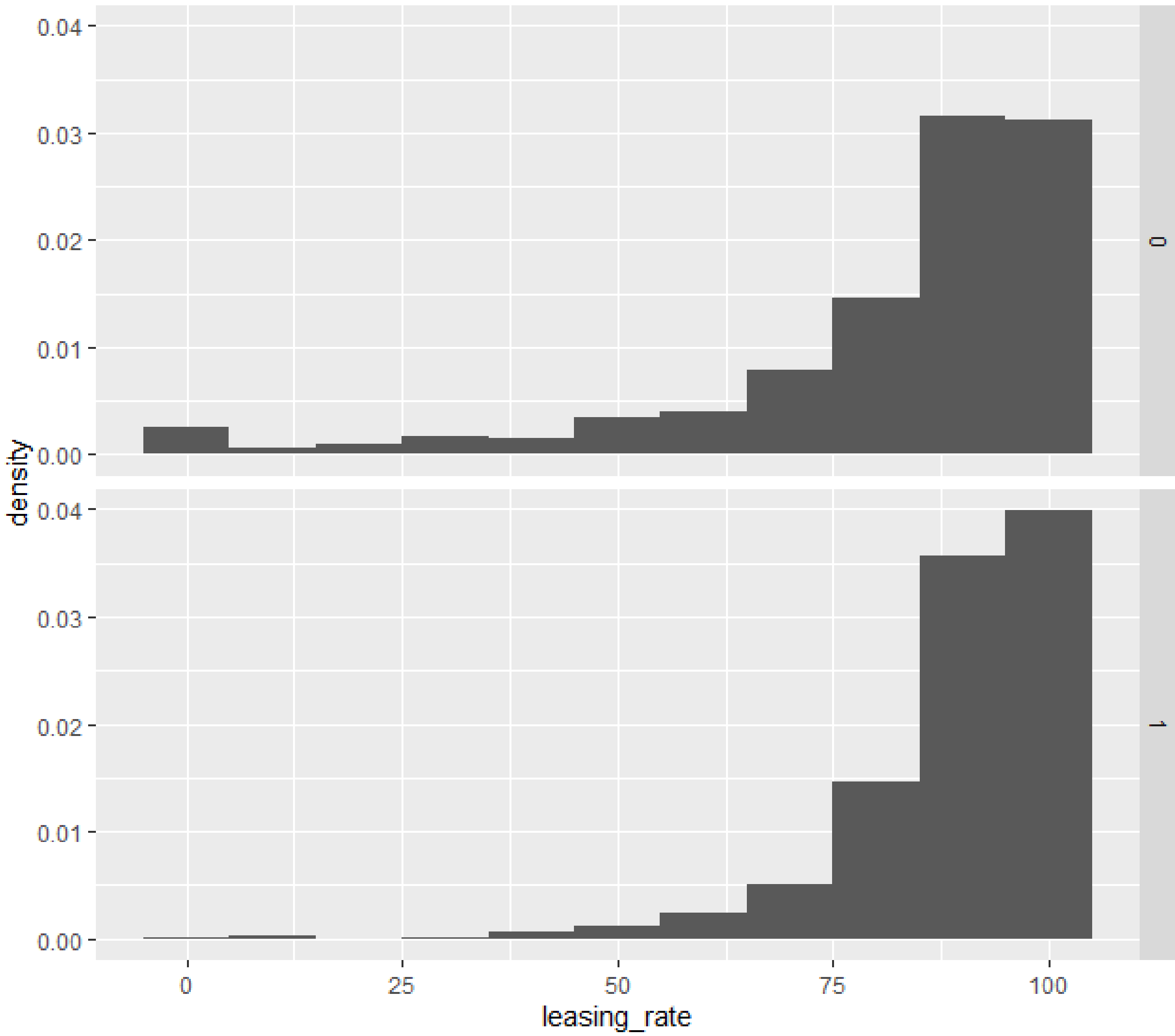
When looking at the environment friendly buildings, its is most important to recognize the economic potential of these product for the investors. Accordingly, I decided to check if a certain variable could predict a trend within the rent variable and also further investigate if there was a difference in non-green buildings and green buildings if this variable showed to be effective.

2) Model Building

First, I worked with the leasing rate variable and checked if the leasing rate variable could predict the rent of the buildings when they were put into a regression model. Below is the regression model of two variables.



As seen from the plot, there seems to be a definite trend between these two variables: rent goes up by 13 cents every leasing rate. So accordingly, I ran two different regression with the leasing rate variables: one with green buildings and another with non-green buildings.



When looking at the two regressions, both buildings looked like they had a similar trend in their relationship with leasing rate: higher the leasing rate, higher the rent.

3) Conclusion

In conclusion, I do agree with the guru's opinion to invest. However, I would be more specific to tell the investor and recommend that it would be more wise to invest in economic buildings with higher lease rate, since they tend to have a much higher rent price than compared those with relative lower lease rates.

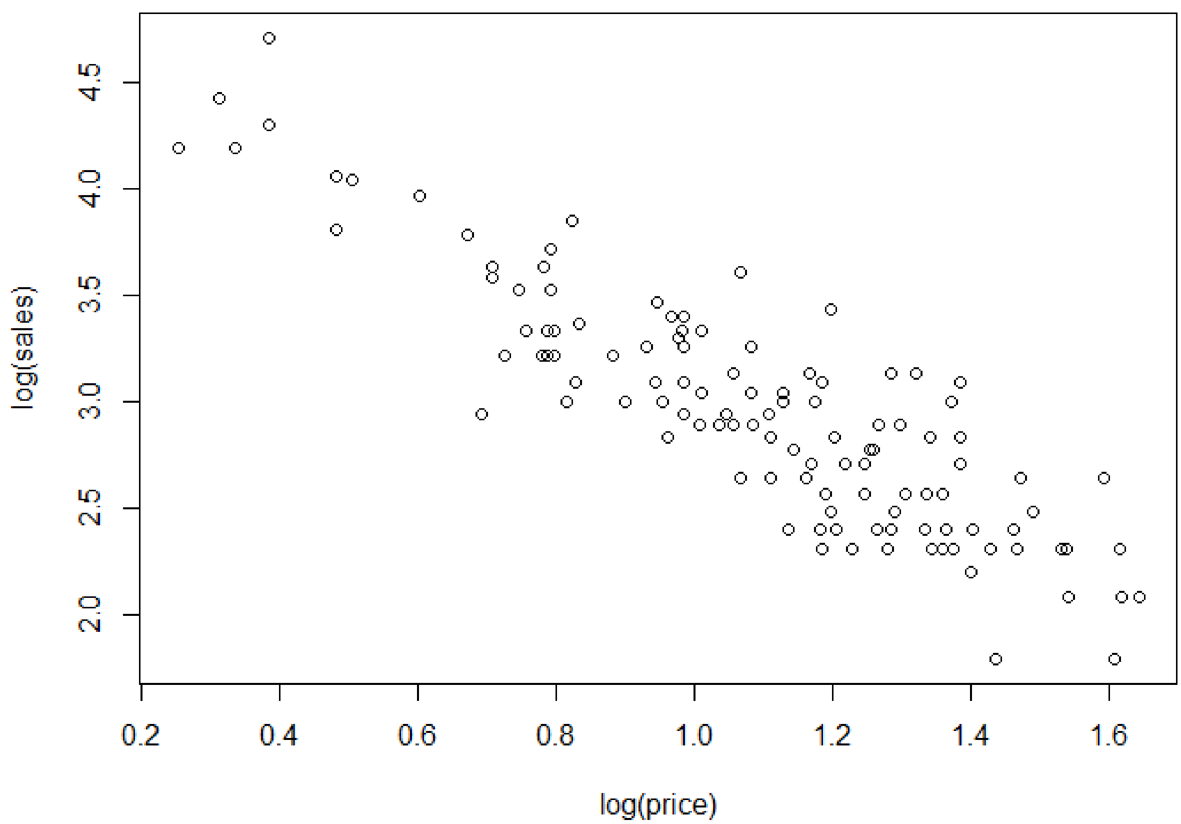
4. Milk Prices

1) Overview

In this analysis, I tried looking for the most optimal price by finding the change in quantity according to price (coefficient) from a log equation, and then putting in the found into a exponential equation to conclude the optimal price for milk and the net profit at the price.

2) Modeling

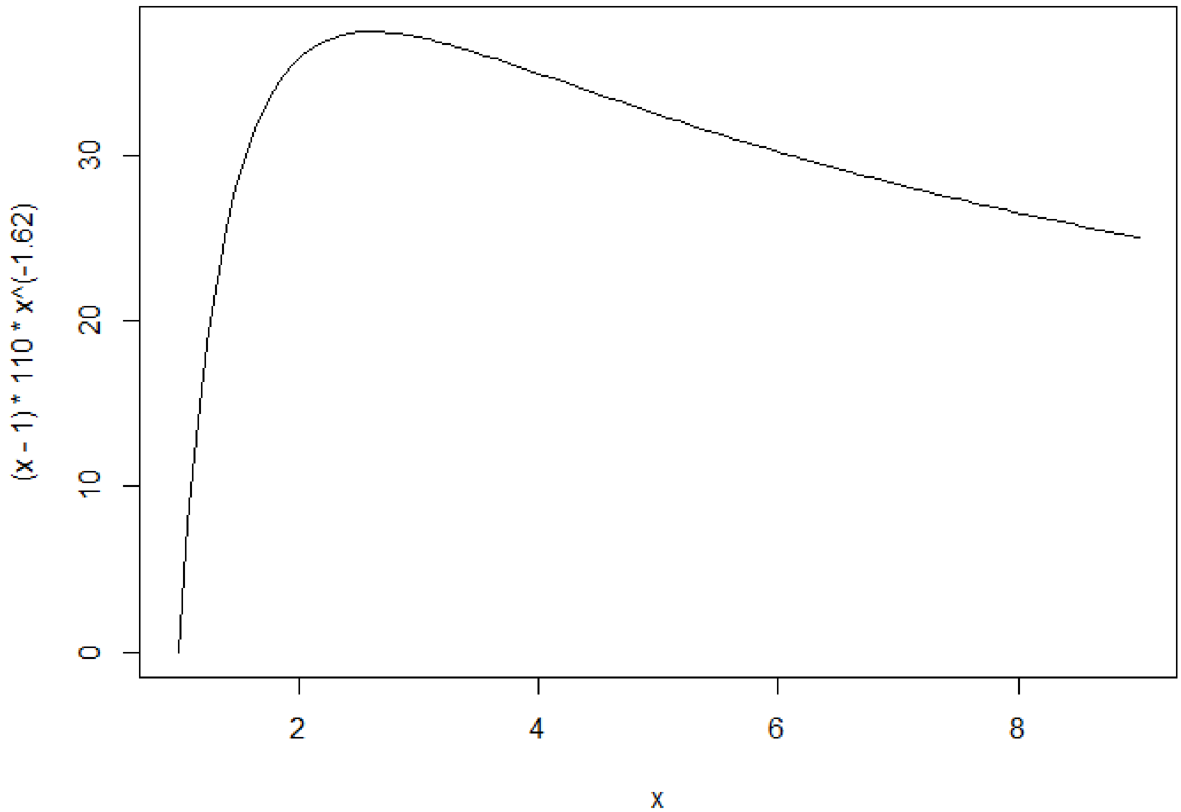
First, I ran a regression between the two log-ed variables - sales and price - to get the coefficient. Below is the relationship plot derived from two variable.



#log(Q)=4.7-1.62*log(P)

#Q=e^4.7*P^-1.62

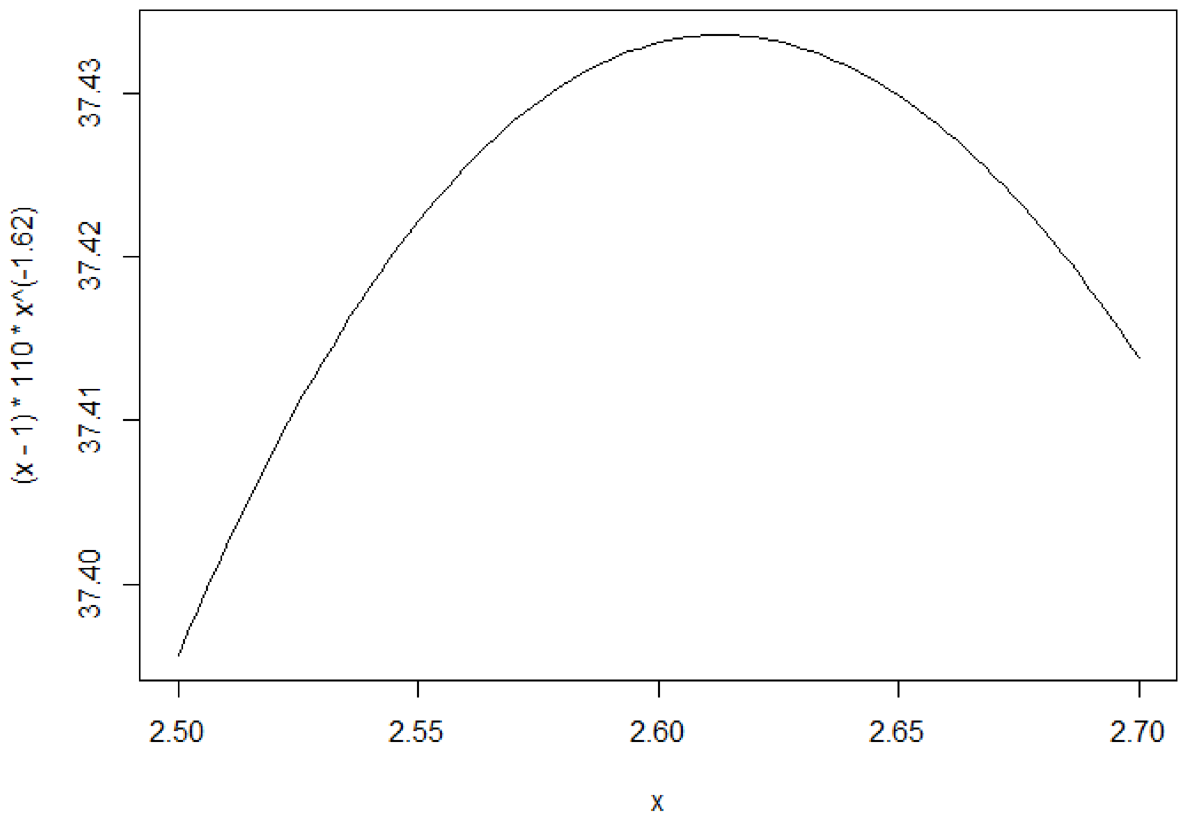
Looking at the following plot and equation, we can now use it to create a new equation that actually predicts the net sales amount at its optimal price. Following is the net sales curve using the price coefficient.



#N=(P-C)*(e^4.7*P^-1.62)

#N=(P-C)*(110*P^-1.62)

From the graph, we could see that the maximum net sales occurred between somewhere from 2.5 to 2.7, so we decided to expand the graph into those specific ranges.



3) Conclusion

From the final graph, we can concluded that our optimal price is around **$2.61**, and the maximum net sales that occurs at the optimal price is around **$38.24**.