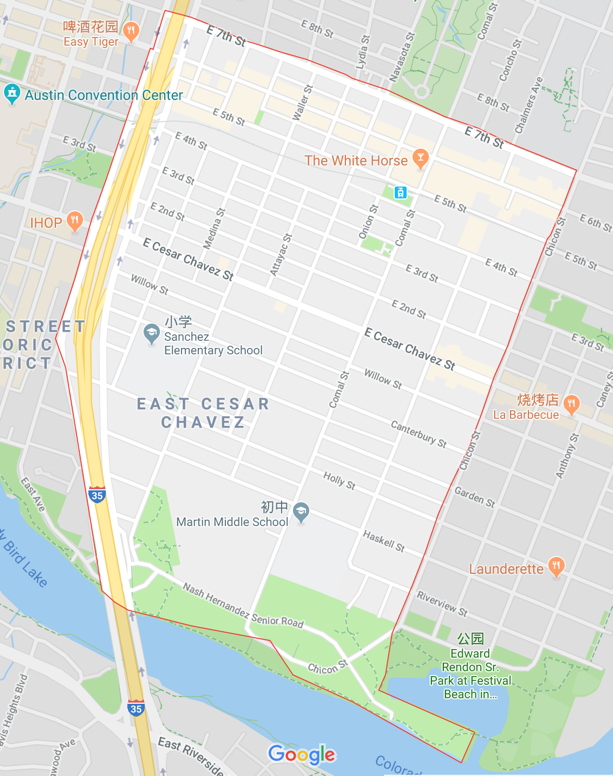
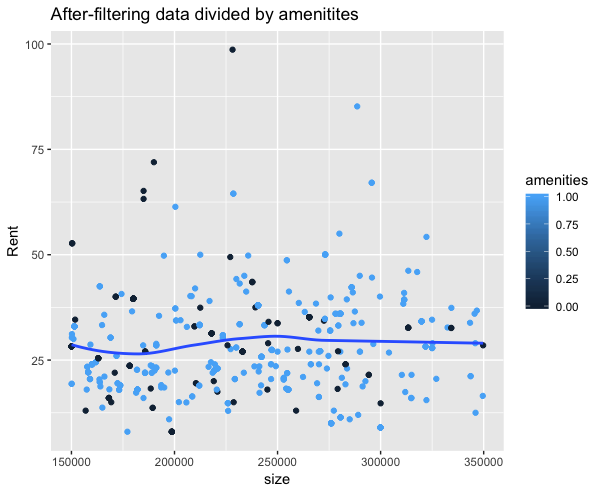
**Part I**

Comparing with analysis by “data guru”, one of the mistake we think is the data he uses is too general, even if his data is big (has many observations), but it cannot reflect the exact result that the developer needs, so before analysis the data, it is better to find out the data which is more suitable and proper for the project. The detailed to filtrate data in our model is shown below:

Austin real-estate developer focuses on 15-story mixed-use building, according to the information, our model limits the range of the stories from 12 to 18, the aim is to use the data which is similar as the developer’s goal. And to test our limitation is right, we run the correlation between Rent and size from all data (includes both green and nongreen buildings) through R. The correlation is 0.116 which means when the stories is higher the rent is bigger, so it is necessary to limit the stories. Same as the size of the building, from the staff’s word, the size of the targeted building is 250,000 square feet, and correlation between size and rent is 0.138, so in our model, we limited size of the building between 150000 feet and 350000 feet. Another filter to the data is because the project located around downtown, and from the google map, we double check about this, it is located in a convenient place. As the graph below shows:



From the detailed of the google map (the above graph does not show), this area has bar, gym, restaurant and so on. So as for the “amenities” indicator, it should be 1. (dummy variable) After the filtering, the rest of data which has 367 buildings will be used for analyzed in our model. And before filtering the amenities, the data is bellowed, the blue one is amenities=1, and black one is amenities=0.



The data shows most of the data comes from the building which possesses good amenities. And mean of the blue point rent is 28.16 dollar(amenity=1), the other one is 30.02dollar. So it is necessary to move out the data which amenity=0.

**Part II**

There is one point we strongly disagree with the “data guru”. He uses difference between the rent between green house and none green house, and work out their total revenue difference each year. And work out how many years can green house recuperate the cost difference.

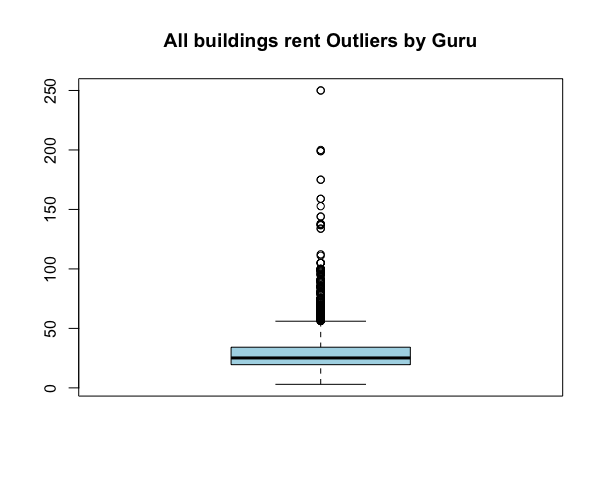
But personally, we think this only using this method is too limited. Developer wants to know “will investing in a green building be worth it, from an economic perspective value.” There are several possible results should analysis think about:

1. If the net profit of greenhouse >0 (Suppose the renting year is at least 30 years , the rest possible results are same) and net profit of the greenhouse> net profit of non-green house,(from economics perspective) investing greenhouse
2. If the net profit of greenhouse>0 but net profit of greenhouse< net profit of non-greenhouse, better not invest greenhouse
3. If the net profit of green house<0 , definitely not consider green house, may consider nongreen house if net profit of it>0

So in our model, we not only analysis the how long can we recuperate the cost difference if we choose green house. And how long can we recuperate the cost if we choose to invest green house. And how long can we recuperate the cost if we choose to invest nongreen house.

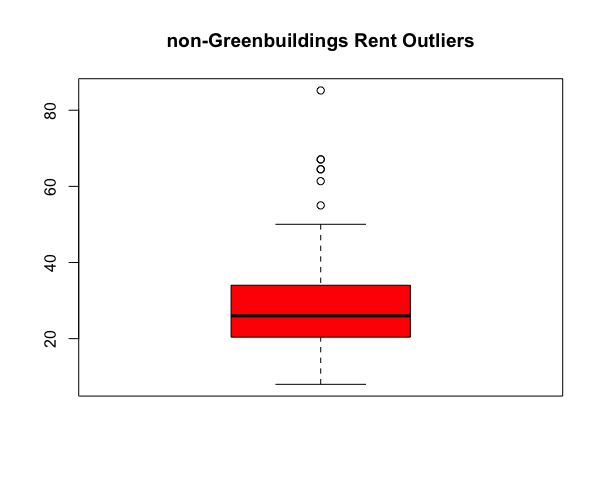
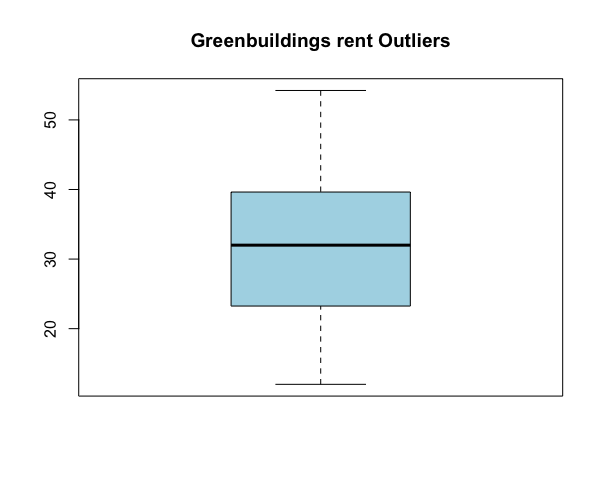
**Part III**

When analyzing the data, guru uses the median of the data to predict its rent, because he thinks there will be many outliers that influence the prediction of the data, so mean cannot be used. The direction that he thinks is right, however, using median still cannot solve the problem. If the outlier only focusses on one side, (outliers are all lying in the high rent or all in low rent, the prediction still has a big problem) Our group using “boxplot” method to find out outliers of the overall data (no filter as guru’s data), the graph is shown below:

****

The graph illustrates that all the outliers are upper, and using R for calculation, we work out that the data has 289 outliers, and all the outliers are bigger than upper quartile+1.5IQR and no outlier is smaller than lower quartile-1.5IQR. Consequently, the estimation method that guru uses is not right. Because all the outliers come from upper side, so the estimation uses median will be bigger than the estimation without outliers. In this case, in order to make an improvement, our group uses boxplot to remove all the outliers and using the rest of data’s mean to estimate the rent.

After using the boxplot for green building and non-green building using the after-filter data. The two box plots are shown as below respectively.



As the graph shows above, the green buildings data does not have any outliers, and non-green buildings has 7 outliers from the observations. And in our model, we remove the 7 outliers. And finally, our model calculates the average mean in each data group respectively. And our result shows mean of the green house rent is 31.74 dollar per square foot per calendar year, and the mean of the non green house rent is 26.77 dollar per square foot per calendar year. So the revenue calculation for each group is shown below:

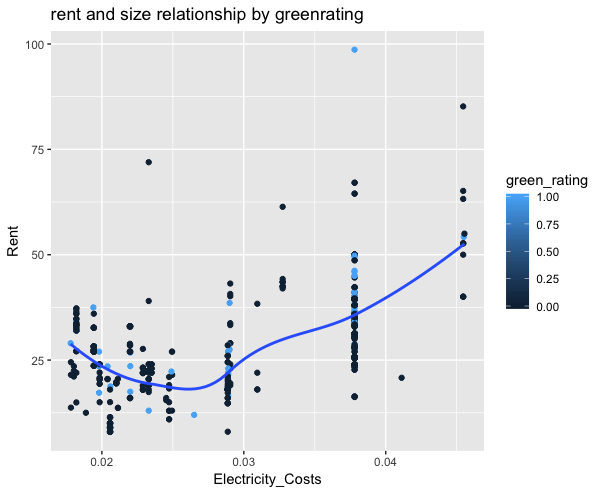
(rent rate using mean of each data sets’ rent rate, working process is in R)

Green house revenue each year = 722679564 , construction cost=105000000

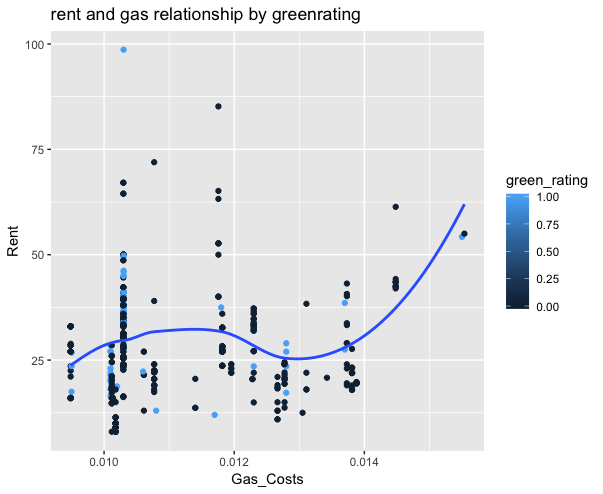
Non Green house revenue each year = 559281853 , construction cost=100000000

So unlevered cash on cash return for green house is 105000000/722679564=14.53 years, this means in 14.53 years the investor’s rent revenue will cover its investment cost. And for non green house revenue, unlevered cash on cash return for the non green house is 100000000/559281853= 17.88 years. So comparing to the these two number, it shows that investing green house is better, because the cost can be covered in 14.53 years which is shorter than nongreenhouse. And also, it is necessary to calculate for how long, can green house recuperate 5 million dollars which is the difference between cost of green-house and non green house. So using5000000/(722679564-559281853)=3.06 years.

Furthermore, for the variable cost. The electricity cost and relationship is shown as below using the “after filtering data”:



We can see that after the rent is more than maybe 20 dollars ,the more the rent is and the more the electricity fee is. And the average rent for non green house is greater than green house, and it is more than 20 dollars, so this means green house’s electricity total cost should be less than non green house. (By green house’s definition, it should also be like that), but the same conclusion can not give to gas cost, which is shown as the graph below:



Because around 27 rent we cannot say if the rent is raise by gas or decrease by gas. So Our group aim to calculate its means by different group, the mean of gas cost in green house grouse is 0.0108 and in non-green house is 0.011. So green house’s cost is a bit cheaper.

From the data above, our group conclusion is if investors want to invest, definitely the greenhouse is better than the non green house. Because only 3.06 years is needed to recuperate the cost difference, and I think investor definitely will not choose less than 3.06 years project, because if the project is less than 14.53 years for green house, investor will make a loss. And also, according to gas and electricity calculation, green house has a cheaper costs in these two area. So definitely greenhouse is the better choice. And if the project is more than 14.53 years, then green house can make profit and more economic than non green house. If the project is less than 14.53 years, investors will make a loss.