April 1st – April 7th

Mathieu and I looked into a few datasets, we searched from Kaggle, corgis and a few other datasets.

We analyzed if the data sets are large enough, whether it contains significant number of rows and columns of data. And if it is cleaned enough so that we can work on it directly without worrying too much about the null value and missing data. We tried to find data sets that is both as interesting and cool to work with. Finally, we decided to take the dataset on Kaggle and use Yelp API to get data and do relation between the length rating, location and inspection score etc.

We get combined json file from

I tried python and R to parse json file, but finally settled on use javascript to parse json file

April 8th – April 14th

April 15th – April 21st

April 22nd – April 28th

27 post codes in sf

April 29th – May 5th

Total ~37,000 data with score

2015 April, 2018 April

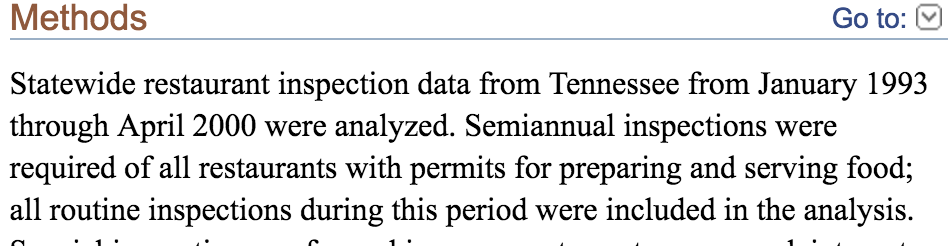
My csv is ~3,800

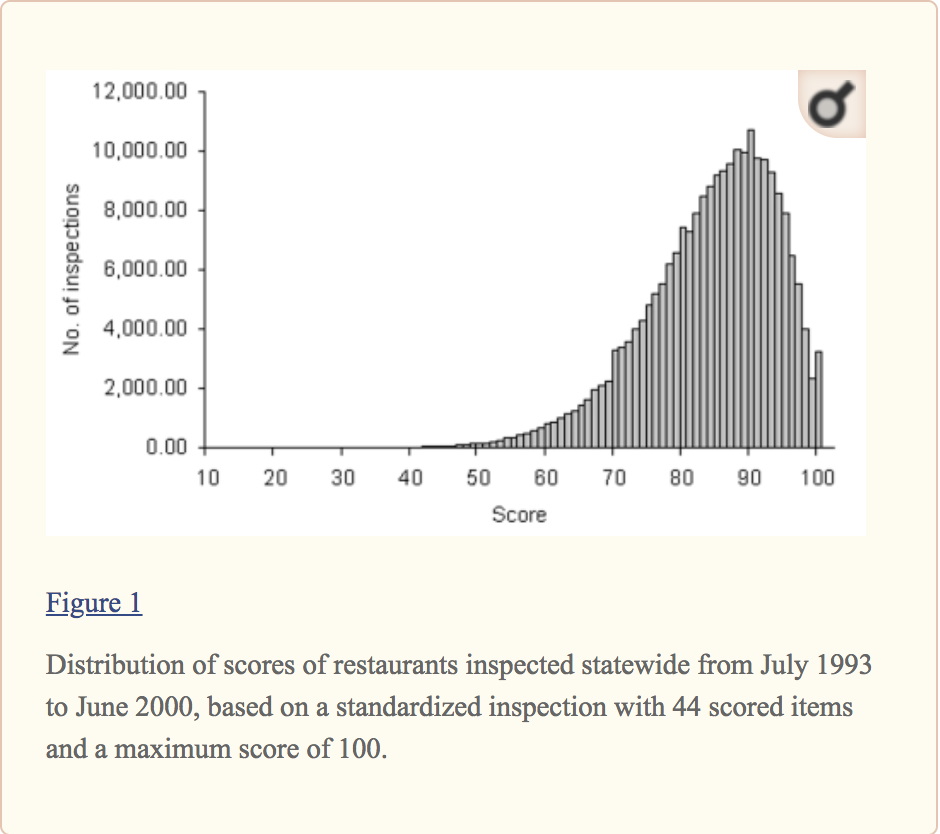
Originally, I suspect the score is normally distributed.

Because I read the GRE and SAT are normally distributed.

But then I read this, so I think this is more gamma or weibull

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3323064/

They took the mean of all score of the latest round all and take out non-toxic ones



https://www.frontiersin.org/articles/10.3389/fmicb.2016.00614/full

https://homes.cs.washington.edu/~yejin/Papers/emnlp13\_hygiene.pdf

The goal of this project is to visualize the restaurant’s find the interesting patterns and relation

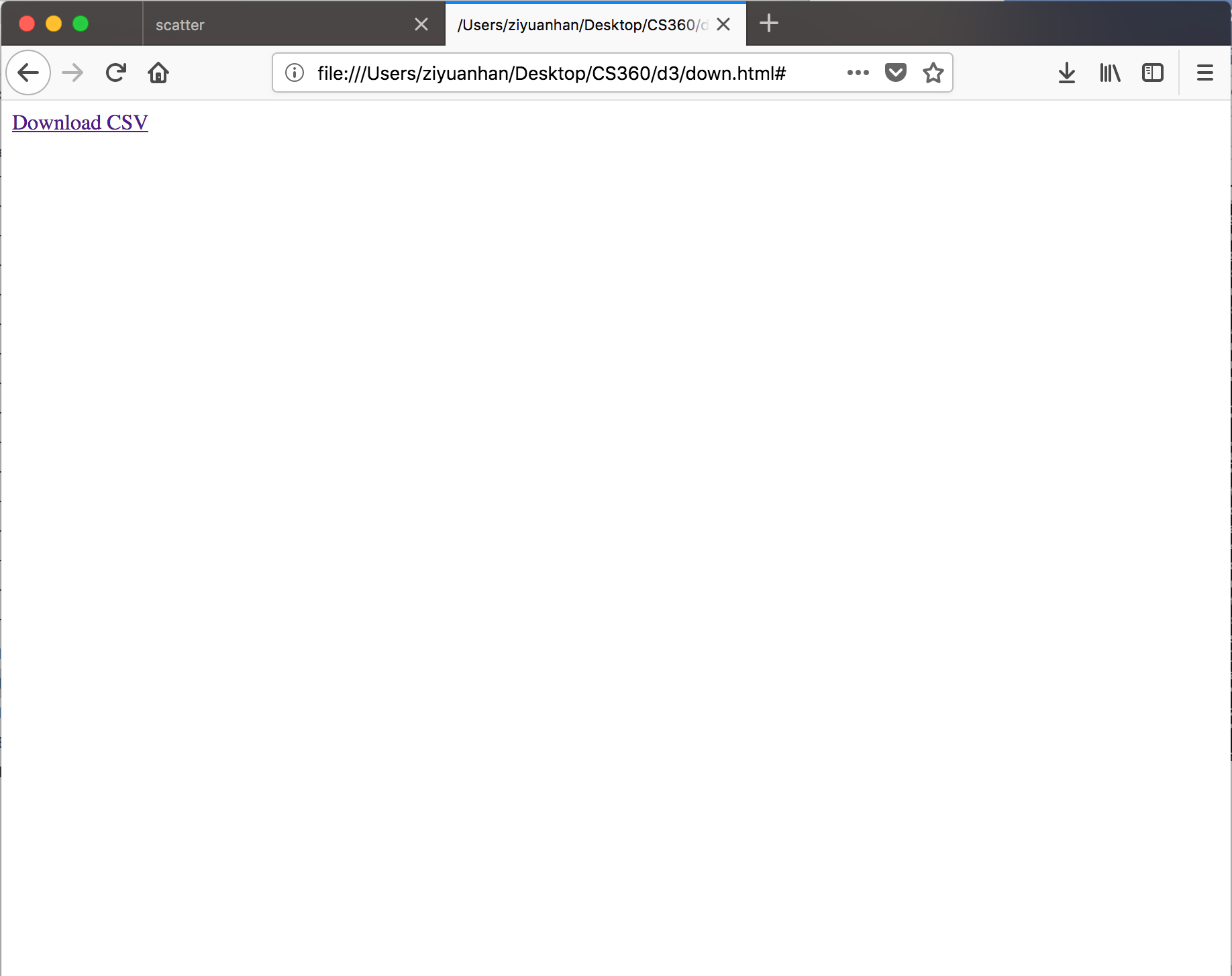
I am in charge of the scatter plot and boxplot, distribution plot and tree plot.

The hardest problem for me is not about the visualization but the process of data

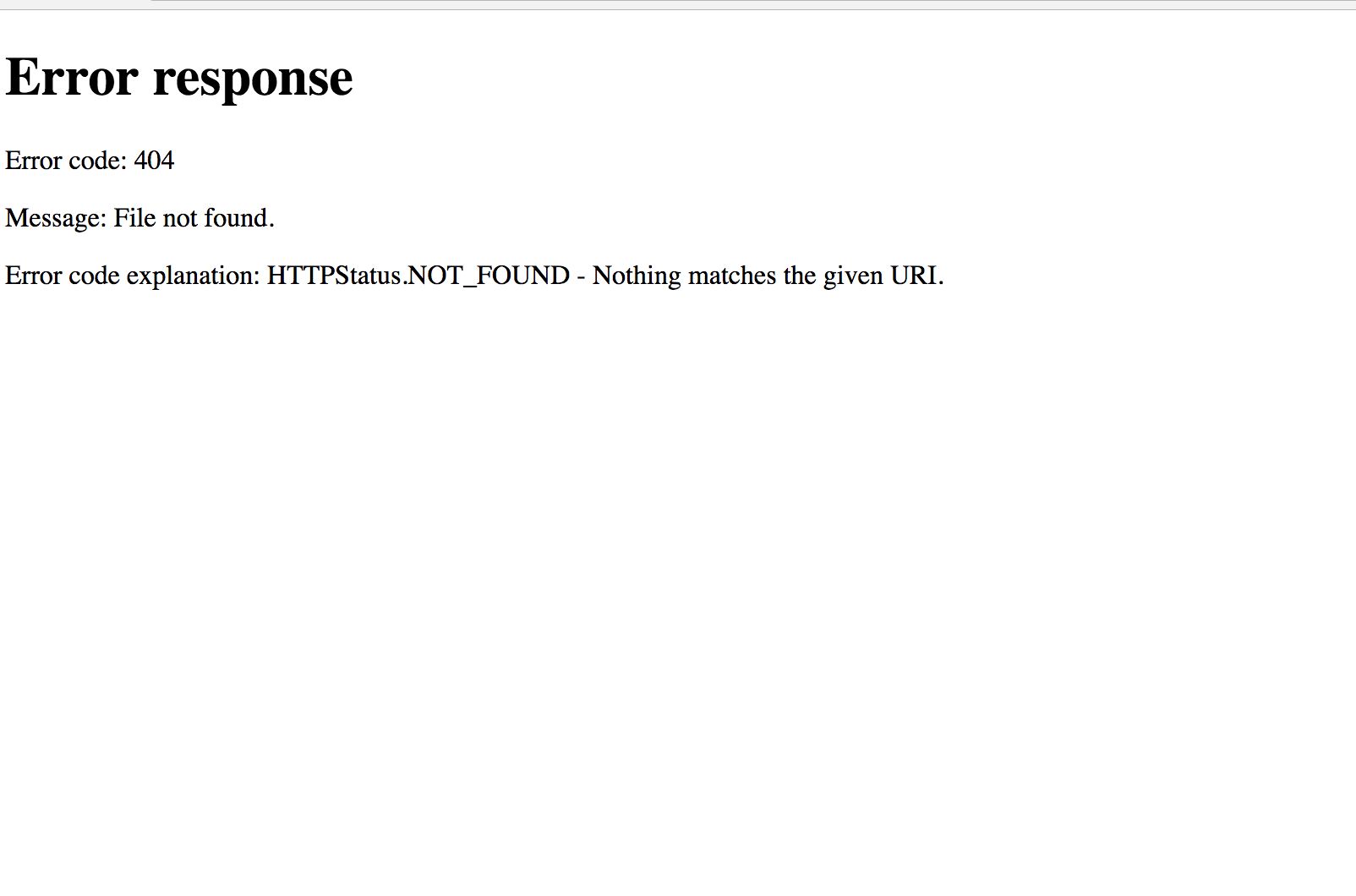
I don’t have much previous experience with server or JQuery or json so I have to re learn everything with server and stuff

The json file we got from the website is like this

The file is not very organized and it does it a lot of time just to read this stuff.



So I managed to find the start of the string and the end of the string.



I found some help from the internet where it teaches me how to download data

But I need to run local python http server in order to obtain use the java script code

The data is not ready yet but I have done some simulated visualization on the data.







