CS5346 Assignment 4 – NEAViz

Student name : Bastian Morath

Matric number : A0195628N

Introduction

The objective of this assignment was to create visualizations about a dataset created by the *National Environment Agency* (NEA), which lists licensed eating establishments in Singapore in 2015, together with metrics such as a health grade, demerit points and suspension history. The visualizations would help to answer analytical questions about this subject.

Supplementary datasets

I used one supplementary dataset, which showed the population of each planning area in Singapore. I got the dataset from the *Department of Statistics Singapore* (https://www.singstat.gov.sg/find-data/search-by-theme/population/geographic-distribution/latest-data -> Basic Demographics Characteristics dataset), where I extracted population data for each planning area only. I then combined this dataset with the NEA dataset by doing an inner join on the *planning area* field.

I chose this dataset since I wanted to compare not only the number of establishments per planning area, but also the *density* of it.

Pre-processing

Apart from the before mentioned pre-processing of the population-dataset, I also changed all 'na' values of the NEA dataset to NULL values, as Tableau handles them better. I also created some new table calculations such as the *density* (Number of eating establishments per 1000 residents), custom color schemes and booleans whether an establishment has already been suspended before.

Analytic questions

1. *Is there a geographic pattern where some regions of Singapore have more Eating Establishments than others? (Figure [1])*

   One can see that in the South and east of Singapore there are more eating establishments than in the North and West. Especially Bukit Merah, Downtown Core, Geyland and Bedok have a lot of eating establishments.

2. *If we also compare the number of establishments to the number of residents in a particular planning area, does the result change? Is there a correlation between the planning area population and the number of eating establishments? (Figure [2.1] and [2.2])*

   The density is more evenly distributed. While the easy still has a relatively high density of establishments, also the west of Singapore has. Note that planning areas with fewer than 200 residents were omitted, such that the results and the colors wouldn't be skewed, since planning areas with only a few residents naturally have a higher density of establishments.

To analyze this question in more detail, I also created a correlation plot between the population of an area and the number of eating establishments. There is a clear tendency (that is correlation) between them. This means that usually a higher population also means more eating establishments, which naturally makes sense since more people have to get fed.

3. *Is there a geographic pattern where some regions of Singapore on average have better grades than others? (Figure [3])*

   While the center and east usually have better grades, the west of Singapore has, on average, worse grades. An explanation could be that in the center, where a lot of people live, the competition among food establishments is much higher and thus the health standards need to be better. In less populated areas, the competition is much smaller, and thus establishments are able to afford having a worse rating.

4. *How are the number of demerit points distributed among the licensees? (Figure [4])*

   One can quickly see that most establishments either have 4 or 6 demerits points. Only a very small amount has 0 or more than 6 demerit points. Also note that in the dataset, there are a lot of licensees that do not have a demerit number. Those are not visualized.

5. *How many suspensions were given by each month in the year 2015? (Figure [5])*

   It's interesting that at the beginning of the year, there are much fewer suspensions given than in the middle of the year. Also interesting is October. It has by far the least suspensions given.

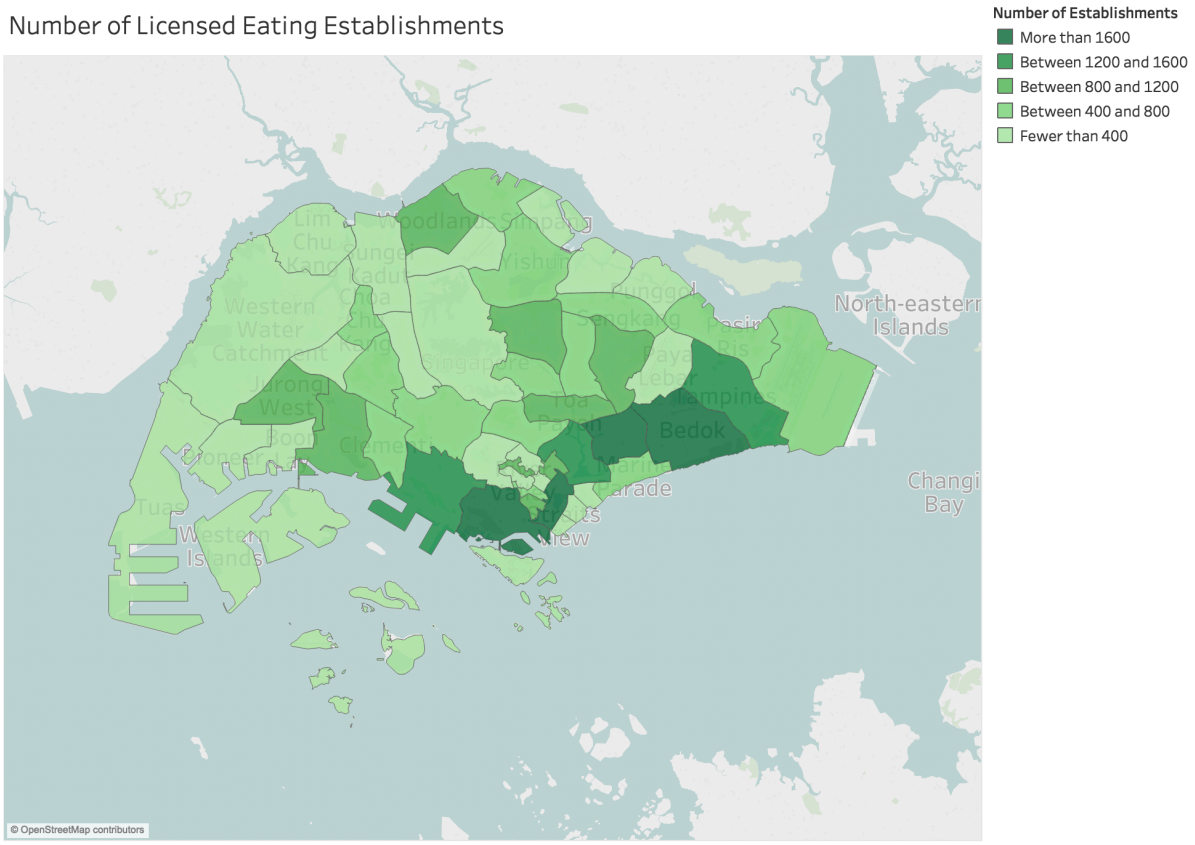
Visual encodings

I mostly used a *map chart*, since it neatly shows different statistics about each planning area. I also used *packed bubbles*, a *line chart* for temporal data, and a *scatter plot* for showing a correlation (in which I also fitted a trend line).

My Visualizations

[1]

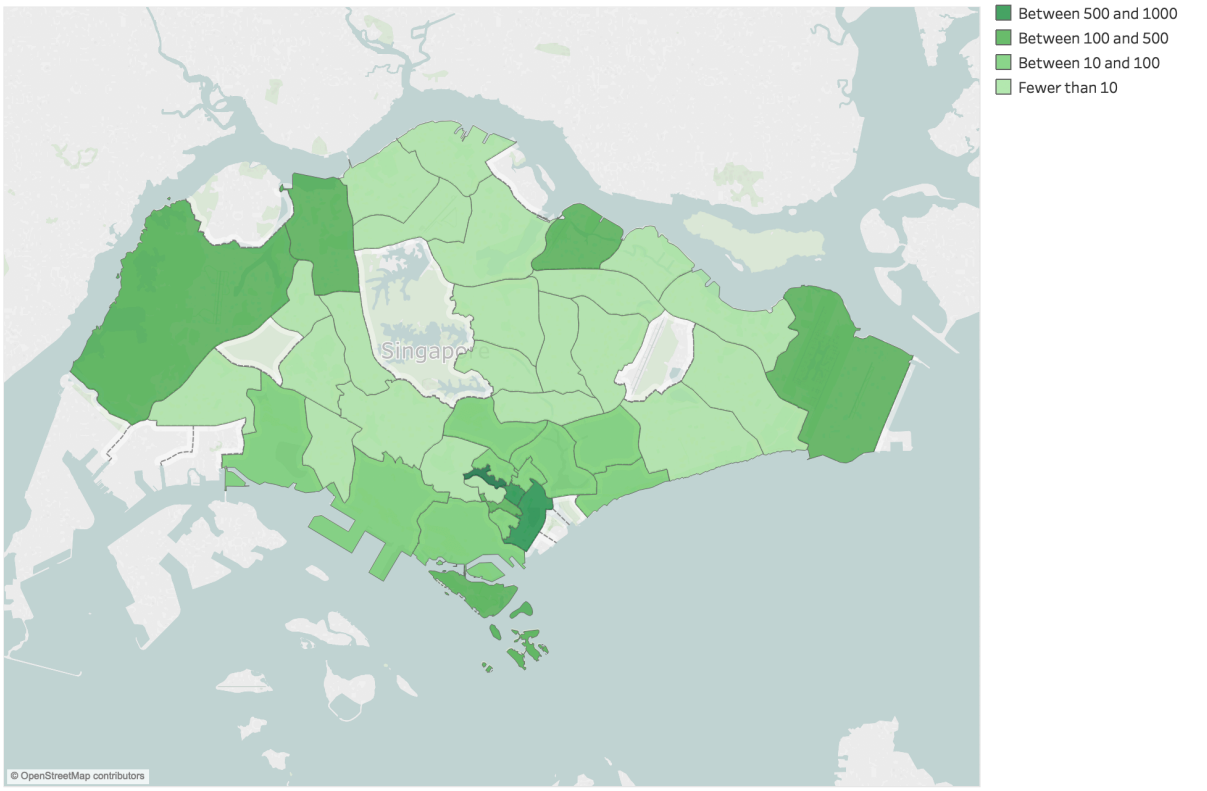
Number of Licensed Eating Establishments



You can see the number of NEA licensed eating establishments per planning area.

[2.1]

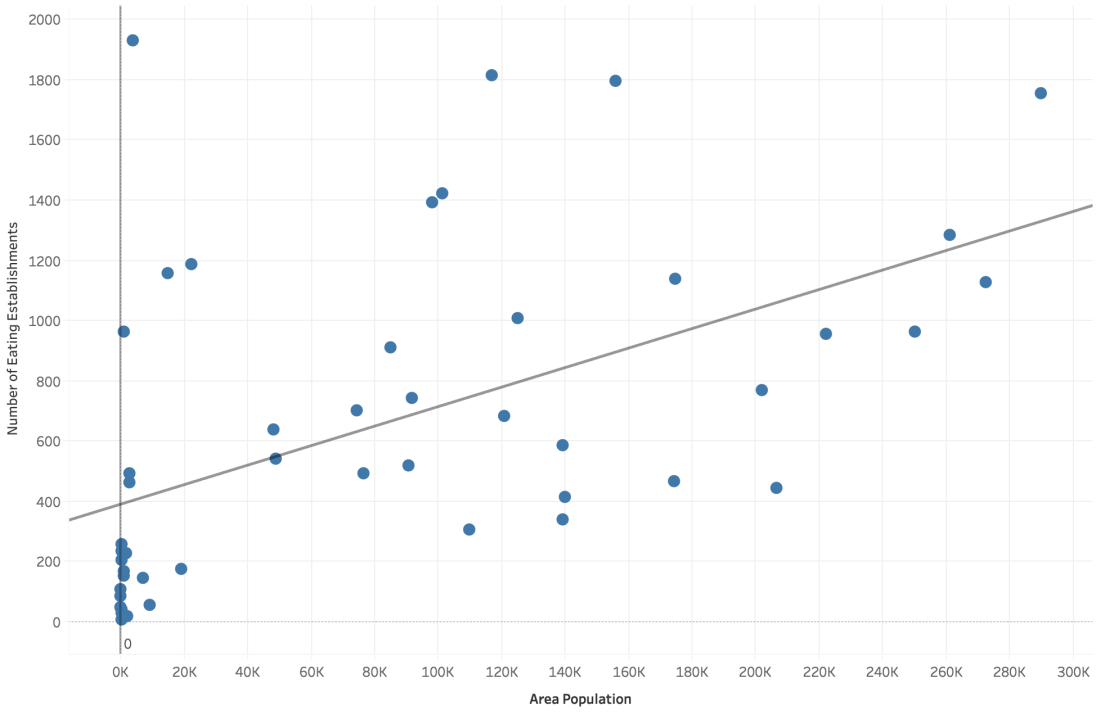
Density of Licensed Eating Establishments per Population



You can see the density of NEA licensed eating establishments per planning area. Areas with fewer than 200 residents were omitted.

[2.2]

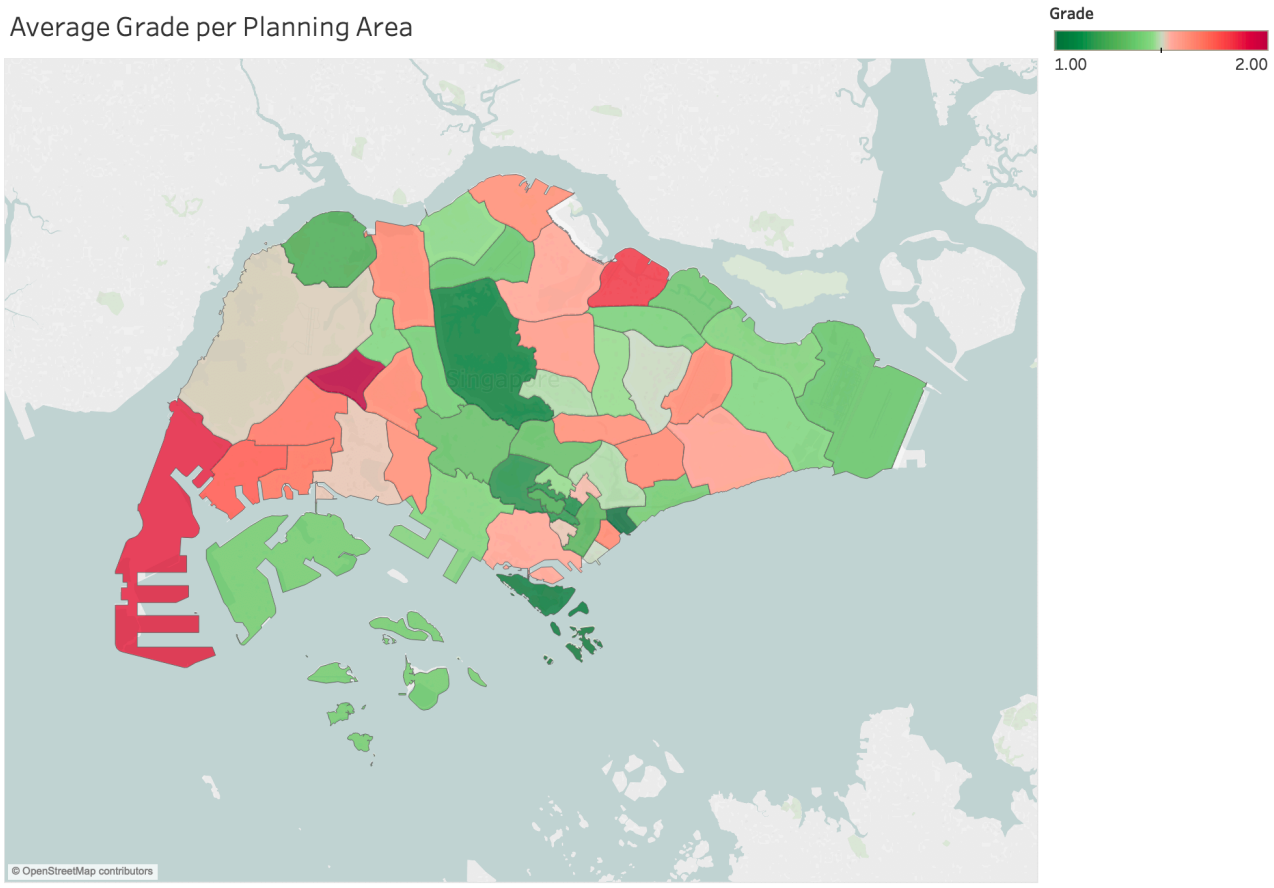
Correlation Population and Number of Establishments



You can see the correlation between the planning area population and the number of licensed eating establishments.

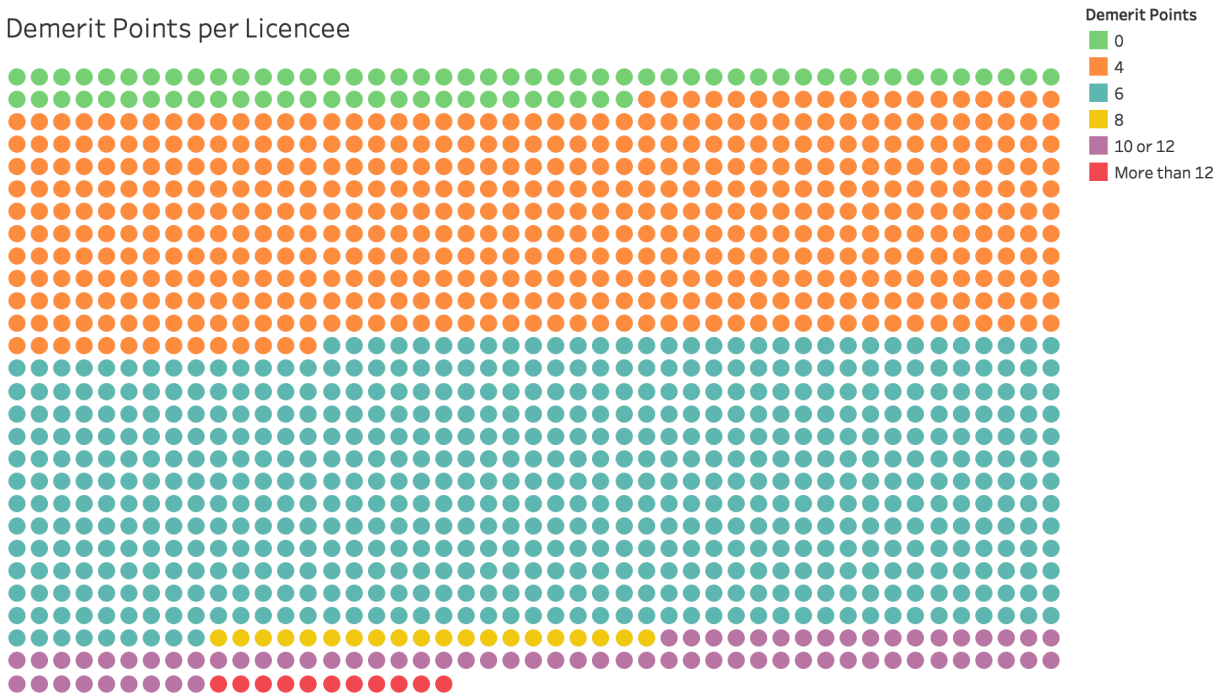
[3]

Average Grade per Planning Area



You can see the average grade (From 1 to 3, 1 being the best) of a particular planning area, given by the NEA.

[4]



You can see the number of demerit points, where each dot represents one particular establishment.

[5]

You can see the number of establishments that were suspended for each month of the year 2015.