

Final Project Hand In

By: Jana Osea

ID: 30016679

Dr. Emma Towlson MW 2-3:15pm

1. Introduction

This project focusses on visualizing all the fire incidents that occurred in Toronto, Ontario from 2011 to 2019. The goal of the visualization is to better understand the effect of certain characteristics on the various target values when a fire happens. The visualization has three main parts that tell the story about the effect of different ignition sources, time of the year, and/or geographical location on various values like number of fire incidents, estimated dollar loss, and civilian casualties.

In this report, we will go through a description of the data, my design process throughout my previous hand ins and final implementation. Overall, my takeaway from this project is that visualization requires the producer to think from the perspective of the viewer and that visualization is a powerful tool that can help provide important insights to various problems.

2. Data description

2.1. Data descriptions

The final data set used for this data set is obtained from the Toronto Open Data [1] website. The original dataset contains ~50 columns and ~79000 rows. The data is very extensive and contains a lot of information. The various columns contain information like when the fire occurred, where it occurred, what caused it, which station responded, etc.

2.2. Pros and cons of data sets

As mentioned above, the data contains so much information as it contains ~50 columns each holding a characteristic about the fire occurrence. In some way this is a strength because I had so much freedom as to which direction I want to take. I can combine different variables to visualize where each combination tells a different perspective. However, in another way this could also be a weakness because it can be overwhelming to know which direction I should go.

2.3. Data set decision

In the end, I decided to use 5 original columns, 1 extracted column and all the rows. The 6 columns consist of

- Estimated dollar loss (original)
- Number of civilian casualties (original)
- Number of fire incidents (extracted – aggregated number of fire records grouped by months)
- Ignition Source (original)
- Incident Date (original)
- Longitude/Latitude (original)

I decided to use this for my final visualization because it provided a good overview of the fire situation as it provided historical and geographical data along with various continuous variables to examine. In addition it also added the categorical variable of ignition source.

3. Design process

3.1. Sketch-able data subsets

Subset 1

Estimated_Dollar_Loss	Ignition_Source	Alarm_Date
200000.00	heater	January 6, 2017
20000.00	appliance	January 26, 2017
20000.00	appliance	February 5, 2017
350000.00	appliance	February 17, 2017

Subset 2

Latitude	Longitude	Clear_Total_Minutes	Ignition_Source
43.68228	-79.33272	4	smoker's article
43.6701	-79.41318	122	smoker's article
43.74283	-79.50161	350	electronic wiring
43.63126	-79.47554	1320	smoker's article

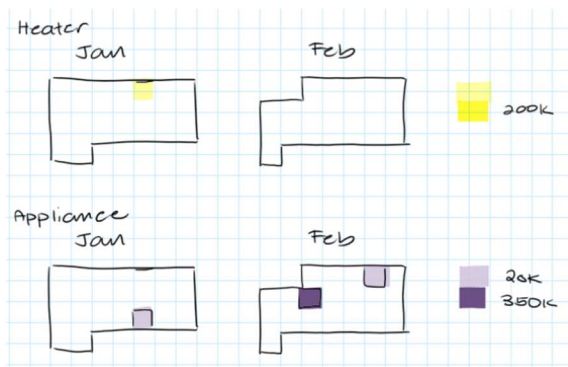
These 2 subsets were chosen because I wanted to explore the effects of ignition source on different values like estimated dollar loss and total clear time. In hand in, I had originally wanted to use total clear time as one of my potential columns to explore, however during my final decision phase, I decided to use other values instead as I found it more interesting to examine the other previously stated values.

3.2. Design direction

Note: The full sketches can be found in the appendix

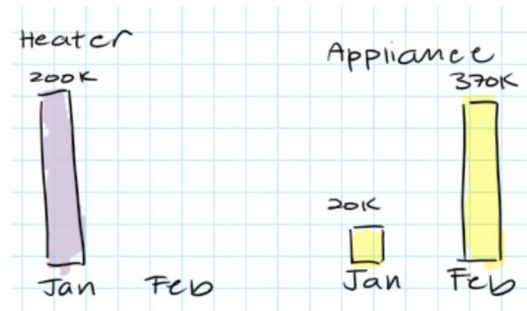
3.2.1. First sketches

Subset 1, Sketch 1



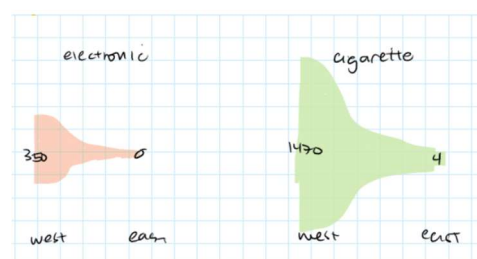
Focusses on historical data

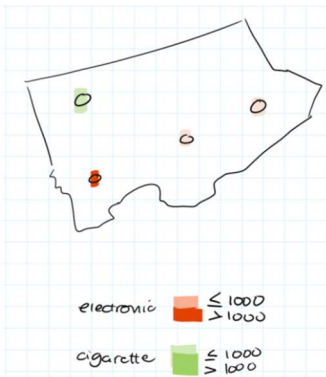
Subset 1, Sketch 2



Subset 2, Sketch 1

Subset 2, Sketch 2

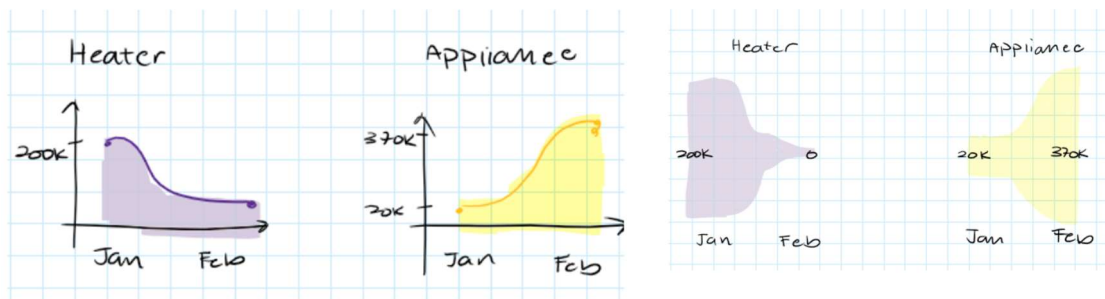




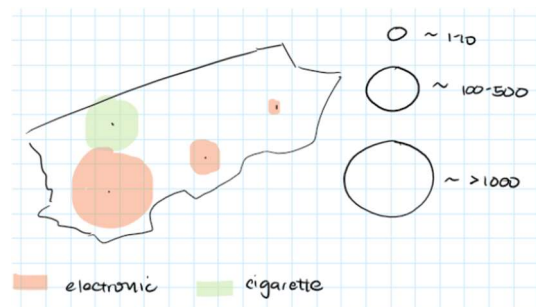
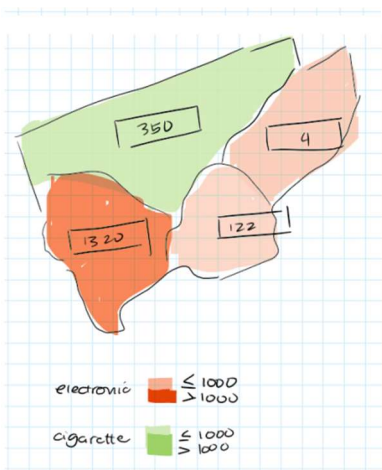
Focusses on geographical data

3.2.2. Variations

The 2 sketches below focus on Subset 1, Sketch 2. The focus was to put the date values in the x-axis and the estimated dollar cost on the y-axis grouped by ignition source.



The 2 sketches below focus on Subset 2, Sketch 1. The focus was to put the fire incidents on a map and use color or shape to visualize the total clear time.



3.3. Process

There are a lot of visualizations that can be produced from these 2 subsets even though there were only 3 main variables used in both sets of data. What came out of the first subset were 10 sketches (that focussed on the effect of date and ignition source to the estimated cost of a fire. The of one sketch focussed on the idea of first separating the ignition source and then separating the date within each ignition source and seeing the effect of that on the estimated cost. As for the second subset, the sketches focussed on the effect of location and ignition source to the total time it takes to clear a fire (in minutes). The sketch that was chosen to be explored for variations was one that had the actual map that portrays the location and not just an aggregate of where the fire was in west or east Toronto.

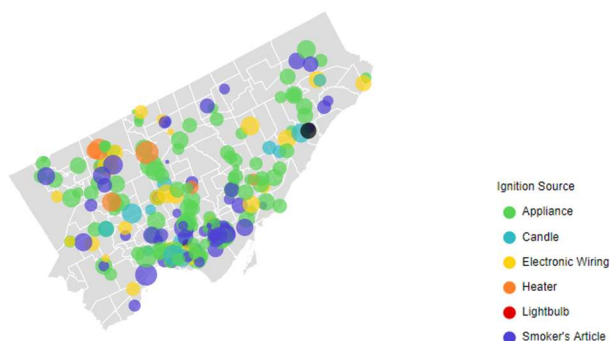
Again, to reiterate I did not use total clear time in my final visualization, but the ideas generated while visualizing the various subsets helped me come up with ideas for the other values instead.

3.4. General design direction

The general design direction was that I want to use a map. I wanted to initially see what the entire data would look like in a map.

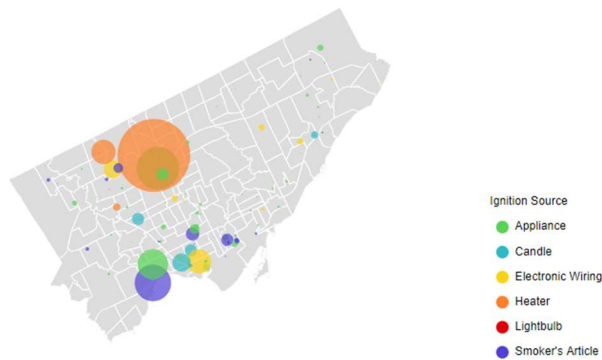
3.5. Prototyping variations

3.5.1. Variation A



For this visualization, I decided to perform the Math $\log(\text{Estimated Dollar Loss})$ and plotted it along all the different locations. Although this drastically helped fit all the data into the map, I find that it does not effectively show the variation in dollar cost.

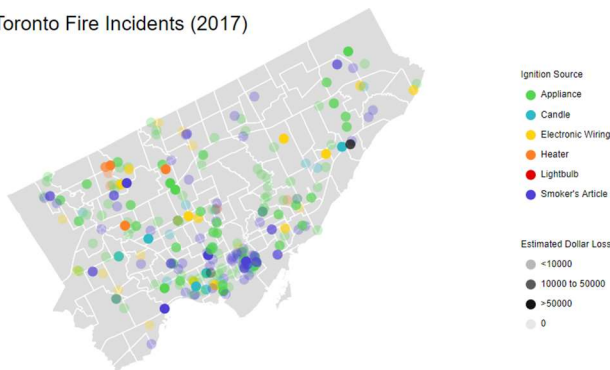
3.5.2. Variation B



This visualization did not have any mathematical manipulations towards the estimated dollar loss. I just fitted the losses according to their real values and scaled it to fit the visualization. I find that although the differences in estimated loss is evident, the ignition sources are not clearly differentiated due to some points being too small.

3.5.3. Variation C

Toronto Fire Incidents (2017)



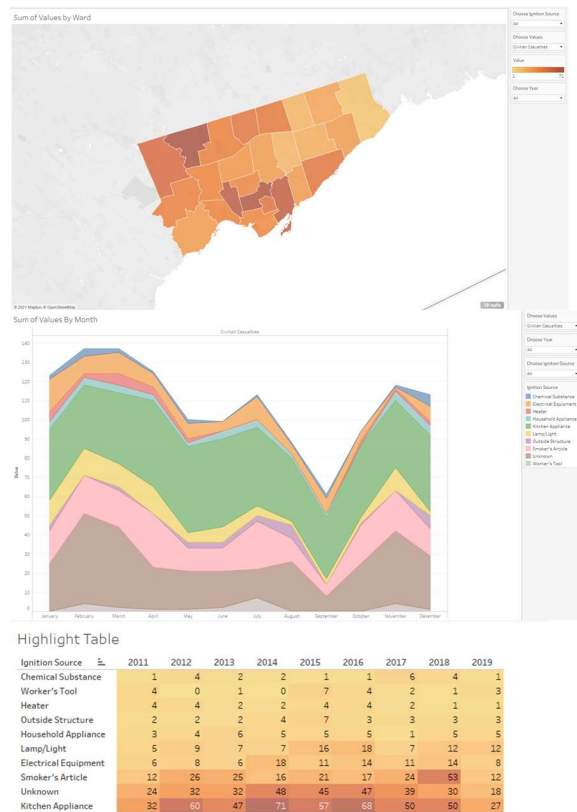
Finally, I believe that my final visualization would appear somewhat like this. The different ignition sources are still quite evident due because the size of the points are a good and visible size. The estimated dollar loss can be differentiated because of the different opacities. However, in the future, I will try to find a better way to categorize the dollar loss so that the opacity is very evidents.

3.6. Implementation process

I found the process of coding the alternatives very helpful in exploring different ways to visualize the data. I found myself using the code to give me ideas as to what area I should manipulate in order to effectively differentiate my data. For example, I was playing around with the different `attr()` in the code by using different sizes, colors, or scales. One thing I found difficult was that the geojson file that I found to create the borders of Toronto was not the same as my dataset. It would be so helpful I could group my geojson borders with my longitude/latitude in order to create a choropleth. However, because it wasn't it was difficult to implement in this hand in.

3.7. Final static design

My final static design consists of 3 different visualizations as follows:



In a static view, this visualization focusses on the total number of civilian casualties from 2011-2019 by ward in Toronto.

In a static view, the height at each month represents the total number of civilian casualties from 2011-2019 grouped by dates and ignition source.

In a static view, the color shade is the total number of civilian casualties from 2011-2019 grouped by year and ignition source.

3.8. Prototyping interaction

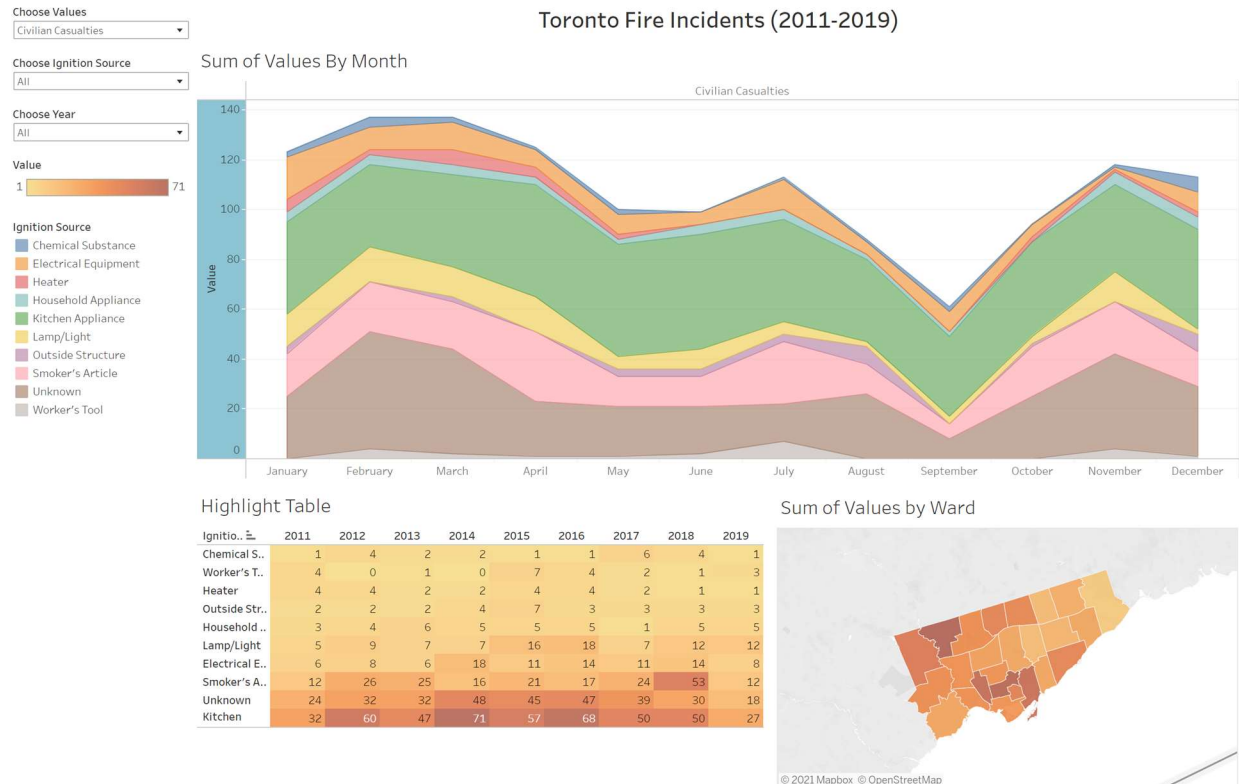
For interaction, I wanted to be able to change the value being examined from not just the total number of civilian casualties but also to the number of fire incidents and the estimated dollar loss. As well, I want to be able to filter out the years and the ignition source. Hence, there are three interactions that I included:

- (1) Change the value to view: Be able to change using a drop down to change the value to view (Estimated Dollar Loss, Civilian Casualties, or Number of fire incidents).
- (2) Filter data by year: Using a drop down to change the year (2011-2019)
- (3) Filter data by ignition source: Using a drop down to change the ignition source.

4. Final implemented visualization

I included all the visualizations into one dashboard with the interactions all on the left side. In addition to the interactions, a summary window shows up every time a mouse hovers over a geographical area that shows the area name.

Here is a screenshot of the visualization

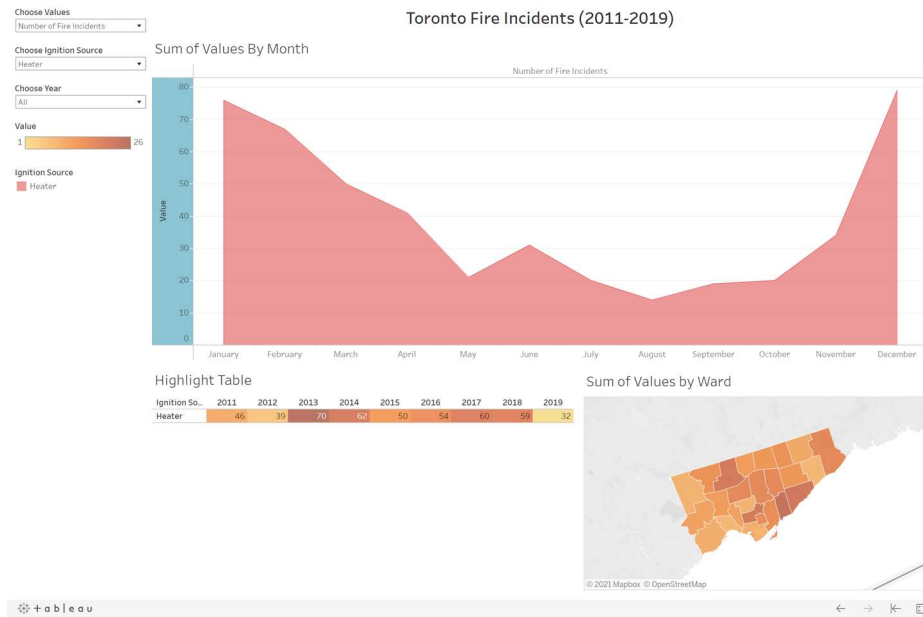


The visualization can be viewed with the interactions on this page:

https://public.tableau.com/profile/jana.osea#!/vizhome/Project_16183800368420/Visualization

4.1. Process reflections

This process of creating this visualization first showed me that Tableau is actually very fun and powerful. This process also helped me realize that there are so many benefits to visualizing data. It can really help create a intuitive understanding of the data. Like for example, we can see that there are more number of fire incidents caused by heater sources as shown below:



5. Discussion

As I have mentioned, I really like my visualization because it conveys a lot of information. I also like the intuitive understanding that it gives. I feel like the visuals that I provided are good. The chart, the highlight table and the choropleth are quite sufficient at giving a good overview of the fire incident situation in Toronto. I do think I could've made the overall visual more "aesthetic" by changing the background color to be somewhat related to fire. However, I wanted the visual to be simple and straight to the point.

6. Conclusion

To conclude, I really like Tableau as it provides a simple, fast, and effective way to create dashboards with easy interactions. This project also helped me realize that visualizing requires a lot of consideration about the viewer and what kind of insight you can make using the visualization. I believe that a good visualization requires a lot of thought from the producer about how the visual is seen by the viewer.

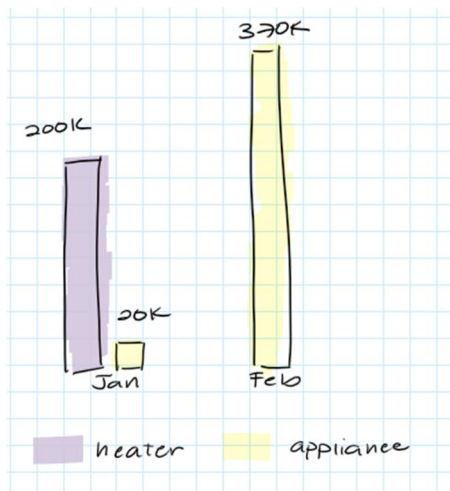
7. References

[1] <https://www.toronto.ca/city-government/data-research-maps/open-data/>

8. Appendix

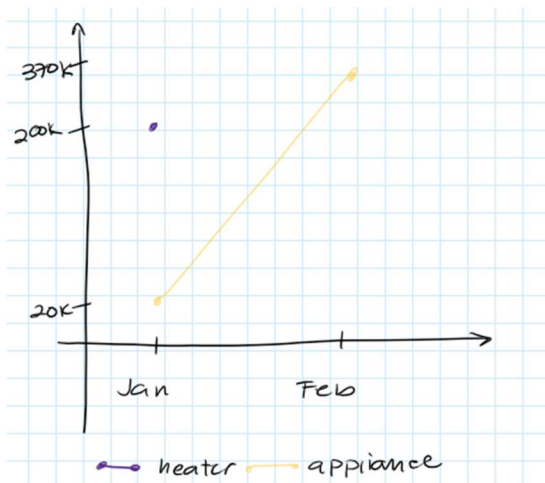
Subset 1: 10 Different Sketches

1.1



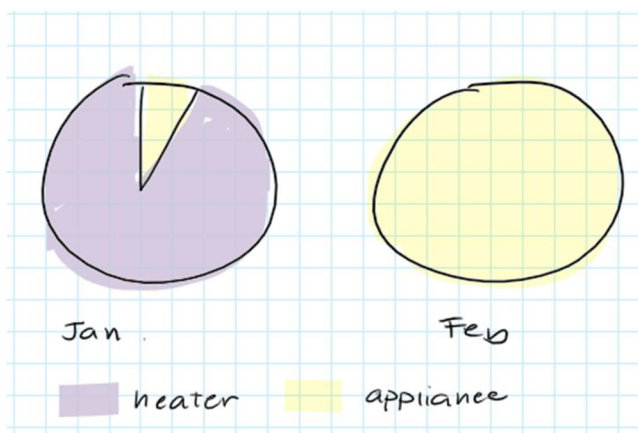
Cost grouped by date and ignition source

1.2



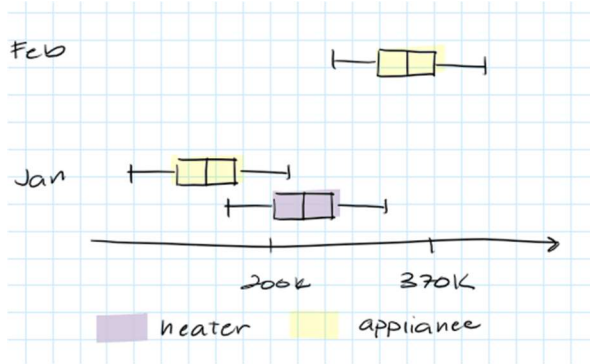
Line graph of cost separated by ignition source

1.3



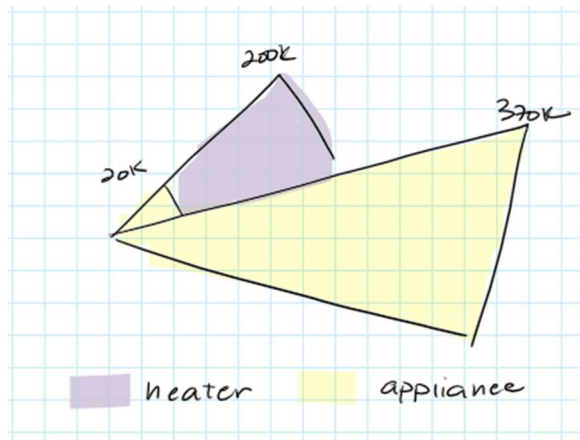
Pie chart of cost percent separated by date

1.4



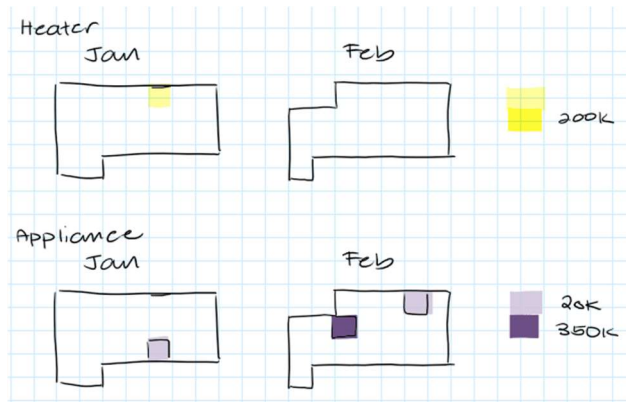
Box and whisker plots of cost separated by dates and grouped by ignition source

1.5



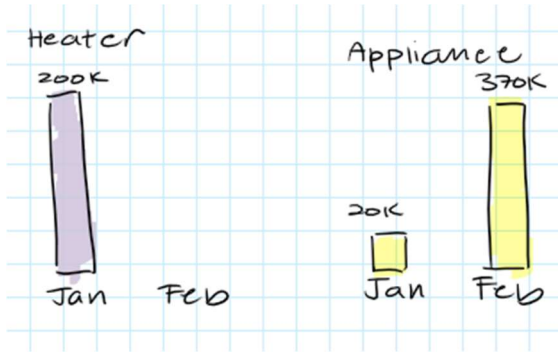
Length represents total cost for each month and the color is the grouping according to ignition source

1.6



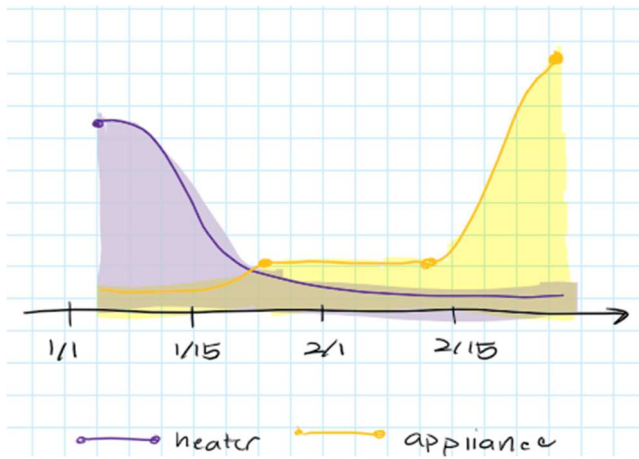
More opaque represents more costly fire incident grouped by ignition source.

1.7



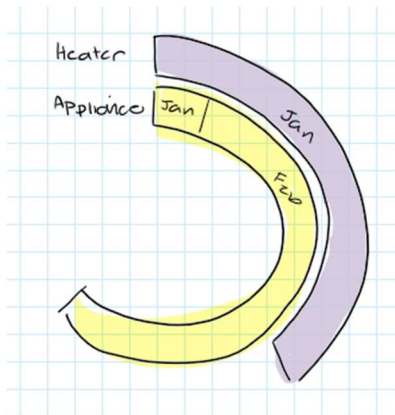
Height represents cost grouped by ignition source and separated by date (please excuse the fact that the 200k is same height as 370k ><)

1.8



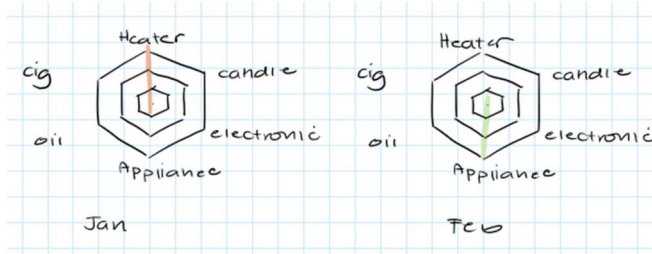
Density of cost

1.9



Length is cost

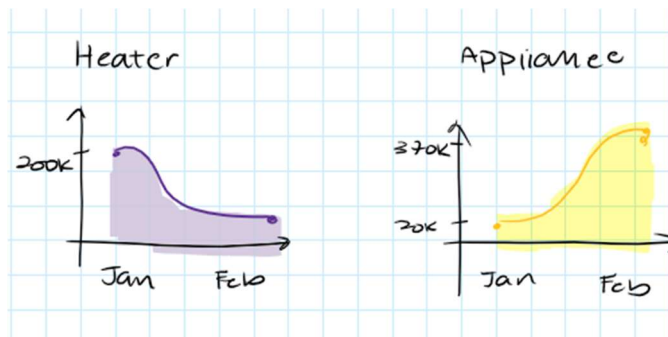
1.10



How close the bar is is the cost

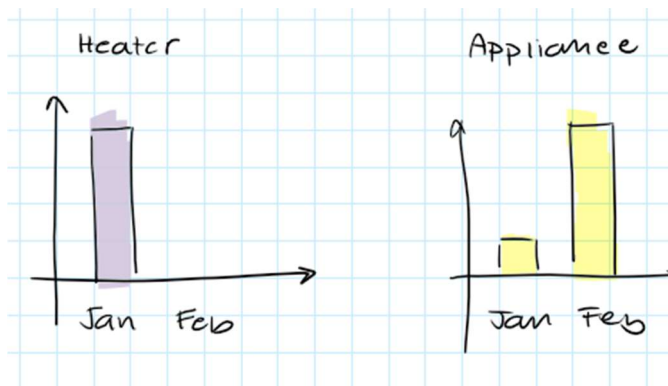
Subset 1: 10 Variations of One Concept

1.7.1



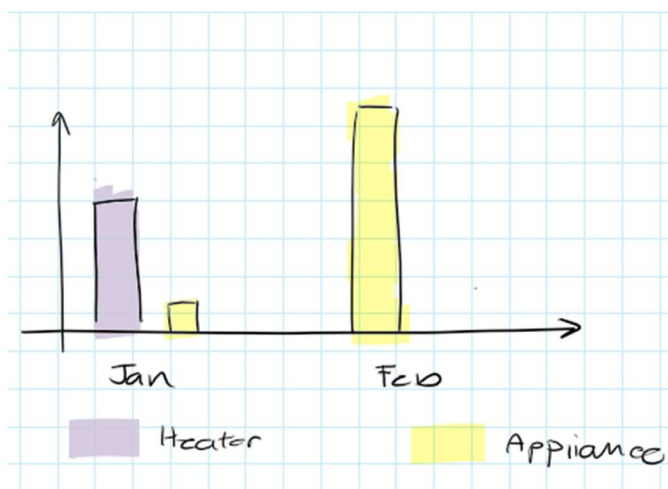
The density is the cost

1.7.2



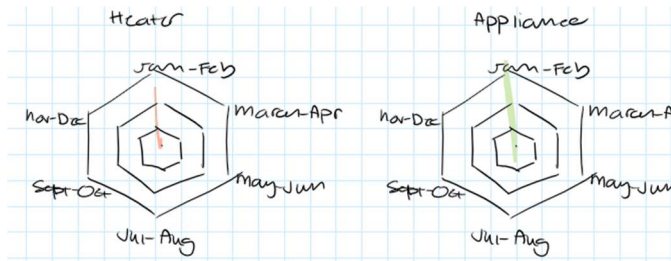
Height is the cost

1.7.3



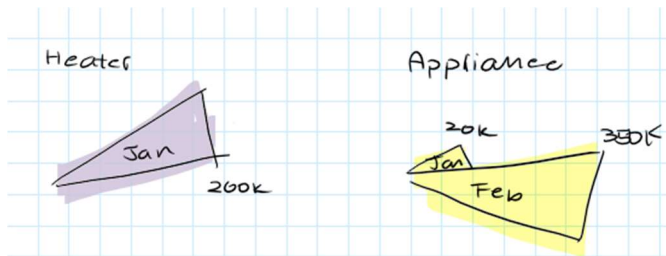
Height is the cost and color is the ignition source

1.7.4



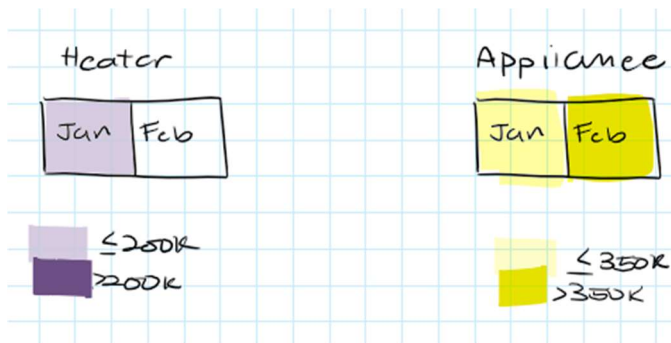
How close the line is to the cost and each point is the ignition source

1.7.5



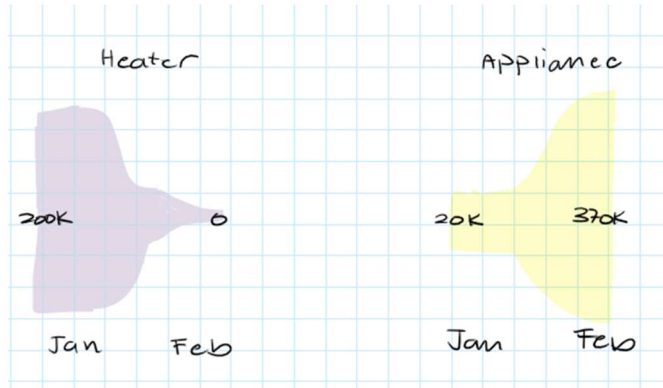
Length is the cost and each "leaf" is the month

1.7.6



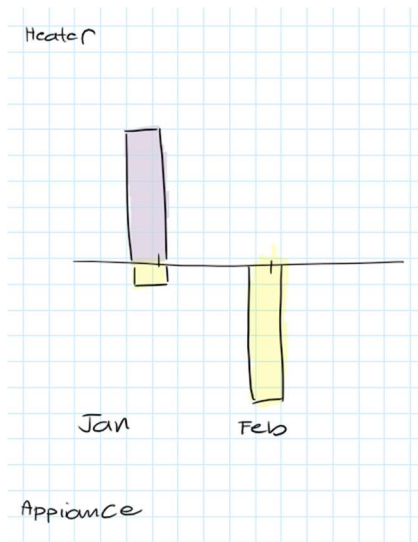
Each box is a month and the opacity is the cost. More opaque the more expensive.

1.7.7



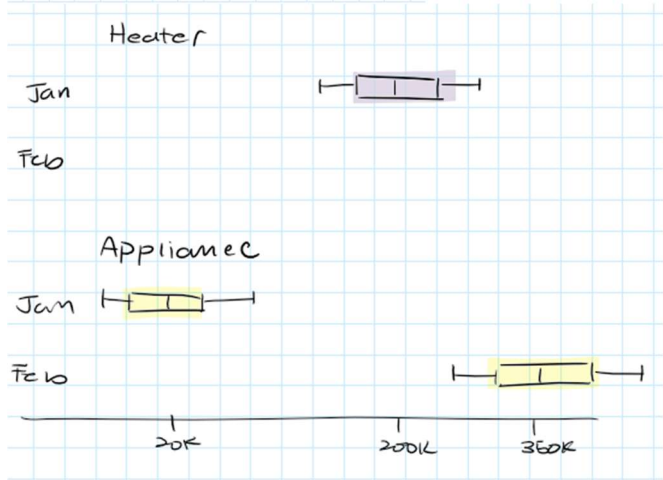
Width is the cost and each occurrence of a new wave (up or down) is a month

1.7.8



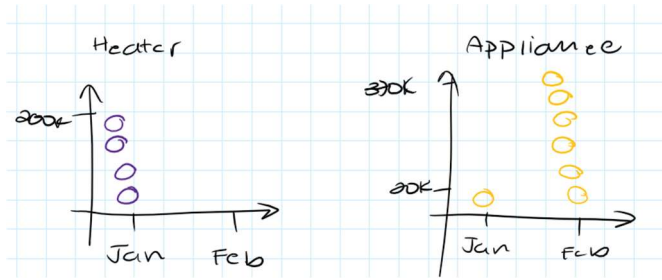
Height is the cost.

1.7.9



Center of each box and whisker is the median and color is the ignition source.

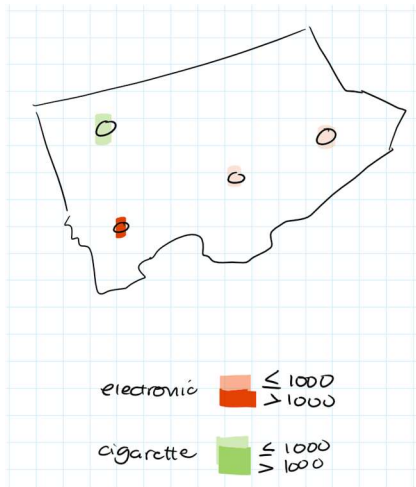
1.7.10



Each circle is an addition to the cost

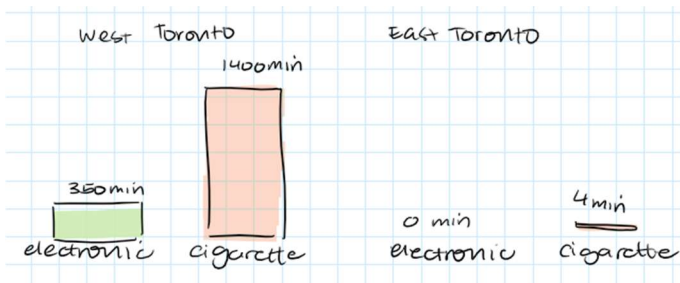
Subset 2: 10 Different Sketches

2.1



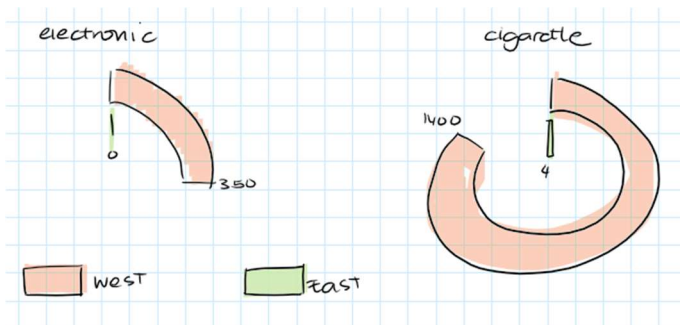
The more opaque, the more total clear time

2.2



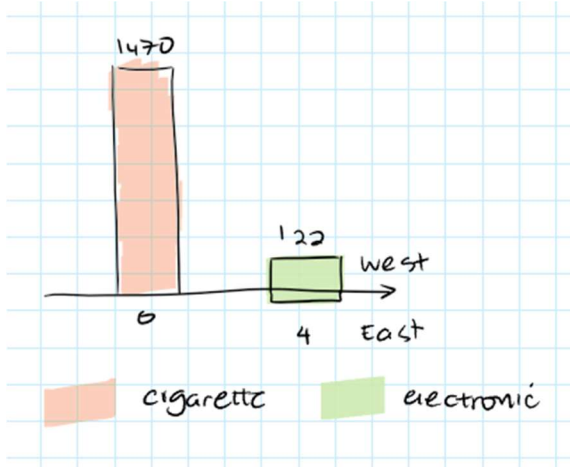
Height is the cost, color is the ignition source and each set of 2 is either west or east Toronto

2.3



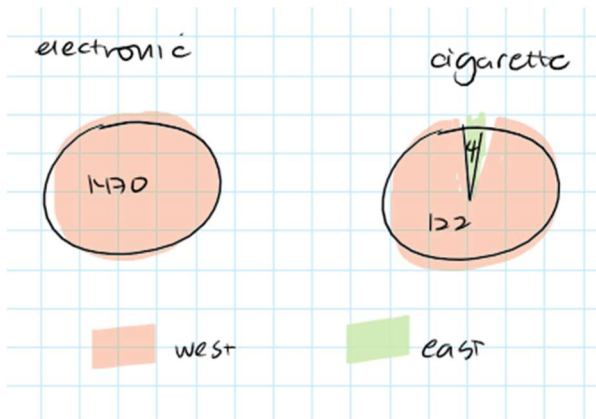
Length is the total clear time and color is the section of the city.

2.4



Height is the total clear time and color is the ignition source. All the bars above x axis is for west and below is for east.

2.5



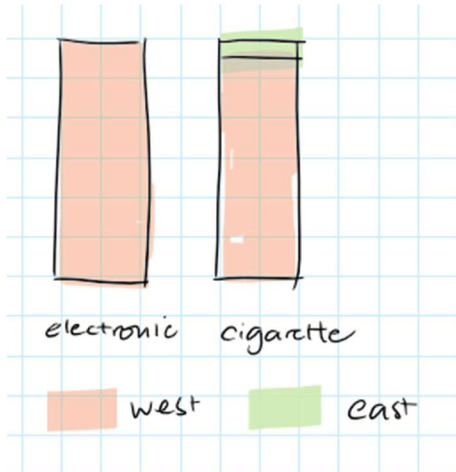
Clear time percentage is the size of the pie slice

2.6



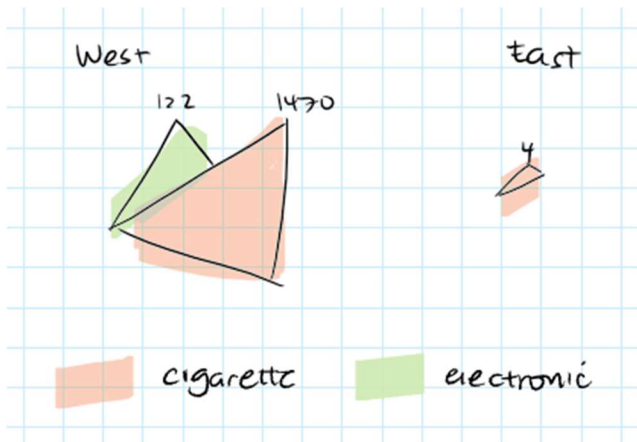
How far out is the total clear time

2.7



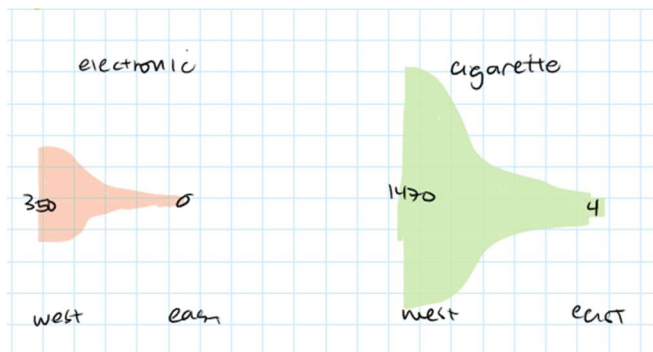
The proportion of total clear time

2.8



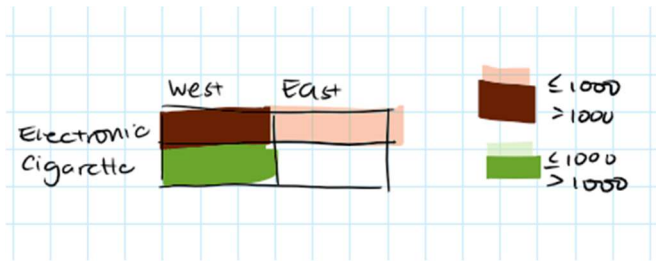
The length of each "leaf" is the total clear time

2.9



Each new wave is the total clear time of west and east separated by ignition source

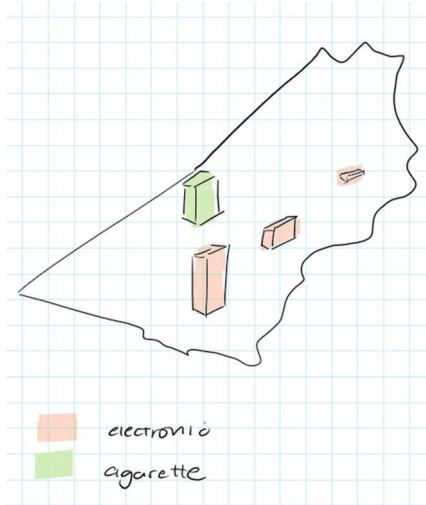
2.10



Heat map. More opaque, more total clear time

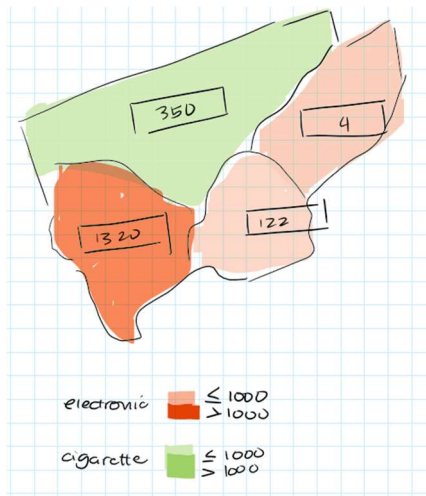
Subset 2: 10 Different Variations of One Concept

2.1.1



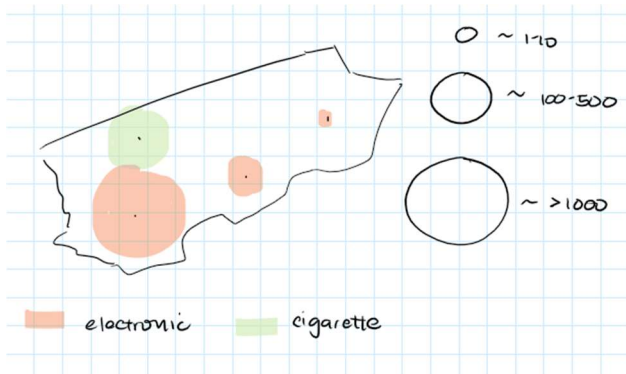
Height of the bars represent the total clear time

2.1.2



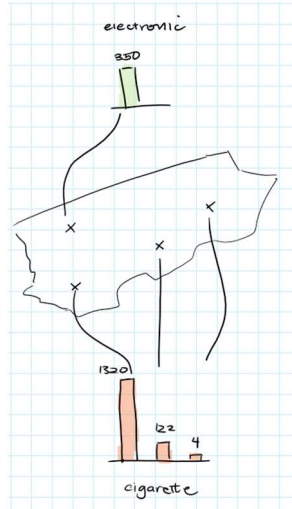
More opaque, more total clear time

2.1.3



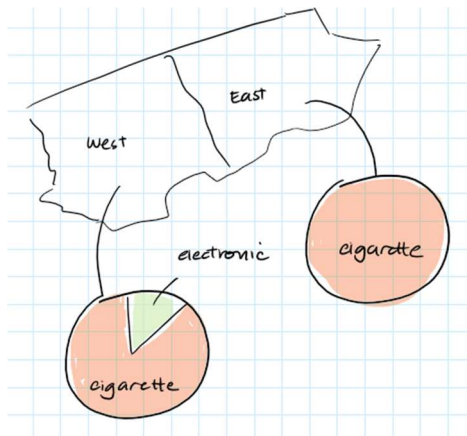
Bigger circle size, more total clear time

2.1.4



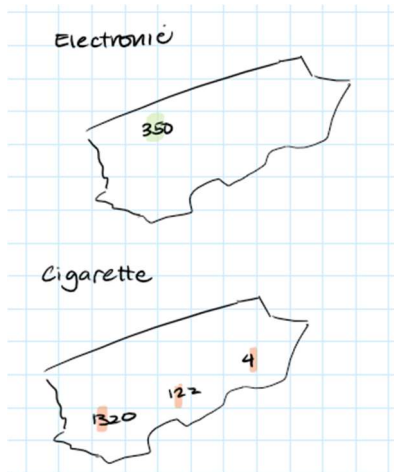
Height of bars is total clear time

2.1.5



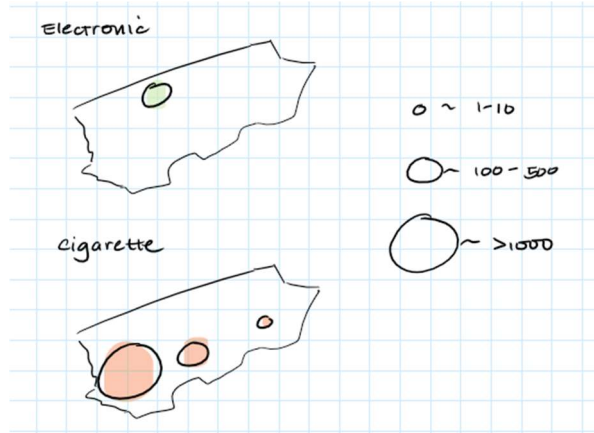
Total clear time percent is the circle area

2.1.6



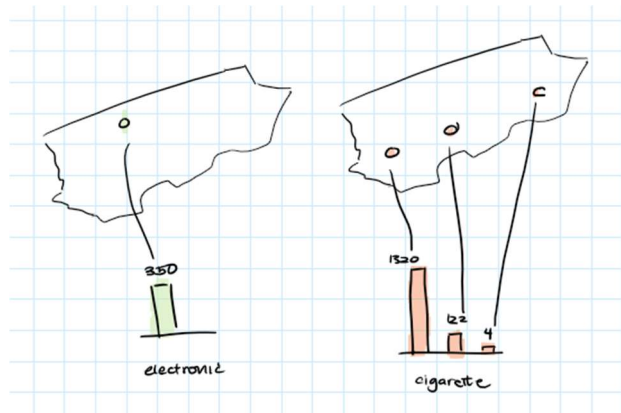
Labelled points colored according to ignition

2.1.7



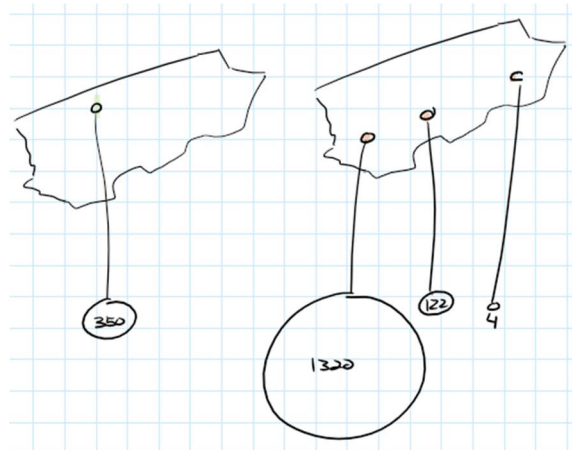
Separated by ignition source,
circle size is the total clear time

2.1.8



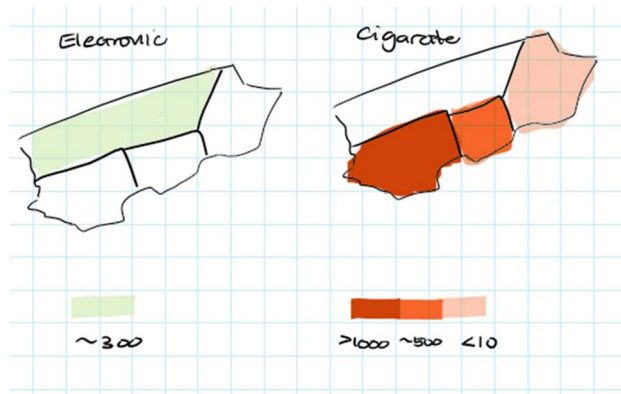
Bar graphs of total clear time

2.1.9



Circle size is the total clear time

2.1.10



More opaque, more total clear
time separated by ignition
source