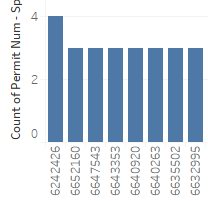
**About the data-**

Dataset- [Building Permits .csv file from Seattle Data Gov](https://data.seattle.gov/Permitting/Building-Permits/76t5-zqzr)

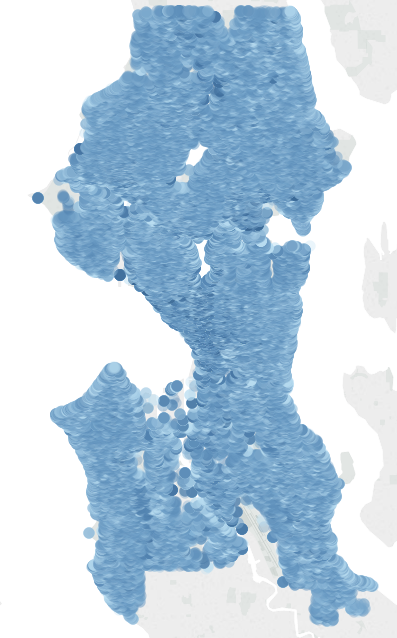
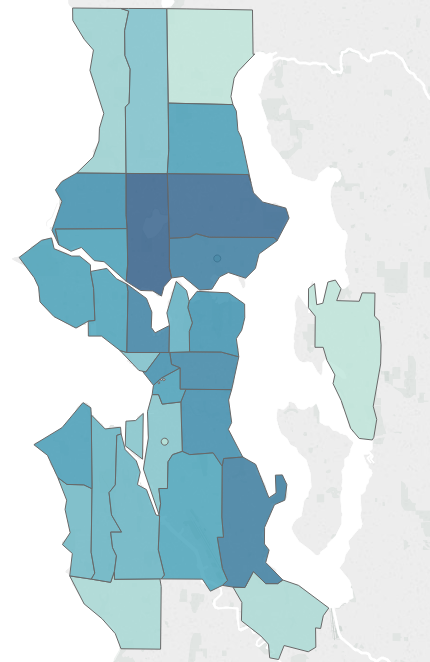
**Data exploration, cleaning, and processing**-

Permit Number should be a unique field- but I noticed it had a trailing set of characters. I split the permit number on the first hyphen ( - ), and to ensure the resulting fields were unique- plotted descending counts for the parsed numeric- the parsed numeric were not unique (picture on left). Further exploration led me to discover some permits had the same leading number- but unique character trails referencing continual permit submissions, or different proposals. I’d lose this uniqueness if I split the permit number, so reverted this step.

Additionally- I had to set the zipcode, city, and state as geographical entities as they were encoded as strings and Figure 1- Sorted permit counts  
numbers initially.

I wanted to check for missing or null values- so plotted graphs (histogram, trends over time, etc.) for various nominal and quantitative fields. Of my dataset of ~120k entries, 6k were missing information on their permit class, and ~14k were missing the permits application date and ~34k the issued date- I didn’t know how to account for this missing data- so either ignored them when creating visualizations, or used the closest matching column with fewer missing fields (e.g.- application date instead of issued date).

**Question 1- Are there spatial trends in the permit file?**

I plotted a heat map on the zip code for the number of applications. I realized zip code is not granular enough (only a few zones)- and the number of applications doesn’t tell us about how money is being spent.

I next plotted every permit’s geo-coordinate data colored by project cost. Adjusting size, and thought adjusting the opacity per dot may help me find some spatial trend. The resulting map suffered from gross overplotting- it is impossible to identify trends in the plot since each there are ~120k entries being plot on the map. The user would need to zoom in per entry- but even then- multiple projects on the same building Figure 2- Attempted visualizations for permit data  
would only reflect the newest one- and granularity on previous buildings is lost.

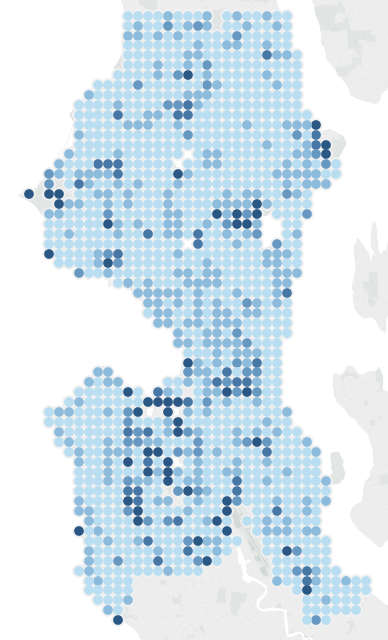
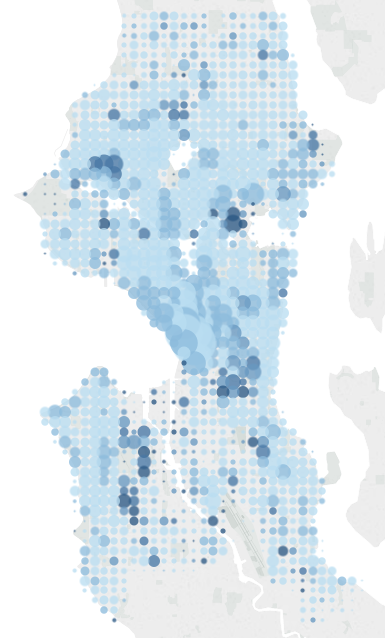
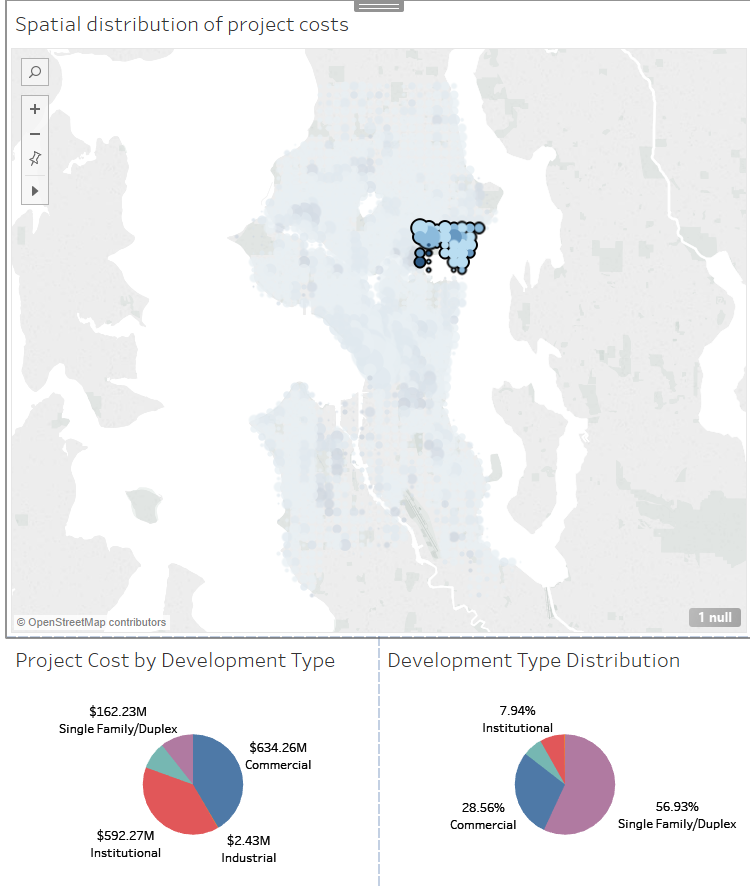
To fix this- I tried binned the geo-coordinates and selected the median project cost by bin be the color. There was an outlier of 88million for the Key Arena making essentially every other dot the same color- so I set the “max” range to be $300k. Some patterns such as the u-District and Downtown having high project costs made sense. I then set the project count to reflect the dot size, and this affirmed that notion that Downtown has seen a lot of development.

Figure 3- Refinement of visualization after binning data

**Question 2- How does the permit class vary by location, and how does correlate with cost?**

I believe Downtown should have a lot of commercial projects- whereas areas like the uDistrict should have more residential development. I wanted to visualize how permits varied by class (e.g.- of all the projects, how many were residential, commercial, institutional, and so on), and to be able to drill down by area.

As a step further- I also wanted to be able to compare the count of a permit with how much money went into the project- are there permits that aren’t applied to as often, but cost a lot more money? How does this change by area?

To implement this, I build two pie charts- one for the count of a given permit class, and another for the total amount of money accounted by projects for that class. I then connected them to my map- so users could select a set of points and see how the permits and their cost breakdown change by areas.

Figure 4- Interaction with dashboard after integrating pie charts visualizing permit data distribution

The picture reflects a selection of points around the uDistrict- notice how although there were only a few applications for institutional development- they still accounted for near half of project costs.

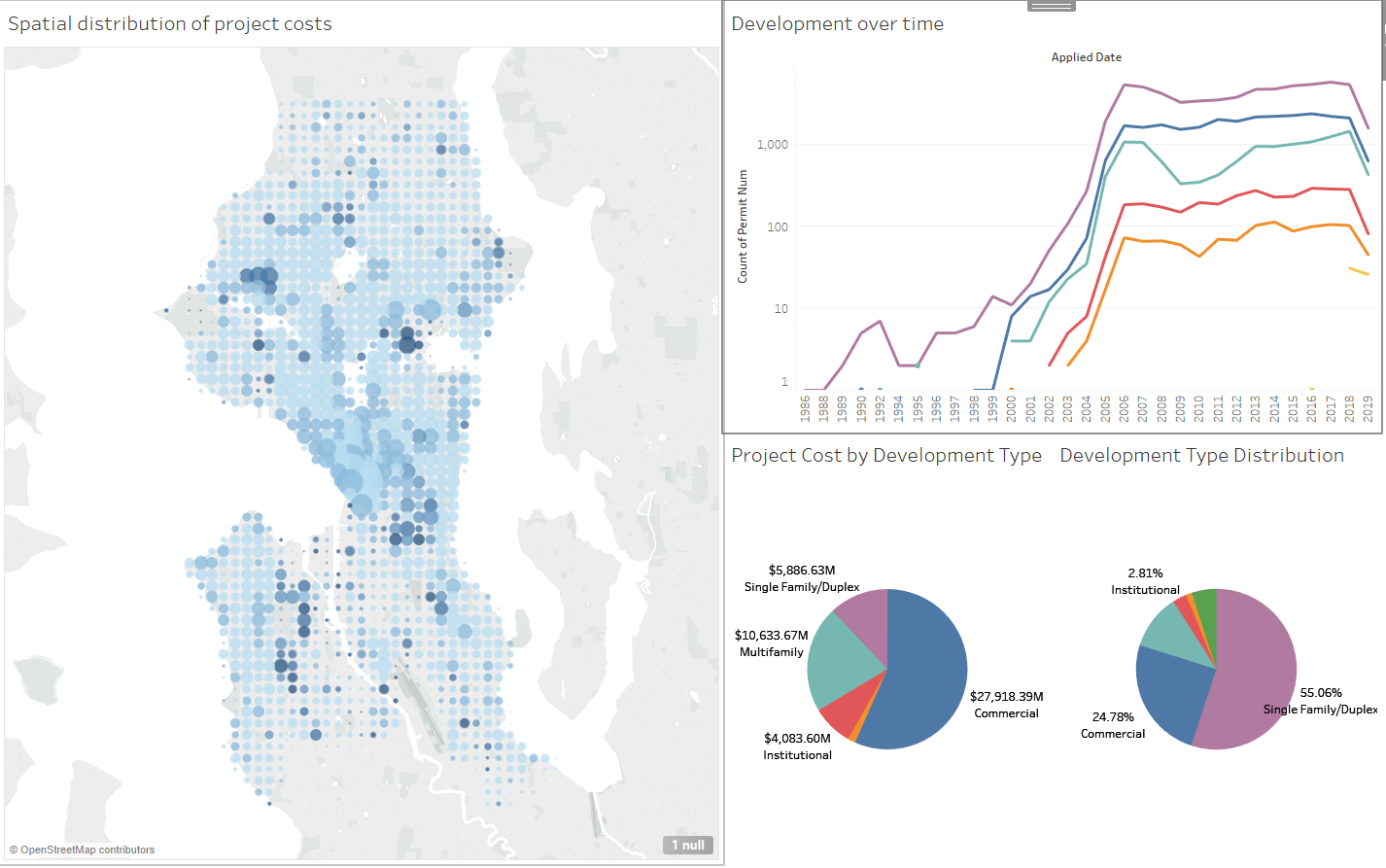
**Question 3- How do applications vary over time?**

Figure 5- Interactive dashboard visualizing permit data for the Seattle housing market

A dashboard for housing permit data for the Seattle area- there is a

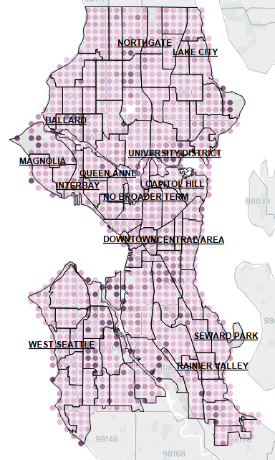
* map visualizing median estimated project costs via color and the number of filed permits via size
* linechart showing temporal trends on how different permits applications (color) vary over time
* pie chart showing the number of permits filed by type- color for permit class, angle for proprtion
* pie chart showing the total project cost by permit type- color for permit class, angle for cost.

This dashboard helps answer by initial question on how do permits vary spatially in Seattle since looking at the map- one can gauge how quickly certain areas are developing by the number of filed permits represented by size- and the scope of the project by color which is mapped to the median estimated project cost.

They can also get additional information by looking at the pie charts to see the different types of permits that have been applied for in Seattle- as well as how much these projects have costed. Selecting a sample of points on the map updates these piecharts to reflect data for only the selected points- so it is possible to see how these permits and their costs vary by county too.

Lastly- there is a linechart visualizing tempral data for the data as well- users can interact with the plot to see how development has changed over time for the area- once again being able to scope it down to an area of interest and use that as a means to compare between locations.

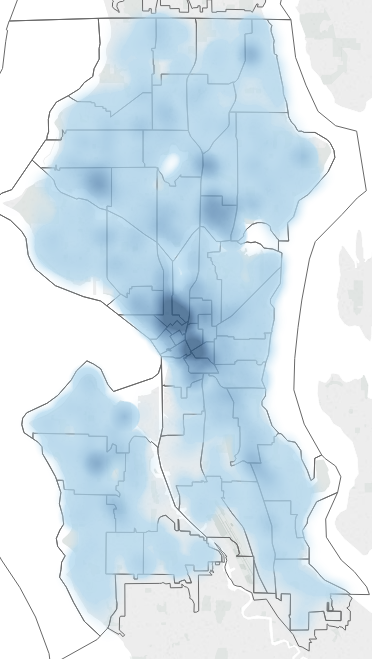
**Here were a few things I tried that didn’t work out-**

**Making the map more intuitive for people unfamiliar with Seattle**

Someone not familiar with the Seattle would be lost to where zones are- and what they are. To fix this, I’m connecting a shapefile with information on different neighborhood and their names. I’m overlaying this atop my map- and I’m also changing the plotted colors to make them stand out more. However- the darker colored circles interfered with the overlayed text and boundaries- making the map look ugly and cluttered.

Figure 6- Color swapped map overlayed with Seattle shapefile and neighborhood names

**Making a density plot overlaid with a shapefile for boundaries**

The dots don’t change dynamically when a user zooms in. I tried making a density plot to fix this- but quickly realized it lacked the affordance that users needed that it was interactable.

Also- my density plots weren’t contained within the shape file itself- some circles overlapped and extended beyond the shape file- this made the map look cluttered and visually unappealing- so was discarded.

Figure 7- Density map for estimated project cost for Seattle

**Changing the piechart for a bargraph to represent the proportion from total for a selected sample of plots**

The pie charts update to reflect the proportion of permits and their estimated cost based on user selection- but users lack the ability to compare that proportional from the total.

I switched the pie chart for a bar graph with an overlayed color that changed based on the user selected region (i.e.- color represents selection; gray represents total from all data), but quickly realized that because the magnitude of plots varied between permit types by an exponential amount- making comparisons between groups would be hard to impossible. In the above plot for instance- almost all the subplots are zero for the selected points which helps inform their proportion vs. the cumulative sum- but the plot does not help us compare how they compare amongst themselves.

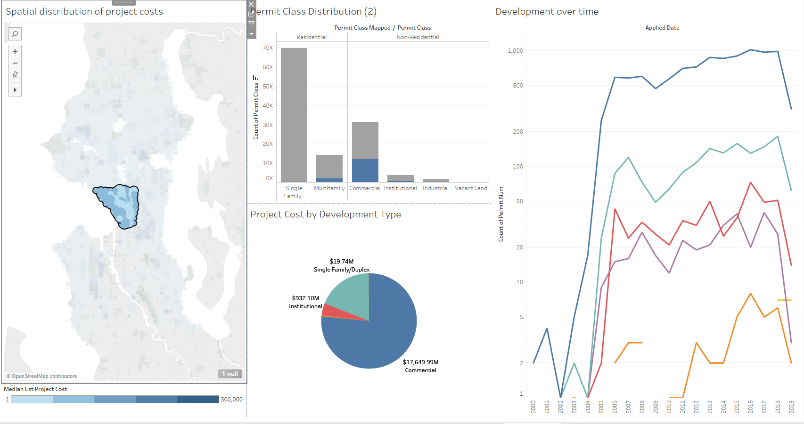


Figure 8- Dashboard alternate- pie chart for permit class distribution by count swapped with bar graph

**Pie chart for contractors by permit by selected sample**

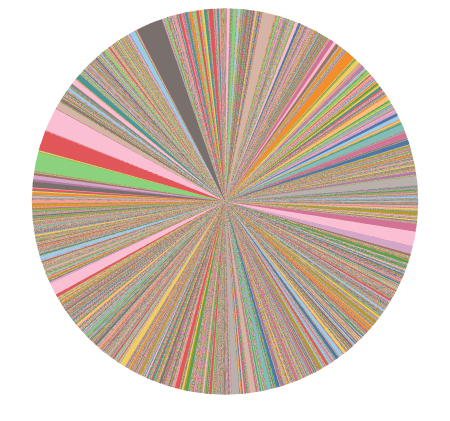
I thought it would be interesting to see if there are spatial trends between which contractor works on which project. However- when plotting this- it was a case of information overload and overplotting- there were too many contractors- even when selecting just a few points in areas like the uDistrict or Downtown which leaned towards a type of permit class. It didn’t appear as though there were spatial trends amongst contractors- so this was omitted.

Figure 9- Pie chart visualizing contractor count for the permit dataset