

Safe Bike Activity Visualization:

Developing a Tool to Highlight and Explore Bike Thefts in
Seattle

John Castro
Fang-Ju (Ruby) Chou
Madeleine Stewart
Xiaochen (Hailey) Yu

Table of Contents

[Abstract](#)

[Introduction](#)

[Previous Works](#)

[Datasets](#)

[Process of Data Cleaning](#)

[Research and Design process](#)

[Persona and Storyboard Development](#)

[Primary Persona:](#)

[Secondary Persona:](#)

[Initial Sketch:](#)

[User Research Method](#)

[First Prototype](#)

[Visualization Interpretation](#)

[User Study for First Prototype](#)

[Feedback from informal usability test:](#)

[Initial Discussion from Peer Review](#)

[Second Prototype](#)

[Visualization Interpretation](#)

[User Study for Second Prototype](#)

[Unsure / Unclear Comments \(Things to improve or consider\)](#)

[Good Features/Positive Feedback:](#)

[Third Prototype](#)

[Visualization Interpretation](#)

[Fourth Prototype](#)

[Visualization Interpretation](#)

[User Studies](#)

[Final Visualization](#)

[Shneiderman's Visualization Tasks](#)

[Evaluation and Further work](#)

[Final Usability Test Results](#)

[Future Work](#)

[Tableau Evaluation](#)

[References](#)

[Appendix 1: Usability Testing Protocol](#)

[Appendix 2: Data Sources and Similar Works](#)

[Datasets](#)

[Similar Works](#)

Abstract

This project explored effective communication of bike theft trends in the Seattle area. Our resulting visualization primarily aims to engage Seattle cyclists with information about bike theft dates and locations, and bike rack locations. This paper discusses the process of developing our visualization, from collecting, cleaning, and augmenting our data, conducting user research and usability studies, and using visualization theories, such as Shneiderman's mantra, to inform our design.

Introduction

The main goal for this project is to assist bicyclists around the Seattle area and promote safety awareness. Seattle was rated the second safest big US city for walking and biking by Alliance for Biking & Walking according to a Seattle Times report (Balk, 2014) and there are a large population of bicyclists in Seattle. However, from Jan. 2015 to Sep. 2015 only, there were 1,109 bike thefts according to our police report dataset, indicating a rate of 4 bike thefts per day. The need for Seattle bicyclists to understand more about bike thefts incidents was not only shown through the number of bike thefts, but also mentioned by the multiple bicyclists we interviewed.

Our intent for this project is to inform bicyclists about trends regarding safe areas for bike parking with access to bike storage while also being able to identify possibilities of theft due to historical records. Through gathering feedback from user research, dataset exploration through Tableau, and incorporation of effective visualization concepts, we were able to develop a visualization that presents information to bike riders and allows our visualization users to interact with data to help make careful decisions and reduce bike theft.

Previous Works

The following section highlights some examples of current works in relation to and drawing inspiration from regarding bicycle commute and safety awareness.

We relied heavily on previous works to guide our design. Initially, we considered either helping cyclists plan routes or informing cyclists on safety trends. When we surveyed current tools used by cyclists, we found that Google Maps, Strava, and similar products were already addressing cyclists' route-planning needs (see Figure 1). Therefore, we decided to focus on bike safety, specifically bike thefts.

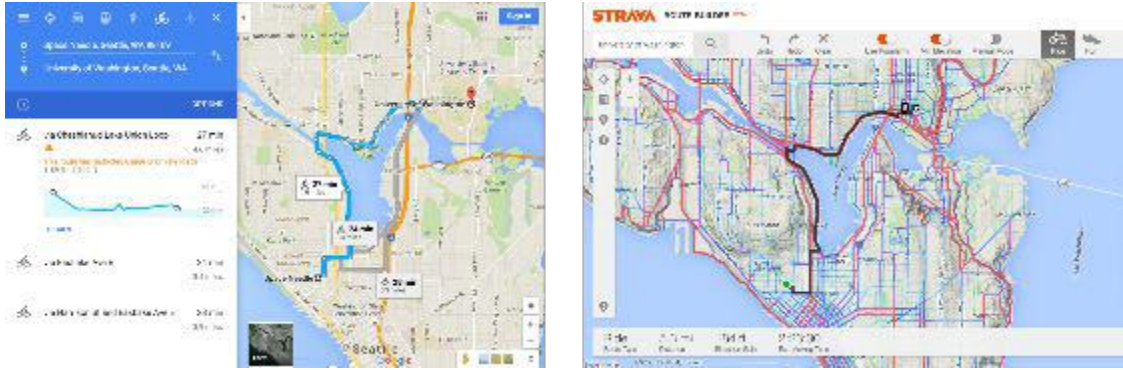


Figure 1: Google Maps and Strava both provide route-planning tools that show information such as incline, traffic, etc.

The interactive map in Figure 2 has static information related to bike routes/trails and bike racks. Users can utilize the sidebar on the left to show/hide information. The map can be zoomed in or out but users cannot click on routes or icons on the map to get more information.

While this information is clearly important to cyclists, we decided not to focus on this information in our visualization. While we initially considered including route-planning information, we chose not to include these dimensions, as our research indicated that people would simply use Google Maps to plan their bike routes. Finally, we wanted to provide details-on-demand to users.

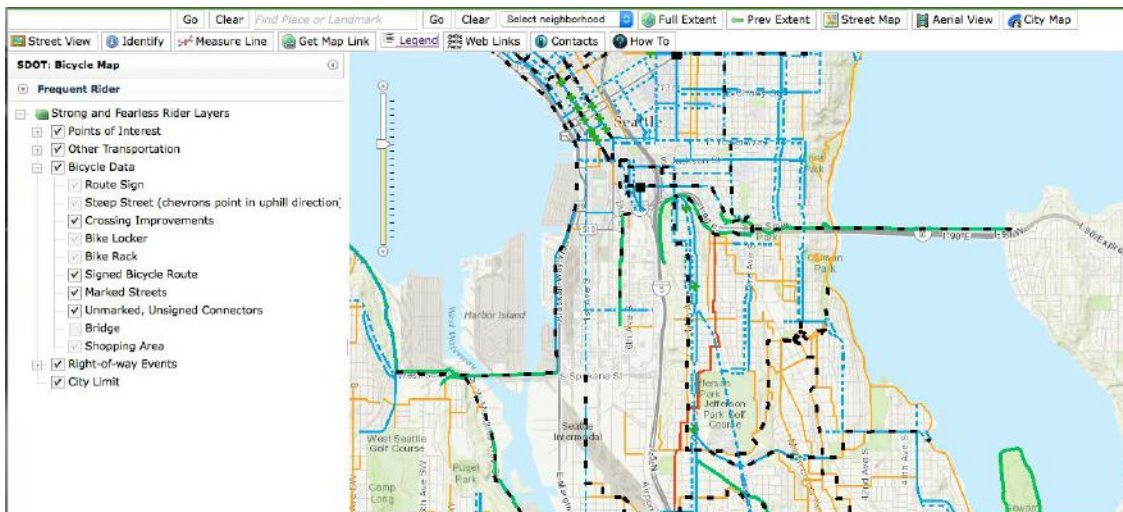


Figure 2: the Seattle Interactive Bicycle Map is a static route-planning tool.

The heatmap in Figure 3 shows the bike thefts in Seattle. Users can filter or sort information using the sidebar on the right. However, the column names in the dataset are unclear and the interface is not user-friendly. The heatmap uses color to show the frequency of bike thefts but the legend is not shown on the map. The use of color is conducive to showing trends by area, but is not useful for understanding trends based on time or for zooming in to see more information about a smaller area or a certain incident. Users cannot click on a certain bike theft and understand more about it.

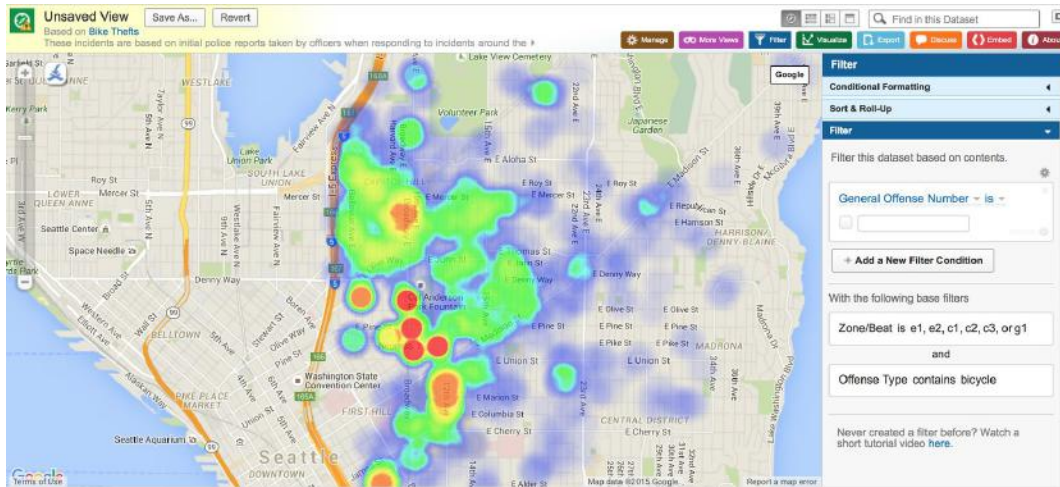


Figure 3: This heatmap shows bike thefts in the city of Seattle using City of Seattle open data.

In our interactive tool, we show similar bike theft information, but instead use scatter plot that allows for zooming and filtering and details on demand. We also added dimensions relating to bike thefts and information about bike rack locations to provide users with a higher density of information.

Datasets

Our primary data source was Seattle Open Data (“Seattle | Open Data”), which is provided through the Socrata platform, a service that makes government data available to the public (“The Data Platform for 21st Century Digital Government”). We augmented our data with geospatial data from the Google Geocoding API and the U.S. Postal Service (USPS), which is discussed further in the Process of Data Cleaning section.

Our dataset on bike racks was provided by the Seattle Department of Transportation via the Seattle Open Data platform. The dimensions used in our visualizations are rack capacity and latitude and longitude, although there were 23 dimensions available. The dataset consisted of approximately 2,500 rows.

Our bike theft data was sourced from “initial police reports taken by officers”, which had been entered into the city Records Management System, which outputs the data to the open data library (“Seattle | Open Data”). The bike theft dataset initially contained over 10,000 rows, with 19 dimensions in total, including information targeted towards police officers (general offense number, district/sector, etc.) and more general information (geographic coordinates, date of theft, etc.). The useful dimensions in the bike theft dataset included geographic coordinates and occurrence date.

Process of Data Cleaning

Our visualization combines data from the Seattle Open Data resource on bike thefts and bike racks, the Google Geocoding API, and the USPS ZIP Code™ Look Up. We used Python scripts to apply changes to our large quantities of data.

1. *Bike Theft Data:* We started with a dataset from the Seattle Open Data resource of approximately 10,500 rows. We noticed that there were duplicate incident reports in the dataset, so we used a Python script to filter incident reports by report number. This reduced the dataset to approximately 6,800 rows.

Our next step was to find a way to separate the data into subsets by geographical region. The simplest and most meaningful way of dividing the data was by zip code. Because all incident reports contained latitude and longitude coordinates, we were able to reverse-geocode the data using a Python wrapper for the Google geocoding API, which takes geographic coordinates and outputs a zip code. The biggest difficulty with this step was running the Python script on the data in batches to avoid hitting the API rate limit, which had a limit of 2,500 requests per day.

We discovered through our user testing that zip codes were not as meaningful as neighborhood names. Therefore, once we had assigned a zip code to each incident of bike theft, we manually compiled a dataset mapping each of the 33 zip codes represented in our dataset to its neighborhood name. We then used a Python script to add the neighborhood name data to the geocoded dataset.

2. *Bike Rack Data:* Similar to the Bike Theft data, we started with a dataset from Seattle Open Data. This dataset did not require any cleaning. We geocoded each bike rack using the Google geocoding API in the same fashion as the bike theft data, and appended neighborhood information to the dimensions of each bike rack.

Finally, we combined the two datasets into one file to show them concurrently in Tableau.

Research and Design process

Persona and Storyboard Development

To get more insight into the need for our visualization, it was essential for us to characterize our users by finding out the type of information that would be useful for them. In order for us to do this, we incorporated the use of personas and storyboards to learn more about our users. According to Alan Cooper, “the best way to successfully accommodate a variety of users is to design for specific types of individuals with specific needs. (Cooper, 2014). For our case and through initial user research, we developed our primary persona who is a cyclist with a particular goal going to a place that is relatively unfamiliar to them and who wants to evaluate the area prior to deciding to park their bike.

Primary Persona:

Who: Casey, a cyclist in Seattle, 23 years old (see Figure 4)

Description: Casey is a student in University of Washington and she loves biking. She normally bikes four to six times a week, mostly for commuting or touring around the city. She enjoys exploring different areas by bike in Seattle during her free time. Casey always worries about her bike being stolen. Although she hasn't experienced a bike theft before, she heard that bike thefts are really common in Seattle. She already has two locks for her bike, but she's still worried, especially when she's in somewhere she's not familiar with.

Major Tasks:

1. When going to a new area, Casey can use the interactive tool to understand if it's safe to leave bike in that area.
2. When parking a bike, Casey can use the tool to decide which bike rack to park with so that her bike would less likely be stolen.



Figure 4: Our primary persona is Casey, an avid cyclist.

Secondary Persona:

Who: Parker, a police officer in Seattle (see Figure 5)

Description: Being a police officer in Seattle for around 10 years, Parker has dealt with a lot of bike theft incidents. Recently, he has received more bike theft reports from citizens in certain Seattle communities. He has begun to worry about the bike theft situation and also would like his team to put more effort into it. He wants to double check the overall trend of bike theft incidents in those communities to confirm his hypothesis about the rising theft rates.

Major Tasks:

1. Parker can use the interactive tool to know if a certain area has a lot of bike theft incidents.
2. Parker can use the tool to see how the number of bike theft incidents in a certain area change over the years.

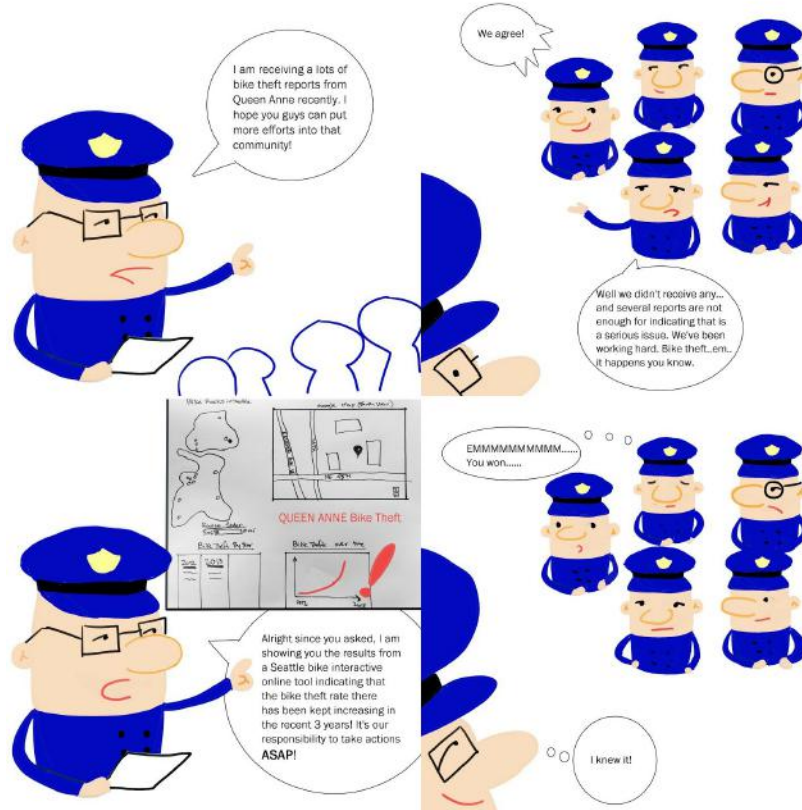


Figure 5: Our secondary personal is Parker, a Seattle police officer.

Initial Sketch:

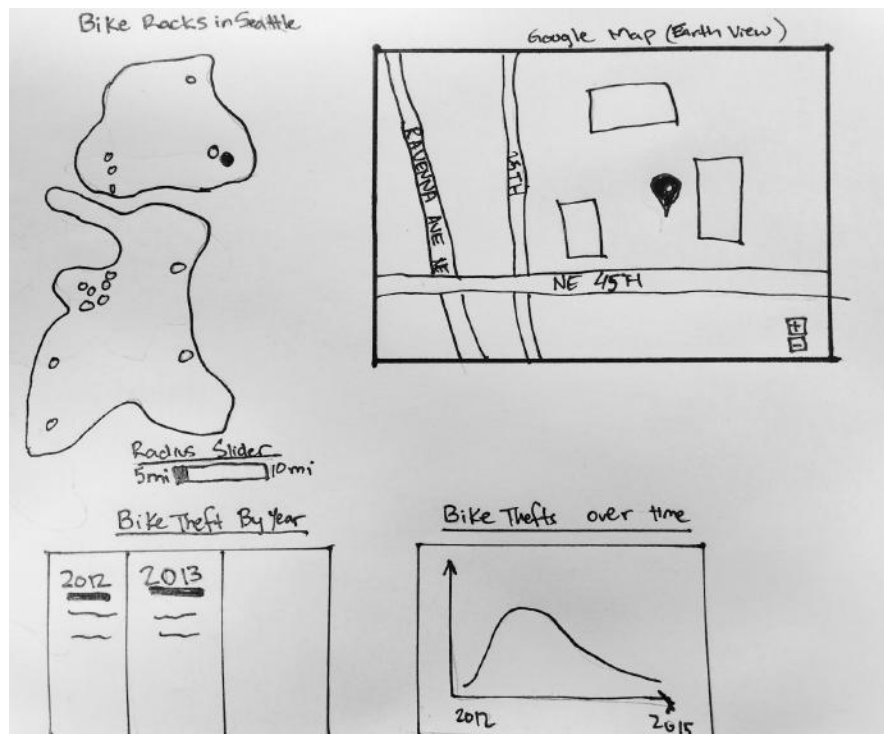


Figure 6: Our initial sketch included a Google Map view, along with a map of bike racks, and several charts illustrating bike theft trends.

Figure 6 shows one of our initial sketches that illustrates the rough structure (dashboard) and look of our intended project. Our idea was to have an overview of the Seattle city region and highlight bike racks from the dataset. The interaction model involves user selection of a specific rack marker triggering details about the bike rack in terms of location and rack capacity. We also envisioned the use of a Google Map connected to the bike rack data to further illustrate what the general area looks like. This would accommodate users not familiar with the area and to help them gauge how busy the selected area appears to be.

The bottom two figures from the sketch are placeholder charts with metrics to indicate theft rates based on the selected bike rack. We particularly wanted to focus on the user quickly understanding the historical bike theft trends relative to the area.

User Research Method

Our project development and research method were loosely based on a RITE (Rapid Iterative Testing and Evaluation) methodology. Our team approached this methodology of “gathering the findings and suggest improvements” and “iterat[ing] on the design as soon as

issues are discovered” as a project management and user research framework (Shirey et al., 2013). This method has helped iterate on our prototype in a way that created constant improvements with testing at each major phase.

To avoid perception bias of our prototypes, we aimed to have different participants at each stage of our prototype phases. Our goal was to elicit qualitative feedback regarding making sense of our concept and ideas as well as task completion of scenarios. Below outlines our general procedure for obtaining user feedback. Our main user testing protocol can be located in Appendix 1.

Usability Testing Highlights:

- General visualization description exercise before interaction
- 3 interactive tasks that are relevant to our personas
- Different participants at each stage of testing

First Prototype

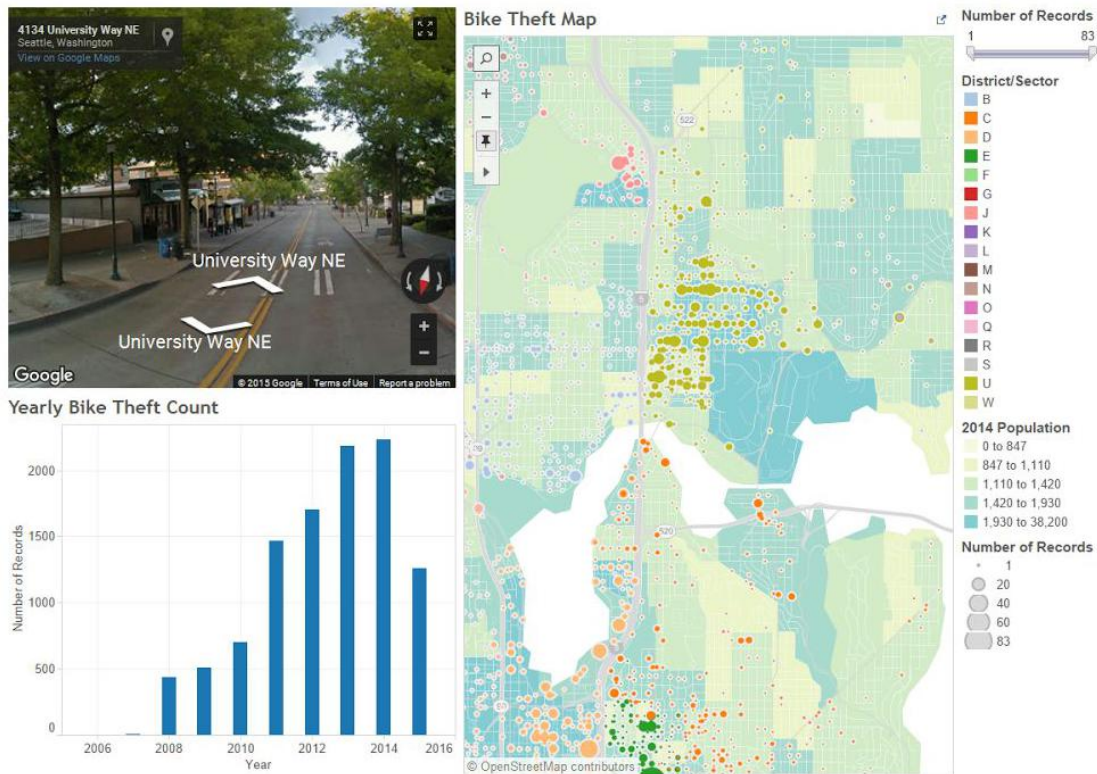


Figure 7: Our first prototype showed a map of bike thefts, a Google street view of each bike theft location, and a bar chart of bike theft counts by year

https://public.tableau.com/views/WOW/BikeTheftVisualization?:embed=y&:display_count=yes&:showTabs=y.

Visualization Interpretation

Based on the ideas from our initial sketch, we began our exploration by implementing the datasets into Tableau Desktop. We firstly experimented with the bike theft dataset from Seattle government into the software, which contains multiple data dimensions about bike theft including the latitude & longitude bike theft locations, bike theft dates, as well as the generated numbers of bike theft cases.

We plotted the bike theft locations in latitude and longitude on a worksheet in Tableau, and the map of Seattle automatically generated a single dot showing the average latitude and longitude of all the bike theft locations (see Figure 7). We were satisfied with this geospatial visualization, and we postulated that users would be comfortable with this visualization based on their general familiarity with Google Maps. After converting the longitude and latitude variables from quantitative measures to dimension variables for spreading out the bike theft locations into multiple points instead of a single spot on the map, we started to plot bike theft occurrences on the map. Using the generated number of bike theft records, Tableau automatically encoded the quantitative dimension into dots with different sizes for all the bike theft locations from Jan 2006 to Sep 2015. In order to add more complexity to our map visualization, we used colors to encode the district information pulled directly from the raw data and inserted the embedded population dataset in Tableau. We made a dashboard and linked our bike theft map with Google map in street view mode and a bar chart showing yearly bike theft counts from 2006 to 2015 when users interact with the map by clicking onto a specific location for details-on-demand information.

We expected our first prototype to show the bike rack locations from another dataset and implemented a radius slider which would obtain surrounding bike theft data in order for Seattle cyclists to pick a bike rack with relatively lower bike theft risk compared the others nearby. However, we were not able to clean both of our datasets and figure out how to connect the two datasets in order to create valuable details-on-demand for users.

We decided to follow Shneiderman's mantra of "overview first, zoom and filter, then details-on-demand" (Heer). We wanted to show an overview of bike thefts in all of Seattle first, to allow users to view overall trends in Seattle bike thefts. Then users could zoom by area to focus on areas relevant to their interests or filter based on the number of thefts. Finally, the users could see details-on-demand, which would show in-depth information about each incident of theft.

User Study for First Prototype

Feedback from informal usability test:

We tested our first prototype with one potential user who is a primary user of our visualization, a UW student who bikes to school occasionally. After briefly introduced the topic of our visualization, we asked him to show us a process of looking for a location around his apartment where he would like to park his bike based on our bike theft prototype in Tableau. He spent several minutes locating his apartment on the map and clicked on the several dots around to view the locations' Google map street view to make sure those are places that he's familiarized with. He narrated that he mainly depended on the size of the location dots to estimate the bike theft risk, and he ultimately picked a dot with relatively smaller size as well as a shorter distance from his apartment compared with some of the other options.

He commented that the straightforwardness of the size of the location dots was very effective in showing the bike theft trends, and the Google Maps street view helped him to confirm the location, as implementing real time location is a technical limitation in Tableau. He did mention that his familiarity with the Google Maps feature helped him to start his exploration with our map prototype. Regarding the drawbacks of our first prototype, he was highly skeptical about the necessity of showing cyclists the yearly bike theft cases bar chart through the last decade, which he didn't use at all during his usability test. However, even before we informed him that our secondary users would be Seattle police officers, he mentioned that police officers could make use of the historical bike theft data in terms of the purpose of their job.

Initial Discussion from Peer Review

The feedback we received from our peers was instrumental in developing our second prototype. They expressed confusion about the direct need of showing population information in terms of bike theft awareness, and they also mentioned that the colors showing population were overlapping with the bike theft dots. Another potential improvement for our project at that phase was how to aggregate the bike theft data geographically. We used the district dimension included in the bike theft dataset in the first prototype encoded with colors, but most feedback from peers suggested that this encoding did not provide any useful information. Our peers also reminded us to reconsider the prioritization of the elements on our dashboard, such as the positioning of the Google Maps Street View compared to the other data visualization elements.

Second Prototype

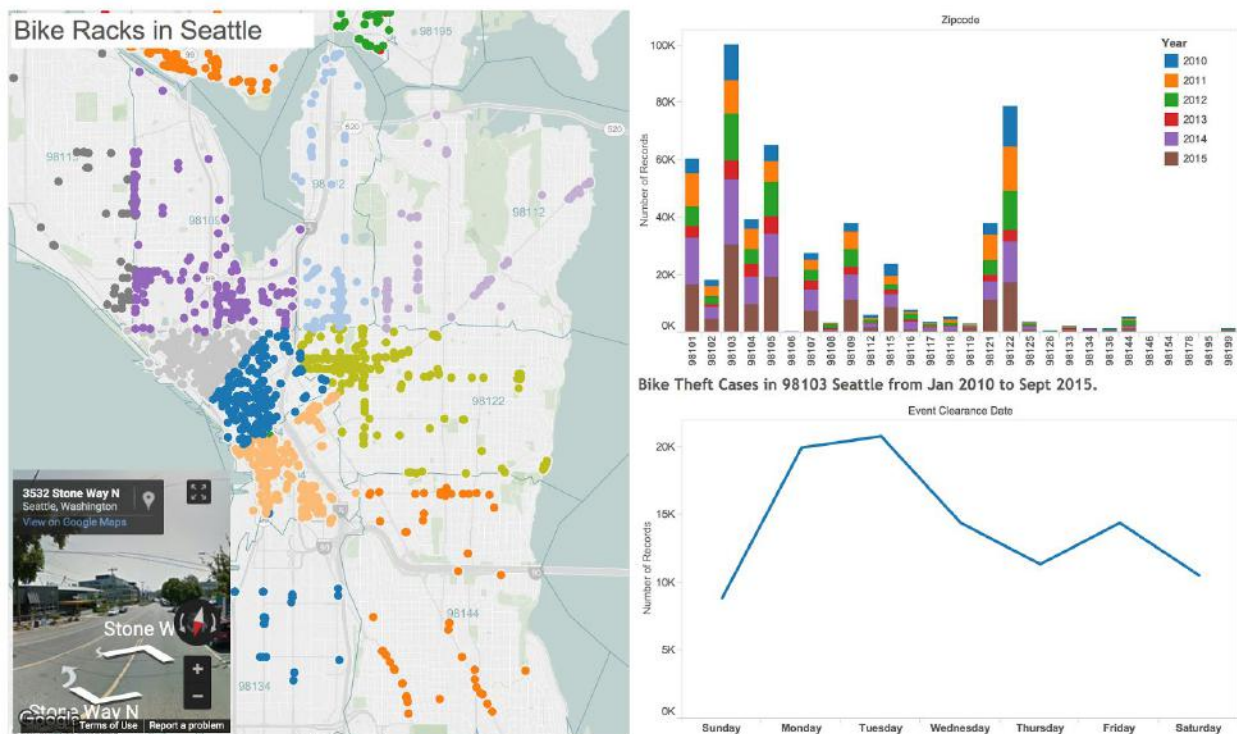


Figure 8: Our second prototype added a color mapping by district to the map, a stacked bar chart relating year, theft count and district, and a line graph of bike thefts by day (<https://public.tableau.com/profile/fang.ju.chou#!/vizhome/SeattleBikeTheft/Dashboard2>).

Visualization Interpretation

After removing the redundant data from our data set, we appended the zip code data from the Google geocoding API to both datasets. We successfully joined the bike theft data with the locations of bike racks based on their shared zip code information, which was also much more standardized than the district/sector data used in the first prototype.

We redesigned the bike racks map in Seattle by labeling the zip codes on the map as well as zip code borders to replace the population information shown in the first prototype, as shown in Figure 8.

We displayed all the bike racks in dots on the map and used colors to encode them based on their zip code area. Mimicking the layout of Google Maps, we placed the Google street view at the lower left corner of the dashboard in a smaller size compared with last prototype. We also replaced the bar chart for yearly bike theft trend with a line chart showing the weekday trend of bike theft, which we thought would be more relevant to cyclists' decisions of

parking their bikes. Another stacked bar chart was added to our second prototype at the upper right corner showing the yearly bike theft cases for each area by zip code.

Users interacted with our visualization by clicking on a bike rack dot to view the weekday trend of the bike theft in the zip code area that the bike locates in, as well as the yearly bike theft case numbers for the past decade.

User Study for Second Prototype

After our midterm presentation our team decided to begin addressing the feedback from our presentation and from peer reviews. Once we had addressed the feedback, we tested our second prototype for its general “look and feel” among some of our identified users. In this study, we were particularly informal in terms of testing protocols and did not have a specific structured interview set up for the process. Essentially, we wanted our participants to explore the visualization further. As far as the user testing protocol for this prototype, three of the group members picked different participants and tested the visualization individually. Outlined below is a summary of comments derived based on Prototype 2.

Unsure / Unclear Comments (Things to improve or consider)

- Participants perceived the dots on map as thefts and use that perception to explore the visualization
- The colors used on maps and the colors used on stacked bars were confusing, and participants tended to connect the colors while there actually wasn't a relationship.
- Participants were confused about how our bike theft data summarized based on zip codes would help a cyclist to compare/contrast between specific bike racks' bike theft risk in order to make a bike rack choosing decision.

Good Features/Positive Feedback:

- Participants found the google street map useful; it provided users with more information about the dot on the map.
- Participants were impressed by the Google map street view, which spiked more interest in our visualization.

Third Prototype

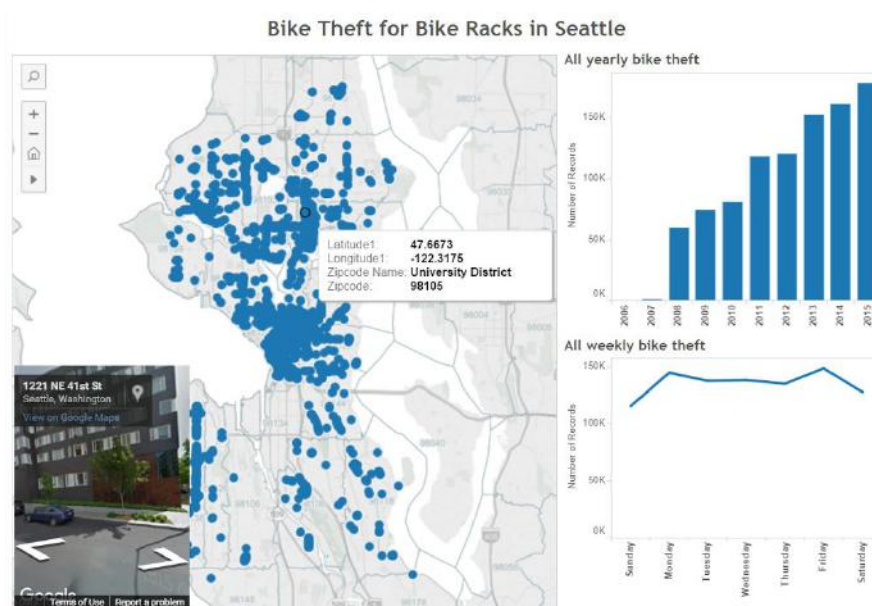


Figure 9: Our third prototype removed the color mappings from the map and removed the neighborhood dimension from the bar chart (<https://public.tableau.com/profile/xiaochen.yu#!/vizhome/Phase2BikeTheft/Dashboard1>).

Visualization Interpretation

Based on the usability study results for our second prototype, we included neighborhood names in our data, using the USPS zip code look-up. In this version of the visualization, we removed the stacked bars and instead used comparative bars only to represent the yearly bike theft information (see Figure 9). As mentioned in Few's book *Show Me the Numbers* (Few, 2004), stacked bars are not suitable for comparisons and showing trends.

We also erased color encoding for neighborhood. Stone (2009) mentioned that in most design situations, using 2 to 3 hues in the color palette is both "aesthetically pleasing and functional." In our previous prototype, the colors used to distinguish between around 30 areas were both confusing and distracting. Thus in this prototype version, we used only blue dots, which largely simplified the visualization.

As Tufte (1983) mentioned in his "graphical excellence" principles, showing greatest amount of information with the least ink is preferred. By largely simplifying our visualization, we can allow our users to focus on the tasks that they are intended to perform, rather than being distracted by overwhelming colors and charts.

Fourth Prototype

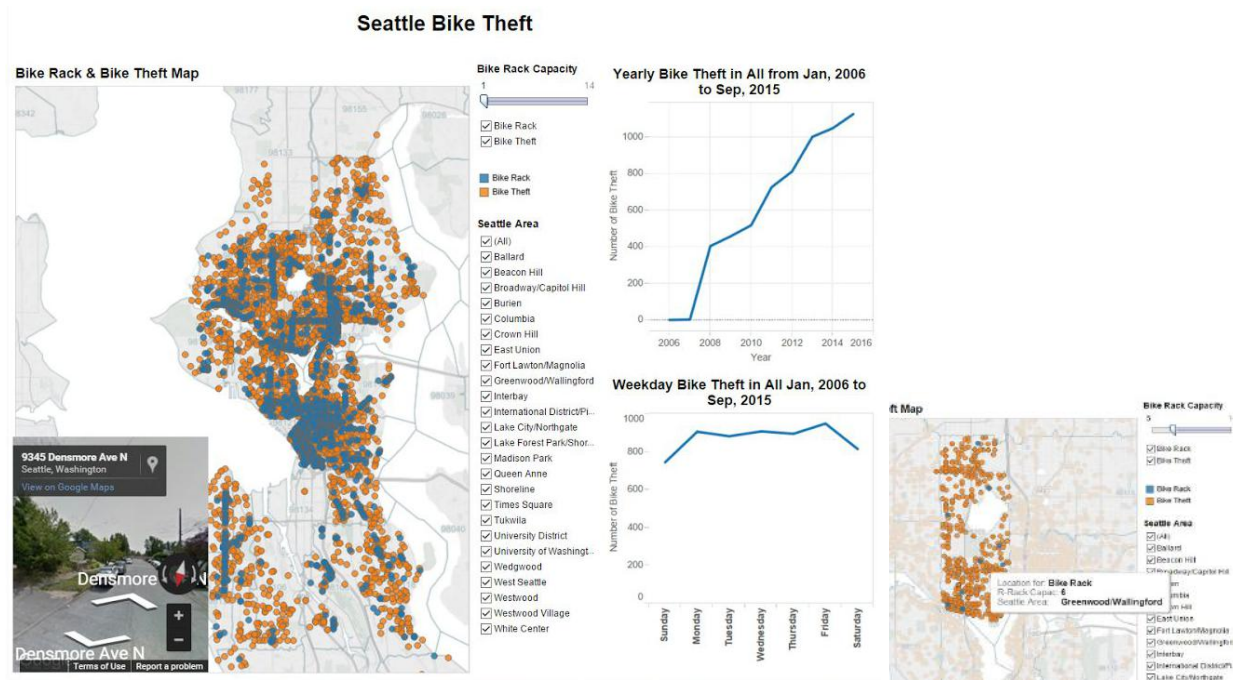


Figure 10: Our fourth prototype included both thefts and bike racks on the same map, changed the bike theft bar chart to a line chart, and added details-on-demand to the map (https://public.tableau.com/views/Phase4/Dashboard1?:embed=y&:display_count=yes&:showTabs=y).

Visualization Interpretation

In this stage, we finally found a way to plot both bike thefts and bike racks on the map by manually merged the two datasets into a single spreadsheet. This was also something that our participants had been talking about throughout our usability tests.

We encoded the two kinds of information with orange (bike thefts) and blue (bike racks), as shown in Figure 10. Orange and blue are complementary colors that have strong color contrast and can be easily separated from each other. These two colors are also color blind-friendly according to Stone (2009).

Clicking on a point on map highlights the zip code area such that users have a clearer view of that particular area. The charts on the right are also updated to only show the information in that area.

We refined the dashboard by adding details, bike rack capacity dimension, filters for location types, and neighborhoods. The two charts on the right were also much more detailed with customized titles and clear axis labels. We also changed the yearly bike theft graph from bar chart to line chart, as the information is time-series and using line chart creates a sense of continuity in change (Few, 2004).

We cleaned the datasets and left only the useful dimensions in Tableau to speed up the reaction time of the visualization. The interactions as well as the presentations of our visualization in this stage were much more complete and refined.

User Studies

The participants in our usability study in this stage had relatively positive feedback our visualization. They mentioned that the layout was clear, and the brushing and linking enabled them to explore more details of the bike theft and bike rack situations in Seattle. They also liked the Google Street View as it provided the actual view of the location.

One major issue that came up during usability testing was that as the line charts on the right changed from overall information to area information, our participants didn't notice that the scale of the y-axis had also changed, which prevented them from accurately perceiving differences between the two charts.

Final Visualization

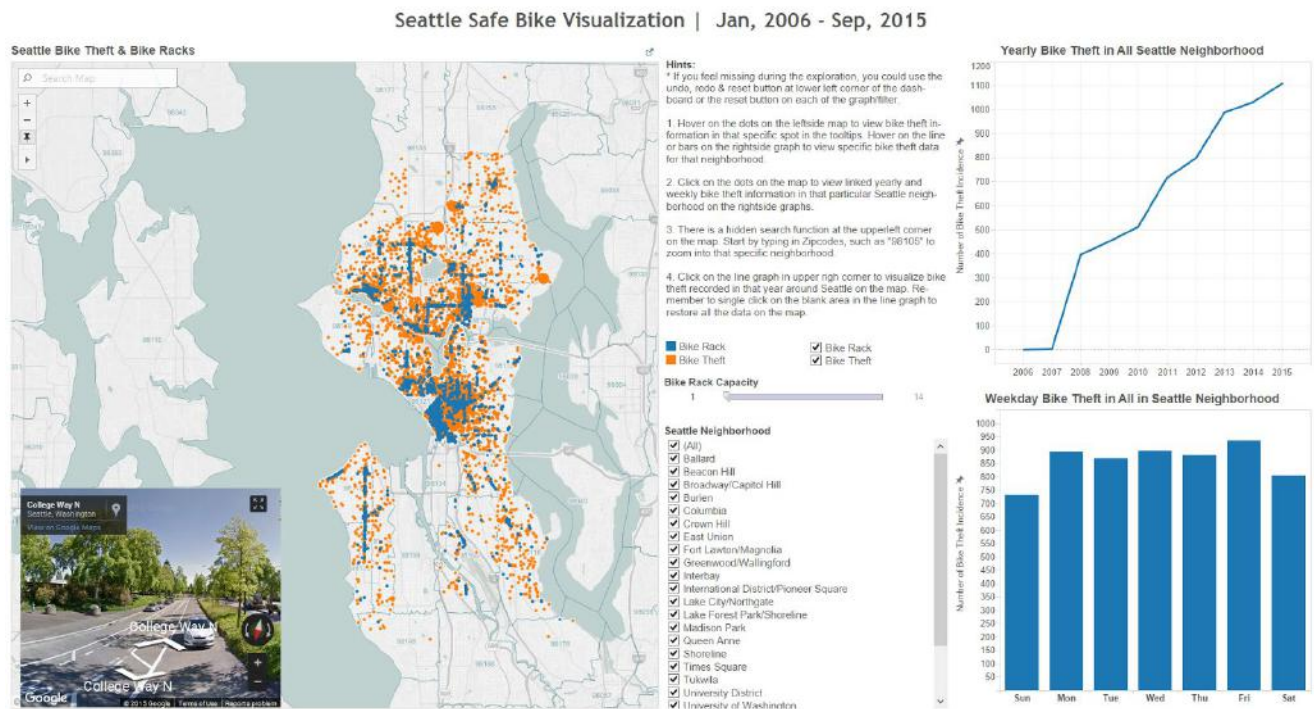


Figure 11: Our final visualization added user instructions, reverted the week day bike theft chart to a line chart, and allowed filtering by bike rack capacity.

Corresponding to the user feedback we received from the last prototype and self-evaluation, we made three changes in our final prototype, shown in Figure 11. Firstly, we added a size encoding to each dot representing a bike theft which mapped number of thefts to

size (a larger dot would represent a location with more thefts). This would give cyclists an easy and straightforward way to perceive the bike theft situation in their interested areas. In addition, we fixed the y-axis scale for both yearly and weekday bike theft charts for precise and honest expression of data. Moreover, we switched the visualization for weekday bike theft from line chart to bar chart, as weekdays do not represent ordinal data and therefore can't be precisely illustrated by line trend in a line chart. We found that the counts were more relevant to the potential users than a trend line. Finally, we refined the details in our dashboard, including adding instructions, shown as "Hints" to guide users through some invisible features in the visualization, customizing the tooltips for details-on-demand, as well as adjusting the overall layout of the dashboard both for visual appeal and functionality.

Shneiderman's Visualization Tasks

In our final version of the visualization, we decided to incorporate Shneiderman's visualization tasks (Shneiderman, 1996) in the following ways:

- **Overview:** the users see an overview of bike theft and bike rack trends upon first viewing the visualization. The map initially shows information for all of Seattle, for all years in the dataset, and for all days of the week.
- **Zoom:** users can use the map's zoom tool to view a close-up of certain areas.
- **Filter:** users can filter to only showing certain neighborhoods, days of the week, or years. They can also filter to show only bike racks of a certain capacity or remove bike racks or bike thefts from the map.
- **Details-on-demand:** users can hover over each point on the map to see further details about the bike rack or incidence of bike theft. Additionally, selecting a point on the visualization will show a Google Maps Street View image of the location.
- **Relate:** our visualization primarily shows two relationships: the relationship between theft and time, and the relationship between theft and location.
- **History:** while our visualization does not provide an "undo" function, it does allow users to see which filters they have selected and a "Reset" button for each graph.
- **Extract:** our visualization does not currently support any form of extraction, as we didn't find saving the data to be immediately relevant to the goal of our visualization. However, extraction could be something explored in our further work.

Evaluation and Further work

Within the constraints of its current format as a Tableau visualization, we find our visualization to be very successful. We conducted a final usability test with four users to gauge the effectiveness of our visualization.

Final Usability Test Results

In our final usability tests, we instructed users to find a safe place to park three bikes at a restaurant in Ballard (a Seattle neighborhood) on a Saturday evening. We were pleased to find that all of the participants accomplished the task successfully. Our participants described our visualization in general as “visually pleasing”, “well-made” and “a good idea”.

From most of their feedback, the visualization filters appeared to be intuitive, and the users did not have a problem finding the destination on the map, or interpreting the information shown. Two of the users remarked on the prevalence of bike thefts in areas with no bike racks. One of the users tended to focus on the map and Google street map view on the left side of the map to gather the information he needed to pick an ideal bike rack.

“The map and the street view are more informative than the other graphs. The dots in different sizes drew all of my attention. The leftside part of the dashboard would be more interesting to look at, which is more straightforward.”

He was especially interested in the Google Map street view.

“Being able to see the bike rack from the Google map is helping me make bike parking decisions. I would like to see what the surroundings of the racks look like, and the reason that I chose this rack than the other, even with less bike theft nearby, is because there seems to be more people in the image. I assume place with much population flow could be safer to park bikes during daytime.”

He also suggested that we make the Google Street view bigger for easier navigation. We explained that our reason for not doing that was to encourage users to focus on the other visualizations, which are more directly related to the bike theft dimension, and that we also perceived the Google Map as an add-on feature than the main focus.

We also received feedback about making the reset button more prominent in the dashboard, which is a default feature in Tableau. Although we put the instructions of how to reset graphs in the hint section, most of the users tended to barely read through the guidelines until they were stuck.

We conducted two of our evaluations on Tableau Public instead of the desktop version. One of our participants was confused by the loading time and thought that she did sometime wrong. She eventually found out that it was a general loading issue because the situation happened repeatedly.

“What happened? Did I do something wrong? Maybe not? I don’t know.”

Another comment regarding the user test using a version in Tableau Public is that it is reliant on stable internet connection for responsive interactions. When resetting a selected neighborhood in our map, one of our participants had trouble resetting the selection due to the choppy internet connection. Perhaps testing in the Tableau application (rather than online) would be a better approach.

Future Work

Based on the feedback from users about our final visualization and the tradeoffs we made during our design iterations, there are some areas of improvement that could be addressed in future work. Extending our visualization into an app would increase mobility and access for our users. To extend our visualization we would have to retool it make it appropriate for mobile views, but would want to retain the same concepts learned from this visualization class. With the knowledge from this class, we are much more prepared to make informed and conscious decisions with our design. We also aim to incorporate the similar interaction methods with the data we presently have. However, we may have to make adjustments to better represent our charts, graphs and maps on a mobile platform.

When considering making this visualization into an app, another feature we would like to implement is to use current location functionality (through GPS) to provide instant access to nearby bike racks. This geolocation feature would enable our mobile users to gain more detailed and accurate data based on location, lowering bike theft risks.

Furthermore, expanding on the location accuracy, allowing users to filter bike racks by radius instead of by zip code was a primary goal that did not make it to our current stage of design. This would have been our ideal way to find bike racks especially in an event when we want to compare differing (or nearby) neighborhoods independent of their zip code boundaries. Another improvement on datasets would be to obtain more accurate addresses of the bike thefts and bike racks to enable a search function for street names. During the production of this project, our datasets included only latitude and longitude that were reverse geocoded to obtain the zip codes; obtaining full street addresses would make the search function more useful should the users elect to search incidents or bike racks by address.

One difficulty users found with our visualization was the inability to compare multiple neighborhoods. This is something we would like to implement in future versions of our visualization.

Lastly, we would like to find a dataset with more accurate bike theft date and time reports. Our current dataset only shows date ranges of reports, which is an inaccurate description of the actual date and time a bike was stolen. Being able to pinpoint the actual date and time of the theft would provide a more accurate representation and more robust trend

analysis to provide our users. If more accurate theft data is not obtainable, we should consider methods of displaying uncertainty to avoid presenting misleading information.

Tableau Evaluation

In general, Tableau was an intuitive tool in creating the current version of our visualization. For most of us, this was the first time using Tableau, but we found that the learning curve of this piece of software was not that steep. For example, it was easy for us to implement the brushing and linking functionality in our visualization to show yearly and weekday bike theft statistics in right side charts after selecting on dots on the map. We also found that the trial and error method was very helpful in terms of experimenting with plotting data in Tableau because it offered unlimited undo changes as well as bookmarking on essential stages.

We think that Tableau is very useful, especially in analyzing the data types in datasets. During the exploration of creating the bike theft visualization, there were several times that we figured out that we couldn't successfully plot certain dimensions as we expected because we hadn't converted the dimension correctly in order to plot them by certain graphs. For example, when we were creating the filter for bike rack capacities, we could only embed that filter into checkbox, dropdown menu, and other filter types for qualitative data, which would not allow the users to filter quantitative data. From the data type columns in Tableau, we immediately figured that out and converted "bike rack capacity dimension" to a "measure" (quantitative data in Tableau), and successfully created a quantitative data slider used in our final visualization.

It is also fair to regard Tableau as a visualization tool that encourages users to customize their work, such as the detailed tooltips, customized chart titles, colors used in data encoding, as well as the whole layout of dashboard. It was also nice to use such a tool that produces many automatic optimizations in terms of the data type categories, and suggested graph types for corresponding data types. However, those embedded optimizations could make users less critical about the data they are visualizing, which would influence the accuracy of the visualization, in some circumstances.

For example, when we were plotting the weekday bike theft statistics for Seattle neighborhoods, Tableau suggested that we to use a line chart, because a weekday is a time dimension. At first, we didn't consider that there could be a better graph type for that dimension, because we believed in the qualification of recommended chart from such a professional visualization software. In the end, we decided to change it from line chart to bar chart, as weekdays do not represent ordinal data (as compared to year). The trend shown in the graph really depends on how we were ordered the weekday on the x-axis. For example, it could make sense if we order the weekday from Sunday to Saturday or from Monday to Sunday, but the resulting trend line would be showing completely different trends to users, which could be confusing and inaccurate as ordinal data. We thought the quantitative summed bike theft

statistics for each weekday could be expressed in a more appropriate way than the trend shown by a line chart. From this experience, we realized that the core of good data interpretation and visualization should still be relying on the critical thinking about the specific data, who the audiences are, and what data we are visualizing for them, rather than blindly depending on data visualization tools without know the interpretations.

References

- Balk, G. (2014, April 18). Report: Seattle ranks No. 2 for cyclist and pedestrian safety. Retrieved December 11, 2015, from http://blogs.seattletimes.com/fyi-guy/2014/04/18/seattle-ranks-no-2-for-cyclist-and-pedestrian-safety/#.U1U5Wm_9OJc.twitter
- Cooper, A., Reimann, Robert, Cronin, Dave, & Noessel, Christopher. (2014). *About face : The essentials of interaction design* (Fourth edition / Alan Cooper, Robert Reimann, David Cronin, Christopher Noessel. ed.).
- Few, S. (2004). Show Me the Numbers. *Analytics Pres.*
- Heer, Jeffrey, and Ben Shneiderman. "Interactive Dynamics for Visual Analysis." *Queue* (2012): 30. Print.
- Seattle | Open Data. (2014). Retrieved November 23, 2015, from <https://data.seattle.gov/>
- Shirey, Jenny, Ann Charng, and Quynh Nguyen. "The RITE Way to Prototype." *The RITE Way to Prototype*. UxMag.com, 19 Mar. 2013. Web. 10 Dec. 2015
- Shneiderman, B. (1996, September). The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* (pp. 336-343). IEEE.
- Stone, M. (2009). Expert Color Choices for Presenting Data. Retrieved December 7, 2015, from <http://www.stonesc.com/pubs/Expert Color Choices.pdf>
- "The Data Platform for 21st Century Digital Government." *Socrata*. Socrata Inc. Web. 23 Nov. 2015.
- Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2, No. 9). Cheshire, CT: Graphics press.

Appendix 1: Usability Testing Protocol

Thank you for participating in our usability testing.

The focus of this testing is to see how users use an interactive visualization about bike related information in Seattle. The session will take about 20 minutes. During the session, I will ask you to complete tasks with the visualization. Please let me know when you think you have successfully completed a task. If you cannot complete a task, you may choose to abandon it.

As you are doing these tasks, please try to think out loud—describing your steps to accomplish a task, what you like or dislike, and whatever is going through your mind. During the session, you may ask questions. However, I may not answer your questions because I would like to see how you would work through the tasks. There is no right or wrong answer. Please understand that we are not testing you, we are testing the usability of this visualization.

Basic information collect (If haven't collected before):

1. How many times do you ride a bike each day?
2. Where do you normally go on bike? What's the purpose of your biking?
3. Have your bikes been stolen before?

Task:

1. First of all, before you interact with the visualization, please try to describe what kind of information do you think this visualization shows in general and in each part of the visualization. What kind of interaction do you think that you can do with the visualization?
2. Please imagine the following situation and interact with the visualization:
 - a. You are planning to have lunch with friends on Tuesday in Wallingford, where you don't really familiar with. You are wondering if it's secure to park your bike in Wallingford at that time and see if there's bike racks that you can park your bike on.
 - b. You are planning to move from U-District to another neighborhood after you graduate. You want to know which neighborhood is safer to park a bike.
 - c. You are planning to go biking to Ballard with four other friends. You are thinking if there are bike racks in Ballard that will be able to park all of your bikes together.
3. Do you have any comments or suggestions on our visualization?

That's all of the usability testing. Let me know if you have any questions about this study or our visualization.

Thank you very much for providing your feedback!

Appendix 2: Data Sources and Similar Works

Datasets

Bike racks: <https://data.seattle.gov/Transportation/City-of-Seattle-Bicycle-Racks/vncn-umqp>

Bike thefts: <https://data.seattle.gov/Public-Safety/stolen-bikes/8rw6-mmz7>

Google Geocoder API: <https://developers.google.com/maps/documentation/geocoding/intro>

USPS ZIP Code lookup: <https://tools.usps.com/go/ZipLookupAction!input.action>

Similar Works

Seattle Interactive Bike Map: <http://web6.seattle.gov/SDOT/BikeMap>

Seattle Bike Theft Heat Map: <https://data.seattle.gov/Public-Safety/Bike-Thefts/uha3-8zjc>