

Visualizing Livability in California Counties with D3.js

Written by: Will Kofski, Jason Stitt, William Wallace, Daniel Gilligan, and Tianyu Lin

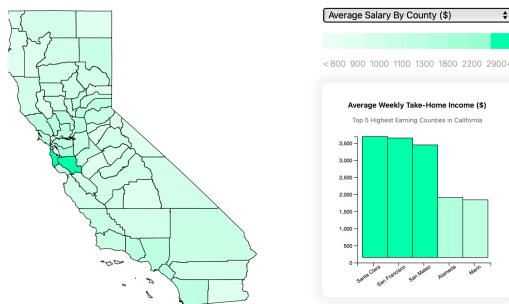
Abstract

When considering where to live, there are a few common factors that are essential to understand before coming to a decision. This visualization aims to provide users with a way to assess how easy it will be for them to live in different areas of California. The visualization will provide a representation of the average rent price, salary range, weather, and crime level for the different counties in California. [GitHub Link](#)

1 Introduction

As a group, we are interested in creating an interactive interface where users can toggle different criteria for the state of California. Overall, we were excited about the chance to synthesize different factors regarding the 'livability' of a certain area. A few factors that we are interested in include weather, education, crime rates, salary, housing costs as well as a composite 'master score'. Given this dashboard the user will be better able to discern the pros and cons of living in a specific area of California. Our group decided to drill down into the county level for our visualization so that we can provide a relevant depiction of costs and quality of life.

While we understand that there are many other rating systems on where to live out there, our group is hoping to bring a unique perspective and novelties to our original equation.



2 Related Work

[Digital Geography](#), a visualization web page, details some techniques for displaying a map using d3 and JavaScript, which will be very helpful for our project. This [paper](#) on California climate change will be helpful in informing our views on weather. Given that our

visualization will likely use color as a channel, this [article](#) regarding rainbow color maps will be integral to our choices. Another helpful resource will be this [paper](#) on the California housing market. We also found some helpful information on how to put our [data into context](#). Additionally, we found another [source](#) about how to synthesize considerable amounts of data together with bar charts as well as heat maps. This Stanford [paper](#) will be helpful when we are tasked with wrangling complex and non-uniform data. One visualization we found in practice is this Bloomberg [article](#) on housing prices. This government [site](#) will be very helpful in determining wage data and such, and it also includes some useful visualizations. This final [source](#) has some really cool line chart visualizations.

3 Use Case

This visualization is intended to be used by anyone who is either interested in how the quality of life compares in different areas of California, or who is considering moving to a new location in California and needs help understanding how living in different counties would affect their life. Visualizations for each dataset that we cover (rent, salary, weather, crime) are often available by themselves. Someone wanting to compare this information needs to find isolated instances of visualizations of each dataset, with no guarantee that they can be easily compared (different chart type, etc.). Our visualization enables anyone to efficiently view information on these attributes through the same medium, in the same place.

4 Data

Rent: [Home Value Data](#)
Salary: [BLS Income Data \(CA\)](#)
Weather: [NOAA County Map \(CA\)](#)
Crime: [Crime Dates by County \(CA\)](#)

5 Task Analysis

Task ID	Domain Task	Analytic Task (low-level, "query")	Search Task (low-level)	Analyze Task (high-level)
1	I want to be able to change between	Summarize	Browse	Discover

	viewing different data within the same view.			
2	I want to be able to compare the overall livability between counties or for a specific county with a single metric.	Compare	Lookup	Derive

The primary consumer of our visualization will be the general public. Specifically, our visualization will be of most interest to members of the public who are either interested in the quality of life in different areas of California or are considering moving within California and would like to better understand their new location.

Our visualization will be primarily developed for discovery consumption. Our visualization will be moreso based around providing functionality for the user to explore the data we have aggregated on their own, as opposed to having written out pre-planned explanations and "findings" of our own. The purpose of our visualization is to provide users with an ability to come to their own conclusions regarding their individual situation instead of drawing a few conclusions and presenting them to all users.

6 Execution Plan & Preliminary Work

In order to create our visualization, we will leverage Javascript and the D3.js library to display multiple visual encodings:

- Heatmap of California, broken up by county*
- A bar chart that will show the selected county(s) statistics*

Accompanying our visual encodings will be a menu in which users are able to 'select' a different attribute for the heatmap to represent. The attributes users can select are 'Rent', 'Salary', 'Weather', and 'Crime'. Our visualization will implement the following interactive components:

- Brushing and Linking on counties*
- Filtering data by attribute*
- Filtering data in bar chart to only include selected county
- Highlighting county that is currently selected in bar chart/map

The visualization and attribute menu will be displayed and interacted with through a webpage. We also may need to use the [Topojson](#) library, which provides an

easier way to create SVG-based maps of states by county.

* Denotes an item we consider necessary for the final project. The other listed components are considered "nice to have", but not necessary to our implementation.

7 Visualization Design

This visualization contains a map of California, next to a bar chart. The map shows data for four different datasets (Rent, Crime, Weather, Salary) by county. The data in the map is represented through the use of color. A user can select multiple counties using a brush, and the bar chart will then show up to 5 counties in the chosen region. The different metrics can be chosen by using the dropdown menu above the bar chart. In addition, users can view the exact statistics of each of the top 5 counties by hovering over the county's respective bar in the bar chart. When a bar is hovered over, the county will be highlighted in the map, and a tooltip will display the statistic.

8 References

Data center. Redfin. Real Estate News. (2021, September 20). Retrieved October 6, 2021, from <https://www.redfin.com/news/data-center/>.

Klammer, P. B. (2014, September 5). D3 Geodata Basics - a map overlay can have many faces. Digital Geography. Retrieved October 6, 2021, from <https://digital-geography.com/d3-geodata-basics-a-map-overlay-can-have-many-faces/>.

Dettinger, M. D. (2005). From Climate-change Spaghetti to Climate-change Distributions for 21st-Century California. San Francisco Estuary and Watershed Science, 3(1), doi:<https://10.15447/sfews.2005v3iss1art6> Retrieved from <https://escholarship.org/uc/item/2pg6c039>

Kosara, R. (2014, November 1). How the rainbow color map misleads. eagereyes. Retrieved October 6, 2021, from <https://eagereyes.org/basics/rainbow-color-map>. Feng, Yalan & Keenan, Donald & Kim, Taewon & Lee, Daniel. (2017). Housing Prices at the Time of QEs in California: Effect of Mortgage Rates. International Journal of Financial Research. 8. 1. 10.5430/ijfr.v8n2p1.

Kosara, R. (2014, November 1). Putting data into context. eagereyes. Retrieved October 6, 2021, from <https://eagereyes.org/basics/putting-data-into-context>

Wrangler: Interactive Visual Specification of Data Transformation Scripts

Sean Kandel, Andreas Paepcke, Joseph Hellerstein, Jeffrey Heer

ACM Human Factors in Computing Systems (CHI), 2011 PDF (1.9 MB) | Video | Website

Group Charter

Group Purpose

Our group was formed mostly because all of our project pitches involved finding some data that would relate to the quality of life in an area, such as salary, housing costs, and weather. Our intended users for our final visualization are going to be both current residents of California, and people that are looking to move there, but are looking for a safer, more desirable place to live.

Group Goals

With this project, all of our group members want to create the highest quality visualization possible, while also learning about different tools and strategies that help generate a top notch data visualization. In addition to this, all of us want to earn an A on the project, since we will be putting in a lot of work into creating this visualization. In order to do this, however, all of our group members have to put forth full effort into all of the project milestones and be willing to meet with the group whenever we need to discuss the project.

Roles & Responsibilities

As a group of five, we all have different roles. Since Dan was the originator of this group, and is a CS and Business major, he has a good understanding of the overall goal of this project, and will serve as Project Manager. Dan will be in charge of determining what kinds of data we might need to search for, and making sure that our team stays on track with the project. Will and Billy are more technical, and will do more work in data cleaning and the actual plotting of the graphs. Will and Billy will be our technical developers. Finally, Jason and Tian will be in charge of the end webpage that we produce, ensuring that it is easily usable and fully implemented. Tian and Jason will serve as our Front-End Developers and Quality Assurance members.

Ground Rules

We decided that we will meet at least once a week over zoom, facetime, or in person to discuss the upcoming project milestone. If there are a lot of things that we have to get done for the week, we can meet more than this if everyone is available. If any of our group members have objections to anything that is going on with the project at the time, then we can discuss it at these meetings. We also have a groupchat where we can discuss anything project related. If we notice that someone is not pulling their weight, then an in-person group meeting will be held so that we can talk about what their issue is and why they aren't able to do their assigned work on time. At the very least, we expect

everyone to meet with the group at the assigned meeting time once a week.

Potential Barriers & Coping Strategies

As a group, we expect full participation and effort from each member. If a member is not contributing an equal amount, the group will hold a meeting and explain the issue with the member and make sure to encourage them to participate more. If necessary, we will point them to the group charter created in order to remind them of each individual's responsibilities. Another barrier we may encounter is one in which a group member feels their opinions and ideas are being disregarded. To solve this, that group member should raise the matter with the group and explain their frustrations, and the group should recognize and understand this, while including some of that member's ideas. We have all experienced a group project in which one person does not contribute an equal share, but by staying on top of the group's dynamics and ensuring the member understands their importance to the project, we can hopefully avoid these circumstances.

We also see a few potential barriers to the actual implementation of the project that we have planned to cope with effectively. The first is that it will end up taking too much time to make a dynamic, toggleable heatmap for each attribute. In order to cope with this barrier, we would create individual heatmaps for each dataset (that will be dynamic by revealing stats on hover, or something similar). A second potential barrier is that it will take too long to implement each of the four datasets that we found. In order to cope with this barrier, we would choose the easiest datasets to work with (salary & rent data), and only display those datasets. Finally, if it is too hard to display the data as a heatmap effectively, we will use bar charts or alternative visualization structures to display the datasets, either toggleable or not depending on the time we have.

Group Update (11/3)

We have all been abiding by the agreed-upon guidelines in this charter. Each member is contributing to every milestone, we have been meeting regularly, and discussions have been professional and inclusive to everyone. We also all feel comfortable with our group roles and responsibilities thus far. There are no problems or issues to report.

Appendix

Interviews

Would you prefer to view each category in its own view (being able to toggle between rent, salary, crime, etc. data from a menu on the chart), or be able to view multiple/all types of information at once within the same view (rent, salary, crime, etc. data all on the same chart at once)?

- I think being able to view multiple categories of information within the same view would be beneficial for analysis and determining relationships amongst the data. Being able to filter down to view only 1 or a few of the categories would also be ideal to get a better sense of a particular category.
- I would say I prefer each category in its own view, because since we are doing it by county and there are so many counties in CA, and if we have to keep track of so many countries and so many factors the information might be overwhelming. In this case a separate category would make things more clear.
- I think having all the categories within one graph will make it easier to see the correlation between the prices and all the factors. If you have them separately displayed then you might need to use different graphs to achieve the same goal.

Please rank the following categories in order of importance to you when considering where to live: avg. area salary, avg. area rent, avg. crime rate, weather info:

- 1: rent, 2: weather, 3: crime, 4: salary
- 1: salary, 2: rent, 3: weather, 4: crime
- 1: rent, 2: crime, 3: salary, 4: weather

What types of weather information are you interested in?

- I would like to see the average high and/or low for each county in a time of year I could specify. Maybe by month or something like a slider for the month and a toggle for high/low
- I would be interested in seeing temperature by county across the 12 months and also humidity levels for each county for a given month.
- I would like to know if this area has some extreme weather, for instance heavy rain / snow or maybe tornadoes. More importantly I need to know if those extreme weather is causing damage and how bad are they.

Are there any other factors that would be of interest to you in this visualization?

- [Population changes](#)?

I would say I am interested in how other factors are going to affect the price or rent of the house in California. My suggestion would be we could add an extra list that has factors that are less related to the house price, for instance we could have surfing data since there are so many coasts in California and that would make our research more interesting.

- One other factor I'd be interested in is potentially the homeless rate by county or poverty rate by county, because I think a lot of people looking for a place to live, myself included, would prefer places with a lower homeless or poverty rate.
- One other factor that I would be interested in is the job distribution. For example I'm a software developer, and I want to know how many software developers are in this area. Because I might want to live in an area that has more colleges.

Appendix B: Data Exploration

Data Review:

- We have 4 separate files for our data, described here:
 - Rent: County column is categorical, and the other 5 columns dealing with different sized units (by bedroom) and their rental prices are quantitative.
 - Salary: Area column is categorical, and employment numbers and average weekly income are quantitative.
 - Weather: County is categorical, “Value” (temperature) is quantitative. We will not be using the other 3 columns in this dataset, but rank is ordinal, and “Anomaly” and the mean from 1901-2000 columns are quantitative.
 - Crime Rates: County is categorical, and the remaining 10 columns depicting different types of crimes & their rates are all quantitative.
- Aside from the column that differentiates each county, which is categorical, all of the data we will be using for this project is quantitative. Some are price-related, others are temperatures, and others are rates per 100,000 residents (crime). Rent, crime rates, weather, and salary are all public data collected using statistics, either from the government or another reliable organization, such as the National Center for Environmental Information (used to obtain weather data).
- We did not notice many biases in our data sets. As stated above, all of this information comes from the U.S government. Someone arrested for a crime will be represented in the Crime data. It’s also unlikely bias would exist in a weather dataset, and the same goes for Rent data collected using rent listings. The only dataset that could contain some sort of bias would be Salary. Salary is likely collected through government surveys, such as the U.S Census. These surveys do not get 100% completion, and because of that, average income amongst counties is likely not exactly what it may be in actuality.

Potential Issues:

- We noticed no potential issues with our data. As discussed, this data is all very reliable and comes from the government or other national institutions. There is no missing data in any of the columns we plan to use, nor are we confused about

any of the variables. Overall, our data is quite straightforward and we do not believe there will be a problem with the actual data itself.

Insights:

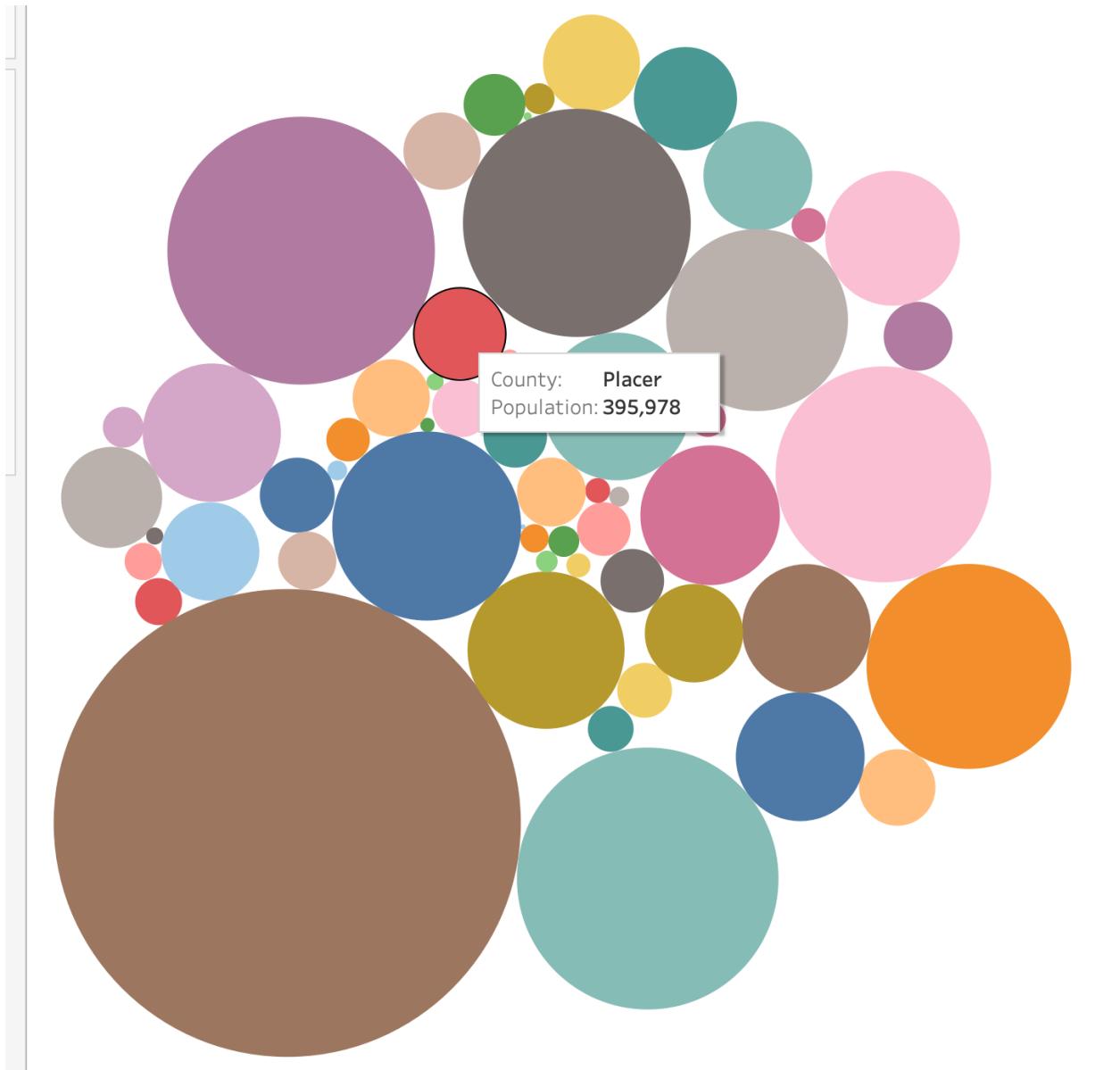
While exploring our data, we came across a lot of relationships amongst columns that we expected. For example, average weekly income and rental prices by county were strongly positively correlated, and San Francisco, as expected, had the highest rent and income levels. One thing that surprised us was that it was difficult to find correlations between different types of crime and other data. We expected to find that there was possibly a relationship between lower income and higher crime rates, or higher income and higher vehicle theft rates, etc. A surprising relationship we found in exploring this data was that average temperature and non-violent crimes had a positive correlation. Obviously, this could be a coincidence that the counties with higher crime rates are generally further south, but it is interesting nonetheless. We did not note any missing/confusing data in our exploration.

Tableau screenshot

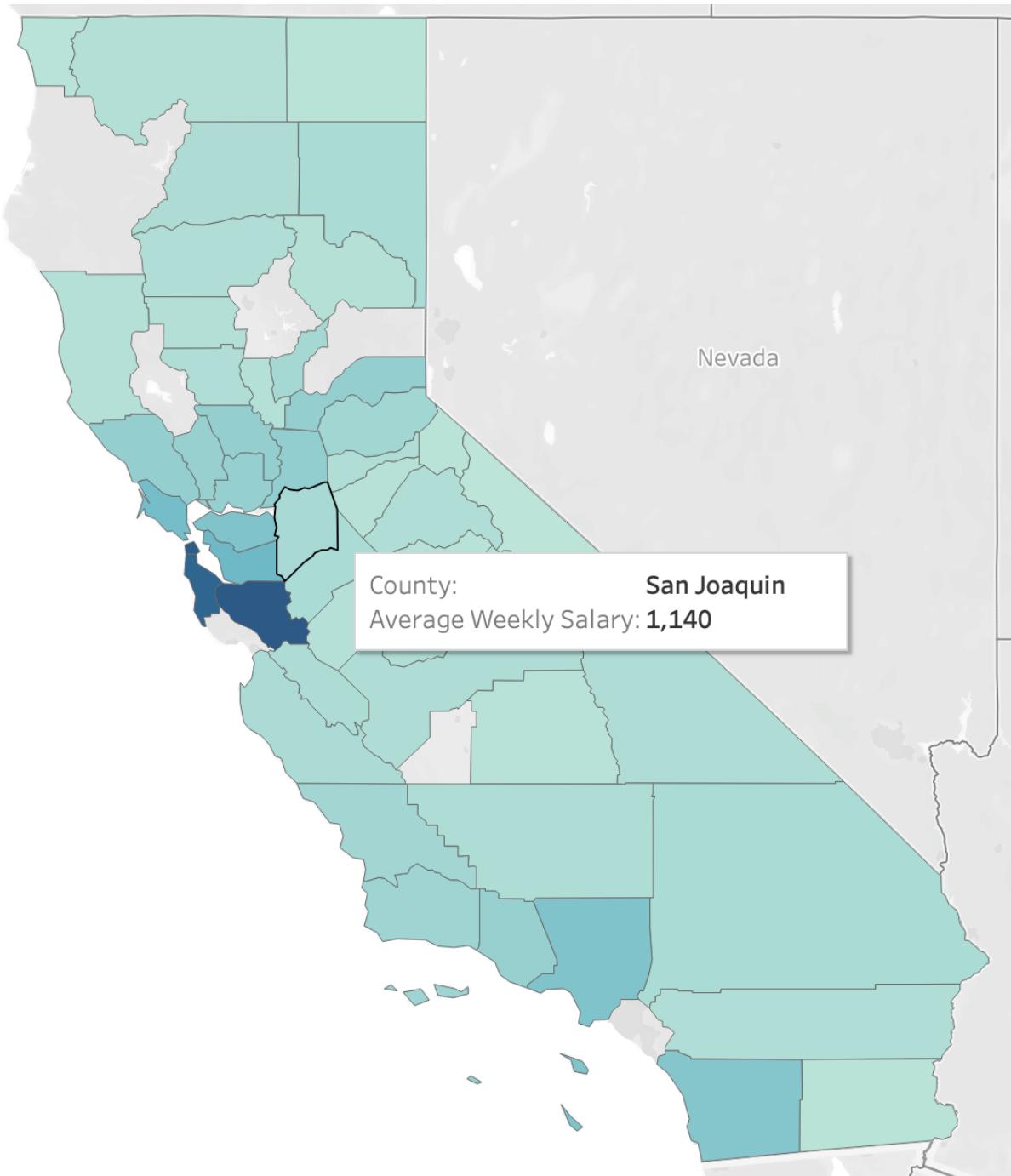
Crime distribution



This visualization describes the crime distribution in terms of types and grouped by counties. The darker the color the higher the crime rate would be. From the graph we can see that Alpine has the highest crime rate.

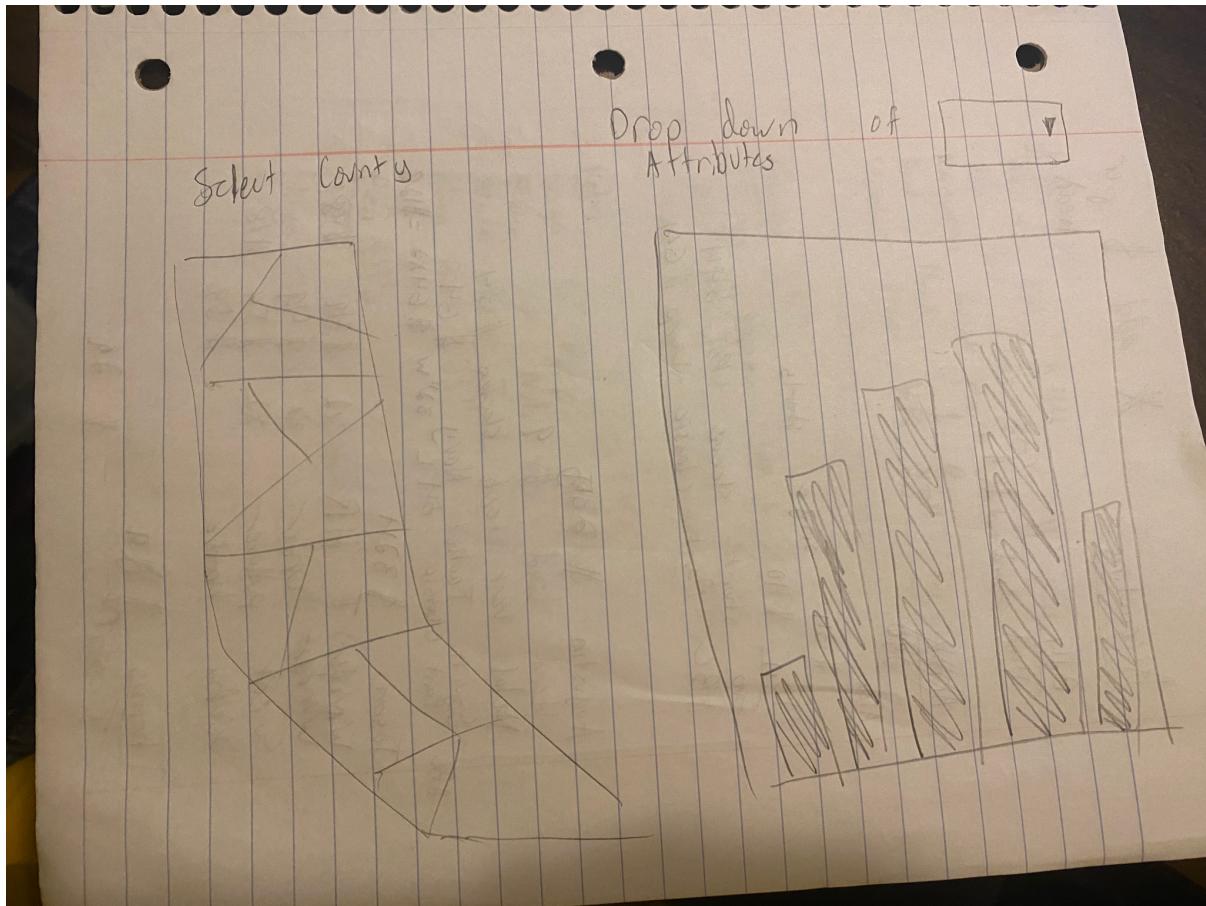


This visualization shows the population date in California. People can easily find which county has the highest population by looking at the size of each circles, if you move your mouse to the circle you can find out the exact population data.



This visualization shows the average weekly salary data. From the graph we can see the county that has higher salary will have darker color. Moreover, if you move your mouse to the map you can see the exact amount of salary.

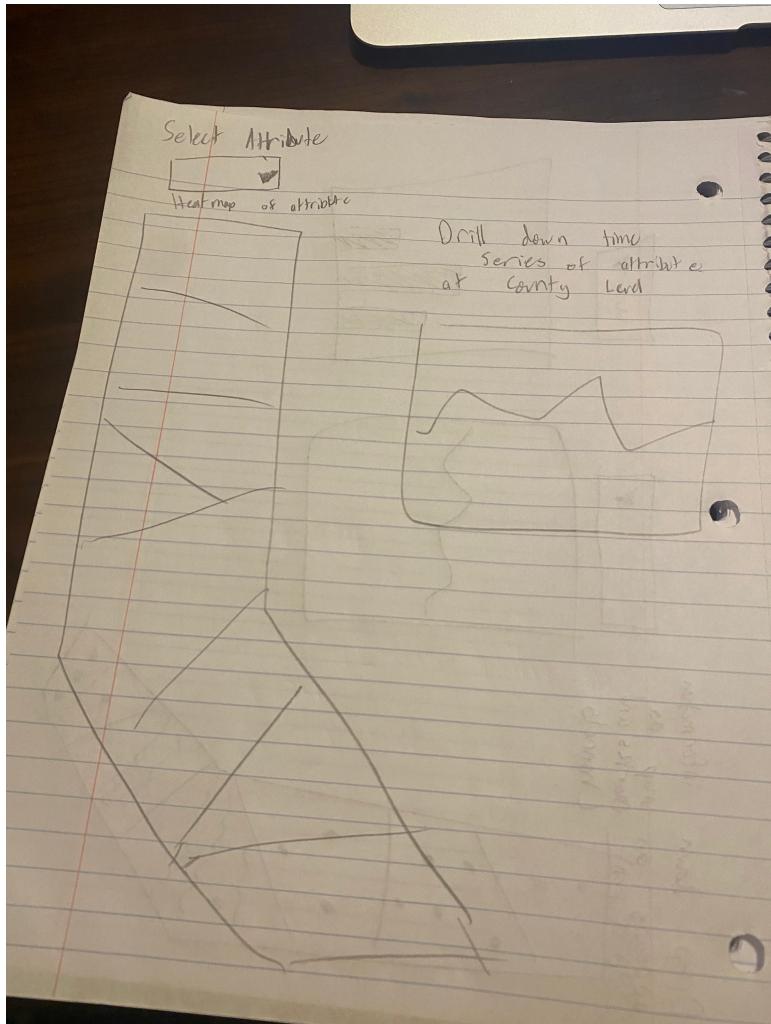
Sketches



Creator: Billy Wallace

Why I chose this sketch: I drew this sketch to be able to illustrate the attributes across each county. This vis allows for a user to drill down into each attribute individually at the county level and also see how each attribute plays out across the entire state.

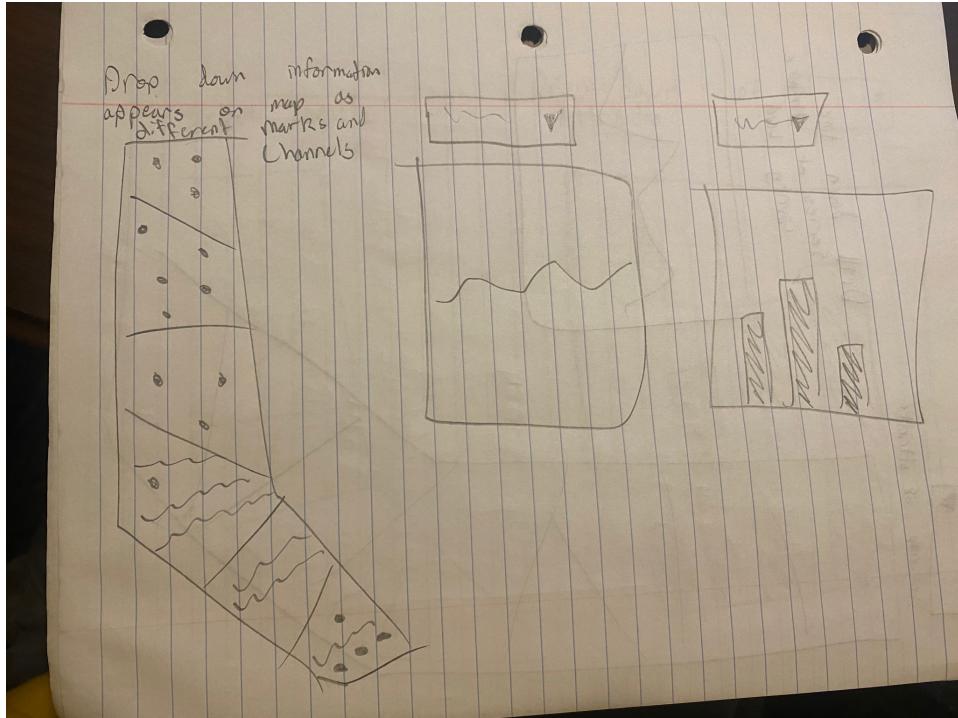
What tasks does it accomplish? This vis accomplishes tasks 1 and 2. Task 1 because the user can change which data is being viewed within the same view. Task 2 because the viewer will be able to toggle to an overall livability metric.



Creator: Billy Wallace

[FAVORITE #1] Why I chose this sketch: I drew this sketch to be able to illustrate the attributes across each county. This vis allows for a user to drill down into each attribute individually, and it's a little more static and simple than the previous drawing. This should make it easier for the user to interact with the vis.

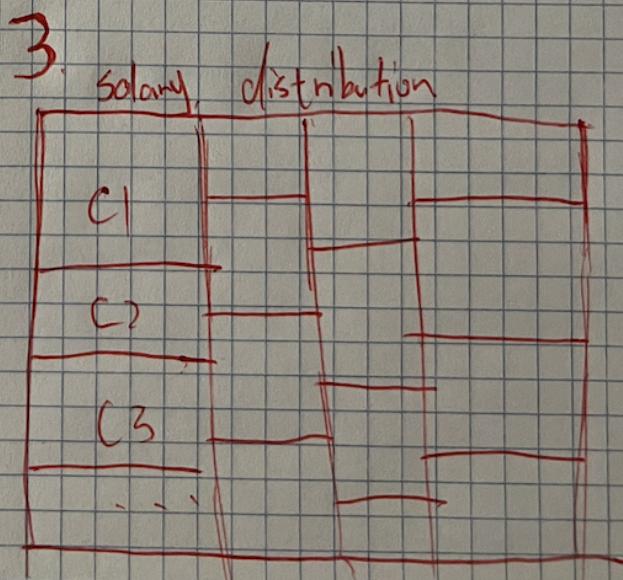
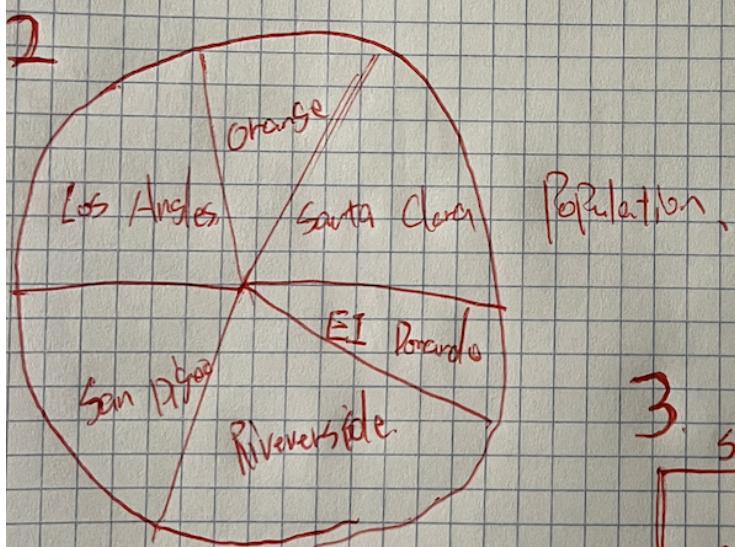
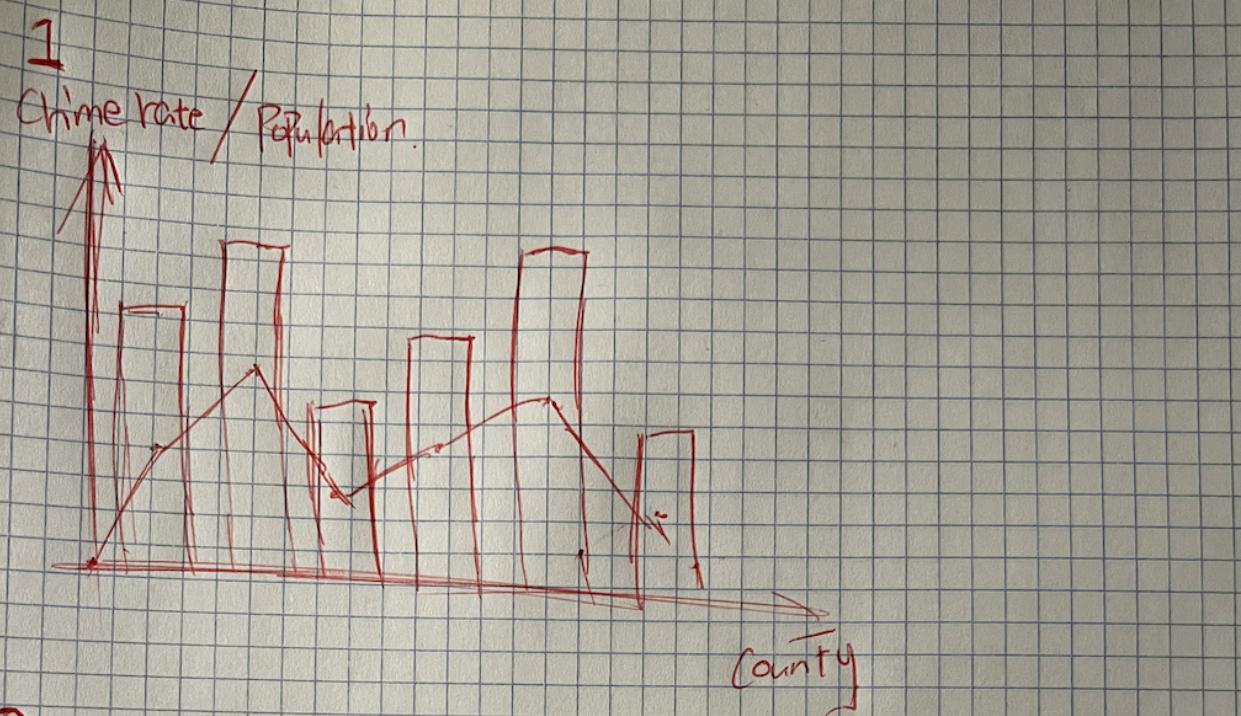
What tasks does it accomplish? This vis accomplishes tasks 1 and 2. Task 1 because the user can change which data is being viewed within the same view. Task 2 because the viewer will be able to toggle to an overall livability metric.



Creator: Billy Wallace

Why I chose this sketch: I drew this sketch to be able to illustrate multiple attributes across each county. This is by far my most challenging vis for a user to be able to learn and interact with. The user will also be able to have two separate charts that they will be able to manipulate using a drop down menu.

What tasks does it accomplish? This vis accomplishes tasks 1, 2 and 3. Task 1 because the user can change which data is being viewed within the same view. Task 2 because the viewer will be able to toggle to an overall livability metric as well as any other supplemental metric. Task 3 is accomplished because the user will be able to view multiple categories of data within the same vis.

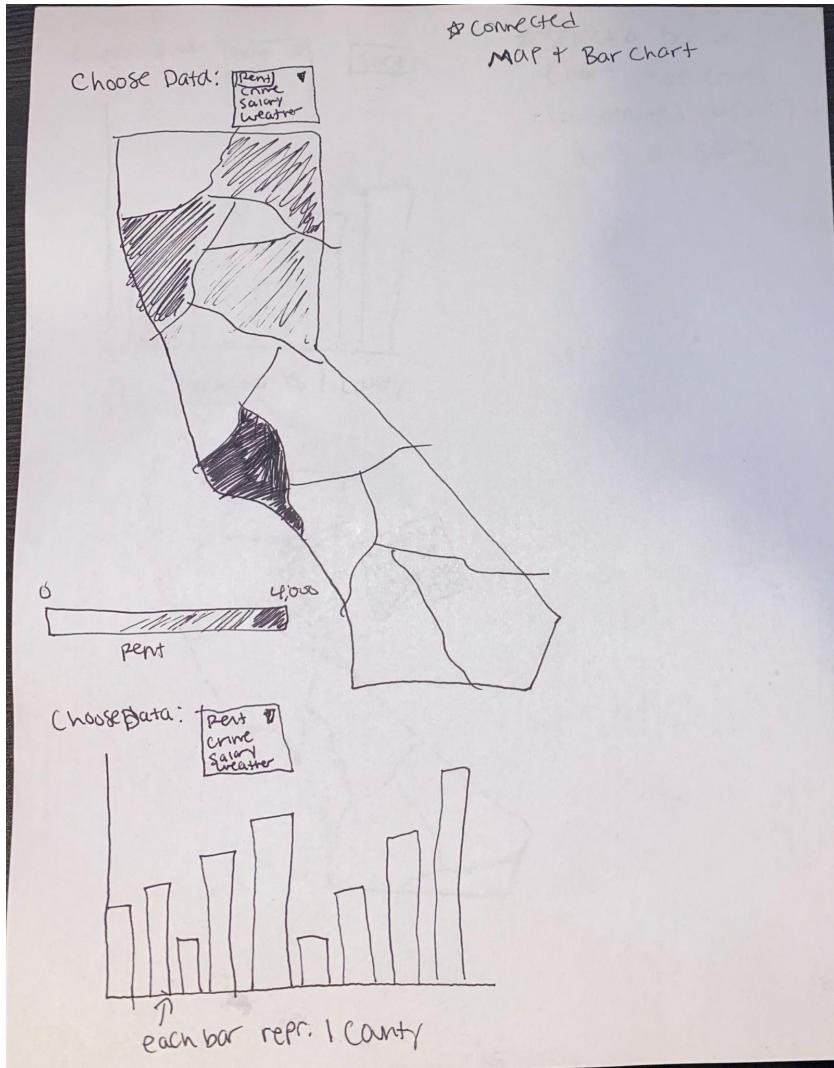


Creator: Tianyu Lin

My first sketch shows the correlation between population and the crime rate, the reason for me to pick that is it can show the correlation between two factors. User may decide their own factors that they want to find the correlation and we will list a branch of options to the user.

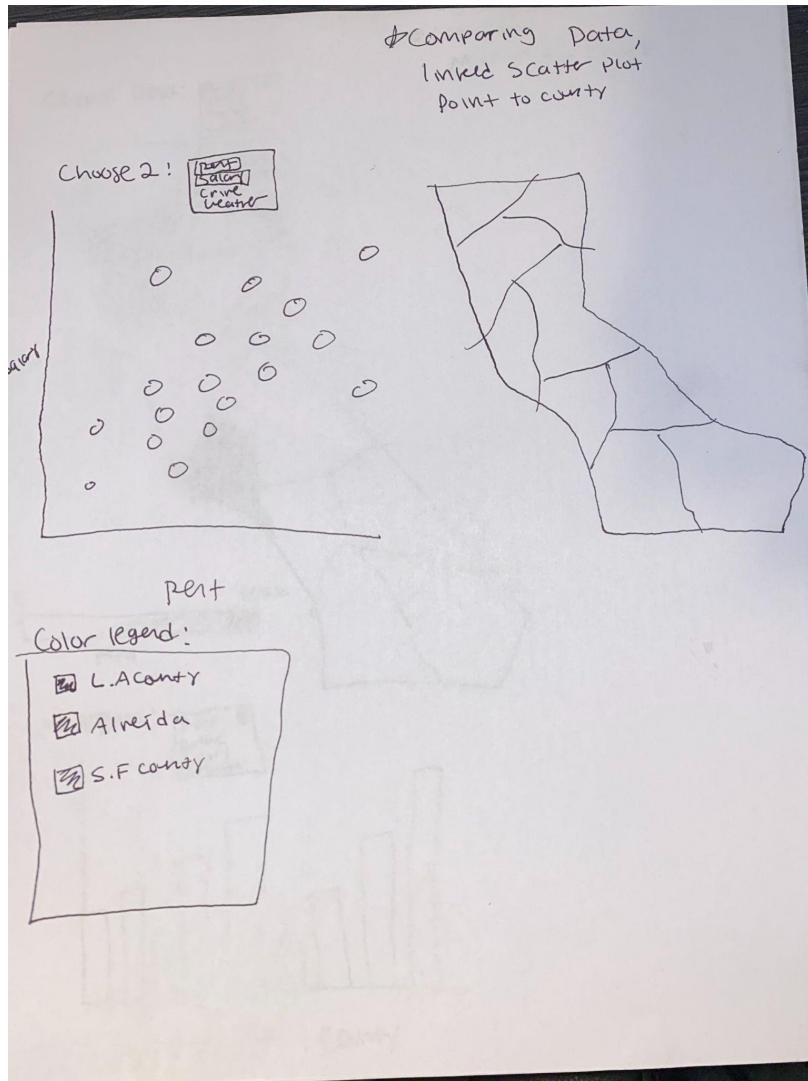
My second sketch shows the population data, and the reason that I picked the bar chart is because people can easily find out which country has the most population, and if user move their mouse to each section of the bar chart they could see the exact number.

The third pic is a treemap that will mainly display the salary data. They will have similar color but with different depth of colour. The darker the color the higher you earn, the brighter the color the less the earn. And the c1 c2 c3 Respectively represent the counties.



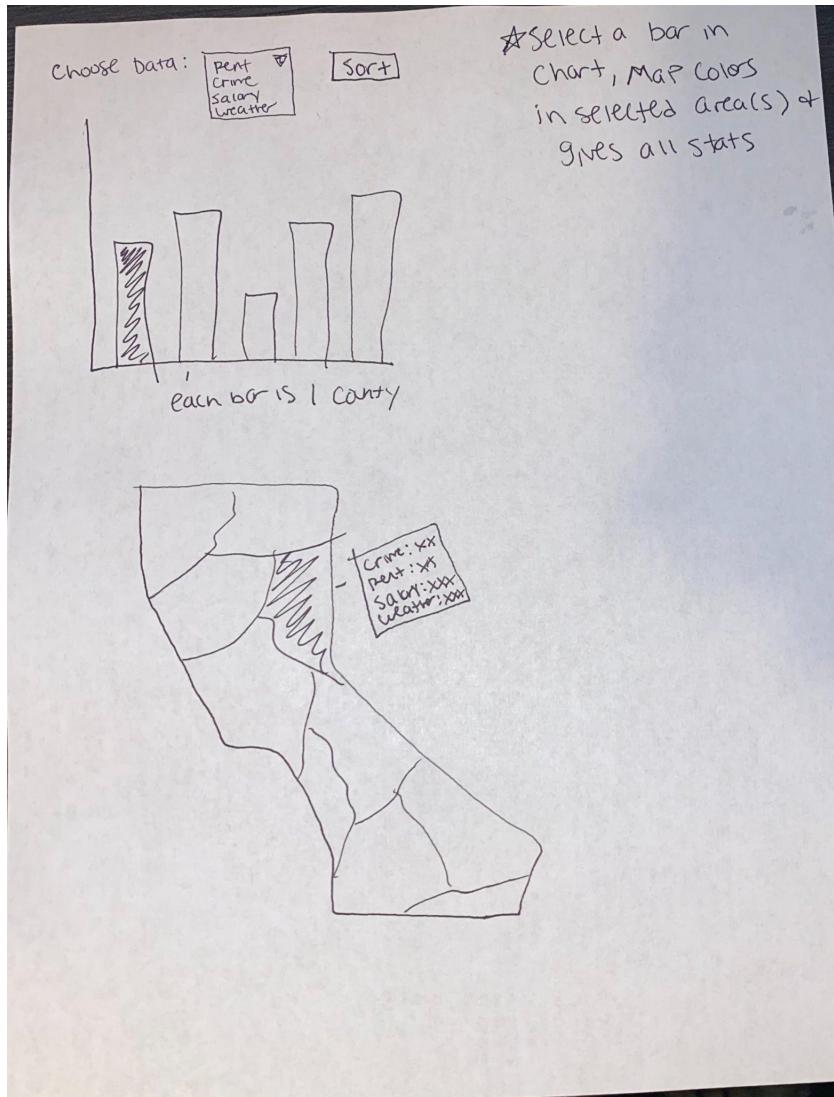
Creator: Jason Stitt

[FAVORITE #2] The bar chart in this vis (on bottom) uses lines for marks and lengths and position for the channels. This enables the user to differentiate between Counties and also view the value of each county compared to one another. The map uses area as a mark and I chose to color each county based on its relative rank in the chosen category. This provides an alternate view for the user to compare counties. The scale is shown below the map of CA, and the different shades represent different colors. This Vis would accomplish tasks 1 and 3, in which the user can view different data and also different categories in the same view by changing using the dropdown bar at the top of the page.



Creator: Jason Stitt

This sketch enables the user to compare two of the different categories of data to each other, using points as marks and horizontal/ vertical position as the channel, along with color to differentiate each county. The horizontal and vertical positions show each county's value on both the y label and x label, which are 2 different data categories. This would also accomplish tasks 1 and 3 in showing all data in one view, with the ability to change to all data using the dropdown bar.

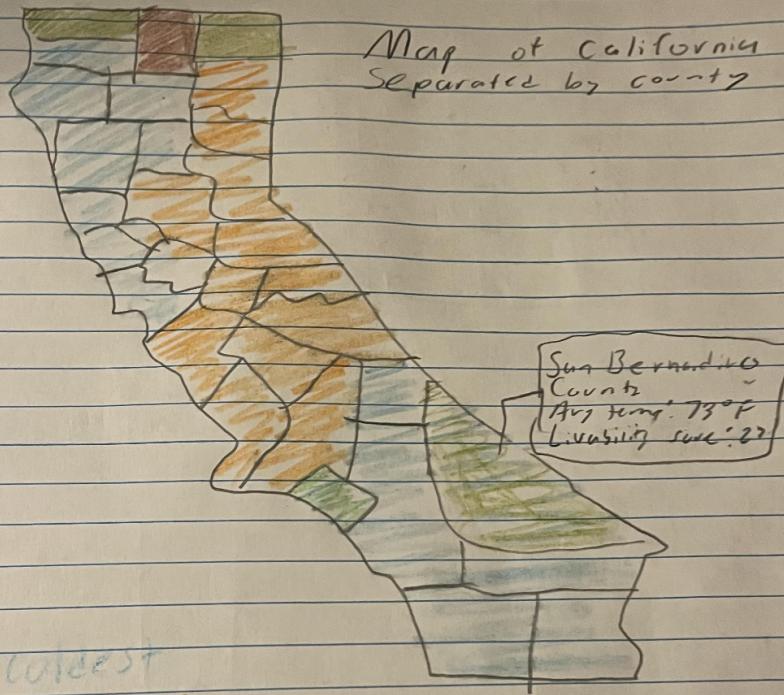


Creator: Jason Stitt

This is another bar chart linked to a map. Again, marks are lines and channels are horizontal position (to differentiate county) and vertical (actual value). The map will use color to show if a given county was selected by the user in the bar, and in this case, shows all the data for that county across the 4 categories. This accomplishes tasks 1 and 3, and enables the user to see all 4 data categories at once by clicking on a county. It would also allow them to sort the data using the sort button next to the bar chart.

Sketch I

Map of California
Separated by county



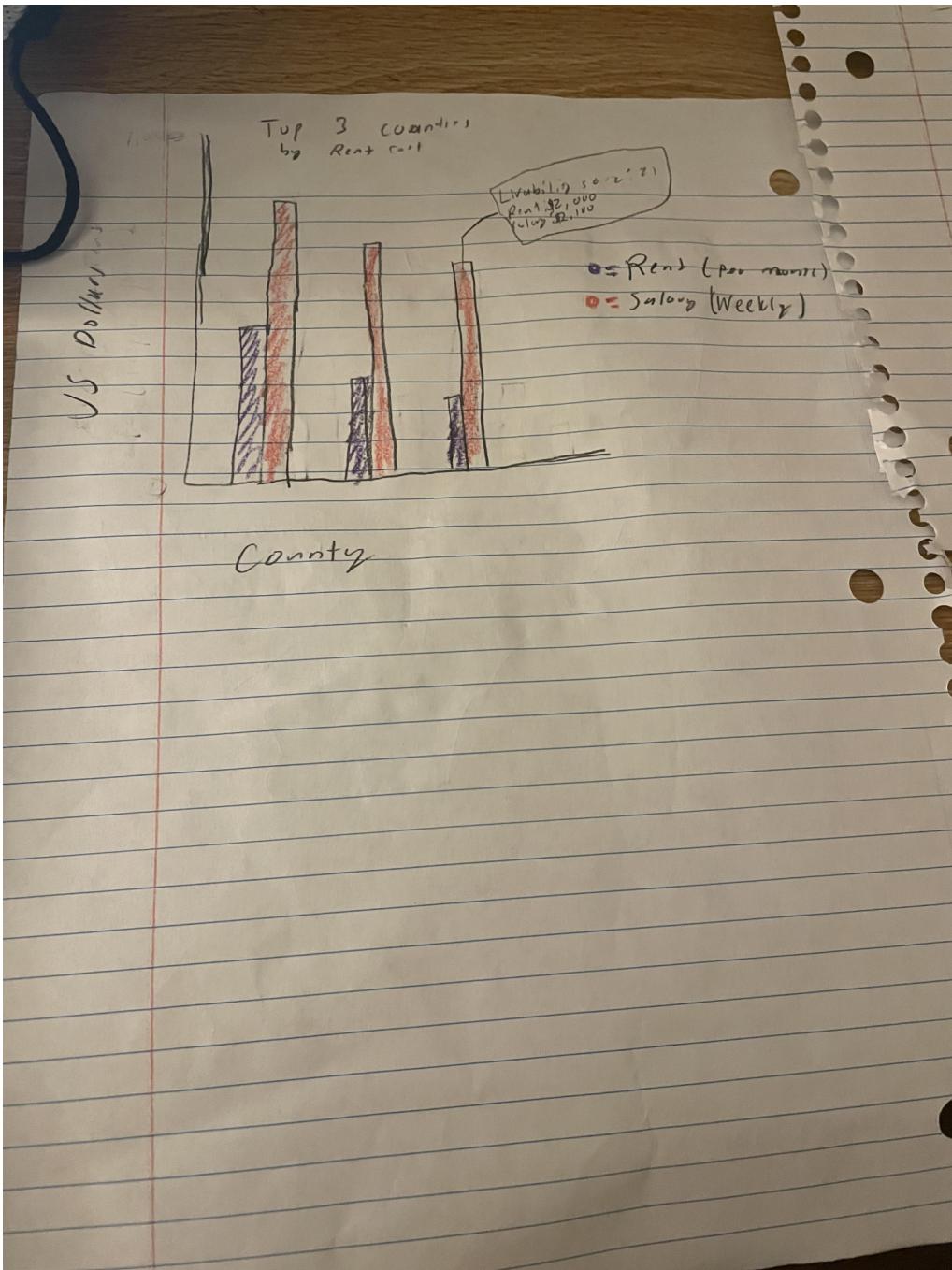
Blue = Coldest

Orange = Warmest

Creator: Dan Gilligan

[FAVORITE #3] Why I chose this design: In order to show the temperature data across countries, it is very useful to be able to visualize where the county is, so that you can see the relationship between temperature and the location of the county in the state. Also, weather data is usually overlaid on some type of map, so I figured that this would be the most fitting way to display the data.

What tasks does this address: This visualization accomplishes tasks 1 and 2. Task 1 is accomplished because the user can change which data is being viewed within the same view, as the user can view the livability score of a county if they click on it Task 2 because the viewer will be able to toggle an overall livability metric by clicking on a county.

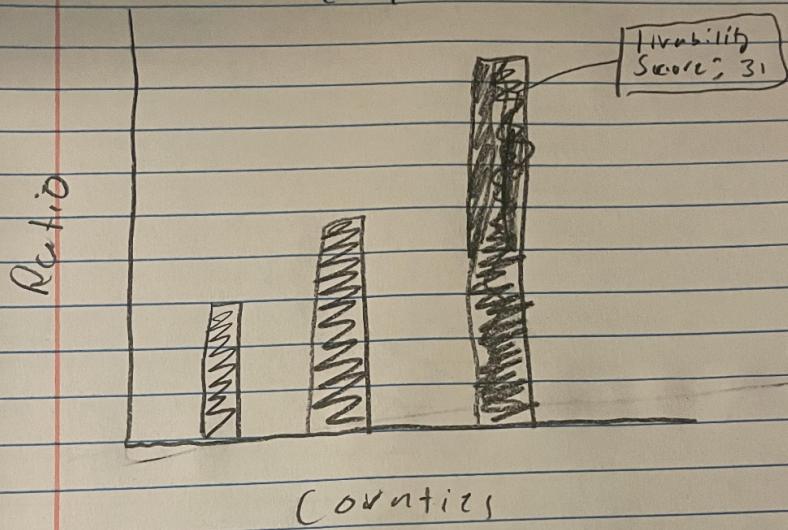


Creator: Dan Gilligan

Why I chose this design: This graph is showing the top 3 counties by monthly rent cost, and it is showing the weekly salary and the monthly rent for each of the top 3 counties. I chose to use a multiple bar graph for this vis because I wanted to show how the high salaries make up for the increased rent costs.

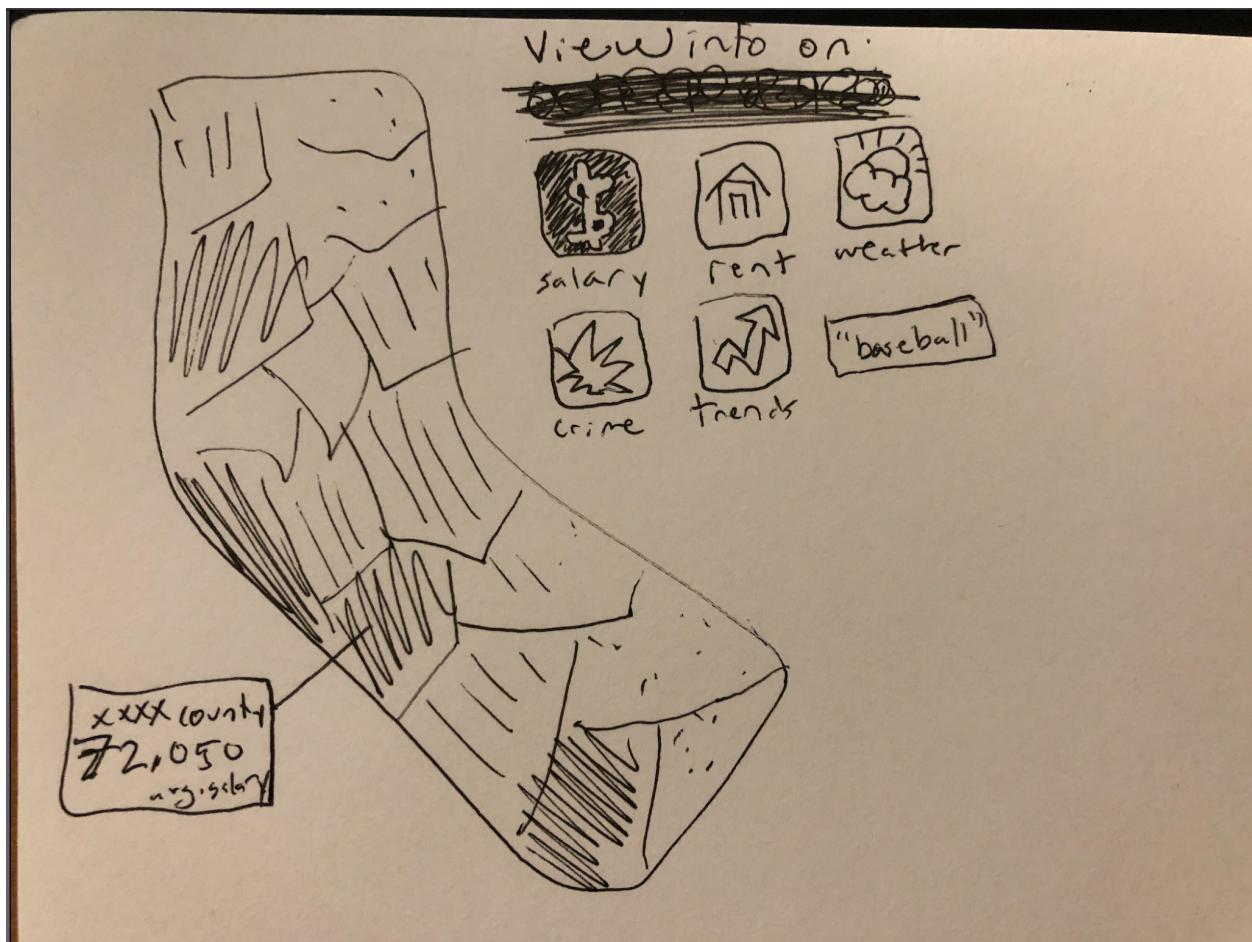
What tasks does this address: This visualization accomplishes tasks 1, 2, and 3. Task 1 is accomplished because the user can change which data is being viewed within the same view, as the user can view the livability score of a county if they click on the bar.. Task 2 because the viewer will be able to toggle to an overall livability metric by clicking on the desired bar. Task 3 is accomplished because we can view both the salary and rent data within the same view.

Lowest Rent to salary ratio
(TOP 3 counties)



Creator: Dan Gilligan

What tasks does this address: This visualization addresses tasks 1 and 2. For task 1, the user is able to change what data is being viewed by clicking on the desired bar so that they can view the livability index. Task 2 is clearly accomplished because the user is able to view a livability index for the desired county.



Creator: Will Kofski

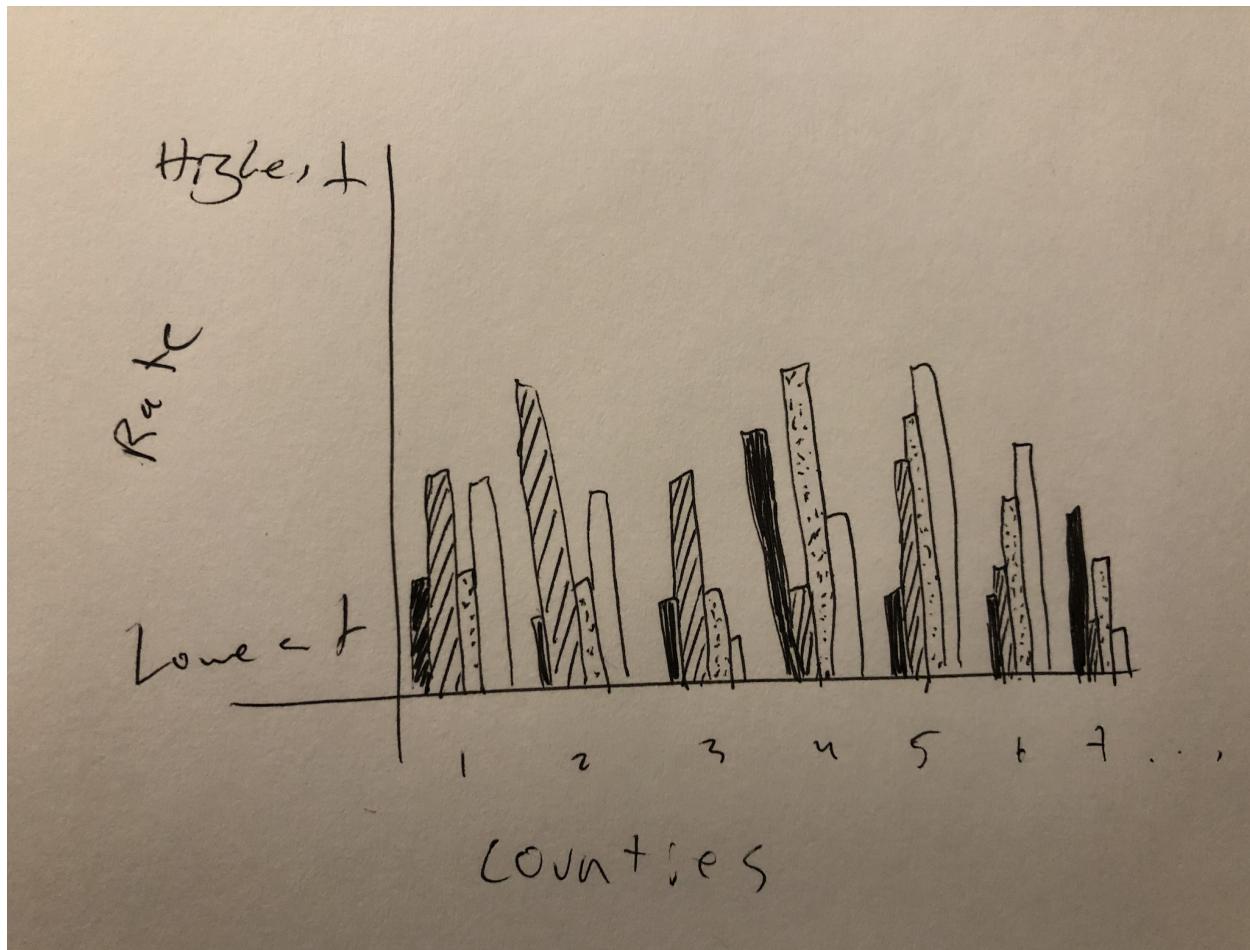
This visualization allows users to view varying levels of different types of data in the form of levels of saturation of each county within the state of California. When a user toggles the attribute they want to view, the color of the counties will adjust to varying levels of the attribute's icon color according to how high the county ranks in the attribute. This visualization addresses the first task listed on our proposal - users are

able to summarize the dataset for each attribute in a user-friendly way across the entire state of California.



Creator: Will Kofski

This visualization allows users to select an area on the state of California representation using brushing, and displays the "total livability index" for the counties that are within the selected area. Having a closely positioned bar chart that dynamically reacts to the area selection allows the user to still view what portion of the state they have selected, while being able to limit the information they can see and helping to scope into what they care about. This visualization addresses the second task we have listed, as it allows a user to compare the "overall livability" of the counties with each other, and more specifically compare the counties they care about.



Creator: Will Kofski

This visualization allows a user to view a select amount of attributes for all of the counties within California all at once within the same figure. There would be a slider at the bottom of the chart that allows the user to navigate horizontally across the different counties and focus on the information that appeals to them. Additionally, the different colorings of the different grouped bars for each county would indicate the attribute they are referring to. This visualization accomplishes our third listed task by allowing a user to view multiple attribute information within the same view.

Group Selection & Motivation

We selected the three visualizations that we feel best accomplish our three listed tasks, provide the best possible user interface and experience, and are the most realistic to implement within the allotted time frame. We want to offer a variety of dynamic visualizations and functionality that allows our data to be able to be explored and adjusted based on a user's own values and use-cases.

The first visualization we chose is a county-based map representation of California whose county colors vary depending on the prominence of a given attribute. The map uses area as a mark and implements color based on a county's attribute level. There is a drop-down that allows a user to view different attribute information within the same frame (Task #1), and within the dropdown as an available attribute is an overall "livability" index (Task #2). Users are able to view a summary of each of our datasets for the entire set of counties, with the dynamic ability of selecting which attribute summary they would like to currently view in order to discover new things.

The second visualization we chose includes a bar chart using lines and lengths to signify the prominence of a given attribute, in addition to a county-based map view that we defined above. The bar chart's information is not directly attached to the information being presented in the map view. With this chart structure, users can effectively compare information between different counties, and now additionally different types of information at the same time. This second visualization accomplished both Task #1 and Task #3, by allowing users to explore summarized information as well as enabling them to be presented with different types of summarized information at the same time.

Our final selection displays a more simplified presentation of only the county-based map view of California, again leveraging area and color to represent level / rank in different attribute datasets. In this chart, the map allows a user to hover over a given county in order to view a more specific definition of its attribute level. Making the visualization more dynamic and detailed leads to a more efficient and enjoyable user experience overall. This chart also addresses Task #1, by letting the user toggle the displayed attribute, as well as Task #2, by offering a total livability index as an option within the attribute selection menu.

Appendix D: Digital Sketches

Click on a county to view its data over time, or drag an area over multiple counties to compare their data:



change dataset being visualized

Crime Rate Data

Top violent crimes per 100 citizens (?)
in **California**



on hover over bar

Click on a county to view its data over time, or drag an area over multiple counties to compare their data:

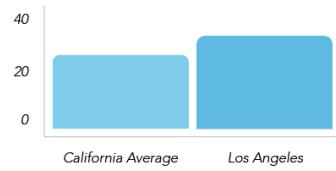


Crime Rate Data

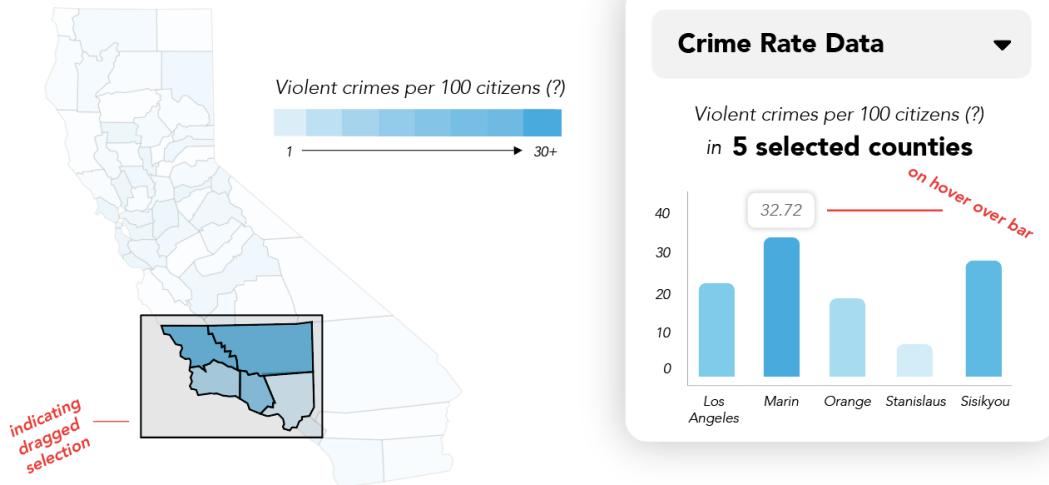
in **Los Angeles County**

28.31

Violent crimes per 100 citizens (?)



Click on a county to view its data over time, or drag an area over multiple counties to compare their data:



Our visualization addresses the first two tasks in our task table, as well as an additional task that we have added and chosen to prioritize over our initial third listed task. With this presentation, users are able to change between different dataset visualizations by county. When selecting a new category, the map on the left of the screen is updated with each county's data - changing color by density / level of the attribute. Secondarily, users can use our visualization to select an "Overall Livability" category from the category dropdown, which will display a ratio index calculated from combining the county's data from each category. Finally, our visualization now allows users to compare the county specific data to an average across the state of California for the category. When selecting an individual county, the right panel will display the county's data for that category compared to an average of the category across the entire state. We added this task to our table and replaced a previous task where a user would be able to view multiple categories of information at the same time. We decided to prioritize the ability to rationalize individual category information against a state average and the ability to easily understand what category of information the user is currently viewing as opposed to allowing the user to view multiple categories of information at the same time. We offer the averaged "Overall Livability" index as a way to compare multiple categories of information at the same time, and did not want to add the complexity of the user having to figure out which categories of data are represented by a given color. For simplicity's sake, only a single category of information and single color will be available on the screen at once.

Appendix E: Reflection

Dan Gilligan --

The first thing that I thought we did well on in this project was keeping the main ideas of our visualizations simple, but still allowing them to convey significant information to the user. I also feel like the graphics were kept clean and to the point, and followed the design rules of thumb very closely, by trying to keep the visualizations easy to read and showing the main idea of our project. The third thing that I thought I did particularly well was starting the project milestones as early as possible, which gave me more time to make sure that my work was up to par with what was asked of us. In addition to the things that I felt I did well on, there were a few things that I would definitely change if I were to do this project again. The first thing I would do differently is make sure that I read the assignment description and rubric thoroughly, as there were a couple assignments where we got points off that could've been easily avoided if the assignment description was read more carefully. Also, I could've read over the code more closely and added more descriptive comments as needed. Some of our comments and documentation could definitely have been improved, and this is something that can be fixed easily if I just went through and did a proofread of the code. Lastly, instead of doing the project just with California, I would have liked to find data sets that pertain to the entire United States, and create an interactive map of the whole country. While this may take much more time, this visualization would have a larger audience than our current one and, in my opinion, be more interesting to view and interact with since it is encompassing the whole country. The most valuable lesson that I learned from this project would be that before starting any big project, whether it be individually or with a group, you should always come up with a plan before implementing and coding the final product. I think that the project milestones and design sketches really helped with this, as it broke the project down into smaller and more manageable steps. While this project has not only improved my ability to brainstorm and design visualizations, it has also made me a much better worker in a group. I think that overall, our group communication has been great, and we are always on the same page about where we are regarding the current project milestone. One thing that we could improve on as a group would be talking about who does what for each pm, as it could get confusing regarding what part of the assignment each of us should do.

Tianyu Lin --

The first thing that I am super satisfied about our project is that we pick a super interesting topic. Our topic is factors that are related to house prices in CA, and we are doing it by countries. This topic is like the intersection between my two majors. The second thing that I'm satisfied with about our project is that we are doing all the work by counties. This is making our graph super clear and user friendly, users can select multiple counties to see the rank of each factor. For example they could see which counties have the highest crime rate etc... The third thing that I am satisfied with is the way we organise the graph. Our graph is super clear and easy to read. There is no redundant information and users can see the thing they need. In terms of things that need to be done differently i would say we could add more states for now we just have CA and we could add more states in the middle or east coast of US, this will make our project much more comprehensive and will be more readable for users from all over the states. The second thing is that we will definitely read the pm description carefully in the future. We lost a lot of points because of not following the pm description. The third thing that needs to change is the way that we select counties. Now we need to use the mouse to select a range of counties and this is very annoying because you can't pick the countries that are not close. We need to add another way to select counties. The most valuable thing that I learned from this project is that we need to keep our visualization clean putting too many things on the same graph is not a good thing. We need to know what we want to say with this graph and make our points clear. A good visualization can help you to get the information that you need without spending a lot of time reading the graph. Overall speaking we did a good job on group communication, we split up the work during the class and if any of us had a problem during the process we would ask questions in the group chat. We have a group of messengers to deal with questions. One thing that we could improve on is we could start the PM earlier. We like to start everything at the end of the due day and this is not a good habit, because we don't have enough time to go through all the stuff and double check if we miss any important information. This could be fixed by starting everything earlier and don't push all the thing to the end of the due day.

Jason Stitt--

I feel that I communicated well with the group in order to give my suggestions and input. I feel I also took in everyone else's suggestions and as a group we incorporated everyone's ideas. I also think I was good at staying on top of the due dates for the project milestones. In terms of things I could've done better, I think I would have thought a little harder about the project goals, because our vis was relatively hard to implement. I also would have maybe chosen 1 or 2 statistics to focus on instead of the 5 that we did to make things more simple for a viewer. Finally, I feel I could have, along with the group, started implementing the visualizations earlier in the week instead of cramming to get it done by the milestone deadline. I think the most surprising thing I learned about visualization design is how difficult a map is to implement with brushing and linking. I feel that when I looked at visualizations before this project I didn't understand the amount of effort required to create them, and I can appreciate that more now. I think our group communicated who would be doing what very well throughout the project. We all understood our roles and felt comfortable giving input as well. If there was one thing that could be improved on, I would say we could have split up work in a more clear and equitable way in order to avoid a case where one person contributed less than another on any given pm.

William Wallace

In this project I believe that we did well in communicating as a group to set out our goals and vision for the final visualization. Another strength of our group was the segmenting of our responsibilities of our team according to our strengths. Some members of the team were focused on writing the code while others worked on the project milestone write ups and such. One more strength of our group was the ability of our team to manage time well and maximize our time spent on the project inside and outside of class. A few things that I believe our group could have done differently would be choosing a more concrete objective. We were unsure about the specifics of our visualization up until the final weeks of the project: it would have been helpful to have a solid idea of what we were aiming for from the beginning. Another thing our team could have done better would be the simplification of our visualization. We have many interactive pieces to our vis, which has ultimately proven challenging to implement. One final piece that we may have improved upon would be picking a more interesting data set to base our project on. At times we contemplated the usefulness of our vis and the potential pieces of data our audience would want to see. I think the biggest lesson learned for me from this project would be the importance of having a clear vision and of the outcome of a data visualization project, complete with proper scoping and prior examples of very similar up to date documentation. If our group had compiled this information more thoroughly, we would have been aware of many of the potential challenges that we ended up facing. With regards to communication within the group, I believe that we have been engaging well with one another via group chat. I think it may have been better if we were all a bit more proactive with our communication as compared to reactive.

William Kofski

The first thing I think we did well was ideating and designing a very clean visualization that was complex to build but easy for our end-user to use. Our end product is easy to understand and seek out information from. Secondly, I think we did a good job of delegating work throughout the project. Each of our members contributed to the project a great deal, and we ended up with a better end result because of this. Finally, I think we especially did a good job soliciting help from each other when in high-pressure situations and things needed to be completed. We did not ever get frustrated or angry with a member, but rather leaned on one another to accomplish our goals.

One thing I would do differently is choose a visualization that is easier to configure and manipulate than an svg-graph of California by county. The odd shapes of the counties were a nightmare to handle brushing and linking with. A second thing I would do differently is spend more time thinking through the structure of our code. Our end visualization works, but our code structure and efficiency could be improved a great deal. Finally, I think I would have chosen a different topic. While our visualization does provide a solution to a unique problem, I would have liked to do something more explorative in an area of my interest (art, movies, music, etc.) as opposed to California county data.

The most valuable lesson I learned is to spend as much time as you can designing, not just to have the best looking product but to have the best looking product that is not also inherently hard to build. Oftentimes, it is easy to want to build something because it looks cool, when it is better to build something based on its utility and the time you can build it in.

I touched on this earlier, but I think in regards to communication the group did a great job working with each other, staying flexible and not getting frustrated when people needed to do the work they had been assigned. I think potentially, we could have done a better job being proactive with the assignments and not letting work build up on us, but we all have been very busy throughout the year with intensive comp. Sci classes, so this is to be expected.