**INFO 4310/5310 HW 2 Report**
Marcus Fang df375

**Describe the data you chose and identify specific insights/use cases for that data that will align with your chosen interactions:**

When choosing with the datasets I found this one is more difficult than I thought. I personally already have a list of tidy datasets and referenced the provided reference link in the homework PDF. However, I found some of those datasets are not suitable for this homework because they are either not complicated enough to support a direct-manipulation style of data visualization or it would be difficult to tell a story.

I chose the global happiness index dataset from Kaggle to visualize because I have been following up with this dataset for a few years. Since I would like to introduce the pop-and-peek style of data visualization on a scatterplot(view a general scatterplot at first, then hover on each datapoint to see the details) and this dataset would align with my selected interactions.

Alternatively, I would like to select the dataset from datasf but found detailed geographical data might not be suitable for this project as geographical data are usually more information-heavy but relatively difficult to interact with besides of selecting and filtering(which filtering is not recommended in this hw).

**Provide storyboards that outline the interactions you will design for your dataset and justify why you are using those particular interaction**s:

Here's the storyboard that initially I was brainstorming, and it looks slightly different to the version submitted. Initially I was still fascinated by the concept of dashboard, but later found out the dashboard might not be the best suitable way for users to 'directly' manipulate the data.

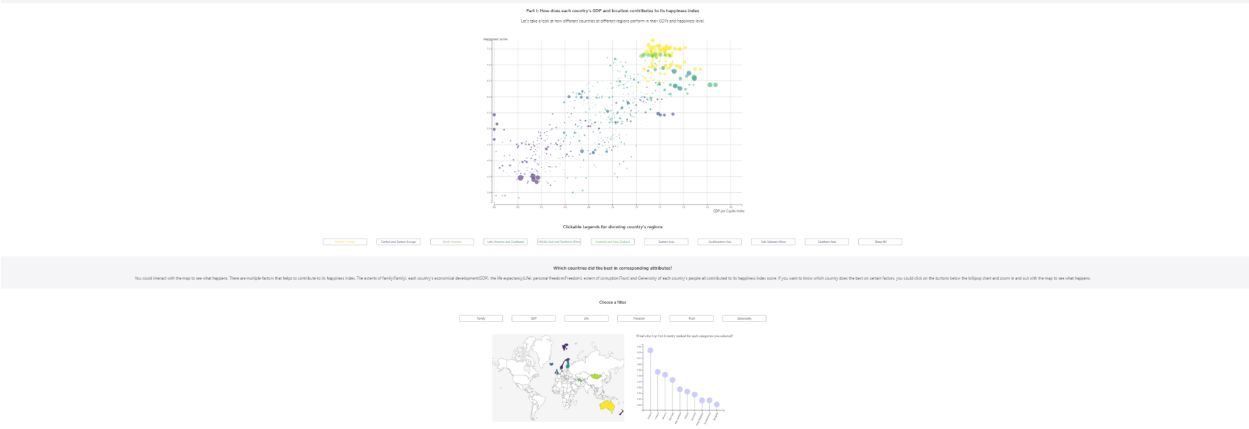
Later I was interested in the 'peek & pop' interactions from iOS 3D Touch gestures and found that it was somehow similar to the recent lectures. I then adjusted my storyboard and it looks much more like the final prototype. Operations include mouseover and clicking each data point for the first graph to see the graphical details, then clicking on the legends to see the data points from each state in detail. Those peek & pop operations would help the users observe the dataset in a higher, generalized level to a detailed level of each datapoint if they need.


**Briefly describe your final interactive visualization application, including a screenshot**

The final implementation of this visualization application includes two major sections with 3 figures. The main visualization would be the scatterplot on top, which is an interactive plot that shows the relationship between the GDP per capita and the happiness index for a country. The users could directly manipulate the dataset by either ‘peek’ into the detail for each datapoint by hovering on or group and select different data points by clicking the button below, where clicking each one would only show up the datapoint from a specific region in order to reduce the clutter and the cognitive load to the users. Additionally, each button would act as the legend for this visualization.

The second part includes 2 figures, an interactive world map that supports zoom in/zoom out features so the users could view the score of different factors that contribute to the happiness index for each country. On the right it’s a barplot that shows the value of the indices that user selected on the top, to show the 10 countries that have the highest scores on different factors.

The screenshot of the visualization is attached below:



**Step back and think about issues or trade-offs associated with the interactions you developed, and how you might alleviate those (or whether they are unavoidable).**

The first tradeoff I think is the implementation of a time slider. Initially this dataset includes all the data from 2010 to 2019 and could support a time-slider feature if I chose to implement it. In the homework instructions it said students should consider carefully on whether implementing the time slider would be beneficial or not. I thought about this for a considerable amount of time and later decided to not add it. Time sliders are convenient for users to see the trends among the time but unless I have a line chart that marks the current year on the visualizations, it's difficult to compare two datapoints at different times. For example I can generate 2 different scatterplots to show datapoints in 2014 and 2019, but it would be difficult for users to compare them directly unless I put all figures on the screen at the same time.

Another tradeoff I was thinking of is the colorset for this visualization. Currently I'm using the commonly used viridis colorset, which is commonly used in information visualization and is one of the 'good' colorsets recommended by Professor Rzeszotarski during one of the INFO 3300 lectures. However, the current colorset selection did not distinctively tell the difference between each continent. Viridis colorset has 4 main tones, yellow, green, blue and purple. It's difficult to distinguish some of the blue/purple colors and 4 main color tones might not be enough for representing data from 7 continents. One example is the colors used on Western Europe and Central&Eastern Europe. Although all 2 regions above are considered in Europe, they have a relatively large gap in GDP index and it would be difficult to tell the

difference between those groups if I used similar colors. I've reviewed the current colorsets and found the viridis one is better than others like schemeCategory10 (which leads to rainbow colorsets), but I'm still not satisfied with the current performance of the viridis colorset in this homework.

Besides of thatI could add a zoom in/zoom out and minimap feature for the scatterplot for better interactivity. However due to multiple constraints I found it might be difficult to add it during this time before submitting the homework. I'll continue exploring that and refining my visualizations after the submission.

**Briefly outline the development process of your tool. Explain how your visualization/interactions changed between storyboarding and final implementation. Comment on any trade-offs or design choices you had to make while developing**

During storyboarding my data visualization looks like a dashboard with drop-down menus. Previously it looked more like an interactive map on a dashboard. I revised my design and order to add more direct manipulation of the data. In retrospect, my initial storyboarding did not have many direct manipulation elements and I could definitely add more of this in the final, more refined product.

Another tradeoff I have to make is to reduce the search bar initial proposed in the storyboard. Searching data points in the search bar is not a direct way of manipulating the data, thus it could be removed(although the efficiency for quick searching is decreased in retrospect).

**Identify how work was broken down in the group and explain each group member's contributions to the project. Give a rough breakdown of how much time you spent developing and which parts of the project took the most time.**

In this project, I solely finished the project on my own and did not find a teammate.