Process Book

Background and Motivation

Medical data is often visualized using charts and graphs. We looked around a number of bio-medical websites and we could only find static images with no or very little interactions. For the physicians/ researchers this is a bottleneck as they are not able to modify the graphs/data for context based analysis. Also the users don't usually have the knowledge of databases or even the datasets which are used in visualization. The objective of this project is to provide an interactive visualization system for the users using which they can visualize, change and compare different datasets.

The primary motivation for doing this project is that we didn't find any good visualization tool online for bio-medical domains. We also discussed with the users about their expectations from a visualization application. The general feedback we got from this exercise is that the existing applications are heavy and involves a learning curve for which non-technical users are reluctant. Our application will be a light weight visualization system with nominal learning curve which enables non technical users to visualize and personalize their data. We will also incorporate basic utilities on data like transform, import and personalize in our application with simple user interface.

Project Objectives

Most of the visualization tools in the medical domains lack ease of use, interaction and intuitiveness and personalization. The main objective along with learning and applying programming languages is to understand and improve the aesthetic and perceptive properties of the visualization. We chose to do this exercise with non-technical users (medical domain) which give us a better insight of how the visualization tools should be.
We classified our objectives into 2 groups; Data based objectives and Aesthetic objectives.
Data objectives:
1. Context based data selection: We will provide utilities to make context based data selection. The users will be allowed to select the filters of their choice.
2. Comparison: Data comparison views will be available for the user.
3. Personalized views (extended objective): Users will be allowed to personalize their views and select the view of their choice.

Aesthetic objectives:
1. Bring intuitiveness in the visualization: We will research over existing methods used in other visualization tools to bring self explanations. We will do our own analysis over the effectiveness of visualization and also consider publically available feedbacks on designs.
2. Ease of use: As stated above our application will have a nominal learning curve. We will ensure that the user interface is easy to learn and use.

We hope to cover not just the technical but cognitive aspects of how the visualization tool should be like. We will learn multiple designs while researching on our design and also the user interactions. We also plan deploy it and get feedback from target users. This will be a good learning exercise and will give us more insights about an effective visualization tool.

Data

For this project we are visualizing flu trends of United States for a period of 2004-2005 to 2013-2014. The data is available on 'Google Flu Trends'.

Data URL : https://www.google.org/flutrends/about/data/flu/us/data.txt

Data description

Google flu trend for United states contains number of flu cases for a period of 11 years. The data is collected on a weekly basis.

For a particular week the flu cases are recorded for states and Human and Health Services regions.

State data: The cities with comparatively more number of flu cases are stored separately and the rest of the flu cases are stored under the State.

Region data: The regions are divided by HHS, a region contains a set of states.

Data Processing

The unwanted text from the data is removed manually.

Visualization data is converted into a master CSV file : 'flu_trends_data.csv' which contains data for all the dates , states and regions.

We created different data structures to save the datasets for different views. The primary datasets are:

- `seasonsData` : stores season wise values for all the states for the selected year.
- `statesData` : stores data for individual states and it's major cities.
- `regionsData` : stores data for given regions
- `yearsData` : stores year wise data for all the states.

Exploratory Data Analysis

There are 2 main aspects of the data that we wanted to show, comparison and multi dimension data. We thought of several designs to show comparison. We tried multi series graph for showing comparison, it gave a very clear intuition of the design. But as the number states grew it became crowded. We decided to go with pie-charts and limited the number of state selections to 10.

We explored our data samples in Microsoft Office charts to come up with the visualizations we wanted to show. We tried different views and also took feedback from others until we were convinced with the graphs.

Design Evolution

The following images describes the design evolution process.

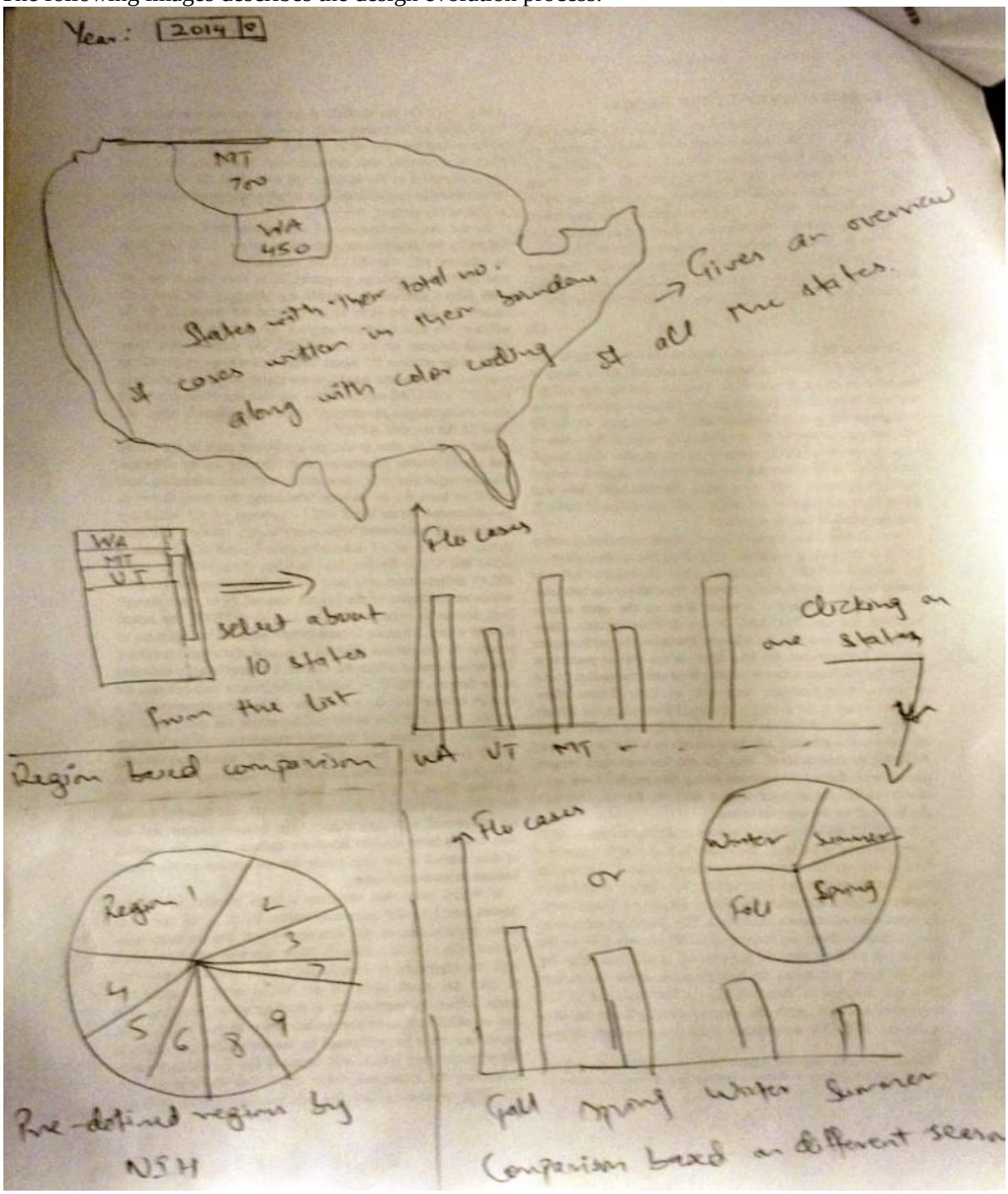


Figure 1: First Version of Design

Figure 1 shows the first version of the design. We started with basic blocks of visualization and kept on modifying keeping ourselves in place of the user. The following 2 figures shows how we landed up with the final design for this visualization.

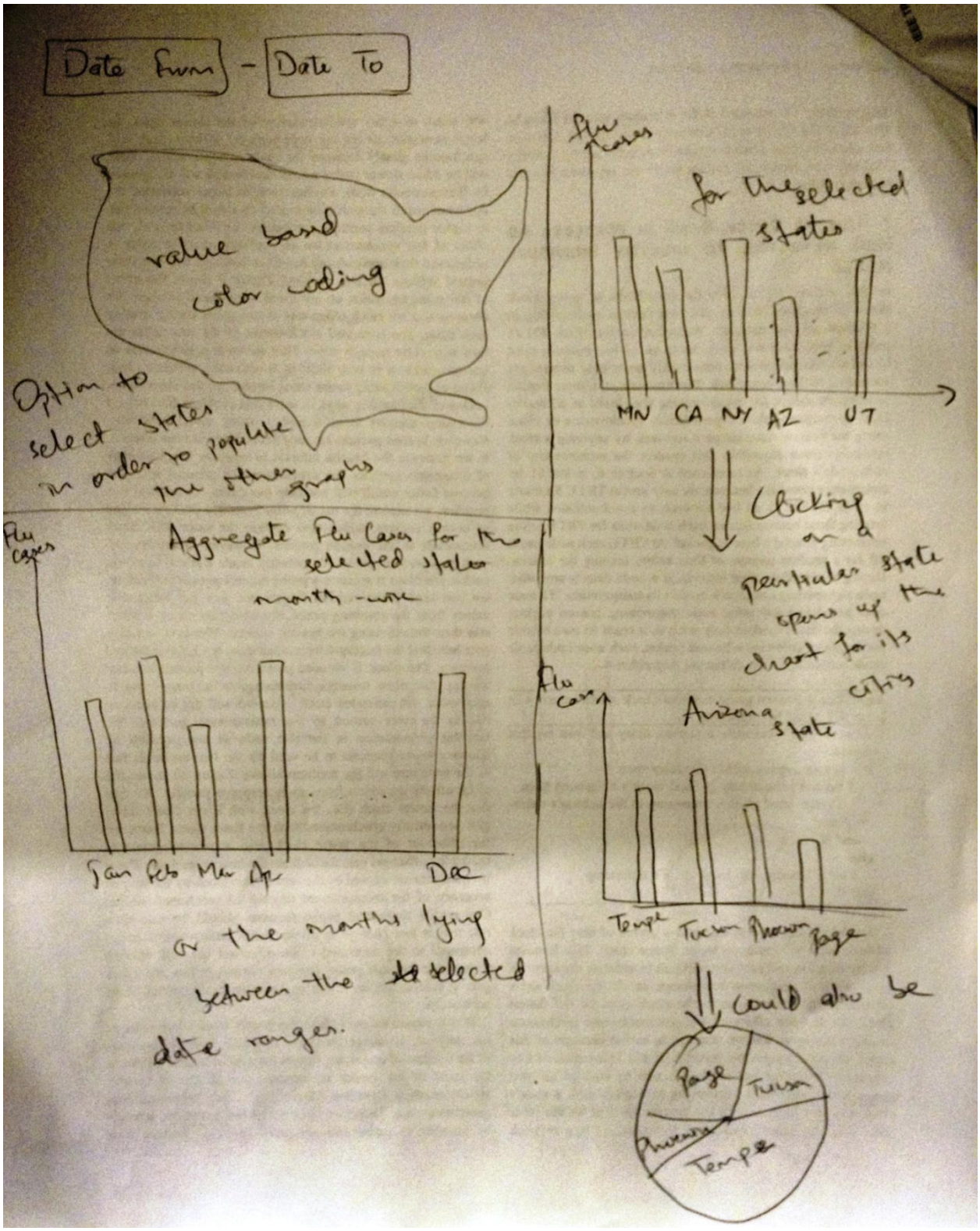


Figure 2: Second Version of Design

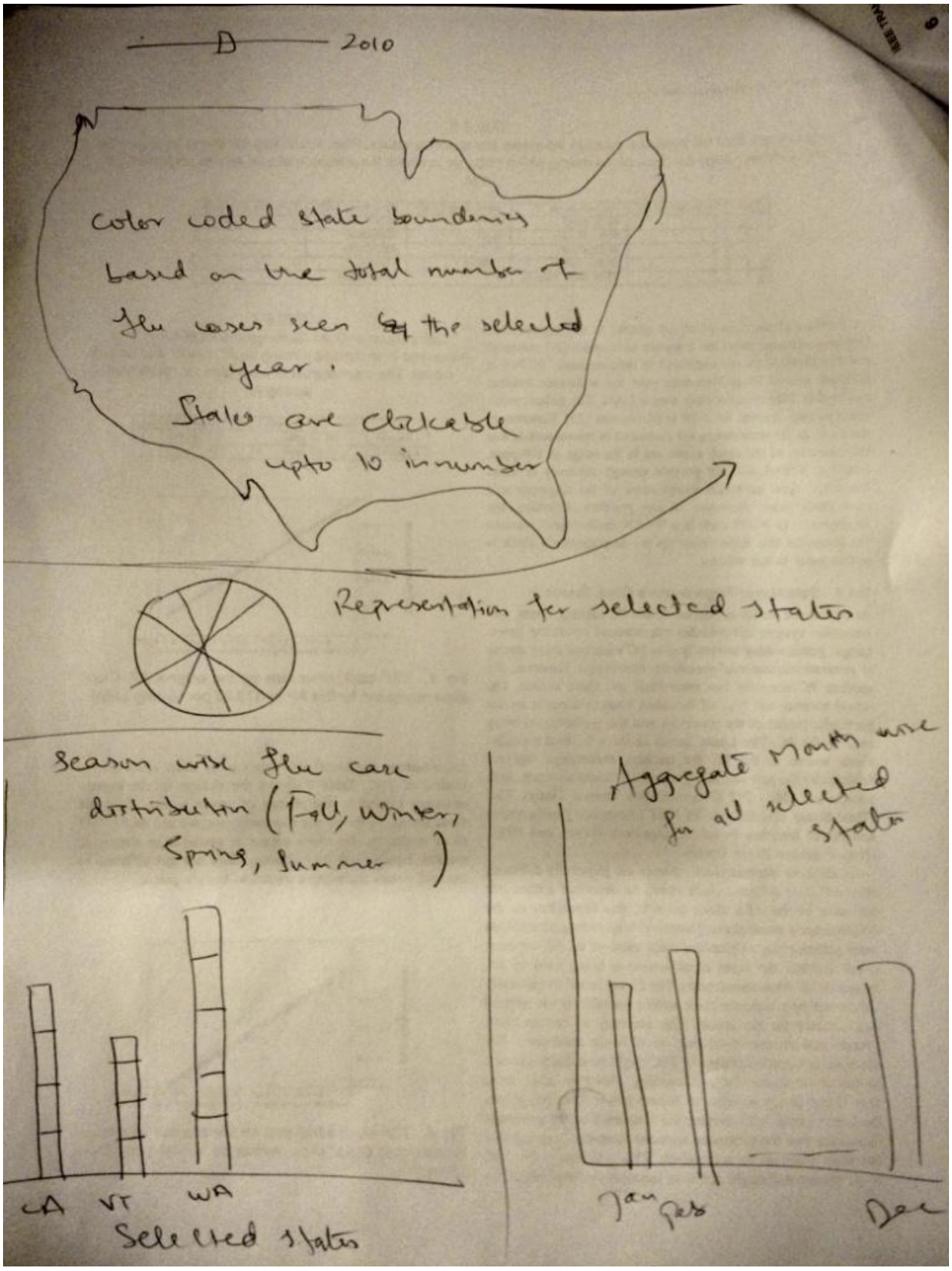
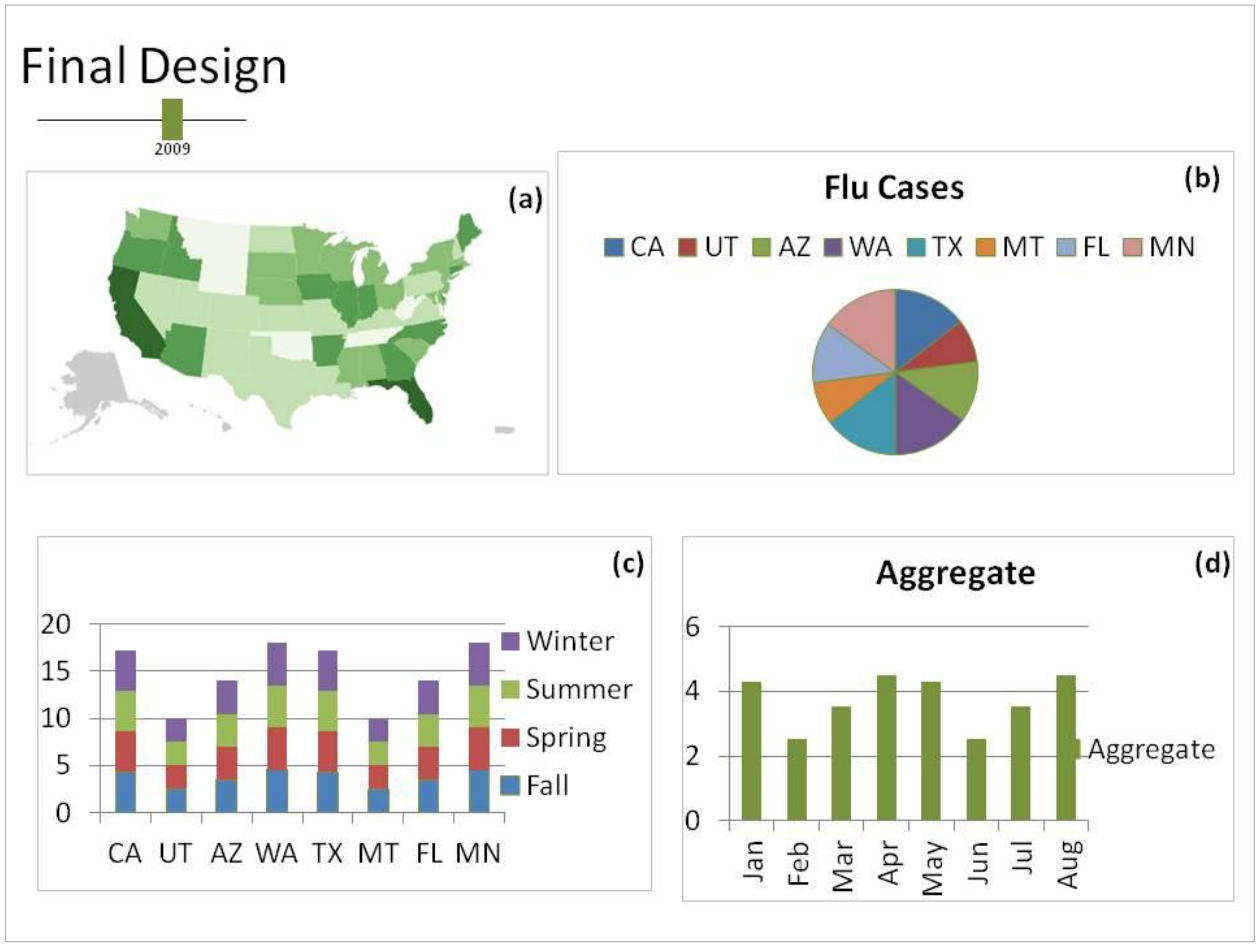


Figure 3: Final Design rough sketch

Selected Design



Design description

a) Map view with states ordered on the flu density for the selected year (default: 2003). User can select multiple states from the map.
b) This view shows a pie chart which better compares the flu cases in the selected states for the year selected using the slider.
c) This view shows the aggregated flu activity during seasons for the selected states. The Stacked graph shows the comparison of number of cases.
d) This view explores the data for the months of selected year. The flu activity of all the selected states is aggregated on the month and displayed in a bar graph.

Final design with extended goals

We implemented the mandatory goals based on the selected design. For the extended goals we implemented two more views:

1. Flu Trends: States Flu Pattern- This view shows the pattern of flu trends for all the years from 2004 to 2014. A line chart with different color codes is implemented for showing this comparison.
2. Flu Trends: Regional Report- Regional report contains the aggregate values for the selected year.

Implemented Design

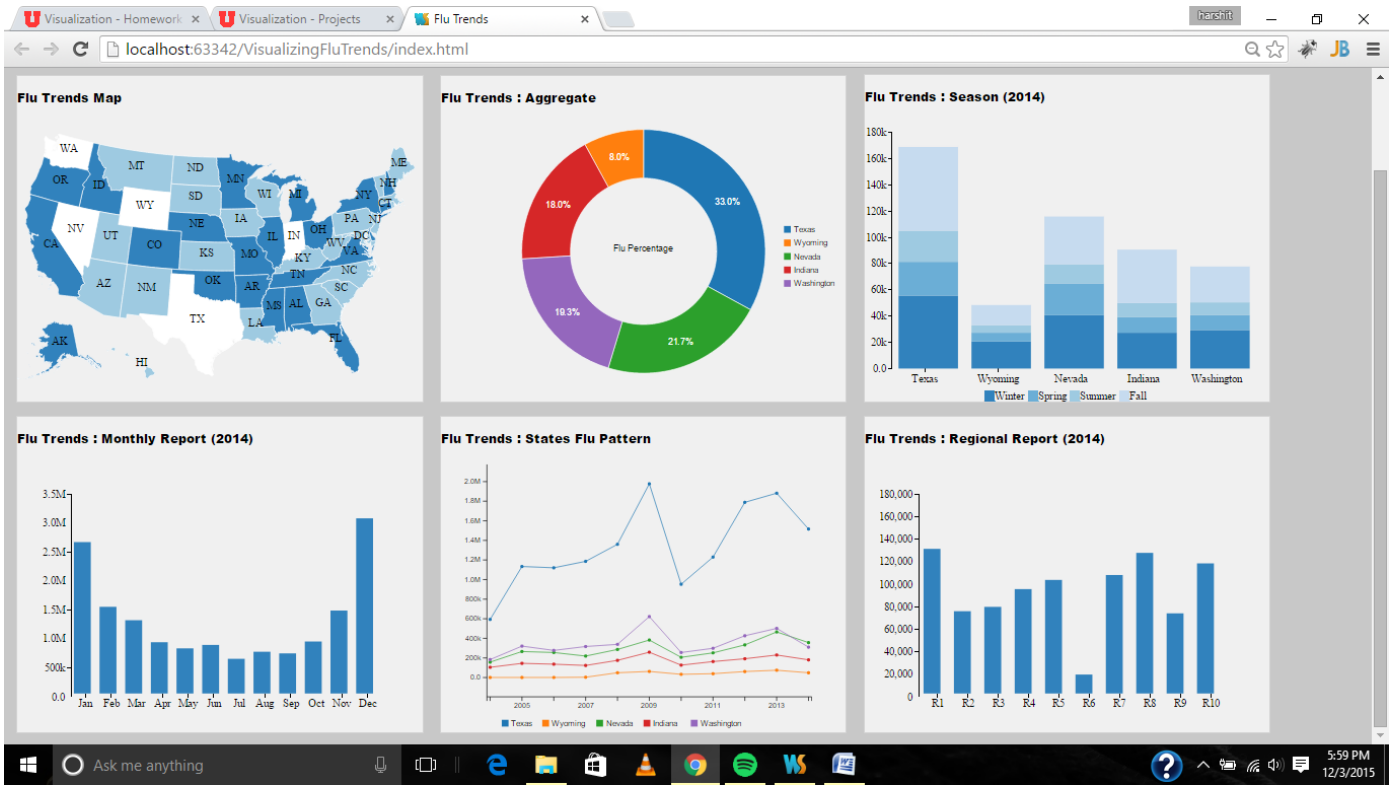


Figure 4: Implemented Design

Figure 4 shows our implementation:

<u>Flu trends map</u>

This view shows the map of United States. The states are assigned colors on the number of flu cases in that state. The map is interactive wherein user can select/unselect the states to populate other views for comparison.

Limitation: The comparison views cannot accommodate so many states. So we have limited the number of selections to 5. If the user wants to see the different states, he will have to unselect older states and select the desired ones.

<u>Flu trends aggregate</u>

This view compares the aggregate flu cases of the selected states and displays as percentages on the donut chart. The view is dynamic and controlled by the selections on the map.

The legends on the chart can be used to highlight the section on the mouse over event.

Flu trends seasons

Seasons chart is a two dimensional comparison in which the number of flu cases are compared for the selected states and seasons for a particular year. The layers on the chart shows the flu cases for Winter, Spring, Summer, Fall. The chart is controlled by the events on the map and the slider.
Cities: The bars on the view can be clicked to see the cities of that state. User can go back to the previous view by clicking on any of the bars again.

Flu trends: Monthly report

This view shows the total number cases for all the states in United States on a monthly basis for the selected year. The view is controlled by the slider. User can view the number of cases using the tooltip.

Flu trends: States Flu Pattern

This view shows the pattern of the flu cases for the all the years from 2004 to 2014 for the selected states. The view is controlled by the selection of states on the map. The legends on the view can be used to highlight pattern for that state.

Flu trends: Regional Report

The regions are sets of states selected by HHS. The view shows the flu cases for the regions in the selected year. The states that belong to that region can be seen on the tooltip. This view is also controlled by the slider event.

Implementation

This section explains the interactive elements and how the views transform on calling the events.

Year selector

This is implemented using HTML5 range selector. The default year selected is 2009 and the user can slide to select the years from 2004 to 2014.

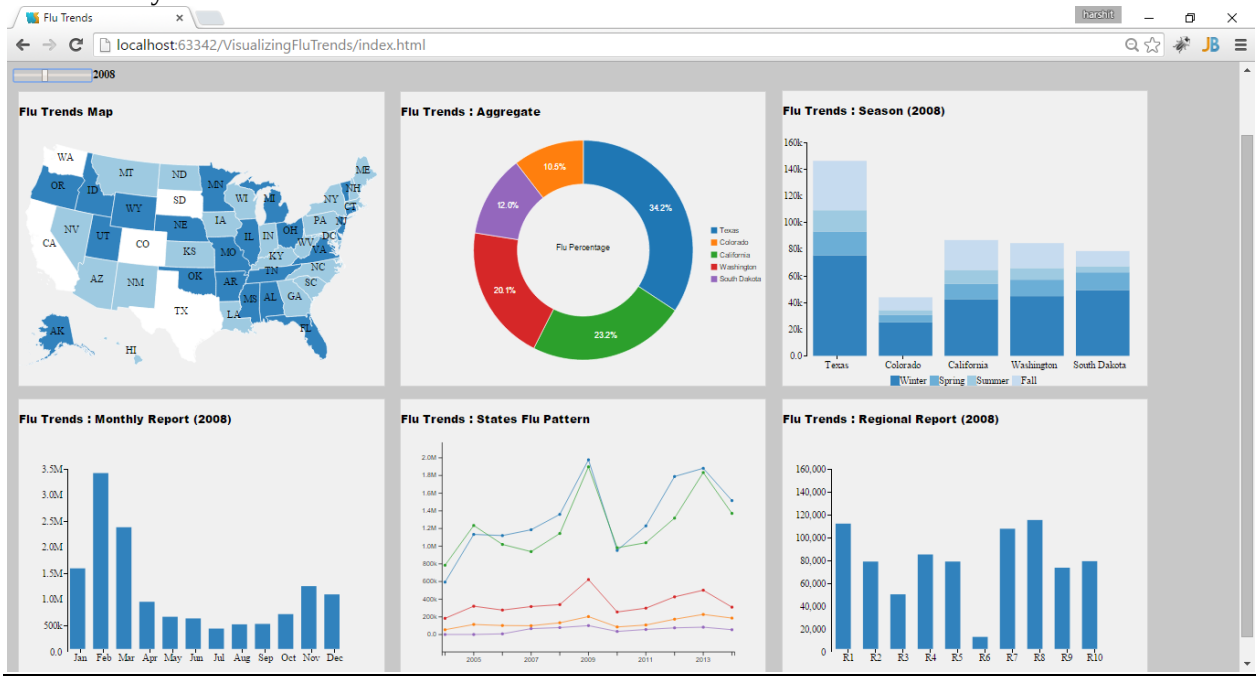


Figure 5: Views for the year 2008

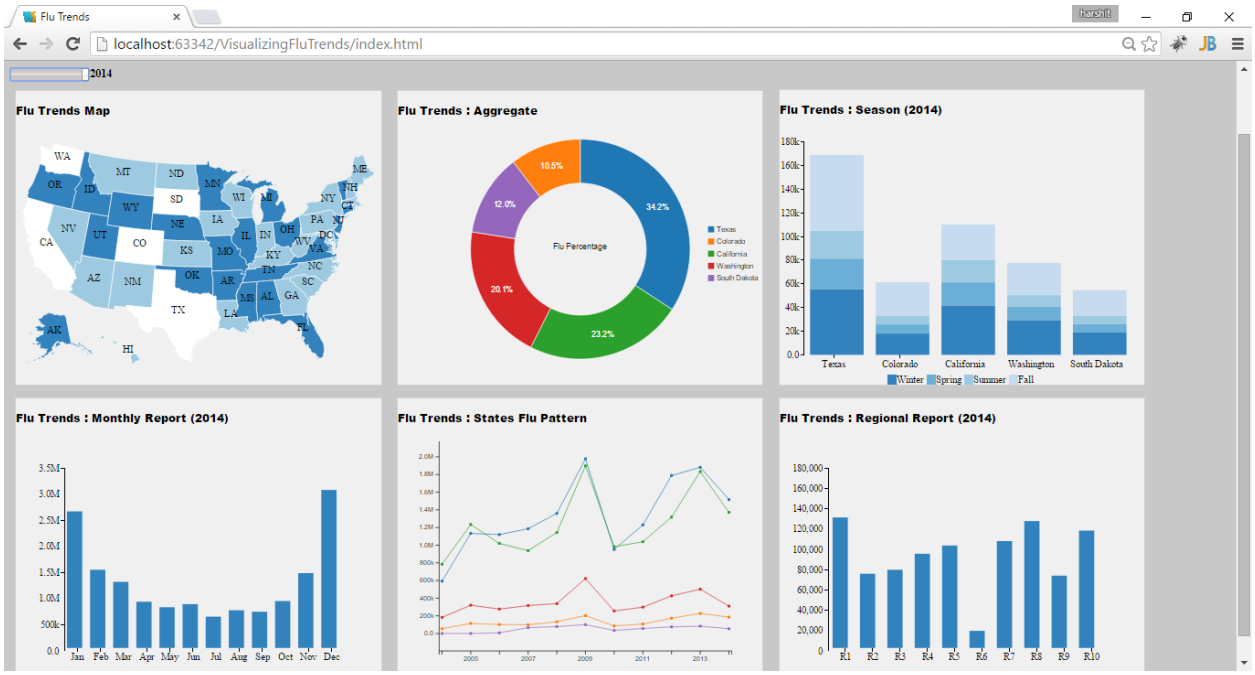


Figure 6: Views for year 2014

Figures 5 and 6 show the visualization for years 2008 and 2014 respectively. Following views are getting updated by the slider

- Flu Trends Map
- Flu Trends: Seasons
- Flu Trends: Monthly Report
- Flu Trends: Regional Report

State selector

The United States map is used to select the states. States can be selected/deselected on the *click* event. There is a limit on the number of selections that a user can make because of the limited space in the views. We have kept this limit to 5. The users can deselect the older states to see the newer states.

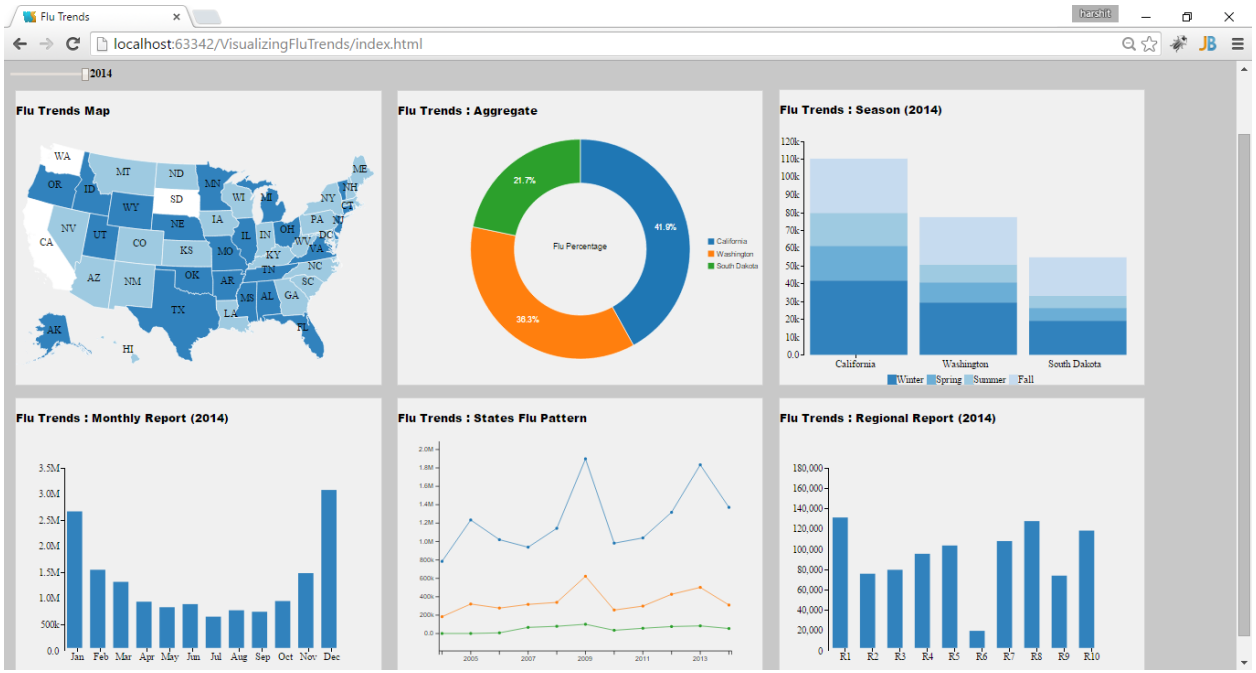


Figure 7: Visualization for states WA, CA and SD

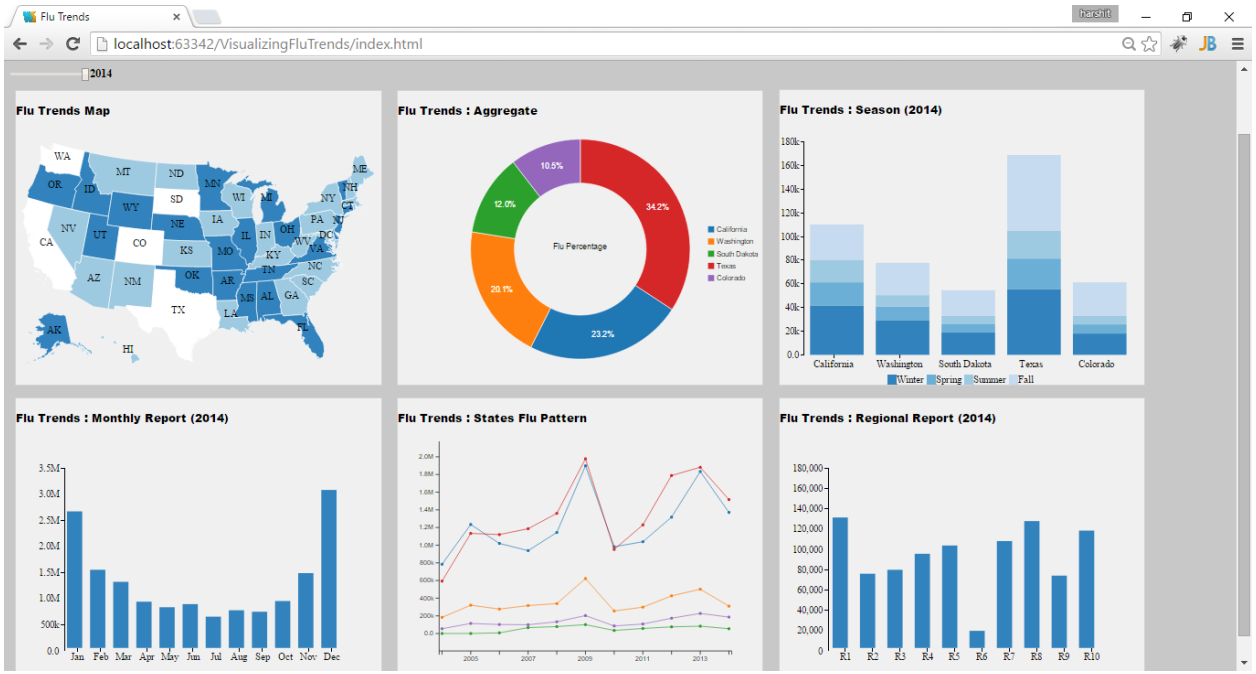


Figure 8: Visualization for states WA, CA, CO, TX and SD

Figures 7 and 8 show the visualization for different number of state selections. The user is notified when he tries to select more than 5 states. Figure 10 shows that scenario.

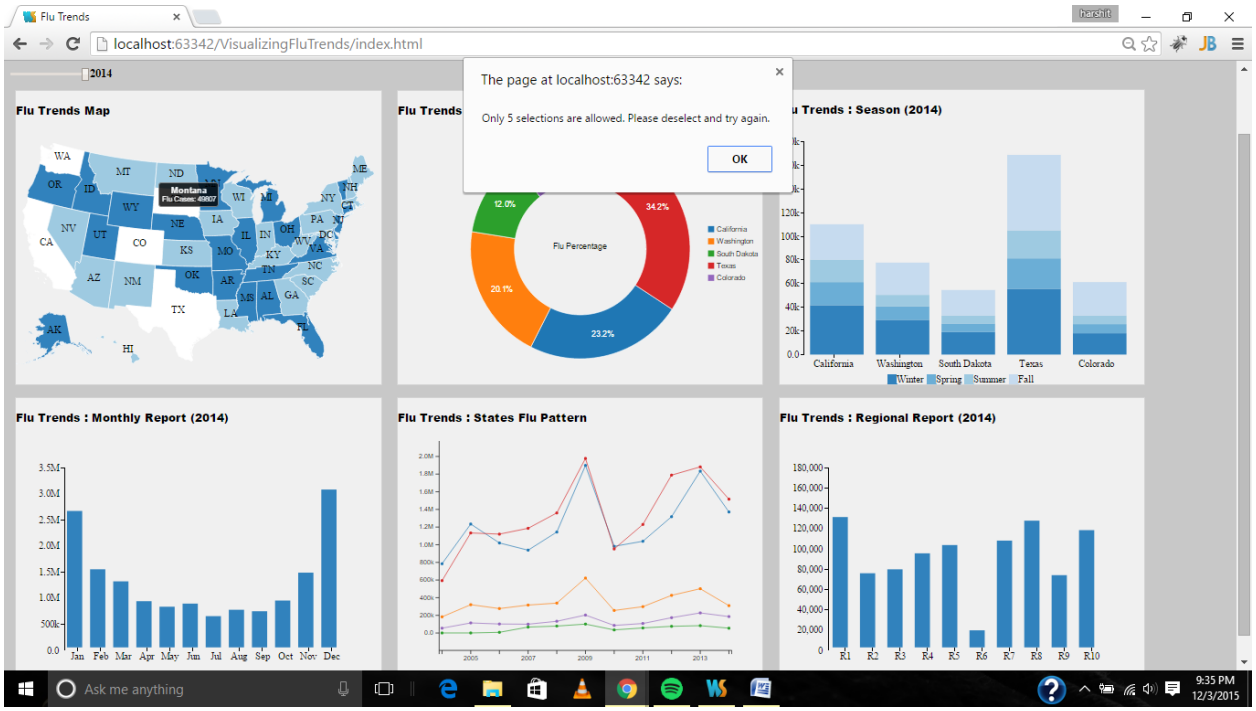


Figure 9: Notification when the number of selections exceed 5

Cities

The data contains flu cases for major cities in that particular state. The seasons view provides the utility of viewing the number of flu cases in the cities of that state. User can click the bars on the Seasons view to see the cities of that state and can go back to the previous view by clicking on any of the bars again.

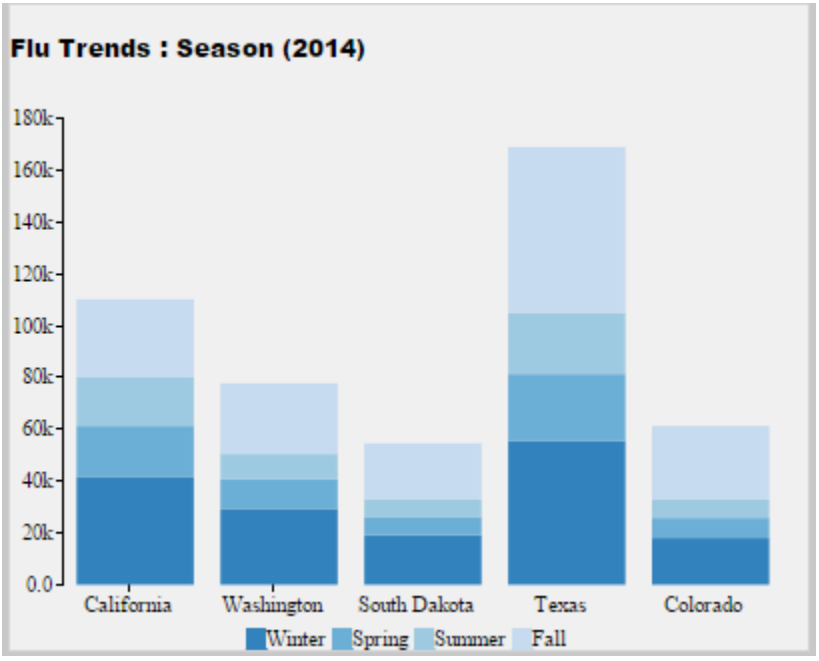


Figure 10: Seasons view for selected states

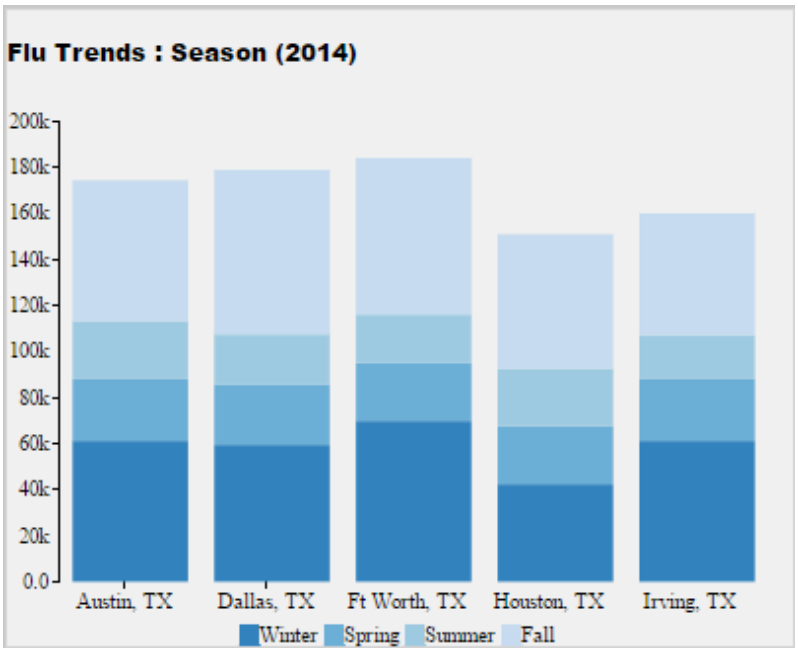


Figure 11: Seasons view when user clicks on Texas Bar

Figure 10 shows the seasons view for the year 2014 and selected states. When the user clicks on Texas Bar, the seasons view gets updated to cities as shown in figure 11. User can go back to the states (figure 10) by clicking on any of the bars again.

Tooltips

The tooltips are enabled on Map, Seasons, Monthly Report and Regional Report views.

- Map view tooltip: Shows the number of flu cases for the selected state and selected year.

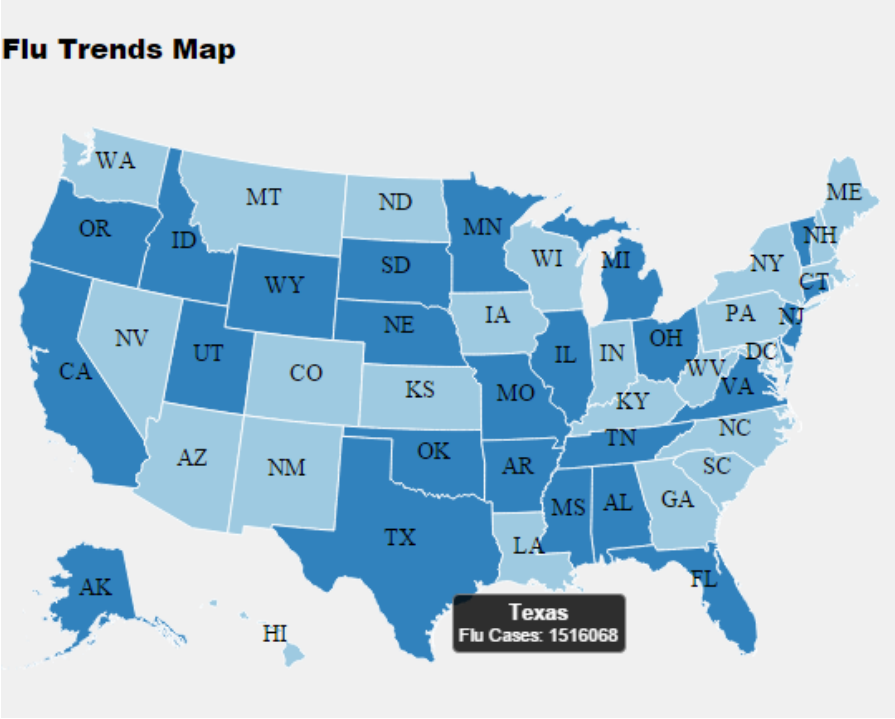


Figure 12: Tooltip on the map

- Season view tooltip: Shows the number of flu cases on the selected season for that year.

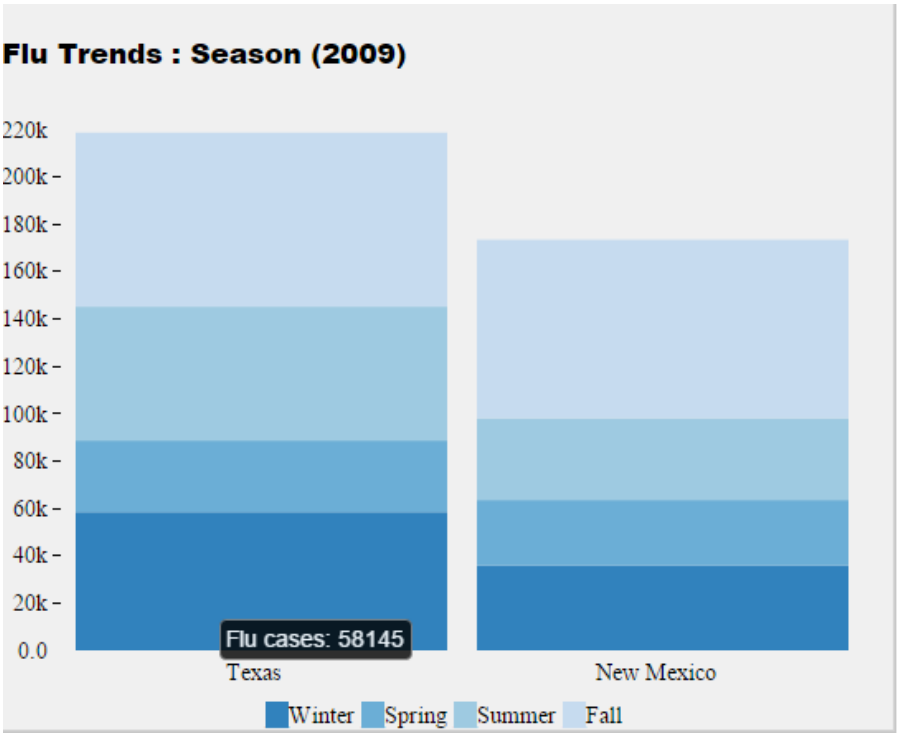


Figure 13: Tooltip showing the flu cases for Texas, Spring 2009

- Monthly Report view tooltip: Shows the number of flu cases for the month in the selected year.

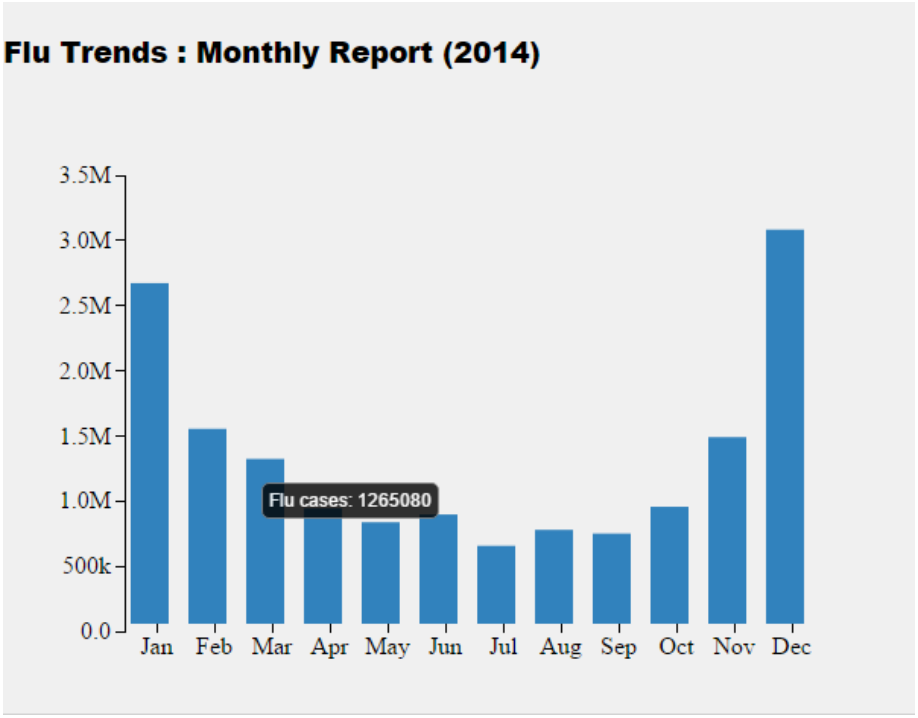


Figure 14: Tooltip showing the flu cases for March 2014

- Regions Report view tooltip: Shows the state codes in the selected region bar.

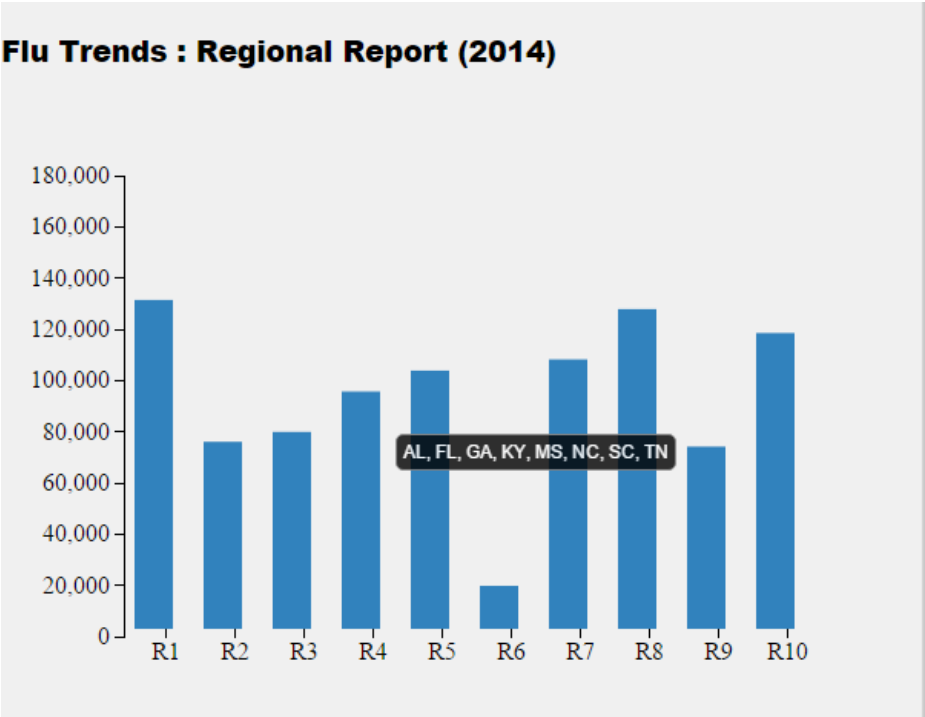


Figure 15: Shows the state codes for R4

Program Flow

Figure 16 shows the event flow of our tool. As discussed above there are primarily 2 events Year selection and State selection that controls the views.

The recursive call on the Seasons view shows that the user can go from States to Cities and vice versa.

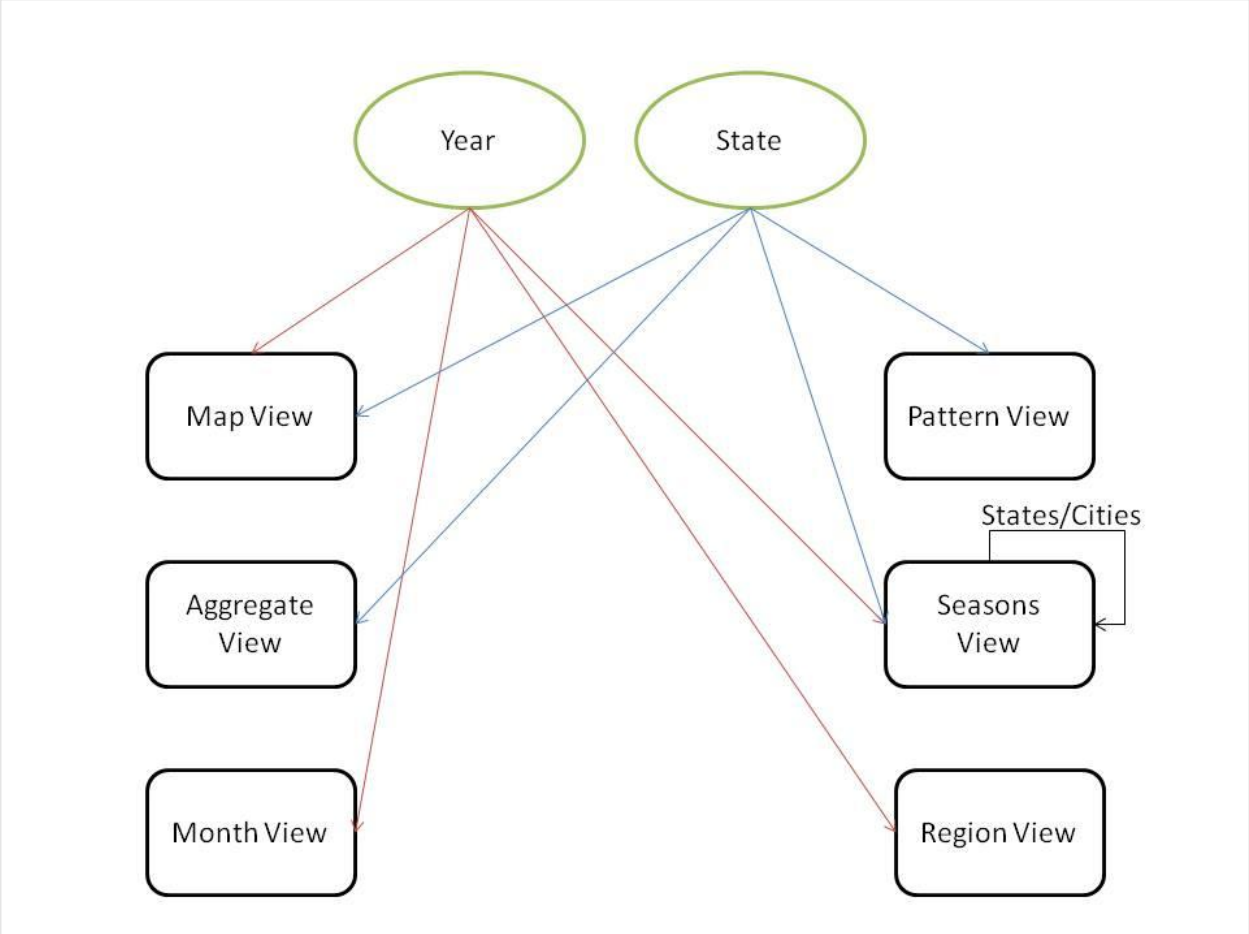


Figure 16: Event Flow

Evaluation

We went through a number of options to design our tool. The scale of data is too large and we wanted to avoid any tabs or multiple visualization designs. In our data set there are multiple dimensions in which the data can be viewed. Years, months, seasons, regions, cities, states and their comparisons, if we had to design a visualization for showing all this information it would have been very difficult to keep it in one page. We prioritized our dimensions and designed based on the more generalized information which the user might find useful. We also limited the number of selections to be made for the states.

There are still improvements which could be made in our design. We are thinking of including as many dimensions in our visualizations yet keeping the tool simple.

Results

The objectives in this project were to identify patterns of flu cases across United States in different time periods and seasons. With our simple design we were able to identify states which are more prone to flu cases. The users can see the flu patterns over a span of 11 years which will help understanding and checking the spread of flu. The lower level details in Seasons view helps to precisely estimate the most vulnerable time period of the year. Our tool is designed for the Google trends data which is released

frequently. Users can use new datasets with our visualizations to learn about flu trends for different time periods.