Basic Info:

Project title : Visualizing USA Consumer Complaints in Finance Sector

Member : Shlok Patel Uid: u1083432 Email: shlokipatel@gmail.com
Madhur Pandey Uid: u1065393 Email: madhur@cs.utah.edu

GitHub Repo : https://github.com/madhur12/Visualizing-Consumer-Complaints

Background and Motivation:

In today's era, everyone uses one or the other kind of financial products/services offered by various financial institutions. Almost everyone has a bank account or a credit card, while some others might also have loan related products. Therefore, it becomes necessary to analyze how satisfied people are with such services and which institutions are doing their best in providing these services to their customers. It would also help if we could analyze how the complaints statistics have changed over the years. So, we decided to choose this project so that it can help us take better decisions in the future, when dealing with finance sector.

Project Objectives:

The main objective of the project is to gain meaningful insights from raw data collected by CFPB, so it can benefit both the consumers and providers of financial products/services.
The primary questions the visualization is trying to answer includes:
- Major issues faced by the consumers when using financial products/services.
- The number of complaints responded/resolved.
- Best/Worst performing institutions based on various criterias
- Regional complaints pattern in the USA.

For the consumers, the visualization can help in determining which institutions are best suited for the service they need.
For the institutions, checking these trends can help to improve their products/services and check how they are performing compared to their toughest competitors.

Data:

The Consumer Financial Protection Bureau (CFPB), a Federal agency responsible from protecting the rights of the consumers, is collecting the consumer complaints data we are using since 2012. The data is available in .csv, .json and .xml formats. All the data is available on the data.gov website at the below link:
https://catalog.data.gov/dataset/consumer-complaint-database#topic=consumer_navigation

Data Processing:

We might need minimal data processing. As of now, we only need to derive new date columns from the existing ones.

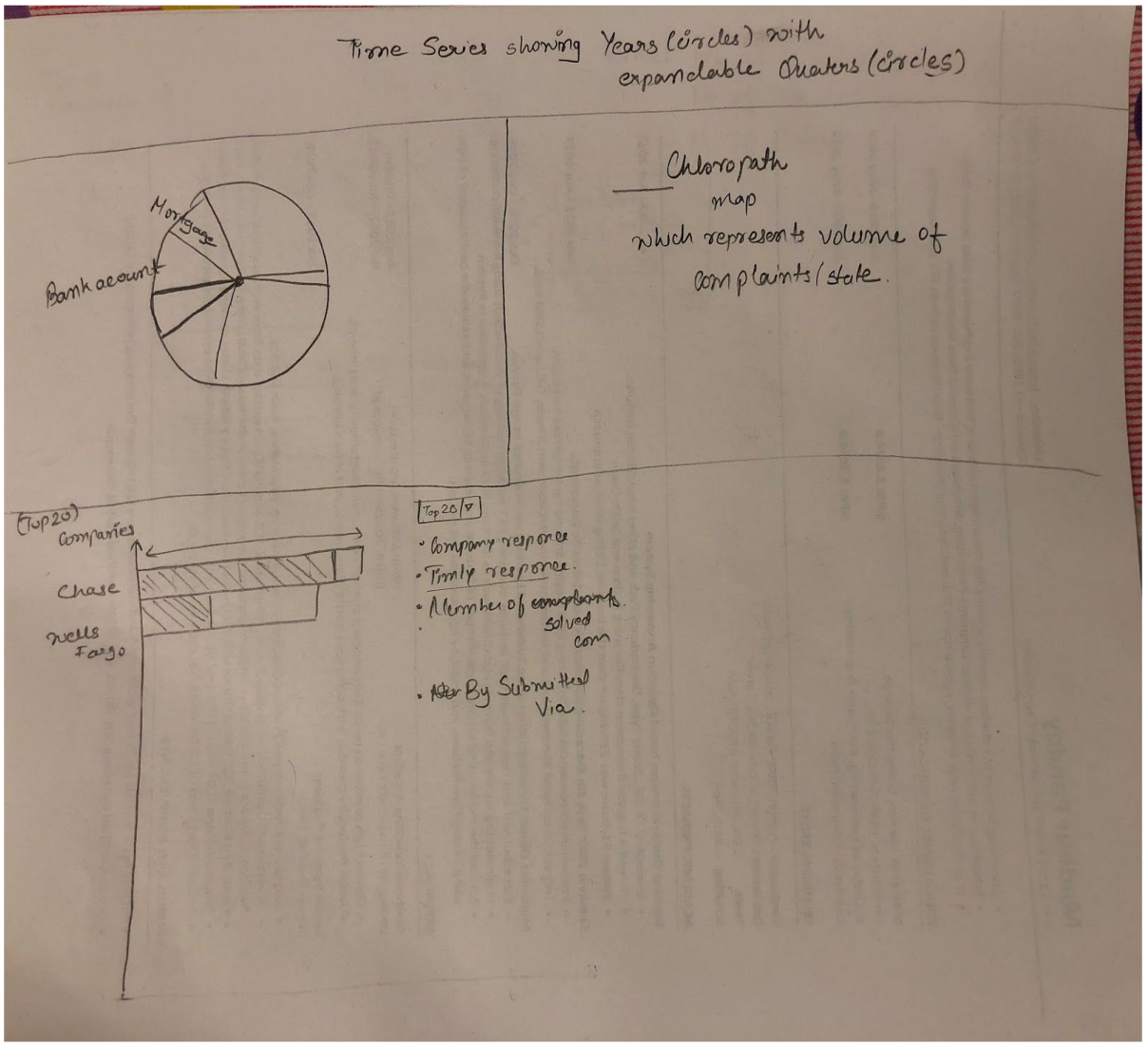
The dataset is very huge and we might have to select a random subset of the data so that it is faster in processing and represents the entire data. However, this is tentative.

<u>Visualization Design</u>:

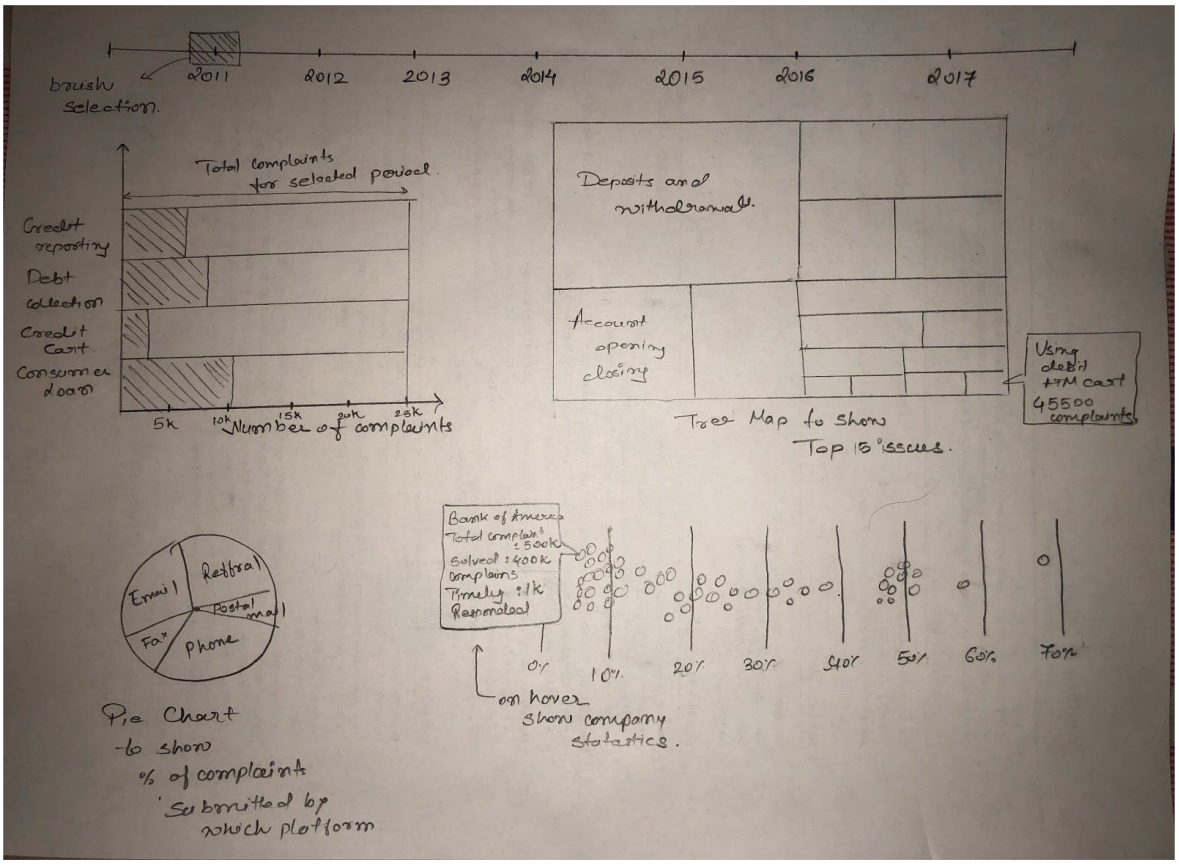
Our visualization design started with brainstorming the various questions we want our visualization to answer. After finalizing the questions, we came up with various initial designs and finally agreed upon a final design. Below are the snapshots of our initial designs and final design.

<u>Initial Designs</u>:

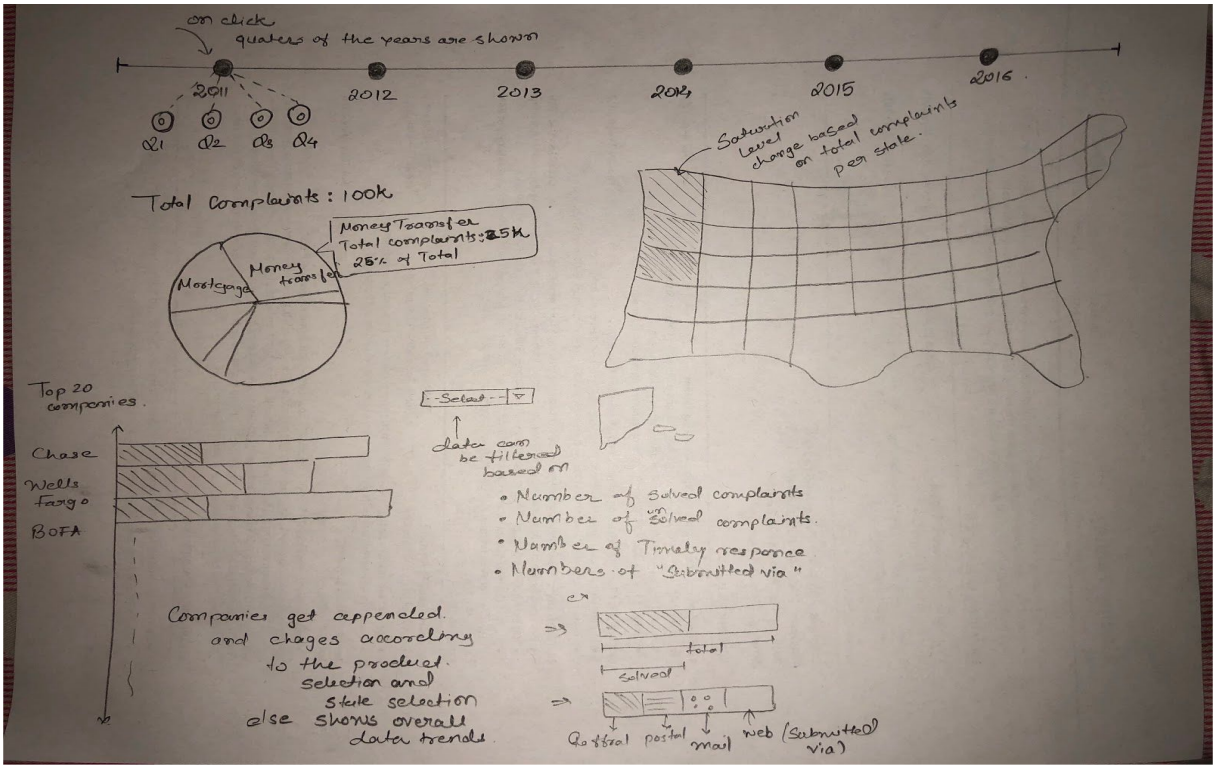
1.



2.


brush selection.
2011
2012
2013
2014
2015
2016
2017
Total complaints for selected period.
Credit reporting
Debt collection
Credit Card
Consumer loans
5k
10k
15k
20k
25k
Number of complaints
Deposits and withdrawals.
Account opening closing
Using debit ATM card
45500 complaints
Tree Map to show Top 15 issues.
Email
Referral
Postal mail
Fax
Phone
Pie Chart
-to show
% of complaints
submitted by
which platform
Bank of America
Total complaints : 500k
Solved : 400k
complains
Timely : 1k
Responsed
0%
10%
20%
30%
40%
50%
60%
70%
on hover
show company statistics.

3:


on click
quaters of the years are shown
2011
2012
2013
2014
2015
2016
Q1
Q2
Q3
Q4
Total Complaints : 100k
Money Transfer
Total complaints : 25k
25% of Total
Mortgage
Money transfer
Saturation level change based on total complaints per state.
Top 20 companies.
Chase
Wells Fargo
BOFA
Companies get appended and chages according to the product selection and state selection else shows overall later trends
Select
later can be filtered based on
Number of Solved complaints
Number of Solved complaints.
Number of Timely response
Numbers of "Submitted via"
total
solved
Referral
postal mail
web (Submitted via)

Final Design:

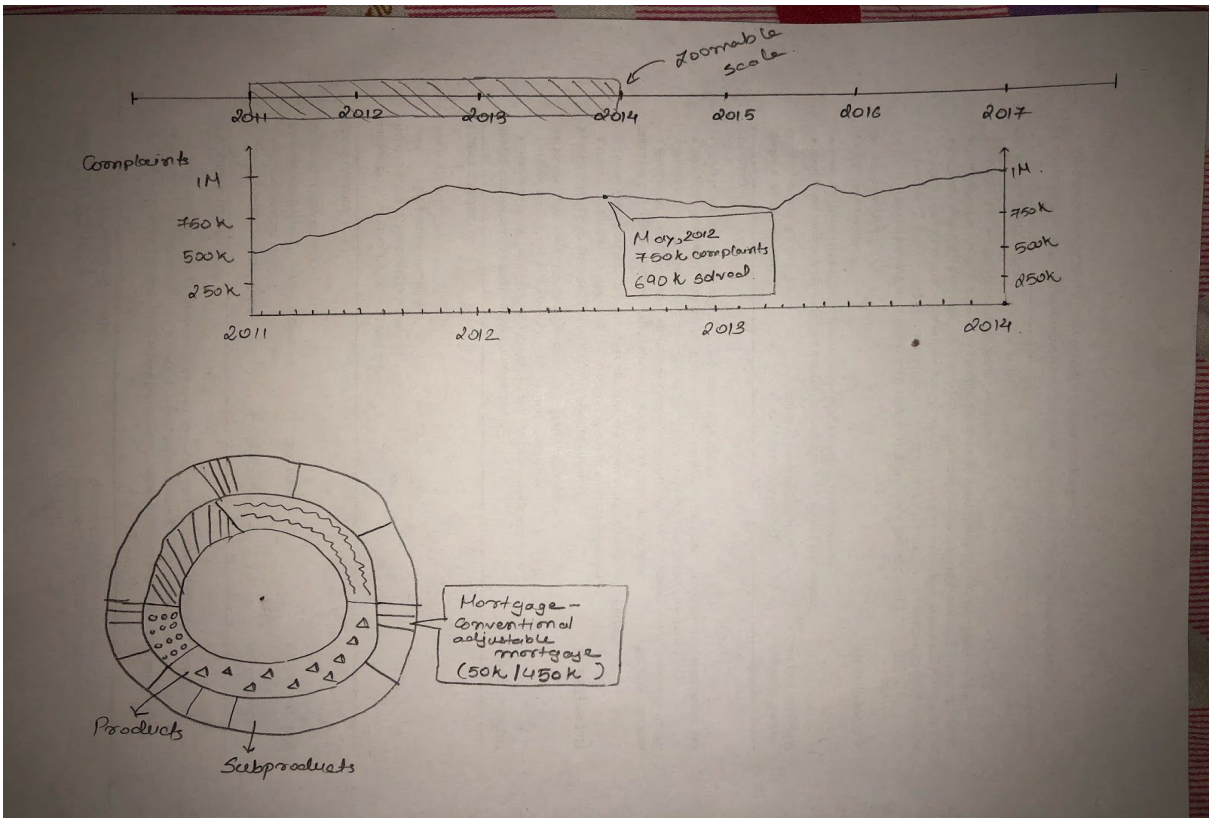


Fig 1: Complaints Trend over year and Product/Subproduct Sunburst Chart

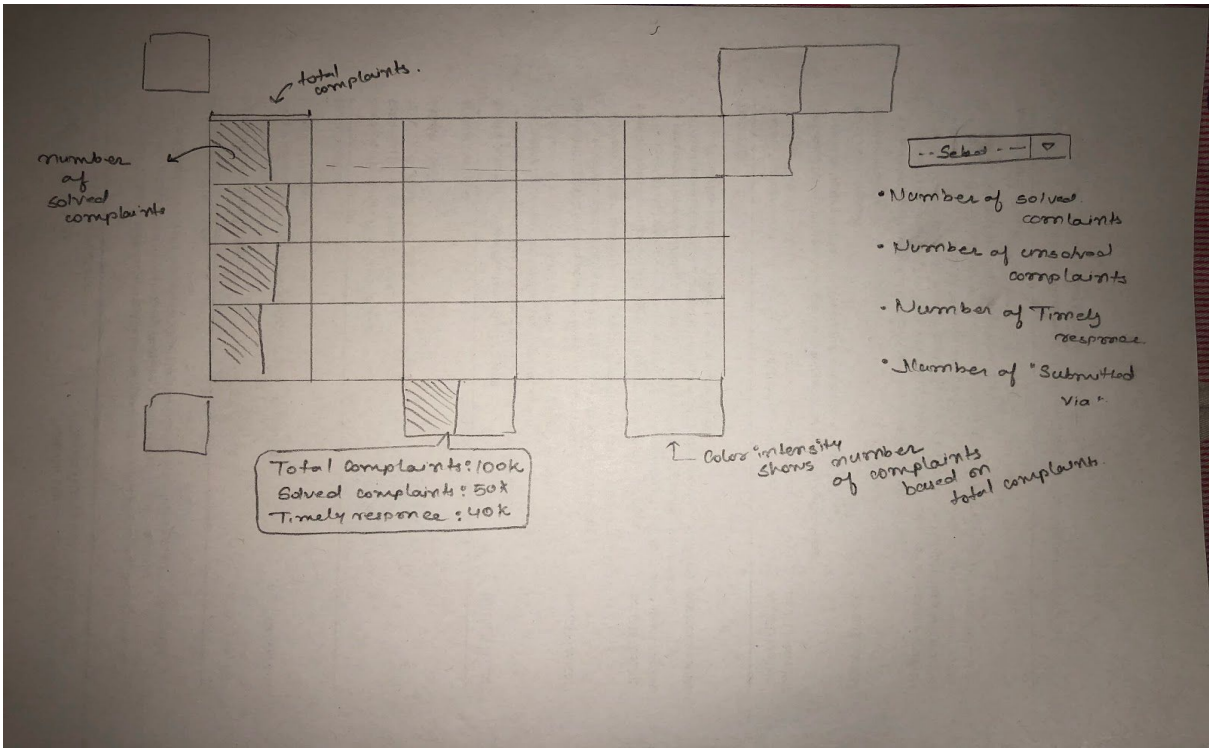


Fig 2: Map with area charts representing the overall state trend for the dropdown selected.

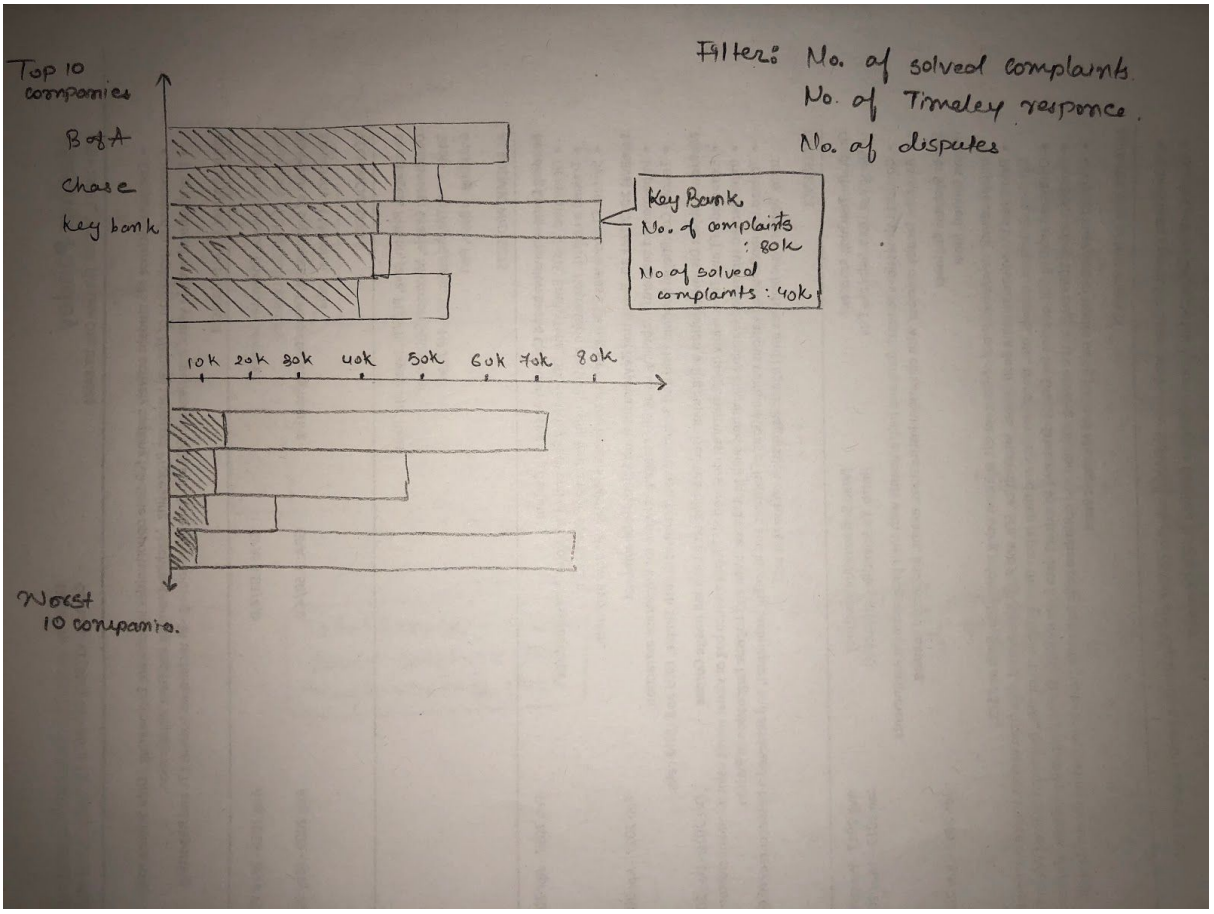


Fig 3: Stacked bar chart showing best/worst companies based on the dropdown value selected.

We plan to implement the visualization using time series, zoomable sunburst chart, customized choropleth and stacked bar chart. We will have brush component over the time series which will be used to select a subset of time and change all the views accordingly. The subset can also be selected by zooming over the timeline.
All the svg components will have tooltips which will display additional information about the data it visualizes. We will also have drop-down for the stacked bar charts to compare the best/worst institutions based on different criteria.
All our views will be interactive in the way that selecting something in one view will update all the corresponding views.
The various channels used are:

Position:
The position on a common scale is used as the channel to encode the year time line. The position channel is the best visual encoding for Ordinal type.

Position/Length:
The complaints trend and the best/worst company charts use position/length as the visual encoding as both the fields are quantitative in nature and quantitative data is best distinguished using the position/length channel.

Hue:
The Sunburst chart uses hue as the channel to distinguish between the various products and subproducts. This is because the products and subproducts are categorical data and hue is a good separator for categorical data.

Arc Length:
The Sunburst chart also uses the Arc length as the channel to encode the number of complaints for each product category, which is quantitative data.

Area:
For each state on the map, we are using Area as the channel to show the proportion of total complaints received vs resolved.

The various marks that our visualization will be using include:

Line:
Lines are used to show connectedness in complaints trend chart. Along with the position channel, it helps in better visualizing the pattern in the chart.

Saturation:
We might (tentative) use saturation to encode the volume of complaints received in each state. This is the saturation of the area chart for map described above.

Must-Have features:
The interactive complaints trend chart and the best/worst companies stacked bar chart, which change based on brush selection are must have feature.

Optional Features:
The Sunburst chart is an optional feature. Also, if time permits, we will implement all companies response time plot as drawn in the bottom right of design-2.

Project Schedule:

Time	Progress	Tasks
Week1	13%	Creating the year timeline with zoomable brush
Week2	40%	Creating the sunburst Chart or Tree map
Week3	66%	Creating the map with selected filters
Week4	90%	Creating a stacked bar chart with selected filters
Week5	100%	Testing the entire functionality