

BIAS IN VEHICULAR TRAFFIC STOPPAGES IN THE USA

CONTENTS

1. Introduction.....	3
2. Design	3-8
2.1. Sheet 1: Brainstorm	
2.2. Sheet 2,3,4: Initial Designs	
2.3. Sheet 5: Realization design	
3. Implementation.....	9-14
3.1. UI Based Libraries	
3.2. “Dashboard” Tab	
3.2.1.Diversity	
3.2.2.Violations	
3.2.2.1. Age Group vs Gender	
3.2.2.2. Hour vs Day of Week	
3.2.3.Violation Rate by Race	
3.2.4.Arrest and Search Count/Rate	
3.2.5.Drugs Search Rate	
3.3. “Explore Locations” Tab	
4. User Guide	14-16
4.1. “Dashboard” Tab	
4.1.1.Diversity	
4.1.2.Violations	
4.1.2.1. Age Group vs Gender	
4.1.2.2. Hour vs Day of Week	
4.1.3.Violation Rate by Race	
4.1.4.Arrest and Search Count/Rate	
4.1.5.Drugs Search Rate	
4.2. “Explore Locations” Tab	
4.2.1.Aggregate Stoppages	
4.2.2.Play Single Day Stoppages	
5. Conclusion	16
6. Bibliography.....	17
7. Appendix.....	18-20

1. Introduction

Are the people of the US somewhat biased among few races? Due to recent happenings and overtime in various states of the United States of America, this becomes an important question in every field/aspect of life, be it job roles, arresting, community rights, voting rights and whatnot.

We through our narrative visualisation project would like to gain the attention of high officials, head of departments of respective states traffic police department. Continuing to answer our questions from the exploration project we want to convey to the audience. After the intended audience realises few evident numbers of prevailing bias, they could take a certain decision over it. It can be to transfer the particular region on-duty traffic police, have one to one word, or if a psychological department exists internally then let them handle the case.

Unfortunately, we don't have recent years data on policing stoppage, but it is worth exploring if this bias has been continuing from the past.

A few of the results which few want to convey to our intended audience includes gender bias, violation analysis, racial bias, spatial analysis by race and many sub narrations which indirectly relates to these.

2. Design

Some of us may have a habit of thinking about designs, draw them and start thinking about the caveats at the same time. This is the human tendency to judge and question at the time. Ideally, one should follow proper logical steps – define, brainstorm, sketch prototype, etc.

We have used the “Five Design Sheet” methodology for data visualisation planning. As the name suggests this involves five stages segregated to its defined sheet. We followed this process so that at the end we are clear about what visualisation design is to be implemented, what resources are required and even set an estimate of the project.

Sheet 1: Brainstorm

The purpose of this stage is to enlarge the possibilities of visualization designs even if the different visualization would convey the same results. One should be open-minded and not criticize any idea.

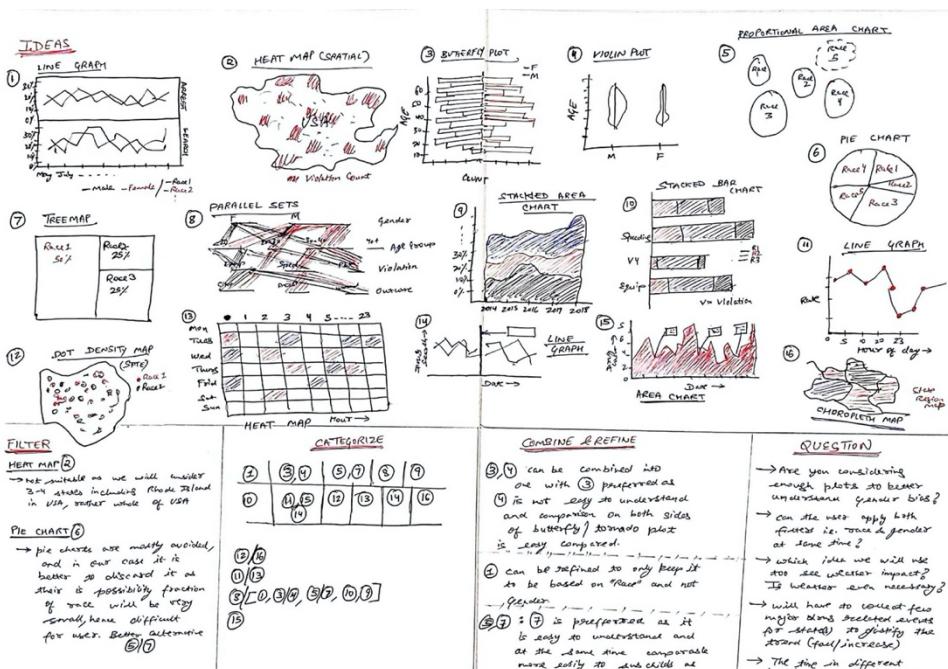


Figure 1: Sheet 1 FDS

We have quite a few results to convey to the user in a simplified and as effective way as possible. At the same time, the user must be able to interact with the visualisations to gain confidence in the findings and drill-down data. After revisiting most of the types of graphs we have learnt and brainstorm, we were able to come up with 16 different types of visualisation plots. The distinct ones were – line graph, heat map, butterfly/tornado plot, violin plot, proportional area chart, pie-chart, tree map, parallel sets, stacked area

chart, stacked bar chart, dot density map, heatmap, area chart and choropleth map. These plots covered all of my analysis which I want to convey related to racial discrimination, gender bias, the impact of time on violations and several more related to demographics of age, location of the traffic stoppages.

The next process involved filtration i.e. discarding of not worthy plots. I discarded the heatmap due to the unavailability of all USA states/county data and generally, heatmap considers one variable by saturation of a single colour. This doesn't fit well as we are planning to depict multi-variate data to get a sense of comparison. The pie chart came next to my find after considering the low proportion of some races, which may make the pie area compact (undesirable).

Categorisation of plots depicts similar results and similar plot types were done, to consider them in our initial design sheets later. The Combine/Refine step was carried out to select the best preferred or combine them. Butterfly plot depicting violation by age-group and gender was preferred over violin plot of the same scenario as violin plot is more of a statistical plot and not well versed to the general audience. Similarly, the tree map was preferred over the proportional area chart to show the population of race in a state/county due to the similar reason of shrinking for small values.

In the end, we needed to ask ourselves questions by taking into consideration the results we want to convey (before even starting the FDS). A few of my questions are:

1. Are we considering enough plots to better understand gender bias specifically?
2. Can the user apply both filters i.e. race and gender at the same time?
3. Is weather data impact will be useful to the client/user? If don't consider it at all.
4. Will collect few major drug events for states to justify the decline/incline during a period of drug search rate during vehicular stoppages.

Sheet 2,3,4: Initial Designs

After we are clear with ideation, it's time to start designing our initial prototype layouts for the dashboard. Sheet 2,3, and 4 are used to try out different plots and layout combinations.

Sheet 2

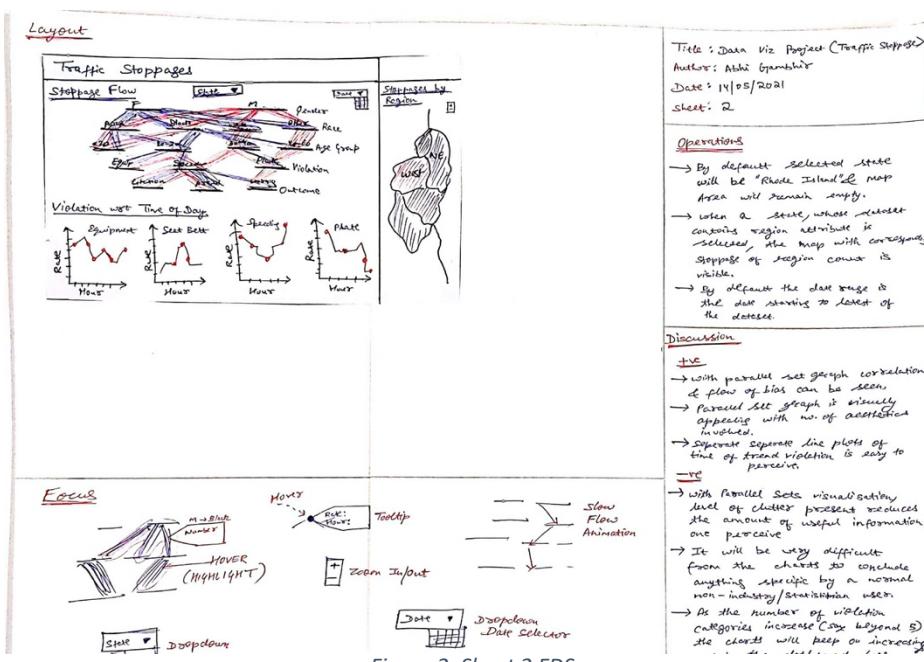


Figure 2: Sheet 2 FDS

For this layout, we chose "stoppage flow" (parallel set plot) to depict the very high-level flow starting from gender level to stoppage outcome level. Intermediate levels may include factors like Race and age group. This will explain the flow and will require careful observation to get the bias (if any).

Another plot that will depict "violation by the hour of the day" (line graph) was subdivided as each line plot for each violation, to avoid visual distortion if we plot more than 5 lines in the same graph.

To depict some spatial data we decided to go with selecting “stoppage by region” (choropleth map). This will give percentage values of stoppages carried out in the region.

Focus

- Hover + highlight on stoppage flow plot, Hover tooltip showing numbers, animation/transition of consecutive level flows (parallel set plot), state dropdown, tooltip of “violation by hour of the day” (line plot) points, zoom in/out map, date dropdown

Operations

- By default selected state will be “Rhode Island” and the map area will remain empty
- When a state, whose dataset contains a region attribute is selected, the map with the corresponding stoppage of region count is visible/plotted
- By default, the date range is the date starting to the latest of the dataset

Discussion

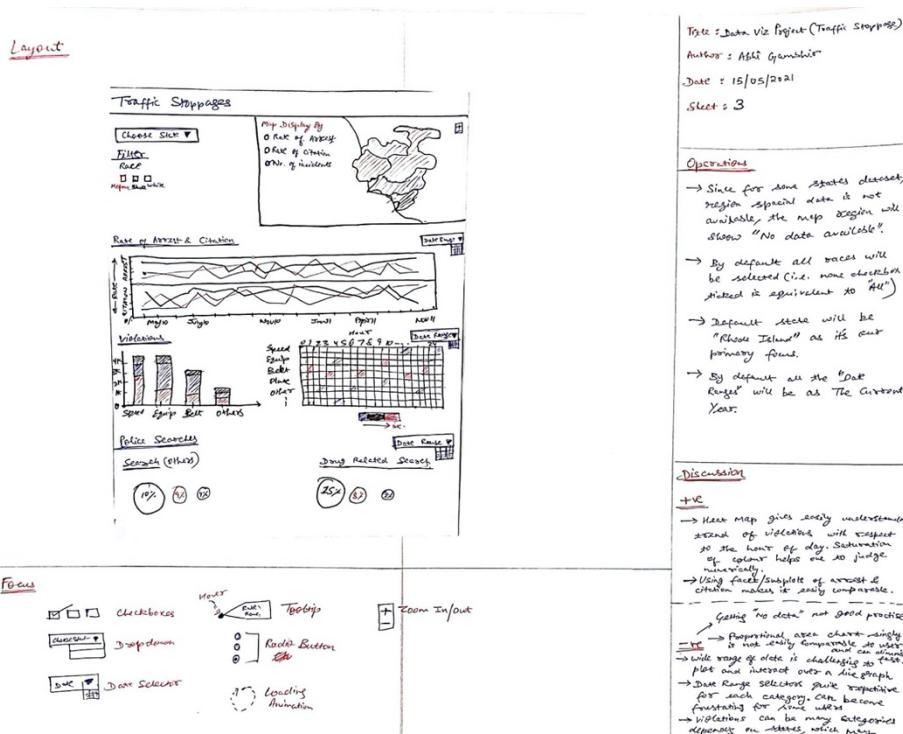
Pros

- With parallel sets visualisation, correlation and flow of bias can be seen
- The parallel set graph is visually aesthetic
- Separate line plots of time of trend violation are easy to perceive

Cons

- With parallel sets, the level of clutter present reduces the amount of useful information one can perceive easily
- It will be very difficult from the charts to conclude anything by a naïve or non-industry user
- As the number of distinct violation categories increase (say beyond 5-6) the plot will seem repetitive taking most of the dashboard space

Sheet 3



For this sheet, we wanted to step towards more user understandable visualisation plots. The first plot we included was “Rate of Arrest and Citation” (synchronised line subgraphs). This will help the user to get the relation and trend of both rates of arrest and citation (ticketing) with the third dimension of driver race.

The second plot, “Violations by race” (stacked bar chart) is used to depict race wise each violation. For depicting “violation by the hour of the day” we use a heatmap, as the saturation of a single colour aesthetically captures user attention at peak/low hour of

Figure 3: Sheet 3 FDS

violation stoppages. A proportional area chart is though not that suitable but let's explore our options by what it can plot. So we chose a proportional area chart of races for driver search and specific driver drug search.

Spatial attributes for "stoppage by region" we continued to use a choropleth map from Sheet 2. This will give percentage values of stoppages carried out in the region.

Focus

- Driver race selection checkboxes, state dropdown, date dropdown, hover tooltip on the heatmap, Radio buttons for region map, Plot loading animation, Zoom in/out of region map

Operations

- Since for some state's dataset, region spatial data is not available, the map region will show "No data available"
- By default, all the available races will be selected from the checkbox
- Default state will be "Rhode Island" as it's our primary focus
- By default, all the date ranges will be as the current year

Discussion

Pros

- Heat map gives easily understandable variances of violation for the hour of the day
- Using facet/subplots of arrest and citation makes it comparable.

Cons

- Proportional area chart solely is not easily comparable to the user and can highly shrink as well
- A wide range of data is challenging to plot and interact over a line graph
- Date range selectors are quite repetitive for each category and can get frustrating for some users
- Violations can be many categories depending on the state, which may take up a lot of space if the bar chart's orientation is kept landscape.

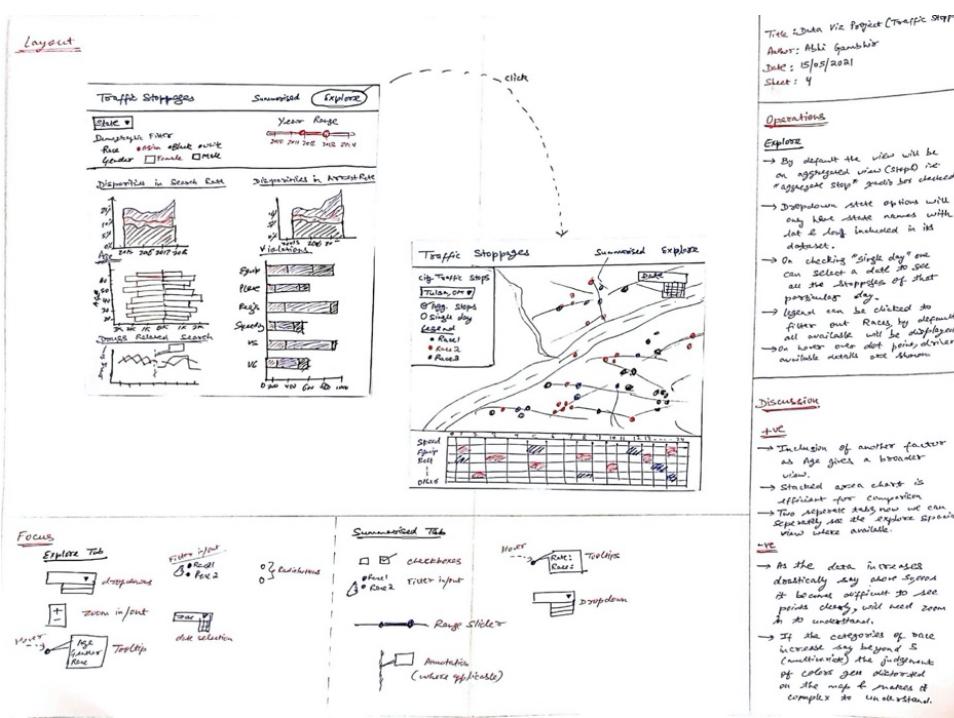


Figure 4: Sheet 4 FDS

Sheet 4

For this sheet's layout we introduced different tab concept, one for summarised dashboard and other primary focused on states with spatial data.

Focus

- Summary Tab: dropdowns, zoom in/out, tooltips, radio buttons
- Explore Tab: checkboxes, filter in/out, range slider, annotation

Summarised Tab:
As the data increases drastically say above 50000 it becomes difficult to see points clearly, user need zoom in to understand.
If the categories of race increase say beyond 5 (including) the grid alignment of colors gets disturbed in the map & number of complex to understand.

Operations

- By default the view will be on aggregated view.
- Dropdown will be dynamically rendered with values containing only those states with latitude and longitude data.
- On checking "single day" markers are plotted on the map for that day.

Discussion

Pros

- Inclusion of another factor as Age gives a broader view
- Stacked area chart is efficient for comparison
- Two separate tabs, now we can separately see spatial data of states (where available)

Cons

- As the data increases drastically say above 50,000s it becomes difficult to see points clearly, will need zoom to understand.
- If the categories of race increase say beyond 5 (multivariate) the judgement of colours get distorted on the map and makes it complex to understand and differentiate.

Sheet 5: Realization design

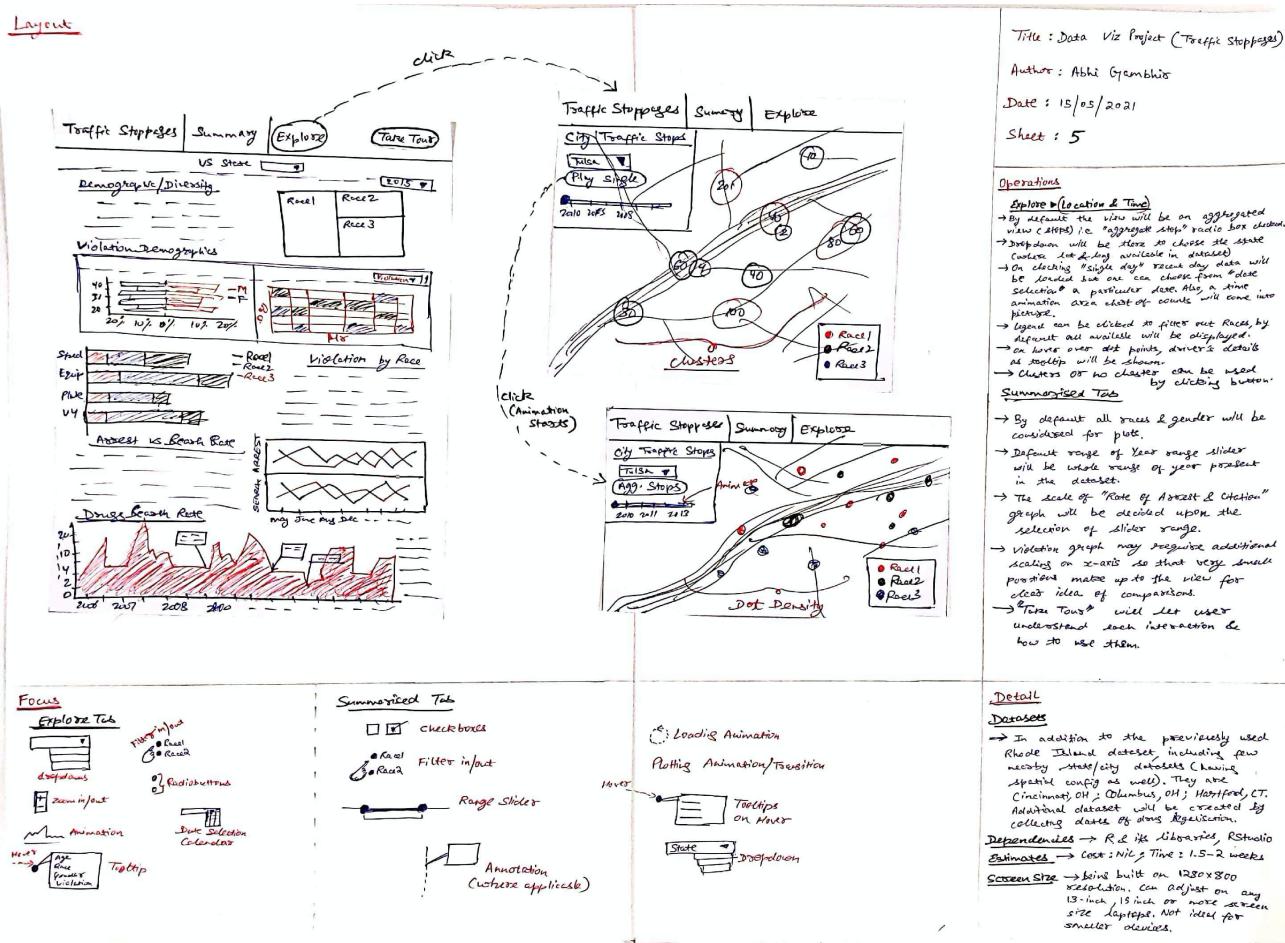


Figure 5: Sheet 5 FDS

The final design sheet collabs various ideas from sheet 2,3,4 considering each pros and cons of each visualisation and design layouts. We chose the following different types of visualisations – tree map, stacked bar chart, butterfly/tornado chart, heat map, area chart, line graph, dot density map.

Focus

- Summary Tab: dropdowns, hover tooltips, zoom in out, annotations, filtering, buttons, sliders, loading animation
- Explore Tab: zoom in out, clicks, buttons, range sliders, plotting animation, loading animation, checkboxes, mini map, drag, dropdowns

Detail

Datasets

- In addition to the previously used Rhode Island dataset, we are including few more states/county (having spatial data, to expand the scope of our project). We are including Vermont and Hartford, CT.

Dependencies:

- R and its libraries, RStudio

Estimates:

- Cost: NIL; Time: 1.5-2 weeks

Screen Size:

- being developed on 1280x800 resolution. Can adjust on any 13 inch, 15 inch laptop. Not ideal for smaller devices.

3. Implementation

It all comes down to implement the final sheet design we have created. We decided to use R Shiny to implement the design. Since we have two tabs in our final design, so let's explore each visualisation in them and how we implemented them.

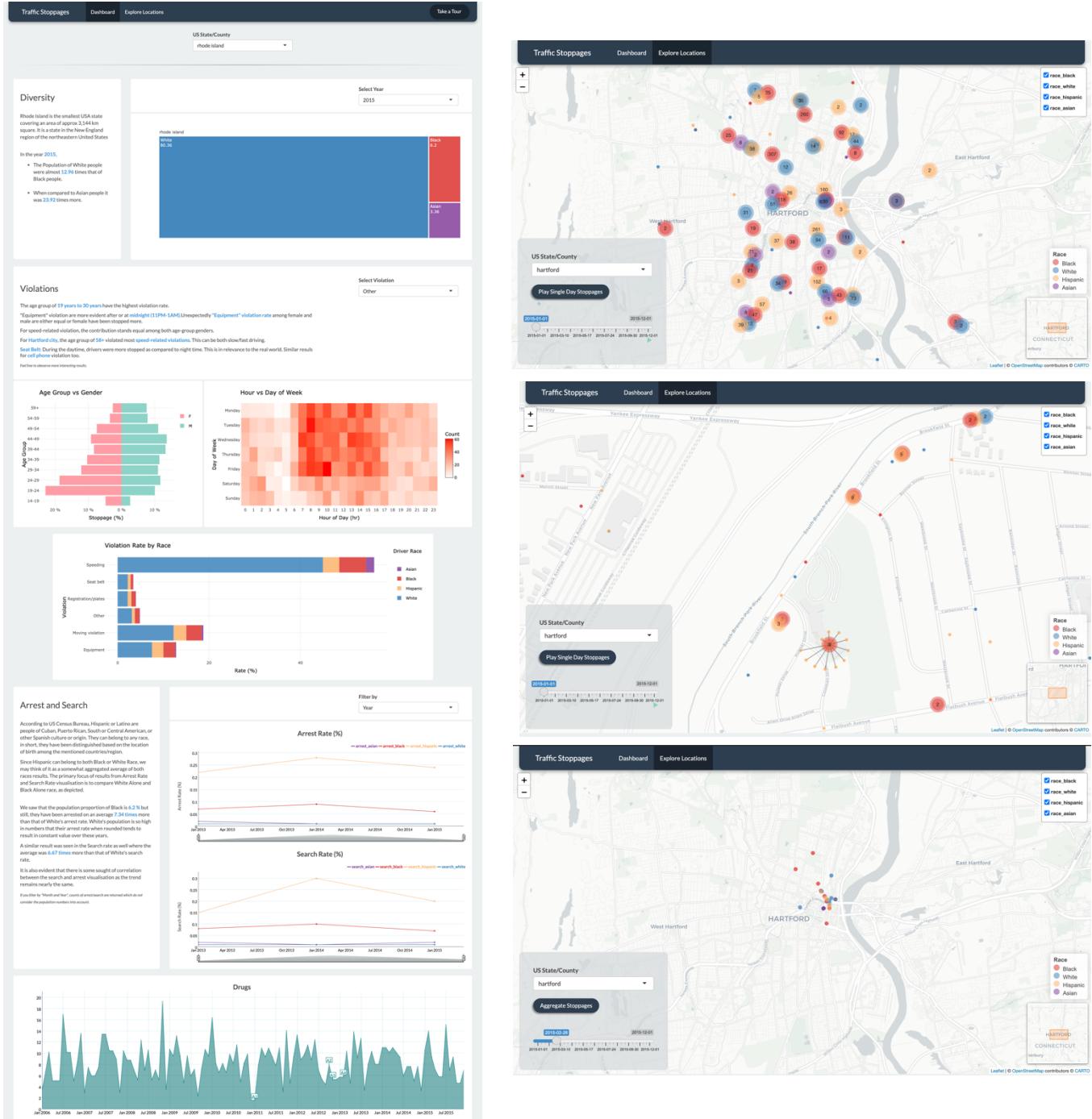


Figure 6: Implementation Screenshots

UI Based Libraries

shinythemes: provides a pre-built theme with default written CSS properties and tags

rintrojs: useful for giving the tour to the user by displaying messages or instructions over elements

shinycustomloader: embeds a loader on plots until it gets load completely.

“Dashboard” Tab

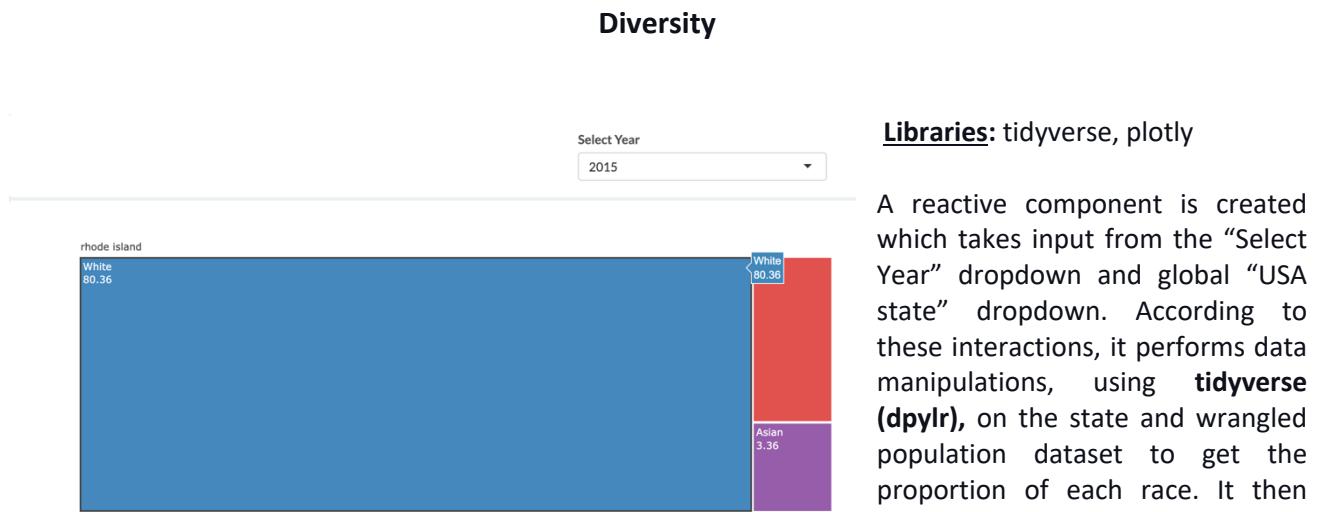


Figure 7: Diversity Tree map Visualisation

Libraries: tidyverse, plotly

A reactive component is created which takes input from the “Select Year” dropdown and global “USA state” dropdown. According to these interactions, it performs data manipulations, using **tidyverse (dplyr)**, on the state and wrangled population dataset to get the proportion of each race. It then returns the final data frame so that we can proceed with plotting.

Plotly is used for visualisation by calling the reactive function and using **Plotly's plot_ly()** to embed into the UI's assigned layout. Labels, values, colours, etc. are used along with the plot. The background is also set to transparent from the user interface perspective.

Challenges: setting out of the parent child relationship structure so as to form the required custom tree map structure.

Alternatives: There is a R package just to plot tree maps by the name of “**treemap**”. We could have used it, but since we could achieve similar visualisation with already loaded library, we chose not to.

Violations

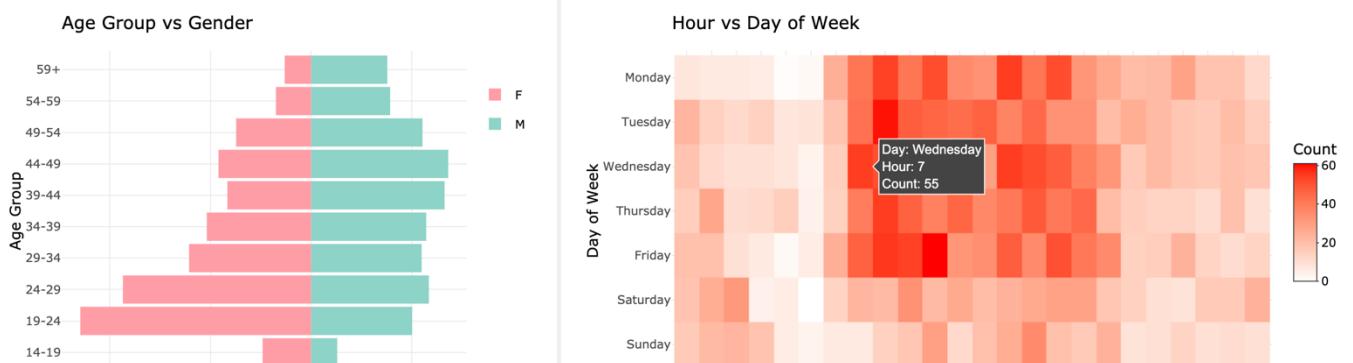


Figure 8: Tornado and Heatmap Visualisation

Age Group vs Gender

Libraries: tidyverse, ggplot2, plotly

A reactive component was created which take selection input from the “Select Violation” dropdown and global “USA state” dropdown. According to the returned values from these interactions, data is manipulated using **tidyverse**. Since we have age recorded as a single entity number for each driver, we created age intervals of 5, starting from (minimum age in the dataset – 1). To display the order of age group correctly levels were assigned to these age groups categories. The population dataset is merged along with the manipulated data and finally, gender-wise stoppage per cent is calculated.

Finally, the tornado/butterfly plot is plotted to the designated area in UI using **ggplot()** and to enable interaction **ggplotly()** on the returned ggplot() object. Since tornado visualisation is a customised plot, not present by default in the mentioned libraries, we had to make modifications and workarounds to get the final result.

Challenges: Since tornado visualisation is a custom visualisation, we had to make modifications to plot the final structure. Briefly we used ggplot’s `geom_bar()` with `fill=gender` and `stat=identity` enabled, flipped the coordinate axis, `scale_fill_manual()` to assign colours, `scale_y_continuous()` with breaks as 0-100% by 10% and labels enabled.

Hour vs Day of Week

Libraries: tidyverse, ggplot, plotly

The “Select Violation” dropdown values were populated dynamically according to the input value of “US state” from the server to the UI element. This was done because several states had different types of violation or even different name to similar ones.

A reactive component is created which according to the state selection manipulated the data by hour-day of the week. This was done using the functions of **tidyverse (dplyr)**.

Before plotting the results weekday names(Mon-Sun), day hour values (0-23) were created manually to capture the results by merge where no value is present. These were also changed to factor datatype with levels so that the real world order is maintained in the visualisation. Tooltip text was created and finally, the heatmap was plotted using the **ggplot** and **ggplotly** (for interaction).

Violation Rate by Race

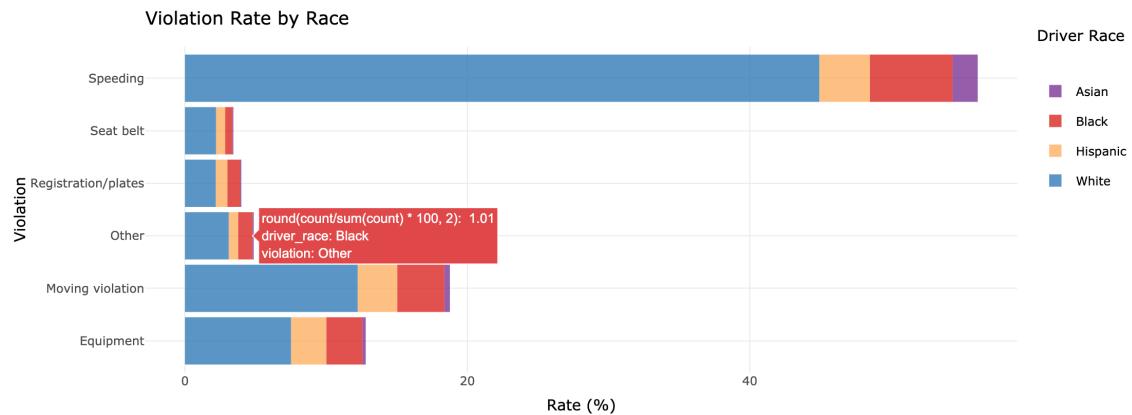


Figure 9: Violation Rate by Races Stacked Bar chart

Libraries: tidyverse, ggplot, plotly

This doesn't need a reactive component, as we are interacting with already made ones used in earlier visualisations. The visualisation makes use of **ggplot's ..count..** and **sum(..count..)** built-in method to get the desired values we need by manipulation. Aesthetics is added with required labels, the coordinate axis are flipped to allow more useable space. Finally, ggplot code is wrapped inside **ggplotly** to make it an interactive visualisation.

Arrest and Search Count/Rate



Figure 10: Arrest and Search Line Graphs with Pointers enabled

Libraries: stats (base package), lubridate, dygraphs, tidyverse

The implementation flow/concepts of both the arrest and search rate visualisation are the same. Let's discuss only in terms of the arrest rate plot.

We have the date and time of the stoppages events, but a lot of days are there when the count can be 0 or not present. To make up for it, as we are planning to plot time series, we use **lubridate** features (`ymd()`, etc.) and create a continuous month-year wise sequence to merge it with the actual dataset with counts. **Tidyverse** is the go-to for every data manipulation, be it filtration, selection, group by, etc. In our project. Finally, the data

frame is converted to time-series type using the base package **stats** `ts()` function. According to the selection input of “filter by”, for each race, we keep updating/saving the `ts` object and finally plot them using **dygraphs** package.

Along with returning time-series `cbind()` of each race, we returned several other parameters like the default date window and label’s text. By enabling `dyRangeSelector()` we get a date range select interaction to fit a huge date period of data in a small space. Since these two plots have been made synchronised, changing filters or the date period of one directs the same to the other.

Challenges: Manipulation of data to make it suitable for conversion to time-series as we had several missing dates; Dynamic labelling of y-axis label on changing the “Filter by” value; Graph synchronisation.

Drugs Search Rate (*available only for “Rhode Island” State*)

Libraries: `stats` (base package), `lubridate`, `dygraphs`, `tidyverse`

The implementation flow/concepts of visualisation are mostly similar to the previous one. New features added to this plot is the annotations of events which may give suitable justification of the trend either increasing or decreasing rate of driver search for drugs reasons. Also, this plot is an area plot, unlike the line plot as that of arrest/search rate.

Challenges: Surfing out the relevant news articles, jurisdiction results related to drugs; Manipulation of data to make it suitable for conversion to time-series as we had several missing dates.

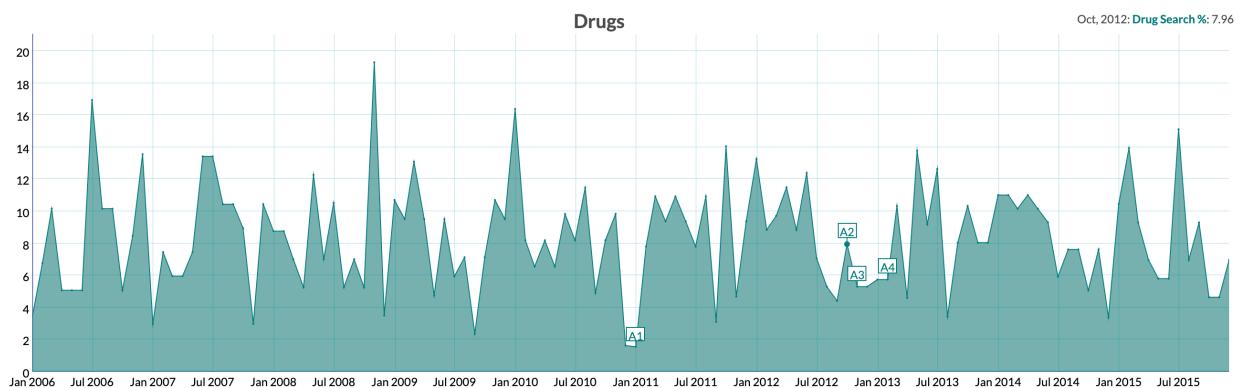


Figure 11: Drug search rate Area chart visualisation

“Explore Locations” Tab (Refer Fig 6)

Libraries: `leaflet`, `tidyverse`

This tab showcases two options of view to the user. One is the “aggregated stoppage” view which enables cluster formation by using plain greedy clustering at the backend. It uses the Distance Grid concept under the hood [7]. This enables high-level interaction by clicking the clusters and drilling down to the exact location of

driver stoppages. We have used both leaflet's static render Leaflet and leaflet proxy with the observed events (buttons, layers, state change in our case).

Another aesthetically pleasing view we offer to our user is animated day-wise stoppages. This can be enabled by clicking the "Play single day stoppages" button. Under the hood, it calls the observe event function clear outs the previously marked markers and add the day-wise markers in the loop. By default, we are taking all the days from the latest year in the dataset. This can be useful for future perspective if somehow we could capture real-time data.

Challenges: The clustering mechanism considers all points by default, to segregate them and make clusters race wise was a bit challenging. But in the end, figured out a way to plot for each race and cluster for each of them. Removing of default's colour of red, yellow, orange (depicting the saturation of the number of points) was again a small challenge. One of the issues I was not able to tackle that well is the overlapping clusters of a different race. The only solution I could implement was to introduce opacity in the colouring scheme.

4. User Guide

It is preferred if the project can be run in browser window, because of extra mouse gestures it allows. Sometimes in Window run of RStudio some zoom in/out using trackpad doesn't work.

The traffic stoppages narrative visualisation dashboard contains two tabs (navbar options) namely: "**Dashboard**" and "**Explore Locations**". Let's take an overview of how to use the features of dashboard.

To take a tour of the dashboard click "**Take a Tour**" button from the nav bar. This will guide you about the interactions available for the visualisation.

"Dashboard" Tab

This is the default tab that opens up after running the shiny app from RStudio. Please wait few seconds to allow the initial plots to load up.

Diversity (Refer Fig 7)

Interactions: US State/County dropdown, Select Year dropdown, Tooltip on hover over plot, Zoom in/out

- 'Select Year' to get Diversity proportion of various Race/Ethnicity of that year.
- Hover over to see the proportion of race.
- Click tree map square to zoom in/out with transition.

Violations (Refer Fig 8)

Renders two plots related to violation as the attribute of filtration and interaction.

Age Group vs Gender

Interactions: US State/County dropdown, Select Violation dropdown, Hover over bars (tooltip), Filter gender

- 'Select Violation' to get age-group wise violation and its frequency of occurrences by weekday-time.
- Hover over the bars to get percentage rate conduct of the violation by gender.
- Feel free to filter out gender by clicking from legend.

Hour vs Day of Week

Interactions: US State/County dropdown, Select Violation dropdown, Hover over square area (tooltip), Zoom

- 'Select Violation' to get age-group wise violation and its frequency of occurrences by weekday-time.
- Hover over the boxes to get the count of violations on weekday-hour basis.
- Select area by dragging mouse with selection (left click) to zoom to that part of weekday + hour.

Violation Rate by Race (Refer Fig 9)

Interactions: US State/County dropdown, Hover over stacked bars, Zoom, Filter race

- Hover over the bars to get driver race, violation and rate of violation.
- Feel free to filter out race(s) by clicking from the legend.
- Zoom by dragging mouse with selection (left click) to select particular view.

Arrest and Search Count/Rate (Refer Fig 10)

Interactions: US State/County dropdown, Filter by dropdown, Hover on line points (tooltip), Synchronised interactions between two plots by using timeframe slider, Highlight on line hover, Zoom

- “Filter by” either “Month and Year” or “Year” to see the data accordingly.
- “Month and Year” shows the Arrest/Search Counts while the “Year” shows Arrest/Search Rate(%).
- To change the timeframe or drill down, feel free to slide timeframe range or zoom in by dragging mouse with left click.
- Hover over the points to see values next to the legend.

Drugs Search Rate (available only for “Rhode Island” State) (Refer Fig 11)

Interactions: US State/County dropdown, Annotation tooltip on hover on event points, Hover on line points (tooltip), Change timeframe

- Select “Rhode Island” state to see the drug rate visualisation plot.
- Some important events which may have caused downfall in driver’s drug search rate are annotated. Feel free to hover them to get the event details.
- Also to change the timeframe, select the area by dragging mouse with selection (left click).
- Hover over the points to see search rate(%) value on top right of plot.

“Explore Locations” Tab

This tab contains the spatial exploration of traffic stoppages by race/ethnicity. One can drill down and find out areas/road/suburbs by zoom-in zoom-out or use mini map, where stoppage was conducted.

Interactions: US State/County selection dropdown, Conditional hide/show of buttons, Animated date slider, Mini map to give a sight of area in view, Checkboxes to add/remove race(s), Zoom in/out, Clusters on click interaction, Marker points animation

There are two views of leaflet map to explore:

Aggregate Stoppages

This view is the default view that loads initially. Its focus is based on clustering of points by races on the basis of greedy clustering by distance.

- Click “Aggregate Stoppages” button (if in view, otherwise it ultimately is the default view)
- Select the “US State/County” from dropdown you want to explore. Let the map plot the clusters.
- Drill down to spatial locations by clicking the cluster and explore the results.
- Drag the mini map area to change the location of actual map.
- Filter out circle markers/clusters to display the race you want from checkboxes on top right.

Play Single Day Stoppages

This is animation based day wise plotting of stoppages points of the latest year present in the dataset. It can be really useful to plot real-time data as a future scope.

- Select the “US State/County” from dropdown you want to explore.
- Click “Play Single Day Stoppages” button (if in view)
- This will initiate the animation marking/plotting of various traffic stoppages along with slider changing values by each day of the latest year in dataset.
- One can drag the slider forward/backward to revisit the markers.
- Drag the mini map area to change the location of actual map.
- Filter out circle markers to display the race you want from checkboxes on top right.

5. Conclusion

Findings Summary

- Population of White is in majority in all the three states or region.
- The age group of 19 years to 30 years have the highest violation rate.
- “Equipment” violation are more evident after or at midnight (11PM-1AM). Unexpectedly “Equipment” violation rate among female and male are either equal or female have been stopped more.
- For speed-related violation, the contribution stands equal among both age-group genders.
- For Hartford city, the age group of 58+ violated most speed-related violations. This can be both slow/fast driving.
- Seat Belt Violation: During the daytime, drivers were more stopped as compared to night time. This is in relevance to the real world. Similar results for cell phone violation too.
- We saw that the population proportion of Black is quite less as compared to White's but still, they have been arrested more than 4 times (depends on state/county) that of White's arrest rate. White's

population is so high in numbers that their arrest rate when rounded tends to result in constant value over years in some state.

- A similar result was seen in the Search rate as well where black's were searched more than 6 times that of White's search rate.
- It is evident that there is bias. Also correlation prevails between the search and arrest rate as the trend remains nearly the same.
- Many other results can be observed by interacting with the visualisations.

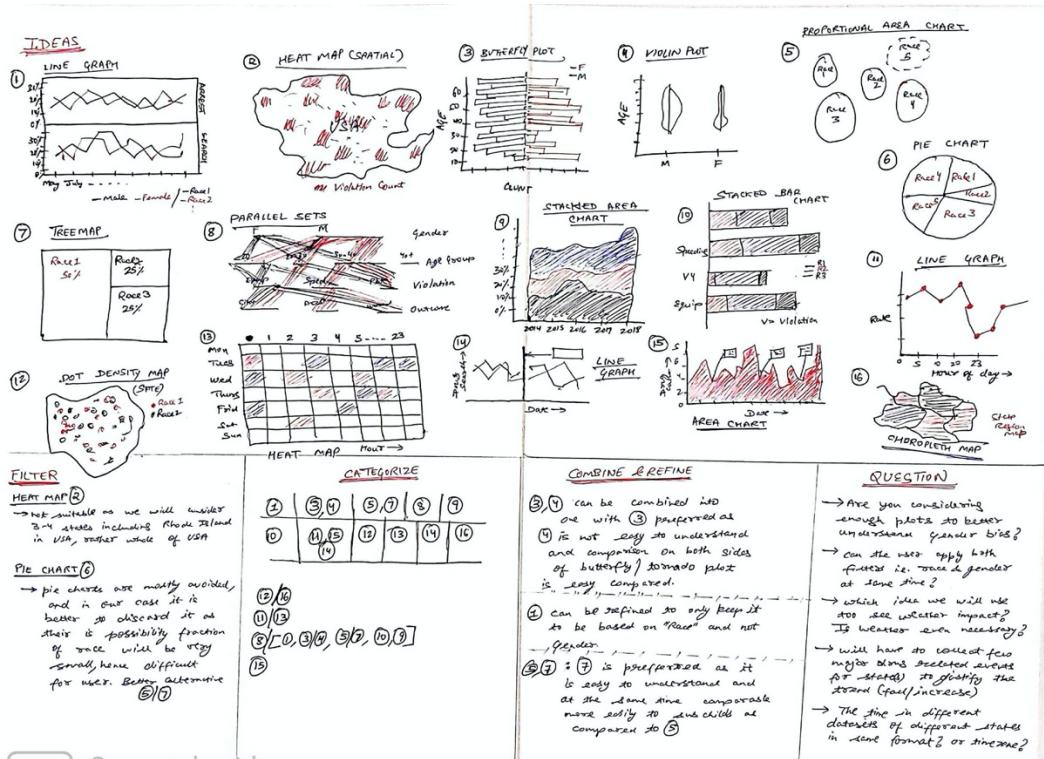
6. Bibliography

- [1] United States Census Bureau. (2017). Providence (city), Rhode Island. Retrieved from <https://web.archive.org/web/20130127054813/http://quickfacts.census.gov/qfd/states/44/4459000.html>
- [2] Deloitte, Datawheel. Population data from <https://datausa.io/>
- [3] Five Design Sheet Methodology. Retrieve from <http://fds.design/>
- [4] District of Rhode Island, 3rd Defendant In Rhode Island's Largest Cocaine Bust Sentenced To 15 1/2 Years In Federal Prison. Retrieved from <https://www.justice.gov/usao-ri/pr/3rd-defendant-rhode-islands-largest-cocaine-bust-sentenced-15-12-years-federal-prison>
- [5] Nicola. (2019). POSXlt and filter() in R. Retrieved from <https://stackoverflow.com/questions/56850331/posxlt-and-filter-in-r>
- [6] The Stanford Open Policing Project, Stanford. Data retrieved from <https://openpolicing.stanford.edu/data/>
- [7] Leaflet Package. Leaflet.MarkerCluster 0.1 Released, The Technical Bits. Understanding from <https://leafletjs.com/2012/08/20/guest-post-markerclusterer-0-1-released.html>
- [8] Plotly, Graphing Libraries. References [from https://plotly.com/r/](https://plotly.com/r/)
- [9] <https://awesomelibrary.org/project/nanxstats/awesome-shiny-extensions#text>
- [10] Dygraphs for R, Package. Reference from <https://rstudio.github.io/dygraphs/>
- [11] The R Graph Gallery, <https://www.r-graph-gallery.com/index.html>
- [12] Color Brewer 2.0, Color Reference from <https://colorbrewer2.org/>

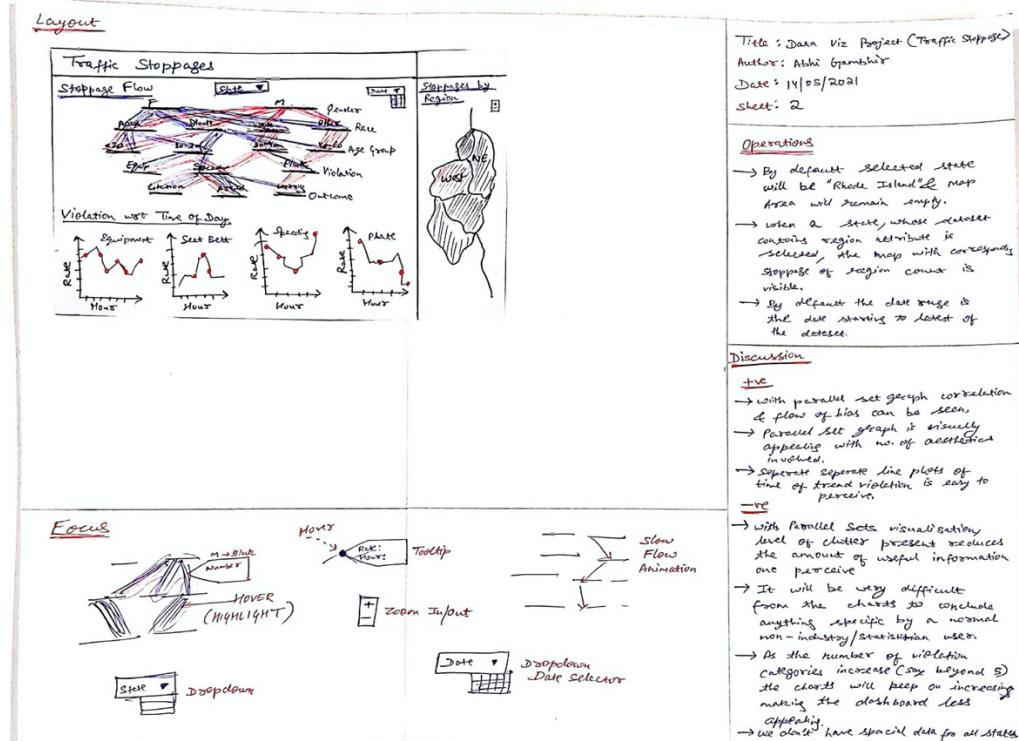
7. Appendix

Five Design Sheets

Sheet 1

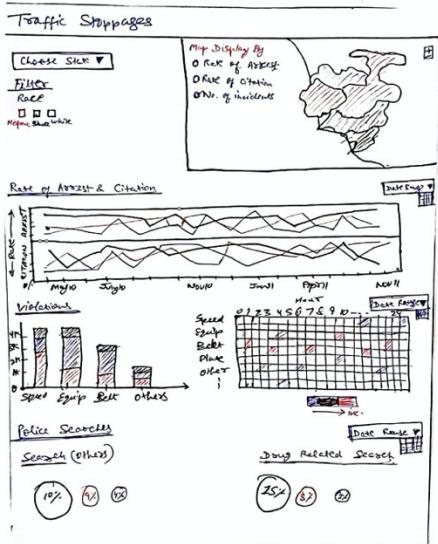


Sheet 2

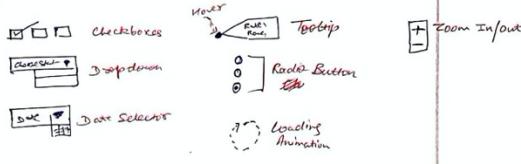


Sheet 3

Layout



Focus



Title: Data Viz Project (Traffic Stoppage)

Author: Abhi Gambhir

Date: 15/05/2021

Sheet: 3

Operations

- Since for some states dataset region specific data is not available, the map region will show "No data available".
- By default all races will be selected (i.e. none checkbox selected is equivalent to "All")
- Default state will be "Rhode Island" as its our primary focus.
- By default all the "Date Range" will be as The Current Year.

Discussion

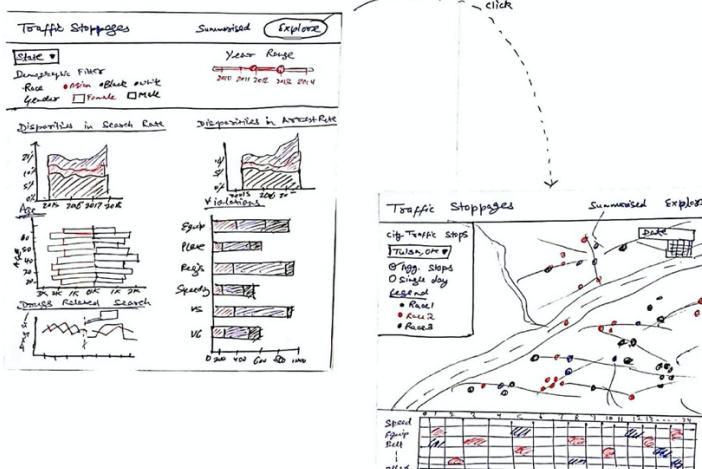
+ve

- Heat Map gives easily understandable trend of violations with respect to the hour of day. Saturations of color helps one to judge numerically.
- Using facets/subplots of arrest & citation makes it easily comparable.

- Getting "No data" not good practice
- Proportional area chart singly is not easily comparable to width and area.
- Wide range of date is challenging to track, plot and interact over a line graph.
- Date Range Selection quite repetitive for each category. Can become frustrating for users when interacting with them.
- Violations can be many categories depends on states, which may take up lot of space if ordination is kept horizontal.

Sheet 4

Layout



Title: Data Viz Project (Traffic Stopper)

Author: Abhi Gambhir

Date: 15/05/2021

Sheet: 4

Operations

Explore

- By default the view will be on aggregated view (Step 1) i.e. aggregate stop counts not checked.
- Dropdown state options will only have state names with lat & long included in its dataset.
- On checking "Single day" one can select a date to see all the stops/offs of that particular day.
- Legend can be clicked to filter out races by displaying all available until the legend is hovering over dot point drivers available details are shown.

Discussion

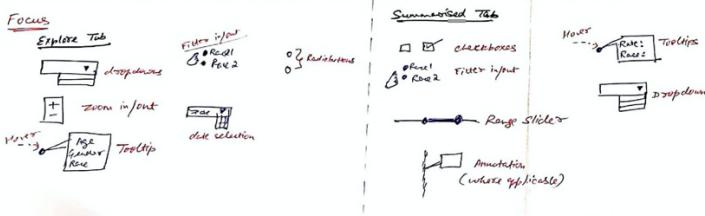
+ve

- Inclusion of another factor as age gives a broader view.
- Stacked area chart is efficient for comparison.
- Two separate tabs now we can separately see the explore section now where available.

-ve

- As the data increases drastically say about 50000 it becomes difficult to see points clearly, will need zoom in to understand.
- If the categories of race increase say beyond 5 (including race) the judgment of colors gets distorted on the map to make it complex to understand.

Focus



Sheet 5

Layout

Title : Data Viz Project (Traffic stops)
Author : Abhi Grambhik
Date : 15/05/2021
Sheet : 5

Operations

Explore (Location & Time)

- By default the view will be an aggregated view (e.g. i.e. aggregated stops per location).
- Dropdown will be used to choose the state (and then specific city/citywide selected).
- On clicking "single day" racial day data will be loaded but one can choose from "date selection" a particular date. Also a time animation over a chart of counts will come into picture.
- Legend can be clicked to filter out races by different colors available will be displayed.
- Hover over dot points, driver's details as tooltip will be shown.
- Clusters or no clusters can be used by clicking button "Summary Tab".

Summary Tab

- By default all races & gender will be considered for plot.
- Default range of year range slider will be whole range of year present in the dataset.
- The scale of "Rate of Arrest & Citation" graph will be decided upon the selection of slider range.
- Violation graph may require additional scaling on x-axis so that we'll have provided marker up to the view for clear idea of comparison.
- "Time Tour" will let user understand each intervention in how it was done.

Focus

Summarized Tab

- checkboxes
- Radio buttons
- Range Slider
- Annotation (Clusters applicable)
- Load Animation
- Plotting Animation/Transition
- Tooltips on Hover
- State dropdown

Detail

Datasets

- In addition to the previously used Rhode Island dataset, including few nearby state/city datasets (having similar config as well). They are Cincinnati, OH ; Columbus, OH ; Hartford, CT.
- Additional dataset will be created by collecting data of stops aggregation.

Dependencies

- RL lib libraries, RStudio
- Estimated : Cost : Nil ; Time : 1.5-2 weeks

Screen Size

- being built on 1280x800 resolution, can adjust on any 13-inch / 15-inch or more screen size laptop. Not ideal for smaller devices.