

Education and Traffic fatalities, what can we expect from “more educated” states when it comes to driving?

Author: David Rios

Introduction

The following visualisation project is oriented to the Government as target audience, more precisely to the National Highway Traffic Safety Administration (NHTSA) which has a division for Vehicle Safety and is in charge of the information related to fatal crashes records in each state.

The main message to be transmitted is which are the states in the United States with a higher amount of traffic fatalities and also the relationship between education and traffic fatalities. This message is very relevant for the government as they are the one in charge of creating new policies that help to reduce this numbers, and taking informed decisions is crucial for reducing the fatalities, specially in the more problematic states.

As it was observed during the exploratory phase of this project, the states of Texas, California and Florida are the most problematic when it comes to traffic fatalities. Interestingly, it was also observed during this part of the project that the education is highly co-related with the observed records for the states mentioned above. The exploration of the reading scores showed that these three problematic states present some of the lowest high school performance for this category.

Taking these facts into consideration, the visualisation project will aim to give the government a general understanding of the current situation of the traffic fatalities with a focus in three main aspects:

- The most dangerous highways across the US with the specific location that present the highest records of traffic fatalities.
- The relationship between education and traffic fatalities per state (focus on TX, CA, FL)
- The type of intersection and hour of the day that is more dangerous across the US.

For additional information about the data sources, please check the appendix 2 Description of data sources.

Design

During the design process the five design sheet methodology was implemented. The resulting design sheets can be observed in the appendix section and a summary of the process is described in this section:

- *First Sheet*: For this first sheet the filtering after the brainstorming process helped to select the levels of detail to be developed for the final visualisation. In this step, the main conclusion was that the graphs should focus on the states of Texas, California and Florida and that the details to be displayed should include the type of intersection where the fatalities occurred, the education information per state, and information about the most dangerous highways in the US with higher detail for the states mentioned before. It is also worth mentioning that during the brainstorming process the *drunk related fatalities* were also identified as relevant to be displayed in the visualisation, as this is something of great interest for the target audience, which in our case is the government.

- *Second, third and fourth Sheets:* In this sheets I took into consideration the main conclusions for the previous sheet in order to develop some visualisations that are focused on the target audience. The following bullet points summarise the points to cover that were exposed through these sheets:
 - The visualisation should present an introductory section that draws the big picture of the states presenting higher number of traffic fatalities.
 - For the details corresponding to other aspects of the message to be transmitted, a set of extra tabs will be implemented in order to create separate visualisations that give the audience the possibility of focusing on one topic without being overwhelmed by many details of multiple topics at the same time.
 - The geo-data would be used with choropleth maps and specially with a tab exclusively focused on the location of the highways presenting the higher fatality rates across the US
 - This maps will contain geo-data and labels with the name of the highway and number of fatalities to give extra information in order for the government to concentrate efforts in specific problematic roads
 - The education is an important element of this project, and it will be included through a scatter plot that displays the performance for each state with the records in traffic fatalities for them.
 - The findings about the type of intersection and hour of day will be revealed through bar-charts that make use of facet-wrapping to show the situation of each state.
 - The influence of drunk drivers will also be taken into consideration across multiple graphs to see the effect of this variable.
- *Alternative designs evaluated:* During the Design process some alternatives were evaluated, first of all the idea of a single page visualisation was considered as it will give the audience a general overview of the situation at a first glance. However, this approach was modified to be separated into multiple tabs as a way to ease the overload of information and give each graph enough space to

Finally, a more familiar design was studied which involved some typical graphs for the audience like bar-graphs and pie charts with some heat-map. This idea was considered as these kind of graphs are widely used in the government projects to communicate their findings, however, it was not selected because it was limited in the use of the geo-data and the message about the specific highways to be tracked was not clear enough in order for the government to take informed actions that allow to control the driving behaviour at specific points.

- *Final Design justification:* In conclusion, the *Sheet: 2 – FDS3* was selected as the final design, this was mainly motivated for the structure of this approach. The fact that it gives each visualisation a separate tab to display the information, represent a big advantage as explained before, in order to give each graph a separate space that allows an effective and strong transmission of the message. The following components are part of the operations to be considered in the selected design:

- Dropdown menus for filtering the states to be displayed
- Checkbox to include/exclude the effect of drunk drivers in the fatality reports

- Multiple tabs with specific contents
- Slider for tracking specific months
- Interactive maps with indicators that display additional information
 - Hover over the indicator to visualise Highway name, State, and fatalities

Implementation

The implementation of the proposed visualisation is supported by the use of the following libraries:

- *shiny*: This library is the base for the adding the interaction interface to the visualisation. Some of the components implemented from this library were the use of *shinythemes*, *tabsetPanel*, *sliders*, *selectInputs*, *checkbox*, *buttons*, etc. All this elements were crucial for materializing the proposed visualisation from the five design sheet.
- *ggplot2*, *leaflet*, *choroplethr*: These libraries were the base for the construction of the actual graphs that are displayed in the shiny interface. With these libraries it was possible to construct the visualisations of the final design.
- *dplyr*, *splitstackshape* and other support libraries: With these final libraries it was possible to reshape the dataset in order to create the inputs for the libraries mentioned before. All the code is contained in one R script that includes the pre-processing required to properly construct the graphs in later stages.

Taking into consideration that the target audience is the government, we want to have a mix of a clear and flexible visualisation that can transmit the message in a robust manner. As it was mentioned in the introduction of this document, the main message to be transmitted is which are the most dangerous states in terms of traffic fatalities in order to give explicit information to the government and assist the organizations in charge of designing policies that can improve the conditions in the states of Texas, California and Florida (which present the highest fatality rate).

With this message in mind, the selected implementation is very convenient because it makes use of different interactive tools from the shiny library which help the user to get a general overview from the beginning and provide extra detail throughout the different tabs, in order to go deeper into the details of each category. With this approach, the message is transmitted with a good balance of complexity and clarity. In the next section we will explain with more details how the visualisation helps the user to go through the different levels of detail in order to get a clear idea of the current situation and give the required information to implement efficient policies based on data.

User Guide

The following User Guide provides the instructions for exploring the narrative visualisation.

1. Visualisation Name

Traffic Fatalities in the US 2015.

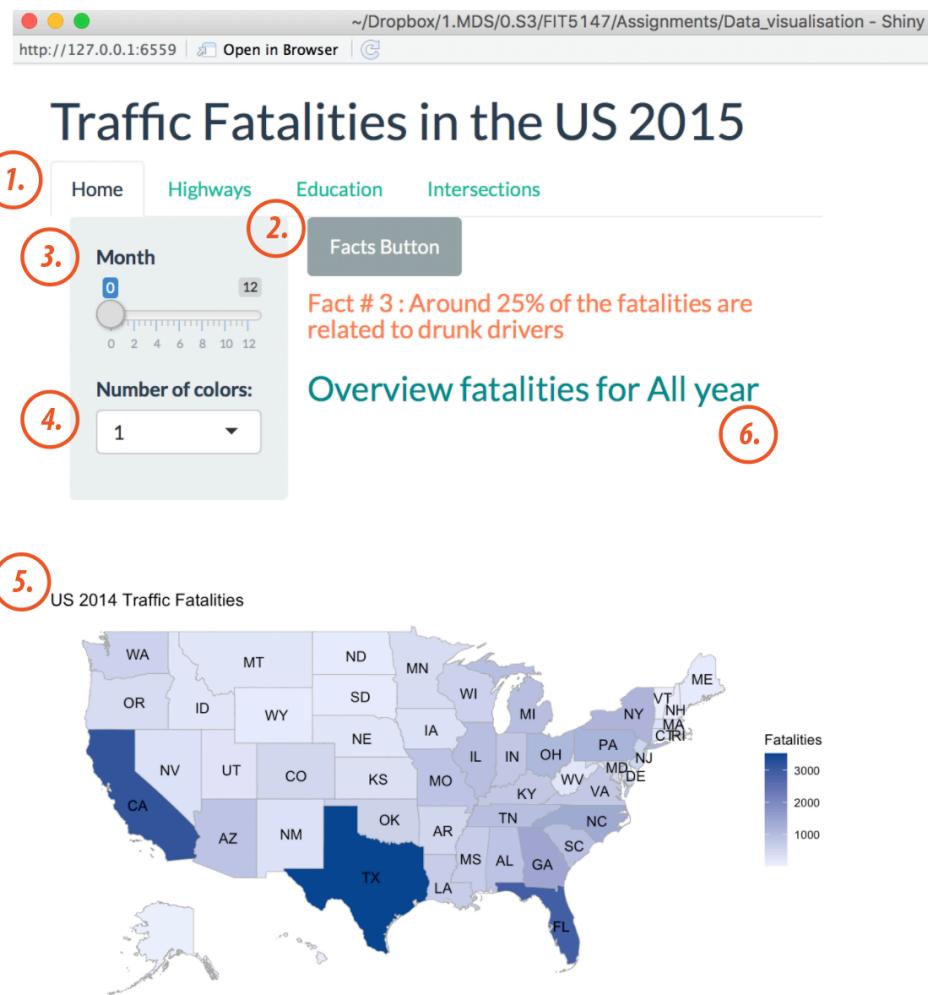
2. Intended Use

The visualisation described in the following user guide is oriented towards the government and aims to provide a big picture of the traffic fatalities in the US for 2015. The main purpose of the visualisation is to exhibit which states are the most problematic in the rates of driving fatalities and supply additional information to implement regulations and policies that allow to improve the safety in the most dangerous spots for these states.

3. UI Description

The *Traffic Fatalities in the US 2015* visualisation is divided in 4 sections as follows:

- a. **Home Section:** In this tab the user will find a choropleth map of the US that provides a general overview of the record of traffic fatalities for 2015. This page contains 6 points that are explain below.



1. Tabs panel with access to other tabs for different levels of detail
2. Facts Button: Provides different facts in order to engage the user with some of the findings of the study.
3. Slider bar to filter by the month or whole year by selecting the option 0

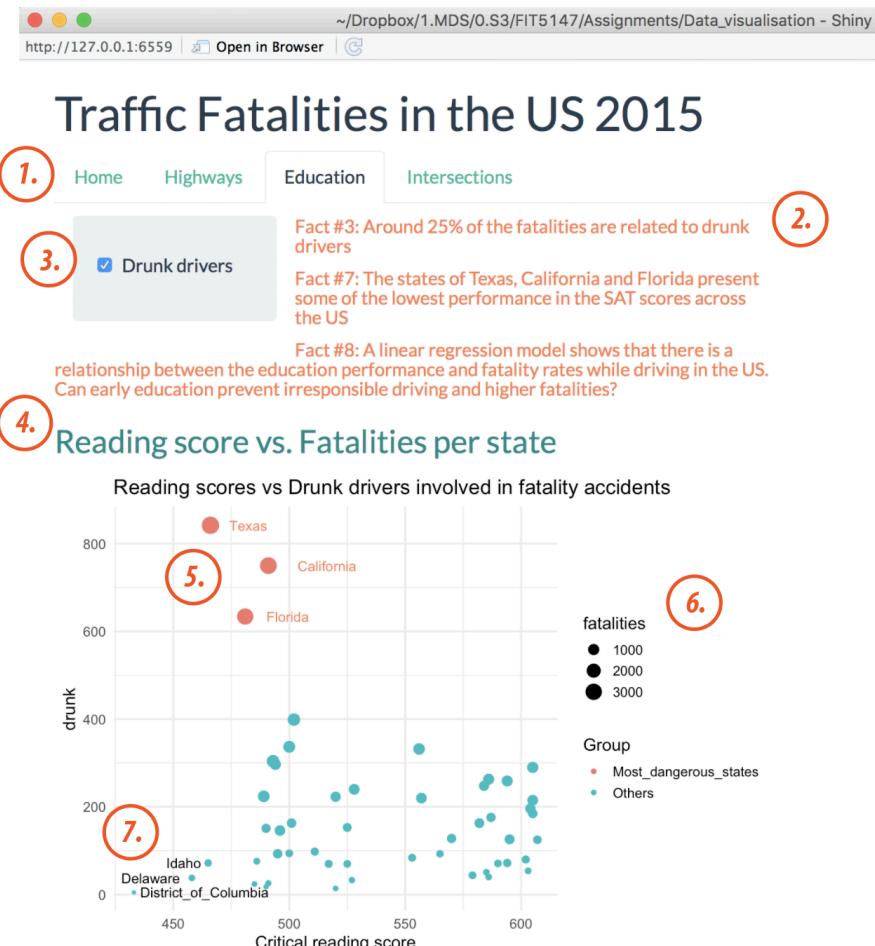
4. Number of colors for the choropleth map. This gives the user the flexibility to use several scales in order to group more states under different ranges for the amount of driving fatalities
5. The Choropleth map adapts to the selection of the user in the steps mentioned above.
6. Interactive title for the section showing the selected month or year.

- b. **Highways Section:** In this tab the user will find all the information related to the specific highways that were identified as the most dangerous in the US. This interface provides the user with several interactive and informative components that are described below the image.



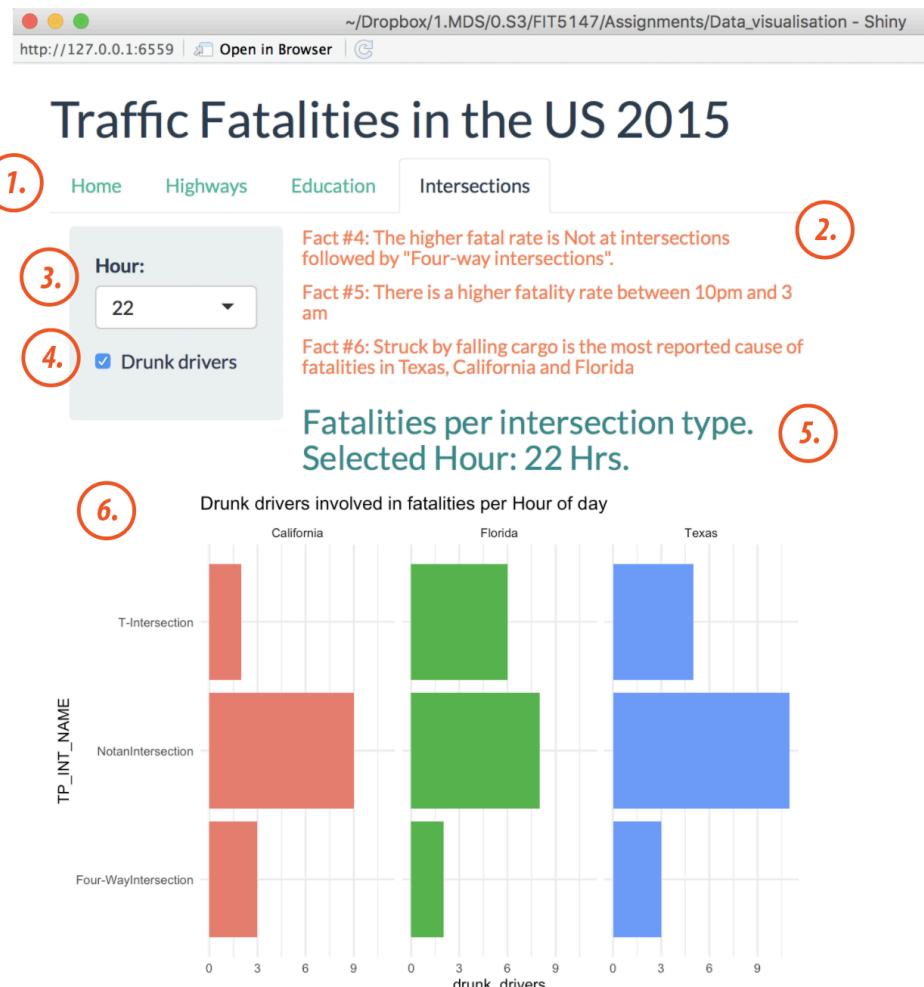
1. Tabset panel again to navigate across the different sections
2. List of the related facts for the Highways section
3. Dropdown menu for filtering the state (or all country)

4. Dropdown menu for selecting different intervals in order to provide more or less detail on the number of fatalities per road
 5. Checkbox to display the labels of the markers in the map with additional information
 6. Interactive title updated according to the selected state or whole country
 7. Leaflet OpenStreetMap with zoom buttons and hover option to show tags of the markers with details
- c. **Education Section:** This tab provides the main insights of the relationship between education and traffic fatalities. The main purpose of this tab is to show to the user how the states of Texas, California and Florida appear to be '*clustered*' in a problematic zone of the graph where the academic performance is low and the drunk driver related fatalities are very high. The main components of the user interface are numerated in the image below and further explained after the image.



1. Tabset panel again to navigate across the different sections
2. List of the related facts for the Highways section
3. Checkbox to include the Drunk drivers effect in the visualisation

- a. Without check: shows the reading score vs. fatalities
 - b. With check: drunk effect is included as y-label and fatalities now becomes an extra dimension (size of the points). Also the labels for states in risk because of low academic performance are displayed
 - 4. Title of the section
 - 5. Cluster of the three most dangerous states with their corresponding label and different colour
 - 6. Legend with the information for colours and and size
 - 7. As mentioned in 3.b, the label for the states at risk for their low academic performance.
- d. **Intersections Section:** The final tab of the visualisation provides the user with specific information related to the intersection type where the fatalities took place and is grouped for the mentioned states (TX, CA, FL). In a similar manner, the following image shows the components of the tab and the explanation is provided below.



1. Tabset panel to navigate across the different sections
2. List of the related facts for the Highways section

3. Dropdown menu for filtering by hour (or full day)
4. Checkbox to display the effect of Drunk drivers only
5. Interactive title updated according to the selected hour of the day
6. Barchart displaying the total fatalities (or drunk related fatalities) per intersection type for each of the 3 problematic states

Conclusion

As a conclusion, this project has provided a visualisation tool aimed to illustrate the most critical states in relation to traffic fatalities in the US. The ultimate goal of the visualisation is to give a better understanding of the critical areas where traffic fatalities are occurring and it tries to provide enough complementary information for the government, as main user of the tool, in order to implement effective policies that help to reduce this records in the most problematic states.

The main findings of the project show that the states of Texas, California and Florida present a much higher amount of traffic fatalities, compared to the rest of the US. In addition, these states present the highest amount of drunk drivers involved in fatalities with a huge gap compared to the rest of the states, and they also present one of the lowest education performance across the country. These two factors are very problematic and should be addressed by the government in order to reduce the amount of traffic fatalities in the long term.

Also, as a short term solution, the information about the most problematic highways in the US was provided. In this map it is possible to observe the exact location where most of the fatalities have occurred and this combined with the information about the most dangerous type of intersection and hour of the day can help the government to implement traffic controls that are highly customized to these problematic areas at specific times of the day, in order to reduce the driving fatalities.

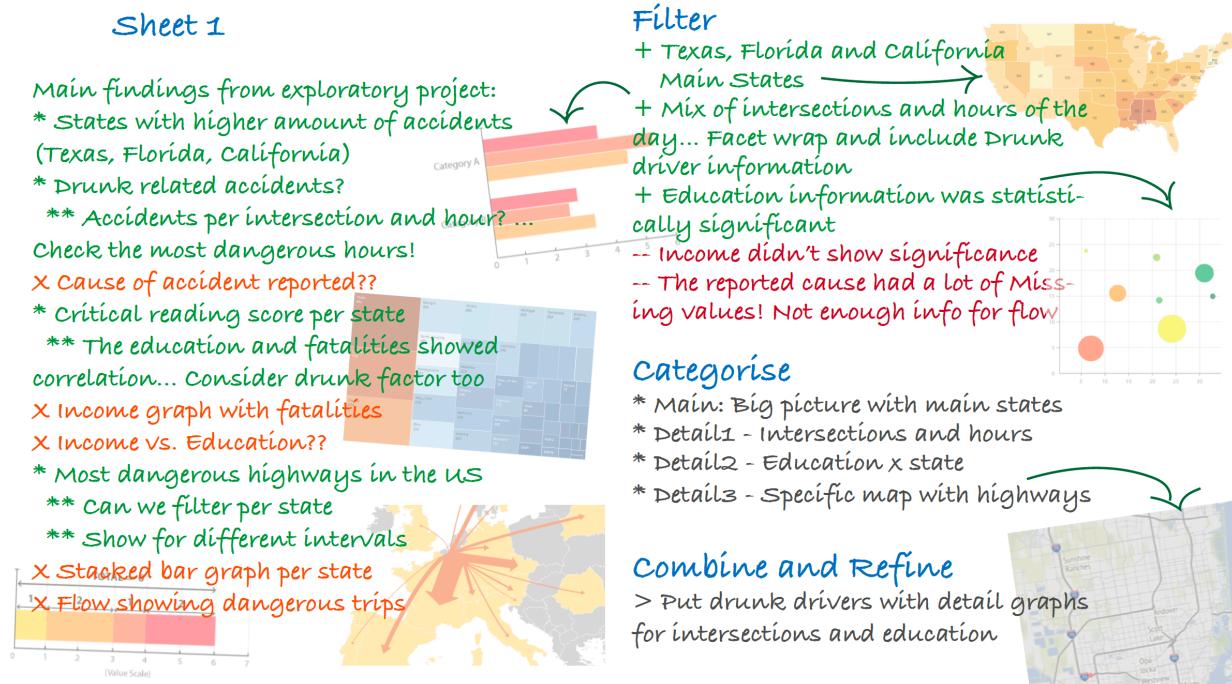
To finalise this section, it is worth mentioning that this project was a good implementation of different interactive tools for visualisation which allowed to expand the use of the graphs in R in order to create a visualisation that can transmit a better message by engaging the user with the data. The use of different libraries allowed the creation of a final tool that combines different types of graphing helping the user to have a better general understanding of the situation and go deeper into some particular details.

As mentioned during the exploratory task, some of the elements that could have been done differently, would be the analysis of some extra information about the different sort of fines drivers receive and the drinking behaviour in the different states of the US. Perhaps some additional information related to the age of the drivers involved in the traffic fatalities could help to establish a stronger relationship between the low education performance at high-school and the high records of fatalities. However, this goes beyond the scope of this initial project.

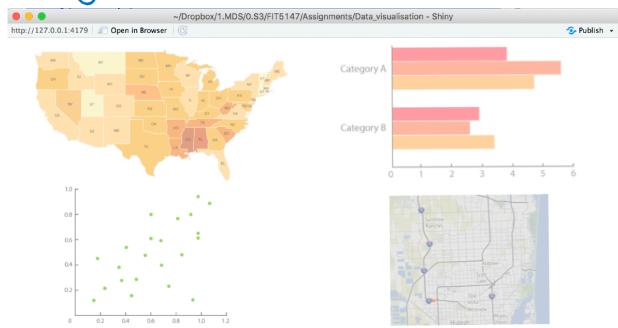
References

- Wagner, I. (2020). Licensed drivers in the U.S. - total number by state 2018. Retrieved from <https://www.statista.com/statistics/198029/total-number-of-us-licensed-drivers-by-state/>
- Golden Oak Research Group LLC, "U.S. Income Database Kaggle". Publication: 5, August 2017. Accessed, 01, April 2020
- Panagiotis D. (2020). Five Design Sheets - A framework for prototyping information visualization interfaces through sketching. Retrieved from <http://pdritsos.com/projects/FDS/>
- The Data visualisation Catalogue. (2020). Retrieved from <https://datavizcatalogue.com/>

Appendix 1 – Five design sheet

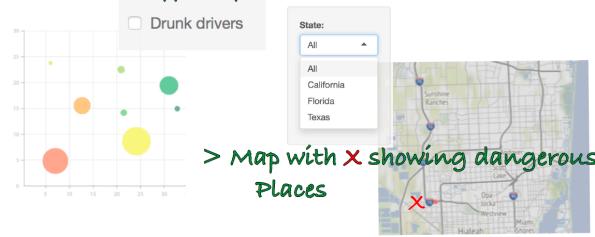


Layout



Focus

> Check the effect of Drunk drivers

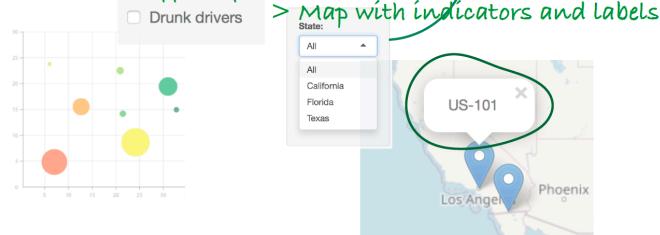


Layout



Focus

> Check the effect of Drunk drivers



Title: USA Traffic Fatalities

Date: 01/06/20 // Sheet: 1 - FDS2

Task: Representation of fatalities with levels of detail in same page

Operations

- * Page with choropleth map, bar graph for intersections, scatter plot education and map of highways
- ** Includes extra detail for all of them
- ** Scroll through the page to visualize the graphs
- *** Add interactive tools in each tab for the user (checkbox, drop menus, Spots marked for the map)

Discussion

- + Multiple graphs displaying different details about the findings of the data
- + General idea at first glance
- + Indicators in the map are not invasive
- The indicators on the map can't provide a lot of information
- The graph could be very heavy with a lot of information displayed at the same time

Title: USA Traffic Fatalities

Date: 01/06/20 // Sheet: 2 - FDS3

Task: Multitab representation of fatalities with the Big picture (Home) and three levels of detail

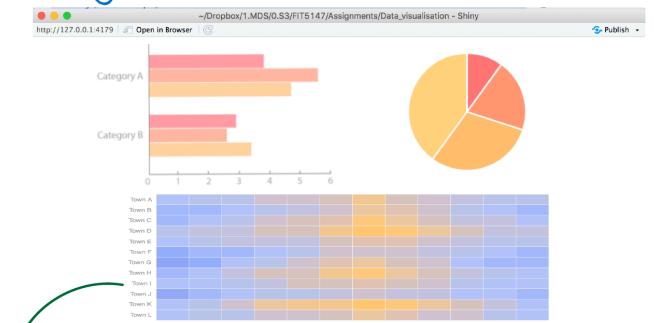
Operations

- * Home page with big picture of the fatalities in the US
- ** Include three levels of detail for the user to select
- ** Separate each detail in an independent tab to make it easier to visualise
- *** Add interactive tools in each tab for the user (checkbox, drop menus, Indicators for the map)
- *** Hover over map indicator for labels with info

Discussion

- + Separation of tabs helps to ease the amount of information displayed
- + Greater detail per tab as they have more space
- + Use of geo-information to detect dangerous roads
- The indicators on the map can be too invasive (filters should be included)
- The graphs could look isolated from the rest

Layout

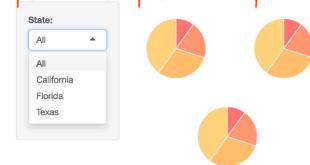


Focus

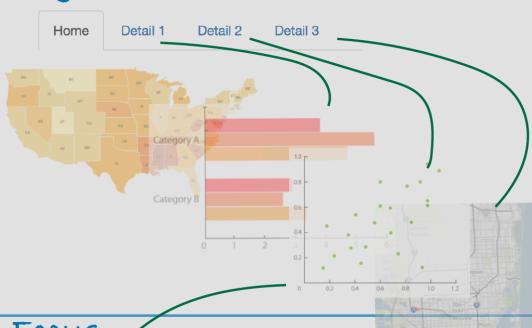
> Check the effect of Drunk drivers



> Specialized pie charts per state



Layout

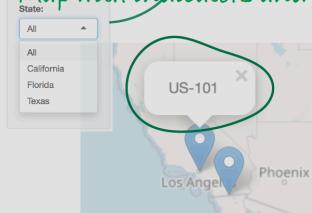


Focus

> Check the effect of Drunk drivers



> Map with indicators and labels



Title: USA Traffic Fatalities

Date: 01/06/20 // Sheet: 3 - FDS4

Task: Traditional line and bar graphs with pie charts

Operations

- * Single page with lighter graphs that are more common and easier to read
- ** Heat maps expanded to show the drunk drivers effect
- ** Separate Pie charts to make it easier to visualise for each state
- ** Add interactive tools for the user (checkbox, drop menus)

Discussion

- + Typical graphs that are easy to read
- + Without the maps content the canvas looks less crowded
- Geo-information is lost without the maps to show the relevance for the highways
- Still a lot of content to display in the zoom options which makes the graphs overwhelming

Title: USA Traffic Fatalities

Date: 01/06/20 // Sheet: 2 - FDS3

Task: Multitab representation of fatalities with the Big picture (Home) and three levels of detail

Operations

- * Home page with big picture of the fatalities in the US
- ** Include three levels of detail for the user to select
- ** Separate each detail in an independent tab to make it easier to visualise
- *** Add interactive tools in each tab for the user (checkbox, drop menus, Indicators for the map)
- *** Hover over map indicator for labels with info

Detail

- + Use Choropleth map for the Home menu which displays the big picture of the dataset (TX, CA, FL)
- + Implement filtering to display the main states
- + ggplot facet-wrap to include the fatalities per hour for all the main states
- + Checkbox, dropdown menu and multiple tabs
- + Leaflet maps with labels to display details
- + shiny, ggplot2, dplyr, leaflet as main libraries

Appendix 2 - Description of data sources

1. Tabular data: 32K rows x 52 columns with spatial and temporal attributes. Several attributes contain codes that need to be searched in the user's manual. (<https://data.world/transportation/2015-traffic-fatalities>)
2. Tabular data in xls with info for each state in 56 rows and 19 columns. The tables contain multiple line heading so a pre-processing task for reshaping must be done. The data contains some additional text at

the end. (https://nces.ed.gov/programs/digest/d16/tables/dt16_226.40.asp)

3. Tabular data with 32K rows and 19 columns. Geo locations, spatial data.

(https://www.kaggle.com/goldenoakresearch/us-household-income-stats-geo-locations#kaggle_income.csv)