**Team Mayhem**

**Our Goal:**

To analyze data to answer the following questions:

1. Have automobile accidents increased with the adoption of mobile phones and smartphones?
2. When were smartphones introduced and what was the rate of adoption?
3. What was the pace of adoption of mobile apps since the onset of the smartphones?
4. Did the rate of automobile fatalities change as a result of the introduction of the smartphone?
5. Did the rate of automobile fatalities change as mobile apps become more prevalent?
6. Has the passing of mobile legislation in states affected the incidence of automobile accidents?
7. Do more densely populated states have higher fatality rates, in general? Do mobile phone adoption rates impact these states more? Meaning more accidents per capita? Per registered voter?

* What impact on accidents has legislation regarding cell phone usage and driving accomplished?

* Is there a significant difference in accident numbers by age?
* Is there a possible correlation with messaging apps and possible correlation with accidents?
* Compare types of accidents versus before the smartphone
* Find the release dates of top messaging apps and see if there is a significant change in total accidents with the adoption of smartphones allowing for messaging applications to be more widely used?

(FB Messenger, Snapchat, WhatsApp, Messenger by Google, Discord - Chat for Games, Google Hangouts, GroupMe, Kik, Skype, Telegram, LINE, WeChat)

**Hypothesis:** U.S. automobile accidents have increased significantly with the introduction of smartphones.

**Data / Datasets:**

1. U.S. automobile accidents
2. SMS and Messaging App release dates market penetration over time

Originally, we were Focusing on:  CA, TX, NY, NH, CO, WY

However, in our initial analysis we found that some states we started with had minimal data that was accessible. Also, that if the state was more Rural there wasn’t much Data.

So, we altered our Focus to the most densely populated states and Colorado but made it possible for 2 additional states to be added to the FARS data pulled by the user of the notebook. CA, FL, IL, NY, OH, PA, TX and CO

**Data Cleanup & Exploration**

Methods of cleanup used: Merging, converting type of data, creating data-frames, manual entry

of data into tables from hard copies (e.g. state numbers from NHTSA site).

**Insights:**

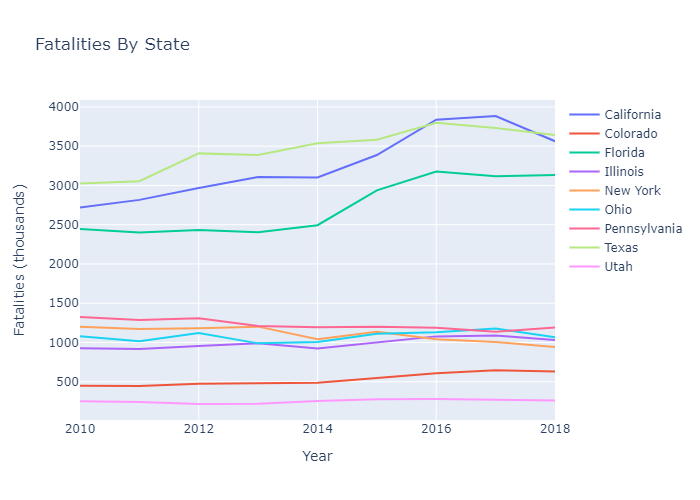
We found that to gather some of the data was either tedious or in a format that would take hours or days to render as they were in formats that we were unable to utilize such as jpeg or PDF formats

**Problems encountered and Resolutions:**

Accessing data from the FARS site was limited to 5,000 requests at a time. This made calling all accidents difficult and very time consuming. This contributed to why we focused on total fatalities per state versus all accidents. We made the assumption that the percent of accidents that are fatalities translates across states (though population density may impact this). Also impacted the number of states’ data we pulled

**Interesting Figures**:

Present and discuss interesting figures developed during exploration, ideally with the help of Jupyter Notebook



The most densely populated states are not at the top of the fatalities list this is most likely because there are better options for public transportation.

**Data Analysis:**

(Discuss the steps you took to analyze the data and answer each question you asked in your proposal)

**API:**

In the API first obstacle that had to be overcome is building a dictionary of states to interface with the FARS API to allow for Multiple states as each state was assigned a number to pull from their database. Once that was complete getting the state number to coincide with the yearly report proved difficult at first which required a nested for loop with each one having it’s own iterator to allow for the years and stated to be iterated separately. Which took reading the json at each step. Once that was accomplished merging that data collected into a usable format at first we used a bunch of series to collect the data but found this extremely inefficient so we then built a dictionary with the API call and built our Data Frame using the .from\_dict function of pandas.

**Findings:**

(Discuss your findings. Did you find what you expected to find? If not, why not? What inferences or general conclusions can you draw from your analysis?)

Katherine – work on PPT, get other state law data, adoption thresholds

Martin – circling with with Zach, plotting FARs data.

Zach – adjusting Jupyter Notebook to yield CSVs per state containing fatalities for each state by year.

Ben – collect licensed drivers per state, put into dataframe, plot drivers per year per state