

**Team Proposal**

Our team is made up of professionals with varying backgrounds enabling us to investigate and study issues from multiple perspectives. In conducting investigative research we use a multifaceted approach allowing us to glean insights from data and translate those insights into easily communicated actionable information.

By combining state of the art data extraction, transformation, loading, and harmonizing algorithms we are able to create an end product which allows for a given hypothesis to be confidently proven or disproven. Effective data collection and analysis combined with well designed and communicated visualizations allow our audiences to make decisions with a high degree of certainty.

The professionals making up our team are:

* Richard Moses
* Dawayne Eason
* Raaghaven Krishnamurthy
* Linh Nguyen
* Geoff King

**What Drives Us**

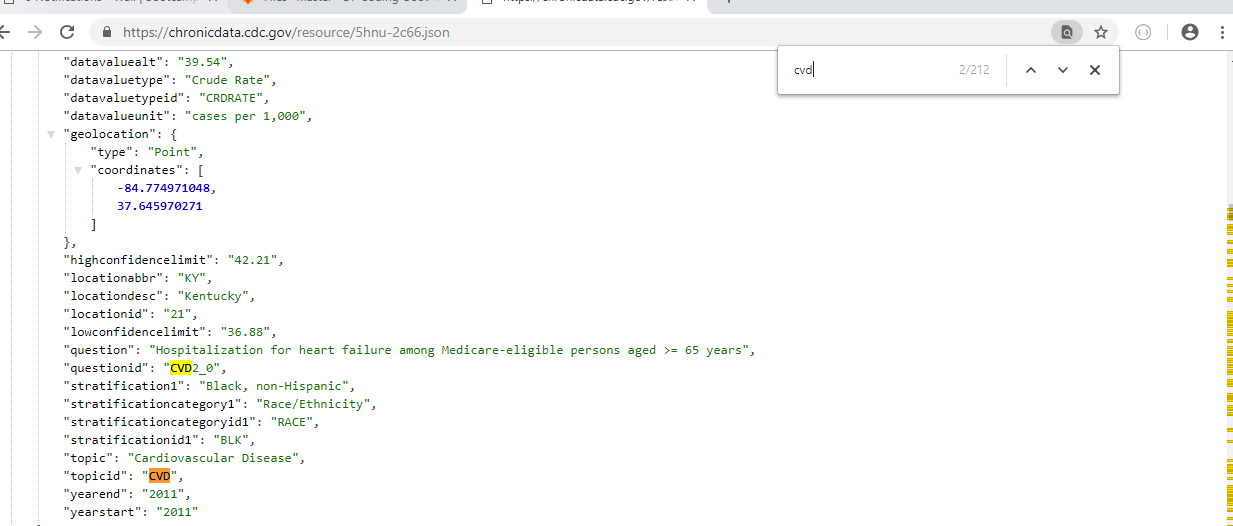
Cardiovascular disease or more commonly known as Heart disease affects the lives of millions of people in the United States with over half a million people dying of the disease in 2017. Sadly many people live their lives without knowing they have the condition or not realizing the seriousness of the condition.

**Proposal**

Our team’s proposal is to investigate the rate of heart disease across the United States and present information related to the areas of the country where the condition is most prevalent. Our plan is to break the data down by gender, ethnicity, age, and location. We will also include information on the different types of heard disease and important risk factors

**Data Sources**

The data sources utilized for this study will be based on freely available data collected by the United States government from sources like CDC.gov, HealthData.gov, HHS.gov and Data.gov. By using these sources our study will not only help keep costs down, but will also utilize the scale of highly reliable data collected by the United States government. Data concerning cardiovascular disease, demographics, and geographical information will be easily accessible by using government resources.



[**https://chronicdata.cdc.gov/resource/5hnu-2c66.json**](https://chronicdata.cdc.gov/resource/5hnu-2c66.json)

**Methods of Extraction**

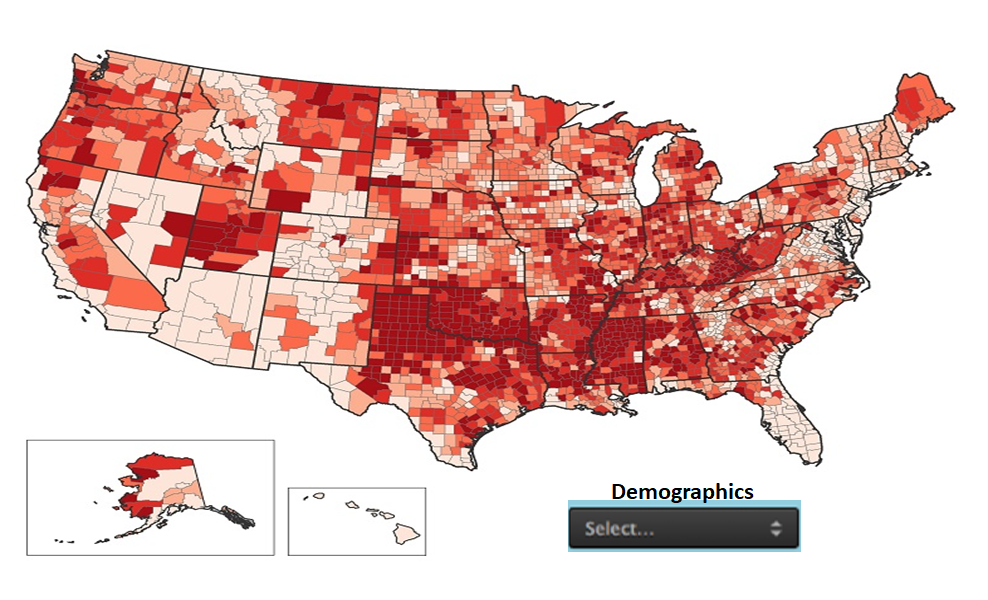
The methods of extraction we will employ will be based on exploiting API and flat file data freely available from the United States government. Whether by making API calls to retrieve data or by downloading data stored in JSON or CSV files, we will retrieve the data sample sizes necessary to appropriately execute our investigation.

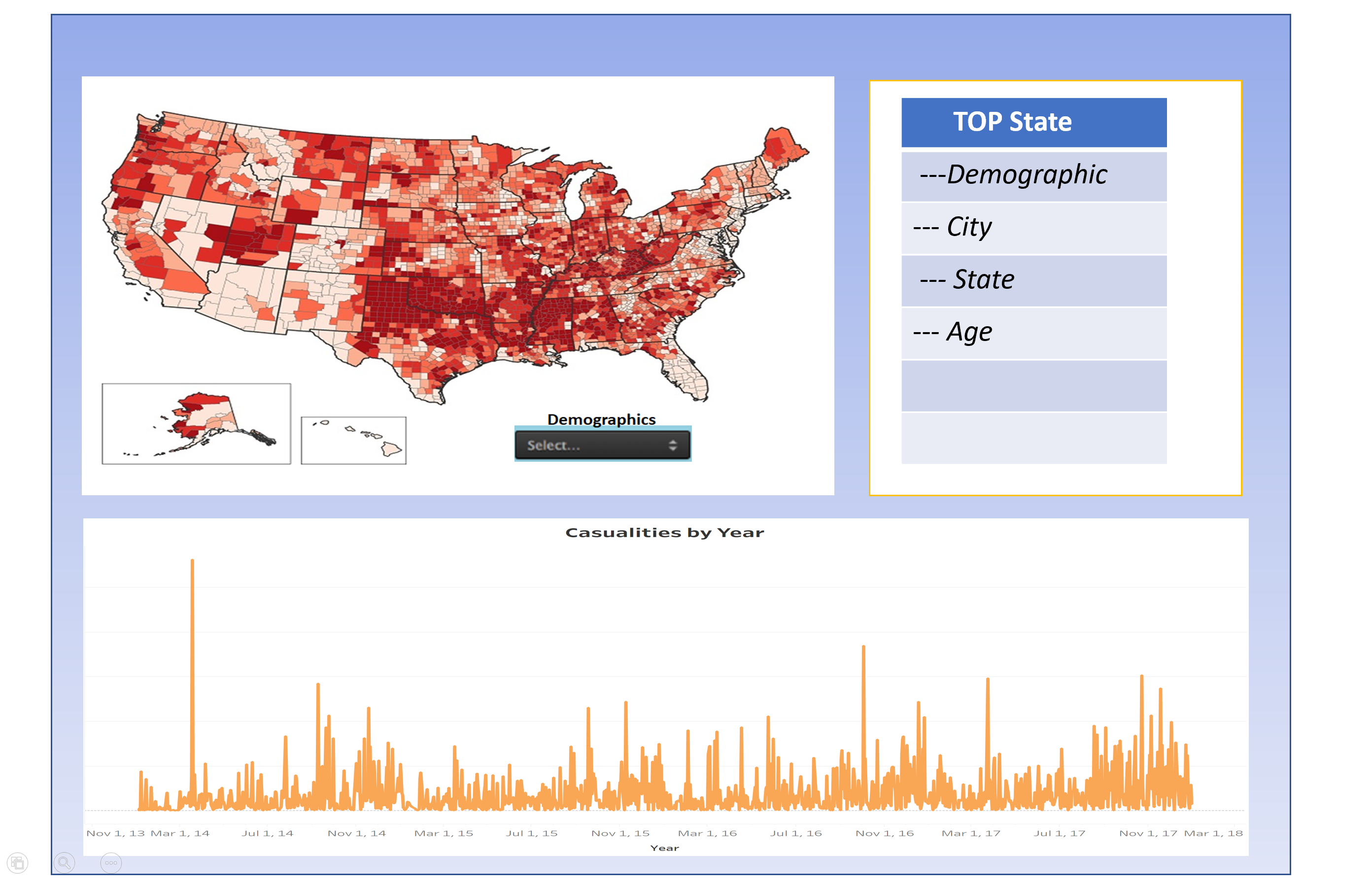
**Methods of Analysis**

By using state of the art data analysis technologies like Jupyter Notebook, Python, Pandas, Numpy, Matplotlib, and MongoDB we will be able to employ coding techniques allowing us to quickly and reliably analyze the data we collect. Once the data has been analyzed we will be able to form the insightful conclusions we believe will save lives.

**Types of Visualizations**

In order to save lives we believe it is our responsibility to effectively communicate information in a manner that is easy to understand. We believe the best way to communicate the results of our study is to use well-designed interactive visualizations. Our efforts at communication will use Choropleth maps, various standard data science based plotting techniques, and a website designed to be easy to navigate. Please see the image below for a preview of one of the visualizations.





**Git Hub Repo Link:**

[**https://github.com/lnguyen1021/BraveHeart**](https://github.com/lnguyen1021/BraveHeart)

**Conclusions**

As stated before, we believe the outcome of our study will be to safe lives. By identifying the areas of the country where heart disease is most prevalent, breaking the data down by demographic information, and highlighting certain risk factors our goal will be achieved.