**Hospital Finder**

**Correlation between diseases caused and neighboring hospitals**

**Team Name:** Titans

**Team Information:**

1. Ankita Patil (asp160730)
2. Krupali Patel (kxp160630)
3. Shivdev Kumar (sxt160530)

**Project Type: Custom**

1. **Introduction**

The project aims to find the top five leading causes of death and once we find it, we address one of leading causes of death and suggest hospitals near to a user for that particular disease so that they can be cured if possible and we can reduce the death rate for that particular disease. For this purpose, data about "Potentially Excess Deaths from the Five Leading Causes of Death(NCHS)" **from National** Vital Statistics System is used and combined with the "Timely and Effective Care- Hospital" and "500 Cities: Local Data for Better Health" for the year 2014.

The top five causes of death

* Unintentional Injury
* Heart Disease
* Stroke
* Cancer
* Chronic Lower Respiratory Disease

In this project, among 5 top five causes of death, we are focusing on ‘Chronic Lower Respiratory Disease’.

1. **Target Audience**

The information obtained from Hospital Finder can be used by general population of the United States to locate nearby specialized hospitals based on their location within a particular radius, so that they can get respective treatments in quickest possible way.

The project can also be used by government officials to address the mismatch between excessive count of deaths and scarcity of hospital for a state. This statistic can help government to better allocate their health funds. With leading cause of death statistic, the government can focus on facilitating more hospitals to cure leading diseases.

1. **Description of Data Sources**
2. **Timely and Effective Care – Hospital**

**Link:** [**https://catalog.data.gov/dataset/timely-and-effective-care-hospital-e4aad**](https://catalog.data.gov/dataset/timely-and-effective-care-hospital-e4aad)

**Domain: Federal**

**Description: This data set includes provider-level data for measures of heart attack care, heart failure care, pneumonia care, surgical care, emergency department care, preventive care, children asthma care, blood clot prevention and treatment, pregnancy and delivery care, and cancer care.**

**Format available: CSV, RDF, JSON, XSL**

1. **Potentially Excess Deaths from the Five Leading Causes of Death**

**Link:**[**https://catalog.data.gov/dataset/nchs-potentially-excess-deaths-from-the-five-leading-causes-of-death**](https://catalog.data.gov/dataset/nchs-potentially-excess-deaths-from-the-five-leading-causes-of-death)

**Domain: Federal**

**Description: This dataset contains Potentially Excess Deaths from the Five Leading Causes of Death in Nonmetropolitan and Metropolitan Areas, United States, 2005-2015. Mortality data for U.S. residents come from the National Vital Statistics System.**

**Format available: CSV, RDF, JSON, XSL**

1. **500 Cities: Local Data for Better Health**

**Link:** <https://catalog.data.gov/dataset/500-cities-local-data-for-better-health-b32fd>

**Domain: Federal**

**Description: This is the complete dataset for the 500 Cities project. This dataset includes 2013, 2014 model-based small area estimates for 27 measures of chronic disease related to unhealthy behaviors, health outcomes and use of preventive services. It includes estimates for the 500 largest US cities and approximately 28,000 census tracts within these cities. These estimates can be used to identify emerging health problems and to inform development and implementation of effective, targeted public health prevention activities.**

**Format available: CSV, RDF, JSON, XSL**

1. **Data Integration**

After getting data from completely different sources, our aim is to extract relevant real-world patterns.

1. Relationship between state-wise population and no. of hospitals
2. State-wise count of death
3. Nearest Specialized Hospital for a particular chronic disease

5. **Data product results**

Below are the screenshots of the visualization made by analyzing and correlating all the three datasets using Google visualization API

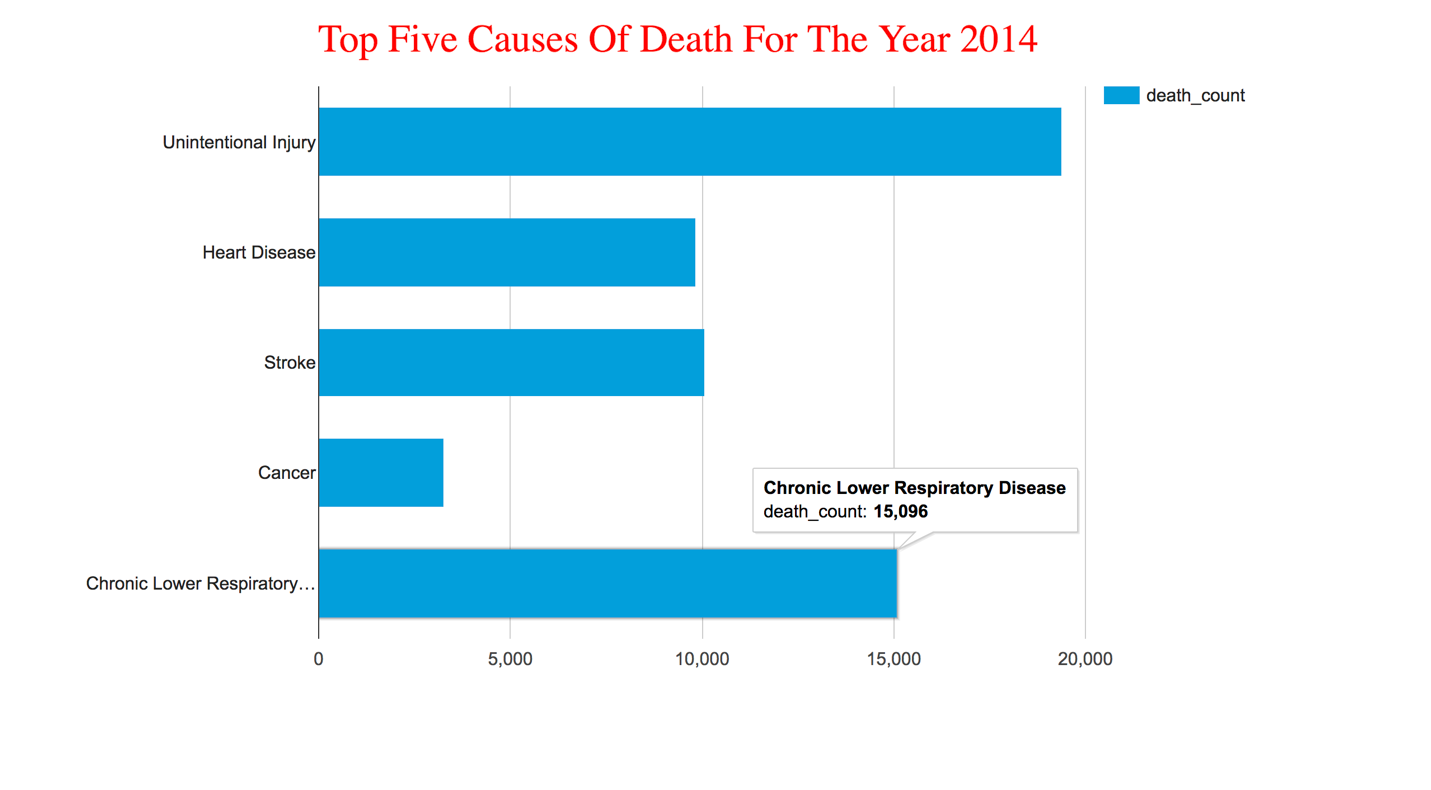


Fig-1. Top Five Causes of Death for the year 2014 in the USA

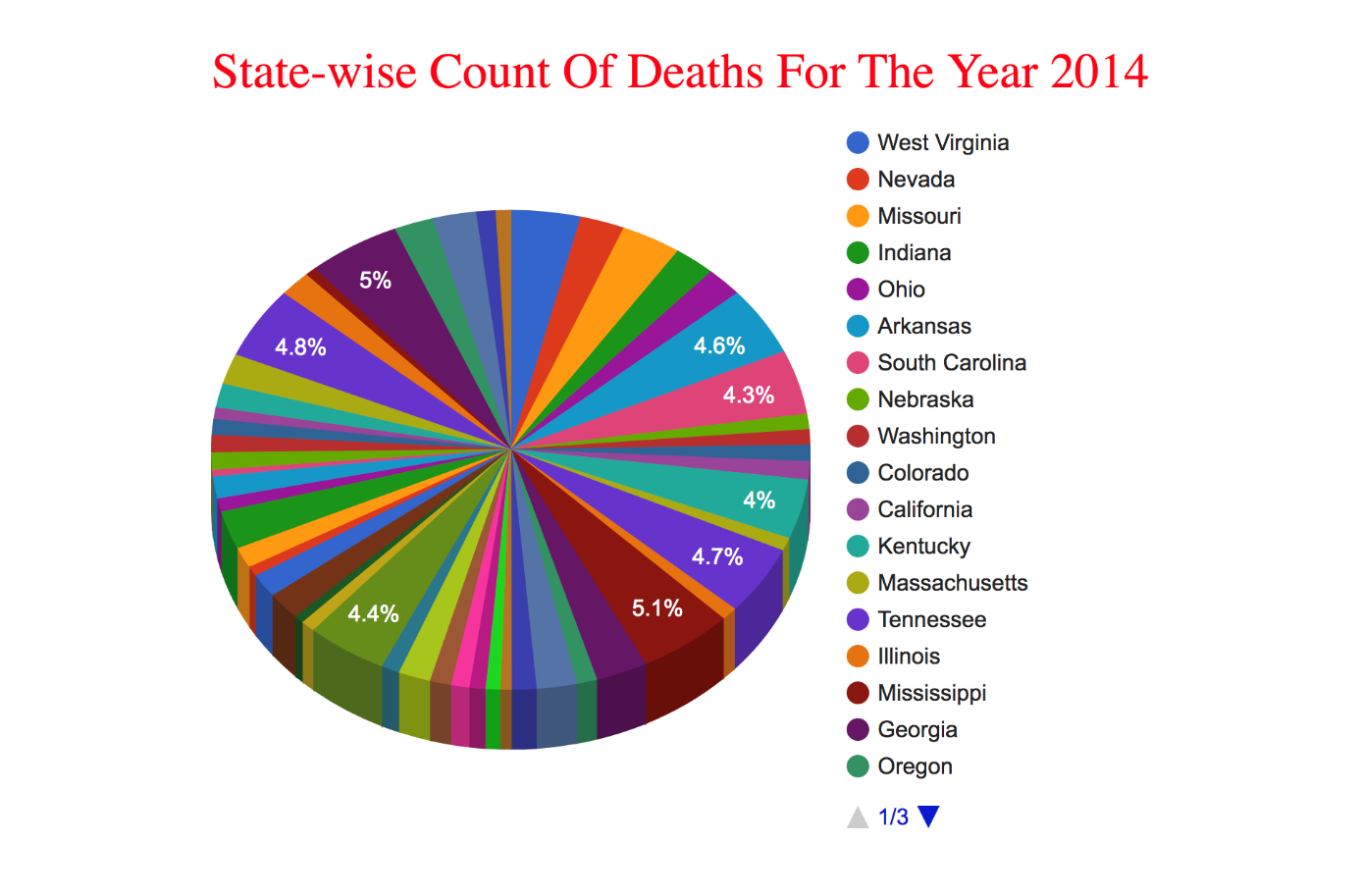


Fig 2. State-wise count of deaths for the year 2014 in the USA

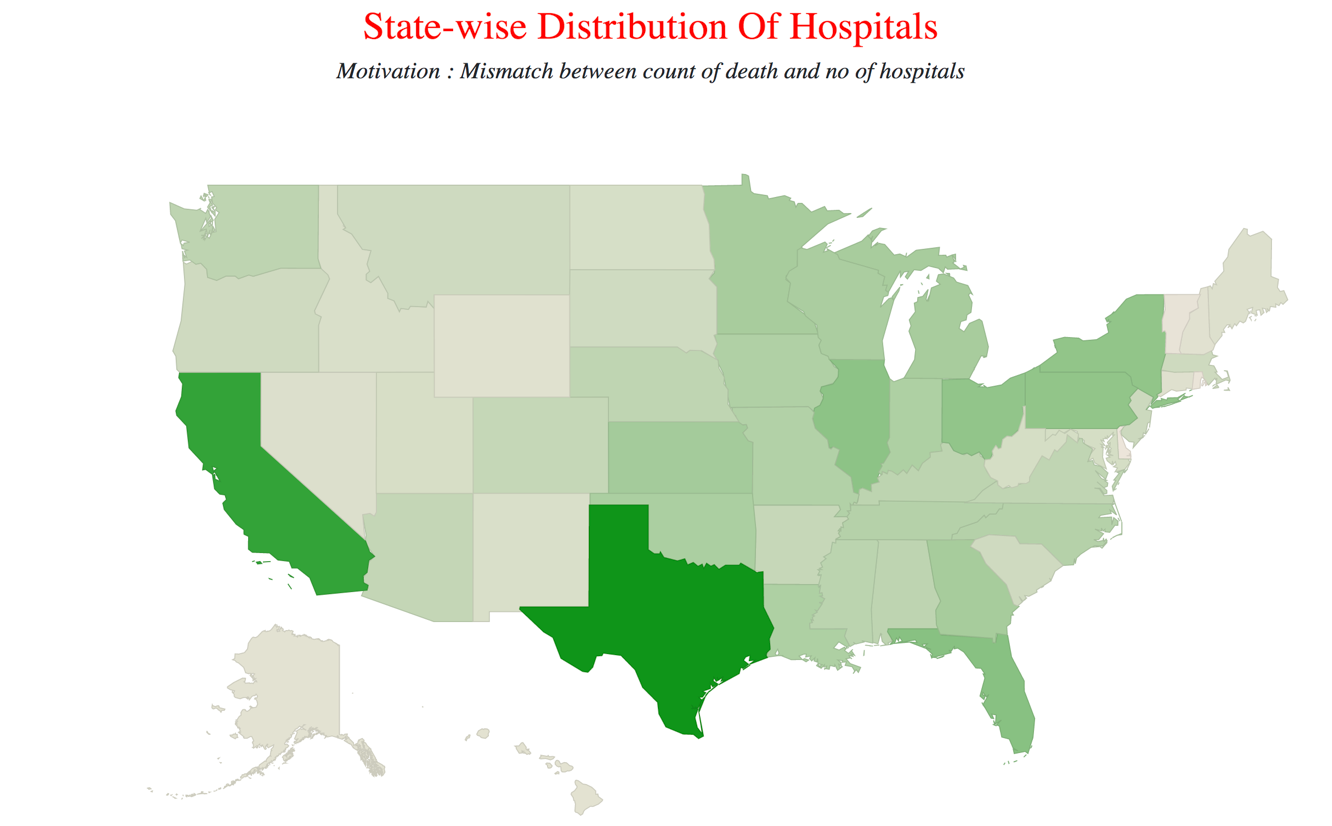


Fig 3. State-wise Distribution of Hospitals

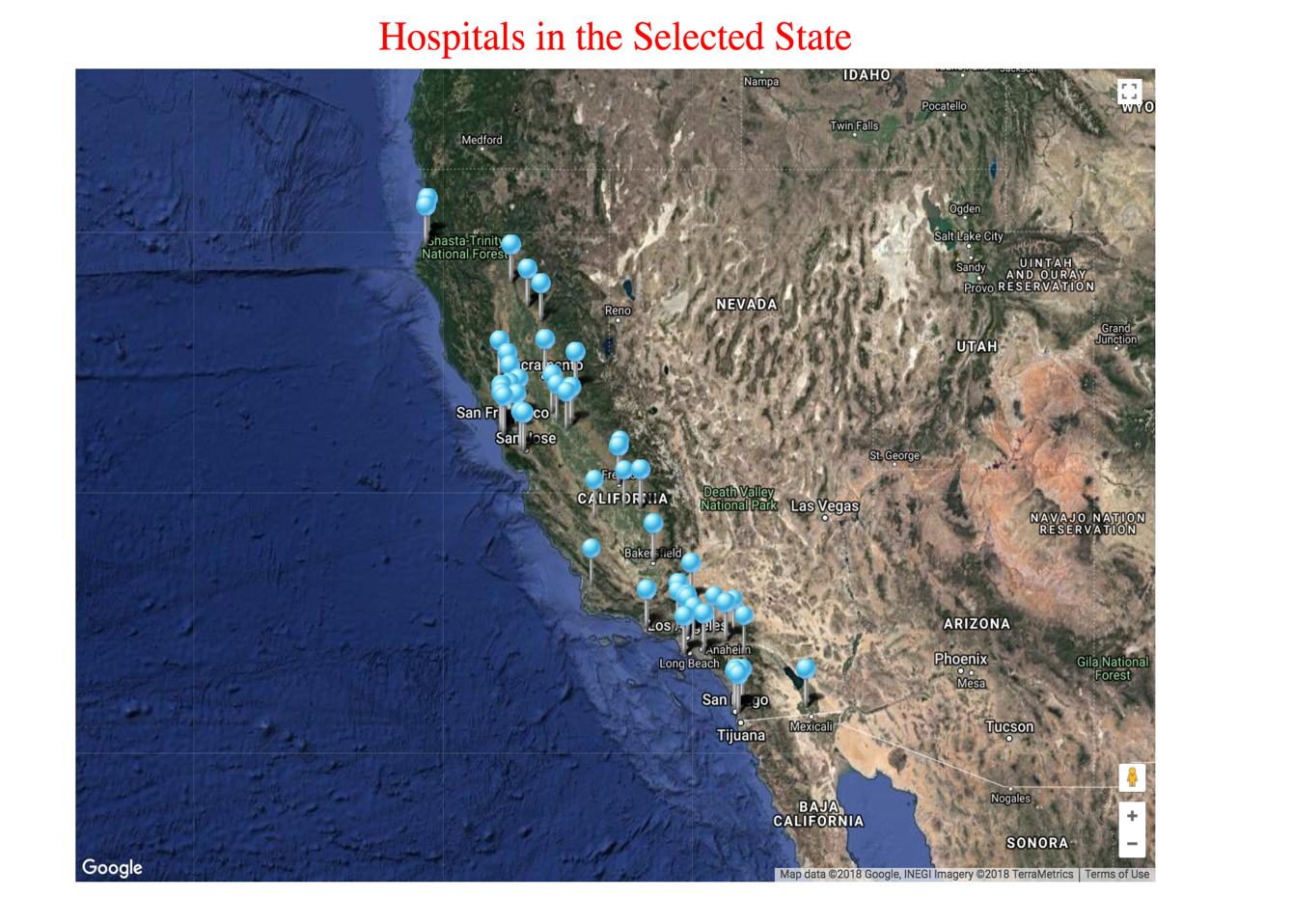


Fig 4. Location of all the hospitals in the selected state (California)

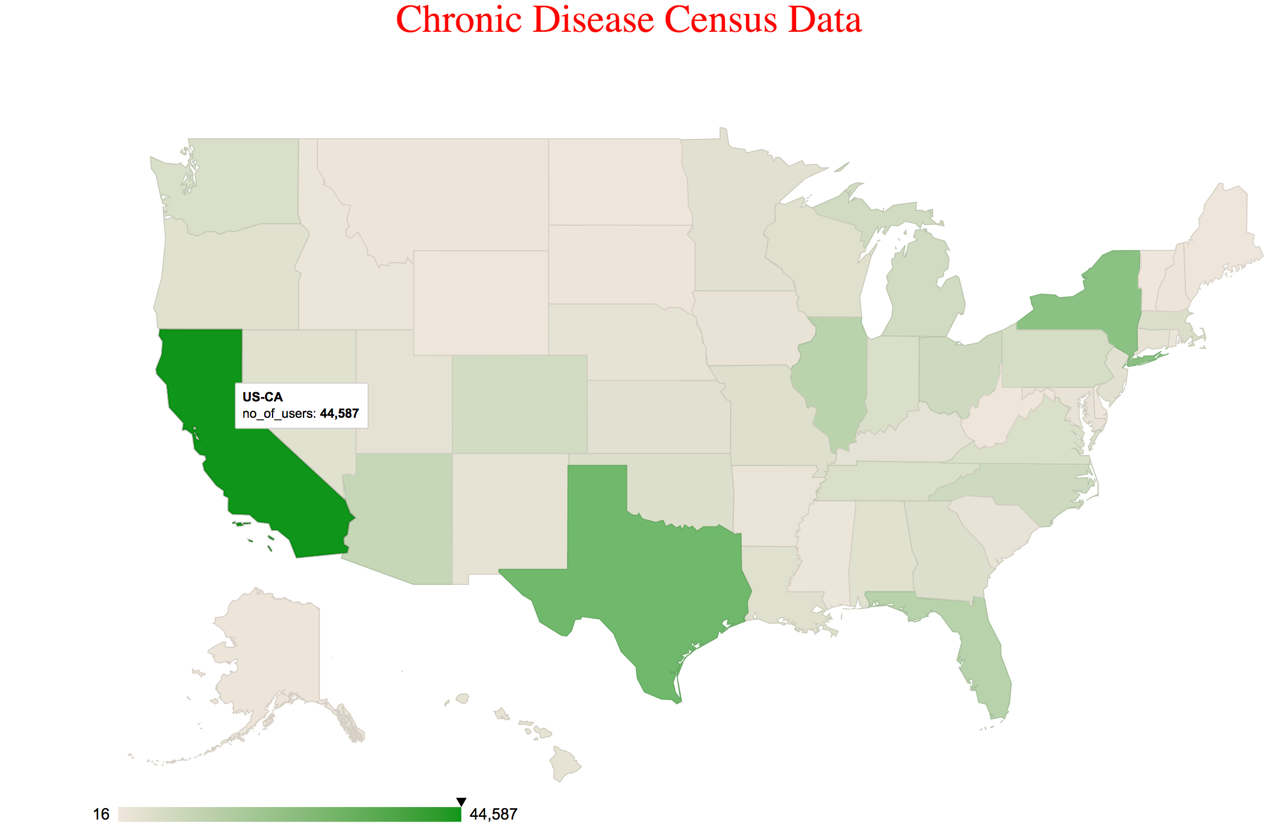


Fig 5. Chronic Disease Census Data for the year 2014 in the USA

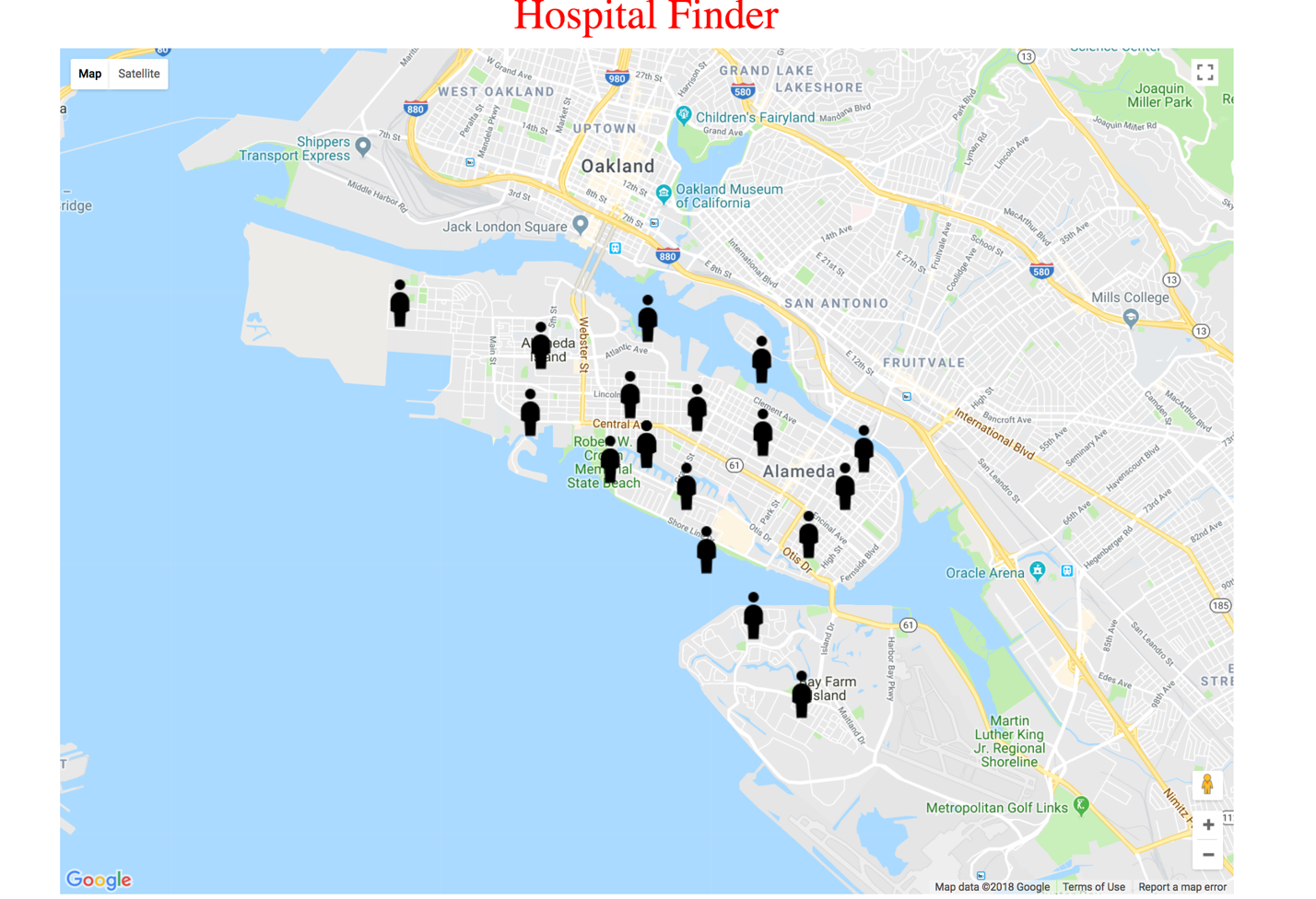


Fig 6. Location of the people who took census

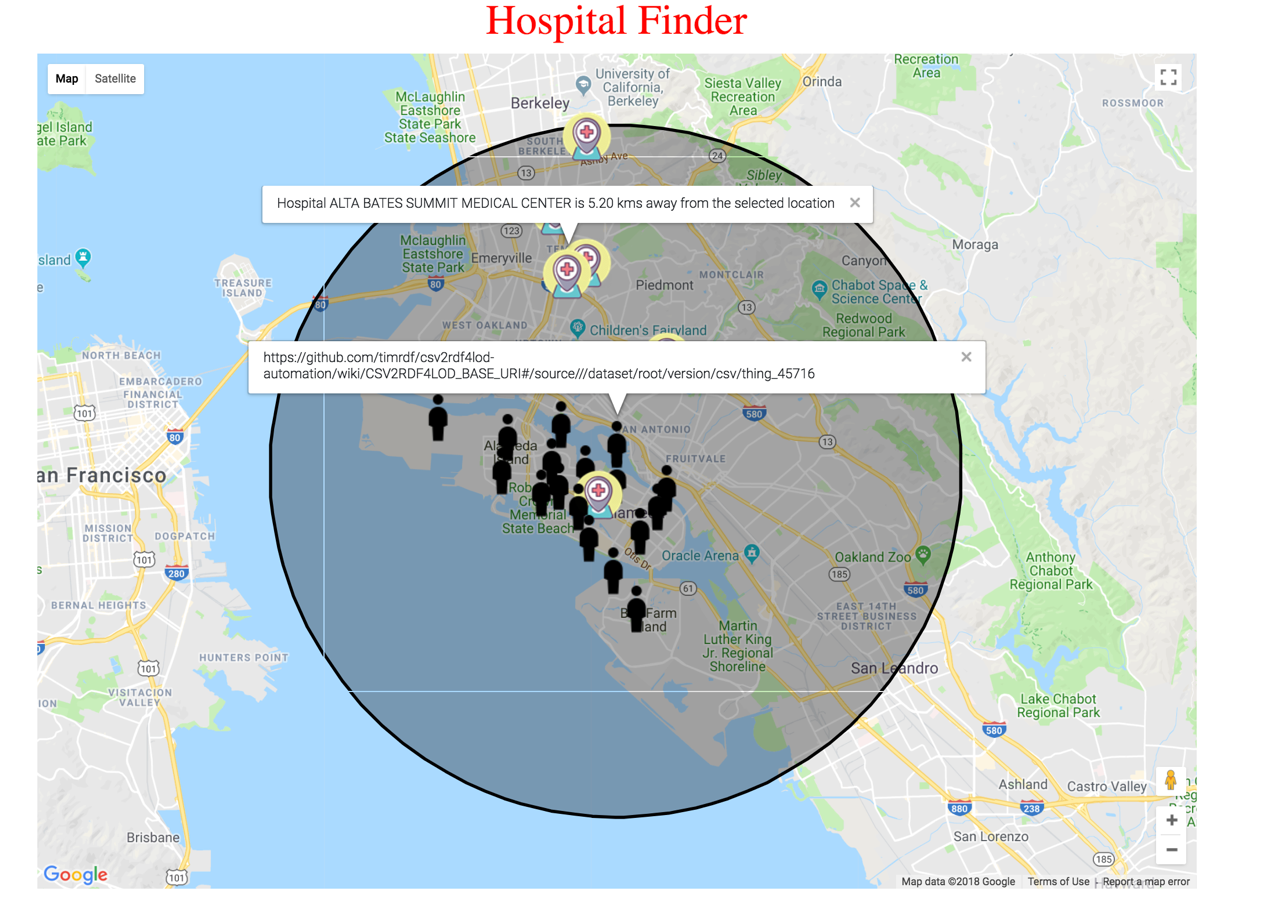


Fig 7. Location of the hospital within 10 km radius from the selected user

1. **Custom Project Justification**

With a single dataset, we could not derive meaningful patterns. Hence, we switched to a custom project to integrate multiple datasets. In this project, we have used three datasets –

- Timely and Effective Care - Hospital

* Potentially Excess Deaths from the Five Leading Causes of Death
* 500 Cities: Local Data for Better Health

where we have analyzed these datasets, and provided comparison between number of hospitals and number of deaths and we have provided a utility to the user to locate nearby hospitals.

1. **Summary**

This project uses three datasets and the datasets are taken from data.gov web portal. They are integrated on the basis of the common attributes such as states, latitude, longitude of users and hospitals. SPARQL is used to query on these data to present user-friendly, easy and readable visualized graph and chart. This project is useful for common people who want to find hospitals which is near to the location he stays. In future, this project can be extended to address all the five causes of death.

**Movie Link:** <https://utdallas.app.box.com/file/289076411668>

**Note:** We have submitted the video in the submission folder (because of the file size constraint, we have compressed it. So, you might experience video with poor quality). For better video quality, please visit <https://utdallas.app.box.com/file/289076411668>