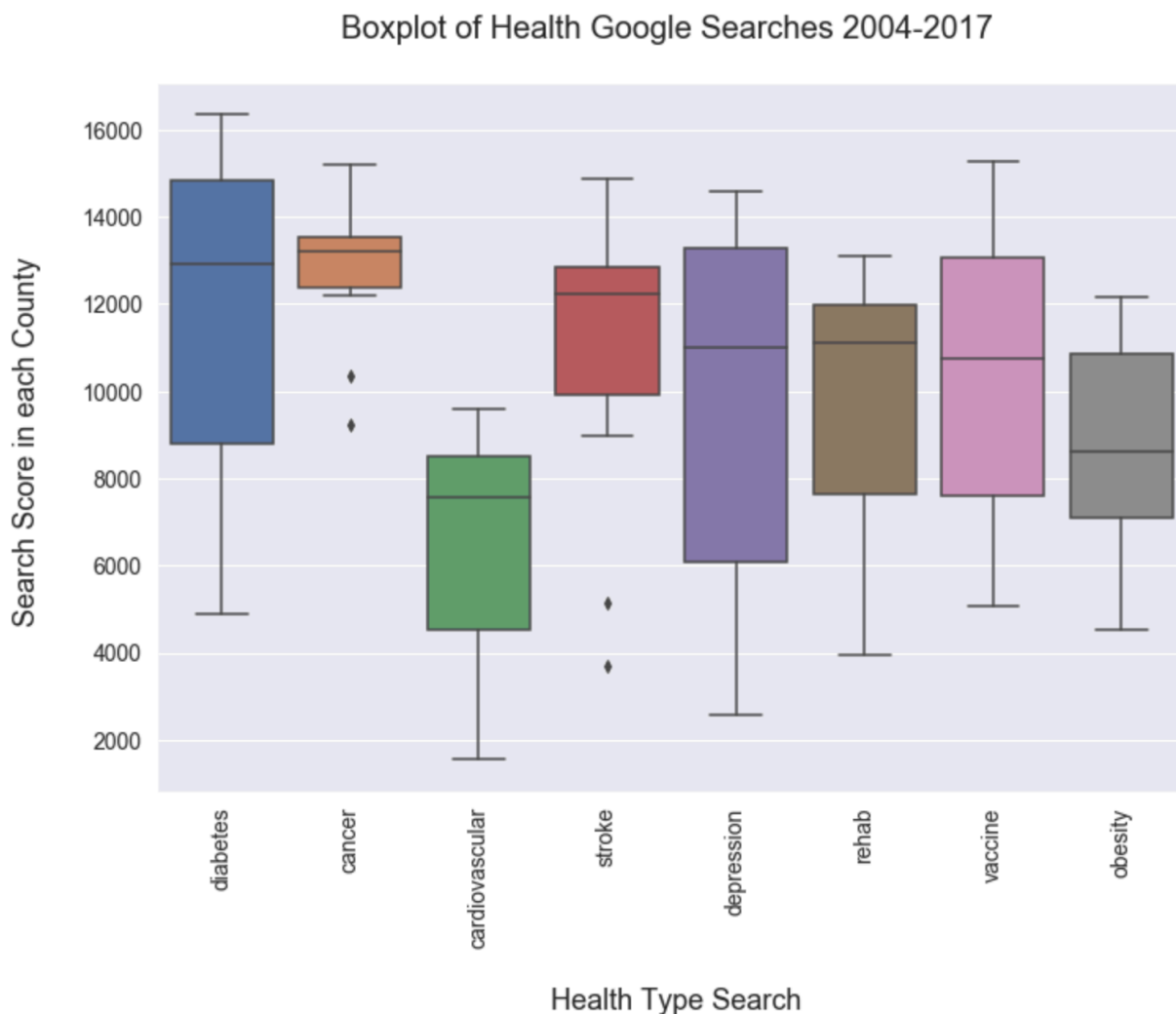


## Introduction

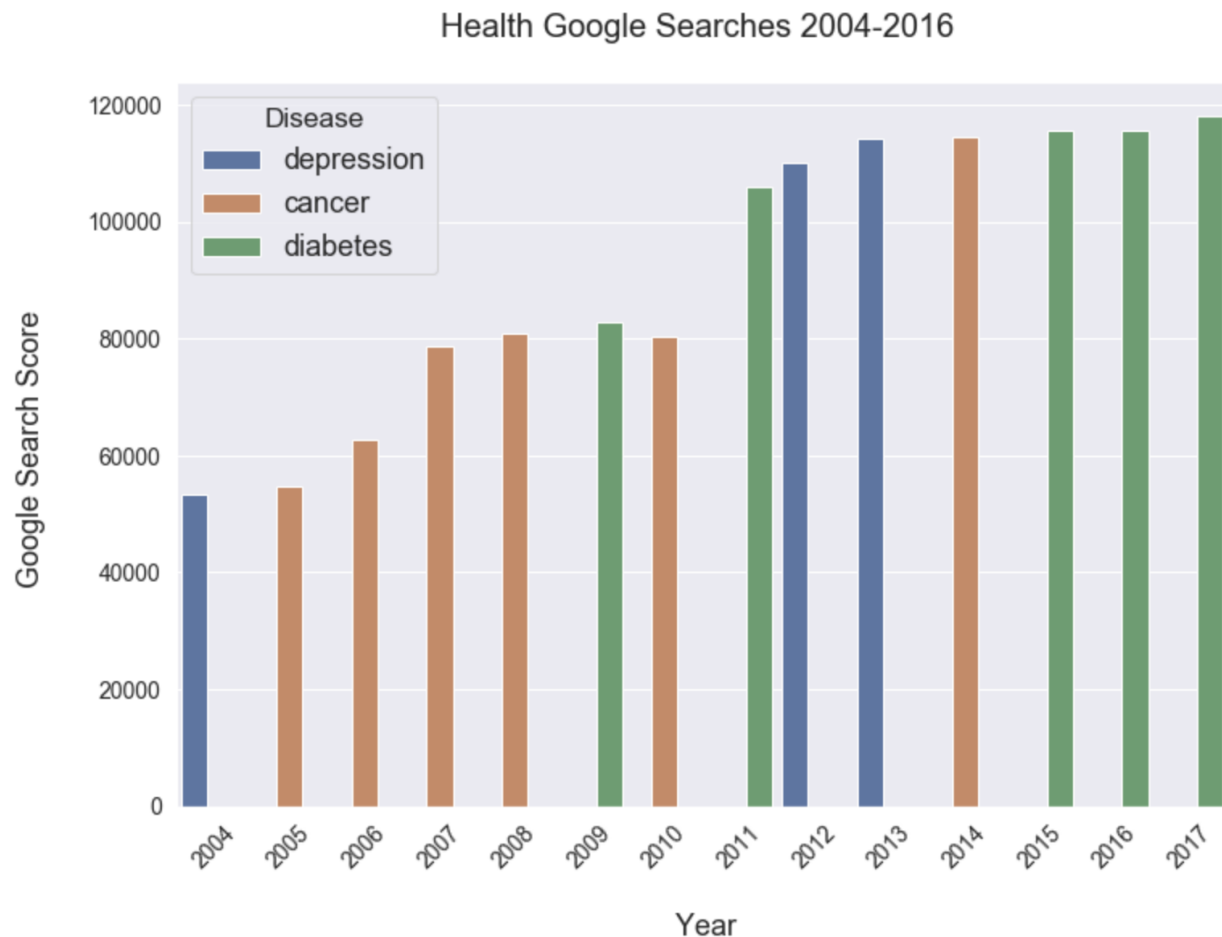
In the data visualization class at the University of San Francisco's Master's in Data Science Program, I was tasked with choosing a dataset to present visualizations. Healthcare accounts for 17.9% of the US economy, and searching healthcare related data sets revealed "Health Searches by US Metropolitan Area" on the Kaggle link at <https://www.kaggle.com/GoogleNewsLab/health-searches-us-county>. This dataset looks at Google Search data and measures the frequency of searches by the public of specific disease categories at the local level. These disease categories include common health issues in the United States, such as Cancer to Diabetes, and reflect the health care information gathering of millions of Americans. The health conditions were selected from the Community Health Status Indicators (CHSI) which provides key metrics for local communities in the United States, and overall trends across the country.

## Boxplot



This boxplot allows us to see the amount variability in the search score (Google's own metric) in health searches from 2004 to 2016. It is interesting there a wide range for both diabetes and depression. The variability can be explained for 1 or 2 regions: 1) variability in the number of searches in each county or 2) variability in the number of searches in each year. Cancer, however, appears to have the least amount of variability as this makes sense - it's been a consistent topic of discussion whereas diabetes and depression/mental health have been more of a discussion in recent years.

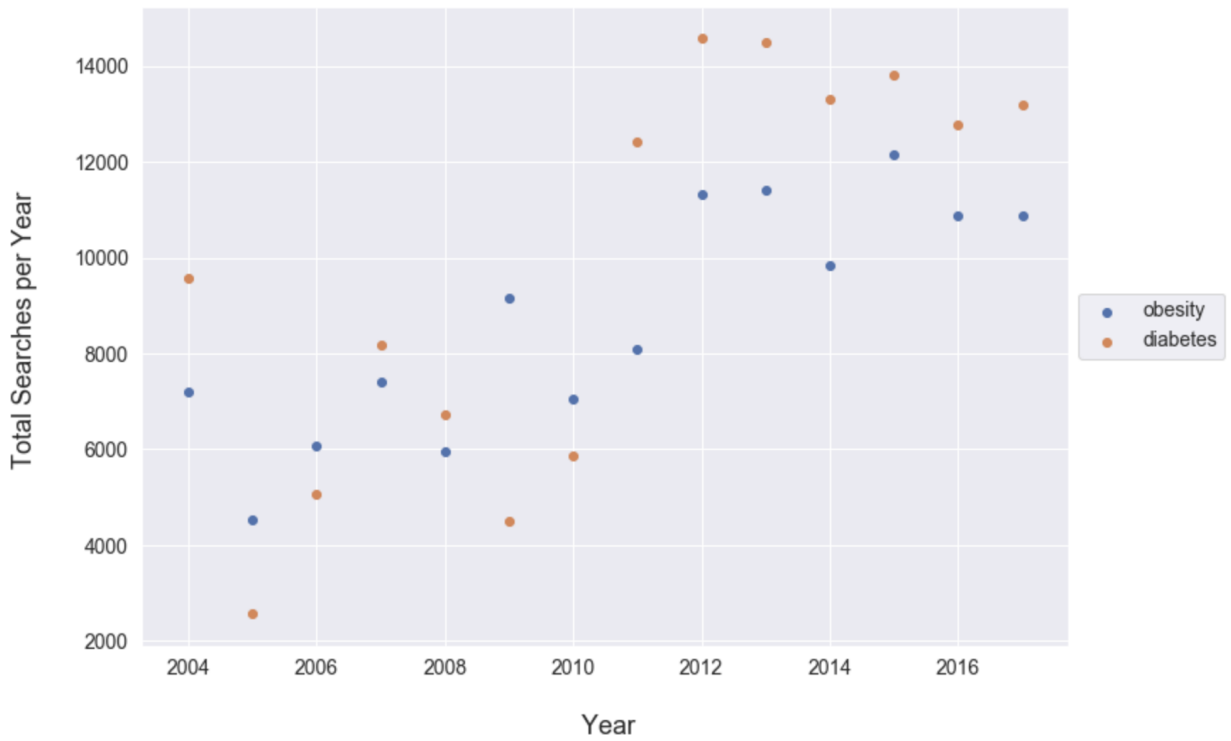
### **Barplot**



As evidenced, health searches have been growing every year. Depression has inconsistently appeared as the most searched health term 3 times, whereas cancer has received the largest searching score the most amount of times. What is the most interesting, is that diabetes has been the most searched term for three years in a row.

## Scatterplots

Health Google Searches 2004-2016

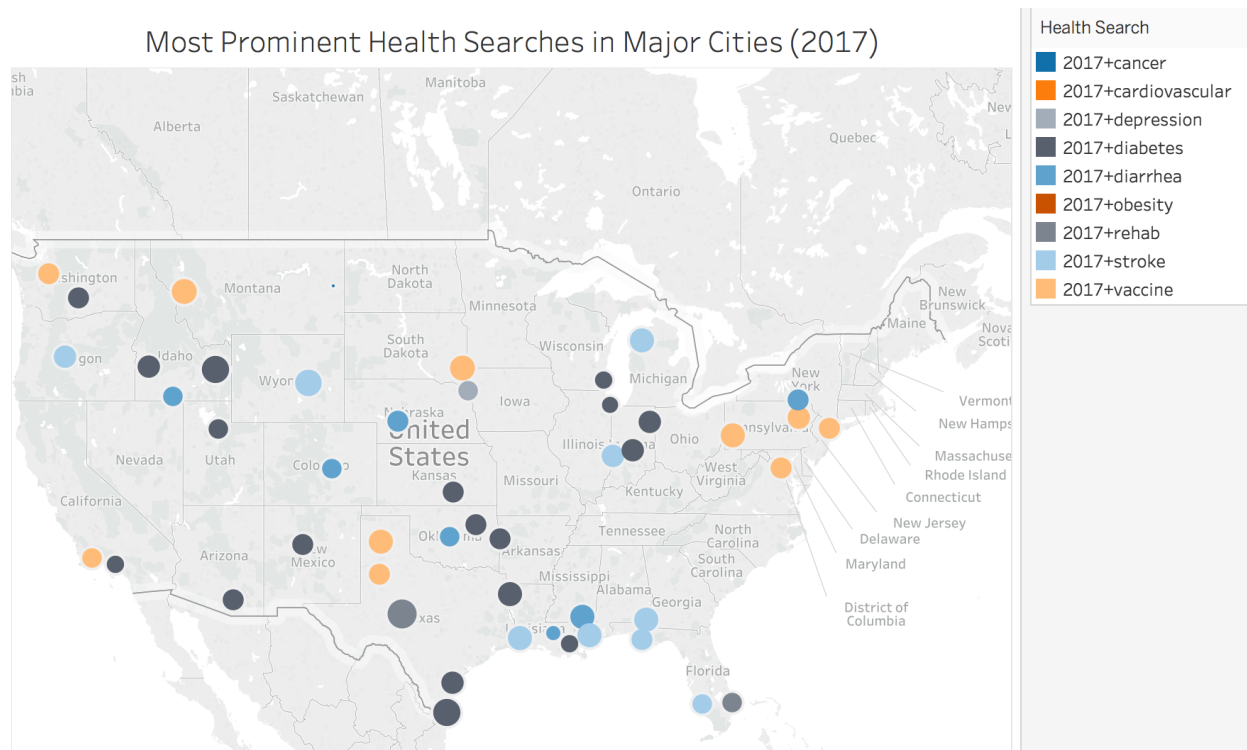


Health Google Searches 2004-2016



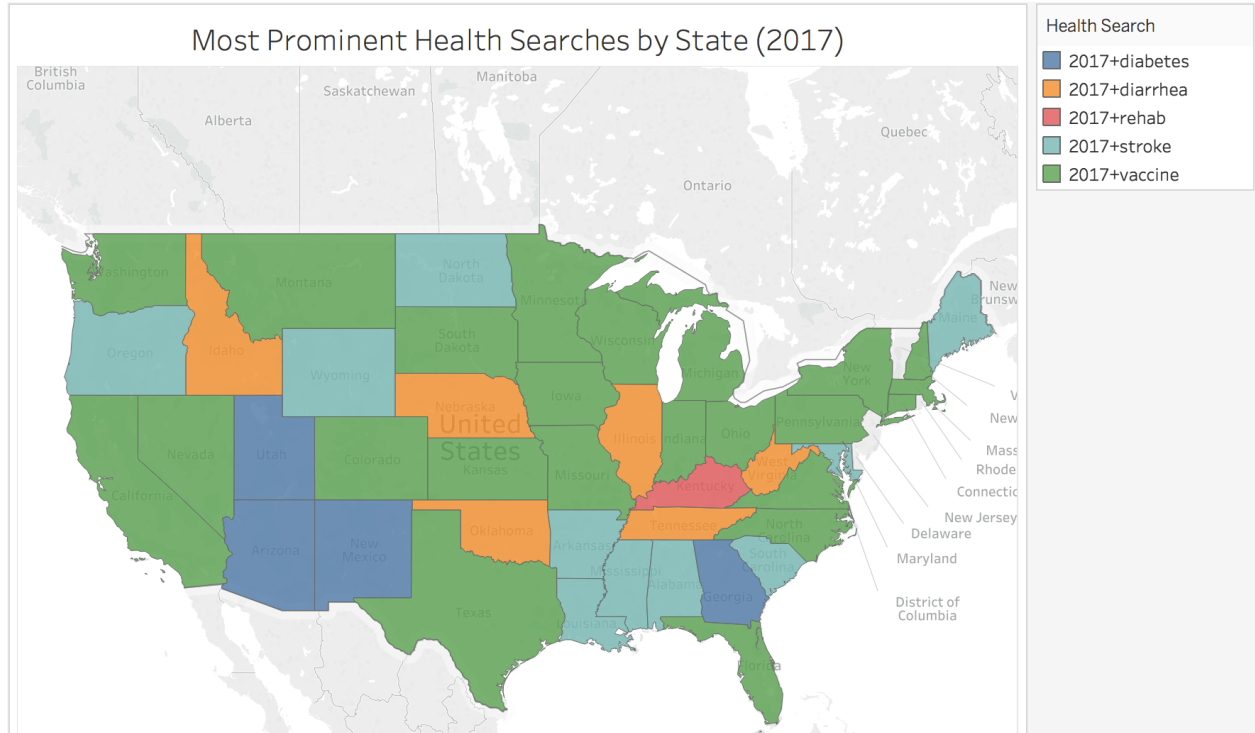
It is well-known that obesity can lead to Type 2 diabetes, and it is interesting that the search results between diabetes and obesity is extremely correlated. To show that this isn't just a result of all searches I displayed cardiovascular diseases to depression, which may be linked in one form or another, however, the search results are not as closely correlated as diabetes and obesity.

### **Bubble Map**



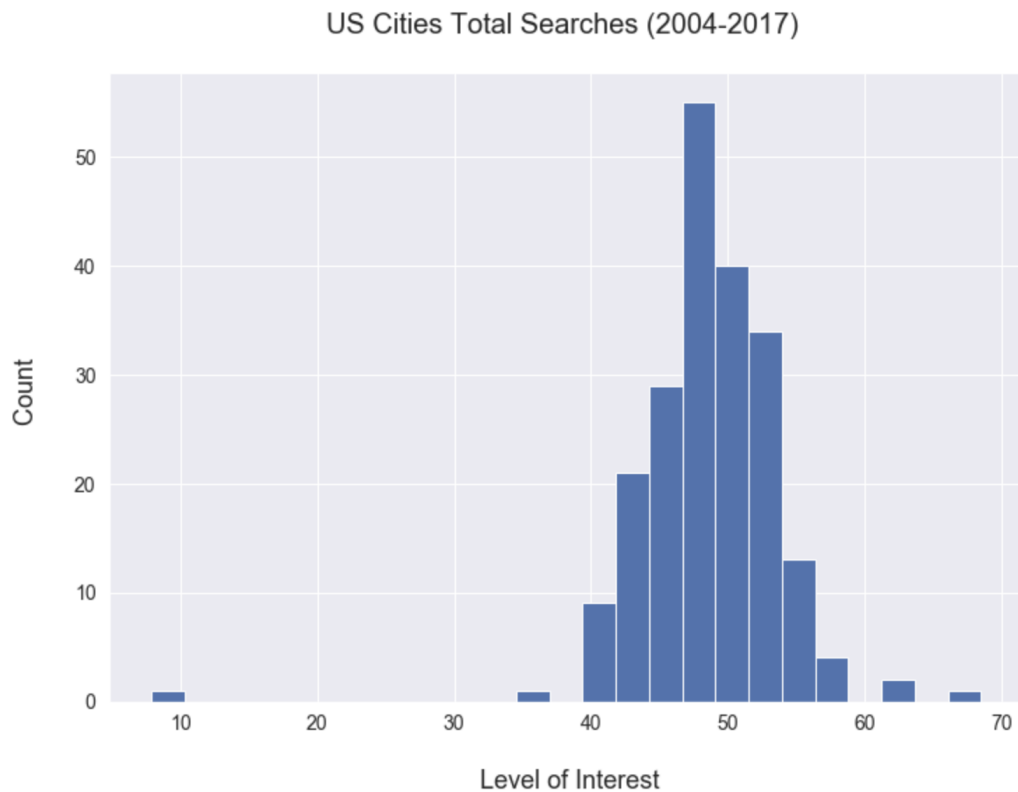
This graph shows the most prominent search in some of the major cities in the United States. The size of the bubble shows the relative amount of searches compared to other cities.

## Choropleth Map



Vaccine has been a leader of statewide searches leading to a hypothesis that smaller counties tend to search for vaccines as their main source of searching.

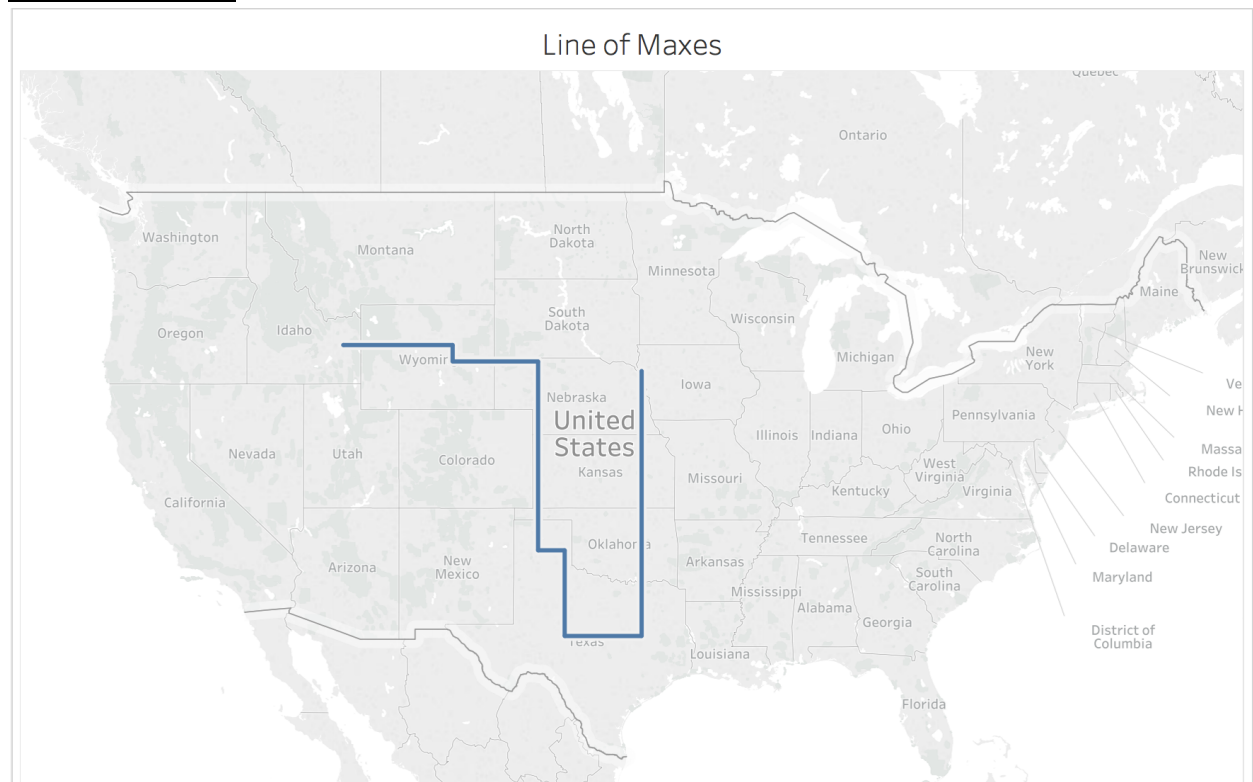
## Histogram



---

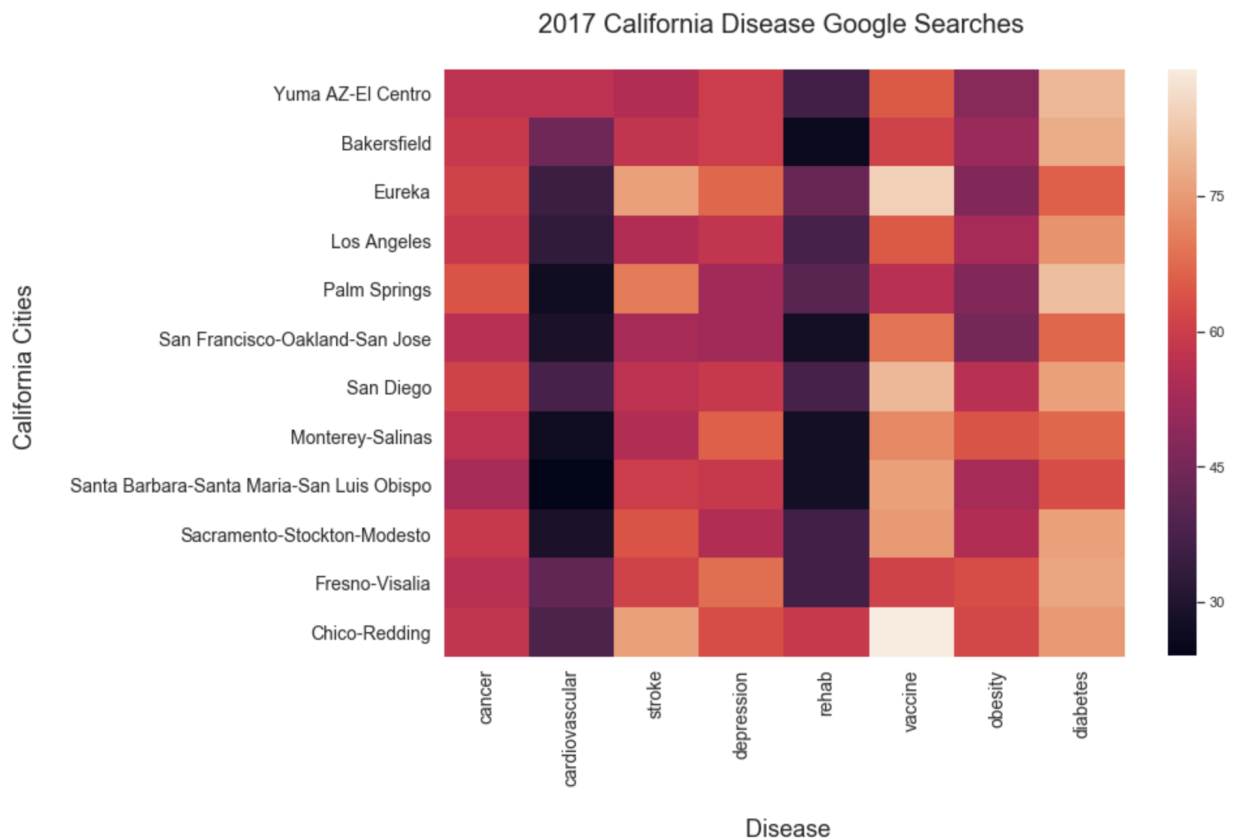
This histogram demonstrates that the city scores for each health term are close to a normal distribution. The search results provide a score rather than cumulative number of searches so smaller counties can still be represented on a scale comparable to larger counties.

## **Connection Map**



This line follows the max of each health search on Google. It is interesting the proximity of search specialties are close together in the midwest region.

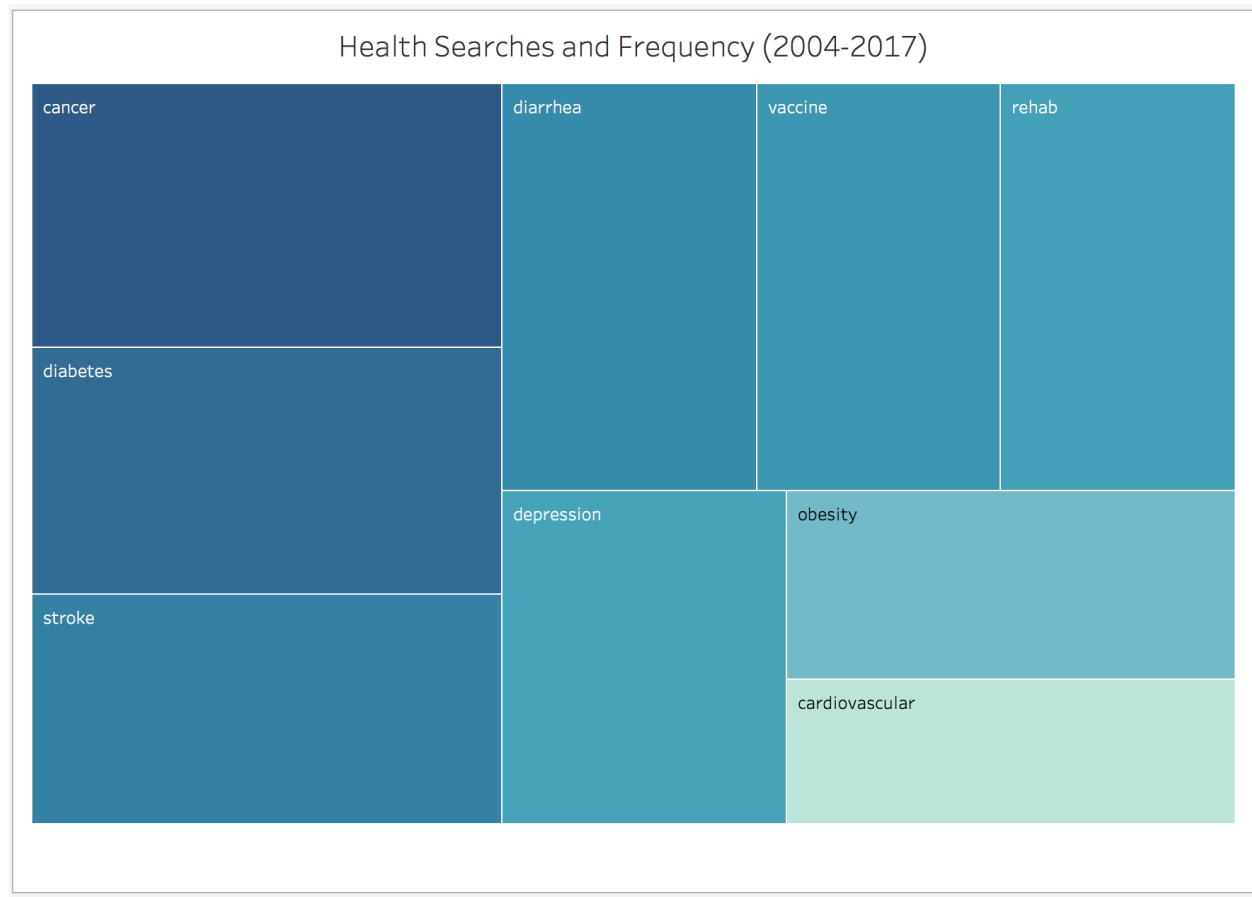
## Heatmap



As I'm from California, I wanted to take a look at the listed cities and see what search terms are most popular in 2017. Diabetes seems to be a frequent search whereas rehab and cardiovascular are less so. For the most part, cities have similar tendencies, however, there are indeed outliers (cardiovascular searches in El Centro)

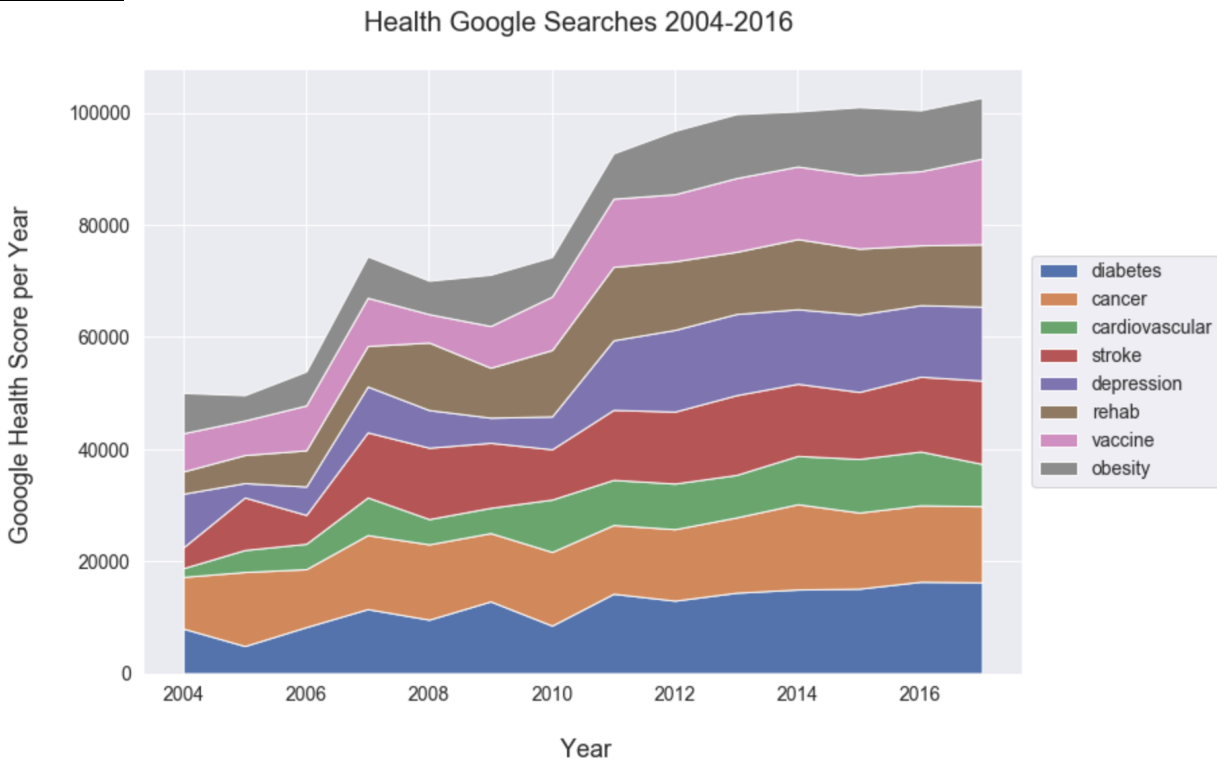


## Treemap



The treemap allows for a big picture of search frequencies for each disease. Cancer searches have lead cumulative Google searches from 2004-2017 followed by diabetes and stroke.

## Storyline:



This stacked area graph presents a dramatic illustration of how the public is increasingly obtaining health care information by searching online. Looking at the graph as a whole, searches since 2004 for the composite group of health conditions has doubled, and interest in all eight individual conditions have also expanded over this time interval. Interestingly, the frequency of searches does not necessarily correlate with the frequency of the condition in the general population, as searches for stroke information are more common than cardiovascular searches even though strokes are more infrequent than cardiovascular disease. Like all data, the trends illustrated in this graph create as many questions as answers. The goal in acquiring more knowledge is to achieve a better outcome, but in medicine, sometimes we can learn just enough to be a danger to ourselves. Even a doctor who treats himself has a fool for a patient.

Source code: [https://github.com/dthomas21/Google\\_Health\\_Searches\\_Visualizations](https://github.com/dthomas21/Google_Health_Searches_Visualizations)