Visualizing heart disease Mortality in California

External Documentation – User Guide to Analytical Tool

Created by Jarod Morris

Landing Page:

You can begin your journey by navigating to the website using the following link:

https://www.heartdiseaseviz.com

You may need to zoom out using Ctrl & Minus.

From the homepage we can see the motivation for creating the website along with a simple visual of a heart. Next you can click on the link to view the data sources. You can also navigate to this page at anytime using the navigation menu at the top.

Data Sources:

From this page you can view the sources for the data visualized in the later pages.

There is an embedded table which details the heart disease related terms that are used in the queries from these relevant data sources.

Next you can navigate to the first visualization for heart disease mortality by demographic using the navigation menu at the top.

Heart Disease Mortality by Demographic:

You are presented with a tableau dashboard with multiple connected graphs. You can use the global filters on the right-hand side of the dashboard.

You can filter by Year, Heart Disease Group, Race or Gender

Mortality by County:

This dashboard contains global filters for Race and Year. Based on these selected filters there 2 visualizations. The first is the Top 10 Counties and the percentage of deaths in the county that are contributed to heart disease out of the total death count. On the right of this table is a heat map of California. Each county is color coded according to a scale of intensity for the percentage of Deaths.

You can easily distinguish which areas of California have the higher or lower percentage based on this color gradient.

Research Output:

In the next dashboard we are greeted with 2 visualizations. On the top is the top 10 teaching hospitals by publication count. This is the affiliation that is associated with the first author on the publication from PubMed. On the bottom is the count of public vs private publications across the years 2015-2019. There is a global filter on public vs private publications.

Impact Factor:

In this dashboard we are visualizing the impact these publications have. In the first of 3 visuals we show the top 10 journals by count of publications, you can see the count of publication along with a horizontal bar graph with the average impact factor of these publication per journal.

Similarly in the next visual we can see the count of publication but for the top 10 affiliations, to the side of each affiliation if a bar chart for the average impact factor of these publications.

On the bottom of the dashboard, we can see inter publication count for all the teaching hospitals in California. This is the number of publications for which the affiliation is working with another on the same publication.

There is a global filter for the publication year and the impact factor at the top.

Research Funding:

In this dashboard we can filter globally by Fiscal Year.

At the top there is a horizontal bar chart visualizing the top 10 federally funded institutions across California.

If you hover over the bar for a specific institution, you can view the average project funding number, total project count and the max project funding amount.

In the lower half of the dashboard, we compare the trendline for total amount of research funding vs publication count vs mortality rate. The mortality trendline contains data from 2015-2019 while the Publication and Research data also includes 2020.

County Maps:

In this dashboard there is 3 maps side by side. The goal of this dashboard is to provide a by county view of the data for research effort, research output and mortality.

On the map you can select specific counties in order to filter out unwanted areas.

There is a filter for the year for each map, you can select between 2015-2019 since this is where there is an overlap in all three of the data sources.

There is a legend for the color below each map which allows us to see the intensity in certain areas of California based on the map.

Prediction Map:

In this dashboard there is a map paired with the up to 3 County prediction model graphs. You can use the selector tool to select up to 3 counties at a time. For each county there is a graph showing the trendline for heart disease related mortality rates in the counties selected. The trendlines visualize the predictions for the Linear Regression, Random Forests, XGBoost and Explainable Boosting Machine models.

You can also hover over each county and view the predication model graphs in the tooltip.

Feature Importance:

In this dashboard there is sheet for each of the prediction models. For each model we have a bar graph showing the weights for each of the 8 features. Each of the graphs are sorted in descending order based on the weight of the feature.