

OLUWASEYITAN AWOJOBI

11087659

EDA Assignment

INTRODUCTION:

In performing an EDA, I used a dataset found on Kaggle through google datasets. This can be accessed directly: <https://1drv.ms/u/s!Auv97m4w1mO9hMNzTnUSb9ovmCCyHw>

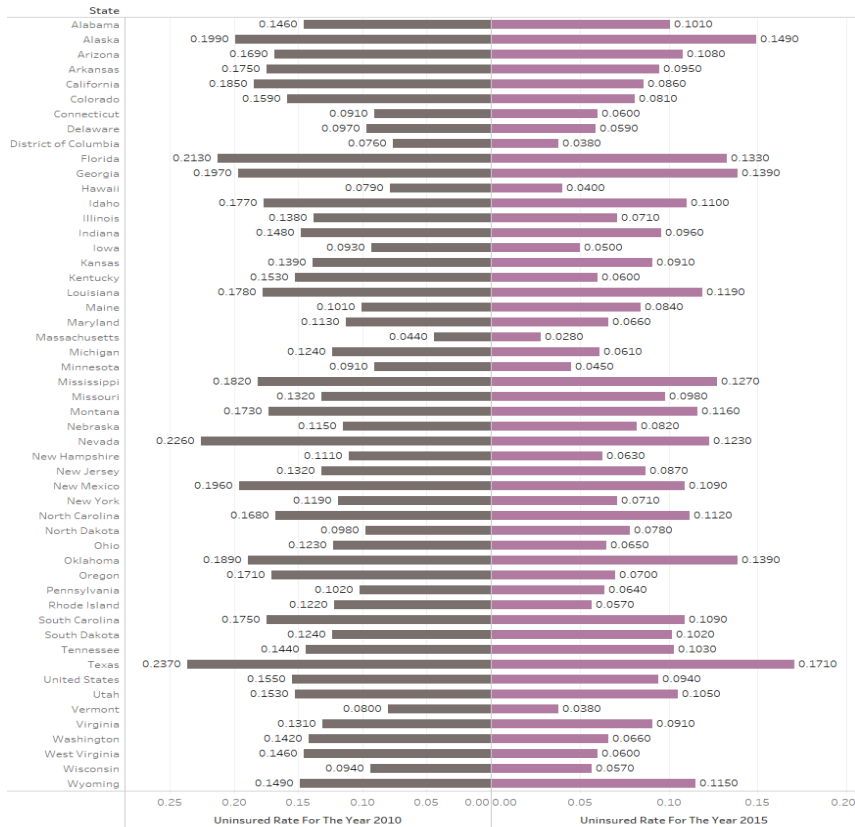
With Kaggle link being: <https://www.kaggle.com/hhs/health-insurance>

This dataset contains information about the various states in America and the health insurance coverage, uninsured rates before and after the Obamacare amongst other valuable information. This dataset was cleaned up hence did not require a lot of data cleansing before using for visualization. For my research questions, the dataset can answer them effectively.

Question 1

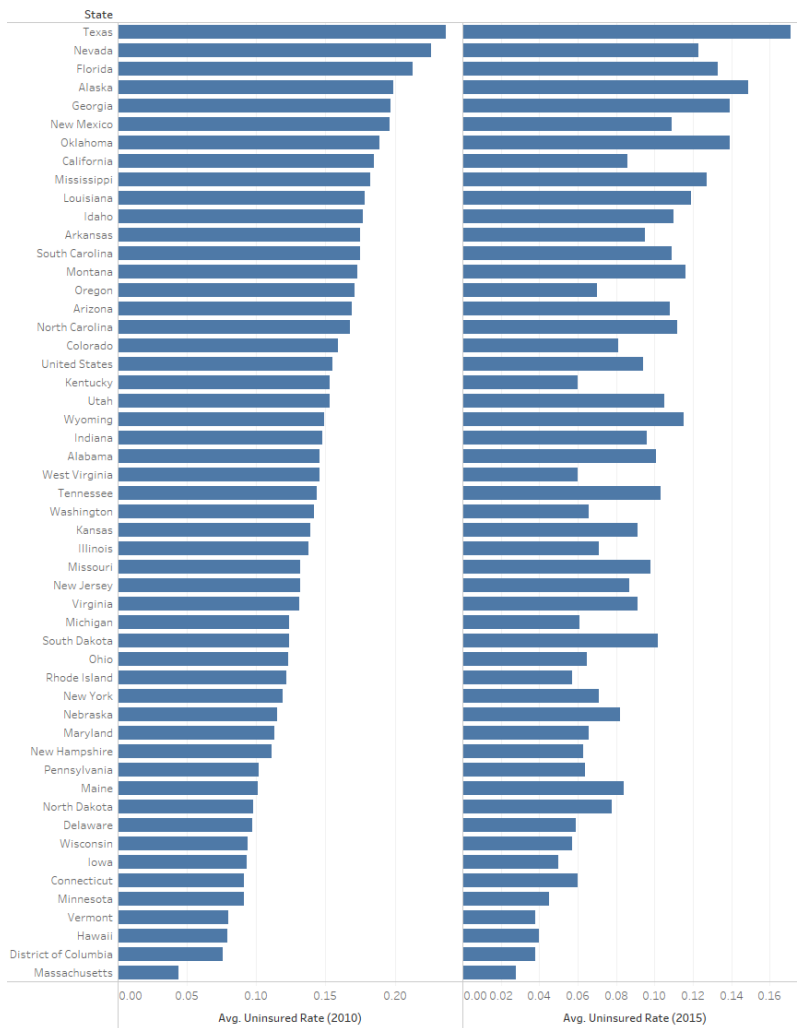
For this dataset, I intend to understand the column Uninsured Rate Change (2010-2015). This shows the number of uninsured individuals we have in the state. In doing this, I am using the bar chart spatial layout and colors grey to depict the year 2010 and purple to depict the uninsured rate for 2015. This is an interactive chart which shows as it is hovered on details of each of the lines.

Health Insurance Visualization



Sum of Uninsured Rate (2010) and sum of Uninsured Rate (2015) for each State.

Uninsured rate decline



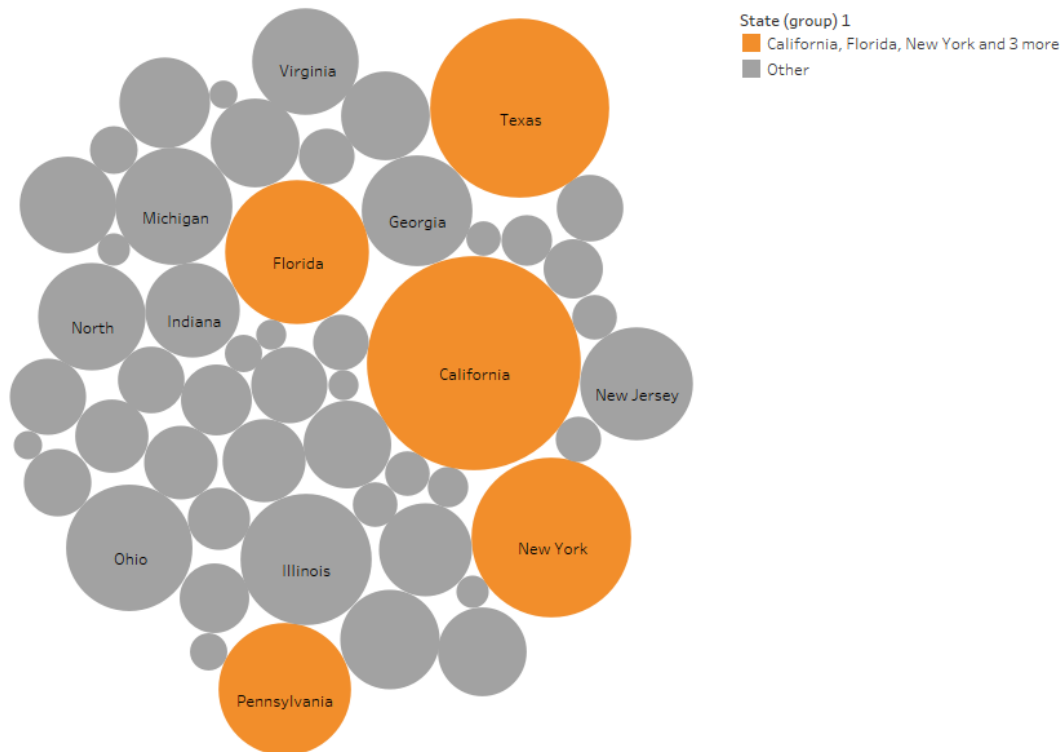
Average of Uninsured Rate (2010) and average of Uninsured Rate (2015) for each State.

From the above visualization, we can see the various rate of Uninsured people in the various states in America and the differences in both years. We see that there are significant reductions in the number of uninsured persons from 2010 to 2015. This could indicate that the Obamacare made a significant difference in the health insurance sector.

Question 2:

Next, I intend to find out the states in America which have the highest employer state insurance. For this question, I use the visualization below to answer it.

Employer Health Insurance coverage and State relationshipSheet 3



State. Color shows details about State (group) 1. Size shows sum of Employer Health Insurance Coverage (2015). The marks are labeled by State. The data is filtered on State (group), which keeps no members. The view is filtered on Exclusions (State,State (group) 1), which keeps 51 members.

From the above, we can see the top 5 states which have the highest Employer health insurance coverage. To answer our research question therefore, the states with the highest employer health insurance are; **Pennsylvania, Florida, California, New York and Texas.**

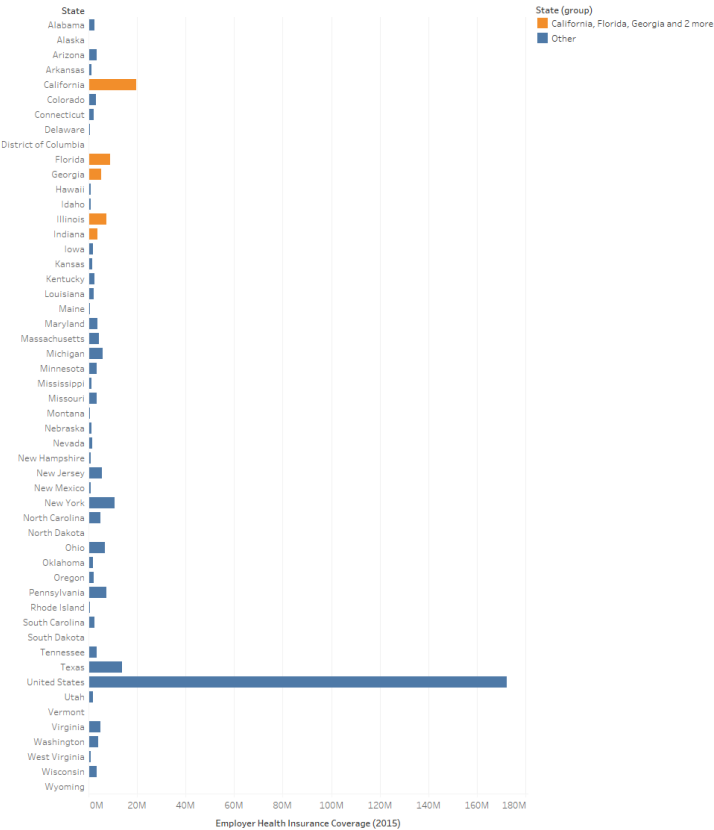
These are highlighted above in orange and have the largest sizes. From this visualization, we could draw assumptions based off the population, and the high cost of living in these cities.

Before Final Result:

Before deciding on this dataset, I had used an insurance dataset with specific questions in mind. The dataset was however very sparse, hence did not have enough information to make give the answers I wanted.

Before arriving on the second visualization result, below is what was first done. This did not accurately depict the information sought hence the one above

Relationship between employer health insurance coverage and the state



Sum of Employer Health Insurance Coverage (2015) for each State. Color shows details about State (group). The view is filtered on State (group), which keeps no members.