

500 Cities Health Data

TEAM COMBAT
SRIRAM
YANHE WU
SRAVANI
NARAHARI

AGENDA

- ▶ Introduction of Target Audience
- ▶ Dataset Introduction
- ▶ Research Questions
- ▶ Data Cleaning Process
- ▶ Data Visualization (using R Plots)
- ▶ Interpretation from Data
- ▶ Recommendations
- ▶ Challenges Faced
- ▶ Outcome & Conclusion

Introduction of Target Audience

- ▶ **Jacob Williams** - Chief Medical Officer, CHI Hospital, Omaha, NE & Director of Division of Public Health
- ▶ Designs medical policies and assists in maintenance of health standards
- ▶ **Interests** - Research on methods for prevention of diseases; Ensure effective medical service by providing training; Promote public health seminars
- ▶ **Concerns** - Current Health awareness programs and campaigns
- ▶ **Challenges** - Predict and prevent diseases by studying trends and bring health awareness
- ▶ Current analysis would help Mr. Jacob in developing effective measures to understand health risks and develop preventive measures

Introduction to Dataset

- ▶ 500 cities health data outlining health outcomes and preventive measure categories
- ▶ Year of Data – 2013
- ▶ Includes population estimates for people affected by various health behaviors and people who have received preventive measures for the health behaviors
- ▶ Data Source - 500 Cities: Local Data for Better Health from <https://catalog.data.gov/dataset/500-cities-local-data-for-better-health-b32fd>

Limitations of Dataset

- ▶ Only one year of data available
- ▶ Lots of redundant and repetitive data
- ▶ Only one column containing a value of measure or significance
- ▶ Minimal information to interpret
- ▶ Highly granular and distributed data

Research Questions

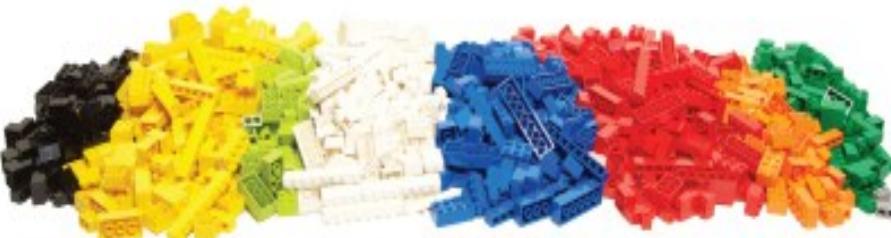
- ▶ What is the estimate of existence of *Health outcomes and Preventive services*, based on different measures in each of the regions in the US ?
- ▶ How does one region compare to the other in terms of population density for Health Outcomes
- ▶ How does one region compare to the other in terms of population density for Preventive Measures provided
- ▶ Which state/region has an imbalance in terms of Health Outcomes and Preventive Measures

Data Cleaning Process

DATA



SORTED

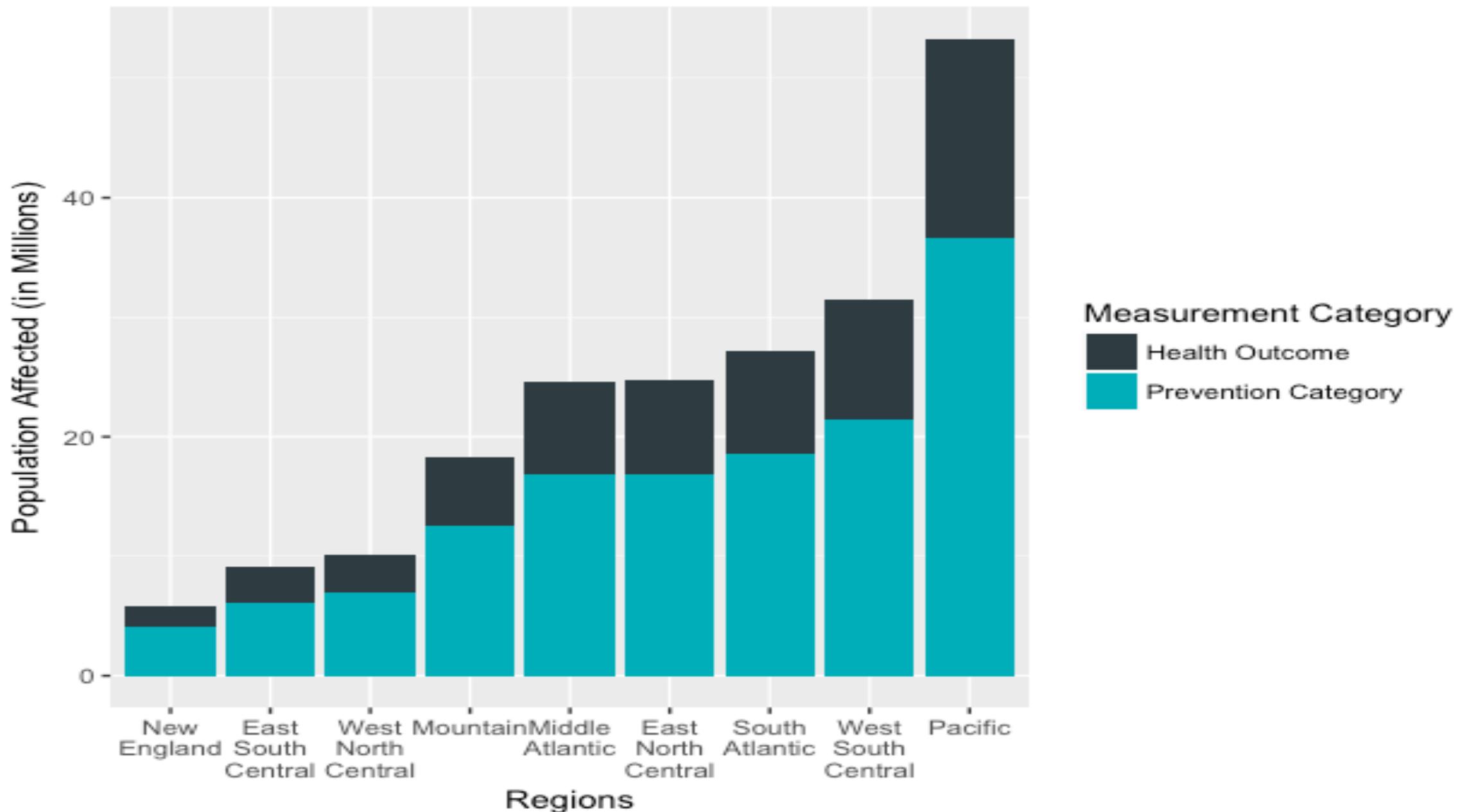


- ▶ Data Validity & relativity
- ▶ Missing values
- ▶ Unstandardized data
- ▶ Irrelevant data
- ▶ Division of data into a scalable set
(Grouping States to Regions)
- ▶ Calculate new columns required for interpretation of data

Data Visualization

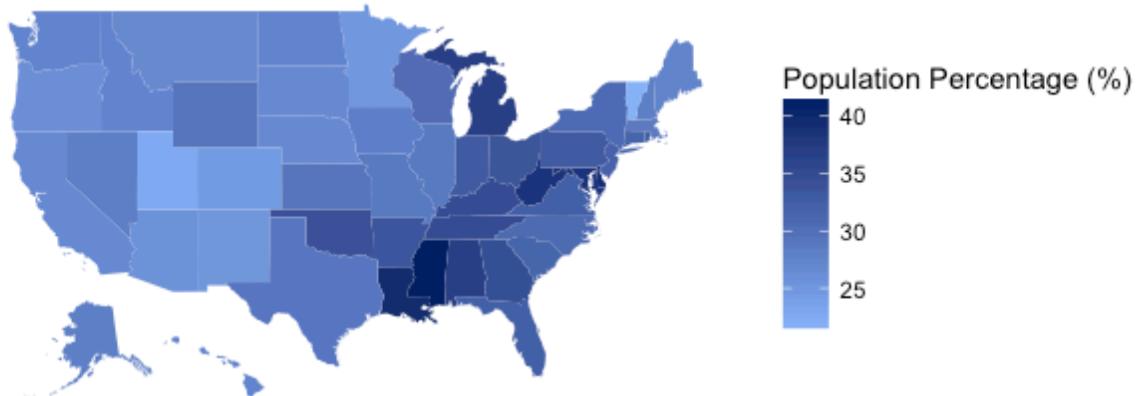
- ▶ To extract information from data
- ▶ Easily communicate on a smaller scale
- ▶ Quick readability
- ▶ Focus on what is required

Box Plot of Population Affected and Regions in the USA

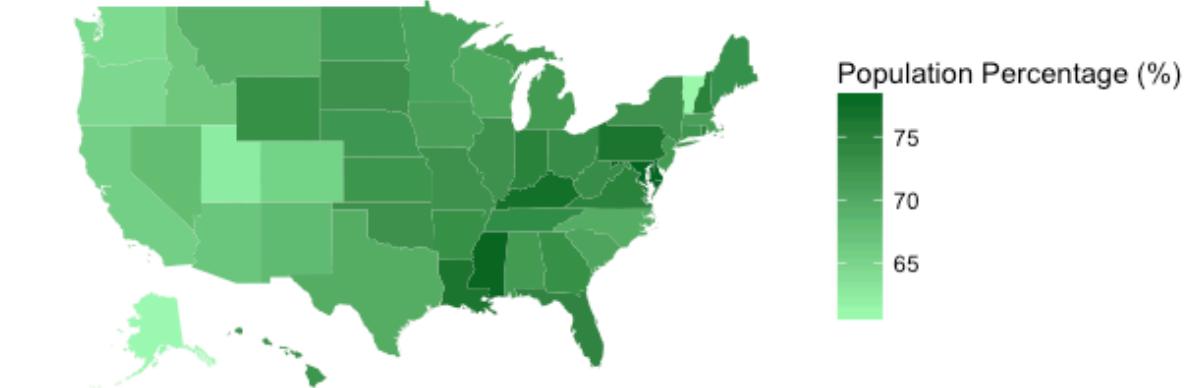


Comparison for High BP

State wise Distribution - Affected Population by High BP (in Percentage)



State wise Distribution - Prevention for High BP (in Percentage)

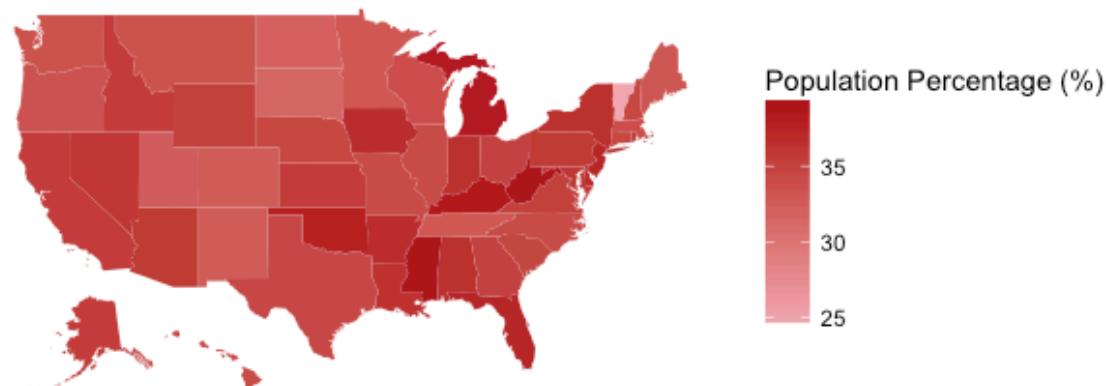


Interpretations from Data

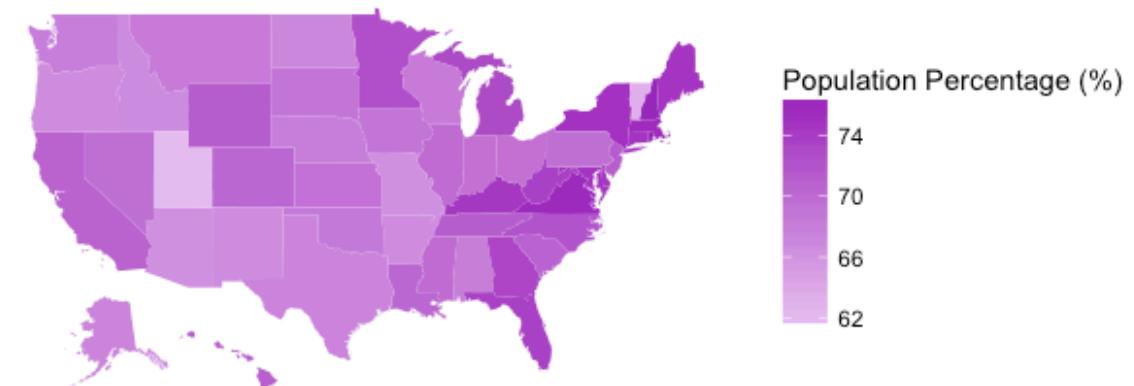
- ▶ Linear relation between health outcomes and preventive measures for all regions in the USA
- ▶ 75% of the Population affected with High BP are provided with some kind of preventive measure in the Southern Regions
- ▶ States with less 50% of population affected by High BP, have about 25% of the population covered by preventive measures (Pacific Region)
- ▶ Linear Relation in terms of High BP health outcomes and preventive measures
- ▶ High Cholesterol is less prevalent than High BP in the USA

Comparison for High Cholesterol

State wise Distribution - High Cholesterol (in Percentage)



State wise Distribution - Prevention for Cholesterol (in Percentage)



Interpretations from Data

- ▶ Southern Region states have more than 35% population affected by high cholesterol
- ▶ However, less than 60% of the affected population has received preventive measures for high cholesterol in these states under the southern region
- ▶ States with 25 – 30% of of population affected by high cholesterol has more than 70% of the population covered by preventive measures
- ▶ For example, the state of Iowa in which about 35% of the population is affected by high cholesterol, around 65% of the affected population have received preventive measure
- ▶ But the state of Maine which less than 30% of the affected population, has more than 70% of the population covered

Recommendations

- ▶ Based on our analysis, we believe that there are still many states which are not completely covered with preventive measures for the health risks faced by them.
- ▶ The population affected by high blood pressure is better covered with preventive measures, than the states facing high cholesterol.
- ▶ For population affected by high cholesterol, more preventive measures must be taken, covering more states.
- ▶ We recommend that CMO should create strategy plan to solve the problem of High Cholesterol in various states.
- ▶ And CMO should also lead their team to investigate why some states with lower levels of high cholesterol have better coverage of preventive measures, than other states with higher level of high cholesterol affected population

Challenges Faced

- ▶ Understanding the dataset – changed at every stage of the project (fortunately for the better)
- ▶ Data cleaning was a continuous process and required updates at every stage – Convert cities to states to regions
- ▶ Only a single data value to work with and create visualizations for interpretations – Required different perspectives
- ▶ Align research questions with data visualization – dataset was not always supportive
- ▶ Limited knowledge of R programming

Outcomes

- ▶ Better understanding of health data and health status of various states of USA
- ▶ Learned R programming and worked well towards creating plots and visualizations
- ▶ Made the most of the dataset in hand and developed useful interpretations
- ▶ Analysis could be extended by having multiple years of data
- ▶ Analysis could be combined with other dataset providing different dimensions
- ▶ Analysis could be improved by better description and understanding of dataset at initial stages of the project



Questions ?

References

- ▶ Dataset source - <https://www.cdc.gov/500cities/>
- ▶ Breakdown by Region done based on -
https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf



Thank You