SUMMARY

Audience Analysis

Our project is written for serval different audiences at the same time.

Primary audience: Chief Medical Officer(CMO).
This project is primarily intended for individual who is senior government official designated head of medical services. The CMO will serve to advise and lead a team of medical experts on matters of public health importance. For the primary audience, they may be male or female, with Doctor/Ph.D. education level on public healthcare and strong professional background on managing cities diseases control project. They also have very excellent executive leadership skill, communication skill and problems solving skill. The primary audience will understand how to read data analysis report and create strategic plan base on the reported outcome. They will be more interested in public healthcare problem than other people. As the requirement needs from their job, they will need our project outcome for their decision making in the public healthcare fields. And the primary audiences can't be the CMO of the country other than the USA. Because the data source of our project is from 500 cities of United State and the project objective is to provide professional analysis to assist cities in planning public health intervention.

Secondary audience: Experts from Center For Disease Control And Prevention or Physician.
With high education level on disease control and prevention, the secondary will have encyclopedic medical knowledge that can be recalled at a moment's notice. Usually, they work as a team member in a team of medical experts and provide a specific solution for disease control and prevention. They are high intelligence, inquisitiveness and have very good communication skill as well as medical problems solving skill. As the requirement of their job, they will need our project outcome for deeper investigation on the disease itself, such as why this type of disease will have higher rate in some area, what solution can be provided to keep people in safe.

Brief description of the source data

It's the complete dataset for the 500 Cities project, available from data.gov. With 21 variables and 810103 observations, it includes 2013, 2014 model-based small area estimates for 27 measures of chronic disease related to unhealthy behaviors (5), health outcomes (13), and use of preventive services (9). It also includes estimates for approximately 28,000 census tracts within 500 largest US cities. After our data cleaning process, we are using the dataset concerning the year 2013 for data analysis. The dataset includes 2013 model-based small area estimates for four measures of chronic disease related to health outcome and prevention with 12 variables and 116024 observations. This dataset is significant to identify emerging health problems and provide information for disease prevention activities.

Variables include: Year, StateDesc, CityName, GeographicLevel, Category, Measure, Data value Type, Data Value, Data_Value(in%), Data Value Footnote, Population Count, CategoryID, MeasureID

Categories: Health outcome, prevention.

Measures of each Category:

A. Health outcome

a. High blood pressure among adults aged equal and larger than 18 Years

b. High cholesterol among adults aged equal and larger than 18 Years who have been screened in the past 5 Years

B. Prevention

a. Cholesterol screening among adults aged equal and larger than 18 Years

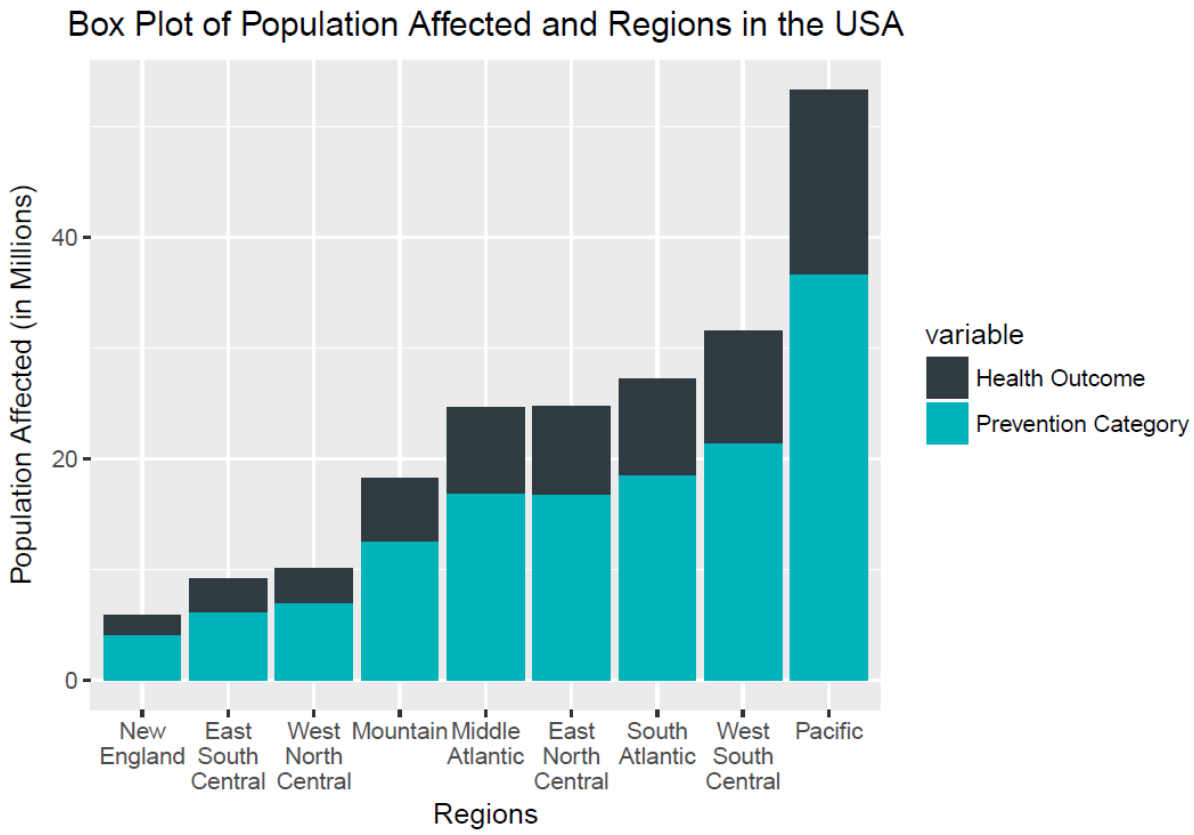
b. Taking medicine for high blood pressure control among adults aged equal and larger than 18 Years with high blood pressure

Data analysis processing

We would like to use four words to summarize the processing of data analysis: inspecting, cleansing, analyzing and visualization. First, inspect data. In this process, our main goal was to the understood dataset and questioned the data to meet end-user data requirement. Then we defined the research questions that the dataset may be able to answer through brainstorm. Second, clean dataset. Maintaining excellent quality data is essential to ensure the data reliability and deliver accuracy in the further data analysis process. During this process, four issues encountered with the data; data validity and relativeness, missing values, unstandardized data, irrelevant data concerning data cleaning and visualization goals. Then Excel and R were used as a tool to solve these four issues. Third, data analyzing. This process is to discover the useful information in the dataset. To answer the research questions, an R script was developed by applying descriptive analysis method. Last but not the least, data visualization. It's a process of presenting analyzed data in a pictorial and graphical format. Server R plots were developed based on the outcome of R Scripts. In the following paragraph, you will see the outcome of the data analysis processing.

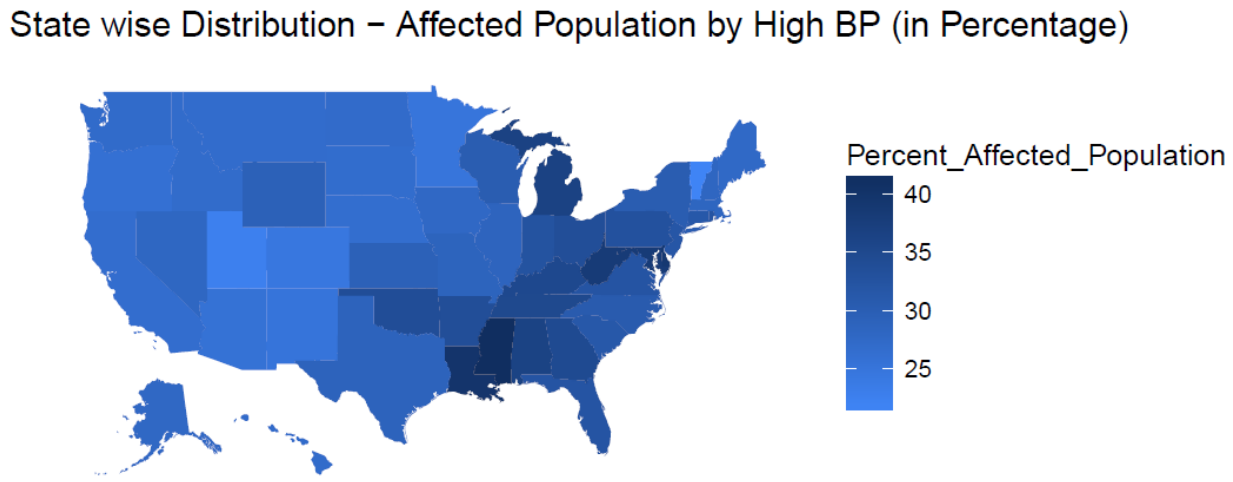
Plots

Plot 1



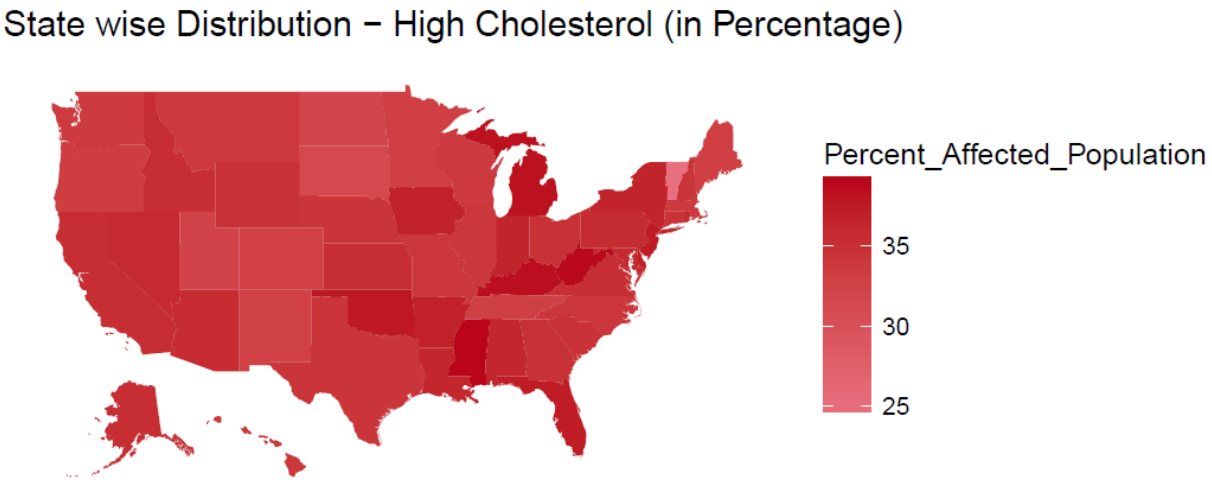
The above plot intends to provide a visualization on the different measure of Categories by which the 500 Cities Local Health Dataset is based upon. The categories, Health Outcomes, and Preventive Measures are combined to be depicted on a single bar graph divided based on Regions in the USA the visualization depicts that there has been a linear relation with the number of Preventive Measures concerning Health outcomes in the 5 Regions in the USA. From the above plot, we could know that Pacific region, which includes Alaska, California, Hawaii, Oregon, and Washington, has more population which is affected by the chronic disease of the categories of health outcome and prevention category than other regions.

Plots 2



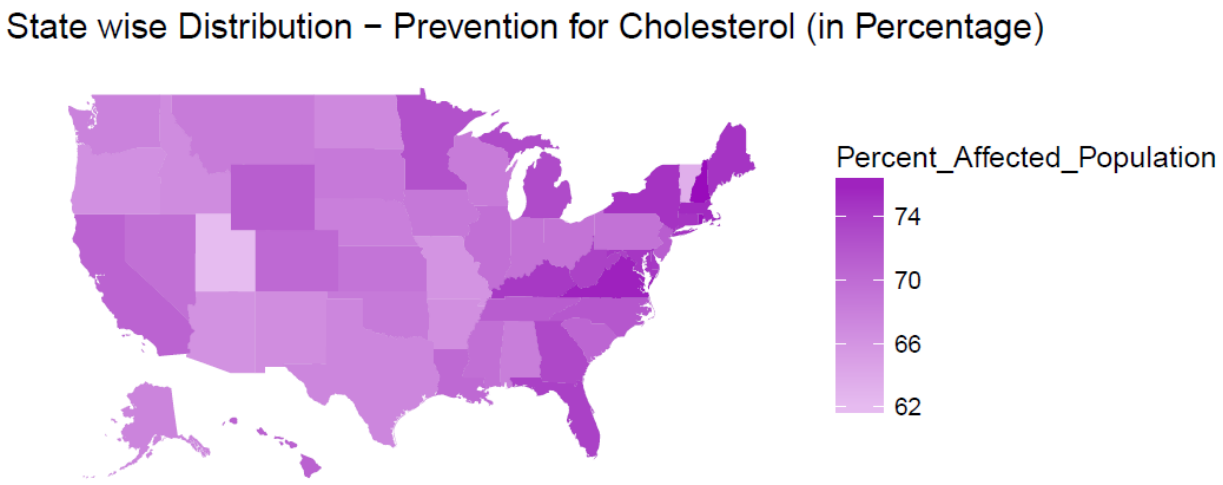
The above plot shows a visualization of the High blood pressure among adults aged equal and larger than 18 Years (High BP) measure (Health Outcome) in all the states in the USA. It can be seen that a state like Mississippi and Louisiana has high population affected by High BP. And the state in South and Northeast region has high population affect by High BP.

Plot 3



The above plot shows a visualization of the *High cholesterol among adults aged equal and larger than 18 Years who have been screened in the past 5 Years* (High Cholesterol) measure (Health Outcome) in all the states in the USA. It can be seen that a state like California has population affected by High Cholesterol.

Plot 4



The above plot shows a visualization of the *Cholesterol screening among adults aged equal and larger than 18 Years* (Preventive Measure) in all the states in the USA. Both state New Hampshire, Virginia have the deep purple color which means they have the highest percentage of the population who is Cholesterol screening among adults aged equal and larger than 18 Years.

Persuasive Argument

Base on the result we had, we recommend that CMO should create strategy plan to solve the problem of High blood pressure among adults aged equal and larger than 18 Years in on south and northeast region. And CMO should also lead their team to investigate why a state like California has higher percent population affected by High cholesterol and provide advice or solution to prevent high cholesterol.

Word Count: 1115

Reference:

500 Cities: Local Data for Better Health. (2016, December 07). Retrieved October 14, 2017, from https://catalog.data.gov/dataset/500-cities-local-data-for-better-health-b32fd

500 Cities: Local Data for Better Health. (2016, December 12). Retrieved December 04, 2017, from https://www.cdc.gov/500cities/definitions/prevention.htm