U.S. CHRONIC DISEASE INDICATORS (CDI) & HEART DISEASE MORTALITY DATA AMONG US ADULTS 35 BY STATE, TERRITORY AND COUNTY

CIS 5270: BUSINESS INTELLIGENCE

Spring 2018

By Tanmai Aurangabadkar

Submitted to: Dr. Shilpa Balan

A) Dataset URL

https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi-e50c9

https://catalog.data.gov/dataset/heart-disease-mortality-data-among-us-adults-35-by-state-territory-and-county

US Chronic Diseases Indicators dataset comes from CDC's Division of Population Health that provides cross-cutting set of 124 indicators that were developed by consensus and that allows states and territories and large metropolitan areas to uniformly define, collect, and report chronic disease data that are important to public health practice and available for

states, territories and large metropolitan areas. In addition to providing access to state-specific indicator data, the CDI web site serves as a gateway to additional information and data resources. The CDI website enables public health professionals and policymakers to retrieve uniformly defined state-level and selected metropolitan-level data for chronic diseases and risk factors that have a substantial impact on public health. These indicators are essential for surveillance, prioritization, and evaluation of public health interventions for chronic disease.

Heart Disease Mortality Dataset is from 2012 to 2014 with 3-year average. Rates are age-standardized and county rates are spatially smoothed. The data can be viewed by gender and race/ethnicity. Data were provided by National Vital Statistics System. Additional data, maps, and methodology can be viewed on the Interactive Atlas of Heart Disease and Stroke.

B) Data Cleaning:

1. Heading in the dataset is not written completely for abbreviation used.

Before

Year	LocationAbbr	LocationDesc
2016	US	United States
2016	AL	Alabama
2016	AK	Alaska

After

Year	LocationAbbreviation	LocationDesc
2016	US	United States
2016	AL	Alabama
2016	AK	Alaska

2. Replacing duplicate values by new values.

Before

After

DataValue	DataValueAlt	Low Confidence Limit
16.9	16.9	16
13	13	11.9
13	13	11.9
18.2	18.2	16

DataValueAlt	Low Confidence Limit	High Confidence Limit
16.9	16	18
13	11.9	14.1
18.2	16	20.6
15.6	14.3	16.9

3. Year column needs to be changed by proper date format

Before

Year	LocationAbbreviation	LocationDesc
2016	US	United States
2016	AL	Alabama
2016	AK	Alaska
2016	AZ	Arizona

After

Year	LocationAbbreviation	LocationDesc
1/1/2016	US	United States
1/1/2016	AL	Alabama
1/1/2016	AK	Alaska
1/1/2016	AZ	Arizona

4. Repetition of year i.e. start and end (twice) which needs to be converted into single column as Year.

Before

Year Start	Year End	LocationA	LocationD	Topic
2016	2016	US	United Sta	Alcohol
2016	2016	AL	Alabama	Alcohol
2016	2016	AK	Alaska	Alcohol
2016	2016	AZ	Arizona	Alcohol
2016	2016	AR	Arkansas	Alcohol

After

Year	LocationA	LocationD	Topic
2016	US	United Sta	Alcohol
2016	AL	Alabama	Alcohol
2016	AK	Alaska	Alcohol
2016	AZ	Arizona	Alcohol
2016	AR	Arkansas	Alcohol
2016	CA	California	Alcohol

5. Replace negative values with positive values.

Before

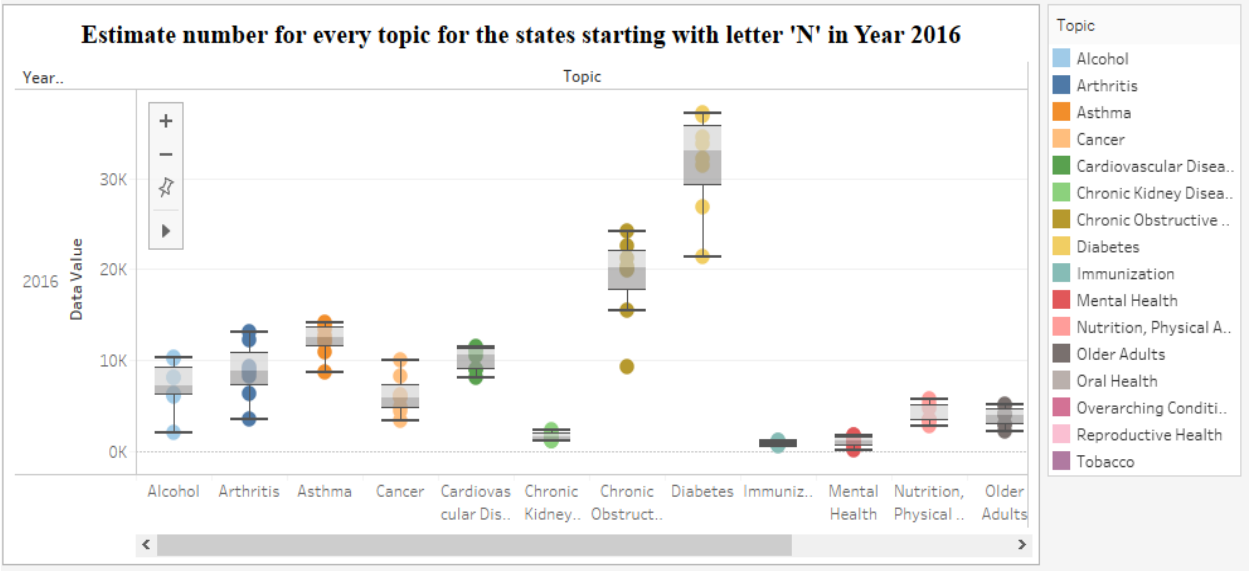
DataValue	DataValue	Low Confi
-16.9	16.9	16
13	13	11.9
18.2	18.2	16
15.6	15.6	14.3

After

DataValue	DataValue	DataValue
Crude Pre	16.9	16.9
Crude Pre	13	13
Crude Pre	18.2	18.2
Crude Pre	15.6	15.6

C) Data Visualization

Question 1: Estimate number for every topic for the states starting with letter 'N' in Year 2016

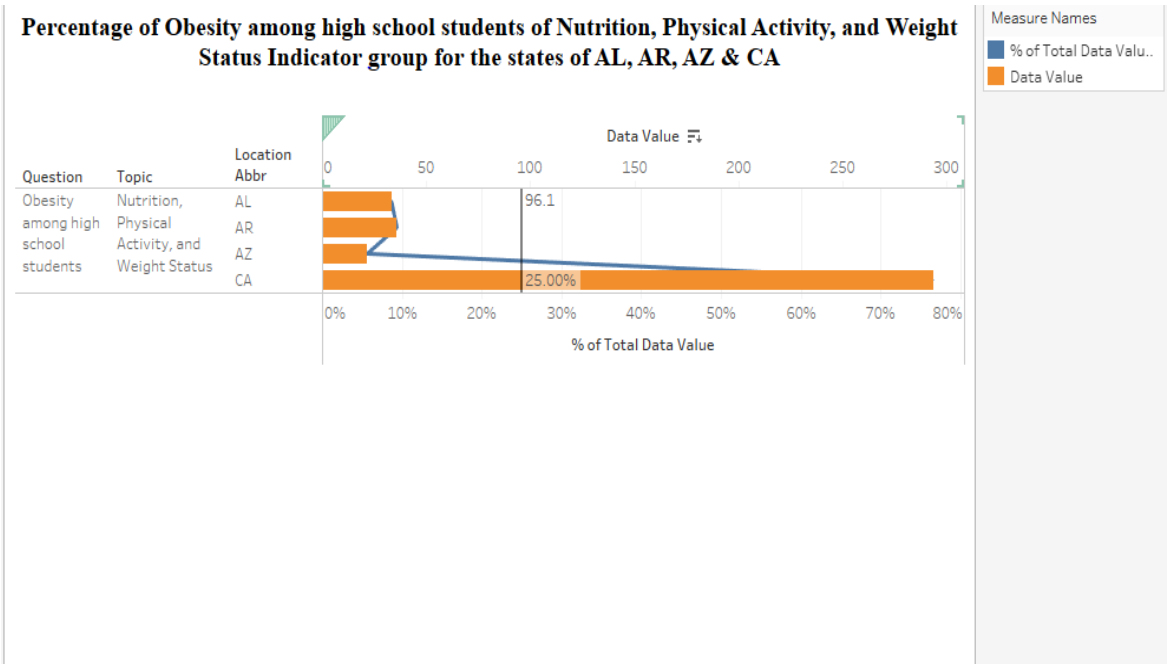


Sum of Data Value for each Topic broken down by Year Year. Color shows details about Topic. Details are shown for Location Desc and Topic. The view is filtered on Location Desc and Year Year. The Location Desc filter keeps 8 of 55 members. The Year Year filter keeps 2016.

[**Tools Used**: **Box & Whiskers Plot**]

There are 16 types of chronic diseases observed in US. This is Box and Whiskers Plot that shows sum of chronic diseases (topic) specifically for year 2016 having location description (states) starting with “N”. From above plot, we can say that total sum of chronic diseases is 37,155 that is highest for New Hampshire seen for diabetes and total sum of chronic diseases is 124 that is lowest for New York seen for Mental health problem. It is required to analyze causes and areas which are having large number of chronic diseases. Looking at the plot, we can conclude that for Cardiovascular Diseases, sum of chronic diseases is exceeding by almost 1000 values for every state starting with “N” and overall average comes to 7992.

Question 2: Obesity among high school students is what percent of Nutrition, Physical Activity, and Weight Status Indicator group for the states of AL, AR, AZ & CA.

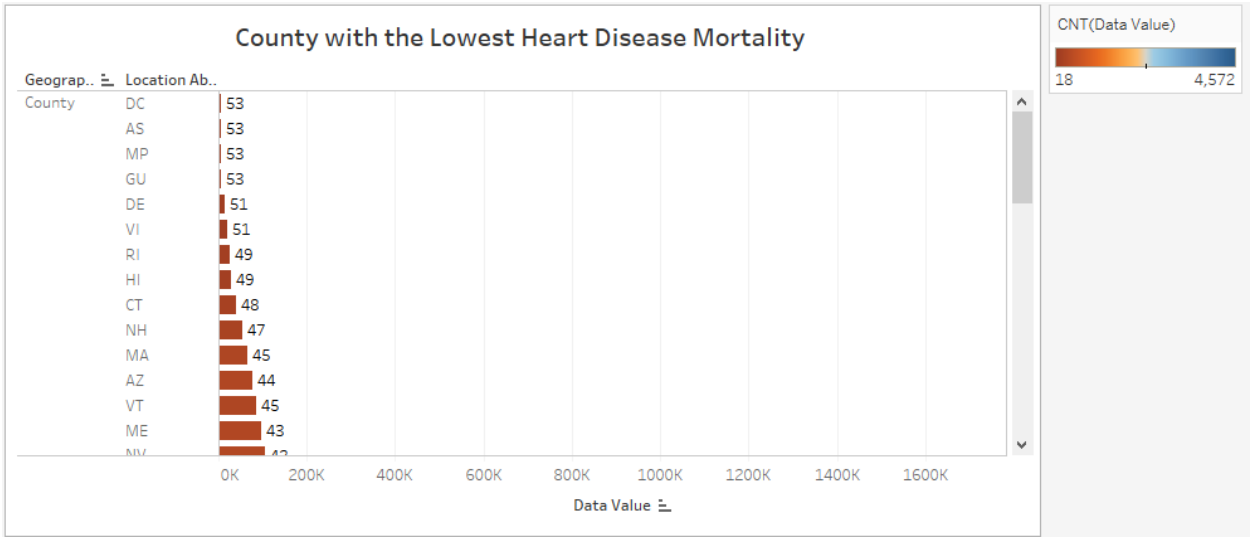


The trends of % of Total Data Value along Location Abbr, Question, Topic and Data Value for Location Abbr broken down by Question and Topic. Color shows details about % of Total Data Value along Location Abbr, Question, Topic and Data Value. The view is filtered on Question and Location Abbr. The Question filter keeps Obesity among high school students. The Location Abbr filter keeps AL, AR, AZ and CA.

[**Tools Used**: **Dual Axis Chart, Reference Line**]

Obesity has emerged as a priority in chronic disease prevention and has been linked to increased risk for heart disease, high blood pressure, arthritis-related disability, and some cancers. Above bar chart shows percentage of obesity among high school students for Nutrition, Physical Activity, and Weight Status Indicator group for the locations: Alabama, Arkansas, Arizona & California. Dual Axis shows comparison between actual data (bar graph) and percentage of data values (Line graph). California is leading the race with maximum summation of data i.e. about 76.42% in terms of obesity among high school students. Least percentage of obesity is seen in Arizona state i.e. about 5.62% followed by Arkansas and Alabama which are close to each other with 8.64 % & 9.32% respectively. Average of obesity data are denoted by Reference Line that comes around 25%.

Question 3: Which county has the lowest Heart Disease Mortality?

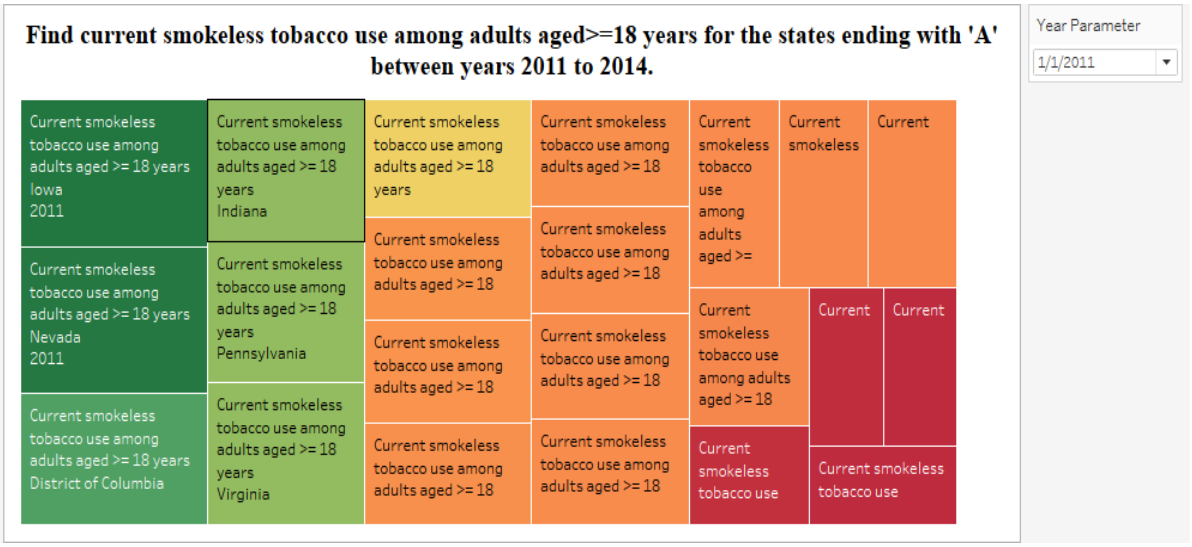


Sum of Data Value for each Location Abbr broken down by Geographic Level. Color shows count of Data Value. The marks are labeled by Rank of Count of Data Value. Details are shown for Location Abbr. The view is filtered on Geographic Level, which keeps County.

[**Tools Used**: **Ranks**]

The most common type of heart disease in the United States is coronary artery disease, which affects blood flow to the heart. Decreased blood flow can cause a heart attack. Above bar graph mentions count for heart disease mortality for county geographical level. All the counties are given ranks to predict the county with lowest heart disease mortality which is District of Columbia. Highest ranking of heart disease mortality is seen for Texas county. Tennessee and Kansas counties are very close to each other in terms of their count for heart disease mortality. Some of the risk factors associated for heart disease mortality can include: High cholesterol, Diabetes, Unhealthy diet, Physical inactivity, high consumption of tobacco and alcohol, obesity, genetics and family history.

Question 4: Find current smokeless tobacco use among adults aged>=18 years for the states ending with 'A' between years 2011 to 2014.



Question, Location Desc and Year Year. Color shows sum of Data Value. Size shows sum of Data Value. The marks are labeled by Question, Location Desc and Year Year. The data is filtered on Topic and Calculation1. The Topic filter keeps Tobacco. The Calculation1 filter keeps True. The view is filtered on Year Year, Question and Location Desc. The Year Year filter keeps 2011, 2012, 2013 and 2014. The Question filter keeps Current smokeless tobacco use among adults aged >= 18 years. The Location Desc filter keeps 22 of 55 members.

[**Tools Used**: **Calculated Field, Parameter**]

Tobacco use remains the leading cause of preventable disease and death in the United States. It is a leading cause of work disability in the United States and one of the most common chronic conditions in the nation. Year is used to create parameter and using this parameter, years can be changed from parameter control to check number of current smokeless tobacco use among adults aged>=18 years for the states ending with 'A' between years 2011 to 2014. Calculated Field is created for Year parameter. Looking at above tree map, we can assume that Iowa is having the highest current smokeless tobacco use among adults aged>=18 years in the year 2011, followed by West Virginia in the year 2012. Iowa is again seen to be the highest in year 2013, followed by North Dakota in the year 2014.

Lowest current smokeless tobacco uses among adults aged>=18 in the year 2011 is seen for North Carolina. North Carolina is again seen to be the lowest in year 2012, followed by Florida in the year 2013.In year 2014, lowest tobacco uses is seen in Arizona.

Question 5: Display the count for every topic for United States in the year 2013.

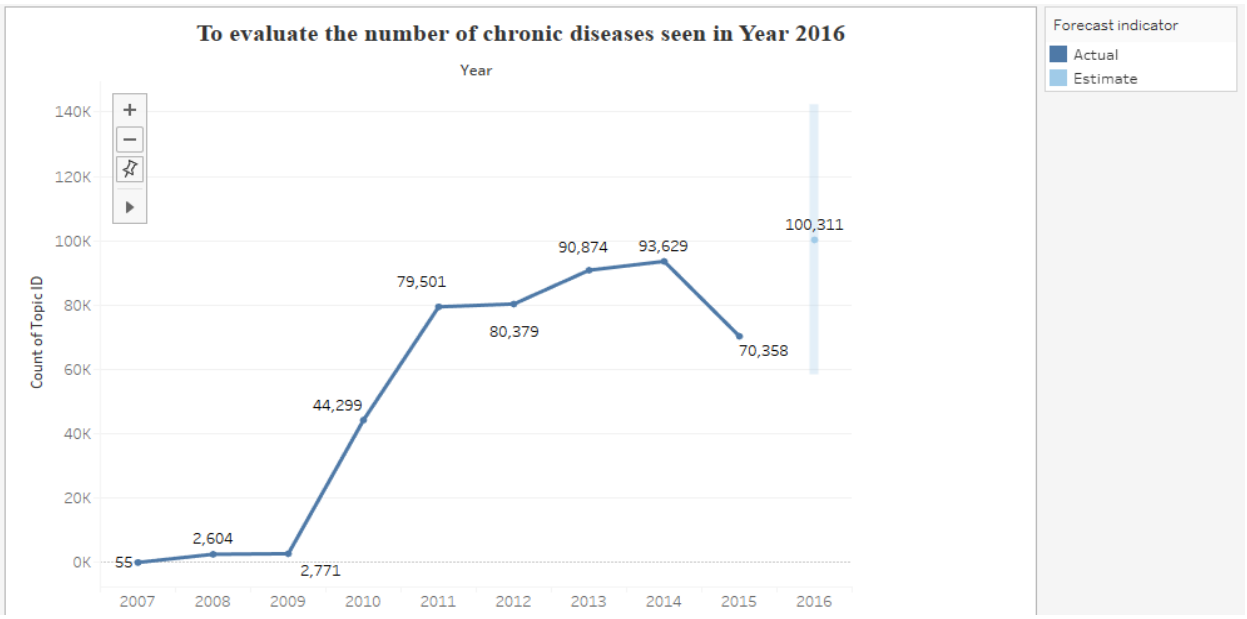
Map based on Longitude (generated) and Latitude (generated) broken down by Year Year. Details are shown for Location Desc and Topic. The data is filtered on Calculation2, which keeps True. The view is filtered on Location Desc and Year Year. The Location Desc filter keeps United States. The Year Year filter keeps 2013.

[**Tools Used: Geographical Maps, Parameter**]

Chronic diseases represented seven of the top 10 causes of death in the United States in 2010. There are 16 topics for Chronic diseases across the United States. By using Latitude and longitude geographical map has been plotted and serves as count for every topic in year 2013. Topic is used to create parameter and using this parameter, topics can be changed from parameter control to check number of chronic diseases associated with every topic in United States in year 2013. Highest count in United states is seen for Chronic Obstructive

Pulmonary Diseases which totals upto 102 whereas Oral Health totals upto just one in year 2013. Chronic disease risk factors, including smoking, poor diet, insufficient physical inactivity, and excessive alcohol consumption, were the leading actual causes of death.

Question 6: To evaluate number of chronic disease in Year 2016?



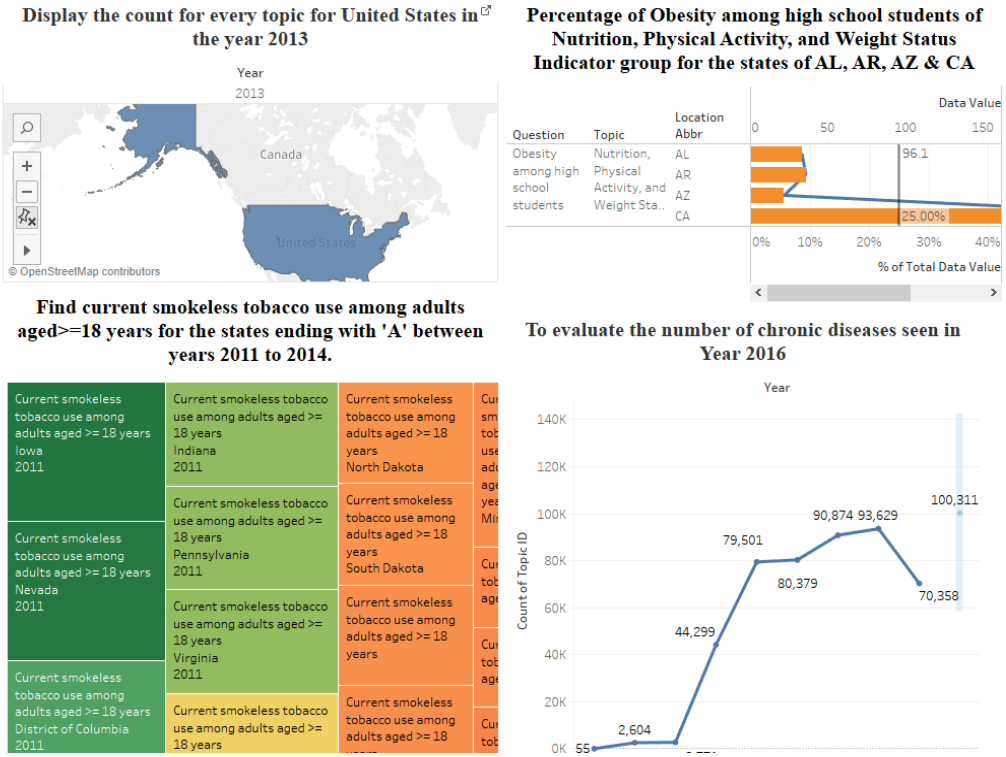
The trend of count of Topic ID (actual & forecast) for Year Year. Color shows details about Forecast indicator. The view is filtered on Year Year, which has multiple members selected.

[**Tools Used: Forecast trend Lines**]

In year 2007, there were only 55 chronic diseases observed. Every year after 2007, there has been a significant increase in the count for chronic diseases. In 2010, 44299 chronic diseases (topics) were observed where in 2013 number got increased by 50% to 90874. This number went gone increasing in the year 2014. By these forecast trend lines, we can estimate downfall trend for count of chronic diseases from year 2014 to year 2015. Count for year 2016 is significantly higher as compared to previous years. In year 2016, we can predict that about

44% chronic diseases will be observed which will be more than count in 2010 and as compared to 2014, those will be just 1% more.

D) Dashboard



E) Storytelling

Chronic diseases are an essential public health issue, which can result in morbidity, mortality, disability, and decreased quality of life. Chronic diseases contribute seven of the top 10 causes of death in the United States in 2010. According to figures from the in the year 2005, 133 million Americans—almost 1 out of every 2 adults—had at least one chronic illness. The chronic disease indicators (CDIs) are a set of surveillance indicators developed by consensus among CDC, the Council of State and Territorial Epidemiologists (CSTE), and the National Association of Chronic Disease Directors (NACDD). [2]

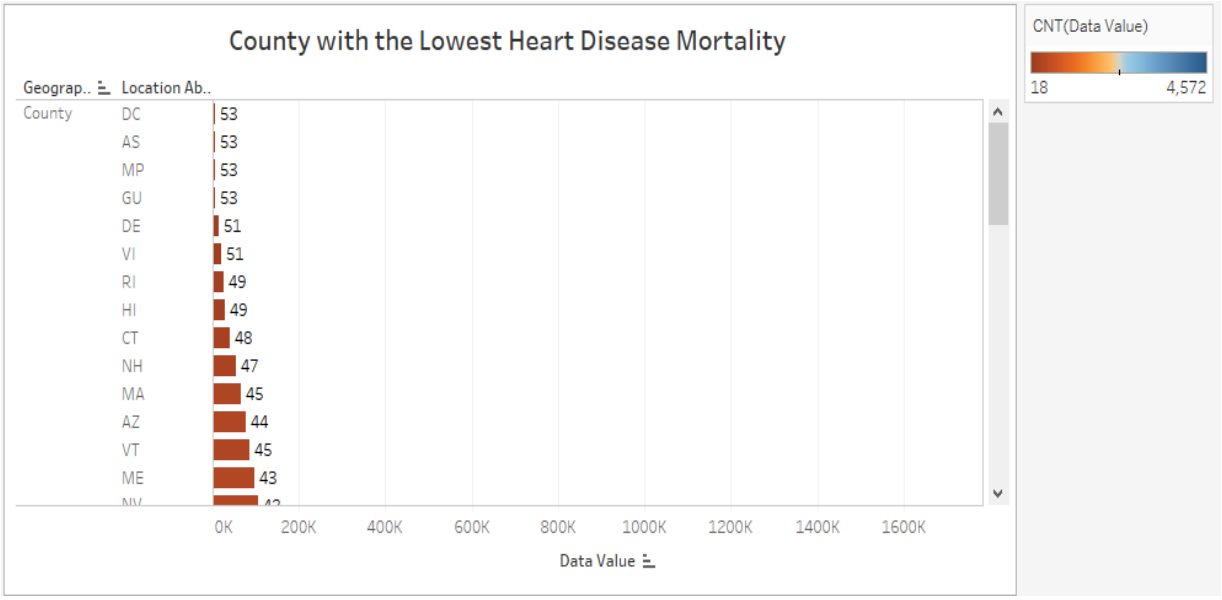
A heart attack, also called a myocardial infarction, occurs when a part of the heart muscle doesn't receive enough blood flow. Heart disease is the leading cause of death for both men and women. More than half of the deaths due to heart disease in 2015 were in men. About 6,30,000 Americans die from heart disease each year- that's 1 in every 4 deaths. Through CDC's Division for Heart Disease and Stroke Prevention, the State and Local Public Health Actions and Good Health and Wellness in Indian Country cooperative agreements help in prevention, management and reduction of the risk factors associated with heart disease and stroke. [7]

As per the sources of population health data, several data concerns for the chronic diseases must be considered. First, in instances in which it is likely that data might be compared across geographic areas and age is an important contributing risk factor, the data should be age standardized. Second, data quality, sample size, and confidentiality considerations might limit the availability of data for certain geographic areas. Last, caution must be exercised when comparing estimates for the same indicator over time. [1]

After analysis as per year 2013, among 16 topics of chronic diseases we can say the highest count in United states can be seen for Chronic Obstructive Pulmonary Diseases which possesses a total count of 102 whereas Oral Health totals upto just one in United States.

To optimize public health's efficiency and effectiveness, the Centers for Disease Control and Prevention (CDC) recommends coordinating chronic disease prevention efforts in four key domains:

- Epidemiology and surveillance: To monitor trends and track progress.
- Environmental approaches: To promote health and support healthy behaviors.
- Health care system interventions: To improve the effective delivery and use of clinical and other high-value preventive services.
- Community programs linked to clinical services: To improve and sustain management of chronic conditions. [4]



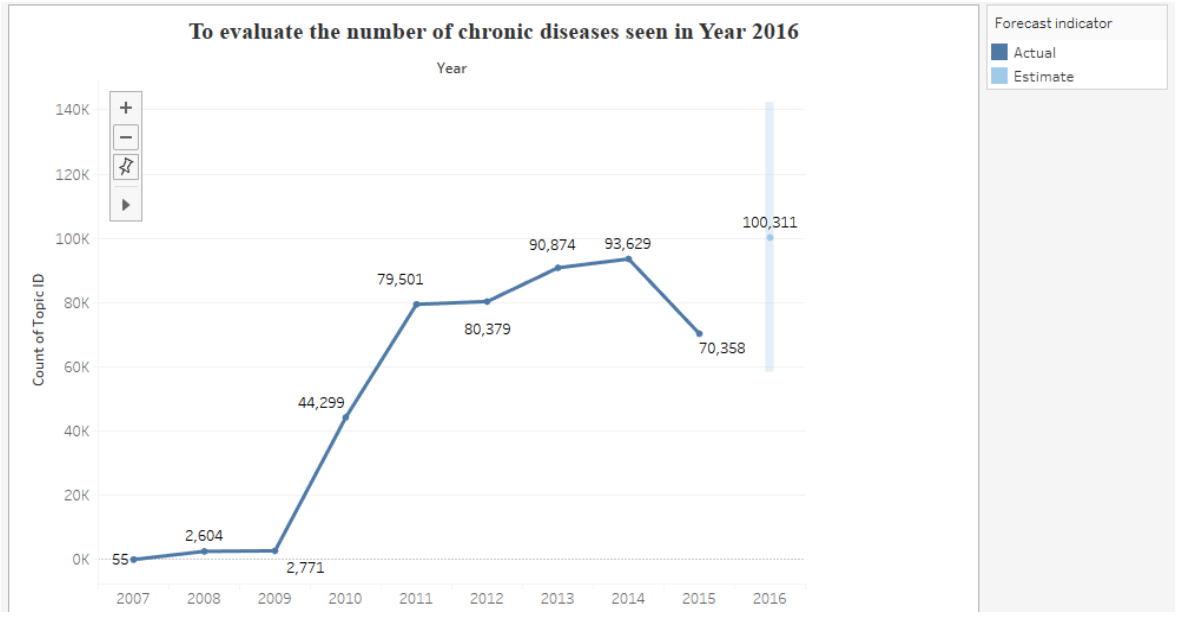
The most common type of heart disease in the United States is coronary artery disease, which affects the blood flow to the heart. Based on the study for heart disease mortality on the geographical level of county, it is observed that the lowest heart disease mortality is seen in District of Columbia and the highest is seen for the Texas county.

Division for Heart Disease Stroke Prevention aims to serve as the nation's public health leader for achieving cardiovascular health for all and for eliminating the disparities in the burden of heart disease and stroke. Following are their goals:

- Prevent risk factors for heart disease and stroke.
- Increase detection and treatment of risk factors for heart disease and stroke.
- Increase early identification and treatment of heart disease and stroke.
- Decrease recurrences of cardiovascular events.
- Foster a skilled and engaged public health workforce to address heart disease and stroke. [5]

The Youth Risk Behavior Surveillance System (YRBSS) is an epidemiologic surveillance system established by the Centers for Disease Control and Prevention (CDC) to monitor the prevalence of youth behaviors that most influence health.[7] As part of the YRBS survey, high school students report on their own behaviors as well as their height and weight, which are used to calculate Body Mass Index (BMI). These self-reported data are useful for tracking the proportion of high school students with obesity (BMI $\geq 95^{th}$ percentile) over time. [6] After scrutinizing data for obesity among high school students for the states of Alabama, Arkansas, Arizona and California, we can deduce that California is leading the race with maximum summation of data i.e. about 76.42%. Also, least percentage of obesity is seen in Arizona state i.e. about 5.62% followed by Arkansas and Alabama which are close to each other with 8.64 % & 9.32% respectively. Average is about 25% for obesity among high school students.

Tobacco use remains the leading cause of preventable disease and death in the United States. According to the 2012 National Survey on Drug Use and Health: About 1 in every 100-youth aged 12–17 years (1.1%) and nearly 4 in every 100 young adults aged 18–25 years (3.9%) were current users of smokeless tobacco and at least one other tobacco product. About 1 in every 100 adults aged 26 years or older (1.2%) were current users of smokeless tobacco and at least one other tobacco product. In 2014, current smokeless tobacco use was highest in the state of Wyoming: nearly 9 in every 100 people (8.8%) and was lowest in the states of Delaware, Hawaii and Massachusetts: more than 1 in every 100 people (1.5%) [8]



The trend of count of Topic ID (actual & forecast) for Year Year. Color shows details about Forecast indicator. The view is filtered on Year Year, which has multiple members selected.

Based on deeper analysis, it is observed for year 2007, about 55 is count for chronic diseases and every year after 2007, there has been a significant increase in count for chronic diseases. In 2010, 44299 chronic diseases were seen where in 2013 number got increased by 50% to 90874. This number went gone increasing in the year 2014. We can also estimate a downfall trend for the count of chronic diseases from year 2014 to year 2015. We can detect that the count for year 2016 is significantly higher as compared to previous years. In year 2016, we can predict that about 44% chronic diseases will be observed which will be more than count in 2010 and as compared to 2014, those will be just 1% more.

References

1. https://www.cdc.gov/mmwr/pdf/rr/rr6401.pdf
2. https://www.cdc.gov/chronicdisease/pdf/four-domains-factsheet-2015.pdf
3. https://www.cdc.gov/mmwr/volumes/66/wr/mm6644a2.htm
4. https://www.cdc.gov/chronicdisease/pdf/2009-power-of-prevention.pdf
5. https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_heart_disease.htm
6. https://www.cdc.gov/dhdsp/strategic_plan.htm
7. ftp://ftp.cdc.gov/pub/data/YRBS/2011/YRBS_2011_National_User_Guide.pdf

8. https://www.cdc.gov/tobacco/data_statistics/fact_sheets/smokeless/use_us/index.htm