

**500 Cities Local Data for Better Health**

By

Era Singh Kajal

CIN: **306605200**

California State University, Los Angeles

SUBMITTED TO

PROFESSOR SHILPA BALAN

1. **URL of data set**

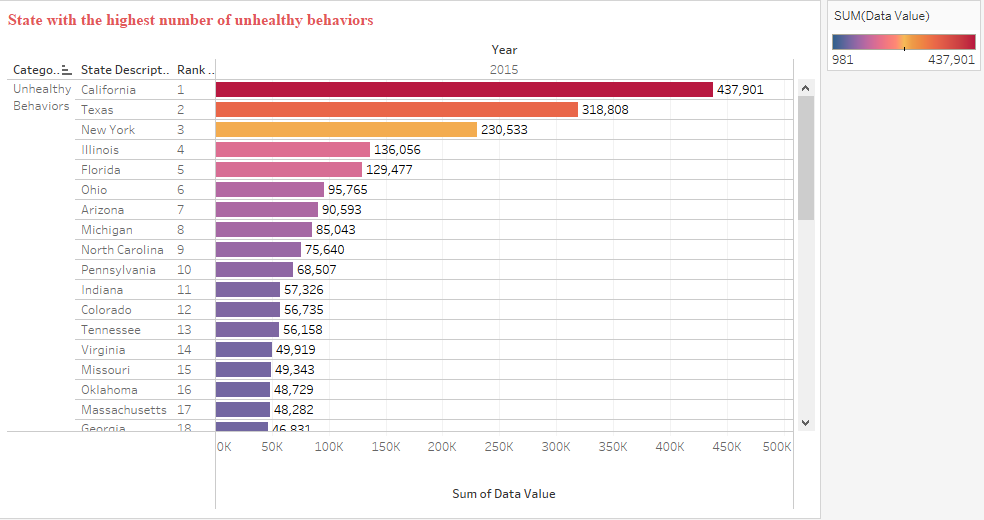
<https://catalog.data.gov/dataset/500-cities-local-data-for-better-health-b32fd>

This is the dataset for the 500 cities in US which includes 2014, 2015 model-based small area estimates for 27 measures of chronic diseases which are related to unhealthy behaviors, health outcomes, and use of preventive services. Data was provided by the Centers for Disease Control and Prevention (CDC), Division of Population Health, Epidemiology and Surveillance Branch. Despite limited data available at the county and metropolitan levels, this dataset represents a first-of-its kind data analysis to release information on a large scale for cities and for small areas within cities. This system complements existing surveillance data mandatory to more fully understand the health issues that affect the residents of that city or census tract. It includes estimates for 500 largest US cities and approximately 28,000 census tracts within these cities. These high-quality, small-area epidemiologic data can be used both by individual cities and group of cities as well as other stakeholders to help develop and implement effective and targeted prevention activities; identify rising health problems; and creating and auditing key health objectives. For example, city planners and elected officials may want to use this data to point neighborhoods with high rates of smoking or other health risk behaviors for effective interventions. Because the small area model cannot identify effects due to local interventions, users are cautioned against using these estimates for program or policy evaluations. Data sources used to generate these measures include Behavioral Risk Factor Surveillance System (BRFSS) data (2014, 2015), Census Bureau 2010 census population data, and American Community Survey (ACS) estimates.

1. **Data Cleaning**

|  |  |  |
| --- | --- | --- |
| **Scope/Issue** | **Dirty Data** | **Cleaned Data** |
| 1. **Duplicate Values** | Contained Duplicate values in the data value and limits column. | Removed the duplicates values from data value and limits column. |
| 1. **Negative Values** | Population Count had negative values which is not possible. | Removed negative values |
| 1. **Abbreviations** | Contained Abbreviations for the “State\_Abbreviation” | StateAbbr: State\_Abbreviation  StateDesc: State\_Description |
| 1. **Wrong Date Format** | Tableau was not taking this field as date and was changing it automatically to 1/1/1900 | Changed the format of the year so that tableau can identify its format |
| 1. **Wrong format** | Measure had â€“ instead of “-” | Replaced â€“with dash symbol. |

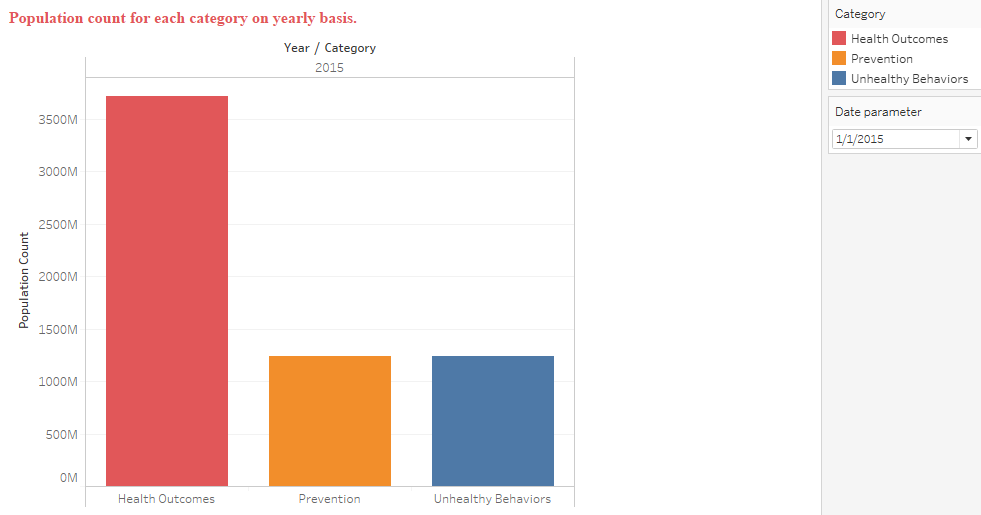
1. **Data Visualization**
2. **Which state has highest number of unhealthy behaviors in the year of 2015.**



[Tools used: Rank]

The above bar chart shows us the name of the state with highest number of people that falls under the category of unhealthy behaviors. Rank is used to tell classification of the states with their data values. It represents the data values for one-year i.e 2015. The data has been arranged in descending manner which represents the value for each state in front of the bars. It is helpful to us in a way that it provides a proper description of all the states with their data.

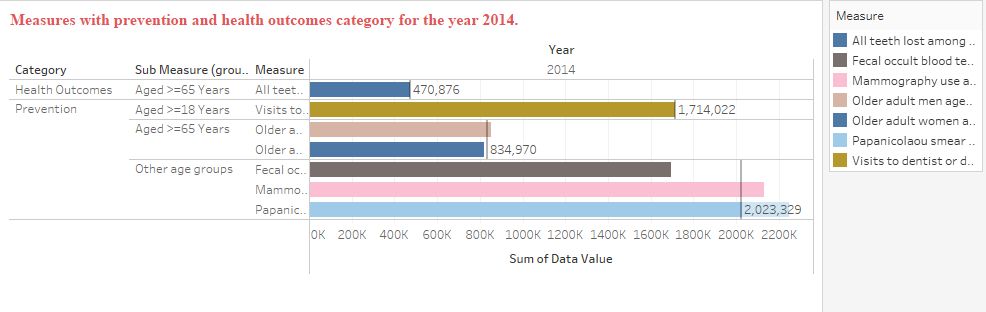
1. **What is the population counted for each category on yearly basis?**



[Tools used: Parameters, dates and calculated field]

This is the year by year stacked bar graph of the population count for different categories. Total number of population for year 2014 was less than 2015. Population count is constantly increasing for every year. Year from date is used to create parameter in the bar graph. By using parameter, date can be changed from parameter control to check population for each year for every different category. It is required to analyze causes and areas which are receiving large number of health problems. Graph shows that year 2014 shows the highest population rate of 1700M for prevention category whereas the year 2015 shows the maximum population rate of 3400M for health outcomes category.

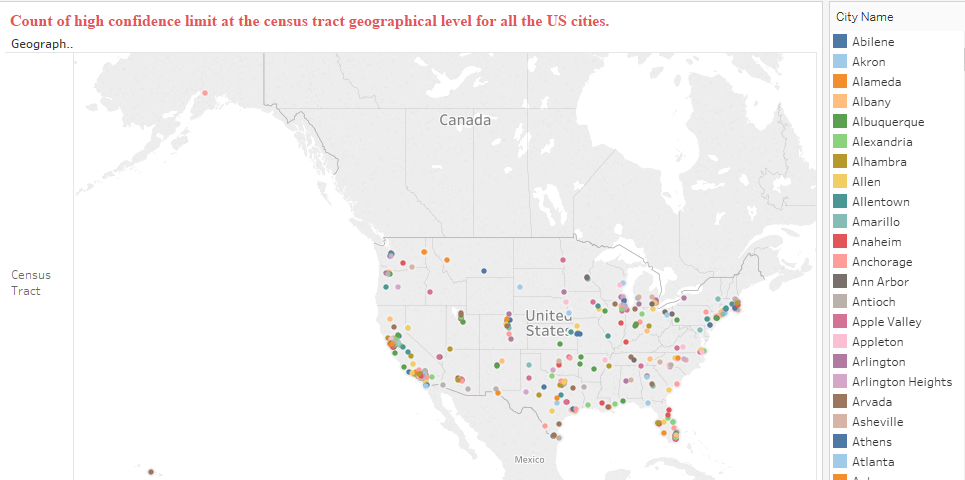
1. **Display the measures with prevention and health outcomes category for the year 2014.**



[Tools used: Grouping and reference line]

There are different types of measures for each category. Groups are categorized based on their functional role. Sub measure(groups) are made based on the different age groups for health outcomes and prevention category. 3 groups have been made for various ages like aged>=65 years, aged>=18 years and other age groups. Also, reference line is used in this bar graph which shows the average of data value for each measure in different categories. Other age groups falling under the category of prevention had the highest average, with 2,023,329, and aged>=65 years under health outcomes had the lowest average with 470,876.

1. **Display the count of high confidence limit at the census tract geographical level for all the US cities.**



[Tools used: Geographical Maps]

There are different US cities representing the count of high confidence limit. By using Latitude and longitude geographical map has been plotted. This geographical map shows us the distribution of all US cities based on geographical level and high confidence limit. There are some states which don’t have that high count of confidence limit. California is leading the race with maximum number of high confidence limit.

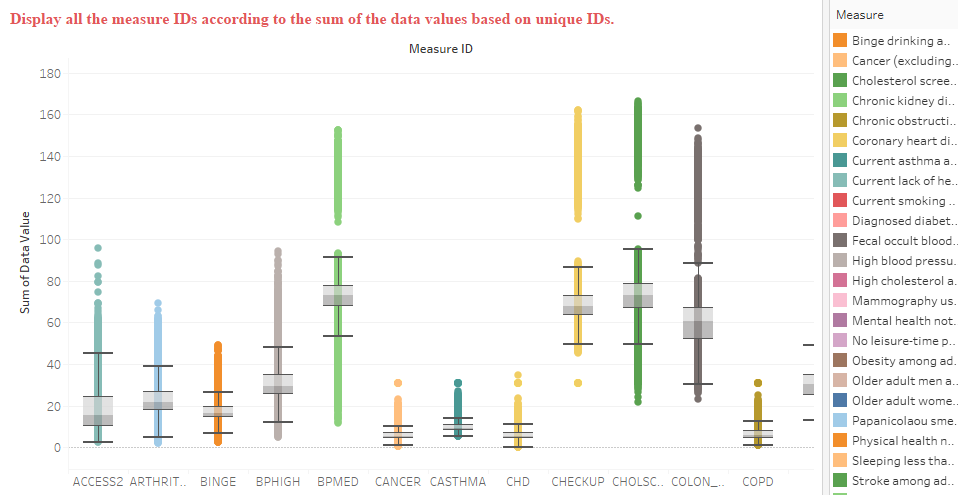
1. **How has the confidence limit changed over the past two years? Is the low confidence limit approximately same as the high confidence limit?**



[Tools used: Dual Axis Chart, Trend Line]

Every year there has been a significant number of increase in the confidence limits. In the dual axis chart, low confidence limit is compared with high confidence limit. In 2014, low confidence limit was 10M whereas in 2015 high confidence limit was 13M.Simultaneously, trend line is also used to show the increase and decrease of the both the confidence limits at a steady rate. This number is significantly higher as compared to previous year.

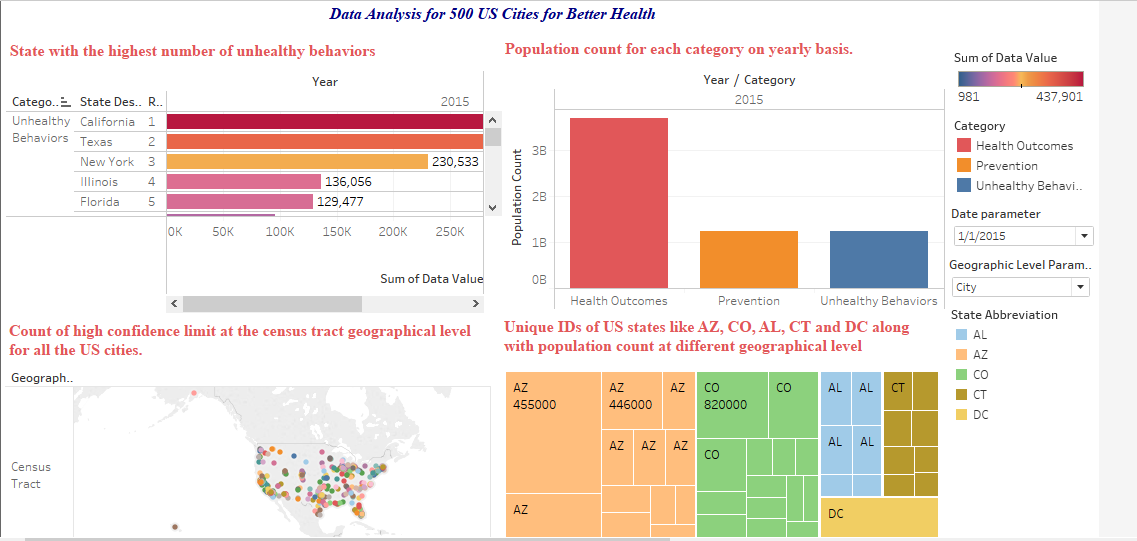
1. **Display all the measure IDs according to the sum of the data values based on unique IDs.**



[Tools Used: Box and whisker plot]

From the above Box & Whisker plot, each measure is represented by a different whisker. Every measure ID is shown along with measure based on their unique ID and sum of data value. Measure ID PAPTEST reaches the highest sum of data value of 176.5 where the unique ID is 2545560 (city geographical level) and the measure ID CHD has the lowest sum of data value of 0.2 where the unique ID is 1703012-17043846411 (census tract geographical level).

1. **Dashboard**



1. **Story Telling**

The Behavioral Risk Factor Surveillance System (BRFSS) is a federally assembled cellular and landline telephone poll conducted by state health departments [1]. Its goal is to tabulate frequencies of chronic illness, any behaviors that may produce them, and the patient outcomes they result in across 500 U.S. cities. In the data analyzed here, 27 illnesses were examined by the Centers for Disease Control and Prevention (CDC), Division of Population Health (DPH), Epidemiology and Surveillance Branch (ESB) [1].

BRFSS Data is collected at the census tract scale [1], a measure created by the United States Census Bureau to organize the decadal census of the U.S. population by neighborhood. 28,000 census tracts were investigated, comprising these 500 cities [1]. The CDC cautions end users of this dataset from making any “program or policy” evaluations due to the BRFSS’ “small area model”. It is instead intended to “identify emerging health problems” and then facilitate data-rich responses to these problems [1].

Reviewing the data visualization, question one ranked states by those with the highest number of unhealthy behaviors in calendar year 2015. The top five were California, Texas, New York, and Illinois, and Florida, corresponding to population at the time of the study [2]. Question two asked for the count within each category, year-by-year. 2014 showed almost identical tallies for the health outcomes and unhealthy behaviors categories; meanwhile, 2015 showed almost identical counts for the prevention and unhealthy behaviors categories.

Question three considered which sub-measures could be broken out within counts of the health outcomes and prevention categories; it was found that questions of health outcome were only asked of those over the age of 65. Meanwhile, questions of prevention were asked of all adults. Question four looked at high confidence limits at the census tract scale, mapping the cities included within the dataset. Question five looked at changes in confidence limits during the study interval, and found that it had increased approximately 30%. Lastly, question six provided a graphical representation of median and 1-2 standard deviations within each category analyzed.

The years analyzed from the study were 2014 and 2015; this timeframe was characterized by the first full years of implementation of the Patient Protection and Affordable Care Act (PPACA). The PPACA represented the largest single expansion of private health insurance coverage in United States history. It also expanded medical coverage for low-income Americans via the joint federal-state “MedicAid” program for childless adults and other previously ineligible populations for the first time [3].

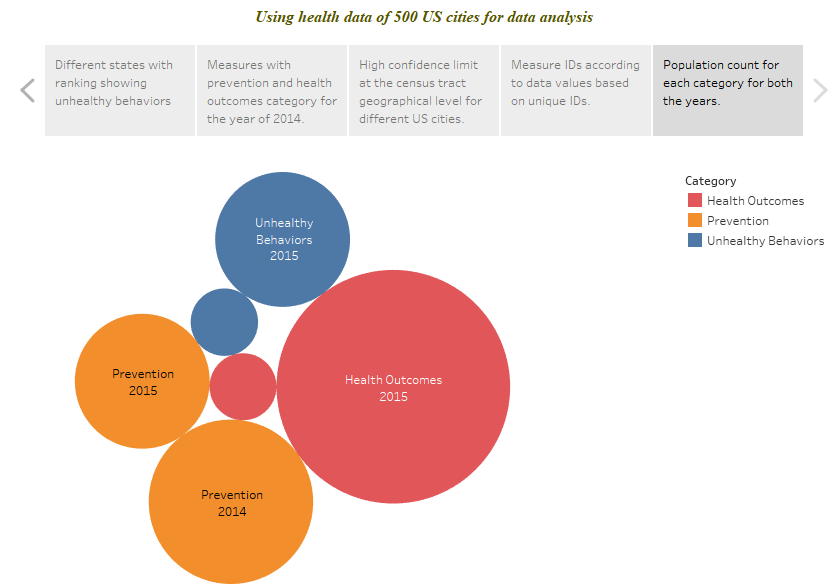
Based on the 2014 and 2015 study findings (Figure below), preventive events decreased by about 50% year-over-year. This is surprising in light of PPACA’s expansion of health coverage to an estimated 17 million previously uncovered individuals in the United States [3]. This number was artificially limited by subsequent judicial rulings that allowed many Republican-controlled states to decline to extend MedicAid the way many Democratic-controlled states eventually did.

Also deteriorating, unhealthy behaviors approximately quadrupled from 2014 to 2015. One possible explanation for this observation, a deteriorating economic climate, does not comport with the fact that economic and job growth were improving during this time. It is unclear whether a rising number of MedicAid patients entering that program affected the BRFSS’ counts significantly in such a way as to cause this (otherwise, unexplained) downward trend.

The overall number of health outcomes increased markedly, by a factor of about 12.5. This may have been due to the rising number of patients entering the insurance and/or hospital space due to the ACA’s expanding pool of covered individuals through both the MedicAid and exchange populations. This increase may then explain all the observed increases in unhealthy behaviors, but it does not explain the decrease in prevention. As-yet-unknown complicating factors may be at play here.

The BRFSS carries significance for a variety of reasons. Its target of analysis is health outcomes in America, and therefore shines a spotlight on the cost-intensive problem of chronic illness. According to the RAND Corporation, 60% of American adults live with at least one chronic illness [4]; 42% have more than one such condition simultaneously. The most common chronic disease diagnoses include cardiovascular disease, stroke, cancer, type II diabetes, and obesity [5]. The most common of these is cardiovascular disease, while obesity is increasing in prevalence most quickly [5]. 30 million adults have a staggering five or more such conditions.

This final cohort accounts for over 40% of all health care spending [4]. In fact, 70% of all deaths in the United States were the result of chronic conditions [5]. Related annual health spending totaled $347 billion at the open of the decade, representing almost 90% of all health care costs. These totals would be even higher, were it not for the fact that 30% of uninsured patients do not have sufficient funds to pay for their treatments. If America wants to get a handle on both its existing deficit problem, and debt overhang, addressing such gaps in the health care system will go a long way towards restraining costs. In turn, the people of the country will become healthier, and its future economic outlook will become brighter, a clear win-win that should be central to any future policy priorities.



**References**

[1] <https://www.cdc.gov/500cities/methodology.htm>

[2] <https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_population>

[3]<https://www.cbpp.org/research/how-health-reforms-medicaid-expansion-will-impact-state-budgets>

[4]<https://www.rand.org/blog/rand-review/2017/07/chronic-conditions-in-america-price-and-prevalence.html>

[5]<https://www.huffingtonpost.com/entry/americas-chronic-illness-crisis_us_577fcd82e4b05b4c02fc5a3a>