**INFO 5709**

**DATA VISUALIZATION PROJECT REPORT**

**FAITH MWANGO CHAMA**

**FALL 2021**

**Child Victims By Age in the U.S.**

**INTRODUCTION**

Child abuse has been rampant over the years, as this is a sensitive but crucial topic, there has been child abuse awareness all around the globe. Child abuse is ambiguous as it can be defined by different terms, abuse on children may be presented in many forms including emotionally, physically and sexually. According to Herzberger (1990) “Variations in the definition emerge because some researchers concentrate on the perpetrator’s act (e.g., Straus, 1979), while others examine the act, the consequences, and the circumstances surrounding the victimization”.

There have been several programs established to reduce the rates of child maltreatment all around the world. However, maltreatment damages development and emotional health more frequently than physical safety (Gilbert, 2008). To handle challenges in awareness of emotional trauma connected to child maltreatment and other sources, it has been said that children should be examined for trauma subjection and indicators at all child-care visits by asking questions that would highlight if the child has undergone any incidence which upset them or the family. If the child agrees and reports experiencing trauma, the child will undergo a brief examination that indicates any symptoms of PTSD( post-traumatic stress disorder). However, the impact this screening has on children in a health setting such as a clinic has not yet been assessed.

**About the Dataset**

The dataset topic is selected from a public dataset called Health Data. Health Data is a subversion of the US federal government, specifically used to provide access to various datasets ranging from the health sector. Health Data dataset search has indexed an estimated 3470 of datasets from all over the US.

The dataset topic “Child Victims by Age” was searched using Health Data; it was a customized search that provides information on the number and rates of child victims by speed and by the state in a single year in the United States of America.

The link below is the chosen topic dataset from the Health Data:

<https://healthdata.gov/dataset/Child-Victims-by-Age/xn3e-yyaj>

This dataset consists of the following attributes:

1. Year; represents the year in which the child victims dataset was recorded, this particular dataset is focused on the year 2019.
2. Age: represents the age of the child victims, for this dataset, the child victims range from 1- 17 years this also includes children who died from child abuse and neglect.
3. State: Represents the states in the US that record child maltreatment.
4. GDP: Represents the gross domestic product for each state in the US.
5. Gender: This dataset does not represent data based wholly on gender.
6. Rate: represents the frequency of each state of child victims ranging from ages 1-17 years.
7. Unborn: this represents a unique count of child victims who are not born, 18-21 yrs or those incidents which were reported but their ages were missing or unknown.
8. Total victims: represents the sum of child victims reported by each state in the year 2019.

**Tools Used in Project Report**

The following tools will be used for this project report:

1. Python;-  is used for the EDA (Exploratory Data Analysis) section.
2. Tableau;- is used for the data visualization section.

**Exploratory Data Analysis (EDA)**

The exploratory data analysis was carried out using python. The python application was accessed through Colab Google. The child victim excel was uploaded into the application. The first step was to check if there were any null values, no null values were found in the dataset, as indicated below:

Four questions are then formulated and its from these questions that the hypotheses are created, the following are the questions that are used for the exploratory data analysis:

1. What is the state with the highest child victims?
2. What is the state with the lowest child victims?
3. Which age group has the highest child victims ?
4. Which age group has the lowest child victims?

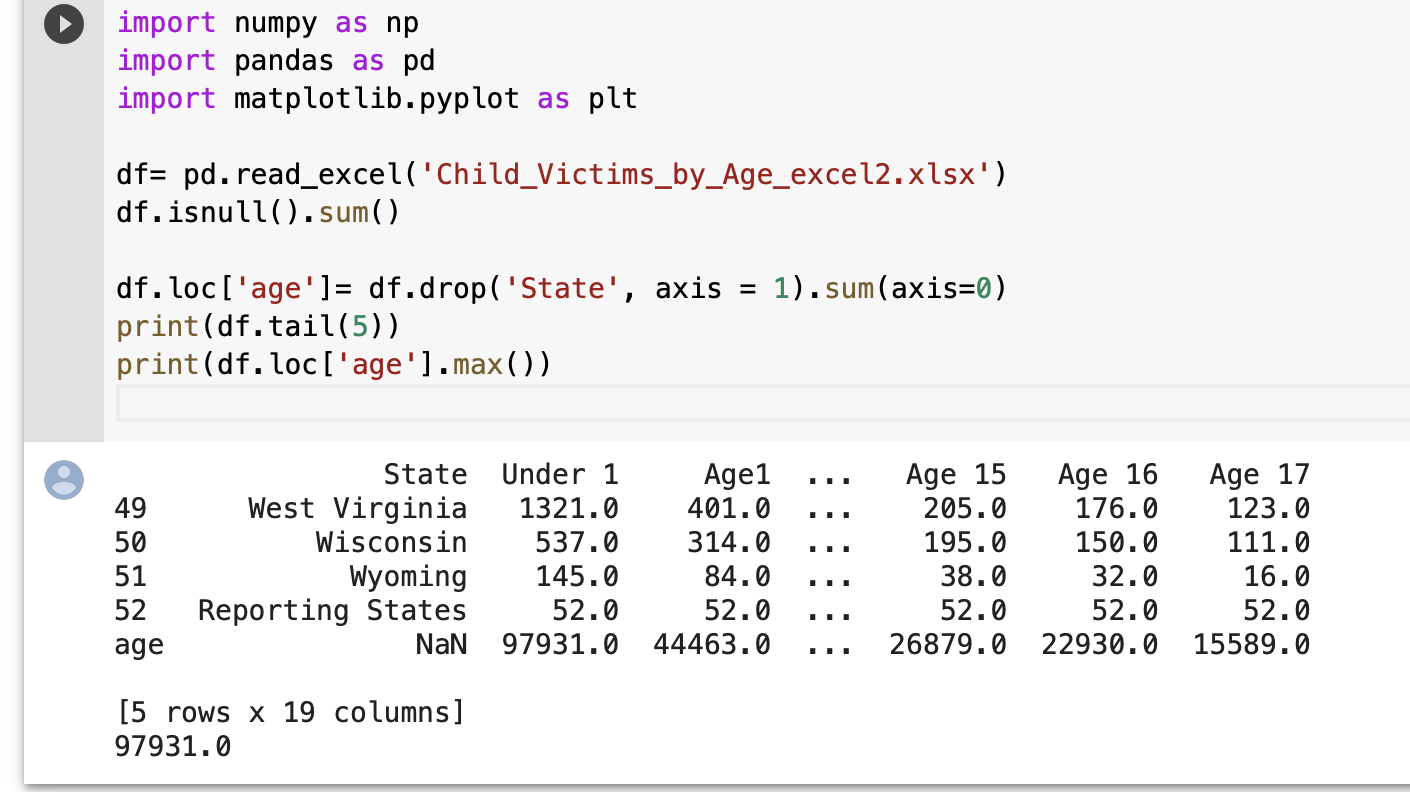
**EDA for Question 1:** What is the state with the highest cases of child victims?

From this analysis, it is shown that the state New York has the highest cases of child victims with a total of 67,127 cases.

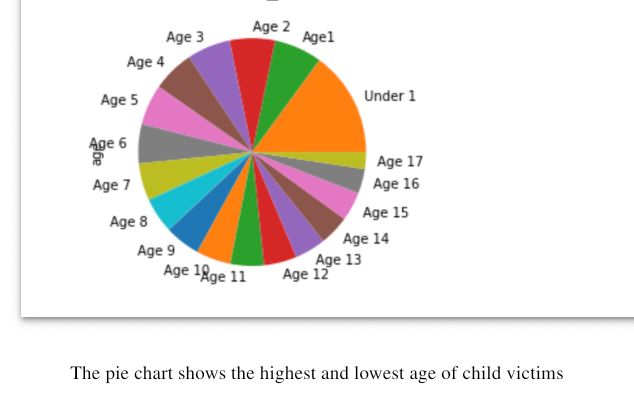
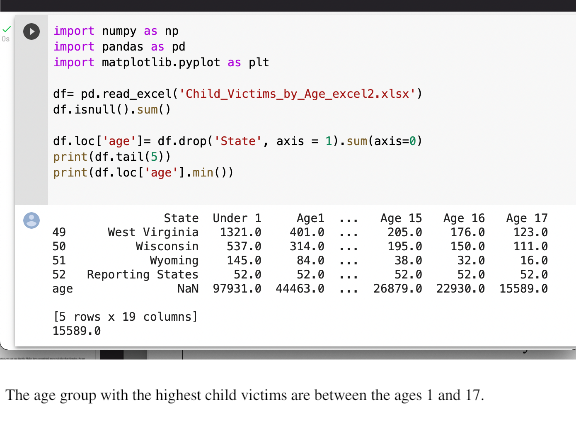
**EDA FOR Question 2:** What is the state with the lowest child victims?

The state of Vermont has the lowest child victims with the total number of cases being 845.

The bar graph below shows the highest and lowest numbers of child victims, state wise. The bar graph answers both of question 1 and question 2.

**EDA for question 3:** Which age group has the highest child victims ?

The age group with the highest number of child victims is under the age of 1.

**EDA for question 4:** Which age group has the lowest child victims?

**Hypotheses**

Based on the EDA results, this dataset is appropriate to answer the hypotheses because the dataset allows a high likelihood in the hypotheses being accurate. The dataset contains characteristics that provide evidence for the plausible hypotheses. The following are the hypotheses generated from the results of the EDA:

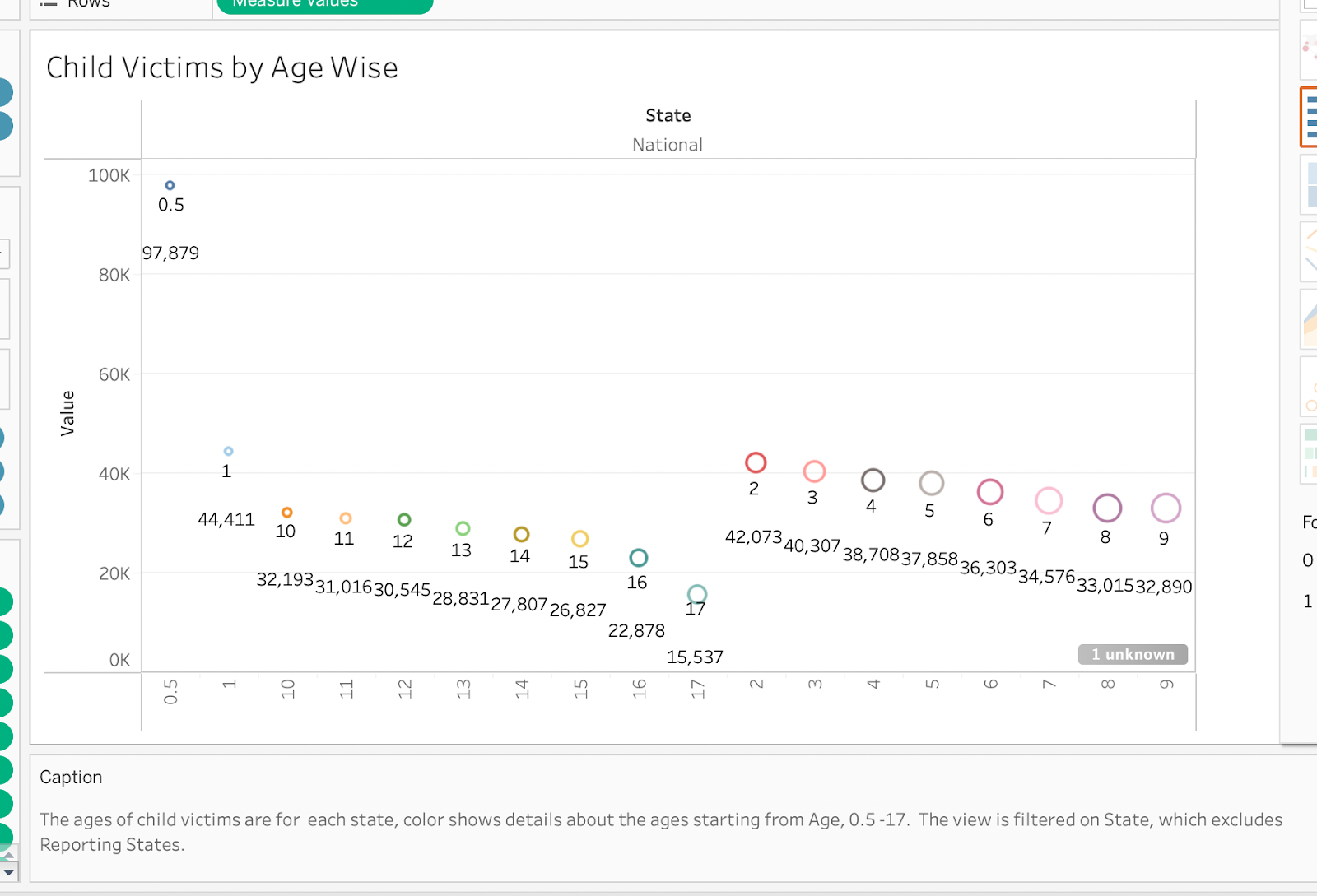
1. The number of child victims cases increases  as the age reduces.
2. The number of cases increases with high GDP in each state therefore the high economic states have the highest cases of child victims.
3. The more crowded the state is the higher of number of child victims.

**Data Visualization for Hypothesis One**

1. The number of child victims cases increases as the age reduces

The data visualization above indicates the child victims according to their age. The dataset was applied to Tableau. On the dashboard, various options show the types of graphs.

**Graph Format:**The graph chosen to answer the first hypothesis is called a scatter plot; the reason for selecting this graph was to show the correlation of child victims' age and the case count, the attributes in this graph include Age, and Total Victims.



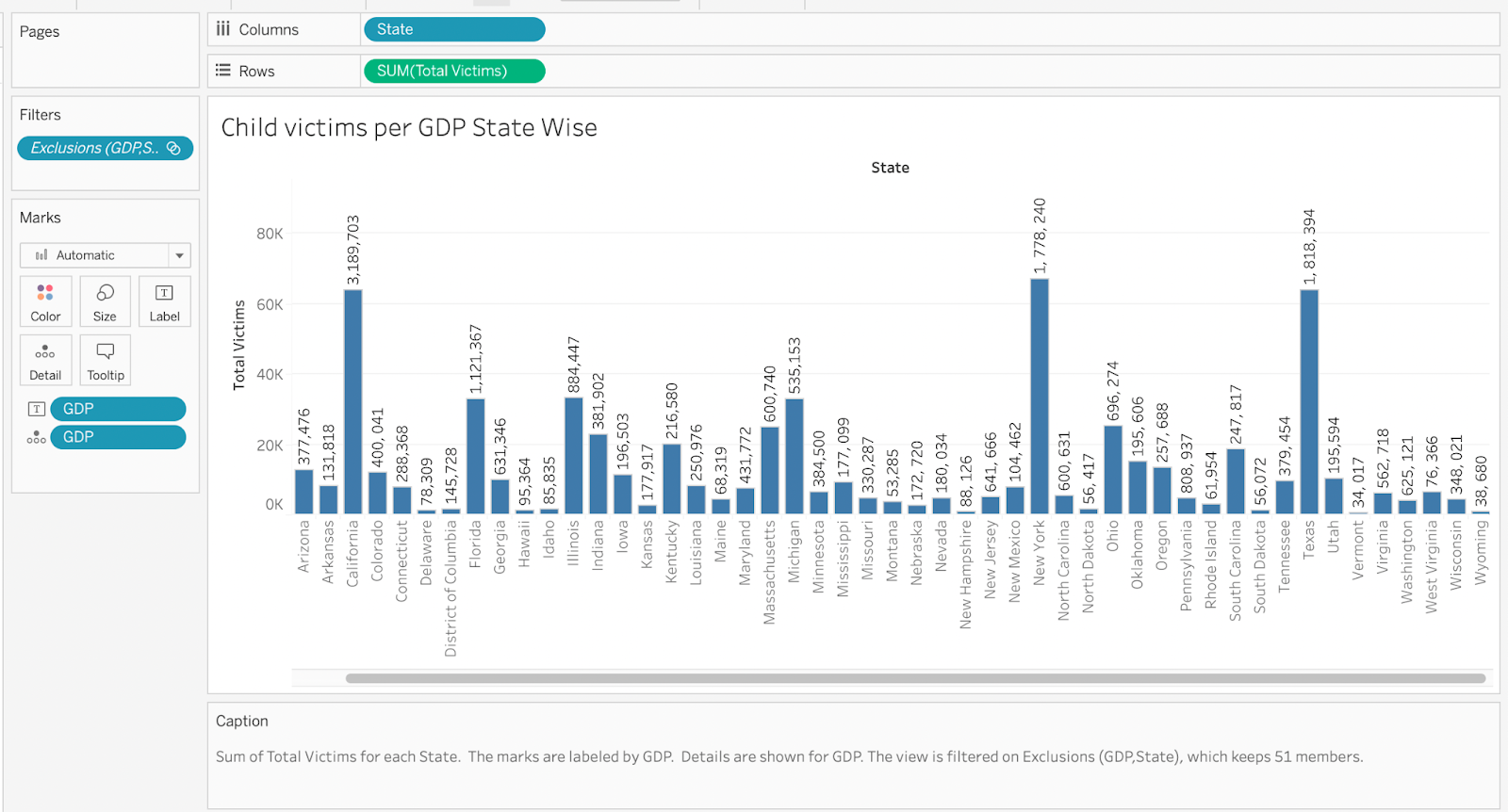
**Result:**After choosing the scatter plot as a visualization format, elementary perceptual tasks were performed. These included filtering the measure names and the measure values, measure names were filtered by color and size, and then measure values were filtered by the label. The reason for choosing the filter feature is because it allows us to view the child's age and the total case count per age. The size of the circles in the scatter plot represents the ages of the child victims. It is apparent in the visualization that as the age reduces, the circles' size also reduces. Therefore, the results of the first hypothesis are correct, and the data visualization indicates the total number of cases per child's age. The higher the child victim case, the lower the age; therefore, the number of child victim cases increases as the child's age reduces.

**Data Visualization for Hypothesis Two**

**2.** The number of cases increases with high GDP in each state. Therefore, the high economic states have the highest cases of child victims.

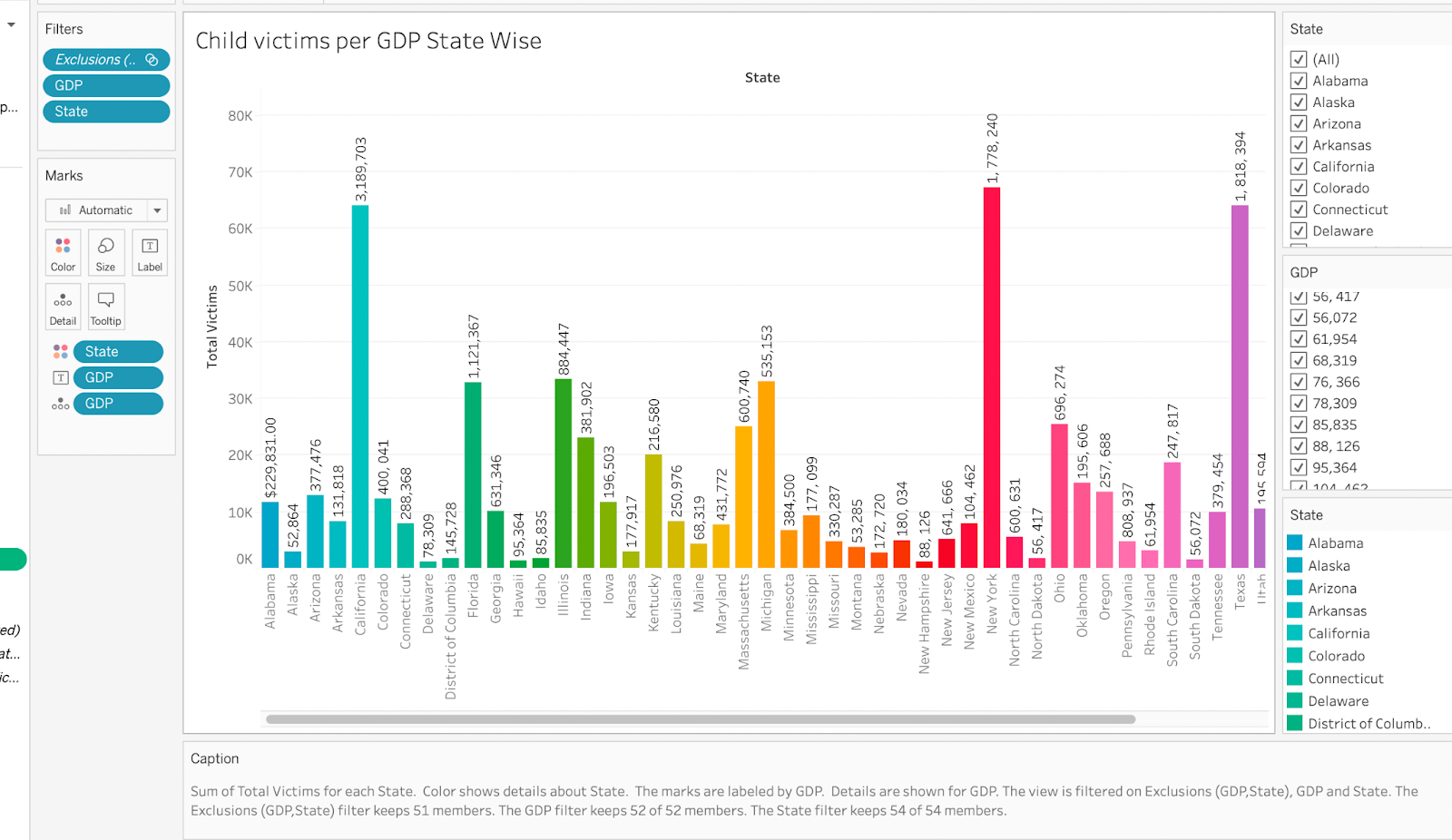
The data visualizations below illuminate the total number of victims in each state according to the states’ GDP.

**Graph Format:**The format used for this visualization is the horizontal bar graph; this bar graph was chosen because it accurately interprets the qualitative data in the dataset. The dataset contains some attributes that are not numerical, and these non-numerical attributes in the dataset are the states. The bar graph is suitable to prove this hypothesis because it emphasizes the size of each of the categories that are being measured, the attributes in this graph include, GDP, Total Victims and State.

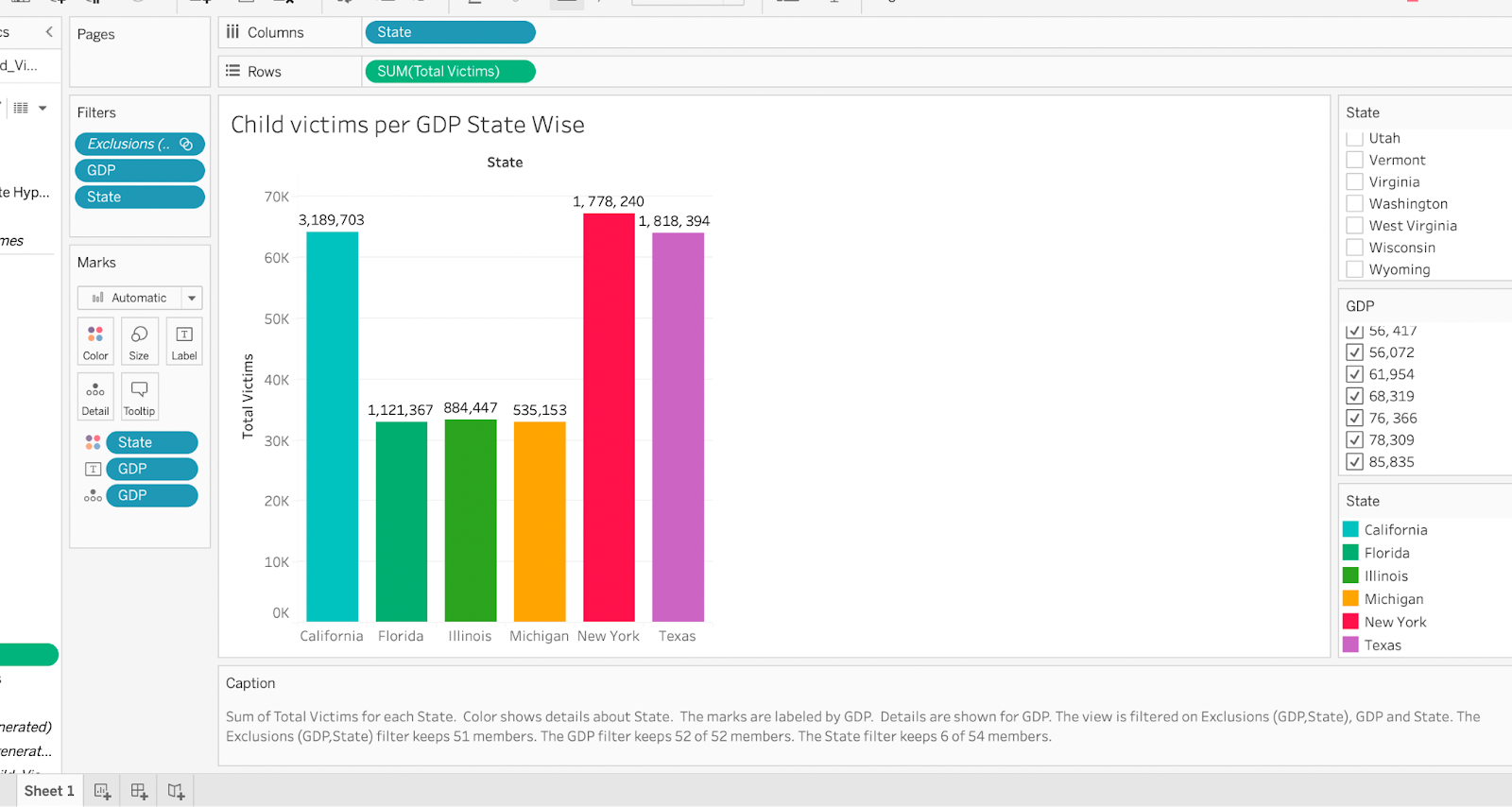


**Results:** After choosing the horizontal bar graph from the map pane that is located on the right side of the dashboard in Tableau, I then picked the geo dimension called state and dropped it in the label filter, this dimension was then again dropped in the text filter (this allows us to see the names of the states with the GDP). Total victims was added so as to measure the number of total victims per GDP State wise (the total victim figures range from 0K to 80K) . As shown in the image above.

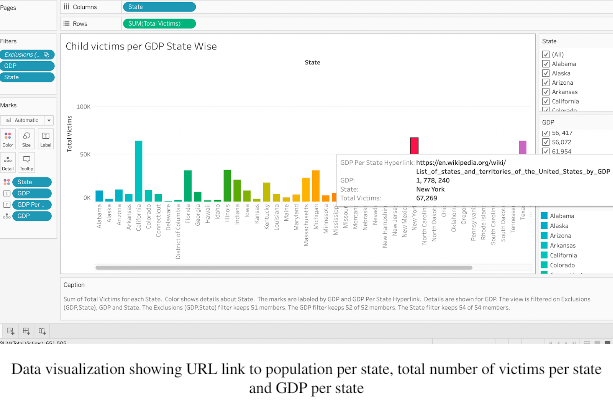
To get a clear picture of the visualization, I dropped State into color, the color feature is important because it produces more effective and readable results. If the states are filtered in color, it will enforce a strong message for the visual display, by looking at the color, we are able to see which states are highlighted as having the highest GDP and the lowest GDP. The color chosen for this visualization is called Hue circle, the states are color-formatted in alphabetical order, as shown below:



The next step was to show which states have the highest GDP and the highest child victim cases, to achieve this results, an interactive filter was applied to the data. I needed to show the filter on the GDP, State and Total Victims. The interactive section for these dimensions appear on the right side of the dashboard. I chose “only relevant values” for each dimensions and the results in the image below the states with the highest GDP that contain the highest child victim cases:



By using this interactive filter, the results above have provided evidence to the hypothesis that indeed the number of cases increases with high GDP in each state, which means the states with high economies have high child victim cases. Lastly, for reference purposes, I included a link to the data in the dashboard that shows the GDP per State in the US. When one clicks on the category in the graph, they will be redirected to the site that contains information of the states GDP as shown in the below:

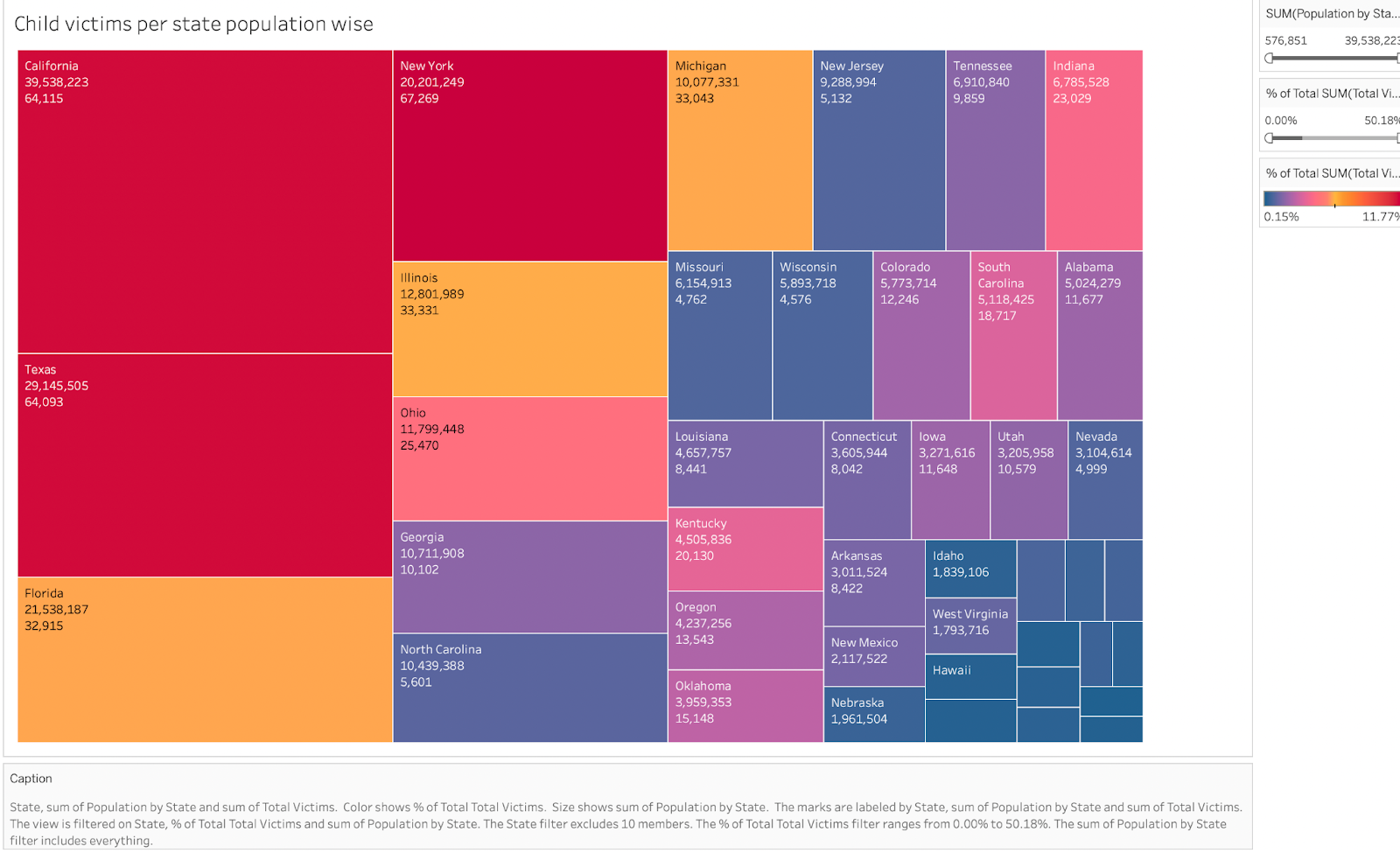


**Data Visualization for Hypothesis Three**

**3.** The more crowded the state is the higher of number of child victims.

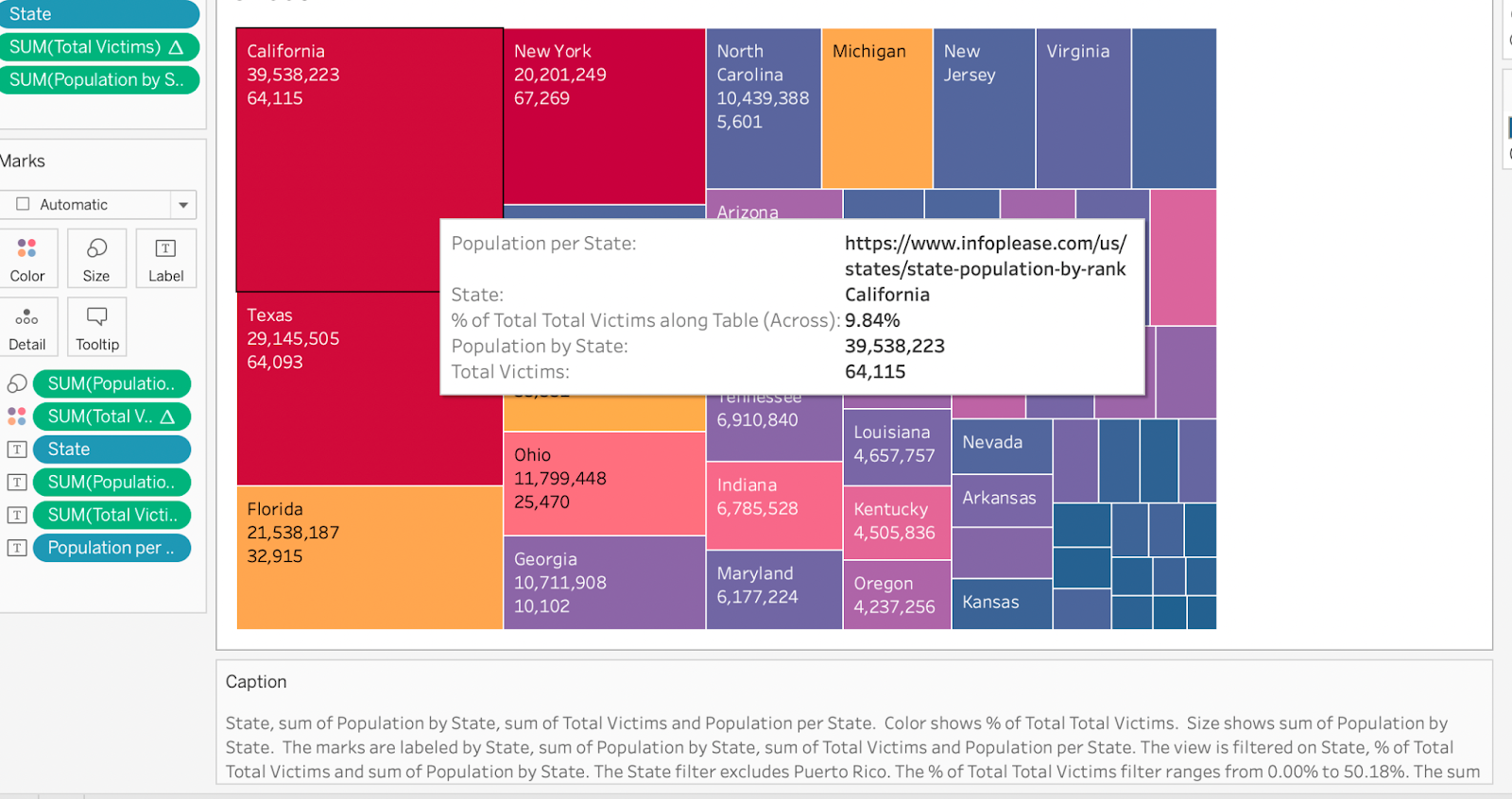
The visualization below tells a story of the child victim cases per state population wise.

**Graph format:** For this type of visualization, the format that was applied on the dashboard is called a tree map. The reason for using this graph format is because the word map represents data that is done either by country, state or city. To answer the third hypothesis a hierarchical structure needed to be formed, the tree map is best for displaying hierarchical structures because it interprets data in an easier way by showing the largest figures to the lowest figures. The tree map contains the dataset attributes such as State, Total Victims and Population Per State. Below is the data visualization used to investigate hypothesis three:



**Results:** The first EPT I performed was to drop population per State into the label filter, then dropped Total Victims into color, for further reference, I double clicked on State and dropped Total Victims into Label; doing this allowed the sheet to present the population figures of the state, figures of total victims in each state and the name of the state. For interaction, I used the quick table calculation and selected percent of total to show the percentage wise of the Total Victims in each state. In this visualization, the sunrise-sunset diverging color was applied because the data on the worksheet contains continuous measures and Geo dimensions.

A URL was added to the worksheet for reference to show the website that can confirm the population figures for each state in the US as per image below:



From the tree map indicated above, the third hypothesis is justified that indeed the more crowded a population is, the higher the child victim cases. The tree map is arranged in a way that highlights the biggest states to the smallest state. For example, in the image, California and New

York have high population and the total number of child victims is high too with a percentile of 9.48% . Therefore the hypothesis (The more crowded the state is the higher of number of child victims) is correct.

**Discussions**

* The analysis from the developed data visualizations shows that most child victim cases are mostly centered in infants under the age of one.
* From the data visualizations, it was further implied that states that have high economy (GDP) also have higher numbers of child victims.
* Further evidence shows that, these states in the US that have high population also have high GDP, and increased cases of child victims. For example, the economic rate for California is 3,189,703, the population for the state is 39, 538, 223 and the total number of child victims reported is 64,115 as compared to the state of Vermont which has a total population of 643,077 and a GDP of 34, 017, Vermont total number of child victim cases is 845. From these figures, the data visualizations were able to prove that all the hypotheses were accurate. The analysis conducted is helpful for users who want to review the child maltreatment case count in the US and for them to get a clear understanding of which states contains the highest and lowest number of child victims.

**Emerging Trends/ Future Work for Child Victims in the U.S.A**

As there have been reports of different types of child maltreatment, the cases continue to surge. Since the beginning of the pandemic, it has been reported that child abuse has increased. According to U.S.News (2021) “exactly what triggered the surge is not fully understood, but similar upticks in child abuse. A pediatrician who was not involved in the new research suspects COVID-19 and pandemic-related stresses created a ‘perfect storm’ for abuse”.

Hence, there are more social services programs being implemented that explore effective delivery systems that may help improve the understanding of different types of child abuse.

Evidence based programs are now on the rise, these programs support high-quality studies of child maltreatment. Evidence based programs, include the use of data visualization as shown in this project. The government continues to create laws and policies that address child maltreatment and capital punishment for perpetrators.

**References**

Gilbert, R., Kemp, A., Thobun, J., Sidebotham, P., Rdford, L., Glaser, D.,

   MacMillan., H.L. (2008). Recognising and responding to Child

   Maltreatment. Published online December 3, 2008 DOI:10.10

    1016/s0140- 6734(08)61707-9

Health Data. (2019) <https://healthdata.gov/dataset/Child-Victims-by-Age/xn3e-yyaj>

Herzberger, S.D. 1990. The American Behavioral Scientist (1986-1994);

   Thousand Oaks Vol. 33, Iss. 5, (May/Jun 1990): 529.

Infoplease. (2021). State Population by Rank. <https://www.infoplease.com/us>

/states/state-population-by-rank

U.S.News. (2021). Study: Child Abuse Rose During COVID Pandemic

<https://www.usnews.com/news/health-news>

Wikipedia. (2021). Lists of U.S. States and Territories by Population

<https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_>

by\_population