Team name and members:

Group 9

Garrett Davis https://github.com/gdavis74343/Project-2-MIST-4610
Chase Bell https://github.com/cbello1618
Megan Hanson https://github.com/megannhanson
Abhi Vadlamudi https://github.com/abhi33ravens/MIST-4610-Group-Project-2

Include information about the name of the team based on the team name assigned to the team on ELC. Also include the names of all team members as well as links to their corresponding github repos that have the project on them.

Describing your dataset and what data it contains:
Where was it obtained, what are the dimensions of it (rows and columns), what are the various columns, data types, etc. Describe it in sufficient detail so that an uninformed reader would understand the dataset.

The dataset is sourced from the Centers for Disease Control (CDC). According to the CDC's website, Alzhiemer's is the most common type of dementia, which is a progressive disease that affects the parts of the brain that control thought, memory, and language. It begins with mild memory loss and eventually leads to a loss of independence due to inability to complete daily tasks. It affects almost 6 million people, indicating that the prevalence of this disease is very high and needs to be better understood. The data collected contains information regarding demographic information, location (whether in the United States or a US territory), and health implications being experienced by people with Alzhiemer's.

The significant columns in our data set include:
YearStart - Year when the Alzhiemer's for the patient began
YearEnd - Year when the Alzhienmer's ended (may be ongoing)
LocationDesc - State where the patient lives
Topic - Health implications experiences by people with Alzhiemers
Stratification1 (Age group) - How old the person is, grouped from 50-64 or 65 and older
Stratification 2 (Gender) - States if the patient is male or female

The 2 questions the team generated and why they are interesting and important:
What are the questions, and why each question is important. Importance can be evaluated and conveyed in a variety of ways including, social, economic, cultural and other factors. Also indicate how they are tied to the data set or sets being used.

Question 1: What health implications are most frequently experienced by people with Alzhiemer's in different age groups? Is there a significant difference in symptom type and frequency for different age groups?

This question is important because it helps to indicate common symptoms of Alzhiemer's and the data provides people with an understanding of a reason why they could be experiencing these symptoms. It may help people to get an earlier diagnosis of Alzhiemer's and receive the necessary treatment to improve their quality of life and get ahead of the disease. Since Alzhiemer's is a disease that mostly affects older segments of the population, it is important for people and health professionals to understand the most common symptoms within different age groups.

Question 2: Which states/regions in the United States have the highest population of patients with Alzhiemers by Ethnicity?

This question is important because it will help health professionals in the United States to understand where Alzhiemer's is most prevalent. The states with the highest population of Alzhiemer's patients will obviously need more healthcare professionals who can care for these patients. According to our analysis, New York and Oregon are the two states that clearly have the greatest number of Alzhiemer's patients who are included in the dataset.

The manipulations applied to the data set as part of the analysis:
Were there any manipulations or calculations that needed to be performed on the data, what were they? Describe the purpose and how they were accomplished.

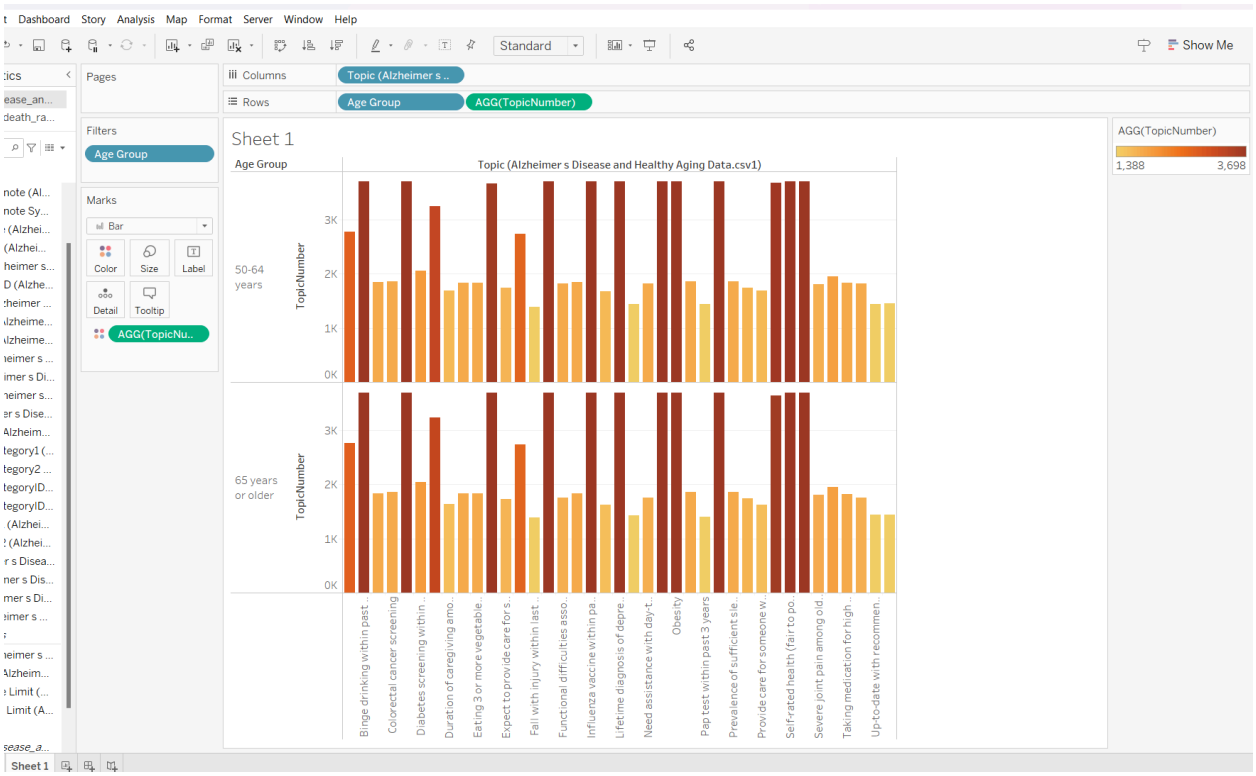
Question 1: Created a Calculated field for the 'Topic' attribute of the data to find a total count of each prevalent health implication. We moved the Age group attribute of the dataset, which separated the patients by ages 50-64 and 65+, onto the y axis of the bar graph, allowing us to seperate the graph into two bar charts with differing age groups on the y-axis, and health implications on the x-axis, with count of number of patients with these health implications indicated by the size of the bars in the chart.

Question 2: Created a Calculated field for the 'LocationDESC' attribute of the dataset, which outlines states of the United States. We also applied a filter to remove variables from the data, including regions such as Midwest and South, as well as territories outside the main 50 states, including Guam, Puerto Rico, and the Virgin Islands. Changed the variable type of 'LocationDESC' to geographic region based on state/province in the United States, which allowed us to utilize the map display function in Tableau. We added stratification 2 into the rows as a dimension. Stratification 2 filters patients with Alzeimers by Ethnicity. There is also a filter

for stratification 2 to take out male and female categories so that we are only looking at ethnicities by each map. Each different map indicates the number of patients with Alzhiemers by state by ethnicity.

Analysis and Results:
Analyze and visualize the results of your analysis and describe the implications of your analysis. Please provide any citations if required as well as supporting visualizations and analysis generated from Tableau.

Question 1
As seen in the bar graphs, the health implications experienced by Alzhiemer's patients in the two different age groups are very similar. The number of patients experiencing the different implications are very similar between age groups, but have slight differences. This indicates that the symptoms and difficulties faced by patients are not necessarily standard to a specific age range; the disease can have an onset in either of the age groups. Although the disease is much more common in adults over age 65 (according to the CDC), the adults within this dataset indicate that symptoms are not necessarily unique to age of onset. This is important for healthcare professionals and the public to be aware of.



Question2

As seen in the map below, Alzheimer's disease is most prevalent in the darker blue states (Oregan, New York, Texas, Utah, Georgia) and least prevalent in the lighter blue states(Florida, Montana, Colorado). It's hard to say that Alzheimer's exists more commonly in a certain region of the country, but we can clearly see the extremities. In Stratification 2, we added race as a dimension to see whether there are any distinct differences between the prevalence of Alzhiemer's between patients of different races in different regions of the United States. We found slight differences. While there are few significant differences, we found the Hispanic population has a lower prevalence of Alzhiemer's disease compared to Black, non-Hispanic, Native Am/Alaskan Native, and White non-Hispanic in the state of Montana.

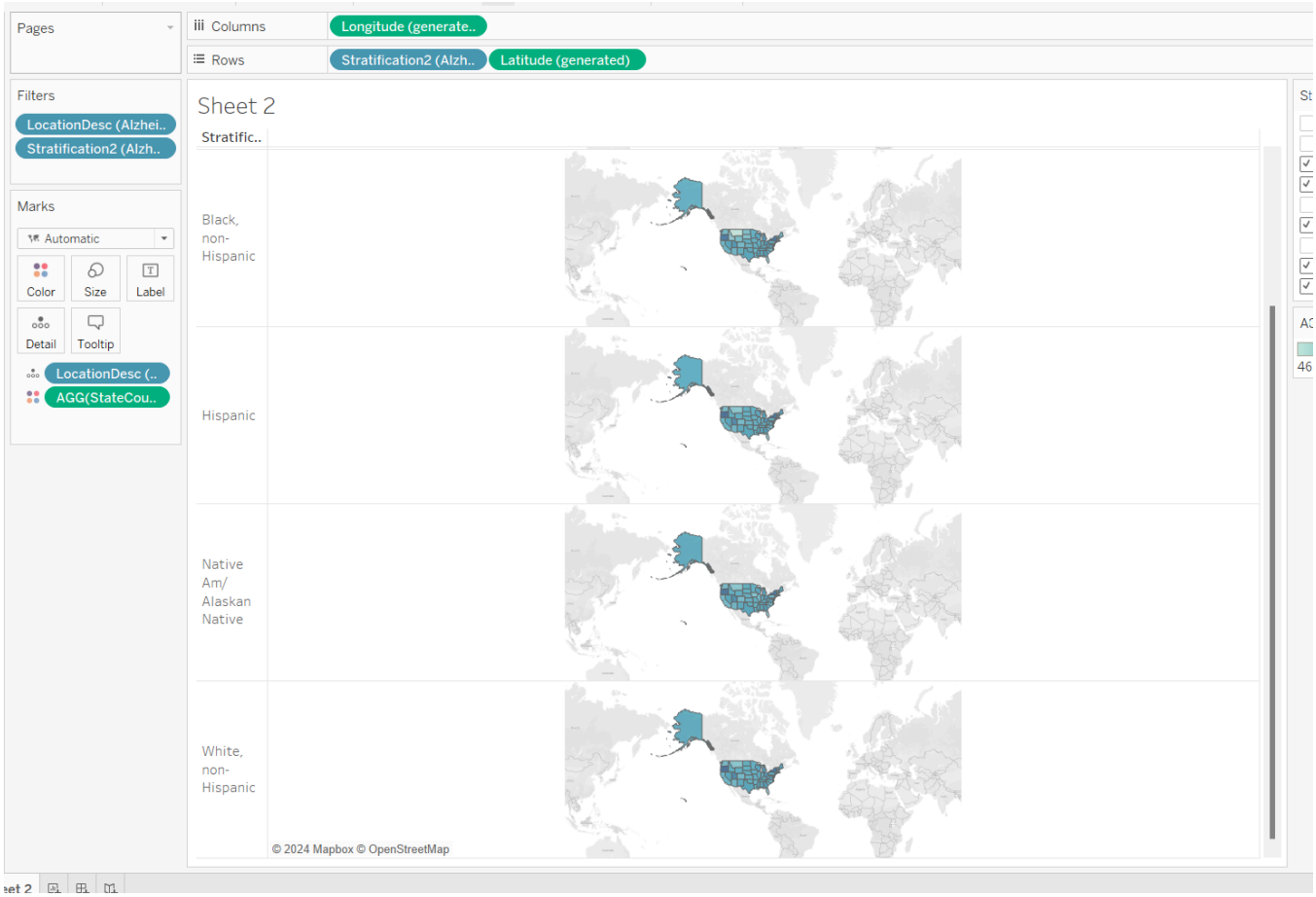


Tableau Packaged Workbook

Save or Export your project as a Tableau packaged workbook file and provide it as part of the github repository.