**Project 1: Written Analysis**

Chronic diseases pose a significant public health challenge in the United States, impacting millions of individuals and straining healthcare resources. The data analyzed in this study is sourced from the Center for Disease Control and Prevention (CDC), a trusted institution that conducts state-based telephone health surveys. These surveys collect information on health-risk behaviors, preventive health practices, and healthcare access, with a focus on chronic diseases and injuries. Our analysis began with a thorough exploration and cleaning of the dataset. After identifying and handling missing values, dropping redundant columns, and reindexing, we analyzed trends in chronic disease prevalence from 2015 to 2022.

According to the data, the key trends in chronic disease prevalence (2015-2022), showed that from 2015 to 2018, most chronic diseases exhibited a relatively stable prevalence, with little variation in crude rates. However, cancer, initially the most prevalent chronic disease in 2015, showed a consistent decline during this period. Starting from 2018, four out of six chronic diseases began to show a noticeable increase in prevalence, continuing until 2021. These diseases include alcohol-related diseases, diabetes, cardiovascular disease, and chronic obstructive pulmonary disease. Interestingly, from 2021 onwards, there was a decline in the prevalence of all chronic diseases analyzed, indicating a potential reversal of the earlier increasing trends.

This analysis was further broken down to effectively analyze the prevalence of chronic diseases per state. This data focused on the crude rate values and therefore required creating a new DataFrame . Analyzing this information, the best visual option was a bar graph, sorted in ascending order to facilitate comparison. The bar graph highlighted that chronic diseases are most prevalent in Oklahoma, with approximately 60,000 cases, while Utah exhibits the lowest prevalence, with around 25,000 cases.

To effectively analyze the prevalence of chronic diseases per state, we focused on the crude rate data values and therefore created a new DataFrame . Analyzing this information, we used the sum function and opted for a bar graph, sorted in ascending order to show the comparison.

The bar graph highlights that chronic diseases are most prevalent in Oklahoma, with approximately 60,000 cases, while Utah exhibits the lowest prevalence, with around 25,000 cases. The significant difference between Oklahoma and other states suggests that the population in Oklahoma might be more inclined to engage in unhealthy lifestyles and behaviors, such as smoking, poor dietary practices, physical inactivity, and excessive alcohol consumption.

Focusing on Oklahoma, the utilization of a pychart was used to illustrate the distribution of chronic diseases within the state. Drawing upon the data from the same DataFrame, we analyzed the occurrence of each chronic condition in this specific region. The top three chronic diseases in Oklahoma, were cardiovascular disease (accounting for 26.7% of cases), chronic obstructive pulmonary disease (accounting for 22.7% of cases), and cancer (accounting for 21.2% of cases). This observation aligns with the findings from question two, confirming that cardiovascular disease is the most prevalent chronic disease.

The crude rate helped us to depict the most prevalent chronic disease in the United States, cardiovascular disease. Over a 3-year study, between 2019-2021, cardiovascular disease was the highest-rated chronic disease, and had a steady increase over the years. No decline or incline was shown, as there was no data represented before 2019 or after 2021. Filtering through different race/ethnicities, allowed us to see the overall effect that chronic cardiovascular diseases had on different demographics. Through this, we noted that the least at-risk were Asian or Pacific Islanders, consisting of 3.8% of the overall study. On the other hand, Black, Hispanic, White, and American Indian or Alaska Native, each contributing 15.4% respectively, were noted as being the highest at-risk.

Through the analysis of the data, the factor of age group became apparent. Through the assumption that the older population was at more of a high-risk, we conducted an analysis of the population of 65 years-of-age or older, against hospitalization rates. Through this study, it showed that hospitalization rates of those aged 65-years or older, that had a principal diagnosis of heart failure, was the highest amongst those living in Wisconsin. At the same time, it can be seen that Hawaii has one of the lowest hospitalization rates for chronic cardiovascular diseases.

When talking about the highest at risk, in relation to race/ethnicity, it is also important to discuss the prevalence of chronic diseases between genders. At the beginning of this section, it was important to get the overall count of males and females in relation to chronic diseases as a whole. After obtaining the count, the data was shown visually through a pie graph, showing females had a slightly higher percentage of having a chronic disease. We then created a data frame that included only the necessary data needed, this being gender, location, and type of chronic disease. Through previous data, it was noted that Utah had the least prevalent rate of chronic diseases, and Oklahoma had the most. With this data known, the next part of the analysis, was to break down all three portions of the data frame, and relate them to each other, that being gender, chronic disease type, and state. This was done for both Oklahoma and Utah respectively, by separating it by gender, male and female, and all six chronic disease types. Once all of the data was collected, two stacked bar graphs were created, separated by chronic disease type, and stacked by gender. One graph represented the data for Oklahoma, and the other for Utah. Both of these graphs showed that cardiovascular diseases were the most prevalent between both Utah and Oklahoma, as well as between both male and female.

During the analysis, it was discovered that the data was nearly identical to each other. To ensure that it was not an error in the code, but how the data was originally brought in, two CSV files were created from both data frames to verify if the data was correct. Upon analysis, the data frames were correct and it was due to how the data was brought in. All data represented in the data frame matched all other visualizations made, which ensured all data was correct and it was just how the data was originally cleaned, before inputted into this project.

In conclusion, chronic diseases present a significant public-health challenge affecting individuals of every demographic. Through the analysis it is evident that a large portion of the population face a heightened risk of developing a chronic cardiovascular disease. Factors such as race, gender, or geographical location show very low influence on the percentage that someone might be susceptible to such diseases. However, with the advancement of medical technology and innovation, the possibility of a decreased potential of developing a chronic cardiovascular disease can occur. This same downward trend is show for all chronic diseases discussed in the analysis. It could be inferred that this downward trend, is due to medical advancements made over the past few years, especially with new medical technology being created or advanced due to the COVID-19 pandemic. As medical research progresses, the possibility of the downward trend could still occur, however this can only be speculative in present-day.