Ridhima Prashanth
Mary Lemmer
STAT 331
Professor Williams

Report on County Data Visualization Application

Purpose and Importance

The County Data Visualization Application provides a comprehensive tool for analyzing and comparing demographic data from counties. By focusing on key variables such as race, gender, age, household size, and income sourced from the American Community Survey (ACS), specifically the 2020 survey conducted by the Census Bureau, this application offers valuable insights into the social and economic landscape of various regions.

For students seeking internships or planning to relocate, this tool serves as an invaluable resource. It enables them to gain a better understanding of the socioeconomic climate of different areas, helping them make informed decisions about potential living arrangements. By exploring the demographic trends and patterns within counties, users can assess whether a particular location aligns with their preferences and needs. This includes considerations such as diversity, economic opportunities, and overall quality of life, ultimately empowering individuals to make well-informed choices about their future endeavors and living arrangements.

Brief Description of the code

The foundation of the application lies in harnessing data from the American Community Survey (ACS) census conducted by the Census Bureau every five years, with a focus on the 2019 survey. The Census Bureau API serves as the primary conduit for accessing this rich demographic dataset. Using the Census Bureau API, the application pulls data from the ACS census, capturing a comprehensive snapshot of demographic indicators from the year 2020. This includes variables pertaining to race, gender, sex, age, household size, and income.

Utilizing a dynamic Shiny frontend allows the user to seamlessly dictate what categories they would like to look at, including specifying specific sub-categories, and choose to evaluate one to three counties at a time. The graphs produced are clear and easy to read including human-legible labels and specific use of color to visually indicate the differences between two counties.

After a location is selected, any change will cause the graphs to update. User choices are parsed by the server that first collects the data from the Census Bureau API for each location desired. Upon obtaining the raw data for each county, the next step involves refining the dataset to extract the pertinent variables for analysis. Functions like filter and mutate are employed to sift through the data and isolate the variables of interest. Variables such as age are

further processed to derive aggregate values. This is because this variable is not split directly on age but on two variables of sex and age.

To facilitate cross-county comparisons, the code utilizes join functions, alongside reduce and anonymous functions, to merge the datasets from different variables of the same county into a cohesive dataframe. This amalgamation enables seamless analysis and visualization of demographic trends across various urban centers. Finally, a crucial aspect of data preparation involves cleaning the dataframe to enhance its usability and clarity. This entails eliminating duplicate columns and standardizing column names for consistency and comprehensibility. Additionally, manual adjustments are made to rename variables and rectify data codes using functions like mutate and stringr. For instance, variables such as B02001_002 are matched with their more descriptive labels, ensuring clarity in the interpretation of demographic attributes.

Role of ChatGPT in Project Development
ChatGPT serves as a valuable resource throughout the project development process. It was utilized to provide aid in troubleshooting code issues, quickly performing repetitive tasks like creating the list of states, give insights on how to accomplish more complicated tasks like wrapping labels or creating custom CSS, and suggesting alternative approaches to lengthy code in an effort to make it more concise. In particular, ChatGPT assists in understanding and utilizing functions that may otherwise be difficult to understand like lapply, enhancing the developer's understanding and proficiency in R programming.

Summary
In summary, the County Data Visualization Application empowers users to explore and compare demographic trends across major cities, contributing to informed decision-making and deeper understanding of societal dynamics. Through efficient data processing and visualization techniques, it provides a user-friendly platform for demographic analysis and interpretation.