Project Report

December 5, 2022

Guiding Questions

There are two questions that this project intended to find answers to:

1. What are the factors that impact the quality of life?
2. Could the quality of life factors vary from area to area?

As I was exploring the datasets that relate to this project, other questions, like what impact does income have on quality of life, emerged. The data gathered during this project can answer all these questions.

Datasets

This project required the use of four datasets. The primary dataset is the Behavioral Risk Factor Dataset. The information gleaned from this dataset is used primarily to find how physical health, mental health, general health, activity limitation, race, gender, and age impact quality of life. This dataset was created by the CDC from responses of individuals to a survey that asked them the number of days they felt mentally and/or physically unwell or limited in their activities. An additional dataset from the Bureau of Economic Intelligence (BEA) detailed the income per capita and personal income for each state from 1929 to 2010. The U.S. Population Zip Code dataset was imported from Kaggle and it contains information about populations for each zip code in America for 2000 and 2010. The Race/Ethnicity dataset was found on DC Data Journalism, a site for data scientists interested in journalism. Both the population and race datasets were used to see how representative the data in the Behavior Risk Factor Dataset is.

All four datasets are cited on the last page of this report.

Discovery

The data used in this project focuses on the following as factors that impact the quality of life: mental health, physical health, general health, activity limitation, race, age, gender, and income. Exploration and analysis of the data reveals that these factors do vary from state to state.

For every year from 1993 to 2010, for every question relating to health, five states had the worst average health values (values aggregated from data in Behavioral Risk Factors dataset). Kentucky, West Virginia, Mississippi, Alabama, and Arkansas consistently had the highest average health values. People in these states had higher numbers of days where they felt mentally unwell, physically unwell, limited in their ability to participate in activities, or generally unwell. Certain states like Hawaii, Iowa, and the District of Columbia had lower average health values than most states, indicating that these residents in these regions have less days of feeling unwell or limited.

Race, age, and gender influence quality of life as well. Native Americans and Alaskan Natives consistently had the most days feeling unwell physically and/or mentally or experiencing activity limitation. Females have higher health average than men. Older people report more days feeling physically unwell, while younger people report more days feeling mentally unwell.

Income has a significant effect on the quality of life per state. Using the BEA income set, I conducted a correlation analysis between income per capita for each state and the average percentage of people with poor self-rated health. The analysis yielded a correlation coefficient of -0.6. This correlation is strong enough to infer that a state with lower incomes per capita have greater number of residents who feel generally unwell.

	Average of IncomePerCapita	*Mean mentally unhealthy days*	*Mean physically unhealthy days*	*Mean physically or mentally unhealthy days*	*Percentage with fair or poor self-rated health*
Average of IncomePerCapita	1				
Mean mentally unhealthy days	-0.470556781	1			
Mean physically unhealthy days	-0.59580225	0.867758472	1		
Mean physically or mentally unhealthy days	-0.545305793	0.955462689	0.960295799	1	
Percentage with fair or poor self-rated health	-0.613399792	0.75896665	0.895924368	0.839072813	1

Visualizations

The visualizations included in Power BI report serve to summarize the answers to the guiding questions concisely and neatly. The first page of visualizations shows the average health values per year and per state. Certain states-aforementioned- and certain years, specifically 2002, have larger average health values than most years. In general, recent years have higher data values. This could be because of an aging population or worsening economic trends. This indicates that time is a factor that influences quality of life as well, but there is no data to explore these theories. The second page has also already been discussed in Discovery. The third page shows correlations between the different topics and average health values. The strongest correlation is between activity limitation and feeling physically unwell, which was expected.

The fourth page includes a scatter plot that compares income per capita to average health value for each state. It illustrates the correlation matrix between income per capita and self-rated health

nicely. The fifth page of visualizations was an attempt at seeing how representative the sample sizes in the Behavioral Risk Factor survey were. The overall sample size is more representative of the total population in 2010 than 2000. White, non-Hispanic individuals are the most represented in each state, while Native American, Alaskan Natives, Asians, and Pacific Islanders are the least represented in the survey with the exceptions of Hawaii and Alaska.

Decisions

A few decisions can be made from this data. First, the CDC should attempt to increase their sample sizes for non-white racial and ethnic groups. Non-white racial groups are staggeringly underrepresented in nearly every state. More representative sample sizes will lead to more reliable data. While I understand that the CDC cannot force participants of certain races to respond to voluntary survey, a greater effort to generate responses from these groups should be made.

The five states with the worst health values consistently should invest more into health care and mental health services. All states should invest more into caring for the physical and mental health of its residents during economic downturns due to the negative relationship between income and health. More resources need to be dedicated to improving the quality of life of minority groups, especially Native Americans and Alaskan Natives.

These are all general conclusions and decisions that can be made at this point. To make more informed and specific decisions, more data and greater analysis is required.

Works Cited

Bureau of Economic Analysis. (2011, September). State Personal Income: Revised estimates for 2010, Version 2. Retrieved December 1, 2022 from https://apps.bea.gov/regional/histdata/releases/0911spi/index.cfm

Centers for Disease Control. (2016, January). Behavioral Risk Factor Data: Health-Related Quality of Life (HRQQL), Version 1. Retrieved November 28, 2022 from https://data.world/cdc/behavioral-risk-factor-hrqol

DC Data Journalism. (2017, January). US Health and Demographic Data: Race_ethnicity.csv, Version 1. Retrieved December 1, 2022 from https://data.world/dc-data-journalism/urban-rural-health-and-demographic-data/workspace/file?filename=Race_ethnicity.csv

Michael Valcic. (2017, January). US Population By Zip Code: Add city, state, longitude, and latitude data, Version 1. Retrieved November 28, 2022 from https://www.kaggle.com/code/mvalcic/add-city-state-longitude-and-latitude-data/script?scriptVersionId=2190548