Pattern of Heart Disease Mortality Rate Within Gender and Race Stratification in California During 2014

Caroline He

12/7/2021

Introduction

Heart disease is the one of the leading cause of death for men, women, and people of different races in the United States. The term "heart disease" includes several types of heart conditions. In the United States, the most common type of heart disease is coronary artery disease (CAD), which can lead to heart attack. In fact, one person dies every 36 seconds in the United States from cardiovascular disease.About 659,000 people in the United States die from heart disease each year, which is in every 4 deaths. Heart disease costs the United States a lot each year, which includes the cost of health care services, medicines, and lost productivity due to death. It would be helpful to examine the pattern of heart disease mortality rate under gender and race stratification to understand the behind association to some extent.

Data Background

The data was obtained from CDC chronic disease and health promotion data & indicators: https://chronicdata.cdc.gov/Heart-Disease-Stroke-Prevention/Heart-Disease-Mortality-Data-Among-US-Adults-35-by/i2vk-mgdh, and it was organized by National Vital Statistics System and focused on heart disease mortality data in US during 2014. The data was collected in county level. Here were the basic information of this dataset:

- 2013 to 2015, 3-year average. Rates are age-standardized. County rates are spatially smoothed. The data can be viewed by gender and race/ethnicity. Data source: National Vital Statistics System. Additional data, maps, and methodology can be viewed on the Interactive Atlas of Heart Disease and Stroke http://www.cdc.gov/dhdsp/maps/atlas

Since it might be too vague to observe the mortality rate from the entire country level, a focus on California data is included in this investigation. The FIPS code of each county was obtained from https://raw.githubusercontent.com/kjhealy/fips-codes/master/state_and_county_fips_master.csv.
The main question is: How were gender and races associated with heart disease death rate in California during 2014?
To better analyze the question step-by-step, several sub-questions are added

- What was the distribution of heart disease death rate under gender and race stratification?
- Which county had relatively higher mortality rate gap between gender groups?
- How was the mortality rate distributed in each county under gender and race classification?

Methods

The data set included the following important variables which will be used later:

- LocationDesc: county name
- Data_Value: heart disease death rate per 100,000 population
- Stratification1: gender categories (Male, Female)
- Stratification2: race categories (White, Black Hispanic, Asian and Pacific Islander, American Indian and Alaskan Native)
- FIPS code: used for locating county

Data Wrangling

The data set was cleaned according to EDA checklists. There were several "NAs" under "Data_value" variable, which was due to insufficient information. It was replaced by 0, which was used for later data visualization. Since the data set included data from entire country, only data in California was selected for analysis. After data wrangling, datasets `CA_race` and `CA_gender` were created and each of them was merged with FIPS code, whose source was introduced in the Introduction session.

The gender data included variables: county name as `LocationDesc`, male mortality rate as `value_male`, female mortality rate as `value_female` and FIPS code for each county as `fips`.

The race data included variables: county name as `LocationDesc`, the White mortality rate as `value_white`, the Hispanic mortality rate as `value_hispanic`, the Black mortality rate as `value_black`, the Asian and Pacific Islander mortality rate as `value_asian_pacific`, the American Indian and Alaskan Native mortality rate as `value_indian_alaskan` and FIPS code for each county as `fips`.

Preliminary Results