Data Curation Project

Samantha Berg
Hernan Razo
Christopher Anthony

Dataset Introduction

Datasets used:

- Health insurance Coverage of the Total Population
  - Provided by KFF, a third party organization focused on health policy analysis and health journalism
  - Gives number of people in the US that are insured or uninsured
  - Breaks down coverages by specific health plan
- Coverage Numbers and Rates by Type of Health Insurance
  - Provided by the United States Census Bureau
  - Gives number of people in the US that are insured or uninsured
  - Breaks down coverages by specific health plan
  - Also breaks down data by state

Domain

- Both datasets fall into the health insurance domain
- They both try to quantify the source of health insurance for the general US population

<u>Methods of Acquirement</u>

- The data from KFF is openly available on their website
- The data from the US Census Bureau is also open to the public on their website

Data Characteristics

- KFF dataset
  - composed of 58 rows and 8 columns
  - Has 8 attributes: Location (state) and various types of insurances (medicaid, medicare, etc.)
- US Census Bureau dataset
  - Composed of 24 rows and 10 columns
  - Has 10 attributes: various types of insurances (medicaid, medicare, etc.)
  - Came in pdf format

Data Quality Problems/Challenges

- The dataset from the US Census Bureau came in pdf format. This means that we had to manually input data values into csv format
- We also had to transform each dataset to only include the common insurance plans between both datasets (Employer, Medicaid, Medicare, Military, uninsured)
- It was hard to tell how each organization counted if someone was insured or not. Discrepancies came up if someone was insured for only portions of the year, what was considered to be uninsured, and discrepancies between ASEC vs. ACS reportings.

Solutions

- We used the Pandas python library to extract all necessary numbers
- We also used python to add all necessary values to get the total number of people that are insured by specific plans.

Findings

- The government reports that more people are insured through their employer and more people are insured through Medicare
  - 23,911,700 more are insured through their employer
  - 13045800 more are insured through Medicare
- KFF reports that more people are insured through the Medicaid or the military. Also, KFF reports more uninsured people
  - 8,751,000 more people insured through Medicaid
  - 1,188,600 more people insured through the military
  - 3,486,100 more people are uninsured

KFF_totals

	Employer	medicaid	medicare	military	uninsured
0	159093300	64602000	45733200	4409600	29597100

gov

	Employer	Medicaid	medicare	Military	Uninsured
0	183005000	55851000	58779000	3221000	26111000