# 1.4 – SOURCING THE RIGHT DATA

## POPULATION BY GEOGRAPHY US CENSUS

### DATA SOURCE:

This is an external data source. The data is collected by the US Census Bureau. This data is owned by the government and is considered trustworthy.

### DATA COLLECTION:

This is administrative data collected by the US Census Bureau. The census is conducted every 10 years, but the Census Bureau publishes population estimates and demographic components of change each year. The data is collected manually and then automatically inputted into the database. There is a time lag of 10 years, but estimates are provided every year by the Census Bureaus in the form of surveys.

### DATA CONTENTS:

The data contains population count for each county in the US from 2009-2017. Variables included are county, year, and total population broken down by age ranges and gender.

### LIMITATIONS:

- The census is conducted every 10 years and although mandatory, it there have been no known penalties given to those who do not participate. Citizens are required to input their information electronically or through mail, which can be prone to human error or families not given the correct count.
- Since the census is conducted decennially, the Census Bureau conducts additional surveys each year to provide estimates. This data could be biased, from the fact that not all surveys are filled out and/or that surveys are given to certain region.

### RELEVANCY:

The data shows the total population count by year from 2009-2017 and is broken down by gender and age range. This data can be compared against the Influenza Death data to provide an estimate on the population to death ratio for each state. As with the Influenza Death data, historical trends can mirror future trends. A project objective is to determine where to send staff, this data set could provide an estimate on the total population of each state and allow for planning additional staff to states with higher populations.

# INFLUENZA LABORATORY TESTS AND PATIENT VISITS

## DATA SOURCE:

This is an external data source and is collected by the CDC. This data is owned by the government and is considered trustworthy.

## DATA COLLECTION:

This is survey data collected as part of a collaborative effort between CDC and its many partners in state, local, and territorial health departments, public health and clinical laboratories, vital statistics offices, health care providers, hospital clinics, emergency departments, and long-term care facilities. Data is updated by medical staff and updated automatically in the database. There is no time lag, as the data is updated once the patient is completed with their appointment.

## DATA CONTENTS:

**Influenza Laboratory Test:**

This data set shows the weekly count of lab test results from 2010-2015 for each state. Variables included are region type, region, year, week, total specimens, percent positive, and the virus type.

**Patient Visits:**

This data set shows the weekly count of influenza-like illnesses visits to a medical provider from 2010-2015 for each state. Variables included are region type, region, year, week, and age ranges.

## LIMITATIONS:

- All influenza activity reporting by public health partners and health care providers is voluntary. In addition, these two data sets are survey data, therefore they're not complete counts of all influenza visits for laboratory tests.
- The need for approval before the data is made available or not.
- Some state data is missing, which can lead to bias due to human error.
- Variables "%weighted" and "%unweighted" needs more clarification as to what they mean.
- The data could be biased depending on who is collecting it as it is manually collected, due to the polarized nature of America in terms of race and discrimination.

## RELEVANCY:

This data provides the necessary information needed to provide analysis as to where to send staff, which is an objective of the project. The data can be analyzed as historical trends to determine flu counts and patient visits for each state. If there is a trend in one state having higher counts of flu-related visits or positive flu test, the staffing firm can plan accordingly.

## CHILDREN FLU SHOTS

### DATA SOURCE:

This is an external data source and data is provided by the National Immunization Surveys (NIS). The University of Chicago runs the surveys on behalf of the CDC. This data is owned by the government and is considered trustworthy.

### DATA COLLECTION:

This is survey data collected through telephone interviews with parents in all U.S. states and territories. The data comes from a random sampling of parents. The demographics are self-reported, but the flu shot information is verified with health providers.

### DATA CONTENTS:

The data contains flu shot data for children 6 months to 17 years. It's categorized by geographic state and contains family demographics including poverty level, race, and parent marital status.

### LIMITATIONS:

- As the data is collected manually through interviews, it is prone to human errors.
- The random sample could be biased depending on how it is determined.

### RELEVANCY:

As the data focuses on the younger population, which is one of the vulnerable populations, it could prove to be useful in the analysis. But because the data is a random sample, it may not provide an accurate reflection of the population as a whole and therefore make it harder to determine where the agency should send additional staff. In addition, the data set provides multiple variables that is not relevant to the project and therefore will not be used for the analysis.