**Exercise 1.4 Sourcing the Right Data**

**US Census Population Data by Geography**

* **Data Source:** This medical staff agency getting data from US Census Bureau and it provided from united state government so it is trustworthy source. This is External source of data.
* **Data Collection:** This data collected by Central Bureau, this is Administrative Data. These data collected through multiple ways like manual, data entered in system, email so there is lag in time. Also this data updated annually.
* **Data Relevance:** The data is relevant because from this data contains dataset for our wish list, we can tell which state has more states with more vulnerable populations. These are the population that is most vulnerable to the flu, which tells us staffing needs during the influenza season.
* **Data Limitation:** As Data collected in time lag of 1 year and have prone to human errors and this is official data provided by government so there is no bias in data.
* **Data Content:** The data is about the total population on a yearly basis from 2009 to 2017. This has information’s about county, year, population, age and gender. It can be say as “Population data by geography”.

**CDC Influenza Laboratory Test and Patients Visit**

* **Data Source:** This data is from CDC which collects, compiles, and analyzes information on influenza activity weekly in the United States. This is external source of data. This source includes both Public health and clinical Laboratories data.
* **Data Collection:** The data collection method involves both manual and automated reporting by the participating providers and laboratories. Given the weekly and annual span of data reporting, there is an inherent time lag that reflects the period between the occurrence of influenza cases and their documentation in the datasets. This is partially administrative data.
* **Data Relevance:** This Data is relevant as this provide data for wish list of relationship between age and influenza. Also this shows timing of flu activity seasons.
* **Data Limitation:** This data’s collected from both Public and Clinical Laboratories, these data’s entered manually so leads to inaccuracy, duplicate and bias in data. These data’s collected over weekly so there time lag occurs.
* **Data Content:** This data has information of both public and clinical Laboratory results of influenza illness, hospitalization, deaths, Vaccination status, age group and patient’s visits from 2010 to early 2019 from all states.

**Children Flu Shots by CDC**

* **Data Source:**  The data source is external because the University of Chicago is collecting the data and sending it to the CDC. These records is accurate and trustworthy because it is verified by healthcare professionals.
* **Data Collection:** The NIS provides current, population-based, state and local area estimates of vaccination coverage among children and teens using a standard survey methodology. The surveys collect data through telephone interviews with parents or guardians in all 50 states, the District of Columbia, and some U.S. territories.
* **Data Relevance:** For one of my hypothesis if people injected with flu shots, so there will be less death. This Data has information of children under 18, this part of the data used to find out if these vaccinations had a helping hand in the morality rate of children.
* **Data Limitation:** These data are collected from phone and mail survey doing manually which leads to Non-Response bias, limited reach though phone and inaccurate due to poor memory or miscommunication.
* **Data Content:** This data has variables includes family demographics, location and most importantly the vaccination status hastypes of vaccinations, number of doses, dates of administration, and other administrative data about the health care facility of children’s and teens in all 50 states.