[Provisional COVID-19 Deaths by Sex and Age - Catalog (data.gov)](https://catalog.data.gov/dataset/provisional-covid-19-death-counts-by-sex-age-and-state)

# Data Source

**#3-5**

The data is external, its owned by the CDC, so it’s probably trustworthy because it’s a government owned organization. The website says the deaths counts are based on the flow of mortality data in the National Vital Statistics System. It also says it can take several weeks for death records submitted to National Center for Health Statistics to be processed, coded, and tabulated. So, the data can be missing a few figures.

The data set shows a start date, end date, is grouped by Month, Year, or total; State, Sex, Age group, Covid-19 Deaths, Total Deaths, Pneumonia Deaths, Pneumonia and Covid-19 Deaths, Influenza Deaths, and Pneumonia, Influenza, or COVID-19 Deaths.

I chose this data set because it looks well organized and easy to use compared to other data sets I viewed. It fits the criteria of the project in terms of row count and location data. It seems like I can find questions to ask and answers for them in this data.

# Data Profile

**#6-8**

As far as ethical limitations it seems to be clean. As I said in the data source section, it’s sourced from the CDC so it is likely trustworthy. It seems like it was automatically collected by a government system and the flow of counts may be delayed by some weeks due to the collection method. There is no PIP data in this so I don’t need to worry about that.

**10.)**

Are some states more effected than others?

Which disease is most deadly?

Which age groups are most vulnerable?

Is there a particular time of year that is most deadly?

Is there a preference for gender in terms of vulnerability?