**Data Set**

|  |  |
| --- | --- |
| Source | Using the Centers for Disease Control and Prevention (CDC) Wonder, a rich ad-hoc query system for analysis of public health data.  [Underlying Cause of Death, 1999-2020 Results (cdc.gov)](https://wonder.cdc.gov/controller/datarequest/D76;jsessionid=4FF4898D04AA66A1DA28BC8B9CED) |
| Collection | The data is administrative data collected as part the National Vital Statistics Cooperative Program. Death Records come from death certificates in which doctor codes the primary cause of death as ICD 100-199 Diseases of the Circulatory System. |
| Contents | The data set contains deaths coded as ICD 100-199 Diseases of the Circulatory System from 2015-2020. The data contains columns of state, race, ten-year age groups, year, gender, deaths, and population. |
| Limitations | The data is from government data which makes it reliable.  The data is from 2015-2020, although not the most recent data, is still relevant.  The biggest flaw of this data is that it does not further break down the ICD code listed on the death certificate. Hence, we are unable to determine how many people are listed to have died from specific ICD codes. |
| Ethics | The data set does not contain personal information that breaks Health Insurance Portability and Accountability Act (HIPAA) rules. |
| Relevancy | This data set is relevant to the projects as it shows me the data for deaths listed as ICD 100-199 and broken down into state, age, race and gender. |

**Data Profile**

* This data contains 7 columns, 9490 rows.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Time Variant/ Invariant** | **Structured/Unstructured** | **Qualitative/Quantitative** | **Further Characteristics** |
| State | Invariant | Structured | Qualitative | Nominal |
| Race | Invariant | Structured | Qualitative | Nominal |
| Ten-Year Age Groups | Invariant | Structured | Qualitative | Ordinal |
| Year | Variant | Structured | Quantitative | Discrete |
| Death | Variant | Structured | Quantitative | Discrete |
| Population | Variant | Structured | Quantitative | Discrete |

**Data Consistency Checks**

* Refer to Jupyter notebook.

**Questions**

1. Are males more likely to die from diseases of the circulatory system than females?
2. Are older people more likely to die from diseases of the circulatory system than younger people?
3. Are certain races more prone to die from diseases of the circulatory system compared to others?
4. Does the geographic location increase the risk of death from diseases of the circulatory system?
5. Are more people dying from diseases of the circulatory system?