Data Miners

Zahra Hatami
Sowjanya Joga
Praseeda Sasanka Pisipati

Agenda

- About Data
- Research Questions
- Data Relevance to Target Audience
- Analysis & Interpretations
- Derived Decisions

About data – Excess deaths and causes

- This data set helps us in evaluating the cause of deaths in various states and increase the health consciousness among the people.
- Represents the following number of deaths in different states in the United States
    - (1) Expected deaths
    - (2) Observed deaths
    - (3) Potentially Excess deaths
- Data collected based on five leading causes of death in metropolitan and non-metropolitan areas
    - (1) Heart disease
    - (2) Cancer
    - (3) Unintentional injury
    - (4) Chronic lower respiratory disease
    - (5) Stroke

About data – Excess deaths and causes

- **Target Audience** for this data set are:

  Doctors

  Health Researchers' Groups

  Medicine Companies

- **Limitations**

  Only has the data for age ranges less than 85

  Only deals with five-leading causes of death

Meta Data

- **State FIPS Code** were numeric and two-letter alphabetic codes defined in U.S. Federal Information Processing Standard Publication ("FIPS PUB")
- **Mortality** data for U.S. residents come from the National Vital Statistics System
- **Cause of death** is based on the International Classification of Diseases, 10th Revision (ICD-10)
- **Locality** (nonmetropolitan vs. metropolitan) is based on the Office of Management and Budget's 2013 county-based classification scheme.
- **Benchmarks** are based on the three states with the lowest age and cause-specific mortality rates.
- **HHS Region** is the number of the region allocated by the Office of Intergovernmental and External Affairs. It hosts ten Regional Offices that directly serve state and local organizations. A President-appointed Regional Director leads each office.
- **Potentially excess deaths** for each state are calculated by subtracting deaths at the benchmark rates (expected deaths) from observed deaths.

Research Questions

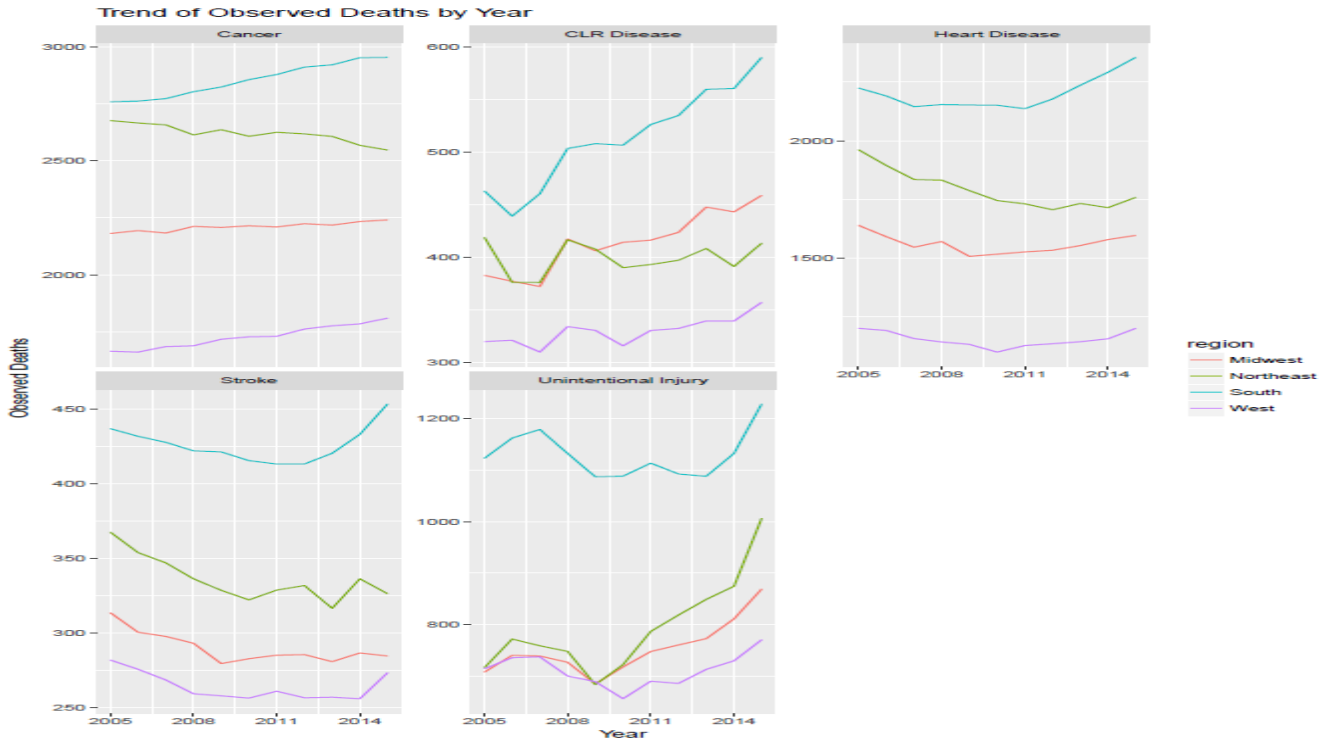
1. What is the trend of observed deaths for all five-leading cause of deaths over time?
2. What is the trend of expected deaths, observed deaths, potentially excess deaths in each age group?
3. Compare the number of deaths for various regions of US. Which region has the maximum and minimum number of deaths recorded for the years 2005 to 2015?
4. Compare the number of deaths for each locality. Which locality has the maximum and minimum number of deaths recorded for the years 2005 to 2015?
5. What is the ratio between the observed deaths and the Population? What is the trend of the ratio over time by region?
6. What is the ratio between the expected deaths and the Population? What is the trend of the ratio over time by region?
7. Are the ratios in questions 5 and 6 correlated?

Data Relevance To Target Audience

- Identify the deaths pattern in metropolitan or non-metropolitan areas
- Identify in which age group are the most deaths observed
- Identify the leading cause for death in each age group
- Increase the health consciousness among public
- Motivate public to practice a healthy lifestyle
- Doctors and Health Research Groups can educate public through the community health programs
- Medicine Companies can plan an effective strategy for sales of medicines

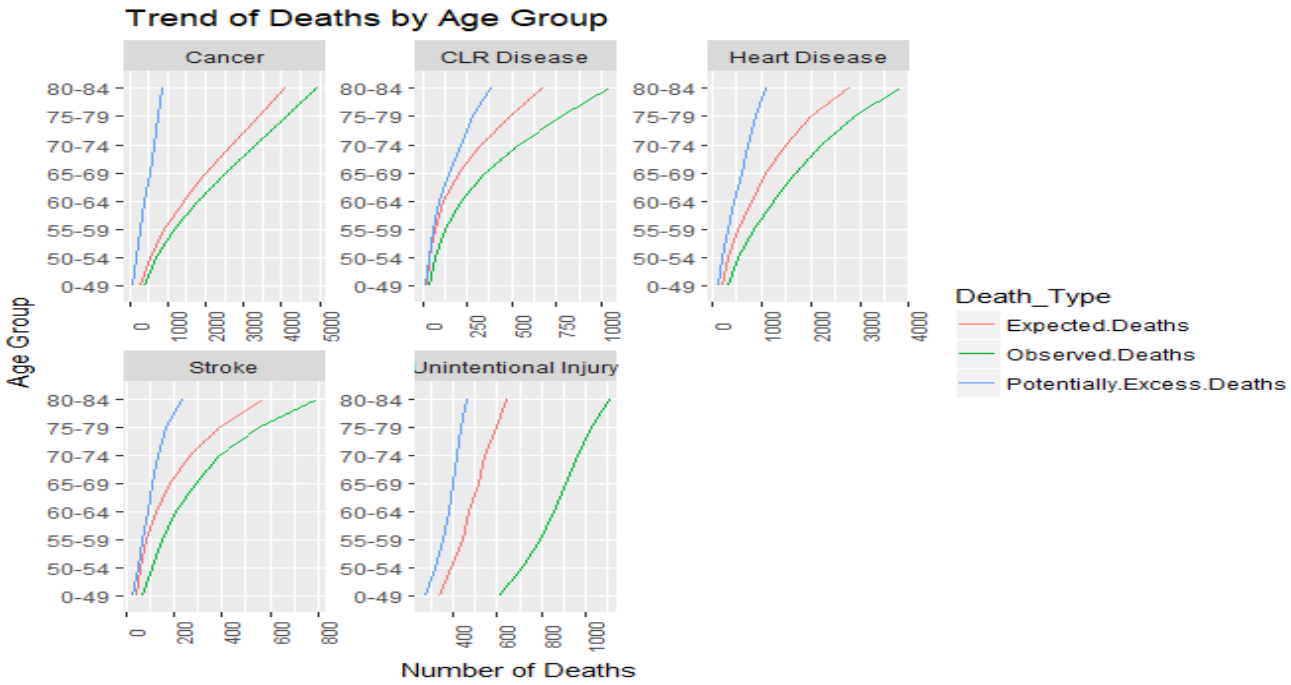
Analysis & Interpretation

Cancer has the highest number of observed deaths as per below trend. It is the only cause of death which didn't show any deviation in its trend. All the other causes recorded a decrease and an increase. And the Southern region has the highest number of deaths recorded in cancer category



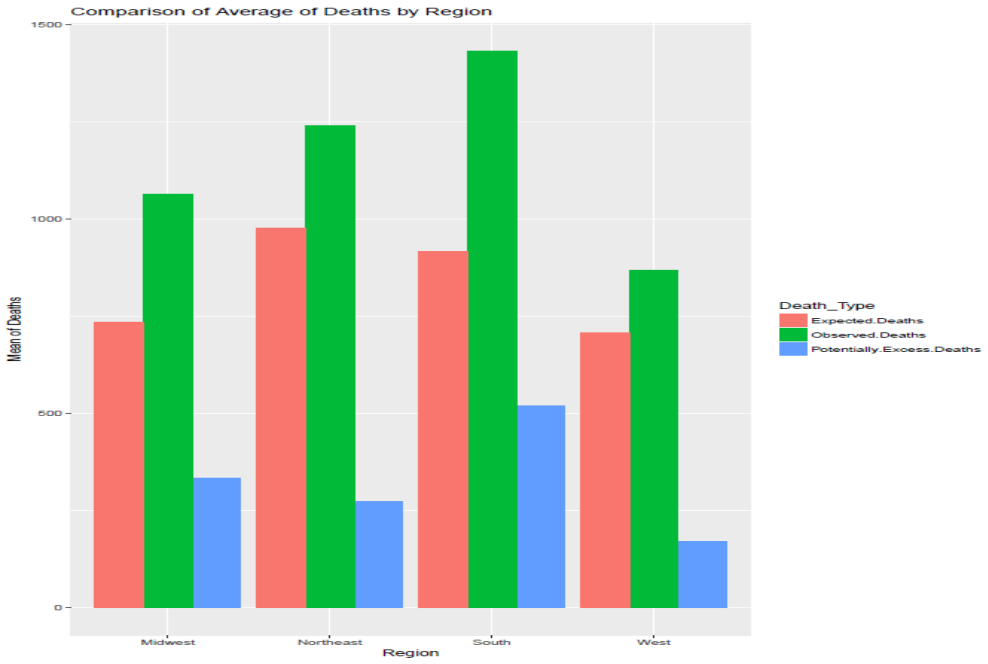
Trend of Deaths by Age Group

The number of observed deaths always exceeded the expected deaths. And also the age and number of deaths are proportional to each other i.e as the age increases, so did the number of deaths.



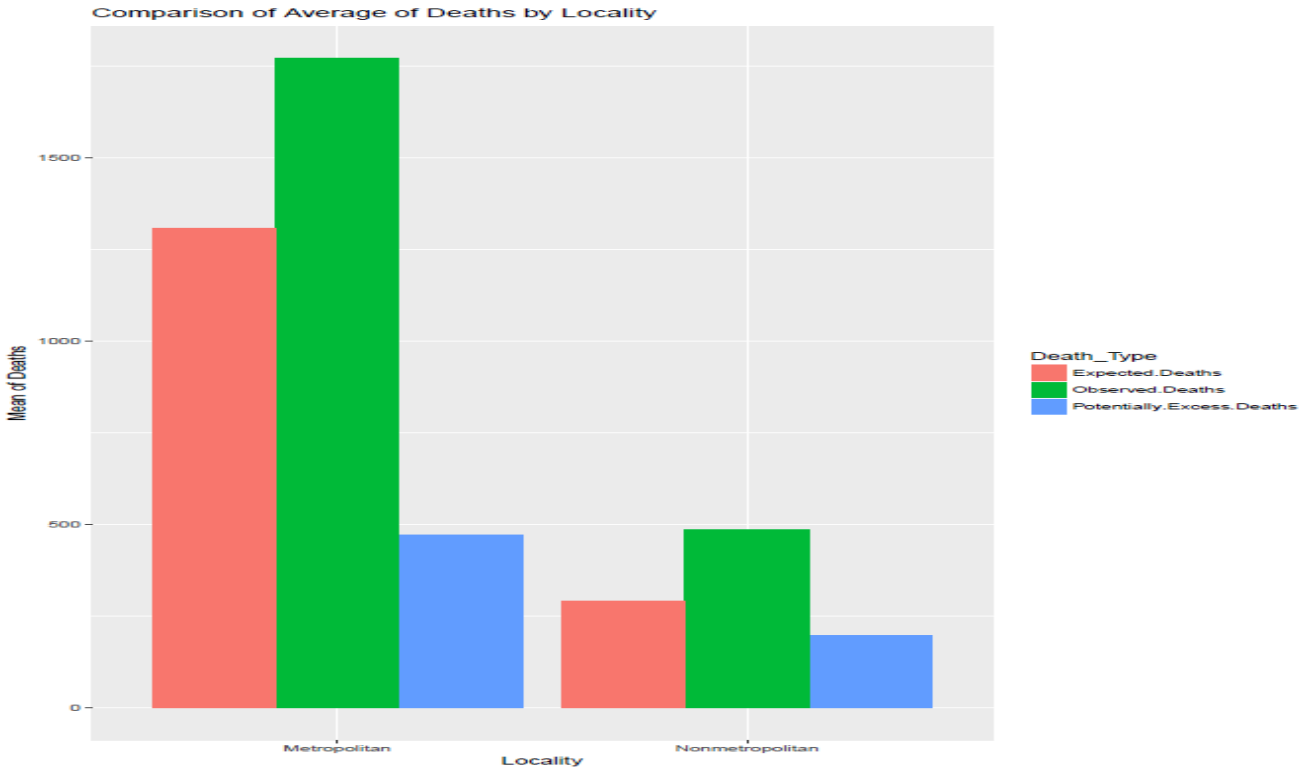
Comparison of Average of Deaths by Region

Observed deaths are more in the southern part of the United States followed by Northeast, Midwest and West. The Expected Deaths average is almost equal in both South and Northeast. South region also has the maximum average for potentially excess deaths followed by Midwest, Northeast and West



Comparison of Average of Deaths by Locality

Considering the population in metropolitan area, it is obvious that the number of deaths recorded in metropolitan would be more than non-metropolitan. But from the data we derived that most number of deaths are recorded in non-metropolitan than the metropolitan area when considering the regions and age group separately for each locality



Derived Decisions

- The increase in number of deaths due to cancer can be due to more tobacco use in non-metropolitan areas. Steps and necessary programs need to be conducted to reduce the tobacco use
- This increase can also be due to less physical activity or due to less travel options for health care and checkups. Doctors and Government need to consider options to improve physical activity and also increase the availability of health care services to public.
- Conduct various research programs to cure the diseases in the initial stages as and when detected
- The uneven trends in all the four-leading cause of deaths except cancer should be carefully examined. If the decrease in death rate is due to any events or programs during that year, a plan should be made to conduct such programs in South and North East regions of US, which recorded the highest number of deaths

Derived Decisions

- The southern states have higher obesity rates when compared to other regions. It also recorded stroke as the major cause for the deaths. The South East regions recorded more in-patient stays in 2012. The South West and Atlantic South are in fifth and seventh position.

Region	States
South East	Alabama, Kentucky, Mississippi, Tennessee
South Atlantic	District of Columbia, Delaware, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia
South West	Arkansas, Louisiana, Oklahoma, Texas
North East- New England	Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island, Vermont
North East- Middle Atlantic	New Jersey, New York, Pennsylvania

THANK YOU
ANY QUESTIONS ?