

AIM

The study of this mortality dataset for the year 2014 is being conducted to find relationships between the different variables (like ethnicity, age, sex etc.) which comprise this dataset. The aim is to derive results and determine the reason for the events taking place. We are trying to find correlations between gender, age, diseases etc.

We find the average age of death in males and females, the top three diseases causing death in males and females and why differences exist. Finding the percentage of deaths in the age group of less than 5 and more than 100 to account for outliers. How manner of death differs in males and females in the age group of 25-34 years and how gender and marital status affect the average age of death.

Determining whether a certain disease is a common cause of death in different age groups. Finding average age based on races to find race with highest life expectancy. Percentage of deaths in blacks and whites based on homicides in 15-24 years and suicides in 25-34 years. Number of deaths due to injury at work based on education. The average age of death for males and females based on education. The percentage of deaths on streets and highways based on race and education.

DATASET SUMMARY

Every year the CDC releases the country's most detailed report on death in the United States under the National Vital Statistics Systems. This mortality dataset is a record of every death in the country for the year 2014.

Information is broken down by sex, age, race, marital status, education, icd10code and many more variables including some recodes. With a total of 2631171 records and 12131 diseases causing death, along with an age range of 1 to 118 with 1299710 males and 1331461 females divided over 16 races, this dataset doesn't forget the manner of death (Accident, suicide, natural etc.) or the education of the person dead or the marital status (Married, widowed, never married etc.).

The marital status of the people dead has no major category as there were **37% married** while **34% widowed** and **15% divorcees** while **12% never married** in the dataset. The education of these people were reported either by education 1989 revision or education 2003 revision. **70%** of the records had education reported by the 2003 revision. The 2003 revision had categories like **Doctorate or professional degree, associate degree, high school or GED completed** etc.

Injury at work tells us whether death was due to an injury at work or not. **90%** of the deaths due to injury at work was in **males**. Manner of death represents whether the cause of death was **self-inflicted** or **natural** or **suicide** etc. Natural deaths were at least **6 times more** than any other category.

TECHNICAL PROCESS

The dataset was obtained from <http://kaggle.com/cdc/mortality> and the files were copied into **HDFS** on Cloudera. The observations in the next section were obtained by analyzing the data using **hive** queries on the **hue** browser. The graphs below were made using the hue browser utility and the more complex ones were made by using the **ggplot2** library in **R**.

OBSERVATIONS

The average age of death in males is **70** while in females is **76**. Clearly females tend to live longer. According to *BBC Future* and *Harvard gazette* the heart rate of females double's during the second menstrual cycle giving them same benefit as moderate exercise. Also the estrogen in females directly leads to increase in good cholesterol thus increasing longevity of life.

The top three diseases causing death in males and females are given below:-

Males	Females
Atherosclerotic heart disease	Unspecified dementia
Malignant neoplasm: Bronchus or lung	Atherosclerotic heart disease
Acute Myocardial infarction	Malignant neoplasm: Bronchus or lung

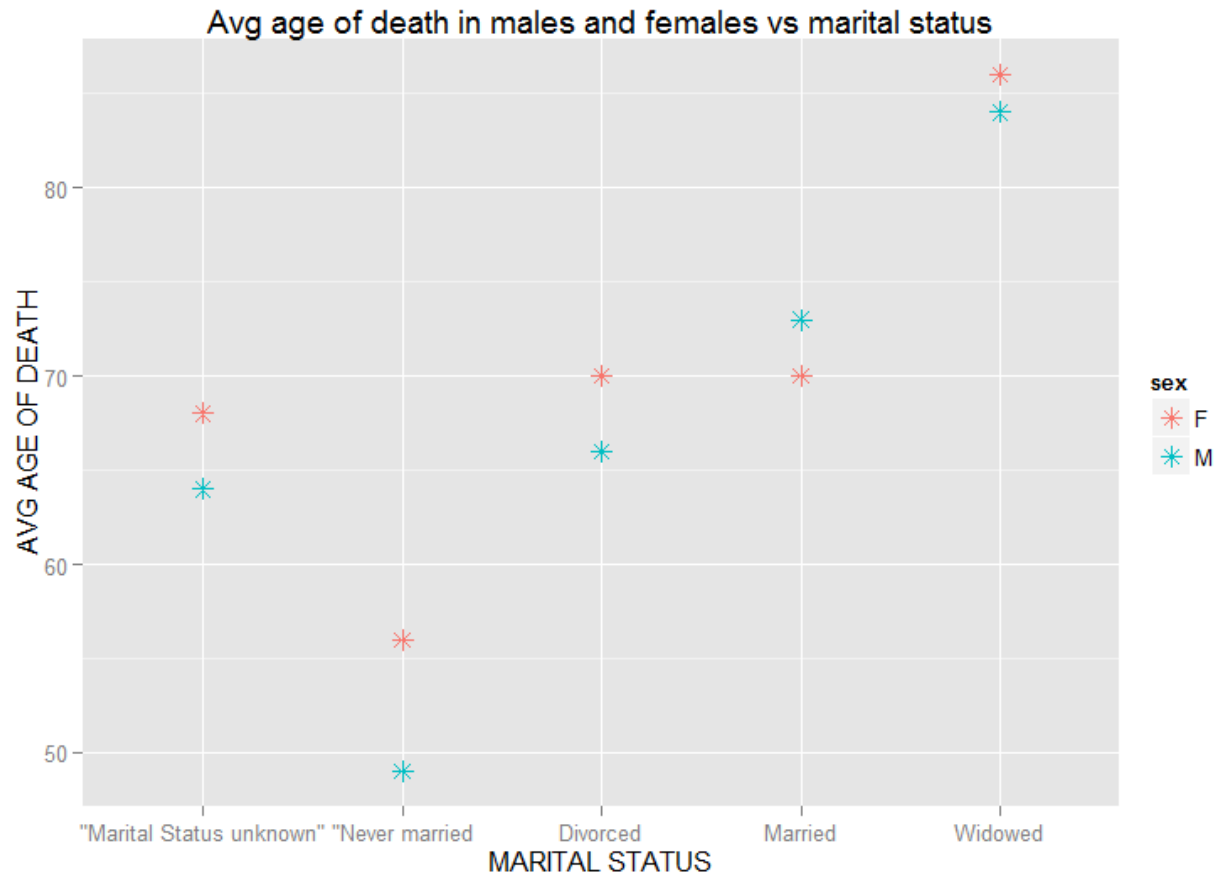
Unspecified dementia is a disease caused by age. As females generally tend to live longer thus it is more common in them. According to *Alzheimer's research study* women tend to care more for their male counterparts having dementia and due to lack of support from friends and family, 20% of these females tend to suffer from dementia as well. **Heart diseases** in females are less due to the good cholesterol increase and increase in heart rate (as mentioned above).

As opposed to many beliefs that more females die before the age of 5 our results show the contrary that is **56%** of the deaths were of males while **43%** were females. The longevity of female age is again portrayed by the fact that, in deaths of people above the age of 100 **83.56% deaths** were of females while only **16.13%** males.

Manner of death in 25-34 years is due to **natural causes** in females while **accidents** in males. According to the *vital statistics: special report* men exclusively tend to do the hazardous work like work on construction sites etc. where machinery may blow up or a blow from a falling object can cause death.

We can see from the graph below that **married** people tend to live longer than **never married**. Also males live longer only when they are married. Females whether married or divorced have the same average age of death while males who are divorced live **8 years less** on average.

According to *Medicaldaily* happily married people live longer because marriages offer emotional support, physical and intellectual intimacy and deeper social ties to family which reduces stress and improve health overall. Usually married people tend to abuse substances very less in comparison to their unmarried counterparts.



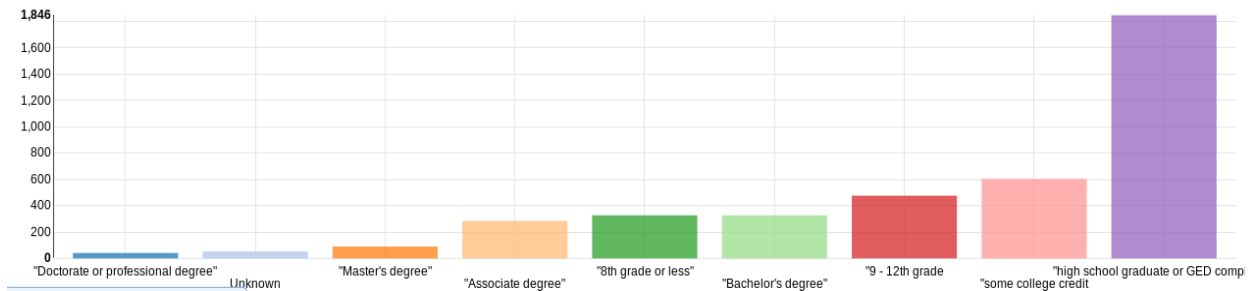
Married men tend to live longer than their divorced or single counterparts because they are less likely to commit suicide, and are more likely to do things like go to the doctor, get their checkups done, stop drinking alcohol and take their medication because wives stay on top of them.

The observations based on major death cause in children are that drowning is the major death cause in children of **1-4 years**. Children who live near open water sources, such as ditches, ponds, irrigation channels, or pools are especially at risk. In **5-14 years** suicide is the major cause of death. The reason for suicides in kids of **5-14 years** is attributed to the following factors: previous suicide attempts, presence of psychiatric disorders and psychopathology, preoccupation with death and family history of psychopathology and suicidal behavior.

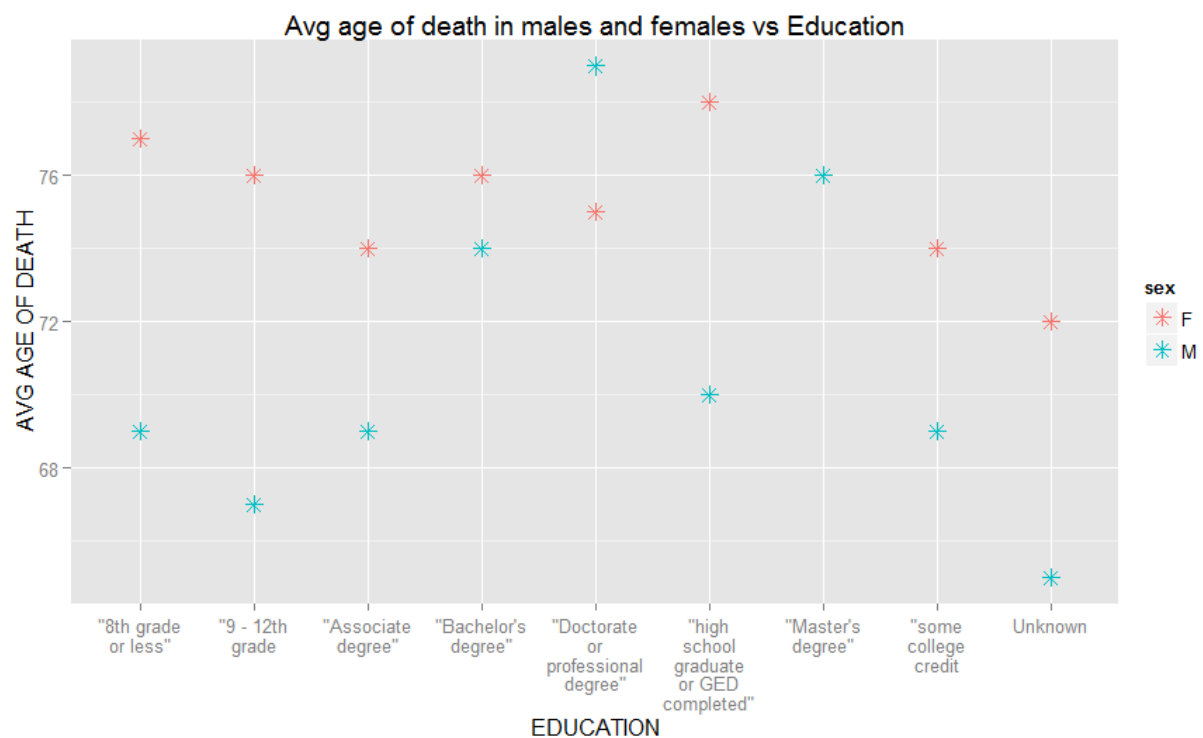
Japanese people in USA have the highest average age of death i.e. **82 years** followed by **77 years** in Chinese. According to *tofugu.com* Japanese live longest because of their diet. For example they prefer fish over red meat because red meat has loads of cholesterol which has its effect in later years in life. According to *livescience.com* Chinese live long because of the large amounts of exercise they do.

Our observations show that homicides in blacks in the age group of **15-24 years** is **8.2 times** more than that of whites. African-Americans tend to die mostly by firearms.

No. of deaths due to injury at work vs Education



From the graph above we can see that doctors died the least due to injury at work while high school graduates died the most. People having less education have to usually work in more hazardous and unsafe environments thus increasing their probability to die due to injury at work. Education helps to improve not just lifestyle but also working environment.



As we can see above in **females** average age of death is mostly between 70-75 while in **males** 69-80 approx. Males having a **master's** or **Doctor's** degree tend to live longer than males with other degrees. Education exerts its direct beneficial effects on health through the adoption of healthier lifestyles, better ability to cope with stress, and more effective management of chronic diseases.

After finding the percentage of deaths on streets and highways we saw that black male's (in 9-12th grade) mortality rate on streets and highways is approximately **5 times** more than whites in the same grade.

INFERENCES TO INCREASE LIFE LONGEVITY

Education is a major beneficiary to the standard of life and thus the age of death. A higher degree ensures a safe working environment along with higher pay which allows regular visits to the doctor. Marriage though considered a legal union between a man and woman is an important factor in increasing life expectancy especially for males. Better social ties ensure less emotional stress. The diet in case of Japanese made their average age of death 5-7 years more than most other races while Chinese prefer remaining fit by exercise. Thus diet and exercise also help increase life.