Death Report in the United States in 2015

Zhenyue Liu | EM 622 |Final Project

#Background

A new year brings U.S. and world population totals. According to the U.S. Census Bureau's annual projections, the United States will enter 2015 with 320,090,857 people while the planet Earth overall will have more than 7.2 billion living souls on it.

Overall the U.S. remains firmly entrenched as the third most populous country in the world, behind China (nearly 1.4 billion) and India (nearly 1.3 billion) and ahead of Indonesia (nearly 256 million). The U.S. population is 2,334,187 people larger than the start of 2014, according to the Census Bureau; that's a 0.73 percent increase from last year, which is pretty much in line since year 2000.

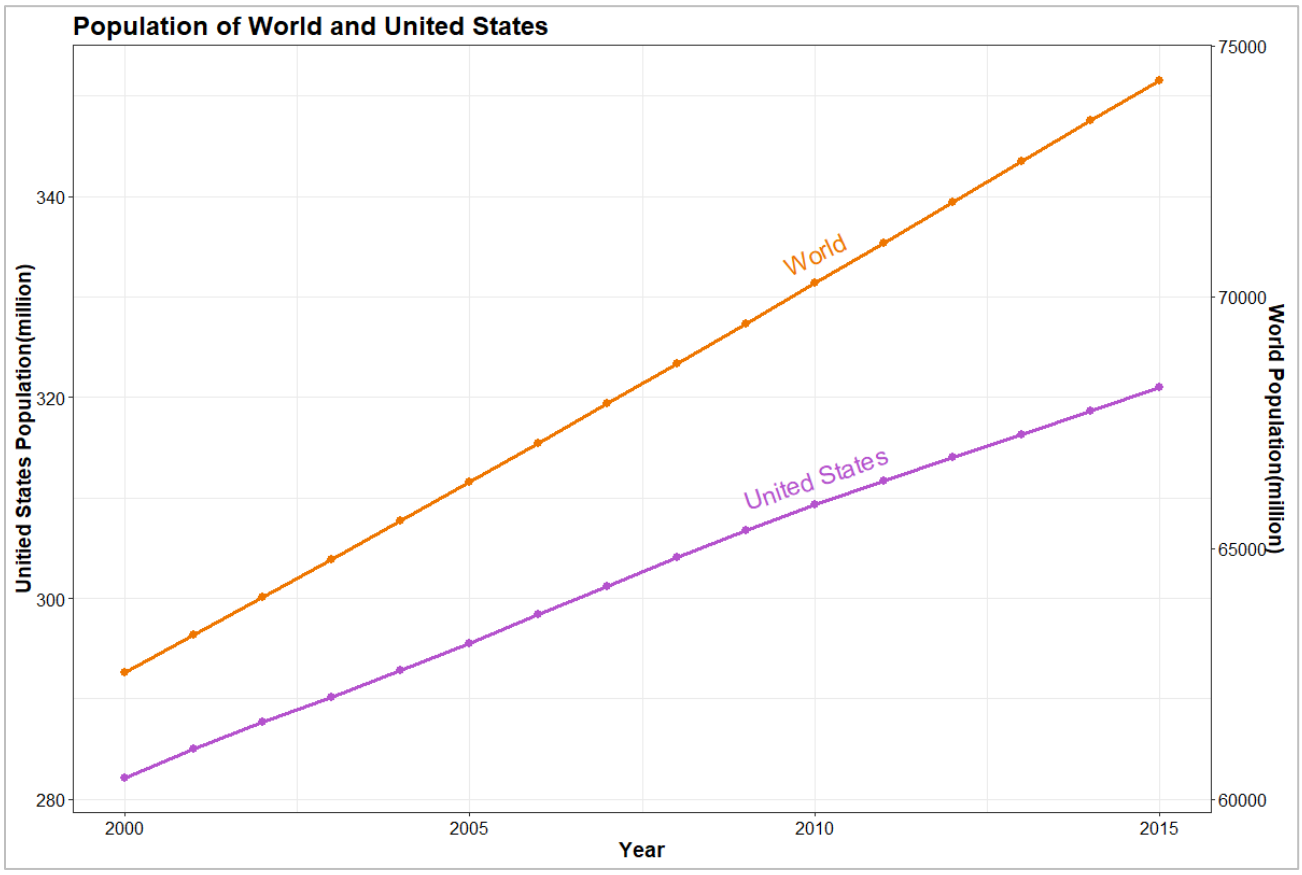

Figure 1 Population of World and the United States

Meanwhile, when we focus on year 2015, the net birth rate also reached to a new peak. The U.S. welcomed 3.98 million births, and experienced 2.71 million people dead. I also added international immigrant statistic, as they are an important component of the U.S. All told – between the births, deaths and immigration – the country's population increases by one person every 16 second, and net population change 2.42 million people compared to 2014.

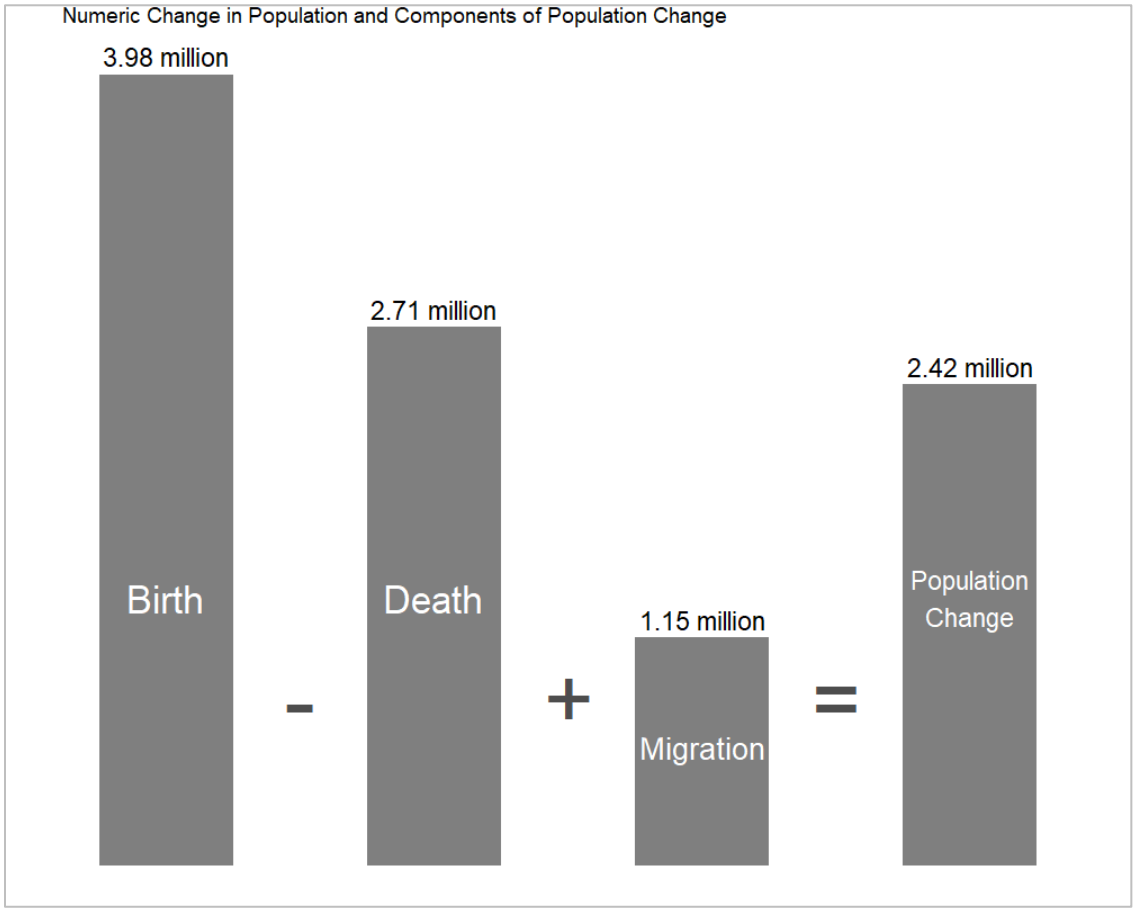

Figure 2 Numeric Change in Population and Components of Population Change

#Introduction

The dataset comes from the "Detailed Mortality" database available on https://wonder.cdc.gov/. The data contain 6940 observations and 12 variables, mostly about the place where people die, and the cause of death. As all the category codes and useless data has been erased from the original data, the five important ones left:

1. Ten.Year.Age.Group: Ages are roughly grouped by each 10 years.
2. Place.of.Deaths: The place where decedents die.

3. ICD.Chapter: Chapter of the International Classification of Diseases

4. ICD.Sub.Chapter: Sub Chapter of the International Classification of Diseases

5. Deaths: The number of people die.

Since there is no specific mission of this project, so my first step of the work is to do some exploration with the data, and find something interesting or abnormal against our common knowledge. Following is the summary of my cleaned data, and I choose to start with the place of death.

```
> summary(data_clean)
  Ten.Year.Age.Groups                          Place.of.Death                                ICD.Chapter
 65-74 years: 843     Medical Facility - Inpatient       :1361   Certain infectious and parasitic diseases : 653
 75-84 years: 836     Decedent's home                    :1179   Diseases of the nervous system            : 638
 85+ years  : 833     Medical Facility - Outpatient or ER:1101   Diseases of the circulatory system        : 604
 55-64 years: 813     Nursing home/long term care        : 871   Diseases of the respiratory system        : 580
 35-44 years: 771     Other                              : 849   External causes of morbidity and mortality: 558
 25-34 years: 671     Hospice facility                   : 796   Diseases of the digestive system          : 552
 (Other)    :2173     (Other)                            : 783   (Other)                                   :3355
                                  ICD.Sub.Chapter     Deaths
 Malignant neoplasms                        :  86   Min.   :    1.0
 Other external causes of accidental injury :  86   1st Qu.:    2.0
 Transport accidents                        :  86   Median :    9.0
 Ill-defined and unknown causes of mortality:  84   Mean   :  393.4
 Assault                                    :  82   3rd Qu.:   57.0
 Other forms of heart disease               :  82   Max.   :70747.0
 (Other)                                    :6434   NA's   :44
```

#Place of Death Frequency

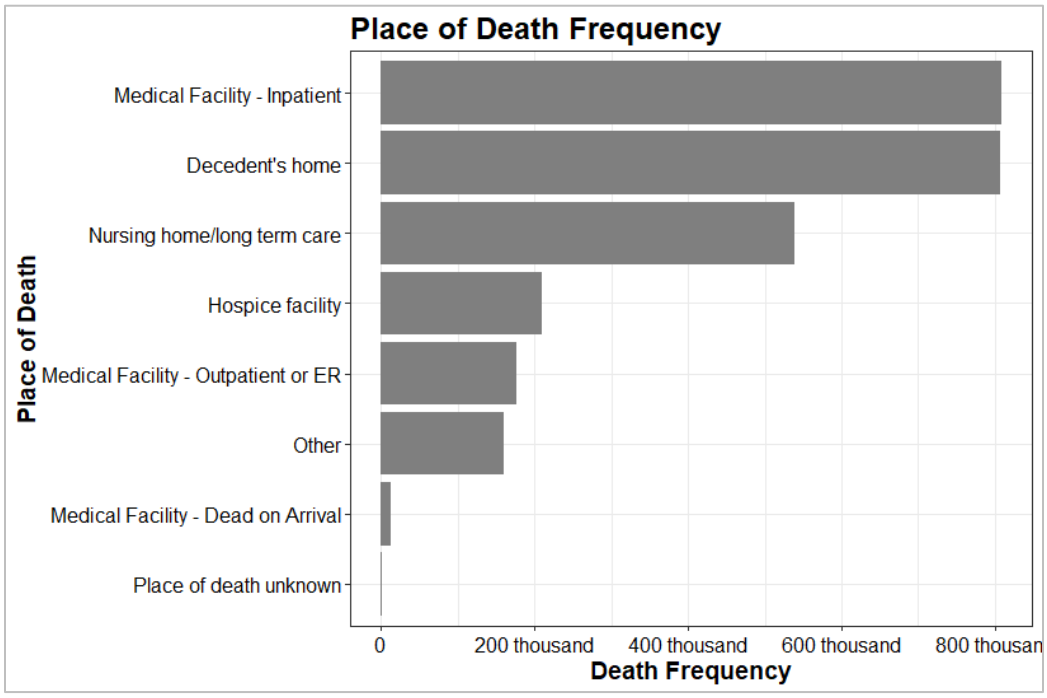

Figure 3 Place of Death Frequency

From Figure 3, we could find out that over half of decedents pass away in medical facility and their own home, and small part of them end their life during long term care at nursing home. The other places just take quite small part of the statistics.

It's easy for us to categorize the place of death with different situation, for example, people who die in medical facilities must have cancer, disease in circulatory system or serious trauma, and the number of these inpatients cannot be small. Meanwhile, the number of decedents who die at their own home is too large to make me surprised. Generally speaking, people who die at home are usually thought to be natural death caused by organ failure, since they must be sent to hospital with better medical condition and environment if they suffered from any diseases.

Therefore, my next step is to explore what kind of diseases take decedents away at their home instead of at hospital.

People die at home by ICD Chapters

All kind of diseases in the dataset are categorized by ICD(International Classification of Diseases). The pie chart from Figure 4 tells us the top two killers are 'Disease of the circulatory system' and 'Neoplasm', both take almost one third of the whole data.

The most common sub chapter of disease of circulatory system is heart disease, especially Ischaemic heart diseases, which also known as sudden heart attack.

In neoplasms chapter, Malignant neoplasms is the largest cause of death. However, most patients of malignant neoplasms often experienced a quite long-term suffering, there's little possibility, generally speaking, that they choose to a place without good medical condition for recovery, unless——the probability to cure the malignant neoplasms is minimal.

Thus, I'd like to see if the heart disease and malignant neoplasms have little probability to cure even in other medical facility or nursing home.

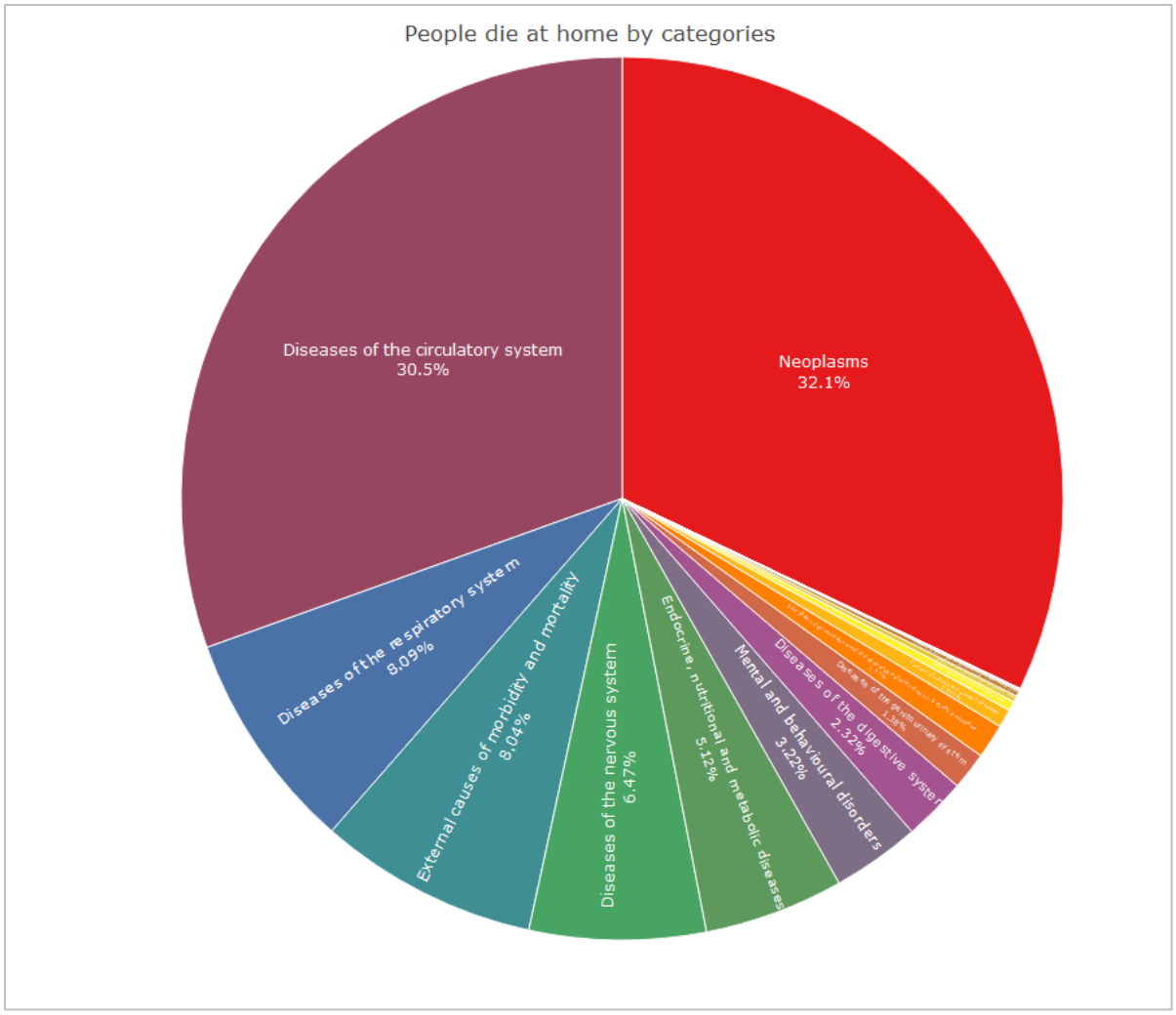

Figure 4 People die at home by ICD Chapters

People die in different places by diseases

Figure 5 mosaic graph connect two categories, one is place of death and the other one is diseases. The size of each sqaure illustrate the number of decedents die where and caused by what kind of disease.

It's quite clear that disease of circulatory system and neoplasms are the two largest causes of death, wherever they are. Therefore, it can be explained partly that lots of people who had neoplasms choose to stay at home instead of going to hospital, because the death rate is too high to get over. At the end of their life, hospital with best medical condition has no comparability to a familiar place surround by family

and friends nearby. However, it's just a unreliable inference from such data, if we are going to confirm it, we need not only larger data but also more variables.

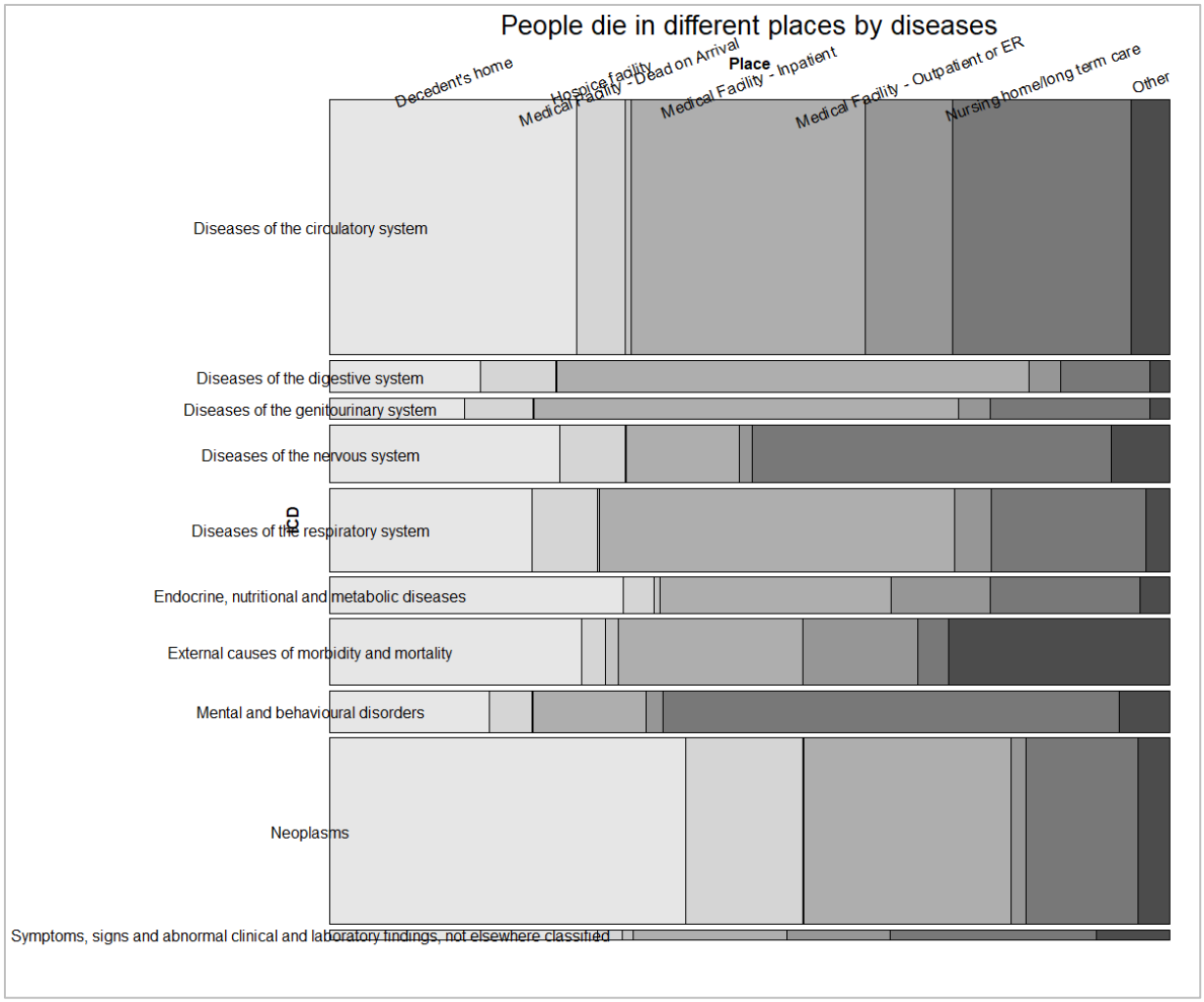

Figure 5 People die in different places by diseases

#Abnormal observation of Age

Except all the graphs presented above, I also did lots of other exploration. When I deal with the age data, something abnormal come to my mind.

Figure 6 shows the relationship between age and disease. Some specific disease, such as 'Certain conditions originating in the perinatal period', 'Congenital malformations, deformations and chromosomal abnormalities', mostly just cause children under 1 years old dead, and 'Pregnancy, childbirth and the puerperium' only bring death to people from 10 to 40 years old——women who are able to

conceive. Except for these four to five kinds of disease, the probability of other diseases causing death all roughly rises with ages.

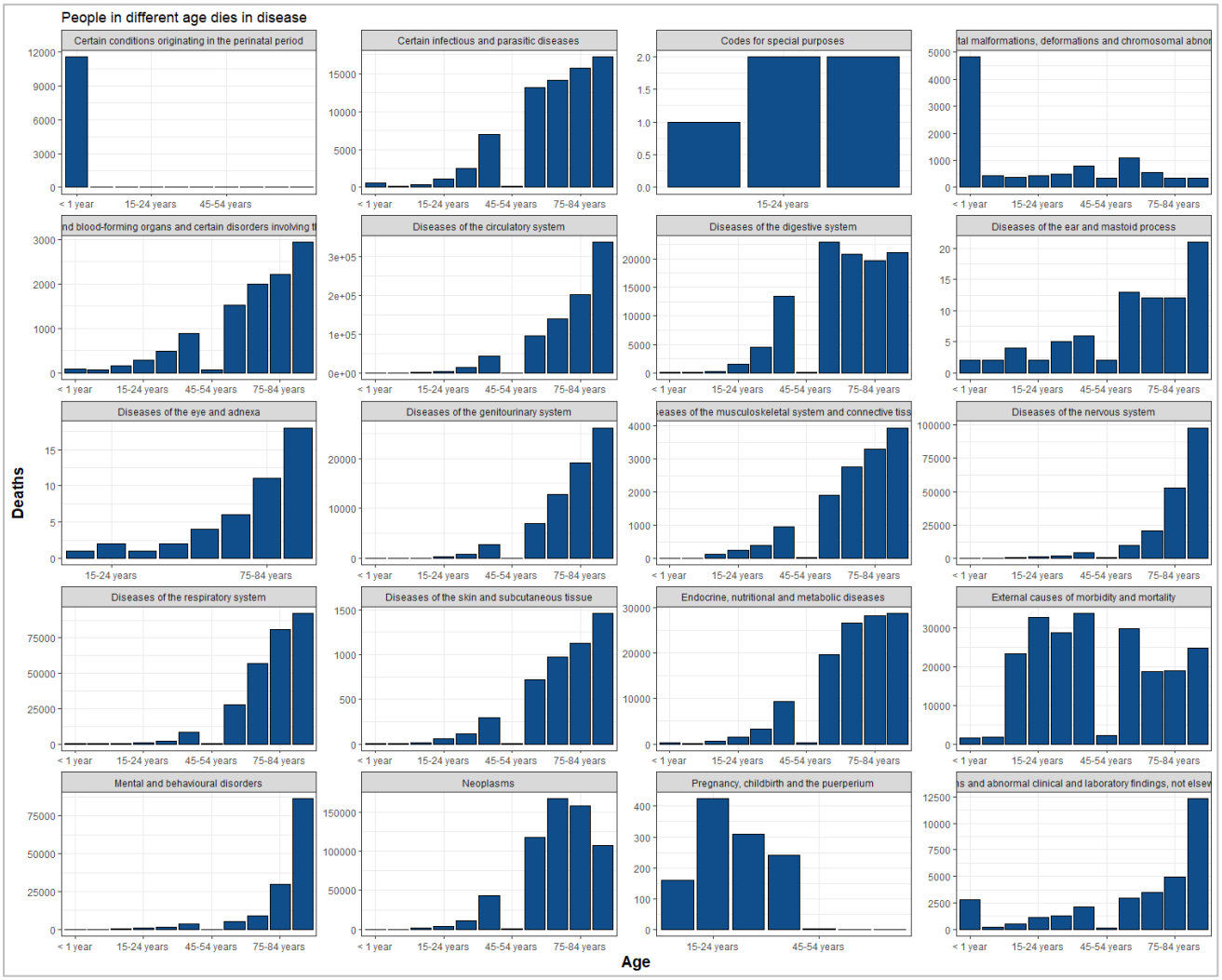

Figure 6 People in different age dies in disease

However, a weird phenomenon in each graph is, every time when the age comes to group ‘45-54 years old’, the number of deaths drops dramatically. Then I made a summary of people die in different age, I am even more confused about the data, since the number of dead people aged from 45 to 54 years old in 2015 is just 5411. Compared to over 170,000 people and 357,000 people die in their 30s and 60s, the death data of ‘45-54 years old’ is quite abnormal.

In my opinion, one reason and the most possible reason is the data has something wrong, though the website is really authoritative. The other reason is that, back to the time they were born, from 1961 to 1970, the birth rate was extremely low, or

large-scale war killed almost all of them when they grew up. All the inference also needs more data supported, it won't be discussed in this report any more.

#Conclusion

The top two killers are 'Disease of the circulatory system' and 'Neoplasm', not only take people away from the world, but also erase the hope of being alive early, leaving patients to die at home.

One 'unexpected' news is that, people aged from 45 to 54 years old seldom get all kind of diseases, and may live even much longer than their young generations——if the data has nothing wrong.