

CSDE 502: Assignment #8

Maxine Wright

March 12, 2020

Instructions:

Using R Markdown, write a brief paper (about two pages) about the country that you've chosen, or make a comparison of two or more countries. Include in this paper:

1. A brief introductory paragraph describing the country (feel free to steal something from the web).
2. Verbiage that describes the data and references the HMD data source.
3. An analysis of the data to describe something interesting about the country.
4. Include at least one of each of the following text with imbedded R:
 - hyperlink
 - table
 - plot
 - some R Markdown thing you've never used before

Introduction:

Chile is a country in South America bordered by Bolivia, Argentina, and Peru. It is a long, narrow country situated next to the Pacific Ocean, with over 6,000km of coastline. The capital of Chile is Santiago, a city in the southern part of the country. One of the main features of the country is the Atacama Desert in the north, which borders the national park and flamingo reserve on the Bolivian border.

The death data included in the Human Mortality Database for Chile cover the *de facto* population and are organized in a 1x1 format. This means that deaths are categorized by one-year age groups (0 to 110+) and year (1992 to 2008). The death total is further broken down by Male and Female. For example, the highest total number of deaths was in 1992 in the age 0 group, at 4209; within that, 1890 of the 4209 deaths were female, while 2319 were male.

Canudas-Romo and Castro (2017) provide information about what is included in the historical mortality series about Chile from the Human Mortality Database here. Most notably, they provide a rationale for the restriction of the mortality data range to the years between 1992 and 2008, saying:

“The historical mortality series for Chile included in the Human Mortality Database (HMD) covers the period from 1992 to 2007. This initial series was selected after assessment of the developments in birth, death and population counts during the second half of the twentieth century. Parts of this evaluation are included in the present document. The main reasons for restricting the HMD series to the period since 1992 are: 1) only the two most recent censuses of 1992 and 2002 show below 3% of incorrect reporting of age; 2) it is only in the last decade of the twentieth century that the percentage of delayed birth registration is 5.4% or lower; and 3) finally, during this period, at least 90% of deaths were certified by a physician.” - Canudas-Romo and Castro (2017), “ABOUT MORTALITY DATA FOR CHILE”, page 2.

Information about data sources used by the Human Mortality Database to compile these demographic statistics is available here.

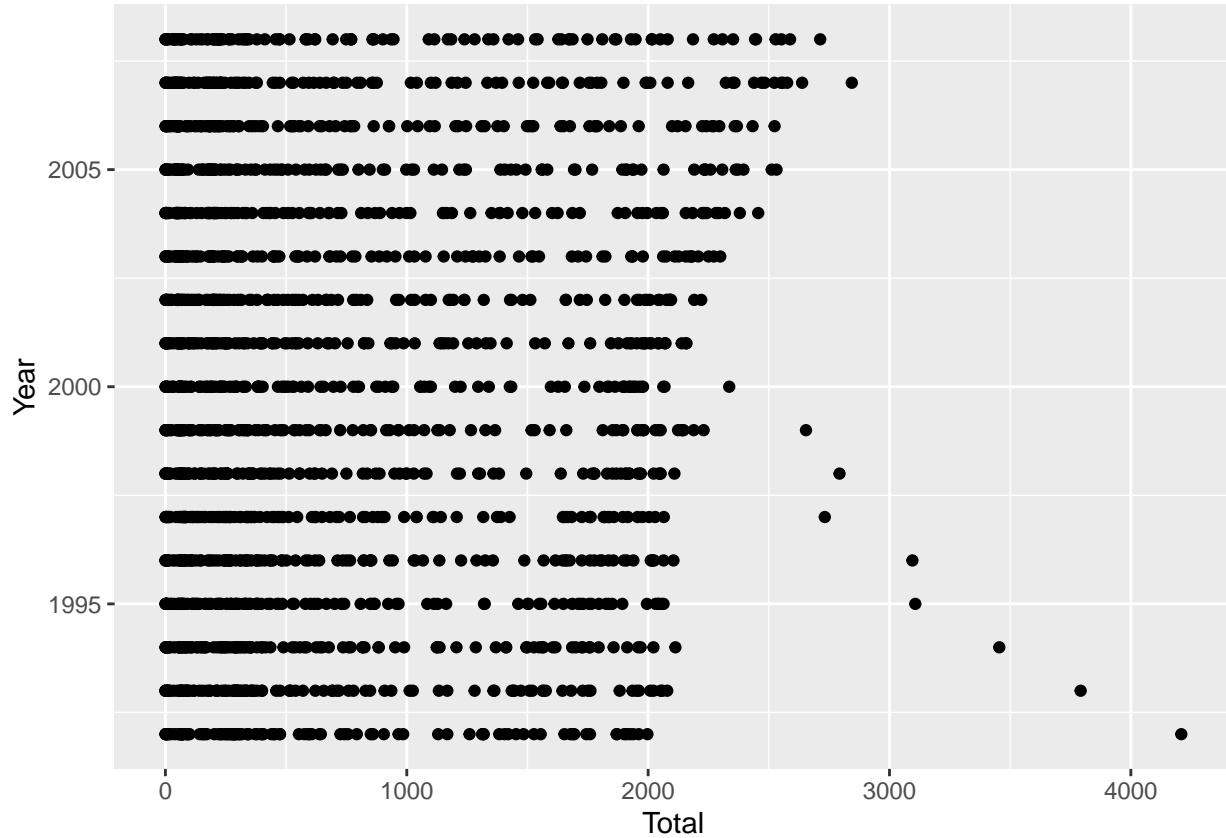


Figure 1: A Flamingo at The Flamingo National Reserve (taken by me!)

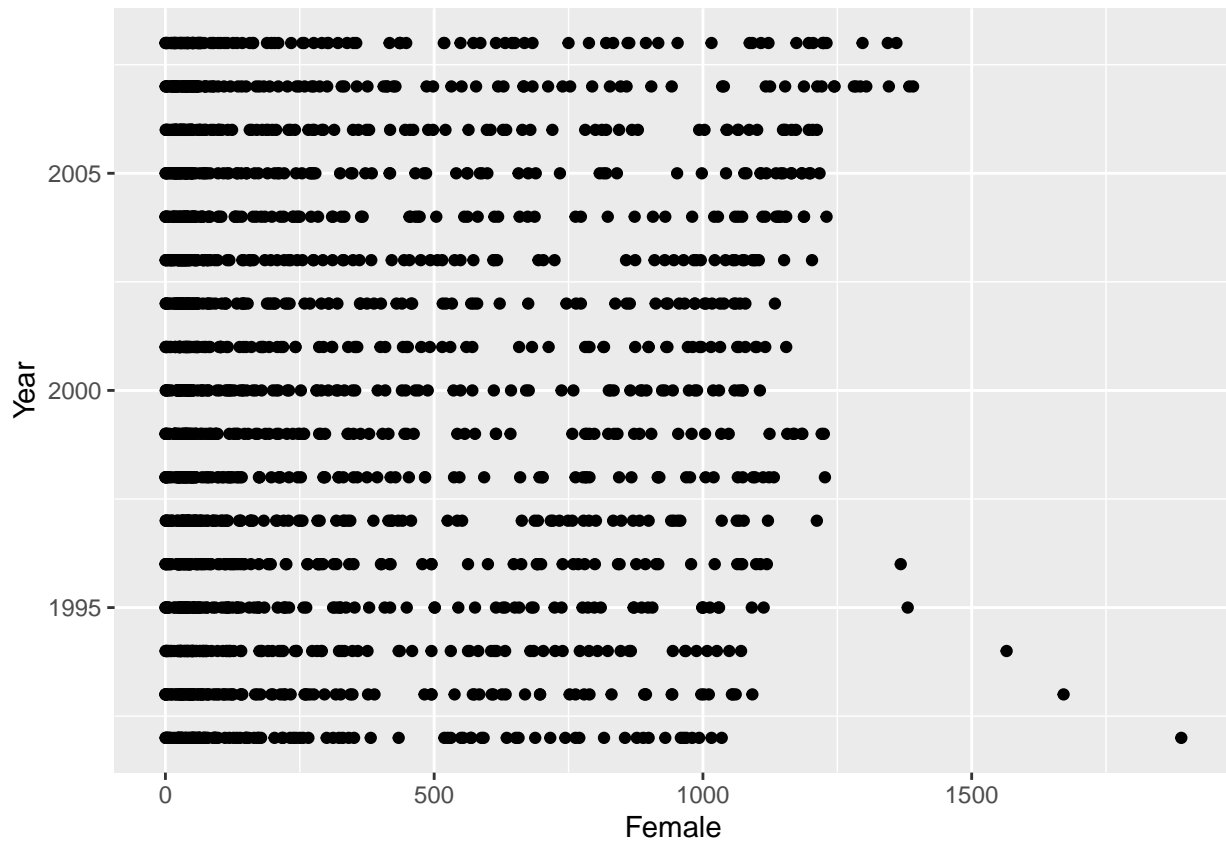
Plots and Tables

The three plots below show the total deaths by year, followed by female deaths and male deaths.

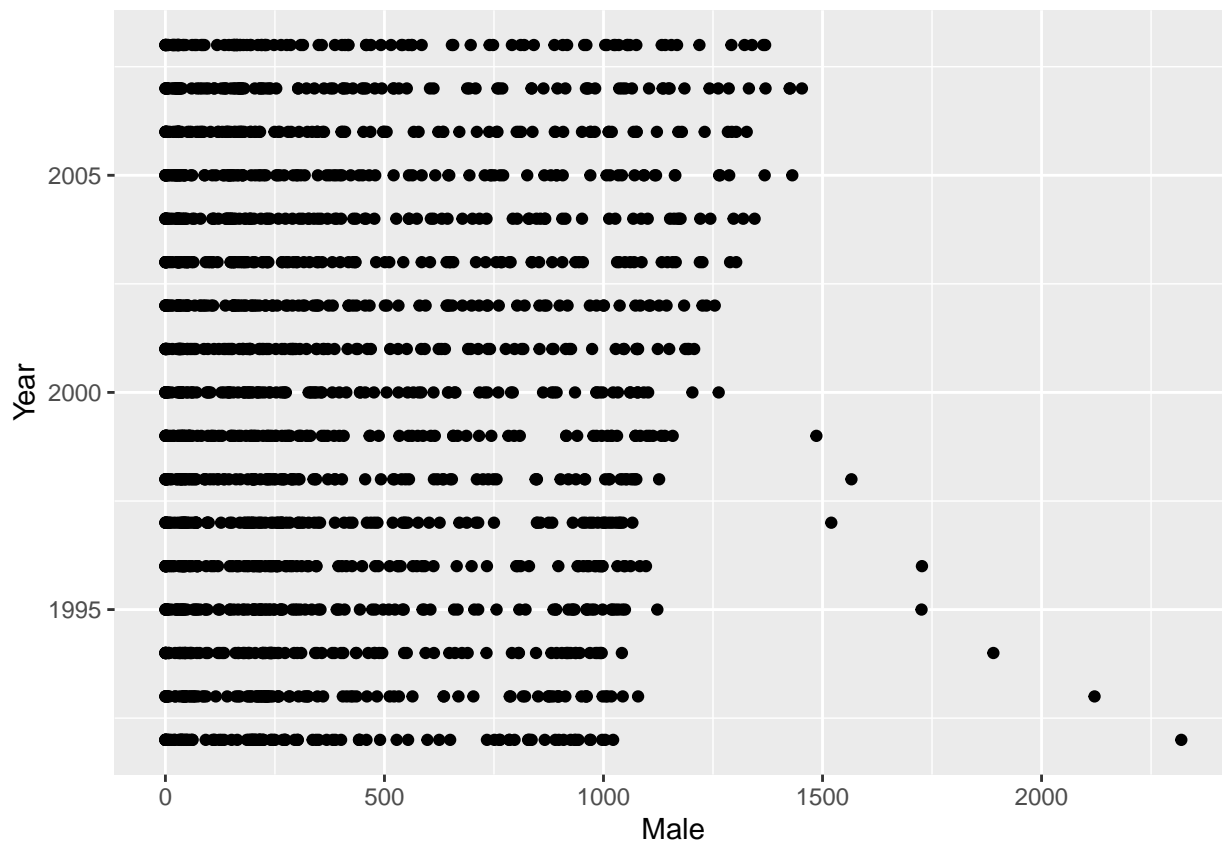
```
ggplot(data=x, aes(x=Total, y=Year, group=1))+  
  geom_point()
```



```
ggplot(data=x, aes(x=Female, y=Year, group=1))+  
  geom_point()
```



```
ggplot(data=x, aes(x=Male, y=Year, group=1))+
  geom_point()
```



As you can see, the number of deaths per year are roughly consistent, with a few outliers. Next, we have the information in grouped table form for male and female deaths:

```
pander(table(x$Female_group))
```

1	2	3	4	5
806	558	349	171	3

```
pander(table(x$Male_group))
```

1	2	3	4	5
515	761	417	187	7