

Causes-of-death-in-the-world

2022-05-03

Causes-of-death-in-the-world

Module project of PSY6422

Data sources

The data is sourced from a page on Our world in data. There are five files of data that I downloaded. The author of the page has divided the raw data into five groups by age, including under 5, 5-14, 15-49, 50-69, and 70+. The columns for each group include Entity, Code of country, year, and a number of deaths by cause of death. There are four variables in total: age, year, region, and cause of death.

These data are collected from a number of institutions, such as, Institute of Health Metrics and Evaluation (IHME), Global Burden of Disease (GBD), World Health Organization (WHO) Global Health Observatory (GHO), Global Terrorism Database (GTD), Amnesty International

Research Questions

In the epidemiological framework of the Global Burden of Disease Study, each death has a specific cause. In their own words. Every death is attributed to a single underlying cause - the cause that initiates the sequence of events that lead to death. Analysis of the causes of death may be useful in identifying risk factors or effective preventive measures to reduce the number of deaths.

According to the official website, approximately 56 million people die each year. This report analyses change over time in the number of deaths caused by different causes of death. It also looks at regional and age differences. In other words, the causes of death are different for people of different ages and in different regions, and the changes over time are also different. This report presents some of the results of this analysis.

Data Preparation

```
#loading data
library(here)
here()
data_under_5 <- read.csv(here("raw","under-5.csv"))
data_5_14 <- read.csv(here("raw","5-14.csv"))
data_15_49 <- read.csv(here("raw","15-49.csv"))
data_50_69 <- read.csv(here("raw","50-69.csv"))
data_70_ <- read.csv(here("raw","70-.csv"))

#clear names#
names(data_under_5) <- names(data_under_5) %>%
  sub('Deaths...', '', .) %>%
  sub('...Sex..Both...Age..Under.5..Number.', '', .) %>%
```


the required data from `data_all` for drawing.

Visualization

line chart

```
#Extract data#
data_line <- data.frame(matrix(0,nrow=30,ncol=37))
for (i in 1:40050){
  for(k in 1990:2019) {
if (data_all$Entity[i] == "World" && data_all$Year[i] == k){
  data_line[k-1989,1:37] <- data_line[k-1989,1:37] + data_all[i,5:41]
}
}
}
names(data_line) <- names(data_all[5:41]) #adjust the name of columns#
data_line$Year <- c(1990:2019) #add Year column#

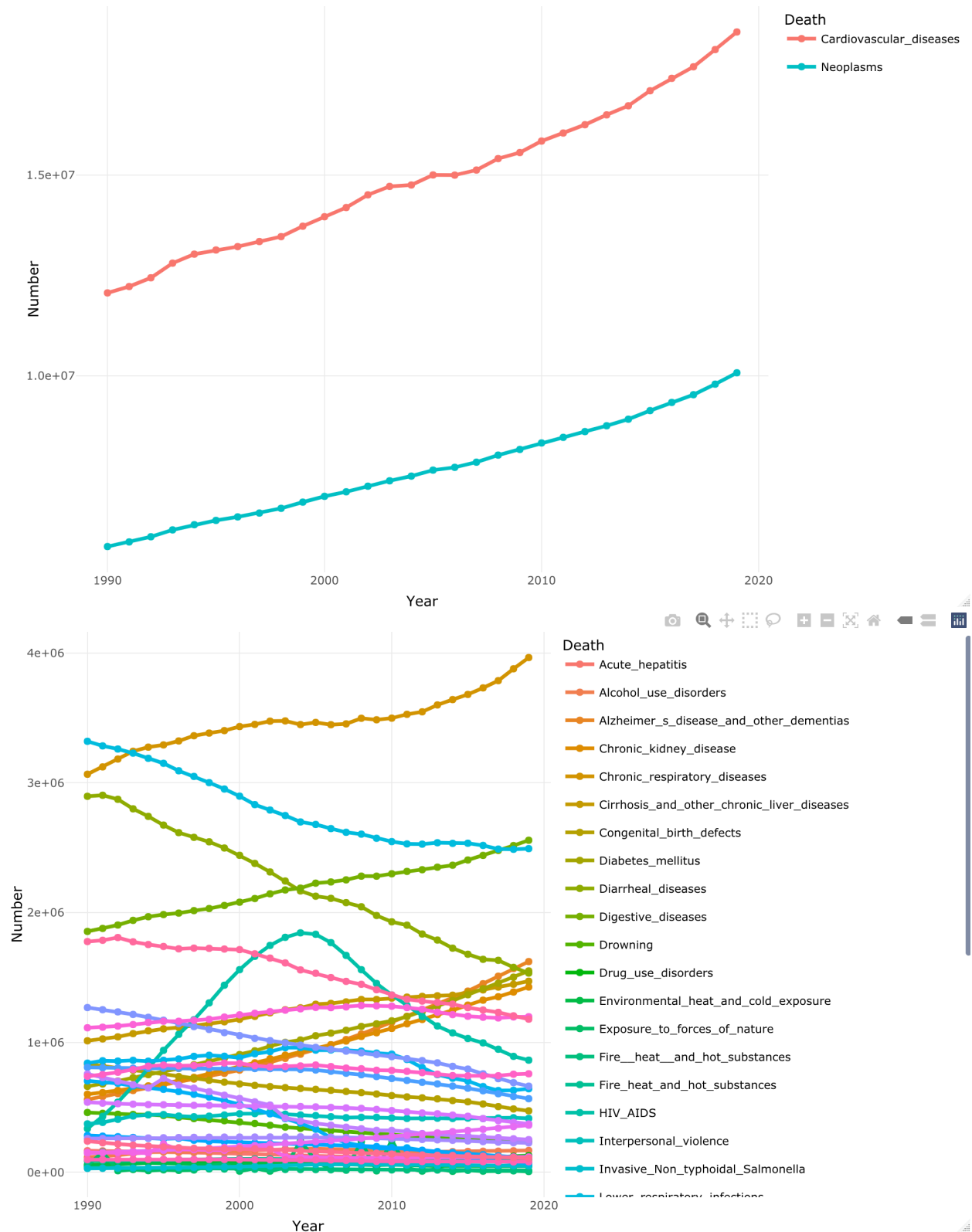
#As the number of patients for Cardiovascular_diseases and Neoplasms is too high, the diagrams are very
#separate two diseases#
data_line_toomany <- subset(data_line,select = c(Year,Cardiovascular_diseases,Neoplasms))
data_line_general <- subset(data_line,select = -c(Cardiovascular_diseases,Neoplasms))

#data gather#
data_line_toomany <- data_line_toomany %>%
  gather(key = "Death", value = "Value", -Year)
data_line_general <- data_line_general %>%
  gather(key = "Death", value = "Value", -Year)

#line chart#
line_toomany <- ggplot(data = data_line_toomany,mapping = aes(x = Year, y = Value, color=Death)) +
  geom_point()+
  geom_line(size=1) +
  xlab("Year")+ #x axis label
  ylab("Number") + #y axis label
  scale_y_continuous(labels = scales::scientific) + #labels using scientific notation
  theme_minimal() #Minimal themes
ggplotly(line_toomany) #add mouse interaction

line_general <- ggplot(data = data_line_general,mapping = aes(x = Year, y = Value, color=Death)) +
  geom_point()+
  geom_line(size=1) +
  xlab("Year")+ #x axis label
  ylab("Number") + #y axis label
  scale_y_continuous(labels = scales::scientific) + #labels using scientific notation
  guides(colour = guide_legend(ncol = 1)) + #The legend has only one column
  theme_minimal() #Minimal themes
ggplotly(line_general) #add mouse interaction

#save image#
ggsave(line_toomany,filename = here("Figures","line_chart_toomany.pdf"),width = 12,height = 9)
ggsave(line_general,filename = here("Figures","line_chart_general.pdf"),width = 12,height = 9)
```



This section, as a whole, looks at the change in all data for all categories of causes of death over the thirty-year period from 1990 to 2019. It can be seen that some diseases have changed very little, such as whooping cough, and others have changed dramatically, such as HIV, cardiovascular disease. When a cause of death

with a relatively large variation in the number of deaths is identified, this one cause of death can be analysed to find its root cause. The cardiovascular disease is not only the most common cause of death, but it has increased very rapidly over the last 30 years, so the next section will analyse cardiovascular disease

bar chart

```
#Extract data#
data_bar <- data.frame(matrix(0,nrow=150,ncol=3))
names(data_bar) <- c("Age","Year","Cardiovascular_diseases")
k=1
for (i in 1:40050){
  if (data_all$Entity[i] == "World") {
    data_bar[k,1:3] <- data.frame(data_all$Age[i],data_all$Year[i],data_all$Cardiovascular_diseases[i])
    k = k+1
  }
}

#bar chart#
bar_chart <- ggplot(data = data_bar, mapping = aes(x = Year, y = Cardiovascular_diseases,fill = Age)) +
  geom_bar(stat = 'identity',position = 'dodge') +
  scale_fill_manual(name = 'Age',
                    values = c('#FF4500','#9400D3','#F4A460','#90EE90','#6495ED')) + #change color
  theme_minimal() #Minimal themes
ggplotly(bar_chart) #add mouse interaction

#save image
ggsave(bar_chart,filename = here("Figures","bar_chart_Cardiovascular_diseases.pdf"),width = 12,height =
```

This section uses bar charts to compare the number of deaths due to cardiovascular disease in different age groups. It is also possible to compare their differences between years. It can be seen that: 1) cardiovascular disease causes a greater number of deaths as people get older, with very few children and young people; 2) the number of deaths due to cardiovascular disease shown in the line graph rises rapidly over time, mainly from the increase in people over the age of 70.

map

```
#loading map#
data_map <- read_sf(here('map','ne_10m_admin_0_countries.shp'))
#change name match the map and data_ll
for (i in 1:40050){
  if (data_all$Entity[i] == "Russia") {
    data_all$Entity[i] = "Russian Federation"
  } else if (data_all$Entity[i] == "Democratic Republic of Congo"){
    data_all$Entity[i] = "Democratic Republic of the Congo"
  } else if (data_all$Entity[i] == "Congo"){
    data_all$Entity[i] = "Republic of the Congo"
  } else if (data_all$Entity[i] == "South Korea"){
    data_all$Entity[i] = "Republic of Korea"
  } else if (data_all$Entity[i] == "North Korea"){
    data_all$Entity[i] = "Dem. Rep. Korea"
  }
}
```

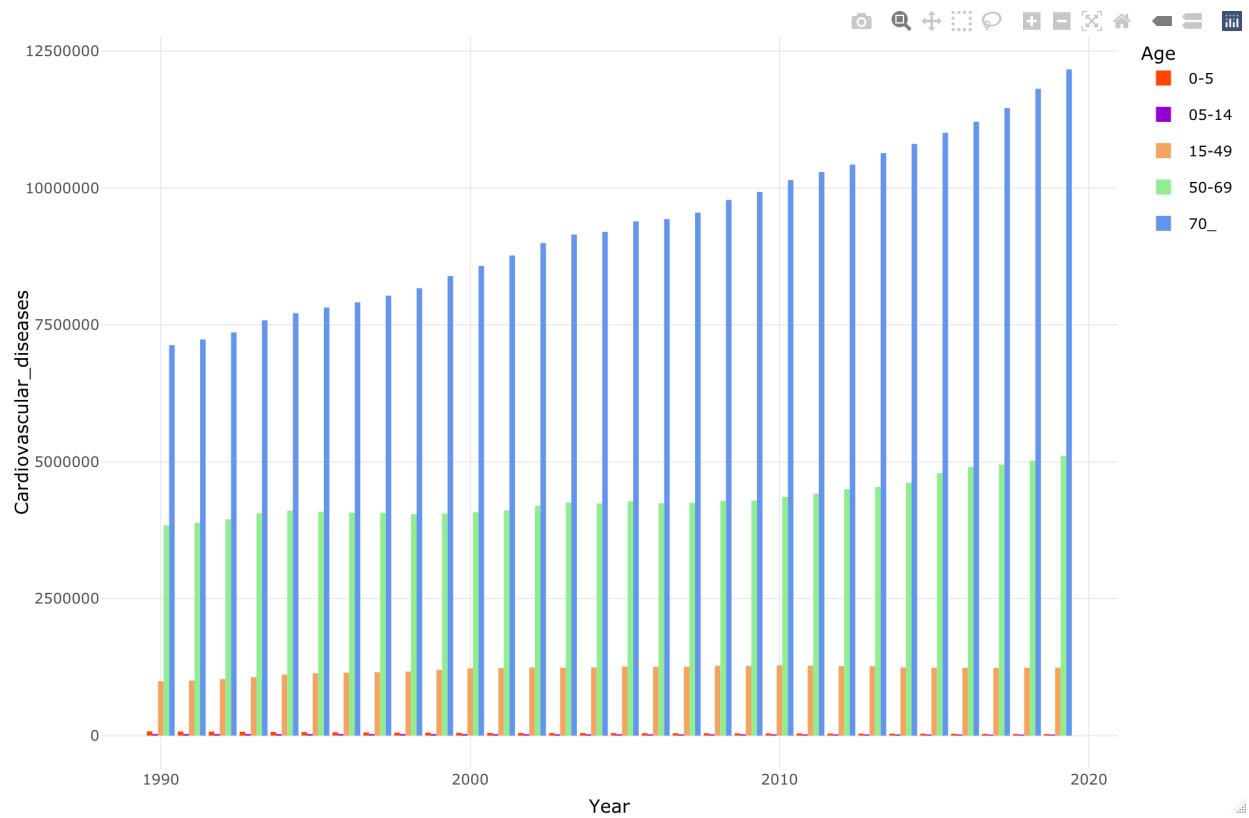


Figure 1: bar_chart

```

#Extract data#
value <- data.frame(NAME_LONG=data_map$NAME_LONG, VALUE = 0)
for (i in 1:40050){
  for (k in 1:258) {
    if (data_all$Entity[i] == value$NAME_LONG[k] && data_all$Year[i] == 2019){
      value$VALUE[k] = value$VALUE[k] + data_all$Cardiovascular_diseases[i]
    }
  }
}
}#This loop may take a few minutes

value$RATIO <- value$VALUE / data_map$POP_EST #Since the total number of people in each country varies,

shape <- merge(data_map, value, by = "NAME_LONG") #Combining information from map and extract data

i_popup <- paste0("<strong>Country: </strong>", shape$NAME_LONG, "<br>", "<strong>Ratio: </strong>", sha

pal <- colorBin(c("darkred", "orangered", "yellow", "darkgreen", "blue"), shape$RATIO, 11) #change the c

map <- leaflet(shape) %>% addTiles() %>%
  addProviderTiles("Esri.WorldStreetMap") %>% #change the background
  setView(0.000000, 0.000000, zoom = 2) %>% #methods to manipulate the map widget
  addPolygons(color = ~pal(shape$RATIO), fillOpacity = 0.8, weight = 1, popup = i_popup) %>% #highlight
  addLegend(pal = pal, values = shape$RATIO, position = "bottomright", title = "Data Map") #adjust legen

map

#save as .html#
saveWidget(map, here("Figures", "map_Cardiovascular_diseases_2019.html"), selfcontained = FALSE)

```

The final section generates a map of the world distribution of mortality due to cardiovascular disease, which allows the differences in mortality rates between regions to be observed. Although only one year can be plotted, any year can be plotted by modifying the parameters of the extracted data

conclusion

Finally, I learnt a lot about generating good looking images, which will help me a lot with the images in my future papers, and reports. For all this data, I can show only a very, very small part of it on this page, the rest of the many causes of death etc. there is plenty to compare or graph to discover some important information. In fact I have also done a world map of HIV.AIDS and found that it is concentrated in the South African region.

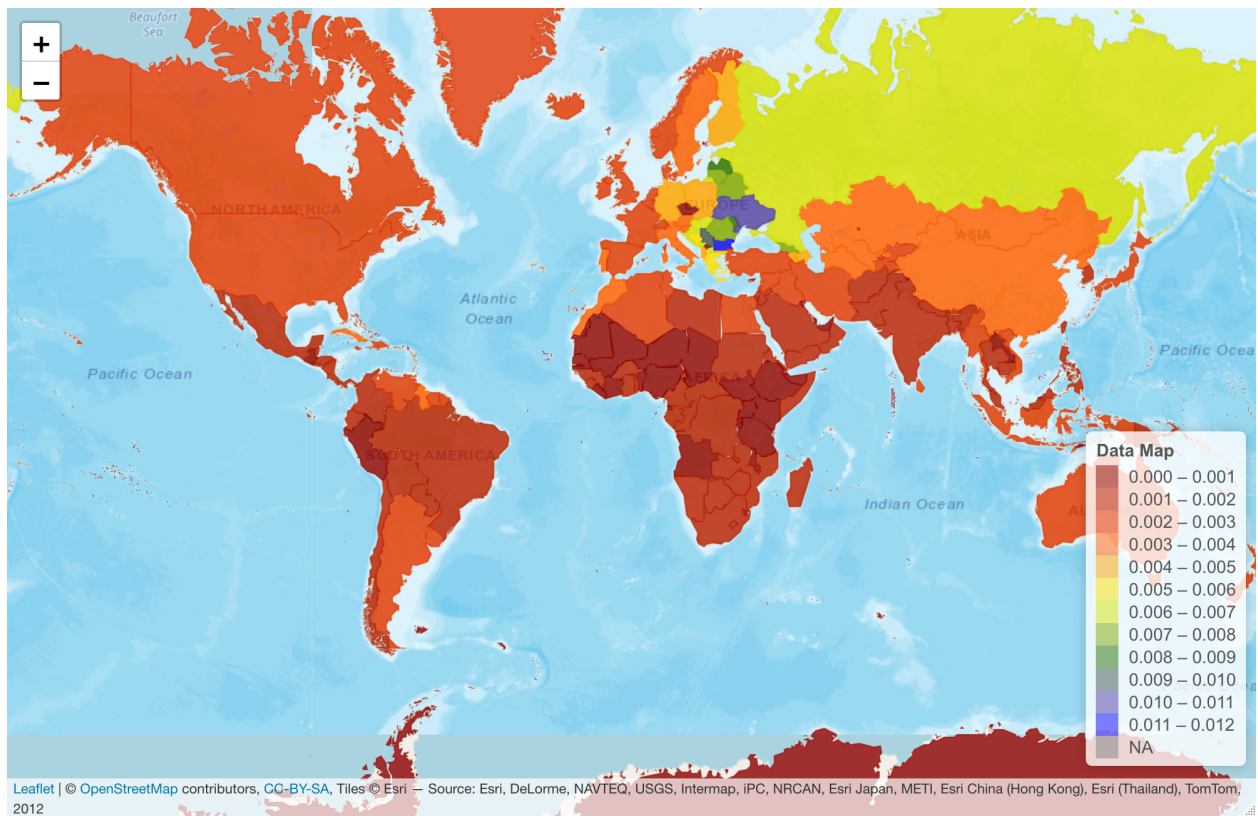


Figure 2: map