Mortality Rates and Causes Across the United States using

Subhalaxmi Rout

2021-03-14

Instruction

I have provided you with data about mortality from all 50 states and the District of Columbia. Please access it at https://github.com/charleyferrari/CUNY_DATA608/tree/master/lecture3/data

Question 1: As a researcher, you frequently compare mortality rates from particular causes across different States. You need a visualization that will let you see (for 2010 only) the crude mortality rate, across all States, from one cause (for example, Neoplasms, which are effectively cancers). Create a visualization that allows you to rank States by crude mortality for each cause of death.

Load libraries

```
library(ggplot2)
library(tidyverse)
library(dplyr)
library(stats)
library(DT)
```

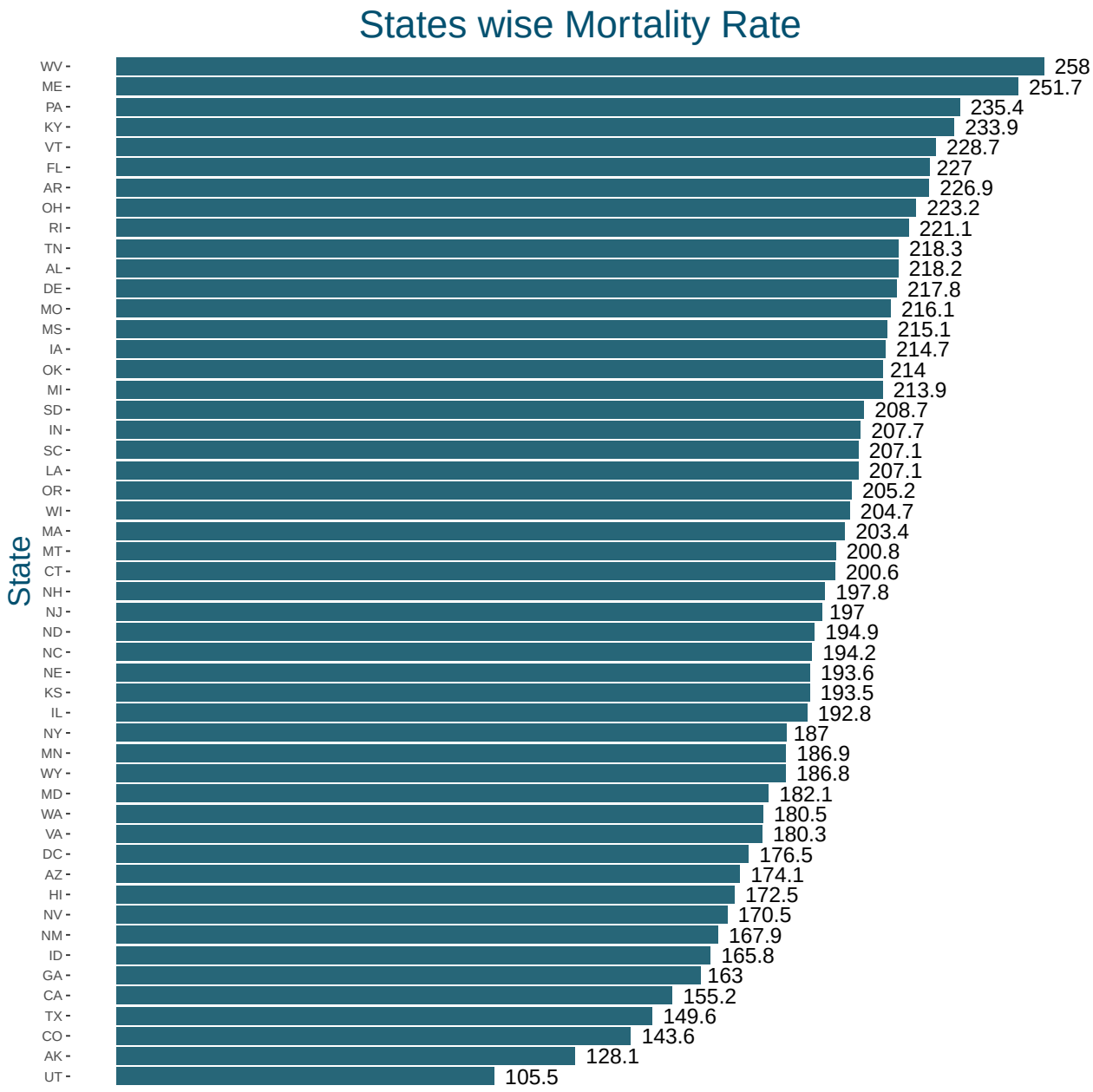
Load data from Git

```
df <- read.csv("https://raw.githubusercontent.com/charleyferrari/CUNY_DATA_608/master/module3/data/clea

head(df)
```

```
##                                ICD.Chapter State Year Deaths Population
## 1 Certain infectious and parasitic diseases    AL 1999   1092    4430141
## 2 Certain infectious and parasitic diseases    AL 2000   1188    4447100
## 3 Certain infectious and parasitic diseases    AL 2001   1211    4467634
## 4 Certain infectious and parasitic diseases    AL 2002   1215    4480089
## 5 Certain infectious and parasitic diseases    AL 2003   1350    4503491
## 6 Certain infectious and parasitic diseases    AL 2004   1251    4530729
##   Crude.Rate
## 1       24.6
## 2       26.7
## 3       27.1
## 4       27.1
## 5       30.0
## 6       27.6
```

The below plot shows state by crude rate filtered with year = 2010, and cause for death = Neoplasms.

```
df2 <- df %>% filter(Year == '2010' & ICD.Chapter == 'Neoplasms') %>% select(State, Crude.Rate)

df2 %>% group_by(State) %>% summarise(Crude.Rate = sum(Crude.Rate)) %>%
  ggplot() + aes(x = reorder(State, Crude.Rate), y = Crude.Rate) +
  ggtitle('States wise Mortality Rate') +
  xlab('State') +
  geom_bar(fill="#276678", stat = "identity") +
  coord_flip() + geom_text(aes(label = Crude.Rate), size = 5,  hjust=-0.20) +
  theme(panel.background = element_rect(fill = "white", color = NA),
        plot.title = element_text(hjust = 0.5, size = 25, colour = "#03506f"),
        axis.title.y = element_text(size = 20, colour = "#03506f"),
        axis.title.x=element_blank(),
        axis.text.x = element_blank(),
        axis.ticks.x=element_blank()
       )
```



Question 2: Often you are asked whether particular States are improving their mortality rates (per cause) faster than, or slower than, the national average. Create a visualization that lets your clients see this for themselves for one cause of death at the time. Keep in mind that the national average should be weighted by the national population.

```
df3 <- df  %>%  filter(ICD.Chapter == 'External causes of morbidity and mortality') %>% group_by(Year) %
  mutate(National.Crude.Rate = sum(Deaths) / sum(Population) * 100000)

df4 <- df %>% filter(ICD.Chapter == 'External causes of morbidity and mortality') %>% group_by(Year)

df5 <- cbind(df4, df3$National.Crude.Rate)

df5 <- df5 %>% filter(State == 'AK')
```

```
colnames(df5)[7] <- "National.Crude.Rate"

df5$National.Crude.Rate = round(df5$National.Crude.Rate,1)
DT::datatable(df5)
```

Show 10 entries Search:

	ICD.Chapter	State	Year	Deaths	Population	Crude.Rate	National.Crude.Rate
1	External causes of morbidity and mortality	AK	1999	474	624779	75.9	54.2
2	External causes of morbidity and mortality	AK	2000	540	626932	86.1	53.8
3	External causes of morbidity and mortality	AK	2001	500	633714	78.9	55.2
4	External causes of morbidity and mortality	AK	2002	546	642337	85	57.1
5	External causes of morbidity and mortality	AK	2003	511	648414	78.8	57.5
6	External causes of morbidity and mortality	AK	2004	544	659286	82.5	58.1
7	External causes of morbidity and mortality	AK	2005	502	666946	75.3	59.7
8	External causes of morbidity and mortality	AK	2006	510	675302	75.5	60.9
9	External causes of morbidity and mortality	AK	2007	580	680300	85.3	61.4
10	External causes of morbidity and mortality	AK	2008	583	687455	84.8	60.4

Showing 1 to 10 of 12 entries Previous 1 2 Next

```
df5 <- df5 %>% select(Year, Crude.Rate, National.Crude.Rate) %>% gather("Crude", "Value", -Year)
head(df5)
```

```
## # A tibble: 6 x 3
## # Groups:   Year [6]
##    Year Crude      Value
##   <int> <chr>      <dbl>
## 1  1999 Crude.Rate  75.9
## 2  2000 Crude.Rate  86.1
## 3  2001 Crude.Rate  78.9
## 4  2002 Crude.Rate  85
## 5  2003 Crude.Rate  78.8
## 6  2004 Crude.Rate  82.5
```

The below plot shows Yearly Crude Rate vs National Crude Rate filtered with State = AK, and cause for death = Certain infectious and parasitic diseases.

```
df5 %>%
  ggplot() + aes(x = Year, y = Value, fill = Crude) +
  geom_col(position = "dodge") +
  ggtitle('Crude Rate vs National Crude Rate') +
  xlab('Year') +
```

```
  ylab('Crude Rate') +
  scale_fill_manual(values = c("#276678", "#999999")) +
  geom_text(aes(label = Value), size = 4, position = position_dodge(width = 1), vjust = -.50) +
  theme(panel.background = element_rect(fill = "white", color = NA),
        plot.title = element_text(hjust = 0.5, size = 15, colour = "#03506f"),
        axis.title.y = element_text(size = 10, colour = "#03506f")
       )
```