Report on COVID19 Data

Praveen Myakala

2023-11-06

Report on COVID19 Data

Import, tidy and analyze the COVID19 dataset from the Johns Hopkins Github site. This is the same dataset I used in class. Feel free to repeat and reuse what I did if you want to. Be sure your project is a reproducible .rmd document which your peers can download and knit. It should contain some visualization and analysis that is unique to your project. You may use the data to do any analysis that is of interest to you. You should include at least two visualizations and one model. Be sure to identify any bias possible in the data and in your analysis.

```
# Download the COVID-19 deaths data

covid_deaths_original_data <- read_csv(
  "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_ti
```

Preapring Data

```
# Prepare the data for analysis
covid_deaths_data <- covid_deaths_original_data %>%
  rename(Country.Region = `Country/Region`,
  Province.State = `Province/State`) %>%
  pivot_longer(cols = -c(Province.State, Country.Region, Lat, Long),
               names_to = "date",
               values_to = "deaths") %>%
  # Convert the date column to a date object
  mutate(date = as.Date(date, format = "%m/%d/%y"),YEAR = year(date), MONTH = month(date)) %>%
  filter(!is.na(deaths)) %>%  # Remove rows with missing values in the deaths column
  mutate(deaths = as.numeric(deaths)) %>%  # Convert deaths column to numeric

  select(date, YEAR, MONTH, Country.Region, deaths) # Select the relevant columns
```

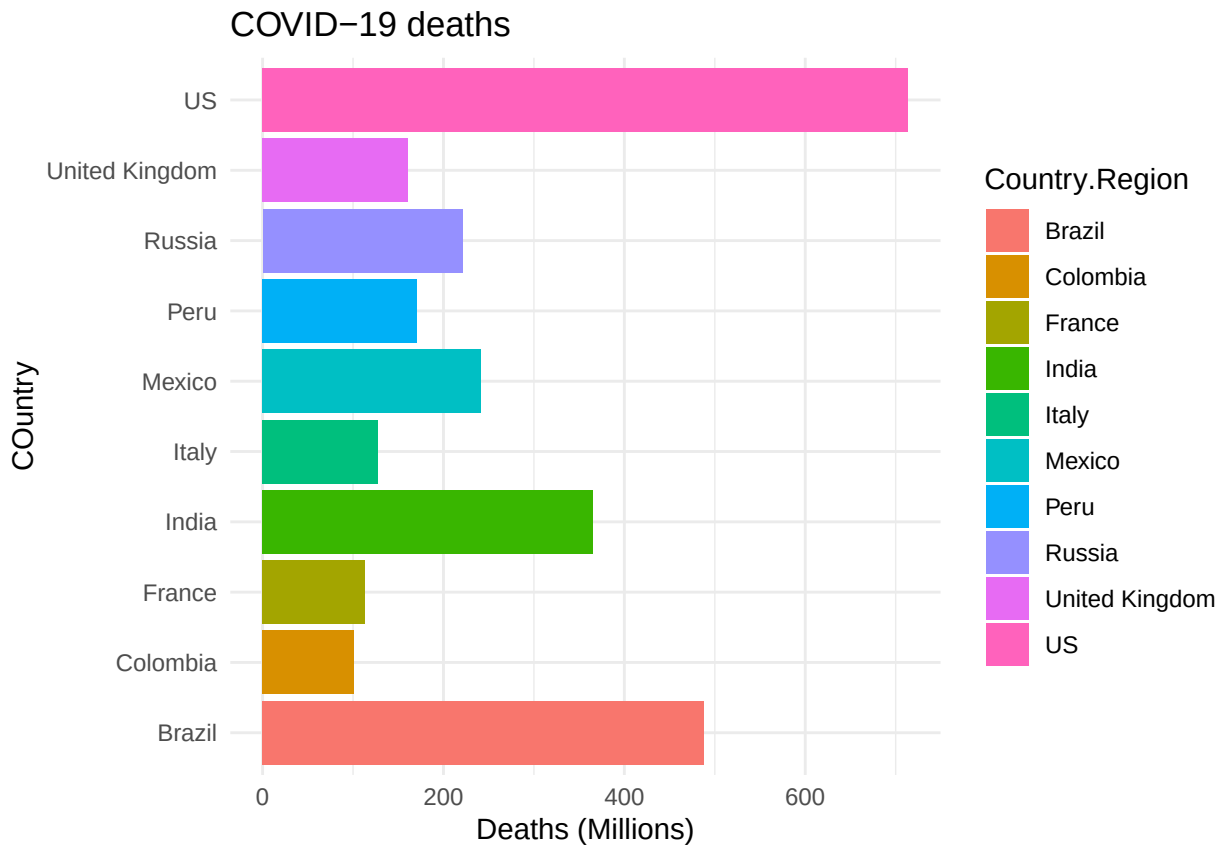
Top 10 Countries in Death

```
# TOP 10 Countries death wise

top_10_countries <- covid_deaths_data %>%
  group_by(Country.Region) %>%
  summarize(deaths = sum(deaths)) %>%
```

```
  arrange(desc(deaths)) %>%
  head(10)
```

```
# Create a line chart of number of COVID-19 deaths by country/region (top 10)
ggplot(top_10_countries, aes(x=deaths, y=Country.Region, fill=Country.Region)) +
  geom_bar(stat="identity") +
  scale_x_continuous(labels = scales::comma_format(scale = 1e-6)) +
  xlab("Deaths (Millions)") + ylab("COuntry") +
  ggtitle("COVID-19 deaths") +
  theme_minimal()
```




```
# Deaths in USA month wise
deaths_in_us_year_wise <- covid_deaths_data %>%
  filter(Country.Region == 'US') %>%
  group_by(YEAR, MONTH) %>%
  summarize(deaths = sum(deaths))
```

```
## `summarise()` has grouped output by 'YEAR'. You can override using the
## `.groups` argument.
```

```
deaths_in_us_year_wise$MONTH_YEAR <- paste0(month.abb[deaths_in_us_year_wise$MONTH], " ", deaths_in_us_

# Create a line chart of the daily number of COVID-19 deaths by country/region
```

```r
ggplot(deaths_in_us_year_wise, aes(y=MONTH_YEAR, x=deaths)) +
  geom_bar(stat="identity") +
  ylab("Month") + xlab("Deaths (Millions)") +
  scale_x_continuous(labels = scales::comma_format(scale = 1e-6)) +
  ggtitle("Deaths in USA - Month wise") +
  theme_minimal()
```