COVID-19

Lin HuiSheng

2024-05-15

##Import library

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(lubridate)
library(stringr)
library(readr)
library(gridExtra)
```

Import Data

Import COVID-19 data from Johns Hopkins Github site

```
url_in <- "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_cov
file_names <-c("time_series_covid19_confirmed_global.csv","time_series_covid19_deaths_global.csv","time_
urls <- str_c(url_in,file_names)

global_cases <- read_csv(urls[1])
global_deaths <- read_csv(urls[2])
us_cases <- read_csv(urls[3])
us_deaths <- read_csv(urls[4])
```

##Tidy global data

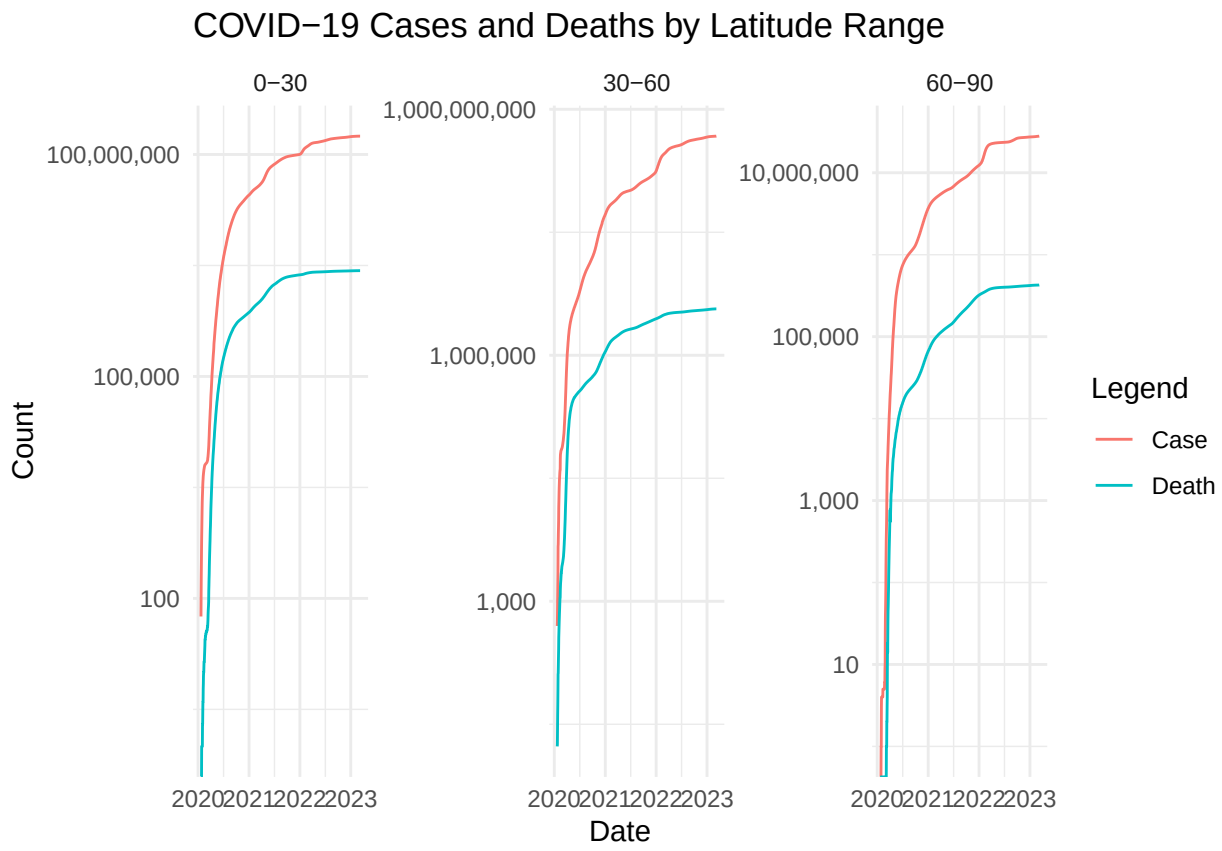
```
#global_cases
global_cases <- global_cases %>%
  pivot_longer(cols = -c(`Province/State`,
                         `Country/Region`,Lat,Long),
               names_to="date",
               values_to="cases")

#global_deaths
global_deaths <- global_deaths %>%
  pivot_longer(cols = -c(`Province/State`,
                         `Country/Region`,Lat,Long),
               names_to="date",
               values_to="deaths")

#combine global_cases & global_deaths
```

```r
global <- global_cases %>%
  full_join(global_deaths) %>%
  rename(Country_Region = `Country/Region`,
         Province_State = `Province/State`) %>%
  mutate(date = as.Date(date, format = "%m/%d/%y"))
```

##Data Segmentation and Initial Analysis First, we divided the data into three latitude groups: low (0-30), medium (30-60), and high (60-90). By analyzing the differences in death rates across these latitude ranges, we discovered that, irrespective of the region, the overall trend in death rates has been decreasing. However, it is notable that in high-latitude regions, the death rate remained relatively stable, with a slight upward trend observed between 2021 and 2022. In contrast, the low and medium latitude regions exhibited a consistent decline in death rates without significant increases.

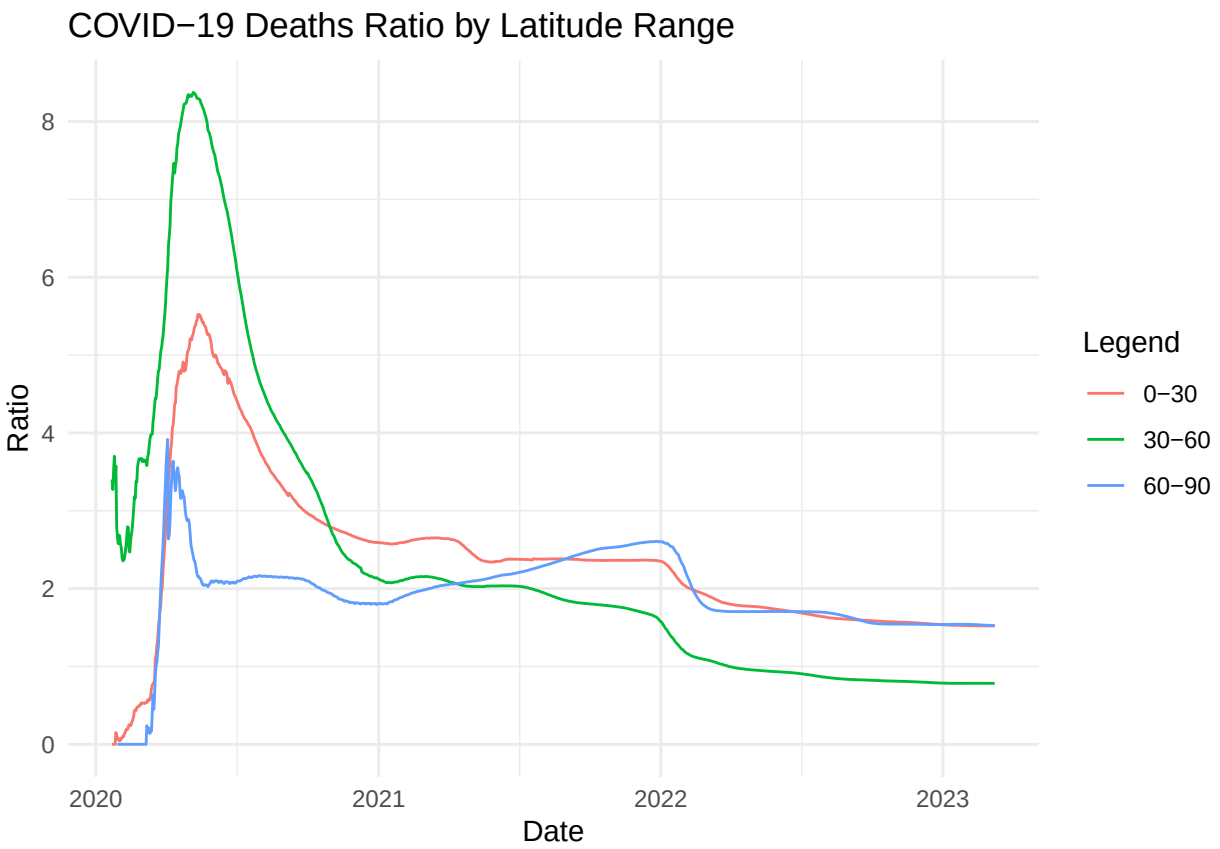
```r
global <- global %>%
  filter(!is.na(Lat))

lat_global <- global %>%
  mutate(latitude_group = case_when(
    between(Lat, 0, 30) ~ "0-30",
    between(Lat, -30, 0) ~ "0-30",
    between(Lat, 30, 60) ~ "30-60",
    between(Lat, -60, -30) ~ "30-60",
    between(Lat, 60, 90) ~ "60-90",
    between(Lat, -90, -60) ~ "60-90",
    TRUE ~ "Other"
  ))

# Calculating Total Cases and Deaths for Each Latitude Range
grouped_data <- lat_global %>%
  group_by(latitude_group, date) %>%
  summarize(total_cases = sum(cases, na.rm = TRUE),
            total_deaths = sum(deaths, na.rm = TRUE),
            ratio=total_deaths/total_cases*100)


# Visualizing the Relationship Between Confirmed Cases and Deaths
ggplot(grouped_data, aes(x = date)) +
  geom_line(aes(y = total_cases, color = "Case")) +
  geom_line(aes(y = total_deaths, color = "Death")) +
  scale_y_log10(labels = scales::comma) +
  facet_wrap(~latitude_group, scales = "free_y") +
  labs(title = "COVID-19 Cases and Deaths by Latitude Range",
       x = "Date",
       y = "Count",
       color = "Legend") +
  theme_minimal()
```

```
## Warning in scale_y_log10(labels = scales::comma): log-10 transformation introduced infinite values.
## log-10 transformation introduced infinite values.
```




```
ggplot(grouped_data, aes(x = date)) +
  geom_line(aes(y = ratio, color = latitude_group)) +
  labs(title = "COVID-19 Deaths Ratio by Latitude Range",
       x = "Date",
       y = "Ratio",
       color = "Legend") +
  theme_minimal()
```

```
## Warning: Removed 7 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

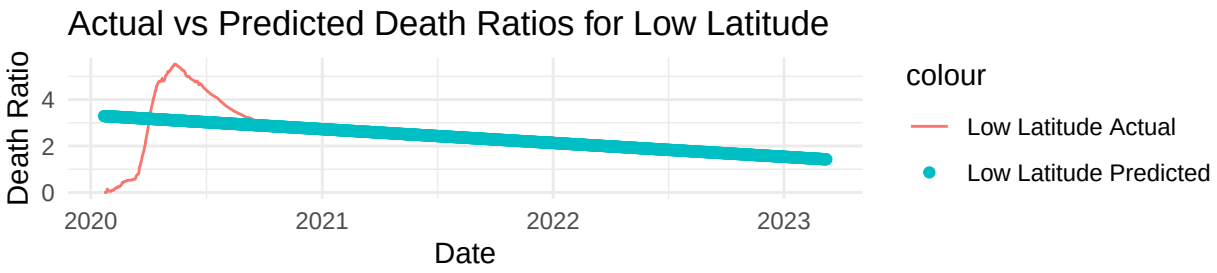


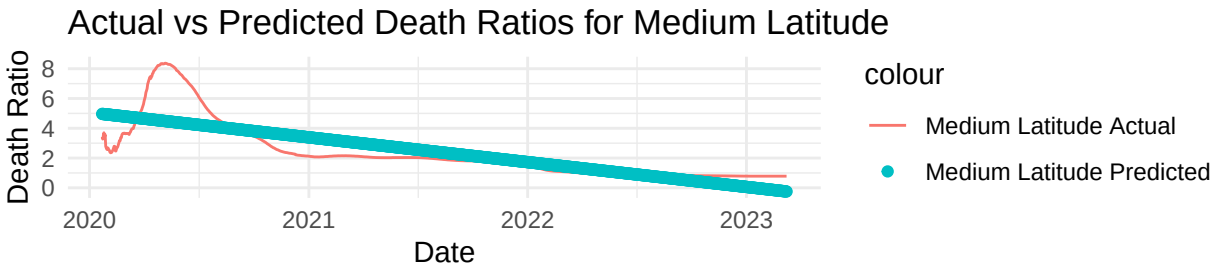
##Predictive Analysis for Each Latitude Group We conducted predictive analyses separately for each latitude group. The results indicated that the trends in low and medium latitude regions showed a declining pattern in death rates. On the other hand, the high-latitude region exhibited a relatively stable death rate. This observation prompted further investigation into whether this stability is related to the level of medical resources and healthcare infrastructure in high-latitude regions.

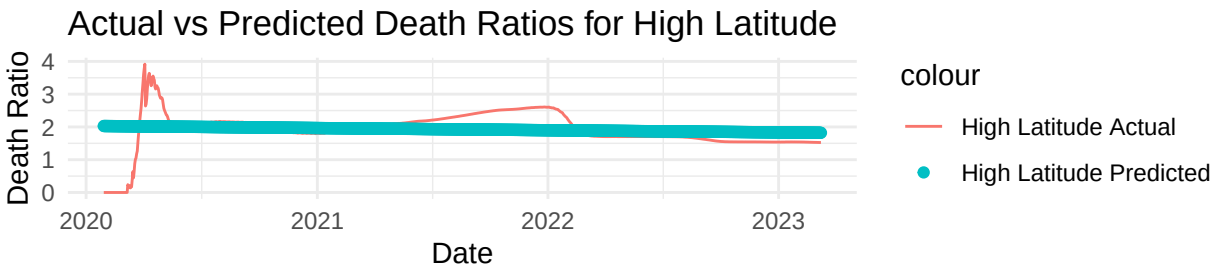
##Initial Predictive Analysis Including 2020 Data It is important to note that the initial surge in death rates globally during the outbreak in 2020 introduced bias into the predictive models. This surge does not accurately reflect the current ability of medical technology to control the pandemic. Therefore, to create a more accurate model, we considered excluding the data from 2020.

```
grouped_data <- grouped_data %>%
  filter(!is.na(ratio))

# pred latitude group
pred_low <- grouped_data %>%
  filter(latitude_group=="0-30")
pred_med <- grouped_data %>%
  filter(latitude_group=="30-60")
pred_high <- grouped_data %>%
  filter(latitude_group=="60-90")

mod <- lm(ratio ~ date , data = pred_low)
Global_low_pred <- pred_low %>% mutate(pred = predict(mod),group = "Low Latitude")

mod <- lm(ratio ~ date , data = pred_med)
Global_med_pred <- pred_med %>% mutate(pred = predict(mod),group = "Medium Latitude")
```

```r
mod <- lm(ratio ~ date , data = pred_high)
Global_high_pred <- pred_high %>% mutate(pred = predict(mod),group = "High Latitude")

plot_low <- ggplot(Global_low_pred)+
  geom_line(aes(x = date, y = ratio, color = "Low Latitude Actual")) +
  geom_point(aes(x = date, y = pred, color = "Low Latitude Predicted"))+
  labs(title = "Actual vs Predicted Death Ratios for Low Latitude",
       x = "Date",
       y = "Death Ratio") +
  theme_minimal()

plot_med <- ggplot(Global_med_pred)+
  geom_line(aes(x = date, y = ratio, color = "Medium Latitude Actual")) +
  geom_point(aes(x = date, y = pred, color = "Medium Latitude Predicted")) +
  labs(title = "Actual vs Predicted Death Ratios for Medium Latitude",
       x = "Date",
       y = "Death Ratio") +
  theme_minimal()

plot_high <- ggplot(Global_high_pred)+
  geom_line(aes(x = date, y = ratio, color = "High Latitude Actual")) +
  geom_point(aes(x = date, y = pred, color = "High Latitude Predicted")) +
  labs(title = "Actual vs Predicted Death Ratios for High Latitude",
       x = "Date",
       y = "Death Ratio") +
  theme_minimal()

grid.arrange(plot_low, plot_med, plot_high, ncol = 1)
```



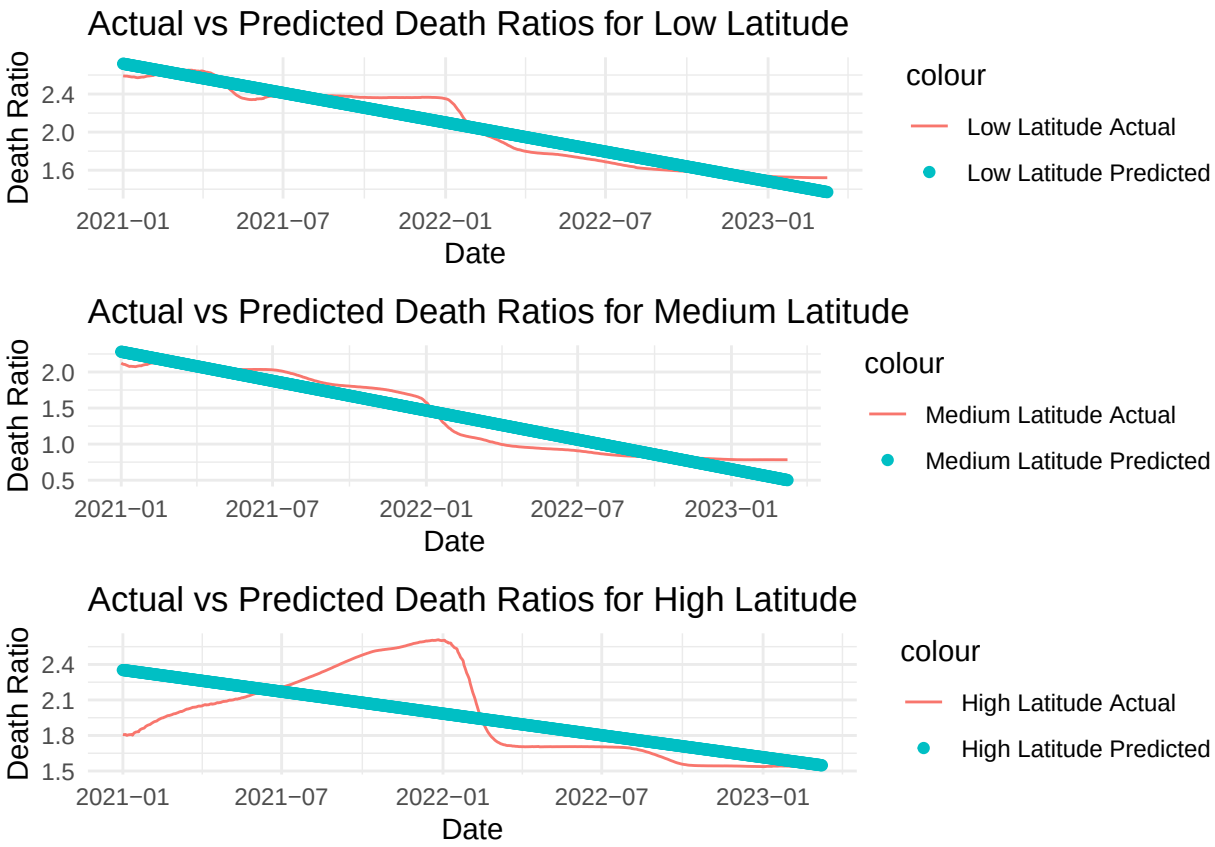




##Predictive Analysis Excluding 2020 Data After excluding the data from 2020 and reanalyzing the models, we found that the death rates showed a clear downward trend. In high-latitude regions, there was an upward trend in death rates in 2021, followed by a sharp decline in 2022. The low and medium latitude regions continued to show a stable decline. However, since 2023, death rates have stabilized, suggesting a trend towards pandemic stabilization. Whether death rates can be further reduced depends on advancements in medical technology and vaccine development.

```
# pred exclude 2020
pred_low <- pred_low %>%
  filter(date>="2021-01-01")
pred_med <- pred_med %>%
  filter(date>="2021-01-01")
pred_high <- pred_high %>%
  filter(date>="2021-01-01")

mod <- lm(ratio ~ date , data = pred_low)
Global_low_pred <- pred_low %>% mutate(pred = predict(mod),group = "Low Latitude")

mod <- lm(ratio ~ date , data = pred_med)
Global_med_pred <- pred_med %>% mutate(pred = predict(mod),group = "Medium Latitude")

mod <- lm(ratio ~ date , data = pred_high)
Global_high_pred <- pred_high %>% mutate(pred = predict(mod),group = "High Latitude")

plot_low <- ggplot(Global_low_pred)+
  geom_line(aes(x = date, y = ratio, color = "Low Latitude Actual")) +
```

```r
  geom_point(aes(x = date, y = pred, color = "Low Latitude Predicted"))+
  labs(title = "Actual vs Predicted Death Ratios for Low Latitude",
       x = "Date",
       y = "Death Ratio") +
  theme_minimal()

plot_med <- ggplot(Global_med_pred)+
  geom_line(aes(x = date, y = ratio, color = "Medium Latitude Actual")) +
  geom_point(aes(x = date, y = pred, color = "Medium Latitude Predicted")) +
  labs(title = "Actual vs Predicted Death Ratios for Medium Latitude",
       x = "Date",
       y = "Death Ratio") +
  theme_minimal()

plot_high <- ggplot(Global_high_pred)+
  geom_line(aes(x = date, y = ratio, color = "High Latitude Actual")) +
  geom_point(aes(x = date, y = pred, color = "High Latitude Predicted")) +
  labs(title = "Actual vs Predicted Death Ratios for High Latitude",
       x = "Date",
       y = "Death Ratio") +
  theme_minimal()

grid.arrange(plot_low, plot_med, plot_high, ncol = 1)
```



##Conclusion The analysis of COVID-19 death rates across different latitude groups reveals significant insights:

Overall Decline in Death Rates: Across all latitude groups, the death rates have generally been decreasing. Stable Death Rates in High Latitudes: High-latitude regions showed relatively stable death rates, with a slight increase observed between 2021 and 2022, followed by a decline. Consistent Decline in Low and Medium Latitudes: Low and medium latitude regions exhibited a consistent and stable decline in death rates. Impact of Excluding 2020 Data: Excluding the initial 2020 data provided a clearer picture of the current trends and the effectiveness of medical interventions. The findings indicate that the pandemic is becoming more controlled, but the future reduction in death rates will depend significantly on the continued progress in medical technology and vaccine development.

By focusing on these trends and refining our predictive models, we can better understand and mitigate the impact of the COVID-19 pandemic across different regions and latitudes.