

COVID19 Analysis

P.Papa

2024-10-09

Introduction

This project aims to analyze the mortality rate of COVID-19 over time for different countries, with a specific focus on Italy. We will use the COVID-19 dataset provided by the Johns Hopkins University.

Data Import and Tidying

Importing the data

```
# Get all data
url_in <- ("https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_co

file_names <-
  c("time_series_covid19_confirmed_global.csv",
    "time_series_covid19_deaths_global.csv",
    "time_series_covid19_confirmed_US.csv",
    "time_series_covid19_deaths_US.csv")

# Concatenate data
urls <- str_c(url_in, file_names)
```

```
# Store each dataset in a variable.
confirmed_global <- read_csv(urls[1])
deaths_global <- read_csv(urls[2])
us_cases_death <- read_csv(urls[3])
us_cases <- read_csv(urls[4])
```

Tidying the data

```
# Transform confirmed cases
confirmed_long <- confirmed_global %>%
  pivot_longer(
    cols = matches("^\\d"),
    names_to = "Date",
    values_to = "Confirmed"
  ) %>%
  mutate(Date = mdy(Date))
```

```
# Transform deaths
deaths_long <- deaths_global %>%
  pivot_longer(
    cols = matches("^\\d"),
    names_to = "Date",
    values_to = "Deaths"
  ) %>%
  mutate(Date = mdy(Date))
```

Merging data

I merged the confirmed cases and death cases into a single table.

```
covid_data <- confirmed_long %>%
  left_join(deaths_long %>% select(-c(Lat, Long)),
    by = c("Province/State", "Country/Region", "Date"))

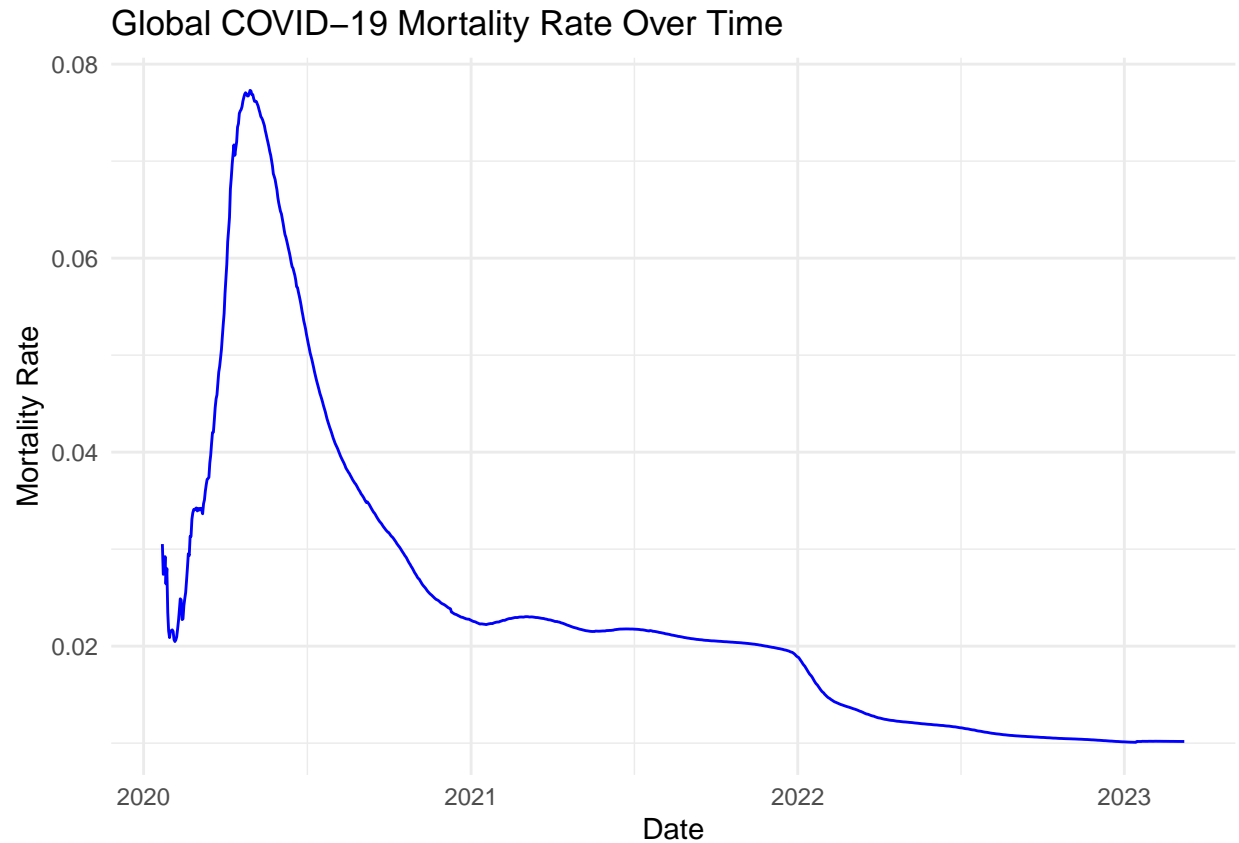
covid_data <- covid_data %>%
  mutate(MortalityRate = Deaths / Confirmed)
```

Visualization

Global mortality rate over time

```
global_mortality <- covid_data %>%
  group_by(Date) %>%
  summarize(
    TotalDeaths = sum(Deaths, na.rm = TRUE),
    TotalConfirmed = sum(Confirmed, na.rm = TRUE)
  ) %>%
  mutate(MortalityRate = TotalDeaths / TotalConfirmed)

ggplot(global_mortality, aes(x = Date, y = MortalityRate)) +
  geom_line(color = "blue") +
  labs(title = "Global COVID-19 Mortality Rate Over Time",
    x = "Date",
    y = "Mortality Rate") +
  theme_minimal()
```



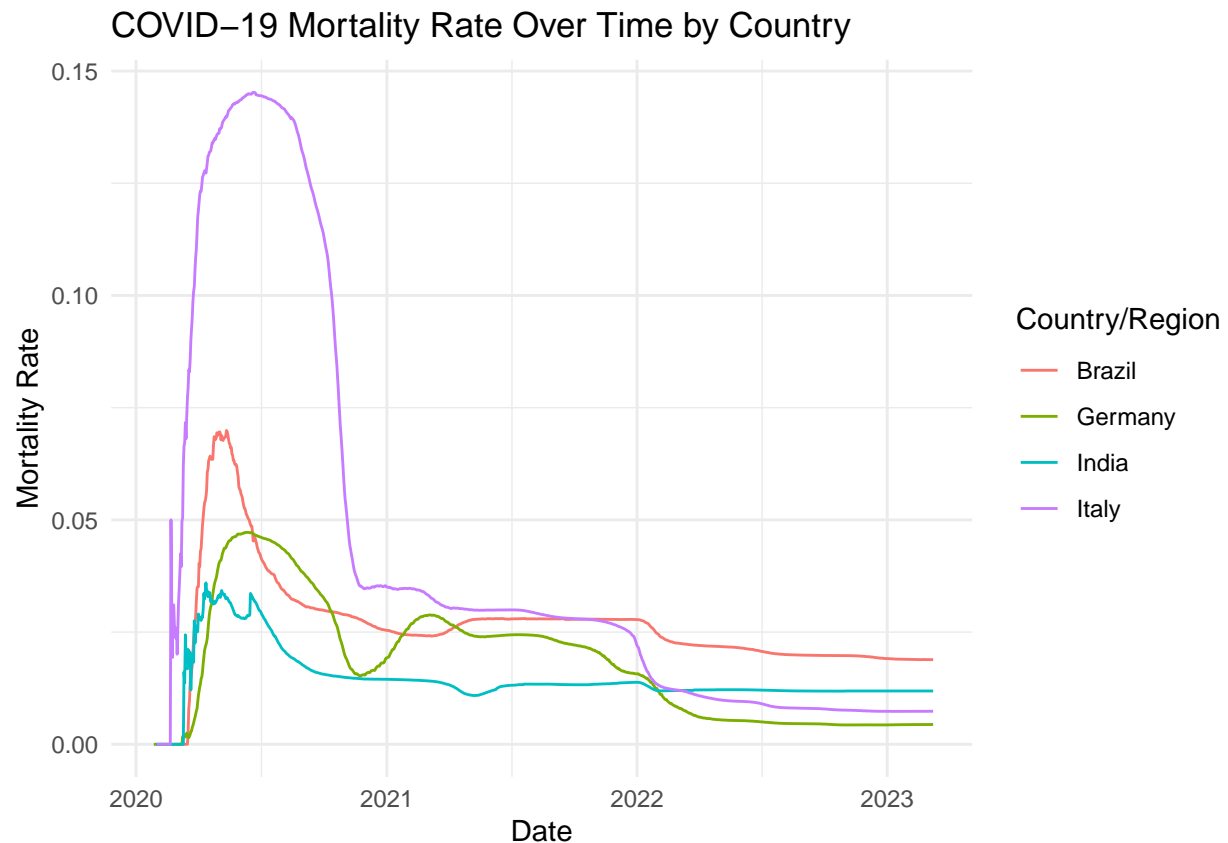
Mortality rate by country I selected some countries from around the world for comparison.

```
countries_of_interest <- c("Italy", "United States", "Brazil", "India", "Germany", "UK")

country_mortality <- covid_data %>%
  filter(`Country/Region` %in% countries_of_interest) %>%
  group_by(`Country/Region`, Date) %>%
  summarize(
    TotalDeaths = sum(Deaths, na.rm = TRUE),
    TotalConfirmed = sum(Confirmed, na.rm = TRUE)
  ) %>%
  mutate(MortalityRate = TotalDeaths / TotalConfirmed)

ggplot(country_mortality, aes(x = Date, y = MortalityRate, color = `Country/Region`)) +
  geom_line() +
  labs(title = "COVID-19 Mortality Rate Over Time by Country",
       x = "Date",
       y = "Mortality Rate") +
  theme_minimal()
```

```
## Warning: Removed 57 rows containing missing values or values outside the scale range
## ('geom_line()').
```



Focus on Italy

Italy experienced a high mortality rate despite its well-regarded healthcare system.

Key dates

I will analyze the data considering key dates for Italy:

- Lockdown Start: March 9, 2020
- Reopening: May 18, 2020
- Mass Vaccine Rollout: March 1, 2021 (first doses administered at the end of 2020, but the effective vaccination campaign began in March 2021)
- Arrival of the Less Severe Omicron Variant: Around December 23, 2021 (28% prevalence in Italy based on global data).

Analysis of Italy's Data

```
italy_data <- covid_data %>%
  filter(`Country/Region` == "Italy") %>%
  group_by(Date) %>%
```

```

summarize(
  TotalDeaths = sum(Deaths, na.rm = TRUE),
  TotalConfirmed = sum(Confirmed, na.rm = TRUE)
) %>%
mutate(
  MortalityRate = TotalDeaths / TotalConfirmed,
  Lockdown = case_when(
    Date >= as.Date("2020-03-09") & Date < as.Date("2020-05-18") ~ "During Lockdown",
    Date >= as.Date("2020-05-18") & Date < as.Date("2020-12-27") ~ "Post Lockdown",
    Date >= as.Date("2021-03-01") ~ "Mass Vaccine Rollout",
    TRUE ~ "Pre Lockdown"
  )
)

```

Plotting Italy's Mortality Rate with Key Dates

```

# Define key dates with labels and positions
key_dates <- data.frame(
  Date = as.Date(c("2020-03-09", "2020-05-18", "2021-03-01", "2021-12-23")),
  Event = c("Lockdown Start", "Reopening", "Mass Vaccine Rollout Start", "Arrival of Less Severe Variants")
)

# Adjust the plotting code
ggplot(italy_data, aes(x = Date, y = MortalityRate)) +
  geom_line(color = "red", size = 1) +
  # Add vertical lines for key dates
  geom_vline(data = key_dates, aes(xintercept = Date), linetype = "dashed", color = "blue") +
  # Add text labels for key dates
  geom_text(
    data = key_dates,
    aes(x = Date, y = max(italy_data$MortalityRate, na.rm = TRUE) * 0.95, label = Event),
    angle = 90,
    vjust = -0.4,
    hjust = 1,
    size = 3,
    color = "blue"
  ) +
  labs(
    title = "COVID-19 Mortality Rate in Italy with Key Dates",
    x = "Date",
    y = "Mortality Rate"
  ) +
  scale_x_date(date_breaks = "3 months", date_labels = "%b %Y") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(face = "bold", size = 14)
  )

```

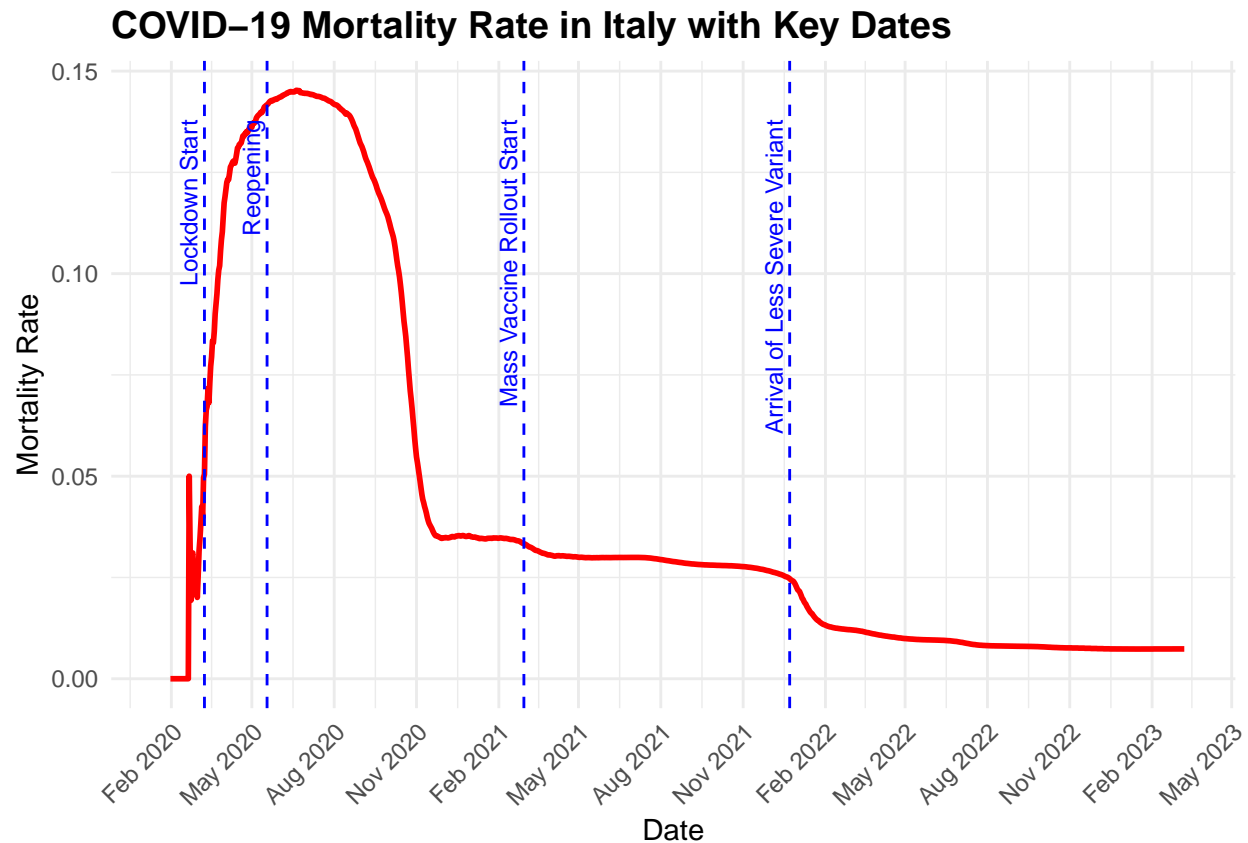
```

## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.

```

```
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

## Warning: Removed 9 rows containing missing values or values outside the scale range
## ('geom_line()').
```



Modeling

I created a model to track the impact of the lockdown, the vaccine rollout, and the Omicron variant.

```
# Update the italy_model_data with new periods
italy_model_data <- italy_data %>%
  mutate(
    Period = case_when(
      Date < as.Date("2020-03-09") ~ "Pre-Lockdown",
      Date >= as.Date("2020-03-09") & Date < as.Date("2020-05-18") ~ "Lockdown",
      Date >= as.Date("2020-05-18") & Date < as.Date("2021-03-01") ~ "Post-Lockdown",
      Date >= as.Date("2021-03-01") & Date < as.Date("2021-12-23") ~ "Vaccine Rollout",
      Date >= as.Date("2021-12-23") ~ "Less Severe Variant"
    )
  )

italy_model_data$Period <- factor(italy_model_data$Period, levels = c("Pre-Lockdown", "Lockdown", "Post-Lockdown", "Vaccine Rollout", "Less Severe Variant"))
```

```

model <- lm(MortalityRate ~ Lockdown, data = italy_model_data)

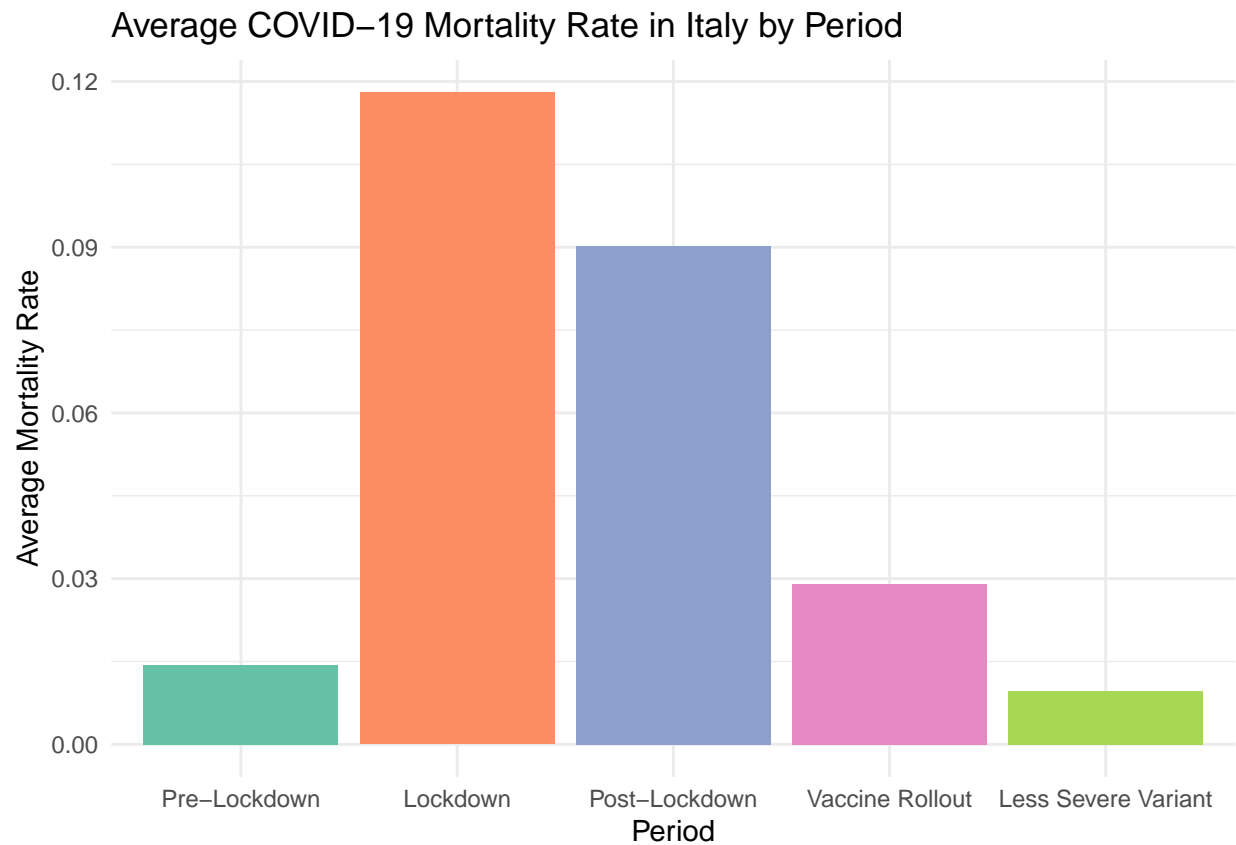
model_period <- lm(MortalityRate ~ Period, data = italy_model_data)
summary(model_period)

##
## Call:
## lm(formula = MortalityRate ~ Period, data = italy_model_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.067563 -0.002336 -0.000289  0.002686  0.055082
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.014393   0.004153   3.466 0.000549 ***
## PeriodLockdown      0.103649   0.005159  20.091 < 2e-16 ***
## PeriodPost-Lockdown  0.075777   0.004420  17.145 < 2e-16 ***
## PeriodVaccine Rollout  0.014661   0.004411   3.324 0.000917 ***
## PeriodLess Severe Variant -0.004698   0.004328  -1.085 0.277934
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0256 on 1129 degrees of freedom
## (9 observations deleted due to missingness)
## Multiple R-squared:  0.6852, Adjusted R-squared:  0.6841
## F-statistic: 614.3 on 4 and 1129 DF,  p-value: < 2.2e-16

# Calculate average mortality rate per period
avg_mortality <- italy_model_data %>%
  group_by(Period) %>%
  summarize(AverageMortalityRate = mean(MortalityRate, na.rm = TRUE))

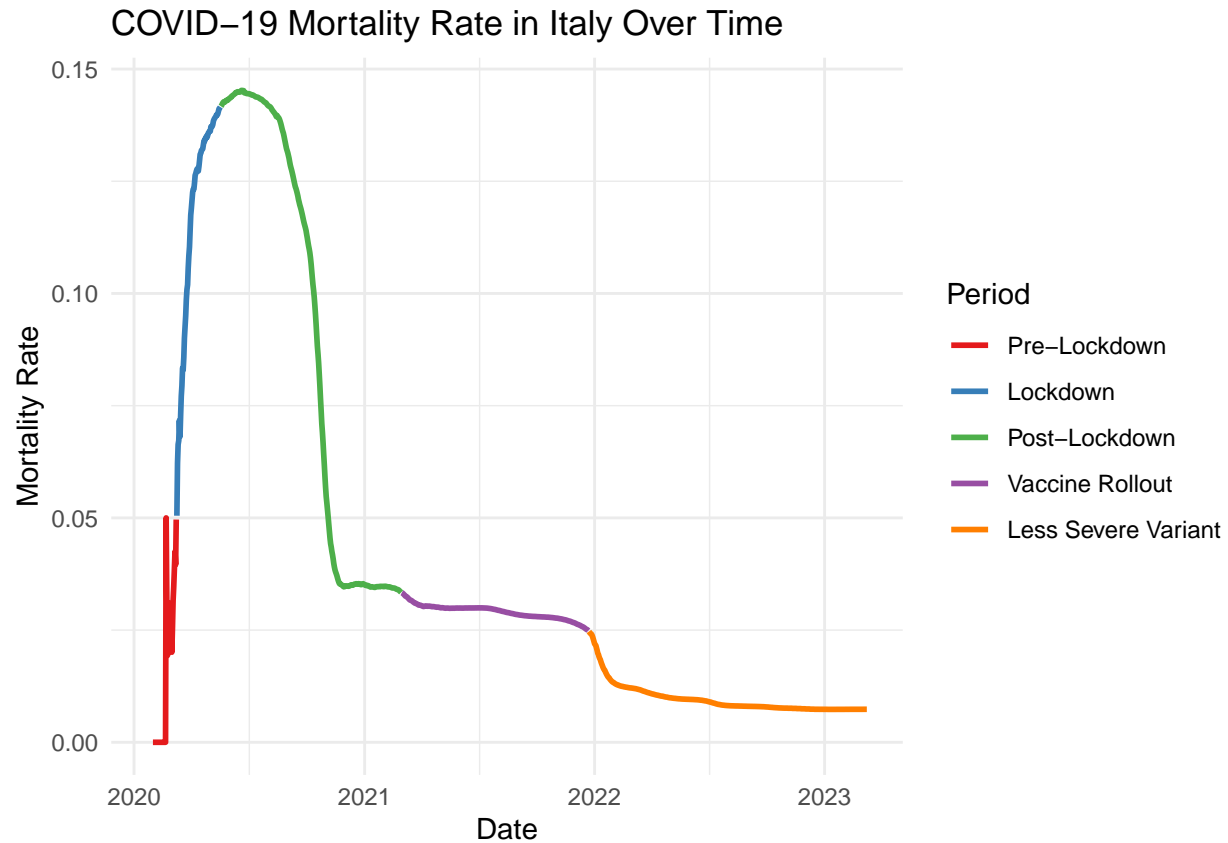
# Bar plot of average mortality rate by period
ggplot(avg_mortality, aes(x = Period, y = AverageMortalityRate, fill = Period)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Average COVID-19 Mortality Rate in Italy by Period",
    x = "Period",
    y = "Average Mortality Rate"
  ) +
  scale_fill_brewer(palette = "Set2") +
  theme_minimal() +
  theme(legend.position = "none")

```



```
ggplot(italy_model_data, aes(x = Date, y = MortalityRate, color = Period)) +
  geom_line(size = 1) +
  labs(
    title = "COVID-19 Mortality Rate in Italy Over Time",
    x = "Date",
    y = "Mortality Rate"
  ) +
  scale_color_brewer(palette = "Set1") +
  theme_minimal()
```

```
## Warning: Removed 9 rows containing missing values or values outside the scale range
## ('geom_line()').
```

Bias

Potential biases in the data and analysis:

- **Underreporting of Cases:** Limited testing capacity, especially in the early stages, may have led to underreporting of confirmed cases, affecting the mortality rate calculation.
- **Variations in Reporting Standards:** Different countries may have different criteria for reporting COVID-19 cases and deaths.
- **Time Lag:** Delays between infection, symptom onset, testing, and reporting can affect the accuracy of the data.
- **Data Gaps:** Missing data or inconsistencies in data collection methods over time.
- **The data does not include vaccination rates or percentages**

Conclusions

The lockdown and post-lockdown periods were marked by high mortality rates (10.36% and 7.58%, respectively). This was largely due to the overwhelmed healthcare system and the ongoing challenges in controlling the pandemic.

During the vaccine rollout, the mortality rate dropped significantly to 1.47%, reflecting the positive impact of vaccination efforts. This reduction is statistically significant compared to earlier periods.

With the emergence of the less severe Omicron variant, the mortality rate further decreased to 0.47%. However, this change may not be statistically significant, as high vaccination coverage and improved treatment protocols likely played a major role in mitigating severity.

References

- Johns Hopkins University Center. Data source: <https://github.com/CSSEGISandData/COVID-19>

Italian Government Official Announcements and Decrees

- Lockdown Start (March 9, 2020): Presidenza del Consiglio dei Ministri. (2020). Decreto del Presidente del Consiglio dei Ministri 9 marzo 2020. Available at: http://www.governo.it/sites/new.governo.it/files/DPCM_20200309.pdf
- Reopening (May 18, 2020): Presidenza del Consiglio dei Ministri. (2020). Decreto del Presidente del Consiglio dei Ministri 17 maggio 2020. Available at: http://www.governo.it/sites/new.governo.it/files/DPCM_20200517.pdf
- Vaccine Rollout Start (December 27, 2020): Italian Ministry of Health. (2020). Start of the COVID-19 Vaccination Campaign in Italy. Available at: <http://www.salute.gov.it>
- Arrival of Less Severe Variant: Istituto Superiore di Sanità (ISS). (2021). SARS-CoV-2 Variants Monitoring Reports. Available at: <https://www.iss.it/variants>
- Omicron coverage (28% on 23 December 2021): <https://www.ilsole24ore.com/art/covid-omicron-italia-avanza-rapidamente-ma-macchia-leopardo-e-282percento-AEkR1N4>
- Project available: <https://github.com/derfel83/COVID19-Italy>