Unraveling Covid-19 Dynamics in the UK: Insights from Multivariate Analysis

Vinothini Balasubramani/22202390, Sreelekshmi Girijadevi Purushothaman/22200034

Vaccination Impact: Modeling Infections, Hospitalization and Mortality

Performing Multivariate linear Regression on Covid-19 Pre Vaccination period data and Vaccination period data to analyse the impact of vaccine and other factors on the infections, hospitalization and mortality rate.

Package and Data Loading:

Loading the required packages:

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(cowplot)
library(tidyverse)
library(caret)
library(car)
library(gganimate)
library(animation)
library(transformr)
library(ggthemes)
library(gridExtra)
```

Loading the data-sets:

```
pre_vaccination_data <- read.csv("UK Pre vaccination period.csv")
vaccination_data <- read.csv("UK Vaccination period.csv")
```

Data Preprocessing:

Data cleaning:

```r
# Replace NA values with zeros
pre_vaccination_data[] <- lapply(pre_vaccination_data, function(x)
                                   replace(x, is.na(x), 0))

vaccination_data[] <- lapply(vaccination_data, function(x)
                               replace(x, is.na(x), 0))
```

Data formatting:

```r
columns_to_format <- c("average_temperature_celsius",
                       "minimum_temperature_celsius",
                       "maximum_temperature_celsius", "rainfall_mm",
                       "dew_point", "relative_humidity")

# Rounding off the columns
pre_vaccination_data[, columns_to_format] <- round(pre_vaccination_data[,
vaccination_data[, columns_to_format] <- round(vaccination_data[,
```

Exploratory Data Analysis:

In this analysis, we will explore and visualize the COVID-19 case data for the United Kingdom during the pre-vaccination and vaccination periods. The data-sets contain various variables related to COVID-19 cases, deaths hospitalizations, vaccinations, mobility, weather, and government policies.

Data structure and summary statistics:

Examining the structure of the data-sets to gain insights about the variables and computing the summary statistics to get an overview about the distribution of data.

```r
# Display the structure and first few rows of the data-sets
str(pre_vaccination_data)
```

```
## 'data.frame':    286 obs. of  28 variables:
##  $ date                          : chr  "3/31/2020" "4/1/2020" "4/2/2020" "4/3/2020" ...
##  $ new_confirmed                 : num  4463 4891 4824 4853 3976 ...
##  $ new_deceased                  : num  662 769 805 840 906 ...
##  $ new_tested                    : num  11896 11947 13623 14629 16080 ...
##  $ cumulative_confirmed          : num  38025 42916 47740 52593 56569 ...
##  $ cumulative_deceased           : num  4434 5203 6008 6848 7754 ...
##  $ cumulative_tested             : num  155174 167237 182352 223578 239658 ...
##  $ new_hospitalized_patients     : num  2834 3565 3284 2932 2922 ...
##  $ cumulative_hospitalized_patients : num  23549 27114 30398 33330 36252 ...
##  $ current_hospitalized_patients : num  12597 13637 13875 15546 17501 ...
##  $ current_ventilator_patients   : num  0 0 1813 2120 2309 ...
##  $ new_persons_vaccinated        : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ cumulative_persons_vaccinated    : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ new_persons_fully_vaccinated     : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ cumulative_persons_fully_vaccinated: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ mobility_retail_and_recreation   : num  -74 -75 -75 -76 -81 -82 -73 -73 -73 -71 ...
##  $ mobility_grocery_and_pharmacy    : num  -30 -32 -32 -30 -36 -41 -31 -29 -27 -20 ...
##  $ mobility_parks                   : num  -21 -29 -24 -25 -38 -29 -20 -7 -14 -13 ...
##  $ mobility_transit_stations        : num  -72 -73 -73 -72 -72 -70 -72 -69 -70 -70 ...
##  $ mobility_workplaces              : num  -70 -69 -70 -69 -57 -54 -70 -70 -70 -70 ...
##  $ mobility_residential             : num  27 28 29 29 19 15 27 27 27 27 ...
##  $ stringency_index                 : num  79.6 79.6 79.6 79.6 79.6 ...
##  $ average_temperature_celsius      : num  5.42 6.17 7.12 5.27 7.62 ...
##  $ minimum_temperature_celsius      : num  2.17 3.29 2.64 1.67 3.89 3.44 4.9 1.19 2.17 4.87 ...
##  $ maximum_temperature_celsius      : num  7.83 8.78 10.46 8.38 10.26 ...
##  $ rainfall_mm                      : num  0.13 0 0.19 0.17 0.76 0 1.37 0 0 0.03 ...
##  $ dew_point                        : num  1.91 1.91 2.13 1.21 3.85 3.19 4.01 2.63 4.05 5.84 ...
##  $ relative_humidity                : num  78.3 74.7 71.4 75.9 77.6 ...
```

```r
str(vaccination_data)
```

```
## 'data.frame':    475 obs. of  28 variables:
##  $ date                             : chr  "1/11/2021" "1/12/2021" "1/13/2021" "1/14/2021" ...
##  $ new_confirmed                    : num  58592 50647 46509 43473 41451 ...
##  $ new_deceased                     : num  1132 1213 1219 1233 1229 ...
##  $ new_tested                       : num  543719 590404 629223 696927 599603 ...
##  $ cumulative_confirmed             : num  3280627 3331274 3377783 3421256 3462707 ...
##  $ cumulative_deceased              : num  86933 88146 89365 90598 91827 ...
##  $ cumulative_tested                : num  61176472 61797627 62445325 63151470 63760165 ...
##  $ new_hospitalized_patients        : num  4310 4580 4225 4100 3684 ...
##  $ cumulative_hospitalized_patients : num  331827 336407 340632 344732 348416 ...
##  $ current_hospitalized_patients    : num  36753 36983 37489 37693 38134 ...
##  $ current_ventilator_patients      : num  3363 3496 3626 3672 3789 ...
##  $ new_persons_vaccinated           : num  145076 207661 278943 316694 341317 ...
##  $ cumulative_persons_vaccinated    : num  2431648 2639309 2918252 3234946 3576263 ...
##  $ new_persons_fully_vaccinated     : num  20768 16065 9745 5257 4208 ...
##  $ cumulative_persons_fully_vaccinated: num  412167 428232 437977 443234 447442 ...
##  $ mobility_retail_and_recreation   : num  -59 -61 -64 -66 -64 -70 -67 -60 -63 -65 ...
##  $ mobility_grocery_and_pharmacy    : num  -25 -24 -27 -29 -23 -26 -27 -25 -27 -28 ...
##  $ mobility_parks                   : num  -16 -7 -30 -28 -15 -31 8 -10 -23 -35 ...
##  $ mobility_transit_stations        : num  -63 -63 -65 -66 -64 -67 -63 -63 -64 -66 ...
##  $ mobility_workplaces              : num  -50 -51 -50 -53 -51 -35 -31 -51 -52 -51 ...
##  $ mobility_residential             : num  20 20 21 22 21 15 11 19 20 21 ...
##  $ stringency_index                 : num  88 88 88 88 88 ...
##  $ average_temperature_celsius      : num  6.38 0.12 -0.11 0.31 -2.31 2.72 3.53 4.49 6.03 5.62 ...
##  $ minimum_temperature_celsius      : num  3.36 -3.03 -3.56 -1.04 -5.48 -1.81 1.23 3.1 3.66 0.78 .
##  $ maximum_temperature_celsius      : num  8.43 4.41 1.83 1.32 0.06 5.87 5.04 5.54 7.66 7.93 ...
##  $ rainfall_mm                      : num  0.87 3.17 0 9.78 4.57 ...
##  $ dew_point                        : num  4.38 -1.8 -1.02 -0.13 -3.08 1.42 1.59 2.63 5.88 5.08 ..
##  $ relative_humidity                : num  87.4 87.2 93.7 96.9 94.5 ...
```

```r
# Compute summary statistics
summary_pre_vacc <- summary(pre_vaccination_data)
summary_vacc <- summary(vaccination_data)
```

```r
# Display the summary statistics
cat("Pre-Vaccination Data - Summary statistics:\n")
```

Pre-Vaccination Data - Summary statistics:

```r
print(summary_pre_vacc)
```

```
##      date            new_confirmed    new_deceased       new_tested
##  Length:286         Min.   :  366   Min.   :   4.00   Min.   : 11896
##  Class :character   1st Qu.: 1158   1st Qu.:  26.25   1st Qu.:103248
##  Mode  :character   Median : 3954   Median : 189.00   Median :183249
##                     Mean   :11149   Mean   : 286.81   Mean   :207692
##                     3rd Qu.:17028   3rd Qu.: 465.00   3rd Qu.:295995
##                     Max.   :83090   Max.   :1099.00   Max.   :682322
##  cumulative_confirmed cumulative_deceased cumulative_tested
##  Min.   :  38025      Min.   : 4434       Min.   :  155174
##  1st Qu.: 268337      1st Qu.:39529       1st Qu.: 4979336
##  Median : 327039      Median :41597       Median :15459148
##  Mean   : 735332      Mean   :43888       Mean   :20415698
##  3rd Qu.:1067564      3rd Qu.:47635       3rd Qu.:34044271
##  Max.   :3222035      Max.   :85801       Max.   :60618780
##  new_hospitalized_patients cumulative_hospitalized_patients
##  Min.   :  72.0            Min.   : 23549
##  1st Qu.: 217.0            1st Qu.:120259
##  Median : 802.5            Median :134497
##  Mean   :1072.7            Mean   :150166
##  3rd Qu.:1590.0            3rd Qu.:175395
##  Max.   :4396.0            Max.   :327517
##  current_hospitalized_patients current_ventilator_patients
##  Min.   :  772                 Min.   :   0.0
##  1st Qu.: 1980                 1st Qu.: 162.8
##  Median : 7461                 Median : 668.5
##  Mean   : 9711                 Mean   : 969.4
##  3rd Qu.:16395                 3rd Qu.:1419.2
##  Max.   :35392                 Max.   :3301.0
##  new_persons_vaccinated cumulative_persons_vaccinated
##  Min.   :0              Min.   :0
##  1st Qu.:0              1st Qu.:0
##  Median :0              Median :0
##  Mean   :0              Mean   :0
##  3rd Qu.:0              3rd Qu.:0
##  Max.   :0              Max.   :0
##  new_persons_fully_vaccinated cumulative_persons_fully_vaccinated
##  Min.   :0                    Min.   :0
##  1st Qu.:0                    1st Qu.:0
##  Median :0                    Median :0
##  Mean   :0                    Mean   :0
##  3rd Qu.:0                    3rd Qu.:0
##  Max.   :0                    Max.   :0
##  mobility_retail_and_recreation mobility_grocery_and_pharmacy mobility_parks
##  Min.   :-92.0                  Min.   :-89.00                Min.   :-47.00
##  1st Qu.:-62.0                  1st Qu.:-19.00                1st Qu.: -4.00
```

```
## Median :-42.0               Median :-13.00              Median : 20.00
## Mean   :-44.3               Mean   :-15.27              Mean   : 29.62
## 3rd Qu.:-27.0               3rd Qu.:-10.00              3rd Qu.: 61.75
## Max.   :  3.0               Max.   : 23.00              Max.   :152.00
## mobility_transit_stations mobility_workplaces mobility_residential
## Min.   :-85.00             Min.   :-85.00      Min.   : 1.00
## 1st Qu.:-58.00             1st Qu.:-51.00      1st Qu.:10.00
## Median :-46.00             Median :-43.00      Median :13.00
## Mean   :-48.91             Mean   :-41.23      Mean   :14.45
## 3rd Qu.:-40.00             3rd Qu.:-33.00      3rd Qu.:19.00
## Max.   :-23.00             Max.   : -6.00      Max.   :32.00
## stringency_index average_temperature_celsius minimum_temperature_celsius
## Min.   :60.19    Min.   :-2.380               Min.   :-7.380
## 1st Qu.:65.74    1st Qu.: 6.985               1st Qu.: 2.893
## Median :69.91    Median : 9.910               Median : 5.825
## Mean   :70.73    Mean   : 9.871               Mean   : 5.921
## 3rd Qu.:75.00    3rd Qu.:13.115               3rd Qu.: 9.250
## Max.   :87.96    Max.   :20.780               Max.   :15.580
## maximum_temperature_celsius  rainfall_mm      dew_point
## Min.   : 0.76                Min.   : 0.000   Min.   :-3.520
## 1st Qu.:10.03                1st Qu.: 0.085   1st Qu.: 3.987
## Median :13.48                Median : 0.940   Median : 6.870
## Mean   :13.32                Mean   : 2.429   Mean   : 6.832
## 3rd Qu.:17.06                3rd Qu.: 3.322   3rd Qu.: 9.818
## Max.   :29.26                Max.   :31.640   Max.   :16.540
## relative_humidity
## Min.   :52.39
## 1st Qu.:75.35
## Median :83.90
## Mean   :81.72
## 3rd Qu.:89.01
## Max.   :97.08
```

```r
cat("Vaccination Data – Summary statistics:\n")
```

```
## Vaccination Data – Summary statistics:
```

```r
print(summary_vacc)
```

```
##      date            new_confirmed      new_deceased      new_tested
## Length:475        Min.   :  1431   Min.   :    4.0   Min.   : 204510
## Class :character  1st Qu.: 10204   1st Qu.:  58.5    1st Qu.: 678121
## Mode  :character  Median : 32109   Median : 129.0    Median : 864788
##                   Mean   : 39679   Mean   : 190.3    Mean   : 898059
##                   3rd Qu.: 47464   3rd Qu.: 195.0    3rd Qu.:1078050
##                   Max.   :275647   Max.   :1370.0    Max.   :2061041
## cumulative_confirmed cumulative_deceased cumulative_tested
## Min.   : 3280627     Min.   : 86933      Min.   : 61176472
## 1st Qu.: 4521555     1st Qu.:128385      1st Qu.:166399348
## Median : 7195114     Median :134587      Median :276524640
## Mean   : 9609425     Mean   :139246      Mean   :286653868
## 3rd Qu.:14075456     3rd Qu.:150556      3rd Qu.:409811567
## Max.   :22069699     Max.   :176190      Max.   :517854849
```

```
##    new_hospitalized_patients cumulative_hospitalized_patients
##    Min.   :  81              Min.   :331827
##    1st Qu.: 555              1st Qu.:464965
##    Median : 909              Median :527957
##    Mean   :1104              Mean   :562905
##    3rd Qu.:1468              3rd Qu.:644490
##    Max.   :4580              Max.   :851929
##    current_hospitalized_patients current_ventilator_patients
##    Min.   :  877                 Min.   : 119.0
##    1st Qu.: 4910                 1st Qu.: 326.5
##    Median : 7980                 Median : 769.0
##    Mean   :10422                 Mean   : 886.9
##    3rd Qu.:14762                 3rd Qu.: 940.5
##    Max.   :39254                 Max.   :4077.0
##    new_persons_vaccinated cumulative_persons_vaccinated
##    Min.   : -2263         Min.   : 2431648
##    1st Qu.: 19582         1st Qu.:35529822
##    Median : 38726         Median :48270113
##    Mean   :107250         Mean   :42184342
##    3rd Qu.:142447         3rd Qu.:51795270
##    Max.   :752308         Max.   :53230188
##    new_persons_fully_vaccinated cumulative_persons_fully_vaccinated
##    Min.   :   818               Min.   :  412167
##    1st Qu.: 17958               1st Qu.:17972468
##    Median : 33674               Median :43455083
##    Mean   :103926               Mean   :33088155
##    3rd Qu.:168109               3rd Qu.:47445252
##    Max.   :556951               Max.   :49756334
##    mobility_retail_and_recreation mobility_grocery_and_pharmacy mobility_parks
##    Min.   :-87.00                 Min.   :-88.0000              Min.   :-49.00
##    1st Qu.:-28.00                 1st Qu.: -3.0000              1st Qu.:  8.00
##    Median :-15.00                 Median :  2.0000              Median : 29.00
##    Mean   :-22.06                 Mean   : -0.2316              Mean   : 34.75
##    3rd Qu.:-10.00                 3rd Qu.:  6.0000              3rd Qu.: 60.00
##    Max.   :  7.00                 Max.   : 42.0000              Max.   :203.00
##    mobility_transit_stations mobility_workplaces mobility_residential
##    Min.   :-81.00            Min.   :-78.00      Min.   :-1.000
##    1st Qu.:-41.00            1st Qu.:-39.00      1st Qu.: 6.000
##    Median :-33.00            Median :-30.00      Median : 8.000
##    Mean   :-36.39            Mean   :-29.21      Mean   : 8.526
##    3rd Qu.:-28.00            3rd Qu.:-23.00      3rd Qu.:11.000
##    Max.   :-12.00            Max.   : -1.00      Max.   :23.000
##    stringency_index average_temperature_celsius minimum_temperature_celsius
##    Min.   :11.11    Min.   :-3.680              Min.   :-6.270
##    1st Qu.:42.13    1st Qu.: 4.265              1st Qu.: 0.375
##    Median :46.76    Median : 7.610              Median : 3.590
##    Mean   :50.88    Mean   : 8.105              Mean   : 4.209
##    3rd Qu.:61.11    3rd Qu.:12.365              3rd Qu.: 8.190
##    Max.   :87.96    Max.   :20.930              Max.   :14.760
##    maximum_temperature_celsius  rainfall_mm       dew_point
##    Min.   :-1.790               Min.   : 0.000   Min.   :-8.490
##    1st Qu.: 7.655               1st Qu.: 0.030   1st Qu.: 1.465
##    Median :10.860               Median : 0.410   Median : 4.650
##    Mean   :11.597               Mean   : 1.992   Mean   : 5.110
```

```
##  3rd Qu.:16.100              3rd Qu.: 2.265    3rd Qu.: 8.875
##  Max.   :26.860              Max.   :20.120    Max.   :15.440
##  relative_humidity
##  Min.   : 0.00
##  1st Qu.:76.47
##  Median :82.45
##  Mean   :81.35
##  3rd Qu.:87.97
##  Max.   :97.25
```

In this section, we computed and examined the summary statistics for both the pre-vaccination and vaccination datasets. These summary statistics provide a comprehensive overview of the key characteristics of the datasets. For the pre-vaccination data, the summary statistics offer insights into variables such as confirmed cases, death cases, hospitalized patients, testing trends, mobility trends, stringency index, and meteorological factors. Similarly, the vaccination data's summary statistics shed light on variables including the persons vaccinated, cumulative persons vaccinated, along with pre vaccination dataset's variables. By analyzing these summary statistics, we gain a better understanding of the distribution, central tendencies, and variations within each dataset's variables, which serves as a foundation step for further analysis and interpretation.

Data visualization:

Exploring the patterns of the data to provide a clear and concise overview of patterns, trends, and insights within the data.

Pre Vaccination: In this section, we analyze the pre-vaccination data to visualize the patterns of new cases, deaths, and hospitalizations. By exploring these visualizations, we aim to identify any trends or factors that may have influenced the COVID-19 metrics during the pre-vaccination period.

Analyzing the trend of cases, deaths and hospitalized patients data in Pre vaccination period:

```r
# Create a sequence of row numbers to use as x-axis
row_numbers <- seq(nrow(pre_vaccination_data))

# Plot the patterns of cases, deaths and hospitalizations
ggplot()+
  geom_line(data = pre_vaccination_data, aes(x = row_numbers,
                                             y = new_confirmed,
                                             color = "New Confirmed"))+
  geom_line(data = pre_vaccination_data, aes(x = row_numbers,
                                             y = new_deceased,
                                             color = "New Deceased"))+
  geom_line(data = pre_vaccination_data, aes(x = row_numbers,
                                      y = new_hospitalized_patients,
                                      color = "New Hospitalized Patients"))+
  labs(title = "Patterns of New Confirmed, New Deceased, and New
       Hospitalized Patients (Pre-Vaccination Period)", x = " ",
       y = "Count", color = "Variable") + theme_minimal()
```

Patterns of New Confirmed, New Deceased, and New Hospitalized Patients (Pre–Vaccination Period)
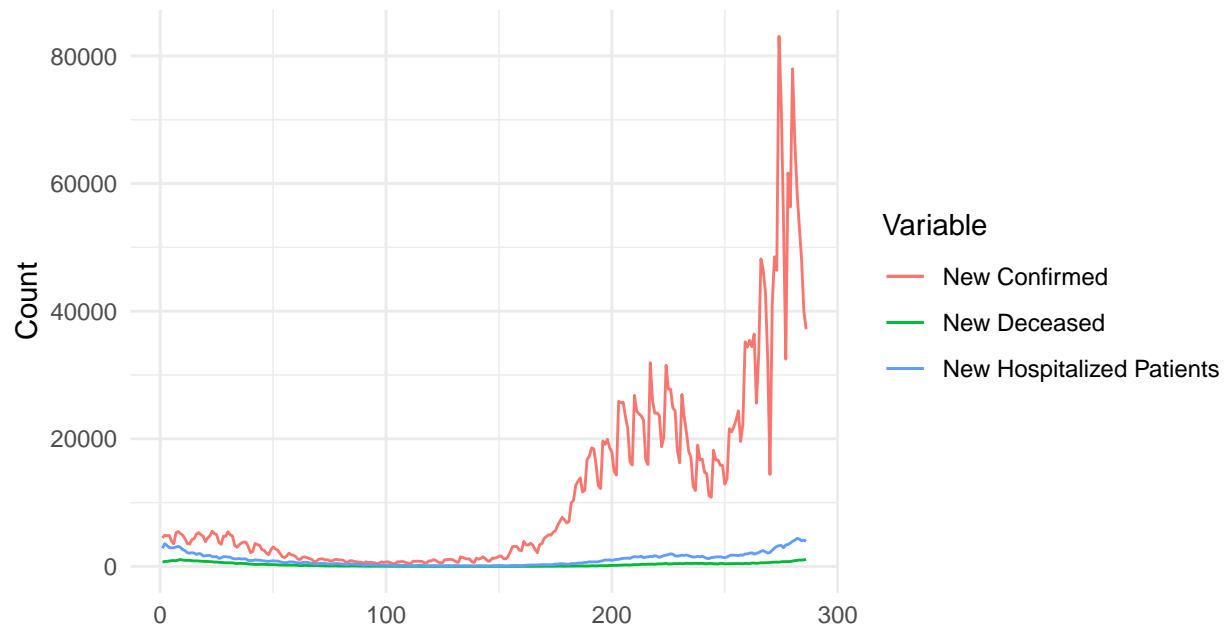
The plot shows **very high number of COVID-19 cases compared to hospitalizations and deaths**, likely due to many individuals opting for home quarantine and subsequent recovery. Severe cases led to hospitalizations, resulting in lower cumulative counts. Fortunately, the majority of cases resulted in survival, highlighting the significance of early detection and proper healthcare measures.

Analyzing the influence of mobility factors on cases, deaths and hospitalized patients data in Pre vaccination period:

```r
# Create long format data for mobility variables
pre_vaccination_data_l <- pre_vaccination_data %>%
                          gather(key = "MobilityFactor",
                                 value = "MobilityValue",
                                 starts_with("mobility_"))

# Create line plot for mobility factors against cases
lp_cases <- ggplot(data = pre_vaccination_data_l,aes(x = MobilityValue)) +
  geom_line(aes(y = new_confirmed, color = "New Cases")) +
  labs(title = "Pre-Vaccination - Mobility Factors vs. New Cases",
       x = "Mobility Value", y = "Count", color = "Variable") +
  facet_wrap(~ MobilityFactor, scales = "free_x") + theme_minimal()

# Create line plot for mobility factors against deaths and hospital counts
lp_deaths_hospital <- ggplot(data = pre_vaccination_data_l,
                             aes(x = MobilityValue)) +
  geom_line(aes(y = new_deceased, color = "New Deaths")) +
  geom_line(aes(y = new_hospitalized_patients,
                color = "New Hospitalized Patients")) +
  # Adjust title for correct alignment
  labs(title =
  "Pre-Vaccination - Mobility Factors vs.Deaths and Hospitalized Patients",
```

```
        x = "Mobility Value", y = "Count", color = "Variable") +
  facet_wrap(~ MobilityFactor, scales = "free_x") + theme_minimal()

# Combine line plots for mobility factors using cowplot
fig_cases <- plot_grid(lp_cases + theme(legend.position="bottom"),
                       ncol = 1, align = 'v', axis = 'lr')

fig_deaths_hospital <- plot_grid(lp_deaths_hospital +
                                 theme(legend.position = "bottom"),
                                 ncol = 1, align = 'v', axis = 'lr')

# Print the two figures
print(fig_cases)
```
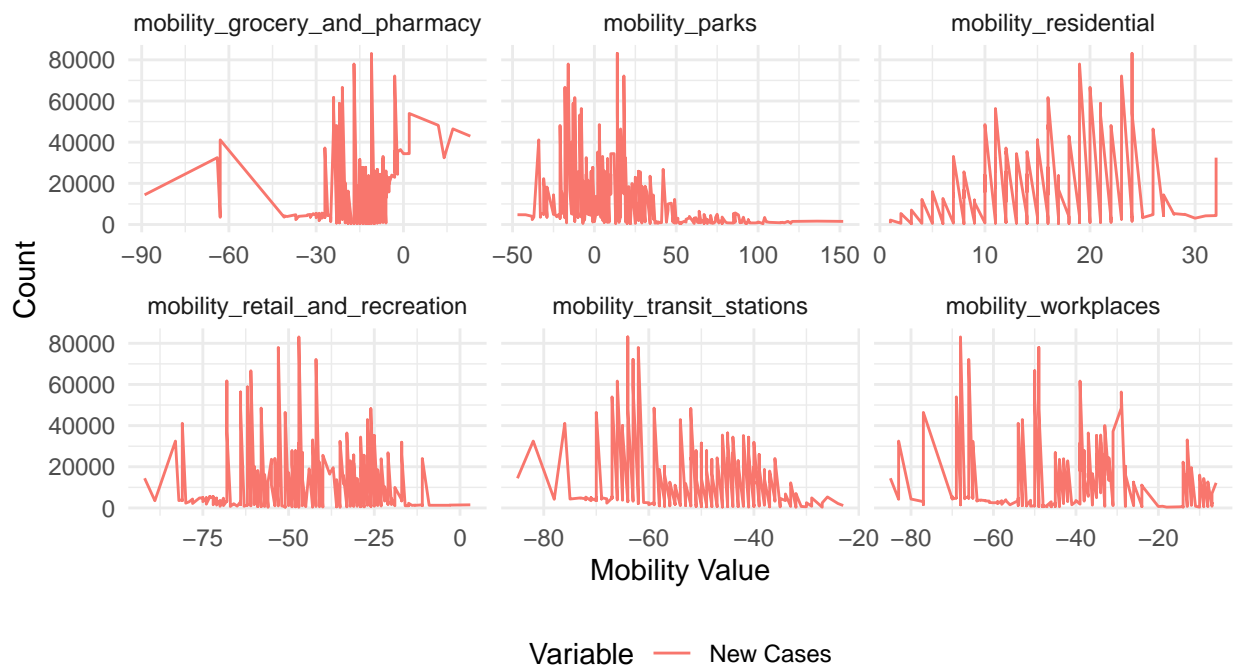

Pre–Vaccination – Mobility Factors vs. New Cases

```
print(fig_deaths_hospital)
```

Pre−Vaccination – Mobility Factors vs.Deaths and Hospitalized Patients
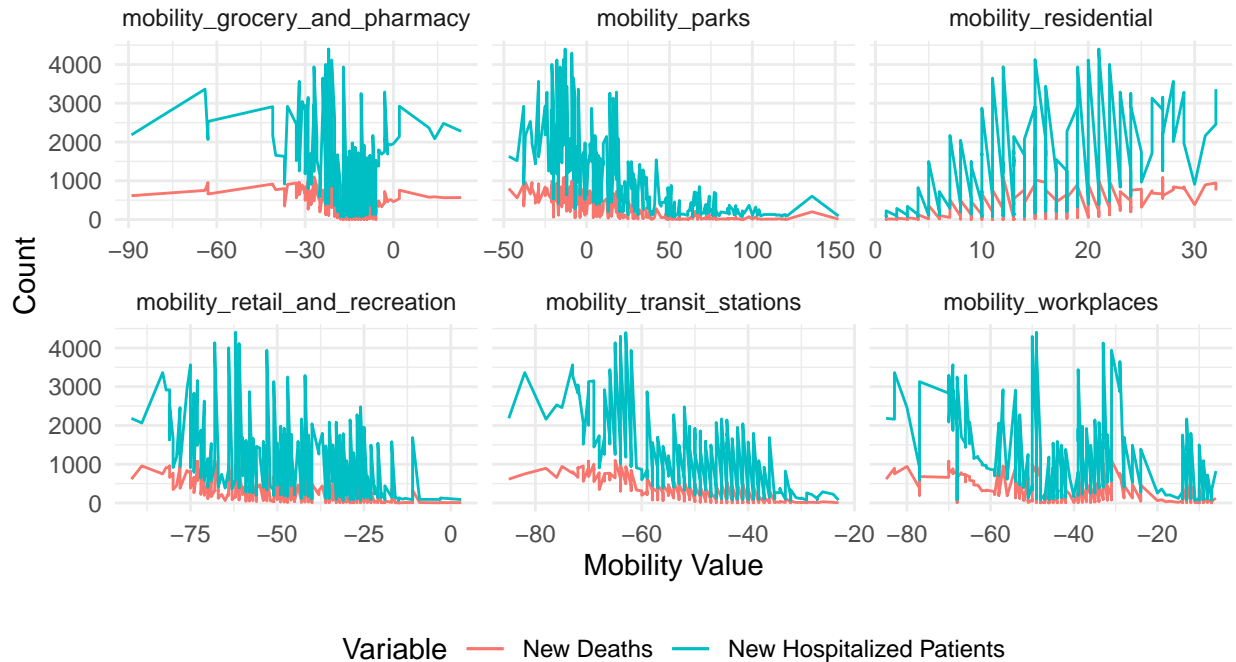
Interpretation: Mobility Variables' Impact on COVID-19 Cases, Deaths, and Hospital Patients:

Mobility Factors - Description:

a. mobility_grocery_and_pharmacy: Percentage change in visits to places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies compared to baseline.

b. mobility_parks: Percentage change in visits to places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens compared to baseline.

c. mobility_transit_stations: Percentage change in visits to places like public transport hubs such as subway, bus, and train stations compared to baseline.

d. mobility_retail_and_recreation: Percentage change in visits to restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters compared to baseline.

e. mobility_residential: Percentage change in visits to places of residence compared to baseline.

f. mobility_workplaces: Percentage change in visits to places of work compared to baseline.

Relationship between Mobility Factors and COVID-19 Cases:

A decrease in COVID-19 cases is observed when all the mobility factors have a negative percentage change. This suggests that reduced mobility in various public places can contribute to lowering the spread of the virus.

Relationship between Mobility Factors and COVID-19 Deaths and Hospital Patients:

o Unlike COVID-19 cases, the relationship between mobility factors and deaths/hospital patients is not as direct. While there is a decrease in deaths/hospital patients for a negative percentage change, there are also some instances of spike in deaths/hospital patients.

o One possible reason for this discrepancy is that hospitalizations and deaths may not have an immediate response to changes in mobility. People may seek hospitalization when their condition worsens after quarantine or self-isolation. Similarly, COVID-19 deaths might occur after a certain period of infection, and mobility may not be the sole factor influencing these outcomes.

Analyzing the influence of weather factors on cases, deaths and hospitalized patients data in Pre vaccination period:

```r
# Create long format data for weather variables
pre_vaccination_data_l <- pre_vaccination_data %>%
                          gather(key = "WeatherVariable",
                                 value = "WeatherValue",
                                 average_temperature_celsius,
                                 minimum_temperature_celsius,
                                 maximum_temperature_celsius,
                                 rainfall_mm,
                                 dew_point,
                                 relative_humidity)


# Create line plot for weather variables against cases
lp_cases_weather <- ggplot(data = pre_vaccination_data_l, aes(x = WeatherValue)) +
  geom_line(aes(y = new_confirmed, color = "New Cases")) +
  labs(title = "Pre-Vaccination - Weather Factors vs. New Cases",
       x = "Weather Value", y = "Count", color = "Variable") +
  facet_wrap(~ WeatherVariable, scales = "free_x") + theme_minimal()


# Create line plot for weather variables against deaths and hospital counts
lp_deaths_hospital_weather <- ggplot(data = pre_vaccination_data_l, aes(x = WeatherValue)) +
  geom_line(aes(y = new_deceased, color = "New Deaths")) +
  geom_line(aes(y = new_hospitalized_patients, color = "New Hospitalized Patients")) +
  # Adjust title for correct alignment
  labs(title =
  "Pre-Vaccination - Weather Factors vs. Deaths and Hospitalized Patients",
       x = "Weather Value", y = "Count", color = "Variable") +
  facet_wrap(~ WeatherVariable, scales = "free_x") + theme_minimal()


# Combine line plots for weather variables using cowplot
fig_cases_weather <- plot_grid(lp_cases_weather + theme(legend.position =
                               "bottom"), ncol = 1, align = 'v',
                               axis = 'lr')


fig_deaths_hospital_weather <- plot_grid(lp_deaths_hospital_weather +
                                    theme(legend.position = "bottom"),
                                    ncol = 1, align = 'v',
                                    axis = 'lr')


# Print the two figures
print(fig_cases_weather)
```
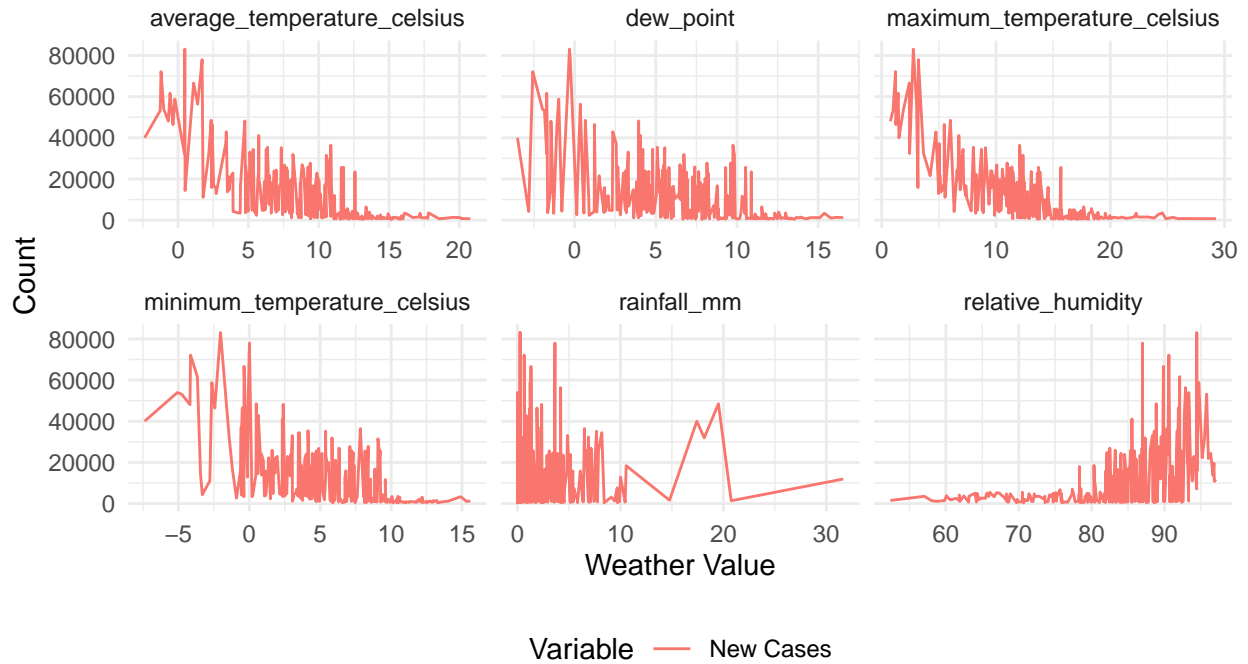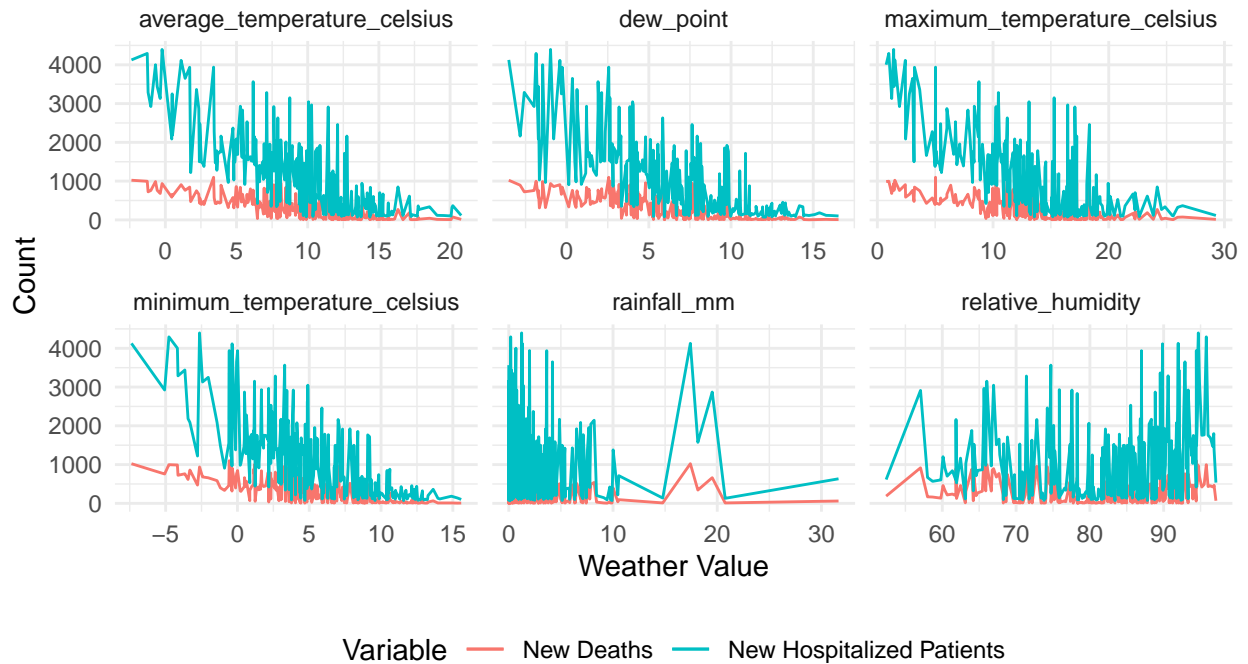
Pre−Vaccination – Weather Factors vs. New Cases

```
print(fig_deaths_hospital_weather)
```


Pre−Vaccination – Weather Factors vs. Deaths and Hospitalized Patients

Interpretation: Weather Variables' Impact on COVID-19 Cases, Deaths, and Hospital Patients:

Relationship between Weather Factors and COVID-19 Cases:

High Temperature: The plot shows that regions with higher temperatures tend to have fewer COVID-19 cases. This observation might be due to the fact that viruses like SARS-CoV-2, which causes COVID-19, can spread less efficiently in hot condition. However, it's essential to consider other factors as well, such as vaccination rates, population density, and public health measures.

Low Temperature: Conversely, areas with lower temperatures appear to have higher numbers of COVID-19 cases. Cold temperatures may promote the survival and transmission of the virus, leading to more infections during colder months.

Dew point: Higher temperature of dew point has lower cases comparing to the lower temperature, indicating that the dew point temperature might play a role in influencing the spread of the virus.

High Humidity: The plot indicates that regions with higher humidity levels show a higher number of COVID-19 cases. This might be surprising, as many respiratory viruses thrive in low humidity. However, high humidity can lead to indoor crowding and reduced ventilation, which can facilitate viral transmission.

Relationship between Weather Factors and COVID-19 Hospital Patients:

The pattern observed in hospital patients follows a similar trend to that of COVID-19 cases. This suggests that weather factors might have an impact on the overall burden of COVID-19 on healthcare facilities. Regions with higher temperatures and lower humidity might experience lower hospitalization rates, while the opposite may be true for areas with lower temperatures and higher humidity.

Relationship between Weather Factors and COVID-19 Deaths:

Temperature Effect: Similar to the cases and hospitalization patterns, regions with higher temperatures show a lower number of COVID-19 deaths. Warmer temperatures might reduce the stability of the virus on surfaces and in the air, potentially leading to fewer infections and subsequent fatalities.

Humidity Effect: Interestingly, the plot shows that humidity has no significant relationship with COVID-19 deaths. This finding might suggest that other factors, such as vaccination rates, healthcare system capacity, and adherence to preventive measures, play a more critical role in determining the mortality rate.

Vaccination period: In this section, we examine the data from the vaccination period to gain insights into the impact of vaccination on COVID-19 metrics. We will focus on visualizing the vaccination variables against new cases, deaths, and hospitalizations. Through these visualizations, we aim to understand the role of vaccination in influencing the trends and outcomes of COVID-19 during this specific period. These visualizations will provide valuable information about the vaccination's overall impact on controlling the pandemic and highlight the significance of vaccination in public health strategies.

```
# Create long format data for vaccination variables
vaccination_data_l <- vaccination_data %>%
                    gather(key = "VaccinationVariable",
                           value = "VaccinationValue",
                           new_persons_vaccinated,
                           new_persons_fully_vaccinated)

# Create line plot for vaccination variables against cases
lp_cases_vaccination <- ggplot(data = vaccination_data_l,
                             aes(x = VaccinationValue)) +
  geom_line(aes(y = new_confirmed, color = "New Cases")) +
  labs(title = "Vaccination period - Vaccination Variables vs. New Cases",
       x = "Vaccination Value", y = "Count", color = "Variable") +
  facet_wrap(~ VaccinationVariable, scales = "free_x") + theme_minimal() +
  scale_x_continuous(labels = scales::comma_format()) +
```

```
    scale_y_continuous(labels = scales::comma_format())

# Create line plot for vaccination variables against deaths and hospital counts
lp_deaths_hospital_vaccination <- ggplot(data = vaccination_data_l,
                                    aes(x = VaccinationValue)) +
  geom_line(aes(y = new_deceased, color = "New Deaths")) +
  geom_line(aes(y = new_hospitalized_patients, color =
             "New Hospitalized Patients")) +
  # Adjust title for correct alignment
  labs(title =
"Vaccination period - Vaccination vs.Deaths & Hospitalized Patients",
       x = "Vaccination Value", y = "Count", color = "Variable") +
  facet_wrap(~ VaccinationVariable, scales = "free_x") + theme_minimal() +
  scale_x_continuous(labels = scales::comma_format())

# Combine line plots for vaccination variables using cowplot
fig_cases_vaccination <- plot_grid(lp_cases_vaccination +
                                  theme(legend.position = "bottom"),
                                  ncol = 1, align = 'v', axis = 'lr')

fig_deaths_hospital_vaccination <- plot_grid(lp_deaths_hospital_vaccination
                                      + theme(legend.position =
                                      "bottom"), ncol = 1,
                                      align = 'v', axis = 'lr')

# Print the two figures
print(fig_cases_vaccination)
```
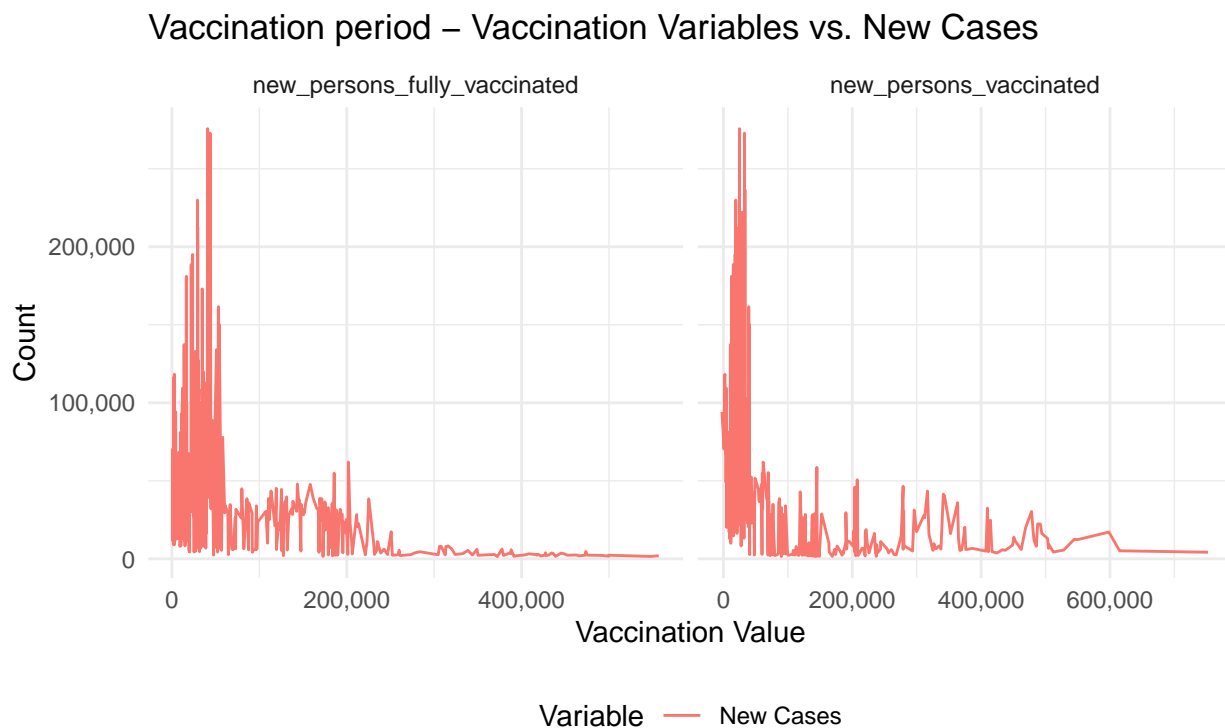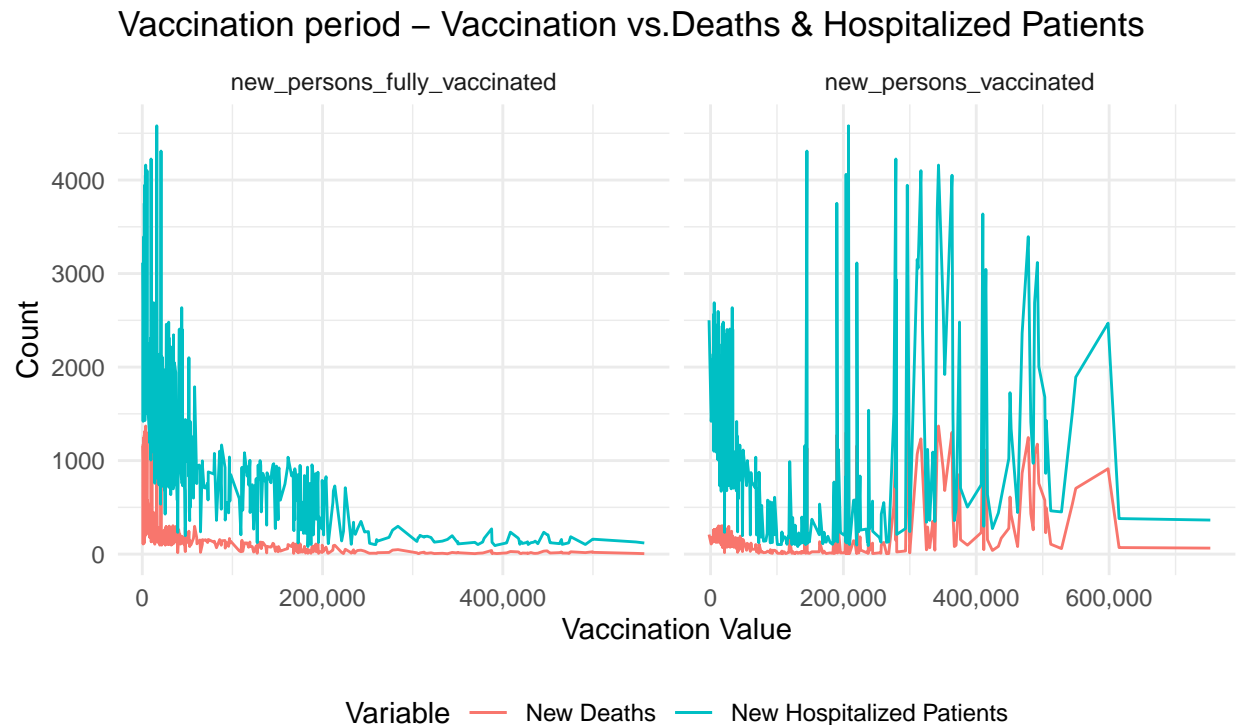
Vaccination period – Vaccination Variables vs. New Cases

```
print(fig_deaths_hospital_vaccination)
```

Vaccination period – Vaccination vs.Deaths & Hospitalized Patients

Interpretation: Vaccination Variables' Impact on COVID-19 Cases, Deaths, and Hospital Patients:

Relationship between Vaccination Factors and COVID-19 Cases:

Both new_persons_vaccinated and new_persons_fully_vaccinated show a general decreasing trend in the number of new COVID-19 cases. This indicates that as the number of vaccinated individuals increases, there is a corresponding decrease in the number of new infections. The decline in cases suggests that the vaccination efforts are contributing positively to controlling the spread of the virus.

Relationship between Vaccination Factors and COVID-19 Hospital Patients:

The line plot for new_persons_fully_vaccinated shows a clear decrease in the number of hospitalized patients as the count of fully vaccinated individuals increases. This suggests that a higher number of fully vaccinated individuals are less likely to require hospitalization when infected with COVID-19. However, the line plot for new_persons_vaccinated does not show a consistent pattern, with ups and downs throughout the plot. This could indicate that the overall effect of new_persons_vaccinated on hospitalizations is not as pronounced or may vary based on other factors.

Relationship between Vaccination Factors and COVID-19 Deaths:

Similar to hospitalizations, the line plot for new_persons_fully_vaccinated shows a decline in COVID-19-related deaths as the count of fully vaccinated individuals increases. This indicates that a higher proportion of fully vaccinated individuals have better protection against severe outcomes of the disease. On the other hand, the relationship between new_persons_vaccinated and deaths is less evident, with fluctuations over time.

EDA - Conclusion:

In conclusion, Our Exploratory Data Analysis (EDA) revealed intriguing insights into the relationship between mobility, weather factors, vaccination variables, and COVID-19 metrics. During the pre-vaccination period, **specific mobility factors exhibited correlations with new cases, and weather factors also showed few associations**. However, the vaccination period showcased promising outcomes, with **both new_persons_vaccinated and new_persons_fully_vaccinated demonstrating a remarkable decrease in COVID-19 cases**. Moreover, these variables also influenced hospitalizations and deaths, although the impact differed.

To gain a more comprehensive understanding of the effects of vaccination and other factors on COVID-19 metrics, we plan to perform multivariate linear regression (MLR) on the pre-vaccination and vaccination data. This will allow us to refine our insights and make evidence-based conclusions.

Multivariate Linear Regression(MLR):

Performing multivariate linear regression on the pre vaccination and vaccination period data-sets separately to find the impact of vaccination and the other factors on the Covid 19 infections, hospitalization and mortality rate.

In this section, we adopt a data splitting approach to divide the data set into two subsets: the training set, and test set. We begin by building the model using the training data, and then we utilize the test data for the final model prediction, ensuring a robust and reliable outcome. This systematic process enables us to create an effective and well-tuned model capable of making accurate predictions based on the given predictor variables.

MLR on Pre vaccination period data:

Data Splitting:

```r
# Define the predictor and response variables
# pvd - pre vaccination data
predictors_pvd <- pre_vaccination_data %>%
  select(new_tested, cumulative_tested, mobility_retail_and_recreation,
         mobility_grocery_and_pharmacy, mobility_parks,
         mobility_transit_stations, mobility_workplaces,
         mobility_residential, stringency_index,
         average_temperature_celsius, minimum_temperature_celsius,
         maximum_temperature_celsius, rainfall_mm, dew_point,
         relative_humidity)

responses_pvd <- pre_vaccination_data %>%
  select(new_confirmed, new_deceased, cumulative_confirmed,
         cumulative_deceased, new_hospitalized_patients,
         cumulative_hospitalized_patients, current_hospitalized_patients,
         current_ventilator_patients)

# Scale the data
predictors_pvd <- scale(predictors_pvd)
responses_pvd <- scale(responses_pvd)

# Convert scaled variables back to a data frame
predictors_pvd <- as.data.frame(predictors_pvd)
responses_pvd <- as.data.frame(responses_pvd)
```

```r
# Data splitting into train and test sets
set.seed(123)  # Set seed for reproducibility
train_index_pvd <- createDataPartition(1:nrow(pre_vaccination_data),
                                        p = 0.7, list = FALSE)
train_data_pvd <- predictors_pvd[train_index_pvd, ]
train_response_pvd <- responses_pvd[train_index_pvd, ]

test_data_pvd <- predictors_pvd[-train_index_pvd, ]
test_response_pvd <- responses_pvd[-train_index_pvd, ]
```

In data splitting, we selected relevant predictor and response variables from the "pre_vaccination_data" data set to build a predictive model for COVID-19 outcomes. The predictor variables encompass a wide range of factors such as testing metrics, mobility trends, and environmental features. To ensure comparability, we applied z-score standardization to scale all variables appropriately. The data was then split into training and testing sets, with **70% used for model training and 30% for evaluation.**

MLR - Model Building:

```r
# Combine response variables into a multivariate response data frame
# pvd - pre vaccination data
train_response_multivariate_pvd <- as.matrix(train_response_pvd)

# Build the MLR model using train data
pvd_model <- lm(train_response_multivariate_pvd ~ ., data = train_data_pvd)
```

Here, **the response variables are combined into a multivariate response data frame called "train_response_multivariate_pvd." This step is crucial as it allows us to build a multivariate linear regression (MLR) model capable of predicting multiple response variables** simultaneously. The "lm" function is then used to create the MLR model using the training data. By specifying the formula "train_response_multivariate_pvd ~ .", the model will use all available predictor variables from the "train_data_pvd" data frame. The model will estimate coefficients for each predictor, indicating the direction and strength of their influence on the multiple response variables, such as new confirmed cases, deaths and hospitalizations.

MLR - Model Evaluation:

```r
# Evaluate the model using test data
# pvd - pre vaccination data
predictions_pvd <- predict(pvd_model, newdata = test_data_pvd)

# Combine response variables into a multivariate response data frame
test_response_multivariate_pvd <- as.matrix(test_response_pvd)

# Calculate Mean Squared Error (MSE) as an evaluation metric
mse_pvd <- mean((predictions_pvd - test_response_multivariate_pvd)^2)
rmse_pvd <- sqrt(mse_pvd)
# Print the MSE
cat("Mean Squared Error (MSE):", mse_pvd, "\n")
```

```
## Mean Squared Error (MSE): 0.1028802
```

17

```
cat("Root Mean Squared Error (RMSE):", rmse_pvd, "\n")
```

Root Mean Squared Error (RMSE): 0.3207495

MLR model's performance is evaluated using the test data. The "predict" function is used to generate predictions for the "test_data_pvd" set. These predictions are then compared to the true response variables, which are combined into a multivariate response data frame called "test_response_multivariate_pvd". To quantify the model's performance, the Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE) values are calculated by measuring the average squared difference between the predicted and true values for all response variables.

The obtained **Mean Squared Error (MSE) value of 0.1028802 and Root Mean Squared Error (RMSE) value of 0.3207495** showcases the effectiveness of the multivariate linear regression model in accurately predicting COVID-19 outcomes based on the selected predictor variables. A low MSE and RMSE indicate that the model's predictions are significantly close to the actual data, highlighting its reliability and robustness in generalizing to new and unseen data. This strong alignment between predictions and actual outcomes underscores the model's high accuracy and ability to provide valuable insights into understanding and forecasting COVID-19 trends. Consequently, we can confidently assert that **this is an effective and dependable model for gaining deeper insights into the factors influencing COVID-19 outcomes.**

MLR - Model Summary:

```
summary(pvd_model)
```

```
## Response new_confirmed :
##
## Call:
## lm(formula = new_confirmed ~ new_tested + cumulative_tested +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_pvd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.04305 -0.15027 -0.02795  0.12222  2.21049
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -0.01406    0.02604  -0.540   0.5899
## new_tested                      0.13059    0.08315   1.571   0.1180
## cumulative_tested               0.69810    0.09204   7.585 1.54e-12 ***
## mobility_retail_and_recreation -0.17975    0.10046  -1.789   0.0752 .
## mobility_grocery_and_pharmacy   0.10042    0.04435   2.264   0.0247 *
## mobility_parks                 -0.09489    0.05612  -1.691   0.0925 .
## mobility_transit_stations       0.23290    0.14537   1.602   0.1108
## mobility_workplaces            -0.26564    0.12827  -2.071   0.0397 *
## mobility_residential           -0.09608    0.11821  -0.813   0.4174
## stringency_index                0.32498    0.04400   7.386 4.93e-12 ***
## average_temperature_celsius     0.41457    0.57402   0.722   0.4711
## minimum_temperature_celsius     0.17537    0.14211   1.234   0.2187
## maximum_temperature_celsius    -0.20966    0.19598  -1.070   0.2861
```

```
## rainfall_mm                        0.00964    0.02909    0.331   0.7408
## dew_point                         -0.39695    0.47433   -0.837   0.4037
## relative_humidity                  0.08864    0.20566    0.431   0.6670
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.368 on 186 degrees of freedom
## Multiple R-squared:  0.8618, Adjusted R-squared:  0.8507
## F-statistic: 77.36 on 15 and 186 DF,  p-value: < 2.2e-16
##
##
## Response new_deceased :
##
## Call:
## lm(formula = new_deceased ~ new_tested + cumulative_tested +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_pvd)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1649 -0.2450 -0.0198  0.2444  1.3821
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    0.002282   0.028276   0.081 0.935776
## new_tested                    -0.363395   0.090295  -4.025 8.30e-05 ***
## cumulative_tested              0.604699   0.099949   6.050 7.79e-09 ***
## mobility_retail_and_recreation 0.096892   0.109101   0.888 0.375636
## mobility_grocery_and_pharmacy  0.029989   0.048161   0.623 0.534257
## mobility_parks                -0.239552   0.060945  -3.931 0.000119 ***
## mobility_transit_stations     -0.439654   0.157870  -2.785 0.005908 **
## mobility_workplaces            0.129251   0.139301   0.928 0.354683
## mobility_residential           0.087090   0.128374   0.678 0.498355
## stringency_index               0.334174   0.047781   6.994 4.67e-11 ***
## average_temperature_celsius    0.984196   0.623378   1.579 0.116078
## minimum_temperature_celsius   -0.272245   0.154326  -1.764 0.079358 .
## maximum_temperature_celsius   -0.332541   0.212830  -1.562 0.119876
## rainfall_mm                   -0.054714   0.031595  -1.732 0.084980 .
## dew_point                     -0.433899   0.515115  -0.842 0.400684
## relative_humidity              0.050227   0.223348   0.225 0.822317
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3996 on 186 degrees of freedom
## Multiple R-squared:  0.8598, Adjusted R-squared:  0.8485
## F-statistic: 76.02 on 15 and 186 DF,  p-value: < 2.2e-16
##
##
## Response cumulative_confirmed :
##
## Call:
```

```
## lm(formula = cumulative_confirmed ~ new_tested + cumulative_tested +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_pvd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.46813 -0.10753  0.00402  0.08988  0.43466
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -0.004633   0.011064  -0.419   0.6759
## new_tested                     -0.044499   0.035333  -1.259   0.2095
## cumulative_tested               1.062628   0.039110  27.170  < 2e-16 ***
## mobility_retail_and_recreation -0.345595   0.042691  -8.095 7.27e-14 ***
## mobility_grocery_and_pharmacy   0.076484   0.018845   4.059 7.27e-05 ***
## mobility_parks                  0.186392   0.023848   7.816 3.92e-13 ***
## mobility_transit_stations       0.028388   0.061775   0.460   0.6464
## mobility_workplaces            -0.039684   0.054508  -0.728   0.4675
## mobility_residential            0.028843   0.050233   0.574   0.5665
## stringency_index                0.143109   0.018697   7.654 1.02e-12 ***
## average_temperature_celsius     0.260808   0.243928   1.069   0.2864
## minimum_temperature_celsius     0.084606   0.060388   1.401   0.1629
## maximum_temperature_celsius    -0.121862   0.083280  -1.463   0.1451
## rainfall_mm                     0.024250   0.012363   1.961   0.0513 .
## dew_point                      -0.251251   0.201564  -1.247   0.2141
## relative_humidity               0.122209   0.087396   1.398   0.1637
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1564 on 186 degrees of freedom
## Multiple R-squared:  0.9774, Adjusted R-squared:  0.9755
## F-statistic: 535.3 on 15 and 186 DF,  p-value: < 2.2e-16
##
##
## Response cumulative_deceased :
##
## Call:
## lm(formula = cumulative_deceased ~ new_tested + cumulative_tested +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_pvd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24453 -0.18426  0.03108  0.20409  0.75154
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -0.004966   0.023579  -0.211   0.8334
```

```
## new_tested                        0.106622   0.075294    1.416    0.1584
## cumulative_tested                 0.950098   0.083344   11.400   < 2e-16 ***
## mobility_retail_and_recreation   -0.582829   0.090976   -6.406  1.19e-09 ***
## mobility_grocery_and_pharmacy     0.085922   0.040160    2.139    0.0337 *
## mobility_parks                    0.398493   0.050820    7.841  3.36e-13 ***
## mobility_transit_stations         0.334226   0.131643    2.539    0.0119 *
## mobility_workplaces              -0.158838   0.116158   -1.367    0.1731
## mobility_residential             -0.017666   0.107047   -0.165    0.8691
## stringency_index                  0.058725   0.039843    1.474    0.1422
## average_temperature_celsius      -0.283097   0.519814   -0.545    0.5867
## minimum_temperature_celsius       0.276755   0.128687    2.151    0.0328 *
## maximum_temperature_celsius       0.133534   0.177472    0.752    0.4527
## rainfall_mm                       0.069966   0.026346    2.656    0.0086 **
## dew_point                        -0.119663   0.429537   -0.279    0.7809
## relative_humidity                 0.155029   0.186242    0.832    0.4062
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3332 on 186 degrees of freedom
## Multiple R-squared:  0.8987, Adjusted R-squared:  0.8905
## F-statistic: 109.9 on 15 and 186 DF,  p-value: < 2.2e-16
##
##
## Response new_hospitalized_patients :
##
## Call:
## lm(formula = new_hospitalized_patients ~ new_tested + cumulative_tested +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_pvd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.96927 -0.20214 -0.04932  0.18694  1.52254
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -0.002849   0.027600   -0.103  0.91791
## new_tested                      -0.112731   0.088135   -1.279  0.20246
## cumulative_tested                0.589470   0.097557    6.042 8.12e-09 ***
## mobility_retail_and_recreation   0.086488   0.106491    0.812  0.41773
## mobility_grocery_and_pharmacy   -0.020182   0.047009   -0.429  0.66818
## mobility_parks                  -0.141389   0.059487   -2.377  0.01848 *
## mobility_transit_stations       -0.294505   0.154093   -1.911  0.05752 .
## mobility_workplaces              0.172198   0.135968    1.266  0.20693
## mobility_residential             0.212715   0.125303    1.698  0.09125 .
## stringency_index                 0.361566   0.046638    7.753 5.71e-13 ***
## average_temperature_celsius      1.555136   0.608463    2.556  0.01139 *
## minimum_temperature_celsius     -0.211968   0.150634   -1.407  0.16104
## maximum_temperature_celsius     -0.583438   0.207738   -2.809  0.00551 **
## rainfall_mm                     -0.033480   0.030839   -1.086  0.27904
## dew_point                       -0.824876   0.502791   -1.641  0.10257
```

```
## relative_humidity                    0.225602    0.218004    1.035   0.30208
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3901 on 186 degrees of freedom
## Multiple R-squared:  0.8625,  Adjusted R-squared:  0.8514
## F-statistic: 77.77 on 15 and 186 DF,  p-value: < 2.2e-16
##
##
## Response cumulative_hospitalized_patients :
##
## Call:
## lm(formula = cumulative_hospitalized_patients ~ new_tested +
##     cumulative_tested + mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_pvd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.85908 -0.13574  0.01609  0.14741  0.52513
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    -0.003522   0.016022  -0.220   0.8263
## new_tested                      0.053928   0.051164   1.054   0.2932
## cumulative_tested               1.013102   0.056634  17.889  < 2e-16 ***
## mobility_retail_and_recreation -0.421119   0.061820  -6.812 1.29e-10 ***
## mobility_grocery_and_pharmacy   0.069318   0.027290   2.540   0.0119 *
## mobility_parks                  0.293614   0.034534   8.502 5.98e-15 ***
## mobility_transit_stations       0.183546   0.089455   2.052   0.0416 *
## mobility_workplaces            -0.086600   0.078932  -1.097   0.2740
## mobility_residential            0.004488   0.072741   0.062   0.9509
## stringency_index                0.067927   0.027075   2.509   0.0130 *
## average_temperature_celsius    -0.125968   0.353227  -0.357   0.7218
## minimum_temperature_celsius     0.180369   0.087446   2.063   0.0405 *
## maximum_temperature_celsius     0.039963   0.120597   0.331   0.7407
## rainfall_mm                     0.046113   0.017903   2.576   0.0108 *
## dew_point                      -0.090641   0.291882  -0.311   0.7565
## relative_humidity               0.108711   0.126557   0.859   0.3915
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2265 on 186 degrees of freedom
## Multiple R-squared:  0.9528,  Adjusted R-squared:  0.949
## F-statistic: 250.5 on 15 and 186 DF,  p-value: < 2.2e-16
##
##
## Response current_hospitalized_patients :
##
## Call:
## lm(formula = current_hospitalized_patients ~ new_tested + cumulative_tested +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
```

```
##       mobility_parks + mobility_transit_stations + mobility_workplaces +
##       mobility_residential + stringency_index + average_temperature_celsius +
##       minimum_temperature_celsius + maximum_temperature_celsius +
##       rainfall_mm + dew_point + relative_humidity, data = train_data_pvd)
##
## Residuals:
##       Min       1Q   Median       3Q      Max
## -0.94255 -0.18162  0.00397  0.17981  0.94561
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -0.0051504  0.0201317  -0.256 0.798360
## new_tested                    -0.2660115  0.0642871  -4.138 5.31e-05 ***
## cumulative_tested              0.7946679  0.0711599  11.167  < 2e-16 ***
## mobility_retail_and_recreation -0.2319745  0.0776762  -2.986 0.003203 **
## mobility_grocery_and_pharmacy  0.1297425  0.0342891   3.784 0.000208 ***
## mobility_parks                -0.0954085  0.0433910  -2.199 0.029125 *
## mobility_transit_stations     -0.3026921  0.1123983  -2.693 0.007728 **
## mobility_workplaces            0.0759785  0.0991772   0.766 0.444595
## mobility_residential           0.0971492  0.0913978   1.063 0.289193
## stringency_index               0.2720695  0.0340187   7.998 1.31e-13 ***
## average_temperature_celsius    0.4263302  0.4438234   0.961 0.338008
## minimum_temperature_celsius   -0.0852042  0.1098748  -0.775 0.439049
## maximum_temperature_celsius   -0.1598879  0.1515276  -1.055 0.292715
## rainfall_mm                   -0.0304601  0.0224945  -1.354 0.177342
## dew_point                     -0.2373113  0.3667442  -0.647 0.518381
## relative_humidity             -0.0007994  0.1590160  -0.005 0.995994
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2845 on 186 degrees of freedom
## Multiple R-squared:  0.9266, Adjusted R-squared:  0.9207
## F-statistic: 156.6 on 15 and 186 DF,  p-value: < 2.2e-16
##
##
## Response current_ventilator_patients :
##
## Call:
## lm(formula = current_ventilator_patients ~ new_tested + cumulative_tested +
##       mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##       mobility_parks + mobility_transit_stations + mobility_workplaces +
##       mobility_residential + stringency_index + average_temperature_celsius +
##       minimum_temperature_celsius + maximum_temperature_celsius +
##       rainfall_mm + dew_point + relative_humidity, data = train_data_pvd)
##
## Residuals:
##       Min       1Q   Median       3Q      Max
## -2.46925 -0.23440 -0.00395  0.21568  1.22179
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -0.004809   0.033806  -0.142 0.887046
## new_tested                    -0.245973   0.107955  -2.278 0.023834 *
## cumulative_tested              0.404160   0.119496   3.382 0.000877 ***
```

```
## mobility_retail_and_recreation -0.016238    0.130438   -0.124 0.901061
## mobility_grocery_and_pharmacy   0.030958    0.057580    0.538 0.591458
## mobility_parks                 -0.295806    0.072865   -4.060 7.23e-05 ***
## mobility_transit_stations      -0.275073    0.188746   -1.457 0.146699
## mobility_workplaces             0.032471    0.166544    0.195 0.845631
## mobility_residential            0.024864    0.153480    0.162 0.871481
## stringency_index               0.312762    0.057126    5.475 1.40e-07 ***
## average_temperature_celsius     0.355919    0.745293    0.478 0.633527
## minimum_temperature_celsius    -0.209749    0.184508   -1.137 0.257084
## maximum_temperature_celsius     0.027602    0.254454    0.108 0.913737
## rainfall_mm                    -0.049825    0.037774   -1.319 0.188785
## dew_point                      -0.234835    0.615858   -0.381 0.703406
## relative_humidity              -0.028515    0.267029   -0.107 0.915073
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4778 on 186 degrees of freedom
## Multiple R-squared:  0.7966, Adjusted R-squared:  0.7802
## F-statistic: 48.57 on 15 and 186 DF,  p-value: < 2.2e-16
```

```r
# Extracting estimates from summary
pvs <- summary(pvd_model)

estimate_pvsnc <- (coef(pvs)[[paste("Response", "new_confirmed")]])[,"Estimate"]
# Create a data frame for the estimates and their labels
estimate_data_pvsnc <- data.frame(
  Variable = names(estimate_pvsnc),
  Estimate = estimate_pvsnc
)


estimate_pvsnd <- (coef(pvs)[[paste("Response", "new_deceased")]])[,"Estimate"]
# Create a data frame for the estimates and their labels
estimate_data_pvsnd <- data.frame(
  Variable = names(estimate_pvsnd),
  Estimate = estimate_pvsnd
)


estimate_pvsnh <- (coef(pvs)[[paste("Response", "new_hospitalized_patients")]])[,"Estimate"]
# Create a data frame for the estimates and their labels
estimate_data_pvsnh <- data.frame(
  Variable = names(estimate_pvsnh),
  Estimate = estimate_pvsnh
)


eplot1 <- ggplot(estimate_data_pvsnc[-1,],
              aes(x = Estimate, y = fct_reorder(Variable, Estimate))) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "gray") +
  geom_point() +
  geom_errorbarh(aes(xmin = Estimate - 2 * abs(Estimate),
                   xmax = Estimate + 2 * abs(Estimate))) +
  labs(title = "Cases - Estimates", x = "Estimate") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 10, hjust = 0),
        axis.title.y = element_blank())
```

```
eplot2 <- ggplot(estimate_data_pvsnd[-1,],
                aes(x = Estimate, y = fct_reorder(Variable, Estimate))) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "gray") +
  geom_point() +
  geom_errorbarh(aes(xmin = Estimate - 2 * abs(Estimate),
                     xmax = Estimate + 2 * abs(Estimate))) +
  labs(title = "Deaths - Estimates", x = "Estimate") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 10, hjust = 0),
        axis.title.y = element_blank())

eplot3 <- ggplot(estimate_data_pvsnh[-1,],
                aes(x = Estimate, y = fct_reorder(Variable, Estimate))) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "gray") +
  geom_point() +
  geom_errorbarh(aes(xmin = Estimate - 2 * abs(Estimate),
                     xmax = Estimate + 2 * abs(Estimate))) +
  labs(title = "Hospitalizations - Estimates", x = "Estimate") +
  theme_minimal() +
  theme(axis.text.y = element_text(size = 10, hjust = 0),
        axis.title.y = element_blank())

# Display the plots
eplot1
```
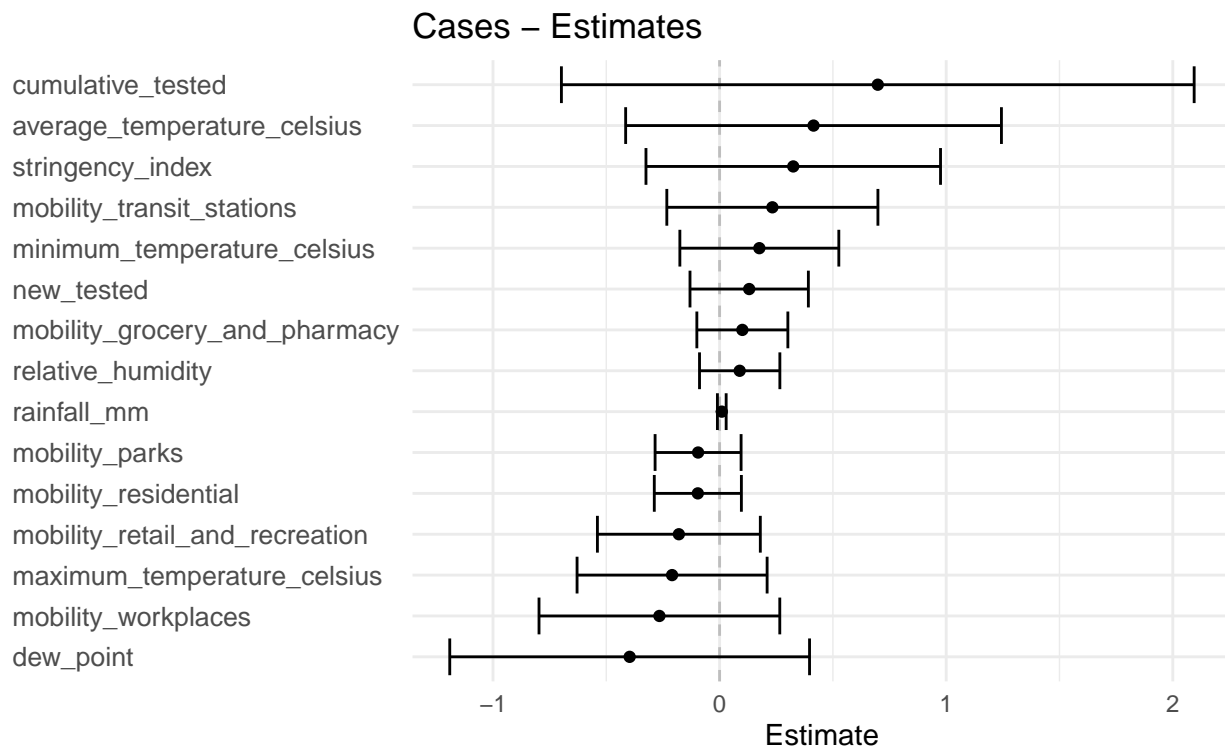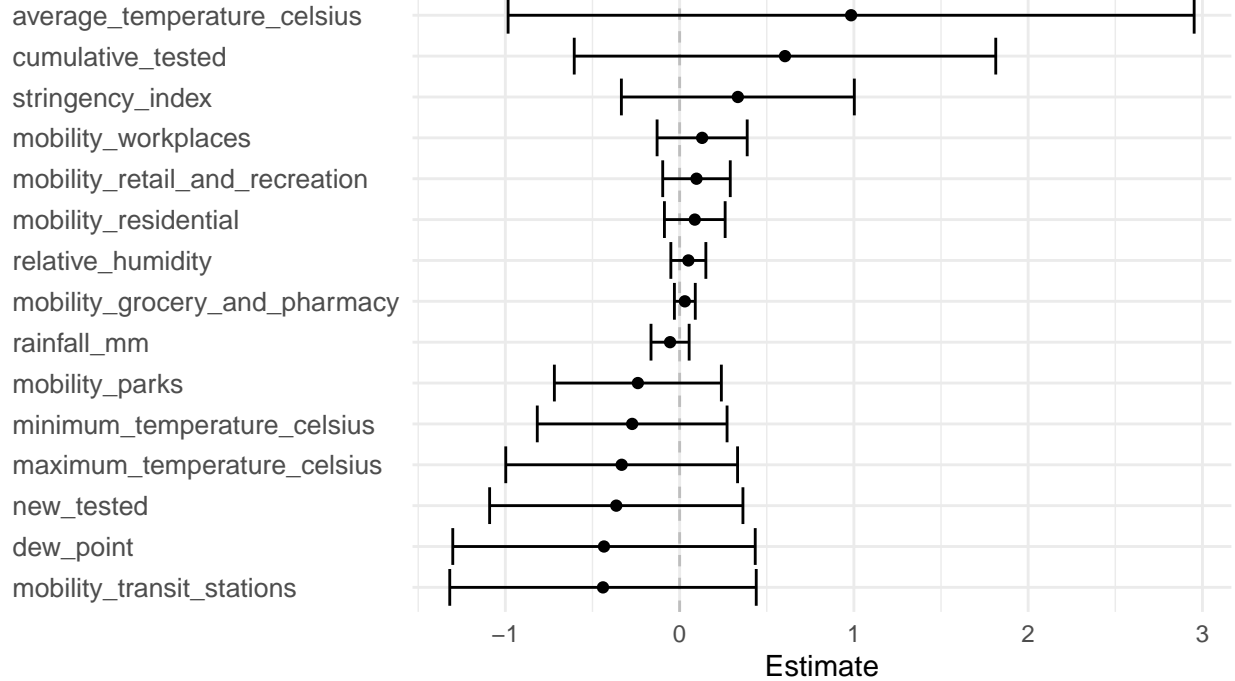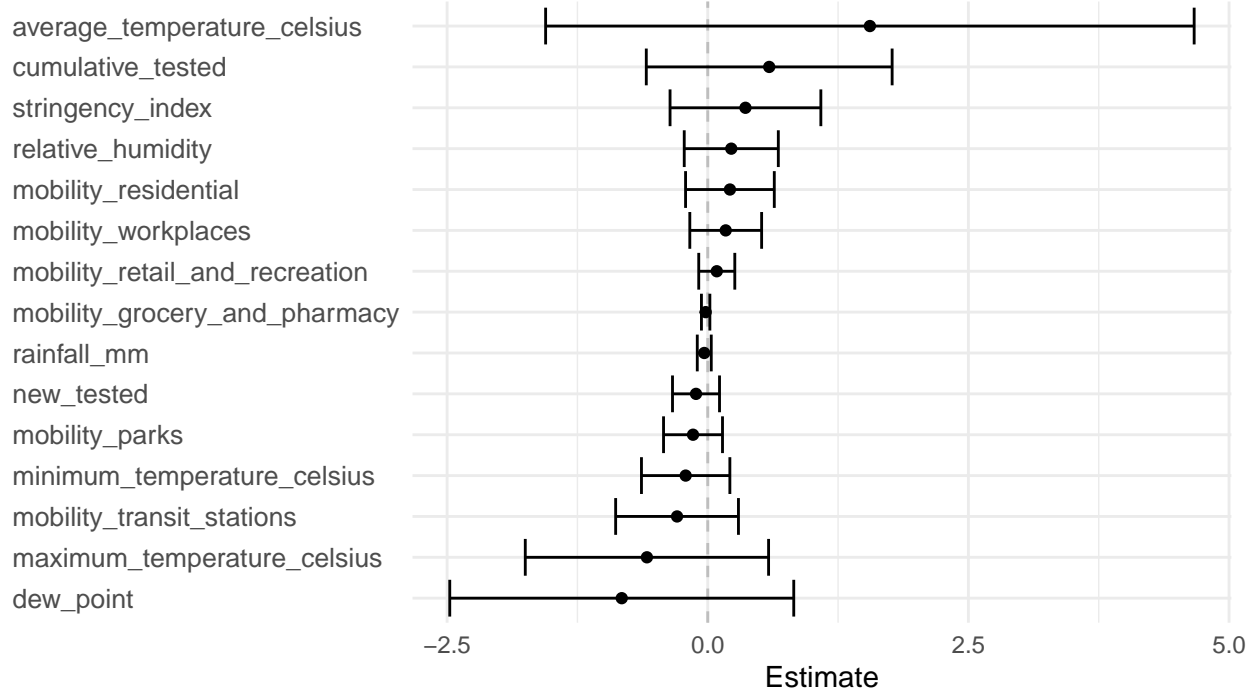

Cases – Estimates

```
eplot2
```

Deaths – Estimates

eplot3

Hospitalizations – Estimates

Interpretation: Summary of Pre Vaccination data - Multivariate Linear Regression model:

Let's interpret each response variable derived from our effective multivariate linear regression model. The results have revealed significant predictor variables, and we will now proceed to thoroughly examine the prediction summary for each response variable. By doing so, we can gain a comprehensive understanding of their respective implications for COVID-19 outcomes. The summary predictions of all response variables demonstrate an adjusted R-squared greater than 0.80, indicating a strong level of explanatory power in the model and excellent predictive performance. This suggests that the model is a highly effective fit for the data.

New Confirmed Cases:

The analysis for new confirmed cases indicates several significant predictor variables. **Cumulative tested, stringency index, and mobility grocery and pharmacy show statistically significant positive relationships** with new confirmed cases. This implies that higher cumulative testing rates and increased mobility to grocery and pharmacy places are associated with higher new confirmed cases and the main factor to note is despite of more stringent restrictions.

Conversely, **mobility in workplaces display significant negative relationships** with new confirmed cases, the interpretation should be cautious and not imply that returning to workplaces without precautions is universally advisable. Many factors can influence the relationship between workplace mobility and COVID-19 cases. For instance, certain workplaces may have strict safety measures in place, and adherence to public health guidelines can vary across different industries and regions. Therefore, it is essential to consider the broader context, the specific nature of workplaces, and the potential impact of various factors on the relationship between workplace mobility and COVID-19 cases.

New Deceased Cases:

The analysis for new deceased cases reveals few significant predictor variables, including test variables, mobility in parks and transit, and the stringency index. The number of new COVID-19 cases tested shows a negative correlation with the number of new deaths but this is not giving a considerable insight. But, on the other hand, **the cumulative number of tested cases demonstrates a positive association** with new deaths, as indicated by the coefficient estimate of 0.605.

The **mobility in parks and transit have negative association** indicating that decreased mobility reflects a increase in new confirmed deaths. But again we have to interpret this cautiously, as it's not advisable to roam around in parks or transit during pandemic.

Cumulative Confirmed Cases:

In the context of cumulative confirmed cases: Foremost, the **cumulative number of tested cases exhibits the most substantial positive impact, with a coefficient estimate of 1.063.** Moreover, we observe significant positive associations between confirmed cases and stringency index.

Interestingly, mobility in retail and recreation demonstrates a significant negative impact, implying that higher mobility in such settings associates with fewer cumulative confirmed COVID-19 cases, but here also we couldn't take this in account. Additionally, we find a moderate positive impact from mobility in grocery and pharmacy.

Cumulative Deceased Cases:

The cumulative number of deceased cases, similar to new deceased cases, is notably and **significantly associated with the cumulative number of tested individuals**. This finding suggests that a considerable proportion of deceased individuals likely underwent testing, indicating a direct correlation. The results imply that the number of confirmed deceased cases is closely related to the number of individuals who sought medical attention and received testing.

Other significant factors with **positive impact are mobility in parks, grocery and pharmacy and transit stations**. Higher mobility in public places is linked to an increase in cumulative deceased cases, possibly due to gatherings and reduced adherence to preventive measures.

An intriguing observation from our analysis is that **higher minimum temperatures and increased rainfall have a positive impact on the number of deaths**, confirming what we previously observed in our exploratory data analysis (EDA) visualizations. This implies that regions experiencing higher minimum temperatures and more rainfall may have a higher mortality rate from COVID-19.

New Hospitalized Patients:

Cumulative tested and stringency index show statistically significant positive relationships with new hospitalized patients. This implies that **higher cumulative testing, and more stringent restrictions may be associated with higher numbers of new hospitalized patients.** Higher average temperatures, as well as maximum temperatures, show positive associations with new confirmed and hospitalized cases. This suggests that warmer weather may contribute to an increase in COVID-19 transmission or affect people's behavior, leading to higher case counts. However, it is essential to approach this correlation cautiously and consider other contributing factors before drawing definitive conclusions.

Cumulative Hospitalized Patients:

Cumulative tested, mobility grocery and pharma, mobility transit and mobility parks show statistically significant positive relationships with cumulative hospitalized patients. This suggests that **higher cumulative testing, increased movement in grocery, pharmacy, parks and transit areas may lead to higher cumulative hospitalized patients**. Conversely, mobility retail and recreation, but again we couldn't come for direct conclusion here, without considering other factors.

Current Hospitalized Patients:

Cumulative tested, mobility grocery and pharmacy, and stringency index show statistically significant positive relationships with current hospitalized patients. This implies that **higher cumulative testing, increased movement in grocery and pharmacy, and more stringent restrictions may be associated with a higher number of current hospitalized patients**. However, mobility parks and mobility transit stations display significant negative relationships with current hospitalized patients, suggesting that increased movement in parks and transit stations may lead to a decrease in current hospitalized patients, and again an unaccountable relationship.

Current Ventilator Patients:

New tested shows a statistically significant positive relationship with current ventilator patients, implying that **higher number of new testing may be associated with a lower number of current ventilator patients**. On the other hand, cumulative tested and stringency index display statistically significant positive relationships with current ventilator patients, suggesting that current ventilator patients have direct correlation with cumulative tested.

MLR on Vaccination period data:

In both the pre-vaccination and vaccination data analyses, we followed a similar procedure to build multivariate linear regression models. While the predictor variables encompassed a wide range of factors such as testing metrics, mobility trends, and environmental features, the crucial vaccination variables are also included in the analysis of vaccination data. This allowed us to explore how vaccination rates may influence COVID-19 outcomes alongside other predictors.

Data Splitting:

```
# Define the predictor and response variables
# vd - vaccination data
predictors_vd <- vaccination_data %>%
  select(new_tested, cumulative_tested, new_persons_vaccinated,
         cumulative_persons_vaccinated, new_persons_fully_vaccinated,
         cumulative_persons_fully_vaccinated,
         mobility_retail_and_recreation,
```

```
            mobility_grocery_and_pharmacy, mobility_parks,
            mobility_transit_stations, mobility_workplaces,
            mobility_residential, stringency_index,
            average_temperature_celsius, minimum_temperature_celsius,
            maximum_temperature_celsius, rainfall_mm, dew_point,
            relative_humidity)

responses_vd <- vaccination_data %>%
  select(new_confirmed, new_deceased, cumulative_confirmed,
         cumulative_deceased, new_hospitalized_patients,
         cumulative_hospitalized_patients, current_hospitalized_patients,
         current_ventilator_patients)

# Scale the data
predictors_vd <- scale(predictors_vd)
responses_vd <- scale(responses_vd)

# Convert scaled variables back to a data frame
predictors_vd <- as.data.frame(predictors_vd)
responses_vd <- as.data.frame(responses_vd)

# Data splitting into train and test sets
set.seed(123)  # Set seed for reproducibility
train_index_vd <- createDataPartition(1:nrow(vaccination_data),
                                       p = 0.7, list = FALSE)
train_data_vd <- predictors_vd[train_index_vd, ]
train_response_vd <- responses_vd[train_index_vd, ]

test_data_vd <- predictors_vd[-train_index_vd, ]
test_response_vd <- responses_vd[-train_index_vd, ]
```

Similar to the pre-vaccination data, the "vaccination_data" is also split into **training (70%) and testing (30%)** sets for model development and evaluation. Predictor and response variables were carefully chosen, including vaccination metrics, testing rates, mobility patterns, and weather attributes. Z-score normalization is also applied to ensure standardization and comparability across all variables.

MLR - Model Building:

```
# Combine response variables into a multivariate response data frame
train_response_multivariate_vd <- as.matrix(train_response_vd)

# Build the MLR model using train data
vd_model <- lm(train_response_multivariate_vd ~ ., data = train_data_vd)
```

For the vaccination data, we adopted a similar model-building approach as for the pre-vaccination data. We transformed the response variables into a matrix format and utilized the "lm" function to create the multivariate linear regression model. Notably, this enabled us to **examine the influence of the vaccination variable alongside other predictors** on the COVID-19 outcomes.

MLR - Model Evaluation:

```
# Evaluate the model using test data
predictions_vd <- predict(vd_model, newdata = test_data_vd)
```

```r
# Combine response variables into a multivariate response data frame
test_response_multivariate_vd <- as.matrix(test_response_vd)

# Calculate Mean Squared Error (MSE) as an evaluation metric
mse_vd <- mean((predictions_vd - test_response_multivariate_vd)^2)
rmse_vd <- sqrt(mse_vd)

# Print the MSE and RMSE
cat("Mean Squared Error (MSE):", mse_vd, "\n")
```

```
## Mean Squared Error (MSE): 0.07249276
```

```r
cat("Root Mean Squared Error (RMSE):", rmse_vd, "\n")
```

```
## Root Mean Squared Error (RMSE): 0.2692448
```

Employed the same evaluation procedure for the MLR model using the test data. Predictions were generated for the "test_data_pvd" set using the "predict" function, and these predictions were compared to the true response variables in the "test_response_multivariate_pvd" data frame. The Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) values were then computed to assess the model's performance, resulting in an **MSE value of 0.07249276 and RMSE value of 0.2692448**.

The **MSE and RMSE for the vaccination data are lower than the error values of pre vaccination data**. This indicates that the multivariate linear regression model performs even more accurately in predicting COVID-19 outcomes for the vaccination data. The model's enhanced accuracy and ability to generalize to new and unseen data during the vaccination era are evident from this lower MSE value, solidifying its efficacy in understanding the factors influencing COVID-19 outcomes amidst the vaccination efforts.

MLR - Model Summary:

```r
summary(vd_model)
```

```
## Response new_confirmed :
##
## Call:
## lm(formula = new_confirmed ~ new_tested + cumulative_tested +
##     new_persons_vaccinated + cumulative_persons_vaccinated +
##     new_persons_fully_vaccinated + cumulative_persons_fully_vaccinated +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_vd)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3346 -0.2868 -0.0641  0.2074  3.3395
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      -0.01440    0.02642  -0.545  0.58603
## new_tested                        0.27529    0.03904   7.052 1.12e-11 ***
```

```
## cumulative_tested                     0.70608    0.13888    5.084 6.35e-07 ***
## new_persons_vaccinated               -0.16239    0.06124   -2.652  0.00842 **
## cumulative_persons_vaccinated        -0.44855    0.14262   -3.145  0.00182 **
## new_persons_fully_vaccinated          0.04961    0.04577    1.084  0.27925
## cumulative_persons_fully_vaccinated   1.19080    0.18661    6.381 6.27e-10 ***
## mobility_retail_and_recreation        0.07172    0.15080    0.476  0.63470
## mobility_grocery_and_pharmacy         0.10764    0.06567    1.639  0.10216
## mobility_parks                        0.11535    0.06993    1.649  0.10006
## mobility_transit_stations            -0.77678    0.12157   -6.389 5.99e-10 ***
## mobility_workplaces                   0.35439    0.11509    3.079  0.00226 **
## mobility_residential                  0.34253    0.14455    2.370  0.01841 *
## stringency_index                      0.55051    0.11964    4.601 6.09e-06 ***
## average_temperature_celsius           0.65564    0.36205    1.811  0.07110 .
## minimum_temperature_celsius          -0.38584    0.16255   -2.374  0.01821 *
## maximum_temperature_celsius          -0.25234    0.21394   -1.180  0.23908
## rainfall_mm                          -0.02454    0.03213   -0.764  0.44566
## dew_point                             0.03251    0.17607    0.185  0.85363
## relative_humidity                     0.03272    0.04960    0.660  0.50991
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.48 on 315 degrees of freedom
## Multiple R-squared:  0.7727, Adjusted R-squared:  0.759
## F-statistic: 56.35 on 19 and 315 DF,  p-value: < 2.2e-16
##
##
## Response new_deceased :
##
## Call:
## lm(formula = new_deceased ~ new_tested + cumulative_tested +
##     new_persons_vaccinated + cumulative_persons_vaccinated +
##     new_persons_fully_vaccinated + cumulative_persons_fully_vaccinated +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_vd)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6556 -0.1414 -0.0112  0.1554  0.9152
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          0.0070387  0.0128684   0.547 0.584783
## new_tested                           0.0542982  0.0190126   2.856 0.004576 **
## cumulative_tested                    0.4734400  0.0676412   6.999 1.56e-11 ***
## new_persons_vaccinated              -0.0998327  0.0298284  -3.347 0.000916 ***
## cumulative_persons_vaccinated       -2.9519100  0.0694606 -42.498  < 2e-16 ***
## new_persons_fully_vaccinated         0.0646797  0.0222934   2.901 0.003978 **
## cumulative_persons_fully_vaccinated  2.2368082  0.0908856  24.611  < 2e-16 ***
## mobility_retail_and_recreation      -0.3685495  0.0734442  -5.018 8.74e-07 ***
## mobility_grocery_and_pharmacy        0.0640660  0.0319819   2.003 0.046013 *
## mobility_parks                       0.1629358  0.0340594   4.784 2.65e-06 ***
```

```
## mobility_transit_stations                   0.1069990   0.0592125    1.807 0.071711 .
## mobility_workplaces                          0.0856314   0.0560562    1.528 0.127616
## mobility_residential                         0.1795517   0.0704028    2.550 0.011235 *
## stringency_index                             0.0031360   0.0582689    0.054 0.957113
## average_temperature_celsius                  0.2756795   0.1763331    1.563 0.118962
## minimum_temperature_celsius                 -0.1042282   0.0791706   -1.317 0.188963
## maximum_temperature_celsius                 -0.1589616   0.1041965   -1.526 0.128114
## rainfall_mm                                  0.0001325   0.0156497    0.008 0.993252
## dew_point                                   -0.0652423   0.0857532   -0.761 0.447336
## relative_humidity                            0.0266304   0.0241554    1.102 0.271104
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2338 on 315 degrees of freedom
## Multiple R-squared:  0.9452, Adjusted R-squared:  0.9419
## F-statistic: 286.1 on 19 and 315 DF,  p-value: < 2.2e-16
##
##
## Response cumulative_confirmed :
##
## Call:
## lm(formula = cumulative_confirmed ~ new_tested + cumulative_tested +
##     new_persons_vaccinated + cumulative_persons_vaccinated +
##     new_persons_fully_vaccinated + cumulative_persons_fully_vaccinated +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_vd)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.195573 -0.033498  0.000393  0.039337  0.142798
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        0.001351   0.003169    0.426 0.670100
## new_tested                        -0.016176   0.004682   -3.455 0.000625 ***
## cumulative_tested                  1.621658   0.016655   97.365  < 2e-16 ***
## new_persons_vaccinated             0.015472   0.007345    2.107 0.035944 *
## cumulative_persons_vaccinated     -0.278767   0.017103  -16.299  < 2e-16 ***
## new_persons_fully_vaccinated       0.009578   0.005489    1.745 0.081981 .
## cumulative_persons_fully_vaccinated -0.453278 0.022379  -20.255  < 2e-16 ***
## mobility_retail_and_recreation    -0.023543   0.018084   -1.302 0.193929
## mobility_grocery_and_pharmacy     -0.005710   0.007875   -0.725 0.468920
## mobility_parks                     0.083498   0.008387    9.956  < 2e-16 ***
## mobility_transit_stations         -0.041184   0.014580   -2.825 0.005034 **
## mobility_workplaces                0.103232   0.013803    7.479 7.50e-13 ***
## mobility_residential               0.101372   0.017335    5.848 1.25e-08 ***
## stringency_index                  -0.067598   0.014348   -4.711 3.70e-06 ***
## average_temperature_celsius        0.032276   0.043419    0.743 0.457821
## minimum_temperature_celsius        0.012174   0.019494    0.624 0.532770
## maximum_temperature_celsius       -0.014277   0.025657   -0.556 0.578283
## rainfall_mm                        0.005886   0.003853    1.527 0.127651
```

```
## dew_point                                 -0.002919   0.021115  -0.138 0.890124
## relative_humidity                         -0.001602   0.005948  -0.269 0.787861
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 0.05757 on 315 degrees of freedom
## Multiple R-squared:  0.997,  Adjusted R-squared:  0.9968
## F-statistic:  5432 on 19 and 315 DF,  p-value: < 2.2e-16
##
##
## Response cumulative_deceased :
##
## Call:
## lm(formula = cumulative_deceased ~ new_tested + cumulative_tested +
##     new_persons_vaccinated + cumulative_persons_vaccinated +
##     new_persons_fully_vaccinated + cumulative_persons_fully_vaccinated +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_vd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.47526 -0.05347 -0.00546  0.06652  0.34114
##
## Coefficients:
##                                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         0.003368   0.006561   0.513  0.60804
## new_tested                         -0.044698   0.009693  -4.611 5.83e-06 ***
## cumulative_tested                   1.263885   0.034486  36.649  < 2e-16 ***
## new_persons_vaccinated              0.080208   0.015208   5.274 2.48e-07 ***
## cumulative_persons_vaccinated       0.733720   0.035413  20.719  < 2e-16 ***
## new_persons_fully_vaccinated       -0.073600   0.011366  -6.476 3.63e-10 ***
## cumulative_persons_fully_vaccinated -1.163389   0.046337 -25.107  < 2e-16 ***
## mobility_retail_and_recreation      0.070328   0.037444   1.878  0.06128 .
## mobility_grocery_and_pharmacy       0.004752   0.016305   0.291  0.77089
## mobility_parks                     -0.039849   0.017365  -2.295  0.02240 *
## mobility_transit_stations           0.021664   0.030188   0.718  0.47353
## mobility_workplaces                -0.092915   0.028579  -3.251  0.00127 **
## mobility_residential               -0.109163   0.035894  -3.041  0.00255 **
## stringency_index                   -0.058035   0.029707  -1.954  0.05164 .
## average_temperature_celsius        -0.121925   0.089901  -1.356  0.17600
## minimum_temperature_celsius         0.016033   0.040364   0.397  0.69149
## maximum_temperature_celsius         0.030832   0.053123   0.580  0.56207
## rainfall_mm                        -0.010489   0.007979  -1.315  0.18960
## dew_point                           0.090612   0.043720   2.073  0.03903 *
## relative_humidity                  -0.022800   0.012315  -1.851  0.06505 .
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 0.1192 on 315 degrees of freedom
## Multiple R-squared:  0.9866, Adjusted R-squared:  0.9858
## F-statistic:  1218 on 19 and 315 DF,  p-value: < 2.2e-16
```

```
##
##
## Response new_hospitalized_patients :
##
## Call:
## lm(formula = new_hospitalized_patients ~ new_tested + cumulative_tested +
##     new_persons_vaccinated + cumulative_persons_vaccinated +
##     new_persons_fully_vaccinated + cumulative_persons_fully_vaccinated +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_vd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20390 -0.22832 -0.00837  0.19783  1.10949
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        -0.0016075  0.0185694  -0.087 0.931069
## new_tested                          0.1039487  0.0274357   3.789 0.000181 ***
## cumulative_tested                   0.9568674  0.0976084   9.803  < 2e-16 ***
## new_persons_vaccinated             -0.1692324  0.0430434  -3.932 0.000104 ***
## cumulative_persons_vaccinated      -2.5047918  0.1002337 -24.990  < 2e-16 ***
## new_persons_fully_vaccinated        0.0412611  0.0321701   1.283 0.200578
## cumulative_persons_fully_vaccinated 1.9545305  0.1311508  14.903  < 2e-16 ***
## mobility_retail_and_recreation     -0.3135056  0.1059823  -2.958 0.003330 **
## mobility_grocery_and_pharmacy       0.0165565  0.0461509   0.359 0.720024
## mobility_parks                      0.2842218  0.0491488   5.783 1.77e-08 ***
## mobility_transit_stations          -0.2410683  0.0854454  -2.821 0.005086 **
## mobility_workplaces                 0.3767056  0.0808908   4.657 4.74e-06 ***
## mobility_residential                0.5147332  0.1015934   5.067 6.91e-07 ***
## stringency_index                   -0.1100235  0.0840838  -1.308 0.191659
## average_temperature_celsius         0.3736718  0.2544541   1.469 0.142960
## minimum_temperature_celsius        -0.1341924  0.1142456  -1.175 0.241044
## maximum_temperature_celsius        -0.1828255  0.1503587  -1.216 0.224922
## rainfall_mm                         0.0004923  0.0225830   0.022 0.982621
## dew_point                           0.0019821  0.1237445   0.016 0.987231
## relative_humidity                   0.0278077  0.0348570   0.798 0.425608
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3374 on 315 degrees of freedom
## Multiple R-squared:  0.8867, Adjusted R-squared:  0.8798
## F-statistic: 129.7 on 19 and 315 DF,  p-value: < 2.2e-16
##
##
## Response cumulative_hospitalized_patients :
##
## Call:
## lm(formula = cumulative_hospitalized_patients ~ new_tested +
##     cumulative_tested + new_persons_vaccinated + cumulative_persons_vaccinated +
##     new_persons_fully_vaccinated + cumulative_persons_fully_vaccinated +
```

```
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_vd)
##
## Residuals:
##       Min       1Q    Median       3Q       Max
## -0.265196 -0.039978  0.002306  0.044578  0.192243
##
## Coefficients:
##                                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)                          0.003569   0.004577   0.780 0.436125
## new_tested                          -0.030991   0.006763  -4.582 6.63e-06 ***
## cumulative_tested                    1.360418   0.024061  56.541  < 2e-16 ***
## new_persons_vaccinated               0.055581   0.010610   5.238 2.97e-07 ***
## cumulative_persons_vaccinated        0.105008   0.024708   4.250 2.82e-05 ***
## new_persons_fully_vaccinated        -0.047745   0.007930  -6.021 4.83e-09 ***
## cumulative_persons_fully_vaccinated -0.711041   0.032329 -21.994  < 2e-16 ***
## mobility_retail_and_recreation      -0.026261   0.026125  -1.005 0.315562
## mobility_grocery_and_pharmacy        0.020953   0.011376   1.842 0.066447 .
## mobility_parks                       0.009218   0.012115   0.761 0.447316
## mobility_transit_stations            0.041893   0.021063   1.989 0.047570 *
## mobility_workplaces                 -0.073135   0.019940  -3.668 0.000287 ***
## mobility_residential                -0.071979   0.025043  -2.874 0.004326 **
## stringency_index                    -0.171429   0.020727  -8.271 3.79e-15 ***
## average_temperature_celsius         -0.041882   0.062724  -0.668 0.504800
## minimum_temperature_celsius         -0.001968   0.028162  -0.070 0.944330
## maximum_temperature_celsius         -0.007380   0.037064  -0.199 0.842299
## rainfall_mm                         -0.004896   0.005567  -0.879 0.379826
## dew_point                            0.057507   0.030503   1.885 0.060313 .
## relative_humidity                   -0.007436   0.008592  -0.865 0.387461
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08316 on 315 degrees of freedom
## Multiple R-squared:  0.9937, Adjusted R-squared:  0.9933
## F-statistic:  2599 on 19 and 315 DF,  p-value: < 2.2e-16
##
##
## Response current_hospitalized_patients :
##
## Call:
## lm(formula = current_hospitalized_patients ~ new_tested + cumulative_tested +
##     new_persons_vaccinated + cumulative_persons_vaccinated +
##     new_persons_fully_vaccinated + cumulative_persons_fully_vaccinated +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_vd)
##
## Residuals:
##       Min       1Q    Median       3Q       Max
```

```
## -0.72503 -0.15435  0.00437  0.14828  0.57009
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        0.010992   0.013130   0.837   0.4031
## new_tested                         0.025058   0.019399   1.292   0.1974
## cumulative_tested                  0.980432   0.069015  14.206  < 2e-16 ***
## new_persons_vaccinated            -0.077979   0.030434  -2.562   0.0109 *
## cumulative_persons_vaccinated     -2.471093   0.070871 -34.867  < 2e-16 ***
## new_persons_fully_vaccinated      -0.052034   0.022746  -2.288   0.0228 *
## cumulative_persons_fully_vaccinated 1.754056  0.092731  18.915  < 2e-16 ***
## mobility_retail_and_recreation    -0.438861   0.074936  -5.857 1.19e-08 ***
## mobility_grocery_and_pharmacy      0.063161   0.032631   1.936   0.0538 .
## mobility_parks                     0.303808   0.034751   8.742  < 2e-16 ***
## mobility_transit_stations          0.047160   0.060415   0.781   0.4356
## mobility_workplaces                0.288023   0.057195   5.036 8.02e-07 ***
## mobility_residential               0.456370   0.071832   6.353 7.37e-10 ***
## stringency_index                   0.055073   0.059452   0.926   0.3550
## average_temperature_celsius        0.217979   0.179914   1.212   0.2266
## minimum_temperature_celsius       -0.128557   0.080778  -1.591   0.1125
## maximum_temperature_celsius       -0.190607   0.106312  -1.793   0.0739 .
## rainfall_mm                       -0.006302   0.015967  -0.395   0.6934
## dew_point                          0.046608   0.087495   0.533   0.5946
## relative_humidity                  0.010569   0.024646   0.429   0.6683
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2385 on 315 degrees of freedom
## Multiple R-squared:  0.9451, Adjusted R-squared:  0.9418
## F-statistic: 285.3 on 19 and 315 DF,  p-value: < 2.2e-16
##
##
## Response current_ventilator_patients :
##
## Call:
## lm(formula = current_ventilator_patients ~ new_tested + cumulative_tested +
##     new_persons_vaccinated + cumulative_persons_vaccinated +
##     new_persons_fully_vaccinated + cumulative_persons_fully_vaccinated +
##     mobility_retail_and_recreation + mobility_grocery_and_pharmacy +
##     mobility_parks + mobility_transit_stations + mobility_workplaces +
##     mobility_residential + stringency_index + average_temperature_celsius +
##     minimum_temperature_celsius + maximum_temperature_celsius +
##     rainfall_mm + dew_point + relative_humidity, data = train_data_vd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.73239 -0.16278  0.02228  0.15236  0.59999
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        0.009462   0.012208   0.775  0.43888
## new_tested                        -0.014149   0.018037  -0.784  0.43338
## cumulative_tested                 -0.440358   0.064172  -6.862 3.61e-11 ***
## new_persons_vaccinated            -0.055003   0.028299  -1.944  0.05283 .
```

```
## cumulative_persons_vaccinated              -1.889291    0.065898 -28.670   < 2e-16 ***
## new_persons_fully_vaccinated               -0.252180    0.021150 -11.923   < 2e-16 ***
## cumulative_persons_fully_vaccinated         1.862901    0.086224  21.605   < 2e-16 ***
## mobility_retail_and_recreation             -0.321759    0.069678  -4.618 5.66e-06 ***
## mobility_grocery_and_pharmacy               0.081938    0.030342   2.701  0.00730 **
## mobility_parks                              0.056340    0.032313   1.744  0.08221 .
## mobility_transit_stations                   0.170223    0.056176   3.030  0.00265 **
## mobility_workplaces                        -0.047769    0.053181  -0.898  0.36975
## mobility_residential                        0.135243    0.066792   2.025  0.04373 *
## stringency_index                            0.136905    0.055281   2.477  0.01379 *
## average_temperature_celsius                -0.016183    0.167290  -0.097  0.92300
## minimum_temperature_celsius                -0.054090    0.075110  -0.720  0.47197
## maximum_temperature_celsius                -0.115608    0.098853  -1.170  0.24309
## rainfall_mm                                -0.032780    0.014847  -2.208  0.02798 *
## dew_point                                   0.088555    0.081355   1.088  0.27721
## relative_humidity                          -0.015714    0.022917  -0.686  0.49339
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2218 on 315 degrees of freedom
## Multiple R-squared:  0.952,  Adjusted R-squared:  0.9491
## F-statistic: 328.6 on 19 and 315 DF,  p-value: < 2.2e-16
```

```r
#Extracting Estimates from summary
vs <- summary(vd_model)

estimate_vsnc <- (coef(vs)[[paste("Response", "new_confirmed")]])[,"Estimate"]
# Create a data frame for the estimates and their labels
estimate_data_vsnc <- data.frame(
  Variable = names(estimate_vsnc),
  Estimate = estimate_vsnc
)


estimate_vsnd <- (coef(vs)[[paste("Response", "new_deceased")]])[,"Estimate"]
# Create a data frame for the estimates and their labels
estimate_data_vsnd <- data.frame(
  Variable = names(estimate_vsnd),
  Estimate = estimate_vsnd
)


estimate_vsnh <- (coef(vs)[[paste("Response", "new_hospitalized_patients")]])[,"Estimate"]
# Create a data frame for the estimates and their labels
estimate_data_vsnh <- data.frame(
  Variable = names(estimate_vsnh),
  Estimate = estimate_vsnh
)


eplot4 <- ggplot(estimate_data_vsnc[-1,],
             aes(x = Estimate, y = fct_reorder(Variable, Estimate))) +
  geom_vline(xintercept = 0, linetype = "dashed", color = "gray") +
  geom_point() +
  geom_errorbarh(aes(xmin = Estimate - 2 * abs(Estimate),
                  xmax = Estimate + 2 * abs(Estimate))) +
  labs(title = "Cases - Estimates", x = "Estimate") +
```
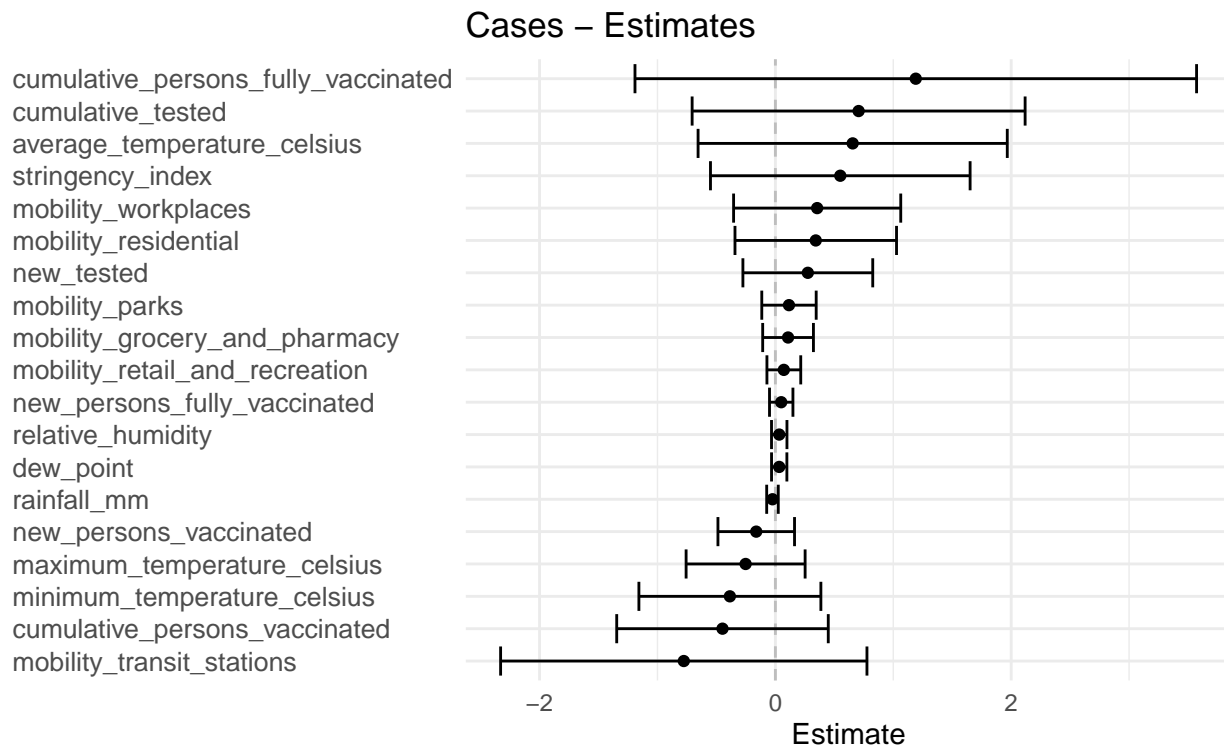
```
    theme_minimal() +
    theme(axis.text.y = element_text(size = 10, hjust = 0),
          axis.title.y = element_blank())

eplot5 <- ggplot(estimate_data_vsnd[-1,],
                 aes(x = Estimate, y = fct_reorder(Variable, Estimate))) +
    geom_vline(xintercept = 0, linetype = "dashed", color = "gray") +
    geom_point() +
    geom_errorbarh(aes(xmin = Estimate - 2 * abs(Estimate),
                       xmax = Estimate + 2 * abs(Estimate))) +
    labs(title = "Deaths - Estimates", x = "Estimate") +
    theme_minimal() +
    theme(axis.text.y = element_text(size = 10, hjust = 0),
          axis.title.y = element_blank())

eplot6 <- ggplot(estimate_data_vsnh,
                 aes(x = Estimate, y = fct_reorder(Variable, Estimate))) +
    geom_vline(xintercept = 0, linetype = "dashed", color = "gray") +
    geom_point() +
    geom_errorbarh(aes(xmin = Estimate - 2 * abs(Estimate),
                       xmax = Estimate + 2 * abs(Estimate))) +
    labs(title = "Hospitalizations - Estimates", x = "Estimate") +
    theme_minimal() +
    theme(axis.text.y = element_text(size = 10, hjust = 0),
          axis.title.y = element_blank())

# Display the plots
eplot4
```
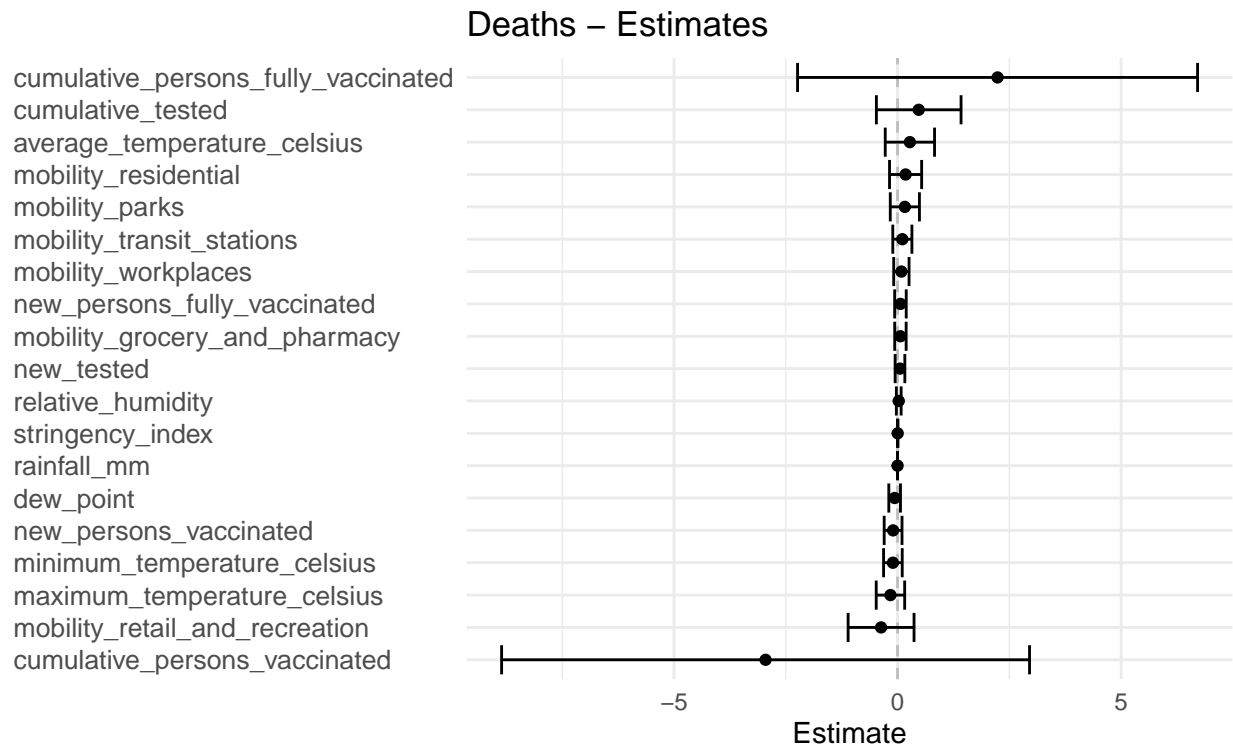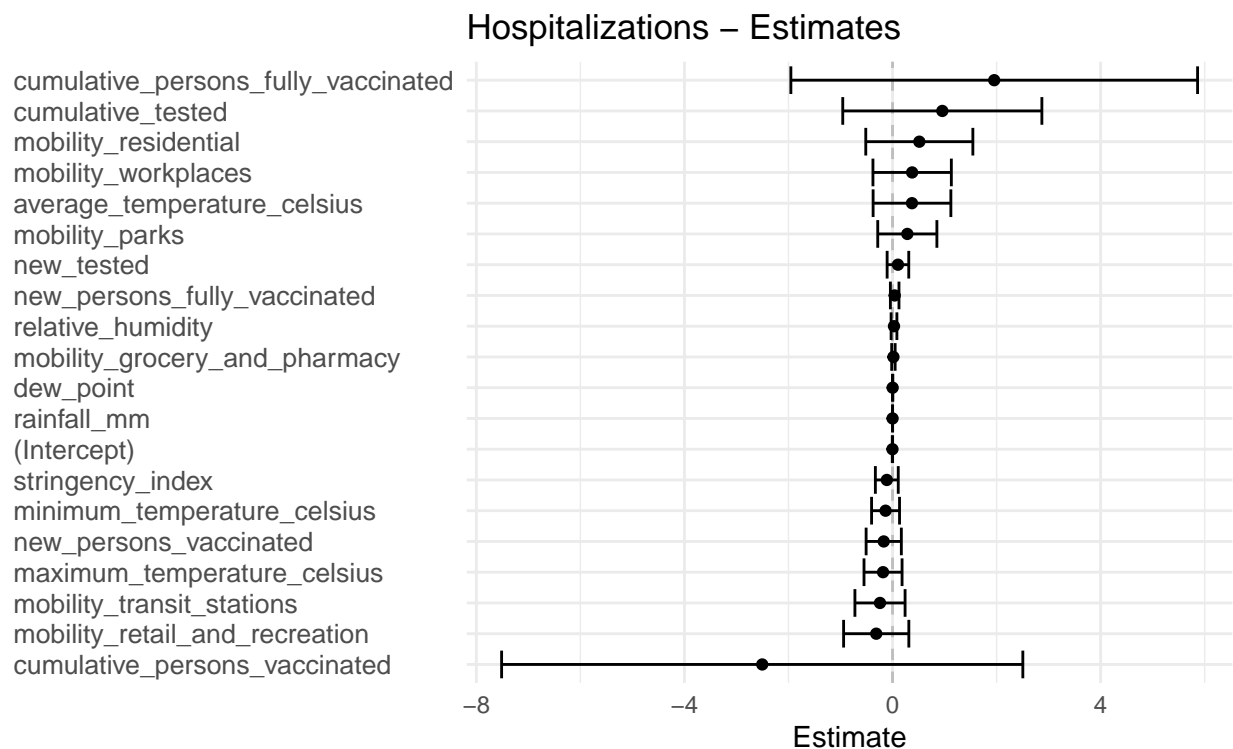

Cases – Estimates

```
eplot5
```

Deaths – Estimates

cumulative_persons_fully_vaccinated
cumulative_tested
average_temperature_celsius
mobility_residential
mobility_parks
mobility_transit_stations
mobility_workplaces
new_persons_fully_vaccinated
mobility_grocery_and_pharmacy
new_tested
relative_humidity
stringency_index
rainfall_mm
dew_point
new_persons_vaccinated
minimum_temperature_celsius
maximum_temperature_celsius
mobility_retail_and_recreation
cumulative_persons_vaccinated

−5   0   5
Estimate

```
eplot6
```

Hospitalizations – Estimates

cumulative_persons_fully_vaccinated
cumulative_tested
mobility_residential
mobility_workplaces
average_temperature_celsius
mobility_parks
new_tested
new_persons_fully_vaccinated
relative_humidity
mobility_grocery_and_pharmacy
dew_point
rainfall_mm
(Intercept)
stringency_index
minimum_temperature_celsius
new_persons_vaccinated
maximum_temperature_celsius
mobility_transit_stations
mobility_retail_and_recreation
cumulative_persons_vaccinated

−8   −4   0   4
Estimate

Interpretation: Summary of Vaccination data - Multivariate Linear Regression model:

The model's summary presents the coefficients of all predictor variables, and it demonstrates excellent statistical measures, particularly notable R-squared values, indicating the model's effectiveness. These results imply that a significant portion of the variance in the response variables can be explained by the predictors. Now, let's delve into each response variable's corresponding significant factors and examine their relationships in detail.

New confirmed:

The predictor variables new_tested and cumulative_tested showed a positive association with new_confirmed, indicating that increased testing is linked to more confirmed cases. On the other hand, **new_persons_vaccinated and cumulative_persons_vaccinated exhibited negative relationships with new_confirmed**, suggesting that higher vaccination rates are associated with fewer new cases. Moreover, mobility_workplaces and stringency_index showed positive associations. Conversely, minimum_temperature_celsius had a negative relationship with new_confirmed.

New Deceased:

The result revealed significant predictors associated with higher new COVID-19 deaths, including the variables new_tested and cumulative_tested. Conversely, **new_persons_vaccinated and "cumulative_persons_vaccinated" were linked to lower fatality rates**. Surprisingly, complete vaccination (new_persons_fully_vaccinated and cumulative_persons_fully_vaccinated) showed substantial positive impacts on new deaths, which could be attributed to an increase in the count due to potential delays in completing the vaccination process. Additionally, reduced mobility in grocery and pharmacy, residential and parks was associated with higher fatalities, while increased retail and recreation mobility showed the opposite trend.

Cumulative confirmed:

Variables new_tested, cumulative_persons_vaccinated and cumulative_persons_fully_vaccinated showed negative coefficients, suggesting that an **increase in testing and vaccination have led to a lower cumulative case count**. On the other hand, new_persons_vaccinated and new_persons_fully_vaccinated variables displayed positive coefficients, indicating that higher vaccination rates were associated with a higher number of confirmed cases, but the coefficient values of these variables are less 0.1 indicating the influence of other factors here. Notably, mobility_parks, mobility_workplaces, and mobility_residential had positive effects on the cumulative confirmed cases. **Stringency index have the negative coefficient** indicating the negative relationship with the cumulative cases confirmed.

Cumulative Deceased:

Predictors **new_tested, new_persons_fully_vaccinated, & cumulative_persons_fully_vaccinated have negative coefficients**. An increase in testing, higher rates of complete vaccinations were associated with a decrease in cumulative deaths. Conversely, cumulative_tested, cumulative_persons_vaccinated, and dew_point displayed positive coefficients, suggesting that the total number of tests conducted, cumulative vaccinations, and dew point temperature may be associated with higher cumulative deaths. Here, we can leave out the positive significance of cumulative testing and vaccination since the new count of testing and vaccination have negative impact on deaths, but the important factor to observe is that the dew point shows positive significance in bot new and cumulative deceased.

New Hospitalized Patients:

The model predicts that new hospitalized patients are **positively correlated with new testing and cumulative testing**. An increase in **new persons and cumulative person vaccinated have negative association** with new hospitalizations. Mobility in workplace, residential, and parks showed positive correlations, suggesting that reduced movement in these areas lead to fewer hospitalizations.

Cumulative Hospitalized Patients:

The model indicates a strong positive relationship between cumulative hospitalized patients and cumulative testing. An **increase in fully vaccinated individuals is negatively correlated with cumulative**

hospitalizations. Mobility in grocery, pharmacy and transit stations demonstrated positive correlations, implying that increased movement in these areas lead to more hospitalizations. Notably, higher stringency index showed a negative correlation, suggesting that stringent measures might limit the overall number of hospitalizations.

Current Hospitalized Patients:

Number of currently hospitalized patients is positively correlated with cumulative testing. An **increase in fully vaccinated individuals is negatively associated with the current number of hospitalizations also**. Mobility in grocery, pharmacy, workplace, residential and parks showed positive correlations, suggesting that reduced movement in these areas lead to fewer current hospitalizations.

Current Ventilator Patients:

Number of current ventilator patients is **negatively correlated with cumulative testing and all the other vaccination factors**, indicating the increase in testing and vaccinations reduces the number of ventilator patients. As like other hospital patients response variables, this too have positive correlation with mobility factors like mobility_transit, mobility_grocery_and_pharmacy and mobility_parks.

Disease Modeling Conclusion:

Key Findings and Interpretations:

Testing: Direct Correlation Across Metrics:

In the pre-vaccination period, the maximum positive coefficient is observed for new_tested and cumulative_tested variables. This positive effect can be attributed to the fact that the number of COVID-19 cases and deaths is likely influenced by the individuals who have undergone testing. Thus, an increase in testing leads to higher reported COVID-19 metrics. In the vaccination period, we observe a mix of positive and negative coefficients, highlighting the combined influence of vaccination and testing. However, testing remains a significant factor in detecting and preventing fatalities even with the vaccination drive in place.

Stringency Index - A Mixed Picture:

In the pre-vaccination phase, the positive coefficients for stringency index highlight the challenges of containing the virus solely through government-imposed measures. While strict restrictions may curb cases to some extent, the influence of other factors contributes to increased COVID-19 metrics. In contrast, the negative coefficients for stringency index in vaccination data indicate that vaccination, along with stringent measures, is crucial for mitigating the infection spread.

Mobility: A Significant Contributor:

Mobility factors, particularly in public areas, significantly impact COVID-19 metrics. While some mobility variables showed positive coefficients, suggesting reduced movement is associated with fewer cases, other mobility factors demonstrated negative coefficients. However, it's essential to recognize that mobility alone cannot be the sole reason for fluctuations, as other contributing factors also play pivotal roles.

Vaccination: A Game-Changer

The vaccination variables showcased substantial negative impacts on COVID-19 metrics. Both Exploratory Data Analysis and Multivariate Linear Regression on Vaccination data revealed that higher rates of vaccination, both partially and fully vaccinated individuals, were associated with lower confirmed cases, deceased cases, hospitalized patients and ventilator patients. Vaccination is a crucial tool in disease control.

Weather: An Indirect Influence:

Weather factors, such as higher temperatures, dew points, and rainfall, demonstrated correlations with COVID-19 metrics. Although not the primary influencers, they may indirectly affect cases, hospitalizations, and deaths. Warmer weather showed a favorable correlation with lower infection rates, where as colder weather positively associated with the infections and deaths.

In summary, this comprehensive disease modeling project unveiled significant predictors and their plausible implications for COVID-19 outcomes.

o **Testing emerged as a direct influencer across metrics.**

o **Vaccination played a pivotal role in mitigating the virus's impact.**

o **Mobility in public areas demonstrated its significance, and the interplay of various factors required cautious consideration.**

o **Government-imposed measures, coupled with vaccination, emerged as effective strategies in disease control.**