Measles, Mumps and Rubella cases and vaccination report

Dr Tyler Charles Baverstock, PhD

Date: 2025-01-11

Personal data portfolio

Contents

```r
#Set CRAN mirror
options(repos = c(CRAN = "https://cloud.r-project.org"))

#Installing required packages
install.packages("tidyverse")
install.packages("readr")
install.packages("dplyr")
install.packages("ggplot2")
install.packages("ggthemes")
install.packages("forecast")
install.packages("rnaturalearth")
install.packages("rnaturalearthdata")
install.packages("tinytex")
install.packages("bookdown")
```

```r
knitr::opts_chunk$set(echo = TRUE)
library(conflicted)
library(tidyverse)
library(readr)
library(dplyr)
library(ggplot2)
library(ggthemes)
library(lubridate)
library(forecast)
library(zoo)
library(rnaturalearth)
library(rnaturalearthdata)
```

```r
Measles_cases <- read_csv("Measles number reported cases.csv")
Mumps_cases <- read_csv("Mumps number reported cases.csv")
Rubella_cases <- read_csv("Rubella number reported cases.csv")

Measles_cases_cleaned <- Measles_cases %>% select(-IndicatorCode, -ValueType,
`Location type`, -`Period type`, -IsLatestYear,
-`Dim1 type`, -Dim1, -Dim1ValueCode, -`Dim2 type`,
-Dim2, -Dim2ValueCode, -`Dim3 type`, -Dim3, -Dim3ValueCode,
-DataSourceDimValueCode, -DataSource, -FactValueNumericPrefix,
-FactValueUoM, -FactValueNumericLowPrefix,
-FactValueNumericLowPrefix,
-FactValueNumericLow, -FactValueNumericHighPrefix,
-FactValueNumericHigh,-FactValueNumericHigh, -FactValueTranslationID,
-FactComments, -Language, -DateModified, -FactValueNumeric)

Mumps_cases_cleaned <- Mumps_cases %>% select(-IndicatorCode, -ValueType,
`Location type`, -`Period type`, -IsLatestYear,
-`Dim1 type`, -Dim1, -Dim1ValueCode, -`Dim2 type`,
-Dim2, -Dim2ValueCode, -`Dim3 type`, -Dim3, -Dim3ValueCode,
-DataSourceDimValueCode, -DataSource, -FactValueNumericPrefix,
-FactValueUoM, -FactValueNumericLowPrefix,
-FactValueNumericLowPrefix, -FactValueNumericLow,
-FactValueNumericHighPrefix, -FactValueNumericHigh, -FactValueNumericHigh,
-FactValueTranslationID, -FactComments,
-Language, -DateModified, -FactValueNumeric)
```

```
Rubella_cases_cleaned <- Rubella_cases %>% select(-IndicatorCode, -ValueType,
`Location type`, -`Period type`, -IsLatestYear,
-`Dim1 type`, -Dim1, -Dim1ValueCode, -`Dim2 type`,
-Dim2, -Dim2ValueCode, -`Dim3 type`, -Dim3, -Dim3ValueCode,
-DataSourceDimValueCode, -DataSource, -FactValueNumericPrefix,
-FactValueUoM, -FactValueNumericLowPrefix,
-FactValueNumericLowPrefix, -FactValueNumericLow,
-FactValueNumericHighPrefix, -FactValueNumericHigh, -FactValueNumericHigh,
-FactValueTranslationID, -FactComments, -Language,
-DateModified, -FactValueNumeric)

Measles_cases_cleaned <- na.omit(Measles_cases_cleaned)
Mumps_cases_cleaned <- na.omit(Mumps_cases_cleaned)
Rubella_cases_cleaned <- na.omit(Rubella_cases_cleaned)

Measles_cases_cleaned$Value <- as.numeric(Measles_cases_cleaned$Value)
Mumps_cases_cleaned$Value <- as.numeric(Mumps_cases_cleaned$Value)
Rubella_cases_cleaned$Value <- as.numeric(Rubella_cases_cleaned$Value)

Measles_cases_cleaned$Indicator <- gsub("Measles - number of reported cases",
"Measles", Measles_cases_cleaned$Indicator)
Mumps_cases_cleaned$Indicator <- gsub("Mumps - number of reported cases",
"Mumps", Mumps_cases_cleaned$Indicator)
Rubella_cases_cleaned$Indicator <- gsub("Rubella - number of reported cases",
"Rubella", Rubella_cases_cleaned$Indicator)
```

```
MCV1_coverage <- read_csv("MCV1_vaccination_coverage.csv")
MCV2_coverage <- read_csv("MCV2_vaccination_coverage.csv")
Rubella_vaccine_coverage <- read_csv("Rubella_vaccination_coverage.csv")

MCV1_coverage_cleaned <- MCV1_coverage %>% select(-GROUP,
-ANTIGEN_DESCRIPTION, -COVERAGE_CATEGORY, -COVERAGE_CATEGORY_DESCRIPTION)

MCV2_coverage_cleaned <- MCV2_coverage %>% select(-GROUP,
-ANTIGEN_DESCRIPTION, -COVERAGE_CATEGORY, -COVERAGE_CATEGORY_DESCRIPTION)

Rubella_vaccine_coverage_cleaned <- Rubella_vaccine_coverage %>% select(-GROUP,
-ANTIGEN_DESCRIPTION, -COVERAGE_CATEGORY, -COVERAGE_CATEGORY_DESCRIPTION)

MCV1_coverage_cleaned <- MCV1_coverage_cleaned %>%
  dplyr::filter(!is.na(COVERAGE)) %>%
  dplyr::filter(!is.na(DOSES))

MCV2_coverage_cleaned <- MCV2_coverage_cleaned %>%
  dplyr::filter(!is.na(COVERAGE)) %>%
  dplyr::filter(!is.na(DOSES))

Rubella_vaccine_coverage_cleaned <- Rubella_vaccine_coverage_cleaned %>%
  dplyr::filter(!is.na(COVERAGE))
```

```
#2000
MCV1_coverage_2000 <- MCV1_coverage_cleaned %>%
  dplyr::filter(YEAR == 2000)  %>%
```

```
  dplyr::rename(MCV1COVERAGE2000 = COVERAGE)

MCV2_coverage_2000 <- MCV2_coverage_cleaned %>%
  dplyr::filter(YEAR == 2000) %>%
  dplyr::rename(MCV2COVERAGE2000 = COVERAGE)

Rubella_vaccination_2000 <- Rubella_vaccine_coverage_cleaned %>%
  dplyr::filter(YEAR == 2000) %>%
  dplyr::rename(RUBELLACOVERAGE2000 = COVERAGE)

top_10_MCV1_coverage_2000 <- MCV1_coverage_2000 %>%
  arrange(desc(MCV1COVERAGE2000)) %>%
  slice_head(n = 10)

top_10_MCV2_coverage_2000 <- MCV2_coverage_2000 %>%
  arrange(desc(MCV2COVERAGE2000)) %>%
  slice_head(n = 10)

top_10_rubella_vaccination_2000 <- Rubella_vaccination_2000 %>%
  arrange(desc(RUBELLACOVERAGE2000)) %>%
  slice_head(n = 10)

#2023
MCV1_coverage_2023 <- MCV1_coverage_cleaned %>%
  dplyr::filter(YEAR == 2023)  %>%
  dplyr::rename(MCV1COVERAGE2023 = COVERAGE)

MCV2_coverage_2023 <- MCV2_coverage_cleaned %>%
  dplyr::filter(YEAR == 2023) %>%
  dplyr::rename(MCV2COVERAGE2023 = COVERAGE)

Rubella_vaccination_2023 <- Rubella_vaccine_coverage_cleaned %>%
  dplyr::filter(YEAR == 2023) %>%
  dplyr::rename(RUBELLACOVERAGE2023 = COVERAGE)

top_10_MCV1_coverage_2023 <- MCV1_coverage_2023 %>%
  arrange(desc(MCV1COVERAGE2023)) %>%
  slice_head(n = 10)

top_10_MCV2_coverage_2023 <- MCV2_coverage_2023 %>%
  arrange(desc(MCV2COVERAGE2023)) %>%
  slice_head(n = 10)

top_10_rubella_vaccination_2023 <- Rubella_vaccination_2023 %>%
  arrange(desc(RUBELLACOVERAGE2023)) %>%
  slice_head(n = 10)
```

1 Preventable disease analysis

1.1 Introduction

This report demonstrates my ability to load, clean, visualize, and analyze data in RStudio. It features a comprehensive analysis of a dataset from the World Health Organization (WHO). The full code for this analysis is available on my portfolio website: www.tylerbaverstockkay.com.

Vaccination is one of the most effective ways to prevent the spread of infectious diseases. The Measles, Mumps, and Rubella (MMR) vaccine, for example, is highly effective in reducing both the incidence and severity of these diseases. However, global vaccination rates for the MMR vaccine remain relatively low due to confusion about side effects, a lack of public understanding, and various social and economic challenges.

Understanding the relationship between vaccine coverage and disease incidence is crucial for assessing the effectiveness of herd immunity and preventing outbreaks. This report investigates vaccination and disease case data to identify connections between countries and populations. By examining factors that affect both vaccination rates and disease cases, we can better understand the public health challenges faced by different regions.

1.2 Infectious diseases over time: Measles, Mumps and Rubella

Measles, Mumps, and Rubella are highly infectious viral diseases that have spread globally. Although vaccines for all three diseases have been available since the late 1900s, the number of cases has fluctuated due to various country-specific factors and global events or influences. In this section, the number of cases for all three diseases will be analysed with reference to time and country, examining how disease cases have changed from the year 2000 to 2023.

```r
#Summarise number of Measles cases per year
Measles_cases_summary <- Measles_cases_cleaned %>%
  group_by(Period) %>%
  summarise(total_cases = sum(Value, na.rm = TRUE))
Measles_cases_summary <- as.data.frame(Measles_cases_summary)

#Summarise number of Mumps cases per year
Mumps_cases_summary <- Mumps_cases_cleaned %>%
  group_by(Period) %>%
  summarise(total_cases = sum(Value, na.rm = TRUE))

#Summarise number of Rubella cases per year
Rubella_cases_summary <- Rubella_cases_cleaned %>%
  group_by(Period) %>%
  summarise(total_cases = sum(Value, na.rm = TRUE))
#Measles/Mumps/Rubella cases per year
ggplot() +
  geom_line(data = Measles_cases_summary, aes(x=Period, y=total_cases,
  color = "Measles"), size = 1.5) +
  geom_point(data = Measles_cases_summary, aes(x=Period, y=total_cases,
  color = "Measles"), size = 3) +
  geom_line(data = Mumps_cases_summary, aes(x=Period, y=total_cases,
  color = "Mumps"), size = 1.5) +
  geom_point(data = Mumps_cases_summary, aes(x=Period, y=total_cases,
  color = "Mumps"), size = 3) +
  geom_line(data = Rubella_cases_summary, aes(x=Period, y=total_cases,
  color = "Rubella"), size = 1.5) +
  geom_point(data = Rubella_cases_summary, aes(x=Period, y=total_cases,
  color = "Rubella"), size = 3) +
  labs(title = "Total Cases",
       x= "Year",
       y= "Cases",
       color = "Disease") +
   scale_y_continuous(labels = scales::label_number(scale = 1e-3, suffix = "k"),
  limits = c(0, 300000)) +
  scale_x_continuous(breaks = seq(min(Measles_cases_summary$Period),
  max(Measles_cases_summary$Period), by = 4)) +
  theme_stata()
```

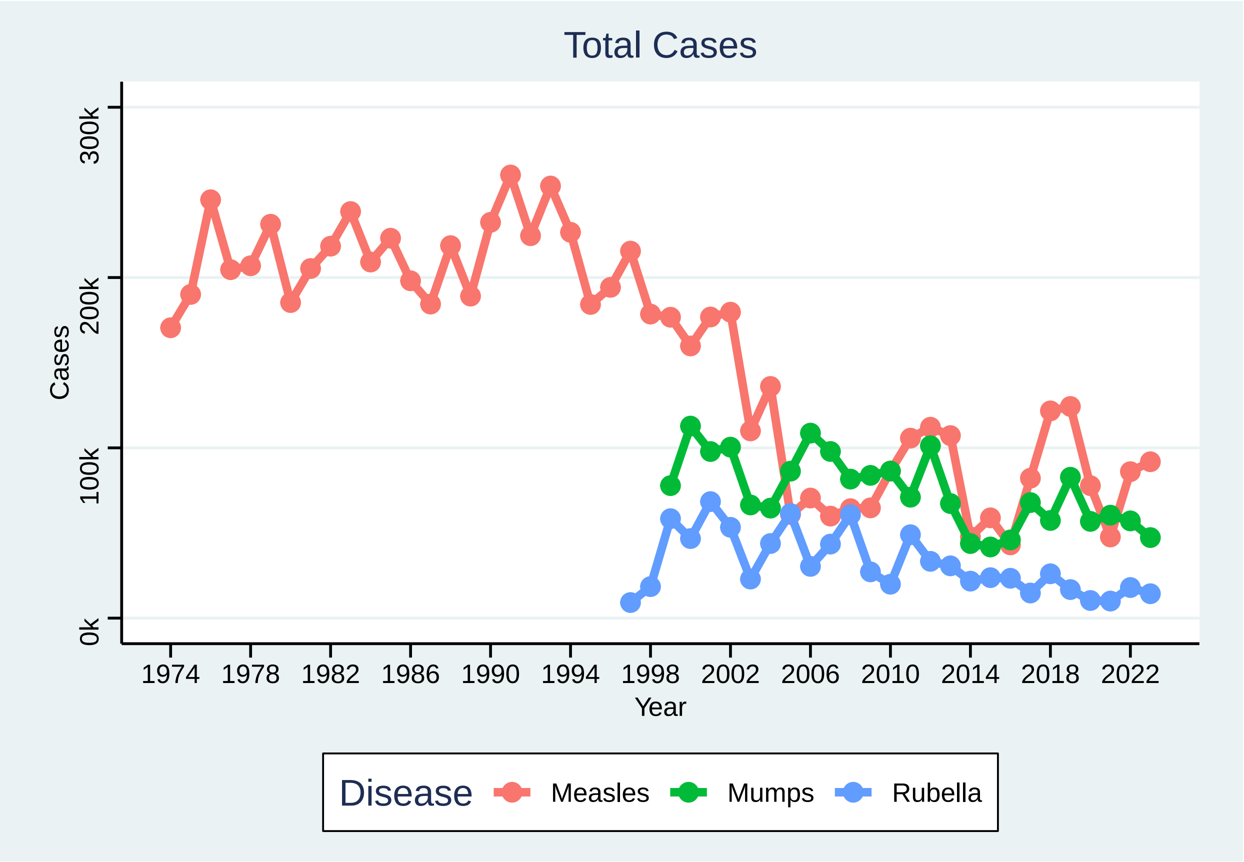


Figure 1: Global cases over time

To begin with, (**figure** 1) shows how the number of global cases of Measles, Mumps, and Rubella has changed over time. The WHO database begins recording Measles cases in 1974, with Rubella starting in 1997 and Mumps in 1999. Global Measles cases were initially between 200,000 and 250,000 until 1999, when a steady decrease in cases is observed, followed by unusual spikes in 2012 and 2019.

Conversely, both Mumps and Rubella cases were predominantly below 100,000, with Rubella consistently having the lowest number of cases among the three diseases in this report. Both diseases, however, show minor spikes in 2012 and 2019.

1.3 Disease cases

To understand a disease, it is important to investigate both the number of cases and their locations. This can provide valuable insight into how a disease spreads, as well as the factors that may increase or decrease transmission.

```r
Measles_cases_2000 <- Measles_cases_cleaned %>%
  dplyr::filter(Period == 2000)

Mumps_cases_2000 <- Mumps_cases_cleaned %>%
  dplyr::filter(Period == 2000)

Rubella_cases_2000 <- Rubella_cases_cleaned %>%
  dplyr::filter(Period == 2000)

top_10_measles_cases_2000 <- Measles_cases_2000 %>%
  arrange(desc(Value)) %>%
  slice_head(n = 10)

top_10_mumps_cases_2000 <- Mumps_cases_2000 %>%
  arrange(desc(Value)) %>%
  slice_head(n = 10)

top_10_rubella_cases_2000 <- Rubella_cases_2000 %>%
  arrange(desc(Value)) %>%
  slice_head(n = 10)

# Reorder the locations within each dataset based on the Value
top_10_measles_cases_2000$Location <-
with(top_10_measles_cases_2000, reorder(Location, Value))
top_10_mumps_cases_2000$Location <-
with(top_10_mumps_cases_2000, reorder(Location, Value))
top_10_rubella_cases_2000$Location <-
with(top_10_rubella_cases_2000, reorder(Location, Value))

# MMR highest number of cases in 2000
ggplot() +
  # Measles
  geom_bar(
    data = top_10_measles_cases_2000,
    aes(x = Value, y = Location, fill = Indicator),
    stat = "identity",
    position = position_dodge(width = 0.8),
    width = 0.6
  ) +
  # Mumps
  geom_bar(
    data = top_10_mumps_cases_2000,
    aes(x = Value, y = Location, fill = Indicator),
    stat = "identity",
    position = position_dodge(width = 0.8),
    width = 0.6
  ) +
  # Rubella
```

```r
  geom_bar(
    data = top_10_rubella_cases_2000,
    aes(x = Value, y = Location, fill = Indicator),
    stat = "identity",
    position = position_dodge(width = 0.8),
    width = 0.6
  ) +
  facet_wrap(~ Indicator, scales = "free_y") +  # Separate rows for each disease
  labs(
    title = "Countries with highest cases of MMR diseases (2000)",
    x = "Cases",
    y = "Country",
    fill = "Indicator"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

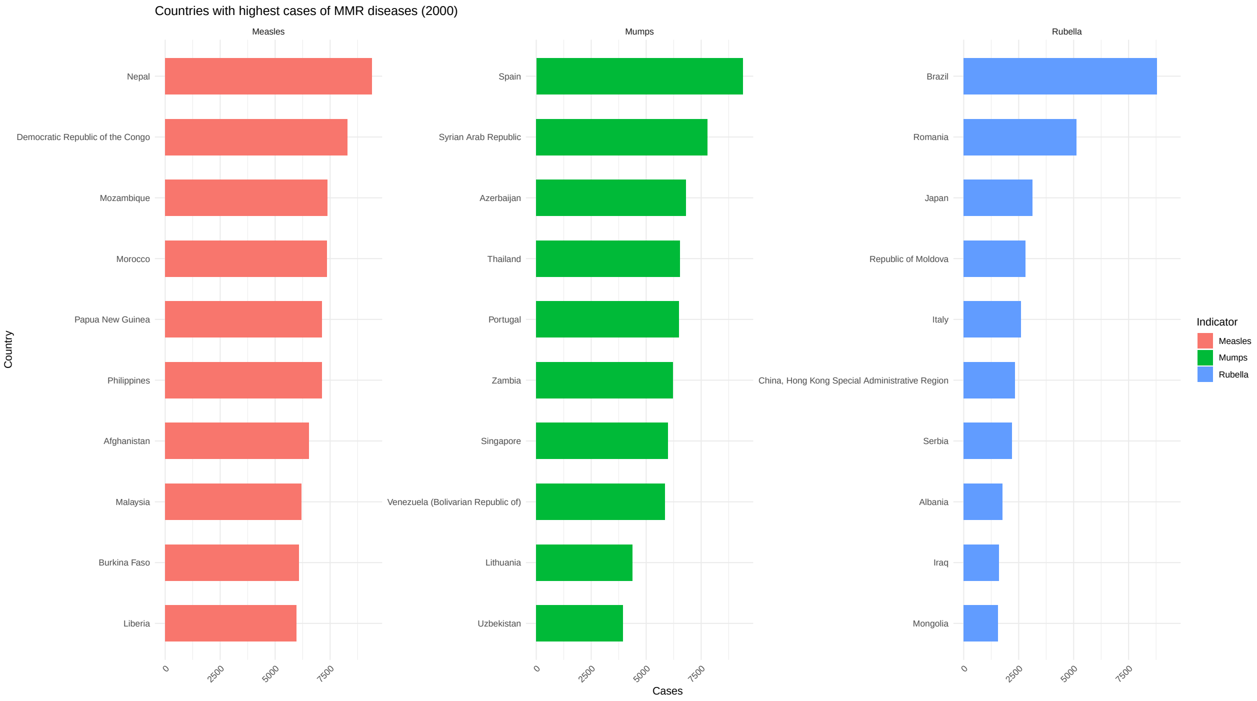


Figure 2: Highest cases per country (2000)

Interestingly, there is no obvious pattern between high disease case numbers and location. In **(figure** 2), a mix of developed and developing countries appear among the highest case locations. This includes Spain, which, as part of the European Union, had the highest number of Mumps cases in 2000. As shown in (**figure** 1), Rubella cases are the lowest of the three diseases, with Brazil having nearly double the number of Rubella cases compared to the second-highest, Romania.

```r
Measles_cases_2023 <- Measles_cases_cleaned %>%
  dplyr::filter(Period == 2023)

Mumps_cases_2023 <- Mumps_cases_cleaned %>%
  dplyr::filter(Period == 2023)

Rubella_cases_2023 <- Rubella_cases_cleaned %>%
  dplyr::filter(Period == 2023)

top_10_measles_cases_2023 <- Measles_cases_2023 %>%
  arrange(desc(Value)) %>%
  slice_head(n = 10)

top_10_mumps_cases_2023 <- Mumps_cases_2023 %>%
  arrange(desc(Value)) %>%
  slice_head(n = 10)

top_10_rubella_cases_2023 <- Rubella_cases_2023 %>%
  arrange(desc(Value)) %>%
  slice_head(n = 10)

# Reorder the locations within each dataset based on the Value
top_10_measles_cases_2023$Location <-
with(top_10_measles_cases_2023, reorder(Location, Value))
top_10_mumps_cases_2023$Location <-
with(top_10_mumps_cases_2023, reorder(Location, Value))
top_10_rubella_cases_2023$Location <-
with(top_10_rubella_cases_2023, reorder(Location, Value))

# MMR highest number of cases in 2023
ggplot() +
  # Measles
  geom_bar(
    data = top_10_measles_cases_2023,
    aes(x = Value, y = Location, fill = Indicator),
    stat = "identity",
    position = position_dodge(width = 0.8),
    width = 0.6
  ) +
  # Mumps
  geom_bar(
    data = top_10_mumps_cases_2023,
    aes(x = Value, y = Location, fill = Indicator),
    stat = "identity",
    position = position_dodge(width = 0.8),
    width = 0.6
  ) +
  # Rubella
  geom_bar(
    data = top_10_rubella_cases_2023,
    aes(x = Value, y = Location, fill = Indicator),
    stat = "identity",
    position = position_dodge(width = 0.8),
```

```
    width = 0.6
  ) +
  facet_wrap(~ Indicator, scales = "free_y") +  # Separate rows for each disease
  labs(
    title = "Countries with highest cases of MMR diseases (2023)",
    x = "Cases",
    y = "Country",
    fill = "Indicator"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

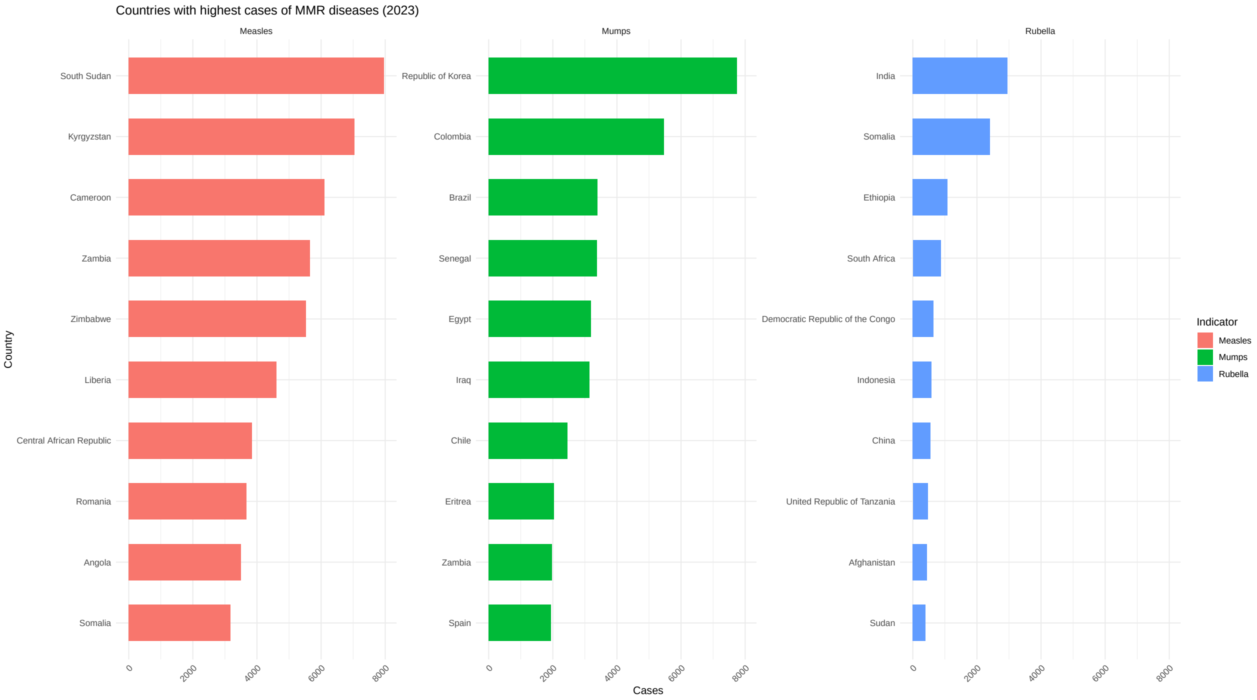


Figure 3: Highest cases per country (2023)

As expected from **figure** 1, the number of cases for each disease has generally decreased, which is also true for the top 10 countries with Mumps and Rubella, as shown in **figure** 3. Interestingly, the top 10 countries for Measles in 2023 have a higher total number of cases than in 2000, which is unsurprising given the numerous measles outbreaks globally.

1.3.1 Measles

Measles is a highly contagious viral disease caused by the Measles morbillivirus, primarily affecting children but also unvaccinated individuals of any age. It spreads through respiratory droplets from coughing or sneezing and can remain infectious in the air or on surfaces for up to two hours. Symptoms include fever, cough, runny nose, conjunctivitis, Koplik spots inside the mouth, and a red blotchy rash that typically spreads from the face downward. While most cases resolve, complications such as pneumonia, diarrhea, encephalitis, and death can occur, particularly in malnourished or immunocompromised individuals. Measles is preventable through the MMR vaccine, which is highly effective when administered in two doses.

Using gradient shading with world map data, the number of global cases in 2000 and 2023 can be visualised in **figure** 4. Although the global measles data for 2000 and 2023 vary in completeness, it is generally evident that global measles cases have decreased, with nearly all countries reporting fewer than 5,000 cases in 2023—an improvement from 2000. This decline is likely due to increased vaccine uptake and better education about the containment and spread of measles. In 2000, cases were widespread across Europe, Asia, and Africa, while in 2023, the majority of cases appear to be concentrated in Africa.

```r
world <- ne_countries(scale = "medium", returnclass = "sf")

Measles_cases_2000_classify <- Measles_cases_2000 %>%
  mutate(Has_Cases = ifelse(Value > 0, "Measles cases", "No cases"))

world_map <- world %>%
  left_join(Measles_cases_2000_classify, by = c("adm0_a3" = "SpatialDimValueCode"))

ggplot(data = world_map) +
  geom_sf(aes(fill = Value)) +
  scale_fill_gradient(
    low = "yellow",
    high = "red",
    na.value = "grey50",
    name = "Measles Cases",
    limits = c(0, 10000)
    ) +
  labs(
    title = "Global Measles cases (2000)",
    fill = "Measles Cases"
  ) +
  theme_minimal()
```

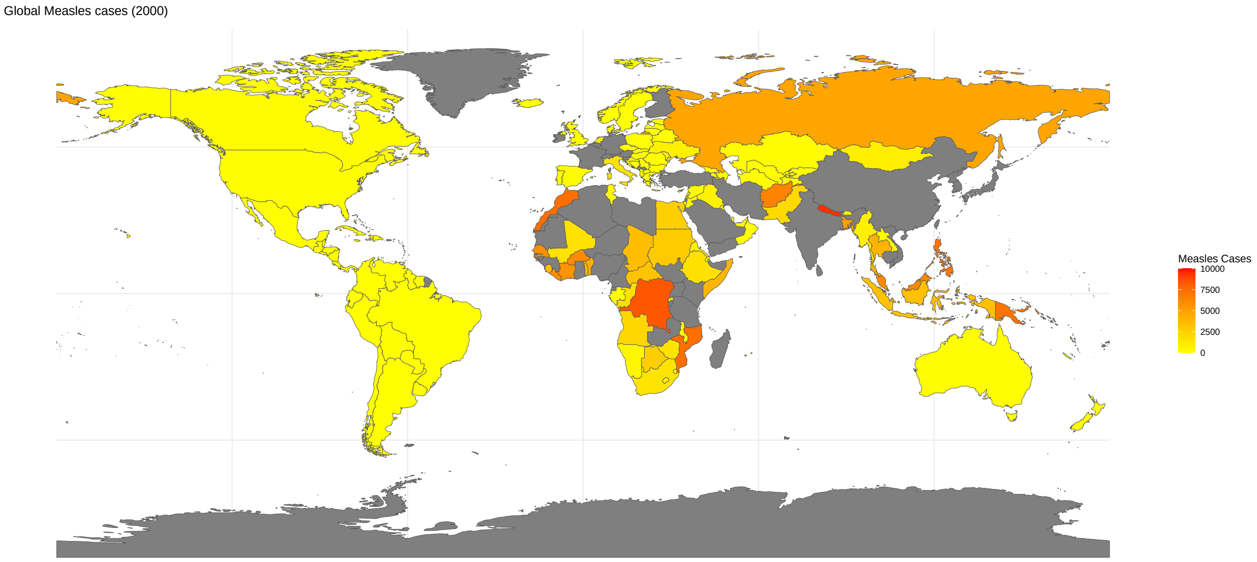
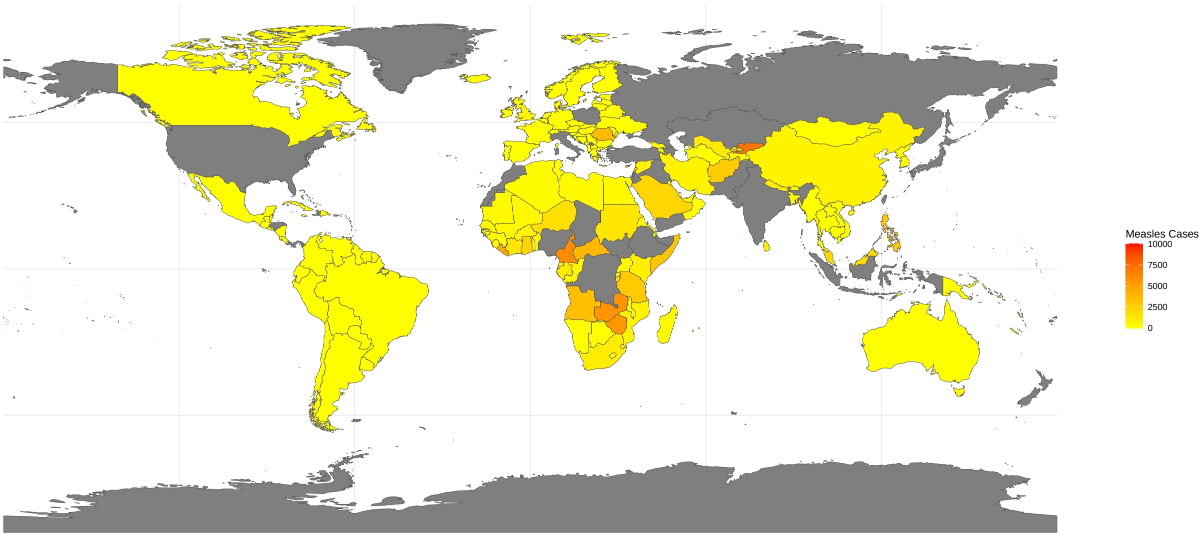


Figure 4: Global measles cases (2000)

```r
world <- ne_countries(scale = "medium", returnclass = "sf")

Measles_cases_2023_classify <- Measles_cases_2023 %>%
  mutate(Has_Cases = ifelse(Value > 0, "Measles cases", "No cases"))

world_map <- world %>%
  left_join(Measles_cases_2023_classify, by = c("adm0_a3" = "SpatialDimValueCode"))

ggplot(data = world_map) +
  geom_sf(aes(fill = Value)) +
  scale_fill_gradient(
    low = "yellow",
    high = "red",
    na.value = "grey50",
    name = "Measles Cases",
    limits = c(0, 10000)
    ) +
  labs(
    title = "Global Measles cases (2023)",
    fill = "Measles Cases"
  ) +
  theme_minimal()
```


Global Measles cases (2023)

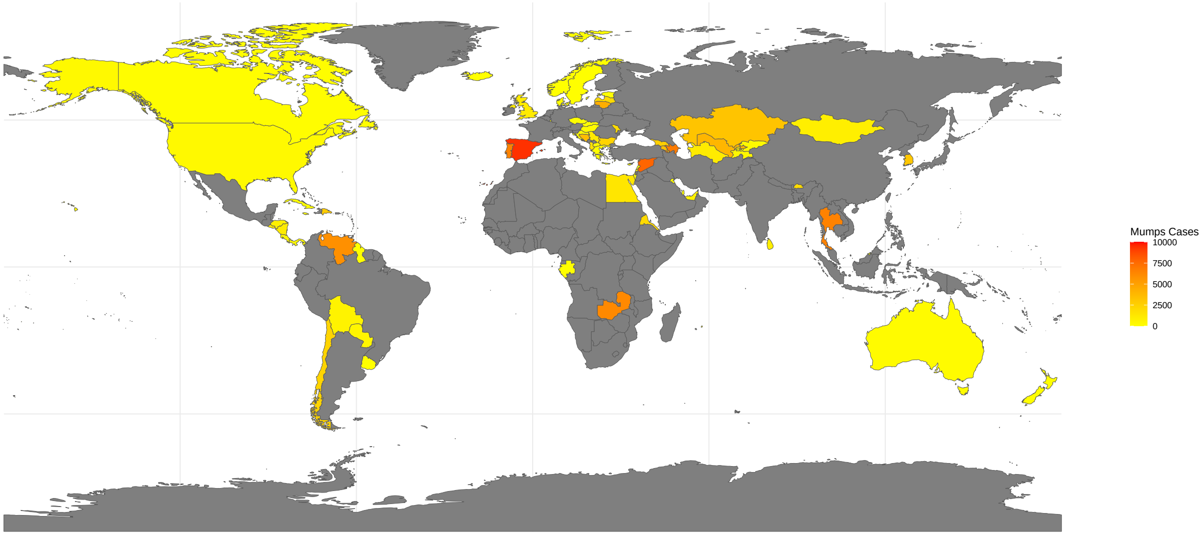
1.3.2 Mumps

Mumps is a contagious viral illness caused by the Mumps rubulavirus, commonly affecting children but also unvaccinated individuals of any age. It spreads through respiratory droplets, direct contact with saliva, or contaminated surfaces. The hallmark symptom is painful swelling of the salivary glands, particularly the parotid glands, often accompanied by fever, headache, muscle aches, fatigue, and loss of appetite. While mumps is usually self-limiting, complications can include orchitis (inflammation of the testicles), oophoritis (inflammation of the ovaries), meningitis, encephalitis, and, rarely, hearing loss. Mumps is preventable with the MMR vaccine, which is highly effective when administered in two doses.

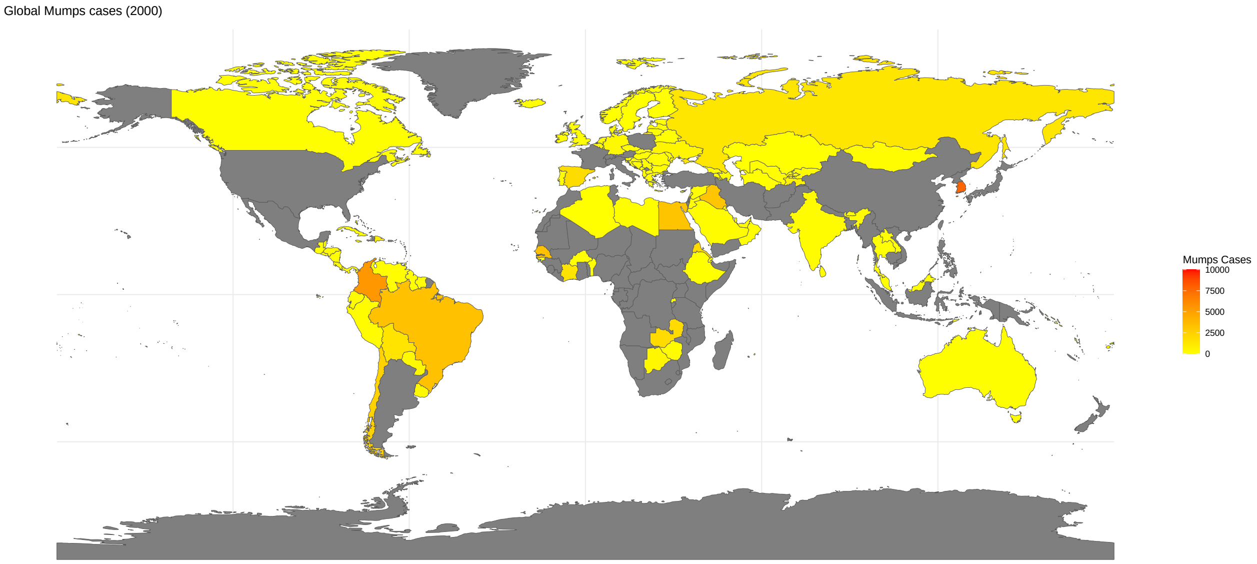
The global WHO data for mumps is the most limited compared to measles and rubella. However, a trend between the years 2000 and 2023 can still be identified, as shown in **figure** 4. In 2000, although data is sparse, it suggests that mumps had high case numbers across various regions, with notable cases in Spain, South America, Asia, and Africa. In 2023, the global data is more complete, showing a general decrease in mumps cases, with the highest case numbers in South America, North Africa, and Asia. This decline is likely due to increased vaccine uptake and greater awareness of the disease, along with the fact that mumps seems to be less transmissible than measles.

```r
world <- ne_countries(scale = "medium", returnclass = "sf")

Mumps_cases_2000_classify <- Mumps_cases_2000 %>%
  mutate(Has_Cases = ifelse(Value > 0, "Measles cases", "No cases"))

world_map <- world %>%
  left_join(Mumps_cases_2000_classify, by = c("adm0_a3" = "SpatialDimValueCode"))

ggplot(data = world_map) +
  geom_sf(aes(fill = Value)) +
  scale_fill_gradient(
    low = "yellow",
    high = "red",
    na.value = "grey50",
    name = "Mumps Cases",
    limits = c(0, 10000)
    ) +
  labs(
    title = "Global Mumps cases (2000)",
    fill = "Mumps Cases"
  ) +
  theme_minimal()
```

Global Mumps cases (2000)



```r
world <- ne_countries(scale = "medium", returnclass = "sf")

Mumps_cases_2023_classify <- Mumps_cases_2023 %>%
  mutate(Has_Cases = ifelse(Value > 0, "Measles cases", "No cases"))

world_map <- world %>%
  left_join(Mumps_cases_2023_classify, by = c("adm0_a3" = "SpatialDimValueCode"))

ggplot(data = world_map) +
  geom_sf(aes(fill = Value)) +
  scale_fill_gradient(
    low = "yellow",
    high = "red",
    na.value = "grey50",
    name = "Mumps Cases",
    limits = c(0, 10000)
    ) +
  labs(
    title = "Global Mumps cases (2000)",
    fill = "Mumps Cases"
  ) +
  theme_minimal()
```

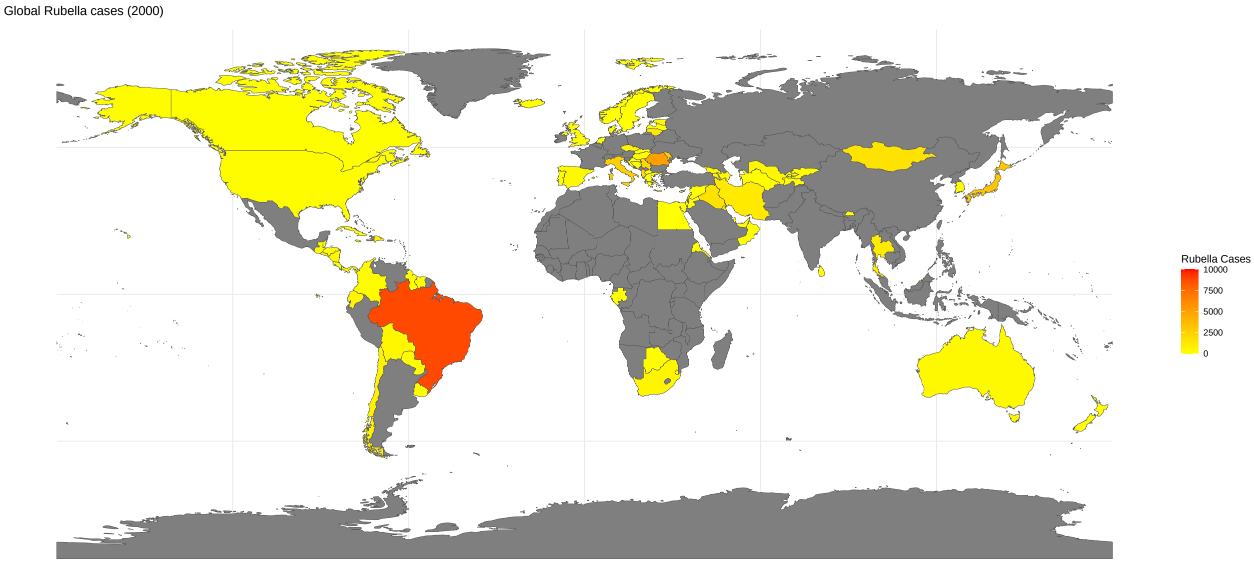


1.3.3 Rubella

Rubella, also known as German measles, is a contagious viral disease caused by the Rubella virus. It primarily affects children and young adults but can occur in unvaccinated individuals of any age. The disease spreads through respiratory droplets from coughing or sneezing and can also be transmitted from a pregnant mother to her fetus. Symptoms are typically mild and include a low-grade fever, sore throat, and a pink or light red rash that begins on the face and spreads to the rest of the body. Swollen lymph nodes, joint pain, and mild conjunctivitis may also occur. While generally not severe, rubella poses a significant risk to pregnant women, as it can cause miscarriage, stillbirth, or congenital rubella syndrome (CRS), leading to serious birth defects. Rubella is preventable with the MMR vaccine, which is highly effective when given in two doses.

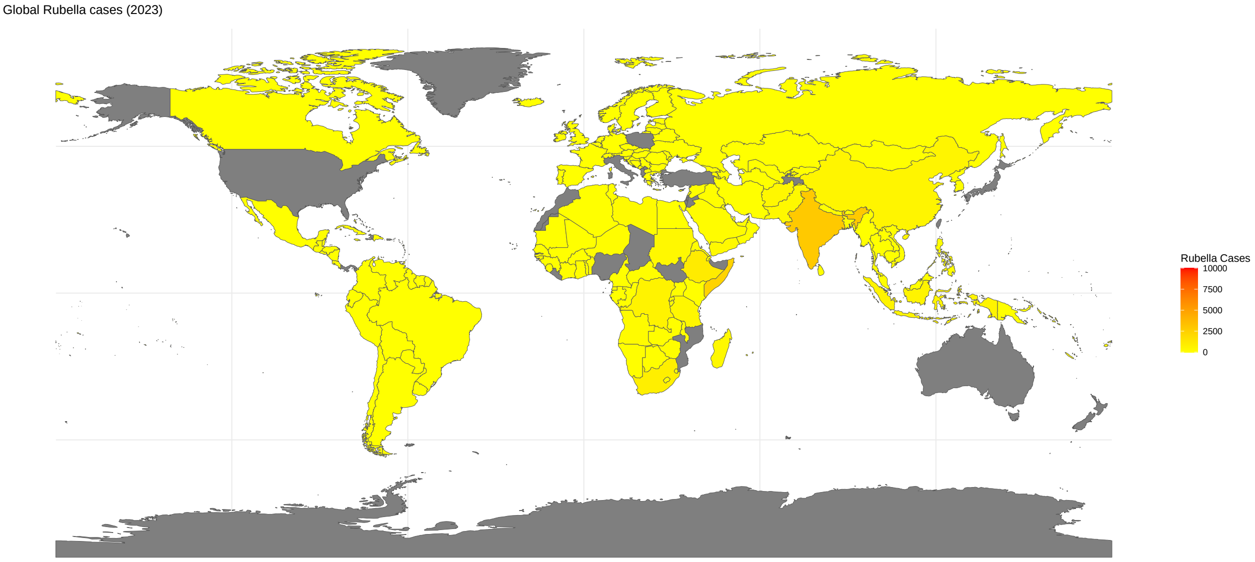
Fortunately, the global data for rubella is more complete compared to the mumps data. As shown earlier in the report in **figure** 4, rubella has the lowest number of global cases compared to measles and mumps. This trend is generally reflected in the global cases for 2000 and 2023; however, Brazil had a notably high number of rubella cases in 2000, with some parts of Europe also affected. In 2023, the data is more complete and shows a low global number of rubella cases, with the exception of India and certain areas of Africa, which still report higher case numbers.

```r
world <- ne_countries(scale = "medium", returnclass = "sf")

Rubella_cases_2000_classify <- Rubella_cases_2000 %>%
  mutate(Has_Cases = ifelse(Value > 0, "Measles cases", "No cases"))

world_map <- world %>%
  left_join(Rubella_cases_2000_classify, by = c("adm0_a3" = "SpatialDimValueCode"))

ggplot(data = world_map) +
  geom_sf(aes(fill = Value)) +
  scale_fill_gradient(
    low = "yellow",
    high = "red",
    na.value = "grey50",
    name = "Rubella Cases",
    limits = c(0, 10000)
    ) +
  labs(
    title = "Global Rubella cases (2000)",
    fill = "Rubella Cases"
  ) +
  theme_minimal()
```



Global Rubella cases (2000)

```
world <- ne_countries(scale = "medium", returnclass = "sf")

Rubella_cases_2023_classify <- Rubella_cases_2023 %>%
  mutate(Has_Cases = ifelse(Value > 0, "Measles cases", "No cases"))

world_map <- world %>%
  left_join(Rubella_cases_2023_classify, by = c("adm0_a3" = "SpatialDimValueCode"))

ggplot(data = world_map) +
  geom_sf(aes(fill = Value)) +
  scale_fill_gradient(
    low = "yellow",
    high = "red",
    na.value = "grey50",
    name = "Rubella Cases",
    limits = c(0, 10000)
    ) +
  labs(
    title = "Global Rubella cases (2023)",
    fill = "Rubella Cases"
  ) +
  theme_minimal()
```



1.3.4 Further analysis

To test the significance of the number of cases between specific years = ANOVA

Case relationship with time, magnitude, prediction, goodness of fit and significance = Linear regression

Predict future cases based on previous data = Forecasting

1.4 Disease vaccines

MCV1 and MCV2 refer to the first and second doses of the measles-containing vaccine, which is typically part of the MMR (measles, mumps, rubella) vaccine. MCV1 is administered to children around 12–15 months of age, providing initial protection against measles. MCV2 is given between the ages of 4 and 6 years as a booster dose, ensuring long-term immunity and nearly 100% protection. The rubella vaccine, also part of the MMR vaccine, is similarly administered in two doses, with the first dose typically given at 12–15 months and the second at 4–6 years. Together, MCV1/MCV2 and the rubella vaccine are highly effective in preventing measles and rubella, both of which can cause serious complications, particularly in pregnant women, and are preventable through immunisation.

The MMR vaccine is a combined vaccine that provides protection against three highly contagious viral diseases: measles, mumps, and rubella. It is typically administered in two doses as part of routine childhood immunisation schedules. The first dose, usually given at 12–15 months of age, provides initial protection against these diseases, while the second dose, given between 4 and 6 years of age, acts as a booster to ensure long-term immunity. The MMR vaccine is highly effective, with a success rate of about 97% for measles and rubella, and 88% for mumps after two doses. By preventing these diseases, the MMR vaccine helps reduce the risk of serious complications, such as encephalitis, pneumonia, infertility, and congenital rubella syndrome, making it a crucial tool in public health.

The source of the global disease and vaccination data for this report comes from the WHO, but only data for the measles and rubella vaccines is available, not for the mumps vaccine. All three vaccines are part of the MMR triple vaccine, but this report only includes data for the measles and rubella vaccines.

```r
#Summarise number of MCV1 vaccinations per year
MCV1_summary <- MCV1_coverage_cleaned %>%
  group_by(YEAR) %>%
  summarise(total_vaccinations = mean(COVERAGE))
#view(MCV1_summary)

#Summarise number of MCV2 vaccinations per year
MCV2_summary <- MCV2_coverage_cleaned %>%
  group_by(YEAR) %>%
  summarise(total_vaccinations = mean(COVERAGE))
#view(MCV2_summary)

#Summarise number of Rubella vaccinations per year
Rubella_vaccination_summary <- Rubella_vaccine_coverage_cleaned %>%
  group_by(YEAR) %>%
  summarise(total_vaccinations = mean(COVERAGE))
#view(Rubella_vaccination_summary)

#Vaccination over time
ggplot() +
  geom_line(data = MCV1_summary, aes(x=YEAR, y=total_vaccinations,
  color = "MCV1"), size = 1.5) +
  geom_point(data = MCV1_summary, aes(x=YEAR, y=total_vaccinations,
  color = "MCV1"), size = 3) +
  geom_line(data = MCV2_summary, aes(x=YEAR, y=total_vaccinations,
  color = "MCV2"), size = 1.5) +
  geom_point(data = MCV2_summary, aes(x=YEAR, y=total_vaccinations,
  color = "MCV2"), size = 3) +
  geom_line(data = Rubella_vaccination_summary,
  aes(x=YEAR, y=total_vaccinations, color = "Rubella Vaccine"), size = 1.5) +
  geom_point(data = Rubella_vaccination_summary,
```

```
aes(x=YEAR, y=total_vaccinations, color = "Rubella Vaccine"), size = 3) +

labs(title = "Measles and Rubella Vaccination over time",
     x= "Year",
     y= "Percentage of vaccination",
     color = "Vaccine") +
scale_y_continuous(limits = c(75, 95)) +
scale_x_continuous(limits = c(2000, 2023)) +
scale_color_manual(values = c("MCV1" = "red", "MCV2" = "orange",
"Rubella Vaccine" = "blue" )) +
scale_x_continuous(breaks = seq(min(Measles_cases_summary$Period),
max(Measles_cases_summary$Period), by = 4))
```

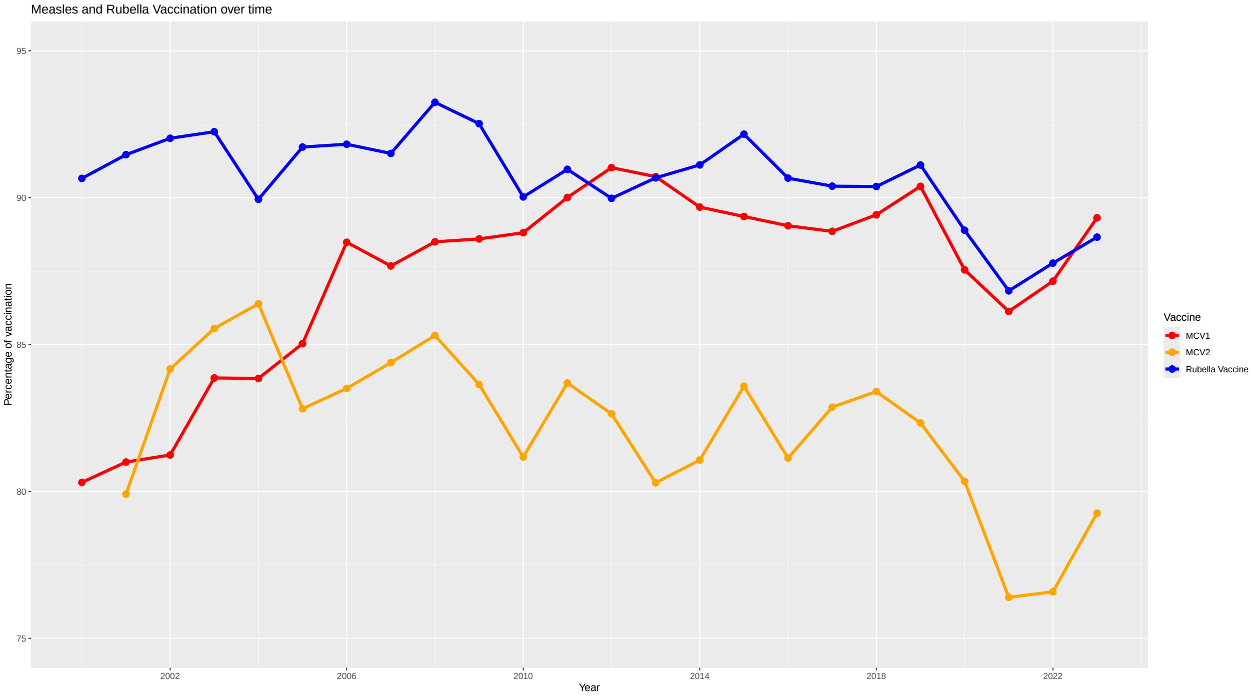


Figure 5: Vaccine coverage 2000-2023

```
theme_stata()
```

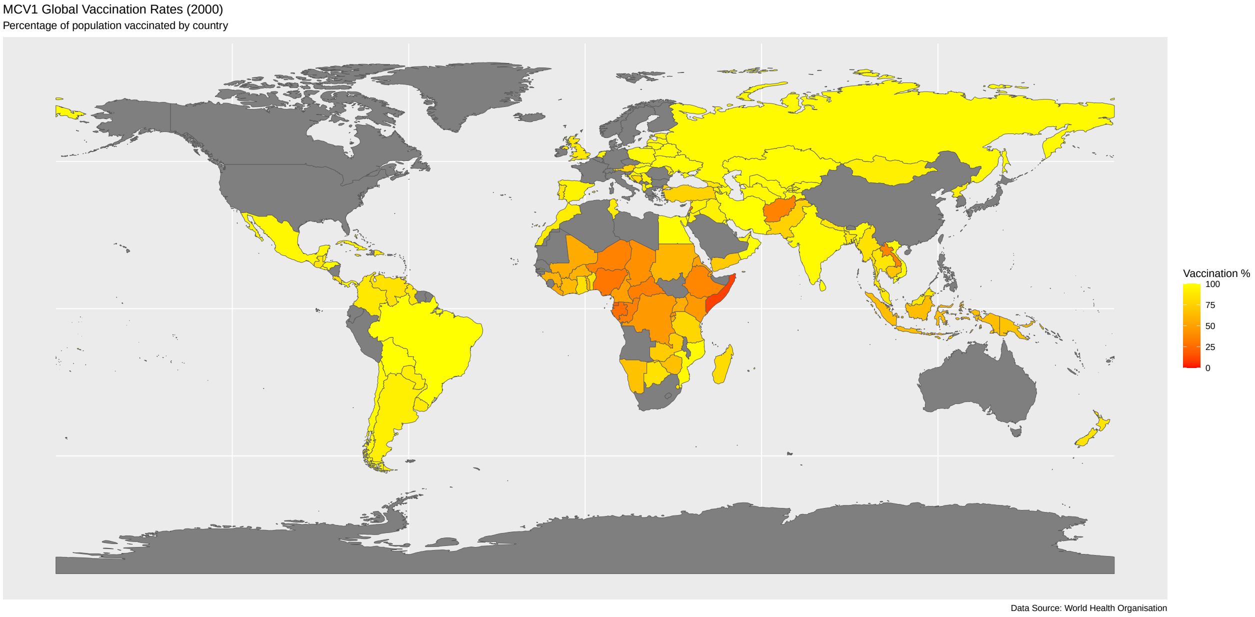
By analysing the average percentage coverage of MCV1/2 and the rubella vaccine, some clear trends emerge. The rubella vaccine has maintained a steady coverage above 90% from 2000 to 2020, whereas both the MCV1 and MCV2 vaccines have consistently been below this percentage. This could help explain why there are more global measles cases compared to rubella, as without 85–90% coverage in the population, herd immunity will be ineffective in preventing new infections in unvaccinated individuals. Another trend observed in this visual (**figure** 5) is a sharp decrease in global vaccination coverage for all three vaccines between 2019 and 2021. This could be attributed to a mixture of vaccine hesitancy, stemming from fears of side effects or media claims. Additionally, the COVID-19 pandemic in 2020 likely limited contact with healthcare services and vaccinations for non-COVID diseases. However, after 2021, vaccination coverage began to increase again, likely due to efforts aimed at reducing outbreaks of measles and rubella.

1.4.1 Vaccination distribution

Vaccine uptake varies by country due to a range of socioeconomic factors. By examining the locations with high and low vaccination coverage, it may be possible to predict future outbreaks. These factors change over time, which will be explored by comparing data from the years 2000 and 2023.

In the year 2000, although several countries have missing data, global MCV1 vaccination coverage is generally above 75%, with lower coverage in Africa and Southern Asia (around 50% and below). The data completeness is poorer for MCV1 in 2000, but based on the available data, South America, Russia, and Europe all show relatively low coverage. The missing data may be due to a reduced deployment of the second MCV2 vaccine, which could also explain the lower vaccine uptake.

```r
world <- ne_countries(scale = "medium", returnclass = "sf")

world_map <- world %>%
  left_join(MCV1_coverage_2000, by = c("adm0_a3" = "CODE"))

ggplot(data = world_map) +
  geom_sf(aes(fill = MCV1COVERAGE2000)) +
  scale_fill_gradient(low = "red",
                      high = "yellow",
                      na.value = "grey50",
                      name = "Vaccination %",
                      limits = c(0,100)) +
  labs(
    title = "MCV1 Global Vaccination Rates (2000)",
    subtitle = "Percentage of population vaccinated by country",
    caption = "Data Source: World Health Organisation"
  )
```

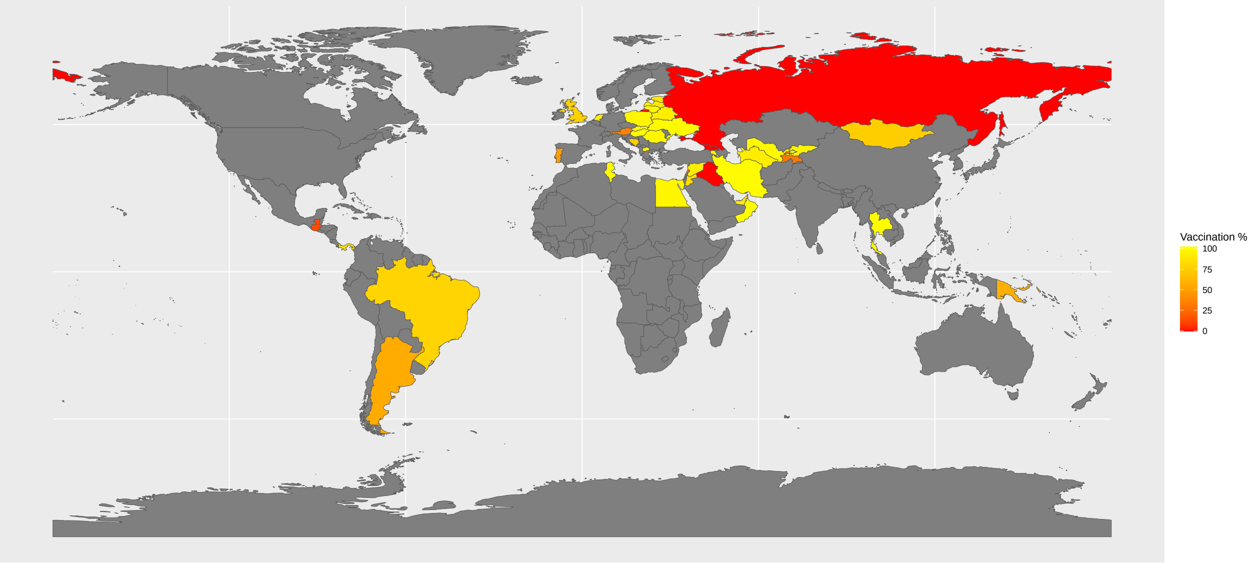


```r
  theme_minimal()
```

```r
world <- ne_countries(scale = "medium", returnclass = "sf")

world_map <- world %>%
  left_join(MCV2_coverage_2000, by = c("adm0_a3" = "CODE"))

ggplot(data = world_map) +
  geom_sf(aes(fill = MCV2COVERAGE2000)) +
  scale_fill_gradient(low = "red", high = "yellow", na.value = "grey50",
  name = "Vaccination %") +
  labs(
    title = "MCV2 Global Vaccination Rates (2000)",
    subtitle = "Percentage of population vaccinated by country",
    caption = "Data Source: World Health Organisation"
  )
```

MCV2 Global Vaccination Rates (2000)

Percentage of population vaccinated by country

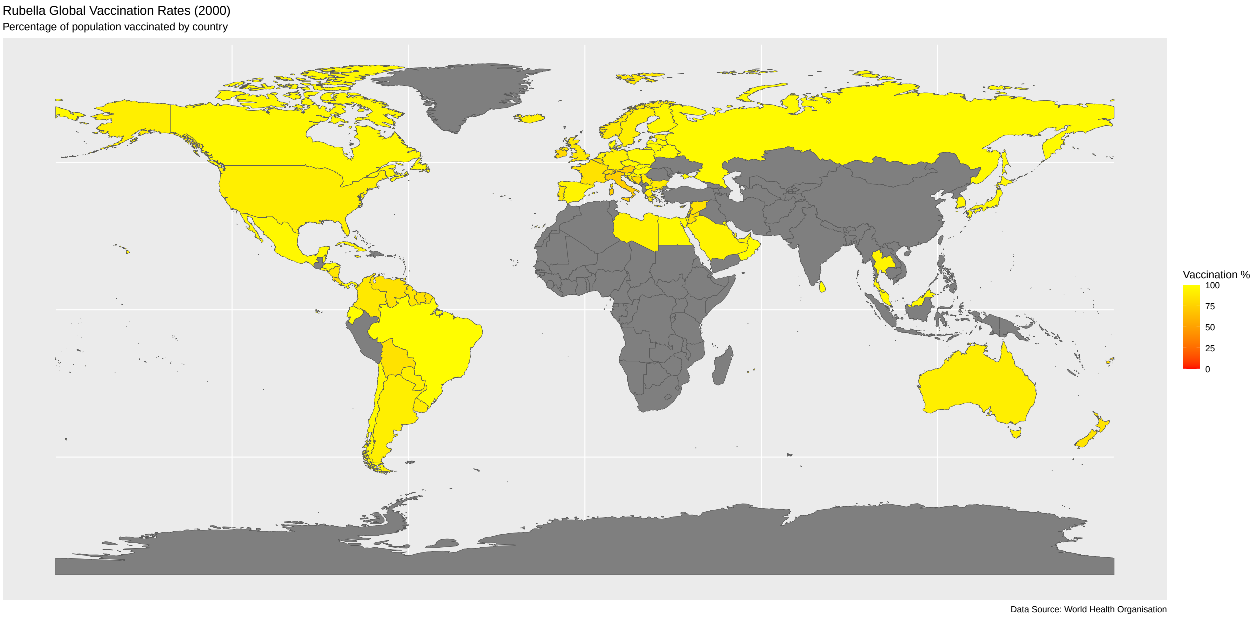


Data Source: World Health Organisation

```r
  theme_minimal()
```

The situation in 2000 for the rubella vaccine is notably better in terms of data completeness and vaccine coverage. The Americas, Europe, North Africa, and Australia have high vaccine coverage, above 75%, while Asia and the rest of Africa lack data. This coverage appears to correlate with lower global cases of rubella compared to measles and mumps, which had lower vaccine coverage in 2000.

```r
world <- ne_countries(scale = "medium", returnclass = "sf")

world_map <- world %>%
  left_join(Rubella_vaccination_2000, by = c("adm0_a3" = "CODE"))

ggplot(data = world_map) +
  geom_sf(aes(fill = RUBELLACOVERAGE2000)) +
  scale_fill_gradient(low = "red", high = "yellow", na.value = "grey50",
  name = "Vaccination %",limits = c(0,100)) +
  labs(
    title = "Rubella Global Vaccination Rates (2000)",
    subtitle = "Percentage of population vaccinated by country",
    caption = "Data Source: World Health Organisation"
  )
```



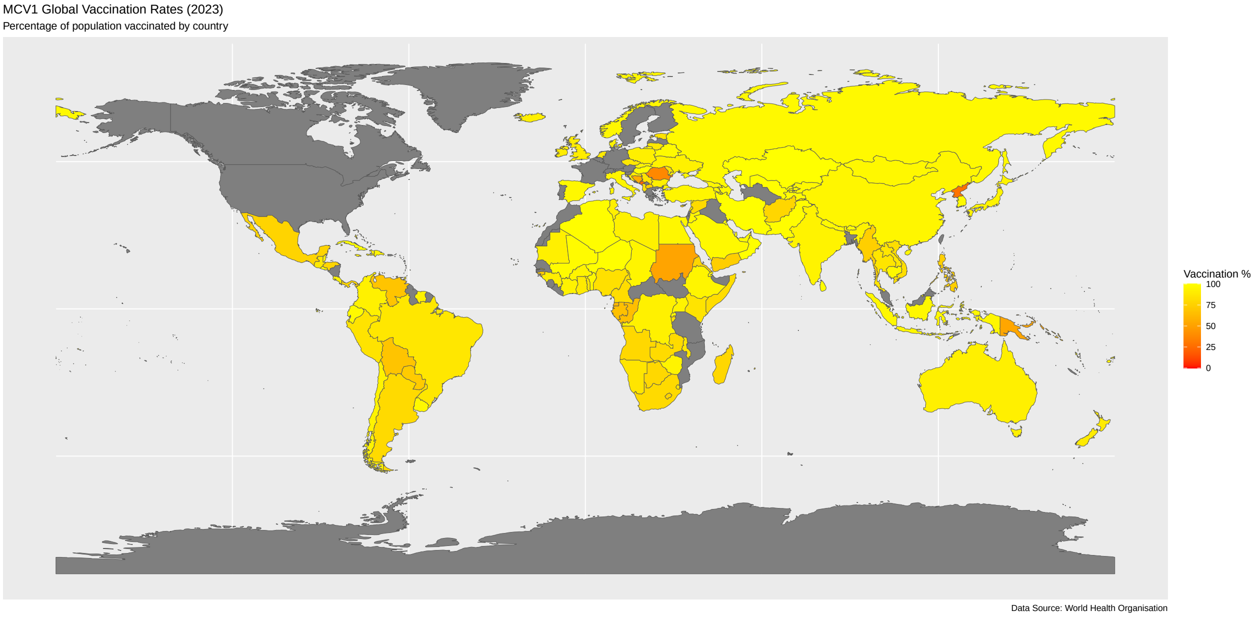
```r
  theme_minimal()
```

The transition from 2000 to 2023 shows positive vaccination uptake in some countries but also troubling decreases in vaccination coverage in others.

Starting with MCV1 in 2023, global vaccination coverage has increased, along with improved data completeness. The most significant improvement is in Africa, where most countries now have vaccination coverage of 75% or higher. In contrast, South America appears to have lower vaccine coverage compared to 2000, with many countries now at around 75% or below.

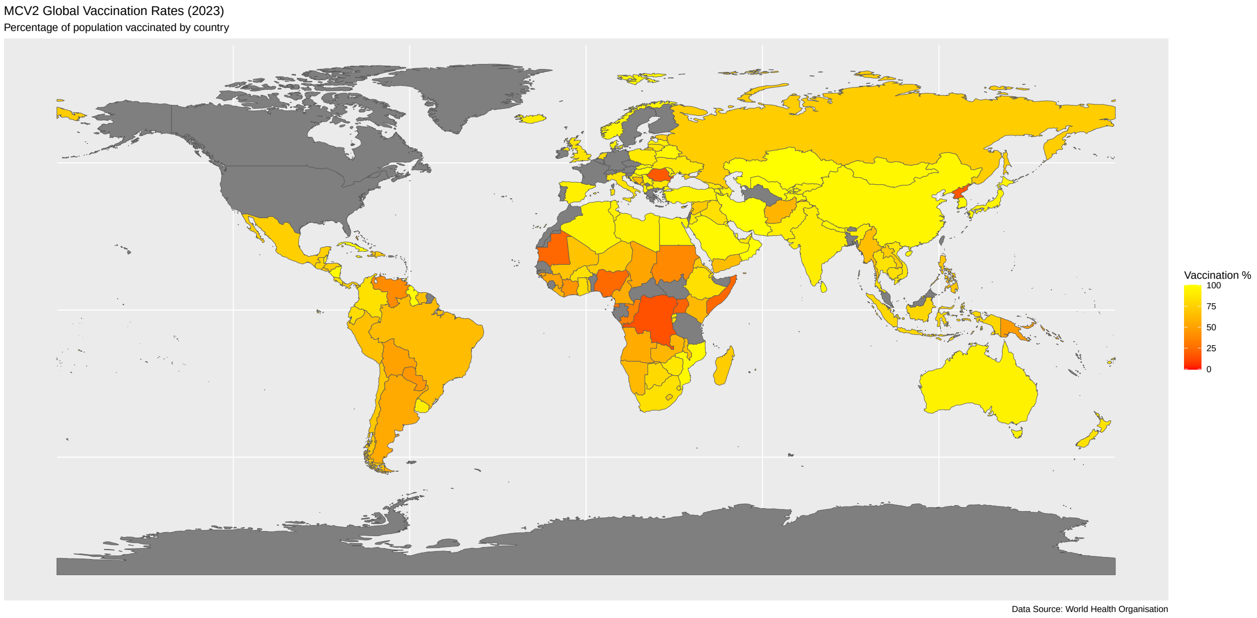
MCV2 coverage in 2023 shows significantly higher data completeness. However, vaccination coverage remains below 75% in Africa, South America, and South Asia, while Europe and mainland Asia have coverage above 75%. It appears that MCV2 uptake is still lower than required for effective herd immunity.

```r
world <- ne_countries(scale = "medium", returnclass = "sf")

world_map <- world %>%
  left_join(MCV1_coverage_2023, by = c("adm0_a3" = "CODE"))

ggplot(data = world_map) +
  geom_sf(aes(fill = MCV1COVERAGE2023)) +
  scale_fill_gradient(low = "red", high = "yellow", na.value = "grey50",
  name = "Vaccination %", limits = c(0,100)) +
  labs(
    title = "MCV1 Global Vaccination Rates (2023)",
    subtitle = "Percentage of population vaccinated by country",
    caption = "Data Source: World Health Organisation"
  )
```



```r
  theme_minimal()
```

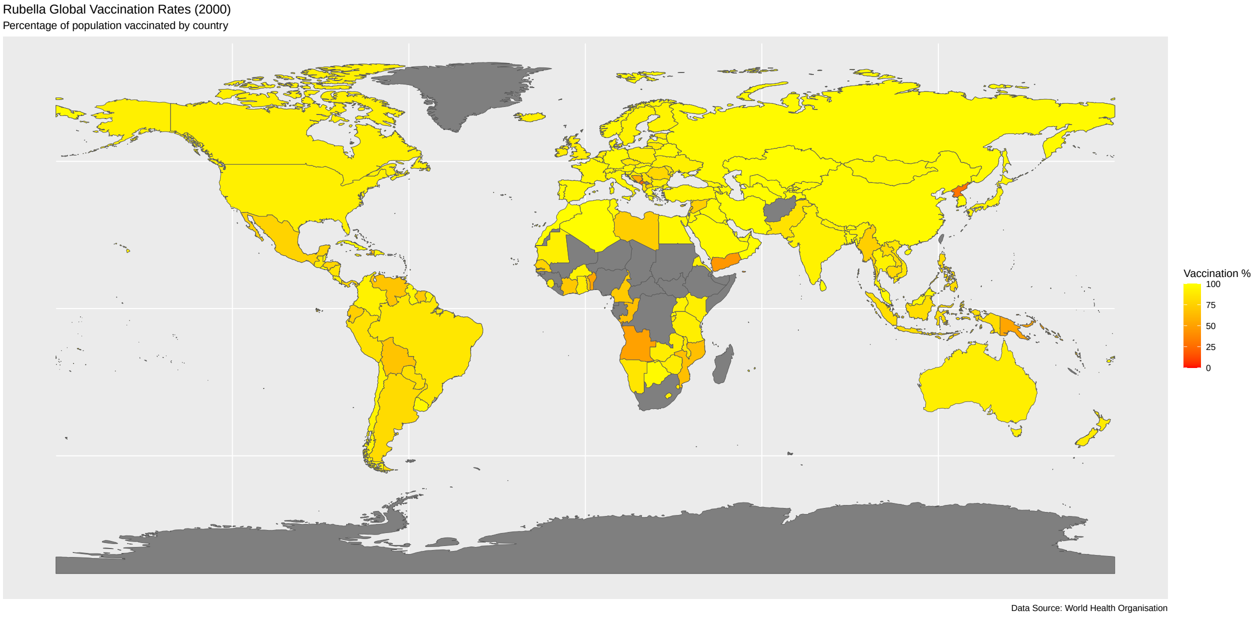
```
world <- ne_countries(scale = "medium", returnclass = "sf")

world_map <- world %>%
  left_join(MCV2_coverage_2023, by = c("adm0_a3" = "CODE"))

ggplot(data = world_map) +
  geom_sf(aes(fill = MCV2COVERAGE2023)) +
  scale_fill_gradient(low = "red", high = "yellow", na.value = "grey50",
  name = "Vaccination %", limits = c(0,100)) +
  labs(
    title = "MCV2 Global Vaccination Rates (2023)",
    subtitle = "Percentage of population vaccinated by country",
    caption = "Data Source: World Health Organisation"
  )
```



```
  theme_minimal()
```

Similar to MCV1/2 in 2023, there is greater data completeness compared to 2000. While there is high vaccine coverage in Europe, Asia, and North America, there is a concerning decrease in vaccination coverage in South America, which has fallen to around 75%, a level that was clearly higher in 2000. This decline could be linked to the 2020 COVID pandemic, which caused a decrease in the availability of healthcare services.

```r
world <- ne_countries(scale = "medium", returnclass = "sf")

world_map <- world %>%
  left_join(Rubella_vaccination_2023, by = c("adm0_a3" = "CODE"))

ggplot(data = world_map) +
  geom_sf(aes(fill = RUBELLACOVERAGE2023)) +
  scale_fill_gradient(low = "red", high = "yellow", na.value = "grey50",
  name = "Vaccination %", limits = c(0,100)) +
  labs(
    title = "Rubella Global Vaccination Rates (2000)",
    subtitle = "Percentage of population vaccinated by country",
    caption = "Data Source: World Health Organisation"
  )
```



```r
  theme_minimal()
```

1.4.2 Vaccine coverage by country

As mentioned in the previous section, while there were increases in vaccination coverage for all three vaccines, there were also minor decreases in some cases. By investigating countries with low vaccination coverage, patterns can be identified.

In 2000, countries with low vaccination coverage for MCV1 tended to be mostly African or Central Asian countries, which are generally less developed and may face challenges such as limited healthcare availability or unstable governments. The same pattern is observed with MCV2, although exceptions include Georgia, Russia, and Malta, which had nearly or almost 0% vaccination coverage for MCV2. This could indicate a lack of vaccine deployment in these countries in 2000. The situation for the rubella vaccine is different, as most countries had high vaccination coverage, with the exception of Iceland, which unusually had coverage in the single digits. A possible explanation for this is that, due to its small population, Iceland may have experienced few rubella outbreaks, leading to a lower demand for vaccination.

```r
bottom_10_MCV1_coverage_2000 <- MCV1_coverage_2000 %>%
  arrange(MCV1COVERAGE2000) %>%
  slice_head(n = 10) %>%
  mutate(NAME = factor(NAME, levels = unique(NAME)))

bottom_10_MCV2_coverage_2000 <- MCV2_coverage_2000 %>%
  arrange(MCV2COVERAGE2000) %>%
  slice_head(n = 10) %>%
  mutate(NAME = factor(NAME, levels = unique(NAME)))

bottom_10_rubella_vaccination_2000 <- Rubella_vaccination_2000 %>%
  arrange(RUBELLACOVERAGE2000) %>%
  slice_head(n = 11) %>%
  mutate(NAME = factor(NAME, levels = unique(NAME)))

# Reorder the locations within each dataset based on the Value
bottom_10_MCV1_coverage_2000$NAME <-
with(bottom_10_MCV1_coverage_2000, reorder(NAME, MCV1COVERAGE2000))
bottom_10_MCV2_coverage_2000$NAME <-
with(bottom_10_MCV2_coverage_2000, reorder(NAME, MCV2COVERAGE2000))
bottom_10_rubella_vaccination_2000$NAME <-
with(bottom_10_rubella_vaccination_2000, reorder(NAME, RUBELLACOVERAGE2000))

# Vaccination coverage in 2000
ggplot() +
  # MSV1
  geom_bar(
    data = bottom_10_MCV1_coverage_2000,
    aes(x = MCV1COVERAGE2000, y = NAME, fill = ANTIGEN),
    stat = "identity",
    position = position_dodge(width = 0.8),
    width = 0.6
  ) +
  # MSV2
  geom_bar(
    data = bottom_10_MCV2_coverage_2000,
    aes(x = MCV2COVERAGE2000, y = NAME, fill = ANTIGEN),
    stat = "identity",
    position = position_dodge(width = 0.8),
```

```
  width = 0.6
) +
# Rubella Vaccine
geom_bar(
  data = bottom_10_rubella_vaccination_2000,
  aes(x = RUBELLACOVERAGE2000, y = NAME, fill = ANTIGEN),
  stat = "identity",
  position = position_dodge(width = 0.8),
  width = 0.6
) +
facet_wrap(~ ANTIGEN, scales = "free_y") +  # Separate rows for each disease
labs(
  title = "Countries with lowest vaccination (2000)",
  x = "Coverage (%)",
  y = "Country",
  fill = "ANTIGEN"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

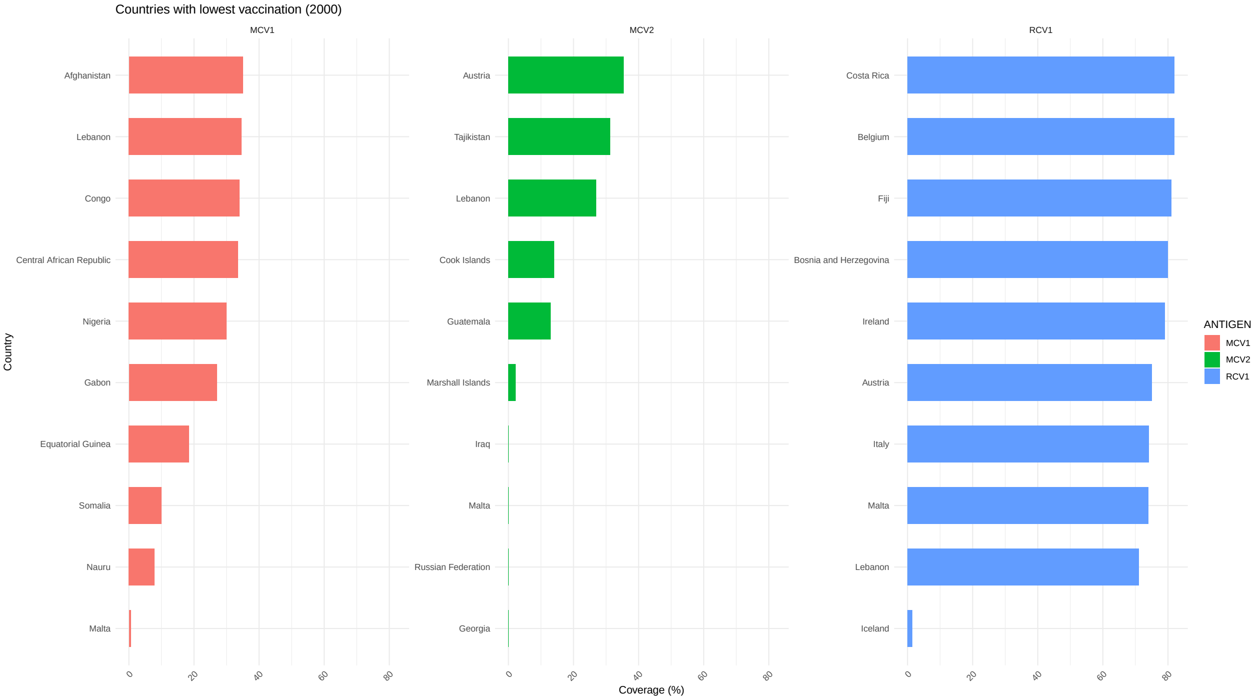
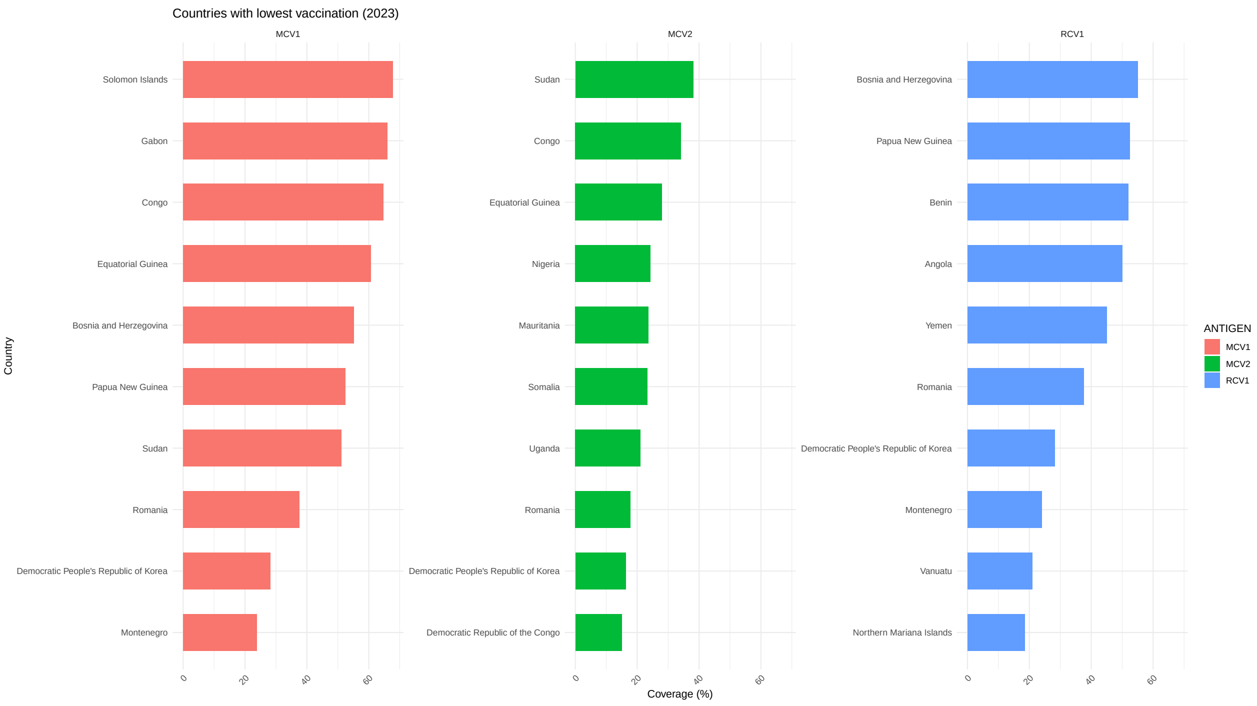


Figure 6: Lowest vaccination coverage (2000)

The trends in vaccination coverage from 2000 largely continue into 2023, with all three vaccines showing lower coverage in African countries, Eastern Europe, and other generally less developed regions. However, higher overall coverage is observed in 2023, as seen in **figure** 6, compared to 2000.

```r
bottom_10_MCV1_coverage_2023 <- MCV1_coverage_2023 %>%
  arrange(MCV1COVERAGE2023) %>%
  slice_head(n = 10) %>%
  mutate(NAME = factor(NAME, levels = unique(NAME)))

bottom_10_MCV2_coverage_2023 <- MCV2_coverage_2023 %>%
  arrange(MCV2COVERAGE2023) %>%
  slice_head(n = 10) %>%
  mutate(NAME = factor(NAME, levels = unique(NAME)))

bottom_10_rubella_vaccination_2023 <- Rubella_vaccination_2023 %>%
  arrange(RUBELLACOVERAGE2023) %>%
  slice_head(n = 17) %>%
  mutate(NAME = factor(NAME, levels = unique(NAME)))

# Reorder the locations within each dataset based on the Value
bottom_10_MCV1_coverage_2023$NAME <-
with(bottom_10_MCV1_coverage_2023, reorder(NAME, MCV1COVERAGE2023))
bottom_10_MCV2_coverage_2023$NAME <-
with(bottom_10_MCV2_coverage_2023, reorder(NAME, MCV2COVERAGE2023))
bottom_10_rubella_vaccination_2023$NAME <-
with(bottom_10_rubella_vaccination_2023, reorder(NAME, RUBELLACOVERAGE2023))

# Vaccination coverage in 2000
ggplot() +
  # MSV1
  geom_bar(
    data = bottom_10_MCV1_coverage_2023,
    aes(x = MCV1COVERAGE2023, y = NAME, fill = ANTIGEN),
    stat = "identity",
    position = position_dodge(width = 0.8),
    width = 0.6
  ) +
  # MSV2
  geom_bar(
    data = bottom_10_MCV2_coverage_2023,
    aes(x = MCV2COVERAGE2023, y = NAME, fill = ANTIGEN),
    stat = "identity",
    position = position_dodge(width = 0.8),
    width = 0.6
  ) +
  # Rubella Vaccine
  geom_bar(
    data = bottom_10_rubella_vaccination_2023,
    aes(x = RUBELLACOVERAGE2023, y = NAME, fill = ANTIGEN),
    stat = "identity",
    position = position_dodge(width = 0.8),
    width = 0.6
  ) +
  facet_wrap(~ ANTIGEN, scales = "free_y") +  # Separate rows for each disease
```

```
labs(
  title = "Countries with lowest vaccination (2023)",
  x = "Coverage (%)",
  y = "Country",
  fill = "ANTIGEN"
) +
theme_minimal() +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

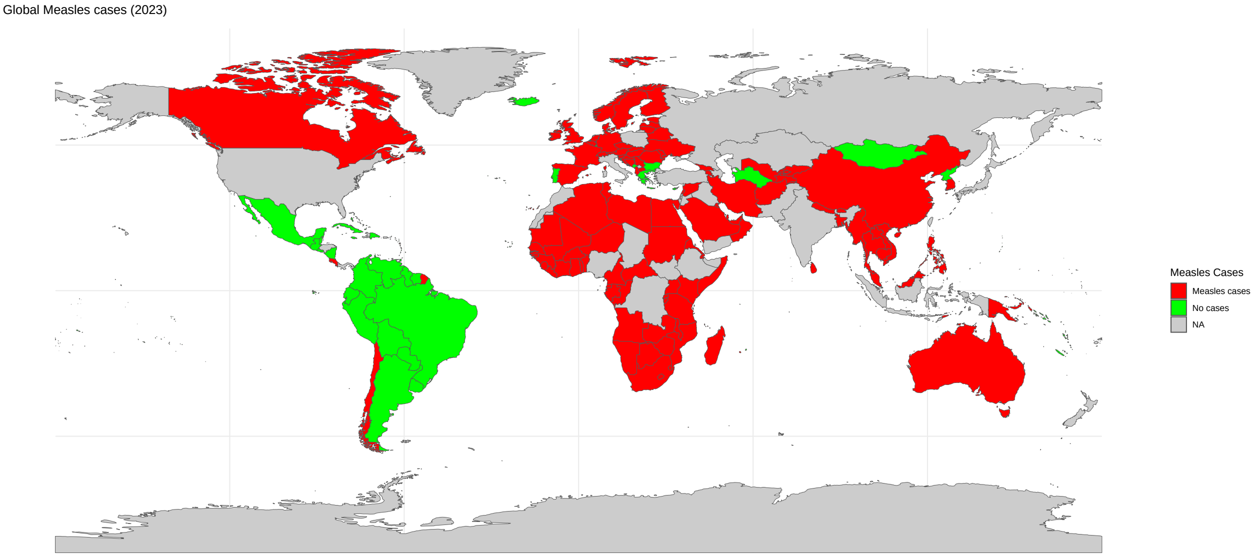


1.5 Disease eradication

To investigate the effectiveness of vaccination and how this has impacted case numbers, shaded maps were created for countries with no reported cases in 2023.

In 2023, there were no reported measles cases in South America (except Chile), Mexico, Iceland, and several other countries. While this is significant, it must be interpreted with some caution, as case reporting could be inaccurate or there may have been misreporting of cases as different diseases.

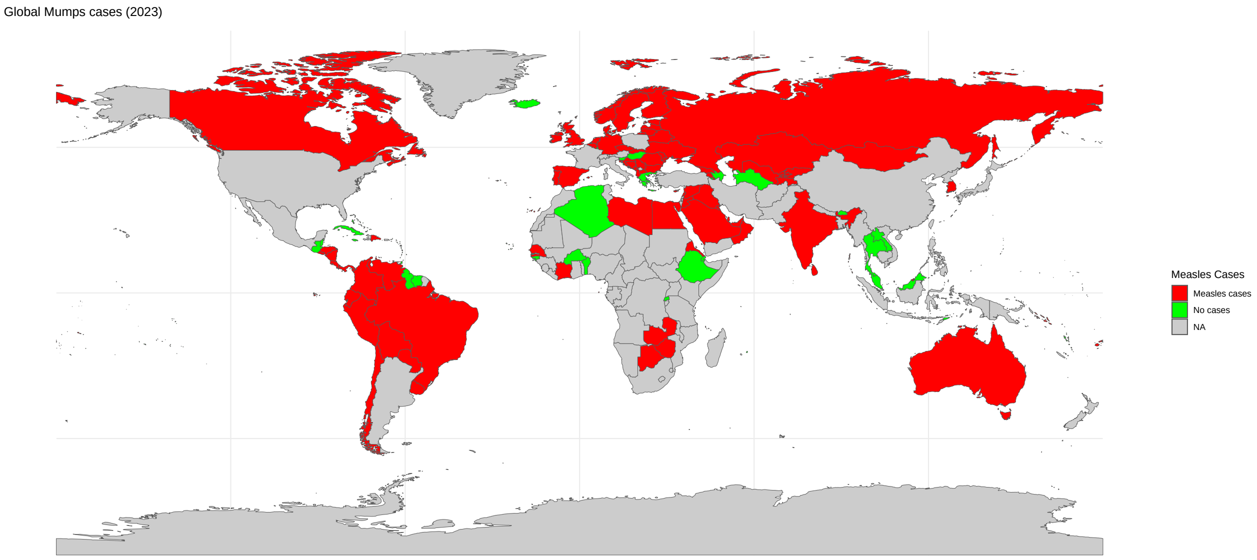
```
world <- ne_countries(scale = "medium", returnclass = "sf")

Measles_cases_2023_classify <- Measles_cases_2023 %>%
  mutate(Has_Cases = ifelse(Value > 0, "Measles cases", "No cases"))

world_map <- world %>%
  left_join(Measles_cases_2023_classify, by = c("adm0_a3" = "SpatialDimValueCode"))

ggplot(data = world_map) +
  geom_sf(aes(fill = Has_Cases)) +
  scale_fill_manual(values = c("Measles cases" = "red", "No cases" = "green"),
  na.value = "gray80") +
  labs(
    title = "Global Measles cases (2023)",
    fill = "Measles Cases"
  ) +
  theme_minimal()
```



Mumps shows a lack of countries with no reported cases, with the only notable exceptions being Iceland and Algeria. While Iceland has a relatively small population, which limits disease transmission, Algeria, with its significantly larger population, has a large landmass, which could reduce the spread of disease. However, like measles, mumps is a highly infectious disease, which explains the absence of many countries with no reported cases.

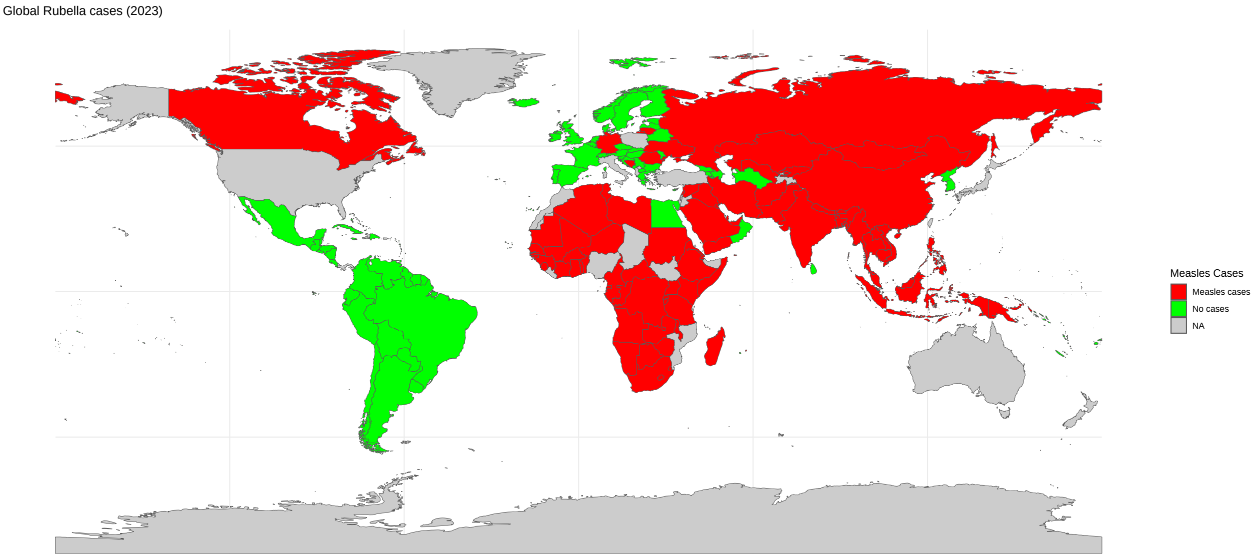
```
world <- ne_countries(scale = "medium", returnclass = "sf")

Mumps_cases_2023_classify <- Mumps_cases_2023 %>%
```

```r
  mutate(Has_Cases = ifelse(Value > 0, "Measles cases", "No cases"))

world_map <- world %>%
  left_join(Mumps_cases_2023_classify, by = c("adm0_a3" = "SpatialDimValueCode"))

ggplot(data = world_map) +
  geom_sf(aes(fill = Has_Cases)) +
  scale_fill_manual(values = c("Measles cases" = "red", "No cases" = "green"),
  na.value = "gray80") +
  labs(
    title = "Global Mumps cases (2023)",
    fill = "Measles Cases"
  ) +
  theme_minimal()
```



Rubella, being less transmissible than measles and mumps, has a significant number of countries reporting no cases. These include the entirety of South America, Western Europe, and the Scandinavian region. As previously mentioned, high global vaccination coverage for rubella has led to a notable reduction in cases, especially when compared to measles and mumps.

```
world <- ne_countries(scale = "medium", returnclass = "sf")

Rubella_cases_2023_classify <- Rubella_cases_2023 %>%
  mutate(Has_Cases = ifelse(Value > 0, "Measles cases", "No cases"))

world_map <- world %>%
  left_join(Rubella_cases_2023_classify, by = c("adm0_a3" = "SpatialDimValueCode"))

ggplot(data = world_map) +
  geom_sf(aes(fill = Has_Cases)) +
  scale_fill_manual(values = c("Measles cases" = "red", "No cases" = "green"),
  na.value = "gray80") +
  labs(
    title = "Global Rubella cases (2023)",
    fill = "Measles Cases"
  ) +
  theme_minimal()
```



1.5.1 Further analysis

To test the significance of the number of vaccinations between specific years = ANOVA

Vaccine uptake relationship with time, magnitude, prediction, goodness of fit and significance = Linear regression

Predict future vaccine uptake based on previous data = Forecasting

1.6 Conclusions

This report investigates and analyses the vaccine-preventable diseases measles, mumps, and rubella. Case numbers and vaccination data were cleaned, summarised, and correlated with time between 2000 and 2023 to assess global changes over 23 years in both vaccination coverage and case numbers, and how they relate to each other. It was found that both measles and mumps consistently had high case numbers globally, while rubella always had the lowest case numbers. Upon closer inspection, this was likely due to high vaccination coverage for rubella. While the cases of measles and mumps have remained high (although a drop in cases during the COVID pandemic is worth noting), vaccination coverage has generally increased since 2000. However, this trend must be maintained to develop effective herd immunity. Proper education and government efforts are necessary to ensure ongoing vaccination and prevent future outbreaks.