EDA

2026-02-02

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(patchwork)
```

```
## Warning: package 'patchwork' was built under R version 4.5.2
```

```r
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.5.2
```

```
## Registered S3 method overwritten by 'quantmod':
##    method            from
##    as.zoo.data.frame zoo
```

```r
df <- read.csv("../df.csv")
```

```r
impact_start = as.POSIXct("2020-03-01")
impact_end   = as.POSIXct("2023-03-01")

df$period = factor(df$period, levels=c("pre", "impacted", "post"))

p1 <- ggplot(df, aes(x=as.POSIXct(month_turn_2), y=rate, color=period)) +
  geom_line(linewidth=1) +
  geom_point(size=1.5) +
  geom_smooth(aes(group=period), method="lm", se=FALSE, linewidth=0.6, color="black")+
  scale_y_continuous(limits = c(60,75)) +
  geom_vline(xintercept =  impact_start, linetype = "dashed", color="gray50")+
  geom_vline(xintercept =  impact_end, linetype = "dashed", color="gray50" ) +
  labs(
```

```
    title = "Measles Vaccination Rate over Time - Slope",
    x = "Month",
    y = "Vaccination Rate",
    color = "Period"
  ) +
  theme_classic(base_size = 13) +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = "right",
    legend.title = element_text(size=11),
    axis.title = element_text(size=12),
    axis.text  = element_text(size=11)
  )

ggsave(p1, filename="../figures/1_whole_plot_with_slop.pdf", width=9, height=4.5, units="in")
```
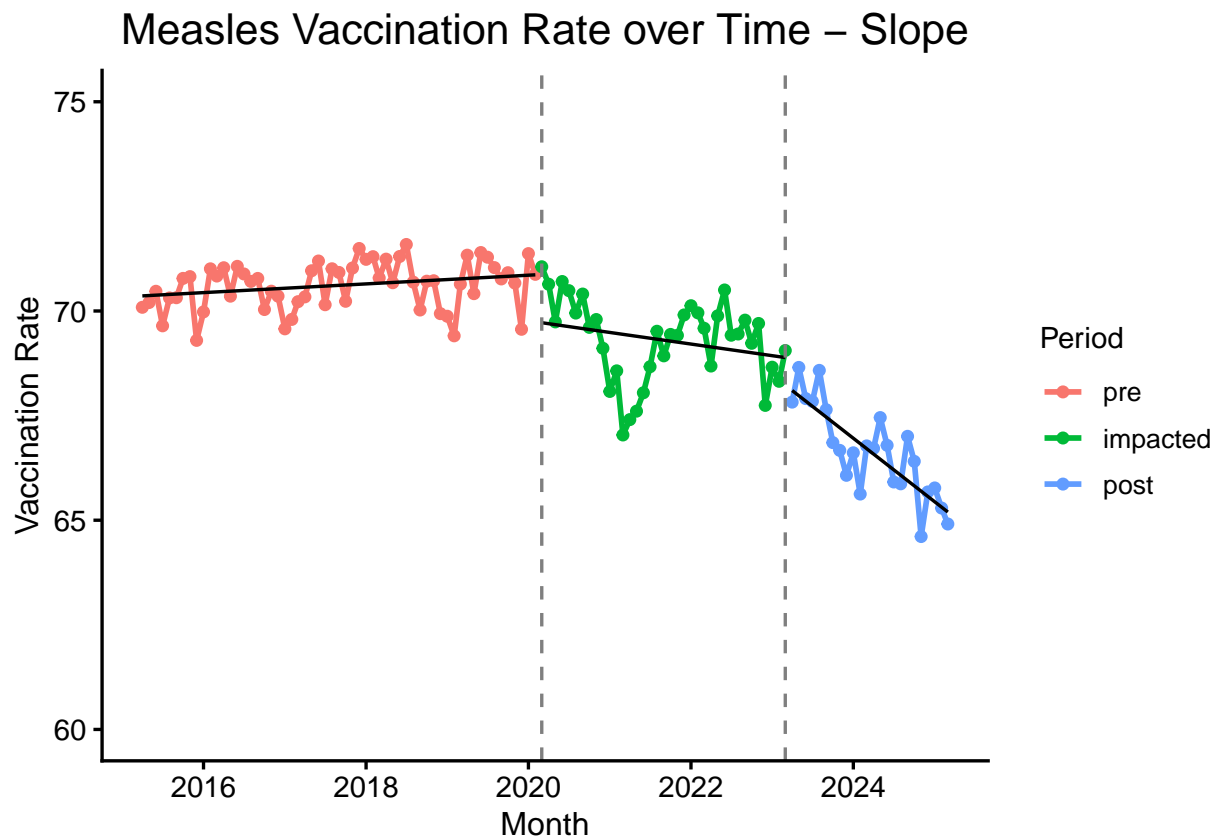
`geom_smooth()` using formula = 'y ~ x'

```
p1
```

`geom_smooth()` using formula = 'y ~ x'

```r
period_means <- df %>%
  group_by(period) %>%
  summarise(
    mean_vax = mean(rate),
    xmin = min(as.POSIXct(month_turn_2)),
    xmax = max(as.POSIXct(month_turn_2)),
    .groups = "drop"
  )

p2 <- ggplot(df, aes(x=as.POSIXct(month_turn_2), y=rate, color=period)) +
  geom_line(linewidth=1) +
  geom_point(size=1.5) +
  geom_segment(data = period_means,
               aes(x=xmin, xend=xmax, y=mean_vax, yend=mean_vax), inherit.aes = FALSE,
               color = "black",
               linewidth=0.6) +
  scale_y_continuous(limits = c(60,75)) +
  geom_vline(xintercept =  impact_start, linetype = "dashed", color="gray50")+
  geom_vline(xintercept =  impact_end, linetype = "dashed", color="gray50" ) +
  labs(
    title = "Measles Vaccination Rate over Time - Average",
    x = "Month",
    y = "Vaccination Rate",
    color = "Period"
  ) +
  theme_classic(base_size = 13) +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.position = "right",
    legend.title = element_text(size=11),
    axis.title = element_text(size=12),
    axis.text  = element_text(size=11)
  )

ggsave(p2, filename="../figures/2_whole_plot_with_mean.pdf", width=9, height=4.5, units="in")

p2
```
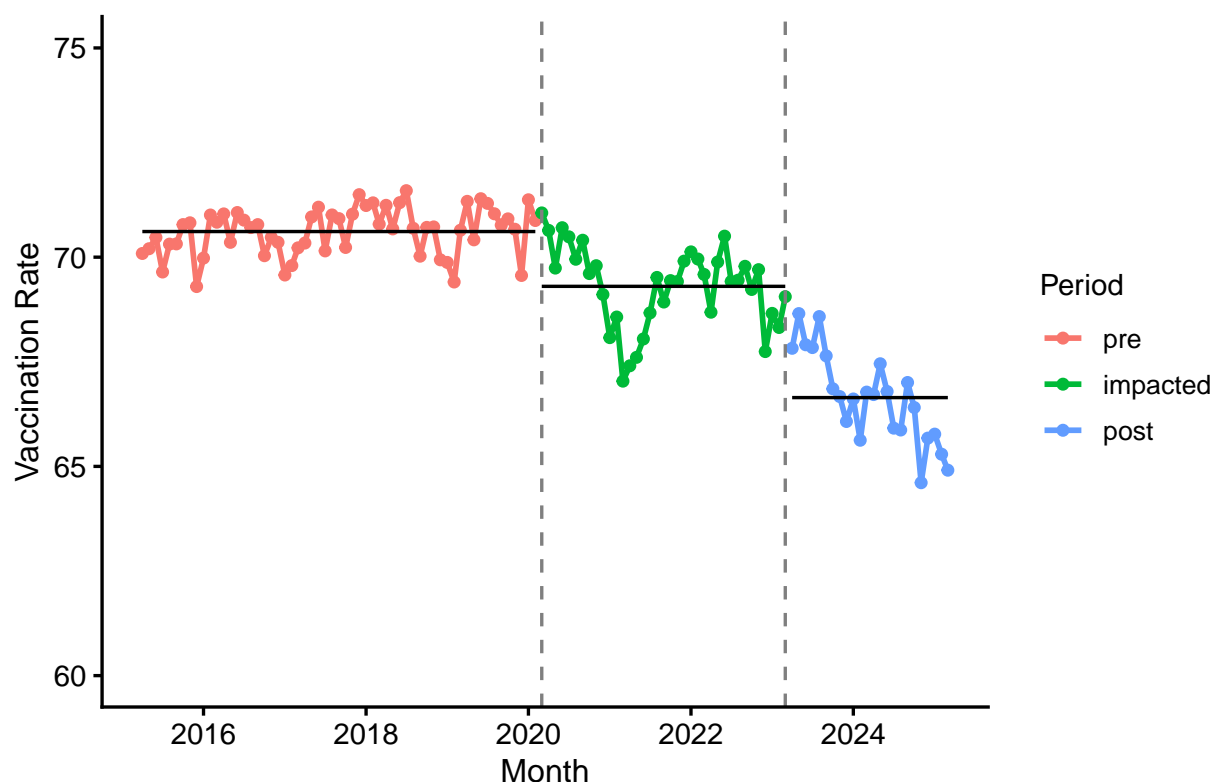
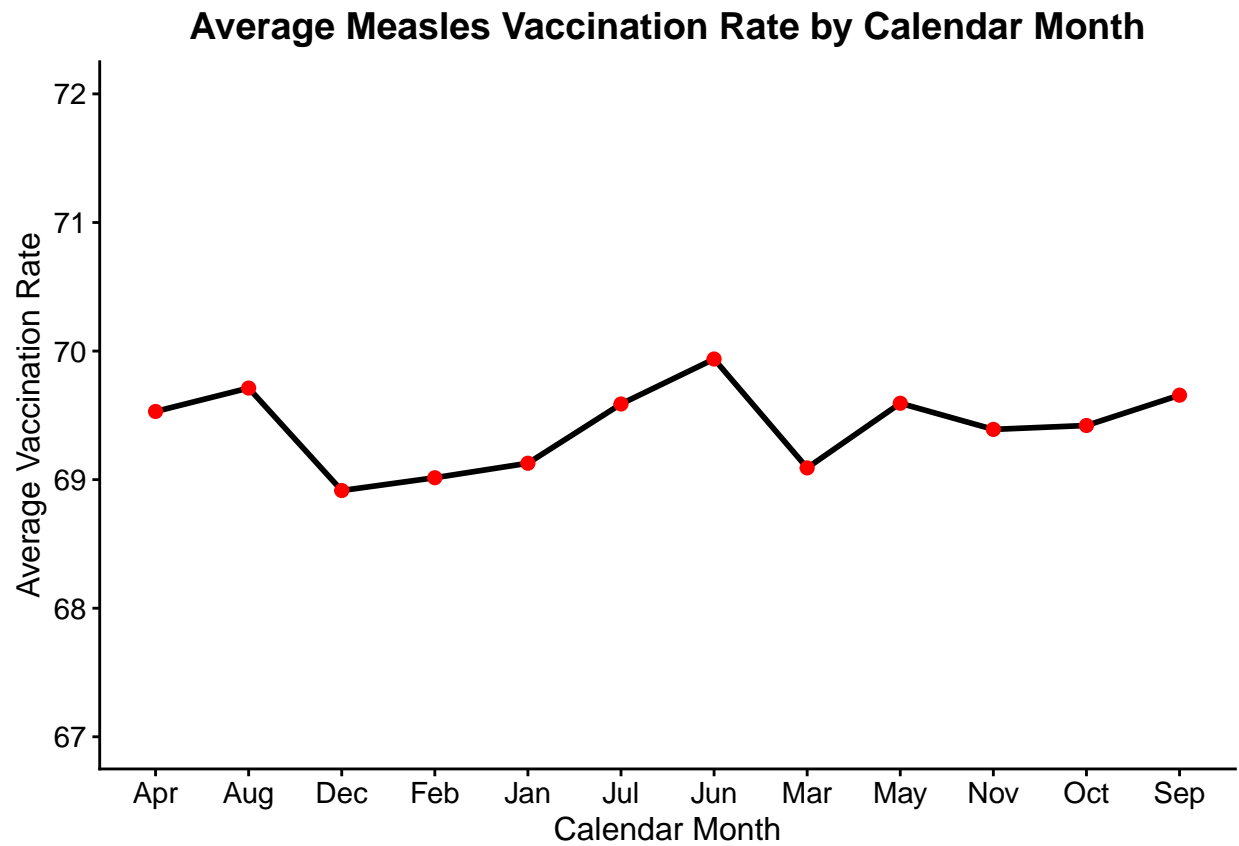# Measles Vaccination Rate over Time – Average


```
# df$calendar_month <- factor(
#   df$calendar_month,
#   levels = 1:12,
#   labels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
#              "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")
#   )


monthly_avg <- df %>% group_by(calendar_month) %>%
    summarise(
      mean_vax = mean(rate),
      .groups = "drop"
    )



p3 <- ggplot(monthly_avg, aes(calendar_month, mean_vax)) +
  geom_line(group=1, linewidth = 1) +
  geom_point(size=2, color="red")+
  labs(title = "Average Measles Vaccination Rate by Calendar Month",
       x    = "Calendar Month",
       y    = "Average Vaccination Rate") +
  scale_y_continuous(limits = c(67,72)) +
  theme_classic() +
  theme(plot.title = element_text(hjust=0.5, size=14, face="bold"),
        axis.title= element_text(size=12),
```

```
        axis.text = element_text(size=11)
        )
p3
```

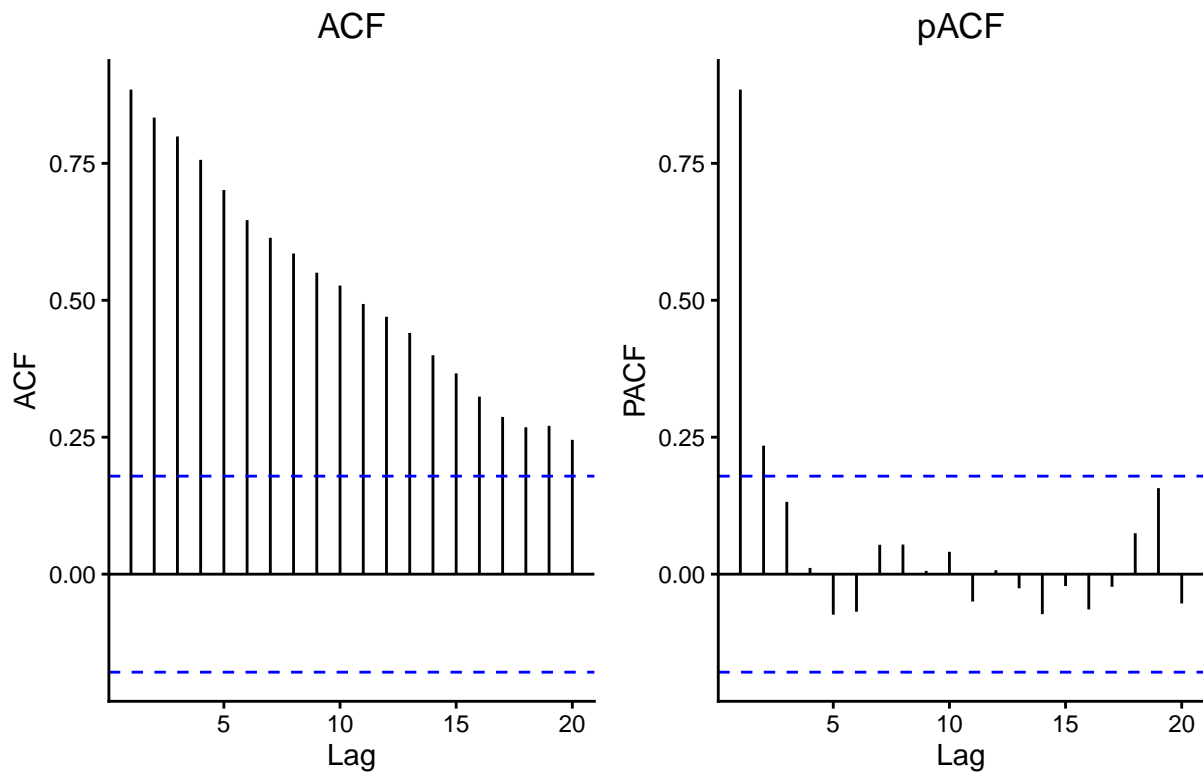**Average Measles Vaccination Rate by Calendar Month**


```
ggsave(p3, filename="../figures/3_seasonality.pdf", width=9, height=4.5, units="in")
```

```
p_acf <- ggAcf(df$rate) + labs(title="ACF") + theme_classic() + theme(plot.title = element_text(hjust=0
p_pacf <- ggPacf(df$rate) + labs(title="pACF") + theme_classic() + theme(plot.title = element_text(hjus

p4 <- (p_acf + p_pacf) + plot_annotation(title="Autocorrelation Diagnosis")
p4
```
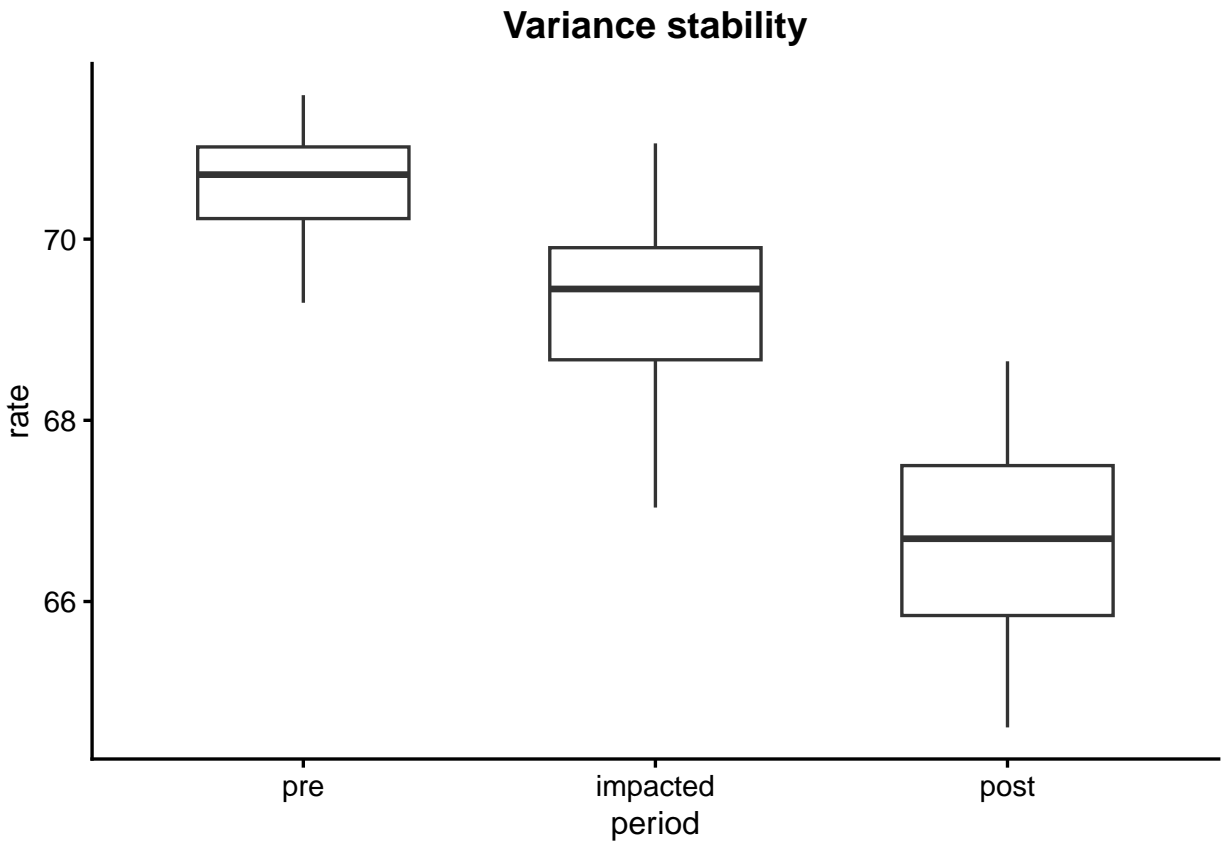
Autocorrelation Diagnosis


```
ggsave(p4, filename="../figures/4_autocorrelation.pdf", width=9, height=4.5, units="in")
```

Variance stability

```
ggplot(df, aes(x=period, y=rate)) +
  geom_boxplot(width=0.6, outlier.size=1.5) +
  theme_classic(base_size = 13) +
  labs(title = "Variance stability") +
  theme(plot.title = element_text(hjust=0.5, size=14, face="bold"),
        axis.title= element_text(size=12),
        axis.text = element_text(size=11)
        )
```

Variance stability

) "'