Autocorrelation function (ACF)

Saurabh

2025-05-04

a) Exploratory data analysis

- Visualization: line plots to visualize trends and scatter plots with a regression line to assess the correlation
- Correlation analysis: Pearson correlation coefficient to measure the relationship's strength.

📈 1. Line Plot: CPI Over Time

```r
# Load required libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```r
library(readr)
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(tidyr)

# Load CPI data
cpi_data <- read_csv("No Header_ConsumerPriceIndex.csv")
```
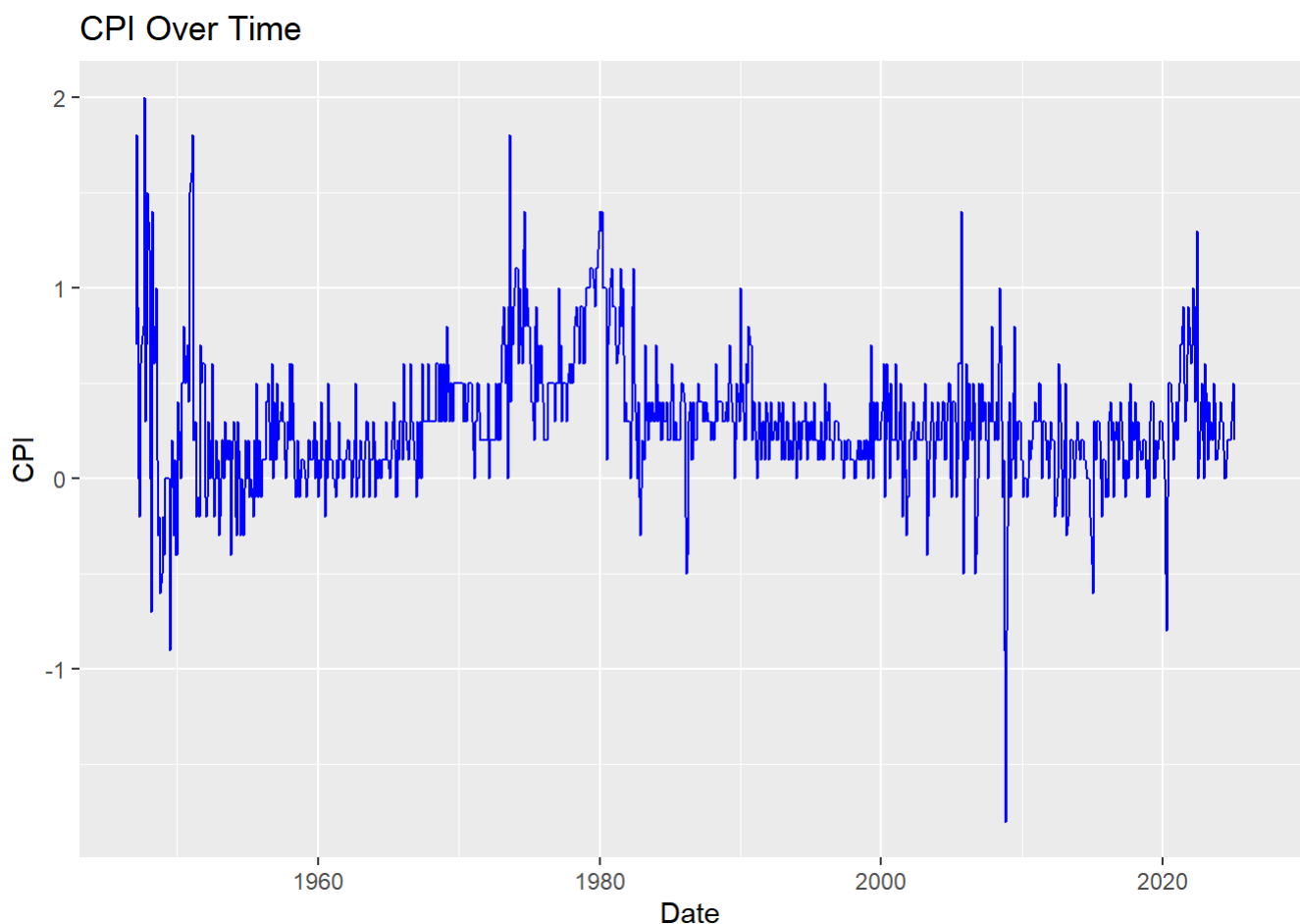
```
## Rows: 79 Columns: 13
```

```
## ── Column specification ──────────────────────────────────────────────
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Transform data
cpi_long <- cpi_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "cpi") |>
  rename(year = Year)

# Create date variable
cpi_long <- cpi_long |>
  mutate(date = as.Date(paste(year, month, "01", sep = "-"), format = "%Y-%B-%d")) |>
  arrange(date)

# Plot CPI over time
ggplot(cpi_long, aes(x = date, y = cpi)) +
  geom_line(color = "blue") +
  labs(title = "CPI Over Time", x = "Date", y = "CPI")
```

```
## Warning: Removed 11 rows containing missing values or values outside the scale range
## (`geom_line()`).
```


CPI Over Time

📈 2. Line Plot: Unemployment Rate Over Time

```r
# Load required libraries
library(ggplot2)
library(readr)
library(dplyr)
library(tidyr)

# Load Unemployment Rate data
unemp_data <- read_csv("NoHeader_UnemploymentRate.csv")
```

```
## Rows: 78 Columns: 13
## — Column specification ————————————————————————————————————
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Transform data
unemp_long <- unemp_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "unemployment_rate") |>
  rename(year = Year)

# Create date variable
unemp_long <- unemp_long |>
  mutate(date = as.Date(paste(year, month, "01", sep = "-"), format = "%Y-%B-%d")) |>
  arrange(date)

# Plot Unemployment Rate over time
ggplot(unemp_long, aes(x = date, y = unemployment_rate)) +
  geom_line(color = "red") +
  labs(title = "Unemployment Rate Over Time", x = "Date", y = "Unemployment Rate")
```

```
## Warning: Removed 10 rows containing missing values or values outside the scale range
## (`geom_line()`).
```

Unemployment Rate Over Time
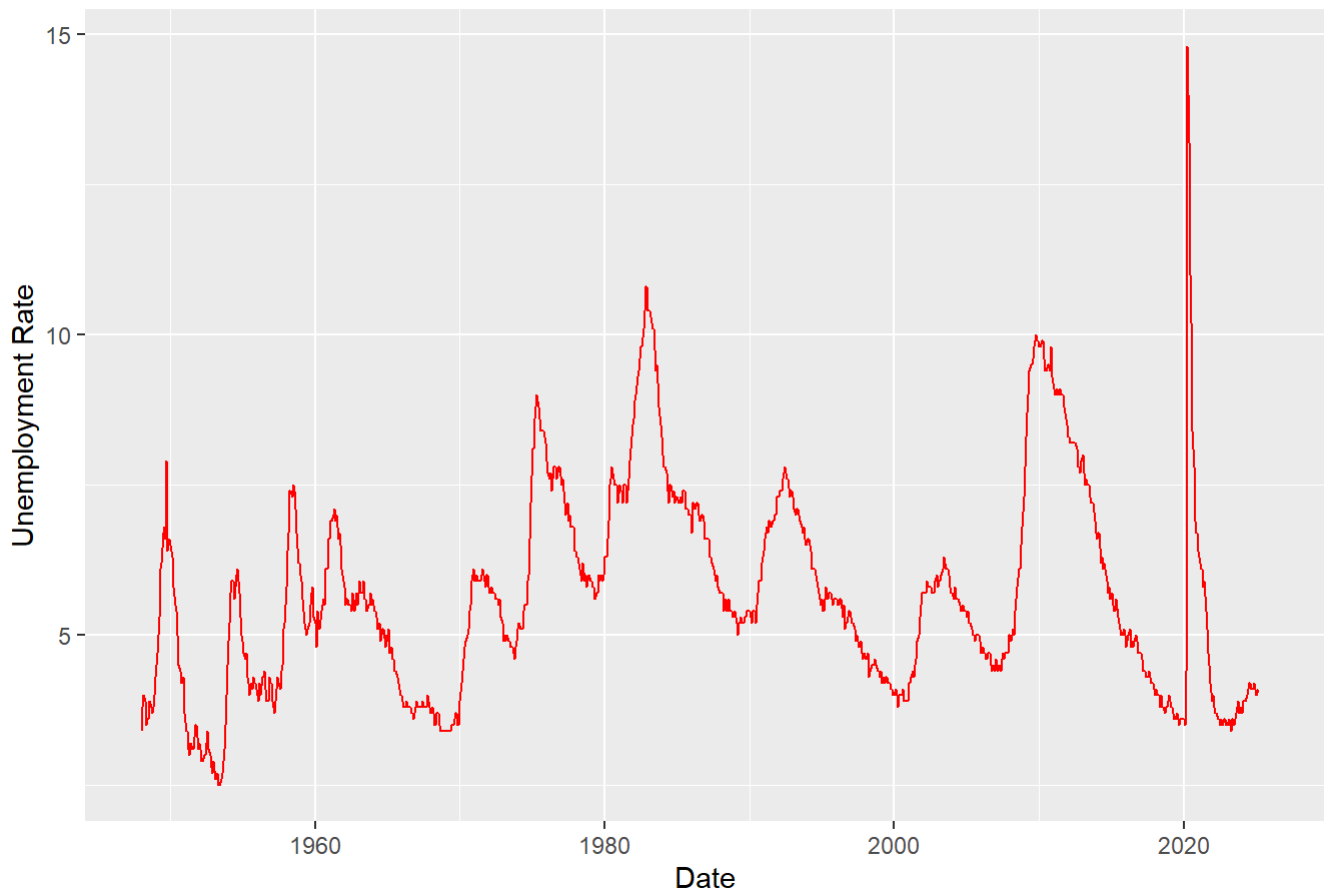
🔵 3. Scatter Plot with Regression Line: CPI vs Unemployment

```r
# Load required libraries
library(ggplot2)
library(readr)
library(dplyr)
library(tidyr)

# Load CPI Data
cpi_data <- read_csv("No Header_ConsumerPriceIndex.csv")
```

```
## Rows: 79 Columns: 13
## — Column specification ————————————————————————————————
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Load Unemployment Rate Data
unemp_data <- read_csv("NoHeader_UnemploymentRate.csv")
```

```
## Rows: 78 Columns: 13
## ── Column specification ──────────────────────────────────────────────────
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Transform CPI
cpi_long <- cpi_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "cpi") |>
  rename(year = Year)

# Transform Unemployment Rate
unemp_long <- unemp_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "unemployment_rate") |>
  rename(year = Year)

# Merge datasets
combined_data <- left_join(cpi_long, unemp_long, by = c("year", "month"))

# Create date column
combined_data <- combined_data |>
  mutate(date = as.Date(paste(year, month, "01", sep = "-"), format = "%Y-%B-%d")) |>
  arrange(date)

# Remove missing values
combined_data <- combined_data |>
  filter(!is.na(cpi), !is.na(unemployment_rate))

# Scatter plot with linear regression
ggplot(combined_data, aes(x = cpi, y = unemployment_rate)) +
  geom_point(color = "darkgreen") +
  geom_smooth(method = "lm", se = TRUE, color = "black") +
  labs(title = "Unemployment Rate vs CPI", x = "CPI", y = "Unemployment Rate")
```
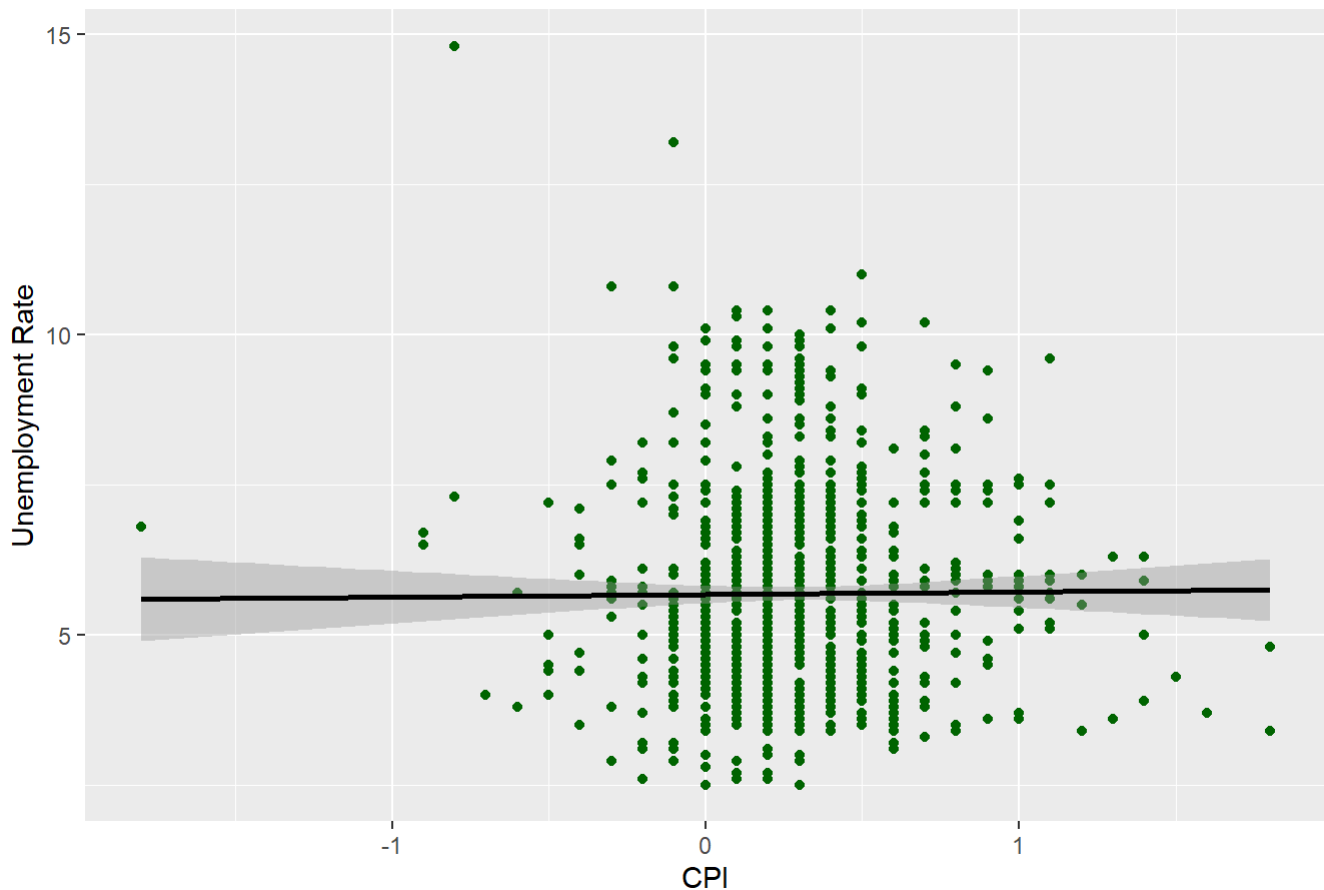
```
## `geom_smooth()` using formula = 'y ~ x'
```

Unemployment Rate vs CPI

📊 4. Pearson Correlation Coefficient

```
correlation <- cor(combined_data$cpi, combined_data$unemployment_rate, method = "pearson")
print(paste("Pearson Correlation Coefficient:", round(correlation, 3)))
```

```
## [1] "Pearson Correlation Coefficient: 0.008"
```

```
# Load necessary libraries
library(ggplot2)
library(tseries)  # For ACF function
```

```
## Warning: package 'tseries' was built under R version 4.4.2
```

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

```
library(dplyr)
library(readr)

# Load and prepare CPI data (assuming CSV files are properly structured)
cpi_data <- read_csv("No Header_ConsumerPriceIndex.csv")
```

```
## Rows: 79 Columns: 13
```

```
## — Column specification ——————————————————————————————————
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Load and prepare Unemployment Rate data
unemp_data <- read_csv("NoHeader_UnemploymentRate.csv")
```

```
## Rows: 78 Columns: 13
## — Column specification ——————————————————————————————————
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Transform CPI data (from wide to long format)
cpi_long <- cpi_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "cpi") |>
  rename(year = Year)

# Transform Unemployment Rate data (from wide to long format)
unemp_long <- unemp_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "unemployment_rate") |>
  rename(year = Year)

# Merge CPI and Unemployment Rate data
combined_data <- left_join(cpi_long, unemp_long, by = c("year", "month"))

# Create a Date column for easier plotting
combined_data <- combined_data |>
  mutate(date = as.Date(paste(year, month, "01", sep = "-"), format = "%Y-%B-%d")) |>
  arrange(date)

# Remove missing values
combined_data <- combined_data |>
  filter(!is.na(cpi), !is.na(unemployment_rate))

# Convert to time series objects
cpi_ts <- ts(combined_data$cpi, start = c(min(combined_data$year), 1), frequency = 12)
unemp_ts <- ts(combined_data$unemployment_rate, start = c(min(combined_data$year), 1), freque
ncy = 12)


# ===============================================
# ACF Plots for CPI and Unemployment Rate
# ===============================================

# ACF for CPI (original series)
acf(na.omit(cpi_ts), main = "ACF of CPI (Original Series)", lag.max = 40)
```
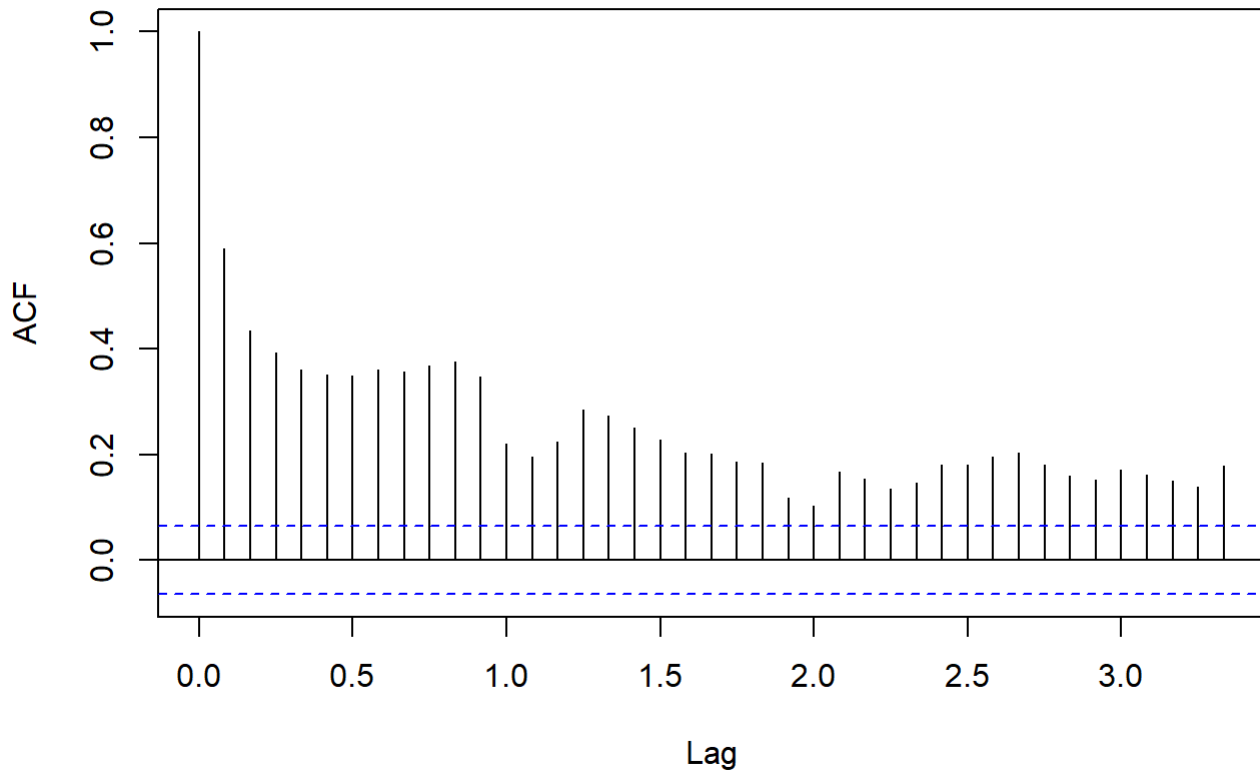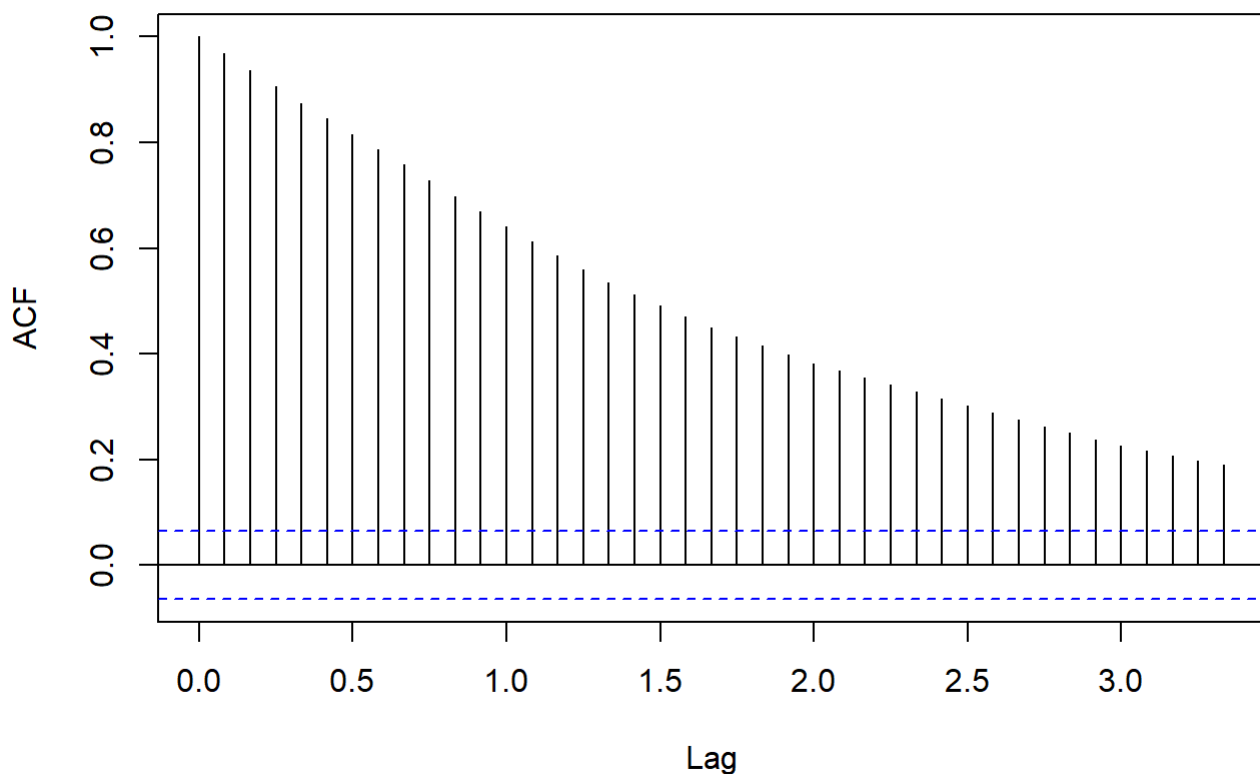
ACF of CPI (Original Series)


```
# ACF for Unemployment Rate (original series)
acf(na.omit(unemp_ts), main = "ACF of Unemployment Rate (Original Series)", lag.max = 40)
```
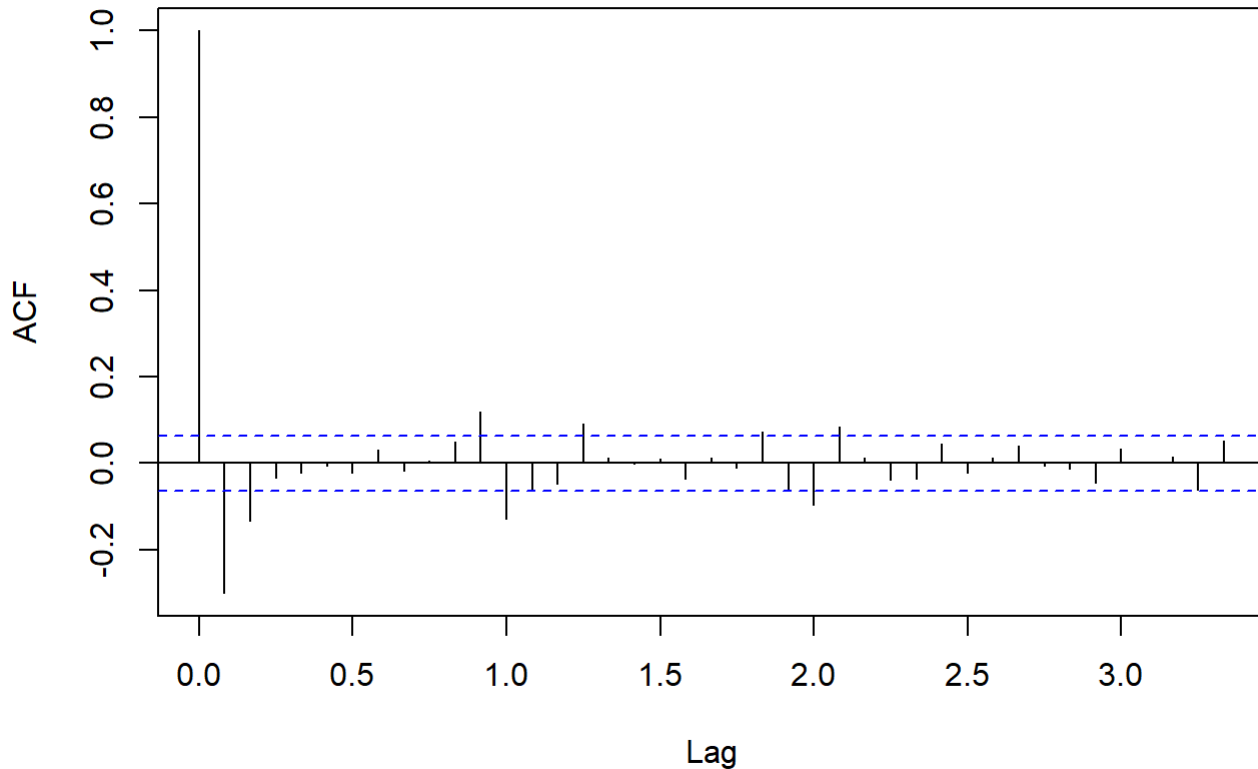
ACF of Unemployment Rate (Original Series)

```
# ACF for differenced CPI (to check for stationarity if needed)
diff_cpi_ts <- diff(na.omit(cpi_ts))
acf(diff_cpi_ts, main = "ACF of Differenced CPI", lag.max = 40)
```

ACF of Differenced CPI


```
# ACF for differenced Unemployment Rate (to check for stationarity if needed)
diff_unemp_ts <- diff(na.omit(unemp_ts))
acf(diff_unemp_ts, main = "ACF of Differenced Unemployment Rate", lag.max = 40)
```

ACF of Differenced Unemployment Rate
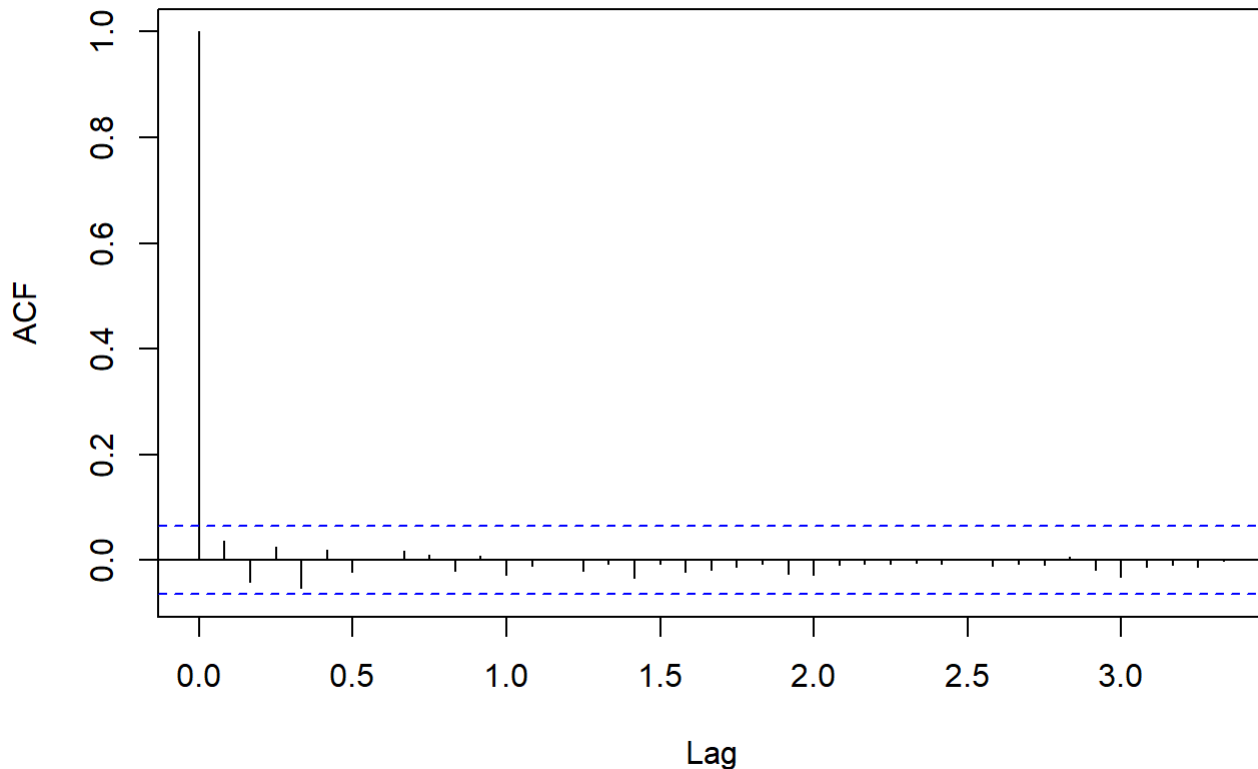

```
# Make sure these libraries are loaded
library(readr)
library(dplyr)
library(tidyr)

# Load CPI data
cpi_data <- read_csv("No Header_ConsumerPriceIndex.csv")
```

```
## Rows: 79 Columns: 13
## ── Column specification ─────────────────────────────────────────────
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Load Unemployment Rate data
unemp_data <- read_csv("NoHeader_UnemploymentRate.csv")
```

```
## Rows: 78 Columns: 13
## ── Column specification ─────────────────────────────────────────────
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Pivot CPI
cpi_long <- cpi_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "cpi") |>
  rename(year = Year)

# Pivot Unemployment
unemp_long <- unemp_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "unemployment_rate") |>
  rename(year = Year)

# Merge datasets
combined_data <- left_join(cpi_long, unemp_long, by = c("year", "month"))

# Create Date column
combined_data <- combined_data |>
  mutate(date = as.Date(paste(year, month, "01", sep = "-"), format = "%Y-%B-%d")) |>
  arrange(date) |>
  filter(!is.na(cpi), !is.na(unemployment_rate))

# Create ts objects
cpi_ts <- ts(combined_data$cpi, start = c(min(combined_data$year), 1), frequency = 12)
unemp_ts <- ts(combined_data$unemployment_rate, start = c(min(combined_data$year), 1), freque
ncy = 12)
```
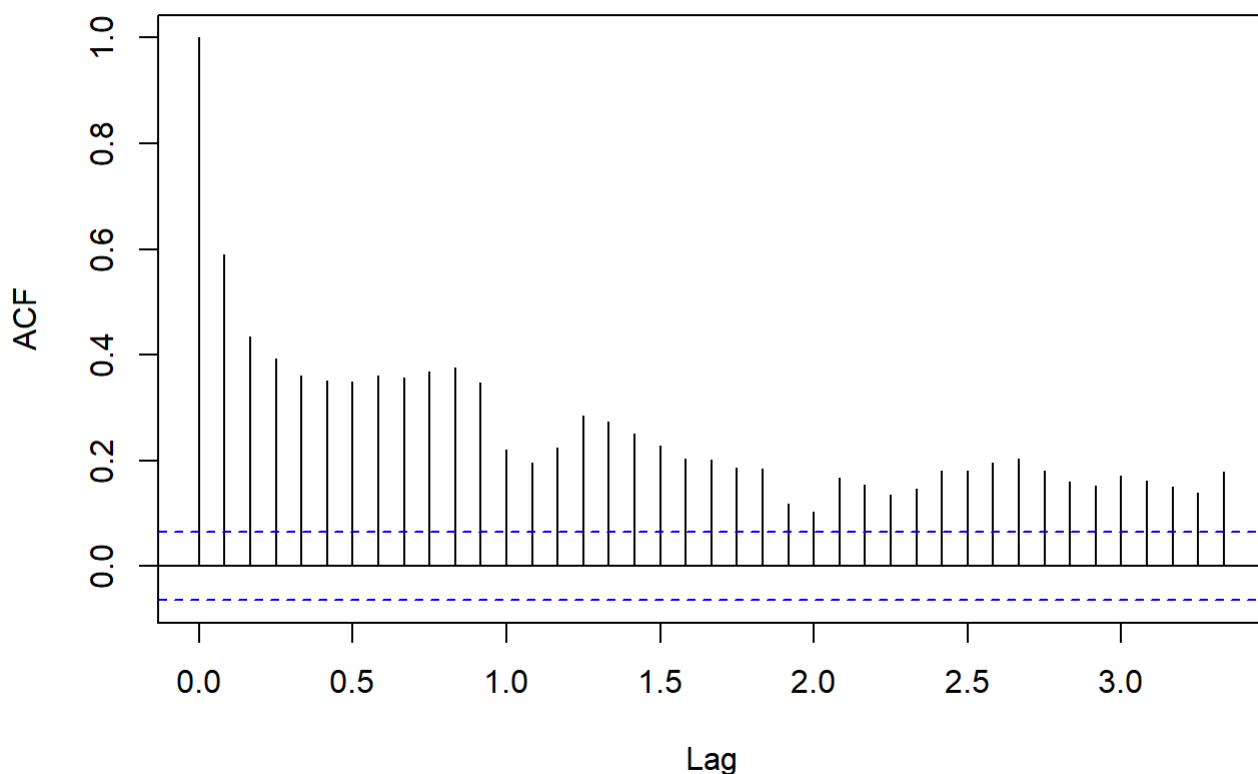
```
# ACF for CPI (original series)
acf(na.omit(cpi_ts), main = "ACF of CPI (Original Series)", lag.max = 40)
```
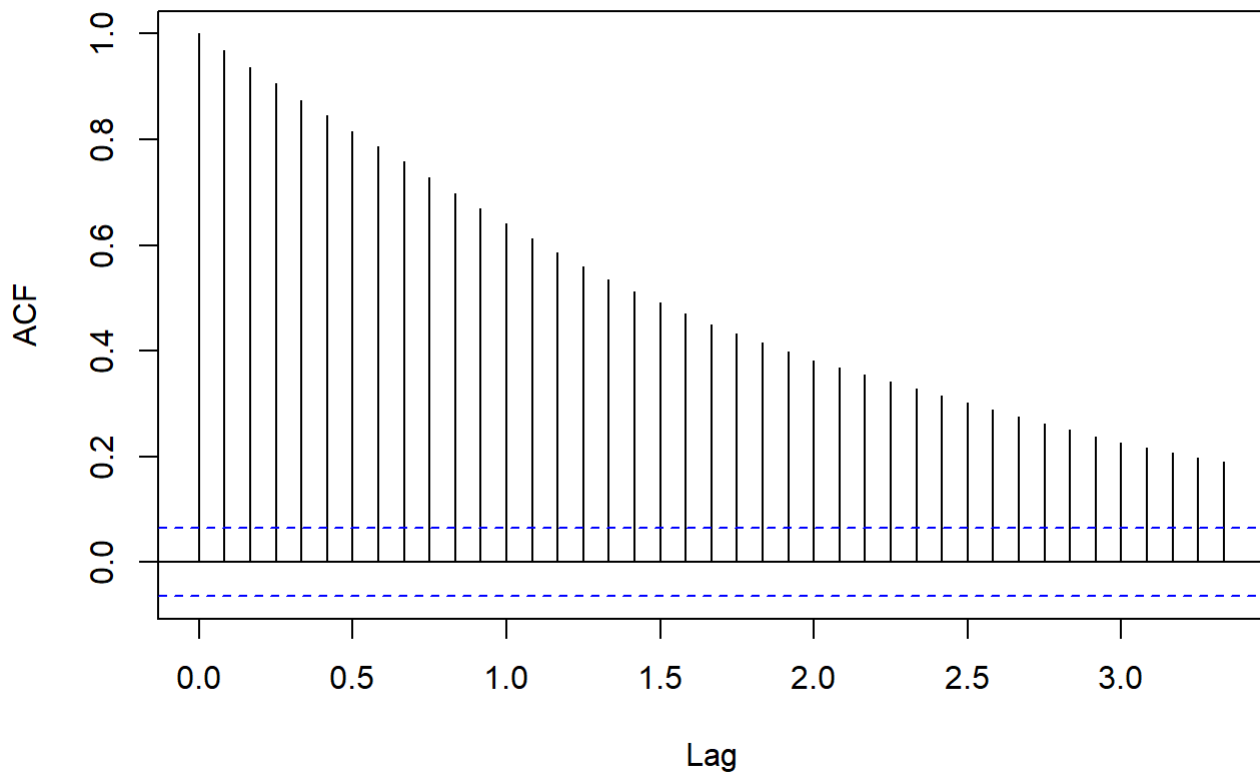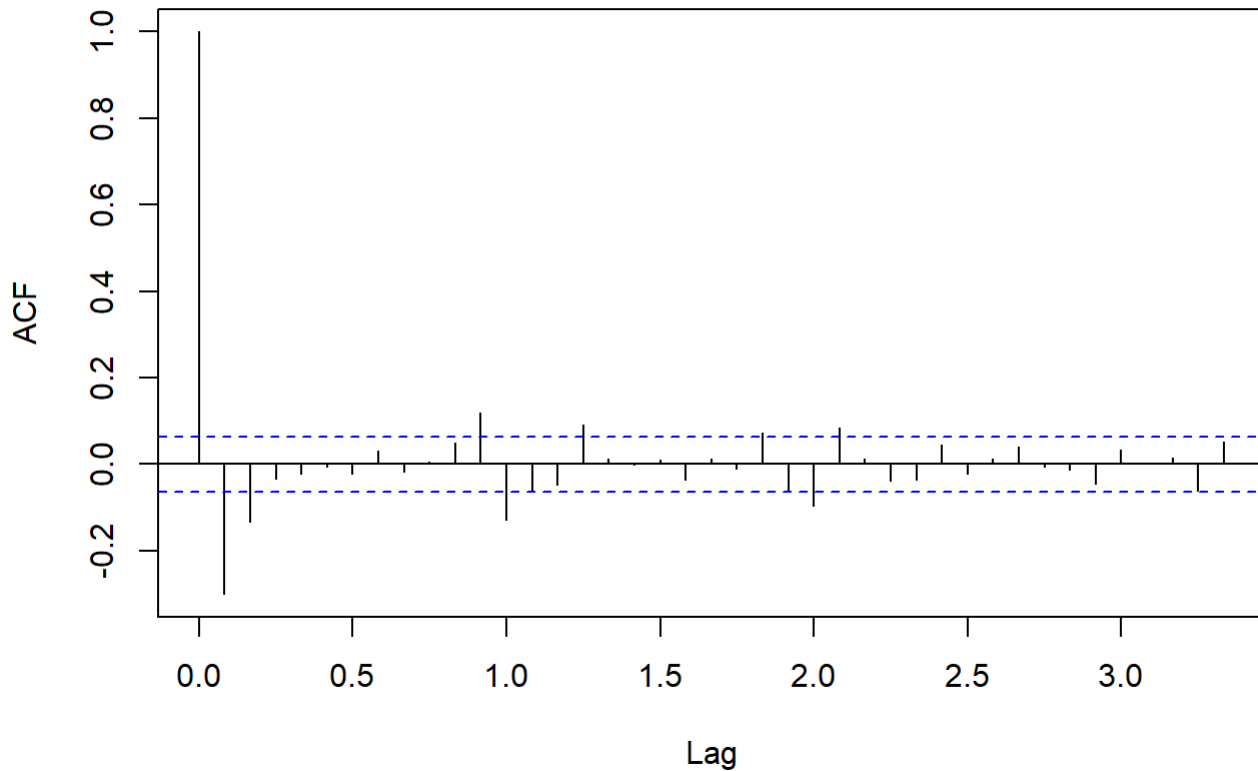
ACF of CPI (Original Series)

```
# ACF for Unemployment Rate (original series)
acf(na.omit(unemp_ts), main = "ACF of Unemployment Rate (Original Series)", lag.max = 40)
```
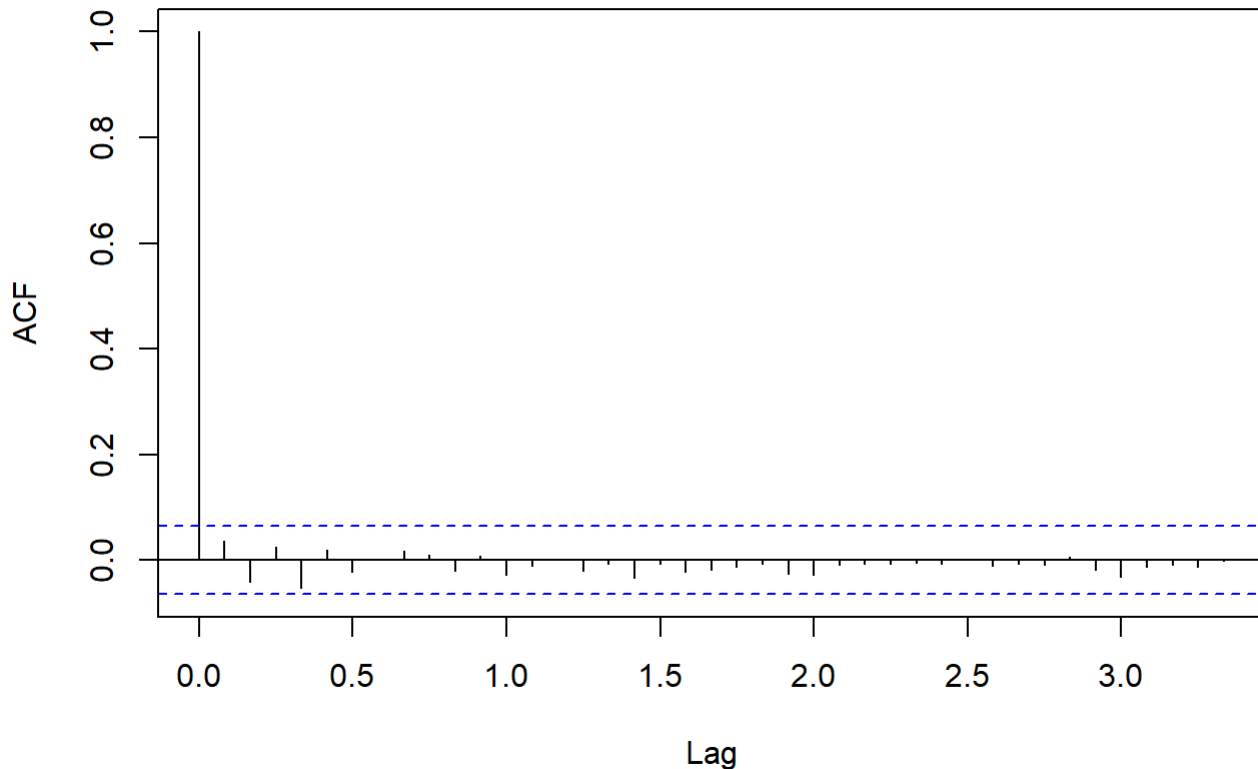
ACF of Unemployment Rate (Original Series)


```
# ACF for differenced CPI
diff_cpi_ts <- diff(na.omit(cpi_ts))
acf(diff_cpi_ts, main = "ACF of Differenced CPI", lag.max = 40)
```

ACF of Differenced CPI


```
# ACF for differenced Unemployment Rate
diff_unemp_ts <- diff(na.omit(unemp_ts))
acf(diff_unemp_ts, main = "ACF of Differenced Unemployment Rate", lag.max = 40)
```

ACF of Differenced Unemployment Rate


```
# ================================================
# Simple Linear Regression: CPI vs Unemployment
# ================================================

# Load libraries (if not already loaded)
library(readr)
library(dplyr)
library(tidyr)

# Load and prepare data (if not already done)
cpi_data <- read_csv("No Header_ConsumerPriceIndex.csv")
```

```
## Rows: 79 Columns: 13
## — Column specification ———————————————————————————————————————
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
unemp_data <- read_csv("NoHeader_UnemploymentRate.csv")
```

```
## Rows: 78 Columns: 13
## ── Column specification ───────────────────────────────────────────────────
## Delimiter: ","
## dbl (13): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Transform data (long format)
cpi_long <- cpi_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "cpi") |>
  rename(year = Year)

unemp_long <- unemp_data |>
  pivot_longer(cols = -Year, names_to = "month", values_to = "unemployment_rate") |>
  rename(year = Year)

# Merge and clean
combined_data <- left_join(cpi_long, unemp_long, by = c("year", "month")) |>
  mutate(date = as.Date(paste(year, month, "01", sep = "-"), format = "%Y-%B-%d")) |>
  arrange(date) |>
  filter(!is.na(cpi), !is.na(unemployment_rate))

# ===============================================
# Run Simple Linear Regression
# ===============================================

# Model: Unemployment Rate ~ CPI
slr_model <- lm(unemployment_rate ~ cpi, data = combined_data)

# View model summary
summary(slr_model)
```

```
##
## Call:
## lm(formula = unemployment_rate ~ cpi, data = combined_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1807 -1.3426 -0.1807  1.0269  9.1672
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.66763    0.07368  76.924   <2e-16 ***
## cpi          0.04355    0.16904   0.258    0.797
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.71 on 924 degrees of freedom
## Multiple R-squared:  7.182e-05,  Adjusted R-squared:  -0.00101
## F-statistic: 0.06637 on 1 and 924 DF,  p-value: 0.7968
```