Inflation: Finding the Correlations to Build a Machine Learning Model for Backtesting and Forecasting

Arthur 'Rich' Richardson

9/5/2022

This Project is dedicated to my two daughters.

I love you both equally, from the furthest Sun our Universe will ever know and back!

Love, Dad.

Installation:

Note: *Due to known loading issues, load the following package(s) dependency(ies), if needed*

```r
install.packages("hms", dependencies = TRUE)
install.packages("gtable", dependencies = TRUE)
install.packages("hexbin", dependencies = TRUE)
install.packages("readr", dependencies=TRUE, INSTALL_opts = c('--no-lock'))
install.packages("caret", dependencies = TRUE)
install.packages("data.table", dependencies = TRUE)
install.packages("tidyverse", dependencies = TRUE)
install.packages("Rcolorbrewer", dependencies = TRUE)
install.packages("gt", dependencies = TRUE)
install.packages("lubridate", dependencies = TRUE)
install.packages("highcharter", dependencies = TRUE)
install.packages("ggpmisc", dependencies = TRUE)
install.packages("janitor", dependencies = TRUE)
install.packages("scales", dependencies = TRUE)
install.packages("quantmod", dependencies = TRUE)
install.packages("forecast", dependencies = TRUE)
install.packages("ggfortify", dependencies = TRUE)
```

Load the following Library(ies)

```r
library(tidyverse)
library(caret)
library(data.table)
library(RColorBrewer)
library(rmarkdown)
library(dslabs)
library(gtable)
library(hexbin)
library(gt)
library(dplyr)
library(ggpmisc)
library(gridExtra)
library(janitor)
library(lubridate)
library(highcharter)
library(viridisLite)
library(broom)
library(scales)
library(xfun)
library(htmltools)
library(mime)
library(quantmod)
library(forecast)
library(tseries)
library(ggfortify)
library(png)
library(jpeg)
library(gtsummary)
library(latexpdf)
library(tinytex)
```

For Lower Bandwidth/ RAM recommend adjusting the timeout settings

```
options(timeout = 320)
```

Depending on your RAM, to free up unused memory, recommend using

```
gc()
```

Required data sets are embedded for download:

```
xfun::pkg_load2(c("htmltools", "mime"))
xfun::embed_files(c('SP5001913.csv',
                    'US CPI.csv', 'Inflation_Rate_Fed_Rate_1913_2017 - Sheet1.csv'))
```

Note: For S&P original file see footnotes. [1]

Load S&P 500 Dataset.

Note: *The document is embedded for download. Insert your file path of where you saved the document*

```
SP500_Data <- read.csv(
  'SP5001913.csv')
```

Note: For Original US CPI file see footnotes [2]

Load US CPI Dataset.

Note: *The document is embedded for download. Insert your file path of where you saved the document*

```
CPI_Data <- read.csv(
  'US CPI.csv')
```

Note: For Original Fed Funds Rate data table see footnotes. [3]

Load US Fed Dataset.

Note: *The document is embedded for download. Insert your file path of where you saved the document*

```
Fed_Data <- read.csv(
  'Inflation_Rate_Fed_Rate_1913_2017 - Sheet1.csv')
```

[1] ***You can download the S&P 500 file from*** (https://datahub.io/core/s-and-p-500/r/data.csv)

[2] ***US CPI file located here*** (https://www.kaggle.com/datasets/varpit94/us-inflation-data-updated-till-may-2021?select=US+CPI.csv)

[3] ***Fed Funds table is located here*** (https://www.thebalance.com/u-s-inflation-rate-history-by-year-and-forecast-3306093)

Reference Section:

1. Bond Yields and Interest Rates: 1900 to 2002. (2003). US CENSUS. Retrieved August 18, 2022, from (https://www2.census.gov/library/publications/2004/compendia/statab/ 123ed/hist/hs-39.pdf) The U.S. Census tracked The 3 Month Bond Yield from 1900 to 2002. The 3 Month Bond Yield is closely correlated with Federal Funds Rate. I used the 3 Month Bond Yield to fill in missing Federal Funds Rate data from 1900-1951.

2. Amadeo, K. (2022, July 27). US Inflation Rate by Year: 1929–2023. The Balance. Retrieved August 18, 2022, from (Https://www.thebalance.com/u-s-inflation-rate-history-by-year-and-forecast-3306093) This project utilized The Balance published report "US Inflation Rate by Year From 1929 to 2023: How Bad Is Inflation? Past, Present, Future." BY KIMBERLY AMADEO, Updated July 27, 2022 REVIEWED BY ROBERT C. KELLY.

3. Irizarry, R. A. (2022, July 7). Introduction to Data Science. HARVARD Data Science. Retrieved August 8, 2022, from (Https://rafalab.github.io/dsbook/) This project utilized "Introduction to Data Science Data Analysis and Prediction Algorithms with R" by our course instructor Rafael A. Irizarry published 2022-07-07 (Chapters 1 through 34).

4. Wheelock, D. C. (2021, September 13). Overview: The History of the Federal Reserve. Federal Reserve History. Retrieved August 8, 2022, from (https://www.federalreservehistory.org/essays/federal-reserve-history) This project utilized the Overview: The History of the Federal Reserve. Published by Federal Reserve Bank of St. Louis in 2021.

5. Julian G.F. (2022, May 10). U.S Inflation - Analysis in R. Kaggle. Retrieved August 8, 2022, from (https://www.kaggle.com/code/fit4kz/u-s-inflation-analysis-in-r) This project utilized the U.S Consumer Price Index (CPI) . This dataset provides average monthly data of CPI for all US cities.

6. Standard and Poor's (S&P) 500 Index Data including Dividend, Earnings and P/E Ratio. (n.d.). DataHub. Retrieved August 8, 2022, from (https://datahub.io/core/s-and-p-500) The data provided is a version of the Economist Robert Shiller data. S&P 500 index data including level, dividend, earnings and P/E ratio on a monthly basis since 1870.

7. Bloomberg. (2022, August 15). Inside the Founding of the Federal Reserve [Video]. YouTube. (https://www.youtube.com/watch?v=0hzdglWpxVM&t=314s) Author and journalist Roger Lowenstein describes the economic crises that led to the founding of the US Federal Reserve in 1913.

8. U.S. Bureau of Labor Statistics. (2022). CPI Home : U.S. Bureau of Labor Statistics. (https://www.bls.gov/cpi/)

9. Standard and Poor's 500 (S&P 500) - Explained. (n.d.). The Business Professor, LLC. Retrieved August 22, 2022, from (https://thebusinessprofessor.com/en_US/investments-trading-financial-markets/standard-and-poors-500-sp-500-definition)

10. Introduction to ARIMA models. (2019). Duke.edu. (https://people.duke.edu/~rnau/411arim.htm)

11. Long, J. (2019, September 26). 14 Time Series Analysis | R Cookbook, 2nd Edition. Retrieved September 5, 2022, from (https://rc2e.com/timeseriesanalysis)

12. Srivastav, A. K. (2022, September 13). Pearson correlation coefficient. WallStreetMojo. Retrieved September 18, 2022, from https://www.wallstreetmojo.com/pearson-correlation-coefficient/

Overview:

Goal One: To examine the data to idenitfy any correlation using Pearson's Correlation Coefficient (r).

Goal Two: Create a forecasting machine learning model using past data from 1929-2017 to predict inflation and the appropriate federal funds rate.

Since 1929, the U.S. has combated inflation. An inflation rate of 2% is believed to be an excellent environment for businesses and consumers. During deflation, corporations and local businesses lose pricing power. Businesses have to shed employees, future investments, and goods to maintain a profit which causes an economic slowdown during deflationary periods. During rising inflation above 2%, business profits rise temporally, but consumer pricing power is eroded over time, and it can lead to hyperinflation/economic crisis/economic slowdown.

To prevent reoccurring economic collapses, deflation, galloping inflation and to fix the lack of synergy with the other 12 regional banks, The U.S. founded the Federal Reserve (the central bank) on December 23, 1913. In this project, I will explore if correlations exist within the Monthly U.S. Consumer Price Index (CPI) average for all U.S. cities, Inflation Rate Year over Year (YoY), geopolitical events, economic events, GDP growth, Federal Funds Rate, and S&P 500 price annualized from 1929 to 2017. I will also examine if one of the Federal Reserve most powerful tools, the Federal Funds Rate, is correlated with several factors listed above. I will create a forecasting algorithm using back dated information to predict inflation and appropriate federal fund rate to combat inflation.

To examine if the United States' geopolitical, domestic, and economic events are correlated with Inflation Rate YoY. I will also examine how the Federal Reserve Fund Rate affects the following: Monthly U.S. Consumer Price Index (CPI) average for all U.S. cities, Inflation Rate YoY, Geopolitical events, Economic Events, GDP Growth, and S&P 500 annualized prices utilizing the Pearson's Correlation Coefficient (r). I will also create a forecasting machine learning model using back dated information to predict inflation and appropriate federal funds rate

Data Wrangling: Clean! Clean! Clean!

Goal: Clean the data. Display the Federal Reserve Fund Rate, the Monthly U.S. Consumer Price Index (CPI) average for all U.S. cities, Inflation Year over Year (YoY), Geopolitical events, Economic Events, G.D.P. Growth, and S&P 500 price annualized from 1929-2017.

The Standard & Poor's earliest origins can be linked to the stock market in 1923. The Standard & Poor's index at the time contained 233 companies. Today, it has 500 companies within its index. It is widely tracked by economists, politicians, investors, and speculators. It is often considered an early indicator of a possible economic expansion or slowdown.

Review S&P 500 Data

```
summary(SP500_Data)
head(SP500_Data, 20)
any(is.na(SP500_Data))
sum(is.na(SP500_Data))
```

Removing all years and data before 1929 and after 2017

```
SP500_Data
New_SP500_Data <- SP500_Data[-c(1:192, 1261:1264),]
New_SP500_Data

SP_rows <-nrow(New_SP500_Data)
SP_rows
```

Annualized the data with an Annual S&P Average Return

```
SP_Mon <- SP_rows/12
SP_Mon
Whole_SP_Mon<-SP_rows%/%12
Whole_SP_Mon

series<-New_SP500_Data$SP500
mon = 12
new= NULL
for (i in 1: Whole_SP_Mon) {
  AnnualData<-series[((i-1)*mon+1):(i*mon)]
  AnnualAverage<-mean(AnnualData)
  new=rbind(new,AnnualAverage)
}
AverageSP500<-new
AverageSP500
```

Convert to years and create a table. Double check the format and convert Annual Closing Price to dollars and create a table.

S&P 500 Annualized Closing Price HardvardX Capstone Project 2022	
Annual Closing Price	Calendar_Year
$26.02	1929
$21.03	1930
$13.66	1931
$6.93	1932
$8.96	1933
$9.84	1934
$10.60	1935
$15.47	1936
$15.41	1937
$11.49	1938

S&P 500 Data is from 1929-2017

Portions of this data is from the Reference Section.

Per the U.S. Bureau of Labor Statistics ("U.S. Bureau of Labor Statistics", 2002), The Consumer Price Index (CPI) is the most widely used measure of inflation and is an indicator of the effectiveness of government policy. CPI is calculated by recording the group of goods, services, and housing that urban consumers purchase and the price average change on a monthly basis.

Load and clean CPI data

```
summary(CPI_Data)
head(CPI_Data, 20)
any(is.na(CPI_Data))
sum(is.na(CPI_Data))
```

Removing all years and data before 1929 and after 2017

```
CPI_Data
New_CPI_Data <- CPI_Data[-c(1:192, 1261:1303), ]

CPI_rows <-nrow(New_CPI_Data)
CPI_rows
```

Annualized the CPI data

```
CPI_Mon <- CPI_rows/12
CPI_Mon
Whole_CPI_Mon<-CPI_rows%/%12
Whole_CPI_Mon

series<-New_CPI_Data$CPI
mon = 12
new= NULL
for (i in 1: Whole_CPI_Mon) {
  AnnualCPIData<-series[((i-1)*mon+1):(i*mon)]
  AnnualCPIAverage<-mean(AnnualCPIData)
  new=rbind(new,AnnualCPIAverage)
}
AverageCPI<-new
AverageCPI
```

Change it to percentage. Make the years manually and create a table

```
New_Avg_CPI <- scales::percent(AverageCPI/100)
New_Avg_CPI
```

```
CalendarYear<- rep(1928+1:length(New_Avg_CPI))
CalendarYear
```

```
Final_CPI <- tibble(New_Avg_CPI)
Final_CPI$new_col <- CalendarYear
colnames(Final_CPI)<- c("Annual_CPI_Average", "Calendar_Year")
```

```
All_CPI <- tibble(Final_CPI)
acp <- head(All_CPI,10)
```

Average U.S. CPI Accumulated Data Annualized	
HardvardX Capstone Project 2022	
Annual_CPI_Average	Calendar_Year
17.1583%	1929
16.7000%	1930
15.2083%	1931
13.6417%	1932
12.9333%	1933
13.3833%	1934
13.7250%	1935
13.8667%	1936
14.3833%	1937
14.0917%	1938
CPI Data is based on data from 1929-2017	
Portions of this data is from the Reference Section	

Since 1929, The United States has experienced different economic and geopolitical events. The Federal Reserve monitors said events and create policies to accommodate the economy to prevent another Great Depression scenario. Federal Reserve utilizes its "set of tools" to help promote a healthy business cycle based on their mandates.

A Business Cycle is the beginning of an expansion(post-recession / post-economic slowdown) period and the beginning of a contraction period (recession/economic slowdown).

The Inflation Rate YoY is the rate of change of inflation yearly. The Inflation rate YoY differs from CPI Annualized data. CPI Annualized data shows how the value of products in 1929 appreciates every year until 2017 based on average inflation accumulated each year. For instance, a gallon of milk in Hawaii on the island of Oahu cost 26 cents in 1929 now a gallon of milk on the island of Oahu cost $5.50. Thats a whooping 2115% of accumulated inflation exceeding the 2017 percentage by 10x. The Inflation Rate YoY shows the change in annual inflation in each specific year vice compounding year after year. The table below will show the changes for this metric.

GDP is the total of all goods produced and sold by a nation over a specific period. This is an indicator of economic growth, stagnation, or slowing down. Let us take a look at U.S. Economic data, Geopolitical Events, and Federal Reserve Data.

Clean U.S. Economic and Event data

```r
summary(Fed_Data)
head(Fed_Data, 20)
any(is.na(Fed_Data))
sum(is.na(Fed_Data))
```

```r
All_Fed <- tibble(Fed_Data)
colnames(All_Fed)<- c("Year", "Inflation Rate YoY", "Fed Funds Rate",
                      "Business Cycle", "GDP Growth",
                      "Events Affecting Inflation" )
afd<- head(All_Fed,10)
```

U.S. Economic, Geopolitical Events and Federal Reserve Data HardvardX Capstone Project 2022					
Year	Inflation Rate YoY	Fed Funds Rate	Business Cycle	GDP Growth	Events Affecting Inflation
1929	0.60%	4.42%	August peak	6.52%	Market crash
1930	-6.40%	2.23%	Contraction	-8.50%	Smoot-Hawley
1931	-9.30%	1.40%	Contraction	-6.40%	Dust Bowl
1932	-10.30%	0.88%	Contraction	-12.90%	Hoover tax hikes
1933	0.80%	0.52%	Contraction ended in March	-1.20%	FDR's New Deal
1934	1.50%	0.26%	Expansion	10.80%	U.S. debt rose
1935	3.00%	0.14%	Expansion	8.90%	Social Security
1936	1.40%	0.14%	Expansion	12.90%	FDR tax hikes
1937	2.90%	0.45%	Expansion peaked in May	5.10%	Depression resumes
1938	-2.80%	0.05%	Contraction ended in June	-3.30%	Depression ended
Federal Funds Rate is based on data from The 3 Month Bond Yield from 1929-1954					
Portions of this data within this table is from Reference Section					

Now that we have a better look at the data, it is hard to discern which economic event, geopolitical event or federal reserve action data correlates with one another. Let us visualize the data to see if we can find an inverse, positive, or no correlation.

Data Visualization: Plotting the Cleaned Data

Goal: Create visualizations with individual data and combined data. Observe any inverse, none, or positive correlations.

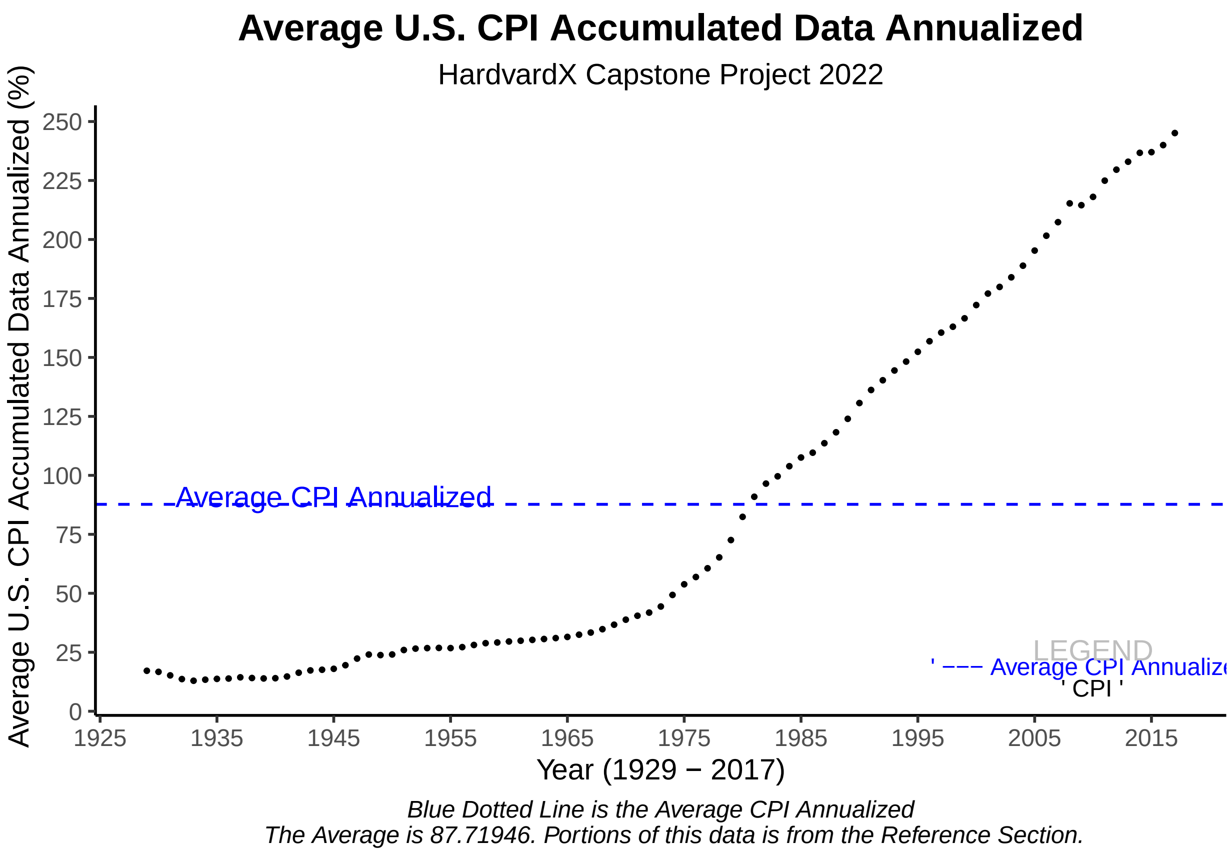
Create a chart of CPI annualized data from 1929 to 2017

```
Final_CPI$Annual_CPI_Average = as.numeric(gsub("[\\%]", "",Final_CPI$Annual_CPI_Average))
Final_CPI
```

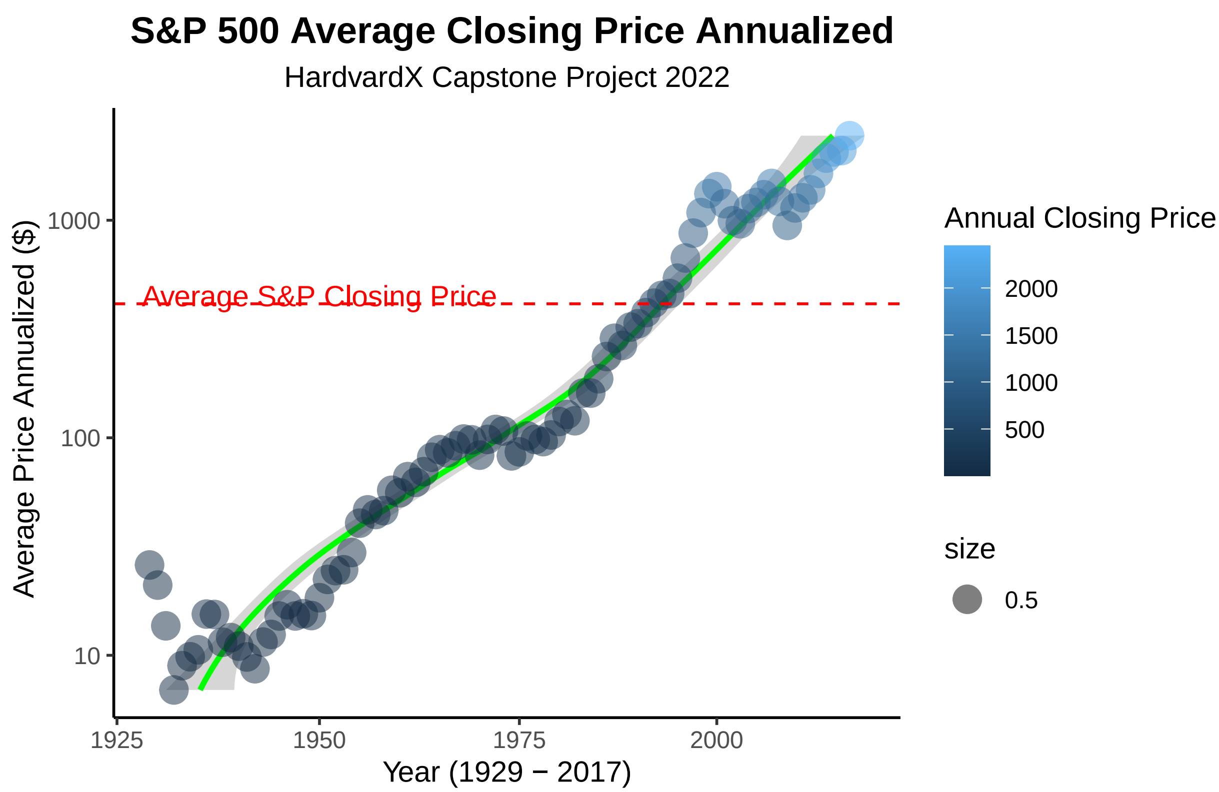
Average Inflation

```
Average_inflation <- mean(Final_CPI$Annual_CPI_Average)
Average_inflation
```

```
## [1] 87.71946
```

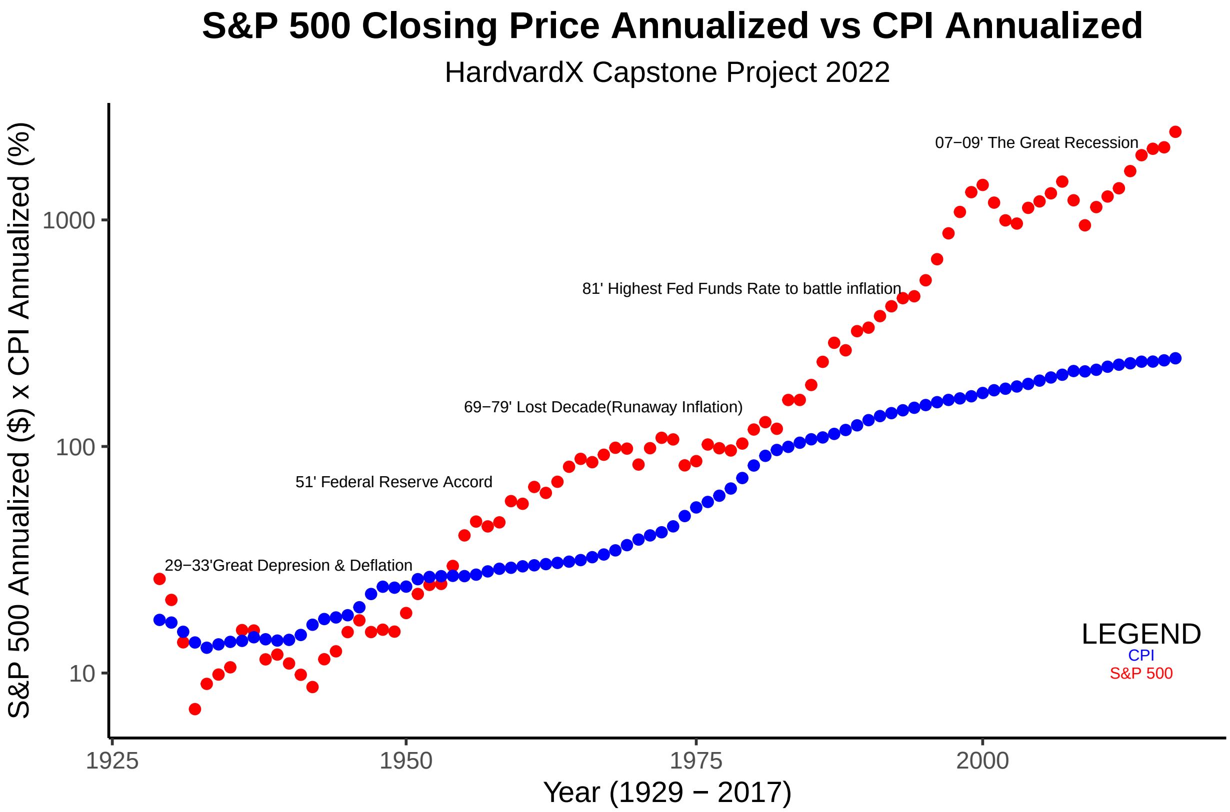


Create a chart of S&P 500 from 1929 to 2017



Red Dotted Line is the Average S&P 500 Closing Price.
The Average is $412.97. Chart Scale is log10. Portions of this data is from the Reference Section.

Create a chart to compare average CPI annualized and S&P 500



Portions of this data is from the Reference Section. Chart scale is log10.

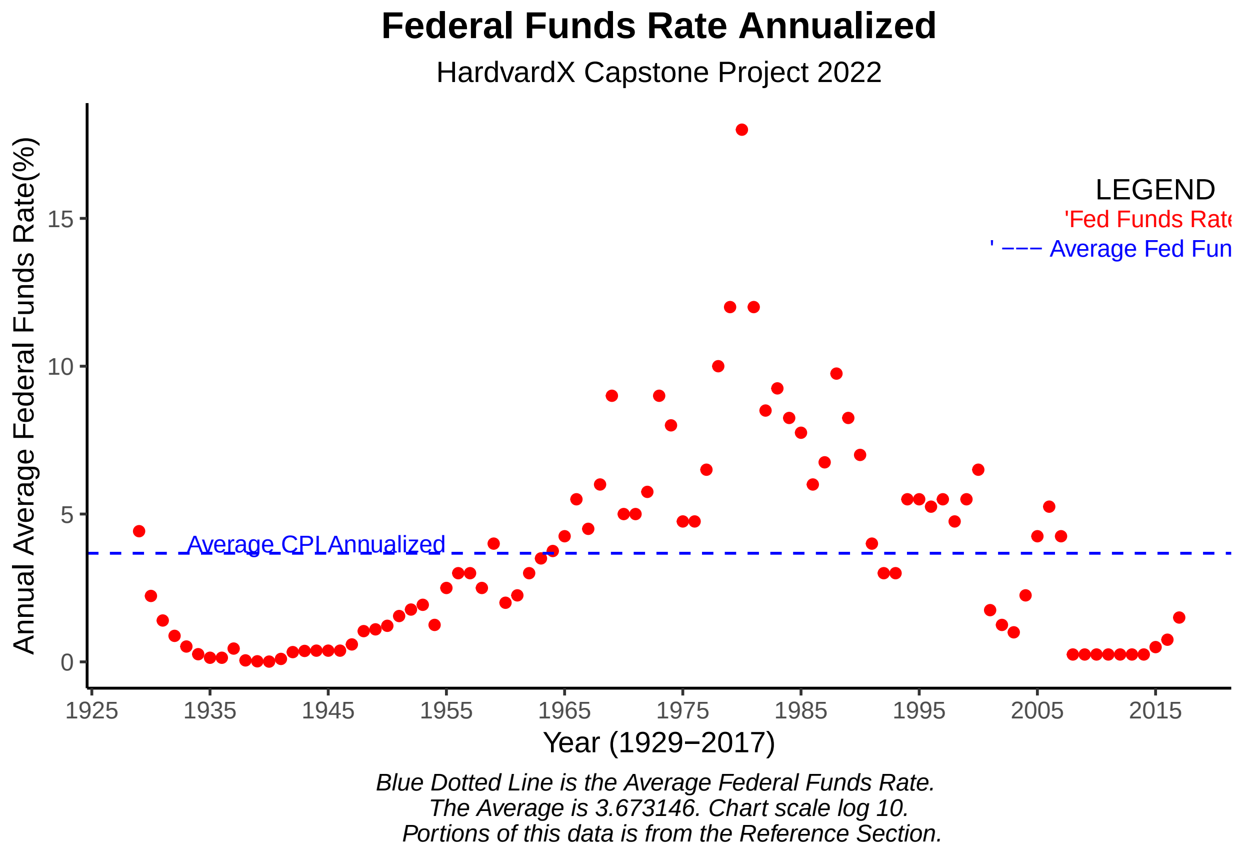
The chart above shows how the Consumer Price Index has grown exponentially over time with the S&P 500. The geopolitical and economic events reflect the S&P 500 negative/positive reactions in some cases and nil in others. As CPI has grew gradually from the late 1970s, the S&P 500 has continued to grow faster in worth over time.

Federal Reserve's Fed Funds Rate is a tool utilized by the Federal Reserve to tackle inflation, economic slowdowns or promote growth in the economy. Chart of Federal Funds Rate from 1929 - 2017.

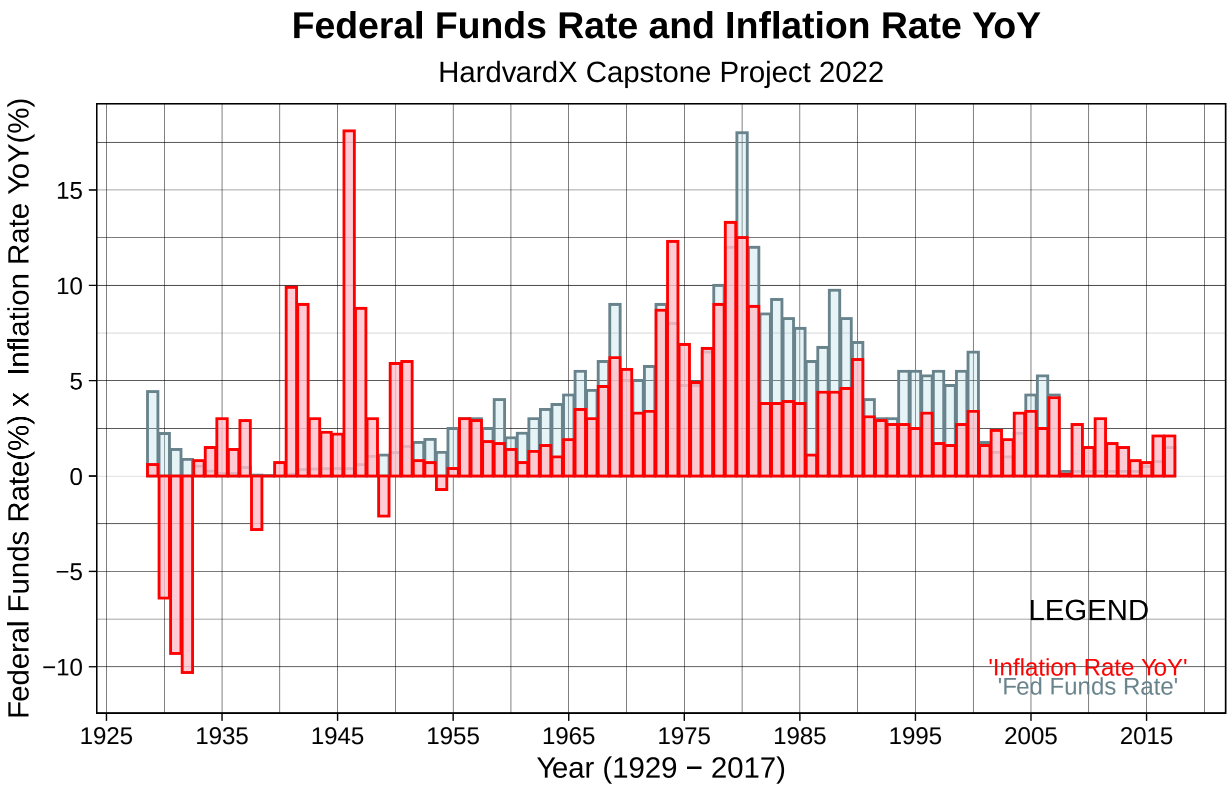
```
All_Fed$`Fed Funds Rate` = as.numeric(gsub("[\\%, ]","", All_Fed$`Fed Funds Rate`))
All_Fed$`Inflation Rate YoY` = as.numeric(gsub("[\\%, ]","", All_Fed$`Inflation Rate YoY`))
All_Fed$`GDP Growth` = as.numeric(gsub("[\\%, ]","", All_Fed$`GDP Growth`))

All_Fed
```

```
Average_Fed_Rate<- mean(All_Fed$`Fed Funds Rate`)
Average_Fed_Rate
```

```
## [1] 3.673146
```



Inflation at high levels is one of the most significant issues that can cause an economic slowdown. Chart of Federal Reserve's Fed Funds Rate and Inflation Rate YoY.



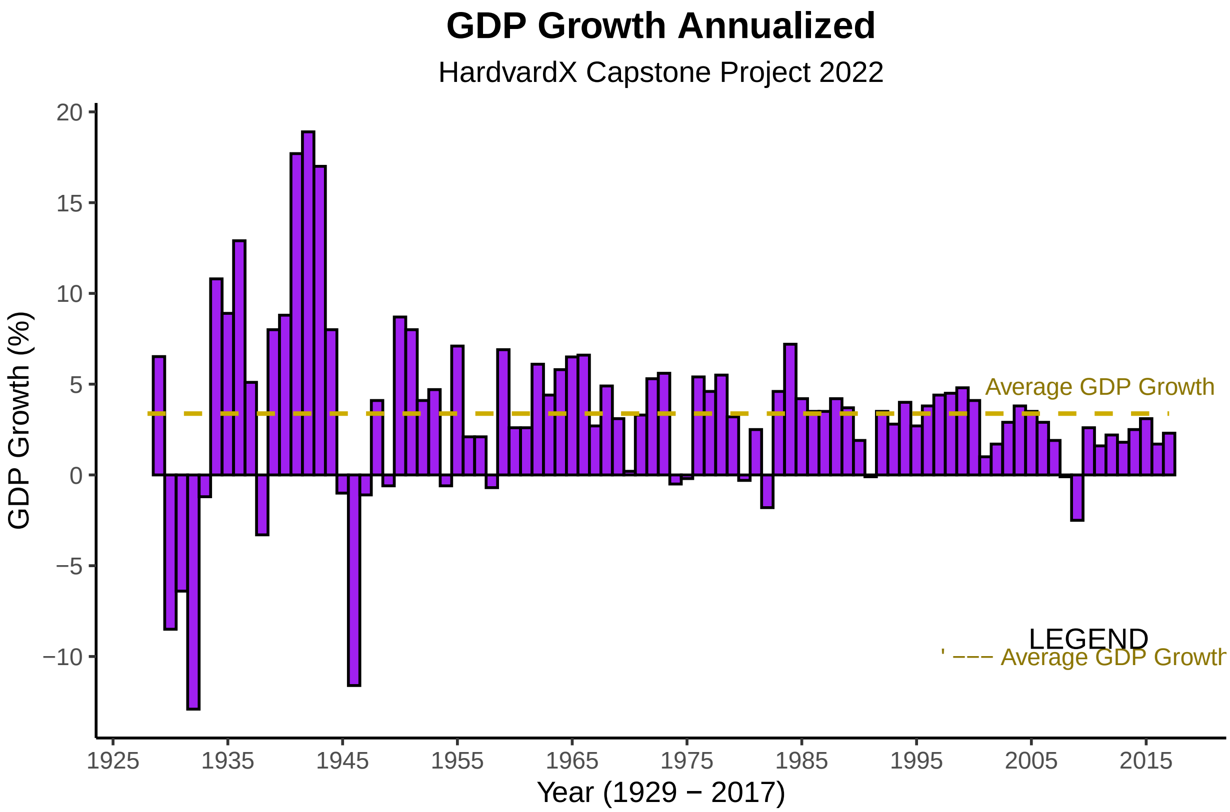
Federal Funds Rate (BLUE) and Inflation Rate YoY (RED).
Portions of this data is from the Reference Section.

Looking at the chart above, we can assess that the Federal Funds Rate and Inflation Rate YoY tend to trend in the same direction annually (the data overlap). We will dig deeper into the data later for confirmation.

Negative GDP signals economic slowdown, a neutral rate indicates economic stagnation, and a positive rising GDP rate signals economic expansion. Let's take a look at GDP. Chart of GDP Growth with the Average GDP.

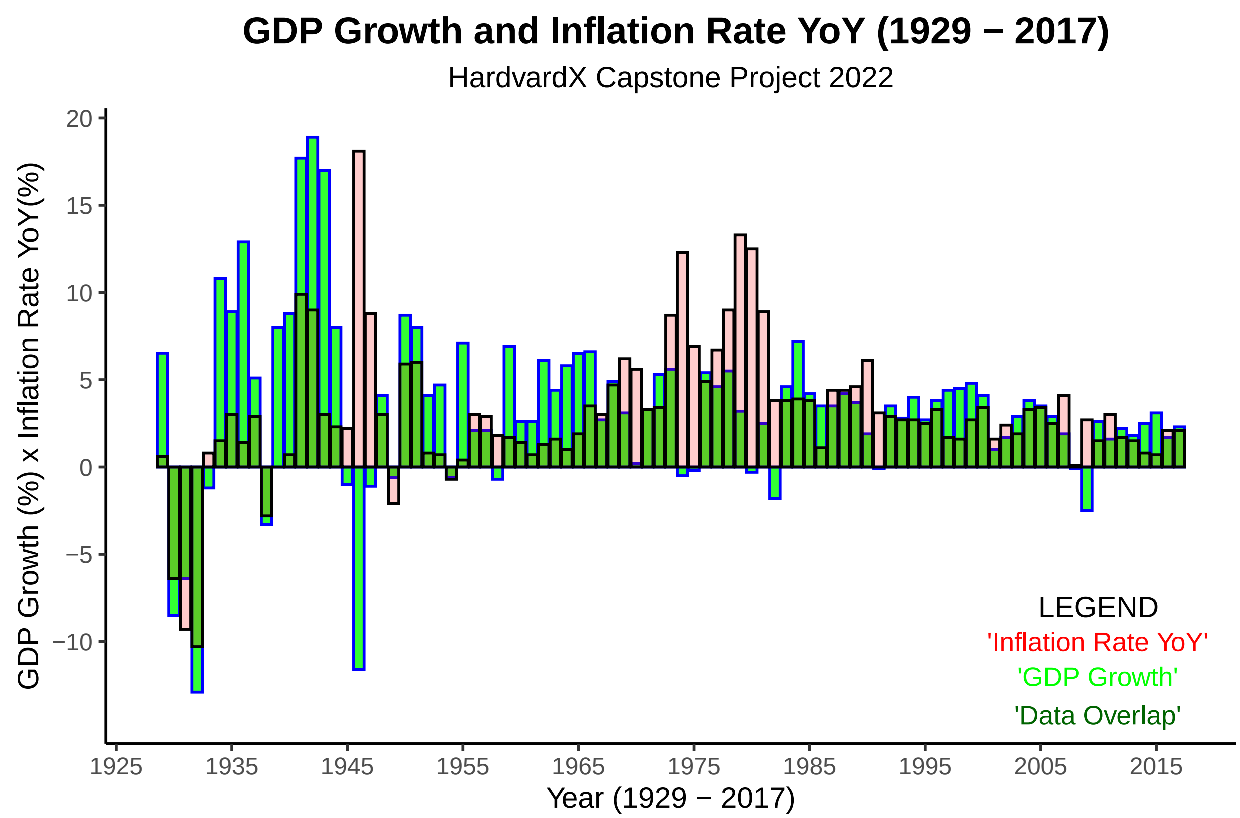
```
Average_GDP_Rate<- mean(All_Fed$`GDP Growth`)
Average_GDP_Rate
```

```
## [1] 3.38
```



Average GDP is 3.38%. Portions of this data is from the Reference Section.

Create a chart of GDP Growth and Inflation Rate YoY



Dark Green color is from GDP Growth and Inflation Rate YoY overlaping. Portions of this data is from the Reference S

After looking at the visualizations, we noticed that some data might have a positive correlation while others have an inverse or no correlation.

I also noticed that U.S. CPI and S&P 500 annualized are more exponential growth over time than the other variables. The Federal Reserve is not mandated to manage the S&P 500 and is banned from buying stocks per the Federal Reserve Act. For this purpose, we will only examine Inflation YoY, GDP Growth, and CPI Average Annualized versus the Federal Reserve's Fed Funds Rate. We will use Pearson's Correlation Coefficient in our Data Analysis - Correlation Section to accurately compute the correlations.

Data Analysis: Discovering Correlations

Goal: To observe if the Fed Funds Rate has a positive, negative or no correlation with Inflation YoY, GDP Growth and CPI Average Annualized from 1929 - 2017.

I will use Pearson's Correlation Coefficient. Pearson's Correlation Coefficient measures the linear correlation between two variables. For the Pearson's Correlation Coefficient, the value "r" represents the correlation.

If r = 0:1, this means an absolute correlation (the variables move in the same direction). If r = 0, this means no correlation between the two variables, and the value r = 0:-1 means a negative correlation (the variables move in the inverse direction). For more information on Pearson's Correlation or any correlation formula, please refer to the Reference Section.

Pearson Correlation Coefficient Formula (Srivastav, 2022): r =

$$\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

r = correlation coefficient. n = number of pairs of scores. x = values of the x-variable in a sample.
y = values of the y-variable in a sample.
$\sum$ = sum of.

Create a table with all the variables needed.

```
All_CorrData <- tibble(All_Fed$Year,
                       All_Fed$`Inflation Rate YoY`,
                       All_Fed$`Fed Funds Rate`,
                       All_Fed$`GDP Growth`,
                       Final_CPI$Annual_CPI_Average
)

colnames(All_CorrData)<- c("Year", "Inflation Rate YoY",
                           "Fed Funds Rate", "GDP Growth", "Annual CPI Average")
any(is.na(All_CorrData))
sum(is.na(All_CorrData))
head(All_CorrData, 20)
All_CorrData
acd<-head(All_CorrData,10)
```

Variables for Pearson's (r) Data Examination				
HardvardX Capstone Project 2022				
Year	Inflation Rate YoY	Fed Funds Rate	GDP Growth	Annual CPI Average
1929	0.6	4.42	6.52	17.1583
1930	-6.4	2.23	-8.50	16.7000
1931	-9.3	1.40	-6.40	15.2083
1932	-10.3	0.88	-12.90	13.6417
1933	0.8	0.52	-1.20	12.9333
1934	1.5	0.26	10.80	13.3833
1935	3.0	0.14	8.90	13.7250
1936	1.4	0.14	12.90	13.8667
1937	2.9	0.45	5.10	14.3833
1938	-2.8	0.05	-3.30	14.0917

All Data is from 1929-2017
Portions of this data is from the Reference Section.

Testing the Pearson's Correlation Coefficient (r) Formula for Federal Funds Rate vs U.S. Inflation Rate YoY

```r
cor.test(All_CorrData$'Fed Funds Rate', All_CorrData$'Inflation Rate YoY')
```

```
##
##  Pearson's product-moment correlation
##
## data:  All_CorrData$"Fed Funds Rate" and All_CorrData$"Inflation Rate YoY"
## t = 4.9002, df = 87, p-value = 4.392e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2843698 0.6138815
## sample estimates:
##       cor
## 0.4650834
```

Since the correlation is .4650834, we can see we have a positive but moderate correlation with Federal Funds Rate and Inflation Rate YoY

To compute the amount of variation between each variable we will utilize R^2 and convert it to a percentage

```r
r2FI<- percent(sqrt(.4650834))
r2FI
```

```
## [1] "68%"
```

With a R^2 of 68% this means that 32% of variance is explained by unknown factors.

Testing the Pearson Correlation Coefficient (r) Formula for Federal Funds Rate vs GDP Growth

```
cor.test(All_CorrData$'Fed Funds Rate', All_CorrData$'GDP Growth')
```

```
##
##  Pearson's product-moment correlation
##
## data:  All_CorrData$"Fed Funds Rate" and All_CorrData$"GDP Growth"
## t = -0.29098, df = 87, p-value = 0.7718
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2378934  0.1782326
## sample estimates:
##        cor
## -0.0311815
```

Since the correlation is -0.0311815, we can see we have a negative to no correlation with Federal Funds Rate and GDP Growth

To compute the amount of variation between each variable we will utilize R^2 and convert it to a percentage (remove the negative number for r will not compute)

```
r2FG<- percent(sqrt(0.0311815))
r2FG
```

```
## [1] "18%"
```

With a R^2 of 18% this means that 82% of variance is explained by unknown factors.

Testing the Pearson Correlation Coefficient (r) Formula for Federal Funds Rate vs Annual CPI Average.

```
cor.test(All_CorrData$'Fed Funds Rate', All_CorrData$'Annual CPI Average')
```

```
##
##  Pearson's product-moment correlation
##
## data:  All_CorrData$"Fed Funds Rate" and All_CorrData$"Annual CPI Average"
## t = 0.23721, df = 87, p-value = 0.8131
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1838068  0.2324491
## sample estimates:
##        cor
## 0.02542312
```

Since the correlation is 0.02542312, we can see we have a positive but little correlation with Federal Funds Rate and Annual CPI Average

To compute the amount of variation between each variable we will utilize R^2 and convert it to a percentage

```
r2FC<- percent(sqrt(0.02542312))
r2FC
```

```
## [1] "16%"
```

With a R^2 of 16% this means that 84% of variance is explained by unknown factors.

Summary Table of the Pearson's Correlation 1929-2017 results

Pearson's (r) Data Variable Variation (%)

HardvardX Capstone Project 2022

Fed Funds Rate vs Inflation YoY	Fed Funds Rate vs GDP Growth	Fed Funds Rate vs Annual CPI Average
68	18	16

The data is from 1929-2017

Portions of this data is from the Reference Section.

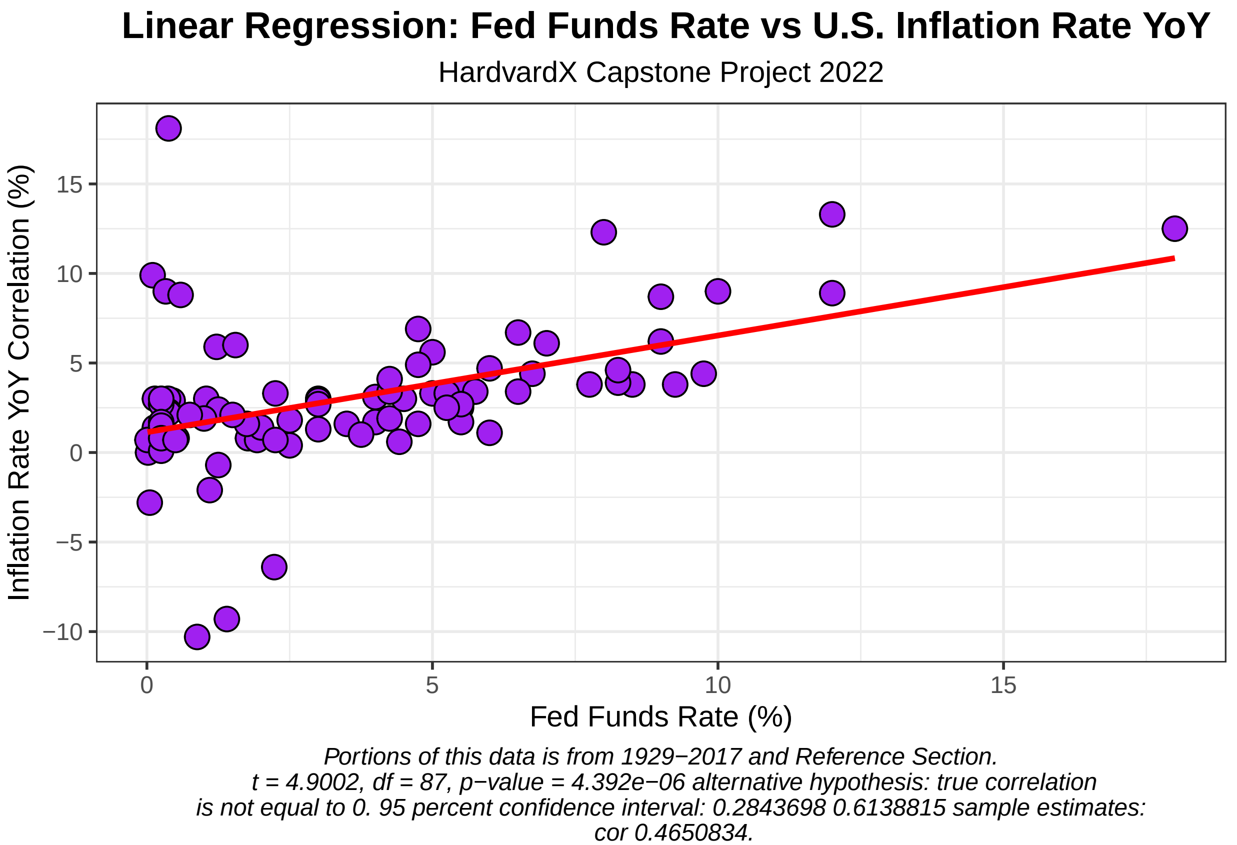
As we can see, our best correlation with the Fed Funds Rate is Inflation Rate YoY. An R^2 of 68% means that unknown factors explain 32% of the variance. The unknown factors could be outliers. Outliers affect the accuracy of a Pearson's Correlation formula. Let's create Linear Regression charts to view the correlations and see if we have any outliers.

Data Analysis: Correlation Visualization

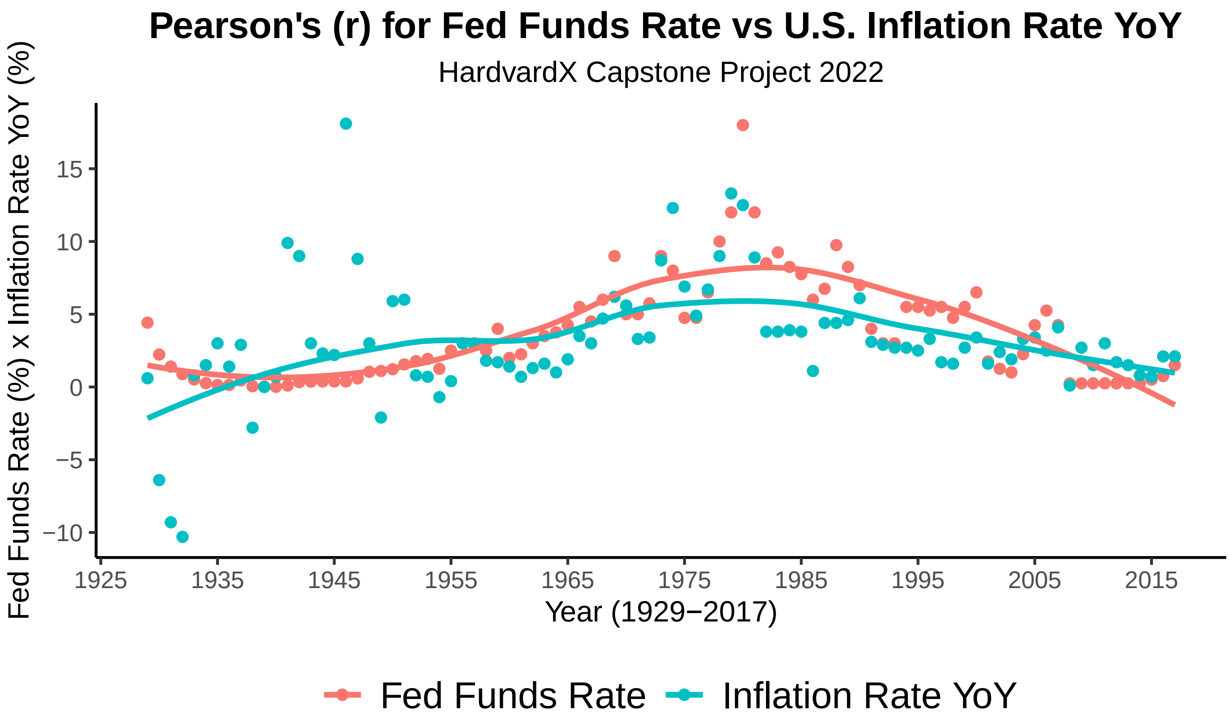
Goal: To visualize the data. Observe what factor(s) are causing the major divergences in correlation.

Scatterplot Pearson's Correlation Coefficient (r) Formula for Federal Funds Rate vs U.S. Inflation Rate YoY

```
data <- data.frame(x= All_CorrData$'Fed Funds Rate',
                   y= All_CorrData$'Inflation Rate YoY')
```



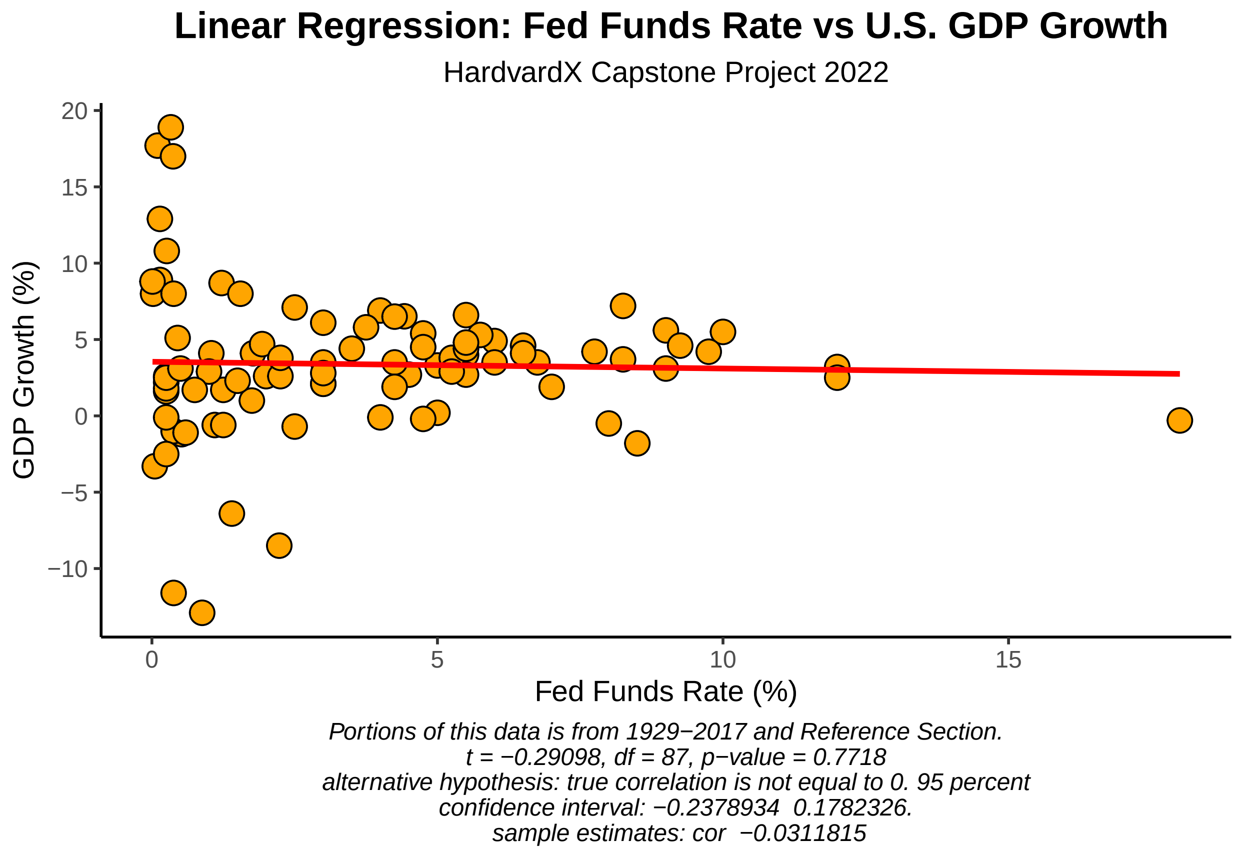
Linear Regression Chart of the Pearson's Correlation Coefficient (r) for Fed Funds Rate and Inflation Rate YoY from 1929-2017



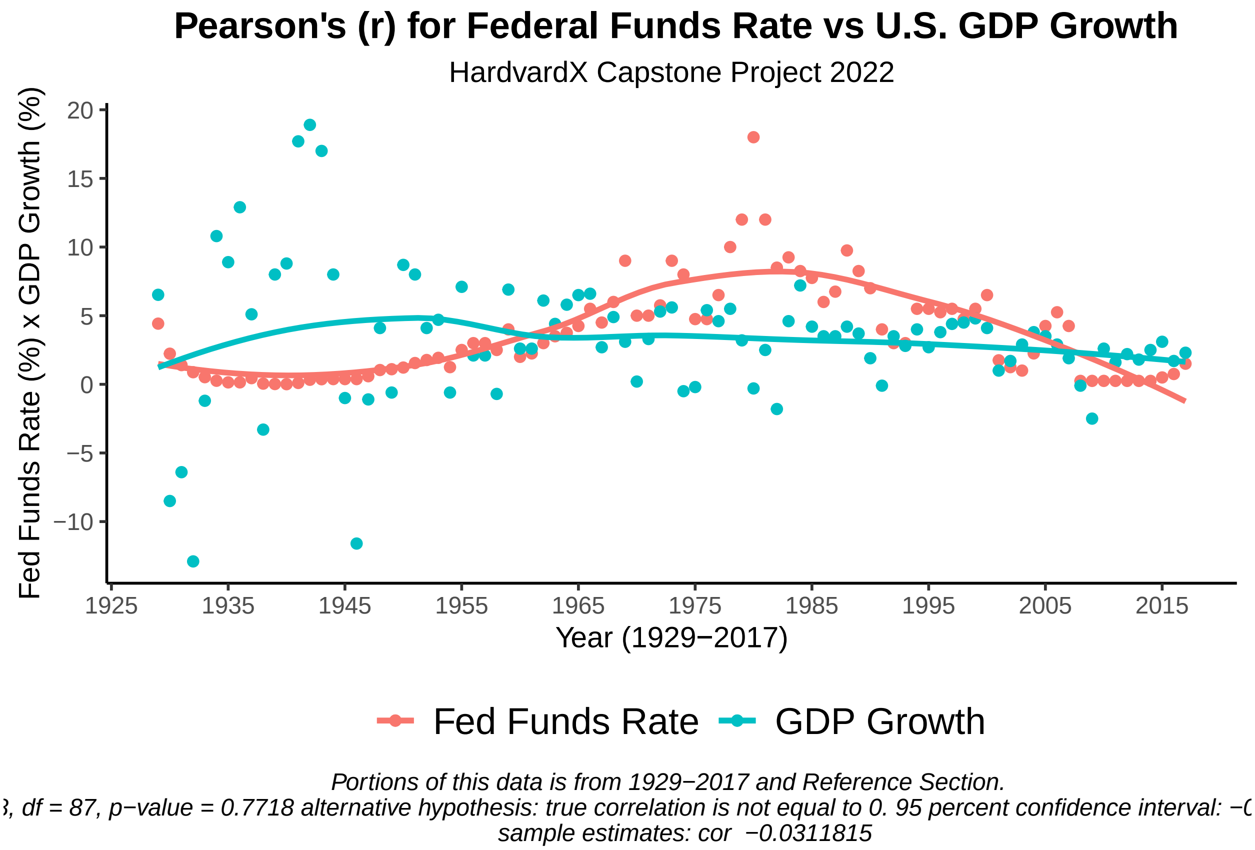
Portions of this data is from 1929−2017 and Reference Section.
t = 4.9002, df = 87, p−value = 4.392e−06 alternative hypothesis: true correlation is not equal to 0. 95 percent confidence interval: 0.2843698 0.6138815 sample estimates: cor 0.4650834.

Scatterplot Pearson's Correlation Coefficient (r) Formula for Federal Funds Rate vs U.S. GDP Growth

```
FvGDPdata <- data.frame(x= All_CorrData$'Fed Funds Rate',
                        y= All_CorrData$'GDP Growth')
```



`Linear Regression Chart of the Pearson's Correlation Coefficient (r) for Fed Funds Rate and GDP Growth f`

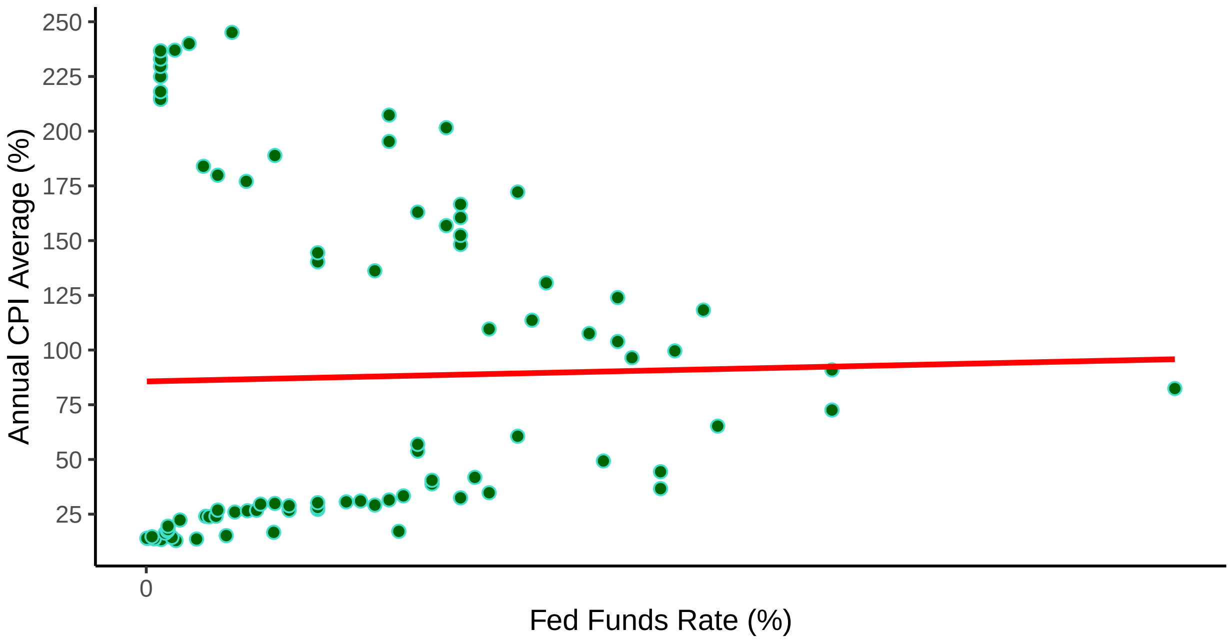


Scatterplot Pearson's Correlation Coefficient (r) Formula for Federal Funds Rate vs CPI Accumulated Average Annualized

```
FvCPIdata <- data.frame(x= All_CorrData$'Fed Funds Rate',
                        y= All_CorrData$"Annual CPI Average")
```

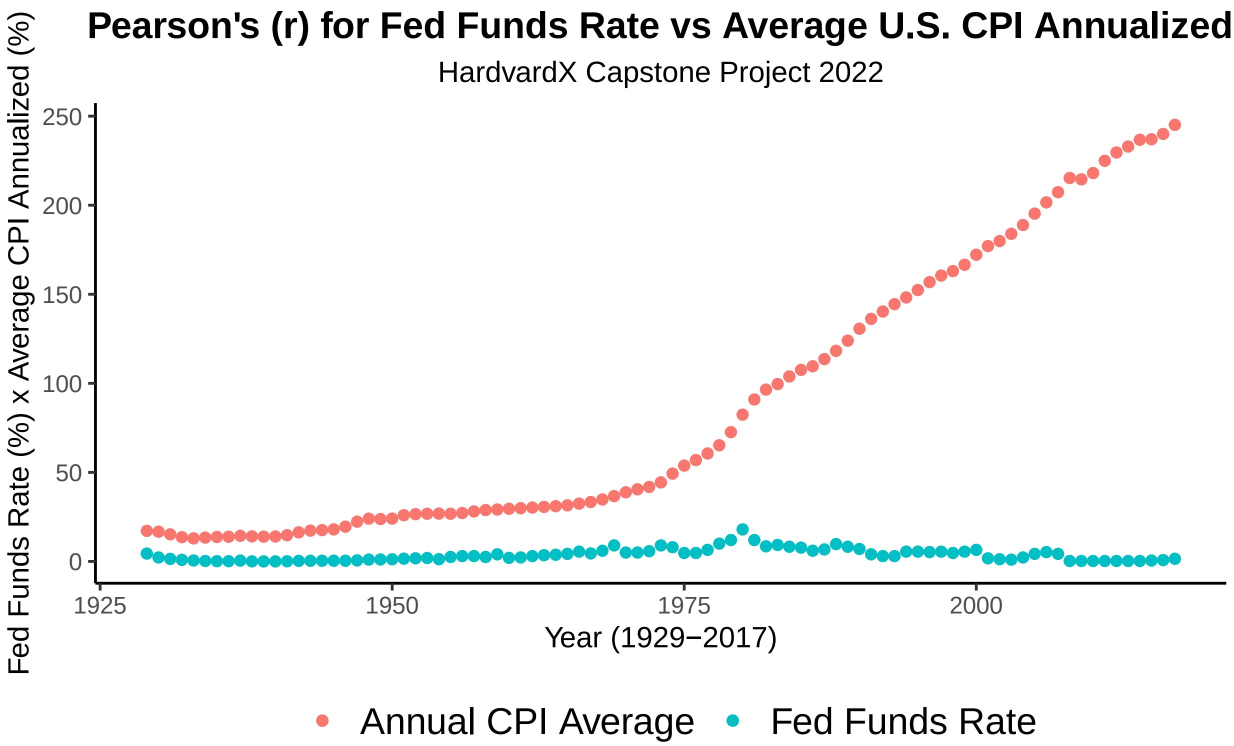
Linear Regression: Fed Funds Rate vs U.S. CPI Annualized

HardvardX Capstone Project 2022



Portions of this data is from 1929−2017 and Reference Section.
t = 0.23721, df = 87, p−value = 0.8131
alternative hypothesis: true correlation is not equal to 0. 95 percent
confidence interval: −0.1838068 0.2324491 sample estimates: cor 0.02542312

```
Linear Regression Chart of the Pearson's Correlation Coefficient (r) for Fed Funds Rate and Annual CPI
```



t = 0.23721, df = 87, p−value = 0.8131 alternative hypothesis: true correlation is not equal to 0. 95 percent confidence interval: −0.1838068 0.2324491 sample estimates: cor 0.02542312 Portions of this data is from the Reference Section.

As depicted in each chart, outliers can affect Pearson's correlation accuracy, but we also can see that all other data displayed in each chart had a distinct or indistinct correlation.

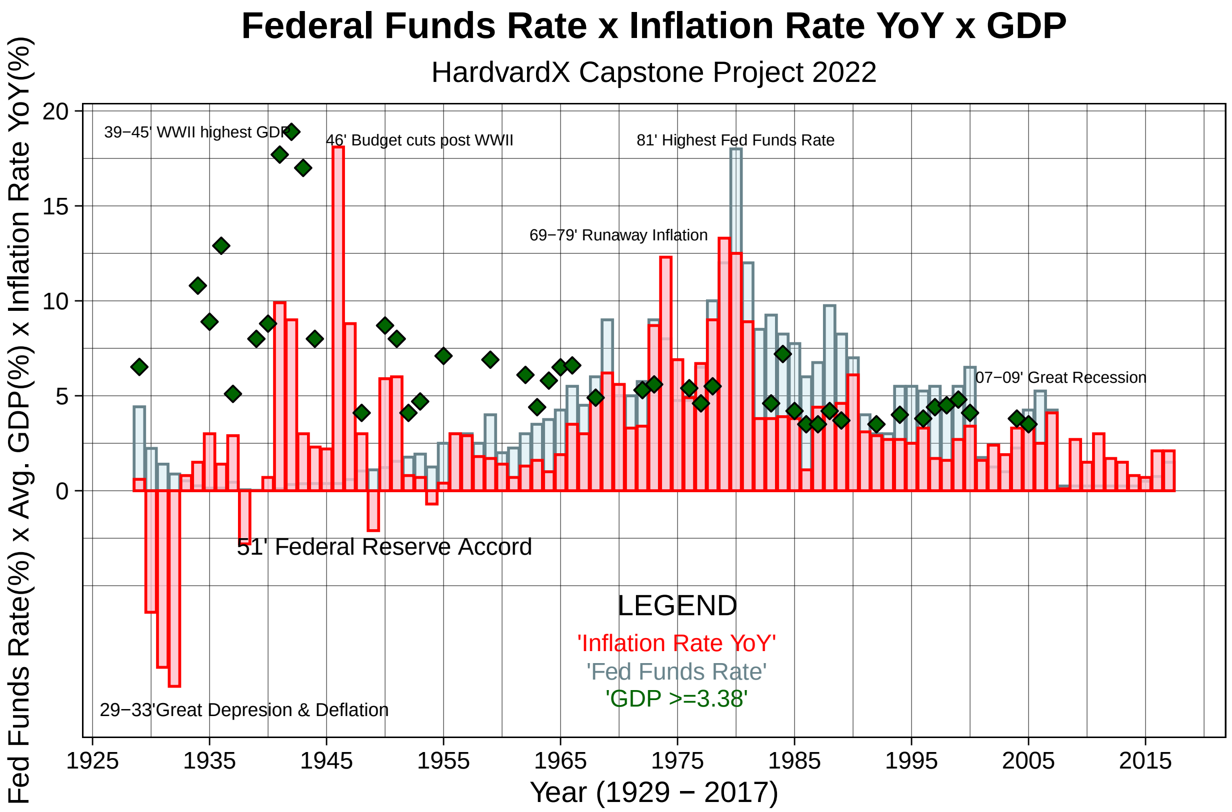
On a good note, we did have a positive but moderate correlation with Federal Funds Rate and Inflation Rate YoY. An R^2 of 68% means that unknown factors explain 32% of the variance.

Let us chart the economic and geopolitical events with Federal Funds Rate and Inflation Rate YoY to see if the "outliers" were economic/geopolitical driven.

```
Conclusion_All_Fed <- tibble(All_Fed)
Conclusion_All_Fed$new_col <- WithDollarSign_SP500$`Annual Closing Price`
colnames(Conclusion_All_Fed)<- c("The_Year", "Inflation Rate YoY", "Fed Funds Rate", "Business Cycle",

print(select_if(Conclusion_All_Fed, is.numeric))
```

In addition let us plot all GDP data points that are greater than or equal to the GDP annualized average of 3.38.

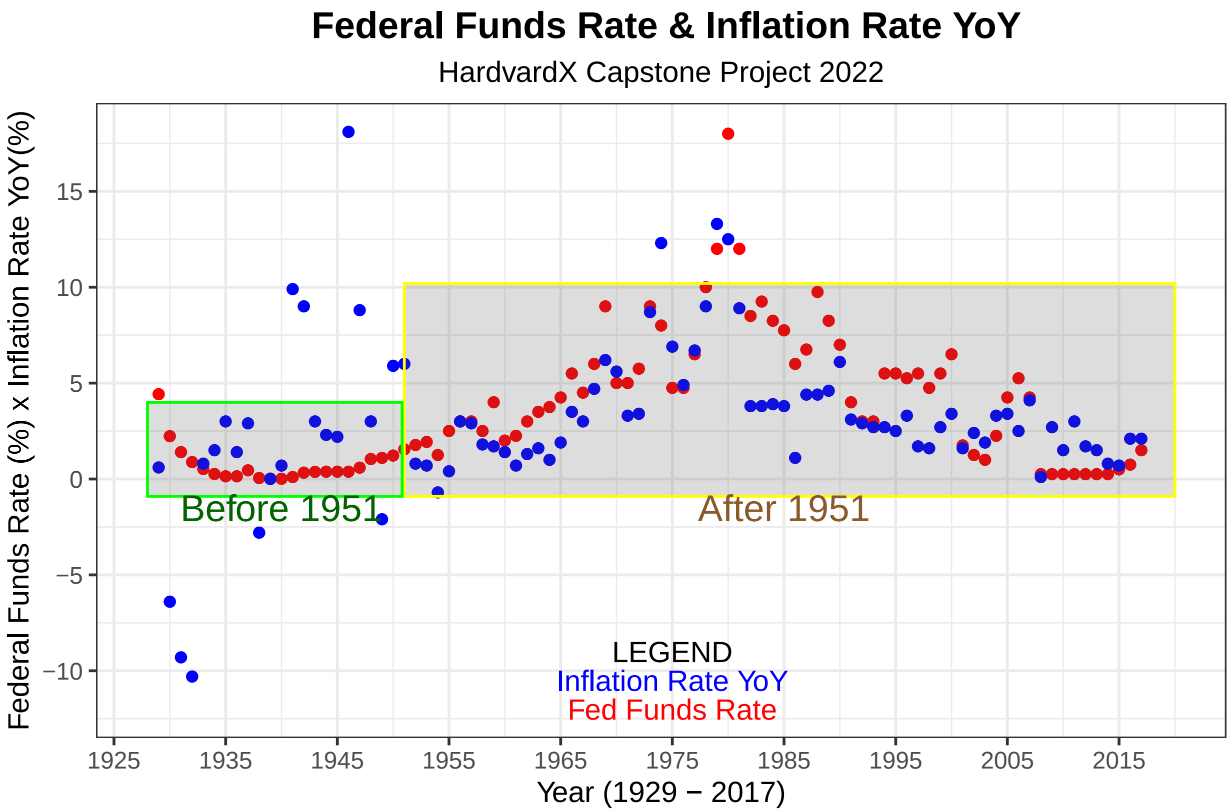
```
GREATER_THAN_Avg_GDP <- subset(Conclusion_All_Fed, `GDP Growth` >= 3.38)
GREATER_THAN_Avg_GDP
```



**Only plotted GDP greater than annualized average of 3.38%. Portions of this data is from the Reference Sectior*

As we can see, outliers are caused by economic and geopolitical factors. These factors can affect inflation and the Federal Reserve's Fund Rate.

Additionally, the majority of the outliers happened prior to 1951. Let us look at a chart highlighting outliers and correlated data for the Federal Funds Rate vs Inflation Rate YoY from 1929-2017.



Portions of this data is from the Reference Section.

As you can see, before 1951, there were 11 outliers and only six after 1951. This may be attributed to the Federal Reserve and U.S. Treasury signing the Accord in 1951. This Accord allowed the Federal Reserve to act independently, utilize its economic tools to fight inflation, and implement monetary policy.

Lets see if the correlation changes if I remove 1929-1950 data. Test the Pearson's Correlation Coefficient (r) Formula for Federal Funds Rate vs U.S. Inflation Rate YoY

```
a1951_Fed <- All_Fed[-c(1:22), ]

cor.test(a1951_Fed$'Fed Funds Rate', a1951_Fed$'Inflation Rate YoY')
```

```
##
##  Pearson's product-moment correlation
##
## data:  a1951_Fed$"Fed Funds Rate" and a1951_Fed$"Inflation Rate YoY"
## t = 9.7623, df = 65, p-value = 2.299e-14
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6515102 0.8532300
## sample estimates:
##       cor
## 0.7710506
```

Since the correlation is 0.7710506, we can see we have a positive but moderate correlation with Federal Funds Rate and Inflation Rate YoY

To compute the amount of variation between each variable we will utilize R^2 and convert it to a percentage

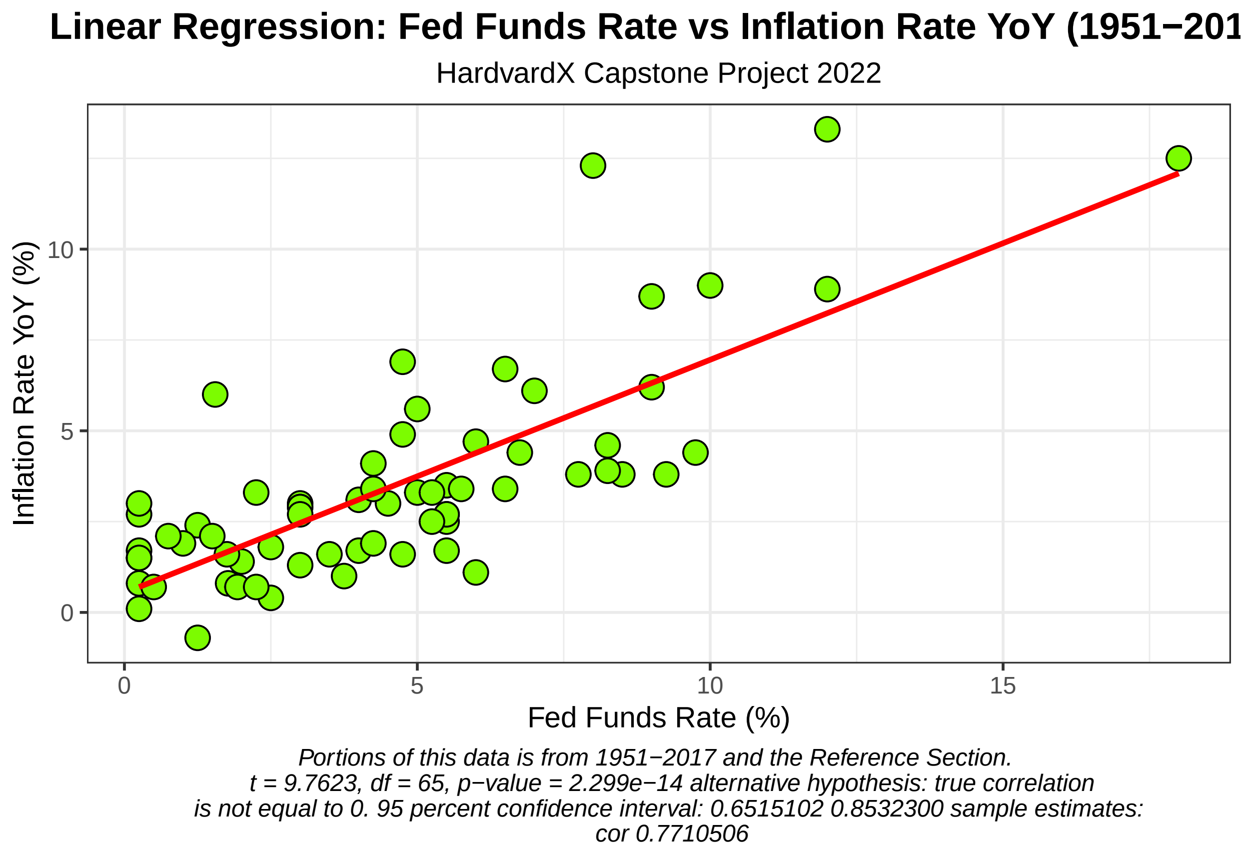
```r
New_r2FI<- percent(sqrt(0.7710506))
New_r2FI
```

```
## [1] "88%"
```

With a R^2 of 88% this means that 12% of variance is explained by unknown factors.

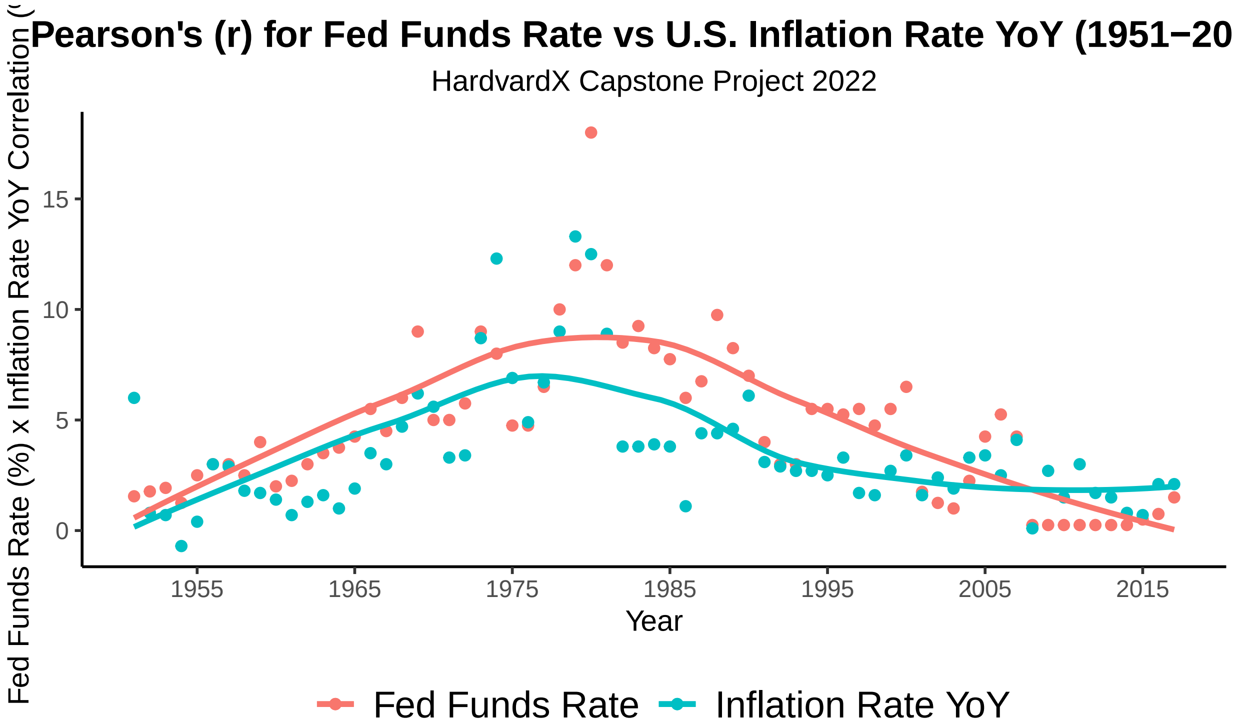
Scatterplot Pearson's Correlation Coefficient (r) Formula for Federal Funds Rate vs U.S. Inflation Rate YoY

```r
NEW_1951_data <- data.frame(x= a1951_Fed$'Fed Funds Rate',
                   y= a1951_Fed$'Inflation Rate YoY')
```



Linear Regression Chart of the Pearson's (r) for Fed Funds Rate and Inflation Rate YoY from 1951-2017

```r
NEW_1951_data
colnames(Final_CPI)<- c("Annual_CPI_Average", "Calendar_Year")
```



Portions of this data is from 1951−2017 and the Reference Section. t = 9.7623, df = 65, p−value = 2.299e−14 alternative hypothesis: true correlation is not equal to 0. 95 percent confidence interval: 0.6515102 0.8532300 sample estimates: cor 0.7710506.

Pearson's Correlation Results:

Final Pearson's (r) Variable Variation Data (%)

HardvardX Capstone Project 2022

Fed Funds Rate vs Inflation YoY after 1951	Fed Funds Rate vs Inflation YoY before 1950	Fed Funds Rate vs GDP Growth	Fed Funds Rate vs Annual CPI Average
88	68	18	16

Portions of this data is from the Reference Section.

```
After focusing on when the Federal Reserve could use the full range of its tools to combat Inflation Ra
```

Ratio Federal Funds Rate vs Inflation Rate YoY to determine Forecasting Model Tolerance.

```r
Model_Tolerance <- mean(All_Fed$`Inflation Rate YoY`)/mean(All_Fed$`Fed Funds Rate`)

Model_Tolerance
```

```
## [1] 0.8488575
```

Lets round the tolerance to the nearest whole number for simplicity. We will use this in our forecasting model.

```r
round(Model_Tolerance)
```

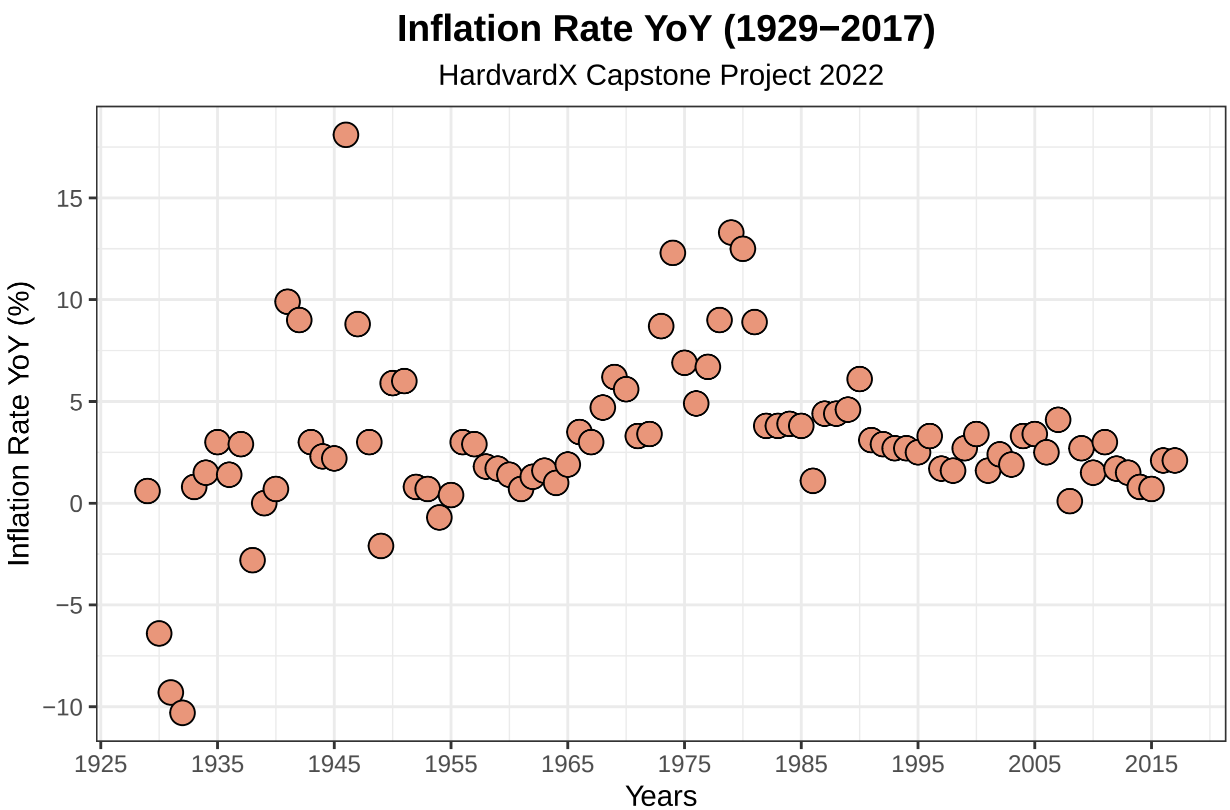
```
## [1] 1
```

Data Analysis: Machine Learning Forecasting and Backtesting Model

Goal: Create a Machine Learning Inflation and Federal Reserve Fund Rate Forecasting and Backtest Model.

Using Pearson's Correlation, we concluded that the Inflation Rate YoY has a strong positive correlation with Federal Reserve's Federal Funds Rate. Let us create a machine learning forecasting model to predict future Inflation Rate YoY and Federal Funds Rate. To make this model sustainable, we will have to backtest with primary data used above. To conduct this, we will use a time series model. Specifically, our machine learning model will use the AutoRegressive Integrated Moving Average (ARIMA). Remember that outliers can drive inconsistencies in our data model, so we want to ensure that our model is within tolerance for most of our data. For this backtesting and forecasting this model we will use a data prediction tolerance of +/- 1 point or 1000 basis points of the original data

Create a table with all the variables needed.



Portions of this data is from 1951−2017 and the Reference Section.

As you can see, the data is scattered about. To use the ARIMA model, we will need to verify the data and see if it is in a time series format.

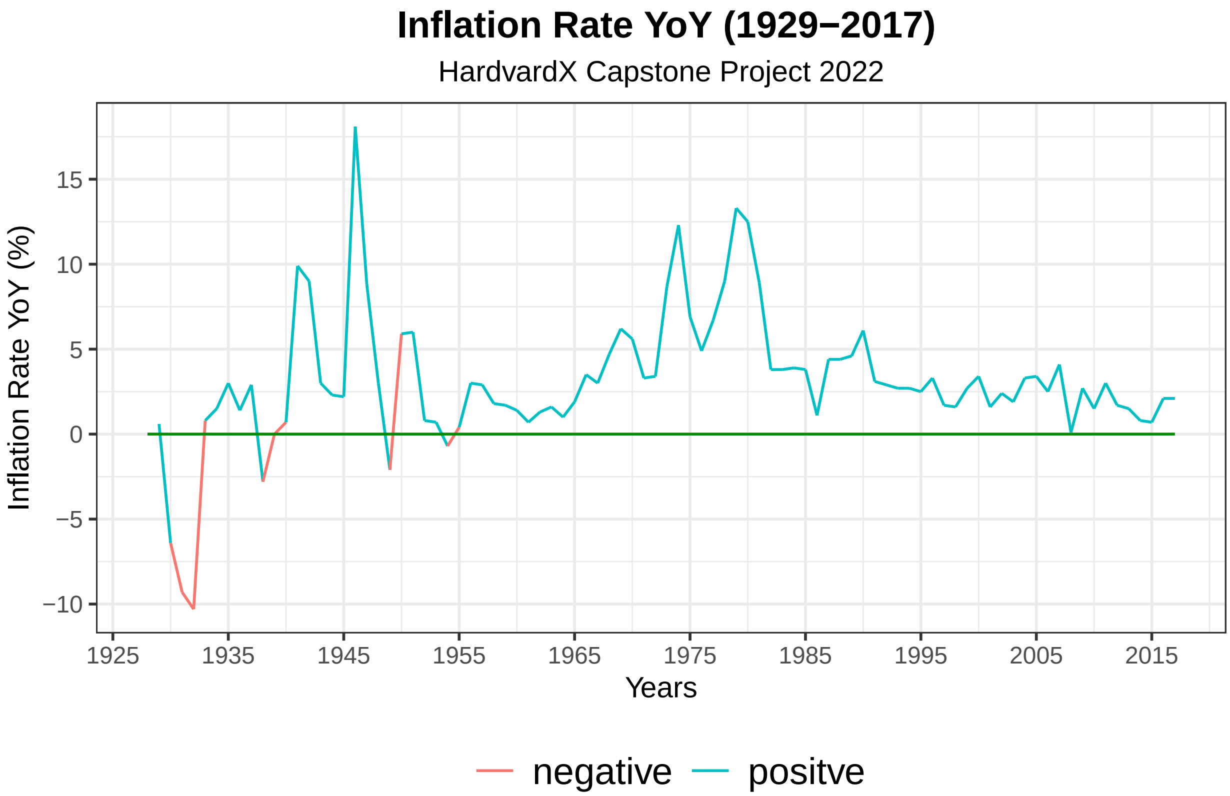
```
class(Inflation_Model)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

Its not so lets convert it

```
Inflation_Model_Time = ts(Inflation_Model$Inflation_YoY, start = min(Inflation_Model$Years), end = max(I

class(Inflation_Model_Time)
```

```
## [1] "ts"
```

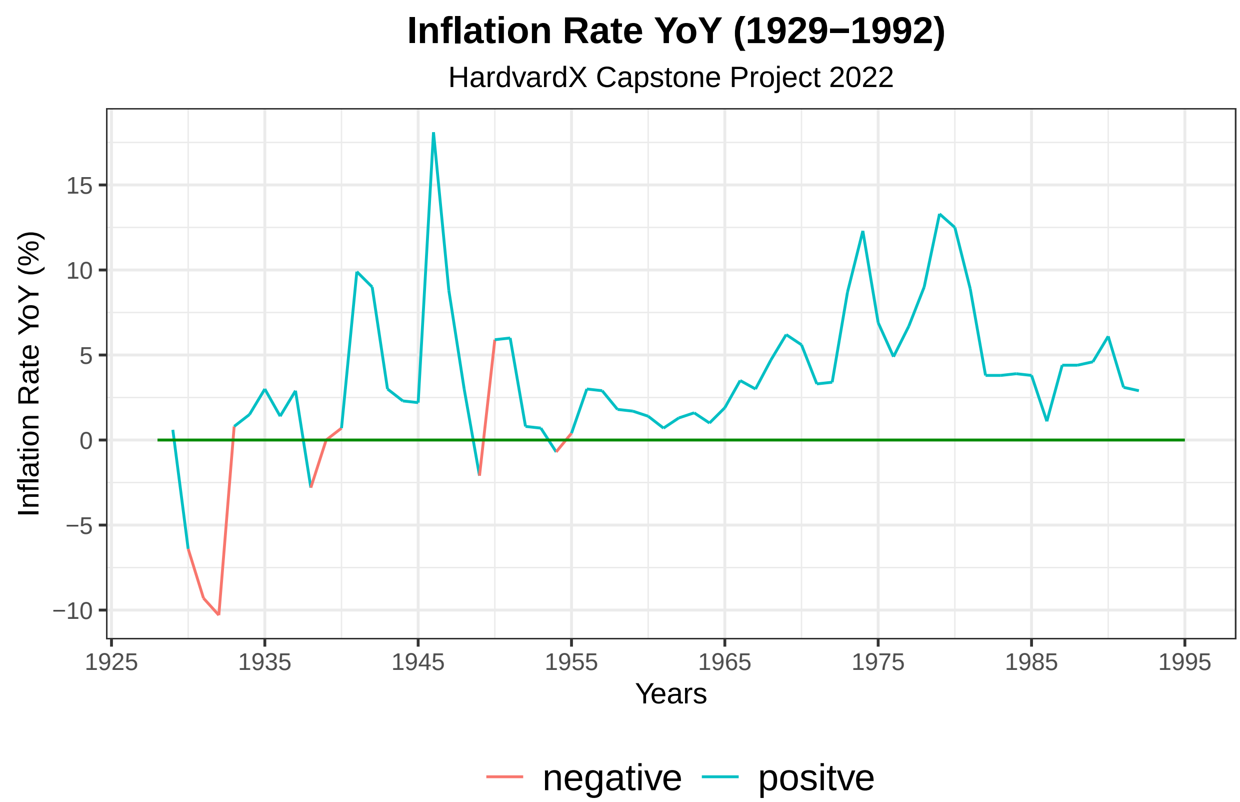


Portions of this data is from 1951−2017 and the Reference Section.

Now that we have the data properly formatted, we have to verify if the data is stationary. Per the Duke research team ("Introduction to ARIMA models," 2019),"A stationary series has no trend, its variations around its mean have a constant amplitude, and it wiggles consistently, i.e., its short-term random time patterns always look the same in a statistical sense. With that being established, the Armia model requires stationary data to predict future values from older data properly."

In this case, we are using data from 1929-2017. We will predict ten years from 2018-2028 and backtest the data from 1993-2003. First, let us backtest the data to find a suitable model to predict future values. We will do a ten year model from 1993-2003.

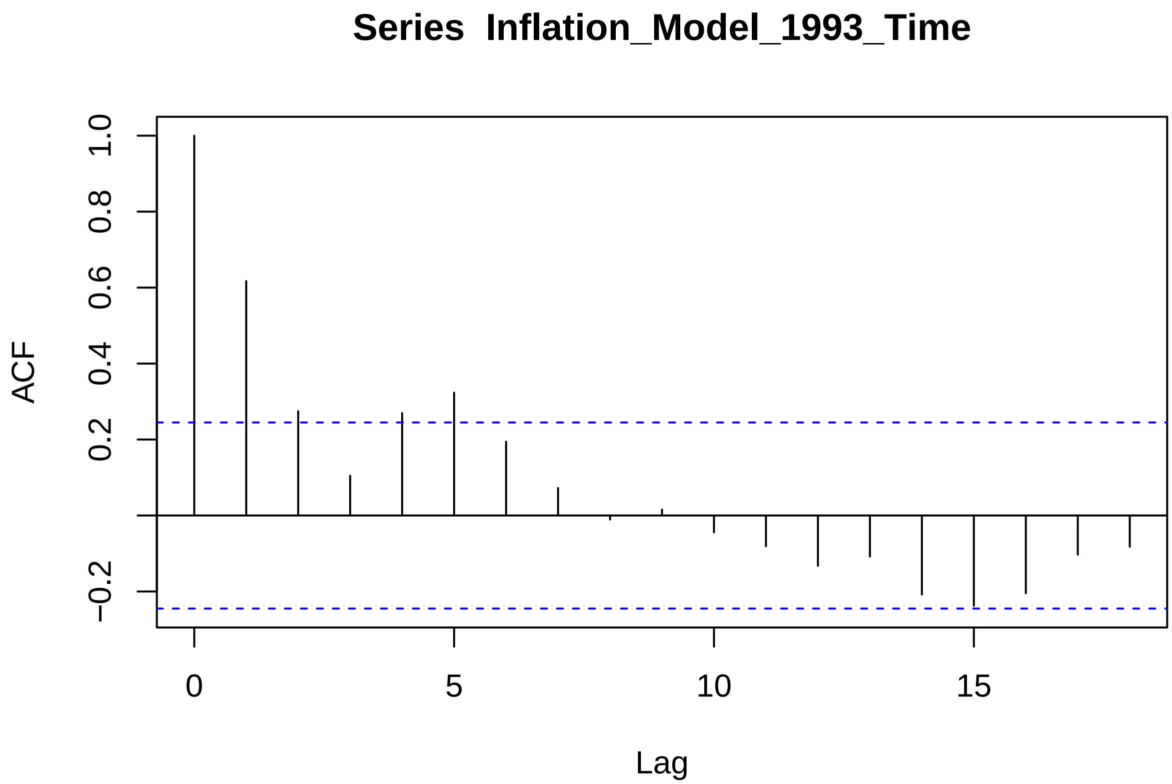
```
## Time Series:
## Start = 1929
## End = 1992
## Frequency = 1
##  [1]   0.6  -6.4  -9.3 -10.3   0.8   1.5   3.0   1.4   2.9  -2.8   0.0   0.7
## [13]   9.9   9.0   3.0   2.3   2.2  18.1   8.8   3.0  -2.1   5.9   6.0   0.8
## [25]   0.7  -0.7   0.4   3.0   2.9   1.8   1.7   1.4   0.7   1.3   1.6   1.0
## [37]   1.9   3.5   3.0   4.7   6.2   5.6   3.3   3.4   8.7  12.3   6.9   4.9
## [49]   6.7   9.0  13.3  12.5   8.9   3.8   3.8   3.9   3.8   1.1   4.4   4.4
## [61]   4.6   6.1   3.1   2.9
```



Portions of this data is from 1929−1992 and the Reference Section.

Verify the data using Auto-Correlation Function (ACF)

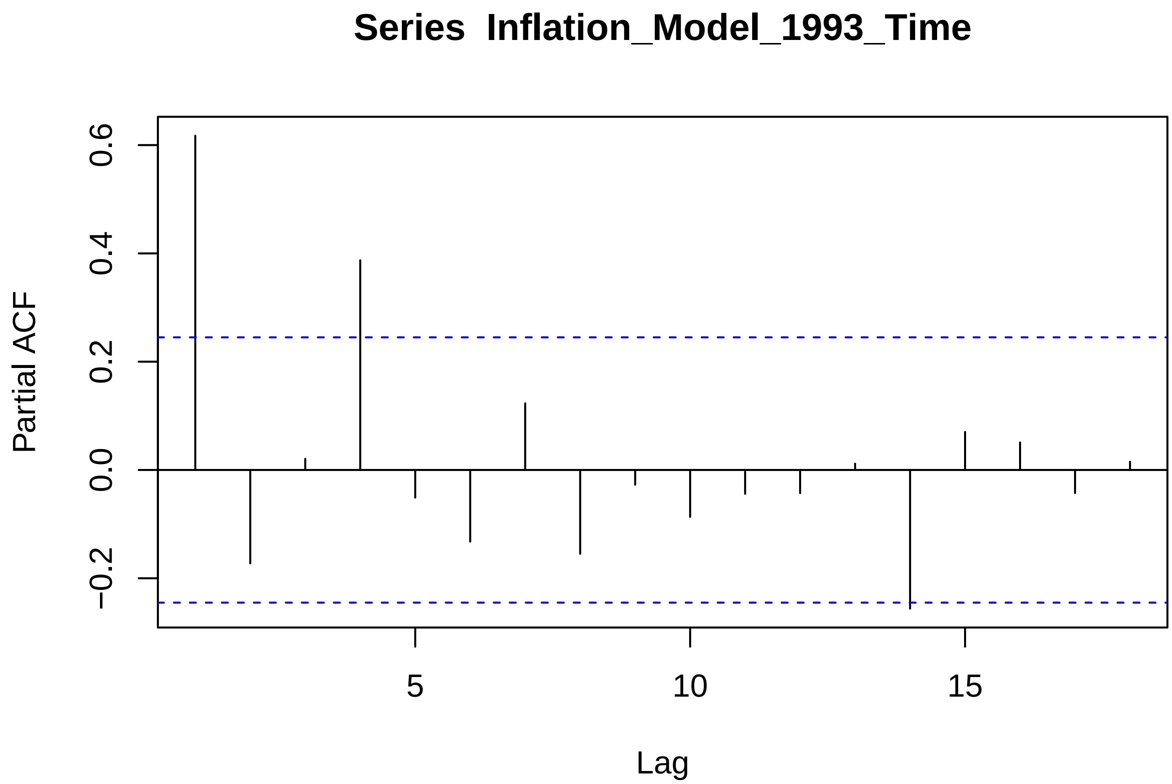
```
acf(Inflation_Model_1993_Time)
```



ACF shows us similarities over time using lagged data in a time series.
The autocorrelations within the blue upper and lower limits are considered significant. The insignificant autocorrelations exceed the blue upper and lower limits. In this ACF data plot, more data units pass the blue upper line, indicating that the data is not as stationary. Lag is a specific period of time; we will reference the lag with a number, i.e., lag 1.

Verify the data using Partial Auto-Correlation Function (PACF)

```
pacf(Inflation_Model_1993_Time)
```



The PACF data is measured by extracting the effects of any shorter lag correlations. In an ARIMA model, PACF can pinpoint the number of autoregression coefficients. PACF test is another verification that the data is not stationary due to the spikes in data that passed the blue upper and lower lines.

Our final verification will be the Augmented Dickey-Fuller Test. This test will determine whether the model data is stationary or nonstationary. P value if less than .05 means it is statistically significant.

Augmented Dickey–Fuller Test

```
adf.test(Inflation_Model_1993_Time)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  Inflation_Model_1993_Time
## Dickey-Fuller = -2.9543, Lag order = 3, p-value = 0.1883
## alternative hypothesis: stationary
```

Our P-Value of .18 (18%) is well above our minimum of .05 (5%). This means we have to alter our confidence level to 82. The ARIMA model we will use comprises three principles. P is the total amount of autoregressive terms, D is the amount of non-seasonal differences needed for the data to remain stationary, and Q is the amount of lagged forecasting errors in the prediction equation. This format is mirrored in the ARIMA (p,d,q). Selecting the correct ARIMA (p,d,q) is critical for this forecasting model. Null Hypothesis means autocorrelation does not exist, and Alternate Hypothesis means autocorrelation does not exist.

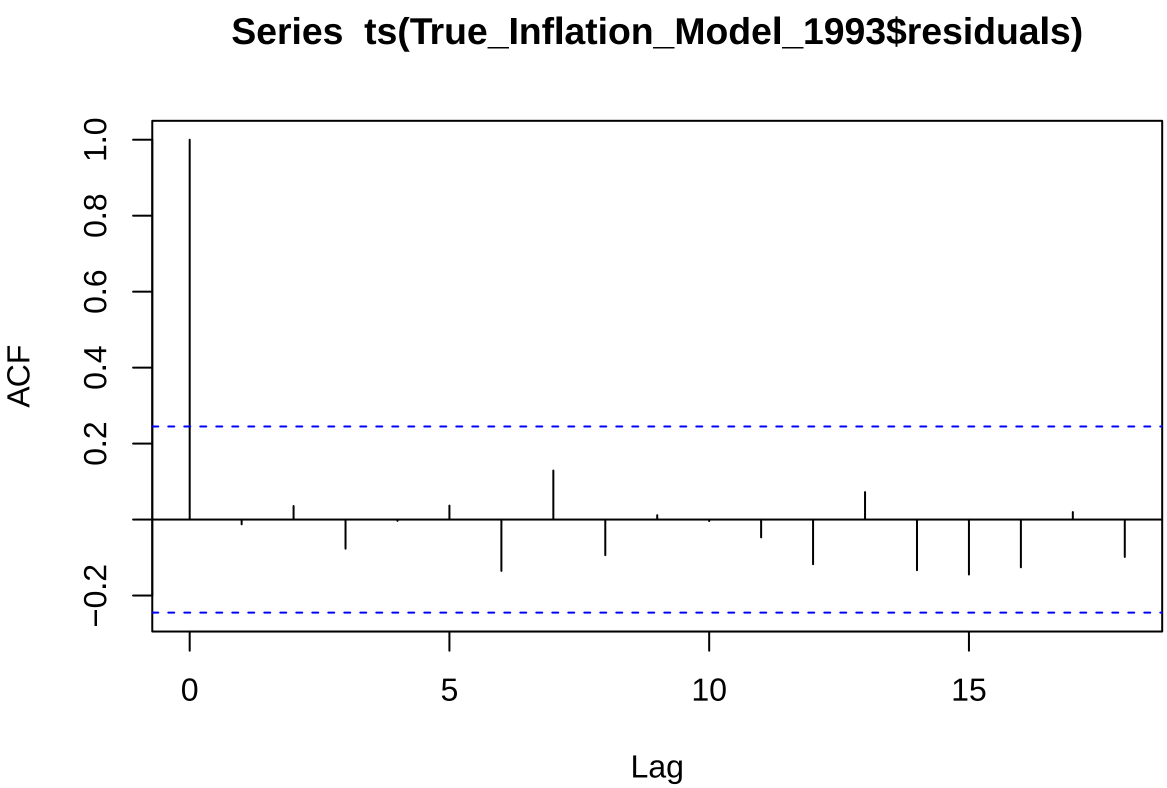
```
True_Inflation_Model_1993= auto.arima(Inflation_Model_1993_Time, ic="aic", trace= TRUE)
```

```
##
##  ARIMA(2,1,2) with drift         : 346.7356
##  ARIMA(0,1,0) with drift         : 359.3801
##  ARIMA(1,1,0) with drift         : 361.0432
```

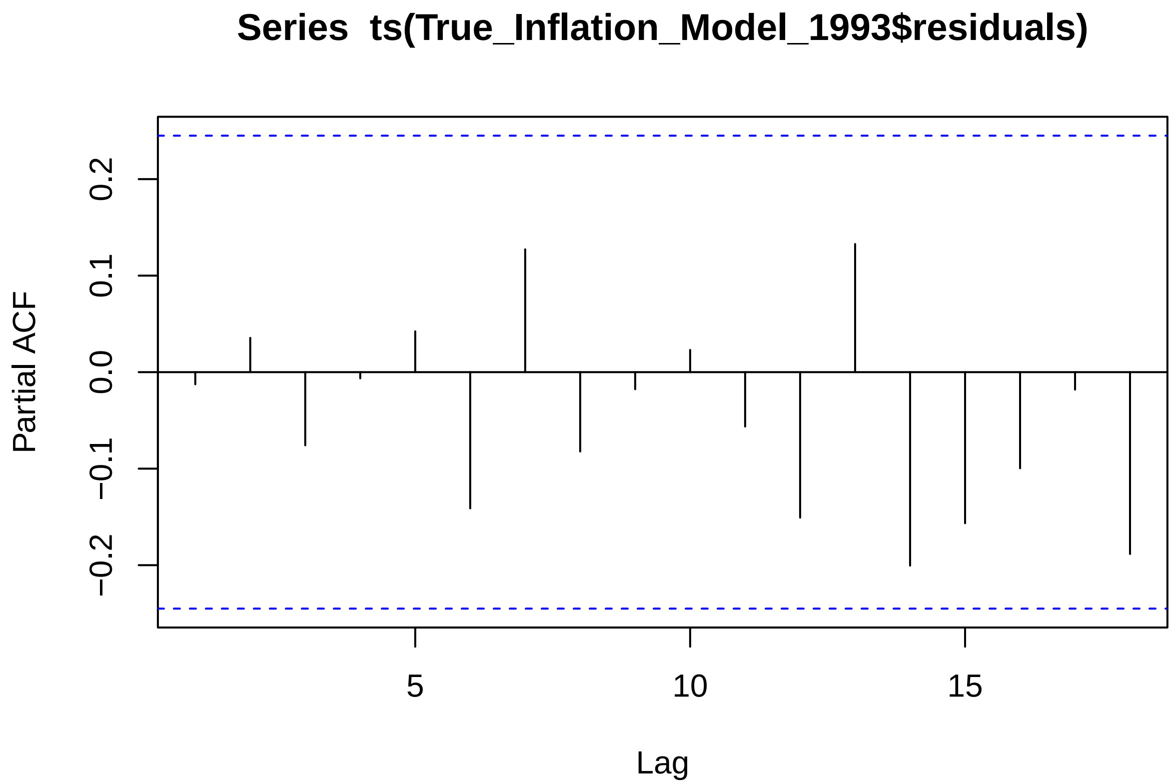
```
##  ARIMA(0,1,1) with drift         : 358.9609
##  ARIMA(0,1,0)                    : 357.3852
##  ARIMA(1,1,2) with drift         : Inf
##  ARIMA(2,1,1) with drift         : 349.6831
##  ARIMA(3,1,2) with drift         : 342.7439
##  ARIMA(3,1,1) with drift         : 340.7826
##  ARIMA(3,1,0) with drift         : 338.9625
##  ARIMA(2,1,0) with drift         : 359.0333
##  ARIMA(4,1,0) with drift         : 340.754
##  ARIMA(4,1,1) with drift         : Inf
##  ARIMA(3,1,0)                    : 337.4522
##  ARIMA(2,1,0)                    : 357.0865
##  ARIMA(4,1,0)                    : 339.2993
##  ARIMA(3,1,1)                    : 339.3246
##  ARIMA(2,1,1)                    : 348.1981
##  ARIMA(4,1,1)                    : Inf
##
##  Best model: ARIMA(3,1,0)
```

The best ARIMA for our model will be ARIMA(3,1,0), verify that the data is stationary and smoothed

```
acf(ts(True_Inflation_Model_1993$residuals))
```




```
pacf(ts(True_Inflation_Model_1993$residuals))
```



Now that the data is smoothed and fits our current ARIMA model, let us forecast what inflation would be in 10 years starting from 1993. **Note:** *h = the number of years in the future. Level 82 = the confidence level for our model.*

```
True_Inflation_Model_1993
```

```
## Series: Inflation_Model_1993_Time
## ARIMA(3,1,0)
##
## Coefficients:
##           ar1      ar2      ar3
##       -0.2472  -0.3188  -0.5458
## s.e.   0.1073   0.1035   0.1049
##
## sigma^2 = 11.26:  log likelihood = -164.73
## AIC=337.45   AICc=338.14   BIC=346.02
```

```
Inflation_Model_Forecast_1993 = forecast(True_Inflation_Model_1993, level = c(95), h = 11)
```

Lets validate the data using the Ljung Box.test to verify that the residuals are not just "white noise"

```
Box.test(Inflation_Model_Forecast_1993, lag = 1, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  Inflation_Model_Forecast_1993
## X-squared = 0.0073422, df = 1, p-value = 0.9317
```

If the P value is less than .05 that means that the data has autocorrelation significance with a 95% confidence level.

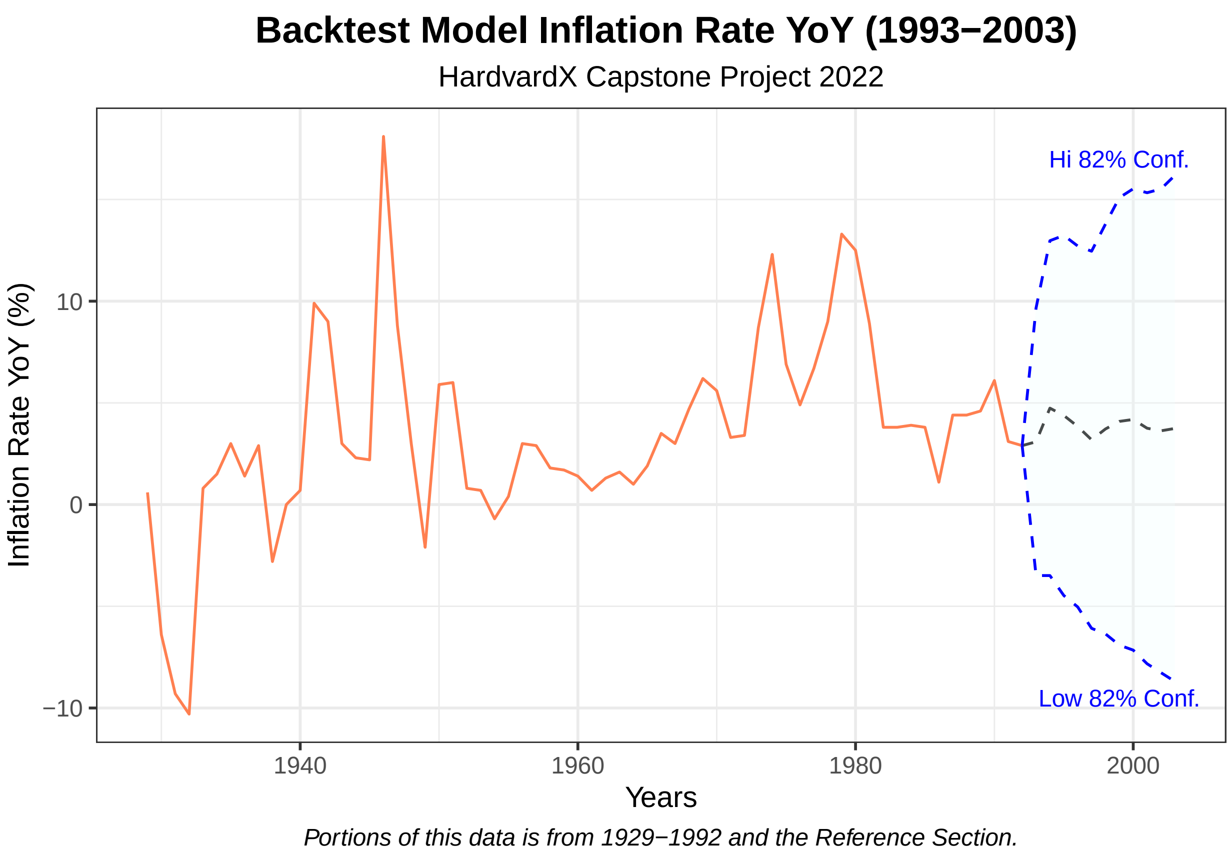
```
Box.test(Inflation_Model_Forecast_1993, lag = 5, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  Inflation_Model_Forecast_1993
## X-squared = 8.9147, df = 5, p-value = 0.1125
```

Lets see the results and chart it

```
Inflation_Model_1993_2003 = auto.arima(Inflation_Model_1993_Time, ic="aic", trace= TRUE,)

Backtest_Inflation_Model_Forecast = forecast(Inflation_Model_1993_2003 , level = c(95), h = 11)
```

```
Backtest_Inflation_Model_Forecast
```

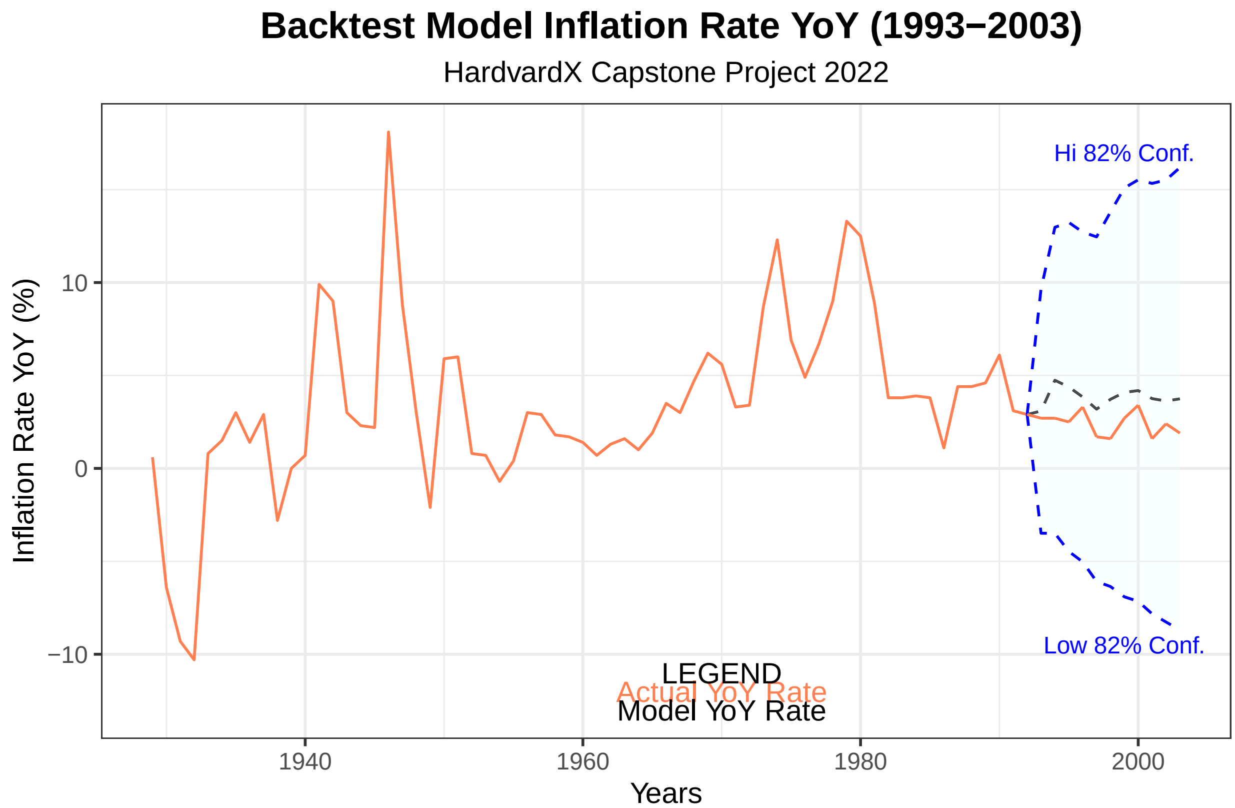


Let's examine how many years are within +/- 1000 basis points of the original Fed Data object Inflation_Model.

Inflation Rate Year over Year Model HardvardX Capstone Project 2022				
Year	Inflation_YoY	Model_Inflation_YoY_(	Predict_Diff	Within_Tolerance
1993	2.7	3.1	0.4	**TRUE**
1994	2.7	4.7	2.0	*FALSE*
1995	2.5	4.4	1.9	*FALSE*
1996	3.3	3.8	0.5	**TRUE**
1997	1.7	3.2	1.5	*FALSE*
1998	1.6	3.7	2.1	*FALSE*
1999	2.7	4.1	1.4	*FALSE*
2000	3.4	4.2	0.8	**TRUE**
2001	1.6	3.8	2.2	*FALSE*
2002	2.4	3.6	1.2	*FALSE*
2003	1.9	3.7	1.8	*FALSE*

Inflation Rate YoY is based on data from 1993-2003
Portions of this data is from the Reference Section

The model has more false results than true. This data is nearly double the inflation rate based on our historical data. We have to make several adjustments moving forward. Let us view the backtest model vs actual data chart.



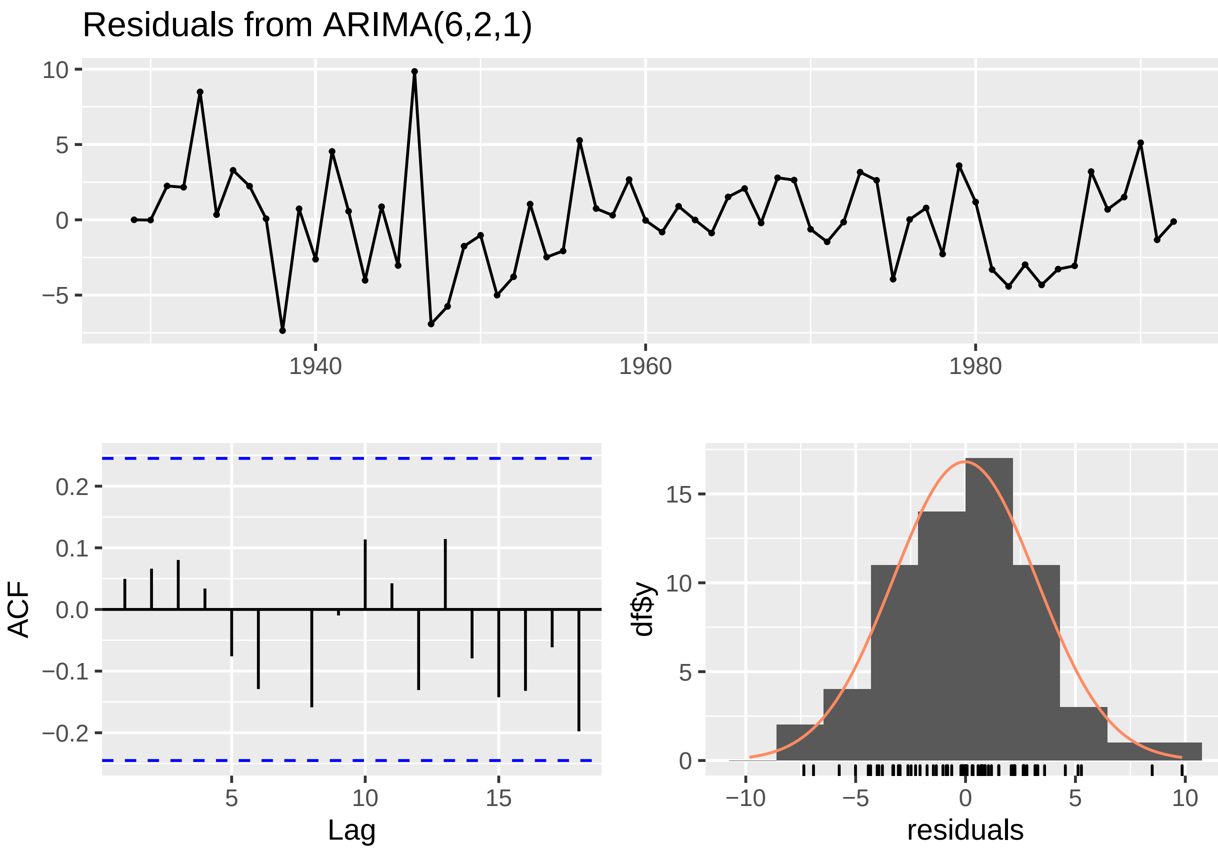
Portions of this data is from 1929−2003 and the Reference Section.

Let's adjust the arima(p,d,q), review and update our P-Value.

```
Inflation_Model_1993_2003 = arima(fixed = NULL, Inflation_Model_1993_Time, order = c(6,2,1), transform.p

Backtest_Inflation_Model_Forecast_1 = forecast(Inflation_Model_1993_2003, level = c(82), h = 11)

Backtest_Inflation_Model_Forecast_1
```

Verify our model by running check residual and update the conf. This is a series of diagnostic tests that we will use to validate our model (Long, 2019).

```
checkresiduals(Backtest_Inflation_Model_Forecast_1)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(6,2,1)
## Q* = 5.5324, df = 3, p-value = 0.1367
##
## Model df: 7.   Total lags used: 10
```

```
Backtest_Inflation_Model_Forecast_1 = forecast(Inflation_Model_1993_2003, level = c(87), h = 11)
```

Great improvement.

Per (Long, 2019), to verify check residuals, we should look for the following:

1. Do the standardized residuals show volatility clusters?
   - If so, readjust the model.
2. Does the ACF show significant autocorrelation between the residuals?
   - If so, readjust the model.
3. Do the residuals look bell-shaped?
   - If so, this suggests they are reasonably symmetrical.
4. Are the p-value in the Ljung–Box test extensive?
   - If so, this indicates the residuals are patternless meaning the model has extracted all the information, and the only noise is left behind.

Let us compare the data and adjust the confidence level to 87% based on the new P-Value of .13 (13%). [4]

Inflation Rate Year over Year Model

HardvardX Capstone Project 2022

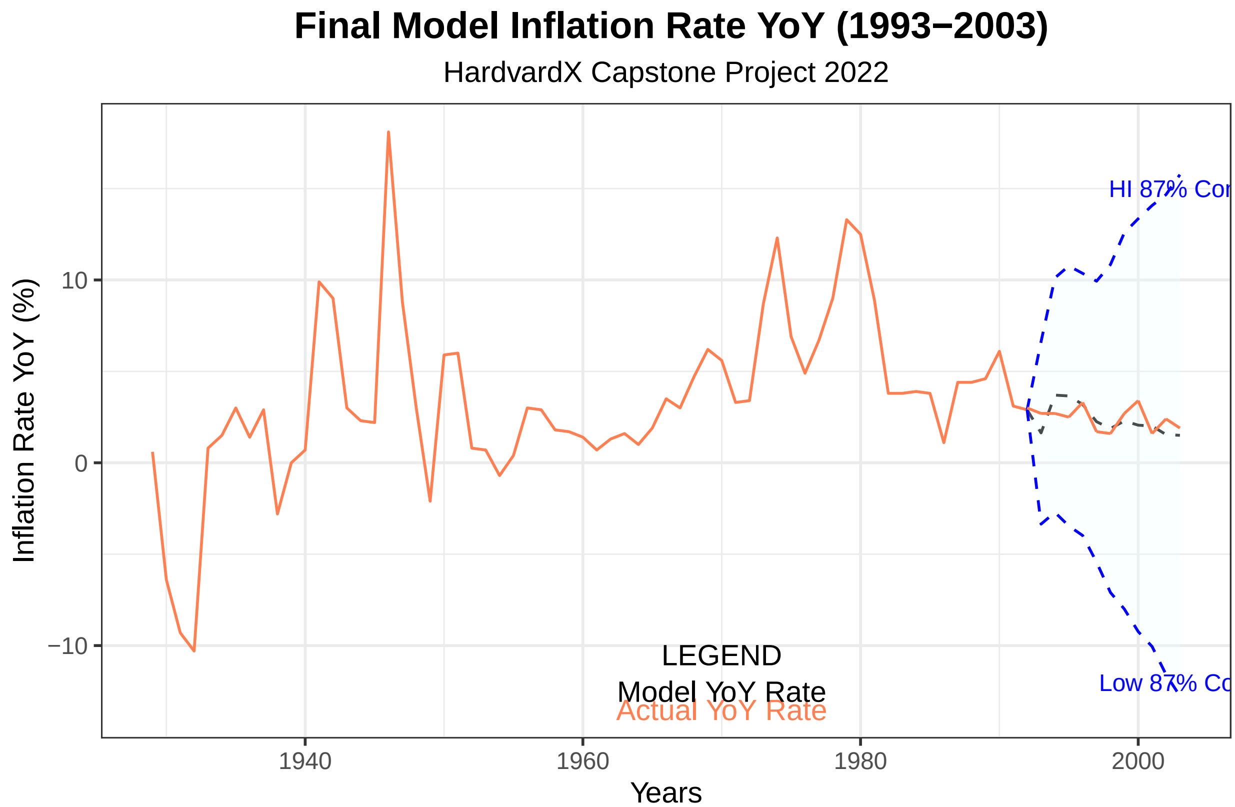
Year	Inflation_YoY	Model_Inflation_YoY	Predict_Diff	Within_Tolerance
1993	2.7	1.6	1.1	*FALSE*
1994	2.7	3.7	-1.0	**TRUE**
1995	2.5	3.7	-1.2	**TRUE**
1996	3.3	3.2	0.1	**TRUE**
1997	1.7	2.2	-0.5	**TRUE**
1998	1.6	1.9	-0.3	**TRUE**
1999	2.7	2.3	0.4	**TRUE**
2000	3.4	2.0	1.4	*FALSE*
2001	1.6	2.0	-0.4	**TRUE**
2002	2.4	1.5	0.9	**TRUE**
2003	1.9	1.5	0.4	**TRUE**

Inflation Rate YoY is based on data from 1993-2003

Portions of this data is from the Reference Section

Amazing! Our model is now closer to our goal. Let us view it in a chart.

[4] ***Correction: row 3 (1995) should be FALSE***

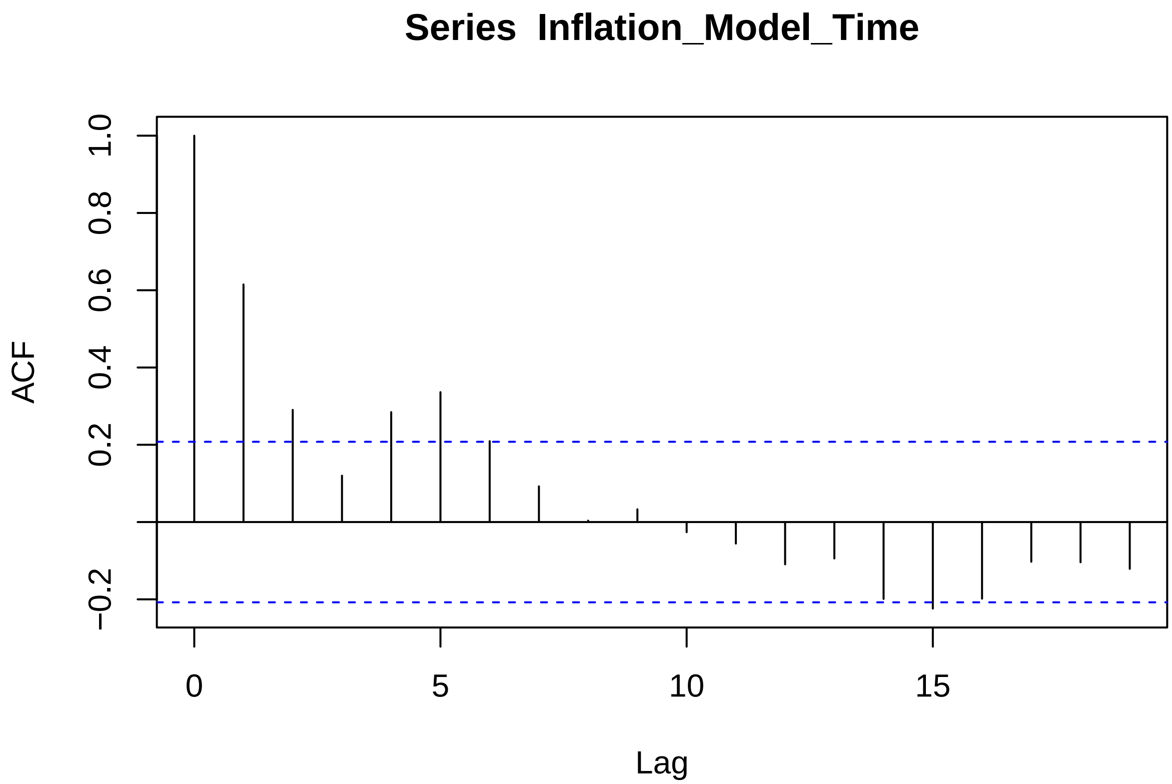


Portions of this data is from 1929−1992 and the Reference Section.

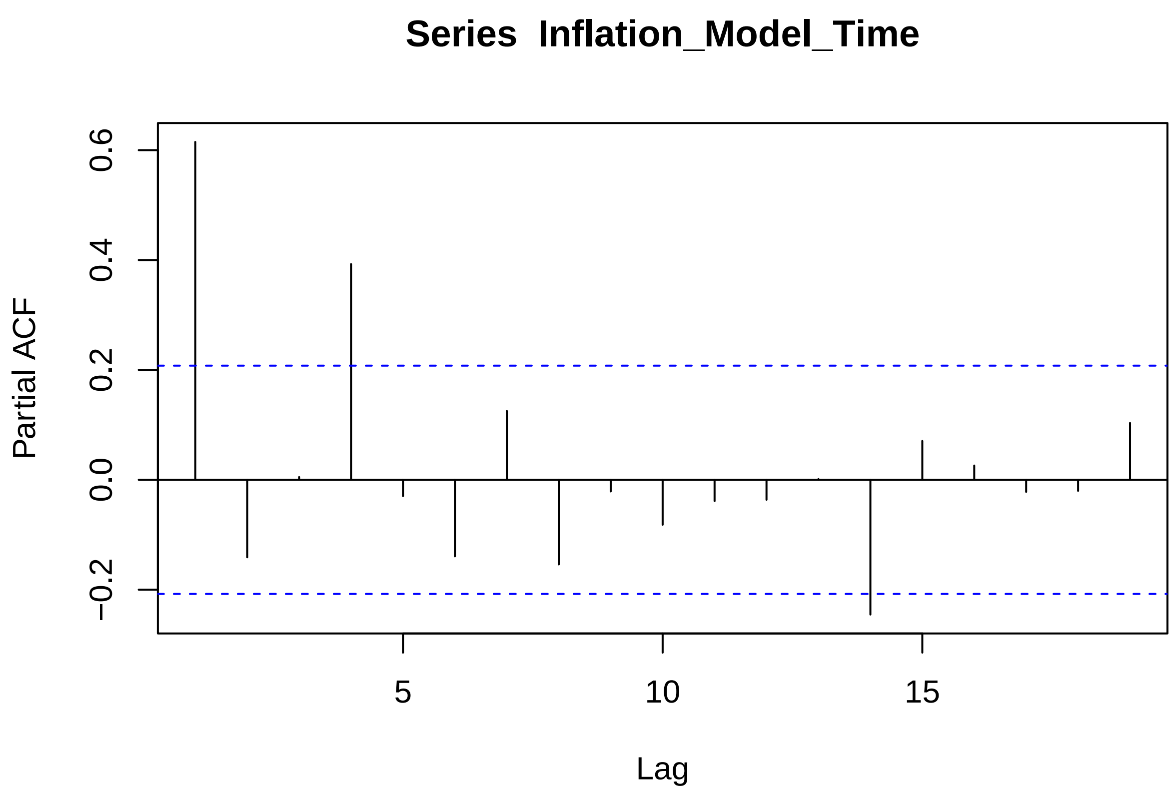
Now that we have a great model we created when we backtested the data, let us create a model that can predict future values of the Inflation Rate YoY.

Verify the data

```
acf(Inflation_Model_Time)
```




```r
pacf(Inflation_Model_Time)
```




```r
adf.test(Inflation_Model_Time)
```

```
##
```

```
##  Augmented Dickey-Fuller Test
##
## data:  Inflation_Model_Time
## Dickey-Fuller = -2.9281, Lag order = 4, p-value = 0.1942
## alternative hypothesis: stationary
```

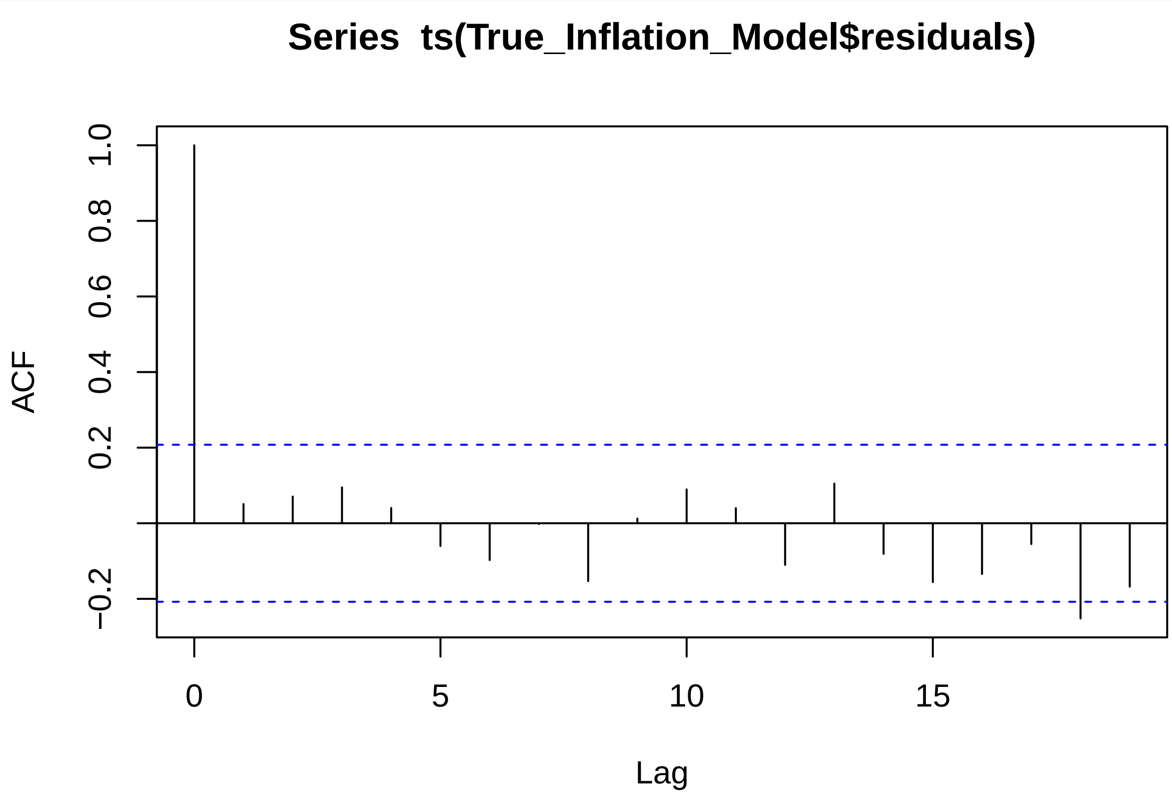
Utilize the ARIMA from our backtest model

```
True_Inflation_Model = arima(fixed = NULL, Inflation_Model_Time, order = c(6,2,1), transform.pars=TRUE)

True_Inflation_Model
```

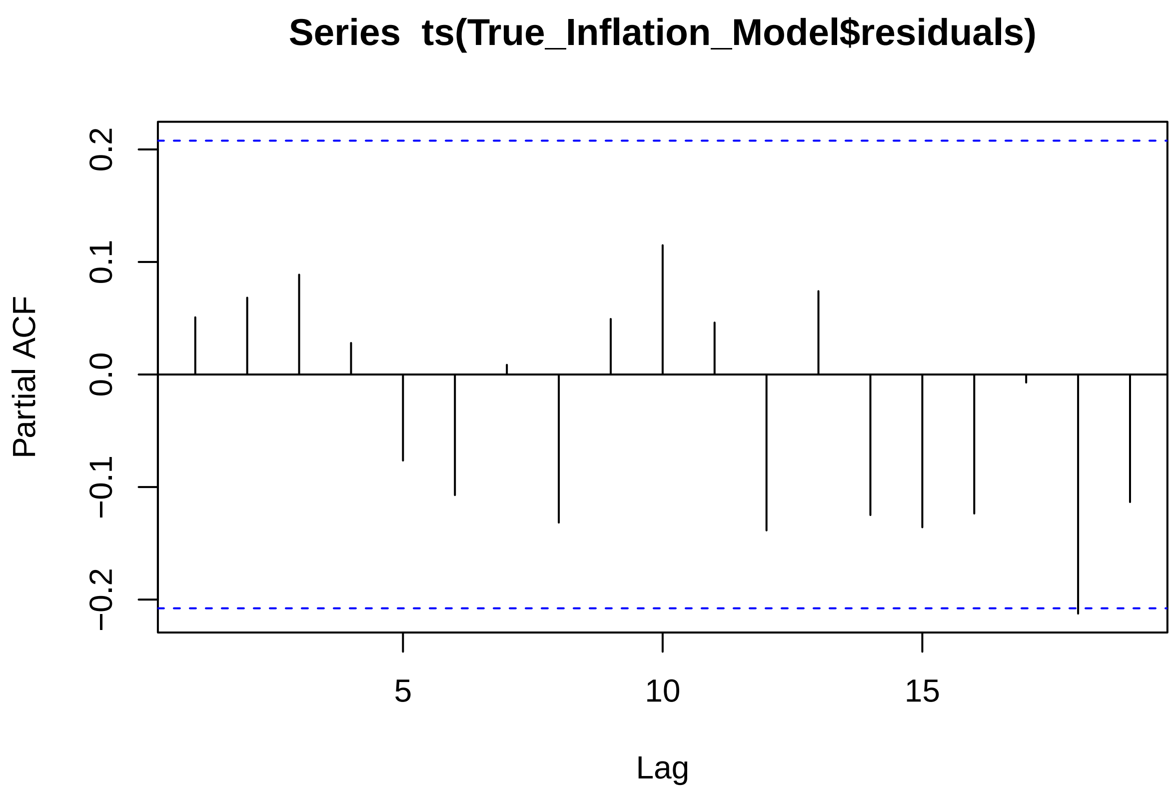
```
##
## Call:
## arima(x = Inflation_Model_Time, order = c(6, 2, 1), transform.pars = TRUE, fixed = NULL)
##
## Coefficients:
##           ar1      ar2      ar3      ar4      ar5      ar6      ma1
##       -0.4207  -0.4801  -0.8078  -0.2887  -0.1219  -0.2351  -0.8095
## s.e.   0.2283   0.2739   0.2967   0.3063   0.2435   0.1663   0.2383
##
## sigma^2 estimated as 8.173:  log likelihood = -217.17,  aic = 450.33
```

verify that the data is stationary and smoothed

```
acf(ts(True_Inflation_Model$residuals))
```




```
pacf(ts(True_Inflation_Model$residuals))
```



Now that the data is smoothed and fits our current ARIMA model lets forecast what inflation would be in 10 years starting from 2018.

```
True_Inflation_Model
```

```
##
## Call:
## arima(x = Inflation_Model_Time, order = c(6, 2, 1), transform.pars = TRUE, fixed = NULL)
##
## Coefficients:
##           ar1      ar2      ar3      ar4      ar5      ar6      ma1
##       -0.4207  -0.4801  -0.8078  -0.2887  -0.1219  -0.2351  -0.8095
## s.e.   0.2283   0.2739   0.2967   0.3063   0.2435   0.1663   0.2383
##
## sigma^2 estimated as 8.173:  log likelihood = -217.17,  aic = 450.33
```

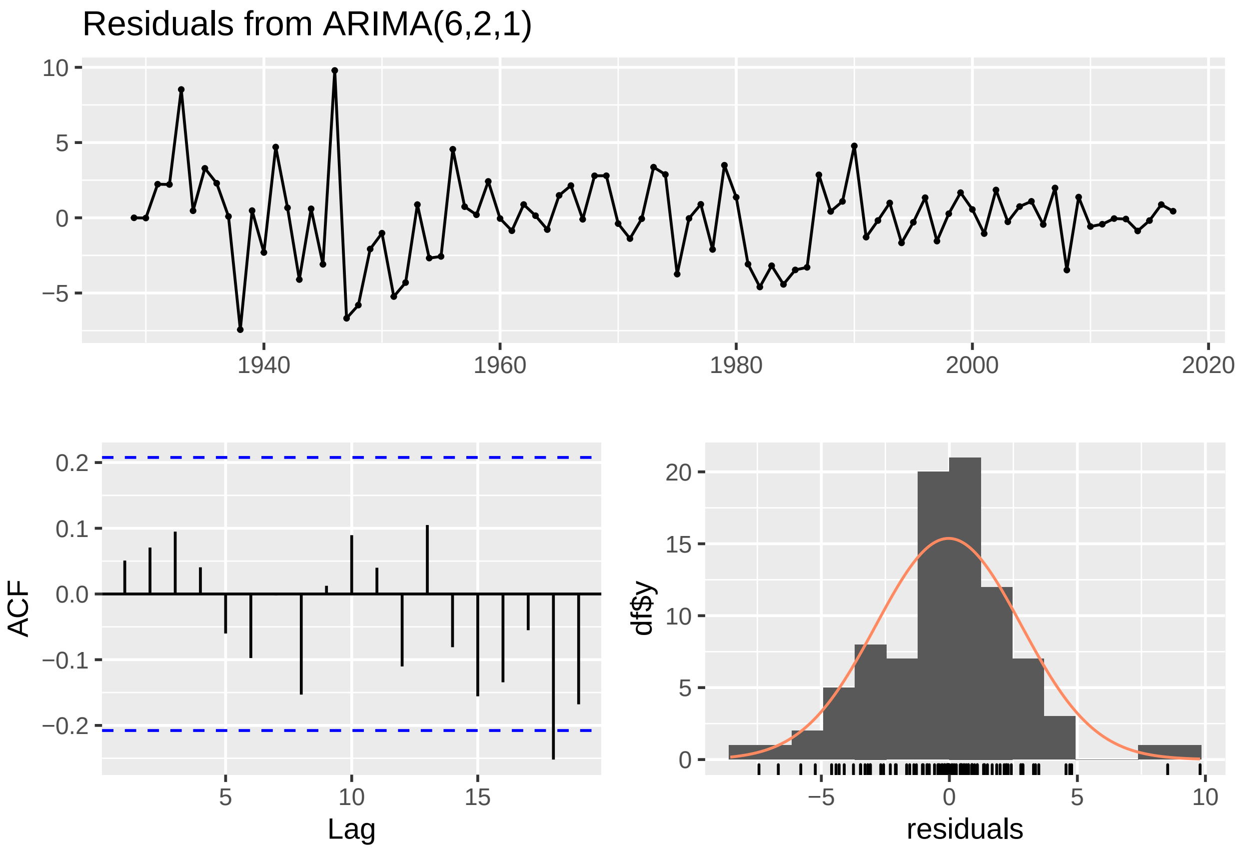
```
Inflation_Model_Forecast = forecast(True_Inflation_Model, level = c(81), h = 11)

Inflation_Model_Forecast
```

```
##      Point Forecast     Lo 81    Hi 81
## 2018      1.7988174 -1.947976 5.545611
## 2019      0.7138813 -4.014452 5.442214
## 2020      0.8455106 -4.353713 6.044735
## 2021      1.1651595 -4.064928 6.395247
## 2022      1.3595859 -4.209642 6.928813
## 2023      1.1259471 -5.298354 7.550248
## 2024      0.7957438 -6.495552 8.087040
## 2025      0.7945696 -7.129842 8.718981
## 2026      0.7742601 -7.623300 9.171820
## 2027      0.7766805 -8.296273 9.849634
```

```
## 2028       0.6224466 -9.199691 10.444584
```

Verify Model

```r
checkresiduals(Inflation_Model_Forecast)
```



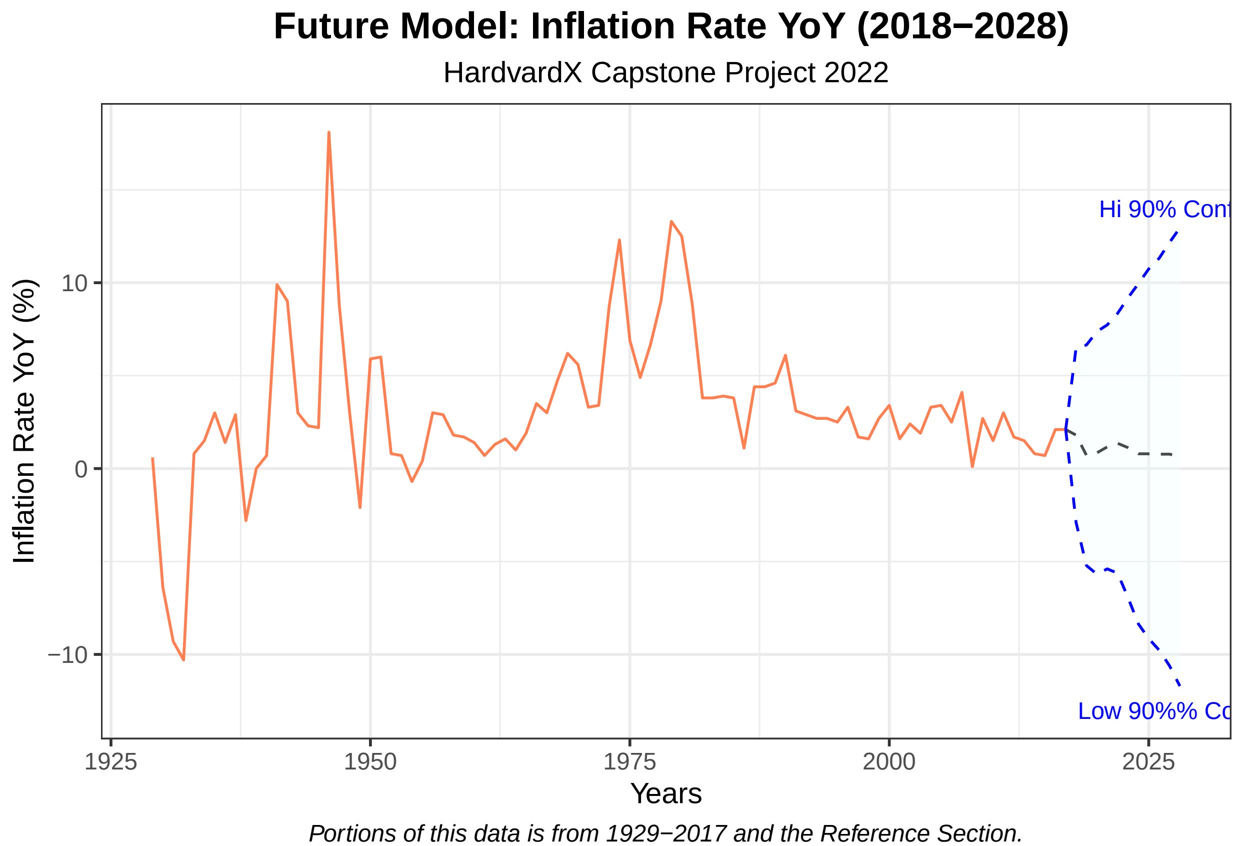

```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(6,2,1)
## Q* = 6.1637, df = 3, p-value = 0.1039
##
## Model df: 7.   Total lags used: 10
```

Lets update the confidence level to 90% since the P Value is .10 (10%)

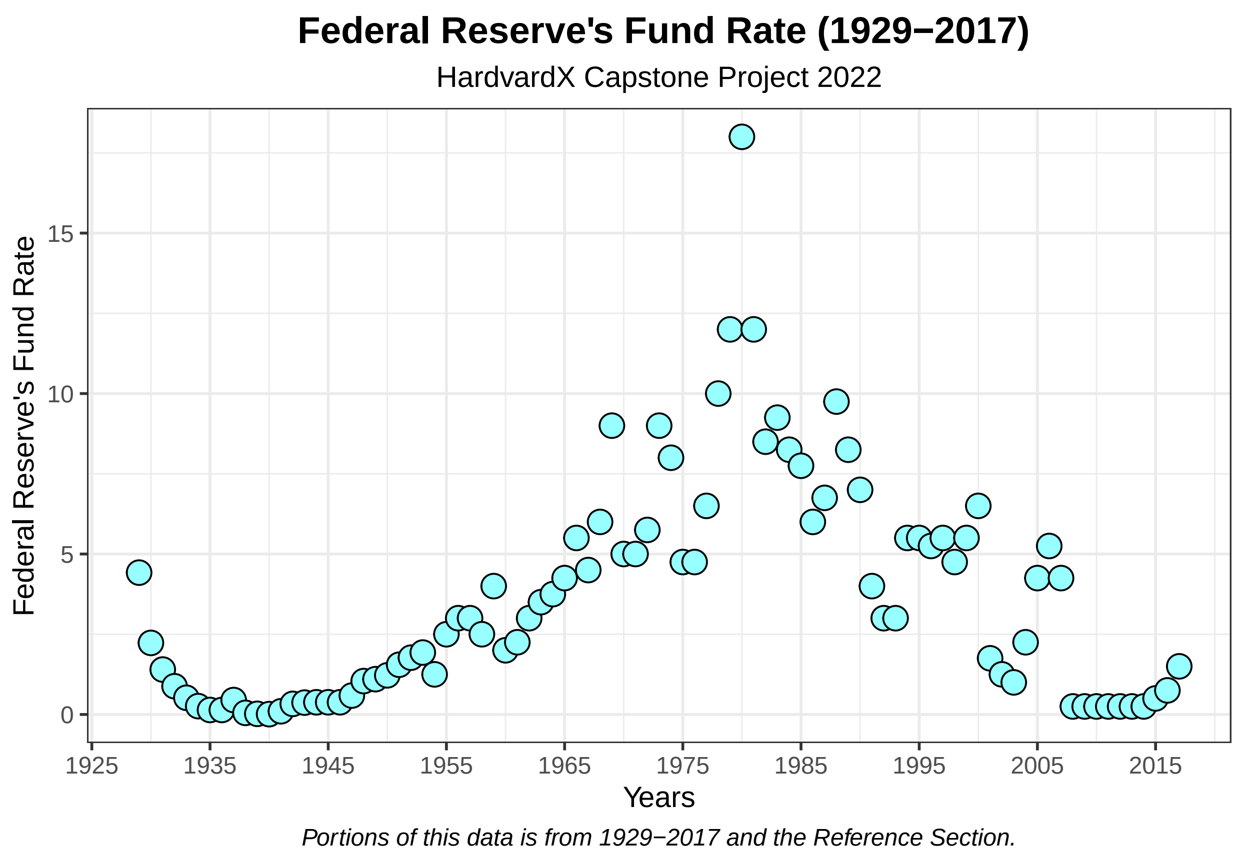
```r
Inflation_Model_Forecast_Update = forecast(True_Inflation_Model, level = c(90), h = 11)

Inflation_Model_Forecast_Update
```

```
##      Point Forecast     Lo 90    Hi 90
## 2018      1.7988174 -2.903628 6.501263
## 2019      0.7138813 -5.220454 6.648216
## 2020      0.8455106 -5.679820 7.370842
## 2021      1.1651595 -5.398907 7.729227
## 2022      1.3595859 -5.630121 8.349293
## 2023      1.1259471 -6.936927 9.188822
```

```
## 2024      0.7957438  -8.355260  9.946748
## 2025      0.7945696  -9.151031 10.740171
## 2026      0.7742601  -9.765170 11.313690
## 2027      0.7766805 -10.610408 12.163769
## 2028      0.6224466 -11.704911 12.949805
```



Let us create a Federal Reserve's Fed Funds Rate model to predict the future rate and backtest prior data. Create a table with all the variables needed.

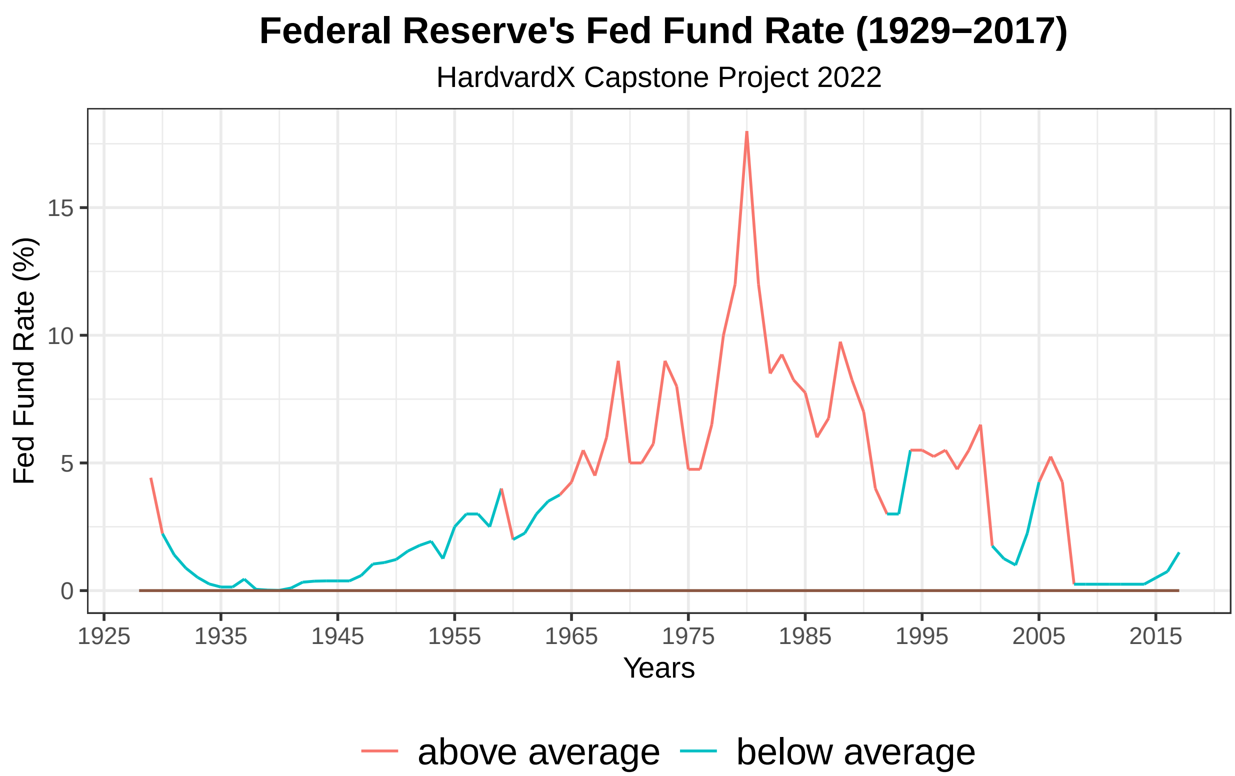


Verify the Model and see if its in a Time Series

```r
class(Fed_Model)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

Its not so lets convert it



Portions of this data is from 1929−2017 and the Reference Section.
The Federal Funds Rate Average is 3.67.

Create a Backtest model

```
adf.test(Fed_Model_1993_Time)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  Fed_Model_1993_Time
## Dickey-Fuller = -2.3703, Lag order = 3, p-value = 0.4249
## alternative hypothesis: stationary
```

```
Fed_Model_1993_2003 = auto.arima(Fed_Model_1993_Time, ic="aic", trace= TRUE,)
```

```
##
##  ARIMA(2,1,2) with drift         : 258.221
##  ARIMA(0,1,0) with drift         : 255.8698
##  ARIMA(1,1,0) with drift         : 257.8299
##  ARIMA(0,1,1) with drift         : 257.806
##  ARIMA(0,1,0)                    : 253.8799
##  ARIMA(1,1,1) with drift         : Inf
##
##  Best model: ARIMA(0,1,0)
```

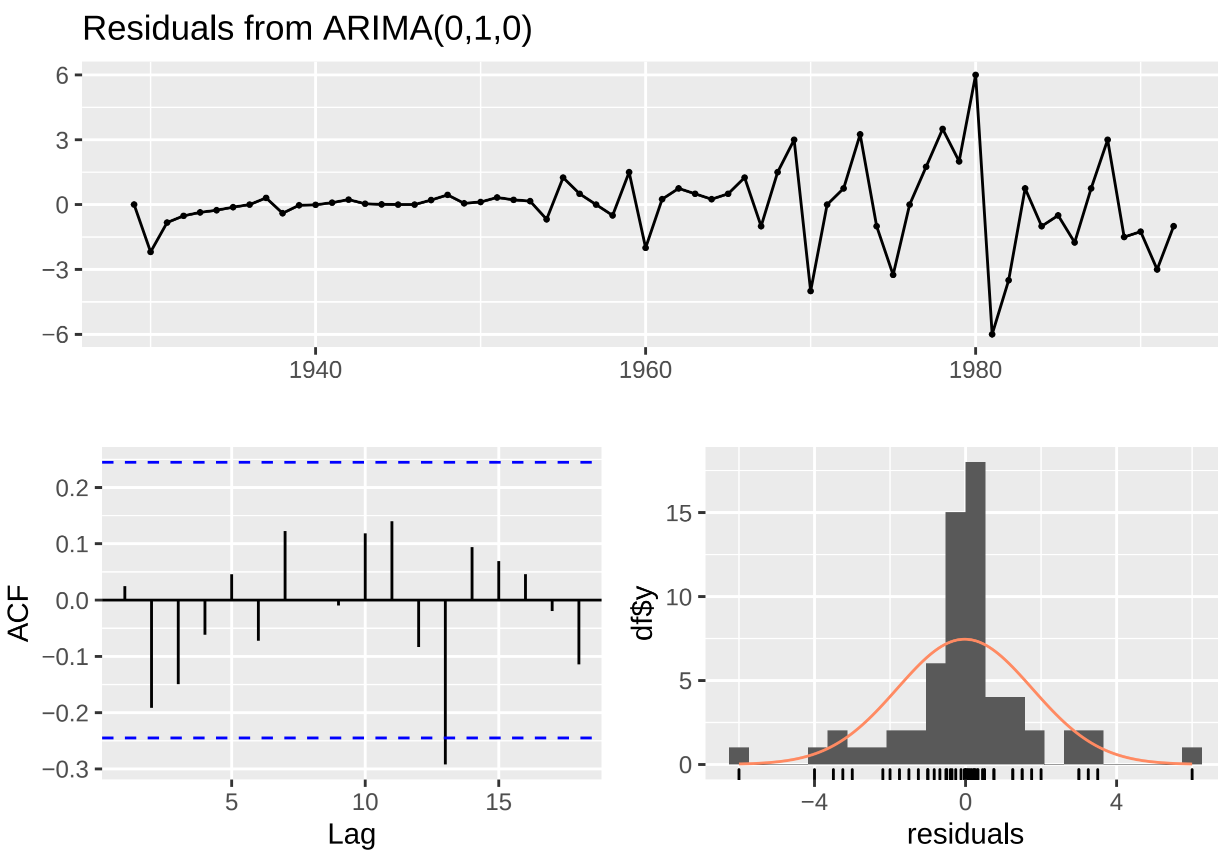
```
Backtest_Fed_Model_Forecast = forecast(Fed_Model_1993_2003, level = c(58), h = 11)

Backtest_Fed_Model_Forecast
```

```
##      Point Forecast     Lo 58    Hi 58
## 1993              3 1.5595451 4.440455
```

```
## 1994               3  0.9628891 5.037111
## 1995               3  0.5050589 5.494941
## 1996               3  0.1190901 5.880910
## 1997               3 -0.2209552 6.220955
## 1998               3 -0.5283796 6.528380
## 1999               3 -0.8110855 6.811086
## 2000               3 -1.0742218 7.074222
## 2001               3 -1.3213648 7.321365
## 2002               3 -1.5551185 7.555118
## 2003               3 -1.7774485 7.777449
```

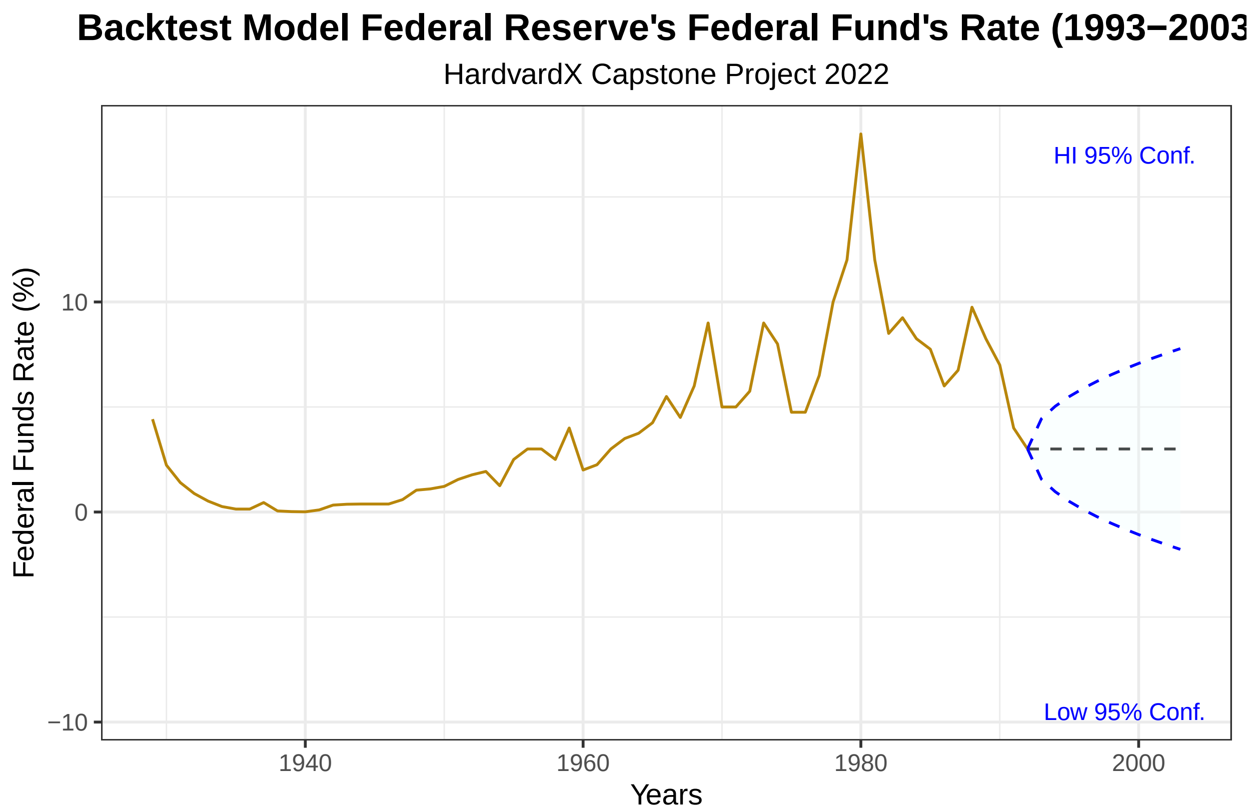
Verify Model

```
checkresiduals(Backtest_Fed_Model_Forecast)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(0,1,0)
## Q* = 7.0945, df = 10, p-value = 0.7165
##
## Model df: 0.   Total lags used: 10
```

Chart the Model



Portions of this data is from 1929−1992 and the Reference Section.

This backtest model is only computing the value 3 as a prediction. We want the prediction within 1 point (1000 basis points) of the original Fed Model object Fed_Fund_Rate. After viewing the results from check residuals, we noticed that the data didn't check all the boxes referenced above and P-Value is .71 (71%) which gives us a confidence level of 29%. Let's manually select the ARIMA.

```
adf.test(Fed_Model_1993_Time)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  Fed_Model_1993_Time
## Dickey-Fuller = -2.3703, Lag order = 3, p-value = 0.4249
## alternative hypothesis: stationary
```

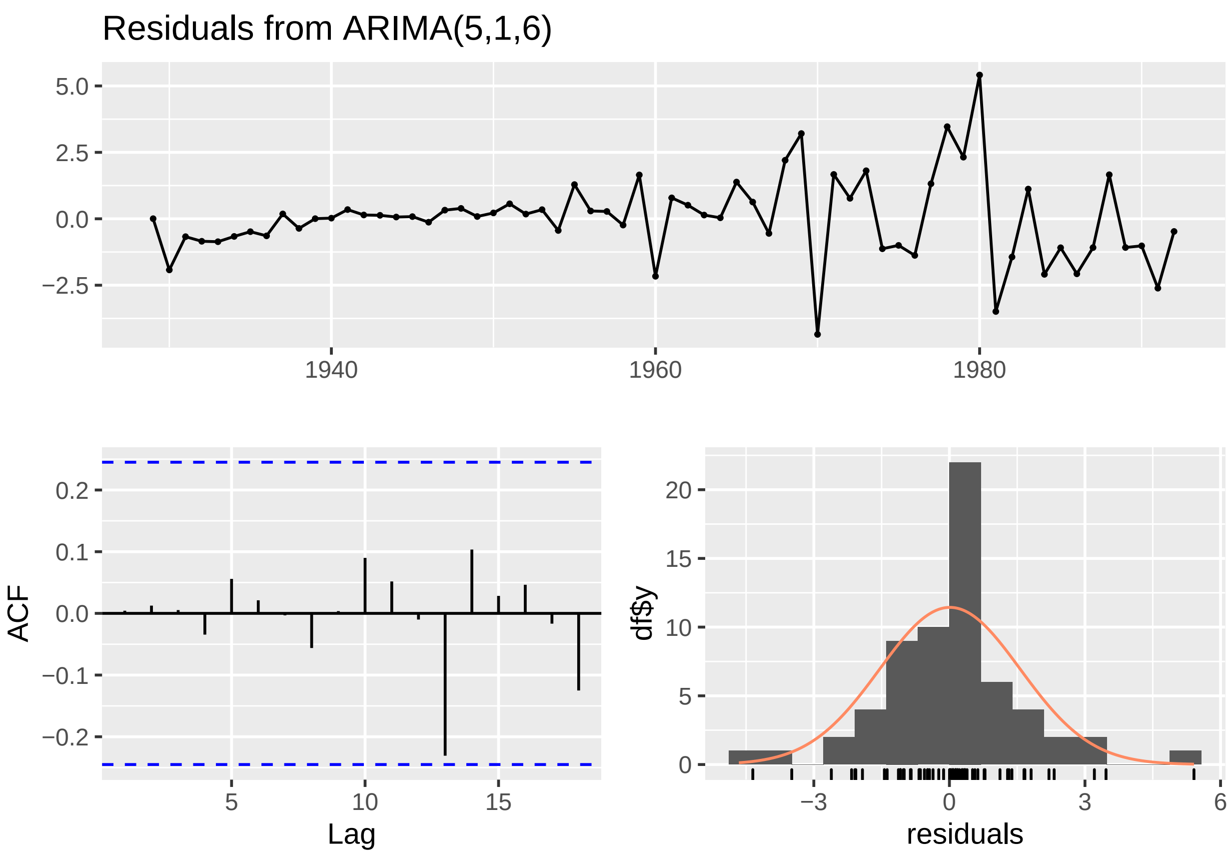
```
Fed_Model_1993_2003 = arima(fixed = NULL, Fed_Model_1993_Time, order = c(5,1,6),  transform.pars=TRUE)

Backtest_Fed_Model_Forecast_Update = forecast(Fed_Model_1993_2003, level = c(30), h = 11)

Backtest_Fed_Model_Forecast_Update
```

```
##      Point Forecast    Lo 30    Hi 30
## 1993       3.659781 3.048595 4.270966
## 1994       4.577001 3.706645 5.447357
## 1995       5.032248 4.032319 6.032176
## 1996       4.873646 3.820492 5.926801
## 1997       5.353607 4.263323 6.443891
## 1998       5.021209 3.864710 6.177708
```

```
## 1999       4.502486 3.305557 5.699416
## 2000       4.312817 3.059936 5.565699
## 2001       4.537356 3.203546 5.871165
## 2002       4.279734 2.863492 5.695976
## 2003       4.452451 2.981904 5.922997
```

Verify Model

```
checkresiduals(Backtest_Fed_Model_Forecast_Update)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(5,1,6)
## Q* = 6.76, df = 3, p-value = 0.07995
##
## Model df: 11.   Total lags used: 14
```

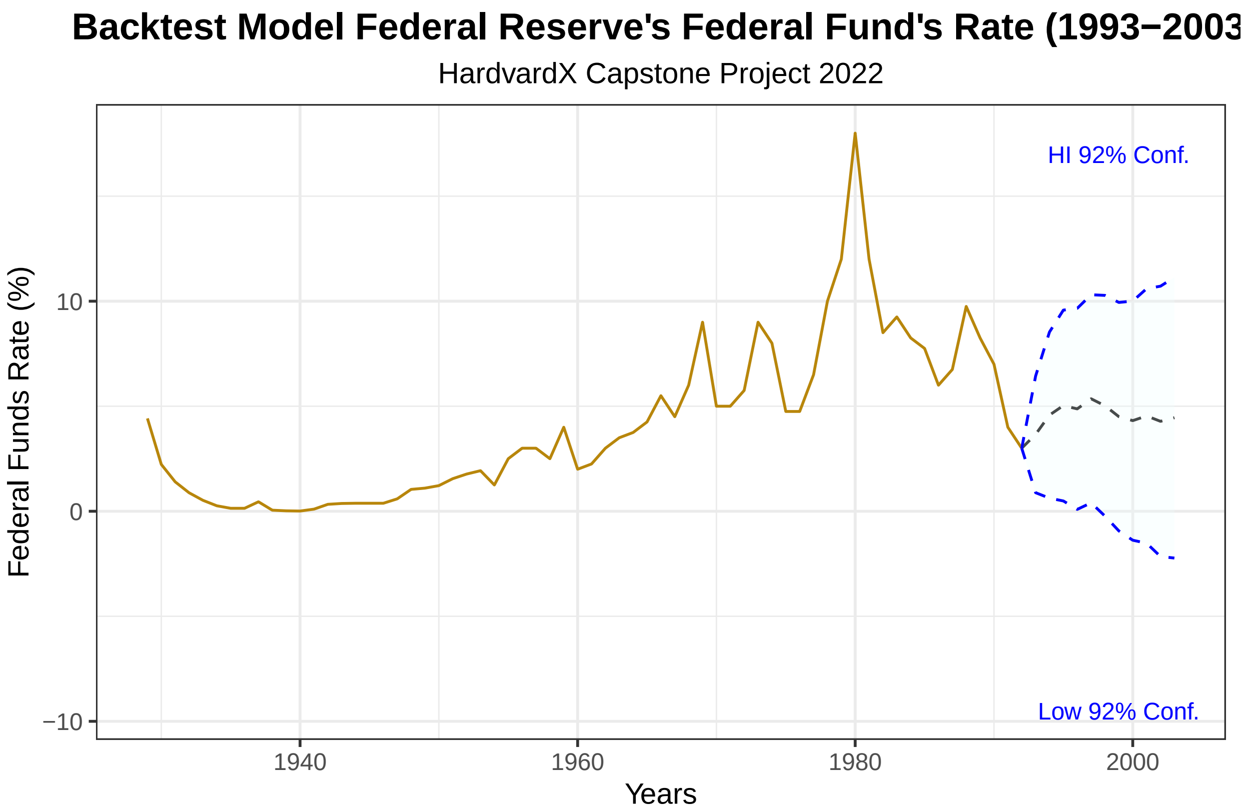
Let's update our conf levels to 92% based on our new p value.

```
Backtest_Fed_Model_Forecast_Update = forecast(Fed_Model_1993_2003, level = c(92), h = 11)

Backtest_Fed_Model_Forecast_Update
```

```
##      Point Forecast      Lo 92    Hi 92
## 1993       3.659781 0.88288706 6.436674
```

```
## 1994       4.577001  0.62257693  8.531425
## 1995       5.032248  0.48911832  9.575377
## 1996       4.873646  0.08868634  9.658606
## 1997       5.353607  0.39995130 10.307263
## 1998       5.021209 -0.23329187 10.275710
## 1999       4.502486 -0.93570761  9.940681
## 2000       4.312817 -1.37959366 10.005228
## 2001       4.537356 -1.52274895 10.597460
## 2002       4.279734 -2.15489770 10.714366
## 2003       4.452451 -2.22891037 11.133812
```

Chart the Data



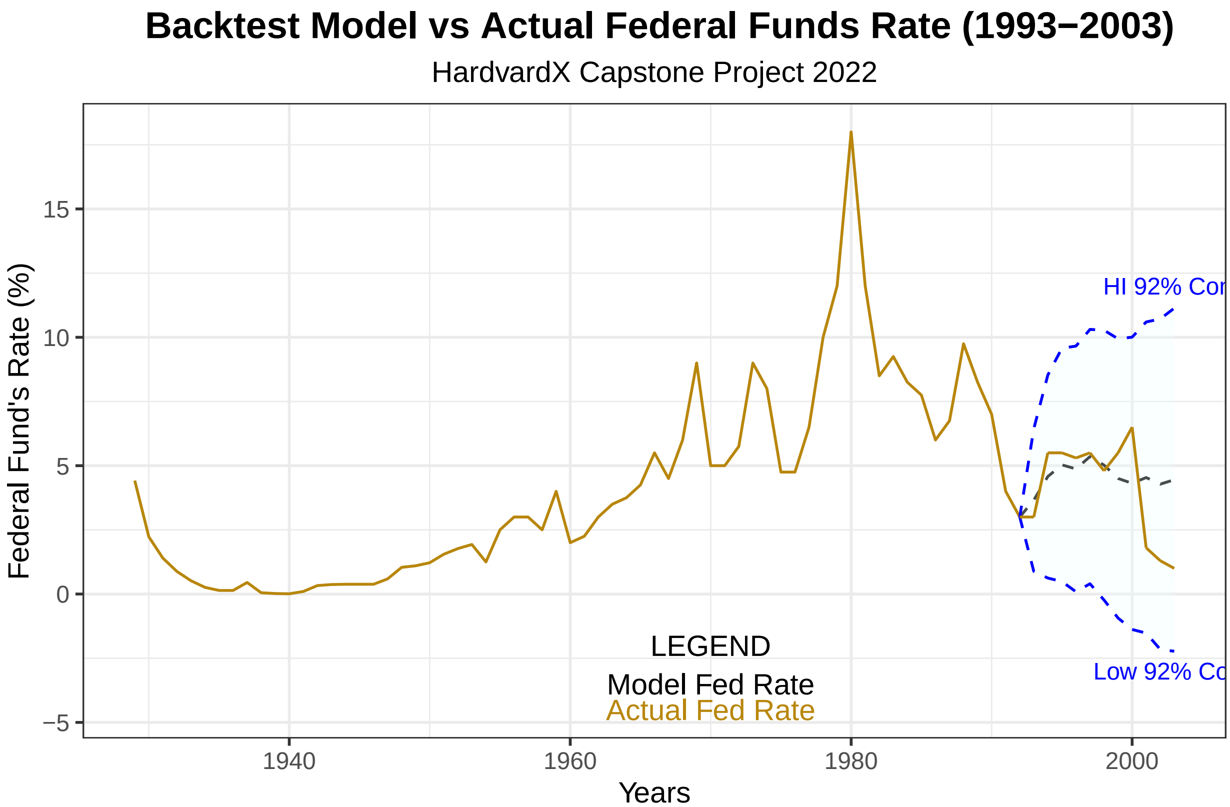
Portions of this data is from 1929−1992 and the Reference Section.

Verify if the model works by using true/false rubric with basis point tolerance .[5]

[5] ***Correction: Row 8 (2000) should be FALSE.***

Federal Reserve's Fed Fund Rate Model HardvardX Capstone Project 2022				
Year	Original_Fed_Rate	Model_Fed_Fund_Rate	Fed_Diff	Within_Tolerance
1993	3.00	3.7	0.70	**TRUE**
1994	5.50	4.6	-0.90	**TRUE**
1995	5.50	5.0	-0.50	**TRUE**
1996	5.25	4.9	-0.35	**TRUE**
1997	5.50	5.4	-0.10	**TRUE**
1998	4.75	5.0	0.25	**TRUE**
1999	5.50	4.5	-1.00	**TRUE**
2000	6.50	4.3	-2.20	**TRUE**
2001	1.75	4.5	2.75	*FALSE*
2002	1.25	4.3	3.05	*FALSE*
2003	1.00	4.5	3.50	*FALSE*
Fed Funds Rate is based on data from 1993-2003 Portions of this data is from the Reference Section				

Create chart with Backtest Model Federal Funds Rate vs Actual Federal Funds Rate (1993-2003).



As we can see the majority of the rates are within our model. In the Year 2000, the U.S. Economy experienced the Dot.com crash. That recessionary event caused the Federal Reserve to cut rates which explains why rates crashed downward.

Create Future Fed Rates Model

```
Fed_Model_True = arima(fixed = NULL, transform.pars=TRUE, Fed_Model_Time, order = c(5,1,6))
Fed_Model_Forecast = forecast(Fed_Model_True, level = c(95), h = 11)
Fed_Model_Forecast
```

```
##      Point Forecast     Lo 95    Hi 95
## 2018      1.0933376 -1.965525 4.152200
## 2019      0.6127404 -3.637121 4.862602
## 2020      0.4882505 -4.302085 5.278586
## 2021      0.3755097 -4.638446 5.389466
## 2022      0.5260666 -4.622014 5.674147
## 2023      0.7419439 -4.629621 6.113509
## 2024      0.8429574 -4.737989 6.423904
## 2025      0.9133718 -4.959976 6.786720
## 2026      0.9607977 -5.315384 7.236979
## 2027      0.8656668 -5.753861 7.485194
## 2028      0.8035516 -6.128491 7.735594
```

Verify Data

```
Box.test(Fed_Model_Forecast, lag = 1, type = "Ljung-Box")
```

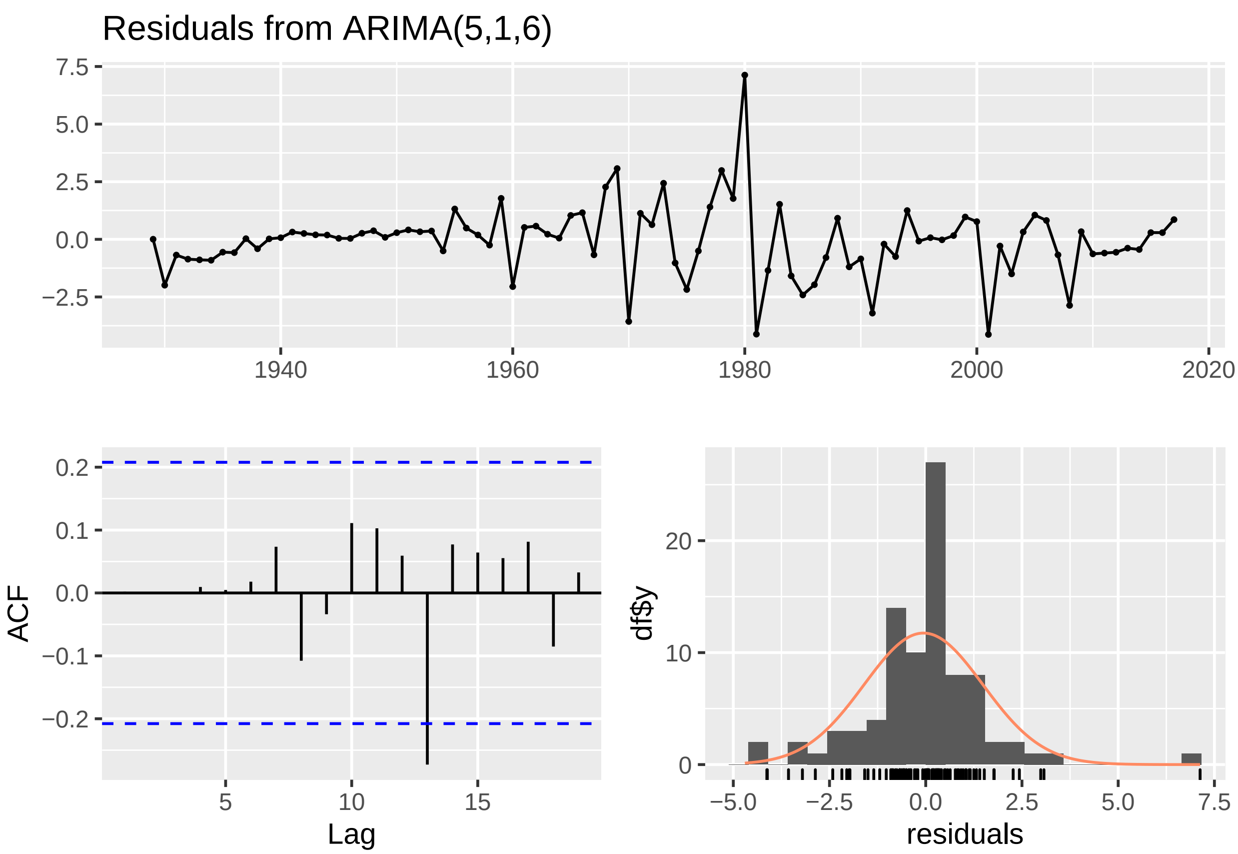
```
##
```

```
##  Box-Ljung test
##
## data:  Fed_Model_Forecast
## X-squared = 3.4737, df = 1, p-value = 0.06235
```

```
Box.test(Fed_Model_Forecast, lag = 5, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  Fed_Model_Forecast
## X-squared = 11.356, df = 5, p-value = 0.04477
```

```
checkresiduals(Fed_Model_Forecast)
```

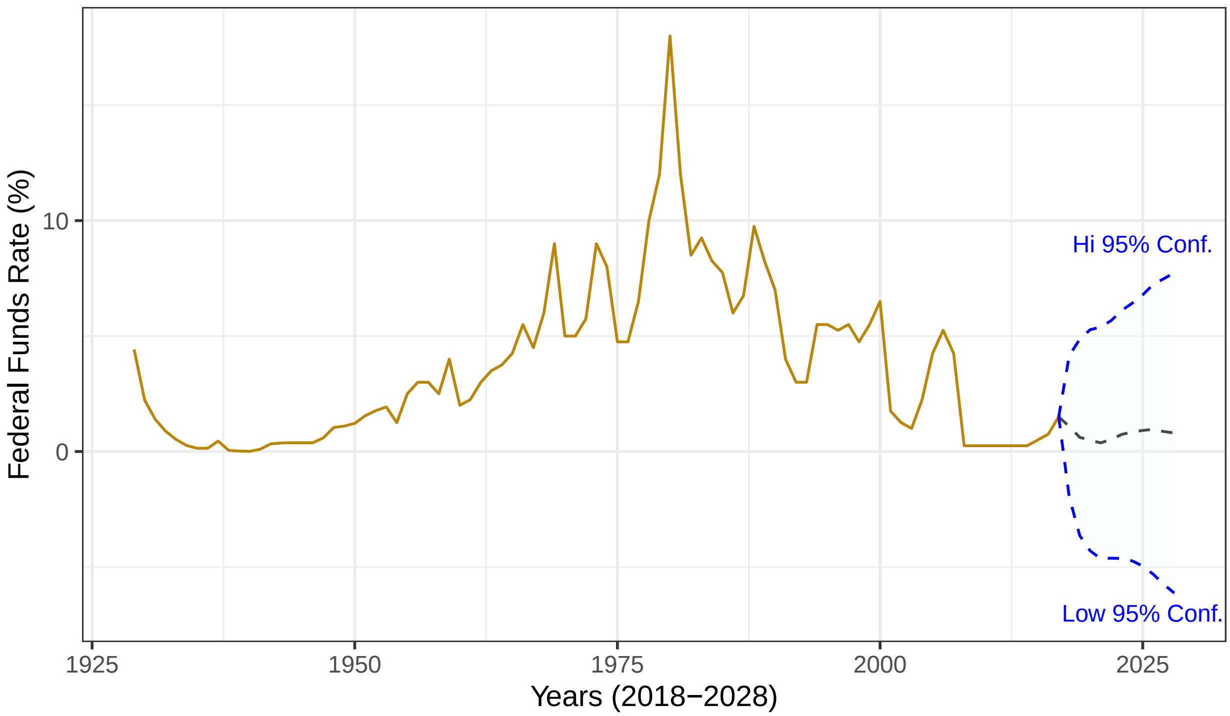



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(5,1,6)
## Q* = 13.163, df = 3, p-value = 0.004297
##
## Model df: 11.   Total lags used: 14
```

Chart Data

Future Model: Federal Reserve's Federal Fund's Rate (2018−2028)

HardvardX Capstone Project 2022



Portions of this data is from 1929−2017 and the Reference Section.

Results:

All of our goals where achieved:

1. We identified the best correlation was a positive correlation with the Federal Funds Rate and Inflation Rate YoY:
   - correlation is ***0.7710506***
   - ***R^2 of 88%*** this means that ***12%*** of variance is explained by outliers deriving from geopolitical and economic factors instead of unknown factors.

2. For our Inflation Rate YoY Machine Learning Model:
   - We identified the best Model for backtesting had an ***ARIMA (6,2,1) with a 87% confidence level.***
   - We identified the best Model for forecasting had an ***ARIMA (6,2,1) with a 90% confidence level.***
   - Backtest Code & Forecast Code:

```
Inflation_Model_1993_2003 = arima(fixed = NULL, Inflation_Model_1993_Time, order = c(6,2,1), transform.p
Backtest_Inflation_Model_Forecast_1 = forecast(Inflation_Model_1993_2003, level = c(87), h = 11)
```

```
    True_Inflation_Model = arima(fixed = NULL, Inflation_Model_Time, order = c(6,2,1), transform.pars=TR
Inflation_Model_Forecast_Update = forecast(True_Inflation_Model, level = c(90), h = 11)
```

3. For our Federal Reserve Federal Funds Rate Model:
   - We identified the best Model for backtesting had an ***ARIMA (5,1,6) with a 92% confidence level.***
   - We identified the best Model for forecasting had an ***ARIMA (5,1,6) with a 95% confidence level.***
   - Backtest Code & Forecast Code:

```
Fed_Model_1993_2003 = arima(fixed = NULL, Fed_Model_1993_Time, order = c(5,1,6),  transform.pars=TRUE)
Backtest_Fed_Model_Forecast_Update = forecast(Fed_Model_1993_2003, level = c(92), h = 11)
```

```
Fed_Model_True = arima(fixed = NULL, transform.pars=TRUE, Fed_Model_Time, order = c(5,1,6))
Fed_Model_Forecast = forecast(Fed_Model_True, level = c(95), h = 11)
```

Conclusion:

Utilizing Pearson's Correlation Coefficient (r) tool can show how different variables may have zero, inverse or positive correlation. In this case, we know that economic and geopolitical factors can drive outliers affecting Pearson's Correlation Coefficient (r) calculations by skewing the calculated results.

Geopolitical and economic outliers can also affect monetary policy, which the Federal Reserve drives. We can also conclude that the Federal Funds Rate and the Inflation Rate YoY have a positive but moderate correlation between 1929-1950 but a significantly positive correlation from 1951-2017 once the Federal Reserve utilized all the tools required to help fight inflation.

Lastly, adding all these factors into our Machine Learning model proved essential in backtesting and predicting future inflation rates YoY and Fed funds rates.