Data info: Inflation rate from 1965 to 2021
View/Download Data: csvdata

TIME SERIES ANALYSIS

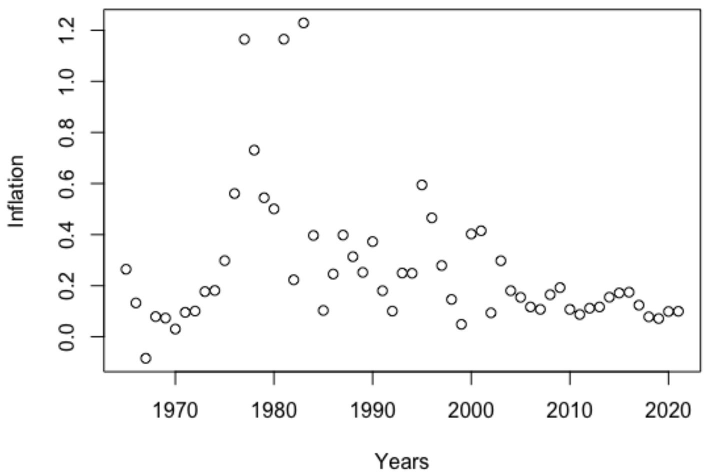
```
library(ggplot2)
library(forecast)
library(tseries)
library(TSA)
library(Kendall)
```

importing data

```
infla =read.csv("csvinflation.csv")
View(infla)
```

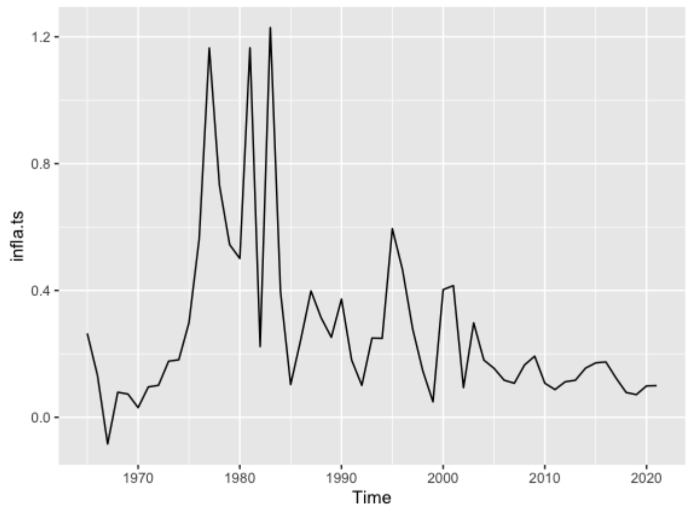
Graphical visualization of data

```
plot(infla)
```



converting original inflation data to time series data

infla.ts = ts(Inflation, start = c(1965,1), frequency = 1)
autoplot(infla.ts)



test for trend, seasonality, normality and correlation

MannKendall(infla.ts) # **Test for trend**

> **Output:** tau = -0.182, 2-sided pvalue = 0.045894
> **Conclusion:** Inflation time series has a monotonic trend since p-value is less than the significance level (0.05)

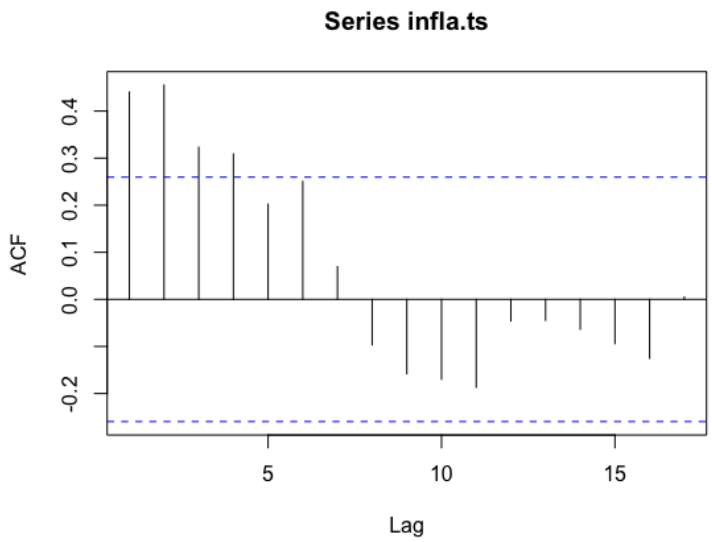
shapiro.test(infla.ts) # **Test for normality**

> **Output:**
> Shapiro-Wilk normality test

data: infla.ts
W = 0.74354, p-value = 1.243e-08

> **Conclusion:** Inflation time series is not normally distributed since p-value is less than the significance level (0.05)

acf(infla.ts) # **Autocorrelation Plot**



stationarity test

```
adf <- adf.test(infla.ts)
pp <- pp.test(infla.ts)
kpss <- kpss.test(infla.ts)
overall <- cbind(adf$p.value, pp$p.value, kpss$p.value)
colnames(overall) <- c("ADF", "PP", "KPSS")
overall
```

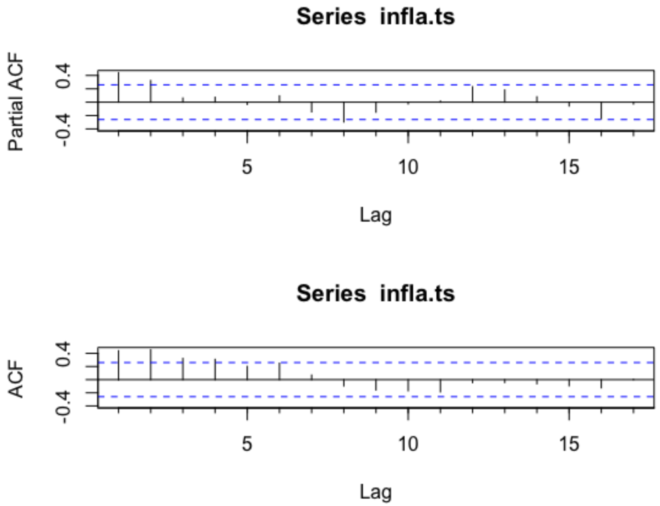
Output:

	ADF	PP	KPSS
[1,]	0.3554373	0.01	0.09596811

Conclusion: PP and KPSS says it's stationary since p-vaue < 0.05 and p-value > 0.05 respectively, while ADF says it is stationary.

Acf & Pacf for model selection (Graphical)

```
Pacf(rain.ts)
Acf(rain.ts)
par(mfrow=c(2,1))
Pacf(rain.ts)
Acf(rain.ts)
```



NB: Since we're getting different conclusions from our unit root test, we use train/test. But we first find the best or most suitable model for the two scenarios; treating it as stationary and differencing it

model selection for UNDIFFERENCED data (automated)

```
m1 = auto.arima(infla.ts, stepwise = FALSE, approximation = FALSE, stationary = TRUE)
m1
```

Output:

Series: infla.ts

ARIMA(2,0,0) with non-zero mean

Coefficients:

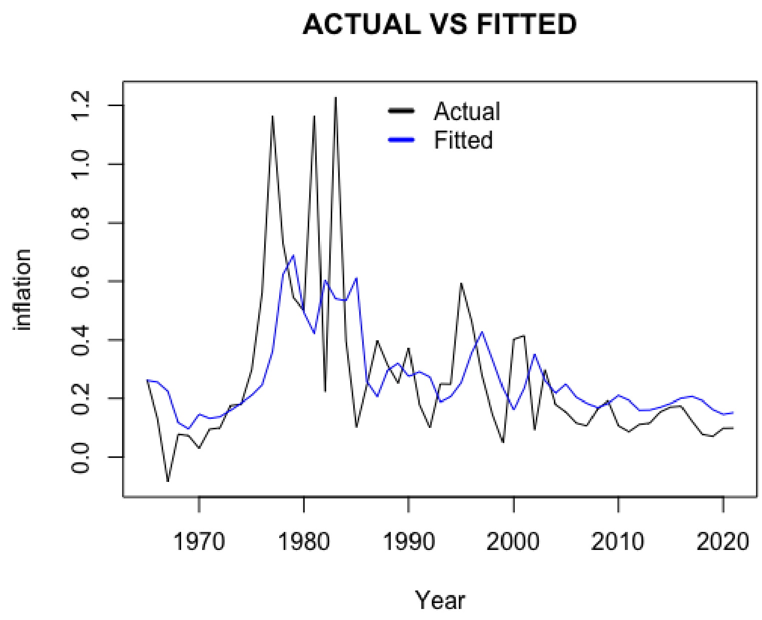
	ar1	ar2	mean
	0.2937	0.3211	0.2612
s.e.	0.1231	0.1234	0.0749

sigma^2 = 0.05414: log likelihood = 3.56

AIC=0.88   AICc=1.65   BIC=9.05

Graphical comparison of Actual and Selected model

```
plot(infla.ts, main="ACTUAL VS FITTED", xlab = "Year", ylab="inflafall Amount", lwd=1)
lines(fitted(m1), col="blue", lwd=1)
legend("top", legend = c("Actual", "Fitted"), col = c("black", "blue"), lwd = 3,seg.len = 1, bty = "n", xpd = TRUE, inset = c(-0.1, 0))
```



Diagnostic of the selected model

checkresiduals(m1)

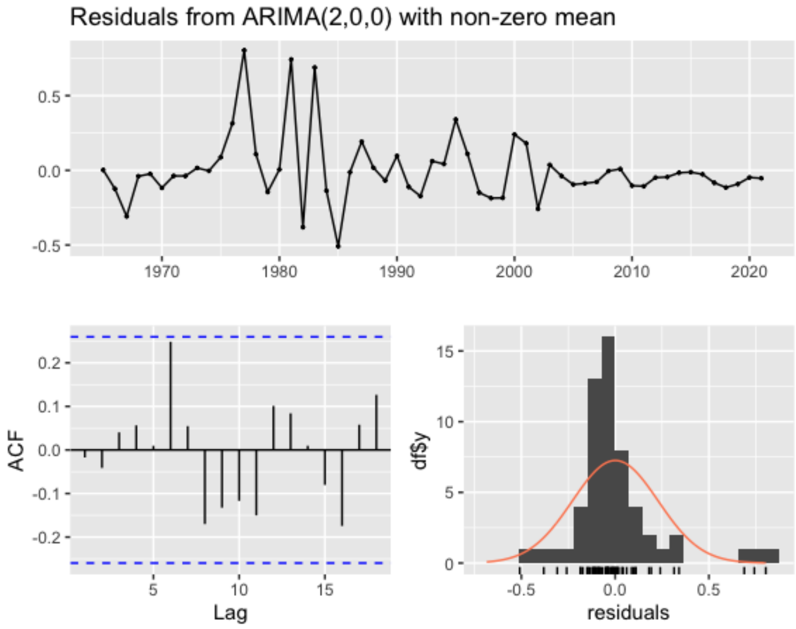
Output:

Ljung-Box test

data: Residuals from ARIMA(2,0,0) with non-zero mean
Q* = 8.8877, df = 8, p-value = 0.3519

Model df: 2. Total lags used: 10

Conclusion: We conclude that Arima(2,0,0) is a suitable model since p-value > 0.05, thus we fail to reject the null hypothesis.



Mean of the residual of Arima(2,0,0)

resid1 = m1$residuals
mean(resid1)

Output: [1] 0.0007906026

Conclusion: Arima(2,0,0) is a suitable model since the mean of the residuals is very close to 0. Thus, it agrees with the Ljung-Box test conclusion.

model selection for DIFFERENCED data

m2 = auto.arima(infla.ts, stepwise = FALSE, approximation = FALSE, stationary = FALSE)
m2

Output:

Series: infla.ts

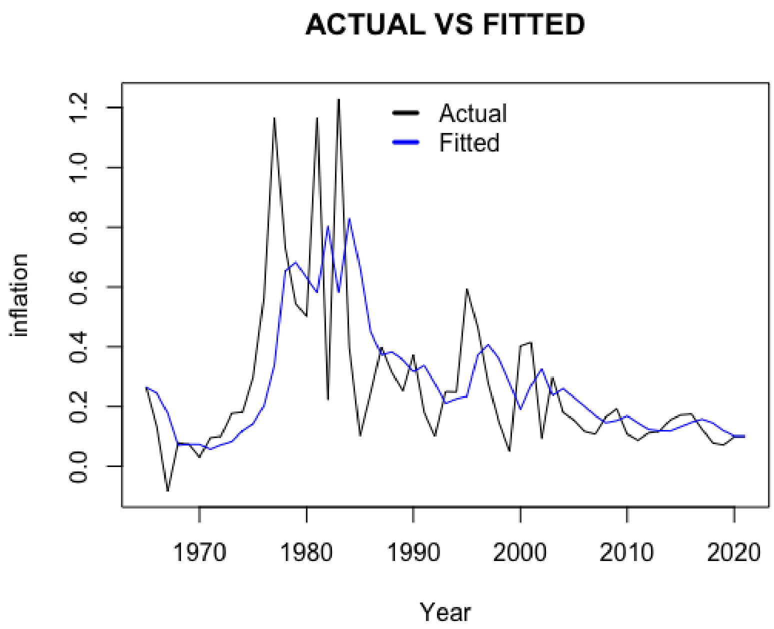
ARIMA(0,1,1)

Coefficients:
ma1
-0.6192
s.e.   0.1034

sigma^2 = 0.05658: log likelihood = 1.22
AIC=1.56   AICc=1.78   BIC=5.61

Graphical comparison of Actual and Selected model

```
plot(infla.ts, main="ACTUAL VS FITTED", xlab = "Year", ylab="inflafall Amount", lwd=1)
lines(fitted(m2), col="blue", lwd=1)
legend("top", legend = c("Actual", "Fitted"), col = c("black", "blue"), lwd = 3,seg.len = 1, bty = "n", xpd = TRUE, inset = c(-0.1, 0))
```



Diagnostic of the selected model

checkresiduals(m2)

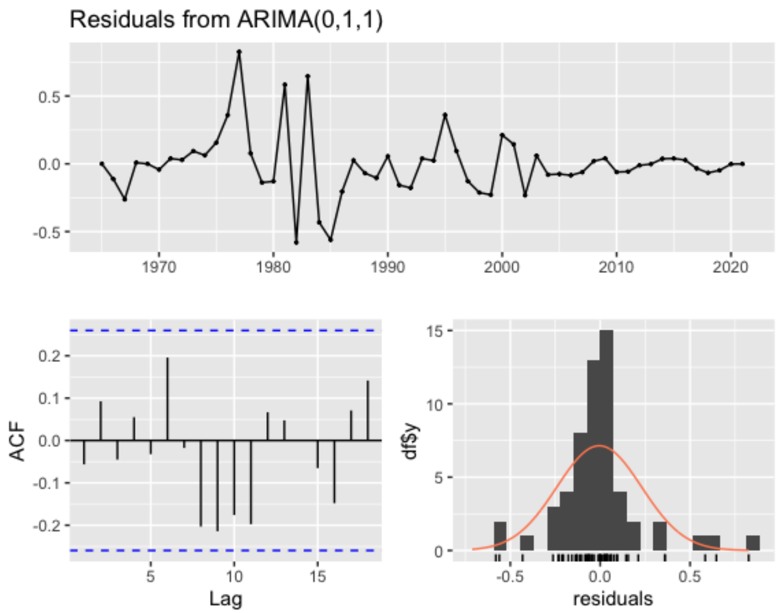
Output:

Ljung-Box test

data: Residuals from ARIMA(0,1,1)
Q* = 11.903, df = 9, p-value = 0.2188

Model df: 1. Total lags used: 10

Conclusion: We conclude that Arima(0,1,1) is a suitable model since p-value > 0.05, thus we fail to reject the null hypothesis.



Mean of the residual of Arima(0,1,1)

```
resid2 = m1$residuals
mean(resid2)
```

Output: [1] -0.005192656

Conclusion: Arima(0,1,1) is a suitable model since the mean of the residuals is very close to 0. Thus, it agrees with the Ljung-Box test conclusion.

Using Train/Test to determine the best model between Arima(2,0,0) and Arima(0,1,1)

Train Set

```
mtrain = window(infla.ts, start = 1965, end = 2011, frequency = 1)
mtrain
```

Output:

```
Time Series:
Start = 1965
End = 2011
Frequency = 1
[1]   0.2644  0.1324 -0.0842  0.0789  0.0732  0.0303  0.0956  0.1007  0.1768  0.1813  0.2982  0.5608  1.1645
[14]  0.7309  0.5444  0.5007  1.1650  0.2230  1.2287  0.3967  0.1031  0.2457  0.3982  0.3136  0.2522  0.3726
[27]  0.1803  0.1006  0.2496  0.2487  0.5946  0.4656  0.2789  0.1462  0.0487  0.4024  0.4151  0.0936  0.2977
[40]  0.1804  0.1544  0.1168  0.1073  0.1649  0.1925  0.1073  0.0873
```

Test Set

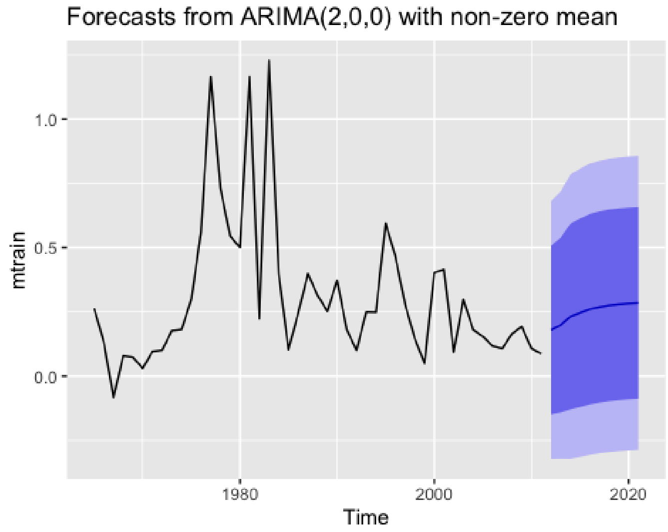
```
mtest = window(infla.ts, start = 2012, frequency = 1)
mtest
```

Output:

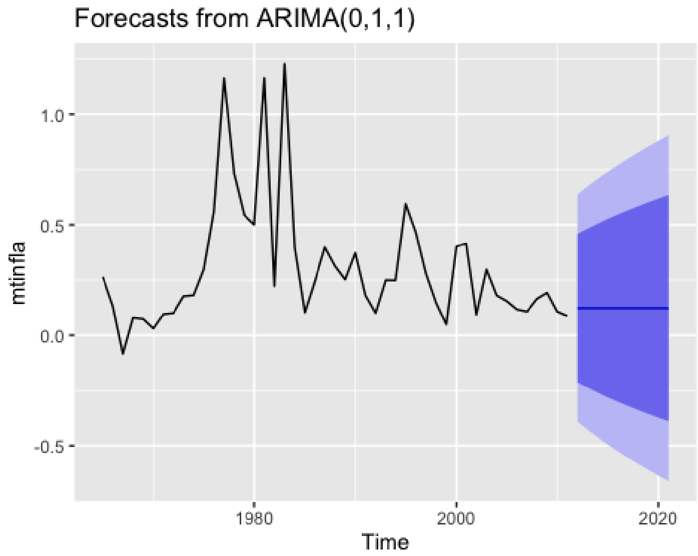
```
Time Series:
Start = 2012
End = 2021
Frequency = 1
[1] 0.1119 0.1167 0.1549 0.1715 0.1745 0.1237 0.0781 0.0714 0.0989 0.0997
```

```
tm1 = Arima(mtrain, order = c(2,0,0))
```

```
fm1 = forecast(tm1, h=length(mtest))
fm1
autoplot(fm1)
```




```
tm2 = Arima(mtrain, order = c(0,1,1))
fm2 = forecast(tm2, h=length(mtest))
autoplot(fm2)
```



Selection of best model by comparing accuracy output for the train/test

accuracy(fm1, mtest)
accuracy(fm2, mtest)

Output(tabular):

Model	RMSE	MAE	ME
Arima(2,0,0)	0.1418473	0.1305226	-0.13052265
Arima(0,1,1)	0.03460538	0.02934484	-0.00264422

Conclusion: We therefore conclude that **Arima(0,1,1)** is the better or more suitable model for our inflation data.