STAT 429 HW 01

Name: Ishita Singh , NetID: is26

```
knitr::opts_chunk$set(echo = TRUE)
library(astsa)
library(ggplot2)
library(ggfortify)
library(fpp2)
```

There are some useful R chunk options that you may use (for this entire semester):

- echo - Display code in output document (default = TRUE)
- include - Include chunk in document after running (default = TRUE)
- message - display code messages in document (default = TRUE)
- results (default = 'markup')
  - 'asis' - passthrough results
  - 'hide' - do not display results
  - 'hold' - put all results below all code
- error - Display error messages in doc (TRUE) or stop render when errors occur (FALSE) (default = FALSE)

See R markdown cheat sheet (inside canvas) for more information.

Question 1.

There are a number of seismic recordings from earthquakes and from mining explosions in `astsa` package. All of the data are in the data frame `eqexp`, but two specific recordings are in `EQ5` and `EXP6`, the fifth earthquake and the sixth explosion, respectively.

(a)

Using RStudio, read the description of `EQ5` data and `EXP6` of `astsa` package using the help function and answer the question. What is the main difference between `EQ5` and `EXP6`?

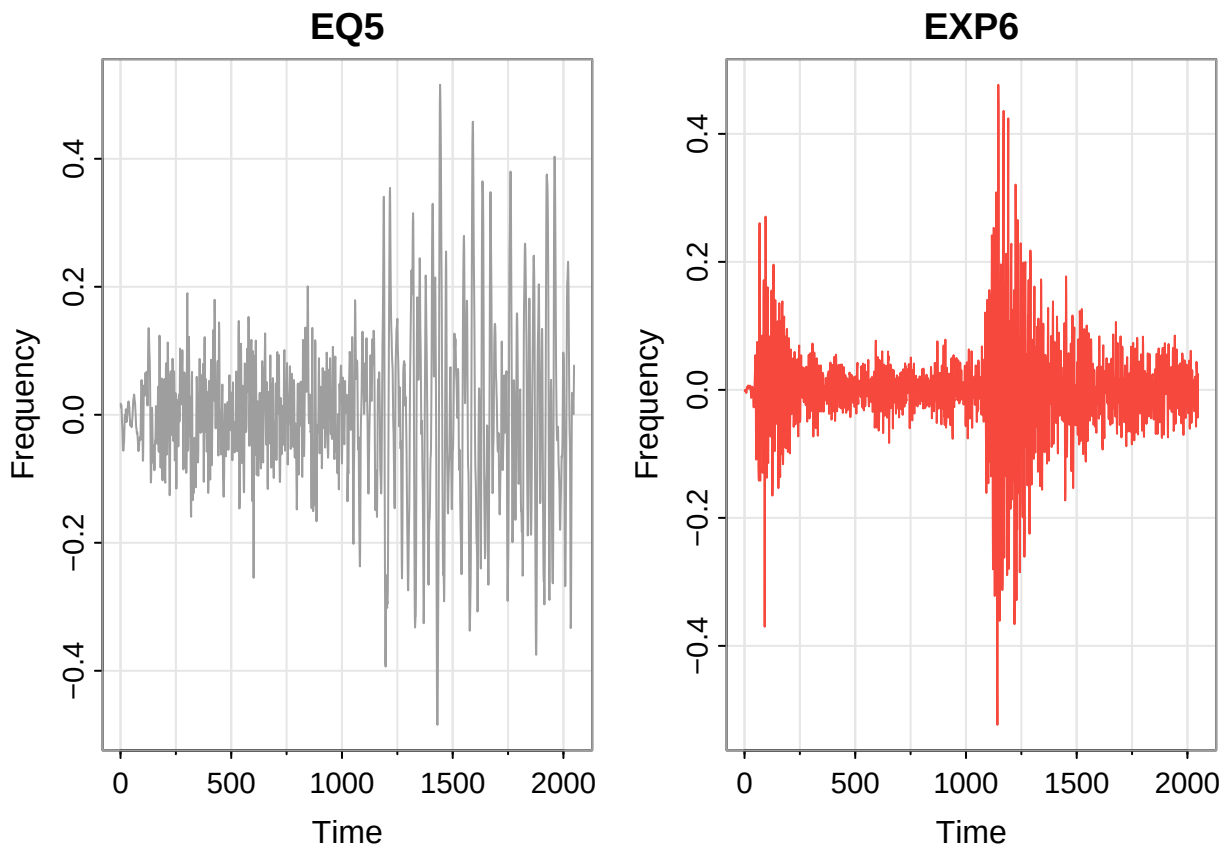
```
#description of EQ5 and EXP6
?EQ5
?EXP6
```

- The major difference is that EQ5 seismic trace from an earthquake and EXP6 is seismic trace from an explosion. Both have 2048 points divided into two phases: primary wave (t=1-1024) and shear wave (t=1025-2048).

(b)

Plot the two series separately in a plot with two columns and one row.

```
data("EQ5")
data("EXP6")

par(mfrow = c(1, 2))

#plot for EQ5
tsplot(EQ5, type = "l", col = 8, lwd = 1, main = "EQ5", xlab = "Time", ylab = "Frequency")

#plot for EXP6
tsplot(EXP6, type = "l", col = 2, lwd = 1, main = "EXP6", xlab = "Time", ylab = "Frequency")
```



(c)

Looking at the plot from (b), in what way are the earthquake and explosion series different? (Any reasonable answer is ok here.)

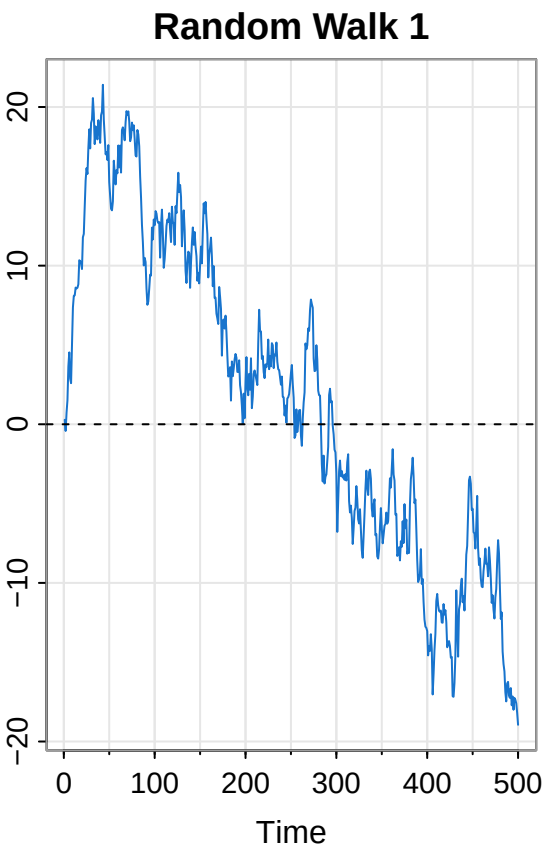
- The distribution of the time series plot is very different. EQ5 is smooth and has less spikes whereas EXP6 has sudden spikes.

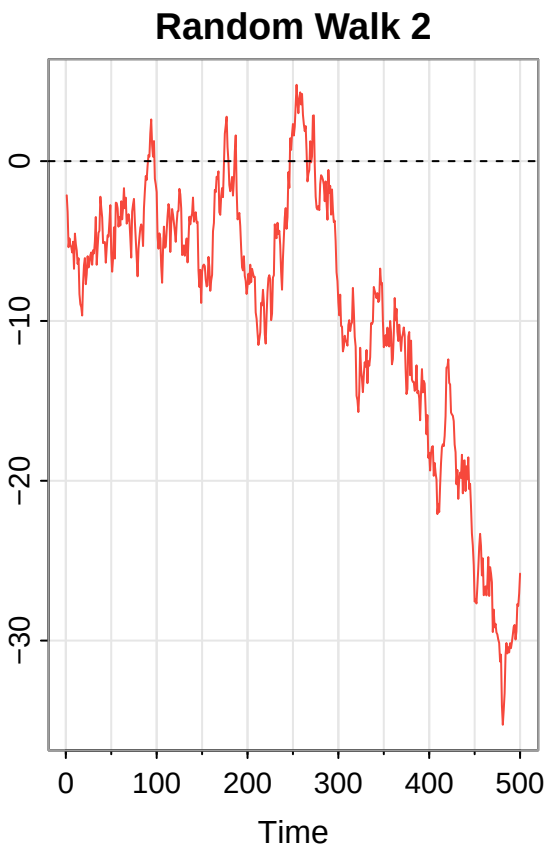
Question 2.

(a)

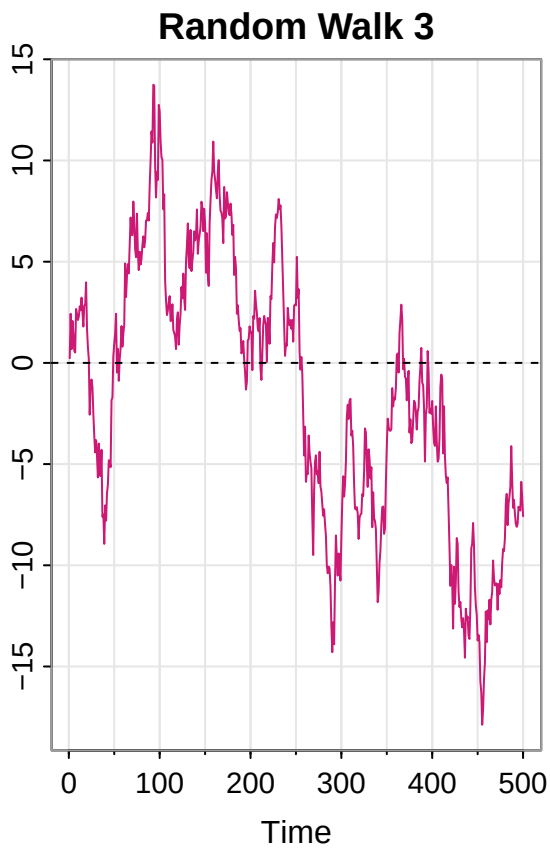
Generate three series that are random walks of length n = 500 without drift ($\delta = 0$) and $\sigma_w = 1.3$. Use random seed 4. Hint: Random walks are different from just randomly generated numbers.

```
set.seed(4)

n <- 500
sigma <- 1.3

#white noise
w1 <- rnorm(n, sd = sigma)
w2 <- rnorm(n, sd = sigma)
w3 <- rnorm(n, sd = sigma)

#random walks
x1 <- cumsum(w1)
x2 <- cumsum(w2)
x3 <- cumsum(w3)

par(mfrow = c(1,2))
tsplot(x1, col = 4, main = "Random Walk 1", ylab = "")
abline(h = 0, lty = 2)
tsplot(x2, col = 2, main = "Random Walk 2", ylab = "")
abline(h = 0, lty = 2)
```






```
tsplot(x3, col = 6, main = "Random Walk 3", ylab = "")
abline(h = 0, lty = 2)
```

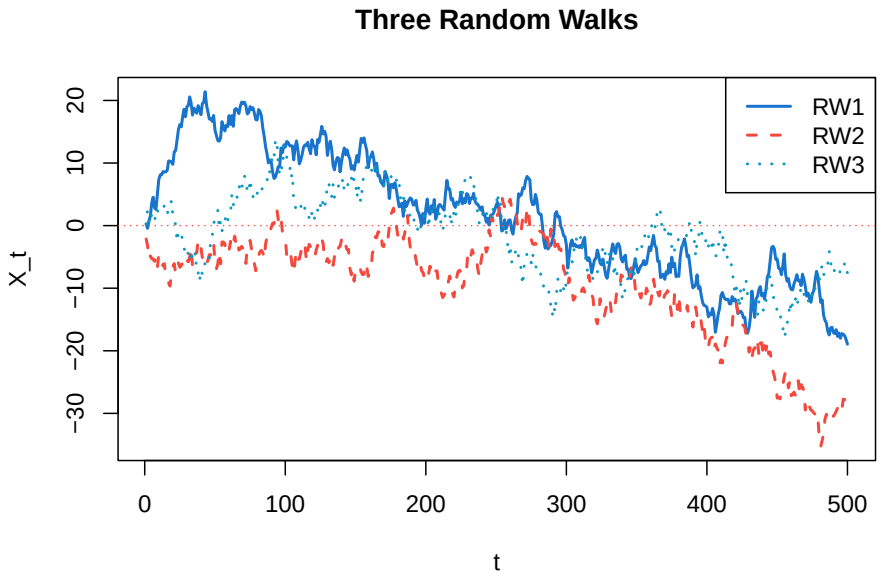


(b)

Use `plot` and `lines` to plot all series in the same graph. Make sure all series are visible, you may want to change y axis range of your plot. Make the plot width to be 75% of the page width. Use different line types and colors for different series.

Hint: https://bookdown.org/yihui/rmarkdown-cookbook/figure-size.html

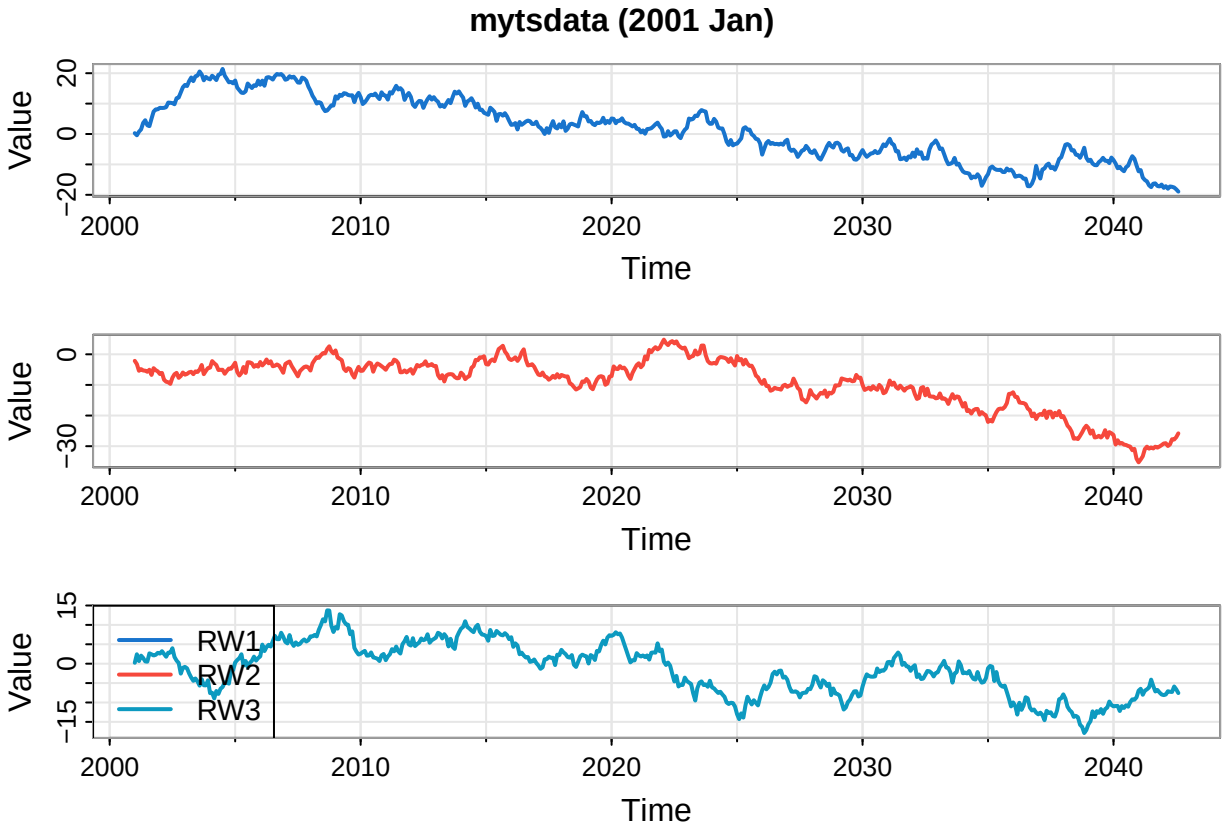
```
set.seed(4)

n <- 500
sigma <- 1.3

w1 <- cumsum(rnorm(n, sd = sigma))
w2 <- cumsum(rnorm(n, sd = sigma))
w3 <- cumsum(rnorm(n, sd = sigma))

# y-axis range
yr <- range(c(w1, w2, w3))

plot(w1, type = "l", ylim = yr, col = 4, lty = 1, lwd = 2, xlab = "t", ylab = "X_t", main = "Three Rando

lines(w2, col = 2,   lty = 2, lwd = 2)
lines(w3, col = 5, lty = 3, lwd = 2)
abline(h = 0, lty = 3, col = 10)
legend("topright", c("RW1","RW2","RW3"), col = c(4,2,5), lty = c(1,2,3), lwd = 2)
```



(c)

Make the data as time series object. Create a monthly time series object that starts in 2001 January (see textbook (TSA4) R.4, Time Series Primer) containing all three series in (a), and name it as `mytsdata`. Then, plot them using `tsplot(mytsdata)`.

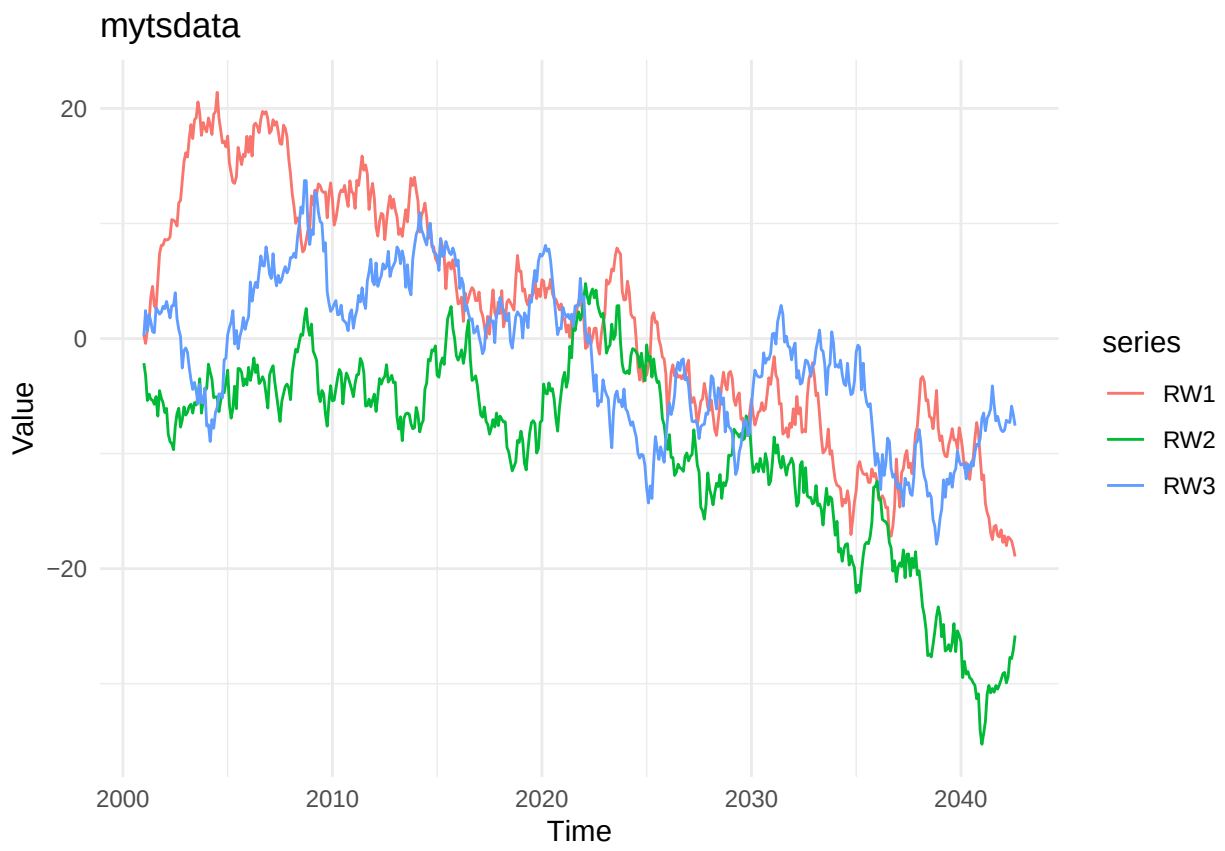
```
#converting the data to a time series data
mytsdata <- ts(cbind(RW1 = w1, RW2 = w2, RW3 = w3), start = c(2001, 1), frequency = 12)

tsplot(mytsdata, col = c(4, 2, 5), lwd = 2, main = "mytsdata (2001 Jan)", ylab = "Value")
legend("topleft", c("RW1","RW2","RW3"), col = c(4,2,5), lwd = 2)
```



(d)

Install (if necessary) `ggfortify` and `ggplot2` packages and load them. Use `autoplot` function to plot `mytsdata`.

```
library(ggplot2)
library(ggfortify)

autoplot(mytsdata) + labs(title = "mytsdata", x = "Time", y = "Value") + theme_minimal()
```


mytsdata
20
0
−20
Value
2000
2010
2020
2030
2040
Time
series
RW1
RW2
RW3

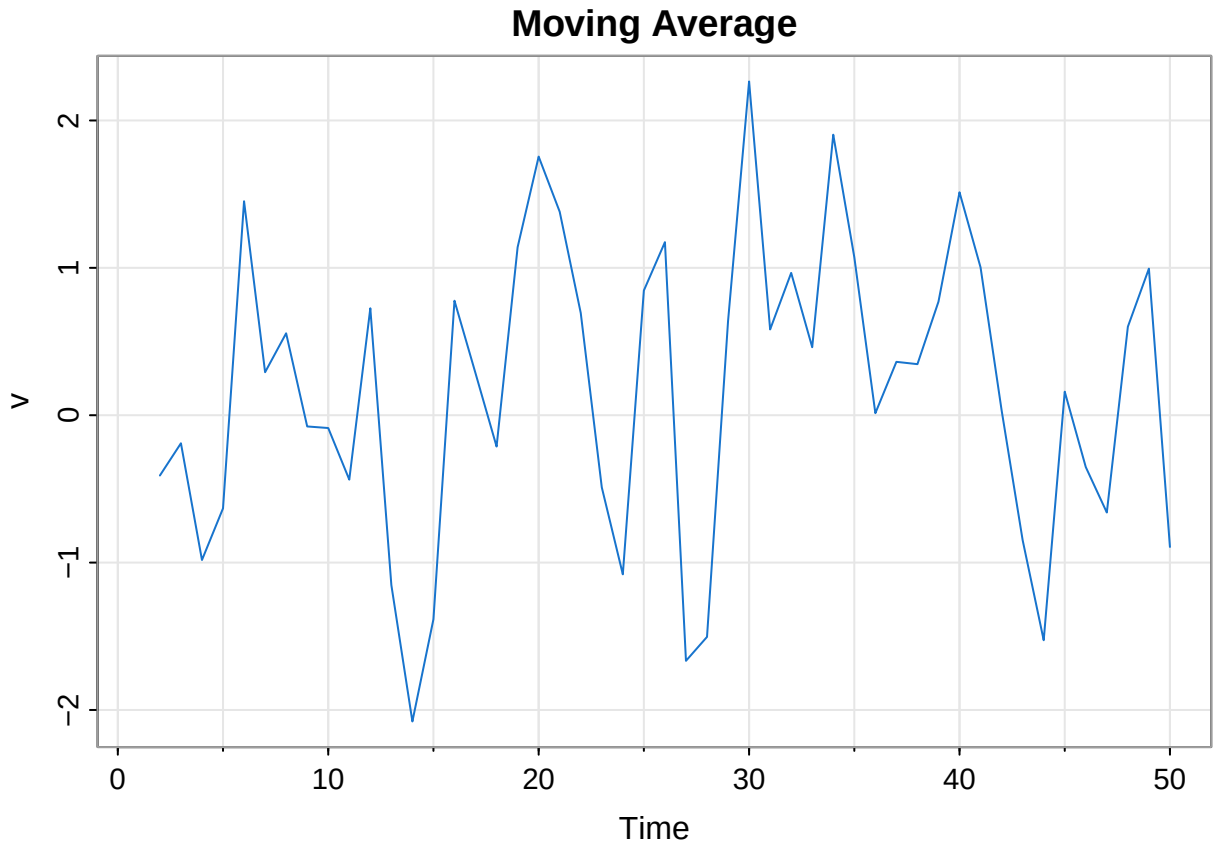
Question 3.

(a)

Generate a moving average time series X_t of length n = 50: $X_t = W_t + 0.3\,W_{t-1}$, with $\sigma_w = 1$.

Note: You can use filter function, or other methods.

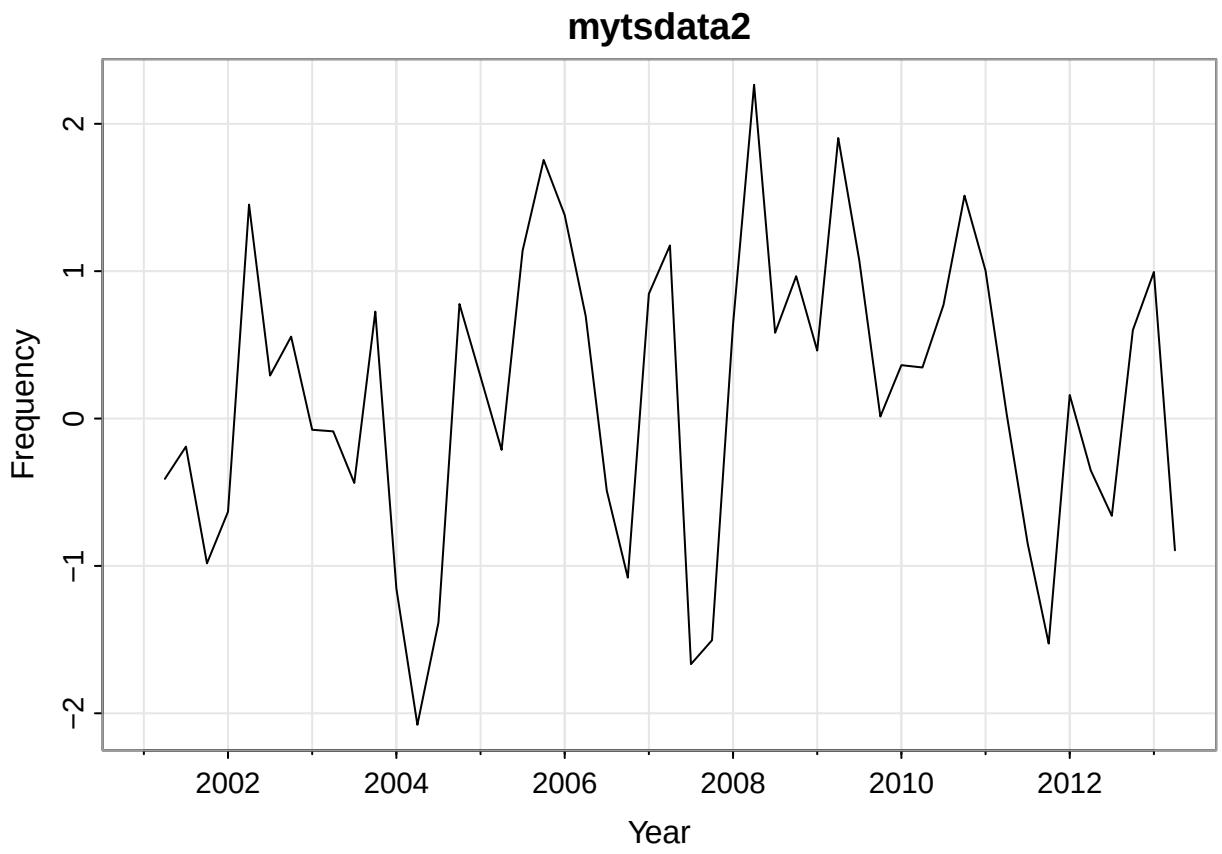
```
n <- 50
w <- rnorm(n)
#X_t = W_t + 0.3 W_{t-1}
v <- filter(w, filter = c(1, 0.3), sides = 1)
tsplot(v, col = 4, main = "Moving Average")
```



(b)

Make the data as time series object. Create a quarterly time series object that starts in 2001 first quarter (see textbook (TSA4) R.4, Time Series Primer), and name it as `mytsdata2`. Then, plot them using `tsplot(mytsdata2)`.

```
mytsdata2 <- ts(v, start = c(2001, 1), frequency=4)
tsplot(mytsdata2, main = "mytsdata2", xlab = "Year", ylab = "Frequency")
```


mytsdata2
Frequency
2
1
0
−1
−2
2002
2004
2006
2008
2010
2012
Year

Question 4.

Find a time series data that interests you. Here are some resources that you can use. (They are examples, you can choose other series)

- ESRL (Earth system research Lab)
  - https://www.esrl.noaa.gov/gmd/dv/data/
- US health data
  - https://www.cdc.gov/nchs/index.htm
- Commodities data
  - https://www.indexmundi.com/commodities/
- NASA
  - https://data.giss.nasa.gov

After choosing the data, save the data and read the data into `R`.

- Print out the first 5 observations (do not print out the entire observation data).

```
df <- read.csv("Dataset.csv")
head(df,5)
```

```
##   Data.As.Of Start.Date   End.Date    Group Year Month         State       Sex
## 1 09/27/2023 01/01/2020 09/23/2023 By Total   NA    NA United States All Sexes
## 2 09/27/2023 01/01/2020 09/23/2023 By Total   NA    NA United States All Sexes
## 3 09/27/2023 01/01/2020 09/23/2023 By Total   NA    NA United States All Sexes
## 4 09/27/2023 01/01/2020 09/23/2023 By Total   NA    NA United States All Sexes
## 5 09/27/2023 01/01/2020 09/23/2023 By Total   NA    NA United States All Sexes
##      Age.Group COVID.19.Deaths Total.Deaths Pneumonia.Deaths
## 1     All Ages         1146774     12303399          1162844
## 2 Under 1 year             519        73213             1056
## 3   0-17 years            1696       130970             2961
## 4    1-4 years             285        14299              692
## 5   5-14 years             509        22008              818
##   Pneumonia.and.COVID.19.Deaths Influenza.Deaths
## 1                        569264            22229
## 2                            95               64
## 3                           424              509
## 4                            66              177
## 5                           143              219
##   Pneumonia..Influenza..or.COVID.19.Deaths Footnote
## 1                                 1760095
## 2                                    1541
## 3                                    4716
## 4                                    1079
## 5                                    1390
```

- Provide a time series plot.

```r
plot(df$COVID.19.Deaths, type = "l", lwd = 2, col = 5, xlab = "X", ylab = "Weekly deaths", main = "COVI
```