Homework5

Megan Jasek, Rohan Thakur, Charles Kekeh

Monday, March 7, 2016

Question1

Part 1

The package has already been installed.

Part 2

```
# Loading the astsa library
library(astsa)
```

```
## Warning: package 'astsa' was built under R version 3.2.3
```

Part 3

```
# Examining series structures
str(EQ5)
```

```
##  Time-Series [1:2048] from 1 to 2048: 0.01749 0.01139 0.01512 0.01477 0.00651 ...
```

```
str(flu)
```

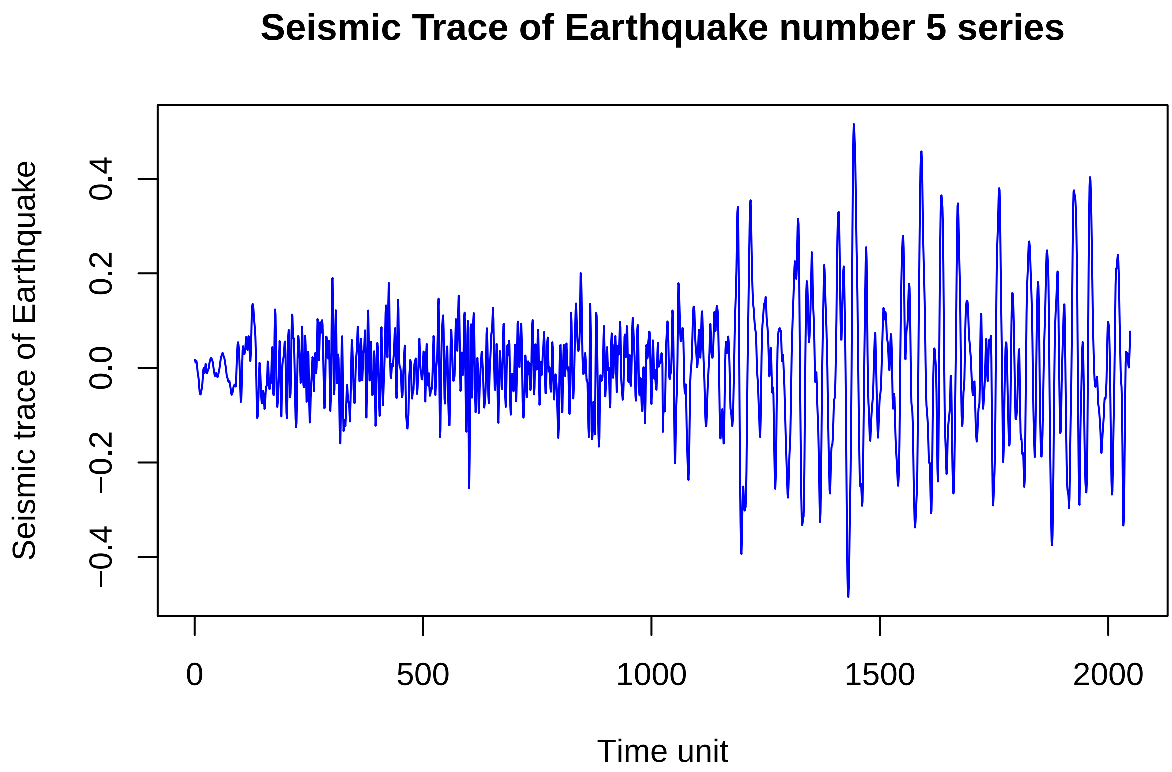
```
##  Time-Series [1:132] from 1968 to 1979: 0.811 0.446 0.342 0.277 0.248 ...
```

```
str(gas)
```

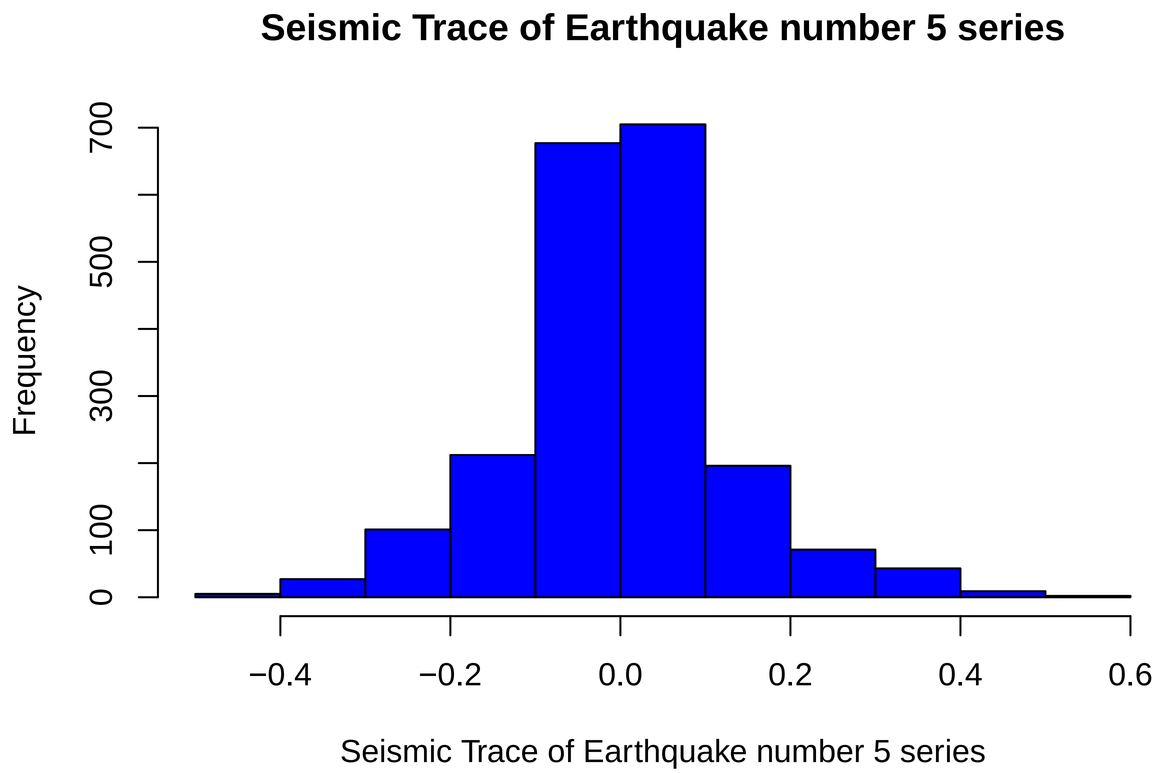
```
##  Time-Series [1:545] from 2000 to 2010: 70.6 71 68.5 65.1 67.9 ...
```

Part 4

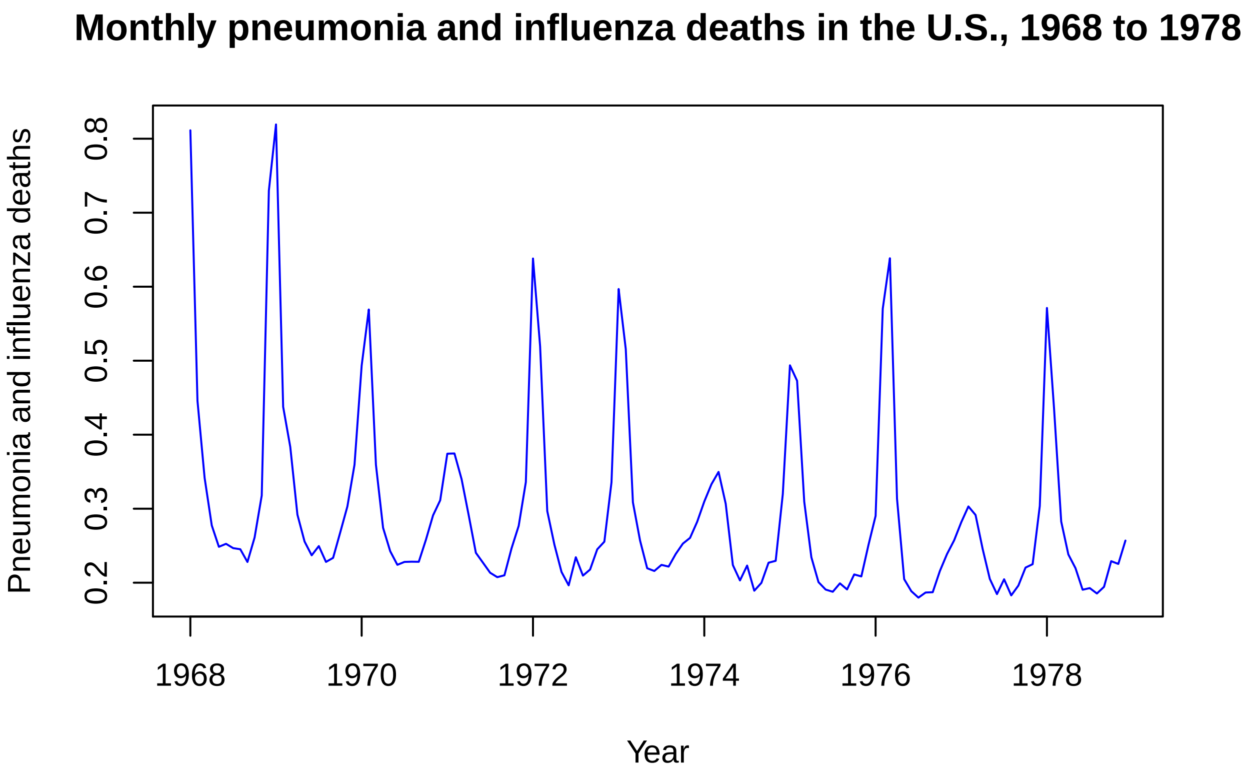
```
# Plotting for EQ5
plot(EQ5, main = "Seismic Trace of Earthquake number 5 series", ylab = "Seismic trace of Earthquake",
    xlab = "Time unit", col = "blue")
```



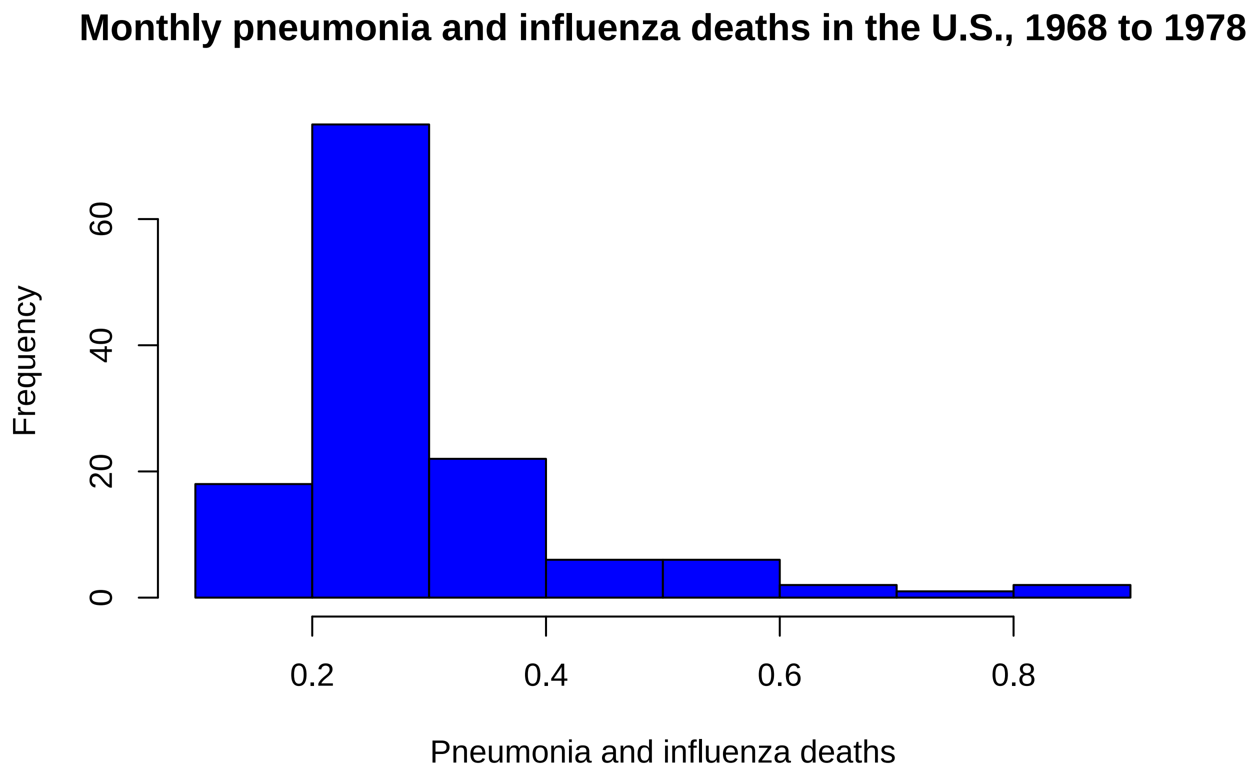

```r
hist(EQ5, main = "Seismic Trace of Earthquake number 5 series", col = "blue",
    xlab = "Seismic Trace of Earthquake number 5 series", ylab = "Frequency")
```



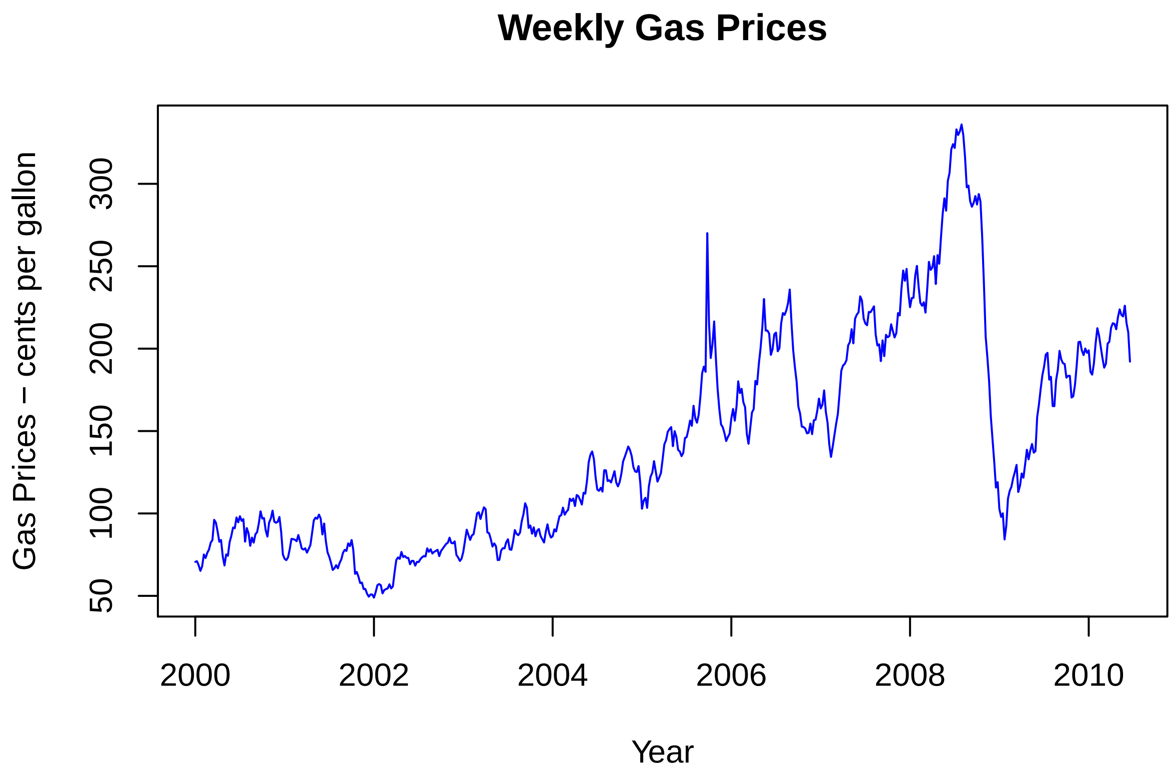

```r
# Plotting for flu
plot(flu, main = "Monthly pneumonia and influenza deaths in the U.S., 1968 to 1978",
    ylab = "Pneumonia and influenza deaths", xlab = "Year", col = "blue")
```



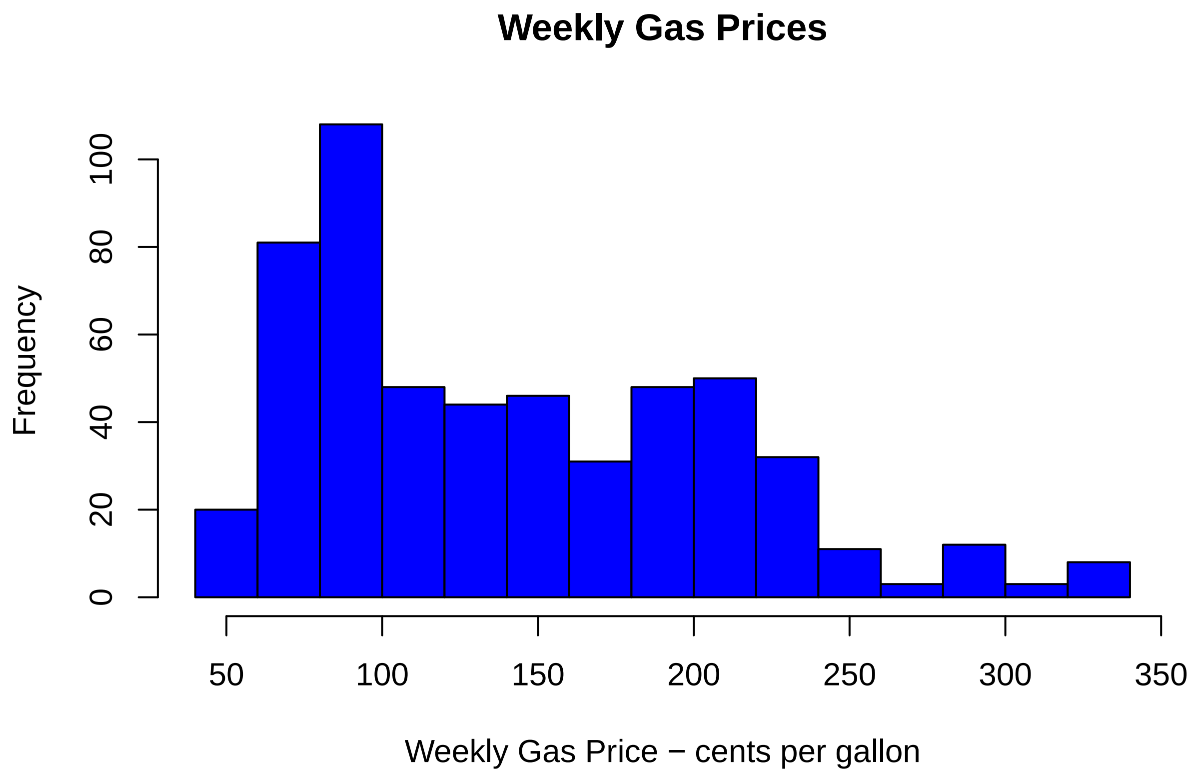

```r
hist(flu, main = "Monthly pneumonia and influenza deaths in the U.S., 1968 to 1978",
    col = "blue", xlab = "Pneumonia and influenza deaths", ylab = "Frequency")
```




```r
# Plotting for gas
plot(gas, main = "Weekly Gas Prices", ylab = "Gas Prices - cents per gallon",
    xlab = "Year", col = "blue")
```




```r
hist(gas, main = "Weekly Gas Prices", col = "blue", xlab = "Weekly Gas Price - cents per gallon",
    ylab = "Frequency")
```



Part 5

EQ5 - Mean of the series stable around 0. Amplitude/Volatility of the series increases sharply around time unit 1000, and fluctuates. Volatility is about twice as much in the 2nd half of the series as in the first.

flu - Seems to be a downward trend from the 60's leading into the 70's. After this, the trend is less pronounced downwards and possibly flat with a steady seasonal variation (during flu season, winter months) with some years worse than others. The spikes of the seasonal variation seem to be trending downward throughout the time series. We can speculate that the downward trends are due to improvements in health programs over the years.

gas - The weekly gas price forms a downward trend from 200-2002, upward trend from 2002 - 2009, a sharp decrease in 2009, and a gradual increase from 2009 onwards. Regular variation around the trend with variation spikes between 2005 and 2006 and a variation dip in 2007. Given the product represented (gasoline), we speculate that the prices in the series followed the evolution of the global prices of oil with increases corresponding to supply/demand mechanisms and world events of importance. Thus the seemingly random nature of the fluctuation. Where the years 2000-2005 showed relatively little fluctuation, years 2005 to 2010 show much more ample fluctuations of the prices.

Question 2

Example 1 -> Number of users adopting new product feature at social media startup (daily data for 1 year). The graph follows an upward trend for three months, corresponding to the initial promotion of the feature, followed by a flattening of the line.

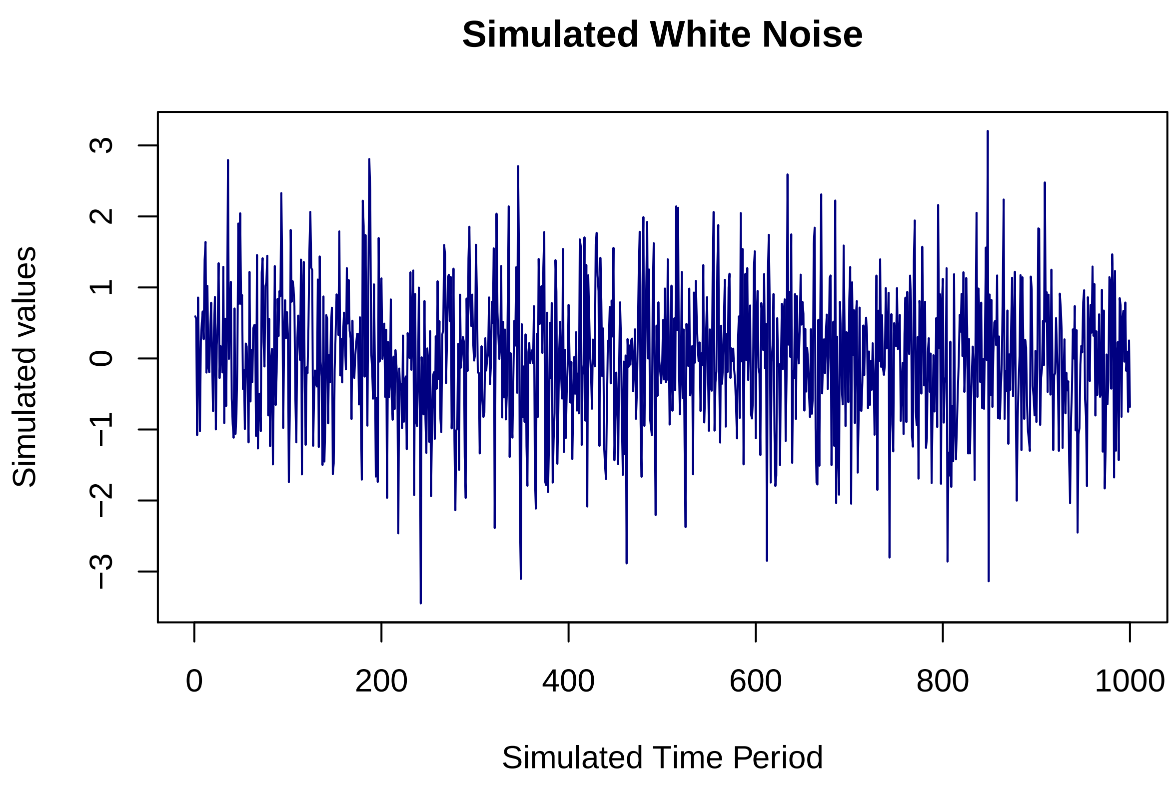
Example 2-> The daily stock price of TSLA in the past 3 months. The price followed a steady downward trend until the past few weeks when it has made an upward trend.

Example 3 -> One example of time series I have come across is that of hardware failures per day in a hosted cloud environment. The series were daily series of failure counts for a number of hardware components (cpu, disk, network), in a datacenter built with low cost utility hardware.

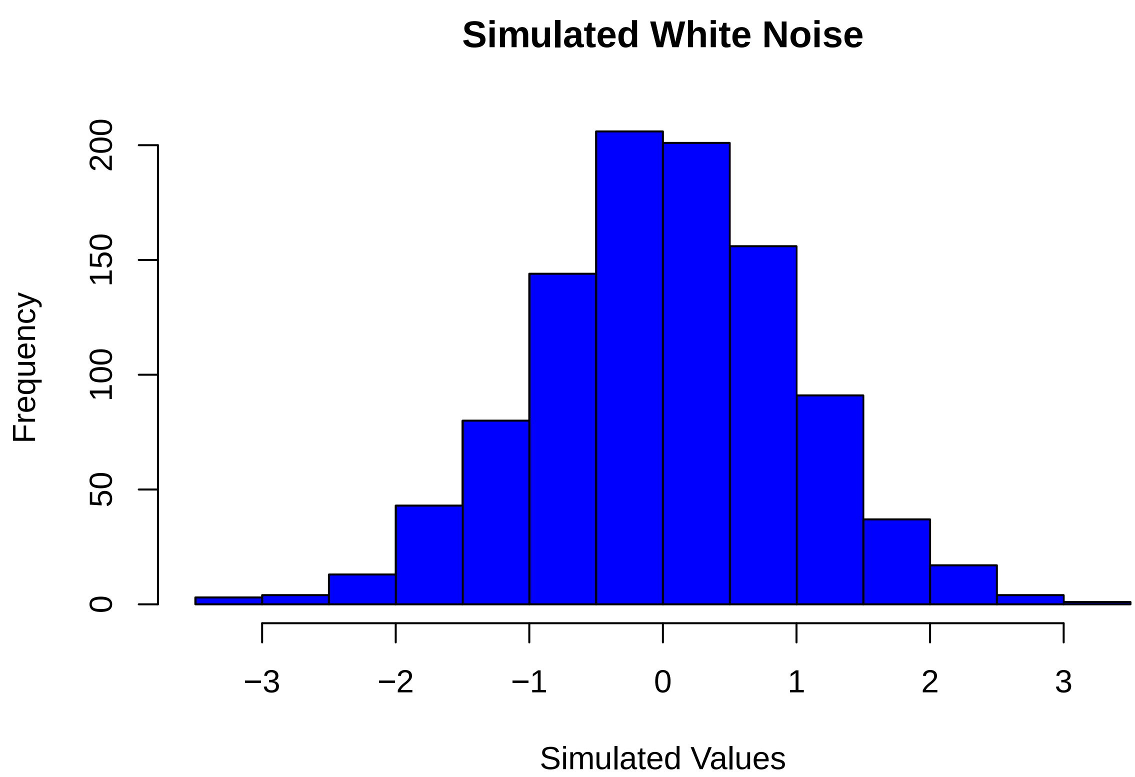
Question 3

```r
# Simulating white noise
white.noise = rnorm(1000, 0, 1)

# Plotting timeplot, histogram
plot.ts(white.noise, main = "Simulated White Noise", col = "navy", ylab = "Simulated values",
    xlab = "Simulated Time Period")
```



```r
hist(white.noise, main = "Simulated White Noise", col = "blue", xlab = "Simulated Values")
```



Question 4

- The AR(1) model with 0.9 has a larger value range than the AR(1) model with 0.2. The parameters used to construct the series are an explanation for the difference in range.
- The AR(1) model with 0.2 fluctuates more in a smaller time period, so its plot appears more dense and where as the AR(1) model with 0.9 appears more sparse. When the AR(1) model with 0.2 goes up, it stays up for longer, when the AR(1) model with 0.9 goes down it stays down for longer.

```r
# Simulating First series
white.noise.series.1 = rnorm(1000, 0, 1)
series.1 <- white.noise.series.1
for (t in 2:length(white.noise.series.1)) {
    series.1[t] <- 0.9 * series.1[t - 1] + white.noise.series.1[t]
}

# Simulating Second Series
white.noise.series.2 = rnorm(1000, 0, 1)
series.2 <- white.noise.series.2
for (t in 2:length(white.noise.series.2)) {
    series.2[t] <- 0.2 * series.2[t - 1] + white.noise.series.2[t]
}
```

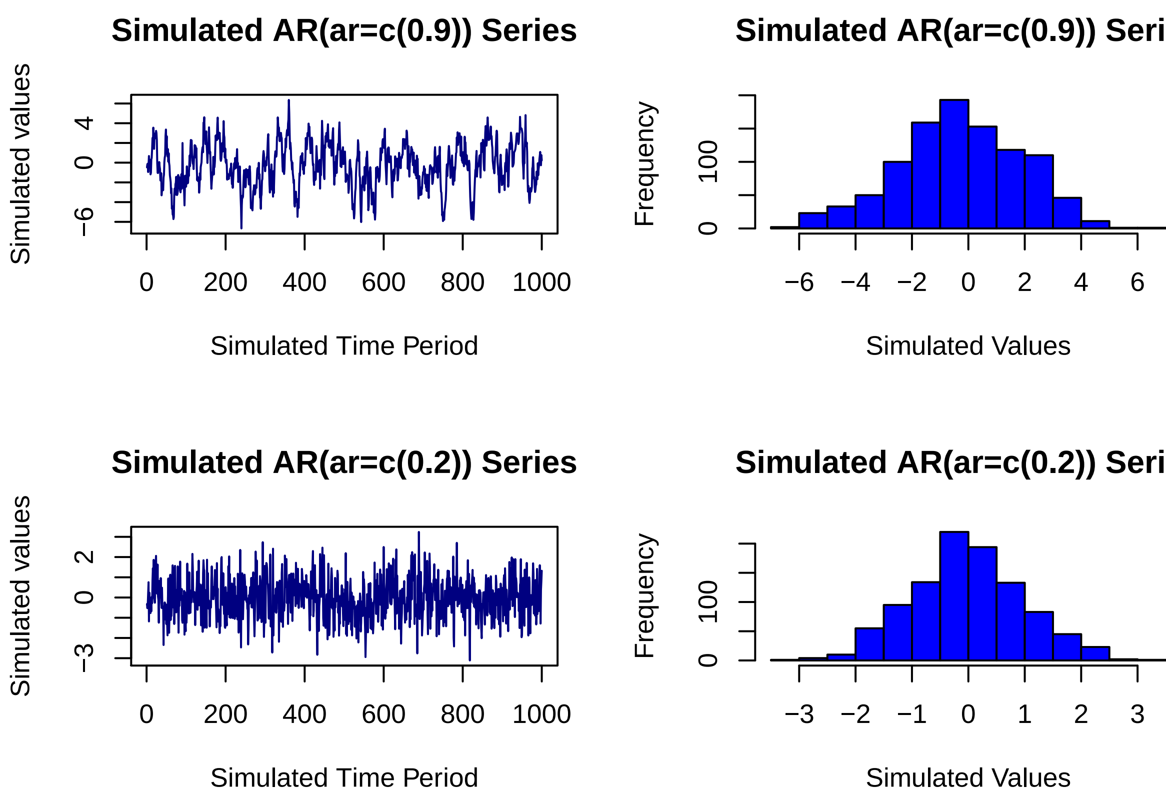
```r
summary(series.1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -6.6660 -1.6300 -0.2827 -0.2862  1.2930  6.3580
```

```r
summary(series.2)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -3.11100 -0.66990 -0.03115 -0.01494  0.64520  3.24200
```

```r
# Plotting Graphs
par(mfrow = c(2, 2))
plot.ts(series.1, main = "Simulated AR(ar=c(0.9)) Series", col = "navy",
    ylab = "Simulated values", xlab = "Simulated Time Period")
hist(series.1, main = "Simulated AR(ar=c(0.9)) Series", col = "blue", xlab = "Simulated Values")
plot.ts(series.2, main = "Simulated AR(ar=c(0.2)) Series", col = "navy",
    ylab = "Simulated values", xlab = "Simulated Time Period")
hist(series.2, main = "Simulated AR(ar=c(0.2)) Series", col = "blue", xlab = "Simulated Values")
```



Question 5

- **The deterministic series:** This series has a mean of 260.25. It is a straight line trending upwards with a slope of 0.5. It has zero variation.

- **Random walk without drift:** This trends flat from 0 to 600 and then has a slight upward trend. There is some variation that is random.
- **Random walk with 0.5 drift:** This series trends upward with a slope about 0.5. There is some variation that is random.

```
white.noise.series.new = rnorm(1000, 0, 1)
r.walk = cumsum(white.noise.series.new)

# Random walk with drift = 0.5
noise.drift = 0.5 + white.noise.series.new
r.walk.drift = cumsum(noise.drift)

mean(0.5 * (1:length(r.walk.drift)) + 10)
```

```
## [1] 260.25
```

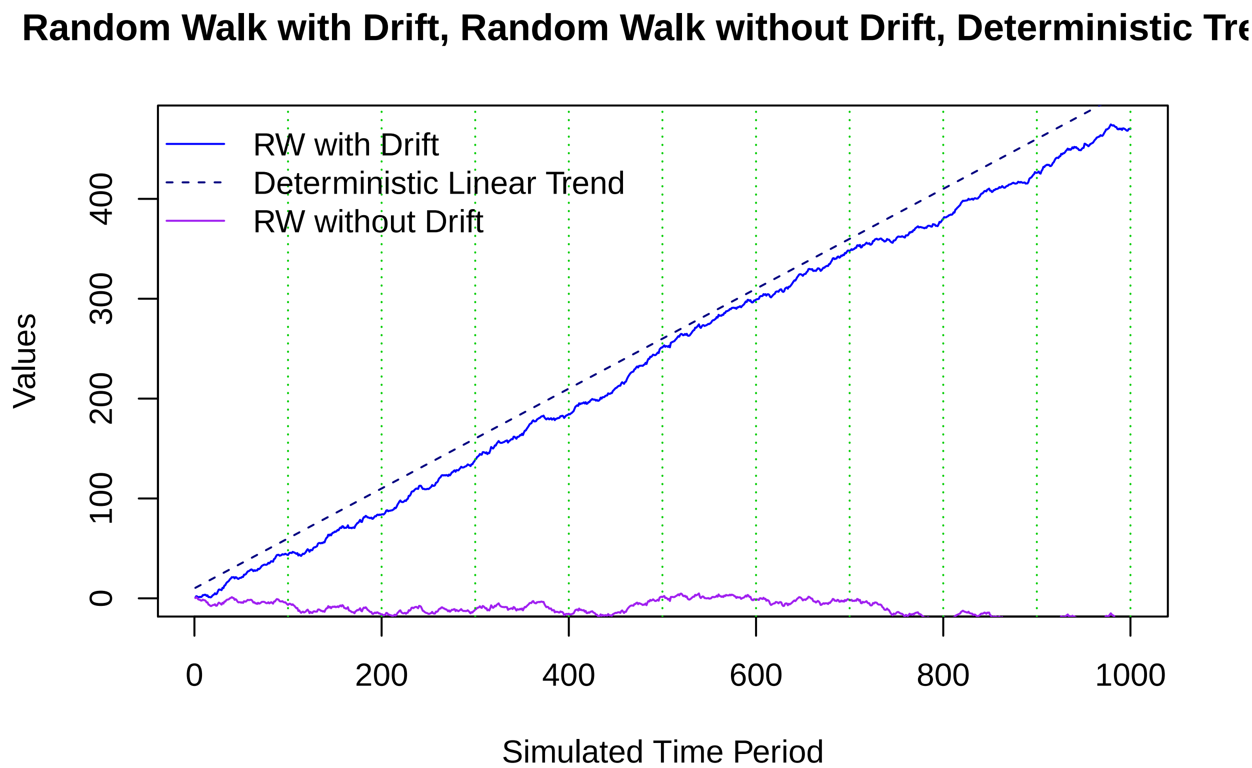
```
mean(r.walk)
```

```
## [1] -10.27767
```

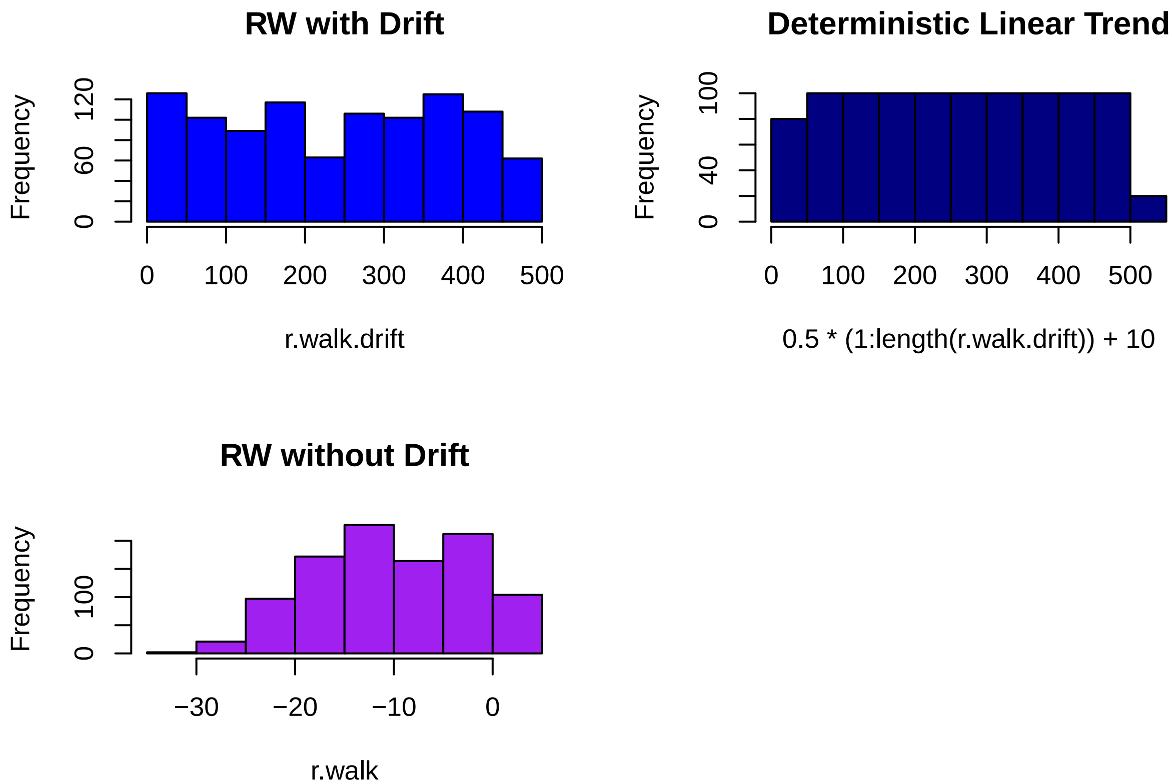
```
mean(r.walk.drift)
```

```
## [1] 239.9723
```

```
par(mfrow = c(1, 1))
plot.ts(r.walk.drift, main = "Random Walk with Drift, Random Walk without Drift, Deterministic Trend",
    col = "blue", ylab = "Values", xlab = "Simulated Time Period", bg = 38)
lines(0.5 * (1:length(r.walk.drift)) + 10, lty = "dashed", col = "navy")
lines(r.walk, col = "purple")
# Add vertical lines
abline(v = c(100, 200, 300, 400, 500, 600, 700, 800, 900, 1000), col = 3,
    lty = 3)
# Add Legend
leg.txt <- c("RW with Drift", "Deterministic Linear Trend", "RW without Drift")
legend("topleft", legend = leg.txt, lty = c(1, 2, 1), col = c("blue", "navy",
    "purple"), bty = "n", cex = 1, merge = TRUE, bg = 336)
```




```r
par(mfrow = c(2, 2))
hist(r.walk.drift, main = "RW with Drift", col = "blue")
hist(0.5 * (1:length(r.walk.drift)) + 10, main = "Deterministic Linear Trend",
     col = "navy")
hist(r.walk, main = "RW without Drift", col = "purple")
```


RW with Drift
Frequency
0
60
120
0
100
200
300
400
500
r.walk.drift
Deterministic Linear Trend
Frequency
0
40
100
0
100
200
300
400
500
0.5 * (1:length(r.walk.drift)) + 10
RW without Drift
Frequency
0
100
−30
−20
−10
0
r.walk