W271-2 – Spring 2016 – **HW 5**

Juanjo Carin, Kevin Davis, Ashley Levato, Minghu Song

March 9, 2016

Contents

Exercises

Question 1

1. **Install the library "astsa" using the function: install.packages("astsa")**

```
# Check if already installed; if not, install it
if (!"astsa" %in% installed.packages()[, "Package"]) install.packages("astsa")
```

2. **Load the library: library(astsa)**

```
# Load the library: library(astsa)
library(astsa)
# Last two commands can be substituted by simply...
if (!require(astsa)) install.packages("astsa")
```

3. **Use the function `str()` to see the information of a particular data series, such as `str(EQ5)` for the Seismic Trace of Earthquake number 5 series**

Since we will have to plot and describe another two time series, we use `str()` for the three of them.

```
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 ...
```

4. **Plot the time series plots and histograms of the following 3 series. Feel free to use the codes provided in the R scripts. Make sure that each of your graph has a title, the axis ticks are clear, the axes are well-labelled, and use color intelligently.**

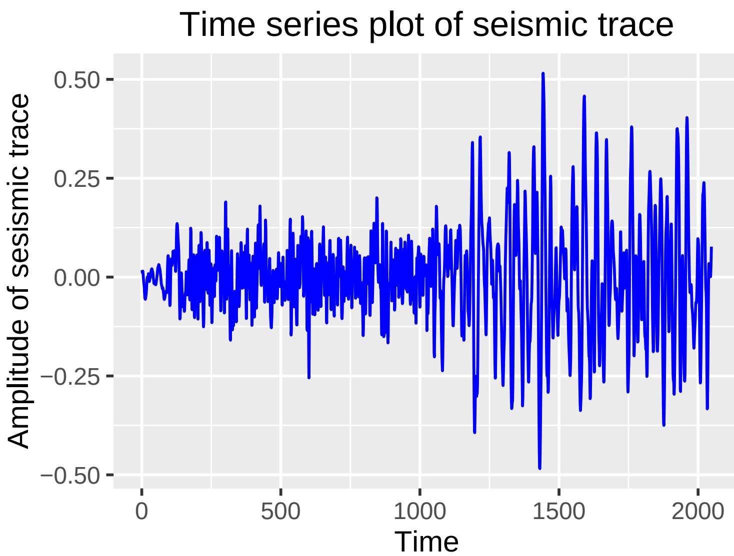

Figure 1: Time series plot of seismic trace EQ5

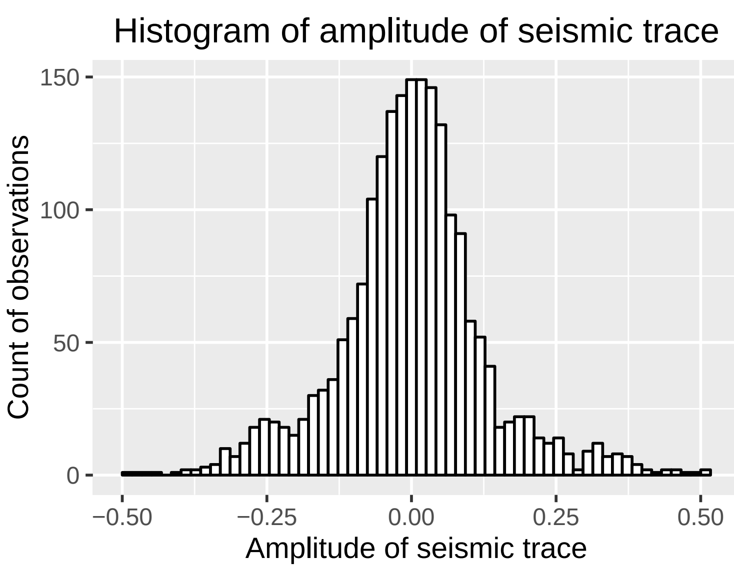

Figure 2: Histogram of amplitude of seismic trace

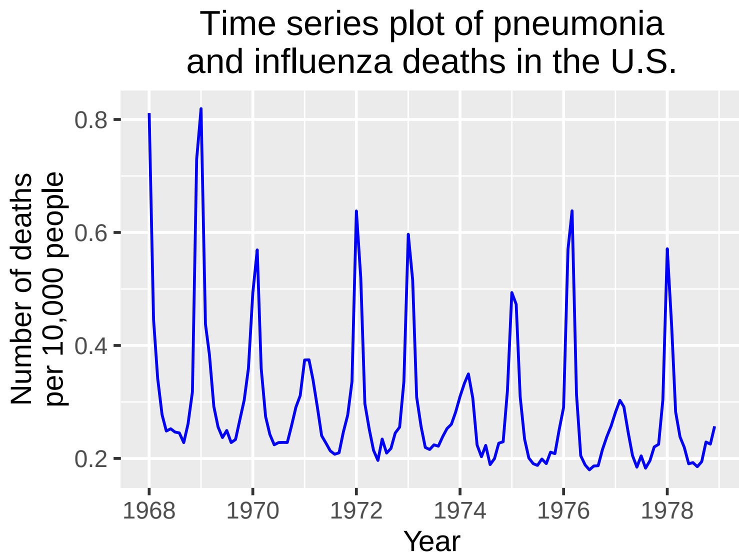

Figure 3: Time series plot of pneumonia and influenza deaths per 10,000 people in the U.S. from 1968 to 1979

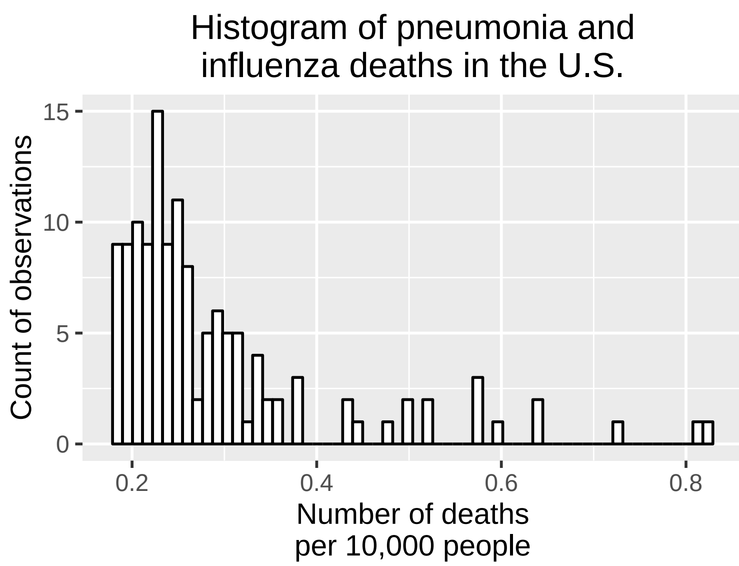

Figure 4: Histogram of pneumonia and influenza deaths per 10,000 people in the U.S. from 1968 to 1979

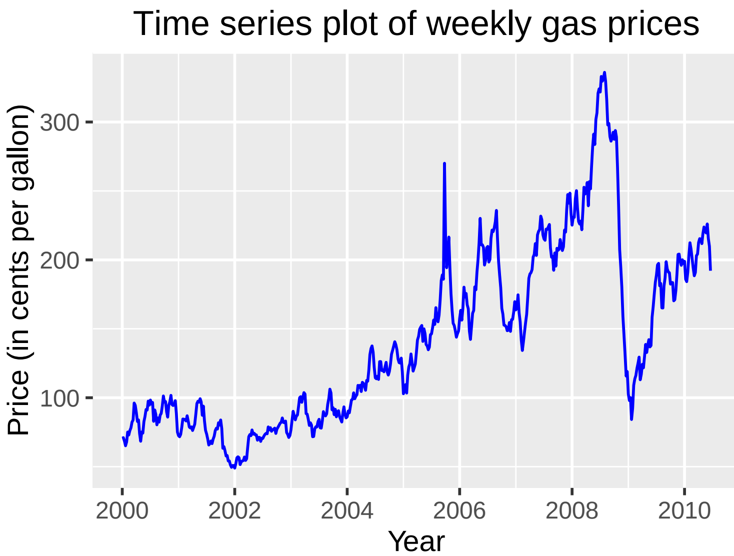

Figure 5: Time series plot of weekly gas prices (in cents per gallon) from 2000 to mid-2010

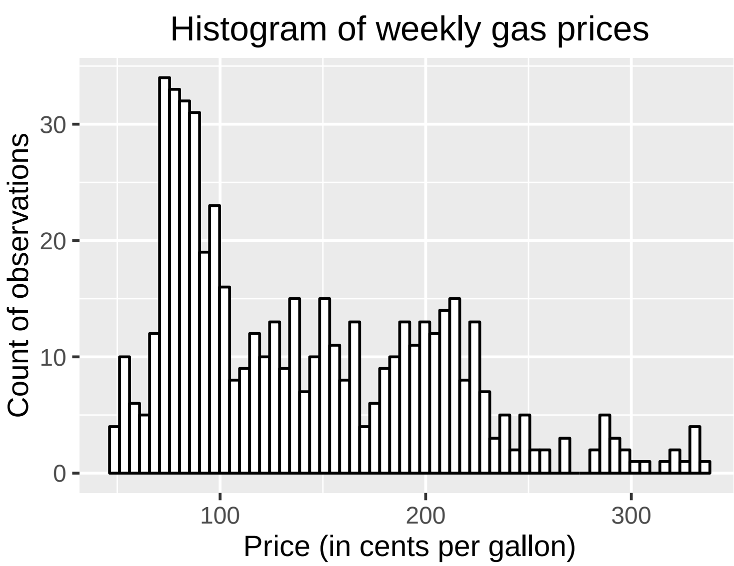

Figure 6: Histogram of weekly gas prices (in cents per gallon) from 2000 to mid-2010

5. **Write a few sentences to describe each of the series.**

- **EQ5**

  According to the package documentation, this time series corresponds to the *Seismic Trace of Earthquake number 5. Seismic trace of an earthquake [two phases or arrivals along the surface, the primary wave (t = 1, . . . , 1024) and the shear wave (t = 1025, . . . , 2048)] recorded at a seismic station.* Figure 1 shows that the amplitude of the primary phase is smaller than the amplitude of the shear phase, both have means around zero, and the shear phase seems to be periodic. Figure 2 shows that the whole dataset has a distribution close to normal.

- **flu**

  According to the package documentation, this time series corresponds to the *Monthly pneumonia and influenza deaths in the U.S., 1968 to 1978. Monthly pneumonia and influenza deaths per 10,000 people in the United States for 11 years, 1968 to 1978.* Figure 3 shows a seasonal pattern: most deaths occur at the beginning of the year (January or February, depending on the year), then the pneumonia and influenza death rate decreases gradually unitl the end of Autumn. Figure 4 shows that the distribution of the time series is right-skewed

- **gas**

  According to the package documentation, this time series corresponds to the *Gas Prices. New York Harbor conventional regular gasoline weekly spot price FOB (in cents per gallon) from 2000 to mid-2010.* Figure 5 shows a seasonal trend (with a period of about 6 months) and an increasing trend, with the exception of a significant drop in gas prices in the last half of 2008. Figure 6 shows that the distribution of gas prices is multimodal (two or three modes can be distinguished), with a long right tail.

Question 2

Describe 3 examples you have used in your work or encounter in real life. Ideally, you can even load at least one of these time series, plot it, and the write a few statements to describe its characteristics.

1. Biotech Stocks: In the first example, we pull out the stock data (from August 2010 to August 2015) of some biotech companies from yahoo financial web links. Biogen's closing stock price is used as the example here. Starting from 2010, it started to gradually go up because of the promising clinical results of aducanumab for Alzheimer treatment in March, 2013. However, the stock began to drop in 2015 due to slowing sales of its multiple sclerosis drug, Tecfidera.

```
# Plot some biotech stocks to watch in 2016.
# Define the variable to get access to the yahoo finacial stock data
biogen_stock_url <- "http://real-chart.finance.yahoo.com/table.csv?s=BIIB
&a=07&b=24&c=2010&d=07&e=24&f=2015&g=d&ignore=.csv"
mdvn_stock_url <- "http://real-chart.finance.yahoo.com/table.csv?s=MDVN
&a=07&b=24&c=2010&d=07&e=24&f=2015&g=d&ignore=.csv"
lexicon_stock_url <- "http://real-chart.finance.yahoo.com/table.csv?s=LXRX
&a=07&b=24&c=2010&d=07&e=24&f=2015&g=d&ignore=.csv"
gilead_stock_url <- "http://real-chart.finance.yahoo.com/table.csv?s=GILD
&a=07&b=24&c=2010&d=07&e=24&f=2015&g=d&ignore=.csv"
enanta_stock_url <- "http://real-chart.finance.yahoo.com/table.csv?s=ENTA
&a=07&b=24&c=2010&d=07&e=24&f=2015&g=d&ignore=.csv"
celgen_stock_url <- "http://real-chart.finance.yahoo.com/table.csv?s=CELG
&a=07&b=24&c=2010&d=07&e=24&f=2015&g=d&ignore=.csv"
# Define function to read financial data through url
yahoo.read <- function(url, variable) {
  dat <- read.table(url, header = TRUE, sep = ",")
  df <- dat[ ,c(1, 5)]
  names(df) <- c('Date', variable)
  df$Date <- as.Date(as.character(df$Date))
  return(df)
}
# Collect the stock data from 2010 to 2016 for those companies
biogen <- yahoo.read(biogen_stock_url, 'Biogen')
medivation <- yahoo.read(mdvn_stock_url, 'Medivation')
lexicon <- yahoo.read(lexicon_stock_url, 'Lexicon')
gilead <- yahoo.read(gilead_stock_url, 'Gilead')
enanta <- yahoo.read(enanta_stock_url, 'Enanta')
enanta2 <- data.frame(Date = biogen$Date)
enanta2$Enanta = ifelse(biogen$Date %in% enanta$Date,
                        enanta$Enanta, NA)
celgen <- yahoo.read(celgen_stock_url, 'Celgen')
# Time series plot for those stocks
whole_dataset <- cbind(biogen, medivation, lexicon, gilead, enanta2, celgen)
whole_dataset <- melt(whole_dataset, variable.name = 'Company', 'Date')
```

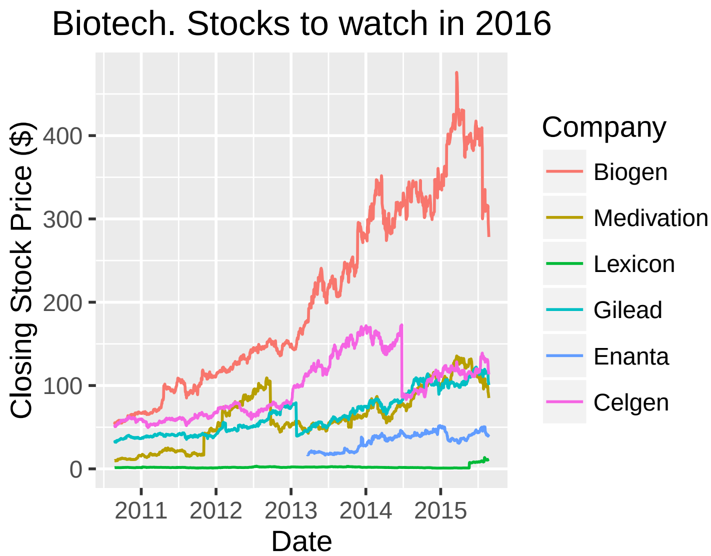

Figure 7: Stock price of some biotech companies from mid 2010 to mid 2015

2. **Daily Averaged Tempreture at JFK airport**: daily averaged temperature recorded at JFK airport weather station. The data is directly pulled out through the `WeatherData` package. Similar to stock price, it is also non-regular time series data. But the temperature at a centain time point would be highly relevant to the tempreture of previous times.

```
# The daily tempreture of 2015 at JFK airport.
# Get access to the weather data through weatherdata package
# (need the scales package as well)
W_KJFK_2015 <- getWeatherForYear("KJFK", 2015)
W_KJFK_2015$Date <- as.Date(W_KJFK_2015$Date, format = "%y-%m-%d")
```

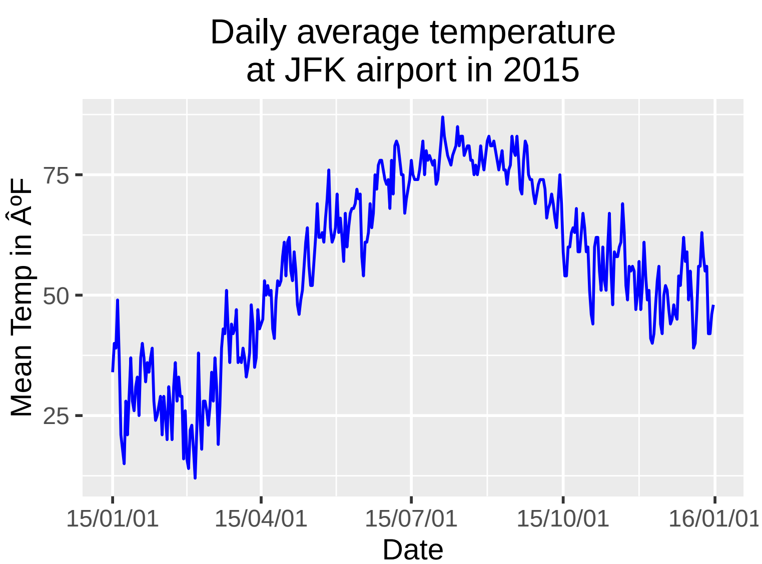

Figure 8: Daily average temperature at JFK airport in 2015

3. **Monthly Averaged Electricity Usage**: The 3rd example shows the monthly averaged electricity usage of my house between 2014 and 2015. This periodic series shows that summers, in general, tend to have lower monthly averaged electricity usage (<20KWh) than the other seasons, e.g. winter.

```
# The daily electricity usage for every month during 2014 Jan. and 2015 Dec.
# Load monthly averaged electricity usage for months in 2014 and 2015
elec_usage = c(35.94, 29.68, 31.83, 31.36, 24.61, 17.91, 18.29, 17.74, 15.70,
               19.33, 23.90, 23.39, 23.30, 28.63, 29.07, 24.27, 20.68, 18.33,
               18.42, 18.41, 24.66, 18.93, 23.03, 23.18)
elec_usage_ts <- ts(elec_usage, start=c(2014, 1), end=c(2015, 12), frequency=12)
elec_usage_df <- data.frame(Date = as.Date(time(elec_usage_ts)),
                             Price = as.numeric(elec_usage_ts))
ggplot(elec_usage_df, aes(Date, Price)) +
  geom_line(colour = 'blue') +
  labs(x = 'Date',  y = 'Daily Electricity Usage (KWh)',
       title = 'Monthyl Electricity Usage in 2014 & 2015') +
  scale_x_date(date_labels = "%b%Y")
```

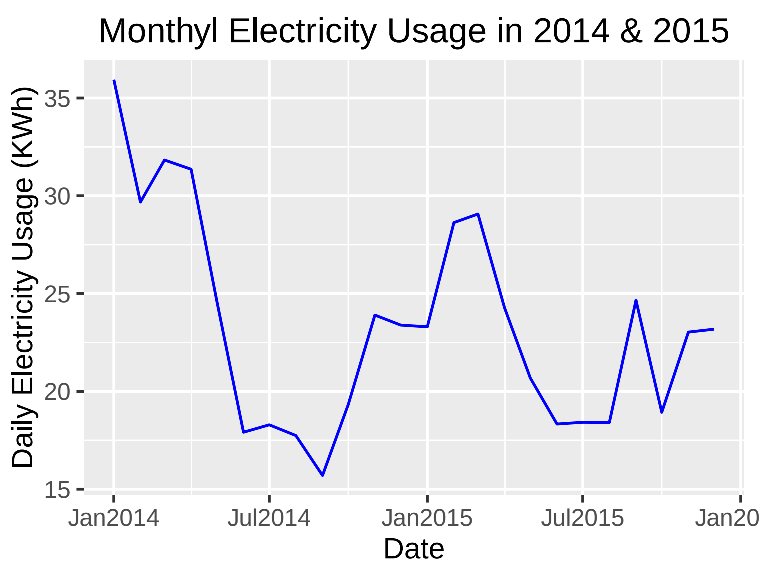

Figure 9: Monthly average electricity usage in 2014 and 2015

Question 3

Simulate a white noise series with 1000 random draws and plot (1) a time series plot and (2) a histogram. The usual reqirements on graphics (described) in Question 1) applied.

```
N <- 1000
wn <- data.frame(Time = 1:N, Level = rnorm(N))
```

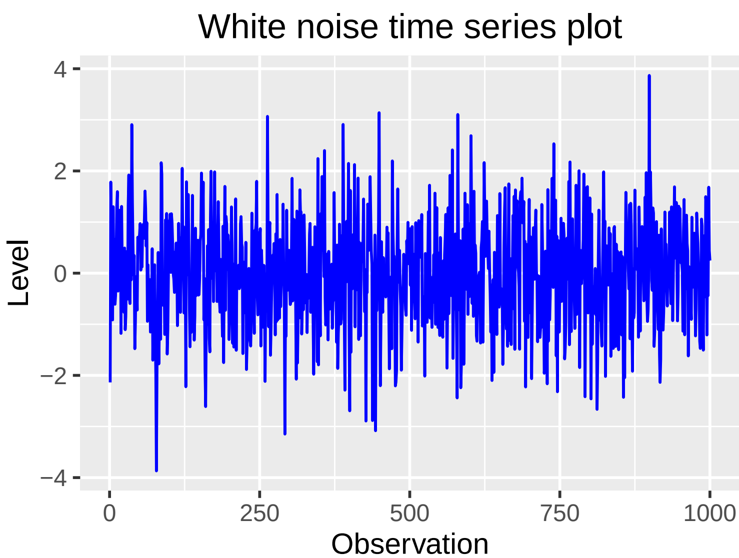

Figure 10: Plot of 1,000 random draws of a white noise series

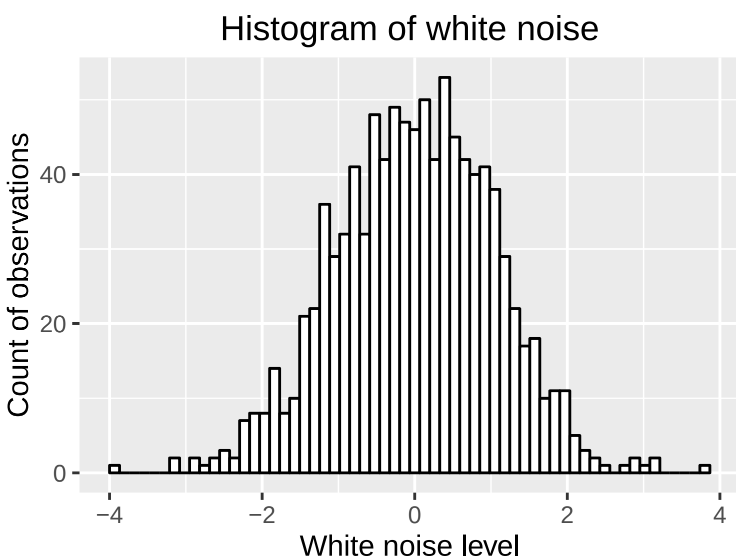

Figure 11: Histogram of 1,000 random draws of a white noise series

Question 4

Simulate (with 1000 random draws) the following two zero-mean autoregressive model with order 1 (i.e. AR(1)) models:

$$\mathbf{y_t = 0.9 y_{t1} + w}$$

$$\mathbf{y_t = 0.2 y_{t1} + w}$$

Plot a time plot for each of the simulated series. Graph a histogram for each of the simulated series. Write a few statements to compare the two series.

```
generate_AR <- function(wn, coef) {
  y <- wn
  for (i in 2:length(wn)) {
  y$Level[i] <- y$Level[i] + coef * y$Level[i - 1]
  }
  return(y)
}
y1 <- generate_AR(wn, 0.9)
```

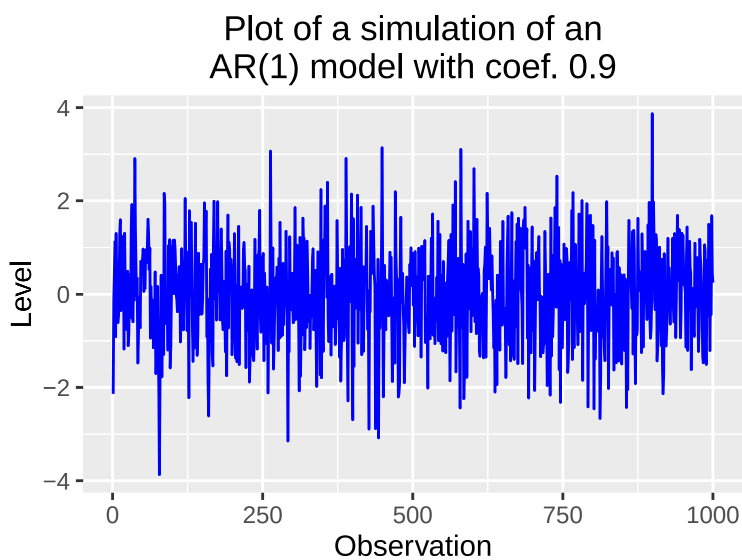

Figure 12: Plot of 1,000 observations of a simulation of an AR(1) model with coefficient 0.9

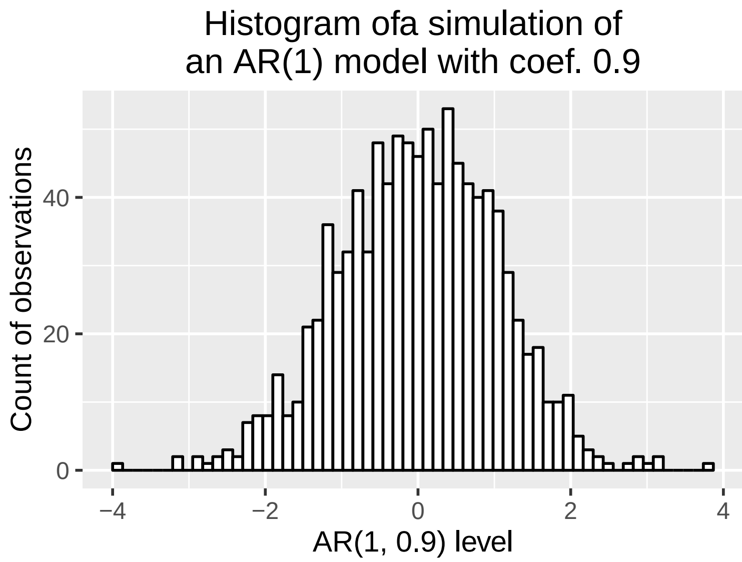

Figure 13: Histogram of 1,000 observations of a simulation of an AR(1) model with coefficient 0.9

```
y2 <- generate_AR(wn, 0.2)
```

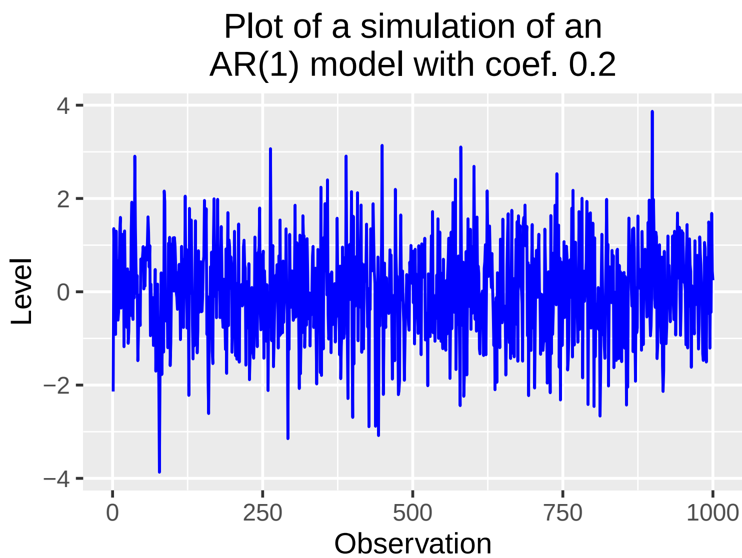

Figure 14: Plot of 1,000 observations of a simulation of an AR(1) model with coefficient 0.2

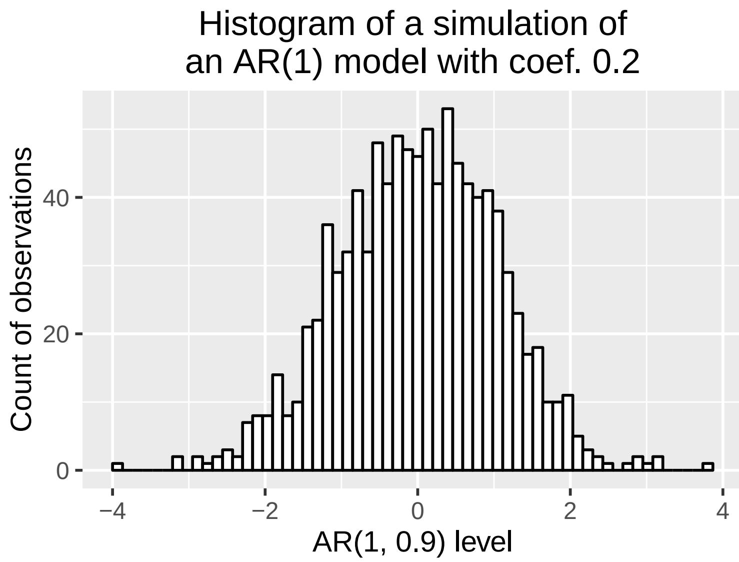

Figure 15: Histogram of 1,000 observations of a simulation of an AR(1) model with coefficient 0.2

Question 5

Simulate (with 1000 random draws) the following 3 models:

1. **A deterministic linear (time) trend of the form:** $y_t = 10 + 0.5t$
2. **Random walk without drift**
3. **Random walk with drift = 0.5**

Plot a time plot for each of the simulated series. Graph a histogram for each of the simulated series. Write a few statements to compare the two series.

```
trend_drift <- 0.5
y1 <- y2 <- y3 <- data.frame(Time = 1:N)
y1$'Trend of slope 0.5' <- 10 + trend_drift * 1:N
y2$'Random walk' <- cumsum(wn$Level)
y3$'Random walk with drift of 0.5' <- cumsum(wn$Level + trend_drift)
whole_dataset <- cbind(y1, y2, y3)
```

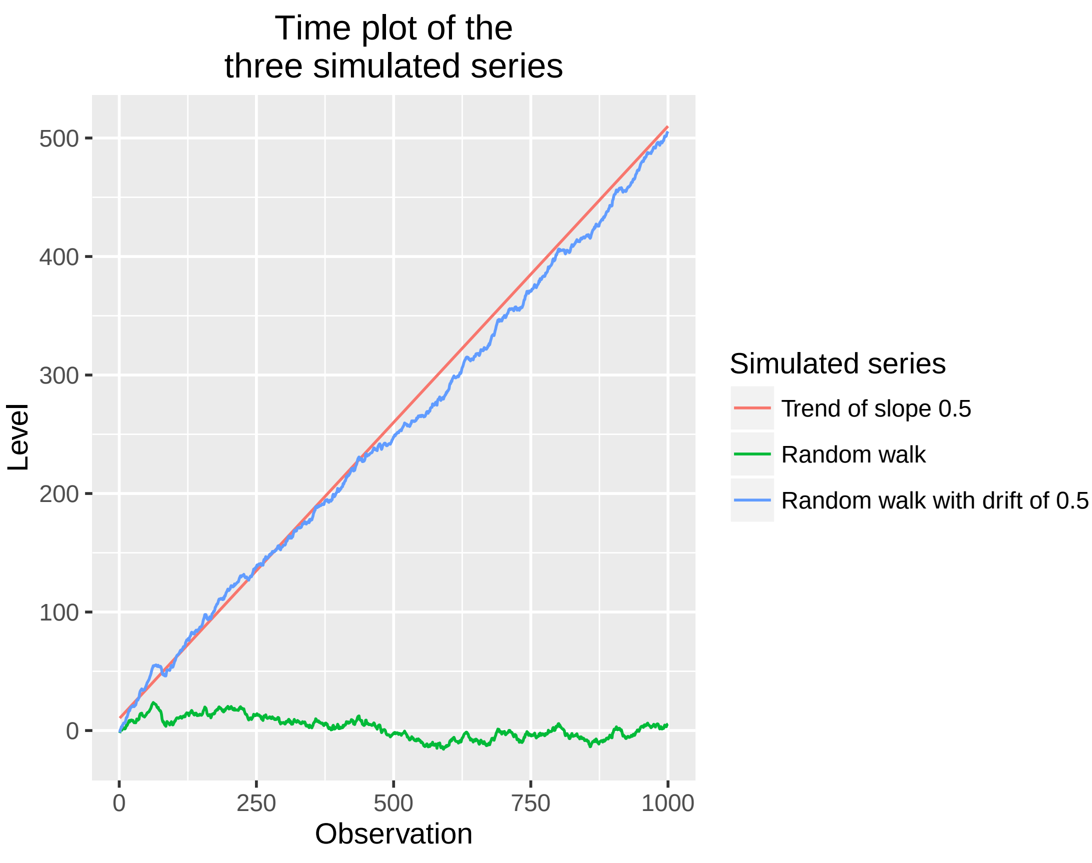

Figure 16: Time plot of the three simulated series

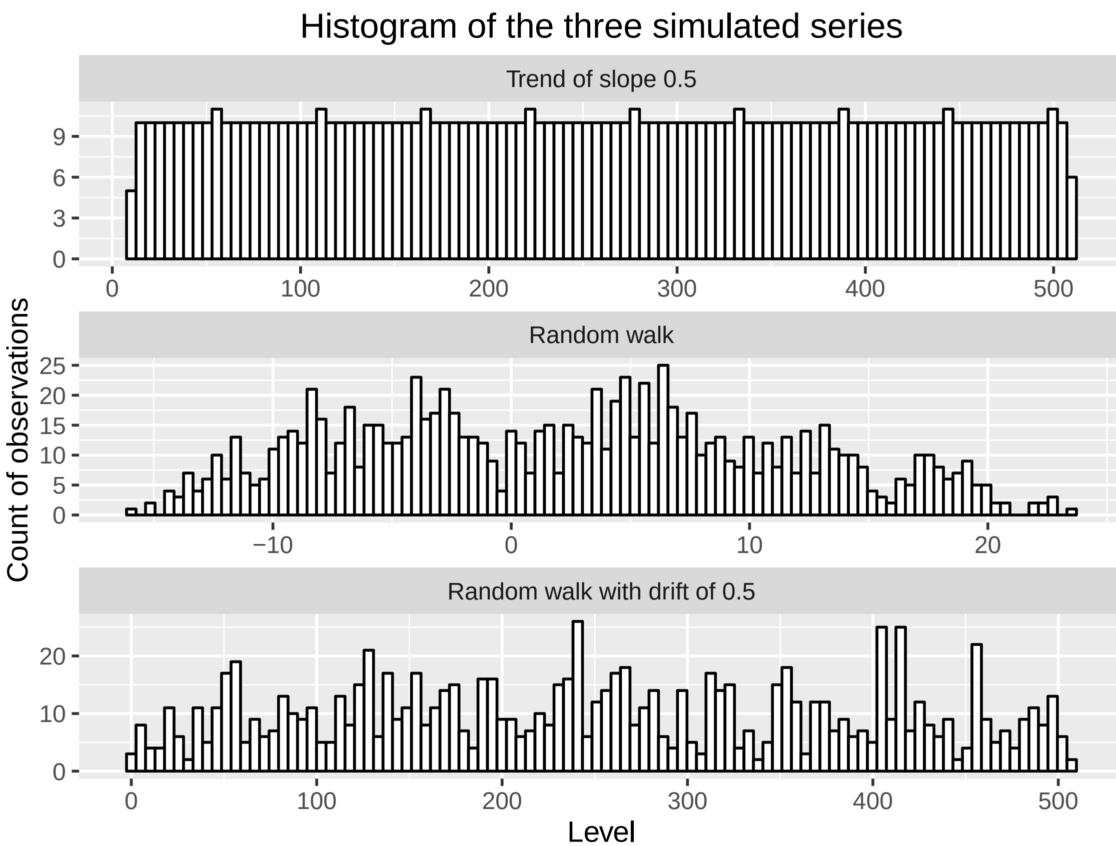

Figure 17: Histogram of the three simulated series

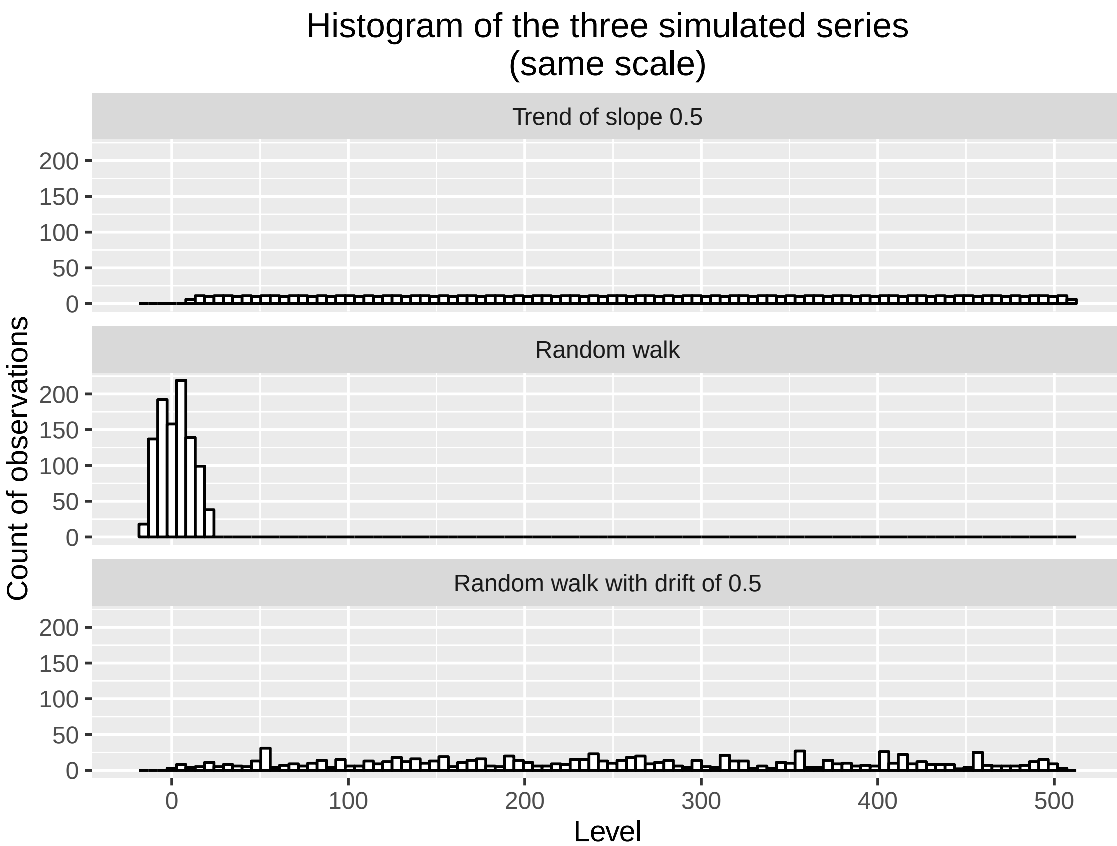

Figure 18: Histogram plot of the three simulated series (same scale)