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**

```
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 ...
```

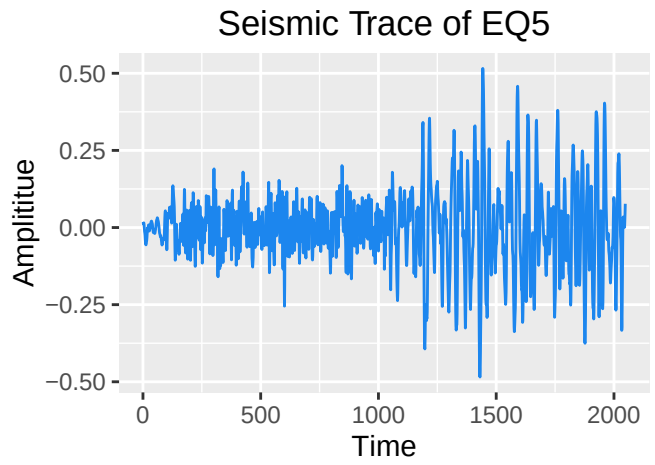
```
str(gas)
```

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

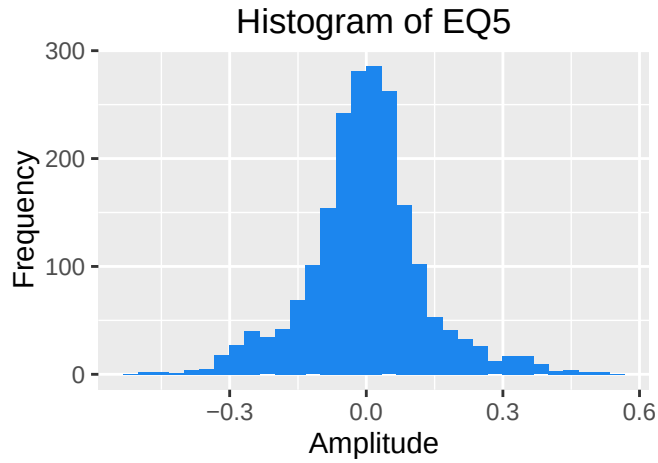
According to that package documentation, `EQ5` corresponds to the *Seismic trace of an earthquake [two phases or arrivals along the surface, the primary wave (t = 1, . . . , 1024) and the shear wave (t = 1025, . . . , 2048)] recorded at a seismic station.*

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.**

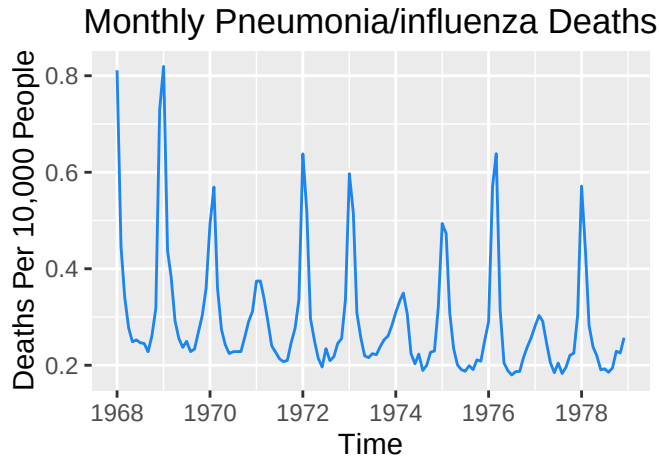
```
# Time series plot and histogram of EQ5 seismic trace
autoplot(EQ5, main='Seismic Trace of EQ5', ts.colour= 'dodgerblue2',
         xlab='Time', ylab='Amplititue')
```



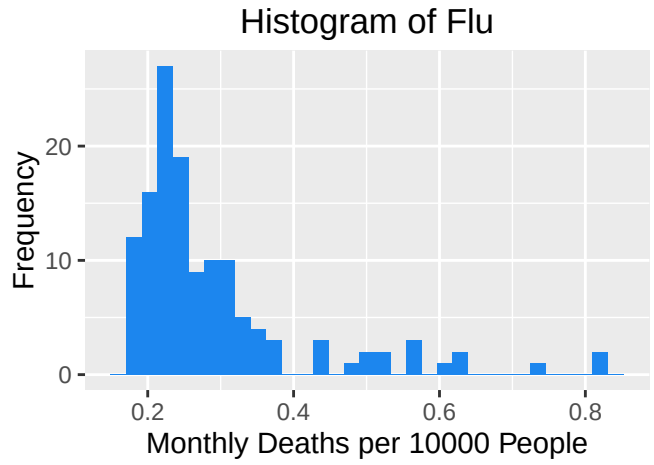

```
qplot(EQ5, geom="histogram",  main='Histogram of EQ5',
      xlab='Amplitude', ylab='Frequency', fill=I("dodgerblue2"))
```



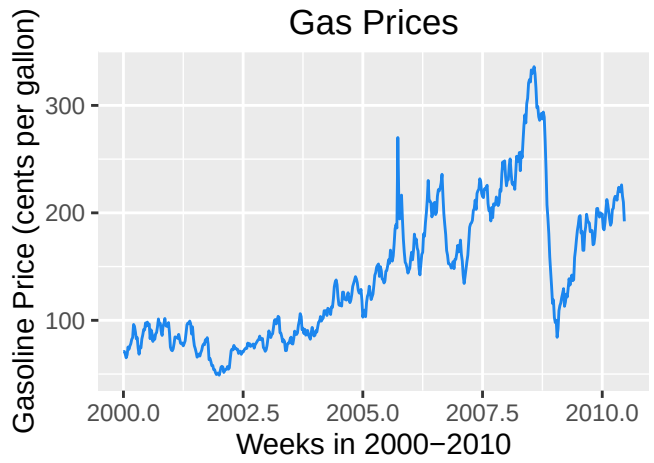

```
# Time series plot and histogram of flu
autoplot(flu, main='Monthly Pneumonia/influenza Deaths', ts.colour = 'dodgerblue2',
         xlab='Time', ylab='Deaths Per 10,000 People')
```



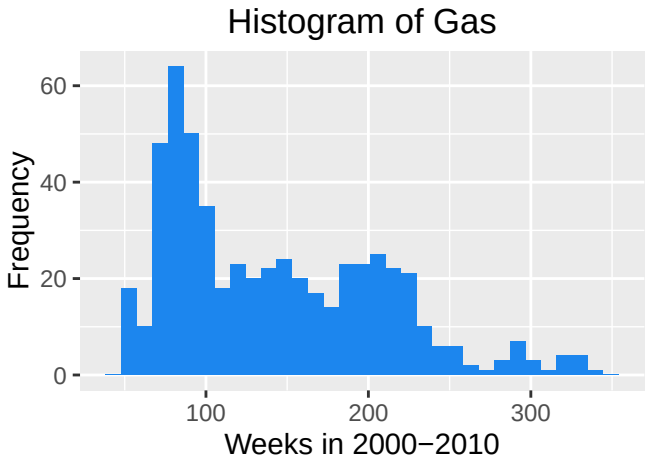
```
qplot(flu, geom="histogram",  main='Histogram of Flu',
      xlab='Monthly Deaths per 10000 People', ylab='Frequency', fill=I("dodgerblue2"))
```




```
# Time series plot and histogram of gas price
autoplot(gas, main='Gas Prices',  ts.colour = 'dodgerblue2', xlab='Weeks in 2000-2010',
         ylab='Gasoline Price (cents per gallon)')
```




```
qplot(gas, geom="histogram",  main='Histogram of Gas',
      xlab='Weeks in 2000-2010', ylab='Frequency', fill=I("dodgerblue2"))
```



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

- **EQ5:** The EQ5 data encodes the seismic trace of an earthquake saved as the ts data class. It include two arrive phases: primary wave or P phase (t=1, . . . ,1024) and the shear wave or S phase (t= 1025, . . . , 2048)]. In this series, the amplitude of the P phase is much smaller than the amplitude of the S phase. In addition, as shown in the above time series plot, it seems that the S phase of EQ5 series is periodic as well because it contains several similar segments with the equal length.
- **flu:** The flu data records monthly deaths per 10,000 people due to pneumonia and influenza in the United States for a period of 132 months(1968-1978). This series also exhibits a periodic pattern that the death rate tends to be the highest in January and then decreases gradually till next year.
- **gas:** The gas data collects the weekly price of gasoline in cents per gallon at New York Harbor from 2000 to mid-2010. According to the data, gas price fist gradually go up, but drop suddently around the 4th quarter of 2009.

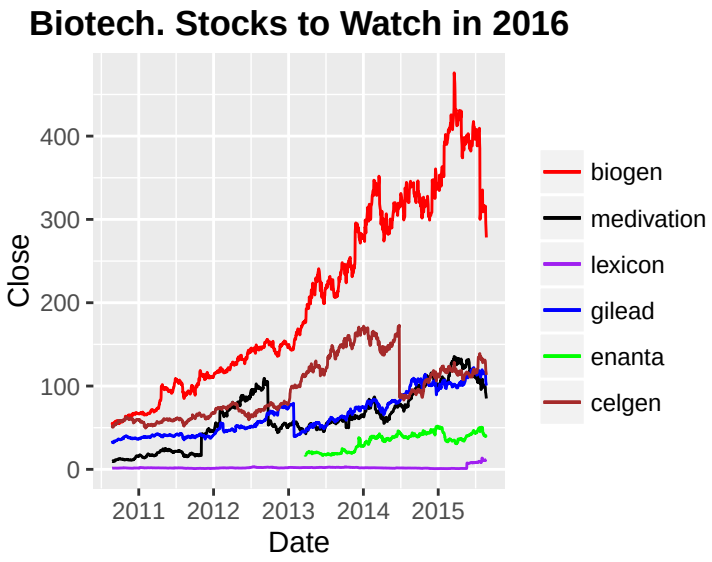
Question 2

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.

- **the Biotech Stock Example:** Stock closing data is one kind of Non-regular time series and they will change every day as a result of market forces. We use Biogen as the example here. Its closing stock price went up and down significantly in 2015 due to the promising clinical results of aducanumab for Alzheimer treatment in March as well as later unexpected slowing sales of its multiple sclerosis drug, Tecfidera.

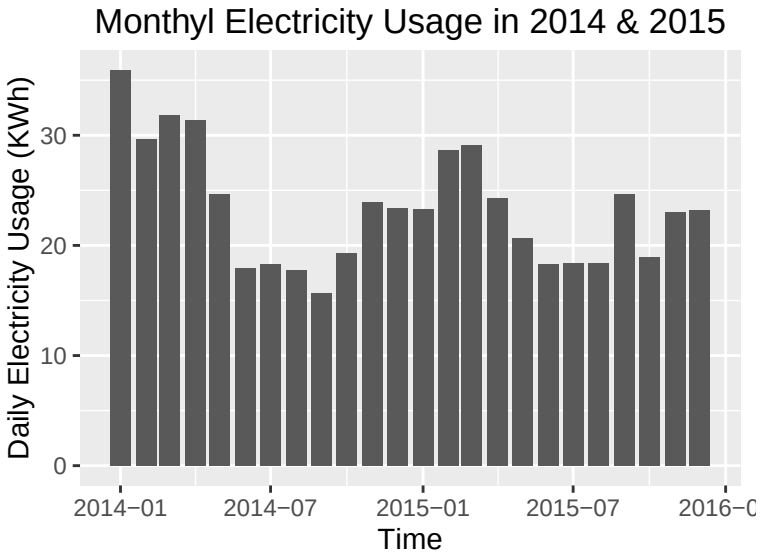
```
# define the variable to get access to the yahoo finacial stock data
biogen_stock_url <- "http://real-chart.finance.yahoo.com/table.csv?s=BIIB
&a=07&b=24&c=2010&d=07&e=24&f=2015&g=d&ignore=.csv"
mdvn_stock_url <- "http://real-chart.finance.yahoo.com/table.csv?s=MDVN
&a=07&b=24&c=2010&d=07&e=24&f=2015&g=d&ignore=.csv"
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){
  dat <- read.table(url,header=TRUE,sep=",")
  df <- dat[,c(1,5)]
  df$Date <- as.Date(as.character(df$Date))
  return(df)}

# grap the stock data from 2010 to 2016 for those companies
biogen  <- yahoo.read(biogen_stock_url)
medivation <- yahoo.read(mdvn_stock_url)
lexicon <- yahoo.read(lexicon_stock_url)
gilead <- yahoo.read(gilead_stock_url)
enanta <- yahoo.read(enanta_stock_url)
celgen <- yahoo.read(celgen_stock_url)

# time series plot for those stocks
ggplot(biogen,aes(Date,Close)) +
  geom_line(aes(color="biogen")) +
  geom_line(data=medivation,aes(color="medivation")) +
  geom_line(data=lexicon,aes(color="lexicon")) +
  geom_line(data=gilead,aes(color="gilead")) +
  geom_line(data=enanta,aes(color="enanta")) +
  geom_line(data=celgen,aes(color="celgen")) +
  labs(color="Legend") +
  scale_colour_manual("", breaks = c("biogen", "medivation","lexicon","gilead","enanta","celgen"),
                      values = c("red", "brown", "green", "blue","purple","black")) +
  ggtitle("Biotech. Stocks to Watch in 2016") +
  theme(plot.title = element_text(lineheight=.7, face="bold"))
```



- **the Monthyl Averaged Electricity Usage Example:** The 2nd examples show the monthly averaged electricity usage of my house between 2014 and 2015. This data shown summers, in general, tended to have lower monthly averaged electricity usage (<20KWh) than the other seasons.

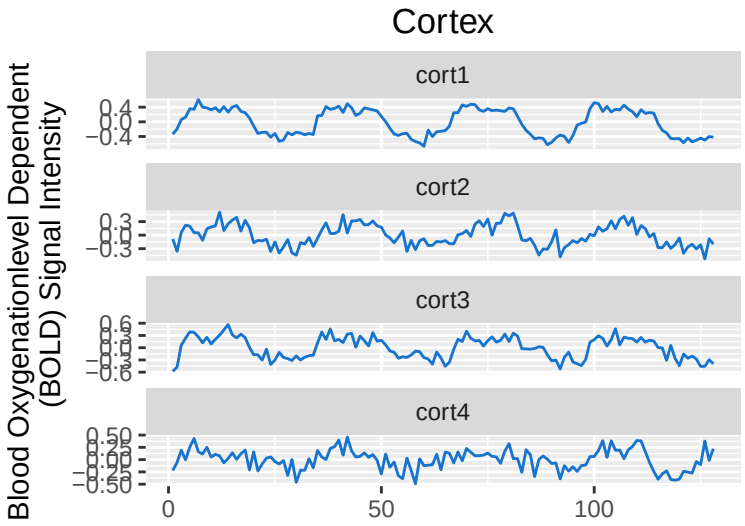
```
# load monthyl averaged electricity usage for months in 2014 and 2015
elec_usage = c(35.94, 29.68, 31.83, 31.36, 24.61, 17.91, 18.29, 17.74, 15.70,
               +19.33, 23.90, 23.39, 23.30, 28.63, 29.07, 24.27, 20.68, 18.33,
               +18.42, 18.41, 24.66, 18.93, 23.03, 23.18)
elec_usage_ts <- ts(elec_usage, start=c(2014, 1), end=c(2015, 12), frequency=12)
autoplot(elec_usage_ts, main='Monthyl Electricity Usage in 2014 & 2015', geom = "bar",
         xlab='Time', ylab='Daily Electricity Usage (KWh)')
```



- **the fMRI Example:** Functional Magnetic Resonance Imaging (fMRI) is a powerful technique to address the functional connectivity within brain. We used the fMRI data in the astsa package as the illustrating example here. The plot below shows the fMRI spectrum measuring the blood

oxygenationlevel dependent (BOLD) signal intensity from the cortex of four differnt subjects. For each subject, singals were sampled every 2 seconds for 256 seconds totally(n=128). During the sampling process, the stimulus was first applied for 32 seconds and then stopped for the subsequent 32 seconds. Therefore, the whole sampling process (256 seconds) covers four periodic components.

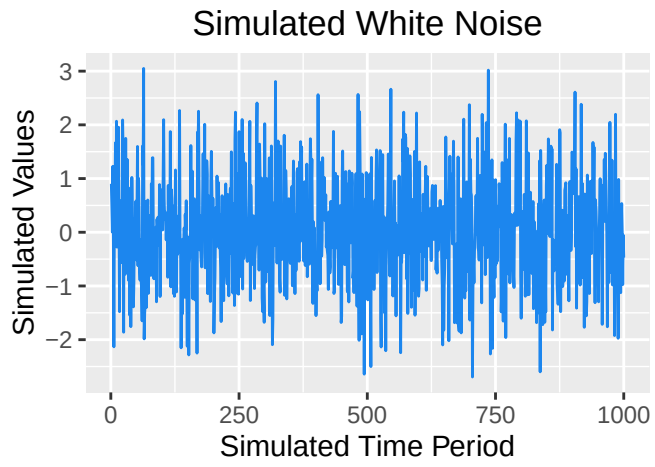
```
autoplot(fmri1[,2:5],  ylab="Blood Oxygenationlevel Dependent
         (BOLD) Signal Intensity", xlab="",main="Cortex",  ts.colour = 'dodgerblue3')
```



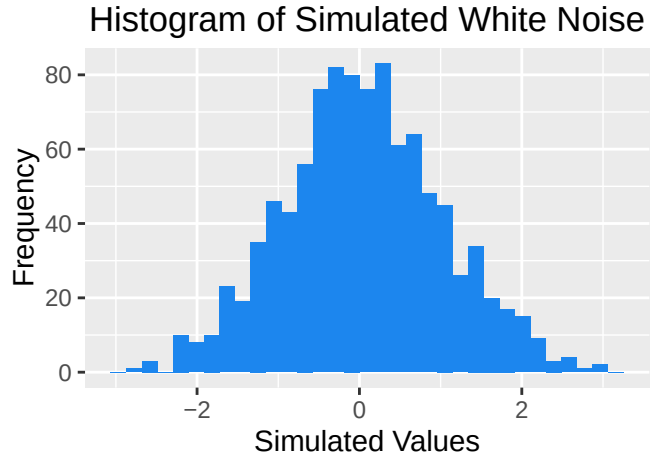
Question 3

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.

```
rand_draw <- rnorm(1000) # 1000 random draw
rand_draw_ts <- ts(rand_draw)
autoplot(rand_draw_ts , xlab = "Simulated Time Period", ylab = "Simulated Values",
         main="Simulated White Noise", ts.colour = 'dodgerblue2')
```




```
qplot(rand_draw_ts, geom="histogram",  main='Histogram of Simulated White Noise',
      xlab='Simulated Values', ylab='Frequency', fill=I("dodgerblue2"))
```



Question 4

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

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

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

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.

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.
