Homework 1

Yeshashvi Munagala

1/19/2022

Problem 1

1a)

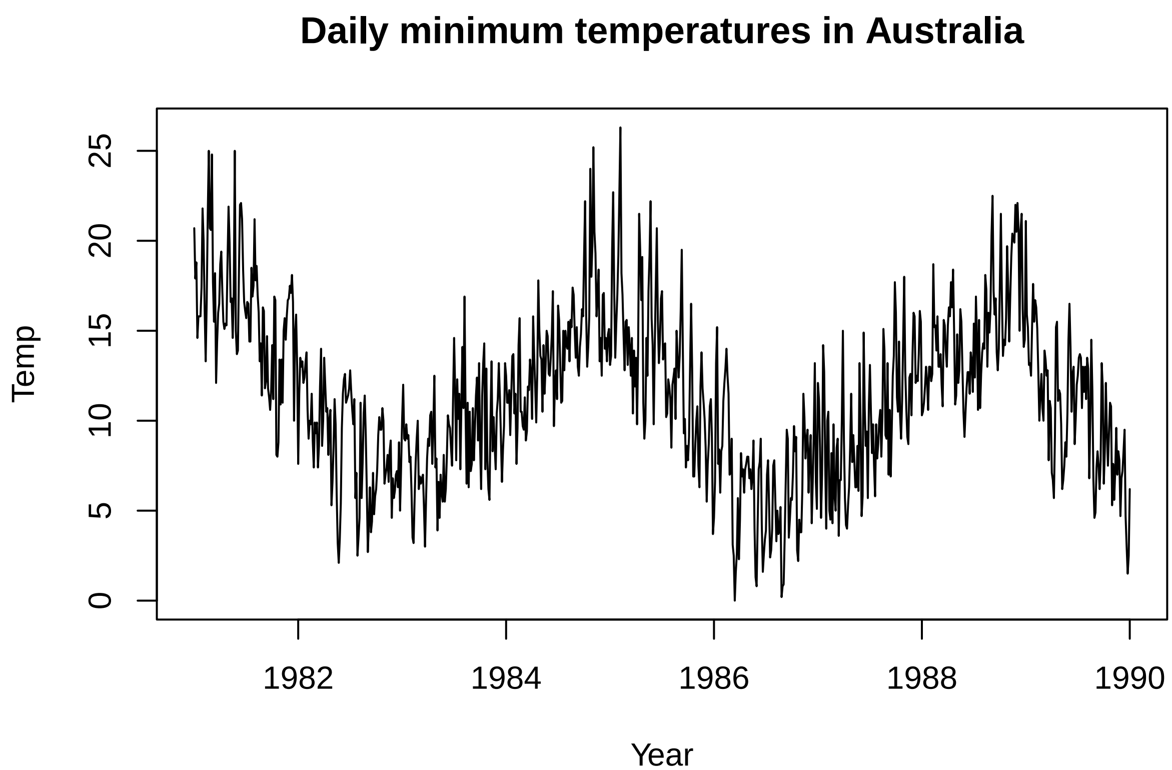
```
Daily minimum temperature dataset source: https://machinelearningmastery.com/time-series-datasets-for-ma

Airplane Fatalities since 1908 dataset source:
https://www.kaggle.com/saurograndi/airplane-crashes-since-1908
```

1b, 1c)

```
library(readr)
temp <- read.csv("~/Downloads/daily-min-temperatures.csv")
tempe <- subset(temp, select = c("Temp"))
head(tempe, n = 10)
```

```
##    Temp
## 1  20.7
## 2  17.9
## 3  18.8
## 4  14.6
## 5  15.8
## 6  15.8
## 7  15.8
## 8  17.4
## 9  21.8
## 10 20.0
```

```
tempe <- ts(data = tempe, start = 1981, end = 1990, frequency = 100)
plot.ts(tempe,ylab = 'Temp',xlab='Year', main = "Daily minimum temperatures in Australia")
```




```r
crash <- read_csv("~/Downloads/Airplane_Crashes_and_Fatalities_Since_1908.csv")
```

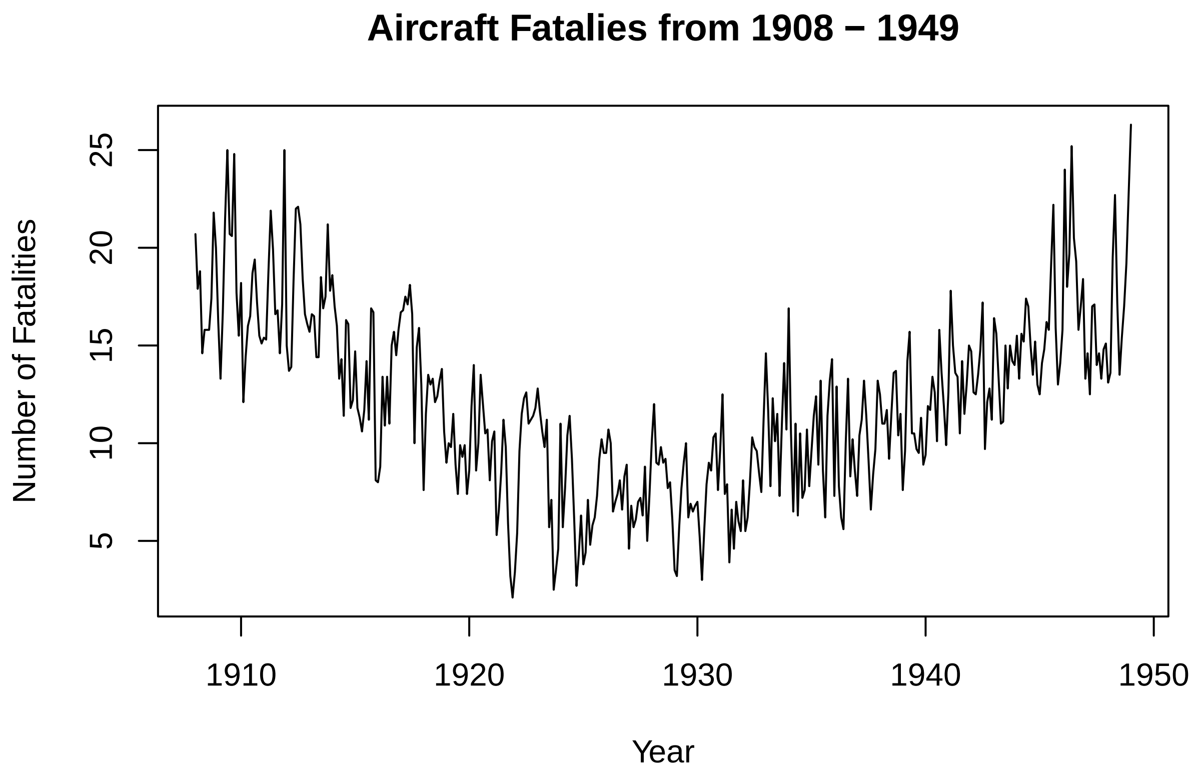
```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Date = col_character(),
##   Time = col_character(),
##   Location = col_character(),
##   Operator = col_character(),
##   `Flight #` = col_character(),
##   Route = col_character(),
##   Type = col_character(),
##   Registration = col_character(),
##   `cn/In` = col_character(),
##   Aboard = col_double(),
##   Fatalities = col_double(),
##   Ground = col_double(),
##   Summary = col_character()
## )
```

```r
crash1 <- subset(crash, select = c("Fatalities"))
head(crash1, n = 10)
```

```
## # A tibble: 10 x 1
##    Fatalities
##         <dbl>
##  1          1
##  2          5
##  3          1
##  4         14
##  5         30
```

```
## 6           21
## 7           19
## 8           20
## 9           22
## 10          19
```

```
crash1 <- ts(data = tempe, start = 1908, end = 1949, frequency = 10)
plot.ts(crash1 ,ylab = 'Number of Fatalities',xlab='Year', main = "Aircraft Fatalies from 1908 - 1949")
```



#

1d)

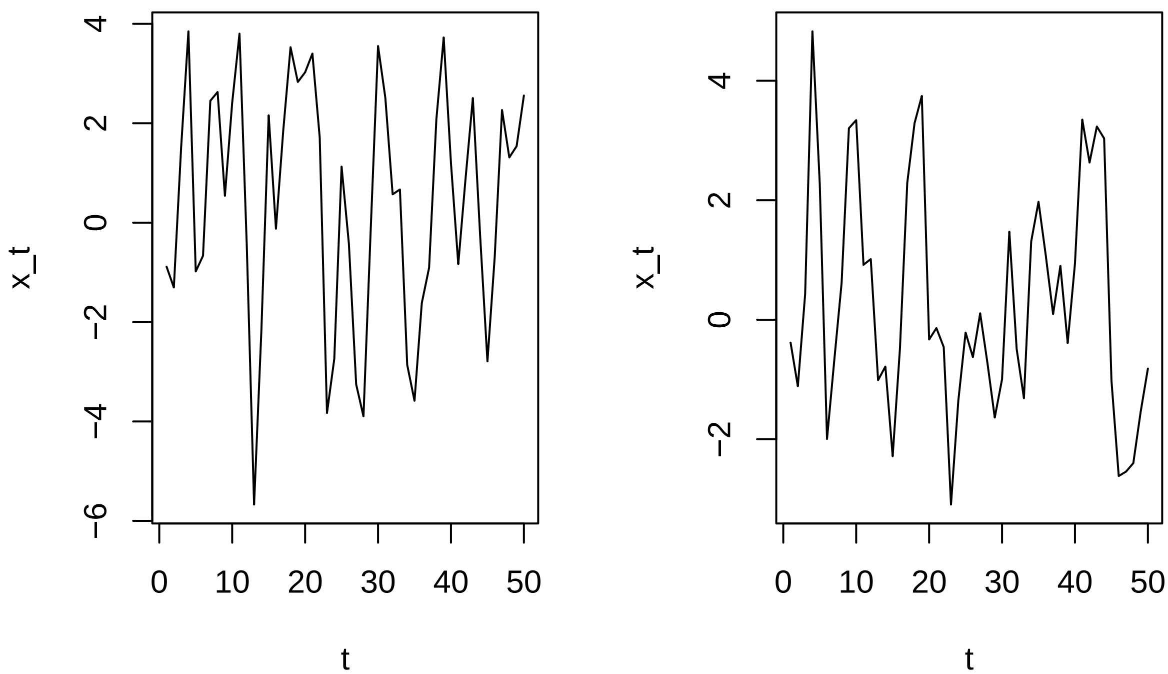
```
Daily minimum temperatures in Australia: This time series model could be used
to study the seasonality of minimum temperatures correlating to the seasons of
the year. This could help in making future predictions for minimum temperatures.

Aircraft Fatalies from 1908 - 1949: Through analysis of this dataset, airline
companies can observe potential trends to see which type of airlines have the
most fatalities and use this to improve their safety protocols. I would also
like to analyze why fatalities were at the lowest point from 1920-1930 but then
again rose after.
```

Problem 2

2a)

```
T <- 50
set.seed(1)

# realization #1 with Normal(0,4)
zt1 <- rnorm(T+1,0,2)
xt1 <- zt1[2:(T+1)] + zt1[1:T]
```

```
# realization #2 with Normal(0, 2)
zt2 <- rnorm(T+1,0,sqrt(2))
xt2 <- zt2[2:(T+1)] + zt2[1:T]

par(mfrow = c(1, 2))
plot.ts(xt1,ylab ='x_t',xlab='t')
plot.ts(xt2,ylab ='x_t',xlab='t')
```



#

2b,c)

```
Theoretical Computations:

E[Xt] = E[Zt + 2Zt-1] = E[Zt] + 2*E[Zt-1] = 3E[Zt] = 3(0) = 0 for both realizations


E[Xt^2] = E[Zt^2 + 4ZtZt-1 + 4Zt-1^2] = E[Zt^2] + 4E[ZtZt-1] + 4E[Zt-1^2] =
5E[Zt^2] + 4E[ZtZt-1] = 5E[Zt^2] + 4(E[Zt])^2 =
    for realization 1: 5(4) + 4(0)^2 = 20
    for realization 2: 5(2) + 4(0)^2 = 10


E[Xt,Xt-1] = E[(Zt + 2Zt-1) (Zt-1 + Zt-2)] = E[ZtZt-1] + 2E[ZtZt-2]
+ 2E[Zt-1^2] + 4E[Zt-1Zt-2] =
  for realization 1: 0(0) + 2(0)(0) + 2(4) + 4(0) = 8
  for realization 2: 0(0) + 2(0)(0) + 2(2) + 4(0) = 4


Corr[Xt, Xt-1] = (E[XtXt-1] - E[Xt]*E[Xt-1]) /
  (sqrt((E[Xt^2 - (E[Xt])^2]))(E[Xt-1^2 - (E[Xt-1])^2])) =
    for realization 1: (8 - 0) / sqrt((20-0)^2) = 0.4
    for realization 2: (4 - 0) / sqrt((10-0)^2) = 0.4
```

```
c(mean(xt1), mean(xt1^2), mean(xt1[-1]*xt1[-T]), cor(xt1[-1],xt1[-T]))
```

```
## [1] 0.4427755 5.7806663 2.6296542 0.4334556
```

```
c(mean(xt2), mean(xt2^2), mean(xt2[-1]*xt2[-T]), cor(xt2[-1],xt2[-T]))
```

```
## [1] 0.2927378 3.7108609 2.1143503 0.5480502
```

2d)

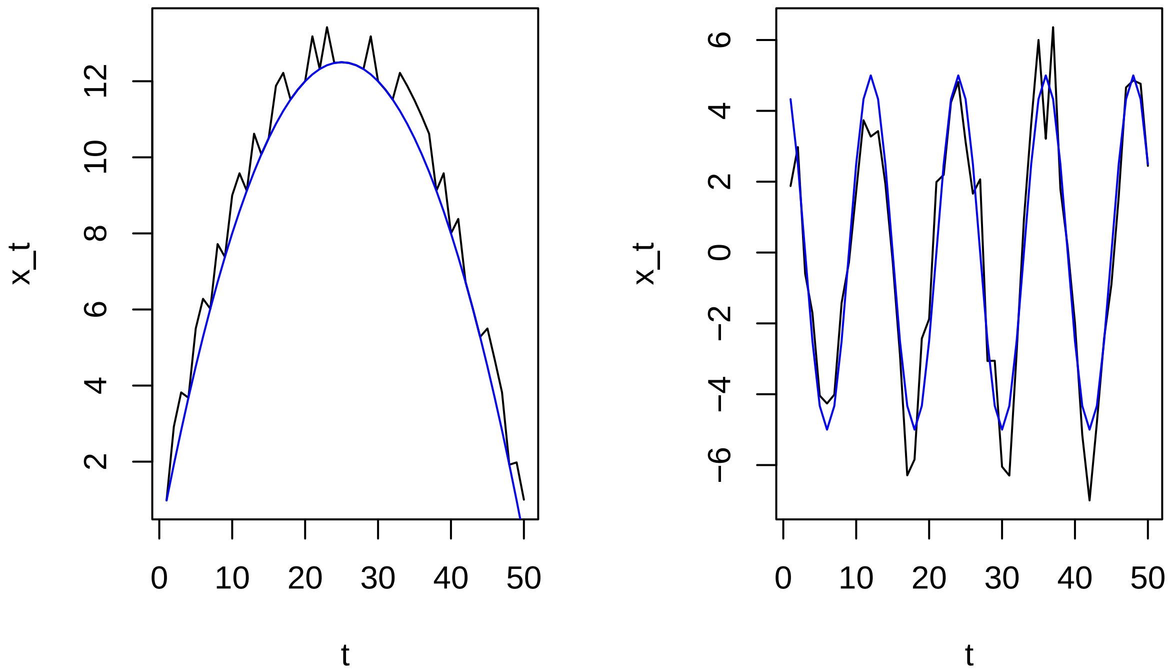
```
Possible values of Xt are {-4, -1, 2}
P(Xt = -4) = P(Zt = -2) * P(Zt-1 = -2) = (1/3)^2 = 1/9

P(Xt = -1) = (P(Zt = -2) * P(Zt-1 = 1)) + (P(Zt = 1) * P(Zt-1 = -2)) =
                                                        2 * (1/3)*(2/3) = 4/9

P(Xt = 2) = P(Zt = 1) * P(Zt-1 = 1) = (2/3)^2 = 4/9
```

Problem 3

3a)

```
The time series model I will be using is:
Xt = mt + Zt with mt = (50 - t^2)/50 and IID noise Zt with Zt ~ Bernoulli(.5)
Xt = st + Zt with st = 5cos((pi*t)/6) and IID noise Zt with Zt ~ Normal(0,1)
```

```
T <- 50
set.seed(2)
tt <- seq(1,T)

# realization 1 with Bernoulli(.5)
mt <- (50*tt-tt^2)/50
zt <- rbinom(T,1,.5)
xt1 <- mt + zt

# realization 2 with Normal(0,1)
st <- 5*cos(pi*tt/6)
zt <- rnorm(T,0,1)
xt2 <- st + zt

par(mfrow = c(1, 2))
plot.ts(xt1,ylab = 'x_t',xlab='t')
lines(1:T,mt,col="blue")
plot.ts(xt2,ylab = 'x_t',xlab='t')
lines(1:T,st,col="blue")
```



3b, 3c)

```
Theoretical Computations:

E[Xt] = E[Z1 + ... + Z50] =
    for realization 1: 50(.5) = 25
    for realization 2: 50(0) = 0


E[Xt^2] = t*E[Z^2] =
    for realization 1: (50 * (.5)^2) + 25^2 = 637.5
    for realization 2: 50 * E[Zt^2] = 50(1) = 50


E[XtXt-1] = (E[Xt])^2 =
    for realization 1: 25^2 = 625
    for realization 2: 1


Corr[Xt, Xt-1] = (E[XtXt-1] - E[Xt]*E[Xt-1]) /
  (sqrt((E[Xt^2 - (E[Xt])^2]))(E[Xt-1^2 - (E[Xt-1])^2])) =
    for realization 1: (625 - ((24.5)^2) / sqrt((25 - 637.5)^2) = 1
    for realization 2: (1 - 0) / sqrt((1 - 0)^2) = 1
```

```
c(mean(xt1), mean(xt1^2), mean(xt1[-1]*xt1[-T]), cor(xt1[-1],xt1[-T]))
```

```
## [1]  8.8300000 91.2957200 92.7035918  0.9646636
```

```
c(mean(xt2), mean(xt2^2), mean(xt2[-1]*xt2[-T]), cor(xt2[-1],xt2[-T]))
```

```
## [1] 6.581034e-04 1.325134e+01 1.070042e+01 7.970467e-01
```