Stat 153 - HW02

Nicholas Lai

September 19, 2018

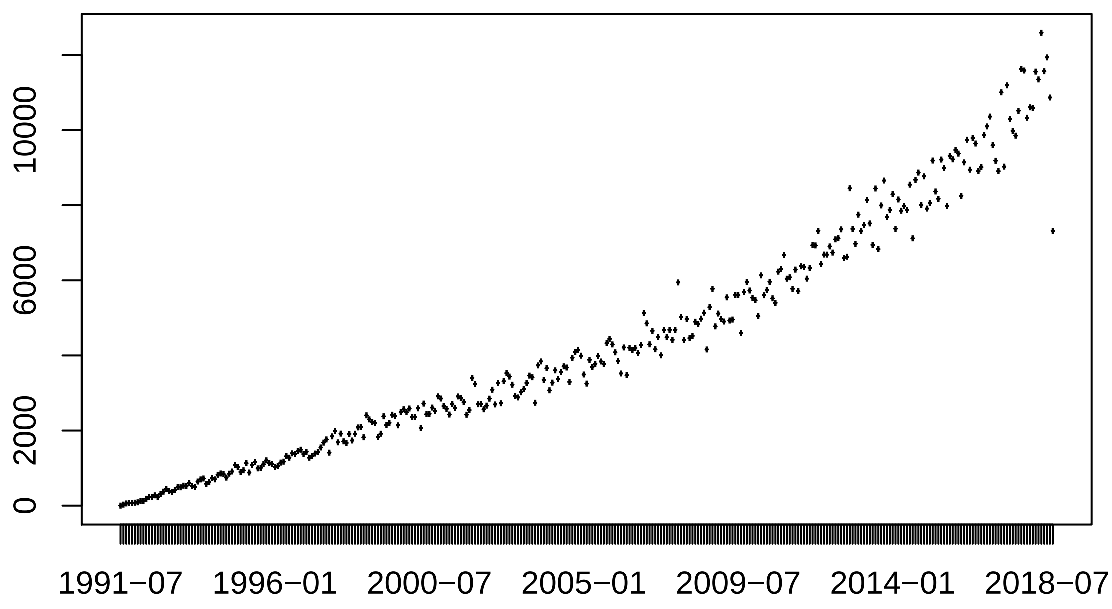
Computer Exercises

Question 1

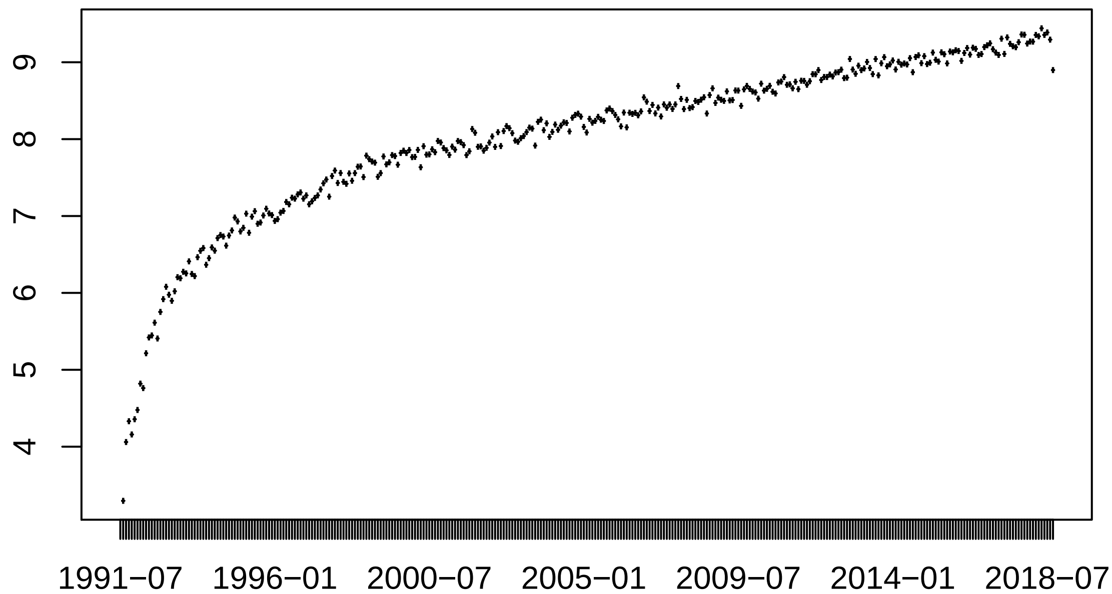
```
dat <- read.csv('get_monthly_submissions')
head(dat)
```

```
##     month submissions historical_delta
## 1 1991-07           0               -2
## 2 1991-08          27               -1
## 3 1991-09          58                0
## 4 1991-10          76                0
## 5 1991-11          64                0
## 6 1991-12          78                0
```

```
plot(dat$month, dat$submissions)
```




```
plot(dat$month, log(dat$submissions))
```



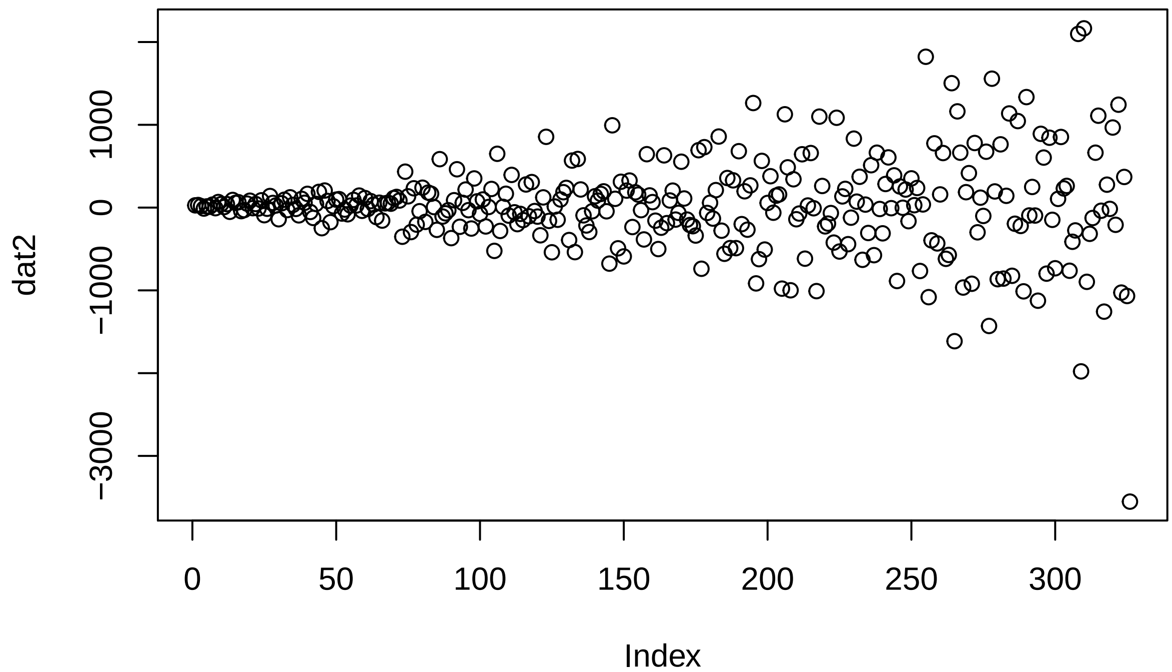
(a)

We should not expect the variance of submissions to remain constant over time, as even if the varinace of individual contributers remained constant over time, growth in the number of contributors would increase variance overall.

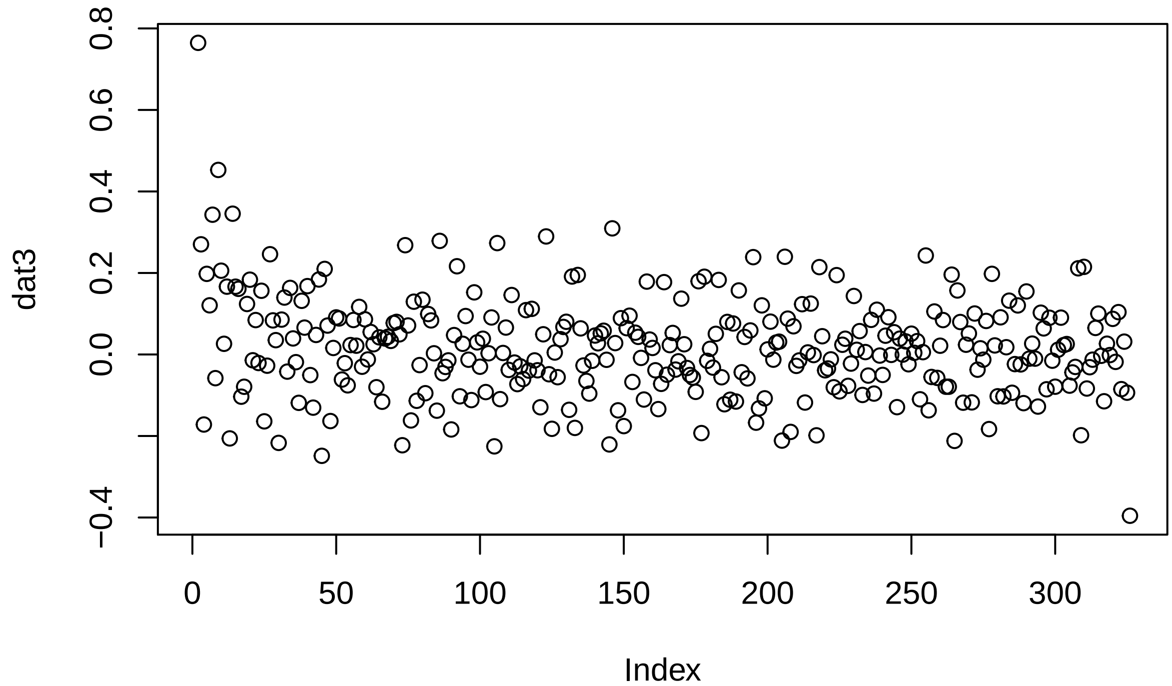
Since the variance of iid observations should increase by a function of expectation squared, a log transform should be a variance stabilizing transform of the data.

(b)

```
dat2 <- diff(dat$submissions)
dat3 <- diff(log(dat$submissions))

plot(dat2)
```




```
plot(dat3)
```



As we can see, the first order differenced data of the transformed series is much more like white noise than the difference of the untransformed data.

(c)

By standard differencing prediction,

$$x_{n+1} = e^{\bar{y}+log(x_n)}$$

```
tail(dat$submissions)
```

```
## [1] 11352 12595 11568 11938 10870  7318
```

```
log <- log(10870)
log
```

```
## [1] 9.293762
```

```
dat3 <- dat3[-1]
log+mean(dat3)
```

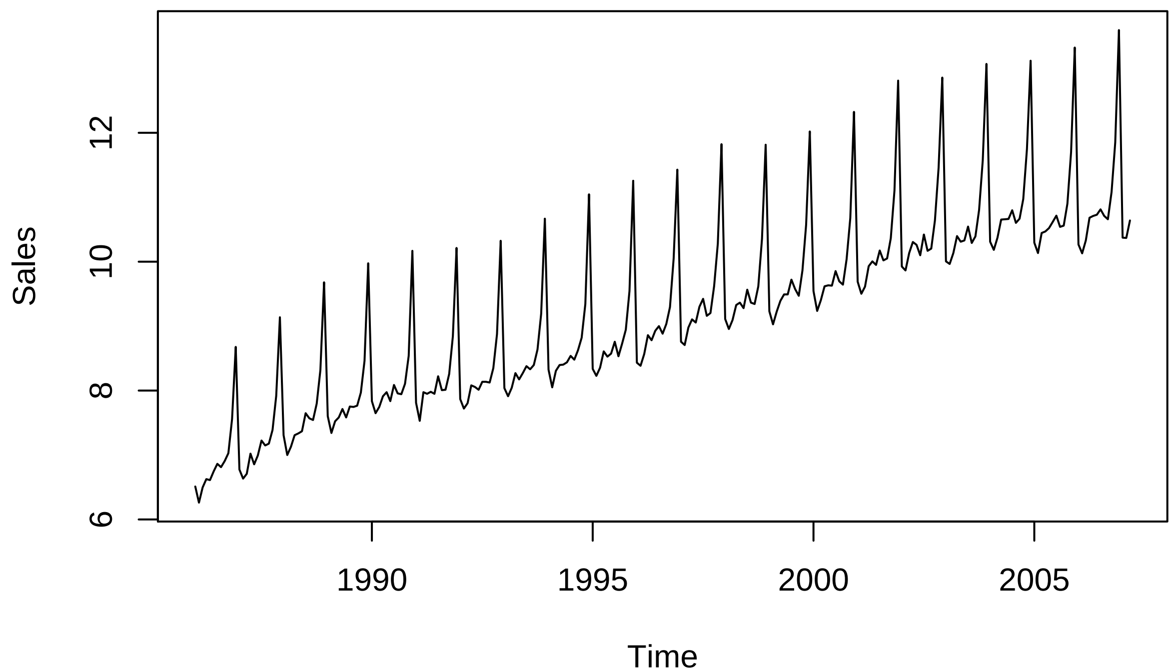
```
## [1] 9.311
```

```
exp(1)^(log+mean(dat3))
```

```
## [1] 11059
```

Question 2

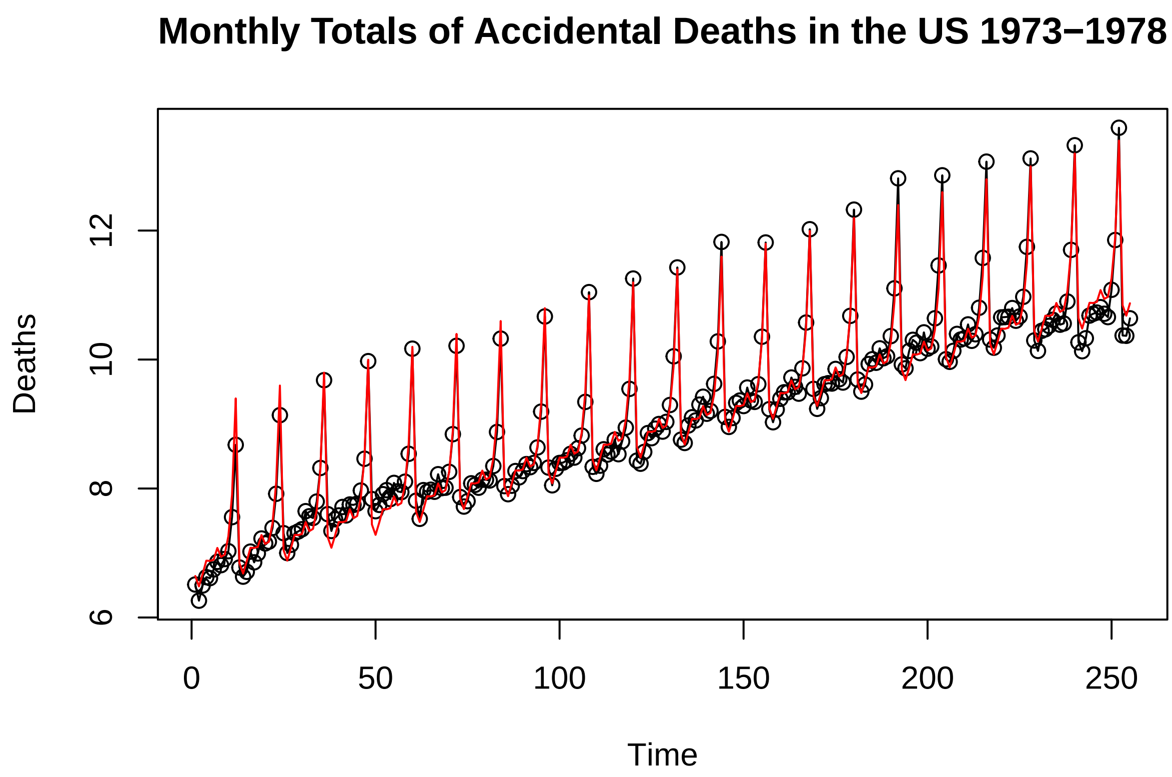
```
data(retail)
sqrt_retail <- (sqrt(retail))
plot(sqrt_retail)
```



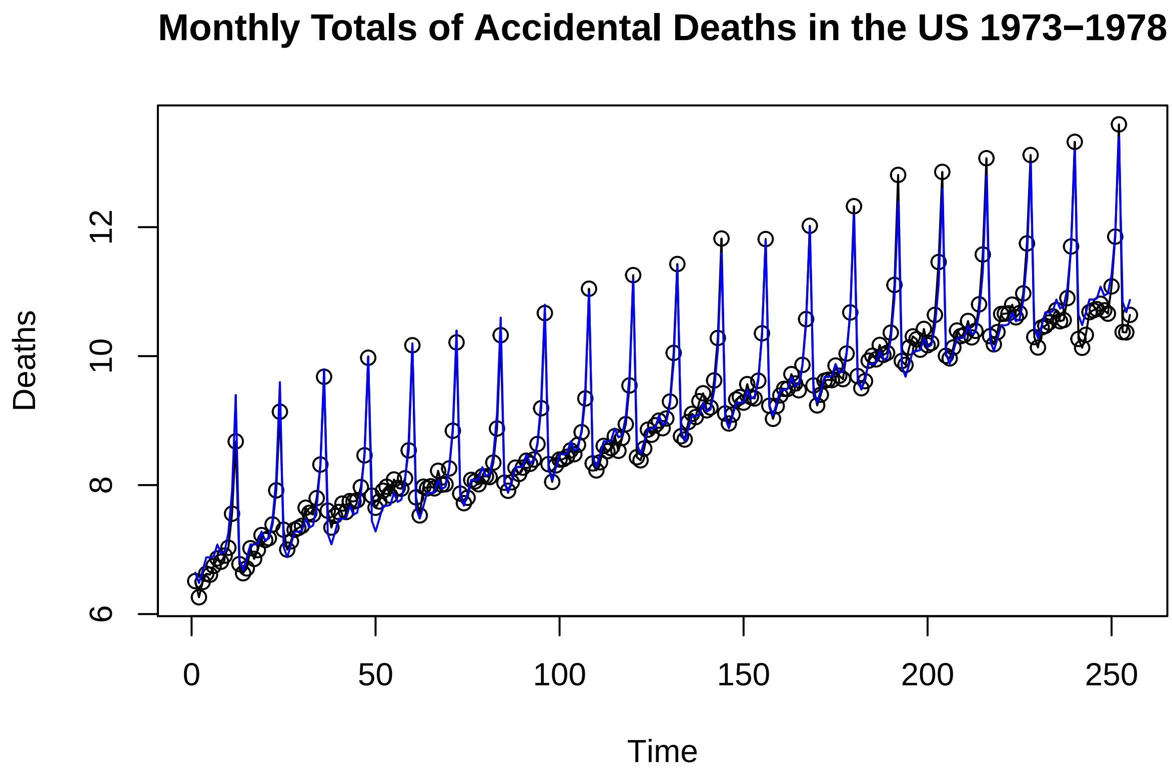
There is a very strong seasonal trend in the data, as well as a seemingly linear trend as well.

```
t = 1: length(sqrt_retail)
f1 = 1
f2 = 2
f3 = 3
f4 = 4
f5 = 5
f6 = 6

d = 12
v1 = cos(2*pi*f1*t/d)
v2 = sin(2*pi*f1*t/d)
v3 = cos(2*pi*f2*t/d)
v4 = sin(2*pi*f2*t/d)
v5 = cos(2*pi*f3*t/d)
v6 = sin(2*pi*f3*t/d)
v7 = cos(2*pi*f4*t/d)
v8 = sin(2*pi*f4*t/d)
v9 = cos(2*pi*f5*t/d)
```

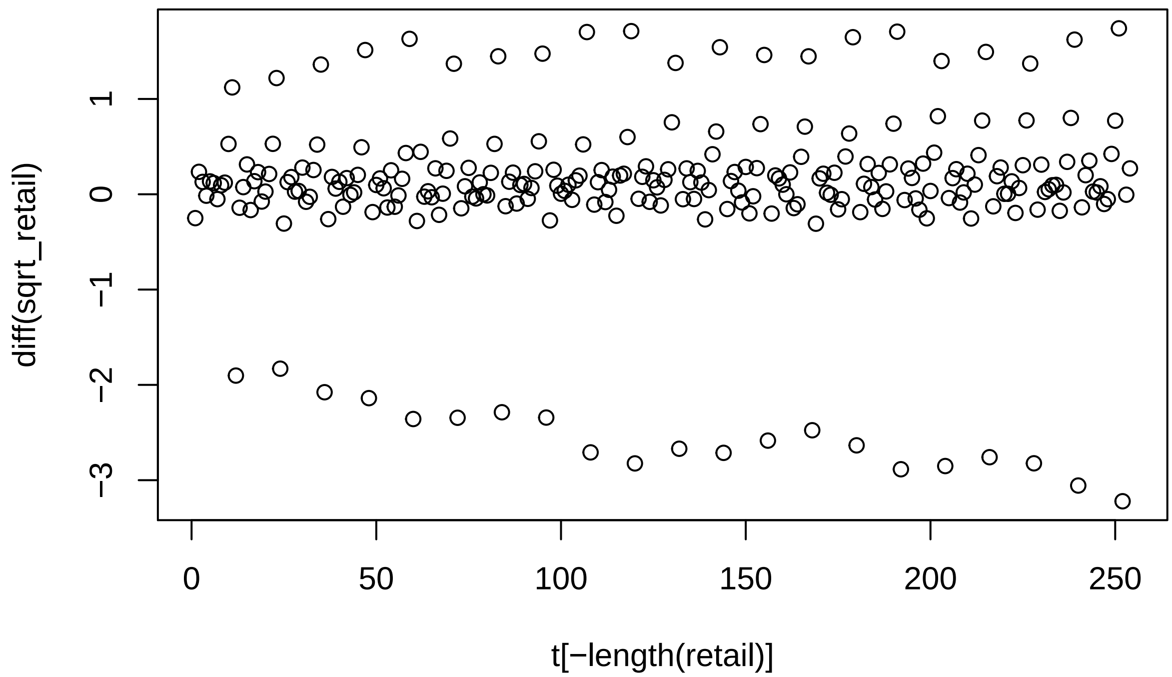
```
v10 = sin(2*pi*f5*t/d)
v11 = cos(2*pi*f6*t/d)

lin.mod = lm(sqrt_retail ~ 1 + t + v1 + v2 + v3 + v4 + v5 + v6 + v7 + v8 + v9 + v10 +v11)
plot(t, sqrt_retail, type = "o", xlab = "Time", ylab = "Deaths", main = "Monthly Totals of Accidental De
points(t, lin.mod$fitted, type = "l", col = "red")
```



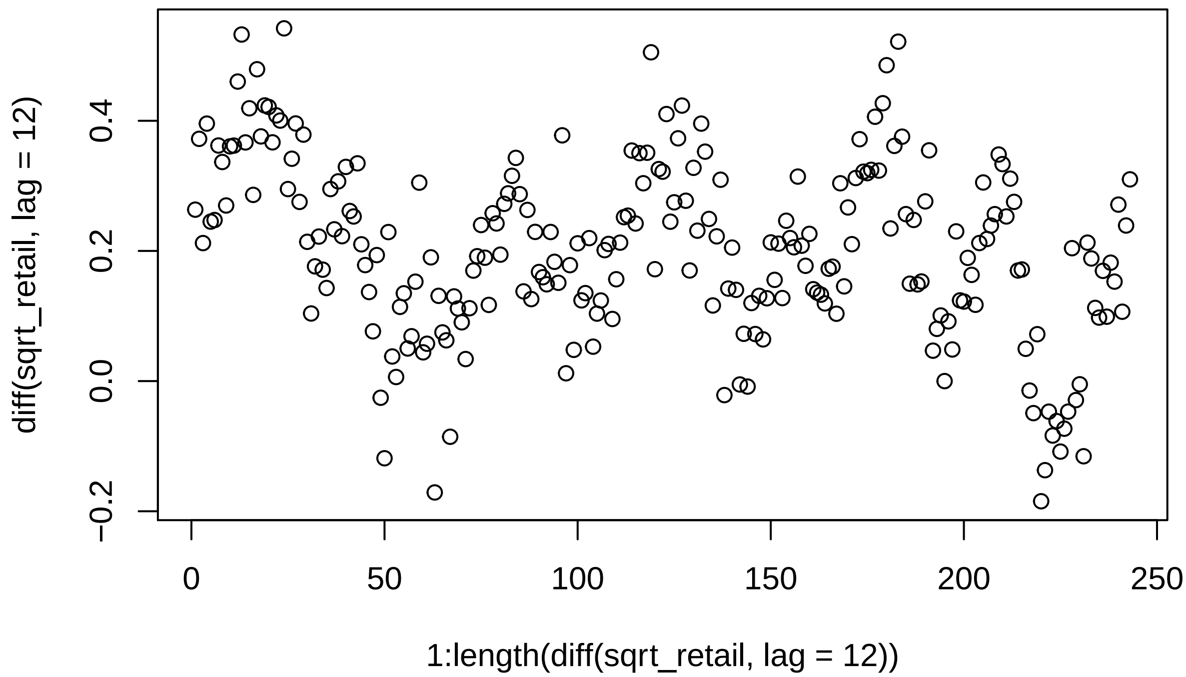

```
month <- seasonaldummy(sqrt_retail)
lin.mod2 = lm(sqrt_retail ~ month + t)
plot(t, sqrt_retail, type = "o", xlab = "Time", ylab = "Deaths", main = "Monthly Totals of Accidental De
points(t, lin.mod2$fitted, type = "l", col = "blue")
```



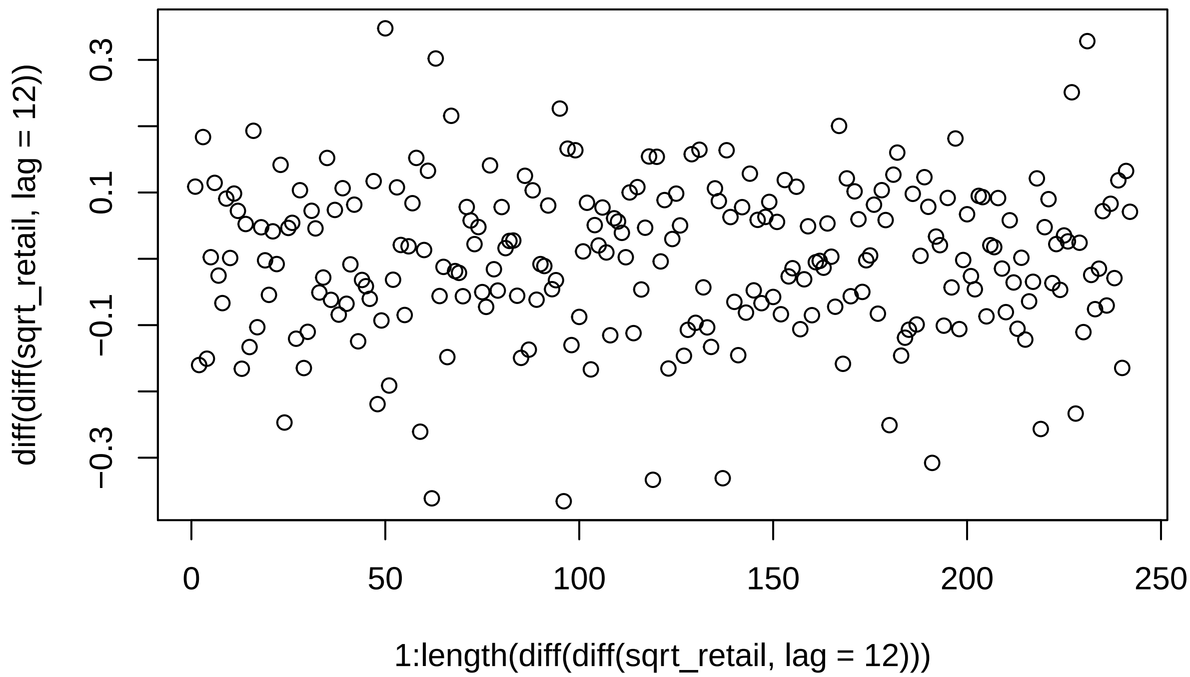

```
plot(x = t[-length(retail)], y = diff(sqrt_retail))
```




```
plot(x = 1:length(diff(sqrt_retail, lag = 12)), y = diff(sqrt_retail, lag = 12))
```




```
plot(x = 1:length(diff(diff(sqrt_retail, lag = 12))), y = diff(diff(sqrt_retail, lag = 12)))
```



iii. looks the most like white noise