analysis

Luca Comba

3/2/2021

Data Cleaning steps

Reading the Data

```
fargo_no_soil <- read.csv("C:/Users/Luca/Desktop/SPRING2021/Statistical Research/Combined data/fargo.csv", encoding="UTF-8")
grand_no_soil <- read.csv("C:/Users/Luca/Desktop/SPRING2021/Statistical Research/Combined data/grand.csv", encoding="UTF-8")

fargo_from_2014 <- read.csv("C:/Users/Luca/Desktop/SPRING2021/Statistical Research/Combined data/fargo_from_2014.csv", encoding="UTF-8")
grand_from_2014 <- read.csv("C:/Users/Luca/Desktop/SPRING2021/Statistical Research/Combined data/grand_from_2014.csv", encoding="UTF-8")
```

adding the data with libridate

```
fargo_no_soil$date <- make_date(year=fargo_no_soil$year, month = fargo_no_soil$month,day=fargo_no_soil$day)

grand_no_soil$date <- make_date(year=grand_no_soil$year, month =grand_no_soil$month,day=grand_no_soil$day)
```

Time Series Analysis

First, let's create date

```
grand_from_2014$date <- make_date(year=grand_from_2014$year, month =grand_from_2014$month,day=grand_from_2014$day)

fargo_from_2014$date <- make_date(year=fargo_from_2014$year, month =fargo_from_2014$month,day=fargo_from_2014$day)
```

After reading the data we are creating a Time Series object.

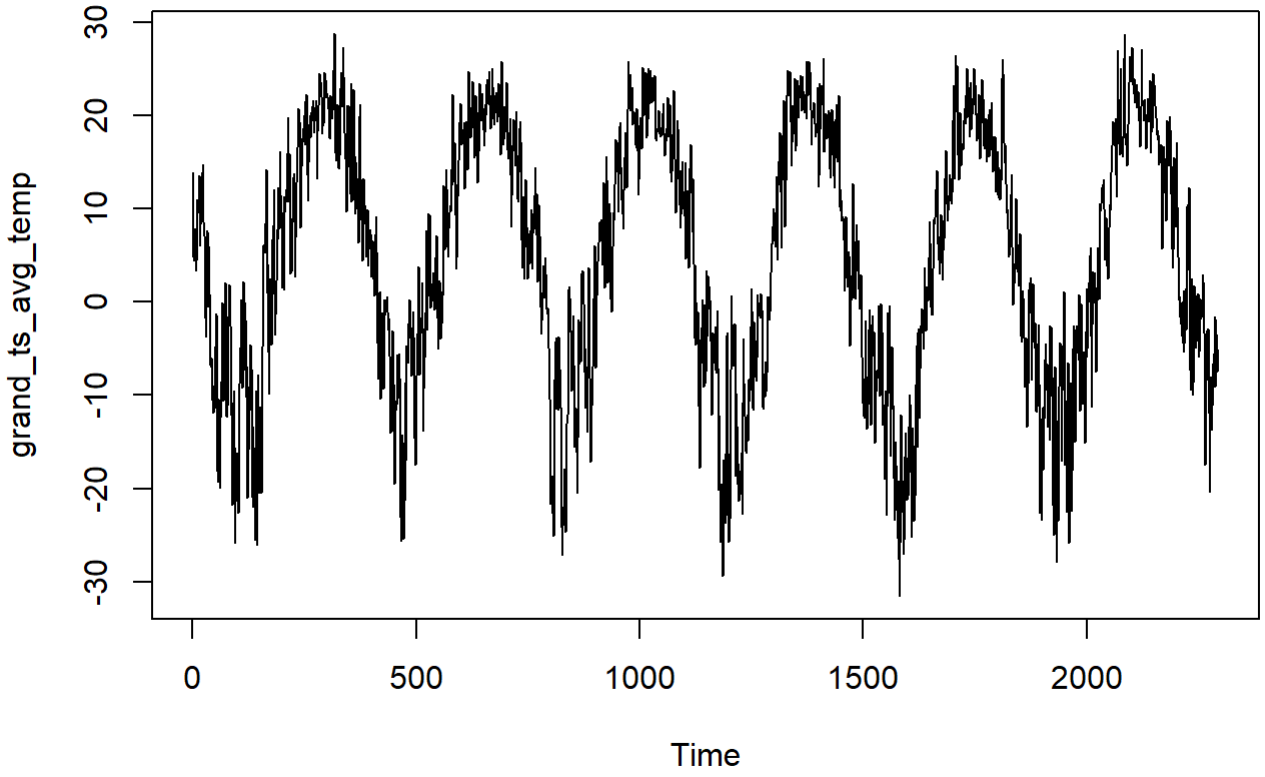
Terms to remember p regular autoregressive terms d regular differences q moving average terms P seasonal autoregressive terms D seasonal differences Q seasonal moving average terms

Average Temperature

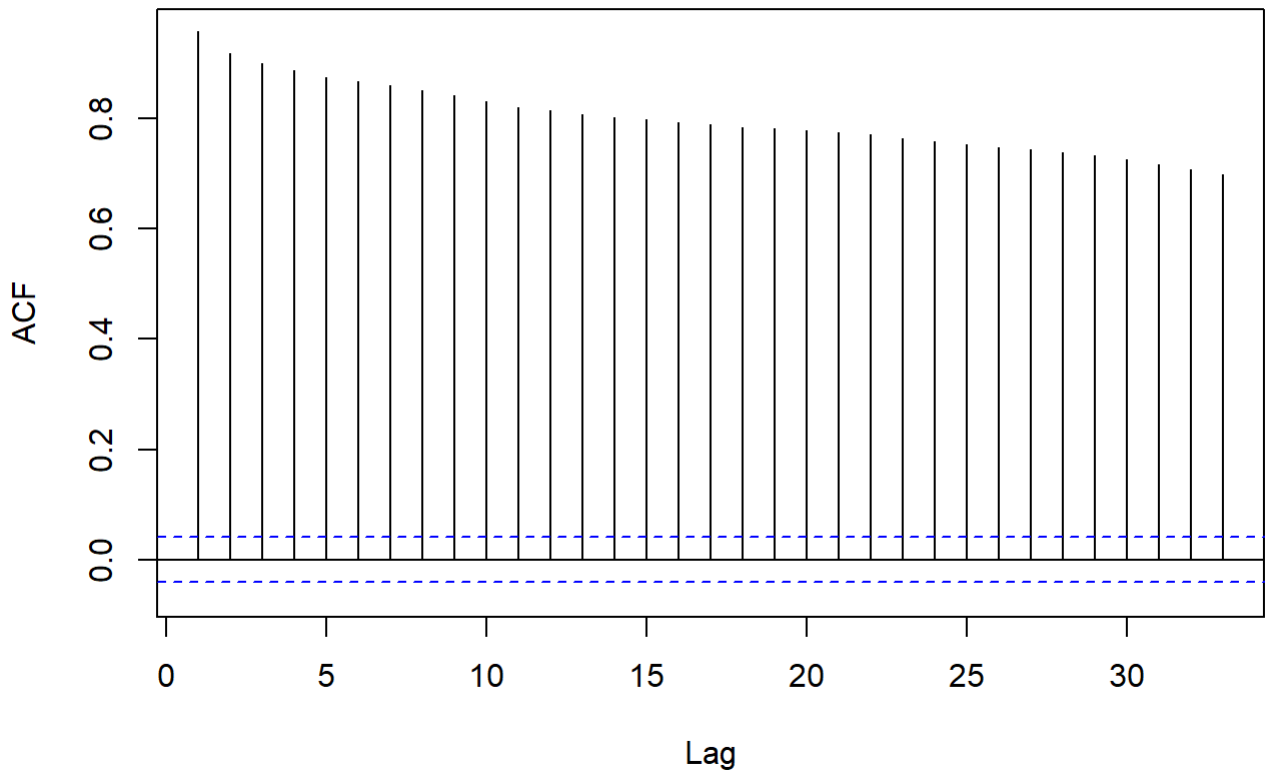
```
grand_ts_avg_temp <- ts(grand_from_2014$avg.temp)
fargo_ts_avg_temp <- ts(fargo_from_2014$avg.temp)

class(grand_ts_avg_temp)
```

```
## [1] "ts"
```

```
plot(grand_ts_avg_temp)
```



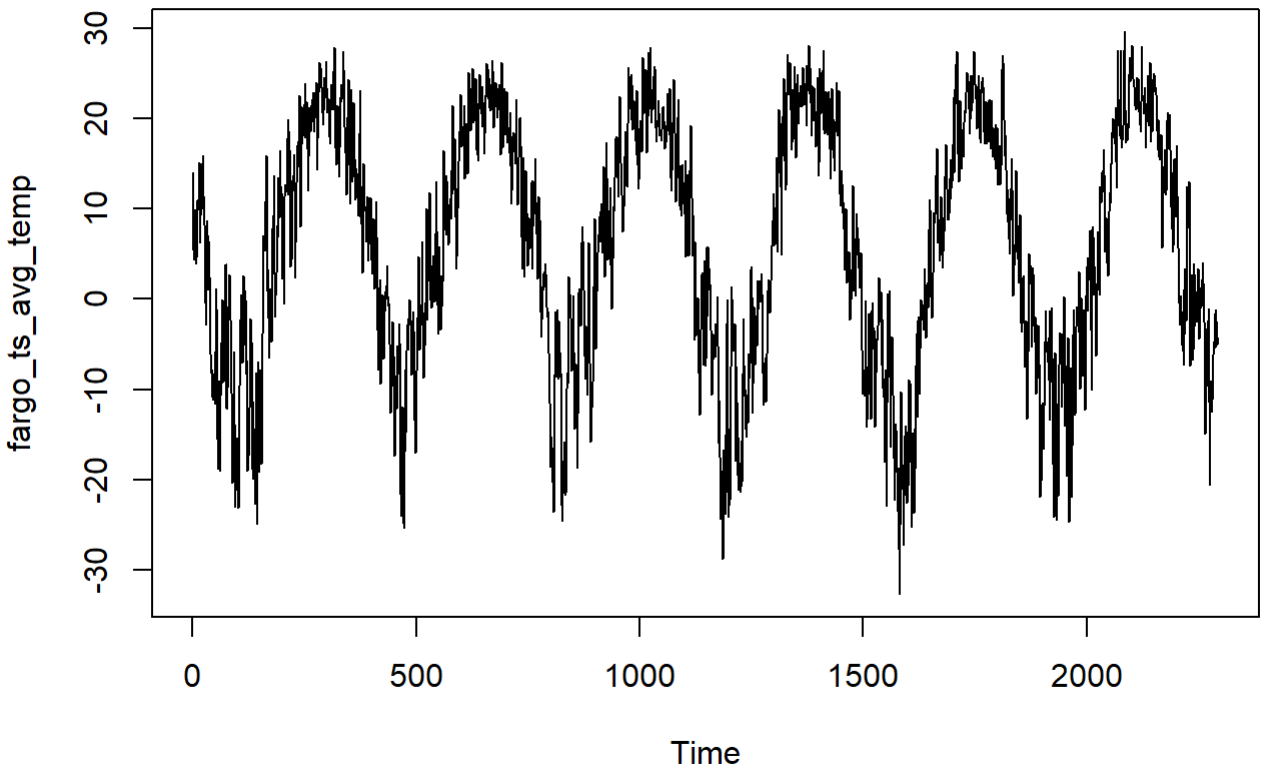

```
Acf(grand_ts_avg_temp)
```



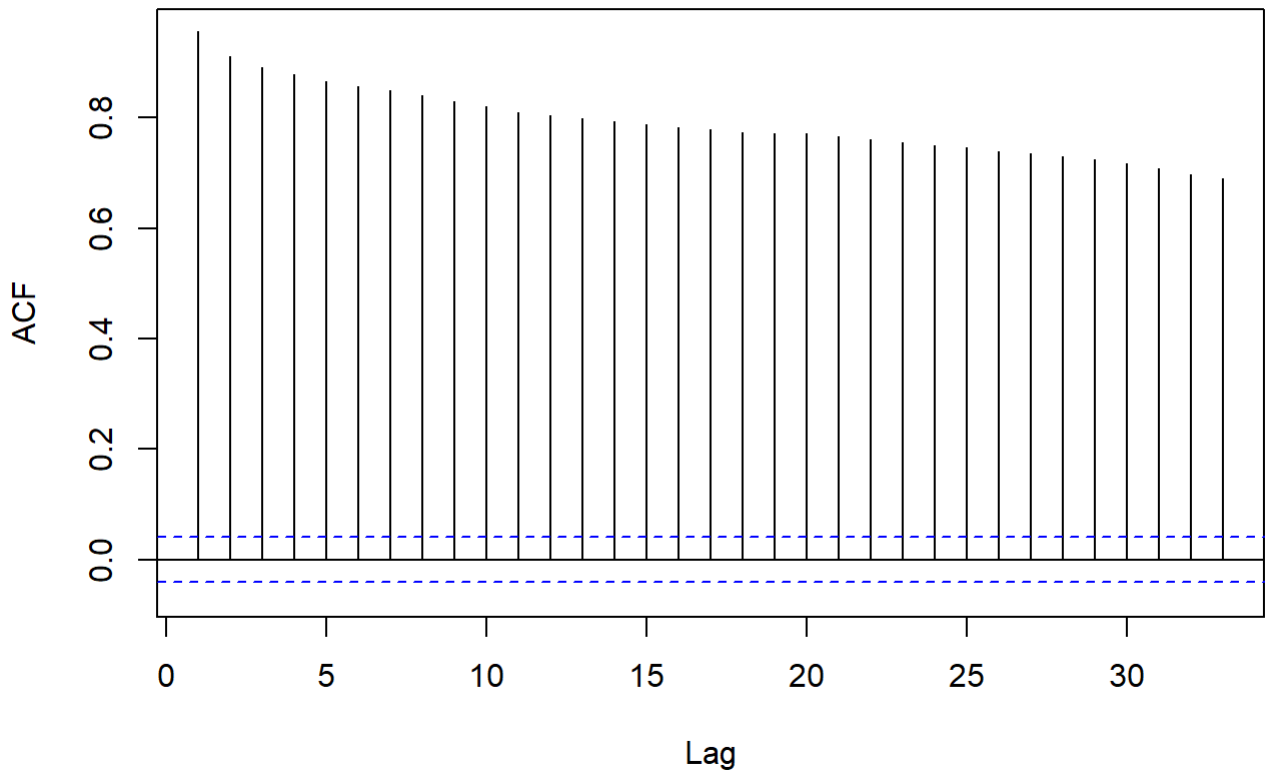

```
class(fargo_ts_avg_temp)
```

```
## [1] "ts"
```

```
plot(fargo_ts_avg_temp)
```




```
Acf(fargo_ts_avg_temp)
```




```
g_temp <- Arima(grand_ts_avg_temp, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), inc
lude.constant = TRUE)
g_temp
```

```
## Series: grand_ts_avg_temp
## ARIMA(0,0,1) with drift
##
## Coefficients:
##           ma1    sar1   drift
##       -0.8792  0.6785  0.0032
## s.e.   0.0164  0.0267  0.0295
##
## sigma^2 estimated as 14.03:  log likelihood=-6280.51
## AIC=12569.03   AICc=12569.05   BIC=12591.98
```

```
coeftest(g_temp)
```

```
##
## z test of coefficients:
##
##         Estimate Std. Error  z value Pr(>|z|)
## ma1   -0.8792456  0.0163713 -53.7066   <2e-16 ***
## sar1   0.6784887  0.0267049  25.4069   <2e-16 ***
## drift  0.0031753  0.0294714   0.1077   0.9142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
f_temp <- Arima(fargo_ts_avg_temp, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), inc
lude.constant = TRUE)
f_temp
```

```
## Series: fargo_ts_avg_temp
## ARIMA(0,0,1) with drift
##
## Coefficients:
##          ma1     sar1   drift
##       0.6989  -0.5787  0.0040
## s.e.  0.0497   0.0581  0.0874
##
## sigma^2 estimated as 15.13:  log likelihood=-6366.78
## AIC=12741.57   AICc=12741.58   BIC=12764.52
```

```
coeftest(f_temp)
```

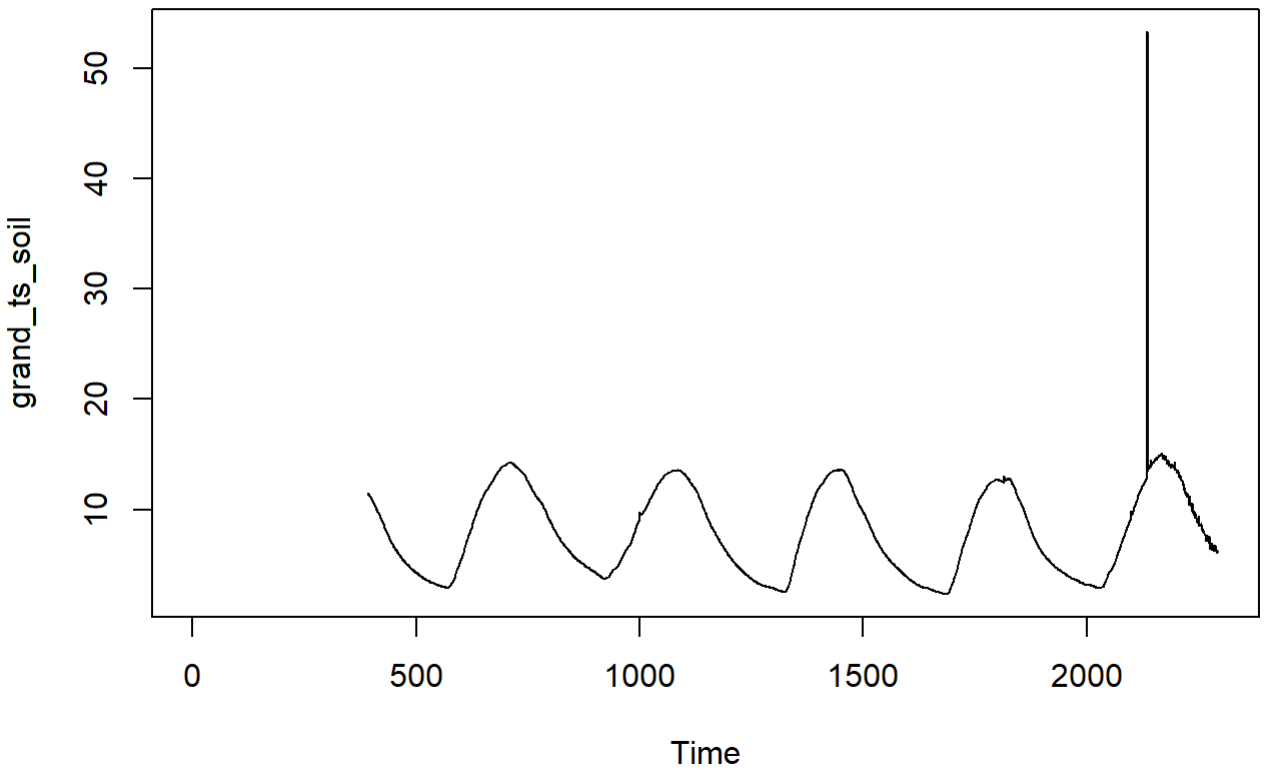
```
##
## z test of coefficients:
##
##         Estimate Std. Error z value Pr(>|z|)
## ma1    0.6989354  0.0497213 14.0571   <2e-16 ***
## sar1  -0.5786983  0.0581328 -9.9548   <2e-16 ***
## drift  0.0039762  0.0873578  0.0455   0.9637
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Soil Temperature

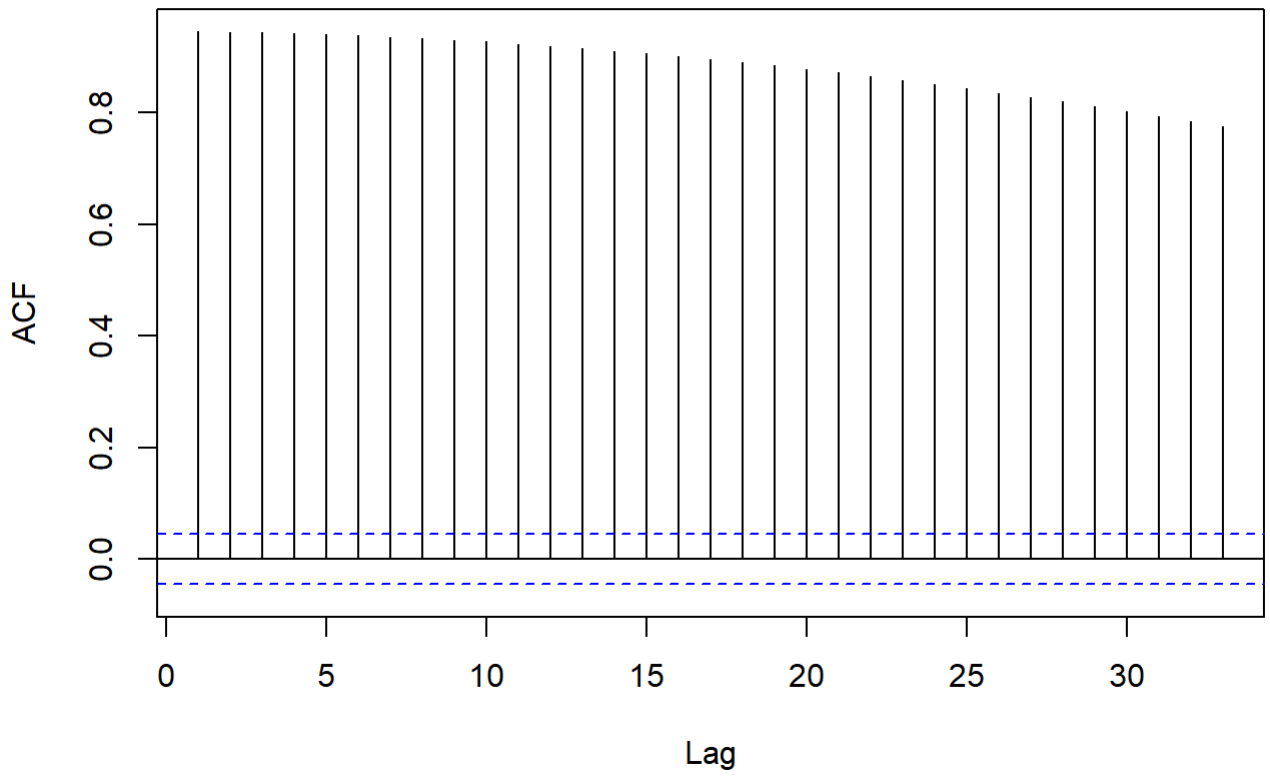
```
grand_ts_soil <- ts(grand_from_2014$soil.200)
fargo_ts_soil <- ts(fargo_from_2014$soil.200)

class(grand_ts_soil)
```

```
## [1] "ts"
```

```
plot(grand_ts_soil)
```



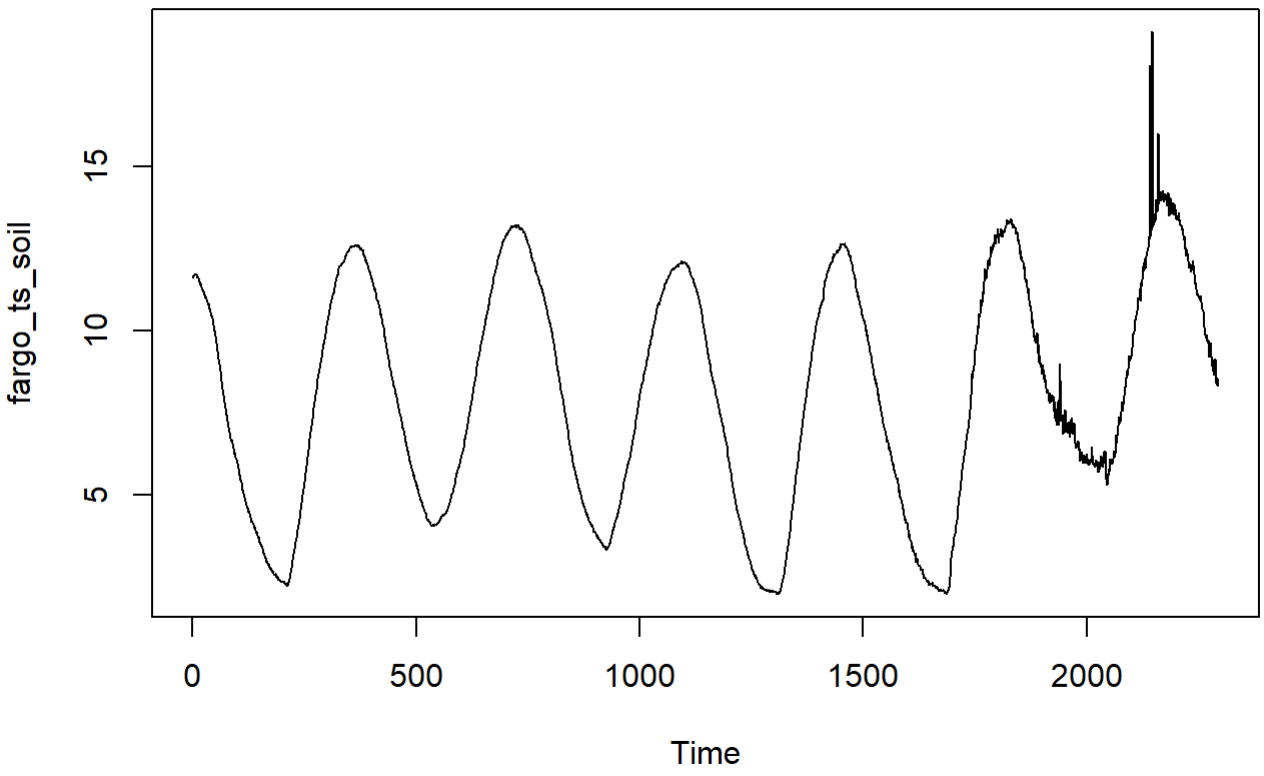

```
Acf(grand_ts_soil)
```



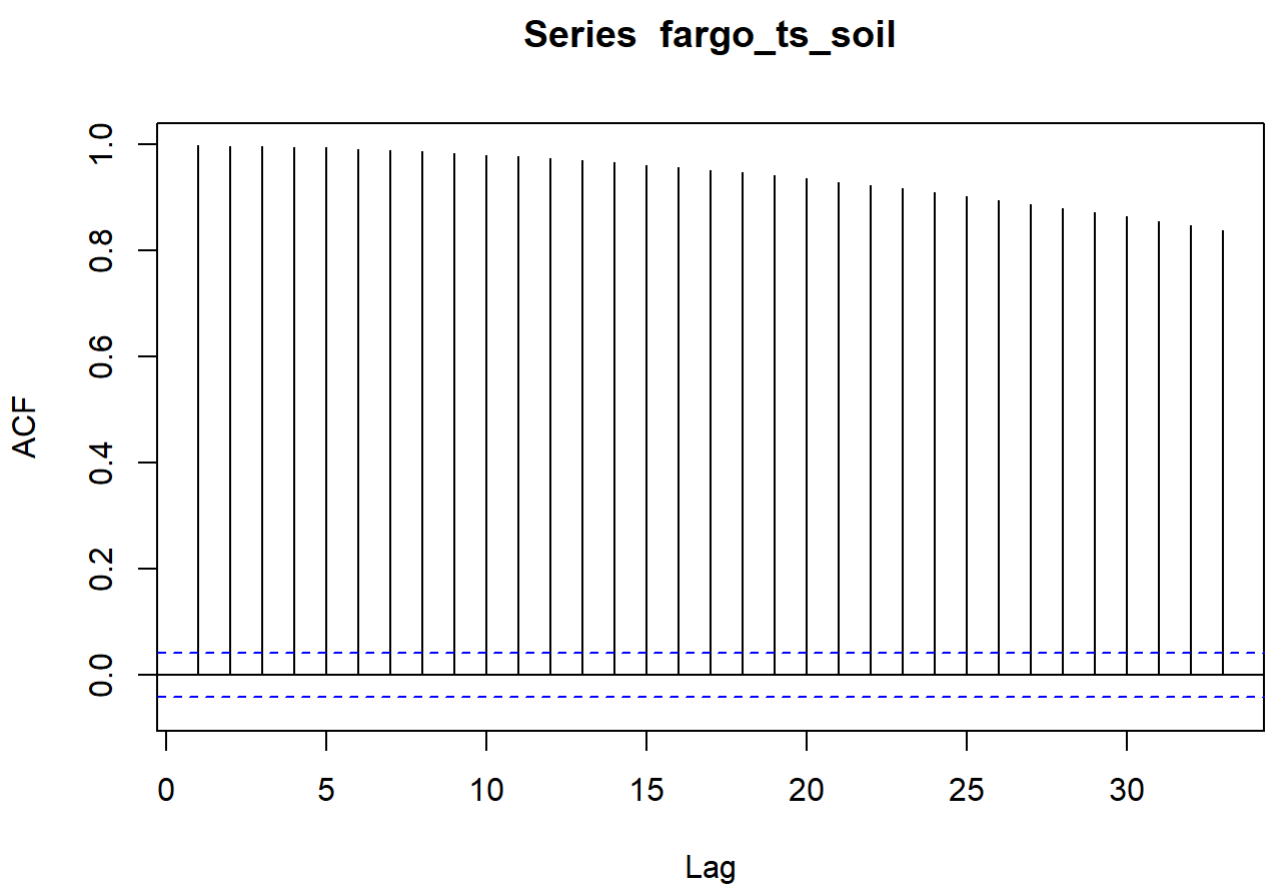

```
class(fargo_ts_soil)
```

```
## [1] "ts"
```

```
plot(fargo_ts_soil)
```




```
Acf(fargo_ts_soil)
```




```
g_soil <- Arima(grand_ts_soil, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), includ
e.constant = TRUE)
g_soil
```

```
## Series: grand_ts_soil
## ARIMA(0,0,1) with drift
##
## Coefficients:
##           ma1     sar1   drift
##       -0.7132  -0.0775  0.0047
## s.e.   0.0170   0.0279  0.0063
##
## sigma^2 estimated as 1.048:  log likelihood=-2745.43
## AIC=5498.85   AICc=5498.87   BIC=5521.06
```

```
coeftest(g_soil)
```

```
##
## z test of coefficients:
##
##         Estimate Std. Error  z value  Pr(>|z|)
## ma1   -0.7132175  0.0169914 -41.9753 < 2.2e-16 ***
## sar1  -0.0774634  0.0279067  -2.7758  0.005507 **
## drift  0.0047178  0.0062948   0.7495  0.453570
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
f_soil <- Arima(fargo_ts_soil, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), includ
e.constant = TRUE)
f_soil
```

```
## Series: fargo_ts_soil
## ARIMA(0,0,1) with drift
##
## Coefficients:
##           ma1     sar1   drift
##       -0.2985  -0.1289  0.0022
## s.e.   0.0352   0.0392  0.0032
##
## sigma^2 estimated as 0.0589:  log likelihood=-5.39
## AIC=18.79   AICc=18.81   BIC=41.74
```

```
coeftest(f_soil)
```

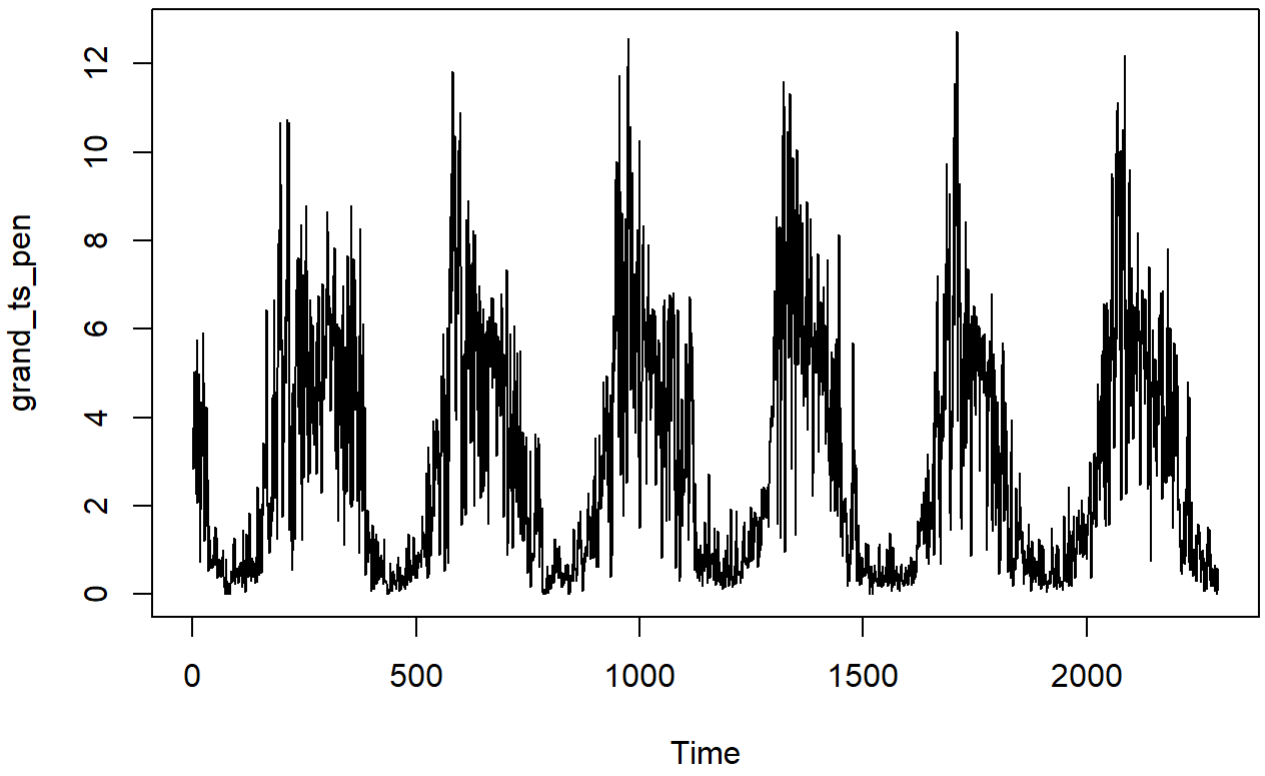
```
##
## z test of coefficients:
##
##         Estimate Std. Error z value  Pr(>|z|)
## ma1   -0.2984586  0.0351882 -8.4818 < 2.2e-16 ***
## sar1  -0.1288951  0.0392152 -3.2869  0.001013 **
## drift  0.0021890  0.0031564  0.6935  0.487986
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

PENman

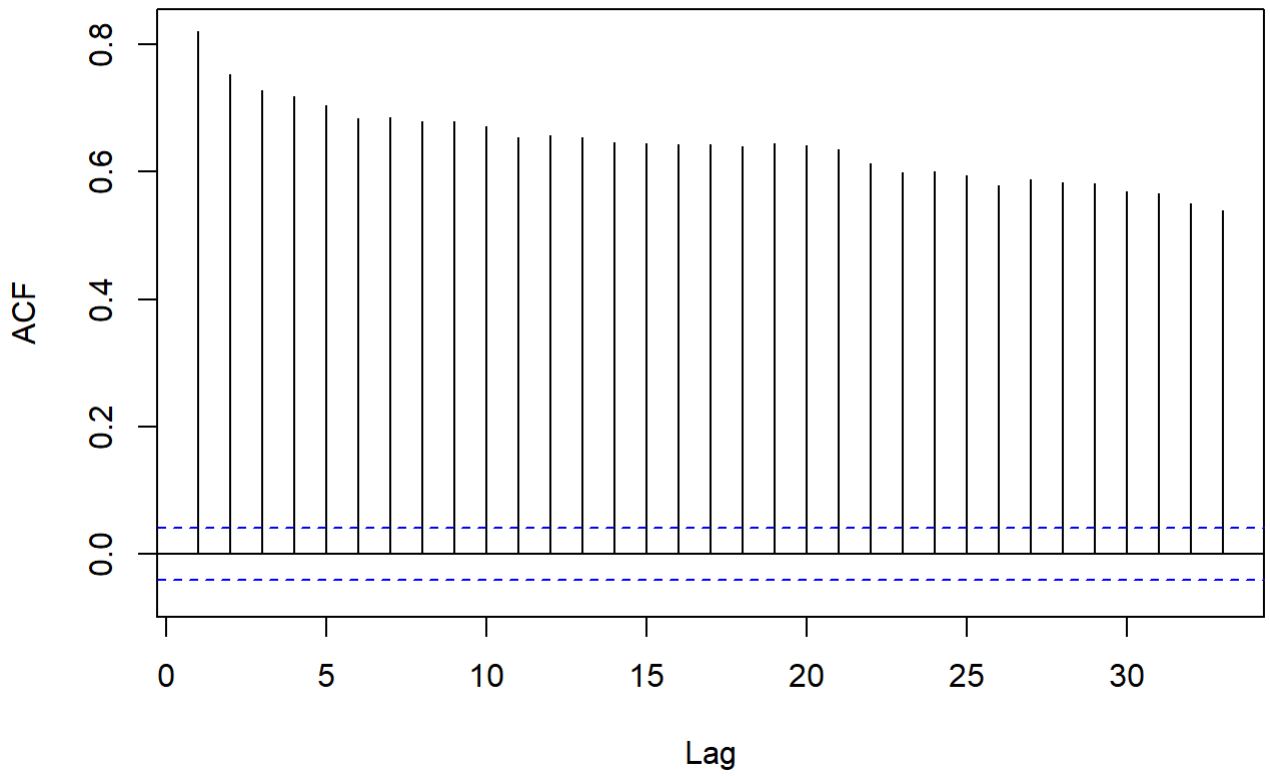
```
grand_ts_pen <- ts(grand_from_2014$penman.pet)
fargo_ts_pen <- ts(fargo_from_2014$penman.pet)

class(grand_ts_pen)
```

```
## [1] "ts"
```

```
plot(grand_ts_pen)
```



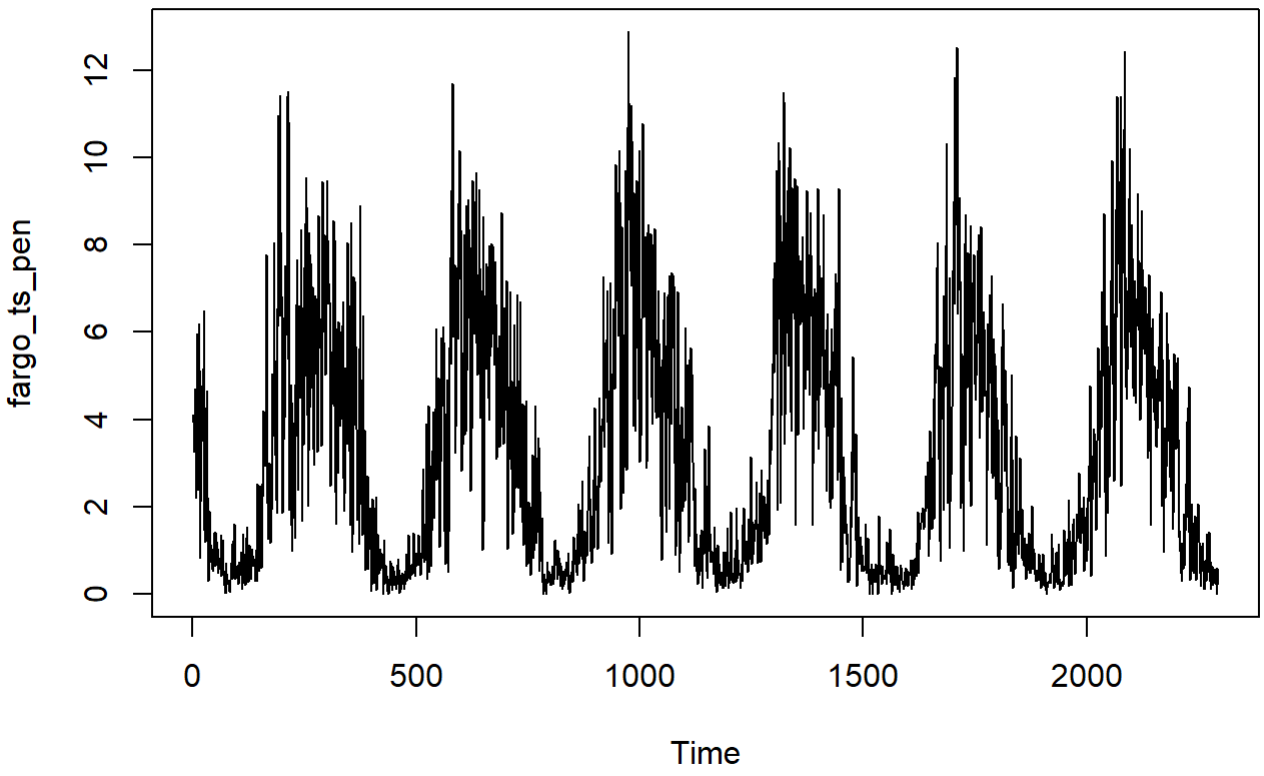

```
Acf(grand_ts_pen)
```



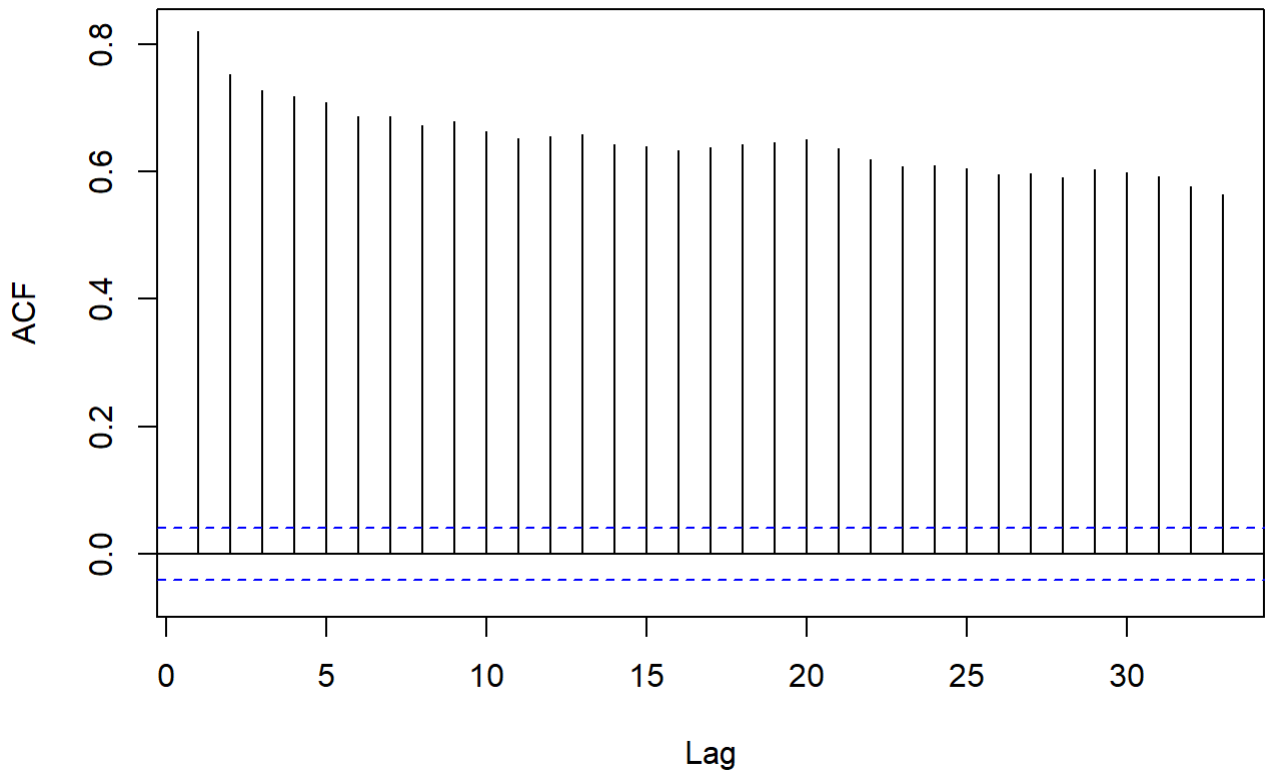

```
class(fargo_ts_pen)
```

```
## [1] "ts"
```

```
plot(fargo_ts_pen)
```




```
Acf(fargo_ts_pen)
```




```
g_temp <- Arima(grand_ts_avg_temp, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), inc
lude.constant = TRUE)
g_temp
```

```
## Series: grand_ts_avg_temp
## ARIMA(0,0,1) with drift
##
## Coefficients:
##           ma1    sar1   drift
##       -0.8792  0.6785  0.0032
## s.e.   0.0164  0.0267  0.0295
##
## sigma^2 estimated as 14.03:  log likelihood=-6280.51
## AIC=12569.03   AICc=12569.05   BIC=12591.98
```

```
coeftest(g_temp)
```

```
##
## z test of coefficients:
##
##         Estimate Std. Error  z value Pr(>|z|)
## ma1   -0.8792456  0.0163713 -53.7066   <2e-16 ***
## sar1   0.6784887  0.0267049  25.4069   <2e-16 ***
## drift  0.0031753  0.0294714   0.1077   0.9142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
f_temp <- Arima(fargo_ts_avg_temp, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), inc
lude.constant = TRUE)
f_temp
```

```
## Series: fargo_ts_avg_temp
## ARIMA(0,0,1) with drift
##
## Coefficients:
##          ma1     sar1   drift
##       0.6989  -0.5787  0.0040
## s.e.  0.0497   0.0581  0.0874
##
## sigma^2 estimated as 15.13:  log likelihood=-6366.78
## AIC=12741.57   AICc=12741.58   BIC=12764.52
```

```
coeftest(f_temp)
```

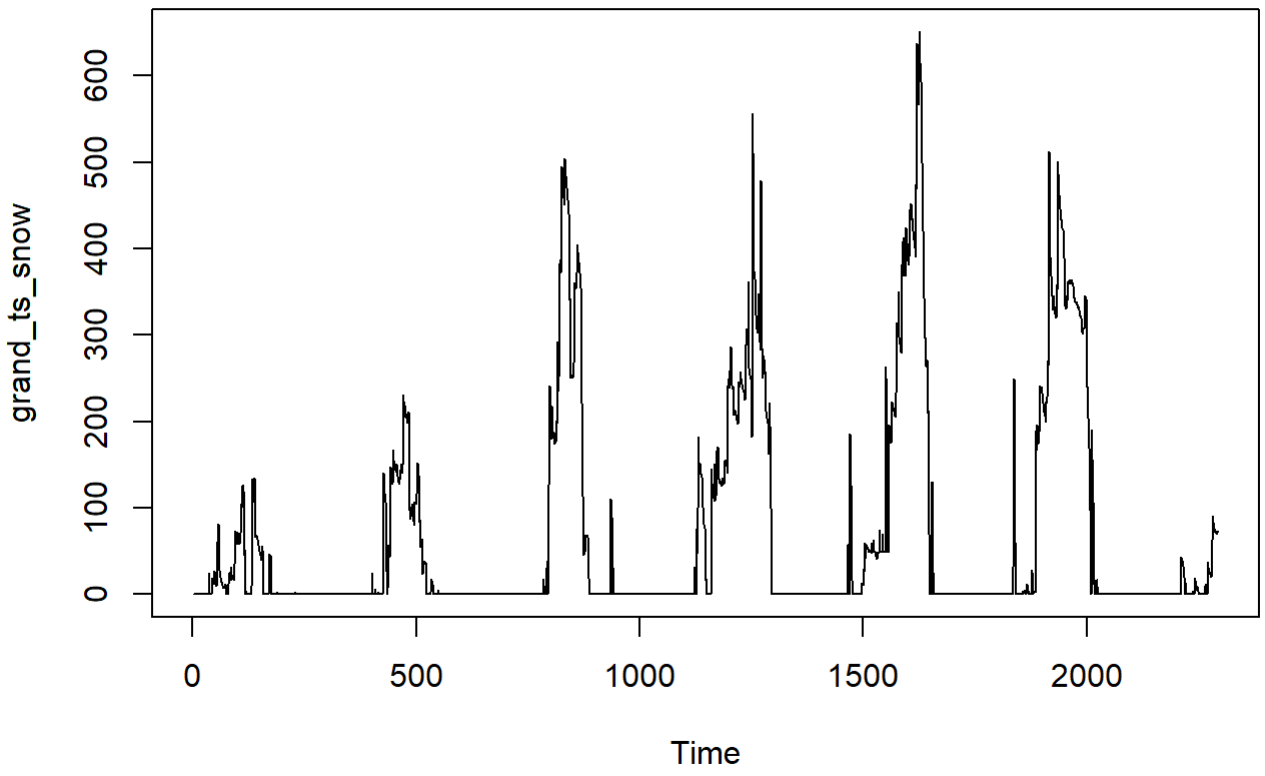
```
##
## z test of coefficients:
##
##         Estimate Std. Error z value Pr(>|z|)
## ma1    0.6989354  0.0497213 14.0571   <2e-16 ***
## sar1  -0.5786983  0.0581328 -9.9548   <2e-16 ***
## drift  0.0039762  0.0873578  0.0455   0.9637
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Snow depth

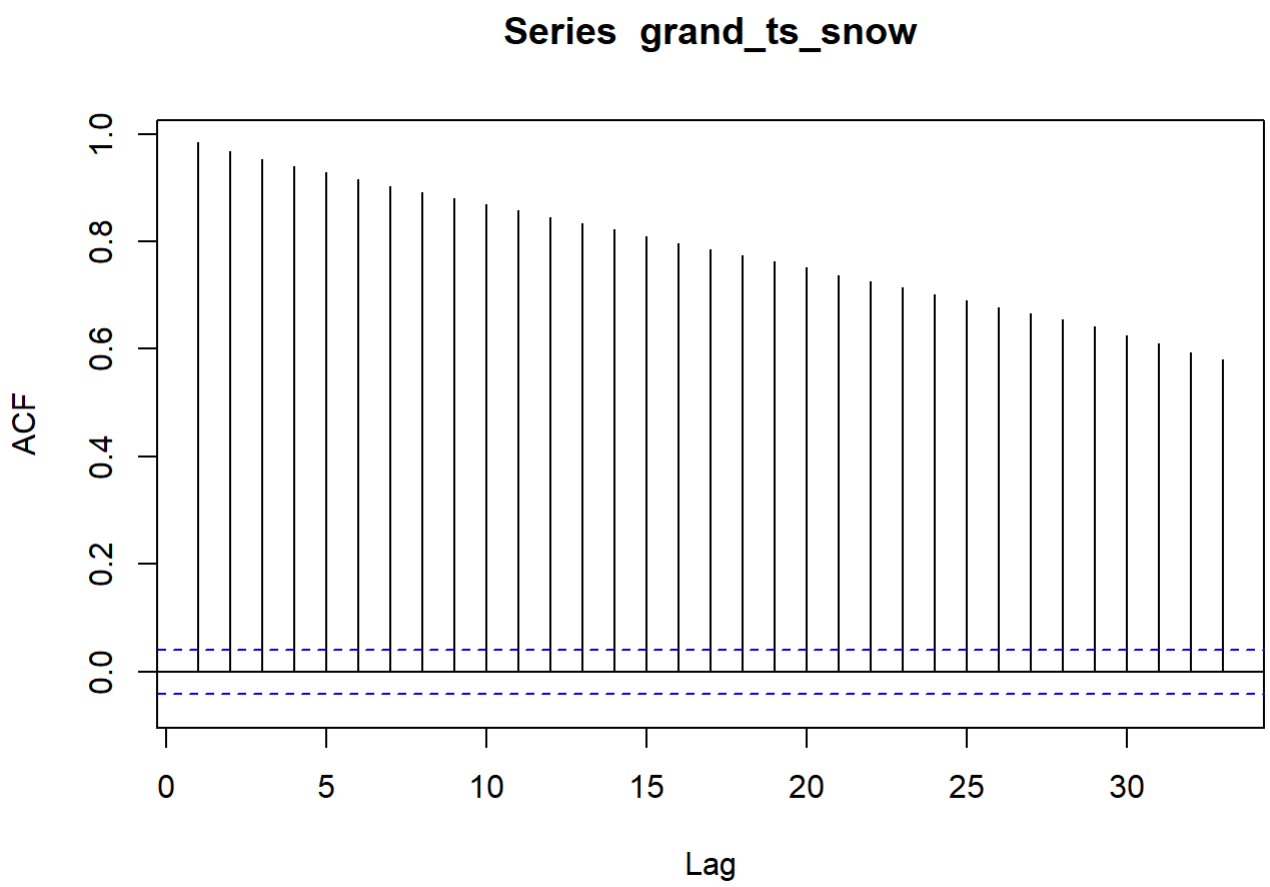
```
grand_ts_snow <- ts(grand_from_2014$snow.depth)
fargo_ts_snow <- ts(fargo_from_2014$snow.depth)

class(grand_ts_snow)
```

```
## [1] "ts"
```

```
plot(grand_ts_snow)
```



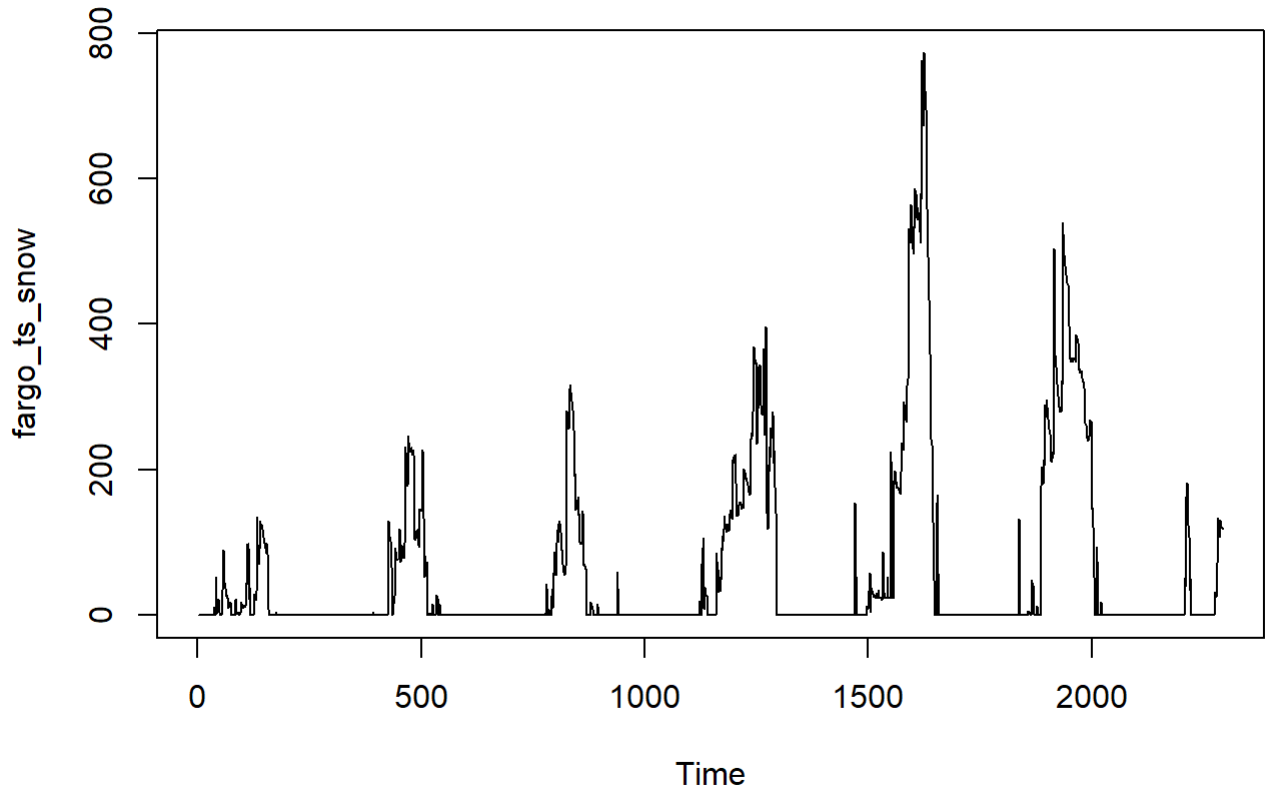

```
Acf(grand_ts_snow)
```



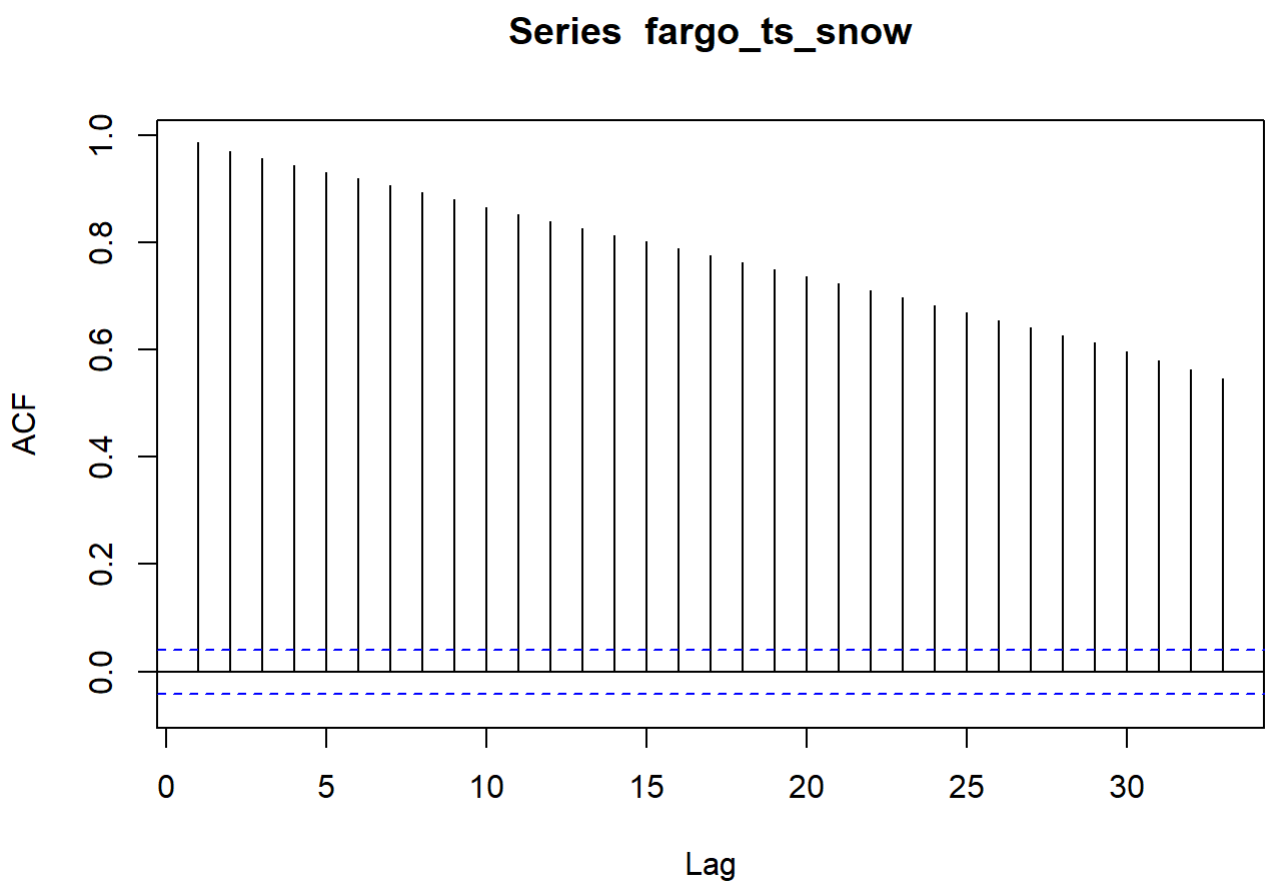

```
class(fargo_ts_snow)
```

```
## [1] "ts"
```

```
plot(fargo_ts_snow)
```




```
Acf(fargo_ts_snow)
```




```
g_snow <- Arima(grand_ts_snow, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), includ
e.constant = TRUE)
g_snow
```

```
## Series: grand_ts_snow
## ARIMA(0,0,1) with drift
##
## Coefficients:
##          ma1     sar1   drift
##       0.7444  -0.7082  0.0525
## s.e.  0.1277   0.1355  0.4813
##
## sigma^2 estimated as 510:  log likelihood=-10399.94
## AIC=20807.88   AICc=20807.89   BIC=20830.83
```

```
coeftest(g_snow)
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error z value  Pr(>|z|)
## ma1    0.744399   0.127675  5.8304 5.529e-09 ***
## sar1  -0.708218   0.135536 -5.2253 1.738e-07 ***
## drift  0.052547   0.481295  0.1092    0.9131
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
f_snow <- Arima(fargo_ts_snow, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), includ
e.constant = TRUE)
f_snow
```

```
## Series: fargo_ts_snow
## ARIMA(0,0,1) with drift
##
## Coefficients:
##          ma1     sar1   drift
##       0.2948  -0.2192  0.0481
## s.e.  0.1438   0.1479  0.4498
##
## sigma^2 estimated as 411.9:  log likelihood=-10155.09
## AIC=20318.17   AICc=20318.19   BIC=20341.12
```

```
coeftest(f_snow)
```

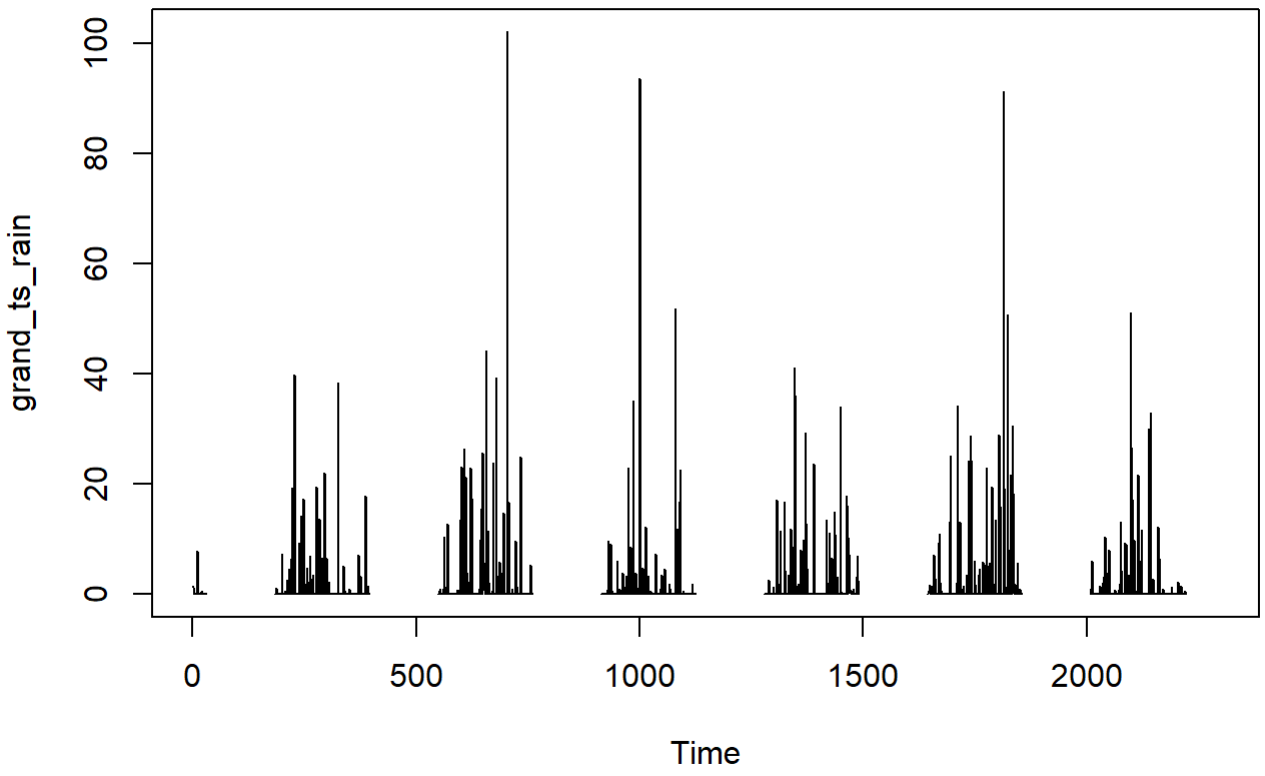
```
##
## z test of coefficients:
##
##        Estimate Std. Error z value Pr(>|z|)
## ma1    0.294755   0.143816  2.0495  0.04041 *
## sar1  -0.219214   0.147930 -1.4819  0.13837
## drift  0.048147   0.449835  0.1070  0.91476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Rainfall

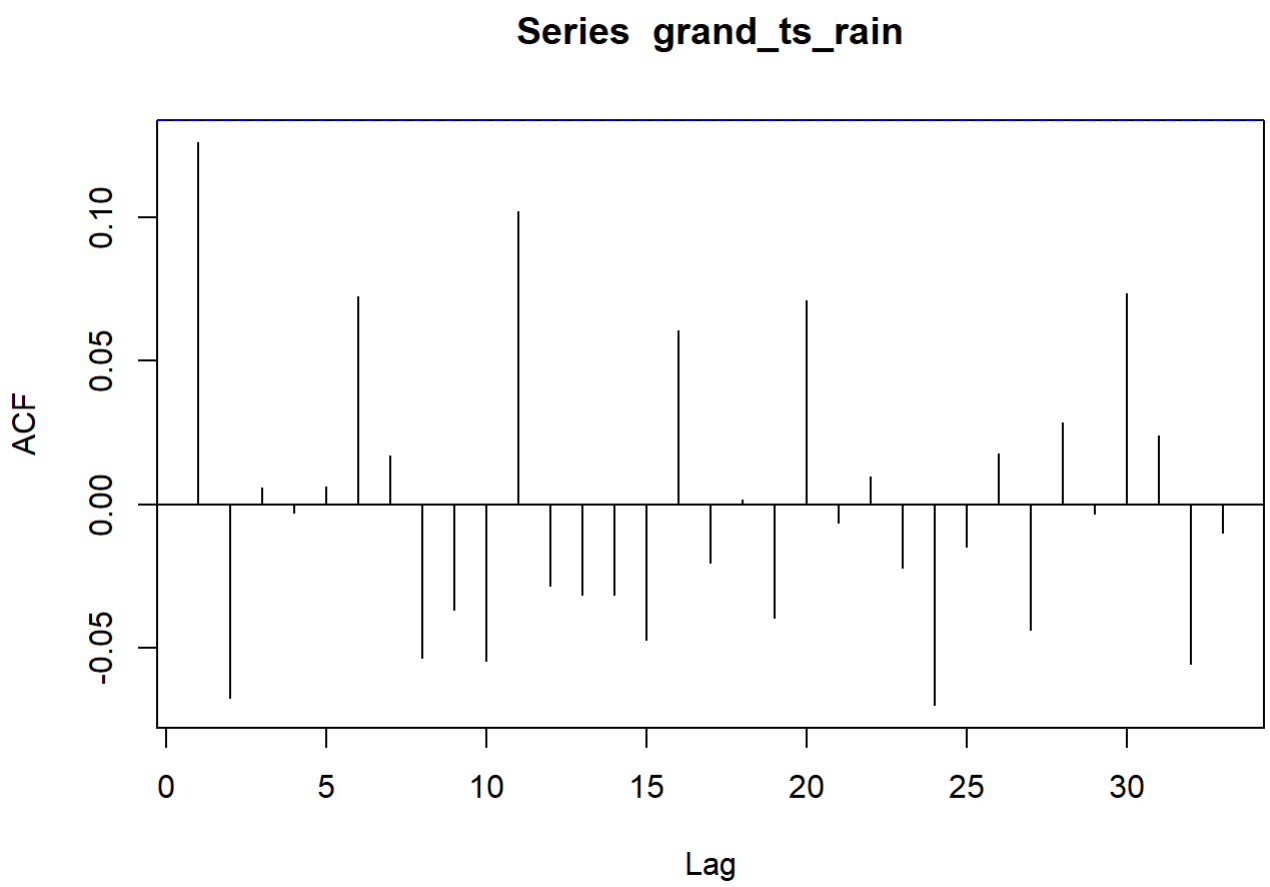
```
grand_ts_rain <- ts(grand_from_2014$rainfall)
fargo_ts_rain <- ts(fargo_from_2014$rainfall)

class(grand_ts_rain)
```

```
## [1] "ts"
```

```
plot(grand_ts_rain)
```



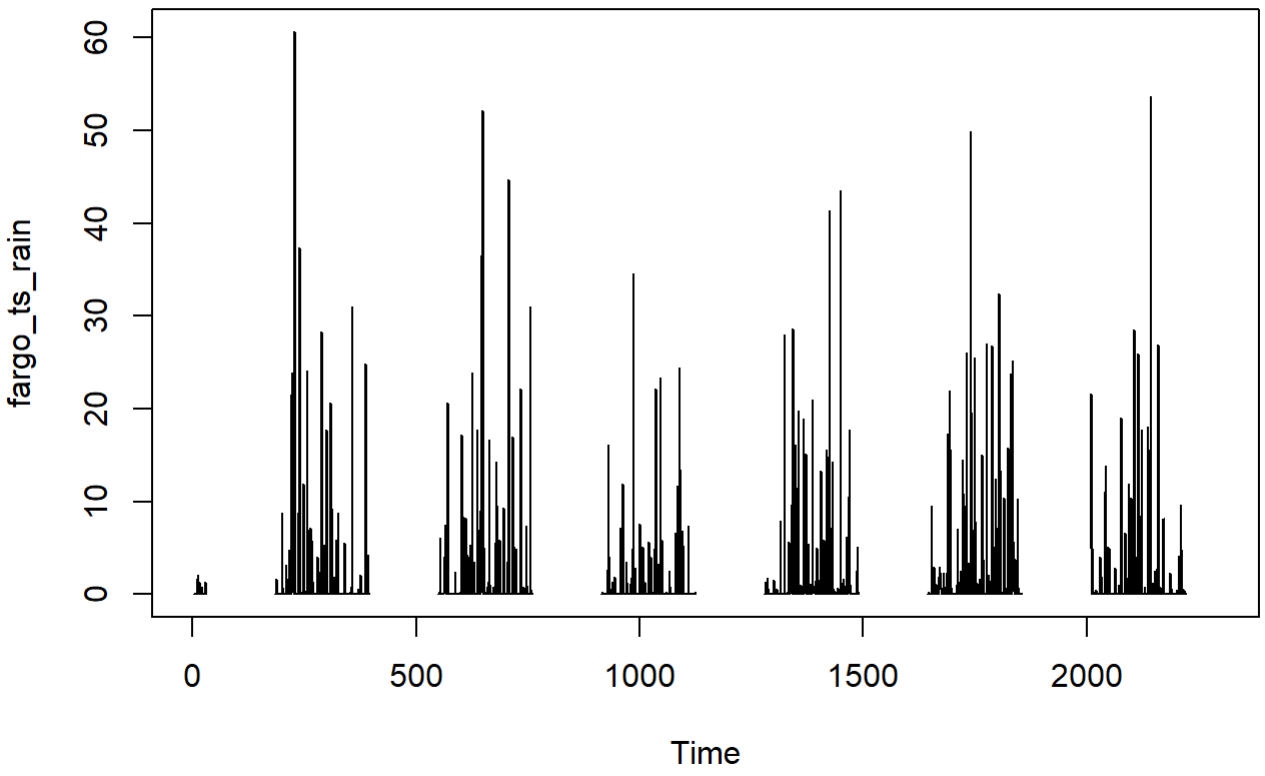

```
Acf(grand_ts_rain)
```



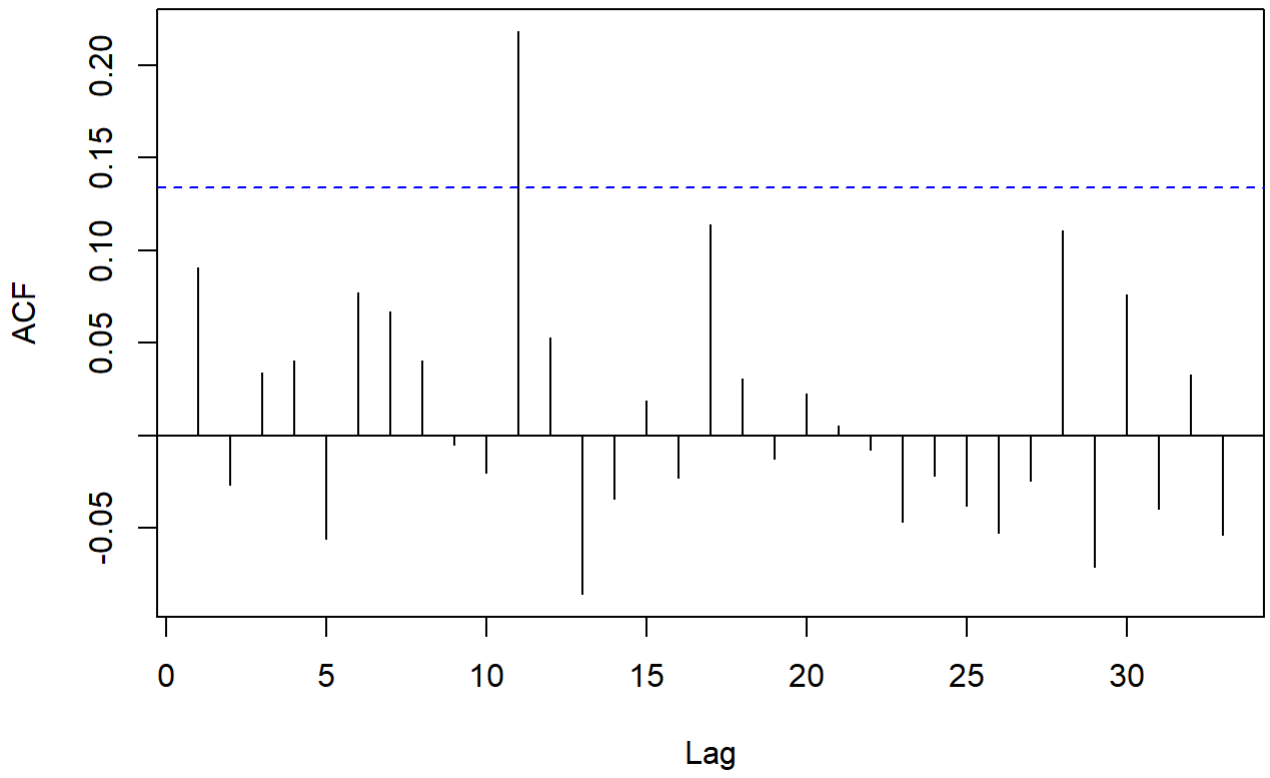

```
class(fargo_ts_rain)
```

```
## [1] "ts"
```

```
plot(fargo_ts_rain)
```




```
Acf(fargo_ts_rain)
```




```
g_rain <- Arima(grand_ts_rain, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), includ
e.constant = TRUE)
g_rain
```

```
## Series: grand_ts_rain
## ARIMA(0,0,1) with drift
##
## Coefficients:
##           ma1    sar1  drift
##       -1.0000  0.0595  2e-04
## s.e.   0.0025  0.0275  3e-04
##
## sigma^2 estimated as 55.24:  log likelihood=-4498.75
## AIC=9005.51   AICc=9005.52   BIC=9028.33
```

```
coeftest(g_rain)
```

```
##
## z test of coefficients:
##
##          Estimate  Std. Error   z value Pr(>|z|)
## ma1   -0.99999990  0.00254141 -393.4828  < 2e-16 ***
## sar1   0.05947071  0.02749611    2.1629  0.03055 *
## drift  0.00022453  0.00033881    0.6627  0.50752
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
f_rain <- Arima(fargo_ts_rain, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), includ
e.constant = TRUE)
f_rain
```

```
## Series: fargo_ts_rain
## ARIMA(0,0,1) with drift
##
## Coefficients:
##           ma1    sar1  drift
##       -1.0000  0.0465  2e-04
## s.e.   0.0023  0.0275  3e-04
##
## sigma^2 estimated as 37.91:  log likelihood=-4251.55
## AIC=8511.1   AICc=8511.12   BIC=8533.92
```

```
coeftest(f_rain)
```

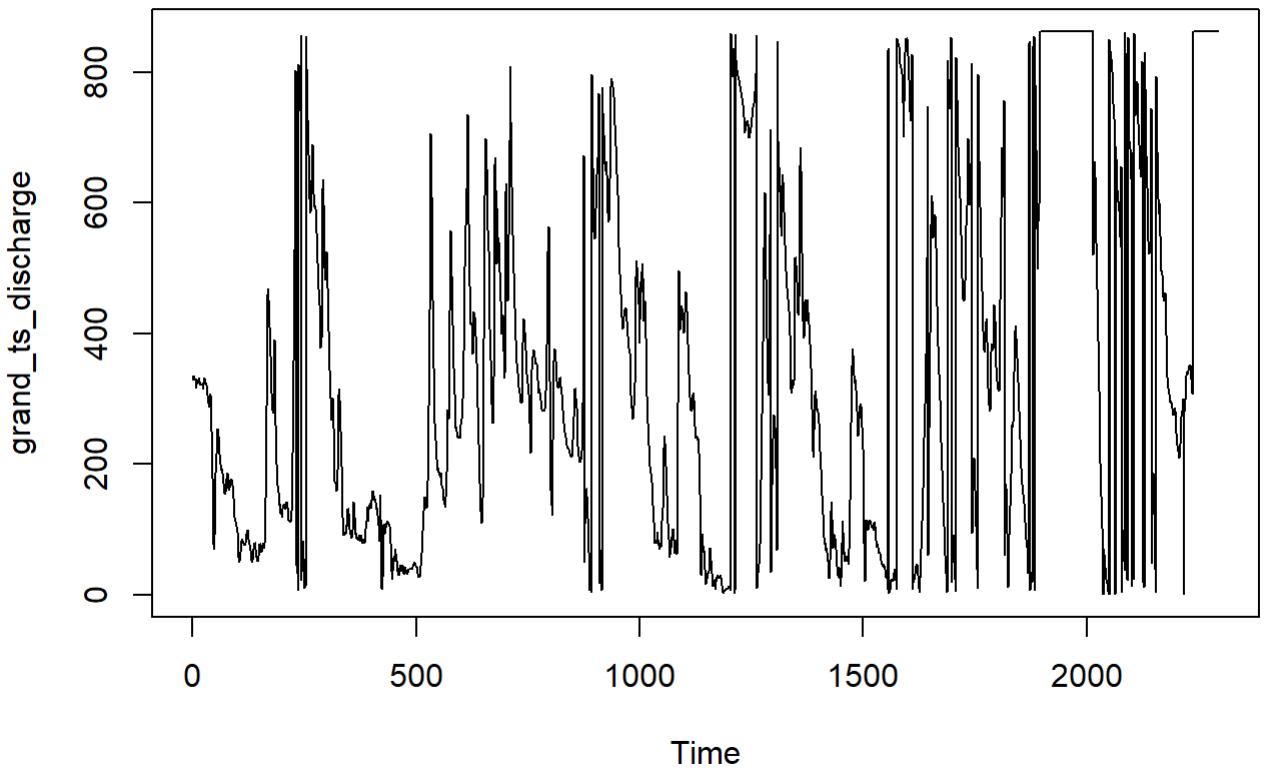
```
##
## z test of coefficients:
##
##          Estimate  Std. Error   z value Pr(>|z|)
## ma1   -0.99999958  0.00231177 -432.5681  < 2e-16 ***
## sar1   0.04649987  0.02752201    1.6896  0.09111 .
## drift  0.00023758  0.00027733    0.8567  0.39163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Discharge

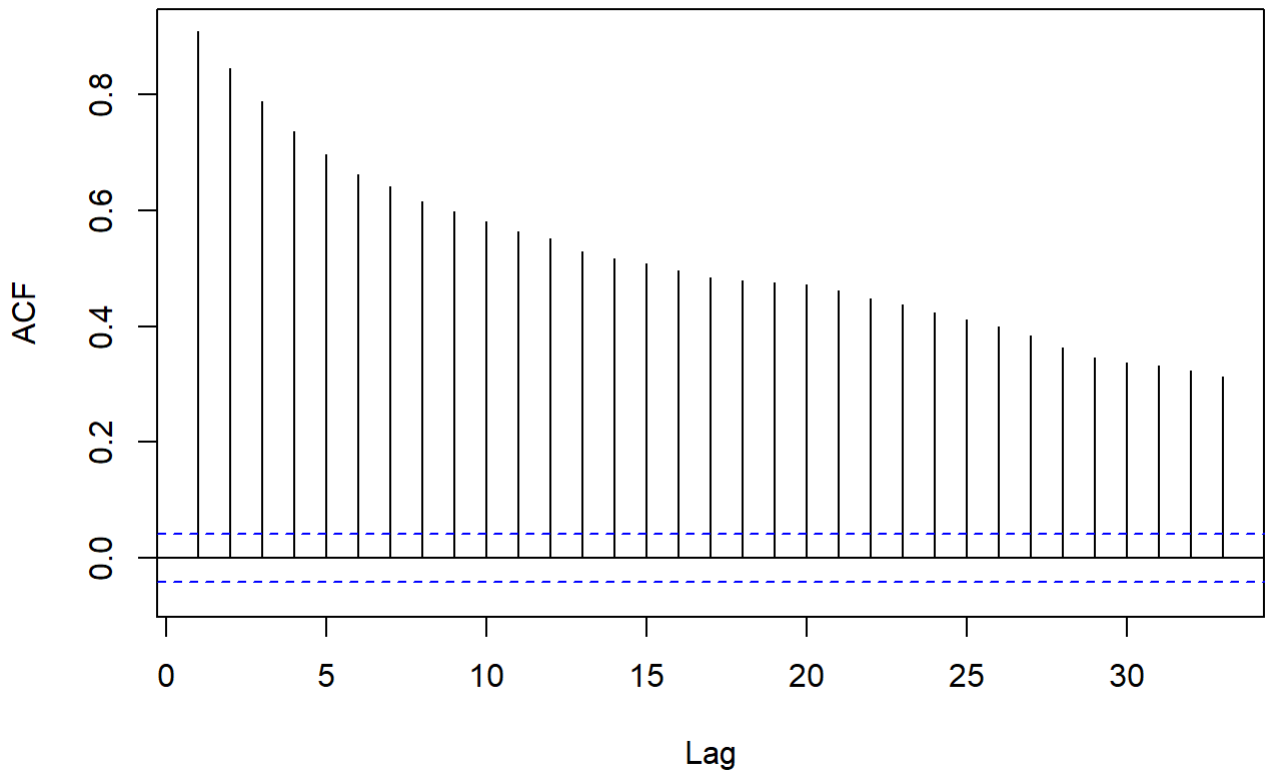
```
grand_ts_discharge <- ts(grand_from_2014$discharge, frequency = 1)
fargo_ts_discharge <- ts(fargo_from_2014$discharge, frequency = 1)

class(grand_ts_discharge)
```

```
## [1] "ts"
```

```
plot(grand_ts_discharge)
```



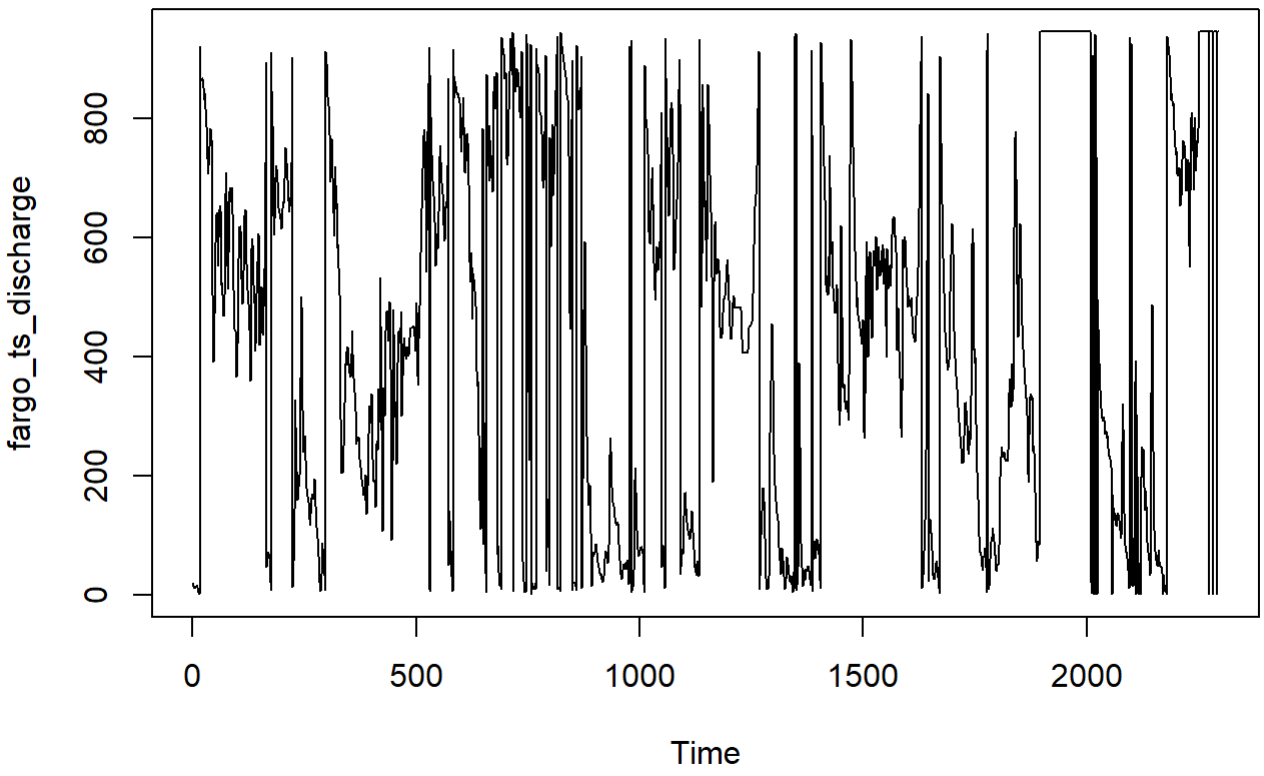

```
Acf(grand_ts_discharge)
```



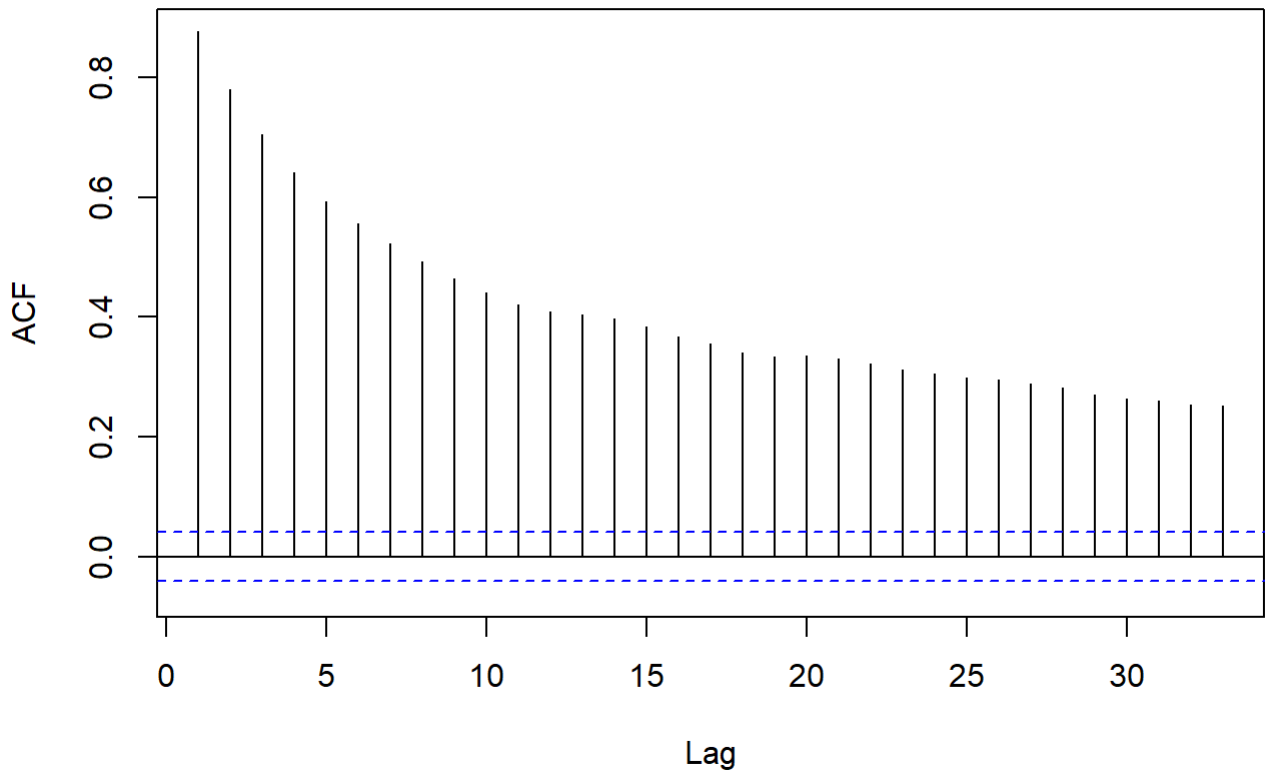

```
class(fargo_ts_discharge)
```

```
## [1] "ts"
```

```
plot(fargo_ts_discharge)
```




```
Acf(fargo_ts_discharge)
```




```
g_dis <- Arima(grand_ts_discharge, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), inc
lude.constant = TRUE)
g_dis
```

```
## Series: grand_ts_discharge
## ARIMA(0,0,1) with drift
##
## Coefficients:
##           ma1    sar1   drift
##       -0.9342  0.7773  0.2701
## s.e.   0.0178  0.0289  0.6952
##
## sigma^2 estimated as 12599:  log likelihood=-14076.89
## AIC=28161.79   AICc=28161.8   BIC=28184.74
```

```
coeftest(g_dis)
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error  z value Pr(>|z|)
## ma1   -0.934248   0.017835 -52.3827   <2e-16 ***
## sar1   0.777350   0.028913  26.8859   <2e-16 ***
## drift  0.270110   0.695169   0.3886   0.6976
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
f_dis <- Arima(fargo_ts_discharge, order = c(0, 0, 1),  seasonal = list(order = c(1, 1, 0)), inc
lude.constant = TRUE)
f_dis
```

```
## Series: fargo_ts_discharge
## ARIMA(0,0,1) with drift
##
## Coefficients:
##           ma1    sar1   drift
##       -0.9555  0.7891  0.2927
## s.e.   0.0107  0.0207  0.6510
##
## sigma^2 estimated as 21371:  log likelihood=-14682.8
## AIC=29373.6   AICc=29373.62   BIC=29396.56
```

```
coeftest(f_dis)
```

```
##
## z test of coefficients:
##
##        Estimate Std. Error  z value Pr(>|z|)
## ma1   -0.955456   0.010682 -89.4460   <2e-16 ***
## sar1   0.789144   0.020675  38.1698   <2e-16 ***
## drift  0.292743   0.651045   0.4497    0.653
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Vector Autoregression Analysis

```
library(vars)
library(mFilter)
library(tseries)
library(TSstudio)
library(forecast)
library(tidyverse)
```

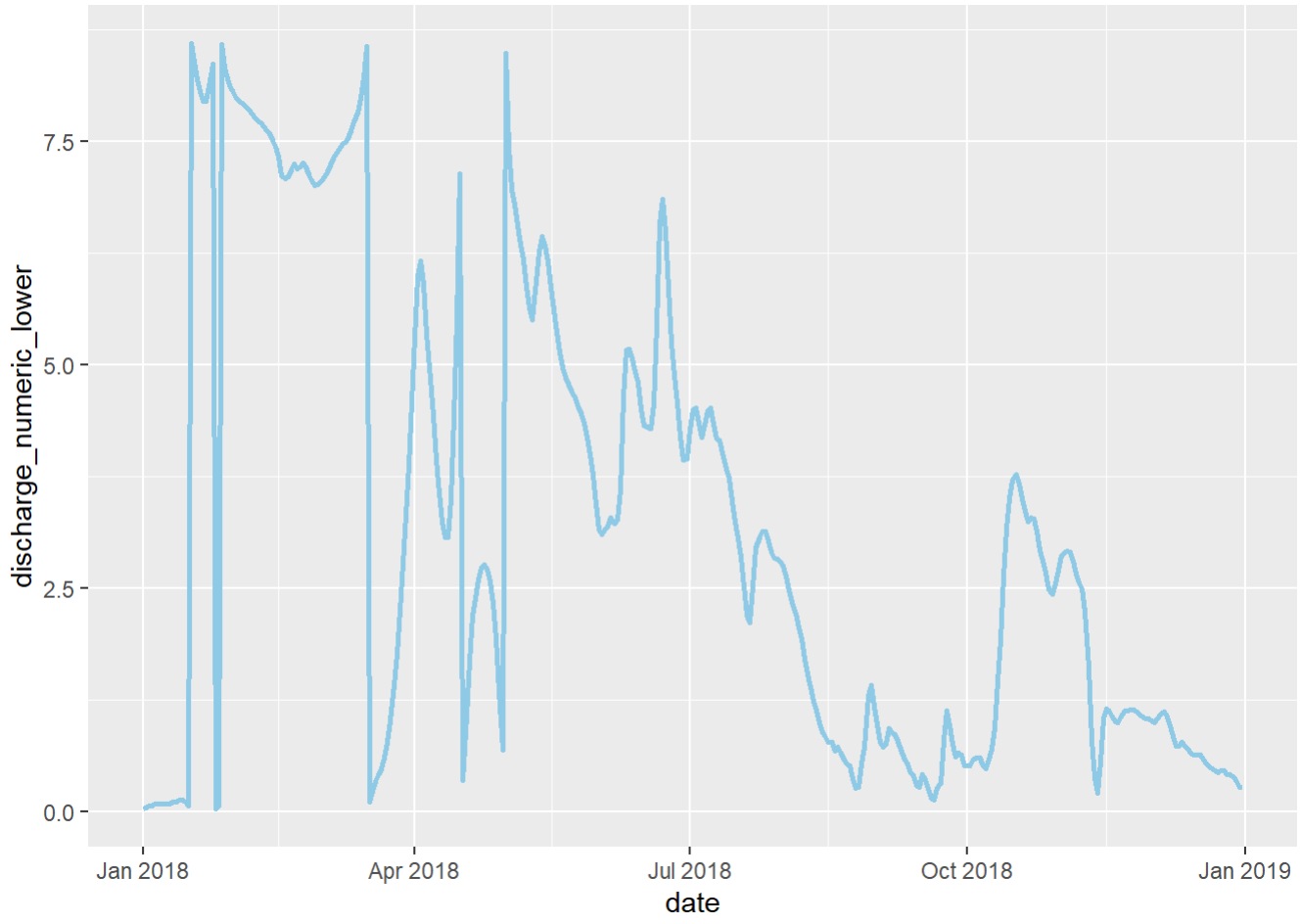
```
g1 <- cbind(g_dis, g_rain, g_snow, g_soil, g_temp)
colnames(g1) <- cbind("discharge", "rainfall", "snowdepth", "soil", "temp")
grand_lagselect <- VARselect(g1, lag.max = 15, type = "const")
grand_lagselect$selection

# Building model 1
m1 <- VAR(g1, p = 2, type = "const", season = NULL, exog = NULL)
summary(m1)
```

Better Visualization

For computation matter let's just take the year of 2018

```
grand_2018 <- grand_from_2014[grand_from_2014$year==2018,]
rownames(grand_2018) <- 1:nrow(grand_2018)

# make discharge continuous
grand_2018$discharge_numeric <- as.numeric(grand_2018$discharge)
grand_2018$discharge_numeric_lower <- grand_2018$discharge_numeric / 100
```

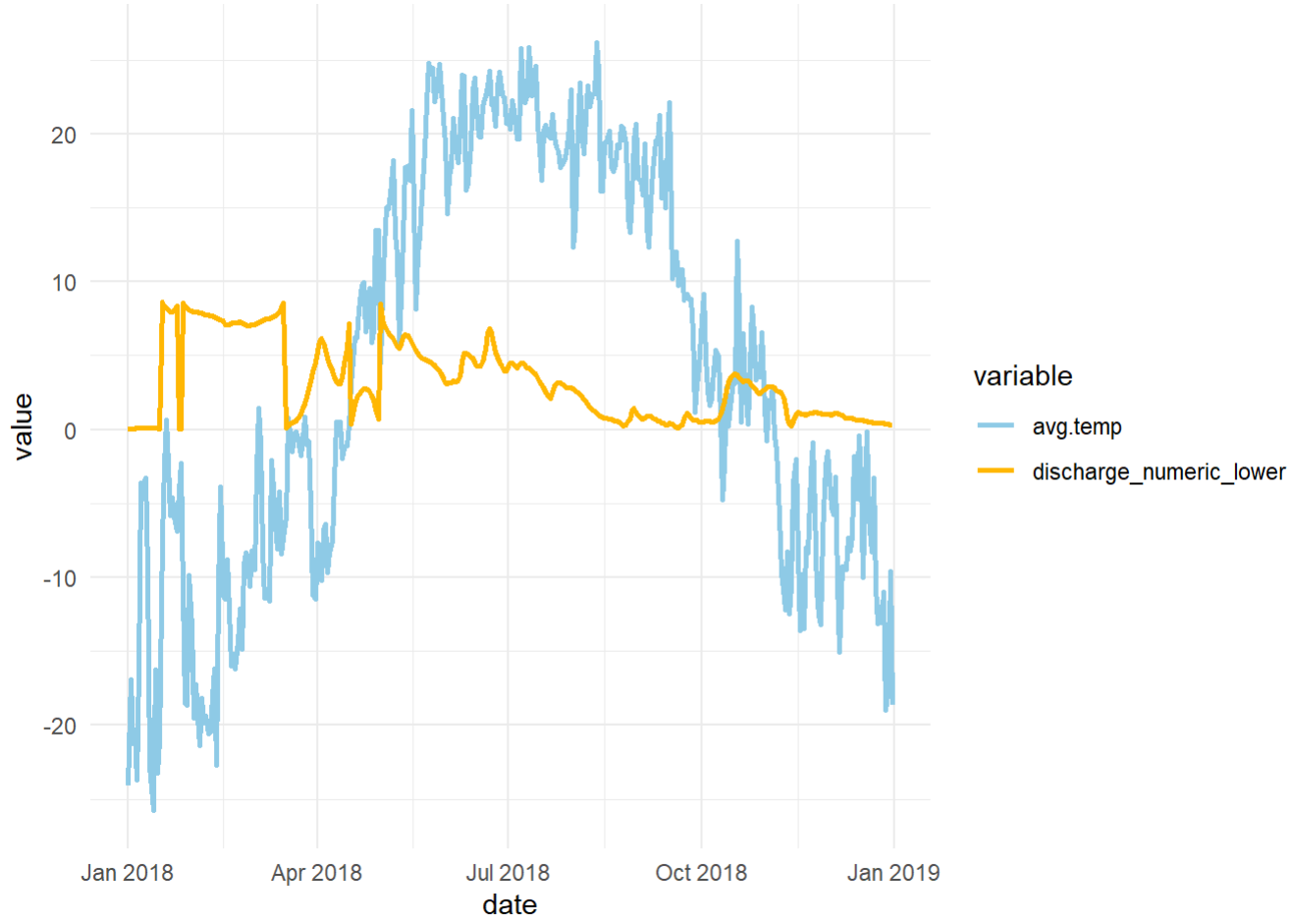
```
ggplot(data = grand_2018, aes(x = date, y = discharge_numeric_lower)) +
  geom_line(color = "#8ecae6", size = 1)
```



```
df <- grand_2018 %>%
  select(date, discharge_numeric_lower, avg.temp) %>%
  gather(key = "variable", value = "value", -date)
```

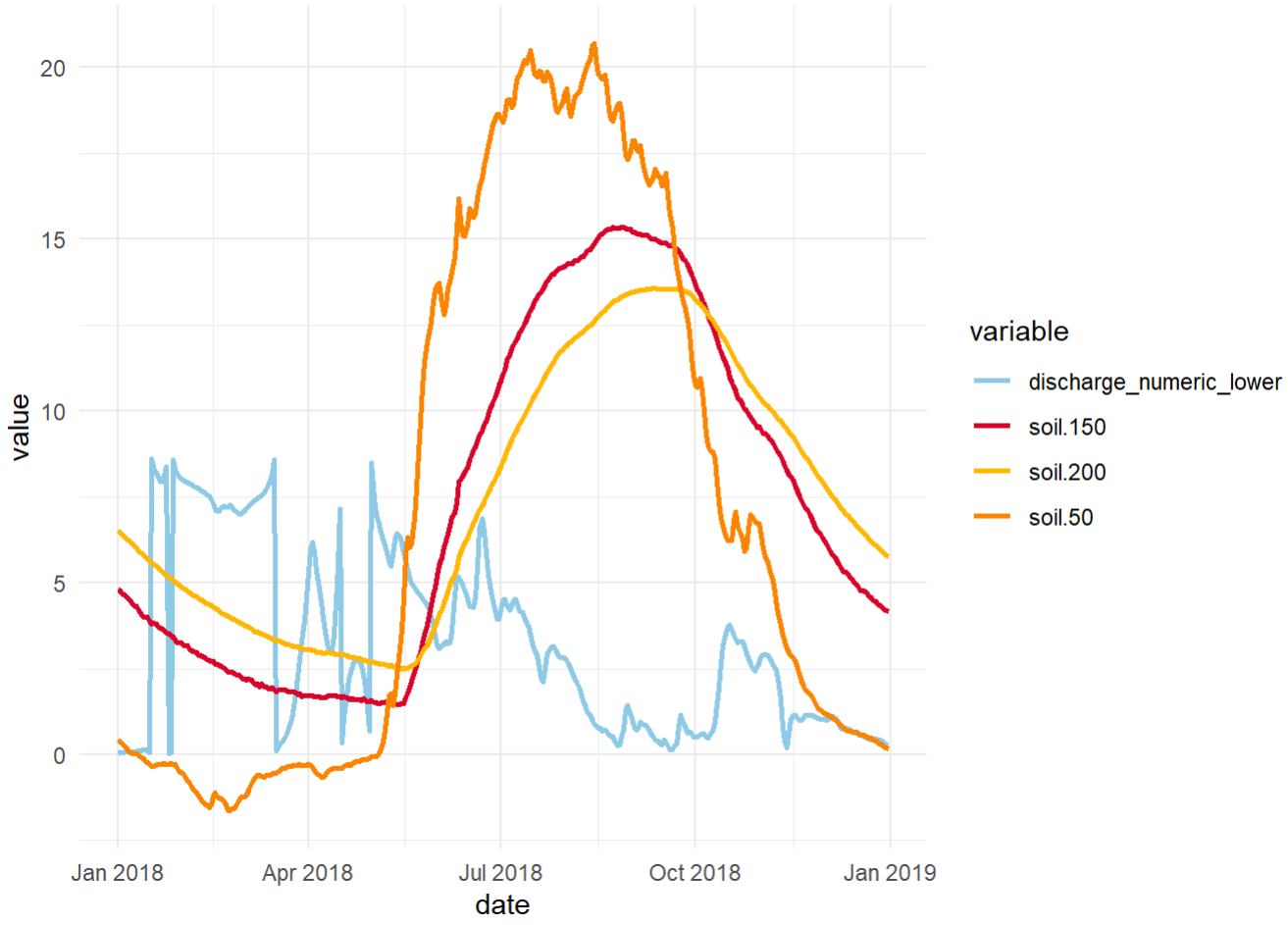
Discharge in relation with the average temperature

```
ggplot(df, aes(x = date, y = value)) +
  geom_line(aes(color = variable), size = 1) +
  scale_color_manual(values = c("#8ecae6", "#ffb703")) +
  theme_minimal()
```



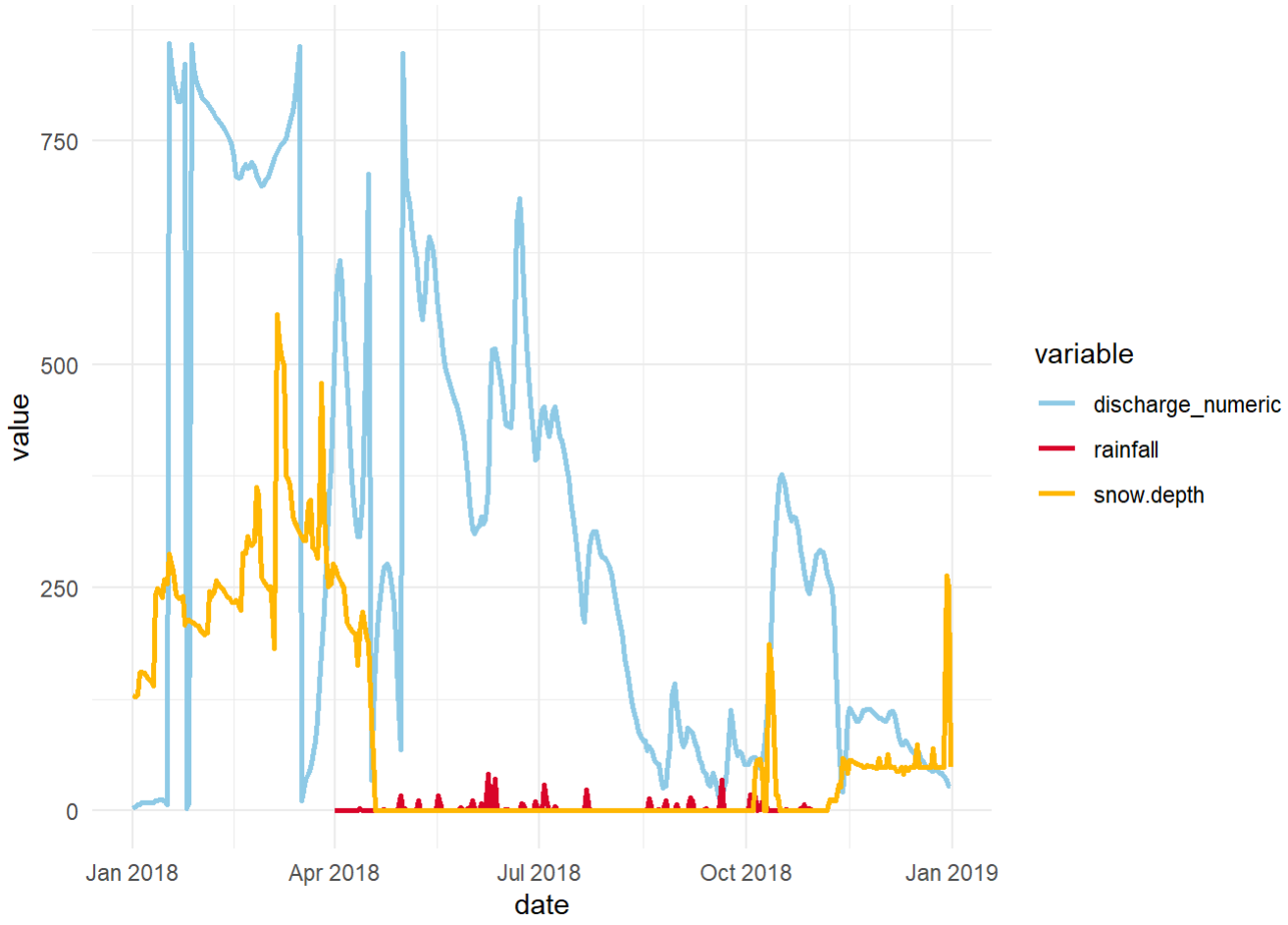
This is with soil temperatures

```
df2 <- grand_2018 %>%
  select(date, discharge_numeric_lower,soil.50,soil.150,soil.200) %>%
  gather(key = "variable", value = "value", -date)


ggplot(df2, aes(x = date, y = value)) +
  geom_line(aes(color = variable), size = 1) +
  scale_color_manual(values = c("#8ecae6", "#d90429", "#ffb703","#fb8500")) +
  theme_minimal()
```




```
df3 <- grand_2018 %>%
  select(date, discharge_numeric, rainfall, snow.depth) %>%
  gather(key = "variable", value = "value", -date)


ggplot(df3, aes(x = date, y = value)) +
  geom_line(aes(color = variable), size = 1) +
  scale_color_manual(values = c("#8ecae6", "#d90429", "#ffb703")) +
  theme_minimal()
```

```
## Warning: Removed 151 row(s) containing missing values (geom_path).
```



Remembering outliers