# ADS506 - Team Project - EDA

Stephen Kuc, Lane Whitmore, Luis Perez

2022-11-07

```
library(fpp2)
library(tidyverse)
library(ggplot2)
library(dplyr)
library(lubridate)
library(zoo)

set.seed(25)
```

```
land.temps <- read.csv("~/USD/ADS506/GlobalLandTemperaturesByCountry.csv")
head(land.temps)
```

```
##           dt AverageTemperature AverageTemperatureUncertainty Country
## 1 1743-11-01              4.384                          2.294  Ã...land
## 2 1743-12-01                 NA                             NA  Ã...land
## 3 1744-01-01                 NA                             NA  Ã...land
## 4 1744-02-01                 NA                             NA  Ã...land
## 5 1744-03-01                 NA                             NA  Ã...land
## 6 1744-04-01              1.530                          4.680  Ã...land
```

```
# let's see what countries we have?
# not going to print all out

head(unique(land.temps$Country), 10)
```

```
##  [1] "Ã...land"         "Afghanistan"    "Africa"         "Albania"
##  [5] "Algeria"        "American Samoa" "Andorra"        "Angola"
##  [9] "Anguilla"       "Antarctica"
```

```
# check structure of dataframe

str(land.temps)
```

```
## 'data.frame':    577462 obs. of  4 variables:
##  $ dt                           : chr  "1743-11-01" "1743-12-01" "1744-01-01" "1744-02-01" ...
##  $ AverageTemperature           : num  4.38 NA NA NA NA ...
##  $ AverageTemperatureUncertainty: num  2.29 NA NA NA NA ...
##  $ Country                      : chr  "Ã...land" "Ã...land" "Ã...land" "Ã...land" ...
```

```r
# converting date into date format

land.temps$dt <- as.Date(land.temps$dt)
head(land.temps)
```

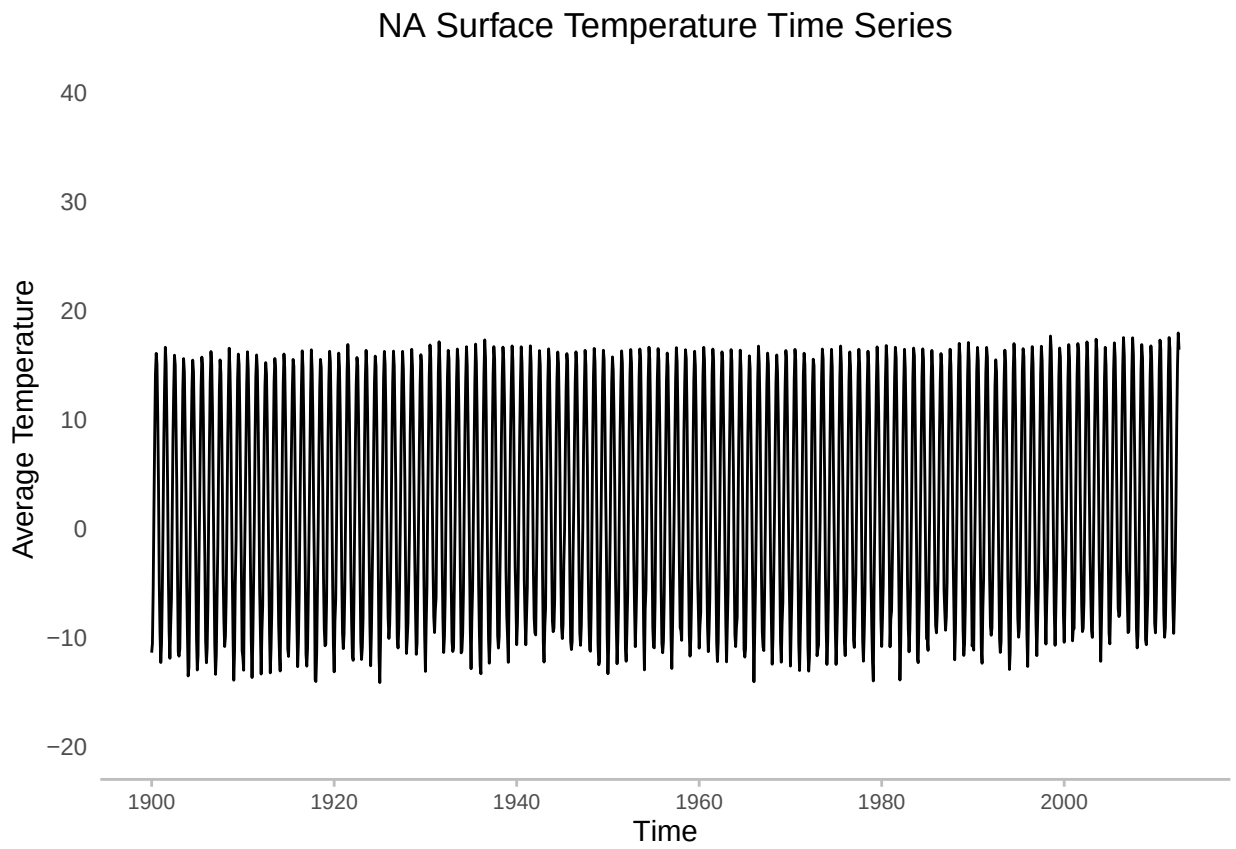
```
##           dt AverageTemperature AverageTemperatureUncertainty Country
## 1 1743-11-01              4.384                         2.294  Ã...land
## 2 1743-12-01                 NA                            NA  Ã...land
## 3 1744-01-01                 NA                            NA  Ã...land
## 4 1744-02-01                 NA                            NA  Ã...land
## 5 1744-03-01                 NA                            NA  Ã...land
## 6 1744-04-01              1.530                         4.680  Ã...land
```

```r
# North America Creation
NA.temps <- as.data.frame(land.temps[land.temps$Country == 'North America',])
rownames(NA.temps) <- 1:nrow(NA.temps)
na.temps <- NA.temps[1577:2940,]
rownames(na.temps) <- 1:nrow(na.temps)
train.na.temps <- na.temps[1:1352,]
val.na.temps <- na.temps[1353:1364,]

# South america
SA.temps <- land.temps[land.temps$Country == 'South America',]
rownames(SA.temps) <- 1:nrow(SA.temps)
sa.temps <- SA.temps[589:1952,]
rownames(sa.temps) <- 1:nrow(sa.temps)
train.sa.temps <- sa.temps[1:1352,]
val.sa.temps <- sa.temps[1353:1364,]

# Africa
AF.temps <- land.temps[land.temps$Country == 'Africa',]
rownames(AF.temps) <- 1:nrow(AF.temps)
af.temps <- AF.temps[601:1964,]
rownames(af.temps) <- 1:nrow(af.temps)
train.af.temps <- af.temps[1:1352,]
val.af.temps <- af.temps[1353:1364,]


# Europe
EU.temps <- land.temps[land.temps$Country == 'Europe',]
rownames(EU.temps) <- 1:nrow(EU.temps)
eu.temps <- EU.temps[1875:3238,]
rownames(eu.temps) <- 1:nrow(eu.temps)
train.eu.temps <- eu.temps[1:1352,]
val.eu.temps <- eu.temps[1353:1364,]

# Asia
AS.temps <- land.temps[land.temps$Country == 'Asia',]
rownames(AS.temps) <- 1:nrow(AS.temps)
as.temps <- AS.temps[1006:2369,]
rownames(as.temps) <- 1:nrow(as.temps)
train.as.temps <- as.temps[1:1352,]
val.as.temps <- as.temps[1353:1364,]
```

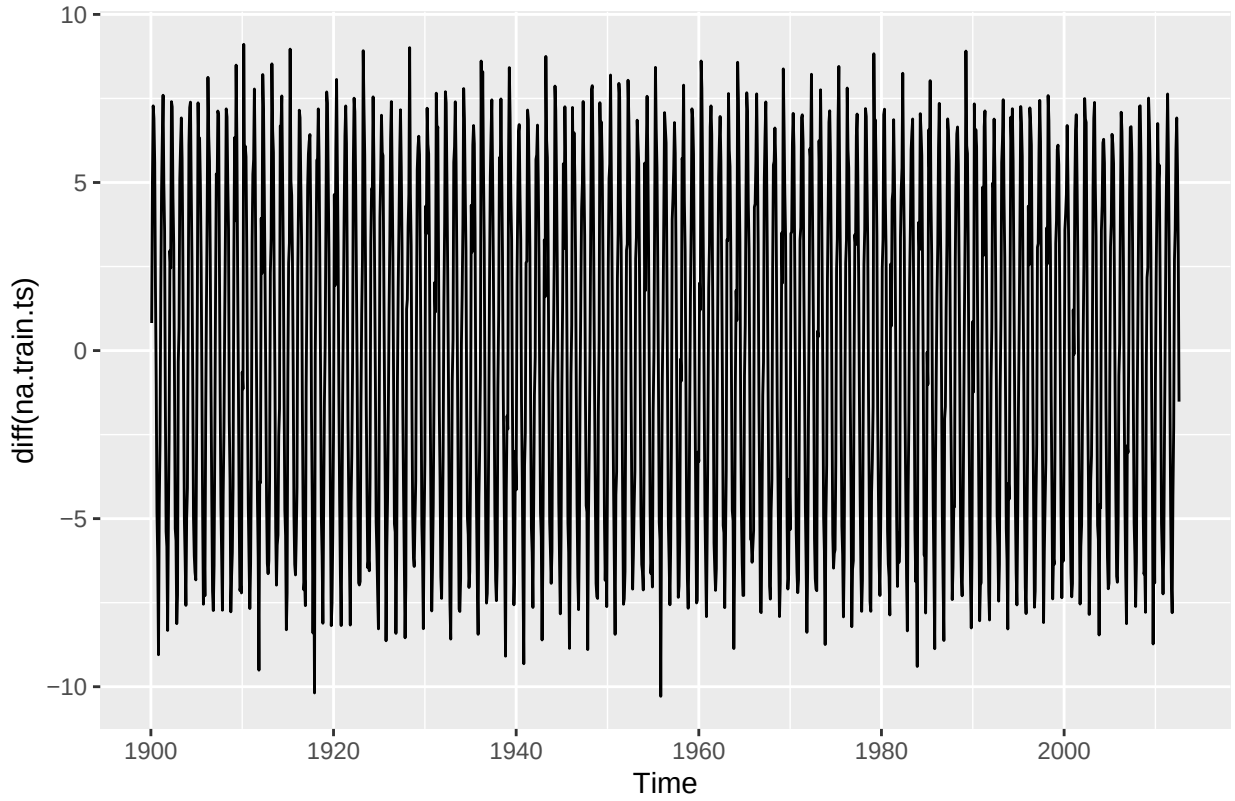
```r
# Australia
AU.temps <- land.temps[land.temps$Country == 'Australia',]
rownames(AU.temps) <- 1:nrow(AU.temps)
au.temps <- AU.temps[571:1934,]
rownames(au.temps) <- 1:nrow(au.temps)
train.au.temps <- au.temps[1:1352,]
val.au.temps <- au.temps[1353:1364,]
```

```r
# Splitting into train and test
# North America
na.train.ts <- ts(train.na.temps$AverageTemperature, start = c(1900,1), frequency = 12)
na.val.ts <- ts(val.na.temps$AverageTemperature, start = c(2012,9), frequency = 12)

# South America
sa.train.ts <- ts(train.sa.temps$AverageTemperature, start = c(1900,1), frequency = 12)
sa.val.ts <- ts(val.sa.temps$AverageTemperature, start = c(2012,9), frequency = 12)

# Africa
af.train.ts <- ts(train.af.temps$AverageTemperature, start = c(1900,1), frequency = 12)
af.val.ts <- ts(val.af.temps$AverageTemperature, start = c(2012,9), frequency = 12)

# Europe
eu.train.ts <- ts(train.eu.temps$AverageTemperature, start = c(1900,1), frequency = 12)
eu.val.ts <- ts(val.eu.temps$AverageTemperature, start = c(2012,9), frequency = 12)

# Asia
as.train.ts <- ts(train.as.temps$AverageTemperature, start = c(1900,1), frequency = 12)
as.val.ts <- ts(val.as.temps$AverageTemperature, start = c(2012,9), frequency = 12)

# Australia
au.train.ts <- ts(train.au.temps$AverageTemperature, start = c(1900,1), frequency = 12)
au.val.ts <- ts(val.au.temps$AverageTemperature, start = c(2012,9), frequency = 12)
```

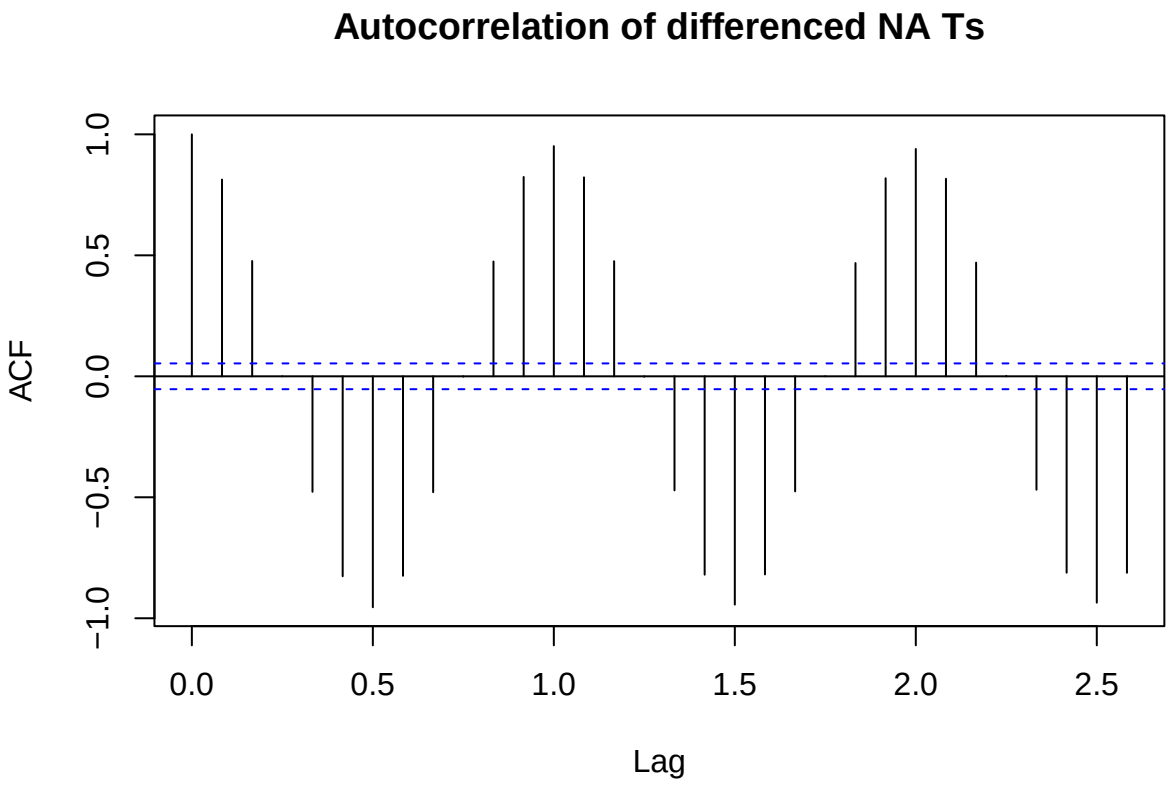
```r
autoplot(na.train.ts, xlab = 'Time', ylab = 'Average Temperature') +
  labs(title = "NA Surface Temperature Time Series") +
  scale_y_continuous(limits = c(-20,40), breaks = c(seq(-20,40,10))) +
  theme(axis.text.x = element_text(vjust = 0.7, size = 8),
        plot.title = element_text(hjust = 0.5),
        panel.background = element_blank(),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_line(colour = 'gray'),
        axis.line.x = element_line(colour = 'gray', size=0.5, linetype='solid'))
```



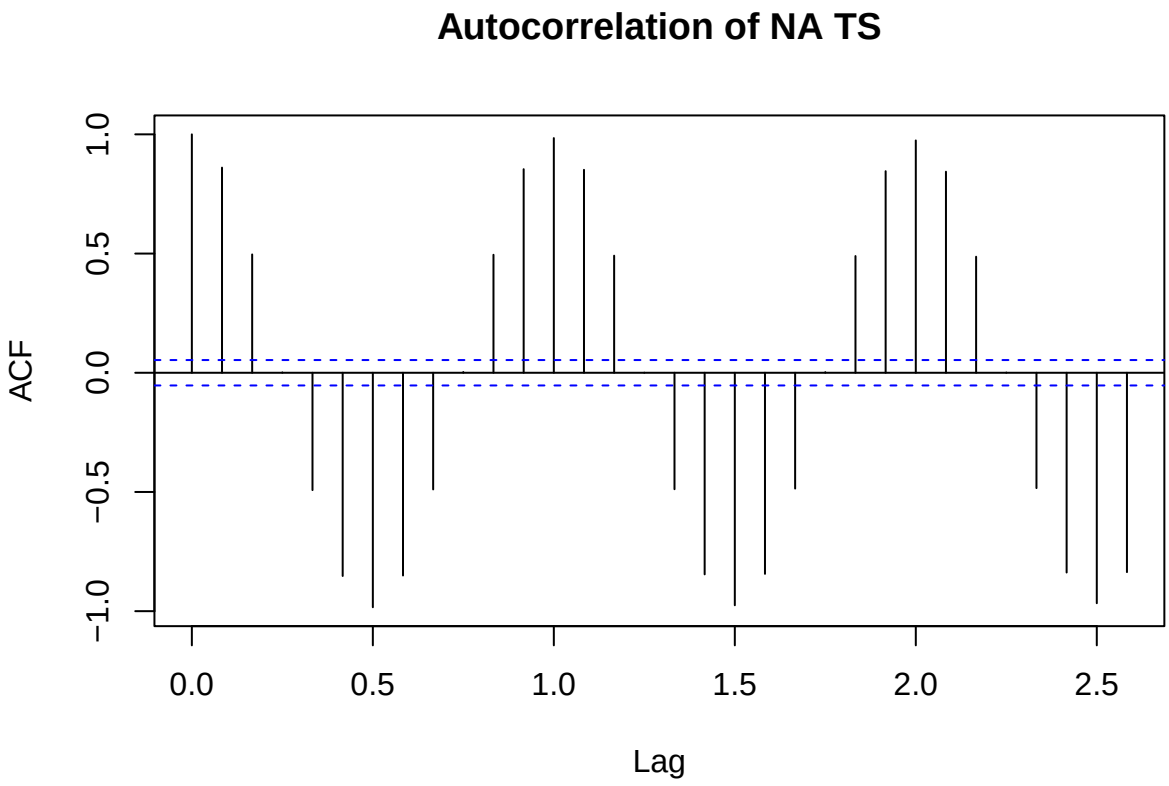

```
# Further EDA into North American Time Series

autoplot(diff(na.train.ts))
```



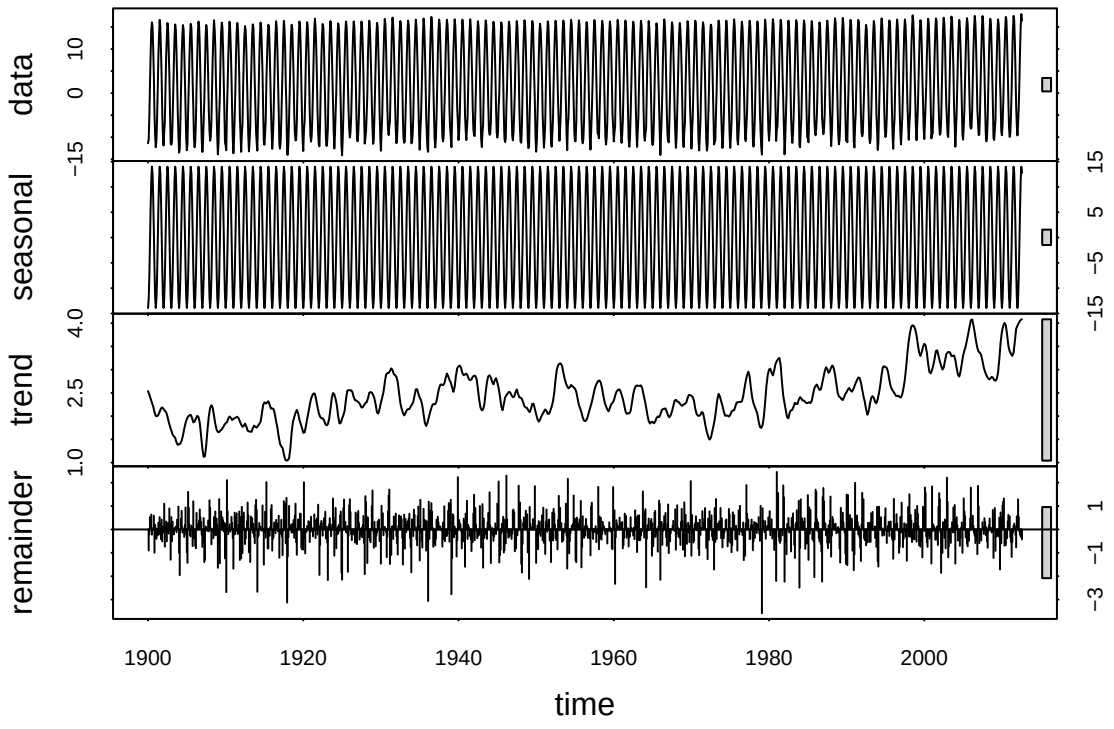

```
acf(diff(na.train.ts), main = 'Autocorrelation of differenced NA Ts')
```



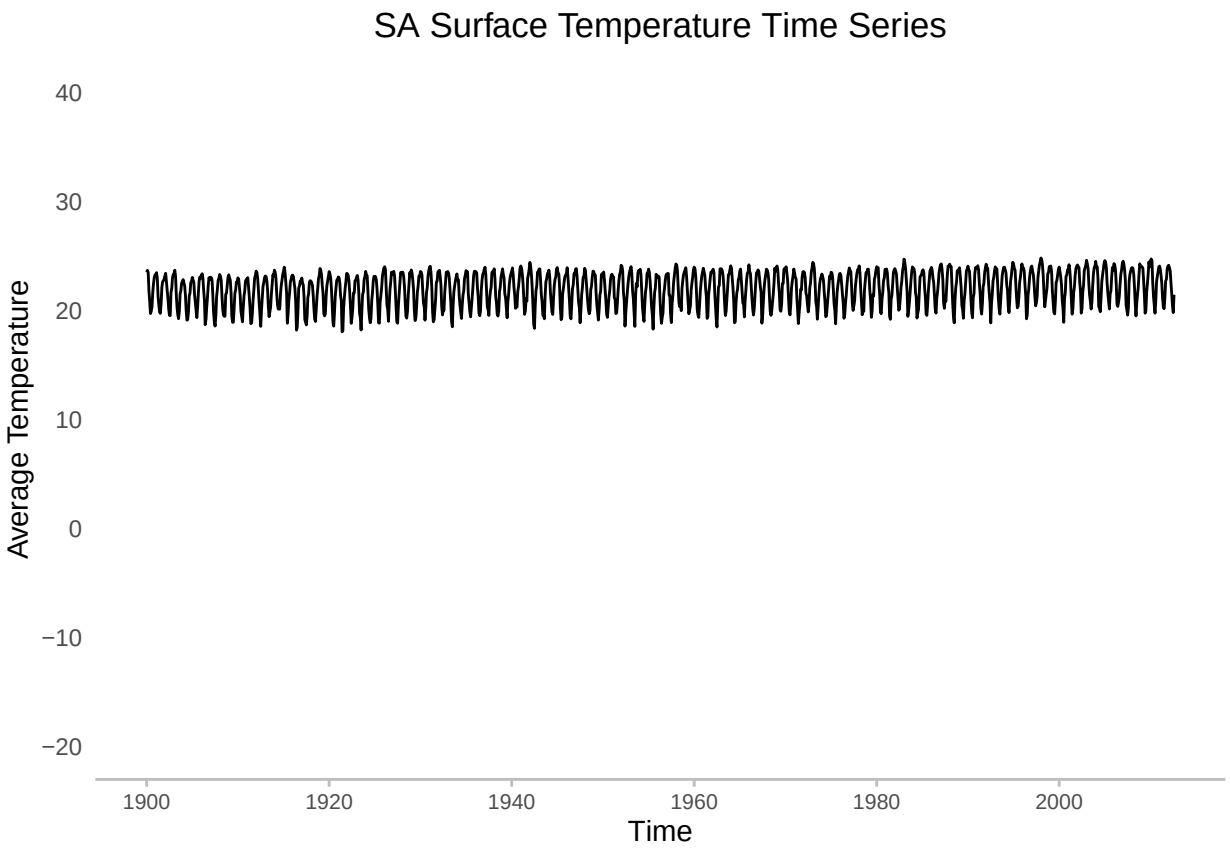

```
acf(na.train.ts, main = 'Autocorrelation of NA TS')
```



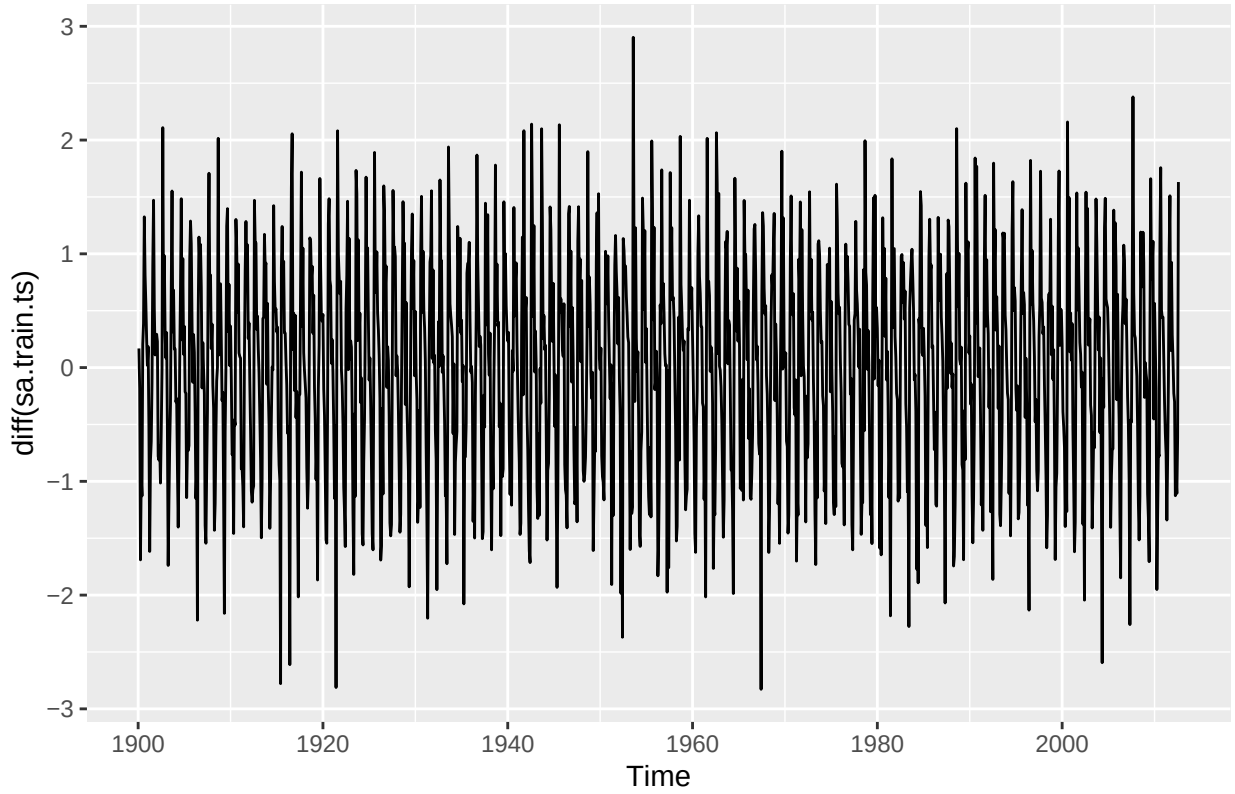

```
# stl() function to see seasonal, trend, and irregular components
na.stl <- stl(na.train.ts, s.window = 'periodic')
plot(na.stl)
```



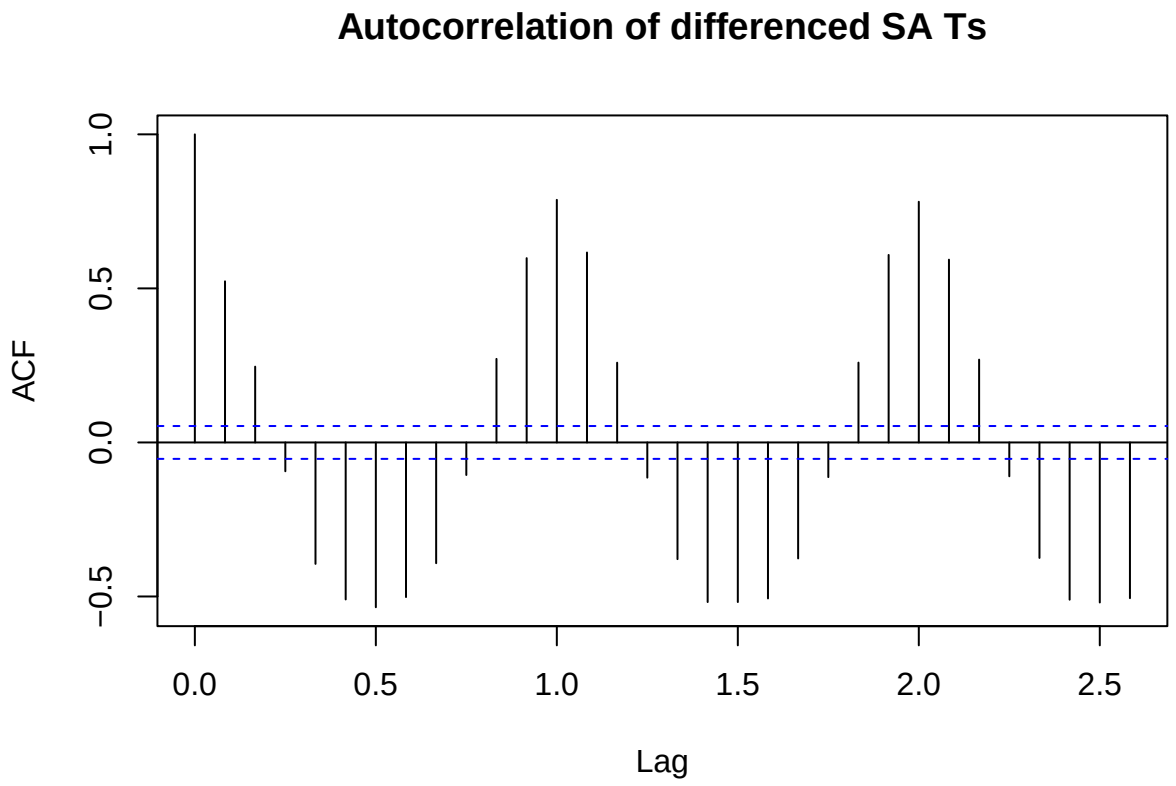

```
autoplot(sa.train.ts, xlab = 'Time', ylab = 'Average Temperature') +
  labs(title = "SA Surface Temperature Time Series") +
  scale_y_continuous(limits = c(-20,40), breaks = c(seq(-20,40,10))) +
  theme(axis.text.x = element_text(vjust = 0.7, size = 8),
        plot.title = element_text(hjust = 0.5),
        panel.background = element_blank(),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_line(colour = 'gray'),
        axis.line.x = element_line(colour = 'gray', size=0.5, linetype='solid'))
```



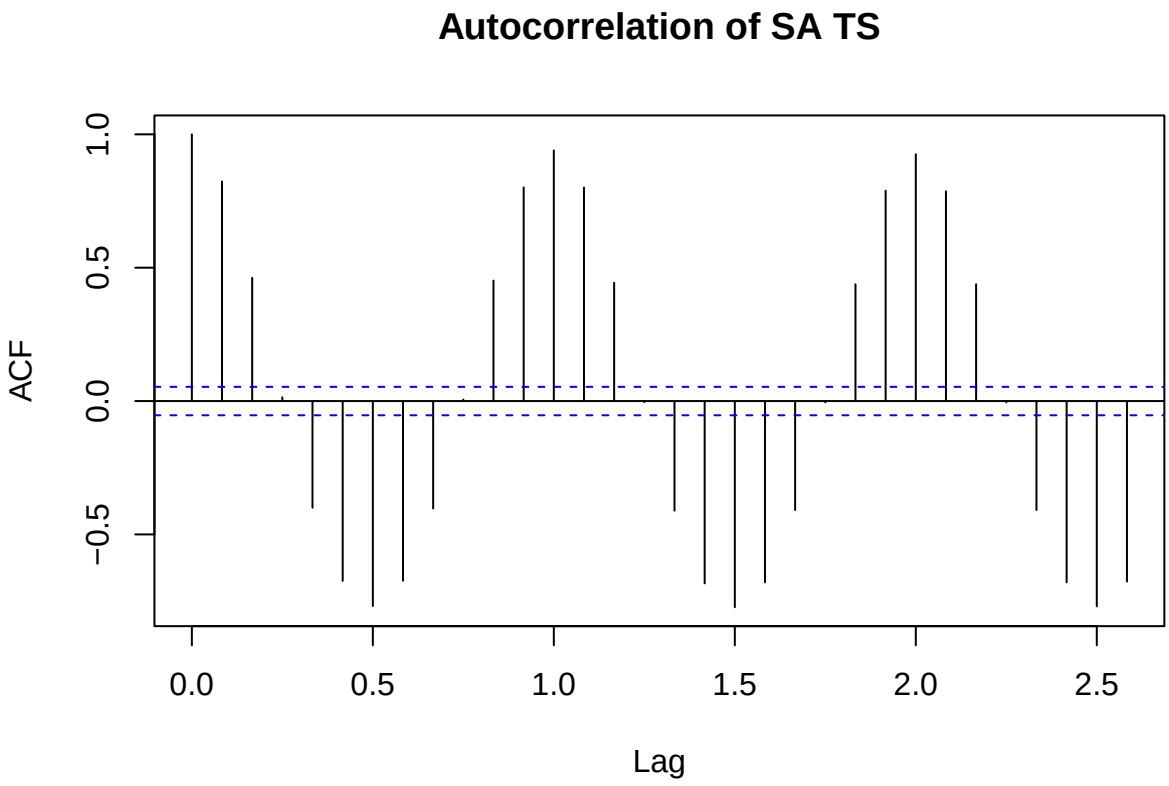

```
# Further EDA into South American Time Series

autoplot(diff(sa.train.ts))
```



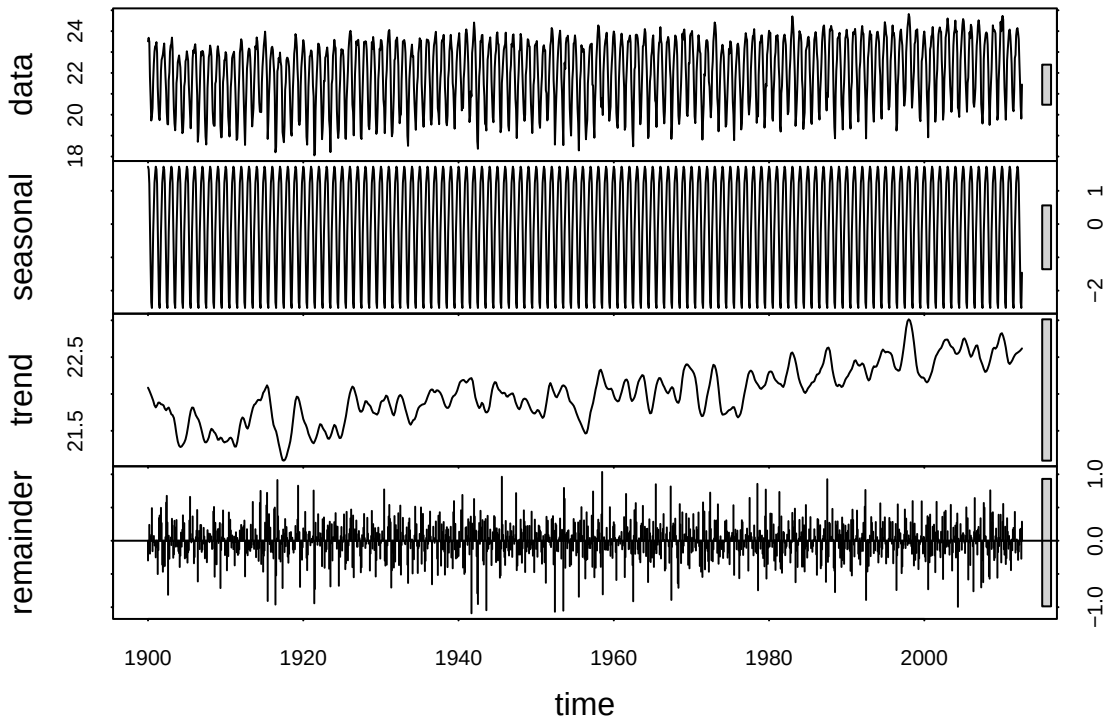

```
acf(diff(sa.train.ts), main = 'Autocorrelation of differenced SA Ts')
```



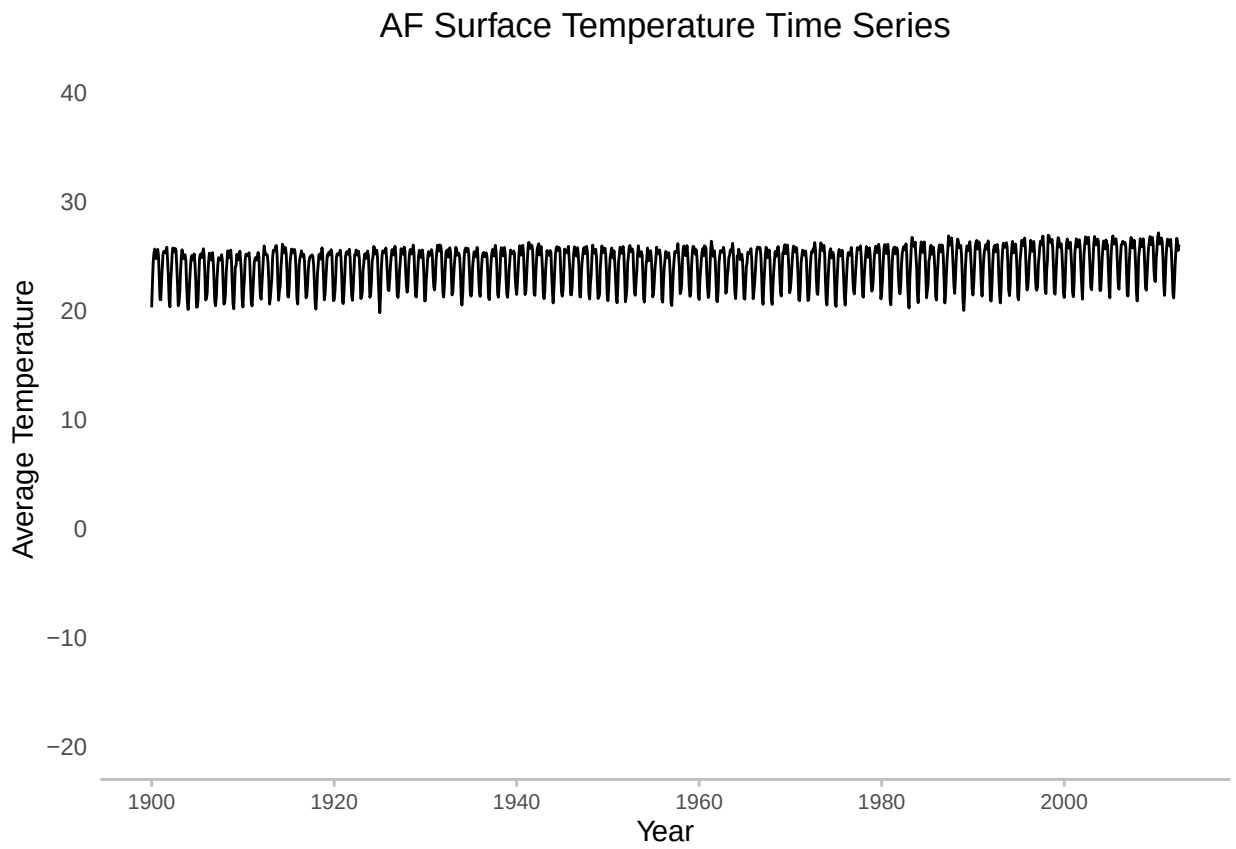

```
acf(sa.train.ts, main = 'Autocorrelation of SA TS')
```



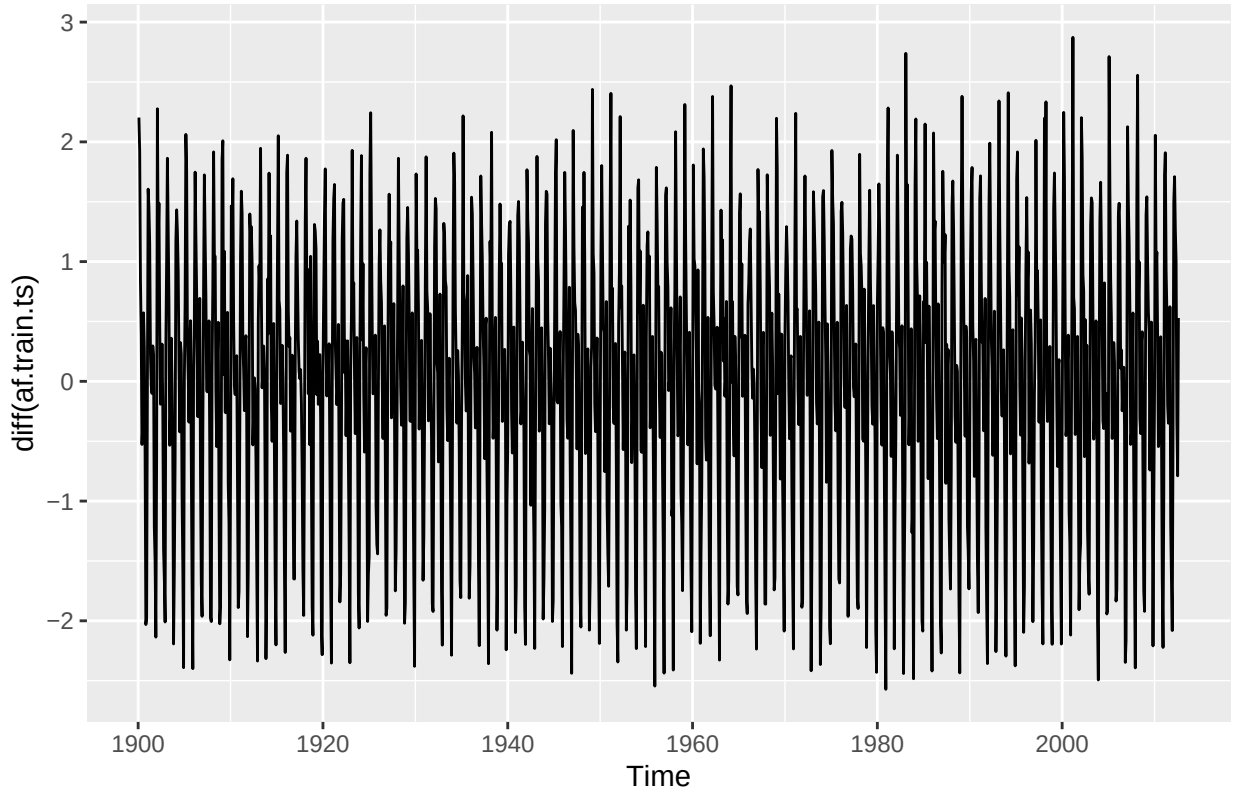

```
# stl() function to see seasonal, trend, and irregular components
sa.stl <- stl(sa.train.ts, s.window = 'periodic')
plot(sa.stl)
```



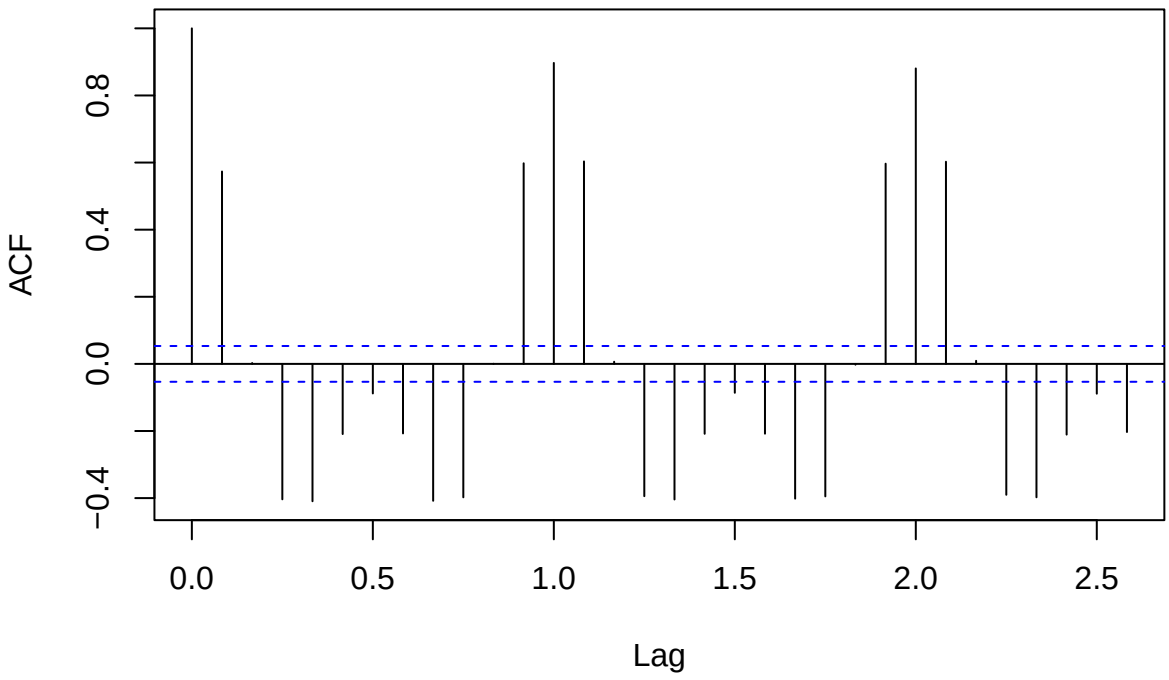

```
# Africa Data

autoplot(af.train.ts, xlab = 'Year', ylab = 'Average Temperature', bty = 'l', ylim = c(-20, 40))+
  labs(title = "AF Surface Temperature Time Series") +
  scale_y_continuous(limits = c(-20,40), breaks = c(seq(-20,40,10))) +
  theme(axis.text.x = element_text(vjust = 0.7, size = 8),
        plot.title = element_text(hjust = 0.5),
        panel.background = element_blank(),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_line(colour = 'gray'),
        axis.line.x = element_line(colour = 'gray', size=0.5, linetype='solid'))
```



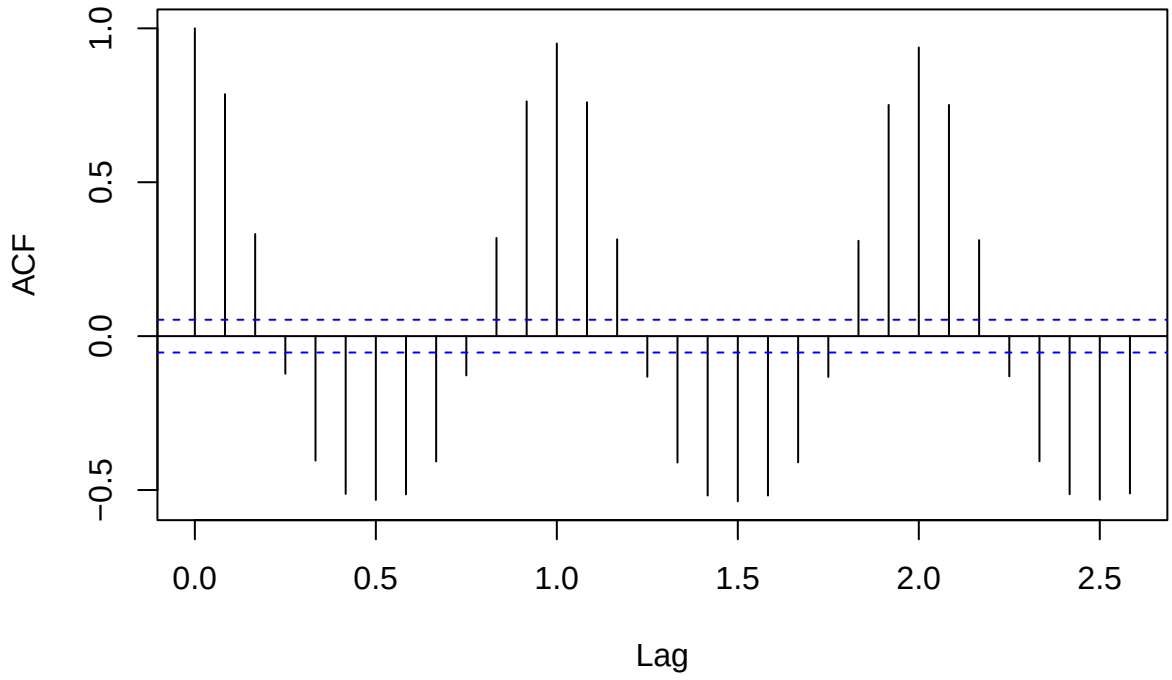

```
# Further EDA of Africa Data

autoplot(diff(af.train.ts))
```



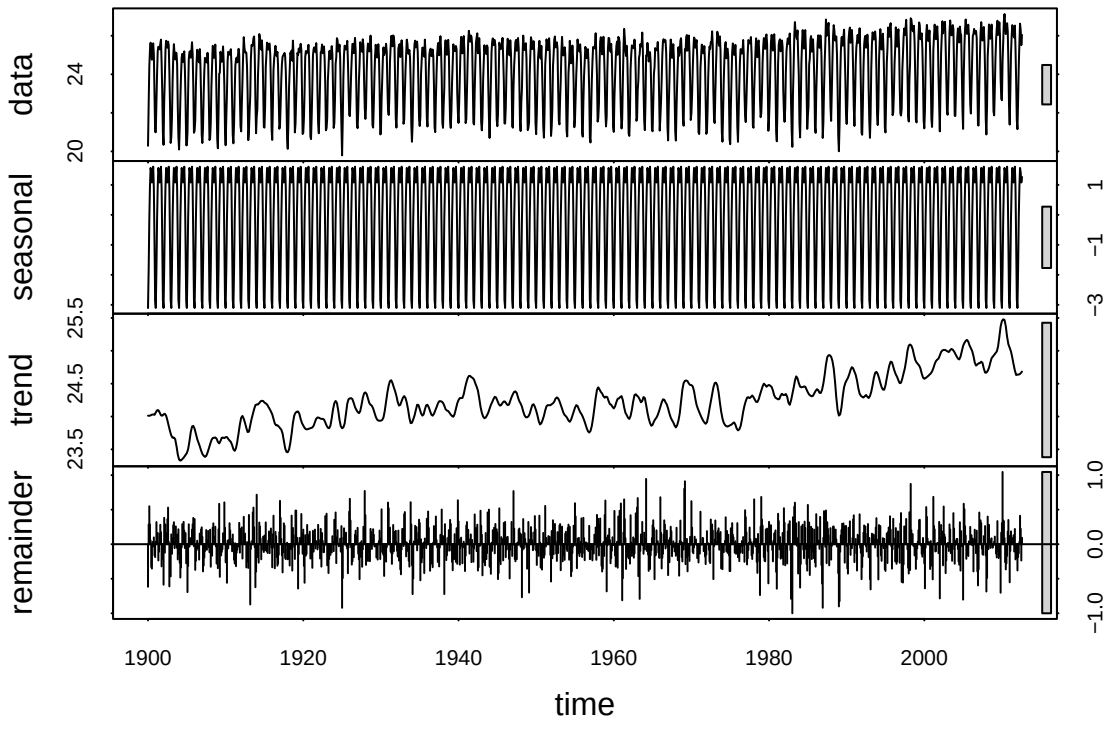

```
acf(diff(af.train.ts), main = '')
```



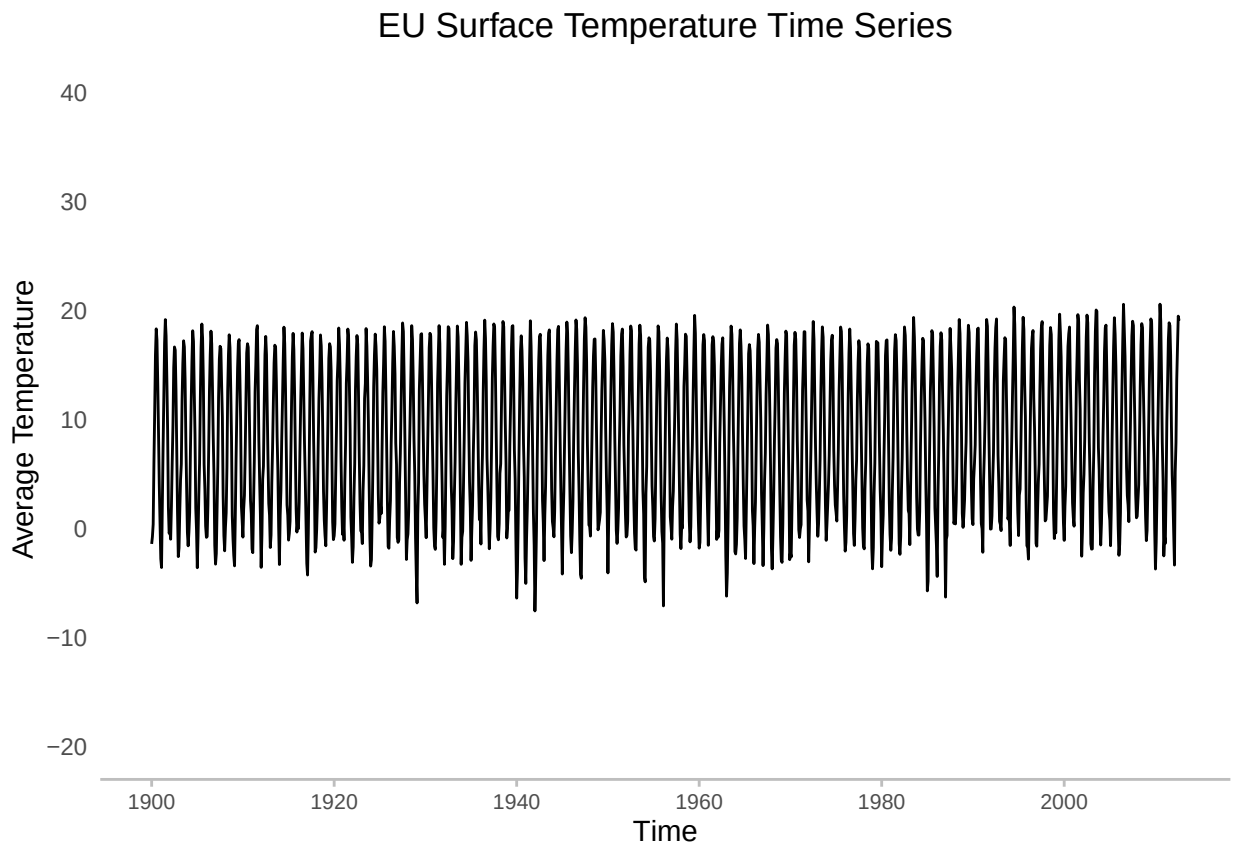

```
acf(af.train.ts, main = '')
```



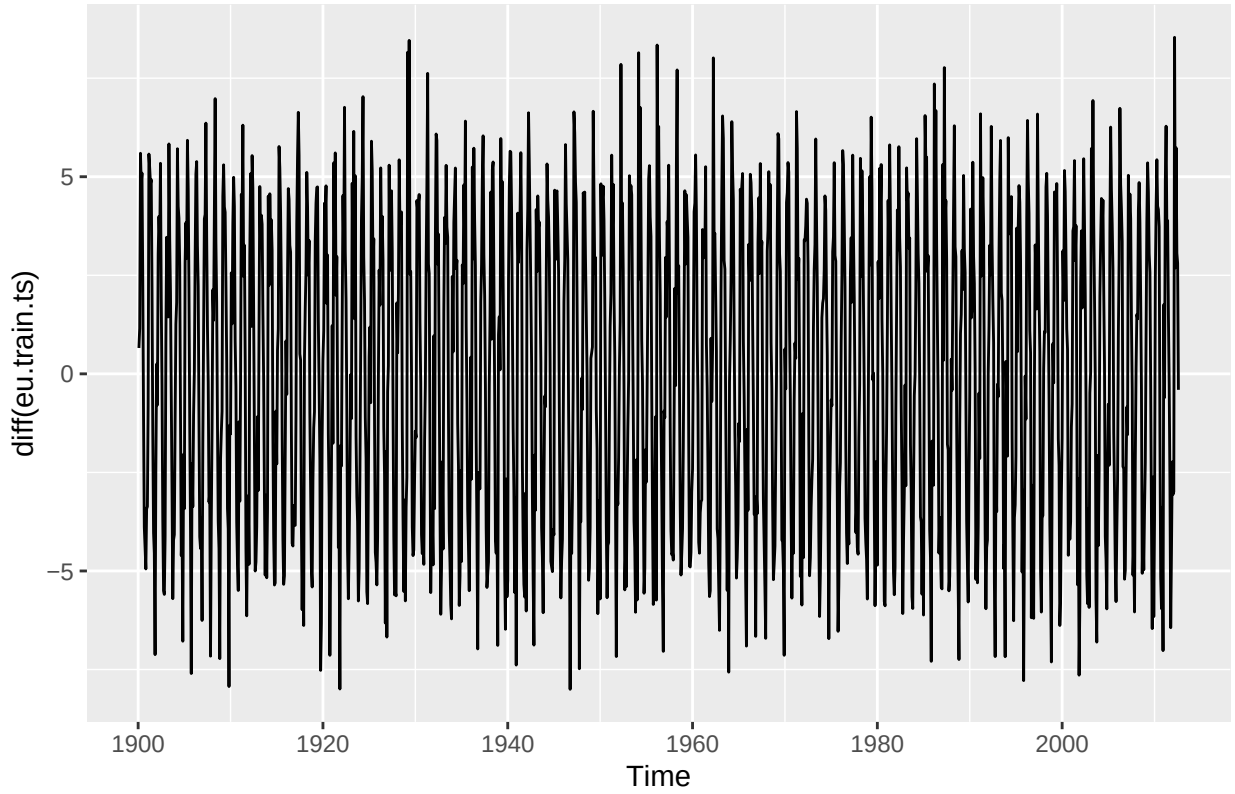

```
af.stl <- stl(af.train.ts, s.window = 'periodic')
plot(af.stl)
```



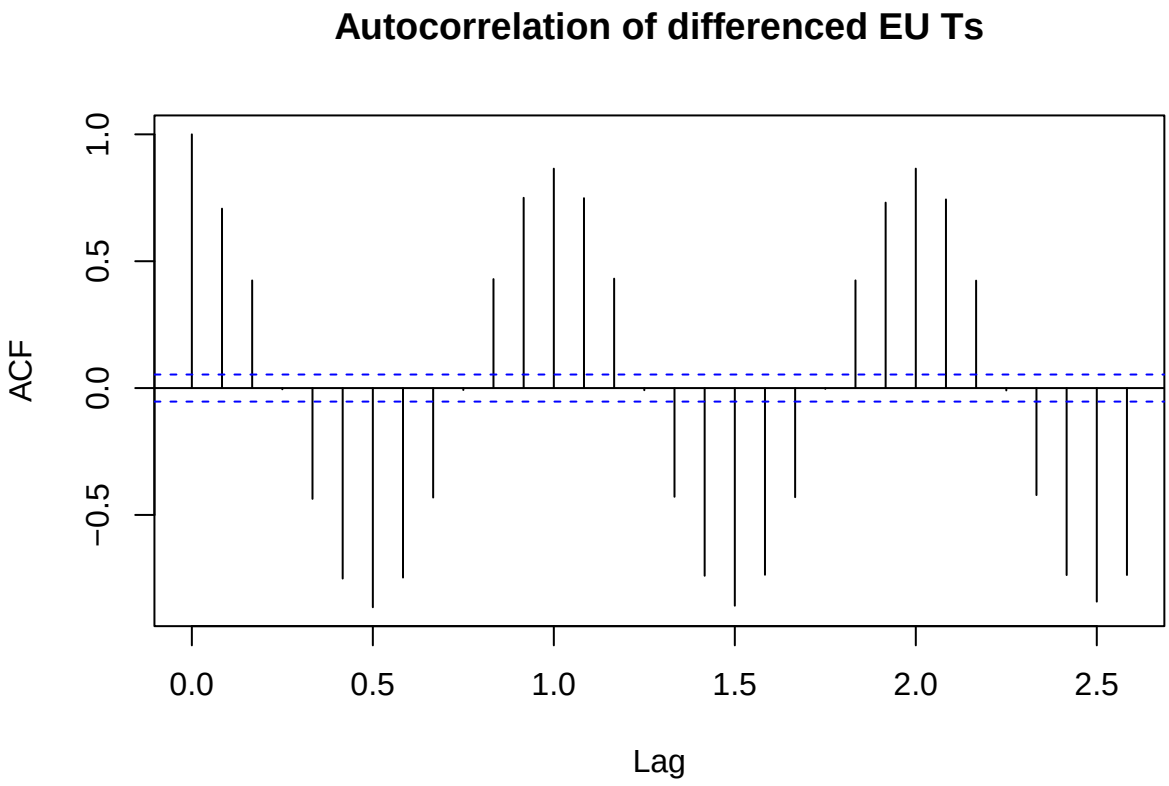

```
# Europe

autoplot(eu.train.ts, xlab = 'Time', ylab = 'Average Temperature') +
  labs(title = "EU Surface Temperature Time Series") +
  scale_y_continuous(limits = c(-20,40), breaks = c(seq(-20,40,10))) +
  theme(axis.text.x = element_text(vjust = 0.7, size = 8),
        plot.title = element_text(hjust = 0.5),
        panel.background = element_blank(),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_line(colour = 'gray'),
        axis.line.x = element_line(colour = 'gray', size=0.5, linetype='solid'))
```



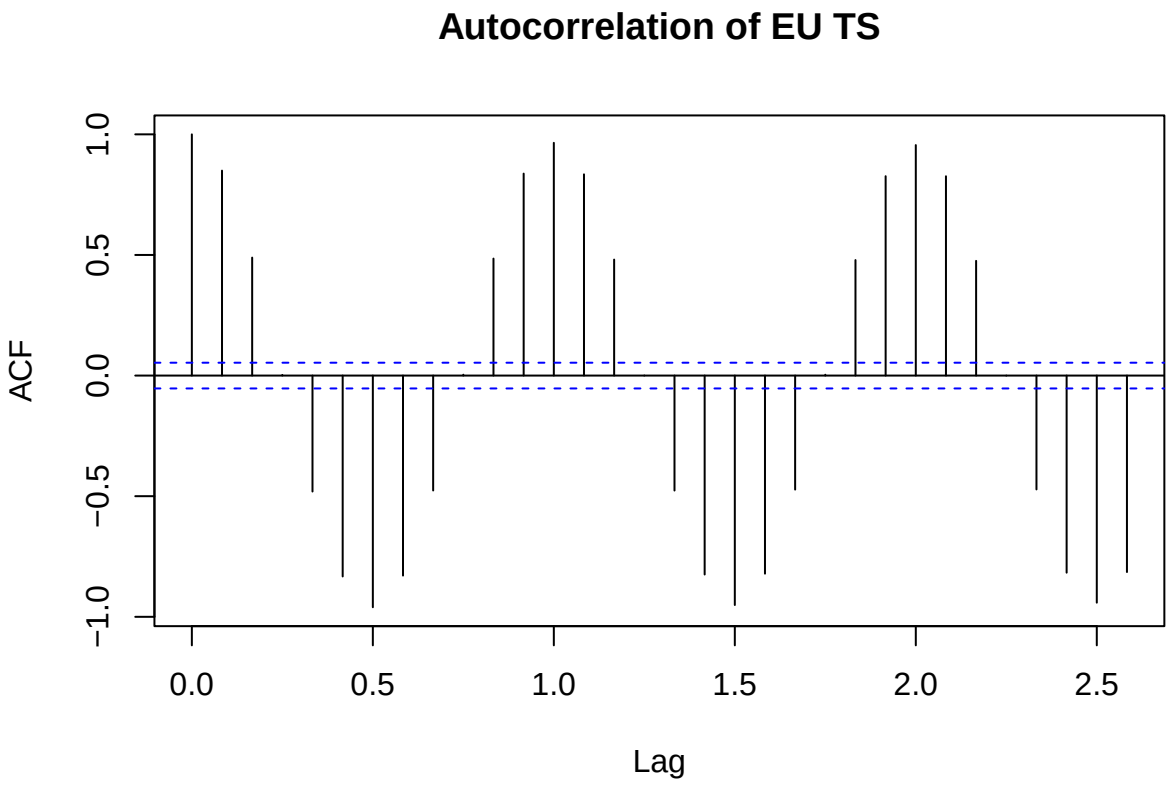

```
# Further EDA into European Time Series

autoplot(diff(eu.train.ts))
```



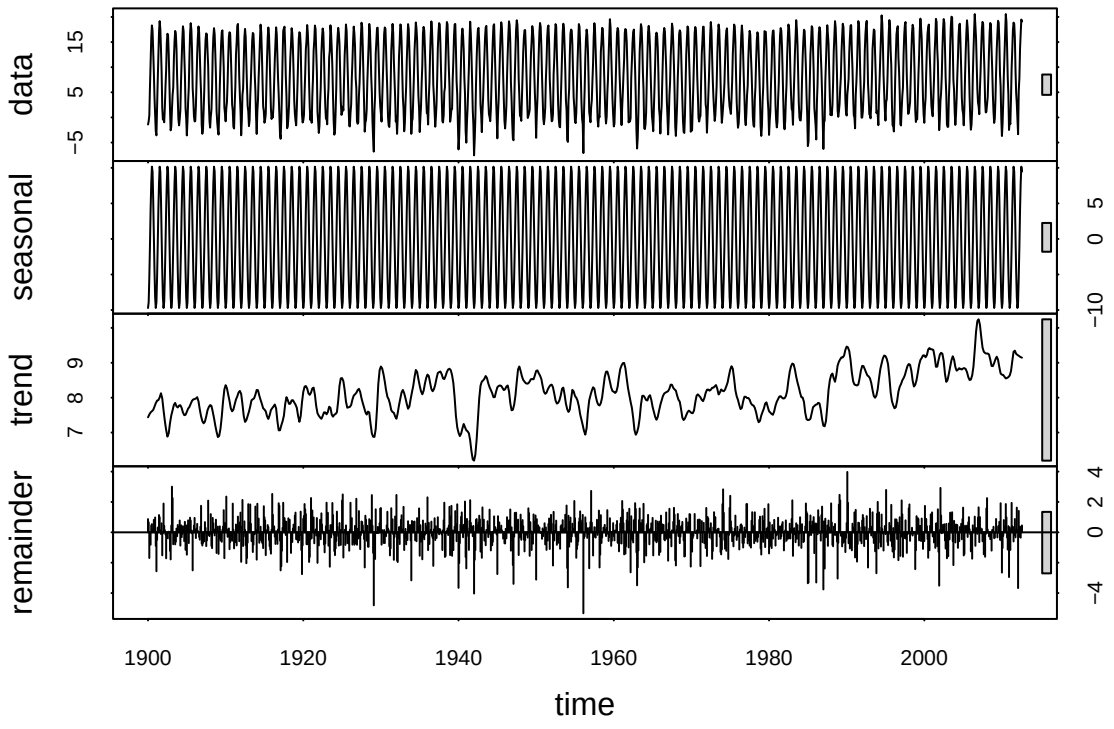

```
acf(diff(eu.train.ts), main = 'Autocorrelation of differenced EU Ts')
```



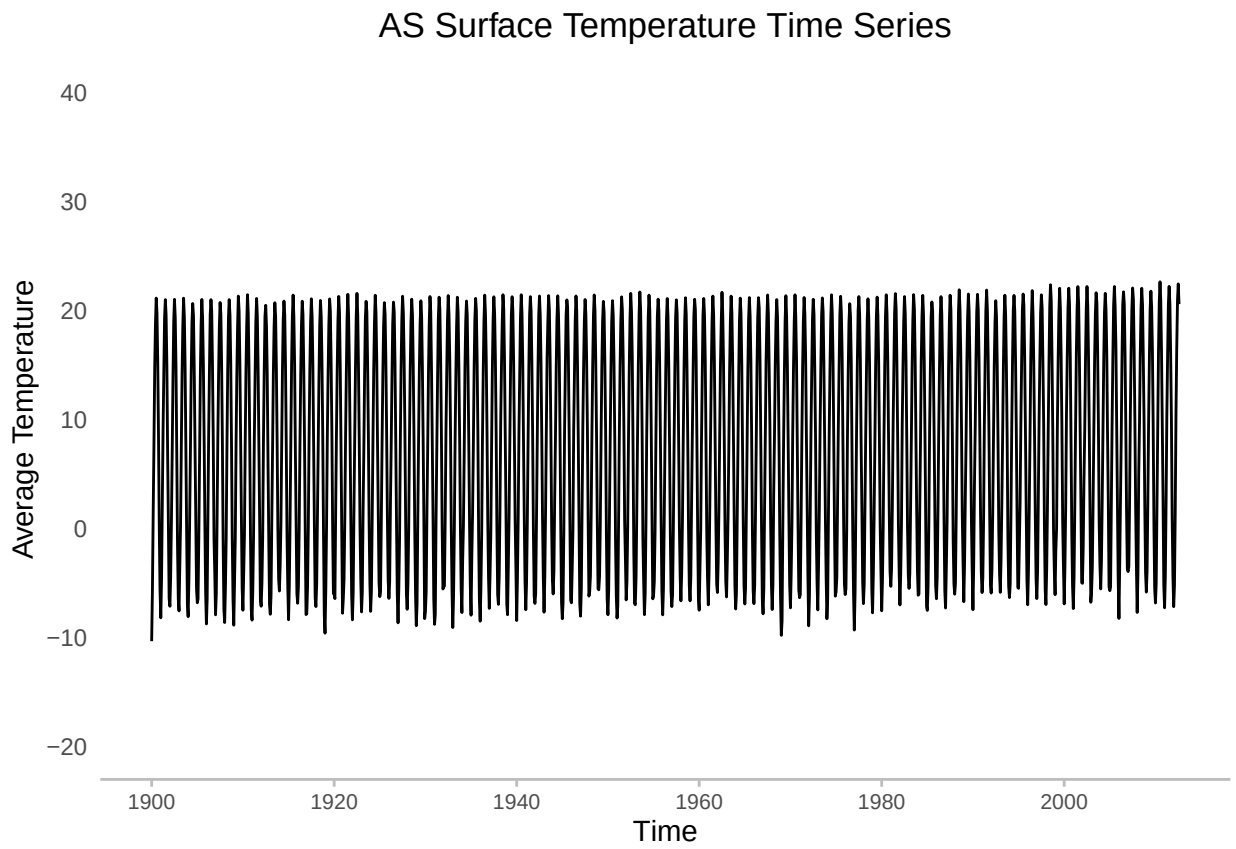

```
acf(eu.train.ts, main = 'Autocorrelation of EU TS')
```



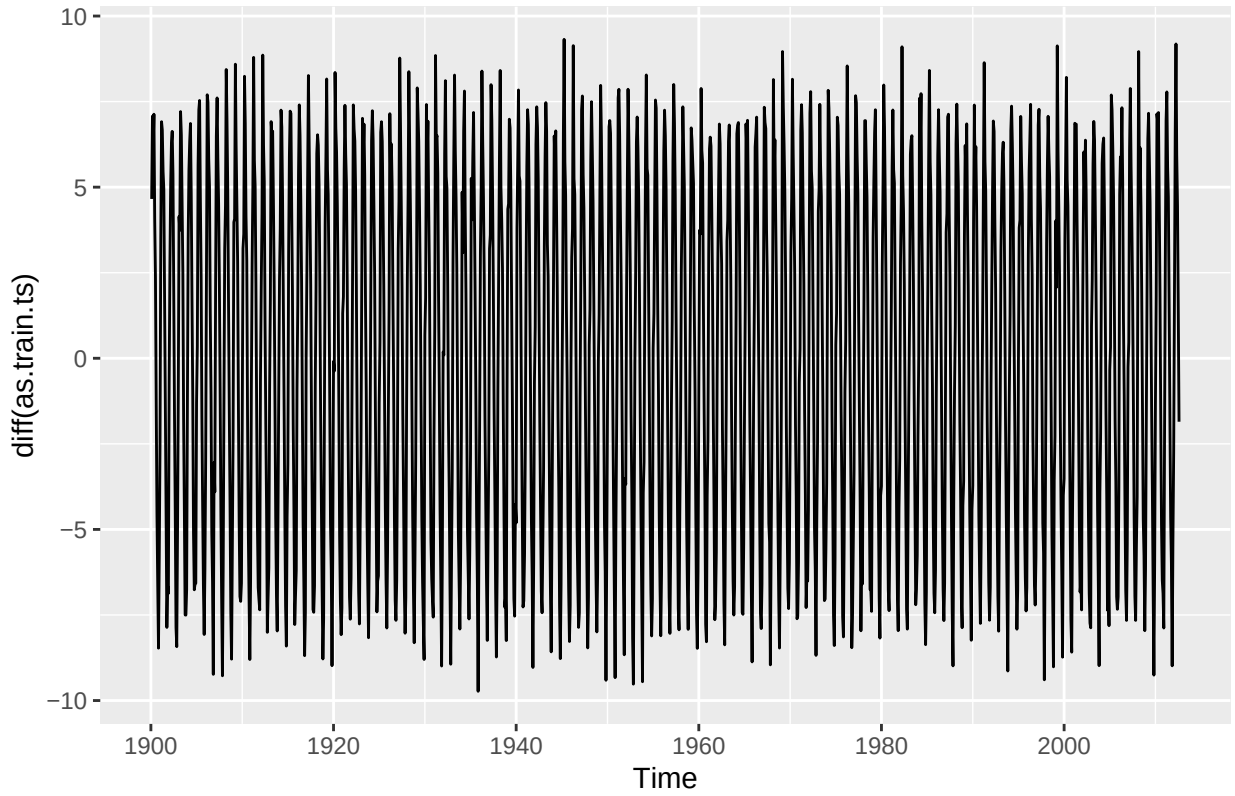

```
# stl() function to see seasonal, trend, and irregular components
eu.stl <- stl(eu.train.ts, s.window = 'periodic')
plot(eu.stl)
```



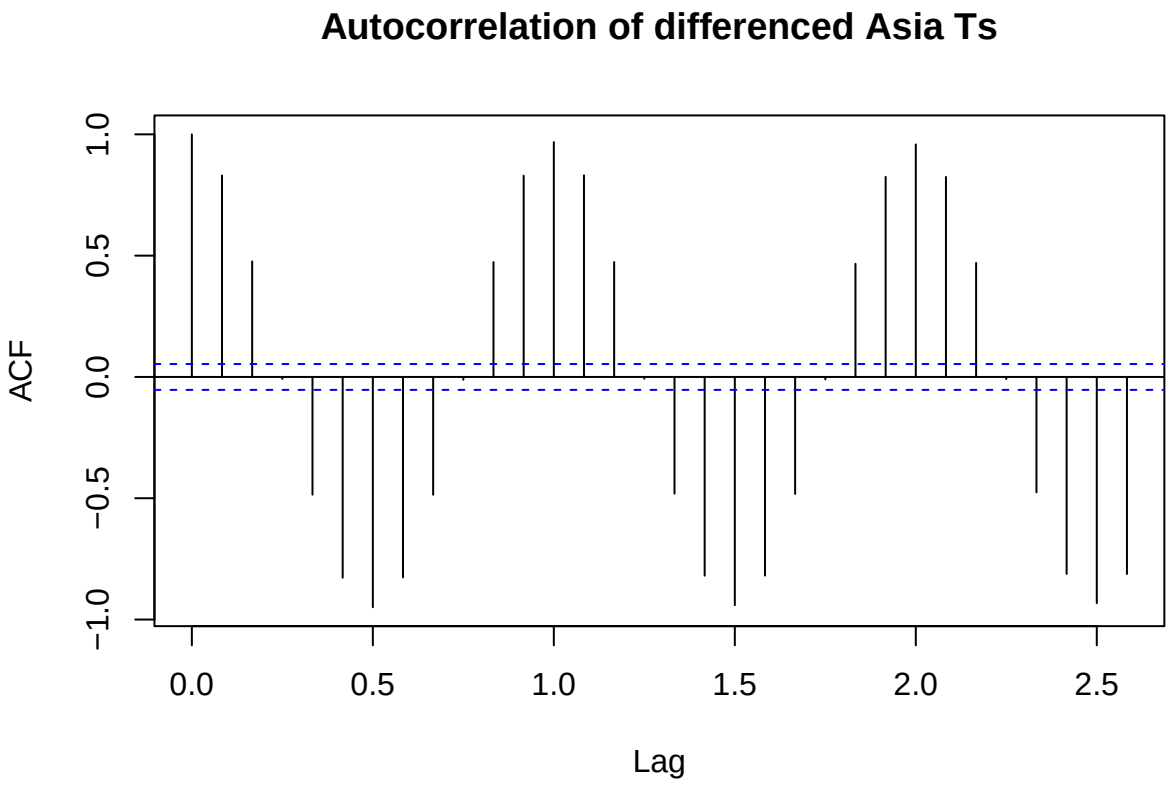

```
# Asia

autoplot(as.train.ts, xlab = 'Time', ylab = 'Average Temperature') +
  labs(title = "AS Surface Temperature Time Series") +
  scale_y_continuous(limits = c(-20,40), breaks = c(seq(-20,40,10))) +
  theme(axis.text.x = element_text(vjust = 0.7, size = 8),
        plot.title = element_text(hjust = 0.5),
        panel.background = element_blank(),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_line(colour = 'gray'),
        axis.line.x = element_line(colour = 'gray', size=0.5, linetype='solid'))
```



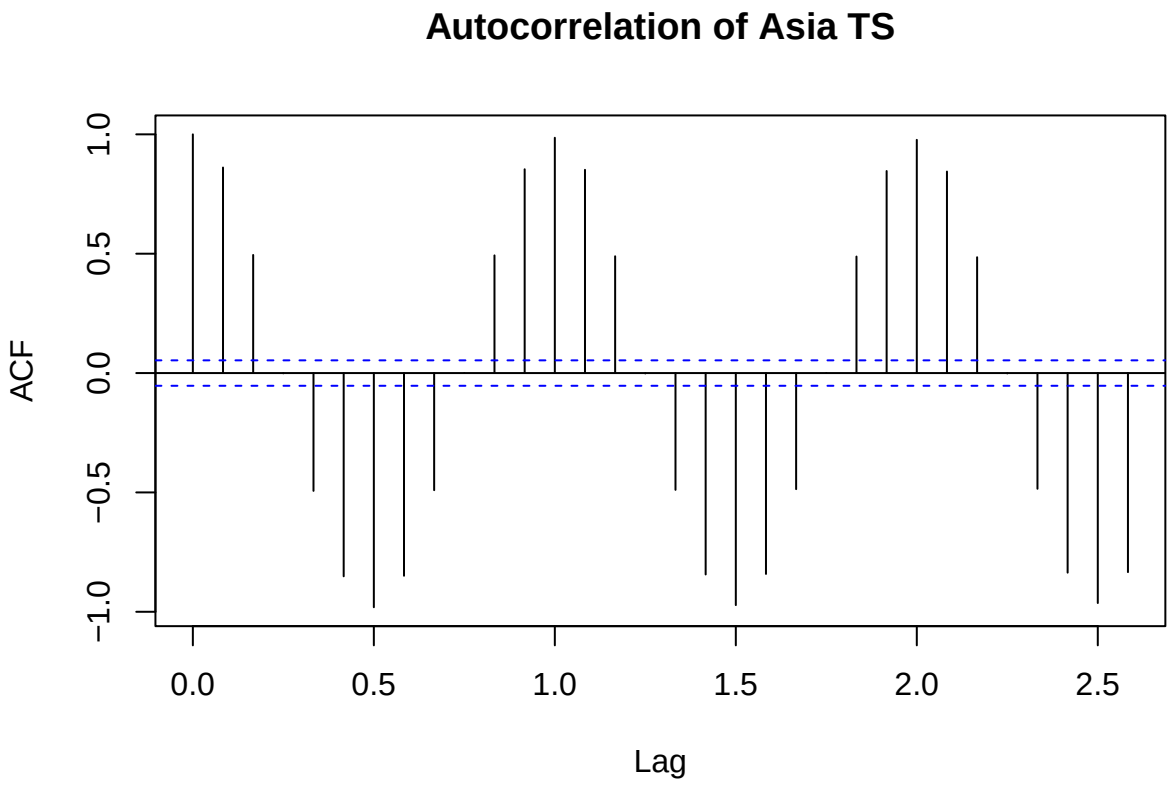

```
# Further EDA into Asia Time Series

autoplot(diff(as.train.ts))
```



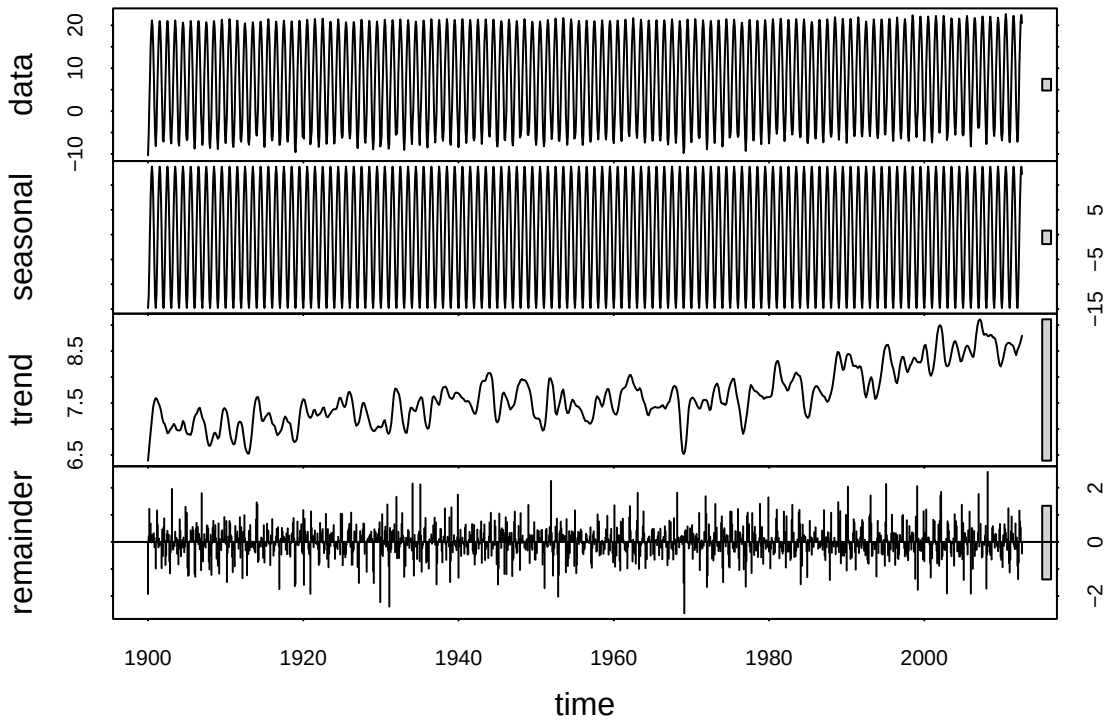

```
acf(diff(as.train.ts), main = 'Autocorrelation of differenced Asia Ts')
```



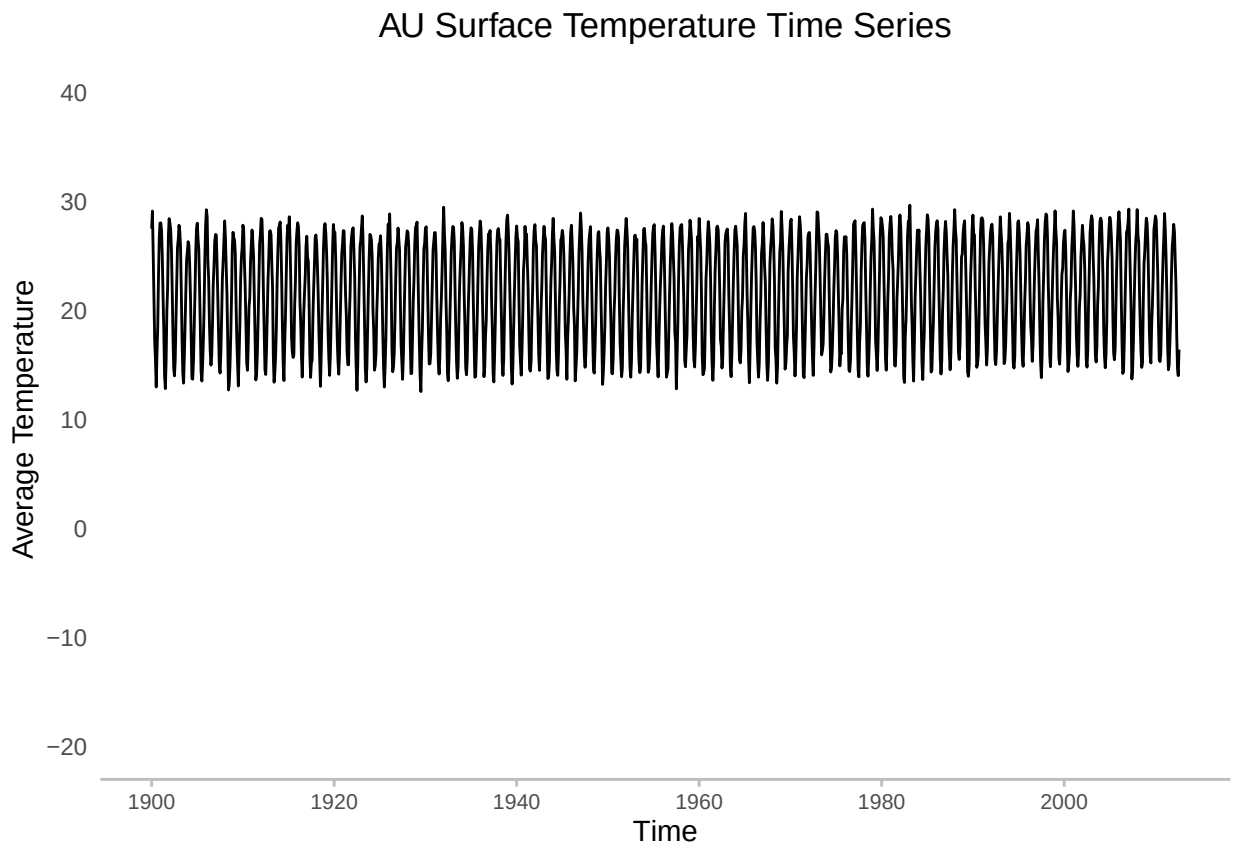

```
acf(as.train.ts, main = 'Autocorrelation of Asia TS')
```



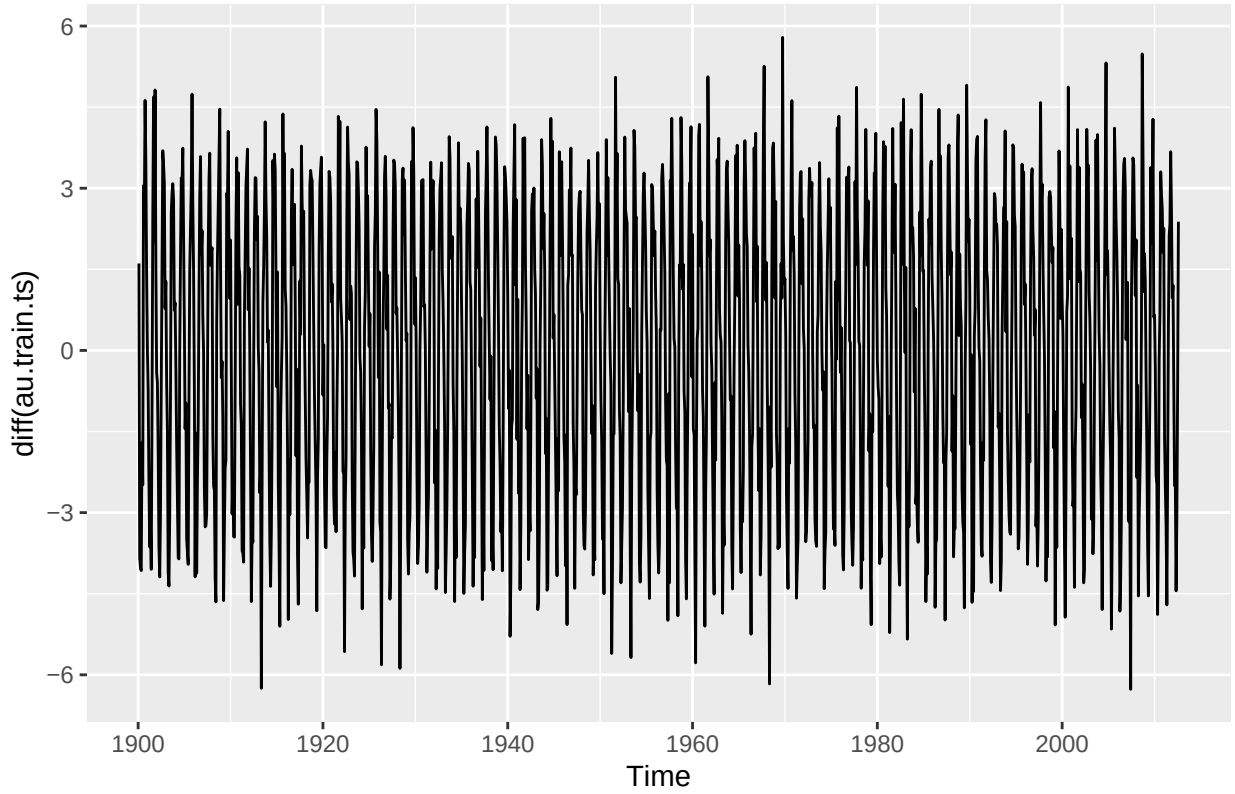

```
# stl() function to see seasonal, trend, and irregular components
as.stl <- stl(as.train.ts, s.window = 'periodic')
plot(as.stl)
```



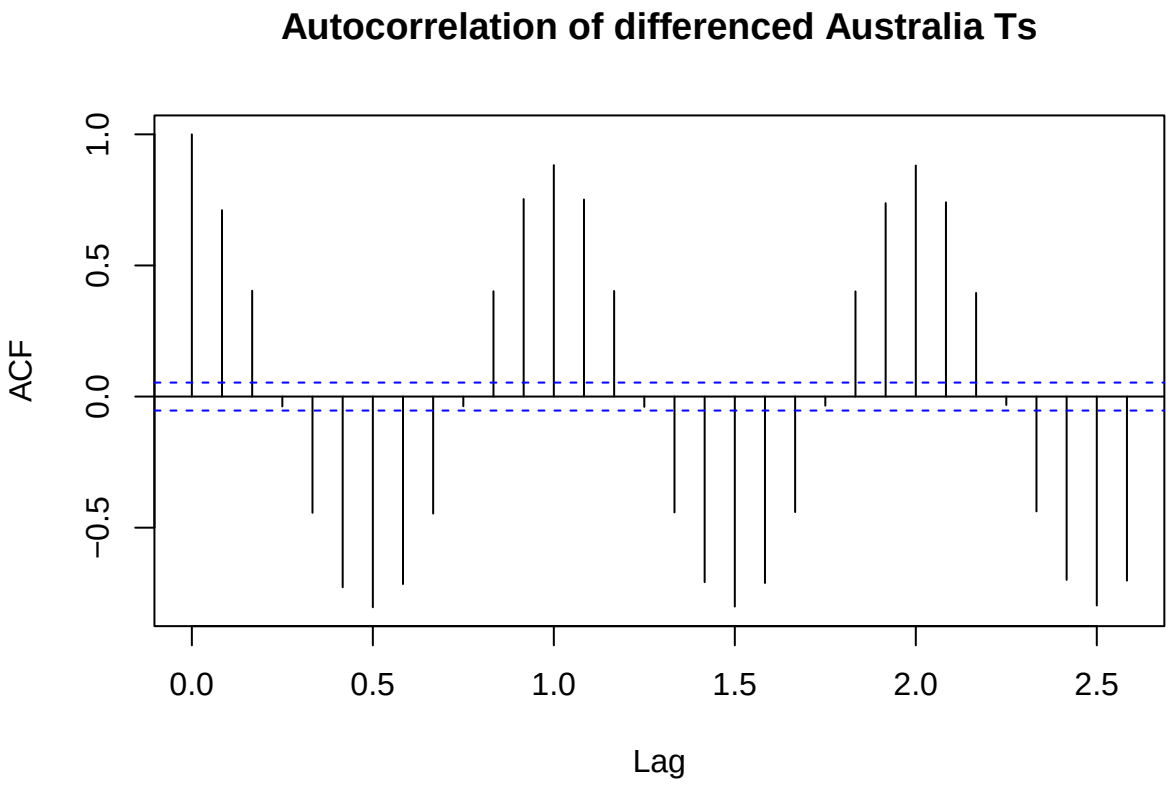

```
autoplot(au.train.ts, xlab = 'Time', ylab = 'Average Temperature') +
  labs(title = "AU Surface Temperature Time Series") +
  scale_y_continuous(limits = c(-20,40), breaks = c(seq(-20,40,10))) +
  theme(axis.text.x = element_text(vjust = 0.7, size = 8),
        plot.title = element_text(hjust = 0.5),
        panel.background = element_blank(),
        axis.ticks.y = element_blank(),
        axis.ticks.x = element_line(colour = 'gray'),
        axis.line.x = element_line(colour = 'gray', size=0.5, linetype='solid'))
```



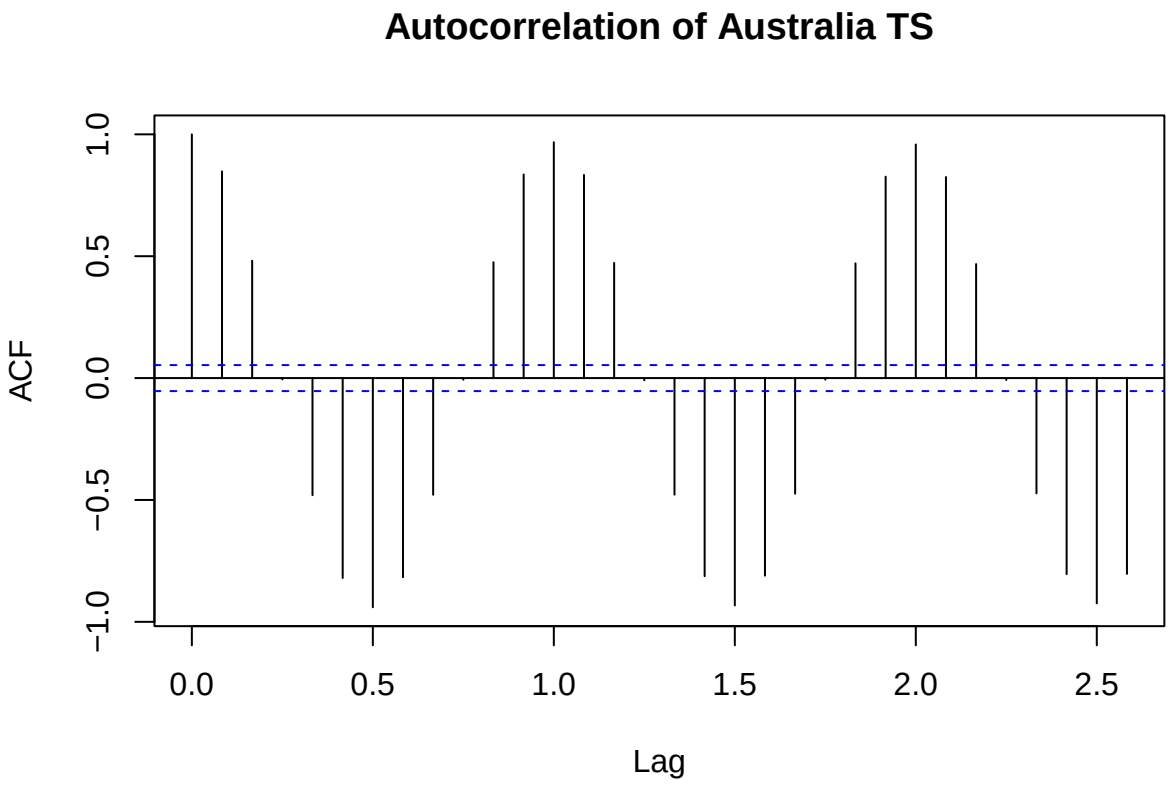

```
# Further EDA into Australia Time Series

autoplot(diff(au.train.ts))
```



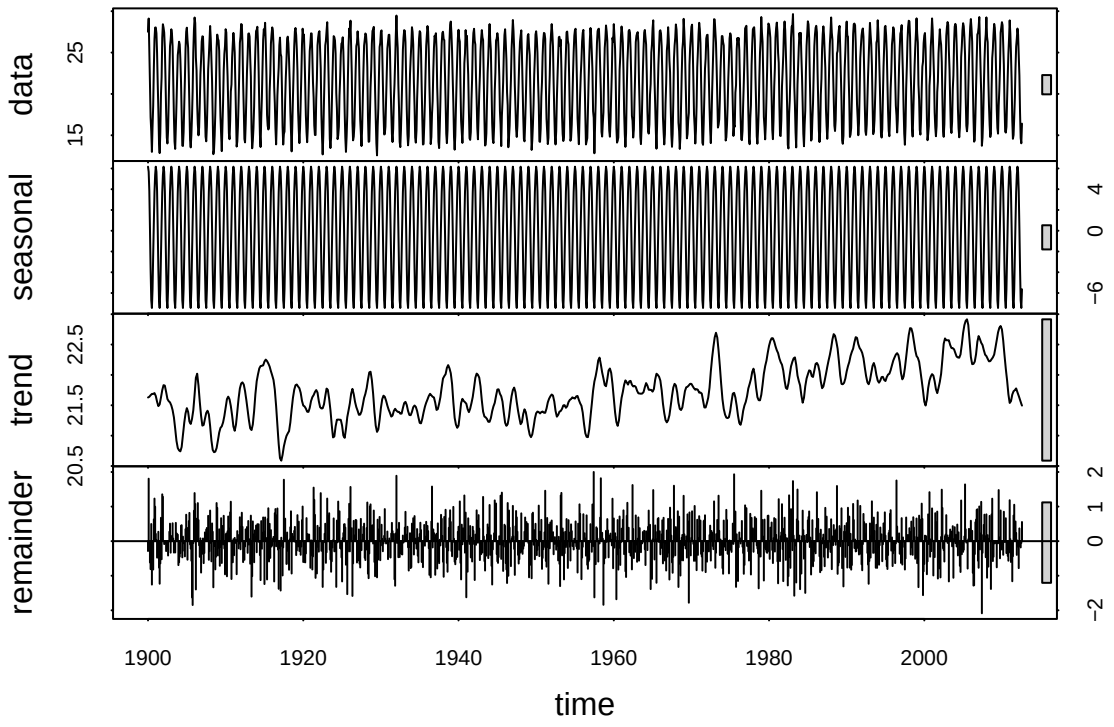

```
acf(diff(au.train.ts), main = 'Autocorrelation of differenced Australia Ts')
```




```
acf(au.train.ts, main = 'Autocorrelation of Australia TS')
```



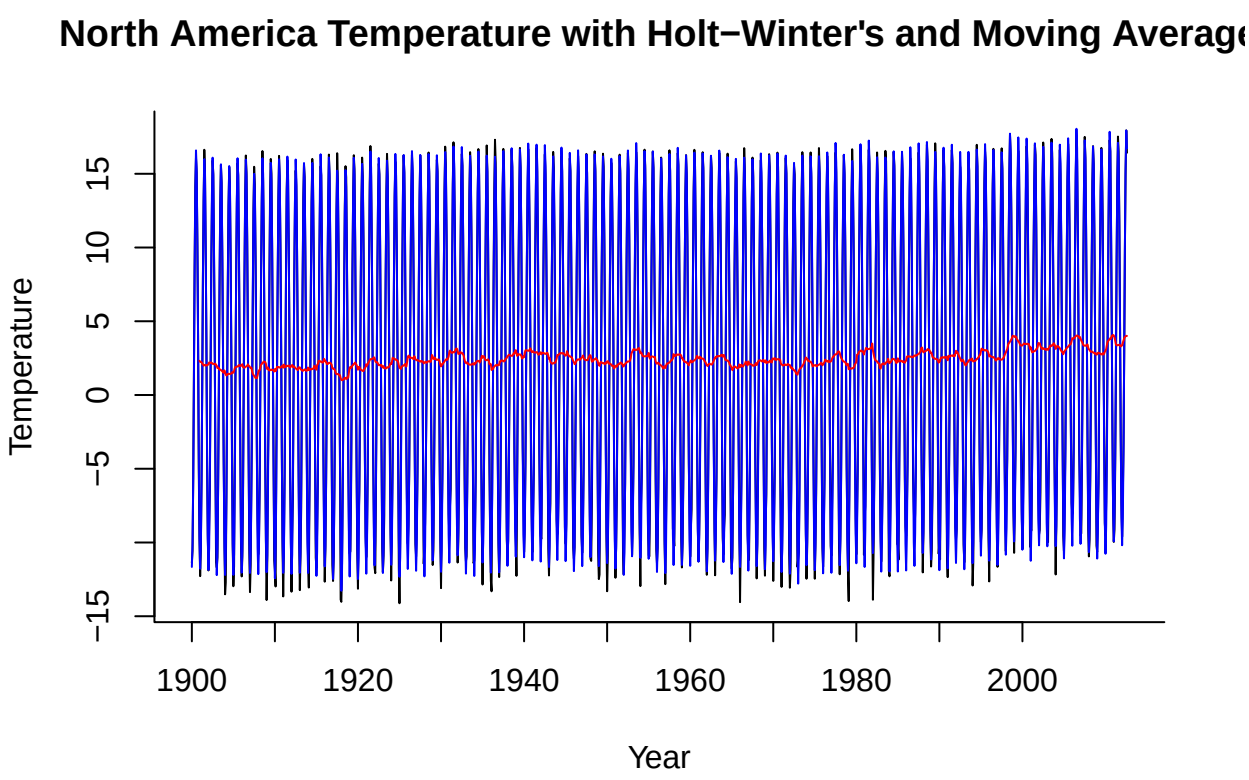

```
# stl() function to see seasonal, trend, and irregular components
au.stl <- stl(au.train.ts, s.window = 'periodic')
plot(au.stl)
```




Some takeaways here are that all continents have some upward trend of temperature. Seasonality is strong from the ACF plots with high postive correlations every 12 months / with strong negative correlations every 6 months.

As there is both trend and seasonality, let's try to do winter's exponential smoothing.

```
# North america;
NA.hw <- ets(na.train.ts, model = "ZZZ", alpha = .2)

NA.ma <- rollmean(na.train.ts, k = 12, align = 'right')

plot(na.train.ts, ylab = "Temperature", xlab = "Year", bty = "l",  xaxt = "n"
     , main = "North America Temperature with Holt-Winter's and Moving Average", flty = 2)
axis(side= 1, at = seq(1900, 2000, 10), labels = format(seq(1900, 2000, 10)))
lines(NA.hw$fitted, col = "blue")
lines(NA.ma, col = "red")
```




This seems to fit fairly well. Let's see a summary of the model

```
summary(NA.hw)
```

```
## ETS(A,N,A)
##
## Call:
##  ets(y = na.train.ts, model = "ZZZ", alpha = 0.2)
##
##   Smoothing parameters:
##     alpha = 0.2
##     gamma = 0.0208
##
##   Initial states:
##     l = 2.2897
##     s = -11.8069 -5.9527 1.4785 8.0475 12.8156 14.0017
##            11.2455 5.7559 -1.3333 -7.8977 -12.4062 -13.9479
##
##   sigma:  0.8436
##
##      AIC     AICc      BIC
## 9300.941 9301.255 9373.872
##
## Training set error measures:
##                       ME      RMSE       MAE       MPE     MAPE      MASE
## Training set 0.006054799 0.8391818 0.6195479 -1.182253 16.23162 0.7280987
```

```
##                      ACF1
## Training set 0.06543514
```

```r
# Europe
EU.hw <- ets(eu.train.ts, model = "ZZZ", alpha = .2)

EU.ma <- rollmean(eu.train.ts, k = 12, align = 'right')

plot(eu.train.ts, ylab = "Temperature", xlab = "Year", bty = "l",  xaxt = "n", main = "Europe Temperatu
axis(side= 1, at = seq(1900, 2000, 10), labels = format(seq(1900, 2000, 10)))
lines(EU.hw$fitted, col = "blue")
lines(EU.ma, col = "red")
```