Global Temperature Forecast

Group6

4/24/2019

```
library(ggplot2)
## Warning: package 'ggplot2' was built under R version
3.5.2
library(readr)
library(fpp2)
## Loading required package: forecast
## Warning: package 'forecast' was built under R
version 3.5.2
## Loading required package: fma
## Loading required package: expsmooth
library(urca)
library(fpp)
## Loading required package: lmtest
## Loading required package: zoo
## Warning: package 'zoo' was built under R version
3.5.2
##
## Attaching package: 'zoo'
## The following objects are masked from
'package:base':
##
##     as.Date, as.Date.numeric
## Loading required package: tseries
##
## Attaching package: 'fpp'
## The following objects are masked from
'package:fpp2':
```

```
##
##     ausair, ausbeer, austa, austourists, debitcards,
departures,
##     elecequip, euretail, guinearice, oil,
sunspotarea, usmelec
```

```
bfproject <- read.csv("~/Documents/Study/Semester2/
Business Forecasting/BF_Project.csv")



dim(bfproject)
```

```
## [1] 8599212       7
```

```
bfdata <- bfproject[8598940:8599211,]    #jan 1991 -
sep 2013

bfdata <- na.omit(bfdata) #removing NA's

#Changing the date format and then dividing it into
year and month respectively

bfdata$dt <-as.Date(bfdata$dt)
bfdata$Year <- format(bfdata$dt,"%Y")
bfdata$Month <- format(bfdata$dt,"%m")

head(bfdata)
```

```
##                  dt AverageTemperature
AverageTemperatureUncertainty   City
## 8598940 1991-01-01              2.942
0.152 Zwolle
## 8598941 1991-02-01             -0.999
0.153 Zwolle
## 8598942 1991-03-01              8.389
0.247 Zwolle
## 8598943 1991-04-01              8.195
0.210 Zwolle
## 8598944 1991-05-01              9.932
0.208 Zwolle
```

```
## 8598945 1991-06-01             12.928
0.169 Zwolle
##             Country Latitude Longitude Year Month
## 8598940 Netherlands   52.24N     5.26E 1991    01
## 8598941 Netherlands   52.24N     5.26E 1991    02
## 8598942 Netherlands   52.24N     5.26E 1991    03
## 8598943 Netherlands   52.24N     5.26E 1991    04
## 8598944 Netherlands   52.24N     5.26E 1991    05
## 8598945 Netherlands   52.24N     5.26E 1991    06
```

```
dim(bfdata) #272
```

```
## [1] 272   9
```

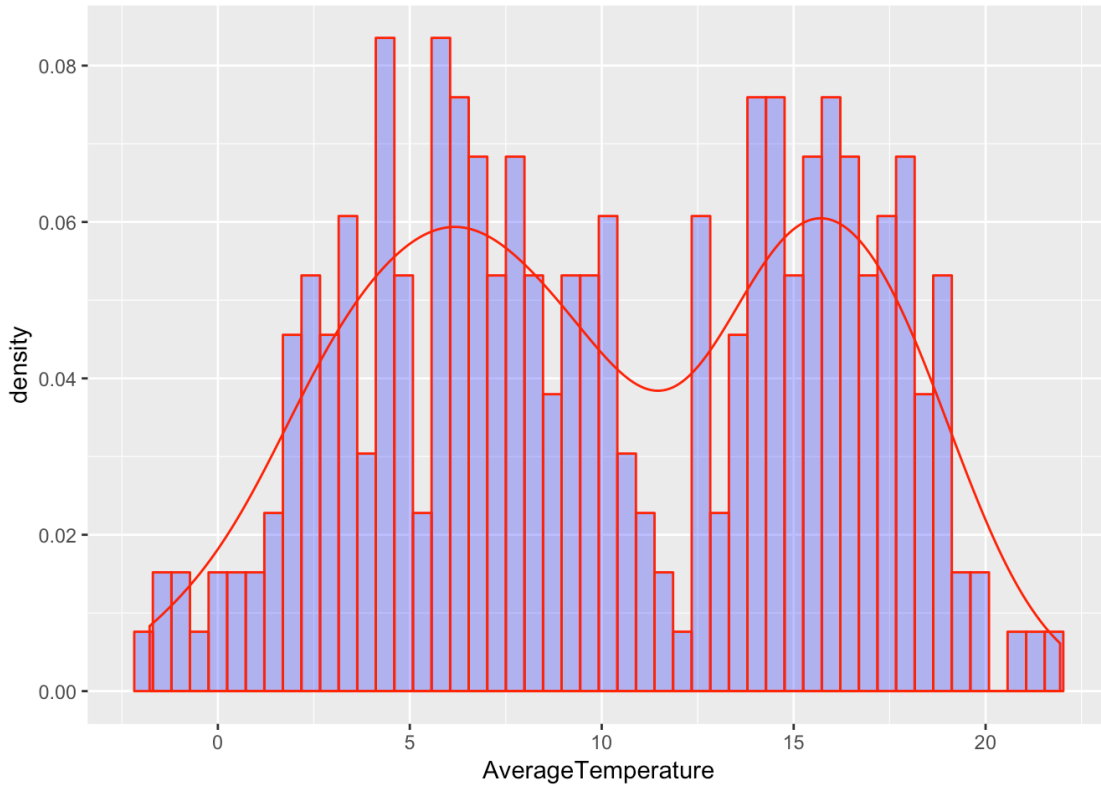
```
str(bfdata)
```

```
## 'data.frame':    272 obs. of  9 variables:
##  $ dt                           : Date, format:
"1991-01-01" "1991-02-01" ...
##  $ AverageTemperature           : num  2.942 -0.999
8.389 8.195 9.932 ...
##  $ AverageTemperatureUncertainty: num  0.152 0.153
0.247 0.21 0.208 0.169 0.217 0.233 0.26 0.136 ...
##  $ City                         : Factor w/ 3448
levels "A Coruña","Aachen",..: 3448 3448 3448 3448 3448
3448 3448 3448 3448 3448 ...
##  $ Country                      : Factor w/ 159
levels "Afghanistan",..: 101 101 101 101 101 101 101
101 101 101 ...
##  $ Latitude                     : Factor w/ 73
levels "0.80N","0.80S",..: 58 58 58 58 58 58 58 58 58
58 ...
##  $ Longitude                    : Factor w/ 1227
levels "0.00W","0.81E",..: 716 716 716 716 716 716 716
716 716 716 ...
##  $ Year                         : chr  "1991" "1991"
"1991" "1991" ...
##  $ Month                        : chr  "01" "02"
"03" "04" ...
```

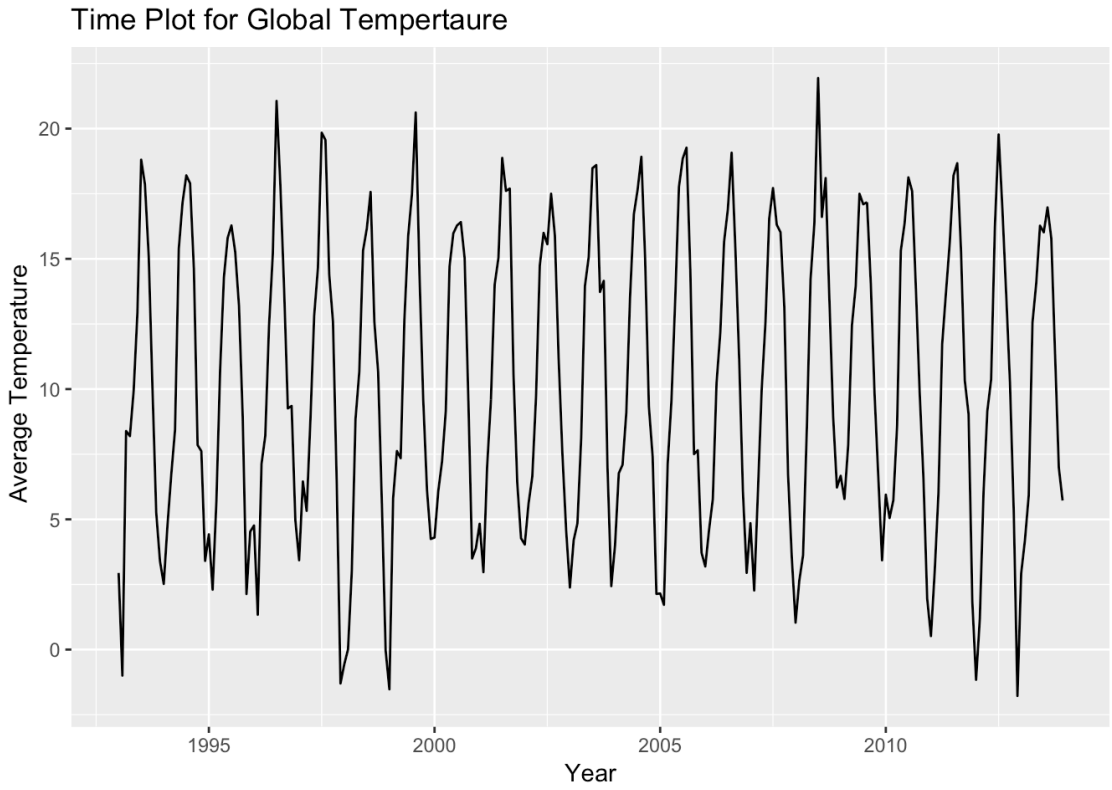
```
summary(bfdata)
```

```
##        dt              AverageTemperature
AverageTemperatureUncertainty
##  Min.   :1991-01-01   Min.   :-1.778     Min.   :
0.0780
##  1st Qu.:1996-08-24   1st Qu.: 5.603     1st Qu.:
0.1708
##  Median :2002-04-16   Median : 9.659     Median :
0.2090
##  Mean   :2002-04-16   Mean   :10.172     Mean   :
0.2130
##  3rd Qu.:2007-12-08   3rd Qu.:15.322     3rd Qu.:
0.2520
##  Max.   :2013-08-01   Max.   :21.941     Max.   :
0.4420
##
##        City            Country       Latitude
Longitude
##  Zwolle  :272   Netherlands:272   52.24N :272
5.26E  :272
##  A Coruña:  0   Afghanistan:  0   0.80N  :  0
0.00W  :  0
##  Aachen  :  0   Albania    :  0   0.80S  :  0
0.81E  :  0
##  Aalborg :  0   Algeria    :  0   10.45N :  0
0.81W  :  0
##  Aba     :  0   Angola     :  0   10.45S :  0
1.15E  :  0
##  Abadan  :  0   Argentina  :  0   12.05N :  0
1.18E  :  0
##  (Other) :  0   (Other)    :  0   (Other):  0
(Other):  0
##      Year              Month
##  Length:272         Length:272
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
```

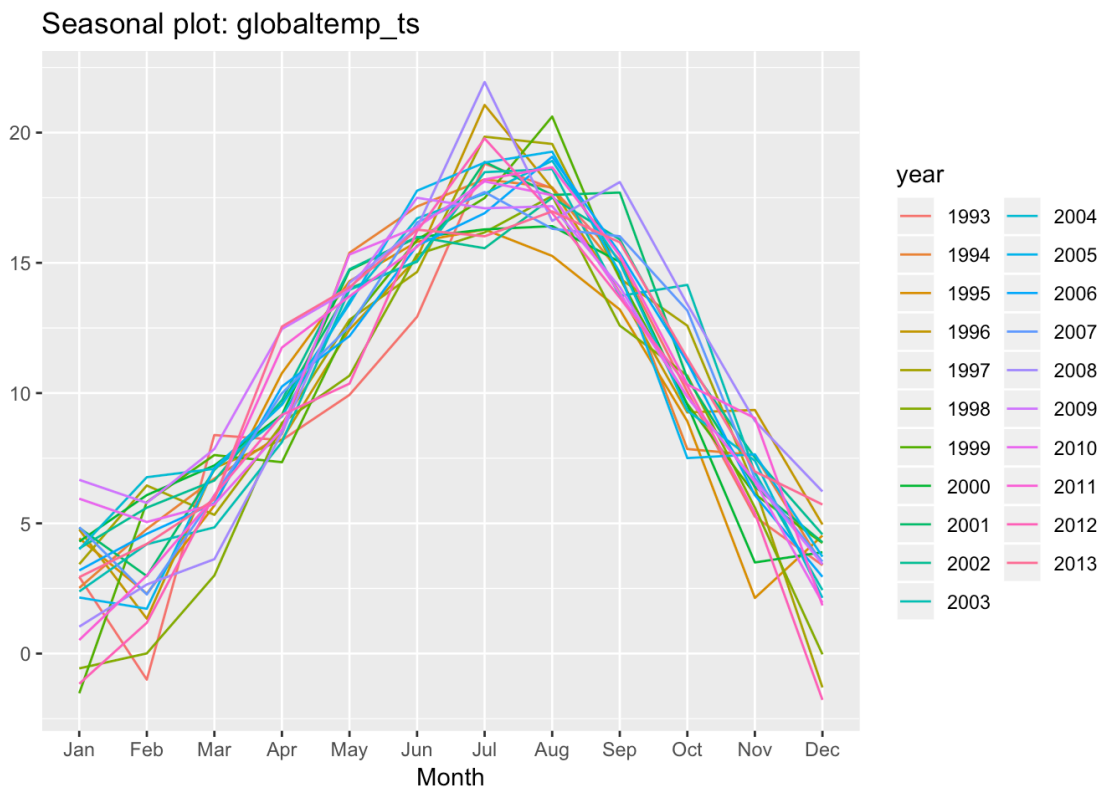
```
##
##
##
#Histogram/Density plot of the data
ggplot(bfdata,aes(x=AverageTemperature))
+geom_histogram(bins = 50, aes(y=..density..),col =
"red", fill="blue", alpha=0.3)+geom_density(col="red")
```



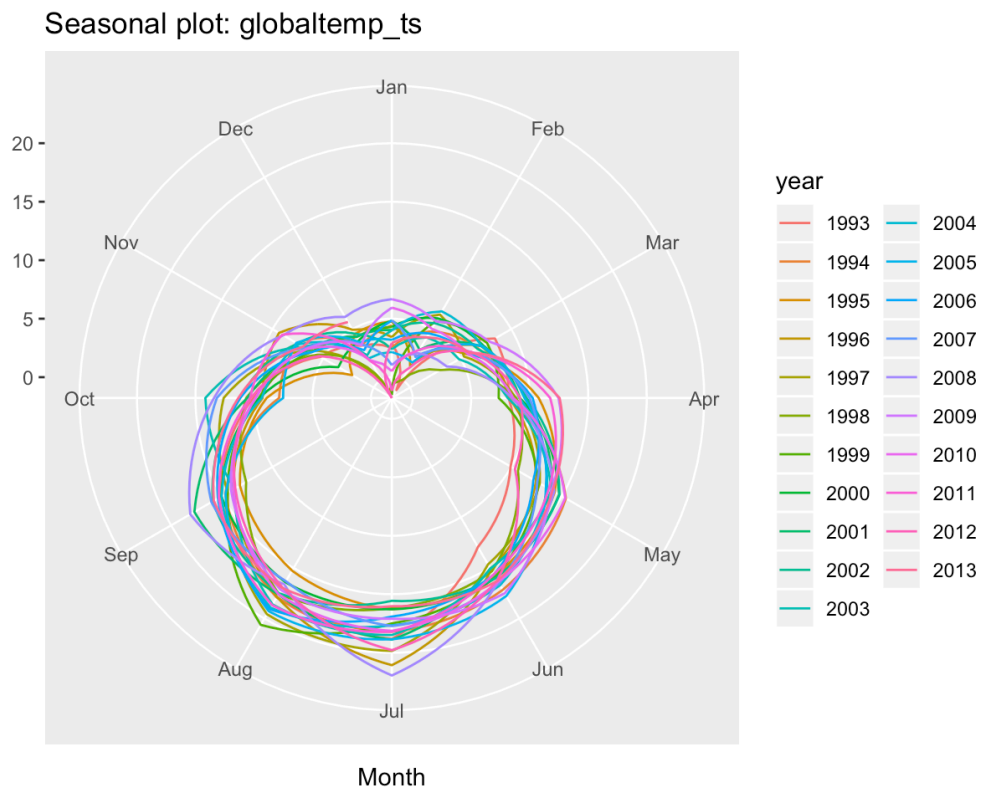

```
#Converting to Time Series data from data frame,
inorder to proceed with time series anlysis and
forecasting.
globaltemp_ts <-ts(bfdata[,
2],start=c(1993,1),end=c(2013,12),frequency=12)

#Plotting the graph to check a pattern in the data
autoplot(globaltemp_ts) + ggtitle("Time Plot for Global
Tempertaure") + xlab("Year") + ylab(" Average
Temperature")
```



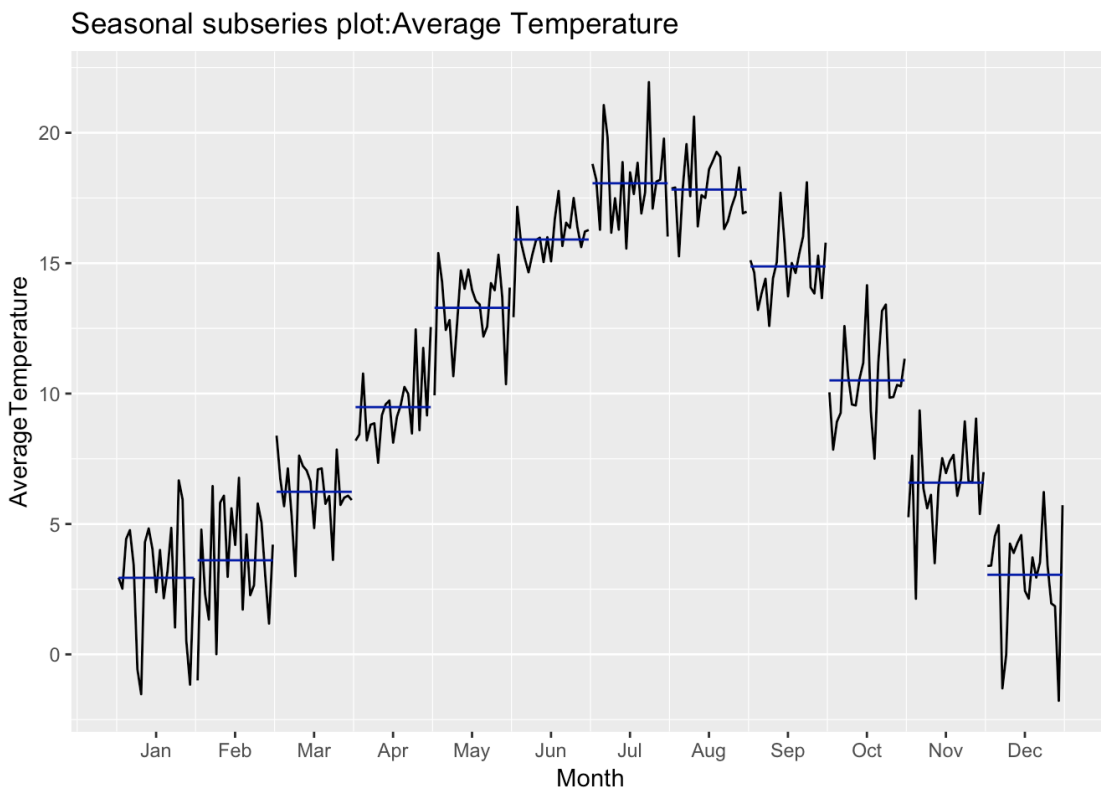

```
#In season plot we observe temperatures rising from Jan
to Aug and then starts decreasing from Sept onwards
#This cycle is followed every year
ggseasonplot(globaltemp_ts)
```



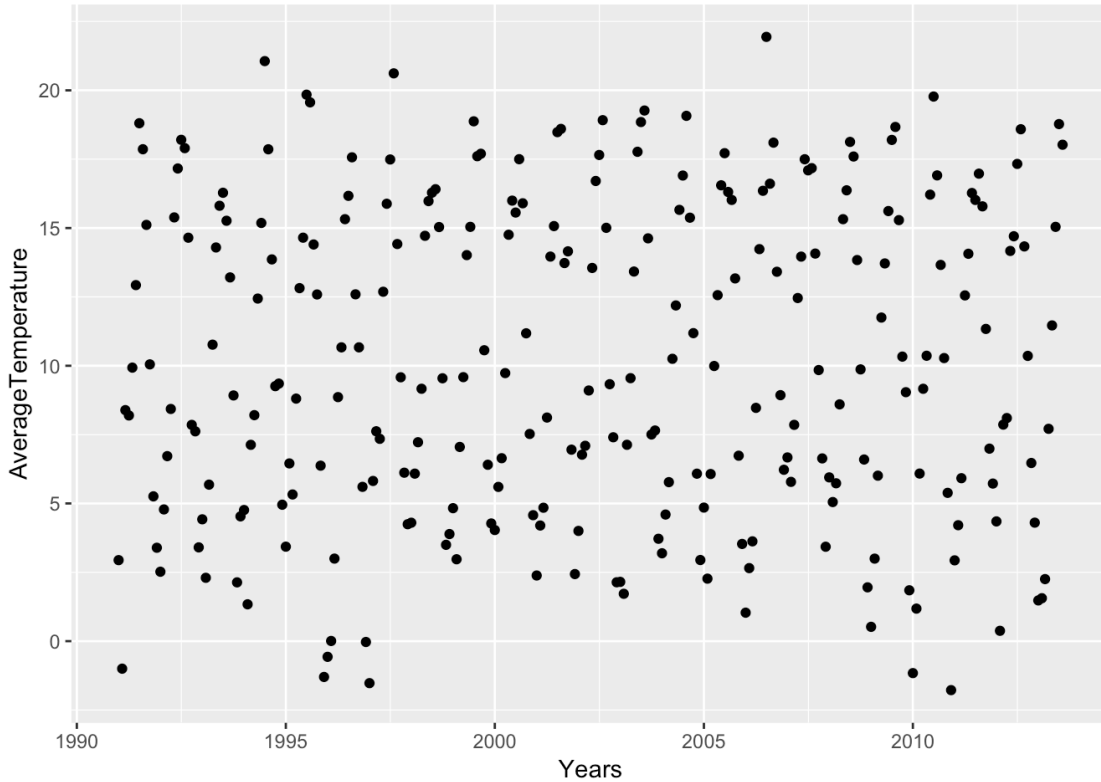

```
#Time Series of the above plot seems to be have both
seasonal and cyclic pattern.
ggseasonplot(globaltemp_ts,polar=TRUE)
```



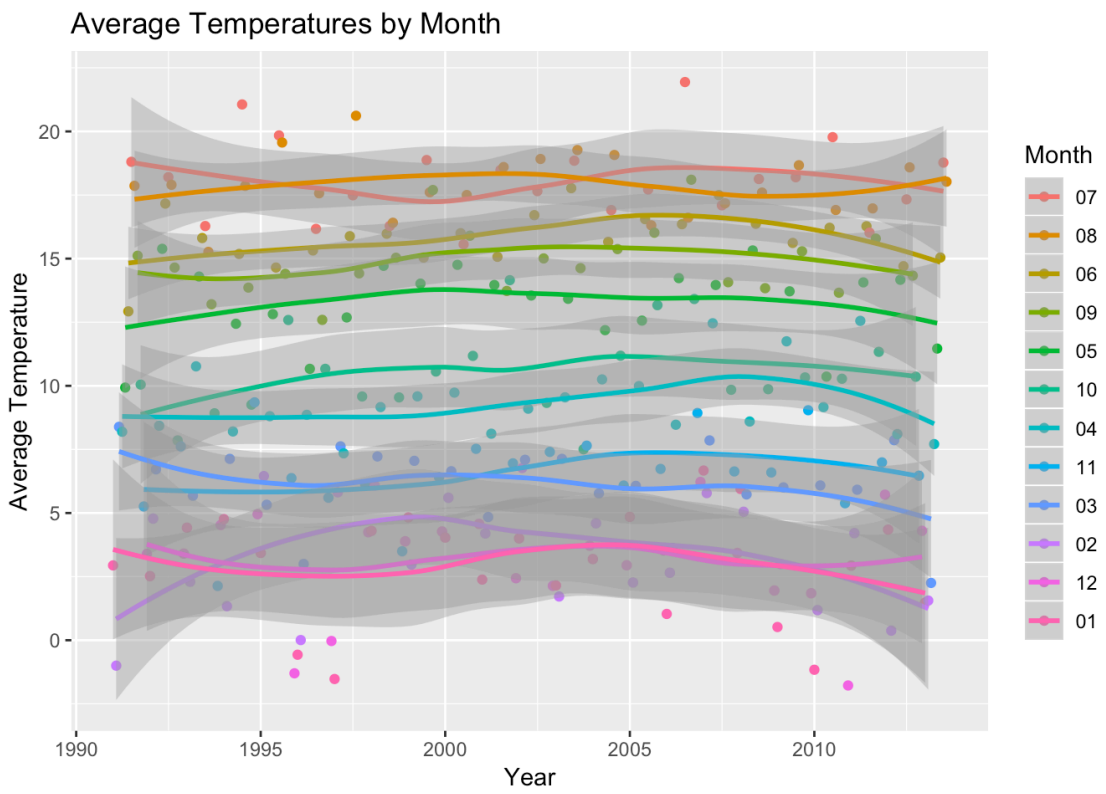

```
#We plot ggsubseriesplot to check for seasonality per
month
ggsubseriesplot(globaltemp_ts) +
ylab("AverageTemperature") + ggtitle("Seasonal
subseries plot:Average Temperature")
```



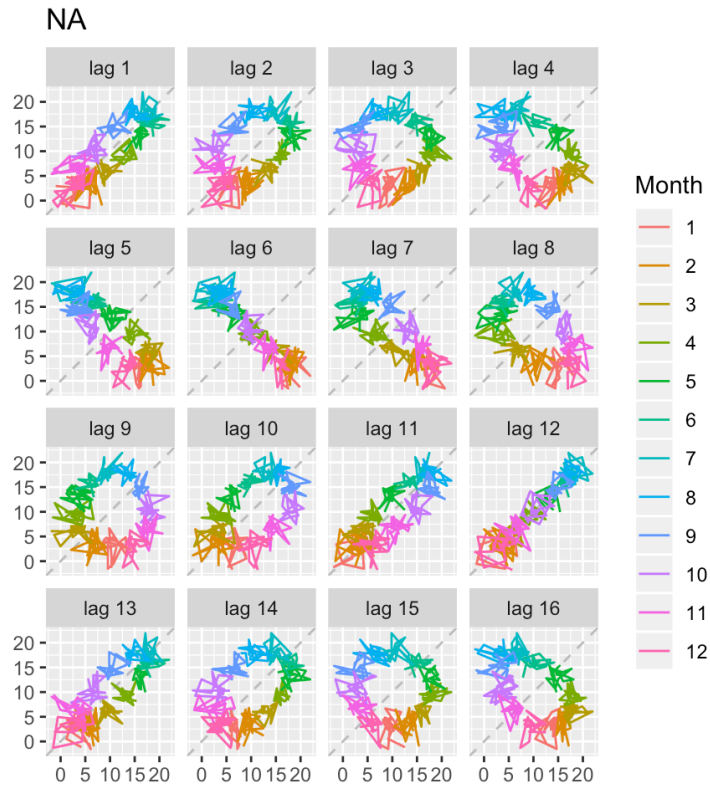

```
#We plot the scatter plot to observe any pattern over
the years
qplot(dt, AverageTemperature, data=(bfdata)) +
ylab("AverageTemperature") + xlab("Years")
```



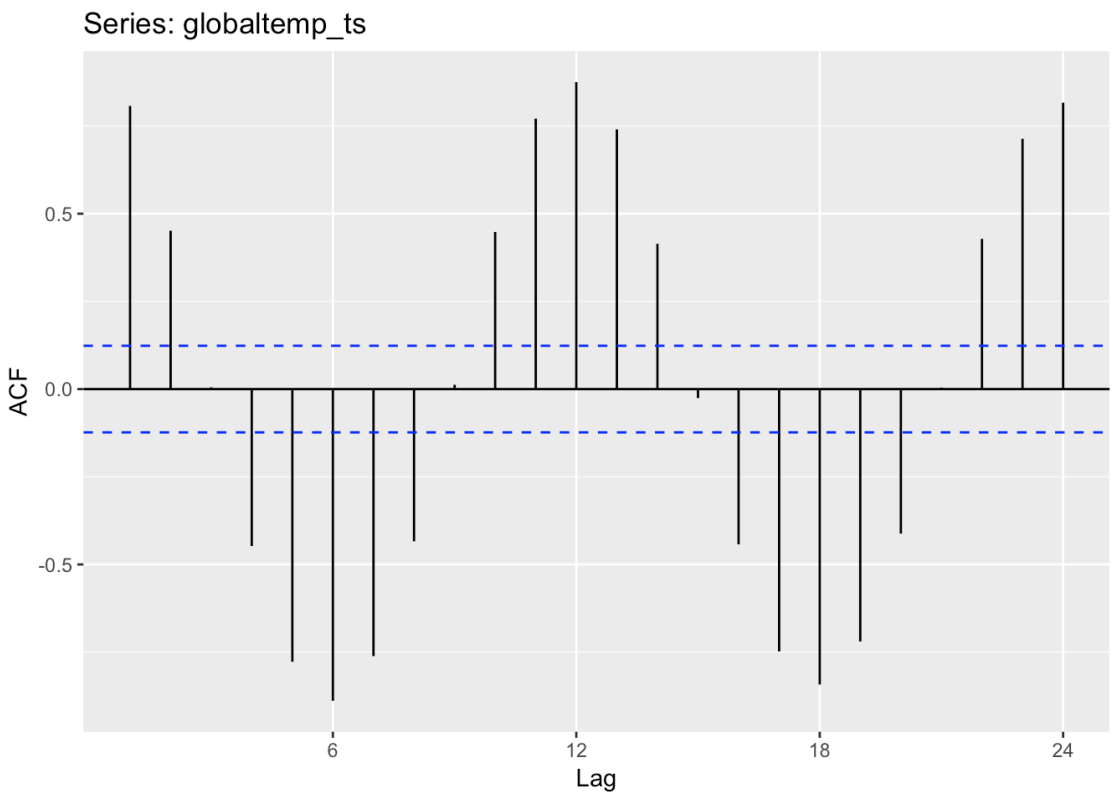

```
# As the time-frame is 20 period we do not get a very
clear picture yet we can observe some sesonality over
the years

#Plotting the average temperatures by month for a span
of 20 years we observe June and August have highest
tempertaures all over the world
ggplot(bfdata,aes(x=dt,y=AverageTemperature,colour=reor
der(Month,-AverageTemperature,mean)))+
  geom_point()+
  geom_smooth(method="loess")+
  labs(title="Average Temperatures by Month",
       x="Year",
       y="Average Temperature",
       colour="Month")
```




```
#Plotting the lag plot for temp vs months
gglagplot(globaltemp_ts)
```



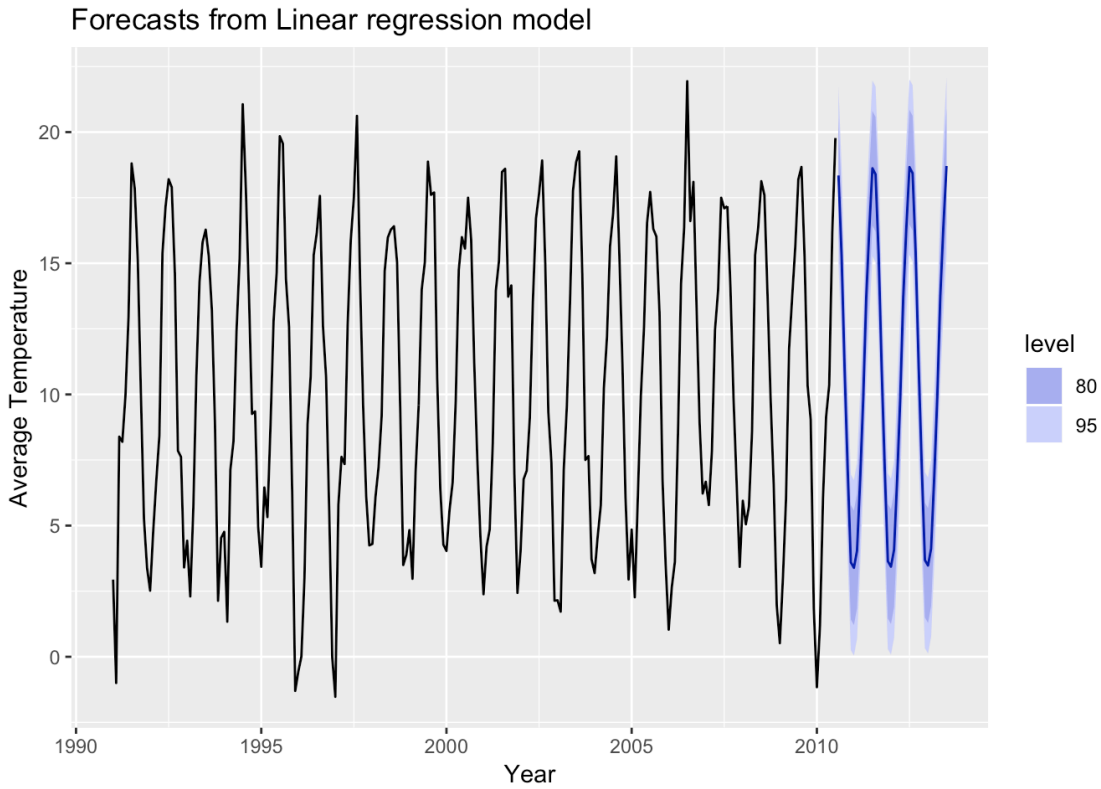

```
#Maximum autocorrelation exists at lag 12
ggAcf(globaltemp_ts)
```



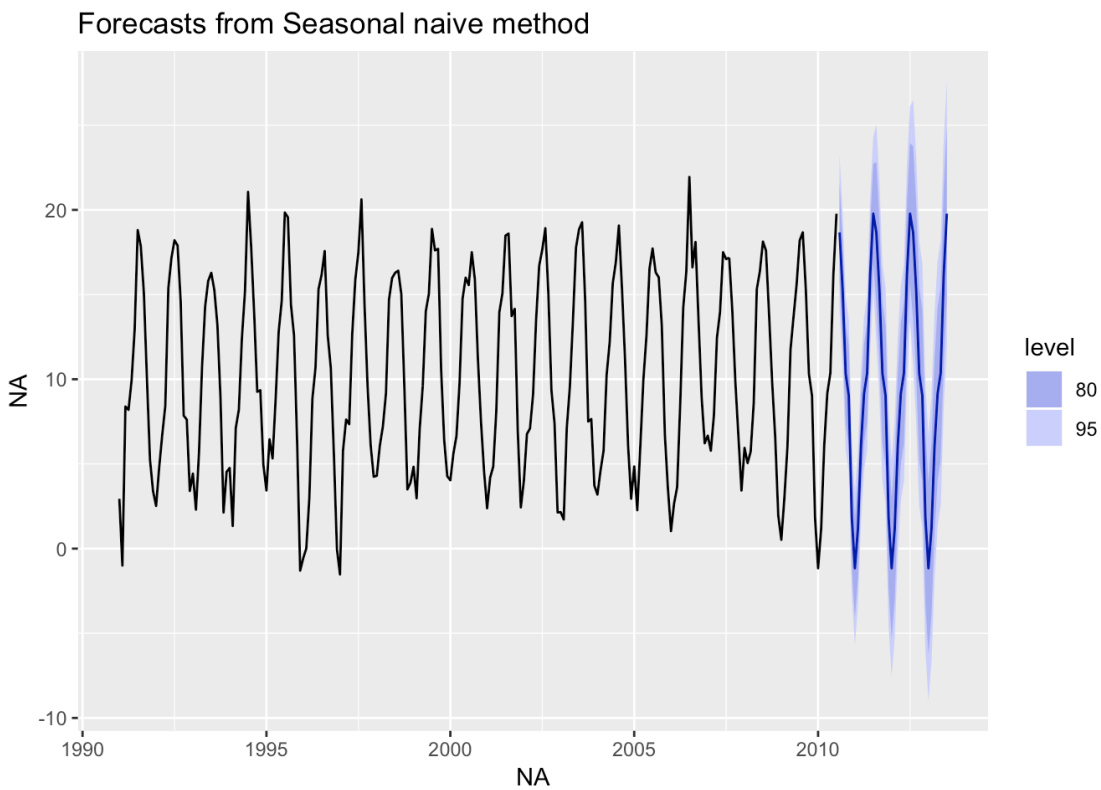

```
#Dividing the dataset into testing and training data
for model application

#Train

temp_trn <- ts(bfdata$AverageTemperature, start =
c(1991,1) , end= c(2010,7), frequency = 12)

#Test
temp_tst <- ts(bfdata$AverageTemperature, start =
c(2010,8) , end= c(2013,08), frequency = 12)


#TSLM - Time Series Regression
lr_model = tslm(temp_trn ~ trend + season)
fc <- forecast(lr_model, h=36)
accuracy(fc, temp_tst)
##                          ME       RMSE       MAE
MPE     MAPE     MASE
```

```
## Training set -4.625714e-17  1.596018 1.237998
-159.36220 277.0025 0.715678
## Test set     -9.696697e-01 10.694403 9.619405
-60.30753 207.4116 5.560912
##                   ACF1 Theil's U
## Training set 0.2237544        NA
## Test set     0.7955392  1.352915
```

```
autoplot(fc) + ylab("Average Temperature") +
xlab("Year")
```




```
#Seasonal Naive Method

Naive_tmp <- snaive(temp_trn, 36)

autoplot(Naive_tmp)
```




```
summary(Naive_tmp)
##
## Forecast method: Seasonal naive method
##
## Model Information:
## Call: snaive(y = temp_trn, h = 36)
##
## Residual sd: 2.2946
##
## Error measures:
##                      ME     RMSE      MAE      MPE
MAPE MASE      ACF1
## Training set 0.02211211 2.289583 1.729825 -435.585
468.4722    1 0.2297885
##
## Forecasts:
##          Point Forecast     Lo 80     Hi 80      Lo
95     Hi 95
## Aug 2010         18.671 15.736781 21.605219
```

```
14.1834992 23.158501
## Sep 2010         15.288 12.353781 18.222219
10.8004992 19.775501
## Oct 2010         10.332  7.397781 13.266219
5.8444992 14.819501
## Nov 2010          9.040  6.105781 11.974219
4.5524992 13.527501
## Dec 2010          1.847 -1.087219  4.781219
-2.6405008  6.334501
## Jan 2011         -1.160 -4.094219  1.774219
-5.6475008  3.327501
## Feb 2011          1.182 -1.752219  4.116219
-3.3055008  5.669501
## Mar 2011          6.085  3.150781  9.019219
1.5974992 10.572501
## Apr 2011          9.164  6.229781 12.098219
4.6764992 13.651501
## May 2011         10.363  7.428781 13.297219
5.8754992 14.850501
## Jun 2011         16.213 13.278781 19.147219
11.7254992 20.700501
## Jul 2011         19.775 16.840781 22.709219
15.2874992 24.262501
## Aug 2011         18.671 14.521388 22.820612
12.3247155 25.017284
## Sep 2011         15.288 11.138388 19.437612
8.9417155 21.634284
## Oct 2011         10.332  6.182388 14.481612
3.9857155 16.678284
## Nov 2011          9.040  4.890388 13.189612
2.6937155 15.386284
## Dec 2011          1.847 -2.302612  5.996612
-4.4992845  8.193284
## Jan 2012         -1.160 -5.309612  2.989612
-7.5062845  5.186284
## Feb 2012          1.182 -2.967612  5.331612
```

```
-5.1642845  7.528284
## Mar 2012          6.085  1.935388 10.234612
-0.2612845 12.431284
## Apr 2012          9.164  5.014388 13.313612
2.8177155 15.510284
## May 2012         10.363  6.213388 14.512612
4.0167155 16.709284
## Jun 2012         16.213 12.063388 20.362612
9.8667155 22.559284
## Jul 2012         19.775 15.625388 23.924612
13.4287155 26.121284
## Aug 2012         18.671 13.588784 23.753216
10.8984206 26.443579
## Sep 2012         15.288 10.205784 20.370216
7.5154206 23.060579
## Oct 2012         10.332  5.249784 15.414216
2.5594206 18.104579
## Nov 2012          9.040  3.957784 14.122216
1.2674206 16.812579
## Dec 2012          1.847 -3.235216  6.929216
-5.9255794  9.619579
## Jan 2013         -1.160 -6.242216  3.922216
-8.9325794  6.612579
## Feb 2013          1.182 -3.900216  6.264216
-6.5905794  8.954579
## Mar 2013          6.085  1.002784 11.167216
-1.6875794 13.857579
## Apr 2013          9.164  4.081784 14.246216
1.3914206 16.936579
## May 2013         10.363  5.280784 15.445216
2.5904206 18.135579
## Jun 2013         16.213 11.130784 21.295216
8.4404206 23.985579
## Jul 2013         19.775 14.692784 24.857216
12.0024206 27.547579
```

```
accuracy(Naive_tmp , temp_tst)
```

```
##                          ME      RMSE       MAE
MPE     MAPE     MASE
## Training set  0.022112108  2.289583  1.729825
-435.58497 468.4722 1.000000
## Test set     -0.007722222 11.744773 10.184111
-55.11688 210.9712 5.887365
##                    ACF1 Theil's U
## Training set 0.2297885        NA
## Test set     0.7871040  1.389437
```

```
#Checking the data
test=ur.kpss(temp_trn)
summary(test)
```

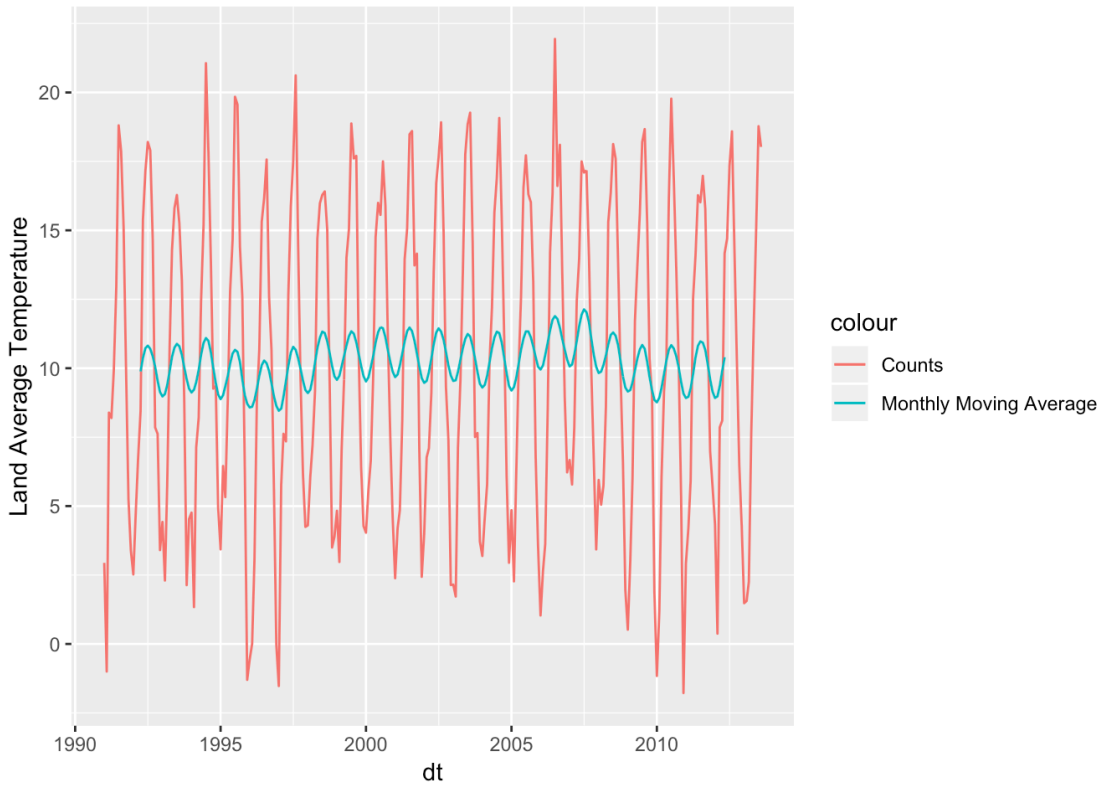
```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 4 lags.
##
## Value of test-statistic is: 0.0343
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

```
ndiffs(temp_trn)
```

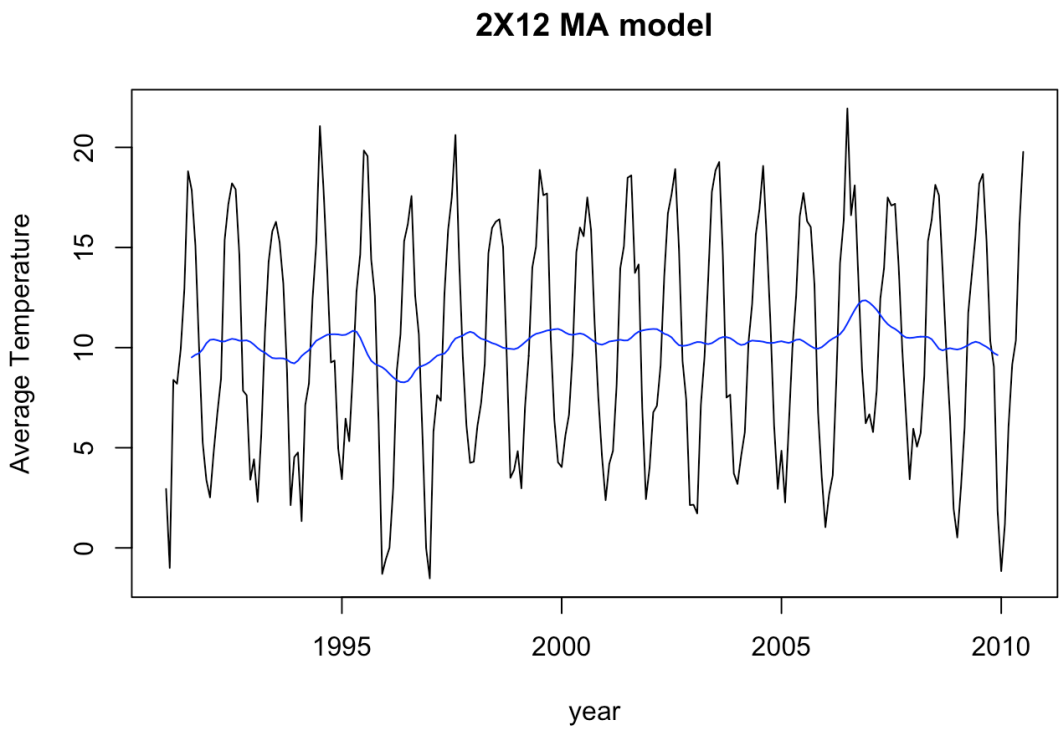
```
## [1] 0
```

```
#Decomposition
#Applying Moving Average for m=30
bfdata$lat_ma30 = ma(bfdata$AverageTemperature,
order=30)
ggplot() +
  geom_line(data = bfdata, aes(x = dt, y =
AverageTemperature, colour = "Counts")) +
  geom_line(data = bfdata, aes(x = dt, y = lat_ma30,
colour = "Monthly Moving Average"))  +
  ylab('Land Average Temperature')
```

```
## Warning: Removed 30 rows containing missing values
(geom_path).
```

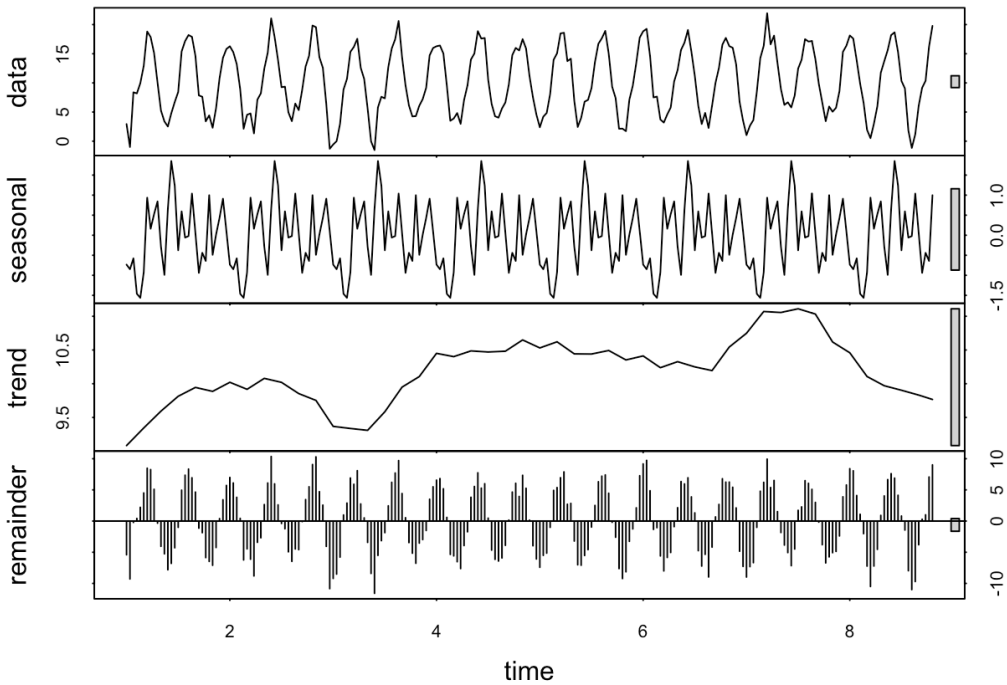



```
# Moving Averages 2x12 model
plot(temp_trn, main ="2X12 MA model", ylab="Average
Temperature", xlab="year")
lines(ma(ma(temp_trn,2),12),col="blue")
```

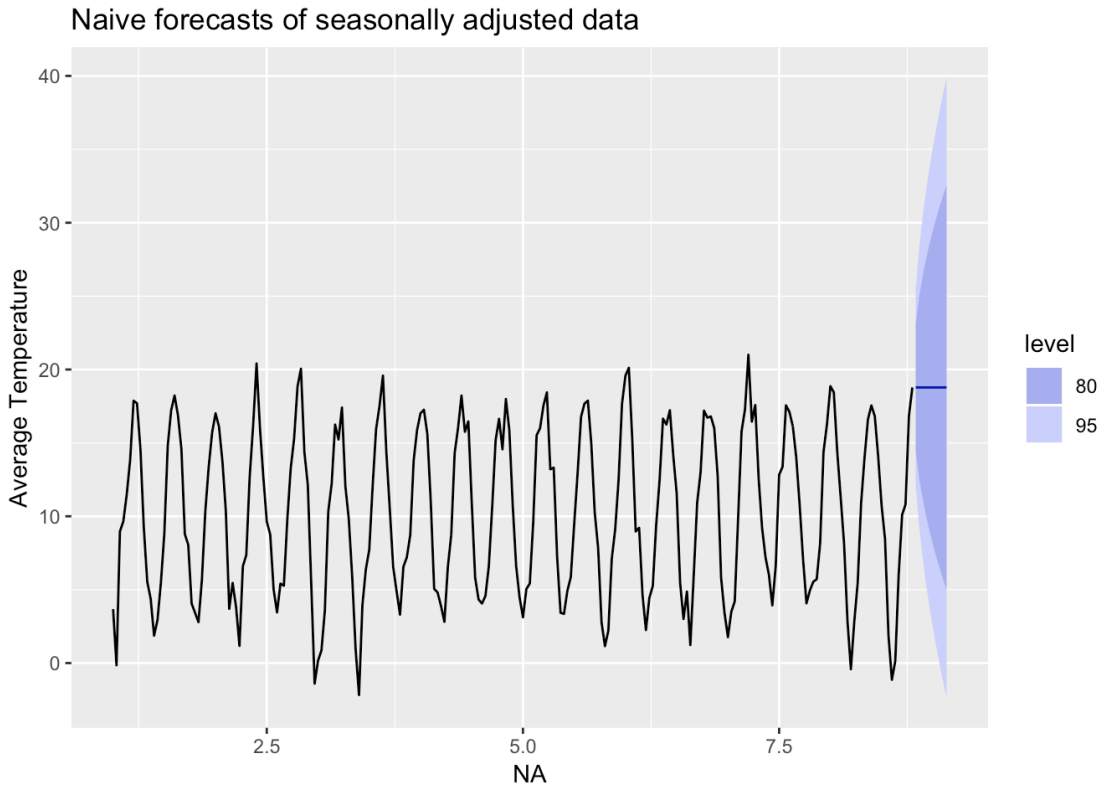



```
ma_12 = forecast(ma(ma(temp_trn,2),12), h = 36)
## Warning in ets(object, lambda = lambda, biasadj =
biasadj,
## allow.multiplicative.trend =
allow.multiplicative.trend, : Missing values
## encountered. Using longest contiguous portion of
time series
accuracy(ma_12, temp_tst)
##                          ME       RMSE        MAE
MPE       MAPE
## Training set -0.0003102815 0.07091146 0.05636221
0.01027199  0.5539328
## Test set      1.9454307585 6.04797133 4.85936384
21.60952494 90.4490964
##                     MASE      ACF1 Theil's U
## Training set 0.07818736 0.4524245        NA
## Test set     6.74105578 0.7326221 0.4754546
#STL decomposition
delat_ma = ts(na.omit(temp_trn), frequency=30)
```

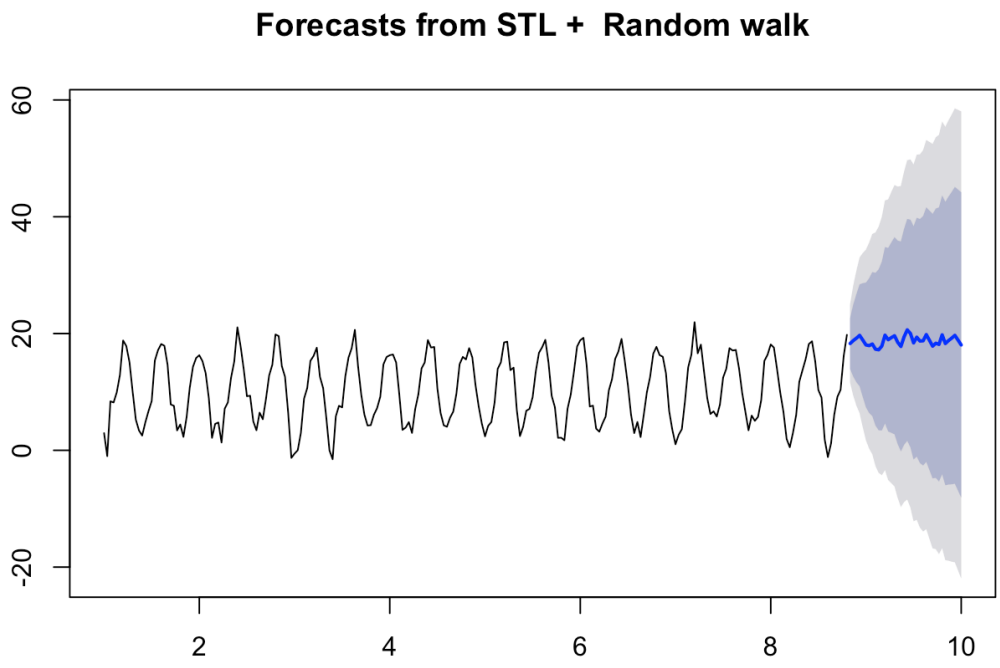
```
decomlat = stl(delat_ma, s.window="periodic")
plot(decomlat)
```



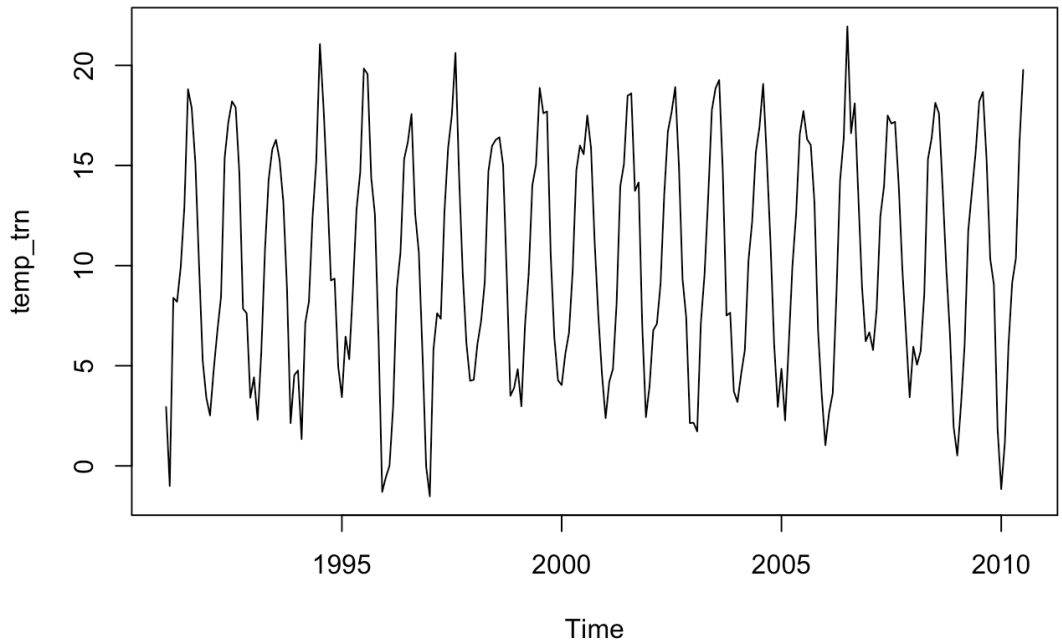

```
#Naive Forecast of seasonally adjusted data
deseasonal_cnt <- seasadj(decomlat)
decomlat %>% seasadj() %>% naive() %>%
  autoplot() + ylab("Average Temperature") +
ggtitle("Naive forecasts of seasonally adjusted data")
```




```
#Forecast of time-decomposition data
fcast <- forecast(decomlat, method="naive", h=36)

plot(fcast)
```

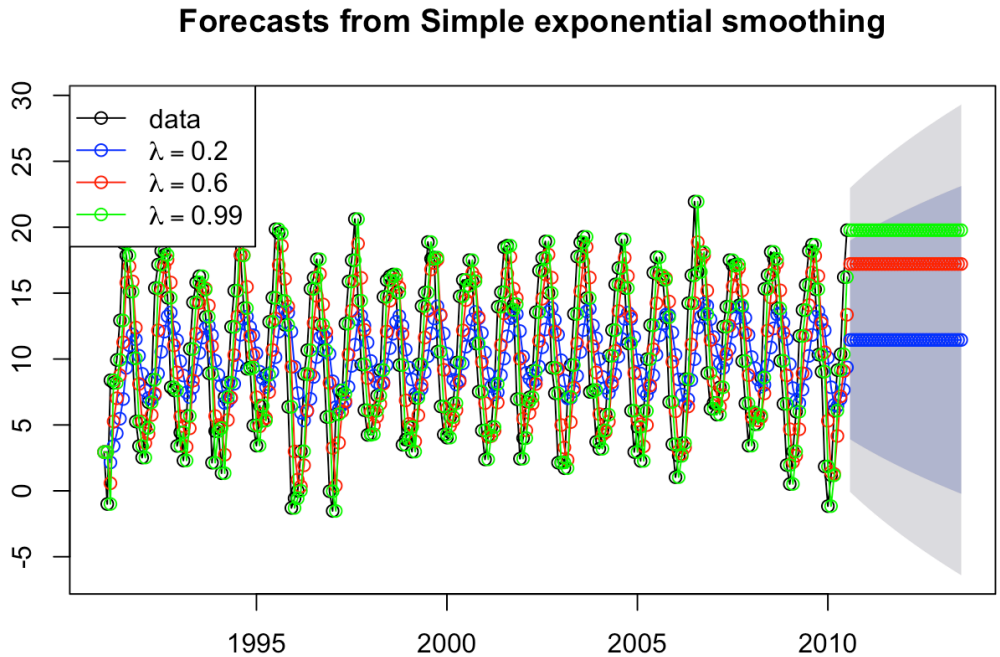



```
#SES: Simple Exponential Smoothing
plot(temp_trn)
```




```
fit1 <-ses(temp_trn, alpha=0.2, initial="simple", h=36)
fit2 <-ses(temp_trn, alpha=0.6, initial="simple", h=36)
fit3 <-ses(temp_trn, h=36)
plot(fit1, plot.conf=FALSE, fcol="white", type="o")
## Warning in plot.window(xlim, ylim, log, ...):
"plot.conf" is not a
## graphical parameter
## Warning in title(main = main, xlab = xlab, ylab =
ylab, ...): "plot.conf"
## is not a graphical parameter
## Warning in axis(1, ...): "plot.conf" is not a
graphical parameter
## Warning in axis(2, ...): "plot.conf" is not a
graphical parameter
## Warning in box(...): "plot.conf" is not a graphical
parameter
lines(fitted(fit1), col="blue", type="o")
lines(fitted(fit2), col="red", type="o")
lines(fitted(fit3), col="green", type="o")
```

```
lines(fit1$mean, col="blue", type="o")
lines(fit2$mean, col="red", type="o")
lines(fit3$mean, col="green", type="o")
legend("topleft",lty=1, col=c(1,"blue","red","green"),
       c("data", expression(lambda == 0.2),
expression(lambda == 0.6),
         expression(lambda == 0.99)),pch=1)
```



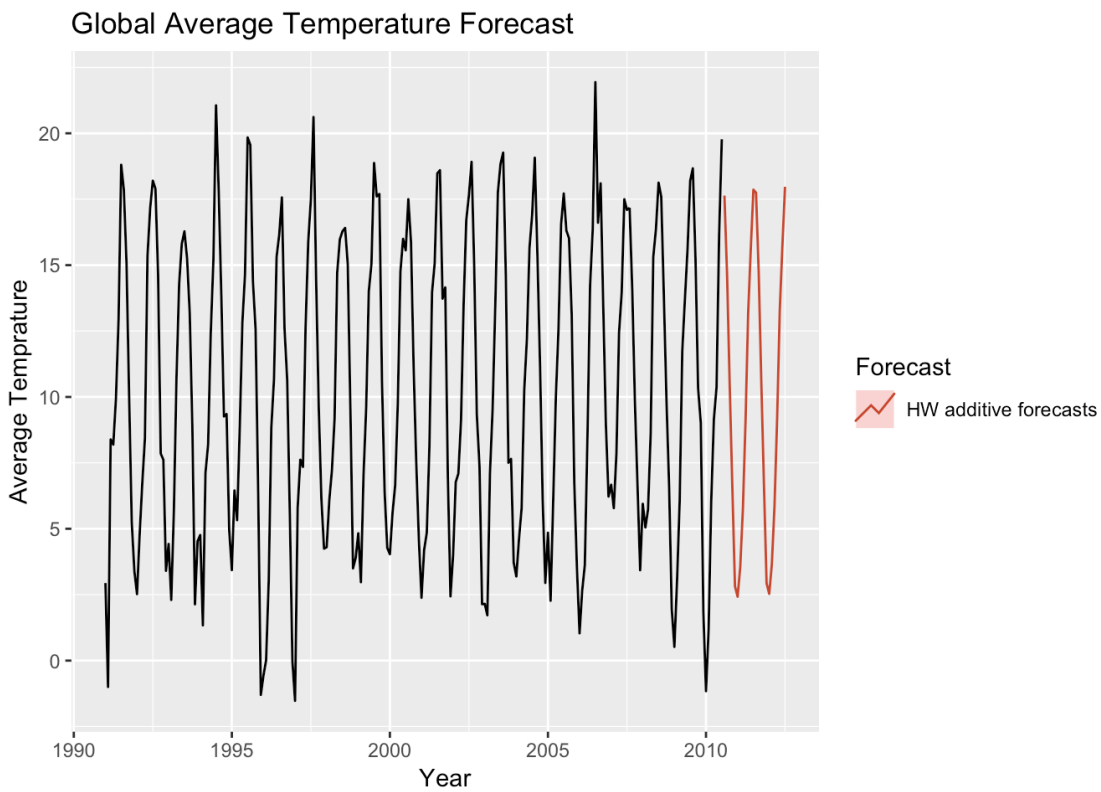

```
fit3$model
## Simple exponential smoothing
##
## Call:
##  ses(y = temp_trn, h = 36)
##
##   Smoothing parameters:
##     alpha = 0.9999
##
##   Initial states:
##     l = 2.938
##
```

```
##   sigma:  3.5119
##
##      AIC     AICc      BIC
## 1877.385 1877.489 1887.764
```

```
x = forecast(fit3, h=36)
accuracy(x, temp_tst)
```

```
##                         ME      RMSE       MAE
MPE     MAPE     MASE
## Training set   0.07165254  3.496901  2.896317
110.2280 162.6176 1.674341
## Test set     -10.04903262 11.503095 10.049033
-136.8725 252.3972 5.809277
##                   ACF1 Theil's U
## Training set 0.4295812        NA
## Test set     0.7532114  2.032675
```

```
#Holt Winter's Seasonal Method
fit1 <- hw(temp_trn,seasonal="additive")

autoplot(temp_trn) +
  autolayer(fit1, series="HW additive forecasts",
PI=FALSE) +
  xlab("Year") +
  ylab("Average Temprature") +
  ggtitle("Global Average Temperature Forecast") +
  guides(colour=guide_legend(title="Forecast"))
```




```
forecast(fit1)
##          Point Forecast       Lo 80     Hi 80
Lo 95     Hi 95
## Aug 2010      17.637801 15.49434337 19.781259
14.35966526 20.915937
## Sep 2010      14.636118 12.46647794 16.805758
11.31793980 17.954297
## Oct 2010      10.436025  8.24047901 12.631572
7.07822698 13.793824
## Nov 2010       6.559996  4.33880996  8.781182
3.16298508  9.957007
## Dec 2010       2.820479  0.57391086  5.067048
-0.61535069  6.256309
## Jan 2011       2.422843  0.15114030  4.694545
-1.05142630  5.897112
## Feb 2011       3.548176  1.25158024  5.844772
0.03583588  7.060516
## Mar 2011       5.732392  3.41113531  8.053648
2.18233635  9.282447
```

```
## Apr 2011       9.036272  6.69058033 11.381964
5.44884605 12.623698
## May 2011      13.160529 10.79061957 15.530438
9.53606554 16.784992
## Jun 2011      15.615626 13.22171189 18.009541
11.95445016 19.276802
## Jul 2011      17.862272 15.44455759 20.279987
14.16469686 21.559847
## Aug 2011      17.746505 15.30514499 20.187865
14.01276713 21.480243
## Sep 2011      14.744822 12.28005505 17.209588
10.97528640 18.514357
## Oct 2011      10.544729  8.05674357 13.032714
6.73968369 14.349774
## Nov 2011       6.668700  4.15767838  9.179721
2.82842405 10.508975
## Dec 2011       2.929183  0.39530370  5.463062
-0.94605095  6.804417
## Jan 2012       2.531546 -0.02501804  5.088110
-1.37838140  6.441474
## Feb 2012       3.656880  1.07779874  6.235960
-0.28748417  7.601243
## Mar 2012       5.841095  3.23966202  8.442529
1.86254640  9.819644
## Apr 2012       9.144976  6.52134977 11.768602
5.13248605 13.157465
## May 2012      13.269232 10.62356921 15.914895
9.22303988 17.315424
## Jun 2012      15.724330 13.05678197 18.391878
11.64466742 19.803992
## Jul 2012      17.970976 15.28169093 20.660260
13.85806963 22.083882
```

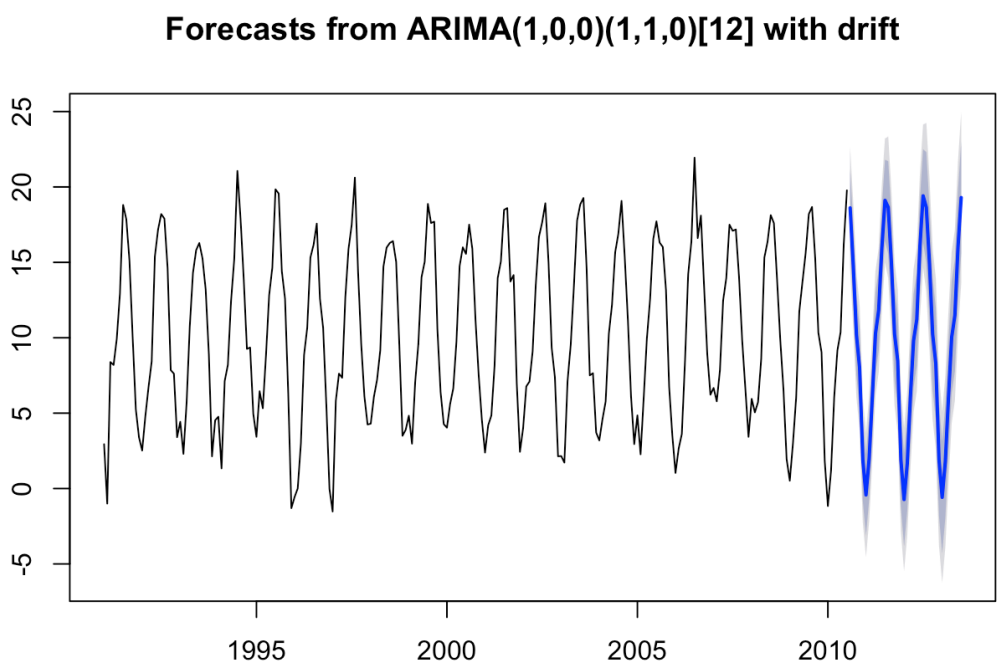
```
accuracy(fit1, temp_tst)
##                       ME      RMSE      MAE
MPE     MAPE      MASE
## Training set -0.05445446  1.614608 1.284018
```

```
-155.78098 241.5204 0.7422821
## Test set     -0.15581259 10.770407 9.629018
-14.65277 206.2453 5.5664690
##                    ACF1 Theil's U
## Training set 0.09590058        NA
## Test set     0.76357509  0.985584
```

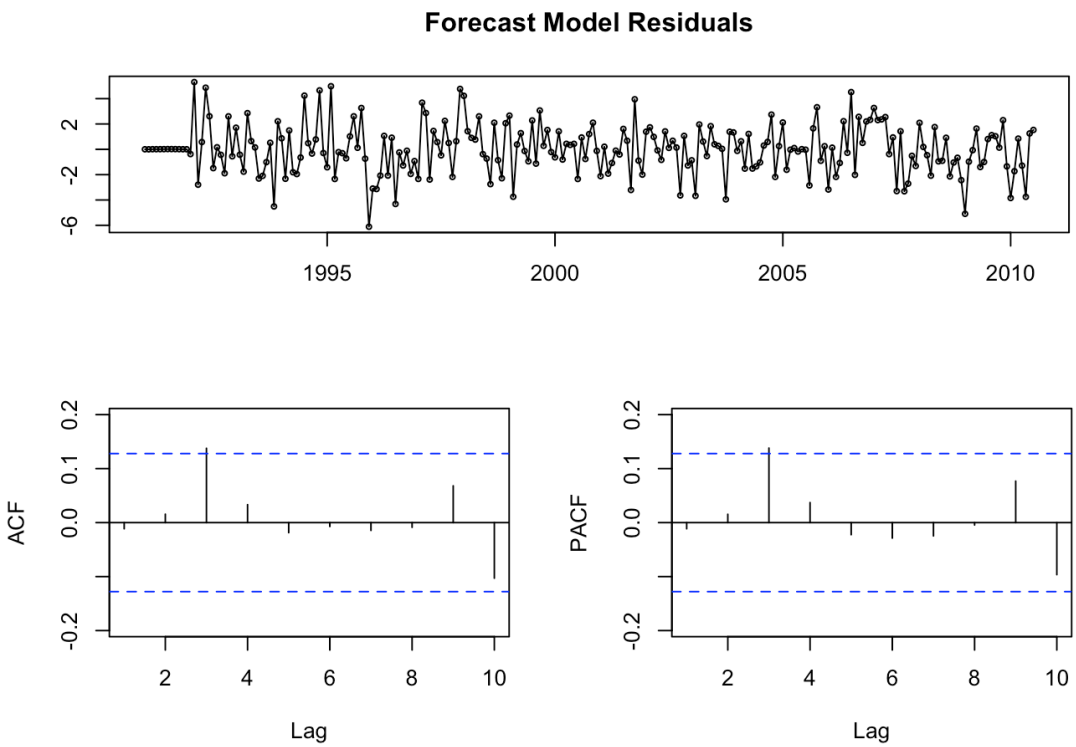
```
#ARIMA
fit <- auto.arima(temp_trn)
fit
```

```
## Series: temp_trn
## ARIMA(1,0,0)(1,1,0)[12] with drift
##
## Coefficients:
##          ar1     sar1   drift
##       0.2448  -0.4301  0.0008
## s.e.  0.0649   0.0619  0.0106
##
## sigma^2 estimated as 4.111:  log likelihood=-473.8
## AIC=955.6   AICc=955.78   BIC=969.23
```

```
#Forescasting the values
fitarima = forecast(fit, h = 36)
plot(fitarima)
```




```
#Accuracy of the ARIMA
accuracy(fitarima, temp_tst)
##                        ME      RMSE      MAE
MPE     MAPE      MASE
## Training set  0.01344832  1.961819 1.499928
-179.88228 236.7830 0.8670979
## Test set     -0.11542019 11.472964 9.985587
-55.69274 206.8493 5.7725989
##                     ACF1 Theil's U
## Training set -0.0114428        NA
## Test set      0.7928919  1.355744
tsdisplay(residuals(fit), lag.max=10, main='Forecast
Model Residuals')
```


Forecast Model Residuals
2
-2
-6
1995
2000
2005
2010
ACF
0.2
0.1
0.0
-0.2
2
4
6
8
10
Lag
PACF
0.2
0.1
0.0
-0.2
2
4
6
8
10
Lag