Project 4: Economic Data Analysis

DA 420

Marjorie Blanco

Part 1:

Use R to develop a Multiple Time Series Plot, as seen in Figure 5.1. Make sure you include the command lines and the output/results. Interpret the graphin details.

Hint:Look at the Exhibit 5.1 in page 62-63.

Data

```
# Economic Data from Federal Reserve Bank of St. Louis (FRED system)
# National Civilian Unemployment Rate (monthly, percentage)
getSymbols("UNRATENSA", src="FRED", return.class = "xts")
```

```
## [1] "UNRATENSA"
```

```
# Manufacturers' New Orders: Durable Goods (millions of dollars)
getSymbols("DGORDER", src="FRED", return.class = "xts")
```

```
## [1] "DGORDER"
```

```
# University of Michigan Index of Consumer Sentiment (1Q 1966 = 100)
getSymbols("UMCSENT", src="FRED", return.class = "xts")
```

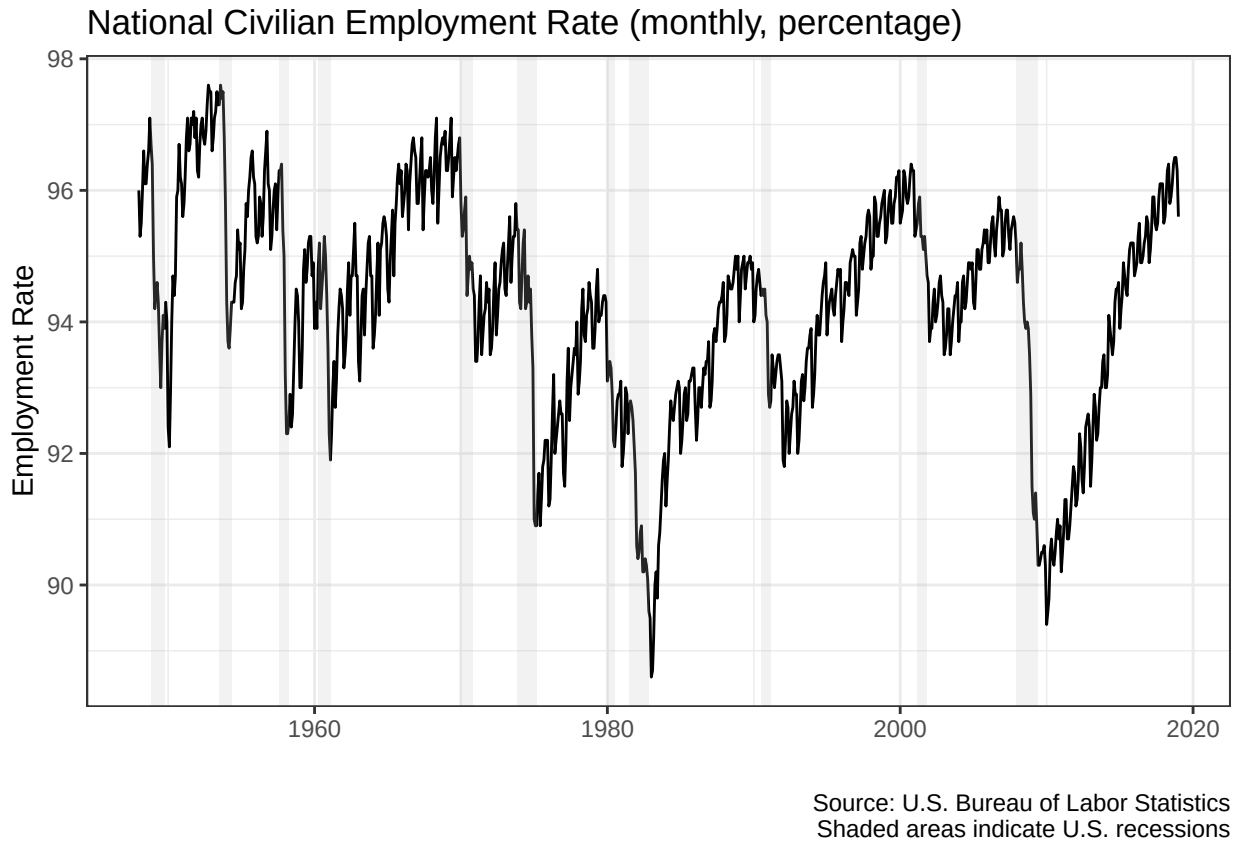
```
## [1] "UMCSENT"
```

```
# New Homes Sold in the US, not seasonally adjusted (monthly, millions)
getSymbols("HSN1FNSA",src="FRED",return.class = "xts")
```

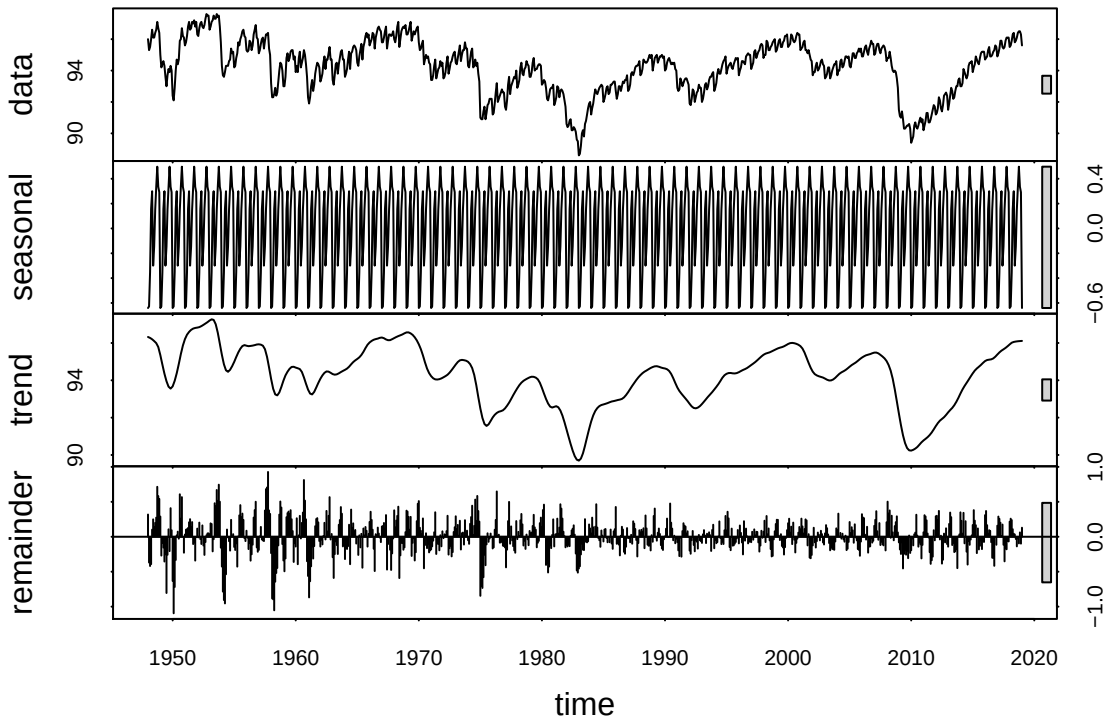
```
## [1] "HSN1FNSA"
```

```
ER <- 100 - UNRATENSA # convert to employment rate
dimnames(ER)[2] <- "ER"
ER.data.frame <- as.data.frame(ER)
ER.data.frame$date <- ymd(rownames(ER.data.frame))

ER.time.series <- ts(ER.data.frame$ER,
  start = c(year(min(ER.data.frame$date)),month(min(ER.data.frame$date))),
  end = c(year(max(ER.data.frame$date)),month(max(ER.data.frame$date))),
  frequency=12)

recessions.trim <- subset(recessions.df, Peak >= min(ER.data.frame$date) )
```

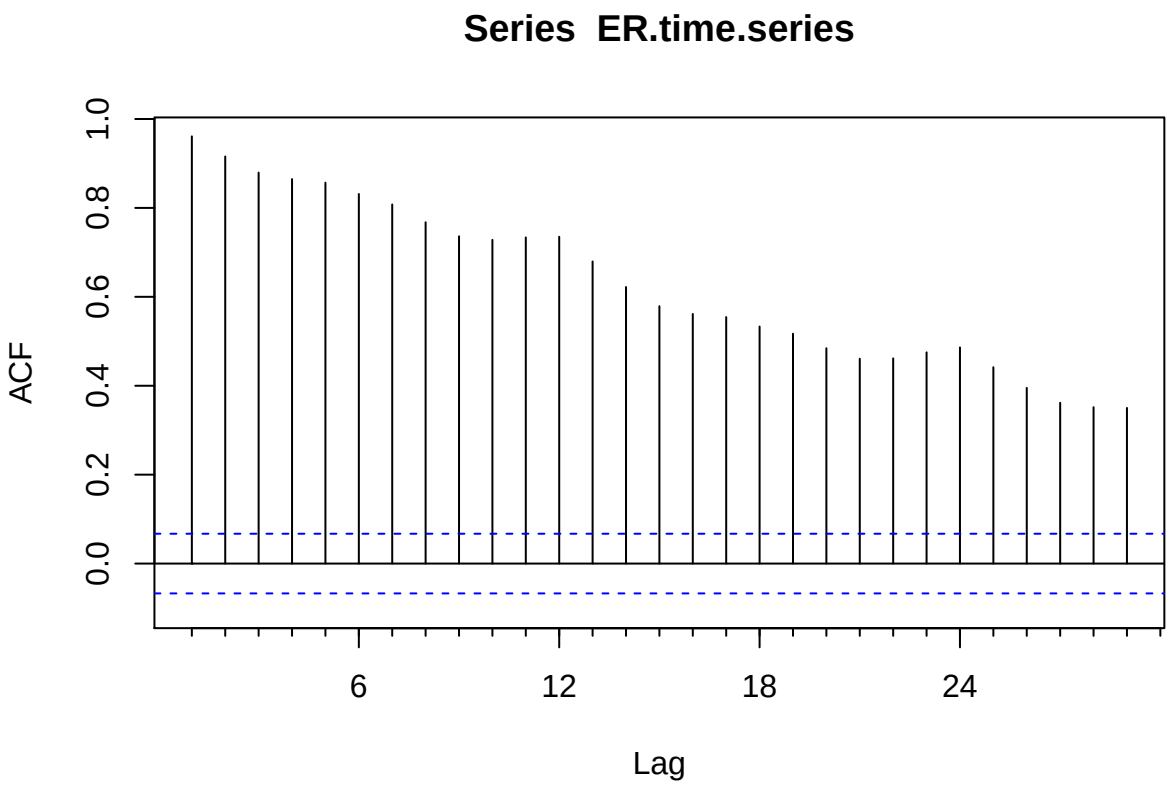
```
ggplot(ER.data.frame) +
  geom_line(aes(x=date, y=ER)) +
  theme_bw() +
  geom_rect(data=recessions.trim, aes(xmin=Peak, xmax=Trough, ymin=-Inf, ymax=+Inf),
            fill='gray', alpha=0.2) +
  labs(title = "National Civilian Employment Rate (monthly, percentage)",
       caption = "Source: U.S. Bureau of Labor Statistics\nShaded areas indicate U.S. recessions",
       x = "", y = "Employment Rate")
```



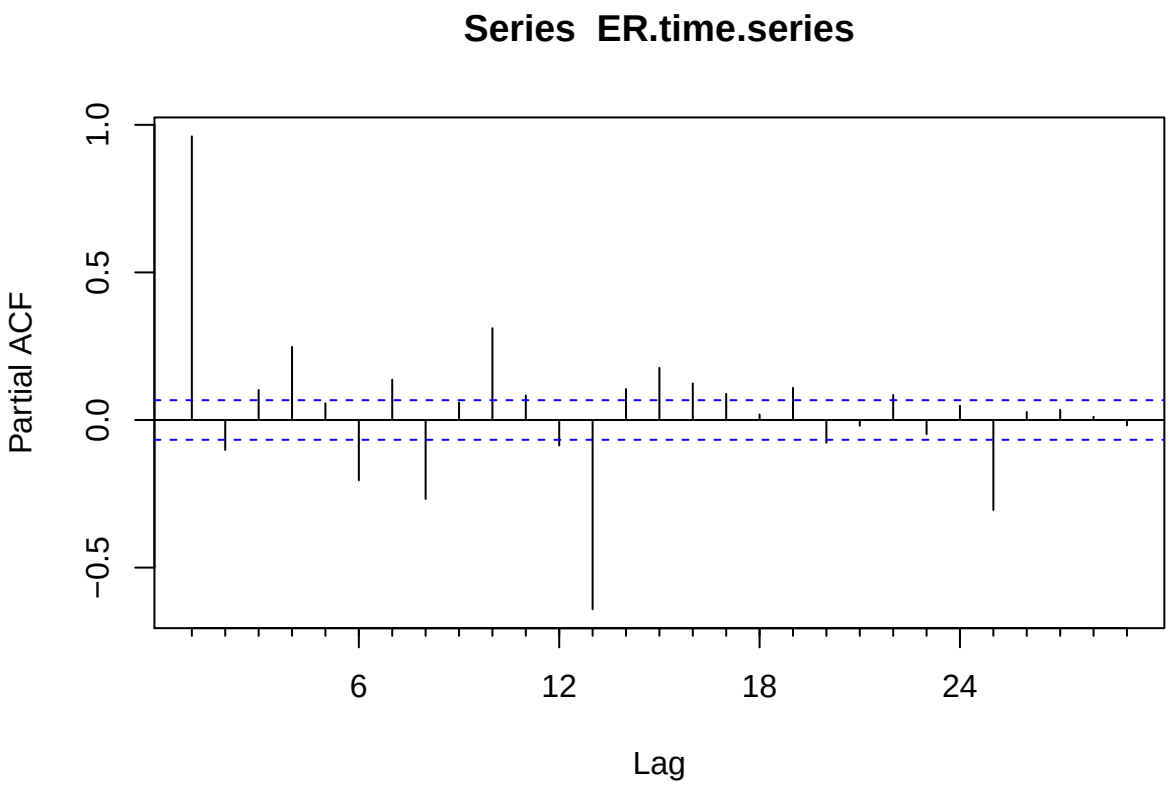

```
decomp = stl(ER.time.series, s.window="periodic")
plot(decomp)
```




```
Acf(ER.time.series)
```

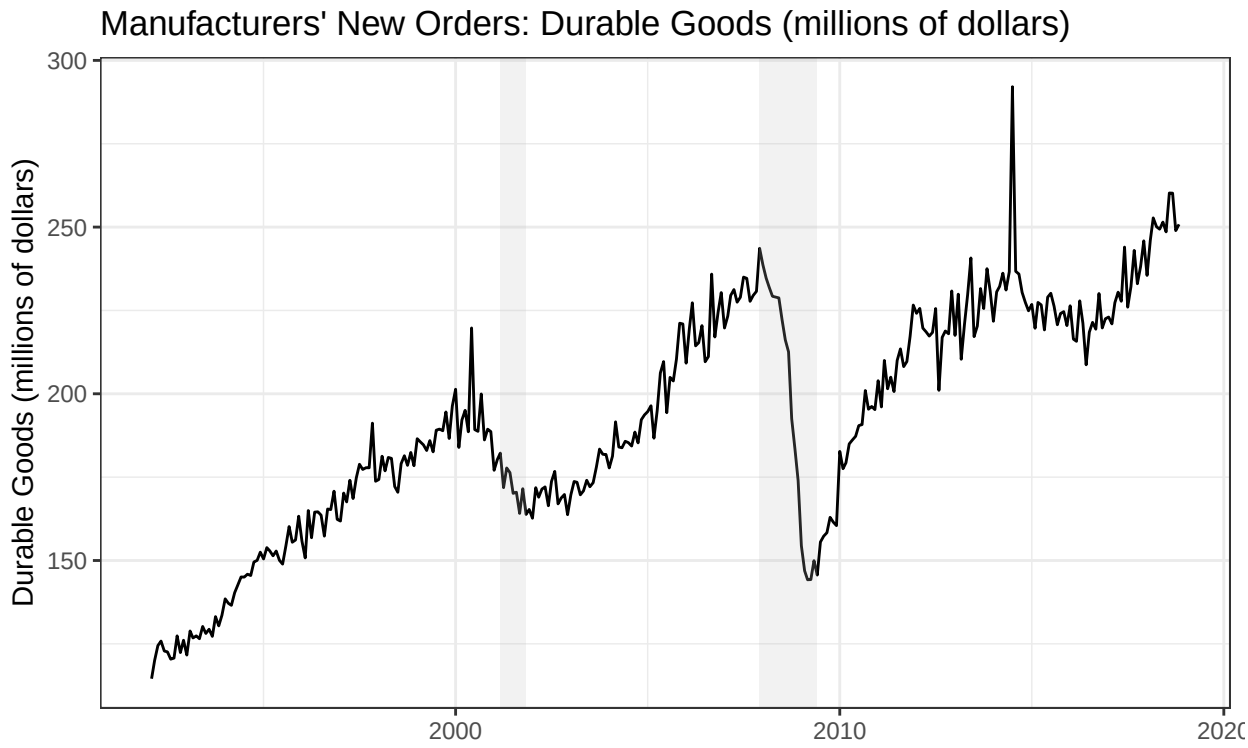



```
Pacf(ER.time.series)
```

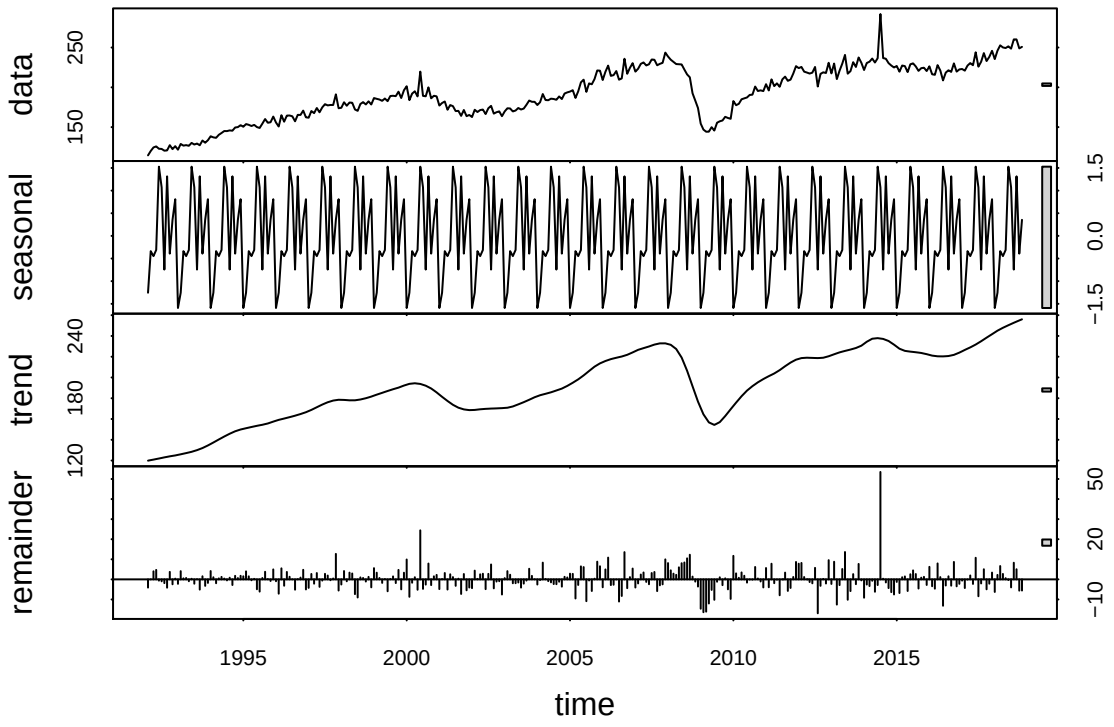


ACF plot shows a spike at lags 1 through 30. PACF plot shows a spike at lags 1 and 13. This chart shows that the employement rate are cyclical related to the recessions. The employement rate shows a upward trend after recession years and downward trend during recession years. The trend and cyclical nature of employment rate is expected.

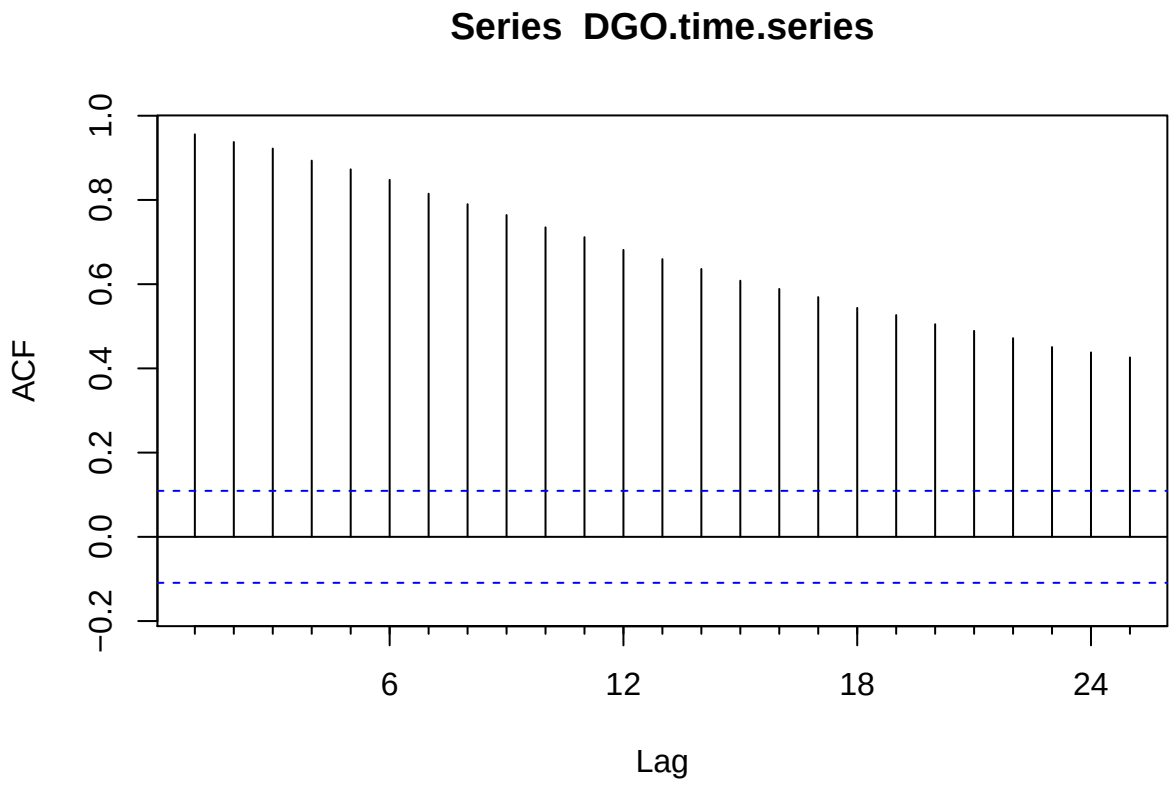
```
DGO <- DGORDER/1000 # convert to billions of dollars
dimnames(DGO)[2] <- "DGO" # use simple name for index
DGO.data.frame <- as.data.frame(DGO)
DGO.data.frame$DGO <- DGO.data.frame$DGO
DGO.data.frame$date <- ymd(rownames(DGO.data.frame))
DGO.time.series <- ts(DGO.data.frame$DGO,
  start = c(year(min(DGO.data.frame$date)),month(min(DGO.data.frame$date))),
  end = c(year(max(DGO.data.frame$date)),month(max(DGO.data.frame$date))),
  frequency=12)

recessions.trim <- subset(recessions.df, Peak >= min(DGO.data.frame$date) )

ggplot(DGO.data.frame) +
  geom_line(aes(x=date, y=DGO)) +
  theme_bw() +
  geom_rect(data=recessions.trim, aes(xmin=Peak, xmax=Trough, ymin=-Inf, ymax=+Inf),
            fill='gray', alpha=0.2) +
  labs(title = "Manufacturers' New Orders: Durable Goods (millions of dollars)",
       caption = "Source: U.S. Bureau of Labor Statistics\nShaded areas indicate U.S. recessions",
       x = "", y = "Durable Goods (millions of dollars)")
```



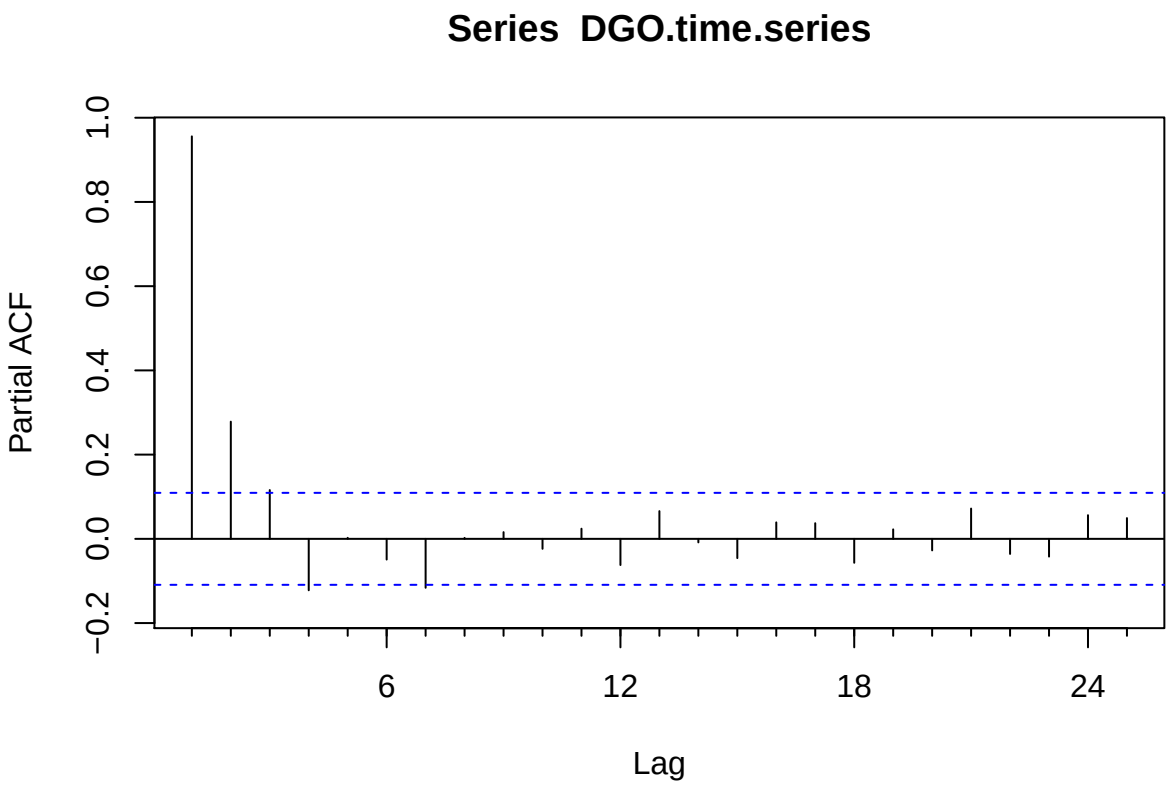

```
decomp = stl(DGO.time.series, s.window="periodic")
plot(decomp)
```




```
Acf(DGO.time.series)
```

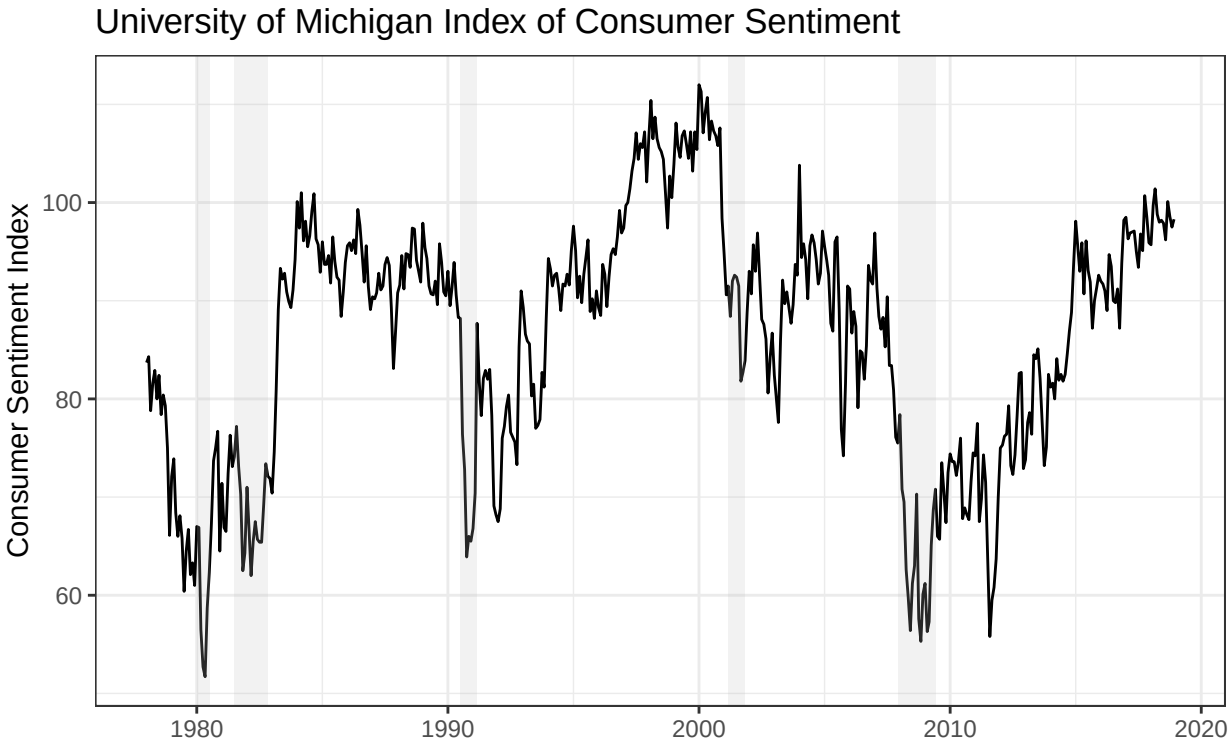



```
Pacf(DGO.time.series)
```



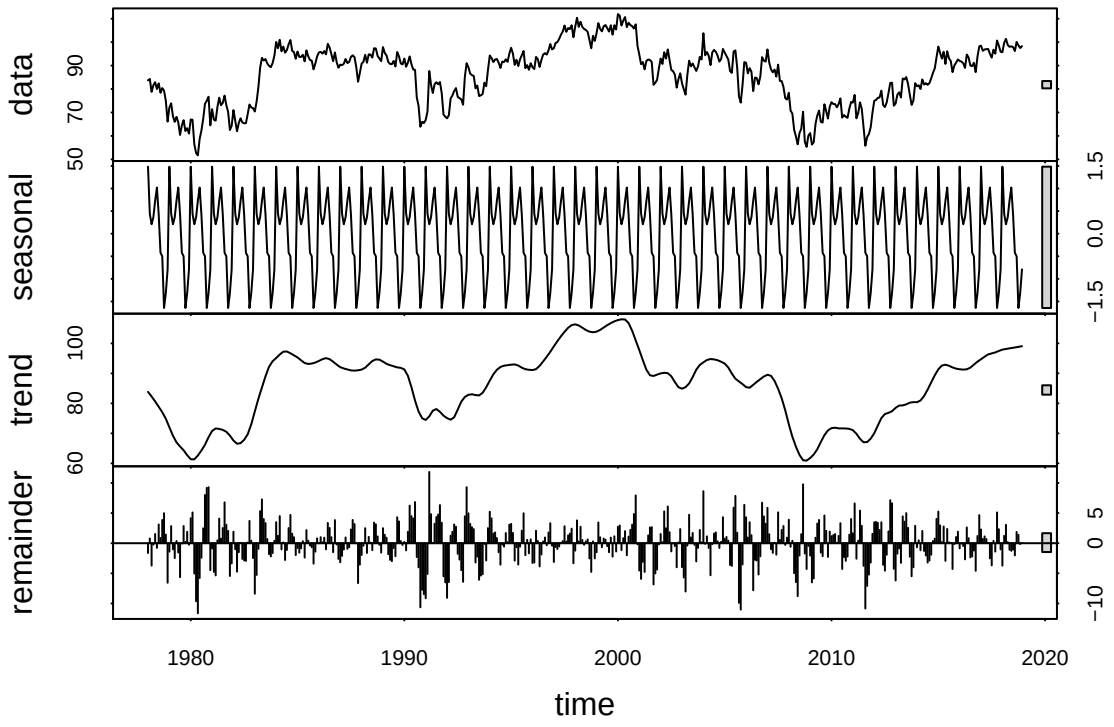
ACF plot shows a spike at lags 1 through 30. PACF plot shows a spike at lags 1 and 2. This chart shows that the manufacturers' new orders are cyclical related to the recessions. The durable goods shows a upward trend after recession years and downward trend during recession years. The trend and cyclical nature is expected.

```
ICS <- UMCSENT # use simple name for xts object
dimnames(ICS)[2] <- "ICS" # use simple name for index
ICS.data.frame <- as.data.frame(ICS)
ICS.data.frame$ICS <- ICS.data.frame$ICS
ICS.data.frame$date <- ymd(rownames(ICS.data.frame))
ICS.data.frame <-  ICS.data.frame[303:nrow(ICS.data.frame),]

ICS.time.series <- ts(ICS.data.frame$ICS,
  start = c(year(min(ICS.data.frame$date)), month(min(ICS.data.frame$date))),
  end = c(year(max(ICS.data.frame$date)),month(max(ICS.data.frame$date))),
  frequency=12)

recessions.trim <- subset(recessions.df, Peak >= ymd("1978-01-01") )

ggplot(ICS.data.frame) +
  geom_line(aes(x=date, y=ICS)) +
  theme_bw() +
  geom_rect(data=recessions.trim, aes(xmin=Peak, xmax=Trough, ymin=-Inf, ymax=+Inf),
            fill='gray', alpha=0.2) +
  labs(title = "University of Michigan Index of Consumer Sentiment",
       caption = "Source: U.S. Bureau of Labor Statistics\nShaded areas indicate U.S. recessions",
       x = "", y = "Consumer Sentiment Index")
```

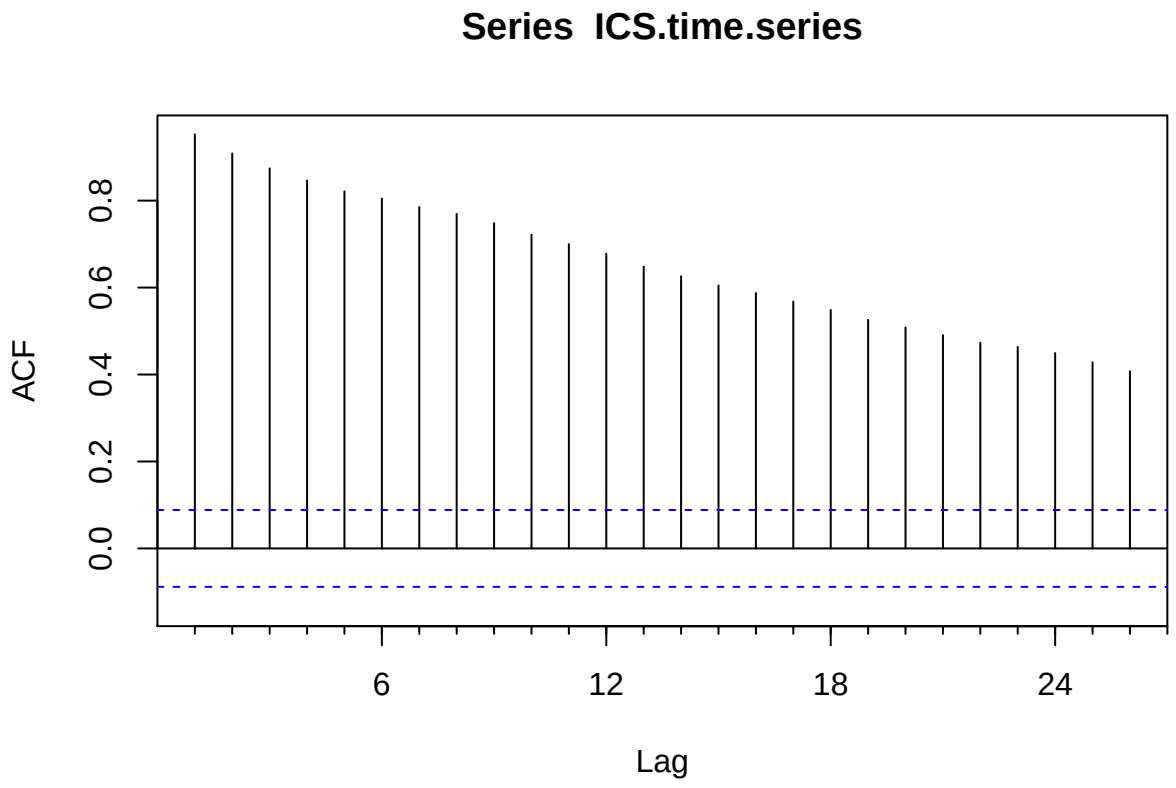


Source: U.S. Bureau of Labor Statistics
Shaded areas indicate U.S. recessions

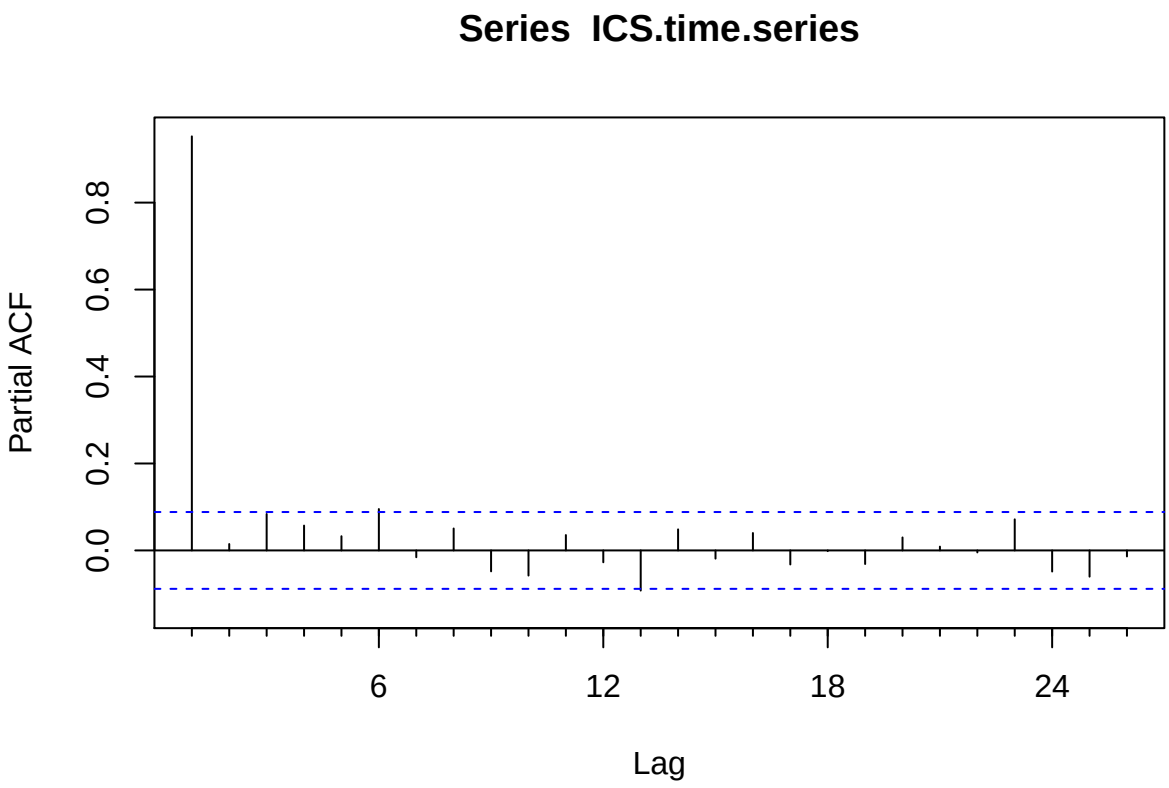
```
decomp = stl(ICS.time.series, s.window="periodic")
plot(decomp)
```




```
Acf(ICS.time.series)
```

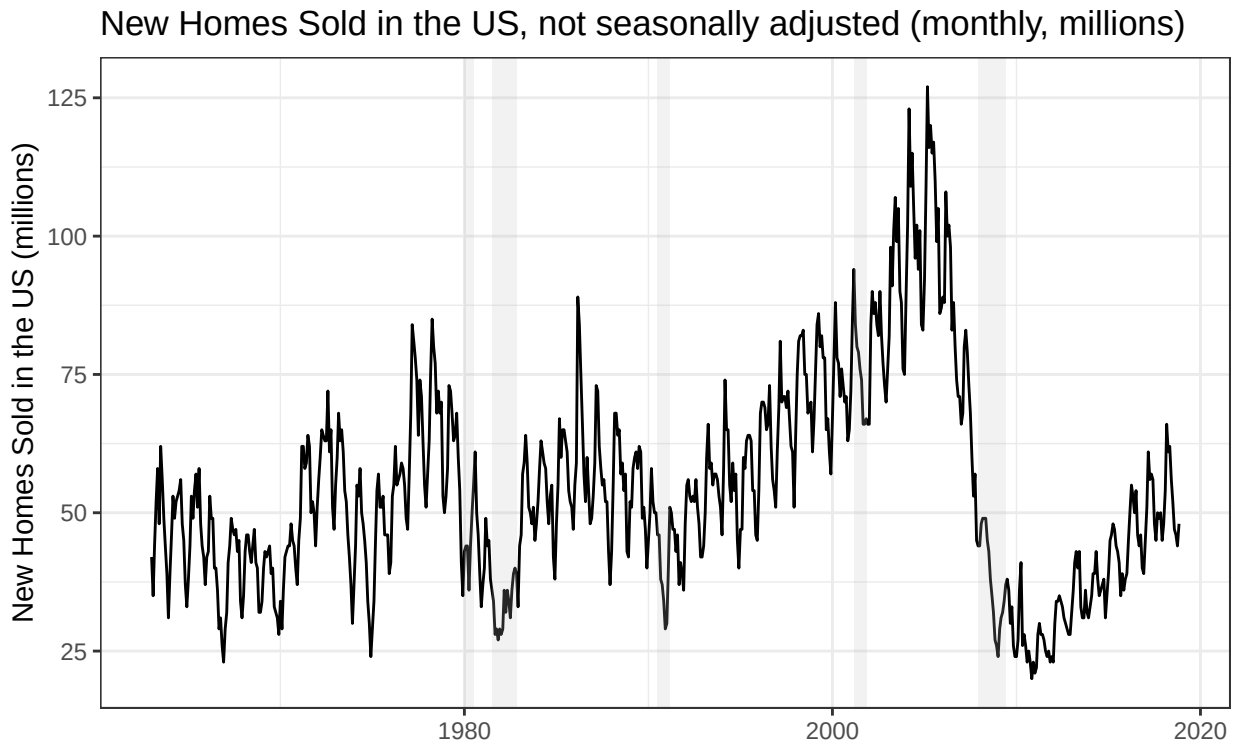



```
Pacf(ICS.time.series)
```



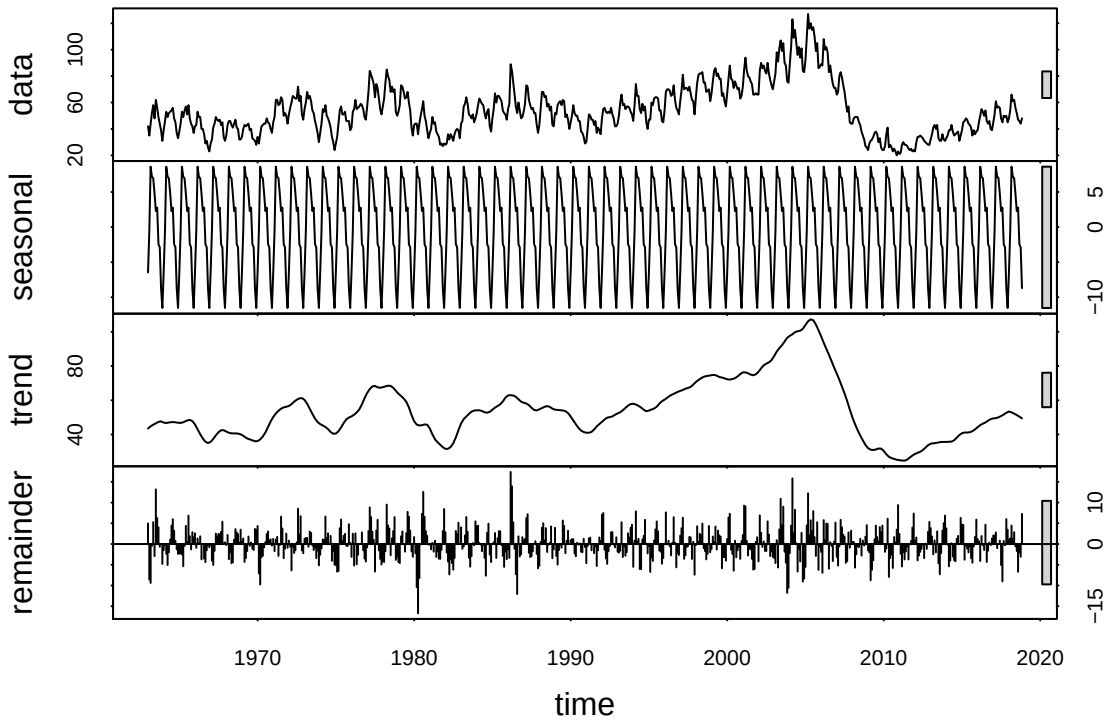
ACF plot shows a spike at lags 1 through 30. PACF plot shows a spike at lags 1. This chart shows that consumer sentiment index is related to the recessions. The durable goods shows a upward trend after recession years and downward trend during recession years. The trend and cyclical nature is expected.

```
NHS <- HSN1FNSA
dimnames(NHS)[2] <- "NHS" # use simple name for index
NHS.data.frame <- as.data.frame(NHS)
NHS.data.frame$NHS <- NHS.data.frame$NHS
NHS.data.frame$date <- ymd(rownames(NHS.data.frame))
NHS.time.series <- ts(NHS.data.frame$NHS,
  start = c(year(min(NHS.data.frame$date)),month(min(NHS.data.frame$date))),
  end = c(year(max(NHS.data.frame$date)),month(max(NHS.data.frame$date))),
  frequency=12)

recessions.trim <- subset(recessions.df, Peak >= ymd("1978-01-01") )

ggplot(NHS.data.frame) +
  geom_line(aes(x=date, y=NHS)) +
  theme_bw() +
  geom_rect(data=recessions.trim, aes(xmin=Peak, xmax=Trough, ymin=-Inf, ymax=+Inf),
            fill='gray', alpha=0.2) +
  labs(title = "New Homes Sold in the US, not seasonally adjusted (monthly, millions)",
       caption = "Source: U.S. Bureau of Labor Statistics\nShaded areas indicate U.S. recessions",
       x = "", y = "New Homes Sold in the US (millions)")
```

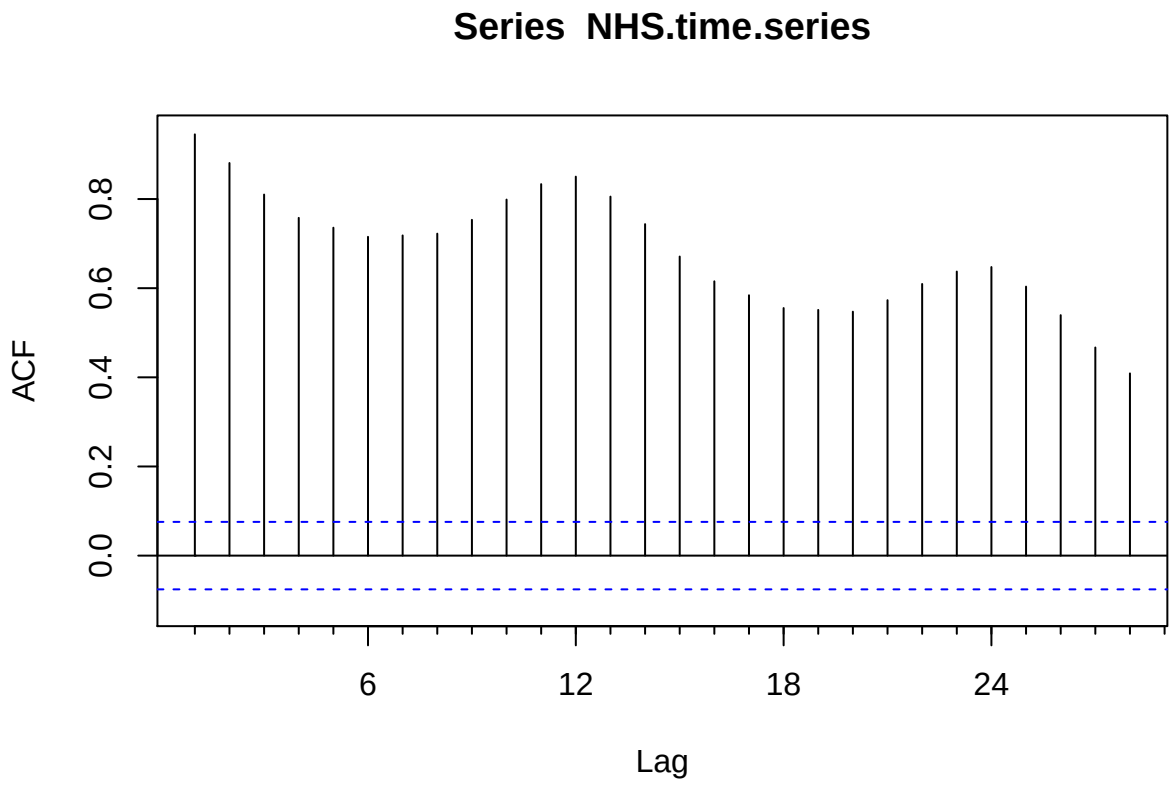


Source: U.S. Bureau of Labor Statistics
Shaded areas indicate U.S. recessions

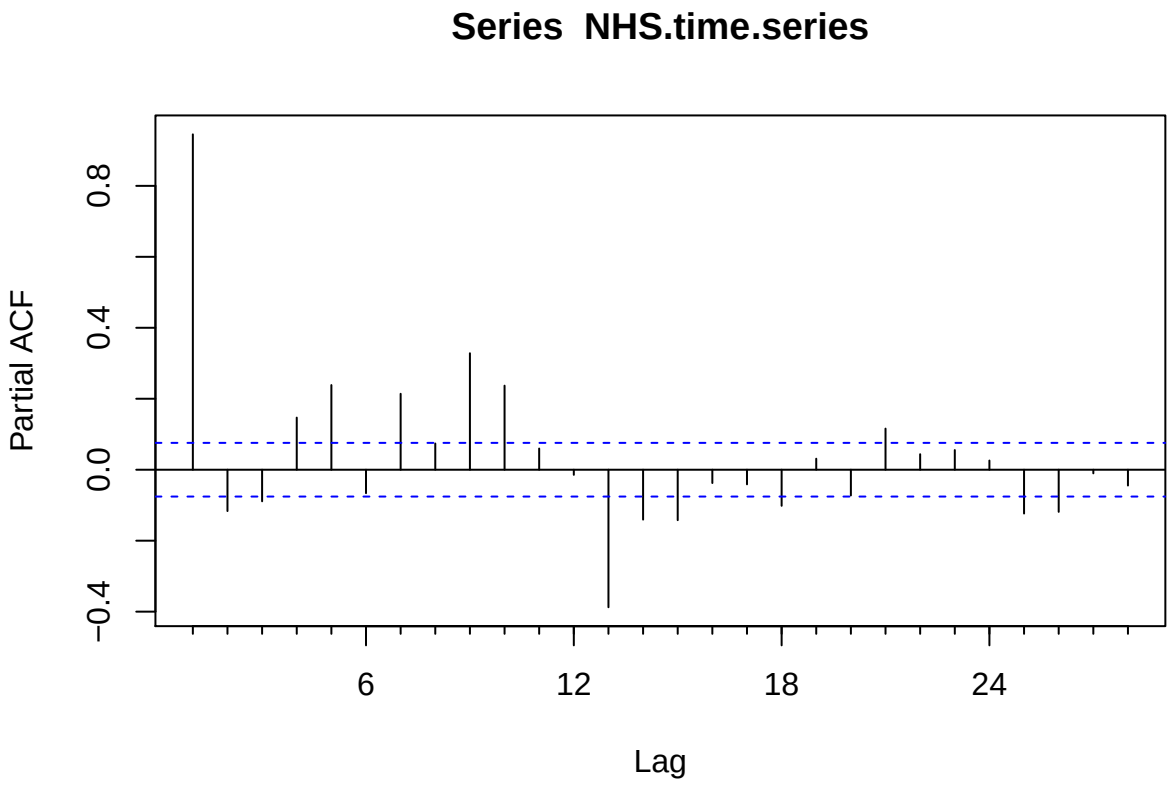
```
decomp = stl(NHS.time.series, s.window="periodic")
plot(decomp)
```




```
Acf(NHS.time.series)
```



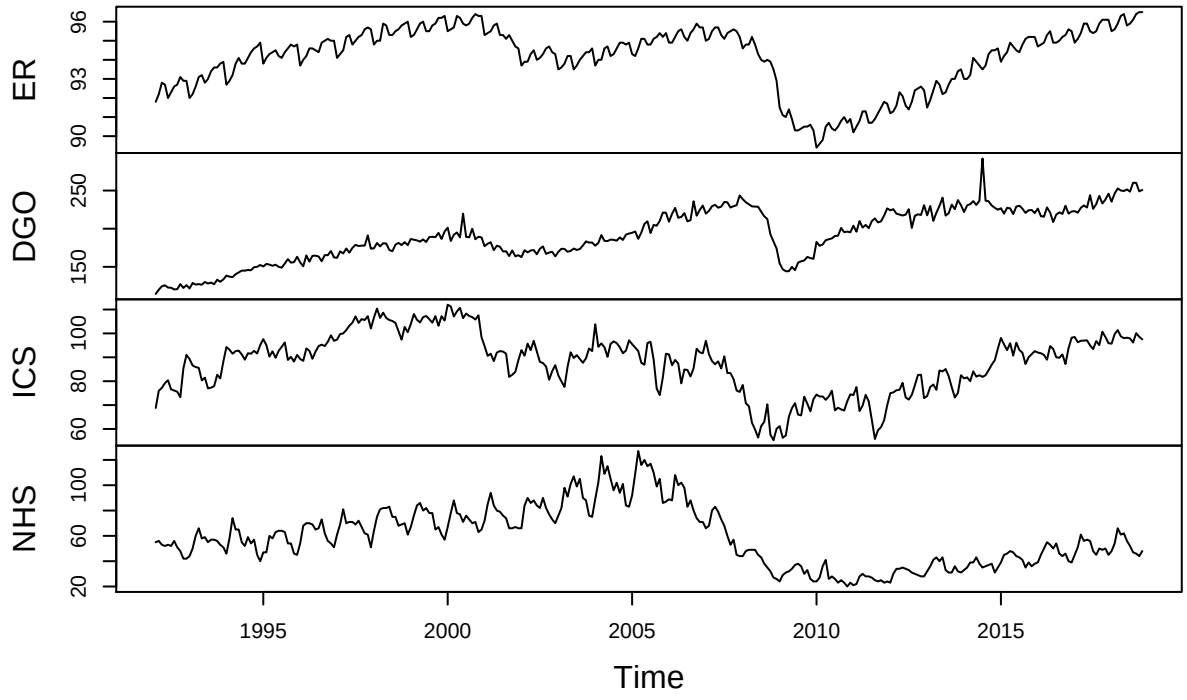

```
Pacf(NHS.time.series)
```




```
min_rate1 <- min(NHS.time.series)
max_rate1 <- max(NHS.time.series)

min_date <- NHS.data.frame %>% filter(NHS == min_rate1)
max_date <- NHS.data.frame %>% filter(NHS == max_rate1)
```

ACF plot shows a spike at lags 1, 12 and 14. PACF plot shows a spike at lags 1 and 13. This chart shows that new homes sold in the US exhibit a seasonal patterns. New homes sold were steadly increasing until 127, 2005-03-01 just before the housing market crash. The lowest home sold in the US was in 20, 2010-11-01.

```
# define multiple time series object
economic.mts <- cbind(ER.time.series, DGO.time.series, ICS.time.series,
  NHS.time.series)
  dimnames(economic.mts)[[2]] <- c("ER","DGO","ICS","NHS") # keep simple names
modeling.mts <- na.omit(economic.mts) # keep overlapping time intervals only

# plot multiple time series
plot(modeling.mts,main="")
```



Data

The source of this data is the U.S. Bureau of Labor Statistics. The rate is released monthly as a percent. The data is seasonally adjusted.

Frequency: Monthly

```
# Economic Data from Federal Reserve Bank of St. Louis (FRED system)
# Civilian Labor Force Participation Rate: Women  (monthly, percentage)
getSymbols("LNS11300002", src="FRED", return.class = "xts")
```

```
## [1] "LNS11300002"
```

```
# Civilian Labor Force Participation Rate: Men  (monthly, percentage)
getSymbols("LNS11300001", src="FRED", return.class = "xts")
```

```
## [1] "LNS11300001"
```

```
# Civilian Labor Force Participation Rate: White  (monthly, percentage)
getSymbols("LNS11300003", src="FRED", return.class = "xts")
```

```
## [1] "LNS11300003"
```

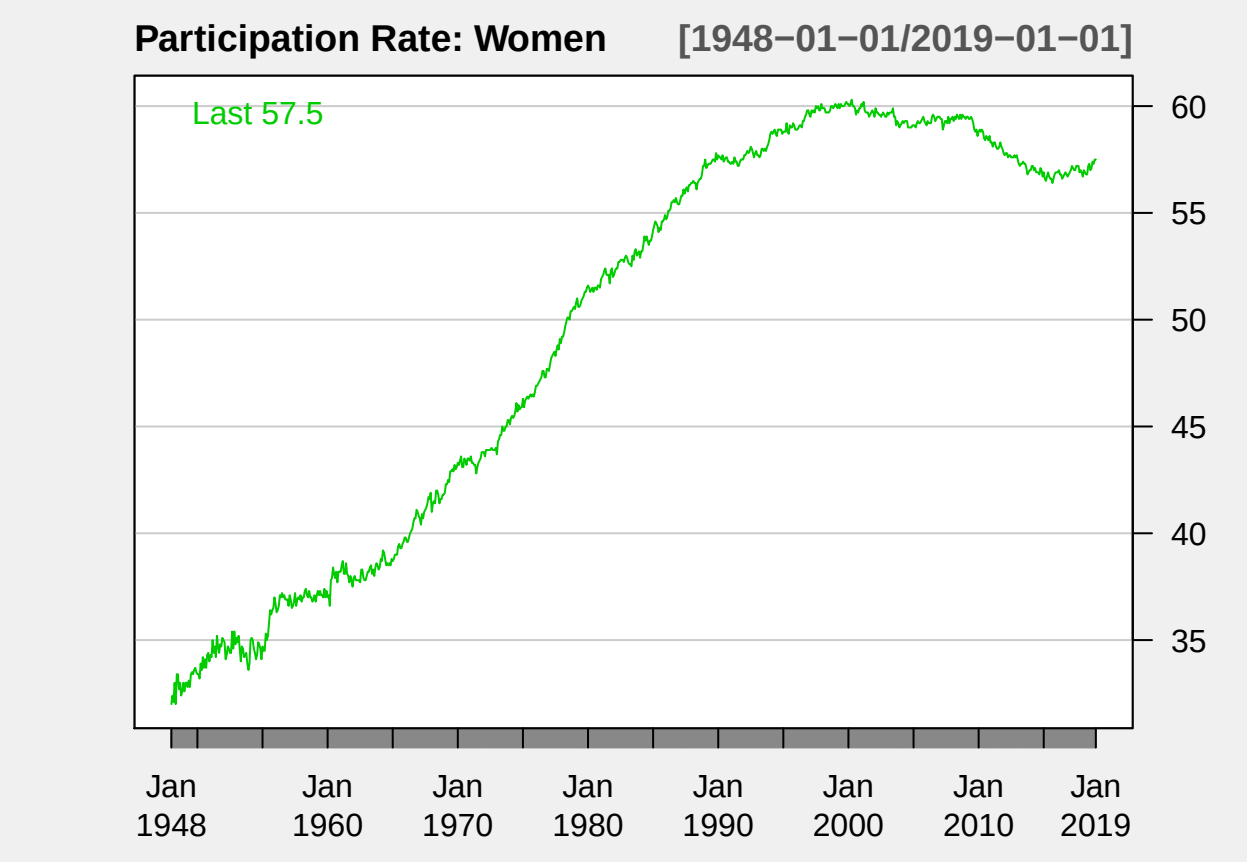
```
# Civilian Labor Force Participation Rate: Black  (monthly, percentage)
getSymbols("LNS11300006", src="FRED", return.class = "xts")
```

```
## [1] "LNS11300006"
```

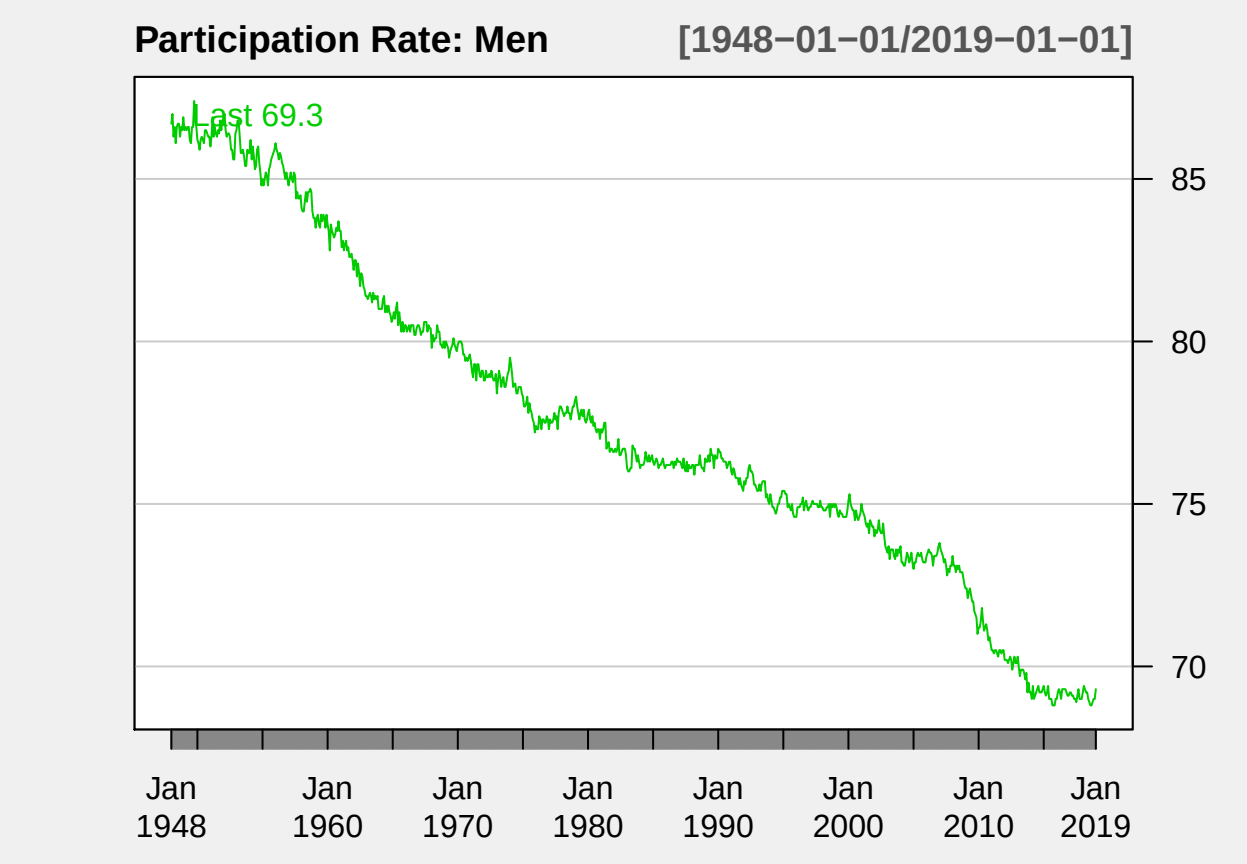
```
# Civilian Labor Force Participation Rate: Hispanic or Latino  (monthly, percentage)
getSymbols("LNS11300009", src="FRED", return.class = "xts")
```

```
## [1] "LNS11300009"
```

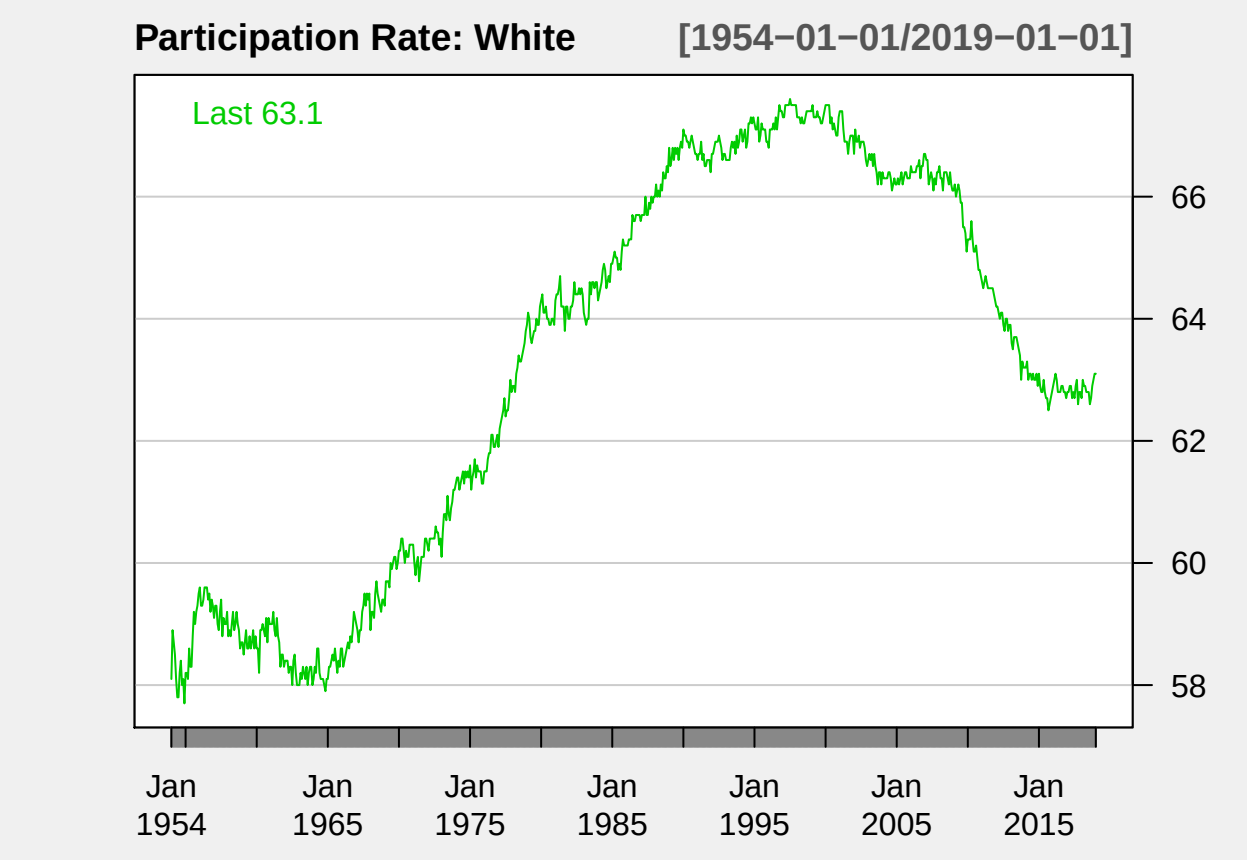
```
PR <- LNS11300002 # convert to participation rate
dimnames(PR)[2] <- "PR"
chartSeries(PR, theme="white", name="Participation Rate: Women")
```



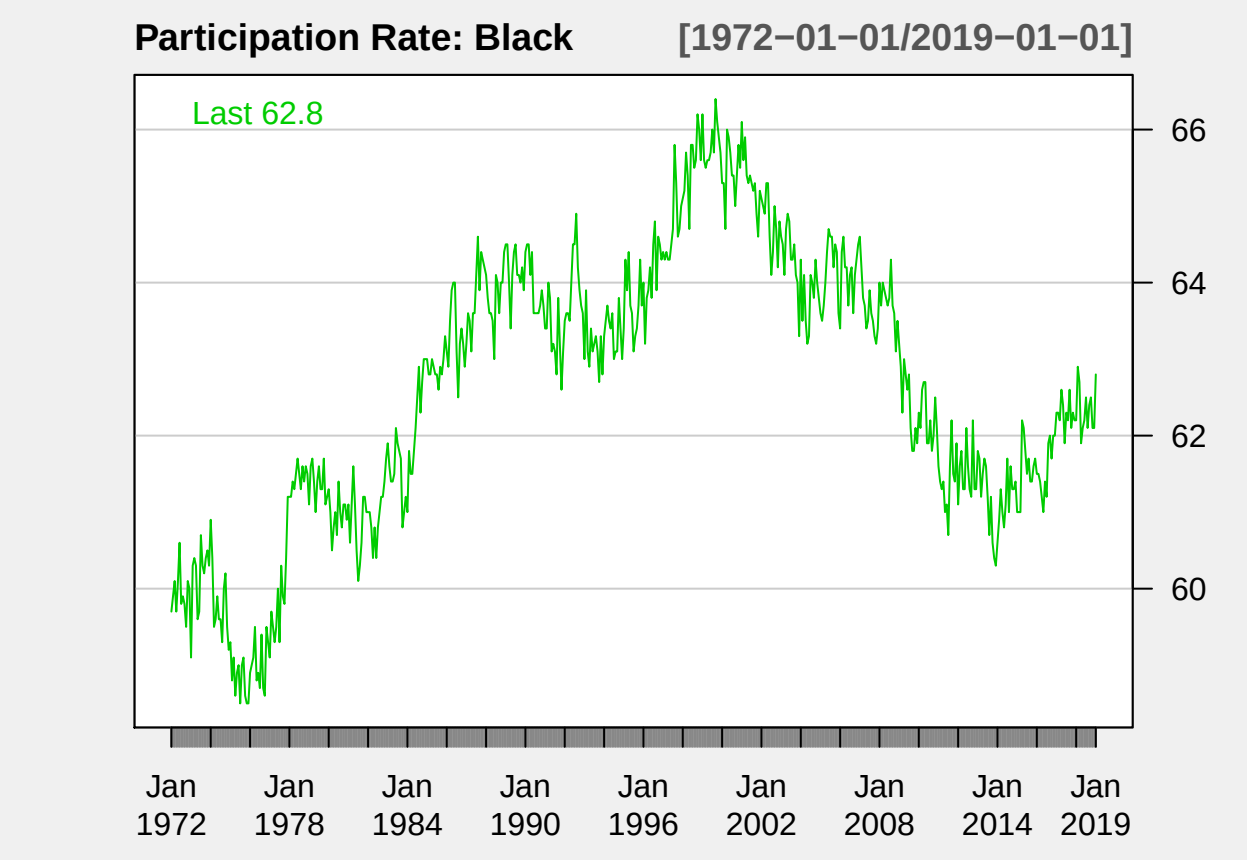

```
PRWomen.data.frame <- as.data.frame(PR)
PRWomen.data.frame$date <- ymd(rownames(PRWomen.data.frame))

PRWomen.time.series <- ts(PRWomen.data.frame$PR,
  start = c(year(min(PRWomen.data.frame$date)),month(min(PRWomen.data.frame$date))),
  end = c(year(max(PRWomen.data.frame$date)),month(max(PRWomen.data.frame$date))),
  frequency=12)

PR <- LNS11300001 # convert to participation rate
dimnames(PR)[2] <- "PR"
chartSeries(PR, theme="white", name="Participation Rate: Men")
```



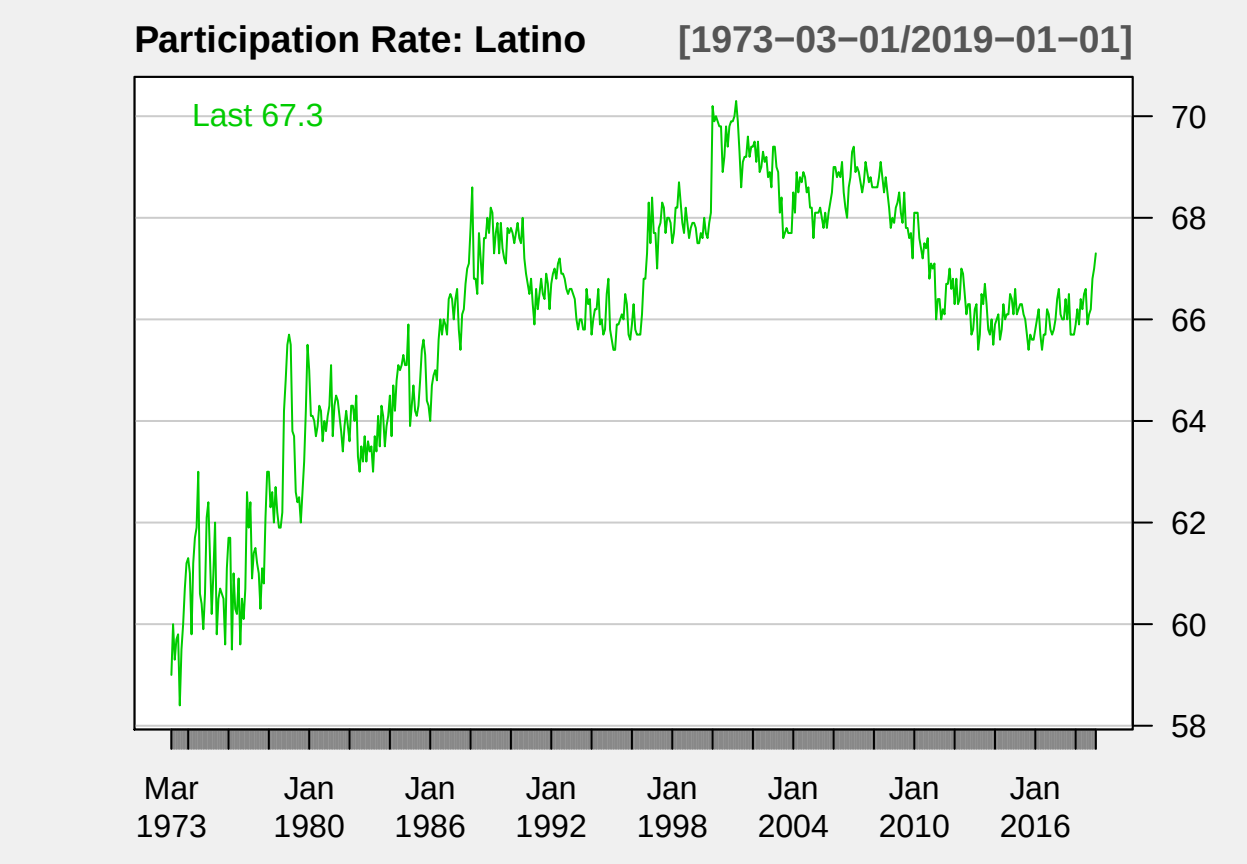

```
PRMen.data.frame <- as.data.frame(PR)
PRMen.data.frame$date <- ymd(rownames(PRMen.data.frame))


PRMen.time.series <- ts(PRMen.data.frame$PR,
  start = c(year(min(PRMen.data.frame$date)),month(min(PRMen.data.frame$date))),
  end = c(year(max(PRMen.data.frame$date)),month(max(PRMen.data.frame$date))),
  frequency=12)

PR <- LNS11300003 # convert to participation rate
dimnames(PR)[2] <- "PR"
chartSeries(PR, theme="white", name="Participation Rate: White")
```




```
PRWhite.data.frame <- as.data.frame(PR)
PRWhite.data.frame$date <- ymd(rownames(PRWhite.data.frame))


PRWhite.time.series <- ts(PRWhite.data.frame$PR,
  start = c(year(min(PRWhite.data.frame$date)),month(min(PRWhite.data.frame$date))),
  end = c(year(max(PRWhite.data.frame$date)),month(max(PRWhite.data.frame$date))),
  frequency=12)

PR <- LNS11300006 # convert to participation rate
dimnames(PR)[2] <- "PR"
chartSeries(PR, theme="white", name="Participation Rate: Black")
```

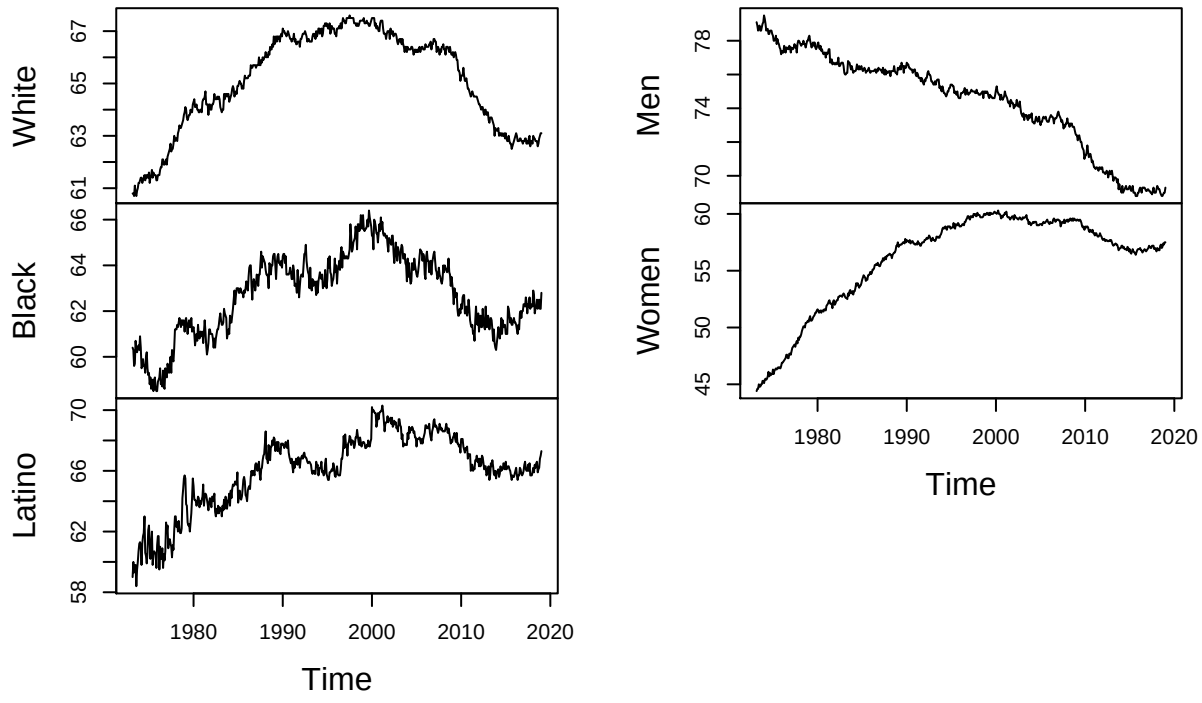



```
PRBlack.data.frame <- as.data.frame(PR)
PRBlack.data.frame$date <- ymd(rownames(PRBlack.data.frame))

PRBlack.time.series <- ts(PRBlack.data.frame$PR,
  start = c(year(min(PRBlack.data.frame$date)),month(min(PRBlack.data.frame$date))),
  end = c(year(max(PRBlack.data.frame$date)),month(max(PRBlack.data.frame$date))),
  frequency=12)

PR <- LNS11300009 # convert to participation rate
dimnames(PR)[2] <- "PR"
chartSeries(PR, theme="white", name="Participation Rate: Latino")
```



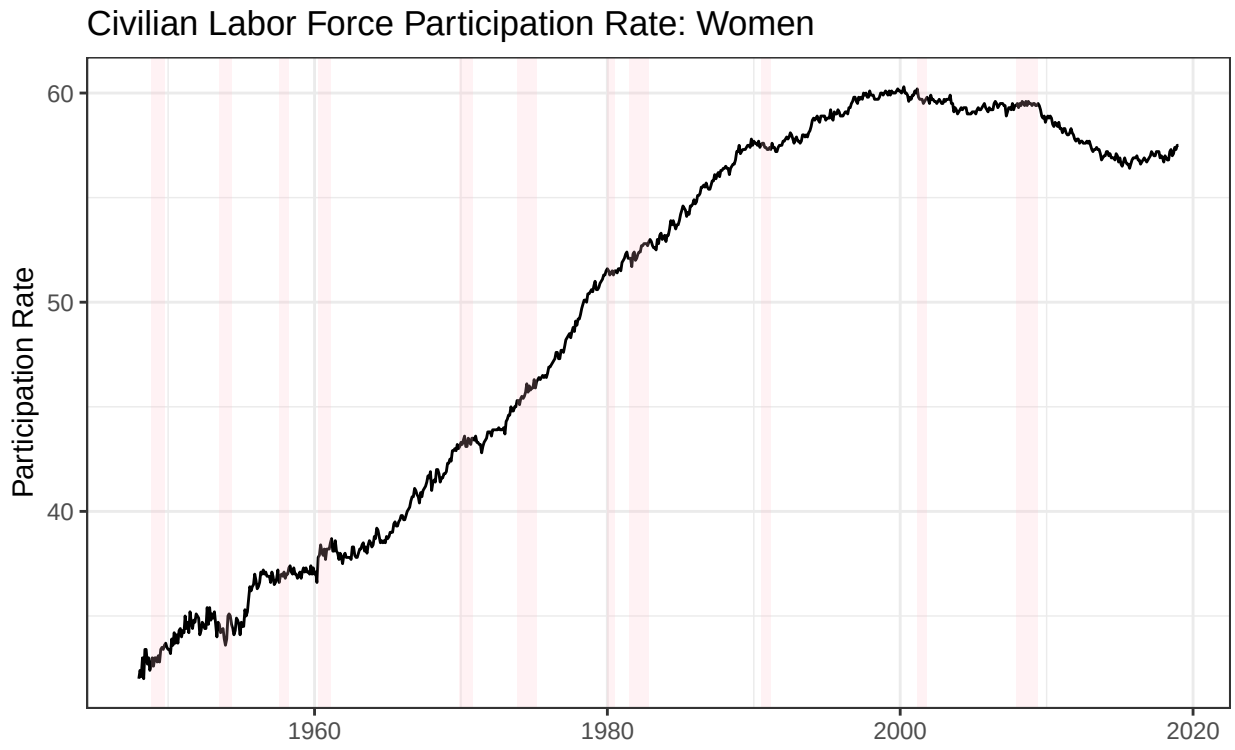

```
PRLatino.data.frame <- as.data.frame(PR)
PRLatino.data.frame$date <- ymd(rownames(PRLatino.data.frame))

PRLatino.time.series <- ts(PRLatino.data.frame$PR,
  start = c(year(min(PRLatino.data.frame$date)),month(min(PRLatino.data.frame$date))),
  end = c(year(max(PRLatino.data.frame$date)),month(max(PRLatino.data.frame$date))),
  frequency=12)
```

Multiple time series



The civilian labor force participation rate for Women, White, Black and Latino has been steadily increasing while the rate for Men has been decreasing. The rates for all groups have been decreasing since the last recession. Data for other races are not available.

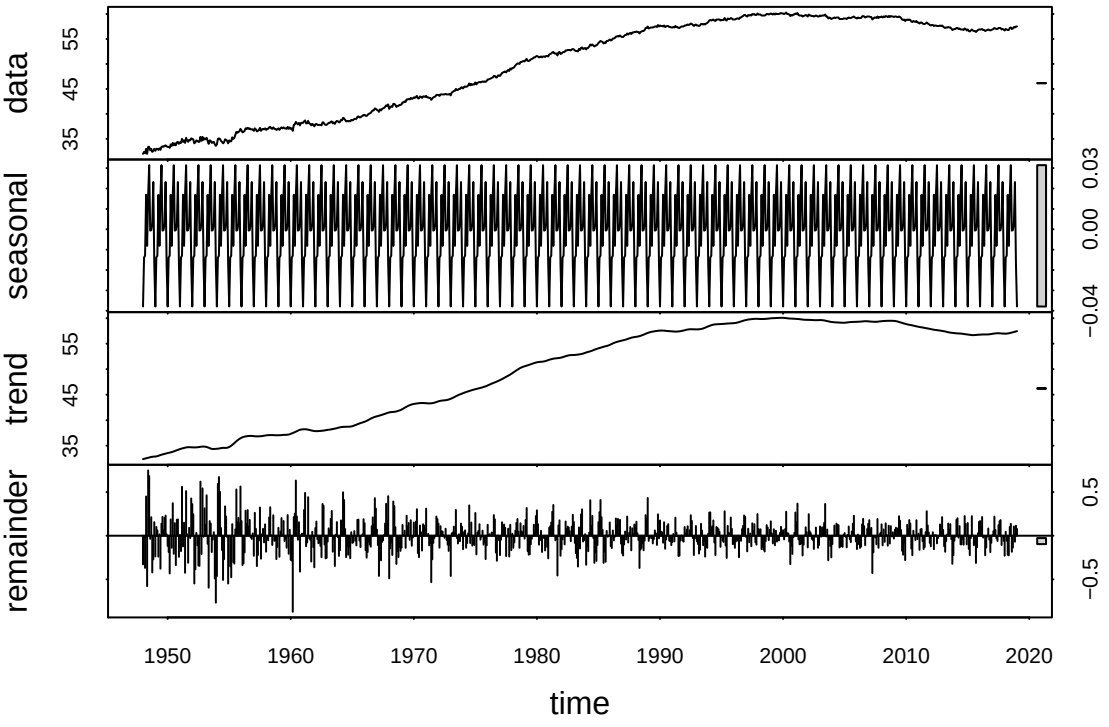
```
ggplot(PR.data.frame) +
  geom_line(aes(x=date, y=PR)) +
  theme_bw() +
  geom_rect(data=recessions.trim, aes(xmin=Peak, xmax=Trough, ymin=-Inf, ymax=+Inf),
            fill='pink', alpha=0.2) +
  labs(title = "Civilian Labor Force Participation Rate: Women ",
       caption = "Source: U.S. Bureau of Labor Statistics\nShaded areas indicate U.S. recessions",
       x = "", y = "Participation Rate")
```



Source: U.S. Bureau of Labor Statistics
Shaded areas indicate U.S. recessions

```
PR.time.series <- ts(PR.data.frame$PR,
  start = c(year(min(PR.data.frame$date)),month(min(PR.data.frame$date))),
  end = c(year(max(PR.data.frame$date)),month(max(PR.data.frame$date))),
  frequency=12)
```

Data decomposition

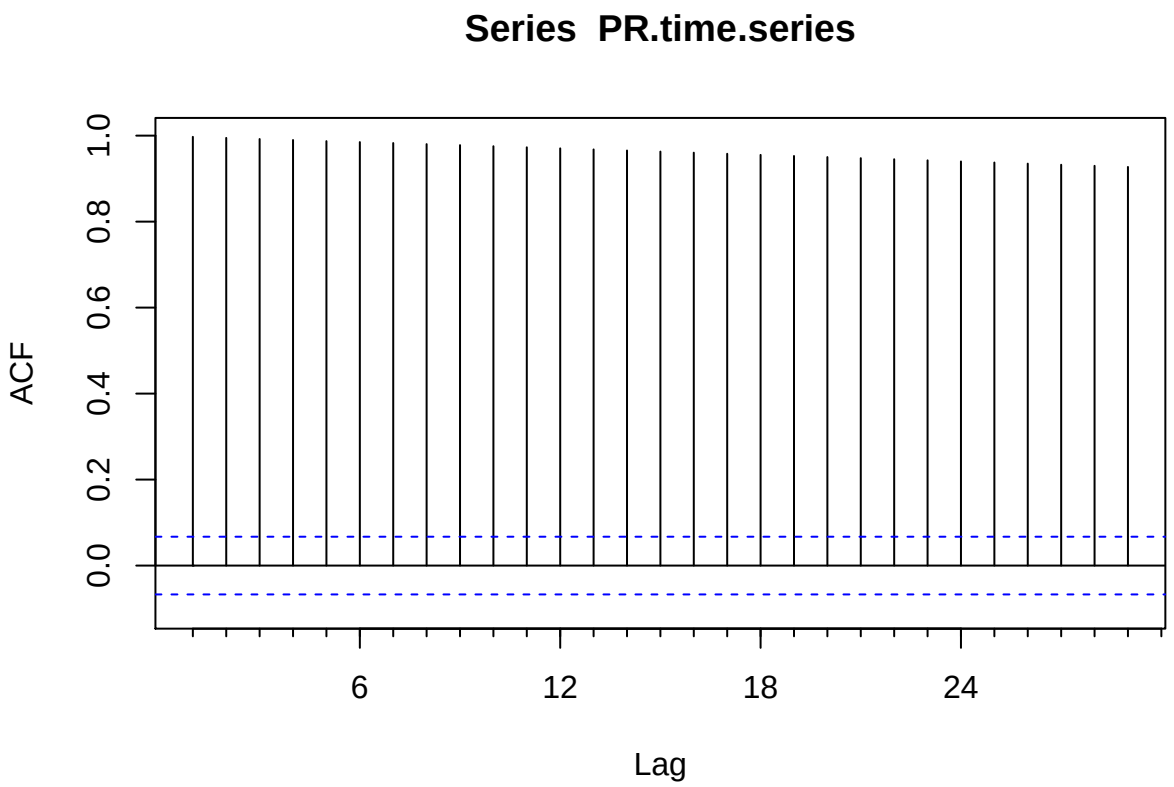


Check for Stationarity

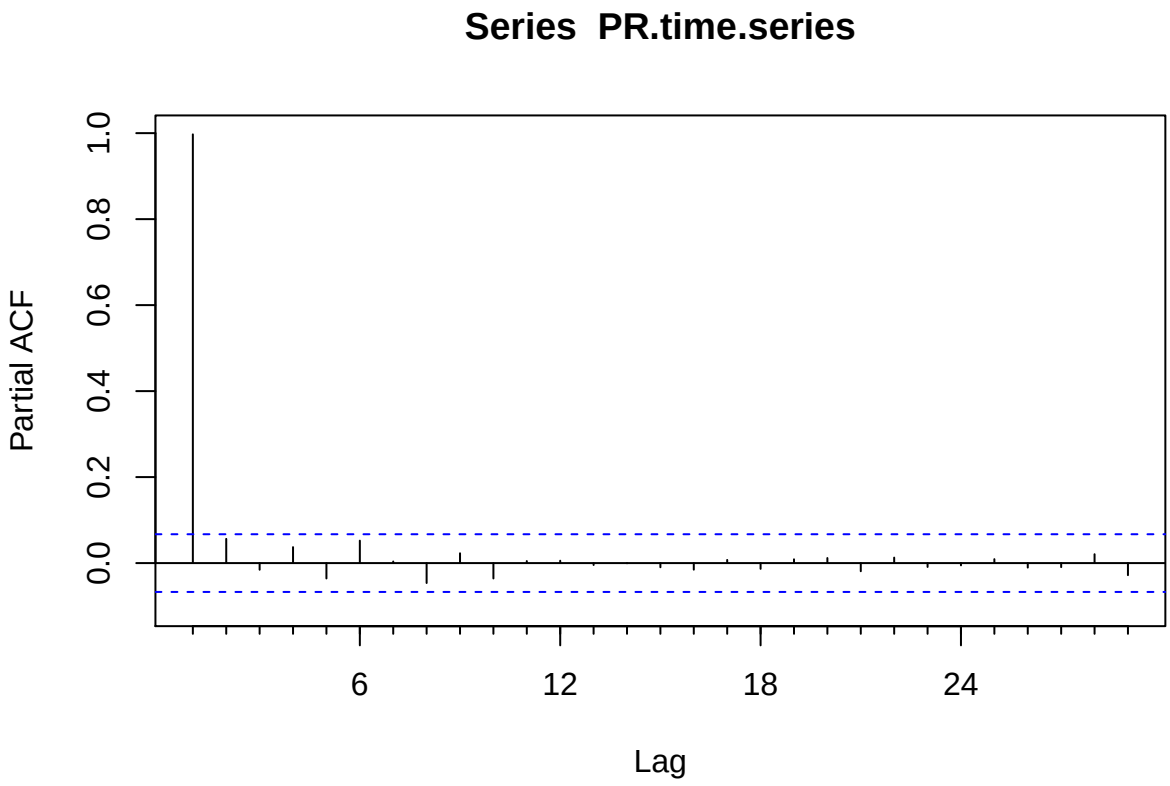
```
##
##  Augmented Dickey-Fuller Test
##
## data:  PR.time.series
## Dickey-Fuller = 1.1954, Lag order = 9, p-value = 0.99
## alternative hypothesis: stationary
```

The data is non-stationary; the labor participation rates changes through time. The ADF test does not reject the null hypothesis of non-stationarity, confirming the visual inspection.

Autocorrelations and Choosing Model Order



ACF plotshows a spike at lags 1 through 30.



PACF plot only shows a spike at lags 1.

Fitting an ARIMA model

```
# ARIMA model fit to the partipation rate data
PR.auto.arima.fit <- auto.arima(PR.time.series, d=NA, D=NA, max.p=5, max.q=5,
  max.P=2, max.Q=2, max.order=3, start.p=1, start.q=1,
  start.P=1, start.Q=1, stationary=FALSE, seasonal=FALSE,
  ic=c("aic"), stepwise=TRUE, trace=FALSE,
  approximation=FALSE, xreg=NULL,
  test=c("kpss","adf","pp"), #seasonal.test=c("ocsb","ch"),
  allowdrift=FALSE, lambda=NULL, parallel=FALSE, num.cores=NULL)
print(summary(PR.auto.arima.fit))
```

```
## Series: PR.time.series
## ARIMA(4,2,3)
##
## Coefficients:
##           ar1      ar2      ar3      ar4      ma1     ma2      ma3
##       -1.3160  -1.4190  -0.5100  -0.1060  -0.0533  0.0513  -0.9595
## s.e.   0.0403   0.0569   0.0552   0.0366   0.0220  0.0179   0.0133
##
## sigma^2 estimated as 0.05325:  log likelihood=41.2
## AIC=-66.39   AICc=-66.22   BIC=-28.42
```

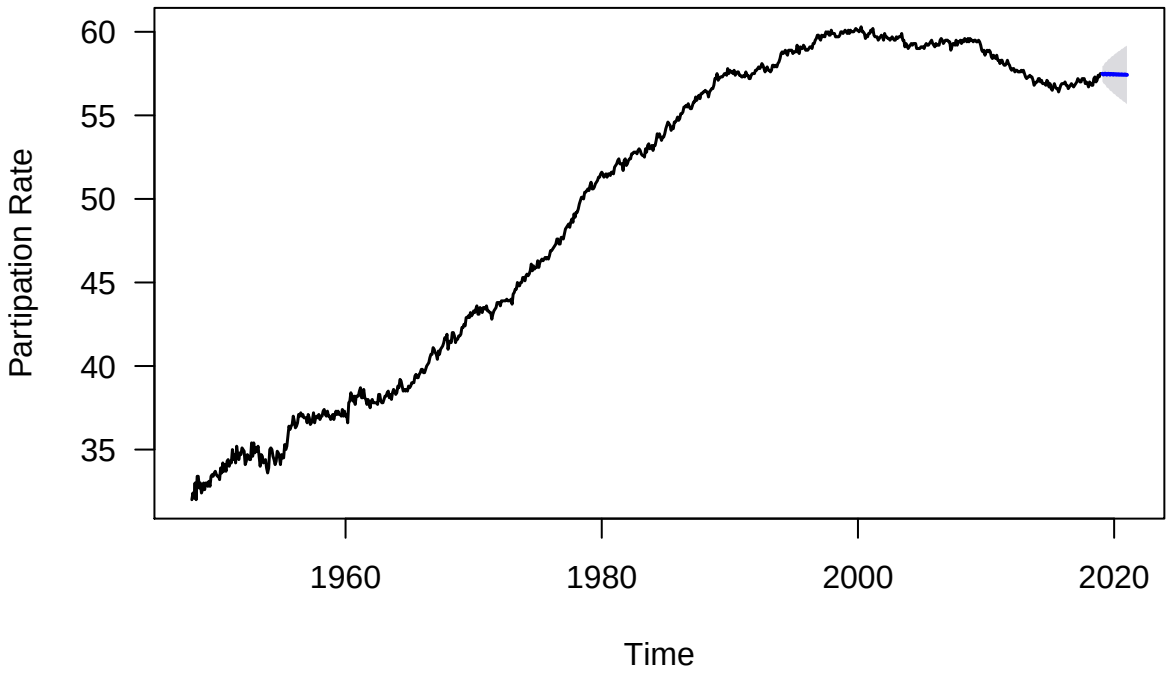
```
##
## Training set error measures:
##                       ME     RMSE       MAE        MPE      MAPE
## Training set -0.008413341 0.229533 0.1706482 -0.0184951 0.3837548
##                    MASE         ACF1
## Training set 0.3194905 -0.005167239
##                       ME     RMSE       MAE        MPE      MAPE
## Training set -0.008413341 0.229533 0.1706482 -0.0184951 0.3837548
##                    MASE         ACF1
## Training set 0.3194905 -0.005167239
```

```
# national partipation rate two-year forecast (horizon h = 24 months)
PR.forecast <- forecast(PR.auto.arima.fit, h=24, level=c(95),
  fan=FALSE, xreg=NULL, bootstrap=FALSE)
```

AR(1) coefficient p = -1.31 tells us that the next value in the series is taken as a dampened previous value by a factor of -1.32 and depends on previous error lag.

Forecast

```
# plot national partipation rate time series with two-month forecast
plot(PR.forecast,main="", ylab="Partipation Rate",
  xlab = "Time", las = 1, lwd = 1.5)
```



	Point Forecast	Lo 95	Hi 95
Feb 2019	57.49107	57.03880	57.94333
Mar 2019	57.44199	56.90726	57.97671
Apr 2019	57.48911	56.86867	58.10955
May 2019	57.49235	56.80330	58.18141
Jun 2019	57.43825	56.67322	58.20328
Jul 2019	57.47707	56.64360	58.31055
Aug 2019	57.48717	56.59842	58.37591
Sep 2019	57.43711	56.48466	58.38955
Oct 2019	57.46566	56.45293	58.47839
Nov 2019	57.48092	56.41800	58.54383
Dec 2019	57.43584	56.31646	58.55522
Jan 2020	57.45531	56.28031	58.63030
Feb 2020	57.47390	56.25149	58.69631
Mar 2020	57.43424	56.15990	58.70858
Apr 2020	57.44596	56.11908	58.77283
May 2020	57.46633	56.09367	58.83898
Jun 2020	57.43221	56.01064	58.85377
Jul 2020	57.43749	55.96558	58.90940
Aug 2020	57.45838	55.94169	58.97508
Sep 2020	57.42970	55.86617	58.99322
Oct 2020	57.42978	55.81754	59.04202
Nov 2020	57.45022	55.79381	59.10663
Dec 2020	57.42668	55.72489	59.12846
Jan 2021	57.42271	55.67346	59.17196

Interpretation

The civilian labor force participation rate for Women has been steadily increasing. The slow decrease/decay in the ACF as the lags increase is due to the upward trend. During the 2007-2009 recession (FROM 2007-12-01 TO 2009-06-01) the partipation rate was between 59.3 and 59.6. After the recession the rate has been steadily decreasing. In the last decade the the participation rate was between 56.4 and 59.5 with lowest participation rate on 9-2015. The 24 month forecast shows the rate to remain steady.