Crystal Mosley

Predict 413, Sec 55

Homework #1

*\*\*All graphs for each question is at end of HW solutions.\*\**

**SECTION 2.10**

**Problem 6—**

library(fma)

library(forecast)

library(fpp)

library(fpp2)

autoplot(hsales)

ggseasonplot(hsales)

ggsubseriesplot(hsales)

gglagplot(hsales)

ggAcf(hsales)

autoplot(usdeaths)

ggseasonplot(usdeaths)

ggsubseriesplot(usdeaths)

gglagplot(usdeaths)

ggAcf(usdeaths)

autoplot(bricksq)

ggseasonplot(bricksq)

ggsubseriesplot(bricksq)

gglagplot(bricksq)

ggAcf(bricksq)

autoplot(sunspotarea)

ggseasonplot(sunspotarea)

ggsubseriesplot(sunspotarea)

gglagplot(sunspotarea)

ggAcf(sunspotarea)

autoplot(gasoline)

ggseasonplot(gasoline)

ggsubseriesplot(gasoline)

gglagplot(gasoline)

ggAcf(gasoline)

Sunspotarea: ggseasonplot() and ggsubseriesplot() couldn’t be used because the data isn’t seasonal.

* Error in ggseasonplot(sunspotarea) : Data are not seasonal
* Error in ggsubseriesplot(sunspotarea) : Data are not seasonal

Gasoline: ggsubseriesplot() couldn’t be used because there isn’t more than 1 observation.

* Each season requires at least 2 observations. This may be caused from

specifying a time-series with non-integer frequency.

1. Data with seasonality: *Gasoline*, *Bricksq*, *Hsales*, and *Usdeaths* all show some seasonality.

Data with cyclicity: *Hsales* and *Usdeaths* both show cyclicity.

Data with trend: *Bricksq* and *Gasoline* show upward trends.

1. **Hsales**: strong positive seasonality in “lag 1” of the lag plot; the ggseasonplot shows a large incline in sales for the first 3 months of the year for over 3 decades.

**Usdeaths**: strong positive seasonality in “lag 1” and “lag 12” of the lag plot; the autoplot shows more white noise than anything; from ’73 – ’78, there was an incline in deaths in the US from Feb-July.

**Bricksq**: the ACF plot shows trend and seasonality.

**Sunspotarea**: no seasonality; stead trend in the autoplot increasing to 1950 and then steadily deacreasing; the lag plot shows positive seasonality for “lag 1”, and negative seasonality for “lag 5” and “lag 6”.

**Gasoline**: the ACF plot shows trend and seasonality.

**SECTION 3.7**

**Problem 9—**

train1 <- window(visnights[, "QLDMetro"], end = c(2015, 4))

fc1 <- snaive(train1, h=11)

fc1.1 <- window(visnights, start=1998)

accuracy(fc1.1, train1)

ME RMSE MAE MPE MAPE ACF1

Test set 3.898585 4.106349 3.898585 36.23672 36.23672 0.006534756

Theil's U

Test set 2.016095

train2 <- window(visnights[, "QLDMetro"], end = c(2014, 4))

fc2 <- snaive(train2, h=11)

fc2.1 <- window(visnights, start=1998)

accuracy(fc2.1, train2)

ME RMSE MAE MPE MAPE ACF1

Test set 3.936922 4.14706 3.936922 36.55018 36.55018 -0.005140185

Theil's U

Test set 2.004495

train3 <- window(visnights[, "QLDMetro"], end = c(2013, 4))

fc3 <- snaive(train3, h=11)

fc3.1 <- window(visnights, start=1998)

accuracy(fc3.1, train3)

ME RMSE MAE MPE MAPE ACF1

Test set 3.955034 4.175647 3.955034 36.63211 36.63211 -0.006798198

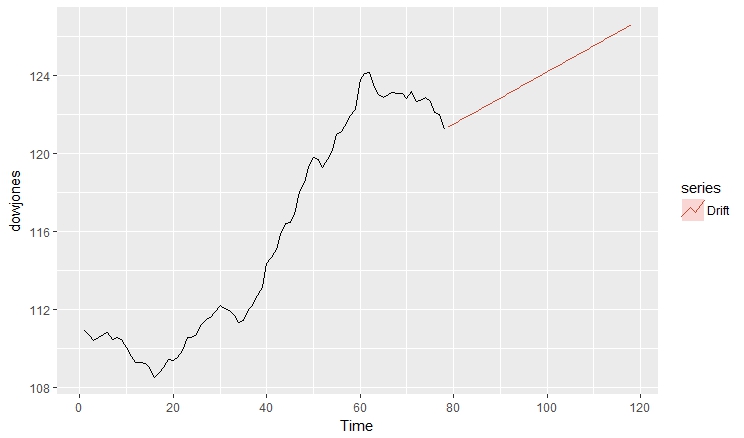
Theil's U

Test set 1.970646

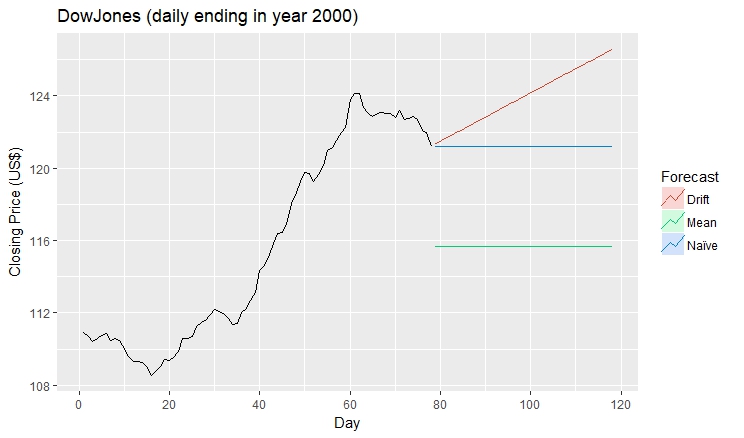
Training set #3’s snaive MAPE shows a better result than the other 2 sets, but only by a little; they’re all 3 relatively the same results.

**Problem 10—**

autoplot(dowjones)+autolayer(rwf(dowjones, drift=TRUE, h=40), series="Drift", PI=FALSE)



autoplot(dowjones) + autolayer(meanf(dowjones, h=40), series="Mean", PI=FALSE) + autolayer(rwf(dowjones, h=40), series="Naïve", PI=FALSE) + autolayer(rwf(dowjones, drift=TRUE, h=40), series="Drift", PI=FALSE) + ggtitle("DowJones (daily ending in year 2000)") + xlab("Day") + ylab("Closing Price (US$)") + guides(colour=guide\_legend(title="Forecast"))

****

Drift or Naïve would be the better methods to utilize because Drift allows the forecasts to increase or decrease over time, and Naïve is good for economic and financial data – the Dow Jones is exactly this.

**Problem 11—**

autoplot(ibmclose, main="Closing Price of IBM Stock")

legend("topright", lty=1, col=c("red", "blue"), legend=c("30 day rolling average","60 day rolling average"))

ibmclose.training <- window(ibmclose, start=1, end=300)

ibmclose.test <- window(ibmclose, start=301)

ibmclose.prediction\_horizon <- length(ibmclose) - 300;

ibmclose.fit.mean <- meanf(ibmclose.training, h=ibmclose.prediction\_horizon)

ibmclose.fit.naive <- naive(ibmclose.training, h=ibmclose.prediction\_horizon)

ibmclose.fit.seasonalNaive <- snaive(ibmclose.training, h=ibmclose.prediction\_horizon)

ibmclose.fit.drift <- rwf(ibmclose.training, drift=TRUE, h=ibmclose.prediction\_horizon)

autoplot(ibmclose.training, main="Closing Price of IBM Stock")

autoplot(ibmclose.fit.mean, plot.conf=FALSE, main="Closing Price of IBM Stock")

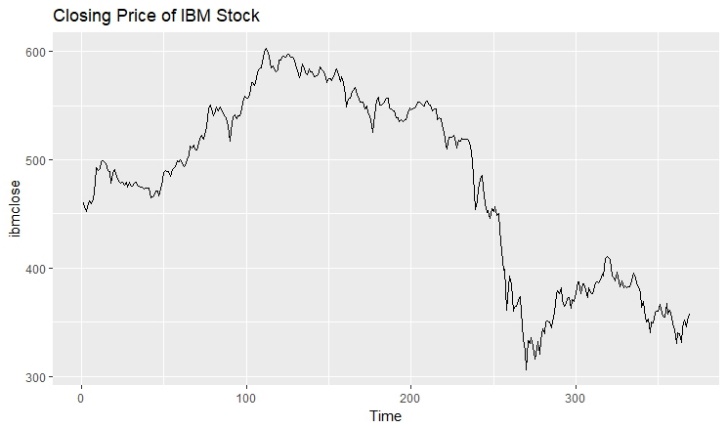
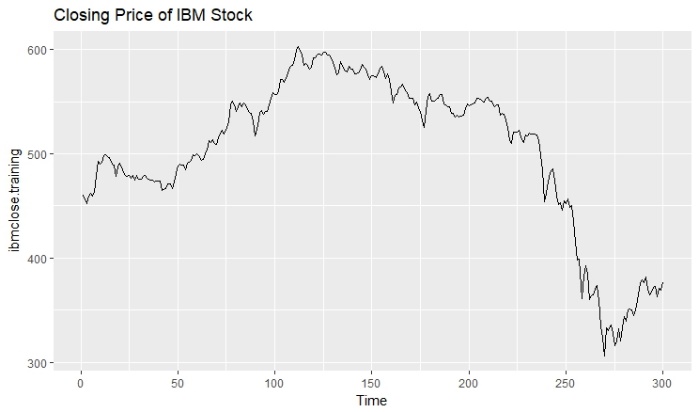
lines(ibmclose.fit.naive$mean, col=2)

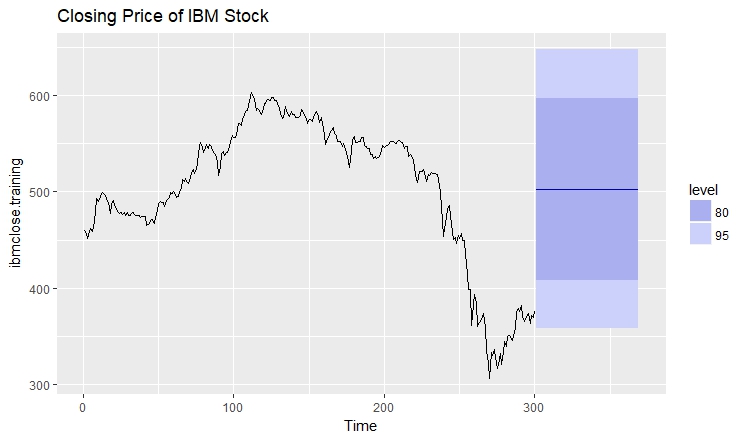
lines(ibmclose.fit.seasonalNaive$mean, col=3)

lines(ibmclose.fit.drift$mean, col=1)

lines(ibmclose)

legend("topright", lty=1, col=c(4,2,3,1), legend=c("Mean method","Naive method","Seasonal naive method", "Drift"))



accuracy(ibmclose.fit.mean, ibmclose.test)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set 1.660438e-14 73.61532 58.72231 -2.642058 13.03019 11.52098 0.9895779 NA

Test set -1.306180e+02 132.12557 130.61797 -35.478819 35.47882 25.62649 0.9314689 19.05515

accuracy(ibmclose.fit.naive, ibmclose.test)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set -0.2809365 7.302815 5.09699 -0.08262872 1.115844 1.000000 0.1351052 NA

Test set -3.7246377 20.248099 17.02899 -1.29391743 4.668186 3.340989 0.9314689 2.973486

accuracy(ibmclose.fit.seasonalNaive, ibmclose.test)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set -0.2809365 7.302815 5.09699 -0.08262872 1.115844 1.000000 0.1351052 NA

Test set -3.7246377 20.248099 17.02899 -1.29391743 4.668186 3.340989 0.9314689 2.973486

accuracy(ibmclose.fit.drift, ibmclose.test)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set -3.916293e-14 7.297409 5.127996 -0.02530123 1.121650 1.006083 0.1351052 NA

Test set 6.108138e+00 17.066963 13.974747 1.41920066 3.707888 2.741765 0.9045875 2.361092

The mean error is too high compared to the rest. It seems that drift overall tends to have lower error values that the naive methods, but are generally close.

**Problem 12—**

autoplot(hsales, main="Sales of new one-family houses in the USA")

ggseasonplot(hsales, col=1:20, pch=19, year.labels=TRUE, year.labels.left=TRUE)

ggsubseriesplot(hsales,ylab="$ million",xlab="Month",xaxt="n", main="Sales of new one-family houses in the USA")

axis(1,at=1:12,labels=month.abb,cex=0.8)

hsales.training <- window(hsales, start=c(1973, 1), end=c(1993, 12))

hsales.test <- window(hsales, start=c(1994, 1), end=c(1995, 11))

hsales.prediction\_horizon <- 23;

hsales.fit.mean <- meanf(hsales.training, h=hsales.prediction\_horizon)

hsales.fit.naive <- naive(hsales.training, h=hsales.prediction\_horizon)

hsales.fit.seasonalNaive <- snaive(hsales.training, h=hsales.prediction\_horizon)

hsales.fit.drift <- rwf(hsales.training, drift=TRUE, h=hsales.prediction\_horizon)

plot(hsales.training, main="Sales of new one-family houses in the USA w/ Forecasts")

plot(hsales.fit.mean, plot.conf=FALSE, main="Sales of new one-family houses in the USA")

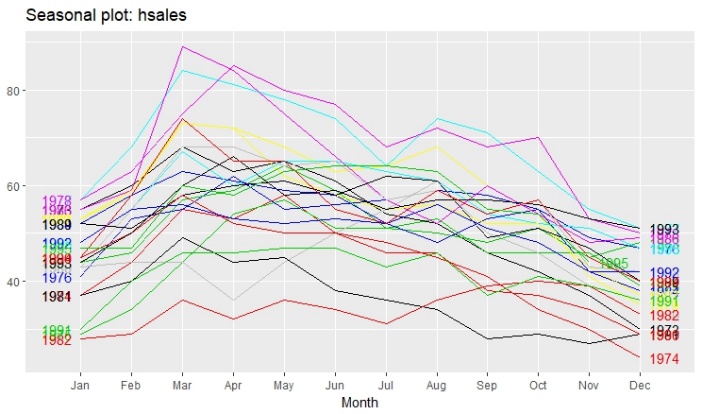
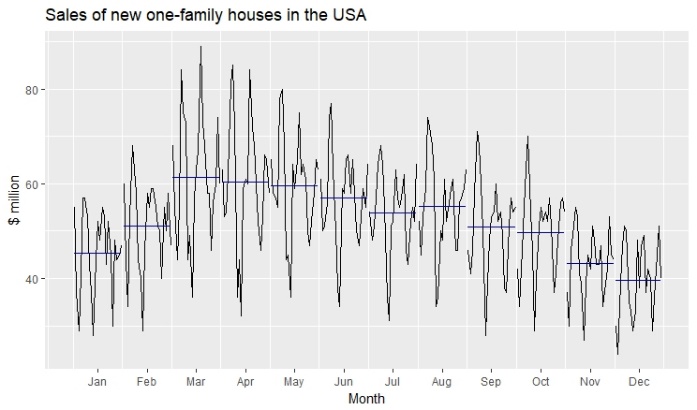
lines(hsales.fit.naive$mean, col=2)

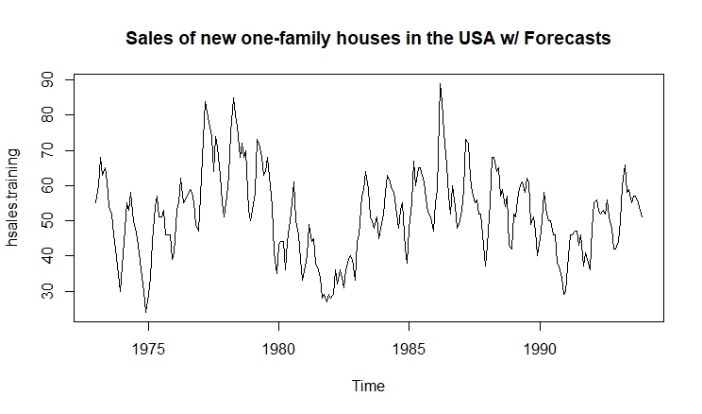
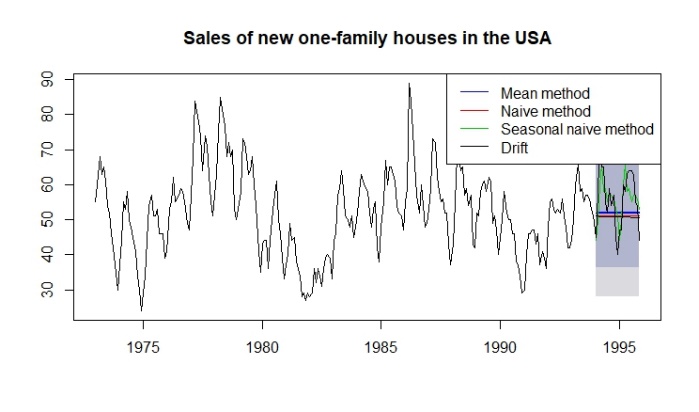
lines(hsales.fit.seasonalNaive$mean, col=3)

lines(hsales.fit.drift$mean, col=1)

lines(hsales)

legend("topright", lty=1, col=c(4,2,3,1), legend=c("Mean method","Naive method","Seasonal naive method", "Drift"))

accuracy(hsales.fit.mean, hsales.test)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set 2.480763e-15 12.138802 9.498898 -6.120182 20.30851 1.119163 0.8661515 NA

Test set 4.051587e+00 9.216133 7.850759 5.074990 13.75973 0.924979 0.5095178 1.13105

accuracy(hsales.fit.naive, hsales.test)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set -0.01593625 6.289813 4.988048 -0.7800232 9.880157 0.5876934 0.1829708 NA

Test set 5.00000000 9.670664 8.304348 6.8080182 14.381673 0.9784210 0.5095178 1.179633

accuracy(hsales.fit.seasonalNaive, hsales.test)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set 0.1375000 10.576113 8.4875 -2.1016380 17.63375 1.0000000 0.838108 NA

Test set 0.3043478 6.160886 5.0000 -0.7312374 9.12828 0.5891016 0.224307 0.8031005

accuracy(hsales.fit.drift, hsales.test)

ME RMSE MAE MPE MAPE MASE ACF1 Theil's U

Training set -1.998380e-12 6.289793 4.987730 -0.7474544 9.87819 0.5876560 0.1829708 NA

Test set 5.191235e+00 9.761548 8.393037 7.1599507 14.50303 0.9888703 0.5083059 1.188562

I’d say the seasonal naive is the better method. Visually, the seasonal naive method overlaps the test data very well, and has the least amount of error for any error metric.

**SECTION 5.10**

**Problem 6—**

gas1 <- window(gasoline, start=1991)

autoplot(gas1) + xlab("Year") + ylab("million barrels per day")

fit.gas <- tslm(gas1 ~ trend + season)

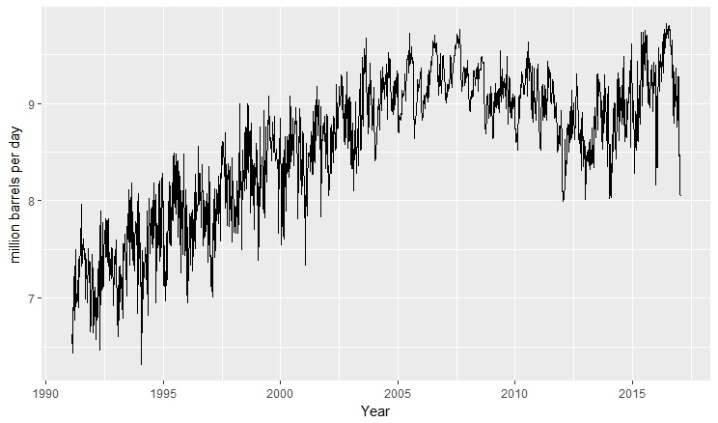
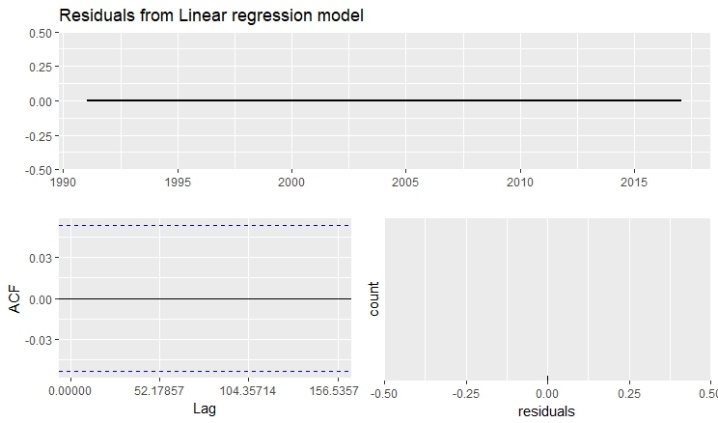
summary(fit.gas)

fourier.gas <- tslm(gas1 ~ trend + fourier(gasoline, K=2))

summary(fourier.gas)

checkresiduals(fit.gas)

fc <- forecast(fit.gas, newdata=data.frame(fourier(gasoline,K=5,h=12)))

 ****

**SECTION 6.9**

**Problem 6—**

bricksq %>%

stl(t.window=13, s.window="periodic", robust=FALSE) %>%

autoplot()

seasonal(bricksq)

seasadj(bricksq)

fit %>% seasadj() %>% naive() %>%

autoplot()

fcast <- stlf(bricksq, method='naive')

bricksq %>%

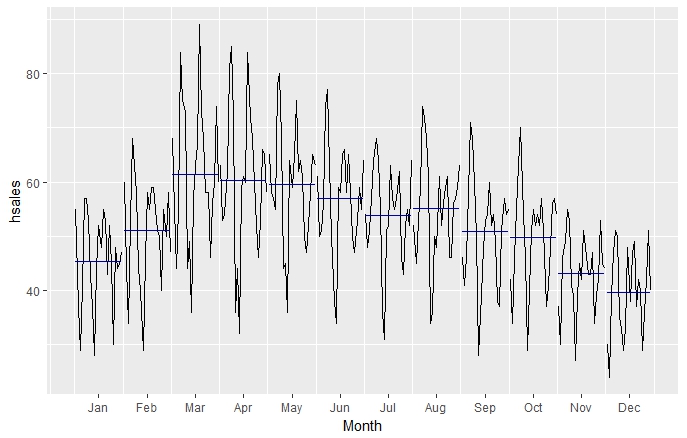
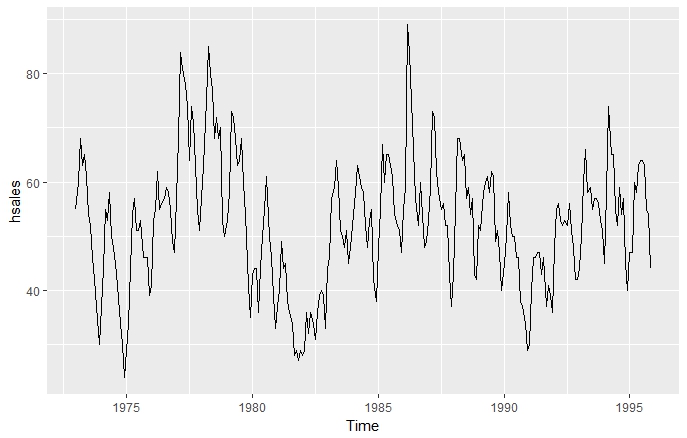
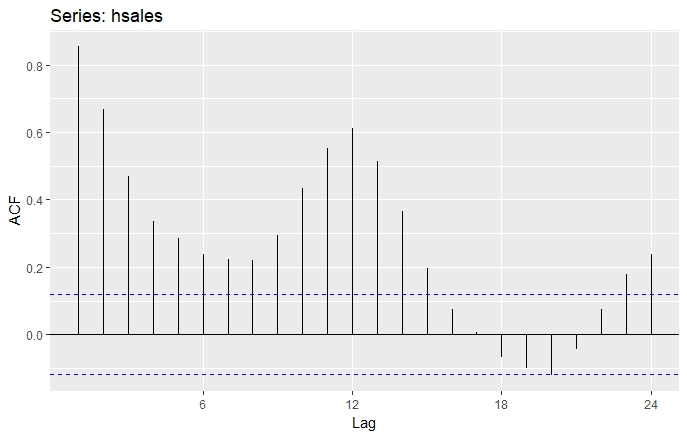
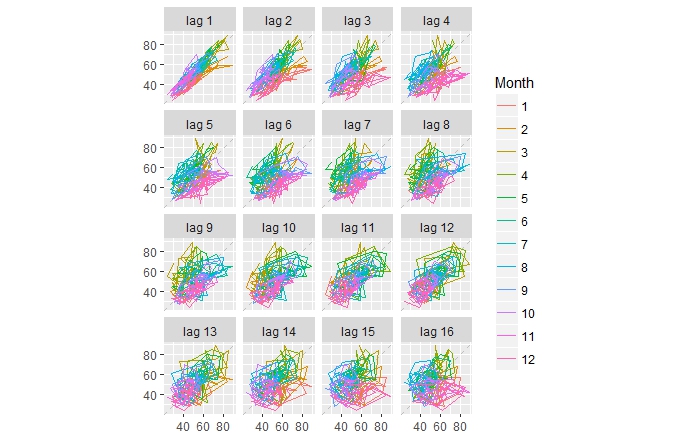
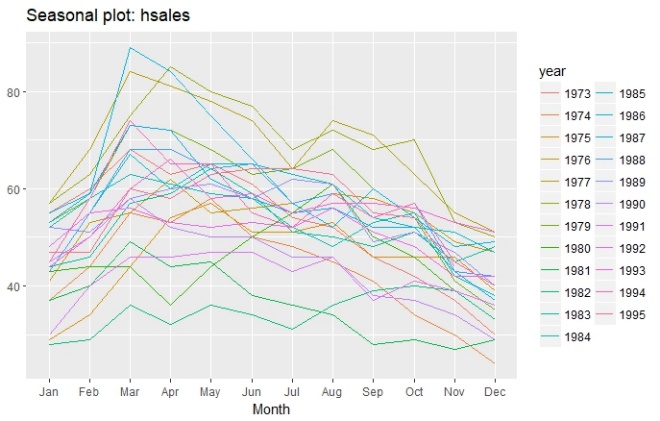
stl(t.window=13, s.window="periodic", robust=TRUE) %>%

autoplot()

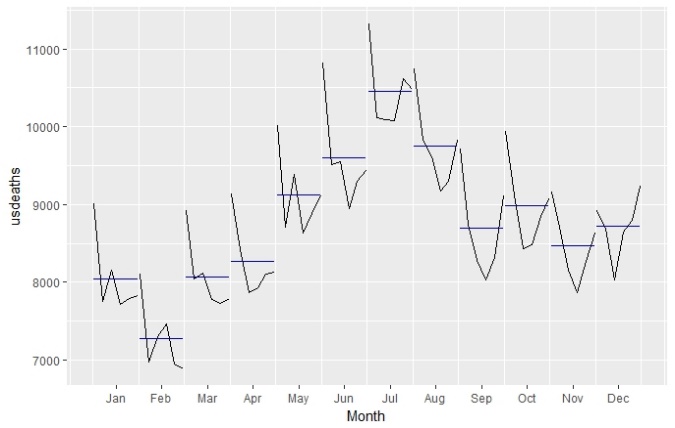
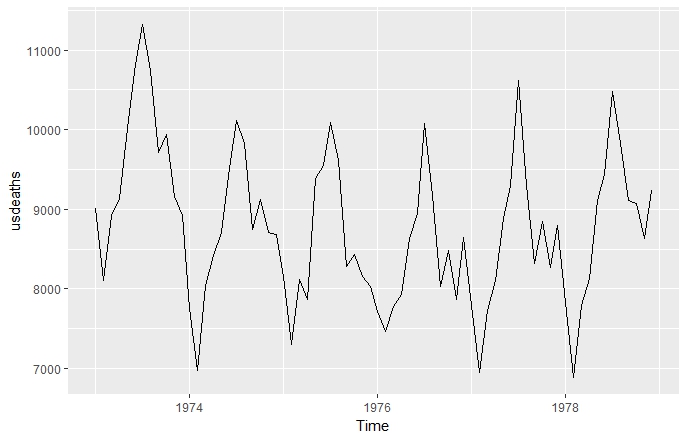
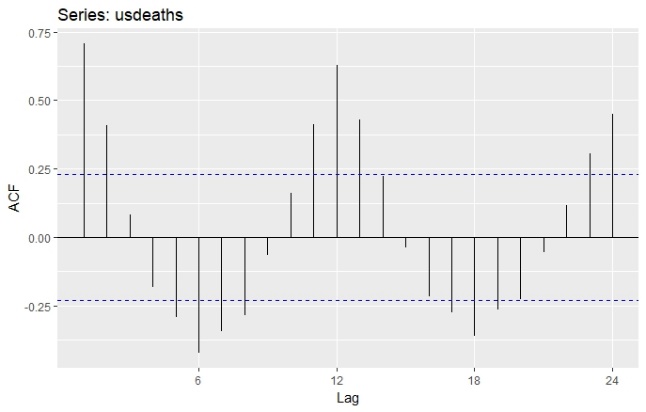
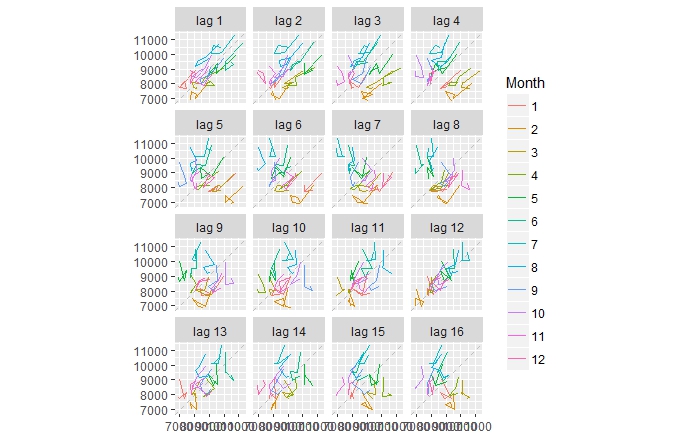
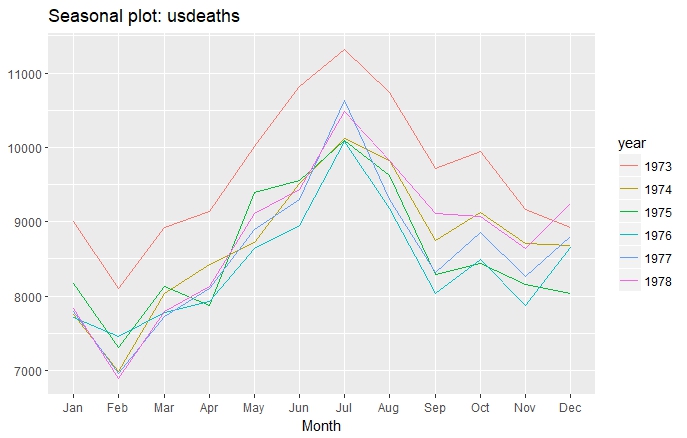
**SECTION 7.8**

**Problem 11—**

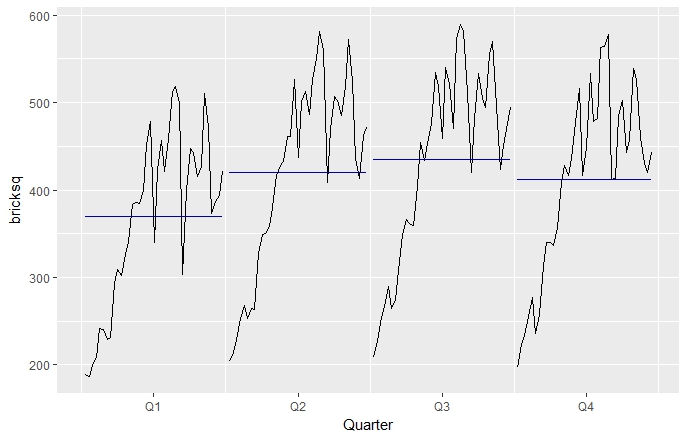
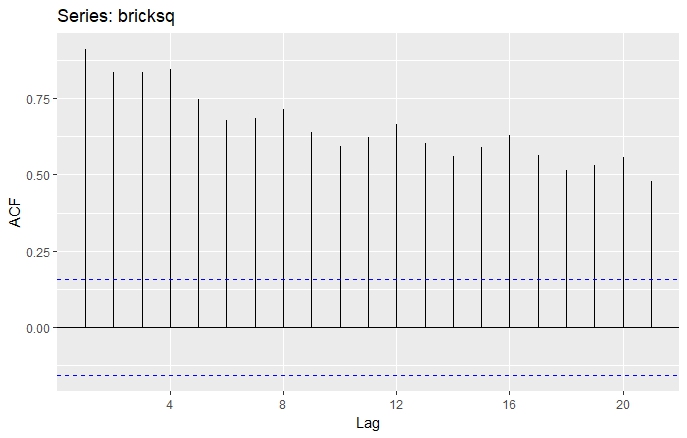
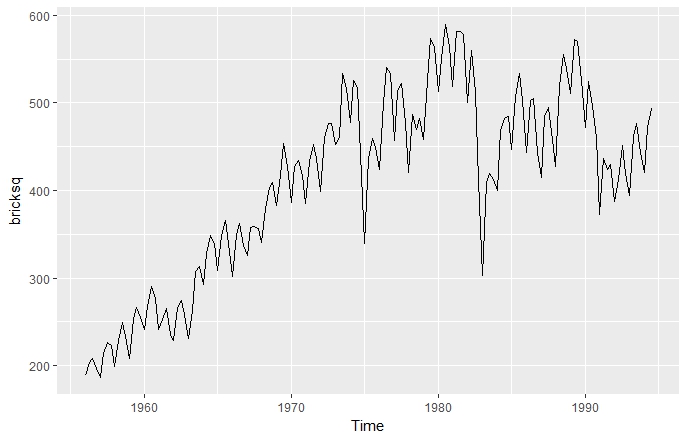
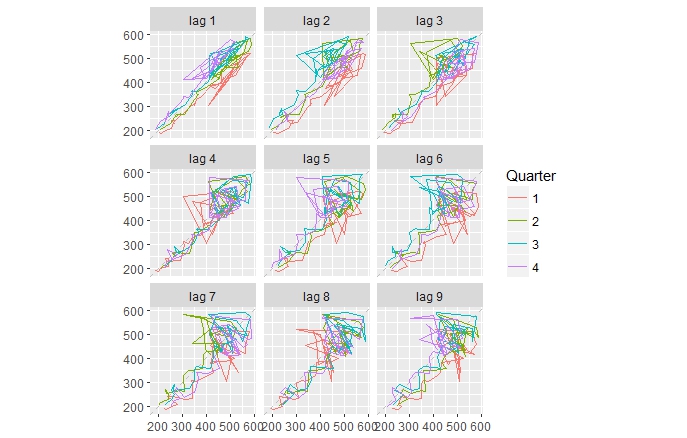
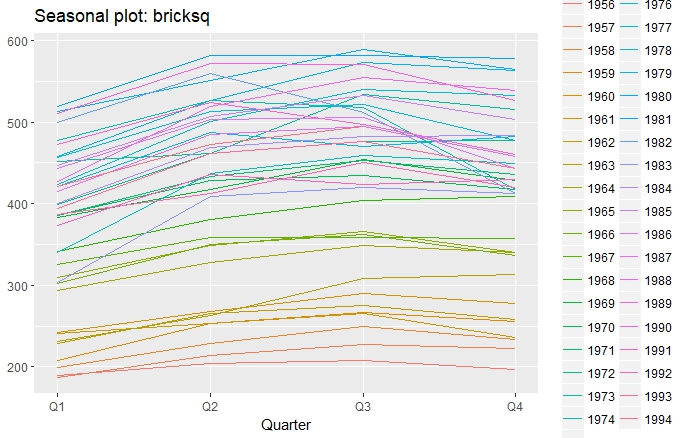
Graphs – “hsales”



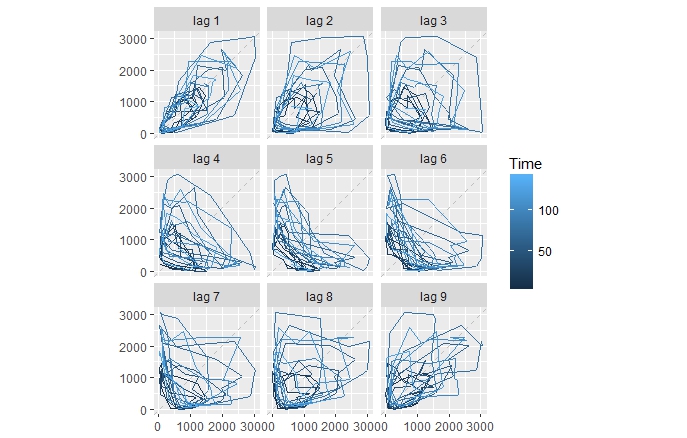
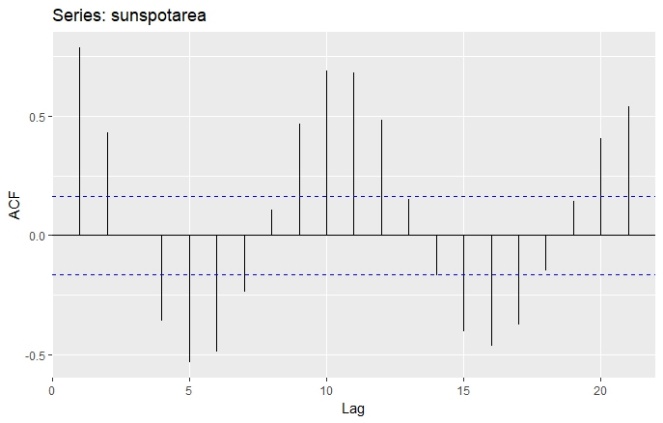
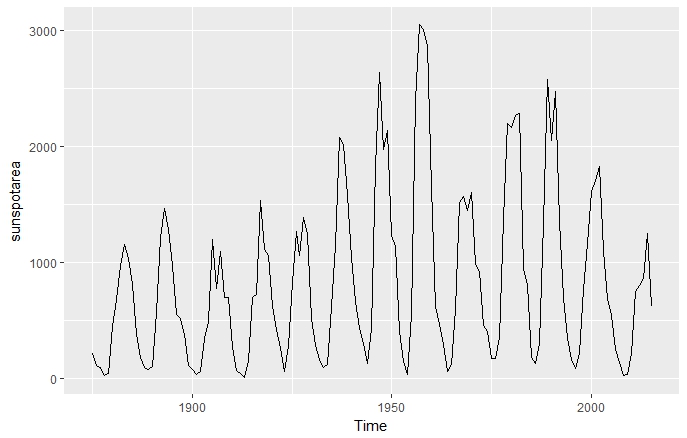
Graphs – “usdeaths”



Graphs – “bricksq”



Graphs – “sunspotarea”



Graphs – “gasoline”

