# Homework 7

Joshua Ingram
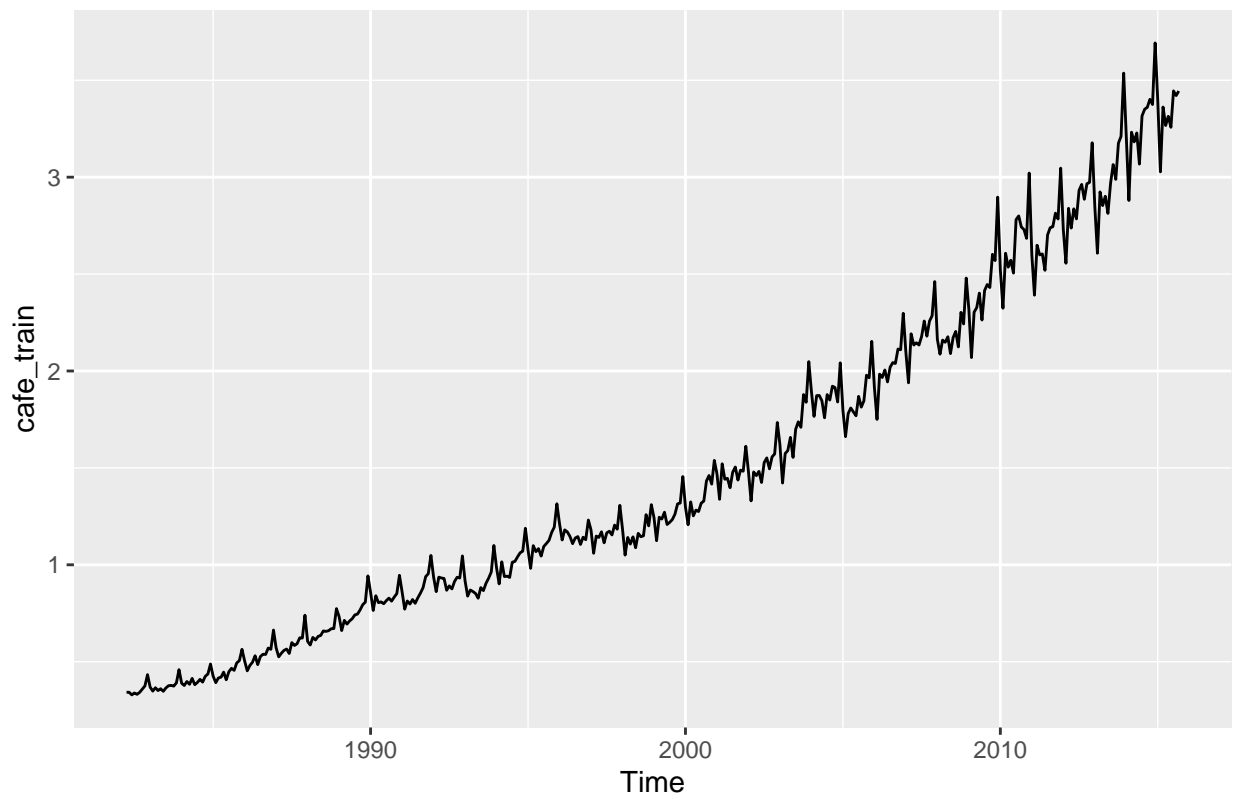
10/20/2020

## Problem 1
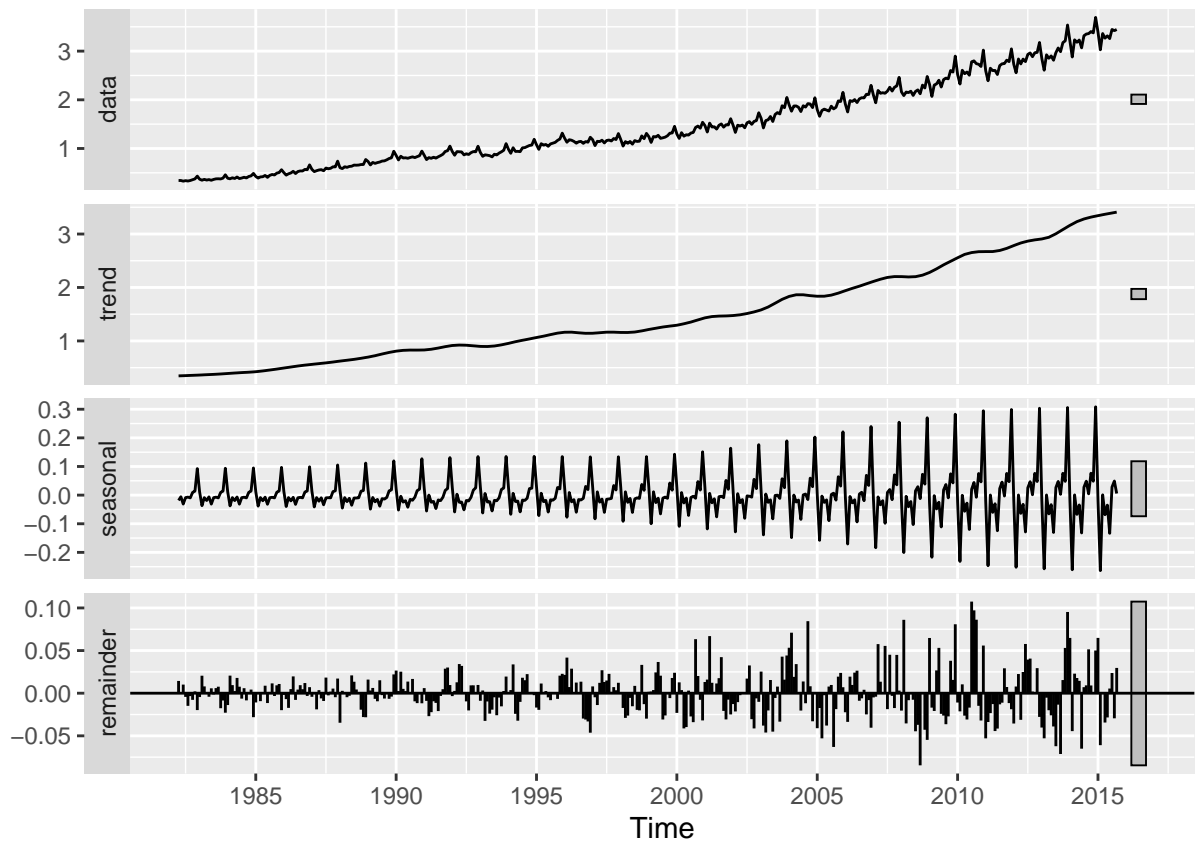
```
cafe_train <- window(cafe, end = c(2015, 9))
```

**1.**

```
autoplot(cafe_train)
```

```
autoplot(stl(cafe_train, s.window=12))
```



There is clearly increasing variance as the total montly expenditure on cafes in Australia increase. This can be seen in both the original time series plot and in the STL decomposition plot, speficially within the seasonal component plot. We should use the Box-Cox method to find the best value of lambda in order to get the transformation with stable variance.

```
lambda_opt <- BoxCox.lambda(cafe_train)
cafe_trans <- BoxCox(cafe_train, lambda = lambda_opt)
autoplot(cafe_trans)
```
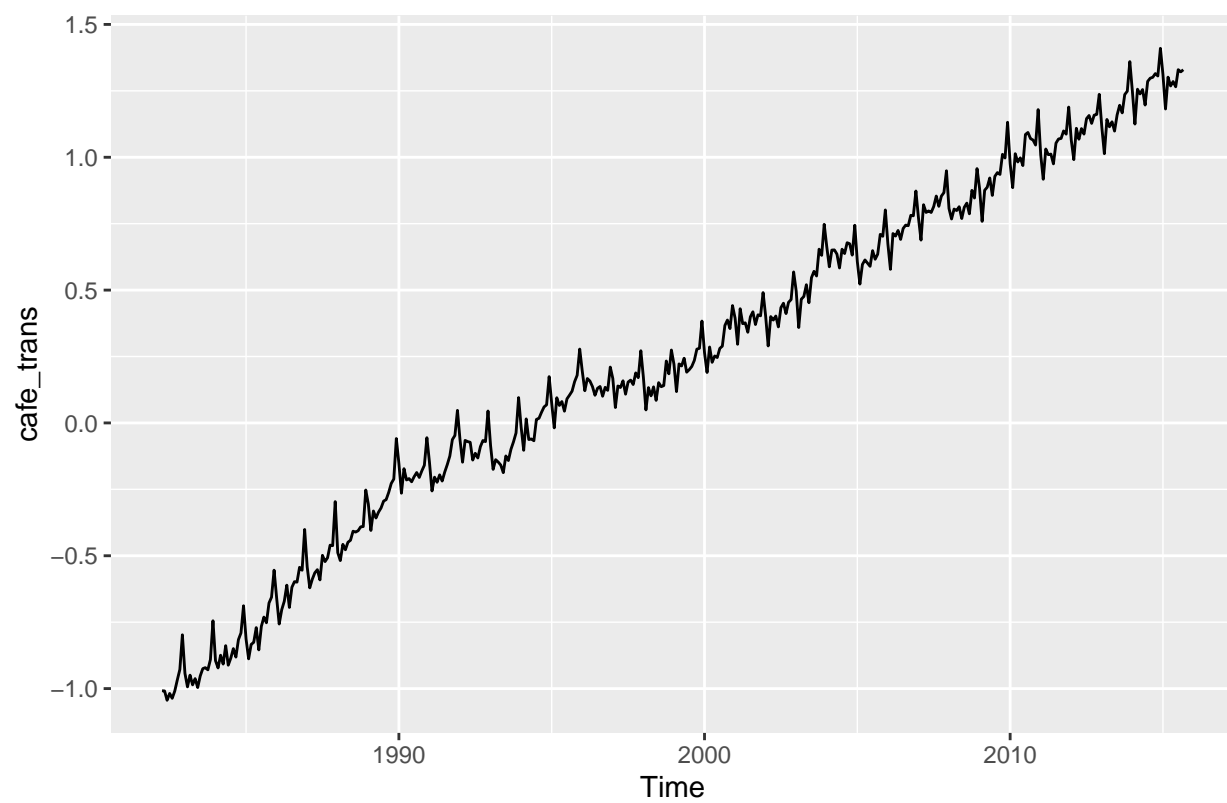
```
autoplot(stl(cafe_trans, s.window = 12))
```
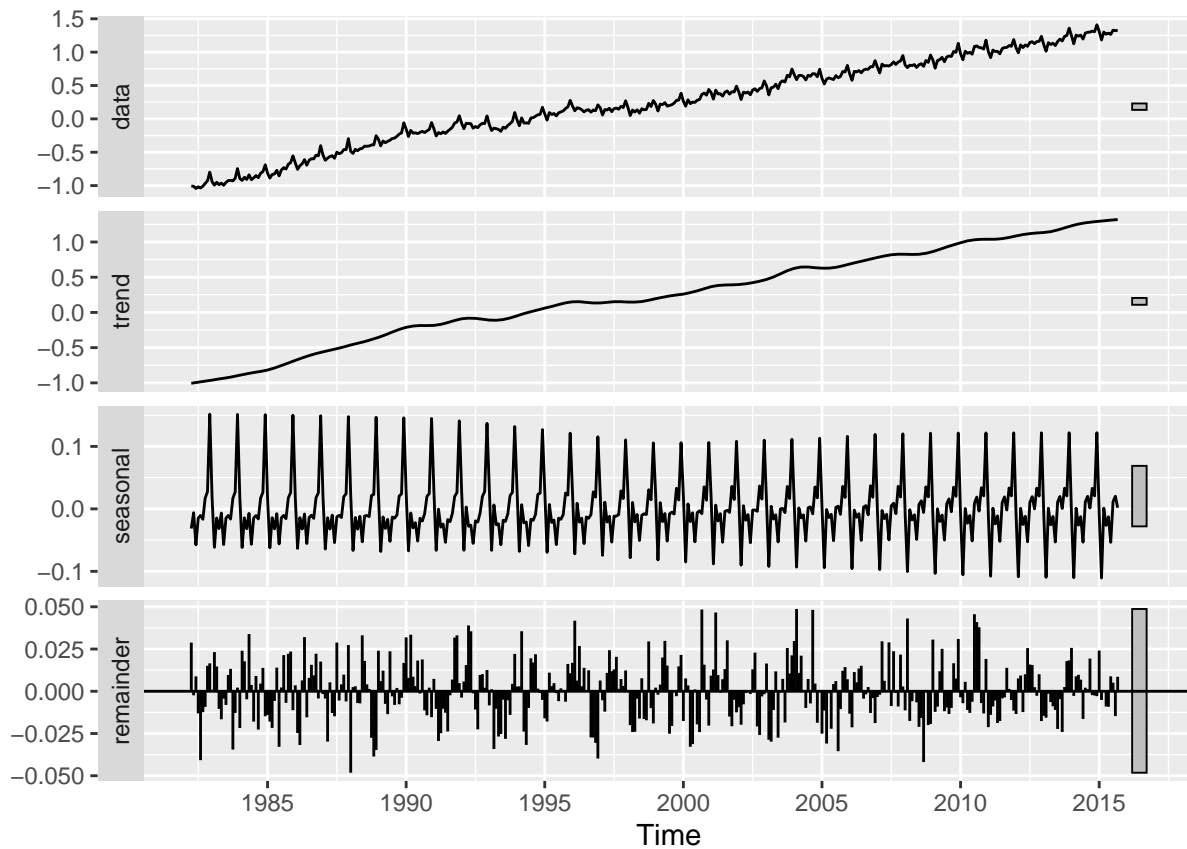
Using the BoxCox.lambda function, the optimal value of lambda is '0.1154731" and is used to transform the data shown in the above graphs. The variance is now stable over time, as can be seen in the graph of the original data and the graph of the seasonal component from the STL decomposition.
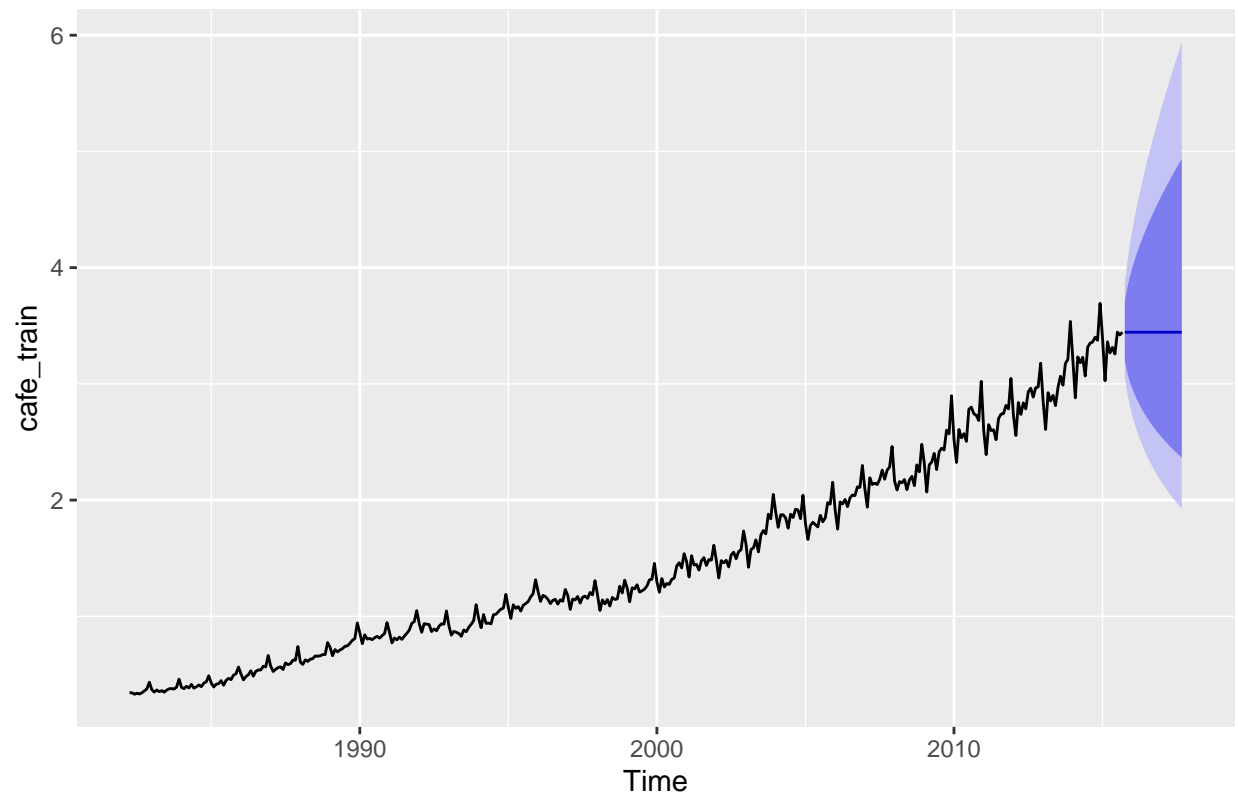
## 2.

**Naive Method**

**a.**

```
fit_naive <- naive(cafe_train, h = 24, lambda = lambda_opt)
autoplot(fit_naive)
```
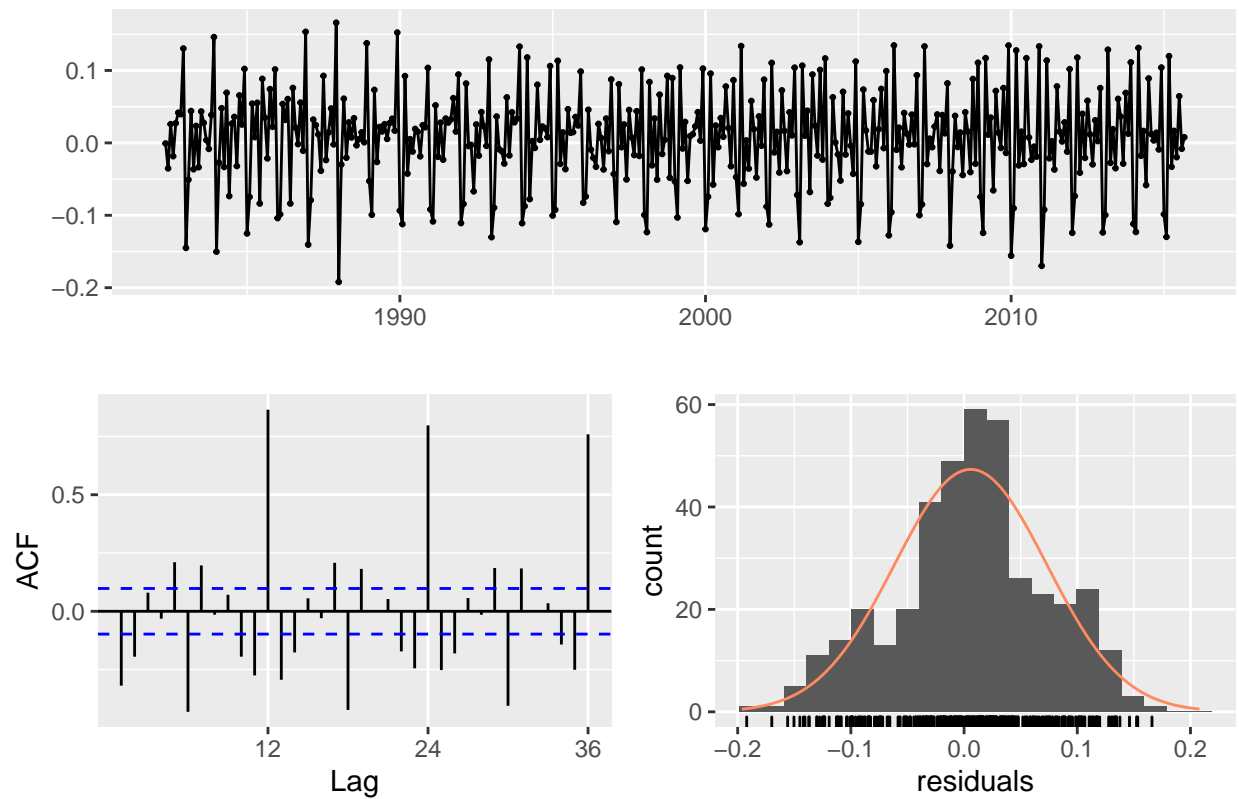
## Forecasts from Naive method



**b.**

```r
checkresiduals(fit_naive)
```

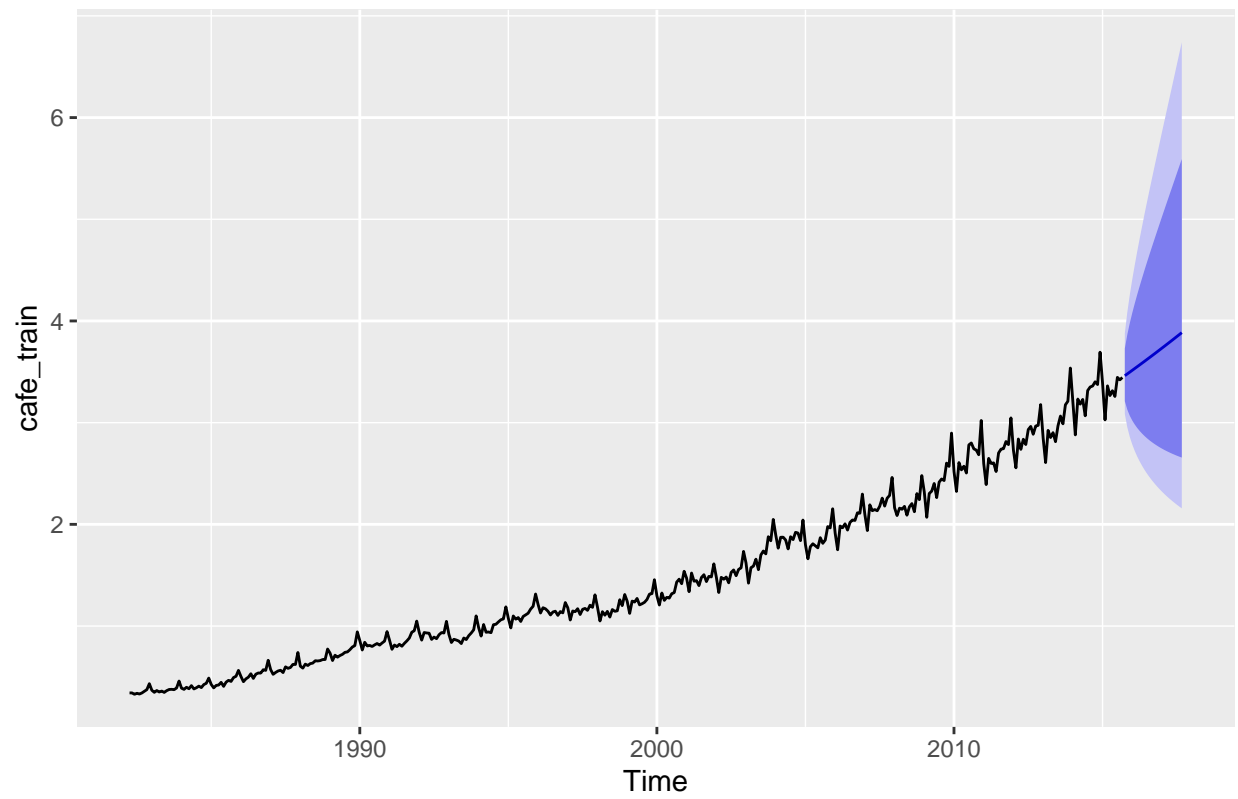## Residuals from Naive method



```
## 
##  Ljung-Box test
## 
## data:  Residuals from Naive method
## Q* = 999.66, df = 24, p-value < 2.2e-16
## 
## Model df: 0.   Total lags used: 24
```

**Drift Method**

**a.**

```
fit_rwf <- rwf(cafe_train, h = 24, drift = TRUE, lambda = lambda_opt)
autoplot(fit_rwf)
```

## Forecasts from Random walk with drift



**b.**

```
checkresiduals(fit_rwf)
```

## Residuals from Random walk with drift
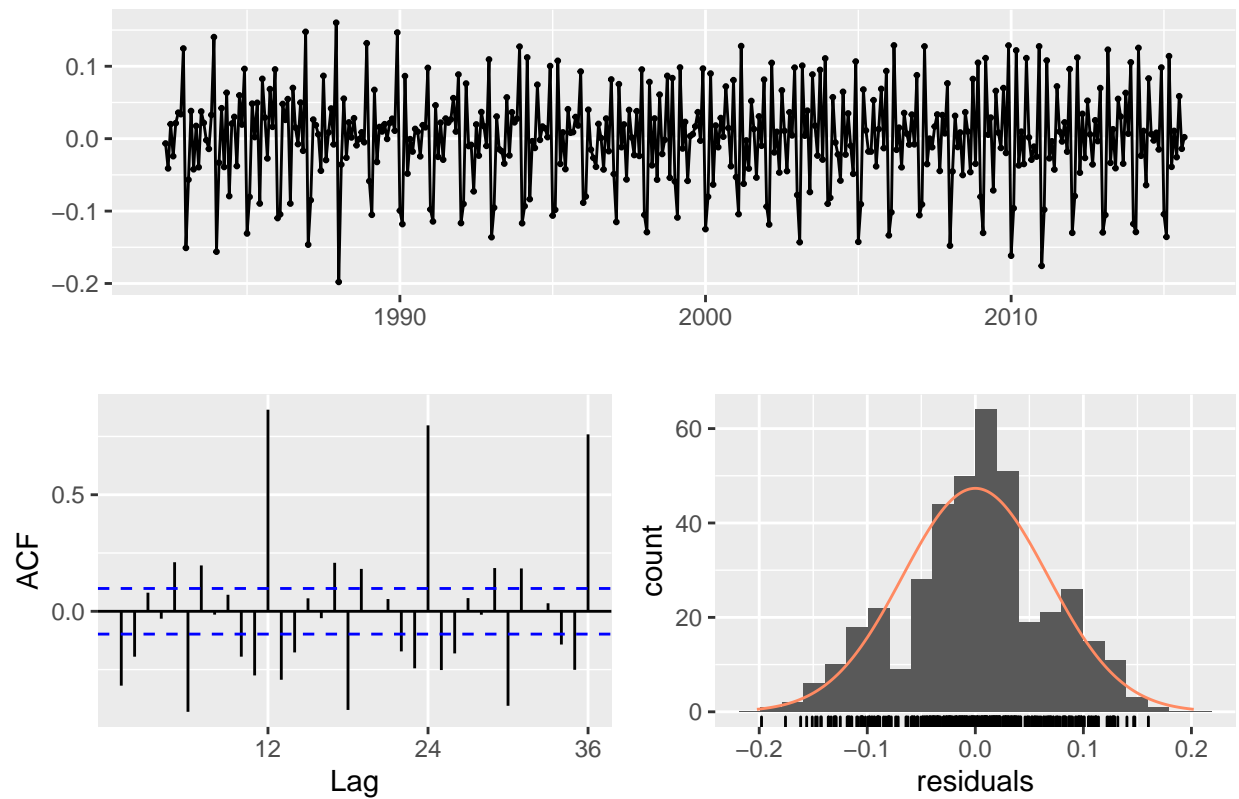


```
##
##  Ljung-Box test
##
## data:  Residuals from Random walk with drift
## Q* = 999.66, df = 23, p-value < 2.2e-16
##
## Model df: 1.   Total lags used: 24
```

**Seasonal Naive Method**

**a.**

```r
fit_snaive <- snaive(cafe_train, h = 24, lambda = lambda_opt)
autoplot(fit_snaive)
```
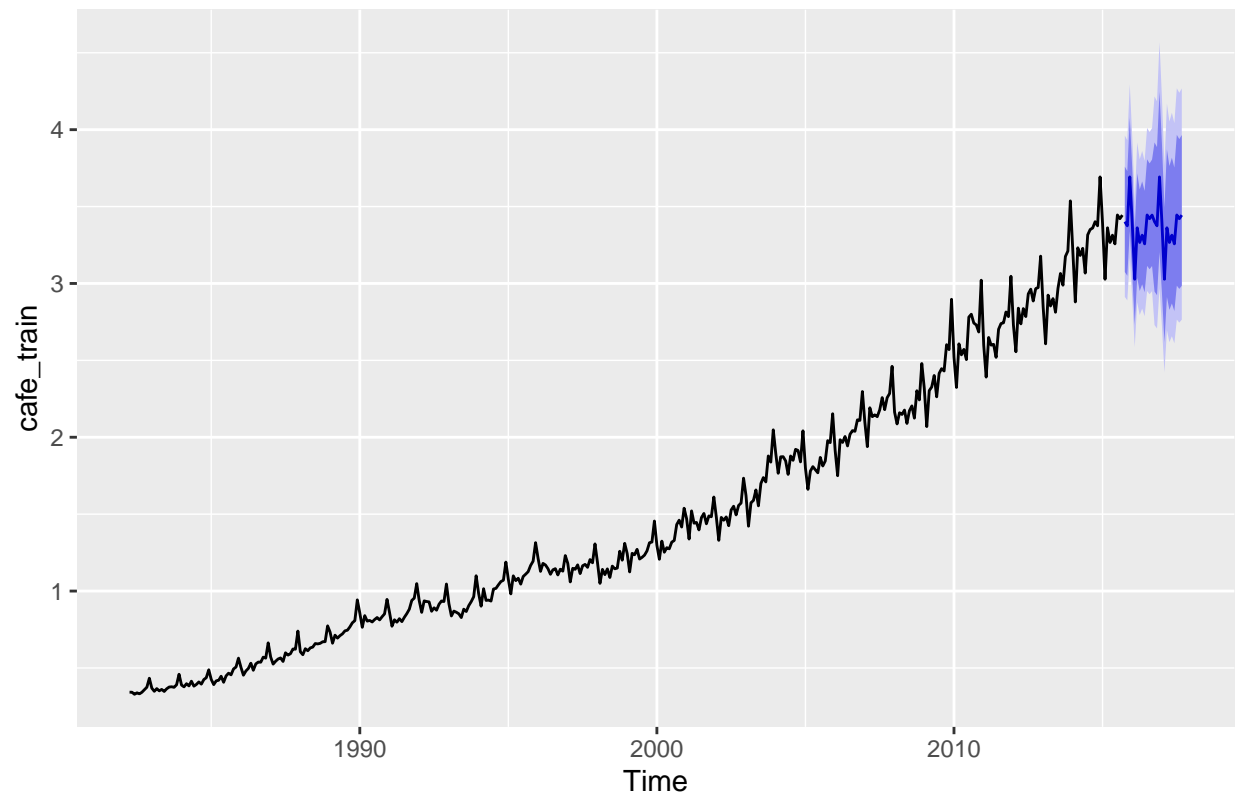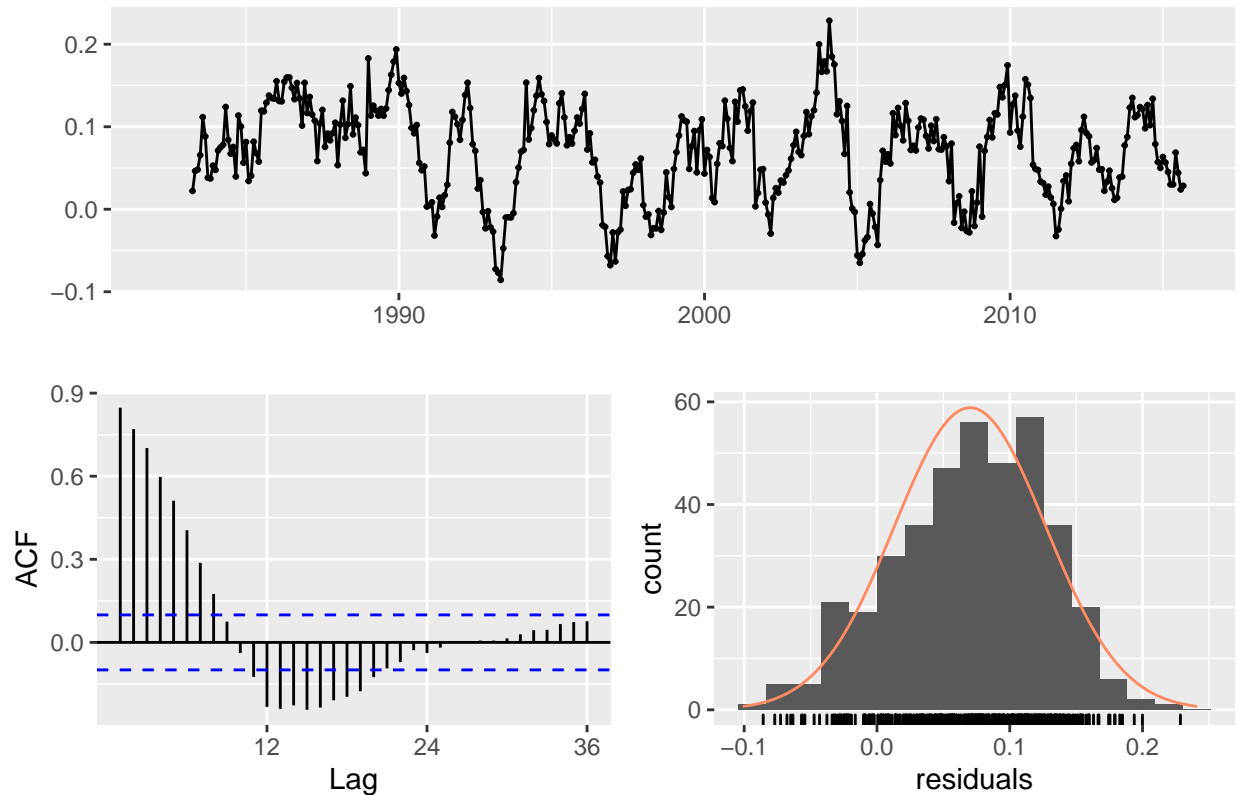
Forecasts from Seasonal naive method

**b.**

```
checkresiduals(fit_snaive)
```

## Residuals from Seasonal naive method



```
## 
##  Ljung-Box test
## 
## data:  Residuals from Seasonal naive method
## Q* = 1247.9, df = 24, p-value < 2.2e-16
## 
## Model df: 0.   Total lags used: 24
```

**Comments**

The drift and naive methods have the widest prediction intervals out of the three, with the seasonal naive method having a narrower prediction interval. When looking at the residual diagnostics, all three methods have residuals that are normally distributed. However, the seasonal naive method's residuals do not have a mean of 0. Additionally, looking at the ACF plots for all three methods, we can see that there is not this "randomness" to the autocorrelation values. There are constant repititions of ACF values outside the confidence bands (indicated by blue lines) for the naive and drift methods. The ACF plot for the seasonal naive method also has several values outside the bands and we see a decreasing, then increasing, trend.

With all of these diagnostics combined, they indicate the residuals do not follow the "white noise" pattern. We should not trust the prediction intervals for this reason, as our assumptions are broken.

**3.**

**a.**

```
cafe_test <- window(cafe, start = c(2015, 10))
# naive method
accuracy(fit_naive, cafe_test)
```

```
##                       ME      RMSE        MAE       MPE     MAPE      MASE
## Training set 0.007735661 0.1107577 0.07576559 0.3586388 5.169650 0.7412365
## Test set     0.130454167 0.2244039 0.18103750 3.3970250 4.953728 1.7711419
##                    ACF1 Theil's U
## Training set -0.3347359        NA
## Test set      0.3088414  1.045942
```

```
# drift method
accuracy(fit_rwf, cafe_test)
```

```
##                        ME      RMSE        MAE        MPE     MAPE      MASE
## Training set -0.0004135923 0.1106912 0.07540539 -0.2111715 5.141544 0.7377126
## Test set     -0.0952577109 0.2048671 0.16318479 -2.8663301 4.623668 1.5964837
##                    ACF1 Theil's U
## Training set -0.3361087        NA
## Test set      0.2978444 0.9722511
```

```
# seasonal naive method
accuracy(fit_snaive, cafe_test)
```

```
##                      ME      RMSE       MAE      MPE     MAPE      MASE
## Training set 0.09262795 0.1312097 0.1022151 6.531361 7.247775 1.000000
## Test set     0.20829583 0.2266826 0.2082958 5.766563 5.766563 2.037818
##                   ACF1 Theil's U
## Training set 0.8662941        NA
## Test set     0.7237032   1.05382
```

The RMSE, Root Mean Squared Error, is the square root of the Mean Squared Error and is scale dependent. The MAPE, Mean Absolute Percentage Error, is calculated by taking the mean ratio between the absolute values of the residual and observed values, then multiplying it by 100. This is another measure of forecast accuracy but is scale independent.

Looking at the measures of accuracy for the test set, overall the drift method performs the best, as shown through the RMSE (0.205) and MAPE (4.624). The RMSE for the both the seasonal naive and the naive methods are both very similar, 0.224 and 0.227, respectively. However, the naive method actually has a lower MAPE than does the seasonal naive method. Overall, the drift method does appear to be more accurate in forecasting than the other two methods.

b.

# Problem 2

1.

```r
set.seed(1)
autos <- c()
for (i in seq(1,500)){

  # creating ith white noise series
  series <- data.frame(1, 0)
  for (i in seq(2, 100)){
  series[i,] <- c(i, rnorm(1, 0, 1))
  }

  # obtaining r_1 for the ith white noise series
  ts_obj <- ts(data = series[2])
  autos <- c(autos, ggAcf(ts_obj)$data[1,3])
}

# sampling distribution of r_1, mean, variance, and expected variance
hist(autos, main = "Sampling Distribution of r_1", xlab = "r_1")
```

## Sampling Distribution of r_1
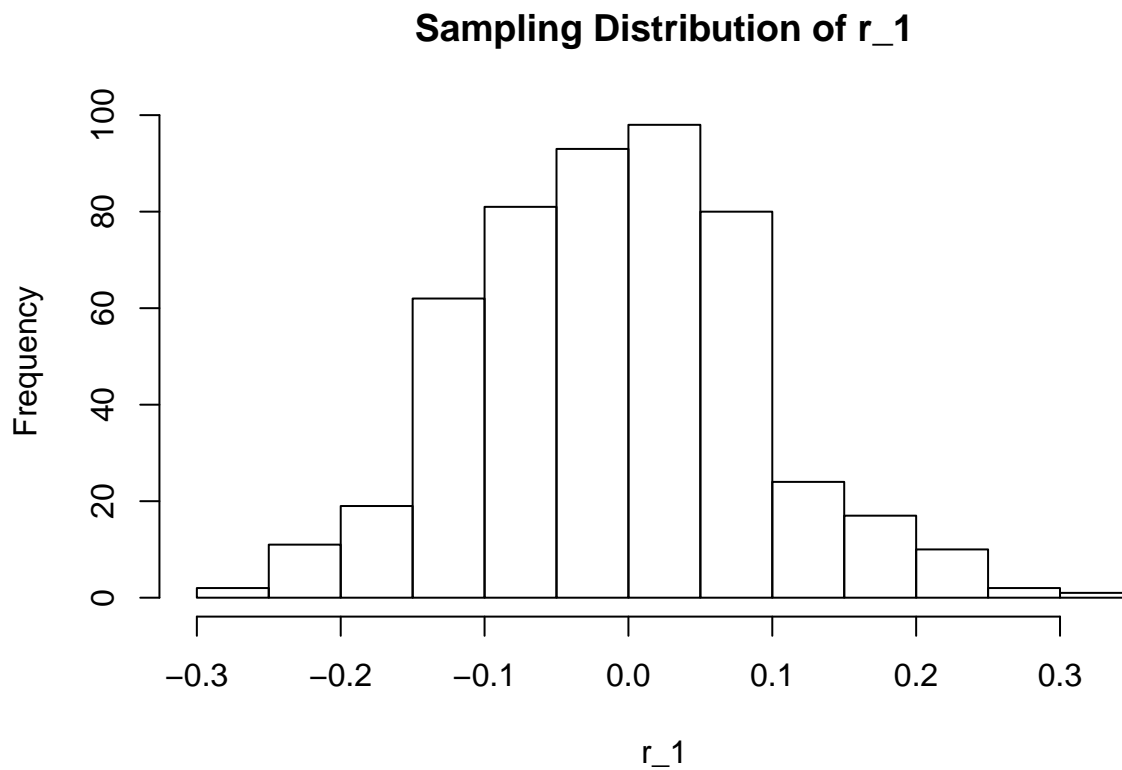
```r
mean(autos)
```

```
## [1] -0.00898618
```

```r
var(autos)
```

```
## [1] 0.009468347
```

```r
1/100
```

```
## [1] 0.01
```

Above is the samplign distribution of $r_1$. Theoretically, we would expect $r_1$ to follow a normal distribution with mean 0 and variance $1/T = 1/100 = 0.01$. The above sampling distribution has a mean of -0.00898618 and variance of 0.009468347. This is extremely close to what we would expect. If the number of samples in the sampling distribution increased, the mean and variance would continue to converge to the expected values.

**2.**

**a.**

Non-stationary. The mean is increasing and variance is non-constant.

**b.**

Non-stationary. The mean is increasing, although the variance is constant.

**c.**

Non-stationary. The mean is constant, but there is some form of seasonality to it.

**d.**

Stationary. There does not appear to be any obvious trend or seasonality. Ther variance is pretty constant.

**e.**

Non-stationary. There appears to be some form of seasonality within each year.

**f.**

Stationary. There does not appear to be any trend and there may be a slgiht change in variance, but it appears to be negligible.