Time Series Modeling Project

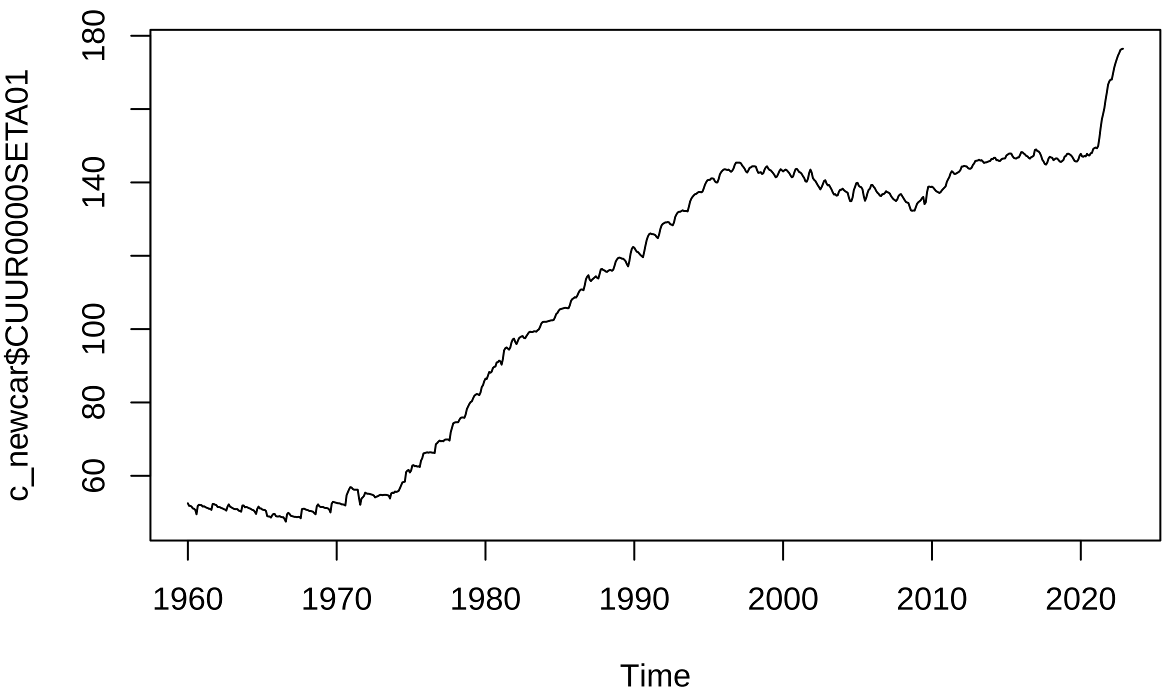
Kyle Wu

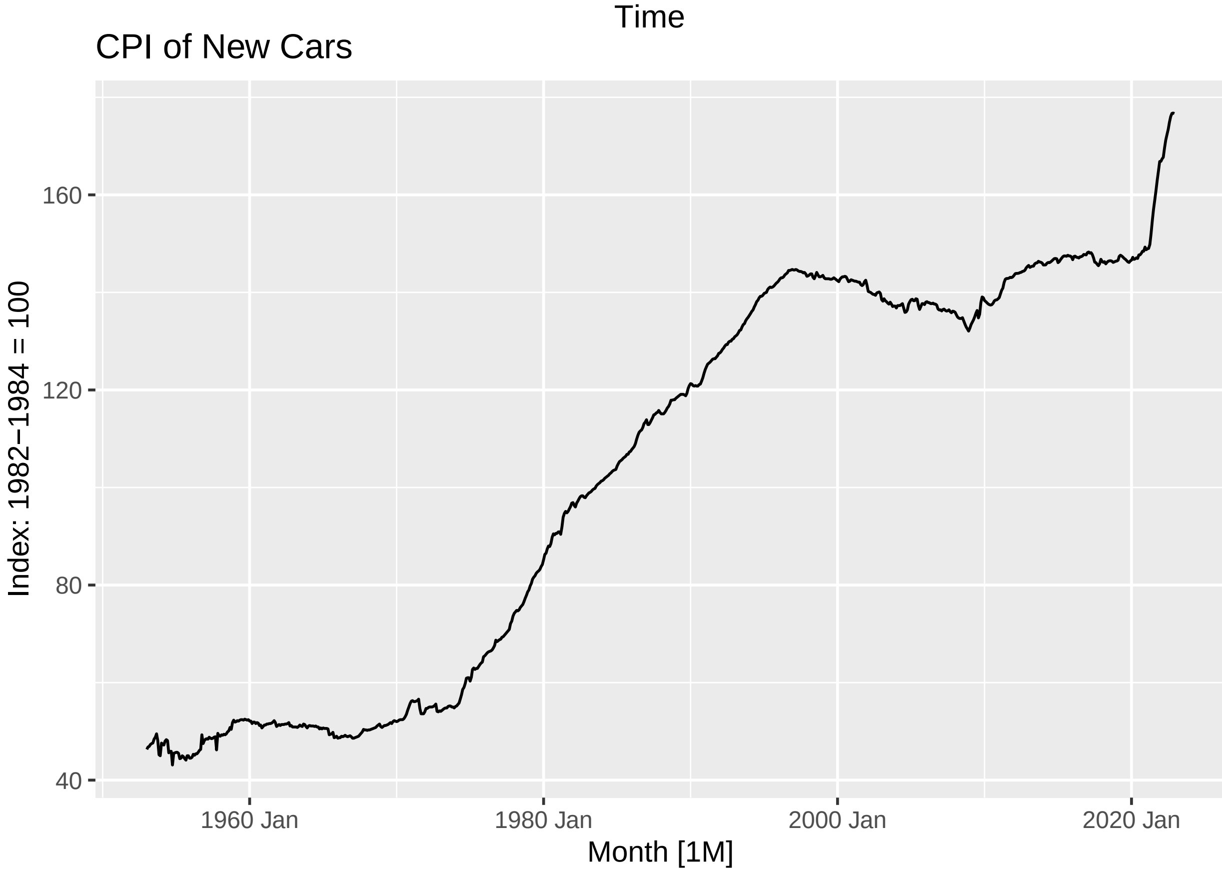
2023-01-30

```
## Rows: 889 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (4): CUSR0000SETA01, DAUPSA, TOTALNSA, CUSR0000SETB01
## dbl  (1): MPRIME
## date (1): DATE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

R Markdown

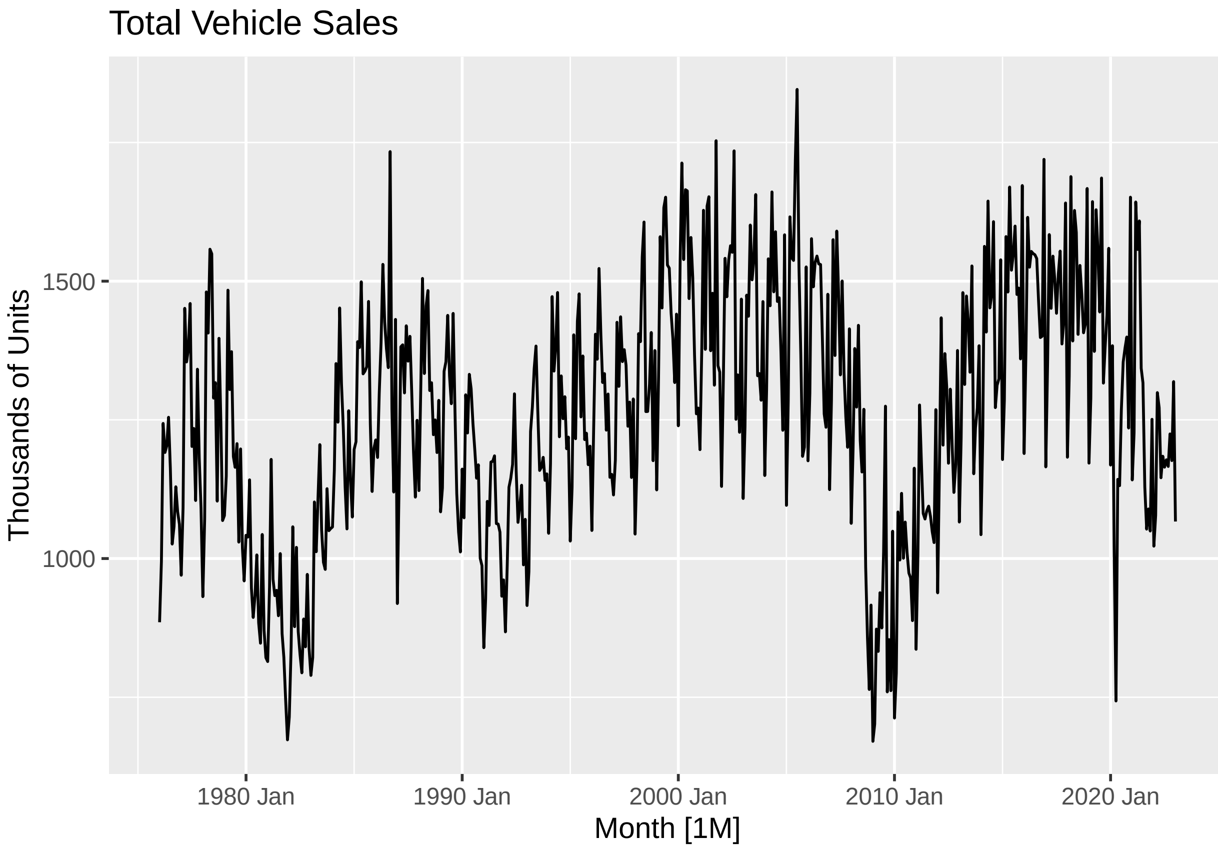
New Car CPI





Prime Rate Loans

Total Vehicle Sales



Domestic Auto Production

GAS CPI

Data Application

In the United States, one of the main modes of transportation is the automobile. To the average consumer, it has seemed that the new car has been slowly getting out of reach, with the average price of a new vehicle currently sitting around $49,500 ("No End in Sight: New Vehicle Transaction Prices End 2022 at Record Highs, According to New Data from Kelley Blue Book" 2023). Despite the high prices of vehicles for many Americans a car is not only a luxury, but a necessity, and many citizens find themselves shelling out a large portion of their paychecks for their transportation. Even when individuals opt to purchase used cars, they are still often faced with prices that would have seemed exorbitant not that long ago.

Since this is the case, studying the United States car market over time will allow us to gain useful knowledge that will be of significance not only to the average consumer, but also for economists trying to understand what trends the American auto market may be facing going forwards and what factors influence automotive sales. Past research into the American auto market have been vital to our understanding of the forces driving the auto market. For example, it is well known that the chip shortage that occurred as a result of COVID-19 shutdowns, among other reasons led to a chip shortage that has in many ways created problems for the world economy ("Inflation and the Auto Industry: When Will Car Prices Drop" 2022). Since cars now heavily rely on computers to work, this resulted in many manufacturers around the world decreasing production projections, which decreased vehicle production, and partially led to the rapid rise in vehicle prices. However, if we look at production figures, we can see that the domestic production of cars had been following a decreasing trend since the 90s, so researchers at Federal Reserve Economic Data (FRED) found that it is hard to say if COVID was fully responsible for the decreased production, or if it would have happened regardless ("Long-Term Trends in Car and Light Truck Sales" 2021). Research by FRED also indicated that despite the increase in population since the mid 1970s, the total number of vehicles sold has remained relatively flat aver the past few decades ("What's Been Drivin the Rise in Auto Prices Since COVID" 2022).

Looking at the data offered by the Federal Reserve could allow us to answer even more questions regarding the American auto market. For example, we could try to understand if it is likely that american automakers would have decreased their production numbers even without the disruptions brought about by COVID or if COVID led to new trends. If we take into account other economic factors, such as interest rate or gas prices, we can then try to measure what economic factors may most affect the sale of motor vehicles. Using the data we obtained and after determing factors that determine automotive sales, we can then create a forecast to determine how each factor relating to the automotive industry will change in the future. For example, we can try to answer the question of whether it is likely new vehicles will continue facing inflation or if it might become stable in the near future. Besides looking at the various factors individually, we can also look at the auto market holistically, asking what the future may be in terms of vehicle purchases in the United States and is it likely that vehicle purchases return to pre-COVID levels. For the average consumer, the questions that will be answered will allow them to perhaps better plan for the expenditure that comes with the purchase of a new car.

Furthermore, studying time series data of vehicles can allow us to better understand how or if certain policy changes may change vehicle prices or purchasing behavior. For example, we could potentially find time periods with varying federal funds rates, which influence bank prime loan rates to see if this changed the overall behavior of consumers.

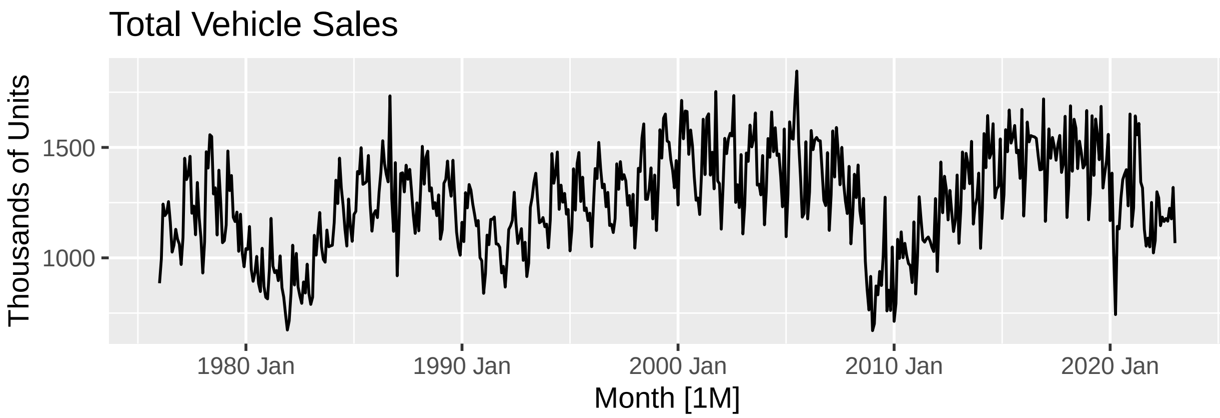
Gathering all this data about the American auto market would then allow us to broadly gain an understanding not only of factors affecting vehicle sales, but also of the health of the American economy due to the fact that vehicles are often the second most expensive possessions of individuals, second only to homes. Increased purchasing of vehicles would indicate that the American has been healthy and following a positive trend, whereas decreased vehicle purchases may indicate that the economy had been following a general downwards trajectory.

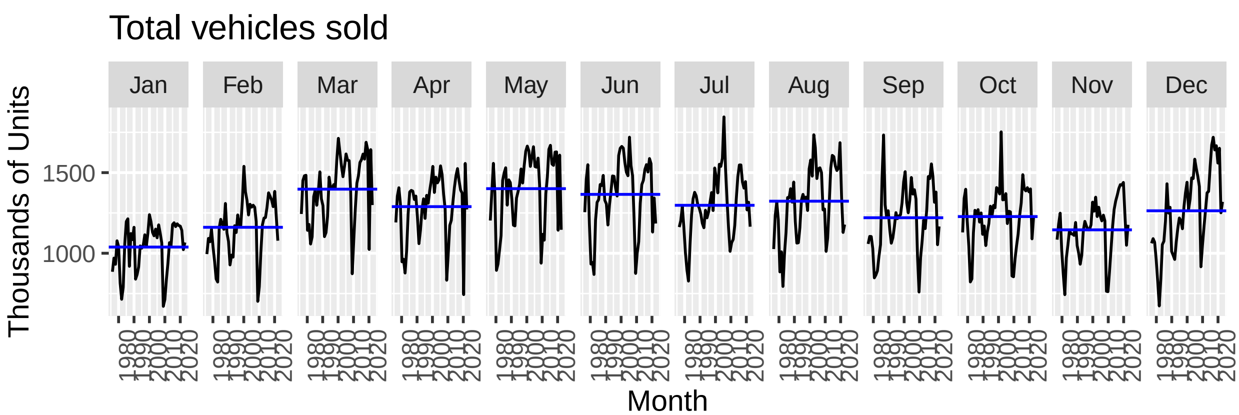
Analysis of Empirical Properties

In all cases, the data I selected came from the Federal Reserve Economic Data database and all variables selected were recorded on a monthly basis and input into a format that was very neat and effective for time series analysis. The variables I have chosen are Total number of vehicles sold, new vehicle consumer price index, domestic auto production, fuel price index, and the bank prime rate. In this study, I will use total number of vehicles sold as the gauge of the american auto market, and the other variables will be used as predictors.

We will first analyze the variables individually before talking about all the factors as they may relate to projecting future car sales.

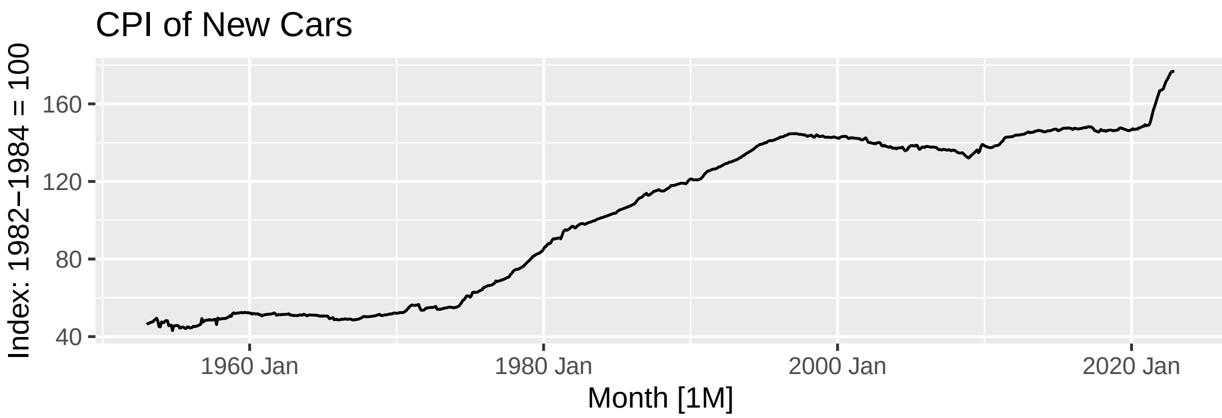
The first variable we will look at is total vehicle sales.

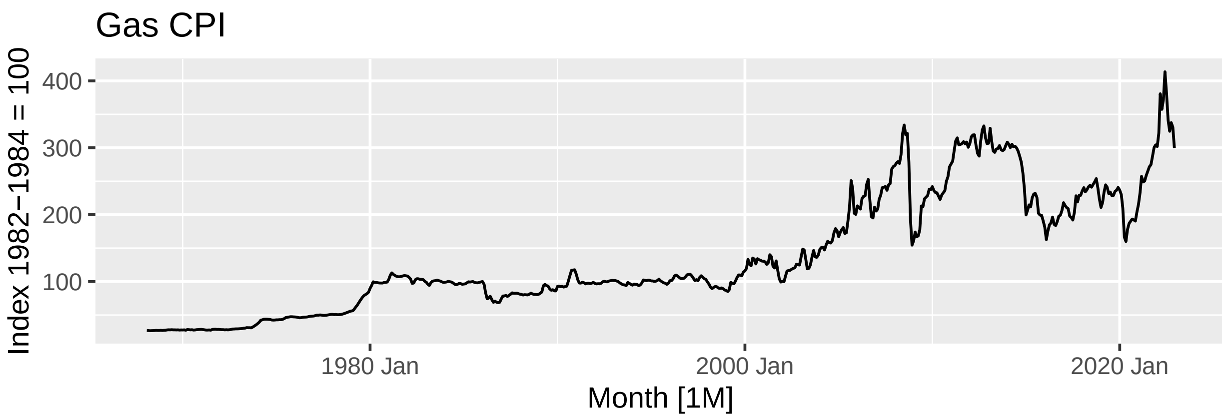




Our data for vehicles sold starts from January 1976, is recorded monthly, and does not have any missing values.The data was collected from the U.S. Bureau of Economic Analysis, which has formatted the data in a easily usable database. From the subseries plot, we see that the data definitely follows a seasonal pattern, as demonstrated by the fluctuating mean lines based on the month of the year. Additionally, from the scalloped shape of the plot, we can see that the data follows a seasonal trend of peaks and troughs every twelve months. Logically this makes sense, as it is well known that vehicle sales usually increase during the summer and tend to decrease during the winter. From the data we can determine the number of vehicles sold over the last 5 decades and whether there has been a general trend in vehicle sales.

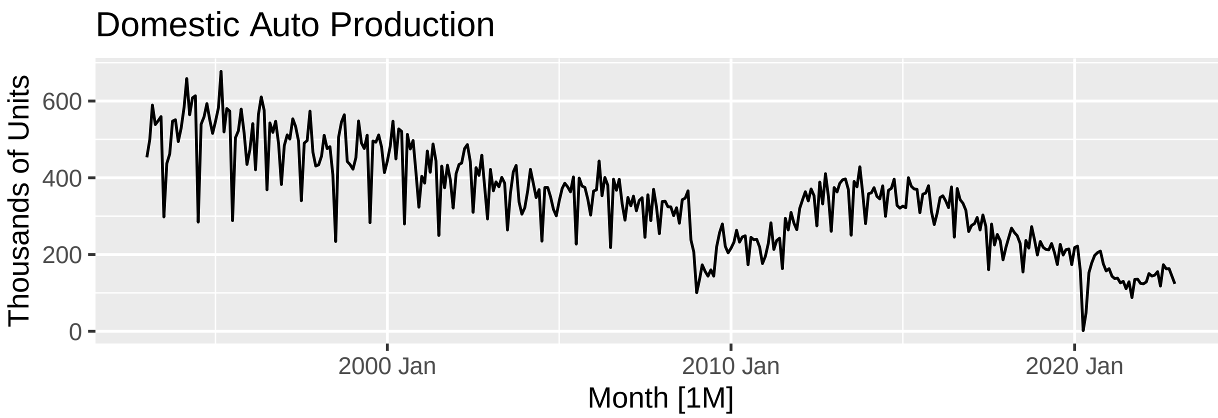
We can now look at the consumer price index of new vehicles in the United States.

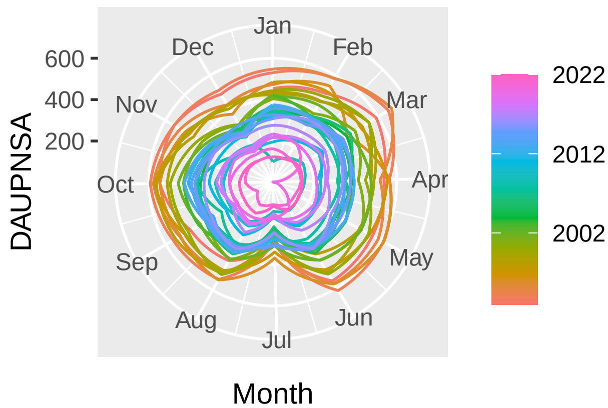


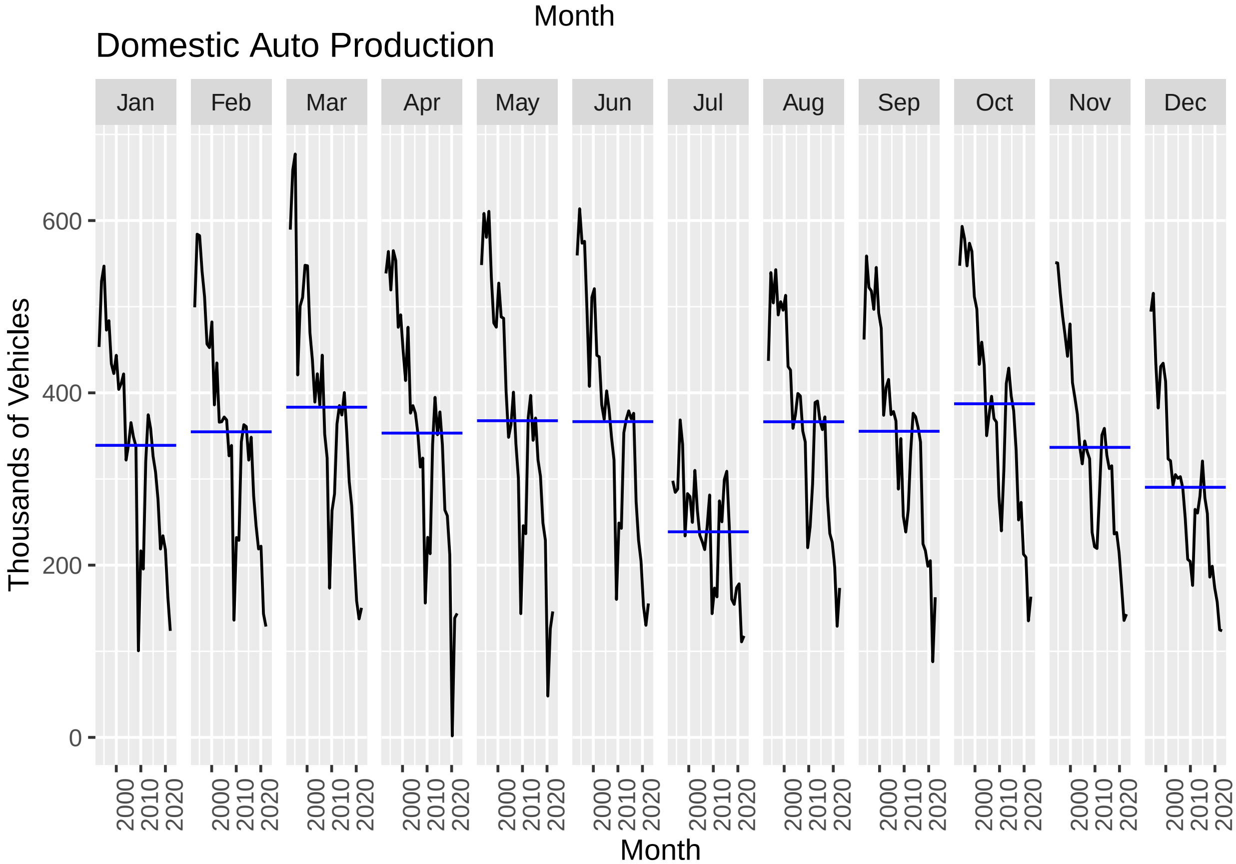


The data for the consumer price index measure of new cars begins in January 1953 and is recorded monthly. This dataset has already been seasonally adjusted so that will not be something that I will have to worry about later. It is worth mentioning that CPI should not be regarded as the price of a vehicle, but is a measure of how much more or less an item cost relative to the base year (1982-1984 in this case). In essence, this CPI measures the spending power consumers have over the good. Overall, there does not appear to be any sharp fluctuations in the data but we do have a couple periods of relatively rapid vehicle price inflation, namely from the 70s to the late 1990s and then again after the arrival of COVID-19 in 2020. It is interesting to see that from the late 1990s to 2020, the CPI value of cars did not change drastically.

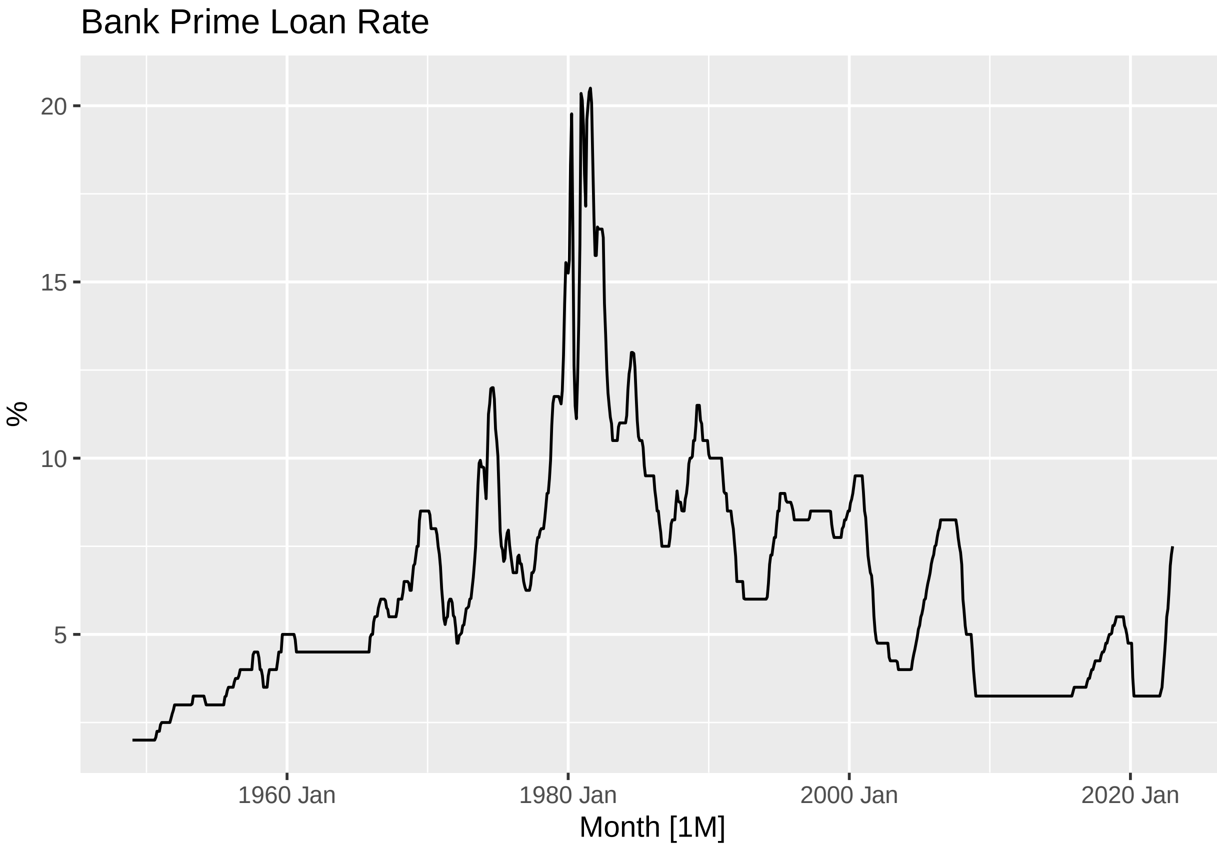
Records for the price index of gasoline start from February 1968 and are recorded monthly by the U.S. Bureau of Labor Statistics with no missingness present. From the data, we see that there has been a general increase in the price of gasoline from 1968 until the early 2000s before prices seem to level off, but with high volatility. We also see that following COVID, there was a large spike in gasoline prices, which has since gradually come down. This data will allow us to understand what trends are present in terms of gasoline prices and whether gasoline prices have an effect on the total number of vehicles sold. In other words, do high gasoline prices lead to less consumers buying vehicles?







For Domestic Auto Production, we have monthly recorded data from the U.S. Bureau of Economic Analysis with no missingess present in the data. Overall, we appear to see volatility month-to-month and pretty consistent seasonal patterns with July being a month with consistently low production numbers in relation to other months and March and October having relatively high production numbers. This data alone can answer a couple questions about the United States auto market. First of all, we can answer if the COVID induced supply shortages drastically impacted U.S. vehicle production numbers or if, as suggested by FRED COVID simply highlighted a long-term trend that would have occurred regardless. Using this data will then also allow us to make forecasts on what the future outlook may be for domestic auto production.



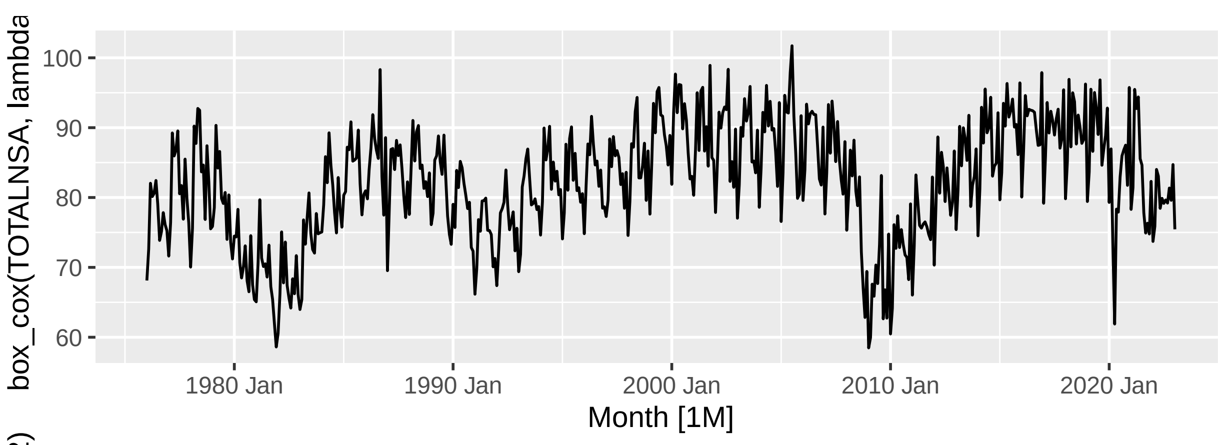
The Bank Prime loan rate is set in relation to the federal funds rate set by the federal reserve and we have data going back to January 1949. The rate is often set based on other determinants of the market so we may not be able to generate too many insights from it individually. However, if we use the bank prime rate in addition to the other factors we discussed earlier, we may be able to create a regression model that will allow us to understand what economic factors influence the car purchasing decisions of Americans. Additionally, studying all the data together will allow us to understand if COVID-19 disruptions altered the purchasing decisions of Americans or if there were vehicle market trends that would have likely occurred regardless of the pandemic.

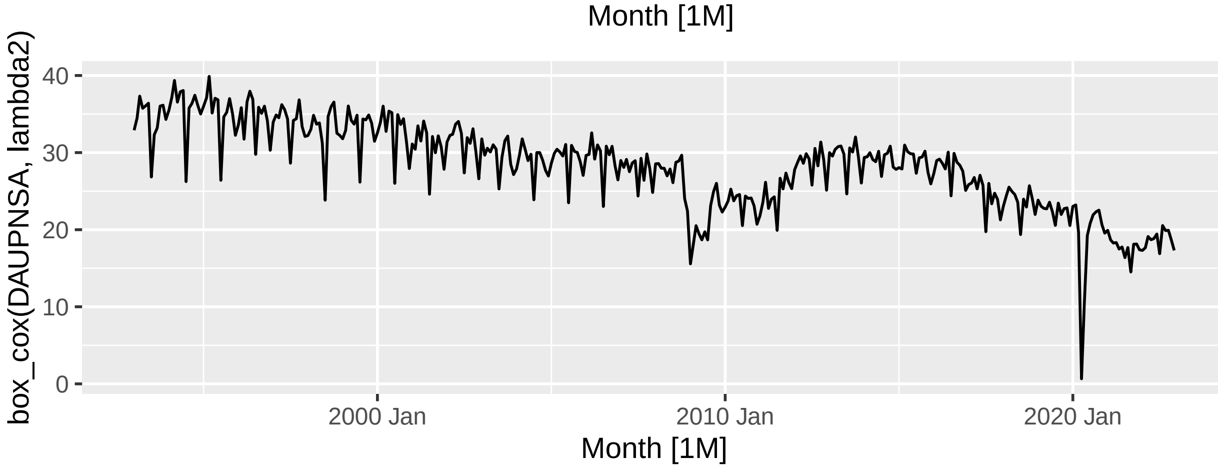
Models for Data Fitting

Throughout this case study, we will aim to fit models that allow us to analyze and forecast future values of our time series using the Box-Jenkins Methodology. There are three main stages to setting up a Box-Jenkins model. The first step is to examine the data and to see which parts of the ARIMA process appear to be the most appropriate to apply. The second step is to estimate the parameters of the chosen models. Dinally, the third step is diagnostic checking where we examine the residuals from the fitted morel to see it they appear to be sufficient. If the model we originally chose is insufficient, then we should try other models until a sufficient model is found.

The first step to fitting our model is to see if any of our time series need to be transformed or adjusted, which often allows us to analyze simpler time series. Adjusting the data will allow us to make patterns more consistent across our data, which will allow us to better model the data, and will lead to better forecasts. For the metrics I used, the only ones that may need transformations due to variations in seasonality are those for `daupnsa` and `totalnsa`. Since these two series show variations that change with the level of the series, we can use Box-Cox transformations to stabilize the variance. Box-Cox transformations depend on a parameter λ and can be defined as follows:

$$y(\lambda) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(y), & \lambda = 0 \end{cases}.$$





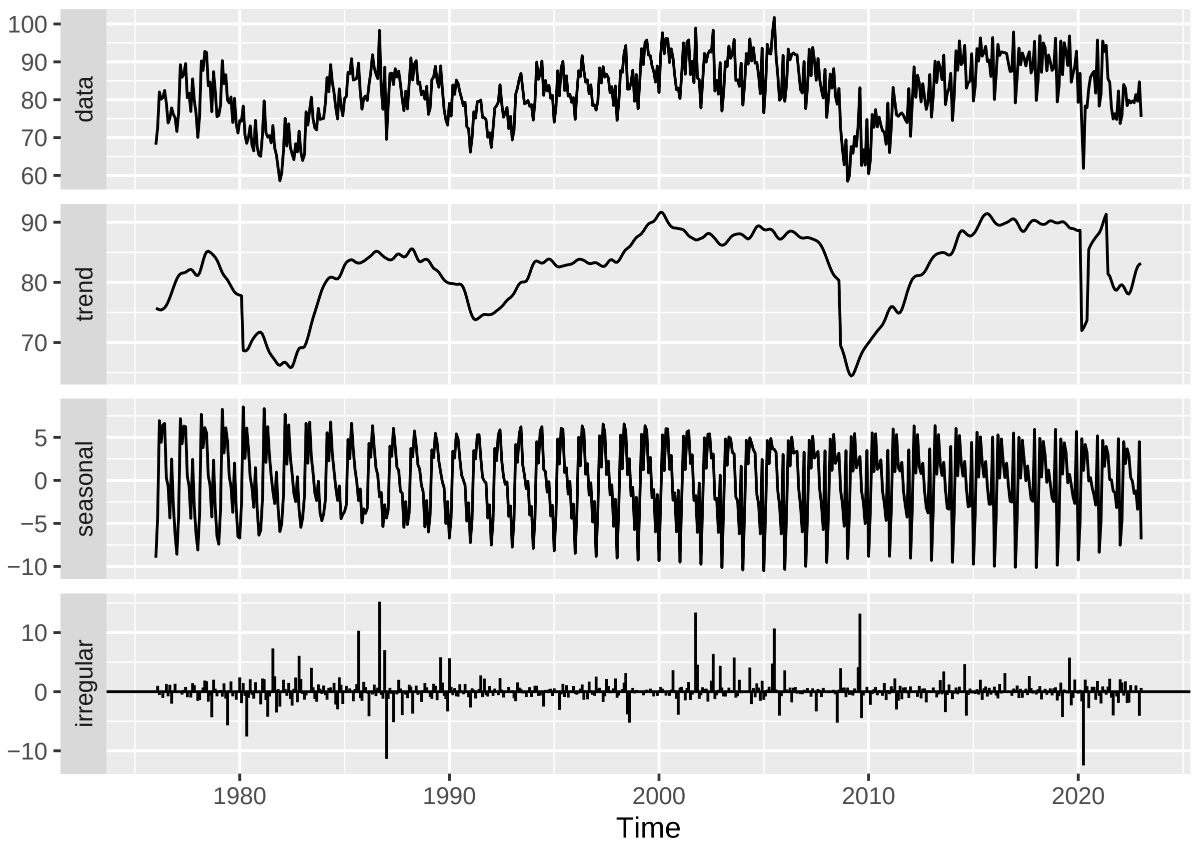
After running the calculations, we find that the appropriate λ value for total auto sales was approximately 0.53 and the appropriate λ value for domestic auto production was 0.45. From the plots that we created above, we also see that we were able to successfully adjust our data since it appears that for the most part, our seasonal effects have been made consistent over the range of out time series.

The next step to fitting our model is to decompose our data into the time series' separate components.

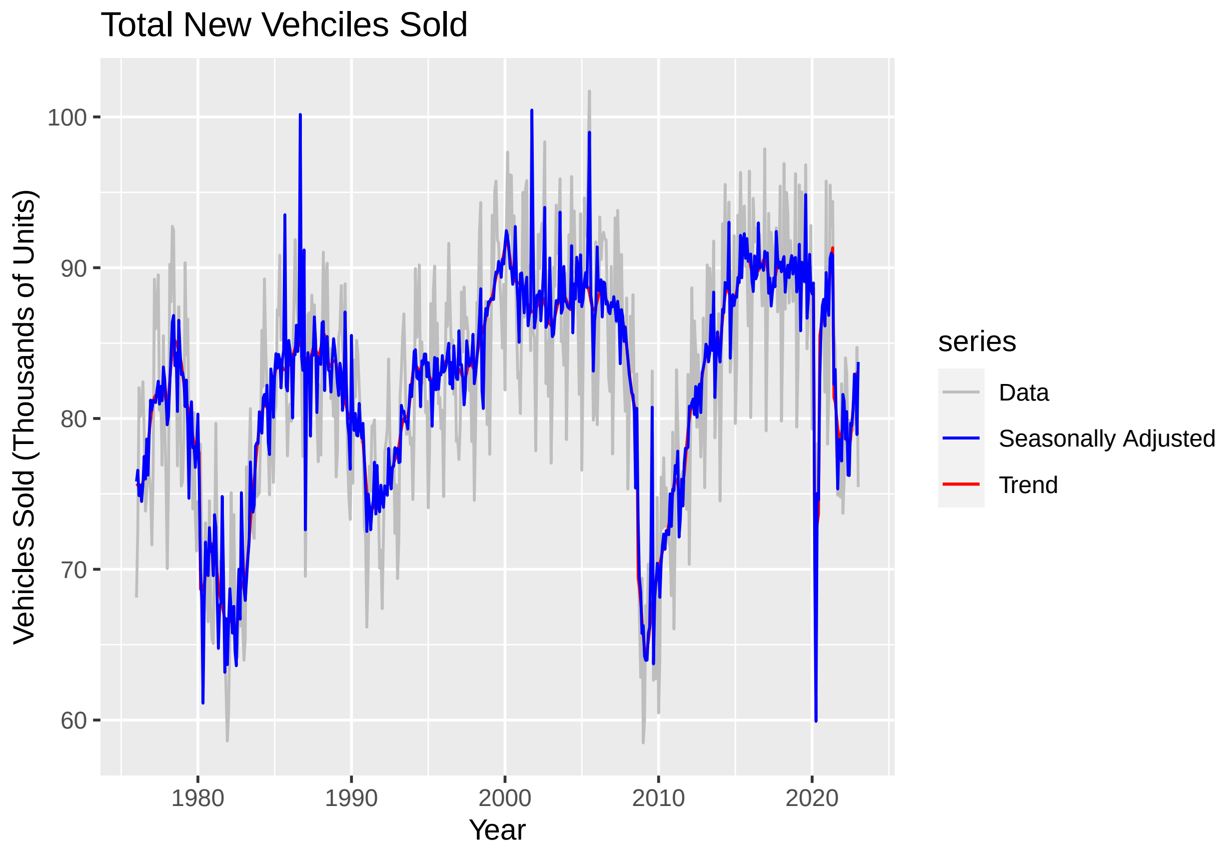
Depending on the type of data we have, there are two possible ways that we can decompose the data. The first possible case of data is a time series with a trend but no seasonal variations, which follows the form $X_t = \alpha + \beta t + \epsilon_t$, where α, β are constants and ϵ_t is a random error term with mean zero. Since our CPI data for gasoline prices and new car prices are both seasonally adjusted, this is the type of decomposition they will undergo.

The second possible case of data is a time series that contains both a trend, and seasonal variation. In this case, there are two possible cases of decomposition we can take. If we assume additive decomposition, then we can use $y_t = S_t + T_t + R_t$, where S_t is the seasonal component, T_t is the trend component, and R_t is the remainder component. The other possible case involves multiplicative decomposition which follows the form $Y_t = S_t \times T_t \times R_t$. We use the multiplicative case if the variation around the trend-cycle appears to be proportional to the level of the time series ("PSTAT 174/274: Time Series Part IV" 2023). The additive case would be if the magnitude of the seasonal fluctuations don't significantly vary over time. Since we used a box-cox transformation to stabilize our variance, we can use the additive case in our decomposition.

In our case, we will use an X-11 decomposition, which is commonly used by the U.S. Census Bureau (include how X-11 decomposition works). An example of X-11 decomposition is shown below



From the decomposition of this time series, along with the other time series, we can get a general idea of the trend and seasonality present within the plot, which will then allow us to better estimate and thus forecast the data moving forwards. Additionally, following the decomposition, we can then see how the trend and seasonally adjusted versions of our data fit with the original data.



For this case study, our goal is to eventually forecast future values of each of our time series data through the use of ARIMA models. The goal of ARIMA models is to describe autocorrelations in the data for forecasting (Rob J Hyndman 2021). When attempting to fit ARIMA models, we attempt to first remove any trend or seasonality present within the data so that we can create stationary time series, which will allow us to attempt to model the remaining residuals. Stationary time series are time series whose mean and variance is constant over time (Glenn, n.d.).

One way to convert non-stationary time series into stationary time series is the method of differencing, where we compute the differenece between consecutive observations in a time series and can be written as follows

$$y'_t = y_t - y_{t-1}$$

Differencing time series allows us to stabilize our means because it removes or reduces trend and seasonality. Furthermore, if our data still does not appear stationary, it may require second-order differencing, which follows the form

$$y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

Before we start differencing our data, we should first test if our data is already stationary, which would then allow us to go straight into further analysis. The test we will use for stationarity is the Kwiatkowski-Phillip-Schmidt_shin (KPSS) test. The KPSS test works by breaking up a time series down into the following decomposition

$$Y_t = r_t + \beta_t + \epsilon_t$$

where r_t is a random walk, β_t is the trend, and ϵ_t is the stationary error. The KPSS test also involves hypothesis testing with $$ H_0: Y_t is trend (or level) stationary \

H_1: Y_t is a unit root process $$ where H_0 and H_1 is the null and alternative hypothesis respectively. Setting our significance level at e can now test each time series to determine whether they will require differencing.

```
##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 6 lags.
##
## Value of test-statistic is: 1.4887
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739

##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 5 lags.
##
## Value of test-statistic is: 4.711
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739

##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 6 lags.
##
## Value of test-statistic is: 2.1478
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739

##
## #######################
## # KPSS Unit Root Test #
## #######################
##
## Test is of type: mu with 6 lags.
##
## Value of test-statistic is: 11.6058
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739

##
## #######################
## # KPSS Unit Root Test #
## #######################
```

```
##
## Test is of type: mu with 6 lags.
##
## Value of test-statistic is: 7.825
##
## Critical value for a significance level of:
##                 10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739

## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1      1.49        0.01

## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1      4.71        0.01

## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1      2.15        0.01

## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1      11.6        0.01

## # A tibble: 1 x 2
##   kpss_stat kpss_pvalue
##       <dbl>       <dbl>
## 1      7.83        0.01
```

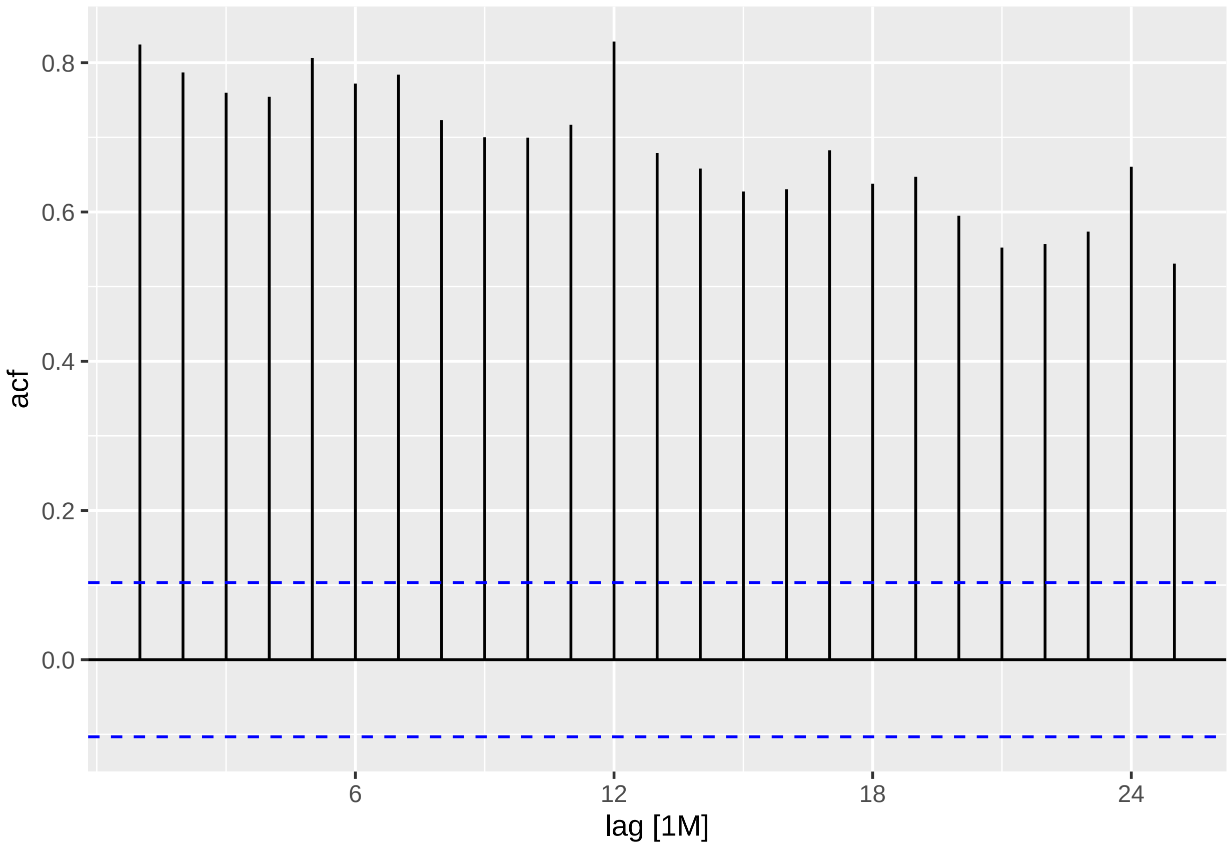
The value of the t-statistic for each test is as follows:

Test Statistic and Significant Values of Time Series

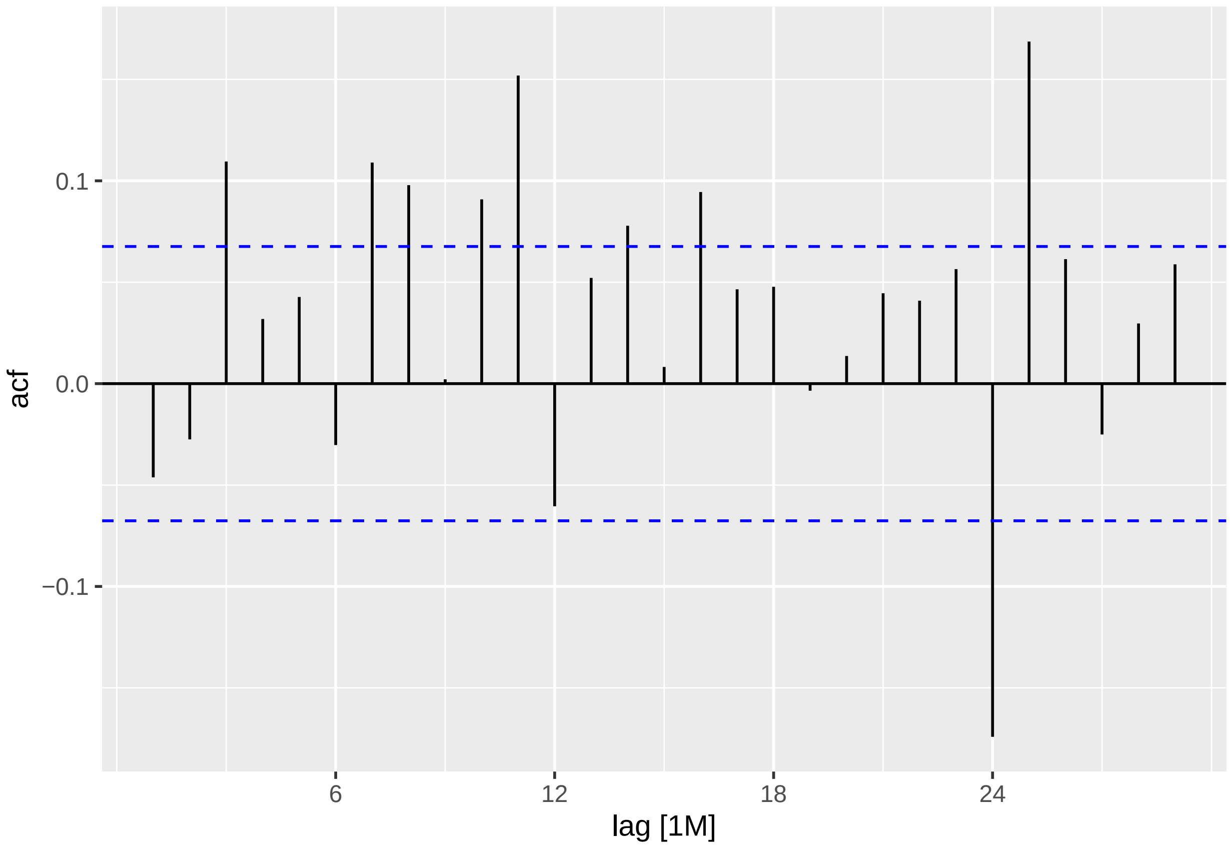
Time Series	Test-Statistic	P-Values
Total Vehicles	1.4887	0.01
Domestic Production	4.7110	0.01
Bank Prime Rate	2.1478	0.01
New Car CPI	11.6058	0.01
Gasoline CPI	7.8250	0.01

Since all of the time series have p-values that are less than 0.05, we can reject the null hypothesis and conclude that it is likely that all our time series are this suggests that differencing will be required in order to make our data stationary. We will demonstrate this process on our time series for new car CPI and for domestic auto production in order to show the process on seasonally adjusted and non-seasonally adjusted data respectively, but the process will be conducted on all time series in this case study.

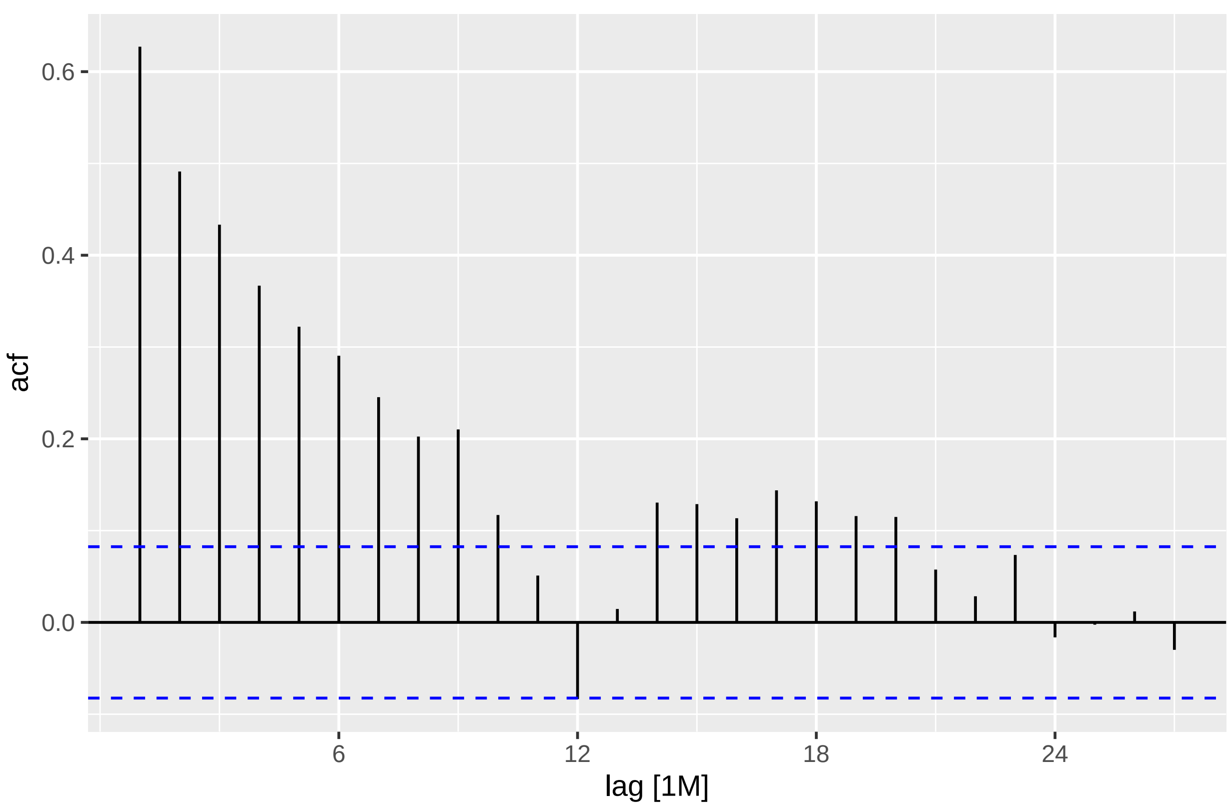
In the case of new car CPI the following procedure will first apply a differencing function and then plot the function to see if the differencing allowed the data to become stationary.



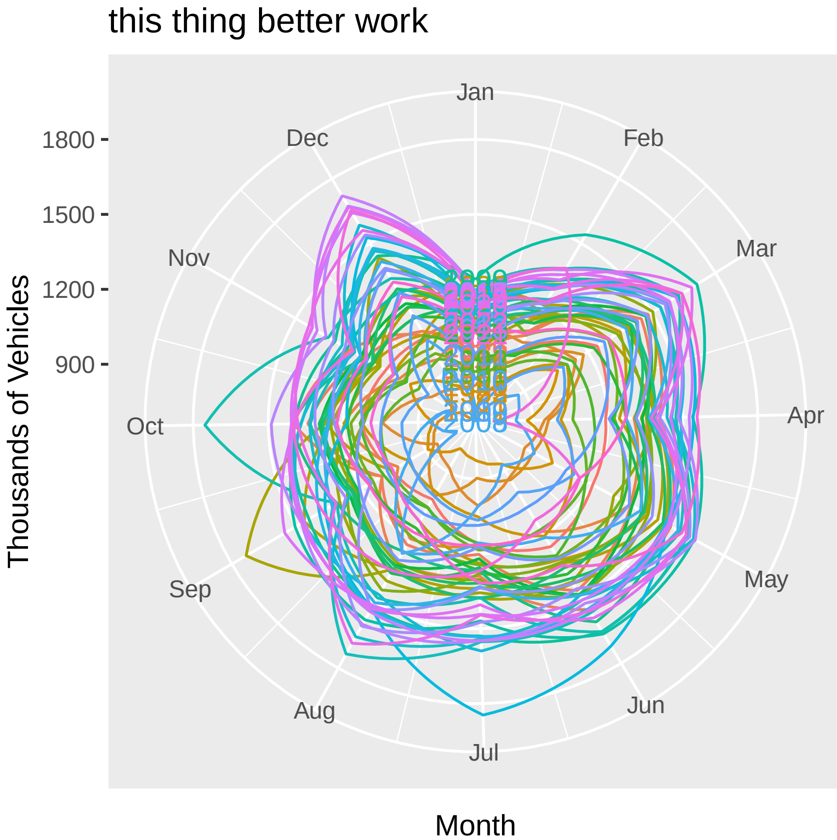

```
## # A tibble: 1 x 2
##   lb_stat  lb_pvalue
##     <dbl>      <dbl>
## 1    45.7 0.00000160
```

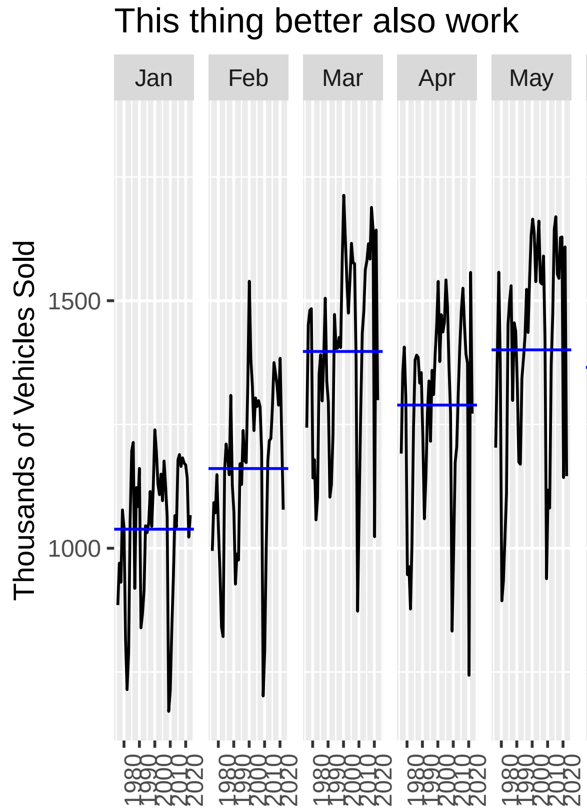


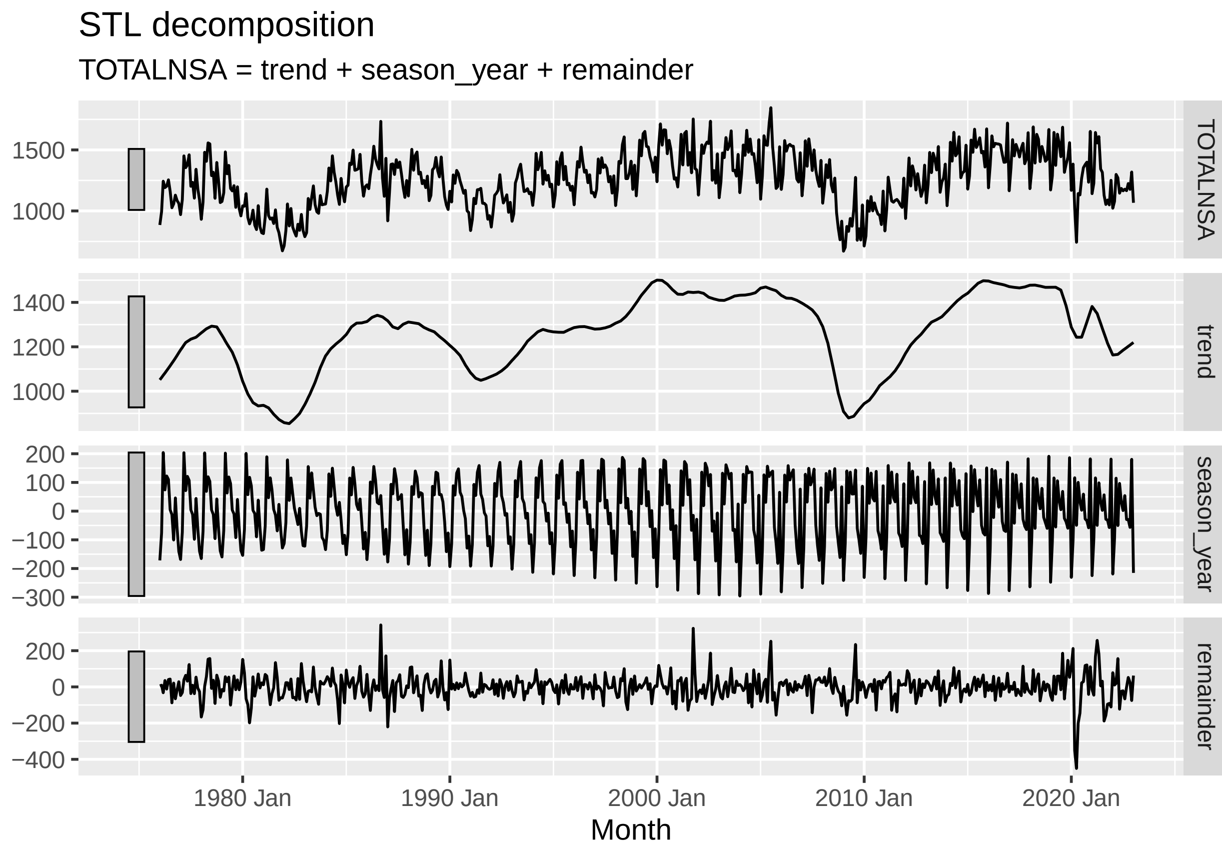
Total Vehicles Sold

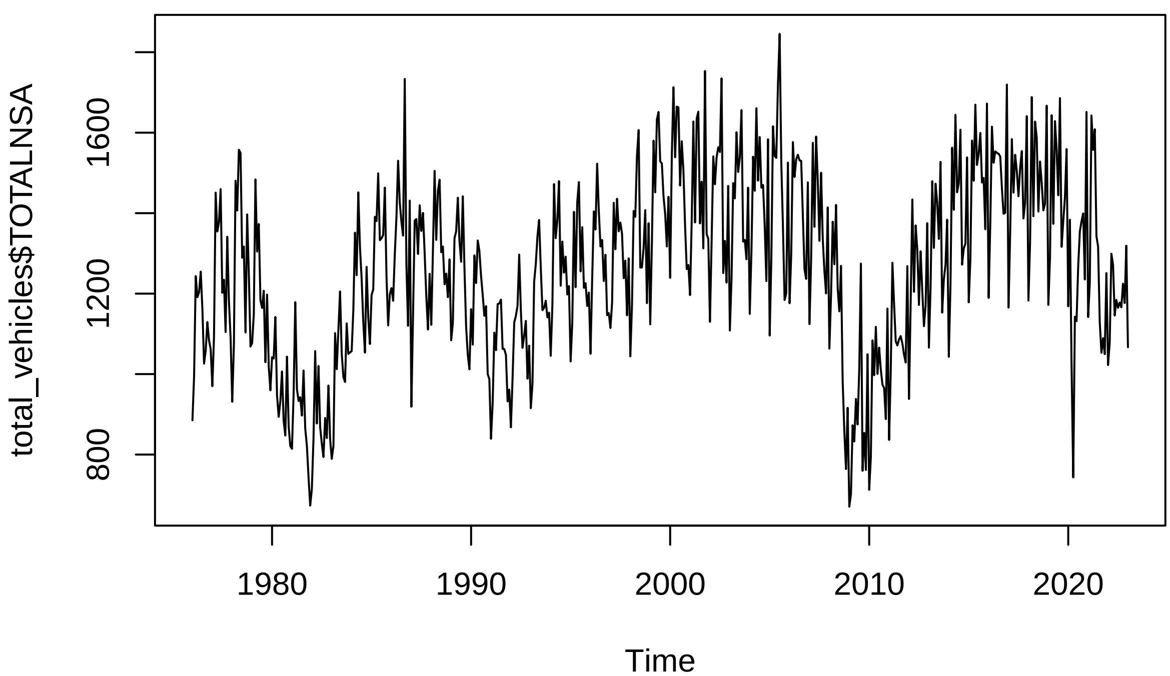


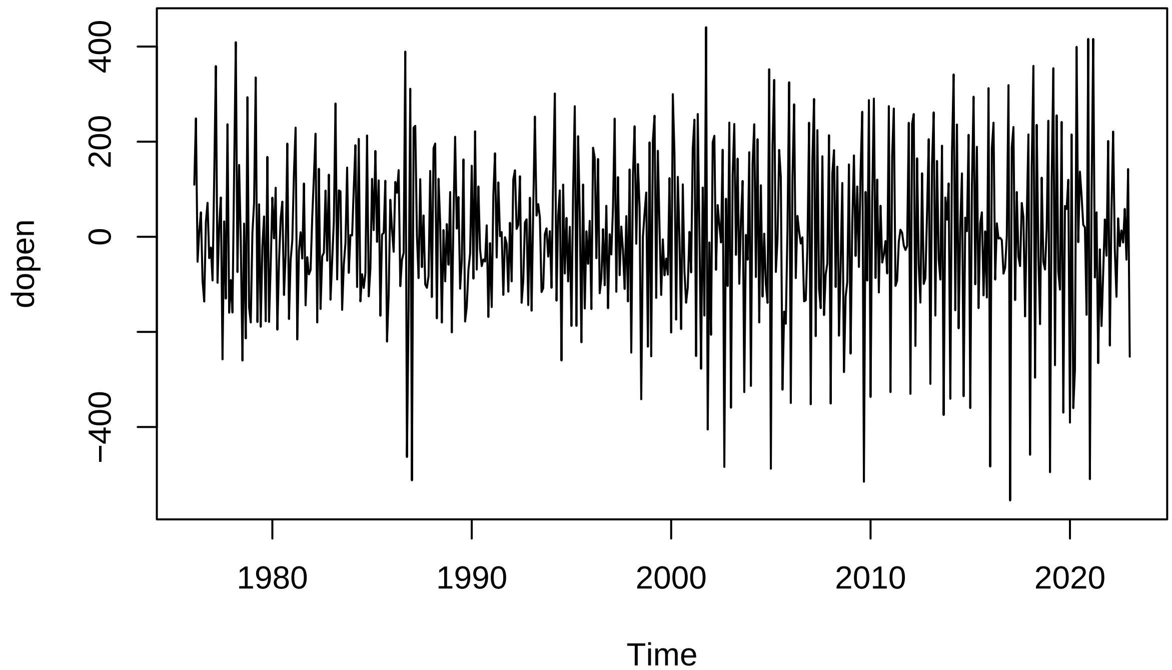
After we have appropriately differenced our models so that they are stationary, we can then begin to consider what type of models to fit to our data.

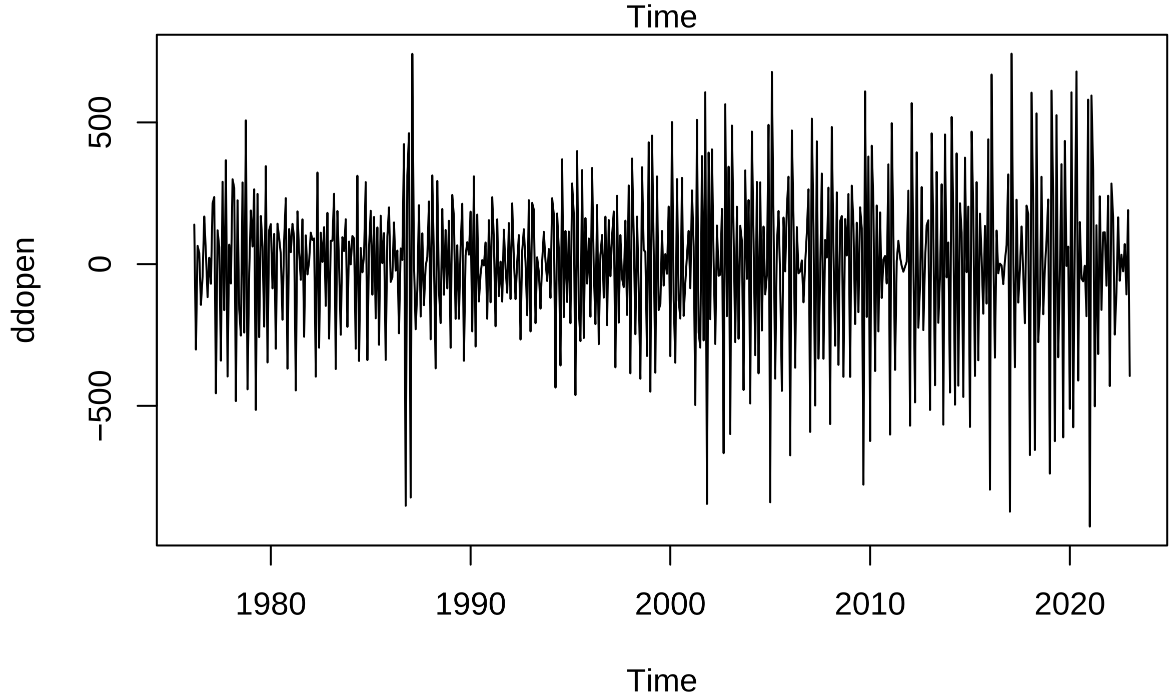




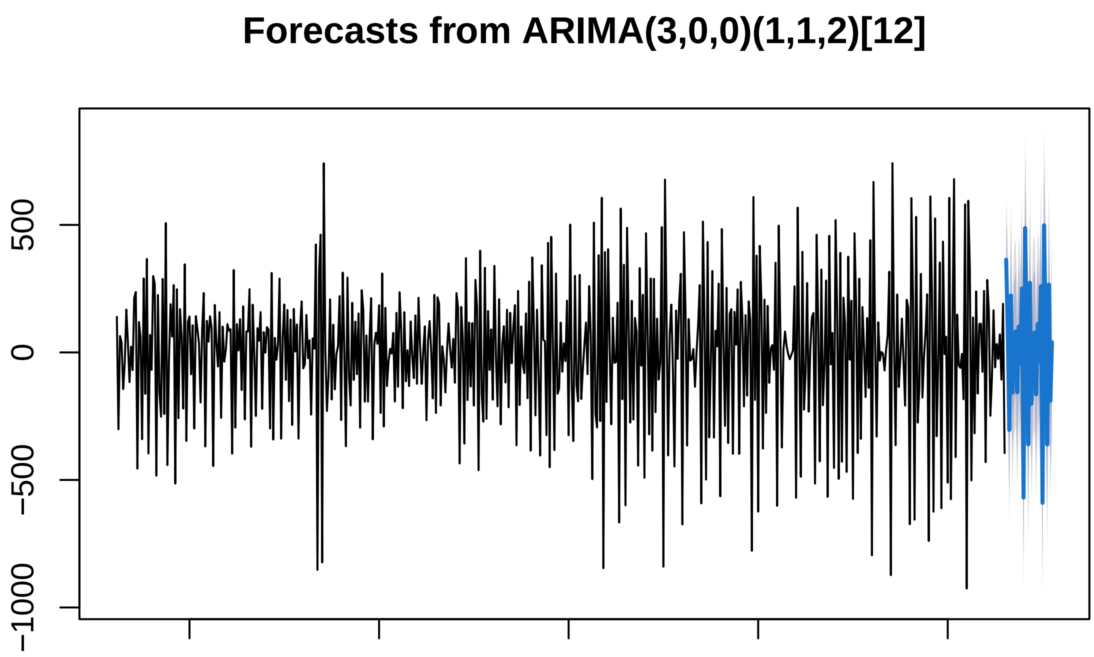

STL decomposition
TOTALNSA = trend + season_year + remainder
1500
1000
TOTALNSA
1400
1200
1000
trend
200
100
0
−100
−200
−300
season_year
200
0
−200
−400
remainder
1980 Jan
1990 Jan
2000 Jan
2010 Jan
2020 Jan
Month

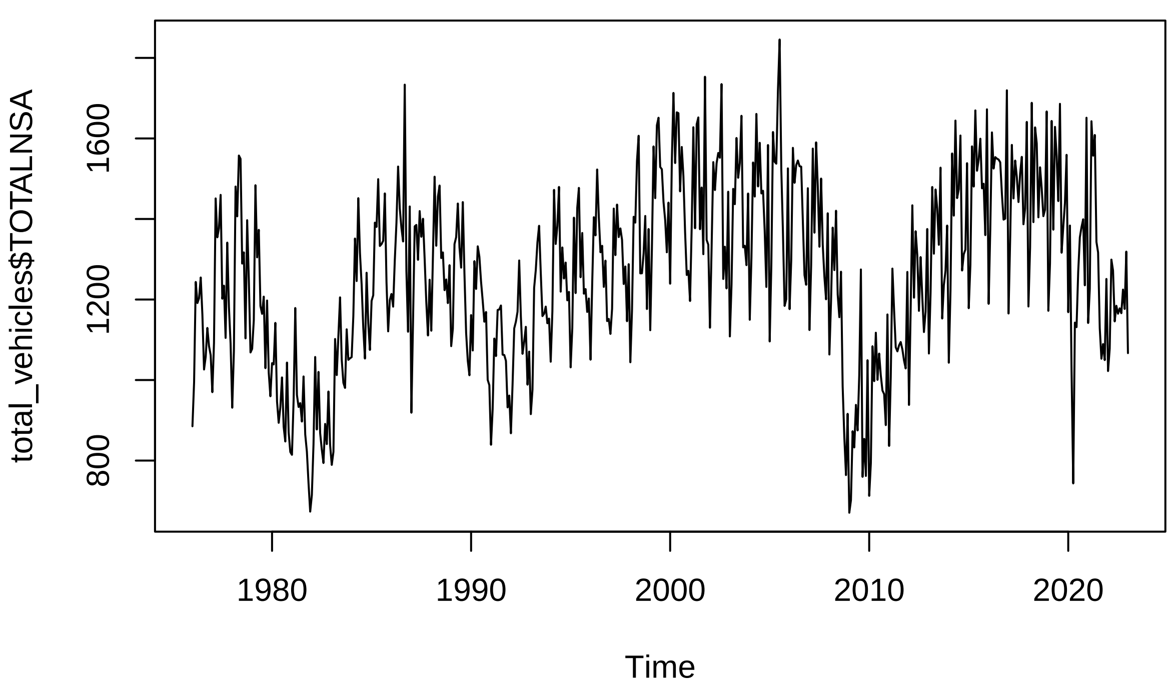

total_vehicles$TOTALNSA
800
1200
1600
1980
1990
2000
2010
2020
Time





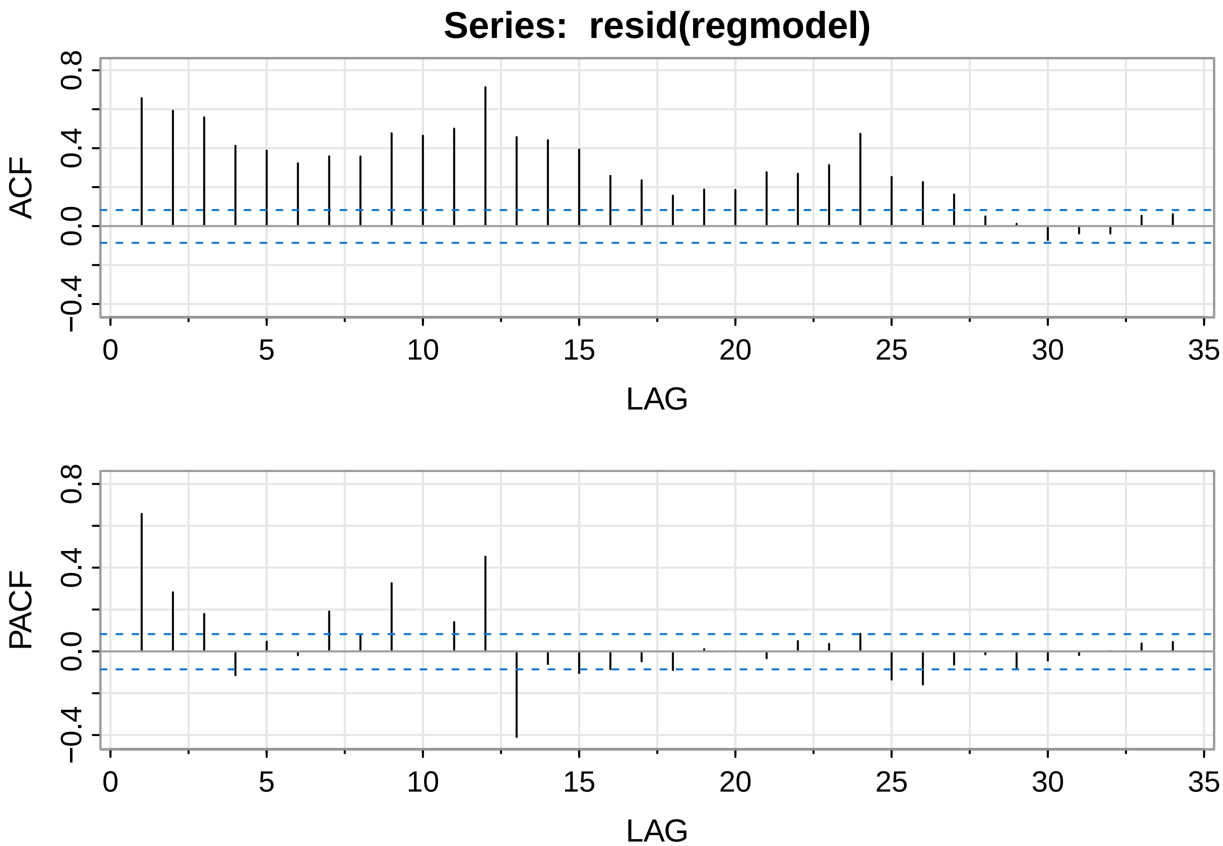

```
## Series: ddopen
## ARIMA(3,0,0)(1,1,2)[12]
##
## Coefficients:
##           ar1      ar2      ar3     sar1     sma1     sma2
##       -1.0416  -0.7815  -0.3270  -0.1110  -0.5883  -0.2286
## s.e.   0.0410   0.0501   0.0403   0.2177   0.2113   0.1710
##
## sigma^2 = 15400:  log likelihood = -3442.57
## AIC=6899.15   AICc=6899.35   BIC=6929.33
```






```
##
## Call:
## lm(formula = total_vehicles$TOTALNSA ~ trend + trend2 + trend3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -665.59 -133.20    6.12  153.30  543.37
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.297e+03  1.324e+01  97.968  < 2e-16 ***
```

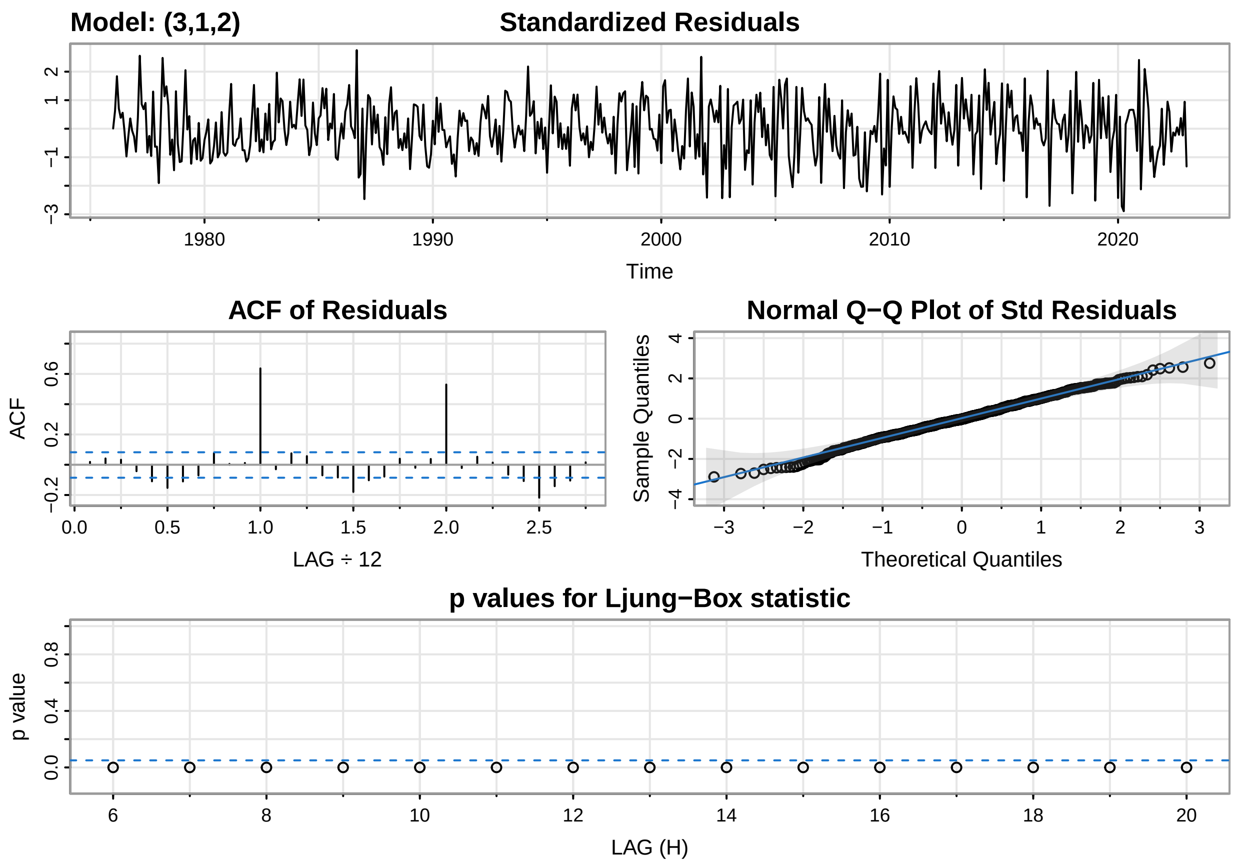
```
## trend        6.297e+00  1.623e+00   3.879 0.000117 ***
## trend2      -1.972e-01  5.341e-02  -3.693 0.000243 ***
## trend3      -3.189e-03  4.474e-03  -0.713 0.476255
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209.8 on 561 degrees of freedom
## Multiple R-squared:  0.1237, Adjusted R-squared:  0.119
## F-statistic:  26.4 on 3 and 561 DF,  p-value: 5.54e-16
```



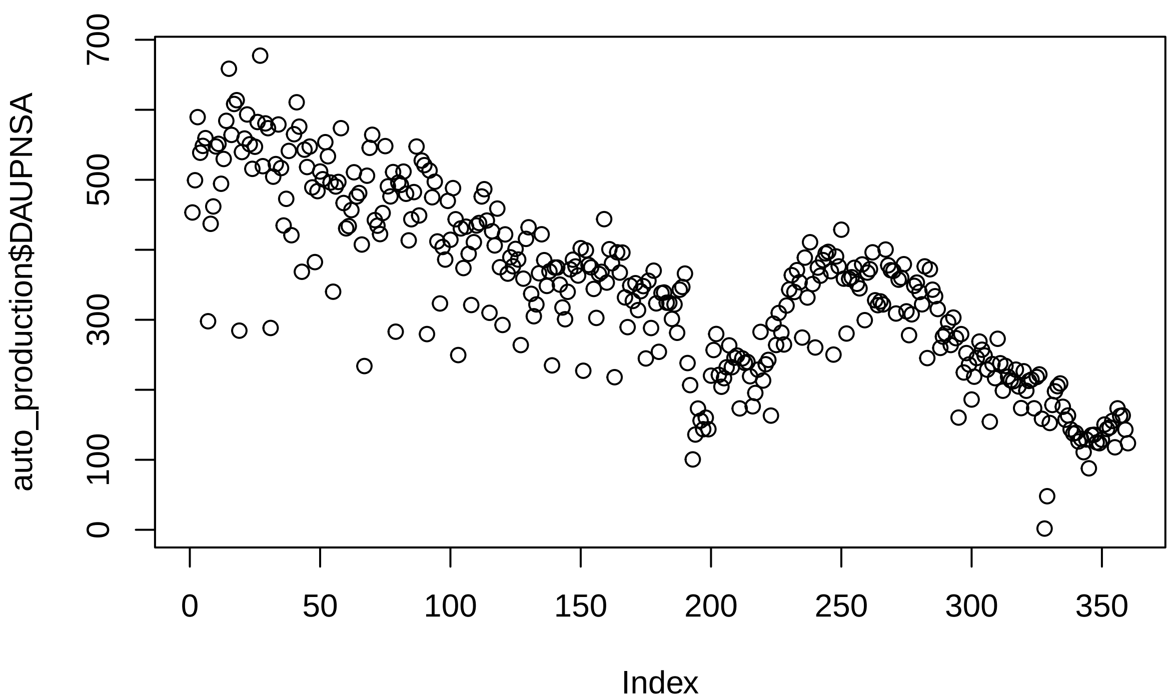

```
##      [,1] [,2] [,3]  [,4] [,5]  [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
## ACF  0.66 0.59 0.56  0.41 0.39  0.32 0.36 0.36 0.48  0.46  0.50  0.71  0.46
## PACF 0.66 0.28 0.18 -0.11 0.05 -0.02 0.19 0.08 0.33  0.00  0.14  0.45 -0.41
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF   0.44  0.39  0.26  0.24  0.16  0.19  0.19  0.28  0.27  0.31  0.47  0.25
## PACF -0.06 -0.10 -0.09 -0.05 -0.09  0.01  0.00 -0.03  0.05  0.04  0.08 -0.14
##      [,26] [,27] [,28] [,29] [,30] [,31] [,32] [,33] [,34]
## ACF   0.23  0.16  0.05  0.01 -0.07 -0.04 -0.04  0.05  0.06
## PACF -0.16 -0.06 -0.01 -0.08 -0.04 -0.02  0.00  0.04  0.05
```

```
## initial  value 5.152535
## iter   2 value 5.071567
## iter   3 value 5.024689
## iter   4 value 5.021687
## iter   5 value 5.012697
## iter   6 value 5.001292
## iter   7 value 4.997668
## iter   8 value 4.991832
```

```
## iter   9 value 4.990246
## iter  10 value 4.990053
## iter  11 value 4.988400
## iter  12 value 4.987975
## iter  13 value 4.987711
## iter  14 value 4.986745
## iter  15 value 4.985365
## iter  16 value 4.984881
## iter  17 value 4.984750
## iter  18 value 4.984739
## iter  19 value 4.984732
## iter  20 value 4.984732
## iter  20 value 4.984732
## final  value 4.984732
## converged
## initial  value 4.984676
## iter   2 value 4.984445
## iter   3 value 4.984391
## iter   4 value 4.984367
## iter   5 value 4.984351
## iter   6 value 4.984347
## iter   7 value 4.984344
## iter   8 value 4.984344
## iter   9 value 4.984343
## iter  10 value 4.984341
## iter  10 value 4.984341
## iter  10 value 4.984341
## final  value 4.984341
## converged
```

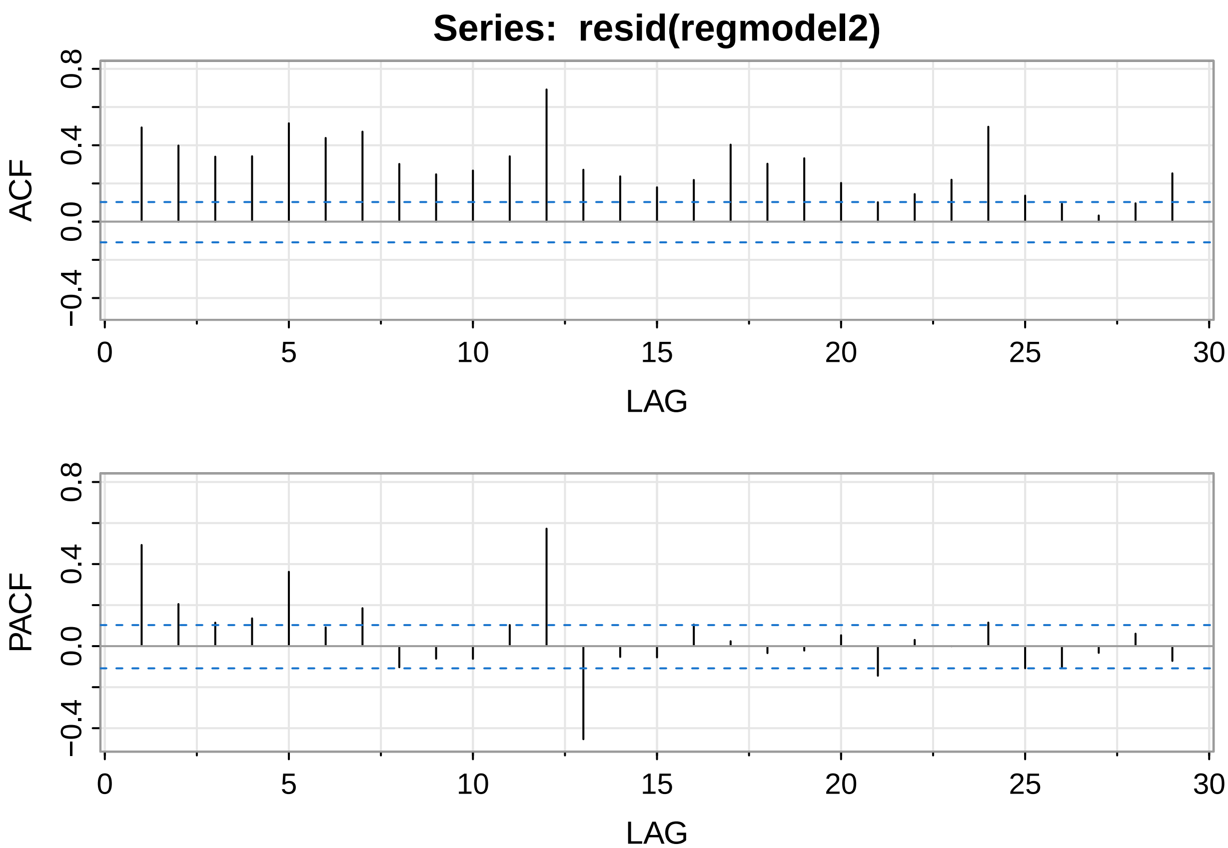



```
##        ar1        ar2        ar3        ma1        ma2      trend     trend2
## -0.3904722  0.3307520  0.2495136 -0.1731236 -0.6554213  1.9584736 -0.3361823
```




```
##
## Call:
```

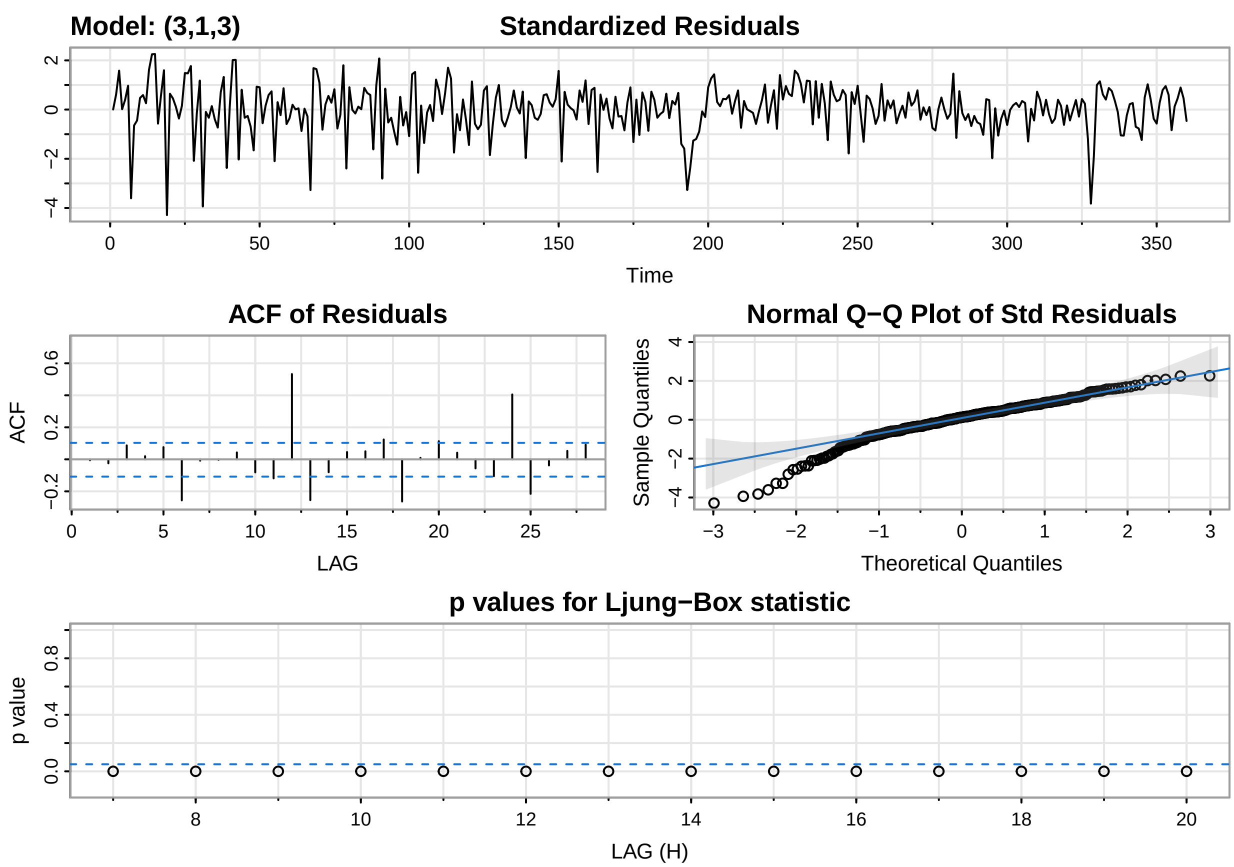
```
## lm(formula = auto_production$DAUPNSA ~ tre)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -231.87  -40.40    5.65   48.52  178.12
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 344.9582     3.9384   87.59   <2e-16 ***
## tre          -1.0047     0.0379  -26.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 74.73 on 358 degrees of freedom
## Multiple R-squared:  0.6625, Adjusted R-squared:  0.6616
## F-statistic: 702.8 on 1 and 358 DF,  p-value: < 2.2e-16
```




```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF  0.49 0.40 0.34 0.34 0.51 0.44 0.47  0.3  0.25  0.27  0.34  0.69  0.27
## PACF 0.49 0.21 0.11 0.14 0.36 0.09 0.18 -0.1 -0.06 -0.06  0.10  0.57 -0.45
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24] [,25]
## ACF   0.24  0.18  0.22  0.40  0.30  0.33  0.20  0.10  0.14  0.22  0.50  0.14
## PACF -0.05 -0.05  0.11  0.02 -0.03 -0.02  0.05 -0.14  0.03  0.00  0.11 -0.11
##      [,26] [,27] [,28] [,29]
## ACF   0.09  0.03  0.10  0.25
## PACF -0.10 -0.03  0.06 -0.07

## initial  value 4.318562
```

```
## iter   2 value 4.157746
## iter   3 value 4.124139
## iter   4 value 4.100119
## iter   5 value 4.098814
## iter   6 value 4.098686
## iter   7 value 4.098150
## iter   8 value 4.097729
## iter   9 value 4.097070
## iter  10 value 4.093929
## iter  11 value 4.090688
## iter  12 value 4.077772
## iter  13 value 4.066611
## iter  14 value 4.054402
## iter  15 value 4.037931
## iter  16 value 4.031472
## iter  17 value 4.028480
## iter  18 value 4.027491
## iter  19 value 4.026483
## iter  20 value 4.022072
## iter  21 value 4.021494
## iter  22 value 4.021187
## iter  23 value 4.021133
## iter  24 value 4.021072
## iter  25 value 4.021020
## iter  26 value 4.020993
## iter  27 value 4.020954
## iter  28 value 4.020947
## iter  29 value 4.020947
## iter  29 value 4.020947
## iter  29 value 4.020947
## final  value 4.020947
## converged
## initial  value 3.993127
## iter   2 value 3.991192
## iter   3 value 3.990562
## iter   4 value 3.989690
## iter   5 value 3.988962
## iter   6 value 3.988546
## iter   7 value 3.988037
## iter   8 value 3.987515
## iter   9 value 3.986954
## iter  10 value 3.985434
## iter  11 value 3.981475
## iter  12 value 3.980978
## iter  13 value 3.978816
## iter  14 value 3.978606
## iter  15 value 3.978409
## iter  16 value 3.978238
## iter  17 value 3.978086
## iter  18 value 3.977774
## iter  19 value 3.977667
## iter  20 value 3.977662
## iter  21 value 3.977660
## iter  22 value 3.977653
```

```
## iter  23 value 3.977651
## iter  24 value 3.977650
## iter  25 value 3.977649
## iter  26 value 3.977647
## iter  27 value 3.977644
## iter  28 value 3.977641
## iter  29 value 3.977637
## iter  30 value 3.977635
## iter  31 value 3.977633
## iter  32 value 3.977632
## iter  33 value 3.977632
## iter  33 value 3.977632
## iter  33 value 3.977632
## final  value 3.977632
## converged
```



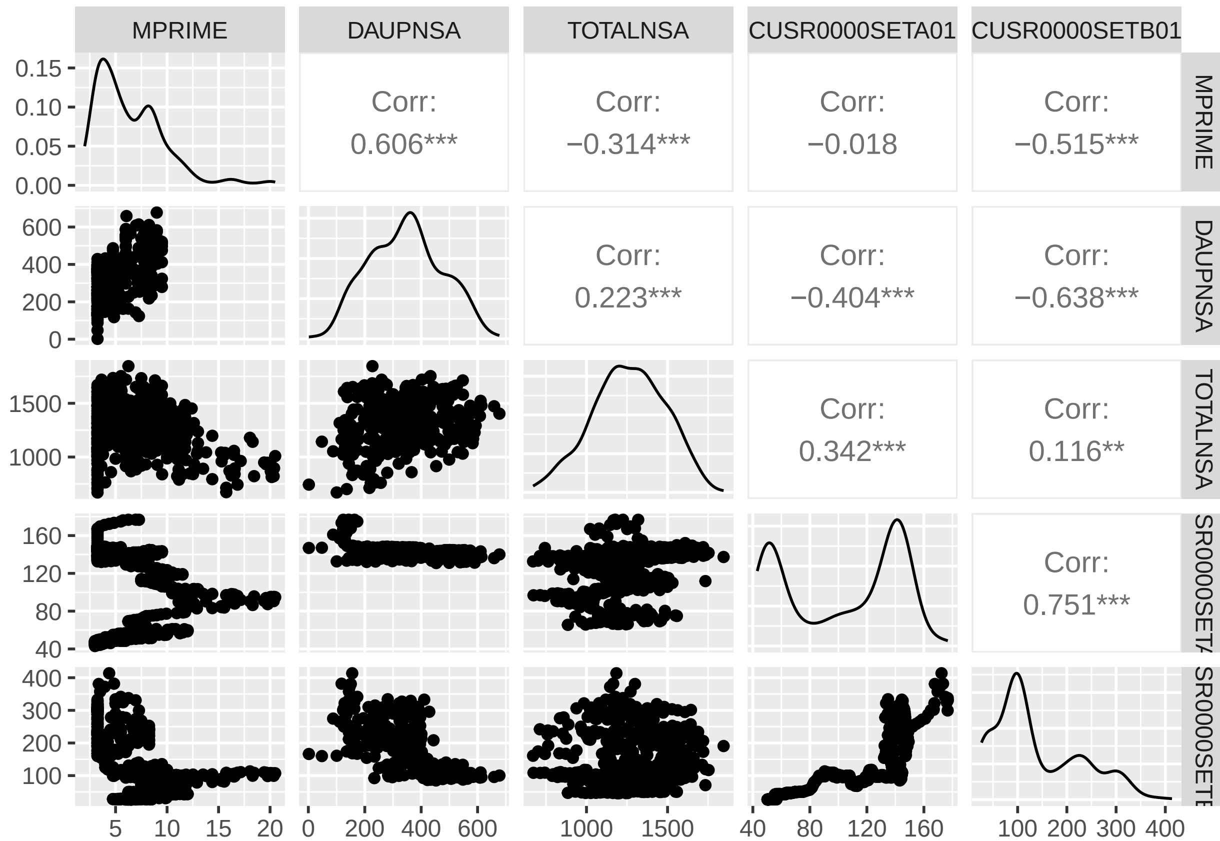

```
##        ar1        ar2        ar3        ma1        ma2        ma3        tre
##  0.8278123 -0.8443521 -0.1524987 -1.6802058  1.6982727 -0.7566727 -1.0375171
```

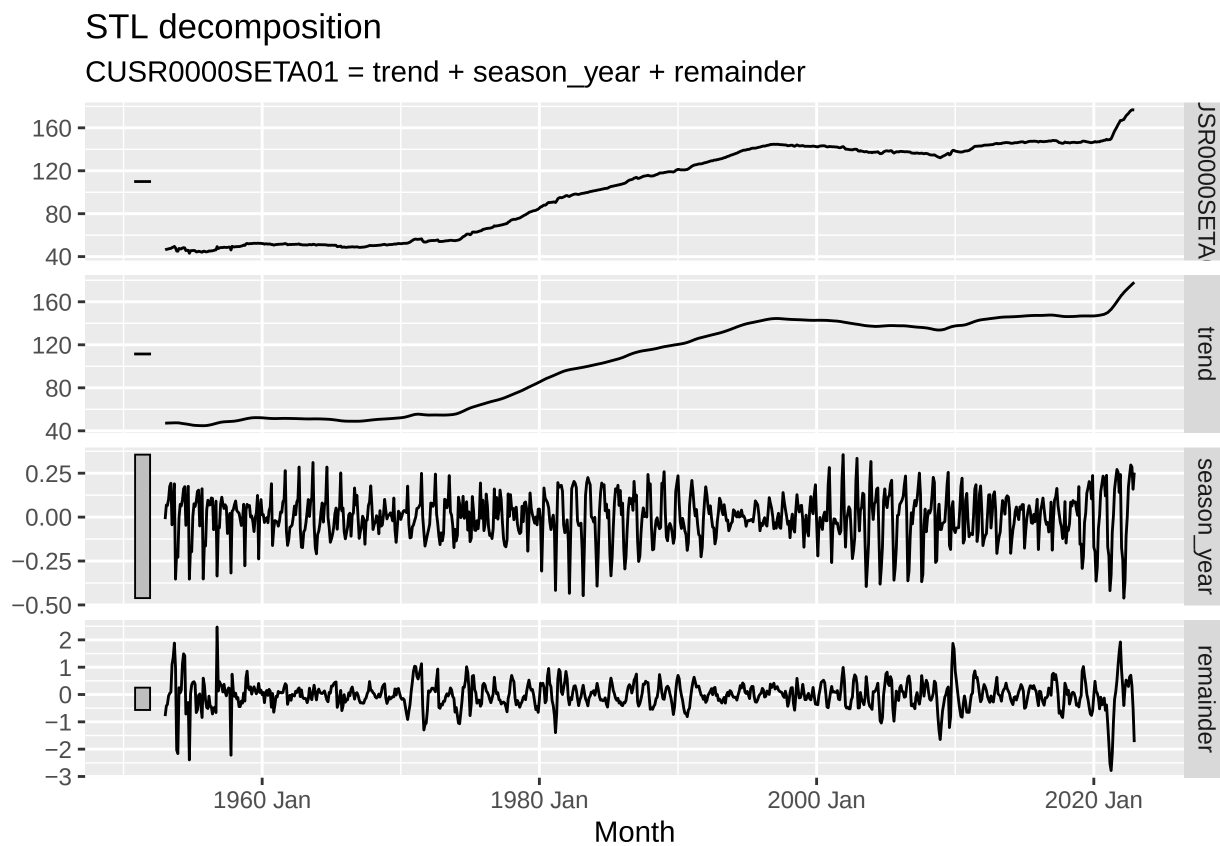
```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 529 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 324 rows containing missing values
```

```
## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 49 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 230 rows containing missing values

## Warning: Removed 529 rows containing missing values (`geom_point()`).

## Warning: Removed 529 rows containing non-finite values (`stat_density()`).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 529 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 529 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 529 rows containing missing values

## Warning: Removed 324 rows containing missing values (`geom_point()`).

## Warning: Removed 529 rows containing missing values (`geom_point()`).

## Warning: Removed 324 rows containing non-finite values (`stat_density()`).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 325 rows containing missing values

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 325 rows containing missing values

## Warning: Removed 49 rows containing missing values (`geom_point()`).

## Warning: Removed 529 rows containing missing values (`geom_point()`).

## Warning: Removed 325 rows containing missing values (`geom_point()`).

## Warning: Removed 49 rows containing non-finite values (`stat_density()`).

## Warning in ggally_statistic(data = data, mapping = mapping, na.rm = na.rm, :
## Removed 230 rows containing missing values

## Warning: Removed 230 rows containing missing values (`geom_point()`).

## Warning: Removed 529 rows containing missing values (`geom_point()`).

## Warning: Removed 325 rows containing missing values (`geom_point()`).

## Warning: Removed 230 rows containing missing values (`geom_point()`).

## Warning: Removed 230 rows containing non-finite values (`stat_density()`).
```


MPRIME
DAUPNSA
TOTALNSA
CUSR0000SETA01
CUSR0000SETB01
Corr:
0.606***
Corr:
−0.314***
Corr:
−0.018
Corr:
−0.515***
Corr:
0.223***
Corr:
−0.404***
Corr:
−0.638***
Corr:
0.342***
Corr:
0.116**
Corr:
0.751***


STL decomposition
CUSR0000SETA01 = trend + season_year + remainder
trend
season_year
remainder
Month

References

Glenn, Stephanie. n.d. "KPSS Test: Definition and Interpretation." https://www.statisticshowto.com/kpss-test/.

"Inflation and the Auto Industry: When Will Car Prices Drop." 2022. https://www.jpmorgan.com/insights/research/when-will-car-prices-drop.

"Long-Term Trends in Car and Light Truck Sales." 2021. https://fredblog.stlouisfed.org/2022/10/whats-been-driving-the-rise-in-auto-prices-since-covid/?utm_source=series_page&utm_medium=related_content&utm_term=related_resources&utm_campaign=fredblog.

"No End in Sight: New Vehicle Transaction Prices End 2022 at Record Highs, According to New Data from Kelley Blue Book." 2023. https://www.coxautoinc.com/market-insights/kbb-atp-december-2022/.

"PSTAT 174/274: Time Series Part IV." 2023. University of California, Santa Barbara.

Rob J Hyndman, George Athanasopoulos. 2021. "Forecasting: Principles and Practice." Monash University, Australia; OTexts.

"What's Been Drivin the Rise in Auto Prices Since COVID." 2022. https://fredblog.stlouisfed.org/2021/03/long-term-trends-in-car-and-light-truck-sales/?utm_source=series_page&utm_medium=related_content&utm_term=related_resources&utm_campaign=fredblog.