

# Time Series Modeling Project

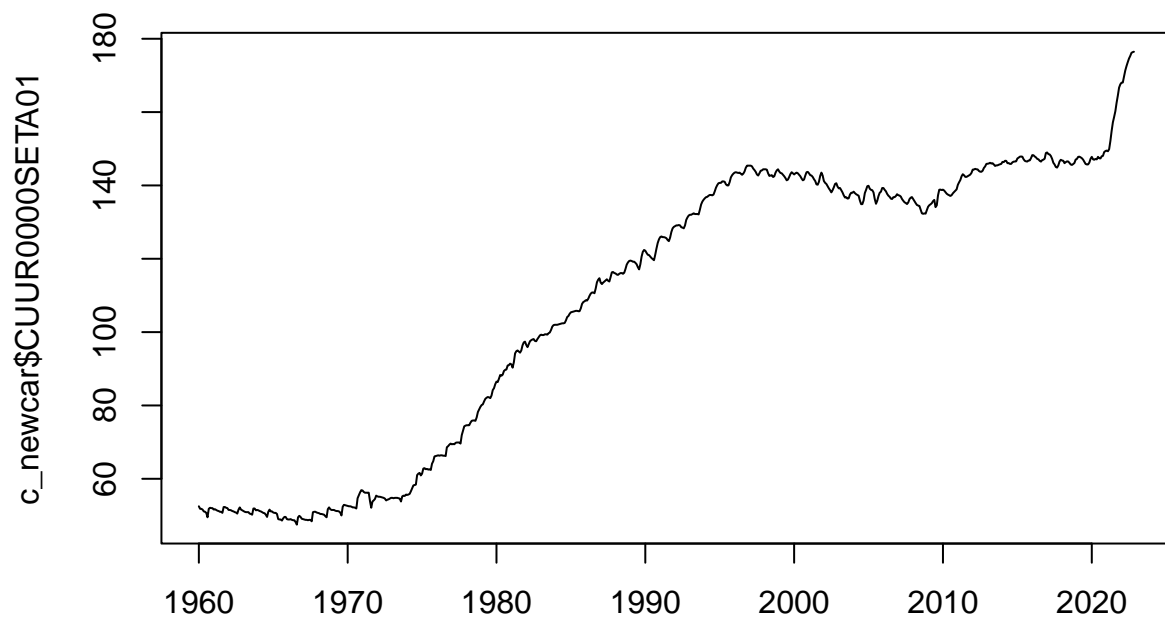
Kyle Wu

2023-01-30

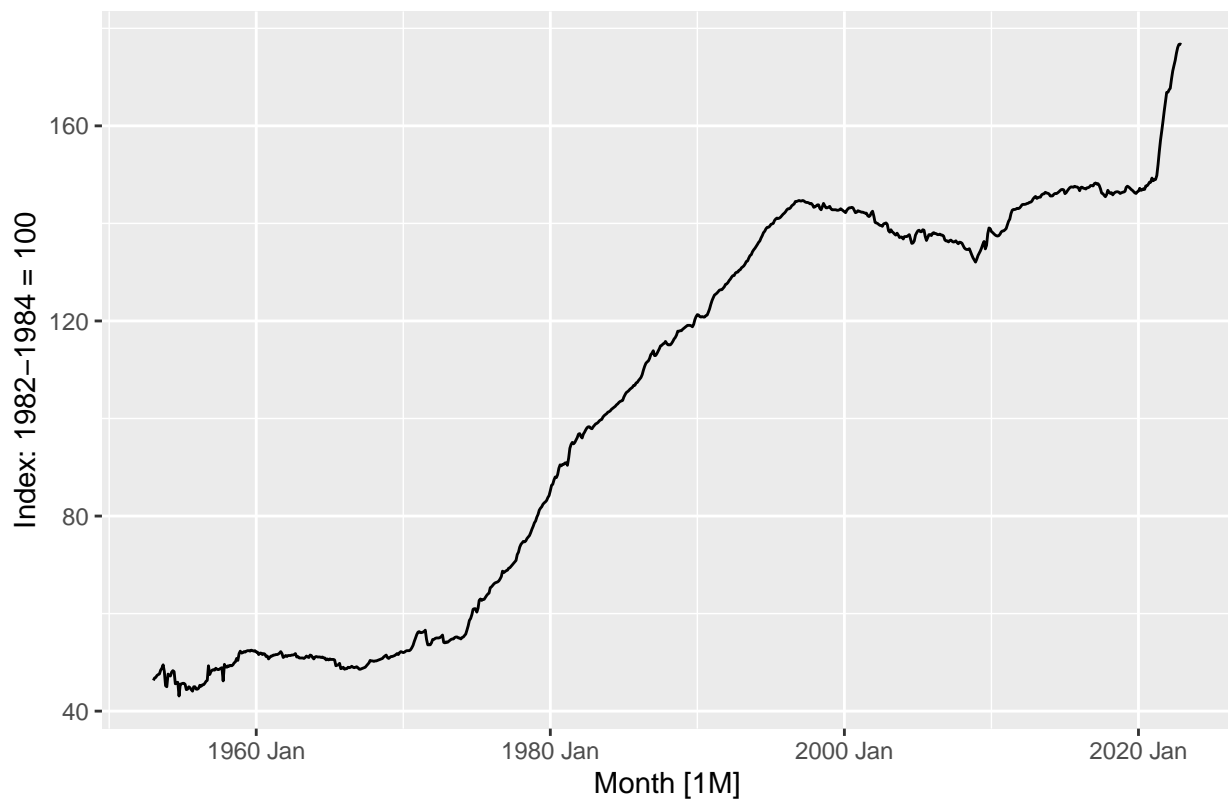
```
## Rows: 889 Columns: 6
## -- Column specification -----
## Delimiter: ","
## chr  (4): CUSR0000SETA01, DAUPSA, TOTALNSA, CUSR0000SETB01
## dbl  (1): MPRIME
## date (1): DATE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types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

**R Markdown**

New Car CPI



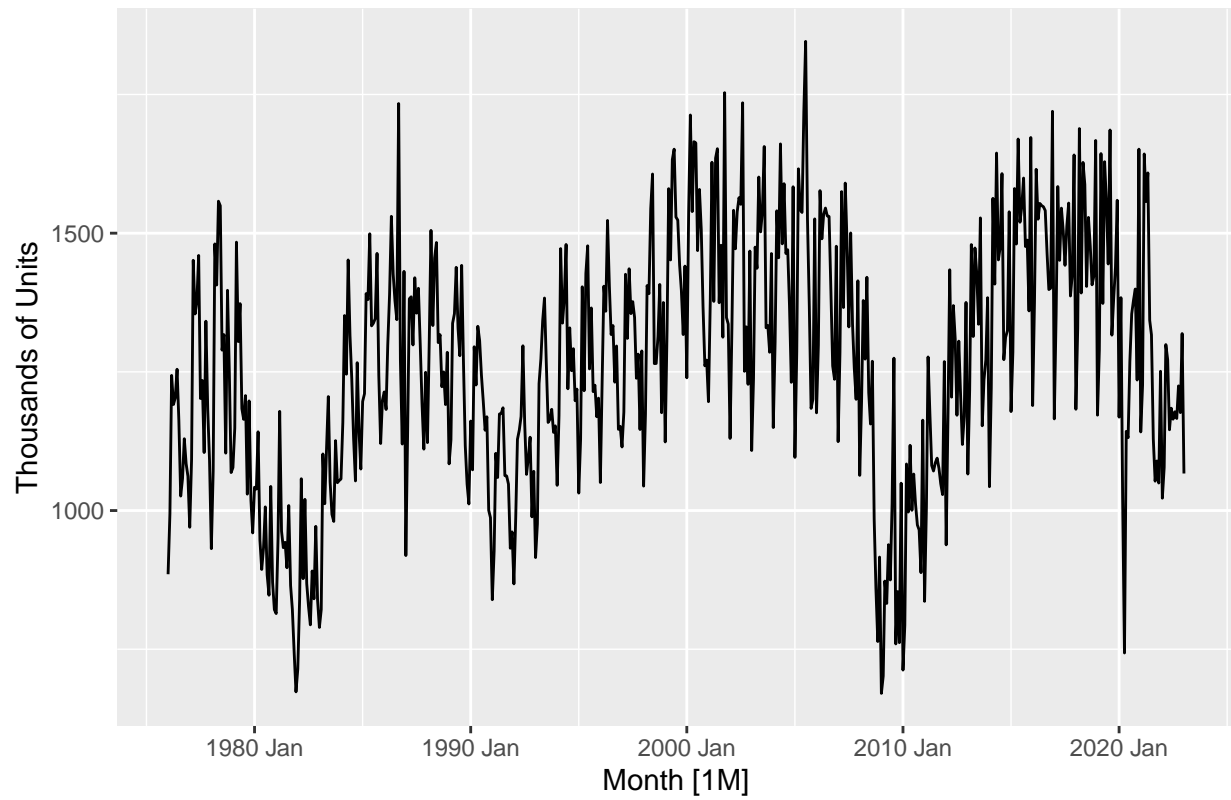
CPI of New Cars



Prime Rate Loans

Total Vehicle Sales

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 over 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 determining 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.

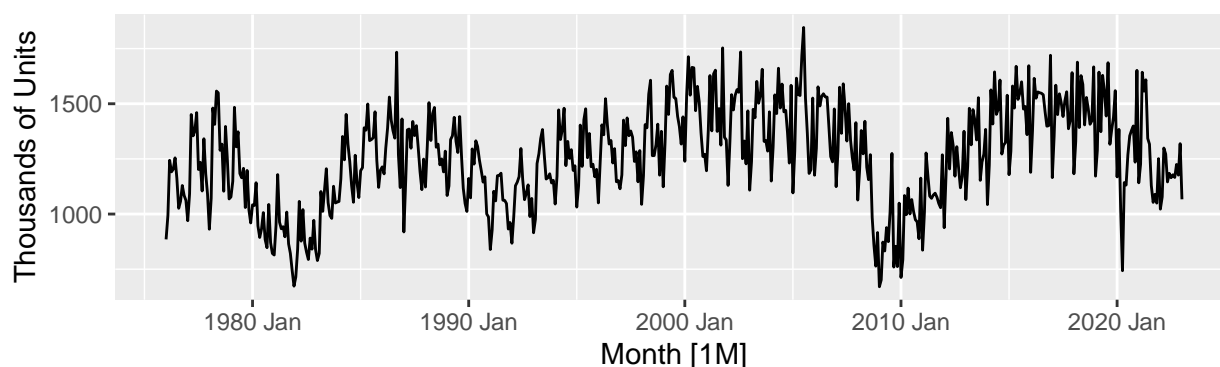
## Analysis of Empirical Properties

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.

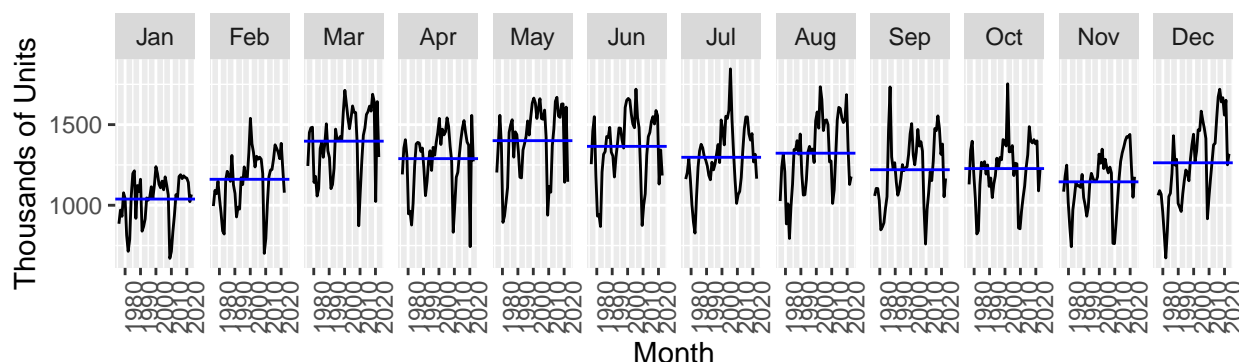
We will first analyze the variables individually before talking about all the factors as they may relate to projecting future car sales.

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

### Total Vehicle Sales

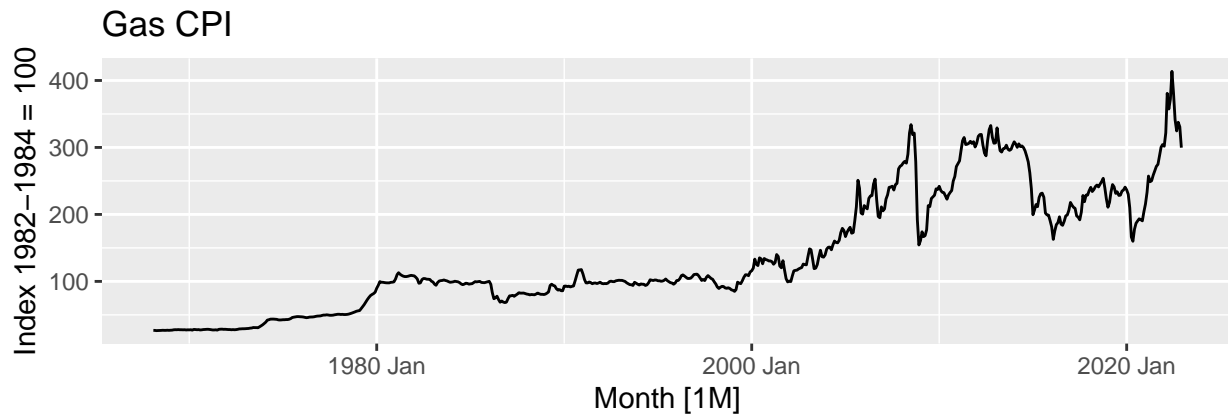
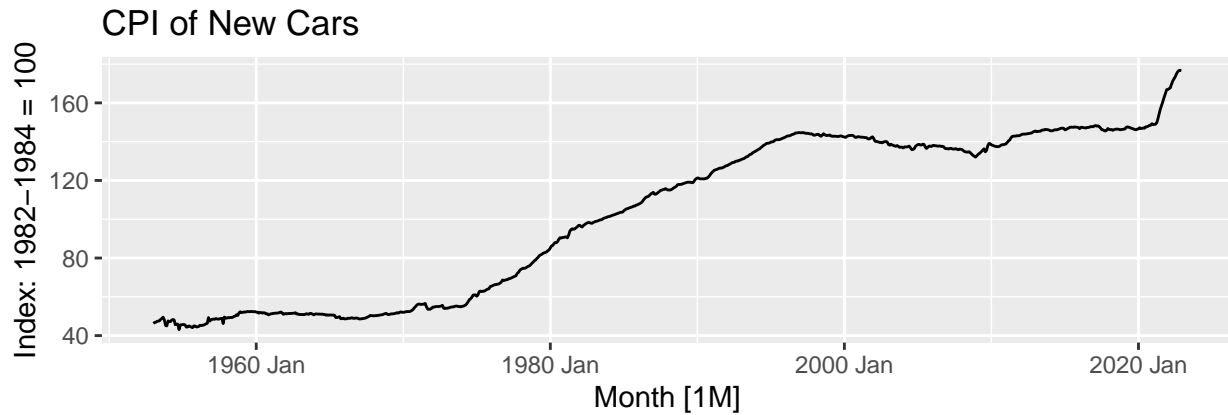


### Total vehicles sold



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.

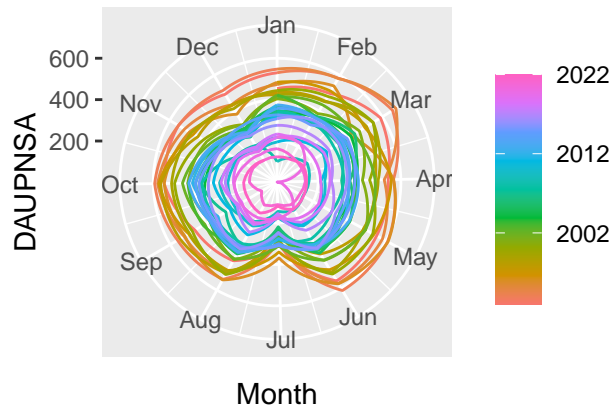
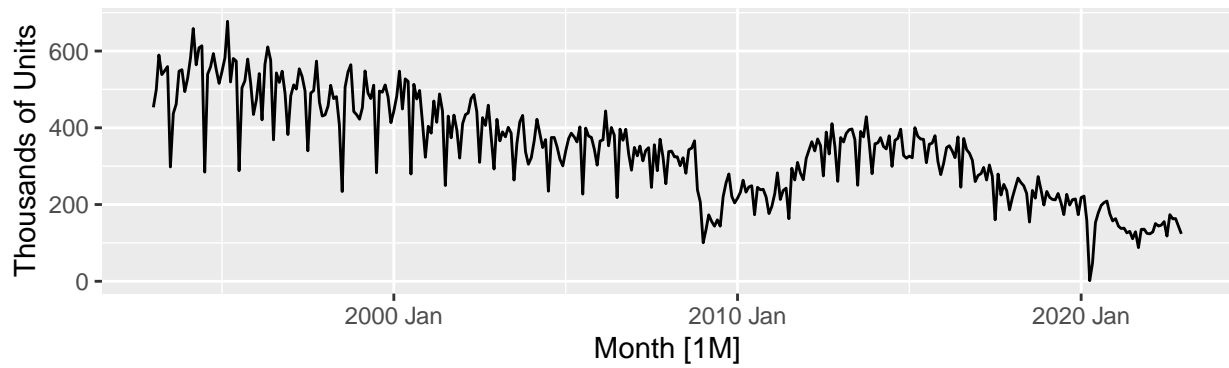
We can now look at the consumer price index of new vehicles in the United States.



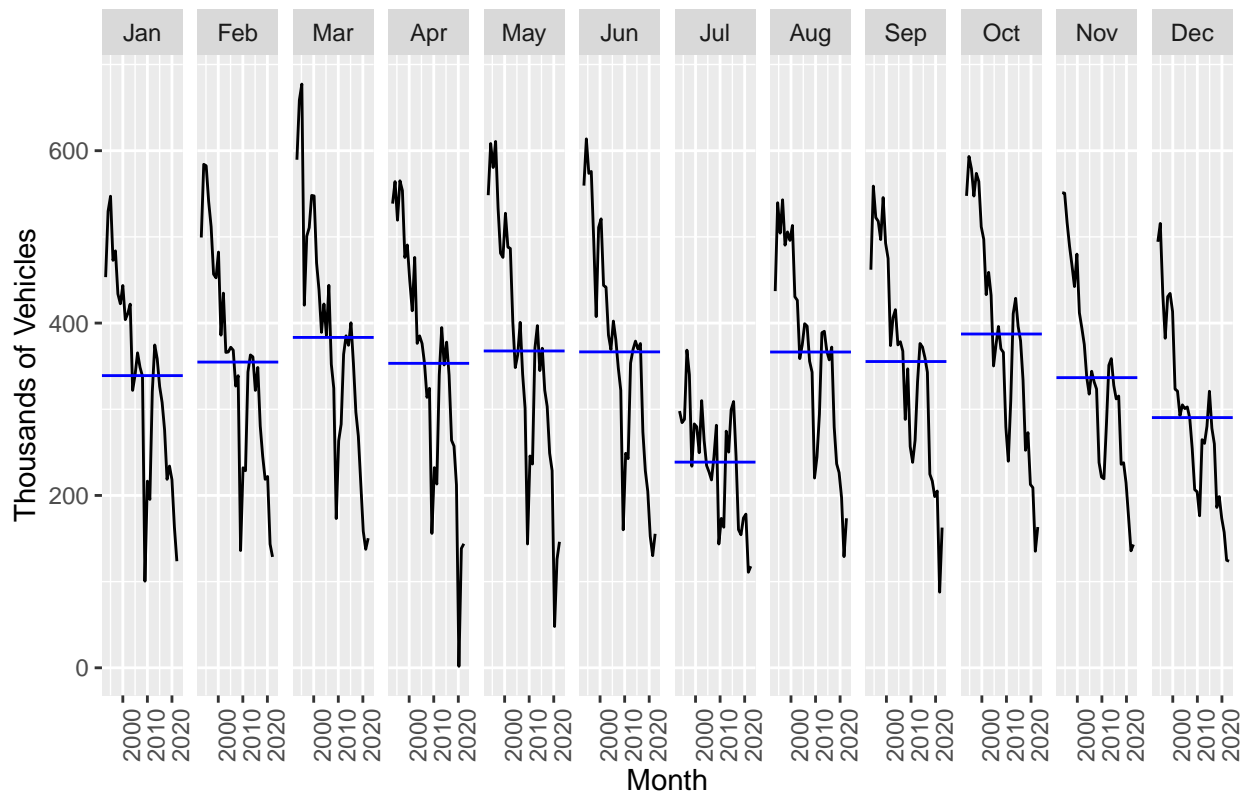
The data for the consumer price index measure of new cars begins in January 1953 and is recorded monthly. This dataset has already been seasonally adjusted so that will not be something that I will have to worry about later. It is worth mentioning that CPI should not be regarded as the price of a vehicle, 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?

Domestic Auto Production

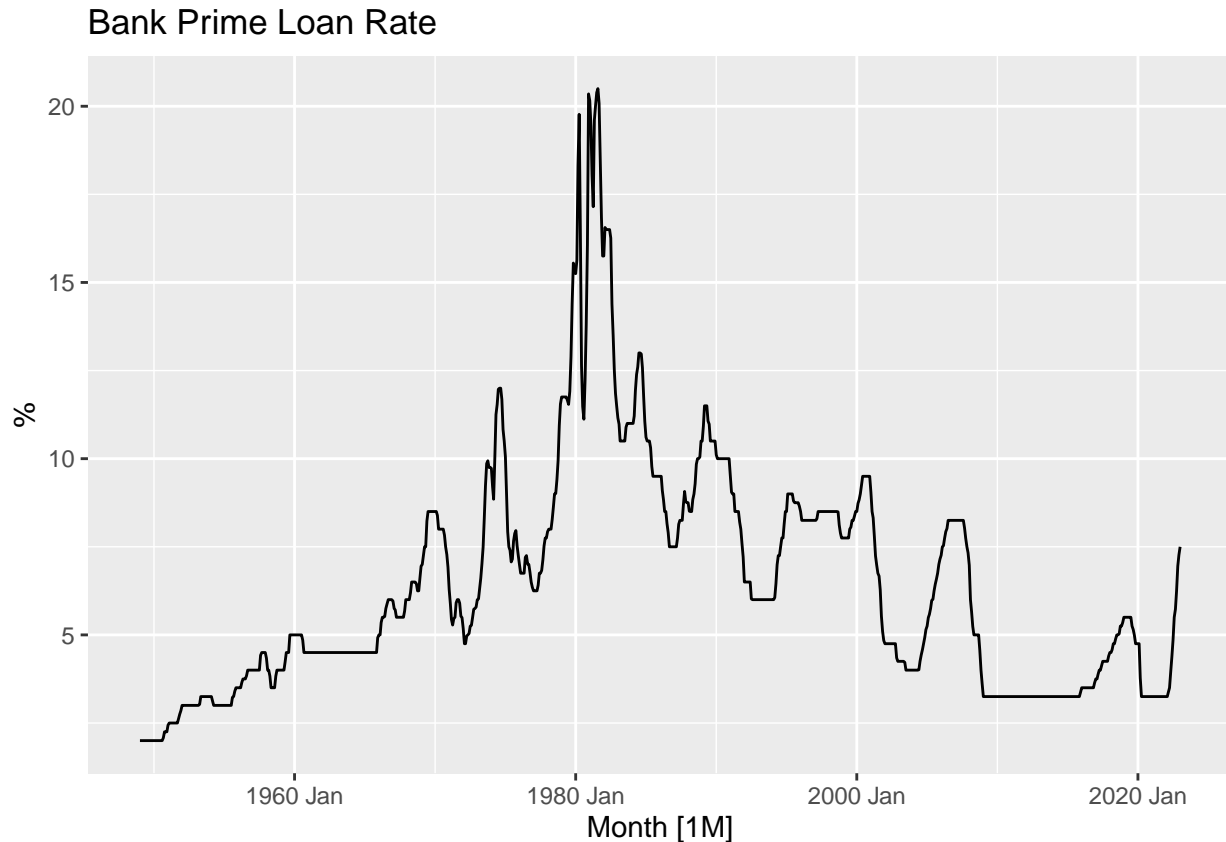


Domestic Auto Production



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.



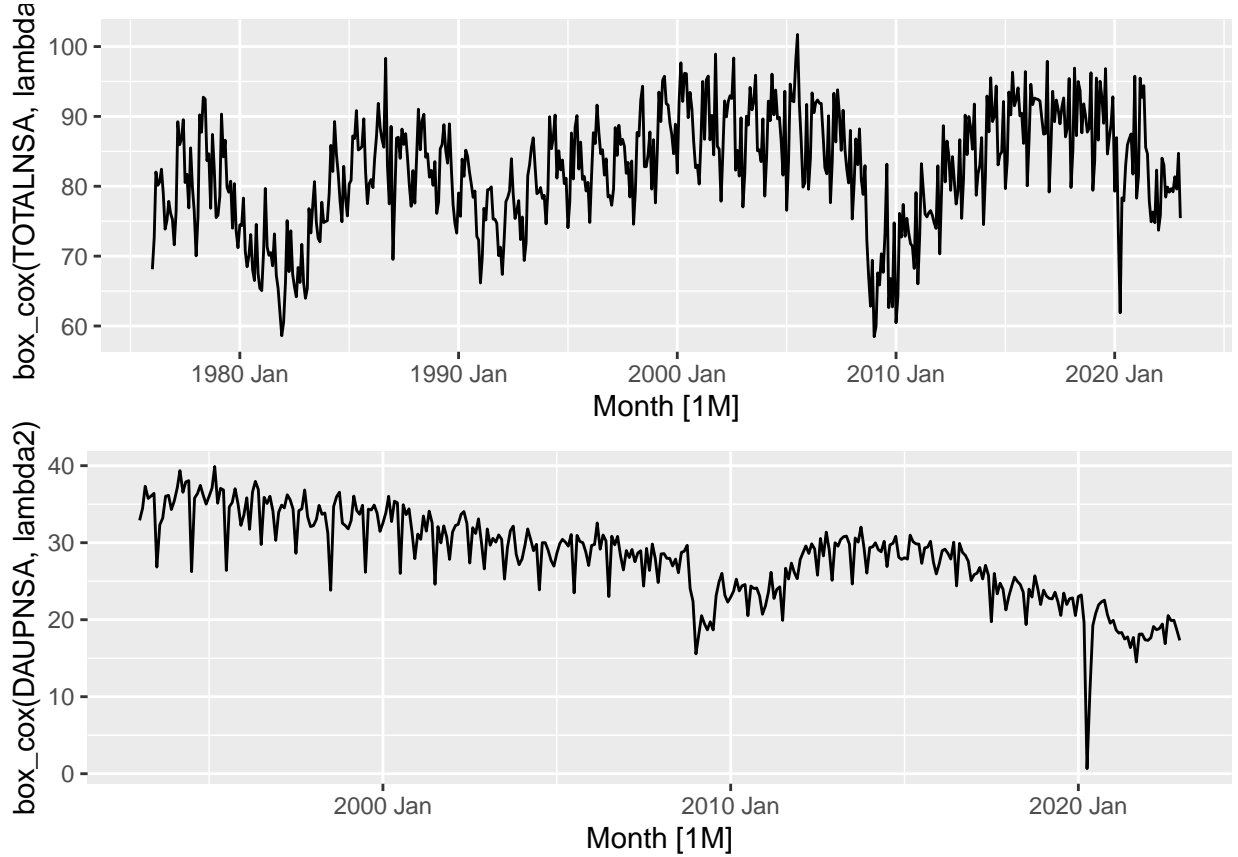
The Bank Prime loan rate is set in relation to the federal funds rate set by the federal reserve and we have data going back to January 1949. The rate is often set based on other determinants of the market so we may not be able to generate too many insights from it individually. However, if we use the 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

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  $\lambda$  and can be defined as follows:



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

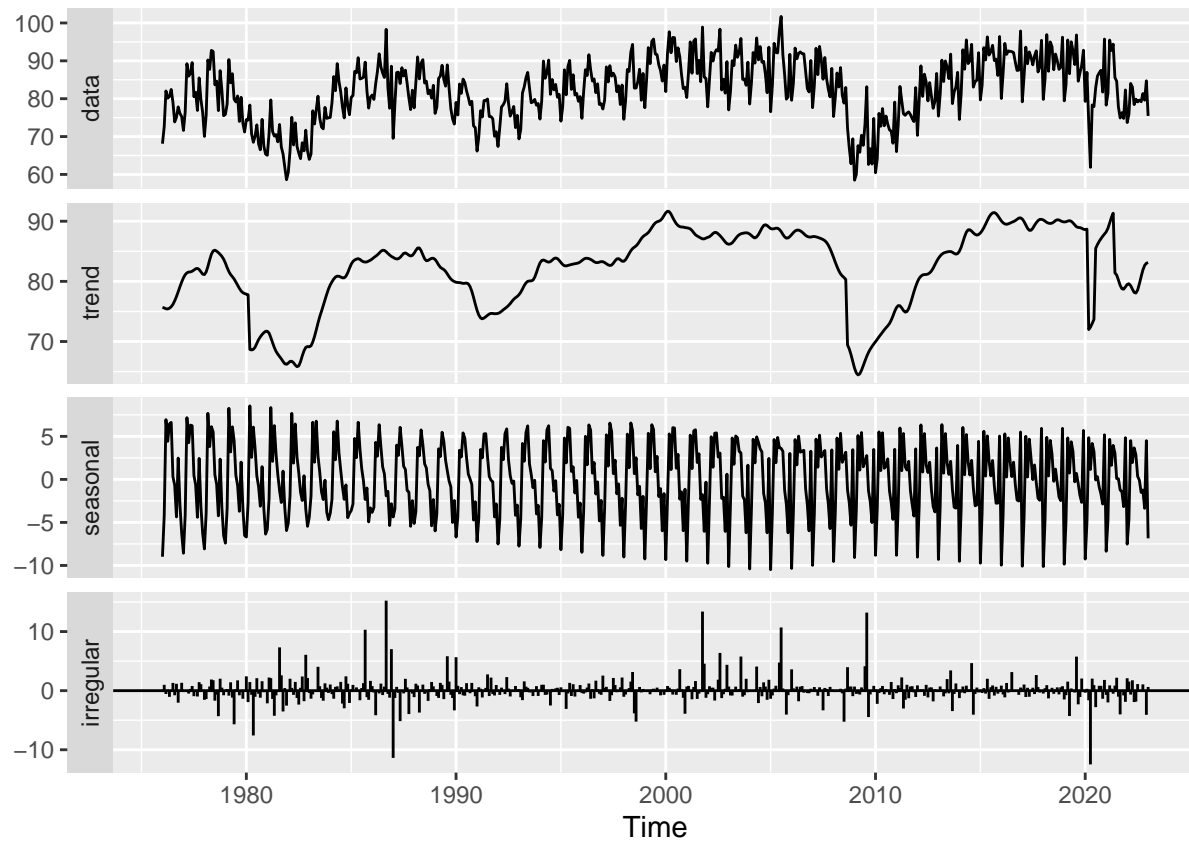


After running the calculations, we find that the appropriate  $\lambda$  value for total auto sales was approximately 0.53 and the appropriate  $\lambda$  value for domestic auto production was 0.45. From the plots that we created above, we also see that we were able to successfully adjust our data since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

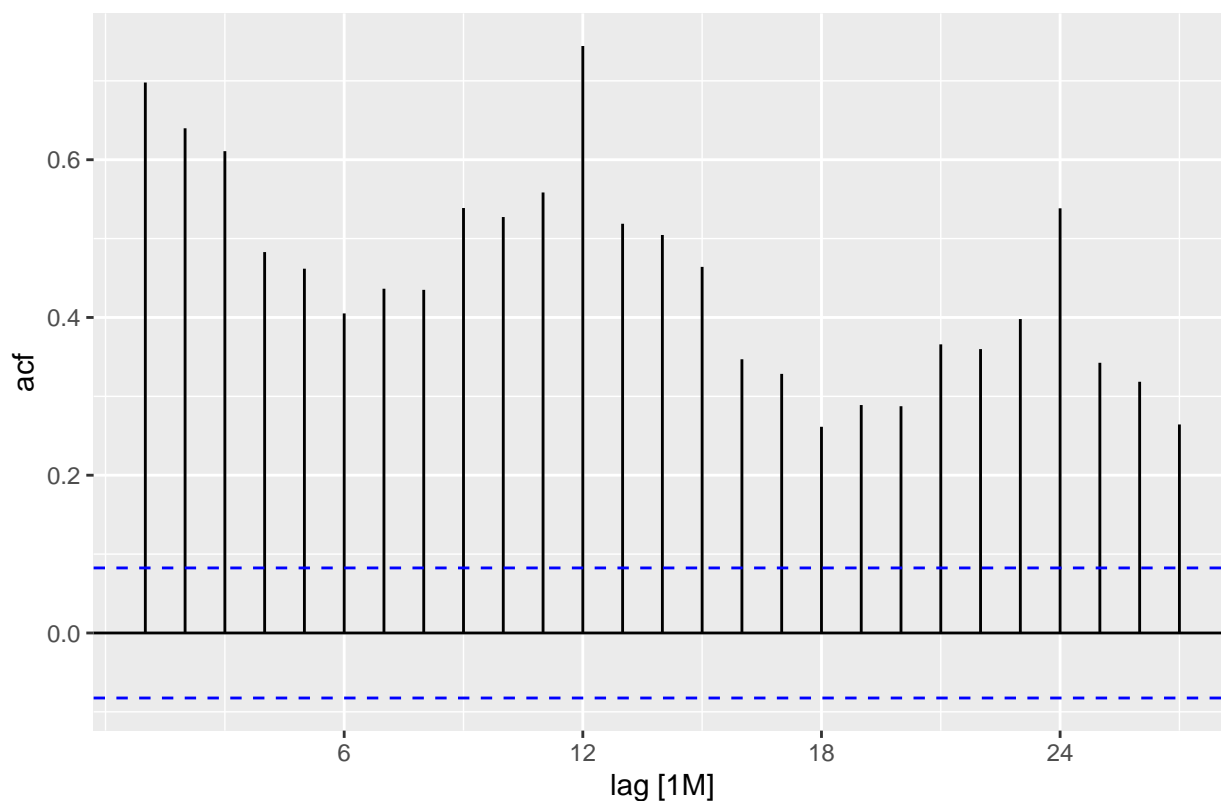
The second possible case of data is a time series that contains both a trend, and seasonal variation. In this case, there are two possible cases of decomposition we can take. If we assume additive decomposition, then we can use  $y_t = S_t + T_t + R_t$ , where  $S_t$  is the seasonal component,  $T_t$  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 (**partiv?**). 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

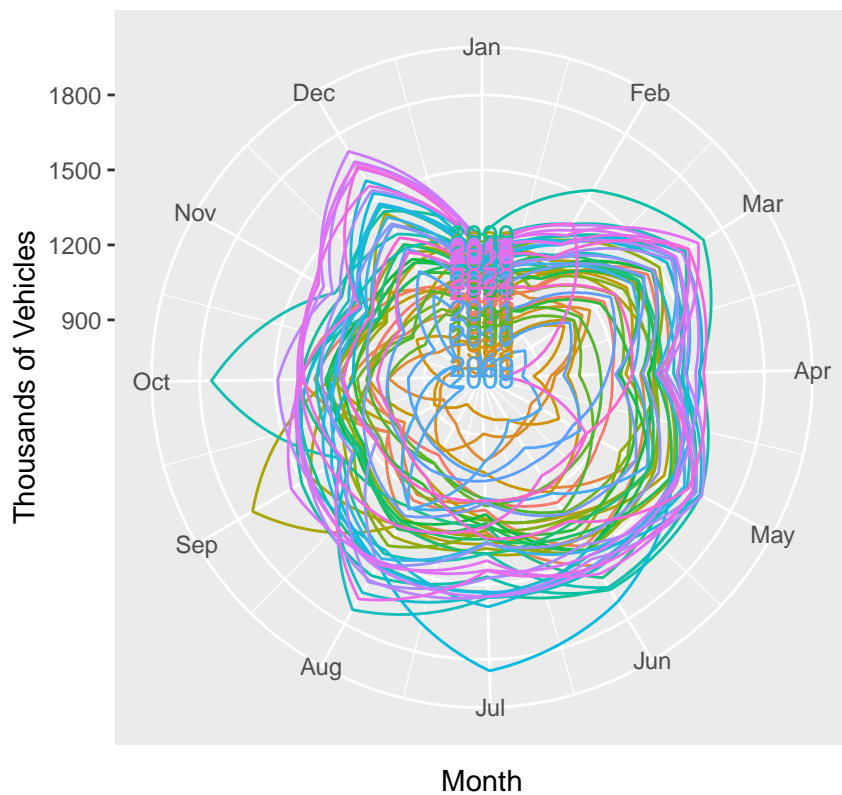


From the decomposition, we can get a general idea of the trend and seasonality present within the plot, which will then allow us to better estimate and thus forecast the data moving forwards.

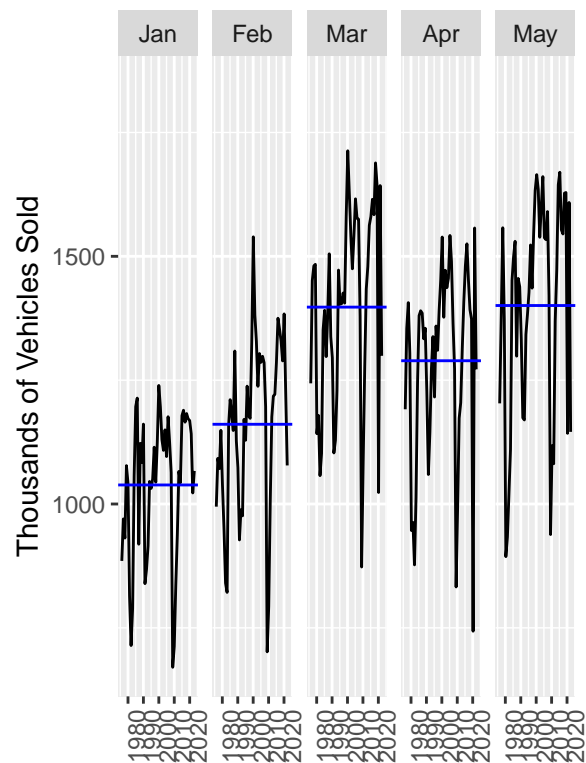
Total Vehicles Sold



this thing better work

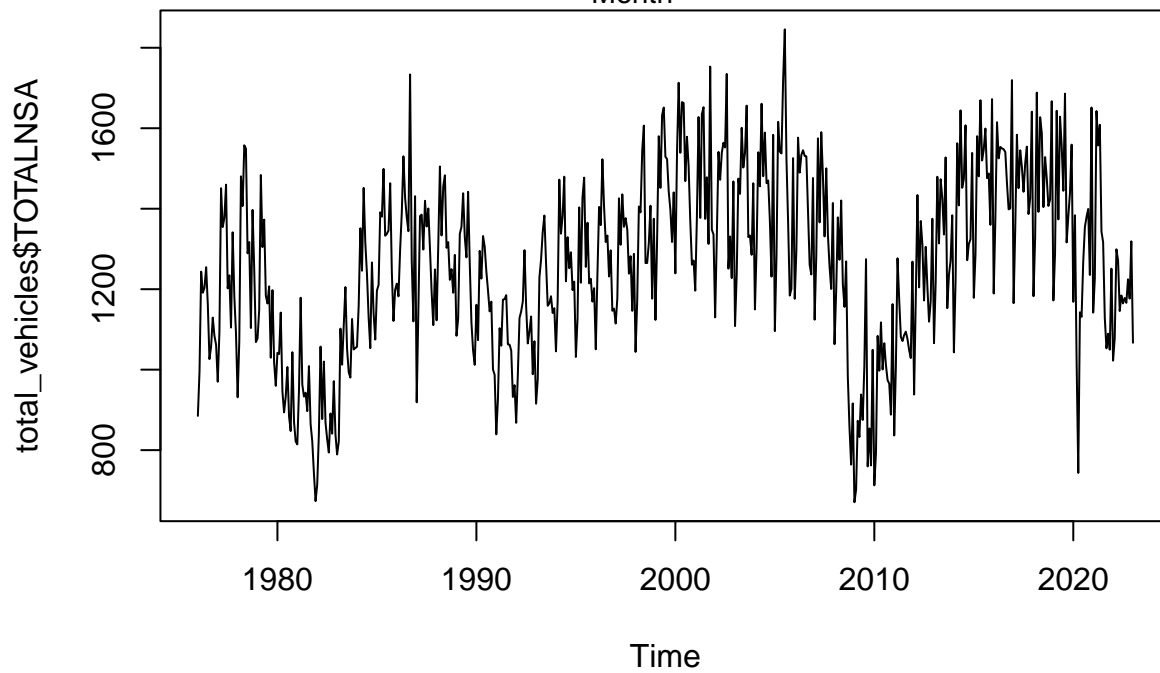
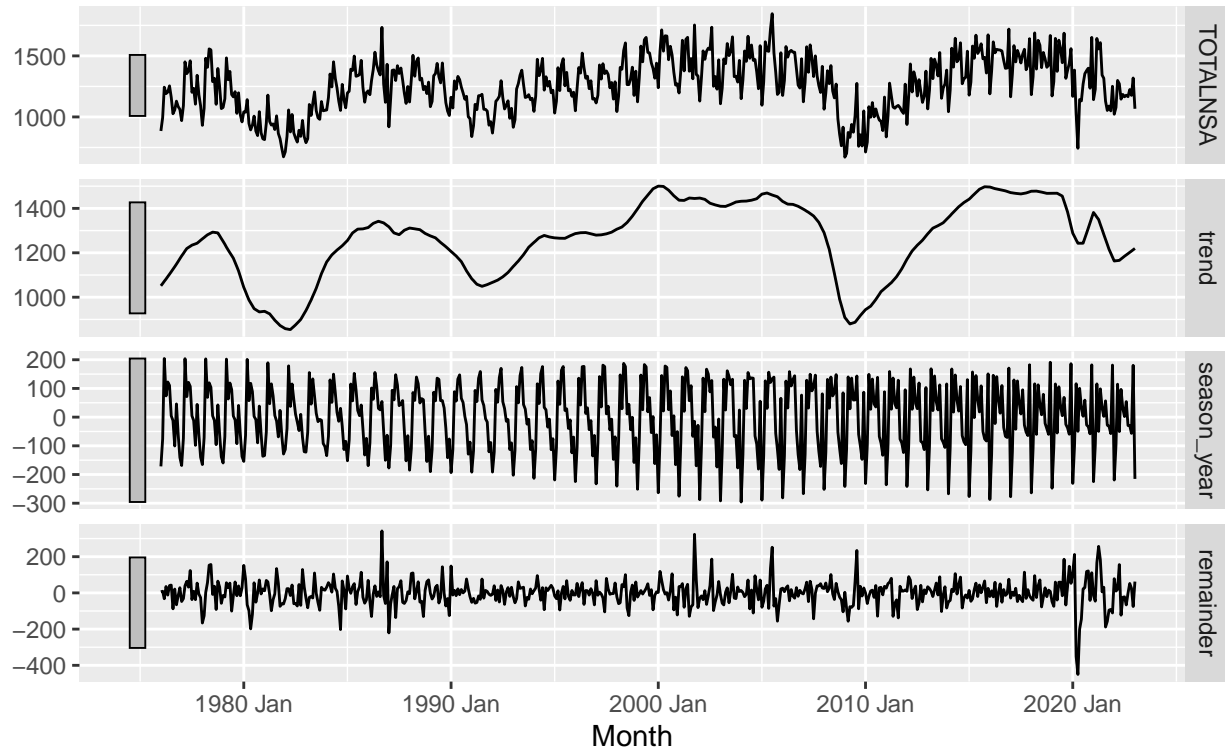


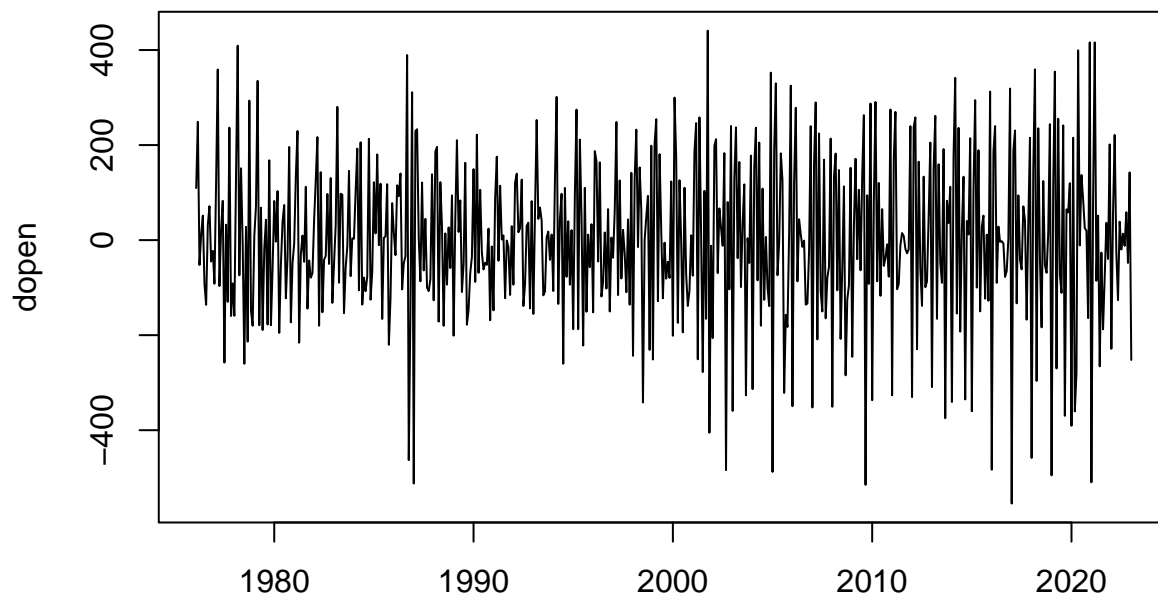
This thing better also work



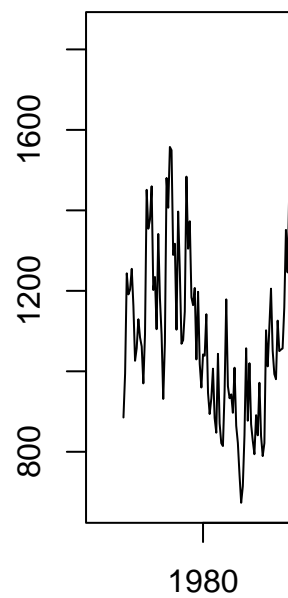
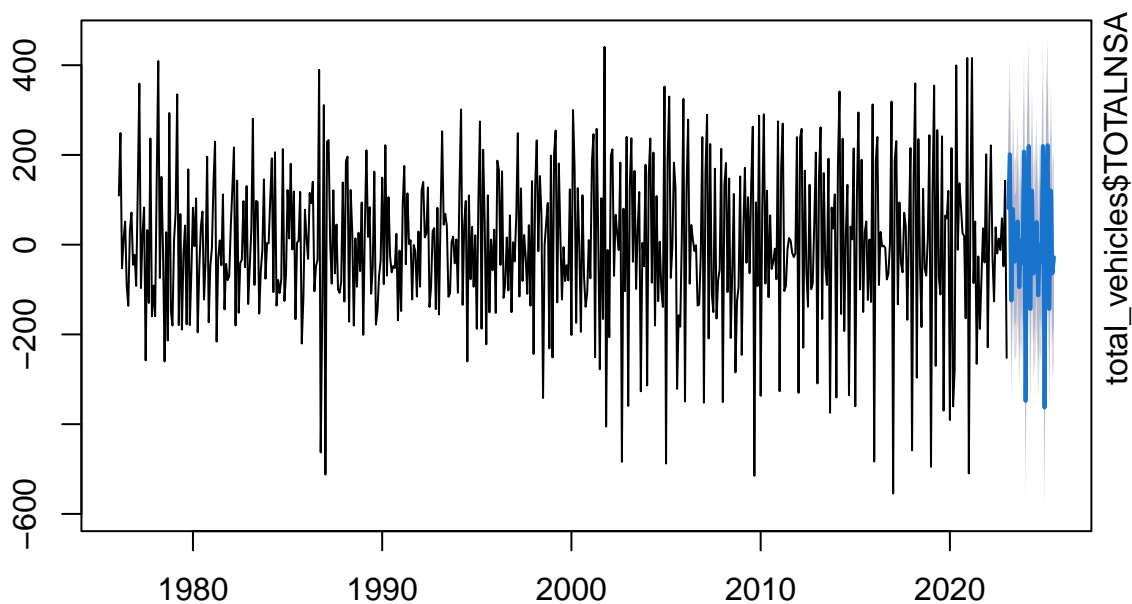
## STL decomposition

TOTALNSA = trend + season\_year + remainder



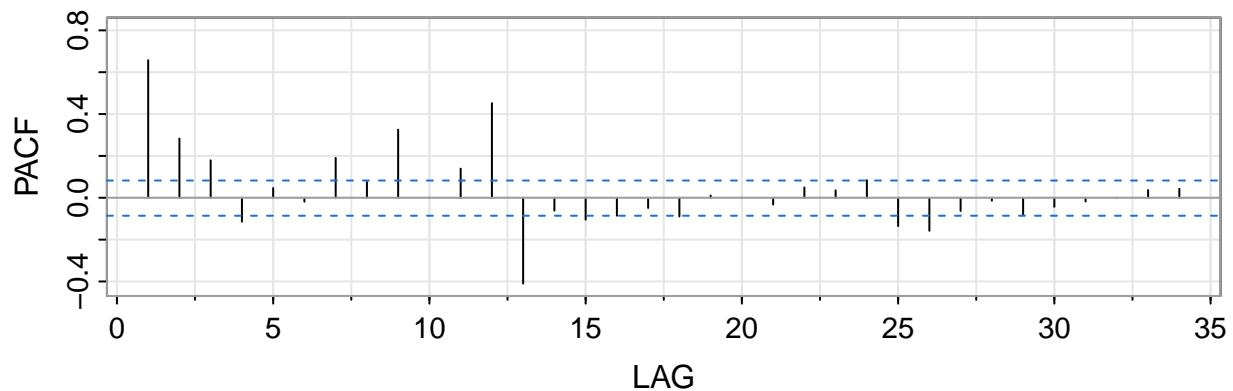
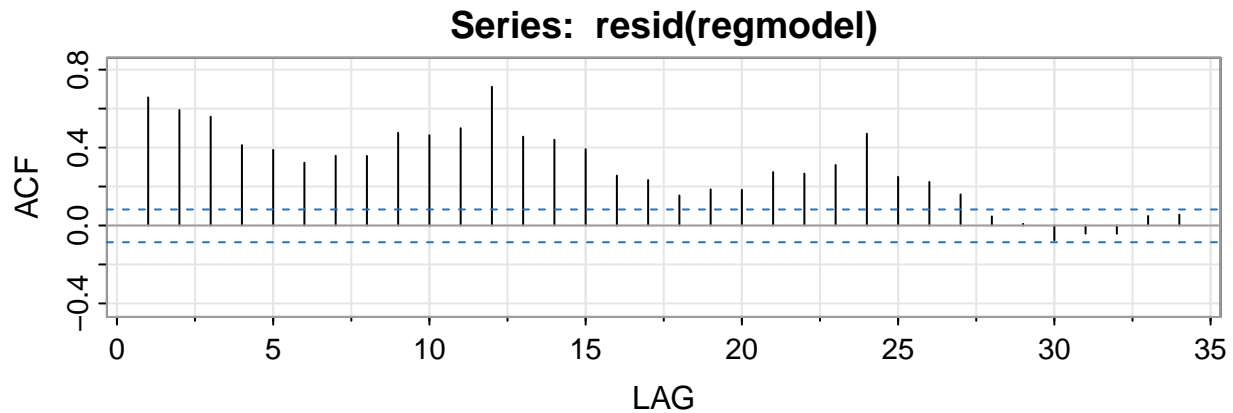


Time  
**Forecasts from ARIMA(0,0,2)(0,1,2)[12]**



```
##
## Call:
## lm(formula = total_vehicles$TOTALNSA ~ trend + trend2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -658.25 -135.85   7.76  151.29  536.50
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1296.77661   13.23092   98.011 < 2e-16 ***
```

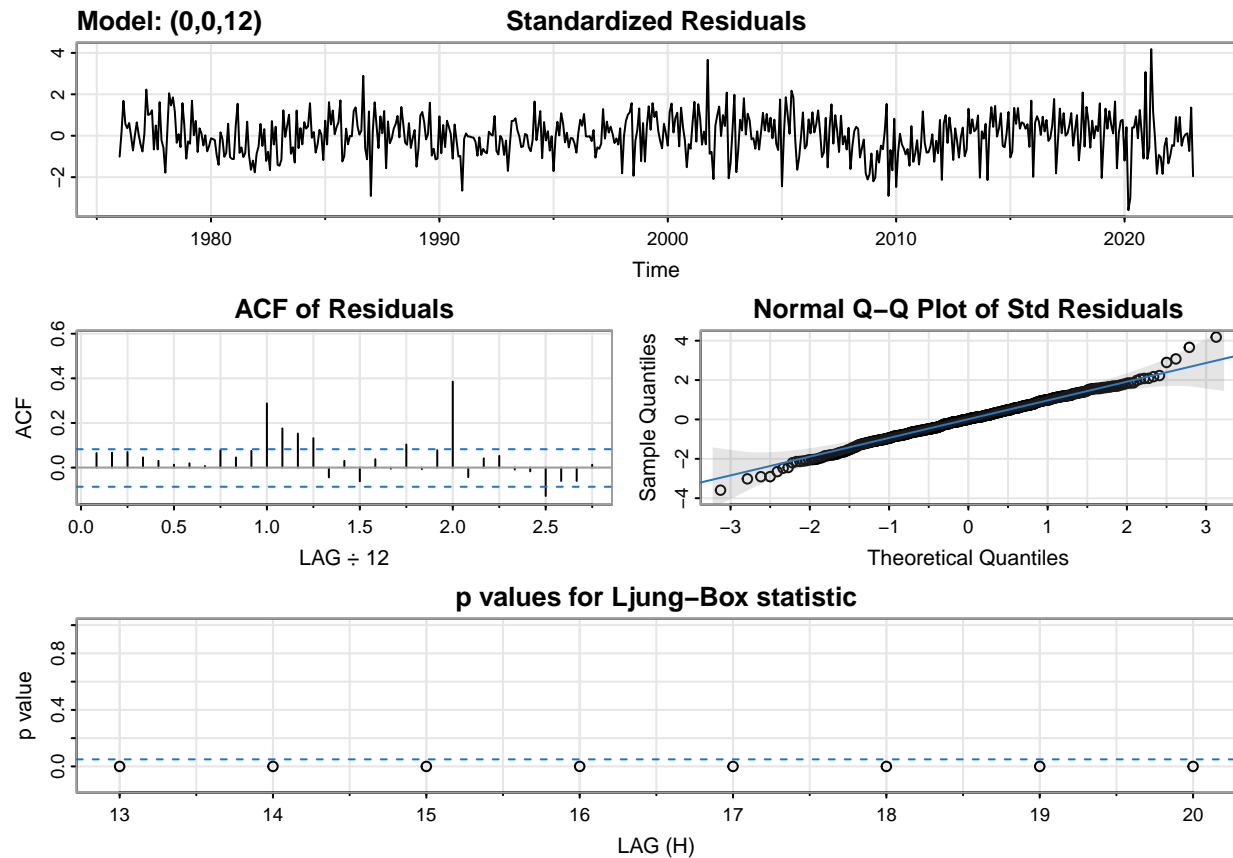
```
## trend      5.23611    0.64897    8.068 4.33e-15 ***
## trend2     -0.19722    0.05338   -3.694 0.000242 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 209.7 on 562 degrees of freedom
## Multiple R-squared:  0.1229, Adjusted R-squared:  0.1198
## F-statistic: 39.37 on 2 and 562 DF,  p-value: < 2.2e-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  0.23  0.15  0.19  0.18  0.27  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  0.22  0.16  0.05  0.01 -0.08 -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  0.04
```

```
## initial value 5.342839
## iter  2 value 4.968384
## iter  3 value 4.928125
## iter  4 value 4.919499
## iter  5 value 4.911679
## iter  6 value 4.894953
## iter  7 value 4.865550
## iter  8 value 4.852656
## iter  9 value 4.844876
```

```
## iter 10 value 4.843347
## iter 11 value 4.843013
## iter 12 value 4.842645
## iter 13 value 4.842488
## iter 14 value 4.842214
## iter 15 value 4.841870
## iter 16 value 4.841667
## iter 17 value 4.841521
## iter 18 value 4.841501
## iter 19 value 4.841495
## iter 20 value 4.841494
## iter 21 value 4.841493
## iter 21 value 4.841493
## iter 21 value 4.841493
## final value 4.841493
## converged
## initial value 4.844275
## iter 2 value 4.844247
## iter 3 value 4.844239
## iter 4 value 4.844236
## iter 5 value 4.844224
## iter 6 value 4.844218
## iter 7 value 4.844208
## iter 8 value 4.844190
## iter 9 value 4.844187
## iter 10 value 4.844185
## iter 11 value 4.844183
## iter 12 value 4.844182
## iter 13 value 4.844181
## iter 14 value 4.844181
## iter 15 value 4.844181
## iter 15 value 4.844181
## iter 15 value 4.844181
## final value 4.844181
## converged
```



	ma1	ma2	ma3	ma4	ma5
##	0.49898538	0.34322417	0.35247943	0.17827417	0.15484370
	ma6	ma7	ma8	ma9	ma10
##	0.05423734	0.10057467	0.05637438	0.19052984	0.16837433
	ma11	ma12	intercept	trend	trend2
##	0.08530509	0.49815898	1298.42481923	5.18436337	-0.21160105

## References

- “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](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/>.
- “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](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).