Forecasting Project 1

Analysis of Amarillo, TX
and
Sacramento, CA Data Sets

Advanced Business Forecasting
CIDM – 6362- 70
Dr. Barthel
Spring 2024

Research and Analysis Completed By:
Miguel M. Sanchez

Section 1

       The two regions I will be analyzing within this forecasting project are the Amarillo, Texas and Sacramento, California regions. Amarillo, Texas is known as being one of the primary hubs of the Cattle Industry in Texas. Amarillo is home to various cattle packing plants, ranches, feeding lots and even auctions. Amarillo's location also makes it a primary transportation route and for logistically speaking a major artery for the cattle industry. Amarillo is also home to very strong agriculture, oil and gas, and even manufacturing sectors. With Amarillo being located within the Texas Panhandle (covering 25,000 miles) its location is not only centralized geographically but also has various transportation routes leading directly into Amarillo. Texas is also the second largest state in the continental United States which allows the state to have a vast array of landscapes, habitats and even climate/temperatures.

       The second region I will be analyzing is the Sacramento, California. Sacramento is considered the agriculture capital of California. Sacramento's economy also consists of government (due to it being the capital of California), technology and even healthcare, to name a few. The single largest contributor to Sacramento's economy is their massive wine industry which comes from the grape production of the city and state. Sacramento is known for having nearly all seasons from summer, spring, fall and winter with varying temperatures and climates. In addition, Sacramento is also considered an economic hub for California and other surrounding cities/states.

       The similarities of both Amarillo, Texas and Sacramento, California include the fact that they are both considered economic nucleuses for their respective states, and they are also the leaders of their respective industries (Cattle for Amarillo and Grape Production for Sacramento). As for their differences, their climates are vastly different with Amarillo being primarily arid and experiencing extreme temperatures while Sacramento experiences all seasons and less arid seasons. The population of Amarillo and Sacramento is vastly different with Sacramento being more densely populated. The population figures for Sacramento are 524,943 while Amarillo has a population of 199,371 (figures as of 2022).
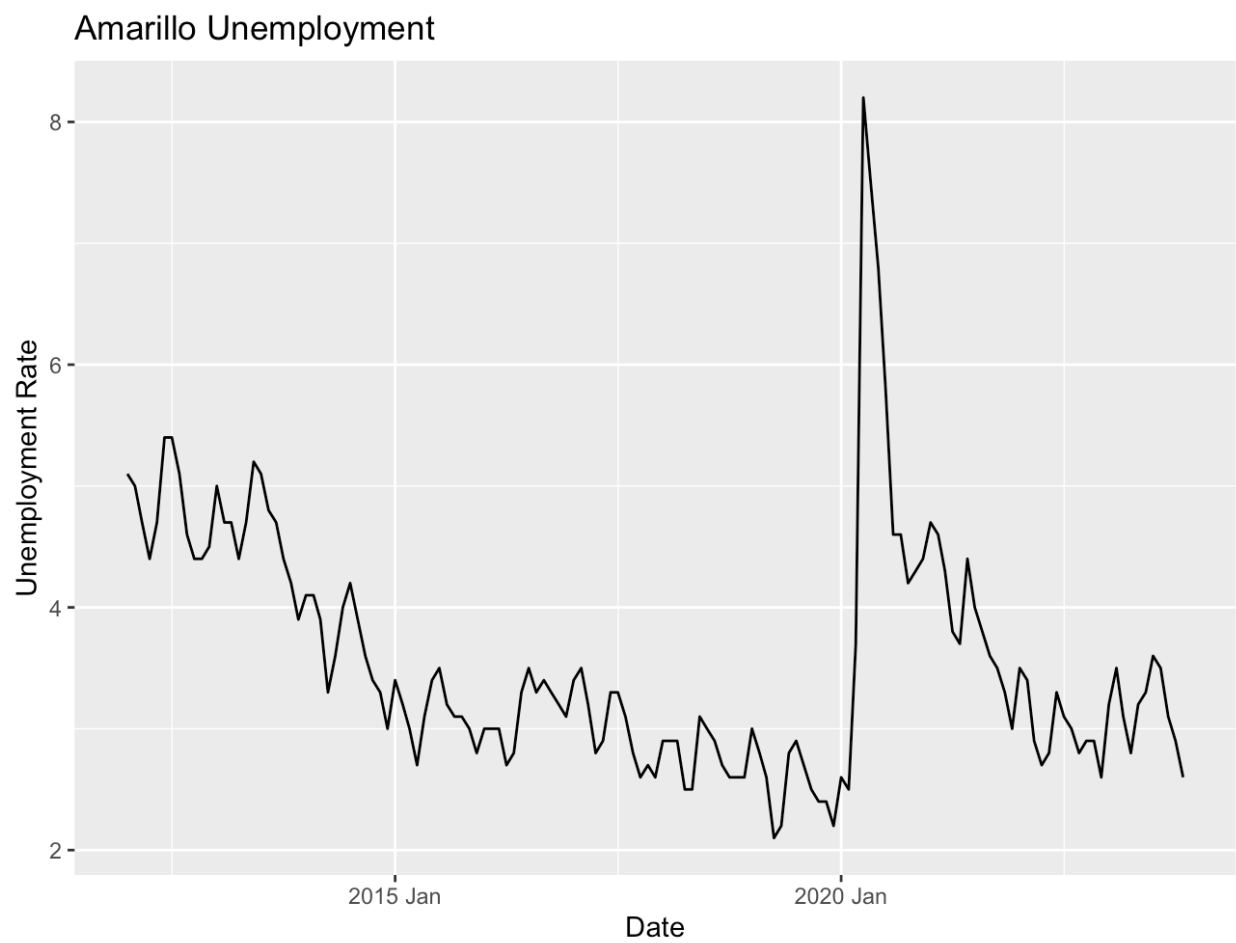
Section 2
Plot the data for all 6 time series. Make sure to label your graphs (axes labels and title for
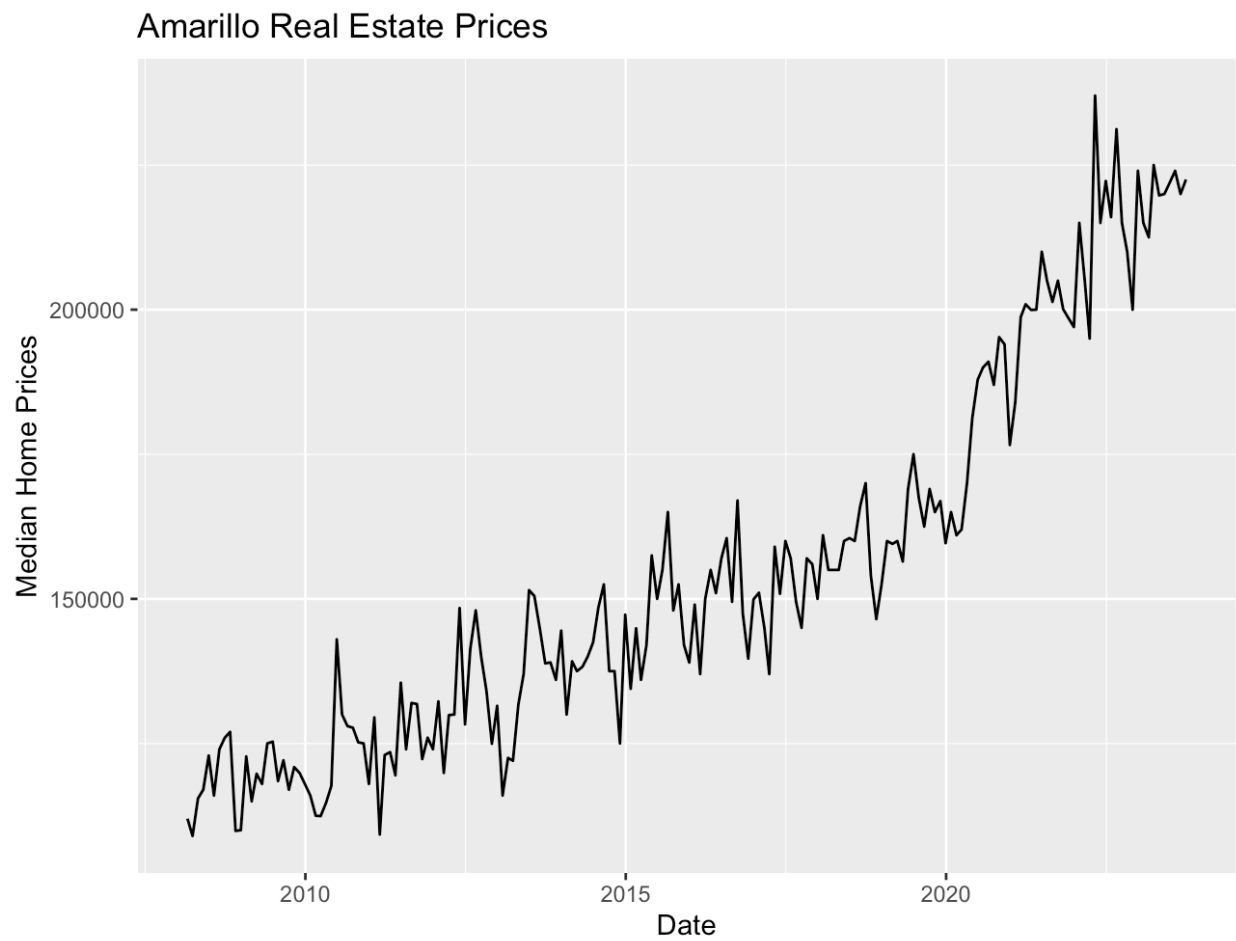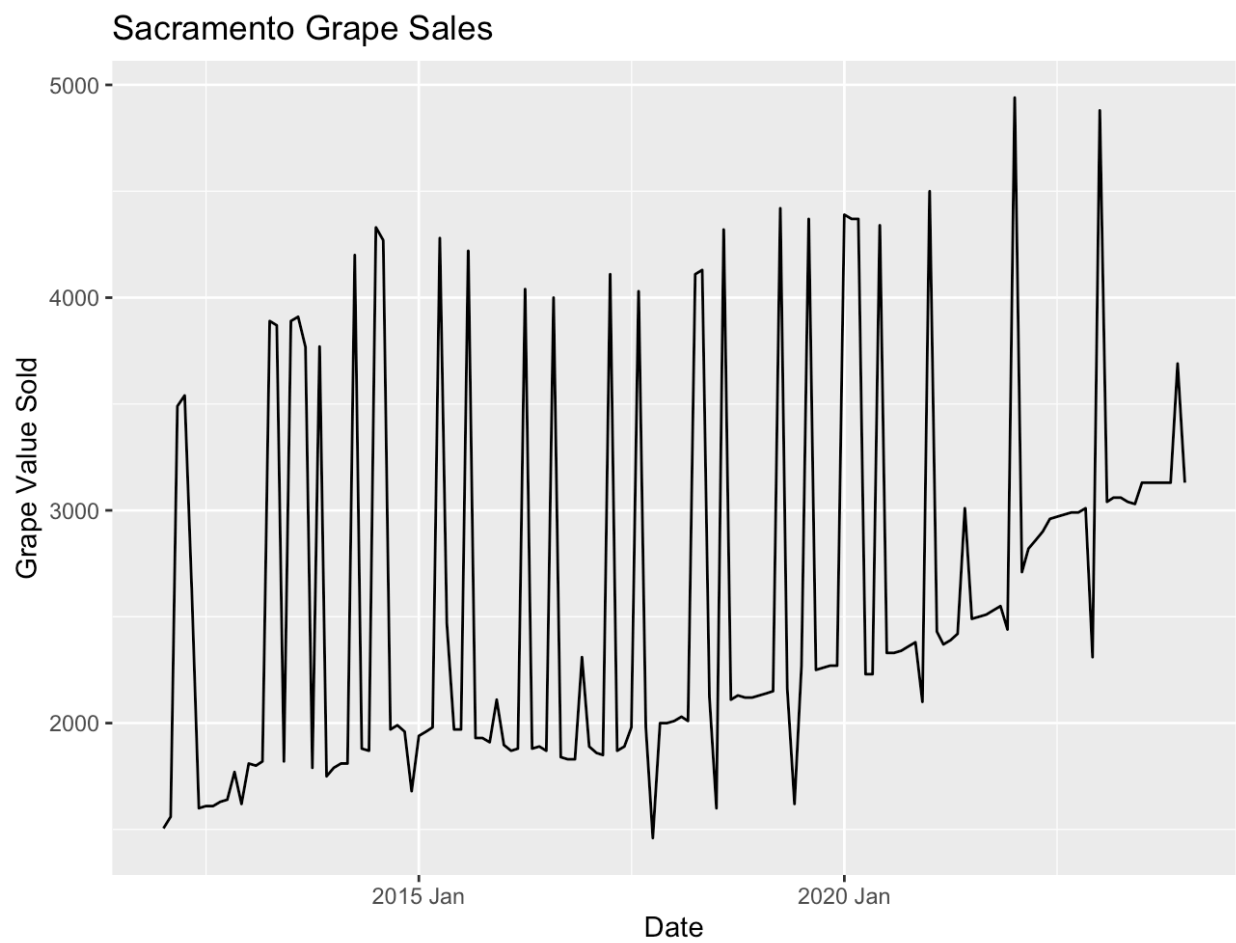each plot).
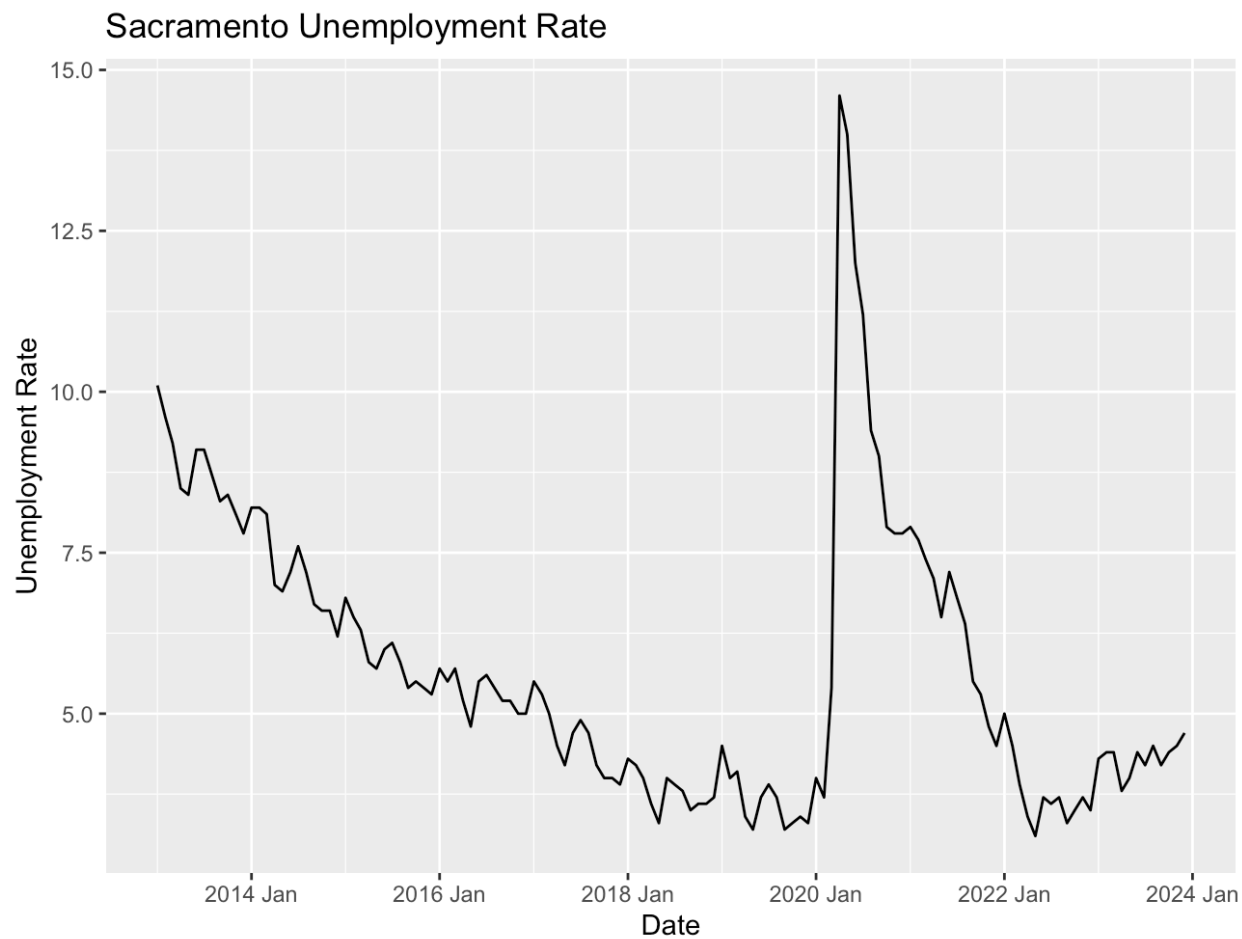
Amarillo Cattle Prices
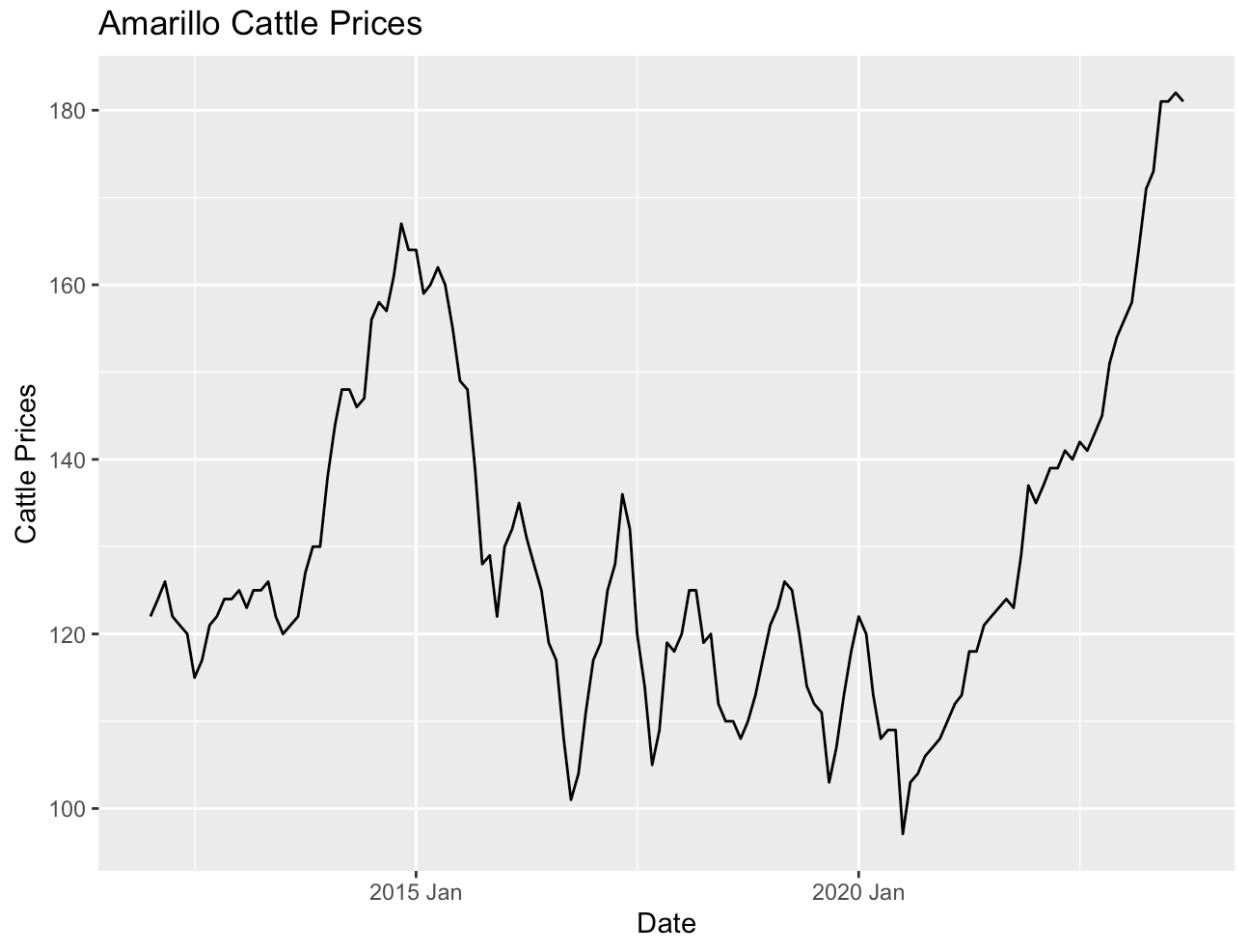
Amarillo Unemployment Rate

Amarillo Unemployment

Amarillo Real Estate Prices

Amarillo Real Estate Prices

Sacramento Grape Sales

Sacramento Unemployment Rate

Sacramento Unemployment Rate

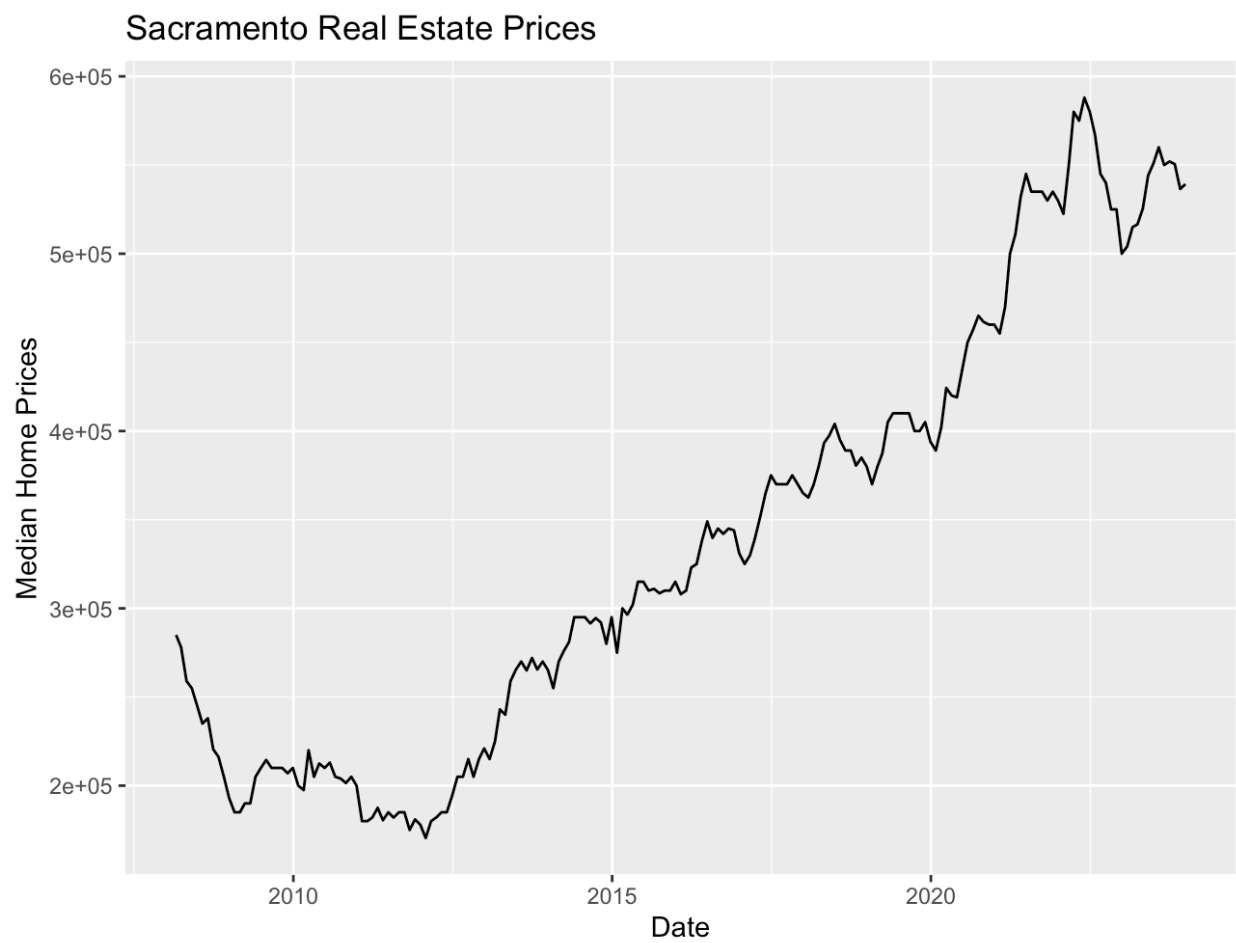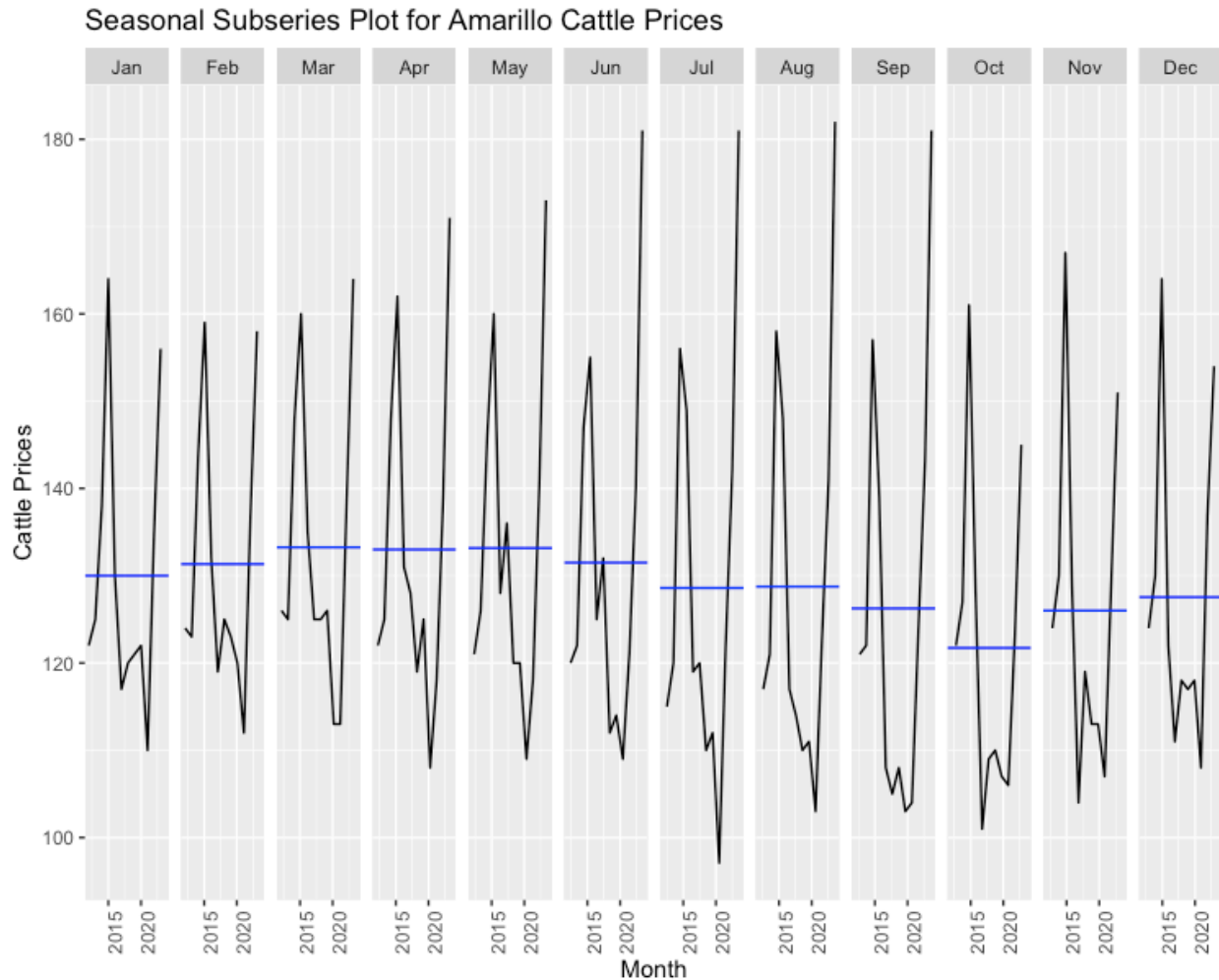Sacramento Real Estate Prices


Sacramento Real Estate Prices

Section 3
Using appropriate tools such as seasonal plots, seasonal subseries plots, or correlograms,
investigate the time series patterns (trend, seasonality and cycles). Include graphs and
describe your findings.

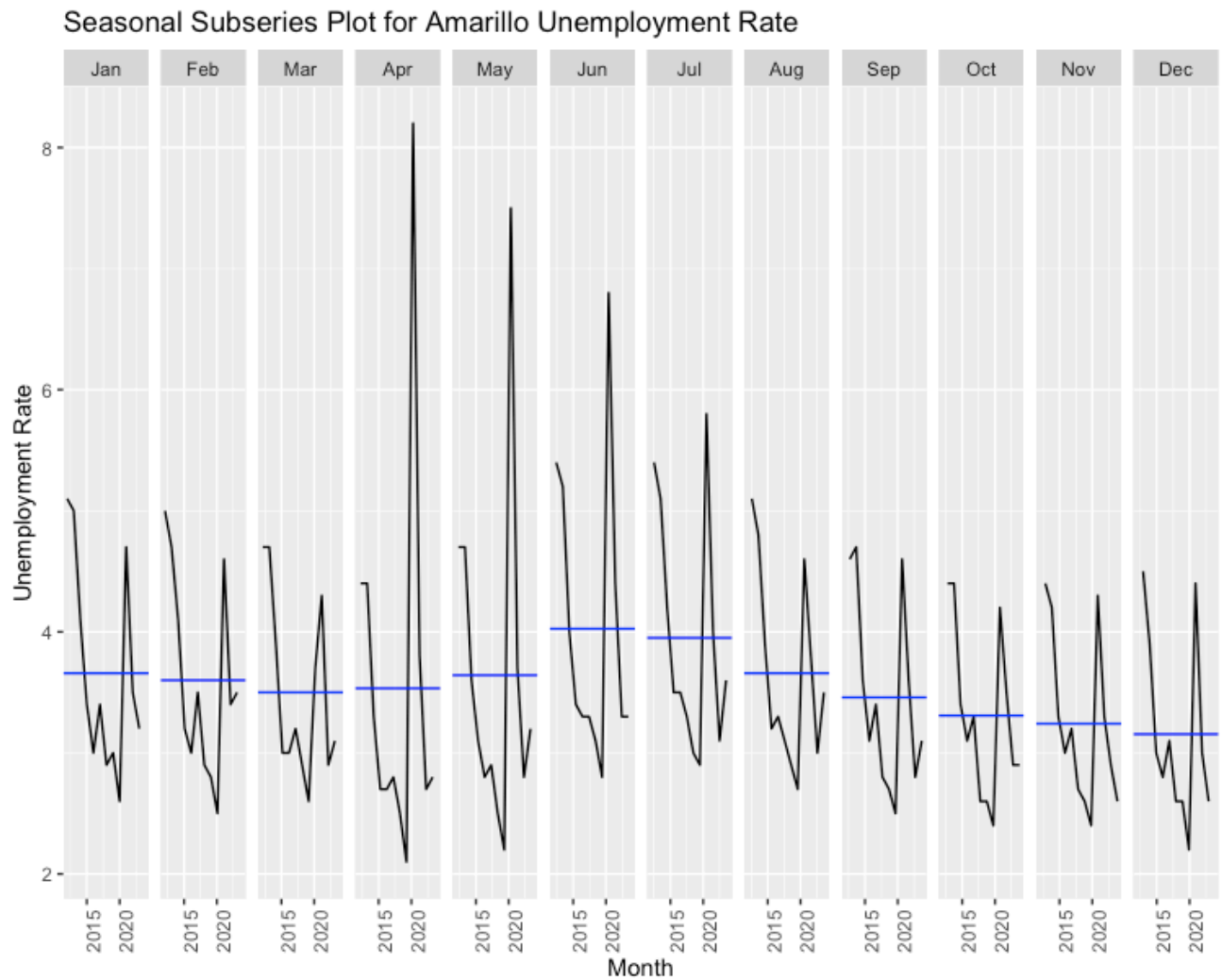Seasonal subseries plot for Amarillo Cattle Prices

When reviewing the Seasonal Subseries plot for Amarillo Cattle prices the
"seasonal average" (the blue line within each month / subseries) shows the average pricing
of that month. Per the data provided it appears that the months where cattle prices are the
highest are during the months of March, April and May. While the lowest months for cattle
prices are September, October and November.

ACF of Amarillo Cattle Price

ACF of Amarillo Cattle Price
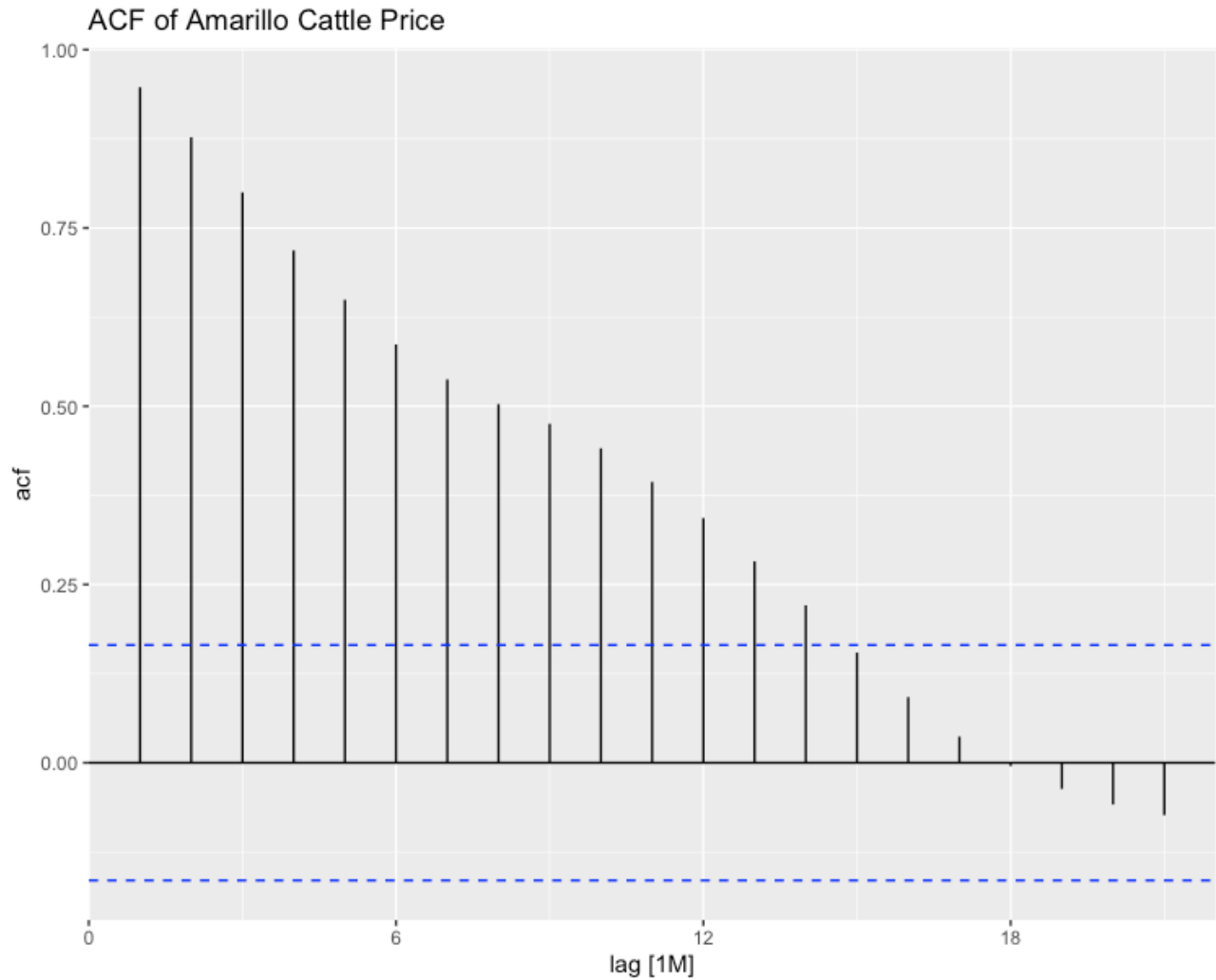
The Auto Correlation Function plot allows us to measure the relationship between lagged values and its corresponding time series, this will be a linear relationship. There appears to be a longer-term trend every 18 months where the price of Cattle appears to have very little to no correlation to the time series before it. Meaning that the cattle prices do not have an impact on each other from a historical sense.

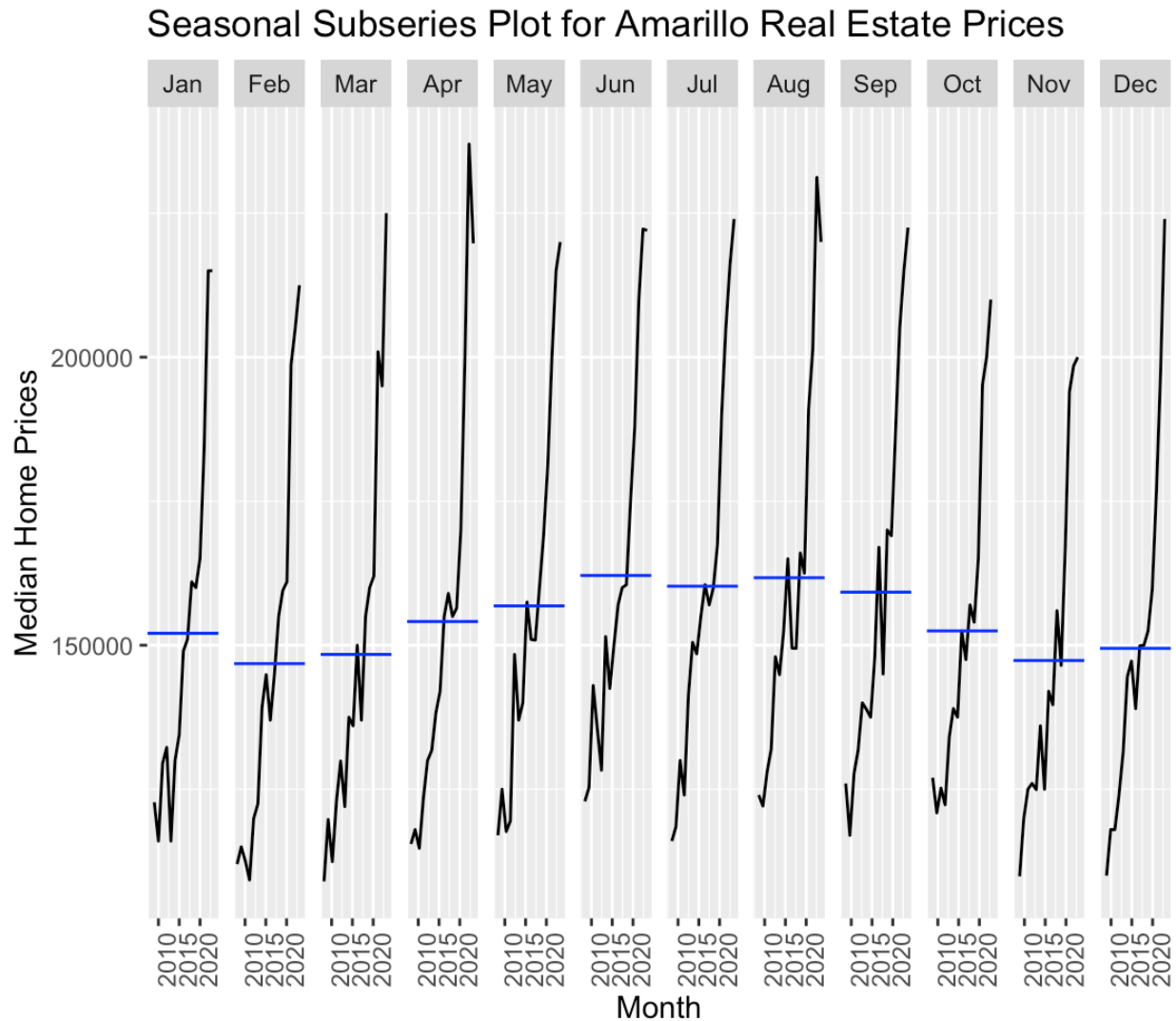Seasonal Subseries Plot for Amarillo Unemployment Rate

When reviewing the Seasonal Subseries plot for Amarillo Unemployment Rates the "seasonal average" (the blue line within each month / subseries) shows the average unemployment rate of that month. Per the data provided it appears that the months where the unemployment rate is the highest are during the months of June and July.

ACF of Amarillo Unemployment Rate

When reviewing the ACF for Amarillo Unemployment there is a linear decrease of the ACF going into lag 18. It appears that every 18 years there seems to be a trend or pattern where every 18 months the unemployment rate increases from a trough to a peak. There is a diminishing value of correlation between the unemployment rate and its next lag period. Meaning long term historical values do not impact the overall value of the Unemployment rate.

Seasonal subseries – Amarillo Real Estate

Seasonal Subseries Plot for Amarillo Real Estate Prices

When reviewing the Seasonal Subseries plot for the Amarillo Housing Market the "seasonal average" (the blue line within each month / subseries) shows the average Median Home price for that month. Per the data provided it appears that the months where the Median Home Prices were at their highest are during the months of June through September. The lowest months are February, March and November.

ACF of Amarillo Real Estate Prices
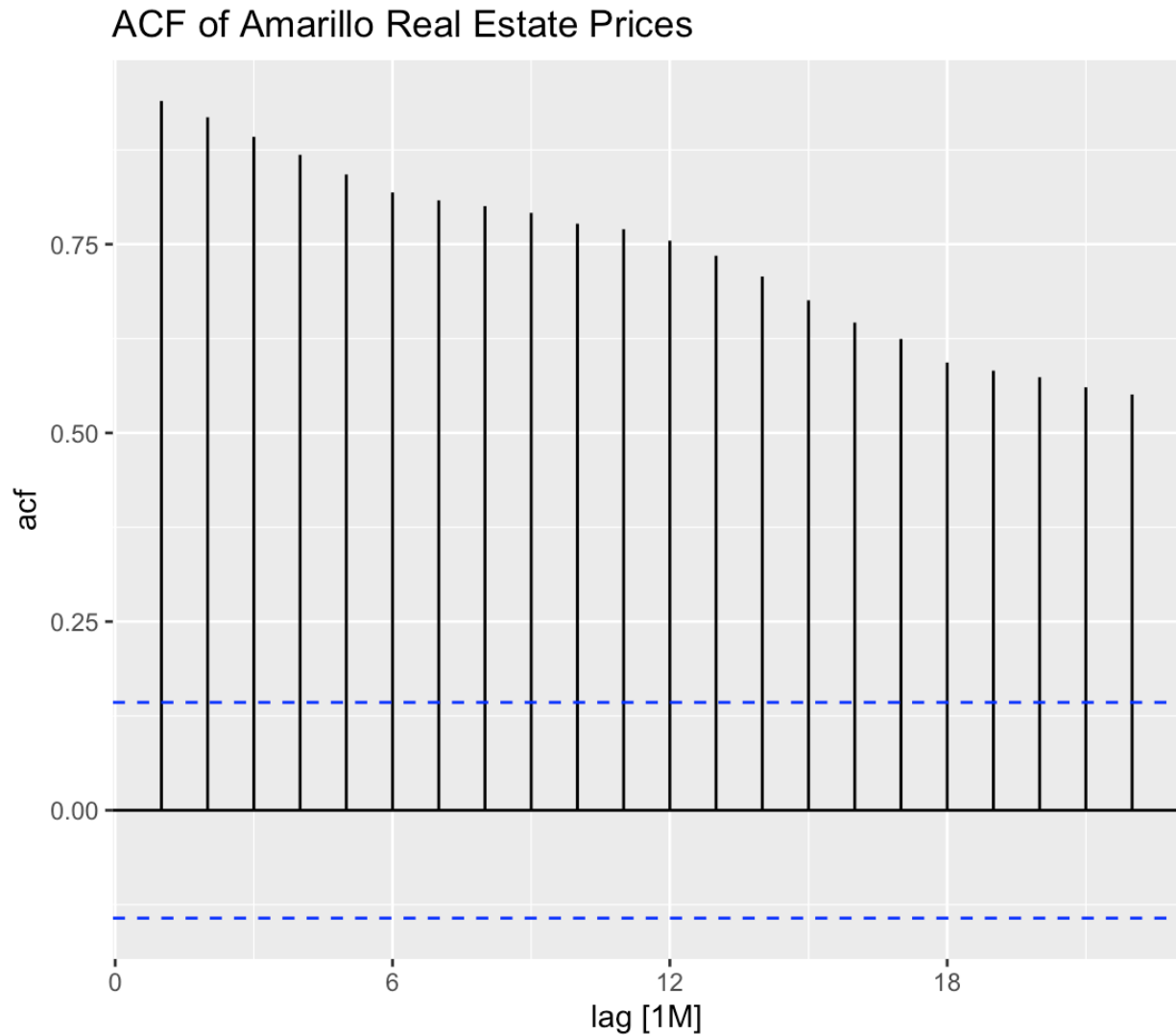
There is a Positive Auto Correlation between each median home price at different lags. Past values of the time series have a positive correlation on future values. This is true because Median Home Prices and real estate typically continue to grow year over year. Prior years impact the future pricing of homes because historical values do influence the present and future values of this specific dataset and series. This is also true for the Amarillo housing market where there also appears to be some seasonality due to the linear and slight curves that take place during lag 7 and going into lag 11 and 12. While the curvature is minimal there is some seasonality which is true within the real estate market in general.
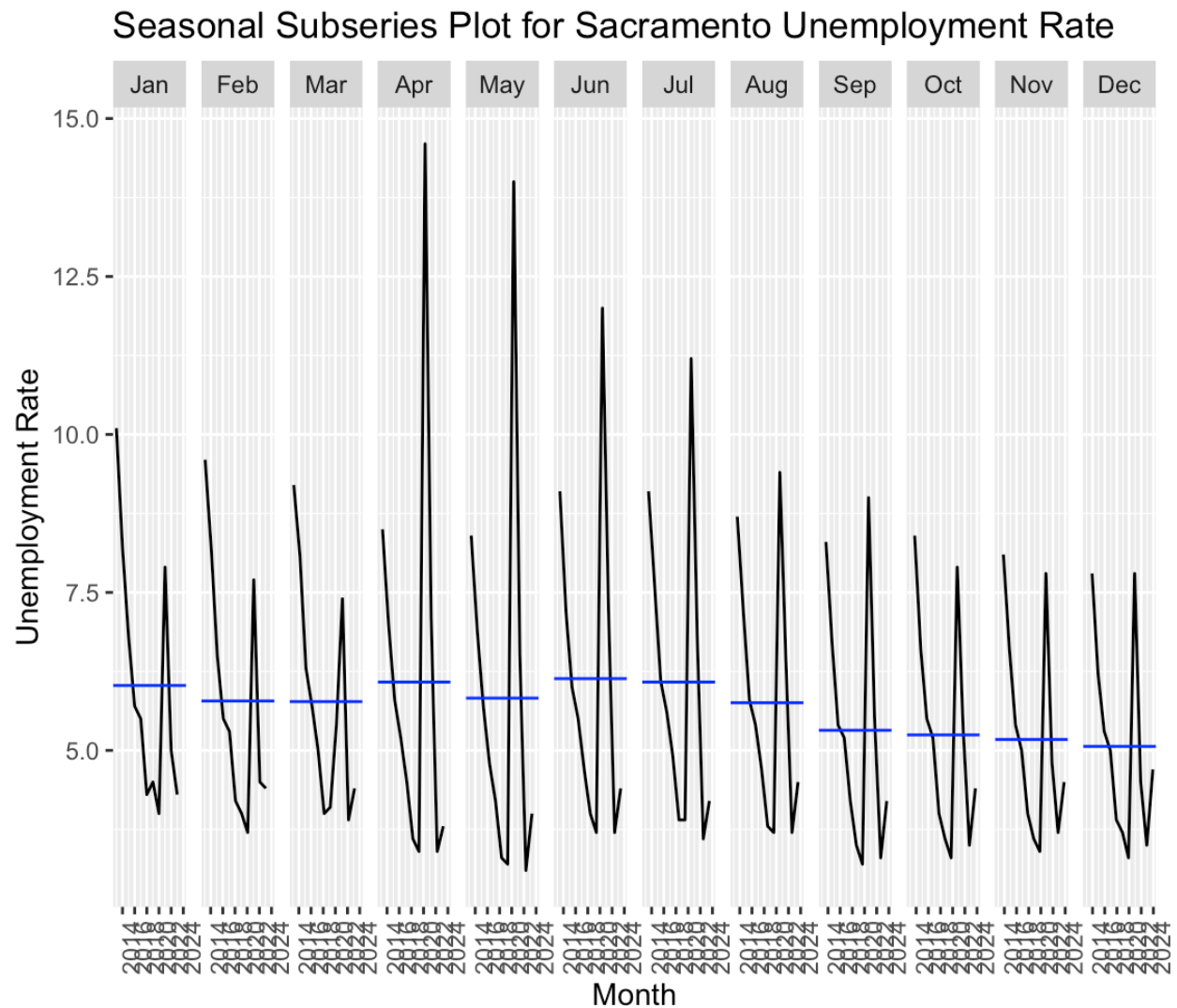
Seasonal Subseries – Sacramento Grapes
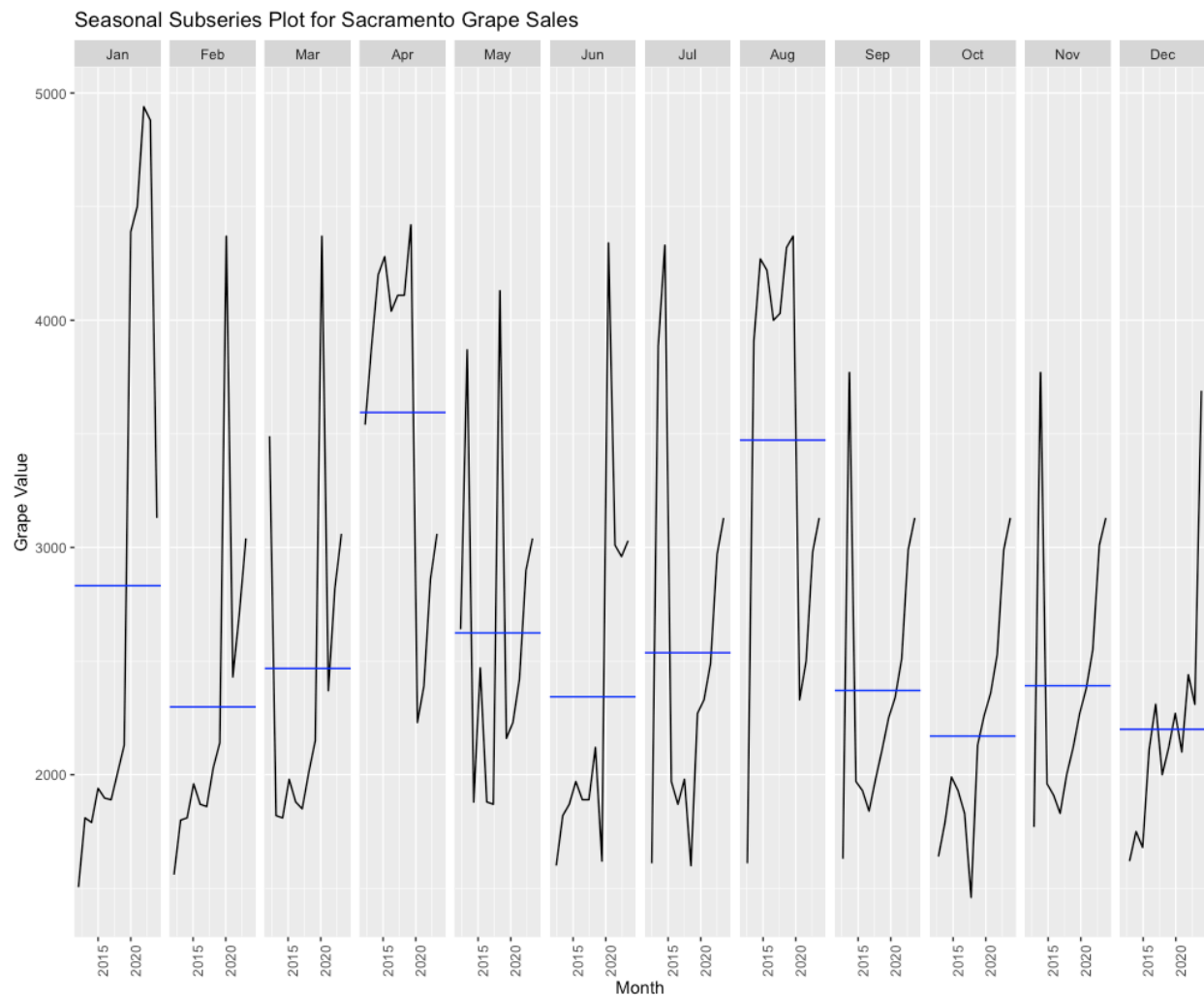
When reviewing the Seasonal Subseries plot for Sacramento Grape Values the "seasonal average" (the blue line within each month / subseries) shows the average Grape value for each month. Per the data provided it appears that the months where the values of grapes were at their highest are during the months of April and August. The lowest months are October and December. This makes sense since the production of grapes peak during the months they are growing and blossoming, hence production increases. This is critical for the wine production for the state of California.

## ACF of Sacramento Grape Sales
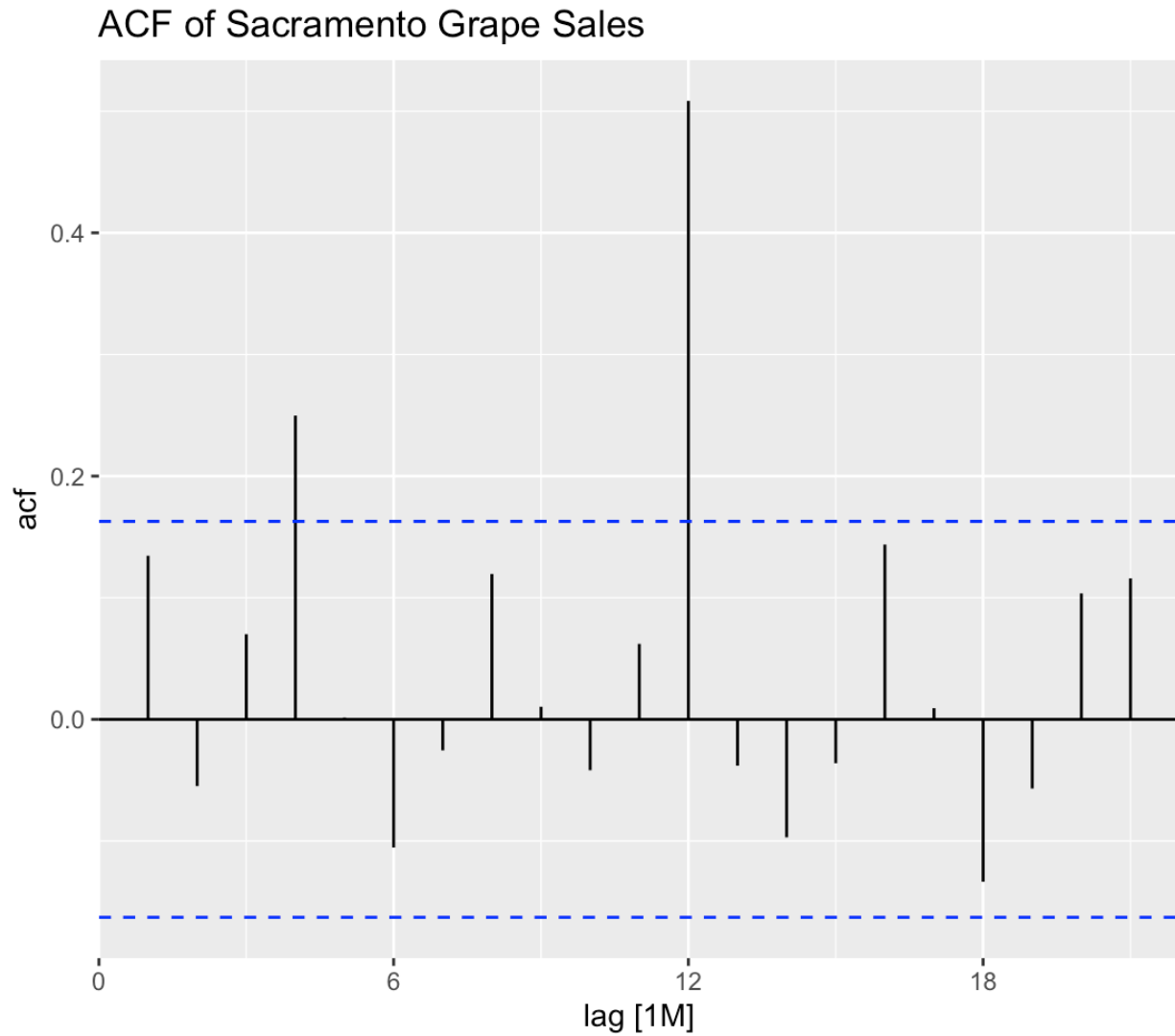
This ACF shoes highly seasonal trends in the value/sales of grapes within the Sacramento Market. There is a positive ACF value at lags 4 and 12 which indicate periodic fluctuations recurring at both lags. As for all other lags these are noise and irregularities within the data reviewed and can be ignored.

Seasonal Subseries – Sacramento Unemployment

Seasonal Subseries Plot for Sacramento Unemployment Rate

When reviewing the Seasonal Subseries plot for the Sacramento Unemployment Rate the "seasonal average" (the blue line within each month / subseries) shows the average Unemployment rate for each month. Per the data provided it appears that the months where unemployment was the highest were during the months of June and July. The lowest months are November and December.

ACF of Sacramento Unemployment Rate
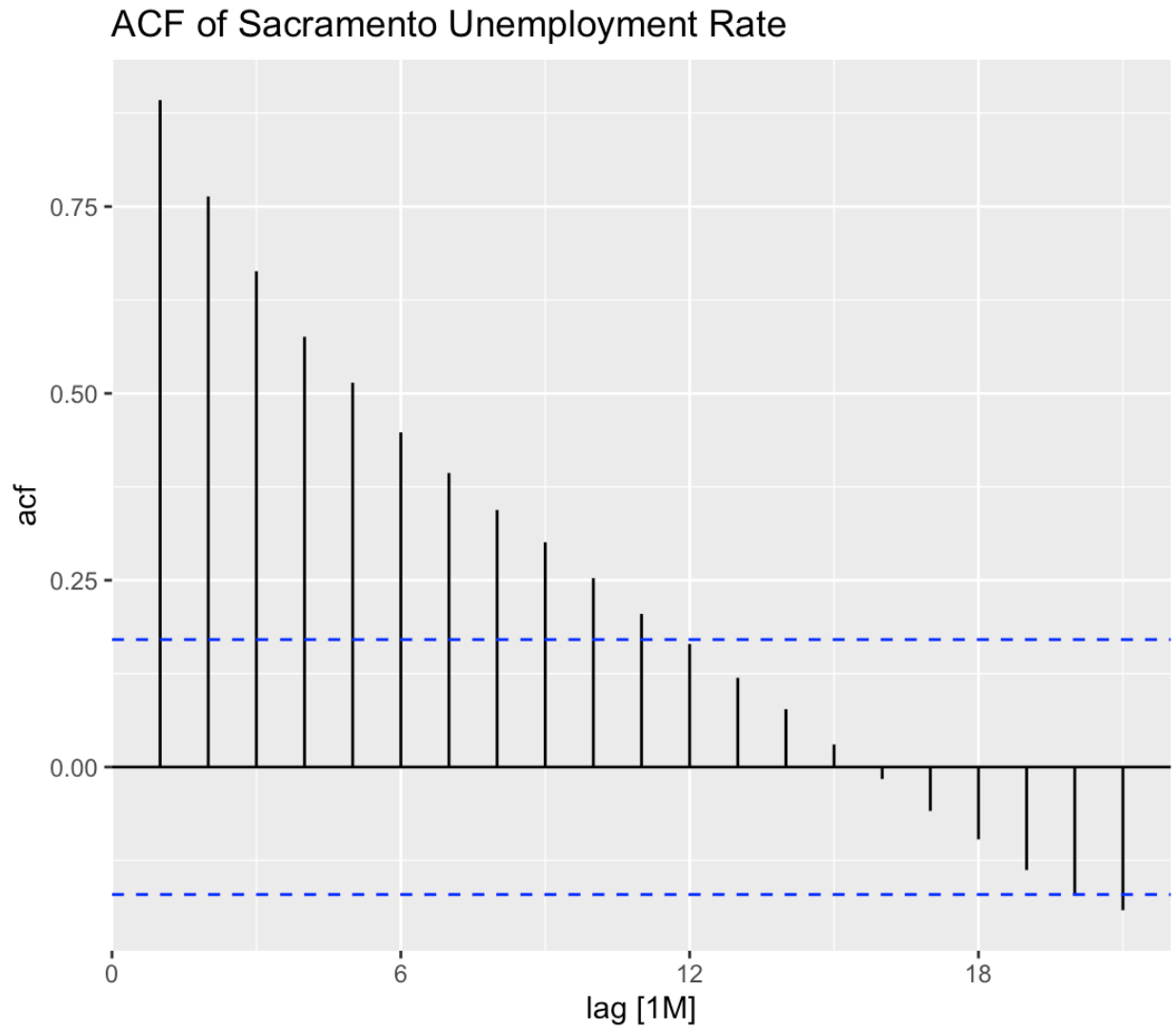
It appears that every 15 periods there may be some seasonal or periodic pattern due to the 15th lag leveling off. From what it appears there is very little to no correlation between the unemployment rate and the lagged values as it increases over time. By lag 15 there is no influence on previous values on the current unemployment rate after lag 15.

Seasonal subseries – Sacramento Housing

Seasonal Subseries Plot for Sacramento Real Estate Prices
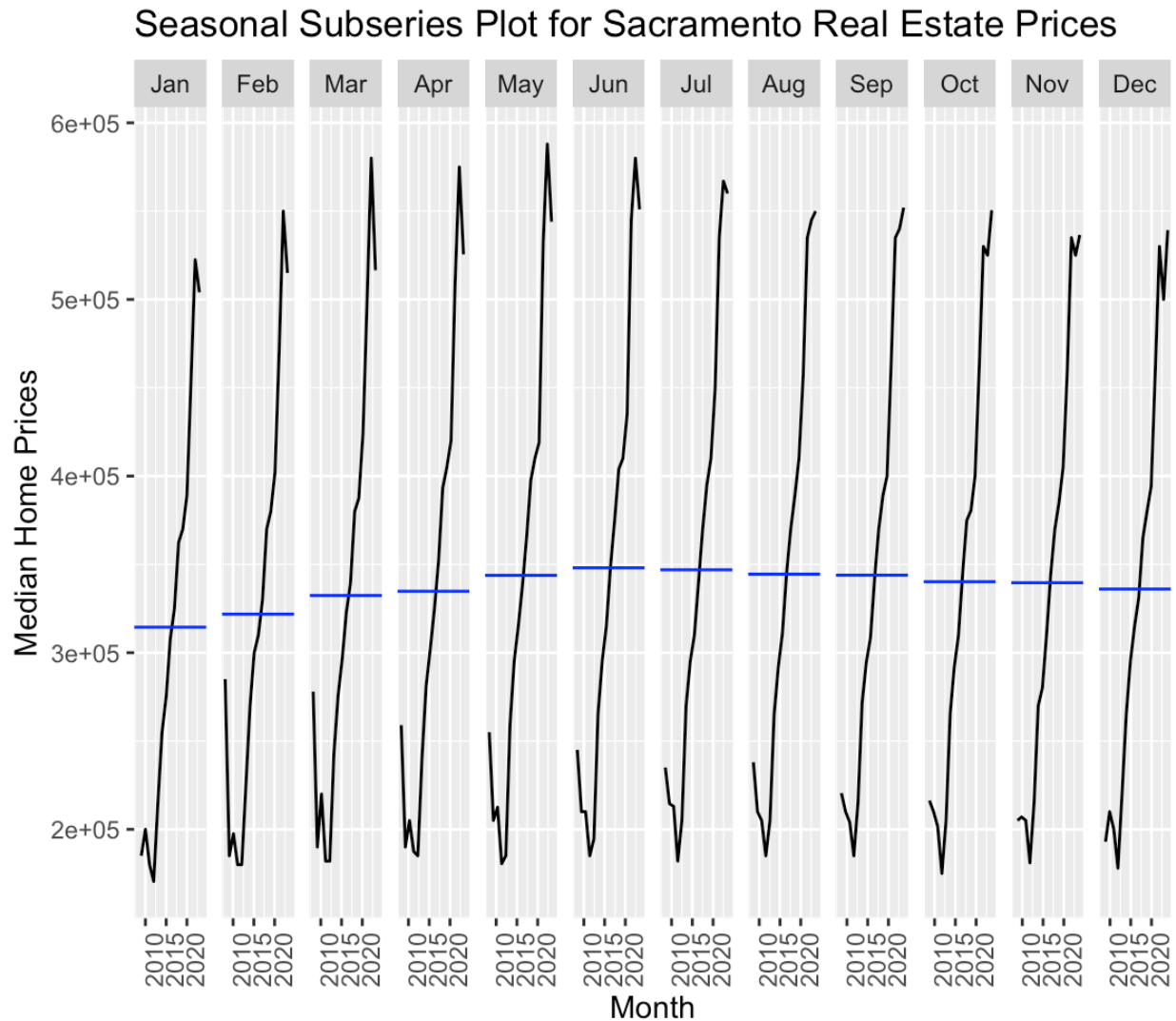
It appears that the Sacramento Housing Market is relatively stable month over month. There appears to be lower median home prices hence sales in the months of January – March but all other months remain relatively flat. There is not much seasonality among the other months excluding the first quarter of every year within this subseries.

ACF – Sacramento Real Estate Prices


ACF of Sacramento Real Estate Prices
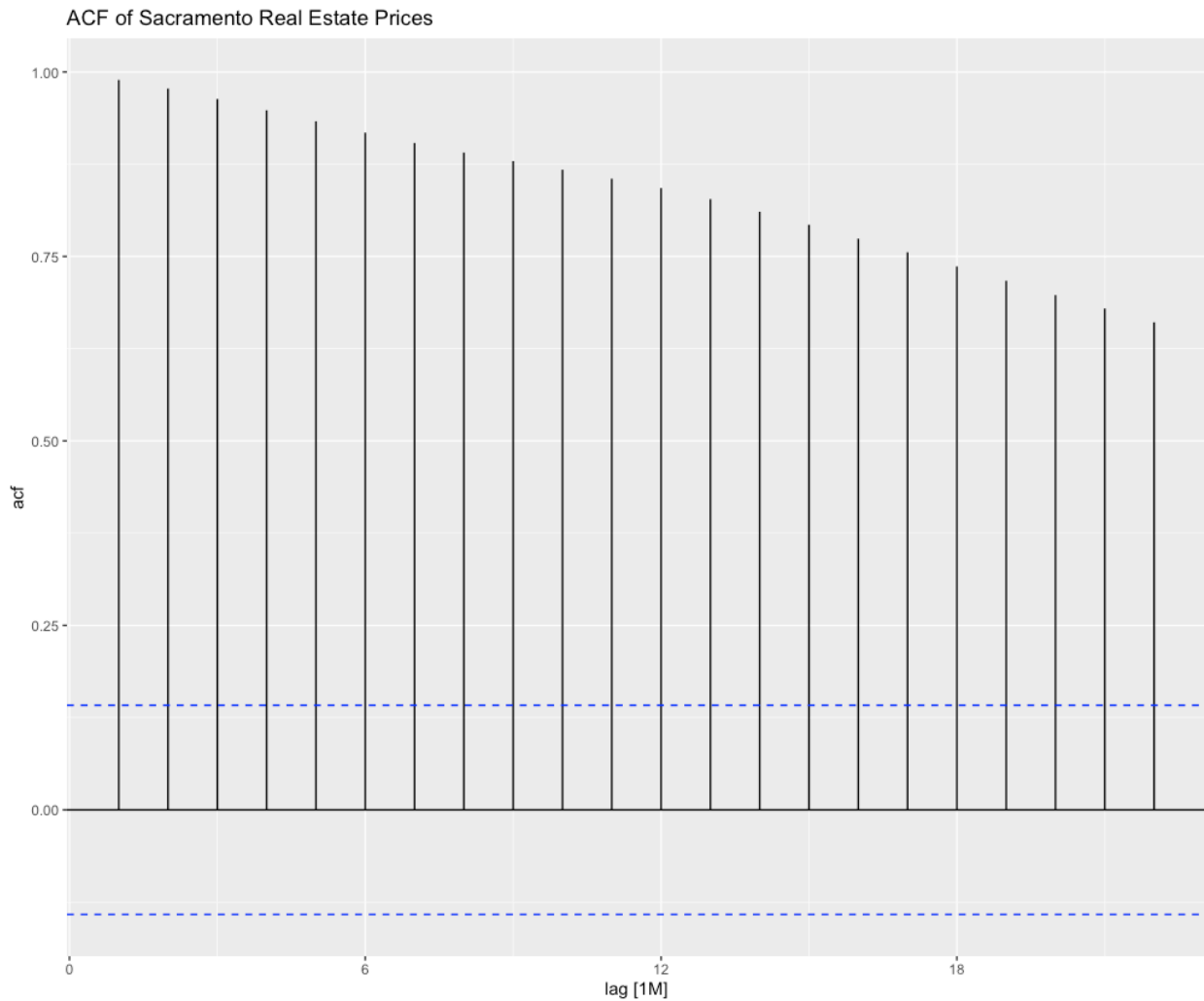
It appears that there is a Positive Auto Correlation between each median home price at different lags. Past values of the time series have a positive correlation on future values. This is true because Median Home Prices and real estate typically continue to grow year over year. Prior years impact the future pricing of homes because historical values do influence the present and future values of this specific dataset and series.

Section 4
Using appropriate models, obtain a suitable time series decomposition for each time series. Justify your choice of decomposition method and include a graph of the result.

  For the Amarillo Cattle Prices Market I am choosing the Multiplicative decomposition time series due to this time series providing the most stable components in my opinion. All other time series create irregularities in the data while the Multiplicative decomposition shows the trend increases and decreases within the time series more accurately.

Multiplicative decomposition

Classical multiplicative decomposition of total Amarillo Cattle Prices
Cattle = trend * seasonal * random
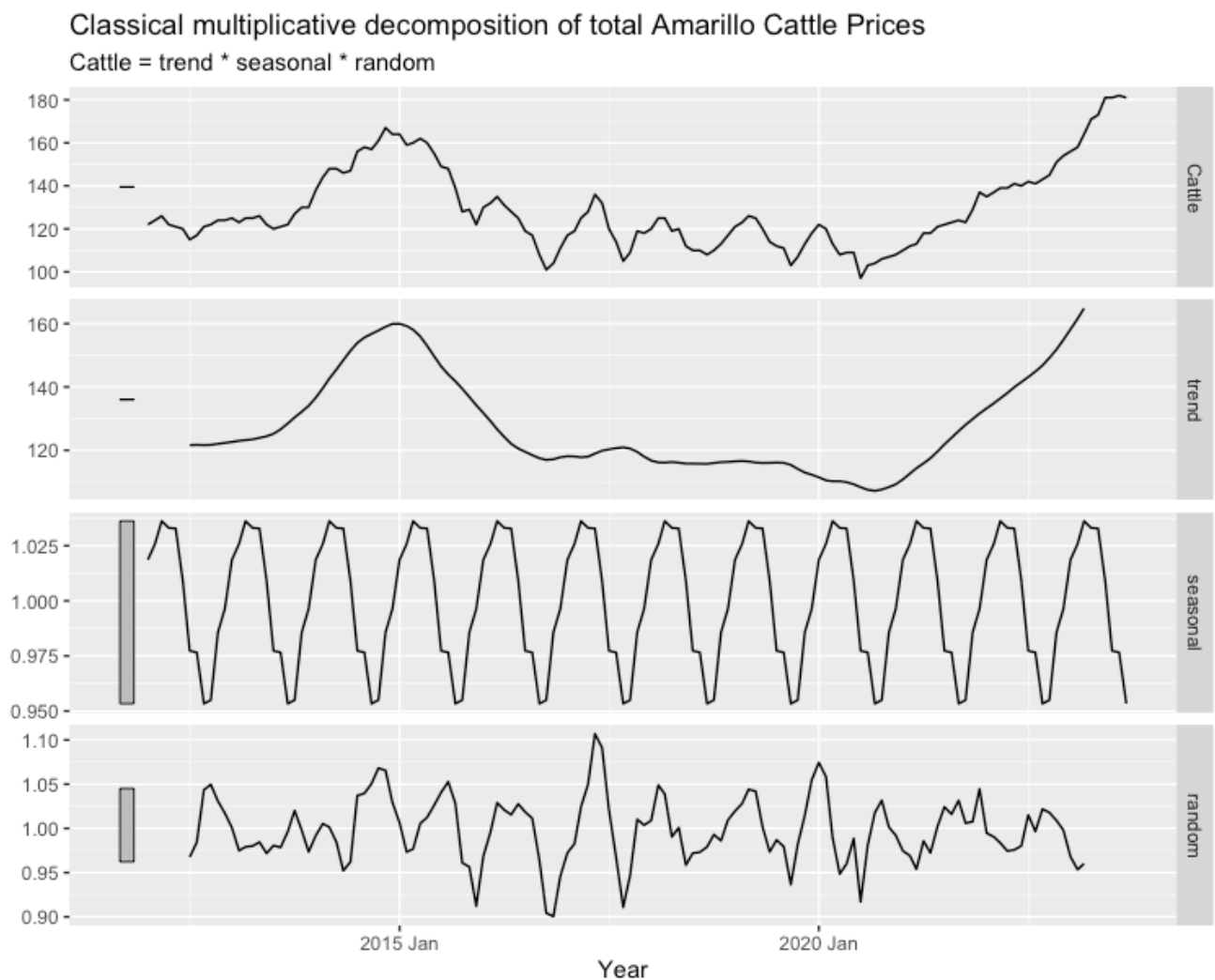
X-13 Arima Seats Decomposition Amarillo Unemployment

When reviewing the Amarillo Unemployment Rate data set the best time-series to tackle this data is the X-13 ARIMA-SEATS decomposition time series. Per our textbook this is a more robust and advanced version of the X-11 time series decomposition that was created by Bank of Spain and is widely used by government agencies around the world. It handles better the adjustments to seasonality and provides a more accurate result.

X-13ARIMA-SEATS decomposition
Unemp_rate = f(trend, seasonal, irregular)
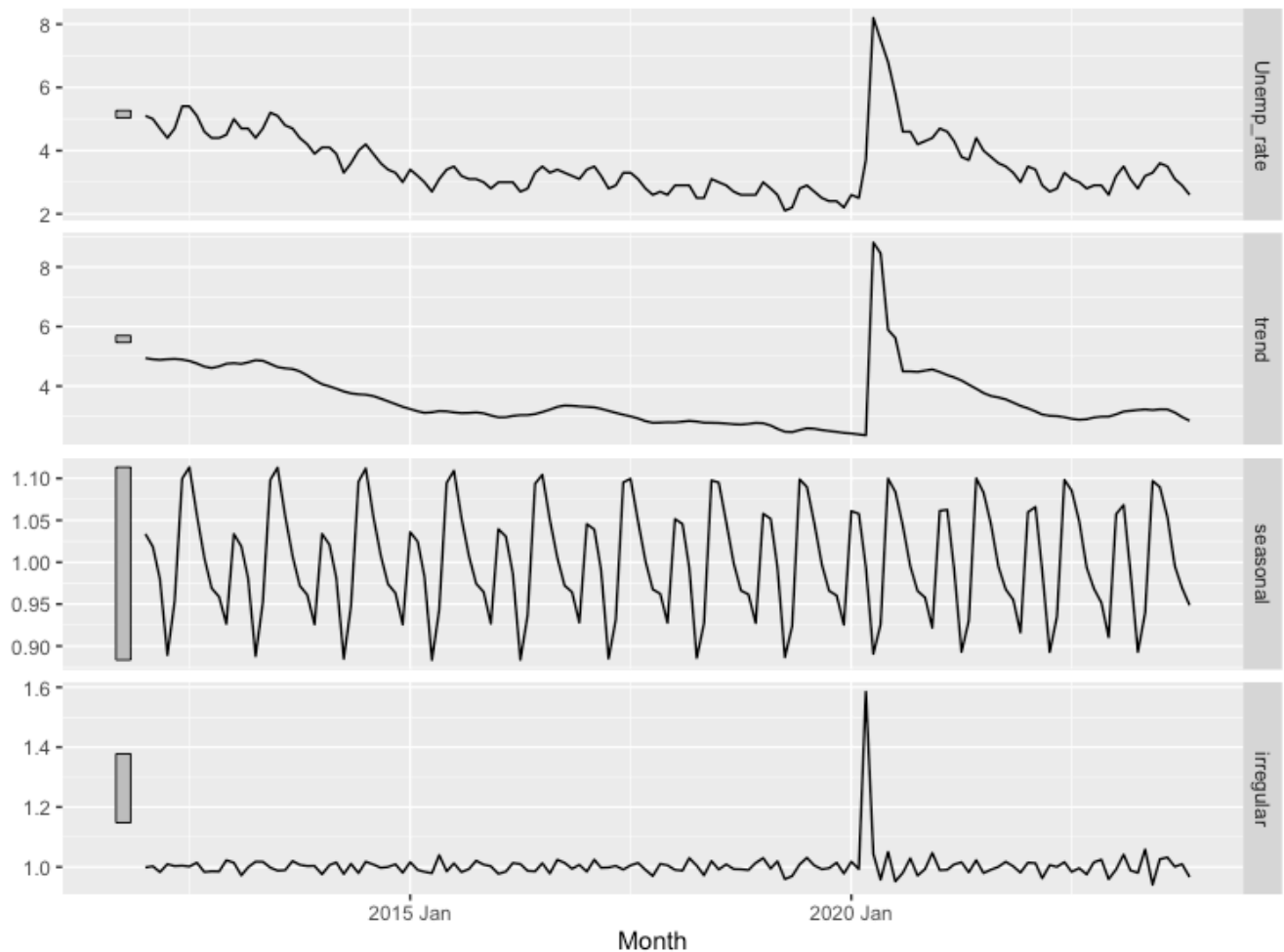
X-11 Decomposition Amarillo Real Estate

When reviewing the Amarillo Real Estate time series decompositions, I believe the best model to choose is the X-11 decomposition model. This model helps consider the upward trend of the value of homes while adjusting the timeseries data based off seasonal fluctuations. This provides a more stable timeseries component as well. It could be argued that X-13 could be better suited but a more robust model like X-13 is not required in my opinion.


X-13ARIMA-SEATS using X-11 adjustment decomposition
Median_sale_price = trend * seasonal * irregular
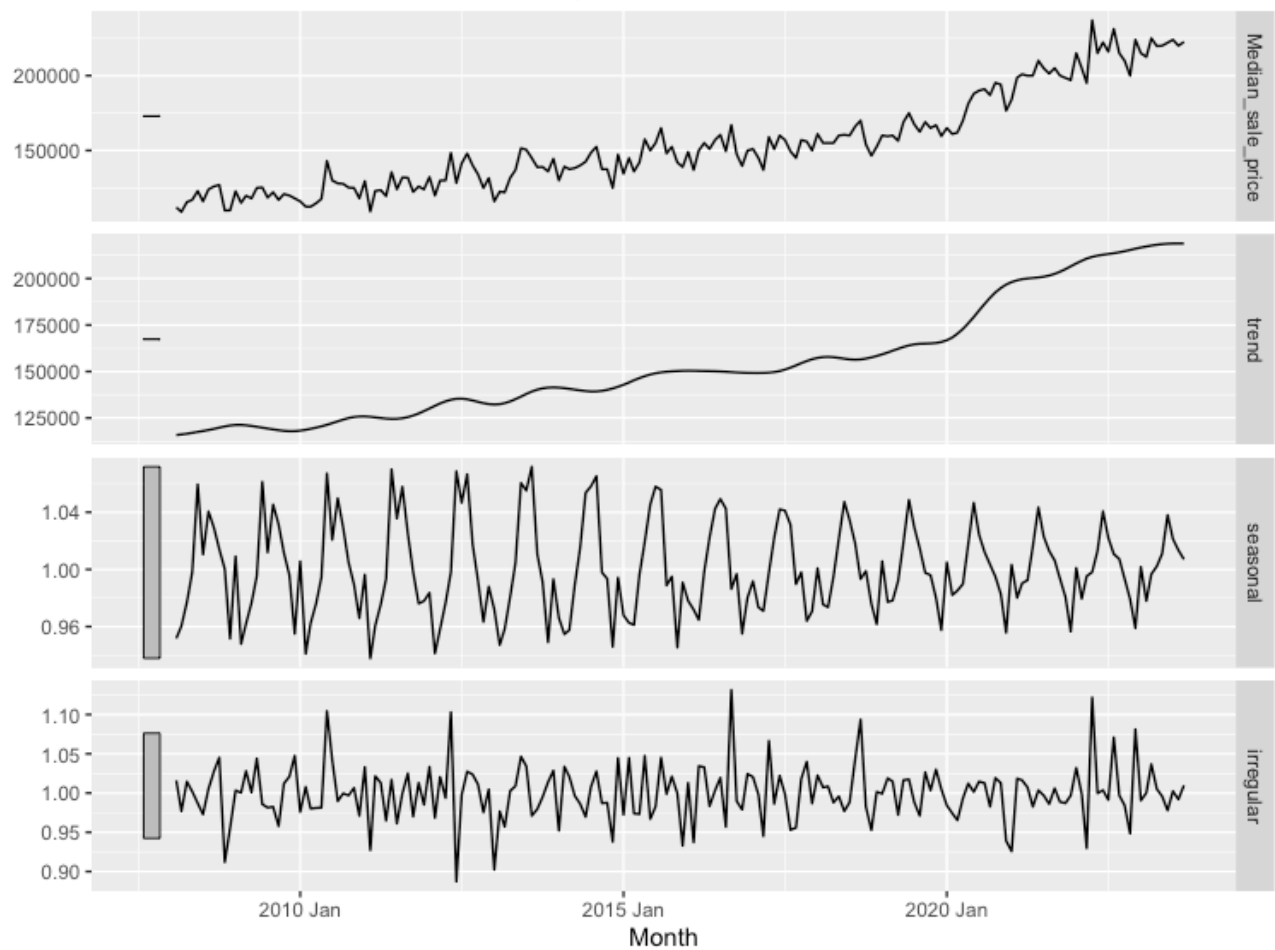
STL Decomposition Sacramento Grapes

This dataset was by far the most irregular of all the data I reviewed for this project. The data shows high seasonality with a constant upward trend of the value of grapes overall. The STL decomposition time series is robust to outliers and provides a smoother and more stable components when reviewing this specific dataset.

Seasonal Decomposition of Sacramento Grape Value
Value = trend + season_year + remainder
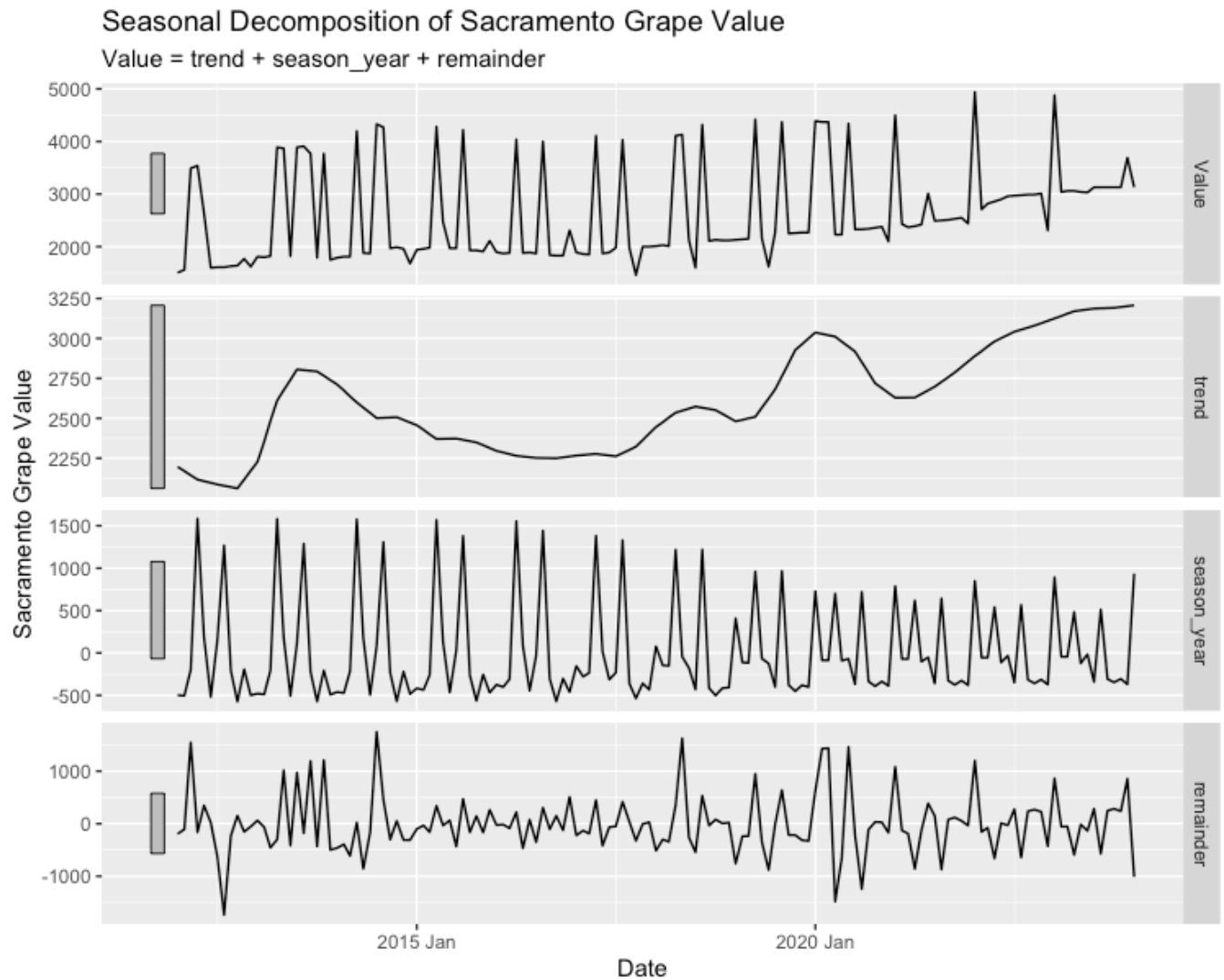
X-13 Decomposition Sacramento Unemployment Rate

     As I previously stated for the Amarillo Unemployment rate I believe the Sacramento Unemployment Rate dataset is also well-suited to be leveraged with the X-13 time series decomposition. This decomposition model was created to be well-suited for large scale economic data and provides more stable components throughout the time-series.

X-13ARIMA-SEATS decomposition

`unemployment rate` = f(trend, seasonal, irregular)
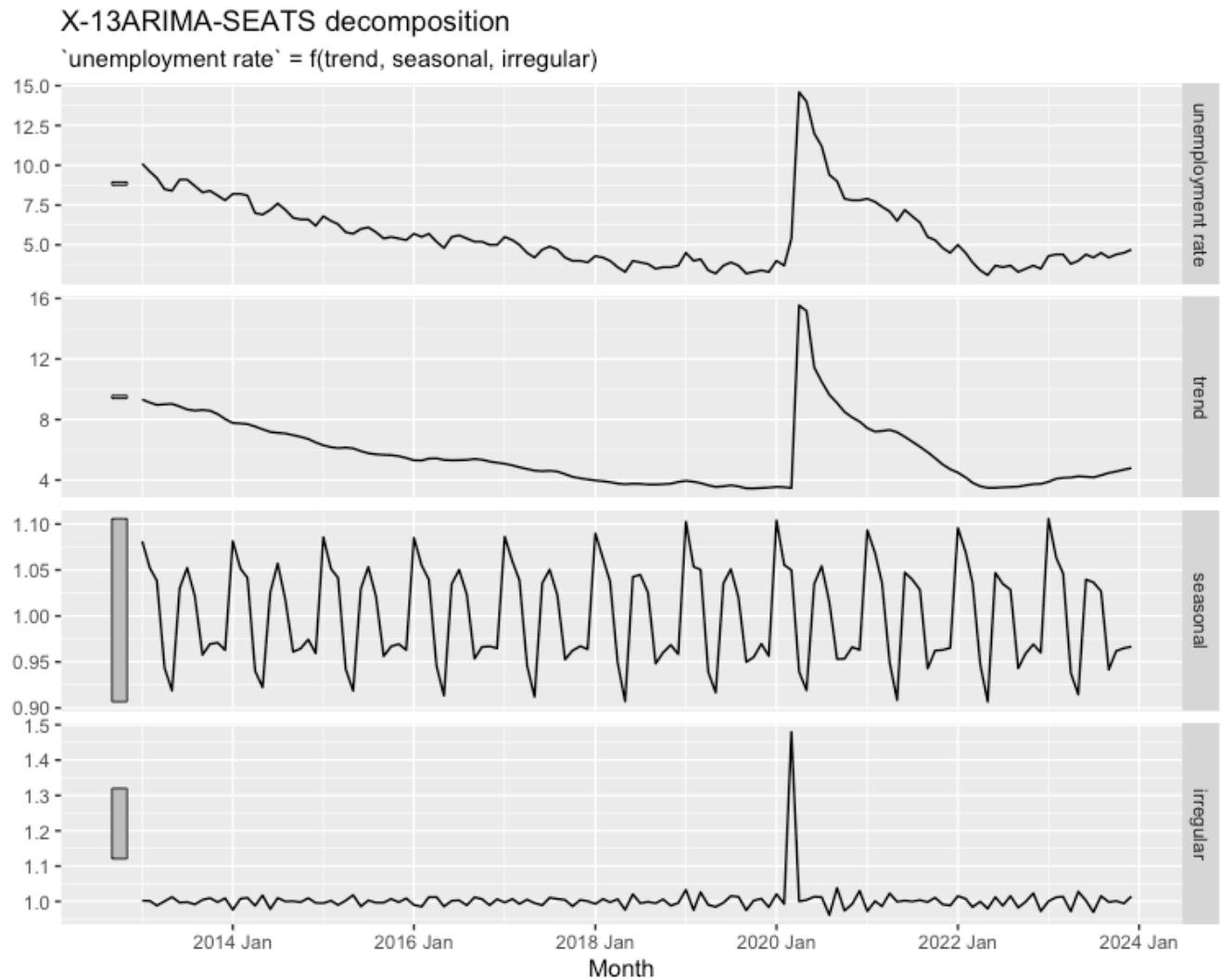
X-11 Decomposition Sacramento Housing Market

   The preferred decomposition time series that I believe should be used for the Sacramento Housing Market is the X-11 decomposition. As previously stated, large scale economic data is better suited for these time series and allows to better track seasonal patterns, trends and the median sales price (which is steadily increasing year over year). This decomposition time series allows for a more stable chart and clearer representation of the time series.


X-13ARIMA-SEATS using X-11 adjustment decomposition
`Median Sales Price` = trend + seasonal + irregular
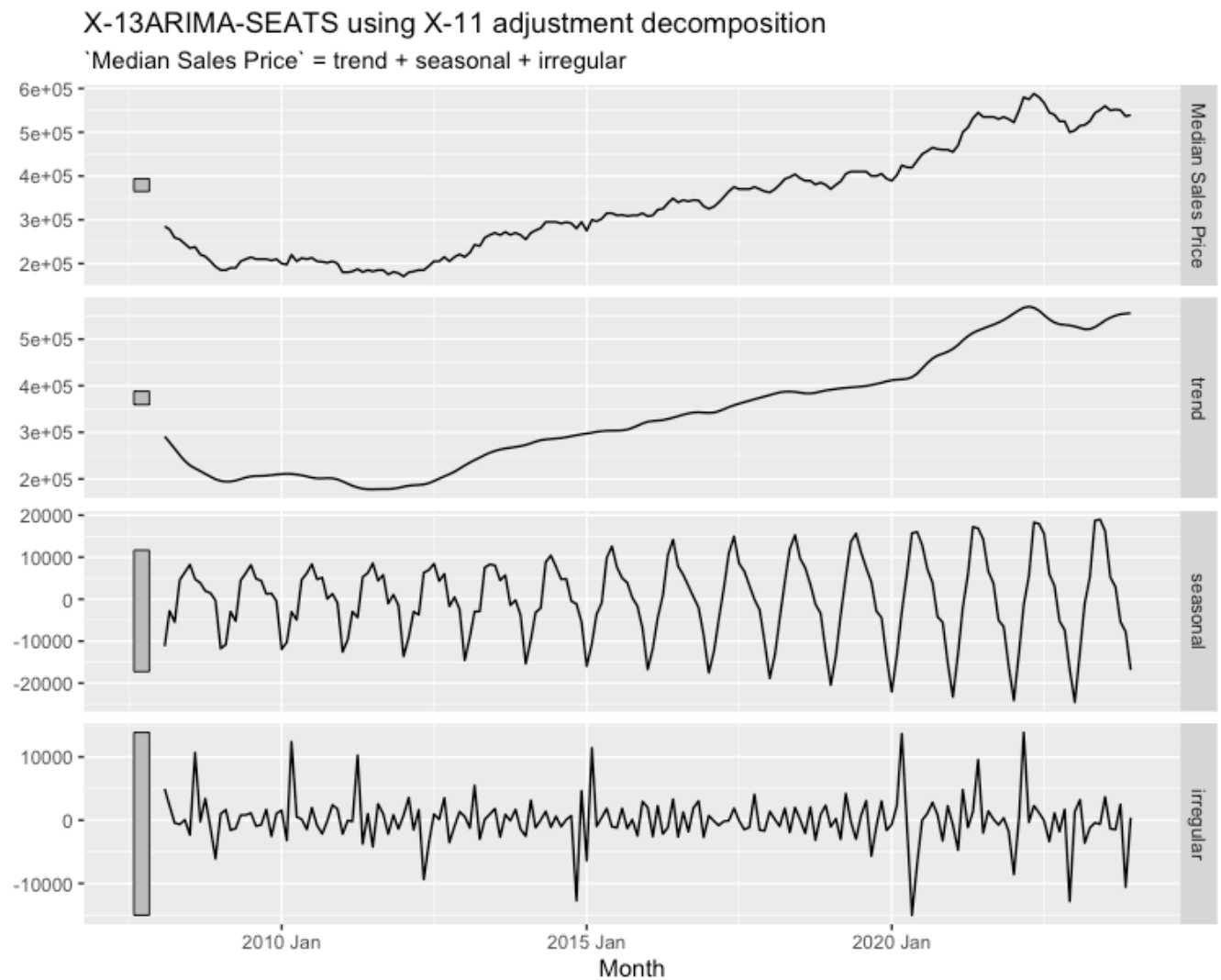
Section 5

When choosing the appropriate forecasting model, I chose the model where the residual diagnostics chart appeared primarily as "white noise" and there was no matching pattern or seasonality to the forecasting method I chose.
RW ~Drift Amarillo Cattle Prices

After reviewing the Amarillo Cattle prices dataset, I have chosen the RW Drift forecast for this specific dataset. There is no clear long-term seasonality of this data. The Cattle Prices in Amarillo are in an upward trend and have continued to grow since 2020 post the Covid crisis.

Amarillo Cattle Prices
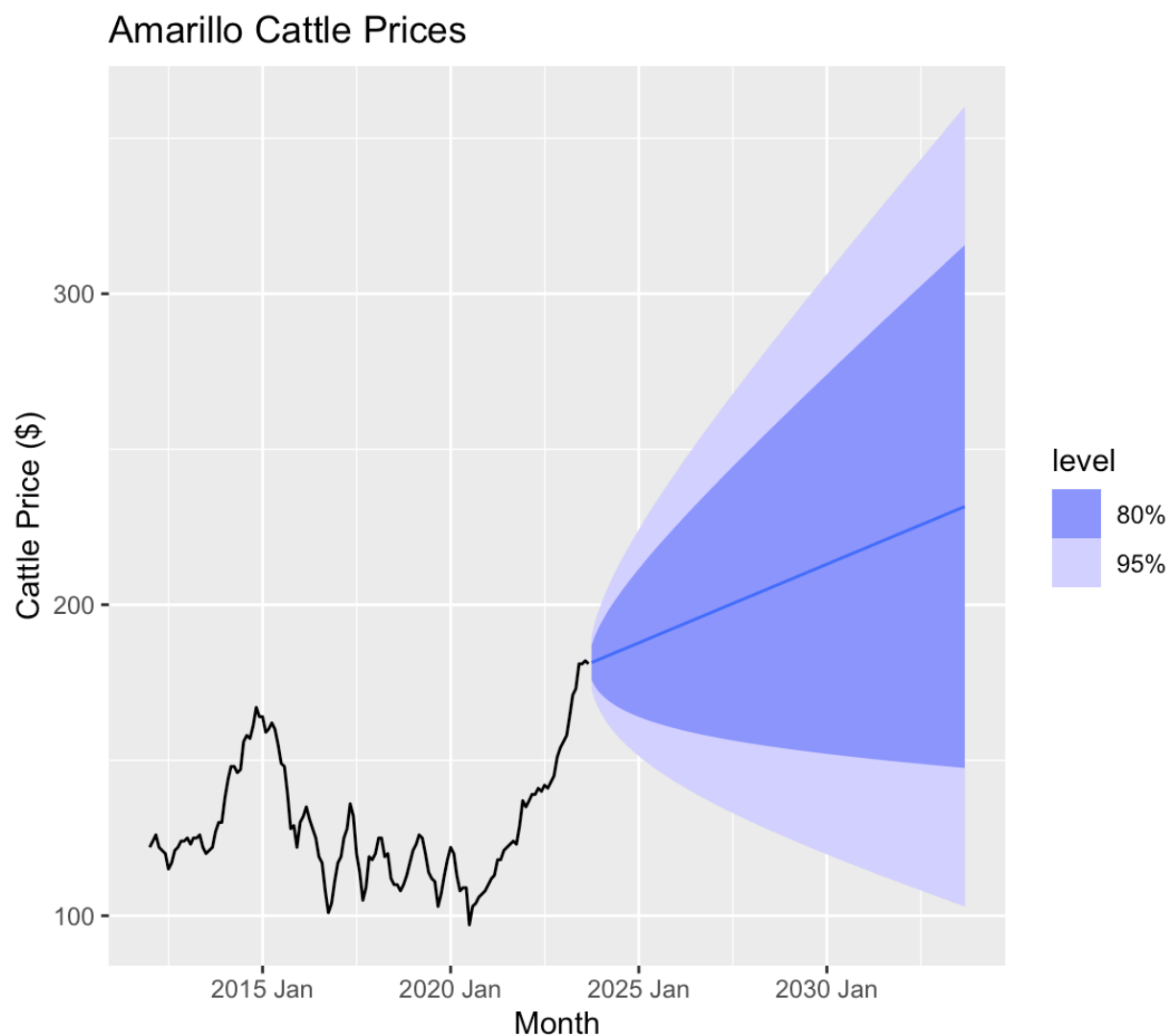
Residual Diagnostics for Amarillo Cattle



Residuals for Amarillo Cattle Prices
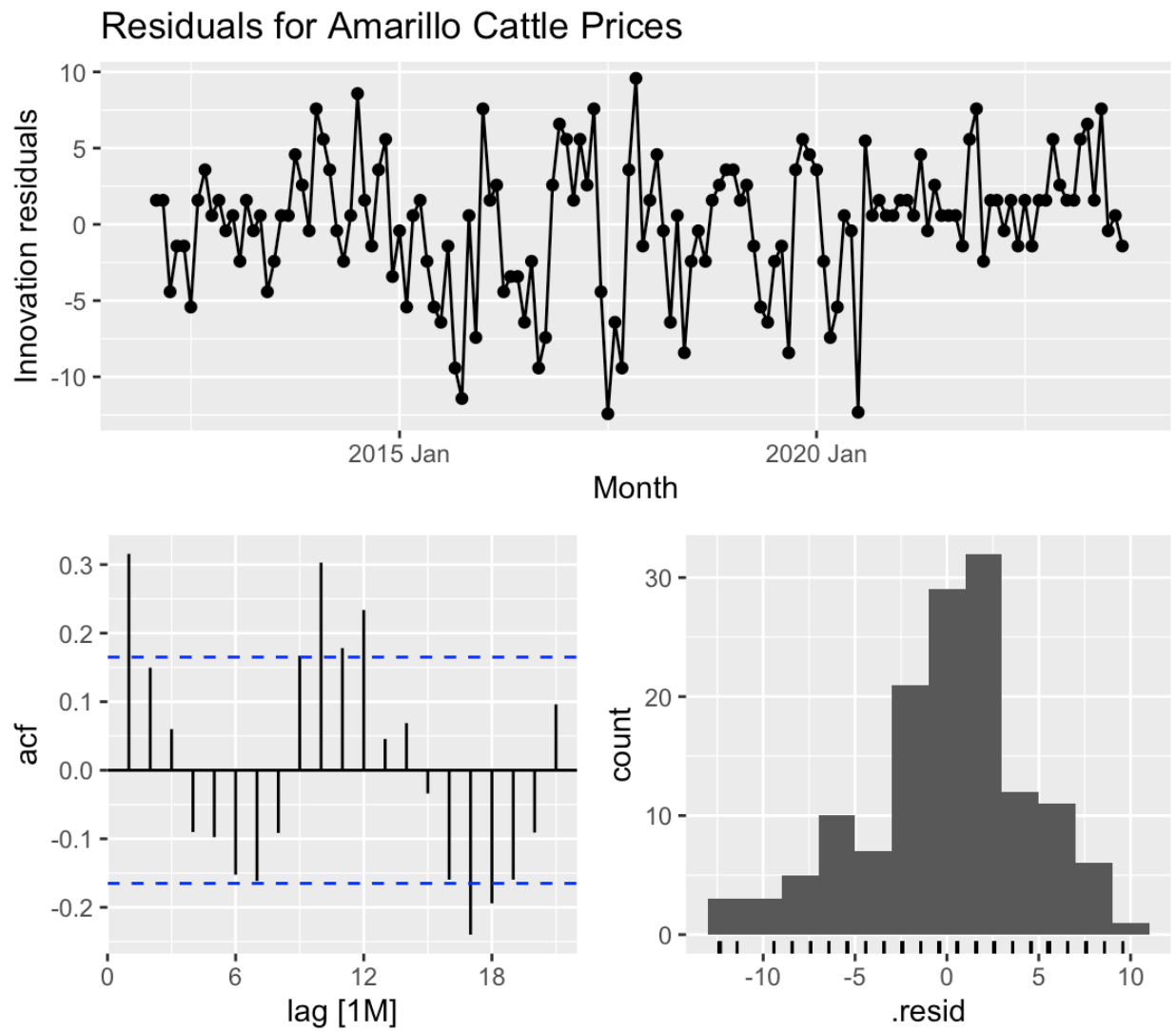
SNAIVE Amarillo Unemployment

       Based off the data set, the best option for Amarillo Unemployment forecasting is the Seasonal Naive model. The Seasonal Naive model allows for the forecast to match the current Unemployment Rate within the confidence intervals highlighted within the 80 and 95 percentiles.
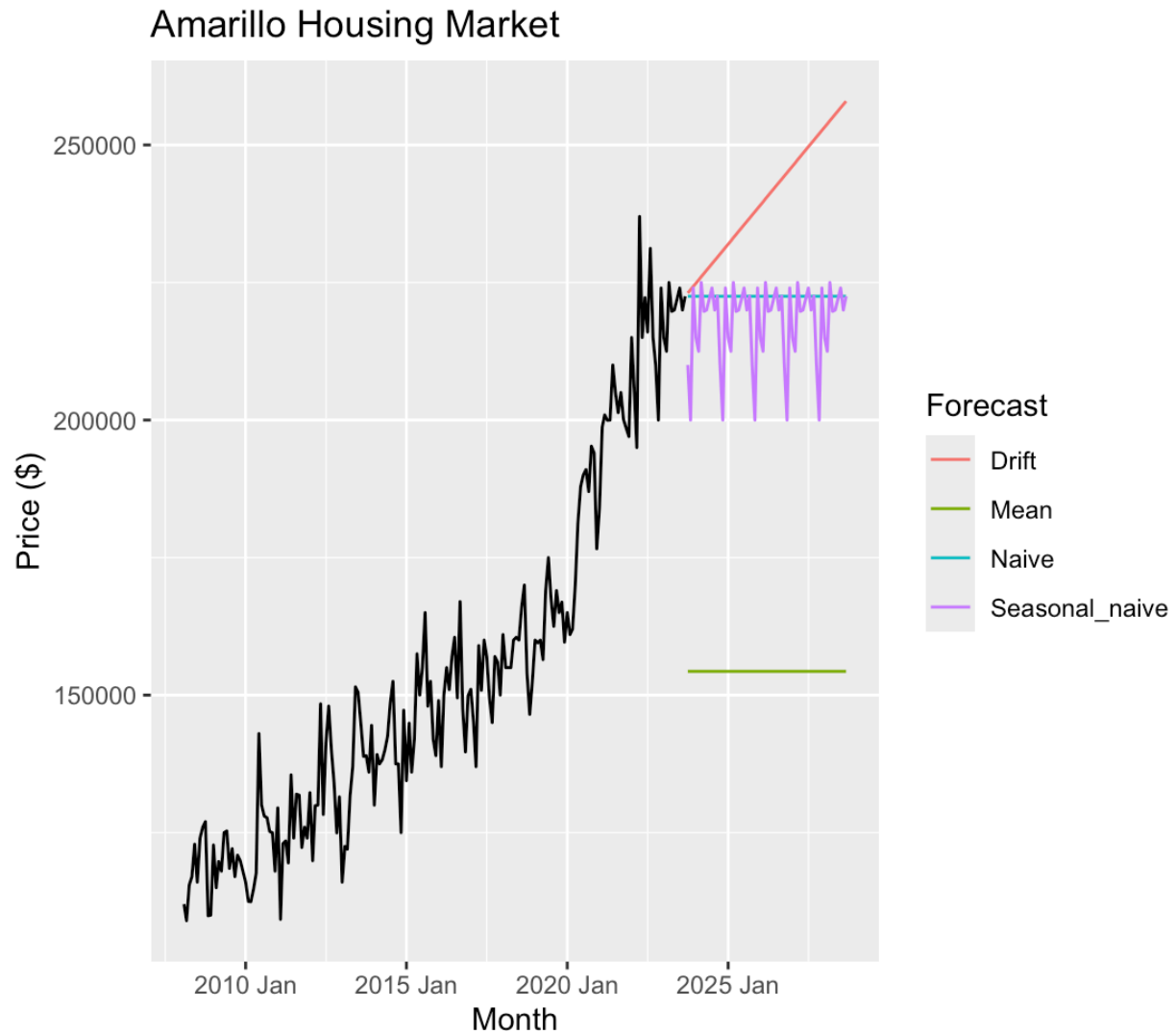
Seasonal Naive Forecast for Amarillo Unemployment Rate

Residual Diagnostics Amarillo Unemployment
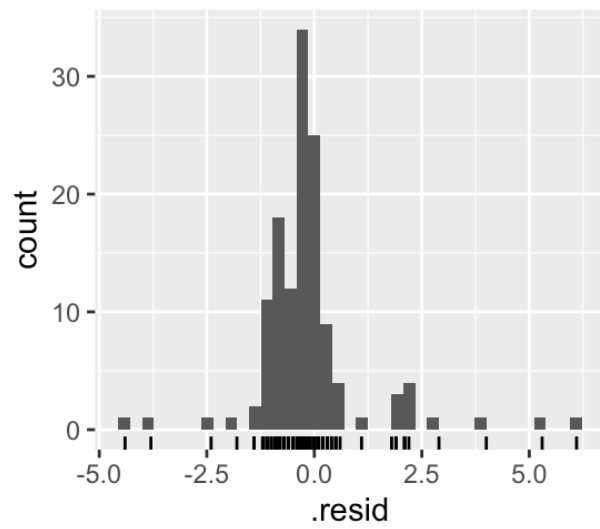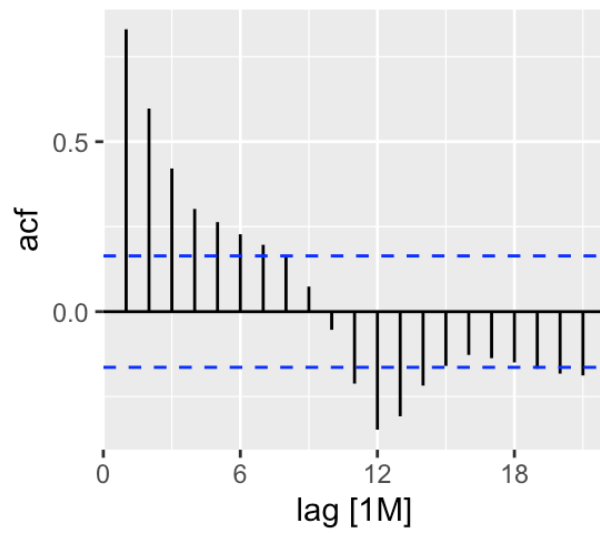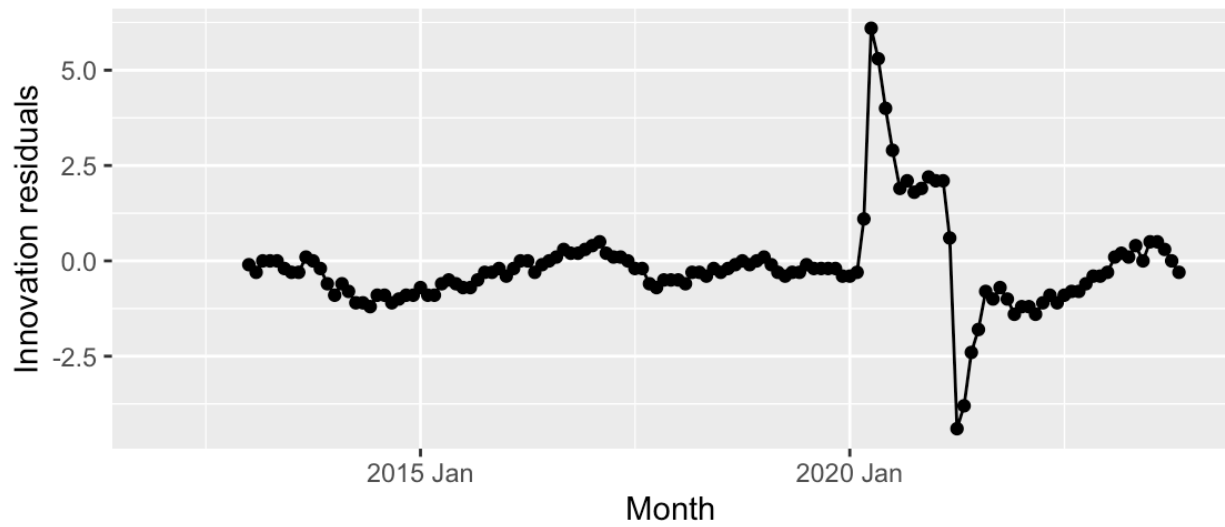
Forecast Models for Amarillo Housing Market

Amarillo Housing Market

I ran this forecasting model to see the different models in relation to the most recent data I have available. When reviewing the models it appears to me that the best used models would be Drift due to the increasing nature of the Home Prices in Amarillo. Additionally, it could be argued that a Seasonal Naive forecast could be used to specifically look at the short-term seasonality trends of the housing market prices but for a long term forecast the Drift method seems to be the best choice.
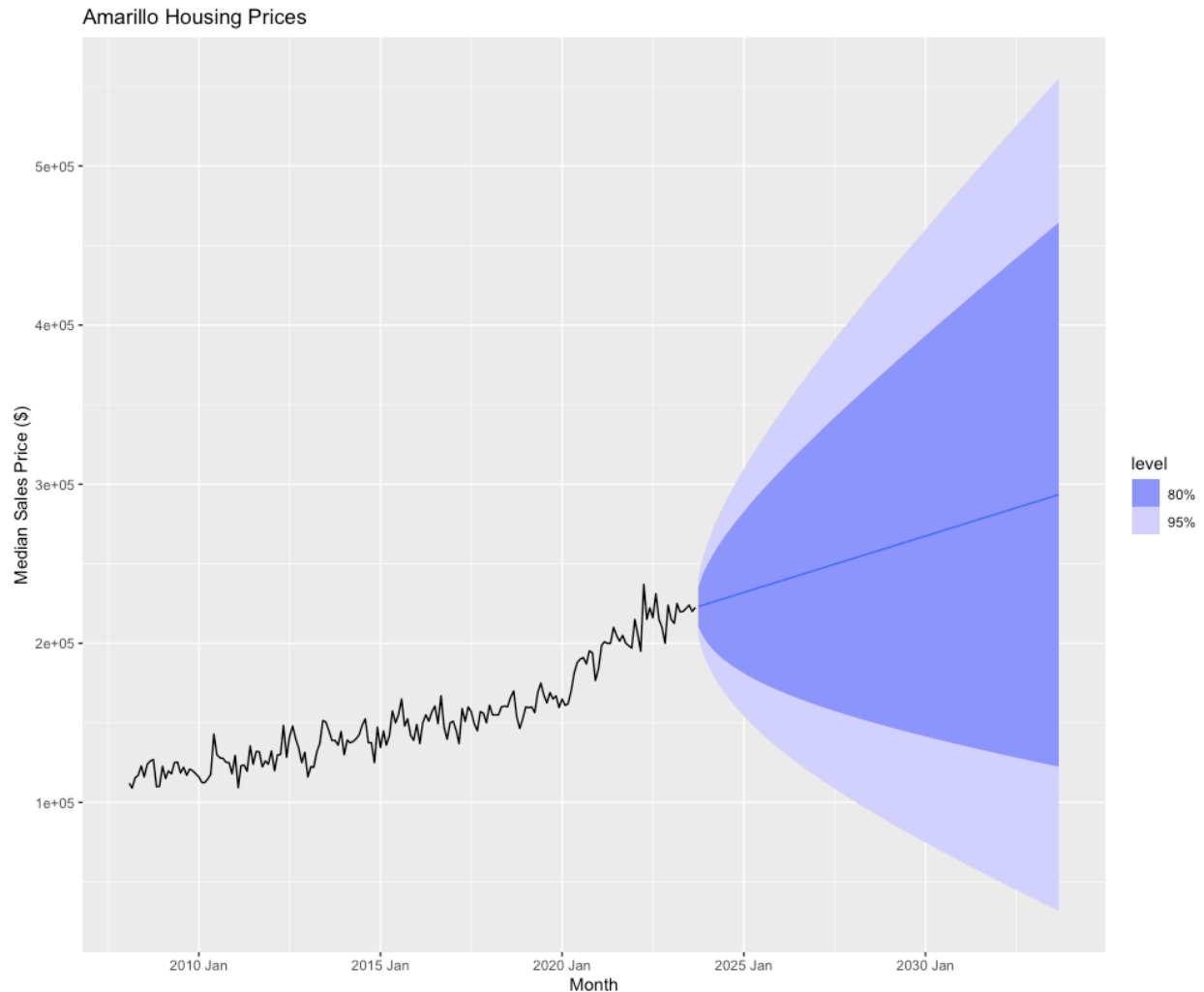
Seasonal Naïve Amarillo Housing Market – Not Preferred Forecasting Method

   This is not the best solution for a long-term forecast of Amarillo Housing prices but could be used to identify the short-term seasonal trends of the housing prices in Amarillo.
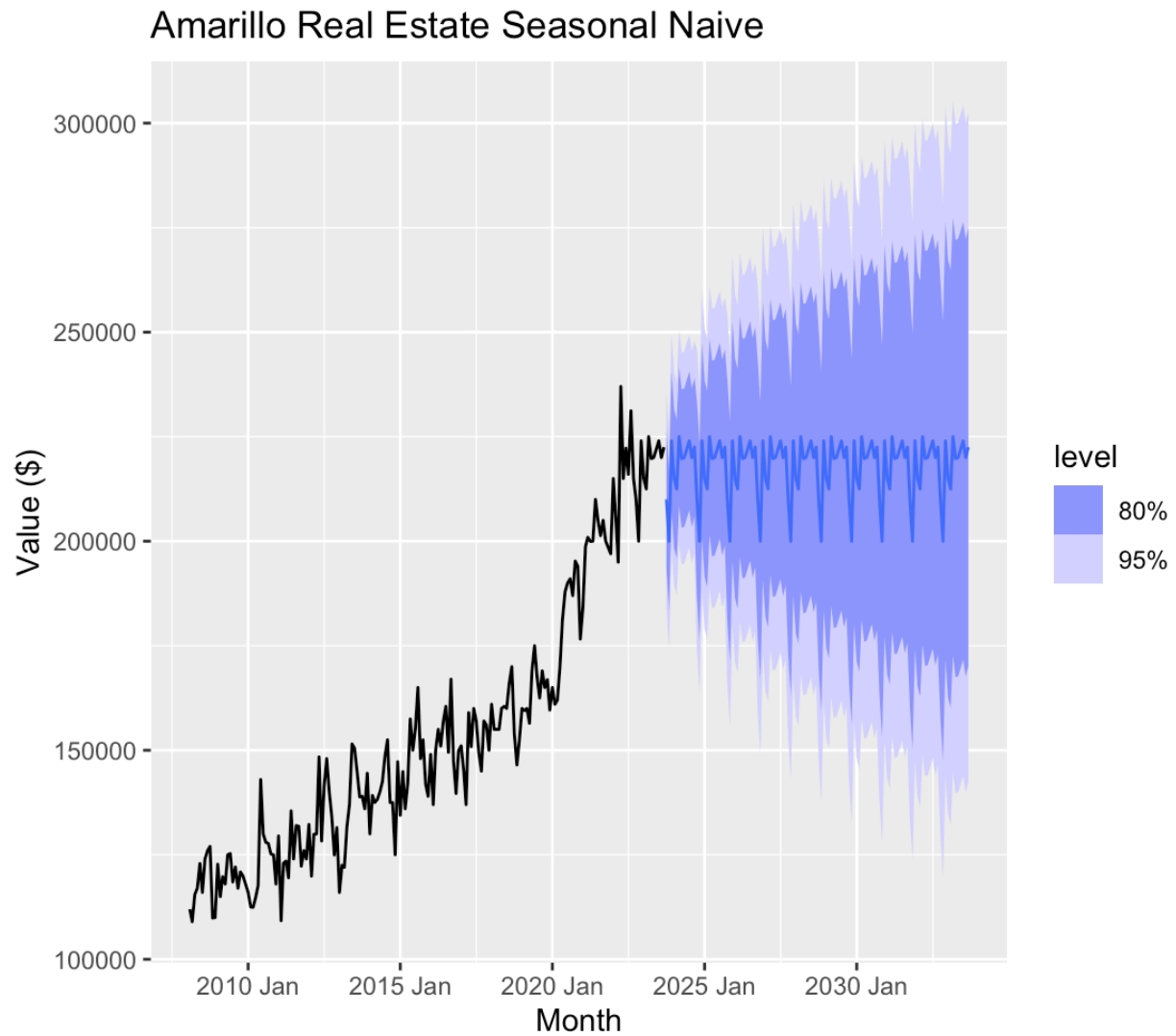
Amarillo Housing Market Drift Forecast – Preferred Forecasting Method

The Drift forecast is the best forecasting method to use for the Amarillo Housing Market dataset. This is due to the lack of long-term seasonality and the constant upward growth of the median price of homes in Amarillo.

It could be argued that a blend of SNAIVE (for short term seasonality) and RW Drift (for long term trend) can be used to forecast the Amarillo Housing Market for a more accurate and robust forecasting outcome.

Amarillo Housing Market Residual Diagnostics

Residuals for Amarillo Housing Market
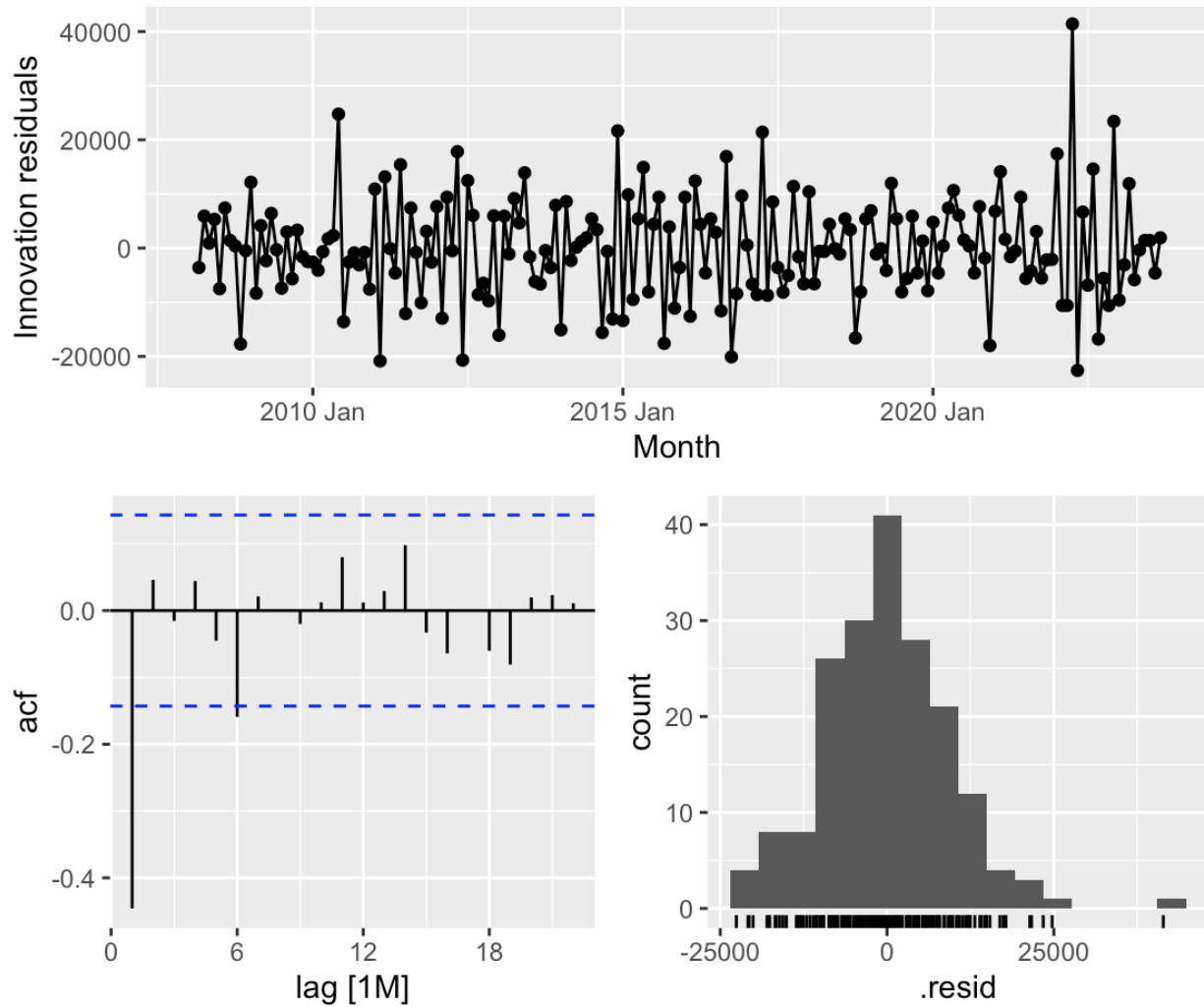
SNAIVE Forecast for Sacramento Grape Value

      Due to the high seasonality of the Sacramento Grape Market the SNAIVE forecasting model is best suited to forecast the Grape Values over the next 10 years. While drift could be argued as well as another model it shouldn't be used due to the seasonality of the grape market. I previously had Drift as my selection but after additional research and consideration I have changed my stance.


SNaive Forecast for Sacramento Grape Value
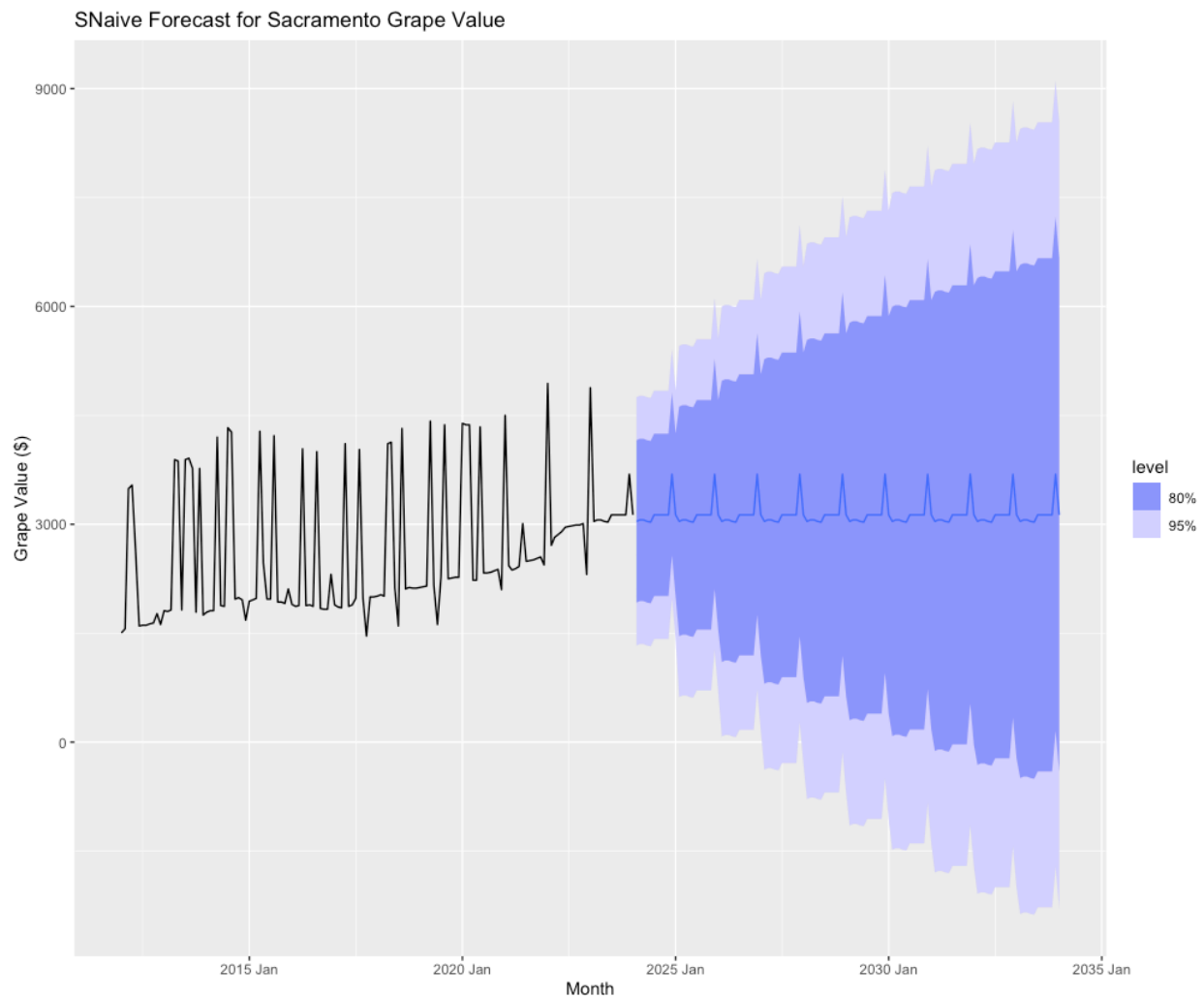
Residual Diagnostics for Sacramento Grape Values

# Residuals for Sacramento Grape Value
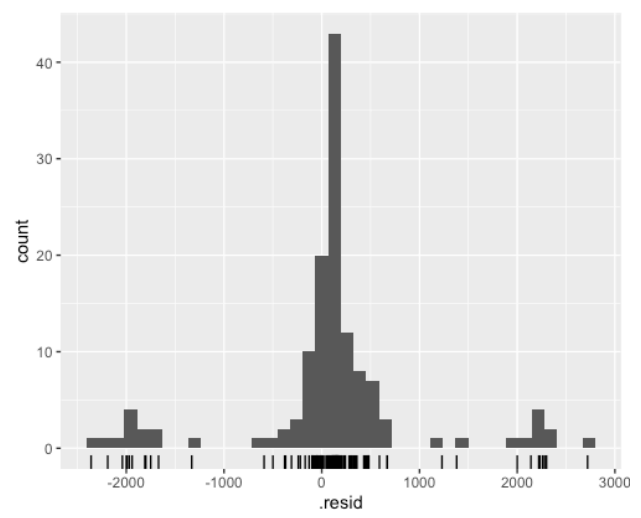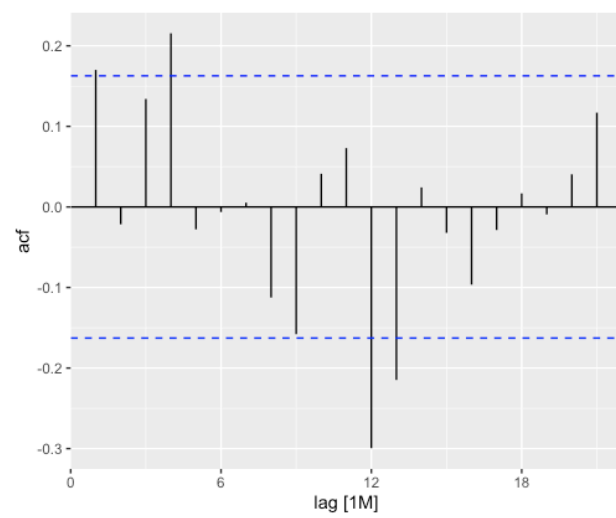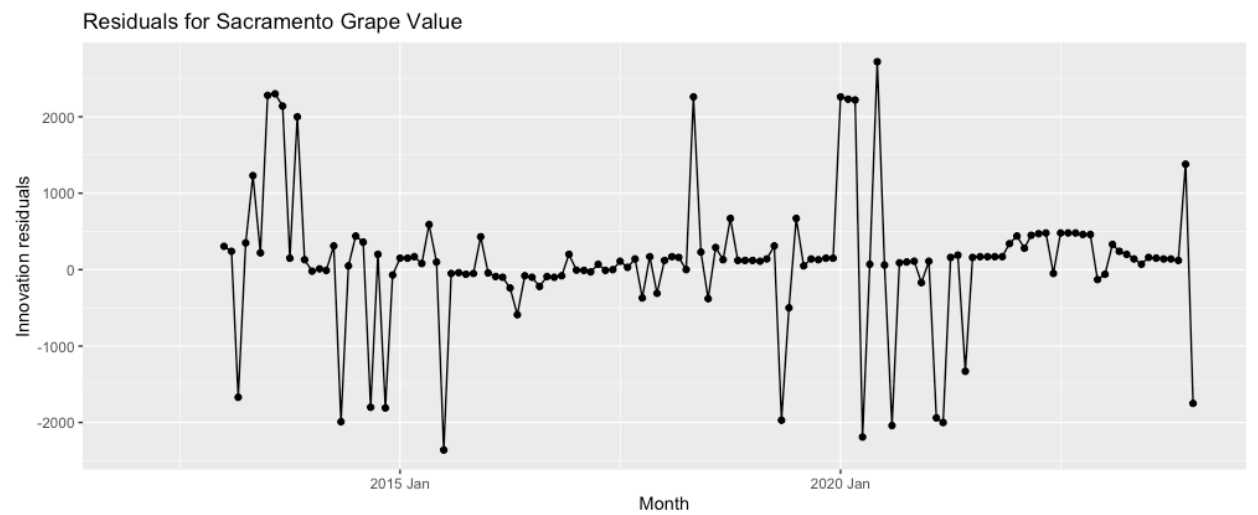
SNAIVE Forecast for Sacramento Unemployment Rate

      In similar fashion to the Amarillo Unemployment Rate, we should use the SNAIVE Forecasting model for to predict the Sacramento Unemployment Rate. This will allow us to consider the seasonality patterns as well as the overall trend of Unemployment rates while remaining sensitive to large fluctuations to the data (i.e. the outliers such as the Covid Pandemic).
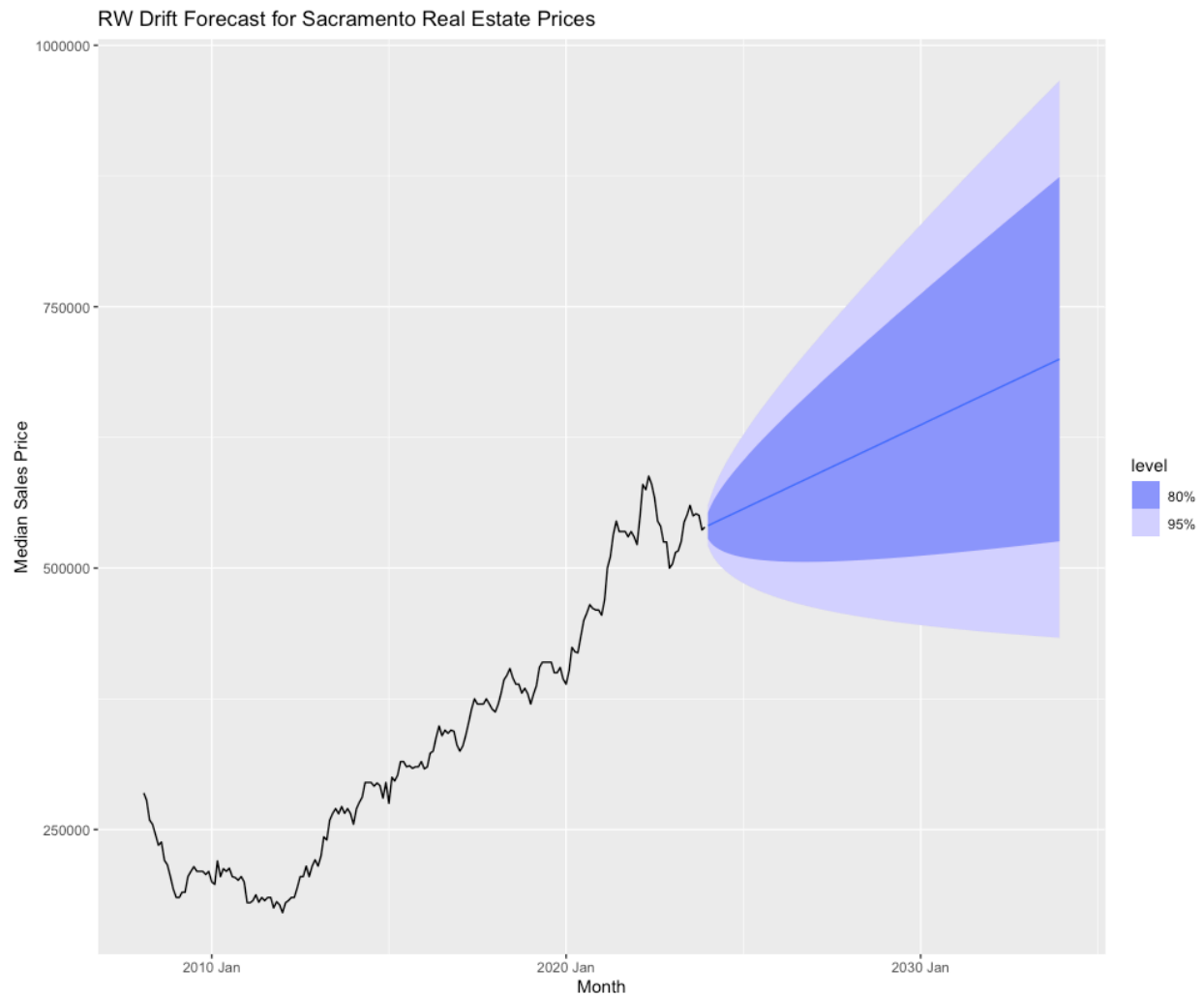

SNaive Forecast for Sacramento Unemployment Rate

Sacramento Unemployment Rate Residual Diagnostics
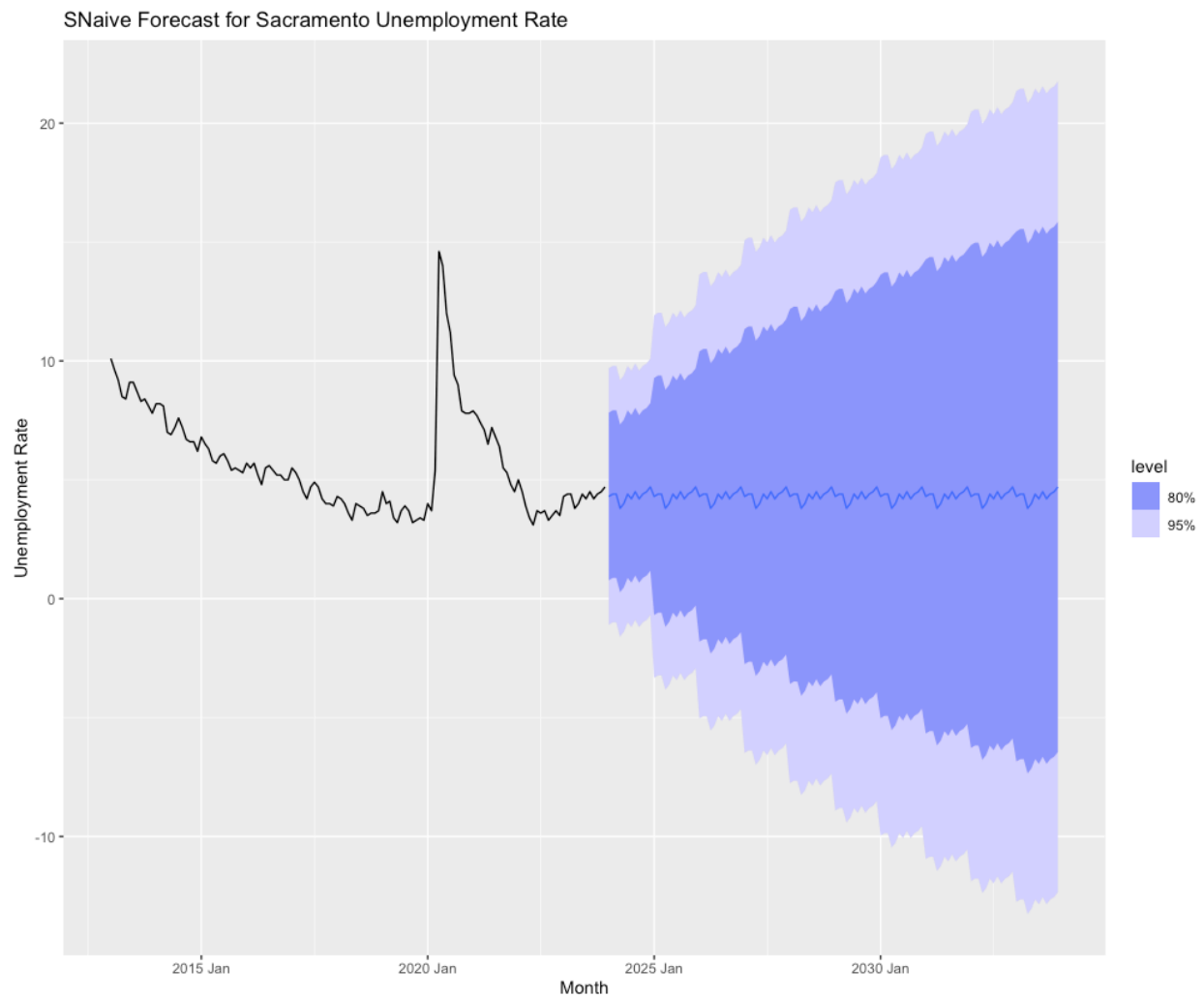
Residuals for Sacramento Unemployment Rate
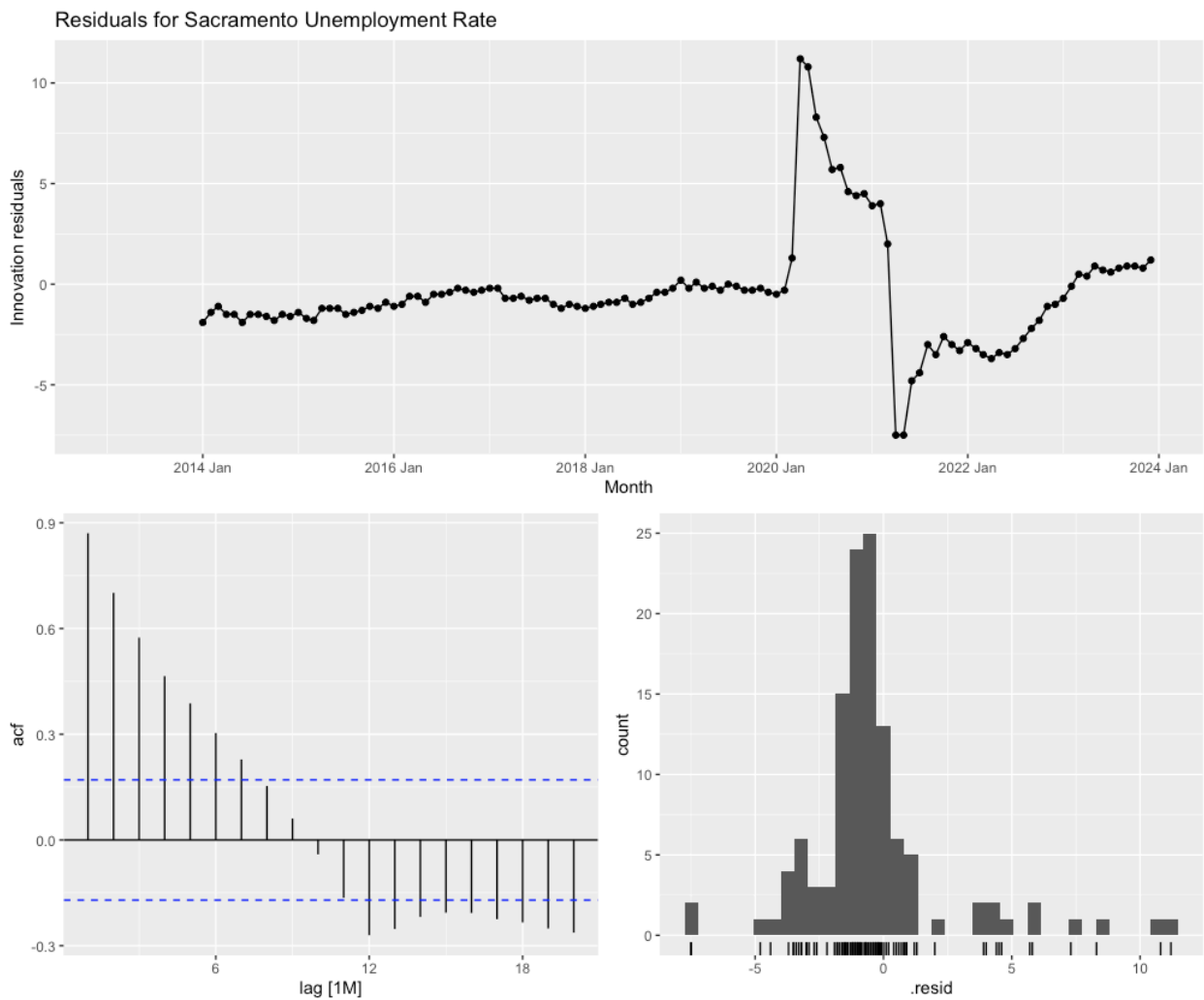
Sacramento Housing Market RW Drift

      When plotting any Real Estate Market data an approach of SNAIVE and Drift can be used to accurately forecast both seasonality and the trend of the forecast. In this specific dataset I have chosen to go with the RW Drift forecasting method. This forecasting method allows us to have a more definitive confidence interval but does not include the seasonality of the housing market in Sacramento. The Sacramento housing market sees less seasonal fluctuations in price when compared to the Amarillo region hence, the usage of Drift is the best forecasting model.

RW Drift Forecast for Sacramento Real Estate Prices

Sacramento Housing Market Residual Diagnostics

Residuals for Sacramento Housing Market