**Abstract:**

**Introduction:**

While catastrophic events such as wars, terror attacks, pandemics, and natural disasters receive the lion’s share of media attention, the simple fact remains that the vast majority of Americans will die from one of two causes: Heart Disease and Cancer. While the speculation on the factors involving the rates of death causes by these two scourges is beyond the scope of this paper, constructing models and making predictions nonetheless has much utility in the context of public health.

**Methods:**

Data were obtained from the CDC’s WONDER database on the total numbers of deaths in the United States, in the years 1999 to 2018. The data were filtered to include deaths from circulatory system issues, representing heart disease, and deaths from neoplasms, representing cancer. The data were collected by year and by month for each. The total population of the United States was available for the yearly but not the monthly data, so it was added to the monthly data, but reflects the population of the US over that entire year, not the month.

In order to remove the effects of the US’s rapid population growth, the *rate* of deaths was analyzed, as opposed to the total number of deaths. Rates were expressed as percentages of the total population dying from each cause per unit time. Because the population data was only collected by year, the percentages could be considered to be only approximations of the total death rates, since intra-year births and deaths were not factored into the counts. However, because the size of the United States’ population is very large compared to the numbers of births and deaths, this approximation is considered to be reasonably accurate for our purposes.

In order to control for general upwards and downwards trends in the data over time, first differences and seasonal differences were taken to see if they stabilized the data. Because both time series appear to have equal variance over time, log and power transformations were not considered. Then auto-correlation functions (ACFs) and partial autocorrelation functions (PACFs) were taken of each time series to inform which starting model was used, and a periodogram was used to confirm periodicity. Starting models included regular differencing and seasonal differencing if shown to be warranted from the initial analysis. The model started as a moving average of degree 1, and degrees were added until they we no longer significant. Then seasonal moving average components were added, one degree at a time, until there were no further significant parameters. Finally, the same was done with autoregressive components. The final model in each case was evaluated with residual autocorrelation, QQ plots. A Shapiro-Wilk test was not considered due to the large sample size of the data.

Due to the complexity of the data, it was possible to produce very complex models that had significant Moving Average and Autoregressive components at high lag times. However, in order to reduce the possibility of overfitting, favor was given to simpler models. This resulted in a compromise whereby models would be complex enough to fully model the data, but models with MA and AR orders greater than 5 were heavily discouraged.

Finally, using the fitted models, predictions were made for the percentage of deaths occurring from both causes for an additional five years. Note that data were only available through year end 2018, and so the data for 2019 and 2020 are predictions. The predictions do not take into account the possibility of medical advancements that would reduce the death rate from both causes, and so if such advancements were made in the prediction window, the rates would not be accurate. The discussion of such advancements is beyond the scope of this paper.

**Results and Discussion:**

**Preliminary:**

**Cardiovascular Deaths:**

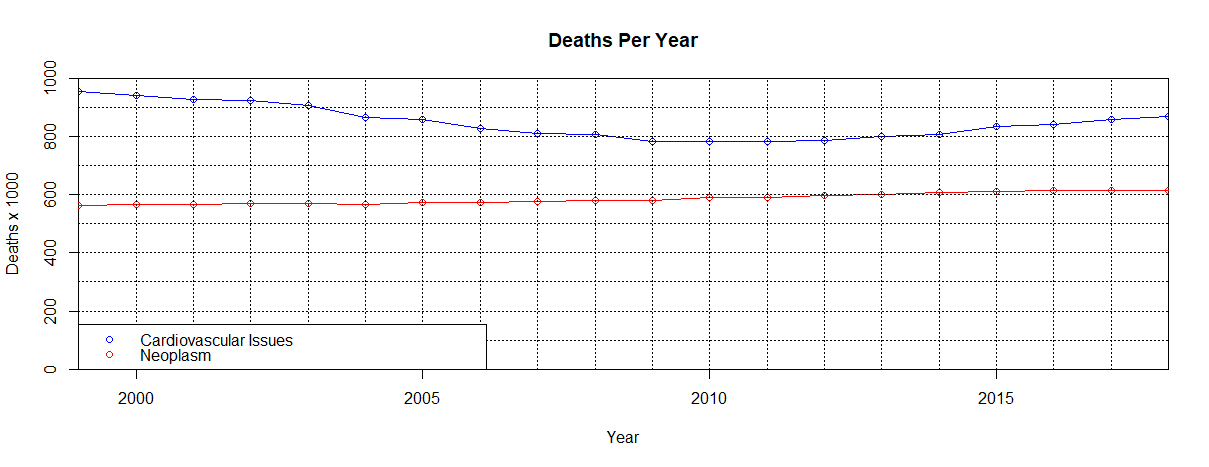
**Neoplasm Deaths:**

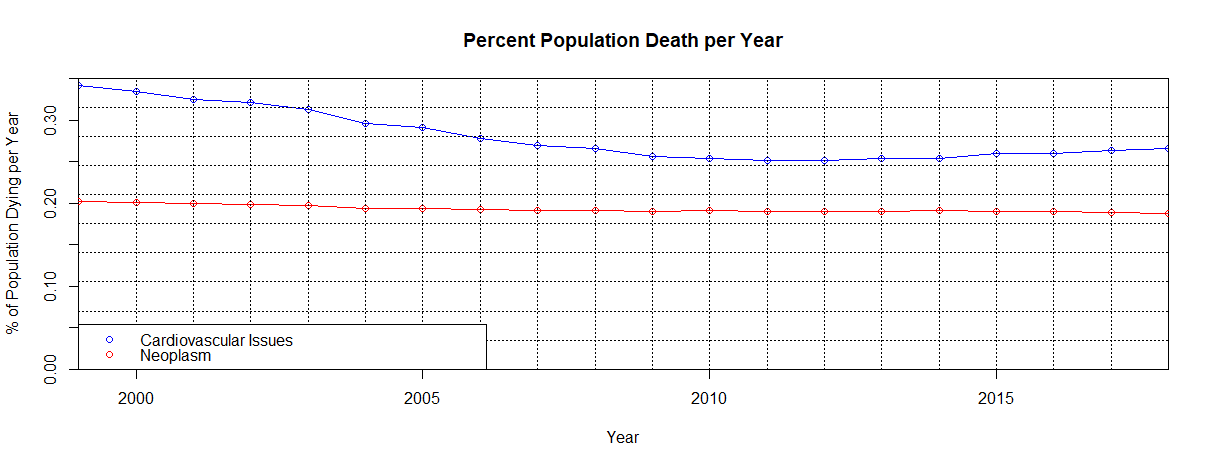
**Predictions:**

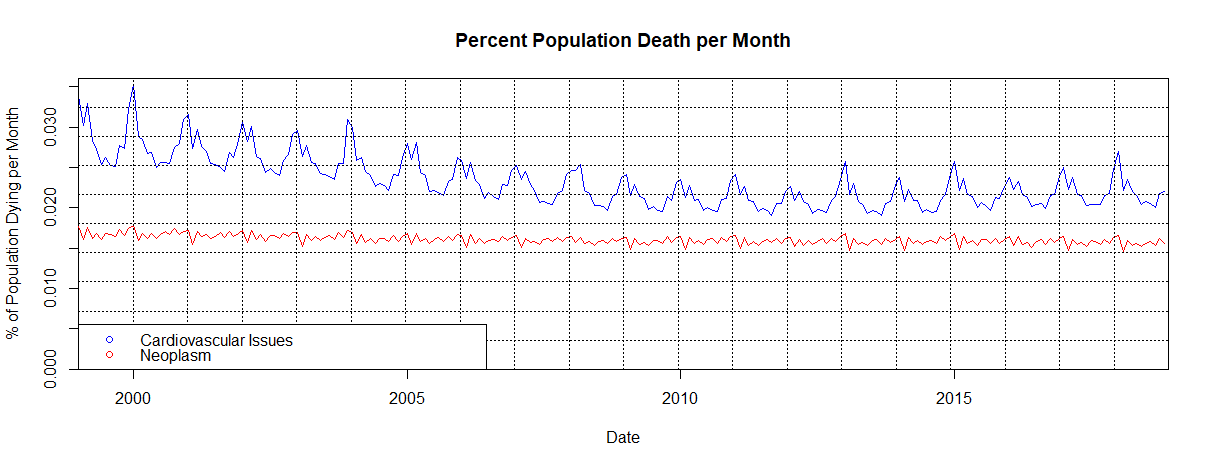
**Conclusion:**

**Appendix:**

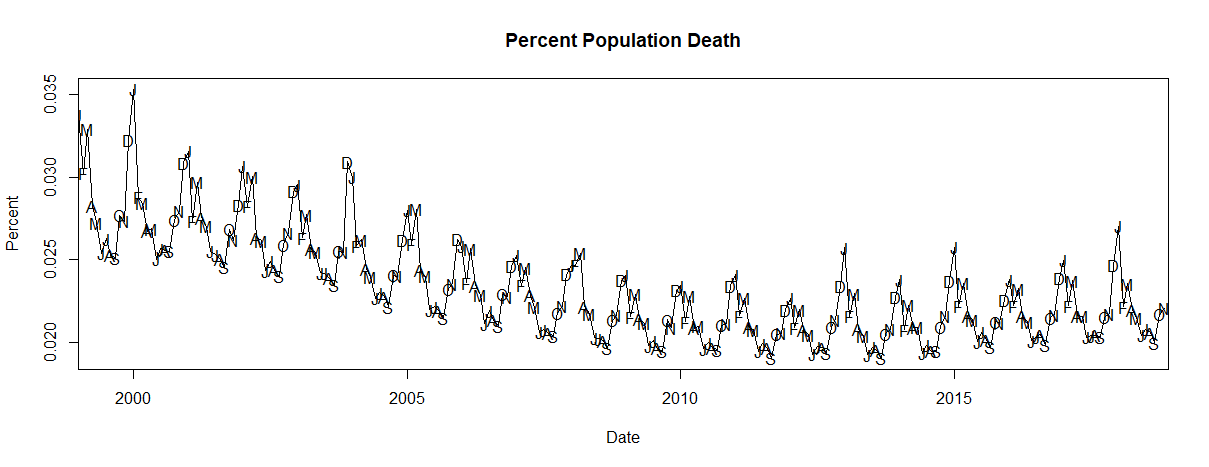
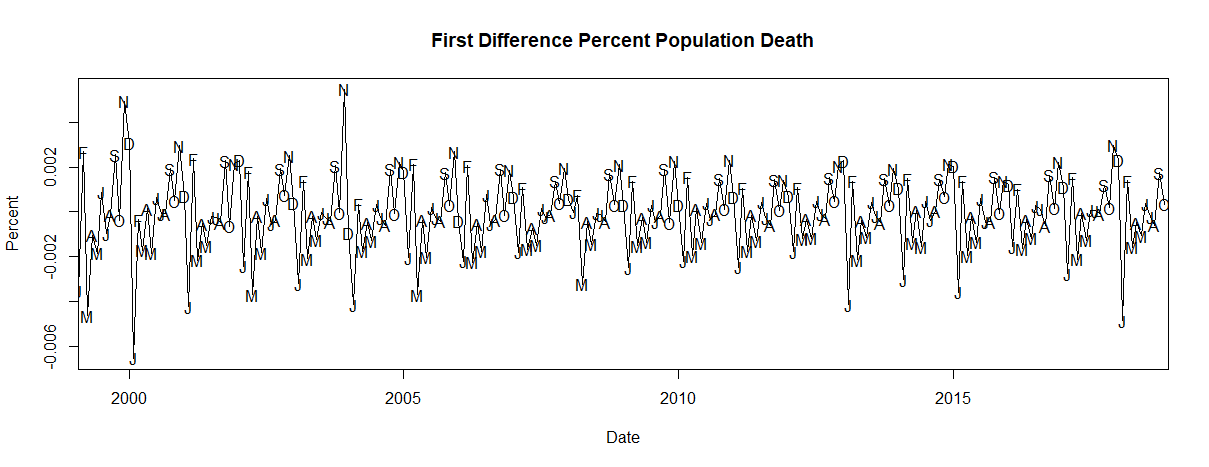
**Appendix 1: Preliminary Analysis:**

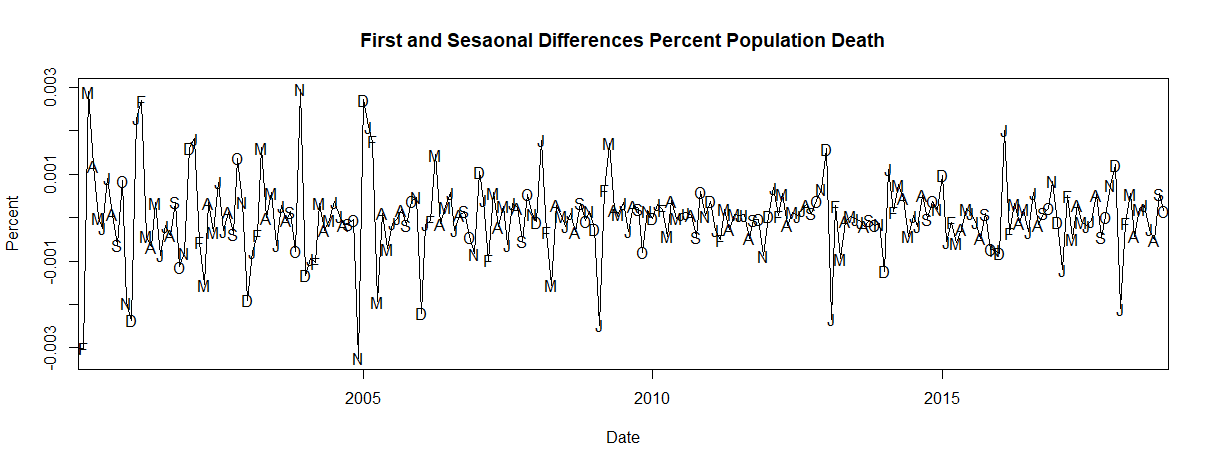
*Figure 1: Total Number of Deaths per Year in the United States from Cardiovascular and Neoplasm issues*

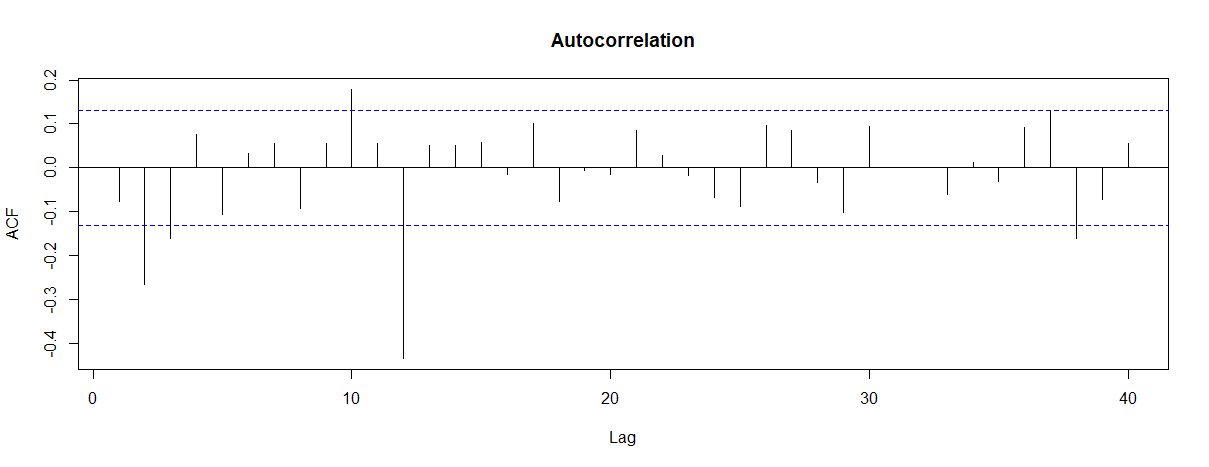
*Figure 2: Percentage of the Population Dying per Year from Cardiovascular and Neoplasm Issues*

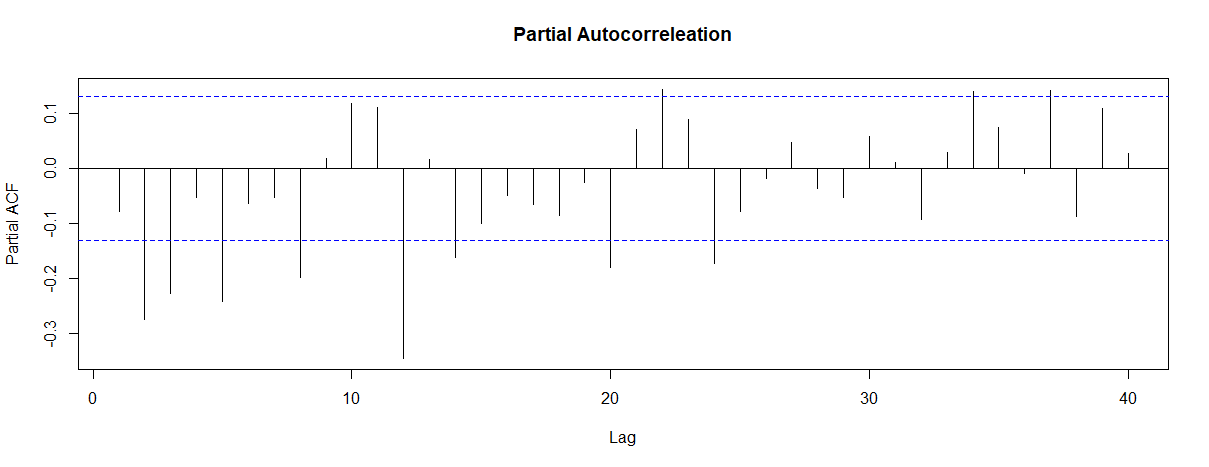
*Figure 3: Percentage of the Population Dying per Month from Cardiovascular and Neoplasm Issues*

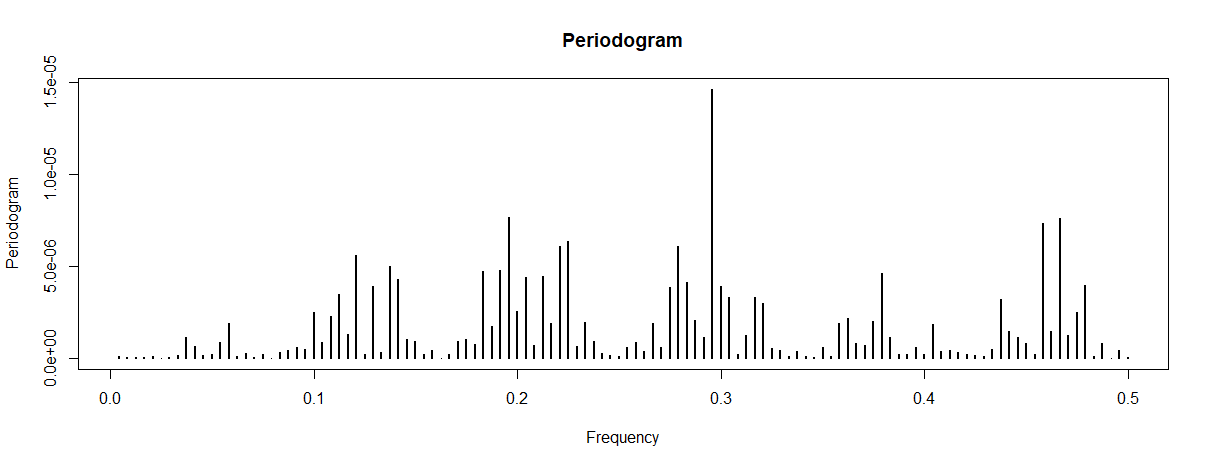
**Appendix 2: Preprocessing - Cardiovascular Deaths:**

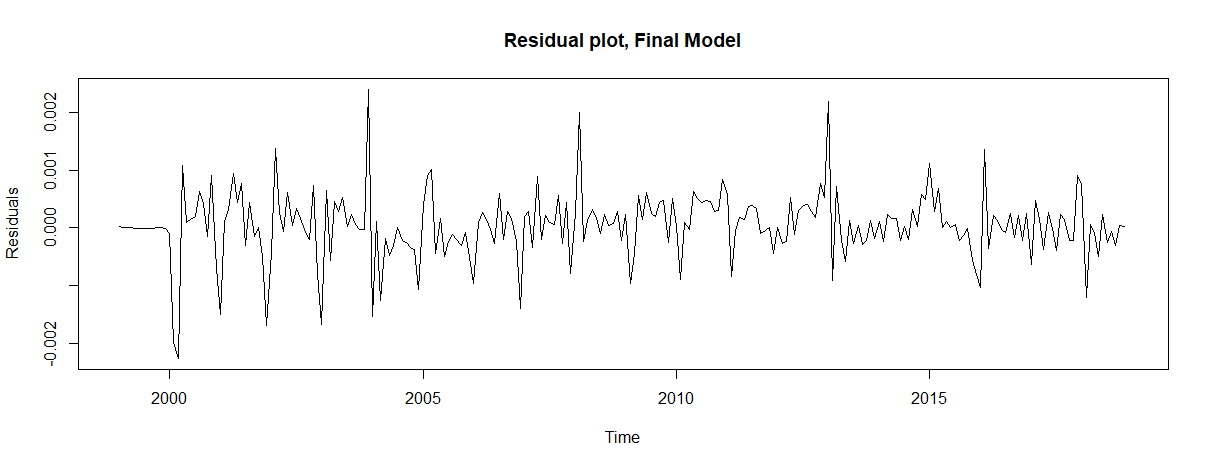
*****Figure 4: Percent of Monthly Population Deaths due to Cardiovascular Issues* *****Figure 5: First Difference of Monthly Population Deaths due to Cardiovascular Issues*

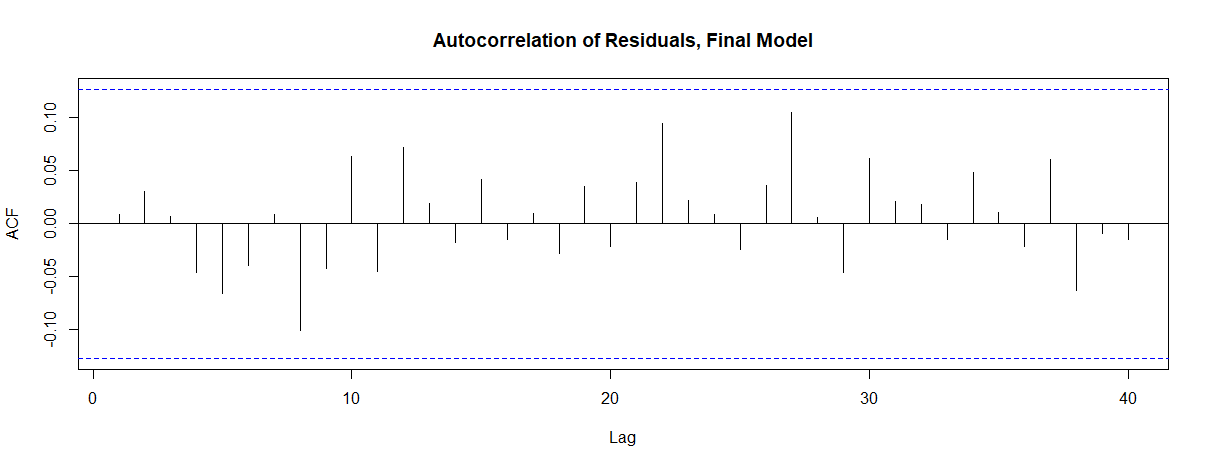
*****Figure 6: First Difference and First Seasonal Difference of Deaths due to Cardiovascular Issues*

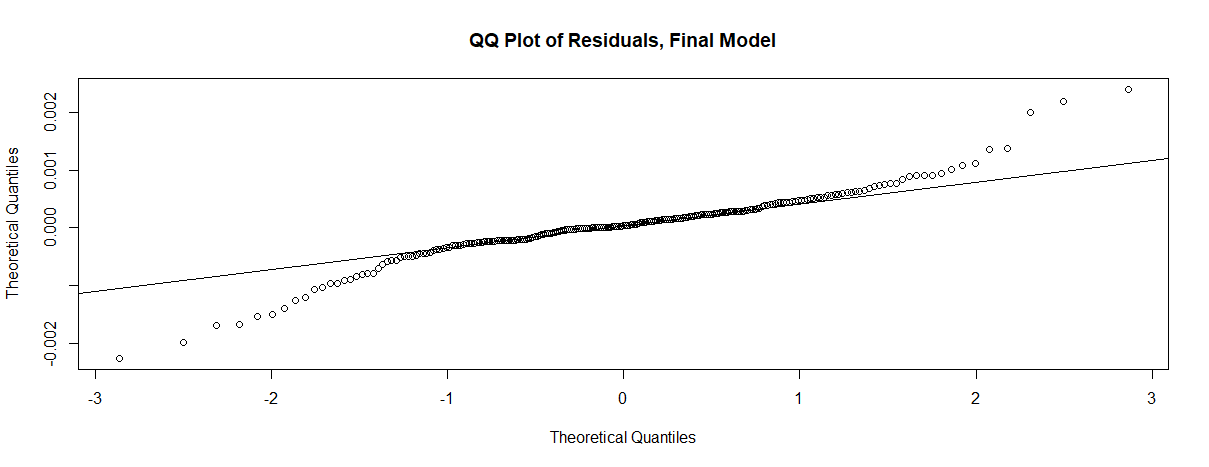
*****Figure 7: Autocorrelation Function of First and First Seasonal Differenced Deaths due to Cardiovascular Issues*

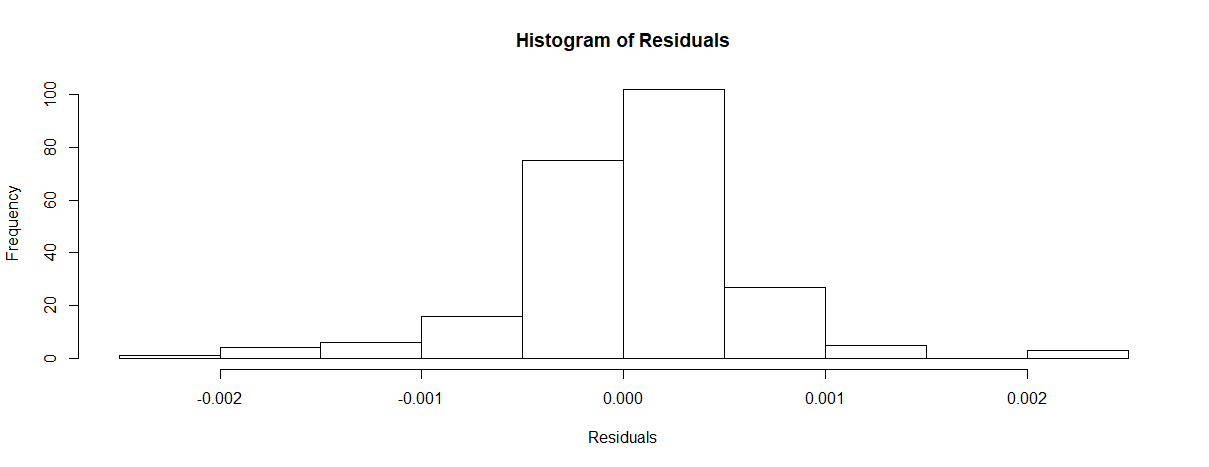
*****Figure 8: Partial Autocorrelation Function of First and First Seasonal Differenced Deaths due to Cardiovascular Issues*

*****Figure 9: Periodogram of First and First Seasonal Differenced Deaths due to Cardiovascular Issues*

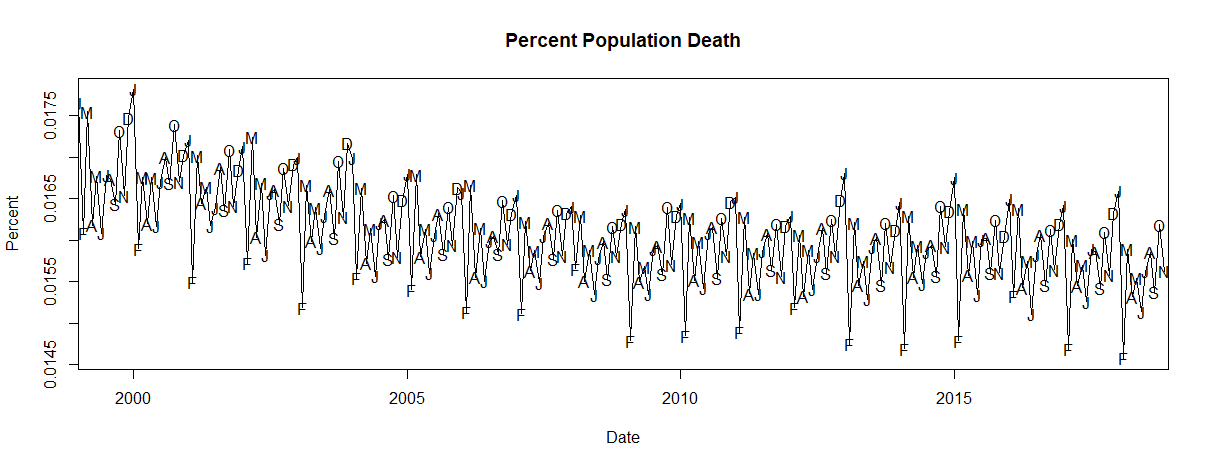
**Appendix 3: Model fit and Diagnostics- Cardiovascular Deaths:***Figure 10: Residual Plot for Final Model of Monthly Deaths Due to Cardiovascular Issues*

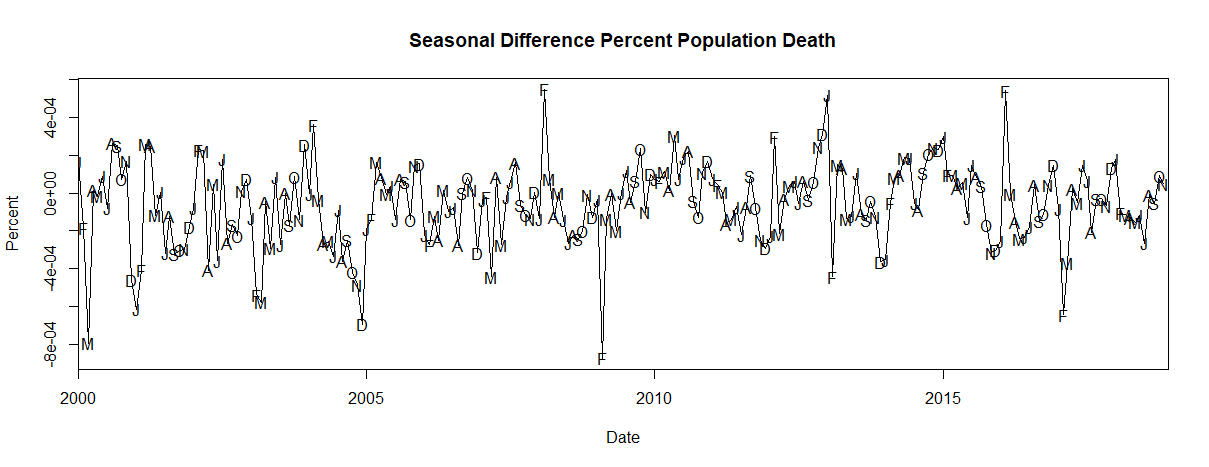
*****Figure 11:* *Autocorrelation Plot of Residuals for Final Model of Monthly Deaths Due to Cardiovascular Issues*

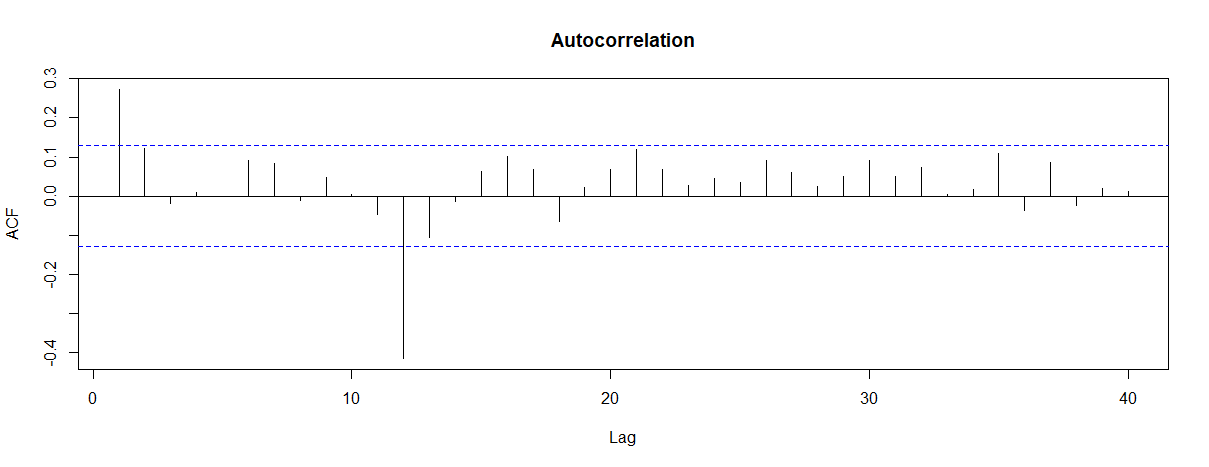
*****Figure 12: QQ plot of Residuals for Final Model of Monthly Deaths Due to Cardiovascular Issues*

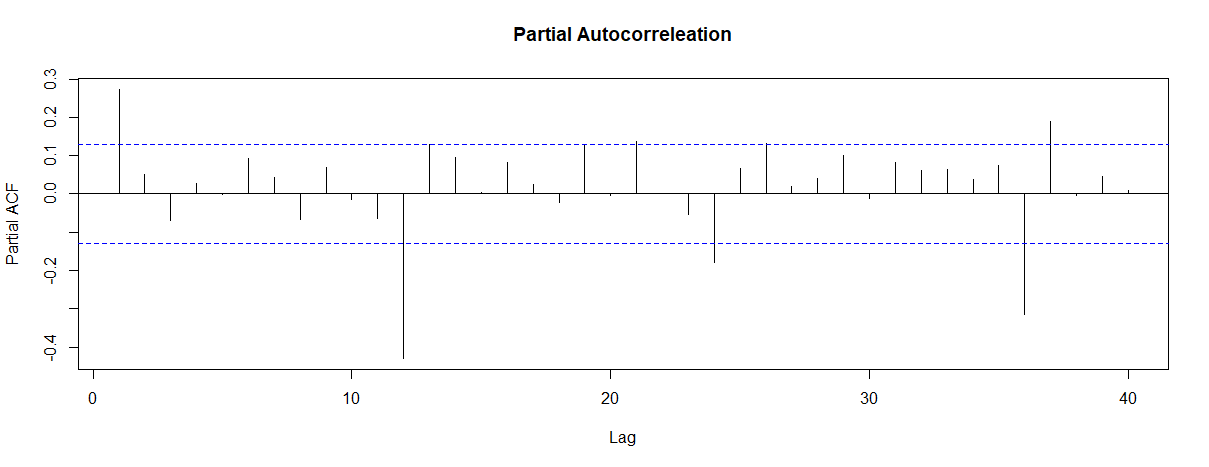
*****Figure 13: QQ plot of Residuals for Final Model of Monthly Deaths Due to Cardiovascular Issues*

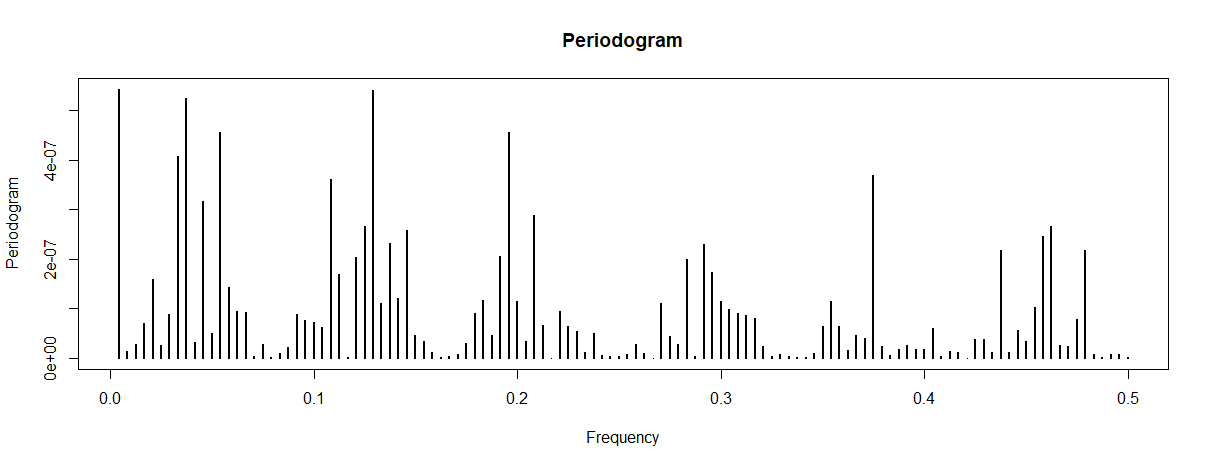
**Appendix 4: Preprocessing – Neoplasm Deaths:**

*****Figure 14: Monthly Percentage of Population Deaths due to Neoplasm Issues*

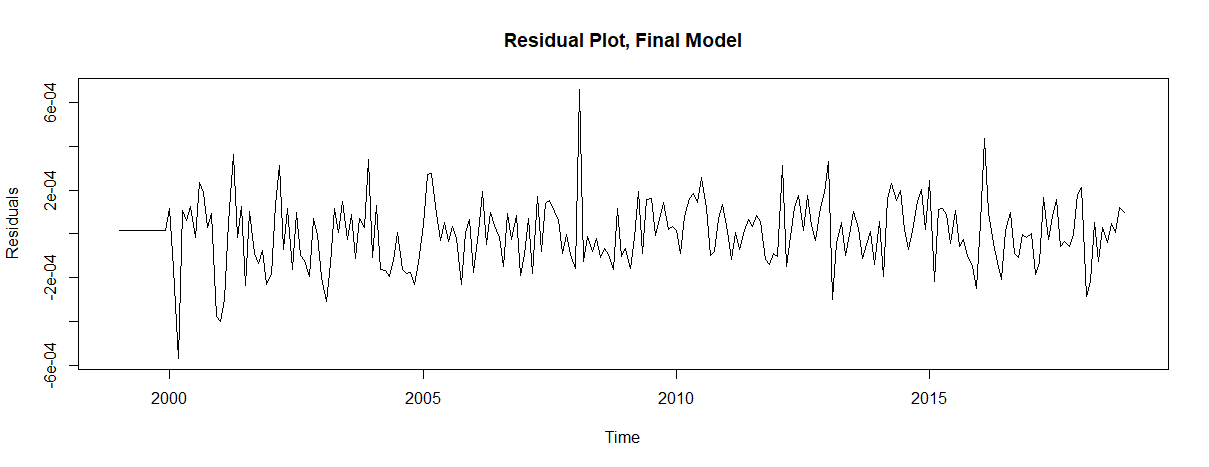
*****Figure 15: First Seasonal Difference of Population Deaths due to Neoplasm Issues*

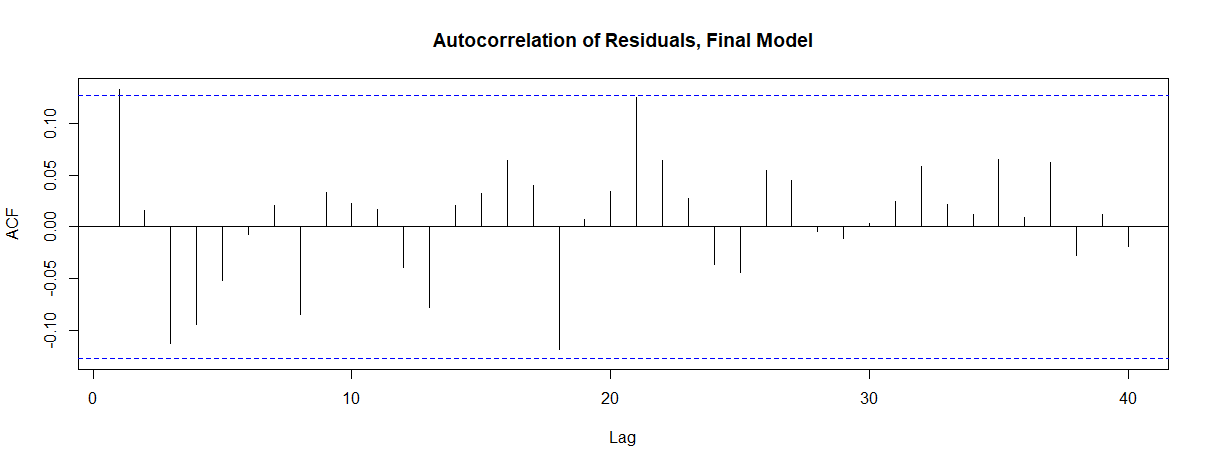
*****Figure 16: Autocorrelation Function of First Seasonal Differenced Deaths due to Neoplasm Issues*

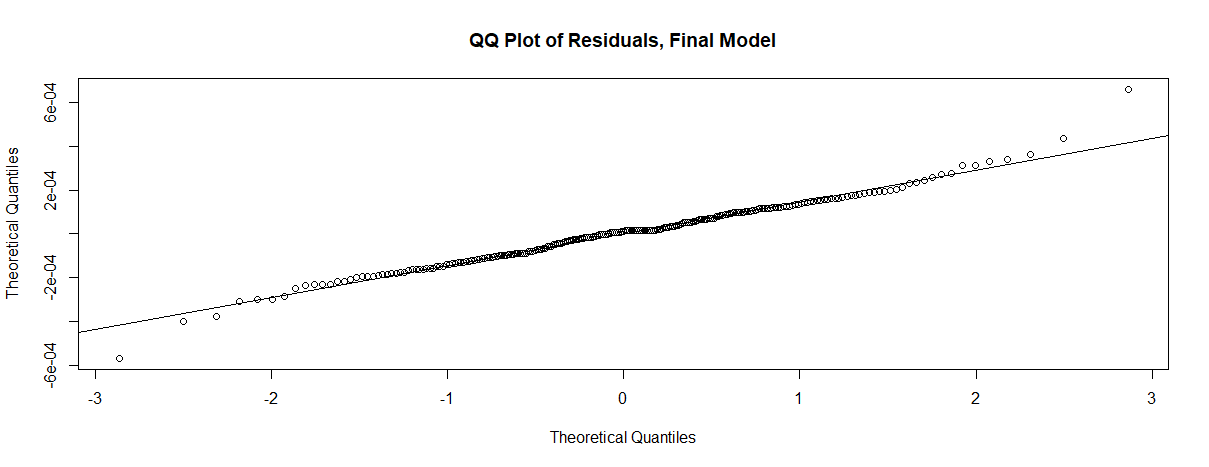
*****Figure 17: Partial Autocorrelation Function of First Seasonal Differenced Deaths due to Cardiovascular Issues*

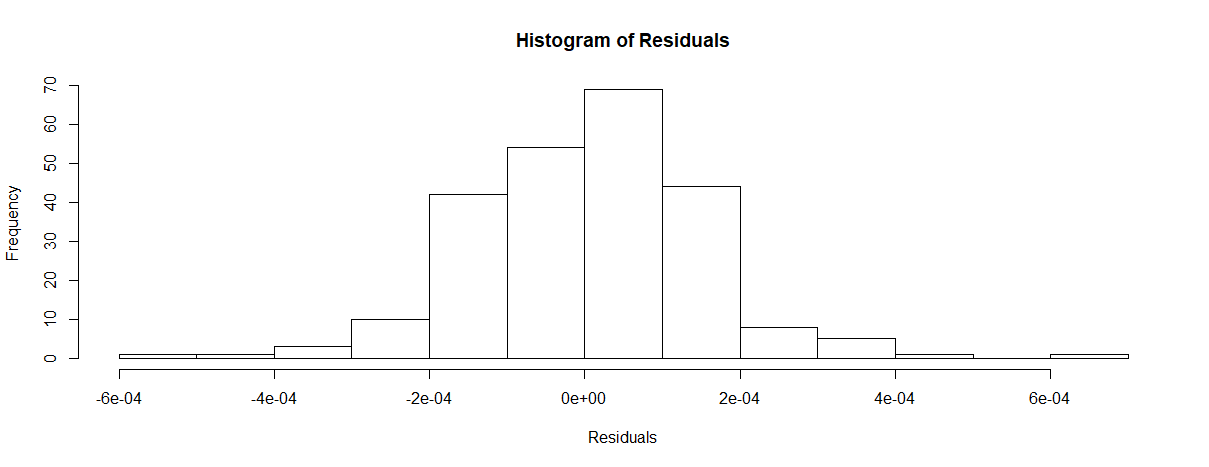
*****Figure 18: Periodogram of First Seasonal Differenced Deaths due to Cardiovascular Issues*

**Appendix 5: Model fit and Diagnostics- Neoplasm Deaths:**

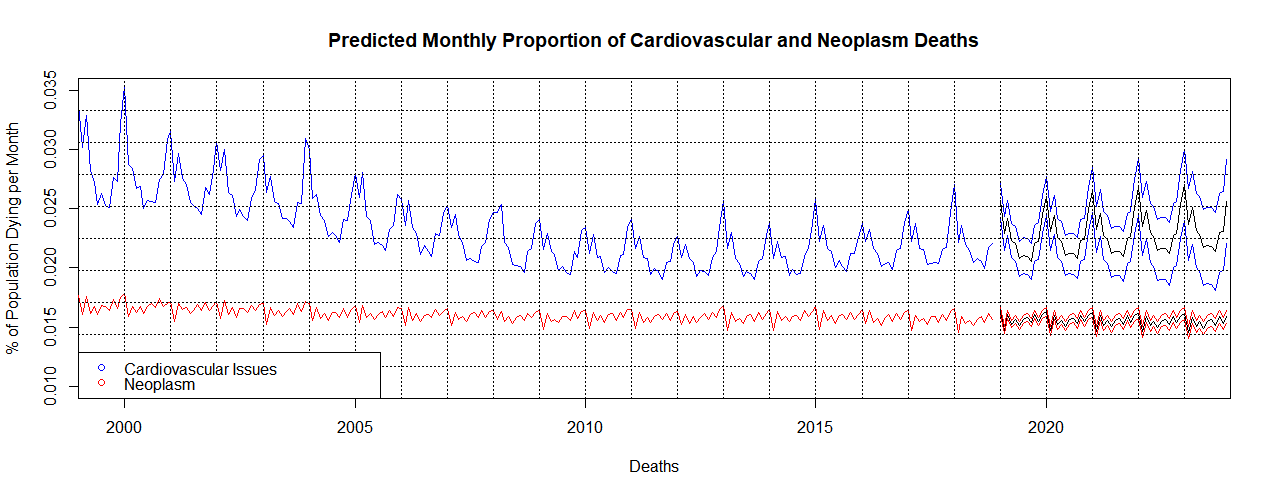
*****Figure 19: Residual Plot for Final Model of Monthly Deaths Due to Neoplasm Issues*

*****Figure 20: Autocorrelation Plot of Residuals for Final Model of Monthly Deaths Due to Neoplasm Issues*

*****Figure 21: QQ Plot of Residuals for Final model of Monthly Deaths Due to Neoplasm Issues*

*****Figure 22: Histogram of Residuals for Final model of Monthly Deaths Due to Neoplasm Issues*

**Appendix 6: Final Predictions:**

*****Figure 23: Plot of Monthly Deaths due to Cardiovascular and Neoplasm Issues, with 60 months predicted and 95% confidence intervals.*