

Time Series Analysis on U.S. Citizen Air Travel to Canada

Leanne Lee

```
## Warning: package 'gridExtra' was built under R version 3.2.4
```

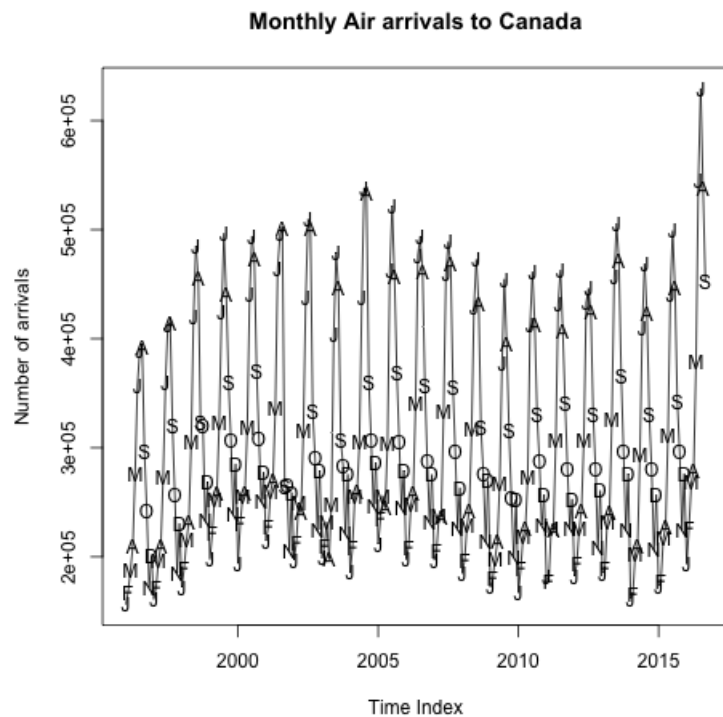
Introduction

The scientific question motivating my work is “is there a seasonal trend for U.S.citizens to travel to Canada by air?” Canada is one of the closest country across the border of the United States. Many U.S. citizen consider Canada a great spot to travel during holiday. In the West Coast, we neighbor near Vancouver and Victoria, while the east coast has Toronto, Montreal and Quebec.

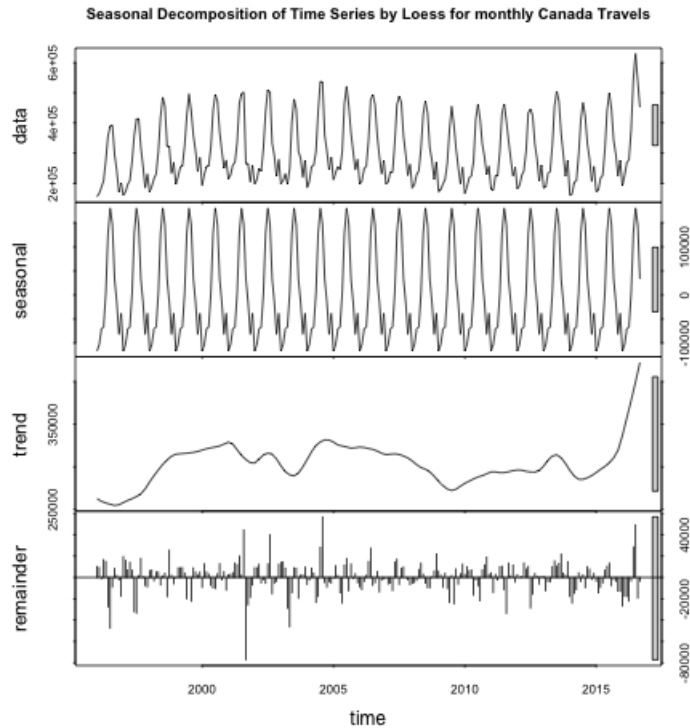
Dataset

The dataset comes from the U.S. Department of Commerce, National Travel & Tourism Office. [NTTO Monthly Tourism Statistics] (<http://travel.trade.gov/research/monthly/departures/index.html>). The dataset contain U.S. outbound travel by world regions from Jan 1996 to Sep 2016. There are 249 monthly data points throughout these twenty years.

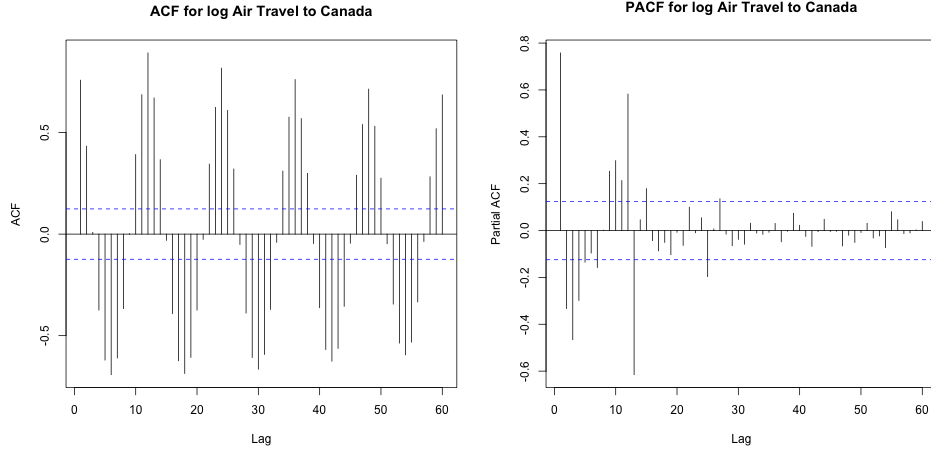
Exploratory Data Analysis



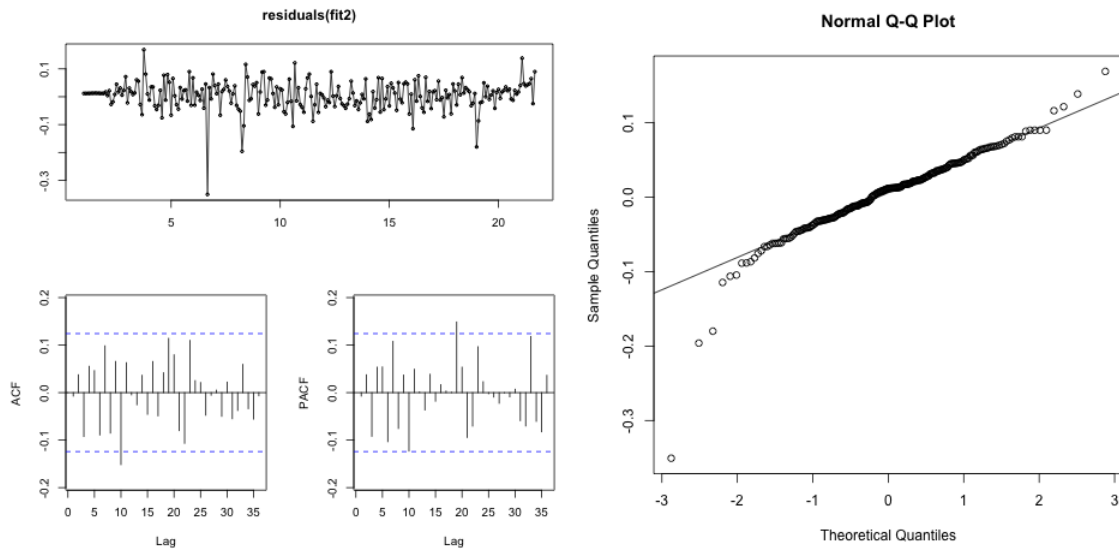
First, I plot the time series for the monthly air arrivals from U.S. to Canada. By examining the time series, the plot looks seasonal with high peak in June, July and August. Students usually travel during their summer break and this is the main drive of the increase in air travel in these months. There is also a smaller peak in between of each cycle. The small peaks take place in December, which is during Christmas break and students tend to go travel more often. Because the graph doesn't get effect when the variation increase with the level of the series, so I didn't consider the log of the time series.



After the basic inspection on the time series, I ran the `stl()` command for a season decomposition of Time Series by Loess. I first inspect the data section. The data looks stationary. I consider the bar on the right hand side as one unit of variation. The bar on the seasonal panel is slightly bigger than the data panel, meaning the seasonal effect is larger than the variation in the data. In the trend panel, it has a larger variation box than the data and seasonal panels, meaning the variation attributed to the trend is much smaller than the seasonal component. It also indicated the trend is not dominating, but there's an upward trend at the end. The spike upward trend of air traveling to Canada in this summer is because the low Canadian dollar, causing a boost to tourism in Canada. The remainder panel shows there are some high residuals in year of 2002 - year of 2005 and year of 2016.

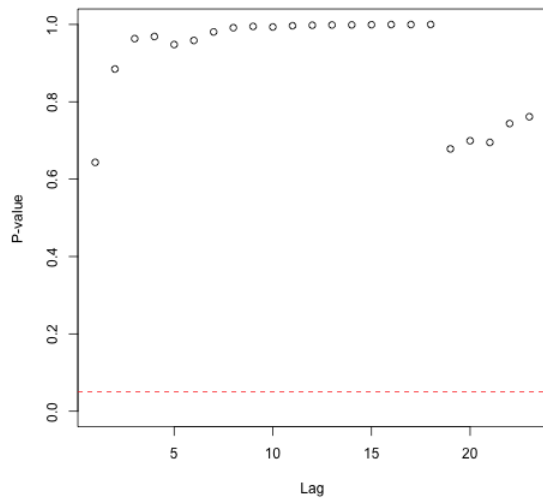


Based on the results on ACF and PACF, I found that there are seasonal autoregressive and moving average effects. So I ran the ARIMA model with changing order of p, d, q and seasonal orders. First, I ran the auto arima to get a rough estimate of my orders of seasonal ARIMA. Since the original series looks stationary, I do not need to consider the high order of differencing(d). Since the seasonal pattern is strong and stable, I use $(0,1,2)$ as a seasonal difference. The best fit ARIMA model with the lowest AIC is **ARIMA** $(2, 0, 1)X(0, 1, 2)_{12}$



After fitting the model, it is important to diagnostic the residuals of the ARIMA model. The blue line indicates the 95% confidence interval under the null hypothesis of white noise. Both ACF and PACF show the majority of residuals are under the blue threshold and it passes

the residual test. Then, I examined the QQ plot to check the normality of the residuals. The plot shows there are a few outliers in the beginning. The most significant one is the first point, which causes by the extreme high value during July 2016.



Before I perform the ARCH-GARCH model, I ran the McLeod-Li test to see if the p-values reject my null hypothesis.