# Time Series Analysis: Citibike NYC

Cesar Rene Pabon Bernal
Stat 715: Professor Dana Sylvan
May 13, 2019

# OUTLINE

◎INTRODUCTION

1. History
2. Data
3. Plot Time Series/Transformations/Auto Correlation Structure

◎METHODS

1. Model Specification
2. Model Fitting
3. Model Diagnostics
4. Forecasting

◎CONCLUSION

# INTRODUCTION: HISTORY

**1967 Amsterdam: Witte Fietse**

**1995 Copenhagen: Bycyklen**

**1996 Portsmouth UK: Bikeabout**

**2007 Paris, France: Vélib & Barcelona, Spain: Bicing**

**2008 Washington DC: SmartBike DC**

**2013 NYC: Citibike Launches w/ 6000 bikes in Manhattan and BK**

**2019 As of Feb, Citibike has:**
**+750 stations**
**+62K rides/day**
**+12K bikes**
**Total revenue for month~$4.7 mill**


1. *"The Real Story Behind the Global Bike-Share Boom." CityLab, www.citylab.com/city-makers-connections/bike-share/.*
2. *Motivate International, Inc. "Citi Bike Monthly Operating Reports." Citi Bike NYC, www.citibikenyc.com/system-data/operating-reports.*

# INTRODUCTION: HISTORY

Why are Bike Sharing Systems Beneficial?

◎Environmental

1. In Shanghai (2018), bike sharing systems cut down CO2 by 25,240 tons. A number big enough to consider as a component when meeting China's Paris Accord greenhouse emissions requirements

◎Economic

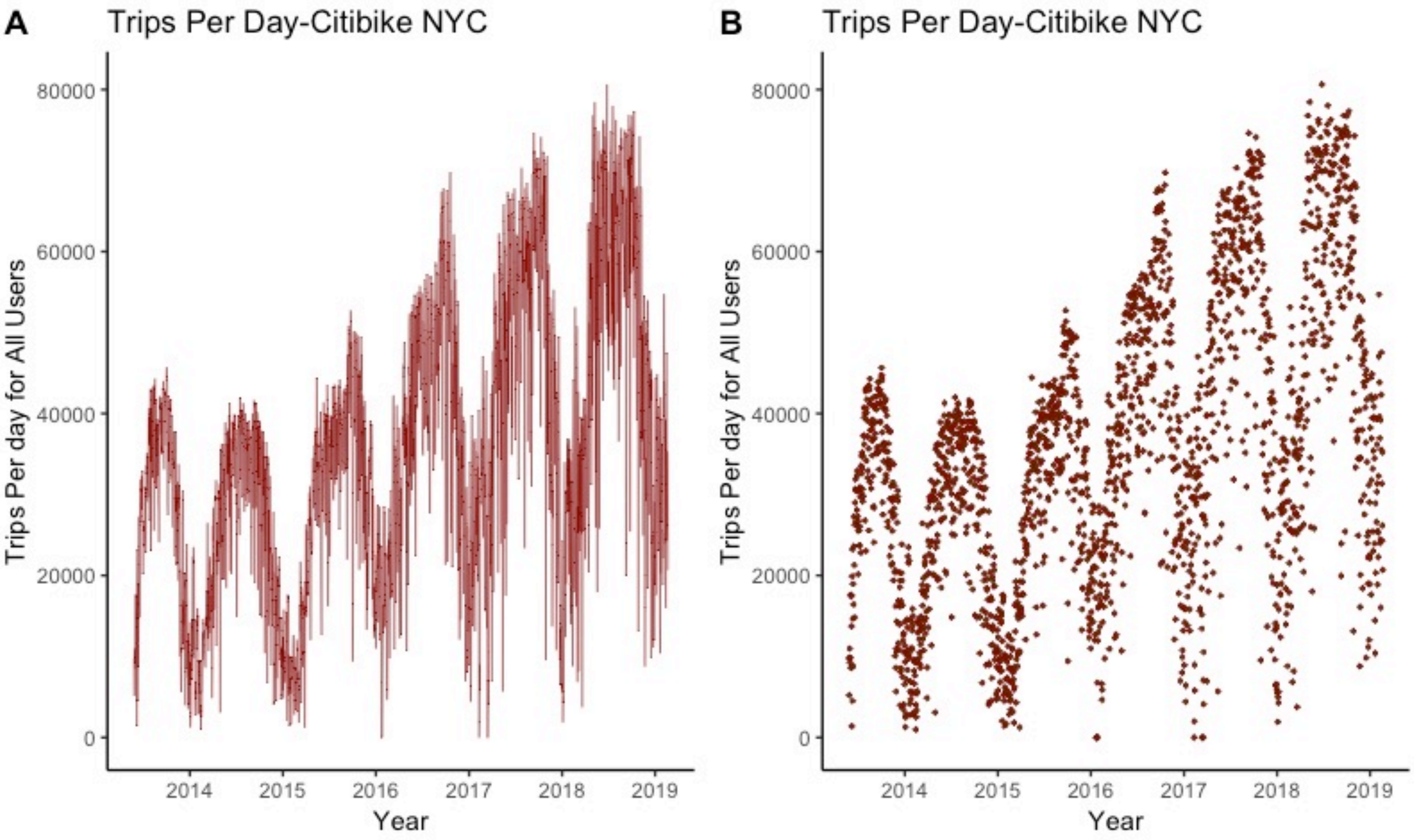
2. If all Bike Sharing trips replaced car trips in 12 major cities in Europe, 73.25 deaths could be avoided each year (225 million Euros saving)


3. *Zhang, Yongping, and Zhifu Mi. "Environmental Benefits of Bike Sharing: A Big Data-Based Analysis." Applied Energy, vol. 220, 2018, pp. 296–301., doi:10.1016/j.apenergy.2018.03.101.*
4. *Otero, I., et al. "Health Impacts of Bike Sharing Systems in Europe." Environment International, vol. 115, 2018, pp. 387–394., doi:10.1016/j.envint.2018.04.014.*

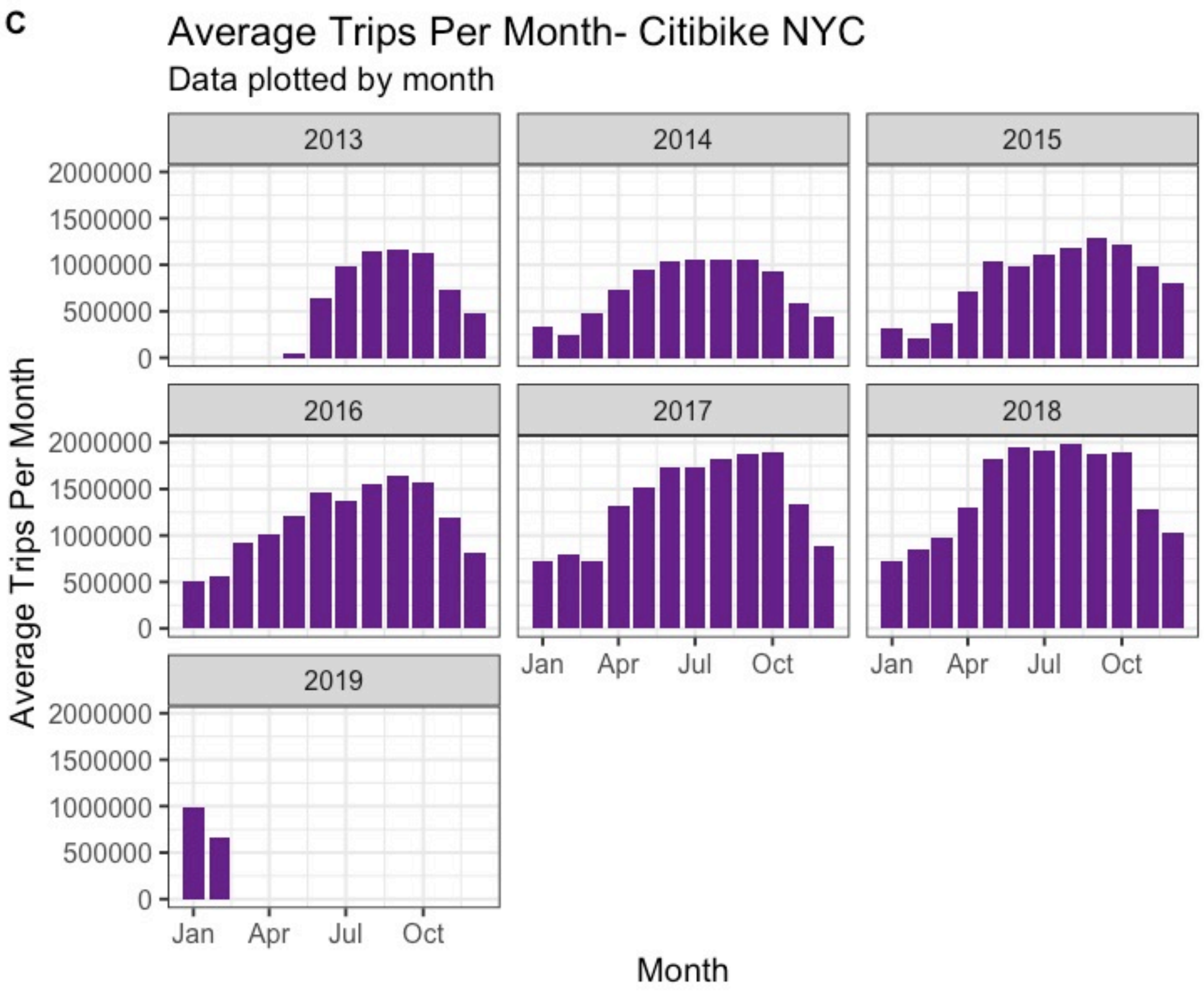
# INTRODUCTION: DATA

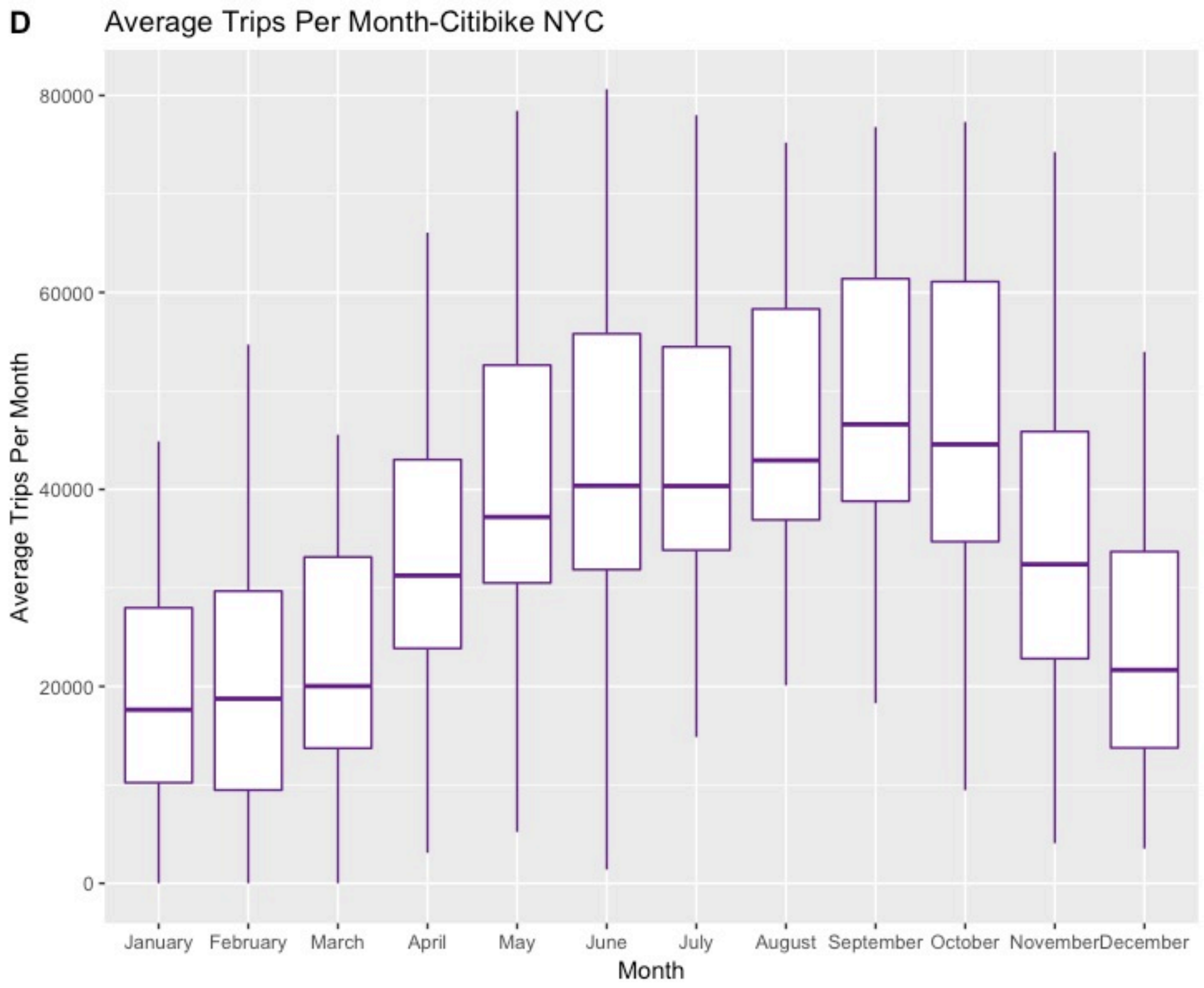
- Online Citibike System Data Repository
- Python:Cleaning, wrangling and exploratory analysis
- R:Statical application
- 2096 observations from 05/27/2013-02/15/2019
  - "Date" "Trips_per_24hrs"
- Hypothesis
  - Ho =Trips per day (for All Users) increase in the summer months

# INTRODUCTION: PLOT TS

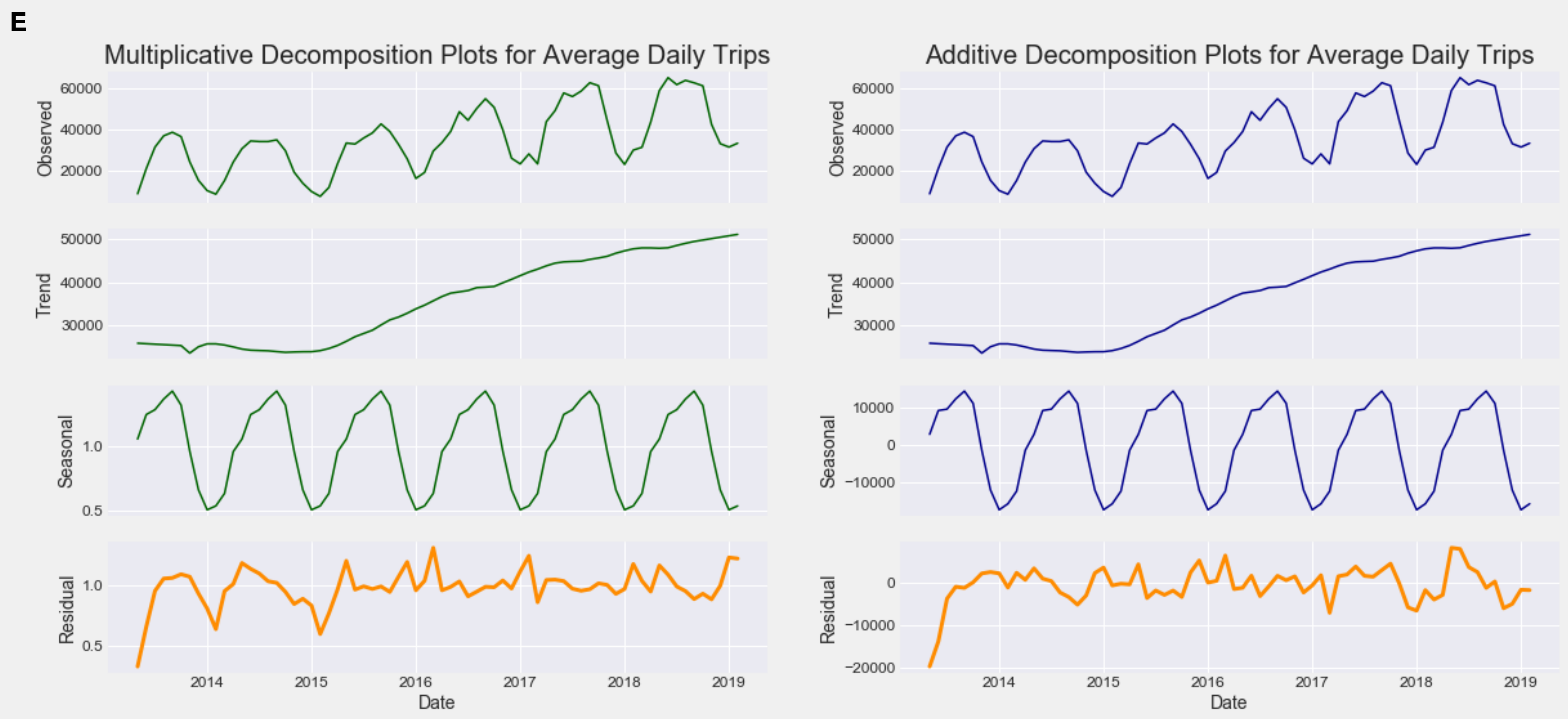


# INTRODUCTION: PLOT TS



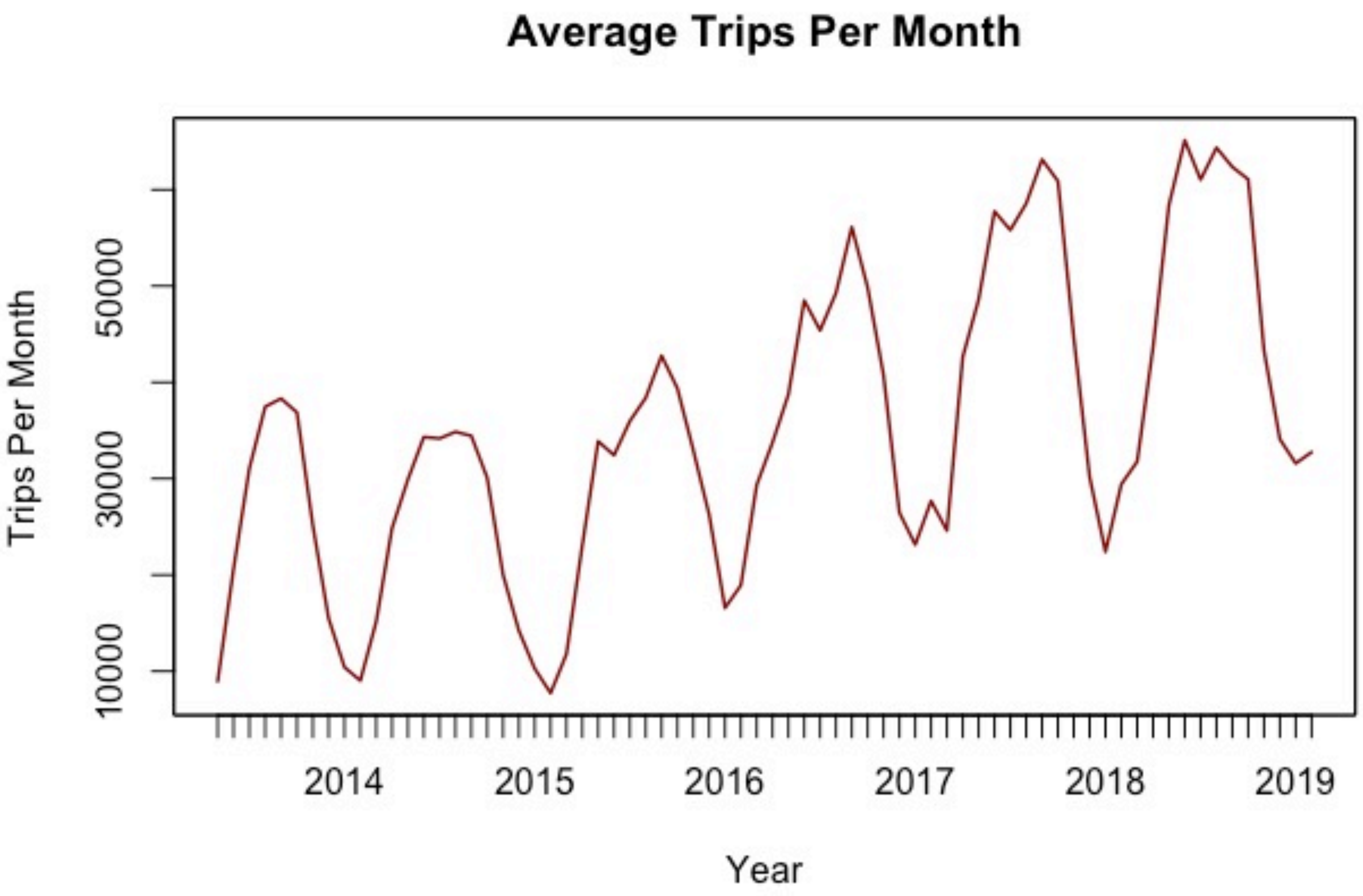



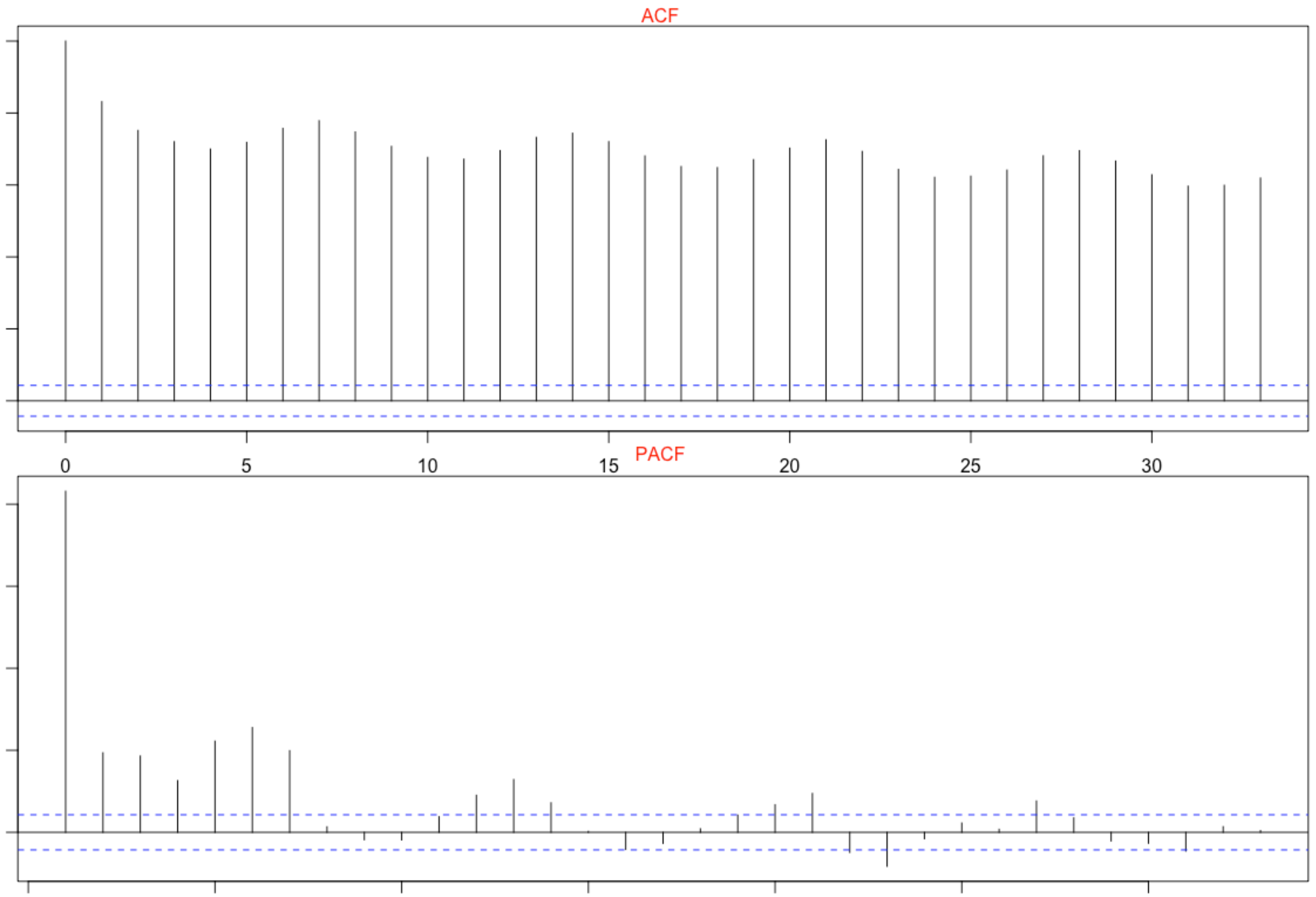
# INTRODUCTION: PLOT TS



# INTRODUCTION: PLOT TS

F

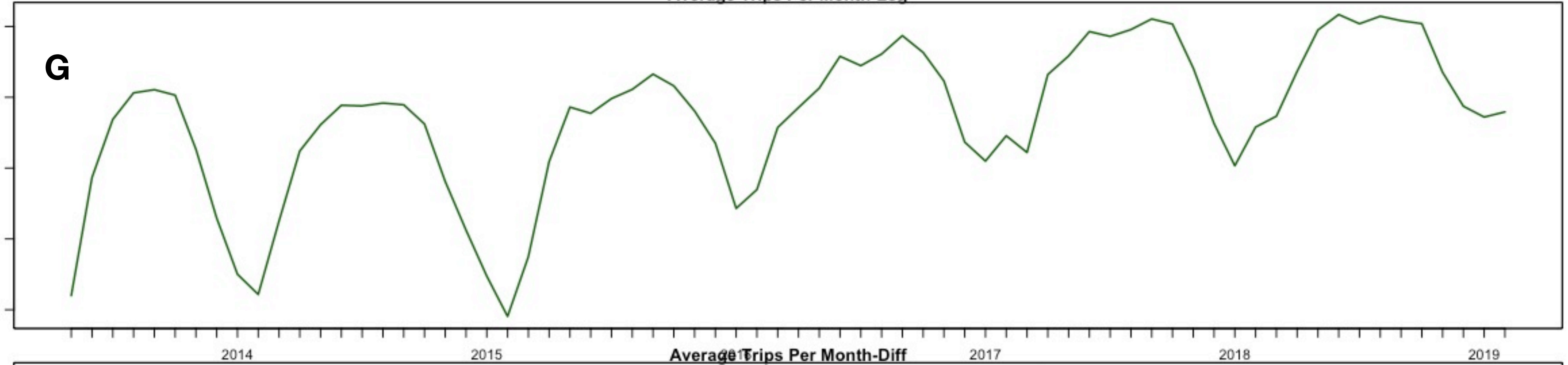



```
Augmented Dickey-Fuller Test

data:  z
Dickey-Fuller = -3.2223, Lag order = 12, p-value = 0.08406
alternative hypothesis: stationary
```

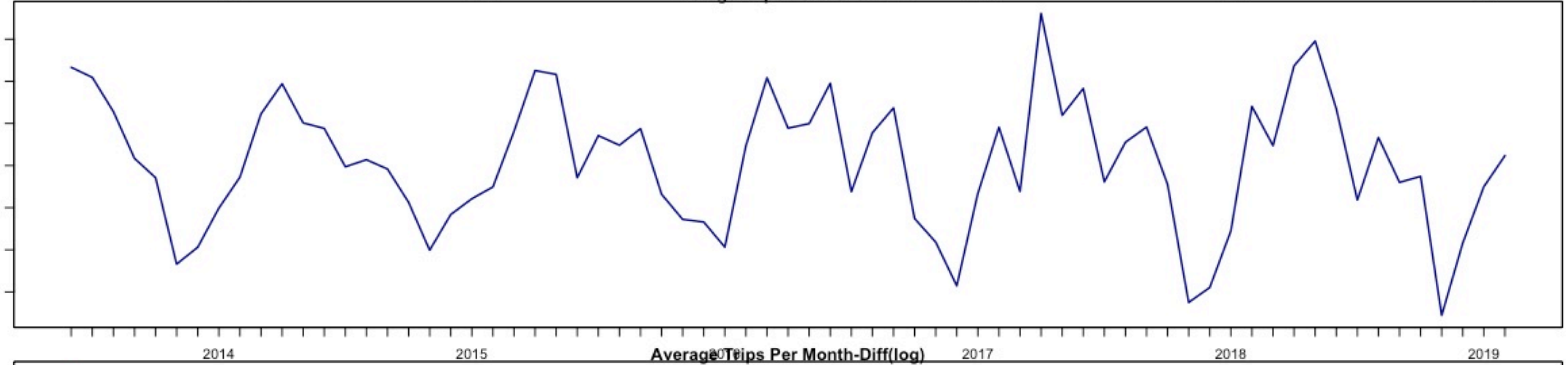
# INTRODUCTION: TRANSFORMATIONS

Average Trips Per Month-Log



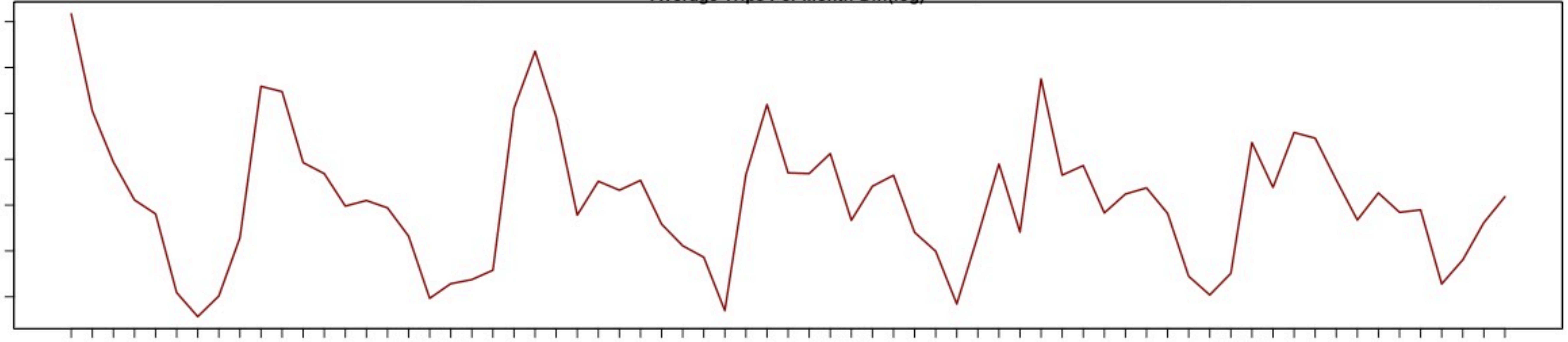

```
        Augmented Dickey-Fuller Test

data:  df_trips.mo.log
Dickey-Fuller = -5.4888, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Average Trips Per Month-Diff




```
        Augmented Dickey-Fuller Test

data:  df_trips.mo.diff
Dickey-Fuller = -5.731, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

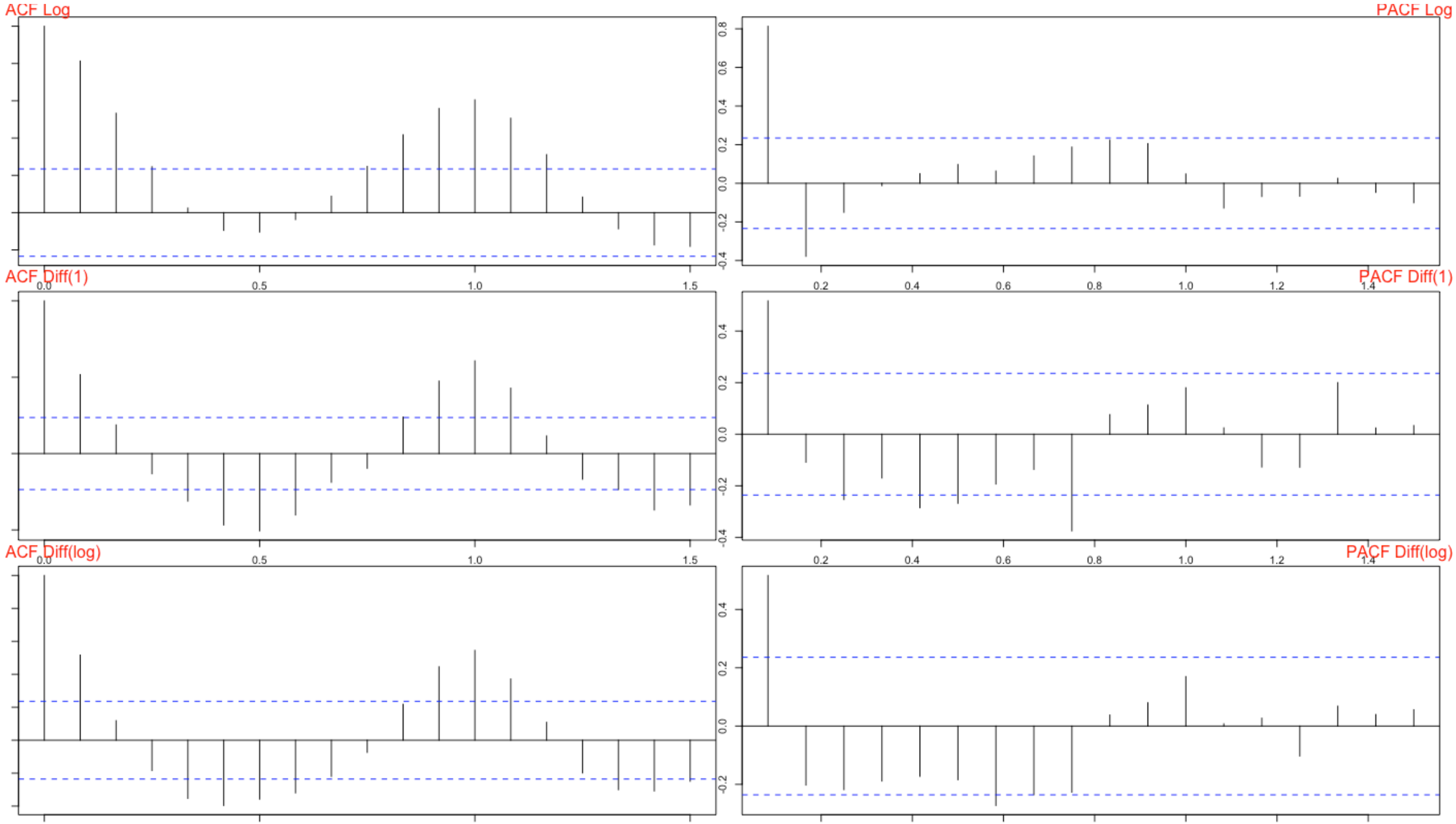
Average Trips Per Month-Diff(log)



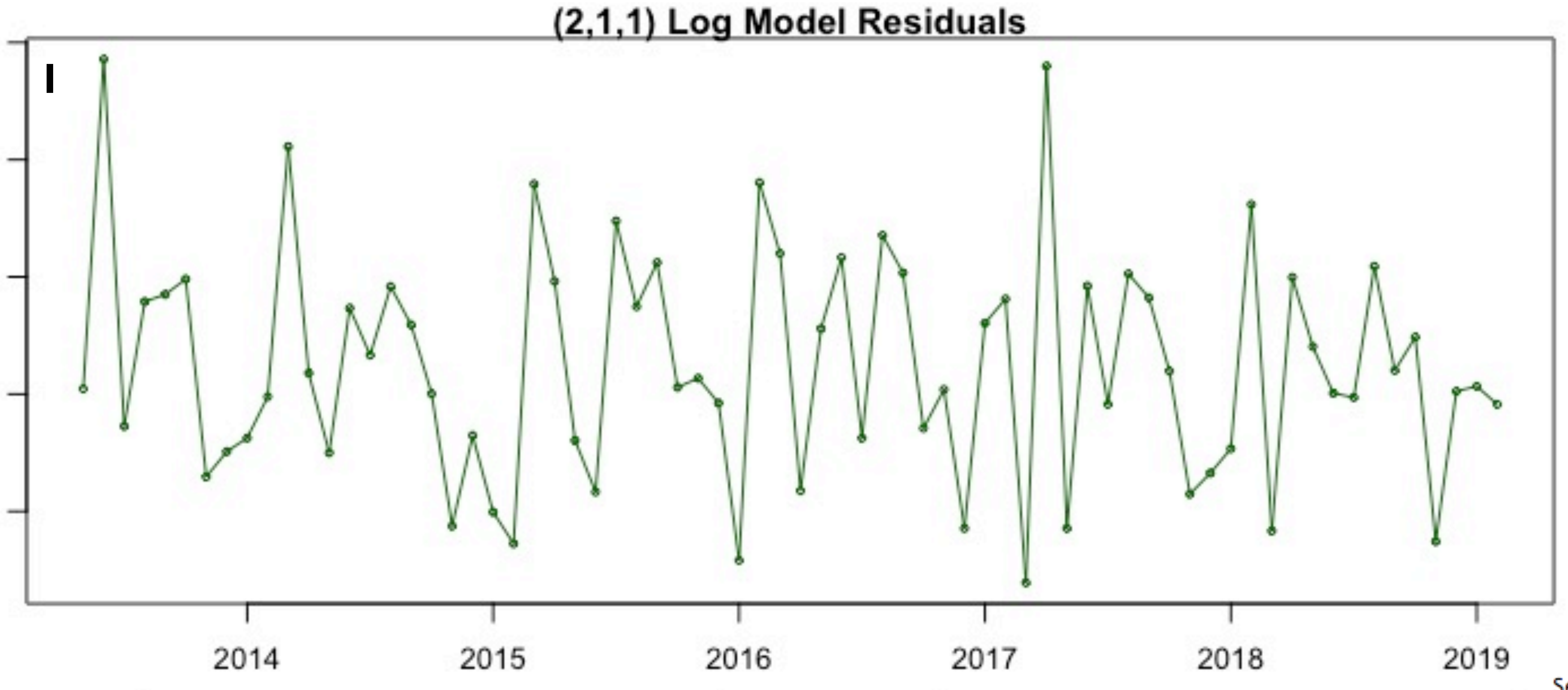
```
        Augmented Dickey-Fuller Test

data:  df_trips.mo.diff_log
Dickey-Fuller = -5.345, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

# INTRODUCTION: ACF STRUCTURES

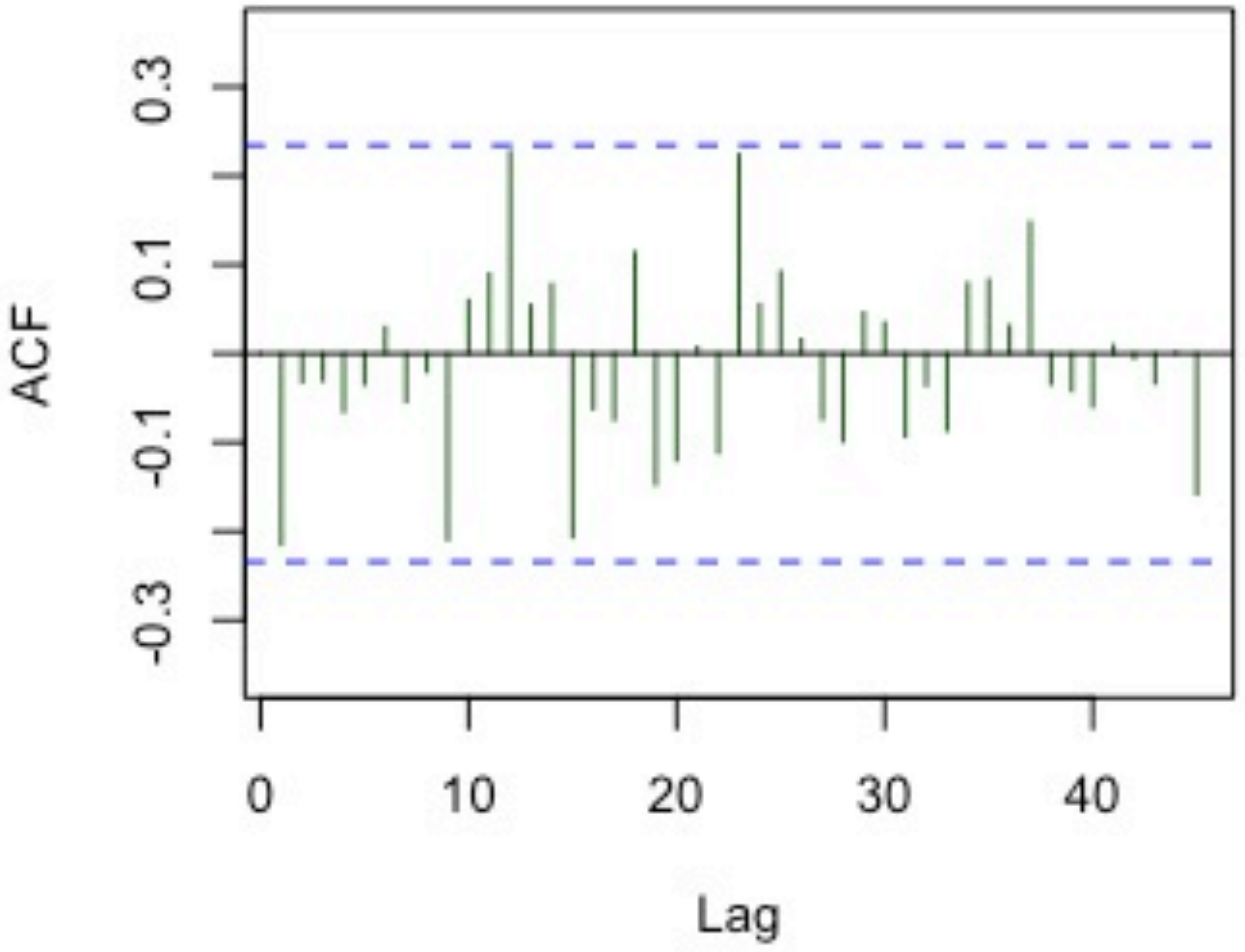
H

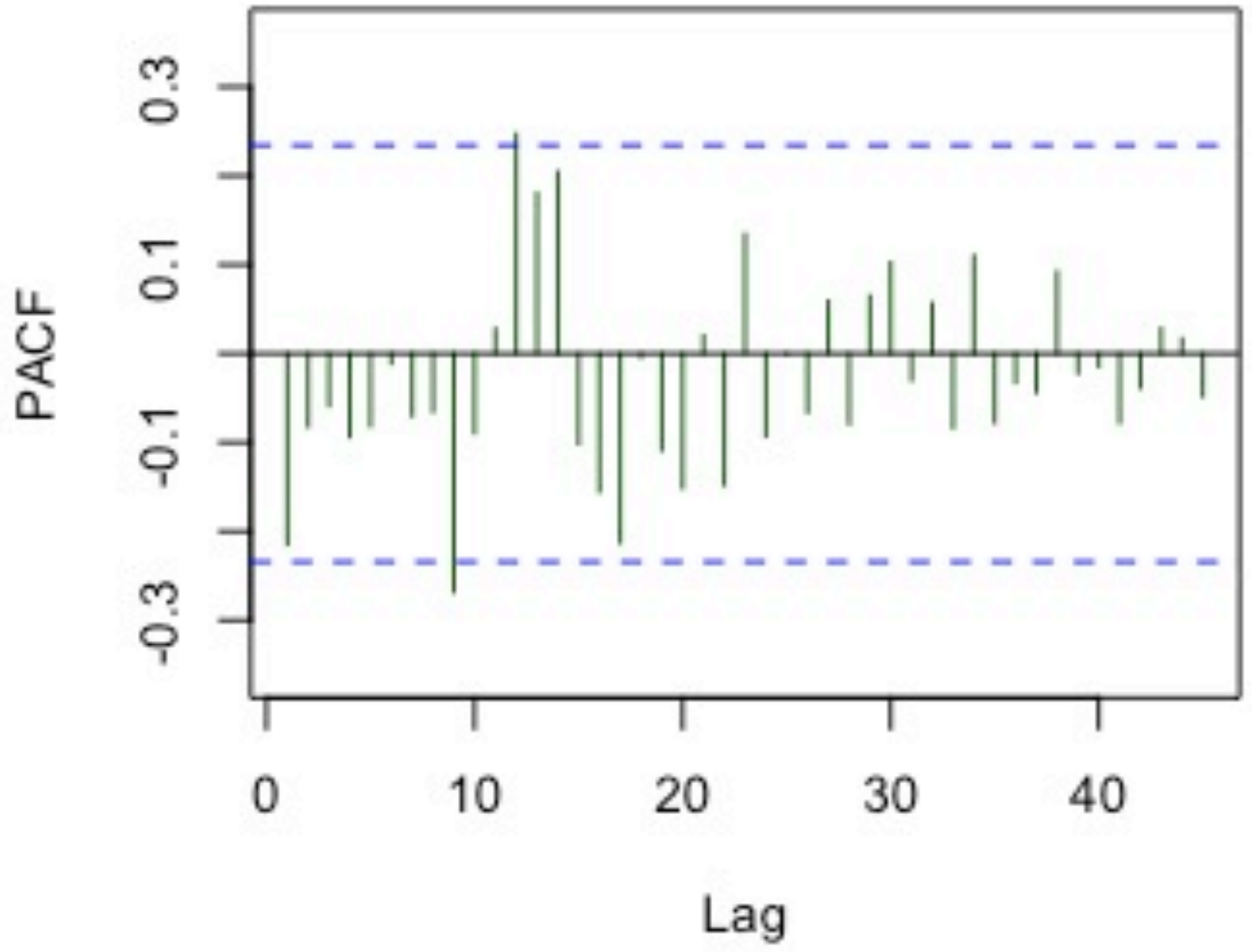


# METHODS: MODEL SPECIFICATION



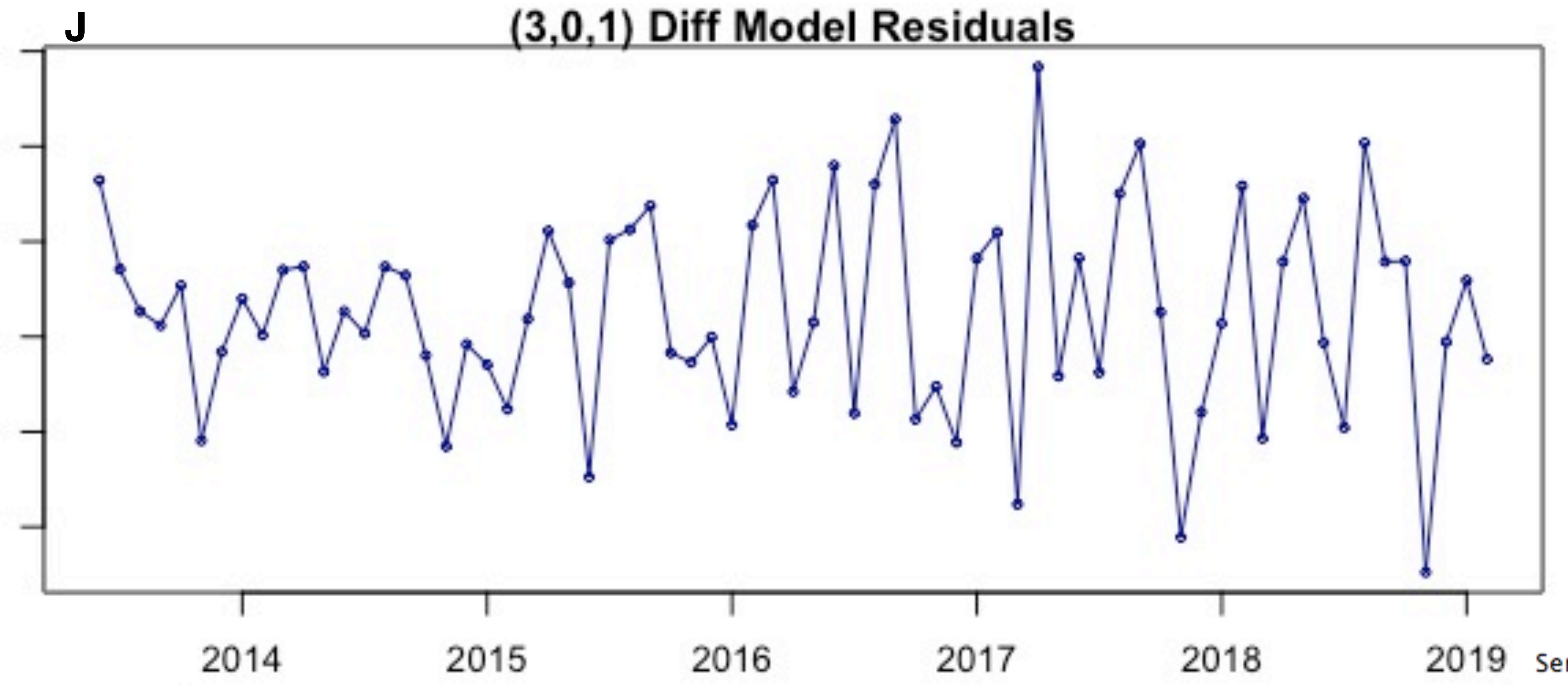

```
Series: df_trips.mo.log
ARIMA(2,1,1)

Coefficients:
         ar1      ar2      ma1
      1.3995  -0.7231  -0.8854
s.e.  0.0920   0.0890   0.0488

sigma^2 estimated as 0.04215:  log likelihood=11.94
AIC=-15.88   AICc=-15.26   BIC=-6.94
```



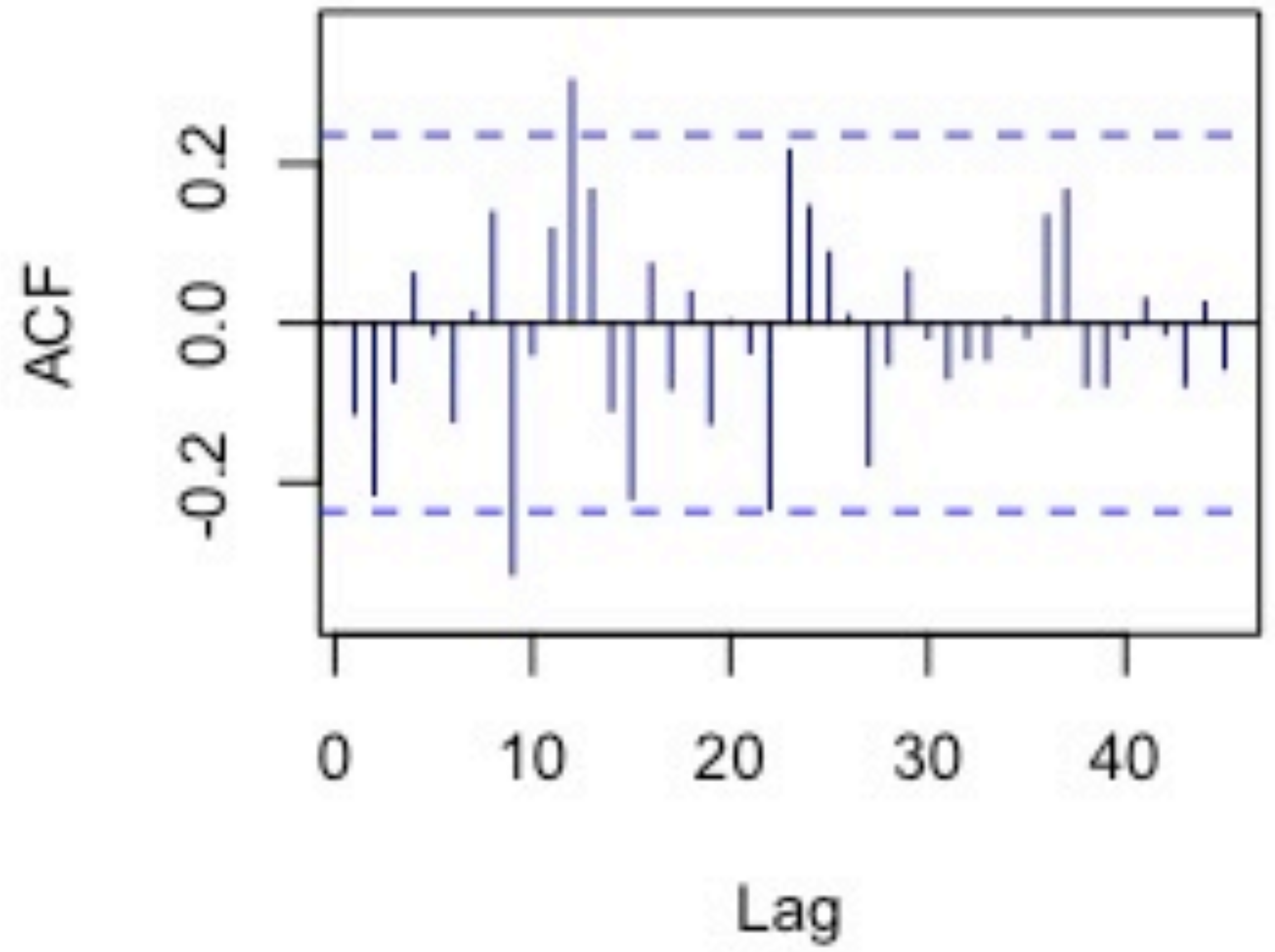



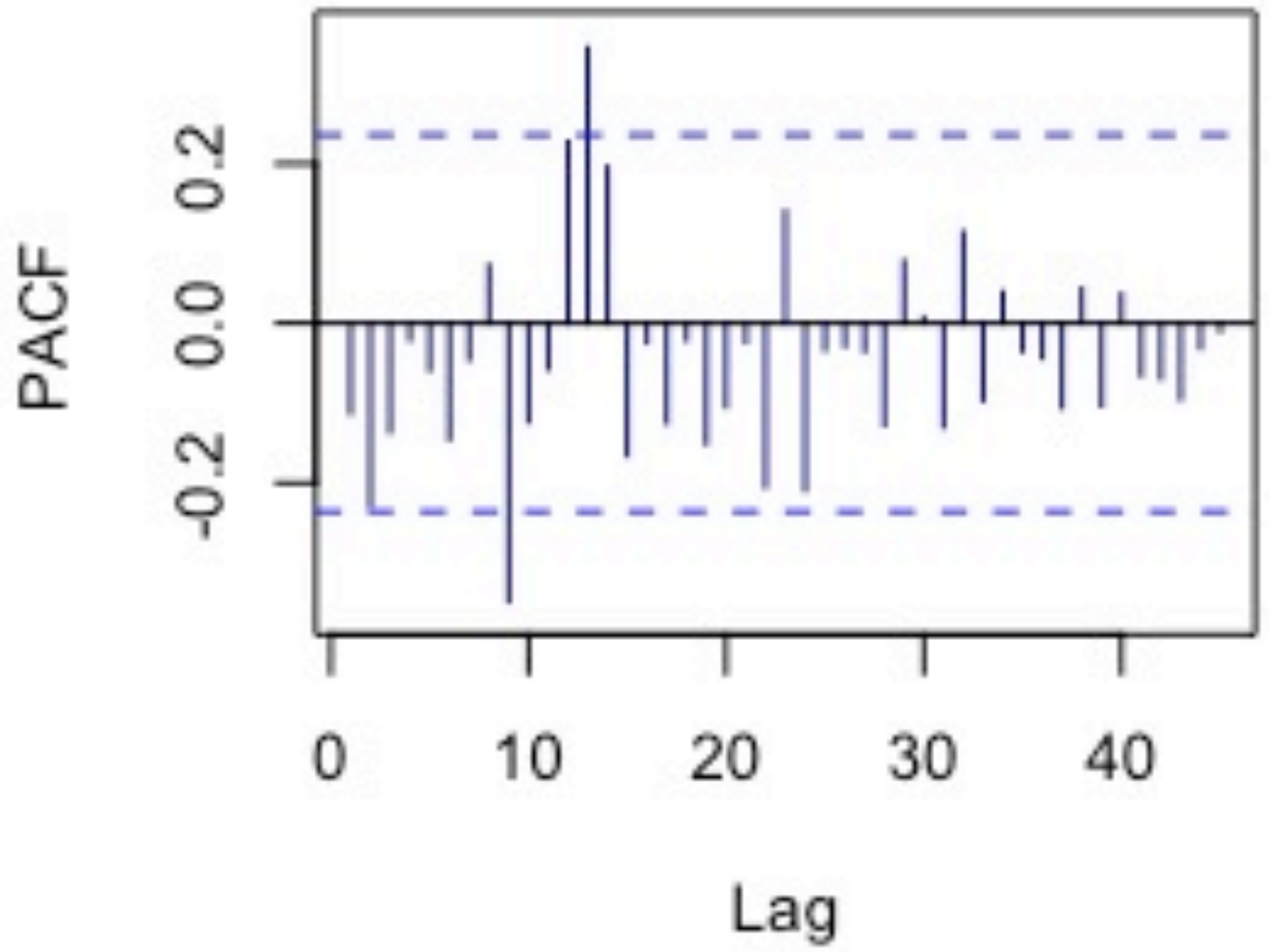

```
Series: df_trips.mo.log
ARIMA(2,1,1)

Coefficients:
         ar1      ar2      ma1
      1.3995  -0.7231  -0.8854
s.e.  0.0920   0.0890   0.0488

sigma^2 estimated as 0.04215:  log likelihood=11.94
AIC=-15.88   AICc=-15.26   BIC=-6.94

Training set error measures:
                     ME      RMSE       MAE      MPE     MAPE      MASE        ACF
Training set 0.04874772 0.1993619 0.1565437 0.454322 1.523958 0.6844499 -0.214706
```

# METHODS: MODEL SPECIFICATION




```
Series: df_trips.mo.diff
ARIMA(3,0,1) with zero mean

Coefficients:
         ar1      ar2      ar3      ma1
      1.1520  -0.2535  -0.3063  -0.8573
s.e.  0.1197   0.1801   0.1191   0.0520

sigma^2 estimated as 31277961:  log likelihood=-692.27
AIC=1394.54   AICc=1395.5   BIC=1405.71
```



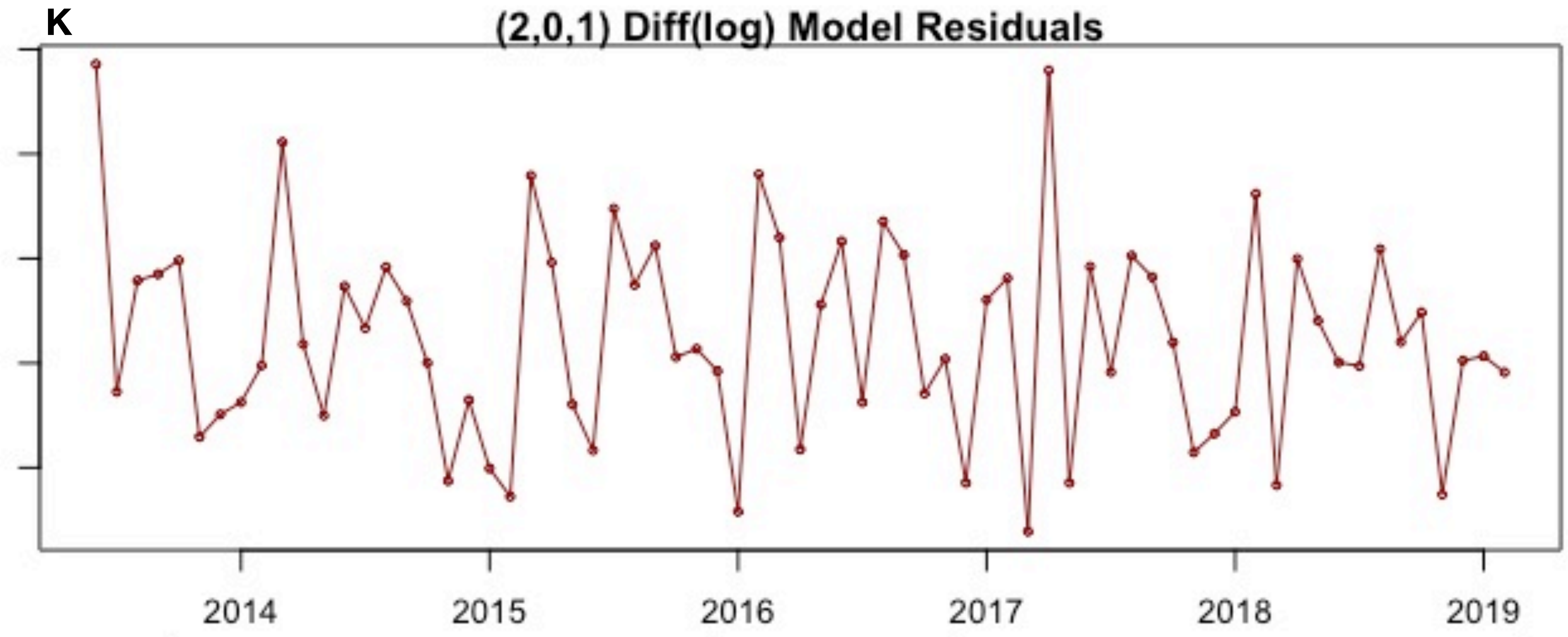


```
Series: df_trips.mo.diff
ARIMA(3,0,1) with zero mean

Coefficients:
         ar1      ar2      ar3      ma1
      1.1520  -0.2535  -0.3063  -0.8573
s.e.  0.1197   0.1801   0.1191   0.0520

sigma^2 estimated as 31277961:  log likelihood=-692.27
AIC=1394.54   AICc=1395.5   BIC=1405.71
```

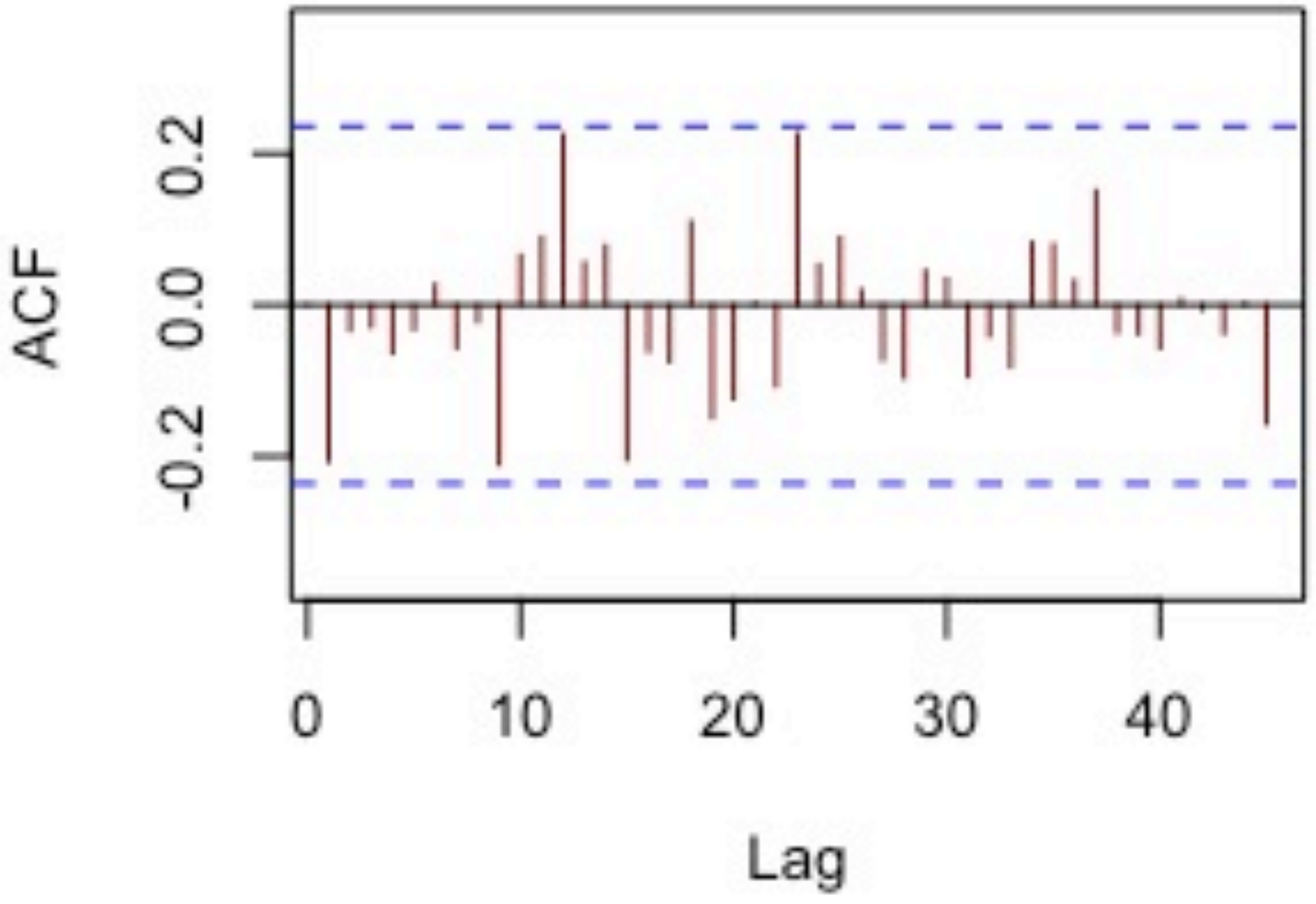

```
Training set error measures:
                   ME     RMSE      MAE    MPE     MAPE     MASE        ACF1
Training set 1250.249 5428.144 4315.528 43.436 112.9727 1.014973 -0.112668
```

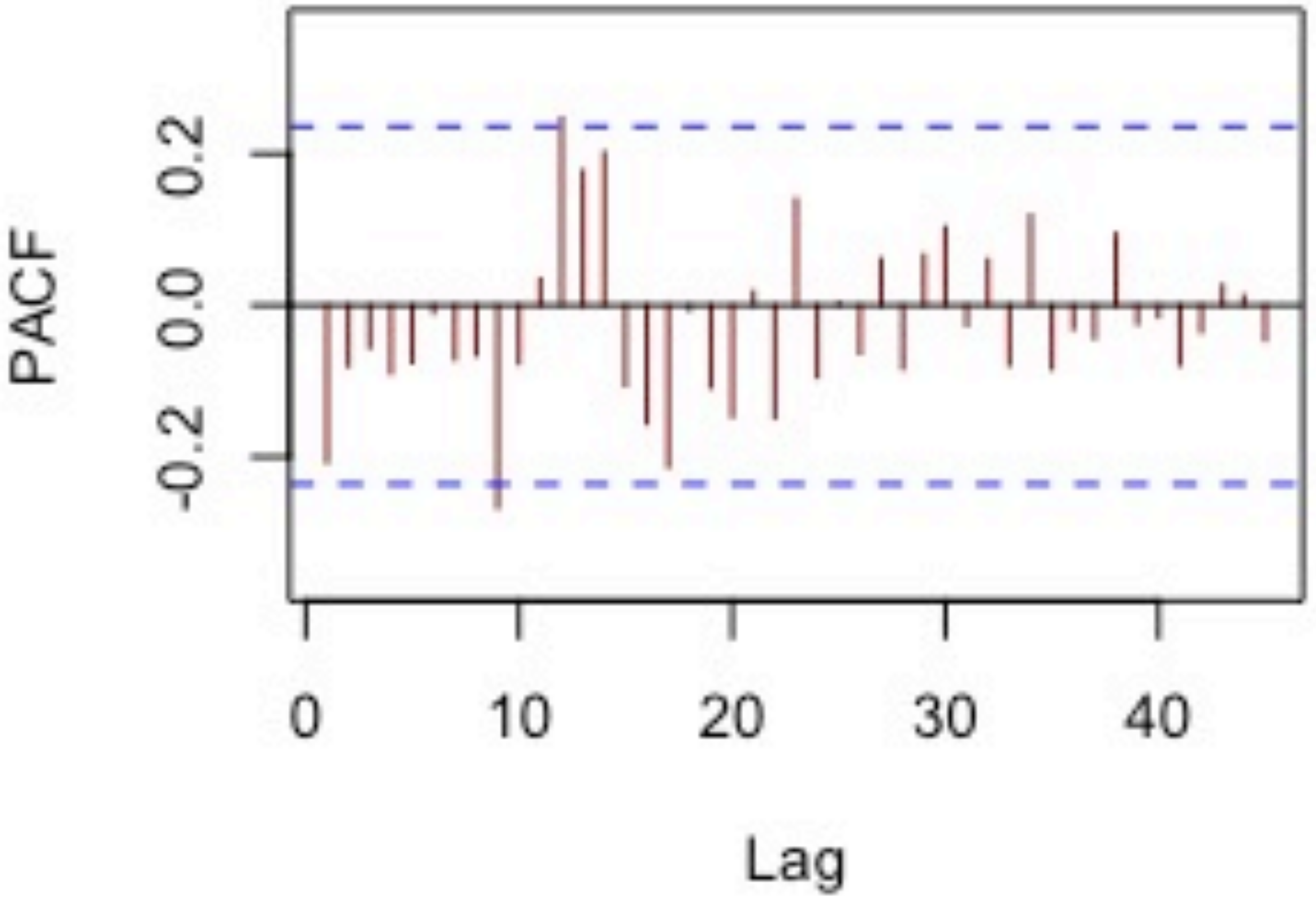
# METHODS: MODEL SPECIFICATION

K




```
Series: df_trips.mo.diff_log
ARIMA(2,0,1) with zero mean

Coefficients:
         ar1      ar2      ma1
      1.3995  -0.7231  -0.8854
s.e.  0.0920   0.0890   0.0488

sigma^2 estimated as 0.04215:  log likelihood=11.94
AIC=-15.88   AICc=-15.26   BIC=-6.94
```



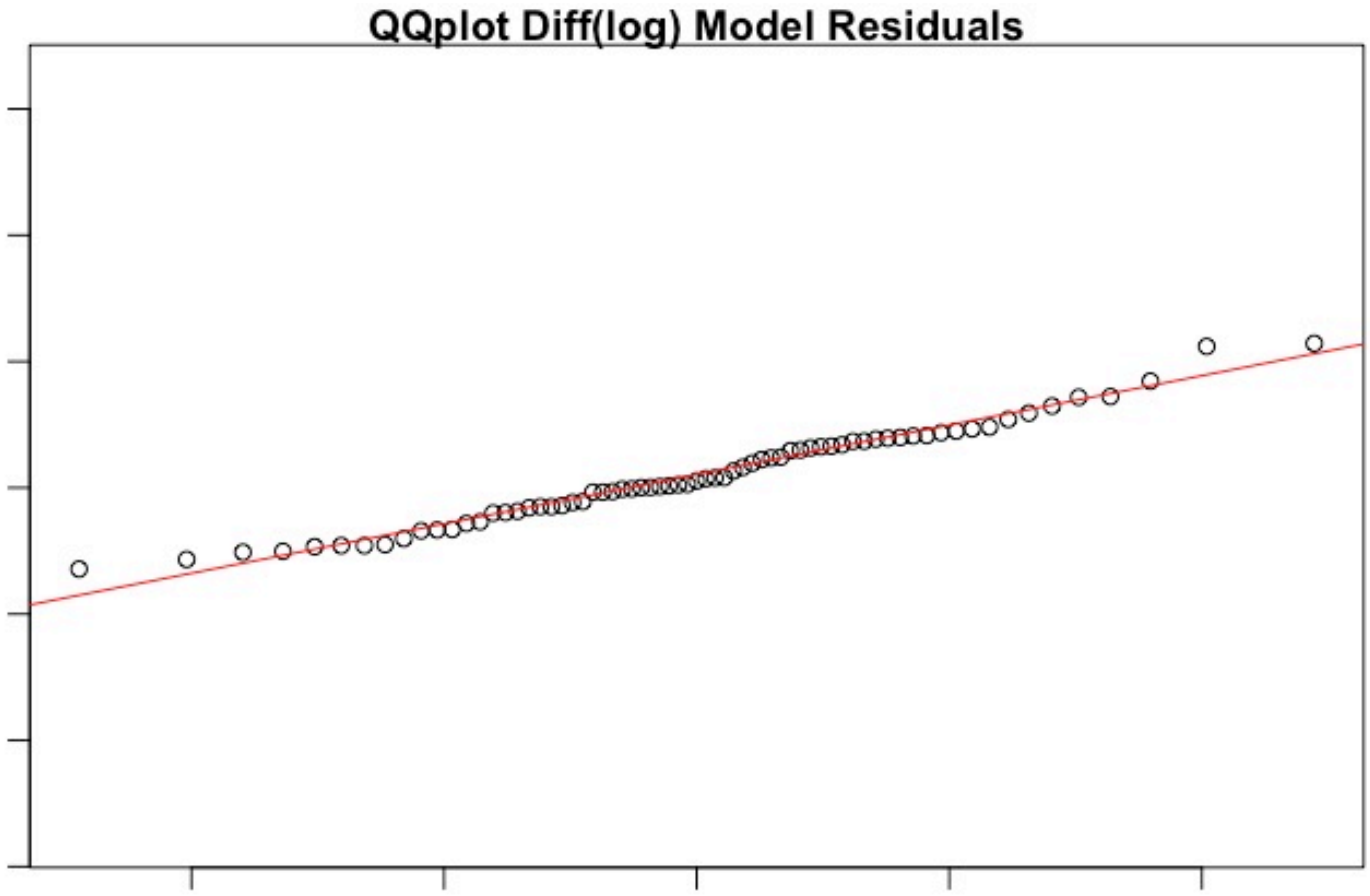





```
Series: df_trips.mo.diff_log
ARIMA(2,0,1) with zero mean

Coefficients:
         ar1      ar2      ma1
      1.3995  -0.7231  -0.8854
s.e.  0.0920   0.0890   0.0488

sigma^2 estimated as 0.04215:  log likelihood=11.94
AIC=-15.88   AICc=-15.26   BIC=-6.94
```


```
Training set error measures:
                     ME      RMSE       MAE      MPE     MAPE     MASE       ACF
Training set 0.04932225 0.2007986 0.1586807 33.86179 166.8318 1.043891 -0.20681
```

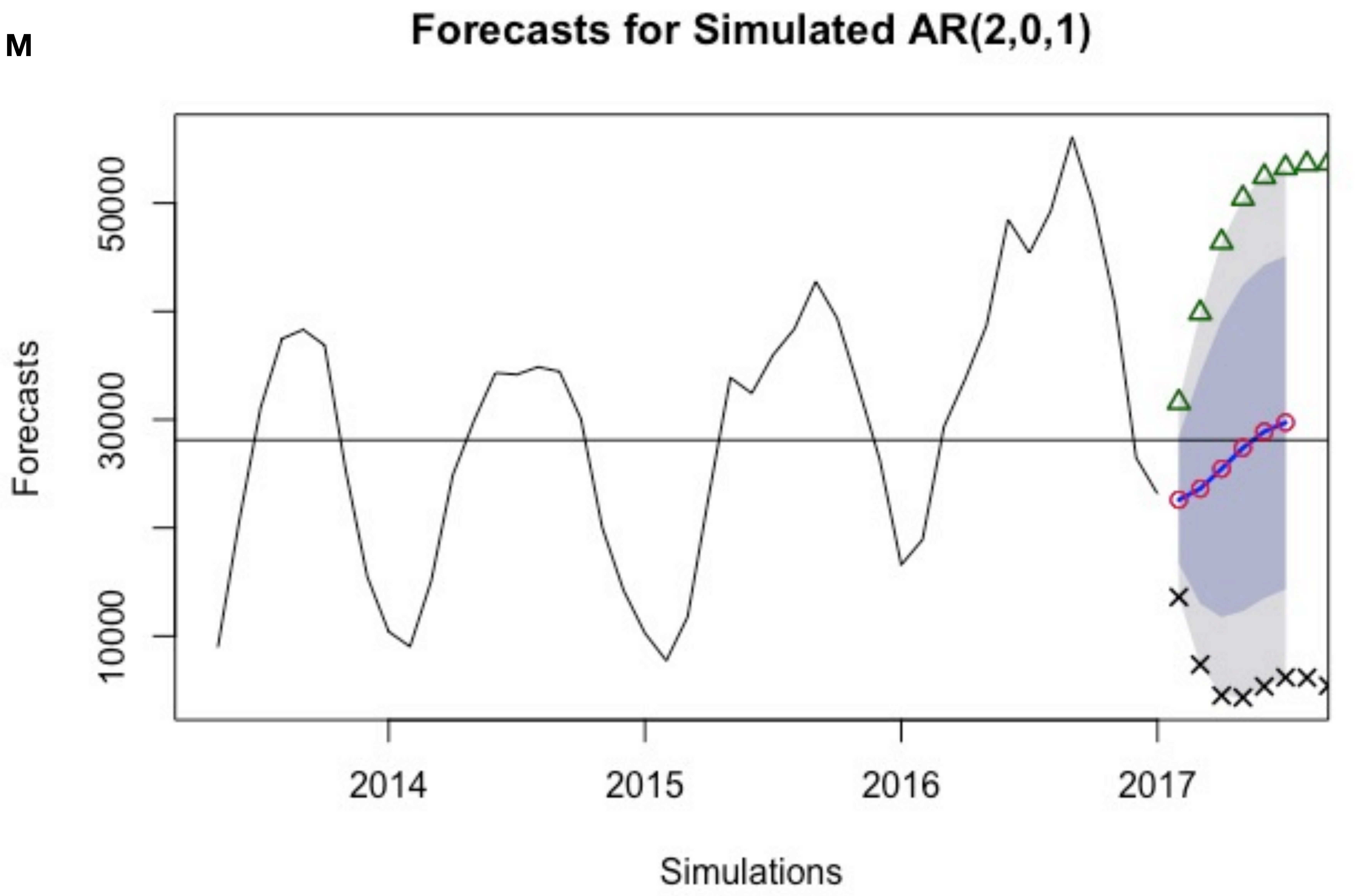
# METHODS: MODEL DIAGNOSTICS




```
        Ljung-Box test

data:  Residuals from ARIMA(2,0,1) with zero mean
Q* = 13.77, df = 11, p-value = 0.246

Model df: 3.   Total lags used: 14
```

# METHODS: FORECASTING

M



# CONCLUSION

- Trips per Day is a Seasonal TS
- AR(2,0,1) of Diff(log) was best model
    1. Best ACF/PCAF plots
    2. Lowest RMSE
- Forecast of Trips per Day confirmed proper selection of model
    - Test values found in lower and upper 95%

# REFERENCES


1. “The Real Story Behind the Global Bike-Share Boom.” CityLab, www.citylab.com/city-makers-connections/bike-share/.
2. Motivate International, Inc. “Citi Bike System Data.” *Citi Bike NYC*, www.citibikenyc.com/system-data.
3. Zhang, Yongping, and Zhifu Mi. “Environmental Benefits of Bike Sharing: A Big Data-Based Analysis.” Applied Energy, vol. 220, 2018, pp. 296–301., doi:10.1016/j.apenergy.2018.03.101.
4. Otero, I., et al. “Health Impacts of Bike Sharing Systems in Europe.” *Environment International*, vol. 115, 2018, pp. 387–394., doi:10.1016/j.envint.2018.04.014.
5. Motivate International, Inc. “Citi Bike System Data.” *Citi Bike NYC*, www.citibikenyc.com/system-data.
6. Citibike Image: “Lined up Rental Citibank Bikes in Lower Manhattan. New York , NY , USA -.” *Shutterstock.com*, 2 Oct. 2018, www.shutterstock.com/image-photo/lined-rental-citibank-bikes-lower-manhattan-1192959820.