Introduction

In the last 10 years, ridehailing companies such as Uber, Lyft, and Didi have revolutionized the world of private passenger transportation services with their smartphone-based platforms. As ridehailing services take hold, they are dramatically altering patterns of human mobility particularly in urban areas in the United States. The easy accessibility of these services and the relative affordability of these services provide ridehailing users with higher levels of mobility than even that afforded by automobile ownership in some respects. Riders may have to wait minutes for service, but they are relived of looking for parking near their destination, a task that is considerably difficult or costly in urban areas where parking is often highly limited.

Indeed, ridehailing services are a transformational force in America's cities, however, they are not without problems. One area of grave concern is that these services are exacerbating traffic congestion as they become more popular. Numerous studies (XXX) have found that these services are not only increasing travel demand (i.e., generating more trips) but they are also competing with more efficient modes of transport, namely public mass transit. Consequently, understanding ridehailing ridership trends is not only an internal concern of ridehailing companies but also of interest to transportation planning agencies.

Public studies about ridehailing services are notoriously rare because ridehailing companies almost never share their data, citing, among other reason, economic and customer privacy concerns. One exception, and the focus of this study, is a trip dataset released by Ride Austin, a non-profit ridehailing company based in Austin. This data set covers the 1.5 million+ trips provided by the company between June 2016 and April 2017 (XXX). It contains several trip characteristics including the time and location of the passenger pick-up and drop-off, the trip duration and distance, vehicle and driver information, and the passenger trip ratings (for the driver).

[Results]

2 Preliminaries

Before starting our analyses, we (1) first inspected and cleaned our data and (2) aggregated trips into uniform time intervals. Beginning with the former, to detect outlier trips (e.g., trips with questionable trip information) we inspected the data set with *kepler.gl*, an open-source geospatial analytic visualization tool developed by Uber (XXX). We also limited our dataset to trips that started and ended in Austin and removed trips with travel distances larger than 100,000 *m*. Consequently, 1,493,671 trips remained after removing 7536 trips.

After cleaning this dataset, we aggregated trips on hourly intervals beginning and ending with each hour. This level of aggregation is used frequently in the transportation modeling literature (e.g., XXX), however, we note that the significance of different time intervals has not yet been thoroughly examined in the transportation literature.

3 Data Exploratory Analysis

3.1 Preliminary Data Exploration

We plotted the resulting time series in **Figure *1*** (below). The time series sample mean was reported to be $\mu = 198.205$ while the sample variance was reported at $\sigma^2 = 56407.620$.

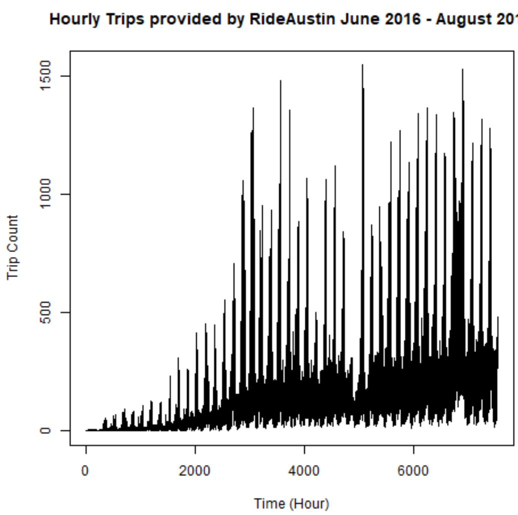


Figure 1

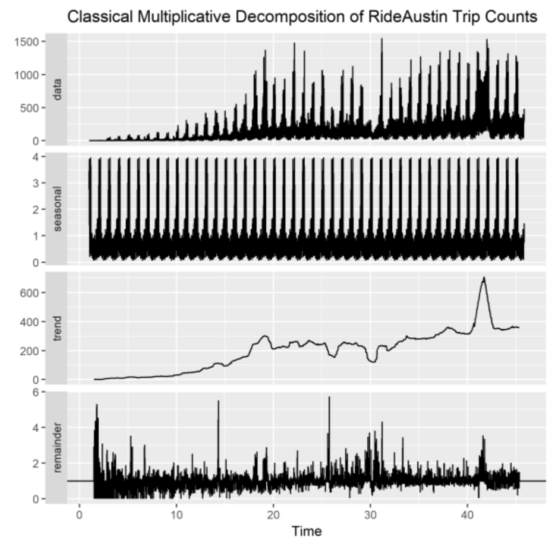
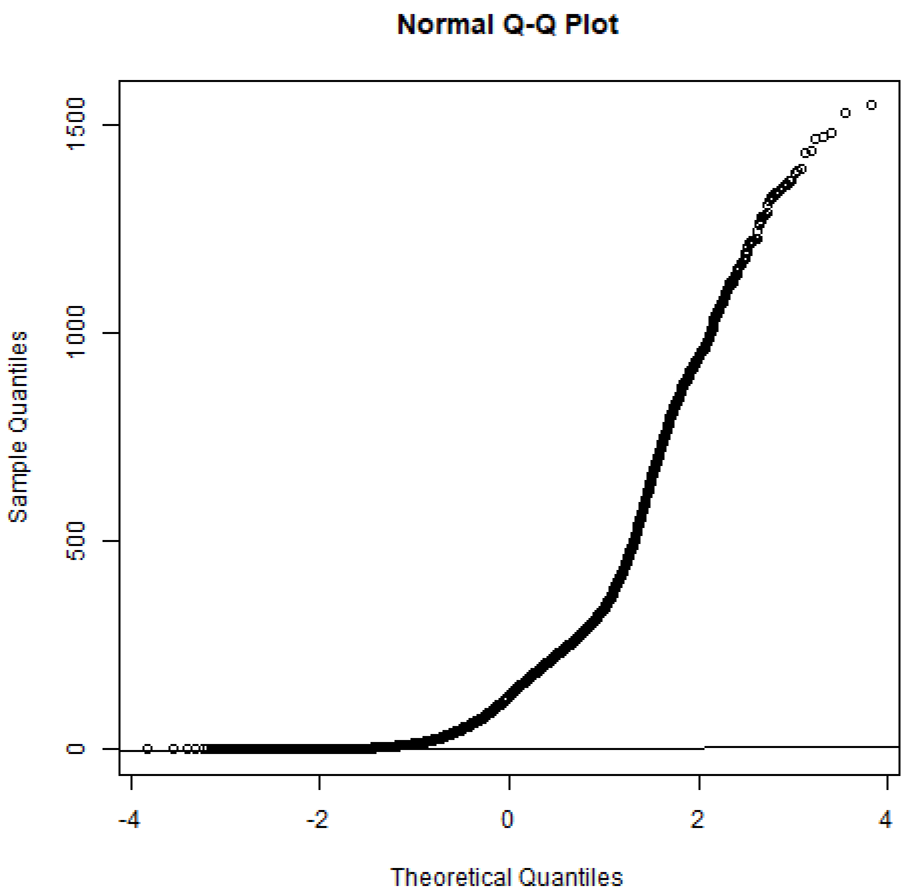


Figure 2

In terms of trends, we observe that there is a general positive trend in the number of trips both on average and in terms of the variability. The most pronounced increase is observed between hours 2000 and 4000 after which the series appears drop between hours 4000 and 6000 before mostly stabilizing soon thereafter.

To show these changes more clearly, we decomposed the time series into a multiplicative model depicted in **Figure 1** **Figure *2*** (above). In this model, $X_t = m_t * s_t * V_t$ where X_t is the trip count at hour t, m_t is the trend component (at hour t), s_t is the seasonal component, and V_t is a stationary process. This model proved to be appropriate than an additive model because of the strong positive relationship between the trend and

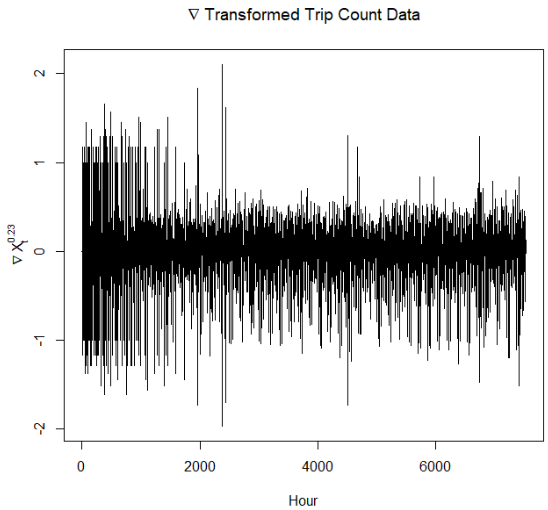
composed of a trend, seasonal, and stationary

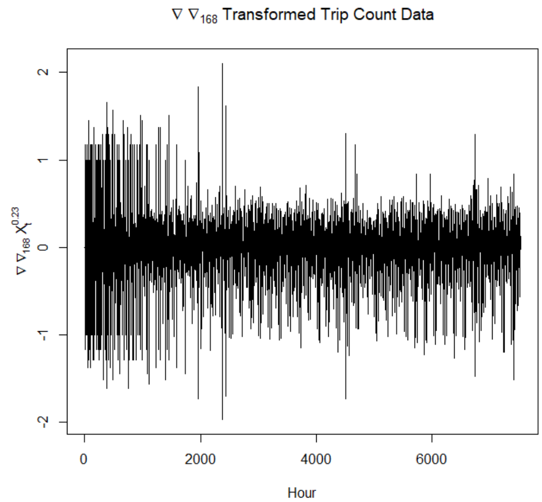


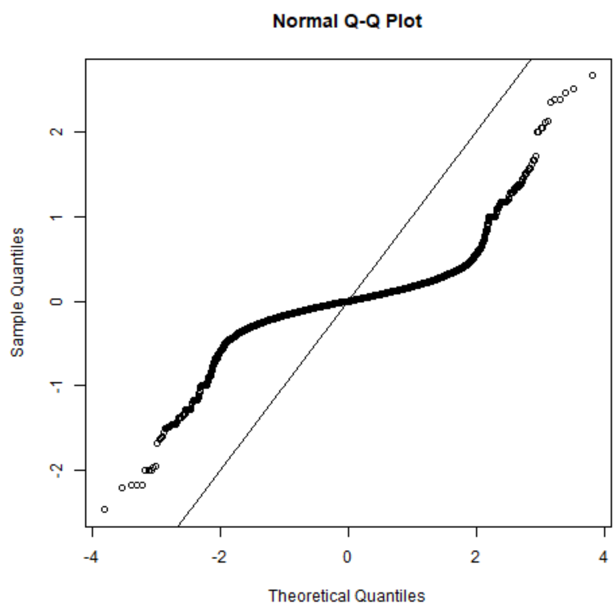
4 Data Transformations

```
yjPower Transformation to Normality
   Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
Y1    0.2314        0.23       0.2181       0.2447

 Likelihood ratio test that transformation parameter is equal to 0
                           LRT df       pval
LR test, lambda = (0) 1243.934  1 < 2.22e-16
```

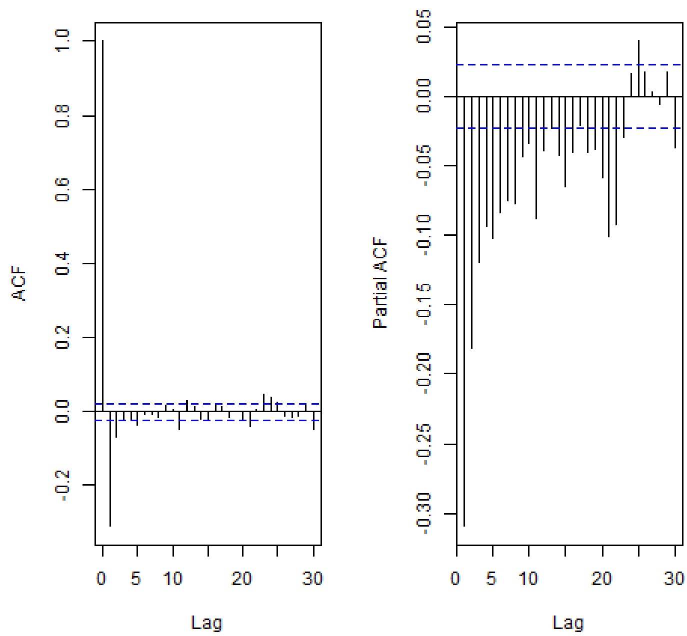






5 Model Identification and Estimation

Series ACF and PACF (lag = 30)



Series ACF and PACF (lag = 175)