

# Analysis of Ford GoBike's Open Data

36315: Statistical Graphics and Visualization

Carnegie Mellon University



Ross Hemmel, Joshua Huang, Julie Kim, Jinhee Lee

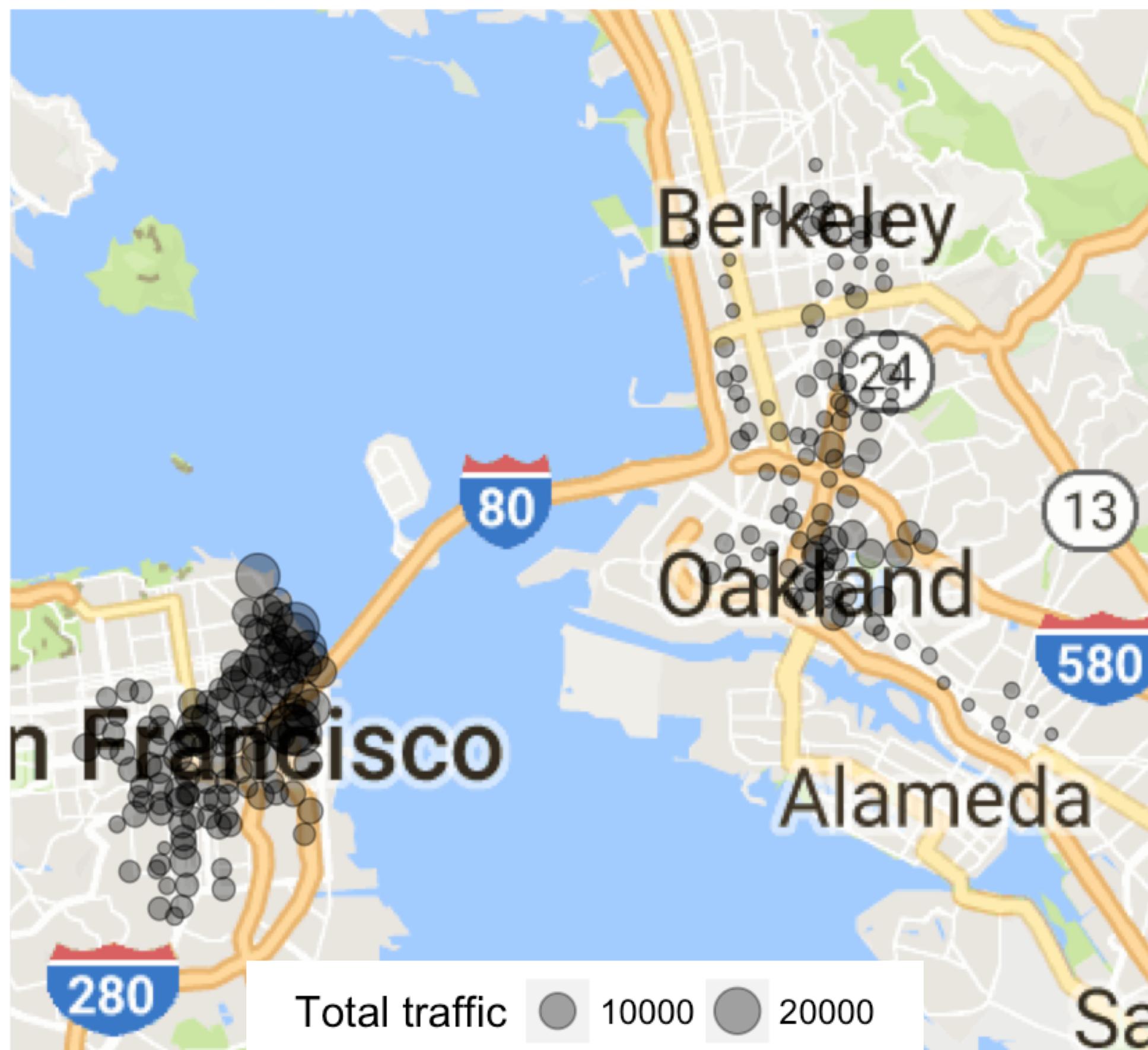
## BACKGROUND

Ford's GoBike program provides an affordable bike-rental system in the San Francisco Bay Area. Users can either pay one-time or membership fees in order to borrow bikes from any station, and return them to any station.

With the open-source data that Ford GoBike offers on their website (<https://s3.amazonaws.com/fordgobike-data/2017-fordgobike-tripdata.csv>), we are looking to use statistical visualization in order to gain insight on customer demographics and biking behaviors. Our map base is pulled from Google.

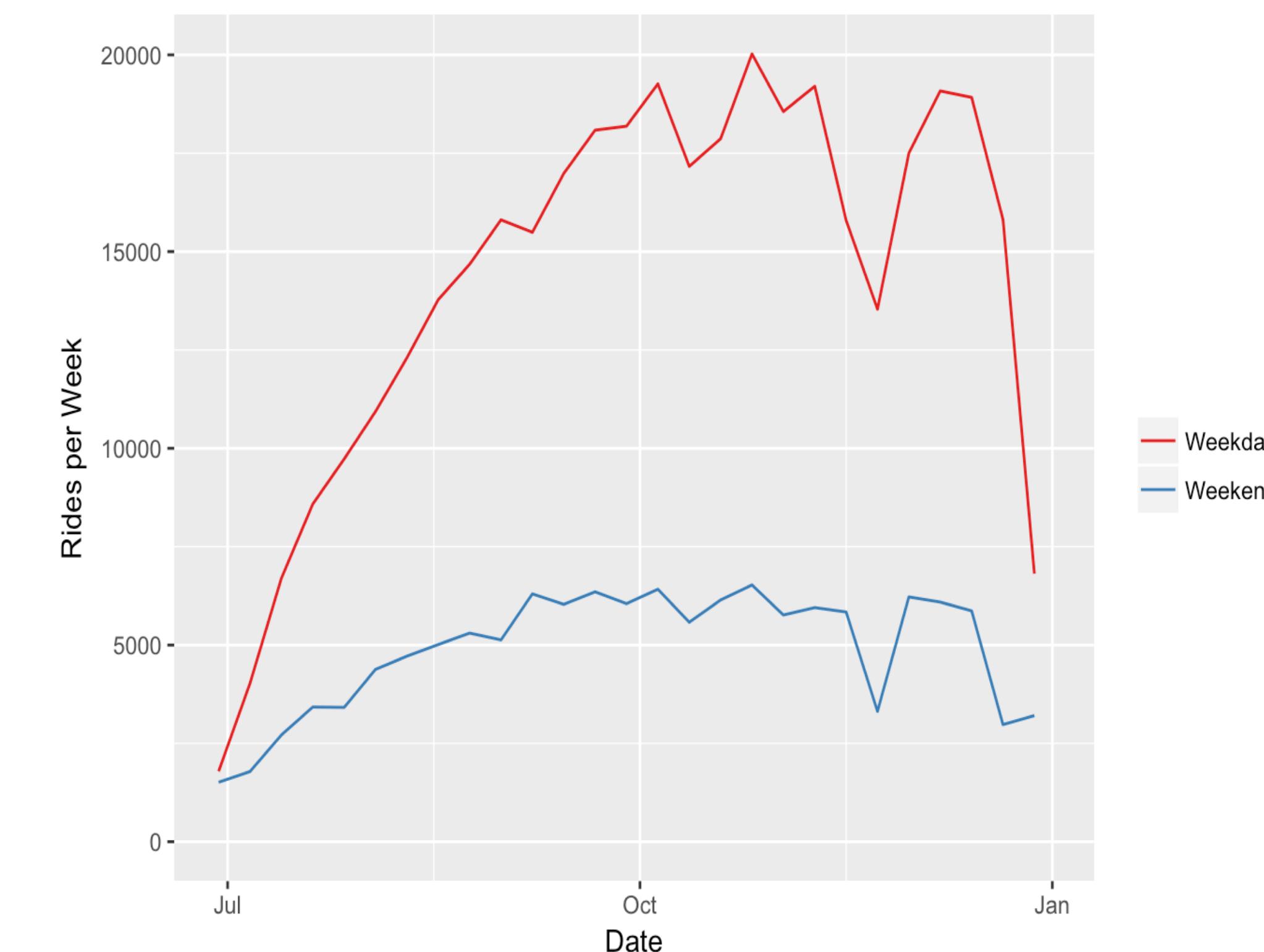
The data involved was collected starting in June of 2017, and includes variables such as bike station geo-coordinates, trip start and end times, member types, and member characteristics. We were able to manipulate quantitative variables such as member birth year to categorical factors by grouping ages together, as you can see in Exhibits 2 and 7. Distances in Exhibit 2 were calculated by inputting longitudinal and latitudinal coordinates into the Haversine formula.

## Exhibit 1: GoBike Stations in the Bay Area



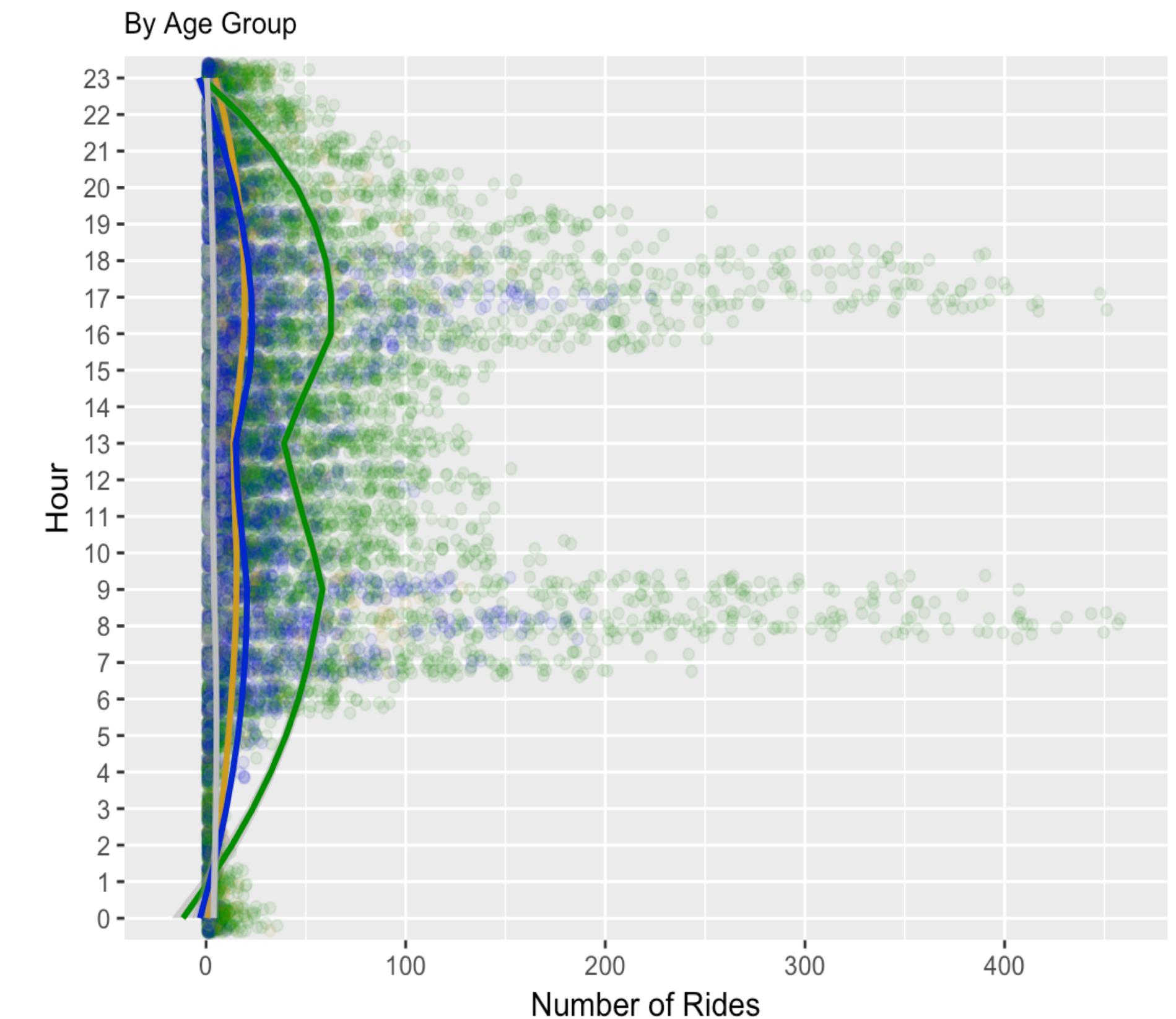
## TIME SERIES

### Exhibit 5: Distribution of Bike Usage over Time

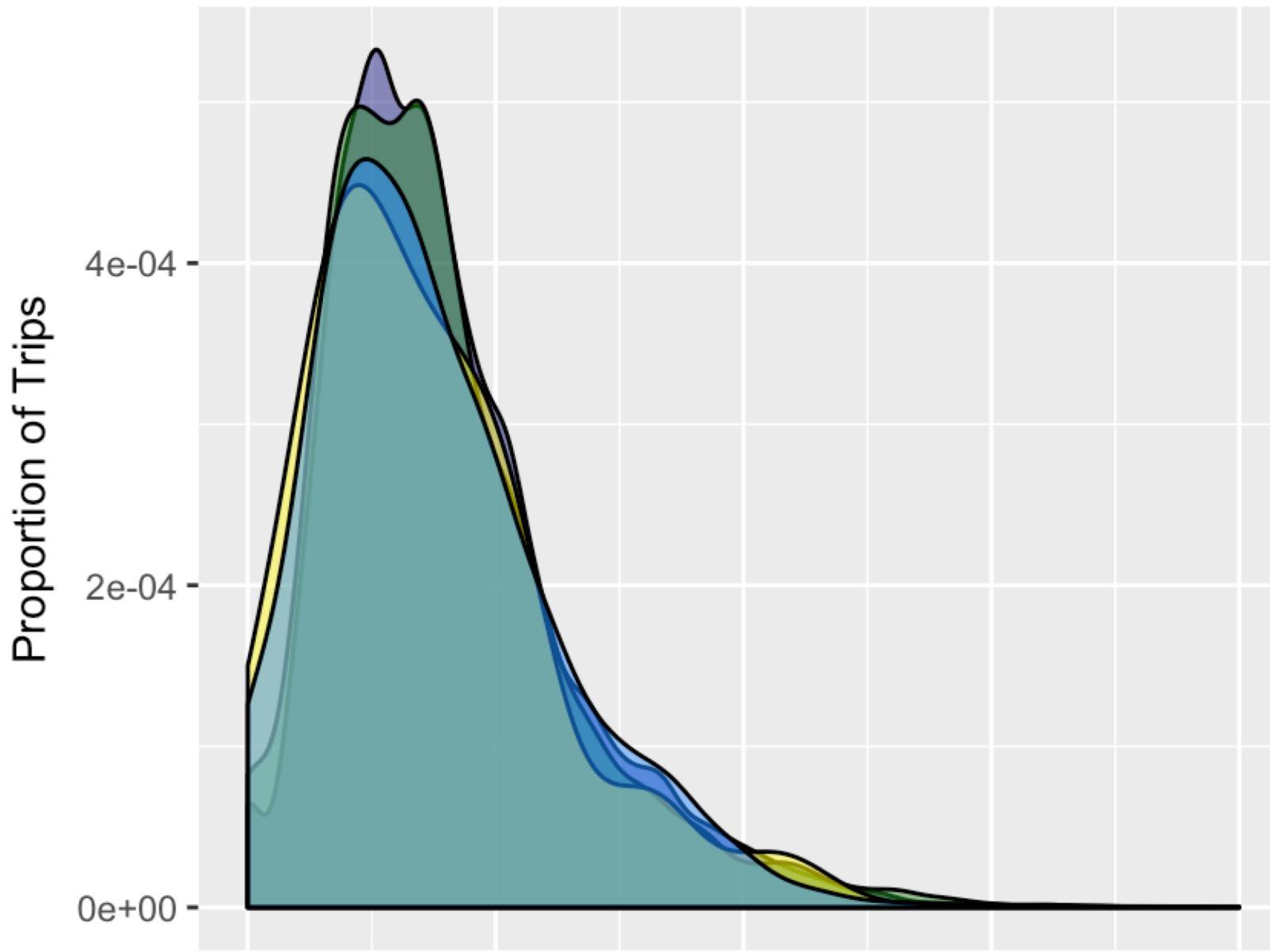


## HOURLY DISTRIBUTIONS

### Exhibit 7: Number of Rides Each Hour of the Day



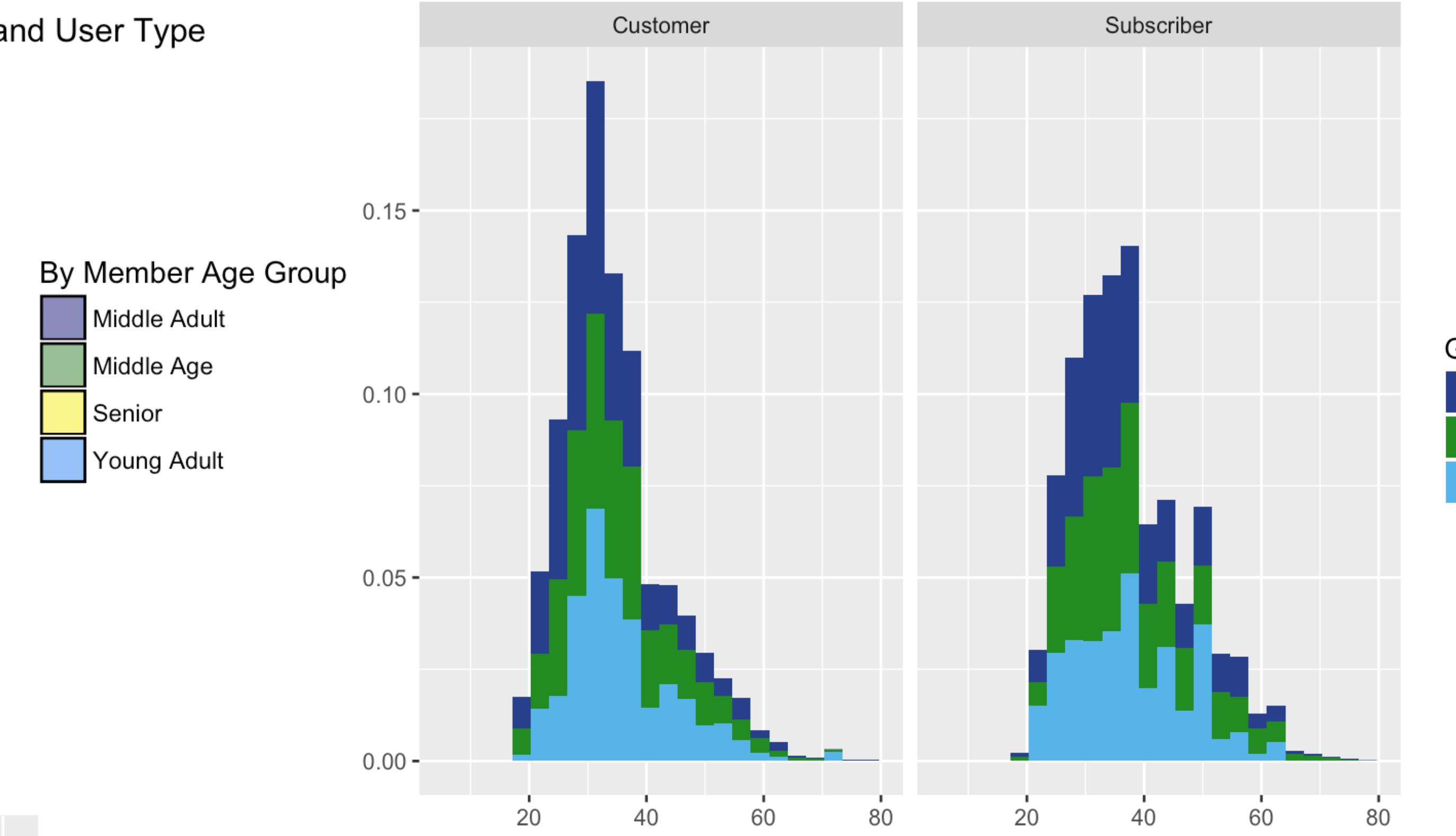
## Exhibit 2: Trip Distance by Age Group, Gender, and User Type



By Member Age Group

- Middle Adult
- Middle Age
- Senior
- Young Adult

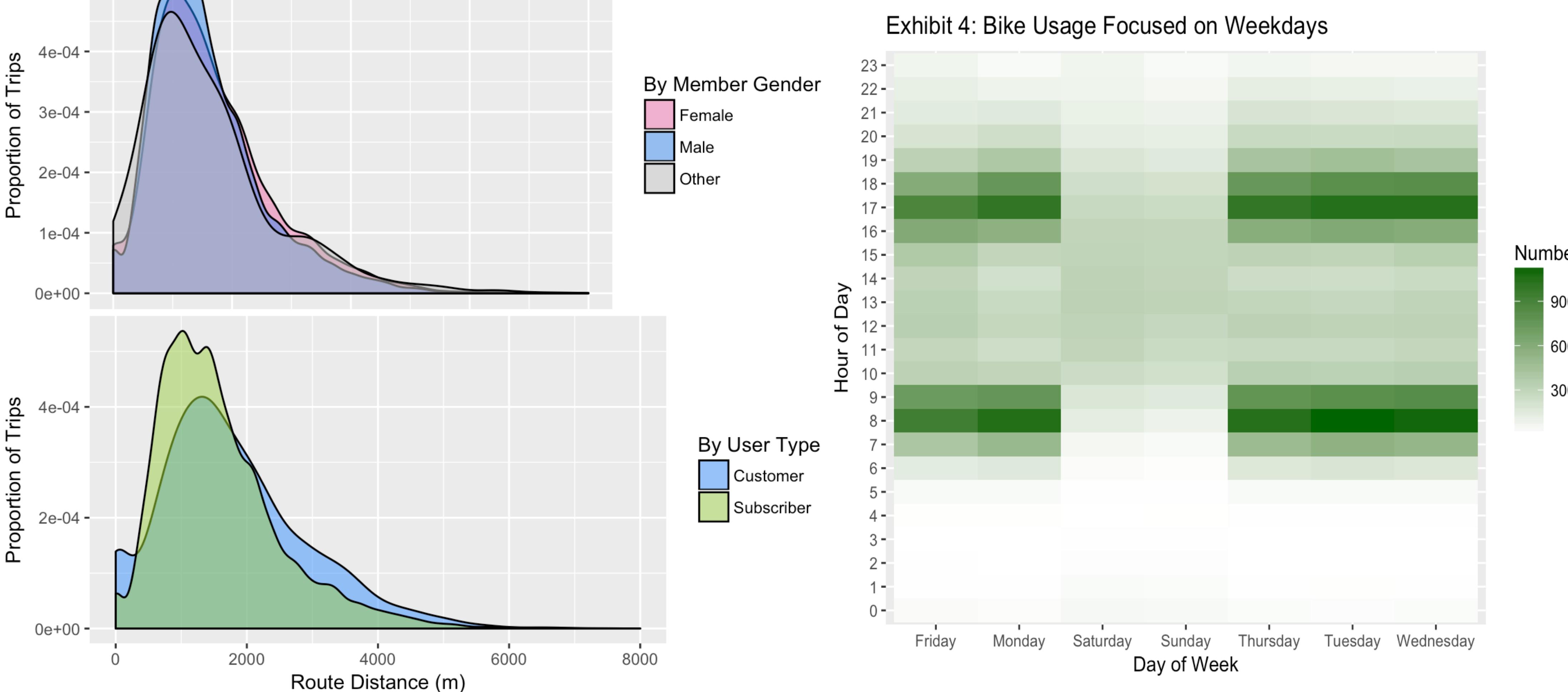
## Exhibit 3: Age Distributions by Subscriber Type



Gender

- Female
- Male
- Other

## Exhibit 4: Bike Usage Focused on Weekdays



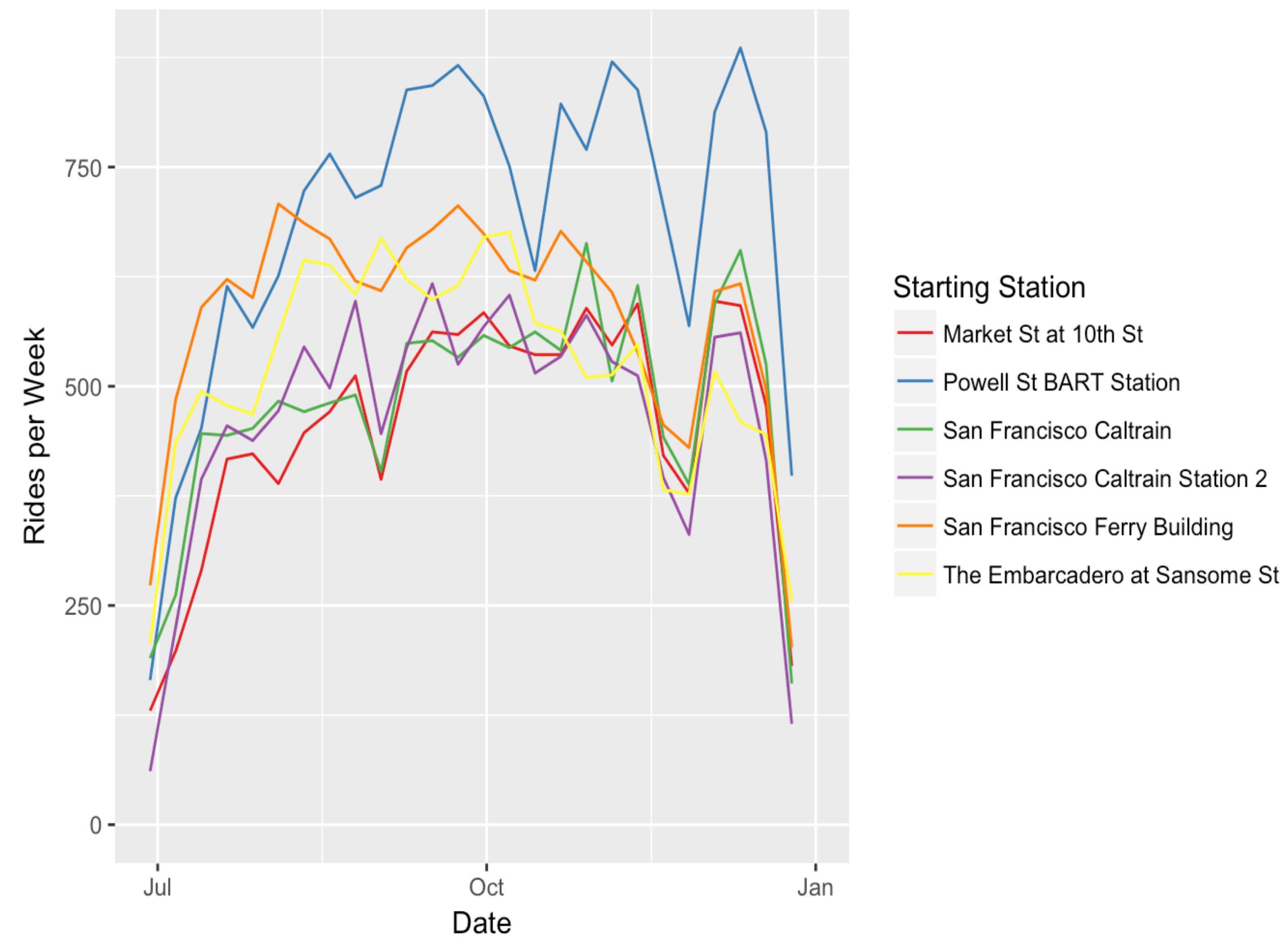
By Member Gender

- Female
- Male
- Other

By User Type

- Customer
- Subscriber

## Exhibit 6: Five Most Popular Stations over Time



## FINDINGS

**Exhibit 2:** Distance travelled is not dependent on the age or gender of members. However, a higher volume of subscribing members ride shorter distances.

**Exhibit 4:** By examining bike usage data by each hour of each day, we found that users most frequently use the service during weekday rush hours of 8 a.m and 5 p.m.

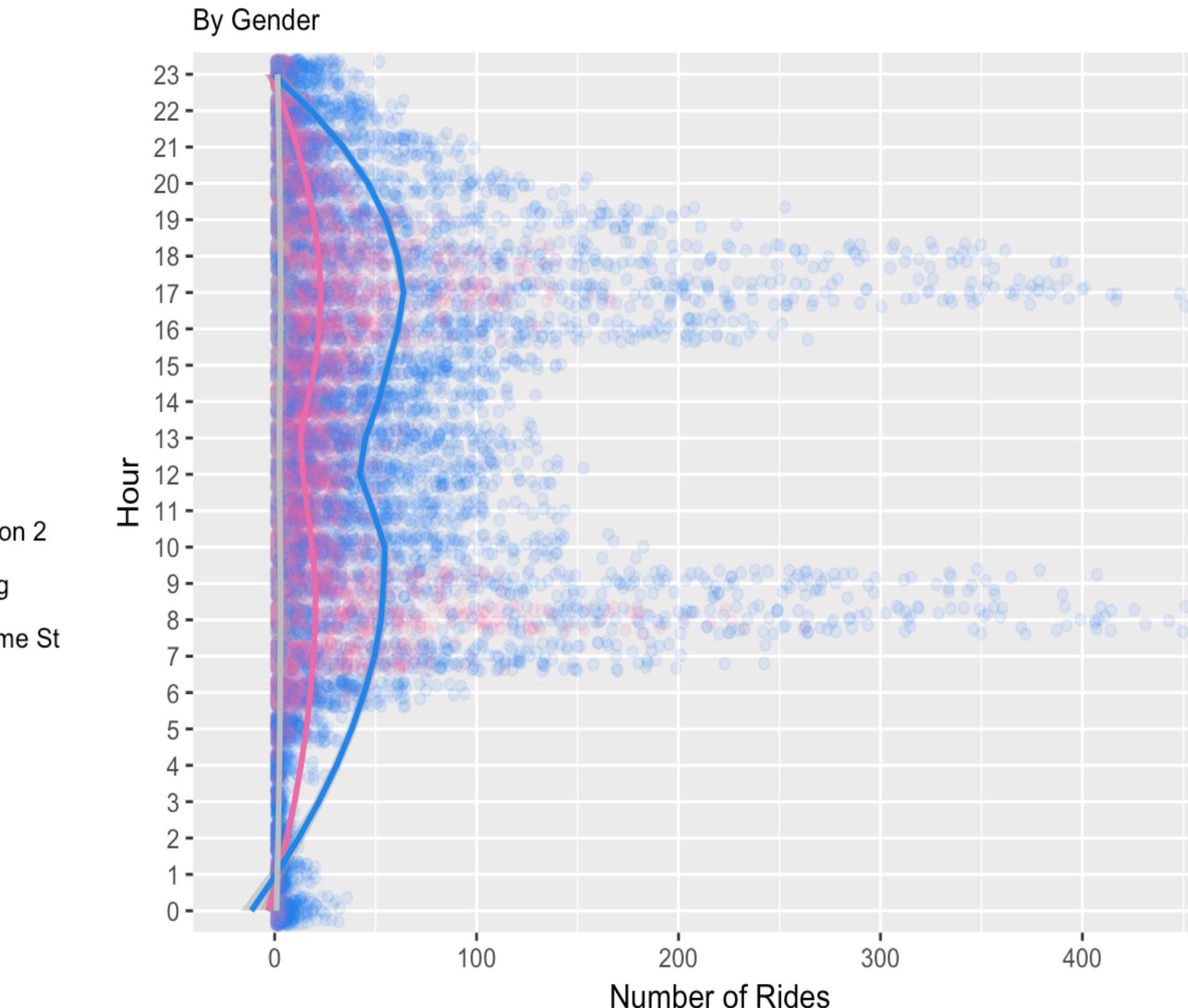
**Exhibit 5:** Weekdays saw a large increase in users over the first few months while weekend riders remained steady.

**Exhibit 6:** A group of 3 stations appear to have very similar trends in user numbers while the other two follow different, sometimes opposite trends.

**Exhibit 7:** The time of trip does not appear to depend on age groups. This conclusion may be uncertain for the Senior (ages 65+) group, due to a much smaller sample size.

**Exhibit 8:** The time of trip does not depend on gender.

## Exhibit 8: Number of Rides Each Hour of the Day



Age Group

- Young Adult
- Middle Adult
- Middle Age
- Senior

Gender

- Female
- Male
- Other