Bike Sharing in Seattle

Zhiwei Zhang, Sam Beadles, Mier Chen, Katherine Qian

Agenda

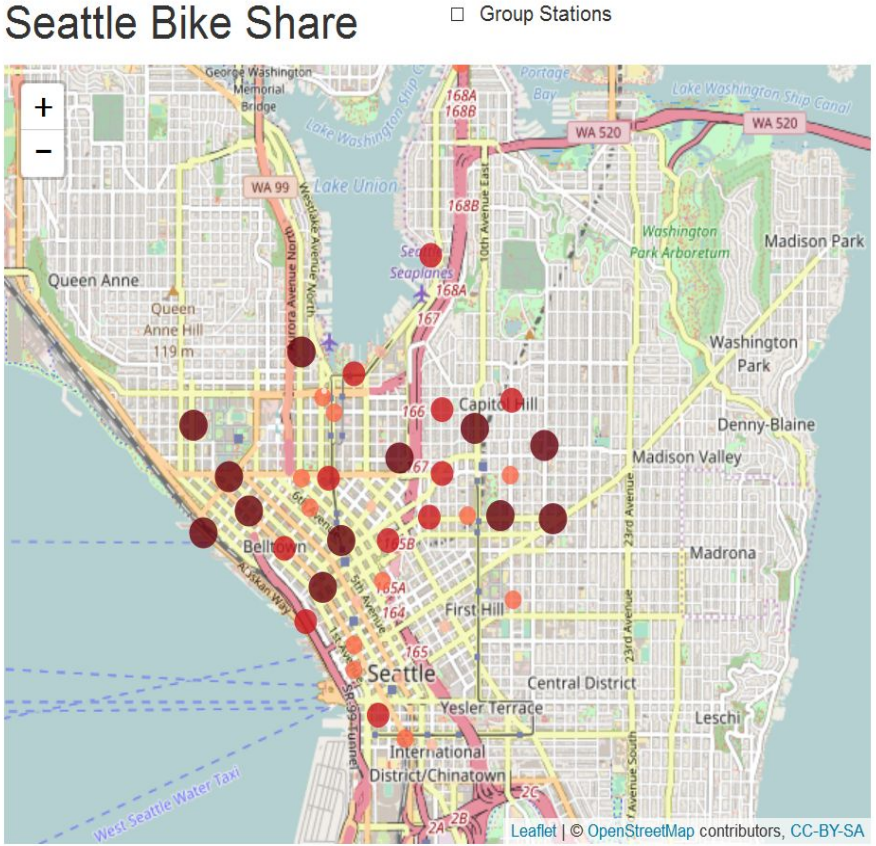
- Introduction, guiding questions
- Description of data
- Data Cleaning
- Data Analysis
- Findings

Description of Data

Our data is based on a bike sharing program in Seattle, WA

- Pronto Bike Share Program:
  - Based in Seattle
  - Annual member vs short-term pass holders
  - 30/45 minute fee structure

58 stations
500 bikes



To do our analysis, we worked from 3 data sets

- station.csv: [“station_id” , “lat” , “long” , “name”]
- trip.csv: [“starttime”,”stoptime”, “tripduration”, “from_station_id”, “to_station_id” , “usertype”, “birthyear”]
- weather.csv: [“Max_Temp”, “Min_Temp” , “Max_Gust_Wind_Speed” , “Events”]

Glimpse of the data

In those 688 days,

There are 236,065 trips,

4,731,578 hours of riding,

With an average of 20 minutes per trip,

Done by 61.9% annual members

Data Cleaning

We had to clean inside the data sets to make them usable

1. Check the accuracy of data: Duplicates in trip_id
2. Datetime format: “10/13/14 10:48” → “2014-10-13 10:48:00”
3. Create columns: “start_year”, “start_weekday” ,”age”,
4. Inconsistency in weather.Events: “Rain, Snow” → “Rain-Snow”
5. Buckets for temperatures
6. Station names:
   Burke Museum / E Stevens Way NE & Memorial Way NE →
   E Harrison St & Broadway Ave E

Left merge on ‘starttime’ and “date”

trip_df + weather_df => trip_weather_df

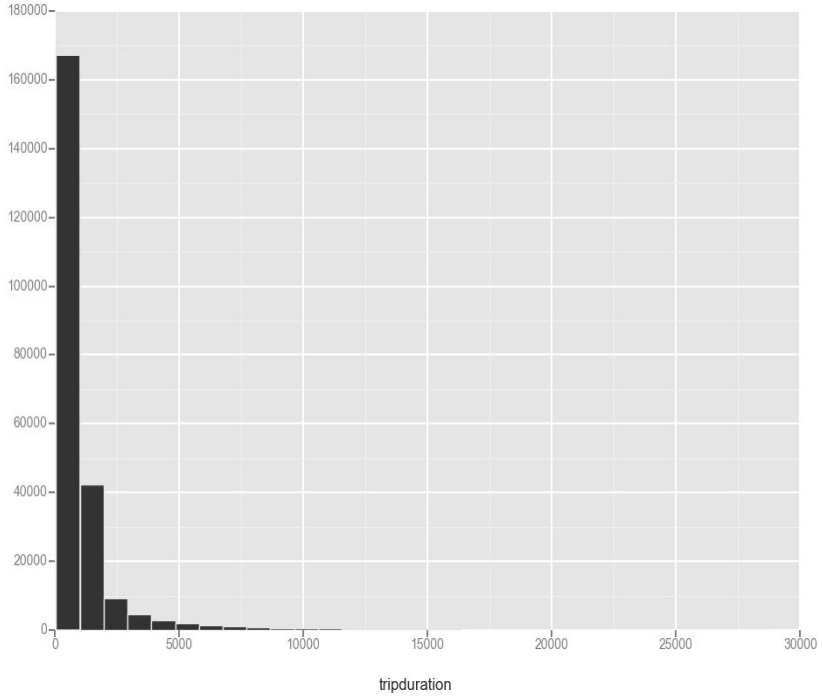
trip_weather_df + from_station_df
+ to_station_df
=> cycle_df

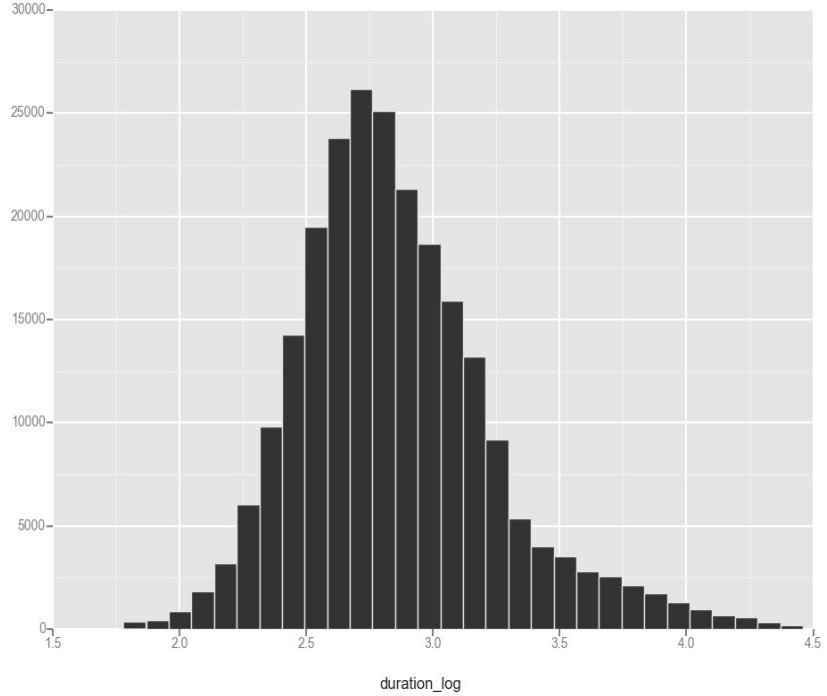
Left merge on
“from_station” and “to_station”

Cycle_df

227954 rows and 65 columns

We eliminated outliers in the trip duration set



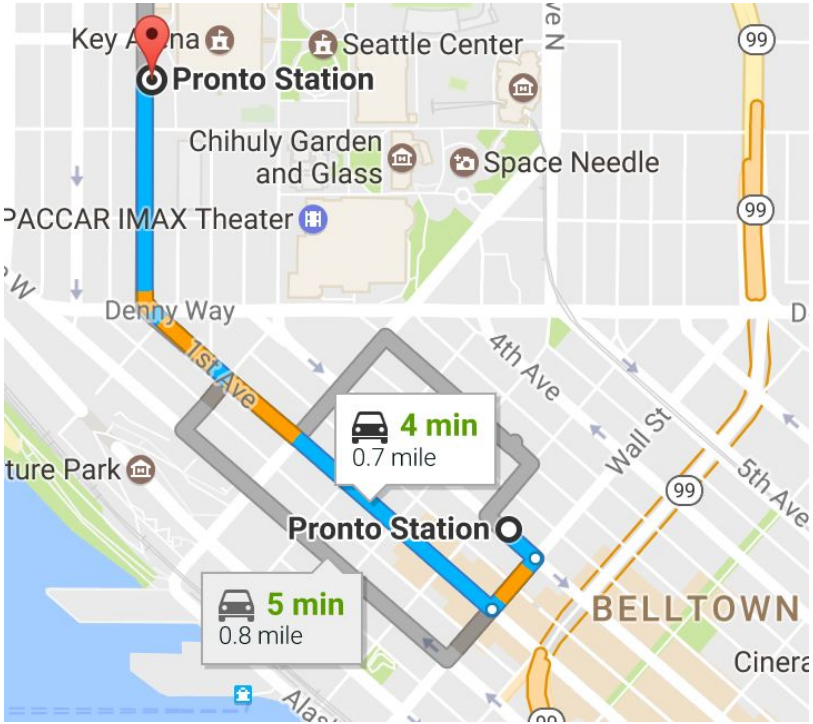


We developed two methods to calculate distance

Google API:

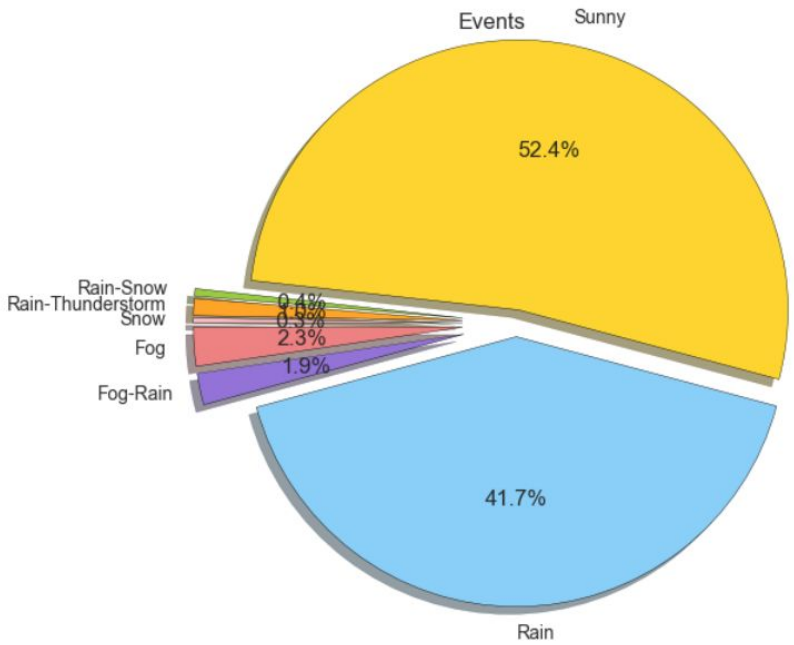

```
{
   "destination_addresses" : [ "New York, NY, USA" ],
   "origin_addresses" : [ "Washington, DC, USA" ],
   "rows" : [
      {
         "elements" : [
            {
               "distance" : {
                  "text" : "267 mi",
                  "value" : 429076
               },
               "duration" : {
                  "text" : "1 day 0 hours",
                  "value" : 86868
               },
               "status" : "OK"
            }
         ]
      }
   ],
   "status" : "OK"
}
```

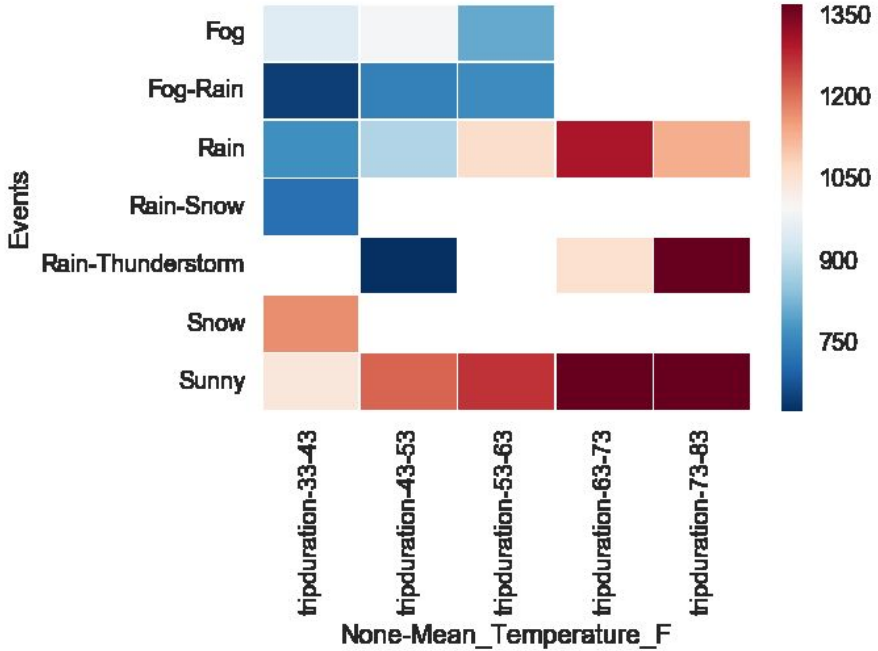

```
from geopy.distance import great_circle
```



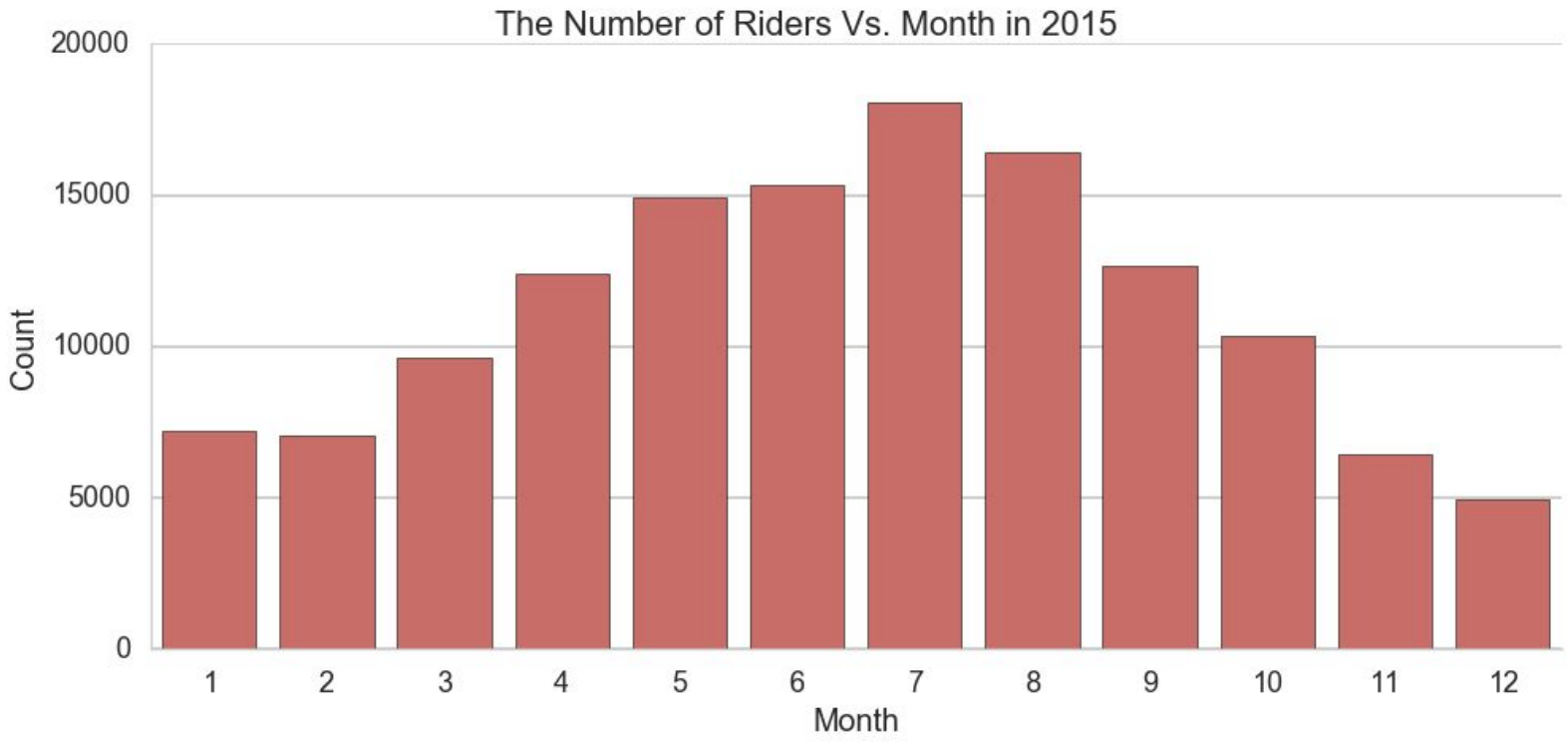
Data Analysis

Weather events are biased toward rain, sun

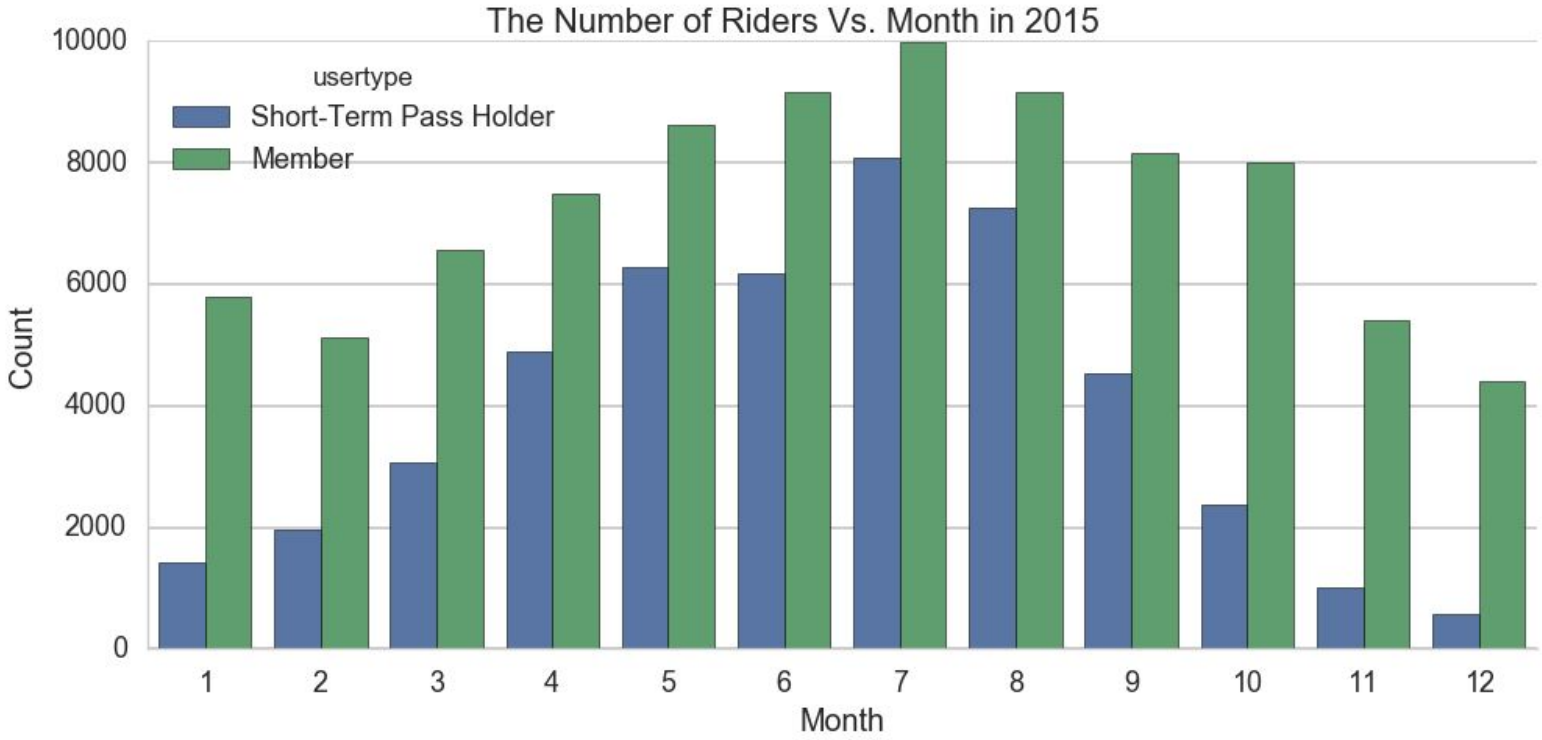




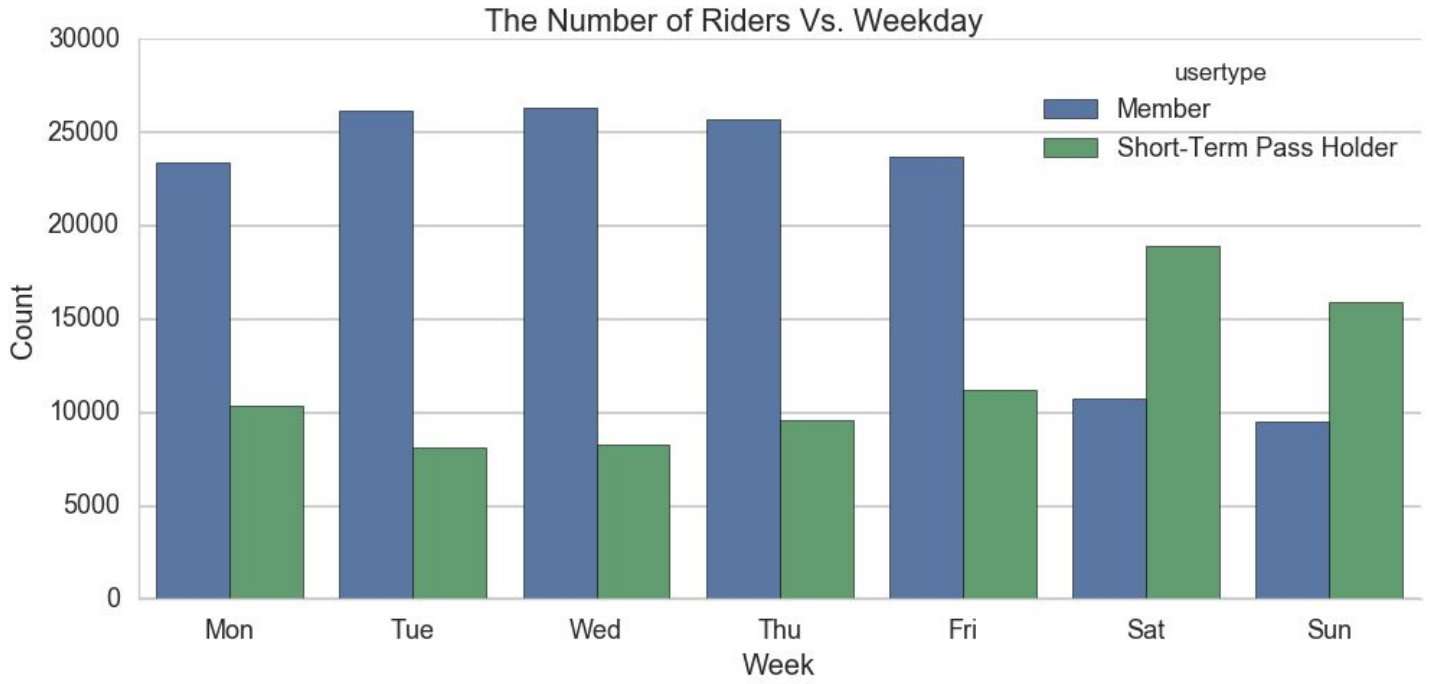
Ridership peaks in summer months of June, July



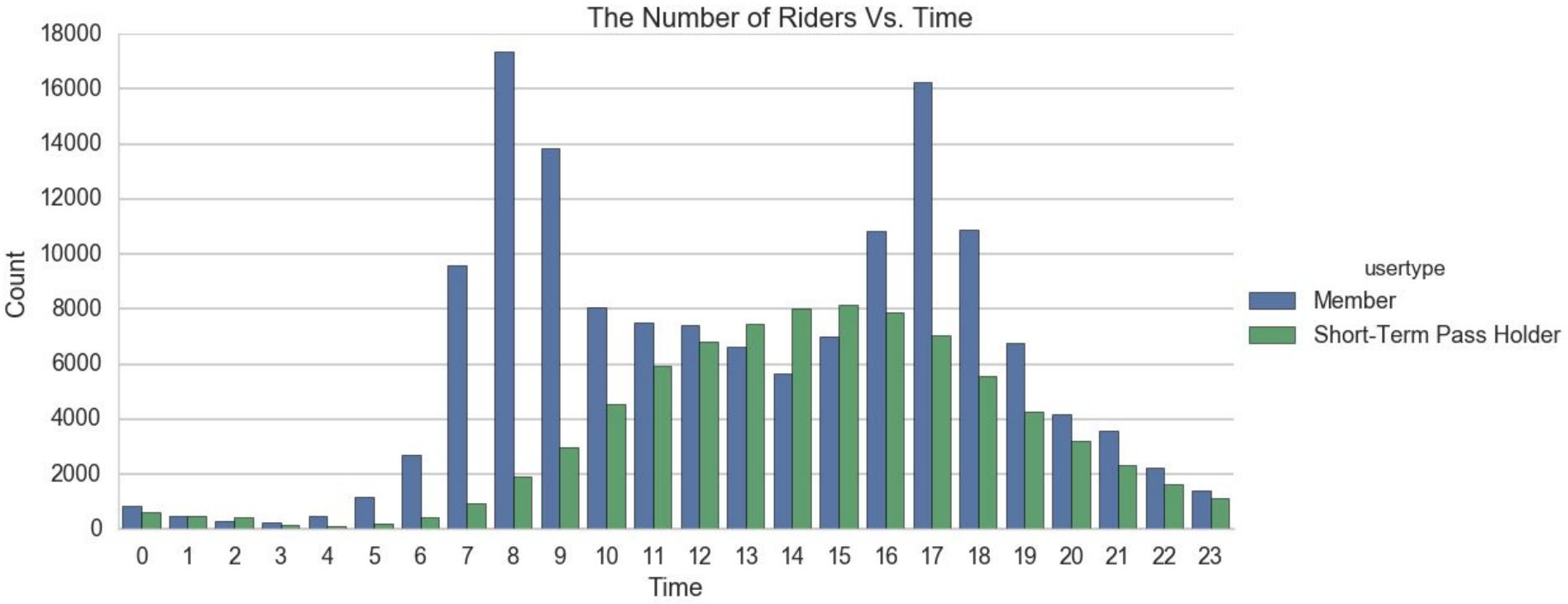
Pass holders have a much more pronounced preference for summer months



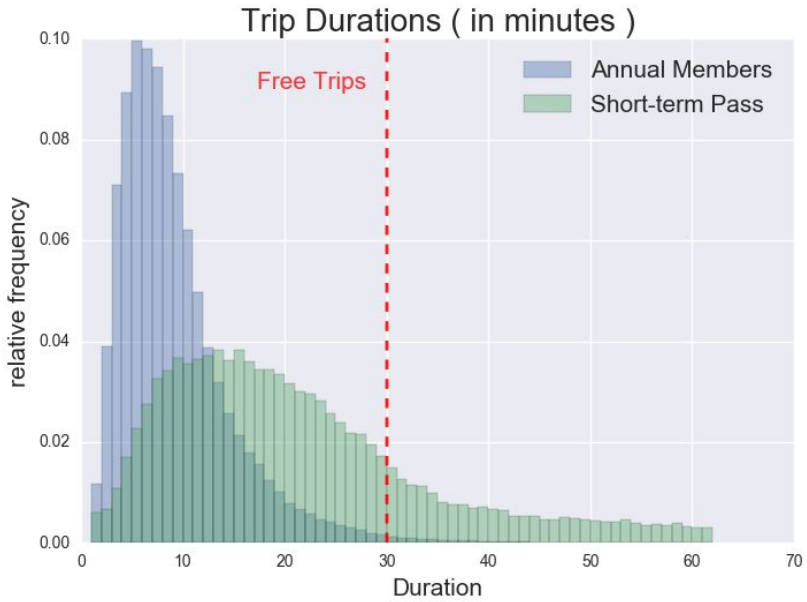
Rider demographics depend on day of the week

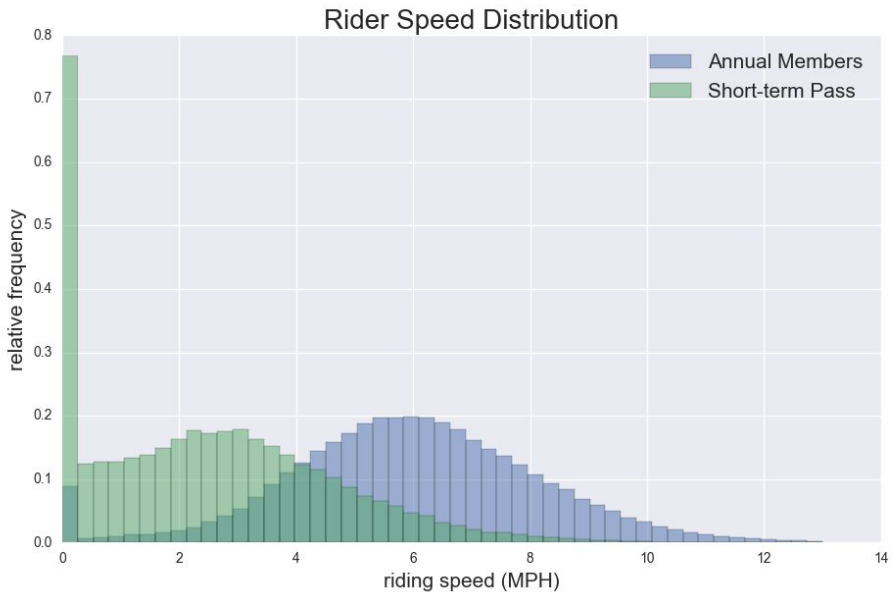


Rider demographics depend on time

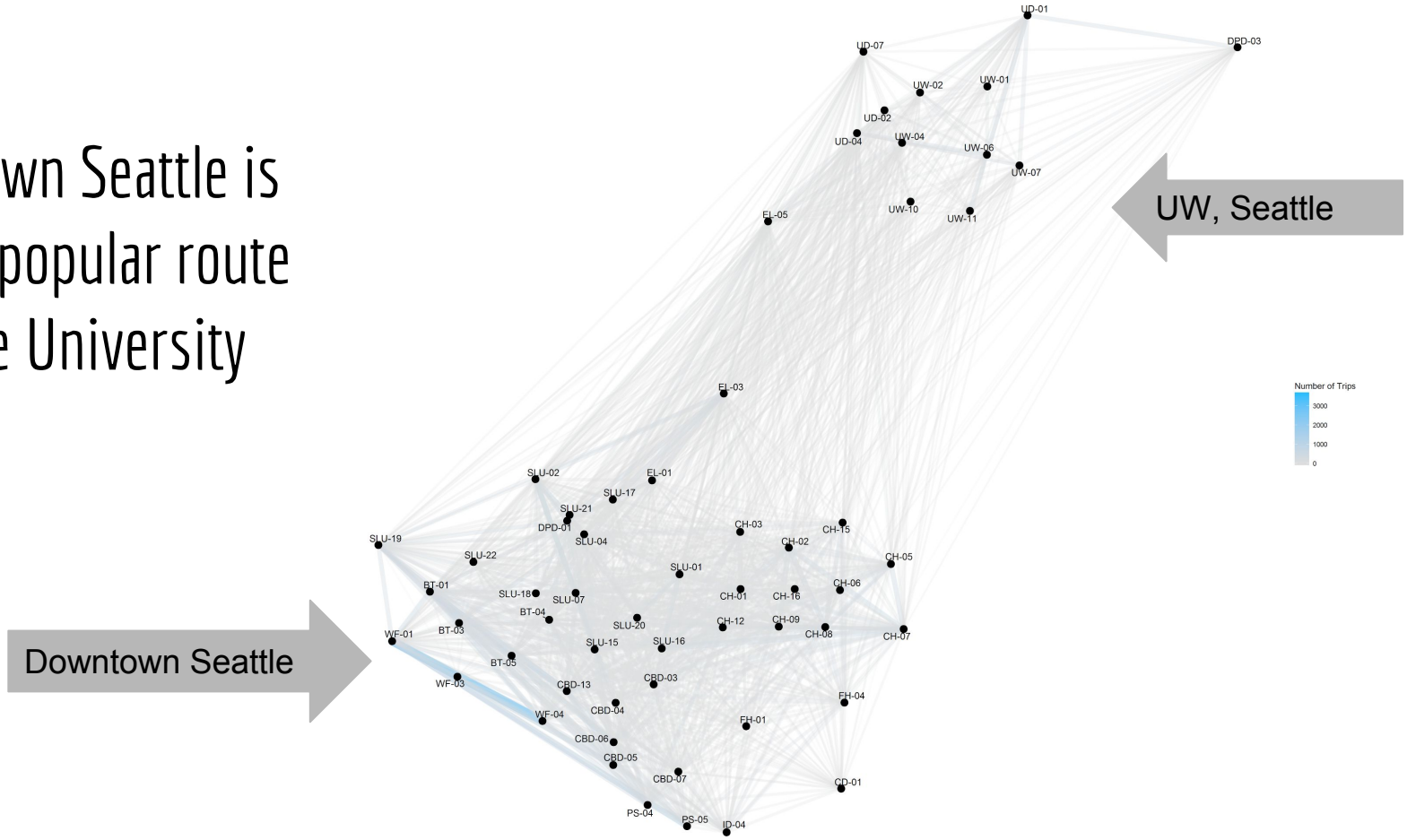


Pass holders are chronic tour-ers

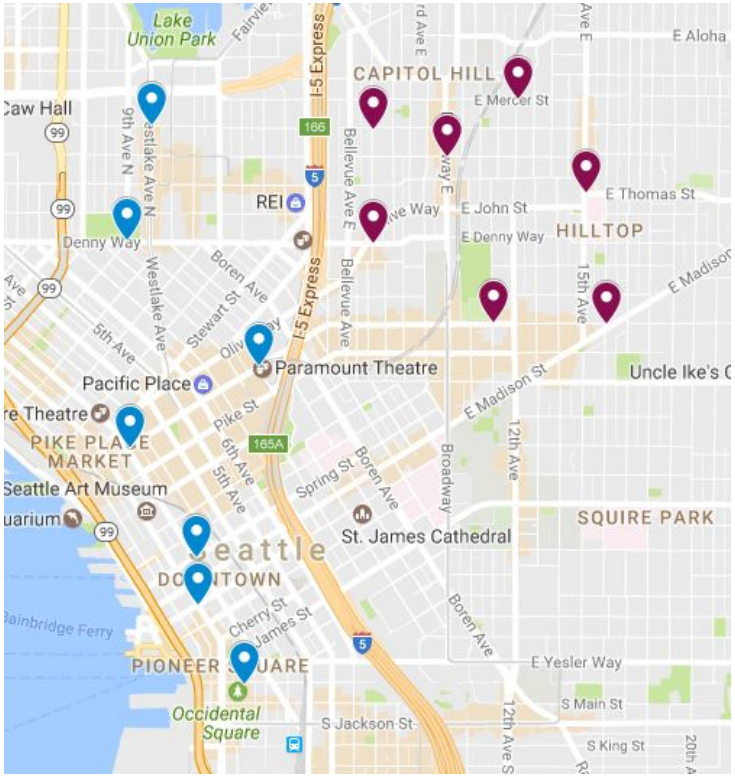


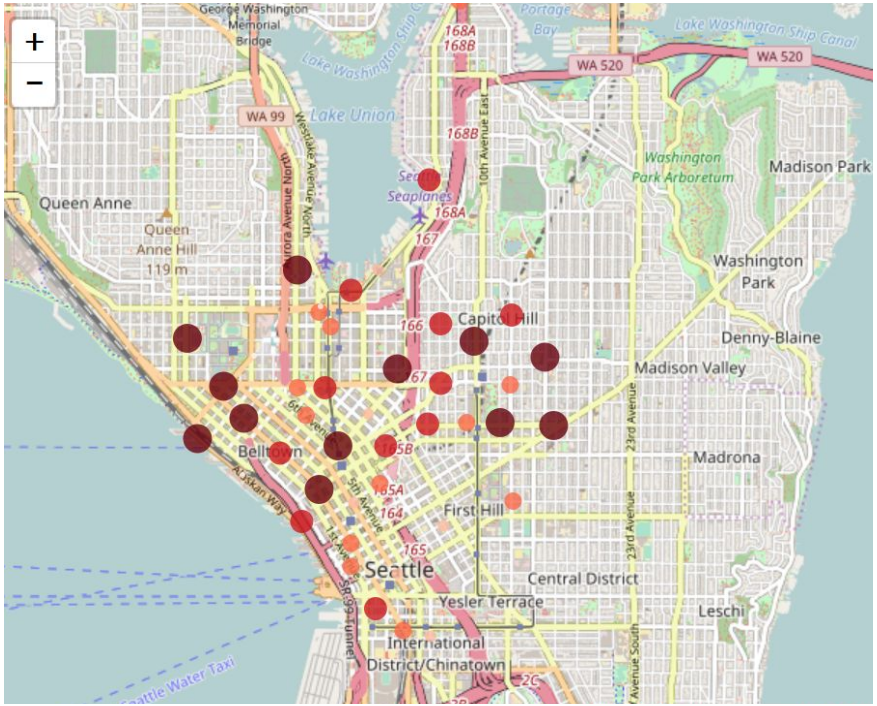


Downtown Seattle is a more popular route than the University



Bikes are unevenly picked up and dropped off across Seattle





Pronto bikes may be able to capitalize on biker demographics

- Charging fees before 30 minutes
  - Low elasticity? Competitors?
- Concentrate upkeep, infrastructure on summer months
- Change pricing structures for pass holders on weekends
- Encourage more even distribution of bikes across Seattle: plan more regular bike transfers from highly-used to less-used.
- Target Washington University students

Questions?