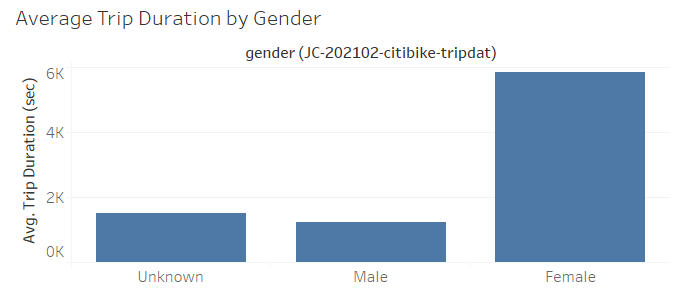
Analysis of Jersey City Citibike Data

Matt Krenicki

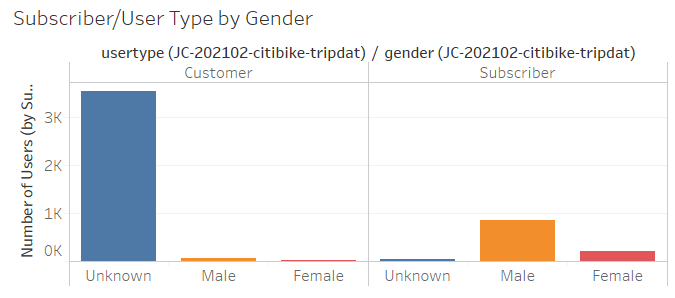
April 8, 2021

**Phenomena Observed #1: Study of relationships between Gender relationship to Trip Duration and User Type**

* The image below shows the relationship between gender and average trip duration
* We see that females appear to take much longer trips on average than males or gender unknown (almost 4X), which is a surprisingly large difference

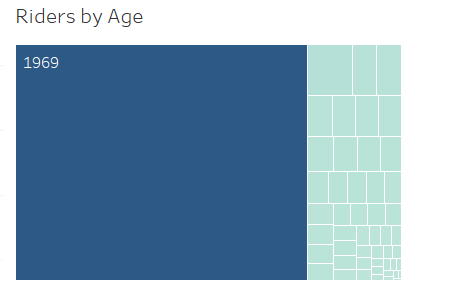


* To look at this further, I looked at how the gender makeup varied across user types
* We see below, that for “Customers”, those without a subscription, nearly all are of gender Unknown
  + This may be happening because this could be a default value when doing a daily rental (i.e. classified as Customer) unless selected out of by the user
* We also see that among “Subscribers” that very few are of gender Unknown. Most have selected Male or Female
* We also see that Females are the smallest group, yet as shown previously their average trip durations are longest
  + So, in summary, fewer females relative to males use the service, but when they use it they tend to take substantially longer trips



**Phenomena Observed #2: Study of relationships between Gender relationship to Trip Duration and User Type**

* Below is a tile visual showing the proportion of riders by age
* We see that nearly 75% of riders were born in 1969
* This seems very strange, so I also looked at how trip duration varied by age to see if we could draw any insights



* Below is a graph depicting the average trip duration vs. age
* We see that for two years, 1964 and 1986, the trip durations are on average much longer than the rest of birth years
* This does not give us any clear insights into why 1969 is so common a birth year for riders, but a possible reason why the average trip durations are so much higher for these birth years is that perhaps they are groups of friends who rode a longer trip together, and thus the average gets inflated. Both birth years had a relatively small number of rides, so the average can easily be inflated by a small group taking a large trip
* Lastly, as far as riders vs. age and the fact that 1969 is so common, it could be that this is the default value for riders if they do not enter their own birth date, similar to what we suspected with regards to gender

