Citibike Analysis - PUI_HW7

Alex Shannon[1]

[1]NYU Center for Urban Science & Progress

November 10, 2017

Abstract – We analyzed Citibike ridership by gender, investigating whether a greater percentage of males relative to females use the bikes when it's dark outside. It has been well documented that men use the service more than women, but we particularly wanted to see if this ratio was exacerbated during non-daylight hours. Using data from June 2016, we organized and plotted data, and ran a unpaired t-test to compare the results. Ultimately we are able to conclude that the relative proportion of male riders does increase after dark, at least with regards for the timeframe we observed.

Introduction – The investigation of ridership by gender before and after normal daylight hours is interesting in that it touches on the culturally embedded notion that it's more dangerous for women to be alone after dark than men. This may be true for numerous reasons, which we will not delve further into at present. However, it is possible that Citibike provides a safer alternative for nighttime travel, as riders more faster and have less human-interaction than walkers. In this instance, there do seem to be reasons that women would be *more likely* to ride after dark than men. Before looking at the data, then, it does seem like there are multiple possible outcomes, all of which tell us a little bit about how people perceive and use Citibike for transportation.

Data – We are using data provided by Citibike's website, covering the month of June 2016. From this data, we transformed start and stop times for individual rides, found their means, and used these means to approximate whether a ride took place during 'light' or 'dark' hours. We defined 'light' as 7am-7pm for middle of the trip, and 'dark' as 7pm-7am for the middle of the trip. We chose to take the middle time of the trip, as, for example, someone starting a ride at 6:50am and riding to 7:40am would be spending more time riding in daylight hours, but would be counted as 'dark' using the start time, and similarly someone starting just before 7pm and peddling for well after would be considered 'light' when most of their time was spent peddling in the dark. Finding the middle time eliminates this issue in most instances, giving an indication whether most of the trip took place in one category or other, and classifying it in this manner.

Methodology – After categorizing the ride-times into 'light' or 'dark,' we separated ridership by gender. We plotted this to get a visual sense or ridership, and then took a ratio of men/women by which to judge how the frequency of gender in ridership. Plotting this, we were able to visually compare the light vs. dark ridership by gender ratio, and we performed a unpaired t-test to gain further statistical insight. We initially thought about performing a Mann-Whitney U test, but due to the low number of observations in our final dataset (12 items in each category), this may have introduced more noise than clarity, and so we refrained.

Conclusions – Both the graph (Figure 4) and low p-value (0.0024 compared to our 0.05 threshold) give strong reason to reject the null hypothesis in this case. Therefore, we are able to conclude that the ratio of men to women ridership is higher than women at night. More analysis is critical before any generalized conclusions are drawn, as this data is from only New York City, and for only one month.

Figures

	tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name	
0	1338	6/1/2015 0:00	6/1/2015 0:22	128	MacDougal St & Prince St	40.727103	-74.002971	2021	W 45 St & 8 Ave	4(
1	290	6/1/2015 0:00	6/1/2015 0:05	438	St Marks Pl & 1 Ave	40.727791	-73.985649	312	Allen St & E Houston St	4(
2	634	6/1/2015 0:01	6/1/2015 0:11	383	Greenwich Ave & Charles St	40.735238	-74.000271	388	W 26 St & 10 Ave	4(
3	159	6/1/2015 0:01	6/1/2015 0:04	361	Allen St & Hester St	40.716059	-73.991908	531	Forsyth St & Broome St	4(
4	1233	6/1/2015 0:02	6/1/2015 0:22	382	University Pl & E 14 St	40.734927	-73.992005	532	S 5 Pl & S 4 St	4(

Figure 1: The data as initially ingested from the Citibike website's CSV source.

	men	women	ratio	dark
0	3240	637	5.086342	True
1	4777	1054	4.532258	True
2	2851	526	5.420152	True
3	1617	259	6.243243	True
4	1120	166	6.746988	True
5	1498	296	5.060811	True
6	8679	1717	5.054747	True
7	22820	5040	4.527778	False
8	41967	12254	3.424759	False

Figure 2: A view of the data after it had been categorized into 'light' and 'dark,' and grouped by gender. The 'ratio' of men/women is also included here.

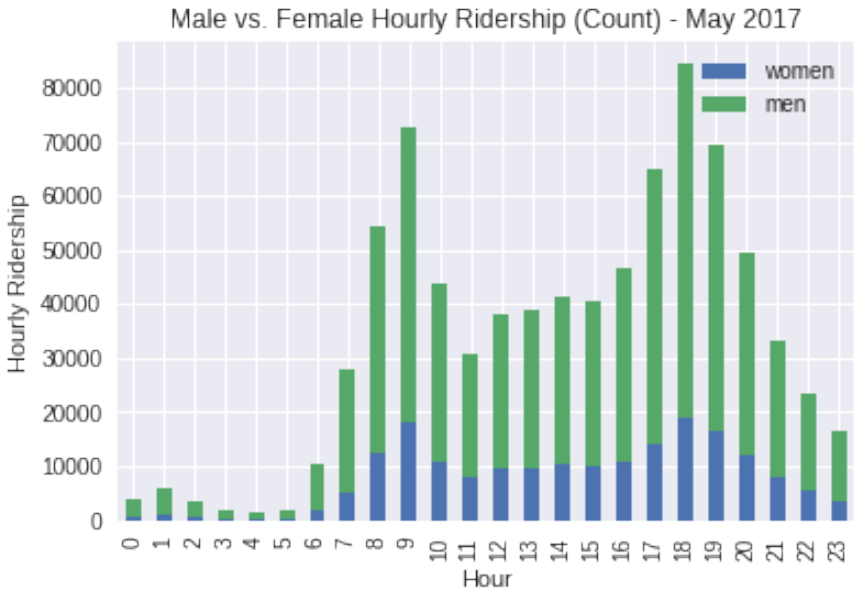

Figure 3: Here we see total ridership by men and women over the course of a month, and sorted by hour. We can see that men ride much more frequently, but cannot quite tell from this image how the ratio varies from hour to hour.

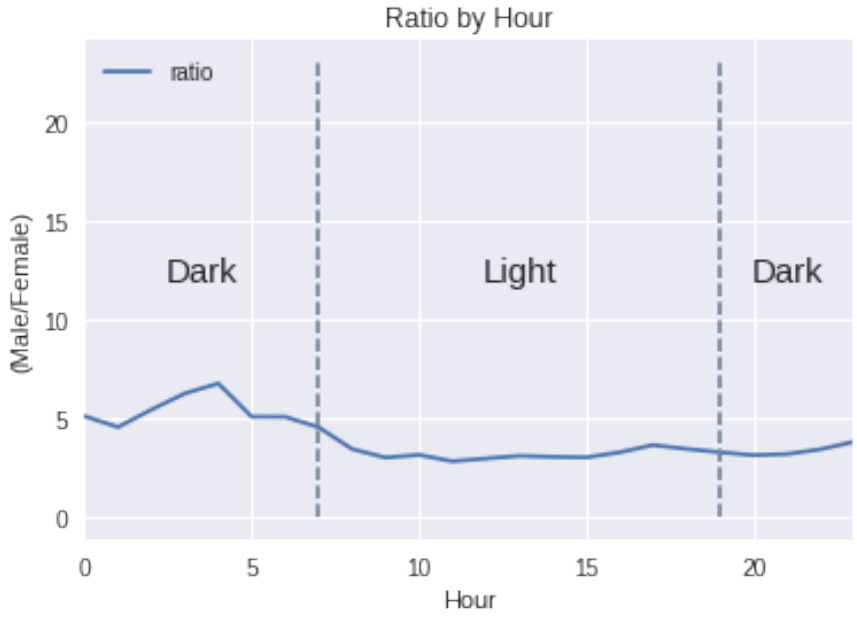

Figure 4: Here we see the ratio (Male/Female) graphed by hour for a month's worth of data. It becomes apparent that the ratio of men increases after dark, particularly in the early morning hours. There is also a relatively high ratio before 10am, but this is likely due to commute preferences, not safety in darkness, and so should be considered differently.