Title

Dana Chermesh[1] and Charlie Moffett[2]

[1]NYU Center for Urban Science & Progress
[2]New York University Shanghai

November 6, 2017

Midnight Citibike rides habbits, Men vs Women

Abstract - Dana and Charlie were curious as to whether per the common stereotype, men move about through the city at night more than women. To investigate, we set up the work for data-driven inference based on CitiBike data. In formulating our Null and Alternative hypotheses (with confidence level of 0.05), we sought to reject the assertion that women ride (in proportion to total female ridership) equally or more so at night than men (in proportion to total male ridership). Originally we had chosen midnight (00:00am-01:00am) as a proxy hour for night time, but later expanded our scope to include the hours between 8pm and 7am upon suggestion of our reviewers. The histograms we plotted demonstrated that our hypothesis was correct - men do ride more at night. We ran a Chi Squared Test to assess independence. This test is applied when you have two categorical variables from a single population. It is used to determine whether there is a significant association between the two variables. Given our confidence level and resultant Chi Squared statistic, we were able to reject our null hypothesis.

Introduction

Citi Bike is New York City's bike share system, and the largest in the nation. Launched in May 2013, Citi Bike had become an essential part of US metropolitans' transportation network (nyc). Citi Bike is providing data that is open to the public, inviting developers, engineers, statisticians, artists, academics and other interested people to use the data for analysis, development, visualization and trends discovering for any question regarding users profile, ride habits and more (nyc). Our research relies on common assumption that women probably ride more during daytime, in comparison to the differences between daytime and nighttime rides of men. This kind of understanding the different habits of the genders could contribute to the understanding men and women mobility / dynamics in the city.

Our research question, null and alternative hypotheses:

Q:

Is the percentage of women rides at midnight from total women rides significantly lower than men's midnight rides percentage?

Null Hypothesis

The percentage of women rides at midnight from total women rides is similar to or greater than men's rides percentage from total men's rides at the same hour?

Alternative Hypothesis

The percentage of women rides at midnight from total women rides is significantly smaller than men's rides percentage from total men's rides at the same hour? *significance level*

Data

For our analytics, we first picked *one month* from Citi Bike open data(dat) to test our hypothesis. We used pandas to read in the Citi Bike files. We extract the hour of the start time of every ride, and aggregated men and women (*1=men and 2=women* in Citi Bike data) for each hour of the day, averaged over the month.

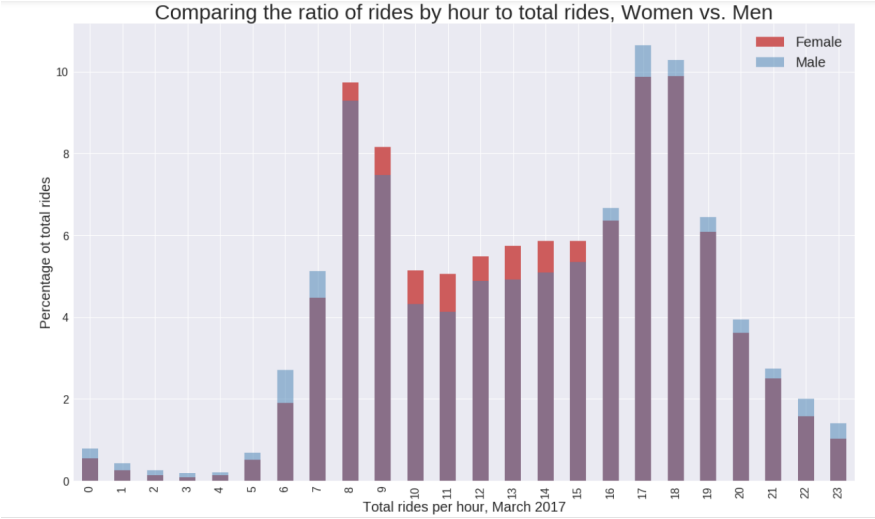

Figure 1: **Comparing the ratio of rides by hour to total rides, Women vs Men, June 2017**

Reviewing the above distribution and looking specifically on midnight (hour = 0), it is obvious that the % of men rides at midnight from total men rides is higher than the % of women rides at midnight from total women rides over the analyzed month. Thus we can *reject the Null Hypothesis.*

Than, we clustered the hours to 'Daytime' and 'Nighttime', 7am-8pm and 8pm-7am respectively.

Reviewing the above distribution it is obvious that the % of men rides during nighttime from total men rides is higher than the % of women rides during nighttime from total women rides over the analyzed mounth. Thus we can *reject the #2 Null Hypothesis.*

At last, we included 5 more months and used the same methodology of our first analysis, reexamine our Null hypothesis based on 6 months data instead of only one month. The results were quite similar to our one-month analysis.

Reviewing the above distribution, whether looking specifically on midnight (hour = 0) or at all nightime hours (8pm-7am in our research), it is obvious that the % of men rides at these hours from total men rides is higher than the % of women rides at the same timeframe from total women rides over the analyzed six months. Thus we can reject the #1 and #2 Null Hypotheses.

Methodology

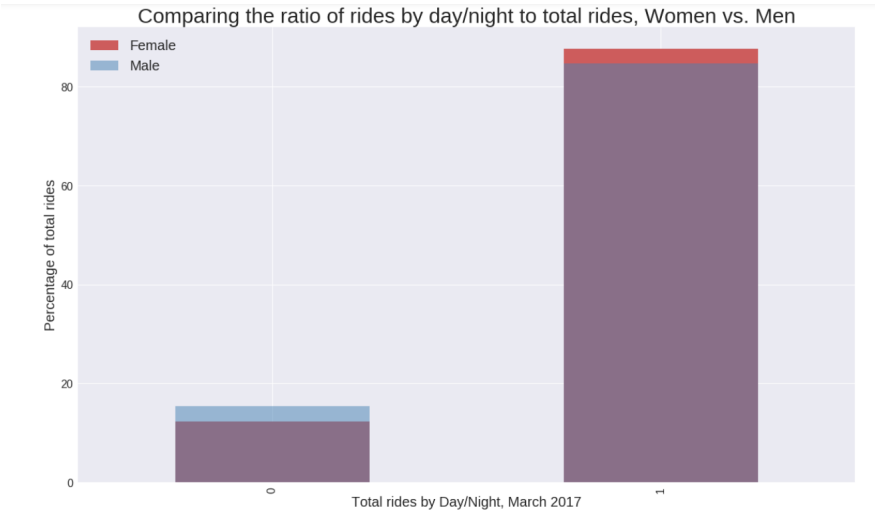

Figure 2: **Comparing the ratio of Daytime and Nighttime rides, Women vs Men, June 2017**

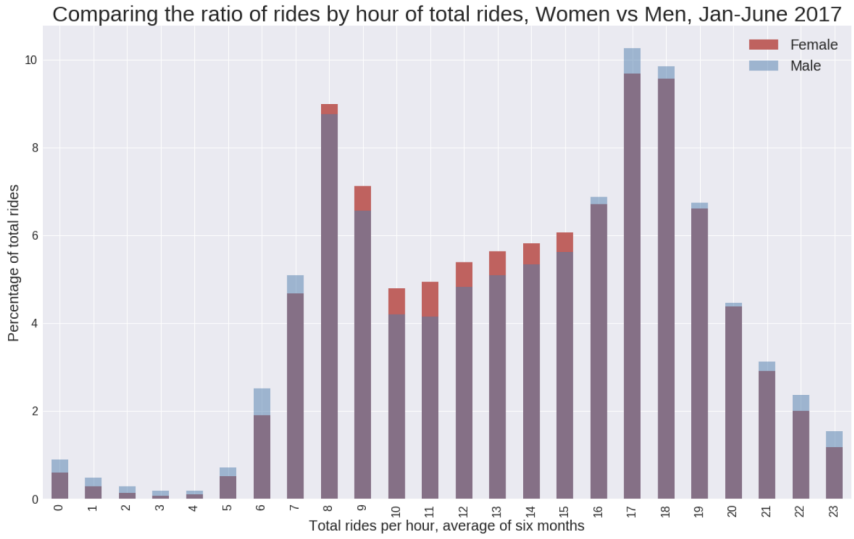

Figure 3: **Comparing the ratio of rides by hour to total rides, Women vs Men, Jan-June 2017**

Using the table from CSUN, one of our reviewers (Rachel) suggested we use Chi Square test to measures the differences between groups. This makes sense if we look at the table from CSUN, given that the male and female variables in our experiment are unpaired and categorical. So we defined a function to perform the Chi Square statistical test and produced a contingency tables of values in order to derive our statistic. With this in hand, we were able to look at the 'Percentage Points of Chi-Square Distribution' table and draw conclusions about the rejection of the Null. The Null hypothesis that women ride at night, measured as: ratio of each gender riding at night to total ridership for each gender, in an equal or higher percentage than men can be rejected at alpha = 0.05 with a chi square statistics of 382.53.

Conclusions

Our research has been shown that, beyond reasonable doubt, women's night rides percentage from total

women rides is consistently lower than men's night rides percentage from total men rides. Our original experiment was pretty narrow in scope, so we added more months to our data and broadened our proxy for nighttime to augment the original results.

References

Citi Bike Open Data. https://s3.amazonaws.com/tripdata/index.html. URL https://s3.amazonaws.com/tripdata/index.html. Accessed on Mon, November 06, 2017.

Citi Bike NYC. https://www.citibikenyc.com/system-data. URL http://www.citibikenyc.com/system-data. Accessed on Sun, November 05, 2017.