A Citi Bike Analysis of Riders Health Conditions by Gender

td928[1]

[1]Affiliation not available

November 11, 2017

Abstract

The purpose of this study is to explore whether females over age fifty is in better physical conditions than their male counterparts by studying their Citi Bike usage. The premise of the study is that more Citi Bike Uses suggest better overall heal conditions. This study finds that the proportion of female Citi Bike riders above age fifty is in fact lower than the proportion of male Citi Bike Riders above age fifty.

Introduction

The Citi Bike is a bike sharing project in New York City. The program allows users to simply self check out bikes at stations and return them to any stations after uses. It supposes to be an integral part of the New York City transportation system providing a sustainable and alternative way of commuting. Citi Bike generated tremendous potential to study urban problems for its temporal and spatial components in the dataset. This study suggests to particularly look at the Citi Bike as an instrument to gauge urban health status. Biking is not just a method of commuting but also can be treated as a form of physical exercise. This study is interested in whether differences in physical conditions of genders in certain age group can be inferred from the Citi Bike data.

Data

The data used for this study is provided by the open data from NYC provided by Citi Bike. The data from the month of June and December for year 2016 for this study. This choice is based on the intuition that the selection of one month each from both summer and winter season could control for some variations due to the weather conditions. Then dataset is further reduced by checking if one of the categories for gender or birth year has missing entry (NaN) and is deleted from the dataset. Then the bikers whose age is above fifty and also born after 1930 is selected to be the final dataset upon which analysis is performed. See **Figure 1** for a snippet of the final dataset.

	gender	birth_year
0	1	1964.0
1	1	1952.0
2	1	1952.0
3	1	1961.0
4	1	1963.0

Figure 1: A sectional look at the final dataset

Methodologies

As suggested by the peer review process, the proportion z test is proposed to perform on the dataset. Since the tests is supposed to show whether the one proportion is significantly higher than the other. A one-tail Z test will make sense for the and it is a one-tail test for female riders proportion should be higher than the male proportion. The chi-square proportion test is also a sensible choice for testing the difference in proportion. The proportion of the female riders over age fifty is calculated by dividing the total female population by the female riders who is born between 1931 and 1966. And the male proportion is calculated likewise. Refer to Fig. 2 and Fig.3 for the plotted percentage of female and male riders by their birth year.

Conclusion

The dataset is indicating that female riders are actually likely. In fact, the dataset is indicating the study that is exact opposite of this research might be worthy to explore. The proportion of male riders over age fifty is in fact higher than the proportion of female riders. But it is waiting to be seen whether this difference is significant in its own right. Some of the limitation for this method might include but not limited to: i) whether Citi Bike usage can reflect the true physical conditions. The choice of age over fifty is completely arbitrary and is not supported by any medical research on physical decline and age. ii) The difference in overall size of female and male riders is quite notable in its own right. The male riders have over 300,000 records while the females 80,000 records. This might suggest some other latent uses pattern might exist for gender difference. For example, women might choose to use the bikes less because they are more likely to wear attires that are inconvenient for bike riding.

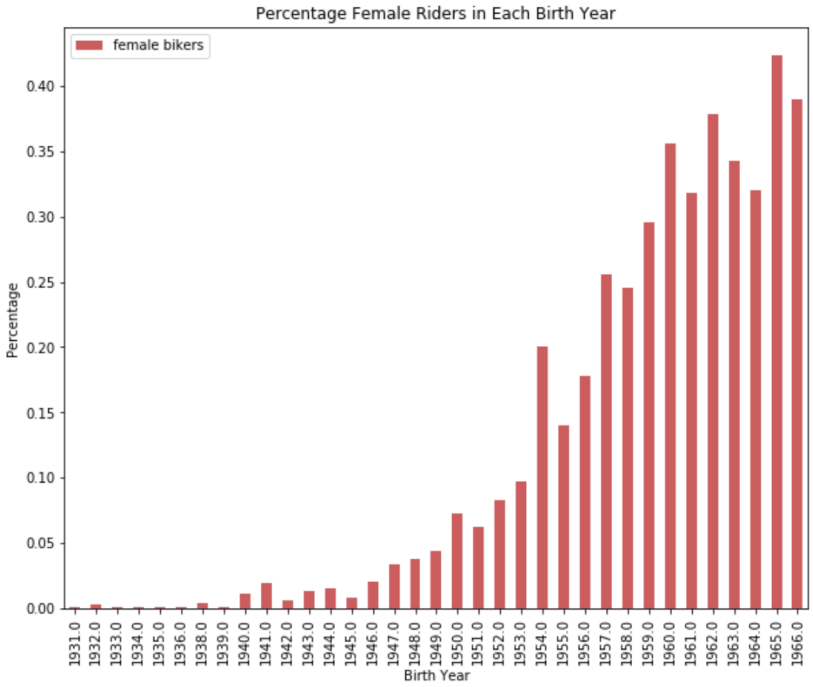

Figure 2: Female bikers histogram by their birth year

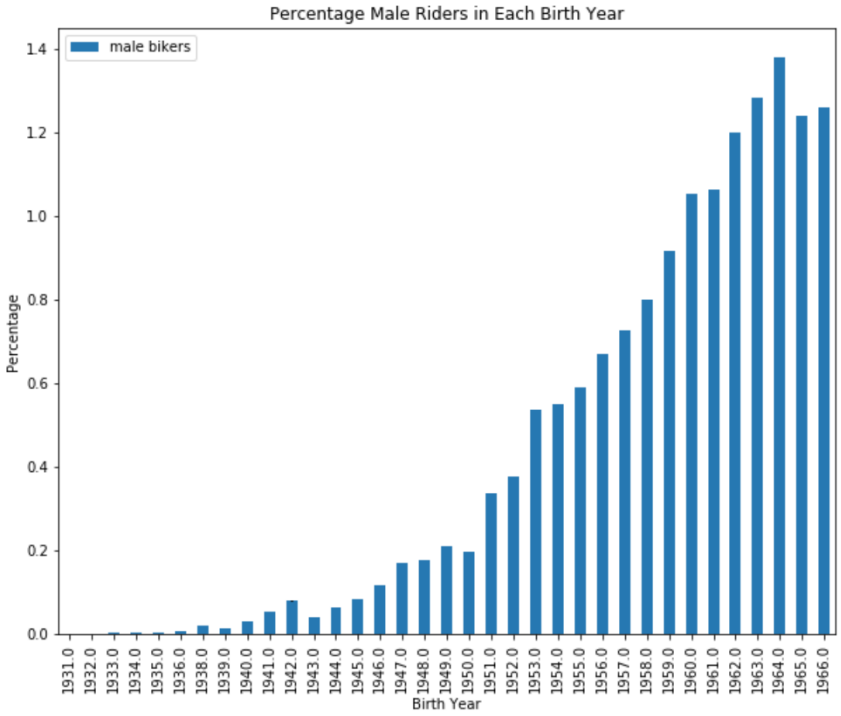

Figure 3: Male bikers histogram by their birth year