**The Citibike data**

New York City launched its bike-sharing program, Citibike, in 2013. To date the program has seen over 20 million rides from individuals of all ages, in all weather conditions, at all points of the year.

New York City posted selected data of its ridership on its website. The data include the day and time of the ride, the age of the rider, and the start and stop stations of the ride, and trip duration.

From this data, we asked whether it is possible infer the rate at which riders slow down as they age. Given that we can narrow down rides that leave from and to the same stations at nearly the same time (within an hour) by age, there is the potential to build out a curve of “slowdown by age.”

Ultimately we narrowed our data down to 1.3 million rides which took the same route within an hour of each other at different ages. However, as discussed below, our results are by and large inconclusive.

One potential explanation is that age is not the determining factor of how fast a Citibike rider goes. The real determinant is likely traffic, both pedestrian and vehicular. There is only so fast a biker can travel in Times Square in rush hour, no matter the age.

Due to the lackluster results from our initial analysis, we also compared the Citibike biking times to the expected biking time as predicted by Google Maps, by age. The initial results appear promising.

**ETL & Parallelization**

The data were stored in a Postgres Database via Amazon Web Services. Due to the cleanliness of the data, this solution provided rapid access and manipulation. Earlier attempts to process the data in hive and spark-sql failed due to the slowness of manipulating the data.

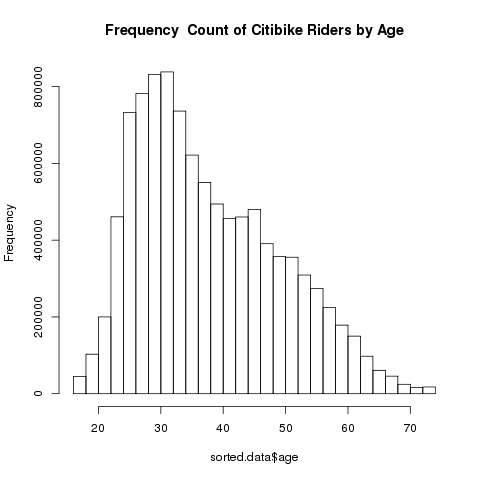
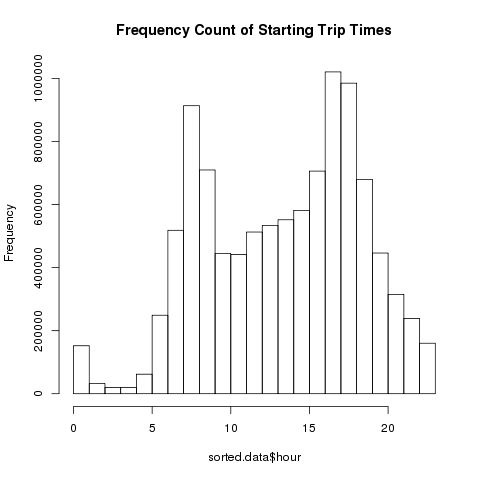
Data manipulation and graphing were conducted in RStudio Server, using the Rpostgres, dplyr, ggplot2 and multiplyr packages. Multiplyr allowed for parallelization of the computations to 7 cores, exponentially speeding the process.

Google maps data were downloaded using the ggmaps package for R, and the results were stored in a CSV file on Amazon Web Services and then subsequently re-loaded into RStudio for processing.

**Data Description**

The total dataset contained 21.3 million rides from 2013 through February 2016 that had all data points needed for this analysis. From this data, we removed all rides where the stop station and the end station were the same, and analyzed the results.

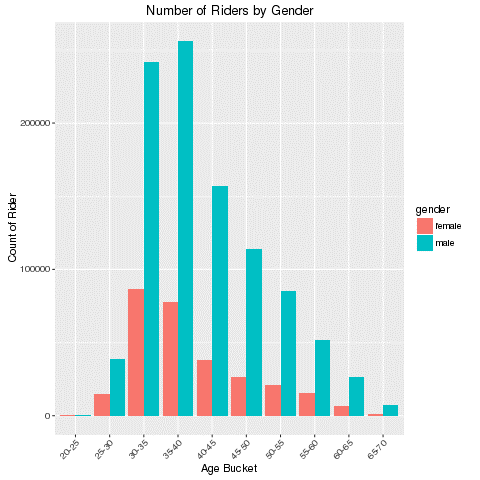
The median age of a Citibike rider is 36 years old, with the peak of the distribution between 28-32 years old. The bulk of the rides occur at rush hour, suggesting that these 28-32 year olds are using Citibike as a commuter device. The median time is only 10.3 minutes: Citibike riders tend to travel to nearby destinations.

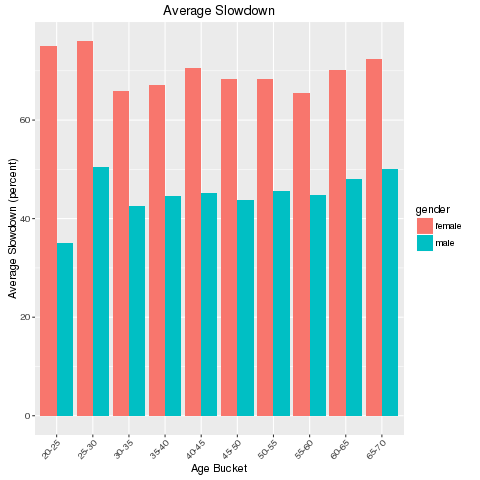
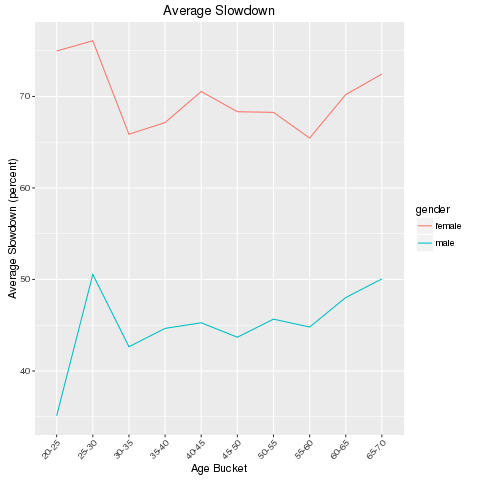
**Data Processing**

To ensure that weather, traffic and other confounding variables besides age were held constant, we narrowed the data set to rides that traveled between the same stations at within the same hour (defined as hour on a clock, i.e. between 8 am – 9 am, not on a rolling 60 minute basis).

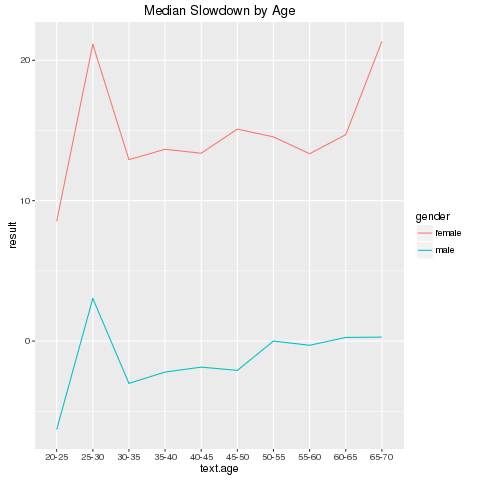
We then divided our ridership into 5 year age buckets, beginning with ages 15-20 and ending with ages 75-80. After narrowing the data, we analyzed 1.27 rides concentrated among men ages 30-35 and 35-40. The count falls to nearly 0 above ages 60-65.

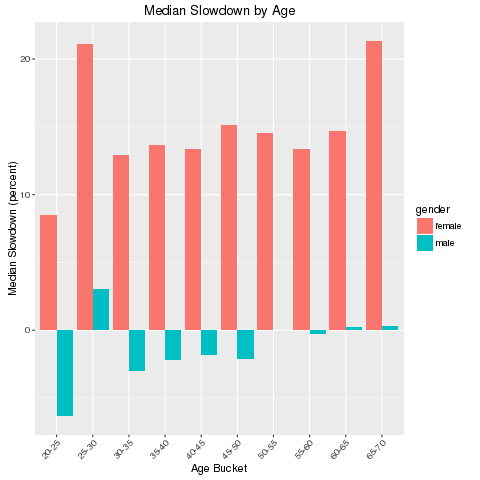


Once the processing was completed, our initial results showed that, on average, men slowed down by roughly 40-50% for every 5 years of aging, while women slowed by 60-70%.



The data show that as people age, we tend to slow down by roughly the same percentage for every 5 years. This conclusion is generally unsatisfactory. It largely defies common sense to believe that the move from 25-30 years old to 40-45 years old is accompanied by a nearly identical level of reduction in physical capacity.

A look at the median slowdowns similarly confirms that our data do not conform with simple common sense. The medians show that the median man actually modestly speeds up as he gets older. Perhaps the large acceleration (negative slowdown) at ages 30-35 is caused by selection bias: only athletic minded men continue to use Citibike after the age of 30. However that explanation requires running to the extremes by the time riders are between 45-50. We would have to believe that men between the ages of 45-50 are traveling faster than your average, already athletic, 30 year old.



**Google Maps Strategy**

Due to the disappointing results of our data, we attempted another strategy using Google Maps. Google maps allows users to download the forecasted biking time between two locations in its database. Users are allowed to download 2,500 free data points, though only half of those if geocodes are used, per day.

For our analysis, we generated 96,000 unique pairs of stations and randomly sampled from those the key pairs we would analyze. We calculated the average *deviation* of Citibike’s trip duration from Google’s expected time.

