Jeremy Neiman

R Core Challenge

Neither of the models (with or without outliers) are good. They both predict a very weak correlation between age and trip duration – with outliers it predicts an increase of 2 seconds/year of age and without outliers it's 1 second/year. In both cases the coefficient of determination is near zero meaning that there is a lot of error in the regression.

Looking at the scatter plots, it appears that there may be a slight parabolic distribution with time decreasing as age goes to the extreme left or right. But there is too much data to make that determination by eye, so based only on the linear regression we did, we can't conclude any correlation between age and trip duration.

Working with the premise that there isn't a correlation between age and trip duration, our model without outliers is better, coming closer to a perfectly flat line. The model with outliers seems to be thrown off a bit by the glut of 115 year old's cruising around the city. Out of curiosity I went to the citibike registration page to see if you were to just put in a random birthday instead of your real birthday, whether it was easiest to put in 1899, such as it being at the top of the drop down. But (at least now) that isn't the case. So why people choose 1899 as a fake birth year when they register remains a mystery!

But it wouldn't be fair to come to the conclusion that the older population can get just as much utility out of Citibike as younger people. The older riders are almost certainly self selecting – that is, the people who use Citibike are the ones that can make good use out of it. And this is backed up by the total number of trips being done by age. The number of trips made by 30 year olds is 1000 times higher than those made by 80 year olds. Basically, Citibike users, regardless of age, take trips of similar times. But Citibike users are prominently people in their 20s to 40s and it alienates people who are older and no longer as capable of riding a bike.

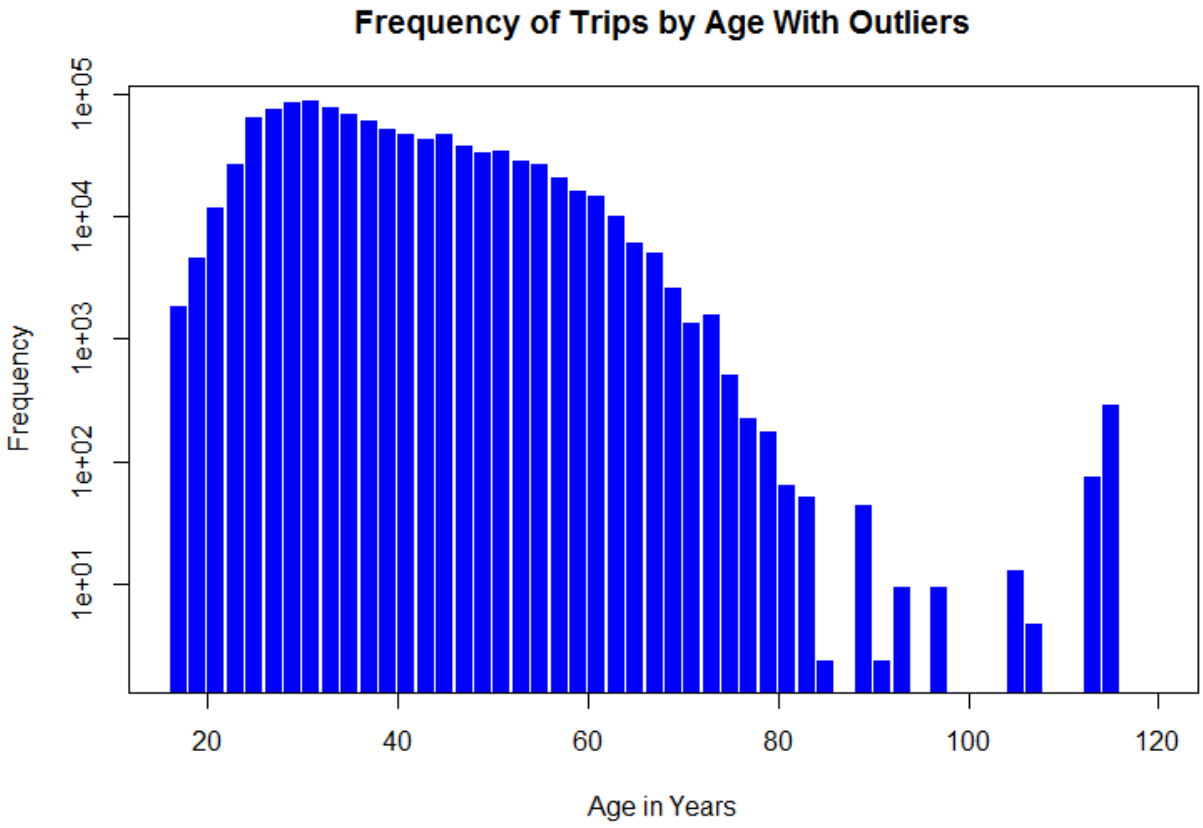
I could certainly recommend a bike share as a part of a multi-modal public transportation solution, which people of all ages could find useful, but it's clear that if it were meant to replace other forms of transportation, such as buses, many people would be alienated.

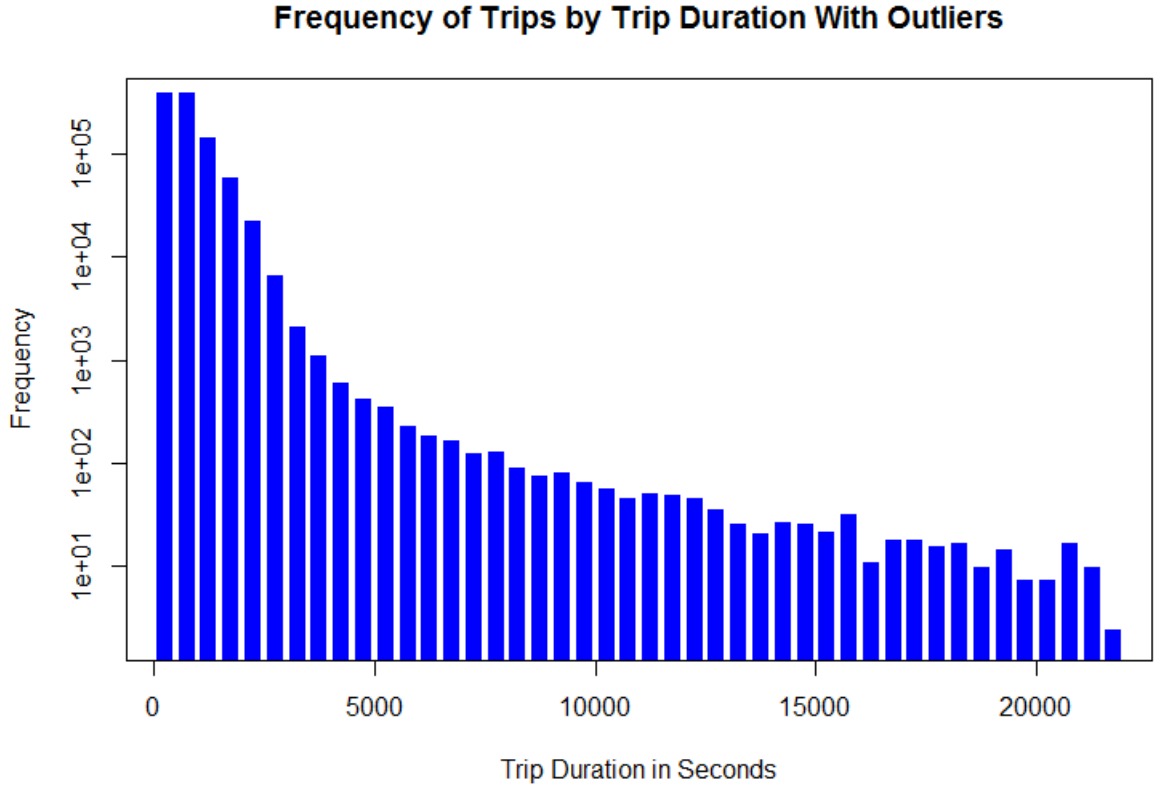
Statistics & Graphs

With Outliers:

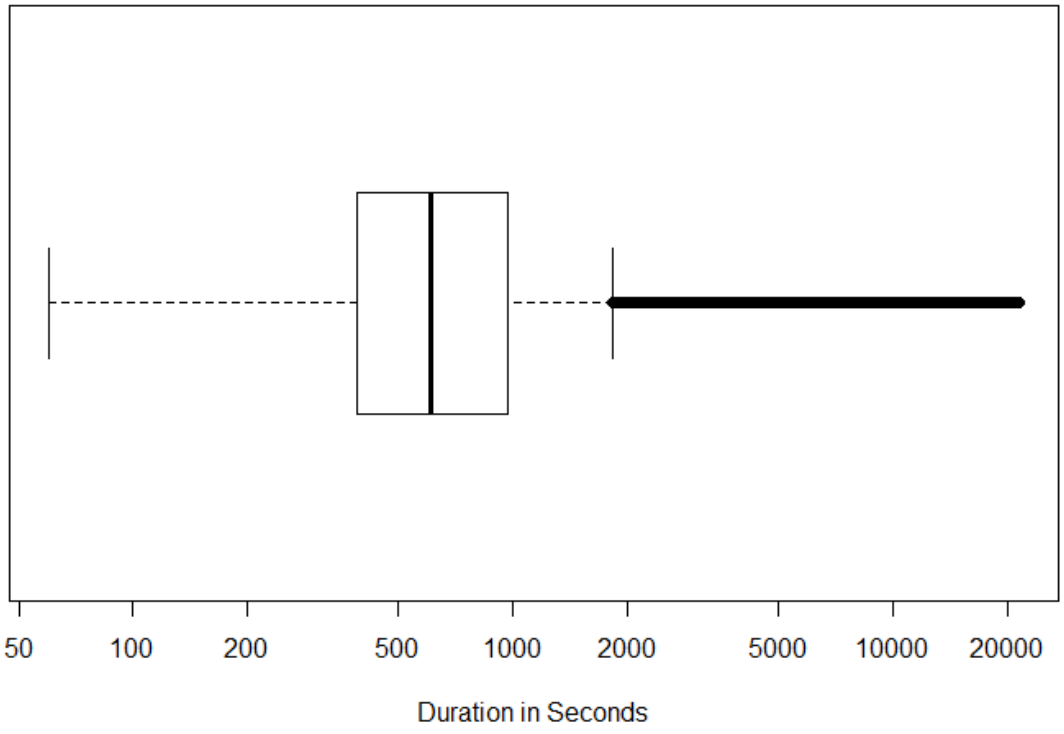
Summary Statistics

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Trip Duration	60	390	606	767.6	965	21560
Age	16	30	36	38.69	46	115

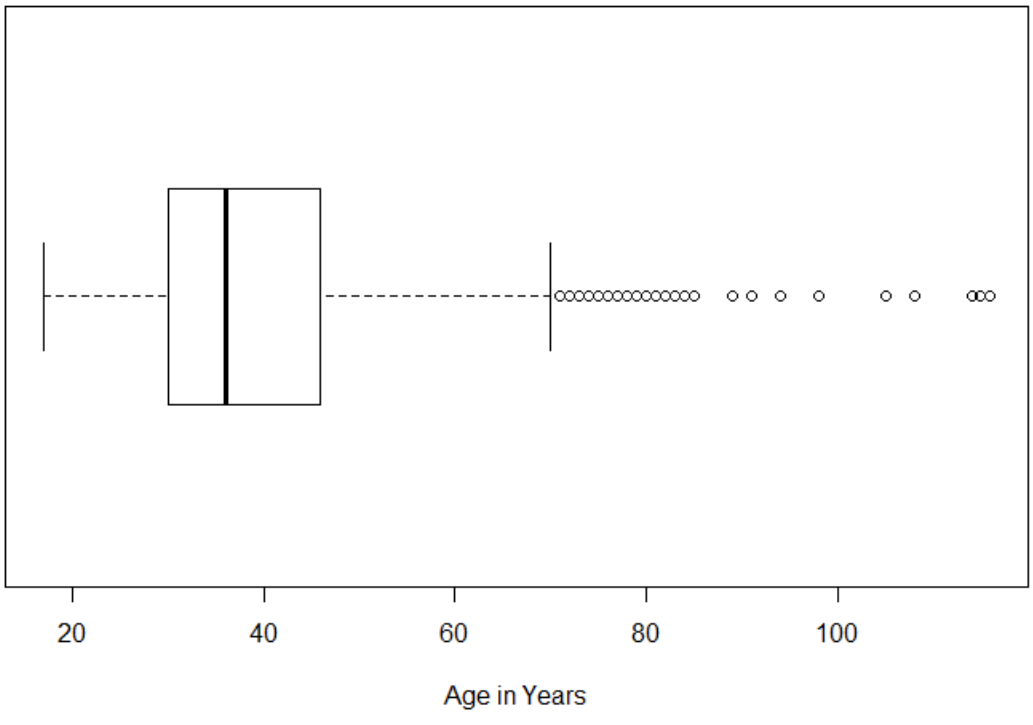

Frequency of Trips by Age With Outliers
Frequency
1e+01
1e+02
1e+03
1e+04
1e+05
20
40
60
80
100
120
Age in Years

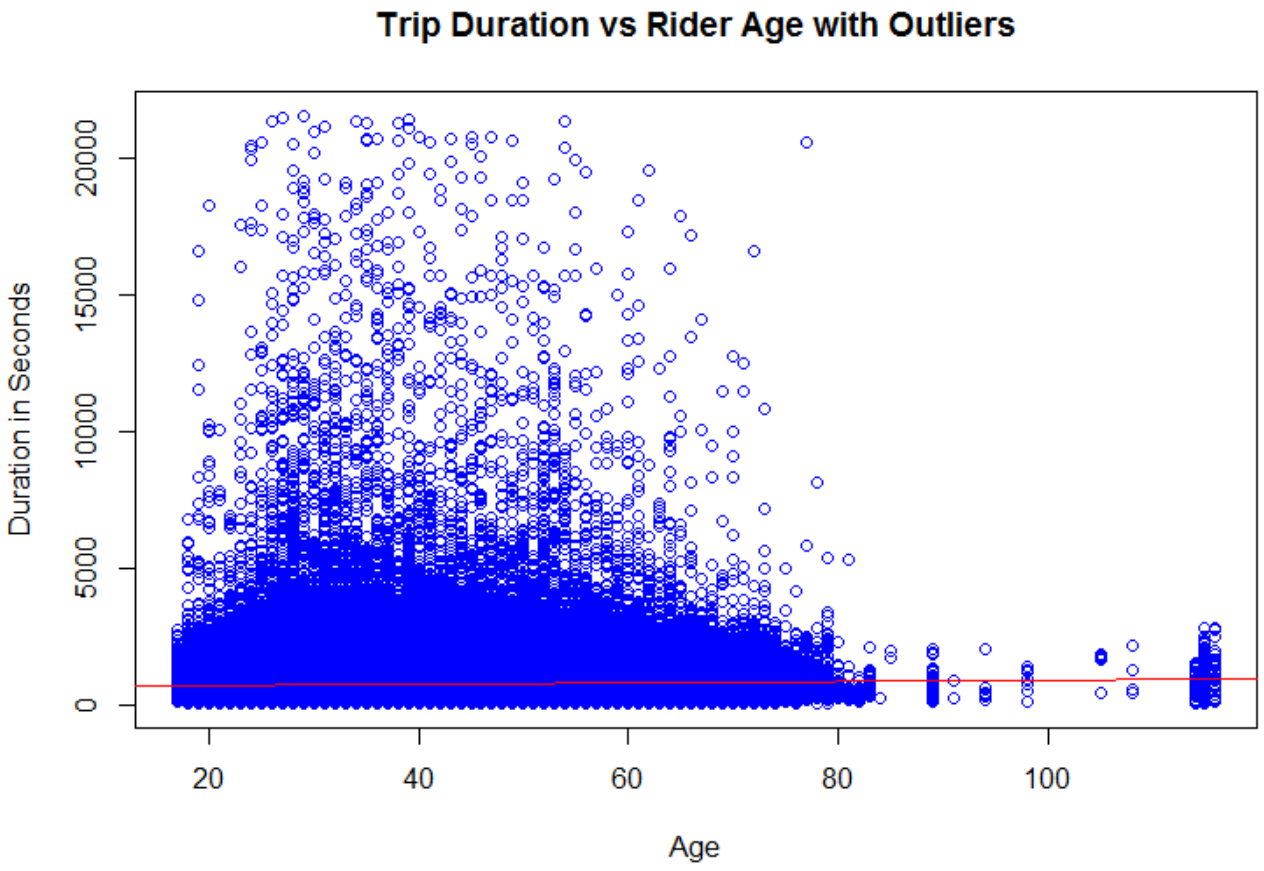

Frequency of Trips by Trip Duration With Outliers
Frequency
1e+01
1e+02
1e+03
1e+04
1e+05
0
5000
10000
15000
20000
Trip Duration in Seconds

Trip Duration with Outliers



Age with Outliers

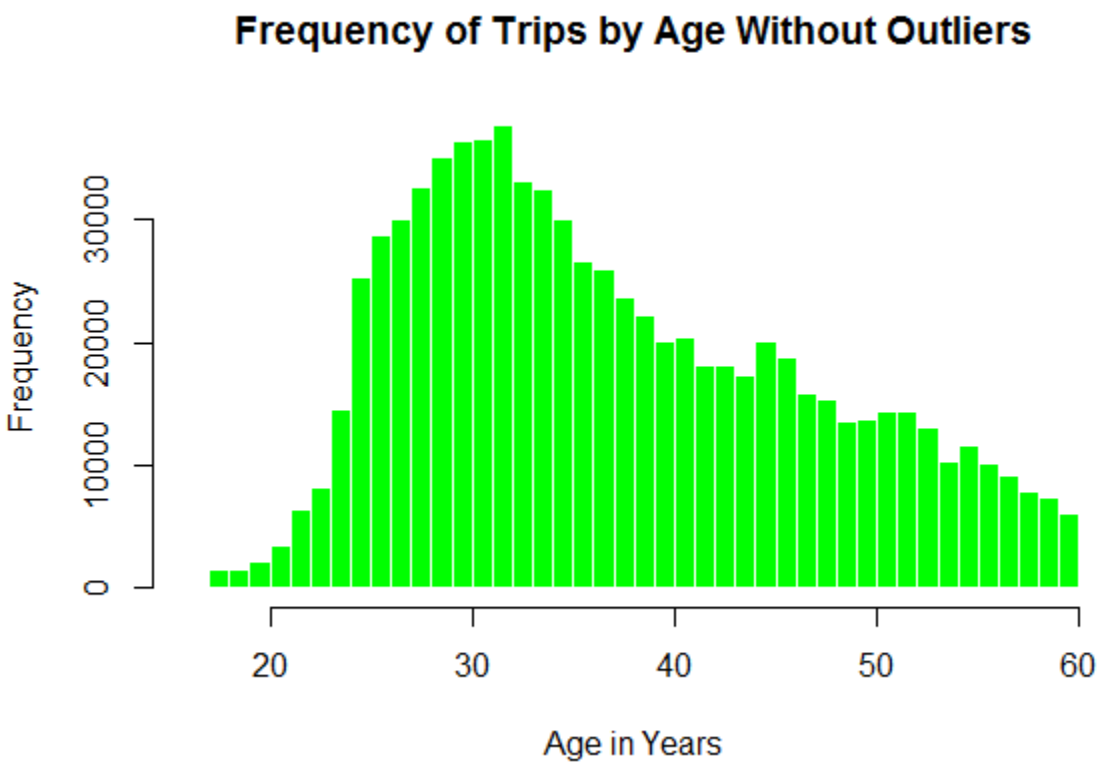


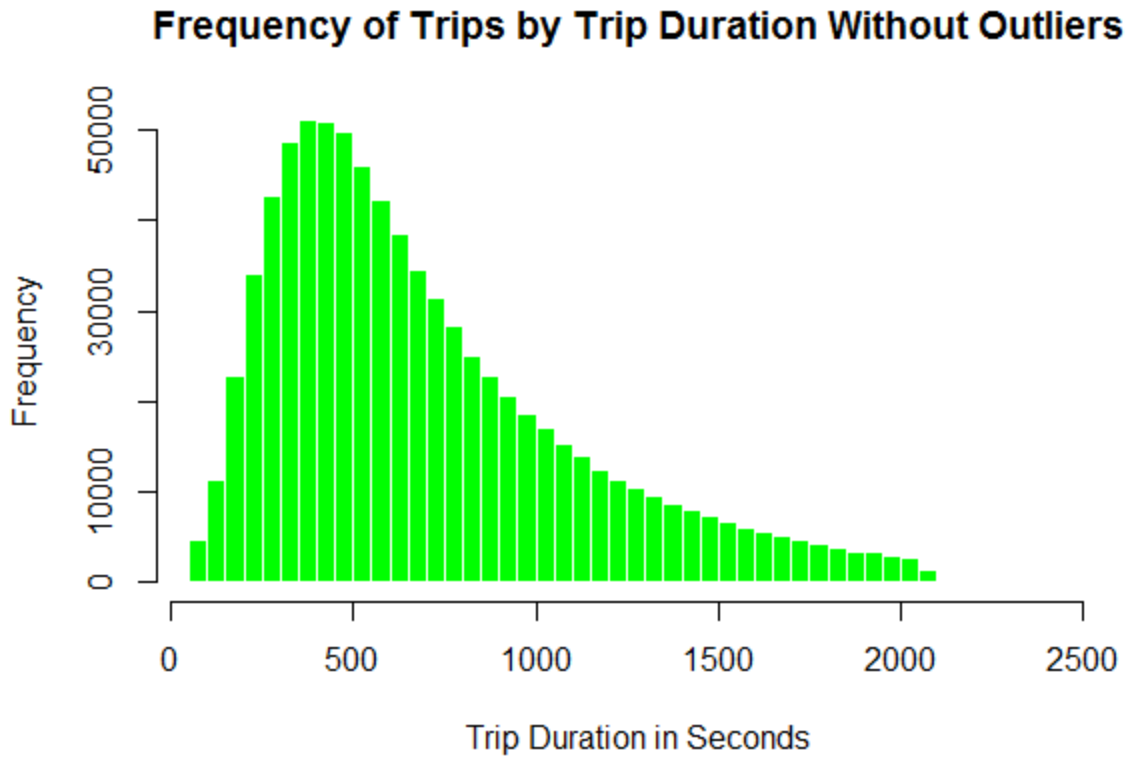


Without Outliers:

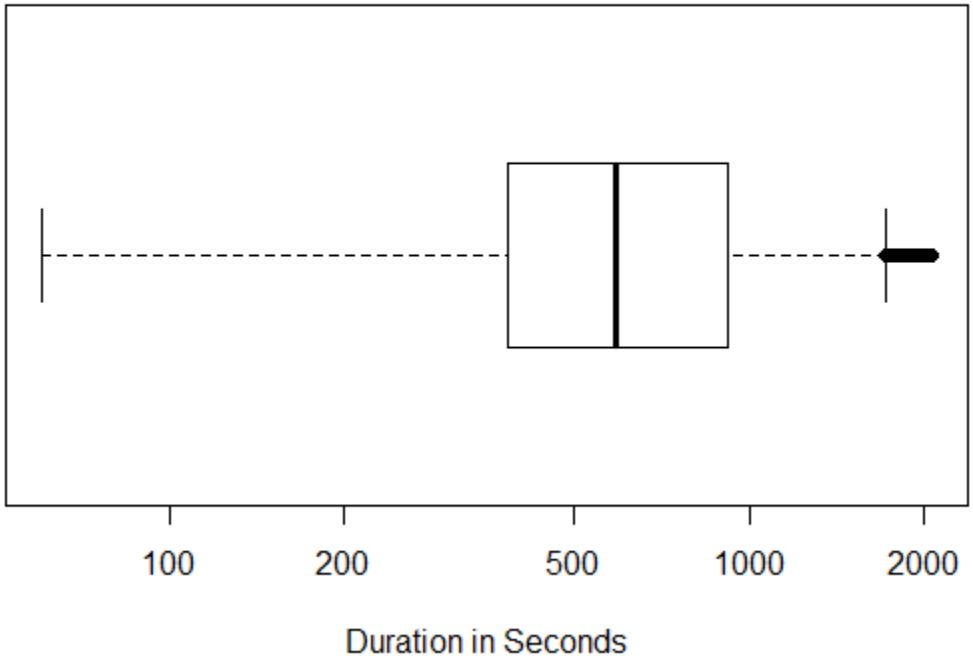
Summary Statistics

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Trip Duration	60	383	589	698	917	2077
Age	16	30	35	37.46	45	60

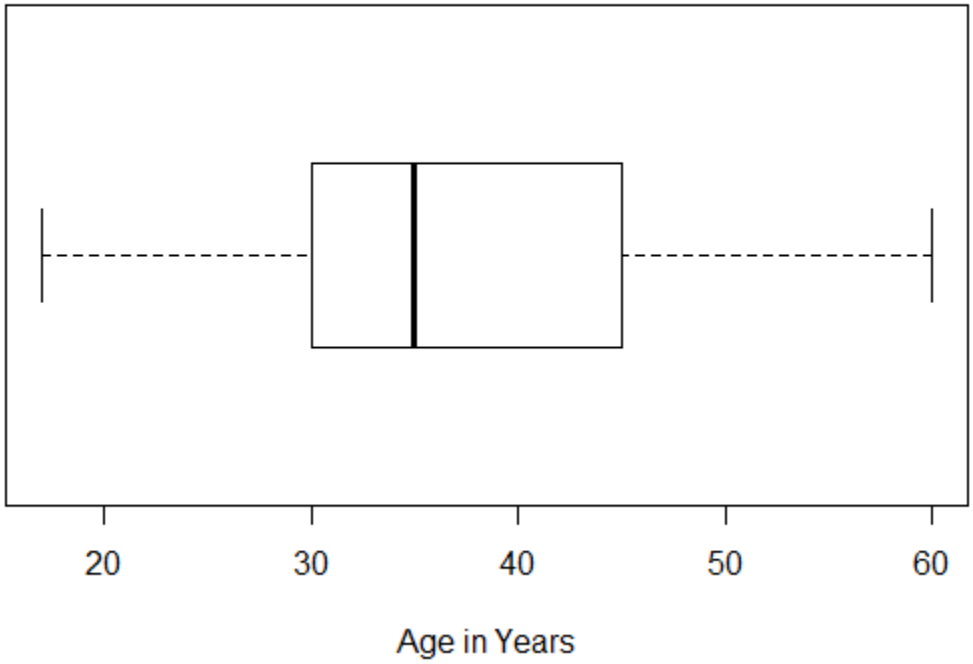

Frequency of Trips by Age Without Outliers
Frequency
0
10000
20000
30000
20
30
40
50
60
Age in Years

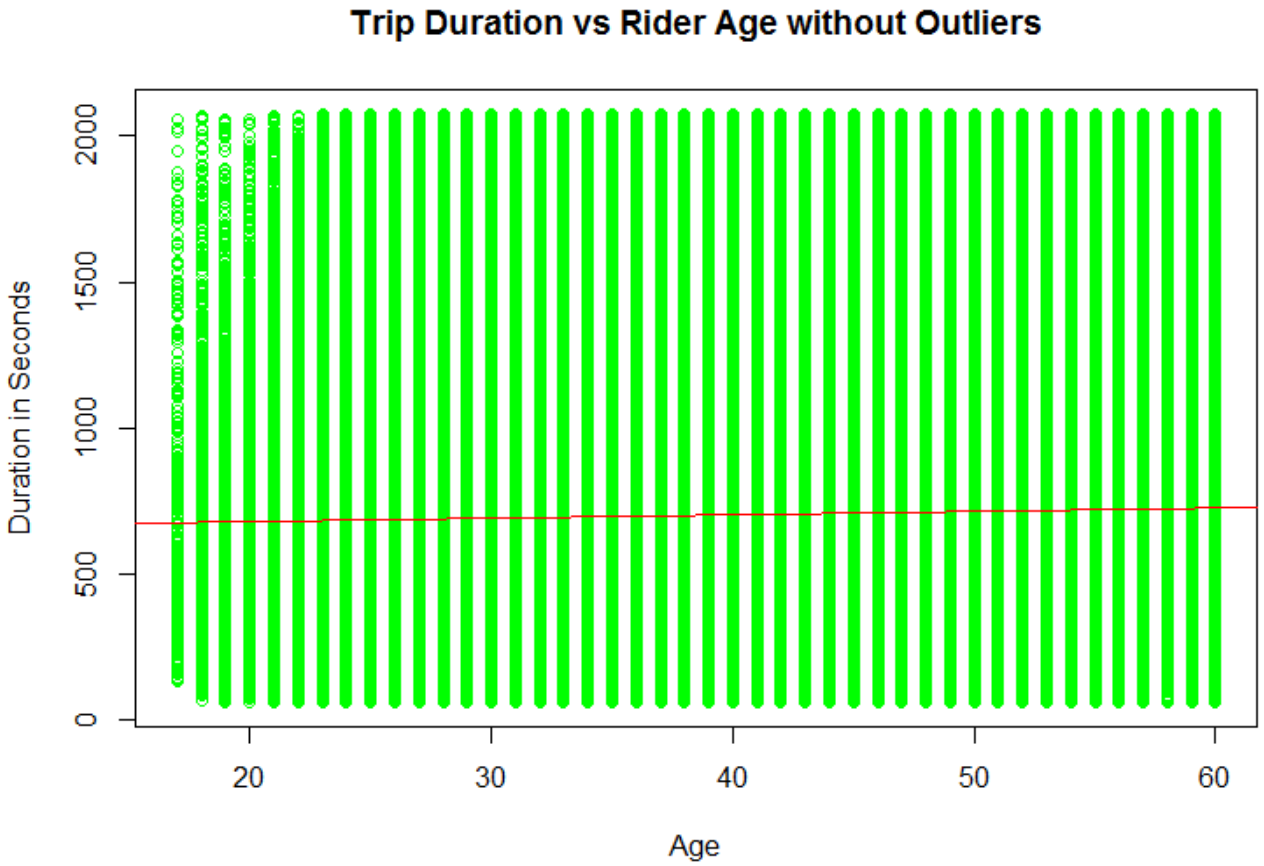

Frequency of Trips by Trip Duration Without Outliers
Frequency
0
10000
30000
50000
0
500
1000
1500
2000
2500
Trip Duration in Seconds

Trip Duration without Outliers



Age without Outliers




Trip Duration vs Rider Age without Outliers
Duration in Seconds
0
500
1000
1500
2000
20
30
40
50
60
Age