The NYC Taxi & Limousine Commission (TLC) tracks data on how taxis and other for-hire-vehicles (FHVs) like Uber and Lyft are used. These data include when and where fares are picked up and dropped off. The number of taxis or FHVs on the road at any given time contribute to the overall number of vehicles on the road. However, the traffic contribution of FHVs may be negligible compared to the traffic contributions of residential vehicles, tourists, or construction.

Because of these unknown factors I chose to use the provided data to get an estimate of average taxi speed. This measure would also be influenced by traffic due to residential vehicles, tourists, and construction and so might give better insight into how traffic has changed since Uber arrived in NYC in May 2011 (https://www.uber.com/blog/new-york-city/uber-nyc-launches-service/). My first step in processing these data was to aggregate the publicly available files down to averages: I wrote a program which calculates summary values for every hour (0-23) of every weekday (Sunday-Saturday) based on a complete month of data. The output is a file which has 168 rows (24 hours/day * 7 days/week) where 0Tuesday = 12 AM Tuesday; each row contains the averages for the number of Yellow Cab fares and associated average distance & time traveled and speed, as well as the average number of FHV fares.

Month Sample Collected	Avg Yellow Cab Fares per week	Avg FHV Fares per week	Avg Yellow Cab Speed (MPH)	Total Yellow Cab & FHV Fares per week
Aug 2010	3,132,044	N/A	10.662	3,132,044
Apr 2011	3,679,743	N/A	11.255	3,679,743
Aug 2017	2,105,516	3,976,475	11.631	6,081,991
Apr 2018	2,326,257	5,268,599	11.280	7,594,856

Table 1: Summary of Yellow Cab Fares/Speed & FHV Fares for April 2011 and August 2017

The most immediately notable parts of table 1 show that Yellow Cab fares dropped steeply after Uber arrived, while Uber now collects on average more fares than Yellow Cab originally did. As a result the overall number of fares being collected has gone up by roughly 2.4 million per week. I was surprised to find that the average taxi speed increased from 2011 to 2017 by 0.376 MPH. While it is a small amount, a paired samples T-test shows this growth to be significant at $p<0.05$. However, the effect size is fairly small at only 0.163. This finding is counter-intuitive, so I decided to re-sample and compare August 2010 against April 2018 in case the month somehow drives increased taxi speed, or if there were some abnormal event that impacted my finding. I found no significant difference between taxi speed from Apr 2011 to 2018. The taxi speed increas from Aug2010 to 2017 was highly significant at $p<0.01$ with a corrected effect size of 0.427.

If we plan to further explore this question I suggest three areas of focus: traffic due to non-commercial private vehicles, traffic in high traffic density areas rather than NYC overall, and traffic as a continuous trend month to month. I did a quick web search to find reliable data on non-commercial vehicles and found that the total vehicle registrations in NYC went from 1.962 to 2.107 million between 2010 and 2015 (http://www.nyc.gov/html/dot/downloads/pdf/mobility-report-2016-screen-optimized.pdf). If we estimate these vehicles are used once per day from home to a destination and back we find an increase equivalent to 2.03 million fares per week; this may contribute as much to traffic as the 2.4 million increase in weekly fares from 2011 to 2017.

As for traffic in high density areas, the aforementioned mobility report points to slowing travel speed in Manhattan, particularly south of 60th street. I would have liked to look at specific areas of high traffic concentration, however there are over 250 locationIds in the provided data and this is meant to be a preliminary exploratory report. Selecting areas to focus on and reporting the findings will take more time and space and I judged it would be better to report my findings and discuss before moving forward. Finally, a continuous month to month trend using simple linear regression could tell us more about how trends in FHV use affect traffic over time. However, each input data set takes about 30 minutes to process, much of this time is consumed in simply reading over the file a single time, so I feel I've already stretched the "few hours" time constraints by spending at least 3 hours processing alone.