[DRAFT]
On-demand public transport is making us mobile

RA

2021-06-23

1 Introduction

The TLC trip record data[1] records Yellow and Green taxi and other "for-hire vehicle" trips in New York City. In the earlier years in particular, Yellow and Green taxi records contain timestamps, trip distance, passenger count, fare, etc., but also approximate locations of pickup and dropoff. Based on those data we give a partial answer to the question

> How many adaptively routed minibuses could service the same demand?

Conventions. The number in the margin refers to the corresponding code listed in §6.1.

2 Data preparation

2.1 Taxi trips[2]

We focus henceforth on May 2016 with ∼12M (Yellow) and ∼1.5M (Green) trip records. The ∼11M "for-hire vehicle" records bear no useful details for our purpose. We keep only the trips that begin and end in Manhattan with reported trip distance between 0.1 and 30 miles. We filter out records that lack geo-coordinates. See Fig. 1 for a net summary. #1

2.2 Road graph[2]

We obtained the road network for Manhattan from the OpenStreetMap Overpass API and #3
filtered for roads that can plausibly sustain public traffic. It is represented as a digraph, #4
i.e. there are one-way roads and the routing $A \to B$ differs from $B \to A$. Edges are broken into bits under 20 m. When modeling individual trips, their reported pickup and dropoff

[1] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page
[2] The codes for this section were mostly written in 2019.

locations are snapped to the nearest node of the road graph; we ignore about 10% of the trips where the discrepancy is over 20 m. The map graphics are from MapBox. #5

2.3 Traffic model

The trip trajectories are not available, only the pickup and dropoff locations (with potential GPS uncertainty of 10 to 100 m). We leverage the reported trip duration to infer plausible mean-field travel speeds (see Fig. 2c) throughout the road graph iteratively as follows. The speeds on all roads are initialized to 5 m/s. For a sample of trips that start at 6-7 pm the quickest trajectories are estimated. The speed of the road bits participating in those trajectories are adjusted toward the reported trip duration. This is repeated. #6

Henceforth, "quickest" means w.r.t. this traffic model. We are assuming that drivers are not sufficiently incentivised for detours (the credence good asymmetry is not significant).

3 Optimization problem

We are facing a so-called vehicle routing problem with these main attributes:

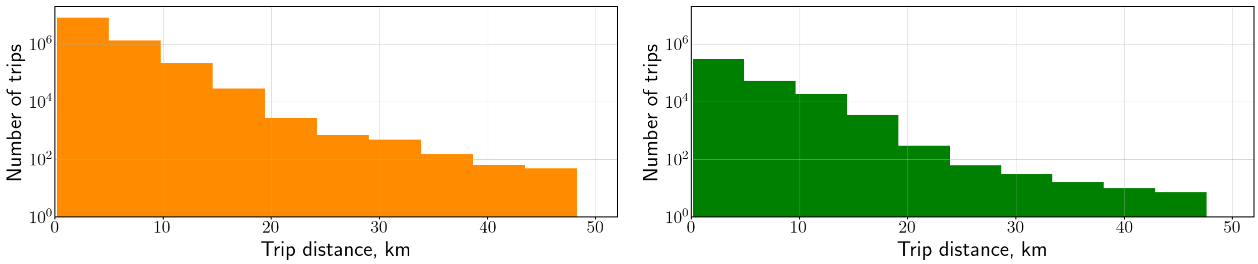
- Capacitated. There are N buses of maximal capacity of C passengers each.
- Time windows. Each passenger has to be picked up within $[-2\,\text{min}, 5\,\text{min}]$ of the recorded pickup time in the trip data (§2.1). The dropoff time window extends to 10 min after the recorded dropoff. To ensure feasibility, a passenger may be ignored at a certain penalty to the optimization objective. The buses are allowed to wait up to 10 min at any location.
- Depot. All buses start and finish at a certain location but have enough time to reach anywhere without compromising feasibility.

We use `ortools`[3] to find reasonable solutions computationally (on modest hardware). We can roughly assess optimality by allotting more time to the solver.

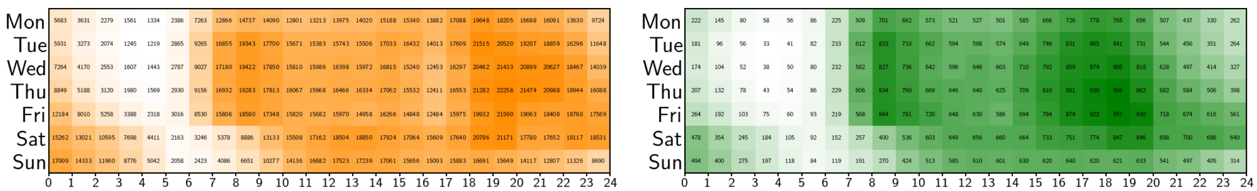
TODO(1): what is the cost TODO(2): initial solution

We focus on a small slice of the trip data at a time, i.e. a few hundred passengers × 1 h × a few square km. We then compare "customer satisfaction" across different fleet sizes N and bus capacities C.

We pretend here that the demand and the traffic conditions are known in advance, whereas some 10 min in advance would be more realistic. Meanwhile, the density of requests in space and time is quite high. Thus, we believe our results remain informative.

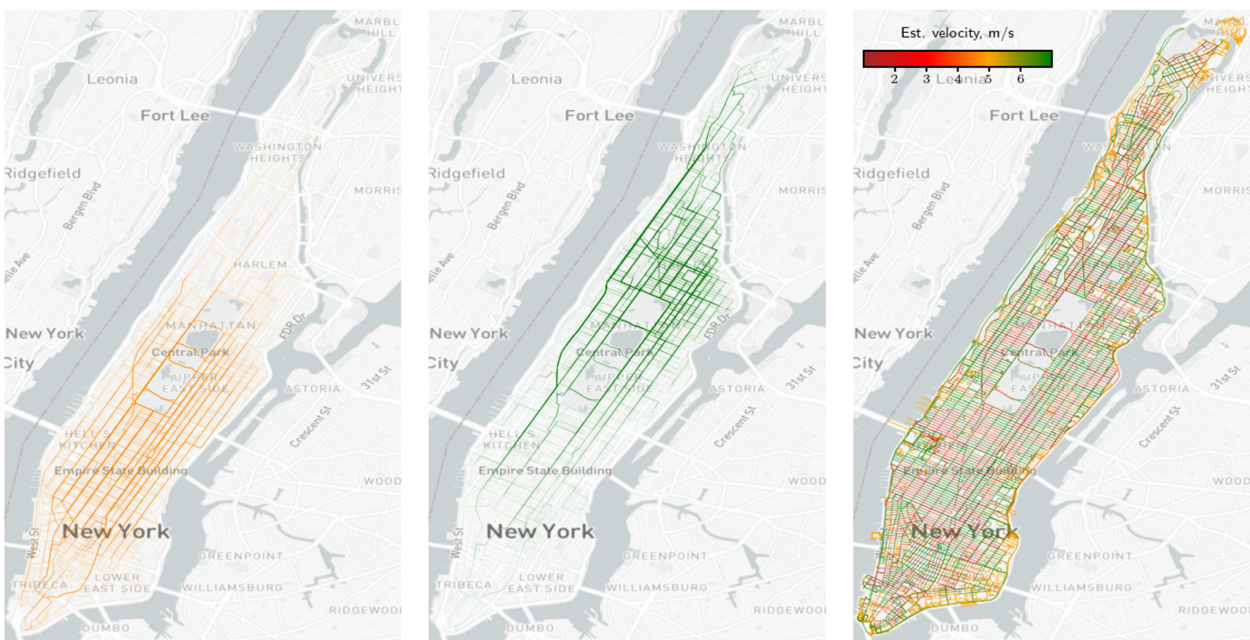


(a) Histogram of the reported trip distance (between 0.1 and 30 miles).



(b) Pickup hour heatmap.

Figure 1: Summary of Yellow (↖) and Green (↗) taxi trips filtered as in §2.1. #2



(a) Sample Yellow taxi trips. (b) Sample Green taxi trips. (c) Inferred mean-field velocity.

Figure 2: A sample of shortest-path trajectories (§2.1/§2.2) and the traffic model from §2.3. #7, #8

4 Case study #9

4.1 Times Square

We take the first n single-passenger trips with reported pickup and dropoff within 1 km of Times Square and within 18:00–19:00 on May 1, 2016. Allowing $n = 400$ requests is already a little difficult to solve ($\sim$1 h $\times$ 2 Gflop/s), so we focus here on

$$n = 100 \text{ requests for } N = 10 \text{ vehicles of capacity } C = 1 \text{ or } C = 8 \ .$$

The traffic model from §2.3 predicts trip times near Times Square well for 6-7 am but is too optimistic for 6-7 pm; we use it here nevertheless. #10

We find that the single-passenger fleet ($C = 1$) can only service about a half of requests, whilst the minibus fleet ($C = 8$) can handle most requests: see Fig. 4. The results for 2 km radius with $N = 20$ vehicles are similar (see code #9 to browse the results).

The minibus fleet runs at about half the capacity on average, see Fig. 5. Since all $n = 100$ pickup requests are from the first 17 min, it would be able to handle more requests per minute when operating over an extended period.

In summary, we estimate that this demand can be serviced by

- $N \approx 20$ single-passenger taxis; or
- $N \approx 10$ minibuses of capacity $C \approx 8$, if some excess travel time is tolerated.

TODO(3): more on this?

4.2 Game-theoretic spin

Consider a passenger (ignoring groups) who chooses between: (A) a single-passenger taxi at cost A that takes the quickest route, or (B) a minibus at cost B that occasionally detours for others. The costs are assumed fixed, since the trips in §4.1 are relatively short. The taxi fare is $A \approx 6\,\$$ for short trips (code #11).

Between 6-7 am and 6-7 pm, the number of requests doubles (Fig. 1b) while trip speeds near Times Square are nearly $3\times$ lower (code #10 and #12). We postulate a causal link and assume the traffic model of §2.3 when all passengers take (B); assume 1/3 the speeds when all passengers take (A); and interpolate linearly inbetween.

We take the $n = 100$ requests from §4.1 and split them randomly into a that take (A) and b that take (B). For simplicity, we keep the minibus fleet at $N = 10$ regardless of a/b.

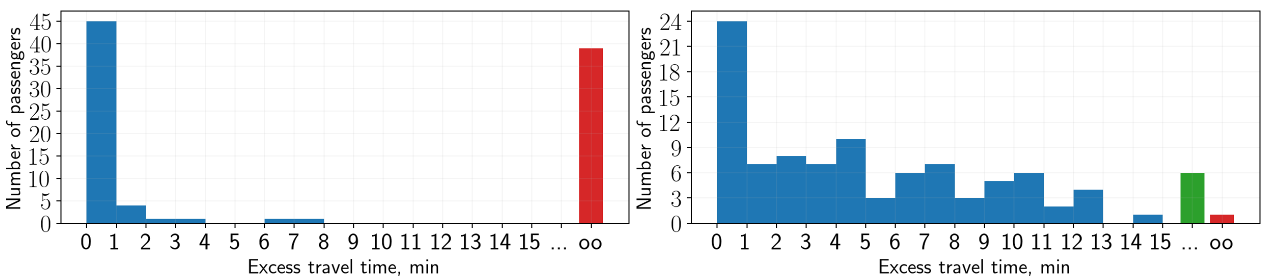
We assume passengers convert excess travel time to dollars. As a proxy for this, we start with the 2019 income census[4], to which the log-normal distribution with $\mu = 11$ and $\sigma = 0.7$

[3] https://developers.google.com/optimization/routing/vrp
[4] https://data.census.gov/cedsci/table?g=1600000US3651000&tid=ACSST5Y2019.S2001

(a) Vehicle capacity $C = 1$.

(b) Vehicle capacity $C = 8$.

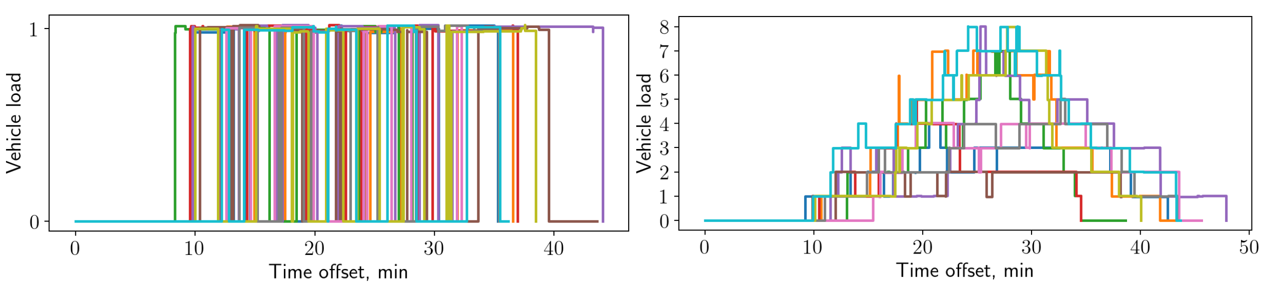
Figure 3: Passenger trajectories colored by excess travel time w.r.t. the quickest route ($n = 100$ requests for $N = 10$ vehicles). Empty red circles are unserviced pickup requests. Cf. §4.1.



(a) Vehicle capacity $C = 1$.

(b) Vehicle capacity $C = 8$.

Figure 4: Histogram of excess travel time w.r.t. the quickest route ($n = 100$ requests for $N = 10$ vehicles). The last bar shows unserviced requests. See §4.1.



(a) Vehicle capacity $C = 1$.

(b) Vehicle capacity $C = 8$.

Figure 5: Load of each vehicle over time ($n = 100$ requests for $N = 10$ vehicles). The time begins at ∼17:50 when the fleet is released from the fictitious depot at Times Square. See §4.1.

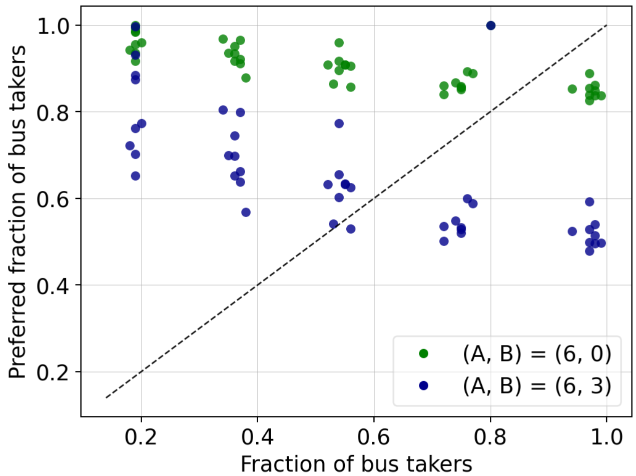


Figure 6: Results of §4.2.
TODO(4): caption

fits reasonably well (see #13); a passenger's income is drawn from this log-normal; division by $52 \times 5 \times 8\text{h}$ gives their hour-to-dollar conversion factor c. #14

For each condition a/b we compute the expected excess travel time e for (B) passengers. Comparing A vs. $B + (e \times c)$ for each c from the income distribution gives a ratio $\bar{a}/\bar{b}$ of those who actually prefer (A) to those who prefer (B). This evolutionary pressure indicates whether (A) or (B) should increase starting from the condition a/b.

The results (Fig. 6) indicate that

- half the people prefer taking the minibus if $A - B = 3\,\$$; and
- most people prefer taking the minibus if $A - B = 6\,\$$.

The conspicuous outlier at $(b, \bar{b}) \approx (0.8, 1)$ is from a solution that is qualitatively different from others.

TODO(5): code refs

5 Conclusions

TODO(6): drop the mic

6 Appendix

6.1 List of codes

	page	https://github.com/numpde/optimum/blob/main/ ...
#1	p.1	code/data/20210610-NYCTLC/a_download.py
#2	p.3	code/data/20210610-NYCTLC/e_explore.py
#3	p.1	code/data/20210611-OSM/a_osm_download.py
#4	p.1	code/data/20210611-OSM/c_road_graph.py
#5	p.2	code/helpers/opt_maps/maps.py
#6	p.2	code/model/20210613-GraphWithLag/b_train.py
#7	p.3	code/data/20210611-OSM/e_explore.py
#8	p.3	code/model/20210613-GraphWithLag/b_train.py
#9	p.4	code/work/20210616-OPT1
#10	p.4	code/model/20210613-GraphWithLag/c_triptime_times_square
#11	p.4	code/data/20210610-NYCTLC/e_explore/trip_fare_vs_distance
#12	p.4	code/data/20210610-NYCTLC/e_explore/trip_speeds_times_square
#13	p.6	code/data/20210621-Income
#14	p.6	code/work/20210616-OPT1/d_postprocess.py

6.2 TODOs

1. p.2. what is the cost
2. p.2. initial solution
3. p.4. more on this?
4. p.6. caption
5. p.6. code refs
6. p.6. drop the mic

2021-06-23 16:05:50+02:00