Metro Math

This presentation at:

kwkelly.com/pres/metro-pres/

Keith Kelly

kwkelly.com

mail@kwkelly.com

@KeithKelly

kwkelly

A Typical Day

You're on your way to work, heading home, or going to meet some friends...

... and you walk into the station and see this:

Waiting for WMATA

Fact:

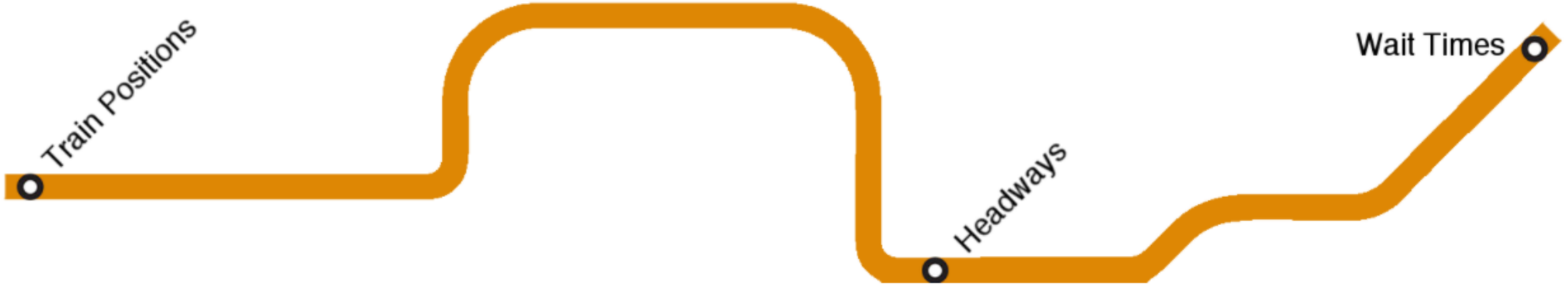
- Riders hate waiting for trains; waiting time is regarded as wasted time.

Question:

- How long do Metrorail riders typically wait for their trains?

Analysis Map

Train Positions → Train Headways → Passenger Waiting Times



Train Positions

- Old rail predictions API 👎
  - Tough to reliably determine if it's a new train at a station
- Newer train positions API 👍
  - CircuitId ← Standard Routes API to determine if circuit is at a station
  - DestinationStationCode
  - DirectionNum
  - LineCode
  - TrainId

Getting Headways

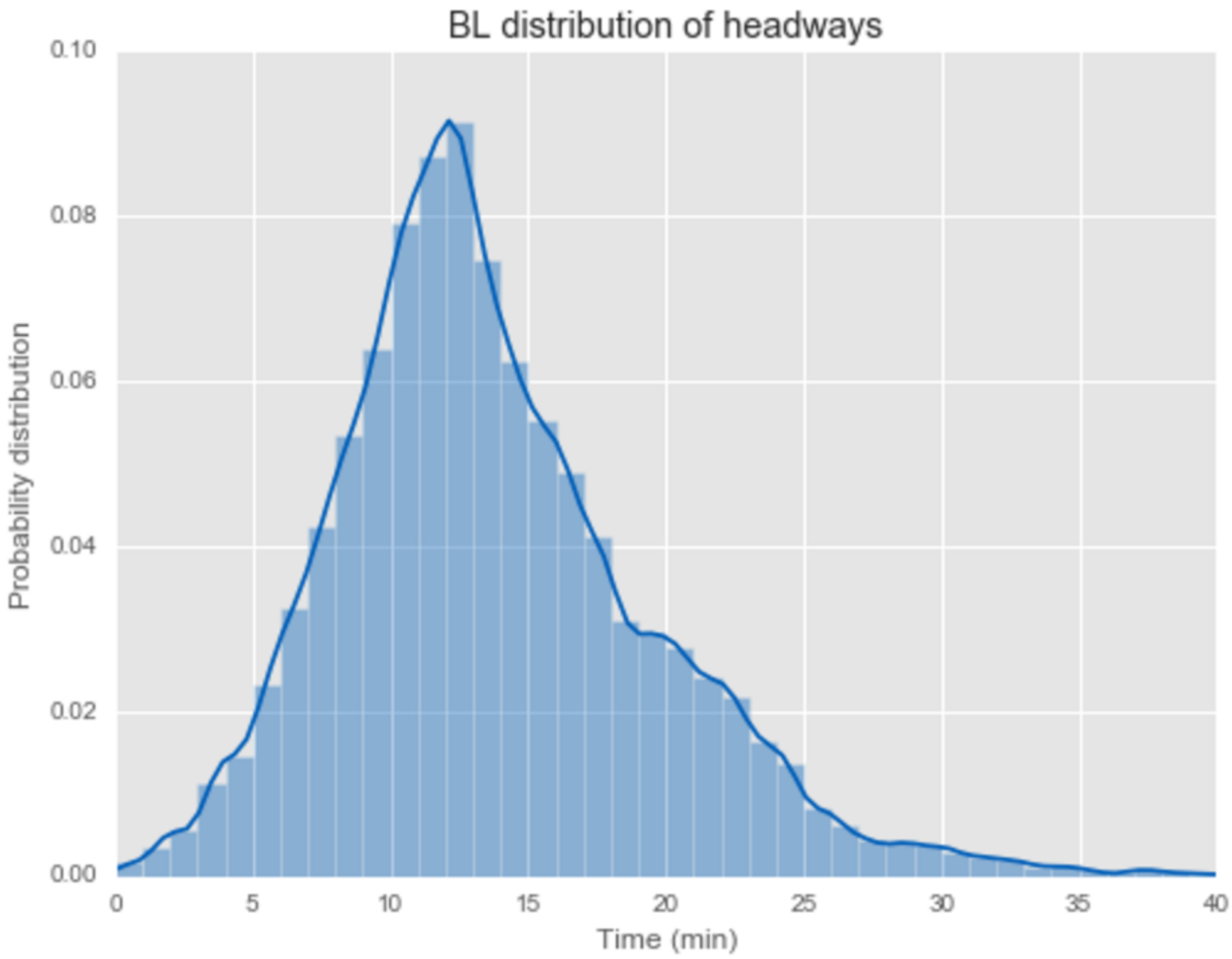
Position data + route data + query time → headways

- Keep data in chronological order
- Simply ignore all position data where a train is not at a station
- Then for a given station, line, and direction, compute the time difference between each row of data

```
# df = all data
df = df[df['StationCode'] == station]
df = df[df['LineCode'] == line_code]
df = df[df['DirectionNum'] == dir_num]
head = np.diff(df['DateTime'])
```

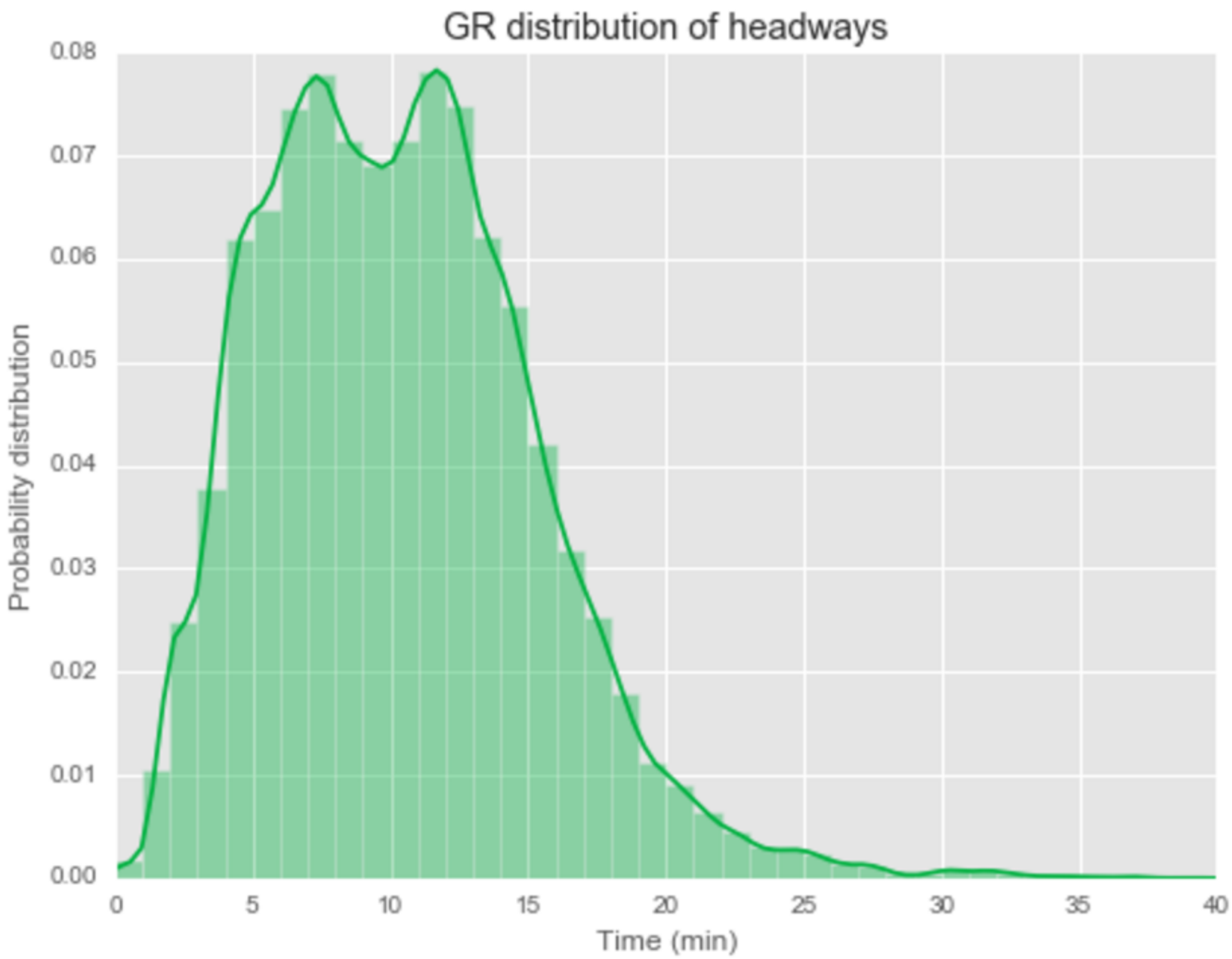
Getting Headways

BL GR RD OR YL SV



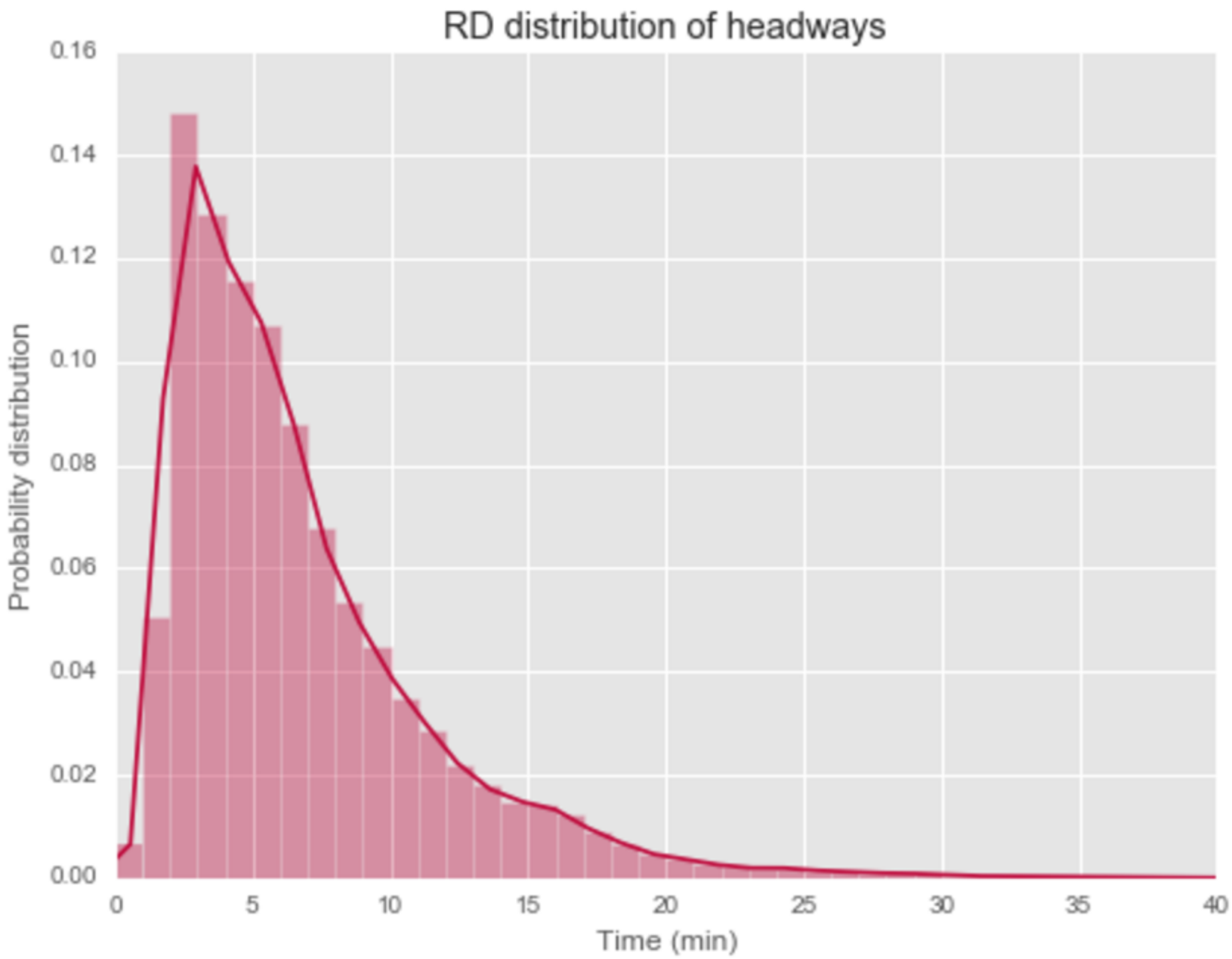
Getting Headways

BL GR RD OR YL SV



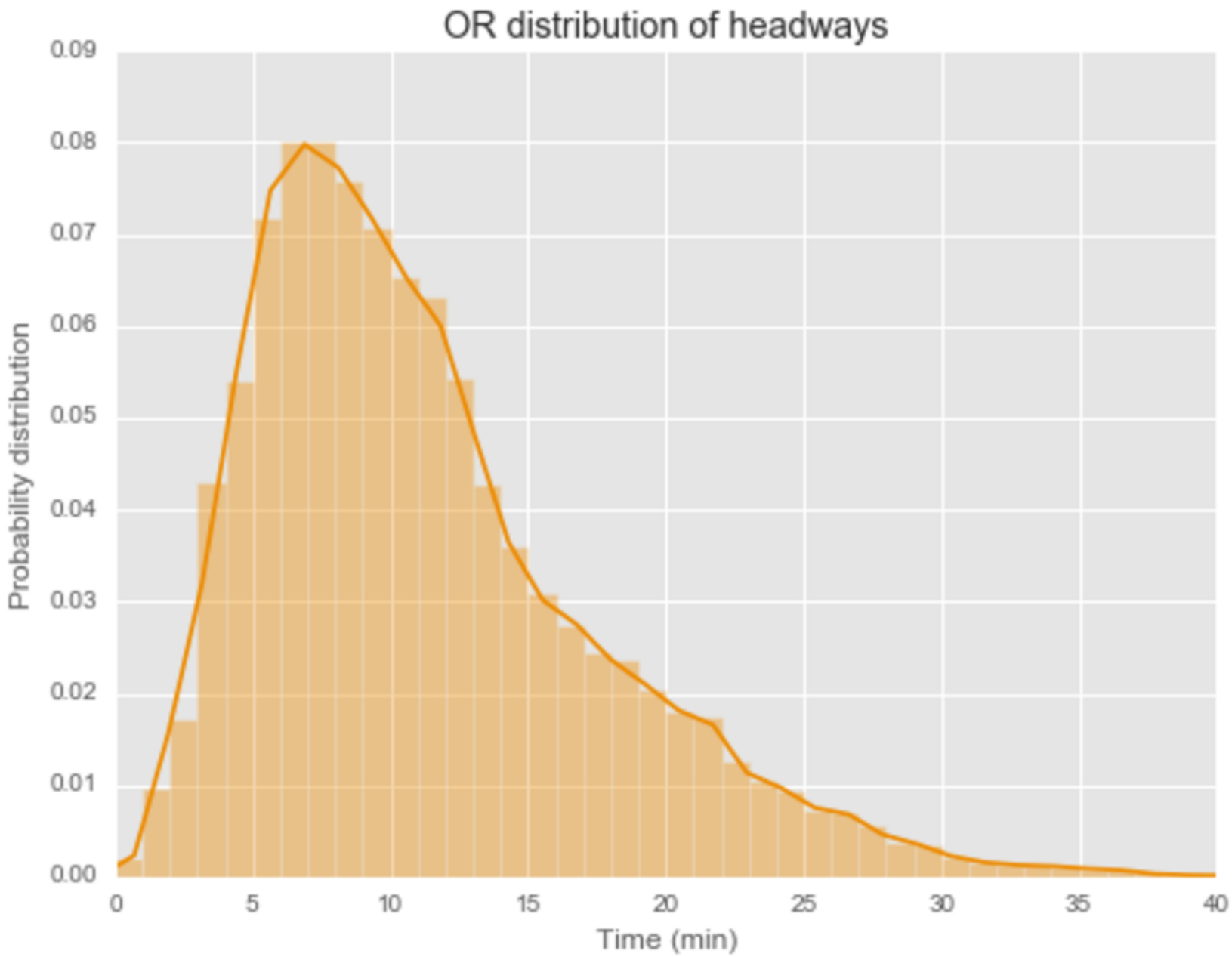
Getting Headways

BL GR RD OR YL SV



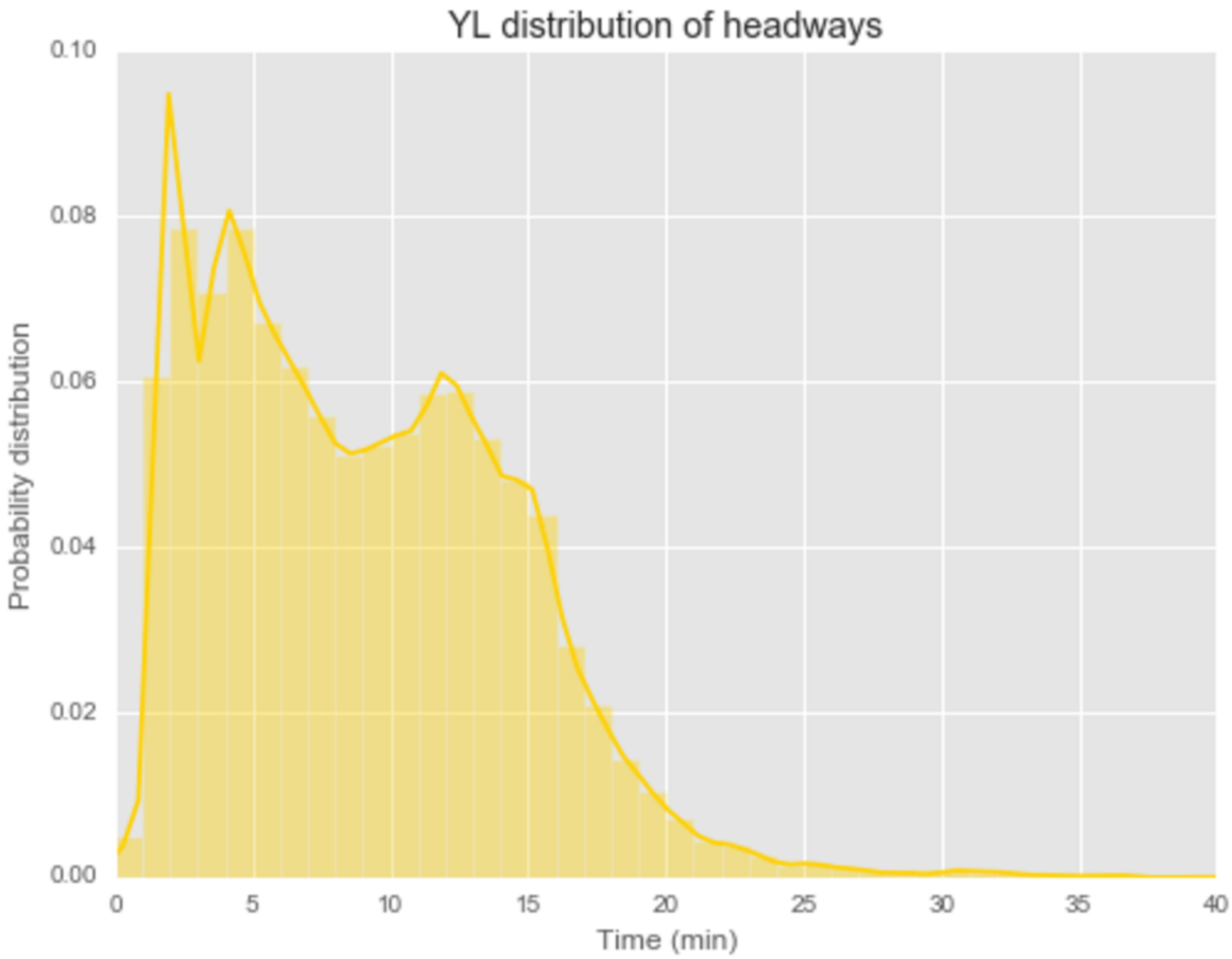
Getting Headways

BL GR RD OR YL SV



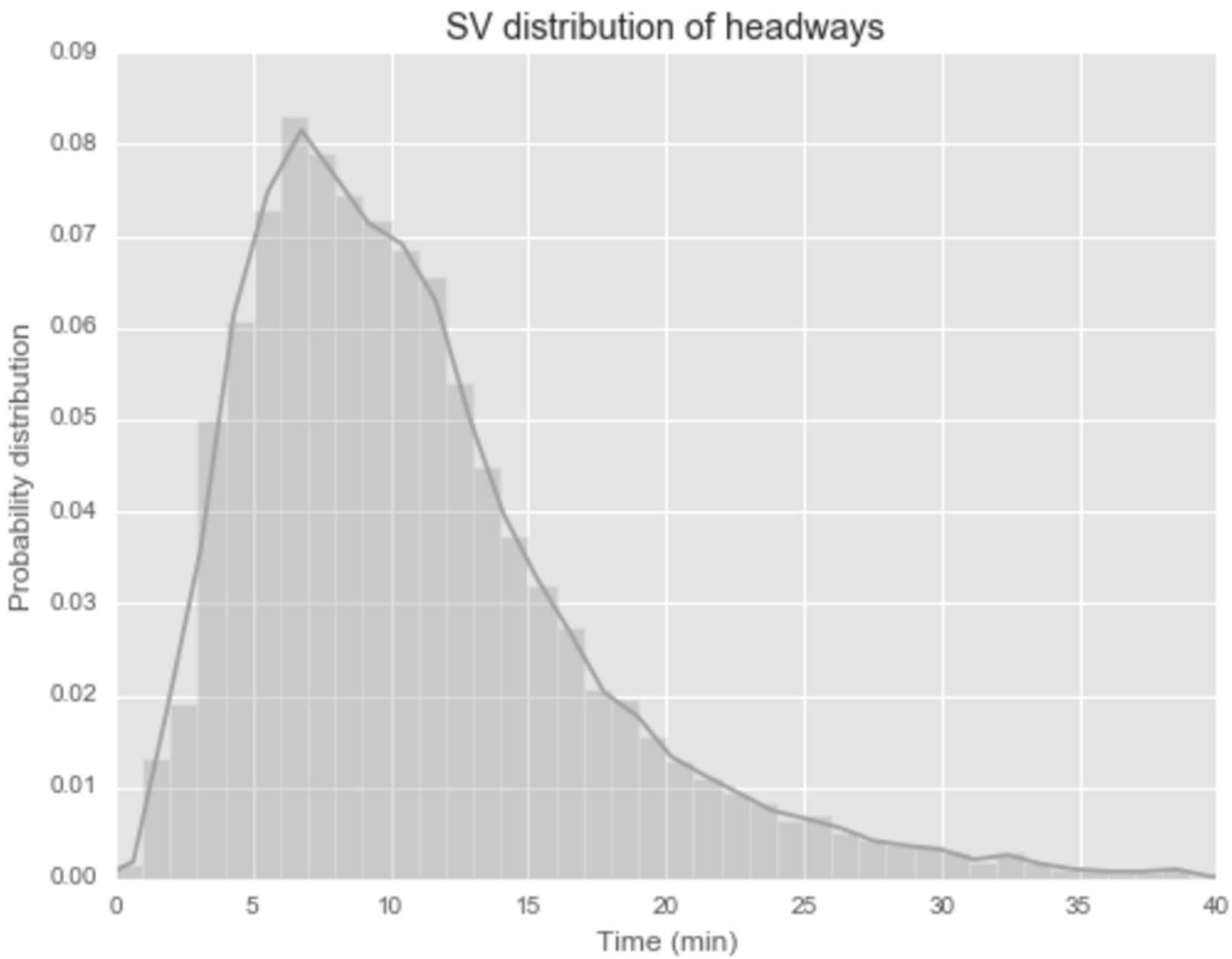
Getting Headways

BL GR RD OR YL SV



Getting Headways

BL GR RD OR YL SV



Estimating Wait Times

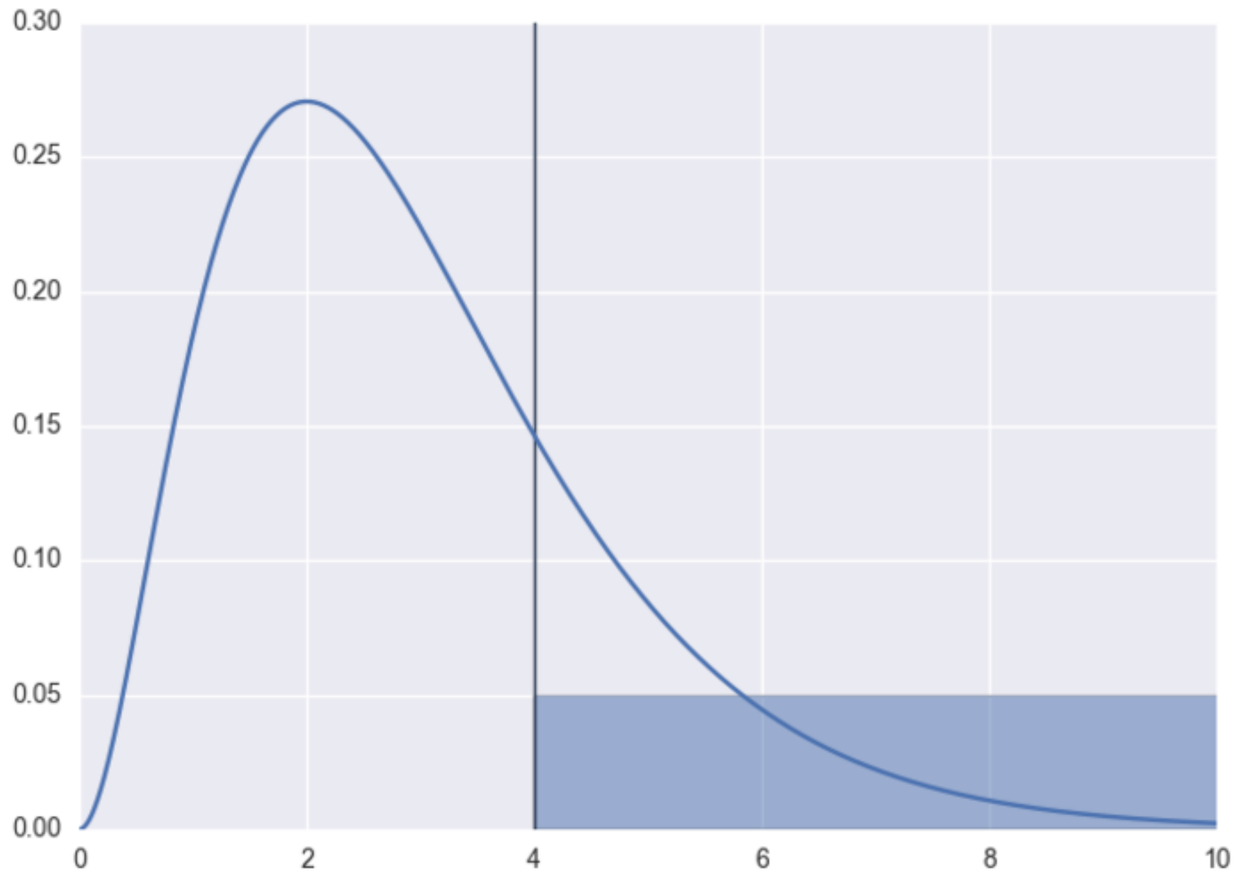
- Headways are about trains, waiting is about people
- We have data about trains, but not people

Assumptions

- No strategic behavior of passengers w.r.t. common lines
- No congestion phenomena
- Passenger arrivals are uniformly distributed

Estimating Wait Times

Since passenger arrivals are assumed to be uniformly distributed, the probability of waiting an amount of time t is proportional to the probability of a headway greater than t.



Estimating Wait Times

Let $h(t)$ be the distribution of headways. Then wait times $w(t)$ can be computed as

$$w(t) \propto \int_t^\infty h(s)\,ds = \bar{H}(t)$$

Finding the constant of proportionality is of course just normalizing the distribution, so that

$$w(t) = \frac{\int_t^\infty h(s)\,ds}{\int_0^\infty \bar{H}(s)\,ds} = \frac{\int_t^\infty h(s)\,ds}{\int_0^\infty s\,h(s)\,ds}$$

But we really don't care about the constant, just the proportion becuase we are using a computer and can normalize after the fact.

Estimating Wait Times

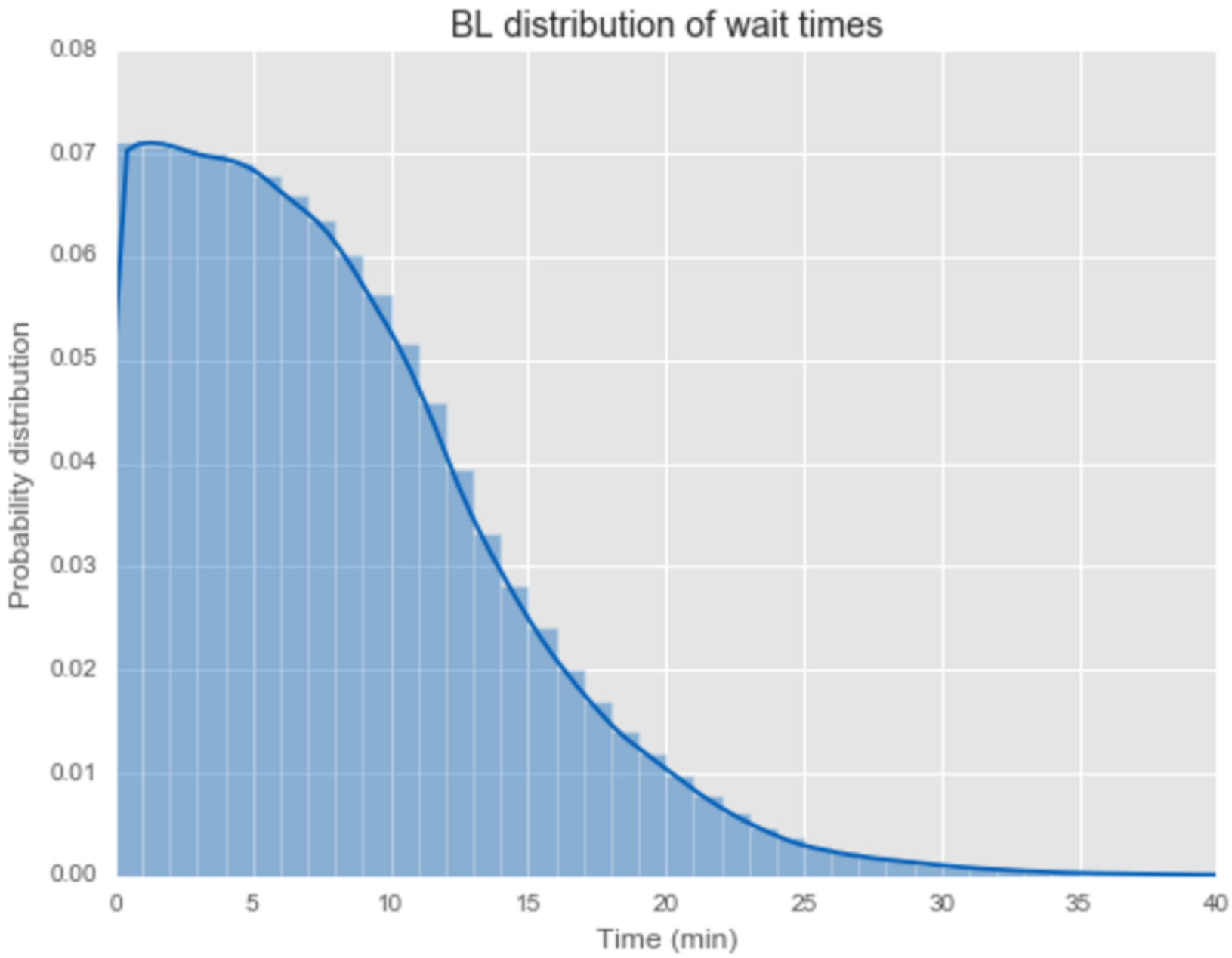
In practice though, we just sample from the headway distribution.

```
# diffs = all headways for a single line
waits = itertools.chain(*[np.arange(0, diff, 0.5).tolist() for diff in diffs])
```

Headway

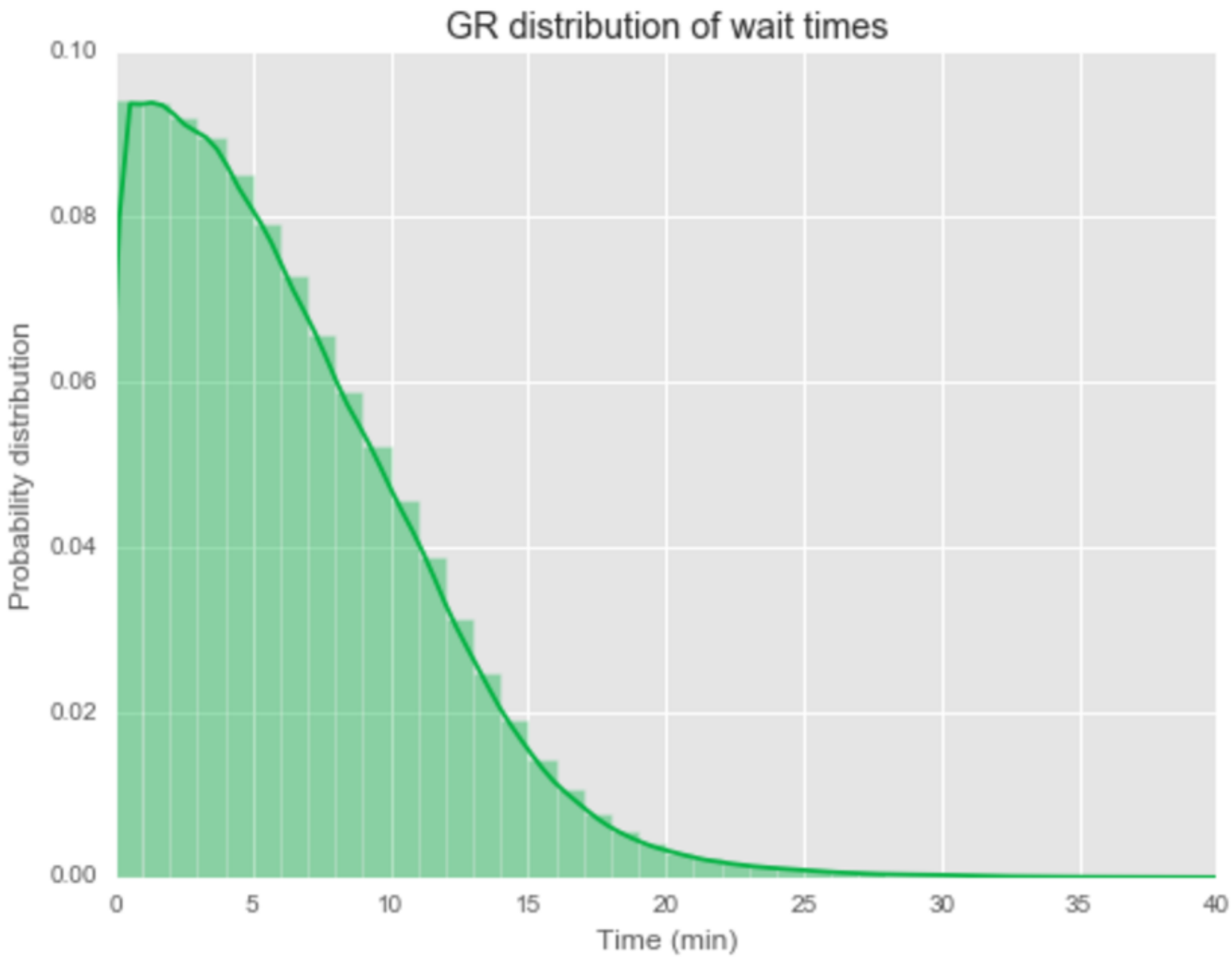
Estimating Wait Times

BL | GR | RD | OR | YL | SV



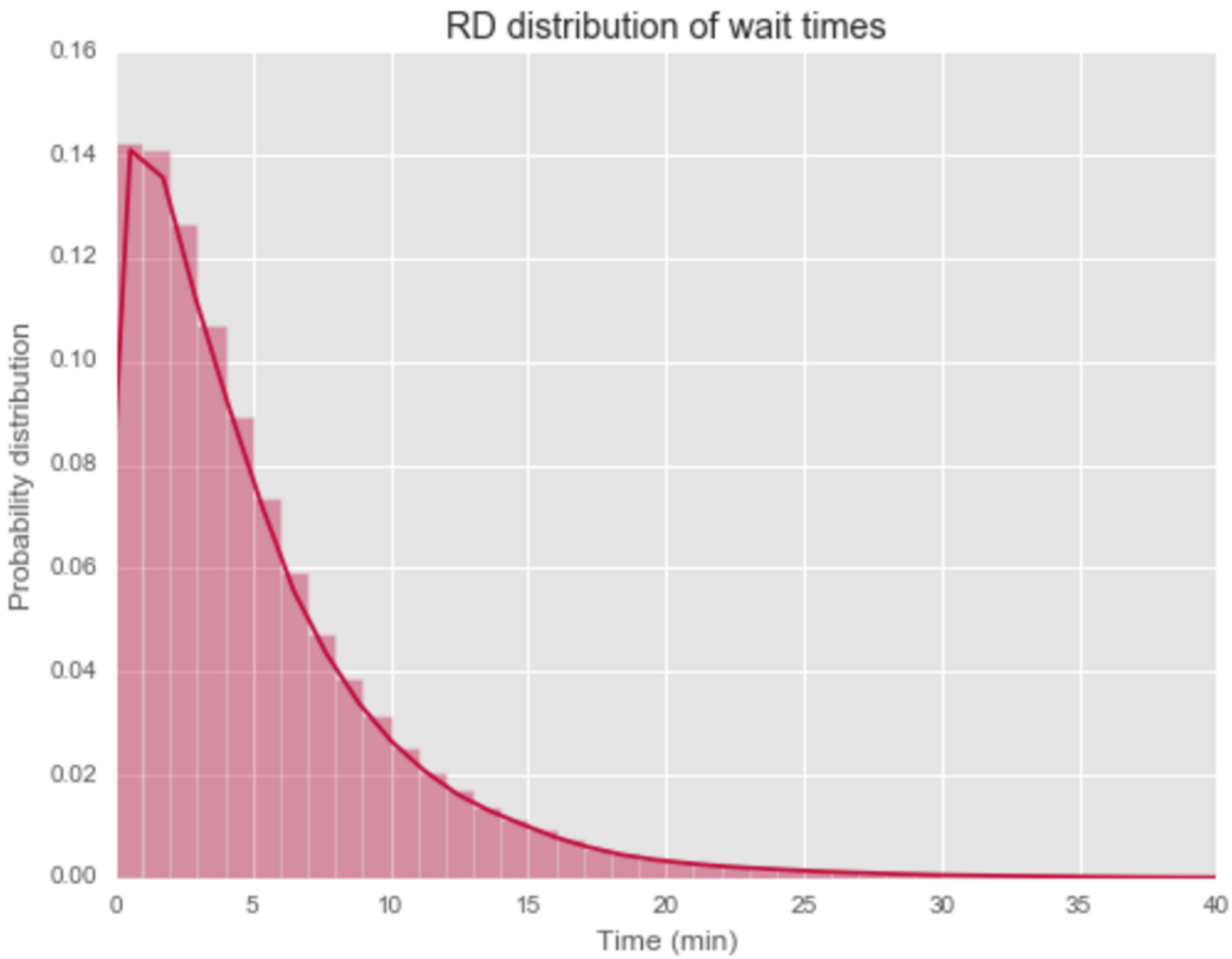
Estimating Wait Times

BL GR RD OR YL SV



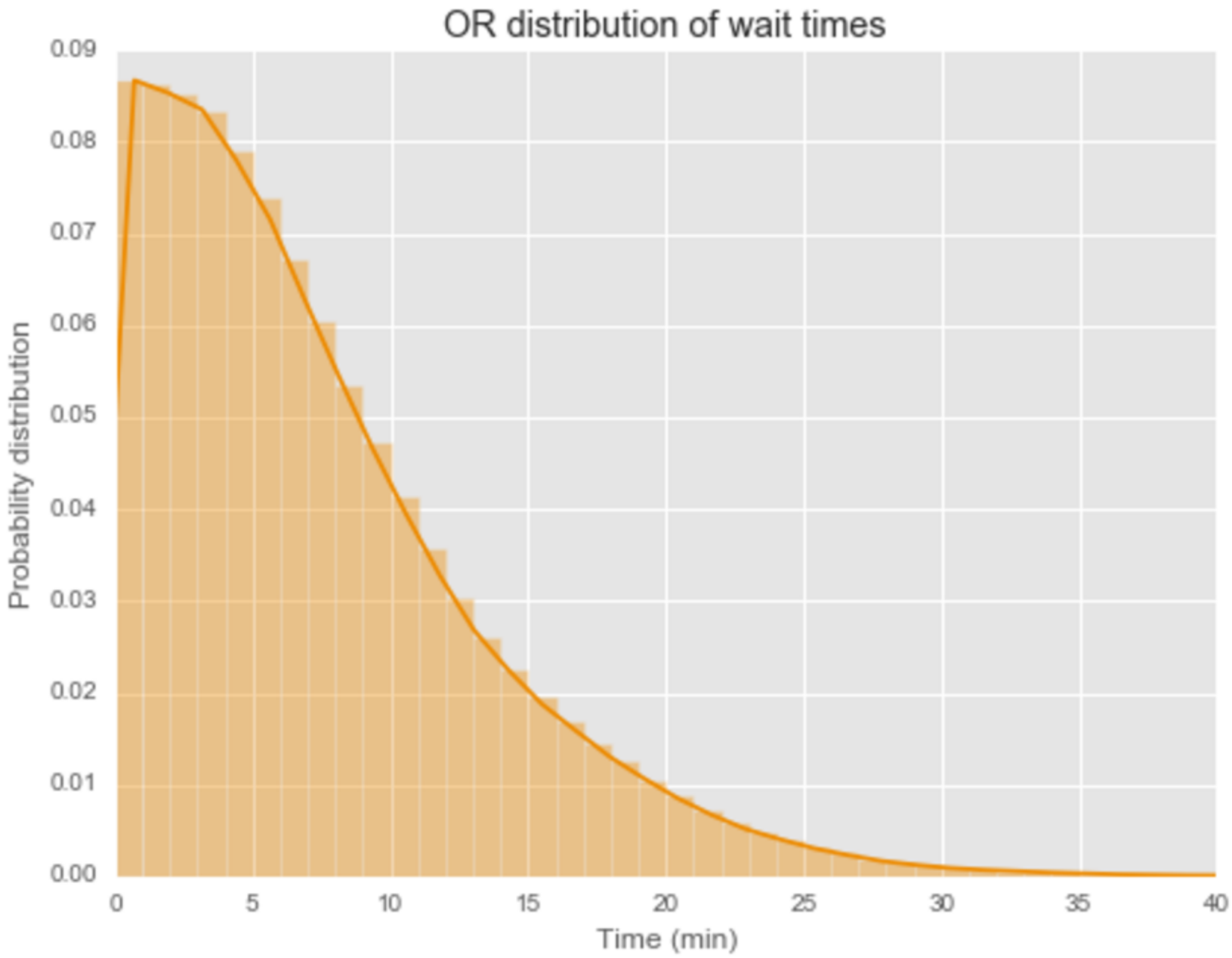
Estimating Wait Times

BL GR RD OR YL SV



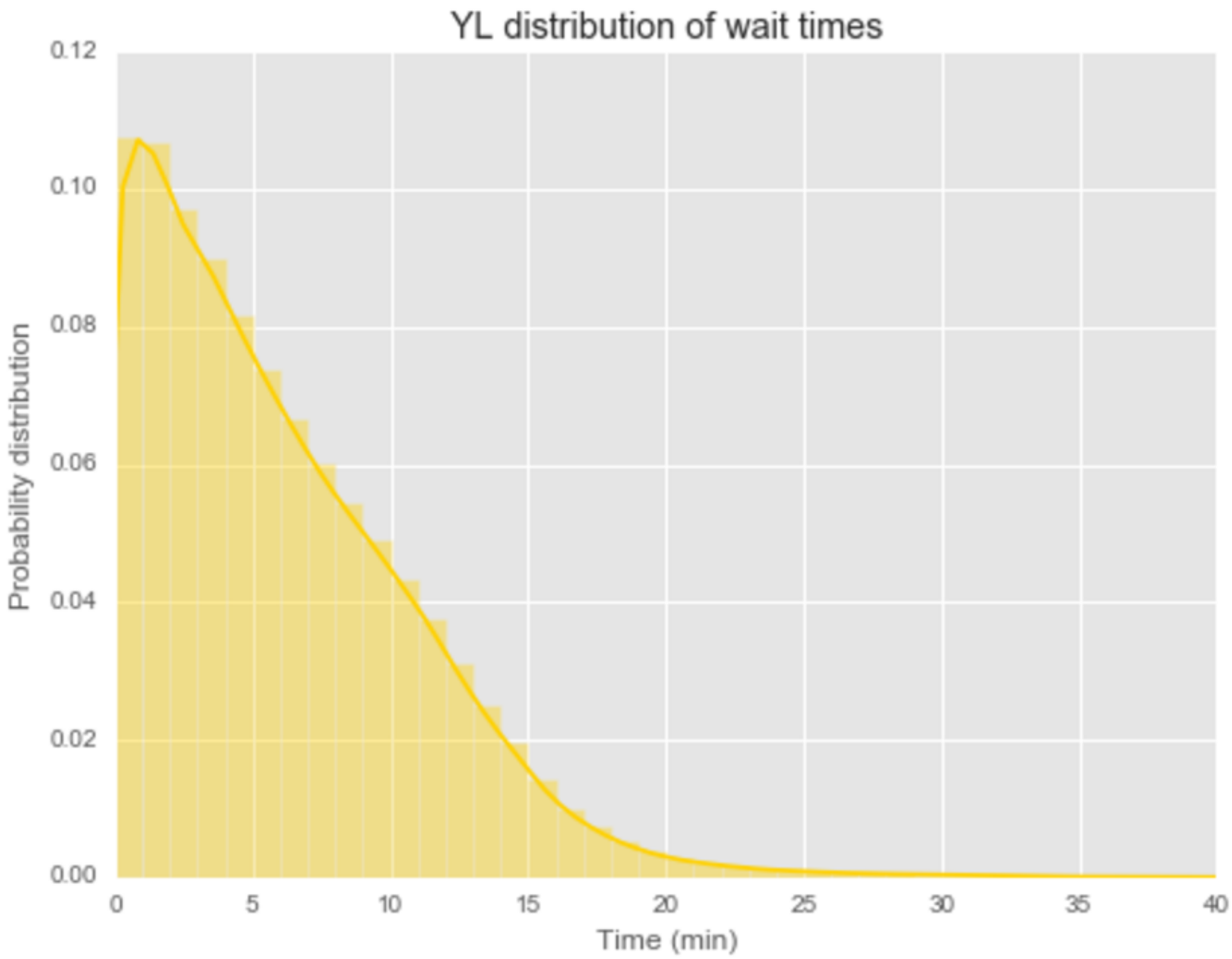
Estimating Wait Times

BL GR RD OR YL SV



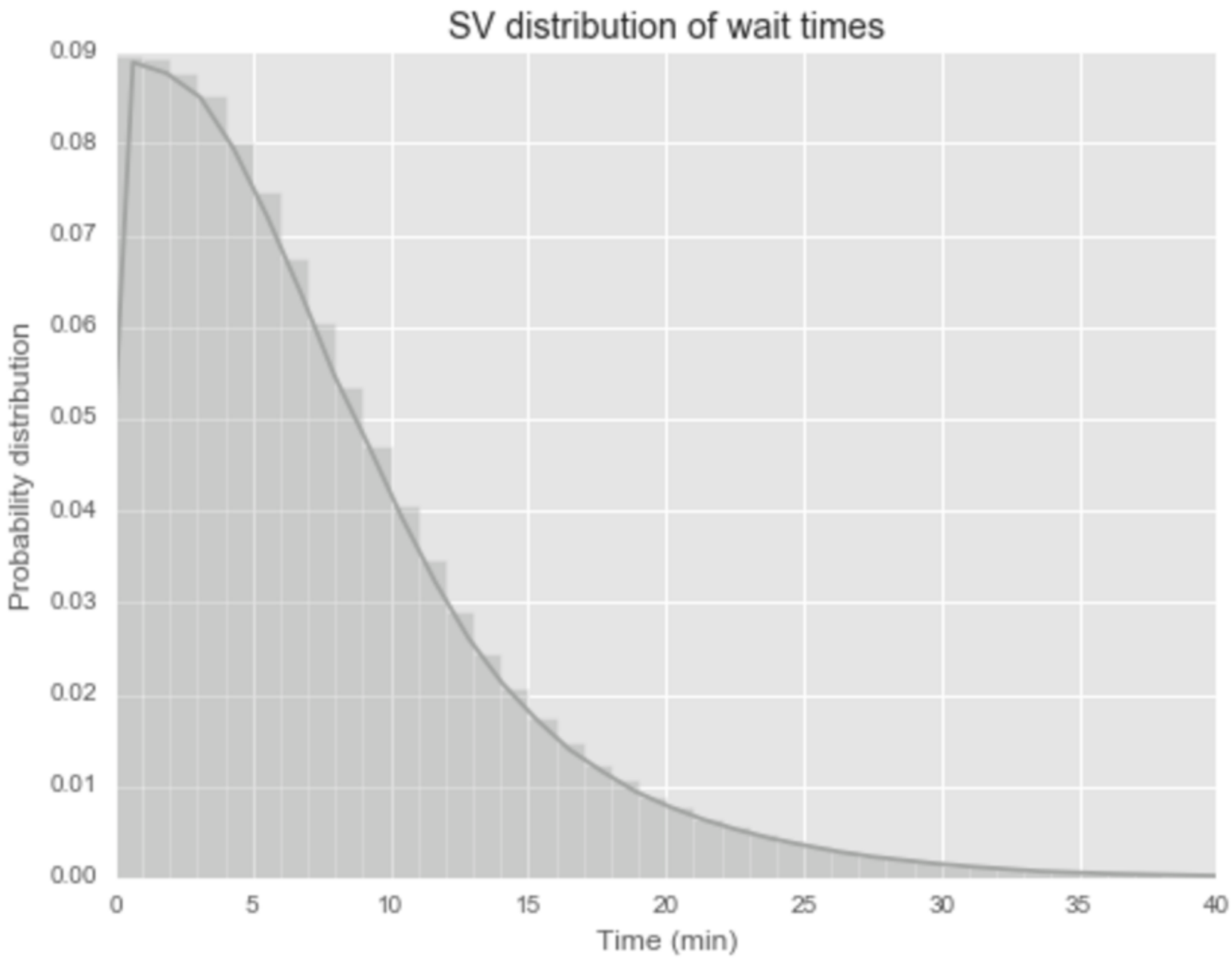
Estimating Wait Times

BL GR RD OR YL SV

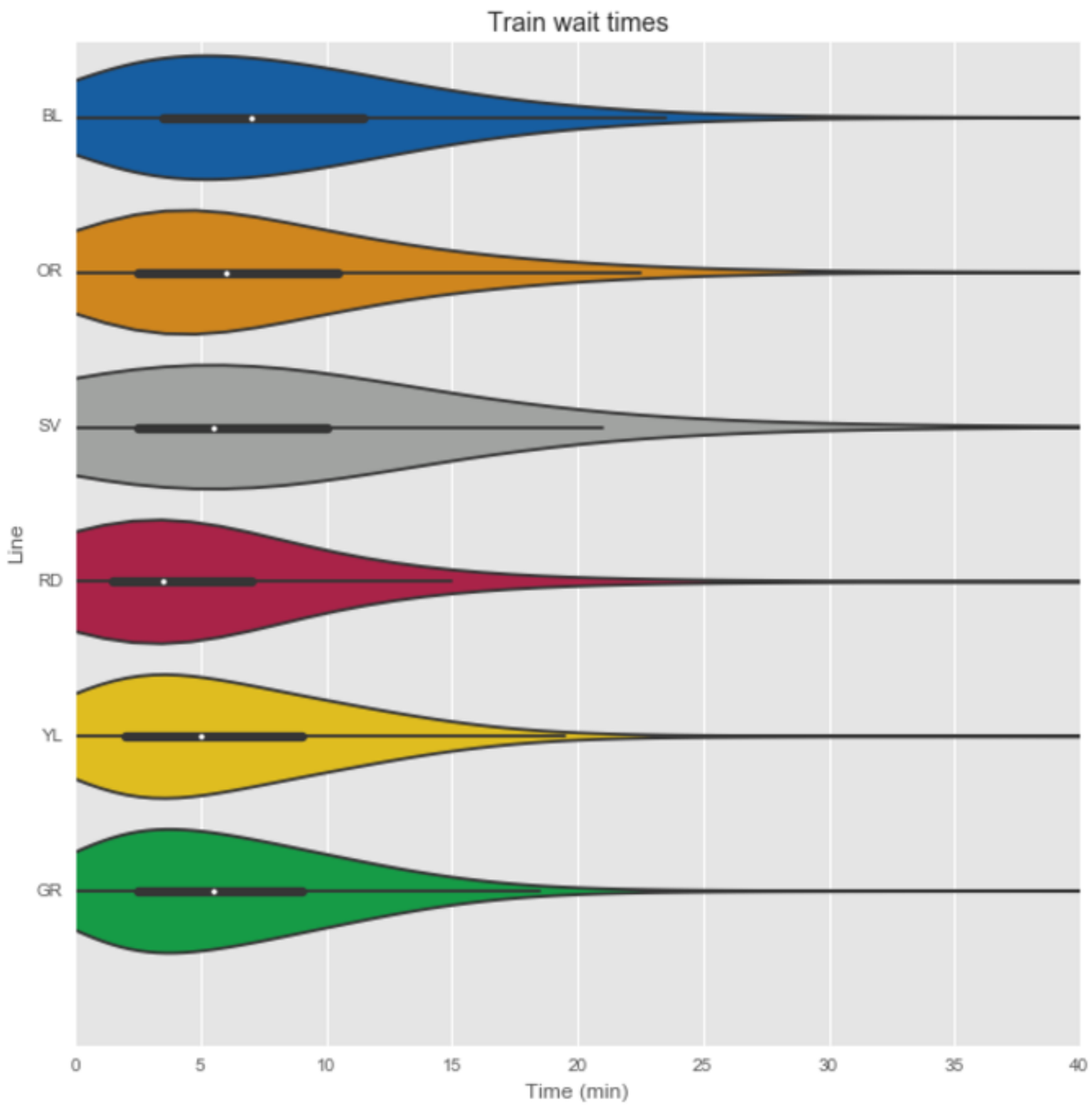


Estimating Wait Times

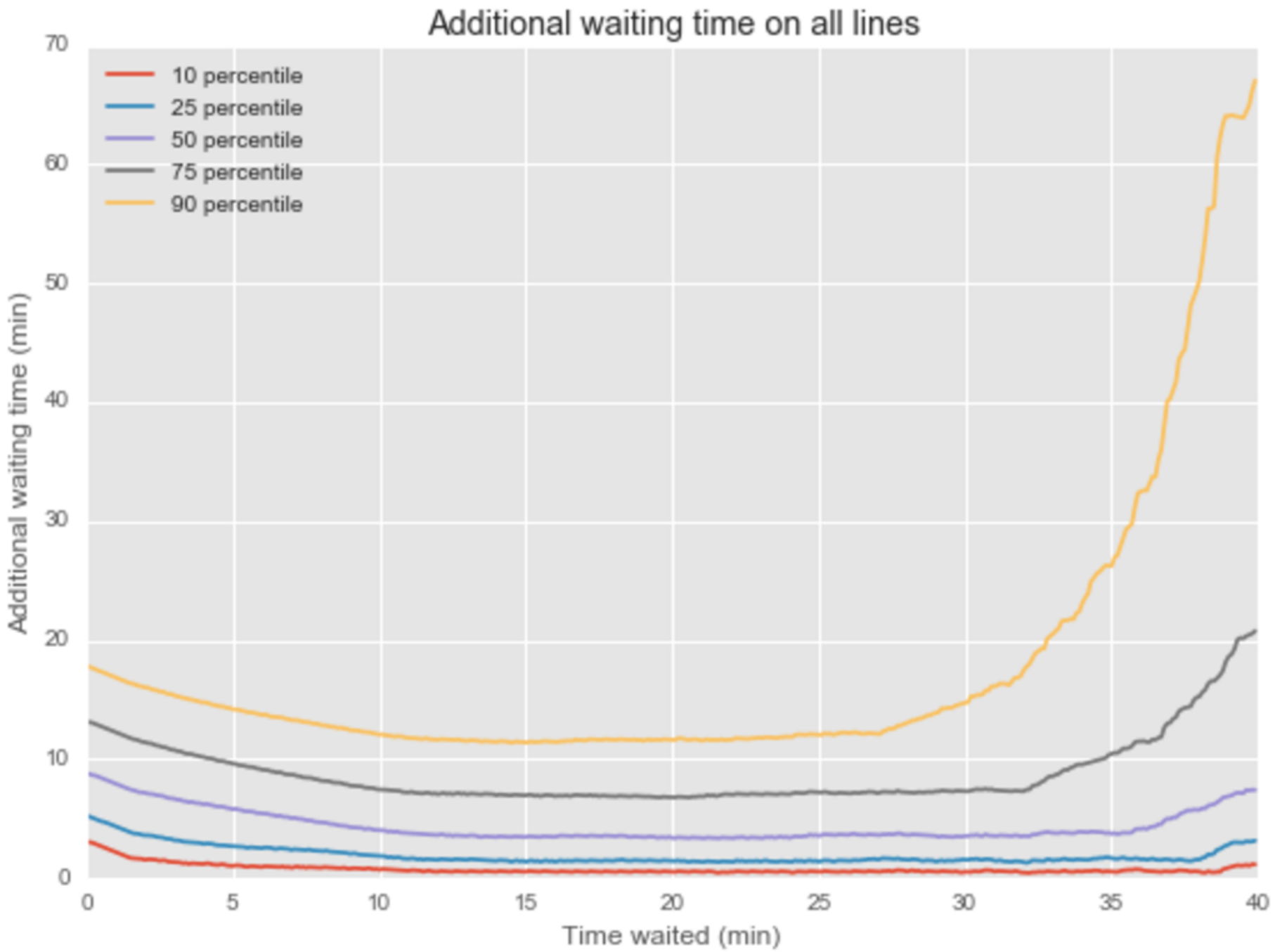
BL GR RD OR YL SV



Estimating Wait Times



Additional Wait Times



Thanks!