

# Methodology for calculating distributions

Joel Miller

July 20, 2020

The central challenge in finding the distribution of wait times for a given demographic is that rows in the original data do not list demographics— instead, it must be inferred from the block group of a request. This is how I estimated that data from the available tables.

## Notation:

- Let  $B$  denote the set of block groups.
- Let  $A$  be a table listing average response times for different requests by block group, such that  $A[b, r]$  gives the average response time for requests of type  $r$  sent out from block group  $b$ .
- Let  $D$  be a table listing demographics by block group, such that  $D[b, d]$  gives the number of people in demographic  $d$  who live in block group  $b$ .
- Let  $N$  be table listing the total number of requests that have come out of each block group, such that  $N[b, r]$  gives the total number of requests of type  $r$  made from block group  $b$ .

First, for a given request  $r$ , demographic  $d$ , and block group  $b$ , we can estimate the total number of requests made by people of demographic  $d$ , who also live in  $b$ , as

$$N[b, r] \times \frac{D[b, d]}{\sum_d D[b, d]}$$

Call this number  $x_{b,d}$ . Now, I estimate the collections of all wait times for members of a demographic  $d$  as

$$\bigcup_{b \in B} \{x_{b,d} \text{ records, each taking time } A[b, r]\}$$

Also note that  $x_{b,r}$  is rounded to the nearest whole number.

From there, I use pandas' `DataFrame.plot.kde` method on that data.