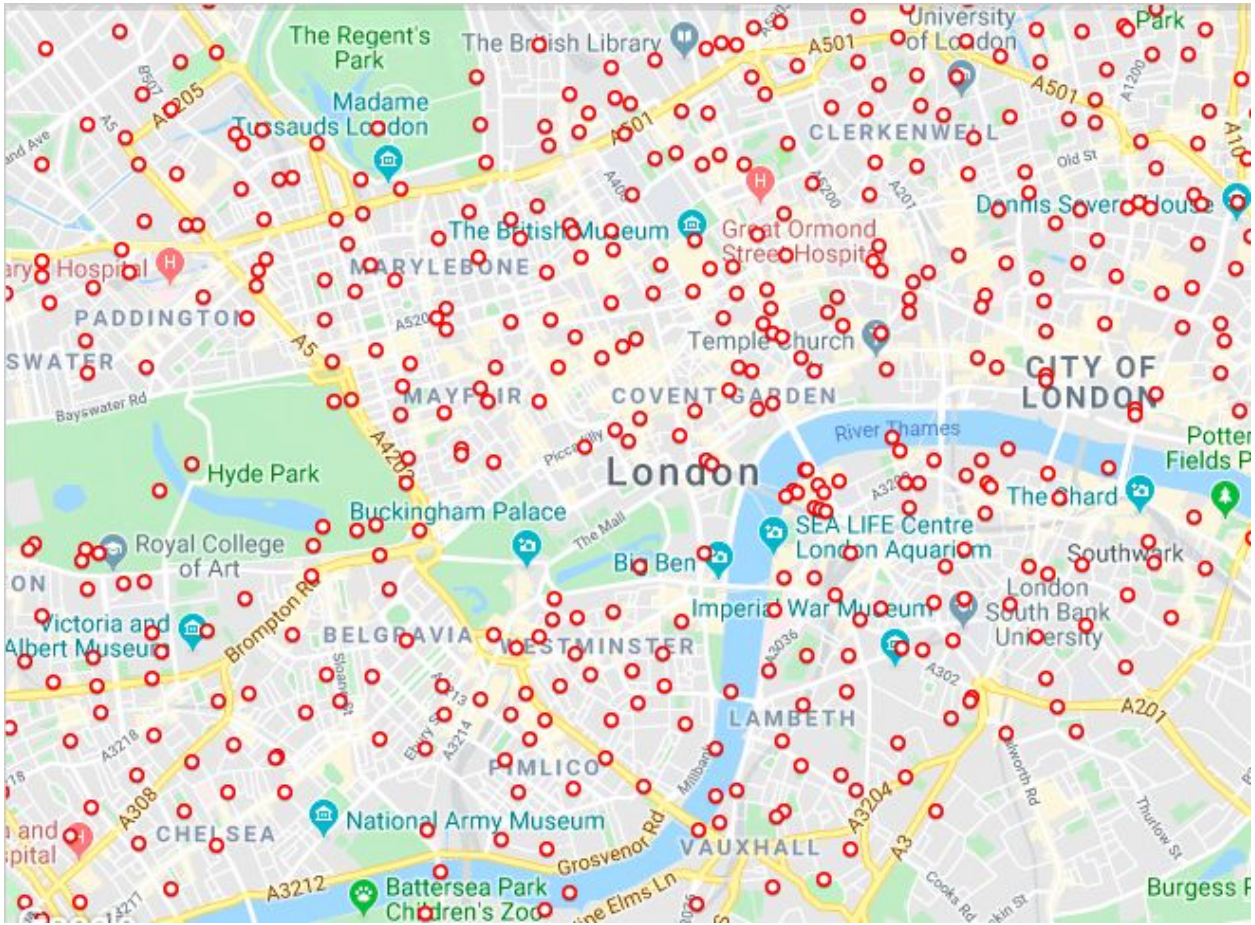
Mod 3 Project

Joe Read and Mohammed Hannan



TFL

Bikepoints

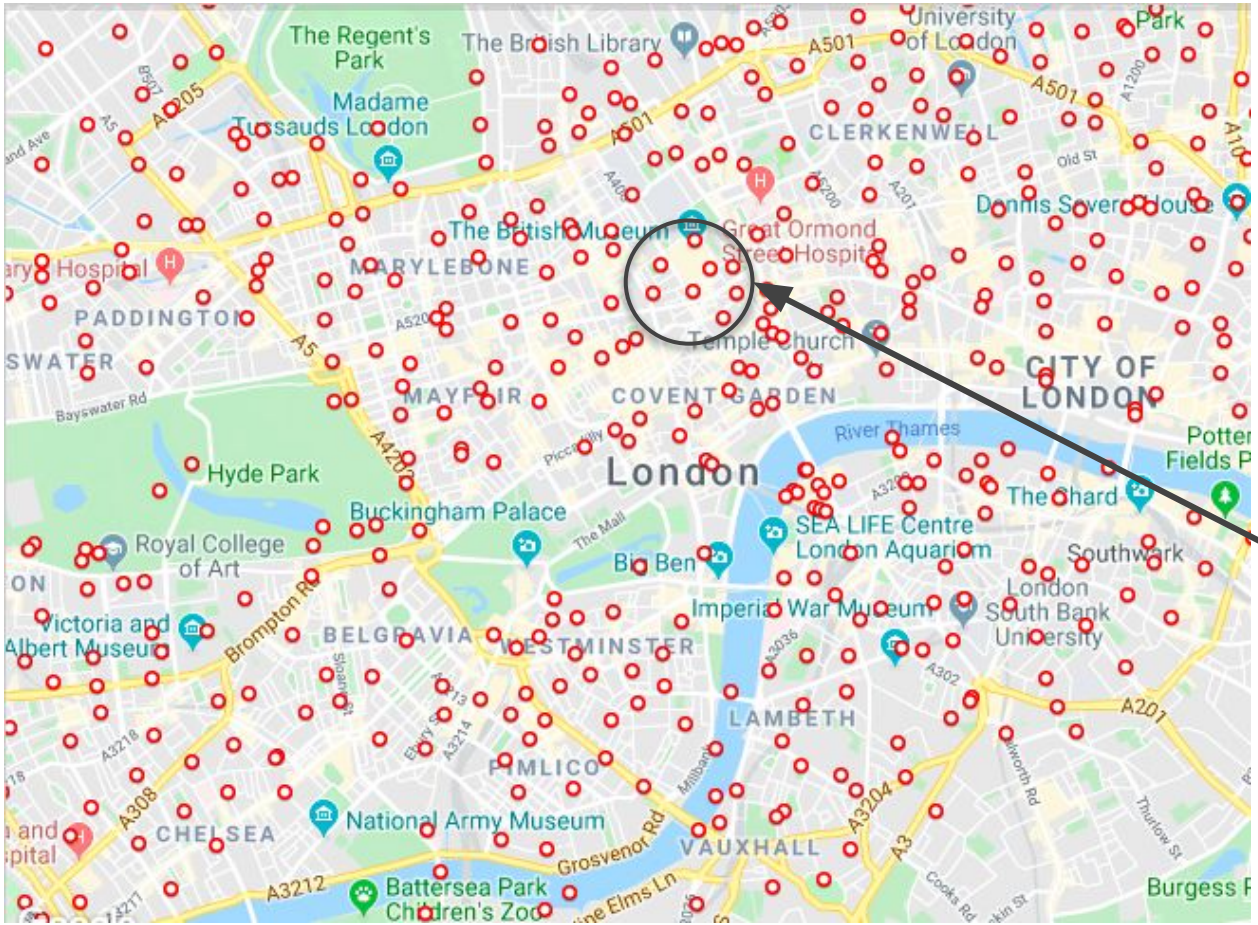
Hypotheses

1. A faulty bikepoints will be less likely to have other faulty bikepoints near it
2. Bikepoints at high elevation have fewer docks
3. Bikepoints in high populated areas are more likely to be faulty
4. High populated areas have larger bikepoints

Methodology

- Our hypotheses allowed the data to be split into boolean catorageries:

  faulty / non-faulty, hill / no-hill, high pop / non-high pop
- We suspected a small effect size so decided to sample 198 times
- Each sample was formed of 50 data points drawn at random
- The resulting t distribution was formed from the means of each sample



Local Features

To check local features we looked at all bikepoints within a certain radius

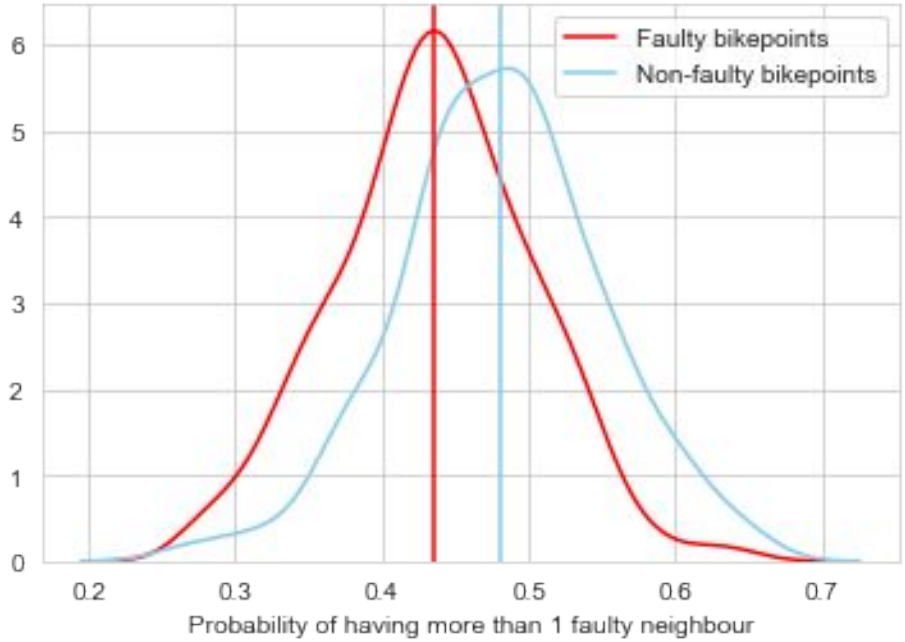
H0 = Faulty bikepoints have the same number of faulty bikepoints nearby as general bikepoints

H1 = Faulty bikepoints have more likely to have more than 1 faulty bikepoints near them than general bikepoints

Statistic: -6.41

Pvalue: 4.22 x 10^-10

Cohen’s d: -0.646



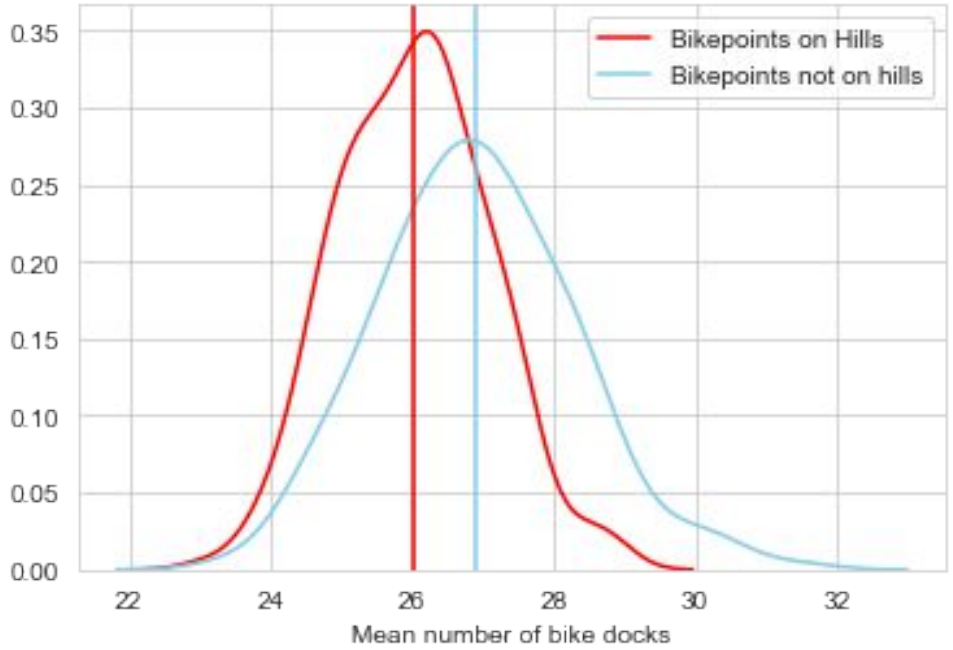
H0 = Bikepoints on hills have the same number of docks as general bikepoints

H1= Bikepoints on hills have fewer docks than general bikepoints

Statistic: -7.15

Pvalue: 4.66 x10^-12

Cohen’s d: -0.721



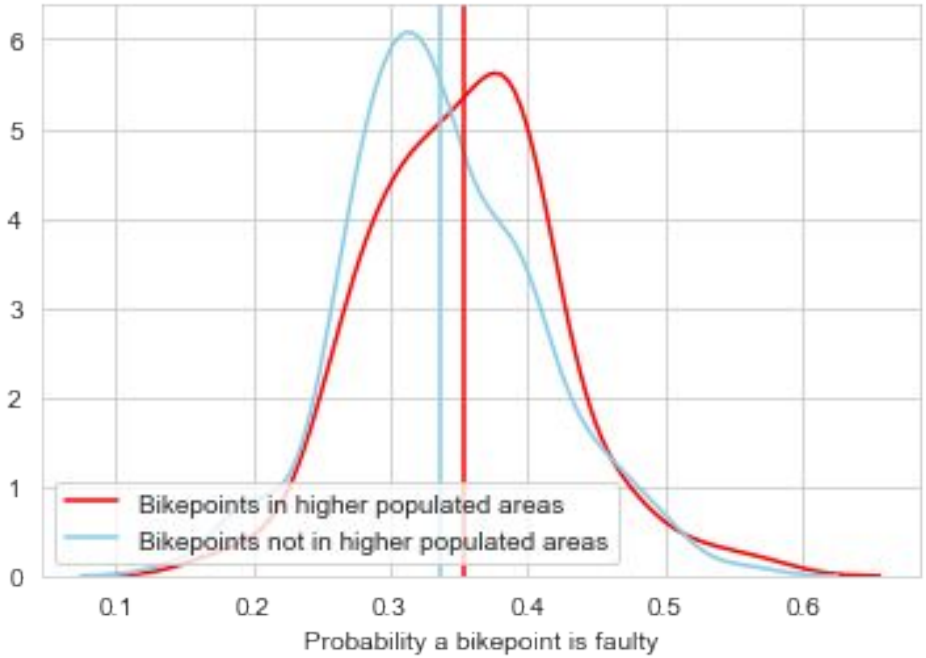
H0 = Bikepoints in highly populated areas have the same probability of having a faulty dock as general bikepoints

H1 = Bikepoints in highly populated areas have a higher probability of having a faulty dock

Statistic: 2.31

Pvalue: 2.14 x10^-2

Cohen’s d: 0.233



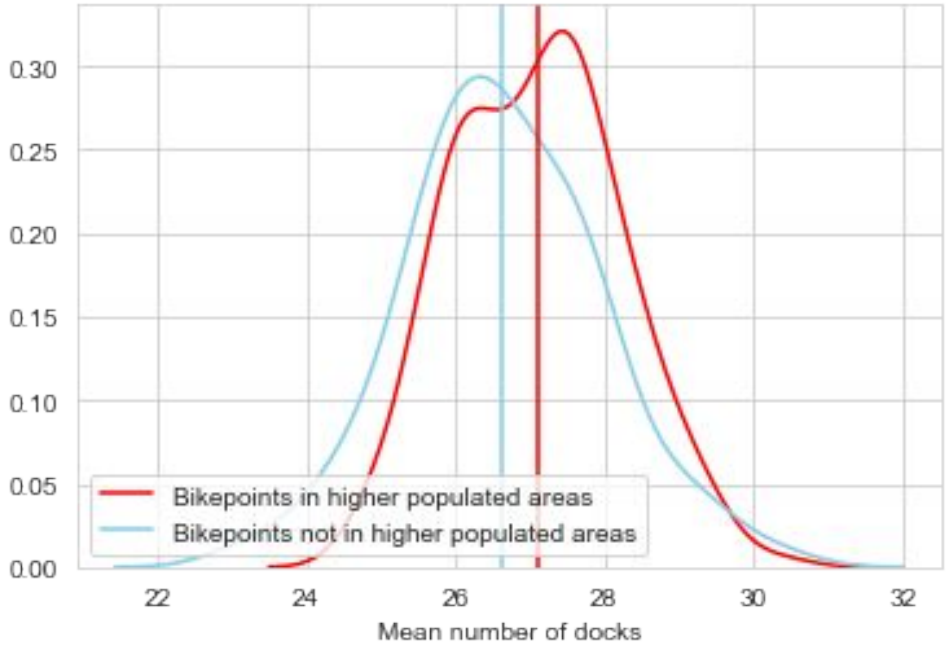
H0 = Bikepoints in highly populated areas have the same number of docks as general bikepoints

H1 = Bikepoints in highly populated areas have more docks than general bikepoints

Statistic: 3.735

Pvalue: 2.16 x 10^-4

Cohen’s d: 0.376



Thanks for listening

Any Questions?