Mini Case Study - Data Linkage

Background

Between June 2015 and May 2016 Google attached air pollution sensors to their street mapping cars in the city of Oakland, California. These sensors measured concentrations of nitrogen oxide (NO), nitrogen dioxide (NO2), and black carbon in the air immediately surrounding the car. (Care was taken so that the sensors did not measure pollution emissions from the car itself.) The purpose of this monitoring campaign was to characterize urban air pollution levels at high spatial resolution. However, it seems plausible that the data could be linked to health outcome data to see if there is an association between concentrations of pollution and poor health.

Data are available from the U.S. Medicaid program on asthma-related emergency room visits for the state of California. The data are aggregated to the ZIP code level and indicate the number rate of emergency room visits in children aged 5–20 in the year 2010.

Data

The Google Cars and Medicaid data are shown in Figure 1 for Oakland, CA.

The Medicaid data come aggregated at the ZIP code leve.

```
# A tibble: 1,708 x 6
  zcta5    ER     n   rate   lat  long
  <int> <int> <int>  <dbl> <dbl> <dbl>
1 90001   210  8535 0.0246  34.0 -118.
2 90002   349  8170 0.0427  33.9 -118.
3 90003   241 10524 0.0229  34.0 -118.
4 90004   134  3657 0.0366  34.1 -118.
5 90005    60  2586 0.0232  34.1 -118.
# ... with 1,703 more rows
```

The Google Cars data are at individual points along roads in Oakland.

```
# A tibble: 21,488 x 5
  Longitude Latitude `NO Value` `NO2 Value` `BC Value`
      <dbl>    <dbl>      <dbl>       <dbl>      <dbl>
1     -122.     37.8       23.4        17.5      0.818
2     -122.     37.8       19.7        20.0      0.551
3     -122.     37.8       23.6        24.0      0.594
4     -122.     37.8       15.7        18.4      0.490
5     -122.     37.8       27.1        25.8      0.739
# ... with 2.148e+04 more rows
```

Linking Datasets

Can these datasets be linked together to answer the question of whether urban air pollution concentrations are associated with asthma ER visits? What assumptions would have to be made? What transformations to the data would need to be made?

Write down three **assumptions** or **transformations to the data** that could be made in order to conduct an analysis of air pollution concentrations and asthma ER visits.

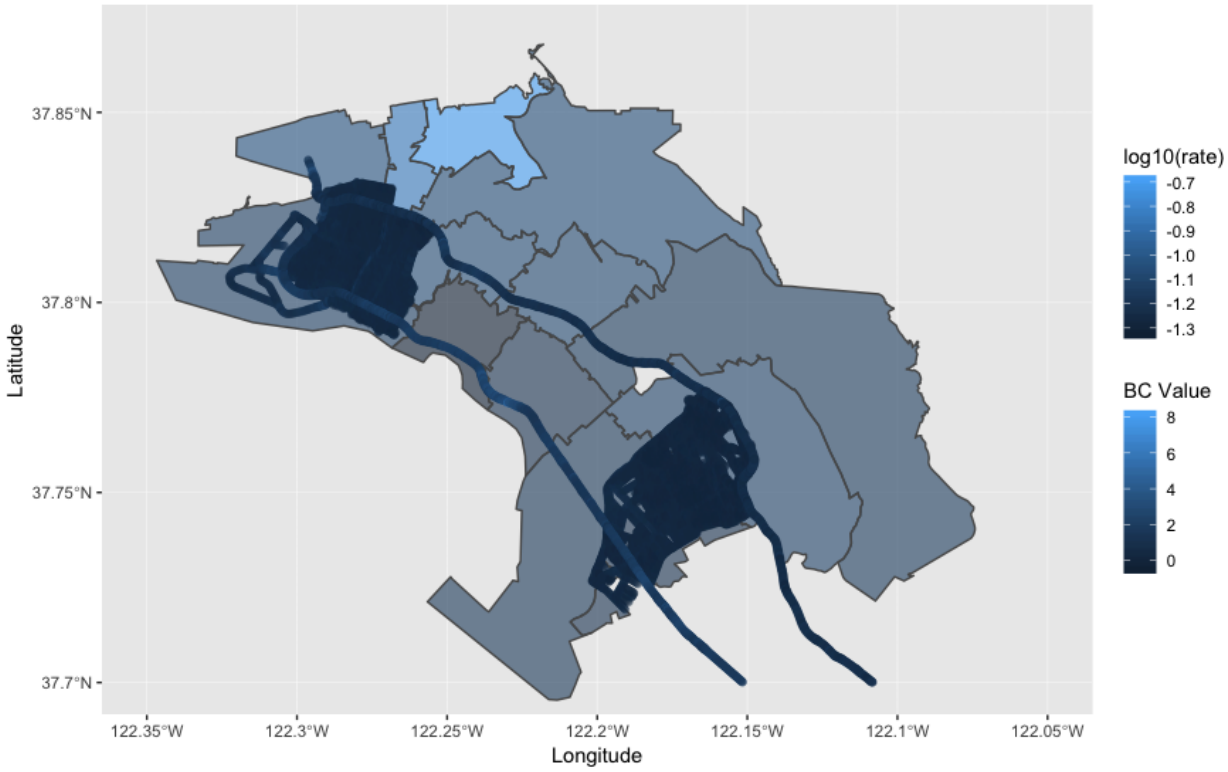

Figure 1: Google Cars and Medicaid Data.