Project: Inspecting Restaurants

Restaurants are generally inspected for health-related practices: food handling and temperature, personal hygiene of the workers, and control of cockroaches and other vermin. Based on the inspection results, local governments may mark the restaurant as passing, put the restaurant on notice, or even close the restaurant.

In New York City, such inspections are conducted and recorded by the Department of Health and Mental Hygiene, which provides a letter grade to each restaurant based on the number and type of violations.

In the era of big data, governments sometimes post their inspection results. In New York City, the record published at `https://data.cityofnewyork.us` includes each individual health violation.

Figure A.21: Orange County, California, restaurant inspection notifications.

Figure A.22: A New York City restaurant displaying its health inspection grade.

Accessing the NYC restaurant inspections data

Execute the commands in this section now, before reading on. It will take about 2 minutes for the last one. Then, while you're waiting, read the rest of this activity. You'll know it's working if `Violations`, `Cuisines`, `ViolationCodes` show up in your account.

You will use four data tables in this exercise:

1. `ZipGeography` in the `DataComputing` package.
2. `Violations` which you must load into R.

   ```
   load(url("http://tiny.cc/dcf/Violations.rda"))
   ```

3. `Cuisines` ..... ditto ...

   ```
   load( url( "http://tiny.cc/dcf/Cuisines.rda" ) )
   ```

4. `ViolationCodes` ..... ditto ...

   ```
   load(url("http://tiny.cc/dcf/ViolationCodes.rda"))
   ```

The data table `Violations` contains information about each health violation.

- `CAMIS` a file code assigned to each restaurant
- `DBA` name of the restaurant.
- `BORO` Which of the five boroughs of New York City the restaurant is located in — Brooklyn, Bronx, Manhattan, Queens, Staten Island — coded as a number 1 to 5.
- `BUILDING`, `STREET`, `ZIPCODE` components of the address of the restaurant.
- `CUISINECODE` the sort of food served by the restaurants. (See the `Cuisines` data set.)
- `INSPDATE` the date of the inspection
- `CURRENTGRADE` the restaurant's grade *after* the inspection.
- `GRADEDATE` the date when the grade was issued
- `VIOLCODE` the violation being recorded. See `ViolationCodes` for explanations.
- `SCORE` the number of demerit points given for the violation.

Quick questions about the data

- What is the meaning of a case in this data table? You can look at the data table with this command: `View( Violations )`.
- How many cases are there?
- How many distinct restaurant names are there?
- Are there any restaurants with the same name but with multiple branch locations? Select one or more variables that plausibly identify a unique branch.
- Which restaurants (individual branches) have the most violations?
- For each restaurant, how many critical violations were reported? Non-critical?
- Make three related charts:

1. the distribution across restaurants of total scores
2. the same, but consider only the score for critical violations.
3. the same, but just for non-critical violations.

In each of the charts, display the score distribution separately for restaurants with different grades.

- Make an interactive map showing the location of restaurants with poor grades. To specify the location, use the restaurant's ZIP code. Get the latitude and longitude of each zip code from `ZipGeography`.

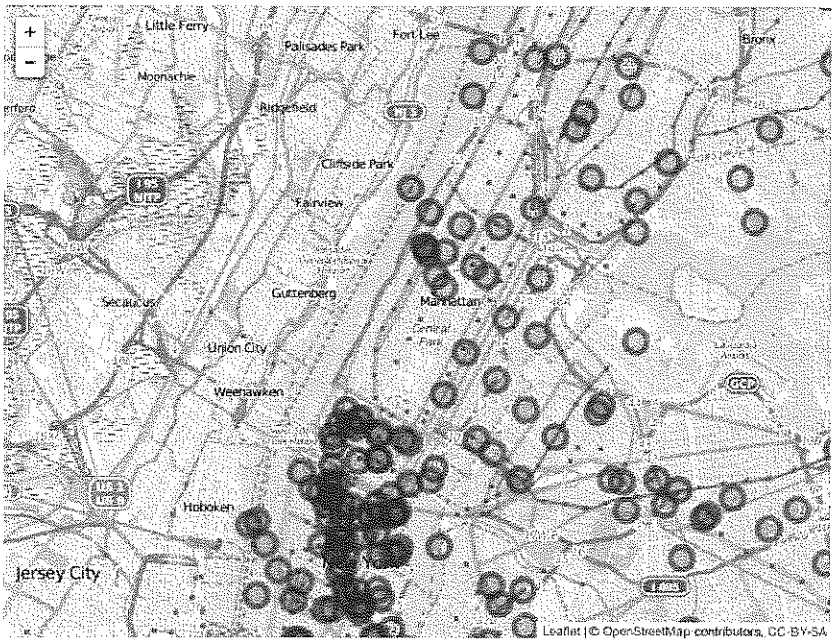

Figure A.23: Locations of C-graded restaurants.

A World of Cuisines

The data table `Cuisines` details the code for each restaurant's cuisine type.

- Judging from the number of restaurants, what's the most common cuisine type in the whole city? In each borough? In each zip code?
- Make an interactive map showing the locations (by ZIP code) of restaurants with a cuisine you select.