Interactive Restaurant Shiny Application

Title:

Capstone Final Project

Author: Craig Lewis

Date: November 15, 2015

Introduction

In 2015, Yelp released a series of datasets as part of the Yelp Dataset Challenge contest. These datasets included anonymized information regarding business reviews, Yelp reviewers, check-in information and tips. The objective was to make this data available to (budding) data scientists and see what can be learned about reviews and business behaviors involving of businesses reviewed on Yelp.

For this capstone project, a Shiny application was developed that allows users to work with an interactive map and see restaurants that match a users-specified selection of criteria. In developing the interactive application, the data was analyzed, cleaned and saved in advance for quick rendering and use of the map. Details of this process along with some summary metrics are provided in this paper.

Methods and Data

There is a large amount of data in the Yelp Dataset Challenge. There are five JSON files that contain different types of information regarding the businesses reviewed on Yelp. For this project, the businesses file was used. The data was read into R and then filtered to identify the categories of businesses and to then extract only the restaurants.

When then map is displayed in the Shiny application we want users to be able to select restaurants by a range of stars in the reviews as well as several other variables of interest. In order to do this, four additional variables are identified and a new dataframe is constructed using a subset of the businesses data set. The businesses dataset is large. There are over 60,000 businesses in the data file and there are dozens of variables used to characterize these businesses. To simlify the Shiny application and speed the loading of the data used to display the maps, we create a new data frame with only the data of interest and then save this data off to a separate file that is loaded by the Shiny application.

Finally, a Shiny applicatno was developed that provided the user with a map of all the restaurants that contain user specified criteria as wellas the average number of stars reviewers have assigned to the restaurant . As the users, select different criteria the map is updated to show the restaurants that match that criteria.

Additional details on this process are discussed below.

Reading JSON Data

The Yelp Dataset Challenge distributes five data sets in JSON format. For the this project, only the businesses dataset is used. The process was ultimately straightforward to do:

biz_dat <- fromJSON(sprintf("[%s]", paste(readLines(json_file), collapse=",")))

where **json_file was the "yelp_academic_dataset_business.json"** file from the Yelp dataset.

Filtering Business Information

There are 61,184 business found in the businesses dataset. This project is only concerned with restaurant information so in this project only the restaurants were extracted.

restaurants<-grep(pattern="Restaurants",biz__dat$categories) bars<-biz__rest<-biz__dat[restaurants,]

Rebuilding a Displayable Dataset

In addition to the large number of businesses found in the input file, there are a large number of columns, some of which are nested (columns within columns). While offering a rich presentation of the data, there is more in the dataset than is required to map the restaurants. Moreover the structure of this data can be cumbersome. It proved easier (ultimately) to filter the data and rebuild a new dataset (data frame) using only the data required in the Shiny application.

The fields extracted from the businesses dataset were:

- **Business ID** - A random set of characters, presumably the anonymized name of the business. Character
- **Stars** - The number average review given by reviewers (number of stars) a number between 1 and 5. Numeric.
- **Longitude** - The geographical longitude of the restaurant location. Numeric.
- **Lattitude** - The geographical lattitude of the restaurant location. Numeric.
- **State** - The state (in the United States) of the restaurant. Ultimate not used. Character.
- **Take out** - Indicates whether the restaurant offers food to-go (take out food). True/False
- **Takes reservations** - Indicates whether the restaurant accepts reservations. True/False.
- **Wi-Fi** - Indicates whether the restaurant offers wi-fi. One of three values: No/Paid/Free.
- **Caters** - Indicates whether the restaurant offers catering. True/False.

This list could be readily extended to include other attributes of interest as necessary.

Once the data was extracted and a new dataframe constructed, the resulting dataframe was written to a file to be read by the Shiny application. Doing this allowed the lengthy preprocessing to be bypassed and the map application to render data quickly.

The code to accomplish this:

bizrates<-data.frame(biz__rest$business_id, ****biz_rest$**stars, biz__rest**$longitude, ****biz_rest$**latitude, biz__rest**$state, ****biz_rest$**attributes**$`Take-out`, biz_rest$**attributes**$`TakesReservations`, biz_rest$**attributes**$`Wi-Fi`, **** biz_rest$**attributes$Caters)**

cc<-complete.cases(bizrates) bizrates<-bizrates[cc,] write.table(bizrates,"bizrates.dat")

Map Rendering with Leaflet

Shiny Application with User Input

Results

Discussion

```r
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```