R Script Examples

Burcin Can Metin

3/1/2017

R Script Examples with R Markdown

This is an R Markdown document for authoring PDF document regarding simple R scripts. The syntax contains how to open and read the json file from Yelp Data Set.

```
library(jsonlite)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Looking into Business Data

```
datapath <- "~/Desktop/DS/dataset"
business <- stream_in(file(file.path(datapath, "yelp_academic_dataset_business.json")))
```

```
## opening file input connection.
```

```
## closing file input connection.
```

```
glimpse(business)
dim(business)
business_df <- select(business, -neighborhood, -address, -postal_code, -latitude, -longitude, -attribut
head(business_df)
```

Looking into Category Data

```
install.packages("tidyr")
```

```
##
## The downloaded binary packages are in
##  /var/folders/cj/q1nnv5j11315zc7s78dkpwf5k6w6pq/T//RtmpEADwSW/downloaded_packages
```

```
library(tidyr)
```

Create a data frame for categories

```
category_df <- select (business_df, name, categories)
```

What categories in the data frame as a list?

```
#results <- sapply(category_df$categories[], function(x){print(x)})
```

A List of first 6 Restaurants from the category data frame

```
head(filter(category_df, categories == "Restaurants"))
```

```
##                         name  categories
## 1 El Paso Bar-B-Que Company Restaurants
## 2                 Edo Japon Restaurants
## 3               Gil's Cafe Restaurants
## 4             Ruby's Diner Restaurants
## 5         Deon's Restaurant Restaurants
## 6                   Freshii Restaurants
```

How many categories are there?

```
length(category_df$categories[])
```

```
## [1] 144072
```

How many unique categories are there?

```
length(unique(unlist(category_df[,2])[!unlist(category_df[,2]) %in% c(NA)]))
```

```
## [1] 1191
```

What are the unique categories?

```
(unique(unlist(category_df[,2])[!unlist(category_df[,2]) %in% c(NA)]))
```

How many Restaurants are in Yelp Business Data Set?

```
number_of_restaurants <- sapply(category_df$categories, function(x){"Restaurants" %in% x})
sum(number_of_restaurants)
```

```
## [1] 48485
```

Create a table of categories

```
table_of_categories <- table(unlist(category_df$categories))
```

What are the most frequent categories?

```
most_frequent_categories <- sort(table_of_categories, decreasing = TRUE)
```

Top 10 Categories

```
top_ten <- sort(table_of_categories, decreasing = TRUE)[1:10]
top_ten
```

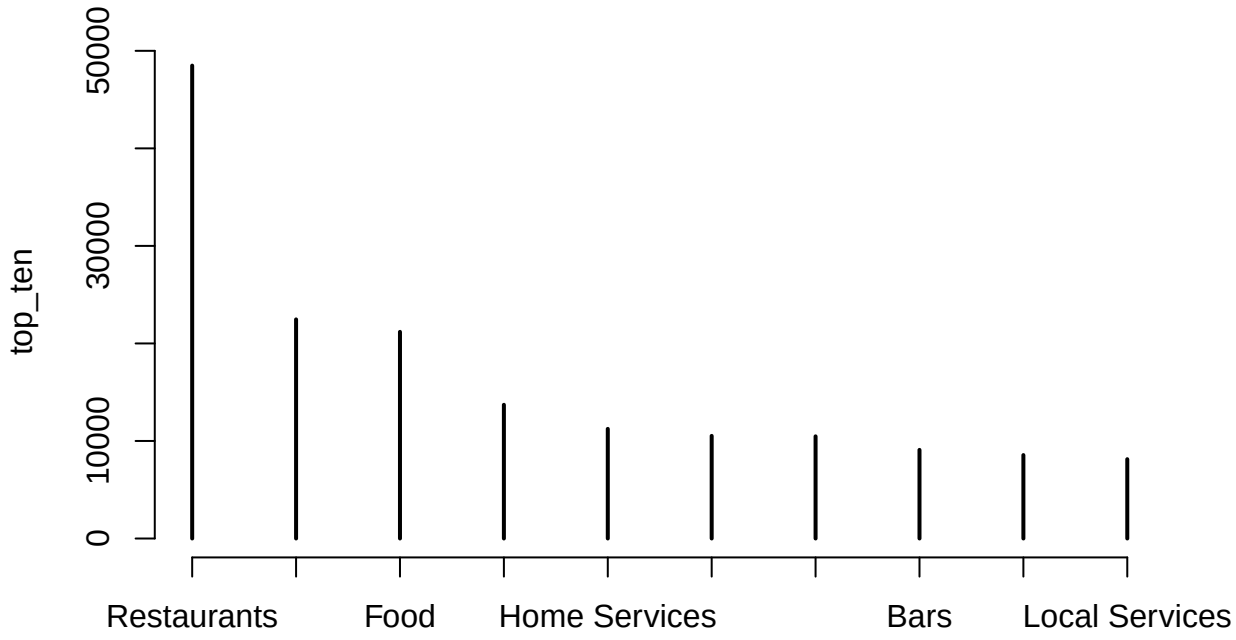
```
##
##      Restaurants         Shopping             Food    Beauty & Spas
##            48485            22466            21189            13711
##    Home Services        Nightlife Health & Medical             Bars
##            11241            10524            10476             9087
##       Automotive   Local Services
##             8554             8133
```

5 Most Popular Categories

Var1	Freq
Restaurants	48485
Shopping	22466
Food	21189
Beauty & Spas	13711
Home Services	11241

A Basic Graph of Top 10 Categories

```
plot(top_ten)
```



Pie Chart of Top 10 Categories

```
names_of_top_ten <- paste(names(top_ten), "\n", top_ten, sep = "")
pie(top_ten, labels = names_of_top_ten, main = "Pie Chart of Top 10 Categories")
```

Pie Chart of Top 10 Categories