Student Name: Henry J. Hu
Course: DSC-611-Z1 Data Visualization
Psychology of Visualization

Introduction

In this assignment, for the purpose of learning how human psychology plays an important role in data visualization, we prepared two dot-maps of New York City using R Studio and Google Maps Static API. The free Points Of Interest data set was provided by NYC Open Data.

Data Source

Description

A spatial data set of New York City points of interests .

Data Size

19,010 observations and 17 fields.

Data Geographic Information

New York City.

Extent

West: -74.276696

East: -73.706565

North: 40.947921

South: 40.453172

Source

Department of Information Technology & Telecommunications (DOITT).

Data Download

https://data.cityofnewyork.us/City-Government/Points-Of-Interest/rxuy-2muj

Table 1: Field Detail

Field Name	Data Type	Description	Unit
SOURCE	Nominal	Agency that defined the CommonPlace location.	Text
SOS	Nominal	Indicates which side of the street the CommonPlace is on.	Text
SEGMENTID	Ordinal	Point is assigned the closest roadbed SegmentID.	Number
SAFTYPE	Nominal	Point is assigned a SAFTYPE if it is a part of a Complex A Alt Address - SAF record same as LION record B Alt Address - SAF record and LION record differ C Ruby Street on Brooklyn-Queens border E Neighborhood name I Named Intersection N NAP assigned to a Stand-alone feature O Out-of-Sequence Address or Opposite Parity Address P Addressable place name S Suffixed house numbers at an intersection V Vanity address M Multiple G Complex NAP X Constituent NAP D Duplicate or overlapping address ranges (real DAPS) F Duplicate or overlapping address ranges (Pseudo DAPS)	Text
PRI_ADD	Ordinal	The Addresspoint ID if the CommonPlace is related to any Addresspoint.	Number
PLACEID	Ordinal	The unique identifier for each CommonPlace point. Links to the point to the FeatureName table.	Number
NAME	Nominal	The name of the CommonPlace. Most name come from Feature name table.	Text
MODIFIED	Interval	Date the data set was modified	Date
LON	Ratio	Longitude degree	Degree
LAT	Ratio	Latitude degree	Degree
FACILITY_T	Nominal	This is a SubType field organizing the CommonPlace points into categories and sets up the domain values for the FACILITY_DOMAINS field. 1 Residential 2 Education Facility 3 Cultural Facility 4 Recreational Facility 5 Social Services 6 Transportation Facility 7 Commercial 8 Government Facility (non public safety) 9 Religious Institution 10 Health Services 11 Public Safety 12 Water 13 Miscellaneous	Text

FACI_DOM	Ordinal	1 Residential 2 Education Facility 3 Cultural Facility 4 Recreational Facility 5 Social Services 6 Transportation Facility 7 Commercial 8 Government Facility (non public safety) 9 Religious Institution 10 Health Services 11 Public Safety 12 Water 13 Miscellaneous	Number
CREATED	Interval	Creation date	Date
COMPLEXID	Ordinal	Point is assigned a ComplexID if it is a part of a Complex	Number
BOROUGH	Nominal	NYC five boroughs. 1 Manhattan 2 Bronx 3 Brooklyn 4 Queens 5 Staten Island 6 Nassau County 7 Westchester 8 New Jersey	Text
BIN	Ordinal	BIN is an abbreviation of Building Identification Number. Point is assigned a BIN if it falls within a building.	Number
B7SC	Ordinal	The Street Code assigned to a CommonPlace .	Number

Development with R

R Markdown was used as the output container for this assignment. The spatial data set was first loaded in R Studio using function read.csv(). Next, we output the descriptive statistics of the data set using functions dim(), str() and summary(). The result revealed that there were missing values in fields SOS, BOROUGH, B7SC, CREATED, MODIFIED, SOURCE, and NAME. However, we only need to use three fields from the data set. They were LON, LAT, and BOROUGH. Since there were missing values in field BOROUGH, we created a subset of the original data set by including only records where BOROUGH > 0. Since field BOROUGH is a numerical representation of the 5 boroughs in New York City, we could not use it to create the legend on the map. A number on the legend has no meaning. Thus, we created another field

named BOROUGHNAME, which was the textual representation of the 5 boroughs. Then, we set the Google Maps API key using function ggmap::register_google(). Then, we defined the center point of the map to be latitude:-73.955 and longitude:40.715-degree latitude. Next, we used the center point together with function get_googlemap() nested inside function ggmap() to retrieve the satellite map and road map from the Google Maps API. The maps were retrieved with a zoom of 13. Then, since the ggsn package has been changed by Github that it no longer works with Google maps API, we had to create a function to draw the scale bar onto the map. Two of these functions were created, each for a different type of Google map. Next, we used package ggplot2 to plot the NYC points of interest onto each map. Then, we used function geom_rect to frame the scale bar onto each map. Also, since the ggsn package no longer could be used with Google maps API, we used functions annotate() and geom_text() to create the North arrow. As for the legend, function geom_point automatically created the legend. Then, we used functions scale_colour_manual(), xlab(), ylab(), ggtitle(), and theme() to apply additional aesthetics to the maps. These two maps are depicted in Figure 1 and Figure 2 below. The hyperlink to the R Markdown page is also included below.

R Markdown page

https://aaacomply.com/data_science/DSC611/Henry_Hu_Moduel_6_Programming.html

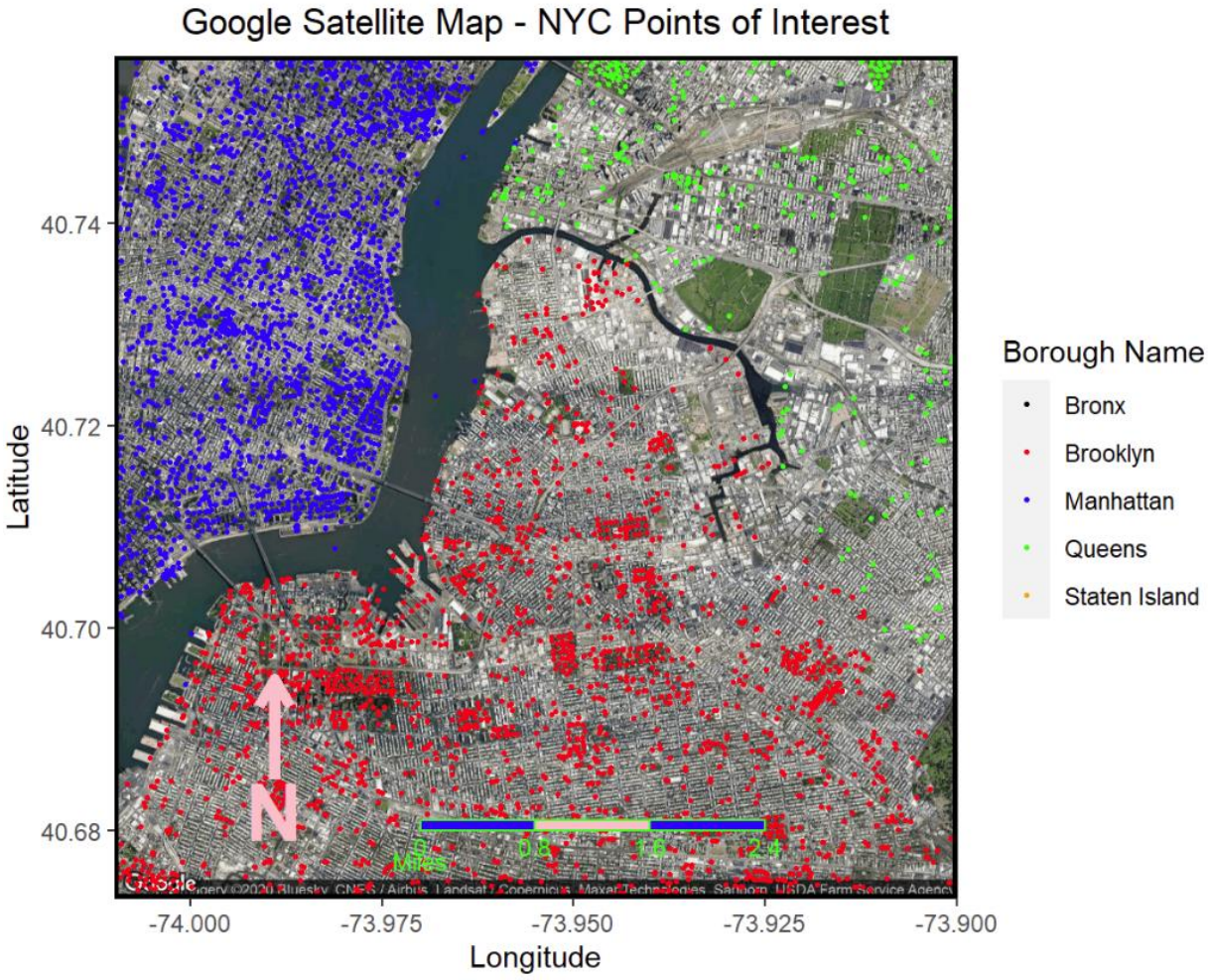

Figure 1: Google Satellite Map – NYC Points of Interest

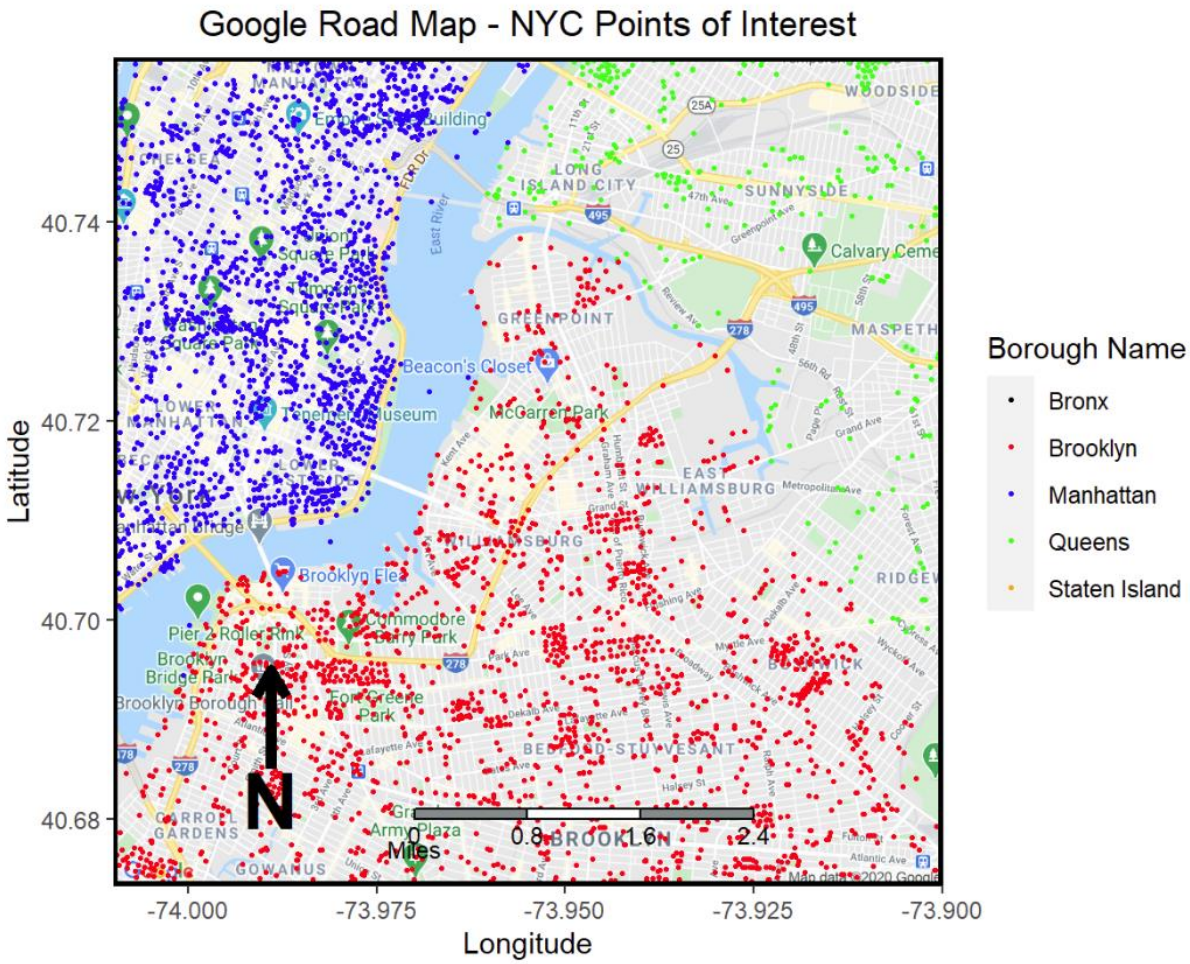

Figure 2: Google Road Map – NYC Points of Interest

Patterns of POI in Brooklyn

Brooklyn is my favorite place in NYC. The interest points at both the top northern region and the lower southern region of the borough are very dispersed. There is not much going on in these two regions. There are many clusters of points at the center region of the borough. One of the clusters is directly on top of the Lew Wallace Public School 284. These clusters seemed to be concentrated horizontally between Brooklyn Bridge Park and Bushwick and vertically between Bedford Stuyvesant and East Williamsburg. The clusters close to the Brooklyn Bridge Park are the densest. It appears that there are no other patterns in the dots except for the clusters.

Differences and Commonalities between The Two Maps

The latitude range, longitude range, geographic midpoint, geographic location, zoom level, elevation, depth, width, height, points of interest, north arrow size and shape, scale bar size and shape, and legend size and shape of both maps are all exactly same. However, the map title, map detail, and different colors are used for north arrow, scale bar, and legend. For example, the satellite map has real details of the city from far above. If one looks closely, one could see the green pasters, railroad tracks, and empty land fields. From the road map, one could not tell the difference between a rail line or a body of water, or the difference between a green paster or an empty land field. It is because the road map is only for finding direction while traveling on the road. I personally prefer the satellite map over the road map. It is because I am very interested in exploring the details when I travel.

Summary

The implementation of these Google maps in R has helped explain most of the concepts covered in the reading for this assignment. After looking at the two maps, one could see that different perceptions are created by the contexts of these maps. The satellite map is more for

an advit traveler to use since it has real details of the neighborhoods, parks, rivers, housing conditions, and public transportation. This type of travel wants to explore the city. The road map is more for a passerby traveler who just wants to quickly find their way through the city. This type of traveler is not interesting in exploring the city. The north arrow on the map helps the traveler decides which direction they should be traveling if they need to reach a certain destination from their location. The scale bar on the map helps the travels approximate the distance in miles they will need to travel to get to their destination. The legend let the traveler know the location of each borough within the city.

Gestalt's principle of proximity helps explained the clusters in the region, horizontally between Brooklyn Bridge Park and Bushwick, and vertically between Bedford Stuyvesant and East Williamsburg. Gestalt's principle of similarity helps explained the distinction between the tiny inland rivers versus the railroad lines. Gestalt's principle of the close region helps explain why the audience saw three boroughs on each map. It is because both the large Hudson river and the small inland rivers helped enclosed the land for each borough. Gestalt's principle of Focal point also helps explain why the north arrow on both maps are the first object any map user will see. It is because the color for the north arrow was specially chosen so it has high contrast against the rest of other colors on the map.

Besides the north arrow and scale bar, maps are very similar to some other types of plots and charts. For instance, an area chart is very similar to a map. The only difference is that a map indicates direction and location, while an area chart indicates the magnitude and volume. However, Gestalt Theory could be used to explain both maps and area charts.

Work Cited/References

Berengueres, J., Fenwick, A., Sandell, M. (2019). Introduction to Data Visualization & Storytelling: A Guide For The Data Scientist (First Edition).
Independently published