Aditya Retnanto

CMSC 3321: Surveillance Aesthetics: Provocations About Privacy and Security in the Digital Age
Assignment #1

**Part 1:**

In this part of the exercise, I was tasked to request data from companies that potentially collect my data. The table below indicate the data obtained.

**Table 1. Table Containing Companies that Returned Data Request**

| Company | Company Sector |
|---|---|
| Google | Multi Sector |
| Snapchat | Social Media – 'Deleted' Pictures |
| Facebook | Social Media |
| Tik Tok | Social Media – Short Video Clips |
| Instagram | Social Media – Based on Pictures |
| Spotify | Music |
| Hinge | Dating |
| Netflix | Film/Video |
| LinkedIn | Job/Recruiting |

As of 10/6/21 the only company that did not respond to my data request was Amazon.

**Part 2:**

Using this dataset. For this project I decided to explore two concepts – one closely related to data visualization and one more artistic. Within the visualization space I aim to create a geospatial dashboard from the perspective those who are doing the surveillance. The medium of the geospatial dashboard aims to question the viewers self-perception on their value of data and highlight the dehumanizing aspect of projecting users as mere bits on a screen, or as instruments to the intentions of the surveyor. Information regarding audio and audiovisual consumption will be displayed, and colors correspond to clusters of locations one frequents. Although one can view the dashboard using other people's data, it is best viewed using one's own information as one can perform the mental cross-validation between one's memory and dashboard visualization. Due to time constraints, the concept of multimedia consumption isn't fully explored but the dashboard provides sample points of when content is consumed. During those periods, users are susceptible to advertisement.

Within the artistic space, a concept I wish to explore is inspired by Jackson Pollock. Although it was not possible to do it digitally, one can transpose the placement of my location dot onto a physical. The color represents the assigned label of above clustering, where the artist can then create gradients and blend on the canvas. Although this might not seem provocative to a

general audience, it does provide evoke the spirit of modern art as it focuses more on the nature of material. This will not be submitted but for future personal art projects.

**Part 3:**

*Design Motivation*

After parsing through the data, I wish to visualize the relationship between space and engagement. Within the Snapchat dataset, I found information regarding my 'rough' location in latitude and longitude. Although that information is not very granular, it still provides a good estimation for 'rough' location. Netflix and Spotify data contain information regarding my content consumption within those apps. I decided to stratify engagement into categories – audio (Spotify) and audiovisual (Netflix). Another key design assumption is that I will be content type agnostic and profiles with multiple personas are lumped into a single persona.

*Algorithmic Considerations*

The first step of the code is to load the datasets in a pandas dataframe in python. The first major roadblock and learnings is that there is no universal standard on how user data is structured. The code individually loads the datasets from the companies and drops most of the content. In the table below, I list the variables that are extracted from the dataset.

**Table 2. Filtered Variables from the User's Data Set**

| Company | Variables |
|---|---|
| Netflix | Start Time, Watch Duration |
| Snapchat | Latitude, Longitude |
| Spotify | End Time, Length |

Location is a key component in the dataset that I wish to construct. To avoid null information, I bound all information within the given time window from snapchat. In the next step I upsampled the above timeseries information to construct a concatenated pandas dataframe. The window of each sample is set at 10 minutes but can be set to whatever the user wants to. With the known durations of Netflix and Spotify content, I can created a Boolean window representing activity if there is overlap between the upsampled times and duration. To account for a dynamic definition of frequently visited clusters, I trained a DBScan model to assign labels and predict what cluster each sample belongs to. Frequent locations were then defined as locations where the cluster density (number of samples in a cluster) is greater than or equal to a uniform cluster density. Folio is used to visualize geospatial information, when clicking on markers information regarding content consumption can be accessed.

The results of this exercise is saved in 'data/personal_tracking.html'

**Appendix:**

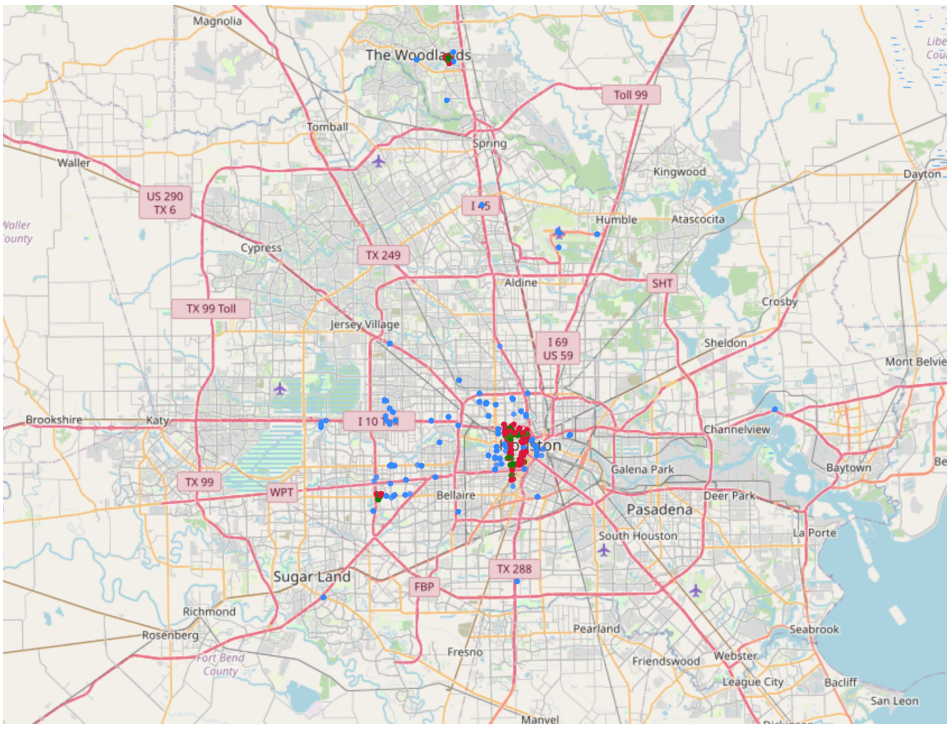


**Figure 1.** Sample Geospatial Partition Containing User Activity

**Figure 2.** Plot of Samples on Longitude and Latitude Axis – for Canvas Reference