Interconnection: NYC Subways and the People Who Fuel Them

Zeana Llamas and John Noonan

April 26, 2018

Project video: https://vimeo.com/266798433

Motivation

Public transit is a vital facet of life in New York City, it's so commonplace that it is at times hard to understand the magnitude of life surrounding it. While the Metropolitan Transportation Authority (MTA) is not often viewed as glamorous, it had ridership as high as 1.757 billion in 2016 [1] and is of utmost importance to those it serves. However, merely observing the numerical representation of a system with such a humanistic effect does not allow the full picture to be seen. Therefore we set out with the goal of visualizing the city as a living connected organism composed of millions of people living their daily lives. To do this we utilised the publically available turnstile data from hundreds of subway stations provided by the MTA. We wish for the viewers of our visualization to feel a sense of introspection and to think about the hidden connections between all city-goers facilitated by the subway system. This project will be a success if after watching our visualization a viewer takes even the slightest moment to stop and think to consider their role and connection to others in the greater scope of the city.

Audience

Anyone and everyone is free to participate of course; however, *Interconnection is primarily* created for all demographics of people with the binding factor of having a life that is directly affected by NYC as a whole, but, specifically those who use the subway system, in order to show how they affect the city around them.

Research Question and Hypotheses

This project is truly about the viewers of the visualization and not the information gleaned from their usage. We want to encourage them to think critically about the subway, how it affects the population of the city as an interconnected whole, and what their role is in all of it. Therefore our research question was whether we could create a visualization that could properly portray the features of the subway system and inspire the introspection we seek. We hypothesize that our distinct visualization form will create interest and hopefully instill mindfulness.

Related Work

Visualizations of NYC Subway turnstile data has been done before with an exemplar being the work done by Chris Whong [2]. Using the same (all though differently formatted) dataset as our project, Whong created a geographically accurate visualization of activity at a majority of the stations in New York. His work simulates movement of individuals by clustering groups of entering and exiting patrons as dots with respective colors to their actions. These dots then move around and fade away over time giving a nature documentary time lapse video feel. As stated earlier, our project uses the same dataset and also seeks to portray the cyclical peaks and lulls of activity. Our projects differ in their intentions as Whong's was created purely to visualize the turnstile data and serves as a visualization of curiosity while ours is less concerned with the simulation type feel and wishes to show the organic nature of usage.

Additionally this project was inspired by The Mill's Uproar [3], an original art piece which pays homage to the resilience of the people of Houston by simultaneously visualizing storm data from Hurricane Harvey and Twitter mentions of #HOUSTONSTRONG. This piece visualizes numerical data through form and forgoes explicit usage of numerals which greatly inspired Interconnection. Their piece is equal parts information and art and strives to tell a story and impart emotional responses from the viewers. We seek to do similar things and use art to tell a data driven story that will allow the viewers to think about themselves and others in the context of connectivity in the City.

Design Evolution

Our idea for the 3D animation of our collected data that would organically represent the activity at each train station went through several steps of design evolution as we developed the method to correlate the change in number of station entrances and exits to the movement of our point clouds.

Figure 1. First attempt to line up geographical data results
(blue points) to reference image of the NYC Subway Station.
Reference image source: http://web.mta.info/nyct/maps/subwaymap.pdf

We decided rather early on that the visualization would be primarily abstract, void of labels and other monikers; therefore we thought it best to have our stations situated in space in relation to each other based on latitudinal and longitudinal data, see: Figure 1.

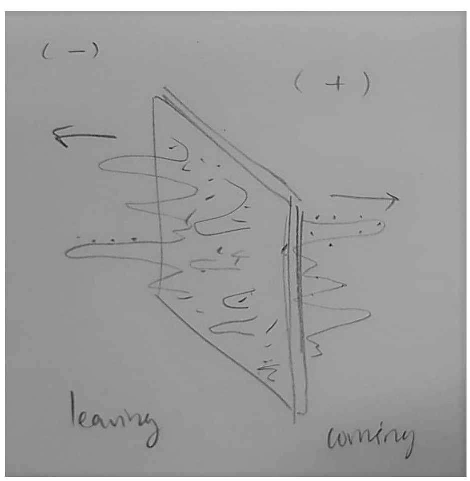


Figure 2. The initial sketch of Idea #1, (top)

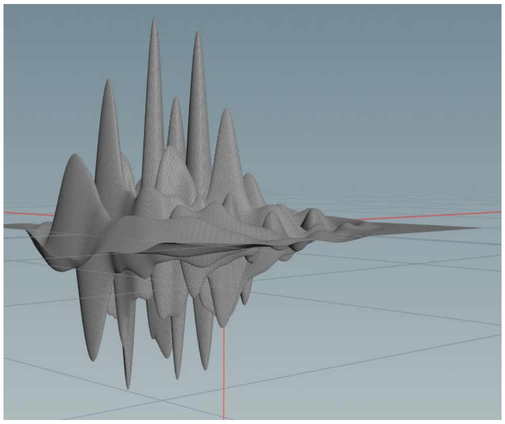
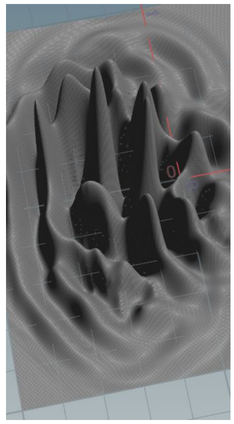
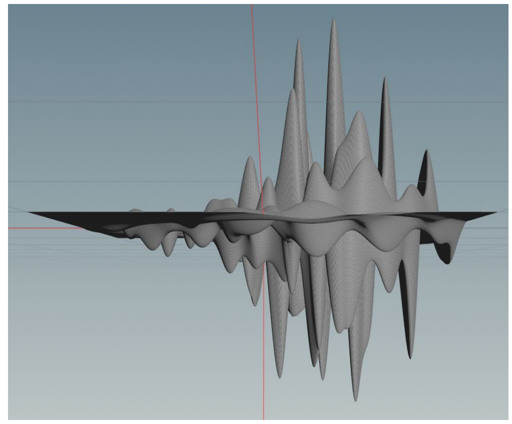




Figure 3. Modeled Idea #1, ¾ view (left)
Figure 4. Modeled Idea #1, original station location data points are visible towards the center and between crevices, bottom view (middle)
Figure 5. Modeled Idea #1, ¾ view (right)

We considered a variety of designs such as Idea #1 of deforming a mesh based on the location points with the peaks in either a positive or negative direction signifying the level of activity at the station. Peaks in one direction (left or down) would signify the a negative delta between arriving and departing patrons at a station, the opposite direction (right or up) would signify the positive change having each station.

Figure 6. Sketch of Idea #2. A frontal view of one of the point clouds that would be generated at a station. The height and color would be indicative of activity (top)

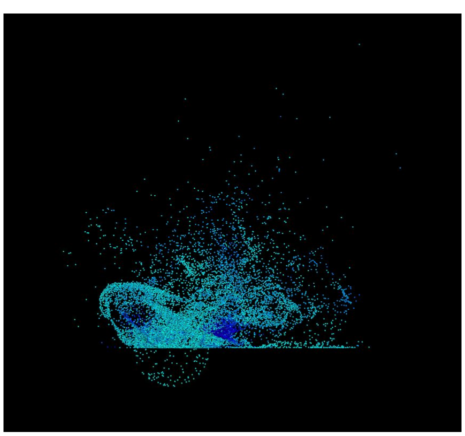
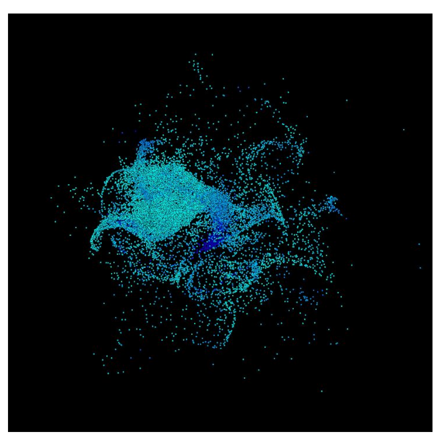



Figure 7. Modeled Idea #2, front view (left)
Figure 8. Modeled Idea #2, top view (right)

However, we opted for a more free flowing option of points moving freely in space as it appeared more organic, see: Idea #2. The first iteration of this is seen in Figure 6-9, the use of noise and color clearly displaying level of activity.

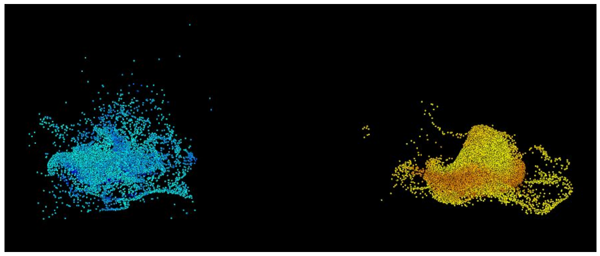


Figure 9. Modeled Idea #2, an example of adjacent stations

Feedback

Unfortunately most of the people we were able to get feedback from did not match our intended audience, therefore, we chose to ask targeted questions about visual aspects of the piece and general comments about the feeling it imbued. On the whole, most of the respondents found our work visually interesting and lively. Some people were confused by the meaning of the particle systems and suggested ways to clarify this such as adding numbers above them and using a real map. We decided not to do either as we felt that it would detract from the message intended. In terms of visual suggestions multiple people agreed that using color to show activity as well growing and shrinking the radius of the point cloud are effective. The use of the radius however had mixed reviews as some believed it might would create too much congestion in areas with high levels of station clustering. The advice for color usage and radius size were both planned to be used despite the few objections as the congestion created fit in with our theme of bustling activity.

Design Evolution [continued]

For our final visualization, drawing from some of our classmates suggestions, at each station the level of activity is controlled by height and roughness of the noise surrounding each station and finally by a color gradient that is mapped from the most active station to the least to further enforce visually the busyness of a station. So it very greatly resembled the version of our visualization presented during our peer review session; however we altered the way that the point clouds were manipulated to achieve a bit more control to achieve a more distinct difference between “calm” (see: Figure 11) and “busy” (see: Figure 12) stations.

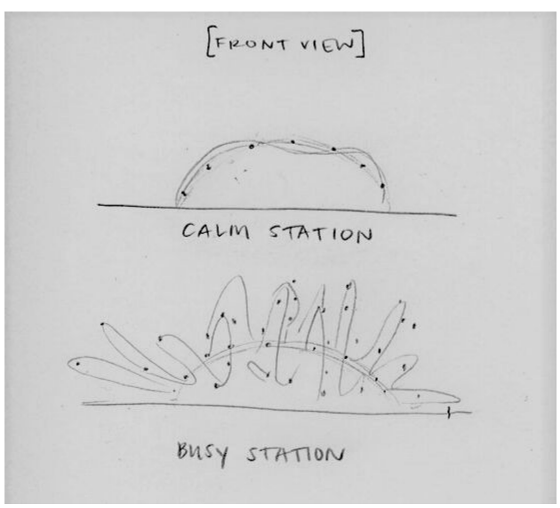
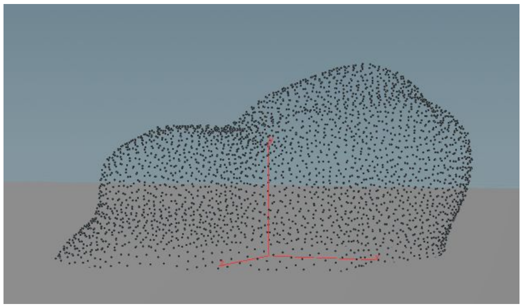


Figure 10. Draft of redesigned point cloud for final version



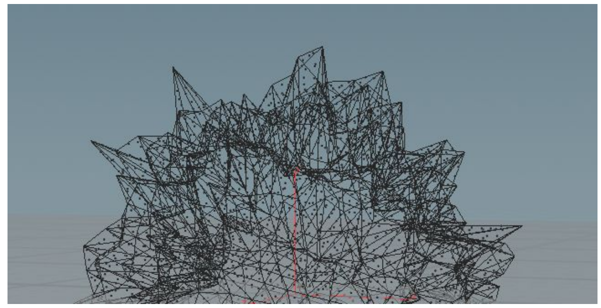


Figure 11. An example of a “calm” point cloud at a station, ie. low levels of activity (left)
Figure 12. An example of a “busy” point cloud at a station, ie. high levels of activity (right)

Contributions

Interconnection was implemented in several parts. First, Python scripts were used to parse and clean the data obtained from the MTA. For this project we used the turnstile data for the week of February 17, 2018 and additional location data of the stations. The input data was stripped of unnecessary elements such as turnstile unit, internal tracking numbers, and train line names, eg. 1,2,Q,B,D trains. Each station was represented with a data structure containing its pertinent information such as its name and entrance and exit values which are sampled every 4 hours.

An immediate challenge of the project was inconsistent naming conventions for the stations across the two data sources. For example in our turnstile data a stations name might be written B'WAY-LAFAYETTE while in the location data it is Broadway-Lafayette St. This incongruity made matching stations to the locations difficult. To fix this we wrote a program that could algorithmically match names of the format “number + number suffix + street type” (i.e. "125th St") as well as concatenations of that type with additional strings or streets (i.e. “116th St-Columbia University”). Unfortunately there were no other patterns in naming and other stations needed to be matched by hand.

In the turnstile input data, the entrance and exit data are tracked cumulatively, that is at time slice n, the entrance value might be 0001767751 and at time n + 1 it might be 0001767859. We wanted to get the number of entrants from slice n to n+1 which was easily calculated by doing entrance at time n + 1 - entrance at time n. This was similarly applied for exits. A problem with this however is that some data was incorrectly reported (i.e. random changes in order of magnitude as well as clearly incorrect data) as well as that certain turnstiles underwent maintenance and would have their counters reset resulting in incorrect delta values for entrance and exits. Unfortunately, without knowing what the true value was any data value with suspect values needed to be algorithmically culled.

Once the data was collected and formatted all visual work was completed in HoudiniFX, a procedural node-based 3D animation software. This was chosen to leverage the relatively quick iteration speed allowed and because it supports great scripting functionality. In the application we mapped the spherical longitude and latitude data of each station to a 2D plane. With this we instantiated geometry at each station and used an organic noise function with parameters fed by the activity of the station which we defined as the number of entering patrons minus those exiting. We used Python and Vex (Houdini’s scripting language) as well as Houdini’s built in nodes to create the animation.

Future Work:

If we had more time and more resources we would have liked more complex interactions between the particles, as if they are travelling between stations in order to further reinforce the connectivity of the system. We would've also liked to make additional renders of the animation with slightly different features which could be edited together, eg. a version with the subway map underlay for reference, or different colors that reflected the color of the train line. We even would have liked to create an online interactive version in which users could tumble and zoom in and out through the animation, enabling other camera angles other than the ones we focused on in our video and encouraging exploration, perhaps of stations that have more personal significance to them.

Work Breakdown

John did initial parsing, and data manipulation so that it could be loaded and used in HoudiniFX. Zeana did data clean-up, formatting, and initial scene set up. Both team members worked on design conception, final paper writing, editing the particle interactions, the final rendering, and video compositing,

References

[1] " Subways," *mta.info | Facts and Figures*. [Online]. Available: http://web.mta.info/nyct/facts/ffsubway.htm. [Accessed: 26-Apr-2018].

[2] C. Whong, "Visualizing the MTA's Turnstile Data," *Chris Whong*, 24-Apr-2013. [Online]. Available: https://chriswhong.com/open-data/visualizing-the-mtas-turnstile-data/. [Accessed: 26-Apr-2018].

[3] "UPROAR," *The Mill*, 08-Feb-2018. [Online]. Available:http://www.themill.com/millchannel/1546/uproar. [Accessed: 26-Apr-2018].