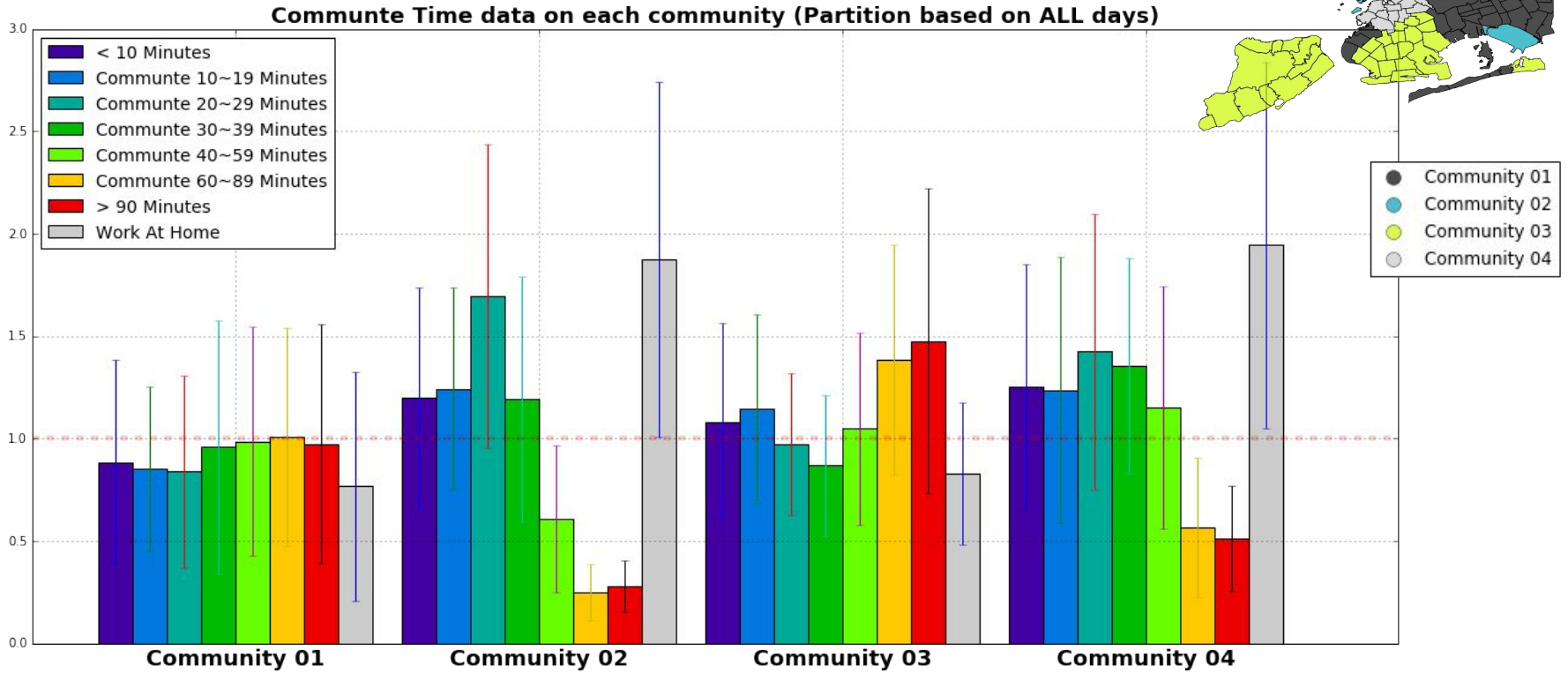
QC Social media

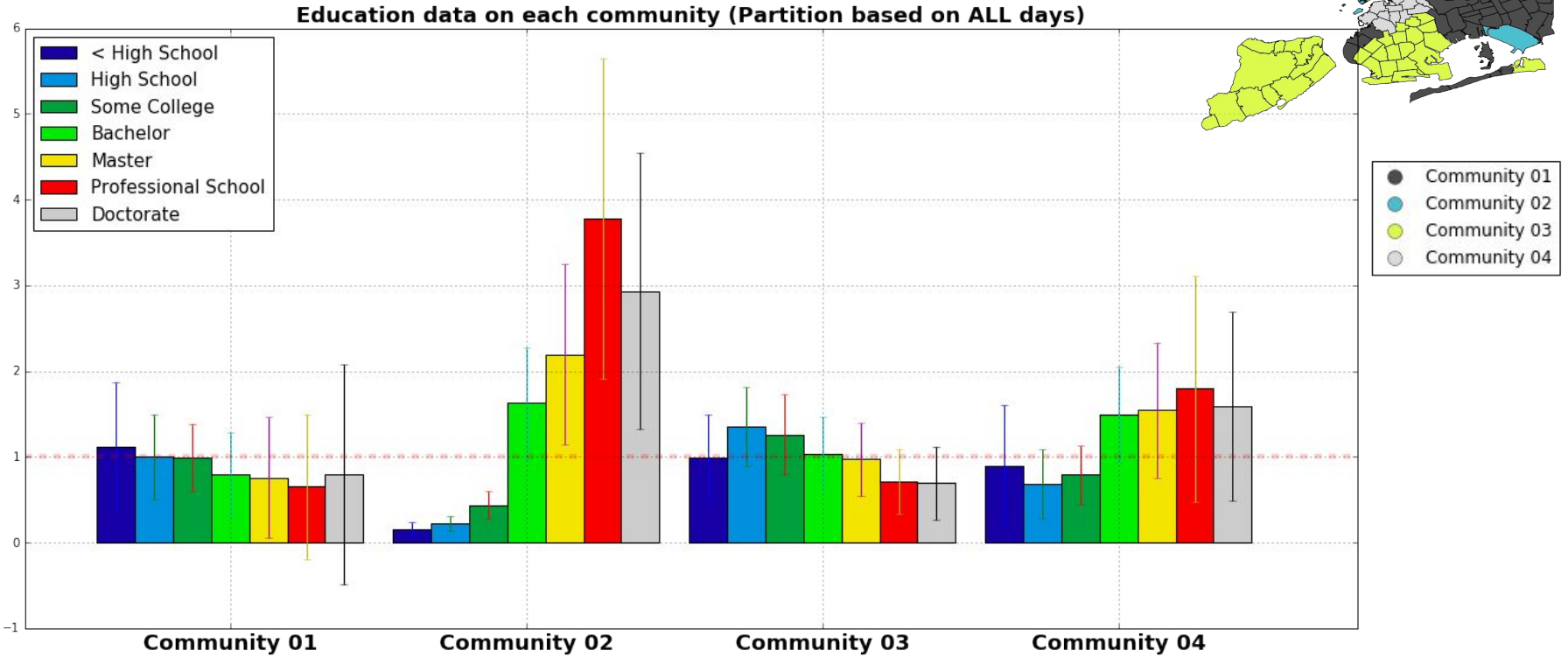
Report 7

4.04.2016

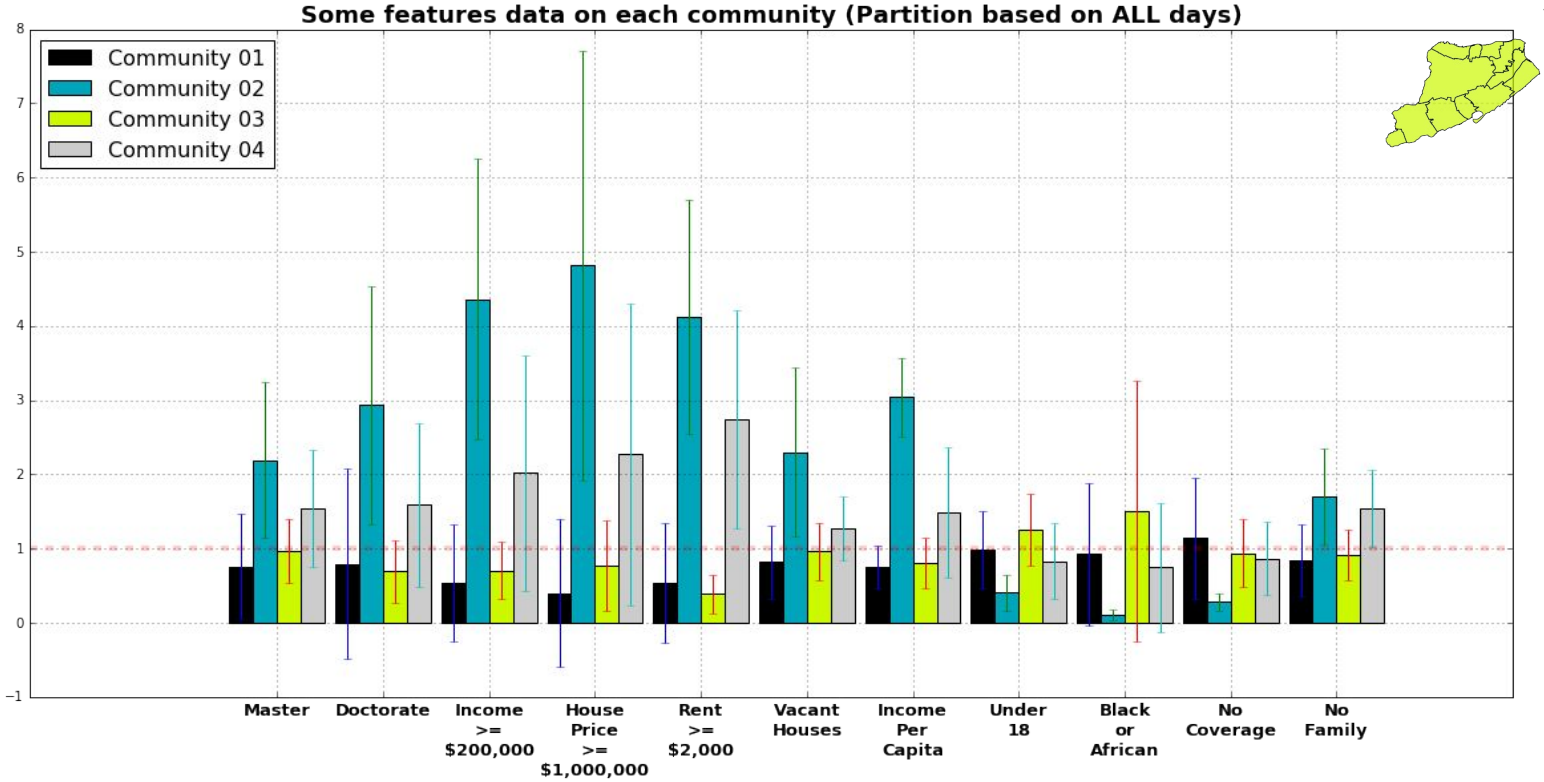
Demographic feature example 1

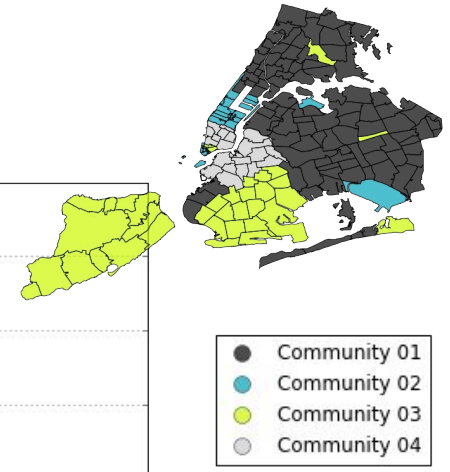


Demographic feature example 2



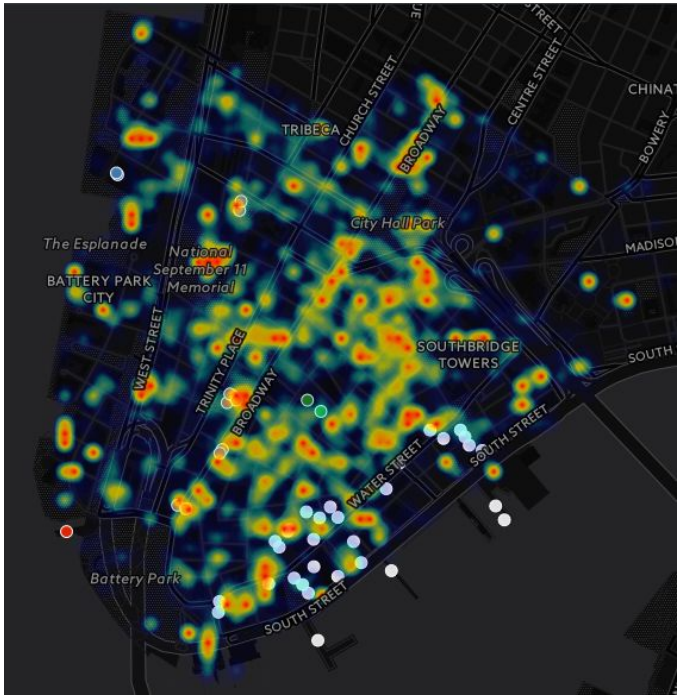
Some selected features



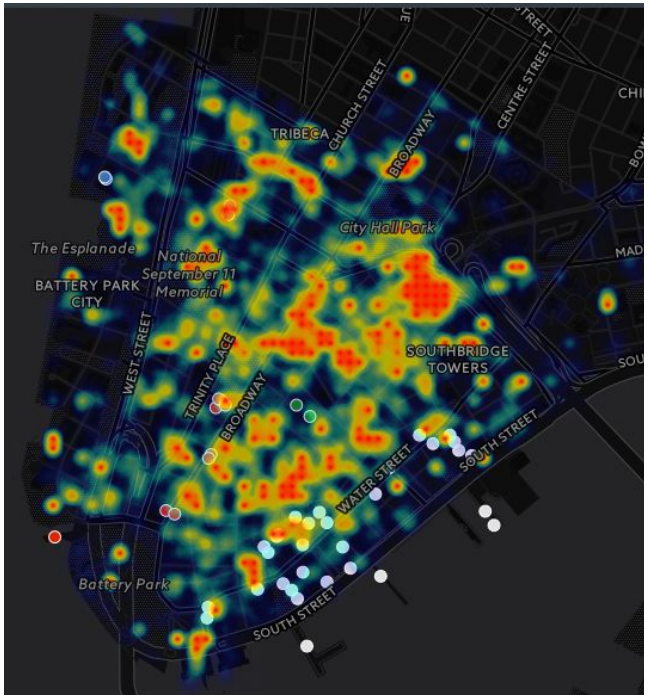


Lower Manhattan Tweets

- BROADWAY WIFI
- DOWNTOWN ALLIANCE 7 WTC
- ONE CHASE PLAZA
- WATER STREET CORRIDOR
- Z-WINTERGARDEN

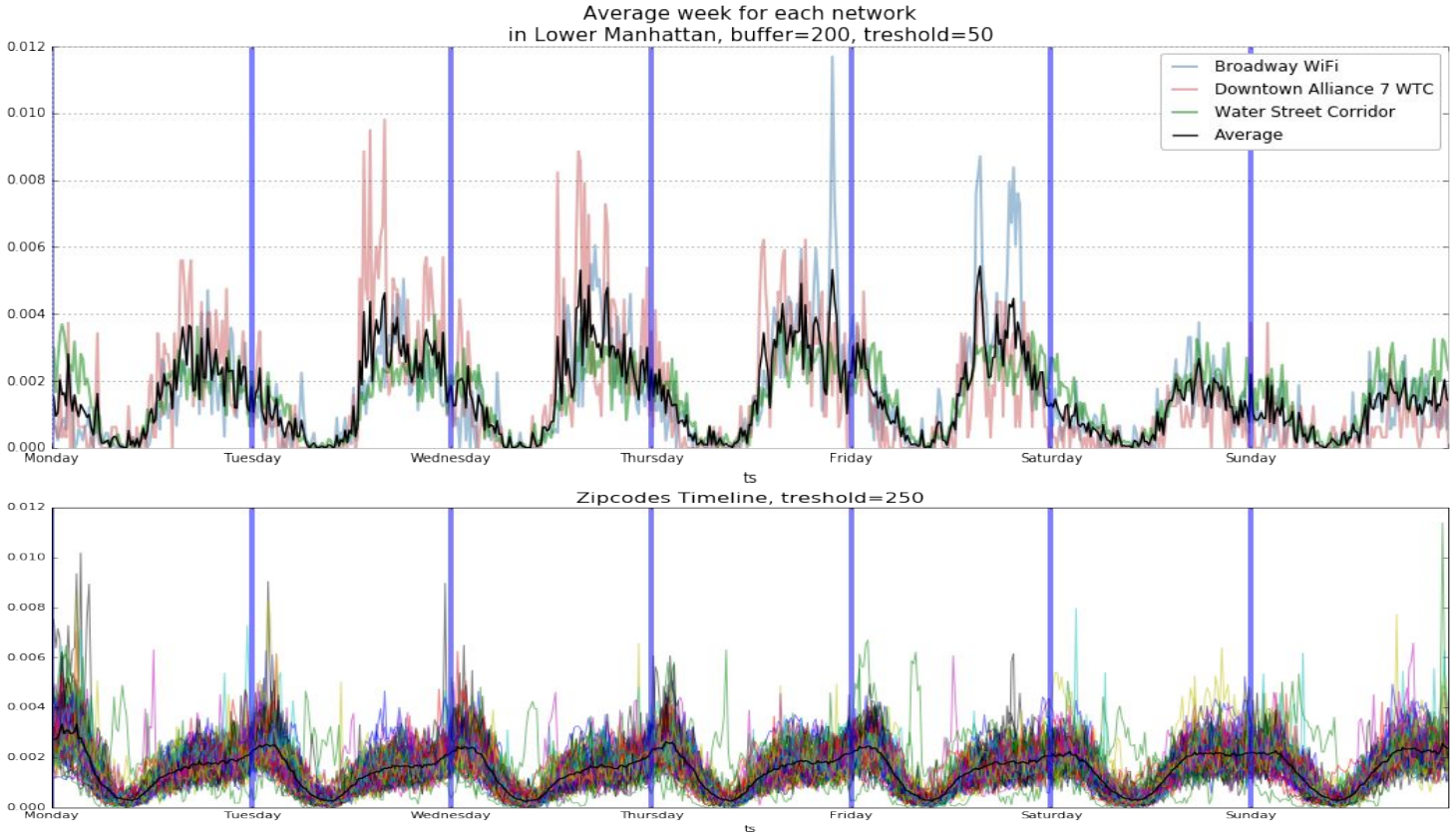


2015.01, 1-15

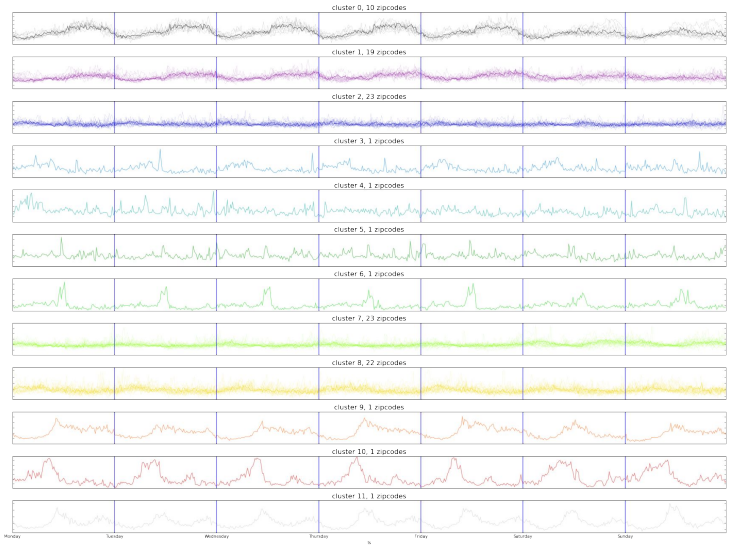


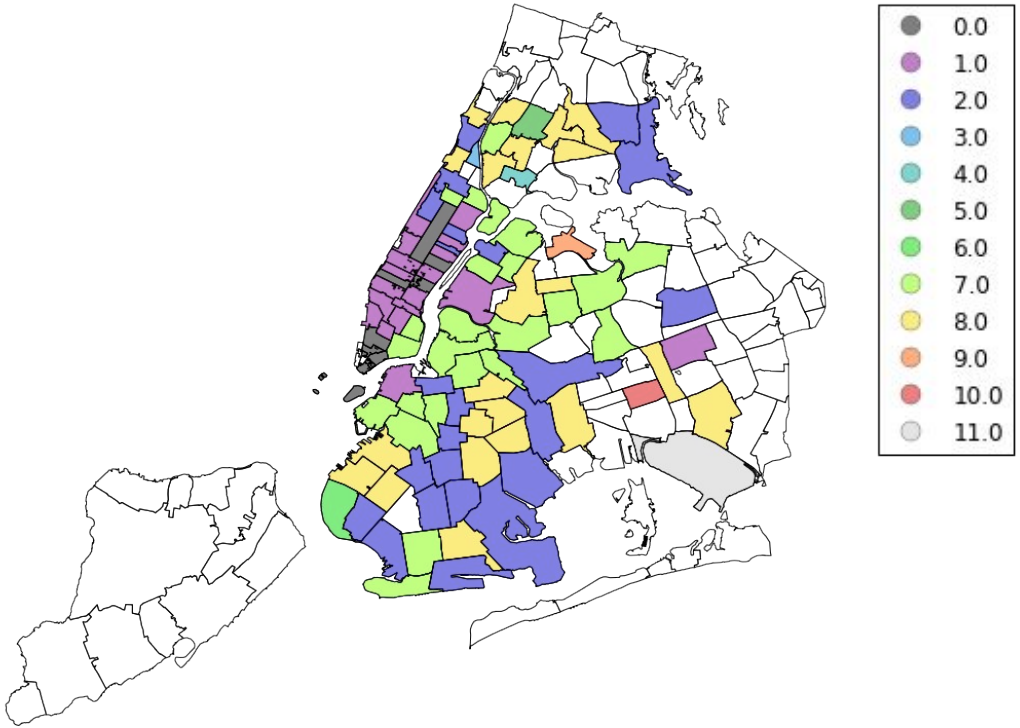
2015.03, 1-15

Lower Manhattan Tweets



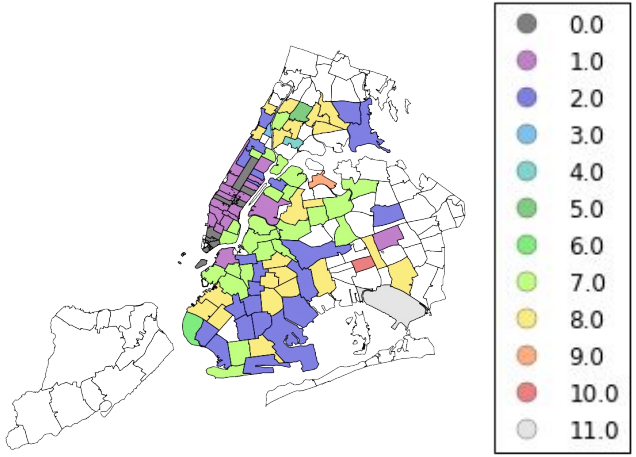
Time Series clustering - demography

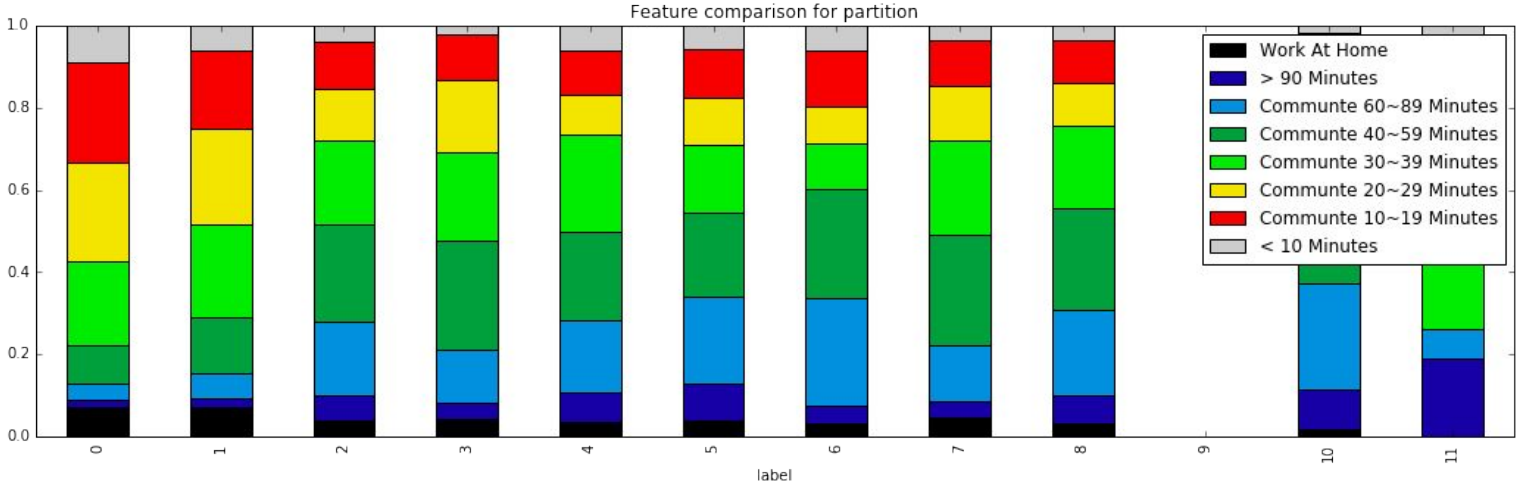




Time Series clustering

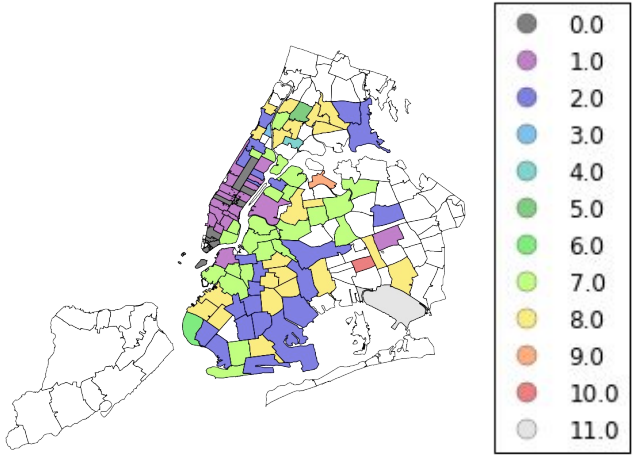
Commute

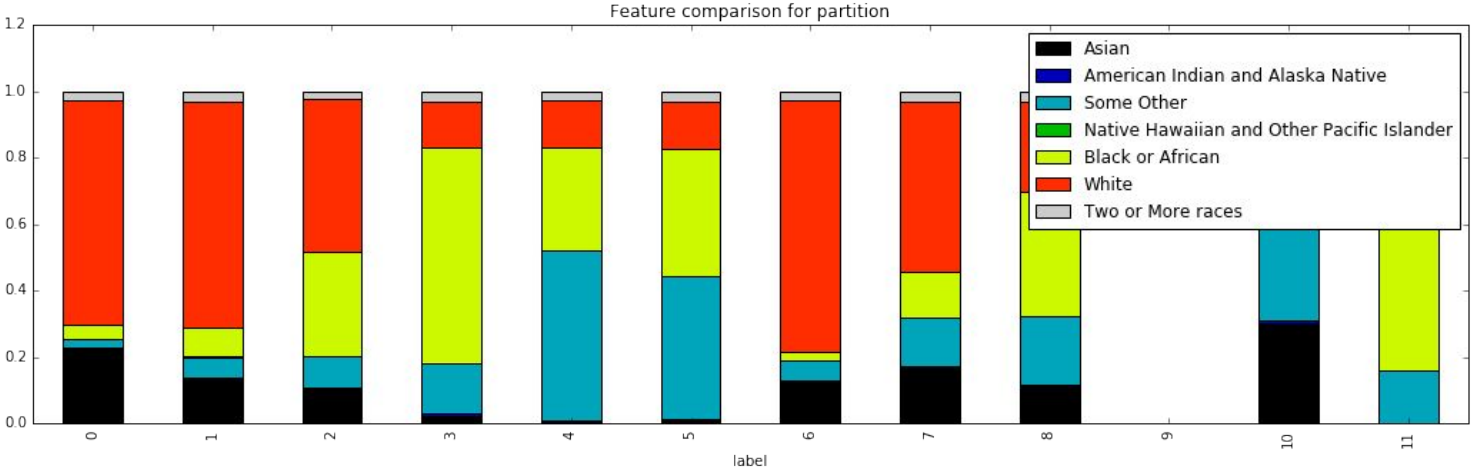




Time Series clustering

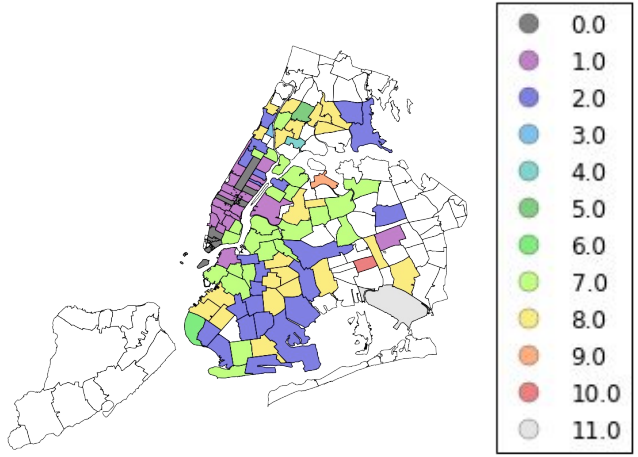
Race

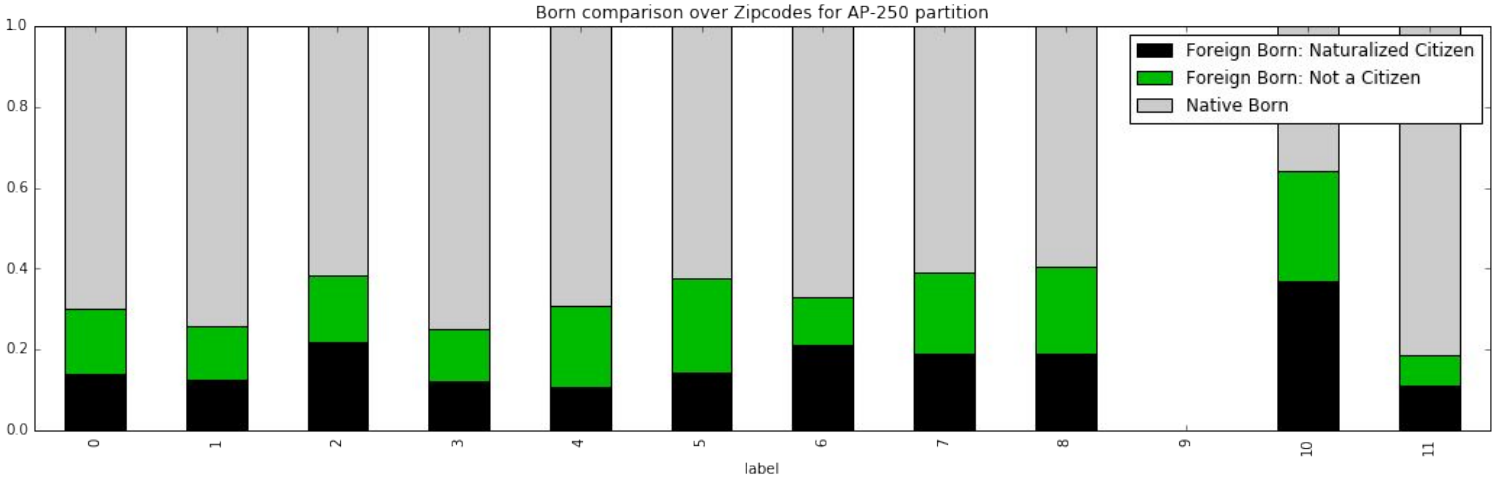




Time Series clustering

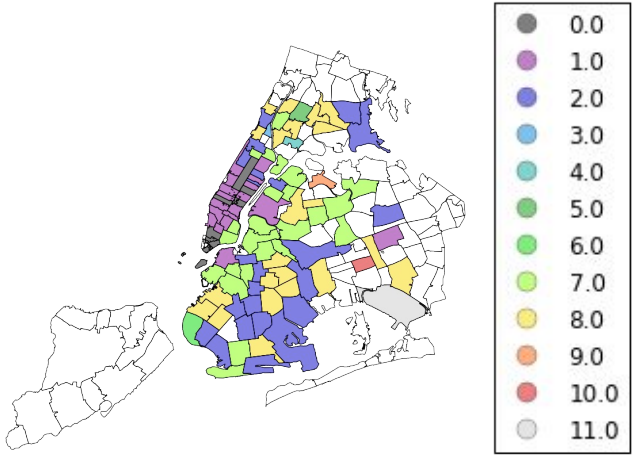
Citizenship

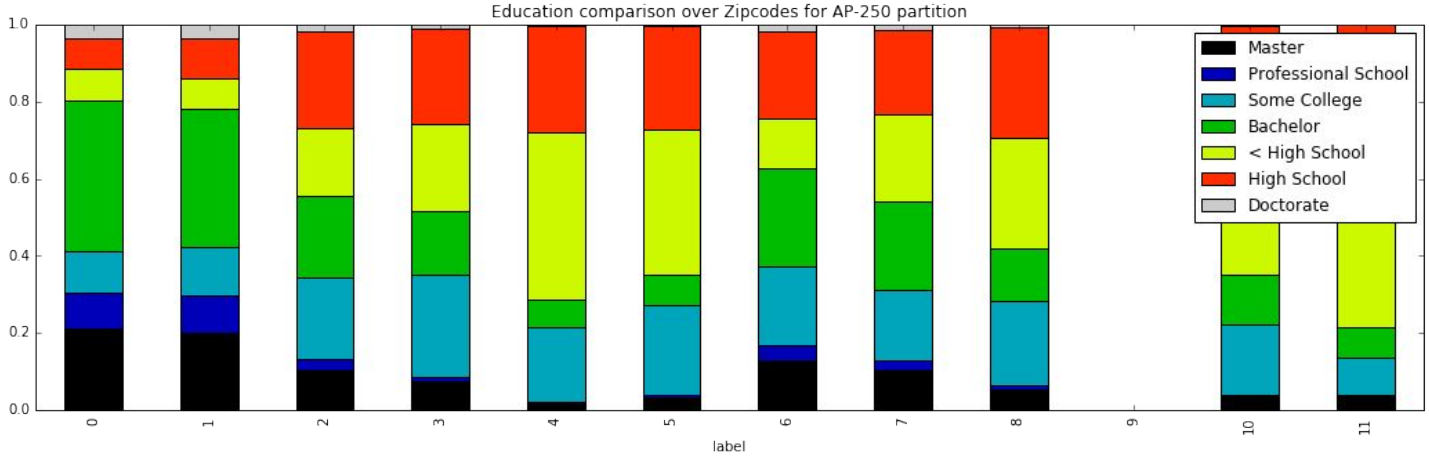




Time Series clustering

Education





Note: Philipp this is for you, please delete after reading

The red dash baseline is the citywide average of our demographic data, weighted by population of each zip code. Then we divide each community's own weighted average by this baseline to obtain a ratio, so that they can share the same scale. To be more specific, in example 2, grey indicates the number, or equivalently, the "density" of PhD in the community. And in community 2 (mid Manhattan near Central Park) it's three times as many as city average.

Next we highlight some features in matching colors, showing what discrepancy among communities our clustering has detected. (Like, Manhattan has significant more people with higher education, higher income, expensive house, better insurance coverage, etc. but less children and less black people, who mainly live in the other 4 boroughs and show some preference to Brooklyn.) We may know some of this from common sense, but the picture not only shows the quantitative comparison other than a qualitative "more or less", but also implies a potential application of detecting some certain type of people quantitatively by simply using Twitter data if demographic data is unavailable.