Impact of the High Population Events on Urban Mobility: Who is Affected the Most?

Jane Roarty, Kieu Giang Nguyen, Son Phan

December 10, 2019

Abstract

This project explores mobility in Washington, D.C. in normal conditions and during highly populated events that create a shock to the transportation system. Washington is significant as the nation's capital and therefore a host of major events. Washington is also a considerably segregated city. Because of the frequent events and segregation, it is important to model mobility throughout the city and the metropolitan area. Specifically, this paper models normal traffic conditions in Washington and conditions during the 2017 inauguration and subsequent Women's March by using Spatial Autoregressive Models.

Introduction

Examining Washington D.C. during the 2017 women's march and presidential inauguration, this paper looks to examine the effect that large scale public events have on urban mobility. On January 20, 2017, Donald Trump was inaugurated in Washington. The following day, January 21, hundreds of thousands of people protested Trump's inauguration through the Women's March. The women's march in Washington has been called the largest protest in the United States with an estimated 470,000 people in attendance. Crowd scientists estimate that the crowd at the Trump inauguration was about around 160,000 people, a third of the size of the protest (Wallace and Parlapiano 2017). A population increase of this size over the course of just one weekend creates a huge shock to the transportation system. This can occur through the influx of people coming into the city as well as road blocks impeding traffic. This study looks at changes in the mean travel time from the center of D.C., with the center point defined as the Washington Monument. Before modeling this impact, however, we first modeled the impact of demographic census data on travel time during normal January conditions. We then were able to observe how demographics impact mobility and how those factors are exacerbated during major shocks to the system. We found that the number of roads, distance to the Washington Monument, and the number of black residents are consistently the most significant predictors of travel time to each polygon from the Washington Monument. Further, we found that ____________ ADD FINDINGS ABOUT SIGNIFICANCE DURING SHOCK!!!

It is important to examine urban mobility in two ways. First, city attributes such as road or transit access make certain areas more accessible. Unfortunately, accessibility or mobility often falls along racial or class lines. Washington is highly segregated and the cause of which is likely in part be due to urban planning that partitions the city into separate demographics.

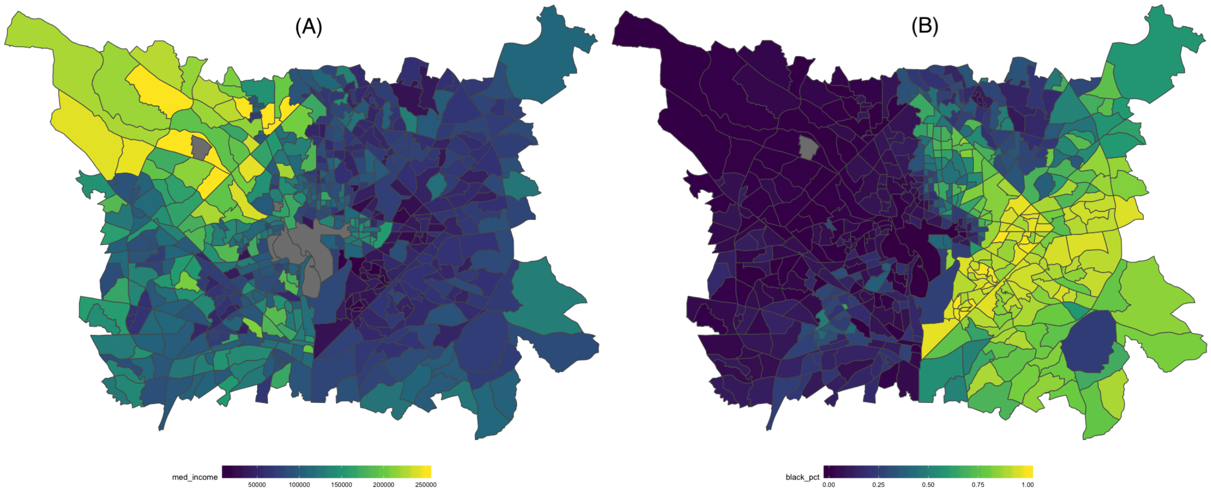

Figure 1: (A) - Median income by census tract; (B) - Percent of Black Citizens by census tract

As shown in figure 1 and 2, it is clear that there are sharp both economic and racial divides as shown by census demographics. Whether the partitions assigned to each demographic group affect their ability to access the center or downtown area of D.C. is a key question to answer if transportation policy is focused on helping different demographics. By using a simultaneous autoregressive model on travel time from the Washington Monument during normal times, we are able to see the impact of race on D.C. area mobility. We observed that the population of black residents in a census tract is one of the few significant variables in predicting travel time to the Washington Monument. Second, major increases in population during special events can shock the city's infrastructure. For this reason, we looked at Washington because it is a city that is especially significant during major national events such as the inauguration and therefore attracts huge numbers of additional visitors. We used a similar simultaneous autoregressive model for the week of the inauguration and Women's March and then modeled the differences between the normal weeks and the inauguration and march to observe the shock on the system. Through modeling both before and during the major events, we were able to observe both demographic significance of travel time and the impact of major population changes on the system.

In order to examine the shock from these massive events, we used areal Uber data which includes travel times between various census tract locations in the DC metropolitan area. To model travel times before and during the event periods, we isolated a location and looked only at the travel time between each census tract in the Washington metropolitan area and the Washington Monument. We chose the Washington Monument as the origin because it is in the middle of the National Mall, where many significant events, including the inauguration and Women's March, took place. Because the spatial features being analyzed are census tracts, we were able to include census information for each polygon through the `tidycensus` package (Walker 2018). This data allowed us to address the impact of demographics on travel time.

Literature Review

Many scholars have studied both mobility and shocks to the system in a variety of contexts. It is important to closely study every city's transportation system which is why many studies have been conducted. Noulas et. al use a network called Foursquare which records peoples locations to track and model mobility in a number of different cities (Noulas et al. 2012). Studies about mobility and accessibility have been conducted in many cities including New York City (Litman 2017) where concepts such as accessibility can be quantified and modeled to demonstrate lapses in the system. Shocks to transportation systems have also been modeled by a number of scholars. Donovan and Work's 2017 study uses taxi data to model the impacts of major events such as hurricane Sandy on New York city (Donovan and Work 2017). Wang also modeled the impact of Hurricane Sandy on New York city by looking at images of people throughout the city (Wang and Taylor 2014). Other major events modeled by scholars include the Olympics (Friedman et al. 2001) and the World

Cup (Menezes and Souza 2014). Transportation scholars have used a variety of data sources to observe the mobility of people including taxi data, satellite imagery, and cell phone data (Calabrese et al. 2010). Uber data is accessible and also provides an indication of mobility since so many people travel in cars. Because of this accessibility, we decided to model the effects of the inauguration and Women's March in Washington, D.C. that happened during the third weekend of January 2017.

Methods

Our goal in modeling the travel times of different census tracts from the Washington monument is twofold. First, we want to see what factors in general influence travel times; and second, we want to see if these factors determine whether a census tract is heavily affected by a signficant event such as the inauguration and Women's March. The census track features we consider during this analysis are the distance to Washington monument, median income, asian, white, and black percentage demographics, percent of people who use cars as their primary mode of transportation versus public transit, number of primary-secondary roads (see ACS Technical Documentation, Page 3-42), and number of transit lines. The purpose of using several features that may distinguish census tracts and specific methods is to find non-confounded estimates and accurately infer what affects travel times.

We first consider whether or not our data is spatially autocorrelated. For example, neighboring census tracts likely have very similar travel times and demographics simply by virtue of being physically near each other. If this correlation was not present, we would see that the travel times between neighboring census tracts have no relation at all. Under spatially autocorrelated conditions, the standard errors of regression become inflated and thus a correlation structure based on neighboring polygons is necessary. Understandably, another equally important choice involved in this analysis is how to define neighbors.

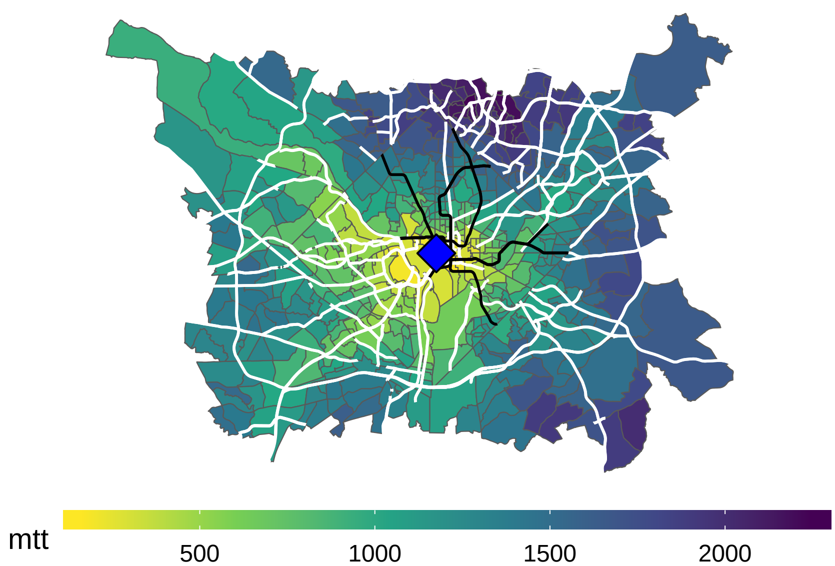

Figure 2: Mean travel time to Washington Monument in seconds. White lines are primary and secondary roads, black lines are Metro Transit rail lines, and the blue diamond labels the monument

As is visible in figure 3, many of the tracts neighboring each other have very similar mean travel times (MTT). The addition of transit and roadways also brings up the notion that neighboring tracts have access to similar roadways; this is especially true for the many roadways that coincide with census tract borders. We further justify the hypothesis of spatial autocorrelation through a Moran's I test. The null hypothesis of Moran's I test is that there is complete spatial randomness (CSR) which implies no spatial autocorrelation. We observe a p-value of 0 for the Moran's I statistic and thus reject the null hypothesis of CSR. In turn, we conclude the presence of spatial autocorrelation and justify the need to impose a neighborhood correlation structure in our regression models. We define two tracts to be neighbors if they simply share a portion of their border. We decide on this neighborhood structure, as opposed to within-distance and k-nearest-neighbors structures, due to the relative size of D.C's census tracts and the lack of tracts that share very minimal borders. Since tracts are relatively small, we can be confident that every bordering neighbor is physically close to any point in a tract. Also because there are not tracts that share minimal borders, we can exclude the possibility of edge cases such as where long and narrow tracts neighbor with each other at one point.

This analysis makes use of simultaneous autoregressive models (SAR). We choose this model over other longitudinal regression because it is able to take into account a neighborhood correlation structure such as the the border-sharing relation we consider. Additionally, we choose this model over other spatial autoregressive models such as CARs because the specification is relatively simple to comprehend.

Travel Times in General

We first look how travel times are related to the selected census track features two weeks before the inaugural weekend. The response, $\mathbf{Y}$, is the mean travel time to each census tract from the Washington Monument aggregated amongst the first two weeks of January. This response is chosen because it is representative of normal travel times in January and can thus be compared to the travel times during the inaugural weekend. The model should allow us to estimate a general travel time link to demographics, roads, and transit. Define $\mathbf{X}$ to be the matrix variables containing distance to the monument, median income, number of primary-secondary roads, number of transit lines, and demographic percentages for all census tracts. Finally, let W be the binary adjacency matrix such that $w_{i,j}$, the i^{th} entry in the j^{th} column, is 1 if i and j are neighbors and 0 otherwise. The SAR model is thus:

$$Y = \underbrace{\lambda WY}_{\text{Autoregressive Term}} + \underbrace{\mathbf{X}\beta}_{\text{Predictors}} + \underbrace{\epsilon}_{\text{Random Error}}, \qquad \epsilon \sim N(0, \sigma^2 I) \tag{1}$$

As shown in equation 1, the estimate of travel time Y involves a term of travel time from all neighbors WY weighted by an autoregressive parameter λ, the independent predictors $\mathbf{X}\beta$, and a random error. This is intuitively simple as travel time in a census tract is a linear function of our chosen predictors and the travel time of its neighbors.

Travel Times as Affected by Inauguration Weekend

When looking at how travel times changed during inauguration weekend we are limited to cross-sectional variation in gauging our parameters. The response in this case is the difference in mean travel time during the third and fourth weeks of January and mean travel time the previous two weeks of January. We chose to model the difference in travel time because the SAR model can not readily capture time series variation of our data. Modeling differences in the SAR model allows us to capture a two-period variation, but limits us in what kinds of precise time-series effects in travel time we can identify. dependent variable represents the aggregated difference before and after inauguration day and the Women's March. Therefore, applying the same specification as in equation 1 would let us estimate the effect of our regressors on how a census tract's travel time is affected by the influx caused by the inaugural weekend.

Results

Referring to table 1, our first spatial autoregressive model looks at how mean travel time is affected by the estimated demographics in each census tract before the inauguration. The coefficients `med_income`, `asian_pct`, `public_transit_pct`, `car_transit_pct`, `n_roads`, `n_transit` are all negative, implying that it takes longer to travel from the Washington Monument to the polygons in which there are lower median income, lower percentage of Asian population, and lower number of public transit, car transit, roads, and metro lines. It is usually the case that the city attributes such as `public_transit_pct`, `car_transit_pct`, `n_roads`, and `n_transit` have an immediate impact on transportation and thus can directly account for the variation in travel times, whereas the link between demographic factors in a region and travel times might not be as clear. One possible explanation is that the areas where people have lower income are often further away from the city center and less accessible by public transport, which altogether results in longer

Table 1: Output of SAR model for travel time before inaugural weekend

	Estimates	Std. Error	Z score	P value
(Intercept)	1.1264e+03	1.7597e+02	6.4009	1.544e-10
med_income	-2.0429e-04	1.7572e-04	-1.1626	0.2450012
asian_pct	-1.8151e+02	1.2842e+02	-1.4134	0.1575314
black_pct	2.2741e+00	7.3356e+01	0.0310	0.9752694
white_pct	-1.8247e+01	7.7101e+01	-0.2367	0.8129190
public_transit_pct	-5.4616e+01	9.6774e+01	-0.5644	0.5725050
car_transit_pct	-1.3932e+02	7.5311e+01	-1.8500	0.0643184
n_roads	-7.2906e+00	1.1736e+00	-6.2122	5.225e-10
n_transit	-2.0245e+01	7.6623e+00	-2.6422	0.0082363
distance	2.2801e-02	6.5935e-03	3.4582	0.0005438

travel times and decreased mobility. The coefficients `white_pct`, `black_pct`, and `distance` are all positive, implying that it is faster to get to the closer polygons or those in which there is a lower percentage of white and black residents from the Washington Monument. Although it makes intuitive sense that distance is positively correlated with time, it does not come as natural to think about why there might be such positive relationship between racial proportion of whites and blacks with travel times. Based on the visualizations, the fact that the predominantly black areas are seen further east while the predominantly white areas are seen further west indicates that the variation observed in mean travel times can be explained by the variation in geographical distance and inherent accessibility in areas dominated by certain races. Furthermore, it is not surprising that `n_roads`, `n_transit`, and `distance` all have significant *p*-values as they are generally the primary predictors that contribute to the estimation of travel times.

```
##
## Call:
## spautolm(formula = MeanTimeDiff ~ med_income + asian_pct + black_pct +
##     white_pct + public_transit_pct + car_transit_pct + n_roads +
##     n_transit + distance, data = travel_time_diff, listw = listW,
##     family = "SAR")
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -369.37822  -14.24845   -0.68549   15.53629  290.00435
##
## Coefficients:
##                      Estimate  Std. Error z value Pr(>|z|)
## (Intercept)         6.9169e+01  2.8516e+01  2.4256  0.01528
## med_income         -1.8596e-05  7.3212e-05 -0.2540  0.79949
## asian_pct          -1.0108e+02  4.8648e+01 -2.0777  0.03774
## black_pct          -2.5959e+01  2.6756e+01 -0.9702  0.33194
## white_pct          -2.2358e+01  2.9300e+01 -0.7631  0.44542
## public_transit_pct  4.6175e+01  3.8256e+01  1.2070  0.22743
## car_transit_pct    -1.3922e+01  3.0224e+01 -0.4606  0.64506
## n_roads             3.7878e-01  4.7338e-01  0.8002  0.42361
## n_transit          -3.0828e-01  3.1525e+00 -0.0978  0.92210
## distance           -9.1644e-04  1.0881e-03 -0.8423  0.39964
##
## Lambda: 0.59874 LR test value: 106.43 p-value: < 2.22e-16
## Numerical Hessian standard error of lambda: 0.048325
##
## Log likelihood: -2879.096
## ML residual variance (sigma squared): 1913, (sigma: 43.738)
```

```
## Number of observations: 550
## Number of parameters estimated: 12
## AIC: 5782.2
```

Table 2: SAR model - difference in travel time

	Estimates	Std. Error	Z score	P value
(Intercept)	6.9169e+01	2.8516e+01	2.4256	0.01528
med_income	-1.8596e-05	7.3212e-05	-0.2540	0.79949
asian_pct	-1.0108e+02	4.8648e+01	-2.0777	0.03774
black_pct	-2.5959e+01	2.6756e+01	-0.9702	0.33194
white_pct	-2.2358e+01	2.9300e+01	-0.7631	0.44542
public_transit_pct	4.6175e+01	3.8256e+01	1.2070	0.22743
car_transit_pct	-1.3922e+01	3.0224e+01	-0.4606	0.64506
n_roads	3.7878e-01	4.7338e-01	0.8002	0.42361
n_transit	-3.0828e-01	3.1525e+00	-0.0978	0.92210
distance	-9.1644e-04	1.0881e-03	-0.8423	0.39964

Now we consider our second spatial autoregressive model shown in table 2 which looks at how mean travel times are affected by the estimated demographics in each census tract during and after inauguration. The coefficients from the second model stay consistent with our observation in the first model, that `white_pct`, `black_pct`, and `distance` are positive while `med_income_pct`, `asian_pct`,`public_transit_pct`, `car_transit_pct`, `n_roads`, `n_transit` are negative. Then, we fit a spatial autoregressive model of the difference in mean travel times between the two time periods with the demographic variables to see how urban mobility is affected by the influx of people coming to Washington D.C. during the shock events. Here, while most of the coefficients match with what we see above, `public_transit`, `n_roads`, and `n_transit` are indeed positive rather than negative. This implies that the difference in mean travel times increases as the number of roads in a polygon increases, which seems counterintuitive at first since we expect that these factors should increase transportation efficiency, however, rising traffic jam due to the surge in people coming to these massive events might be the potential cause of decreased mobility regardless. Overall, we find no indicator of a significant difference in travel times as well as a direct impact of the demographics on travel times before and after the shock events in Washington D.C.

Conclusions

The insignificant difference in travel times during these major shocks can be explained by two main reasons: the unchanging state of demographics in a short time span and the efficiency of transportation system in Washington D.C. Since we only look at travel times in a two-week span, it is likely that the demographic census data stays unchanged and thus not having any effect on the city mobility during those times. Additionally, Washington D.C. is known to have one of the safest, cleanest, and most efficient transportation systems nationwide so the increase in population from the large scale public events is also unlikely to act on its urban mobility. Nonetheless, the models give us interesting insights into the role that socioeconomic status and racial background play in geographical mobility, specifically in the case of Washington D.C.

There are a couple of limitations to our model that limit its performance. First, there are infinitely many unobservable characteristics that might affect urban mobility and our data might not perfectly capture those aspects, which leads to variable bias in these results. More importantly, we did not take into accouunt the fact that Washington D.C. is considerably segregated when computing the spatial covariance matrix, so the estimates do not fully represent the correlation between demographics and urban mobility. All in all, the contextual limitations of our chosen model outlined above will bias the estimates and push them further from the truth.

Calabrese, Francesco, Francisco C Pereira, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. 2010. "The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events." In *International Conference on Pervasive Computing*, 22–37. Springer.

Donovan, Brian, and Daniel B Work. 2017. "Empirically Quantifying City-Scale Transportation System Resilience to Extreme Events." *Transportation Research Part C: Emerging Technologies* 79. Elsevier: 333–46.

Friedman, Michael S, Kenneth E Powell, Lori Hutwagner, LeRoy M Graham, and W Gerald Teague. 2001. "Impact of Changes in Transportation and Commuting Behaviors During the 1996 Summer Olympic Games in Atlanta on Air Quality and Childhood Asthma." *Jama* 285 (7). American Medical Association: 897–905.

Litman, Todd. 2017. *Evaluating Accessibility for Transport Planning.* Victoria Transport Policy Institute.

Menezes, Thais Rubens de, and João Figueira de Souza. 2014. "Transportation and Urban Mobility in Mega-Events: The Case of Recife." *Procedia-Social and Behavioral Sciences* 162. Elsevier: 218–27.

Noulas, Anastasios, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. 2012. "A Tale of Many Cities: Universal Patterns in Human Urban Mobility." *PloS One* 7 (5). Public Library of Science: e37027.

Walker, Kyle. 2018. "Tidycensus: Load Us Census Boundary and Attribute Data as' Tidyverse'and'sf'-Ready Data Frames." *URL Https://CRAN. R-Project. Org/Package= Tidycensus. R Package Version 0.4* 1.

Wallace, Tim, and Alicia Parlapiano. 2017. "Crowd Scientists Say Women's March in Washington Had 3 Times as Many People as Trump's Inauguration." *The New York Times.* The New York Times. https://www.nytimes.com/interactive/2017/01/22/us/politics/womens-march-trump-crowd-estimates.html?searchResultPosition=1.

Wang, Qi, and John E Taylor. 2014. "Quantifying Human Mobility Perturbation and Resilience in Hurricane Sandy." *PLoS One* 9 (11). Public Library of Science: e112608.