Segregation Empirical Work

Erik B. Johnson

23 April, 2019

Description

Data work and documentation for the Richmond transportation/segregation paper.

Background

Really seems that we should focus on the commuting component of transportation (empirical and theoretical reasons.)

This is a list of possible sources to help motivate our paper.

- Commuting to Opportunity
- Commuting in America
- Low income commuters and Cycling
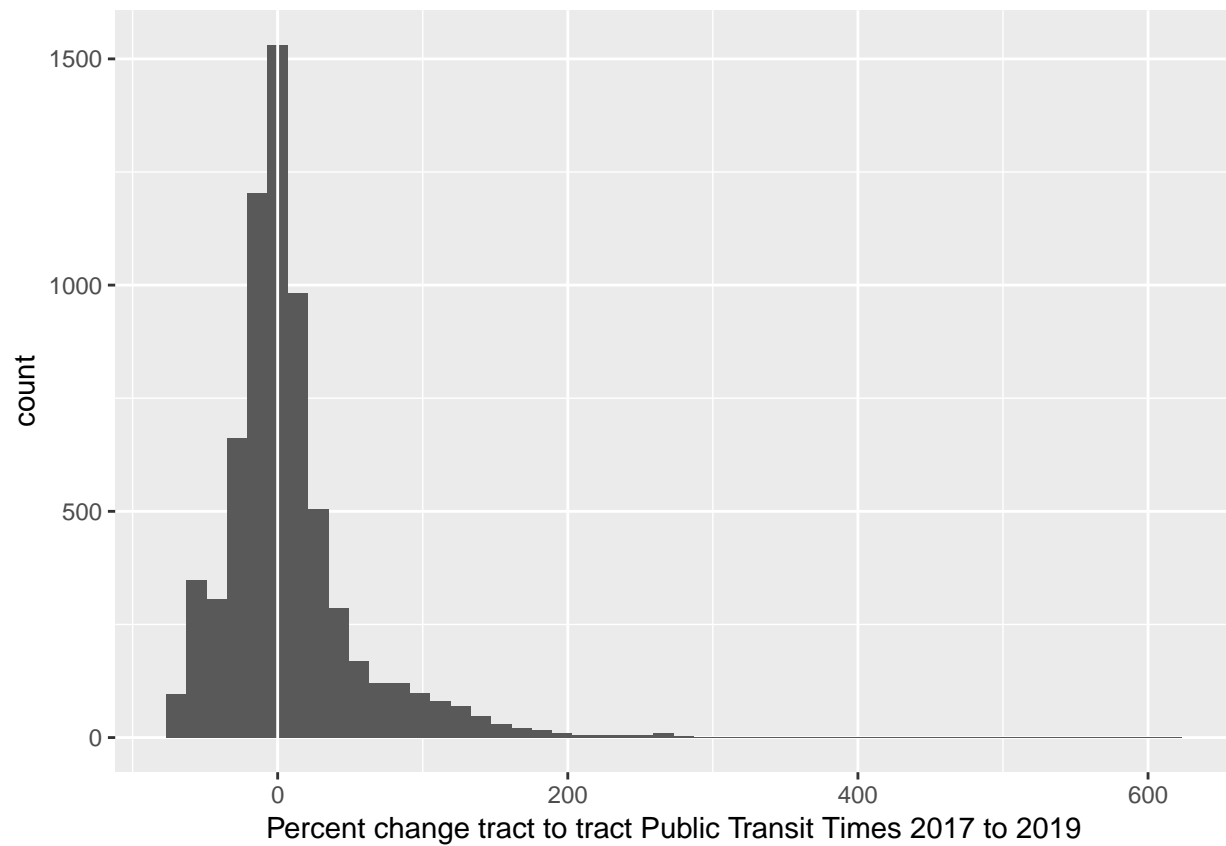
Data

Change in Travel times

First, use google travel times to build **dists.richmond** for distances (meters) and travel times (seconds) by mode to and from all census tracts. Based on tract-centroids to tract-centroids. Distance is non-euclidean. For google distance and time calculation documentation see: Google distance api documentation

Summary statistics for **dists.richmond** (where NA transit values default to walking values):

Table 1: Pairwise Distance Summary

Statistic	Mean	St. Dev.	Min	Max
driving_03-2017	1,347	514	95	3,286
driving_03-2018	1,347	514	95	3,286
transit_03-2017	14,107	9,022	406	47,240
transit_03-2018	14,176	9,132	406	49,558
walking_03-2017	15,505	8,284	406	47,240
walking_03-2018	15,505	8,284	406	47,240

Change in Travel times


```
## Saving 6.5 x 4.5 in image
```

Which tracts have the largest changes in travel times?
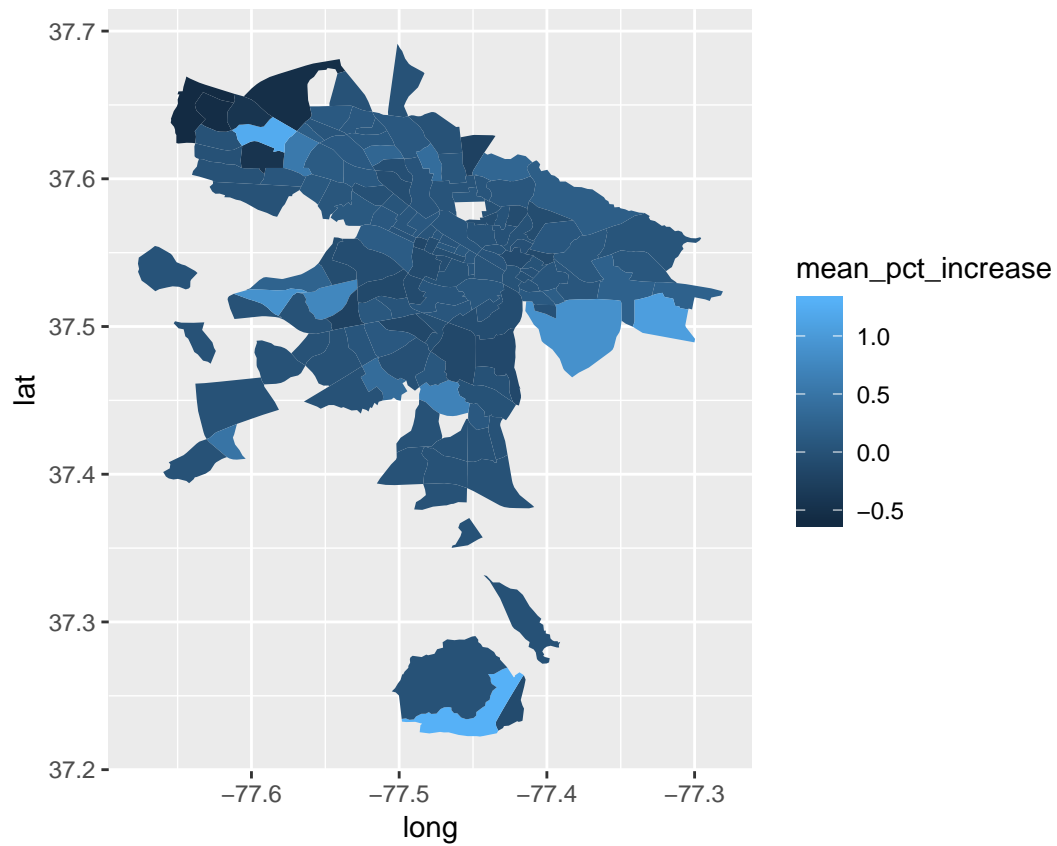
Census [census]

The **census** dataset consists of the following fields:

The construction of the census table is documented in Table **??**. For more specifics see *funCensus.income, funCensus.commute, funCensus.race* in the file `functions.segregation.R`.

Measuring segregation

We begin by estimating the amount of segregation in the city with a variety of traditional segregation measures (from the R library `seg`)[1]. Interestingly, there have been a variety of measures which include a variety of spatial terms that uses information on neighbors and shared borders. These measures are, however, fundamentally different from our new one since spatial distance is a matrix that incorporates a variety travel times between tracts over the entire city.

Dissimilarity

We begin by calculating a simple dissimlarity index between two groups X and Y in locations i described in Equation (1). Higher values of dissimilarity imply more wihtin tract race distributions. Note again that this measure is inherently aspatial and only uses the tract level census data. Note that the 'nb' term in `seg` library scales the interaction of the iteraction is normalized to 1 and not appropriate for our application. Additional information on the library can be found at the Stanford Dissimilarity. Empirical results are shown in Table **??**. We can see that the most spatially dissimilar races according to this measure are with a value of

$$D = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{x_i}{X} - \frac{y_i}{Y} \right| \tag{1}$$

Wasserstein Measure

```
l.wasserstein <- funMeasures.wasserstein(census, dists.richmond)
```

The two main drawbacks of the D measure are the lack of spatial information (distance) between populations and the fact that there is no direction implied in the relationship. Next, we measure dissimilarity through a Wasserstein measure, which will include both the spatial information and has the ability to infer directional relationships in the form of an asymmetric graph. This is a two stage process and requires careful selection of counterfactuals. We begin with the most simple formulation[2].

---

[1]Documentation and explaination at (http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0113767) and Reardon and O'Sullivan (2004)

[2]The Wasserstein measure is found in the R `transport` package.