Segregation Empirical Work

Erik B. Johnson and Santiago Pinto

13 February, 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

Google [dists.richmond]

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	Pctl(25)	Pctl(75)	Max
driving.meters	23,347	12,335	597	13,776.8	31,400	71,106
driving.seconds	1,347	514	95	979	1,675	3,286
transit.meters	22,333	11,447	536	14,351.5	28,845.2	98,403
transit.seconds	14,107	9,022	406	5,783.8	20,470.2	47,240
walking.meters	20,665	11,084	536	12,070.5	27,806	62,833
walking.seconds	15,505	8,284	406	9,085	20,866.2	47,240

Correlation matrix for **dists.richmond**

Table 2: Distance Correlation Matrix

	driving.m	driving.s	transit.m	transit.s	walking.m	walking.s
driving.meters	1	0.945	0.898	0.903	0.967	0.968
driving.seconds	0.945	1	0.868	0.870	0.933	0.933
transit.meters	0.898	0.868	1	0.793	0.927	0.927
transit.seconds	0.903	0.870	0.793	1	0.939	0.939
walking.meters	0.967	0.933	0.927	0.939	1	1.000
walking.seconds	0.968	0.933	0.927	0.939	1.000	1

```
knitr::kable(correlation.matrix)
```

	driving.m	driving.s	transit.m	transit.s	walking.m	walking.s
driving.meters	1.0000000	0.9453170	0.8976362	0.9025457	0.9673928	0.9682031
driving.seconds	0.9453170	1.0000000	0.8677223	0.8696202	0.9331407	0.9331496
transit.meters	0.8976362	0.8677223	1.0000000	0.7932233	0.9270210	0.9272311
transit.seconds	0.9025457	0.8696202	0.7932233	1.0000000	0.9392204	0.9385725
walking.meters	0.9673928	0.9331407	0.9270210	0.9392204	1.0000000	0.9998509
walking.seconds	0.9682031	0.9331496	0.9272311	0.9385725	0.9998509	1.0000000

	driving.m	driving.s	transit.m	transit.s	walking.m	walking.s

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.