Data Needs for Corridor Study

Jamaal Green

October 17, 2017

As we move into expanding our study site beyond the Portland pilot we need to have the appropriate data available for analysis. What follows is a brief description of the necessary data sources. Many of these datasets are public, but a few are proprietary or your city may not normally make them available. In such a case, we would ask your assistance in gaining this data in a timely manner.

Spatial Data

What follows are datasets that are explicitly *spatial* in nature and we would require them in an existing spatial format such as in a shapefile or GeoJSON file:

1. Infrastructure

- Bike network- bike network data that includes the construction data for segments
- Improved sidewalk/road calming- road calming or improved sidewalk segments if of interest

2. Buildings and Neighborhood

- Building outline/parcel data- assuming we have point data for establishments we can more easily designate improved corridors by joining establishments to existing parcel and building data. This is also important for trying to identify first floor businesses
- Neighborhood or zip code boundary files for the city- neighborhood boundaries allow for additional identifcation variables for corridors and help with overall final map products

Tabular Data

These are data that do not need to be given to us in a spatial format. What follows are largely economic/business related

1. Economic and Sales Data

- QCEW/ES-202- QCEW (Quarterly Census of Employment and Wages) is a quarterly survey administered by each state's employment office for establishments that pay into unemployment insurance and submitted to the Bureau of Labor Statistics. Some states allow the usage of the confidential data they collect for research purposes.
  - For example, Oregon allows qualified researchers to use the confidential data and it includes an establishment ID number, an address, XY coordinates, and employment and salary estimates for each month. This data would be preferable for estimating employment impacts because it is at the establishment level and longitudinal level and is public. Additionally, many employment departments map employment using this data meaning it may exist in a spatial format. If so, the spatial format would be preferred.
  - This example table shows what this dataset may look like. Note the case id number specific to the establishment to allow for tracking over time, full address, monthly emplyment numbers, and XY coordinates for mapping.

Table 1: Sample QCEW Data (continued below)

CaseID	NAICS	BizName	Address	City	State	Zip
1098	586901	Coffee Galore	123 Fairfield	Portland	OR	97212

CaseID	NAICS	BizName	Address	City	State	Zip
2304	586902	Doughnuts Now	346 Broadway	Portland	OR	97210
3704	586901	Bar Bar	769 Burnside	Portland	OR	97210

Jan	Feb	...Dec	X	Y
12	20	18	122.1	45.2
4	6	5	122.1	45.3
12	13	13	122.1	45.4

- Retail sales tax data- for cities that do collect a sales tax, sales tax data offers a rich alternative data source for examining economic impacts. Such data exists at the establishment level and is longitudinal. While this may not be available for all cities, gathering such data for cities where it is applicable gives us another metric to test the impact of streetscape improvements that may not be immediately apparent using employment data
  - Simlar to QCEW table example, sales tax data should ideally have an ID that follows particular establishments over time, address information, ideally either XY coordinates or a building parcel number for mapping, and estimates of sales tax collected, at minimum, annually

Note that while we have divided this into two forms of data- tabular and spatial- that the QCEW and sales tax data are also "spatial" in the sense that they include address information and even XY coordinates for mapping. In fact, for Oregon, the QCEW data is often shipped as a shapefile so cities can better map changes in the spatial distribution of economic activity and to support planning efforts. In this case, there is no real distinction between spatial and tabular data.