Exam 2 Solutions

PUBPOL 2130 / INFO 3130
Moon Duchin, Spring 2025

(1) In the `Week6` (elections) notebook, we start working with Massachusetts data and at a certain point our geodataframe looks like this:

	TOWN	WARD	PRECINCT	VTD	geometry
0	Braintree	None	5B	5B	POLYGON ((242364.517 883741.979, 242359.642 88...
1	Braintree	None	6A	6A	POLYGON ((243479.798 881994.897, 243477.538 88...
2	Braintree	None	6B	6B	POLYGON ((241614.364 881551.551, 241602.936 88...
3	Chelsea	1	1	1-1	POLYGON ((239281.036 904133.287, 239277.513 90...
4	Chelsea	1	2	1-2	POLYGON ((238408.578 904480.138, 238393.140 90...
...	...	...	...	...	...

Then we have the following code block.

```
ma_ward_precinct_gdf["Name"] = (
ma_ward_precinct_gdf["TOWN"] +
" Ward " + ma_ward_precinct_gdf["WARD"].astype(str) +
" Precinct " + ma_ward_precinct_gdf["PRECINCT"].astype(str)
)

len(ma_ward_precinct_gdf["Name"].unique()) == len(ma_ward_precinct_gdf)
```

(a) What would be the 0-indexed entry of this new "Name" column?

Braintree Ward None Precinct 5B

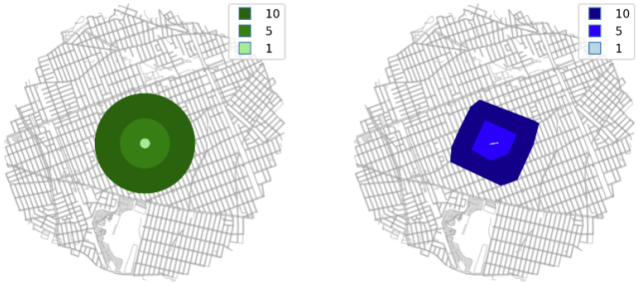
(b) What are we doing here (including in the last line), and why is it important for the workflow?

We are creating a *unique identifier*. The point is that the initial index column (0,1,2,...) is not informative, and none of the other individual columns is in a 1-1 relationship with the precincts. The "precinct" column is sure to have repeats, as many towns will have a Precinct 1. So we create a compound name that uniquely identifies the precincts.

In the last line, `len` is length, and we're checking that the length of the list of unique Names is the same as the overall number of precincts by using an assertion of equality. If it returns `True`, that confirms that there are no repeats.

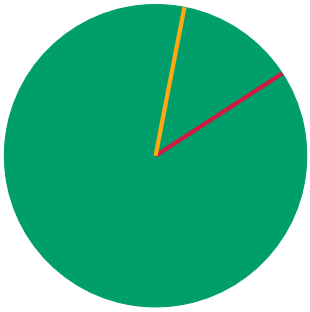
Name: _____

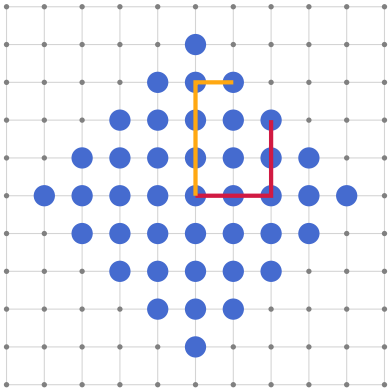
(2) In the `Week9` (transit) notebook, we produce these plots.



One of these used the OSMnx package to construct a plot with *network* data. Which one, and how is that related to one set of isochrones looking round, while the other is roughly diamond-shaped?

A "network" is a graph – it has nodes and edges. In this case, the plot on the right treats the city as a network of *road segments* connecting *intersections*. We find areas reachable in 1, 5, and 10 minutes by assuming a constant walking speed. It is a diamond because this part of NYC is nearly a grid (see pic on the right, which shows where you can get in 4 steps on a grid). On the other hand, if you used "crow-flies" (Euclidean) distance, then all the places you can get within a fixed distance would form a round disk.





Name: _____

For questions 3-7, your answer should be BRIEF—about three sentences! Succinct answers preferred.

(3) On the Decennial Census (short form), respondents can fill in options within Asian identity, such as a checkbox for "Korean." But that level of detail doesn't appear in the released Census data, which just uses the coarser category of Asian. How would we access that data so that we can merge it with other sources to study Korean-American migration, employment, voting, etc?

This kind of detailed, raw data is called **microdata**, and the most obvious place to access it is through PUMS (public use microdata samples). There is an institute called IPUMS based at the University of Minnesota which cleans and releases these data, aggregated to larger geographies called PUMAs.

(4) The Great Migration saw millions of Black people move out of the South; since 1970, there is a smaller but very clear "reverse migration," with Black migration back to Southern locations, especially cities. Enrico Moretti (in *The inequality of mobility and cost of living*) has several theories of why people do and do not move. Does Moretti's framework fit better with the exodus from the South, or with the return to the South?

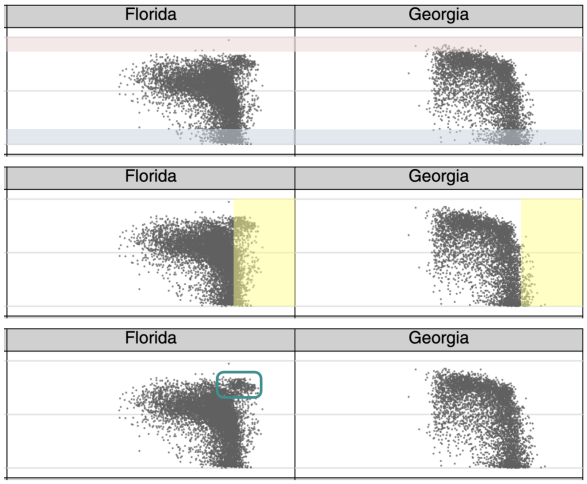
Here's a long answer for thoroughness! Making just a few of these points would earn a full score.

Isabel Wilkerson gave us a description of the reasons driving the exodus from the South in which she emphasized that alongside the search for employment (largely factory employment in Northern and Midwestern cities) there was also a desire to escape violence and discrimination under Jim Crow laws and practices. On the other hand, Frey's report on "The New Great Migration" shows that the South has seen a large population influx since 1970, and within that, Black people are somewhat more likely than White people to choose the South as a new home. This new migration is led by college-educated individuals, and Frey calls it a "brain gain."

Moretti tells us that educated people with disposable income are more likely to move to better employment markets or those with low cost of living; he also says that family ties and cultural expectations of closeness to family can be a major factor. This is not an obvious fit for the Great Migration, which was driven by need and uprooted people from their family historical homes; we have no evidence of higher education levels among those migrants, and poverty/sharecropping meant likely not much disposable income. It is easier to fit this framework to the Great **Return** because of Frey's stats on education and the low cost of living attracting new Southward migrants. However, none of that explains why Black people are moving South more than White folks are, so it would also be reasonable to say that **neither** shift is especially Morettian.

Name: _____

(5) In these plots from Jonathan Rodden's work (discussed in class), the x axis is log population density and the y axis is the vote share for George W. Bush (Republican presidential candidate) in 2000. First, give a short explanation of what the scatterplots show. Then, discuss one point or argument from Rodden's book chapter or interview that is closely related to the plot.



These plots have a lot of information!! Here are three observations that I've marked up the plots to highlight. (a) There are many, many block groups that vote nearly 100% D, but very few in either state that vote near 100% R. (b) Interestingly, the conventional story of a straightforward correlation between density and partisanship is not the main story here. The "L" shape of these plots means that higher-density places have a *range* of voting behavior, while the least dense places favor Bush. The most densely populated block groups in Georgia are consistently D overall, but in Florida it's less clear. (c) There's a cluster of block groups in Florida that is very dense but clearly votes Republican (circled in the plot).

For connections to Rodden, you've got lots of choice: (a) Democrats are "wasting votes" in both cities and the countryside! They're winning by a lot in the cities, but they're also casting lots of losing votes in rural areas. (b) The areas voting 90 or 95% D tend to be urban and post-industrial. Dense housing and good infrastructure drew some left-leaning voters and *made* others. (c) One of only two dense neighborhoods Rodden mentions in the whole world that votes with the Right is in Florida—it's Little Havana, Miami!

(6) Kieran Healy writes that "Commensuration is *the expression or measurement of characteristics normally represented by different units according to a common metric.*" We used the concept of commensurability earlier, when discussing how SIPRI's TIV metric makes different heavy weapons comparable. Healy is pointing out that it is hard to measure the value of a "gift," because there is no clear common metric that takes all of the factors into account. When it comes to organs, what are some factors he discusses that make it hard to measure the value of an organ donation?

Here are a few concrete Healy points: There *is no* production market for human blood and organs. (This can complicate priceability because the market is often used to set a price.) Huge variability in availability (around the world, and year-to-year) is a sign that we don't understand people's incentives and "bundles of motives." And because bodies are intimate and spiritual, authorities have an incentive to avoid commodification, which can lead to exploitation.

Name: _____

(7) In class (Lectures 13-14) we discussed *linear programming* as a style of problem-solving that people from math and OR often use to solve policy and planning problems. We went over at least three examples of problems that researchers had solved with LP (or the related techniques used for integer programming). What does the phrase **objective function** mean? What objectives were used in Sommer Gentry's work? (If you don't remember, then give the objective function for any other planning problem we discussed, like the diet problem or COVID hospital/dorm matching.)

A linear program is a problem in which you seek to optimize a linear expression, subject to some constraints, and the *objective function* is the quantity to be optimized. Sommer Gentry proposed a redistricting solution for liver donation that was based on linear programming, where she discussed two objectives: minimize pre-transplant deaths (save lives) or minimize geographic variation (be most fair). Her solution was to focus on saving lives, and she found that her team got greatly improved geographic equity without having to take it as an explicit goal.

Name: _____