Research Question

The high level research question is to look at the effect of reinforcing identity within Hispanic populations on their schooling outcomes. Specifically, I'll be using the influence of Spanish language television as the channel by which identity is reinforced, and look at how it affects everything from graduation rates to disciplinary action taken to math abilities and English proficiency for Hispanic students in public schools. In short, if I have access to more programming from my home country, does this make me less engaged in school (perhaps because there are more distractions or because it socially ostracizes me etc.), or does this make me perform better (perhaps because l have more role models or because I have something to talk with peers about in school, and hence motivation to attend/perform)?

Why does this matter?

There's good reason to believe that identity, as reinforced through mass media, has a large effect on the lives people lead. Oberholzer-Gee, Waldfogel (AER 2009) demonstrate that the presence of Spanish language local news increases Hispanic voter turnout, while Yanigazawa-Drott (QJE 2014) shows that radio broadcasts in Rwanda contributed to the violence and genocide that took place in the 90s. It would be reasonable to think then, that there could be a meaningful effect of Spanish language TV on education.

Method

To isolate the causal effect of Spanish language television, I adopt the technique used in Newman, Velez (AJPS 2019) and generalize it from three counties to the entirety of the US. Newman and Velez exploit a FCC (Federal Communications Commission) regulation which determines the distance from a TV station in which the station's broadcast signal is protected from interference. This creates a natural regression discontinuity, where the decaying strength of a signal over distance is combined with this cutoff in broadcast protection to create a split among people just inside and outside these coverage 'contours' that are presumably comparable save for their access to broadcast TV.

In the case of Spanish language TV in particular, this should allow me to examine its causal effect on Hispanic populations for spatially located outcomes, such as public schooling results. It's worth noting that these contours are purely determined by an algorithm that looks at things like local elevation and antennae strength, so that the cutoffs are located in more or less random locations, and that coverage is large enough that these contours tend to cut across towns and suburbs, rather than cities. Finally, regressions using US census data indicate that Hispanic people do not migrate across counties in response to these contours.

A standard regression thus looks like restricting the universe of schools to only those within a small radius of the contour boundary, where the key independent variable of interest is an indicator for the school being inside or outside the boundary.

Data Sources

I have already collected all (or probably all) of the data that I will need for this project.

Data for the instrument comes from both the FCC and TMS (a telecommunications company that was kind enough to let me use their API for free), and the instrument is fully constructed. The relevant data here is essentially just the coverage contour spatial data and the broadcast language of the station.

The data on public schools comes from the US government's CRDC (Civil Rights Data Collection) dataset. It's a very large dataset with a ton of outcome/control variables, but importantly, it breaks down all major variables of interest by ethnicity. These variables includes graduation rates, chronic absenteeism, suspensions, expulsions, arrests, bullying, AP test results, English proficiency, math class performances, gifted program enrolment etc., so I can look at effects on both the top and bottom end of the distribution, and examine potential mechanisms driving outcomes. These are all at the school level, and so I can run this through ArcGIS to get physical locations of these schools as well.

I get additional controls for things like population, income, density of Hispanic population etc. at the county level from IPUMS.

I don't have preliminary results yet, but am close to getting the school data in a clean enough shape where I can start running regressions.