POLS 7012 Final Exam Practice

YOUR NAME HERE

University of Georgia

email@email.email

Joseph T. Ornstein

University of Georgia

jornstein@uga.edu

December 08, 2020

Abstract

In this practice final exam, we will replicate some results from "Local demographic changes and US presidential voting, 2012 to 2016" (Hill, Hopkins, and Huber 2019). You can find the paper in the `papers/` folder, and the data you'll need in (you guessed it) the `data/` folder. Write you code in the chunks provided, then knit the document to a PDF.

1   Introduction

Hill, Hopkins, and Huber (2019) investigate whether support for Republican candidates in US presidential elections is correlated with demographic changes at the precinct level. They find, contrary to their expectations, that precinct-level increases in the Hispanic share of a population are not significantly associated with shifts in vote share from Obama in 2012 to Trump in 2016. Before you begin the replication, skim over the paper a bit. We'll be replicating Figure 1 and Table 1. Follow the instructions below, and for an extra challenge, you can complete any steps listed as "an extra challenge".

2   Data

The precinct-level data is a Stata file called "PrecinctData.dta". I downloaded it here. Start by reading the data file into memory. Consult the codebook for variable definitions.

3 Results

3.1 Replicating Figure 1

Figure 1 plots the relationship between changes in Hispanic population and changes in Republican vote share for a subset of the precincts, along with a fitted LOESS curve (Locally Estimated Scatterplot Smoothing; the default in `geom_smooth()`). To recreate this figure, we'll need to complete the following steps:

- Create four plots, named `p1`, `p2`, `p3`, and `p4`.
- Add a `geom_point()` and `geom_smooth()` layer to each plot with the appropriate variable mappings
- Color the LOESS fit green
- Add axis labels to the plot to match the labels from the paper

For an extra challenge, you can complete any combination of the following:

- Before creating each subplot, filter the data so that it only includes only the interior 5% to 95% quantiles of the x variable. (They mention this step in a footnote on page 4. Sneaky sneaky.) The `quantile()` function is useful here.
- Draw a random sample of 2,000 precincts for the scatter plot, like in the original figure (but keep the LOESS fit for the full dataset).
- Make the points semi-transparent

The `patchwork` library helps stitch together a bunch of ggplots into a grid, like in the original paper.[1] If you created `p1`, `p2`, `p3`, and `p4` correctly, the included code chunk will recreate Figure 1.

3.2 Replicating Table 1

In Table 1, the authors estimate a series of eight linear models to test whether the relationships from Figure 1 remain after conditioning on potential confounders. In each model, change in Hispanic share is negatively associated with change in Republican vote share from 2012 to 2016. We'll replicate the first three columns of that table. To do so, take the following steps:

- The first model, `lm1`, is a linear model with the change in GOP vote share as the outcome variable and change in the share of Hispanic population between 2011 and 2016 as the explanatory variable.
- The second model, `lm2`, includes "County Fixed Effects". This just means including a indicator (dummy) variable for each county. Recode `countyid` as a character or factor and include it in the model.

---

[1]I previously showed you `cowplot`. That would work too, but I wanted to try something new here.
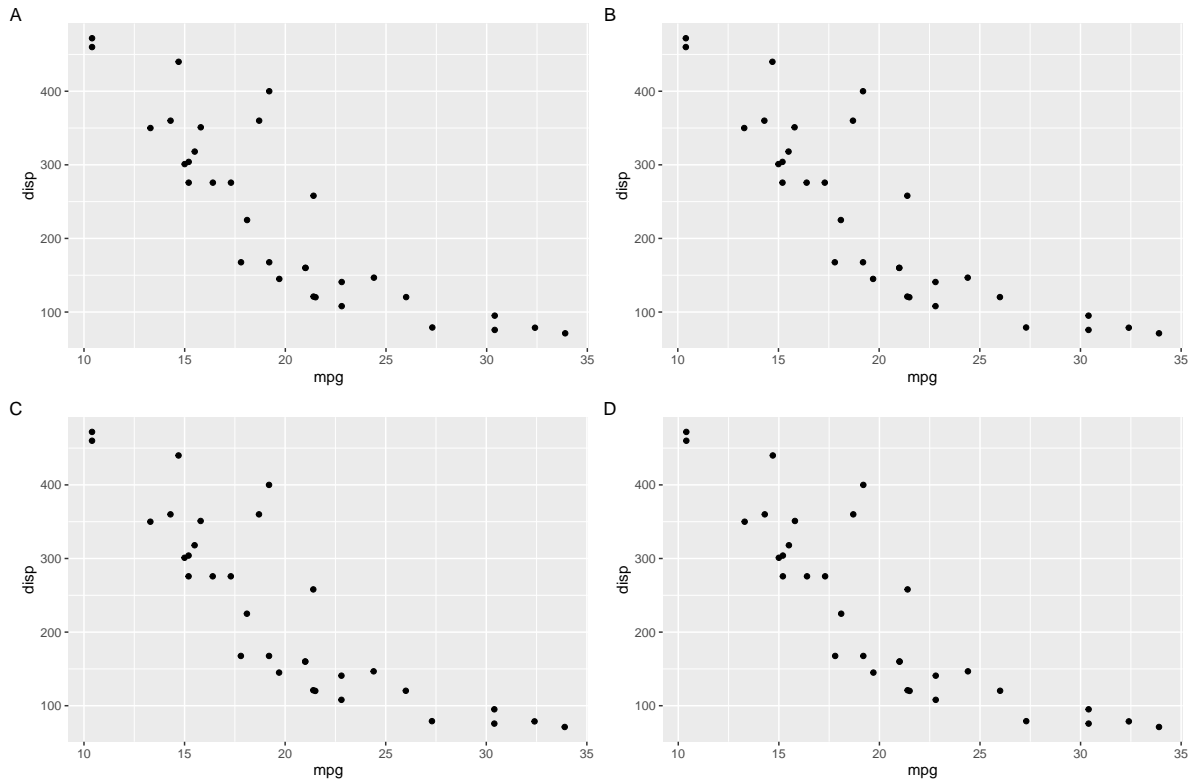
Figure 1: Change in Republican vote share, 2012 to 2016, and change in Hispanic population. Note: Points are random samples of 2,000 precincts. Loess lines are generated from all observations. Points are shaded corresponding to density, with darker colors indicating more precincts.

- `lm2` also includes a set of "Additional Census Controls" and "Controls For Levels". The full list of variables is on page 4; cross-reference the codebook for variable names.
- The third model, `lm3`, includes all the covariates as `lm2`, plus a set of indicator variables measuring the Republican 2012 vote share decile.
- Note that their models are estimated using weighted least squares (WLS) instead of orindary least squares (OLS). This procedure gives more weight to precincts with larger populations when estimating the best fit equation. It turns out not to matter much; you'll get pretty similar estimates with OLS. But to replicate their results, add the argument `weight = weights` to `lm()`.
- Don't worry if your standard errors don't exactly match up. You'll learn more about "robust" standard errors next semester.

The following code chunk uses the `stargazer` package to output a pretty regression table like the one in Table 1. I provide it to you free of charge. All you need to do is make sure that the model objects `lm1` through `lm3` are correctly specified. For an extra challenge, you can do any combination of the following:

- Try to replicate columns (4) through (8).
- Play with the `stargazer()` options to edit the layout and/or the covariate labels.
- Figure out why the number of observations and coefficient estimates are *slightly* different in `R` than the results the authors got from Stata. I genuinely have no idea. They may have copy-and-pasted wrong.

Table 1: Change in Republican vote share 2012 to 2016 and change in Hispanic population, various time intervals

	(1)	(2)	(3)
Observations	32	32	32
R^2	0.472	0.472	0.472

Note: *p$<$0.05; **p$<$0.01; ***p$<$0.001

References

Hill, Seth, Daniel J. Hopkins, and Gregory Huber. 2019. "Demographic Change, Threat, and Presidential Voting: Evidence from U.S. Electoral Precincts, 2012 to 2016." *PNAS*, 1–6. `https://doi.org/10.2139/ssrn.3351950`.