

PS6 homework

Ahmed

03/22/2022

Cleaning process

DIVISION	STATE	COUNTY	STNAME	CTYNAME	CENSUS2010POP	ESTIMATESBASE2010
6	01	000	Alabama	Alabama	4779736	4780118
6	01	001	Alabama	Autauga County	54571	54582
6	01	003	Alabama	Baldwin County	182265	182263
6	01	005	Alabama	Barbour County	27457	27454
6	01	007	Alabama	Bibb County	22915	22904
6	01	009	Alabama	Blount County	57322	57322

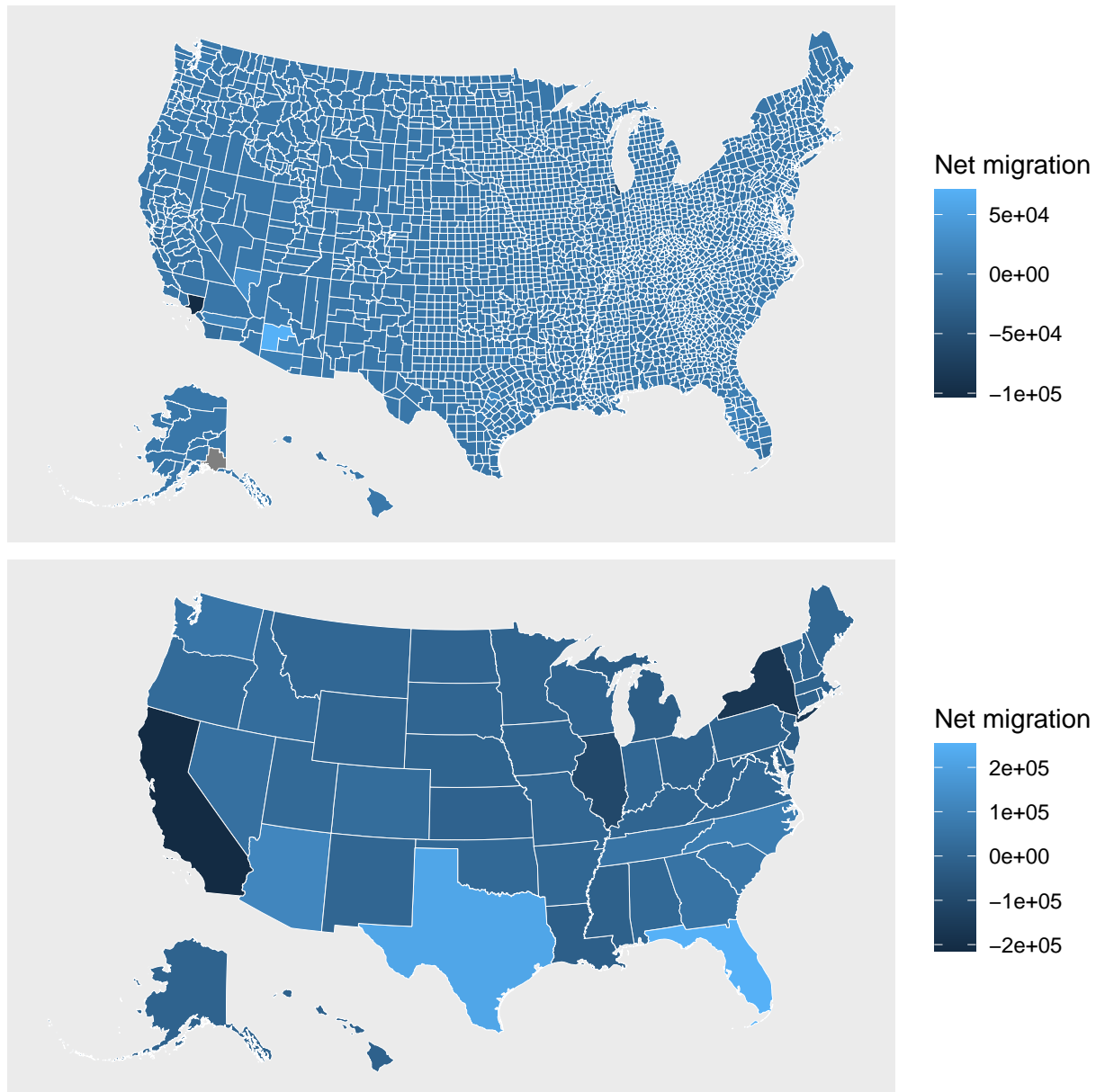
The raw data from census comes in a wide format where years are attached to the variable names (see the sample of raw data above). The first step is to first gather the 180 columns using “gather” function in R. Next, we create a year variable by parsing or extracting the year from the string categories. We then omit the years in the category variable to end up with unique list of variables. We can keep our data in the long format or spread it again using “spread” function; I actually prefer the latter format as it makes it easy to see and sort variables in R studio. The cleaned data looks as follow:

year	county_fips	STNAME	CTYNAME	POPESTIMATE	BIRTHS	RBIRTH	DEATHS
2010	01001	Alabama	Autauga County	54761	151	NA	157
2010	01003	Alabama	Baldwin County	183121	514	NA	534
2010	01005	Alabama	Barbour County	27325	70	NA	131
2010	01007	Alabama	Bibb County	22858	44	NA	32
2010	01009	Alabama	Blount County	57372	181	NA	132
2010	01011	Alabama	Bullock County	10876	37	NA	53

Net migration

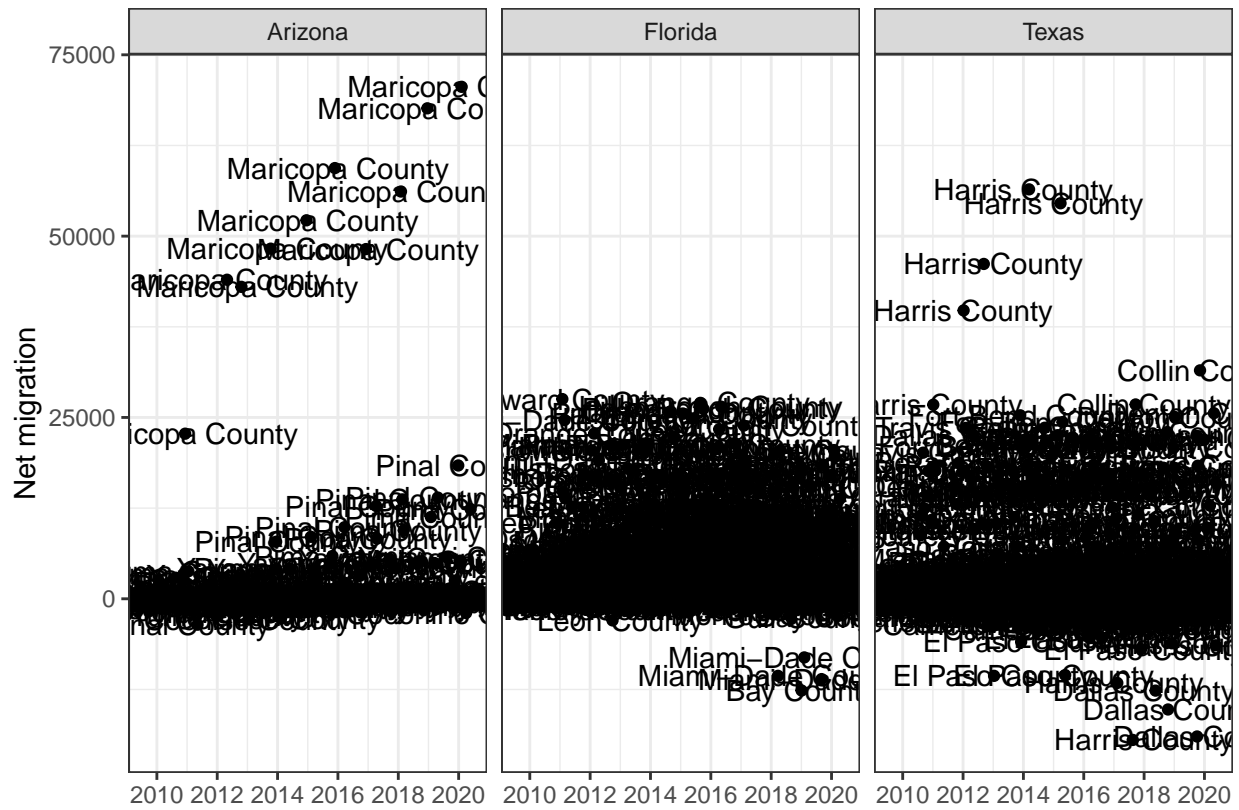
While I cleaned this data primarily to merge it with my project data for college enrollment and use only population and migration as controls, I think it’s interesting to visualize the net migration and see which destinations, counties and states, attract more migration. Choropleth map can be of a good use for this purpose. Thus, I used “urbnmapr” along with “ggplot”.

Visuals

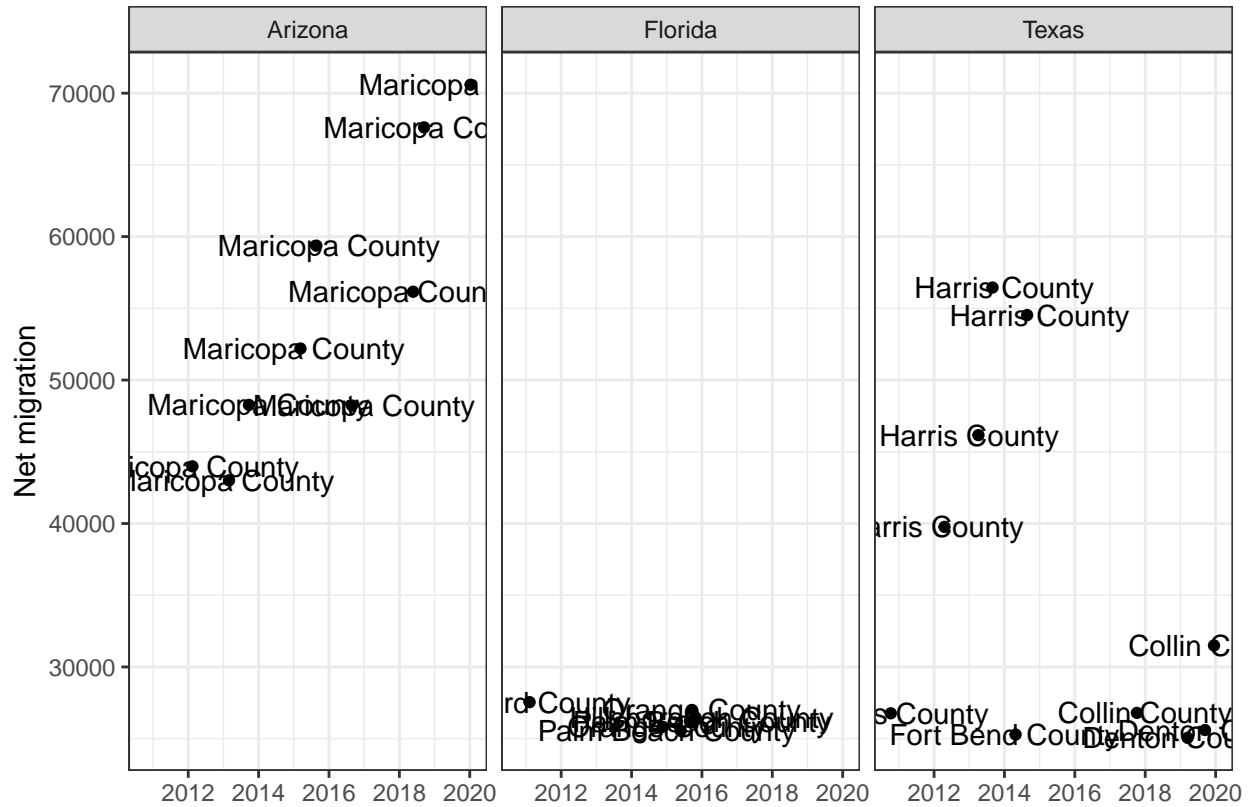


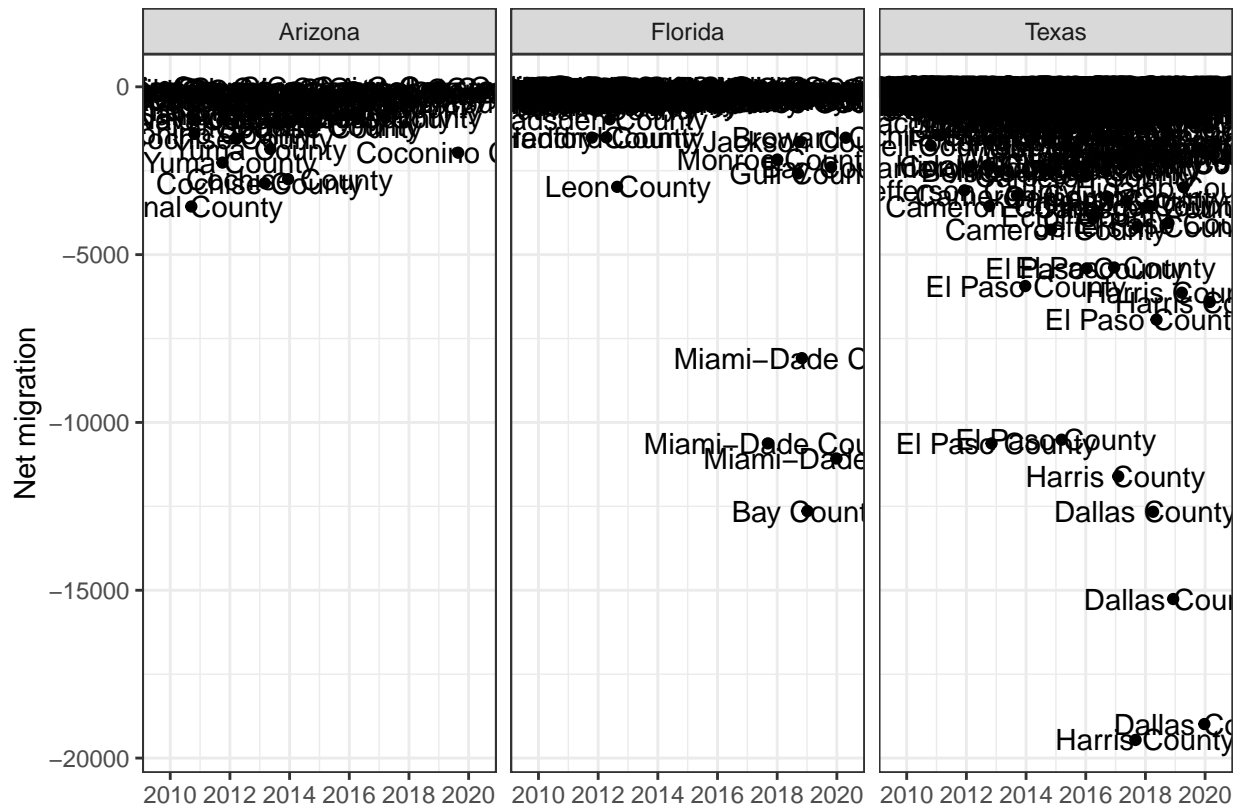
The maps show the net migration per county and state in 2020. Florida, Texas and probably even Arizona seem to have higher net migration compared to any other states. The county map doesn't clearly show which counties in these three states have higher migration. We can though see that Maricopa county in Arizona, and Collin and Harris counties in Texas are the most attractive migration destination.

Let's now closely look at which counties are driving these states high migration.



The above plot looks too busy especially for Florida, so let's focus on the upper and lower migration counties.





Arizona seem to have fewer counties with negative net migration. Miami Dade and Dallas counties have gradually net negative migration throughout the period. Harris county seem to have ups and downs in net migration.

Future research ideas may entail learning about factors that drive higher net migration in the discussed states. Some of the reasons are probably already obvious, high prices and unemployment rate.