

PROC GMAP to visualize 30 years of US Census data

Evian Fernandez Garcia, Astellas Pharma Europe BV, Leiden, The Netherlands

ABSTRACT

Data from the US Decennial Census from 1980 to 2010 was included for the most southward states: Alabama, Arizona, Arkansas, California, Colorado, Florida, Georgia, Kansas, Louisiana, Mississippi, Nevada, New Mexico, North Carolina, Oklahoma, South Carolina, Tennessee, Texas and Utah. Population Count and Race are included: the Hispanic population of any race. Then the non-Hispanic population is divided by race in other three levels White, Blacks and Other race.

The 30 years census data shows a decline in the percentage of the White populations, and an increase of the Hispanic, closing the gap from 60% to 30%. If the current trend continues, during the next 100 years the Hispanic might surpass the White as the biggest population in those 18 states.

SAS statistical models and PROC GMAP visualisation of the data and predictions for the next 30 years are used to compare the time evolution of the percentages of White and Hispanic populations.

INTRODUCTION

The original purpose of this exercise was to learn how to plot data in SAS that contain geographical information. For years, I was doing that in R, which have more graphical capabilities. Then at the time to choose, the data I decided to use the data from the Decennial US Census, because is accessible in their website and because have topics of my interest: population growth and migration.

The US Decennial Census started collecting data about the Hispanic population in 1980. Until 2000, among other data, the gender, race and the Hispanic Origin (Mexican, Puerto Rican, Cuban or Other) was included. In 2010, the Hispanic Origin was expanded to the Country of Origin. This data is collected at different geographical levels, for this exercise, we consider County and State levels only.

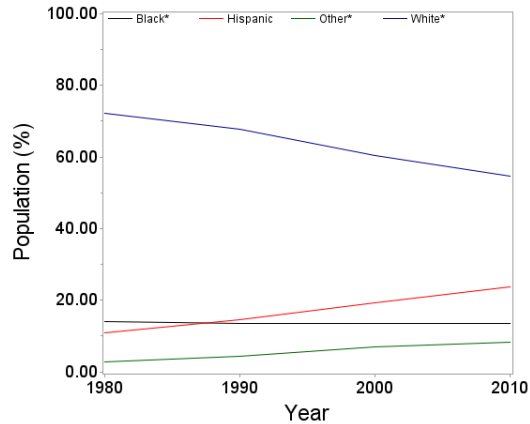
The data is available at the US Decennial Census website. However, the data for the States: South Carolina, New Mexico and Oklahoma for the 1980 Census at County level was missing.

In the county maps of 1980, the county of La Paz is not displayed. This county only existed after 1983 when was created out of the county of Yuma.

THIRTY YEARS OF CENCUS

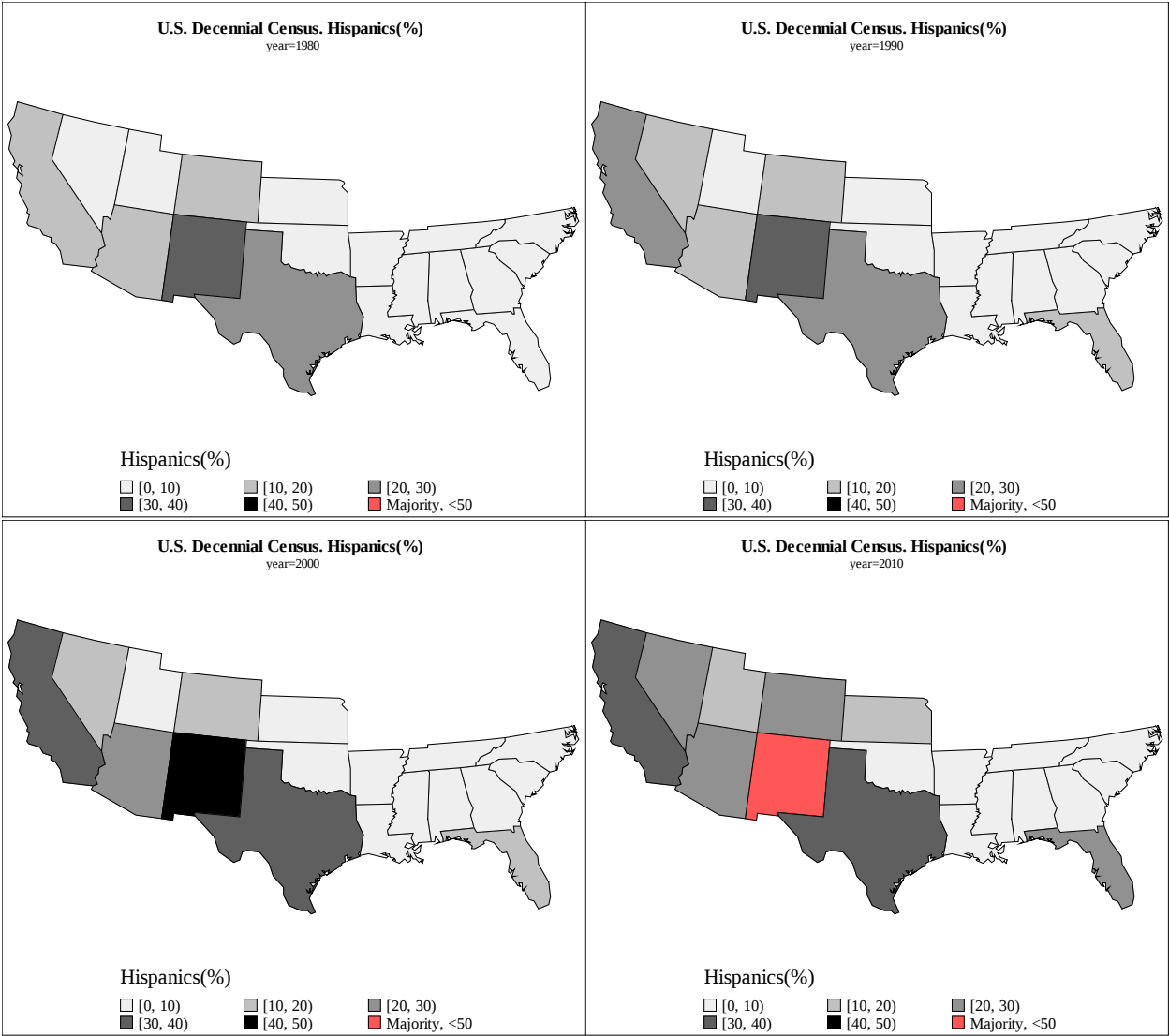
Let us begin with an overview of the over the years. As a percentage of the total population, the Non-Hispanic Blacks (Blacks*) have been slowly declining from 14.1% to 13.5% and the Non-Hispanic Whites (Whites*) have been steeply declining from 72.2% to 54.6%. The minorities that have been increasing in the last 30 years are the Other Non-Hispanics (Other*), going from 2.9% to 8.2%, and the Hispanics, going from 10.9% to 23.7%.

This shows that the Hispanics have surpass the Blacks* as the first minority in the region and in the last 30 years have been closing the gap with the Whites* from 61.3% to 30.8%.



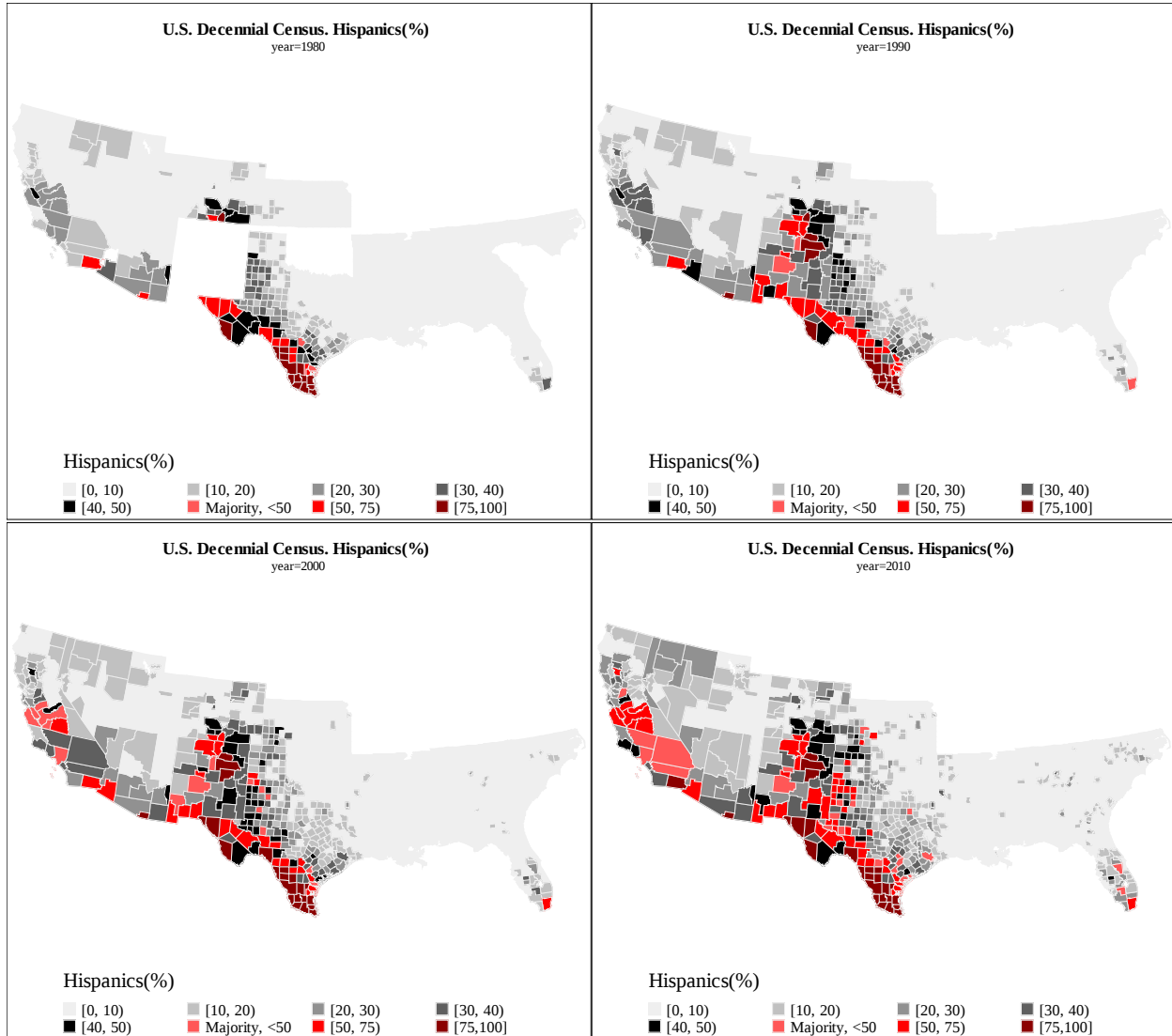
HISPANIC POPULATION AT STATE LEVEL

This increase of the Hispanic population is not the same throughout the eighteen states included in this article. The state of Florida from 8.8% to 22.5%, Nevada from 6.8% to 26.5%, Arizona from 16.3% to 29.7%, California from 19.2% to 37.6%, Texas from 21.0% to 37.6% and New Mexico from 36.6% to 46.3%, achieving majority. The previous six states account for the 55.7% of the total population in 1980 and 60.7% in 2010.



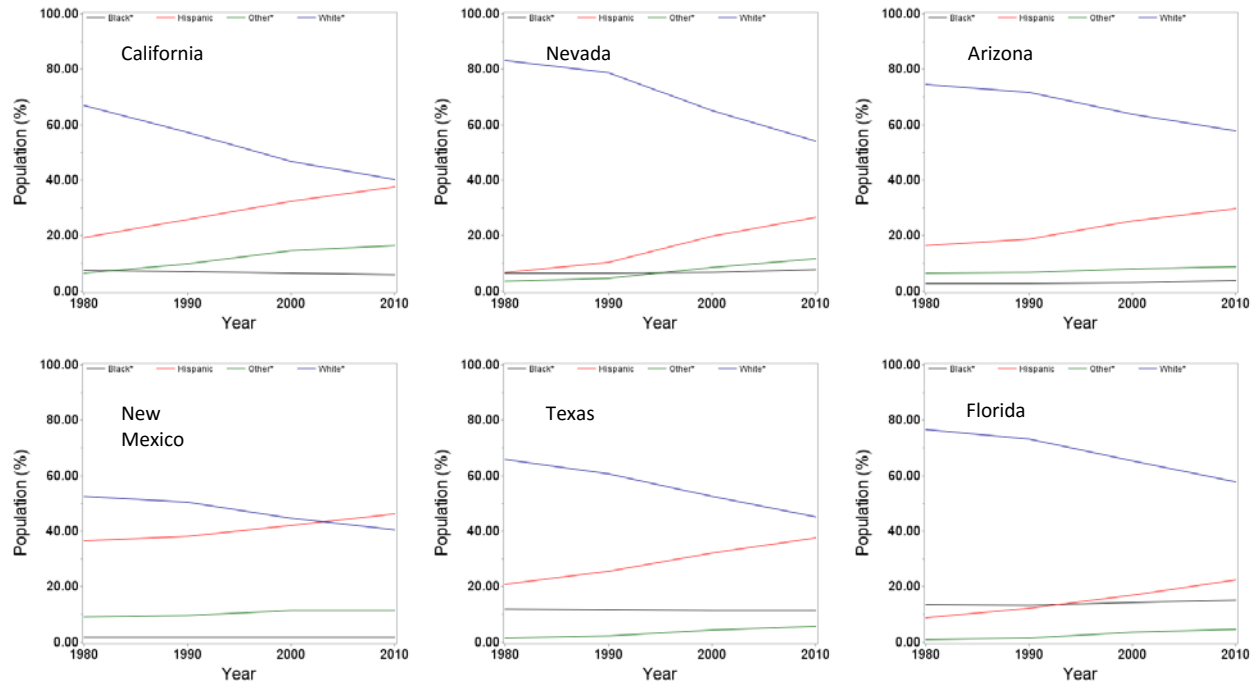
HISPANIC POPULATION AT COUNTY LEVEL

Going a geographical level deeper: the County level, we can see in more detail the development in certain areas. Starting south: the border with Mexico and the southern tip of Florida and going north. We can also see the increase in the percentage of Hispanics, reaching more than 75% in twenty-six counties: one in Arizona, one in California, three in New Mexico and twenty-one in Texas. Moreover, reaching more than 90% in eight counties, all of them in Texas.



CHANGES IN A SUBSET OF STATES

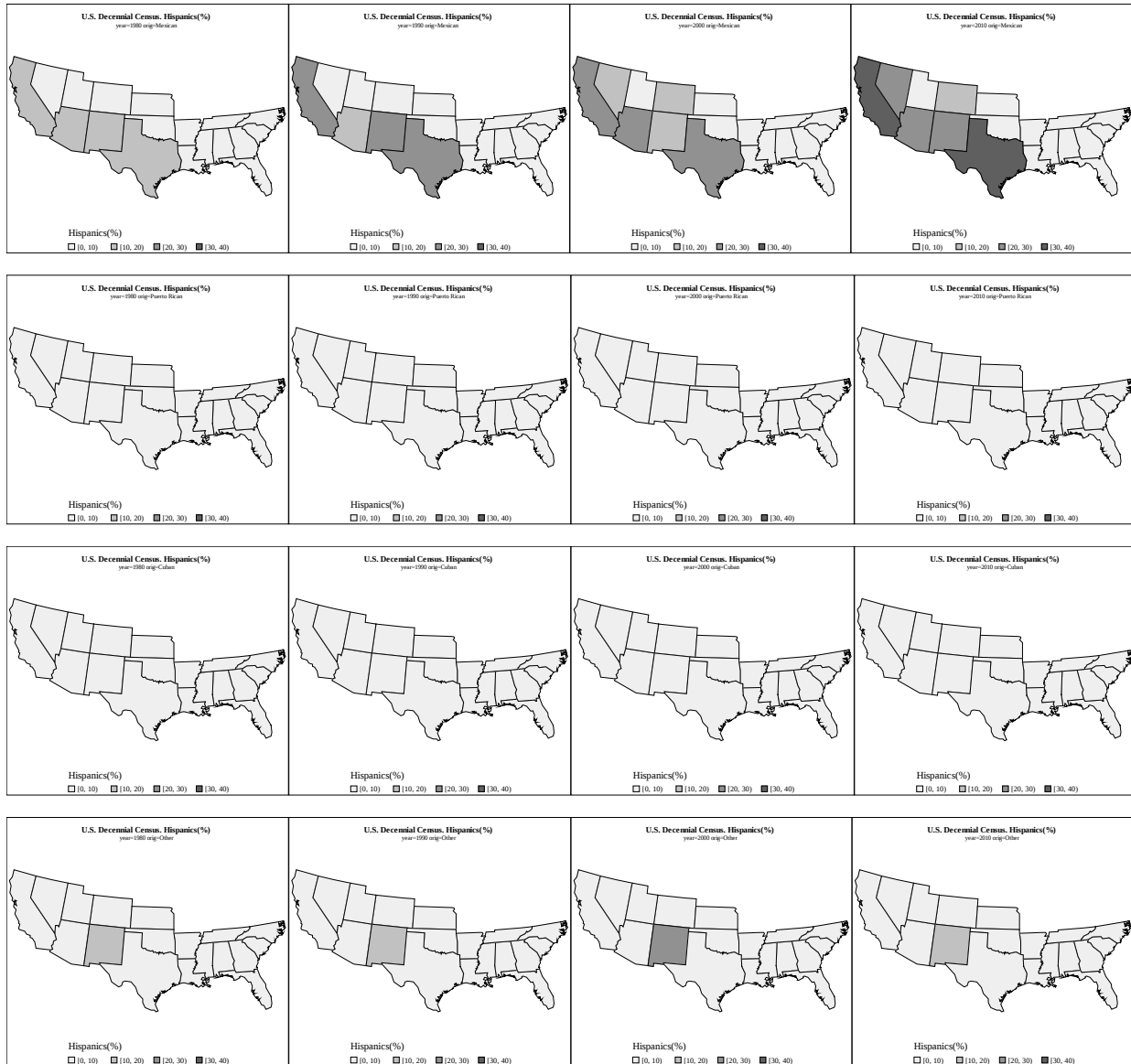
Selecting the States with the highest percentage of Hispanics: California, Nevada, Arizona, New Mexico, Texas and Florida, we can see the time trend of every population.



HISPANIC ORIGIN AT STATE LEVEL

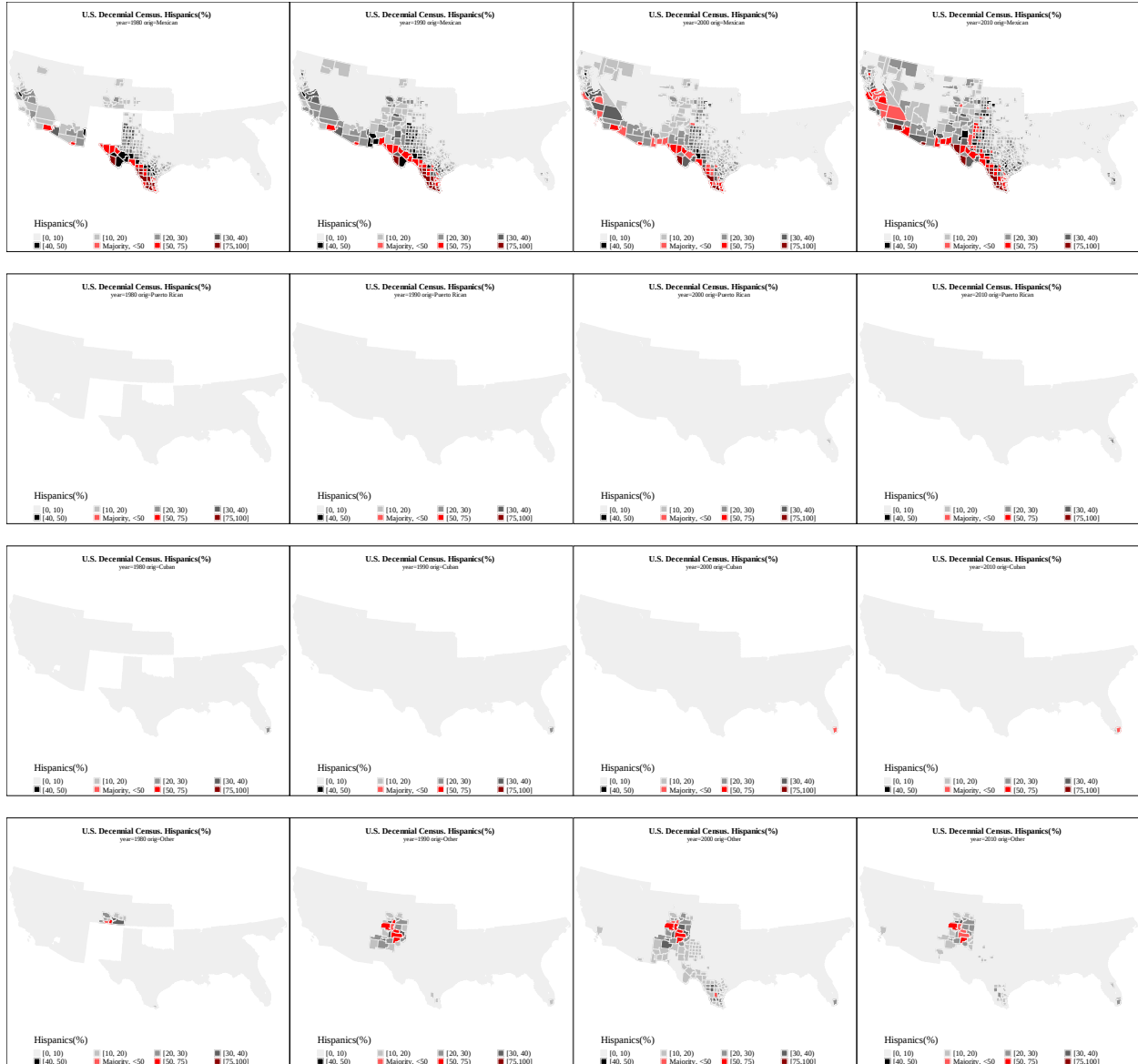
Inside the Hispanic population, the Origin: Mexican, Puerto-Rican, Cuban and Other Origin, was another of the variables recorded. These four origins have been growing in numbers but maintaining the same proportion: Mexican 74.1% to 72.1%, Puerto-Rican 2.6% to 4.2%, Cuban 5.6% to 4.1% and Other Origin 17.7% to 19.6%.

Geographically, at State level Puerto-Ricans, Cubans and Other Origins stayed and the same level. However, Mexicans have increased their percentage in most of the western states.



HISPANIC ORIGIN AT COUNTY LEVEL

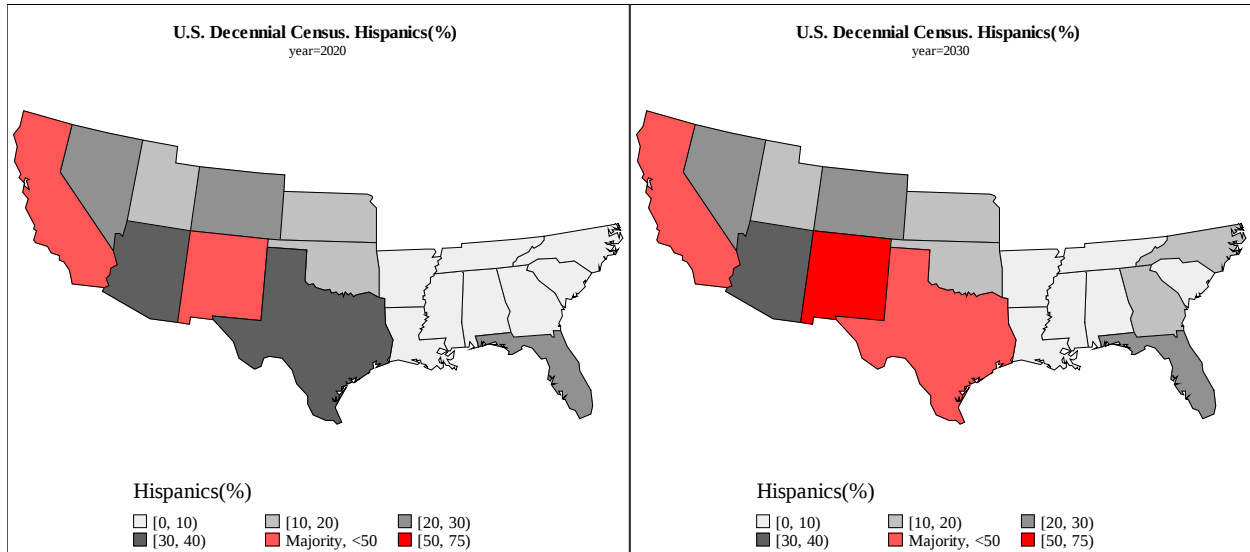
Going deeper, at County level we can see in more detail the increase of the Mexican origin in the west half, reaching majority in 68 counties: two in Arizona, twelve in California, one in Colorado, one in Kansas, four in New Mexico and 48 in Texas. The Puerto-Rican increase in central Florida and the Cuban increase in two counties in south Florida reaching majority in one of them.



PREDICTIONS AT STATE LEVEL

Previously we have seen that the Hispanic population reached majority in the state of New Mexico in 2010 with 46.3%. To explore future trends for the rest of the region we decided to fit longitudinal mix models with random intercepts, the Population Count as the response variable, Race and Race*Year as covariates and the data used was at the County level. Two models were fitted, one for the region and another by State.

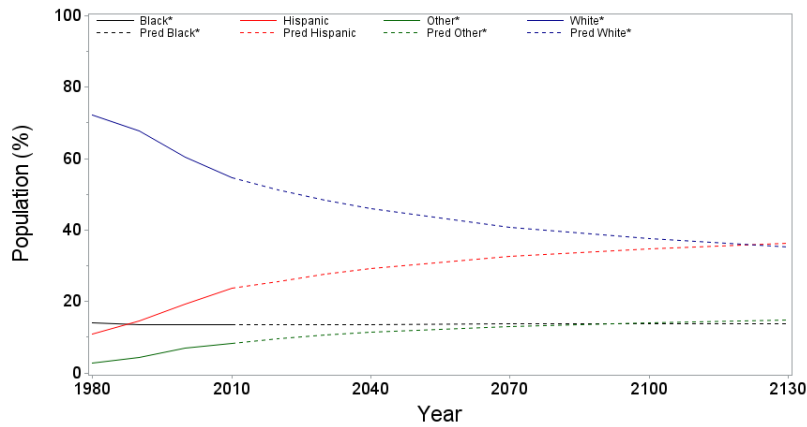
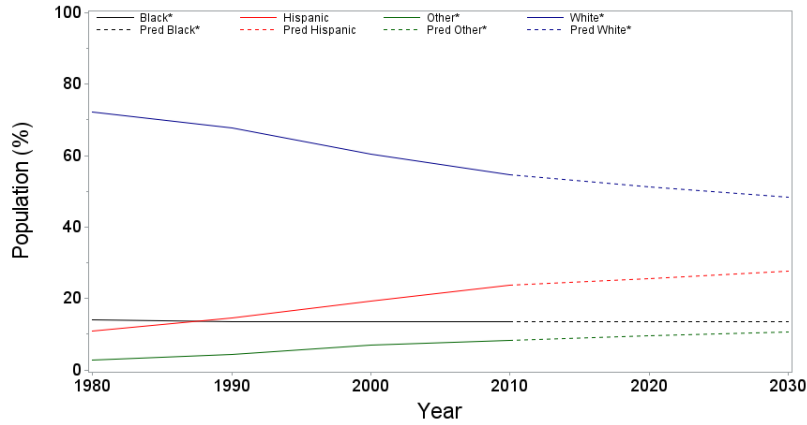
The By-State model estimate that the Hispanic population will reach majority in state of California in 2020 with 40.6% and in 2030 in the state of Texas with 42.0%.



PREDICTIONS AT REGIONAL LEVEL

The Region model estimate that the Hispanic in 20 years will further close the gap with the Whites* to 20.8% and in 120 years the Hispanics will reach majority in the region.

If we repeat the Region model using the six states with the highest percentage of Hispanics: California, Nevada, Arizona, New Mexico, Texas and Florida, the Hispanics will reach majority in 2040 with a 40.0%.



SAS PROCEDURE PROC GMAP

Using the SAS procedure PROC GMAP was quite straightforward, but this procedure lack certain options that could enhance the display. Still a lot can be achieved using PROC GMAP if combined with other procedures and other programming blocks in SAS.

```
PROC GMAP DATA= stat3 MAP= mapssas.us;  
    ID statecode;  
    CHORO percc / LEVELS= 8 UNIFORM LEGEND= legend1 coutline= black;  
    BY year;  
RUN;
```

Continuation of body – after source code.

SHORTCOMINGS OF PROC GMAP

- Not possible (in an easy way) to merge two of more areas inside a geographical level, e.g. two counties, and plot them as such. This can be done using PROC GREMOVE.
- At county level, the SAS map `mapssas.uscounty` do not contain the names of the counties and states, only the respective code numbers.
- There are no options to combine different maps. This have to done through programming, modifying coordinates.
- There are no options to plot different type of lines at different geographical levels, e.g., black outline for the US and grey outline for the States. This can be done using the ANNO option in CHORO.
- There are no options to label geographical areas. This can be done using the ANNO option in CHORO.

CONCLUSION

This SAS procedure is not for the beginner programmer if we want to use several advance options.

REFERENCES

<https://www.census.gov/>

https://support.sas.com/sassamples/graphgallery/PROC_GMAP.html

CONTACT INFORMATION

(In case a reader wants to get in touch with you, please put your contact information at the end of the paper.)

Your comments and questions are valued and encouraged. Contact the author at:

Evian Fernandez Garcia
Astellas Pharma Europe BV
evian.fernandezgarcia@astellas.com

Brand and product names are trademarks of their respective companies.