

SESUG Paper 203-2019
Data Visualizations using Census Shapefiles, PROC GMAP, SAS/GRAPH Animation, and BISG
Martha D. Pohl¹

ABSTRACT

The methodology to compute a proxy for missing race and ethnicity released in the summer of 2014 by the Consumer Financial Protection Bureau (CFPB) uses the surname of SMITH and the state of California as an example of this computation. This methodology uses Bayes' Theorem, the Census Surname List by Race Ethnicity from the Decennial Census, and the racial/ethnic composition of the population by geography from the Census Summary File 1 (100% sample). It is called the Bayesian Improved Surname Geocoding (BISG) proxy method.

This presentation will look at the Surname of SMITH (the most frequently occurring surname in the United States) as well as other surnames at the County and Zip Code (ZCTA) levels using the data visualization technique of SAS/GRAPH animation to examine these changes by geographic location and concentration. Racial/ethnic composition of the population by county over time using the American Community Survey (ACS) yearly Estimates will be visualized nationwide.

This methodology called BISG is used to compute a proxy for race/ethnicity when it is not available using the surname and geographic location. SAS/GRAPH Templates for SAS[®] PROC GMAP using Shapefiles from the Census Bureau and PROC GSLIDE for SAS animation will be provided.

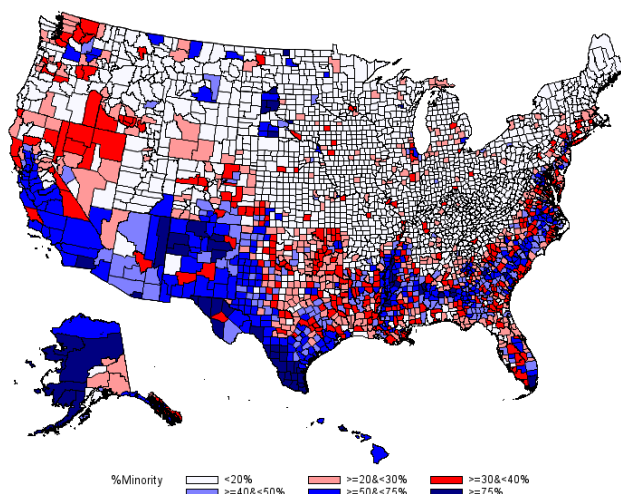
INTRODUCTION

Data Scientists continually look to enhance the capability of analyzing data. This can be done by dynamic mapping over time to visualize changes. This presentation explores this method using SAS/GRAPH animation and Census data at the county, block and zip code levels. The calculations and visualizations were done using SAS², SAS/GRAPH, PROC GMAP and PROC GSLIDE. The SAS code for these procedures is included as well as three examples of the animated SAS GIF files.

PROC GMAP: AMERICAN COMMUNITY SURVEY (ACS) 5-YEAR ESTIMATES PNGS: COUNTY LEVEL

2013-2017 ACS 5-year Estimates by County for %Minority; %Black, Not Hispanic; %Hispanic, regardless of race; %Asian, Not Hispanic³; %Native American, Not Hispanic using SAS/GRAPH PROC GMAP.

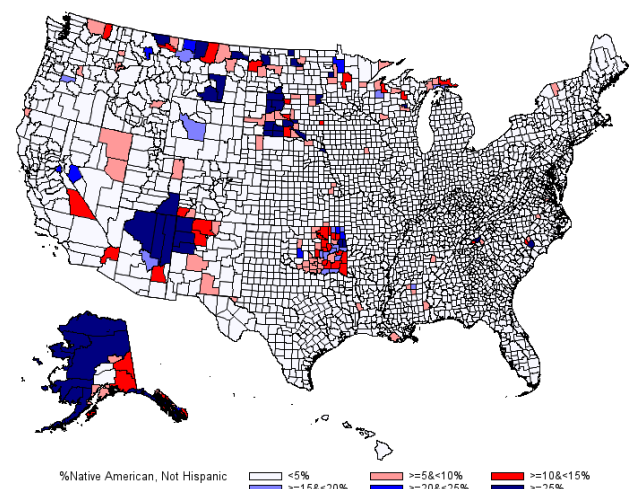
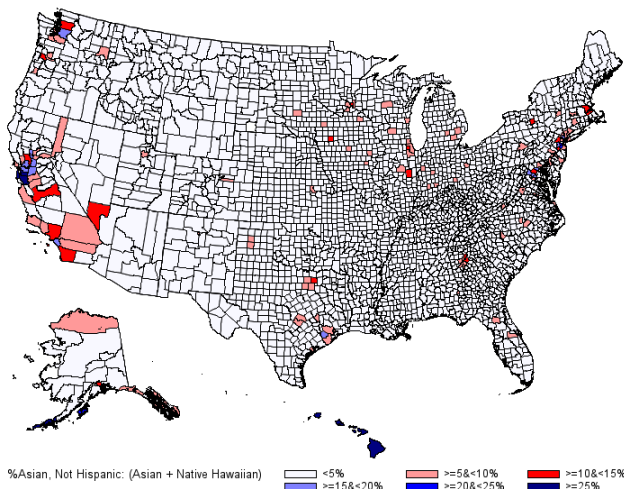
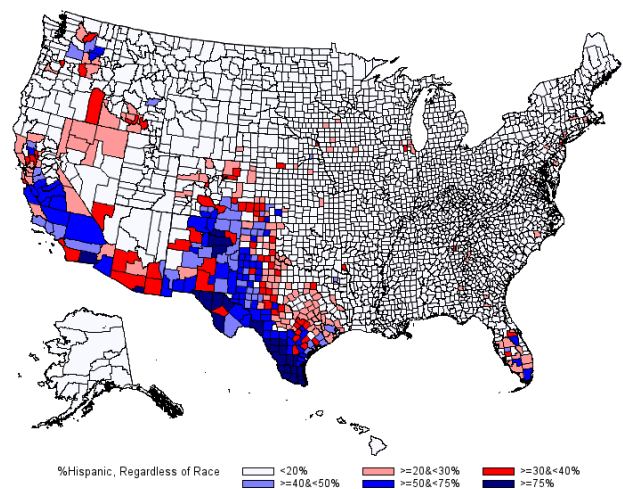
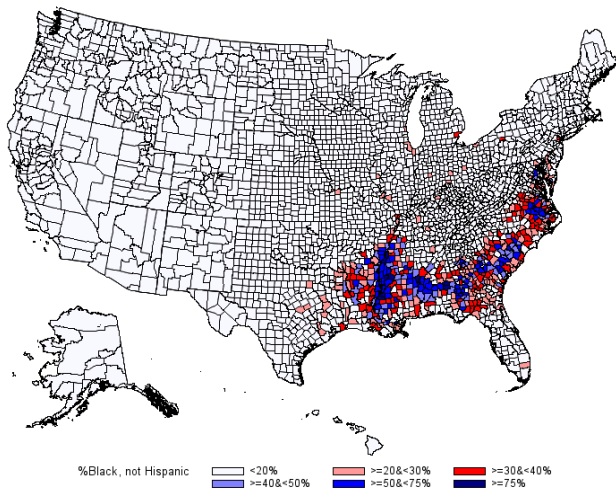
Percent Minority by County 2017 ACS 5-year estimate, Source: www.census.gov
M.D. Pohl SESUG 2019



¹ U.S. Department of Justice, retired. Disclaimer: Any opinions expressed in this paper are those of the author and do not constitute policy or opinion of the U.S. Department of Justice or any of its subcomponents.

² SAS 9.4 TS Level 1M4 X64_7 PRO platform.

³ Asian=Asian plus Native Hawaiian.



SAS CODE TO GENERATE PNGs: SAS/GRAPH PROC GMAP AT THE COUNTY LEVEL

```
DATA ACSi;
input @10 ST $2.0 @12 CT $3.0 @15 T W B I A NH O T2 H;
cards;
;
;
```

```
DATA ACSip; SET ACSi;
STATE=(ST*10)/10; COUNTY=(CT*10)/10;
MINI=T-W; MINIP=(MINI/T)*100;
KEEP STATE COUNTY MINIP;
```

```
PROC FORMAT;
VALUE MINIPP
1= '<10%'
2= '>=10&<20%'
3= '>=20&<30%'
4= '>=30&<50%'
5= '>=50&<75%'
6= '>=75%';
```

```
DATA ACSa; SET ACSip; IF state=72 then delete;
IF MINIP lt 10 THEN MINIPP=1;
IF MINIP Ge 10 AND MINIP LT 20 THEN MINIPP=2;
IF MINIP Ge 20 AND MINIP LT 30 THEN MINIPP=3;
IF MINIP Ge 30 AND MINIP LT 50 THEN MINIPP=4;
IF MINIP Ge 50 AND MINIP LT 75 THEN MINIPP=5;
IF MINIP Ge 75 THEN MINIPP=6;
```

```

PROC FREQ DATA=ACSa; TABLES MINIPP;
FORMAT MINIPP MINIPP.; run;

TITLE1 "Percent Minority by County 2009 ACS 5-year estimate, Source: www.census.gov";
TITLE2 "M.D. Pohl: SESUG 2019";

PATTERN1 V=S C= GhostWhite;
PATTERN2 V=S C= CXFF9999;
PATTERN3 V=S C= CXFF0000;
PATTERN4 V=S C= CX8080FF;
PATTERN5 V=S C= CX0000FF;
PATTERN6 V=S C= CX000080;

DATA states alaska hawaii;
set mapsgfk.us_counties;
if state=15 then output hawaii;
else if state=2 then do; if county ^= 16 then output alaska; end;
else output states; run;

proc gproject data=states out=stproj degrees eastlong latlong;
id state county;
proc gproject data=alaska out=akproj degrees eastlong latlong;
id state county;
proc gproject data=hawaii out=hiproj degrees eastlong latlong;
id state county;

data akproj(drop=x y);
set akproj;
length newx newy 5;
newx=(x-.55)*.50;
newy=(y-.35)*.50;

data hiproj(drop=x y);
set hiproj;
length newx newy 5;
newx=x-.12;
newy=y-.20;
run;

data all;
set stproj akproj(rename=(newx=x newy=y))
hiproj(rename=(newx=x newy=y));
run;
proc sort; by state county;

proc gremove data=all out=states;
by state;
id county;
run;

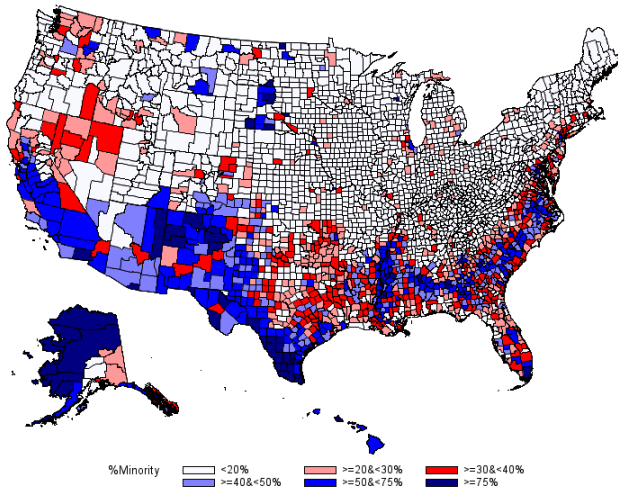
data stoutline;
set states;
by state segment;
retain xsys ysys '2' when 'a' color 'black' style 'mempty' size 1.5;
length function $8;
if first.state or first.segment then function='poly';
else function='polycont';
run;

proc gmap data=ACSa map=all all;
ID STATE COUNTY;
CHORO MINIPP / DISCRETE LEVELS=6 coutline=black anno=stoutline;
FORMAT MINIPP MINIPP.;
LABEL MINIPP='%Minority';
run;
quit;

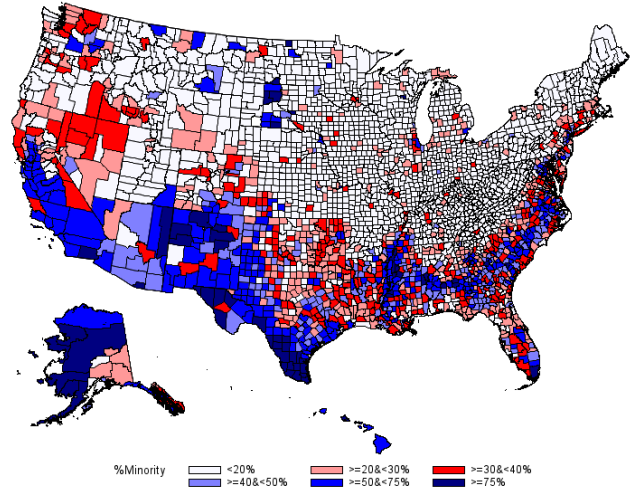
```

SAS/GRAPH ANIMATION USING PROC GMAP AND PROC GSLIDE: %MINORITY⁴ (3 CATEGORIES) 2005-2009 & 2013-2017 ACS 5-YEAR ESTIMATES SHOWING AN INCREASE IN COUNTY CONCENTRATIONS.

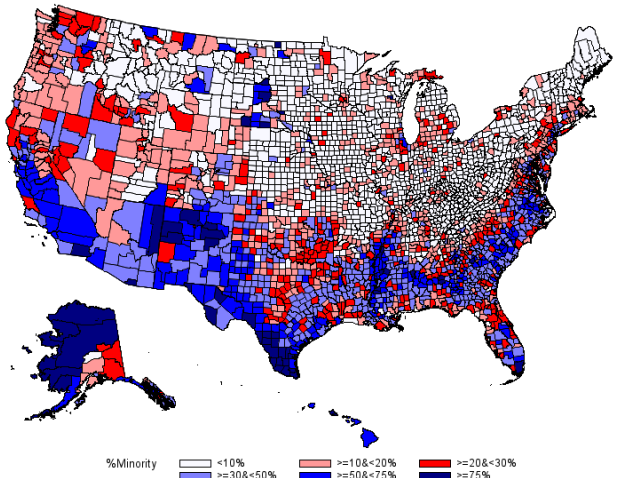
Percent Minority by County 2009 ACS 5-year estimate, Source: www.census.gov
M.D. Pohl: SESUG 2019



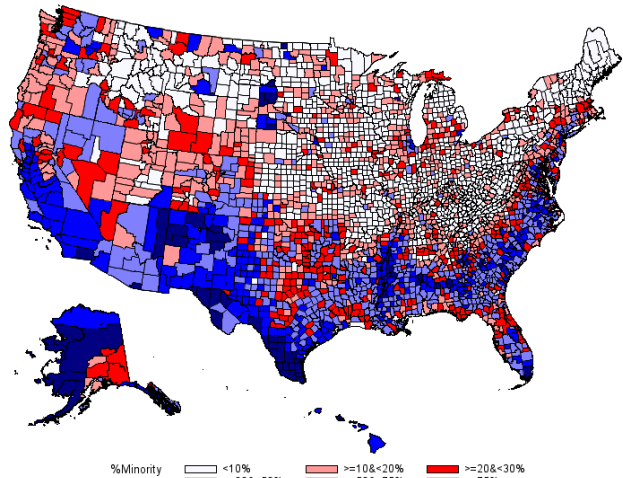
Percent Minority by County 2017 ACS 5-year estimate, Source: www.census.gov
M.D. Pohl: SESUG 2019



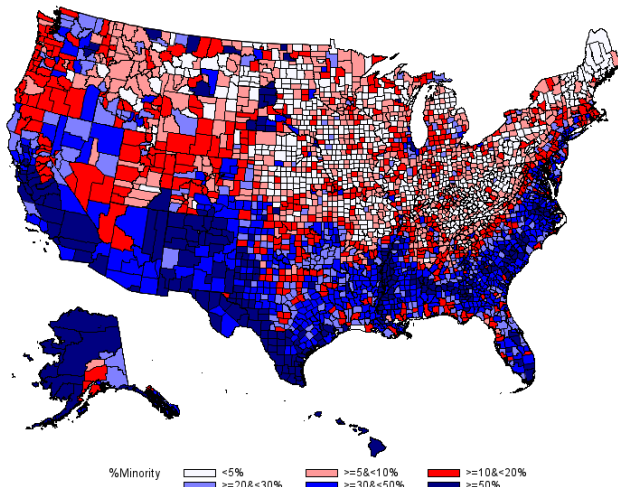
Percent Minority by County 2009 ACS 5-year estimate, Source: www.census.gov
M.D. Pohl: SESUG 2019



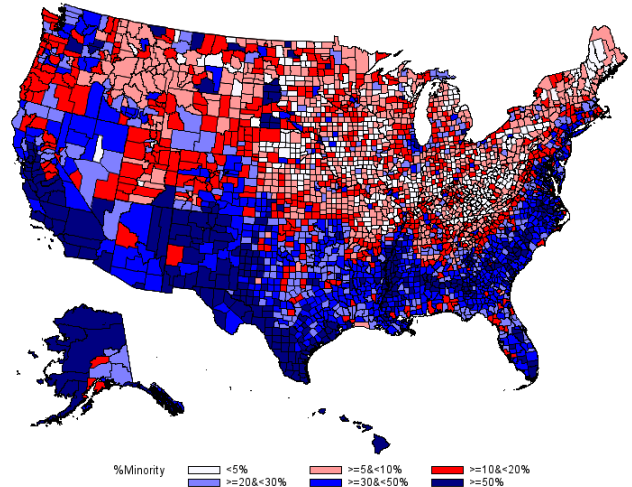
Percent Minority by County 2017 ACS 5-year estimate, Source: www.census.gov
M.D. Pohl: SESUG 2019



Percent Minority by County 2009 ACS 5-year estimate, Source: www.census.gov
M.D. Pohl: SESUG 2019



Percent Minority by County 2017 ACS 5-year estimate, Source: www.census.gov
M.D. Pohl: SESUG 2019



⁴ Total Population minus White, Not Hispanic.

SAS/GRAPH ANIMATION CODE: PROC GSLIDE 2009 TO 2017 AMERICAN COMMUNITY SURVEY (ACS) 5-YEAR ESTIMATES

I created three SAS/GRAPH animated GIFs for %Minority at the county level for the following 9 PNGs over time: 2005-2009 ACS 5-year Estimates to 2013-2017 ACS 5-year Estimates⁵.

American Community Survey (ACS):

2013-2017 ACS 5-year Estimates	2012-2016 ACS 5-year Estimates	2011-2015 ACS 5-year Estimates
2010-2014 ACS 5-year Estimates	2009-2013 ACS 5-year Estimates	2008-2012 ACS 5-year Estimates
2007-2011 ACS 5-year Estimates	2006-2010 ACS 5-year Estimates	2005-2009 ACS 5-year Estimates

EXTERNAL ATTACHMENT: THREE SAS/GRAPH ANIMATED GIFS

ACS1.gif, ACS2.gif, and ACS3.gif.

SAS CODE TO GENERATE ANIMATED GIFS PROC GSLIDE:

```
data anno1;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2009.png"; output; run;

data anno2;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2009.png"; output; run;

data anno3;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2009.png"; output; run;

data anno4;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2009.png"; output; run;

data anno5;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2009.png"; output; run;

data anno6;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2009.png"; output; run;

data anno7;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2010.png"; output; run;
```

⁵ Starting at 2013, an additional county: Bedford City VA exists (State code 51; County Code 515).

```

data anno8;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2011.png"; output; run;

data anno9;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2012.png"; output; run;

data anno10;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2013.png"; output; run;

data anno11;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2014.png"; output; run;

data anno12;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2015.png"; output; run;

data anno13;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2016.png"; output; run;

data anno14;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2017.png"; output; run;

data anno15;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2017.png"; output; run;

data anno16;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2017.png"; output; run;

data anno17;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2017.png"; output; run;

data anno18;
  length function $8 imgpath $100;
  xsys='3'; ysys='3'; when='a';
  x=0; y=0; function='move'; output;
  x=100; y=100; function='image'; style='fit';
  imgpath="E:\TEST\ACSa2017.png"; output; run;

```

```

options reset=all device=PNG xpixels=800 ypixels=600 nodisplay;
goptions reset=all device=GIF;
goptions xpixels=800 ypixels=600;
goptions cback=white gunit=pct;
goptions noborder;

/*options for GIF animation*/
options nodate nonumber
/*orientation=landscape*/ /*papersize=("11in" "8.5in")*/ /*;*/
animduration=.4 animloop=/*0*/ /*1*/ 0
animoverlay=no animate=start center; run;

/*output*/
ods _all_ close;
ods html gpath='E:\test\' (url=none)
    file="E:\test\cum.html"
    style=htmlblue;

/*PROC GSLIDE to read the PNGs*/
goptions gsfmodes=replace; /*1st PNG, gsfmodes=replace*/
proc gslide anno=anno1 des=''; run; quit;
goptions gsfmodes=append; /* For rest of PNGs*/
proc gslide anno=anno2 des=''; run; quit;
proc gslide anno=anno3 des=''; run; quit;
proc gslide anno=anno4 des=''; run; quit;
proc gslide anno=anno5 des=''; run; quit;
proc gslide anno=anno6 des=''; run; quit;
proc gslide anno=anno7 des=''; run; quit;
proc gslide anno=anno8 des=''; run; quit;
proc gslide anno=anno9 des=''; run; quit;
proc gslide anno=anno10 des=''; run; quit;
proc gslide anno=anno11 des=''; run; quit;
proc gslide anno=anno12 des=''; run; quit;
proc gslide anno=anno13 des=''; run; quit;
proc gslide anno=anno14 des=''; run; quit;
proc gslide anno=anno15 des=''; run; quit;
proc gslide anno=anno16 des=''; run; quit;
proc gslide anno=anno17 des=''; run; quit;
proc gslide anno=anno18 des=''; run; quit;
ods html close;
ods listing;

```

SAS VISUALIZATION SAS/GRAPH PROC GMAP USING CENSUS SHAPEFILES AT THE CENSUS BLOCK LEVEL (CENSUS REDISTRICTING FILE PNG EXAMPLES).

I selected the Beverly Hills and Los Angeles areas using zip codes (ZCTAs) locations and the Census relationship files that provide a crosswalk between ZCTAs and Census Tracts. 27 Census Tracts in Beverly Hills and 115 in East Los Angeles. I zoomed in to 7, 5, 3 and 1 Census Tract of the 27 and used % Minority 2010 Census Redistricting data (Public Law 94-171) Summary File. I downloaded Census Shapefiles at the Census Block level for the State of California (over 40,000,000 records) to produce the SAS/GRAPH PROC GMAP data visualization.

Selected Tracts in Los Angeles County for the Beverly Hills area:

```

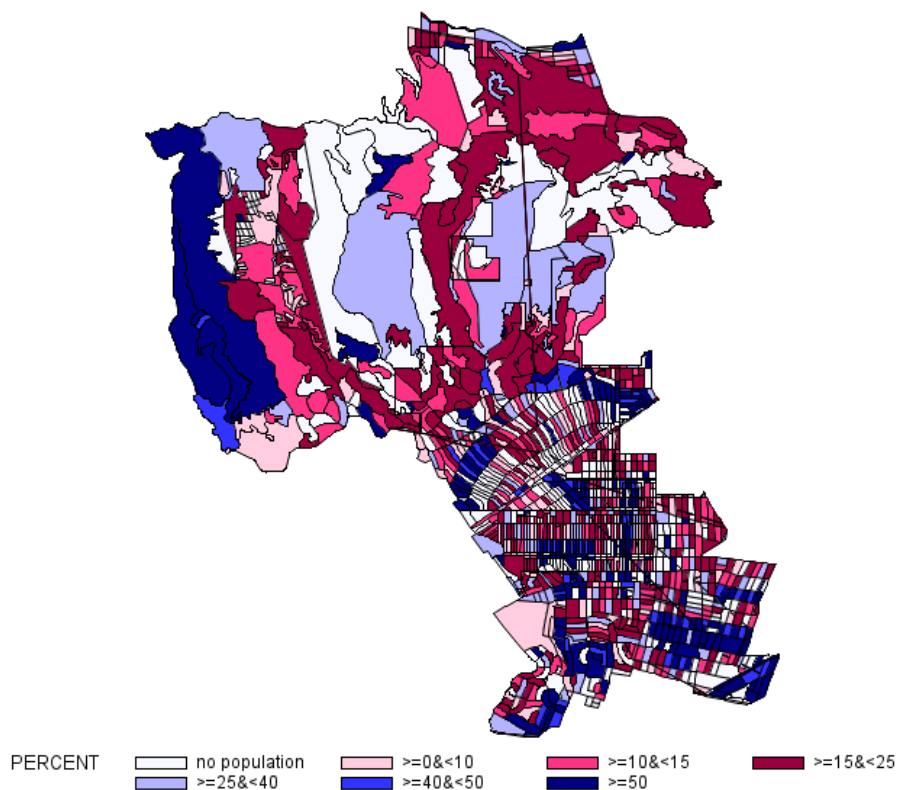
IF TRACTCE10= 143901 OR TRACTCE10= 143902 OR TRACTCE10= 214901 OR TRACTCE10= 216401 OR TRACTCE10= 216402
OR TRACTCE10= 216700 OR TRACTCE10= 216800 OR TRACTCE10= 217001 OR TRACTCE10= 217002 OR TRACTCE10= 261101
OR TRACTCE10= 261102 OR TRACTCE10= 261200 OR TRACTCE10= 262100 OR TRACTCE10= 269000 OR TRACTCE10= 269100
OR TRACTCE10= 269500 OR TRACTCE10= 269601 OR TRACTCE10= 269602 OR TRACTCE10= 270300 OR TRACTCE10= 700501
OR TRACTCE10= 700600 OR TRACTCE10= 700700 OR TRACTCE10= 700801 OR TRACTCE10= 700802
OR TRACTCE10= 700901 OR TRACTCE10= 700902 OR TRACTCE10= 701000; /*27 tracts*/

IF TRACTCE10= 700600 OR TRACTCE10= 700700 OR TRACTCE10= 700801 OR TRACTCE10= 700802 OR TRACTCE10= 700901
OR TRACTCE10= 700902 OR TRACTCE10= 701000; /*7 tracts*/

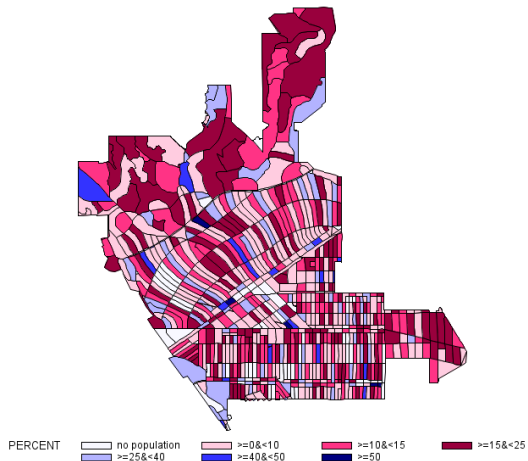
IF TRACTCE10= 701000 OR TRACTCE10= 700801 OR TRACTCE10= 700802 OR TRACTCE10= 700901 OR TRACTCE10= 700902;
/*5 tracts*/
IF TRACTCE10= 701000 OR TRACTCE10= 700901 OR TRACTCE10= 700902; /*3 tracts*/
IF TRACTCE10= 700902; /*1 tract*/

```

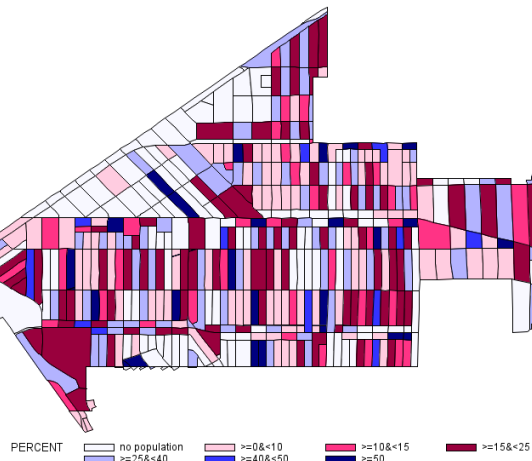
Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1
 Beverly Hills Blocks in 27 Census Tracts M.D. Pohl SESUG 2019



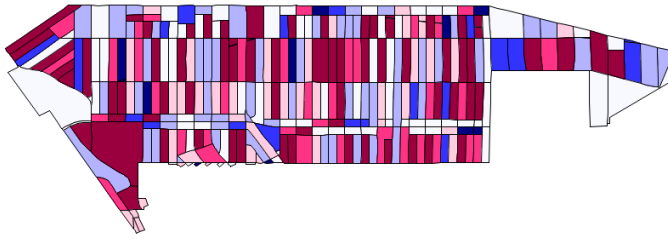
Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1
 Beverly Hills Blocks in 7 Census Tracts M.D. Pohl SESUG 2019



Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1
 Beverly Hills Blocks in 5 Census Tracts M.D. Pohl SESUG 2019

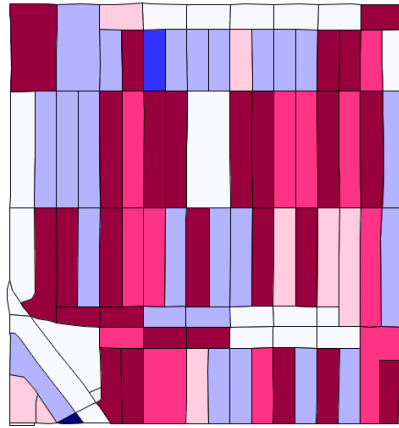


Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1
Beverly Hills Blocks in 3 Census Tracts M.D. Pohl SESUG 2019



PERCENT
 no population
 >=25%<40
 >=0%<10
 >=40%<50
 >=10%<15
 >=50
 >=15%<25

Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1
Beverly Hills Blocks in 1 Census Tracts M.D. Pohl SESUG 2019



PERCENT
 no population
 >=25%<40
 >=0%<10
 >=40%<50
 >=10%<15
 >=50
 >=15%<25

SAS/GRAPH PROC GMAP CODE USING SHAPEFILES AT THE CENSUS BLOCK LEVEL:

```
PROC MAPIMPORT DATAFILE='F:\test\tl_2010_06_tabblock10.shp' OUT=RR1; RUN;
PROC IMPORT datafile='F:\test\DEC_10_PL_P2LA.xlsx' DBMS=XLSX OUT=BB; RUN;
```

```
options reset=all;
options nodate orientation=landscape /*ls=118 ps=32000*/;
```

```
PROC FORMAT;
VALUE PCT
1='no population'
2='>=0%<10'
3='>=10%<15'
4='>=15%<25'
5='>=25%<40'
6='>=40%<50'
7='>=50';
```

```
DATA BB1; SET BB; LENGTH STATEFP10 $2 COUNTYFP10 $3
TRACTCE10 $6 BLOCKCE10 $4;
STATEFP10=SUBSTR(GEO,10,2); COUNTYFP10=SUBSTR(GEO,12,3);
TRACTCE10=SUBSTR(GEO,15,6); BLOCKCE10=SUBSTR(GEO,21,4);
```

```
IF COUNTYFP10='037';
IF TRACTCE10= 143901 OR TRACTCE10= 143902 OR TRACTCE10= 214901
OR TRACTCE10= 216401 OR TRACTCE10= 216402 OR TRACTCE10= 216700
OR TRACTCE10= 216800 OR TRACTCE10= 217001 OR TRACTCE10= 217002
OR TRACTCE10= 261101 OR TRACTCE10= 261102 OR TRACTCE10= 261200
OR TRACTCE10= 262100 OR TRACTCE10= 269000 OR TRACTCE10= 269100
OR TRACTCE10= 269500 OR TRACTCE10= 269601 OR TRACTCE10= 269602
OR TRACTCE10= 270300 OR TRACTCE10= 700501 OR TRACTCE10= 700600
OR TRACTCE10= 700700 OR TRACTCE10= 700801 OR TRACTCE10= 700802
OR TRACTCE10= 700901 OR TRACTCE10= 700902 OR TRACTCE10= 701000;
```

```
DATA BBB1; SET BB1;
ASIAN=A+NH; MIN= T-W; MINP=(MIN/T)*100;
HP=(H/T)*100; ASIANP=(ASIAN/T)*100;
BP=(B/T)*100; IP=(I/T)*100; WP=(W/T)*100; OP=(O/T)*100;
MULTI=T2; MULTIP=(MULTI/T)*100;
IF HP=. THEN HP=0; IF BP=. THEN BP=0;
IF WP=. THEN WP=0; IF IP=. THEN IP=0;
IF OP=. THEN OP=0; IF MULTIP=. THEN MULTIP=0;
IF T NE 0 AND MIN=0 THEN MINP=0;
IF T=0 THEN MINP=.;
DATA BEVH; SET BBB1;
```

```
IF MINP=. THEN PERCENT=1;
IF MINP GE 0 AND MINP LT 10 THEN PERCENT=2;
IF MINP GE 10 AND MINP LT 15 THEN PERCENT=3;
IF MINP GE 15 AND MINP LT 25 THEN PERCENT=4;
IF MINP GE 25 AND MINP LT 40 THEN PERCENT=5;
IF MINP GE 40 AND MINP LT 50 THEN PERCENT=6;
IF MINP GE 50 THEN PERCENT=7; RUN;
```

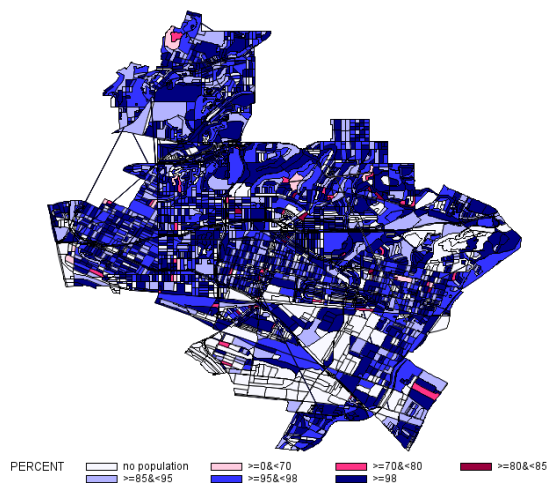
```
DATA RR2; set RR1; IF COUNTYFP10='037';
IF TRACTCE10= 143901 OR TRACTCE10= 143902 OR TRACTCE10= 214901
OR TRACTCE10= 216401 OR TRACTCE10= 216402 OR TRACTCE10= 216700
OR TRACTCE10= 216800 OR TRACTCE10= 217001 OR TRACTCE10= 217002
OR TRACTCE10= 261101 OR TRACTCE10= 261102 OR TRACTCE10= 261200
OR TRACTCE10= 262100 OR TRACTCE10= 269000 OR TRACTCE10= 269100
OR TRACTCE10= 269500 OR TRACTCE10= 269601 OR TRACTCE10= 269602
OR TRACTCE10= 270300 OR TRACTCE10= 700501 OR TRACTCE10= 700600
OR TRACTCE10= 700700 OR TRACTCE10= 700801 OR TRACTCE10= 700802
OR TRACTCE10= 700901 OR TRACTCE10= 700902 OR TRACTCE10= 701000; /*27 CENSUS TRACTS BEVERLY HILLS*/
```

```
PATTERN1 VALUE=MS COLOR= GHOSTWHITE;
PATTERN2 VALUE=MS COLOR= CXFFCCE0;
PATTERN3 VALUE=MS COLOR= CXFF3385;
PATTERN4 VALUE=MS COLOR= CX99003D;
PATTERN5 VALUE=MS COLOR= CXB3B3FF;
PATTERN6 VALUE=MS COLOR= CX3333FF;
PATTERN7 VALUE=MS COLOR= CX000080;
```

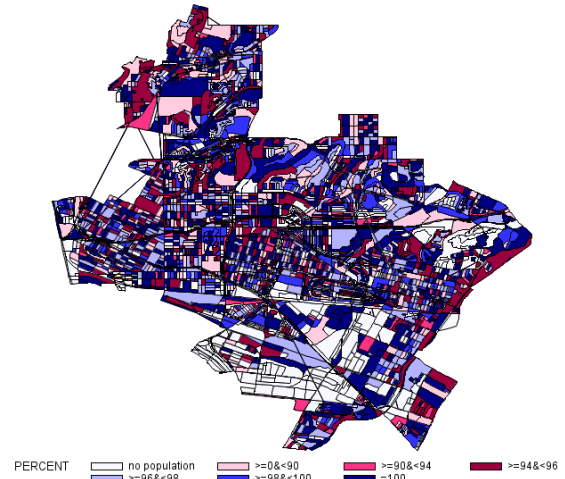
```
title1 "Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1";
title2 "Beverly Hills Blocks in 27 Census Tracts M.D. Pohl SESUG 2019";
PROC GMAP DATA=BEVH MAP=RR2 ALL ;
FORMAT PERCENT pct.; ID BLOCKCE10 ;
CHORO PERCENT / coutline=black discrete; RUN; QUIT;
```

```
PROC FREQ DATA=BEVH; TABLES PERCENT; FORMAT PERCENT pct.; RUN; QUIT;
PROC UNIVARIATE DATA=BEVH; VAR MINP; RUN; QUIT;
```

Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1
East LA Blocks in 110 Census Tracts M.D. Pohl SESUG 2019



Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1
East LA Blocks in 110 Census Tracts M.D. Pohl SESUG 2019



SAS GRAPH COLOR SELECTION⁶:

⁶ SAS/GRAPH® 9.4: Reference, Fifth Edition

<https://documentation.sas.com/api/docsets/graphref/9.4/content/graphref.pdf?locale=en#nameddest=n161ukdyz9wpfsn1nh8sihforvyq>

<https://documentation.sas.com/?docsetId=graphref&docsetTarget=p0edl20cvxxmm9n1i9ht3n21eict.htm&docsetVersion=9.4&locale=en>

Predefined Colors SAS/GRAPH® 9.4: Reference, Fifth Edition

<https://documentation.sas.com/?docsetId=graphref&docsetTarget=n161ukdyz9wpfsn1nh8sihforvyq.htm&docsetVersion=9.4&locale=en>

Color-Naming Schemes SAS® 9.4 Graph Template Language: User's Guide, Fifth Edition

<https://documentation.sas.com/?docsetId=grstatug&docsetTarget=p0edl20cvxxmm9n1i9ht3n21eict.htm&docsetVersion=9.4&locale=en>

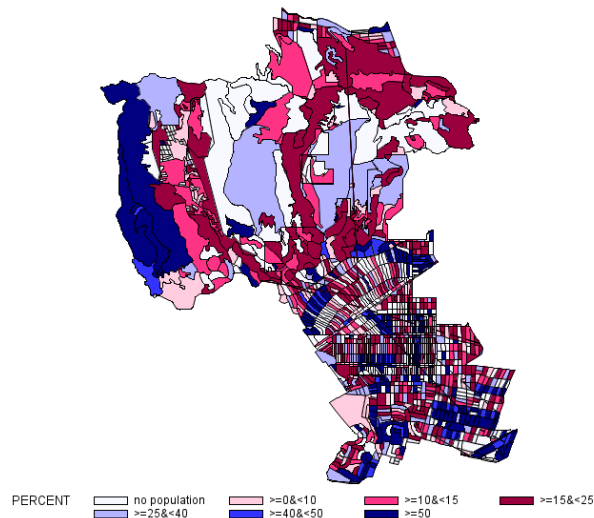
Appendix C: Color Names Defined by SAS/GRAPH

https://support.sas.com/content/dam/SAS/support/en/books/pro-template-made-easy-a-guide-for-sas-users/62007_Appendix.pdf

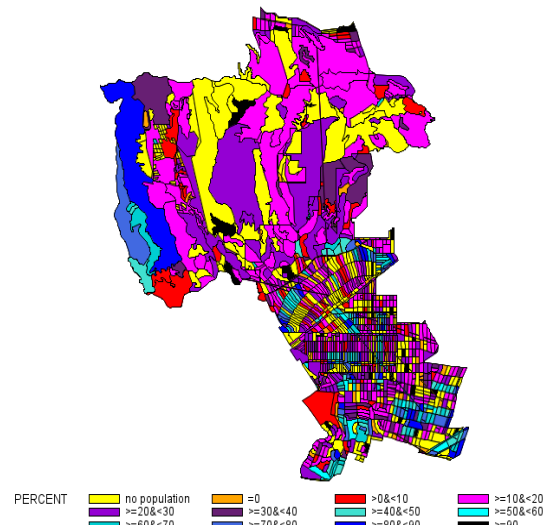
SAS/GRAPH Colors : Color-naming Schemes - SAS OnlineDoc, V8

<https://v8doc.sas.com/sashtml/gref/zgcscheme.htm>

Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1
Beverly Hills Blocks in 27 Census Tracts M.D. Pohl SESUG 2019



Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1
Beverly Hills Blocks in 27 Census Tracts M.D. Pohl SESUG 2019



```
PATTERN1 VALUE=MS COLOR= GHOSTWHITE;
PATTERN2 VALUE=MS COLOR= CXFFCCE0;
PATTERN3 VALUE=MS COLOR= CXFF3385;
PATTERN4 VALUE=MS COLOR= CX99003D;
PATTERN5 VALUE=MS COLOR= CXB3B3FF;
PATTERN6 VALUE=MS COLOR= CX3333FF;
PATTERN7 VALUE=MS COLOR= CX000080;
```

TITLE1 "Selected Tracts in Los Angeles County CA: Percent Minority 2010 Census SF 1";
TITLE2 "Beverly Hills Blocks in 27 Census Tracts M.D. Pohl SESUG 2019";

BAYES' THEOREM:

Bayes' Theorem, was named after the 18th-century British mathematician, Thomas Bayes (1701–1761). Richard Price edited Bayes' unpublished manuscript which was posthumously published in "An Essay towards solving a Problem in the Doctrine of Chances" in 1763. In 1774, the French mathematician Pierre-Simon Laplace reproduced and expanded Bayes' results. Laplace was unaware of Bayes' work and developed in large part, the Bayesian interpretation of probability.

Bayes' Theorem is a mathematical formula for determining conditional probability. It is as follows:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}, \text{ where } A \text{ and } B \text{ are events and } P(B) \neq 0.$$

Both $P(A/B)$ and $P(B/A)$ are conditional probabilities (also called posterior or revised probabilities).

$P(A/B)$ is the likelihood of event A occurring given that B is true.

$P(B/A)$ is the likelihood of event B occurring given that A is true.

Both $P(A)$ and $P(B)$ are marginal probabilities (also called a priori or prior probabilities).

$P(A)$ is the probability of event A independent of event B.

$P(B)$ is the probability of event B independent of event A.

SAS VISUALIZATION USING THE BISG (BAYESIAN IMPROVED SURNAME GEOCODING) PROXY

Consumer Financial Protection Bureau (CFPB) published "Using publicly available information to proxy for unidentified race and ethnicity: A methodology and assessment" in the Summer of 2014⁷. This publication has an application of Bayes' Theorem called BISG (Bayesian Improved Surname Geocoding proxy method). This proxy for race/ethnicity is used when the race/ethnicity information is not available or not required to be reported. To arrive at a

⁷ http://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf

single probability for each race/ethnicity, the methodology uses both the racial/ethnic composition of the geographic location and the surname. This methodology is more efficient than using geographic location alone or surname alone. This publication uses the Census 2000 Surname List (released 2007) instead of the 2010 Census Surname List (released in December 2016) since this the 2010 was not available at the time of publication. However, the BISG paper used the 2010 Census Population 18 and over for the geographic component.

SMITH is the most frequent surname in the United States in both Decennial Censuses. The BISG publication contains an example of the computation of the surname of SMITH and the State of California using the Census 2000 Surname List for the total population 18 years and older from the 2010 Census. According to the Census 2000 Surname List, there were 2,376,206 individuals with the last name "Smith"⁸. The example uses non-Hispanic Black. According to the Census 2000 Surname list, 22.22% of the population in the U.S. have a surname of Smith reporting to be non-Hispanic Black while in the 2010 Census the percentage was 23.11. I computed the BISG for SMITH for non-Hispanic Black using the State of California for both the Census 2000 and the 2010 Census Summary File 1 for the total population 18 years and older. It is 16.61% and 16.80%, respectively.

STATE OF CALIFORNIA FOR SMITH M.D. Pohl SESUG 2019 USING 2010 CENSUS POPULATION 18 AND OVER: CENSUS SURNAME LISTS: 2000 AND 2010

2000/2010 Census Surname List							BISG computation for the State of California						
YEAR	PW	PB	PA	PI	PT2	PH	surname	BISGW	BISGB	BISGA	BISGI	BISGT2	BISGH
2000	73.35	22.22	0.40	0.85	1.63	1.56	SMITH	72.00	16.61	1.65	0.83	3.54	5.37
2010	70.90	23.11	0.50	0.89	2.19	2.40	SMITH	67.68	16.80	2.01	0.84	4.63	8.04

SAS CODE TO COMPUTE BISG (BAYESIAN IMPROVED SURNAME GEOCODING) PROXY EXAMPLE.

```

/*CENSUS 2000 SURNAME LIST FOR SMITH*/
/*PW=73.35/100; PB=22.22/100; PA=0.40/100; PI=0.85/100; PT2=1.63/100; PH=1.56/100;*/
/*CENSUS 2010 SURNAME LIST FOR SMITH*/
/*PW=70.90/100; PB=23.11/100; PA=0.50/100; PI=0.89/100; PT2=2.19/100; PH=2.40/100;*/
TITLE1 "STATE OF CALIFORNIA FOR SMITH M.D. Pohl SESUG 2019";
TITLE2 "USING 2010 CENSUS POPULATION 18 AND OVER: CENSUS SURNAME LISTS: 2000 AND 2010";
DATA IPD; input YEAR $1-5 @10 surname $14. @26 PW PB PA PI PT2 PH;
CARDS;
2000 SMITH 73.35 22.22 0.40 0.85 1.63 1.56
2010 SMITH 70.90 23.11 0.50 0.89 2.19 2.40
;

/*USES TOTAL POPULATION 18 AND OVER*/
DATA CA1; SET IPD;
T1= 234564071; T= 27958916;
W1= 157444597; W= 12461055;
B1= 27464591; B= 1655298;
A1= 11901269; A= 3968506;
I1= 1609046; I= 126421;
T21= 2797866; T2= 490137;
H1= 33346703; H= 9257499;
/*PROC PRINT DATA=HELLO; RUN;*/
DATA CA2; SET CA1;
T0=(T/T1); W0=(W/W1); B0=(B/B1); I0=(I/I1); A0=(A/A1); T20=(T2/T21); H0=(H/H1);
/*PROC PRINT DATA=HIHI; RUN;*/
DATA CA3; SET CA2;
BISGH=((PH*H0)/((PW*W0)+(PB*B0)+(PA*A0)+(PI*I0)+(PT2*T20)+(PH*H0))) *100;
BISGW=((PW*W0)/((PW*W0)+(PB*B0)+(PA*A0)+(PI*I0)+(PT2*T20)+(PH*H0))) *100;
BISGB=((PB*B0)/((PW*W0)+(PB*B0)+(PA*A0)+(PI*I0)+(PT2*T20)+(PH*H0))) *100;
BISGA=((PA*A0)/((PW*W0)+(PB*B0)+(PA*A0)+(PI*I0)+(PT2*T20)+(PH*H0))) *100;
BISGI=((PI*I0)/((PW*W0)+(PB*B0)+(PA*A0)+(PI*I0)+(PT2*T20)+(PH*H0))) *100;
BISGT2=((PT2*T20)/((PW*W0)+(PB*B0)+(PA*A0)+(PI*I0)+(PT2*T20)+(PH*H0))) *100;
DENOM=(PW*W0)+(PB*B0)+(PA*A0)+(PI*I0)+(PT2*T20)+(PH*H0);
PROC PRINT DATA=CA3; VAR YEAR PW PB PA PI PT2 PH SURNAME BISGW BISGB BISGA BISGI BISGT2 BISGH;
FORMAT PW PB PA PI PT2 PH BISGW BISGB BISGA BISGI BISGT2 BISGH 6.2; RUN;

```

The computation uses the Census population 18 and over using 2010 Census SF 1 for the geographic component; and the 2010 Census Surname list with additional examples at the County and 5-digit zip code areas (ZCTAs). I selected

⁸ https://www.census.gov/topics/population/genealogy/data/2010_surnames.html
https://www.census.gov/topics/population/genealogy/data/2000_surnames.html

additional surnames from the 2010 Census Surname list that are "Heavily" (75% or more – White, Not Hispanic; Black, Not Hispanic; Asian, Not Hispanic; Native American, Not Hispanic; Hispanic, regardless of race; and not "Heavily" (which includes "SMITH"). Below are the results for the State of California.

STATE OF CALIFORNIA FOR OTHER SELECTED SURNAMES: USING 2010 CENSUS POPULATION 18 AND OVER: 2010 CENSUS SURNAME LISTS: CATEGORY: 75% OR MORE: HEAVILY M.D. Pohl SESUG 2019

2010 Census Surname List							BISG computation for the State of California						
HEAVILY	PW	PB	PA	PI	PT2	PH	surname	BISGW	BISGB	BISGA	BISGI	BISGT2	BISGH
White	94.89	0.52	0.66	0.34	1.26	2.33	BECKER	86.76	0.36	2.54	0.31	2.55	7.47
White	94.83	0.46	0.77	0.49	1.38	2.07	CARLSON	86.82	0.32	2.97	0.45	2.80	6.65
White	95.05	0.37	0.65	0.57	1.44	1.92	ERICKSON	87.56	0.26	2.52	0.52	2.94	6.20
White	94.62	0.69	0.64	0.30	1.26	2.50	GALLAGHER	86.26	0.48	2.46	0.27	2.54	7.99
White	94.79	0.49	0.74	0.62	1.41	1.95	LARSON	87.08	0.34	2.86	0.57	2.87	6.28
White	94.78	0.63	0.00	0.00	2.09	1.88	MCMURPHY	89.01	0.45	0.00	0.00	4.34	6.19
White	94.84	0.52	0.71	0.29	1.37	2.28	MEYER	86.58	0.36	2.73	0.26	2.77	7.30
White	84.11	10.76	0.54	0.66	1.77	2.17	MILLER	78.78	7.67	2.13	0.61	3.67	7.13
White	94.76	0.40	0.66	0.72	1.43	2.03	OLSON	87.06	0.28	2.55	0.66	2.91	6.54
White	95.15	0.37	0.61	0.34	1.19	2.34	SCHMIDT	87.15	0.26	2.35	0.31	2.41	7.52
Black	5.17	87.53	0.30	0.68	3.78	2.54	WASHINGTON	5.68	73.22	1.39	0.74	9.19	9.79
Black	14.05	80.09	0.18	0.18	3.42	2.07	WEATHERSPOON	15.47	67.16	0.84	0.20	8.34	8.00
Black	13.75	80.40	0.25	0.15	3.11	2.35	RUFFIN	15.06	67.06	1.15	0.16	7.54	9.03
Black	1.26	95.95	0.20	0.20	1.33	1.06	OWUSU	1.54	89.07	1.03	0.24	3.59	4.53
Black	3.72	90.21	0.25	0.26	2.78	2.78	GADSON	4.15	76.64	1.18	0.29	6.87	10.88
Black	11.09	83.26	0.12	0.42	2.93	2.19	MCKOY	12.38	70.78	0.56	0.47	7.24	8.57
Black	7.05	86.74	0.31	0.92	2.23	2.75	PIERRE	7.84	73.47	1.45	1.02	5.49	10.73
Black	13.76	79.83	0.22	0.40	3.28	2.51	ALSTON	14.97	66.12	1.01	0.43	7.90	9.58
Black	12.93	79.33	0.54	0.52	3.48	3.20	ARTIS	13.60	63.55	2.39	0.54	8.10	11.81
Black	14.58	76.89	0.64	2.25	3.05	2.58	ANTOINE	15.53	62.38	2.87	2.38	7.19	9.64
Asian	2.37	0.87	88.84	0.03	4.18	3.71	CHANG	0.59	0.17	93.66	0.01	2.32	3.26
Asian	1.40	0.30	96.12	0.02	1.64	0.52	CHEN	0.34	0.06	98.28	0.00	0.88	0.44
Asian	0.90	0.11	97.14	0.02	1.42	0.40	HUANG	0.22	0.02	98.66	0.00	0.76	0.34
Asian	6.89	3.86	81.25	0.54	5.27	2.20	KHAN	1.85	0.79	92.00	0.14	3.14	2.07
Asian	2.52	0.39	94.47	0.02	1.96	0.65	KIM	0.62	0.07	97.68	0.00	1.06	0.56
Asian	1.79	0.17	95.62	0.02	1.86	0.54	LIU	0.44	0.03	98.06	0.00	1.00	0.46
Asian	0.95	0.12	96.45	0.03	1.83	0.63	NGUYEN	0.23	0.02	98.23	0.01	0.98	0.53
Asian	2.10	0.38	94.78	0.65	1.65	0.44	PATEL	0.52	0.07	97.98	0.16	0.90	0.38
Asian	0.90	0.20	96.33	0.03	1.81	0.72	PHAM	0.22	0.04	98.16	0.01	0.97	0.61
Asian	4.29	4.48	82.77	1.15	4.81	2.50	SINGH	1.14	0.90	92.50	0.30	2.82	2.33
Asian	1.35	0.10	96.00	0.03	1.70	0.81	TRAN	0.33	0.02	98.05	0.01	0.91	0.69
Asian	2.59	0.29	95.24	0.02	1.50	0.35	WANG	0.63	0.05	98.19	0.00	0.81	0.30
Asian	1.37	0.18	96.37	0.02	1.59	0.48	WU	0.33	0.03	98.37	0.00	0.85	0.41
Asian	0.72	0.22	98.19	0.00	0.66	0.00	XIE	0.17	0.04	99.44	0.00	0.35	0.00
Asian	1.03	0.20	96.81	0.02	1.50	0.45	YANG	0.25	0.04	98.53	0.00	0.80	0.38
Asian	0.99	0.16	98.06	0.02	0.62	0.15	ZHANG	0.24	0.03	99.27	0.00	0.33	0.13
NatAm	0.00	0.00	0.00	98.00	0.00	1.45	APACHITO	0.00	0.00	0.00	95.03	0.00	4.97
NatAm	1.85	0.00	0.00	93.02	1.42	3.60	BECENTI	1.68	0.00	0.00	83.98	2.86	11.48
NatAm	1.28	0.00	0.00	97.81	0.00	0.00	BEDONIE	1.30	0.00	0.00	98.70	0.00	0.00
NatAm	1.95	0.44	0.29	93.84	1.70	1.78	BEGAY	1.83	0.31	1.15	87.33	3.53	5.85
NatAm	0.91	0.00	0.00	95.65	1.10	2.21	BEGAYE	0.86	0.00	0.00	89.54	2.30	7.31
NatAm	8.03	1.51	1.51	82.54	2.21	4.19	BIA	6.86	0.98	5.43	69.99	4.18	12.55
NatAm	1.83	0.00	0.00	94.14	1.10	2.75	CALABAZA	1.70	0.00	0.00	87.04	2.27	8.98
NatAm	13.90	1.15	0.00	81.38	2.72	0.00	CLY	13.68	0.86	0.00	79.53	5.93	0.00
NatAm	0.51	0.00	0.00	96.00	1.23	2.05	ETSITTY	0.48	0.00	0.00	90.14	2.58	6.80
NatAm	11.97	0.51	0.37	76.52	7.87	2.76	HARJO	10.23	0.33	1.33	64.94	14.89	8.28
NatAm	0.90	0.00	0.00	91.50	1.45	5.61	MANUELITO	0.79	0.00	0.00	79.25	2.80	17.17
NatAm	2.04	0.00	0.00	93.76	3.26	0.68	MANYGOATS	1.95	0.00	0.00	88.88	6.89	2.28
NatAm	1.18	0.00	0.00	95.00	1.50	2.23	TSOSIE	1.11	0.00	0.00	88.44	3.11	7.34
NatAm	1.52	0.00	0.00	96.28	0.00	1.35	WAUNEKA	1.49	0.00	0.00	93.86	0.00	4.65
NatAm	7.58	1.59	0.00	84.66	3.88	2.29	WHITEHORSE	6.93	1.11	0.00	76.78	7.85	7.34
NatAm	1.32	0.00	0.00	94.46	2.11	2.11	YELLOWHAIR	1.23	0.00	0.00	87.50	4.36	6.91
Hispanic	4.03	0.26	1.64	0.52	0.23	93.32	AGUILAR	1.19	0.06	2.04	0.15	0.15	96.42
Hispanic	4.29	0.33	0.53	0.23	0.20	94.42	ALVARADO	1.27	0.07	0.66	0.07	0.13	97.80
Hispanic	5.18	0.60	1.16	0.38	0.23	92.45	ALVAREZ	1.54	0.14	1.46	0.11	0.15	96.60
Hispanic	7.29	0.66	3.74	0.24	0.64	87.42	CASTRO	2.20	0.15	4.75	0.07	0.43	92.40

STATE OF CALIFORNIA FOR OTHER SELECTED SURNAMES: USING 2010 CENSUS POPULATION 18 AND OVER: 2010 CENSUS SURNAME LISTS: CATEGORY: 75% OR MORE: HEAVILY M.D. Pohl SESUG 2019

2010 Census Surname List							BISG computation for the State of California						
HEAVILY	PW	PB	PA	PI	PT2	PH	surname	BISGW	BISGB	BISGA	BISGI	BISGT2	BISGH
Hispanic	5.02	0.21	0.65	0.94	0.23	92.96	CHAVEZ	1.50	0.05	0.82	0.28	0.15	97.21
Hispanic	5.19	0.65	1.19	0.16	0.25	92.56	DIAZ	1.54	0.15	1.49	0.05	0.16	96.60
Hispanic	4.87	0.42	2.08	0.34	0.36	91.94	FLORES	1.44	0.09	2.60	0.10	0.24	95.53
Hispanic	4.86	0.57	1.02	0.38	0.25	92.92	LOPEZ	1.44	0.13	1.28	0.11	0.16	96.87
Hispanic	5.28	0.67	0.83	0.26	0.23	92.72	NUNEZ	1.57	0.15	1.04	0.08	0.15	97.00
Hispanic	5.34	0.52	1.09	0.31	0.24	92.50	PENA	1.59	0.12	1.37	0.09	0.16	96.67
Hispanic	4.96	0.45	1.17	0.20	0.26	92.95	PEREZ	1.47	0.10	1.46	0.06	0.17	96.73
Hispanic	3.89	0.31	0.93	0.20	0.19	94.48	RAMIREZ	1.14	0.07	1.15	0.06	0.12	97.45
Hispanic	6.49	0.98	4.89	0.19	0.65	86.79	RAMOS	1.94	0.22	6.17	0.06	0.43	91.18
Hispanic	4.99	0.44	0.43	0.32	0.22	93.61	RIOS	1.48	0.10	0.54	0.09	0.14	97.64
Hispanic	5.40	0.95	2.04	0.18	0.37	91.06	RIVERA	1.61	0.22	2.56	0.05	0.24	95.31
Hispanic	4.75	0.54	0.57	0.18	0.18	93.77	RODRIGUEZ	1.41	0.12	0.71	0.05	0.12	97.58
Hispanic	5.17	0.44	1.03	0.22	0.24	92.89	RUIZ	1.54	0.10	1.29	0.06	0.16	96.85
Hispanic	5.62	0.27	1.69	0.46	0.30	91.65	SALAZAR	1.67	0.06	2.12	0.14	0.20	95.81
Not Heavily	65.38	28.42	0.56	0.76	2.41	2.46	TAYLOR	62.81	20.79	2.27	0.72	5.12	8.29
Not Heavily	66.41	27.74	0.48	0.68	2.34	2.34	MOORE	64.09	20.39	1.95	0.65	5.00	7.92
Not Heavily	70.90	23.11	0.50	0.89	2.19	2.40	SMITH	67.68	16.80	2.01	0.84	4.63	8.04

GEOGRAPHIC AND SURNAME COMPONENTS:

The racial/ethnic categories as reported in the 2010 (and 2000) Census geographic component are listed below⁹:

Total:

Not Hispanic or Latino:

White alone

Black or African American alone

American Indian and Alaska Native alone

Asian alone

Native Hawaiian and Other Pacific Islander alone

Some Other Race alone

Two or More Races

Hispanic or Latino:

White alone

Black or African American alone

American Indian and Alaska Native alone

Asian alone

Native Hawaiian and Other Pacific Islander alone

Some Other Race alone

Two or More Races

The Census surname lists report the following racial/ethnic categories: pctwhite; pctblack; pctapi; pctaian; pct2prace; and pctHispanic. These pertain to: Not Hispanic: "White alone"; "Black or African American alone"; "Asian alone"; and "Native Hawaiian and Other Pacific Islander alone"; "American Indian and Alaska Native alone"; and Two or More Races"¹⁰. Please note that "Asian Alone" and "Native Hawaiian and other Pacific Islander alone" are aggregated to one category as "Asian". The sixth category is "Hispanic or Latino" (regardless of race).¹¹

The 2010 Census SF1 has tabulations that report counts of Hispanics and non-Hispanics by race. CFPB reallocated the "Some Other Race" counts to each of the remaining racial/ethnic categories using an "Iterative Proportional Fitting" procedure to make geography based demographic categories consistent with those on the census surname list. The CFPB has created three files for population age 18 and over for levels: block group, census tract and zip code (ZIP Code

⁹ 2010 Census Summary File 1: P5 HISPANIC OR LATINO ORIGIN BY RACE; Universe: Total population

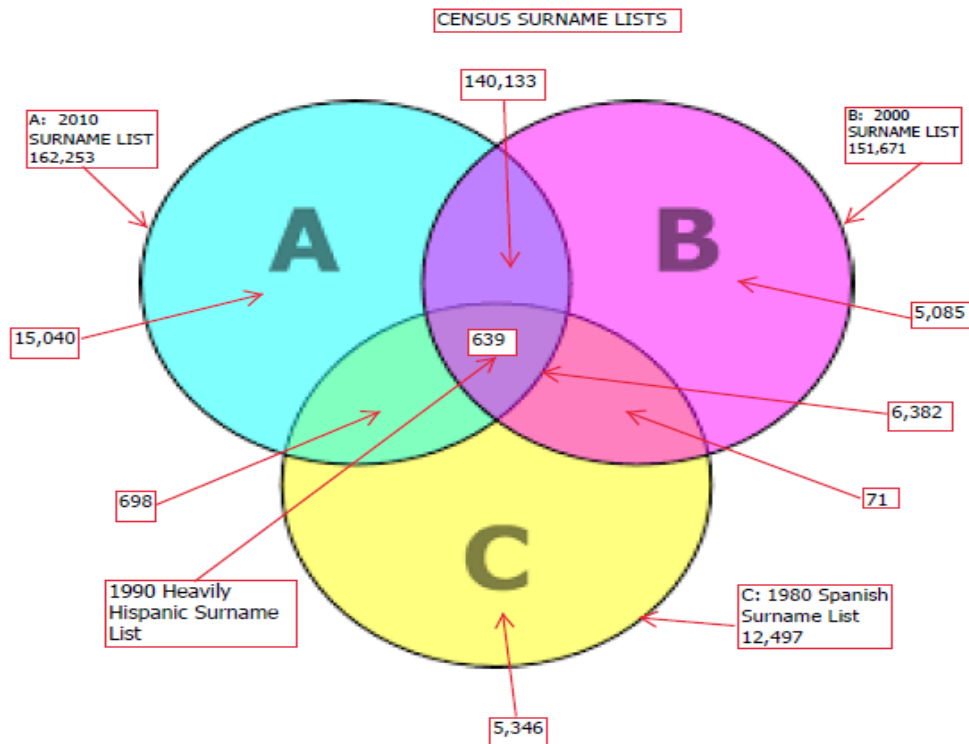
¹⁰ Also referred to as "Multiracial".

¹¹ As you can see, there is one category left out in the Census Surname Lists that would make the tally of these six categories equal to the Total Population: "Some Other Race alone".

Tabulation Areas (ZCTAs)) as defined by the U.S. Census Bureau. These three files are used in geographic probabilities for the proxy¹².

CENSUS SURNAME LISTS: NAMES FILES

- #1. 2010 Census Surname List ("2010 List") consisting of 162,253 observations of those surnames appearing on the 2010 Census at least 100 times;
- #2. Census 2000 Surnames List ("2000 List") consisting of 151,671 entries of those surnames appearing 100 times or More in the Census 2000;
- #3. 1990 Census Heavily Hispanic List contains 639 surnames "Heavily Hispanic List";
- #4. 1990 Census Surname List: Frequently Occurring Surnames from Census 1990 – Names Files with 88,799 entries;
- #5. 1980 Census "Passel-Word Spanish surname list with 12,497 entries.



Both the 2010 and the 2000 Census Surname lists ("Census names files") report those surnames with a frequency of 100 or more by race/ethnicity. The two Census names files have in common 146,515 surnames (140,133 + 6,382). The 2000 List has 5,156 surnames (5,085 + 71) not contained in the 2010 List; while the 2010 List has 15,738 surnames (15,040 + 698) not contained in the 2000 List. The 1980 Census Spanish Surname List contains 12,497 surnames (5,346 + 71 + 6,382 + 698)¹³. The 1980 Census Spanish Surname List contains 5,346 Spanish Surnames in neither the 2010 nor the 2000 Census List. The 1990 Heavily Hispanic Surname List containing 639 surnames are included in all three lists. The 639 are part of the 6,382 surnames which these 3 Lists have in common. These four Census Lists have a total of 172,755 surnames, of which 5,346 do not have corresponding probabilities by race/ethnicity from any Decennial Census. These other two Hispanic Surname Lists may be used to capture information lost due to data suppression.

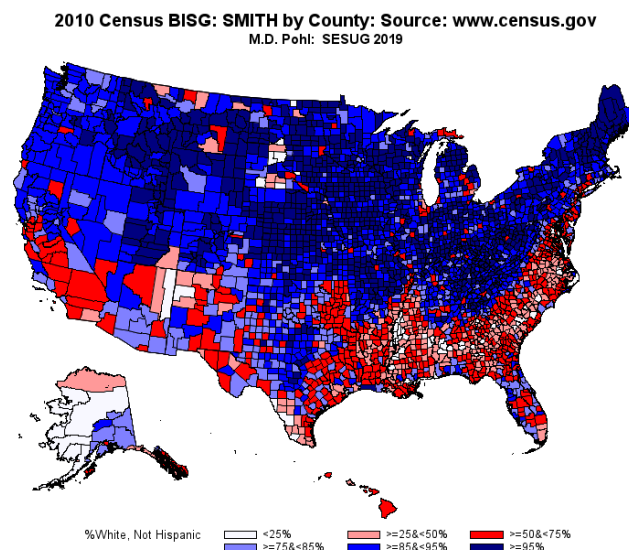
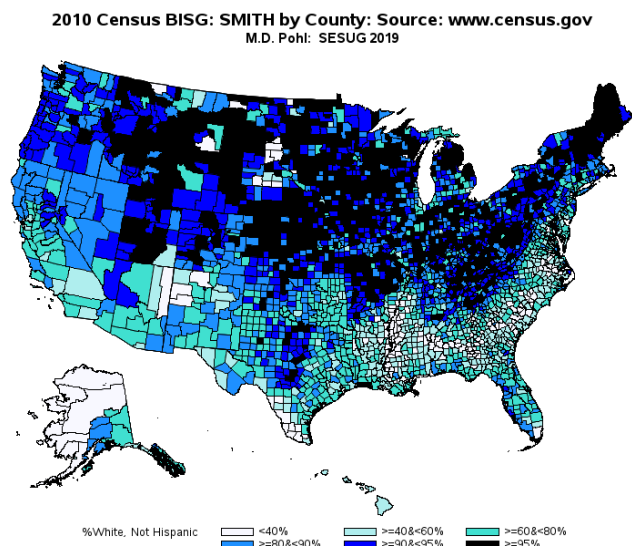
¹² They are available on the CFPB GitHub and are labeled:

- /input_files/created/blkgrp_attr_over18.dta
- /input_files/created/tract_attr_over18.dta
- /input_files/created/zip_attr_over18.dta

¹³ Jeffrey Passel and David Word produced a list of common Hispanic surnames for the 1980 Decennial Census. This resulted in 12,497 Hispanic surnames called the "Passel-Word (PW) Spanish surname list".

The Decennial Census Surname Lists have three types of data suppression – those surnames with a frequency of less than 100; cell suppression at the race/ethnicity level for a particular surname; and those surnames reporting Other Race, Not Hispanic. The Decennial Surname Lists have individual entries of “(S)” in the 2010 and 2000 Census Surname Lists representing “data suppression”. Since no information was provided on that particular entry, for calculation purposes, these entries were set to zero; but for calculating the average entry these were set to missing. The 2010 Census Surname List contains 162,253 names and represent “Frequently Occurring Surnames in the 2010 Census: Surnames Occurring at Least 100 Times Nationally”.

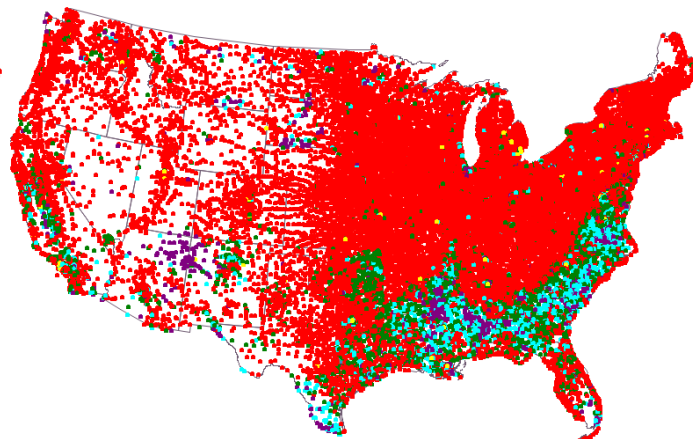
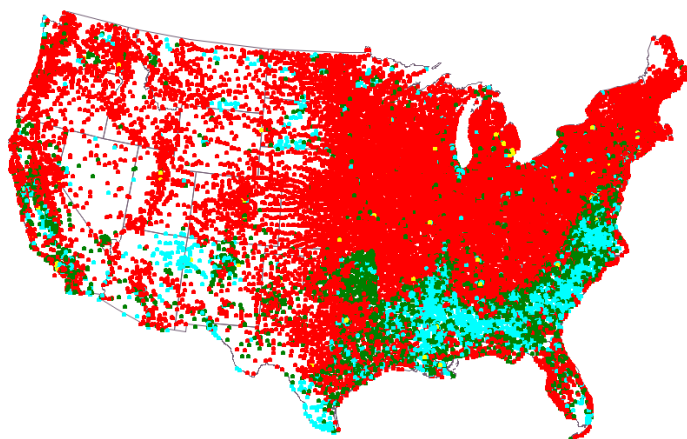
SAS VISUALIZATION USING THE BISG AT THE COUNTY LEVEL PNG USING SMITH.



PROC GMAP: BISG AT ZCTA LEVEL PNGS: VARIOUS CATEGORIES FOR SELECTED RACIAL/ETHNIC CATEGORIES: FOR SELECTED SURNAMES: SMITH, PEREZ, WEATHERSPOON

BISG: SMITH ZCTA: 2010 CENSUS %WHITE, NOT HISPANIC
YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%
M.D. Pohl: SESUG 2019

BISG: SMITH ZCTA: 2010 CENSUS %WHITE, NOT HISPANIC
YELLOW: MISSING OR ZERO; PURPLE: >0-<25%
CYAN: >25-<50%; GREEN: >=50-<75% RED: >=75%; M.D. Pohl: SESUG 2019

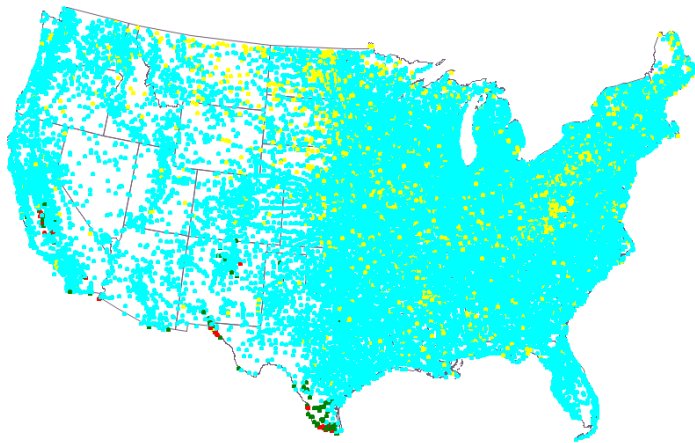


SMITH %White Not Hispanic: YELLOW: . or 0; CYAN: >0-<50%; GREEN: >=50-<80%; RED: >=80%

SMITH %White Not Hispanic: YELLOW: . or 0; PURPLE: >0-<25%; CYAN: >=25-<50%; GREEN: >=50-<75%; RED: >=75%

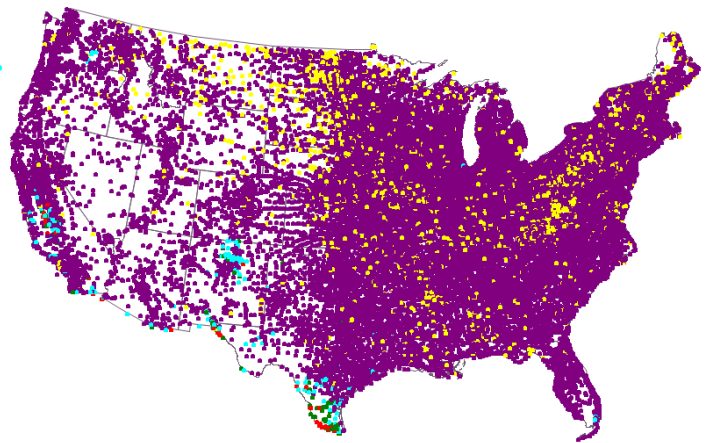
BISG: SMITH ZCTA: 2010 CENSUS %HISPANIC, REGARDLESS OF RACE

YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%
M.D. Pohl: SESUG 2019



BISG: SMITH ZCTA: 2010 CENSUS %HISPANIC, REGARDLESS OF RACE

YELLOW: MISSING OR ZERO; PURPLE: >0-<25%
CYAN: >25-<50%; GREEN: >=50-<75% RED: >=75%; M.D. Pohl: SESUG 2019

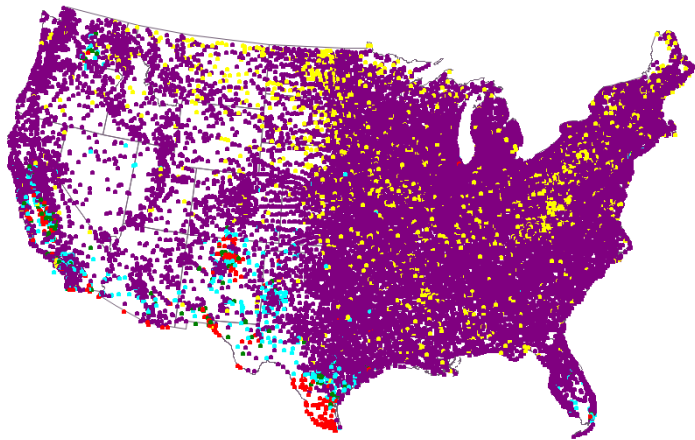


SMITH %Hispanic regardless of race YELLOW: . or 0; CYAN: >0-<50%; GREEN: >=50-<80%; RED: >=80%

SMITH %Hispanic regardless of race YELLOW: . or 0; PURPLE: >0-<25%; CYAN: >=25-<50%; GREEN: >=50-<75%; RED: >=75%

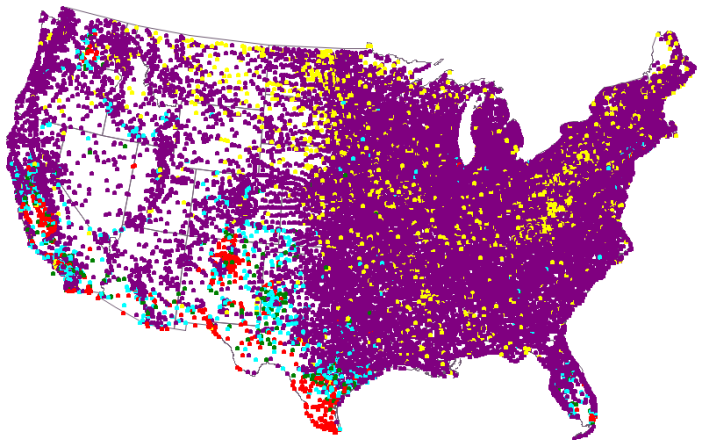
BISG: SMITH ZCTA: 2010 CENSUS %HISPANIC, REGARDLESS OF RACE

YELLOW: MISSING OR ZERO; PURPLE: >0-<10%
CYAN: >=10-<20%; GREEN: >=20-<30% RED: >=30%; M.D. Pohl: SESUG 2019



BISG: SMITH ZCTA: 2010 CENSUS %HISPANIC, REGARDLESS OF RACE

YELLOW: MISSING OR ZERO; PURPLE: >0-<5%
CYAN: >5-<10%; GREEN: >=10-<15% RED: >=15%; M.D. Pohl: SESUG 2019

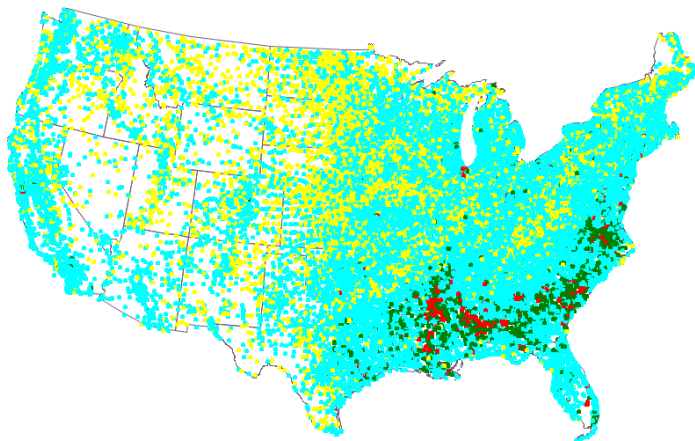


SMITH %Hispanic regardless of race YELLOW: . or 0; PURPLE: >0-<10%; CYAN: >=10-<20%; GREEN: >=20-<30%; RED: >=30%

SMITH %Hispanic regardless of race YELLOW: . or 0; PURPLE: >0-<5%; CYAN: >=5-<10%; GREEN: >=10-<15%; RED: >=15%

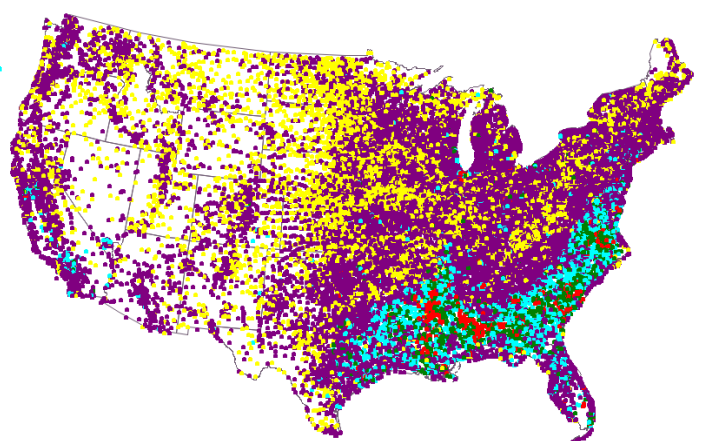
BISG: SMITH ZCTA: 2010 CENSUS %BLACK, NOT HISPANIC

YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%
M.D. Pohl: SESUG 2019



BISG: SMITH ZCTA: 2010 CENSUS %BLACK, NOT HISPANIC

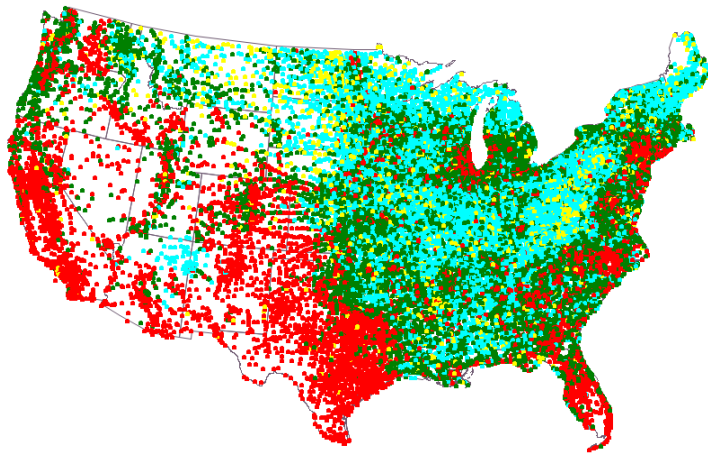
YELLOW: MISSING OR ZERO; PURPLE: >0-<25%
CYAN: >25-<50%; GREEN: >=50-<75% RED: >=75%; M.D. Pohl: SESUG 2019



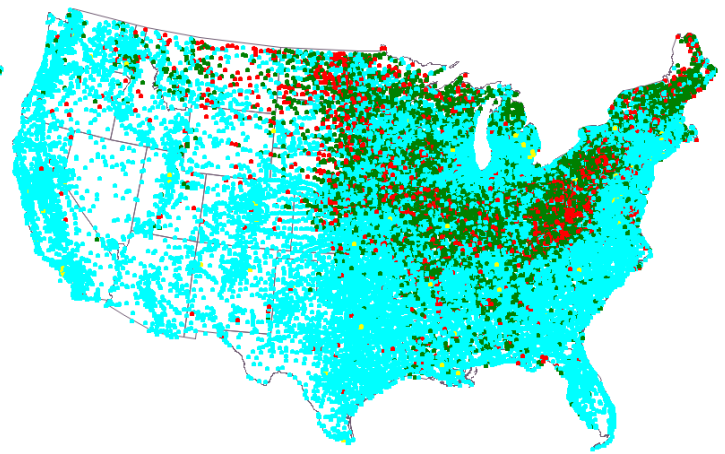
SMITH %Black Not Hispanic: YELLOW: . or 0; CYAN: >0-<50%; GREEN: >=50-<80%; RED: >=80%

SMITH %Black Not Hispanic: YELLOW: . or 0; PURPLE: >0-<25%; CYAN: >=25-<50%; GREEN: >=50-<75%; RED: >=75%

BISG: PEREZ ZCTA: 2010 CENSUS %HISPANIC, REGARDLESS OF RACE
 YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%
 M.D. Pohl: SESUG 2019

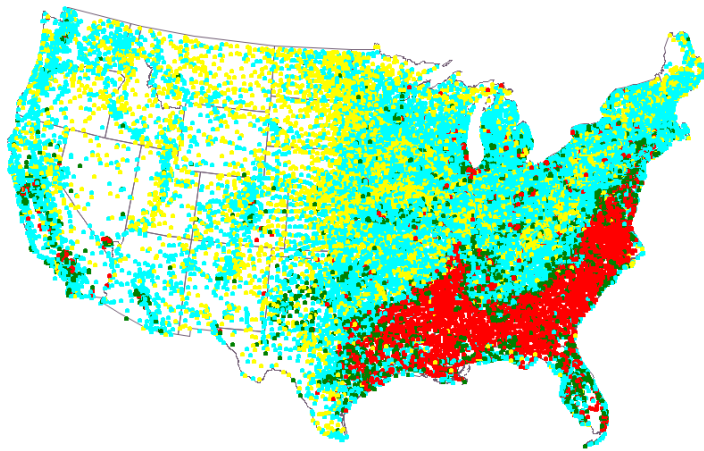


BISG: PEREZ ZCTA: 2010 CENSUS %WHITE, NOT HISPANIC
 YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%
 M.D. Pohl: SESUG 2019

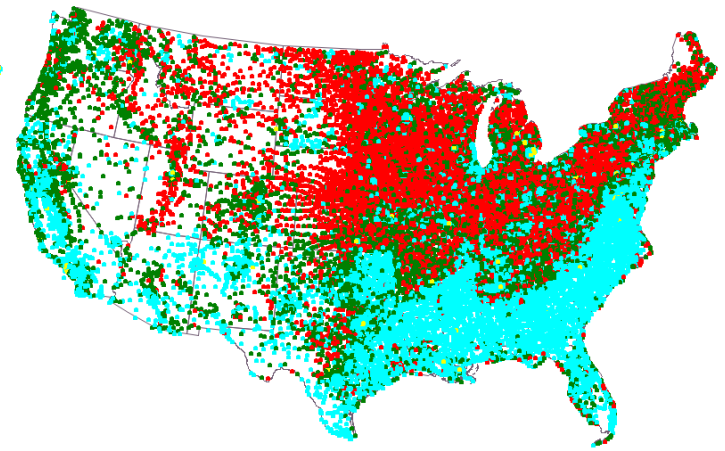


PEREZ %Hispanic, regardless of race: YELLOW: . or 0; CYAN: >0-<50%; GREEN: >=50-<80%; RED: >=80%
 PEREZ %White, Not Hispanic: YELLOW: . or 0; CYAN: >0-<50%; GREEN: >=50-<80%; RED: >=80%

BISG: WEATHERSPOON ZCTA: 2010 CENSUS %BLACK, NOT HISPANIC
 YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%
 M.D. Pohl: SESUG 2019



BISG: WEATHERSPOON ZCTA: 2010 CENSUS %WHITE, NOT HISPANIC
 YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%
 M.D. Pohl: SESUG 2019



WEATHERSPOON %Black, Not Hispanic: YELLOW: . or 0; CYAN: >0-<50%; GREEN: >=50-<80%; RED: >=80%
 PEREZ %White, Not Hispanic: YELLOW: . or 0; CYAN: >0-<50%; GREEN: >=50-<80%; RED: >=80%

SAS/GRAPH PROC GMAP CODE: BISG AT ZCTA LEVEL: SAS CODE TO GENERATE PNGS: SAS PROCS: FORMAT, GPROJECT, GMAP ¹⁴

```
goptions reset=all border;
data myzip;
input zip iii    BISGB    BISGA    BISGI    BISGT2    BISGH;
datalines;
. . . . .
run;
proc sort data=myzip;  by zip; run;

data longlat;
  merge myzip(in=mine)
        sashelp.zipcode(rename=(x=long y=lat)keep=x y zip state);
  by zip;
  if mine;
  x=atan(1)/45*long;  y=atan(1)/45*lat;
  x=-x;
  keep zip iii x y state; run;

data anno;
```

¹⁴ /*FOR OTHER SPECIAL CHARACTERS*/ proc gfont nobuild name=Special sr; run;


```

set longlat;
retain xsys ysys '2' function 'label' size 1.3 flag 1 when 'a';
style='special'; text='J'; /*M*/

IF III = . OR III=0 THEN COLOR='YELLOW';
IF III GT 0 AND III LT 50 THEN COLOR='CYAN';
IF III GE 50 AND III LT 80 THEN COLOR='GREEN';
IF III GE 80 THEN COLOR='RED';
output; run;
TITLE1 "BISG: SMITH ZCTA: 2010 CENSUS PERCENT WHITE, NOT HISPANIC";
TITLE2 "YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%";
TITLE3 "M.D. Pohl: SESUG 2019";
data all;
set maps.states anno;
where state not in (2,72,15);
run;

proc gproject data=all out=allp dupok; id state;
run; quit;

data map dot;
set allp;
if flag=1 then output dot;
else output map;
run;
pattern1 v=me c=grp r=50;

proc gmap data=map map=map; id state;
choro state / anno=dot nolegend;
run; quit;

proc sort data=myzip; by zip; run;

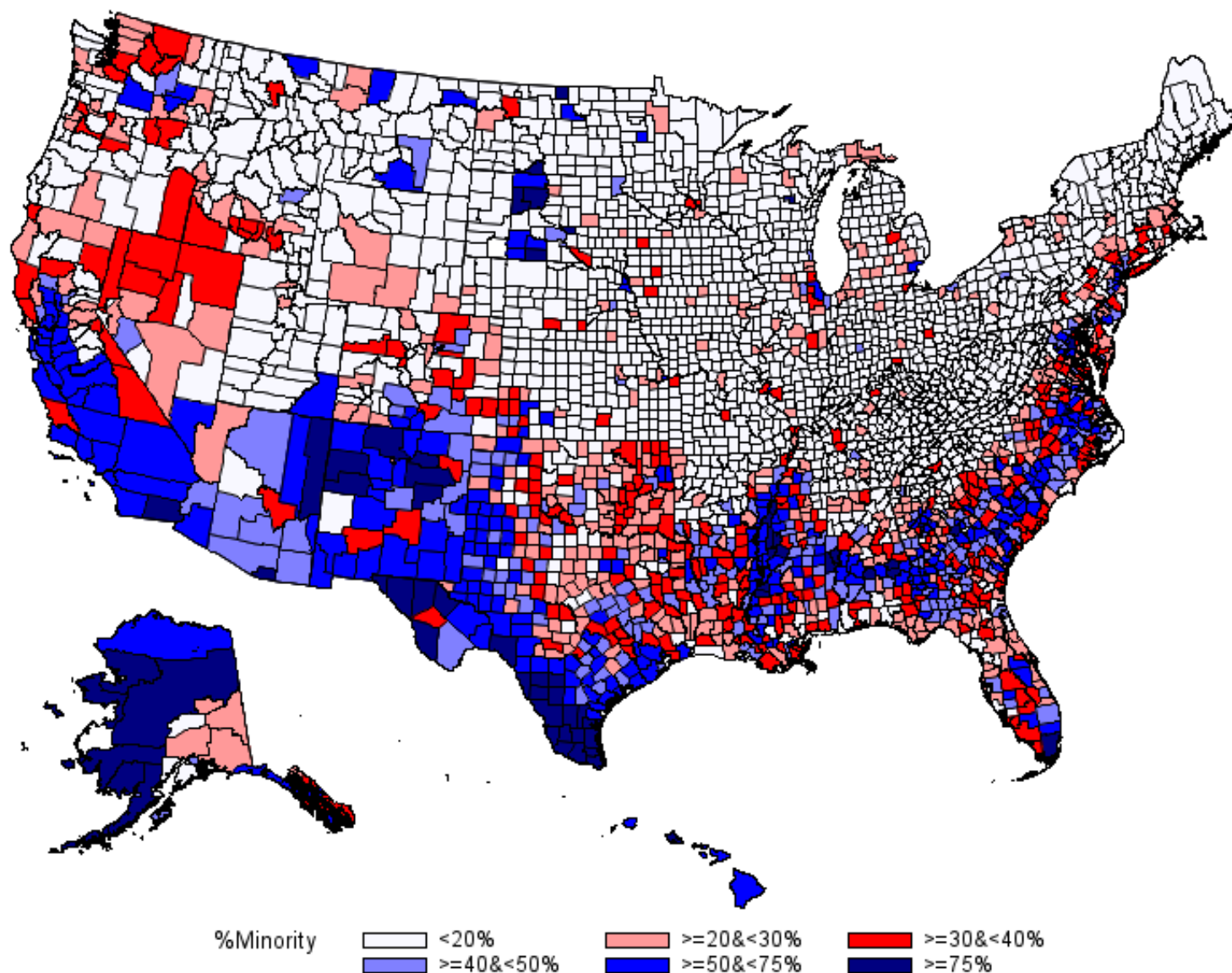
data longlat;
merge myzip(in=meme)
sashelp.zipcode(rename=(x=long y=lat)keep=x y zip state);
by zip;
if meme;
x=atan(1)/45*long; y=atan(1)/45*lat;
x=-x;
keep zip iii x y state;
run;

data anno;
set longlat;
retain xsys ysys '2' function 'label' size 1.3 flag 1 when 'a';
style='special';
text='J'; /*H*/ /*M*/
IF III = . OR III=0 THEN COLOR='YELLOW';
IF III GT 0 AND III LT 50 THEN COLOR='CYAN';
IF III GE 50 AND III LT 80 THEN COLOR='GREEN';
IF III GE 80 THEN COLOR='RED';
output; run;
TITLE1 "BISG: SMITH ZCTA: 2010 CENSUS PERCENT WHITE, NOT HISPANIC";
TITLE2 "YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%";
TITLE3 "M.D. Pohl: SESUG 2019";
data all;
set maps.states anno;
where state not in (2,72,15);
run;
proc gproject data=all out=allp dupok; id state;
run;
quit;

data map dot;
set allp;
if flag=1 then output dot;
else output map;
run;
pattern1 v=me c=grp r=50;
proc gmap data=map map=map; id state;
choro state / anno=dot nolegend;
run; quit;

```

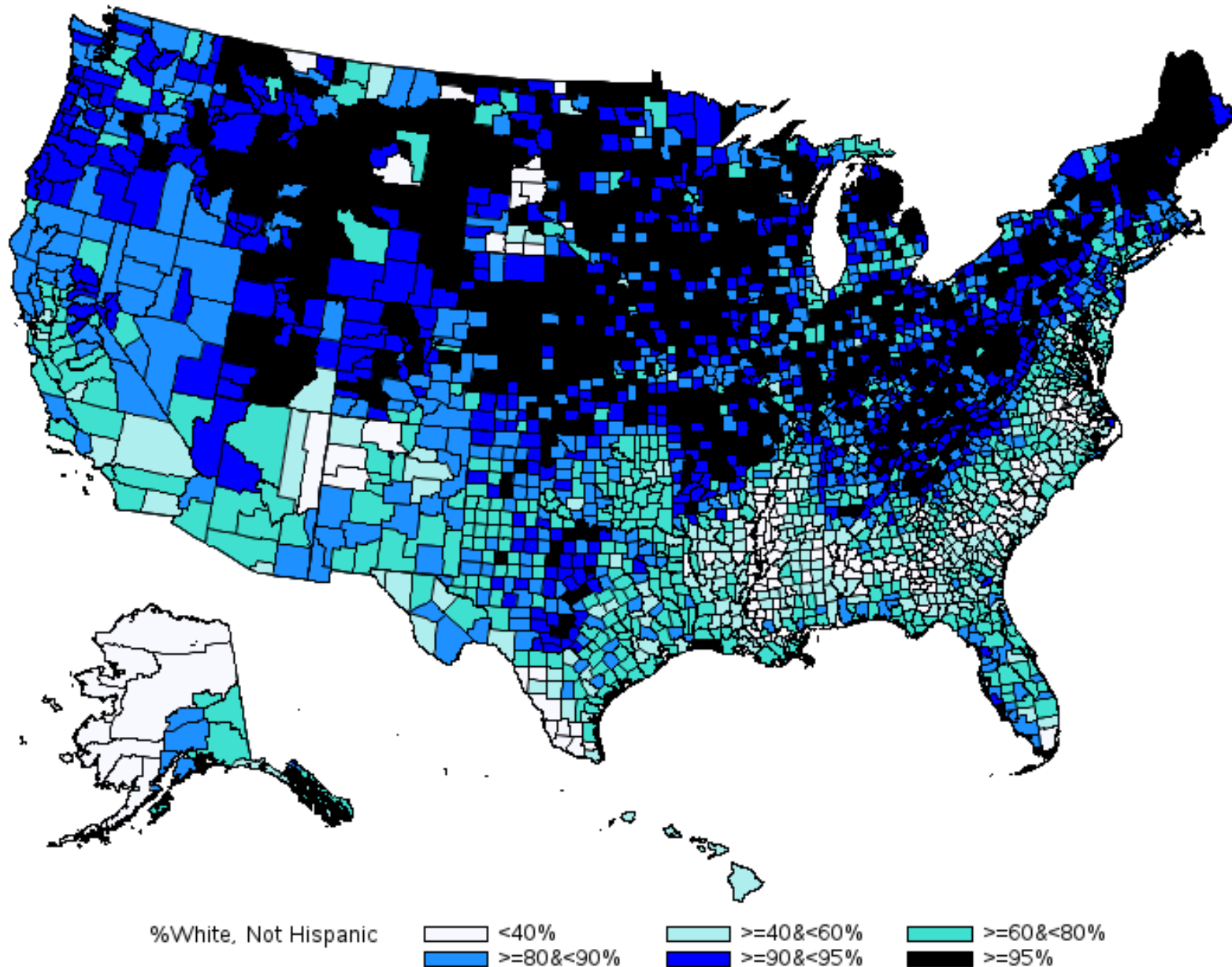
Percent Minority by County 2017 ACS 5-year estimate, Source: www.census.gov
M.D. Pohl SESUG 2019



American Community Survey (ACS):

2013-2017 ACS 5-year Estimates

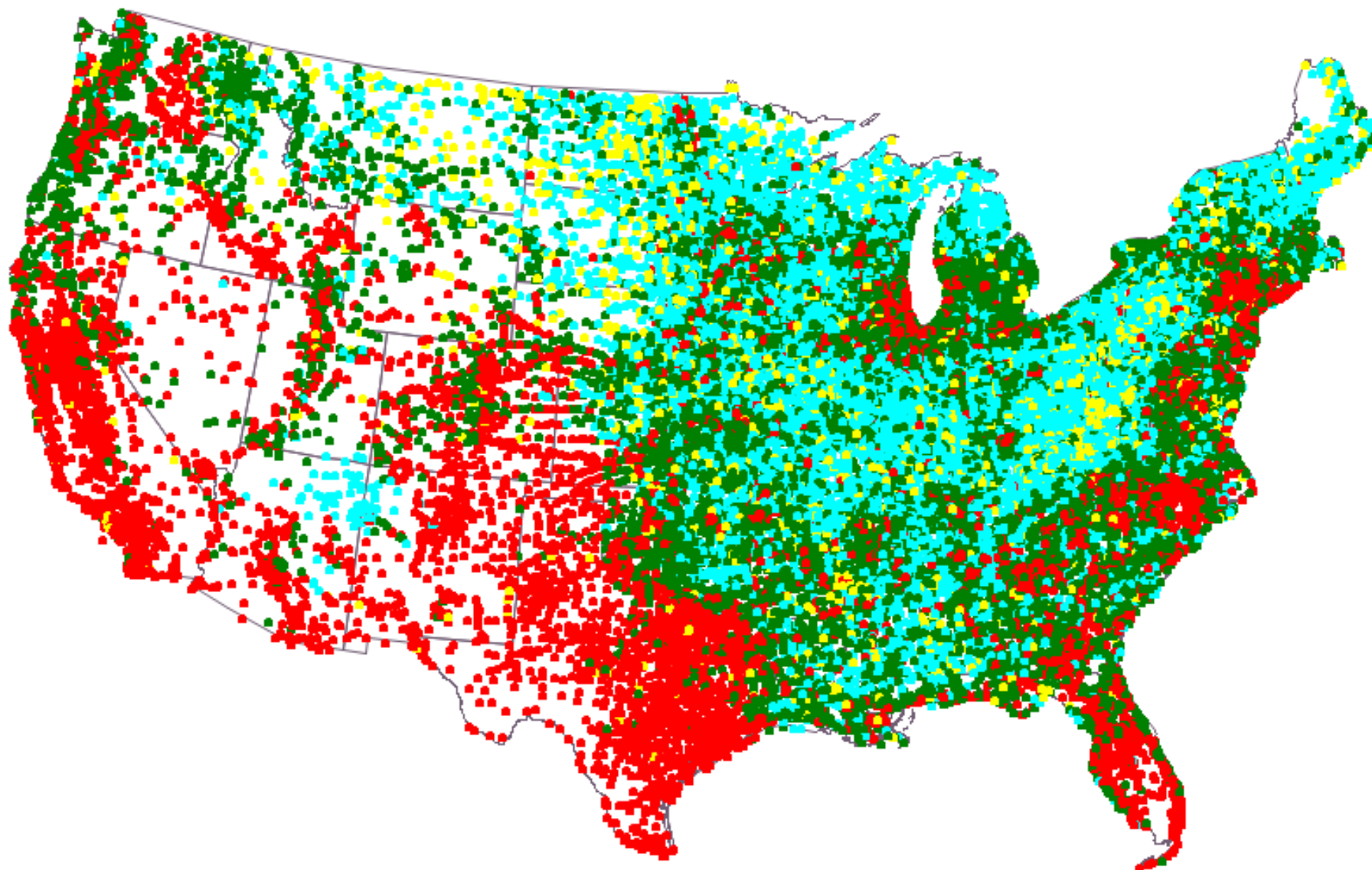
2010 Census BISG: SMITH by County: Source: www.census.gov
M.D. Pohl: SESUG 2019



BISG: PEREZ ZCTA: 2010 CENSUS %HISPANIC, REGARDLESS OF RACE

YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%

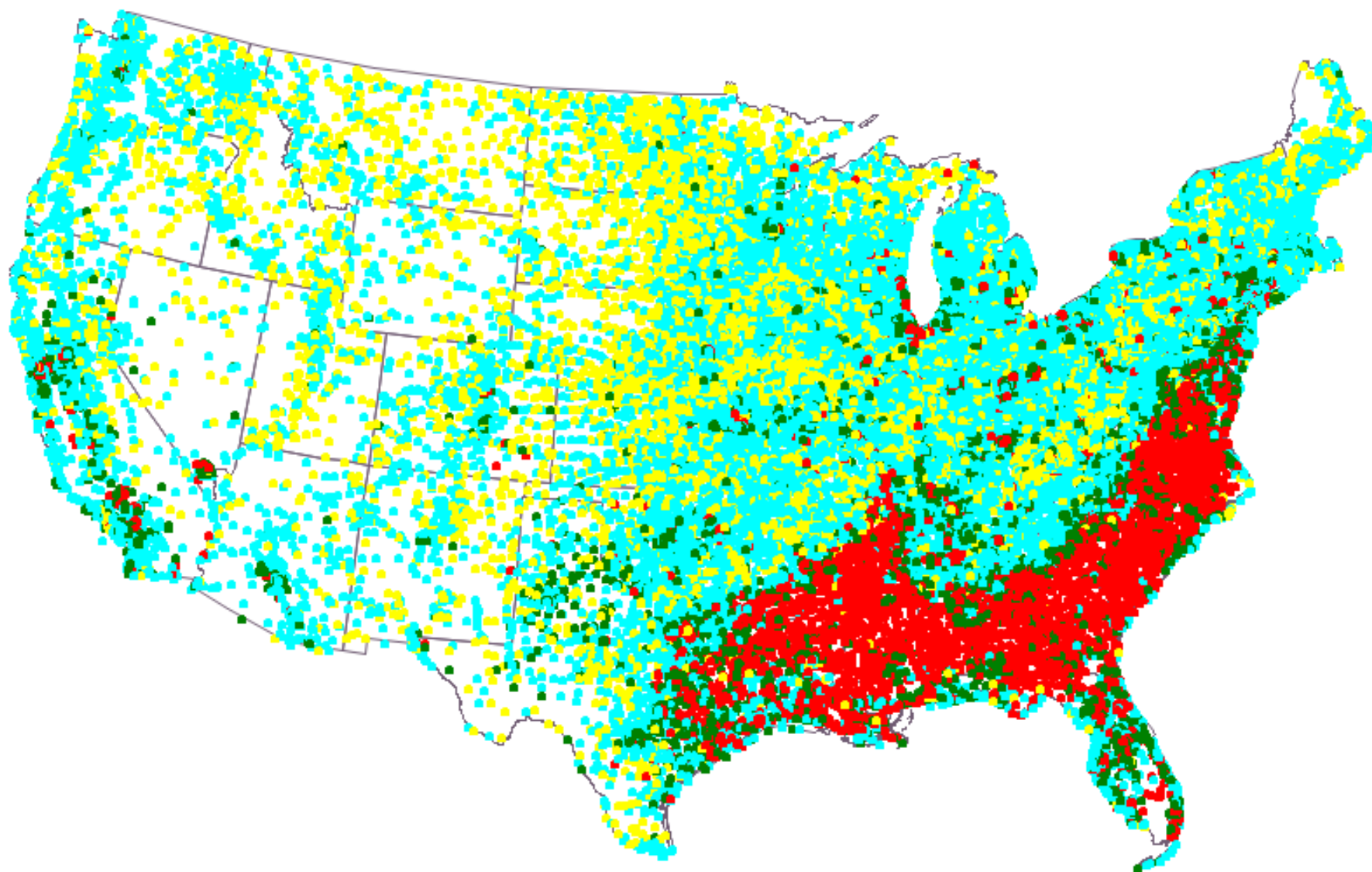
M.D. Pohl: SESUG 2019



BISG: WEATHERSPOON ZCTA: 2010 CENSUS %BLACK, NOT HISPANIC

YELLOW: MISSING OR ZERO CYAN: >0-<50% GREEN: >=50-<80% RED: >=80%

M.D. Pohl: SESUG 2019



CONCLUSION

The BISG proxy uses two components to arrive at an estimate. The surname component is a national frequency while the geographic component varies by geographic level and area. It is reported that if assignment of the race/ethnicity for an individual entry is to the highest probability, then this will have an accuracy of 80%. This probability can be used as a proxy for race/ethnicity and is more efficient than using geographic location alone or surname alone. An analysis of the racial/ethnic distribution by geographic area does have an impact on the BISG estimates and are changing over time.

REFERENCES

2010 CENSUS ZCTA TO TRACT RELATIONSHIP FILE:

https://www2.census.gov/geo/docs/maps-data/data/rel/zcta_tract_rel_10.txt?

Census Relationship Files <https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html>

2010 CENSUS TRACT REFERENCE MAPS: Last Revised: October 18, 2018

<https://www.census.gov/geographies/reference-maps/2010/geo/2010-census-tract-maps.html>

2010 CENSUS TRACT REFERENCE MAP: Los Angeles County, CA

https://www2.census.gov/geo/maps/dc10map/tract/st06_ca/c06037_los_angeles/DC10CT_C06037_000.pdf

"Map Created by Geography Division: December 01, 2010 "

GITHUB: BISG_RACE_ETHNICITY dated April 18, 2017

<https://github.com/cfpb/proxy-methodology/blob/master/README.md>

GITHUB CFPB <https://github.com/cfpb/proxy-methodology>

cfpb/proxy-methodology. INPUT FILES dated April 14, 2017.

https://github.com/cfpb/proxy-methodology/tree/master/input_files

Frequently Occurring Surnames from the 2010 Census, Last Revised: December 27, 2016

https://www.census.gov/topics/population/genealogy/data/2010_surnames.html

Frequently Occurring Surnames from the Census 2000, Last Revised: September 15, 2014

https://www.census.gov/topics/population/genealogy/data/2000_surnames.html

Color Selection: SAS/GRAPH® 9.4: Reference, Fifth Edition

<https://documentation.sas.com/api/docsets/graphref/9.4/content/graphref.pdf?locale=en#nameddest=n161ukdyz9wpfsn1nh8sihforvyq>

<https://documentation.sas.com/?docsetId=graphref&docsetTarget=p0edl20cvxxmm9n1i9ht3n21eict.htm&docsetVersion=9.4&locale=en>

Predefined Colors SAS/GRAPH® 9.4: Reference, Fifth Edition

<https://documentation.sas.com/?docsetId=graphref&docsetTarget=n161ukdyz9wpfsn1nh8sihforvyq.htm&docsetVersion=9.4&locale=en>

Color-Naming Schemes SAS® 9.4 Graph Template Language: User's Guide, Fifth Edition

<https://documentation.sas.com/?docsetId=grstatug&docsetTarget=p0edl20cvxxmm9n1i9ht3n21eict.htm&docsetVersion=9.4&locale=en>

Appendix C: Color Names Defined by SAS/GRAPH

https://support.sas.com/content/dam/SAS/support/en/books/pro-template-made-easy-a-guide-for-sas-users/62007_Appendix.pdf

SAS/GRAPH Colors : Color-naming Schemes - SAS OnlineDoc, V8

<https://v8doc.sas.com/sashtml/gref/zgscheme.htm>

RECOMMENDED READING:

The Royal Society, London, England: An Essay Towards Solving a Problem in the Doctrine of Chances. Letter is dated Nov. 10, 1763. <https://royalsocietypublishing.org/doi/pdf/10.1098/rstl.1763.0053>

Encyclopedia Britannica: Bayes's theorem: PROBABILITY, WRITTEN BY: Richard Routledge
<https://www.britannica.com/topic/Bayess-theorem>
<https://www.britannica.com/topic/Bayess-theorem/media/1/56808/218912>

Bayes formula. A.N. Shiryayev (originator), Encyclopedia of Mathematics. Last modified February 7, 2011.
https://www.encyclopediaofmath.org/index.php/Bayes_formula

CFPB: "Using publicly available information to proxy for unidentified race and ethnicity A methodology and assessment" dated summer 2014 https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf

FREQUENTLY OCCURRING SURNAMES IN THE 2010 CENSUS, Joshua Comenetz, October 2016
<https://www2.census.gov/topics/genealogy/2010surnames/surnames.pdf>

"Demographic Aspects of Surnames from Census 2000"
by David L. Word, Charles D. Coleman, Robert Nunziata and Robert Kominski.
<https://www2.census.gov/topics/genealogy/2000surnames/surnames.pdf?#>

Census: Genealogy <https://www.census.gov/topics/population/genealogy.html>

Frequently Occurring Surnames from the 1990 Census
https://www.census.gov/topics/population/genealogy/data/1990_census.html

Frequently Occurring Surnames from Census 1990 – Names Files
https://www.census.gov/topics/population/genealogy/data/1990_census/1990_census_namefiles.html

"Building a Spanish Surname List for the 1990's— A New Approach to an Old Problem"
by David L. Word and R. Colby Perkins, Jr. dated March 1996
<https://www.census.gov/population/documentation/twpno13.pdf>

"1980 Census of Population and Housing Spanish Surname List Technical Documentation" Revised 1993.
<https://www2.census.gov/prod2/decennial/documents/D1-D80-SPSN-14-TECH.pdf>

"APPENDIX E CENSUS LIST OF SPANISH SURNAMES"
<https://fcds.med.miami.edu/downloads/DataAcquisitionManual/dam2014/25%20Appendix%20E%20Census%20List%20of%20Spanish%20Surnames.pdf>

"Measuring Race and Ethnicity Across the Decades: 1790–2010 Mapped to 1997
U.S. Office of Management and Budget Classification Standards" by Beverly M. Pratt, Lindsay Hixson, and Nicholas A. Jones. https://www.census.gov/data-tools/demo/race/MREAD_1790_2010.html

Census: Appendix A. Geographic Terms and Concepts https://www2.census.gov/geo/pdfs/reference/GTC_10.pdf

Census: Geographic Terms and Definitions
<https://www.census.gov/programs-surveys/popest/about/glossary/geo-terms.html>

Census: Geographic Overview <https://www2.census.gov/geo/pdfs/reference/GARM/Ch2GARM.pdf>

Standard Hierarchy of Census Geographic Entities <https://www2.census.gov/geo/pdfs/reference/geodiagram.pdf?#>

Redistricting Data Program Management
<https://www.census.gov/programs-surveys/decennial-census/about/rdo/program-management.html>

Decennial Census P.L. 94-171 Redistricting Data, MAY 08, 2017

<https://www.census.gov/programs-surveys/decennial-census/about/rdo/summary-files.html>

2010 Census P.L. 94-171 Redistricting Data Summary Files, Technical Documentation PL/10-2 (RV) Issued January 2011

https://www2.census.gov/programs-surveys/decennial/rdo/about/2010-census-programs/2010Census_pl94-171_techdoc.pdf?#

Cartographic Boundary Files - Shapefile

<https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html>

2018 Cartographic Boundary Files Naming Convention

https://www2.census.gov/geo/tiger/GENZ2018/2018_file_name_def.pdf?#

Evaluating the Passel-Word Spanish Surname List: 1990 Decennial Census

Post Enumeration Survey Results, R. Colby Perkins, Population Division, U.S. Bureau of the Census, Washington, DC 20233, August 1993, Population Division Working Paper No. 4

<https://www.census.gov/population/www/documentation/twps0004.html>

Technical Working Paper No. 13, U.S. Census Bureau March 1996, Building a Spanish Surname List for the 1990's— A New Approach to an Old Problem, by David L. Word and R. Colby Perkins Jr., Population Division, U. S. Bureau of the Census, Washington D.C. <https://www.census.gov/population/documentation/twpno13.pdf>

1980 Census of Population and Housing, Summary Tape File 3. Technical Documentation

<https://www2.census.gov/prod2/decennial/documents/D1-D80-S300-14-TECH.pdf>

"Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity" dated October 30, 1997.

<https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf>

"Office of Management and Budget (OMB) DIRECTIVE NO. 15 Race and Ethnic Standards for Federal Statistics and Administrative Reporting (as adopted on May 12, 1977)". dated January 5, 2016.

<https://wonder.cdc.gov/wonder/help/populations/bridged-race/directive15.html>

CONTACT INFORMATION:

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

Martha D. Pohl

703-403-4249

maze44@verizon.net