

TAP TO GO  
BACK TO  
KIOSK MENU

# SAS<sup>®</sup> GLOBAL FORUM 2020

MARCH 29 - APRIL 1  
WASHINGTON, DC



USERS PROGRAM

Wendy B. Dickinson, Ph.D.  
Department of Educational and Psychological Studies



<https://youtu.be/Bz7lOwmpISI>

## Abstract

SAS provides an array of powerful tools to analytically examine-- and graphically depict-- large-scale datasets. This project explored the PROC FREQ functionality to produce mosaic plots for categorical data. Mosaic plots, as described by Hartigan and Kleiner (1981), are formed when numbers in a contingency table are represented by rectangles of areas proportional to the numbers, with shape and position rendered to expose deviations from independence models. The resulting visual depiction (collection of rectangles for the contingency table) is called a mosaic (Hartigan & Kleiner, 1981, 1984; Friendly, 1994, 2000, 2002). The colors and patterns displayed within the mosaic plots illustrate and define the relationships displayed by the categorical variable values.

The American Community Survey, 2018 Release, was utilized as the empirical data source. The American Community Survey (ACS) is conducted annually by the United States Census Bureau. Social, housing, economic, and demographic data comprise the 4 major survey sections of the ACS. Within these four major survey sections, the ACS yields rich, personal data covering more than 40 topic areas, such as educational level of attainment, commute time to work, type of housing, and ethnicity. By combining SAS functionality with varying combinations of categorical information from a large-scale national dataset, mosaic plots were generated to create a visual snapshot-in-time of the ACS respondents. This work thus provides a contemporary window into the daily lives of millions of Americans across our country.

## Abstract

Introduction

Methods

Results 1

Results 2

Conclusion

Please use the headings above to navigate through the different sections of the poster

## Introduction

In their most common form, “mosaic plots visualize relative cell frequencies from a two-variable contingency table as areas of rectangles” (Gromping, 2014, p. 108).

Mosaic plots are state-of-the-art graphical displays utilized for multivariate categorical data in statistical visualization (Unlu et al, 2013). As Hartigan & Kleiner (1984) related,

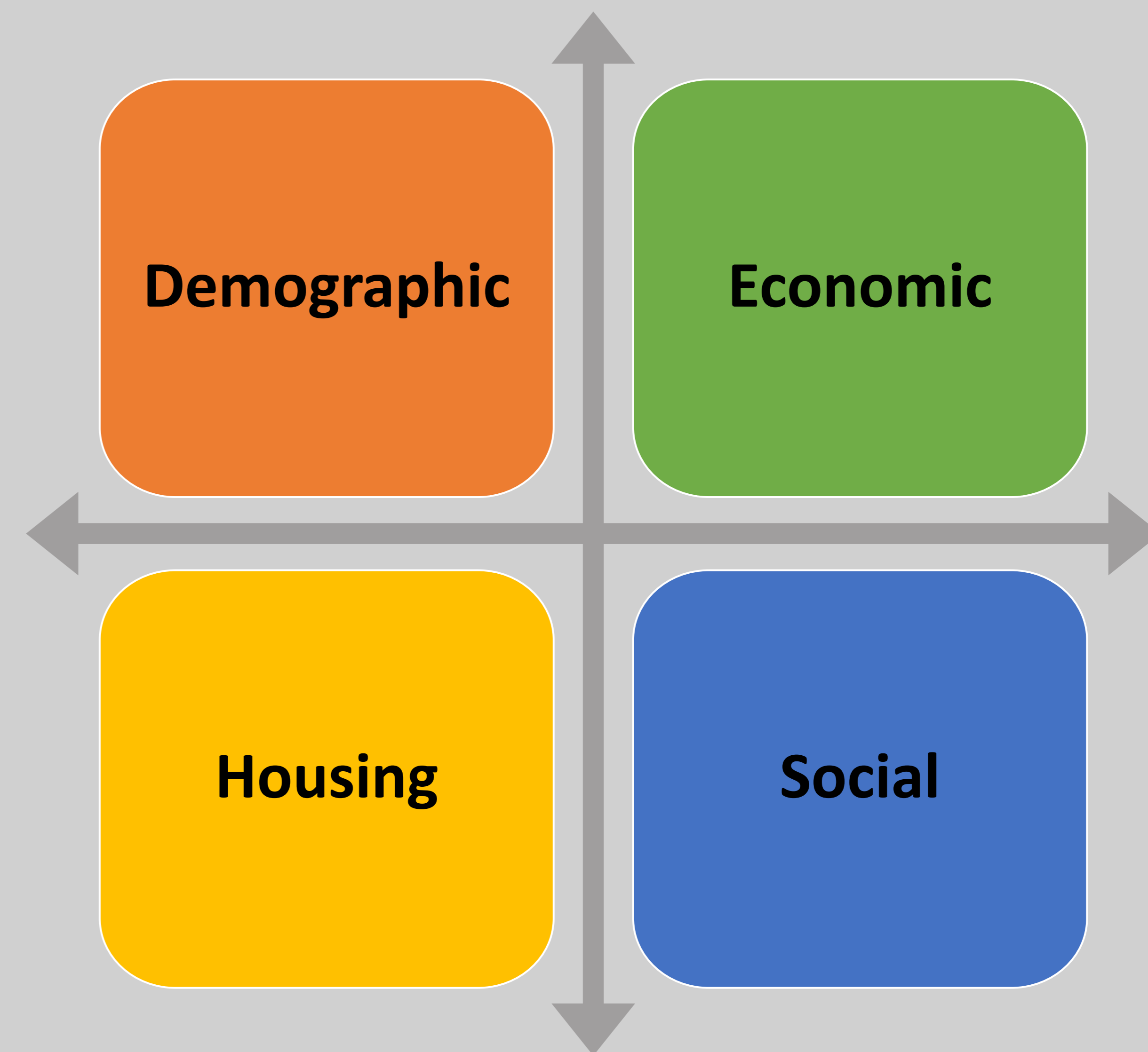
*A mosaic is a graphical display of cross-classified data in which each count is represented by a rectangle of area proportional to the count. The positions and sides of the rectangles are set to encourage comparisons between counts in the figures. Mosaics are useful for discovering unusually high or small counts and for discovering dependencies between variables* (Hartigan & Kleiner, 1984, p. 32).

Theus (2012) referred to mosaic plots as the “Swiss Army knife of displays for categorical data” (Theus, 2012, p.191). Indeed, it is this versatility of mosaic plots to display categorical relationships that continues to inspire their use across multiple disciplines. By combining SAS functionality with varying combinations of categorical information from a large-scale national dataset, we visually depicted variable relationships discovered through survey responses.

## Objective

The objective was to generate mosaic plots to create a visual snapshot-in-time of the American Community Survey respondents. This work thus provides a contemporary window into the daily lives of millions of Americans across our country. To provide an interactive learning experience, my graduate students were invited to contribute a mosaic plot for this presentation.

### Four major topic areas of the American Community Survey



Abstract

Introduction

Methods

Results 1

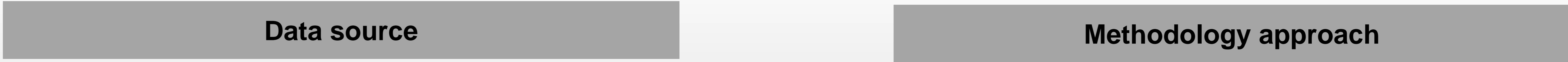
Results 2

Conclusion

Please use the headings above to navigate through the different sections of the poster

- Abstract
- Introduction
- Methods**
- Results 1
- Results 2
- Conclusion

Please use the headings above to navigate through the different sections of the poster



I. Variables of interest selected from the ACS 2018 item responses

**Demographic**

- Age
- Sex
- Group Quarters Population
- Hispanic or Latino Origin
- Race
- Relationship to Householder
- Total Population

**Economic**

- Class of Worker
- Commuting (Journey to Work) and Place of Work
- Employment Status
- Food Stamps/Supplemental Nutrition Assistance Program (SNAP)
- Health Insurance Coverage
- Income and Earnings
- Industry and Occupation
- Poverty Status
- Work Status Last Year

**Housing**

- Bedrooms
- Computer and Internet Use
- House Heating Fuel, Kitchen Facilities
- Occupancy/Vacancy Status
- Occupants per Room
- Plumbing Facilities
- Rent, Rooms
- Selected Monthly Owner Costs
- Telephone Service Available
- Tenure (Owner/Renter)
- Units in Structure, Value of Home
- Vehicles Available
- Year Householder Moved Into Unit
- Year Structure Built

**Social**

- Ancestry
- Citizen Voting-Age Population
- Citizenship Status, Disability Status
- Educational Attainment
- Fertility
- Grandparents as Caregivers
- Language Spoken at Home
- Marital History, Marital Status
- Migration/Residence 1 Year Ago
- Place of Birth, School Enrollment
- Undergraduate Field of Degree
- Veteran Status; Period of Military Service
- Year of Entry

II. SAS code written to invoke the mosaicplotparm template

III. Mosaic plot output and graphs examined for visual relationships and patterns



- Abstract
- Introduction
- Methods
- Results 1
- Results 2
- Conclusion

Please use the headings above to navigate through the different sections of the poster

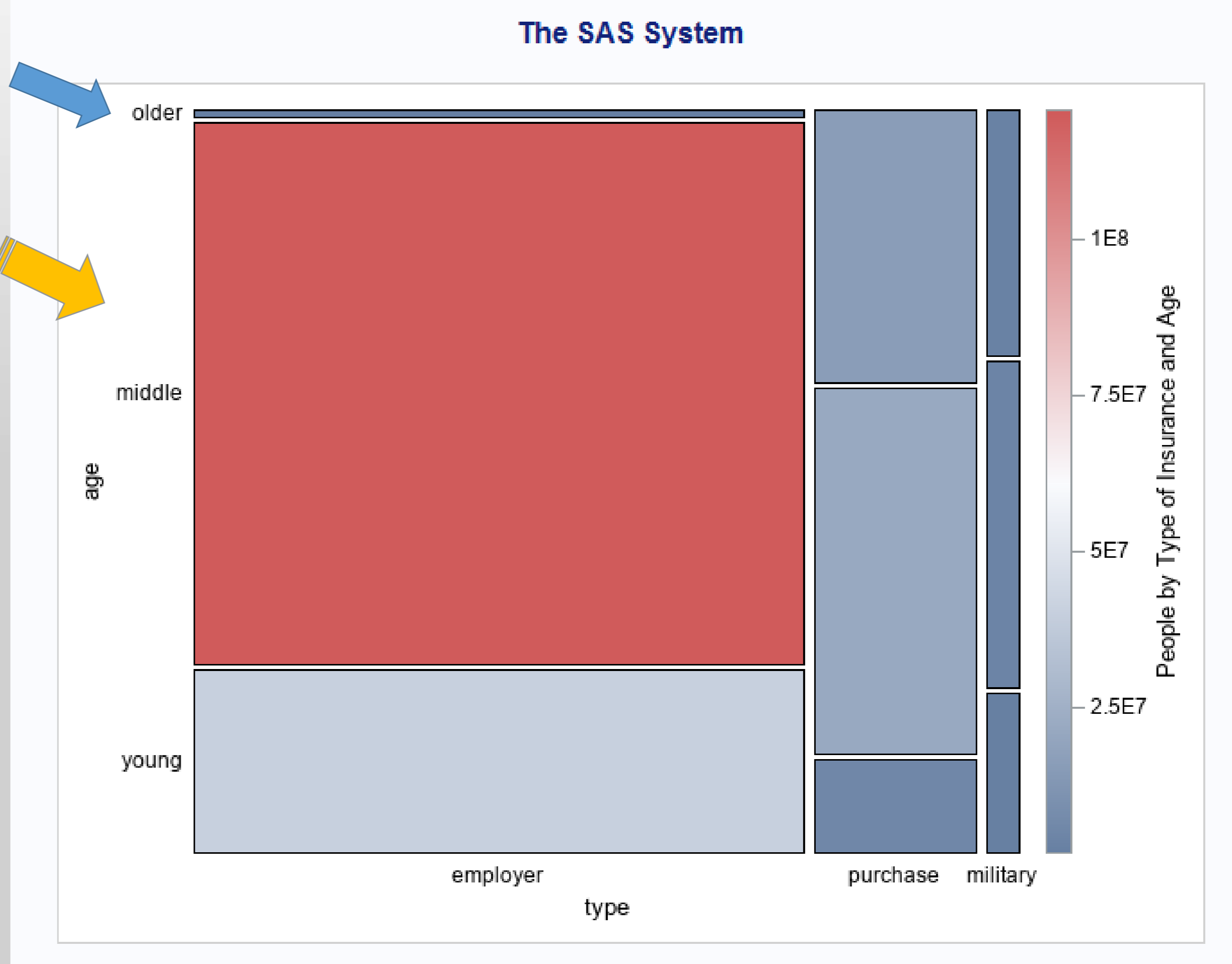
## SAS syntax to invoke and define mosaicplotparm template

```
proc template;
define statgraph mosaicPlotParm;
begingraph;
layout region;
mosaicPlotParm category =(type age)count=number/
name = "mosaic" colorresponse=number;
continuouslegend "mosaic"/title = "People by Type of
Insurance and Age";
endlayout;
endgraph;
end;
run;

proc sgrender data = insurance template = mosaicplotparm;
run;
```

ACS Variable	SAS variable	Levels of variable
Type of health insurance coverage	Type	<ul style="list-style-type: none"> <li>Employer-provided</li> <li>Direct purchase by consumer</li> <li>Military/Tri-Care</li> </ul>
Age (in years) of respondent; ACS defined levels utilized	Age	<ul style="list-style-type: none"> <li>Young (&lt;19)</li> <li>Middle (19-64)</li> <li>Older (≥65)</li> </ul>

## Mosaic plot output



From the mosaic plot output above, we can see the relationship between the type of insurance coverage and people’s age group. The yellow arrow is pointing to the largest red color tile; which represents the middle group for age (19-64 years) with employer-provided health insurance. The blue arrow is pointing to the smallest color tile, which represents older respondents (≥ 65 years) with employer-provided health insurance.

This graph tells us that the majority of all survey respondents between the ages of birth and 64 years reported having employer-provided health coverage. Consequently, for people in this group who lose their job, this means they will also lose their health insurance provider.

- Abstract
- Introduction
- Methods
- Results 1
- Results 2
- Conclusion

Please use the headings above to navigate through the different sections of the poster

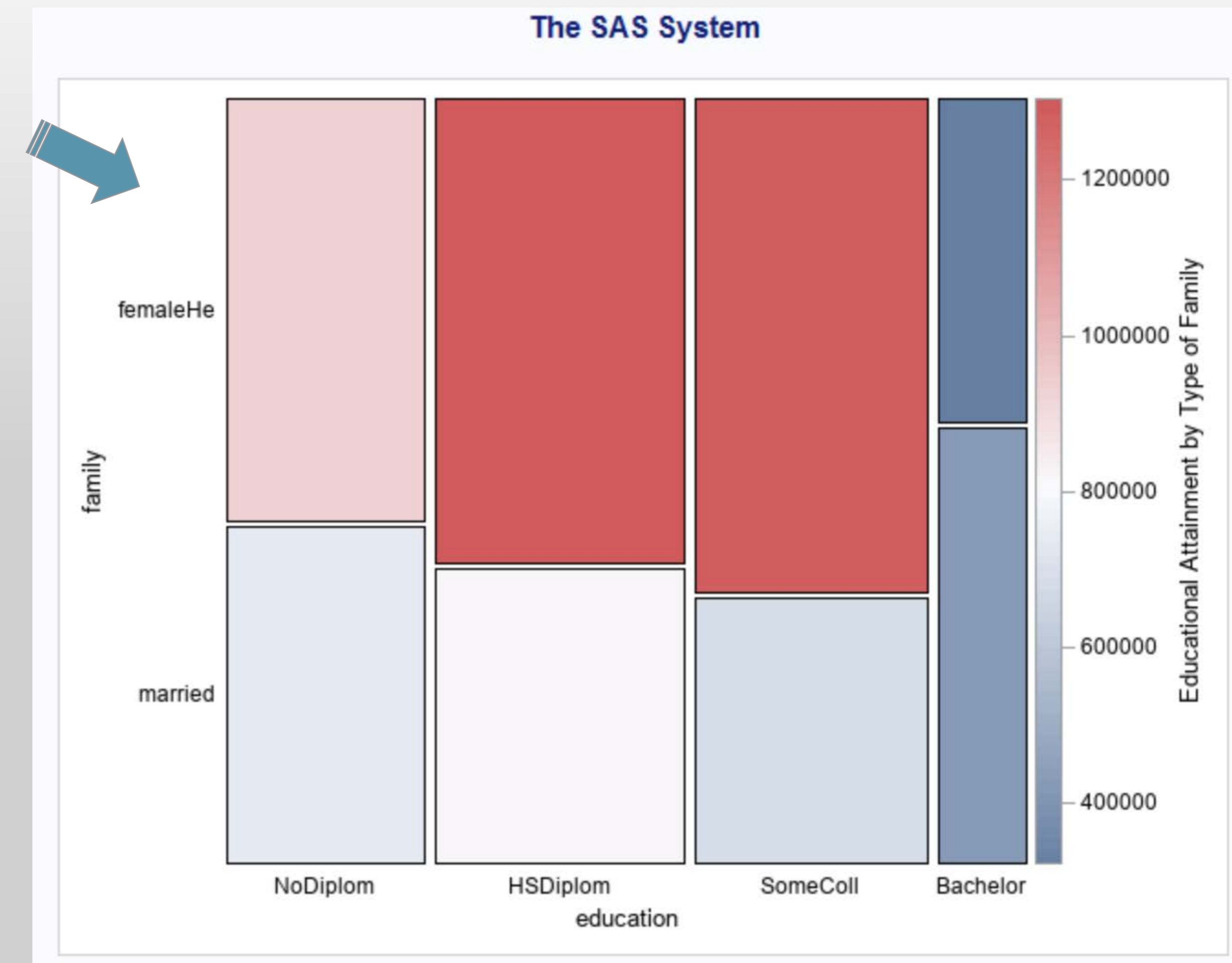
## Poverty level and Family status

### American Community Survey question: Poverty Status in the Past 12 Months of Families

Completing one's education is more difficult when one has a family. This data looks at the highest level of education reported, and the level of poverty of two different family types: married couples and female heads of household.

ACS Variable	SAS variable	Levels of variable
Highest reported educational level attained	level	<ul style="list-style-type: none"> <li>• No diploma</li> <li>• HS diploma</li> <li>• Some College</li> <li>• Bachelors or higher</li> </ul>
Type of family	type	<ul style="list-style-type: none"> <li>• Married couples</li> <li>• Female head of household</li> </ul>
ACS definitions utilized		

## Mosaic plot output



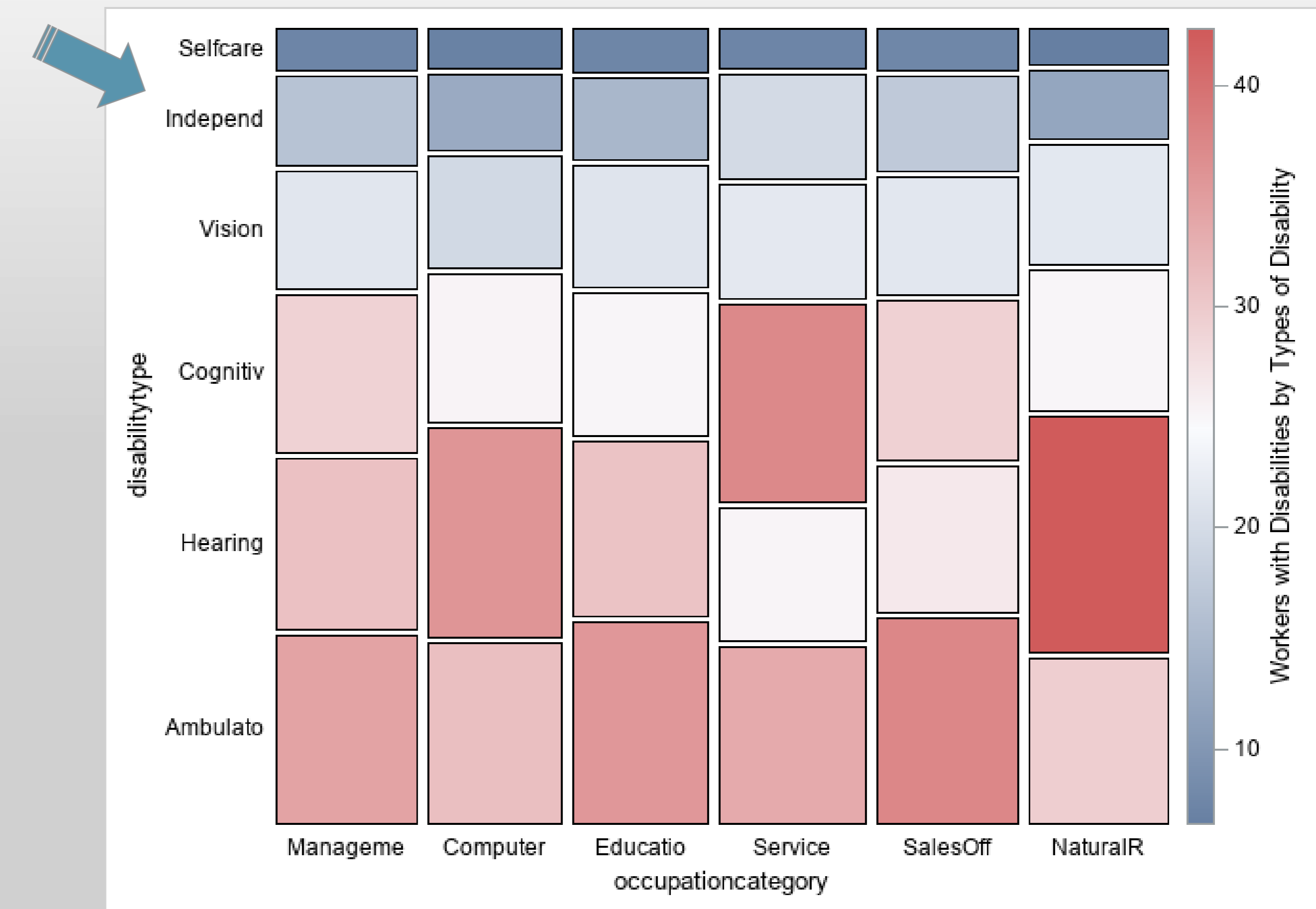
For this mosaic plot, we can see that for survey respondents that live below the poverty line, women with children (female head of household) were likely to have their high school diploma, and likely to have some college experience. Respondents with a bachelor's degree were the least likely to live below the poverty line (both married couples and female head of household).

It is also notable that single mothers with a bachelor's degree were the least likely to live below the poverty line. Those most likely to live below the poverty line were women that attended some college or had an associates degree. In order to live above the poverty line, people should obtain a bachelor's degree.

## Mosaic plot output

### Type of disability and occupational industry reported by ACS respondents

ACS Variable	SAS variable	Levels of variable
Occupational industry reported	Occupation	<ul style="list-style-type: none"> <li>• Computer engineering science</li> <li>• Education/legal/community services/arts/media</li> <li>• Management/business</li> <li>• Natural resources/construction</li> <li>• Sales office</li> <li>• Service</li> </ul>
Disability type ACS defined levels utilized	Type	<ul style="list-style-type: none"> <li>• Ambulatory</li> <li>• Cognitive</li> <li>• Hearing</li> <li>• Independent living</li> <li>• Selfcare</li> <li>• Vision</li> </ul>



From the mosaic plot output above, we can see the relationship between the type of disability, and the occupation category reported by the survey respondents. For example, within the natural resources and construction occupations, the predominant reported disability was hearing.

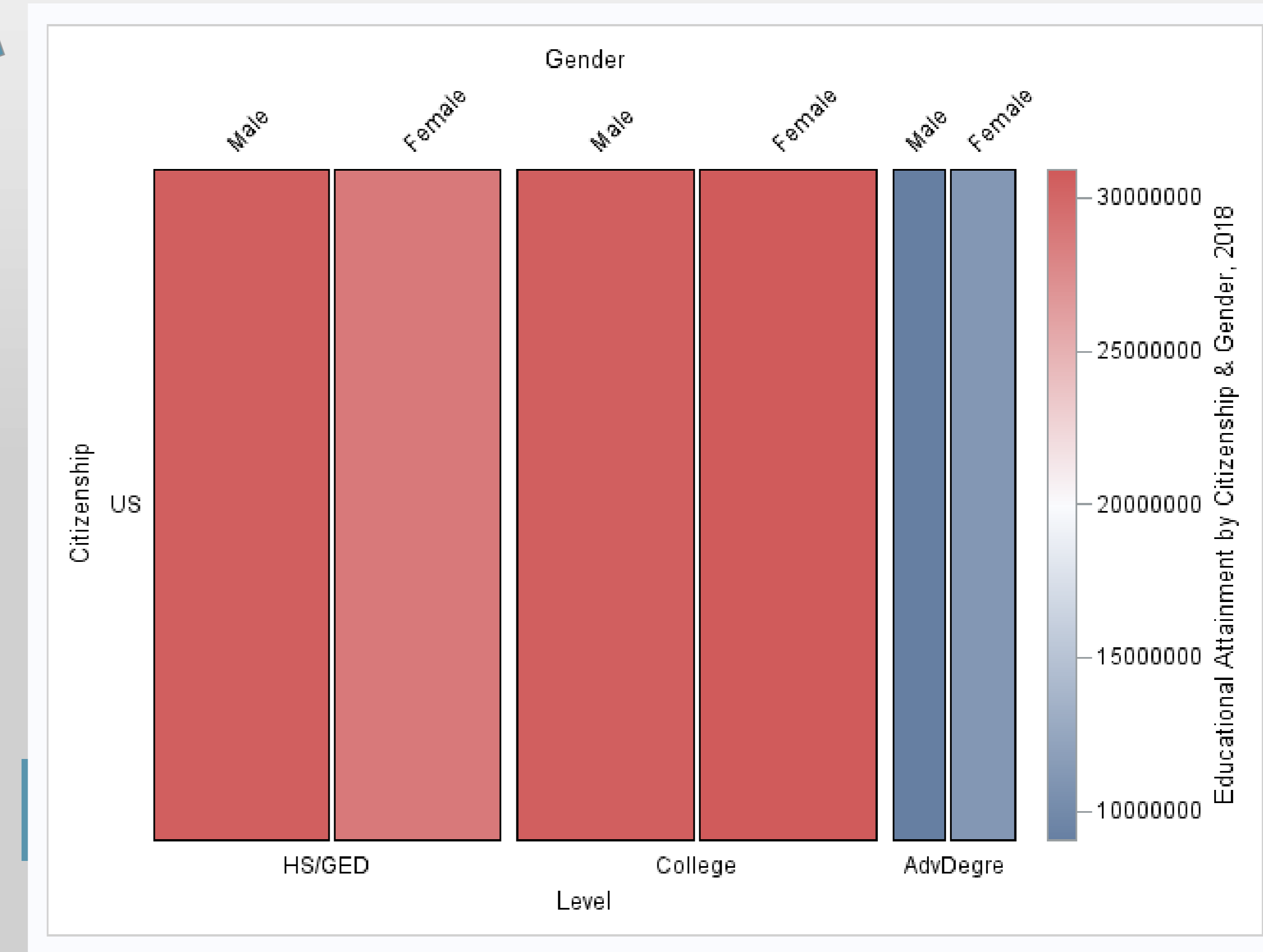
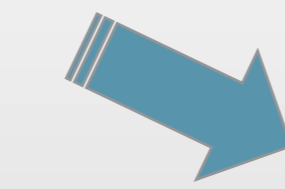
- Abstract
- Introduction
- Methods
- Results 1
- Results 2**
- Conclusion

Please use the headings above to navigate through the different sections of the poster

## Mosaic plot output

### Educational attainment by gender and United States citizenship status as reported by ACS respondents

ACS Variable	SAS variable	Levels of variable
Level of educational attainment	Level	<ul style="list-style-type: none"> <li>High school diploma-GED</li> <li>College degree</li> <li>Advanced college degree</li> </ul>
Gender: ACS defined levels utilized	Gender	<ul style="list-style-type: none"> <li>Male</li> <li>Female</li> </ul>
U.S. citizenship status: ACS defined levels utilized	Citizenship	<ul style="list-style-type: none"> <li>U.S. citizen born in the U.S.</li> <li>U.S. citizen, born abroad to U.S. citizen parents</li> <li>Non-U.S. citizen</li> <li>U.S. territories</li> <li>Naturalized citizen</li> </ul>



From the mosaic plot output above, we can see the relationship between the level of educational attainment, gender, and the citizenship status as reported by the survey respondents. This graph shows the relationship between gender and educational attainment for United States citizens. On the left, we can see that females are less likely to graduate from high school than males. On the far right, we see more females hold advanced college degrees, as compared to males.

- Abstract
- Introduction
- Methods
- Results 1
- Results 2**
- Conclusion

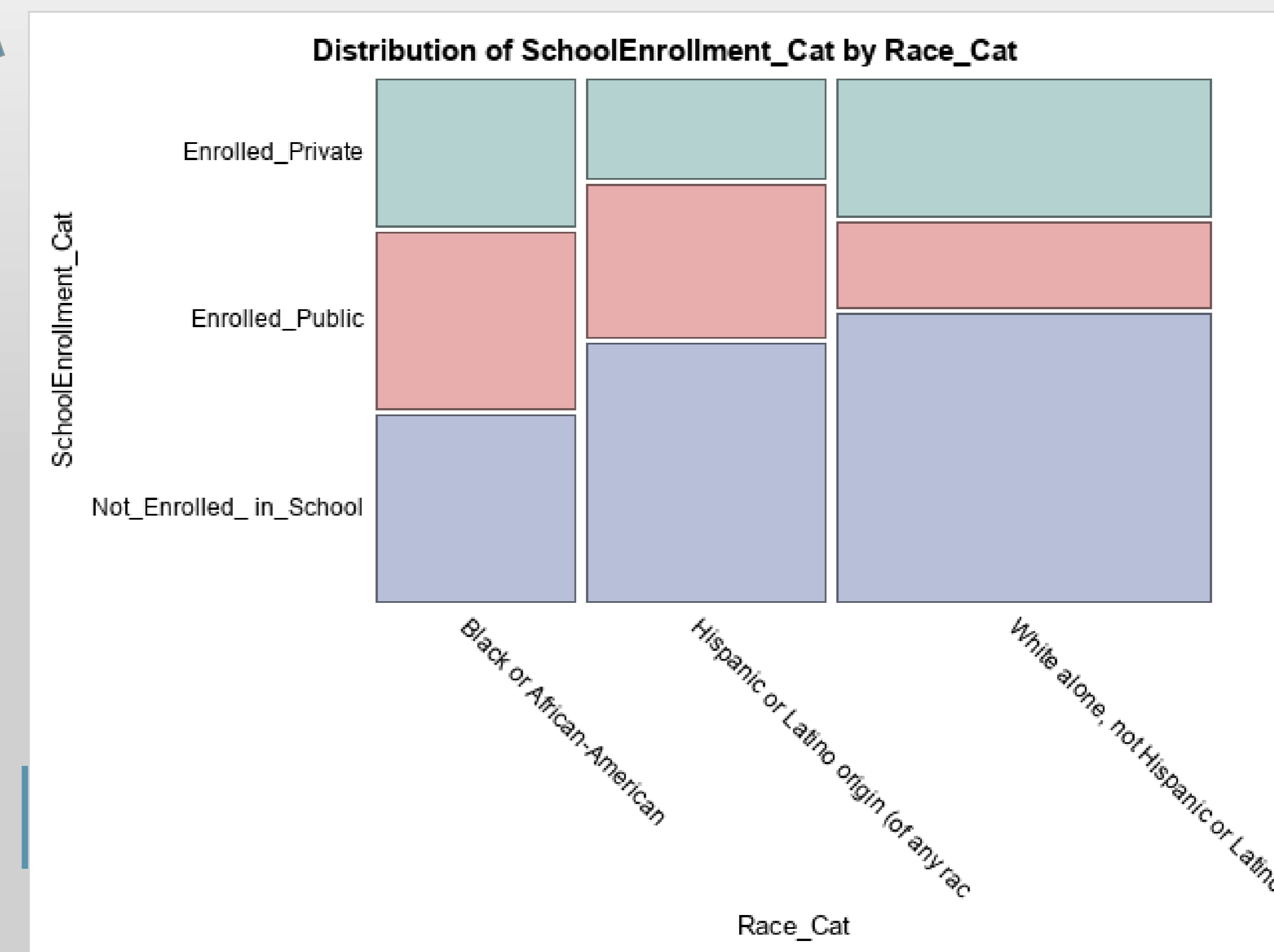
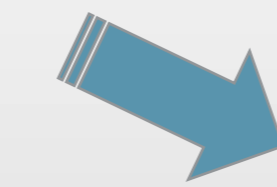
Please use the headings above to navigate through the different sections of the poster



## Mosaic plot output

### School enrollment distribution by race category: Characteristics of teenagers 15-19 years old

ACS Variable	SAS variable	Levels of variable
Level of school enrollment	Level	<ul style="list-style-type: none"> <li>Enrolled in private school</li> <li>Enrolled in public school</li> <li>Not enrolled in school</li> </ul>
Race category: ACS defined levels utilized	Race	<ul style="list-style-type: none"> <li>Black or African-American</li> <li>Hispanic or Latino origin</li> <li>White alone, not Hispanic or Latino</li> </ul>



At the intersection of School Enrollment and Race in the American Community Survey (ACS), respondent groups report the varying reality of their experiences. For youth ages 15-19, with respect to being "Not Enrolled in School (either in Private or Public)", those who self-identified as "White alone (not Hispanic or Latino)" showed a higher response level versus their "Black or African-American" or "Hispanic or Latino origin (of any race)" counterparts. Comparatively, those who self-identified as "Black or African-American" show a higher response level about being "Enrolled in School (Public)".

- Abstract
- Introduction
- Methods
- Results 1
- Results 2**
- Conclusion

Please use the headings above to navigate through the different sections of the poster

## References

- Friendly, M. (1994). Mosaic displays for multi-way contingency tables. *Journal of the American Statistical Association*, 89, 190-200.
- Friendly, M. (2000). *Visualizing categorical data*. Cary, NC: SAS Publishing.
- Friendly, M. (2002). A Brief History of the Mosaic Display. *Journal of Computational and Graphical Statistics*, 11(1), 89. <https://www-jstor-org.ezproxy.lib.usf.edu/stable/1391129>
- Grömping, U. (2014). Mosaic Plots are Useful for Visualizing Low-Order Projections of Factorial Designs. *American Statistician*, 2, 108. <https://doi.org/10.1080/00031305.2014.896829>
- Hartigan, J. A. & Kleiner, B. (1981). Mosaics for contingency tables. *Computer Science and Statistics: Proceedings of the 13<sup>th</sup> Symposium on the Interface*, NY: Springer-Verlag, 268-273.
- Hartigan, J. A. & Kleiner, B. (1984). A Mosaic of Television Ratings. *The American Statistician*, 38(1), 32. <https://doi-org.ezproxy.lib.usf.edu/10.2307/2683556>
- SAS Institute. (2016). *SAS® 9.4 Graph Template Language: Reference, Fifth Edition*. SAS Institute, Cary: NC.
- Theus, M. (2012). Mosaic plots. *WIREs Computational Statistics*, 4, 191-198. <https://doi.org/10.1002/wics.1192>
- United States Census Bureau (2018). *American Community Survey*. <https://www2.census.gov/programs-surveys/acs/methodology/questionnaires/2018/quest18.pdf>
- United States Census Bureau (2019). *American Community Survey Information Guide*. <https://www.census.gov/programs-surveys/acs/about/information-guide.html>
- Ünlü, A., Sargin, A., & Viehrig, M. (2013). Mosaics for Visualizing Knowledge Structures. *Journal of Data Science*, 11, 451–478.

Abstract  
Introduction  
Methods  
Results 1  
Results 2

Conclusion

Please use the headings above to navigate through the different sections of the poster

## EME 6346 // DATA VISUALIZATION



Samantha



Yaël-Alexandra



Umi



Victoria



Adel



Chantal

Abstract  
Introduction  
Methods  
Results 1  
Results 2

Conclusion

Please use the headings above to navigate through the different sections of the poster

The background of the banner features a scenic view of the Washington Monument at dusk, reflected in the water of the Tidal Basin. The sky is a mix of blue, purple, and orange. In the foreground, there are cherry blossom trees with pink and white flowers, and a stone walkway. A dark blue rectangular box is centered over the image, containing the event title in white and teal text.

# SAS<sup>®</sup> GLOBAL FORUM 2020

USERS PROGRAM

MARCH 29 - APRIL 1 | WASHINGTON, DC | #SASGF

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. <sup>®</sup> indicates USA registration. Other brand and product names are trademarks of their respective companies.