Visualizing the Census and American Community Survey: An Educational Module for Undergraduates

Andrew C. Thomas, Samuel L. Ventura, Mikhail Popov, Rebecca Nugent

Department of Statistics, Carnegie Mellon University

Statistical Graphics & Visualization

Undergraduate course offered yearly by the Department of Statistics

Taken by 50-70 majors in statistics, information science, computer science, and others

Teaches how to identify and visualize the structure and features of low and high dimensional data

Uses the programming language R

Visualization Project

Visualize the distribution of population characteristics in the metropolitan area(s) of interest.

Two guiding questions:

1. How does the population distribution compare over time, from 2000 to 2010, particularly by race?

2. How do characteristics like age, income, and household size correlate to population distribution and race in 2010?

Project Structure and Specifications

Students were split into **12 groups** of 5 students

Each group was **assigned 1-2 cities** to analyze at **tract and block group** levels

Students used **UScensus2000** and **UScensus2010** R packages

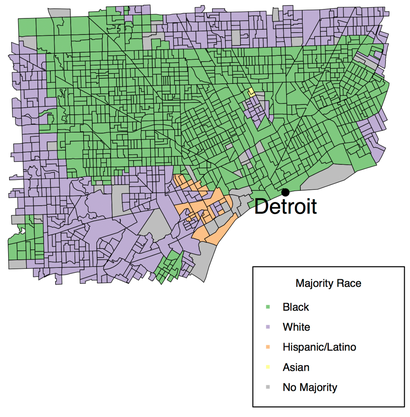
Some students incorporated **2011 American Community Survey** and, in the case of Chicago, crime and arrests data

Students had to **create 8-10 graphics** (with explanations) and display them on a **static website** or a **dynamic Shiny application**

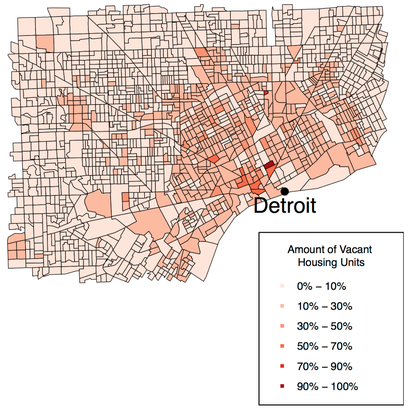
Mikhail gave a presentation on how to make Shiny applications in R (see poster), which inspired some students to develop Shiny applications

Selected Student Submissions

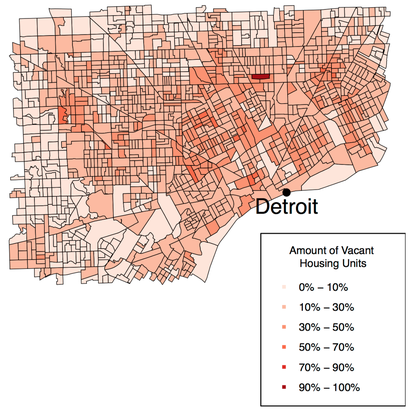
Race Distribution – 2010



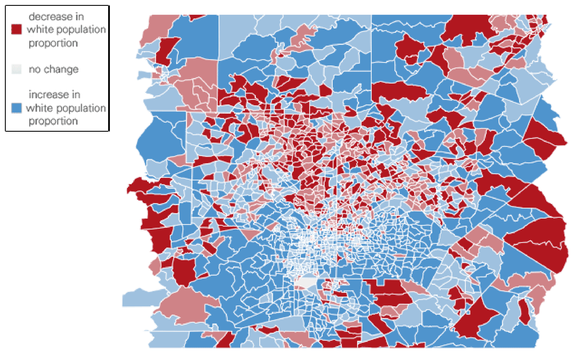
Proportion of Vacant Houses in Detroit in 2000



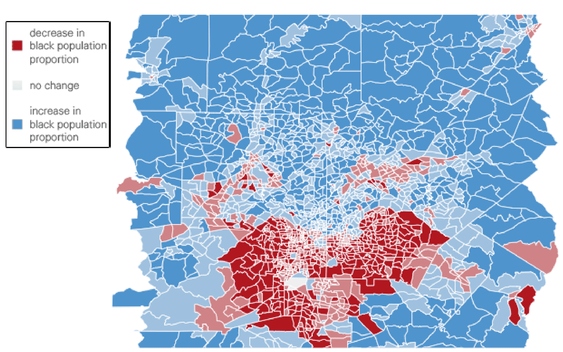
Proportion of Vacant Houses in Detroit in 2010



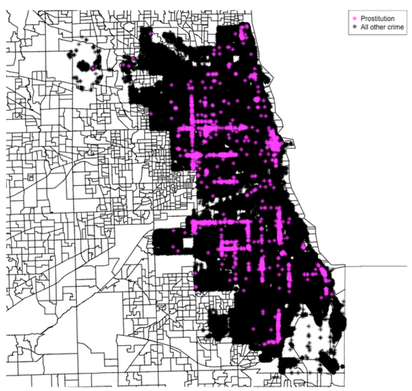
Change in Proportion of Whites in Atlanta



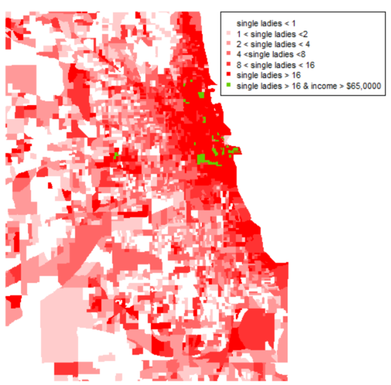
Change in Proportion of Blacks in Atlanta



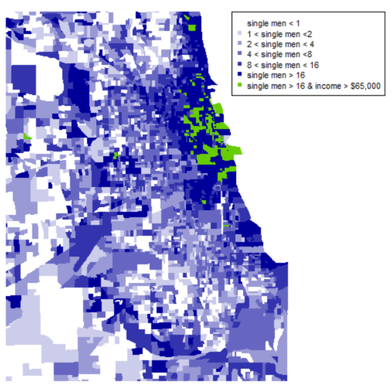
Prostitution Arrests in Chicago, IL



All The Single Ladies in Chicago, IL



Single Men in Chicago, IL



Student Takeaways

Students learned how to identify temporal and geographical differences in the US population with choropleth maps using Census/ACS data

Students practiced combining data from alternative sources (e.g. crimes in Chicago) with Census data to generate interesting hypotheses and come to insightful conclusions

Lessons Learned

Students who tried to tell a story about their city generally performed better and connected with the data more closely

e.g. children and families in Philadelphia, hispanics in Miami, population migration in Atlanta

Changes in the US Census from 2000 to 2010 (blocks, tracts, available data fields, etc) made it more difficult for students to create advanced graphics or test advanced hypotheses

Recommendations

The group project format worked very well in this setting. Students were able to pool their skills and ideas to create great statistical graphics.

It was worthwhile to spend class time to explain how the Census works (e.g. explaining the difference between blocks, tracts, counties, etc). This is especially true for classes with a high number of foreign students.

Personnel

Rebecca Nugent, Associate Teaching Professor
www.stat.cmu.edu/∼rnugent/

Mikhail Popov, Graduate Student (Master's)
www.mpopov.com

Andrew C. Thomas, Research Statistician
www.acthomas.ca

Samuel L. Ventura, Graduate Student (PhD)
www.stat.cmu.edu/∼sventura/

Acknowledgments

NSF # SES-1130706