**Reallocating ACS Data from Census Geographies to Neighborhood Boundaries.**

**ETHICAL CONDISERATIONS AND REGULATORY: HUMAN SUBJECTS IN CLINICAL RESEARCH**

**A THESIS IN**

**Bioinformatics**

Presented to the faculty of the University of Missouri-

Kansas City in partial fulfilment of the

requirements for the degree.

MASTER OF SCIENCE

by

SATYASREE GOLLA

M.S., University of Missouri-Kansas City, 2021

Kansas City, Missouri

2021

**Introduction**

Proposed solution=re-allocating information published at the Bureau of the Census and aggregated to census block groups to neighborhood and communities instead. Weighted average approach based on the number of people or the number of housing units.

**Literature review**

The United States Census Bureau, officially the Bureau of the Census, is a principal agency of the U.S. Federal Statistical System, responsible for producing data about the American people and economy. The Census Bureau mission is to serve as the nation’s leading provider of quality data about its people and economy. It runs under Title 13 and Title 26 of the U.S. Code. The Census Bureau collects the information about American Community Survey (ACS), Census of Governments, Decennial Census of Population and Housing, and Economic Census. It deals with different sectors such as Business and economy, Education, Health and Research, Income and poverty, Population and Housing, Employment, International trade in the nation. It includes photos, publications, videos, fact sheets, blogs and working papers related to data. It includes different surveys and programs like American Community Survey, American Housing survey, American business survey.

The American Community Survey (ACS) is a demographics survey program conducted by the U.S. Census Bureau. It helps local officials, community leaders, and businesses understand the changes taking place in their communities. It is the premier source for detailed population and housing information about our nation. It is a yearly basis survey and gives information about how federal and state funds are distributed each year. It holds different tables like Selected social characteristics, selected economic characteristics, selected housing characteristics, Demographic and Housing estimates. It regularly gathers information previously held only in the long form of the decennial census.

Every 10 years, the U.S. Census Bureau conducts a census to figure out the number of people living in the United States. The data collected by the decennial census are used to apportion the number of seats each state has in the U.S. House of representatives. The data collected by ACS is used by many public-sector, private-sector and not-for-profit stakeholders to distribute funding, track shifting demographics, plan for emergencies, and learn about local communities.

Census tracts are small, relatively permanent statistical subdivisions of a county. It has 2,500 to 8,000 people. Block Groups (BGs) are statistical divisions of census tracts, are generally defined to contain between 600 and 3,000 people and are used to present data and control block numbering. Statistical areas bounded by visible features such as roads, streams, and railroad tracks, and by nonvisible boundaries such as property lines, city, township, school district, county limits and short line-of-sight extensions of roads.

Various neighborhood and community organizations would like to use ACS data to understand information that can help improve the places where they live. [Elaborate on neighborhood and community boundaries not coinciding with census tracts and block groups].

**Somewhere in this section, put background about TIGER files.**

**Methods**

The Census provides two sources of data, the shapefiles for various geographic regions and data on these regions collected through the ACS.

**TIGER and Shapefiles:**

The TIGER/Line Shapefiles are the fully supported, core geographic product from the U.S. Census Bureau. They are extracts of selected geographic and cartographic information from the U.S. Census Bureau’s Master.

Geographic regions are defined as shapefiles. A shapefile is a geospatial data format for use in geographic information system (GIS) software. Shapefiles spatially describe vector data such as points, lines, and polygons, representing, for instance, landmarks, roads, and lakes. The Environmental Systems Research Institute (Esri) created the format for use in their software, but the shapefile format works in additional Geographic Information System software as well.

**Census shapefiles:**

The shapefiles of Census Geographies are divided into individual states. Each state has its own FIPS code (20 – Kansas; 29 – Missouri). States are divided into counties/parishes which are then divided into Census Tracts. Census Tracts have 2500 to 8000 people. Census Tracts are divided into Block Groups. Block Groups have 600 to 3000 people. Census Block Groups are divided into Census Blocks. Census Blocks are the smallest geographical areas. In urban areas the Census Block can be a small as a single city Block but can be larger in other regions.

**Neighborhood and Community District shapefiles:**

In addition, we have shapefiles for the community districts and neighborhoods in the Kansas City metropolitan area. The two geographies, community district and neighborhood, have slightly different sources. The neighborhood geography was originally developed by the Kansas City Missouri planning department in the 1980's as part of the user defined geography initiative of US Census Department. These neighborhood boundaries, though contested at the margin in some places, were based on the pre-existing (pre-1980) social geography. The Wyandotte County neighborhoods have a similar history. The North Kansas City, Independence and Raytown neighborhoods were developed as part of the KC Health CORE initiative. Slight changes were made to the North Kansas City shapefiles to make them continuous (minimal gaps, no overlaps).

The Community District geography was developed as part of the Center for Economic Information's neighborhood and urban development work. Following the work of the Chicago School of Urban Sociology (Burgess, McKenzie, etc.), CEI recognized the need for a mid-level geography, between the neighborhood and city level, that was rooted in the living patterns of residents. The construction of the community district geography is discussed at length in Dr. Bowles dissertation and a slightly modified version of the community district geography is used by the KCMO planning department.

**American Community Survey:**

The data related to American Community Survey, programs and related articles are present in census bureau website. The information is derived from these publications. The mapping of subdivisions counties, tracts, block groups, pumas and census blocks are plotted using R studio by TIGER line/Shapefiles. Shapefile data from 2007 onward was used to develop Neighborhood and Community District boundaries. Boundaries derived prior to 2007 used TIGER lines.

Census Bureau also has all the data about the population, housing, sources for The ACS data can be found in data.census.gov website. The data includes population, number of housing units in the counties and other subdivisions.

**Describe the process of evaluating which census blocks are partially inside versus completely inside a community district or neighborhood boundary.**

Bibliography (References?)

U.S. Census Bureau. (2017). *U.S. Census Bureau at a Glance*. U.S Census Bureau. <https://www.census.gov/about/what/census-at-a-glance.html>

<https://www.census.gov/data/what-is-data-census-gov/about.html>

<https://www.census.gov/programs-surveys/acs>

<https://www.census.gov/programs-surveys/acs/about.html>

<https://www2.census.gov/geo/pdfs/reference/GARM/Ch8GARM.pdf>

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

Bowles, D "Geographic information systems-based analysis of metropolitan development, decline, and recovery" University of Missouri Kansas City, 2005. Dissertation in Economics and Social Science. [https://www.proquest.com/openview/6b1c099f66f2fe4d8364d6265131e882/1?pq-origsite=gscholar&cbl=18750&diss=y](https://nam02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.proquest.com%2Fopenview%2F6b1c099f66f2fe4d8364d6265131e882%2F1%3Fpq-origsite%3Dgscholar%26cbl%3D18750%26diss%3Dy&data=04%7C01%7Csg4cd%40mail.umkc.edu%7C1e9a933b8fc84d97b11608d9fd6b1a31%7Ce3fefdbef7e9401ba51a355e01b05a89%7C0%7C0%7C637819458708571574%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000&sdata=N7cr2fkmLvjVrmoSD7VTG%2FNTanqvOwrHg%2BgLmgHQtgM%3D&reserved=0)

Appendix A. R code

[List the program names]

Appendix B. Data dictionaries.