SEATTLE GENTRIFICATION DATASET

Data Nutrition Label

ABOUT

This dataset was created by Aileen Kuang and Victoria Zhuang to analyze and compare the progression of gentrification within two Seattle neighborhoods: Wallingford and the Chinatown-International District. In the past, gentrification has been linked with the racial demographics of a neighborhood; usually, neighborhoods with more people of color are more likely to face gentrification and displacement. This dataset, a combination of US Census data (collected by the US Census Bureau) and the Built Units since 2010 data set (from the City of Seattle GIS Program) aim to help identify whether such a relationship exists, and the severity of the impact on Wallingford and Chinatown. Both original data sets are public records provided by government websites. The merged dataset focuses on different built units from 2013-2019 -- including building type, location, permits -- and the corresponding demographics, split into subpopulations for race in that neighborhood. Hopefully, this dataset could also be saved for future reference and research into gentrification and displacement in Seattle.

USE CASES

- What is the relationship between built units and population change in the International District? In Wallingford?
- How do existing population demographics influence the rate at which new developments appear?
- How do new developments affect demographics?
- Can we predict future demographic trends based on current demographics and built units?
- What types of units are a sign of gentrification in specific areas?
- How does gentrification vary between different neighborhoods?
- Can we determine those changes based on racial demographics?

Nutrition Facts

Racial demographics and gentrification data set

Sample

Data Integrity

Bias	2
Inaccessibility	2
Generalization	3

- **Bias:** sampling bias through non-responses, creators' bias in connecting census and Built Units.
- **Ambiguity**:
- **Inaccessibility**: variables and observations from Built Units data set are difficult to understand, and not all variables from the original data set are present
- **Generalization**: aggregates Built Units data to the block group level, not all built units may be present in the data set, assuming population demographics based on US Census data

CHANGES MADE

- Columns from the Built Units Since 2010 data set were removed before merging with the US Census data set.
- Observations dated 2020-2023 from the Built Units data set were removed before merging with the US Census data set.
- Values in the GEOID column from the Built Units data set were edited -- previously, they represented census blocks, but they have been aggregated to the census block group level.
- New columns were added: percent of white and non-white people in each census block for each year from 2013-2019, total number of non-white people, and the percentage of the dominant racial demographic in each census block group for each year from 2013-2019.

ALERTS

- Units built does not indicate current residence
- Units built does not cover all housing units built within Seattle
- May contain selection bias in both Census and built units
- Merged dataset does not include all data from both built units and current residence, which may result in selective bias

DANGER

- Could be used to provide further guidance in gentrifying different neighborhoods in the future
- Could be used to target certain subpopulations of people
- Original Built Units dataset could potentially reveal address' of residents
- Certain columns are not easily understandable from the Built Units dataset if the reader is not professionally involved in those fields