



# Self Segregation caused by homophily preference

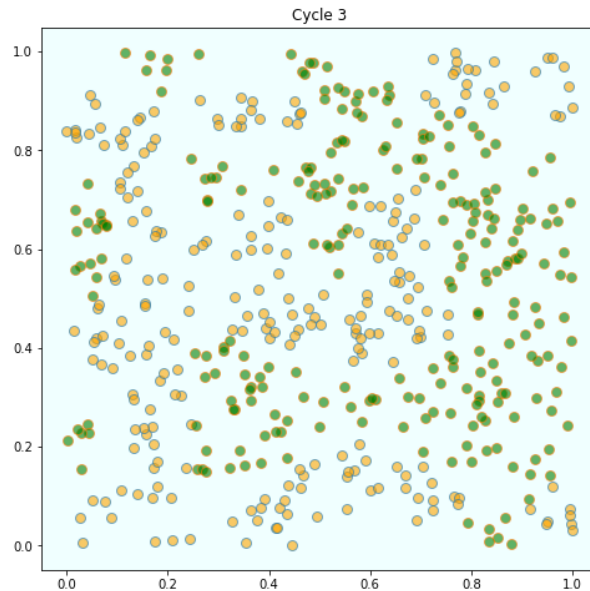
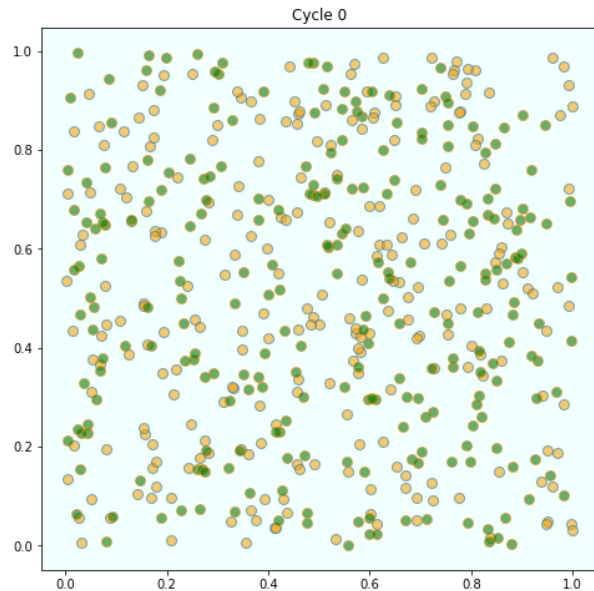
Research question: how large is the effect of people's homophily preference on population migration within a city

Zeyu Xu



# Motivation

Schelling's Segregation Model: People like to have similar neighbors



cited from <https://lectures.quantecon.org/py/schelling.html>

# • Theories of population migration •

There are many factors affecting people's decision on moving, they have different effect at different levels, here's some important ones.

## Income

- Almost a self-evident truth

## Climate

- Graves(1976)
- Environmental considerations are important in migration decision

## Public goods

- Tiebout(1956)
- Vote with their feet

## Homophily preference

- Schelling (1969)
- Micro homophily preference leads to macro segregation



# Model

---

- A block-level study: focus on the effect of income and homophily preference, ignore effects of all other variables.
- A Markov chain: calculate the probability of moving from one region to another based on the state at that period (income & population composition).





# Model

---

- **Probability calculation:**

$$p_{i,j,t} = \frac{1}{1 + e^{-r\Delta u_{i,j,t}}}$$

$$\Delta u_{i,j,t} = u_{j,t} - u_{i,t}$$

$$u_{i,t} = \beta_0 + \beta_1 \times income_{i,t} + \beta_2 \times homophily_{i,t}$$

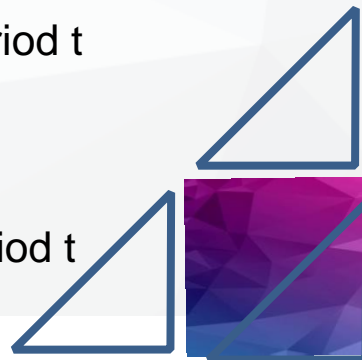
- **Variable interpretation:**

$u_{i,t}$  : The utility of living in block i at period t

$income_{i,t}$  : The average annual income of people living in block i at period t

$homophily_i$  : The similarity of people in block i to the agent at period t

$p_{i,j}$  : The probability that the agent moves from block i to block j at period t





# Model

- **Population mobility pattern:**

$$population_{i,t+1} = population_{i,t} + \sum_{j \neq i} p_{j,i,t} \times population_{j,t} - \sum_{k \neq i} p_{i,k,t} \times population_{i,t}$$

- **Loss Function:**

$$L_t = \frac{1}{n} \sum_i^n (population_{i,t} - \widehat{population_{i,t}})^2$$

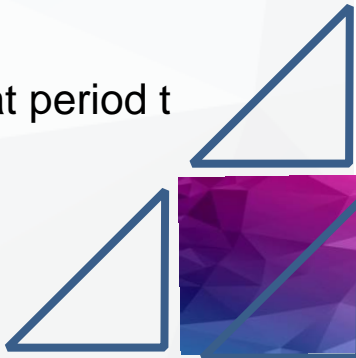
- **Variable interpretation:**

$population_{i,t}$  : The population (of specific ethnic group) of block i at period t

$\widehat{population_{i,t}}$  : The predicted population (of specific ethnic group) of block i at period t

$n$  : The number of blocks in Chicago

$L_t$  : The sum of squared error of prediction on population at period t





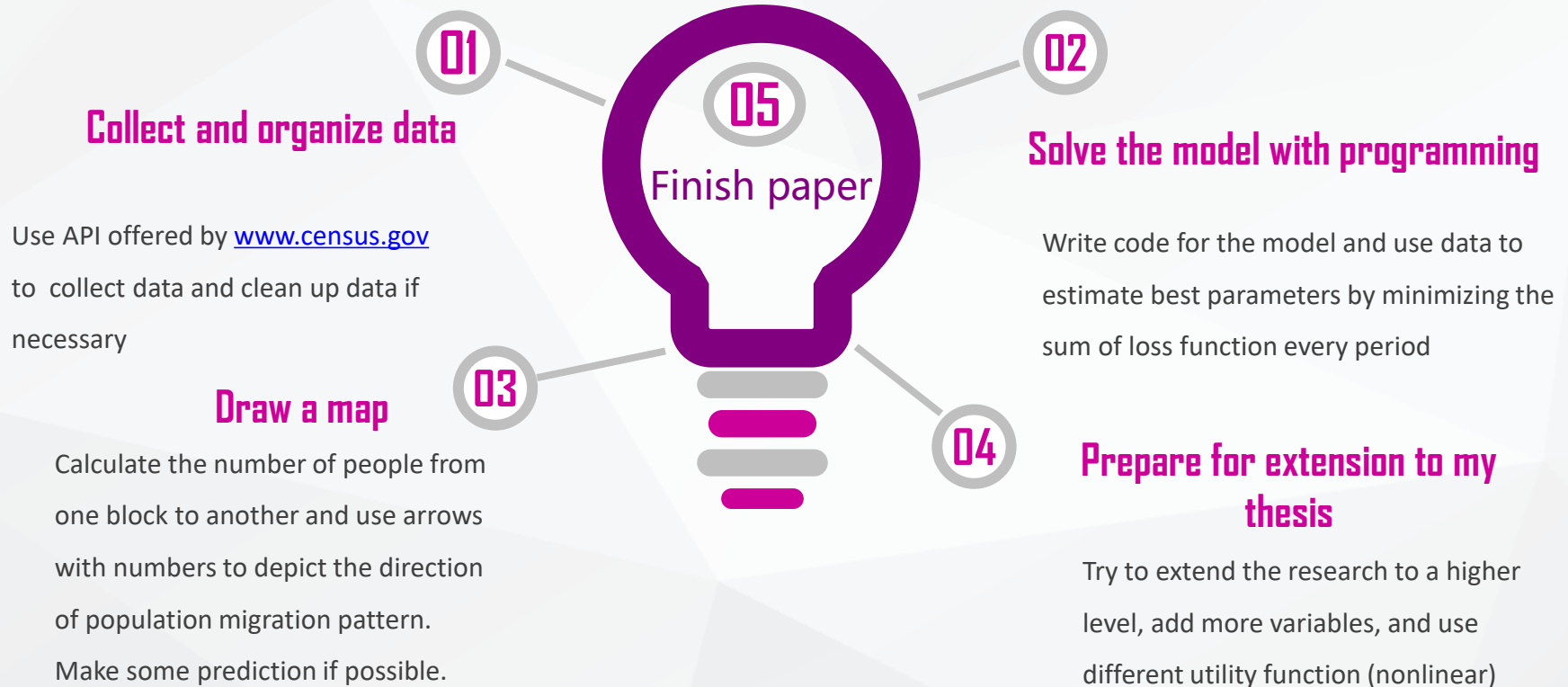
# Database

---

- Source: United States Census Bureau <https://www.census.gov/>
- Population: American Community Survey (5-year survey for year 2011-2013, 5-year estimates for year 2014-2017)
- Average annual income: American Community Survey (5-year survey for year 2011-2013, 5-year estimates for year 2014-2017)
- The data of year X comes from the survey or estimate from X-4 to X, for example, data of 2011 comes from 2007-2011 American Community Survey



# My research plan





# Limitation

- Focus only on block-level to make model simple, unable to compare the effect with other factors like climate, public goods and so on.
- About half data comes from the result of survey while the other half comes from estimates based on survey.
- Ignore people moving in and out the city to simplify the question.

# Extension

1

Use data of higher level to see if the effect becomes weaker

2

Add more variables to get more accurate estimation

3

Try to find more consistent data

4

Try to internalize income, so that we can simulate the migration with initial state

# Limitation

- Focus only on block-level to make model simple, unable to compare the effect with other factors like climate, public goods and so on.
- About half data comes from the result of survey while the other half comes from estimates based on survey.
- Ignore people moving in and out the city to simplify the question.

# Extension

1

Use data of higher level to see if the effect becomes weaker

2

Add more variables to get more accurate estimation

3

Try to find more consistent data

4

Try to internalize income, so that we can simulate the migration with initial state