

ECON 0150 | Economic Data Analysis

The economist's data analysis skillset.

Part 1.6 | Relationships In Space

Geographic Data

Some types of relationships in space

- *Geographic data is data organized on three axes (latitude, longitude, altitude)*
- *We typically only use latitude and longitude*
- *Geo data is often combined with other variables like population*
- *Two main types of geo data: points, shapes*
- *We sometimes observe points, but most data comes in groups*

Example: Restaurants by Zipcode

Are there fewer restaurants further from downtown Pittsburgh?

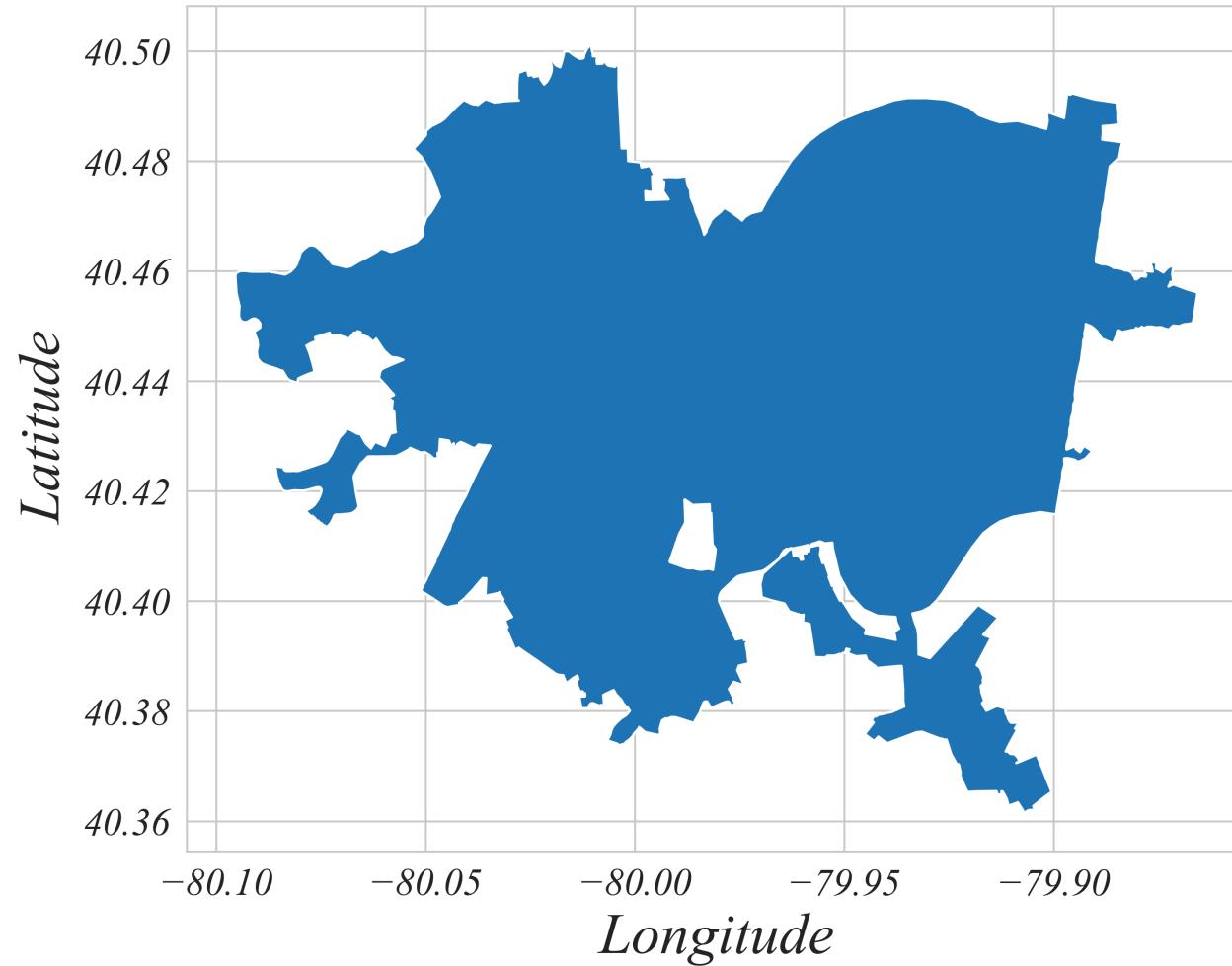
We're going to use a Census maps and openly available data on restaurant locations to answer this question.

- *Data: Census Shapefiles and Open Street Maps*

Geographic Data

Maps are (typically) plots on two axis

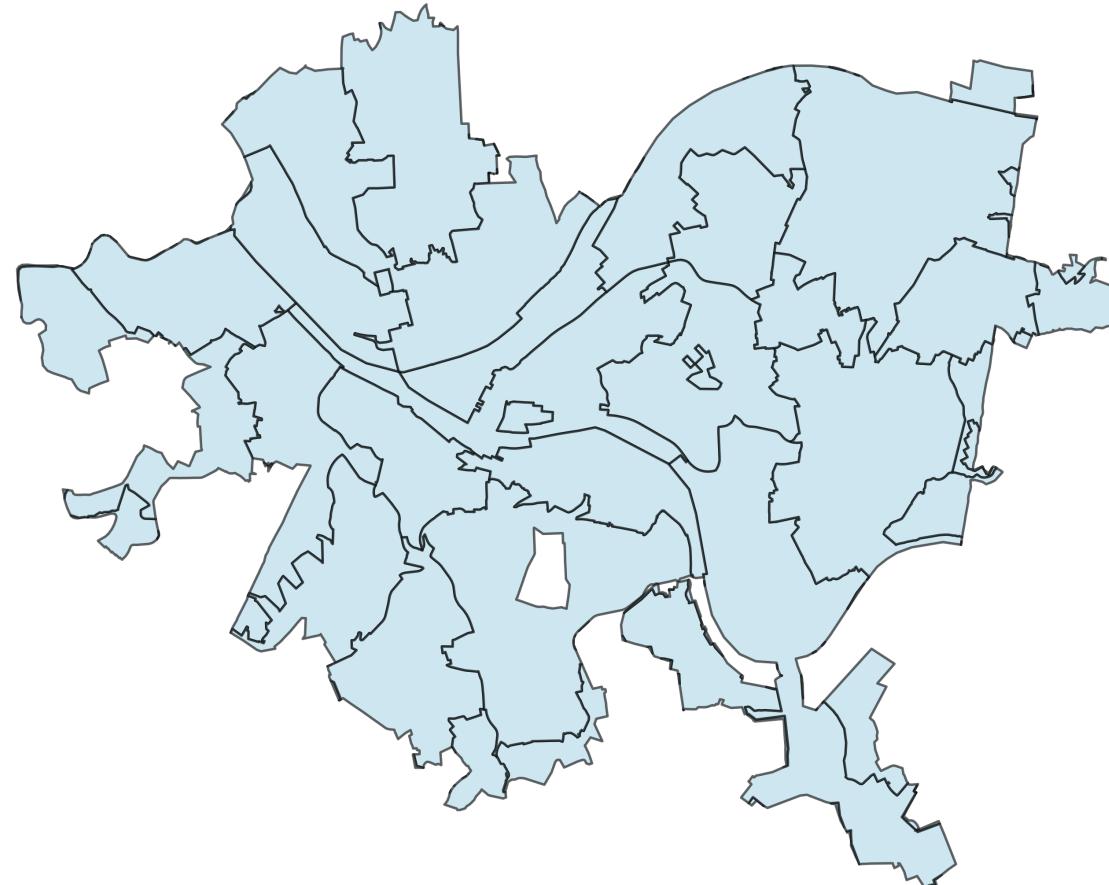
> a basic map of Pittsburgh



Geographic Data

Maps can show any level of detail available in the data

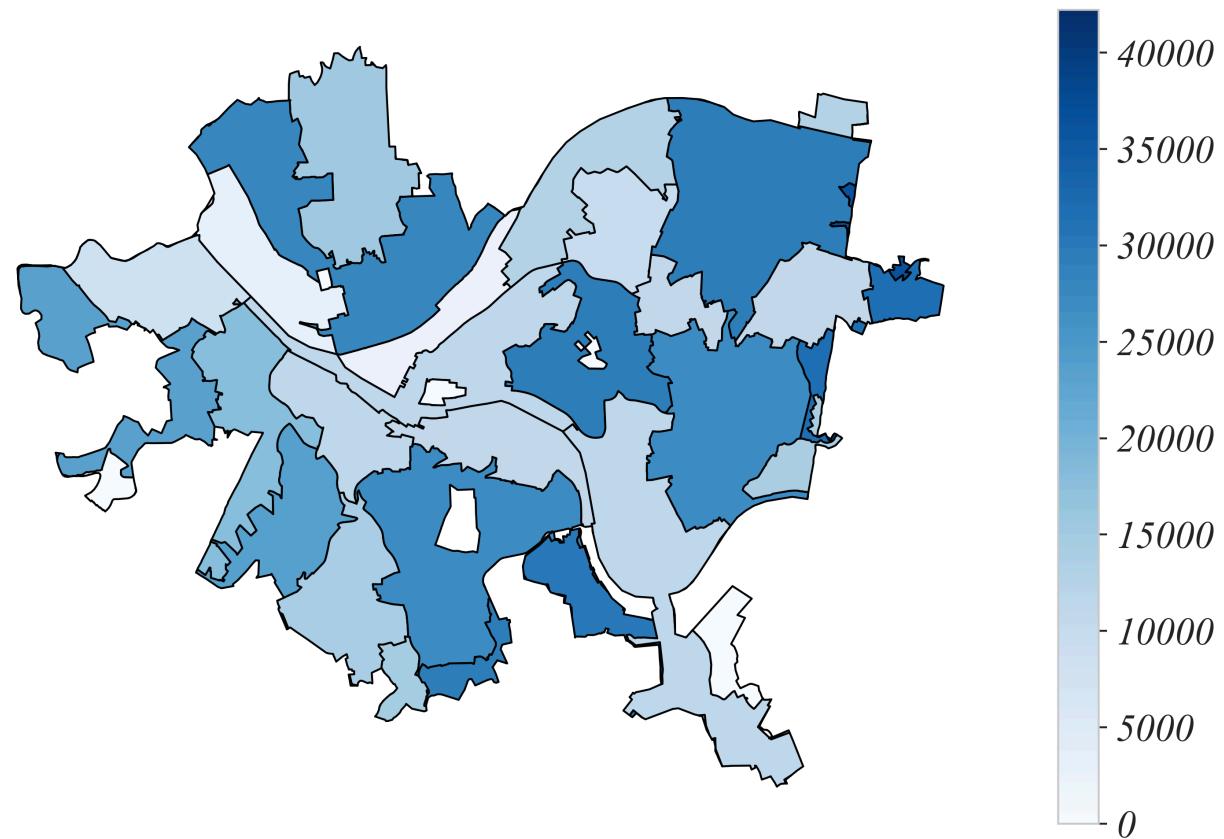
> *a map of Pittsburgh Zipcodes*



Geographic Data

We can add information: colors

> a map of Pittsburgh Zipcode populations



Geographic Data

We can add information: colors

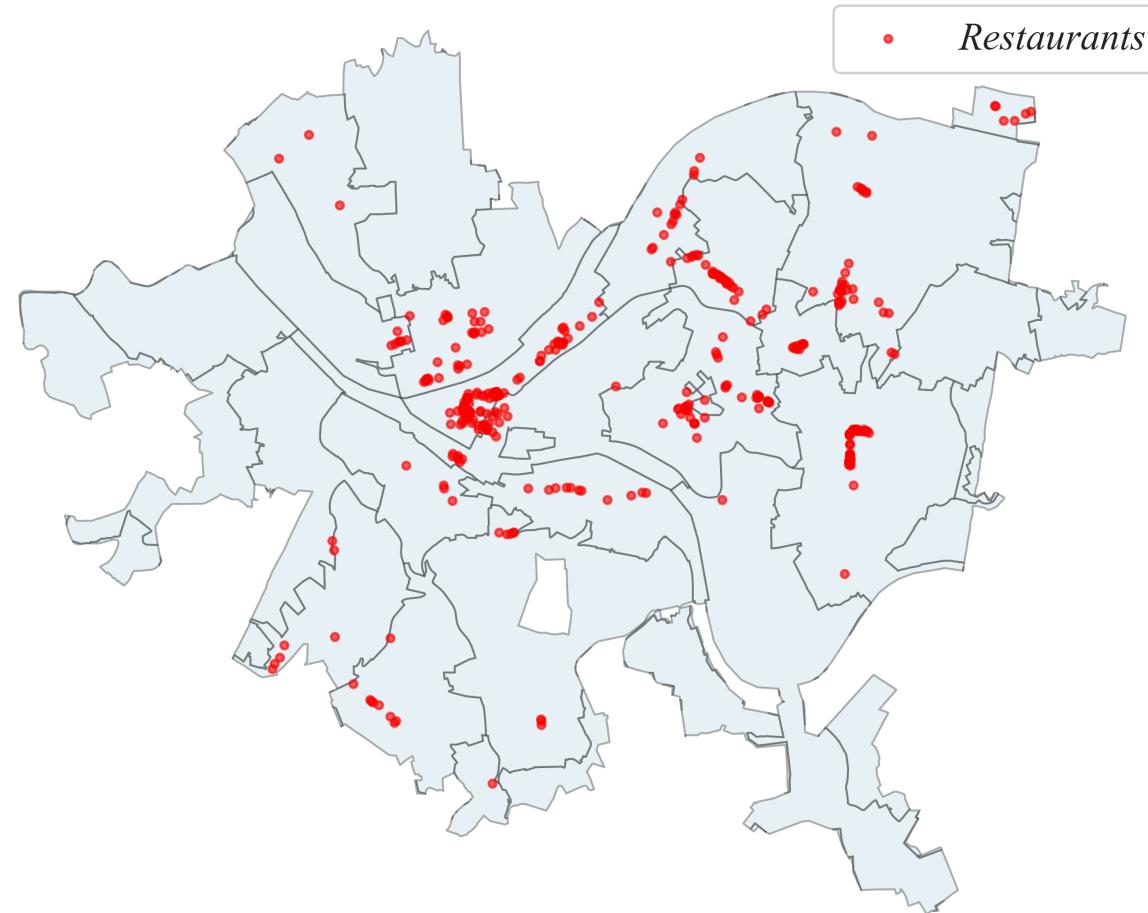
> *a map of Pittsburgh Zipcode populations: interactive!*



Geographic Data

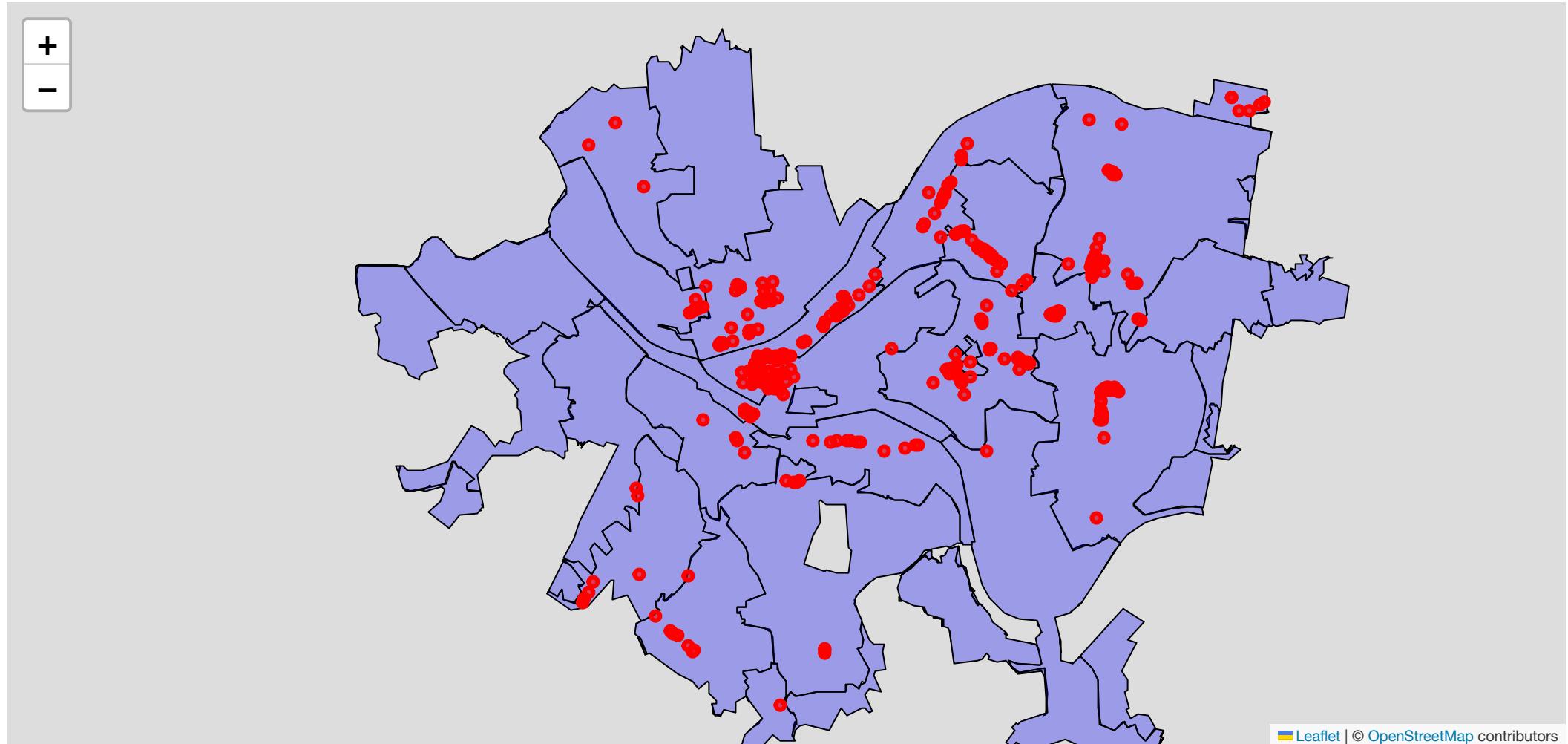
Maps can also show points

> *some restaurants in Pittsburgh!*



Geographic Data

Maps can also show points



Geographic Data

Maps can also show points

- *Points are nice but we typically can't use them raw*
- *Some point transformations: distances between points; group by area; etc*
- *We can also relate points to other variables (eg. zipcode population for each restaurant)*

Geographic Data Example: Nunn (2008)

Did the historical trade of enslaved people impact modern economic development in Africa?

Method: Uses historical data and the distance from major ports

Findings: Areas more disrupted by enslavement have lower GDP today, due to:

- *Weakened institutions (political fragmentation, mistrust).*
- *Disrupted societies (population loss, economic stagnation).*

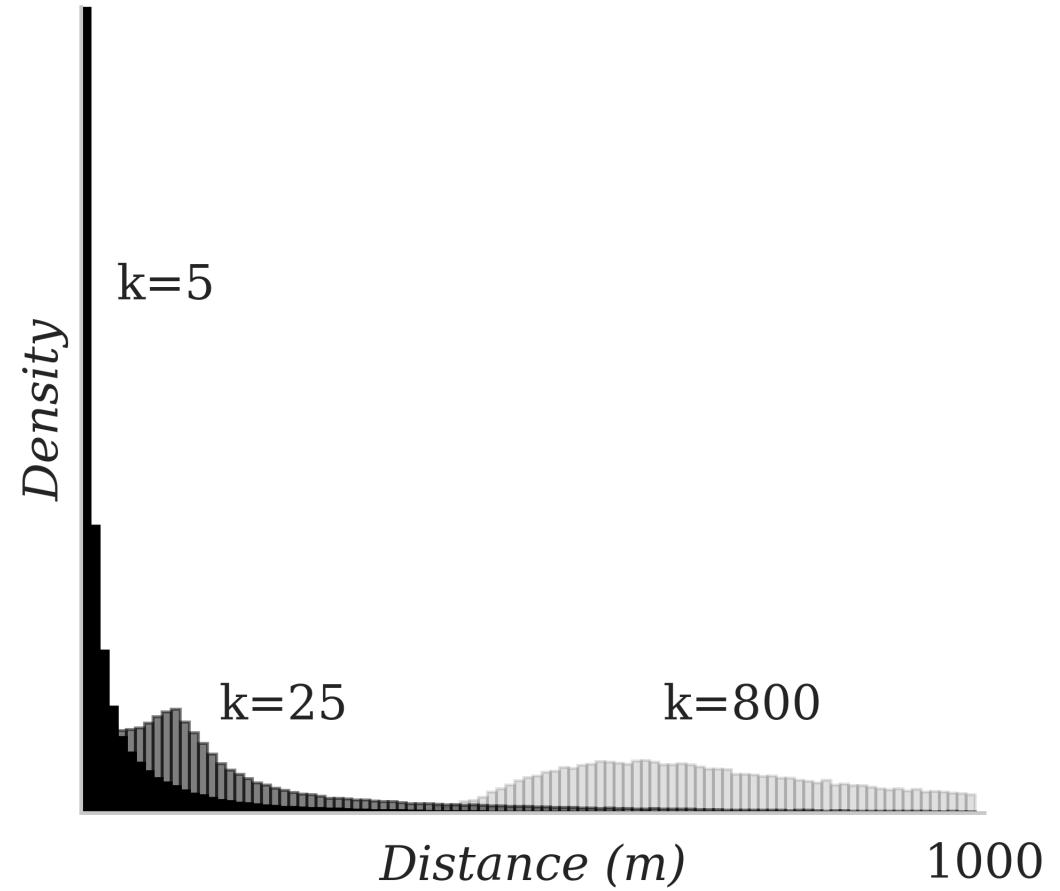
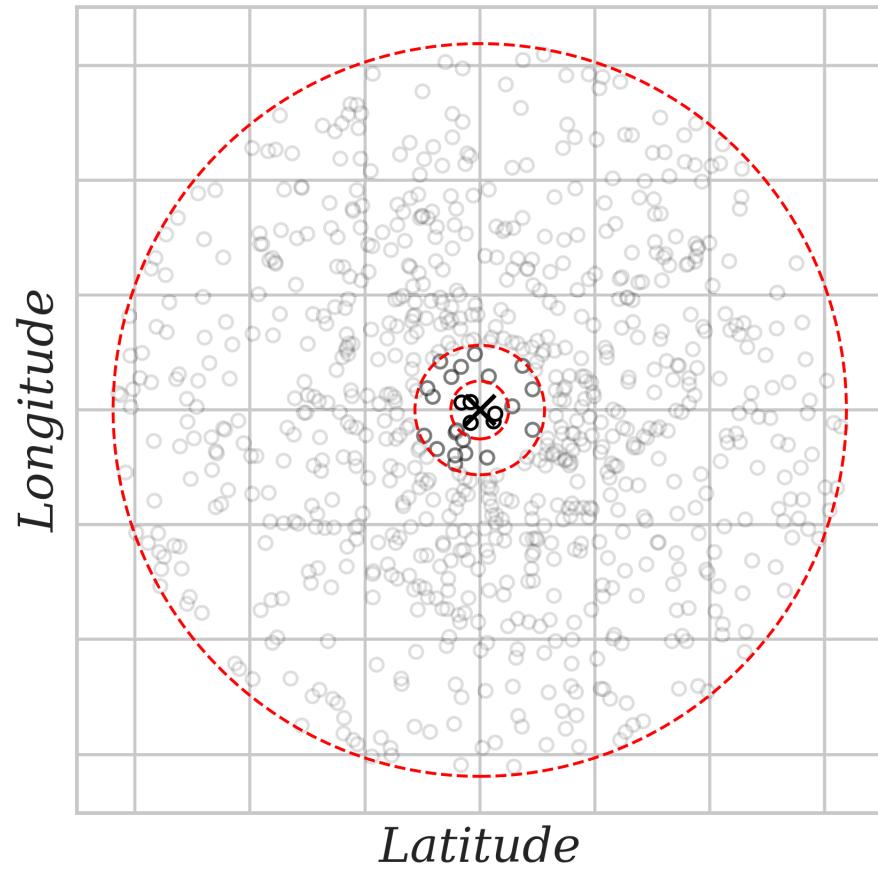
Implication: Historical shocks can have persistent economic effects.

Geographic Data Example: Weidman (2024)

Does the party of your neighbors impact your decision to vote?

Geographic Data Example: Weidman (2024)

My dissertation involved measuring distances between voters



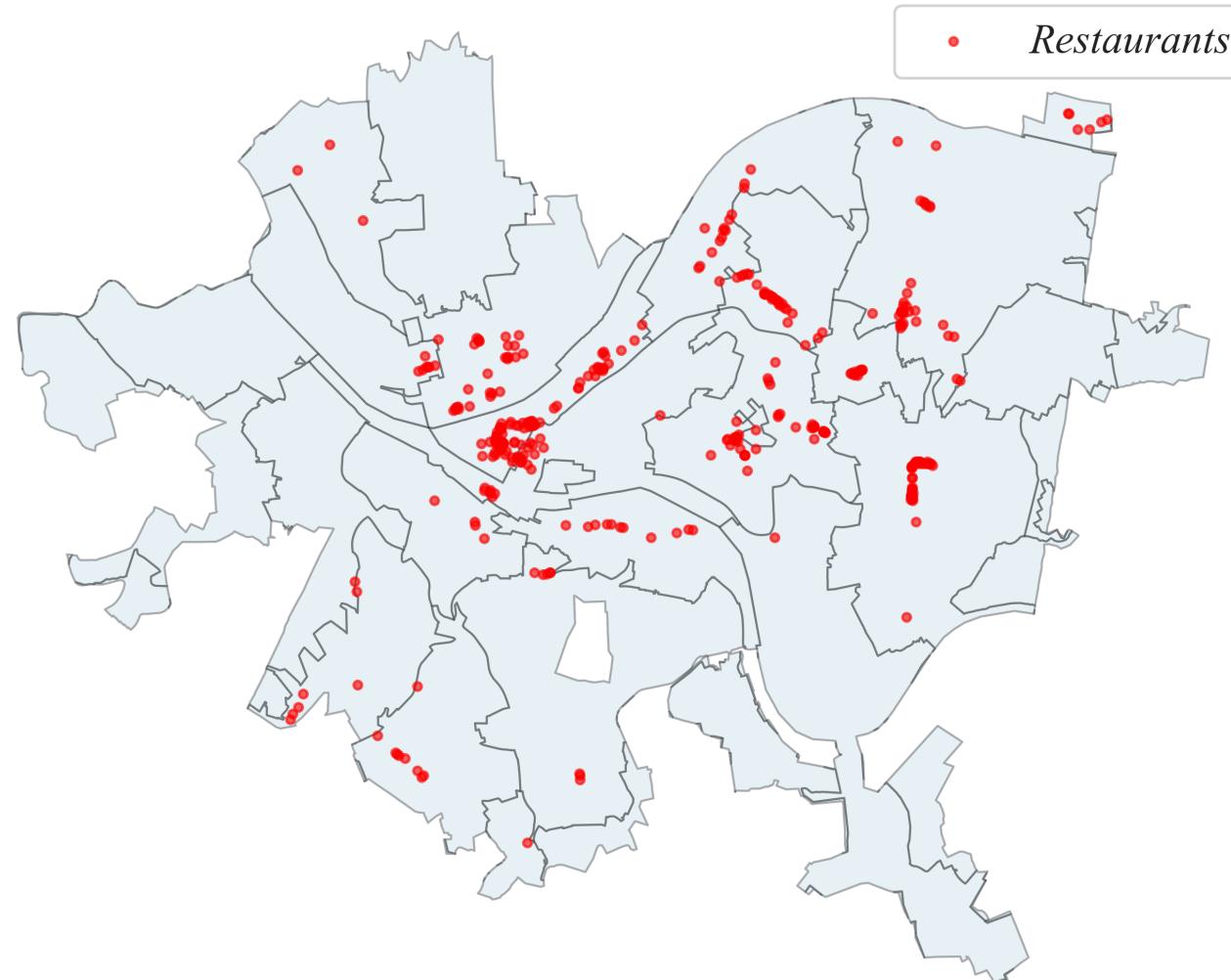
Geographic Data

Are there fewer restaurants further from downtown Pittsburgh?

> lets get back to our question!

Geographic Data

Are there fewer restaurants further from downtown Pittsburgh?



Geographic Data

Are there fewer restaurants further from downtown Pittsburgh?

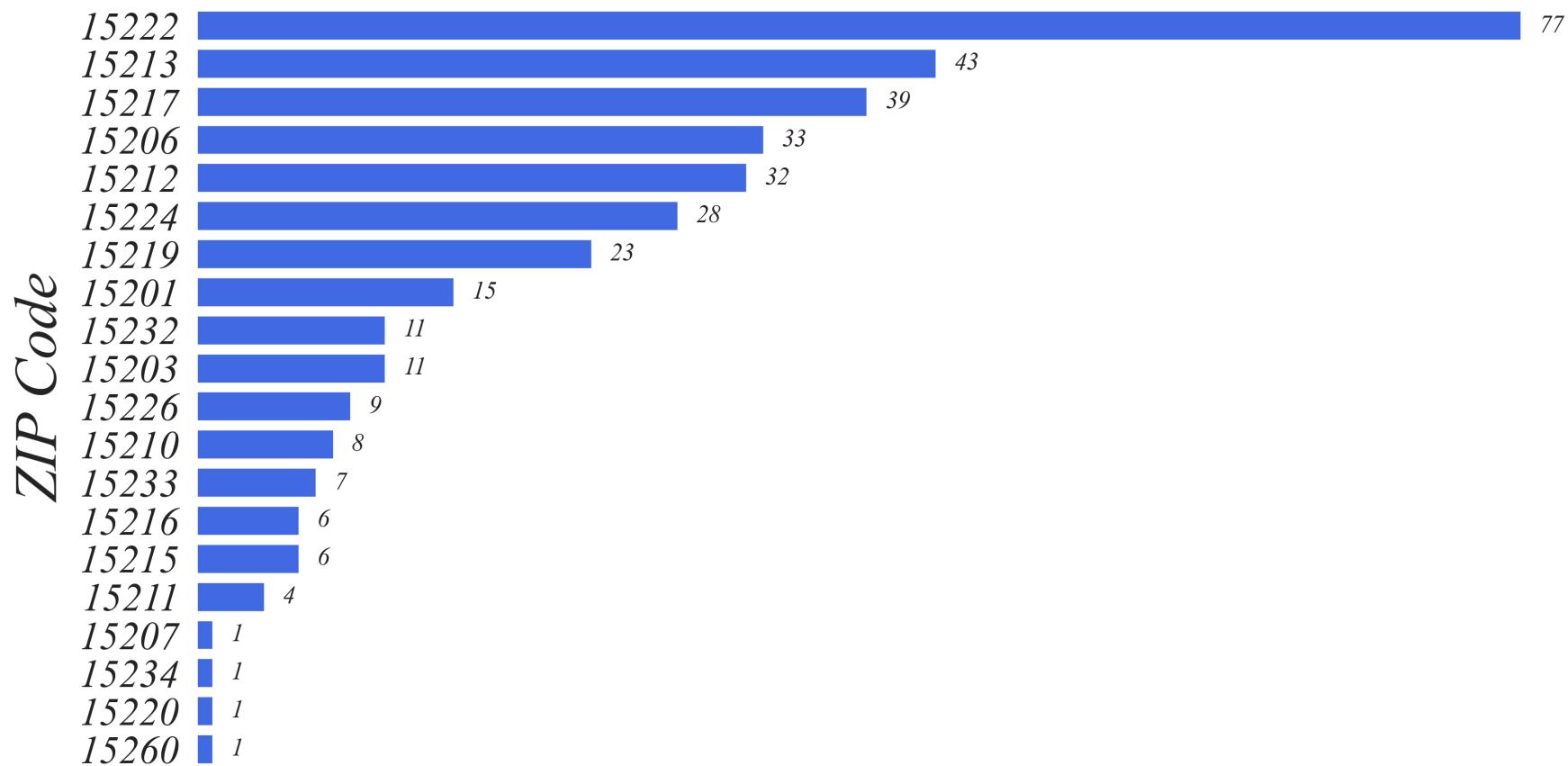
Steps:

- *Measure points by zipcode area*
- *Measure distances between groups (take the centroid, etc)*

Geographic Data

Subquestion 1: how many restaurants are in each Pittsburgh zipcode?

Number of Restaurants by ZIP Code in Pittsburgh



Geographic Data

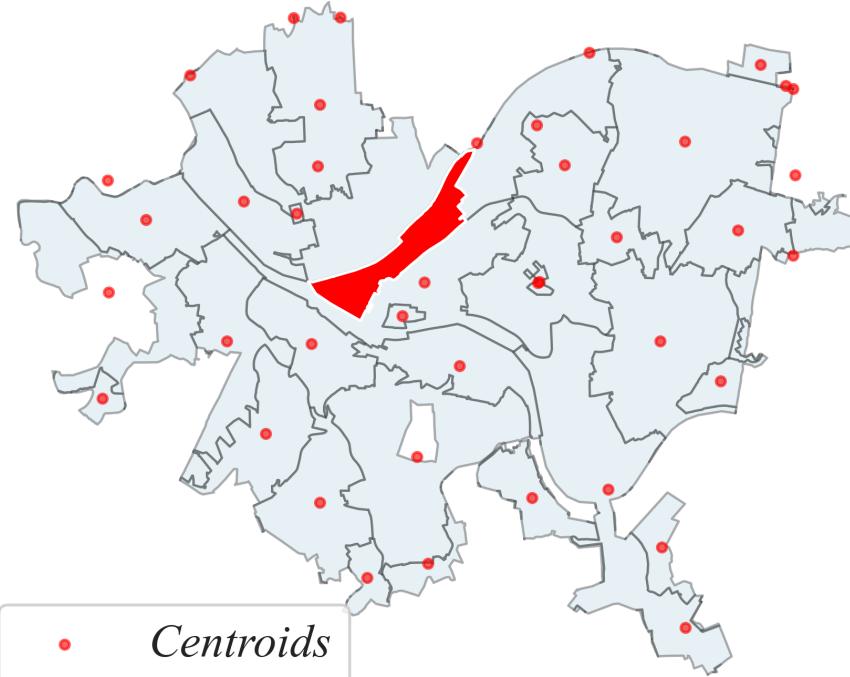
Subquestion 2: how far is each zipcode from downtown?



> measure from the center (*centroid*) of the zipcode

Geographic Data

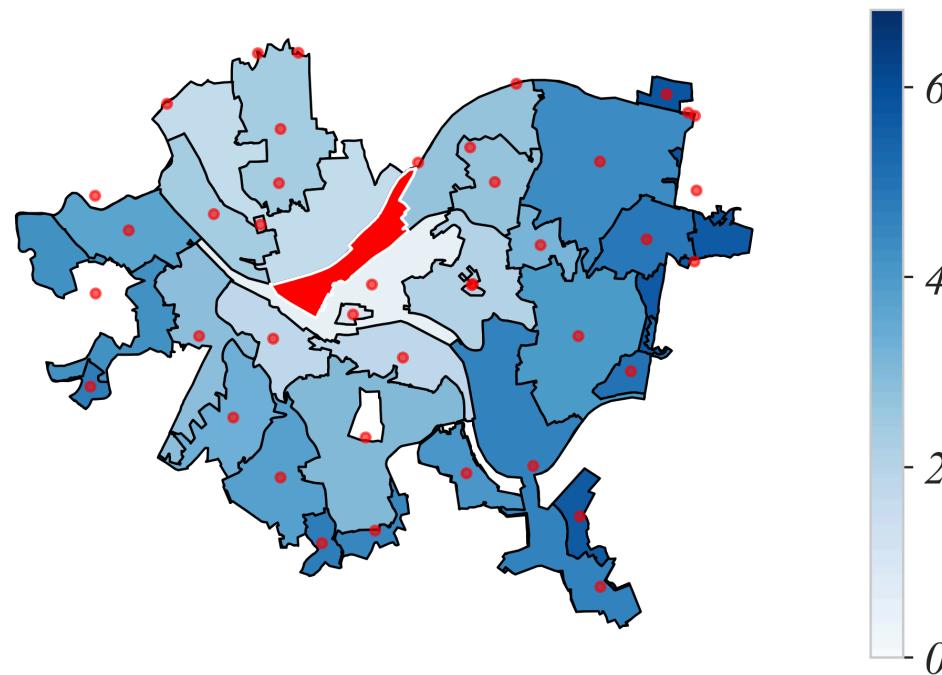
Subquestion 2: how far is each zipcode from downtown?



> measure from the center (centroid) of the zipcode

Geographic Data

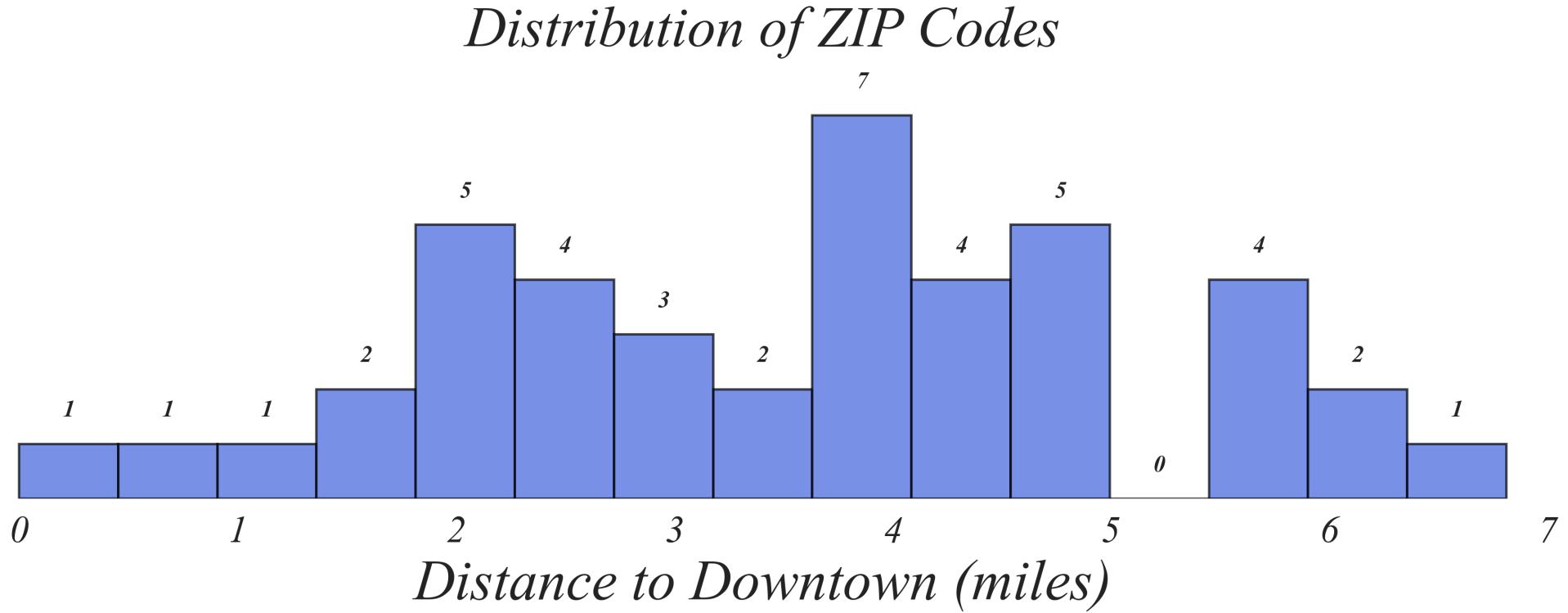
Subquestion 2: how far is each zipcode from downtown?



> what's the distribution?

Geographic Data

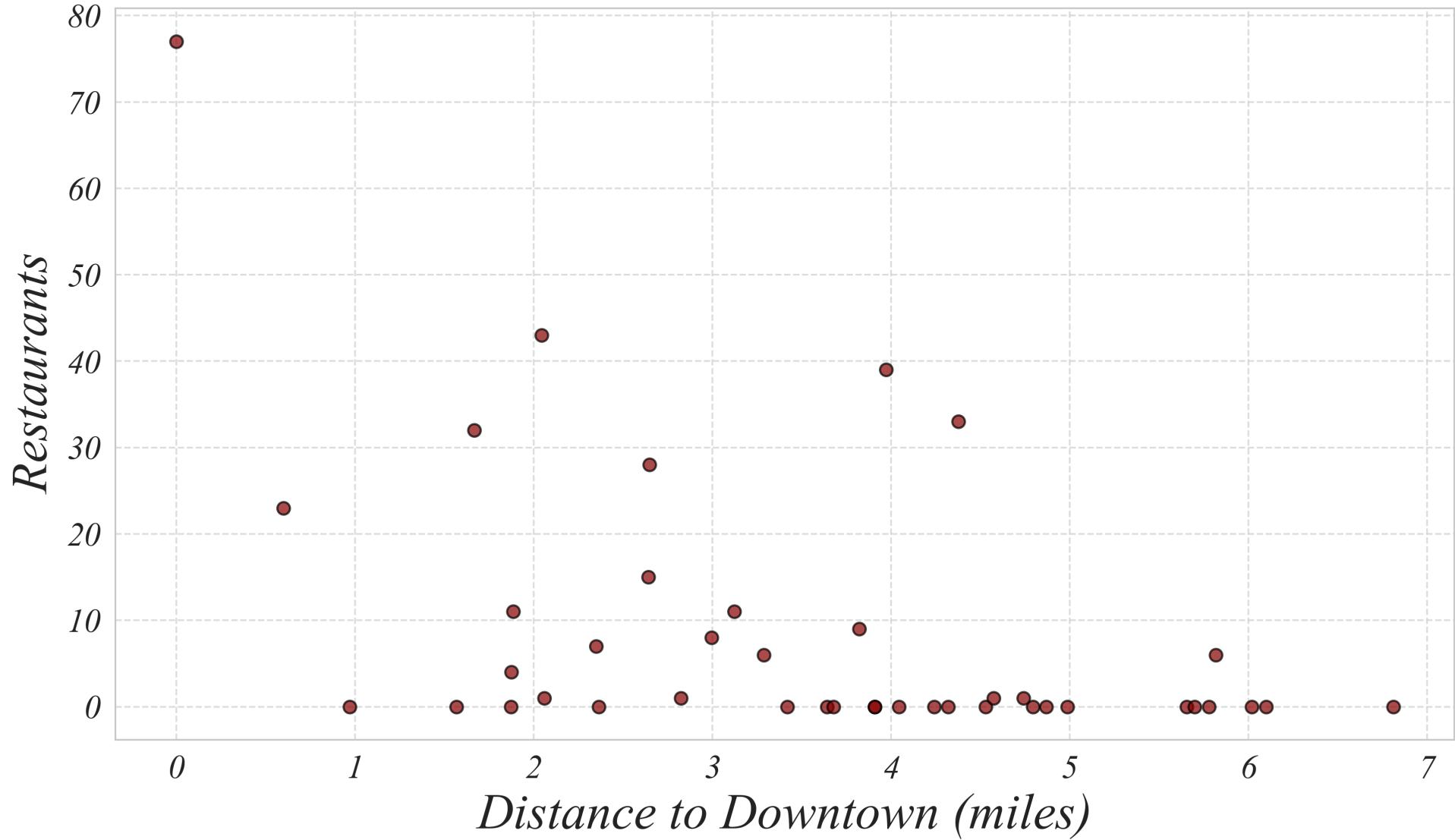
Subquestion 2: how far is each zipcode from downtown?



> we now have enough to answer our main question!

Geographic Data

Are there fewer restaurants in areas further from downtown Pittsburgh?



Homework 1.6 | US City Population and Temp

We're going to use data on locations (`lat`, `lng`), `population`, and temperature (`avg_temp`) of US cities to map temerature and examine whether there is a relationship between latitute (north/south) is related to temperature.

- *Data: `US_Cities.csv` and `Eastern_Cities.csv`*

Part 1 Wrap Up

Exercise 1: Employment Status

What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on the employment status of individuals.

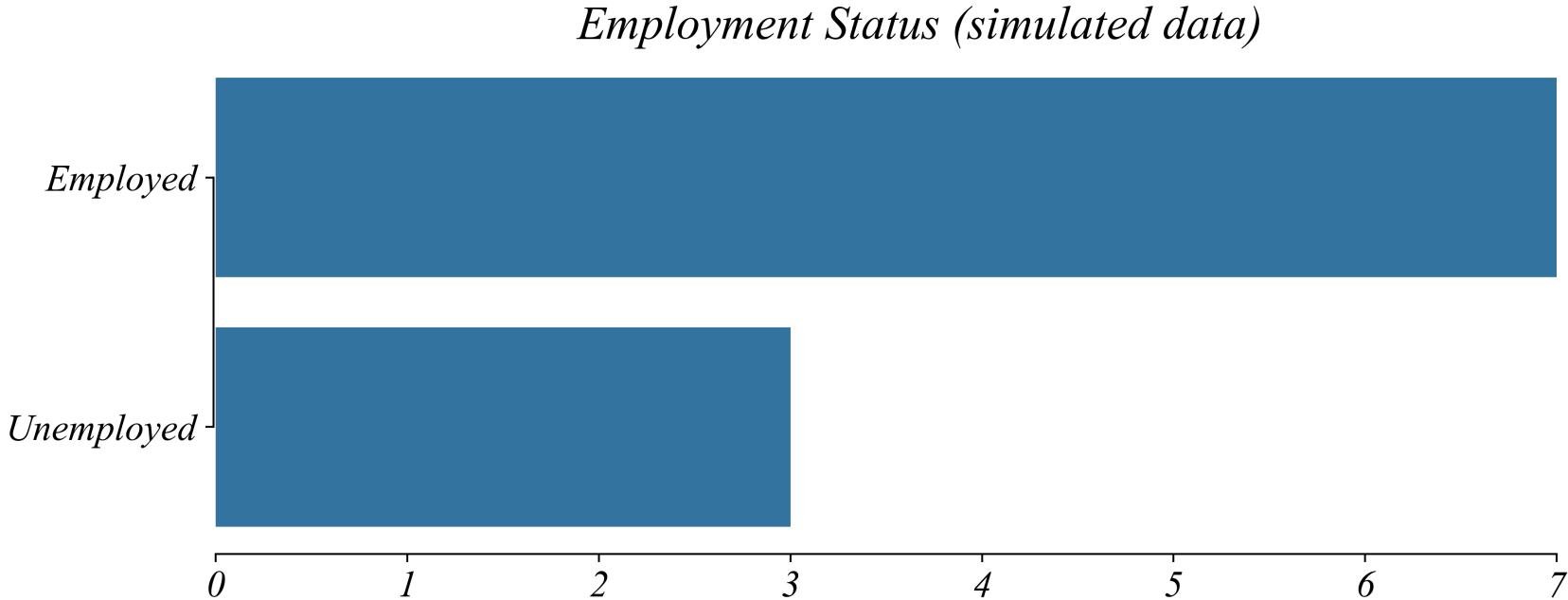
id	status
1	Employed
2	Unemployed
3	Employed
4	Employed
5	Unemployed

Data Dimensions:

- 1. Data Structure:** Cross-Sectional
- 2. Number of Variables:** Univariate
- 3. Variable Type(s):** Binary Categorical

Exercise 1: Employment Status

What 1) are the dimensions of this dataset, and 2) an effective visualization?



A **Barplot** effectively visualizes

1. *Cross-Sectional*
2. *Binary Categorical*
3. *Univariate*

Exercise 2: Employment Industry

What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on the employment sector of individuals.

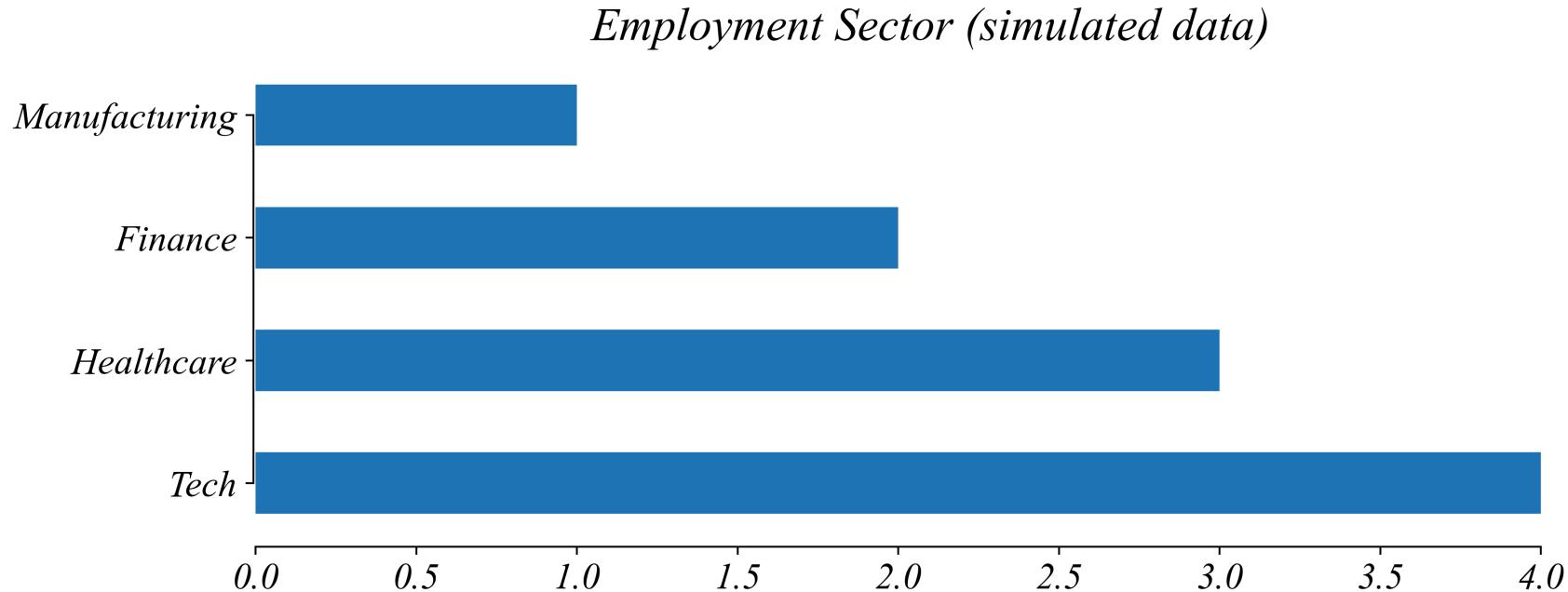
sector
Tech
Healthcare
Finance
Tech
Manufacturing

Data Dimensions:

- 1. Data Structure:** Cross-Sectional
- 2. Number of Variables:** Univariate
- 3. Variable Type(s):** Nominal Categorical

Exercise 2: Employment Industry

What 1) are the dimensions of this dataset, and 2) an effective visualization?



A **Barplot** effectively visualizes

1. *Cross-Sectional*
2. *Nominal Categorical*
3. *Univariate*

Exercise 3: Educational Attainment

What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on the educational attainment of individuals.

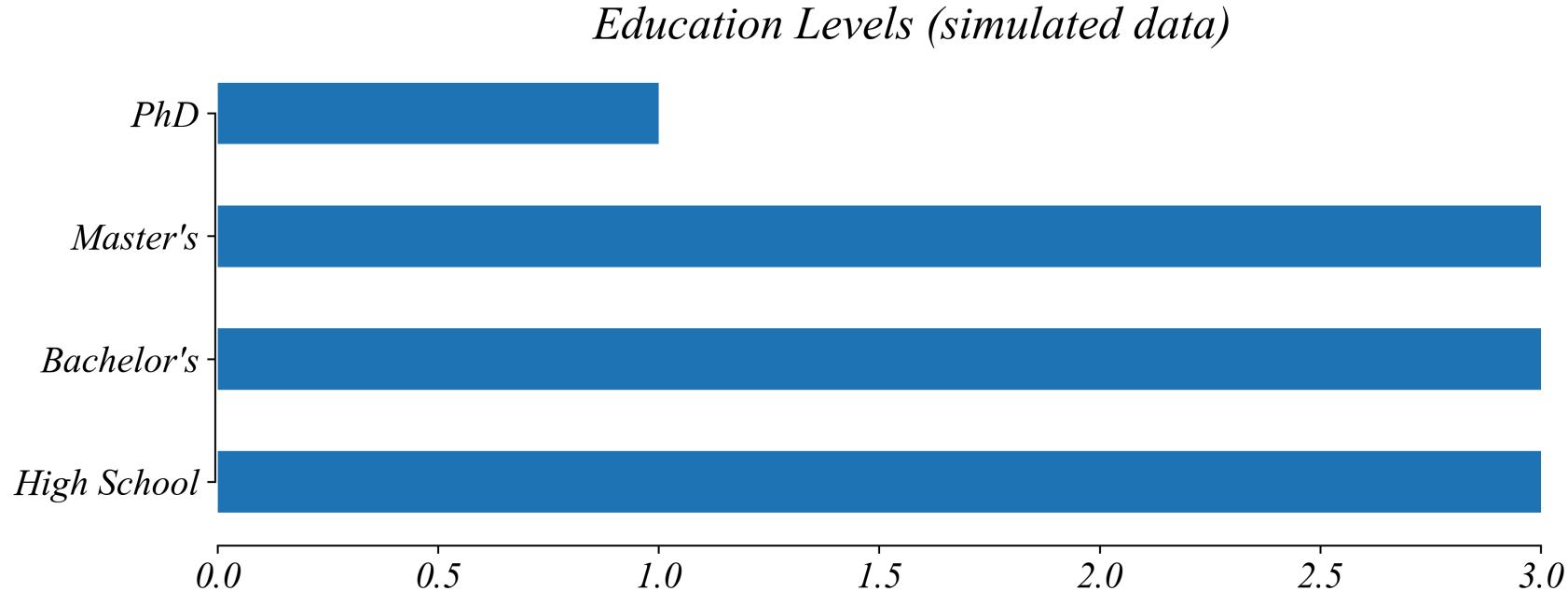
level
High School
Bachelor's
Master's
High School
Bachelor's

Data Dimensions:

- 1. Data Structure:** Cross-Sectional
- 2. Number of Variables:** Univariate
- 3. Variable Type(s):** Ordinal Categorical

Exercise 3: Educational Attainment

What 1) are the dimensions of this dataset, and 2) an effective visualization?



A **Barplot** (or histogram) effectively visualizes

1. *Cross-Sectional*
2. *Ordinal Categorical*
3. *Univariate*

Exercise 4: Annual Income

What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on annual individual income.

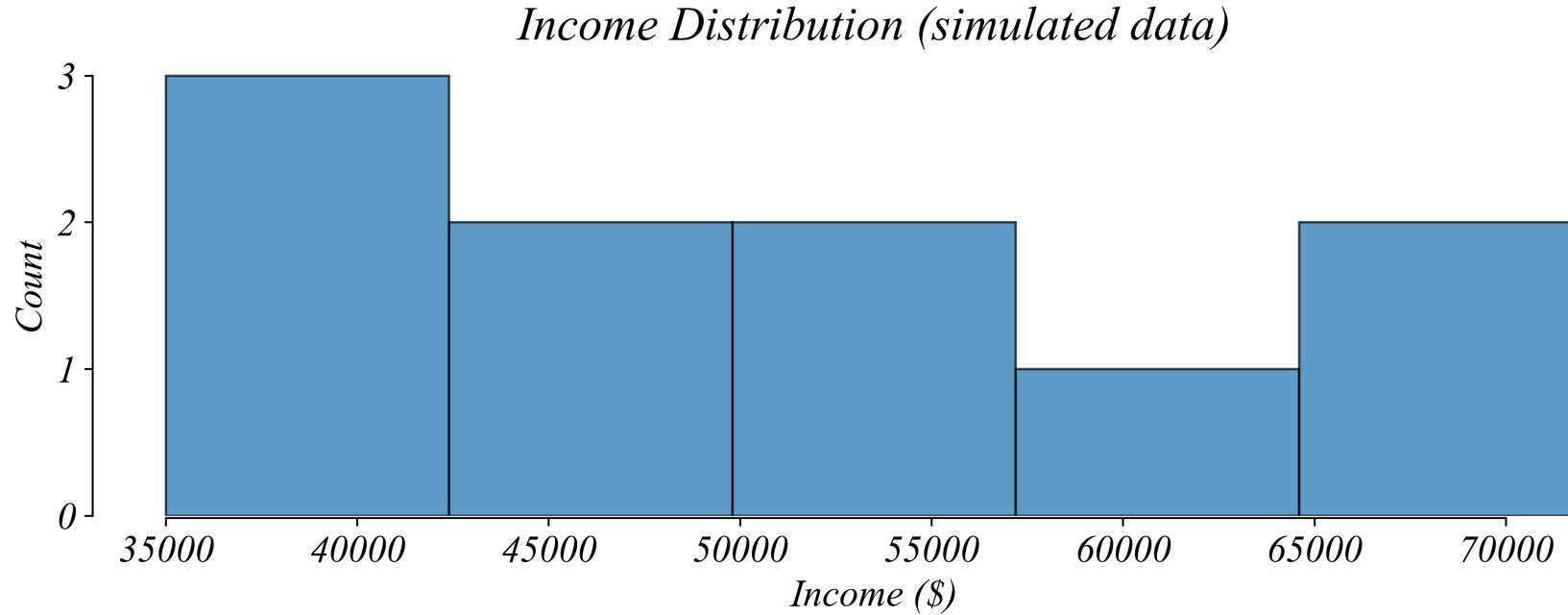
income
45000
52000
38000
65000
41000

Data Dimensions:

- 1. Data Structure:** Cross-Sectional
- 2. Number of Variables:** Univariate
- 3. Variable Type(s):** Numerical

Exercise 4: Annual Income

What 1) are the dimensions of this dataset, and 2) an effective visualization?



A **Histogram** (or boxplot) effectively visualizes

1. Cross-Sectional
2. Numerical
3. Univariate

Exercise 5: Employment by Education

What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on employment by education level.

education	employed
High School	Yes
Bachelor's	Yes
Master's	No
High School	No
Bachelor's	Yes

Data Dimensions:

- 1. Data Structure:** Cross-Sectional
- 2. Number of Variables:** Bivariate
- 3. Variable Type(s):** Categorical by Categorical

Exercise 5: Employment by Education

What 1) are the dimensions of this dataset, and 2) an effective visualization?

We didn't cover how to visualize categorical by categorical bivariate data.

Exercise 6: Income by Education

What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on annual individual income by education.

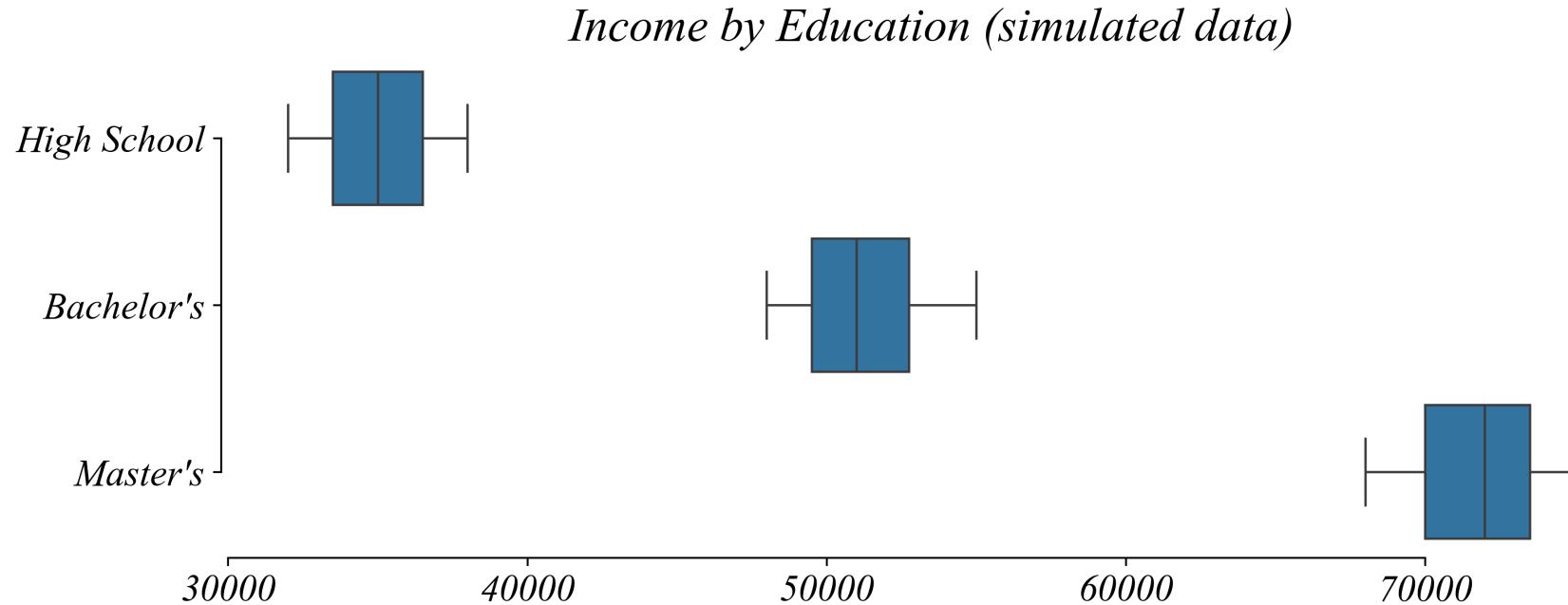
education	income
High School	35000
Bachelor's	52000
Master's	68000
High School	38000
Bachelor's	55000

Data Dimensions:

- 1. Data Structure:** Cross-Sectional
- 2. Number of Variables:** Bivariate
- 3. Variable Type(s):** Numerical by Categorical

Exercise 6: Income by Education

What 1) are the dimensions of this dataset, and 2) an effective visualization?



A **Multi-Boxplot** (or multi-linegraph-histogram) effectively visualizes

1. Cross-Sectional
2. Numerical by Categorical
3. Bivariate

Exercise 7: Income by Age

What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on annual individual income by age.

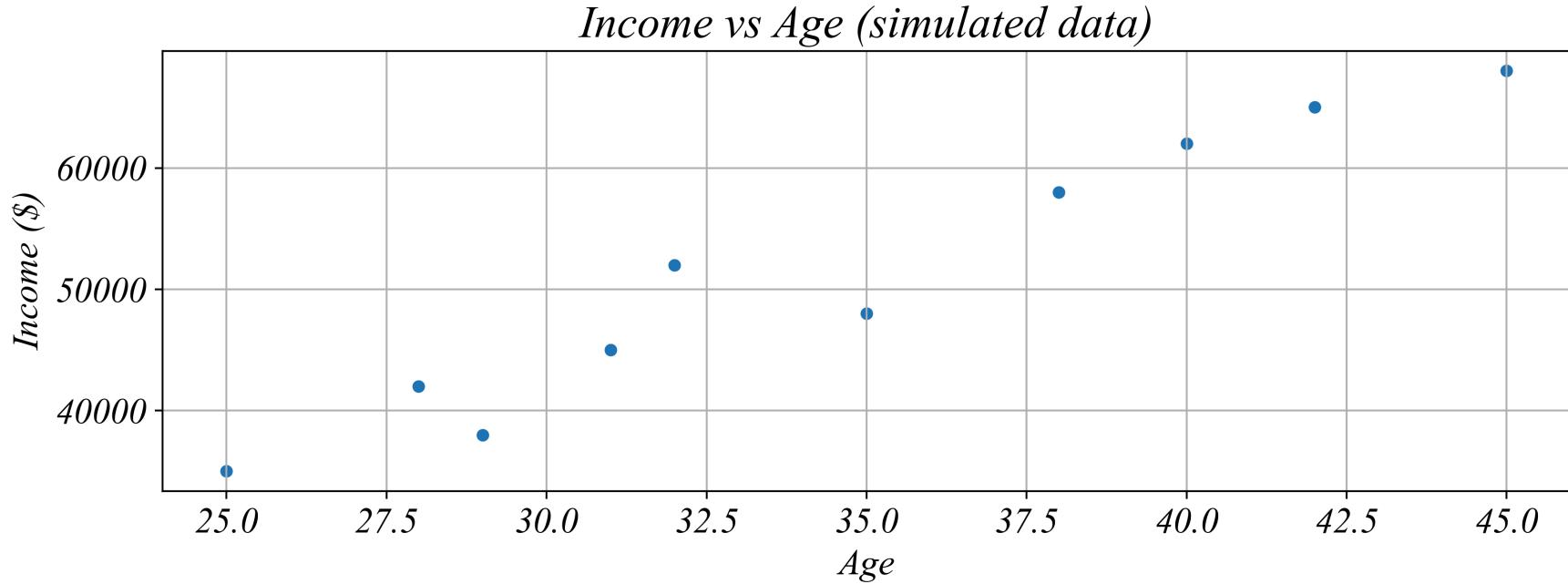
age	income
25	35000
32	52000
28	42000
45	68000
38	58000

Data Dimensions:

- 1. Data Structure:** Cross-Sectional
- 2. Number of Variables:** Bivariate
- 3. Variable Type(s):** Numerical by Numerical

Exercise 7: Income by Age

What 1) are the dimensions of this dataset, and 2) an effective visualization?



A *Scatterplot* effectively visualizes

1. *Cross-Sectional*
2. *Numerical*
3. *Bivariate*

Exercise 8: GDP After 2015

What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on US GDP between 2015 and 2019.

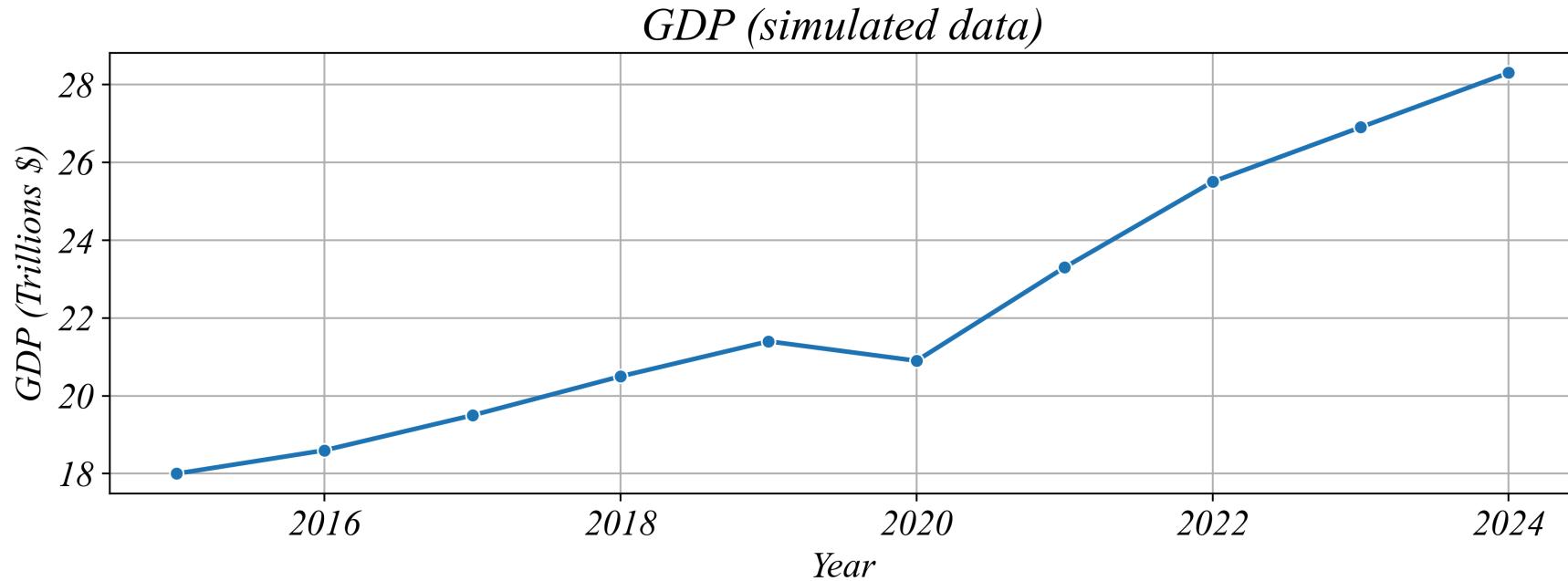
year	gdp
2015	18.000000
2016	18.600000
2017	19.500000
2018	20.500000
2019	21.400000

Data Dimensions:

- 1. Data Structure:** Timeseries
- 2. Number of Variables:** Univariate
- 3. Variable Type(s):** Numerical

Exercise 8: GDP After 2015

What 1) are the dimensions of this dataset, and 2) an effective visualization?



A **Linegraph** effectively visualizes

1. *Timeseries*
2. *Numerical*
3. *Univariate*

Exercise 9: Inflation and Unemployment

What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on inflation and unemployment after 2015.

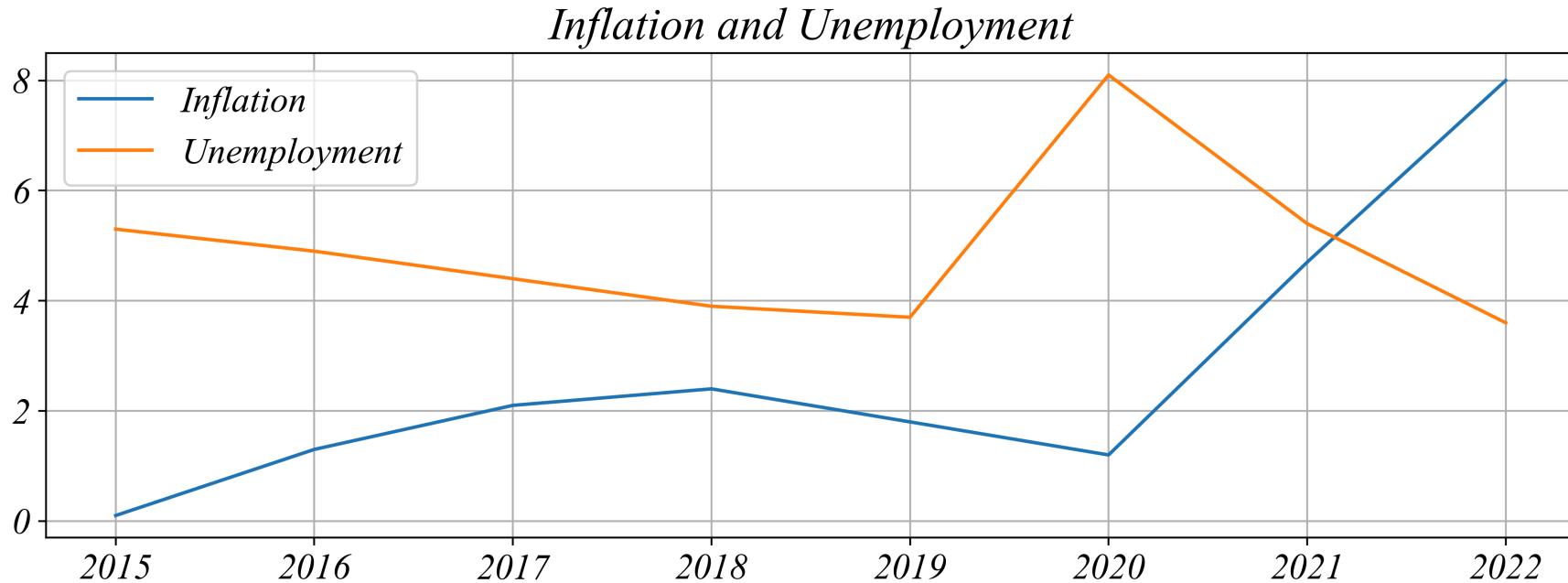
year	unemployment	inflation
2015	5.300000	0.100000
2016	4.900000	1.300000
2017	4.400000	2.100000
2018	3.900000	2.400000
2019	3.700000	1.800000

Data Dimensions:

- 1. Data Structure:** Timeseries
- 2. Number of Variables:** Bivariate
- 3. Variable Type(s):** Numerical

Exercise 9: Inflation and Unemployment

What 1) are the dimensions of this dataset, and 2) an effective visualization?

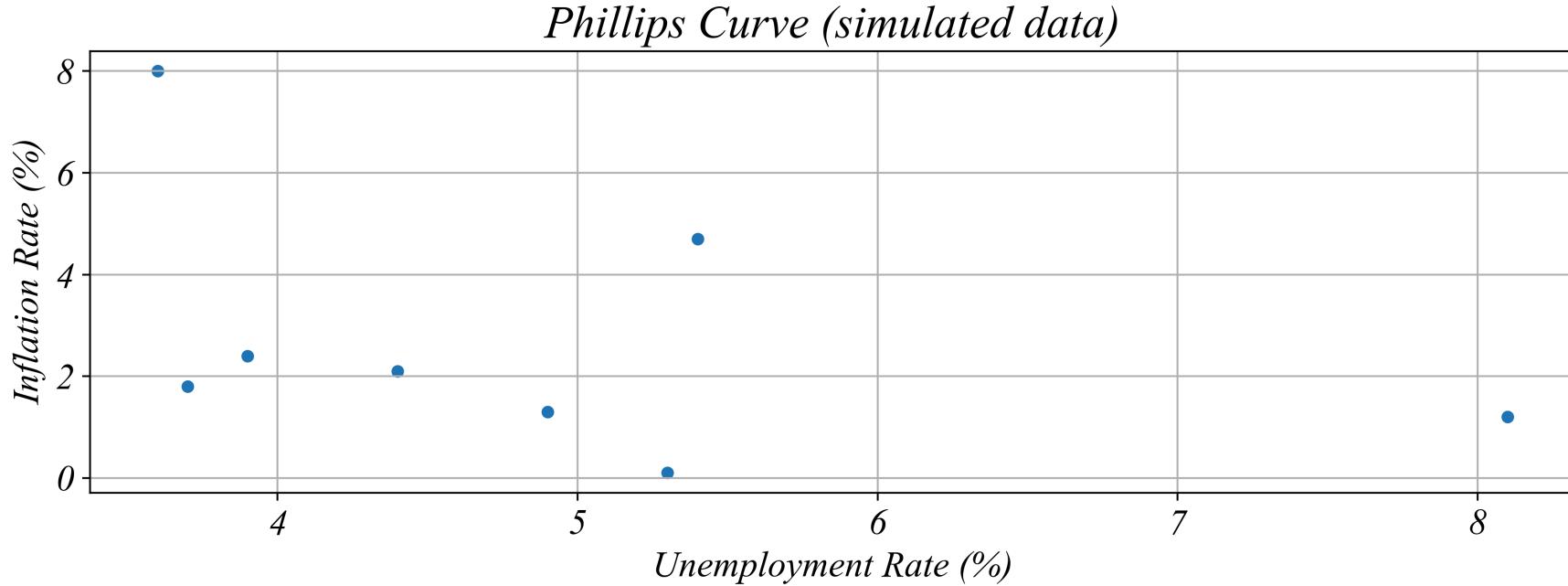


A *Scatterplot* (or sometimes a multilinegraph) effectively visualizes

1. Timeseries
2. Numerical
3. Bivariate

Exercise 9: Inflation and Unemployment

What 1) are the dimensions of this dataset, and 2) an effective visualization?



A **Scatterplot** (or sometimes a multiline graph) effectively visualizes

1. Timeseries
2. Numerical
3. Bivariate

Exercise 10: GDP Growth by Country

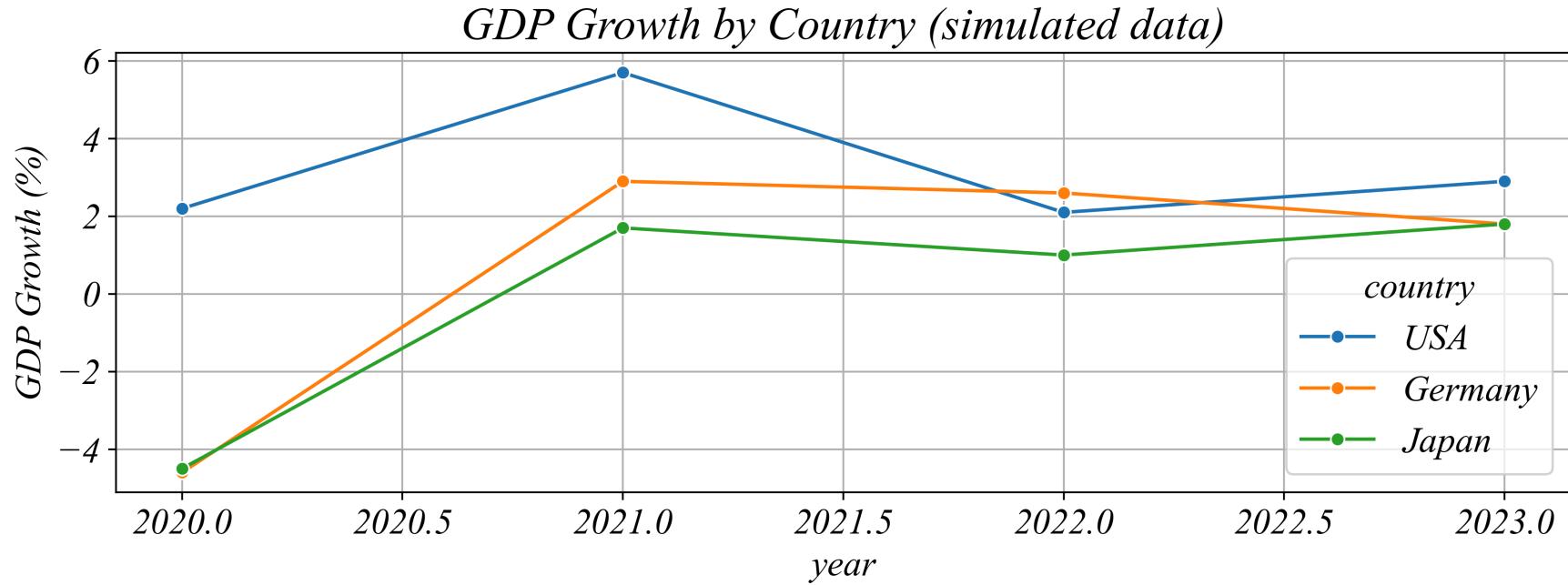
What 1) are the dimensions of this dataset, and 2) an effective visualization?

A dataset on GDP growth by country after 2020.

country	year	gdp_growth
USA	2020	2.200000
USA	2021	5.700000
USA	2022	2.100000
USA	2023	2.900000
Germany	2020	-4.600000

Exercise 10: GDP Growth by Country

What 1) are the dimensions of this dataset, and 2) an effective visualization?



A **Multi-Linegraph** (or sometimes a scatterplot) effectively visualizes

1. Panel
2. Numerical
3. Univariate