ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

Part 1.3 | Data Structures

Part 1.3 | Data Structures
...three main relationships between data points.

The way we understand data depends on the relationship between the data points.

	Cross-Sectional	**Time-Series**	**Panel Data**
Focus	Multiple units, one time point	One unit, many times	Multiple units, many time points
Shape	Wide format	Long format	Long format
Ex.	Household income, 2025	US GDP, 10 years	Household income, 10 years

> *we'll introduce geographic data structures later.*

Exercise: Data Structures

Lets identify the data structure for each dataset.

- *Dataset 1:* `us_real_gdp.csv`
- *Dataset 2:* `household_savings.csv`
- *Dataset 3:* `household_incomes.csv`

Exercise: Dataset 1

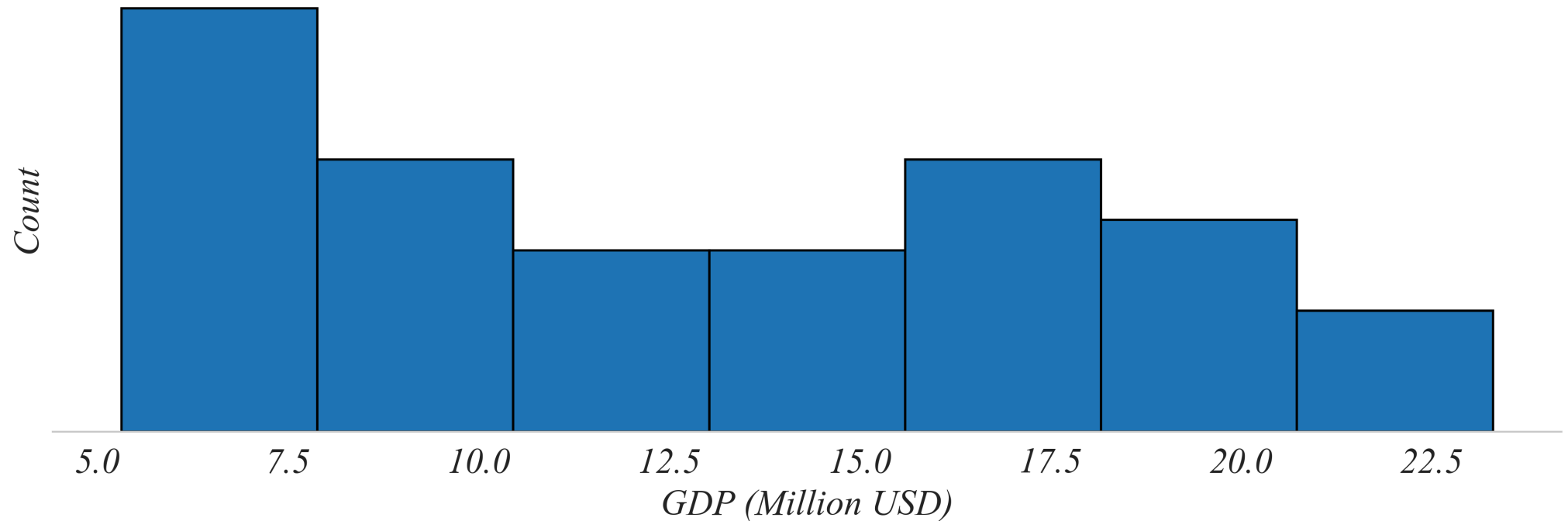*us_real_gdp.csv* *is a Time-Series Dataset*

Clue: Each year's GDP (the unit of observation) is connected through time.

Exercise: Dataset 1

us_real_gdp.csv *is a Time-Series Dataset*

Clue: Each year's GDP (the unit of observation) is connected through time.
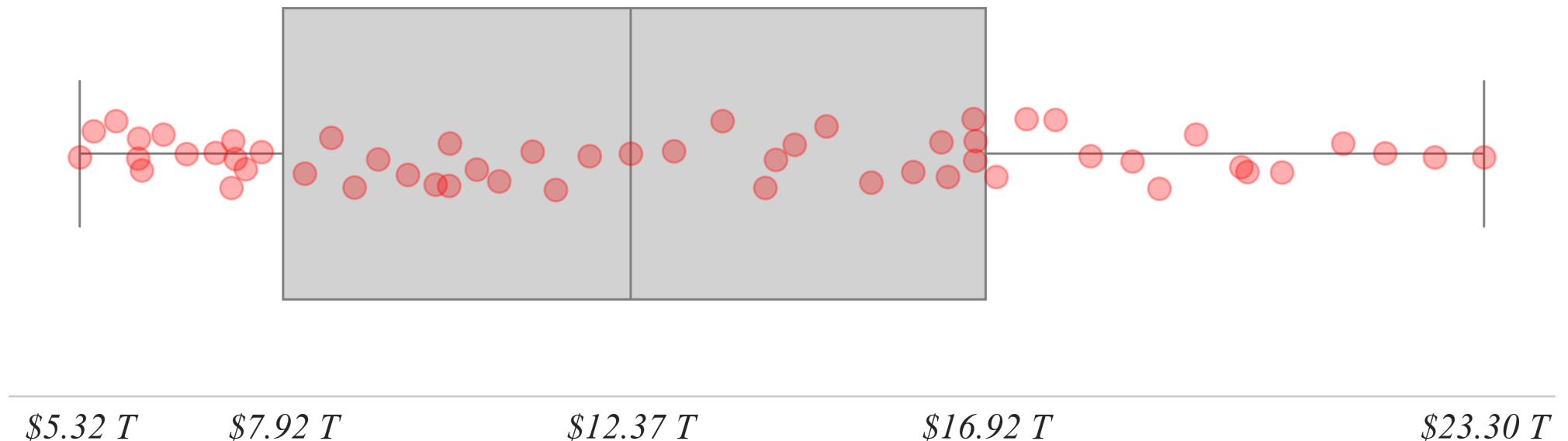

Distribution of GDP

Show the ***distribution*** using a ***histogram*** or a ***boxplot***.

Exercise: Dataset 1

us_real_gdp.csv is a Time-Series Dataset

Clue: Each year's GDP (the unit of observation) is connected through time.
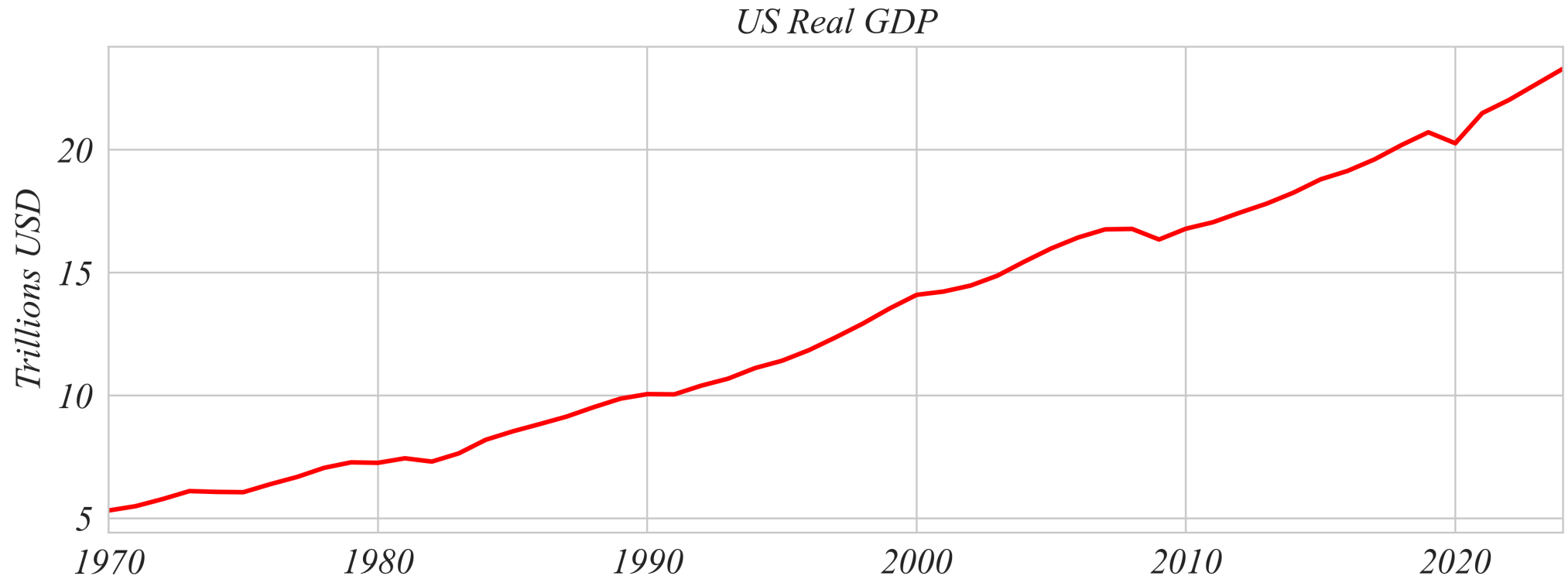
GDP (Million USD)

| $5.32 T | $7.92 T | $12.37 T | $16.92 T | $23.30 T |

Show the ***distribution*** using a ***histogram*** or a ***boxplot***.

Exercise: Dataset 1

us_real_gdp.csv is a Time-Series Dataset

Clue: Each year's GDP (the unit of observation) is connected through time.

US Real GDP

Focus on the ***relationship through time*** using a linegraph.

Exercise: Dataset 2

household_savings.csv is a Cross-Sectional Dataset

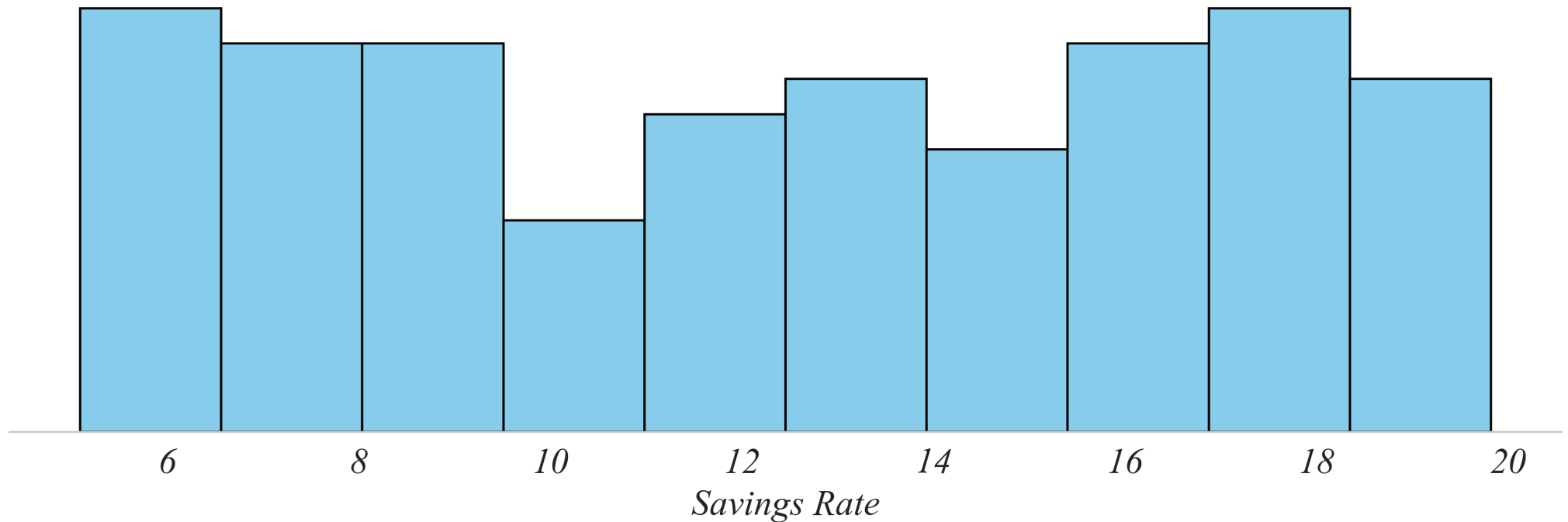**Clue**: Each household (the unit of observation) appears only once.

Exercise: Dataset 2

household_savings.csv is a Cross-Sectional Dataset

Clue: Each household (the unit of observation) appears only once.
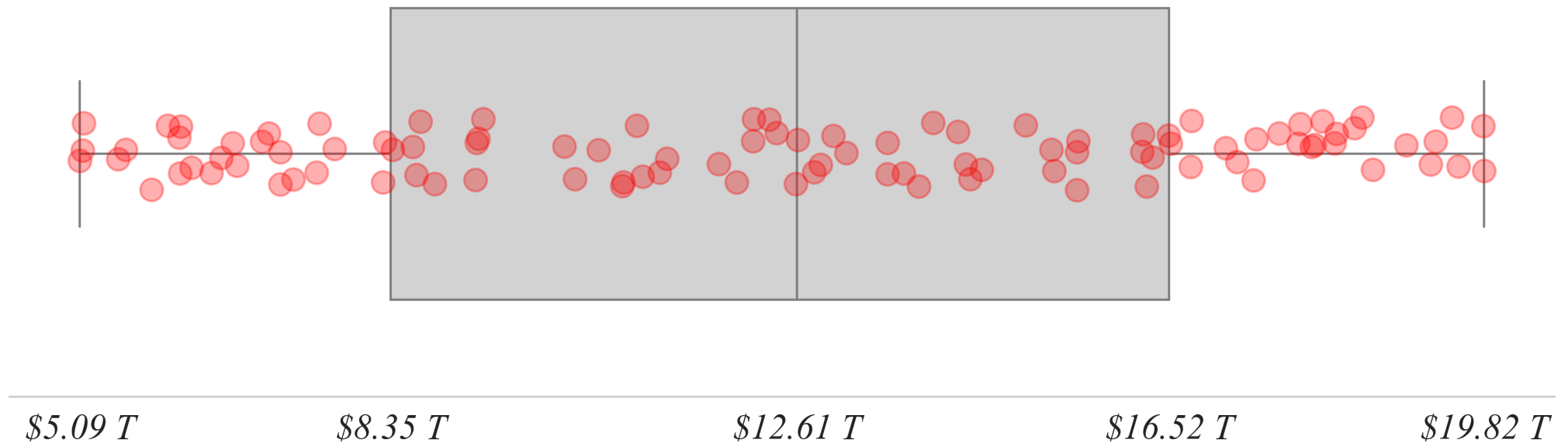
Household Savings Rates

Savings Rate

Focus on the *distribution* using a *histogram* or a *boxplot*.

Exercise: Dataset 2

household_savings.csv is a Cross-Sectional Dataset

Clue: Each household (the unit of observation) appears only once.

Household Savings Rates

| $5.09 T | $8.35 T | $12.61 T | $16.52 T | $19.82 T |

Focus on the ***distribution*** using a ***histogram*** or a ***boxplot***.
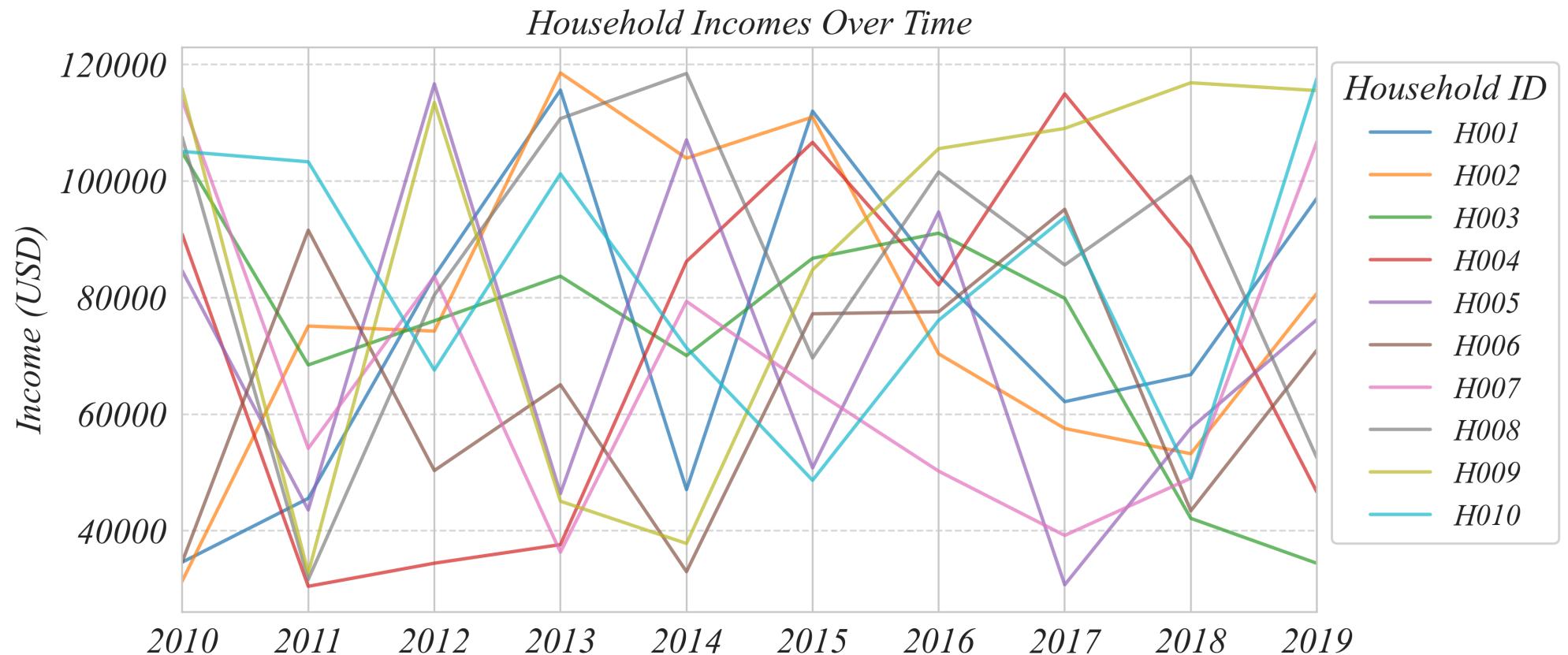
Exercise: Dataset 3

household_incomes.csv *is a Panel Dataset*

Clue: Each household (the unit of observation) has many appearances.

Exercise: Dataset 3

`household_incomes.csv` *is a Panel Dataset*

Clue: Each household (the unit of observation) has many appearances.
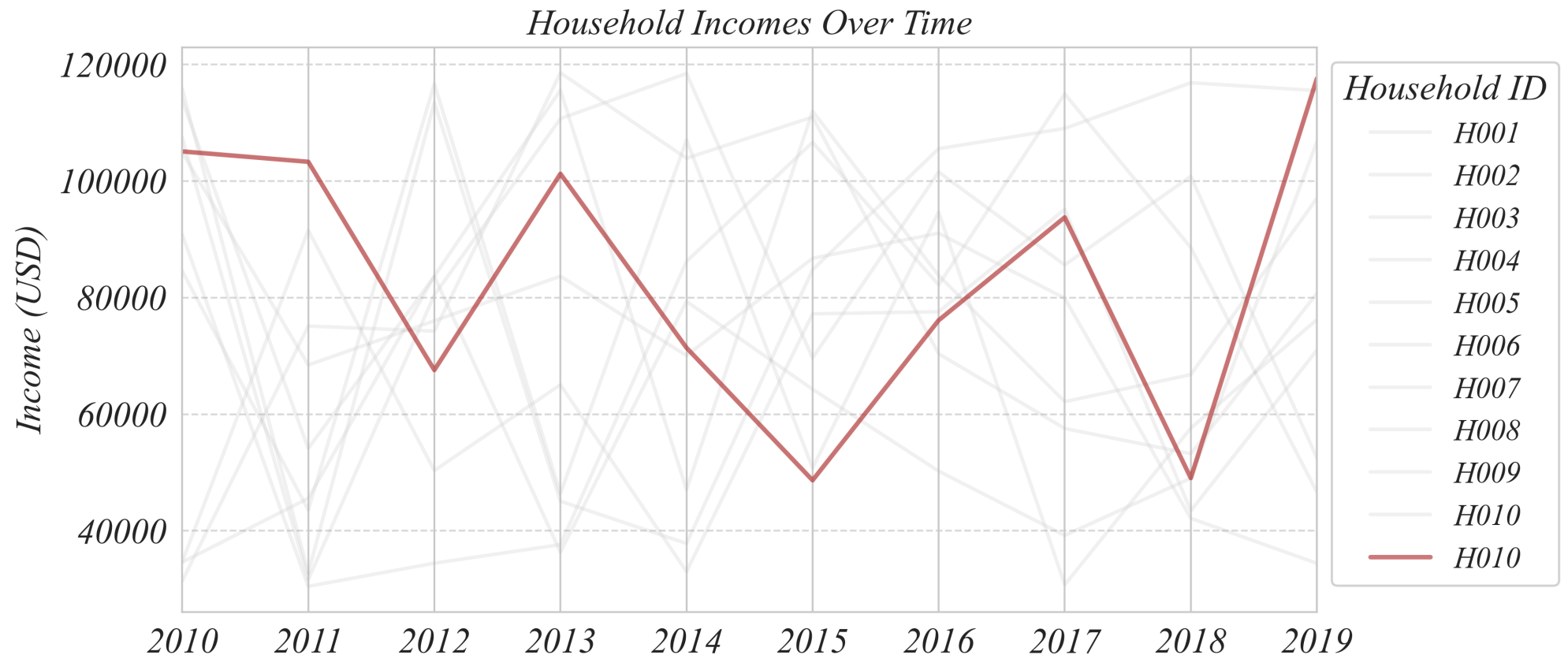

Household Incomes Over Time

> *this is too noisy to communicate well*

Exercise: Dataset 3

household_incomes.csv *is a Panel Dataset*

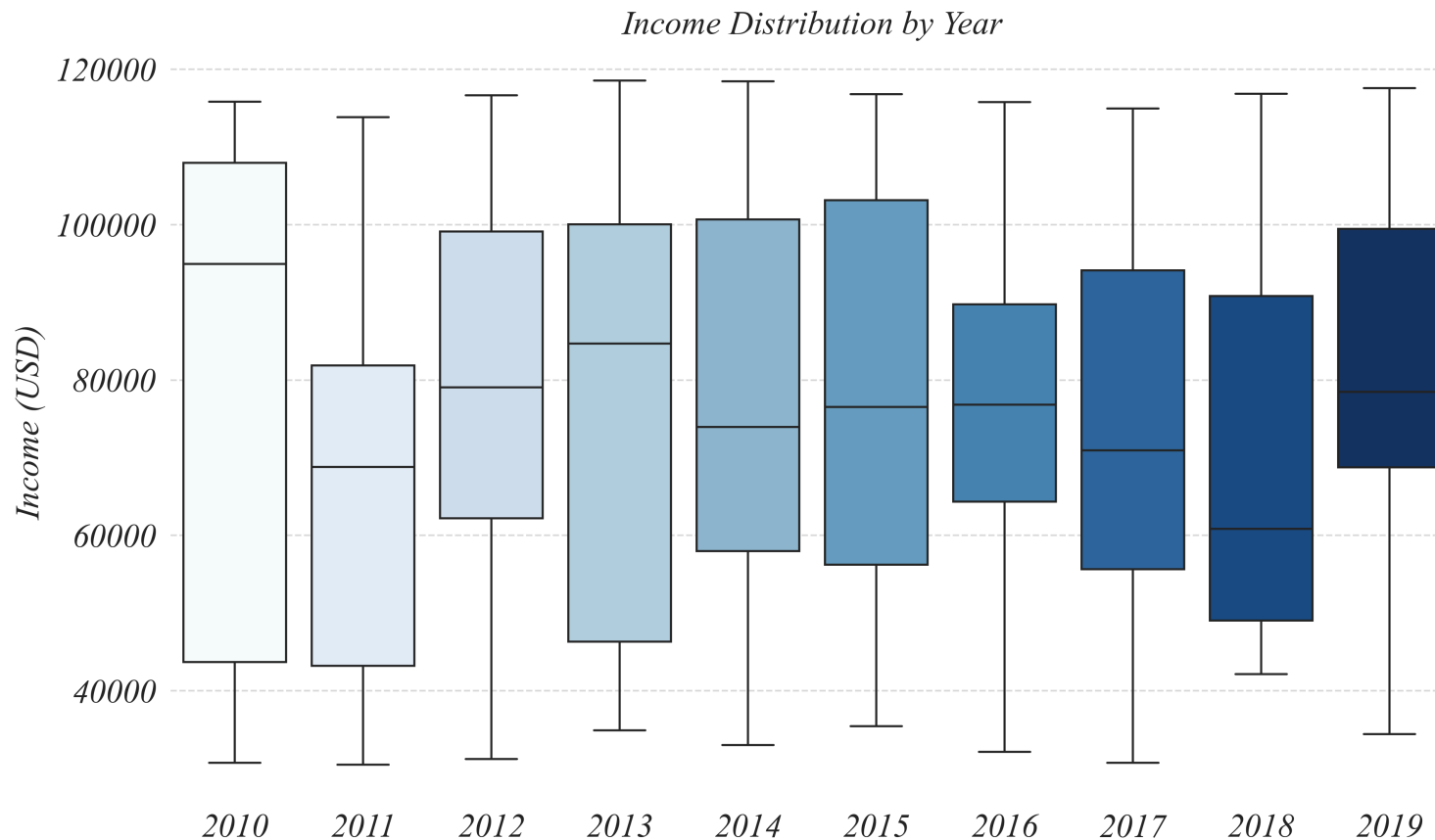**Clue**: Each household (the unit of observation) has many appearances.

Household Incomes Over Time

> *this is a little better*

Exercise: Dataset 3

`household_incomes.csv` *is a Panel Dataset*

Clue: Each household (the unit of observation) has many appearances.
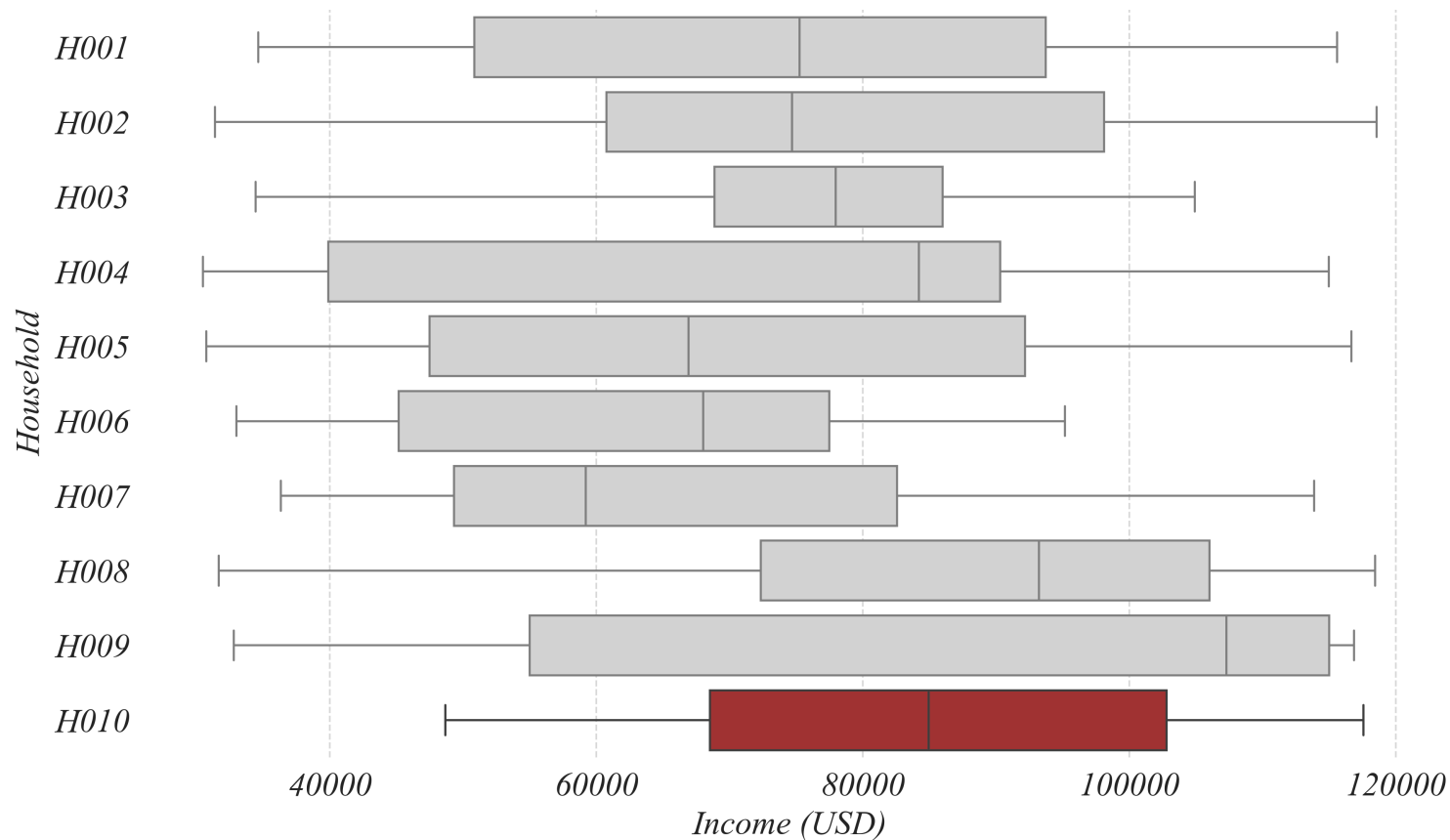

Income Distribution by Year

Focus on the variation *among the households* within the year using a ***boxplot***.

Exercise: Dataset 3

household_incomes.csv *is a Panel Dataset*

Clue: Each household (the unit of observation) has many appearances.

Focus on the variation *within the household* using a ***boxplot***.