

# ECON 0150 | Economic Data Analysis

*The economist's data analysis pipeline.*

## *Part 1.0 | Variable Types*

# *Part 1.0 | Variable Types*

We cannot typically understand our data without summarizing it.



The main differentiators between a good and bad summarization tools is the dimension of the data and what you're trying to understand about the data.

# Dimensions of Data

*Data comes in all shapes, sizes, and types.*

## Variable Type

- *Categorical data*
- *Numerical data*

## Data Structure

- *Cross-sectional data*
- *Time series data*
- *Panel data*

## Number of Variables

- *Univariate ( $n=1$ )*
- *Bivariate ( $n=2$ )*
- *Multivariate ( $n>2$ )*

# Variable Types: Categorical

*... data that's best recorded in categories*

**Binary:** only two categories

- *Economics major (YES/NO)*
- *Human (YES/NO)*

**Nominal:** categories cannot be ordered / ranked

- *Blood types (A, B, AB, O)*
- *Colors (Red, Blue, Green)*

**Ordinal:** categories have order / rank but not a meaningful scale

- *Education levels (High School, Bachelor's, Master's, PhD)*
- *Survey responses (Strongly Disagree, Neutral, Agree, Strongly Agree)*
- *Size categories (Small, Medium, Large)*

# Variable Types: Numerical

*... data that's best recorded in numerical form*

**Discrete:** countable numbers with meaningful intervals

- *Number of children in a household (1, 2, ... )*
- *Number of siblings (1, 2, ... )*

**Continuous:** quantities measurable on the reals

- *Household income*
- *US GDP*

# Exercise 1.0 | Variable Types

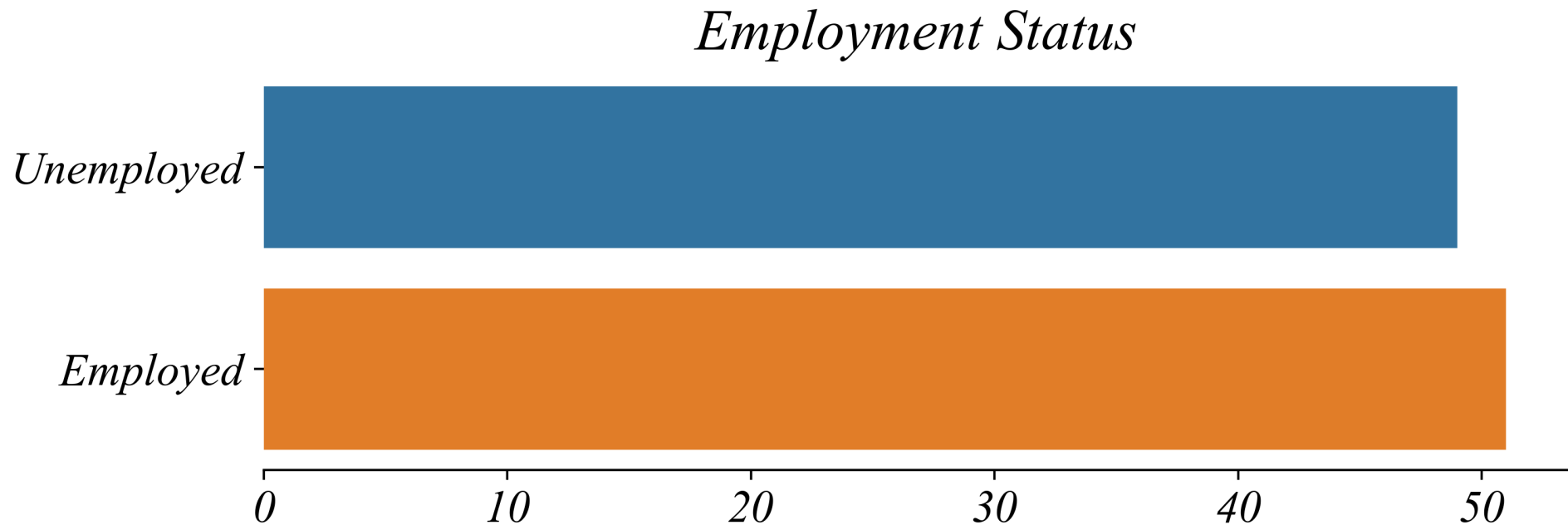
Lets identify the variable type for each dataset.

- *Dataset 1: [employment\\_status.csv](#)*
- *Dataset 2: [employment\\_sector.csv](#)*
- *Dataset 3: [economic\\_optimism.csv](#)*
- *Dataset 4: [household\\_size.csv](#)*
- *Dataset 5: [household\\_income.csv](#)*

# Exercise: Employment Status

*What type of variables are contained in `employment_status.csv`?*

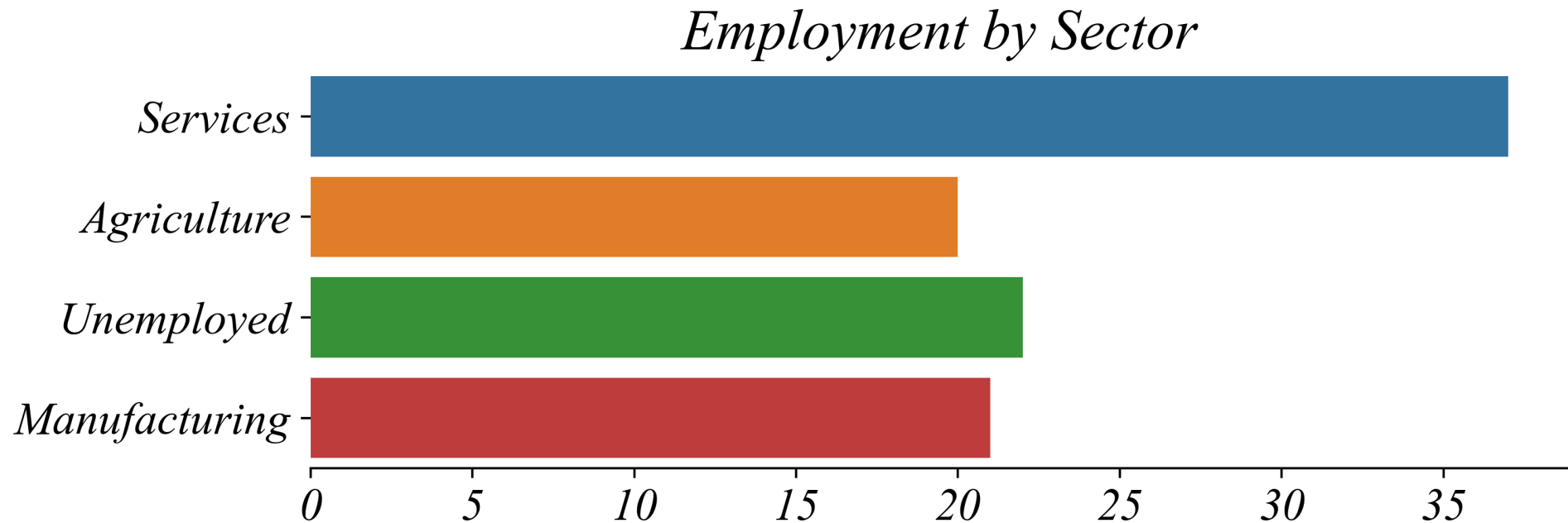
**Binary Categorical:** two categories (*Employed*, *Unemployed*)



# Exercise: Employment Sector

*What type of variables are contained in `employment_sector.csv`?*

**Nominal:** no inherent order (*Agriculture, Services, Unemployed*)

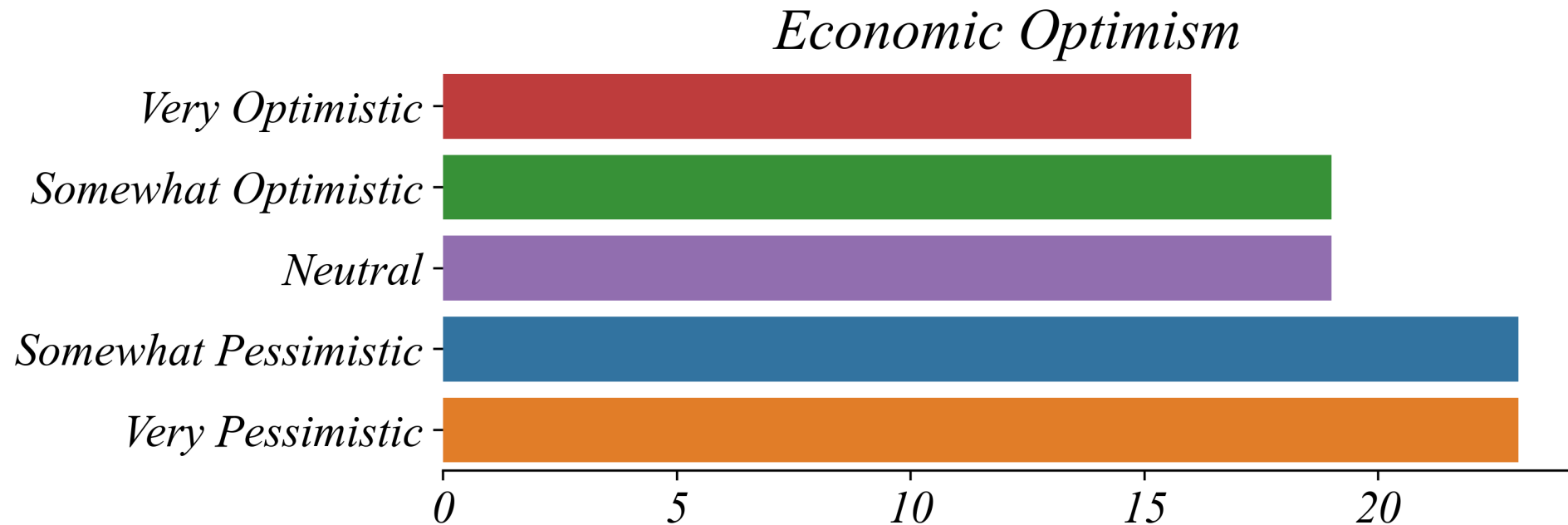




# Exercise: Economic Optimism

*What type of variables are contained in `economic_optimism.csv`?*

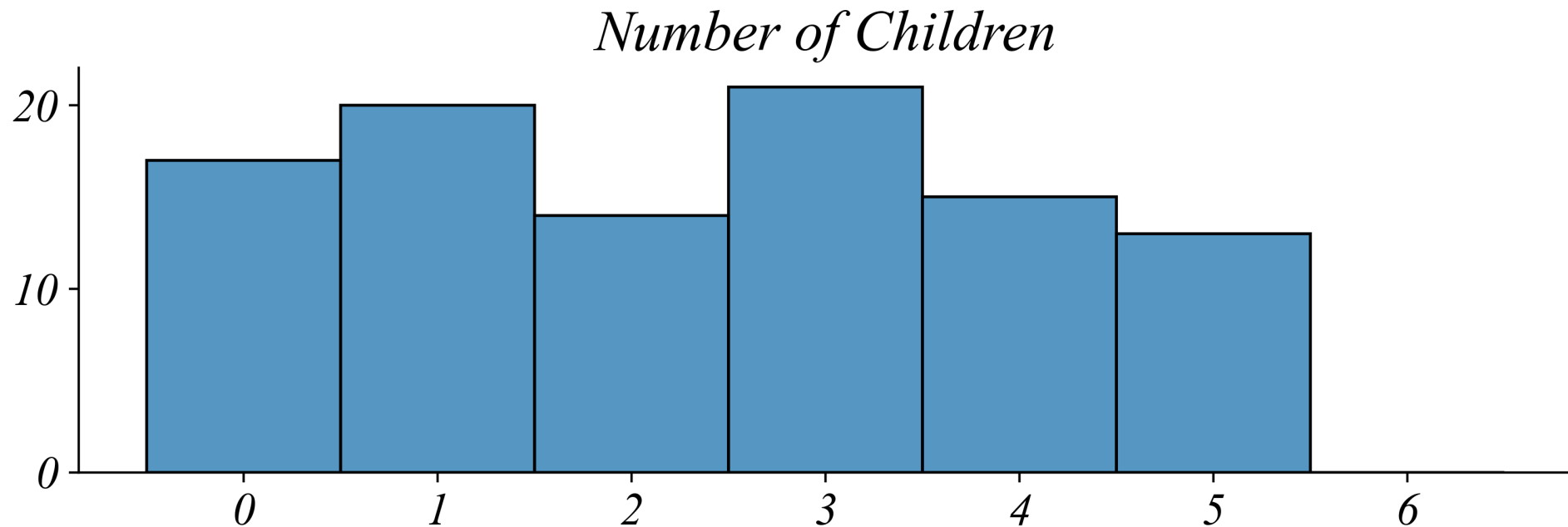
**Ordinal:** meaningful order without meaningful intervals (*Very Pessimistic to Very Optimistic*)



# Exercise: Household Size

*What type of variables are contained in `household_size.csv`?*

**Discrete Numerical:** countable numbers with meaningful intervals (*Number of Children in a Household*)



# Exercise: Household Income

*What type of variables are contained in `household_income.csv`?*

**Continuous Numerical:** quantities measurable on the reals (*Income in USD*)

*Household Income (USD)*

