

ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

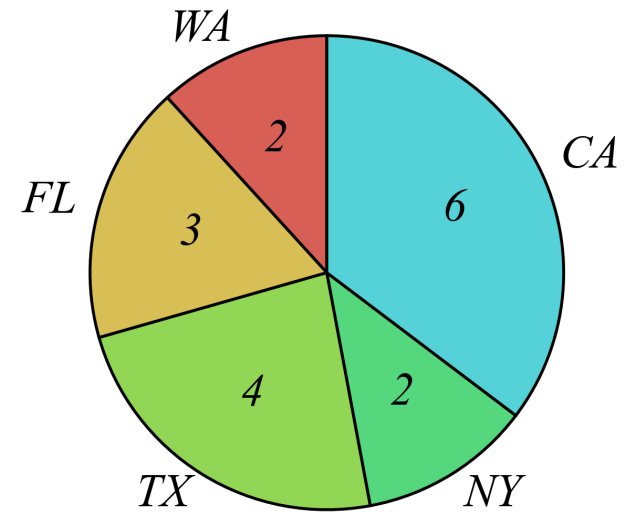
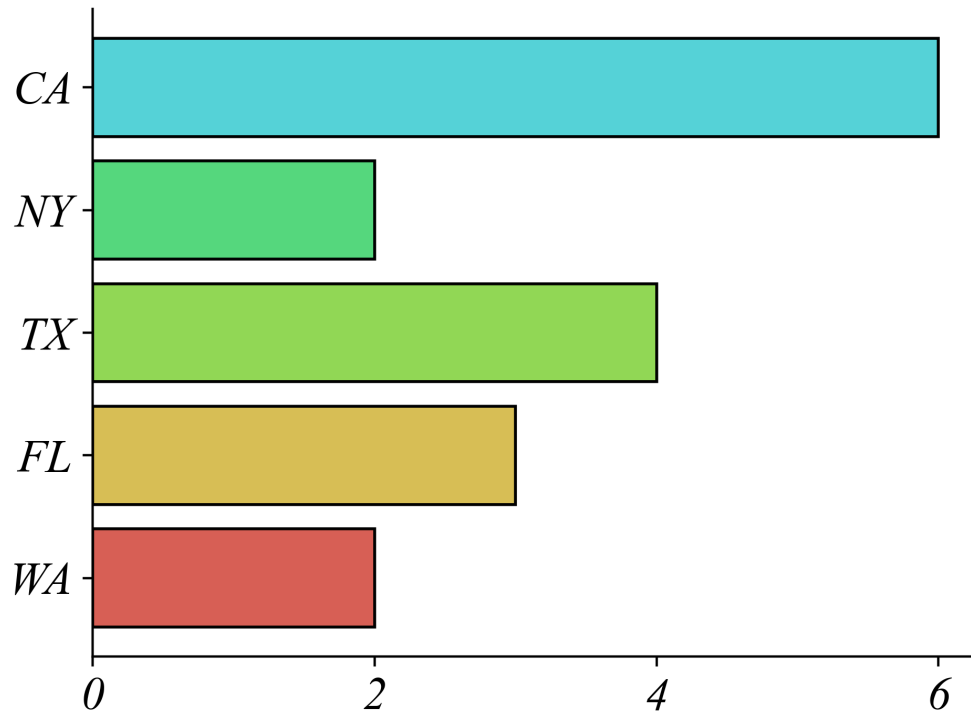
Part 1.1 | Summarizing Categorical Variables

Summarizing Categorical Variables

... use the appropriate summary tool for the variable type

Catagorical Variables: Visualizations

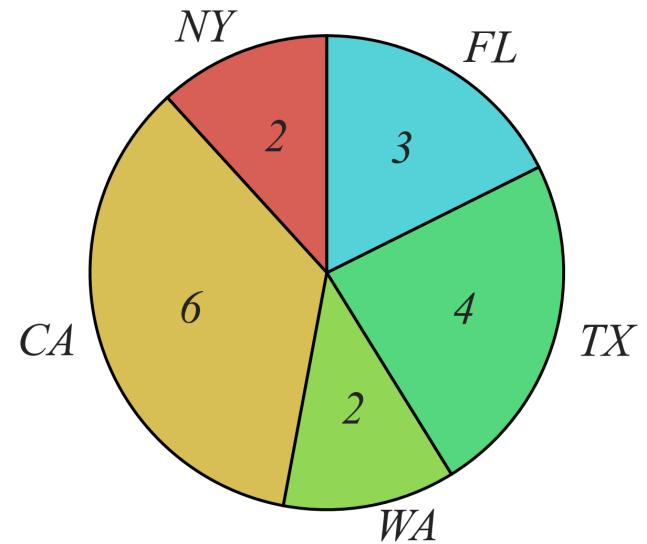
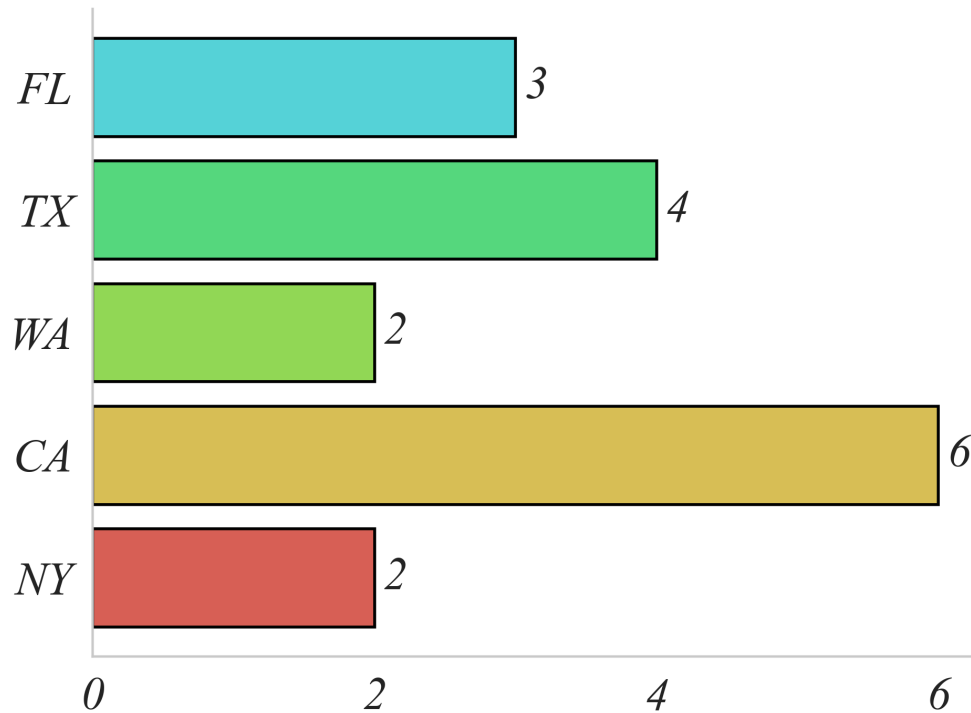
Q. Which state has the second most locations?



> pay attention to which of these two figures is easier to answer the question

Catagorical Variables: Bar Plots

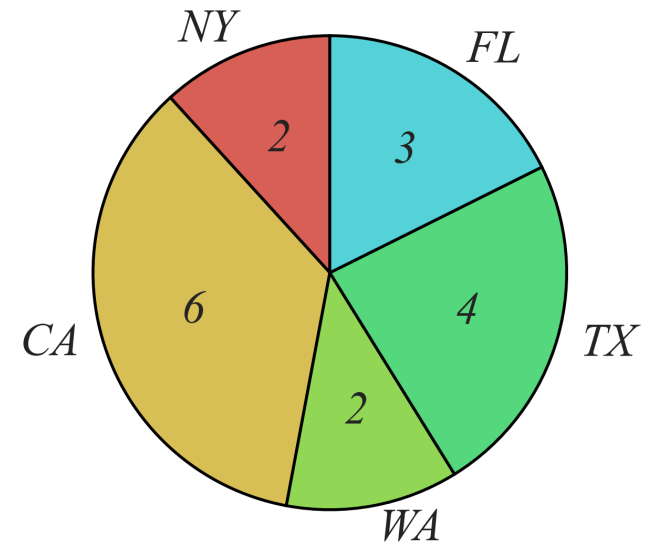
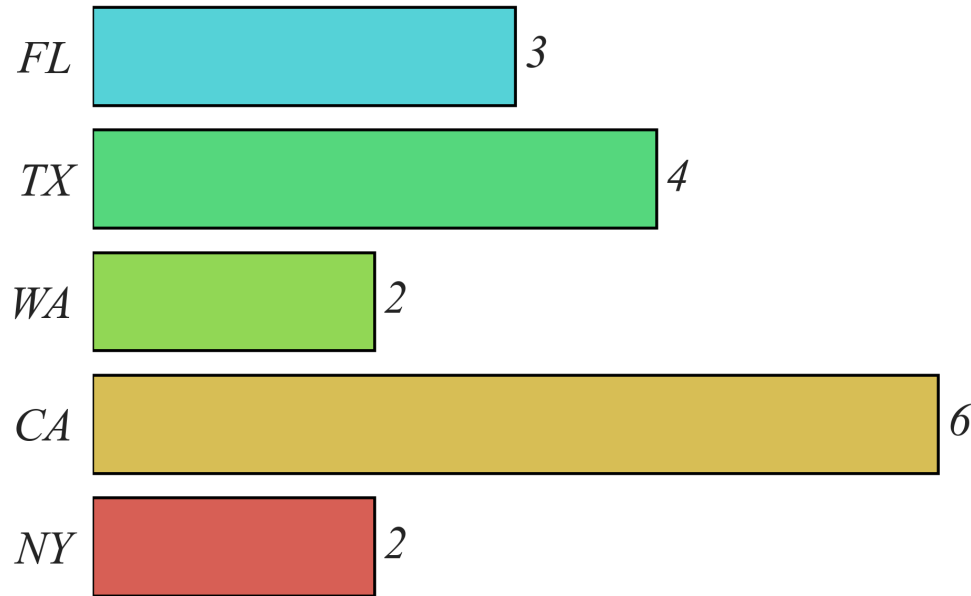
Q. Which state has the second most locations?



> make the bar graph even easier to read by placing the number near the bar

Catagorical Variables: Remove Clutter

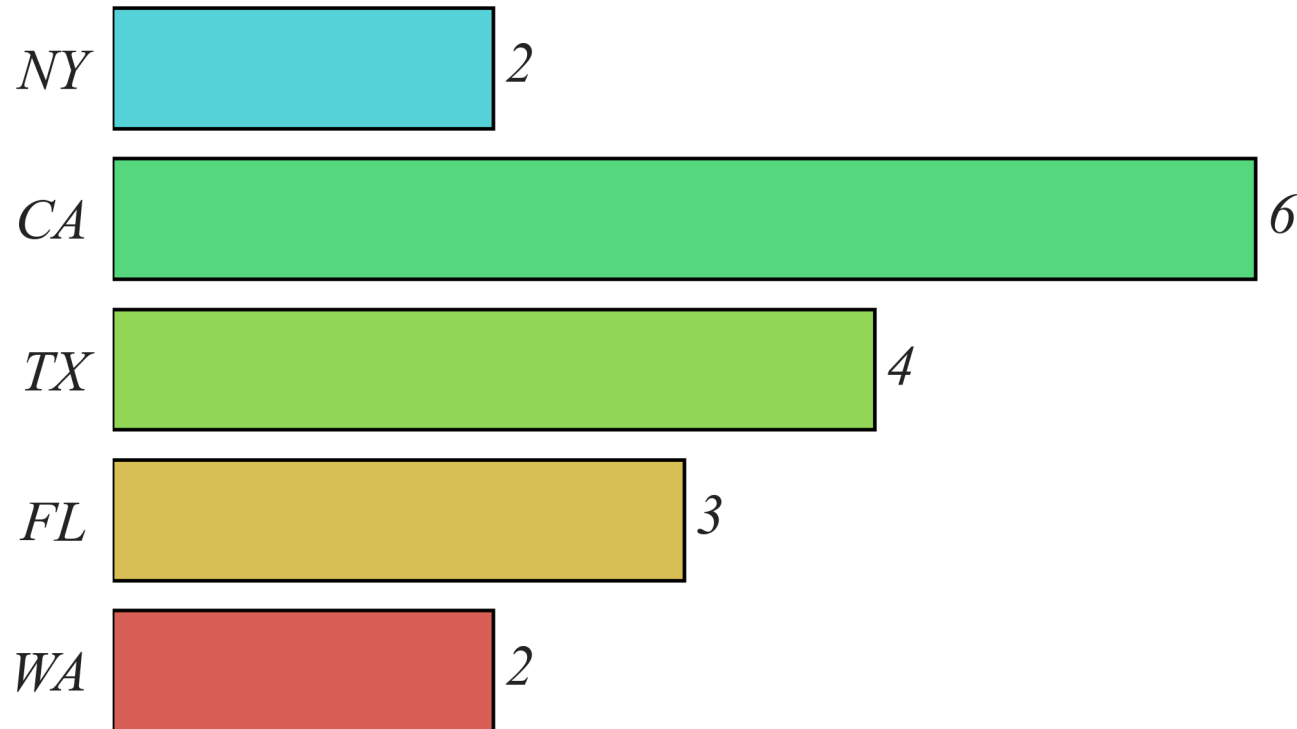
Q. Which state has the second most locations?



> removing clutter also guides your eye to the important information

Catagorical Variables: Remove Clutter

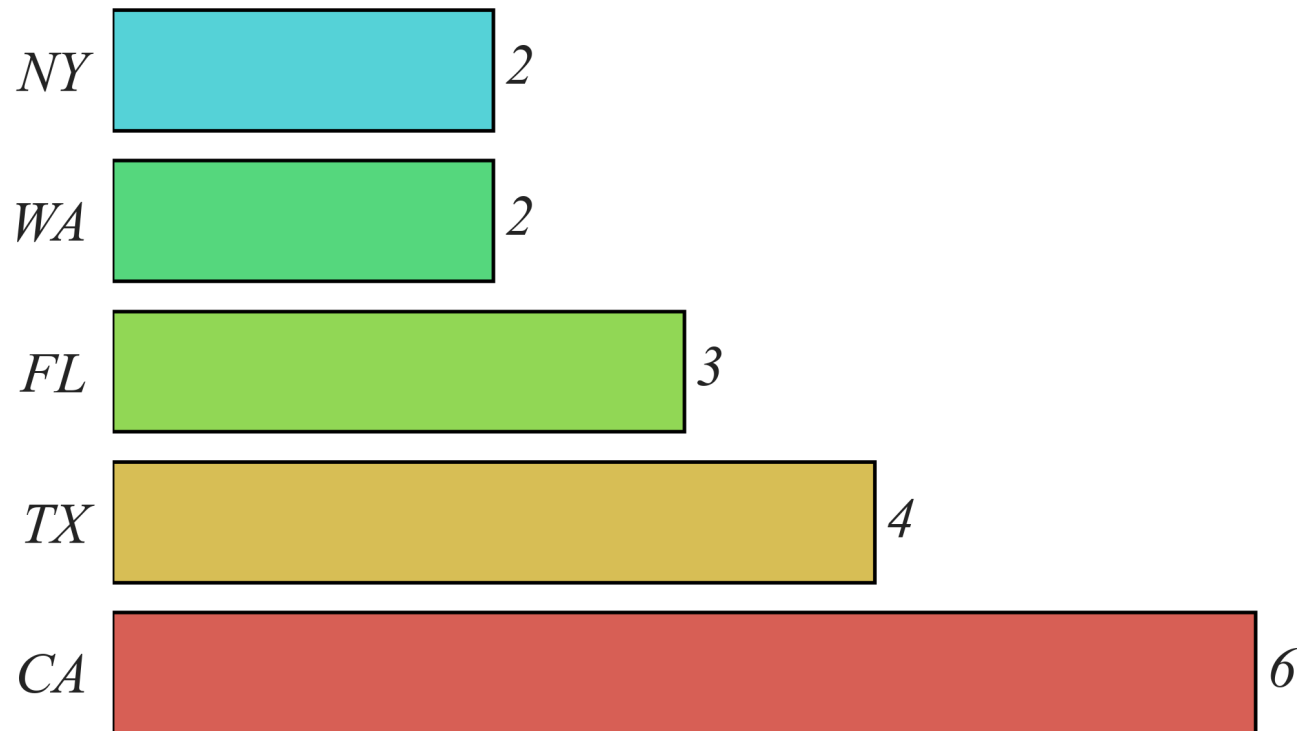
Q. Which state has the second most locations?



> removing clutter also guides your eye to the important information

Catagorical Variables: Order by Size

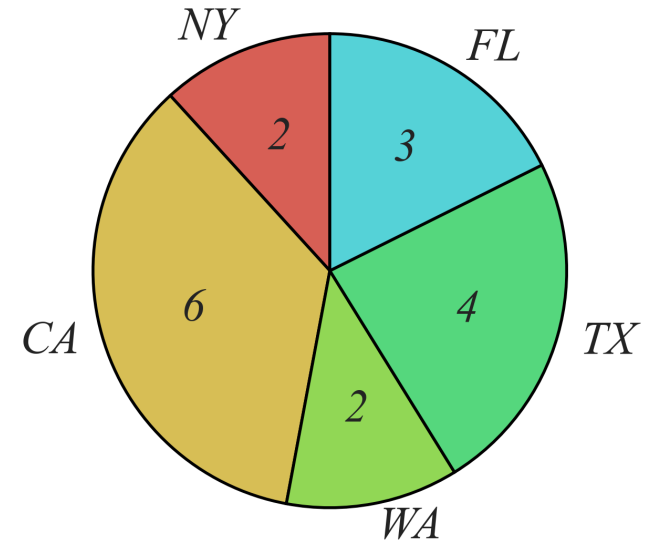
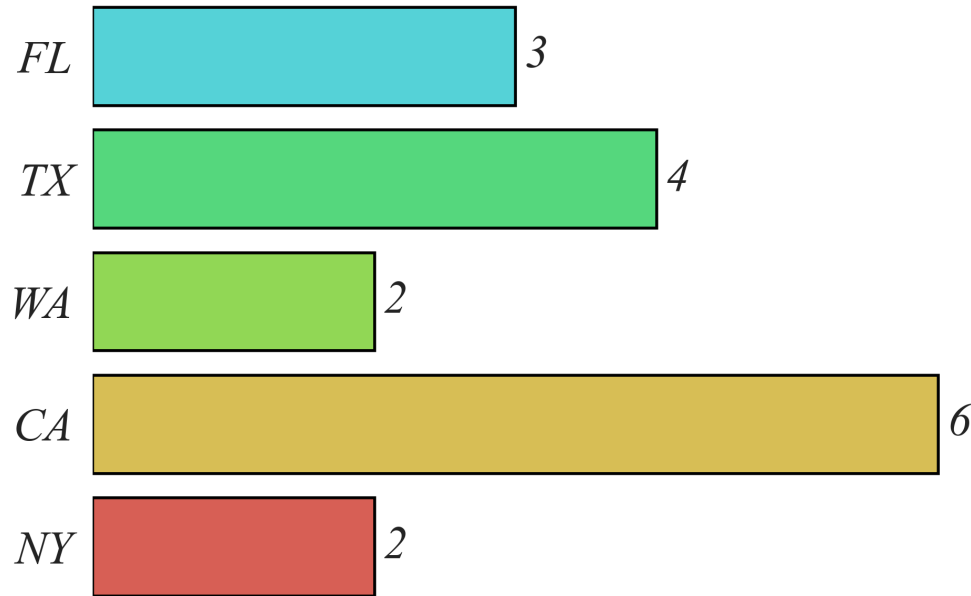
Q. Which state has the second most locations?



> states have no inherent order, but sorting can make comparisons easier

Example: Coffee Shops by State

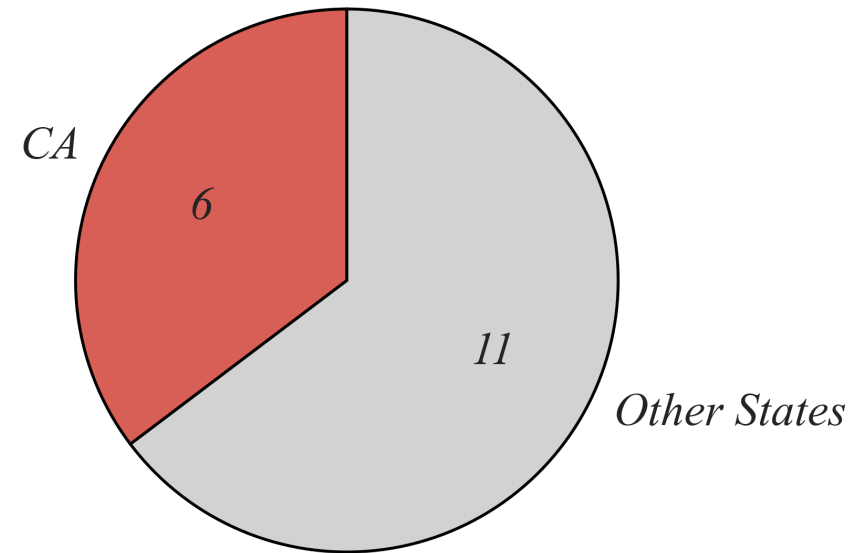
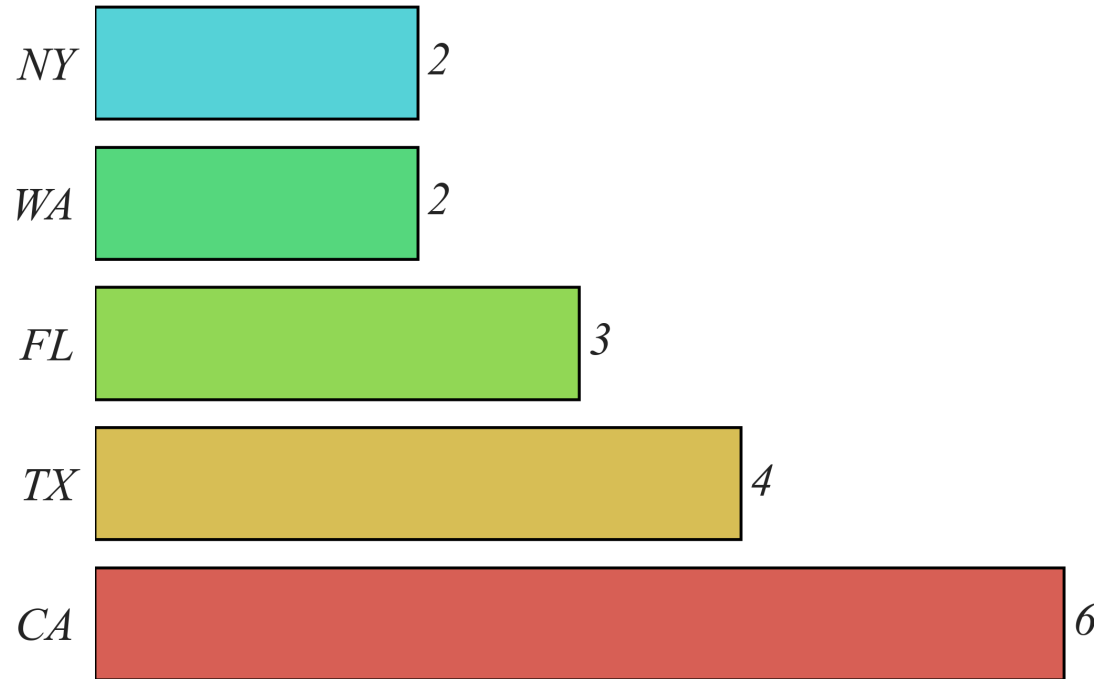
Q. How does CA compare to the whole?



> instead of a nominal categorical variable, this is a binary categorical variable (CA / Other)

Example: Coffee Shops by State

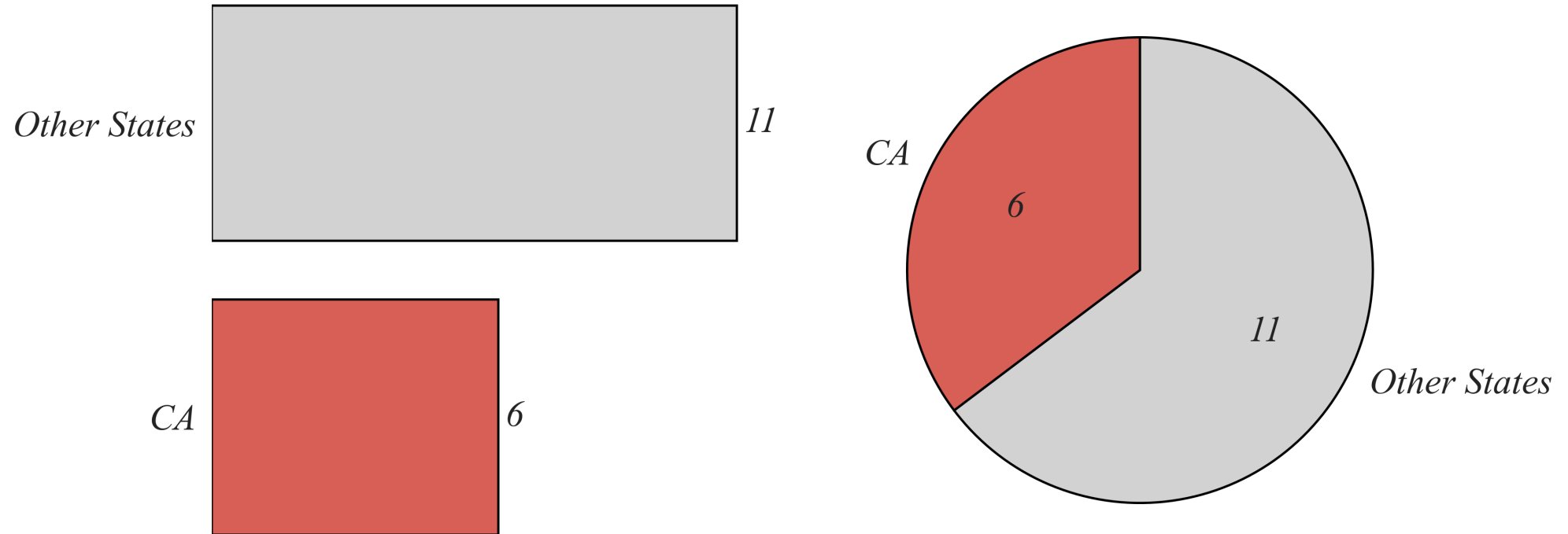
Q. How does CA compare to the whole?



> this question is much easier to see when visualizing as a binary categorical variable

Example: Coffee Shops by State

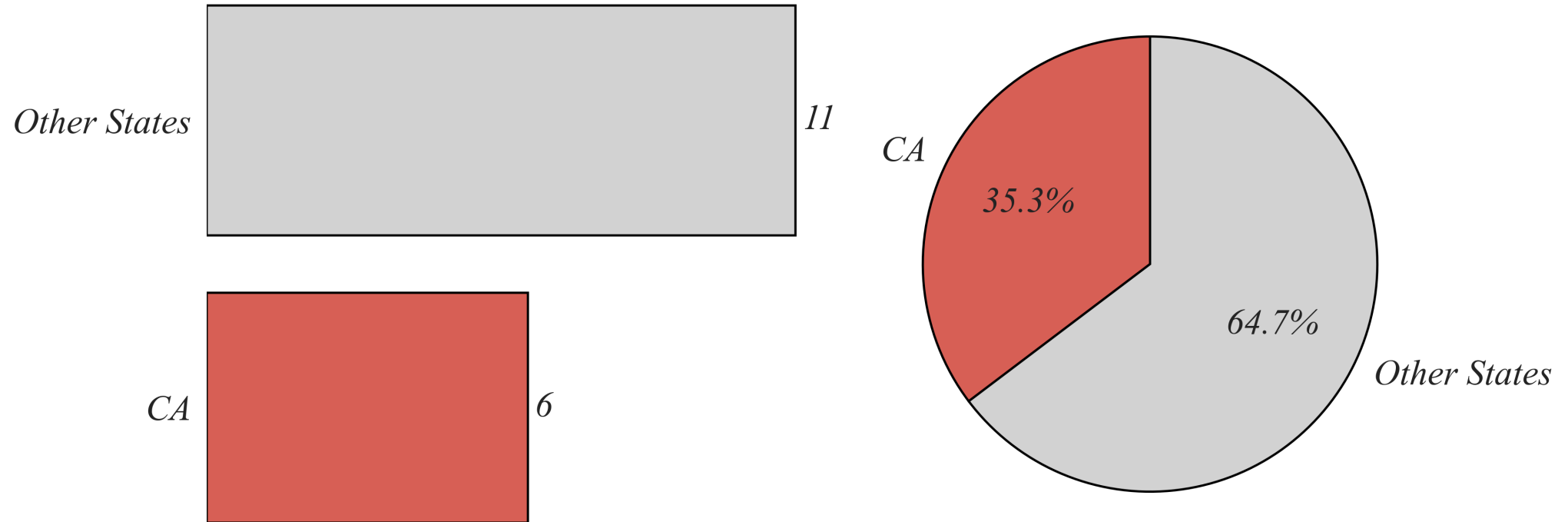
Q. How does CA compare to the whole?



> *here both the pie and the bar communicate the data effectively*

Example: Coffee Shops by State

Q. How does CA compare to the whole?



> but if the question is about shares of a whole, percentages with a pie chart may work best

Takeaways

... use the right summary tool for the variable type

- *Binary Categorical Variables: use a **pie chart** or **bar graph***
- *Nominal (unordered) Categorical Variables: use a **bar graph**; maybe order by value*
- *Ordinal Categorical Variables: use an **ordered bar graph***
- *Remove clutter; keep it simple*
- *Place information near their object*

Exercise 1.1 | Categorical Variables

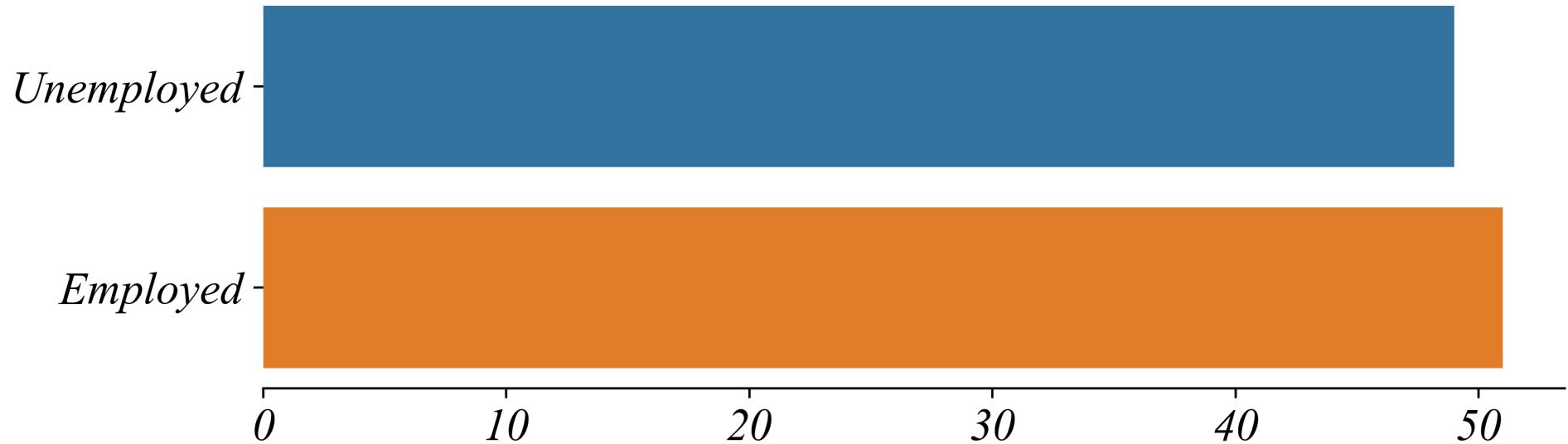
Lets visualize the data in each dataset.

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

Exercise: Dataset 1

Summarize *employment_status.csv*.

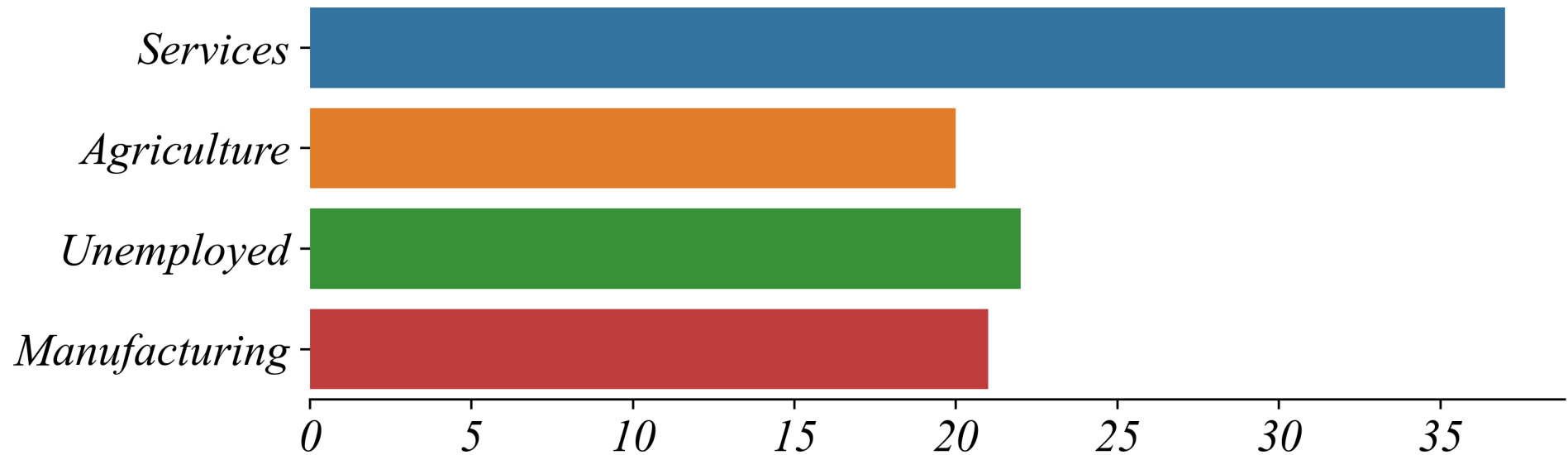
Employment Status



Exercise: Dataset 2

Summarize [household_savings.csv](#).

Employment by Sector



Exercise: Dataset 3

Summarize [household_incomes.csv](#).

