

# COR 142 C: Section 2.4

Dr. Chris Kao

# Summarizing Categorical Distributions

Since categorical data has **no inherent order**, the measure of **shape, center, horizontal spread** do not appropriately describe categorical data.

To describe categorical data, we use \* **Mode**, which is the most frequent observed value. \* **Variability**, which describes the diversity of different observed values.

# Mode

- The category or the observed value that **occurs most frequently**.
- Modes for numerical and categorical data:
  - Numerical: Peaks do not need to be the same height.
  - Categorical: Modes must be **roughly the same height**.

# Variability

- Diversity in the observed data values.
- High variation: Lots of diversity in observed values.
- Low variation: Low diversity in observed values.

## Note

Variability in a categorical distribution is more about the **occurrence of many distinct observed values** rather than the frequency of each observed value.