

# Tables

## Categorical Numerical Summaries

Grinnell College

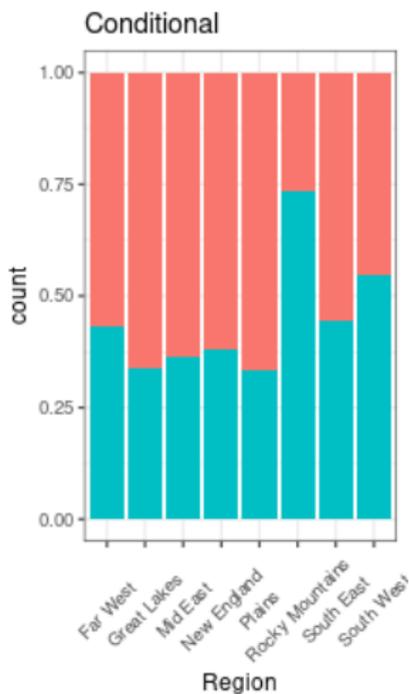
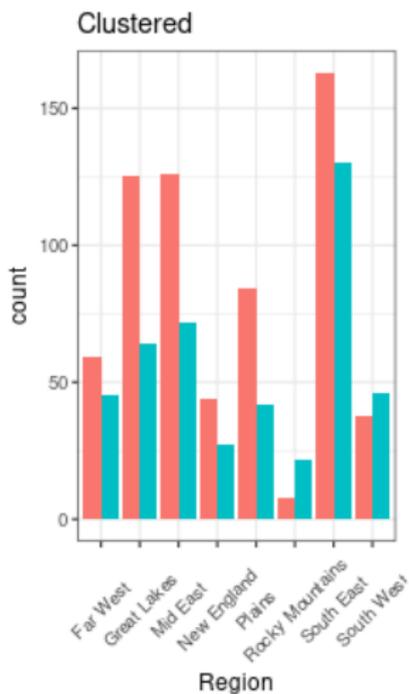
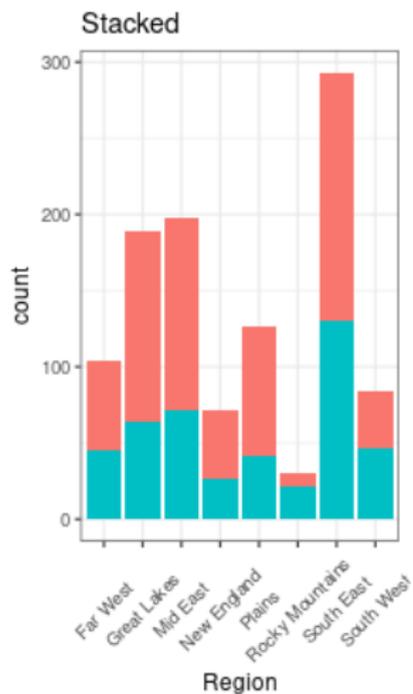
Last few sessions:

- ▶ Describing quantitative variables (histograms and boxplots)
  - ▶ center
  - ▶ spread
  
- ▶ Order statistics (skewed, outliers)
  - ▶ median
  - ▶ IQR
  
- ▶ Moment statistics (symmetric, no outliers)
  - ▶ mean
  - ▶ standard deviation

# Today's Outline

- ▶ (Review-ish) Different ways to represent multiple categorical variables using bar charts
- ▶ Using tables to organize data
- ▶ Relating tables and barcharts to each other
- ▶ "Conditional Statistics"

# Bar Charts



Type ■ Private ■ Public

Other names: clustered = dodge, conditional = proportional

# Descriptive Statistics – Categorical Variables

Information for categorical variables are often presented in *tables*

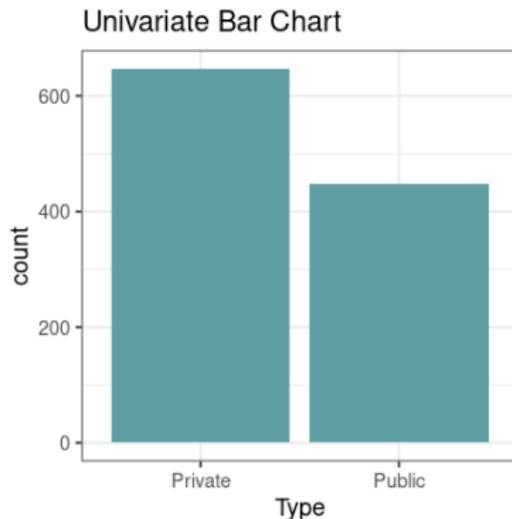
- ▶ **Frequencies:** counts how many of each case belongs to a particular category
- ▶ **Proportions:** fractions based upon frequencies, also called *relative frequencies*

Frequency table:

	Frequency
Private	647
Public	448

Table of proportions:

	Proportion
Private	0.591
Public	0.409



# Bivariate Bar Charts

Just as we did when looking at graphical summaries, we tend to designate variables as being either *explanatory* or *response* variables

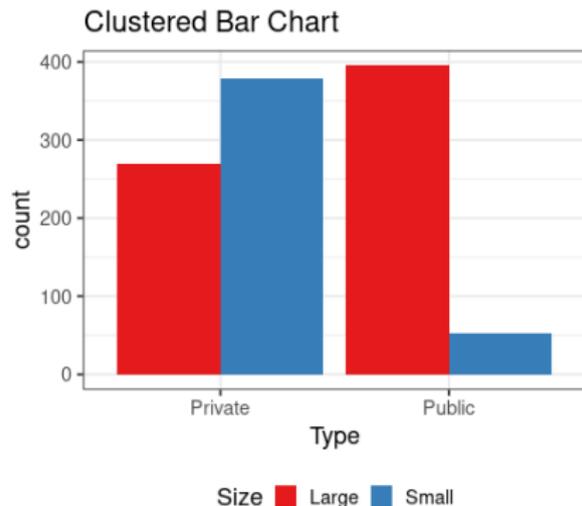
Again, this is **not** causal

We tend to think of these relationships *conditionally* when discussing categorical variables, which means that we focus on combinations of the various categories

# Descriptive Statistics – Categorical Variables

Two-way frequency table:

	Small	Large
Private	378	269
Public	53	395

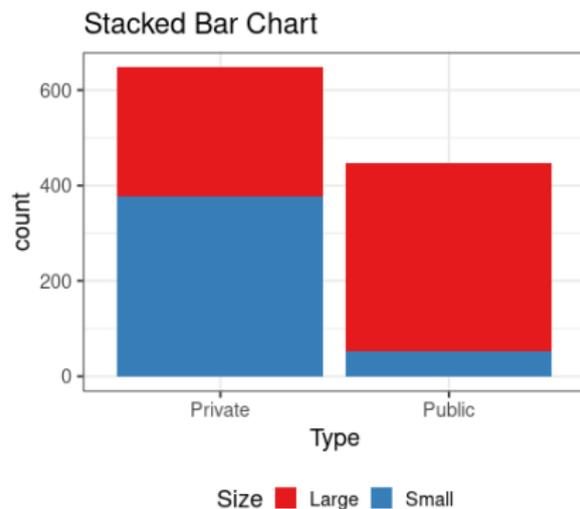


The counts in a table like this correspond to a dodged/clustered bar chart

# Descriptive Statistics – Categorical Variables

Often these tables include margin sums as well

	Small	Large	Sum
Private	378	269	647
Public	53	395	448
Sum	431	664	1095



## Descriptive Statistics – Categorical Variables

Two-way table of proportions (overall proportions)

	Small	Large
Private	0.3452	0.2457
Public	0.0484	0.3607

*“36% of all schools are large public schools”*

# Conditional Statistics

A **conditional statistic** is a statistic derived from one or more variables for all observations sharing a value of another variable

- ▶ “What is the predicted weight of an individual *given* that they are 6ft tall”
- ▶ “What is the proportion of public schools *given* that we are looking at the Plains region” → only look at Plains colleges

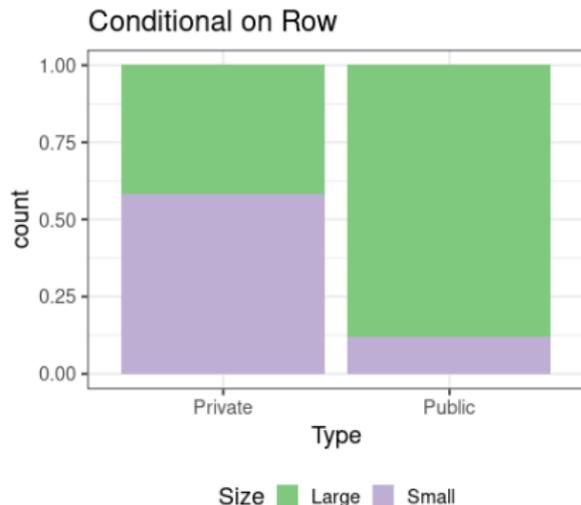
Note that we typically condition on the *explanatory* variable

## Descriptive Statistics – Row Proportions

*“88% of public schools are considered large”*

*“Given that a school is a public school, 88% of them are considered large”*

	Small	Large
Private	0.5842	0.4158
Public	0.1183	0.8817

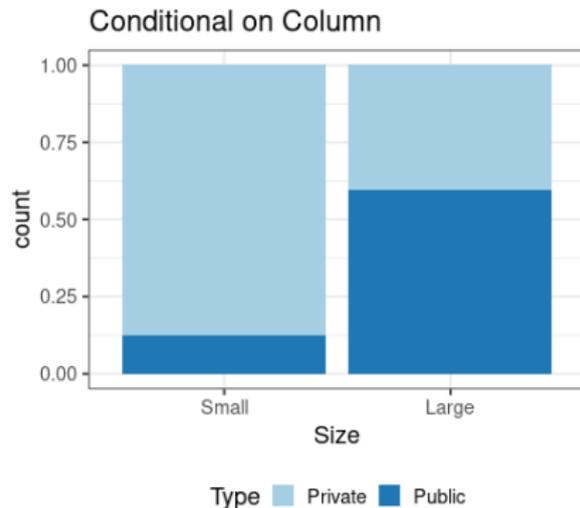


To find out whether rows/columns are conditioned on, see which adds up to 1.00

# Descriptive Statistics – Column Proportions

*“12% of small colleges are public”*

	Small	Large
Private	0.8770	0.4051
Public	0.1230	0.5949



## Example

The two-way table below describes the survival of crew members and first class passengers aboard the Titanic

	Survived	Died
Crew	212	673
First Class	203	122

1. Given that an individual survived, is it more likely that they were a crew member or a passenger in first class?
2. Given that an individual was a crew member, is it more likely that they survived or died?
3. Which group was more likely to survive the shipwreck?

# Summary

- ▶ Types of charts
  - ▶ Stacked
  - ▶ Clustered
  - ▶ Conditional
- ▶ Types of Tables
  - ▶ One and two-way tables
  - ▶ Frequency and proportions
  - ▶ Tables match with specific barcharts
- ▶ Association for categorical variables
  - ▶ See how 'conditional' percentages look