Applied Statistical Analysis

EDUC 6050

Week 2

Finding clarity using data

Today

1. Working with Data
2. Overview of Statistics
3. Intro to Statistical Terminology
4. Intro to Jamovi

Reading

Data in Spreadsheets

What did you like? Not like?
Things you thought were useful? Confusing?

Data in Spreadsheets

2 Be Consistent
3 Choose good names for things
4 Write dates as YYYY-MM-DD
6 Put just one thing in a cell
7 Make it a rectangle
8 Create a data dictionary

Why Learn Statistics?

It is the language of understanding data

- Allows you to complete your thesis!
- Helps you communicate with other data people you work with
- Gives you power to convince stakeholders with evidence
- Opens up job opportunities

Data and Statistics

Statistics helps us understand our data

Summarize the data easily

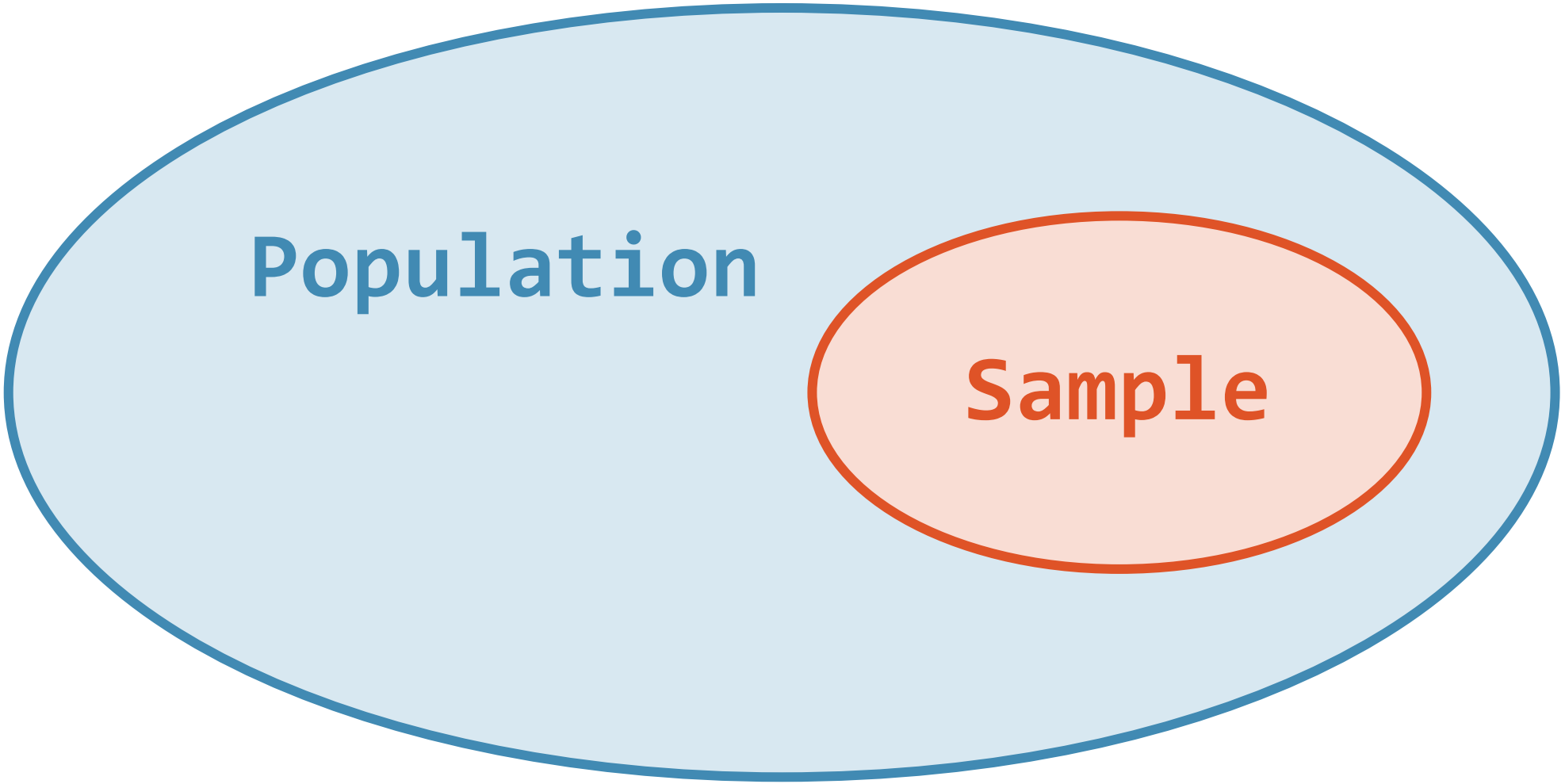
Ask questions about what the data mean

Statistics

A **statistic** is some sort of summary of the data

- The **average** is a statistic
- A **frequency** (**count**) is a statistic

The Vocabulary of Statistics



The Vocabulary of Statistics

Descriptive Statistics

Describing the data that you have (your sample)

Inferential Statistics

Understanding what your data say about the population

The Vocabulary of Statistics

Independent Variables

“predictors” or “IV”

These are the variables that we think are causing or influencing the outcome

Dependent Variables

“outcomes” or “DV”

These are the variables that we think are caused by an independent variable

The Vocabulary of Statistics

Hypothesis Testing (Inferential Statistics)

"Null Hypothesis Significance Testing"

Gives us an idea about what the population may look like based on our sample (accounts for **sampling error**) = "significance"

The Vocabulary of Statistics

Hypothesis Testing (Inferential Statistics)

"Null Hypothesis Significance Testing"

Effect Sizes

"Magnitude of the effect"

Tells us how big the effect is = "meaningfulness"

Scales of Measurement

"The way a variable is measured determines the kinds of statistical procedures that can be used" (pg 10)

Want measures that:

1. Are reliable
2. Are valid
3. Are meaningful
4. Have a high degree of information

Scales of Measurement

4 General Types (see pg. 11)

Scale	*Definition*	*What the scale allows you to do*
Nominal	Categories based on qualitative similarity (no order to the categories)	Count the number of things in the categories
Ordinal	Like nominal, but the categories can be ranked	Count and rank the number of things in each category
Interval	Quantify how much of something	Count, rank, and quantify how much of something there is (zero does not mean there's nothing)
Ratio	Quantify how much of something (zero means there is none of that thing)	Count, rank, and quantify how much of something there is with a meaningful zero

Scales of Measurement

4 General Types (see pg. 11)

Scale	*Definition*	*What the scale allows you to do*
Nominal	Categories based on qualitative similarity (no order to the categories)	Count the number of things in the categories
Ordinal	Like nominal, but the categories can be ranked	Count and rank the number of things in each category
Interval	Quantify how much of something	Count, rank, and quantify how much of something there is (zero does not mean there's nothing)
Ratio	Quantify how much of something (zero means there is none of that thing)	Count, rank, and quantify how much of something there is with a meaningful zero

Scales of Measurement

4 General Types (see pg. 11)

Scale	*Definition*	*What the scale allows you to do*
Nominal	Categories based on qualitative similarity (no order to the categories)	Count the number of things in the categories
Ordinal	Like nominal, but the categories can be ranked	Count and rank the number of things in each category
Interval	Quantify how much of something	Count, rank, and quantify how much of something there is (zero does not mean there's nothing)
Ratio	Quantify how much of something (zero means there is none of that thing)	Count, rank, and quantify how much of something there is with a meaningful zero

Increasing degree of information

Scales of Measurement

4 General Types (see pg. 11)

Scale	*Definition*	*What the scale allows you to do*
Nominal	Categories based on qualitative similarity (no order to the categories)	Count the number of things in the categories
Ordinal	Like nominal, but the categories can be ordered	Count and rank the number of things in the categories
Interval	Quantify how much of something	Count, rank, and quantify how much of something there is (zero does not mean there is nothing)
Ratio	Quantify how much of something (zero means there is none of that thing)	Count, rank, and quantify how much of something there is with a meaningful zero

Team Challenge:
What are some examples of each type?

Scales of Measurement

These lie on a spectrum from qualitative to quantitative

Scales of Measurement

Discrete

Cannot be broken down into smaller units

Number of siblings, racial groups, have the disease or not

Continuous

Can be broken into smaller units

Time to finish an exam, height of a person

Break Time

Graphing Data

A VERY IMPORTANT part of data analysis

It is useful for both:

1. **Understanding patterns** in the data
2. **Communicating** results in a much more meaningful way

Takes some practice

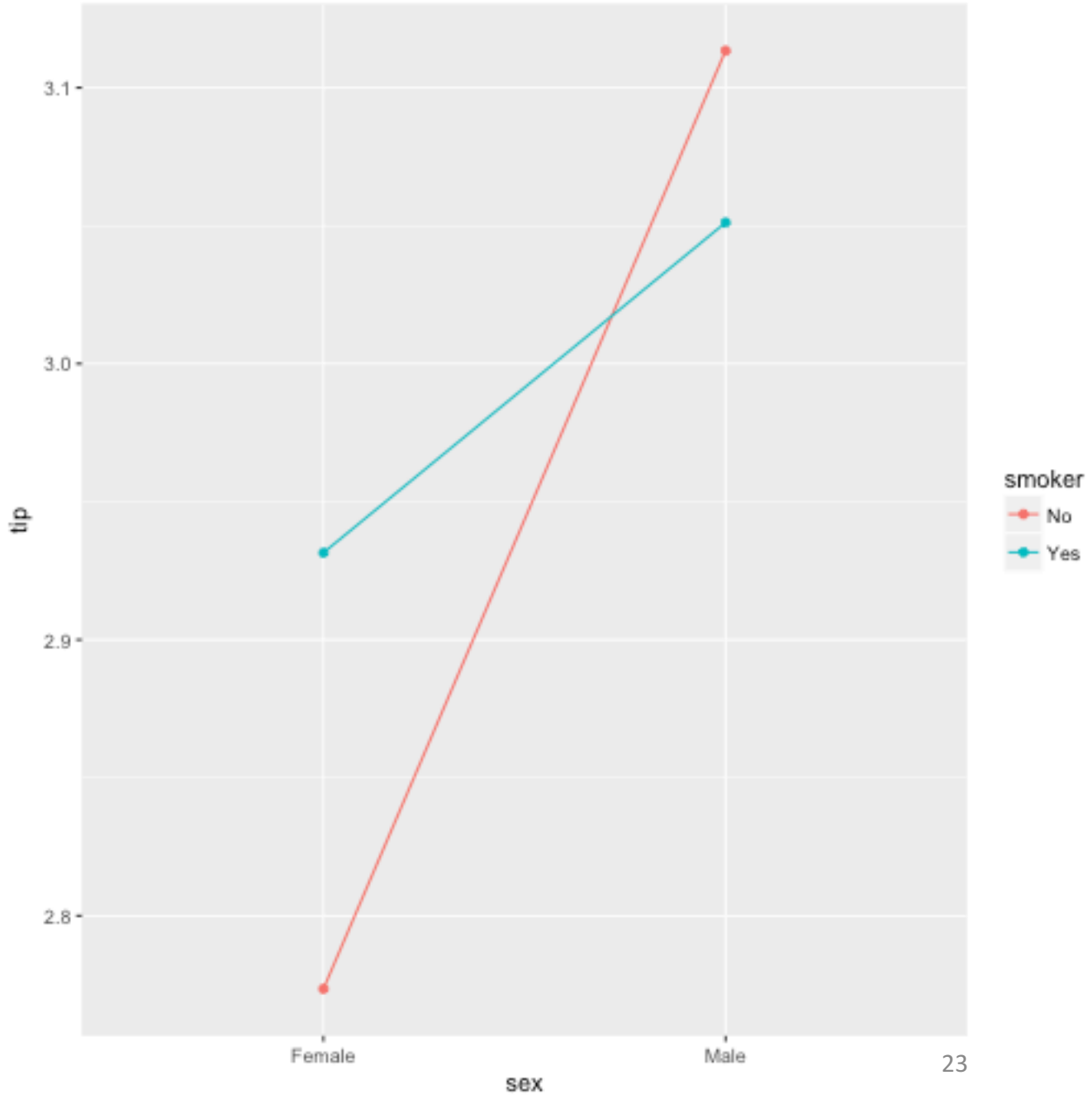
Some Types of Data Graphics

Each provide different insights into the data

1. Line Graphs
2. Bar Graphs and Histograms
3. Scatterplots
4. Boxplots

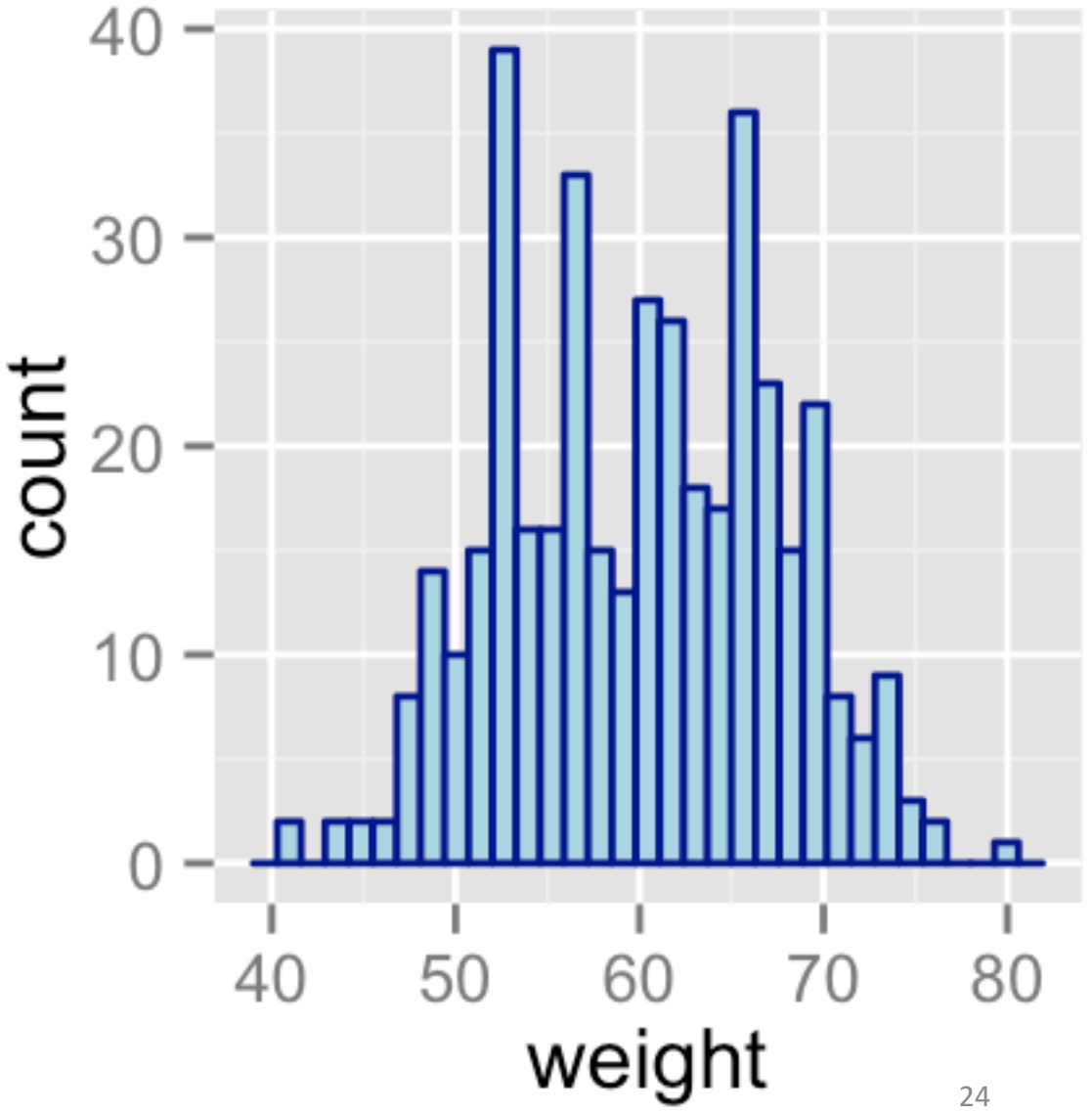
Line Graphs

Generally shows trends and patterns across groups



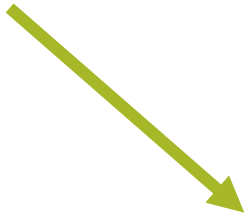
Bar Graphs and Histograms

These help us understand distributions and frequencies



Bar Graphs and Histograms

These help us understand distributions and frequencies

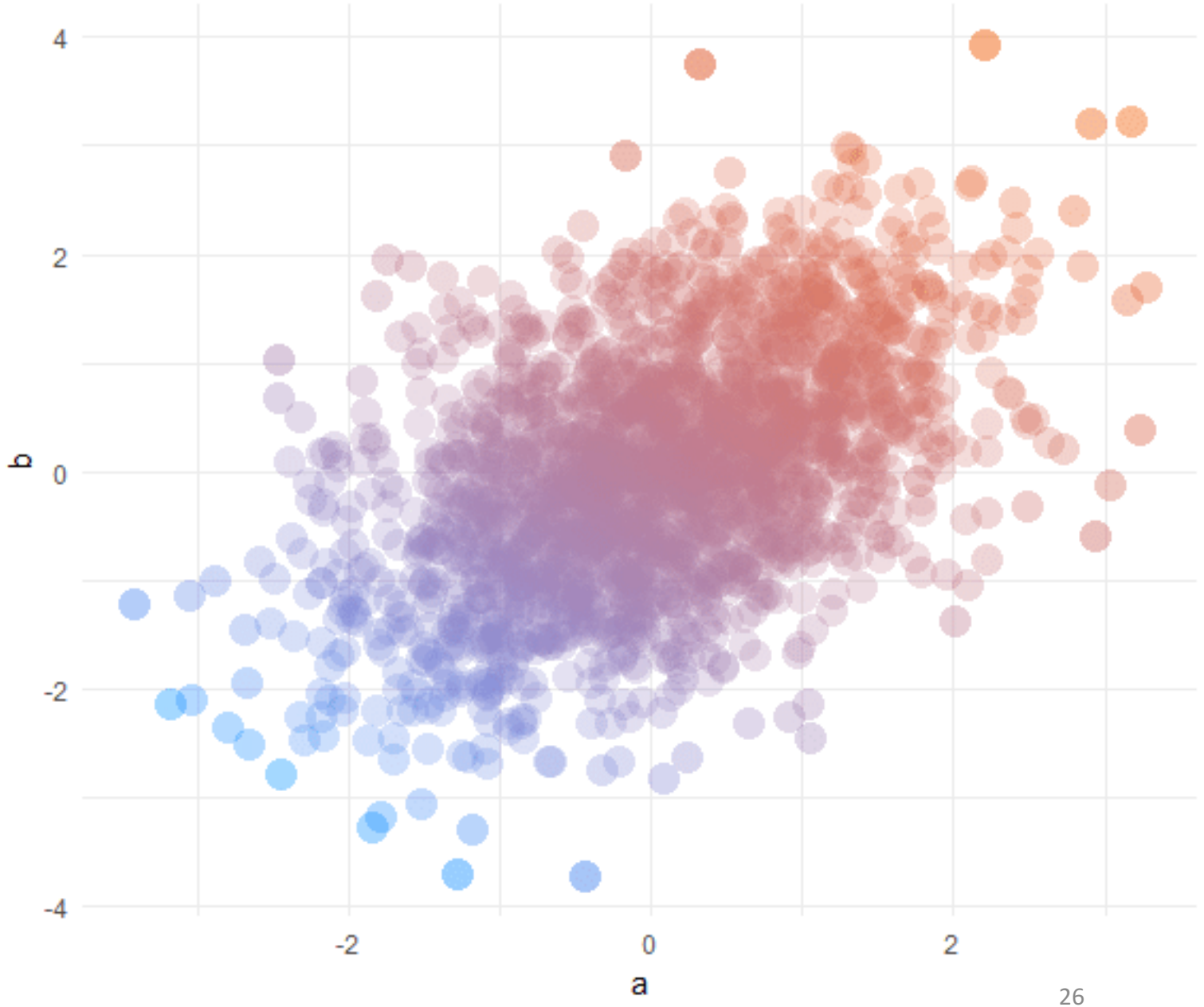


Skew
Kurtosis

Symmetric vs. Asymmetric
Unimodal vs. Multimodal
Short-tailed vs. long-tailed

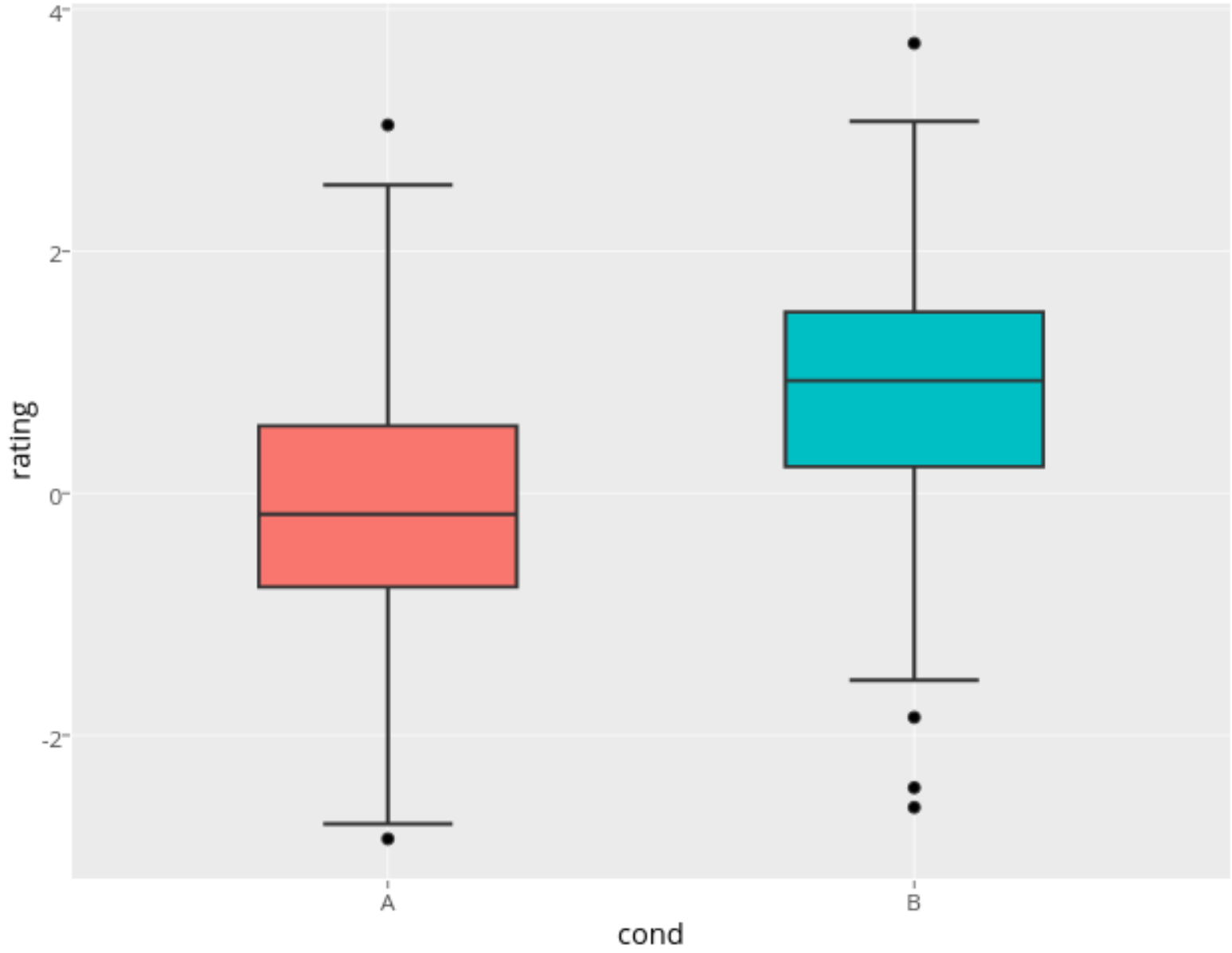
Scatterplots

Show us how two (or more) variables are related



Boxplots

Show us the range and where most values are for a variable (usually across groups)



Frequency Tables

Tables can also be very valuable to understand patterns in the data

Level	Frequency	Percent	Cumulative Percent
A	10	25.0%	25.0%
B	5	12.5%	37.5%
C	20	50.0%	87.5%
D	5	12.5%	100%

Frequency Tables

Tables can also be very valuable to understand patterns in the data

Level	Frequency
A	10
B	5
C	20
D	5

What plot could be used to show this information?

Review

1. Name one thing you liked from Broman et al.
2. What is a statistic?
3. What is the difference between a population and a sample?
4. True or False. Independent variables are also known as outcomes.
5. Which contain more information: ordinal or ratio variables?

Review

6. What information does a boxplot give us?
7. What about a scatterplot?
8. What is the difference between a bar graph and a histogram?
9. Graph the data from the table:

Score	Frequency
1	0
2	3
3	2
4	5
5	8
6	6
7	3
8	1
9	6
10	8

Let’s jump into Jamovi

Questions?

Next week:

1. Statistics terminology (Hypothesis, IV and DV, Measurement, Validity and Reliability, Correlation and Experimentation, Distributions, Central Tendency and Variability)
2. Chapters 2 and 3 in Book
3. Start looking for articles