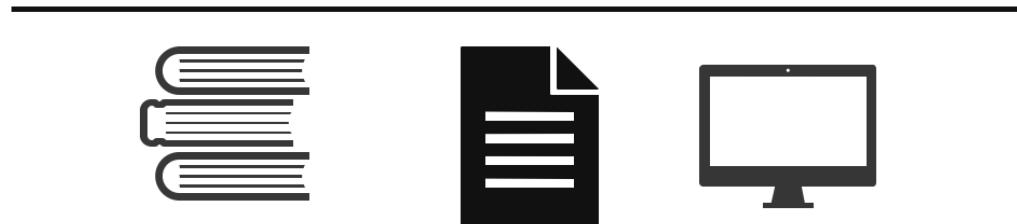
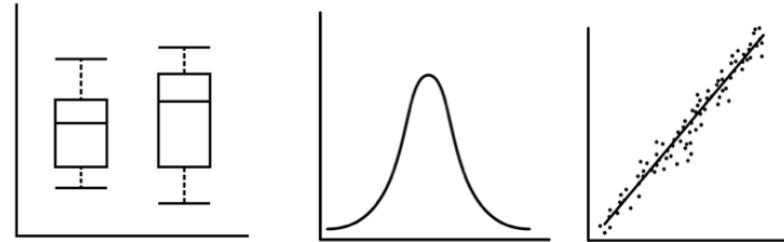


PSYC 2300

Introduction to Statistics



Lecture 03: Visualizing Data Using Graphs

Updates: Course Calendar

Date	Week	Class	Topic	Reading	What's due
1/3/22	1	1	Introduction. Why study statistics in Psychology?	Syllabus, Ch.1	
1/5/22		2	Measures of central tendency and variability	Ch.2, Ch.3	
1/10/22	2	3	Visualizing data using graphs	Ch.4	
1/12/22		4	Correlations, Reliability, and Validity	Ch.5, Ch.6	
1/17/22	3	5	NO CLASS; MLK JR DAY		
1/19/22		6	Using hypotheses to test questions	Ch.7	Quiz 1 (ch.1-5)
1/24/22	4	7	Probability and normal curve	Ch.8	
1/26/22		8	Introduction to statistical significance	Ch.9	Quiz 2 (ch.6-8)
1/31/22	5	9	Midterm review		
2/2/22		10	Midterm Exam		Mini-report
2/7/22	6	11	Differences from the population	Ch.10	
2/9/22		12	Differences between two groups I	Ch.11	Quiz 3 (ch.9-10)
2/14/22	7	13	Difference between two groups II	Ch.12	
2/16/22		14	Difference between many groups	Ch.13	Quiz 4 (ch.11-12)
2/21/22	8	15	Difference between many factors	Ch.14	
2/23/22		16	Testing relationships using correlations	Ch.15	Quiz 5 (ch.13-14)
2/28/22	9	17	Making predictions using regression	Ch.16	
3/2/22		18	Analyzing data using JASP I		Quiz 6 (ch.15-16)
3/7/22	10	19	Analyzing data using JASP II		
3/9/22		20	Final Review		Application Project
3/14/22	11	21	Final Exam		

Announcements: Guides

- *Guides* posted on Canvas
- As we learn a new calculation, I will post a new *Guide*

The screenshot shows a PDF document titled "guide-01-standard-deviation.pdf" (Page 1 of 3). The document is titled "PSYC 2300: Introduction to Statistics" and "Guide 01: Standard Deviation". It is a guide for calculating sample standard deviation. It starts by stating: "In this document, I provide a complete solution for calculating the standard deviation of a sample of data. To begin, let's start with some data: 7, 9, 8, 4, 3, 3, 5, 6, 7, 8". It then explains the formula for sample standard deviation: $s = \sqrt{\frac{\sum(X_i - \bar{x})^2}{n-1}}$. It provides two steps: "Step 1: Find the mean" and "Step 2: Calculate deviations from the mean".

PSYC 2300: Introduction to Statistics Guide 01: Standard Deviation

Calculating Sample Standard Deviation

In this document, I provide a complete solution for calculating the standard deviation of a sample of data. To begin, let's start with some data:

7, 9, 8, 4, 3, 3, 5, 6, 7, 8

Recall the formula for sample standard deviation. Make sure to note the differences between the formulas for sample standard deviation, population standard deviation, sample variance, and population variance; these formulas are all similar and involve similar steps to compute, but they have slight differences in the final steps that will result in different answers!

Sample Standard Deviation

$$s = \sqrt{\frac{\sum(X_i - \bar{x})^2}{n-1}}$$

Step 1: Find the mean

You can do this by using, for example, $\bar{x} = \frac{\sum x}{n}$. Put simply, you just add up all the values and divide by the number of values you added up. For this dataset, the mean = 6. Now that you have your mean, you can begin building a table to help solve this problem.

$$\bar{x} = \frac{\sum x}{n} = \frac{7 + 9 + 8 + 4 + 3 + 3 + 5 + 6 + 7 + 8}{10} = 6$$

Step 2: Calculate deviations from the mean

You can do this by subtracting the mean you calculated above from each value. A way of representing what you're doing mathematically is as follows: $x_i - \bar{x}$.

3

Outline for Today

- Review parts of last class
- Basics of data visualization
- Variability practice problem
 - Have  and  ready
- Descriptive statistics in JASP
 - Have JASP installed and open
 - Download "[Stats Class 3 Dataset \(Pets\).csv](#)" from Canvas Week 02



Review Parts of Last Class

Measures of Central Tendency

Measures of central tendency: Numbers that represent the *center* or *middle* of a distribution of data

Mean

Median

Mode

Scales of Measurement

Mean: The sum of a set of scores divided by the total number of scores in the set

Median: The point at which half (50%) of the values are above and half (50%) of the values are below

Mode: The value in a distribution of data that occurs most frequently

Scales of Measurement

Scales of measurement: describes the nature of the information contained in a given set of data

Nominal

Ordinal

Interval

Ratio

Scales of Measurement

Nominal

- Non-numerical
- Each item in the set belongs to a class or category

Ordinal

- Items are *ordered* in a meaningful direction
- Distance between items is *not necessarily* equal

Interval

- Distance between points is *equal* and *meaningful*
- But relationship between points is *not* meaningful

Ratio

- All the qualities of the interval scale plus a *true zero point*
- Relationship between points is *meaningful*

Measures of Variability

In statistics, **measures of variability** describe how scores in a given dataset differ from one another (e.g., the spread or clustering of points)

Range

Standard
Deviation

Variance

The Range

The range is the simplest measure of variability (or dispersion), and is defined as follows:

Range

$$r = h - l$$

h = highest score in the set

l = lowest score in the set

Standard Deviation: Calculation

i	x	$X_i - \mu$	$(X_i - \mu)^2$
1	1	-1.6	2.56
2	2	-0.6	0.36
3	1	-1.6	2.56
4	4	1.4	1.96
5	3	0.4	0.16
6	3	0.4	0.16
7	6	3.4	11.56
8	1	-1.6	2.56
9	2	-0.6	0.36
10	3	0.4	0.16

$$\mu = 2.6$$

$$\sum(X_i - \mu)^2 = 22.4$$

Sum of Squares

$$\sum(X_i - \mu)^2$$

Standard Deviation and Variance

Population Parameter

Standard Deviation

$$\sigma = \sqrt{\frac{\sum(X_i - \mu)^2}{N}}$$

Variance

$$\sigma^2 = \frac{\sum(X_i - \mu)^2}{N}$$

Sample Statistic

Standard Deviation

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{(n-1)}}$$

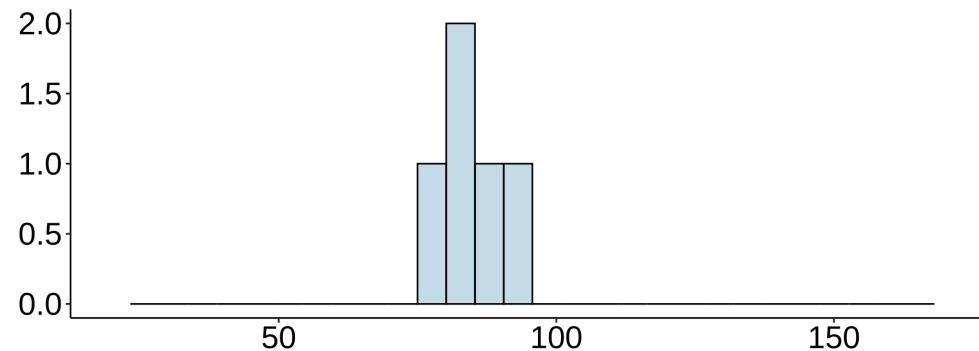
Variance

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{(n-1)}$$

Basics of Data Visualization

Same Mean, Different Variability

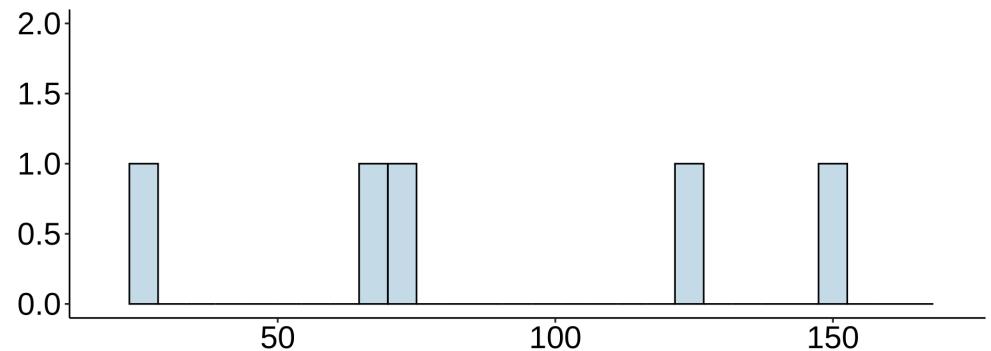
80, 85, 85, 90, 95



$$\bar{x} = 87$$

$$r = 15$$

25, 65, 70, 125, 150

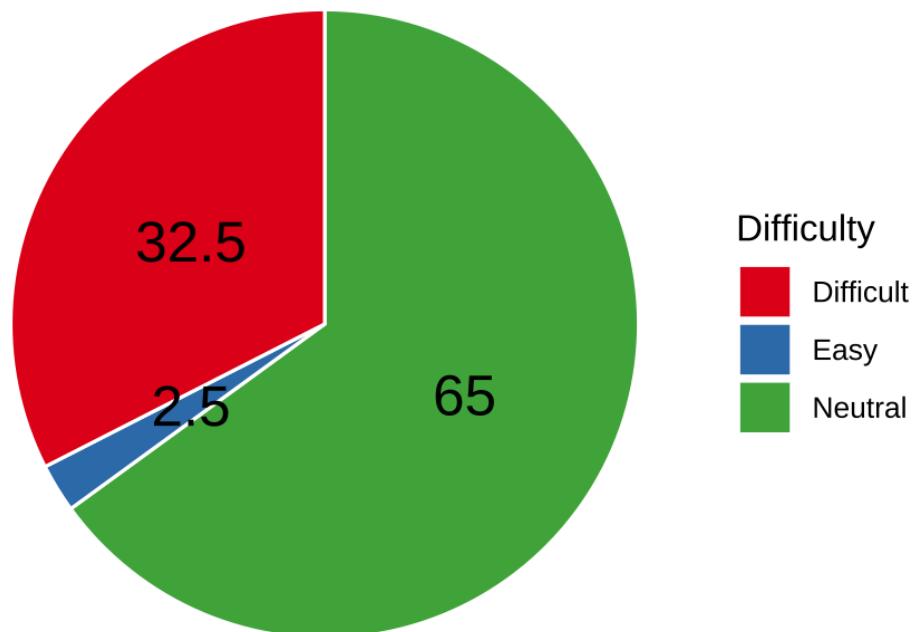


$$\bar{x} = 87$$

$$r = 125$$

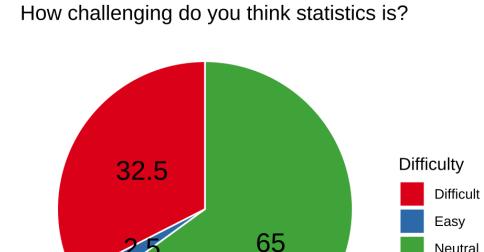
Pie Chart

How challenging do you think statistics is?

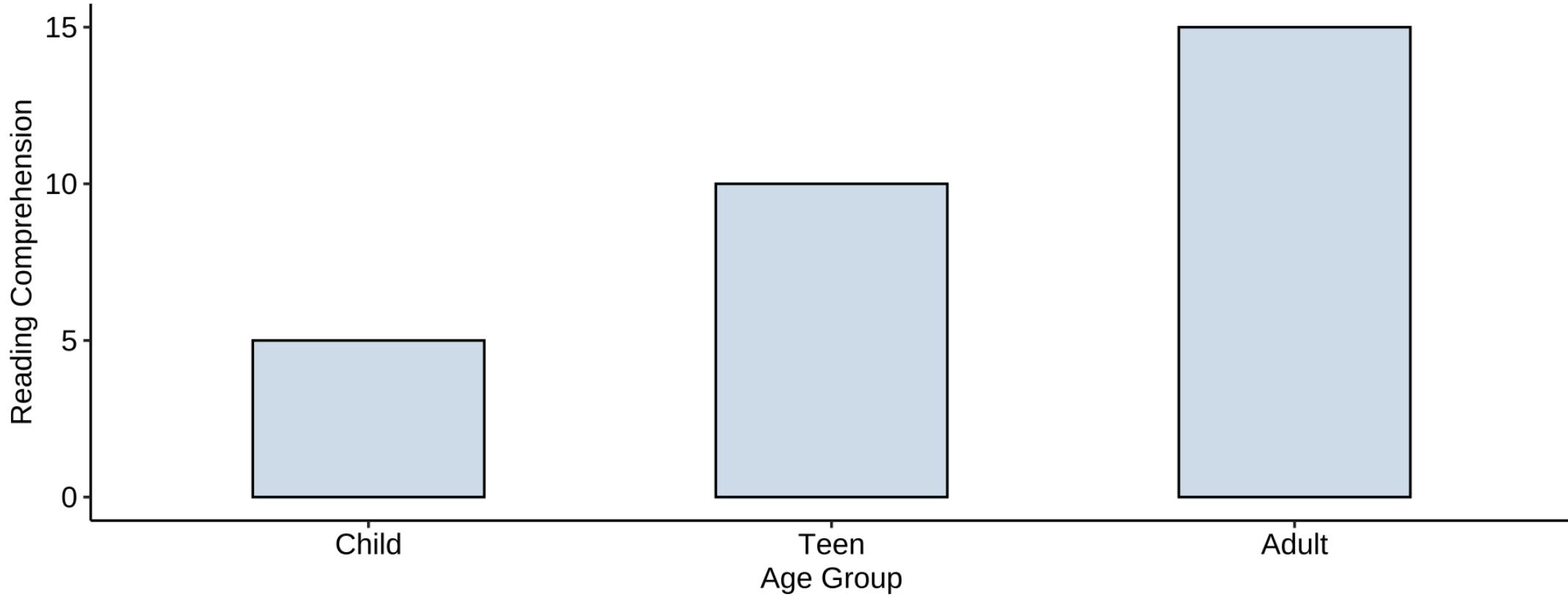


Pie Chart

- Appropriate for **nominal** scale variables
- Appropriate for **ordinal** scale variables
- You can indicate values or percentages for each level of the nominal variable

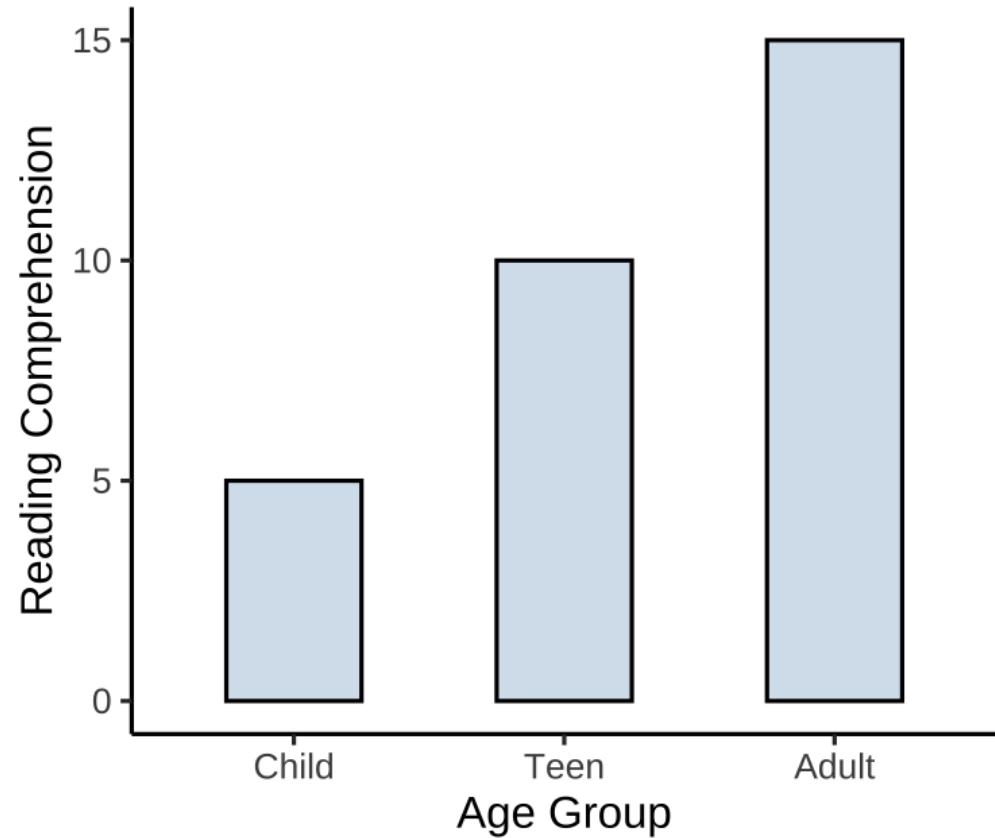


Bar Graph

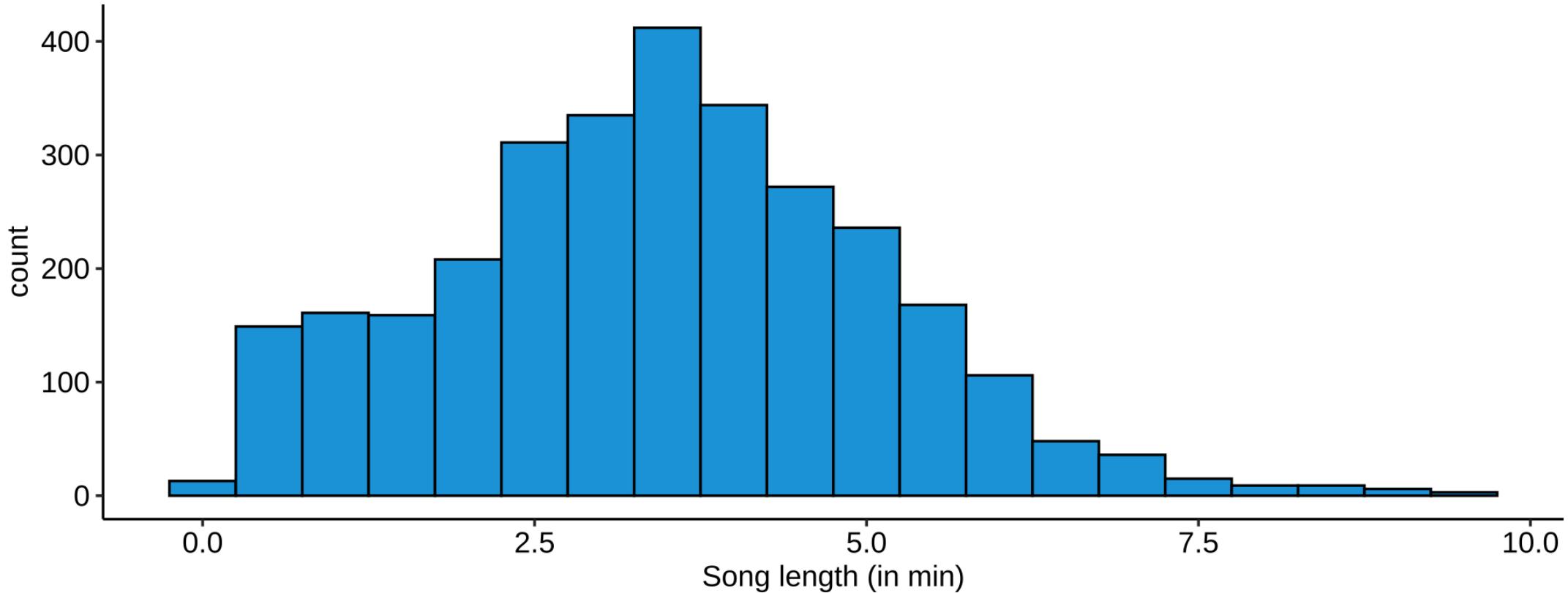


Bar Graph

- Appropriate for **nominal** and **ordinal** data
- Not appropriate for **ratio** (continuous) data
- Simple and visually appealing

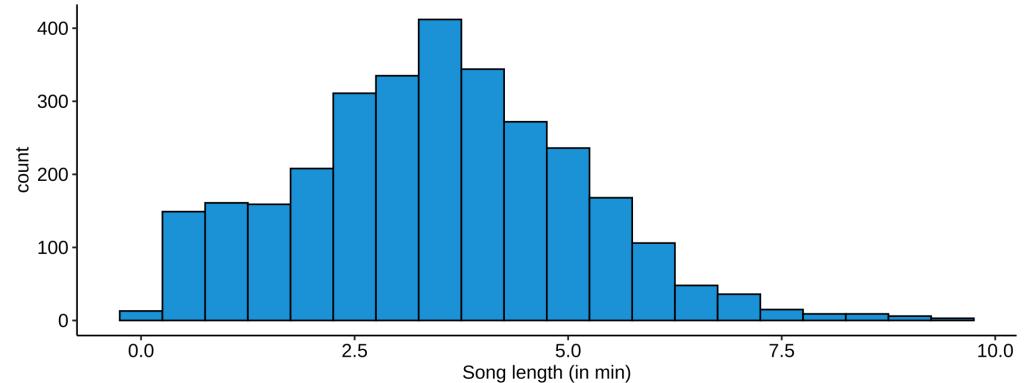


Histogram

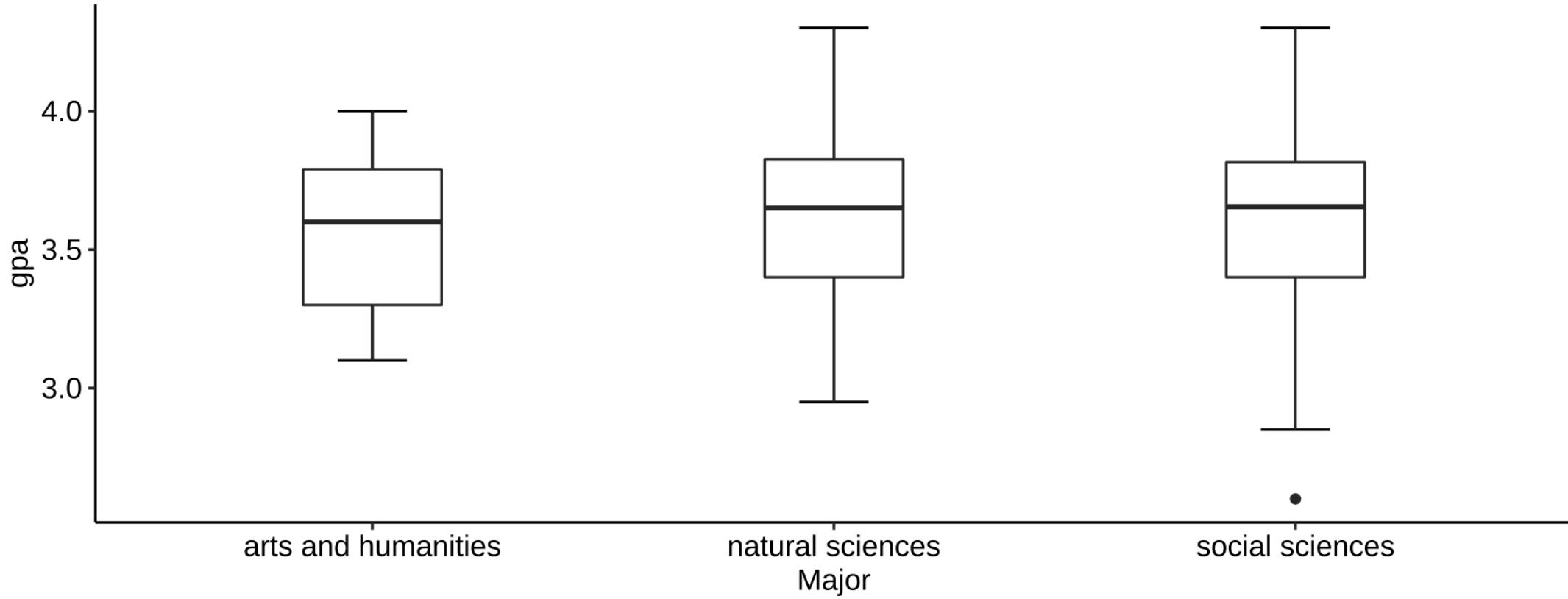


Histogram

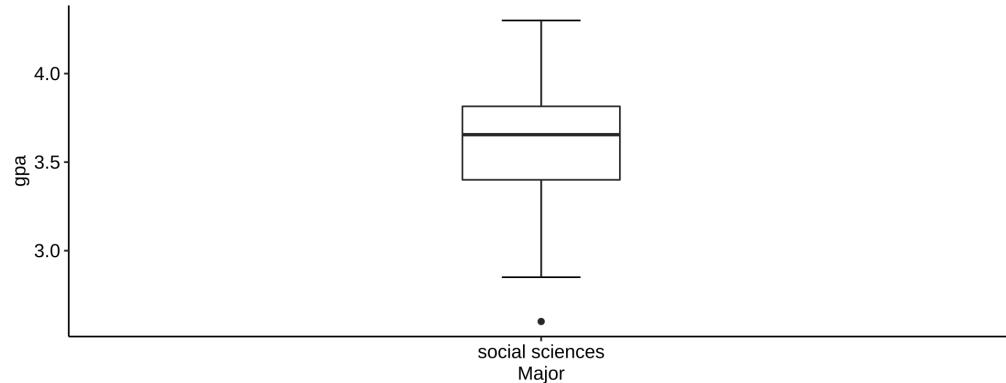
- Appropriate for **Ratio** scale variables
- Not appropriate for nominal or ordinal data
- Values are "binned"
- Bars are touching



Boxplots



Boxplots



First quartile = 25th percentile
Median = 50th percentile
Third quartile = 75th percentile

Social Sciences Major

```
##      gpa          major
##  Min.   :2.600   arts and humanities: 0
##  1st Qu.:3.400   natural sciences    : 0
##  Median :3.655   social sciences     :114
##  Mean   :3.589
##  3rd Qu.:3.815
##  Max.   :4.300
```

The Box

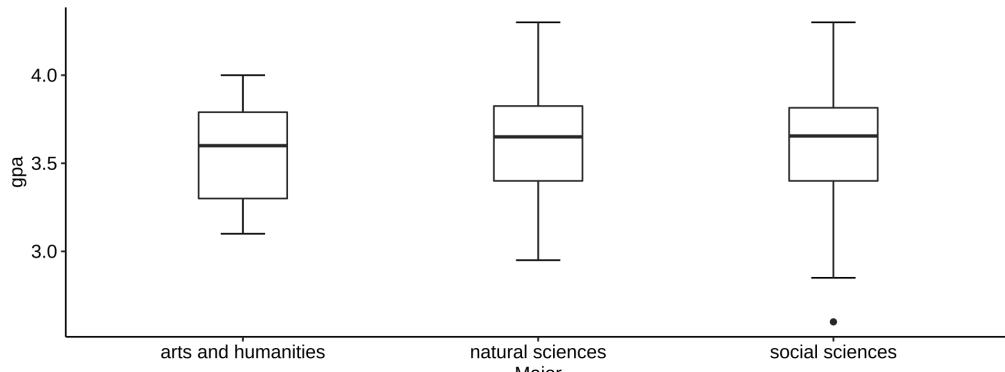
First quartile: 3.40
Median: 3.66
Third quartile: 3.82

The Whiskers

Lowest value: 2.60
Highest value: 4.30

Boxplots

- Appropriate for **Ratio** scale variable
- Appropriate to use with ratio scale and **Nominal** scale variable (if more than one group)



Show 4 entries

Search:

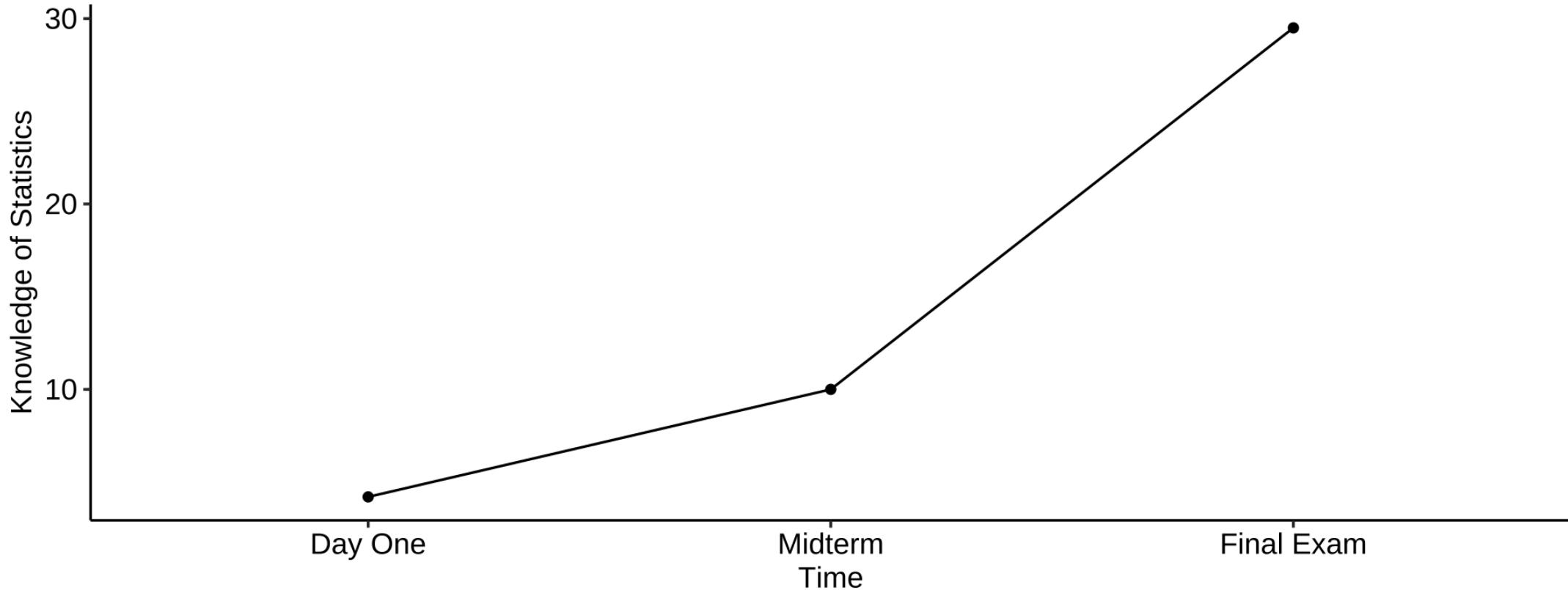
	gpa	major
1	4	social sciences
2	3.8	social sciences
3	3.93	social sciences
4	3.4	natural sciences

Showing 1 to 4 of 218 entries

Previous 1 2 3 4 5

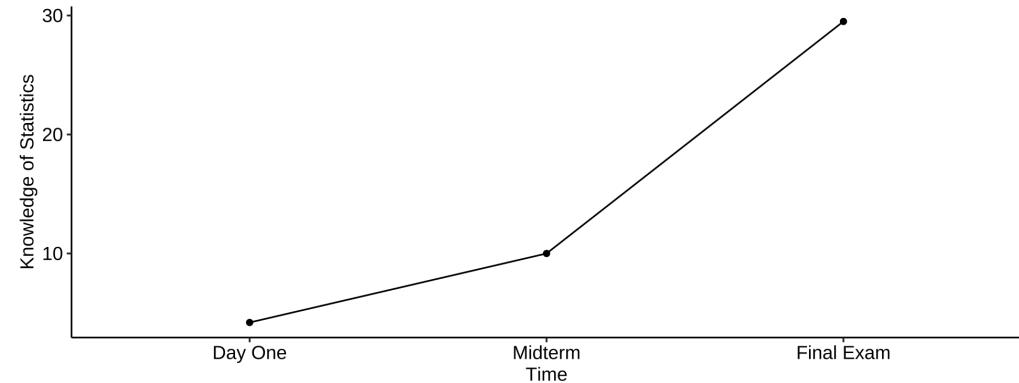
... 55 Next

Line Graphs

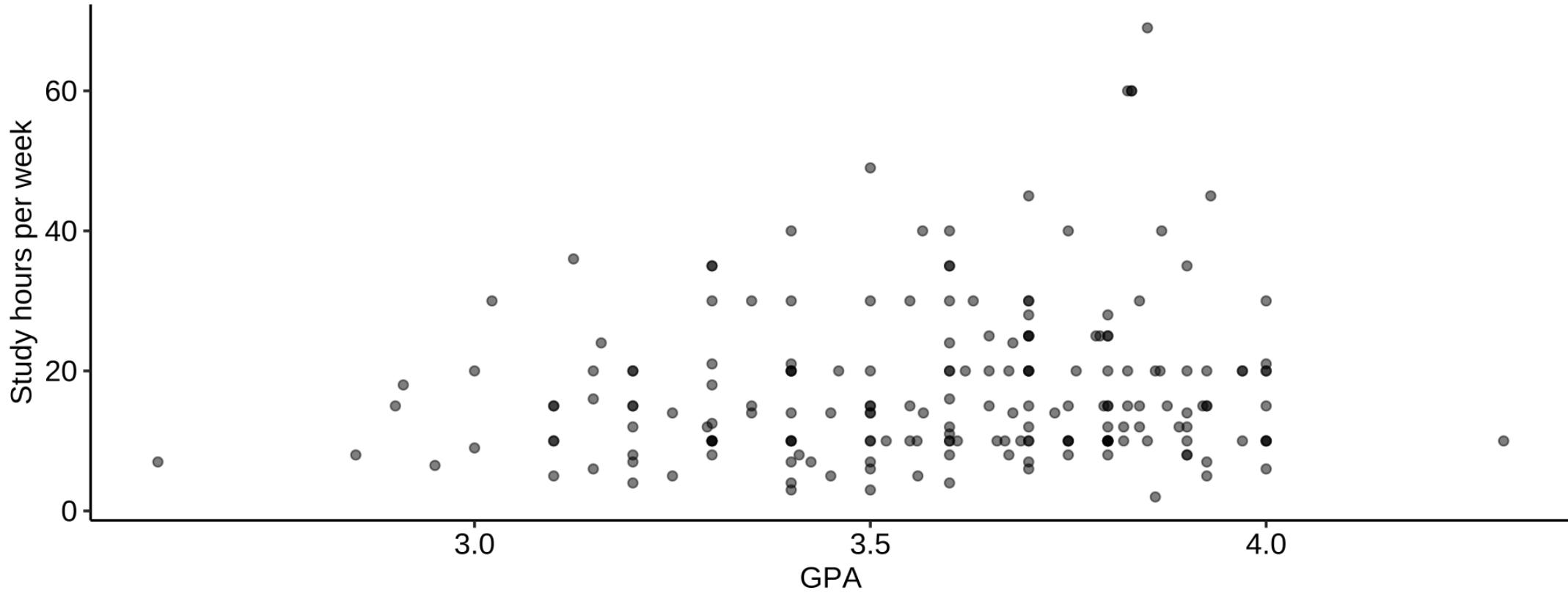


Line Graphs

- One **nominal** scale variable on the x axis
- One **ratio** scale variable on the y axis
- Most commonly used to see change over time

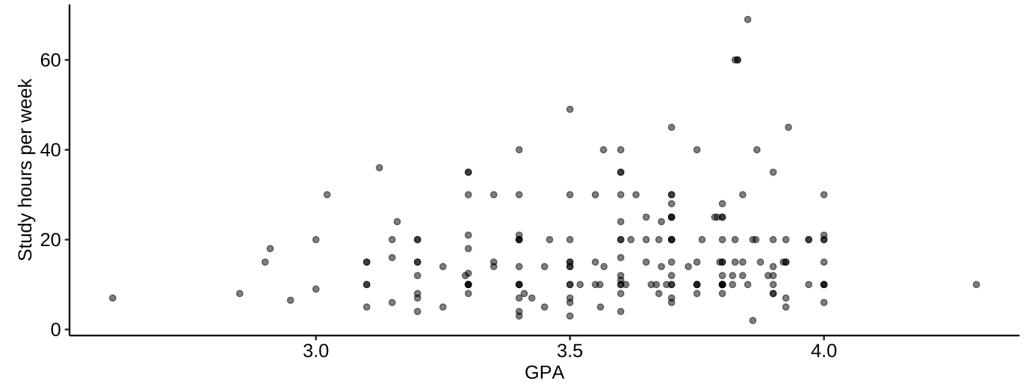


Scatterplot

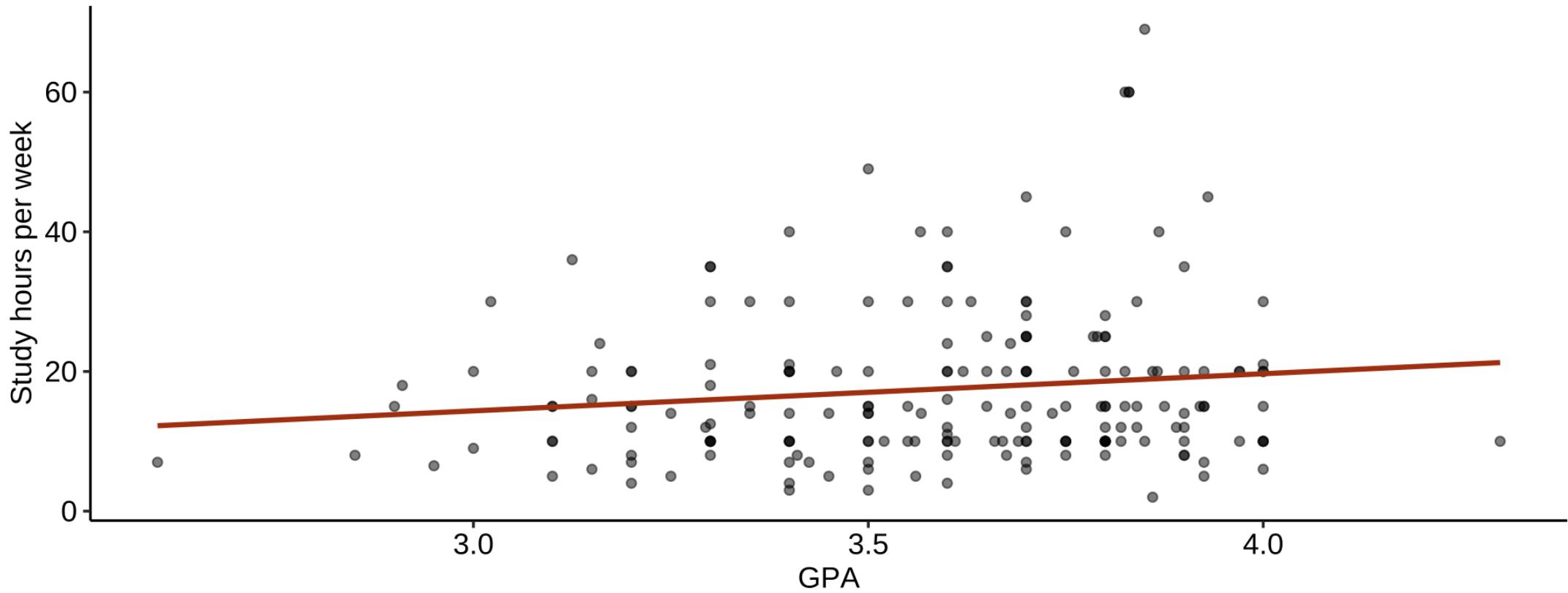


Scatterplot

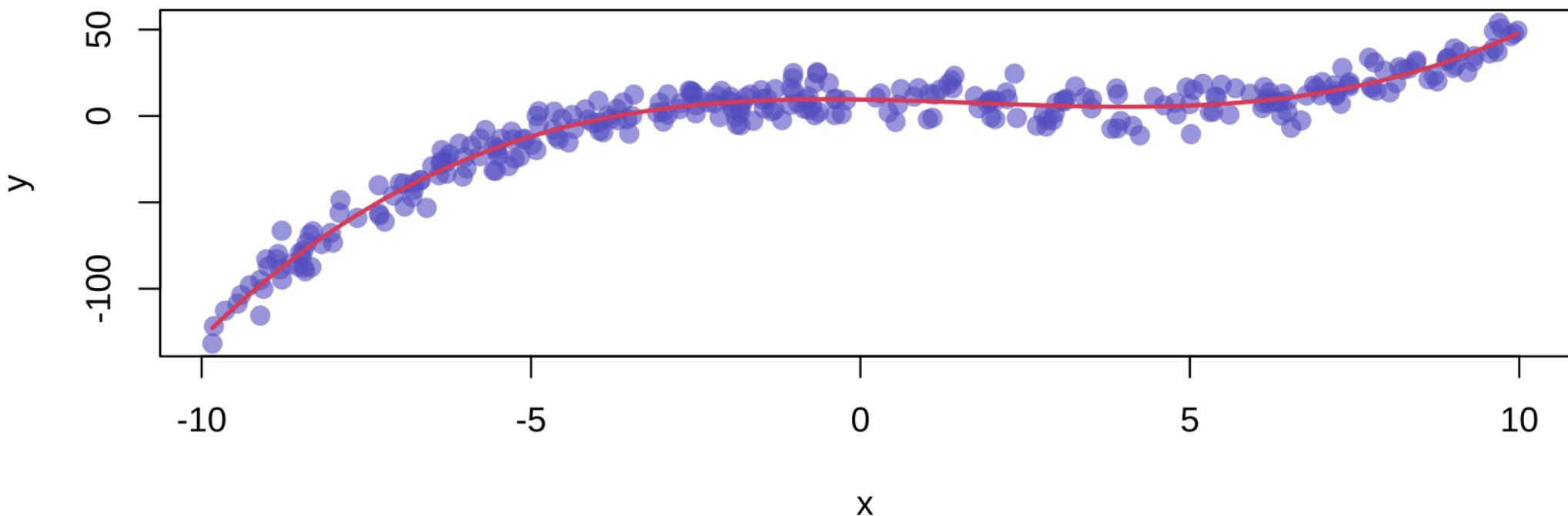
- Two **ratio** scale variables
- Each point in the dataset are represented
- Visualize how two variables relate



Scatterplot: Linear Fit



Scatterplot: Curvilinear Fit



And Many More Options



<https://www.data-to-viz.com>

Variability Practice Problem

Variability Practice Problem

Calculate the sum of squares (SS) of the following dataset:

3, 6, 7, 3, 4, 4, 7, 8, 10

Then, use this value to find:

- The sample standard deviation: s
- The population standard deviation: σ
- The sample variance: s^2
- The population variance: σ^2

05 : 00

Variability Practice Problem

Sum of Squares

$$\sum(X_i - \mu)^2 = 47.56$$

Population Standard Deviation

$$\sigma = 2.30$$

Population Variance

$$\sigma^2 = 5.28$$

Sample Standard Deviation

$$s = 2.44$$

Sample Variance

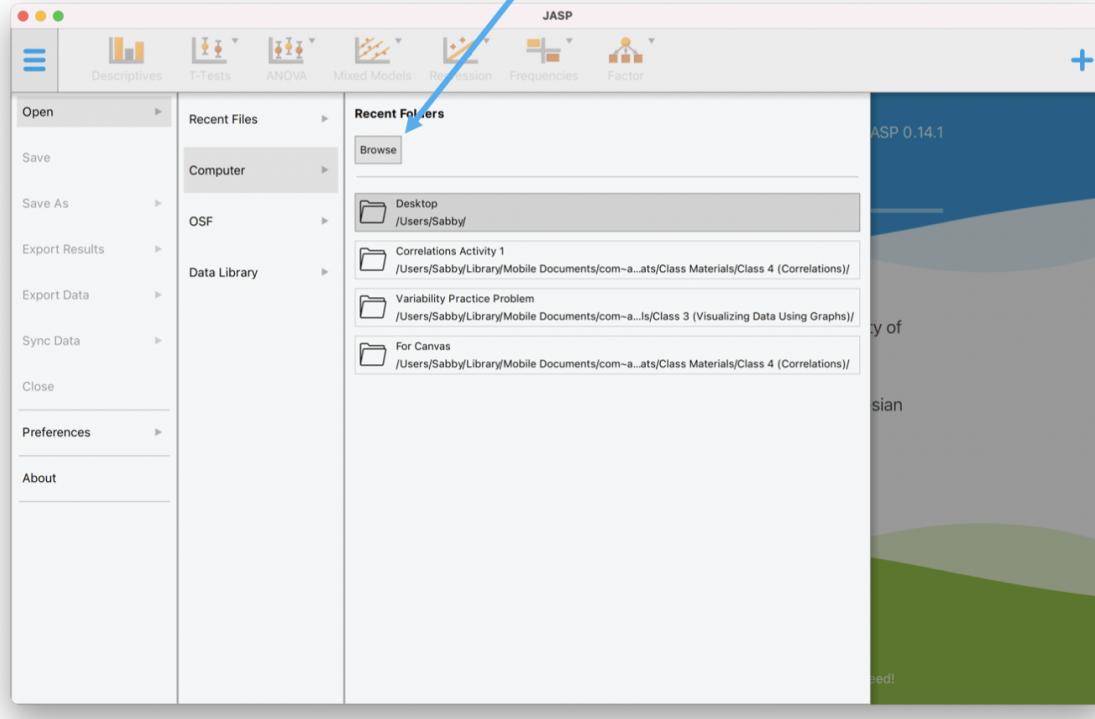
$$s^2 = 5.94$$

Descriptive Statistics in JASP

JASP: Open the File



Click on “Browse” and open a .csv or .jasp file



JASP: Click Descriptives



Click on “Descriptives” to calculate measures of central tendency and variability, etc.

The screenshot shows the JASP software interface. At the top, there is a toolbar with several icons: Descriptives (selected), T-Tests, ANOVA, Mixed Models, Regression, Frequencies, and Factor. Below the toolbar is a table titled "Stats Class 3 Dataset (Pets)* (/Users/Sabby/Desktop)". The table has three columns: ID, #Pets, and PetOwner. The data is as follows:

ID	#Pets	PetOwner
1	8	Yes
2	7	Yes
3	7	No
4	7	No
5	7	No
6	4	No
7	7	Yes
8	6	No
9	6	Yes
10	9	Yes
11	9	Yes
12	5	No
13	7	No
14	6	No
15	6	Yes
16	6	No
17	8	Yes
18	7	Yes
19	5	Yes
20	3	Yes
21	6	No
22	7.5	Yes

JASP: Drag Variables



In the new analysis pane, select your variables and move (i.e., drag and drop) them into the “Variables” box

The screenshot shows the JASP software interface. At the top, there is a toolbar with various statistical analysis icons: Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, and Factor. Below the toolbar, the title bar indicates the dataset is "Stats Class 1 Dataset (Pets)*" located at "/Users/Sabby/Desktop".

The main window is divided into two main sections: "Analysis" on the left and "Results" on the right.

Analysis Section: This section contains a tree view under "Descriptive Statistics". A blue arrow points from the text "In the new analysis pane, select your variables and move (i.e., drag and drop) them into the ‘Variables’ box" to the "Variables" box in the tree view. The "Variables" box contains three items: Happiness, V2orMorePets, and #Pets. There is also a "Split" button below the variables.

Results Section: This section displays the results of the descriptive statistics analysis. A table titled "Descriptive Statistics" is shown, with columns for Happiness, V2orMorePets, and #Pets. The table includes the following data:

	Happiness	V2orMorePets	#Pets
Valid	38	38	38
Missing	0	0	0
Mean	6.67		1.95
Std. Deviation	1.35		1.68
Minimum	3.00		0.00
Maximum	9.00		7.00

Note: Not all values are available for Nominal Text variables

JASP: Additional Descriptives



Expand the “Statistics” menu and select additional descriptive statistics you want information about (e.g., variance, medians)

The screenshot shows the JASP software interface with the title "Stats Class 3 Dataset (Pets) (/Users/Sabby/Desktop)". The top navigation bar includes Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, and Factor. The main window has tabs for Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, and Factor. The Descriptives tab is active. On the left, there are sections for Frequency tables, Plots, and Statistics. The Statistics section is expanded, showing options for Percentile Values (Quartiles, Cut points for 4 equal groups, Percentiles), Central Tendency (Mean, Median, Mode, Sum), Dispersion (S.E.mean, Std.deviation, MAD, MAD Robust, IQR, Range, Maximum, Variance, Minimum), and Distribution (Skewness, Kurtosis, Shapiro-Wilk test). The "Median" and "Variance" checkboxes are circled in blue. The "Results" panel on the right displays a table of Descriptive Statistics for three variables: Happiness, V2orMorePets, and #Pets. The table includes rows for Valid, Missing, Mean, Median, Std. Deviation, Variance, Minimum, and Maximum.

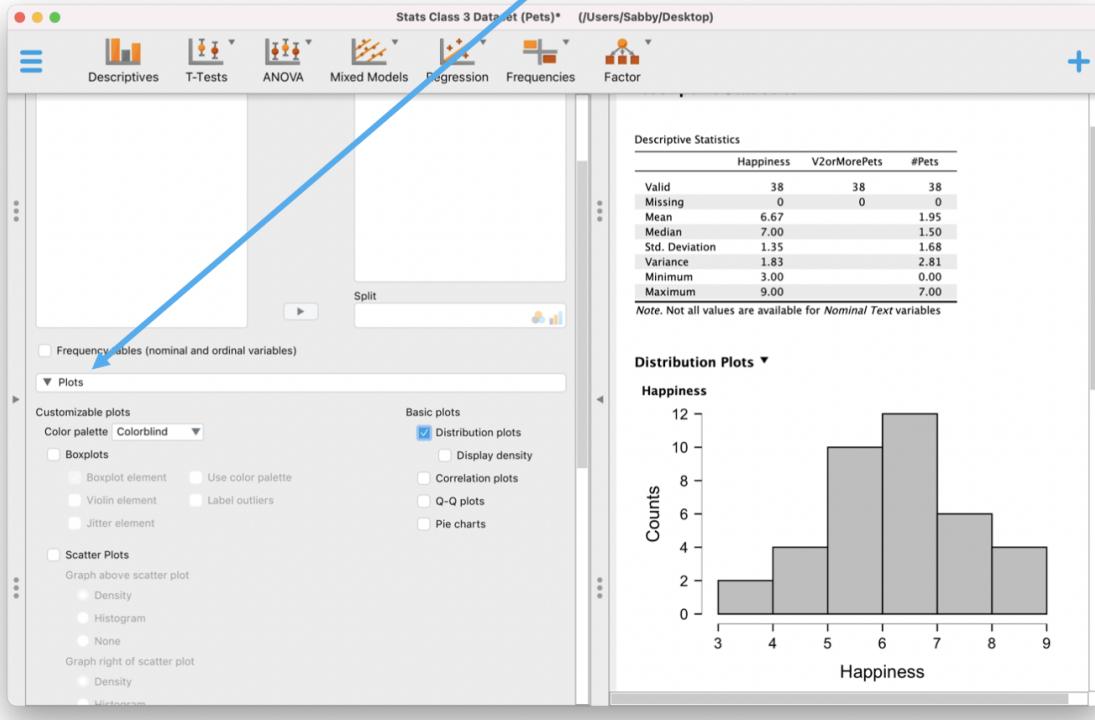
	Happiness	V2orMorePets	#Pets
Valid	38	38	38
Missing	0	0	0
Mean	6.67	1.95	
Median	7.00	1.50	
Std. Deviation	1.35	1.68	
Variance	1.83	2.81	
Minimum	3.00	0.00	
Maximum	9.00	7.00	

Note. Not all values are available for Nominal/Text variables

JASP: Plots



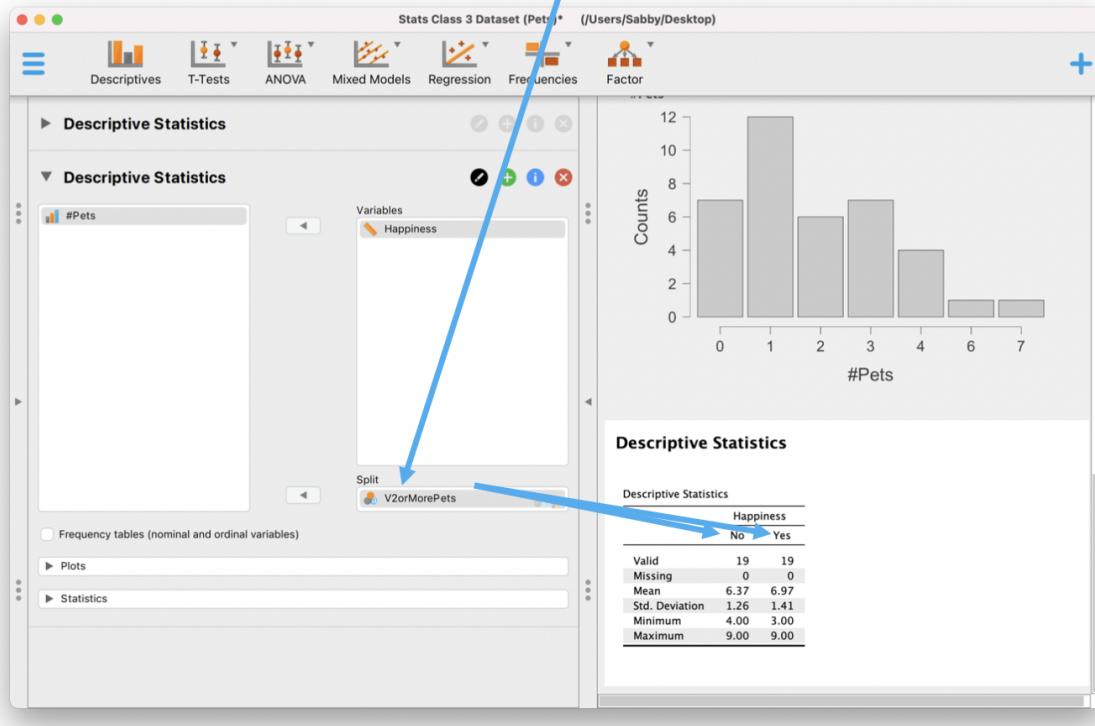
Finally, expand the “Plots” menu and select whatever plots you want JASP to generate (e.g., histograms or box plots)



JASP: Split by Nominal Variable



If you “Split” by a categorical variable, JASP will output separate results for the other variables for each category



JASP: Save your Work



If you save, your “.jasp” file will contain all three components of your work: the dataset, the analyses, and the results

The screenshot displays the JASP software interface with the following components:

- Dataset View:** On the left, a table shows a dataset with two columns: "#Pets" and "Happiness". The "#Pets" column ranges from 0 to 5, and the "Happiness" column has values "Yes" and "No".
- Analysis Configuration:** The central panel shows the "Descriptive Statistics" analysis selected. Under "Variables", "#Pets" is listed. Under "Split", "V2orMorePets" is selected. There are also sections for "Plots" and "Statistics".
- Results View:** On the right, there are two results panels. The top panel shows a histogram of "#Pets" with counts on the y-axis (0 to 12) and categories 0, 1, and 2 on the x-axis. The bottom panel is a "Descriptive Statistics" table:

	Happiness	
	No	Yes
Valid	19	19
Missing	0	0
Mean	6.37	6.97
Std. Deviation	1.26	1.41
Minimum	4.00	3.00
Maximum	9.00	9.00

Next time

Lecture

- Correlations, Reliability, and Validity

Reading

- Chapter Four

