Intro Stats Project Directions:
Analyzing Categorical and Quantitative Data

Use with Intro Stats with Support Textbook
Chapters 2,4,& 5 / Updated Spring 2026

GRADING RUBRIC FOR THE DATA ANALYSIS PROJECT *(100 points total)*

- **First and Last Name, Title, Section Number, Instructors Name, Semester and Year:** *2 points*
- **Non-cheating statement:** *2 points*
- **Five StatKey Pictures (Bar Chart, Categorical Summary Statistics, Histogram, Box Plot, Quantitative Summary Statistics):** *4 points each*
- **Answer to Question#3:** *3 points*
- **Answers to Sentence Questions:** *3 points each*
- **All Formula Calculations:** *2 points each*
- **All Other Questions:** *2 points each*

IMPORTANT NOTE: You must have the bar chart from your categorical column with summary statistics (counts and proportions) included. You must have a 5-bar histogram from your quantitative column with summary statistics (sample size, mean standard deviation, etc.) included. You must have a box plot from your quantitative column with summary statistics (sample size, mean standard deviation, etc.) included. All of the graphs and statistics must NOT be too small. They must be readable. Without readable StatKey graphs and statistics, your instructor will NOT be able to grade your report!

PROJECT QUESTIONS: Answer the following questions. This is what you will turn in to your instructor to be graded. There are 100 points possible.

Put the following title at the top of your report.
First and Last Name
Intro Stats Data Analysis Project
Instructor's Name
Section# for your class
Semester and Year

Confirm that you did not cheat. If true, write the following:
I did not cheat on this project. I did the work myself. I did not use any AI like ChatGPT on this project. No one else did this work for me.

PROJECT PART 1: CATEGORICAL DATA ANALYSIS DIRECTIONS

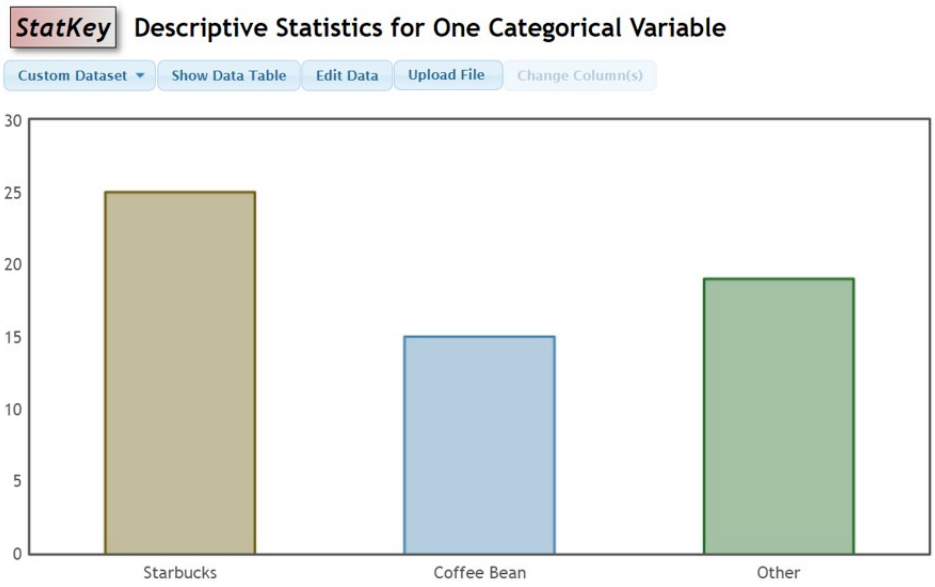
Put the categorical column of data (column of words) **assigned to you into StatKey.**

- **Go to www.lock5stat.com, click on "StatKey", and then click on "One Categorical Variable" under the descriptive statistics and graphs menu.**
- **Click the "Edit Data" button. Push Control A and delete to delete out any data listed.**
- **Highlight your column of categorical** *(words)* **data. Then push "control C" to copy. Or you can right-click and copy.**
- **Go back to the edit data screen in StatKey, and paste the assigned column of categorical data into StatKey. This is raw categorical data, so you will need to check the box that says "raw data". If you have the title, check the box that says "Data has a header row". If you do not have the column title, do not check that box that says "Data has a header row". Then push the "OK" button.**

Intro Stats Project Directions:
Analyzing Categorical and Quantitative Data

Use with Intro Stats with Support Textbook
Chapters 2,4,& 5 / Updated Spring 2026

- **You should now see a bar chart and the summary statistics with the counts and proportions for each category. Take a picture of your bar chart and a picture of the categorical summary statistics. Do not make the picture too small. Make sure the statistics and categories are readable.**



Summary Statistics

	Count	Proportion
Starbucks	25	0.424
Coffee Bean	15	0.254
Other	19	0.322
Total	59	1.000

Intro Stats Project Directions:
Analyzing Categorical and Quantitative Data

Use with Intro Stats with Support Textbook
Chapters 2,4,& 5 / Updated Spring 2026

PROJECT PART 1: CATEGORICAL DATA ANALYSIS QUESTIONS TO ANSWER.

1. **What Categorical data was assigned to you?** *(Example: The categorical data is a random sample of COC stat students favorite coffee shop.)*
2. **Show the StatKey picture showing your summary statistics and bar chart for your categorical data.** *(Look at your categorical data bar chart and summary statistics counts and proportions to answer the following.)*
3. **Convert all of your proportions listed in your summary statistics into percentages. Do <u>not</u> round.** *(Example:*
   Percentage of COC stat students in the data that prefer Starbucks = 0.424 x 100% = 42.4%.
   Percentage of COC stat students in the data that prefer Coffee Bean = 0.254 x 100% = 25.4%.
   Percentage of COC stat students in the data that prefer Other coffee shop = 0.322 x 100% = 32.2%.)
4. **Which category answer was answered the most?** *(Example: The most popular coffee shop for the COC stat students in the data was Starbucks.)*
5. **How many people gave this most common answer?** *(Example: 25 COC stat students in the data said that Starbucks is their favorite coffee shop.)*
6. **What was the proportion for the most common answer?** *(Example: 0.424 of the COC stat students in the data said that Starbucks is their favorite coffee shop.)*
7. **What was the percentage for the most common answer?** *(Example: 42.4% of the COC stat students in the data said that Starbucks is their favorite coffee shop.)*
8. **Which category answer was answered the least?** (i.e. had the lowest bar.)
9. **How many people gave this least common answer?** *(Example: The least popular coffee shop for the COC stat students in the data was Coffee Bean.)*
10. **What was the proportion for the least common answer?** *(Example: 0.254 of the COC stat students in the data said that Coffee Bean is their favorite coffee shop.)*
11. **What was the percentage for the least common answer?** *(Example: 25.4% of the COC stat students in the data said that Coffee Bean is their favorite coffee shop.)*
12. **Look at your categorical data bar chart and summary statistics counts, proportions, and percentages. List something you found surprising or interesting? Explain why you found them surprising or interesting.** *(You cannot say you found nothing surprising or interesting. If you say that, you will get no points.)*

PROJECT PART 2: QUANTITATIVE DATA ANALYSIS DIRECTIONS

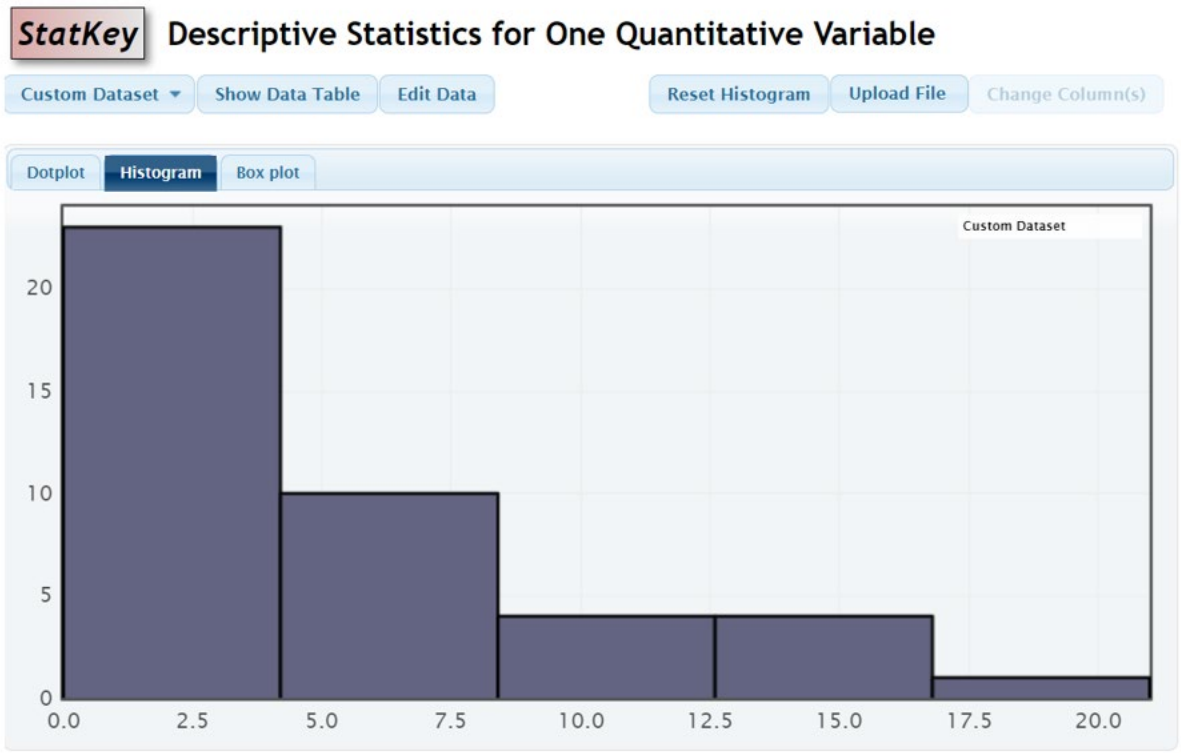
Put the quantitative column of data assigned to you (column of numbers) **into StatKey.**

- **Go to www.lock5stat.com, click on "StatKey", and then click on "One Quantitative Variable" under the descriptive statistics and graphs menu.**
- **Click the "Edit Data" button. Push Control A and then delete to delete out any data listed.**
- **Go back to your Excel or Tables spreadsheet. Highlight your assigned column of quantitative (numerical measurement) data. Then push "control C" to copy. Or you can right-click and copy.**
- **Go back to the edit data screen in StatKey, and paste the assigned column of quantitative data into StatKey. This is one column of numerical data, so you do not have an "identifier" column. Make sure the box that says "First Column is Identifier" is <u>not</u> checked. If you have the title, check the box that says "Data has a header row". If you do not have the column title, do <u>not</u> check that box that says "Data has a header row". Then push the "OK" button.**

Intro Stats Project Directions:
Analyzing Categorical and Quantitative Data

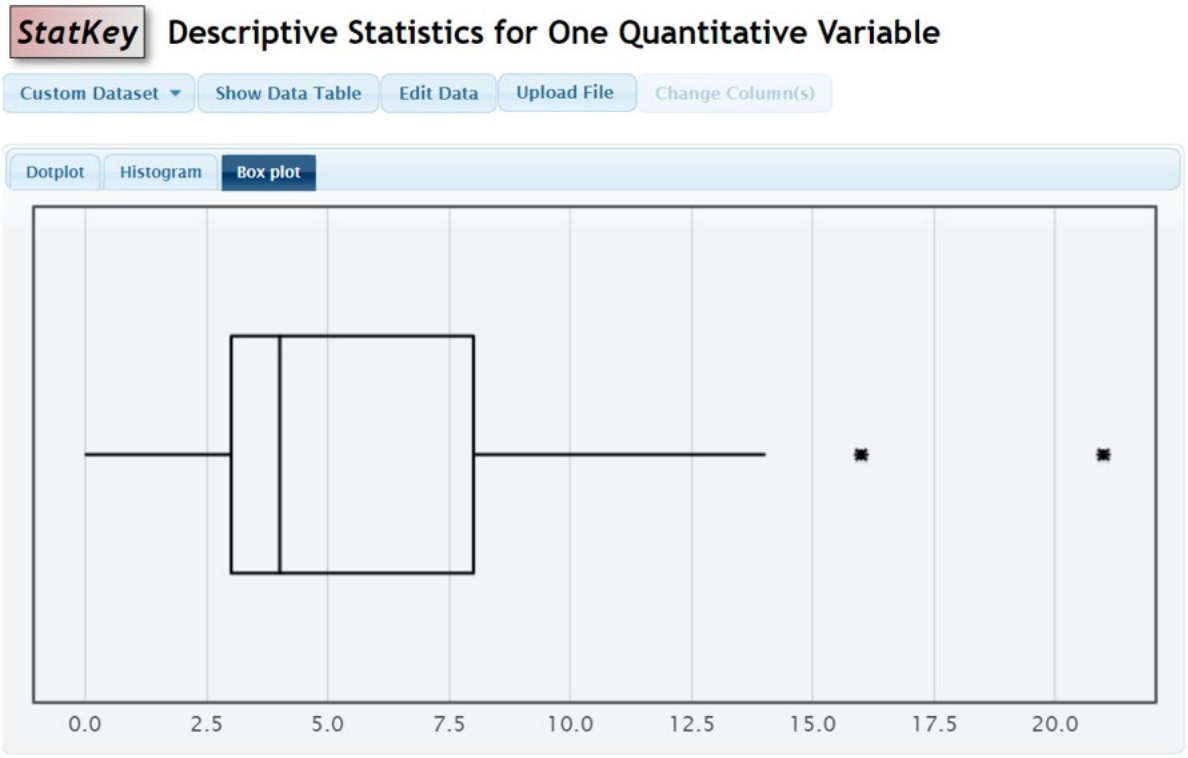
Use with Intro Stats with Support Textbook
Chapters 2,4,& 5 / Updated Spring 2026

- **Click on the "Histogram" button. Use the slider on the bottom right to change the number of buckets (bars) to 5 bars. You may need to keep sliding the buckets slider back and forth. Make sure your histogram has 5 bars. Take a picture of your histogram. If your highest bar is in the middle, we can classify the data as "nearly normal" and use the mean as our best average and standard deviation as our best measure of spread. If the highest bar is on the far right, we can classify the data as skewed left. If the highest bar is on the far left, we can classify the data as skewed right. For skewed data, we will use the median as our best average and IQR as our best measure of spread. Do not make the picture too small. Make sure the numbers are readable.**
- **Click on the "Box plot" button. Take a picture of your box plot. Do not make the picture too small. Make sure the numbers are readable.**
- **Next to the Histogram and Box Plot you will see the quantitative summary statistics window with the mean, standard deviation, sample size, median, min, max etc. Take a picture of the quantitative summary statistics window. Do not make the picture too small. Make sure all of the numbers are readable.**



Intro Stats Project Directions:
Analyzing Categorical and Quantitative Data

Use with Intro Stats with Support Textbook
Chapters 2,4,& 5 / Updated Spring 2026



Summary Statistics

Statistic	Value
Sample Size	42
Mean	5.976
Standard Deviation	4.672
Minimum	0
Q_1	3.000
Median	4.000
Q_3	8.000
Maximum	21

Histogram Controls Set Limits

Number of buckets: 5

lower bound	upper bound	frequency

Intro Stats Project Directions:
Analyzing Categorical and Quantitative Data

Use with Intro Stats with Support Textbook
Chapters 2,4,& 5 / Updated Spring 2026

PROJECT PART 2: QUANTITATIVE DATA ANALYSIS QUESTIONS TO ANSWER.

13. **What Quantitative data was assigned to you?** *(Example: The quantitative data is a random sample of exercise hours per week for COC stat students.)*
14. **What were the units of your quantitative data?** *(Example: Hours of Exercise per Week)*
15. **Show the StatKey picture showing your summary statistics and 5-bar histogram for your quantitative data.**
16. **Show the StatKey picture showing your summary statistics and box plot for your quantitative data.**
17. **Look closely at the histogram and box plot. Was your data normal** (close to bell shaped)**, skewed right** (long right tail)**, or skewed left** (long left tail)**?** *(Example: The COC Stat student exercise data was skewed right.)*
18. **Based on the shape, should we use the <u>mean average</u> or the <u>median average</u> as our most accurate average?** *(Example: Since the COC stat student exercise data is skewed right, we should use the median average.)*
19. **Look at mean and median calculated in your summary statistics for the quantitative data. Write a sentence using the most accurate average in context with units.** *(For example: COC stat students in my data exercise a median average of 8 hours per week.)*
20. **Look at the first quartile (Q1) and the third quartile (Q3) calculated in your summary statistics for the quantitative data. Calculate the interquartile range (IQR) using the formula** IQR = Q3 – Q1**.** *(Example: IQR = 8 – 3 = 5 hours per week)*
21. **Based on the shape, should we use the <u>standard deviation</u> or the <u>interquartile range (IQR)</u> as our most accurate measure of spread?** *(Example: Since the COC stat student exercise data is skewed right, we should use IQR.)*

22. **Look at the mean and the standard deviation listed in your summary statistics for the quantitative data. Calculate the following two values using the provided formulas.**

    mean – standard deviation = ?
    mean + standard deviation = ?

    (Example: mean – standard deviation = 5.976 – 4.672 = 1.304 hours per week of exercise
    mean + standard deviation = 5.976 + 4.672 = 10.648 hours per week of exercise)

23. **Based on the shape, which formula should we use for typical values**
    <u>Between Q1 & Q3</u> **OR** <u>Between mean ± standard deviation</u>?

    (Example: Since the COC stat student exercise data is skewed right, typical values should be between Q1 and Q3.)

24. **Write a sentence in context with numbers and units explaining the two cutoffs where typical values will be in between.** *(Example: Typical COC stat students in my data exercise between 5 and 12 hours per week.)*

Intro Stats Project Directions:
Analyzing Categorical and Quantitative Data

Use with Intro Stats with Support Textbook
Chapters 2,4,& 5 / Updated Spring 2026

25. Look at the mean and the standard deviation listed in your summary statistics for the quantitative data. Calculate the following two values using the provided formulas. Make sure to do the multiplication before doing the addition or subtraction. *(Note: It is common for the lower outlier cutoff to be negative. It does not mean you did something wrong.)*

mean – (2 x standard deviation) =
mean + (2 x standard deviation) =

(Example: mean – (2 x standard deviation) = 5.976 – (2 x 4.672) = -3.368 hours per week of exercise
mean + (2 x standard deviation) = 5.976 + (2 x 4.672) = 15.320 hours per week of exercise)

26. **Look at the answer for #20 and the Q1 and Q3 listed in your summary statistics for the quantitative data. Calculate the following two values using the provided formulas. Make sure to do the multiplication before doing the addition or subtraction.** *(Note: It is common for the lower outlier cutoff to be negative. It does not mean you did something wrong.)*

Q1 – (1.5 x IQR) =
Q3 + (1.5 x IQR) =

Example: Q1 – (1.5 x IQR) = 3 – (1.5 x 5) = -4.5 hours per week of exercise

Q3 + (1.5 x IQR) = 8 + (1.5 x 5) = 15.5 hours per week of exercise

27. Based on the shape of the data, which formula should you use for the low outlier cutoff mean – (2 x standard deviation) **OR** Q1 – (1.5 x IQR)**?**

(Example: Since the COC stat student exercise data is skewed right, the low outlier cutoff should be Q1 – (1.5 x IQR).)

28. **Write a sentence in context with the number and units about the low outlier cutoff. The number should coincide with the formula you picked for the low outlier cutoff.**
(For example: It is unusual for COC stat students in my data to work out less than -4.5 hours per week. This means zero hours of exercise per week was not unusual.)

29. Based on your low outlier cutoff, write a sentence in context listing all of the low outliers in the data set. If there are no low outliers, just say that "There are no unusually low values in my data set". *(For example: There were no unusually low hours of exercise per week in my data.)*

30. **Based on the shape of the data, which formula should you use for the high outlier cutoff** mean + (2 x standard deviation) **OR** Q3 + (1.5 x IQR)**?**

(Example: Since the COC stat student exercise data is skewed right, the high outlier cutoff should be Q3 + (1.5 x IQR).)

Intro Stats Project Directions:
Analyzing Categorical and Quantitative Data

Use with Intro Stats with Support Textbook
Chapters 2,4,& 5 / Updated Spring 2026

31. **Write a sentence in context with units about the high outlier cutoff.** *(The number should coincide with the formula you picked in #31. For example: It is unusual for a COC stat student in the data to work out more than 15.5 hours per week.)*

32. **Based on your high outlier cutoff, write a sentence in context listing all of the high outliers in the data set. If there are no high outliers, just say that "There are no unusually high values in my data set".** *(For example: In my data, there were two unusually high values that COC stat students exercise per week: 16 hours per week and 21 hours per week.)*

33. **Look at your quantitative data graphs, statistics and calculations. List something you found surprising or interesting? Explain why you found them surprising or interesting.** *(You cannot say you found nothing surprising or interesting. If you say that, you will get no points.)*
