Math 140 & Math 140X Project#1
Directions & Grading Rubric: Collecting Your Own Data
Use with Teachout Textbook Sections 1A,1B & 1C
Updated Spring 2025

Project#1 Directions *(You will be collecting your data, typing the data into two columns in Excel, creating an "Other" category if needed, custom sorting your data, and then answering 15 questions below.)*

Collecting Data

I. **Choose one categorical question and one quantitative question from the lists provided at the "projects" page at www.matt-teachout.org or on your instructor's home page in Canvas.**
    - This is called paired data since you will be asking each person you talk to both questions. Choose your two questions from one of the various lists provided on your home page in Canvas. Each list has a different topic emphasis *(business/marketing, food/drink, health, sports, politics, other miscellaneous topics).*
    - Let your instructor know the paired-data number you chose. Each student must use different paired data. It is ok if one of the questions is the same, but both questions cannot be the same as another student. So, if you see on the list that the paired data says "taken", that means another student has chosen that pair and you will have to choose another. First come first serve.

II. **Collect your categorical and quantitative data.**
    - Identify your large population of interest. The population must be large and cannot be your friends and family, people in a class, or a few people at your work. Examples of the population of interest might be all people in the Santa Clarita Valley, all COC students, or chips sold in stores in Reseda, CA.
    - Get data from 50 or more people or objects in your population. This is called a sample. If your population is all COC students, you cannot ask people outside of COC. If your population is all people in the San Fernando Valley, you cannot ask COC students or people in other areas. Your data must be a sample, not a census, so you cannot get data from everyone in your population.
    - You will need to ask each person you talk to BOTH of the questions listed in your paired data choice. Or you can choose objects to get the two bits of data from.
    - Only keep data from people that answered both questions**.** Keep track of how many people refused to answer one or both of the questions. This will be asked in your report and is part of non-response bias. However, do not keep data from people that only answered one of the questions. The only data you will type into your spreadsheet, are those that answered both questions.
    - You need to get data from at least 50 people or objects. Keep track of the number of people that refused to answer one or both of the questions. You may need to ask 75 people to get at least 50 that answered both questions. Remember, more data is better so it is better if you can get data from more than 50 people.

Type your Data into Excel

III. **Type your data into an Excel spreadsheet (Windows) or another spreadsheet program.**
    - Your spreadsheet should have two columns, one for the categorical data (words) on the left column and one for the quantitative data (count or numerical measurement) on the right column.
    - Make sure both of your columns have a title describing what was asked.

- When typing in the categorical data words, make sure the words for each answer are typed exactly the same. For example, if you ask people their favorite sports team, you should not have "dodgers", "Dodgers", "LA Dodgers", "Los Angeles Dodgers.", etc. in the same column. They should all be spelled and capitalized the same. In this case, people that answered Los Angeles Dodgers should all say "Dodgers". The reason for this is that the computer will see "Dodgers" and "dodgers" as separate different answers. They need to be spelled exactly the same for the computer to recognize it.

Create an Other Category (Only if this applies to your data.)

IV. **If you have many different categorical answers, create an "Other" category.**
- Note: If your categorical question has set number of answers like "Yes or NO" or "Hamburgers, Tacos or Pizza", or "small, medium, large and extra-large", then DO NOT make an "Other" category. Just keep the answers you have.
- If you have many categorical answer options, then you will want to create an "Other" category. You will want to keep answers that many people answered. Answers given by only a few people should be changed to say "Other".
- For example, suppose the categorical question was favorite coffee shop. Suppose many people answered "Starbucks", "Honu", "Peets", and "Coffee Bean". Unusual answers were "McDonalds" and "Home Coffee". These unusual answers can now become the "Other" category.

Favorite Coffee Shop	$ spent on gas per week
Starbucks	50
Coffee Bean	40
Starbucks	0
Honu	25
Starbucks	60
Peets	75
Coffee Bean	20
Home Coffee	35
Honu	80
Peets	100
McDonalds	65

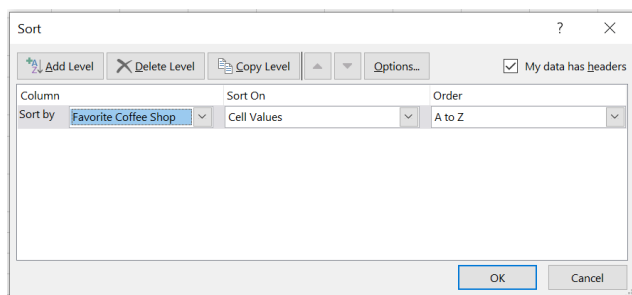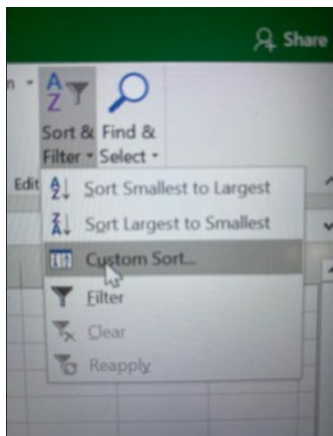| Favorite Coffee Shop	$ spent on gas per week
Starbucks	50
Coffee Bean	40
Starbucks	0
Honu	25
Starbucks	60
Peets	75
Coffee Bean	20
Other	35
Honu	80
Peets	100
Other	65

Custom Sort your Data

> **V.** **Custom Sort the paired categorical and quantitative data you collected by category.**
> - Highlight <u>both</u> of the two columns of data. Here is an example using the favorite coffee shop and $ spent on gas per week questions.

Favorite Coffee Shop	$ spent on gas per week
Coffee Bean	40
Coffee Bean	20
Honu	25
Honu	80
Other	35
Other	65
Peets	75
Peets	100
Starbucks	50
Starbucks	0
Starbucks	60

> - **Custom Sorting Directions in Excel:** After highlighting both of the columns, click on "sort and filter" and then click on "custom sort". The sort window will open. Make sure you check the box that says "My data has headers". Under "Sort by" click on your categorical column. In this example, that is "Favorite Coffee Shop".   Then push "OK". Your categorical data should now be in alphabetical order without losing the pairing to your quantitative data. <u>Remember you have to highlight both columns before sorting</u>. <u>Do NOT just sort the left column alone. If you do that you lose the pairing and mess up your data</u>!

Favorite Coffee Shop	$ spent on gas per week
Coffee Bean	40
Coffee Bean	20
Honu	25
Honu	80
Other	35
Other	65
Peets	75
Peets	100
Starbucks	50
Starbucks	0
Starbucks	60

- You should now see your left column in alphabetical order, but you have not lost the pairing to your quantitative right column of data. Notice the numbers next to Coffee Bean still say 40 and 20. The numbers next to Starbucks still say 50, 0 and 60. The numbers next to Honu still say 25 and 80. The number next to Peets still say 75 and 100. The numbers next to "Other" still say 65 and 35.

Save Excel Spreadsheet File

VI.   **Save Excel File. (Will need to take a picture of the two columns of sorted data if you used a different spreadsheet program.)**

- Save your two columns of sorted data in Excel or Tables. You will need this data for future projects in the class. If you used Excel, you can turn in the Excel file. If you used a Numbers spreadsheet with a MAC, you will need to take a picture of it to prove you have done it. You will attach either the Excel file or the picture or the Tables file with your report to prove you have typed the data into a spreadsheet. Remember, you should have at least 51 rows including the title. For categorical data with many answers, you should have narrowed it down to your top answers. Unusual answers should be changed to say "Other". The data should then be in alphabetical order without losing the pairing by using a custom sort. Save the spreadsheet. You will need it to complete other projects in the class.

Favorite Coffee Shop	$ spent on gas per week
Coffee Bean	40
Coffee Bean	20
Honu	25
Honu	80
Other	35
Other	65
Peets	75
Peets	100
Starbucks	50
Starbucks	0
Starbucks	60

Answer Report Questions #1-15 below

Project#1 Report Questions: Turn in the Excel Spreadsheet with two columns of sorted data, name and project title, anti-cheating statement, and answers to these 15 questions. There are 100 points possible.

A. **Attach a picture of your Excel or Tables spreadsheet to prove you have typed the two columns of data into Excel or Tables.** <u>Notes</u>: <u>The data must be custom sorted by category but not lose the connection to the quantitative numerical question answer.</u> <u>Unusual categorical answers should be changed to "Other".</u> <u>I prefer your data to be in Excel. If your two columns are in Excel, you can attach the Excel file with your report. If you use a different spreadsheet program, you will need to submit a picture of your two columns of data.</u>

B. **Put the following title at the top of your report.**
   First and Last Name
   Math 140X Project#1 Collecting Data
   Instructor's Name
   Section# for your class
   Semester and Year (Spring 2025)

C. **Confirm that you did not cheat. If true, write the following:**
   <u>I did not cheat on this project. I collected the data and did the work myself.</u>
   <u>No one else did this work for me.</u>

D. **Answer the following fifteen questions.**

   1. **What was your categorical question?** *(Answer was words)*

   2. **What was your quantitative question?** *(Answer was numerical measurement or count.)*

   3. **What was your unit of measurement in your quantitative question?** *(For example: dollars, pounds, inches, # of pets, # of times put gas in the car per month,…)*

   4. **What was your pair data number corresponding to these two questions on the lists provided on your instructor's home page?**

   5. **<u>How</u> <u>many</u> people answered both questions?** *(Or how many objects did you get paired data from?)*

   6. **<u>How</u> <u>many</u> people did not answer one or both of the questions for any reason?** *(Or how many objects were you <u>not</u> able to get paired data on?)*

   7. **What is your population of interest?** *(i.e. What large group could you apply your sample data to?) For example: If you collect data from 47 COC students you attend class with, your population of interest might be <u>all</u> students at College of the Canyons. If you collect data from 33 people on Instagram, your population of interest may be <u>all</u> people on Instagram.*

8. **Describe how you collected the data? Be specific and give details.** *Here are examples: "I walked around the Valencia mall and asked people I met both questions." "I posted my questions on Instagram and kept data from only those people that answered both of my questions". "I sat in the COC cafeteria and asked COC students I met both of my questions". "I asked all of the students in the COC classes I am enrolled in during the semester."*

9. **What method of collecting data did you use: convenience, voluntary response, simple random sample, random cluster, non-random cluster?** (Note: You cannot have a census. The data you collect must be a sample from a small group that you will attempt to apply to a larger population.)

10. **Do you think that sampling bias is present in your data?** *(Data has sampling bias if the sample data was not collected randomly. Do not answer yes or no. Your answer should say "There is sampling bias in my data" or "There is not sampling bias in my data".)* **Explain why** you think your data **does** or **does not** have sampling bias.

11. **Do you think that question bias is present in your data?** *(Data has question bias if the question is unclear or phrased with extra information. Do not answer yes or no. Your answer should say "There is question bias in my data" or "There is not question bias in my data".)* **Explain why** you think your data **does** or **does not** have **question** bias.

12. **Do you think that response bias may be present in your data?** *Data has response bias if the people answering the questions lie or cannot give an accurate answer. There is also a response bias if the paired data cannot be measured accurately. Do not answer yes or no. Your answer should say "There is response bias in my data" or "There is not response bias in my data".)* **Explain why** you think your data **may** or **may not** have **response bias.**

13. **Do you think that non-response bias may be present in your data?** *(Data has non-response bias if there are people that refused to answer the questions or for any reason were not able to answer the questions. There is also non-response bias if the paired data cannot be collected from certain objects. Do not answer yes or no. Your answer should say "There is non-response bias in my data" or "There is not non-response bias in my data".)* **Explain why** you think your data **may** or **may not** have **non-response bias.**

14. **Do you think that deliberate bias is present in your data?** *(Data has deliberate bias if there are groups in the population that were left out and not represented in the data. Data can also have deliberate bias if the person collecting the data deletes or falsifies the data after collecting it. Do not answer yes or no. Your answer should say "There is deliberate bias in my data" or "There is not deliberate bias in my data".)* **Explain why you think your data does or does not have deliberate bias.**

15. **List at least three specific questions that could be answered by analyzing your data?**

Project #1 Grading Rubric *(100 points total)*

- **Collect data on both questions from people or objects.** *(24 points total)*
- **Typing two columns of data into Excel or Tables spreadsheet.** *(20 points)*
- **Create an "Other" category if needed.** *(5 points)*
- **Custom sort your paired data so that the categorical data is in alphabetical order but you have not lost the pairing.** *(5 points)*
- **Name and Title, Cheating Statement, Answers #1-7,9** *(2 points each, 20 points total)*
- **Answers #8,15** *(3 points each, 6 total points)*
- **Answers #10-14** *(4 points each, 20 total points)*