Organizing Data

How is data displayed to make it meaningful?

Why?

Scientists rely on data to describe nature and uncover relationships. The raw data—measurements taken in the lab—are most useful when they are organized in a way that makes the relationships clear. In this activity you will explore two common ways that scientists organize data to help in analysis.

Model 1 – Copper Samples

Group Number	Volume (cm^3)	Mass (g)	Substance
1	2.0	17.92	Copper
2	6.0	50.89	Copper
3	10.0	93.45	Copper
4	8.0	79.30	Copper
5	14.0	125.44	Copper
6	4.0	39.80	Copper
7	12.0	103.85	Copper

Room Temperature: 21.7 °C

1. What substance were the students working with to obtain the data in Model 1?

   Copper

2. What variables did the students measure to produce the data in Model 1?

   Volume and Mass

3. Briefly describe an experiment that the class might have done on the day that the data in Model 1 was collected. Discuss your answer with your group members to be sure there is consensus.

   Yield of copper extraction from a solution, or maybe measurement of relationship of volume and mass of copper.

4. Consider the data in Model 1.

   a. Which variable was the **independent variable** in the experiment, and why do you think it was the independent variable?

   Volume because it measures the copper solution or maybe copper itself.

   b. Which variable was the **dependent variable** in the experiment, and why do you think it was the dependent variable?

   Mass because it measures total yield of the copper after extraction.

   c. List two **controlled variables** in the experiment?

   Temperature and copper.

5. Consider the data in Model 1.

   a. How is the data organized?

   In a spreadsheet but values are random.

   b. Is the table in Model 1 organized in a way that helps determine a relationship between the independent and dependent variables in the experiment? Explain.

   Not really, it just lists the data for some given variable V. It doesn't describe the experiment or anything of the like really. Plus the data is kinda organized at random so it's hard to tell.

6. Propose a better way to organize the data in Model 1, and transcribe the data into the table below.

Group Number	Volume (cm^3)	Mass (g)	Substance
1	2.0	17.92	Copper
6	4.0	39.80	Copper
2	6.0	50.89	Copper
4	8.0	79.30	Copper
3	10.0	93.45	Copper
7	12.0	103.85	Copper
5	14.0	125.44	Copper

7. The data table in Question 6 should allow you to state a relationship between the variables involved in the class's experiment. Complete the following statement:

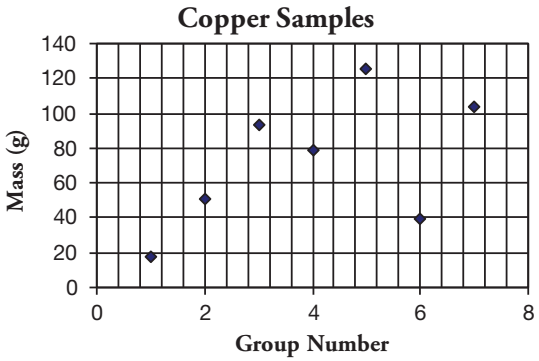
   As the volume of copper increases, the mass of copper increases as well.

Read This!

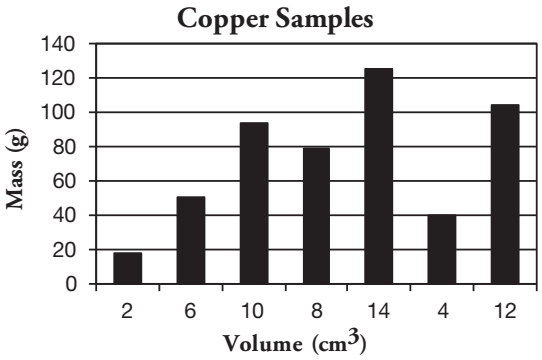
When scientists design an experiment they are usually looking for a cause-and-effect relationship between the independent variable and the dependent variable. Therefore, organizing the data by the independent variable is the easiest way to reveal a relationship. When the data is not organized, the relationships are not apparent.

Model 2 – Graphs for Copper Data

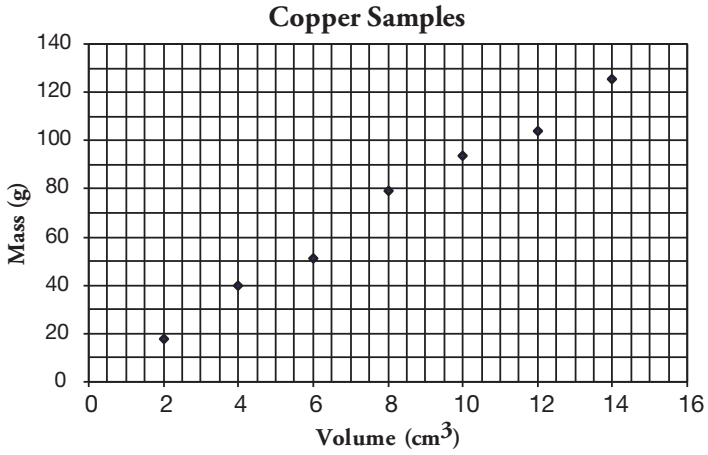
Graph A



Graph B



Graph C



8. Identify each of the graphs in Model 2 as a bar graph or a scatter plot.

   A scatter plot
   B Bar
   C scatter

9. One of the data points in graph B indicates that a volume of 8 cm^3 has a mass of 80 g. Which other graph in Model 2 shows this same data?

   C

10. Of the three graphs in Model 2, which illustrates the relationship between the variables that you stated in Question 7 most clearly?

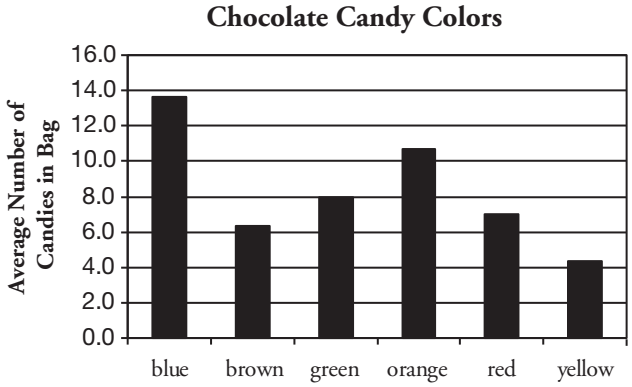
    Graph C

Read This!

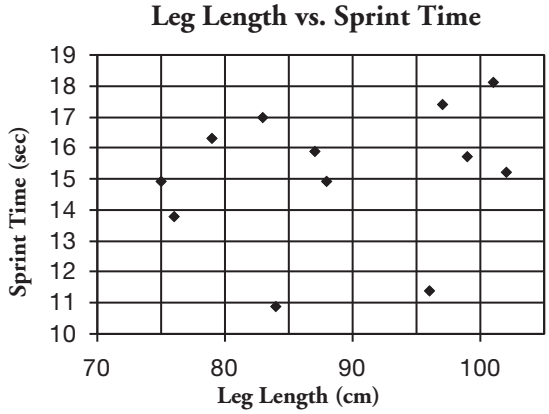
Scientists use graphs to clearly illustrate whether or not there is a relationship between variables. In most cases a scatter plot is used. Bar graphs are sometimes used if the independent variable is limited to specific numeric values (where the values in-between are not possible) or is non-numeric. A special type of bar graph called a histogram is used in cases where the scientist wants to show how often something happens.

Model 3 – More Examples of Graphs

Graph D



Graph E



11. Identify the independent variable and dependent variable for each of the graphs in Model 3.

	Graph D	Graph E
Independent Variable	color	leg length
Dependent Variable	# of candies	sprint time

12. Match the experimental questions below to the appropriate graph from Model 3.
    a. "Is the number of candies in a bag of chocolates dependent on the color of the candy?"
    Graph no
    b. "Does the length of a person's leg affect the time it takes them to sprint 60 yards?"
    Graph Maybe
13. Why was the data for Graph D plotted in a bar graph?

    Color cannot be quantized

14. Using the graphs in Model 2 and Model 3 as examples of proper graphs, identify the axis (*x* or *y*) where you would usually plot the independent variable.

    X

15. For each of the following experiments, choose "scatter plot" or "bar graph" as the most appropriate way to display the data. Justify your answer.

a. Students heated oil on a hot plate at the #4 setting for different amounts of time. They wanted to answer the question "How long do you need to heat an oil bath to reach a given temperature?"

Volume Oil (mL)	Hot Plate Setting	Initial Temp. of Oil (°C)	Time Heated (min)	Final Temp. of Oil (°C)
250	#4	21	0	21
250	#4	21	5	30
250	#4	21	10	38
250	#4	21	15	47
250	#4	21	20	57

scatter because independent var is quantized

b. Students measured the height of each student in class. They wanted to answer the question "What is the most common height among 10th grade students?"

Height Range	Number of Students
under 4′ 0″	1
4′ 1″ to 4′ 6″	3
4′ 7″ to 5′ 0″	5
5′ 1″ to 5′ 6″	9
5′ 7″ to 6′ 0″	3
over 6′ 0″	1

Bar because hight range sections cannot be quantized.

c. The Fish and Wildlife agency measured the size of Pacific salmon for 1 year and recorded the average weight for each species.

Salmon Species	Average Weight (lbs)
King	15
Sockeye	8
Silver	12
Chum	15
Humpback	5

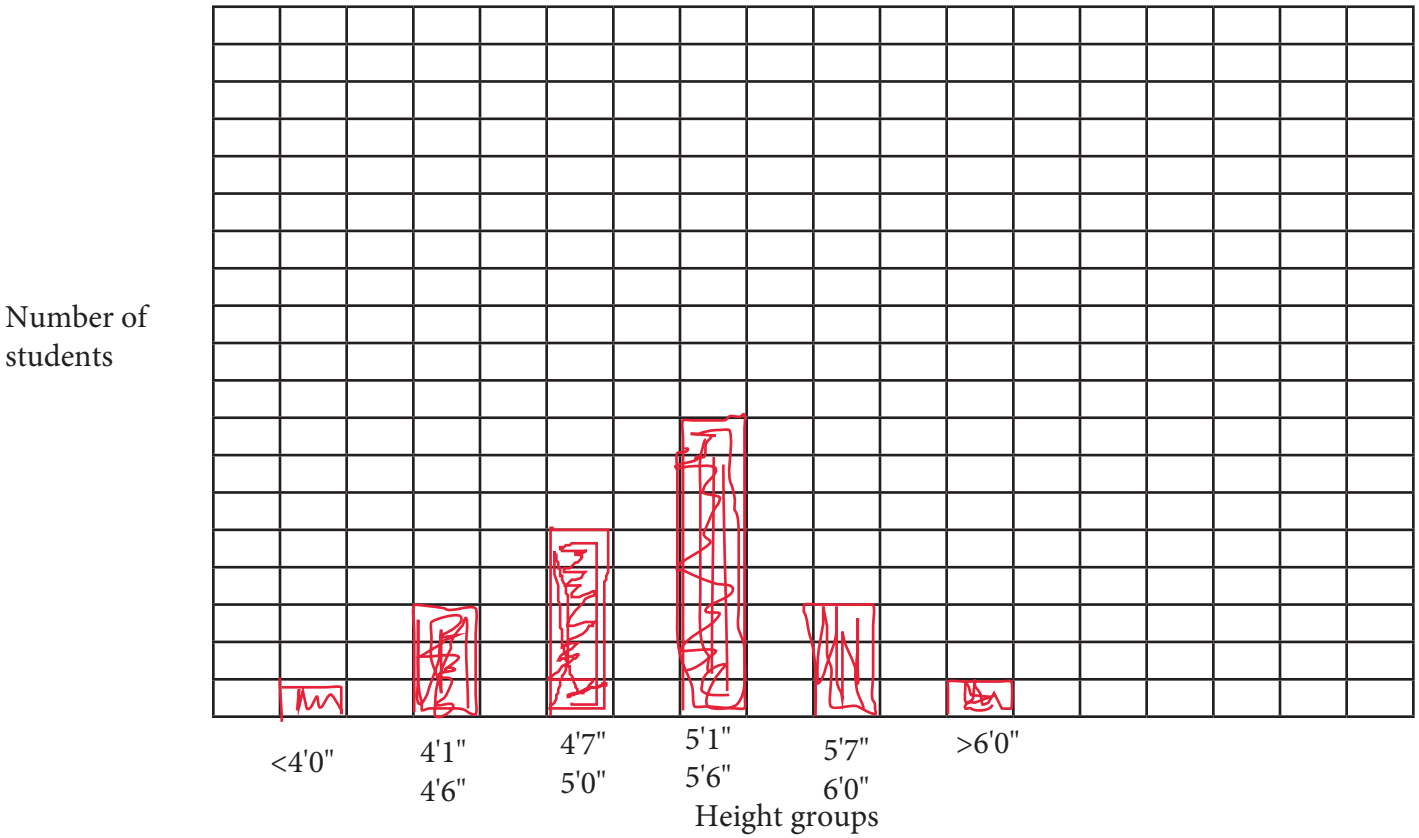
bar because salmon species names cannot be quantized

d. The National Oceanic and Atmospheric Administration measured the pressure of the atmosphere at various altitudes.

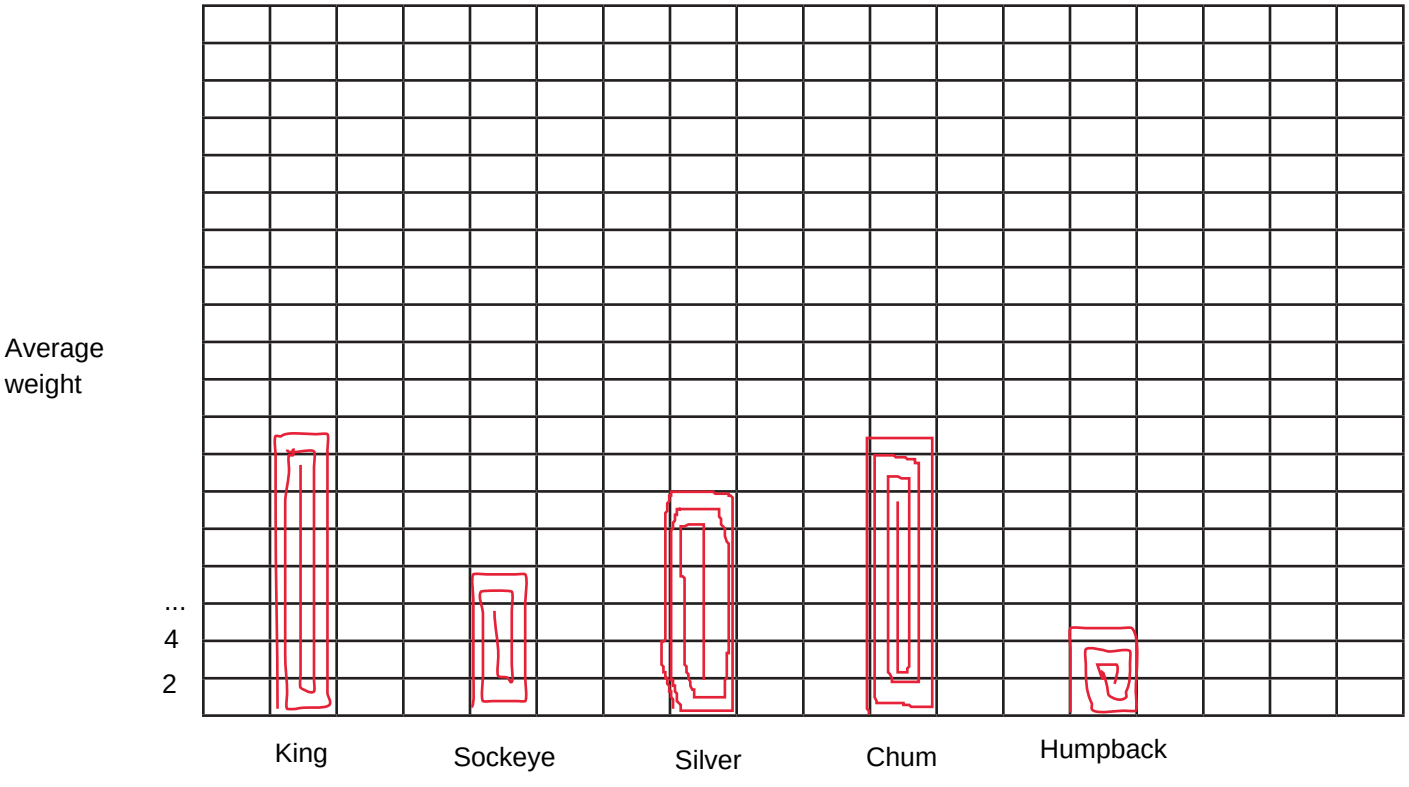
Altitude (m)	Atmos. Pressure (atm)	Altitude (m)	Atmos. Pressure (atm)
0	1.000	16,132	0.100
2750	0.750	30,901	0.010
5486	0.500	48,467	0.001
8376	0.333		

scatter bec invar is quantized

16. Choose one of the data sets in Question 15 that you selected as appropriate for a scatter plot and graph it here. Remember to label the axes.



17. Choose one of the data sets in Question 15 that you selected as appropriate for a bar graph and graph it here. Remember to label the axes.



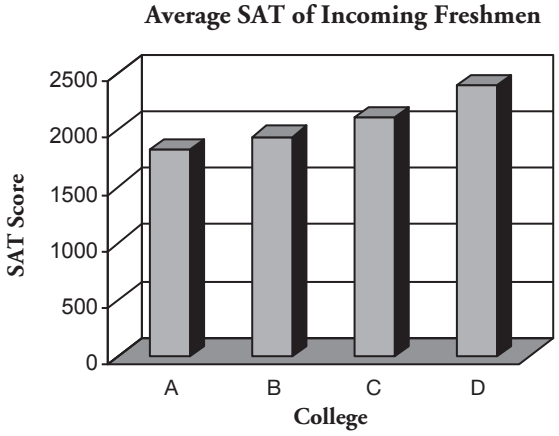
18. Send one representative of your group to another table to share the two graphs constructed in Questions 16 and 17. Does the other group agree that your graphs are properly constructed? If not, what adjustments do you think need to be made?

I dont think we can do this because we are not in class, but the graphs I made look good.

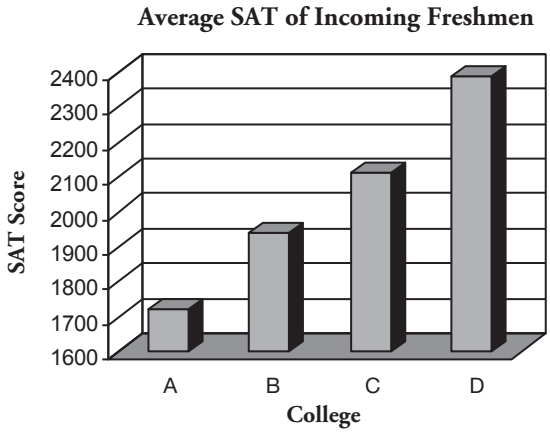
Extension Questions

Model 4 – SAT Scores

Graph F



Graph G



19. Describe the independent and dependent variables for the data that is displayed in Graphs F and G in Model 4.

    College is the independent, SAT scores is the dependent.

20. When you look at Graph F, what message is communicated by the relative lengths of the bars to prospective students about College D's average SAT scores compared to the other three schools?

    They are relatively the same.

21. When you look at Graph G, what message is communicated by the relative lengths of the bars to prospective students about College D's average SAT scores compared to the other three schools?

    There is a decent difference between college D and the other schools.

22. For each of the graphs in Model 4, estimate the average score for each college represented by the height of the bar. Is the data being displayed in the two graphs the same or different? Support your answer with evidence from the graph.

    Yes, they are the same even if Graph G's Y axis is inflated. Same variables.

23. A student takes a quick look at Graph G and says "Based on the size of these bars, it looks to me as though College D had entering freshman with SAT scores nearly four times higher than College A." Explain to this student what mistake he has made in processing the information presented in Graph G.

    Check y axis for correct numbers.