9.1 Constructing Scatter Plots

A **scatter plot** is a plot on the coordinate plane used to compare two sets of data and look for a correlation between those data sets. An **association** is a relationship or dependence between data. For example, the price of oil and the price of gasoline have a strong association. The daily price of oil and the number of penguins swimming in the ocean on that day most likely have no association at all. However, to find this association we need to make a scatter plot.

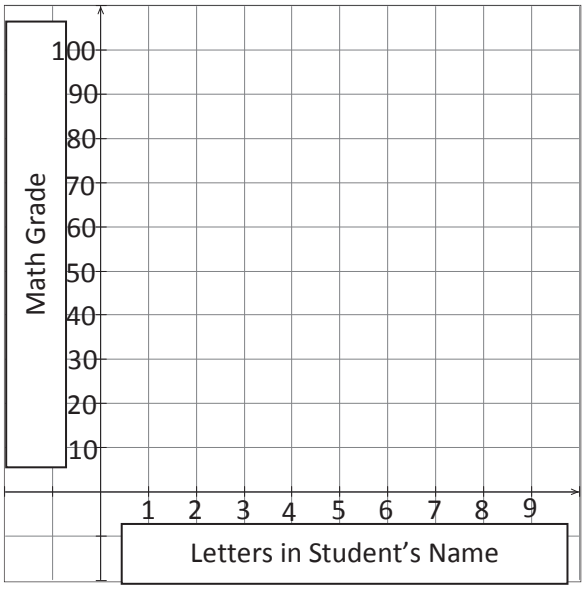
Start with the Data

Before we can make a scatter plot, we need two sets of data that we want to compare. For example, we might compare the number of letters in a student's first name and their math grade. Do people with shorter names tend to score higher in math? Do people with the lowest grades have longer names? These are questions of relationship, or correlation, that we can explore with a scatter plot once have some data. That data set might look like this:

Name	Nichole	Josiah	Kame	Gungar	Roberto	Frank	John	Herman	Sami	Daimon
Letters	7	6	4	6	7	5	4	6	4	6
Grade	58	83	61	70	31	76	81	70	72	57

Name	Yolina	Johanne	Karolinea	Kurt	Addison	Ian	Dennis	Ophelia	Kristina	Bradford
Letters	6	7	9	4	7	3	6	7	8	8
Grade	77	90	87	83	76	78	87	87	80	41

Prepare the Coordinate Plane



Now that we have our data, we need to decide how to put this data on the coordinate plane. We can let the x-axis be the number of letters in a student's name and the y-axis be the students overall math grade. Once we have decided this we should label our axes.

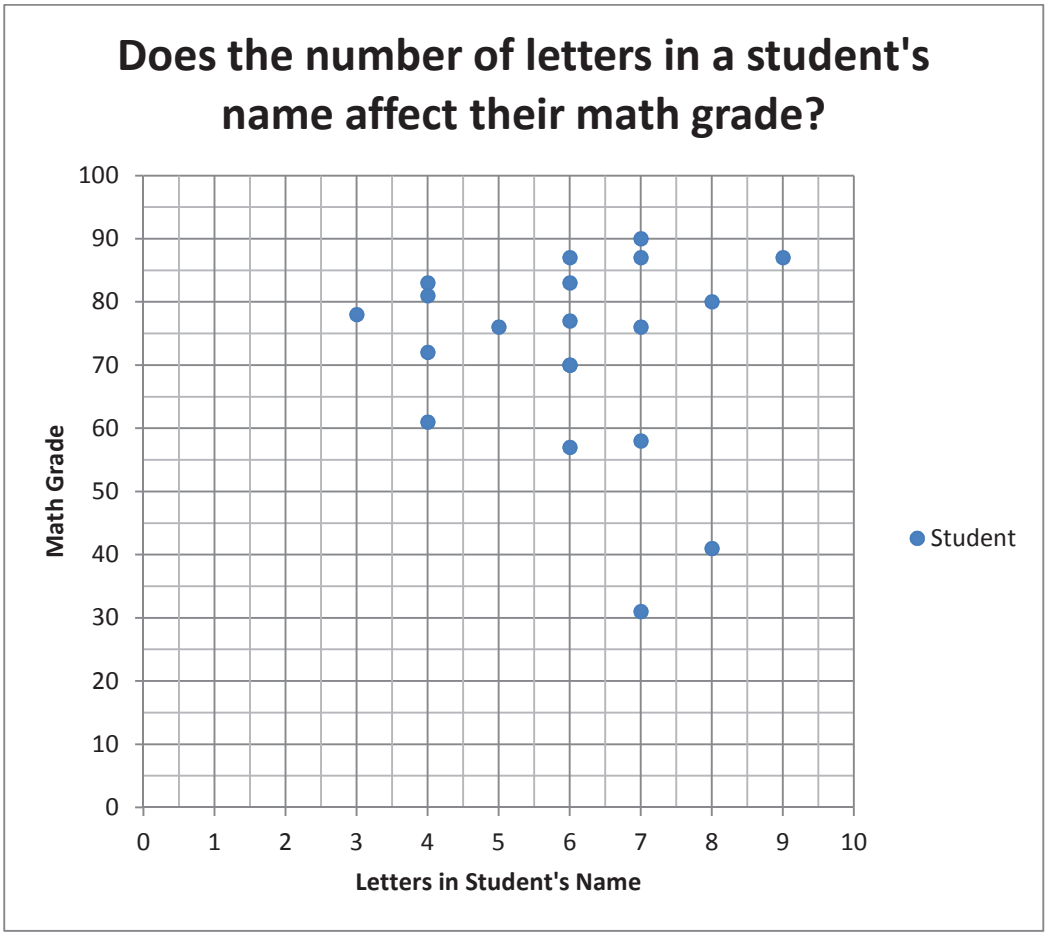
Next we'll need to decide on a scale and interval. The scale is the low to high number on the axis and the interval is what we count by. Notice first of all that we're only looking at Quadrant I because we won't have negative amounts of letters or negative grades. Since the grades can be from zero to one hundred, we might choose to count by tens on the y-axis giving us a scale of 0-100 and an interval of 10. Since the letters range from three to nine, we might count by ones on the x-axis. This gives us a scale of 0-10 with an interval of 1.

When to use a broken axis

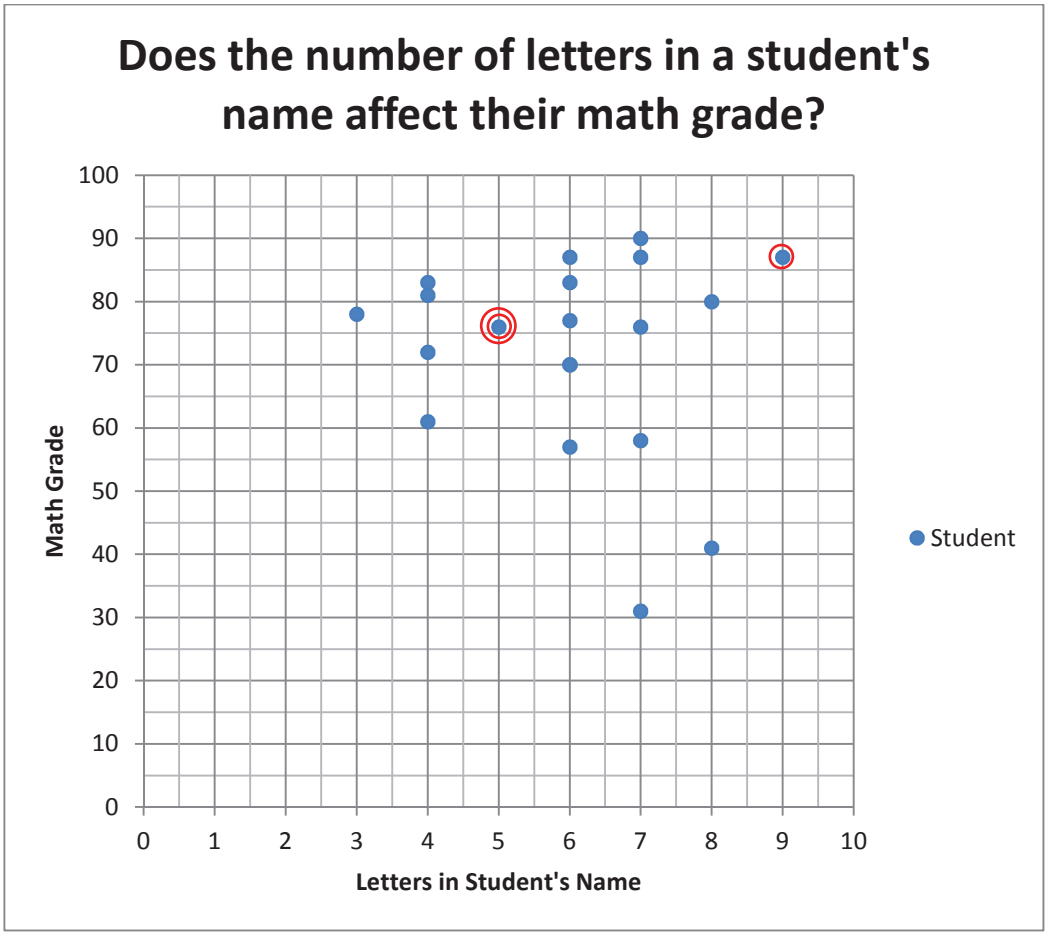
A broken axis is useful whenever more than half of the area of the scatter plot will be blank. Nobody likes to see a blank graph with all the data in one tiny area. So instead, we zoom use by using a broken axis. If the range of your data is less than the lowest data point, a broken axis may be useful. For example, in our math test situation above if everyone scored above a 60%, then we might break the y-axis and begin counting at 60. We could then count by 4's to make it up to 100%.

Plot the Points

Finally we would then plot each person on the graph. So Nicholas will be the point $(7, 58)$, Josiah the point $(6, 83)$, and so forth. Using Excel to make our scatter plot, the final scatter plot might look like the following. Notice that each dot on the graph represents a person. While the labeling is not necessary, it may be useful in some circumstances.



Many times on a scatter plot you may have the same data point multiple times. One way to represent this fact is to put another circle around the data point. Let's add a few new students to our data set: Johnathan (9 letters and 87 math score), Jacob (5 letters and 76 math score), and Helga (5 letters and 76 math score). The new graph could look like this:



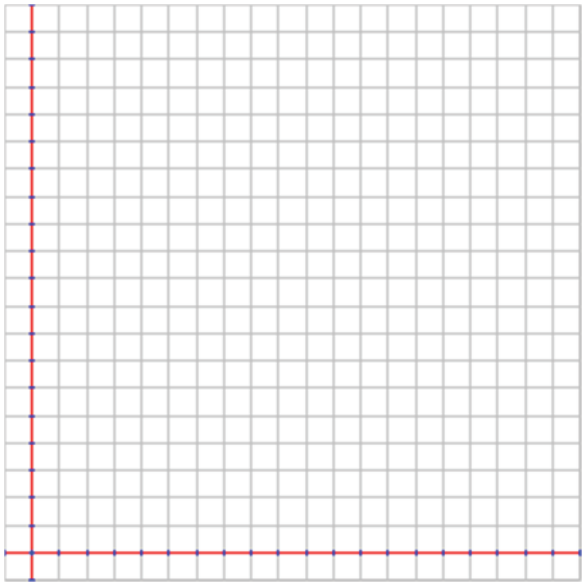
While this practice is not necessarily standard, it can be useful as a visual representation of what is happening with the data. We can more easily see the multiple data points this way. In Excel, you wouldn't get the red circles. Those would have to be put in by hand.

Lesson 9.1

Use the given data to answer the questions and construct the scatter plots.

Pathfinder Character Level vs. Total Experience Points

Level	2	3	6	9	10	11	14	15	17	20
XP	15	35	150	500	710	1050	2950	4250	8500	24000



1. Which variable should be the independent variable (x-axis) and which should be the dependent variable (y-axis)?

2. Should you use a broken axis? Why or why not?

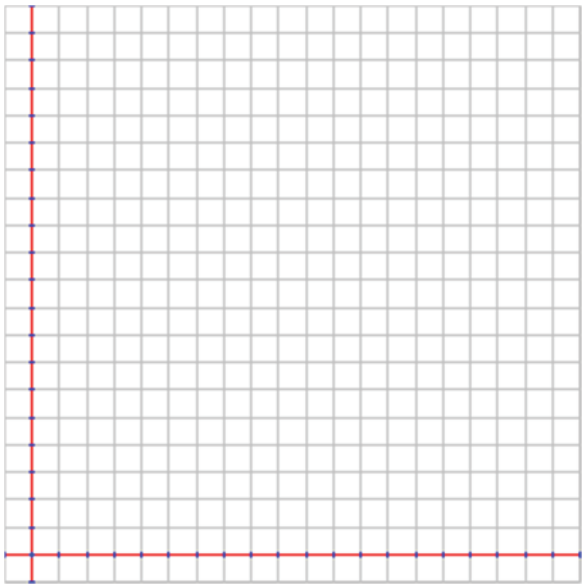
3. What scale and interval should you use for the x-axis?

4. What scale and interval should you use for the y-axis?

5. Construct the scatter plot.

Age vs. Weekly Allowance

Age	12	12	13	13	14	14	15	15	16	16
Allowance	0	5	5	8	10	15	20	20	25	30



6. Which variable should be the independent variable (x-axis) and which should be the dependent variable (y-axis)?

7. Should you use a broken axis? Why or why not?

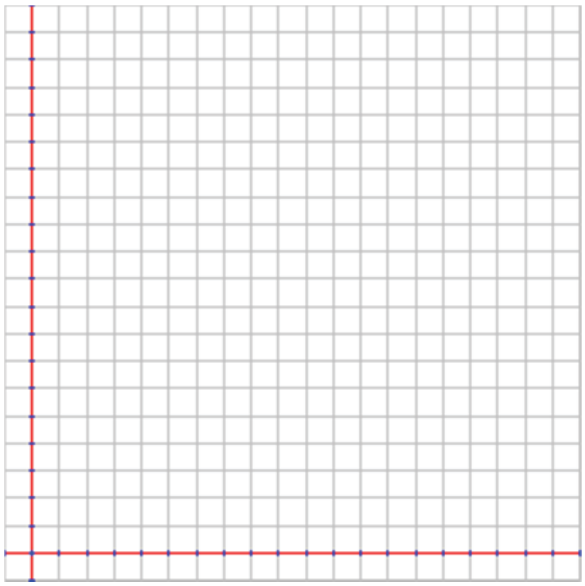
8. What scale and interval should you use for the x-axis?

9. What scale and interval should you use for the y-axis?

10. Construct the scatter plot.

Age vs. Number of Baby Teeth

Age	5	6	7	7	8	9	10	11	11	12
Baby Teeth	20	19	17	15	10	10	8	4	2	2



11. Which variable should be the independent variable (x-axis) and which should be the dependent variable (y-axis)?

12. Should you use a broken axis? Why or why not?

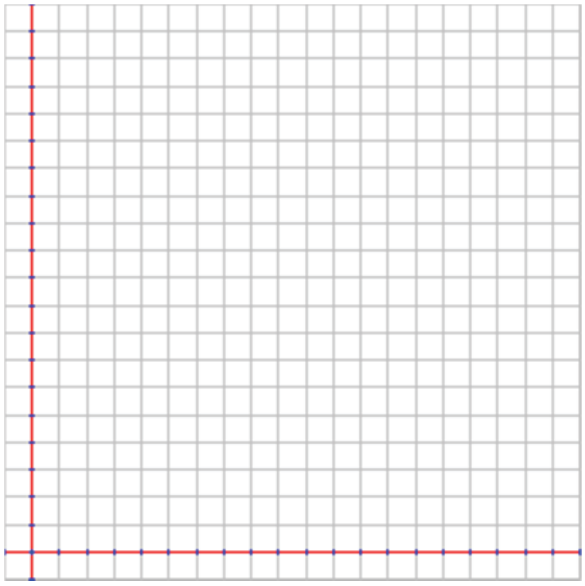
13. What scale and interval should you use for the x-axis?

14. What scale and interval should you use for the y-axis?

15. Construct the scatter plot.

Car Speed (in mph) vs. Gas Mileage (in mpg)

Speed	20	25	35	40	45	55	65	80	90	100
Mileage	25	27	28	30	31	32	30	29	25	22



16. Which variable should be the independent variable (x-axis) and which should be the dependent variable (y-axis)?

17. Should you use a broken axis? Why or why not?

18. What scale and interval should you use for the x-axis?

19. What scale and interval should you use for the y-axis?

20. Construct the scatter plot.