1 Lesson 3: Presenting Data Graphically

1.1 Types of graphs

Once data is organized and arranged, it can be presented. Graphic representations of data are called graphs, plots or charts. There are an untold number of graph and chart types, and a myriad of names for these types. We will discuss a few of the more popular or common types, and we will use the names for them in many spreadsheets.

1.1.1 Scatter Plots

The simplest type to start with are scatter plots. These are very popular in mathematics applications, and are a good choice when your data comes as pairs of numbers. (Each datum is a pair of numbers.) In a real application, the more data points drawn that can be plotted on a computer the better; in a hand drawn classroom exercise, the fewer the better.

First a middle school example:

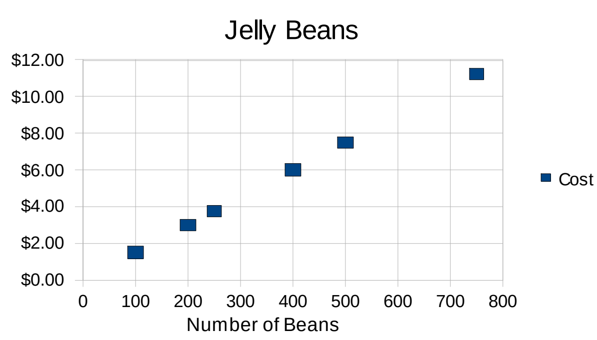
Bags of jelly beans come in various sizes, and cost various amounts of money. The cart below gives the costs of bags of various sizes.

Number of Beans	100	200	250	400	500	750
Cost	\$1.50	\$3.00	\$3.75	\$6.00	\$7.50	\$11.25

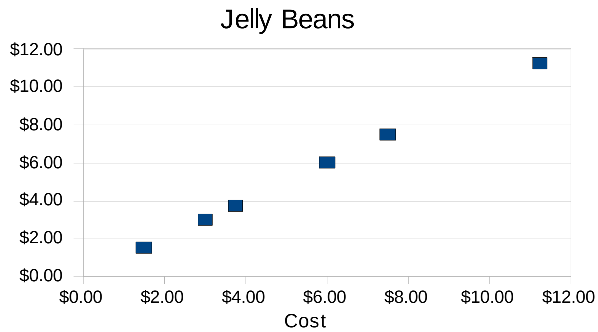
First we must make a note of the organization of the data:

- The data is organized as pairs of numbers written in columns, the first column only being labels.
- All the entries are numbers.
- The data is arranged in increasing order along each row.
- There is a functional relationship in both directions of the pairs.

A scatter plot of this data might look like



This has the cost on the vertical axis and the number on the horizontal axis. There is another choice which switches these positions.



Notice, to get this second plot, you would say that we have rearranged the data. We might have rearranged the data by switching the two rows in our data table to

Cost	\$1.50	\$3.00	\$3.75	\$6.00	\$7.50	\$11.25
Number of Beans	100	200	250	400	500	750

Even if we did not actually do it on paper, it should be considered a rearrangement.

What is happening here is that we are choosing between the two functional relationships we have available. The cost of a bag of jelly beans is in a functional relationship with the cost of the bag. The "$x$variable" is typically drawn on the horizontal axis, so we might say that the first graph illustrates this functional relationship. If we look at the second graph, the cost is on the horizontal axis. We can interpret this as illustrating the functional relationship the number of jelly beans has with the cost. All we are doing is giving a mathematical context to the only slightly different graphical representations of the same thing. Mostly the distinction is just your point of view, but mathematicians do have a their own preference when it comes to functions.

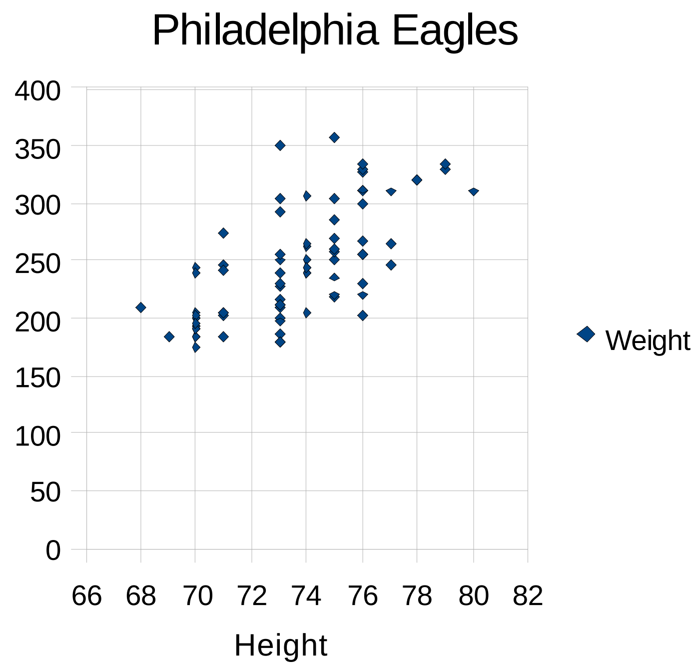
What if there is no functional relation in a table? Then a scatter plot is still a good choice. Consider the 2008 Philadelphia Eagles football team again. There are 80 players on the team mostly, but not all, big men. We can collect a table of the heights and weights of all these players. Without looking at the table, we know how it might be organized:

- The data is organized as columns of (First Name, Second Name, Height, Weight)..
- The first two entries in a datum are names, the second two entries are numbers.
- The data is arranged alphabetical order by last name.

- There is a functional relationship in both directions of the pairs.
- The players' heights are in a functional relationship with the full names.
- The players' weights are in a functional relationship with the full names.
- The players' heights are not in a functional relationship with the weights.
- The players' weights are not in a functional relationship with the full heights.

These last four observations are worth a look at. No two players have the same first and last names, so a player's full name uniquely identify him and thus both his height and weight. That is what we need for the first two functional relationships with the player's full name. Without actually looking at the data, we are only guessing that players' heights are not in a functional relationship with the weights. But them it seems rather likely that in a list of 80 people, two might have the same weight but have different heights. One example of that is enough to eliminate a functional relationship. The last observation is on more solid ground. We already saw that Andrews and Baskett are the same height, but different weights. So we know for certain that the players' weights are not in a functional relationship with their heights.

If we are interested in the player's sizes, and not the players themselves, we can rearrange the data by leaving out the players' names. (Remember dropping data is just a rearrangement.) Now this leaves a data set made up of numbers, and better still the numbers are in pairs. Unfortunately, there are no functional relationships in these number pairs. A scatter plot still works. In fact, the good side of not having a functional relationship in pairs of numbers is that there is basically only one type of chart that we can use to display that data: a scatter plot.

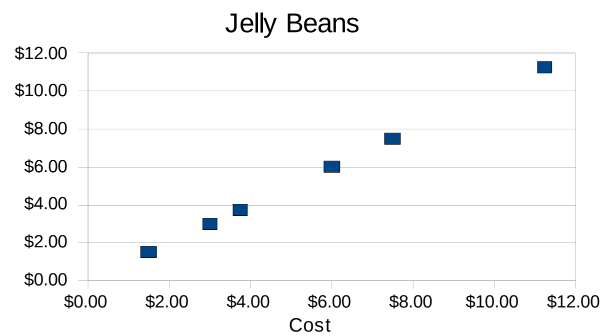


Each diamond represents a player; its place along the horizontal axis is the players height; it place along the vertical axis represents his weight. We can see that there are a number of players who are 6'1" tall, and their weights vary from about 170 pounds to about 350. So this there is not a functional relationship in this data, but the scatter chart gives a very clear picture of how large these players are. The scatter plot does not need a functional relationship is provide good visual information. (Notice how clearly the scatter plot shows the curious fact that no player is exactly 6 foot tall.)

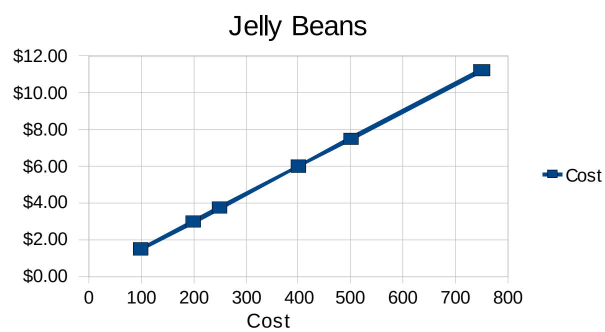
In summary, scatter graphs as good for data made up of pairs of numbers. They work whether or not there is a functional relationship.

1.1.2 Line graphs

The example of the costs of jelly beans led us to the scatter plot



There is such a strong pattern in this data, that it almost jumps off the graph. These points line up! If we were to accentuate this by drawing a line between the points, we would have the line graph



This line graph is possible because the number of jelly beans is in a functional relationship with the cost. For a line graph to work, there must be a functional relationship. The data of Eagles football players' heights and weights has nothing resembling a functional relationship, and there is no way of connecting the points in its scatter plot into a line graph. Line graphs are only possible when there is a functional relationship in pairs of numbers in the data. When the "input" data in the functional relationship has a natural order, line graphs are often a good choice.

There only has to be a functional relationship in one direction for a line chart to work. In the last lesson we considered the grades on three tests earned by a group of students.

Name	Test 1	Test 2	Test 3
April	55	71	64
Barry	63	67	63
Cindy	88	90	91
David	97	92	87
Eileen	58	55	75
Frank	90	89	96
Gena	88	100	85
Harry	71	70	71
Ivy	65	75	85
Jacob	77	70	65
Keri	75	88	85
Larry	88	92	92
Mary	95	95	100
Norm	86	82	80

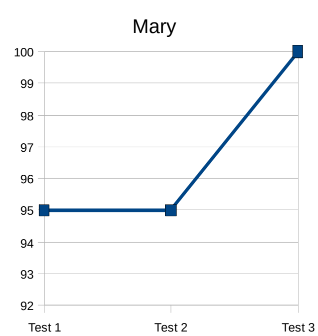
At first this data does not seem to fit in the pattern of either a scatter plot or a line plot. However, if we look at one datum from this, we have a different situation.

Name	Test 1	Test 2	Test 3
Mary	95	95	100

Here the data has the form

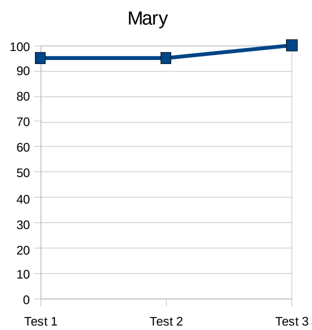
- Each datum is a pair of the form (Test #, Score)
- The score is in a functional relationship with the test number.
- The test number is not in a functional relationship with the score.

Here is a set of data that can be illustrated using a line chart.



Notice that we placed the tests along the horizontal axis. Typically the "input" item in a functional relationship is set along the horizontal axis, and the "output" item is placed on the vertical.

Now Mary is a good student, but this graph makes it look she improved quite a bit on the last test. However, she was an A student throughout, and this graph may not give the correct impression. We can change the scale of grades to correct this impression, although the result may be just as bad.

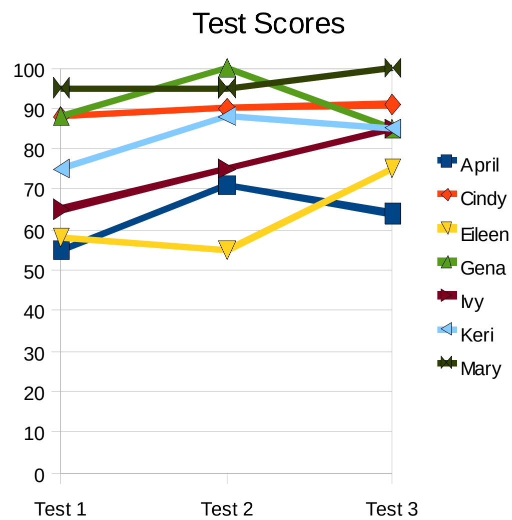


For now, we will not worry that much about adjusting a chart or graph visually, but rather concentrate on simply selecting the best type of chart. Later we will revisit the idea of formatting a chart for visual effect.

Still the scale from 0 to 100 might work better for other students, and certainly if we want to compare several students. Consider the girls:

Name	Test 1	Test 2	Test 3
April	55	71	64
Cindy	88	90	91
Eileen	58	55	75
Gena	88	100	85
Ivy	65	75	85
Keri	75	88	85
Mary	95	95	100

We can give each girl her own line graph, and place them all in one chart.

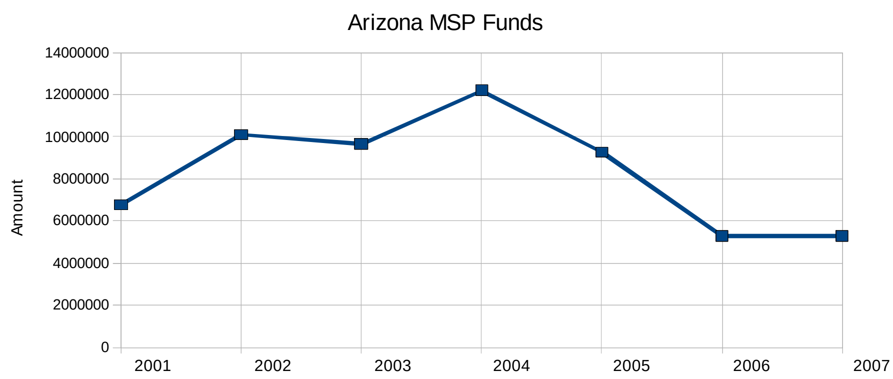


For another example, consider the Department of Education program called Math and Science Partnerships. The Department distributes funds to the states for specific projects, and the funds received by Arizona over since 2001 are given in the table:

	2001	2002	2003	2004	2005	2006	2007
Arizona	$6,759,013	$10,114,346	$9,655,054	$12,202,519	$9,278,899	$5,291,697	$5,290,464

- The data is organized as pairs written in columns; and so each row is a datum.
- The data can be thought of as made up of pairs of numbers.
- The money is in a functional relationship with the year.
- The data is arranged in increasing order of years.

Since we have pairs of numbers, we could use a scatter plot. But there is a functional relationship, so we could just as well use a line plot. Since the years have a natural order, a line plot is a good choice:



Notice, the only lines in this graph are the ones we added. All the line graph does is connect points in the scatter plot with straight line segments. In many cases, a line graph is a visual presentation of the data more than a mathematical representation.

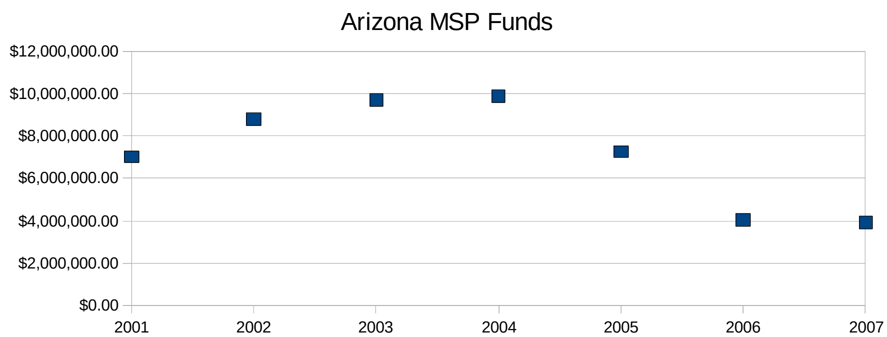
It is possible, and sometimes useful, to smooth out the plot by curving the lines through the points in the scatter plot. The results is still just a variation of the line graph. There are times when shading in the area below the lines in the line graph is a good idea for visual effect, but mathematically such an "area" chart is only a variation of a line graph. Line graphs are very useful, and there are any number of major and minor variations of their basic idea that can come in handy.

1.1.3 Column charts and bar charts

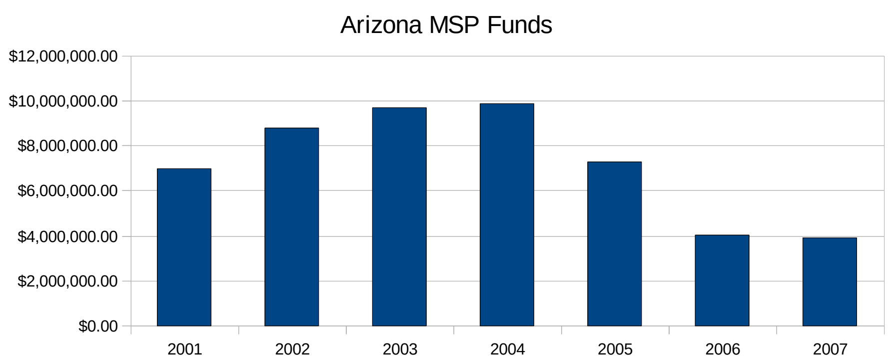
Consider the Arizona MSP data a second time.

	2001	2002	2003	2004	2005	2006	2007
Arizona	$6,759,013	$10,114,346	$9,655,054	$12,202,519	$9,278,899	$5,291,697	$5,290,464

The amount of money is in a functional relationship with the year. The lines between points in the Arizona MSP funding chart give the impression that the funds arrived uniformly during the year. This is not necessarily wrong, but it might not be the impression we want to give. A simple scatter plot looks too sparse, and may be hard to read.

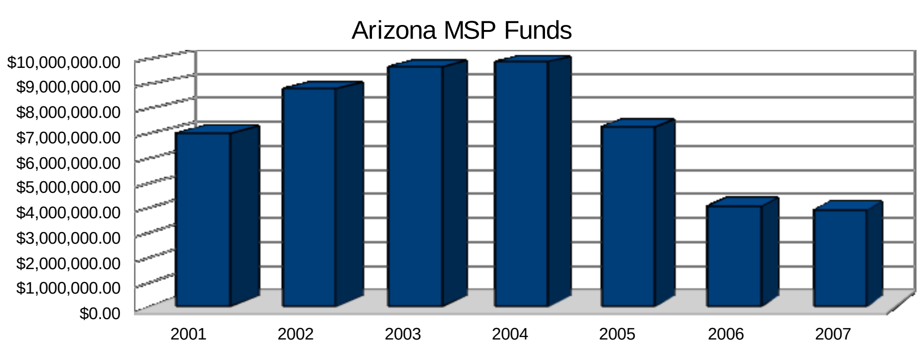


But because there is a functional relationship, and we can use a column chart.

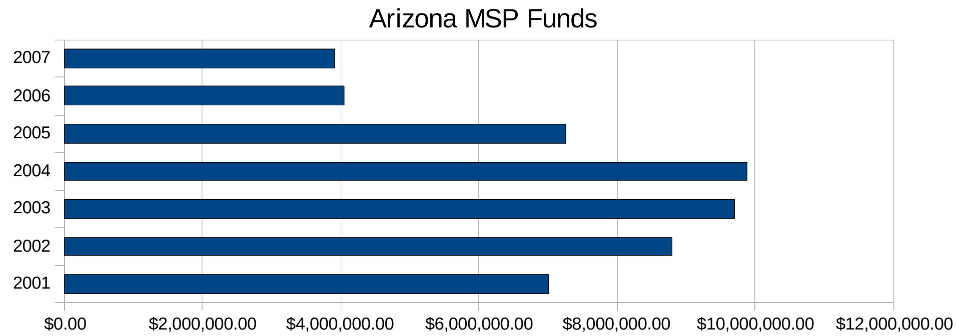


The nice thing about column charts is that the "input" data in the functional relationship does not need to be numbers. You can use column charts when the data along the horizontal axis are numbers, names, words, or anything. These data need a functional relationship with numbers, but the results can be quite nice.

There is no rule that the columns need to be simple; all sorts of shapes can be used. Three dimensional bars work



We could just a well use cylinders, pyramids, cones, dollar signs, people in cap and gown, or basically anything we can draw to scale. We can draw it sideways and turn it into a bar chart:



Column and Bar charts work very well when there are only a few data items to illustrate, and even more so when the numbers represent a quantity of some sort. They are less useful when one numerical value is significantly larger than all the others. It is also very important to note the scale used when reading one of these charts. The scale used in a column chart may be chosen to visually exaggerate a difference in the data. This is not necessarily bad, but it does require the reader to be alert. Any column or bar chart that does not have clear numerical labels in its scale should be viewed with scepticism

1.1.4 Pie Charts

Pie charts are a great choice for displaying percentages. Take, for example, the MSP funding again:

	2001	2002	2003	2004	2005	2006	2007
Arizona	$6,759,013	$10,114,346	$9,655,054	$12,202,519	$9,278,899	$5,291,697	$5,290,464

Note again: The data is organized as pairs written in columns; and so each row is a datum.

- The data can be thought of as made up of pairs of numbers.
- The money is in a functional relationship with the year.
- The data is arranged in increasing order of years.

Before we build a pie chart, we need to rearrange the data. First, we add a total column

	2001	2002	2003	2004	2005	2006	2007	Total
Arizona	$6,759,013	$10,114,346	$9,655,054	$12,202,519	$9,278,899	$5,291,697	$5,290,464	$58,591,992

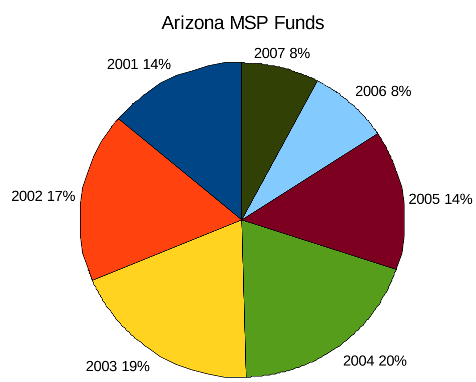
A pie chart needs a functional relationship to work, but more important, the output of the function needs to be a number. In this example, the amount of funds is in a functional relationship with the year. Each year has its own funding amount, and that will be represented with its own slice of a pie. The size of the slice is proportional to the amount of money received in that year. To draw the pie chart, we need to rearrange our data to include percentages:

	2001	2002	2003	2004	2005	2006	2007	Total
Arizona	$6,759,013	$10,114,346	$9,655,054	$12,202,519	$9,278,899	$5,291,697	$5,290,464	$58,591,992
Percentage	12%	17%	16%	21%	16%	9%	9%	100%

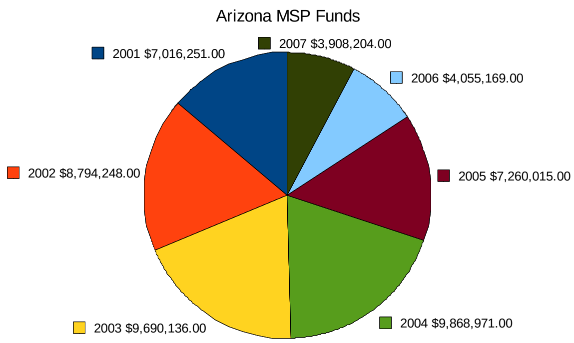
In a pie chart, the pie is made up entirely by the slices. In our case, each slice represents the funding in one year. The area of each slice is the same percentage of the whole pie as the percentage of the amount obtained in that year. Now the areas in pie slices are proportional to the degrees in their angles. Since a whole pie comes from an angle of 360° we rearrange the table one more time to give angles of the corresponding percentage:

	2001	2002	2003	2004	2005	2006	2007	Total
Arizona	$6,759,013	$10,114,346	$9,655,054	$12,202,519	$9,278,899	$5,291,697	$5,290,464	$58,591,992
Percentage	12%	17%	16%	21%	16%	9%	9%	100%
Angle	42°	62°	59°	75°	57°	33°	33°	360°

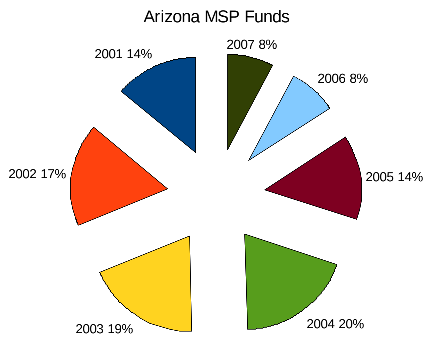
After this, it is just a matter of how you like to draw it. A protractor helps, but not as much as a computer program:



There are lots of ways to draw and label the same data:

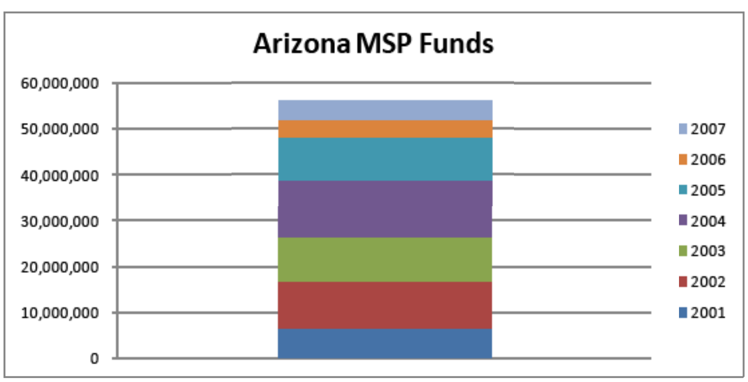


Or if you want to split it up a bit,



We can easily see that the largest percentage of funds came in 2004, and the least in 2006.

Now a pie chart does not need to be a pie, it can be a rectangle:



1.1.5 Stem and Leaf Plots; Frequency Graphs and Histograms

There are many other types of graphs than scatter plots, line graphs, column and row charts, and pie charts. There are also various versions of each of these types. There are plenty of choices for representing data graphically. We do not have time to do more than outline the basic types above. There is, however, one last general type of graphic associated with data. we will consider three forms of this general type: Stem and Leaf Plots, Frequency Graphs, and Histograms.

A "Stem and Leaf Plot" provides a graphic way to assemble or collect data in the form of numbers. Many times a completed stem and leaf plot will give at least some visual information about the data. In truth, a stem an leaf plot is tabulation of data more than a graphic representation of the data. However, it has a graphic element to it, and it can be seen as a primitive form of a histogram.

In its simplest form a stem and leaf plot is used on data consisting of two digit numbers. In this case the "stem" of the graph is the first digit of the number, and the "leaf" is the second. The stem numbers are most often stacked vertically, and the leaves are placed to one side horizontally.

Consider our list of test scores.

April	55
Barry	63
Cindy	88
David	97
Eileen	58
Frank	90
Gena	88
Harry	71
Ivy	65
Jacob	77
Keri	75
Larry	88
Mary	95
Norm	86

If we are interested in the test more than the student's scores, we rearrange the table by dropping the names so that each datum is just a test score. All we care about is the scores, so we rearrange the data by forgetting about the names in these rows.

$$55, 63, 88, 97, 58, 90, 88, 71, 65, 77, 75, 88, 95, 86.$$

We end up with the most straightforward of all data sets: a simple list of numbers.

Stem and Leaf Plots only work for data sets made up of numbers. Using the first digit as the stem and the second digit as the leaf, we get a plot like

Stem	Leaf
5	5 8
6	3 5
7	1 5 7
8	6 8 8 8
9	0 5 7

We can easily see where Barry's score of 63 appears in this plot, its stem is 6 and leaf is 3. Unfortunately, we no longer know it belongs to Barry. As we can see, this works quite well on data sets made up of two digit numbers. But it is easy to see how to adapt it to other types of numbers.

Consider the basic idea behind this stem and leaf plot. We start with a data set that is either given or being collected. The datum in the set are single objects, in the example, two digit test scores between 50 and 99. When we chose to use the first digit of each number as a stem, we have actually broken the data into classes: 50-59, 60-69, 70-79, 80-89, and 90-99. Once these classes have been assigned as the stems, the rest of the datum became the leaf. Since in the example the leaves are digits (numbers) we used the digits themselves to represent the leaves, and arranged them in increasing order. When numbers are involved, proper etiquette calls for the leaves to be listed in order.

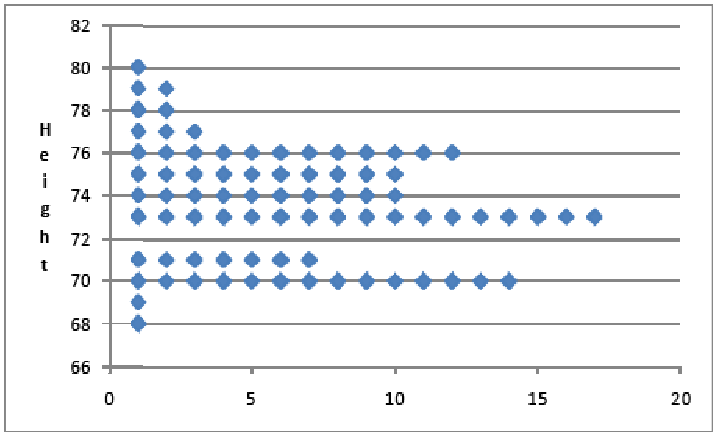
This general description of a stem and leaf plot leads us to a "Class and Frequency Plot." Data is divided into classes, and the classes play the role of the "stems." Since the remainder of a datum may not be a digit, or even a number, we might use any symbol to denote each leaf. Each piece of data appears as a symbol in the correct stem. In our example of test data, we might end up with something like

Stem	Leaf
5	$* *$
6	$* *$
7	$* * *$
8	$* * * *$
9	$* * *$

We definitely lose information this way, and this is probably not an example where this is the best idea. However, there is information left that might be useful. The number of leaves in each stem shows the frequency that data occurs in this class. Thus the "Stem and Leaf Plot" has become a "Class and Frequency Plot."

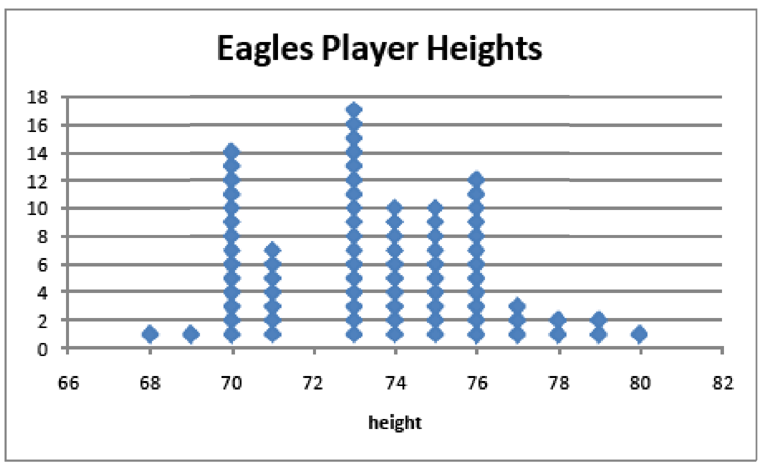
Consider a better example, namely the Eagles football team. Suppose that we have collected data in the form (First Name, Last Name, Height, Weight.) for all 80 players. We notice that the players are all between 68 and 80 inches tall. We will assign every measurement in this range as a class. The classes refer to only on part of each datum, and the rest of the datum is a bit complicated.

Still we will just denote the rest of each datum with a blue ♦. The result is a true class and frequency graph:

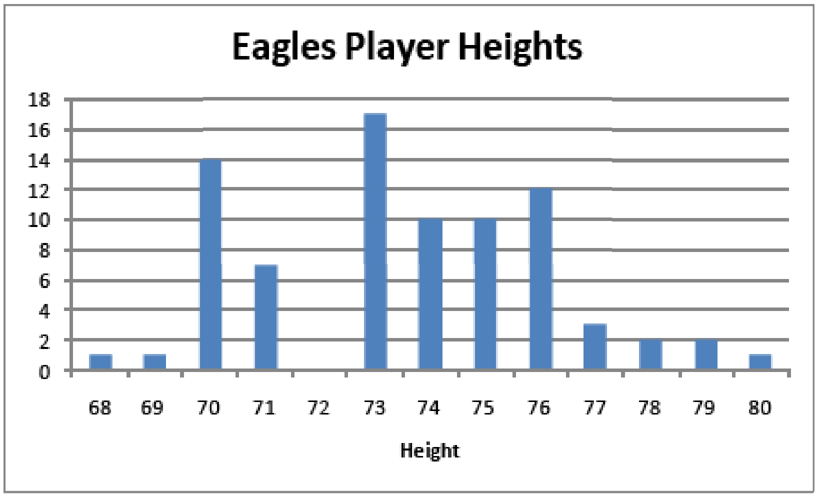


We easily see that 3 players are 77 inches tall.

If we tilt it on its side, it looks like

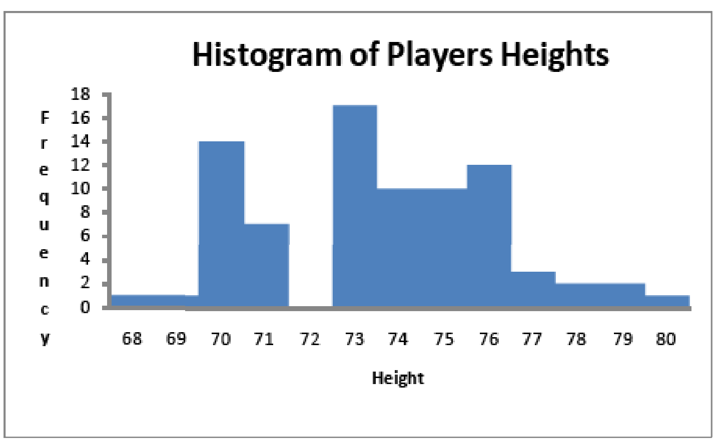


If we merge the diamonds into columns, we get

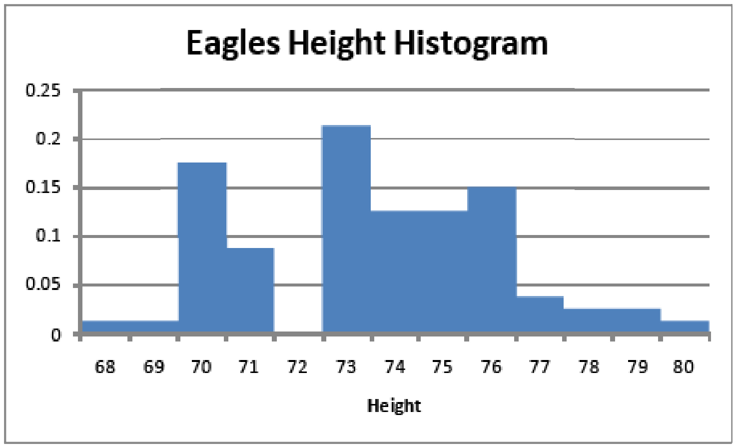


Notice all of these are class and frequency graphs drawn as columns or rows. Thus class and frequency graphs can appear as bar charts and column charts. In this example, the classes are the heights and are listed across one axis. The columns measure a count of the number of times that a datum falls in a particular class; that is to say, it gives the frequency of that class in the whole.

Once we have a class and frequency chart, we are very close to our last type of chart, a histogram. In a proper histogram, we also select classes, and separate the data into those classes. Again we highlight the frequency with which each class occurs in the data. In a histogram, we use a shaded area above a class label - rather than a height - to indicate the frequency. Usually, but not strictly, the scale marking the frequency is either a percentage or a proportion out of 1. To emphasize the area aspect of the graph, regions of the graph are run together unless a class has a zero frequency of occurrence.



Here we used the actually numbers to scale the frequencies. When the units of the frequencies are uniform like this, there is no real distinction between the height and area. Thus relabeling to a proportion of 1 does not change the shape of the graph;



Prepared by: Daniel Madden and Alyssa Keri: May 2009