

Graphical Display Examples

1. Frequency distributions

While frequency distributions are not graphical in nature, they do provide a useful means of displaying raw scores in a distribution of scores. For example, below are final course average grades in an introductory educational research class (rounded to the nearest whole number).

Final Course Average Grades in Introductory Educational Research

85	74	64	93	65	83	81	84	72
91	74	75	84	92	87	82	79	86
88	93	85	72	67	75	96	78	69

The numbers above present a disorganized picture of class performance, but they can be presented in a manner that is easier to understand such as through frequency displays as illustrated below.

This first example shows two command lines for the statistical analysis software Stata (for more information on this software, see www.stata.com). These commands were used to open a data file and to generate a frequency table with frequencies, relative frequencies, and cumulative frequencies.

```
. use "C:\Documents\GSU\COURSES\EDUR 8131\graphical display examples.dta",
clear
. tabulate Grade
```

Final Course Grade Average	Freq.	Percent	Cum.
64	1	3.70	3.70
65	1	3.70	7.41
67	1	3.70	11.11
69	1	3.70	14.81
72	2	7.41	22.22
74	2	7.41	29.63
75	2	7.41	37.04
78	1	3.70	40.74
79	1	3.70	44.44
81	1	3.70	48.15
82	1	3.70	51.85
83	1	3.70	55.56
84	2	7.41	62.96
85	2	7.41	70.37
86	1	3.70	74.07
87	1	3.70	77.78
88	1	3.70	81.48
91	1	3.70	85.19
92	1	3.70	88.89
93	2	7.41	96.30
96	1	3.70	100.00
Total	27	100.00	

The next display shows these same data, but now these data are grouped by letter grade ranges, i.e., 60 to 69 is D, 70 to 79 is C, and so on. This grouped frequency display allows one to quickly determine performance levels of the class by grade ranges. For example, 5 students earned A's (5 students in the average grade range of 90 to 99), 10 students earned B's, etc.

```
. egen Grade_cat = cut(Grade), at(50 60 70 80 90 100)
. tabulate Grade_cat
```

-> tabulation of Grade_cat

Grade_cat	Freq.	Percent	Cum.
60	4	14.81	14.81
70	8	29.63	44.44
80	10	37.04	81.48
90	5	18.52	100.00
Total	27	100.00	

2. Stem-and-leaf Displays

Similar to frequency displays are stem-and-leaf displays. These displays show frequency by building “leaves” from the raw scores. The longer the leaf, the greater the frequency associated with that set of scores. Reading the display below, note that one student scored a 64, which is displayed as “6* | 4,” and three students scored between 65 and 69. Their specific scores were 65, 67, and 69, and these are displayed as “6. | 579.”

```
. stem Grade
Stem-and-leaf plot for Grade (Final Course Grade Average)
```

```
6* | 4
6. | 579
7* | 2244
7. | 5589
8* | 12344
8. | 55678
9* | 1233
9. | 6
```

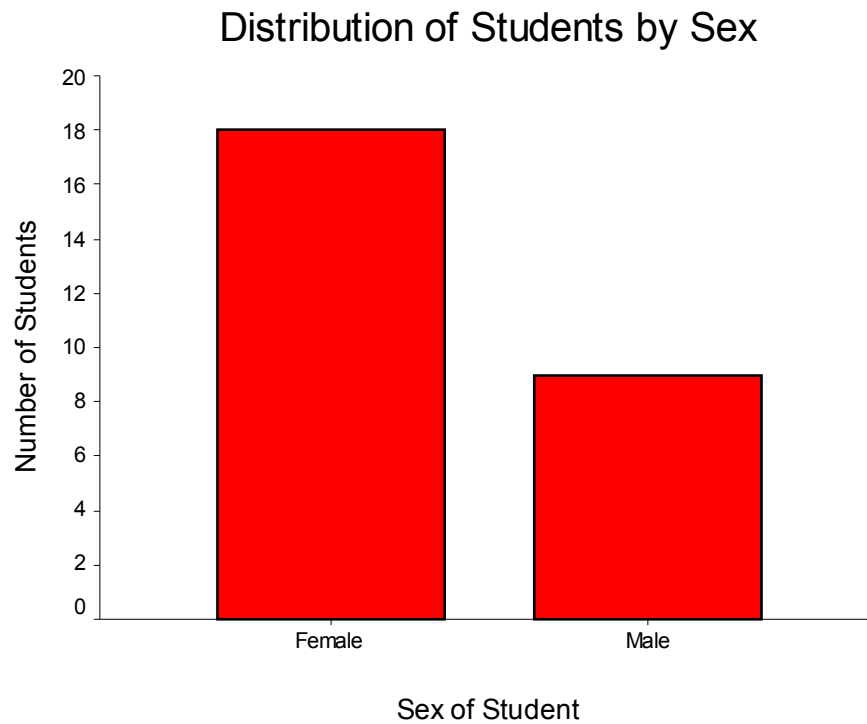
This version shows scores grouped by letter grade ranges (e.g., 60 to 69, 70 to 79, etc.).

```
. stem Grade, digits(2)
Stem-and-leaf plot for Grade (Final Course Grade Average)
```

```
6* | 4579
7* | 22445589
8* | 1234455678
9* | 12336
```

3. Bar Charts

Bar charts are traditionally used to display frequency counts for qualitative variables. Below is an example showing the sex distribution of students in the educational research class. As the frequency display below the bar chart shows, there were 18 females and 9 males enrolled in the class.

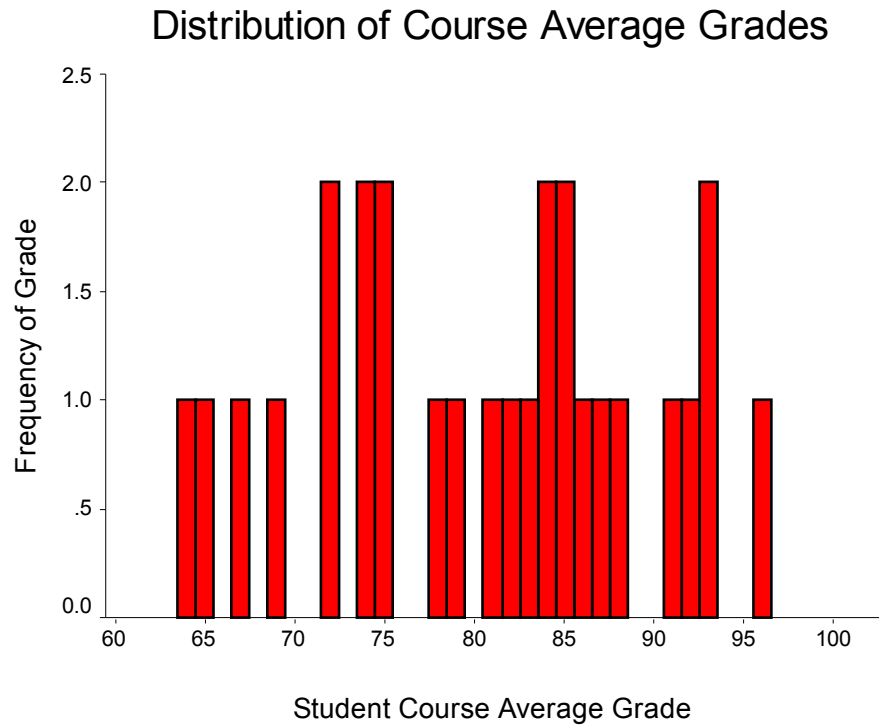


```
. tabulate Sex
```

Sex of Student	Freq.	Percent	Cum.
f	18	66.67	66.67
m	9	33.33	100.00
Total	27	100.00	

4. Histograms

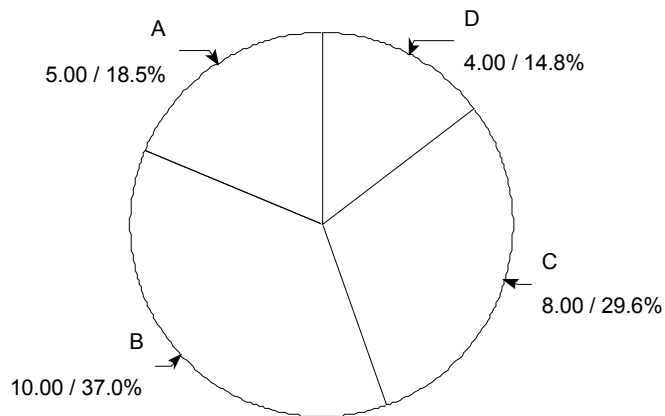
Similar to bar charts and stem-and-leaf displays, histograms may be used to show frequency information for quantitative variables. The primary difference between histograms and bar charts is that histograms are designed for quantitative data so the bars are allowed to touch when consecutive scores are presented. When gaps are present between bar in a histogram, that signals a frequency of zero for that particular score. Below is an example of a histogram for student grades. Smooth histograms are often used to present distributional shapes such as normal, F, t, chi-square, etc.



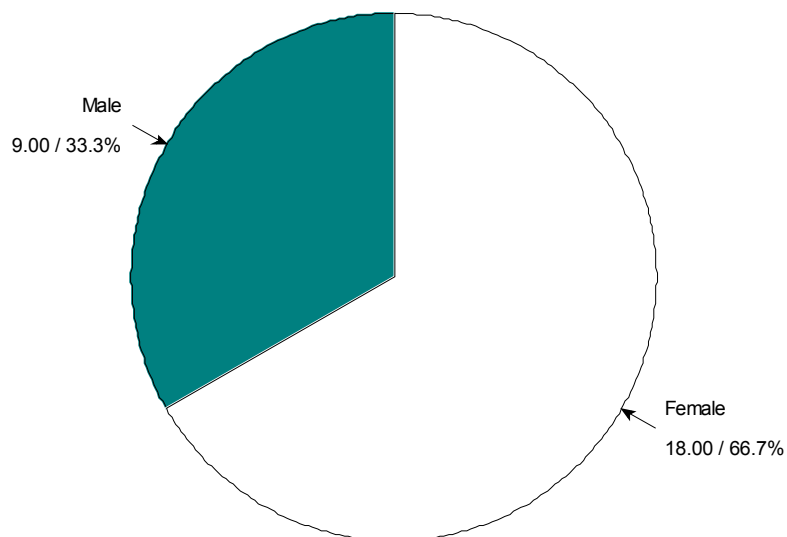
5. Pie Charts

Pie charts are another commonly used tool to display distributions. Pie charts can be used to display both qualitative and quantitative data. The first example below shows course letter grade distribution and the second shows sex distribution.

Letter Grade Distribution in Educational Research Course



Distribution of Students by Sex

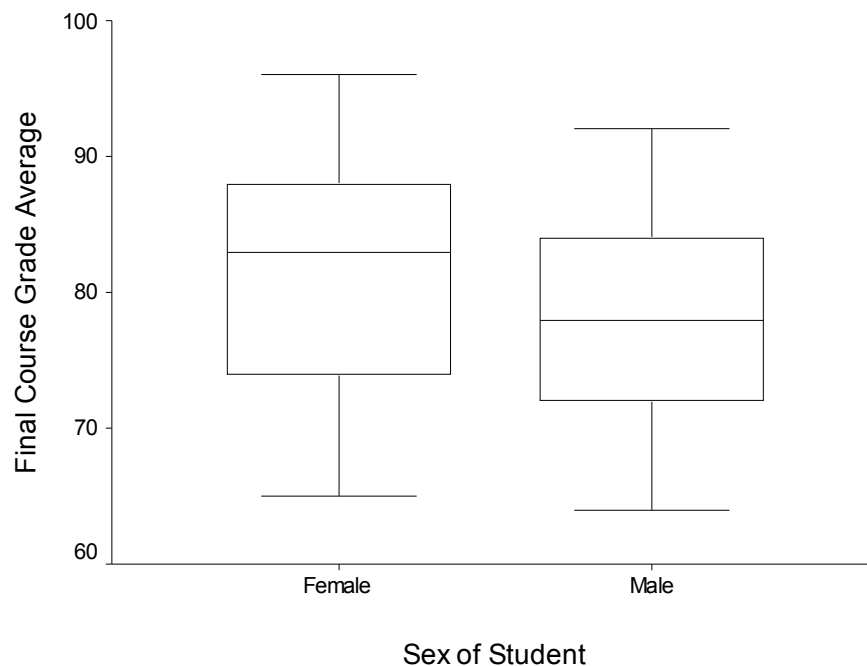


6. Box-and-whisker (or simply Box) Plots

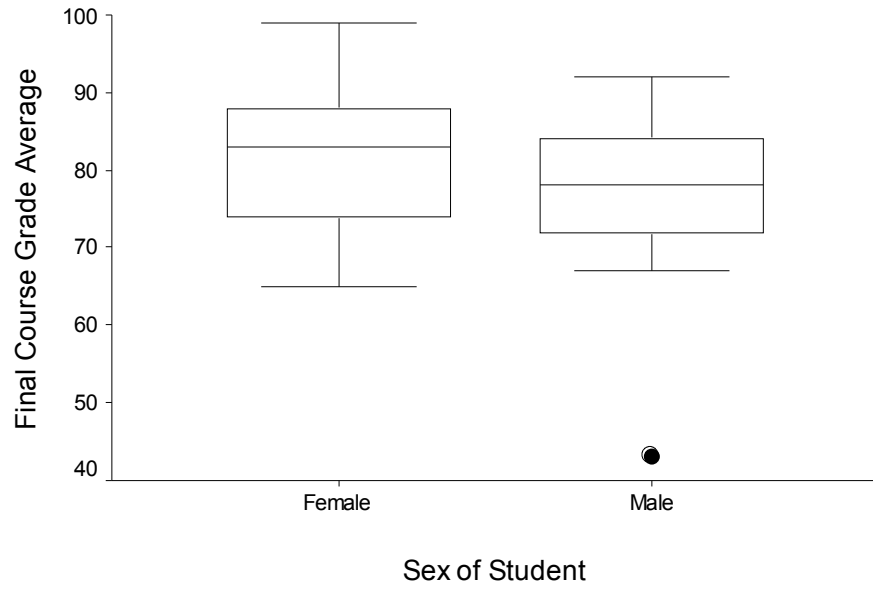
These graphs are designed to display several summary indicators of data. For example, for females, the bottom of the box shows the score at the 25th percentile (symbolized as P_{25} or Q_1 , which is roughly 74 in this sample); the top of the box is the 75th percentile (P_{75} or Q_3 , a score in this sample of about 88); and the thick line in the middle of the box represents the median (50th percentile, P_{50} or Q_2). Note that the box is designed to describe the middle 50% of scores in the distribution.

The whiskers extending from the box may represent several different things depending upon how they are implemented for given software. For the example listed below, the whiskers appear to show the upper and lower range for the sample of scores. The the bottom whisker shows the lower range for the distribution of scores (a lower score of about 64); and the top whisker shows the upper range for the distribution of scores (a top score of about 96). In some software applications, whiskers extend to P_{10} and P_{90} , and any scores beyond this range are represented as dots. The second box plot below illustrates a score, denoted by the black dot, that extends below the range of P_{10} .

Box Plot of Grade Distribution by Sex

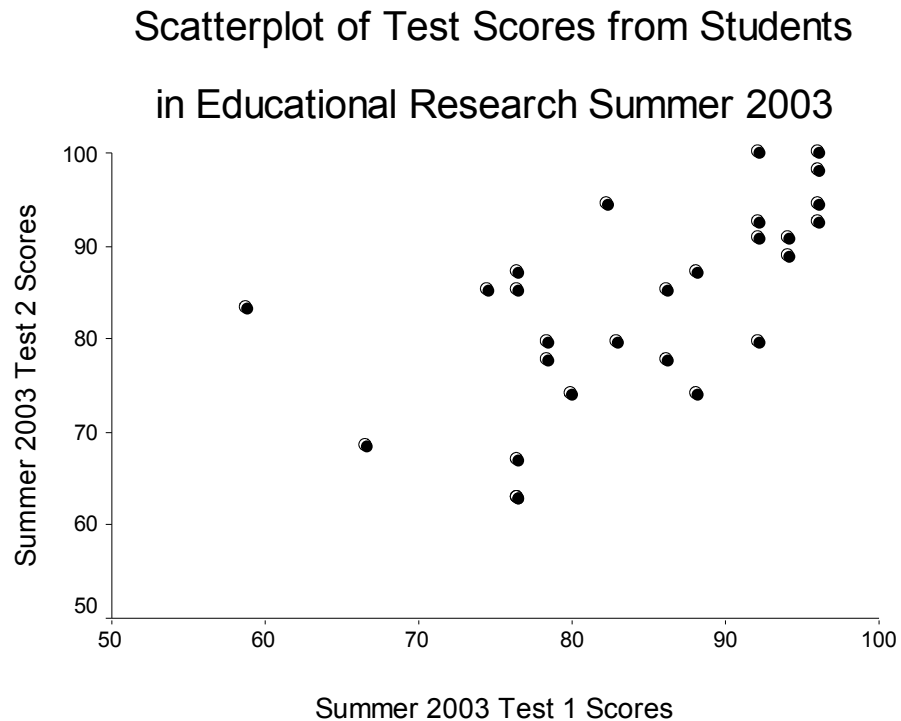


Box Plot of Grade Distribution by Sex
(with one extreme score added)

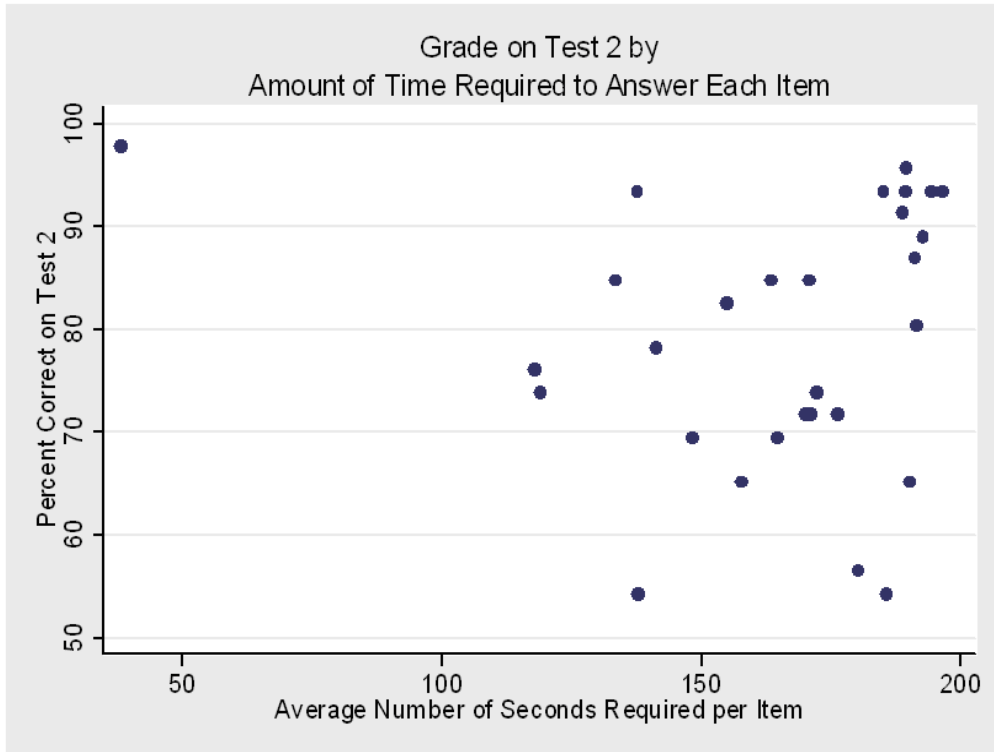


7. Scatterplots

These graphs are useful for displaying the nature of relation between two quantitative variables. As the first scatterplot shows, there is a positive, linear trend between scores from tests 1 and 2 in educational research during the summer of 2003. Students who did well on test 1 tended also to perform well on test 2; similarly, those who performed poorly on test 1 also tended to perform poorly on test 2. There are, however, several exceptions to this trend. Note the student who scored just under 60 for test 1, but score over 80 for test 2. This student is the isolated dot to the left of other dots in the scatter.



The next scatterplot, displayed below, shows information pertaining to student performance on a test in educational research. The two variables considered are the average number of seconds spent per item completing the test and test score. The scatter of data to the right of the graph shows a slight positive relation between time spent on items and test score. Generally speaking, those students who spent more time per item tended to perform better on the test, although this pattern is not strong. Most students took between 120 and 195 seconds to answer each item (that's 2 to 3.5 minutes per item). There is one very clear exception to these data and that exception is symbolized by a student who spent an average of 38 seconds per item and scored 98% correct on this test. This student's performance represents what is known as an outlier, an observation that is clearly discrepant from other observations (data) in the distribution of scores.



The next scatterplot displays data from an agricultural experiment in which grapefruit were treated with two types of fungicides and with varying amounts of the active ingredient (copper). The outcome of interest is the severity of the infection on grapefruit. The line in the graph represents a prediction line and can be used to estimate the change in severity of infection according to differing amounts of copper used.

