

Stat 150 - Day 6 Graphs in Minitab

Example 1: Pursuit of Happiness

The General Social Survey (GSS) is a large-scale survey conducted in the U.S. every two years. One of the questions asked concerns how happy the person feels with his/her life. Results for the 2004 survey are:

Very happy	Pretty happy	Not too happy
405	726	175

The simplest graph with a categorical variable is a **bar graph**.

Type the category names/labels into one column in Minitab and the counts into another column. Also name these columns by clicking in the gray box above row 1. Then select `Graph > Bar chart` from the menu. Then change “bars represent” to “values from a table” and keep the other choice on “simple.” Click OK. Then double click on the appropriate column of counts to put it in the “graph variables” box, and double-click on the column of category names/labels for the “categorical variable” box. Click on “chart options” and select “show Y as percent.” Click OK twice.

- What does the graph reveal about Americans’ self-reported levels of happiness?

Example 2: Pursuit of Happiness (cont.)

The GSS also asked people about their political viewpoints. The results of happiness level and political viewpoint are summarized in the following two-way table:

	Liberal	Moderate	Conservative
Very Happy	90	128	187
Pretty Happy	177	291	258
Not Too Happy	51	76	48

Open the Minitab worksheet `happiness.mtw`, available under the “Datasets” link from our course webpage. Notice that the first two columns of this worksheet contain the raw data, one row for each of the 1306 people surveyed and one column for each of the two variables.

Select `Stat > Tables > Cross-tabulation and chi-square`. Enter `c2` for rows and `c1` for columns. Click OK, and the original table (with one small tweak) should be reproduced. Then repeat these steps, but click on “column percents” instead of “counts.”

- What does the table of column percents reveal about any association between happiness levels and political viewpoint?

To display the relationship between two categorical variables, use a **segmented bar graph**.

Select Graph> Bar chart. Keep the first option on “counts of unique values” and select “stack.” Click OK. Double-click on c1 and then on c2. Click on “chart options” and choose “show Y as percent” and then “within categories at level 1.” [A segmented bar graph can also be made from the table of counts: Choose Graph> Bar chart, and then change “bars represent” to “values from a table” and also change to “Two-way table” and “Stack.” Click OK. Then enter c6–c8 in the “graph variables” box and c5 in the “row labels” box. Click on “chart options” and select “show Y as percent” and “within categories at level 1.” Click OK twice.]

- Does this graph reveal any relationship between political viewpoint and happiness level?

Example 3: Rower Weights

The Minitab worksheet `rowers04.mtw` contains the weights of rowers on the U.S. men’s Olympic rowing team in 2004.

The simplest graph for a quantitative variable is a **dotplot**.

Choose Graph> Dotplot, and select c2 as the graph variable. Click OK.

- What does this graph reveal about these rower weights?

The shape of this graph is skewed to the left, because the values are clustered to the right (high values) with a long tail to the left (low values).

Example 4: Geyser Eruptions

The data in the Minitab worksheet `geyser.mtw` contain times (in minutes) between eruptions of Old Faithful geyser over a period of two weeks in August of 1978.

A similar display to a dotplot, especially useful with large datasets, is a **histogram**.

Open the worksheet, and choose `Graph> Histogram`. Select `c1` as the graph variable. Click OK.

- Describe what this graph reveals about the distribution of inter-eruption times. In particular, what unusual shape does the histogram reveal?

Example 5: February Temperatures

The Minitab worksheet `FebTemps.mtw` contains the daily high temperatures for three cities in February of 2006. Open this worksheet. Then select `Graph> Dotplot` and select “multiple Ys, simple.” Enter `c2`, `c3`, `c4` as the graph variables, and click OK.

- What do the dotplots reveal about which of these three cities tends to have the highest temperatures in February? Which tends to have the lowest temperatures?

- What do the dotplots reveal about which of these three cities has the most variability in February temperatures? Which has the least variability?

Example 6: Cloud Seeding

Scientists have long investigated whether humans can help nature to produce more rainfall. In one study researchers in southern Florida explored whether injecting silver iodide into cumulus clouds would lead to increased rainfall. On each of 52 days that were judged to be suitable for cloud seeding, a target cloud was identified and a plane flew through the target cloud in order to seed it. Randomization was used to determine whether or not to load a seeding mechanism and seed the target cloud with silver iodide on that day. Radar was used to measure the volume of rainfall from the selected cloud during the next 24 hours. The results (measured in volume units of acre-feet, “height” of rain across one acre) are presented below (from Simpson, Alsen, and Eden, 1975):

Unseeded:

1.0	4.9	4.9	11.5	17.3	21.7	24.4	26.1	26.3	28.6	29.0	36.6	41.1
47.3	68.5	81.2	87.0	95.0	147.8	163.0	244.3	321.2	345.5	372.4	830.1	1202.6

Seeded:

4.1	7.7	17.5	31.4	32.7	40.6	92.4	115.3	118.3	119.0	129.6	198.6	200.7
242.5	255.0	274.7	274.7	302.8	334.1	430.0	489.1	703.4	978.0	1656.0	1697.8	2745.6

Open the Minitab worksheet `CloudSeeding.mtw`. Notice that the observational units (days) are in rows and the variables (whether the cloud was seeded, rainfall amount) are in columns. Select `Graph > Dotplot` and then “one Y, with groups.” Enter `c1` as the graph variable and `c2` as the categorical variable. Click OK.

- Which type of cloud (seeded or unseeded) tends to produce more rainfall?
- Which type of cloud has more variability in its rainfall amounts?
- What shape do both rainfall distributions follow?
- Do the distributions appear to have any outliers (unusually large or small observations)?

When comparing results on a quantitative variable between two or more groups, **boxplots** can be effective displays. Boxplots are based on the **five-number summary**, which “reduces” the distribution of data to these values:

- Minimum, smallest value in data set
- Lower quartile, also known as the 25th percentile, a value that 25% of the data falls below and 75% above (after the data have been arranged in order)
- Median, or 50th percentile, the middle value when the data have been arranged in order so that that half of the data falls below this value and half above
- Upper quartile, or 75th percentile, a value that 75% of the data falls below and 25% above (after the data have been arranged in order)
- Maximum, largest value in data set

Boxplots also indicate as outliers any value that is more than $1.5 \times \text{IQR}$ away from its nearer quartile, where IQR is the difference between the quartiles. You can also mark the mean (arithmetic average) on a boxplot.

Select Graph> Boxplot and then “with groups.” Enter c1 as the graph variable and c2 as the categorical variable. Click on “scale” and select “transpose value and category scales.” Click OK twice.

- Comment on what these boxplots reveal about rainfall amounts, comparing seeded with unseeded clouds.