**STAT4S: Exercise Using SPSS to Explore Graphs and Charts**

Author:   Ed Nelson
Department of Sociology M/S SS97
California State University, Fresno
Fresno, CA 93740
Email:  **ednelson@csufresno.edu**

**Note to the Instructor:** The data set used in this exercise is gss14\_subset\_for\_classes\_STATISTICS.sav which is a subset of the 2014 General Social Survey. Some of the variables in the GSS have been recoded to make them easier to use and some new variables have been created.  The data have been weighted according to the instructions from the National Opinion Research Center.  This exercise uses FREQUENCIES and EXPLORE in SPSS to explore different ways of creating graphs and charts.  A good reference on using SPSS is *SPSS for Windows Version 23.0 A Basic Tutorial* by Linda Fiddler, John Korey, Edward Nelson (Editor), and Elizabeth Nelson.  The online version of the book is on the [**Social Science Research and Instructional Council's Website**](http://ssric.org/node/582).  You have permission to use this exercise and to revise it to fit your needs.  Please send a copy of any revision to the author. Included with this exercise (as separate files) are more detailed notes to the instructors, the SPSS syntax necessary to carry out the exercise (SPSS syntax file), and the SPSS output for the exercise (SPSS output file). Please contact the author for additional information.

I’m attaching the following files.

* [**Data subset**](http://ssric.org/files/gss14_subset_for_classes_STATISTICS.sav) (.sav format)
* [**Extended notes for instructors**](http://ssric.org/files/Extended_Notes_for_Instructors_for_STAT4S.docx) (MS Word; docx format).
* [**Syntax file**](http://ssric.org/files/SPSS_Syntax_for_STAT4S.sps) (.sps format)
* [**Output file**](http://ssric.org/files/SPSS_Output_for_STAT4S.spv) (.spv format)
* [**This page**](http://ssric.org/files/STAT4S.docx) (MS Word; docx format).

**Goals of Exercise**

The goal of this exercise is to explore different ways of graphing frequency distributions. The exercise also gives you practice in using FREQUENCIES and EXPLORE in SPSS.

**Part I – Pie Charts**

A pie chart is a chart that shows the frequencies or percents of a variable with a small number of categories.  It is presented as a circle divided into a series of slices.  The area of each slice is proportional to the number of cases or the percent of cases in each category.  It is normally used with nominal or ordinal variables (see Exercise STAT1S) but can be used with interval or ratio variables which have a small number of categories.

Run FREQUENCIES in SPSS for the variables p1\_partyid, p4\_polviews, and d12\_childs.  (See Chapter 4, Frequencies in the online SPSS book mentioned on page 1.)  Click on “Charts” and select “Pie charts.”  Notice that there is an option called “Chart Values” that allows you to select whether you want your table to include “Percentages” or “Frequencies.”  Usually you want to select “Percentages.”

Once SPSS has displayed the pie chart in the output window, you can double click anywhere inside the pie chart to open the “Chart Editor.”  Once you have opened the “Chart Editor” right-click anywhere inside one of the pie slices in the “Chart Editor” and you will see a list of different ways you can edit your pie chart.  Click on “Show Data Labels” and then click on the “Data Value Labels” tab.  If “Percent” is not listed in the “Displayed” box, move it to that box and click on “Apply” and then “Close.”  If it is listed in the “Displayed” box, just click on close.  This will close the “Properties” box.  Click anywhere outside the “Chart Editor” and you will see your edited pie chart.  There are lots of other ways you could edit your chart.  Explore some of them if you are curious.

If you are wondering why you shouldn’t use pie charts for variables with a large number of categories, create a pie chart for d1\_age and you’ll see why.

**Part II – Bar Charts**

A bar chart is a chart that shows the frequencies or percents of a variable and is presented as a series of vertical bars that do not touch each other.  The height of each bar is proportional to the number of cases or the percent of cases in each category.  It is normally used with nominal or ordinal variables.

Run FREQUENCIES for the variables p1\_partyid and p4\_polviews.  This time click on “Charts” and select “Bar charts.”  Select “Percentages” to display percents in the chart.

**Part III – Histograms**

A histogram is a graph that shows the frequencies or percents of a variable with a larger number of categories.  It is presented as a series of vertical bars that touch each other.  The height of each bar is proportional to the number of cases or the percent of cases in each category.  It is used with interval or ratio variables.

Run FREQUENCIES for the variables d1\_age, d4\_educ, and d12\_childs**[[1]](http://ssric.org/node/496/edit%22%20%5Cl%20%22_ftn1%22%20%5Co%20%22)**.  Click on “Charts” and select “Histogram.”

Look at the histogram for d1\_age.  Let’s say you want to redefine the width of each vertical bar.  Double-click anywhere inside the histogram which will open the “Chart Editor.”  Now right click anywhere inside the rectangles in the “Chart Editor” and click on “Properties Window.”  This will open the “Properties” box. Click on the tab for “Binning.”  Click on “Custom” and “Interval width” under “X Axis.”  Enter 10 in the “Interval width” box indicating that you want each vertical bar to represent an interval width of ten years.

Where do we want the first interval to start?  We could let SPSS decide but let’s make the decision ourselves.  Click on “Custom value for anchor” and enter 10 in the box.  Click on “Apply” and look at your histogram.  Does it look how you want it to look?  Is there any further editing you want to do?  If you are satisfied, click on “Close” to close the “Properties” box.  Click anywhere outside the “Chart Editor” box and you will see your edited histogram.

**Part IV – Box Plots**

A box plot is a graph that displays visually a number of characteristics of a frequency distribution:

* the third quartile (Q3),
* the first quartile (Q1),
* the interquartile range (IQR),
* the median,
* the range,
* outliers, and
* extreme values.

Run EXPLORE for d1\_age, d4\_educ, and d12\_childs.  (See Chapter 4, Explore in the online SPSS book.)  You can use the default settings for EXPLORE so all you have to do is click “OK” after you have selected your variables.

The first thing you will see is various descriptive statistics for each variable.  You’re probably familiar with most of these.  Then you’ll see the stem-and-leaf display which we’re not going to discuss.  The last thing you’ll see is the box plot.  Let’s look at the boxplot for d1\_age.  The box is bounded at the top by the third quartile (Q3) and at the bottom by the first quartile (Q1).  The height of the box (Q3 – Q1) is the interquartile range.  The horizontal line inside the box represents the median.  There are two vertical lines coming out of the box.  This line extends upward to the maximum value and downward to the minimum value.  The difference between the maximum and minimum values is the range.

You can also learn about skewness from the box plot.  In a non-skewed distribution, the median will be in the middle of the box halfway between the third and first quartiles.  In a skewed distribution the median will be either higher or lower in the box.  Notice that for d1\_age and d4\_educ the median is in the middle of the box suggesting that these distributions are not very skewed but for d12\_childs the median is in the upper part of the box suggesting that this is a positively skewed distribution.

Now look at the box plots for d4\_educ and d12\_childs.  Here you’ll see some circles and numbers.  The circles represent outliers which are values that lie between 1.5 and 3.0 box lengths above the third quartile or below the first quartile.  A box length is just another name for the interquartile range since the height of the box is the interquartile range.  The numbers are the case numbers in SPSS.  Extreme values are values that are more than 3.0 box lengths from the first or third quartiles.  There aren’t any extreme values in these distributions.

Sometimes you want to compare box plots for two or more groups of respondents.  Let’s look at the box plot for d1\_age and compare the box plots for men and women.  Run EXPLORE for d1\_age but this time put d5\_sex in the “Factor List” box.  Your output should now show the box plots for men and women side-by-side.

**Part V – Conclusions**

We have talked about four different types of graphs – pie charts, bar charts, histograms, and box plots.  There are other types of graphs you could use but these are the four most commonly used graphs.  There are other ways to construct graphs in SPSS that your instructor might want to talk about.  You can click on “Graphs” in the menu bar at the top of the SPSS screen and then on “Chart Builder” but we aren’t going to go into that in this exercise.

**[[1]](http://ssric.org/node/496/edit%22%20%5Cl%20%22_ftnref1%22%20%5Co%20%22)** There is a small problem with d12\_childs.  One of the categories is “eight or more” children.  That means we don’t know what these values actually are.  They could be 8 or 10 or 12 or 14 or something else.  Since there are so few cases in this category we’re going to ignore this problem.