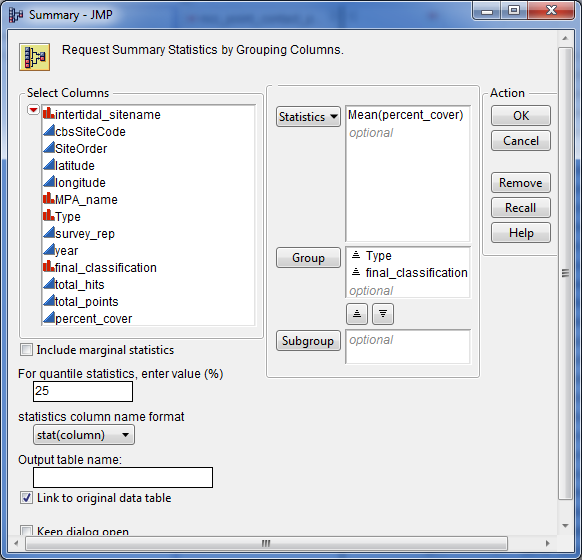
JMP Pro Introduction worksheet. Preliminary Step: make a new folder where you can keep work. Try to save your work to this folder. If you are using a class laptop make the folder in “My Documents” and call it **“your name BIO 286 WORK”**

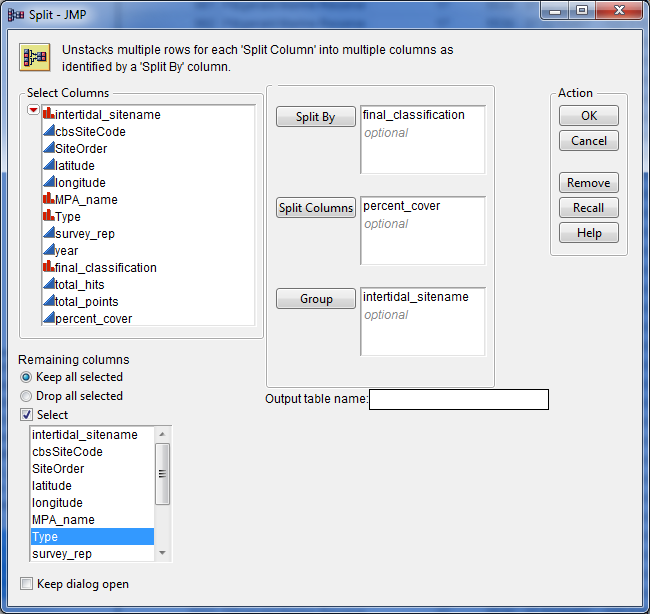
ALL CAP = COMMAND

‘bbb’ = file or variable name’

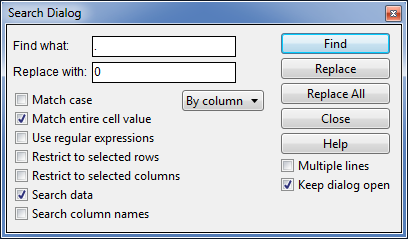
1. Opening and saving data
   1. Open the data file ‘ncc\_biodiversity\_summary\_data.xlsx’. Click on the down arrow next to OPEN to select the worksheet that you desire to work on ‘ncc\_point\_contact\_per\_cover’. These are data from the assessment of the north central California coast Marine Protected areas. The key variables for the worksheet are MPA name (marine protected area name), Type (reserve or reference area), final-classification (species name), percent\_cover (percent cover of each species at each site.
   2. Save the datafile back to your directory as a JMP file. FILE, SAVE
2. Manipulating data – TABLES
   1. Summarize data. Lets assume that we want to calculate the average percent cover for all species based on the type of area (reserve vs reference). Click on the TABLES tab on the main menu, then click on SUMMARY. Now click on ‘percent\_cover’ to highlight it, then on STATISTICS (next to the top window), then on MEAN. ‘Mean(percent\_cover)’ will appear in the window. (Notice the other summary options).
      1. Now we want to calculate these values based on species and type of area sampled to add ‘Type’ and ‘final classification’ to the GROUP window.



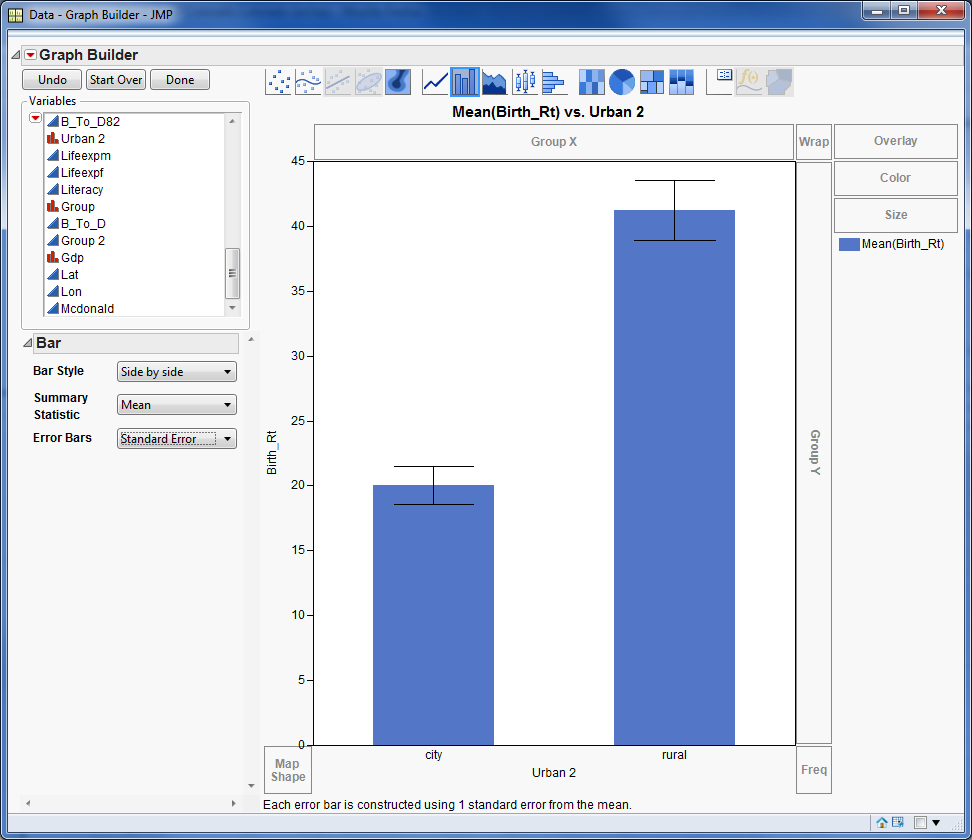
* + 1. Click OK and you will get a new table that shows the mean percent cover as a function of species and area type. You can save this as a JMP file (FILE, SAVE). The default name for this is ‘ncc\_point\_contact\_per\_cover By (Type, final\_classification)’. You could save it to an alternative file type (eg. XLS, CSV) by using FILE, SAVE AS
  1. Re-structure data. Once of the great attributes of JMP is its tools to restructure data. For example assume that we want to restructure the file ‘ncc\_point\_contact\_per\_cover’ so that each species is a variable (that is each species in a separate column). Such a structure is extremely useful for graphing and some types of analyses (especially multivariate stats).
     1. Go back to ‘ncc\_point\_contact\_per\_cover’. Click on TABLES, SPLIT.
     2. Put ‘final\_classification’ in the SPLIT BY window. This indicates that you want to split this single variable into separate species.
     3. Put ‘percent\_cover’ in the SPLIT COLUMNS window. This indicates that the data to be associated with the species is the percent cover.
     4. Put ‘intertidal\_sitename’ in the GROUP window. This will produce a table where each line represents a site. Click SELECT, KEEP ALL SELECTED on the REMAINING COLUMNS option and highlight ‘Type’. This will allow this variable to be included in the re-structured datafile (in Windows use CTRL button if you want to highlight multiple variables; command on a Mac.)



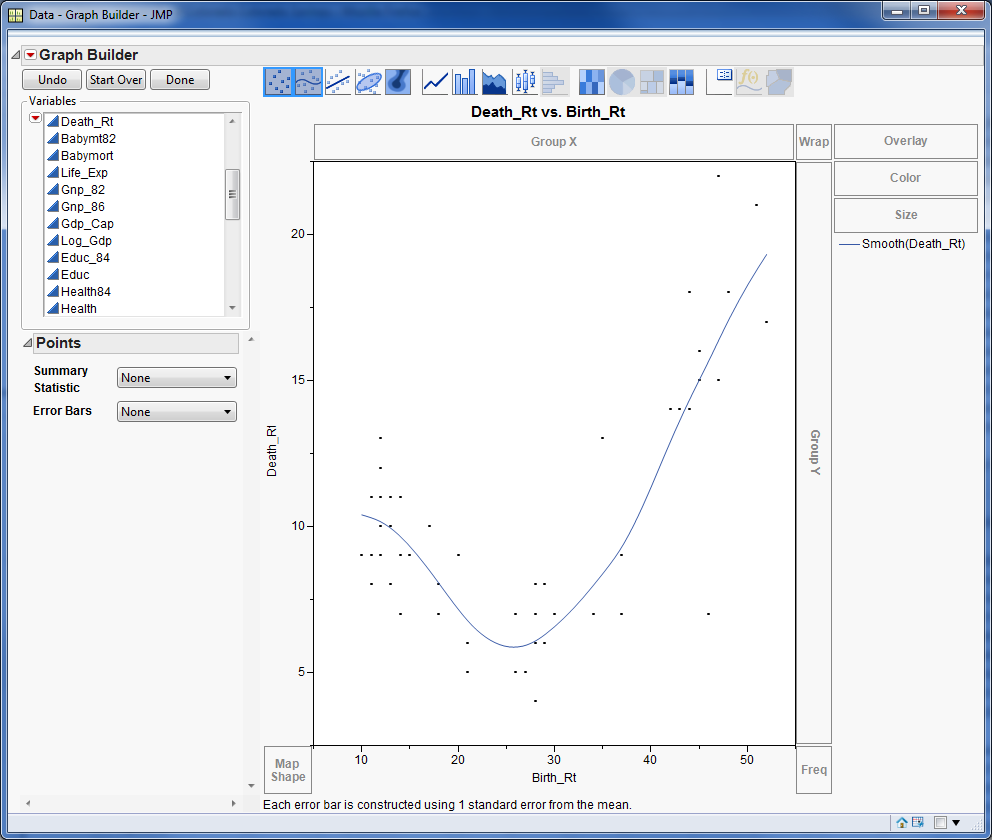
* + 1. Click OK and you will get a new table that shows the percent cover for each species, where each species is in its own column. You can save this as a JMP file (FILE, SAVE). Here you will need to give the file a name. I usually call a split file something like ‘file\_split’ where ‘file’ is a description of the dataset
    2. You will notice that there are many . in the new datafile. These are missing values and are the result of the original file being unpopulated. This term refers to files that only include non-zero observations, which is very common for large datasets. In order to use the data for analyses and graphs it is often important to populate the file. Note an unpopulated file is explicitly one where true zeros have been left out to constrain size. True missing values (e.g. you lost some data or forgot to sample a species) should never be replaced in a datafile. We need to replace all the missing values with zeros. To do this click EDIT, SELECT ALL on the split datafile. Then click EDIT, SEARCH, FIND. In the FIND window put ‘.’ In the FIND WHAT window and ‘0’ in the REPLACE WITH window and click REPLACE ALL. This will replace all missing values with zeros. Resave the file.



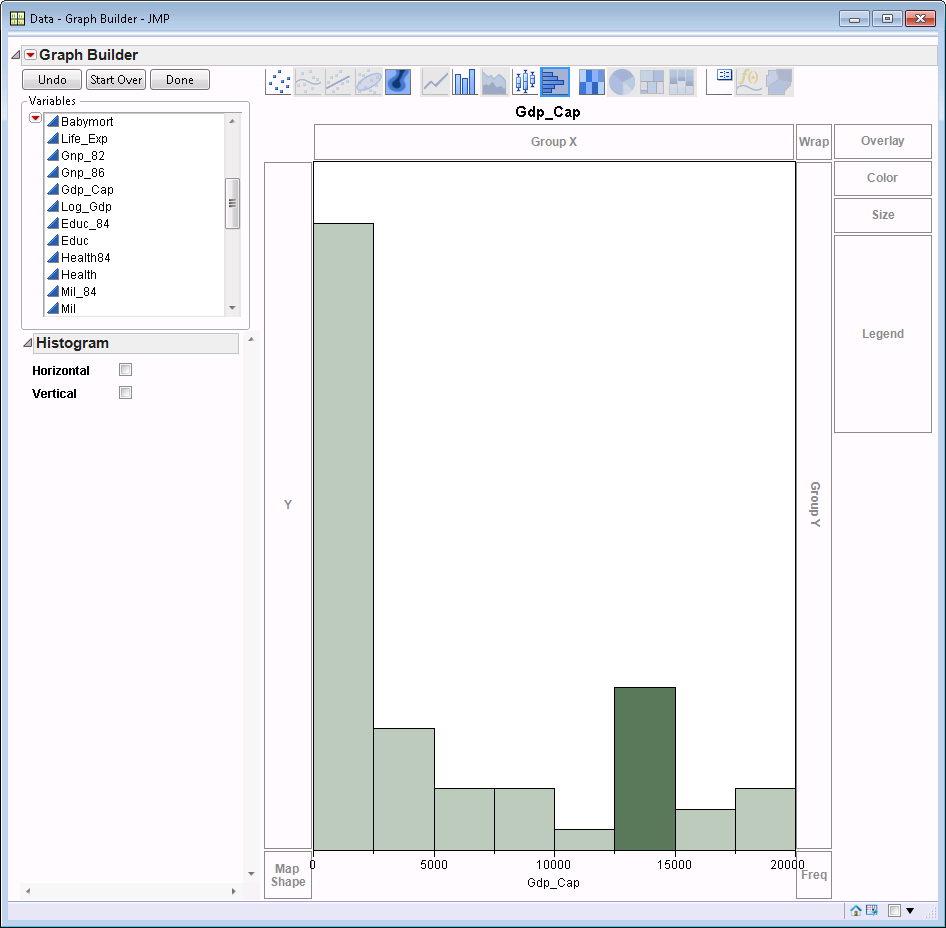
1. Elementary graphing - here you will get familiar with some of JMP graphing modes
   1. Data snooping. Open the file ‘ourworld.xlsx’. This is a file that describes demographic and other attribute of countries of the world. Click GRAPH, GRAPH BUILDER. This is the generalized graphing utility in JMP. It is very useful for snooping but there are better alternatives in JMP for detailed or complex graphs.
      1. Assume you want to look at the relationship between Birth rate and whether a country is Urban or Rural. This would be a bar graph and to make this graph drag the variable ‘Urban 2’ to the X AXIS and ‘Birth\_Rt’ to the Y AXIS. Note the red bar icon means categorical variable and the blue triangle icon means continuous variable.
         1. The default graph type is scatterplot, change this to a bar graph by clicking on the BAR icon on the tip of the frame (the HISTOGRAM icon looks like a sideways bar graph)
         2. Add error bars by clicking ERROR BAR and choosing STANDARD ERROR.



* + - 1. Play with other options to see the abilities of GRAPH BUILDER. For example try putting ‘Group’ in the overlay box.
    1. Assume you want to look at the relationship between ‘Birth\_Rt’ and ‘Death\_Rt’. Put ‘Birth\_Rt’ in the X AXIS and ‘Death\_Rt’ in the Y AXIS. Now click the SCATTERPLOT icon with the curved line. This represents a local smoother (LOESS Smoother) that represents the general fit to the data



* + 1. Now assume that you want to look at the distribution of the per capita GDP (gross domestic product) ‘Gdp\_Cap’’ and to see if it is normally distributed.
       1. In the GRAPH BUILDER window put ‘Gdp\_Cap’ in the X AXIS, then click on the HISTOGRAM icon (which looks like horizontal bars). This is the histogram of the data.
          1. Click on one of the bars and it will become highlighted. Then look to the dataset – those observations that are in the bar will be highlighted. This is a great way to examine data and also is useful for data selection.



* + - 1. Now you can do something very cool - In the upper left corner of GRAPH BUILDER (and this is true for all JMP windows) is a red triangle. This indicates options that can be invoked. Click on the red triangle and then on LAUNCH ANALYSIS. This will launch what JMP thinks is the most likely analysis that would be done, given the graph. Here the DISTRIBUTION platform is launched.
      2. You should see a Histogram of the data (vertical orientation) and next to it a box plot. Click on the GDP\_CAP red triangle and then on HISTOGRAM OPTIONS then on PROB AXIS. Follow this procedure again to show the COUNT AXIS. These show the probability of occurrence for each bar value (eg 50% chance and 30 cases of Gdp\_Cap being between 0-2500). Below the histogram is a table of quantiles and then summary statistics
      3. To find out if this is a normal distribution click on the red triangle for GDP\_CAP and then on CONTINUOUS FIT then on NORMAL. Now click on the red triangle for FITTED NORMAL then DIAGNOSTIC FIT. This produces a probability plot, which shows the data vs the data if they were normally distributed. The points should be on the straight red line, which is the expected fit if the data are normally distributed. DO the data look normally distributed? Often you use this device to visually assess normality.
      4. If you are interested in testing for normality click on the red triangle for FITTED NORMAL and then on GOODNESS OF FIT. You will now get a test for normality. There are many tests and they will often differ depending on distribution to fit and sample size. Don’t worry, JMP uses the most robust test for each case.
      5. Now lets see if the fit would be improved by a transformation. Based on the histogram and the probability graph, a log transformation may be useful.
      6. Click on the red triangle for GDP\_CAP and then on CONTINUOUS FIT then on LOGNORMAL. Now click on the red triangle for FITTED LOGNORMAL then DIAGNOSTIC FIT. Dies this look better than the graph This produces a probability plot, which shows the data vs the data if they were Log normally distributed. The points should be on the straight red line, which is the expected fit if the data are Log normally distributed. DO the data look Log normally distributed?
      7. Now lets test the fit. Click on the red triangle for FITTED LOGNORMAL and then on GOODNESS OF FIT. You will now get a test for goodness of fit to Log Normality. Compare this p-value to that for the normal distribution.
      8. What is your conclusion - should the data be transformed?

1. Additional graphing – do these if you have time. In all cases explore graph options in GRAPH BUILDER module. Play with this module and its options . Make sure to ask question if you get stuck or simply want to find out if something can be done that you can’t figure out how to do. Open “data for analytical graphing handout.xlsx”. There will be 5 sheets in that workbook and in JMP you will be asked if you want to import all of them. You should do this. This will open all of them in separate sheets in in JMP. Use the appropriate sheet for each graph. Think about graph type (eg bar, line, scatterplot, density) for each dataset. The data sets are:
   1. *Blue whale abundance*. Here we want to see how Period and Location are associated with blue whale numbers. You may need to restructure the dataset to get the graph you want.
   2. *Abalone landings*. This sheet shows the commercial landings in pounds for all species of abalone in CA up to the point of fishery closure. There are many options about whatyou may want to plot (eg by species, in total etc).
   3. *Size and age of seastars*. Here the relationship is shown for the relationship between size and age of seastars. Importantly this dataset is the modeled relationship.
   4. *VO2 vs runtime*. These data are from individuals where two measurements were made: VO2 max (an estimate of the fitness or ability to use O2) and the time it took that person to run a mile. Here we are looking at raw data and trying to understand the relationship between two measured variables.
   5. *Limpet size*. Here we are interested in the size frequency distribution of limpets