Stat 301 – Lab 1

This lab has two parts:

1) an illustration of sampling (Sampling rocks)

2) an introduction to JMP

**Part 1: Sampling rocks:**

Question: Geologists would like to know the total weight of a collection of rocks. This can be calculated from the number of rocks and the mean weight (gm/rock) in the population. For various reasons, it is impossible to weigh all rocks together and impractical to weigh each rock individually. Hence, we want to estimate the mean weight. The data will come from a sample of rocks.

**Materials:**

1) There are two different populations of rocks. One has rocks numbered 1 to 50; the second has rocks numbered 51 to 100. Please choose to work with one of the two populations (please divide somewhat evenly).

2) Two ten-sided dice that should be shared with others.

3) A data sheet to record information on.

**Activity:**

Please work individually.

You will each use four different ways to sample the rocks. For each method, record the rock number (or numbers) on your data sheet.

1) Look at the rocks. Determine which single one is the most representative of the population (remember we are interested in the weight). That is the “representative rock” on the data sheet.

2) Now determine the five rocks that you believe collectively are representative of the population. Those numbers are the “5 representative rocks” on the data sheet.

3) Use the dice to randomly select a rock from your population. Roll the dice. 00 is rock # 100. If the number you rolled is not in the population you’re working with (e.g. your population is rocks 1-50 and you rolled 62), roll again until you get a number appropriate for your population. That number is the “random die rock” number on the data sheet. Repeat five times for a total of five “random die rocks”. It ***is not a problem*** if one or more rocks are duplicates (e.g. you roll 10, 25, 32, 49, 25, so rock 25 occurs twice in your random sample).

4) Please start working on the next part (JMP). After all are done with rocks activities 1-3, we will put all the rocks into a box. Sometime before you leave lab, come up and “randomly” grab a rock from the box. Record its number as one of the “random grab rock” numbers on the data sheet. Replace that rock in the box and “randomly” grab another. Record its number. Again, it’s not a problem if that rock is the same as one you’ve drawn before. Repeat for a total of five “randomly grabbed” rocks.

Turn in your data sheet before you leave lab. We have the weights of each rock and will calculate the mean weight for each of your four samples.

We will discuss the results in Friday’s lecture.

**Part 2: an introduction to JMP.**

Work at your own pace. You may leave whenever you feel comfortable starting JMP, reading data, and drawing basic plots, and computing summary statistics. We will circulate to answer questions.

We will use the pondice2 data set. This is the pondice data set described in problem 1.16 with an extra variable. To quote from the textbook:

“The National Snow and Ice Data Center (NSIDC) collects data on the albedo, depth, and physical characteristics of ice melt ponds in the Canadian Arctic. Environmental engineers at the University of Colorado are using these data to study how climate impacts the sea ice. Data for 504 ice melt ponds location in the Barrow Strait in the Canadian Arctic are saved in the PONDICE file. One variable of interest is the type of ice observed for each pond. Ice type is classified as first-year ice, multiyear ice, or landfast ice.”

The extra variable is Season, which is Early if the pond was evaluated early in the season (Julian Day < 157) and Late if the pond was evaluated late in the season (Julian Day > 157).

**Guiding principles:**

To do something in JMP (draw a graph, calculate statistics, fit models), you:

1) load the data set into JMP (File menu items)

2) select what you want JMP to do (usually from the Analyze or Graph menus)

3) select the variables JMP is to use and specify how it is to use them

(from a Select Columns box followed by a role for each variable)

4) tell JMP to do the calculations / draw the plot that you want

5) modify it by providing options (menus under the red triangles)

Here are the details to draw histograms.

Goal:

1) draw a histogram of the number of ponds with each type of ice.

2) draw stacked histograms of the number of each type of ice, classified by season (Early or Late).

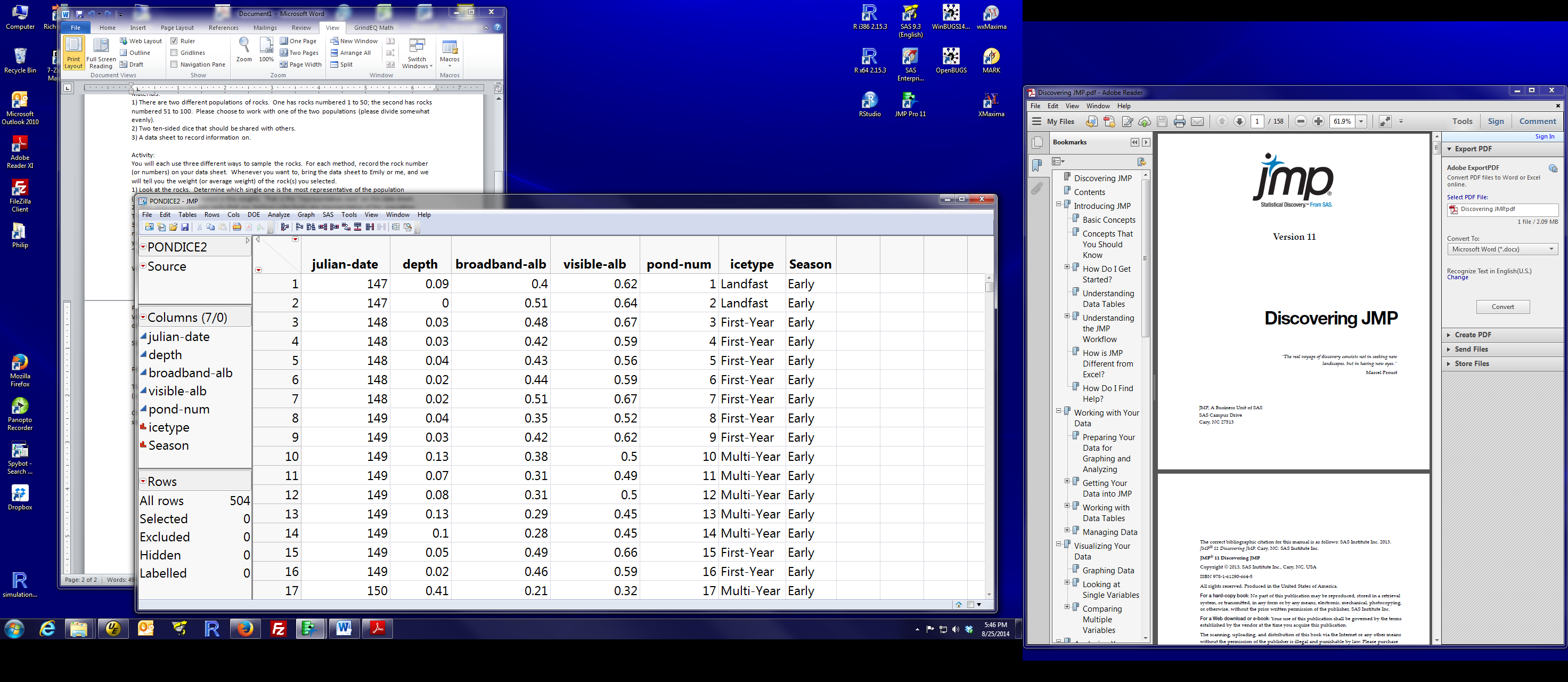
**Starting JMP:**

double click on the JMP Pro 12 icon or select JMP Pro 12 / JMP Pro 12 in the start button menu.

**Reading a text data file:**

Data files for class are on the data sets page of the class web site. Download the pondice2.txt file. This is the pondice data set from the book with an extra variable. Save the file in a convenient location (e.g. the desktop, your myfiles space, or a flash drive).

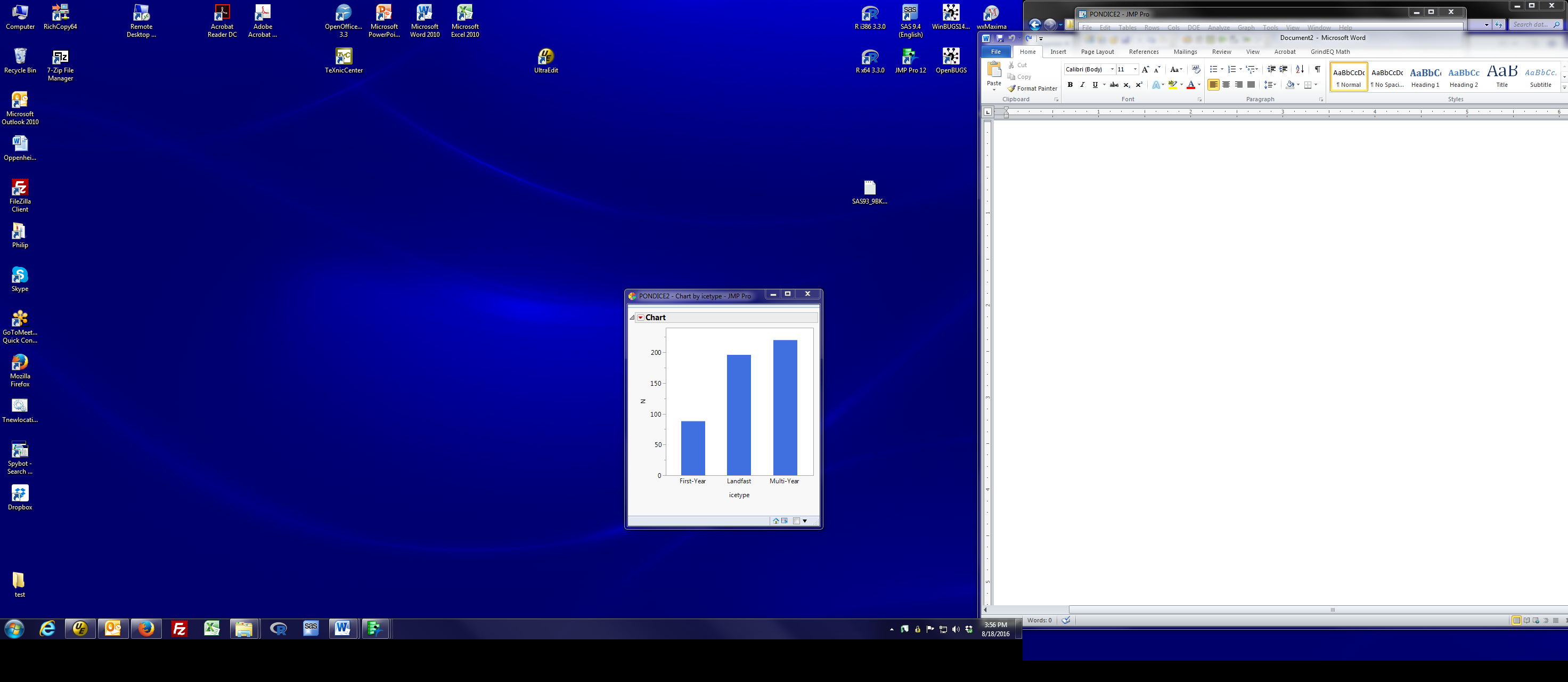
In JMP, select File/Open. In that dialog box, change the folder (top box) to your convenient location. Change the file type (bottom right) to text files. Select pondice2.txt. Make sure the “Open as:” button is set to “Data, using Text Import preferences”, then click Open. A new window looking like:



should open up. This provides access to various data manipulation tools. We will discuss some of these later in the semester.

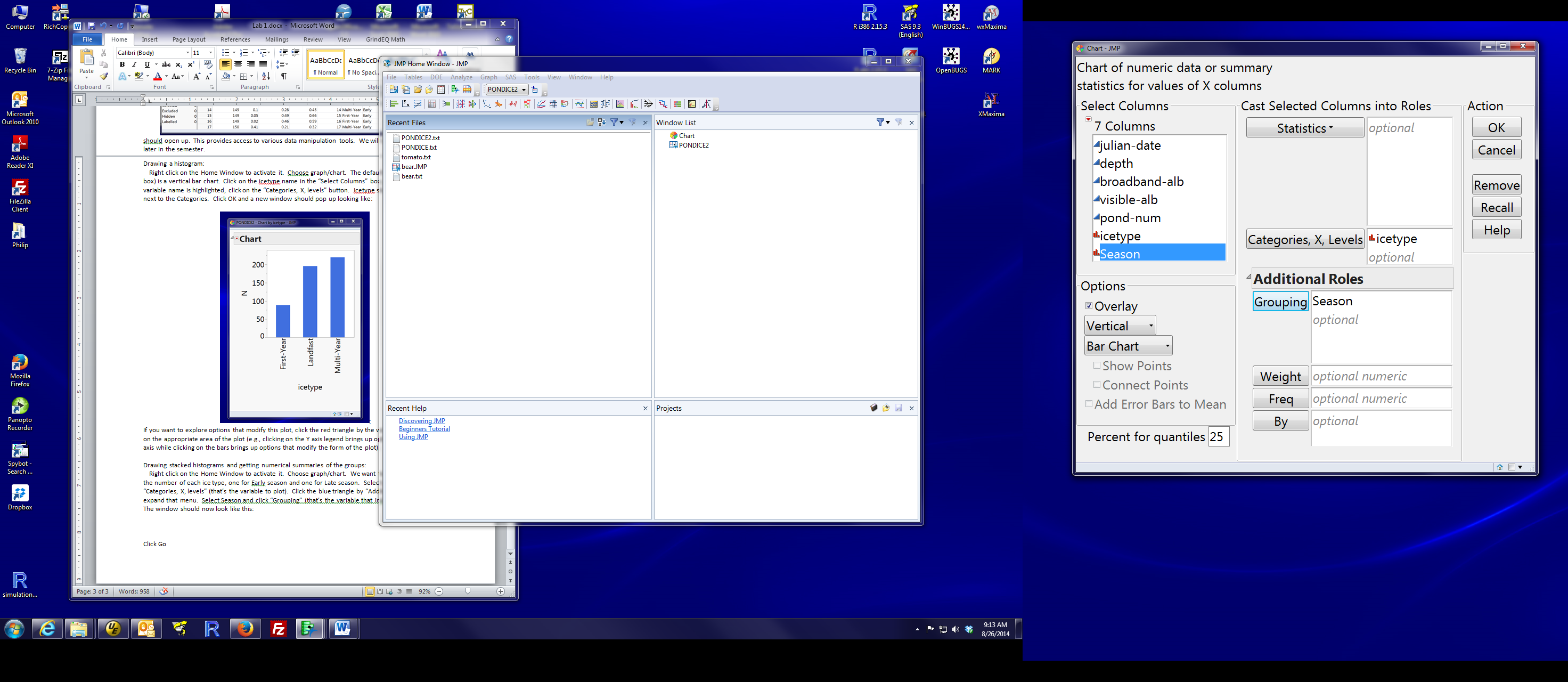
**Drawing a histogram:**

Right click on the Home Window to activate it. Choose graph/chart. The default option (lower left box) is a vertical bar chart. Click on the icetype name in the “Select Columns” box, then while that variable name is highlighted, click on the “Categories, X, levels” button. Icetype should appear in box next to the Categories. Click OK and a new window should pop up looking like:



If you want to explore options that modify this plot, click the red triangle by the word Chart or right-click on the appropriate area of the plot (e.g., clicking on the Y axis legend brings up options that modify that axis while clicking on the bars brings up options that modify the form of the plot).

Drawing stacked histograms:

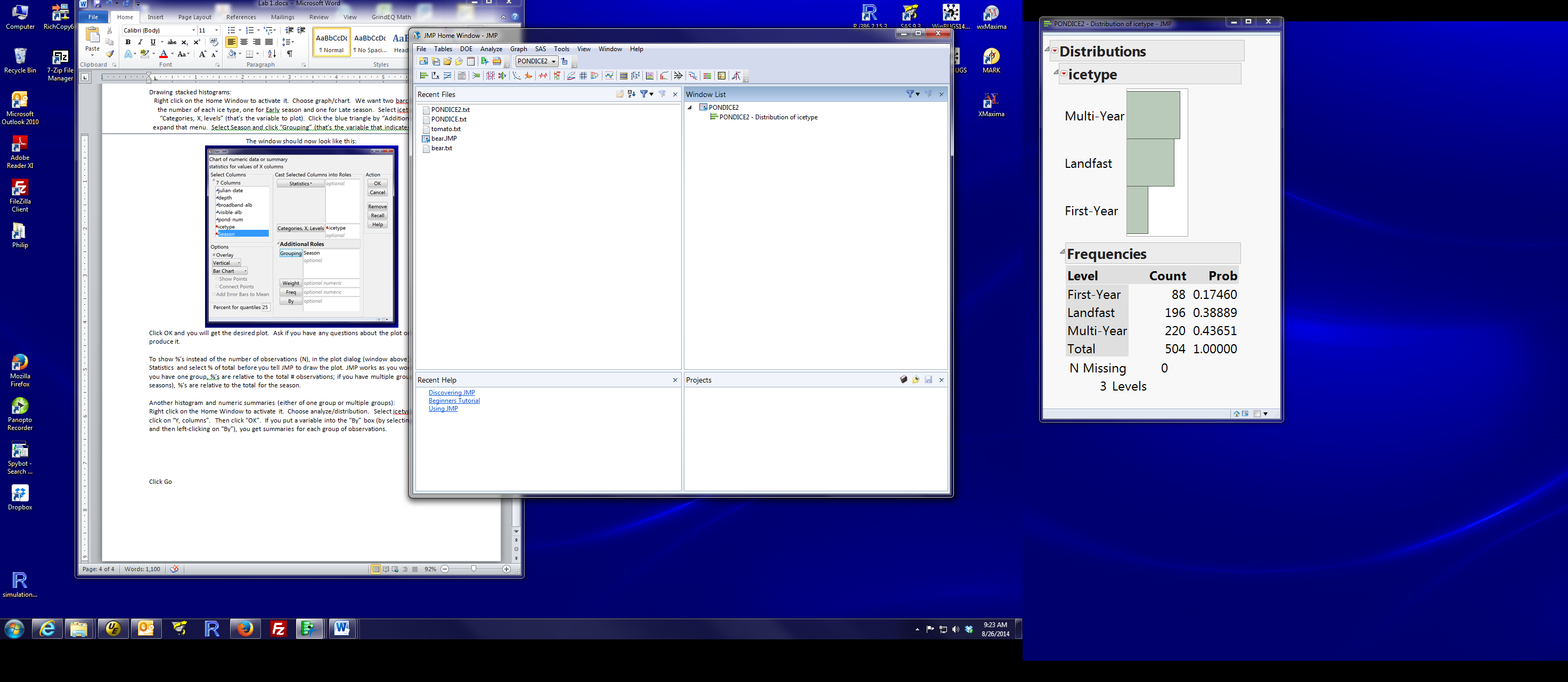
Right click on the Home Window to activate it. Choose graph/chart. We want two barcharts showing the number of each ice type, one for Early season and one for Late season. Select icetype and click “Categories, X, levels” (that’s the variable to plot). Click the blue triangle by “Additional Roles” to expand that menu. Select Season and click “Grouping” (that’s the variable that indicates data groups). The window should now look like this: 

Click OK and you will get the desired plot. Ask if you have any questions about the plot or how to produce it.

To show %’s instead of the number of observations (N), in the plot dialog (window above), left-click on Statistics and select % of total before you tell JMP to draw the plot. JMP works as you would expect: if you have one group, %’s are relative to the total # observations; if you have multiple groups (e.g. seasons), %’s are relative to the total for the season.

**Another histogram and numeric summaries (either of one group or multiple groups):**

Right click on the Home Window to activate it. Choose analyze/distribution. Select icetype then left-click on “Y, columns”. Then click “OK”. The results window will look like:



The top part is a sketch histogram of the counts. The bottom table is the # of each type (“Count”) and their proportion (“Prob”). You also get a count of the number of observations that are missing the icetype value.

If you put a variable into the “By” box (by selecting the variable and then left-clicking on “By”), you get summaries for each group of observations.

**Self-assessment:**

The EARTHQUAKE.csv data set is the date and magnitude of the 2,929 aftershocks following the 17 January 1994 Northridge (Los Angeles CA) earthquake. The variable SIZE categorizes the magnitude of the aftershock as very weak (< 2.5 on the Richter scale), weak (2.5 – 3.5) or strong (> 3.5).

1) Draw a histogram of the size of the aftershocks. Were most aftershocks, Strong, Weak, or Very weak? Briefly explain your answer.

2) How many strong aftershocks were there?

3) What percent of aftershocks were strong?

The variable Month is 1 for January and 2 for February.

4) How many strong aftershocks were there in January? In February?

5) What percent of January aftershocks were strong?

Answers:

1) Most were very weak – the height of the bar for very weak greatly exceeds that of the other two sizes.

2) 109

3) 3.7%

4) January: 105. February: 4

5) 4.4%