LAB 2 - Descriptive Statistics

Analyzing a Scatter plot
Let’s continue using the Polycity.JMP dataset from last time.

1. Which variable(s) are nominal (qualitative)?
2. Which variable(s) are ordinal (qualitative, but ordered)?
3. Which variable(s) are quantitative (specify discrete or continuous)?

Plot ozone (y) by population (x). Analyze → Fit Y by X. Next, try changing the title of your plot by double clicking on the grey bar at the top of the plot. Change the title to ”Ozone by Population”, then click enter. If you wish to also change the names of either axis you can click on the variable name for that axis to change it. Try this by changing the x-axis from ”POP” to ”Population”.

4. Describe your plot. Do you see a pattern in the data? (Does it form a straight line? Are there any clusters of observations? Any points that ”stray” from the general pattern?)

Move the pointer over various observations: the city name should appear on the screen to correspond to that particular observation. The reason we’re able to see the city name corresponding to a particular observation is because ”city” has been labeled by default.


6. What city has a population of approximately 3500 units and an ozone level of approximately 0.18 units?

Constructing a Histogram
Now let’s use the dataset Students.JMP, which can be found under JMP IN Data.

7. Let’s get familiar with this new dataset: repeat questions 1-3 from above. In our last dataset, ”city” was our unit of observation-what is our unit of observation?

We want to use ”sex” as a labeling variable in plotting height (y) by weight (x). To do this, right click on the ”sex” variable at the leftmost side of your screen, and select ”label”. Notice the tag next to the ”sex” variable. Now let’s plot height (y) by weight (x).

8. Is the observation at approximately height = 53 units and weight = 195 units male or female?

9. Of the four tallest students, how many are male?

10. Of the three shortest students, how many are female?

11. What percent of students are age 13?

In order to this, use Analyze → Distribution of Y and choose ”age”. This gives us a histogram of ”age”. If you wish to see this histogram on a horizontal scale, click on the red arrow next to ”age”, then go to Display Options → Horizontal Layout.
Variable Summaries, Box plots and Stem and Leaf plots

Moving right along to the Cereal.JMP dataset found in JMP Extra Data, let’s go through the following questions:

12. What is the average, standard deviation, minimum, and maximum of "calories"? Analyze → Distribution of Y.

13. What is the median and quartiles of "sugars"?

14. Using the histogram, box plot, and stem and leaf plot, describe the distribution of "complex carbo".

Click on the red arrow next to "complex carbo" and select Stem and Leaf. Describe the distribution in terms of location, spread, shape, any gaps, or potential outlying observations. Next, click on the hand icon on the tool bar at the very top of your screen. Use this to alter the bin widths of your histogram. Do this by clicking on any given bar of the histogram and dragging the hand up or down to change the widths. By pulling the bars up to the top, you will have the maximum number of bars. Now, click back on the arrow on the top toolbar. Use the arrow to select the leftmost bin of your histogram. This will highlight the observation in the stem and leaf plot as well as the dataset itself.

15. What cereal does this outlier correspond to?