CHAPTER 2

Graphs

JMP has many tools available to construct a variety of statistical graphs. Depending on your definition of the type of data there are some graphs that may be constructed. As you studied during lecture and in your textbook, several types of graphs are suitable for certain types of data as shown in Table 2.1. The choice of the type of graph depends on which characteristics of the data you want to highlight.

Table 2.1

<table>
<thead>
<tr>
<th>Type of data</th>
<th>Type of Graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate: Numeric – Continuous</td>
<td>• Histogram</td>
</tr>
<tr>
<td></td>
<td>• Frequency polygon</td>
</tr>
<tr>
<td></td>
<td>• Dot plot</td>
</tr>
<tr>
<td></td>
<td>• Stem and leaf plot</td>
</tr>
<tr>
<td></td>
<td>• Box Plot</td>
</tr>
<tr>
<td>Univariate: Character – Nominal (counts for each category)</td>
<td>• Bar graph</td>
</tr>
<tr>
<td></td>
<td>• Pareto chart</td>
</tr>
<tr>
<td></td>
<td>• Pie chart</td>
</tr>
<tr>
<td>Bivariate: Var 1) Numeric – Continuous vs. Var 2) Numeric – Continuous</td>
<td>• Scatter plot</td>
</tr>
<tr>
<td>Bivariate: Var 1) Numeric – Continuous vs. Var 2) Categorical</td>
<td>• Parallel boxplots</td>
</tr>
<tr>
<td>Multivariate: Several variables: Numeric – Continuous</td>
<td>• Scatterplot matrix</td>
</tr>
<tr>
<td>Multivariate: Var 1) Categorical- counts per category Var 2) Categorical- counts per category</td>
<td>• Mosaic Plot</td>
</tr>
<tr>
<td>Numeric – Ordinal</td>
<td>• Dot plot</td>
</tr>
<tr>
<td></td>
<td>• Stem and leaf plot</td>
</tr>
</tbody>
</table>

Class Exercise: Practice creating some statistical graphs from the file MOVIES:

First, open the JMP file MOVIES following the procedure described on Chapter 1. Then look at the “columns info” by right-clicking at the heading of each column. Verify that the type of data is...
compatible with the kind of graph that you are planning to do. If that is not the case, change the type of data to match the type of graph that you intend to create according to Table 2.1.

We are going to create a histogram using a numeric-continuous variable, for example let’s use the variable “Budget” to construct a histogram.

Use the “Analyze” menu and then choose “Distributions” as follows:

Figure 2.1

Then select the variable “Budget” and click over “Y Columns” as described below:

Figure 2.2
click “OK” and you can see a new window with a histogram as follows:

Figure 2.3

![Vertical Histogram](image)

This is a vertical histogram, but you can change the orientation of the histogram by selecting the red triangle on the left side of “Budget”. Then select “Display options” and click over “Horizontal Layout” as shown below.

Figure 2.4

![Display Options](image)
you get a histogram in a horizontal layout and a boxplot above the histogram. From this new window you may select a bar (left click over it) and the data that corresponds to that bar is highlighted. This is a practical way that allows you to identify the individual values that corresponds to each bar.

Figure 2.5

We can also generate a stem and leaf plot. In the same window, let’s click on the red triangle on the left side of “Budget” and select the “Stem and Leaf” Option in that menu as shown in the figure below:
Then the stem and leaf plot is displayed.

Now, let’s try a Pareto chart. You need to open the menu “Graph” and then select “Pareto Plot” as shown below:
In this case we need to select a categorical variable, for that purpose we select the variable "MPAA Rating" and click over "Y, Cause" as shown below:
Then click “OK” and you get a Pareto chart. Please remember from lecture that this chart contains bars ordered in decreasing order from the highest to the lowest frequency. In addition to the definition of a Pareto chart seen during lecture, this graph also contains a cumulative frequency line over the Pareto chart, and it shows two vertical scales, on the left side we can see a frequency scale, on the right side there is a percentage scale. Not surprisingly, the last point in the cumulative frequency line reaches the total number of observations on the frequency scale (left side) or equivalently, it reaches 100 percent on the percentage scale (right side).

Figure 2.9

We can also construct a pie chart by selecting the “Chart” option from the “Graph” menu shown below:
Then, we select “Pie Chart” from the drop down menu labeled “Vertical” as shown below
next, select the variable “MPAA Rating” and click over “Categories, X, Levels”

Figure 2.12
as a result the pie chart is shown in the next figure. There are many options to add labels to the pie chart, but I am going to let you explore this options. Please click over the red triangle and play these options.

Figure 2.13
Graphs for bivariate data.

Next, we are going to generate graphs for bivariate data (paired data), that is for pairs of observations, such as \((x,y)\) for example. Usually, in this type of graph we intend to explore the strength of the association between these two variables. We are going to explore the relationship between the variables “Budget” and “Gross” in the data set MOVIES, by using an appropriate type of graph. A scatter plot is a powerful tool to explore this kind of relationship. Let’s graph the pair of variables using the proposed scatter plot. First you have to select the option “Fit Y by X” from the “Analyze” menu as shown below.

Figure 2.14

Let’s select “Budget” and click over “X, factor” this is the independent variable. Then select “Gross” and click over “Y, Response” as shown in Figure 2.15. “Gross” is the dependent variable,

Figure 2.15
let’s take a look at the box located on the lower left corner of the previous window. This window summarizes the types of graphs and analyses that are available according to definition of the type of data. The type of analyses available will be discussed in detail in posterior chapters. A scatter plot, located on the upper left corner, can be generated by using two continuous variables as in this case.

Figure 2.16

![Bivariate Analysis Options](image)

after selecting these options, you click over “OK”. The resulting scatterplot can be seen on the next window,

Figure 2.17

![Scatterplot of Budget vs. Gross](image)
We can also generate parallel boxplots to compare continuous measurements that correspond to different categories. This is a very typical situation that is commonly found in research. In order to generate this graph we select the option “Fit Y by X” from the “Analyze” menu as shown below.

Figure 2.18

We want to compare the “Gross” variable for movies, according to their type of rating “PG”, “PG-13” and “R”, then, let’s select “MPAA Rating” as the “X, Factor” (the independent variable) and “Gross” as “Y, Response” (the dependent variable), as follows:

Figure 2.19
click over “OK” and you get the graph shown below. This graph shows the numeric observations classified in different groups according to the MPAA Rating. It may be an interesting research question to determine if there is a difference on gross sales across the different groups defined by their MPAA rating,

Figure 2.20

Furthermore, we can also generate a special type of graph that is called a Box Plot (see chapter 3 of your textbook) by clicking over the red triangle located on the left side of “One way Analysis of Gross by MPAA Rating”. A new menu is displayed, then select “Display Options”, another menu appears, then select “Box Plots”. 
this graph shows parallel boxplots; each boxplot corresponds to a different level of the classification variable, in this case the “MPAA Rating”

Figure 2.22
The width of the box is proportional to the number of observations included in that level of the classification variable.

Next, we are going to explore multiple scatter plots arranged side by side in a matrix form that is called a scatter plot matrix. This kind of graph is useful to explore data sets found in a multivariate environment.

First, select “Scatter Plot Matrix” from the “Graph” menu as shown below:

Figure 2.23

Then, select the set of variables that you want to include in the scatter plot matrix. It could be any number of variables, then click over “X”, and you will see the window below

Figure 24.
repeat the same procedure for the variables of interest, in this case, select the same variables and click over “Y Columns” as shown below, then click over “OK”

Figure 2.24

now, you can see the generated scatter plot matrix. The result is an arrangement of scatter plots that shows a graph for every combination of variables. Notice that the lower left triangle (below the diagonal) is a mirror image of the upper right triangle (above the diagonal).

Figure 5.24
Class exercise: Practice by doing all possible graphs from the file BEARS, select every variable and obtain a suitable type of graph for each variable. You can also graph bivariate data by selecting pairs of variables that you think might be related and follow the procedures described above. You can also obtain also a scatter plot matrix for all continuous variables.

Team Assignment:

At this point you should have a good understanding of several types of graphs that can be generated using JMP.

1- Open your random sample that you obtained during the previous lab session from the Excel file “Smalltown_TX.xls”. If you have not done this activity yet please refer to the lab manual for Chapter 1, obtain your random sample and save it as described there.

2- Determine the type of graph that would be appropriate for each variable in the dataset.

3- Generate univariate graphs for every numeric-continuous variable in the dataset.

4- Get Pie Charts or Pareto charts for all categorical variables in the data set.

5- Generate parallel boxplots for men and women for the variables salary and height. Hint: Use using the menu “Analyze” and “Fit Y by X” then use the option “By” and select the variable SEX in the menus to separate men and women. Make sure that the variable SEX is defined as a categorical variable.

6- Generate scatter plots for pairs of continuous variables that may be associated using the menu “Analyze” and “Fit Y by X” (You can also use a scatter plot matrix)
   a. Weight vs. Height
   b. Income vs. Gifts
   c. Age vs. Sys_BP
   d. And some others that you suspect might be related.

7- Write a report for your lab instructor summarizing all your findings and graphs. Make comments about every interesting finding in your graphs.

Chapter Summary

You already practiced constructing the following graphs during this chapter:

- Histograms
- Stem and leaf plots
- Pareto charts
- Pie charts
- Scatter plots (bivariate data)
- Parallel boxplots (continuous variables separated by a classification variable)
- Scatter plot matrix (several continuous variables)