

Lecture Notes: Writing and figures

The creation of a good figure is somewhat of a creative process. It is definitely not trivial. It is not sufficient to use a simple “plot” command and do nothing else. This will produce a figure, but it will not guarantee that the data you desire to show are clearly visible, or that everything is scaled and sized appropriately. As the creator of a figure, you need to know what you are trying to show. The figure-creating tool is designed to help you, but it cannot figure out for you which piece of your data is most critical to show. Be prepared to spend some time using the features of the figure-creating tool to adjust your figure, fine tuning its appearance and content to get to the desired result.

Recall that it is up to **you** to explain the figure to your reader. The figure is not a “data dump”. The reader does not want to have to sift through all your results to determine what is important. The purpose of a figure is to help you highlight to your reader the important results. Sometimes it is helpful to sketch out a figure using pen and paper before using the figure-creating tool. This can help the creative process and give you a goal towards which to work. Other times, you may find yourself iterating the figure multiple times within the tool. The first figure created may show some of what you want, but not all. Use the tool to explore variations until you get to the desired output.

There are many pitfalls that lessen the quality of a figure. The following examples are all taken from student papers. I scanned them from hardcopy printouts, so that the version displayed in this document was generated from a raster. As we know from previous lectures this is bad because it lessens the quality of the figure. But even so, most of these figures were originally produced using raster graphics and for some the original rasterization is very obvious. I am omitting figure captions from my document in order to show the original figure without confusion. Thus, if a figure caption was shown in the original document, I have included it here for reference.

In class, besides going through these figures, I will demonstrate some examples of how they could all be made better. I have also posted additional links at the class web site on guides to figure creation and a “top 10 worst figures published in research” that is quite humorous.

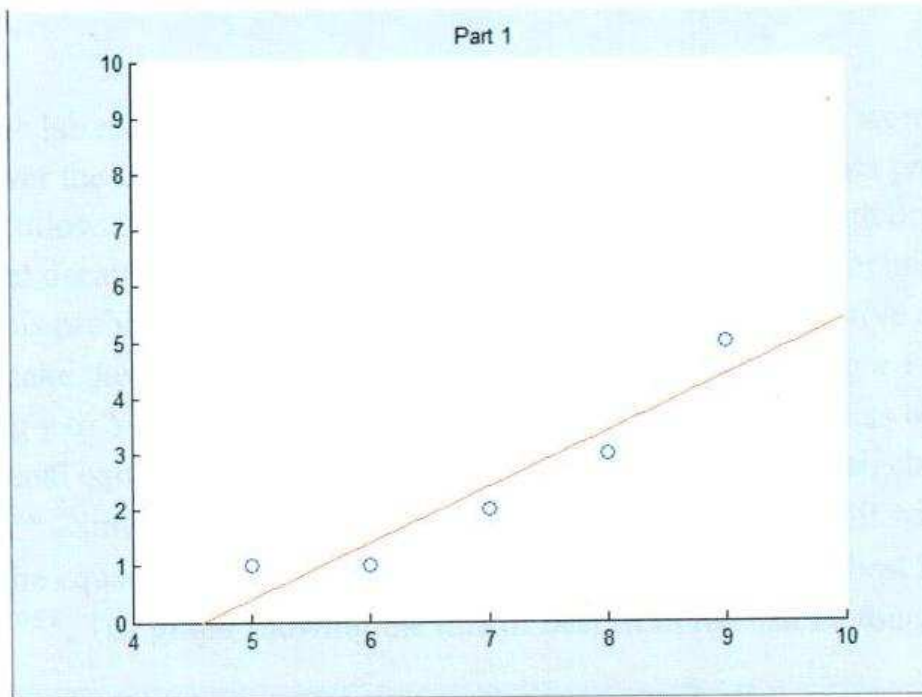


Figure 1 Best Fit Line for Part 1 Lab 1

This figure illustrates several problems. First, what is the purpose of the grey-shaded area around the graph? It is superfluous and wastes space. A simple black bounding box would be better. Second, the axes are unlabeled and there is no legend. The figure caption identifies a “line” but what are the circles? It would be appropriate to either label the circles as “data” and the line as “fitted line” in a legend, or to describe both in the figure caption. Note that having both a legend and details in the figure caption would be redundant, but this information should be in one or the other. Third, a heading “Part 1” is visible at the top and is repeated in the figure caption. This is redundant. There is no reason to have the heading. This information is generally best left to the caption. Fourth, the line and circles are fuzzy and difficult to see, clearly an artifact of rasterization. Fifth, color was used for the line and circles. Greyscale or simple black-and-white is generally preferred for superior contrast, inexpensive printing, and for clarity for the color-blind.

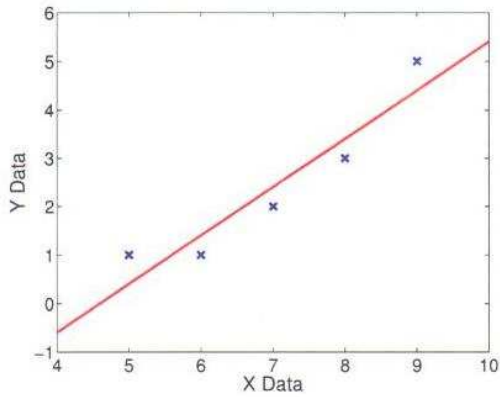


Fig. 1. Part 1 Results

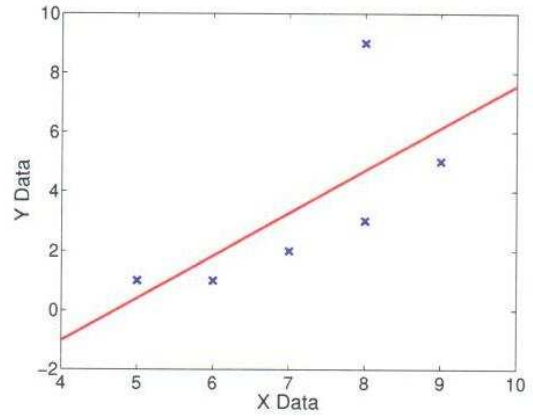


Fig. 2. Part 2 Results

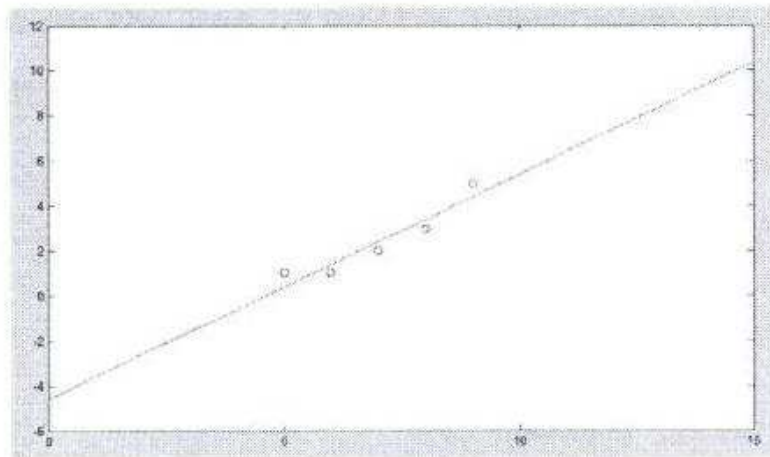
These figures have some of the same problems as the previous example. The use of color is unwarranted. There is no explanation of the symbols in either a legend or figure caption. A third problem is less obvious on first glance. Both plots have been auto-scaled, so that the scale of the Y-axis is different between the figures. Yet the point of these figures is to show what happens to the fitted line after an outlying point has been added. At first glance it looks like nothing happens to the line. Both figures appear to show a line with the same slope. Only after noticing the change in the Y-axis scale can one infer that the slope of the line changed quite a bit. How much it changed is difficult to visualize without mental warping. The moral of this example is to never use auto-scaling when showing multiple plots of similar (or the same) data. Fix the scale manually and keep it the same between the figures.

parameters using the Normal Equations. Report the calc for the line model. Show a plot of the raw data points an

The line that best fits the data is : $Y=aX+b$

The calculated value is: $a=1.0000$; $b=-4.6000$

So the line is: $Y=X-4.6$



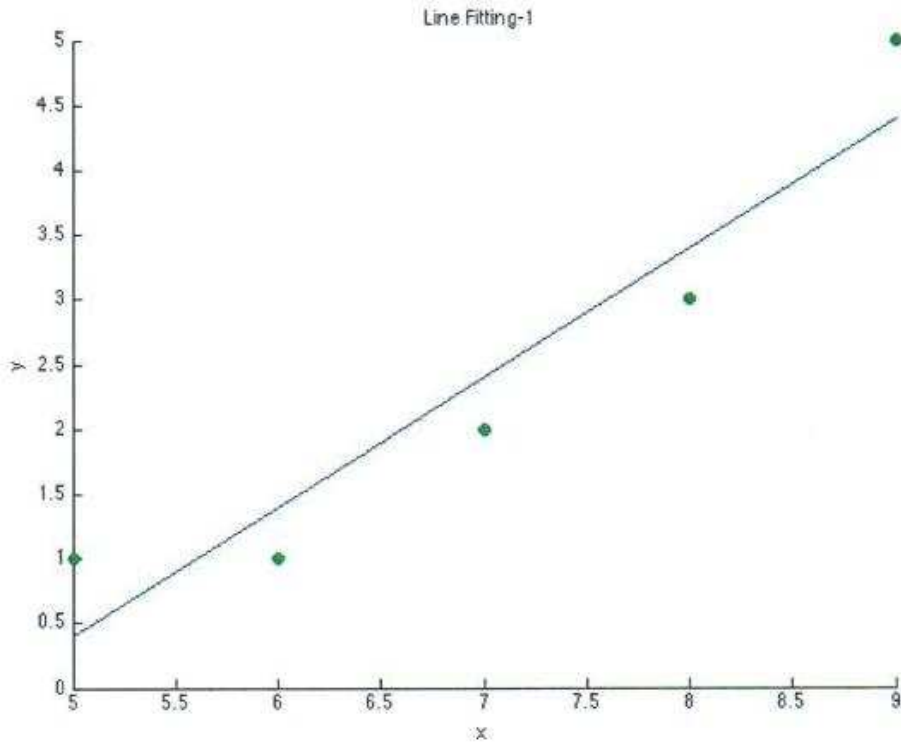
Code for the first part:

```
X1=[5,6,7,8,9]
```

```
Y1=[1,1,2,3,5]
```

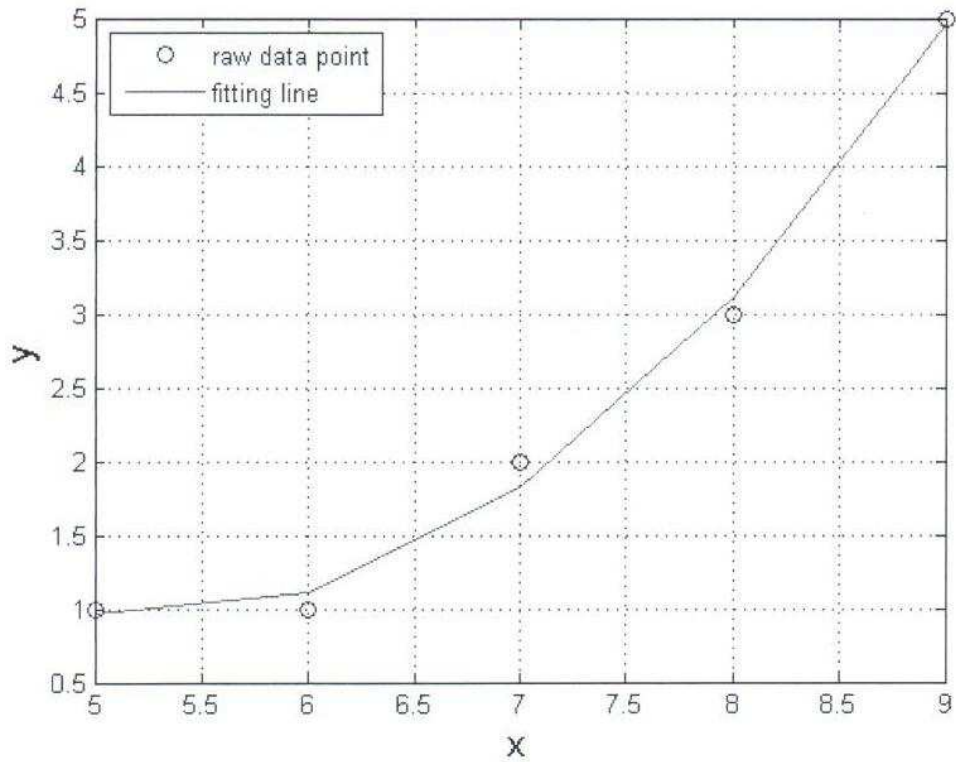
I grabbed this figure with a patch of its surrounding text. The figure has some of the same problems already discussed. It has a superfluous shaded grey boundary. It used low-resolution rasterization making everything fuzzy. The line and circles are too thin and faint. There is no legend or caption and the axes are not labeled. The surrounding text highlights another problem: the font used in the figure is too small. Compared to the fonts used in the surrounding text it is barely legible at the same scale. In general, fonts used in figures should be roughly the same size as the font used in the surrounding text.

On another note, at least 3 fonts can be seen to have been used in the surrounding text. This is also bad. Font use within the body of a document should be stable and consistent.



Graph of the line over given data

This example again shows some of the same problems: the use of color, the lack of a legend of similar information in the caption, the use of a redundant heading, and too small of a font. The figure caption has no number which makes referencing it difficult. Extreme rasterization can also be seen in the fuzziness of the line and the axes.



Another example showing extreme rasterization (fuzziness). The font size is good and the symbols are identified in the legend, but there is no caption. The reason for including a grid is not apparent. In general a grid is helpful for locating data more precisely; it is not clear why that is necessary for this figure. This is part of a larger issue of using “fancy graphics” just because they are available. In general that is a bad idea.

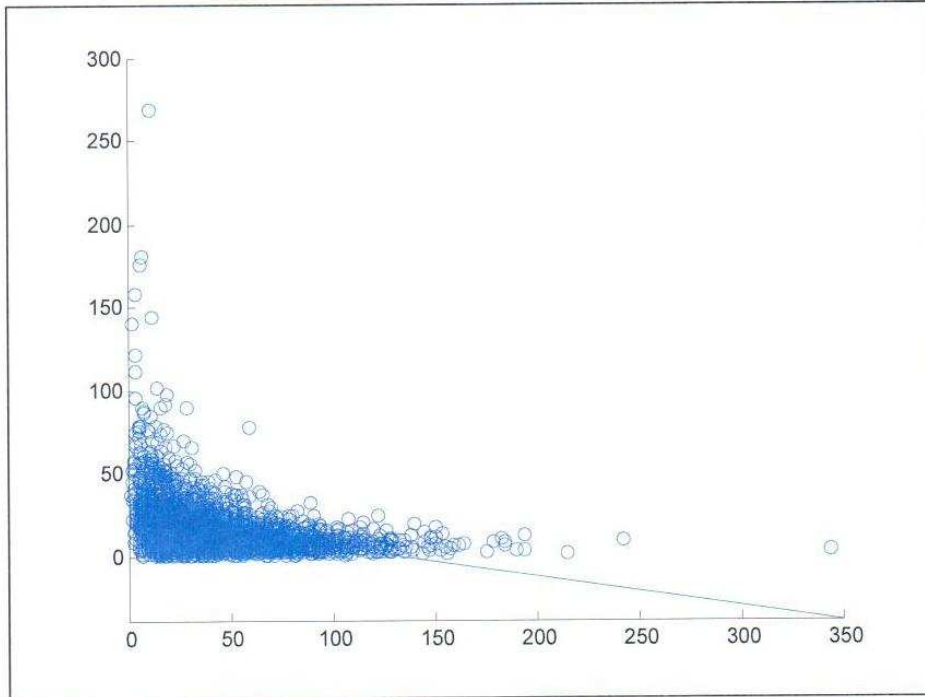
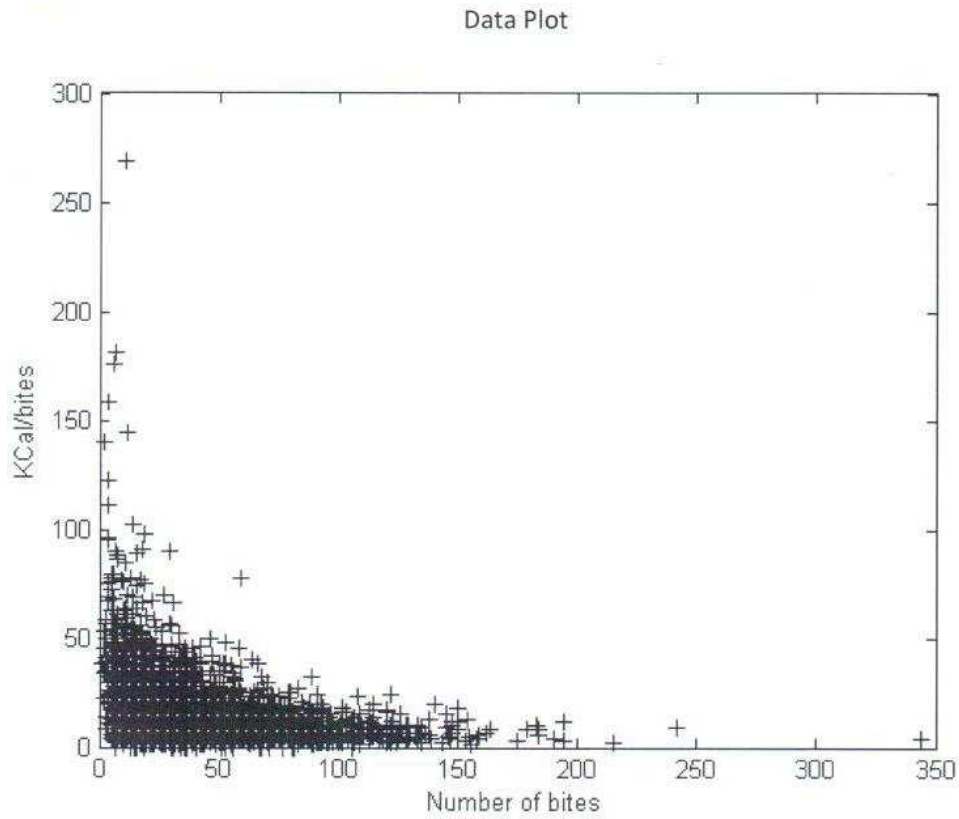
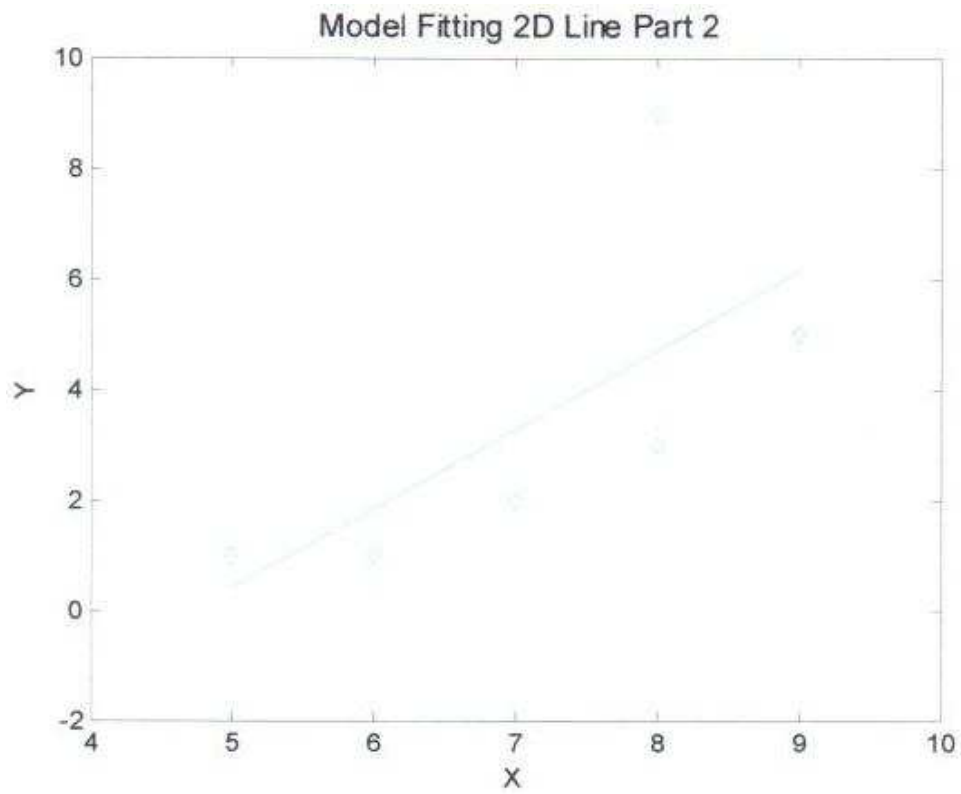


Fig3: Data set 3, with solution $(-0.1771, 23.4417)$

This figure has some of the same problems already discussed: the use of color and rasterization. The font for the axes units is appropriately sized, but the font for the figure caption is too small. The axes are not labeled. The figure caption includes numbers but their meaning is unclear. Default precision of four decimal places is given, but it is not clear why so many digits are needed. The symbol chosen for the data points is a hollow blue circle, but it is too large for the given amount of data. Printing the given symbol at every given data location results in a large blob that is indecipherable. The fitted model uses a thin line of the same color, which cannot be seen inside the large blob of data. Autoscaling has been used, causing most of the figure to consist of empty white space. The figure would probably be better served by manually scaling the X and Y axes to zoom in to the majority of the data, at the expense of not displaying every point. One might wonder if it is okay to clip some of the data, but you should answer that question with the following question: what is the point of the figure? Is it to show all the data, or show the model that best fits the majority of the data? Other ideas would be to wean the data set (manually) so that it is not so dense near the origin, for example stating in the caption that only the first 100 data are shown (for clarity). Other ideas are to use a smaller symbol for each point, or transform it to a different scale. There is also unnecessary white space surrounding the figure (but still inside the bounding box), further reducing the size of the space for the actual figure, which is already too crowded.



This example is similar to the last and shows many of the problems already discussed. There is no figure caption but there is a redundant heading. Autoscaling causes most of the figure space to be empty. There is severe crowding for the given amount of data and symbol chosen to display each point. There is a pattern of “dots” in the blob of data near the origin due to the spacing of the symbols. This may be interesting in that it shows a quantization of the data, but we cannot tell at the given scale.



This figure shows many of the usual problems. But the most obvious problem is that the data and line are barely visible.