A picture is worth a thousand words. Before formal analysis, it is often a good idea to plot the data to get a feel and gain some insights.

For categorical data, standard plots include **bar charts** and **pie charts**.

For numerical data, **histograms** and **scatter plots** are commonly used. Side-by-side **boxplots** are effective in contrasting multiple 1-D distributions. **Scatter plot matrices** should be looked at before multiple regression is attempted.

Scatter plots and variants can be customized to serve a great variety of practical needs. For the plots to be visually telling, one should adhere to the following principles, of which some are set as defaults in `R` graphics.

- Make the data stand out. Minimize clutter.
- Box the plotting region with scale lines, and put tickmarks on the outside. Do not overdo the number of tickmarks.
- Add reference lines when necessary.
- Set edges of the plotting box just outside of the data ranges to "fill" the box.
- Transform highly skewed data to spread the scatter more evenly.
- Use colors and plotting symbols to distinguish superposed data.
- Overlapping plotting symbols should be visually distinguishable; jitter if necessary.
- Add connecting lines to mark groups, to show trends, etc.
- Put keys and markers away from the data region. Put notes in the legend or text.
- Use the log scale for multiplicative factors such as percent changes.

Average brain weights (in g) and body weights (in kg) were collected for 28 species of land animals. A simple linear regression line,

$$\log_{10}(\text{brain weight}) = 0.9339 + 0.7523 \log_{10}(\text{body weight}) + \epsilon,$$

fits the 25 living species reasonably well.

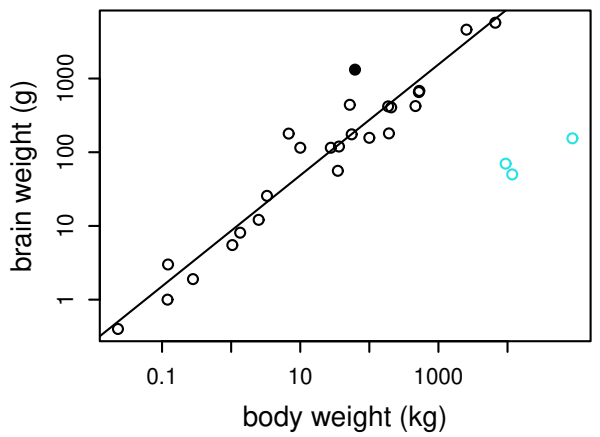

Figure 1: Data and fit. Dinosaurs are faded.

The weights are all positive and vary over several orders of magnitude, so we work on the log scale with both. As seen in Figure 1, three of the data points do not belong to the group, which happen to be the dinosaur species *Triceratops*, *Dipliodocus*, and *Brachiosaurus*.

It appears that the brain weight is roughly proportional to the 3/4 power of the body weight. As one may expect, human, the filled circle in Figure 1, has the largest relative brain size $\frac{\text{brain}}{(\text{body})^{3/4}} = 59.7$, a cut above the next two at 42.5 (Rhesus Monkey) and 22.7 (Chimpanzee).