Babraham
Bioinformatics

Exercises:
Plotting Complex Figures Using R

Version 1.1

Licence

Exercise 1: Customising simple plots

- The file weight_chart.txt contains data for a growth chart for a typical baby over the first 9 months of its life. Use the plot function to draw this as a point an line graph with the following changes
  - Change the point character (pch) to be a filled square
  - Change the plot point size to be 1.5x normal size
  - Change the line thickness to be 2 pixels
  - Change the y-axis to scale between 2 and 10kg
  - Change the x-axis title to be Age (months)
  - Change the y-axis title to be Weight (kg)
  - Add a suitable title to the top of the plot

- The file feature_counts.txt contains a summary of the number of features of different types in the mouse GRCm38 genome. Can you plot this as a barplot but with the following changes
  - The bars should be horizontal rather than vertical.
  - The count axis should be labelled.
  - The feature names should be added to the y axis.
  - The plot should be given a suitable title.
  - The text labels should all be horizontal (par las parameter)
  - The margins should be adjusted to accommodate the labels (par mar parameter)

- **[Extension if you have time**] Use this hist function to plot out the distribution of 10000 points sampled from a standard normal distribution (rnorm) along with another 10000 points sampled from the same distribution but with an offset of 4.
  - Example: `c(rnorm(10000),rnorm(10000)+4)`
  - Find a suitable number of breaks to make the plot appear

Exercise 2: Using colour

- The file male_female_counts.txt contains a time series split into male and female count values.
  - Plot this as a barchart
  - Make all bars different colours using the rainbow function
  - Replot, and make the bars for the males a different colour to those for the females

- The file up_down_expression.txt contains an expression comparison dataset, but has an extra column which classifies the rows into one of 3 groups (up,down or unchanging). Plot this as a scatterplot (plot) with the up being red, the down being blue and the unchanging being grey.
  - Read in the file
  - Determine the order of the levels for the ‘State’ column
  - Use the palette function to set the corresponding colours you want to use
  - Set the State column as the col parameter to divide the data by this column when plotting

- The file colour_to_value_map.r contains a function to map a value from a range to a colour from a predefined palette. The file expression_methylation.txt contains data for gene body methylation, promoter methylation and gene expression. Construct a scatterplot of promoter methylation vs gene body methylation and colour this by gene expression.

  - You will need to construct a colour palette function from grey to red using `colorRampPalette` and then call this function to generate a palette of 100 colours.
  - You will need to use the `range` function on the expression data to get the range of values you need to cover.
  - You can then call the `map.colours` function you've been given with the set of expression values, the value range and the palette to be given a set of colours which you can pass to the col parameter.

- **[Extension if you have time**] Modify the previous plot so that the colour range goes from 2 to 8 and values outside this get the colour for 2 or 8. Change the colour scheme to be a diverging scheme centred on 5.

Exercise 3: Using overlays

- The file `chromosome_position_data.txt` contains positional count data for 3 different datasets (a WT and two mutants). Plot this as a line graph showing the 3 different datasets overlaid.
  - You'll need to do an initial `plot` with `type="l"` specified followed by two additional layers using the `lines` function.
  - Remember to calculate the full `range` of values across all 3 datasets when doing the initial plot so that all of the data will fit into the plot area.
  - For the colours generate a 3 colour palette from the `"Set1" RColorBrewer` set using the `brewer.pal` function.
  - Add a `legend` to the `topleft` of the plot with the data names and the corresponding colours.
  - Make the lines 2px thick (`lwd`)
  - Add suitable labels to each axis.

- The file `brain_bodyweight.txt` contains data for the log10 brain and bodyweight for a range of species, along with an SEM measure for each point. Plot these data on a scatterplot and add the following annotation
  - Darkgrey `arrows` with flat ends to indicate the confidence intervals for the measures. You'll need to draw 4 `arrows` to show the positive and negative SEM intervals for each of the measurements. Use `angle=90` to make the arrow ends flat and `length=0.05` to make the ends a sensible size.
  - Add the names of the species just below each data point. You'll need to use the `text` function to do this. Make the font size (`cex`) 0.7 so the text is smaller.

- The file `different_scale.txt` contains two datasets, both of which need to be plotted as line graphs. The data are on very different scales. Can you plot these out two ways
  - As two separate line graphs next to each other in a multi-panel plot
  - As two lines in a single graph, but with the second data series being on a different scale to the first. Add a second `axis` to the right (side=4) of the plot for the second data set. Add a label to the secondary axis using `mtext(side=4, line=3)`