

Assignments to Lecture 5 – Plots

Danny Arends

Plots

The data in the genotype.txt is a subset of the data we collected using the MegaMuga mouse genotyping array. For the different advanced intercross line (AIL) individuals weights are measured at several days, which can be found in the phenotypes.txt file. The genetic map of the markers used is found in map.txt. Please note *this data is from a current analysis, and is not public (yet)*.

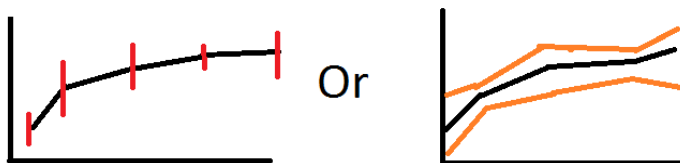
1) Load in the different datasets (genotypes and phenotypes) using the `read.table()` or `read.csv()` function, make sure to load in headers, rownames, set the correct separator and colClasses.

Basic curves

Lets analyse some of the data, we'll start with the phenotypes. The d<number> measurements stand for day of weight measurement. The column WG2 is the litter size (wurfgröße) after two days (since in the first couple of days new born mice tend to die or get eaten by the mother). Be sure to setup the drawing area of the different plots beforehand using the `plot()` function, then add the data using the `points()` function.

2a) Create a global growthcurve showing the mean weight for all the animals combined, try out the different plotting types, and select one that you like the best.

2b) Add to this plot 1) lines show the variation deviation at the measurement day, or 2) error bars that show the same thing (see the sketch), don't forget to add a legend describing the colors, and add good figure axis and a main title



2c) Create a new line plot that shows the median at each measurement date and in light gray all the measurements of the different individuals.

Multiple plots in the same window

3a) We're now going to split up the different WG groups, divide the animals in 2 groups ($WG < 10$ and $WG \geq 10$). Create a mulitplot that shows the growth curves per group, in either the style of 2b or in the style of 2c. Can you see a difference ?

3b) Create 4 x 2 notched boxplots (one for each timepoint), showing the combined data of individuals in the 2 groups, label the groups correctly in the plot, what do we learn about the effect of litter size on bodyweight ? make sure the axis are readable, and individual boxplots have a main title which mentions the day.

- Image

4) Using the data from the `genotypes.txt` file, first use the data from `map.txt` to subset the genotypes for a single chromosome. After you created a subset for e.g. chromosome 1 create an image for the chromosome showing the different SNPs in the order in which they occur on the map. This means:

- a) Sort the map on basepair location, using the `order()` or the `sort()` function
- b) Iterate through the chromosomes and select the SNPs (e.g. using their `rowname`) from the genotype matrix, that are on that chromosome
- c) Convert the genotypes to numeric values using `as.factor / as.numeric`
- d) Open a device for plotting (e.g. PNG or JPG) (use the chromosome name to create unique filenames)
- e) Create a plot that shows the different genotypes on the chromosome
- f) Close the plotting device

- Chromosome plot

5) From the data create a chromosome plot, showing the locations of the markers on the chromosomes. Use the code that was given as an example during the lecture.

Additional:

6) Color the different markers on the chromosome based on the majority of the genotypes at that marker (A, H or B) in case of genotypes having the same amount (e.g. no majority) mix the colors. The easiest here is to choose red, green and blue as the colors for the different genotypes, and then use the `rgb()` function to mix them.