

Lin & Jan

Question	Result	Points	Lin & Jan
1	Loading data; <code>dim()</code> , <code>str()</code> , <code>names()</code>	5	2
	Using <code>min()</code> , <code>max()</code> , <code>quantile()</code>	5	3
	Histogram, density of samples, also on log-transformed data	5	3
	Histogram, density of means, also on log-transformed data	5	3
	Notice: different distributions per class	10	0
2	Scatterplot	5	5
	Notice: mostly $x = y$, some points away (diff. expr.)	10	10
	Notice: one highly expressed gene, GAPDH	5	0
3	Histogram of differences	10	10
	Notice: more underexpressed genes, verify by counting	10	10
4	Boxplots per gene	10	10
	Notice: no diff. expr., underexpressed, overexpressed	10	10
	Grade	10	7,6

Q1

- You should always perform a sanity check of your data frame using `dim()`, `str()`, `names()` etc. This will help you to really truly understand what kind of data you are working with
- Using `quantile()` in addition to `range()` or `min()` and `max()` gives you a better idea of the distribution of your data points
- It is generally wise to produce figures containing log-transformed data in addition to figures containing untransformed data, especially when working with data which show large differences in scale, such as gene expression levels.
- The general distributions of gene expression are different between the two conditions, as opposed to similar as you wrote in your report.

Q2

- Be mindful of outliers: they may correspond to particularly interesting observations or indicate some form of technical or biological bias. A scatterplot of the untransformed means would have made the outlier clearer