Real-world data wrangling with R

1. Merge ASV tables from two different sequencing runs
2. Remove all samples that have been sequenced by “Novogene” (hint: column names, PXXXSX)
3. Remove all ASVs with zero abundance across all remaining samples (hint: keep an eye on number of samples/ASVs removed etc.! → code annotation)
4. Inspect raw metadata table containing clinical data and find column to match it to the ASV table
5. Adjust sample names in ASV table to match metadata
6. Inspect miss-matches between samples in ASV tables and samples in metadata → consult with Mel what do do about them!
7. Using final metadata table, create some useful summary statistics:
    a. How many samples are we working with?
    b. From how many patients?
    c. How many samples/patient, samples/time point etc.?
    d. How many responders vs non-responders/NE vs non-NE patients/gut decontamination vs no gut decontamination patients are contained in the data? Any differences between chemo-cycles?
    e. Mean age of patients (overall/split by sex)?
    f. Any other summaries you might find useful...

Note: All of the code (and output) for 7. will potentially be very useful for your report writing!

/nfs/course/551-1119-00L_masterdata/tutorials/data_wrangling