Data description (TCGA): <https://cancergenome.nih.gov/cancersselected/thyroid>

BioConductor:

* biocLite()
* library(SummarizedExperiment)
* load SE object form .rds

Pre-processing:

* quality assessment and normalization using DGEList
* examine sequencing depth of maped reads (bar plot)
* CPM scaling
* Plot log2 CPM
* Identify samples with distinctive RNA composition
* Identify lowly expressed genes
* Filter out genes and compare both approaches
* MA-plot
* TMM method (adjust library size)
* Some comparison of un-normalized with normalized
* Normalization between-sample (see example in slides)
* Multidimensional plots
* Filter Sammples with problems

Batch Effect:

Consider adding some plots that appear in slides ex: hitmaps comparing with batch effect and once filtered.

* Identify Batch effect
* Examine the cross-classification of outcome of interes with the possible surrogate variables
* do a hierarchical clustering
* multidimensional scaling
* Adjuste for batch effect
  + SVA
* If it does not work (do also and compare), remove batch effect
  + ComBat
  + QR decomposition
  + SVD

(compare the three cases)

Differential Expression Analysis

* See possible blockings in experimental design
* use logarithmic scale
* compare expression values between two groups gene by gene (fold-change)
* rank genes by decreasing value of log fold change (significance)
* check slides to see the example on multiple testing corrections
  + Bonferroni
  + FDR

(choose which one is better in our case)

* Volcano plot
* Consider non-specific filtering (see slides)

To be continued…