Exercises: PCA, tSNE

1. Plot the two-dimentional "twodimentional.Rdata". Conduct PCA and plot the data in new coordinates (prcomp). What are the new coordinates?

2. Calculate the covariance matrix for the "twodimentional" (cov). How much variance is explained by each of the original coordinates $e_1 = (1, 0)$ and $e_2 = (0, 1)$? How much variance is explained by a vector $v = (a, b)$? Can you interpret the first principal component?

3. Plot 'data3', run PCA with and without centering. What is the difference?

4. Plot 'data4', run PCA with and without scaling. What is the difference?

5. (Single cell) Download 'singleCellCounts.Rdata'. These are transcriptional profiles of mouse brain cells, an integer represents the numbers of molecules of a particular gene in a particular cell.

   (a) Perform PCA on raw counts, color cells by library size (total number of molecules per cell). What can you say?

   (b) Perform PCA on normalized counts, try different normalizations.

   (c) Perform tSNE on properly normalised data. How many cluster of cells can you see? Change the perplexity parameter in tSNE and compare the results.

   (d) Download 'MarkerGenes.Rdata' table and try to identify which cell are likely to be glutamatergic neurons, which cell are GABA neurons, and which cells are not neurons.