

Practical instructions Essentials Course Unsupervised learning day 1

General overview this week

Day 1

- 10:00-10:30: short introductory lecture
- 10:35-11:20 lecture clustering
- 11:20-13:30: clustering practical
- 13:30-14:00: self-supervised learning lecture
- 14:00-16:00 continue clustering practical. Free to go when done!

Learning goals: understanding and working with K-means clustering and hierarchical clustering, able to explain what self-supervised learning is

Day 2

- 10:00-10:45 lecture dimensionality reduction
- 10:45-15:00 dim. reduction practical (as far as you get, no problem if not finished!)
- 15:00-15:30 practical discussion
- 15:30-16:00 answering questions paper

Learning goals: understanding the reason for, and working with, dimensionality reduction tools such as PCA, t-SNE and UMAP.

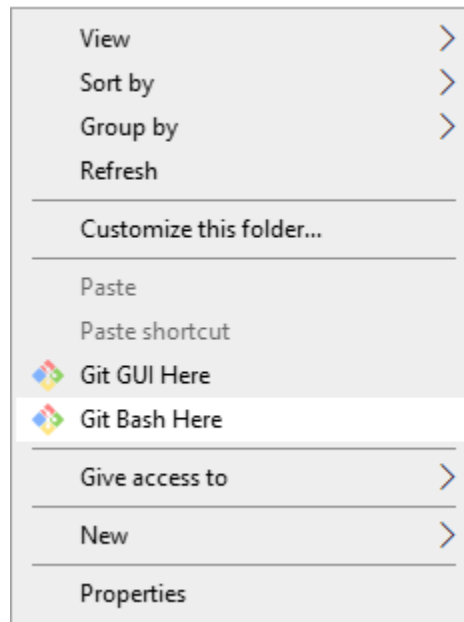
Practical today

- Get your hands (ever so slightly) dirty with k-means and hierarchical clustering.

Clustering practical

Getting materials

You first need to download the practical materials to your computer. They are in a repository on Github. A simple way to download this data is to install [Git Bash](#) (if on Windows), go to the folder where you want to download the materials, and open Git Bash there. You can open it in any folder by right clicking:



Then, simply type git clone

<https://github.com/DieStok/PracticalEssentialsCourseUUSeptember2023.git>. This should work also on Mac or Linux (you don't need Git Bash then, git is just in your terminal).

If that's not for you, you can also go [here](#) and download all the files in a zip.

Working on the practical

The practical is self-paced and should be self-explanatory. It starts with k-means clustering and showing how it comes about, while showcasing the usage of the ggplot2 and gganimate libraries. Then comes hierarchical clustering. That's it for the first part.

The second part, which we deal with on Thursday, discusses k-means and PCA in the context of dimensionality and noise. We then do a small comparison of PCA, t-SNE and UMAP on some data (where you are also required to watch videos about t-SNE and UMAP). What follows is a part about Eigenfaces. This is the most non-essential, though quite interesting, part of the practical. It can be skipped if you don't have the time. We end with **optionally** trying to recreate a paper figure subpanel from actual published RNAseq data. Optionally means it's also skippable, so rejoice!

Please go through the practical individually, but be sure to discuss, compare, and contrast with your neighbours. Please help others who might have had less exposure to R if you think you can. If there are any questions, please raise your hand and we will diligently come over to exterminate any and all questions.

Good luck!