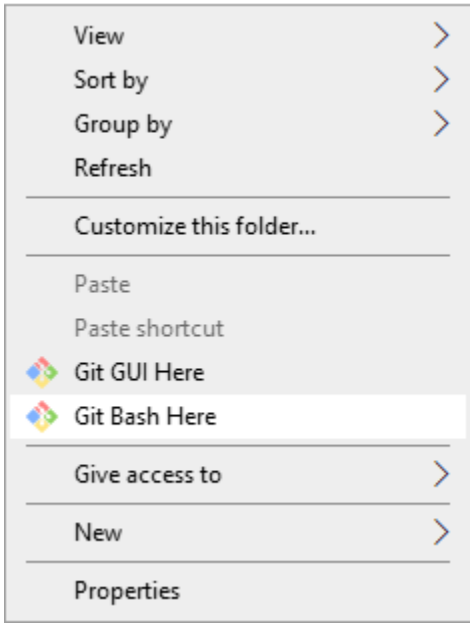
Practical instructions Essentials Course Unsupervised learning 2024

Lots of concepts, little doing. Now there will be some doing.

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/BIBCEssentials2024.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

1. Go sit with/close to the team with whom you shared your paper assignment. Then make sure one group does practical 1, and the other does practical 2 for about 1.5 hours (i.e. you get how far you get in 1.5 hours), and inform each other on what you learned.

2. As a final exercise, there's a small patient dataset of gene expression for 49 cancer patients and 49 healthy controls **In the Teams Files/Hands-on folder**. See if you can come up with some meaningful clustering/dimensionality reduction visualisation figure that gives insight into this data.

Practical 1 deals with k-means clustering and showing how it comes about, while showcasing the usage of the ggplot2 and gganimate libraries (although gganimate is optional and sometimes finicky so feel free to ignore that). Then comes hierarchical clustering.
Practical 2 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 the MNIST data.

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 I will diligently come over to exterminate any and all questions.

Good luck!