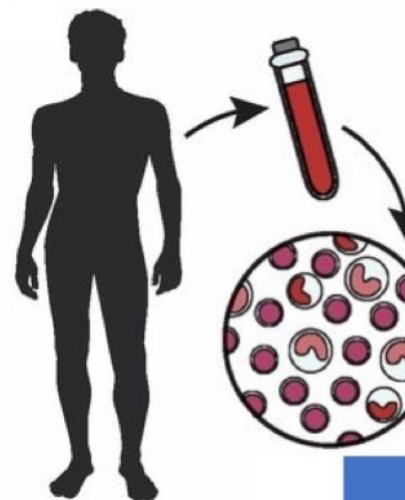
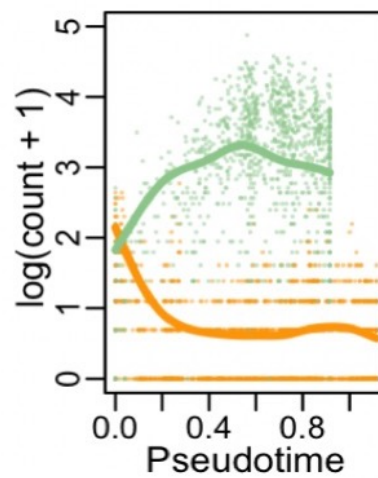
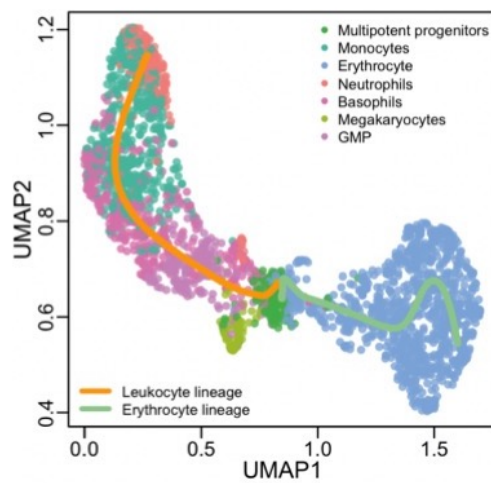


High Dimensional Data Analysis 2021 (HDDA21)

Contents



	Cell 1	Cell 2	Cell 3	...
g1	0	10	0	...
g2	8	0	0	...
g3	0	15	0	...
g4	3	0	20	...
...

	bu
g1	34
g2	80

0.0.1 Course Description

Modern high throughput technologies easily generate data on thousands of variables; e.g. genomics, chemometrics, environmental monitoring, ... Conventional statistical methods are no longer suited for effectively analysing such high-dimensional data. Multivariate statistical methods may be used, but for often the dimensionality of the data set is much larger than the number of (biological) samples. Modern advances in statistical data analyses allow for the appropriate analysis of such data. Methods for the analysis of high dimensional data rely heavily on multivariate statistical methods. Therefore a large part of the course content is devoted to multivariate methods, but with a focus on high dimensional settings and issues. Multivariate statistical analysis covers many methods. In this course, only a few are discussed. A selection of techniques is made based on our experience that they are frequently used in industry and research institutes (e.g. principal component analysis, cluster analysis, classification methods). Central in the course are applications from different fields (analytical chemistry, ecology, biotechnology, genomics, ...).

0.0.2 Prerequisites

The prerequisites for the High Dimensional Data Analysis course are the successful completion of a basic course of statistics that covers topics on data exploration and descriptive statistics, statistical modeling, and inference: linear models, confidence intervals, t-tests, F-tests, anova, chi-squared test.

The basis concepts may be revisited in my online course <https://gtpb.github.io/PSLS20/> (English) and in <https://statomics.github.io/statistiekCursusNotas/> (Dutch).

A primer to R and Data visualisation in R can be found in:

- R Basics: <https://dodona.ugent.be/nl/courses/335/>
- R Data Exploration: <https://dodona.ugent.be/nl/courses/345/>

0.0.3 Issues

If you encounter any problems related to the course material (e.g. package installation problems, bugs in the code, typos, ...), please consider posting an issue on GitHub.

Questions related to the course contents can be asked by posting on UFora, by contacting the teachers by email or during the lectures or practical sessions.

0.0.4 Topics

0.0.4.1 1. Introduction

- Slides: Lecture 1: Intro
- Software: Install and Launch Statistical Software
- Introduction [PDF]
- Linear regression Geometric Interpretation

0.0.4.2 2. Singular value decomposition

- Lecture 2-3: Singular value decomposition [PDF]
- Geometric Interpretation
- MDS Link Gram Distance Matrix

0.0.4.3 3. Prediction with High Dimensional Predictors

- Lecture 4-5: Prediction Theory [PDF]

0.0.4.4 4. Sparse Singular Value Decomposition

- Lecture 5: Sparse SVD [PDF]

0.0.4.5 5. Linear discriminant analysis

- Lecture 5: Linear Discriminant Analysis Theory [PDF]

0.0.4.6 6. Large Scale inference

- Lecture 6: Large Scale Inference [PDF]

0.0.4.7 Instructors

- Lieven Clement
- Milan Malfait