

Applied Multivariate Statistics

Fall-Winter-Semester 2020

University of Mannheim
Department of Economics
Chair of Statistics
Dr. Toni Stocker

Applied Multivariate Statistics (AMS) - Content

Introduction to AMS	1
Matrix Algebra	20
Multivariate Samples	57
Principal Component Analysis (PCA)	77
Biplots	129
Factor Analysis	141
Multidimensional Scaling (MDS)	152
Cluster Analysis	170
Linear Discriminant Analysis (LDA)	183
Binary Response Models	194
Statistical Methods for Data Science	212

Introduction to AMS

General Course Information

Prerequisites

- Students in Economics from Mannheim: no problem
- All other students:
should have attended two or more courses in Statistics
(descriptive statistics, estimating and hypothesis testing)
- A course in *Basic Econometrics* or *Linear Algebra* would be helpful but is not strictly required.
- The statistical software R will intensively be used throughout this course. Students who are not yet familiar with R should work through chapters 1-5 of the *R introduction* (see course folder) on their own by October 5 at the latest.
- Though R is easy to learn, you need to invest some time at the beginning. But you may benefit from it for a long time.

General Course Information

Time and Locations

Format	Day	Time	Method
Lecture	provided till Friday noon		Video (not live)
Group Exercise Session 1	Thursday	17:15-18:45	Online, Live (Zoom)
Group Exercise Session 2	Friday	08:30-10:00	Online, Live (Zoom)
R-Tutorial (optional)	Friday	15:30-17:00	Online, Live (Zoom)

Choose one of the two compulsory GE-Sessions! They start in the 2nd week.
The R Tutorial is not compulsory in terms of grading. It starts in the 1st week.

Contact

Office Hour: Wednesday, 3:00-4:30 p.m. only online (Zoom)

Office: L7, 3-5, 1st floor, 143

Phone: 0621-181-3963

Email: stocker@uni-mannheim.de

General Course Information

Course Material

Slides (Lecture), **Assignments** (Tutorials), **'Introduction to R'** (see p. 2)

Material will be updated weekly (Friday) to find in course folder at **Studierendenportal (ILIAS)**

References

- R. Johnson, D. Wichern (2007):
Applied Multivariate Statistical Analysis; Pearson Intl. Ed. Main Reference
- A. Rencher (2002):
Methods of Multivariate Analysis; Wiley.
- W. Härdle, L. Simar (2003):
Applied Multivariate Statistical Analysis; Wiley.
- A. J. Izenman (2008):
Modern Multivariate Statistical Techniques; Springer.
- P. Hewson (2009):
Multivariate Statistics with R; Open Text Book.

Examination

Exam + Assignments:

80% written exam (120 minutes) + 20% Exercises (individual and collaborative part) in terms of points to earn in total.

Example:

	Points
Written Exam:	60 (from 80)
<u>Exercises:</u>	<u>18 (from 20):</u>
Total:	78 (from 100)

=> Grading will be based on 78 points (from 100)

Minimum for passing: ≤ 40

Assignments:

Need to submit homework and attend tutorial. To get full points (20) you need to work at least on 10 assignments (out of 12) in a meaningful way.

(See *Guidelines for Assignments*)

What is it about?

Issues of Applied Multivariate Statistics (AMS)

Multivariate analysis consists of a collection of methods that can be used when several measurements are made on each individual or object in one or more samples.

See Renchner (2002), p.1

Objectives

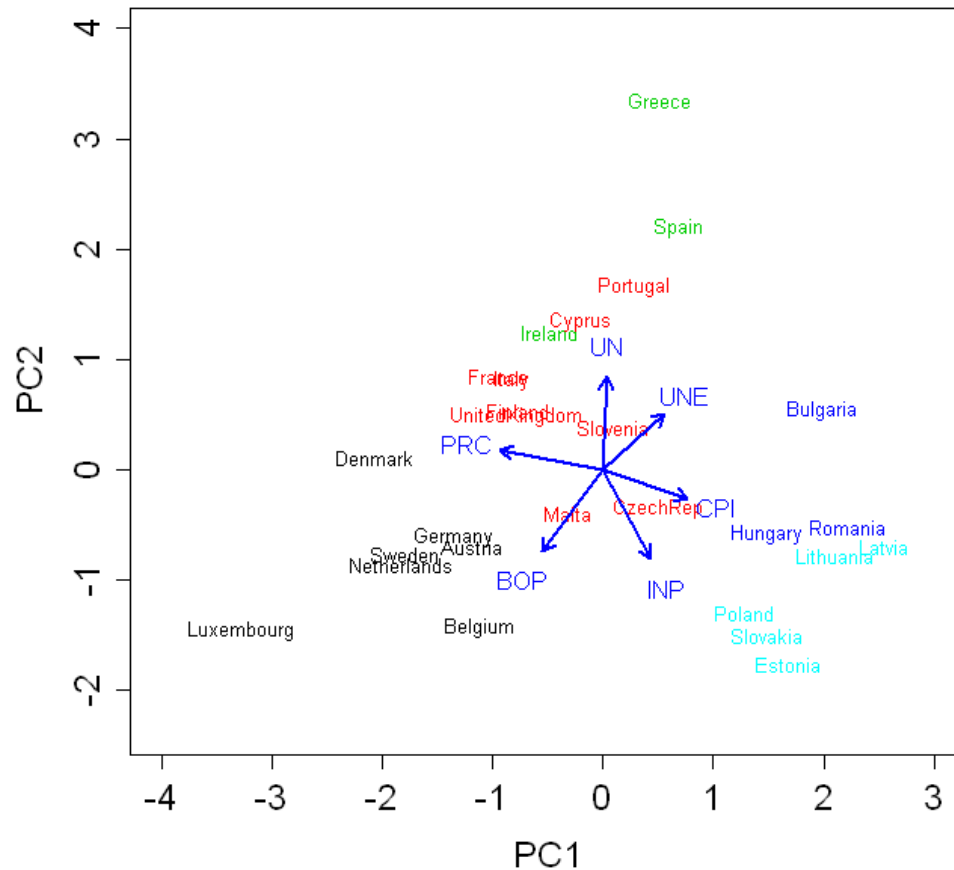
- Dimension reduction and structural simplification
- Visualization of high-dimensional data
- Investigation of the dependence among variables
- Grouping, discrimination and classification
- Close link to other areas such as Exploratory Data Analysis (EDA) and Data Mining

(see also J+W (2007), p.2)

What is it about?

Example 1: Dimension Reduction

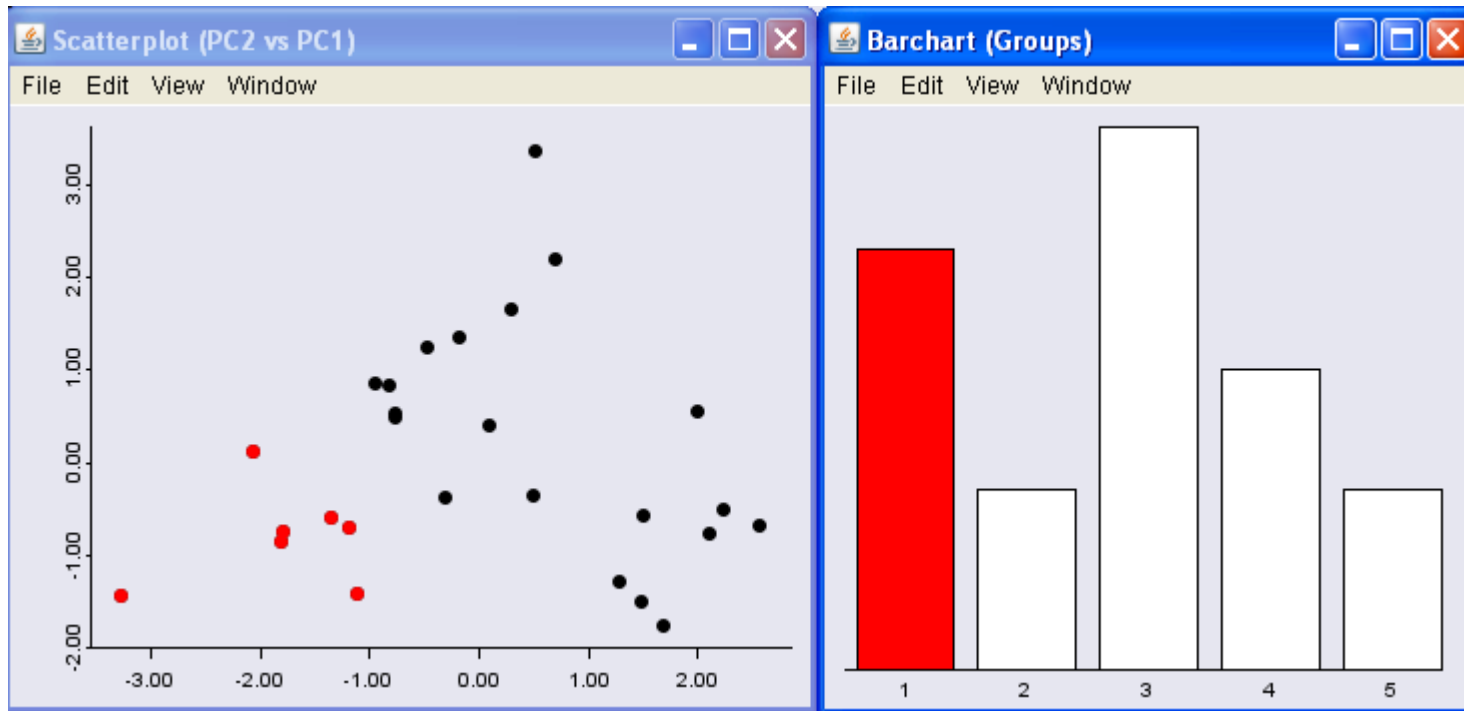
Economic Indicators for the 27 European Union Countries in 2011
(see *WIREs Comput Stat* 2012, 4:399–406. doi: 10.1002/wics.1200)



What is it about?

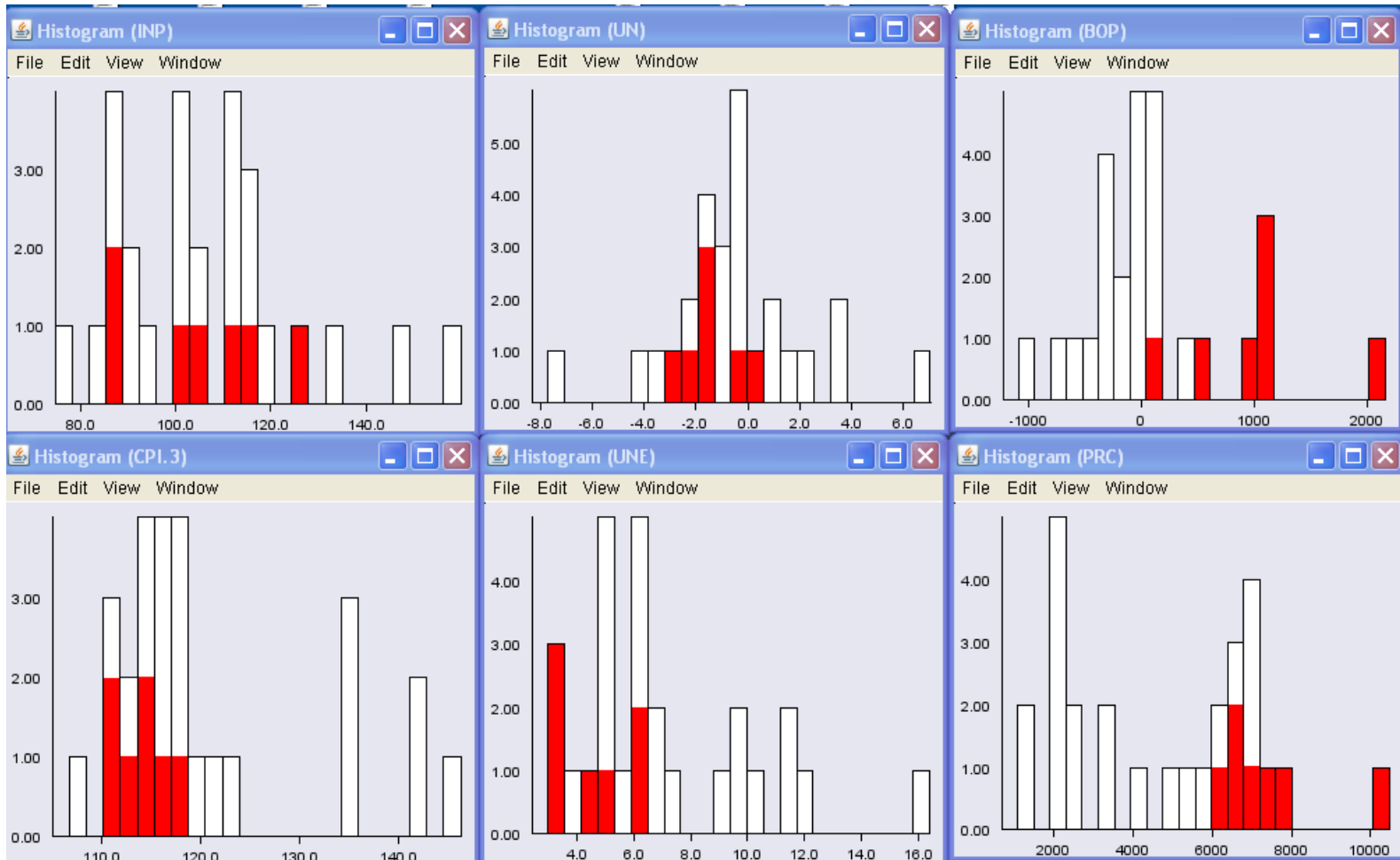
Example 2: Brushing and Linking

... using R (iplots)



What is it about?

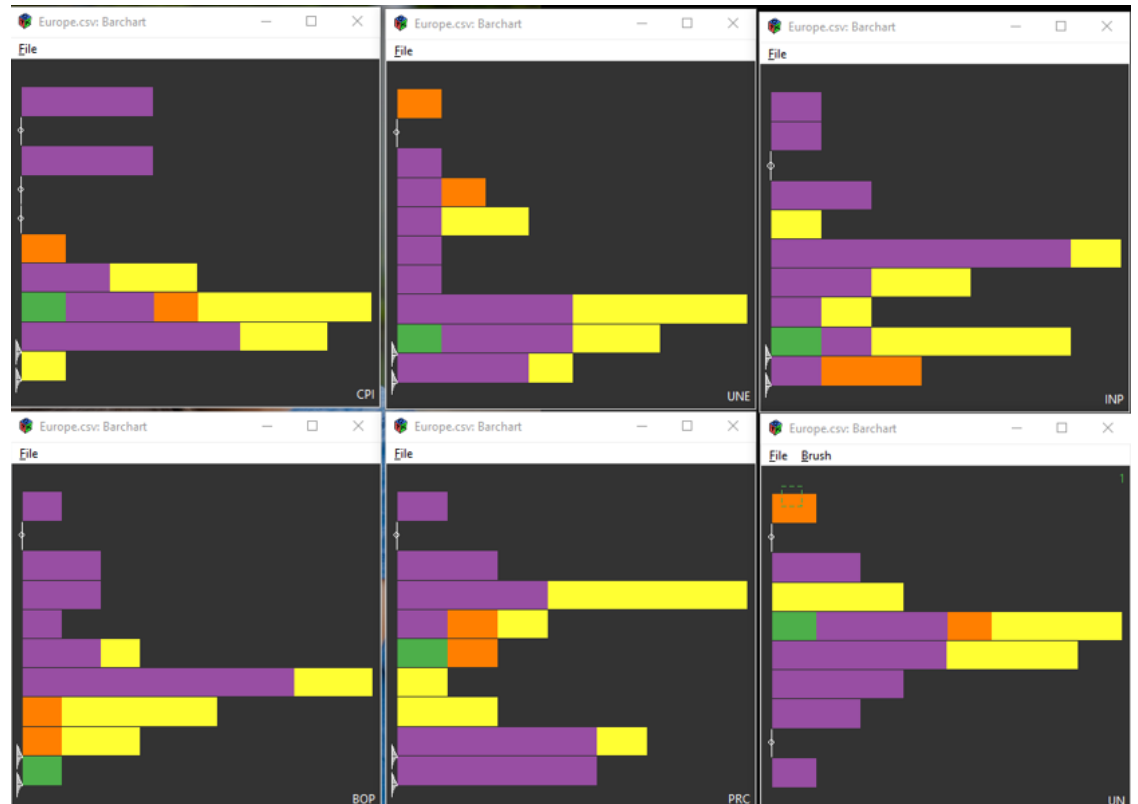
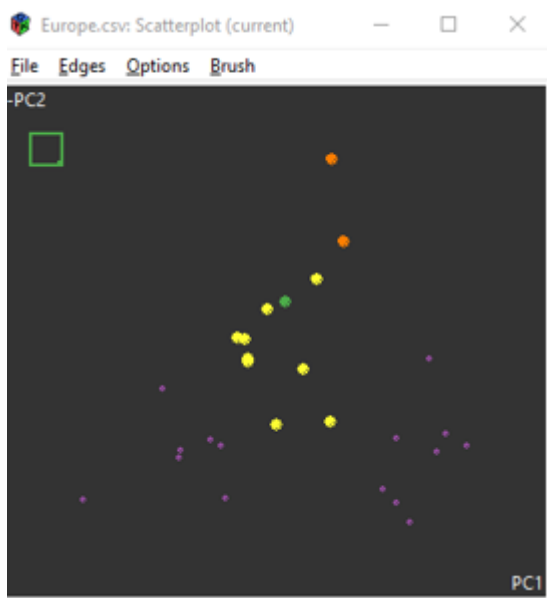
Example 2 ...



What is it about?

Example 2 ...

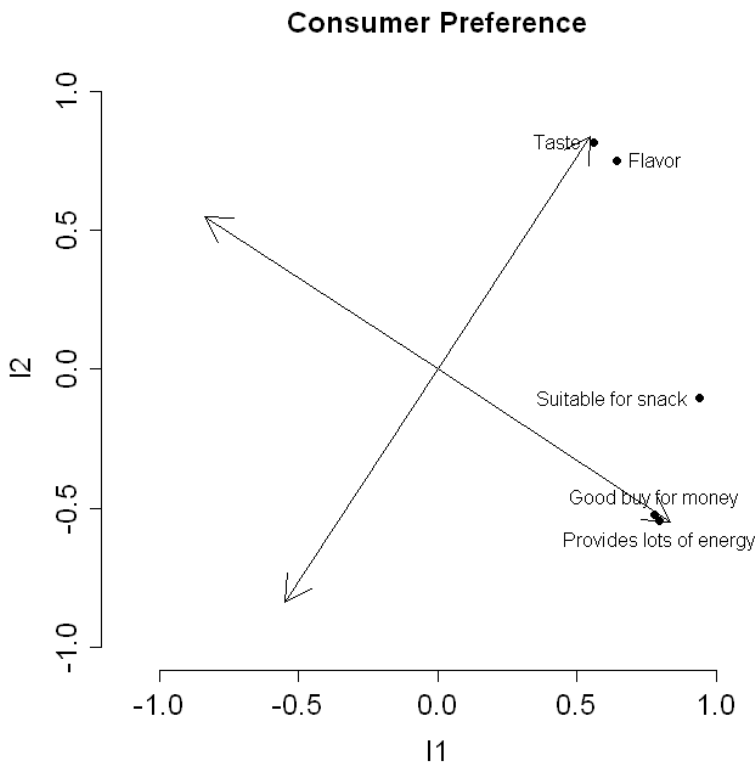
... or using Ggobi



What is it about?

Example 3: Factor Analysis

Consumer Preference (J&W, example 9.9, p. 508)



$$\mathbf{R} = \begin{pmatrix} 1 & 0.02 & 0.96 & 0.42 & 0.01 \\ 0.02 & 1 & 0.13 & 0.71 & 0.85 \\ 0.96 & 0.13 & 1 & 0.50 & 0.11 \\ 0.42 & 0.71 & 0.50 & 1 & 0.79 \\ 0.01 & 0.85 & 0.11 & 0.79 & 1 \end{pmatrix}$$

Taste

Good buy for money

Flavor

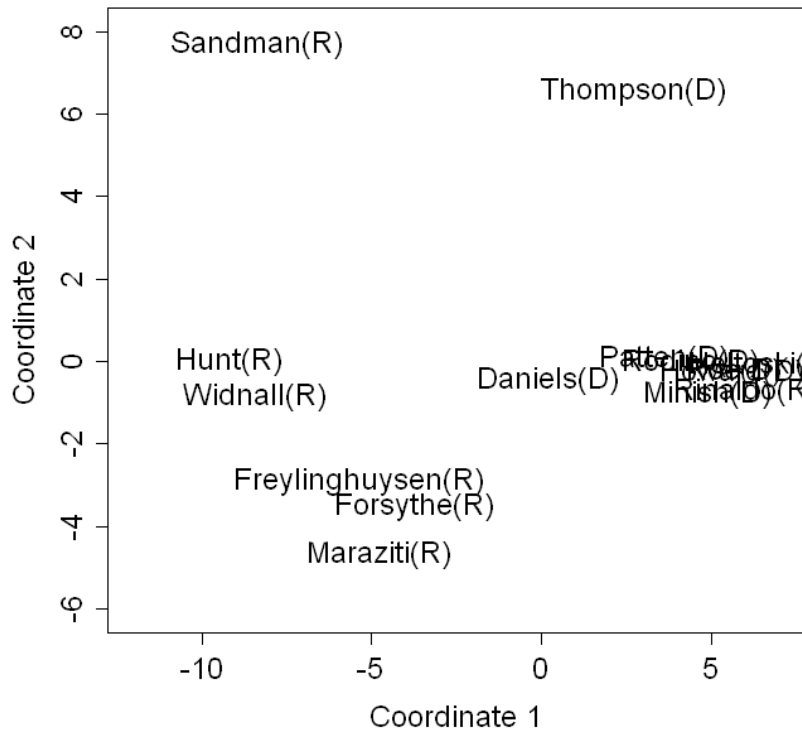
Suitable for snack

Provides lots of energy

What is it about?

Example 4: Distances

Voting results for 15 congressmen from New Jersey
(example from R package HSAUR)



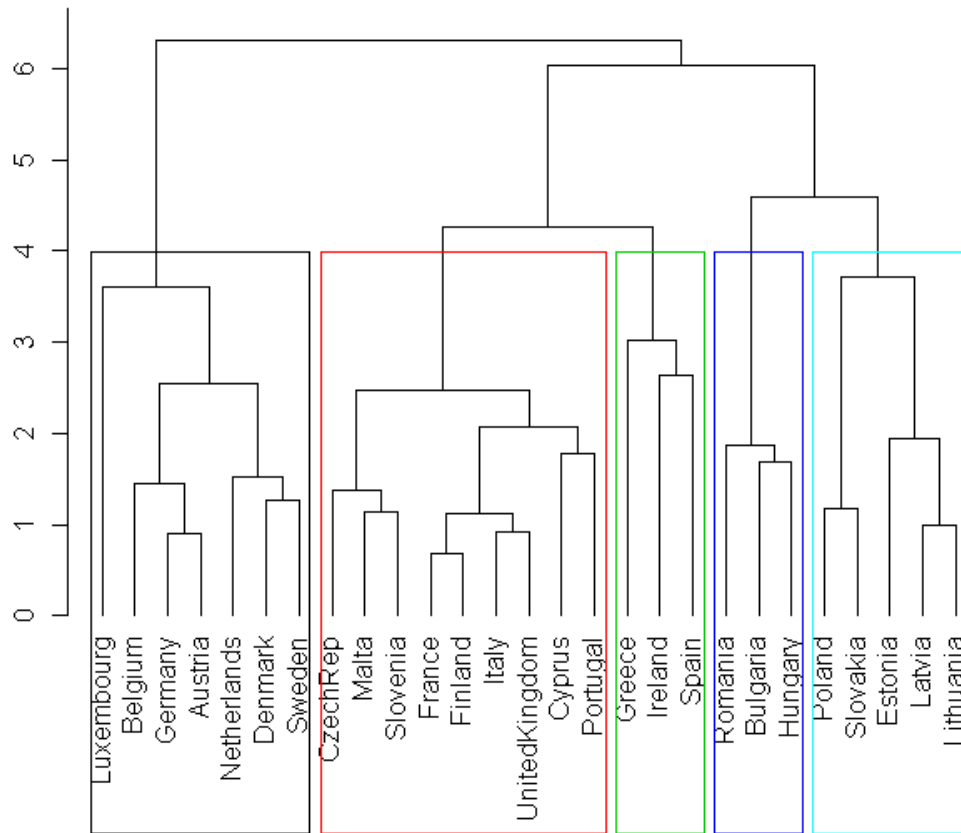
Extraction from the distance matrix ...

	Hunt(R)	Sandman(R)	Howard(D)
Hunt(R)	0	8	15
Sandman(R)	8	0	17
Howard(D)	15	17	0

What is it about?

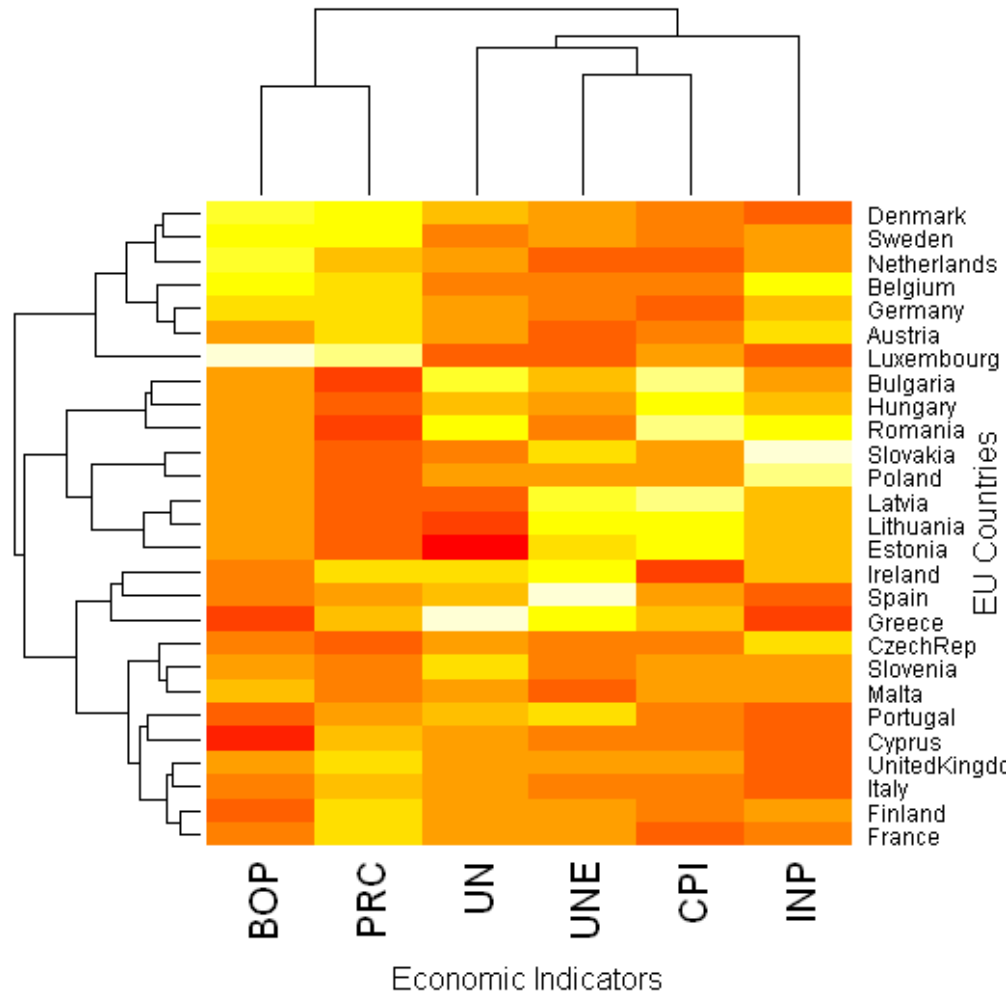
Example 5: Grouping

Cluster Dendrogram



What is it about?

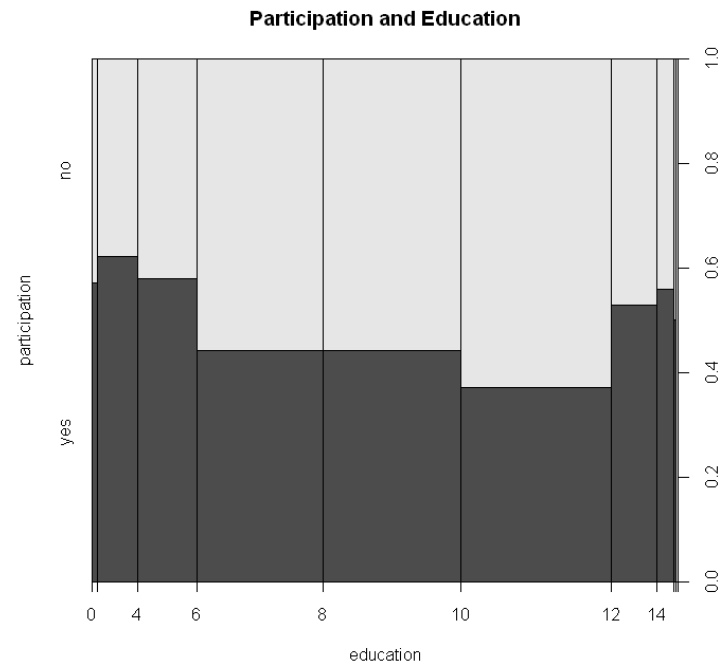
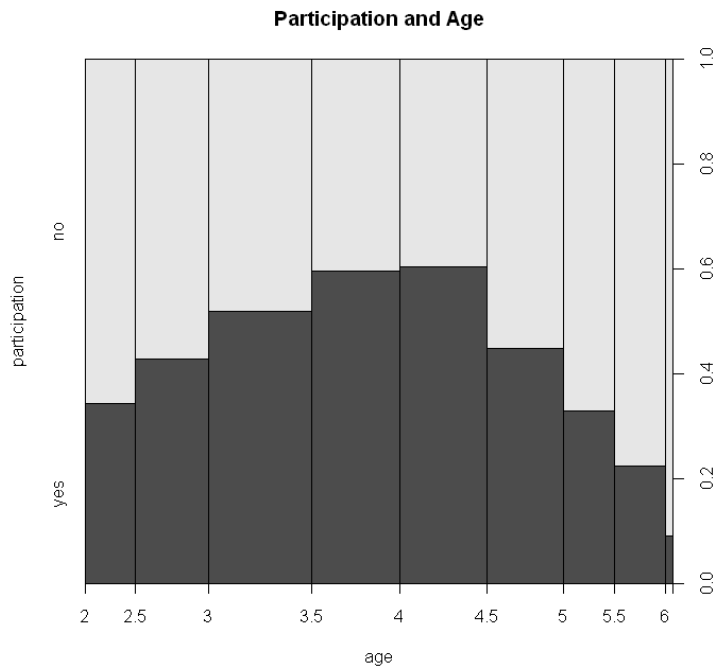
Example 5 ...



What is it about?

Example 6: Classification

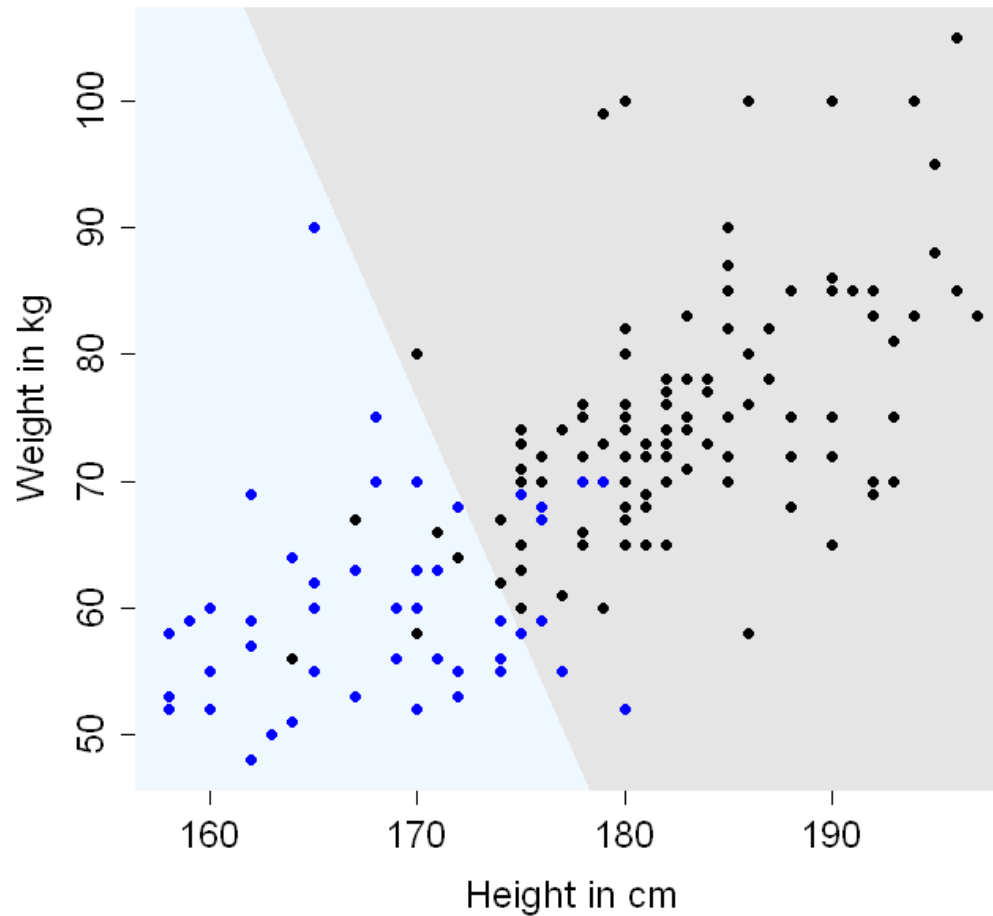
Labor Market Participation of Married Women in Switzerland (1981)
(example from R package AER)



What is it about?

Example 7: Discrimination

Heights and weights of students



What is it about?

Course Outline

Generally: Chapters 1-4, 8, 9, 11, 12 from Johnson and Wichern (J&W)

Timetable and Contents

- Lecture 1: Introduction
 - Lecture 2: Matrix Algebra (part 1)
 - Lecture 3: Matrix Algebra (part 2)
 - Lecture 4: Multivariate Samples
 - Lecture 5: Principal Component Analysis (part 1)
 - Lecture 6: Principal Component Analysis (part 2)
 - Lecture 7: Biplots
 - Lecture 8: Factor Analysis
- } **Week 1 (Ass. 0+1)**
- Week 2 (Ass. 2)
- Week 3 (Ass. 3)
- } **Week 4 (Ass. 4+5)**
- Week 5 (Ass. 6)
- Week 6 (Ass. 7)

What is it about?

- Lecture 9: Multidimensional Scaling Week 7 (Ass. 8)
- Lecture 10: Cluster Analysis Week 8 (Ass. 9)
- Lecture 11: Linear Discriminant Analysis Week 9 (Ass. 10)
- Lecture 12: Binary Response Models Week 10 (Ass. 11)
- Lecture 13: Statistical Methods for Data Science Week 11 (Ass. 12*)

Note: This is just a plan! Topics may be skipped; order may be changed; lecture topics may overlap

* interactive lecture, no homework

What is it about?

Main Objectives

... at the end of the semester you

- know and (hopefully) understand most common methods for analyzing multivariate data and their theoretical background
- can proficiently use **R** when using multivariate techniques: data import, constructing graphics, inference, model diagnosis and assessment
- have experienced the possibilities and limitations of multivariate methods on the basis of real data examples



Generally: This is an introductory and applied course. Modern multivariate techniques based on machine learning algorithms will hardly be covered.