STA 437/2005:
Methods for Multivariate Data

Week 4: Principal Component Analysis

Piotr Zwiernik

University of Toronto

PCA teaser

Example 1: Decathlon

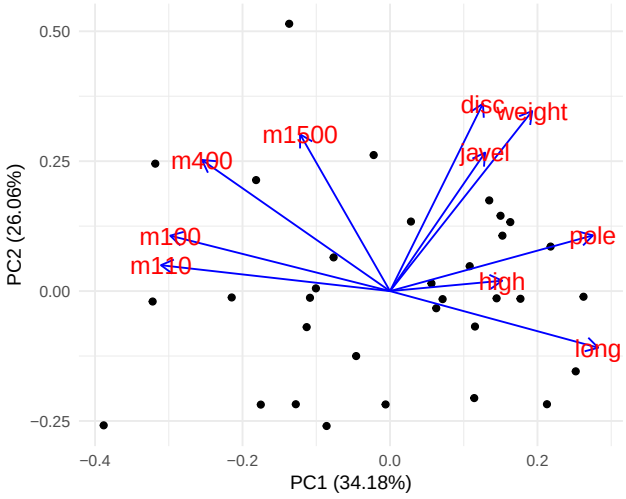
The columns are a subset of gene expression measurements, they correspond to 156 genes that show differential expression between cell types:

```
> data("olympic", package = "ade4")
> athletes = setNames(olympic$tab,
+   c("m100", "long", "weight", "high", "m400", "m110", "disc", "pole", "javel", "m1500"))
> head(athletes)
```

```
   m100 long weight high  m400  m110  disc pole javel  m1500
1 11.25 7.43  15.48 2.27 48.90 15.13 49.28  4.7 61.32 268.95
2 10.87 7.45  14.97 1.97 47.71 14.46 44.36  5.1 61.76 273.02
3 11.18 7.44  14.20 1.97 48.29 14.81 43.66  5.2 64.16 263.20
4 10.62 7.38  15.02 2.03 49.06 14.72 44.80  4.9 64.04 285.11
5 11.02 7.43  12.92 1.97 47.44 14.40 41.20  5.2 57.46 256.64
6 10.83 7.72  13.58 2.12 48.34 14.18 43.06  4.9 52.18 274.07
```

PCA Biplot for Decathlon data

PCA Biplot of Olympic Athletes



Example 3: Pottery

Chemical analysis data on Romano-British pottery made in three different regions (kiln 1, kilns 2-3, and kilns 4-5):

```
> data("pottery", package = "HSAUR2")
> head(pottery)
```

	Al2O3	Fe2O3	MgO	CaO	Na2O	K2O	TiO2	MnO	BaO	kiln
1	18.8	9.52	2.00	0.79	0.40	3.20	1.01	0.077	0.015	1
2	16.9	7.33	1.65	0.84	0.40	3.05	0.99	0.067	0.018	1
3	18.2	7.64	1.82	0.77	0.40	3.07	0.98	0.087	0.014	1
4	16.9	7.29	1.56	0.76	0.40	3.05	1.00	0.063	0.019	1
5	17.8	7.24	1.83	0.92	0.43	3.12	0.93	0.061	0.019	1
6	18.8	7.45	2.06	0.87	0.25	3.26	0.98	0.072	0.017	1

Question: Do the chemical profiles of each pot suggest different types of pots and if any such types are related to kiln or region.

PCA Biplot for Pottery data

PCA Biplot of Olympic Athletes