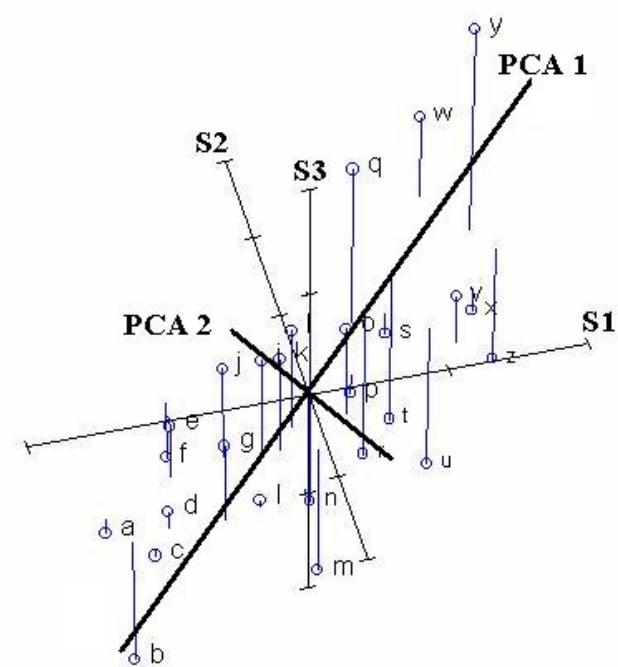
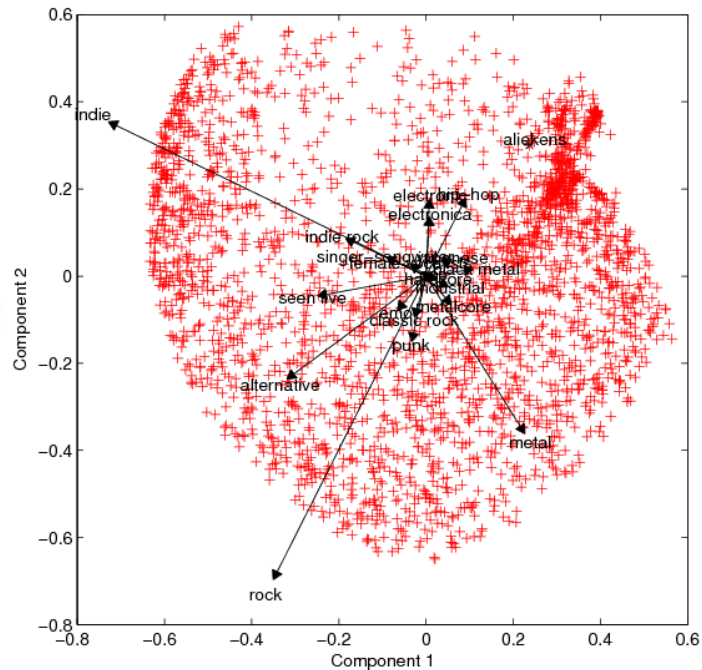
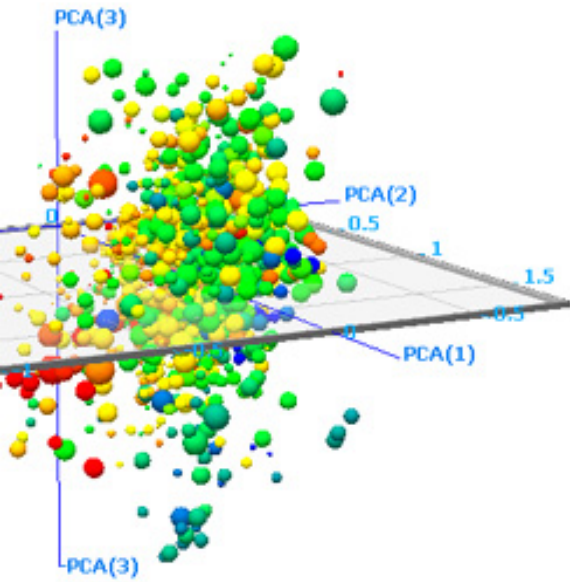


Principle Component Analysis



What is PCA?

- **method of simplifying multivariate data**
- **reduce number of variables**
- **detect structure**
- **start with variables that correlate to varying degrees**
- **generate linear combinations that capture most variance**
- **the best ones are then factors which may better describe the data**

What is PCA when compared to EFA

- **Similarities:**

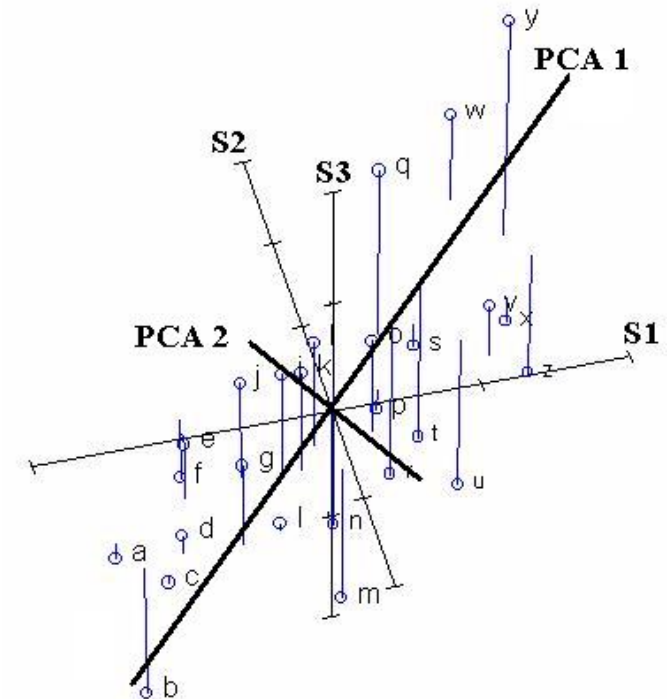
- linear relation between observed variables
- normal distribution for each variable
- measurement scale is interval or ratio level
- both variable reduction techniques
- outcome of PCA and FA should be similar
- PCA mainly for reduction
- FA mainly for structure

- **Differences:**

- PCA account for maximal amount of variance / EFA account for common variance
- PCA minimizes perpendicular distance to the component axes / EFA estimates factors which influence responses on observed variables
- In PCA component scores are a linear combination of the observed variables weighted by eigenvectors / in EFA Observed variables are linear combinations of the underlying and unique factors

PCA as multivariate extension of regression

- if you have 2 variables you can plot a scatterplot
- if there is a correlation you can fit a line
- line represents the what's common (covariance)
- how do you get what's common out of multiple variables?



Some examples of calculating principal components

call	sqrt(eigenvalues)	eigenvalues	eigenvectors	rotated data
prcomp(ph,scale.=T)	sdev		rotation	scores
princomp(scale(ph))	sdev		loadings	x
eigen(cor(ph))		values	vectors	scale(ph) %*% eig\$vector
svd(cor(ph))		d	u	scale(ph) %*% sv\$vector

- **Anxiety Related Behaviours Questionnaire**

- **1 gbh201/2** Often makes comments critical of him/herself
- **2 gbh441/2** Tends to blame him/herself
- **3 gbh401/2** Has low self-confidence
- **4 gbh291/2** Many worries or often seems worried
- **5 gbh621/2** Anxious that bad things will happen
- **6 gbh591/2** Asks for reassurance that s/he is OK
- **7 gbh351/2** Often unhappy, down-hearted or tearful
- **8 gbh151/2** Doesn't enjoy him/herself
- **9 gbh561/2** Seems keyed up, on edge or tense
- **10 gbh061/2** Insists on doing something over and over so that it interferes with day to day life
- **11 gbh261/2** Complains or whines a lot
- **12 gsickf1/2** Does either twin complain of stomach-aches, headaches or sickness and if so, how often?
- **13 gbh341/2** Is extremely distressed by changes to routine or familiar arrangements
- **14 ghtics** Does either twin have repeated tics or twitches (e.g. forceful eye blinking or a rapid head jerk)?
- **15 gbh281/2** Is afraid of small closed spaces, heights, water or the dark
- **16 gbh131/2** Many fears or easily scared
- **17 gbh091/2** Is afraid of animals or insects (like dogs, spiders, snakes, or insects)
- **18 gbh121/2** Strongly refuses or resists sleeping alone
- **19 gbh031/2** Is afraid of medical procedures such as going to see the doctor or dentist
- **20 gbh241/2** Tends to check that some things are done exactly 'right'
- **21 gbh431/2** Fussy or over particular
- **22 gbh511/2** Tends to be shy or timid
- **23 gbh481/2** Is afraid in social situations
- **24 gbh551/2** Takes a long time to warm to strangers
- **25 gbh601/2** Nervous or clingy in new situations, easily loses confidence in new situations

Run PCA

- `arb1 <- na.omit(arb)`
- `pca <- prcomp(arb1, scale=T)`

Interpretation: eigenvalues

- **remember it's the variance**
- **how many components do you pay attention to?**
- **2 popular rules-of-thumb**
- **variance > 1**
- **scree inflection**

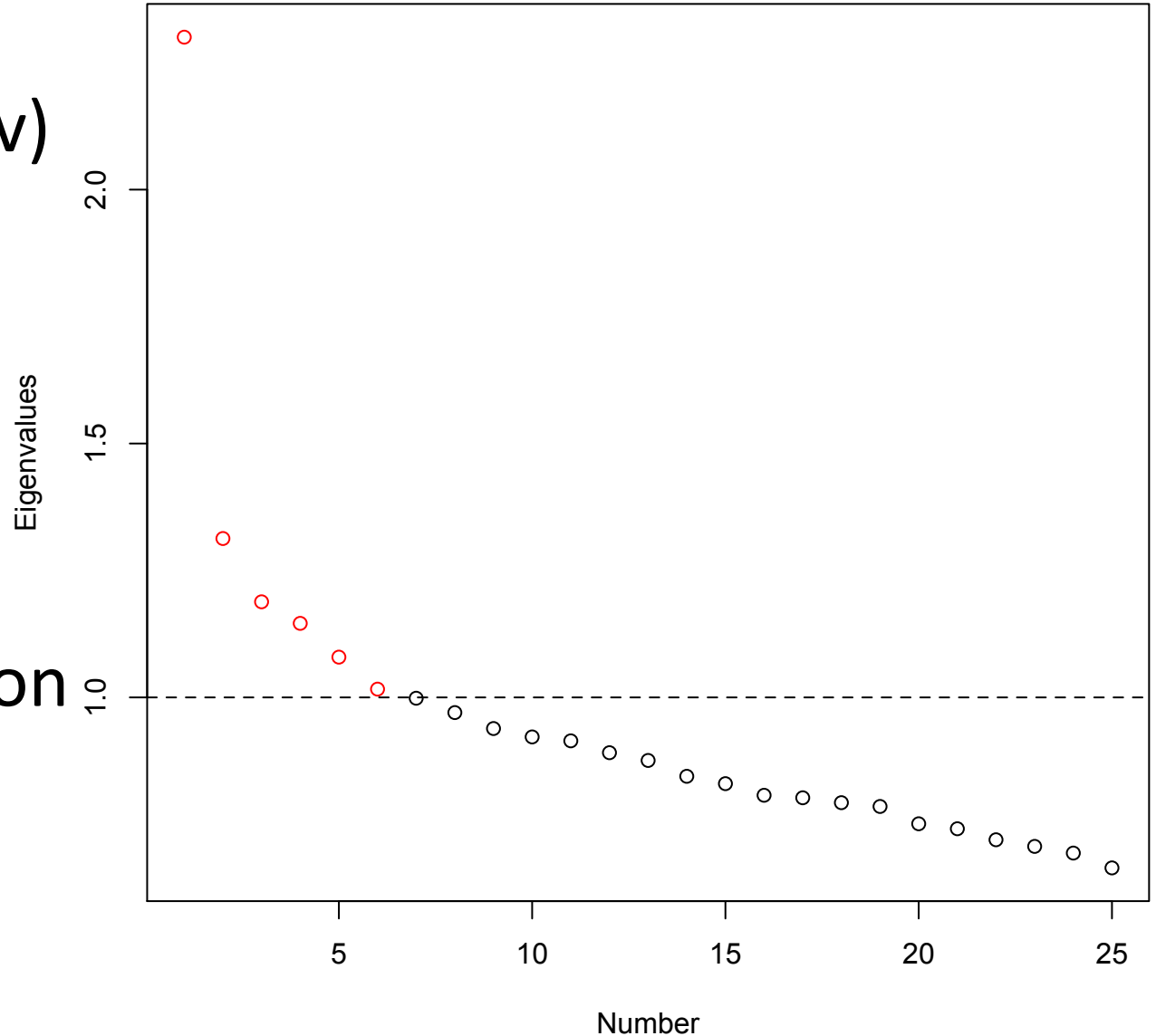
'Inflection'



Scree-plot PCA

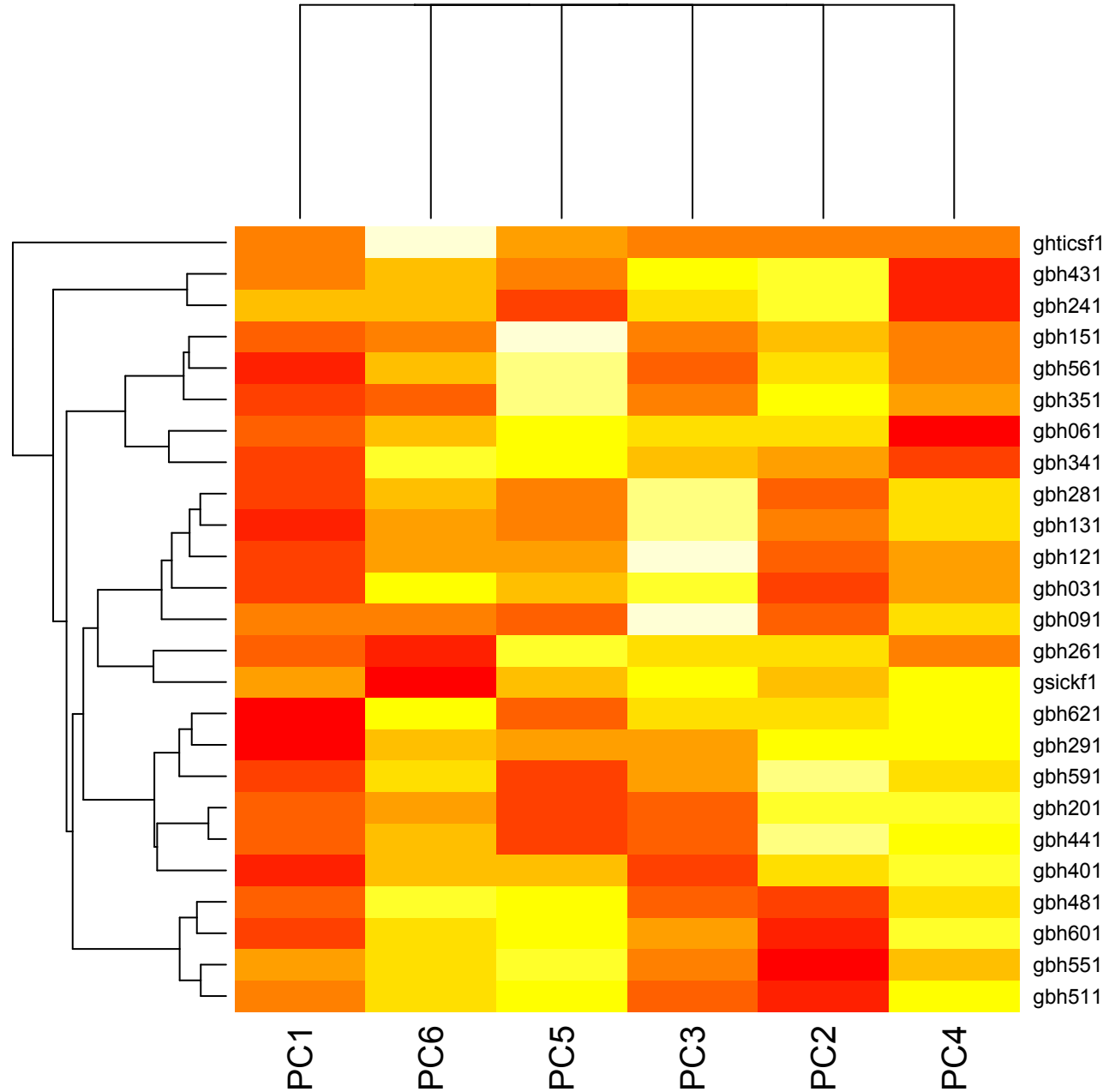
- `plot(pca$sdev)`
- `abline(h=1)`

- Kaiser Criterion



Plotting

- `heatmap(pca
$rot[,2:6])`



Variance explained

- `summary(pca)`

- Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	
Standard deviation	2.3002	1.31290	1.18827	1.14581	1.0793	1.01631	
Proportion of Variance	0.2116	0.06895	0.05648	0.05252	0.0466	0.04132	
Cumulative Proportion	0.2116	0.28059	0.33707	0.38958	0.4362	0.47750	
	PC7	PC8	PC9	PC10	PC11	PC12	
Standard deviation	0.99837	0.97012	0.93869	0.92217	0.91445	0.89110	
Proportion of Variance	0.03987	0.03765	0.03525	0.03402	0.03345	0.03176	
Cumulative Proportion	0.51737	0.55501	0.59026	0.62427			

Rotation

- We have a reason to believe that the underlying factors could be related, thus use 'oblique' instead of 'orthogonal' rotation
- `promax(arb$rot[,1:6])$load`

- `promax(pca$rotation[,c('PC1','PC2', 'PC3', 'PC4', 'PC5', 'PC6')])$load`

- Loadings:

	PC1	PC2	PC3	PC4	PC5	PC6	
• gbh201	-0.545						Makes comments critical of him/herself
• gbh441	-0.521						Tends to blame him/herself
• gbh401	-0.314	-0.146		0.139	0.122		Has low self-confidence
• gbh291	-0.266		0.121		0.122		Many worries
• gbh621	-0.280		0.226			0.113	Anxious bad things will happen
• gbh591	-0.320			-0.171			Asks for reassurance
• gbh351				0.101	0.472	-0.141	Often unhappy, tearful
• gbh151			-0.103	0.107	0.480		Doesn't enjoy him/herself
• gbh561					0.394		Seems keyed up, tense
• gbh061	0.146			-0.262	0.365	0.148	Insists on doing something over and over
• gbh261					0.381	-0.334	Complains or whines
• gsickf1			0.110			-0.535	Complains of stomach or headaches
• gbh341				-0.127	0.214	0.218	Distressed by changes to routine
• ghticsf1						0.660	Tick or twitches
• gbh281			0.532		-0.103		Afraid of closed spaces, heights, dark
• gbh131			0.464				Many fears
• gbh091			0.349		-0.152	-0.191	Afraid of animals or insects
• gbh121	0.118		0.430				Refuses to sleep alone
• gbh031			0.284				Is afraid of medical procedures
• gbh241				-0.695	-0.130		Everything has to be exactly right
• gbh431				-0.590			Fussy
• gbh511		-0.541					Shy or timid
• gbh481		-0.435				0.102	Afraid of social situations
• gbh551	0.136	-0.508					Takes long to warm up to strangers
• gbh601		-0.443					Nervous or clingy in new situations

Exercise 1

- Dataset showing incidence of crimes and percentage of urban population in 1973 America

library(stats)

?USArrests

- What is the relationship between urban population and the three sorts of crime?

Exercise 2

- Genom-wide mouse expression data
 - `load(url('http://sgdp.iop.kcl.ac.uk/oleo/stress.RData'))`
 - `pca1 <- prcomp(yann.rma@exprs)`
 - `plot(pca$rotation[,1:2])`
-
- Use plotting character and color arguments to discover 'hidden' information about the data