

CS 360: Machine Learning

Prof. Sara Mathieson

Fall 2019



Admin

- In lab today: **project meetings** with all groups
- No office hours tomorrow, but feel free to **make an appointment**

Outline for December 5

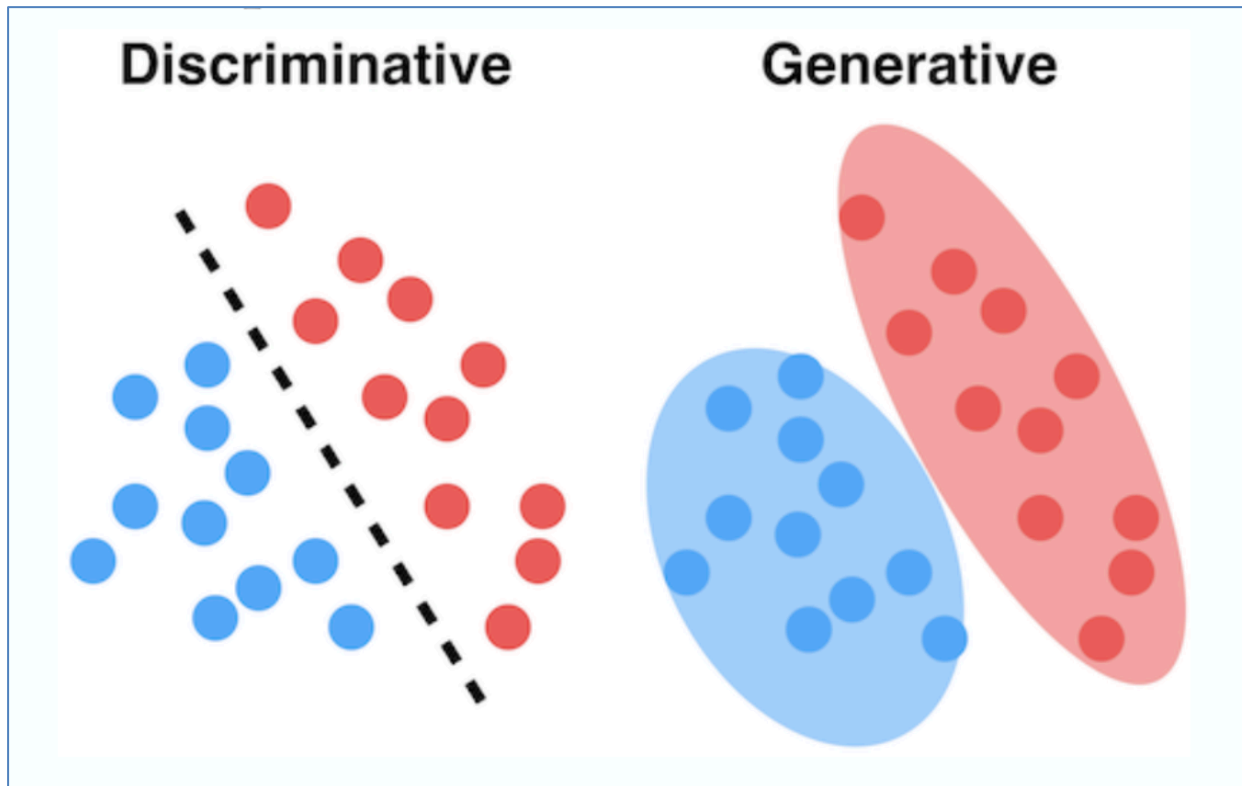
- Finish: Gaussian Mixture Models (GMM)
- Hierarchical clustering algorithms
- Dimensionality reduction
- Principal Component Analysis (PCA)

Outline for December 5

- **Finish: Gaussian Mixture Models (GMM)**
- Hierarchical clustering algorithms
- Dimensionality reduction
- Principal Component Analysis (PCA)

Discriminative vs. Generative

- Discriminative: finds a decision boundary
 - Logistic regression, K-means
- Generative: estimates probability distributions
 - Naïve Bayes, Gaussian Mixture Models



GMM

EM algorithm

Goal: model cluster size
• estimate model params



• π_k = prob. of cluster k
("size")

begin: $\pi_k = \frac{1}{K} \quad \forall k=1 \dots K$

• $\vec{\mu}_k$ = mean of cluster k

begin: choose random point.

• σ_k^2 = variance
"spread"
of cluster
 k

begin: sample
variance of
all points
closest to

$\vec{\mu}_k$

E-step "Soft assignment"

w_{ik} = prob. that \vec{x}_i came from cluster k

$$w_{ik} = P(k | \vec{x}_i) = \frac{\overset{\text{prior}}{p(k)} p(\vec{x}_i | k)}{\underset{\text{posterior}}{p(\vec{x}_i)} \underset{\text{evidence (data)}}{p(\vec{x}_i)}}$$

$$w_{ik} = \frac{\pi_k \mathcal{N}(\vec{x}_i; \vec{\mu}_k, \sigma_k^2)}{\sum_{k'=1}^K \pi_{k'} \mathcal{N}(\vec{x}_i; \vec{\mu}_{k'}, \sigma_{k'}^2)}$$

Gaussian / normal dist.

normalization



$$W = \begin{bmatrix} \ell_1 & \ell_2 & \ell_3 & \vec{x}_1 \\ 0.5 & 0.3 & 0.2 & \end{bmatrix}_{n \times K}$$

K-means

$$W = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

weight on cluster 2

\sum all entries = n

M-step

update params

$$M_k = \sum_{i=1}^n w_{ik}$$

"# points assigned to cluster k"

$$\Rightarrow \textcircled{1} \pi_k = \frac{M_k}{n} \star$$

$$\textcircled{2} \vec{\mu}_k = \frac{1}{M_k} \sum_{i=1}^n w_{ik} \vec{x}_i$$

weighted average

Sum of all weights

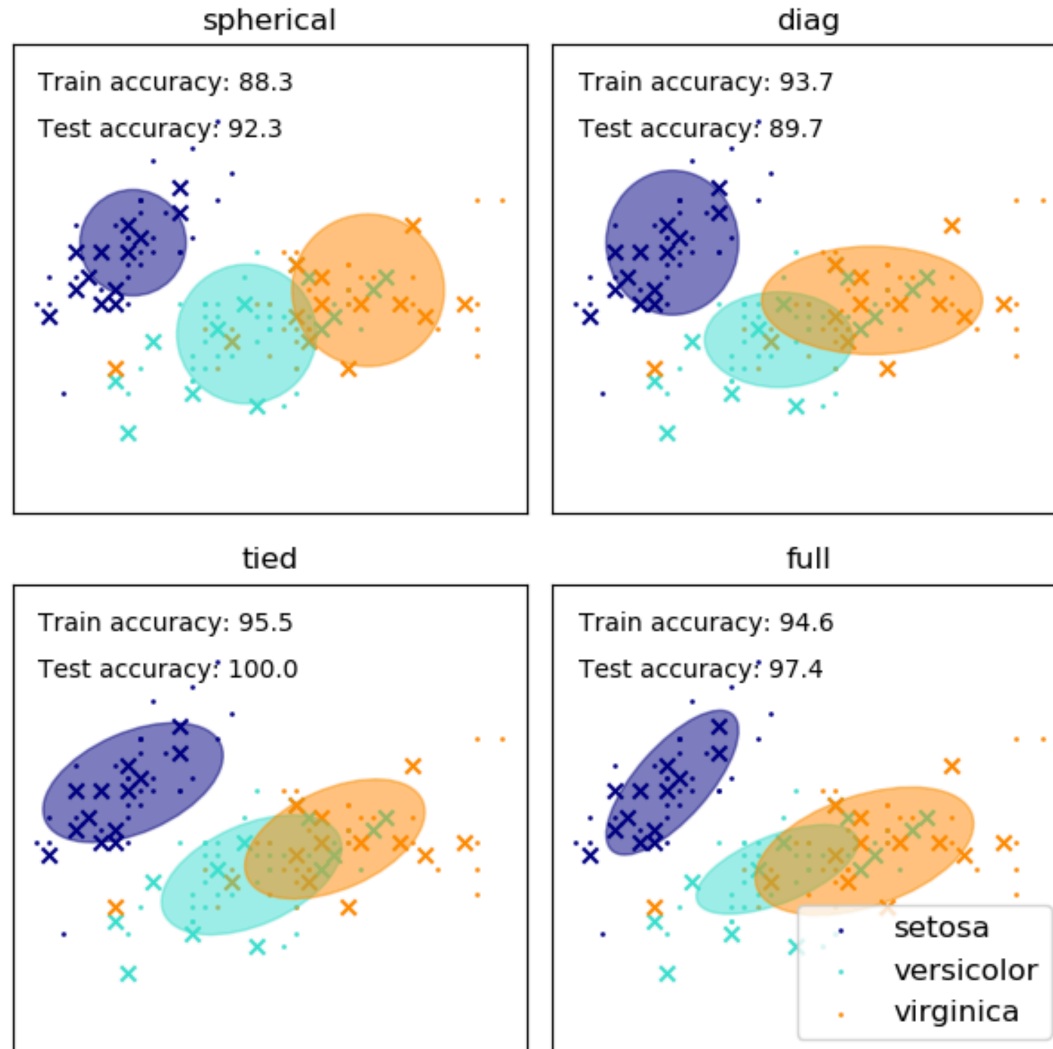
$$\frac{\frac{1}{3}x + \frac{6}{2}y}{\frac{1}{3} + \frac{6}{2}}$$

$$\textcircled{3} \sigma_k^2 = \text{weighted sample variance}$$

Generative Process

- ① choose a cluster k using $[\pi_1, \pi_2, \dots, \pi_k]$
- ② use $\vec{\mu}_k$ & σ_k^2 to draw data point \vec{x}

Example of GMMs with different covariance constraints on the Iris flower data

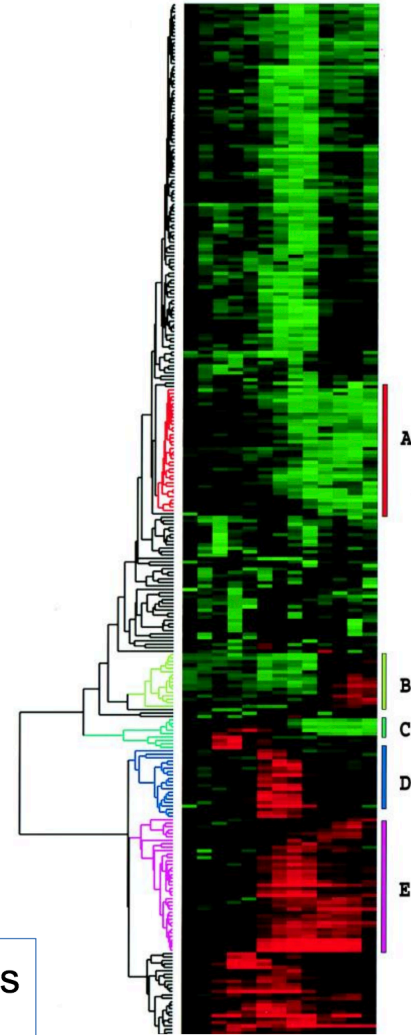


Outline for December 5

- Finish: Gaussian Mixture Models (GMM)
- Hierarchical clustering algorithms
- Dimensionality reduction
- Principal Component Analysis (PCA)

Applications of clustering in ML

- Cluster genes with similar expression patterns

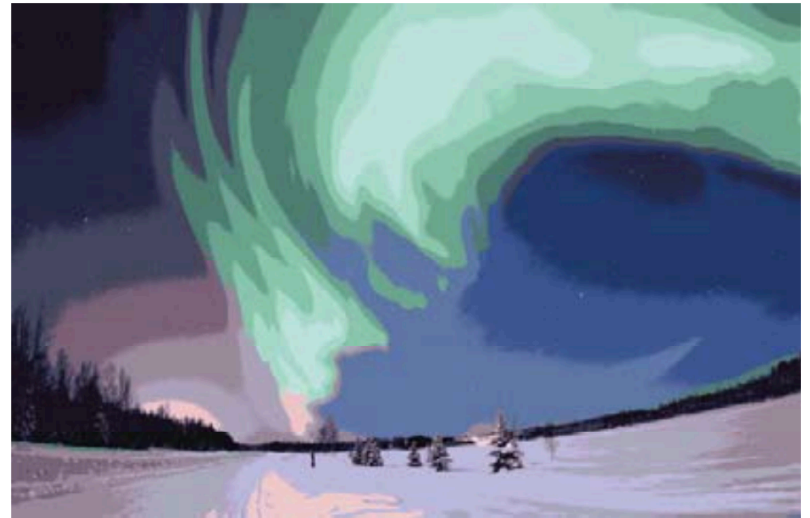


Cluster analysis and display of genome-wide expression patterns

[Michael B. Eisen](#),^{*} [Paul T. Spellman](#),^{*} [Patrick O. Brown](#),[†] and [David Botstein](#)^{*‡}

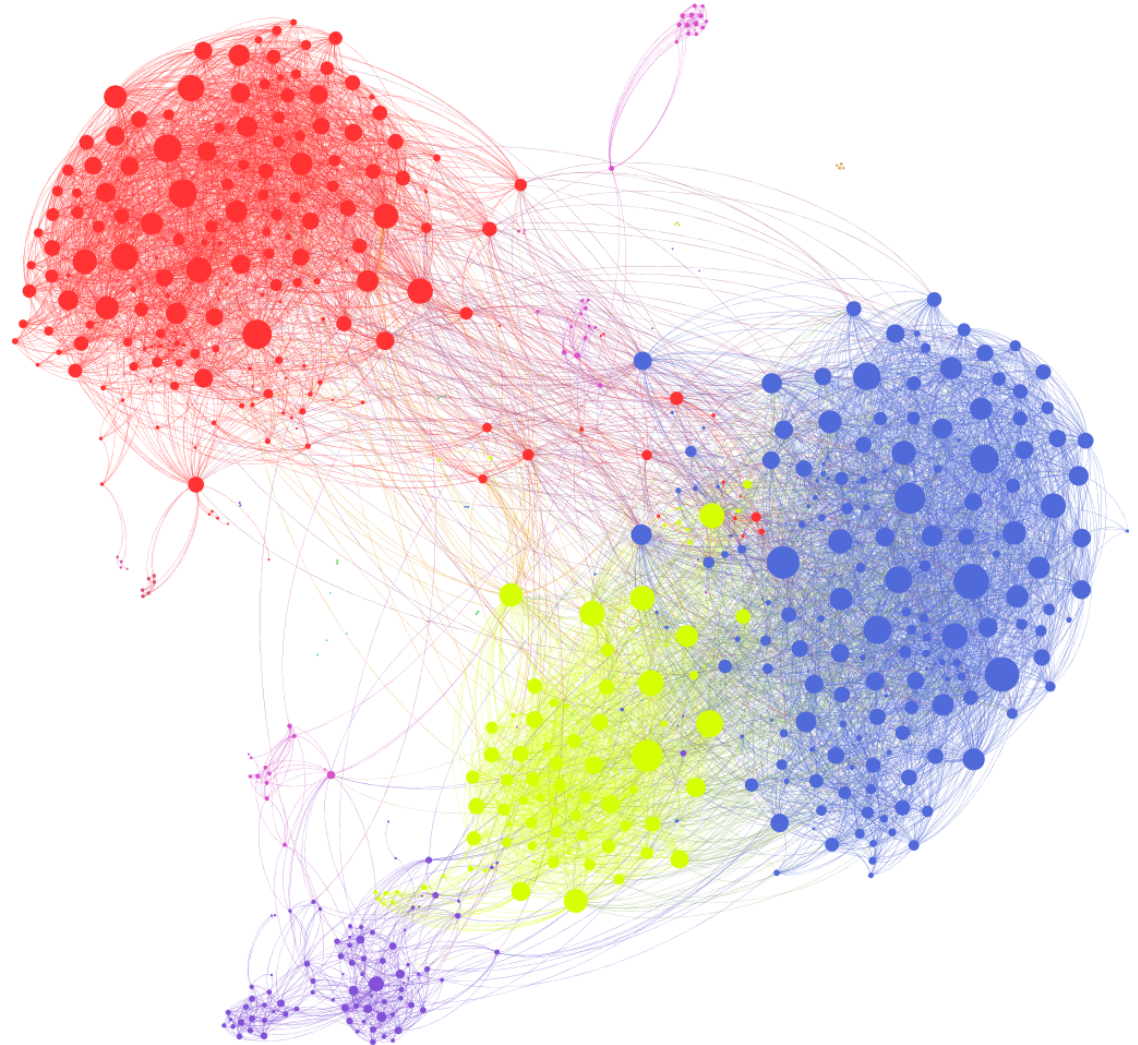
Applications of clustering in ML

- Image segmentation: cluster similar regions of an image



Applications of clustering in ML

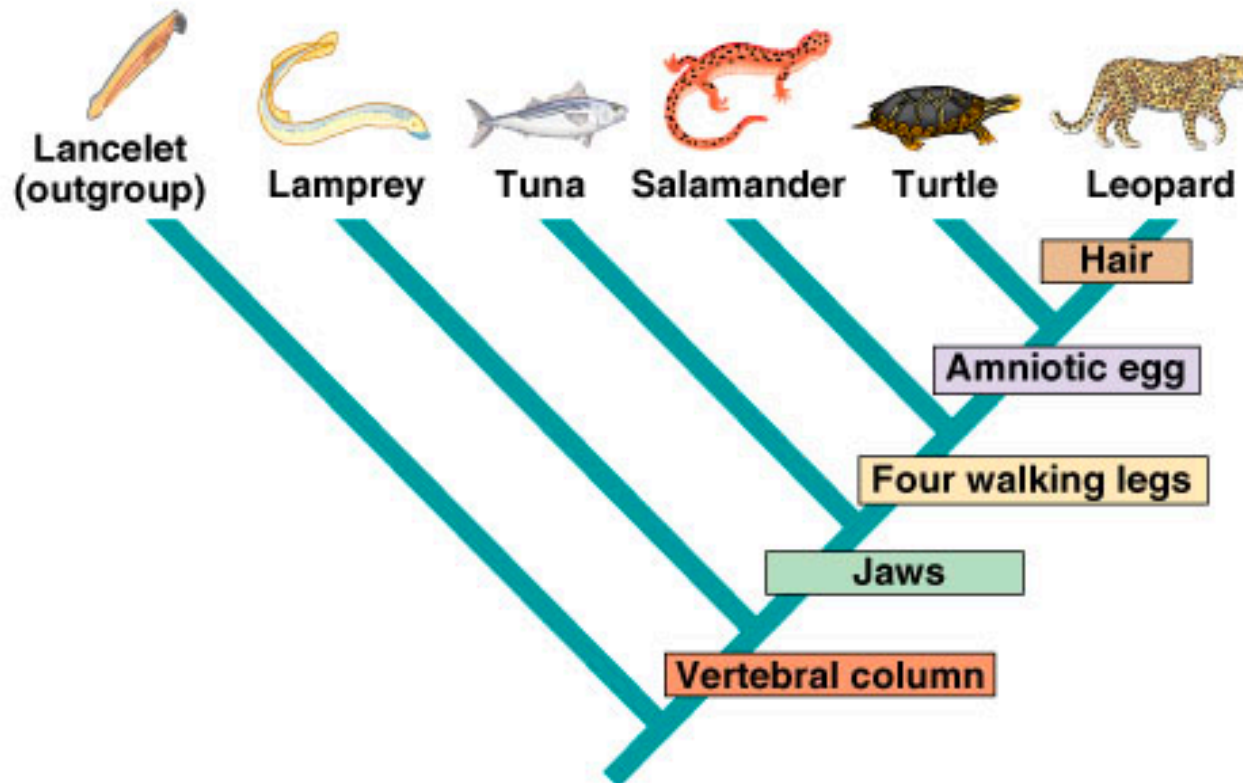
- Clustering in social graphs



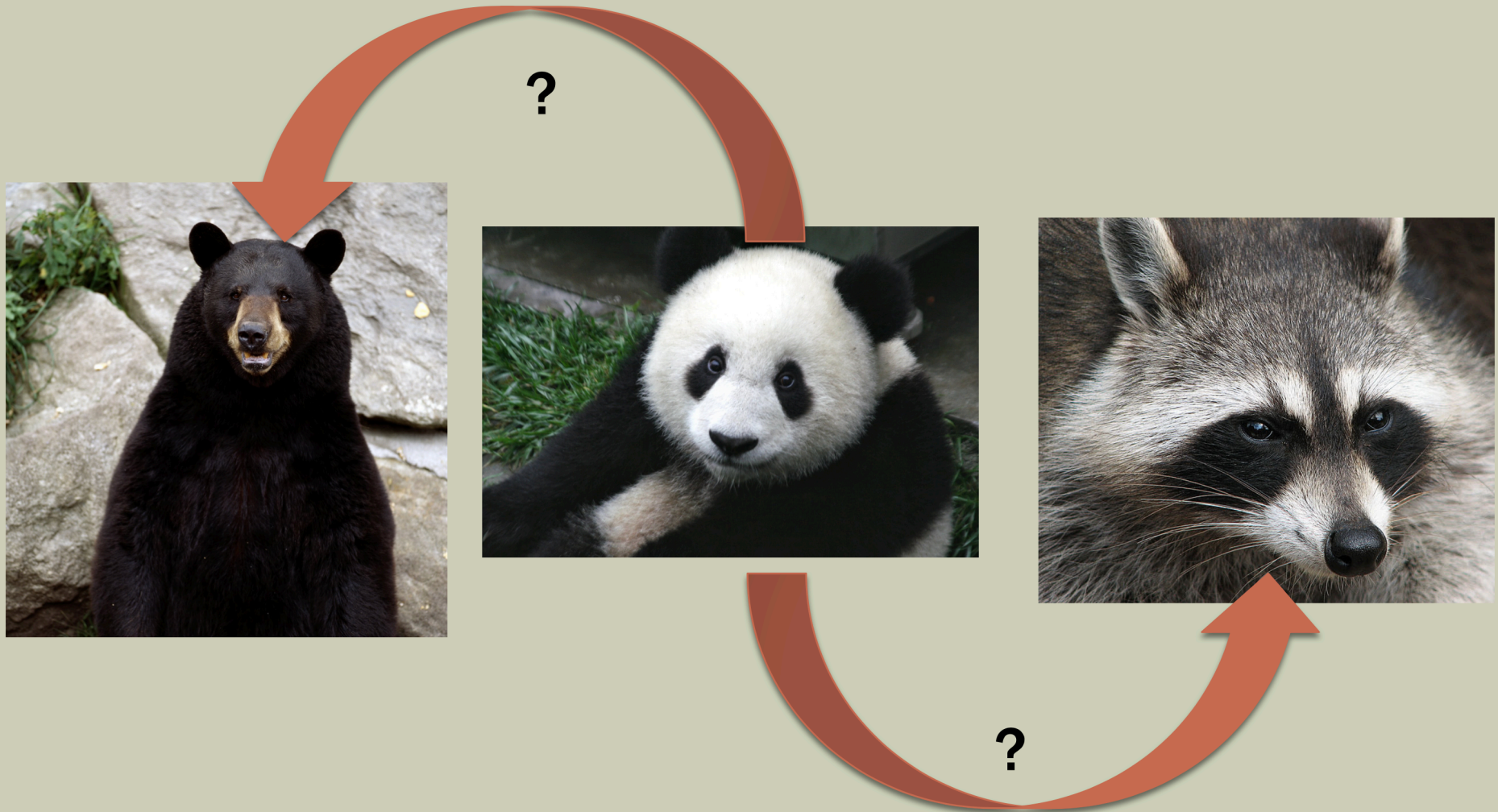
Two main types of clustering

- Flat/Partitional:
 - K-means
 - Gaussian mixture models
- Hierarchical:
 - Agglomerative: bottom-up
 - Divisive: top-down
 - Examples: UPGMA and Neighbor Joining

Hierarchical clustering example: trees



Are pandas more closely related to bears or raccoons?



UPGMA and Neighbor Joining

- Start with a dissimilarity map between examples (symmetric matrix)
- Say our examples are: A,B,C,D,E

δ	A	B	C	D	E
A	0	1	3	6	6
B		0	2	5	5
C			0	5	5
D				0	2
E					0

Hierarchical clustering example (UPGMA)

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

A D B F G C E

Hierarchical clustering example (UPGMA)

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	8.00	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

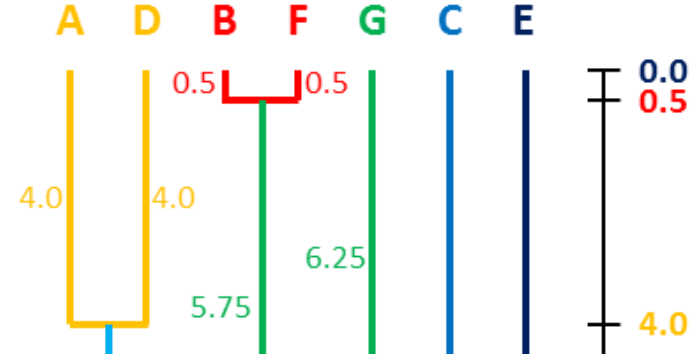


Hierarchical clustering example (UPGMA)

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	8.00	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	12.50	29.00	28.00



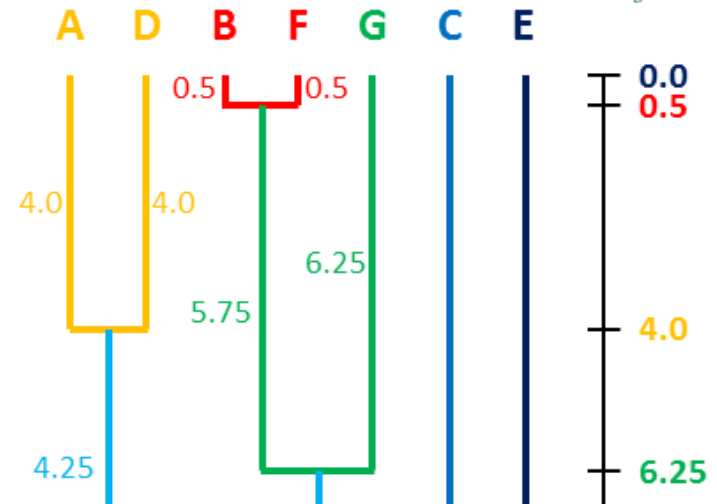
Hierarchical clustering example (UPGMA)

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	8.00	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	12.50	29.00	28.00

	AD	BFG	C
BFG	16.50		
C	26.50	30.67	
E	32.00	33.00	41.00



Hierarchical clustering example (UPGMA)

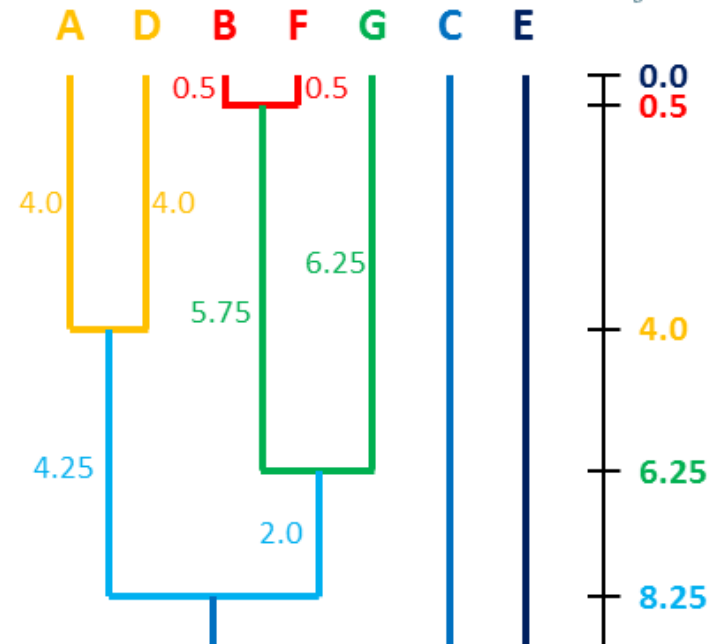
	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	8.00	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	12.50	29.00	28.00

	AD	BFG	C
BFG	16.50		
C	26.50	30.67	
E	32.00	33.00	41.00

	ADBFG	C
C	29.00	
E	32.60	41.00



Hierarchical clustering example (UPGMA)

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

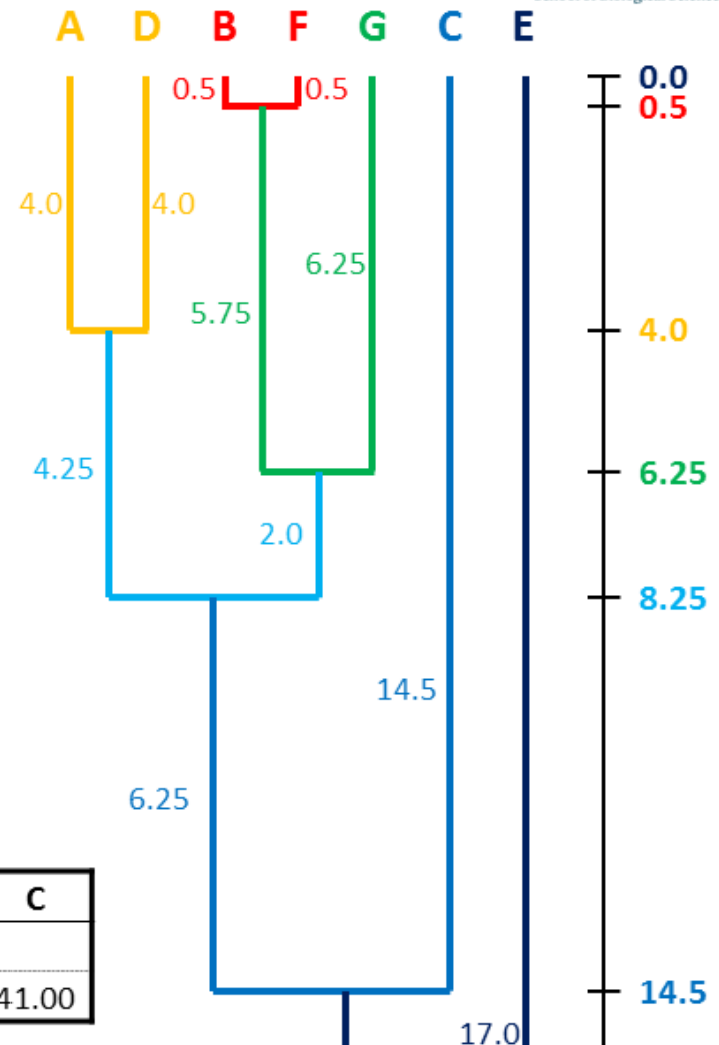
	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	8.00	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	12.50	29.00	28.00

	AD	BFG	C
BFG	16.50		
C	26.50	30.67	
E	32.00	33.00	41.00

	ADBFG	C
C	29.00	
E	32.60	41.00

	ADBFGC
E	34.00



Hierarchical clustering example (UPGMA)

	A	B	C	D	E	F	G
A							
B	19.00						
C	27.00	31.00					
D	8.00	18.00	26.00				
E	33.00	36.00	41.00	31.00			
F	18.00	1.00	32.00	17.00	35.00		
G	13.00	13.00	29.00	14.00	28.00	12.00	

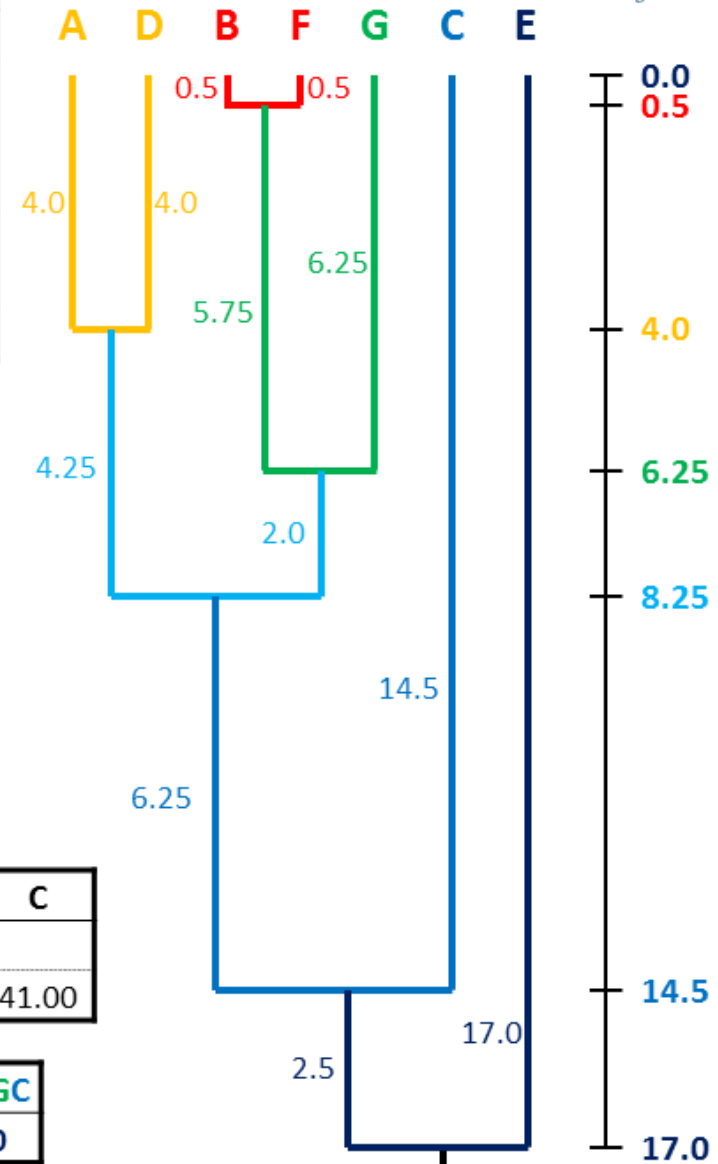
	A	BF	C	D	E
BF	18.50				
C	27.00	31.50			
D	8.00	17.50	26.00		
E	33.00	35.50	41.00	31.00	
G	13.00	12.50	29.00	14.00	28.00

	AD	BF	C	E
BF	18.00			
C	26.50	31.50		
E	32.00	35.50	41.00	
G	13.50	12.50	29.00	28.00

	AD	BFG	C
BFG	16.50		
C	26.50	30.67	
E	32.00	33.00	41.00

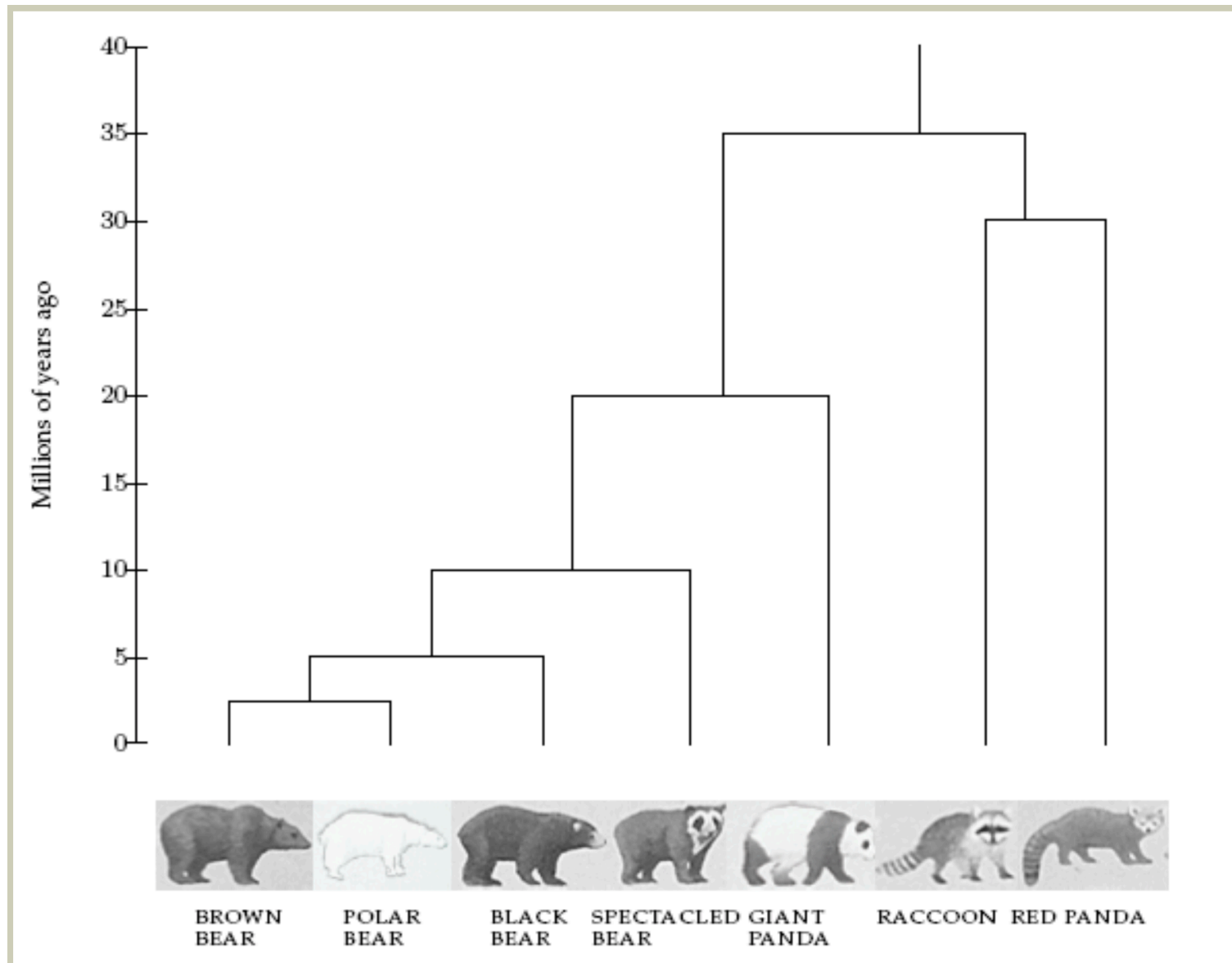
	ADBFG	C
C	29.00	
E	32.60	41.00

	ADBFGC
E	34.00



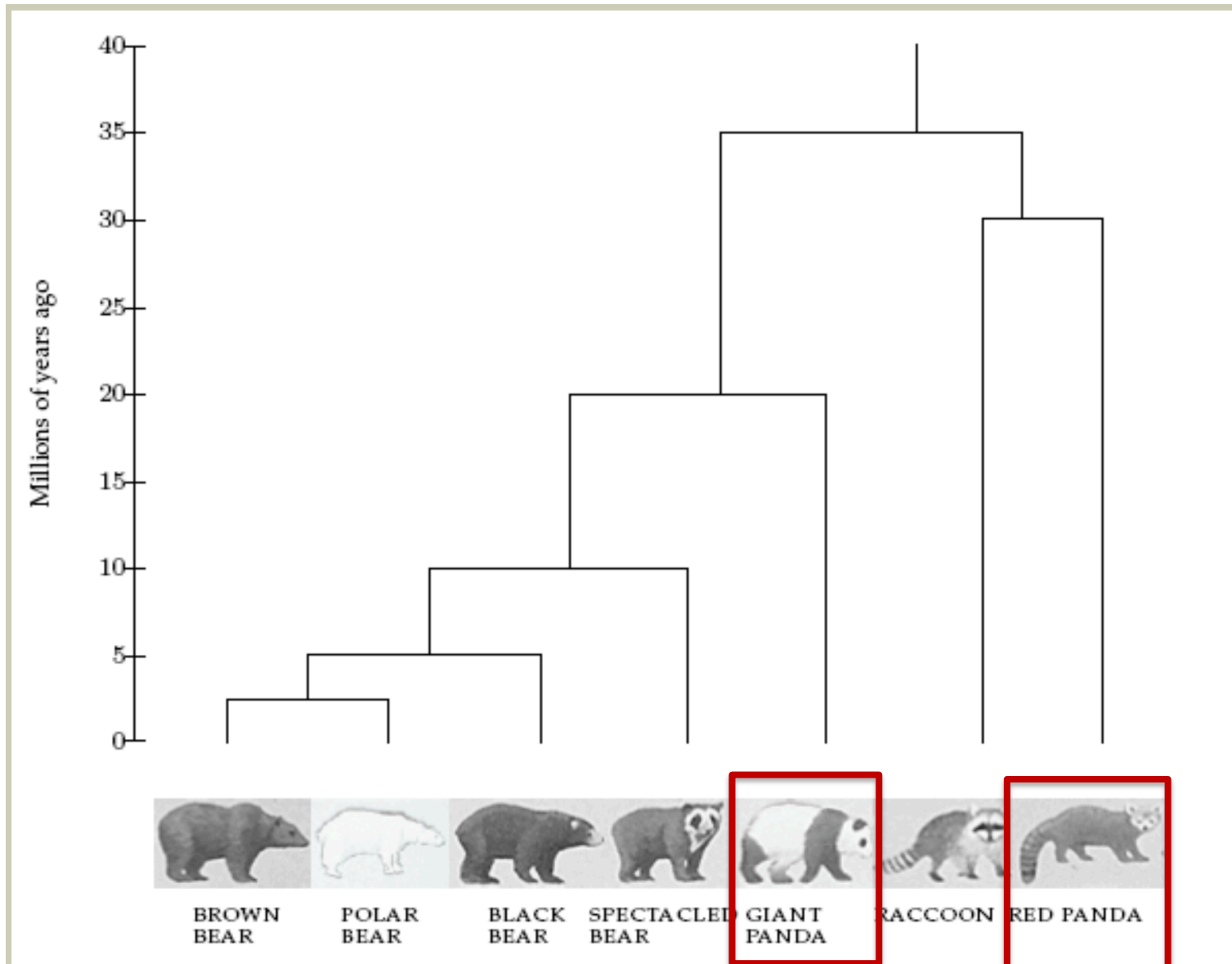
Back to the pandas....

Back to the pandas....



Credit:
Ameet
Soni

Back to the pandas....



*Credit:
Ameet
Soni*

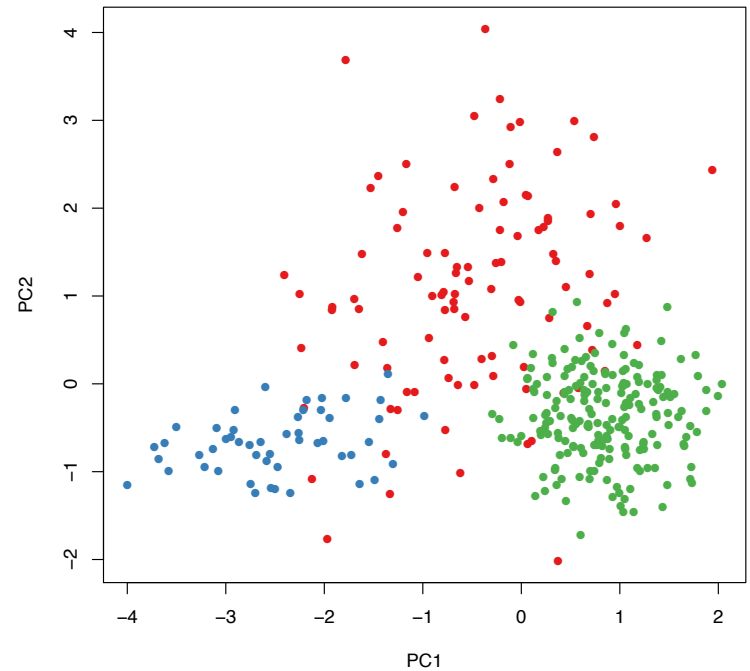
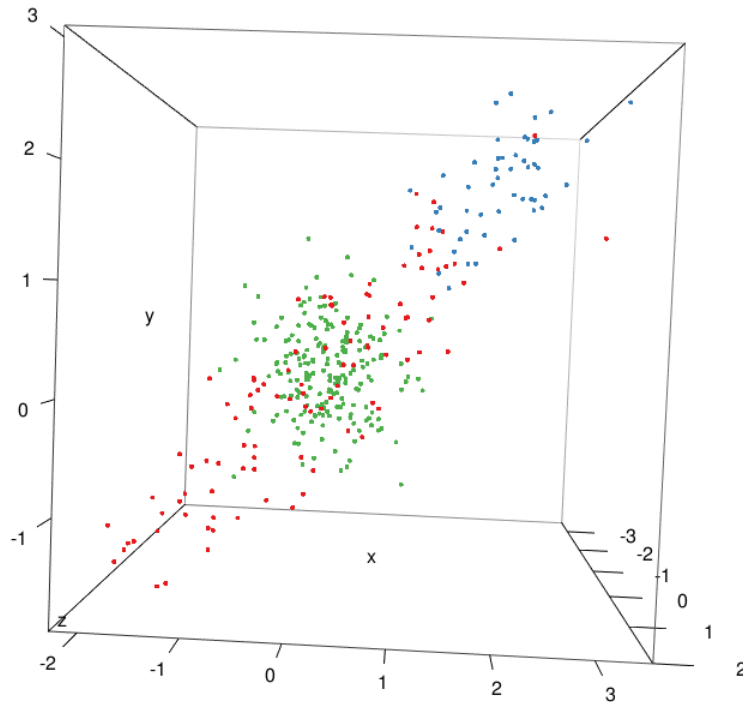
Outline for December 5

- Finish: Gaussian Mixture Models (GMM)
- Hierarchical clustering algorithms
- Dimensionality reduction
- Principal Component Analysis (PCA)

Principal Components Analysis (PCA)

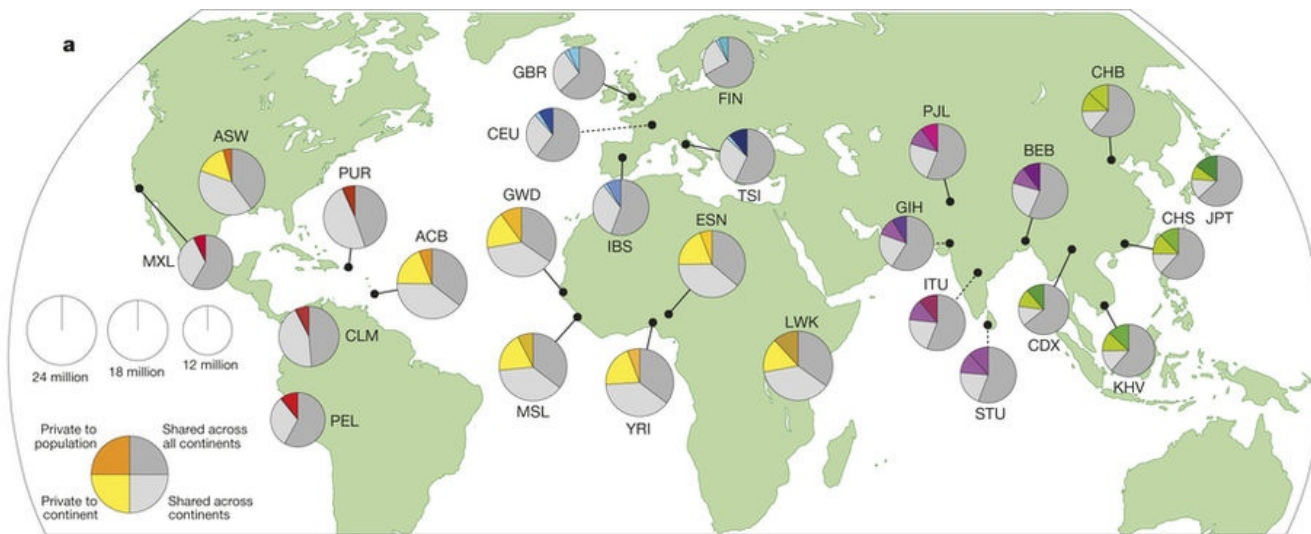
- Transforms p -dimensional data so that the new first dimension explains as much of the variation as possible, the new second explains as much of the remaining variation as possible, and so on
- Typically, we look at the first few dimensions of the transformed data and use as a means of dimensionality reduction
- PCA is a linear transformation
- PCA is often used for:
 - Data visualization
 - Infer qualitative relationships between groups

Principal component analysis



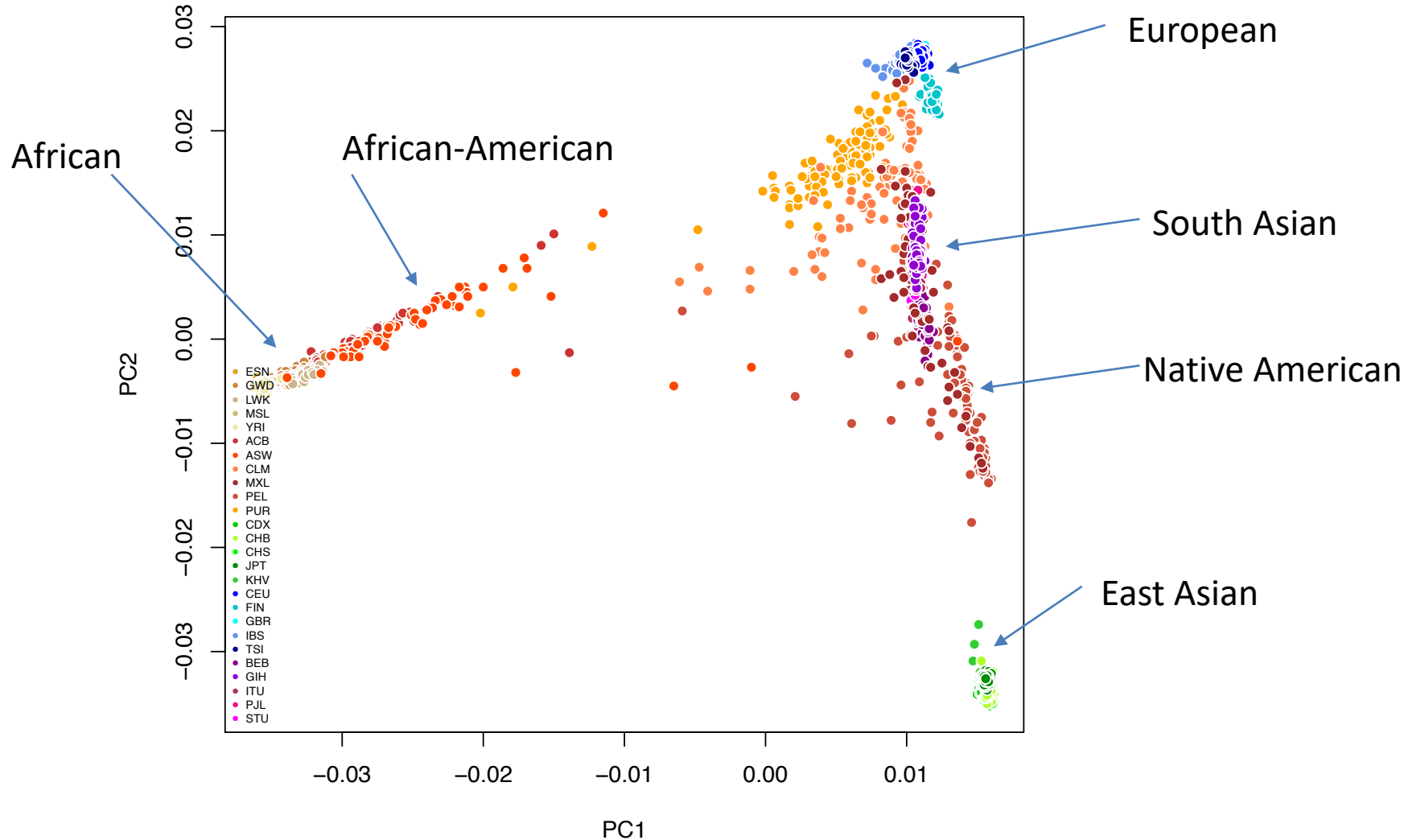
The 1000 Genomes project

- Whole-genome **sequence data** from 2504 individuals from 26 populations
- A catalog of human genetic variation, useful as a reference or **imputation** panel
- Completely public. Download from <ftp://ftp-trace.ncbi.nih.gov/1000genomes/>



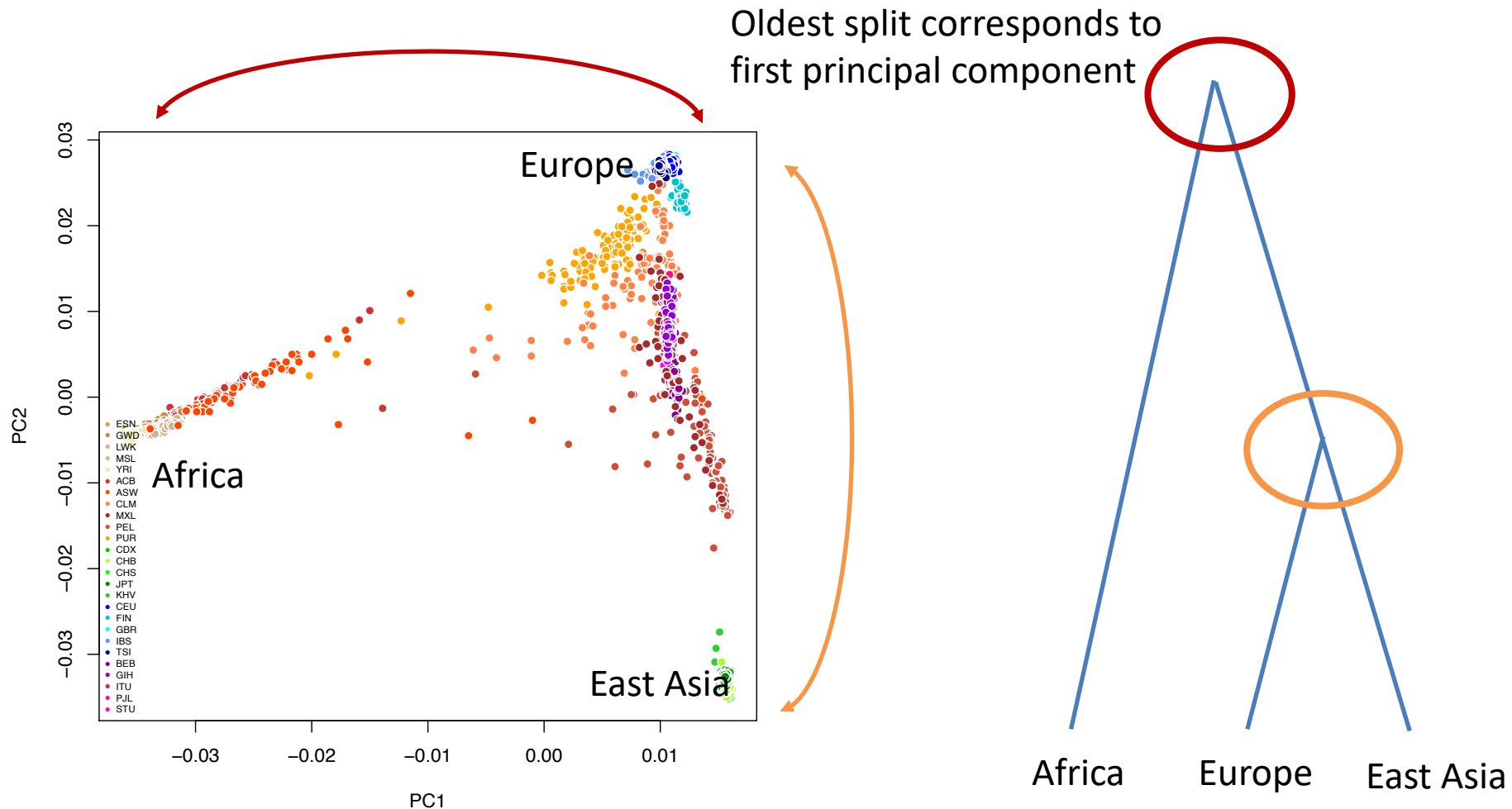
```
##ALT=<ID=CN120,Description="Copy number allele: 120 copies">
##ALT=<ID=CN121,Description="Copy number allele: 121 copies">
##ALT=<ID=CN122,Description="Copy number allele: 122 copies">
##ALT=<ID=CN123,Description="Copy number allele: 123 copies">
##ALT=<ID=CN124,Description="Copy number allele: 124 copies">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##bcftools_annotateVersion=1.6+htslib-1.6
##bcftools_annotateCommand=annotate -x INFO 20130502_phase3_final/ALL.chr20.phase3_shapeit2_mvncall_integrated_v5.20130502.genotypes.vcf.gz; Date=Fri Jan 19 19:20:16 2018
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT HG00096 HG00097 HG00099 HG00100 HG00101 HG00102 HG00103 HG00105 HG00106 HG00107 HG00108 HG00109 HG00110 HG00111
20 60343 . G A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60419 . A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60479 rs149529999 C T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60522 rs150241001 T TC 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60568 . A C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60571 rs116145529 C A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60579 . G A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60649 . A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60778 . A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60795 rs184056664 G C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60808 . G A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60810 . G GA 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60826 . A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60828 rs187713677 T G 100 PASS . GT 0|0 0|0 0|1 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60864 . G A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60895 . A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 60916 . G T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61044 . C A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61070 . C T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61098 rs6078030 C T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|1 0|0 1|0 0|0 0|0 0|0 0|1
20 61118 . A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61138 rs140305189 C CT 100 PASS . GT 0|0 0|0 0|1 0|0 0|0 0|1 0|0 0|0 0|0 0|0 0|0 0|0
20 61270 rs143291093 A C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61271 . T A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61272 . C A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61279 rs189899941 C T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61329 rs182162684 C T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61388 rs146681064 T C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61409 rs139103017 A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61437 . A C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61450 . T C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61517 rs187280035 C T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61538 . A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61638 . C T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61651 rs76553454 C A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61711 rs369824431 G T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61724 rs142532139 A C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61795 rs4814683 G T 100 PASS . GT 1|0 0|0 0|0 0|0 0|0 0|1 0|0 1|0 0|0 0|1 0|0 0|1
20 61955 . C T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 61972 . T C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62100 rs6047235 T C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62174 . AGATCAGTCCTTT A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62255 rs192879424 T C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62283 . T C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62348 rs141113228 A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62387 . T A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62420 rs185326153 A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62461 . C T 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62471 rs188652106 G A 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62478 rs192812899 A G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62545 rs150267191 C G 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
20 62553 rs114190700 T C 100 PASS . GT 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0 0|0
```

Global population structure



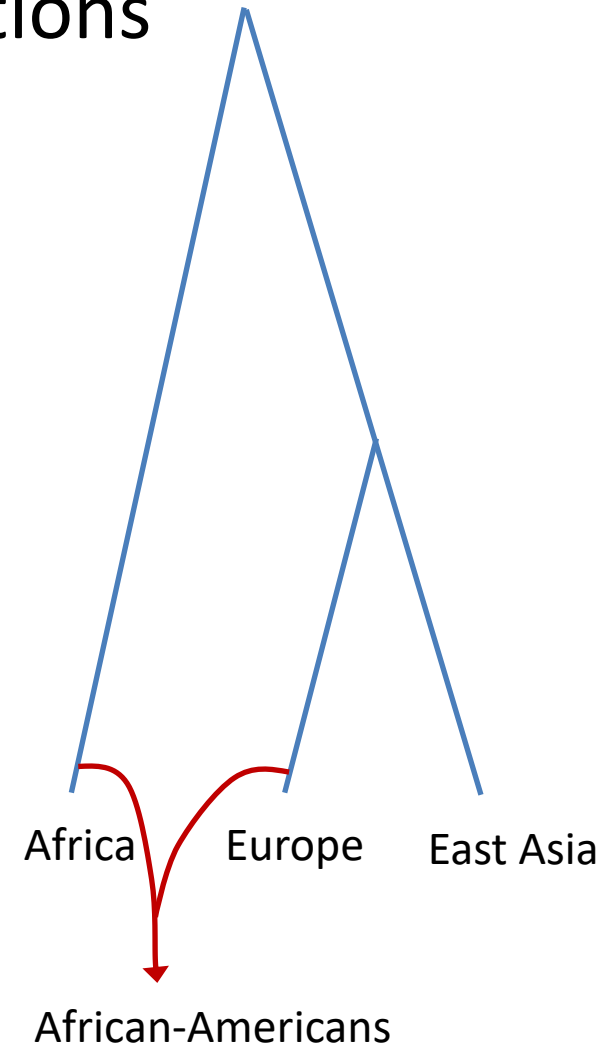
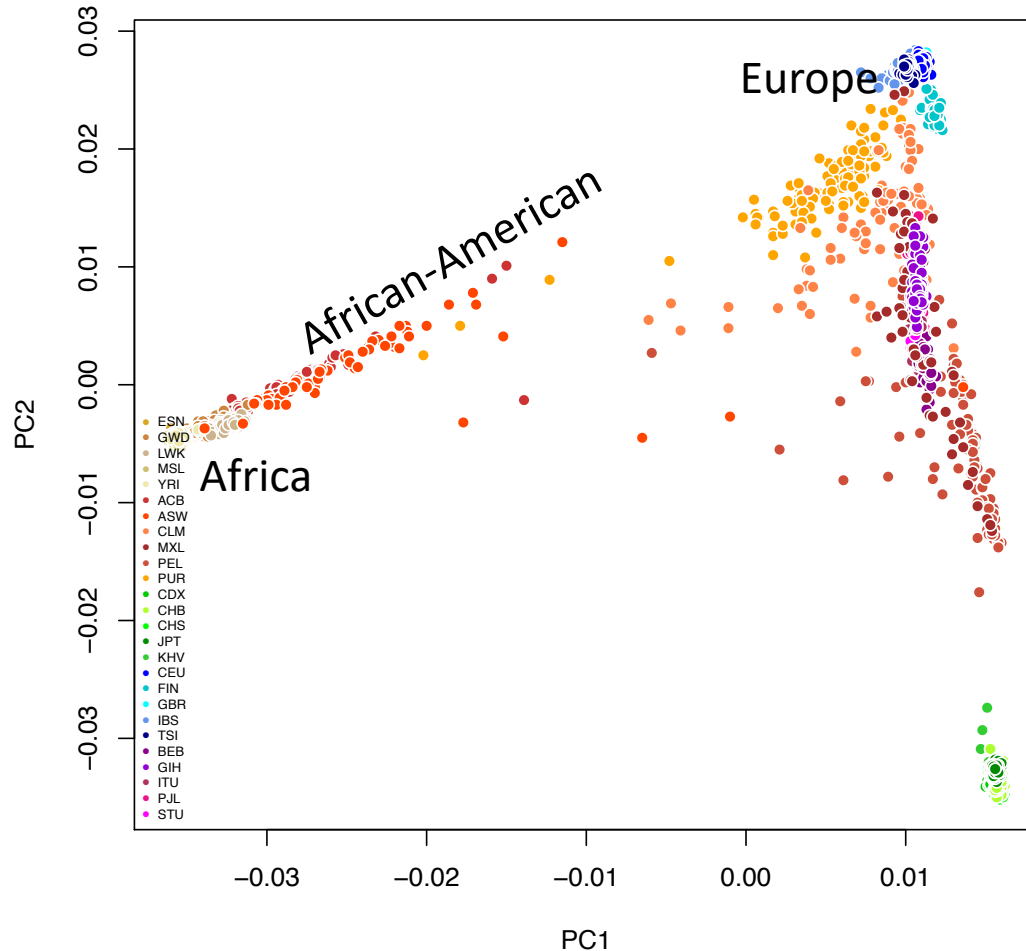
What causes these patterns?

1. Populations **splits** separate populations

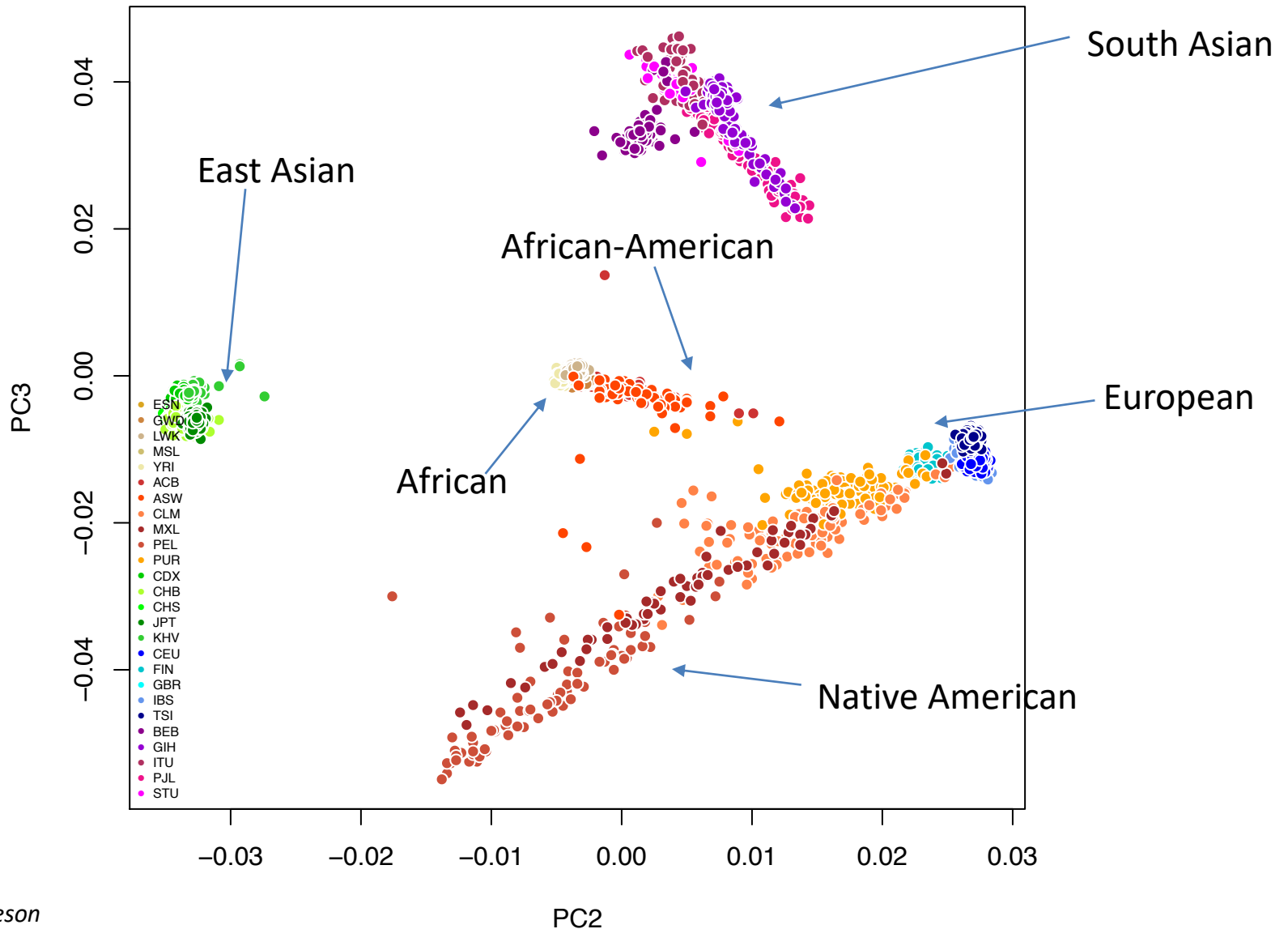


What causes these patterns?

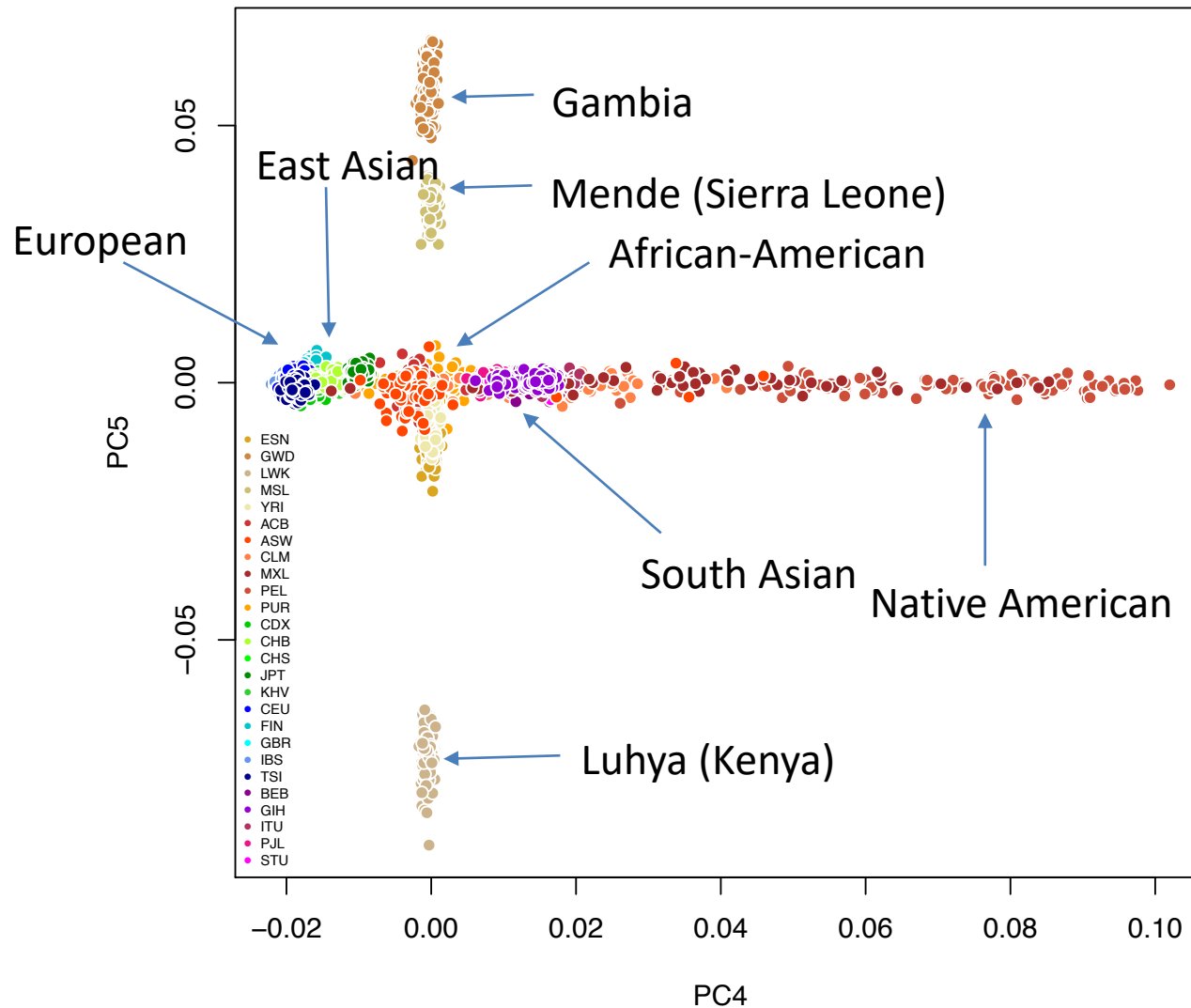
2. **Admixture** merges populations




Global population structure



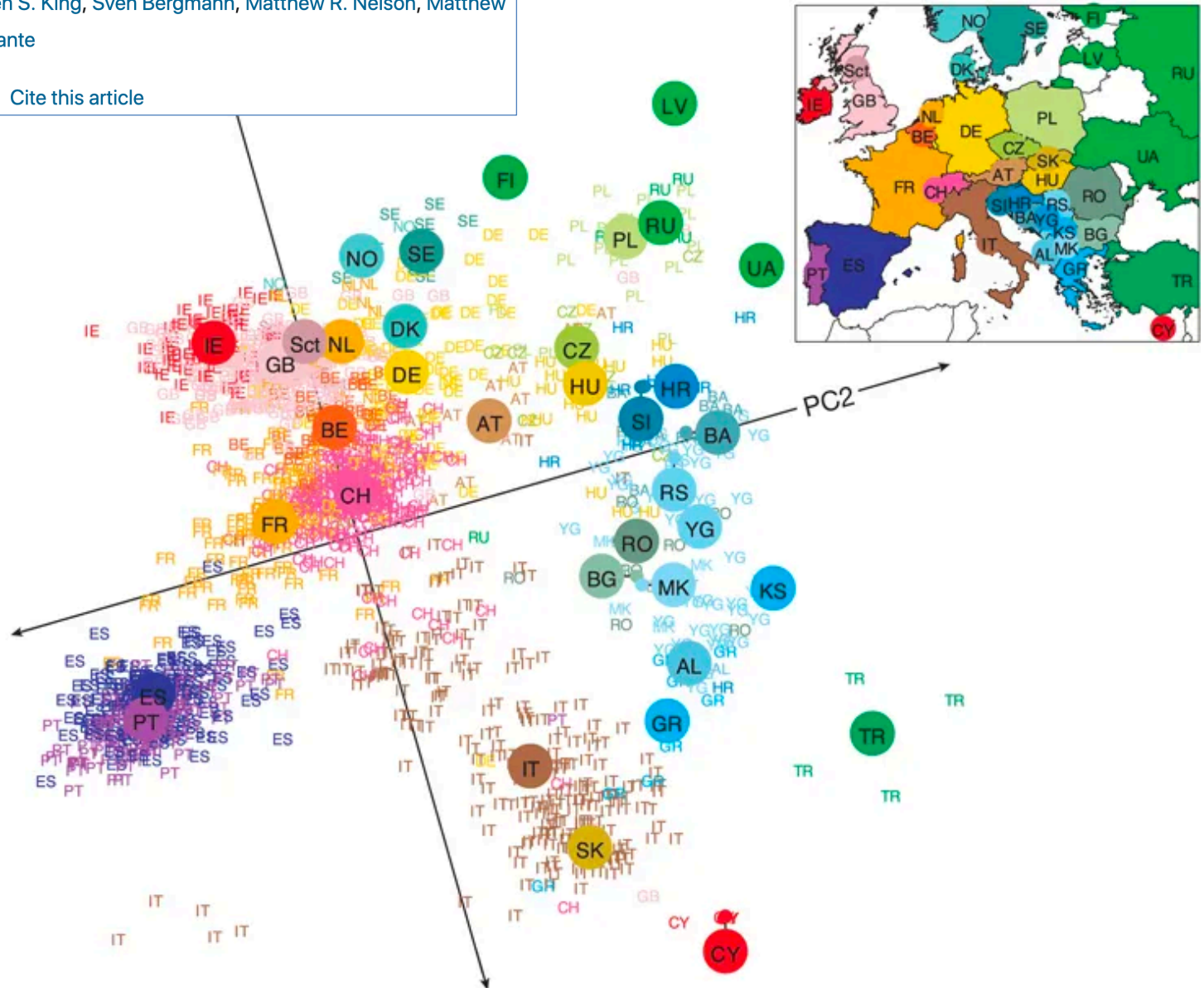
Global population structure



Genes mirror geography within Europe

John Novembre , Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R. Boyko, Adam Auton, Amit Indap, Karen S. King, Sven Bergmann, Matthew R. Nelson, Matthew Stephens & Carlos D. Bustamante























Nature **456**, 98–101(2008) | [Cite this article](#)



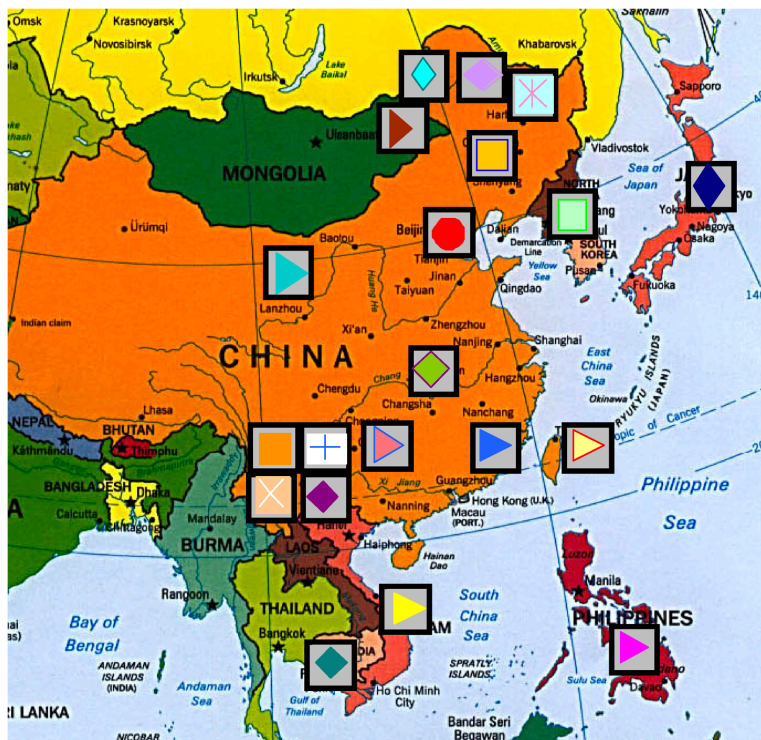
Analysis of East Asia Genetic Substructure Using Genome-Wide SNP Arrays

Chao Tian, Roman Kosoy, Annette Lee, Michael Ransom, John W. Belmont, Peter K. Gregersen, Michael F. Seldin 

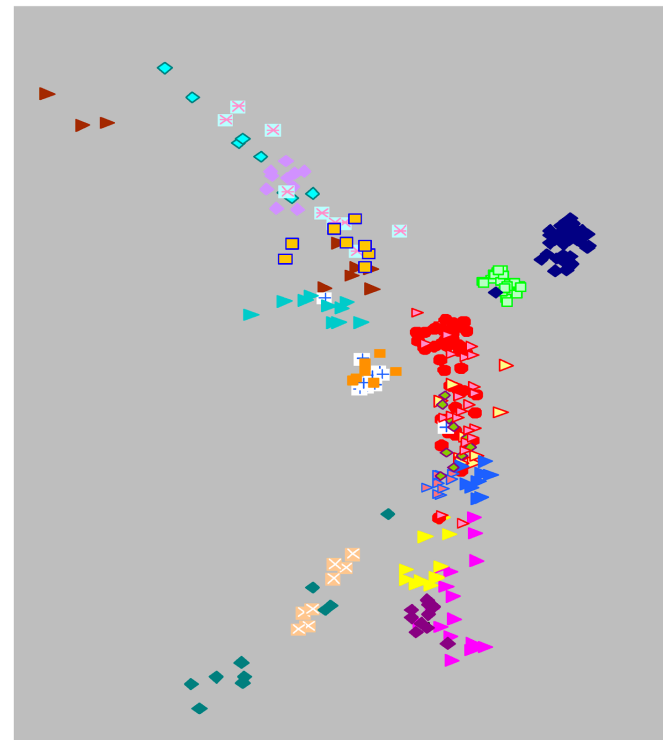
Published: December 5, 2008 • <https://doi.org/10.1371/journal.pone.0003862>

-  FIL
-  VIET
-  LAHU
-  DAI
-  CAMB
-  CHB
-  MGL
-  ORQ
-  DAUR
-  KOR
-  TWN
-  YI
-  HEZ
-  MIAO
-  NAXI
-  SHE
-  TU
-  TUJ
-  XIBO
-  CHA
-  JPT
-  YAK

C

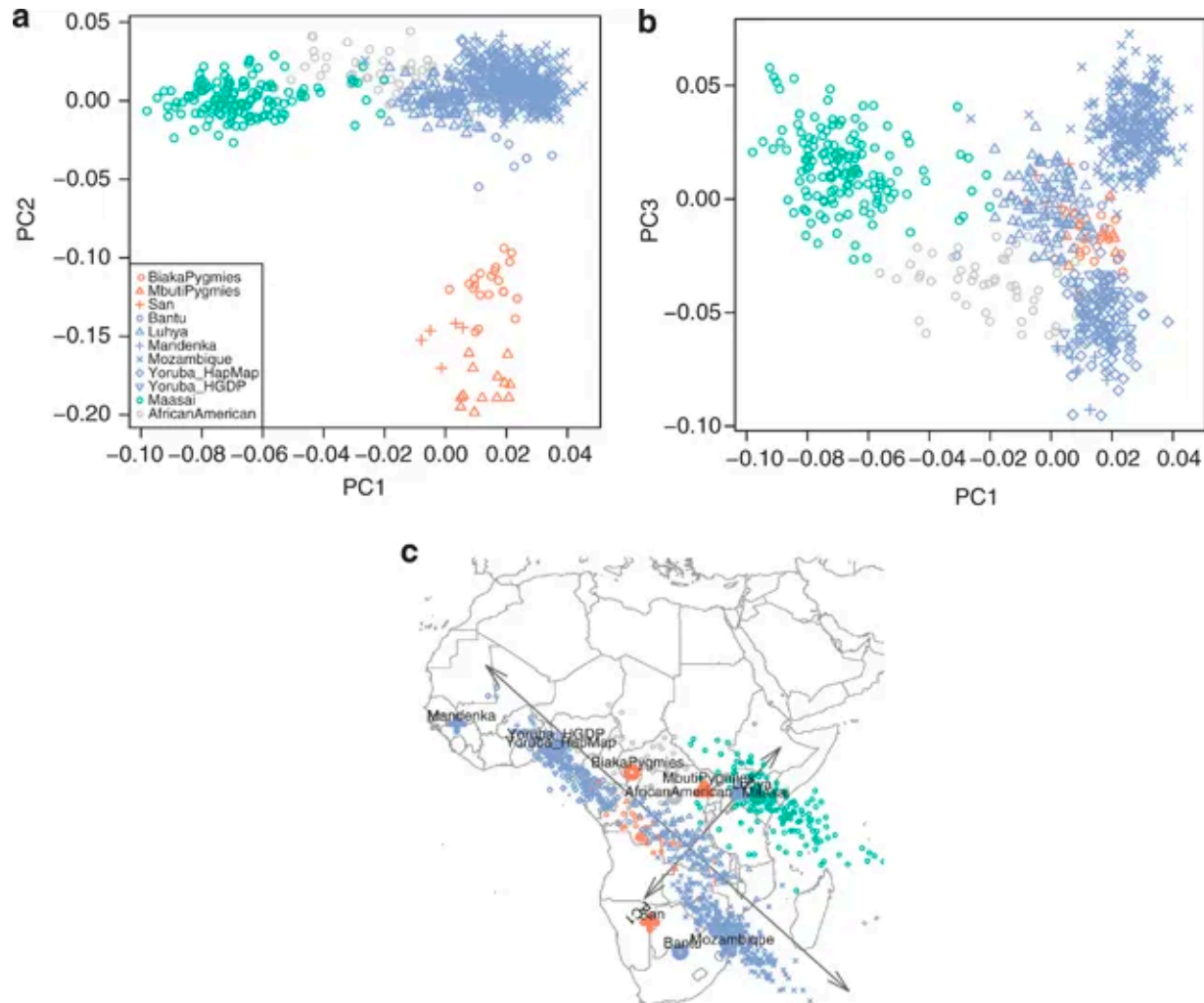


D



A genomic analysis identifies a novel component in the genetic structure of sub-Saharan African populations

Martin Sikora, Hafid Laayouni, Francesc Calafell, David Comas & Jaume Bertranpetit 



PCA

Step 1

input data

$$X_{\text{orig}} = \left[\begin{array}{c} \text{feature } j \\ \vdots \\ x_{ij} \\ \vdots \end{array} \right] \left. \vphantom{\begin{array}{c} \text{feature } j \\ \vdots \\ x_{ij} \\ \vdots \end{array}} \right\}^n \text{ } \left. \vphantom{\begin{array}{c} \text{feature } j \\ \vdots \\ x_{ij} \\ \vdots \end{array}} \right\}^{n \times p}$$

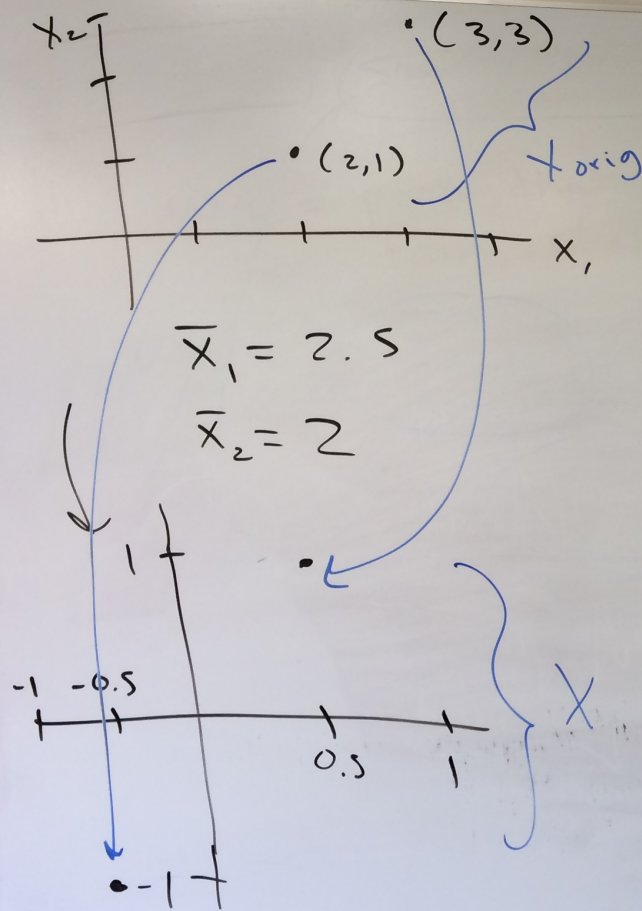
p

no label
"y"!

Step 2

Subtract off the column-wise mean from each column (feature).

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \Rightarrow \bar{x}_j - x_{ij}$$



Step 3

compute covariance matrix A

$$\text{cov}(f, g) = \frac{1}{n-1} \sum_{i=1}^n (f_i - \bar{f})(g_i - \bar{g})$$

$$\text{cov}(f, f) = \text{var}(f) = \frac{1}{n-1} \sum_{i=1}^n (f_i - \bar{f})^2$$

all zero for \bar{f}

$$A = \begin{bmatrix} \text{cov}(f, f) & \text{cov}(f, g) \\ \text{cov}(g, f) & \text{cov}(g, g) \end{bmatrix}$$

Symmetric

Step 4

compute eigenvalues & eigenvectors of A .

$$A\vec{v} = \lambda\vec{v}$$

\Rightarrow sort by eigenvalue

$X_{n \times p} A_{p \times p} \rightarrow n \times p$

Step 5

$$W = \begin{bmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_r \\ | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \dots & \vec{v}_r \\ | & | & & | \end{bmatrix} \begin{matrix} \leftarrow \text{sorted} \\ \text{high} \rightarrow \text{low} \end{matrix}$$

r is my new dimension.

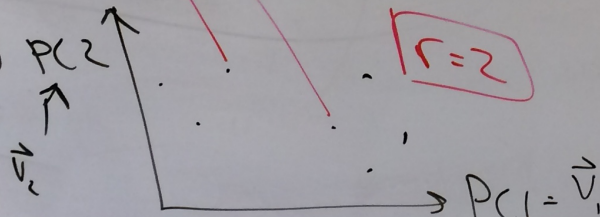
$p \times r$

$$T = X W$$

$n \times p$ $p \times r$

$n \times r$

Step 6



Handout 22

$$\bar{f} = \frac{1}{2}, \quad \bar{g} = \frac{1}{2}$$

$$\Rightarrow X = \begin{bmatrix} -1/2 & 1/2 \\ -1/2 & 1/2 \\ -1/2 & 1/2 \\ 1/2 & -1/2 \\ 1/2 & -1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

Step 3

$$\text{var}(f) = \frac{1}{5} \cdot 6 \cdot \frac{1}{4} = \frac{3}{10}$$

$$\text{cov}(f, g) = \frac{1}{5} \cdot 6 \cdot \left(-\frac{1}{4}\right) = -\frac{3}{10}$$

Handout 22

$$A = \begin{bmatrix} 3/10 & -3/10 \\ -3/10 & 3/10 \end{bmatrix}$$

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

$$\det \left(\overbrace{\begin{bmatrix} 3/10 & -3/10 \\ -3/10 & 3/10 \end{bmatrix}}^A - \overbrace{\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}}^{\lambda I} \right) = 0$$

$$\left(\frac{3}{10} - \lambda \right)^2 - \left(\frac{3}{10} \right)^2 = 0$$

$$\begin{cases} \lambda_1 = \frac{3}{5} \\ \lambda_2 = 0 \end{cases}$$

sort
descending

$$\cancel{\left(\frac{3}{10} \right)^2} - 2 \left(\frac{3}{10} \right) \lambda + \lambda^2 - \cancel{\left(\frac{3}{10} \right)^2} = 0$$

$$\lambda \left(\lambda - \frac{3}{5} \right) = 0$$