LECTURE 18

Recall SVD, $\|A\|_F$.

Def The operator (a.k.a. spectral) norm:

$$\|A\| = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A) = \max_x \frac{\|Ax\|_2}{\|x\|_2}$$

by optimization form (last class)

FACTS

- (a) $\|A\| \le \|A\|_F \le \sqrt{r}\,\|A\|$ where $r = \text{rank}(A)$ **HW?**

(b) (Diagonal) $A = \begin{bmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{bmatrix} \Rightarrow$ by SVD, $(\sigma_i(A)) = (|d_i|) \Rightarrow \|A\| = \max_i |d_i|$

(c) (Orthog. invariance) $\|UA\| = \|AV\| = \|A\|$ $\forall$ orthogonal U, V
$[\,\|UAx\|_2 = \|Ax\|_2\,]$

(d) (Quadratic form) If A is symmetric, $\boxed{\|A\| = \max_{x \text{ unit}} |x^T A x|}$ **HW**

(e) $\|A+B\| \le \|A\| + \|B\|$ (Δ ineq.) **HW**

(f) $\|AB\| \le \|A\|\,\|B\|$ $[\,\|ABx\|_2 \le \|A\| \cdot \|Bx\|_2 \le \|A\| \cdot \|B\| \cdot \|x\|_2\,]$

(g) $\|A^T\| = \|A\|$

RANDOM VECTORS

$X \in \mathbb{R}^d$

- Mean $\mathbb{E}X = \mu$
- Variance in 1D: $Var(X) = \mathbb{E}(X-\mu)^2 = \mathbb{E}X^2 - \mu^2$

$\downarrow$

Covariance matrix in kD: $\boxed{Cov(X) = \mathbb{E}(X-\mu)(X-\mu)^T = \mathbb{E}XX^T - \mu\mu^T.}$

- Entries:

$$Cov(X)_{ij} = \mathbb{E}(X-\mu)_i (X-\mu)_j = \mathbb{E}(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j) = Cov(X_i, X_j)$$

Diagonal entries:

$$Cov(X)_{ii} = Cov(X_i, X_i) = Var(X_i)$$

- Ex: $X \in \mathbb{R}^2$ $\quad Cov(X) = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) \\ Cov(X_2, X_1) & Var(X_2) \end{bmatrix}$

- $Cov(X)$ is a $\underbrace{d \times d \text{ symmetric}}_{\text{Obvious}}$ $\underbrace{\text{PSD matrix}}_{\text{Proof}}$.

$\forall v \in \mathbb{R}^d$: $\quad v^T \Sigma v = v^T \mathbb{E}XX^T v = \mathbb{E}\underbrace{v^T X}_{=\langle v, X\rangle} \underbrace{X^T v}_{=\langle X, v\rangle} = \mathbb{E}\langle X, v\rangle^2 \geq 0.$ $(*)$

EXAMPLE: M.D. NORMAL DISTRIBUTION

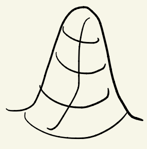
- **Def** $Z=(Z_1,\dots,Z_d)\in\mathbb{R}^d$ is a standard normal r.vector if $Z_i\sim N(0,1)$ independent

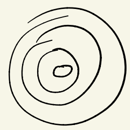
- $\mathbb{E}Z=0$, $\mathrm{Cov}(Z)=\begin{bmatrix}1 & & 0\\ & 1 & \\ 0 & & \ddots\ 1\end{bmatrix}=I_d$ $\Rightarrow$ notation $\boxed{Z\sim N(0,I_n)}$

- pdf: $f(x)=f(x_1)\cdots f(x_d)=\frac{1}{\sqrt{2\pi}}e^{-x_1^2/2}\cdots\frac{1}{\sqrt{2\pi}}e^{-x_d^2/2}=\boxed{\frac{1}{(2\pi)^{d/2}}e^{-\|x\|_2^2/2}}$ $x\in\mathbb{R}^d$.

- $f(x)$ is rotationally invariant: $f(x)=f(Ux)$ $\forall$ orth. U $\Rightarrow$

Prop (Rotation invariance) If $Z\sim N(0,I_n)$ then $UZ\sim N(0,I_n)$ $\forall$ orthogonal matrix U





Def (General normal)
$X=(X_1,\dots,X_d)\in\mathbb{R}^d$ is a normal r.vector if X is an affine transformation of a standard normal r.vector $Z\sim N(0,I_d)$, i.e.

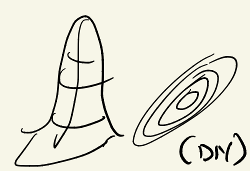
$$X=AZ+\mu \qquad \text{for some invertible } A\in\mathbb{R}^{d\times d},\ \mu\in\mathbb{R}^d.$$

- $\mathbb{E}X=\mu$,

$$\mathrm{Cov}(X)=\mathbb{E}(AZ)(AZ)^T=\mathbb{E}\left[AZZ^TA^T\right]=A\cdot\underbrace{\mathbb{E}[ZZ^T]}_{=I_d}\cdot A^T=\boxed{AA^T=:\Sigma}$$

- **Ex**: pdf of X is

$$f(x)=\frac{1}{(2\pi)^{d/2}\sqrt{\det(\Sigma)}}\exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right),\quad x\in\mathbb{R}^d$$

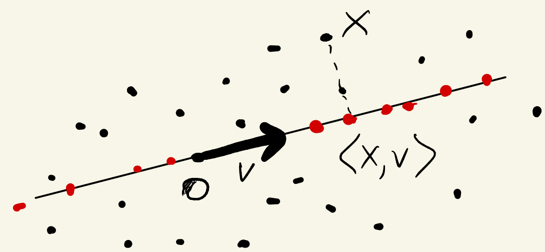


(DN)

$\Rightarrow$ pdf is determined by μ,Σ. $\Rightarrow$ notation $\boxed{X\sim N(\mu,\Sigma)}$

Principal Component Analysis

- Let $X \in \mathbb{R}^d$ be a random vector with mean 0 (for simplicity - by translation) and $Cov(X) = \Sigma$.



- Reduce dimension? project X onto 1D line that "best explains" the variation in X.

  $\Leftrightarrow$ find a unit vector v that maximizes

$$Var(\langle X, v\rangle) = \mathbb{E}\langle X, v\rangle^2 = v^T \Sigma v \qquad (* \text{ p. } 2)$$

$$\max_{v \text{ unit}} Var(\langle X, v\rangle) = \max_v v^T \Sigma v = \boxed{\lambda_1(\Sigma)}, \quad \text{argmax} = \boxed{v_1(\Sigma)}$$

- Next best direction:

$$\max_{\substack{v \perp v_1 \\ \text{unit}}} Var(\langle X, v\rangle) = \boxed{\lambda_2(\Sigma)}, \quad \text{argmax} = \boxed{v_2(\Sigma)}$$

  $\Rightarrow$ project X onto 2D plane spanned by v_1, v_2.

PCA

- In order to reduce dimension of X from d to k, project X onto the subspace spanned by the top k eigenvectors of $Cov(X)$. "Principal components" of X.
- This subspace explains the variability of X the best among all k-dimensional subspaces.

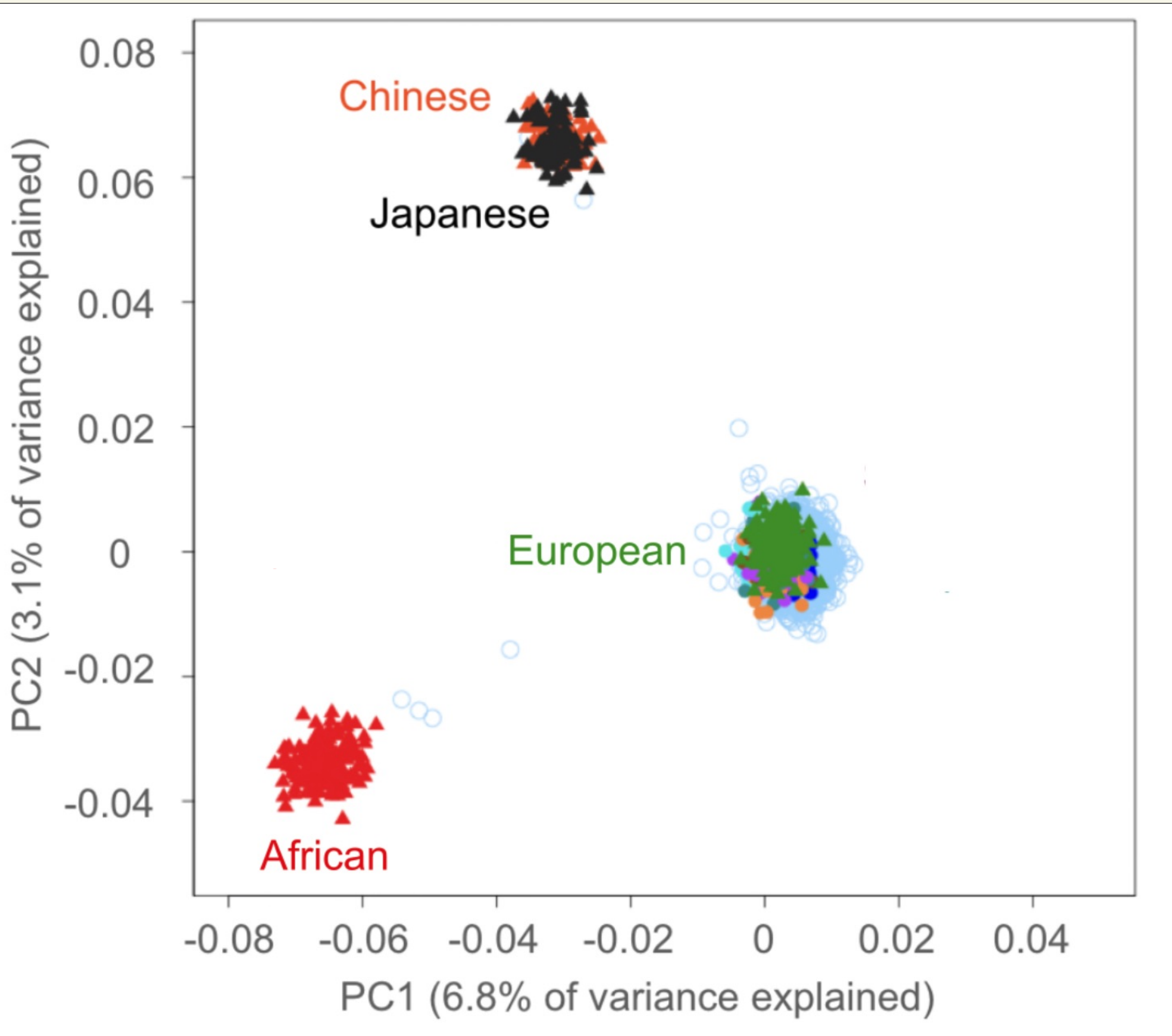
- Exploratory data analysis. Unsupervised learning.
- Advantages & disadvantages.

APPLICATION: GENETIC PCA

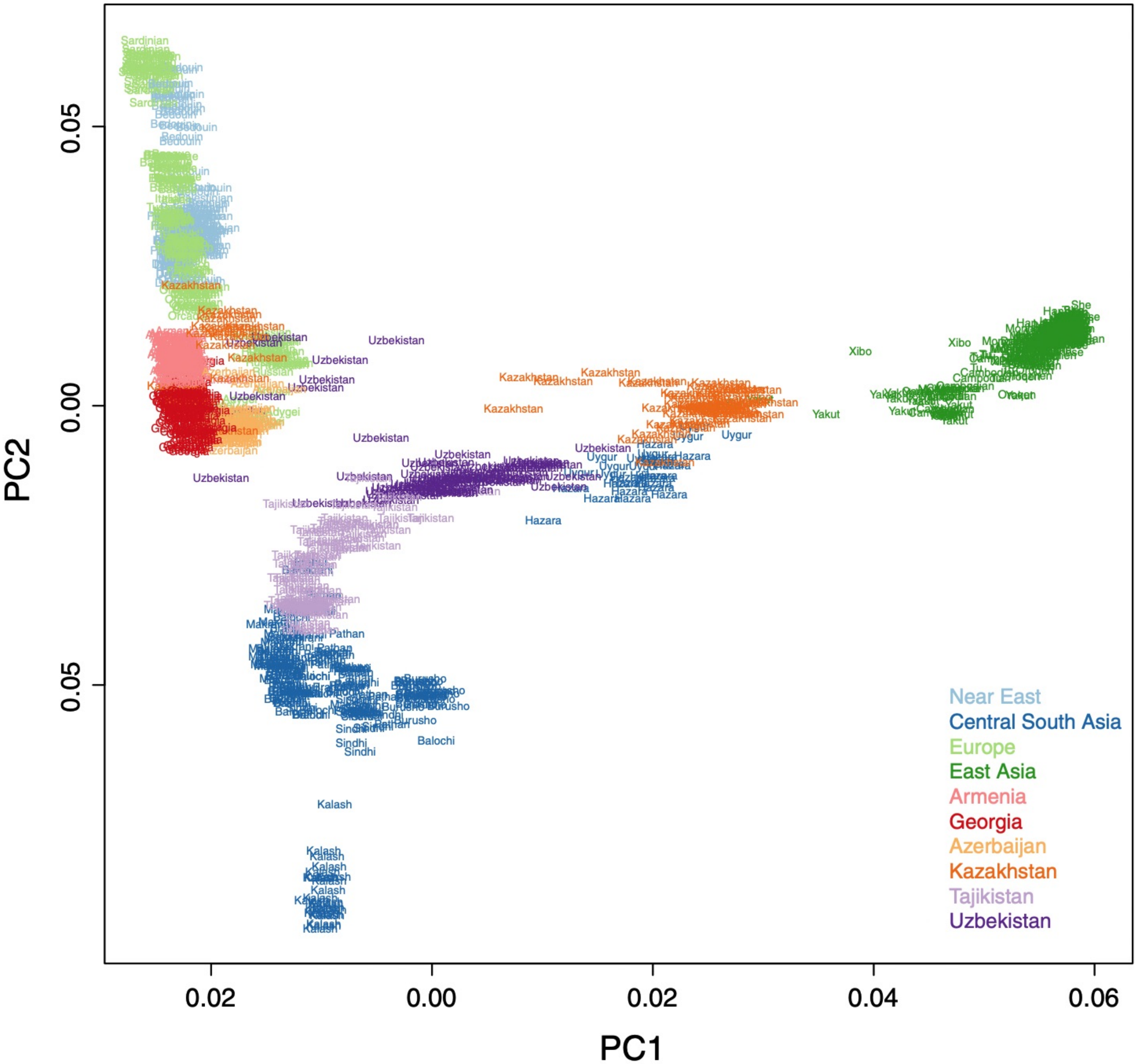
$X = (x_1, \ldots, x_d)$ of a randomly chosen person

gene expressions

WORLD:

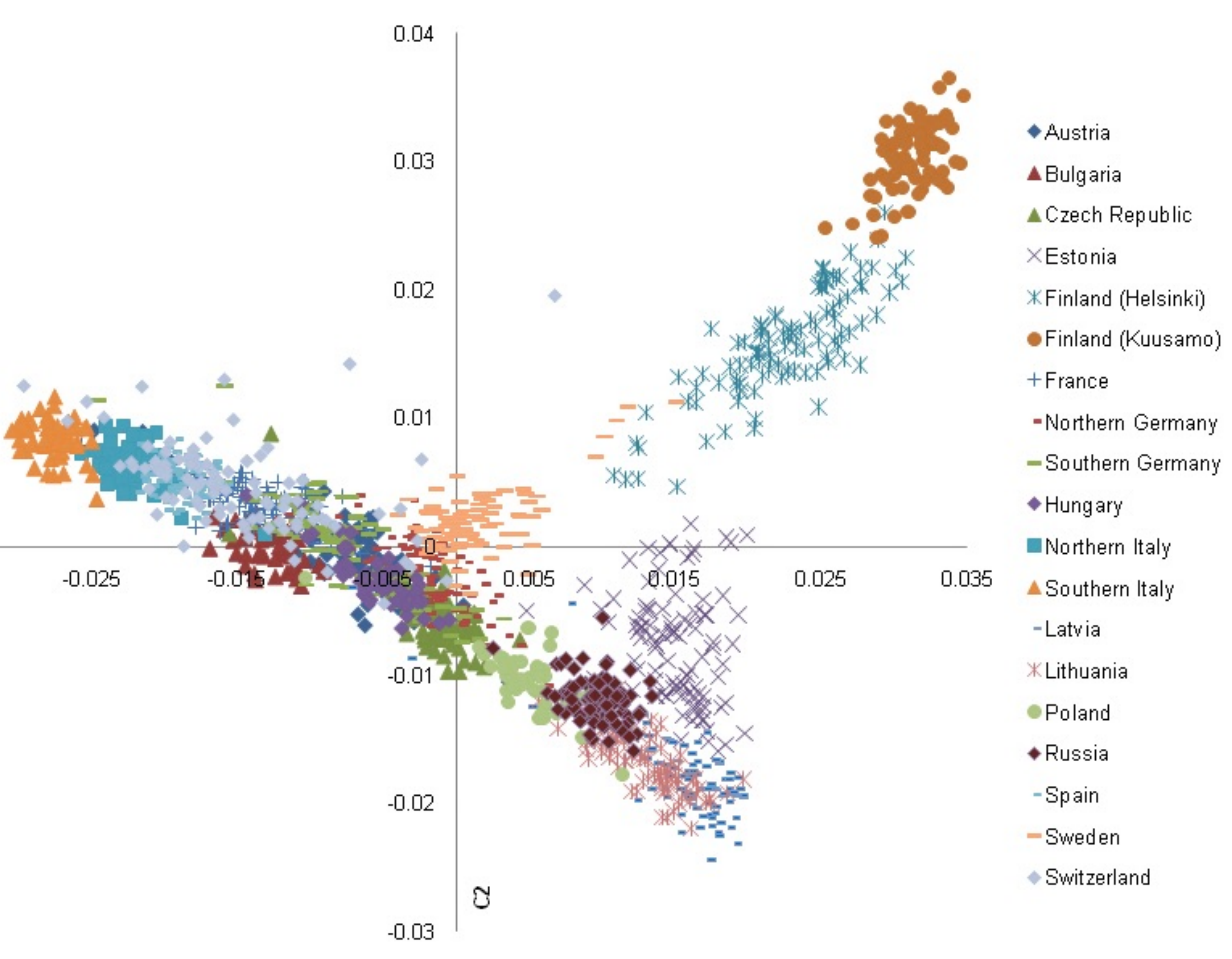


ASIA



PC1
PC2
0.05
0.00
0.05
0.02
0.00
0.02
0.04
0.06
Near East
Central South Asia
Europe
East Asia
Armenia
Georgia
Azerbaijan
Kazakhstan
Tajikistan
Uzbekistan

EUROPE



0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.025
-0.015
-0.005
0.005
0.015
0.025
0.035
C2
Austria
Bulgaria
Czech Republic
Estonia
Finland (Helsinki)
Finland (Kuusamo)
France
Northern Germany
Southern Germany
Hungary
Northern Italy
Southern Italy
Latvia
Lithuania
Poland
Russia
Spain
Sweden
Switzerland