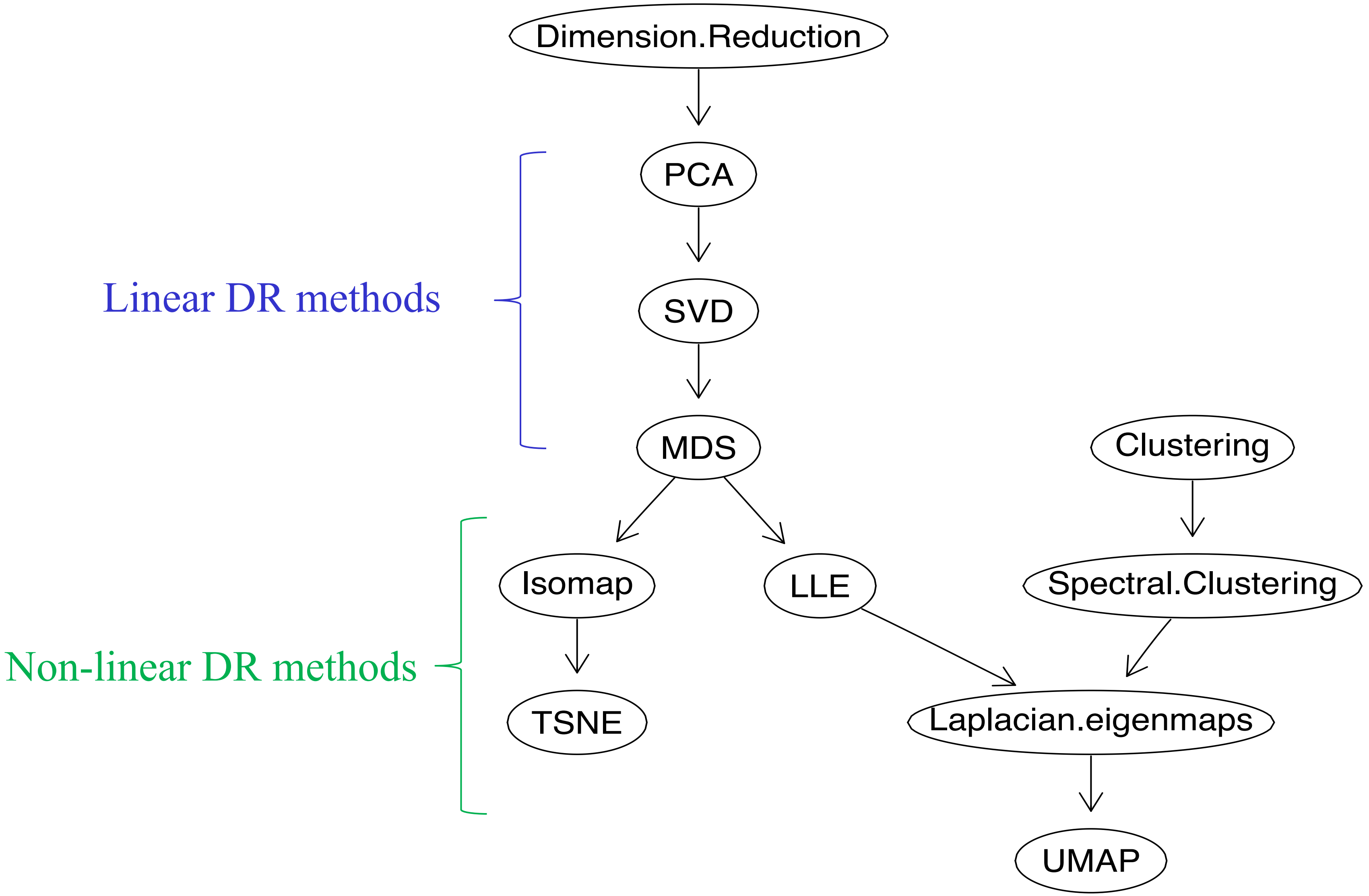
Dimension Reduction Methods:
From PCA to TSNE and UMAP

Maxwell Lee

High-dimension Data Analysis Group
Laboratory of Cancer Biology and Genetics
Center for Cancer Research
National Cancer Institute

April 23, 2020

Outline for Dimension Reduction Methods



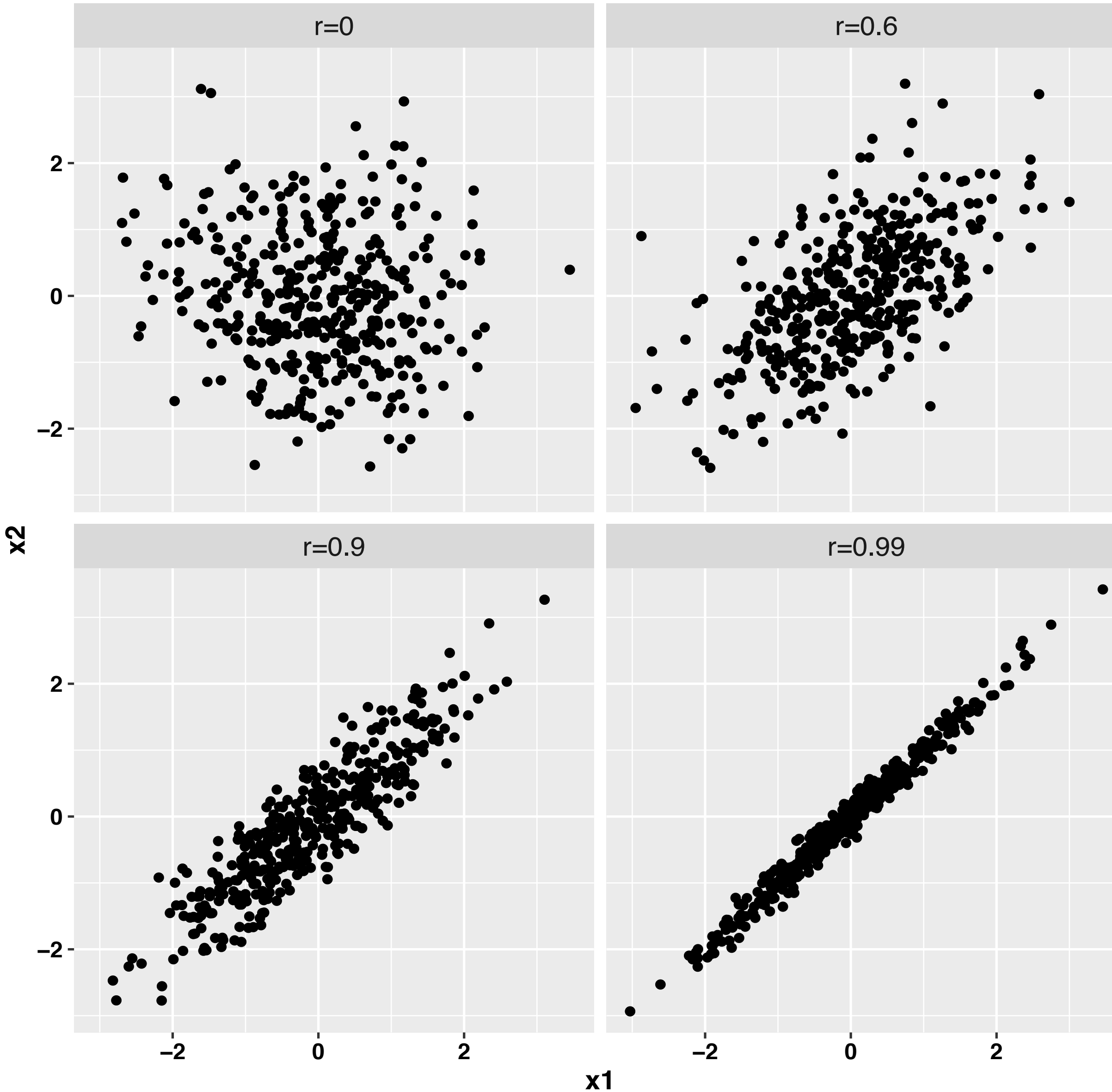
The Presence of Correlation Between Variables Is the Reason Why We Can Reduce Dimension by PCA

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

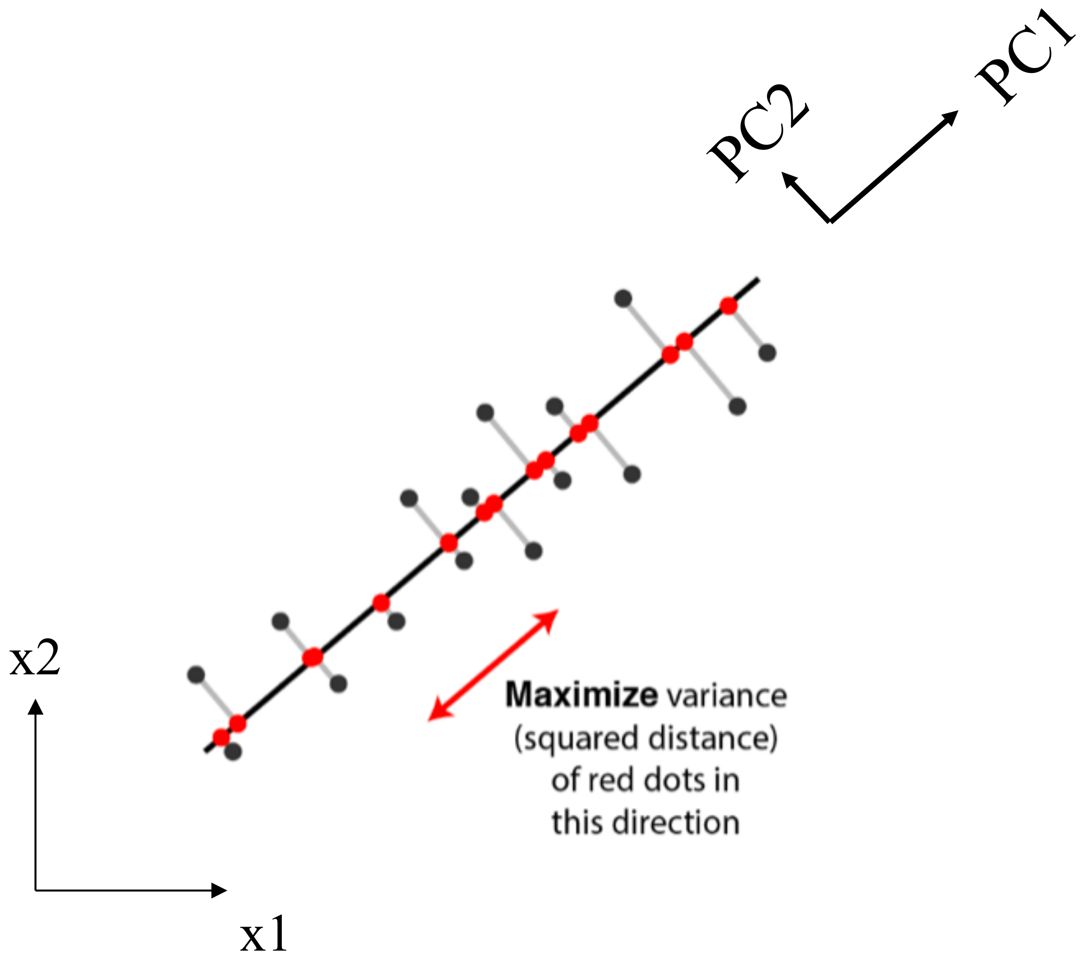
$r = \rho$

	μ_1	μ_2	σ_1	σ_2	ρ
d1	0	0	1	1	0
d2	0	0	1	1	0.6
d3	0	0	1	1	0.9
d4	0	0	1	1	0.99

$n = 400$



Principal Component Analysis (PCA)



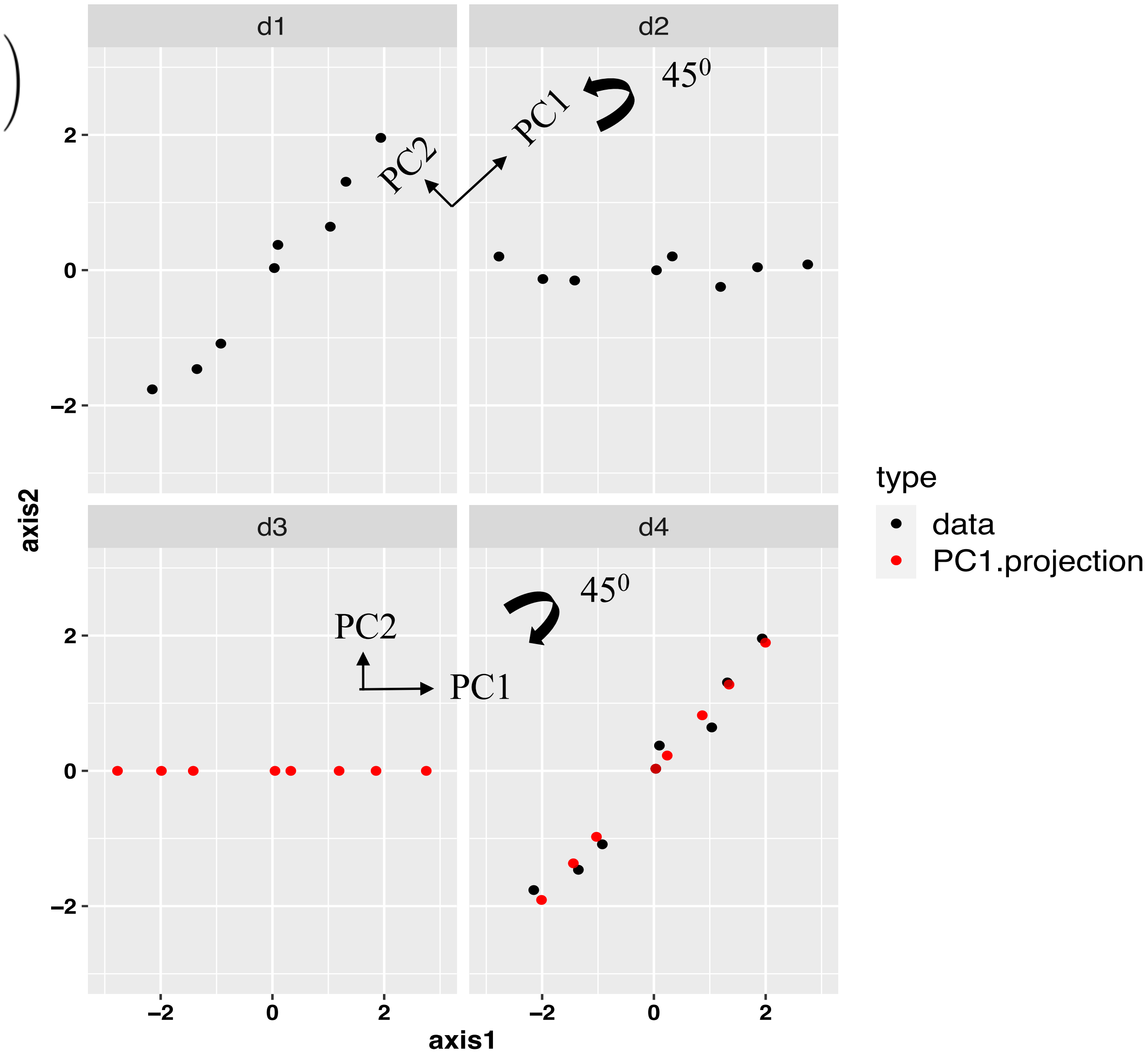
Karl Pearson 1901; Harold Hotelling 1933-1936

Geometric View of PCA: Rotation of Coordinates

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

$\rho = 0.9$

$n = 8$

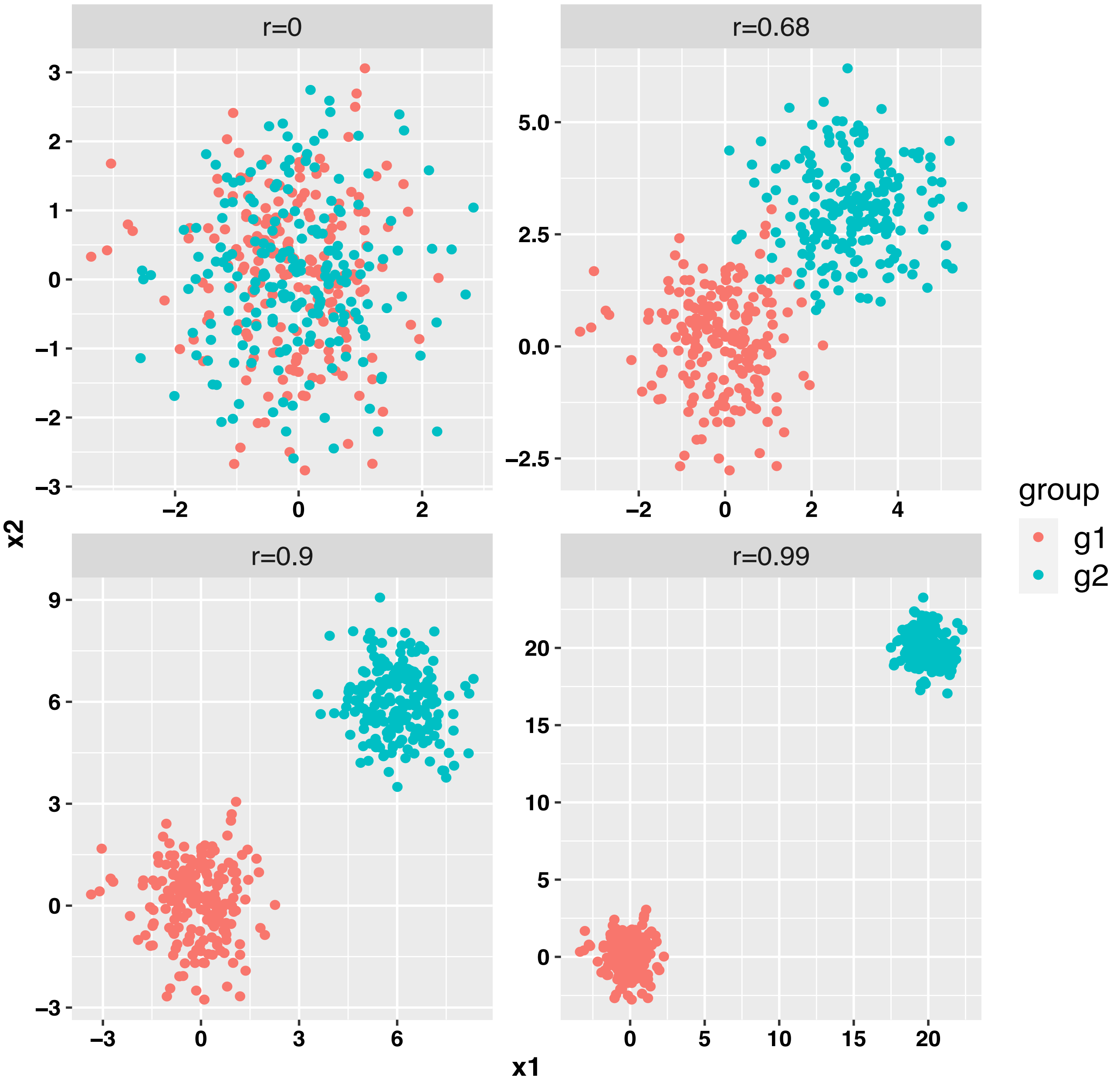


Correlation Between Variables Can Result from Heterogeneity in Sample

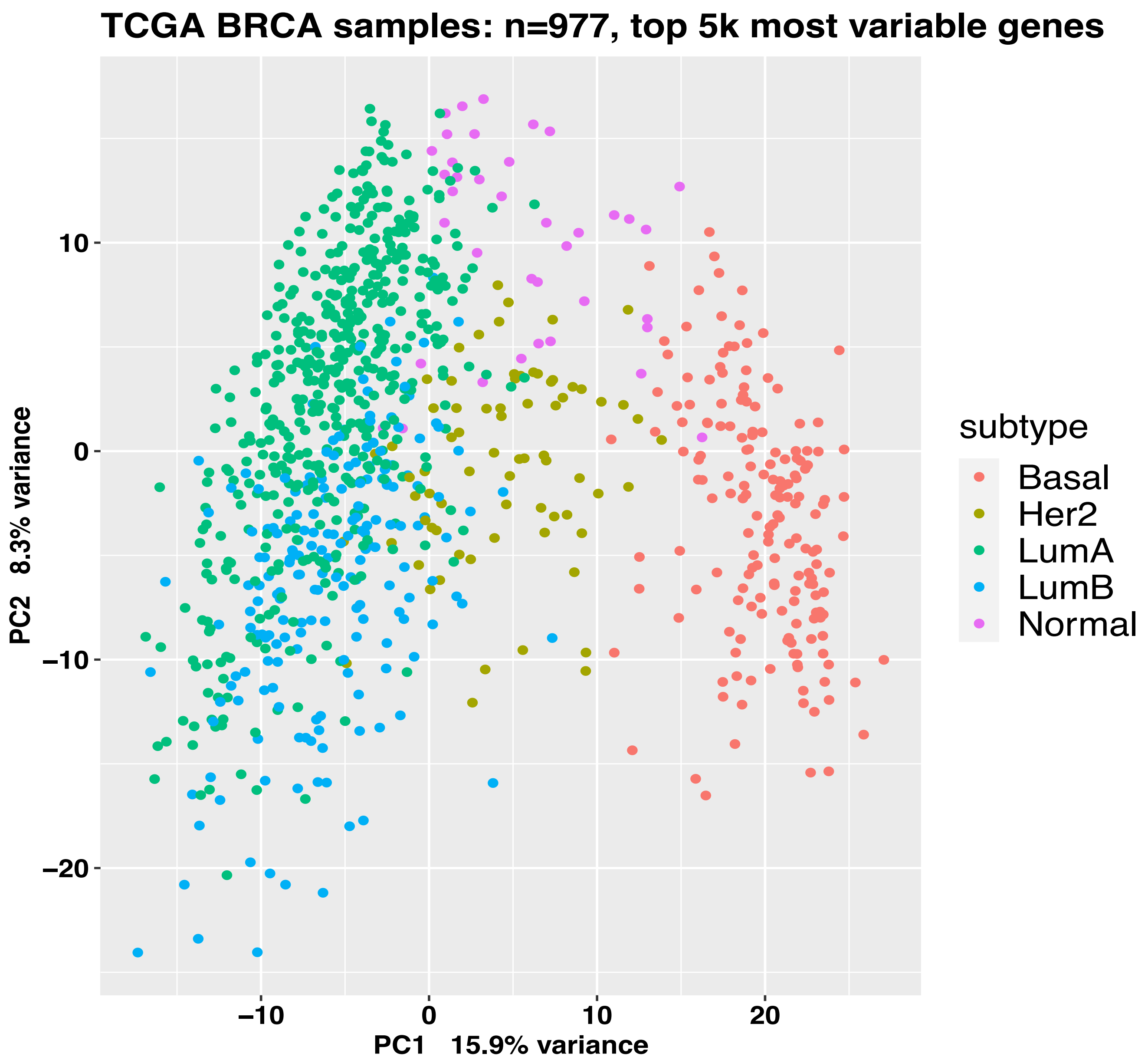
$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

$$\rho = 0$$

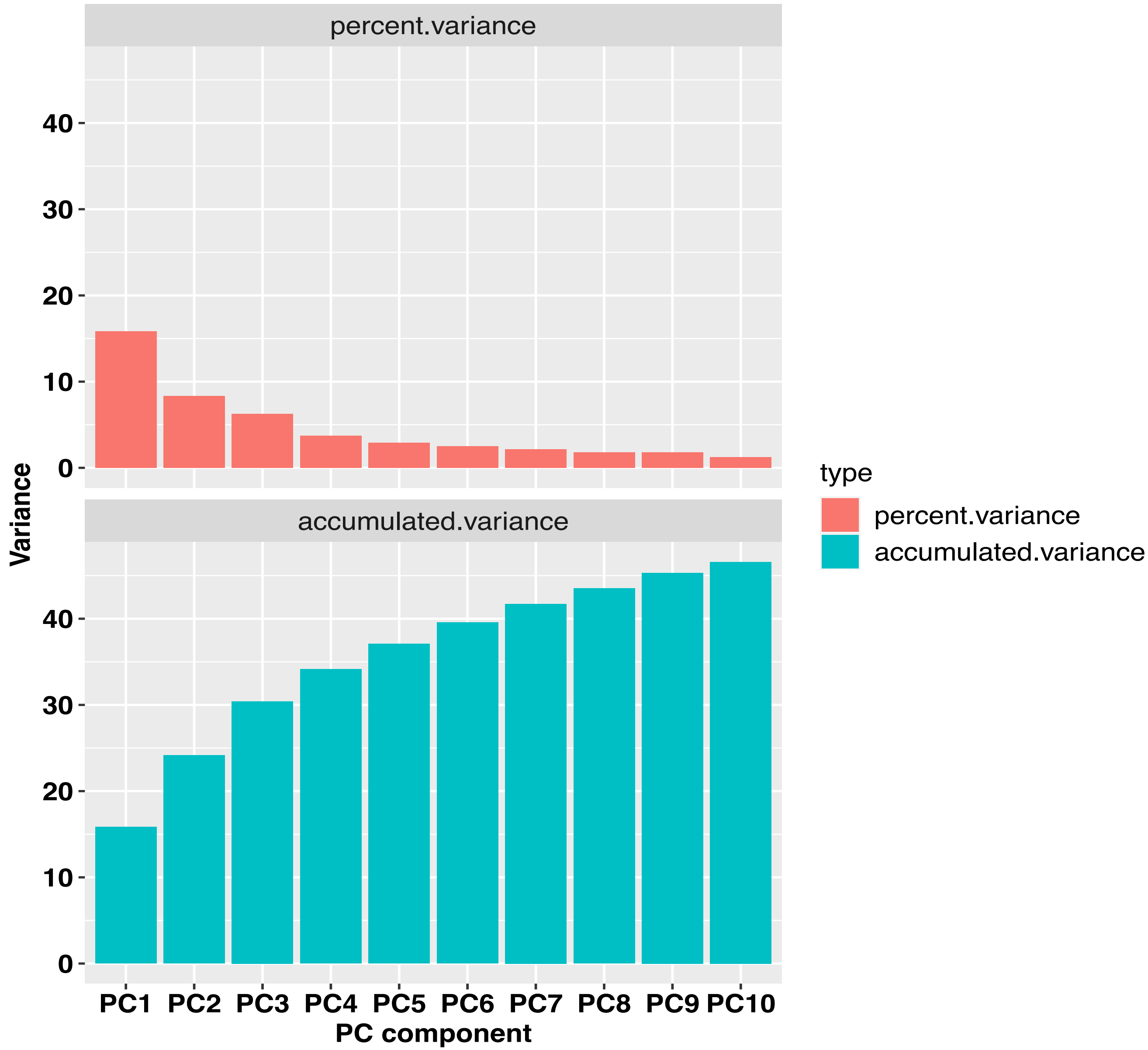
	Group1		Group2	
	μ_1	μ_2	μ_1	μ_2
d1	0	0	0	0
d2	0	0	3	3
d3	0	0	6	6
d4	0	0	20	20



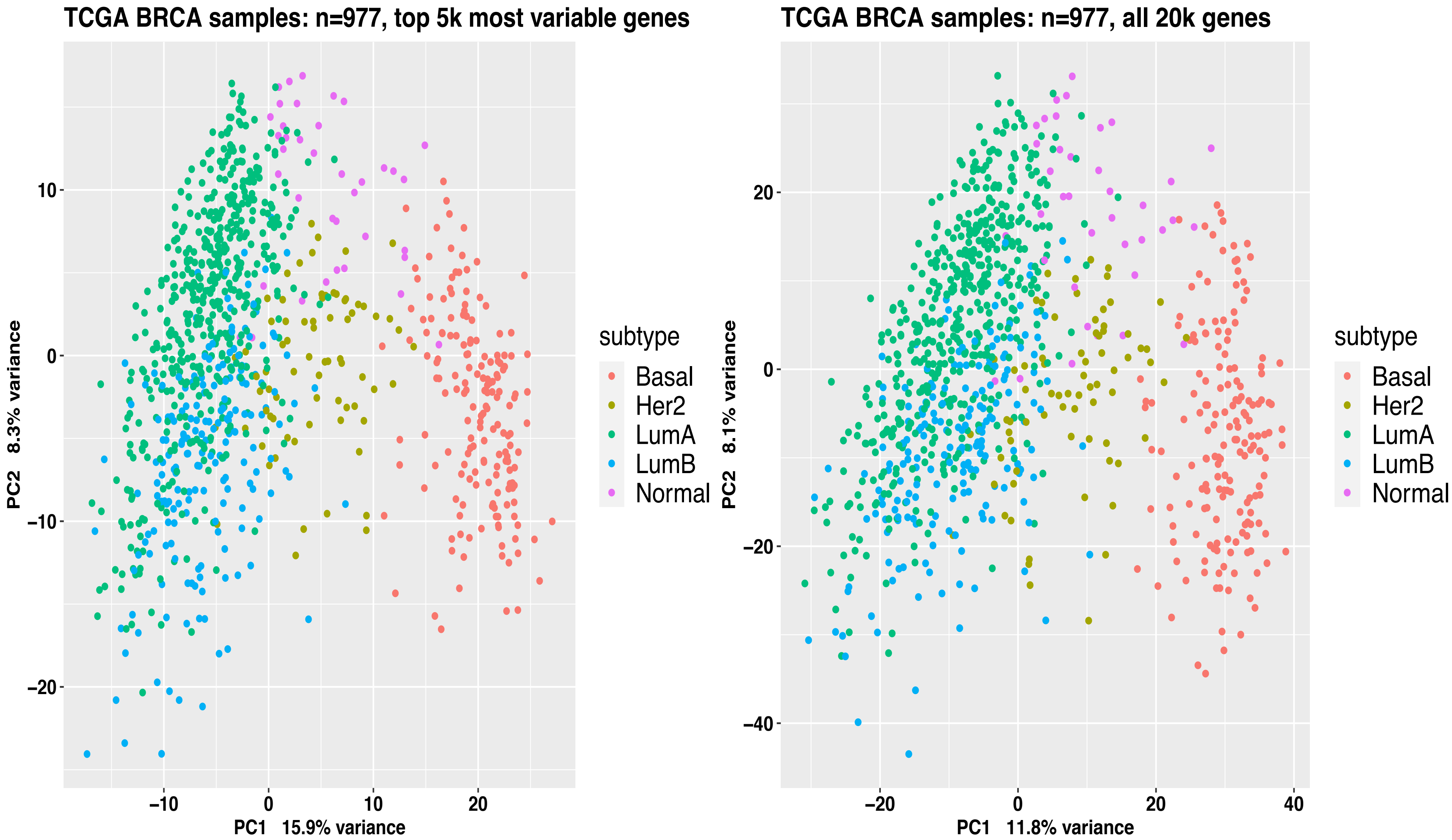
PCA Analysis of TCGA Breast Cancer RNAseq Data



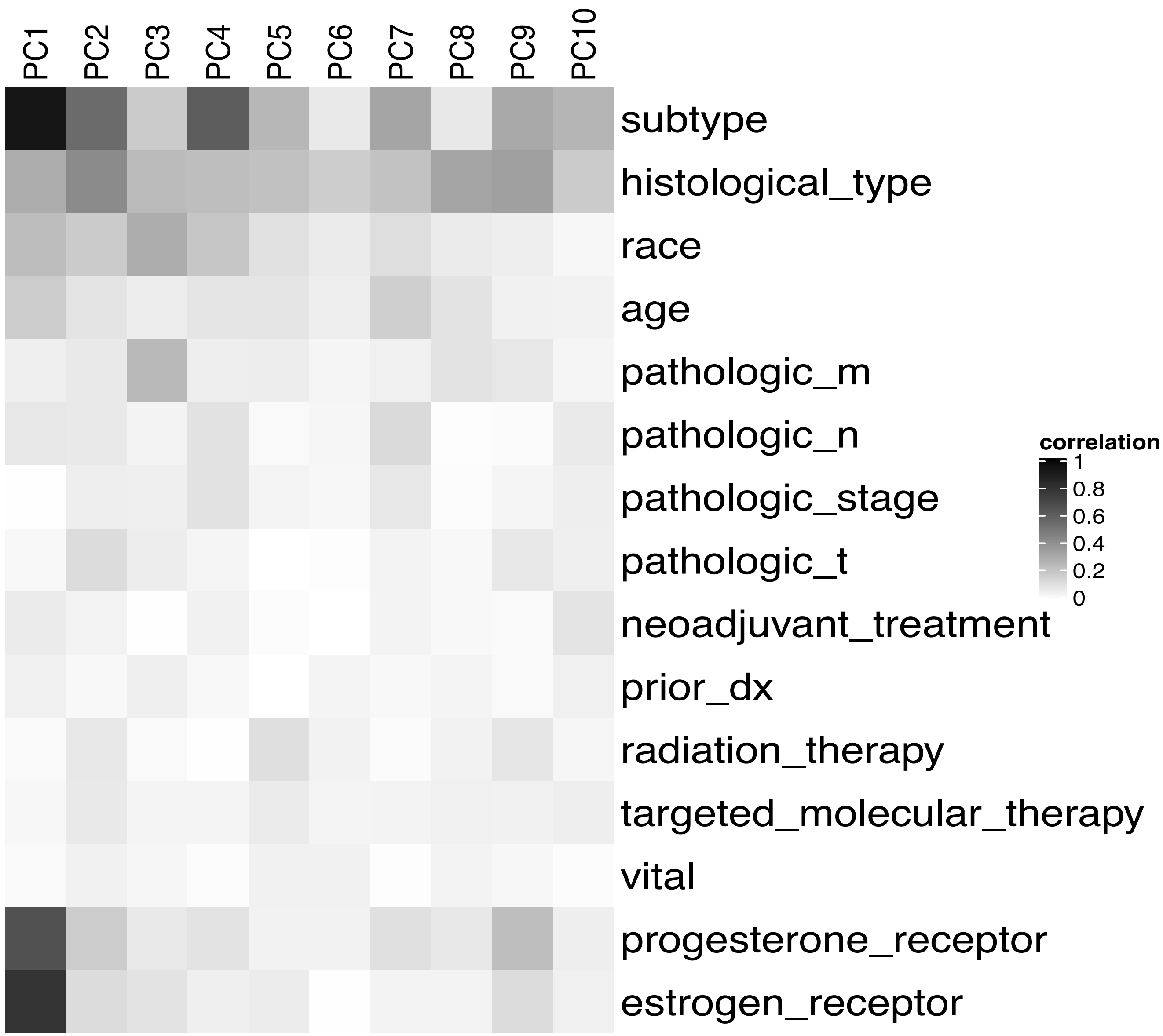
Variance of Principal Components Are Ranked from the Highest to the Lowest



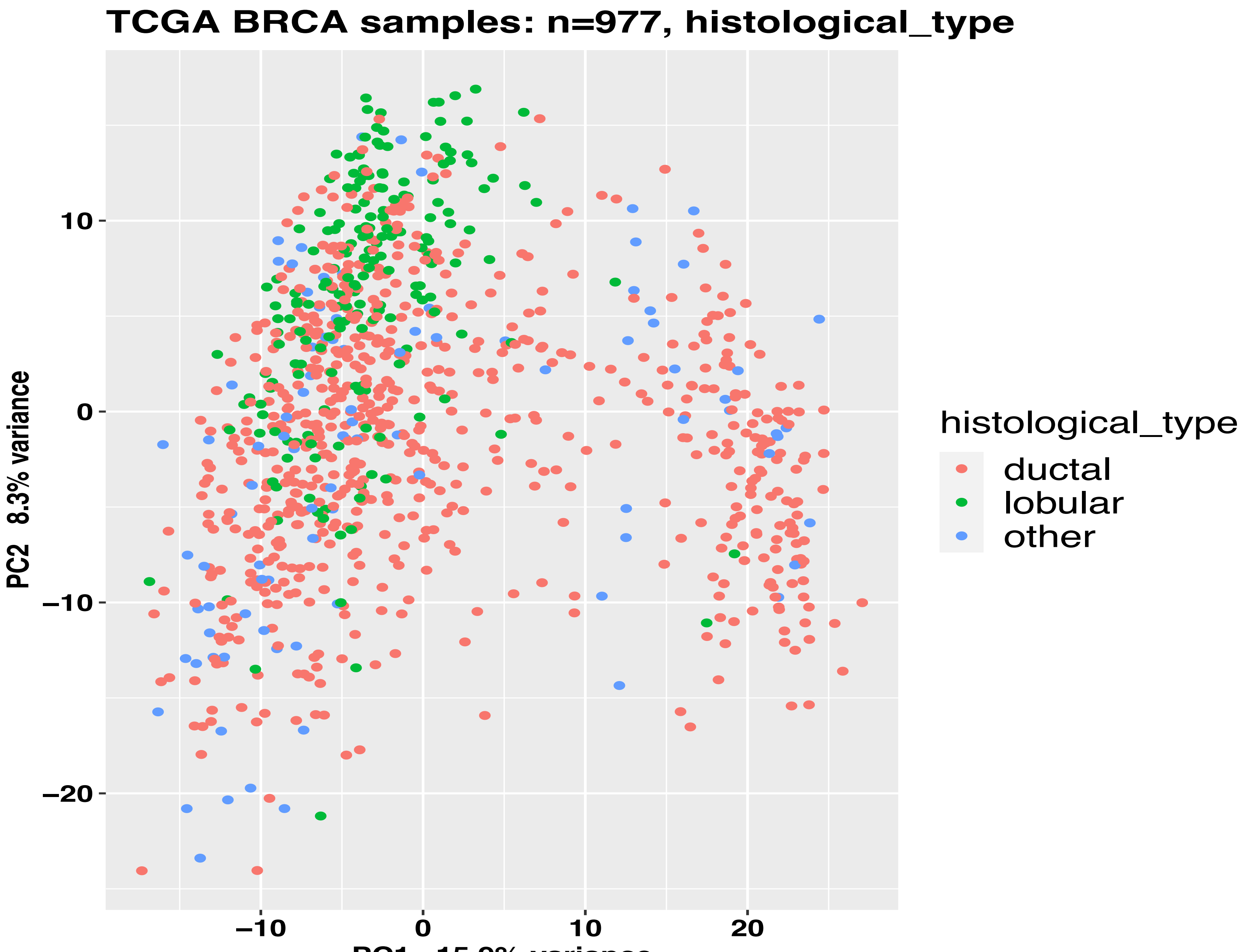
Filtering Out Genes of Low Variance Increases Percent of Variance Accounted for by PC1



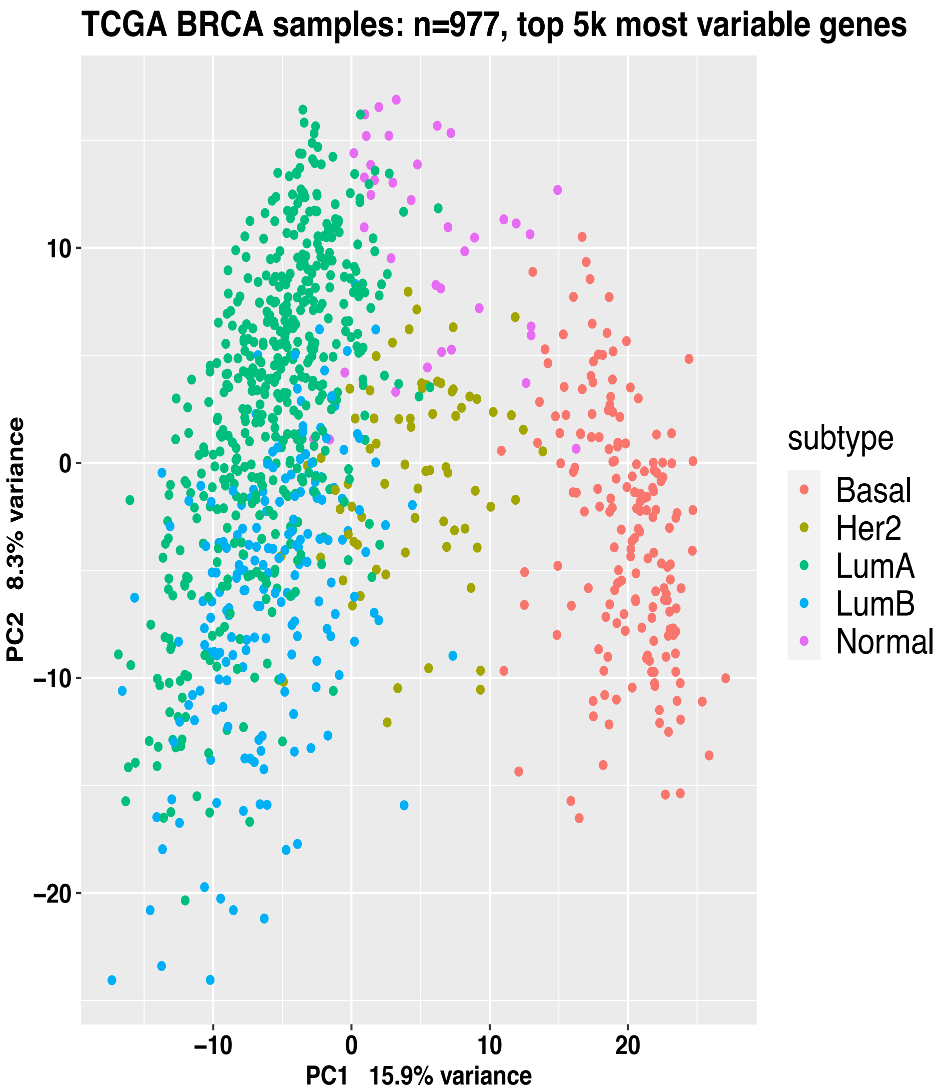
Correlation Between Principal Components and Phenotypes of Breast Cancer Data

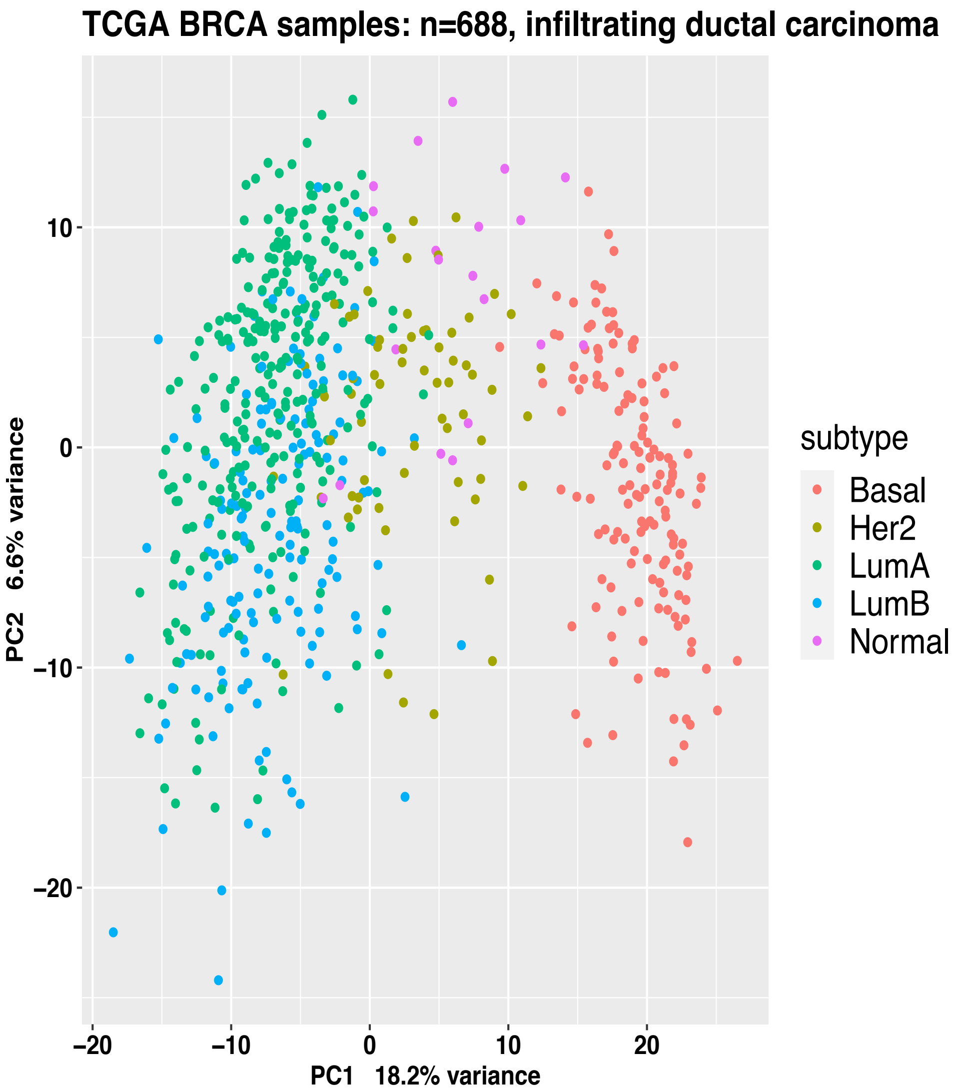


Variation in Histological Type Is Associated with PC2

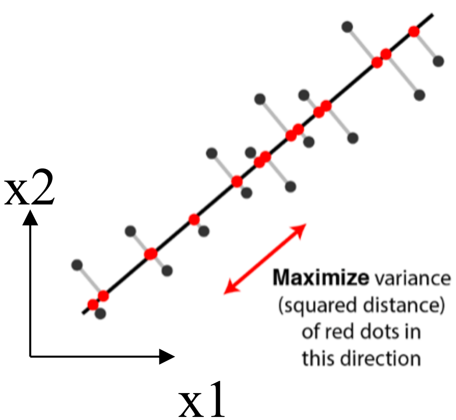


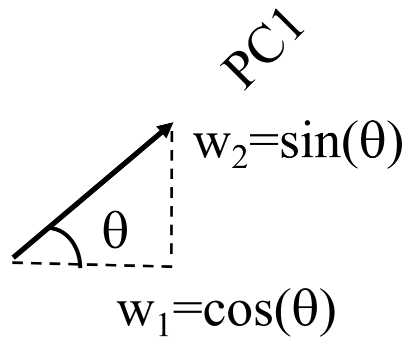
Removing Heterogeneity in Histological Type Reduces PC2 Variance and Increases PC1 Variance





Algorithm of PCA: How Does PCA Find the Direction of PC1?





$$\underbrace{\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ . & . \\ . & . \\ x_{n1} & x_{n2} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{Xw} = \underbrace{\begin{bmatrix} w_1 x_{11} + w_2 x_{12} \\ w_1 x_{21} + w_2 x_{22} \\ . \\ . \\ w_n x_{21} + w_2 x_{n2} \end{bmatrix}}_{z}$$

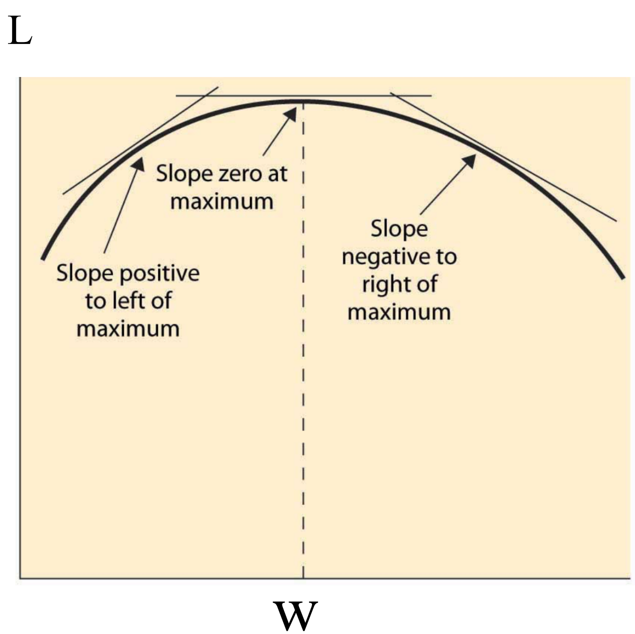
$$z = Xw$$

$$\mathrm{var}(z) = (Xw)^T Xw$$

$$\mathrm{var}(z) = w^T X^T X w = w^T S w$$

Choose w to maximize $w^T S w$
subject to $w^T w = 1$

The Direction of PC1 Is the Eigen Vector with the Highest Eigen Value



Choose w to maximize $w^T S w$
subject to $w^T w = 1$

$$L(w, \lambda) = w^T S w - \lambda(w^T w - 1)$$

$$\frac{\partial L}{\partial \mathbf{w}} = 2Sw - 2\lambda w$$

$$Sw = \lambda w$$

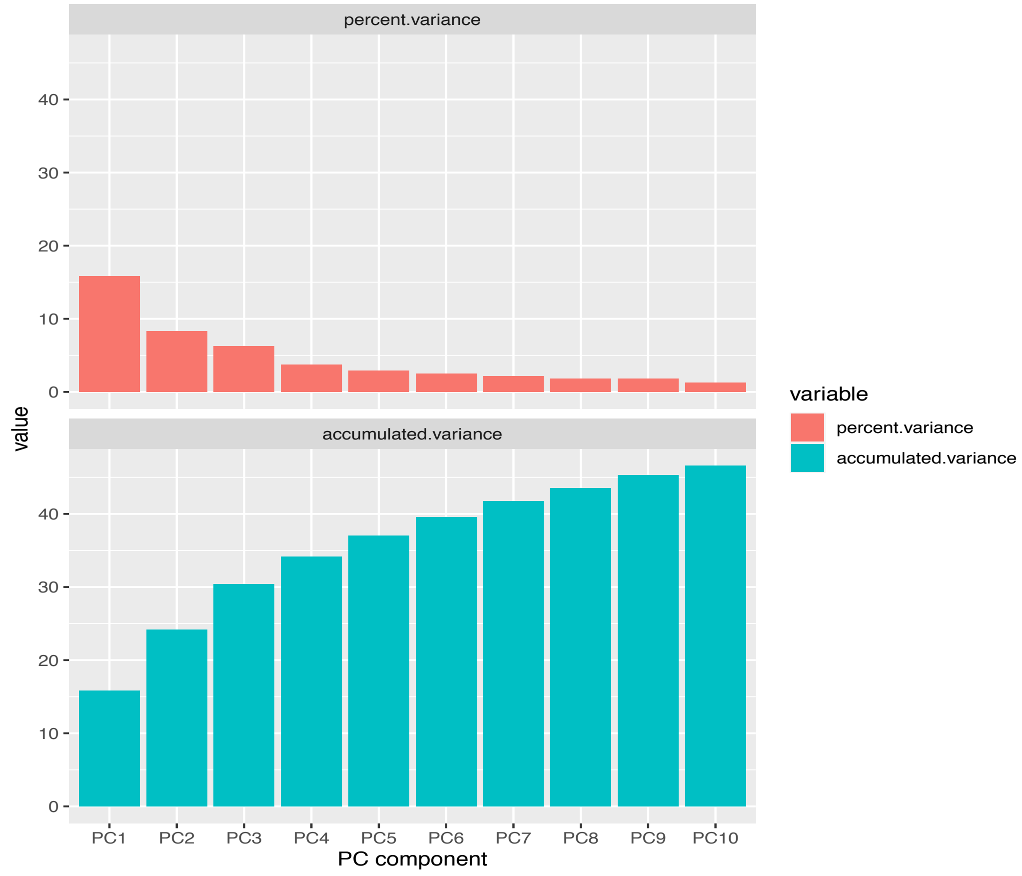
w is the eigen vector and λ is eigen value

Variance of PCs Are Eigen Value and Are Additive

$$\begin{aligned} var(z) &= w^T S w \\ &= w^T \lambda w \\ &= \lambda \end{aligned}$$

There are p pairs of eigen vectors and eigen values

$$var(Z) = \lambda_1 + \lambda_2 \ldots + \lambda_p$$



Singular Value Decomposition (SVD)

$$Z = XW$$

$$Z_s = XW\Lambda^{-1/2}$$

$$Z_s\Lambda^{1/2}W^T = X$$

$$X = Z_s\Lambda^{1/2}W^T$$

$X = U\Sigma V^T$ from standard SVD notation

Eugenio Beltrami and Camille Jordan, 1873-1874

Right and Left Singular Vectors of SVD

p column vectors
n row vectors

$$
\begin{bmatrix}
x_{11} & x_{12} & \ldots & x_{1p} \\
x_{21} & x_{22} & \ldots & x_{2p} \\
. & & & . \\
. & & & . \\
x_{n1} & x_{n2} & \ldots & x_{np}
\end{bmatrix}
$$

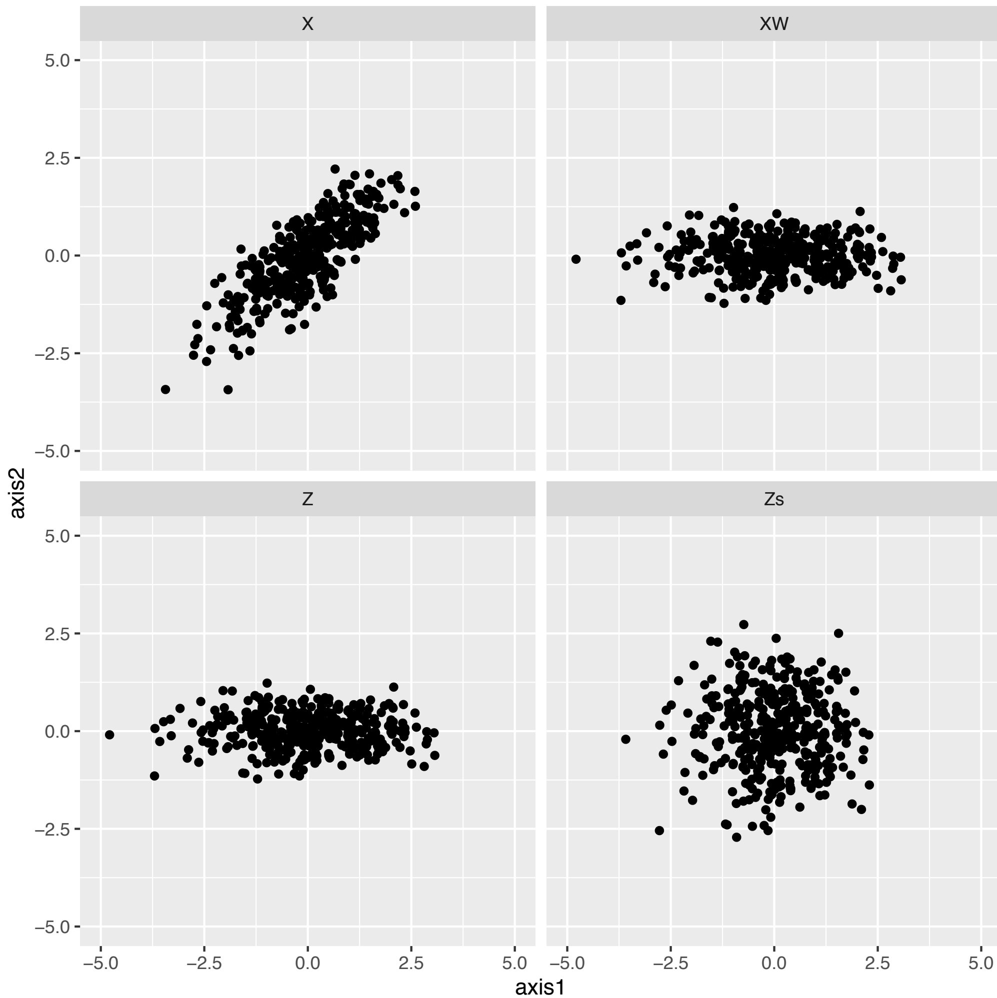
$$X = U\Sigma V^T$$

$$X^TX = V\Sigma U^TU\Sigma V^T$$
$$= V\Sigma^2V^T$$

$$XX^T = U\Sigma V^TV\Sigma U^T$$
$$= U\Sigma^2U^T$$

Interconversion Between Principal Components and Standardized Principal Components

$$Z_s = XWD^{-1/2}$$



$$Z = Z_s D^{1/2}$$

Three Dimensional Object and Its Two dimensional Image: Ellipse As Shadow of Sphere

Europe Map on a Globe and on Google Map

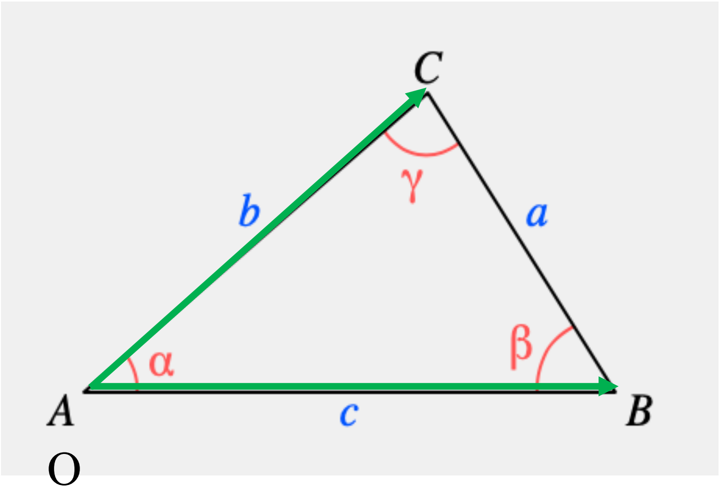
Eight European Cities Pairwise Distance Matrix

	Athens	Berlin	Dublin	London	Madrid	Paris	Rome	Warsaw
Athens	0	1119	1777	1486	1475	1303	646	1013
Berlin	1119	0	817	577	1159	545	736	327
Dublin	1777	817	0	291	906	489	1182	1135
London	1486	577	291	0	783	213	897	904
Madrid	1475	1159	906	783	0	652	856	1483
Paris	1303	545	489	213	652	0	694	859
Rome	646	736	1182	897	856	694	0	839
Warsaw	1013	327	1135	904	1483	859	839	0

Law of Cosines:
Dot Product of Two Vectors Can Be Expressed By the Difference Between the Squared Distances

Law of cosine

$$a^2 = b^2 + c^2 - 2bc\cos(\alpha)$$



$$2bc\cos(\alpha) = b^2 + c^2 - a^2$$

$$bc\cos(\alpha) = -\tfrac{1}{2}(a^2 - b^2 + c^2)$$

$$\mathbf{b \cdot c} = bc\cos(\alpha)$$

$$\mathbf{b \cdot c} = -1/2(a^2 - b^2 - c^2)$$

Warren Torgerson, 1958

Eigen Decomposition Can Generate Projection Map From Kernel Matrix Derived From the Pairwise Distance Between Two Cities

$$\mathbf{b \cdot c} = -1/2(a^2 - b^2 - c^2)$$

$$\begin{bmatrix} k_{11} & k_{12} & \ldots & k_{1n} \\ k_{21} & k_{22} & \ldots & k_{2n} \\ . & & & . \\ . & & & . \\ k_{n1} & k_{n2} & \ldots & k_{nn} \end{bmatrix}$$

$$K = U \Lambda U^T$$

$$Z = U \Lambda^{1/2}$$

MDS and PCA Are Equivalent

$Z = U\Lambda^{1/2}$ from MDS

$X = U\Sigma V^T$

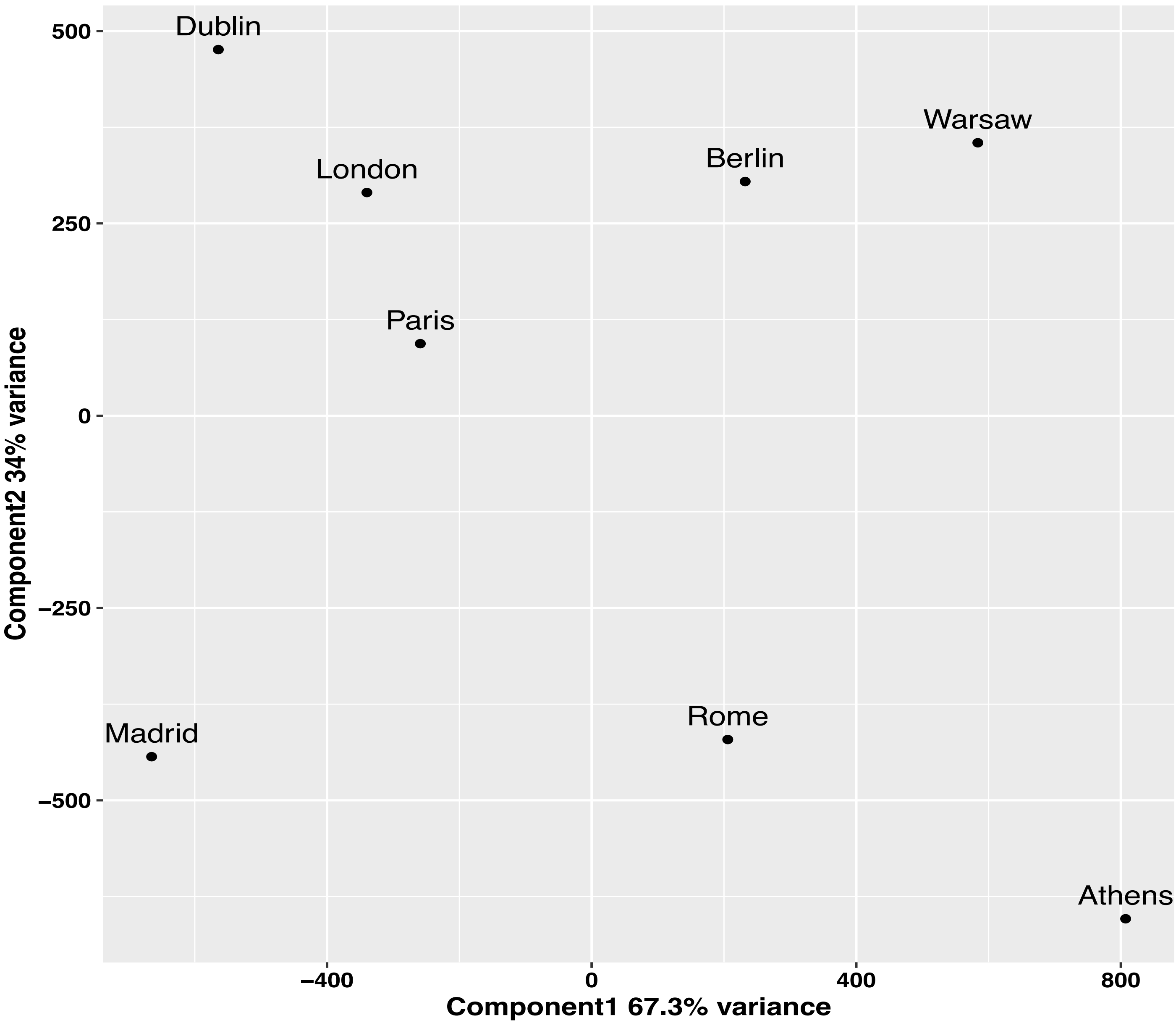
$XV = U\Sigma$

$Z = U\Lambda^{1/2}$

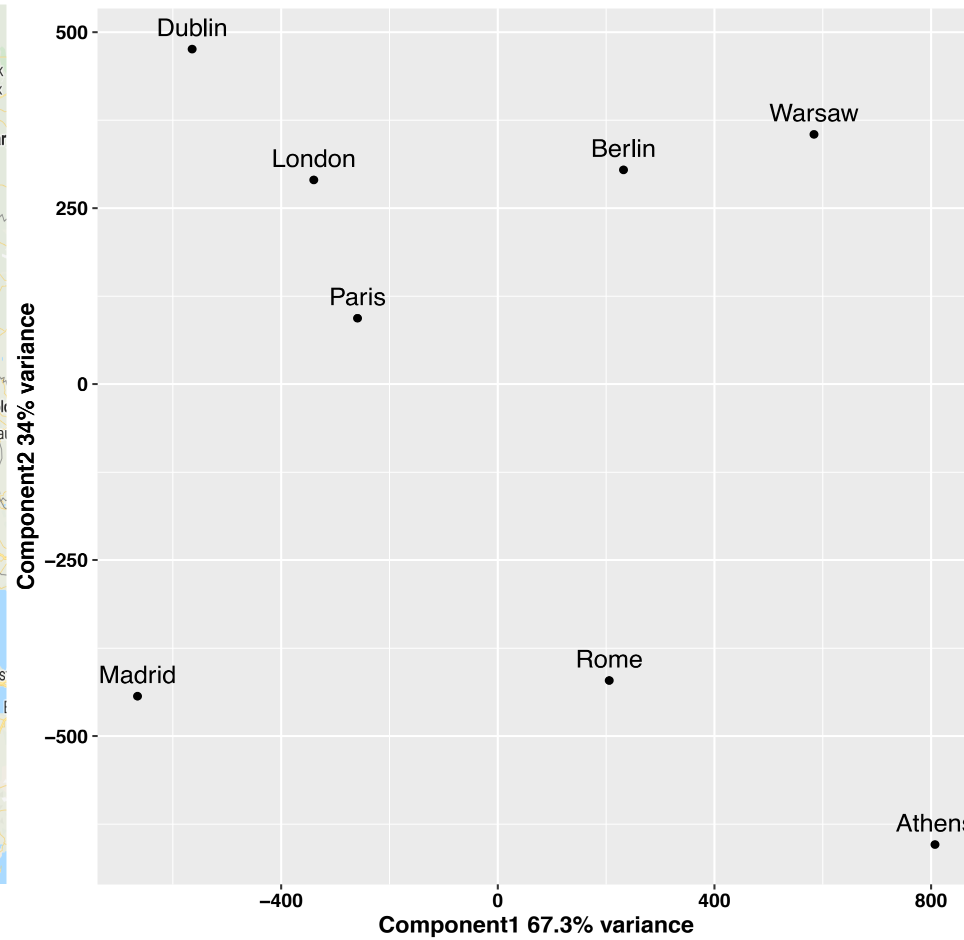
$K = U\Lambda U^T$

K was derived from the pairwise distance matrix without X

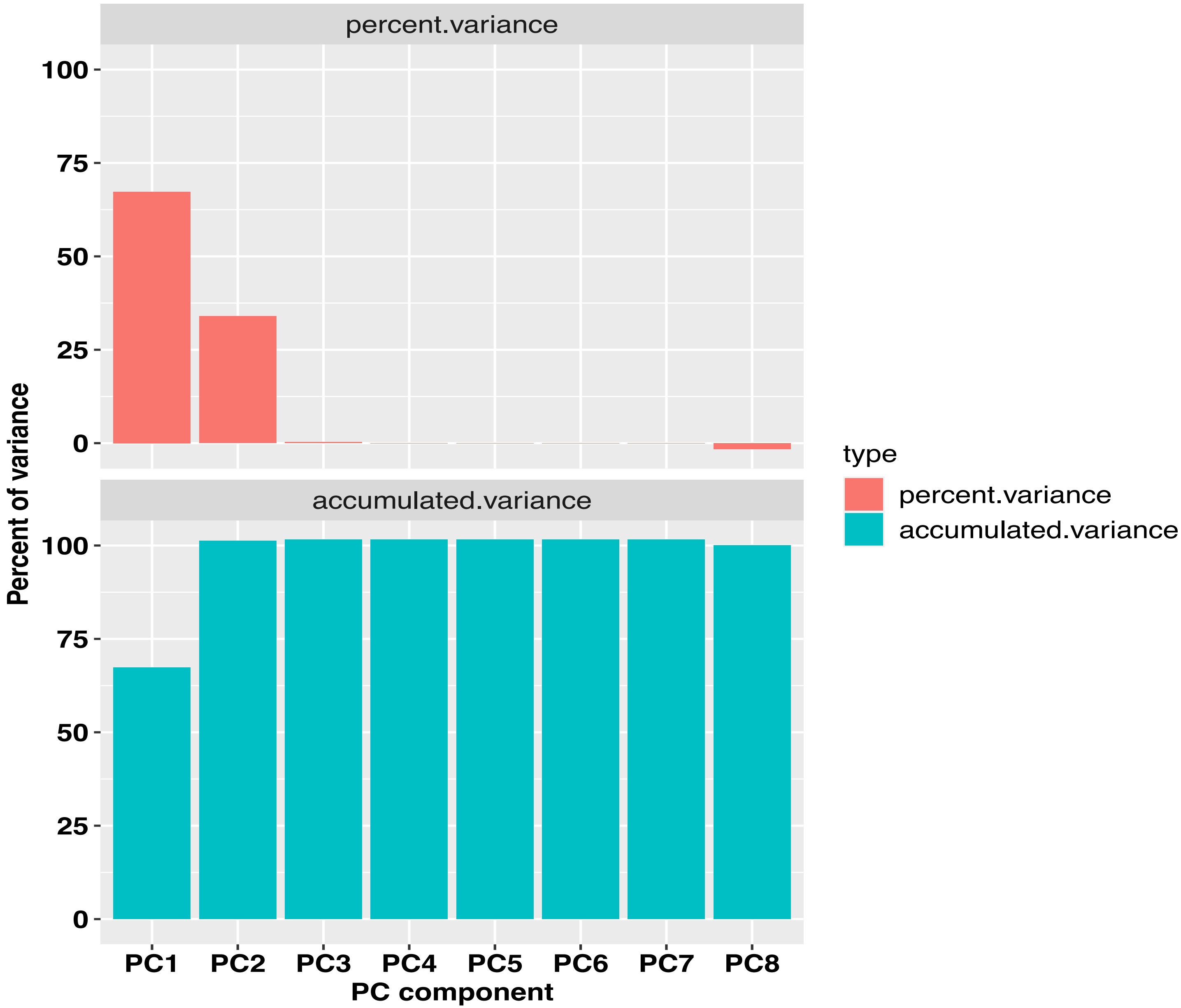
Two Dimensional Map Generated with MDS



Comparison Between Google Map and MDS Projection



Variance of MDS Components



Outline for Dimension Reduction Methods