Principal Component Analysis
Linear Dimension Reduction Technique

PCA performs a linear mapping of the data to a lower-dimensional space in such a way that the variance of the data in the low-dimensional representation is maximized.

1 Representations of Individuals

1. For the standardized data, $\frac{1}{n-1}\mathbf{X^T X}$ is the correlation matrix.
(If not standardized, $\frac{1}{n-1}\mathbf{X^T X}$ is the covariance matrix.)

2. **Loadings**: $\mathbf{V}$ is the $p \times r$ unit-norm matrix with eigenvectors of $\frac{1}{n-1}\mathbf{X^T X}$.
   - $\frac{1}{n-1}\mathbf{X^T X} = \mathbf{V D}^2 \mathbf{V^T} = \mathbf{V \Lambda V^T}$, where
   - Each $\mathbf{v_k}$ reflects how much each variable $\mathbf{X_j}$ loads in the PC k.

3. **Eigenvalues**
   - $\mathbf{D}$ is the $r \times r$ diagonal matrix of $\frac{1}{n-1}\mathbf{X}$ consisting of r positive eigenvalues (descending ordered).
   - $\mathbf{\Lambda}$ is the $r \times r$ diagonal matrix of $\frac{1}{n-1}\mathbf{X^T X}$, where each λ_k captures the projected inertia (i.e. variation) on each extracted dimension.
   - $\mathbf{\Lambda} = \mathbf{D}^2$.
   - r is the rank of $\mathbf{X}$.
   - When X is standardized, $\sum_{k=1}^{r} \lambda_k = p$, and the proportion of of variation captured by each PC is λ_k / p.

4. **Scores** or **Principal Components**: $\mathbf{Z}$ is the $n \times r$ matrix in which each column is the projection of all points on that axis/ PC.
   - $\mathbf{Z} = \mathbf{XV}$
   - PCs are linear combinations of the variables

2 Looking at Variables

1. To project the cloud of standardized variables, the projection of variable j onto an axis k, is equal to the cosine of the angle θ_{jk}, so this criteria maximizes $\sum_{i=1}^{p} cos^2(\theta_{jk}) = \sum_{i=1}^{p} corr^2(\mathbf{x_j}, \mathbf{z_k})$.

2. **Variable Factors**: $\mathbf{Q}$ is the $p \times r$ matrix in which each column is a linear combination of the objects.

- Projection of variable j on axis H_1 generated by vector $\mathbf{u_1}$ is $\mathbf{x_j^T u_1}$.
- $\mathbf{Q} = \mathbf{X^T U}$

3. **Eigenvectors**: $\mathbf{U}$ is the $p \times k$ matrix.

- Each column $\mathbf{u_1}$ is an eigenvector of $\frac{1}{n-1}\mathbf{XX^T}$
- $\frac{1}{n-1}\mathbf{XX^T}$ is the matrix of inner products between individuals
- $\frac{1}{n-1}\mathbf{XX^T} = \mathbf{U\Lambda U^T} = \mathbf{UD}^2\mathbf{U^T}$.

3 Relationship between the representations of Individuals and Variables

1. $\frac{1}{\sqrt{n-1}}\mathbf{X} = \mathbf{UDV^T} \Rightarrow \frac{1}{\sqrt{n-1}}\mathbf{XV} = \mathbf{UDV^TV} = \mathbf{UD} \Rightarrow \mathbf{U} = \frac{1}{\sqrt{n-1}}\mathbf{XVD^{-1}}$.
Since $\mathbf{Z} = \mathbf{XV}$, $\mathbf{U} = \frac{1}{\sqrt{n-1}}\mathbf{ZD^{-1}}$

2. $\frac{1}{\sqrt{n-1}}\mathbf{X} = \mathbf{UDV^T} \Rightarrow \frac{1}{\sqrt{n-1}}\mathbf{U^TX} = \mathbf{U^TUDV^T} = \mathbf{DV^T} \Rightarrow$
$\mathbf{V} = \frac{1}{\sqrt{n-1}}\mathbf{X^TUD^{-1}} \Rightarrow \mathbf{V} = \frac{1}{\sqrt{n-1}}\mathbf{QD^{-1}}$.

3. $\mathbf{Z} = \mathbf{XV} = \mathbf{X}\frac{1}{\sqrt{n-1}}\mathbf{QD^{-1}} = \frac{1}{\sqrt{n-1}}\mathbf{XX^TUD^{-1}} = \frac{1}{\sqrt{n-1}}\mathbf{U\Lambda U^TU\Lambda^{-1/2}} =$
$\frac{1}{\sqrt{n-1}}\mathbf{U\Lambda\Lambda^{-1/2}} = \frac{1}{\sqrt{n-1}}\mathbf{U\Lambda^{1/2}}$.

4 Visualize PCA Results

There is a really good article providing practical guide for PCA.

You can also use the following link: http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials