Almost everything you need to know about PLS

Part 1: Background, Theory, and Examples

Jenny Rieck & Derek Beaton

October 24, 2017

The BIG outline

- Part 1: Background & Examples

The BIG outline

- Part 1: Background & Examples
  - RIGHT NOW

The BIG outline

- Part 1: Background & Examples
  - RIGHT NOW
  - Introduce everything we need

The BIG outline

- Part 1: Background & Examples
  - RIGHT NOW
  - Introduce everything we need
- Part 2: PLS in Matlab & R

The BIG outline

- Part 1: Background & Examples
  - RIGHT NOW
  - Introduce everything we need
- Part 2: PLS in Matlab & R
  - Tuesday November 21, 10:00-12:00 Worstman Hall

The BIG outline

- Part 1: Background & Examples
  - RIGHT NOW
  - Introduce everything we need
- Part 2: PLS in Matlab & R
  - Tuesday November 21, 10:00-12:00 Worstman Hall
  - Put knowledge into practice

Part 1 outline

- Theory & Background

Part 1 outline

- Theory & Background
  - Principal component analysis (PCA)

Part 1 outline

- Theory & Background
  - Principal component analysis (PCA)
  - Partial least squares (PLS)

Part 1 outline

- Theory & Background
  - Principal component analysis (PCA)
  - Partial least squares (PLS)
- Short break

Part 1 outline

- Theory & Background
  - Principal component analysis (PCA)
  - Partial least squares (PLS)
- Short break
- Examples (with ADNI)

Part 1 outline

- Theory & Background
  - Principal component analysis (PCA)
  - Partial least squares (PLS)
- Short break
- Examples (with ADNI)
  - Standard PLS (& inference)

Part 1 outline

- Theory & Background
  - Principal component analysis (PCA)
  - Partial least squares (PLS)
- Short break
- Examples (with ADNI)
  - Standard PLS (& inference)
  - Discriminant PLS

Part 1 outline

- Theory & Background
  - Principal component analysis (PCA)
  - Partial least squares (PLS)
- Short break
- Examples (with ADNI)
  - Standard PLS (& inference)
  - Discriminant PLS
  - “Seed” PLS

Part 1 outline

- Theory & Background
  - Principal component analysis (PCA)
  - Partial least squares (PLS)
- Short break
- Examples (with ADNI)
  - Standard PLS (& inference)
  - Discriminant PLS
  - “Seed” PLS
- And beyond!

Principal Component Analysis

Background

Where did PCA come from?

- Modern form

Where did PCA come from?

- Modern form
  - Hotelling (1933)

Where did PCA come from?

- Modern form
  - Hotelling (1933)
- Traces back to

Where did PCA come from?

- Modern form
    - Hotelling (1933)
- Traces back to
    - Cauchy (1829)

Where did PCA come from?

- Modern form
  - Hotelling (1933)
- Traces back to
  - Cauchy (1829)
  - Galton (1859)

Where did PCA come from?

- Modern form
  - Hotelling (1933)
- Traces back to
  - Cauchy (1829)
  - Galton (1859)
  - Pearson (1901)

Where did PCA come from?

- Modern form
  - Hotelling (1933)
- Traces back to
  - Cauchy (1829)
  - Galton (1859)
  - Pearson (1901)
- Modern overviews

Where did PCA come from?

- Modern form
  - Hotelling (1933)
- Traces back to
  - Cauchy (1829)
  - Galton (1859)
  - Pearson (1901)
- Modern overviews
  - S. Wold et al., (1987)

Where did PCA come from?

- Modern form
  - Hotelling (1933)
- Traces back to
  - Cauchy (1829)
  - Galton (1859)
  - Pearson (1901)
- Modern overviews
  - S. Wold et al., (1987)
  - Jollife (2002)

Where did PCA come from?

- Modern form
  - Hotelling (1933)
- Traces back to
  - Cauchy (1829)
  - Galton (1859)
  - Pearson (1901)
- Modern overviews
  - S. Wold et al., (1987)
  - Jollife (2002)
  - Abdi & Williams (2010)

What is PCA?

- Visualize high dimensional data

What is PCA?

- Visualize high dimensional data
- Orthogonal transformation

What is PCA?

- Visualize high dimensional data
- Orthogonal transformation
- Dimensionality reduction

What is PCA?

- Find “components”

What is PCA?

- Find “components”
  - Components are new variables that are combinations of old variables

What is PCA?

- Find "components"
  - Components are new variables that are combinations of old variables
- Components explain maximum possible variance

What is PCA?

- Find “components”
  - Components are new variables that are combinations of old variables
- Components explain maximum possible variance
  - Conditional to orthogonality

Visual example

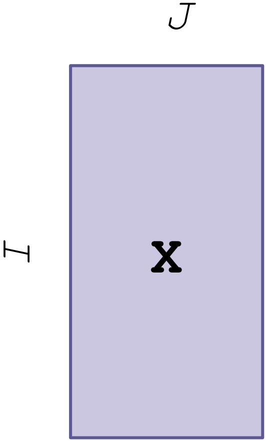
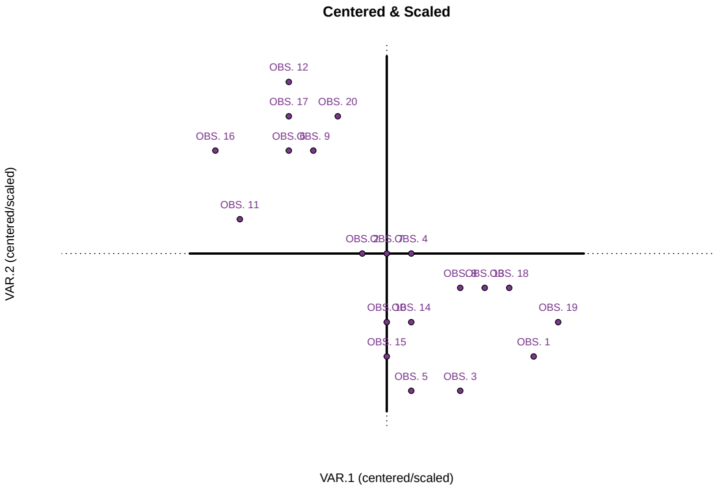


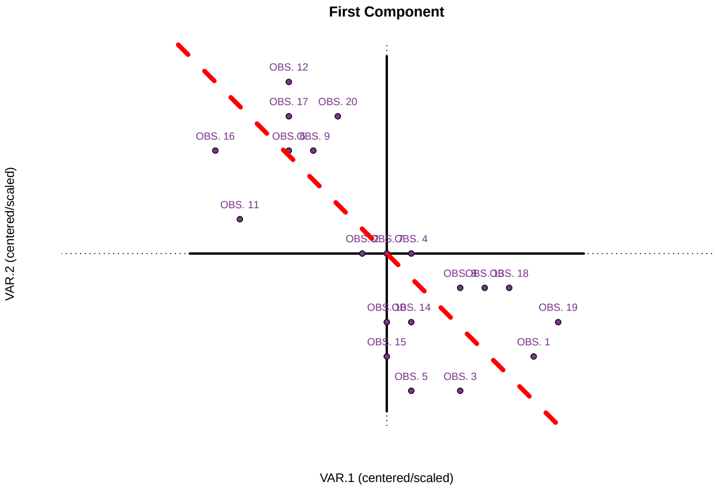
Figure 1: The kind of data we usually expect for PCA

Visual example

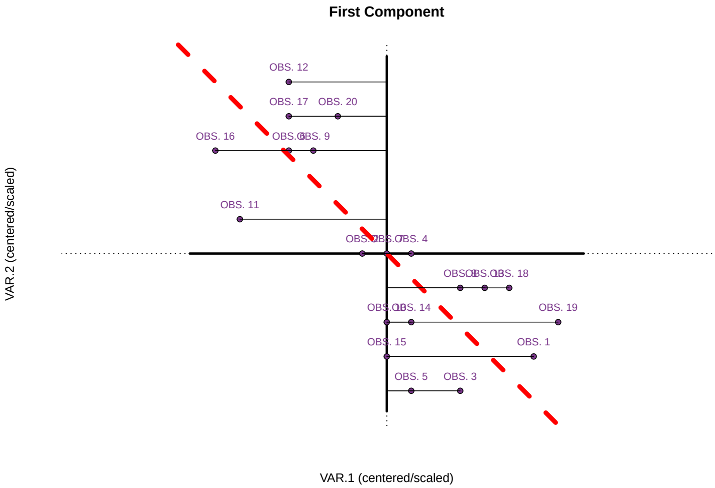
Visual example



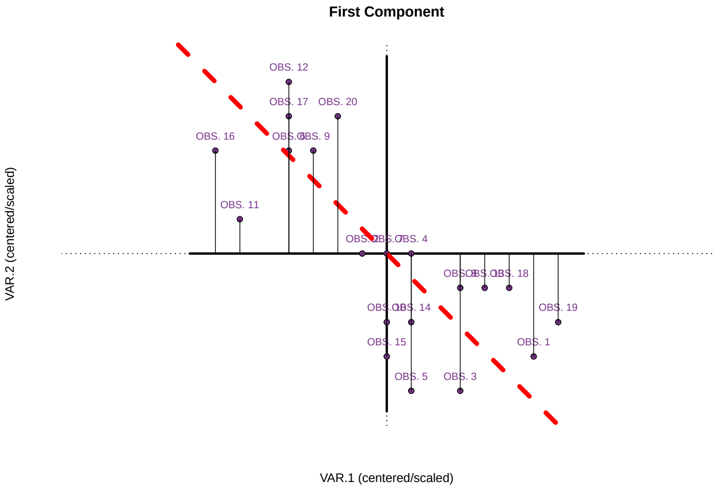
Visual example



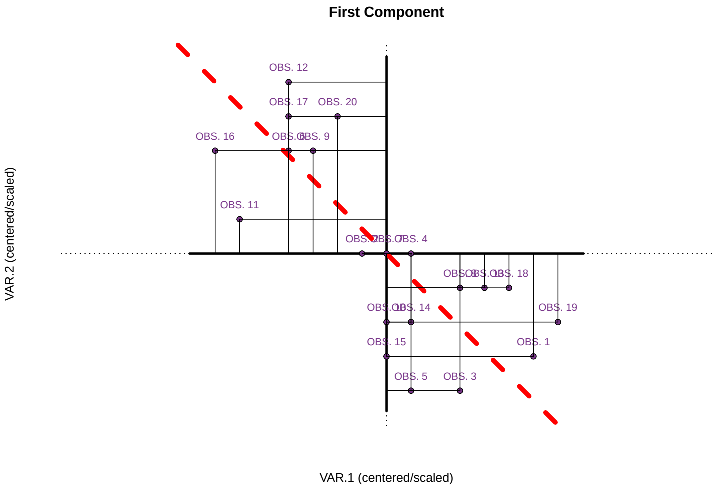
Visual example



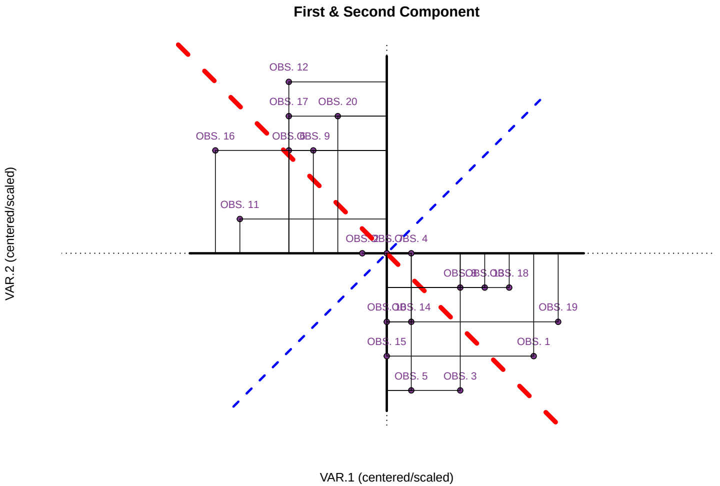
Visual example



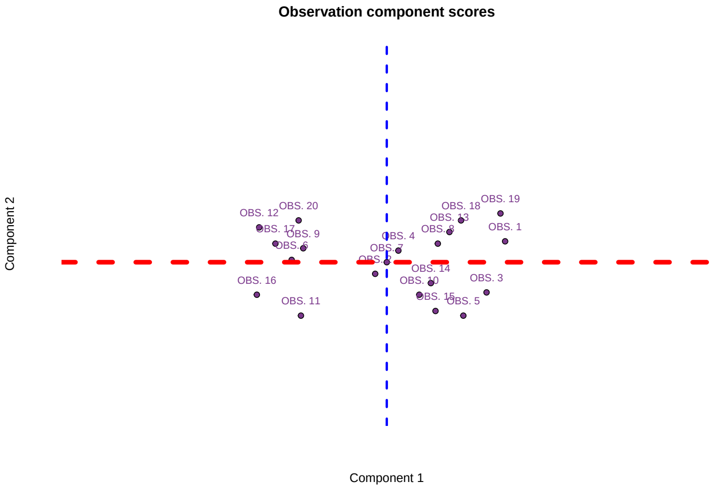
Visual example



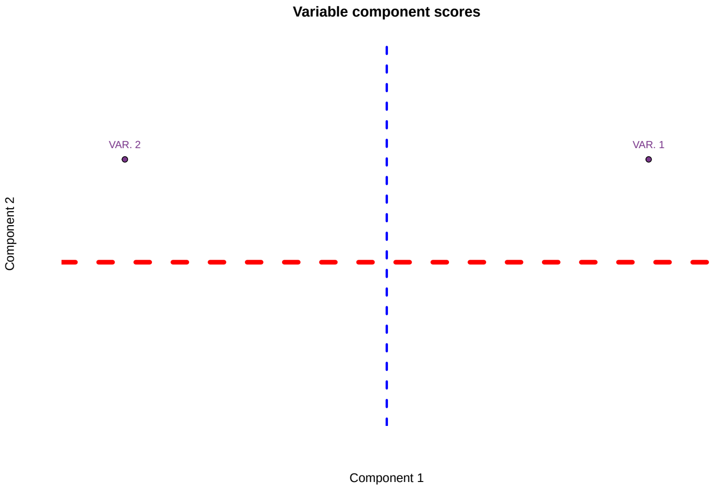
Visual example



Visual example



Visual example



What is PCA?

- The basis of modern techniques

What is PCA?

- The basis of modern techniques
  - Factor analyses

What is PCA?

- The basis of modern techniques
  - Factor analyses
  - Independent components analysis

What is PCA?

- The basis of modern techniques
  - Factor analyses
  - Independent components analysis
  - Partial least squares

What is PCA?

- The basis of modern techniques
  - Factor analyses
  - Independent components analysis
  - Partial least squares
  - Discriminant analyses

What is PCA?

- The basis of modern techniques
  - Factor analyses
  - Independent components analysis
  - Partial least squares
  - Discriminant analyses
  - Multi-table (e.g., MFA, GCCA)

What is PCA?

- A special case of the singular value decomposition (SVD)

What is PCA?

- A special case of the singular value decomposition (SVD)
- Which means (almost) everything else is, too

Formalization

Singular value decomposition

The SVD is one of the most ubiquitous and important tools

SVD

- We'll go into *enough* formalization

SVD

- We'll go into *enough* formalization
- If you want more:

SVD

- We'll go into *enough* formalization
- If you want more:
    - S. Wold et al., (1987)

SVD

- We'll go into *enough* formalization
- If you want more:
    - S. Wold et al., (1987)
    - Jollife (2002)

SVD

- We'll go into *enough* formalization
- If you want more:
    - S. Wold et al., (1987)
    - Jollife (2002)
    - Abdi & Williams (2010)

SVD

- We'll go into *enough* formalization
- If you want more:
  - S. Wold et al., (1987)
  - Jollife (2002)
  - Abdi & Williams (2010)
  - And many others. . .

SVD

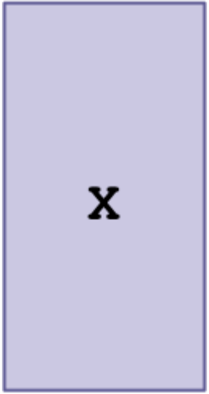


Figure 2: The shape of the data

SVD

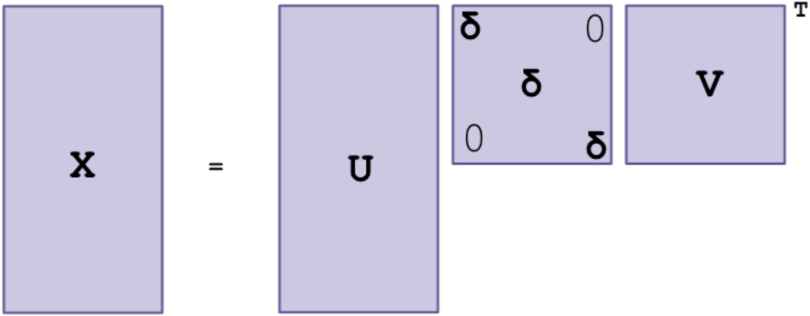


Figure 3: SVD breaks down the data

SVD

Notation

SVD

Notation

- x - a scalar

SVD

Notation

- x - a scalar
- $\mathbf{a}$ - a vector

SVD

Notation

- x - a scalar
- $\mathbf{a}$ - a vector
- $\mathbf{A}$ - a matrix

SVD

Notation

- x - a scalar
- $\mathbf{a}$ - a vector
- $\mathbf{A}$ - a matrix
- $\mathbf{A}^T$ - transpose

SVD

Notation

- x - a scalar
- $\mathbf{a}$ - a vector
- $\mathbf{A}$ - a matrix
- $\mathbf{A}^T$ - transpose
- $\mathbf{AB}$ - multiplication

SVD

- Think back to PCA

SVD

- Think back to PCA
- We want to find *the* principal component

SVD

- Think back to PCA
- We want to find *the* principal component
    - a.k.a. maximum source of variance

SVD

Given a matrix $\mathbf{X}$ we generally assume that

SVD

Given a matrix $\mathbf{X}$ we generally assume that

- column-wise centered

SVD

Given a matrix $\mathbf{X}$ we generally assume that

- column-wise centered
- column-wise scaled (e.g., z-scores or sums of squares $= 1$)

SVD

Given $\mathbf{X}$ of size $I \times J$

SVD

Given $\mathbf{X}$ of size $I \times J$

We want to find vectors

SVD

Given $\mathbf{X}$ of size $I \times J$

We want to find vectors

- $\mathbf{u}$ of size $I \times 1$
- $\mathbf{v}$ of size $J \times 1$

SVD

Given $\mathbf{X}$ of size $I \times J$

We want to find vectors

- $\mathbf{u}$ of size $I \times 1$
- $\mathbf{v}$ of size $J \times 1$

such that

$$\underset{\mathbf{u},\mathbf{v}}{\operatorname{argmax}} \delta = \mathbf{u}^T\mathbf{X}\mathbf{v} \text{ conditional to } \mathbf{u}^T\mathbf{u} = 1 = \mathbf{v}^T\mathbf{v}$$

SVD

Which gives us the following equivalencies:

- $\mathbf{Xv} = \mathbf{u}\delta$
- $\mathbf{X}^T\mathbf{u} = \mathbf{v}\delta$

SVD

Which gives us the following equivalencies:

- $\mathbf{X}\mathbf{v} = \mathbf{u}\delta$
- $\mathbf{X}^T\mathbf{u} = \mathbf{v}\delta$

where

- $\mathbf{X}_1 = \delta\mathbf{u}\mathbf{v}^T$

SVD

Which gives us the following equivalencies:

- $\mathbf{Xv} = \mathbf{u}\delta$
- $\mathbf{X}^T\mathbf{u} = \mathbf{v}\delta$

where

- $\mathbf{X}_1 = \delta\mathbf{uv}^T$
  - $\mathbf{X}_1$ is $\mathbf{X}$ as represented by *source of maximum variance*

SVD

Given $\mathbf{X}_1 = \delta \mathbf{p}\mathbf{q}^T$ we can find the other components

SVD

Given $\mathbf{X}_1 = \delta \mathbf{p}\mathbf{q}^T$ we can find the other components

- $\mathbf{X}_{2:L} = \mathbf{X} - \mathbf{X}_1$
- $\mathbf{X}_{2:L}$ is orthogonal to $\mathbf{X}_1$

SVD

Given $\mathbf{X}_1 = \delta \mathbf{p}\mathbf{q}^T$ we can find the other components

- $\mathbf{X}_{2:L} = \mathbf{X} - \mathbf{X}_1$
- $\mathbf{X}_{2:L}$ is orthogonal to $\mathbf{X}_1$

$$\underset{\mathbf{p},\mathbf{q}}{\text{argmax}} \delta = \mathbf{p}^T \mathbf{X}_{2:L} \mathbf{q} \text{ conditional to } \mathbf{p}^T \mathbf{p} = 1 = \mathbf{q}^T \mathbf{q}$$

SVD

The SVD of $\mathbf{X}$ of size $I \times J$:

SVD

The SVD of $\mathbf{X}$ of size $I \times J$:

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}} \tag{1}$$

SVD

The SVD of $\mathbf{X}$ of size $I \times J$:

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathsf{T}} \tag{1}$$

$\mathbf{U}$ and $\mathbf{V}$ are orthonormal such that

SVD

The SVD of $\mathbf{X}$ of size $I \times J$:

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathsf{T}} \tag{1}$$

$\mathbf{U}$ and $\mathbf{V}$ are orthonormal such that

$$\mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathsf{T}}\mathbf{V} \tag{2}$$

SVD

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathsf{T}} \text{ such that } \mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathsf{T}}\mathbf{V}$$

SVD

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}} \text{ such that } \mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathbf{T}}\mathbf{V}$$

- where rank is L (i.e., number of extractable components)

SVD

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathsf{T}} \text{ such that } \mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathsf{T}}\mathbf{V}$$

- where rank is L (i.e., number of extractable components)
- $\mathbf{U}$ is $I \times L$ (left singular vectors; rows of $\mathbf{X}$)

SVD

$\mathbf{X} = \mathbf{U\Delta V^T}$ such that $\mathbf{U^T U} = \mathbf{I} = \mathbf{V^T V}$

- where rank is L (i.e., number of extractable components)
- $\mathbf{U}$ is $I \times L$ (left singular vectors; rows of $\mathbf{X}$)
- $\mathbf{V}$ is $J \times L$ (right singular vectors; columns of $\mathbf{X}$)

SVD

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}} \text{ such that } \mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathbf{T}}\mathbf{V}$$

SVD

$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}}$ such that $\mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathbf{T}}\mathbf{V}$

- $\boldsymbol{\Delta}$ is $L \times L$ diagonal matrix

SVD

$\mathbf{X} = \mathbf{U}\mathbf{\Delta}\mathbf{V}^{\mathsf{T}}$ such that $\mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathsf{T}}\mathbf{V}$

- $\mathbf{\Delta}$ is $L \times L$ diagonal matrix
- $\mathrm{diag}\{\mathbf{\Delta}\} = \boldsymbol{\delta}$ are singular values (decreasing)

SVD

$\mathbf{X} = \mathbf{U\Delta V^T}$ such that $\mathbf{U^TU} = \mathbf{I} = \mathbf{V^TV}$

- $\mathbf{\Delta}$ is $L \times L$ diagonal matrix
- $\mathrm{diag}\{\mathbf{\Delta}\} = \boldsymbol{\delta}$ are singular values (decreasing)
- $\boldsymbol{\lambda} = \boldsymbol{\delta}^2$ are the eigenvalues (variance)

SVD

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathsf{T}} \text{ such that } \mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathsf{T}}\mathbf{V}$$

SVD

$\mathbf{X} = \mathbf{U\Delta V^T}$ such that $\mathbf{U^T U} = \mathbf{I} = \mathbf{V^T V}$

- $\mathbf{X}_1 = \mathbf{U}_1 \mathbf{\Delta}_1 \mathbf{V}_1^T$

SVD

$$\mathbf{X} = \mathbf{U\Delta V^T} \text{ such that } \mathbf{U^TU} = \mathbf{I} = \mathbf{V^TV}$$

- $\mathbf{X}_1 = \mathbf{U}_1\mathbf{\Delta}_1\mathbf{V}_1^T$
- $\mathbf{X}_{2:L} = \mathbf{U}_{2:L}\mathbf{\Delta}_{2:L}\mathbf{V}_{2:L}^T$

SVD

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}} \text{ such that } \mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathbf{T}}\mathbf{V}$$

SVD

$\mathbf{X} = \mathbf{U\Delta V^T}$ such that $\mathbf{U^TU} = \mathbf{I} = \mathbf{V^TV}$

- $\mathbf{F}_I = \mathbf{U\Delta}$ (row component scores)

SVD

$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}}$ such that $\mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathbf{T}}\mathbf{V}$

- $\mathbf{F}_I = \mathbf{U}\boldsymbol{\Delta}$ (row component scores)
- $\mathbf{F}_J = \mathbf{V}\boldsymbol{\Delta}$ (column component scores)

SVD

$$\mathbf{X} = \mathbf{U\Delta V^T} \text{ such that } \mathbf{U^TU} = \mathbf{I} = \mathbf{V^TV}$$

SVD

$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}}$ such that $\mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathbf{T}}\mathbf{V}$

- $\mathbf{F}_I = \mathbf{U}\boldsymbol{\Delta} = \mathbf{X}\mathbf{V}$

SVD

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}} \text{ such that } \mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathbf{T}}\mathbf{V}$$

- $\mathbf{F}_I = \mathbf{U}\boldsymbol{\Delta} = \mathbf{X}\mathbf{V}$
- $\mathbf{F}_J = \mathbf{V}\boldsymbol{\Delta} = \mathbf{X}^T\mathbf{U}$

SVD

- Phew.

SVD

- Phew.
- Enough nerd stuff

SVD

- Phew.
- Enough nerd stuff
- Let's get back to PCA

Toy example

- In R with ExPosition packages

- In R with ExPosition packages
  - Sneak preview of newest version

- In R with ExPosition packages
  - Sneak preview of newest version
- https://github.com/derekbeaton/ExPosition-Family/

- In R with ExPosition packages
  - Sneak preview of newest version
- https://github.com/derekbeaton/ExPosition-Family/
- https://cran.r-project.org/web/packages/ExPosition/index.html

PCA Toy Dataset - Wine

- 36 different wines (e.g., USA red cab., CAN rose syrah)

PCA Toy Dataset - Wine

- 36 different wines (e.g., USA red cab., CAN rose syrah)
- 4 objective measures (e.g., “Alcohol”, “Acidity”)

PCA Toy Dataset - Wine

- 36 different wines (e.g., USA red cab., CAN rose syrah)
- 4 objective measures (e.g., “Alcohol”, “Acidity”)
- 9 subjective measures (e.g., “sweet”, “acidic”)

PCA Toy Dataset - Wine

- 36 different wines (e.g., USA red cab., CAN rose syrah)
- 4 objective measures (e.g., "Alcohol", "Acidity")
- 9 subjective measures (e.g., "sweet", "acidic")
    - We'll use this one for PCA

PCA Toy Dataset - Wine

- 36 different wines (e.g., USA red cab., CAN rose syrah)
- 4 objective measures (e.g., "Alcohol", "Acidity")
- 9 subjective measures (e.g., "sweet", "acidic")
  - We'll use this one for PCA
  - We'll use both for PLS

PCA Subjective Measures

	fruity	floral	vegetal	spicy
Chili_red_merlot	6	2	1	4
Chili_red_cabernet	5	3	2	3
Chili_red_shiraz	7	1	2	6
Canada_red_pinot	4	2	3	1
Canada_white_chardonnay	4	3	2	1
Canada_white_sauvignon	8	4	3	2
USA_rose_cabernet	8	3	3	3
USA_rose_pinot	6	1	1	2
USA_rose_syrah	9	3	2	5

PCA - Wine

```
pca.res <- epPCA(
              wine.subjective,
              scale = T,
              center = T,
              graphs = F)
```

PCA - Wine

- PCA maximizes variance

PCA - Wine

- PCA maximizes variance
- So let's visualize the variance per component

PCA - Wine

- PCA maximizes variance
- So let's visualize the variance per component
    - Use the eigenvalues

PCA - Wine

- PCA maximizes variance
- So let's visualize the variance per component
  - Use the eigenvalues
  - "Scree plot"

PCA - Wine

- Many ways to present the results

PCA - Wine

- Many ways to present the results
- We prioritize visualization

PCA - Wine

- Many ways to present the results
- We prioritize visualization
    - Use numbers and tables when needed

PCA - Wine

- Many ways to present the results
- We prioritize visualization
    - Use numbers and tables when needed
- PCA is half

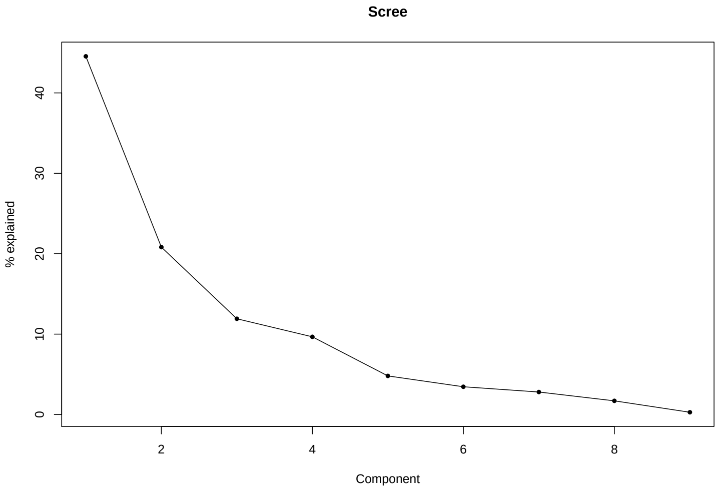
PCA - Wine

- Many ways to present the results
- We prioritize visualization
  - Use numbers and tables when needed
- PCA is half
  - Stats

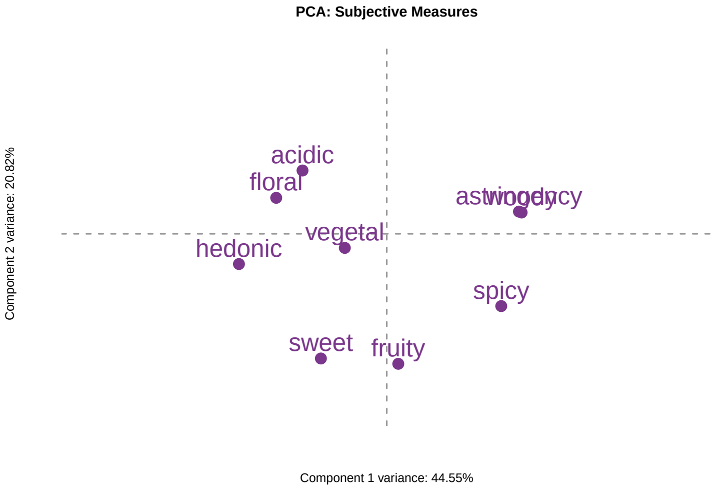
PCA - Wine

- Many ways to present the results
- We prioritize visualization
  - Use numbers and tables when needed
- PCA is half
  - Stats
  - Art

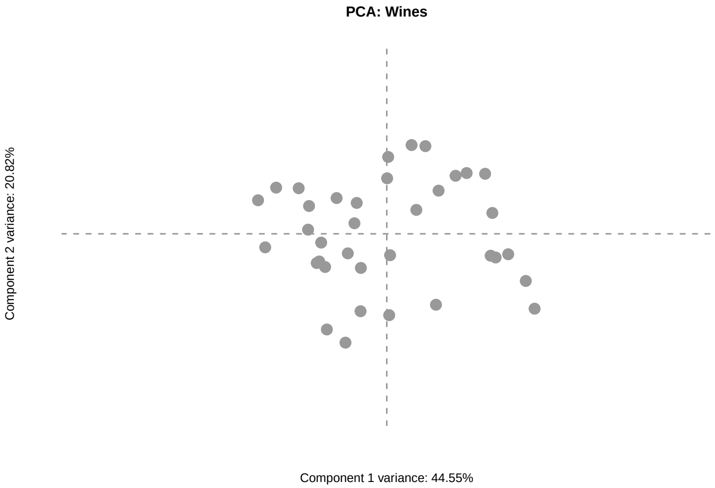
PCA - Wine



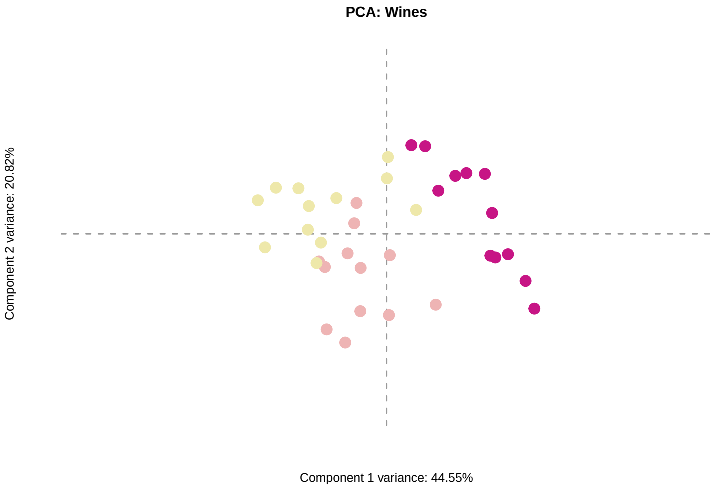
PCA - Wine


PCA: Subjective Measures
acidic
floral
astringency
woody
vegetal
hedonic
spicy
sweet
fruity
Component 2 variance: 20.82%
Component 1 variance: 44.55%

PCA - Wine


PCA: Wines
Component 2 variance: 20.82%
Component 1 variance: 44.55%

PCA - Wine


PCA: Wines
Component 2 variance: 20.82%
Component 1 variance: 44.55%

PCA - Wine

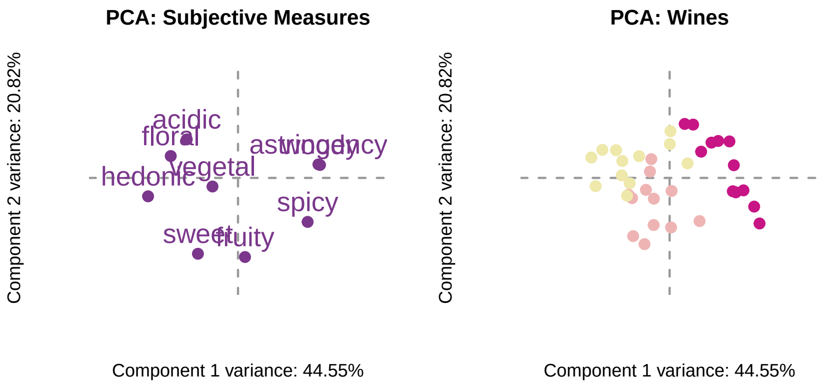


Figure 4: Variables & Observations

PCA

- If you know PCA you know about 90% of the multivariate stats in use

Partial least squares

Background

PLS means a lot of things

- Projection onto latent structures

PLS means a lot of things

- Projection onto latent structures
    - Probably the most accurate name

PLS means a lot of things

- Projection onto latent structures
    - Probably the most accurate name
    - But also probably too broad a definition

PLS means a lot of things

Partial least squares sounds like ordinary least squares

- When we have two matrices: $\mathbf{X}$ and $\mathbf{Y}$

PLS means a lot of things

Partial least squares sounds like ordinary least squares

- When we have two matrices: $\mathbf{X}$ and $\mathbf{Y}$
- OLS: $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$

PLS means a lot of things

Partial least squares sounds like ordinary least squares

- When we have two matrices: $\mathbf{X}$ and $\mathbf{Y}$
- OLS: $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$
- PLS: $\mathbf{X}^T\mathbf{Y}$

PLS means a lot of things

- Partial least squares path modelling (PLS-PM)

PLS means a lot of things

- Partial least squares path modelling (PLS-PM)
- Partial least squares regression (PLSR)

PLS means a lot of things

- Partial least squares path modelling (PLS-PM)
- Partial least squares regression (PLSR)
- Partial least squares correlation (PLSC)

PLS means a lot of things

- Partial least squares path modelling (PLS-PM)
- Partial least squares regression (PLSR)
- Partial least squares correlation (PLSC)
  - This is the one we'll talk about today

PLSC has many. . .

Names

- Inter-battery (factor) analysis (Tucker, 1958)

PLSC has many...

Names

- Inter-battery (factor) analysis (Tucker, 1958)
- Covariance between two fields (Bretherton, Smith, & Wallace, 1992)

PLSC has many. . .

Names

- Inter-battery (factor) analysis (Tucker, 1958)
- Covariance between two fields (Bretherton, Smith, & Wallace, 1992)
- PLS-SVD (Tenenhaus, 2005)

PLSC has many. . .

Names

- Inter-battery (factor) analysis (Tucker, 1958)
- Covariance between two fields (Bretherton, Smith, & Wallace, 1992)
- PLS-SVD (Tenenhaus, 2005)
- co-inertia analysis (Dray, 2014)

PLSC has many. . .

Friends

- Reduced Rank Regression

PLSC has many. . .

Friends

- Reduced Rank Regression
- Canonical Correlation Analysis

PLSC has many. . .

Friends

- Reduced Rank Regression
- Canonical Correlation Analysis
- (Fisher's) Linear Discriminant Analysis

PLSC has many. . .

Friends

- Reduced Rank Regression
- Canonical Correlation Analysis
- (Fisher's) Linear Discriminant Analysis
- PLS-correspondence analysis

PLSC

History

- McIntosh, Bookstein, Haxby, & Grady (1996)

PLSC

History

- McIntosh, Bookstein, Haxby, & Grady (1996)
- Bookstein (1992)

PLSC

History

- McIntosh, Bookstein, Haxby, & Grady (1996)
- Bookstein (1992)
- Tucker (1958)

PLSC

Modern overviews

- McIntosh & Lobaugh (2004)

PLSC

Modern overviews

- McIntosh & Lobaugh (2004)
- Krishnan et al., (2011)

PLSC

It's effectively just PCA applied to the cross product of two matrices measured on the same observations:

PLSC

It's effectively just PCA applied to the cross product of two matrices measured on the same observations:

- $\mathbf{X}$ which is $I \times J$

PLSC

It's effectively just PCA applied to the cross product of two matrices measured on the same observations:

- $\mathbf{X}$ which is $I \times J$
- $\mathbf{Y}$ which is $I \times K$

Formalization

PLSC via the SVD

Given $\mathbf{X}$ of size $I \times J$ and $\mathbf{Y}$ of size $I \times K$

PLSC via the SVD

Given $\mathbf{X}$ of size $I \times J$ and $\mathbf{Y}$ of size $I \times K$

We want to find vectors

PLSC via the SVD

Given $\mathbf{X}$ of size $I \times J$ and $\mathbf{Y}$ of size $I \times K$

We want to find vectors

- $\mathbf{u}$ of size $J \times 1$
- $\mathbf{v}$ of size $K \times 1$

To define latent variables

PLSC via the SVD

Given $\mathbf{X}$ of size $I \times J$ and $\mathbf{Y}$ of size $I \times K$

We want to find vectors

- $\mathbf{u}$ of size $J \times 1$
- $\mathbf{v}$ of size $K \times 1$

To define latent variables

- $\mathbf{l_X} = \mathbf{Xu}$ of size $I \times 1$
- $\mathbf{l_Y} = \mathbf{Yv}$ of size $I \times 1$

PLSC via the SVD

Given $\mathbf{X}$ of size $I \times J$ and $\mathbf{Y}$ of size $I \times K$

We want to find vectors

- $\mathbf{u}$ of size $J \times 1$
- $\mathbf{v}$ of size $K \times 1$

To define latent variables

- $\mathbf{l_X} = \mathbf{Xu}$ of size $I \times 1$
- $\mathbf{l_Y} = \mathbf{Yv}$ of size $I \times 1$

such that

$$\underset{\mathbf{u},\mathbf{v}}{\operatorname{argmax}} \delta = \mathbf{u}^T\mathbf{X}^T\mathbf{Y}\mathbf{v} \text{ conditional to } \mathbf{u}^T\mathbf{u} = 1 = \mathbf{v}^T\mathbf{v}$$

PLSC via the SVD

Compute the relationship between **X** and **Y**

$$\mathbf{R} = \mathbf{X}^{\mathsf{T}}\mathbf{Y} \tag{3}$$

PLSC via the SVD

Compute the relationship between $\mathbf{X}$ and $\mathbf{Y}$

$$\mathbf{R} = \mathbf{X}^{\mathbf{T}}\mathbf{Y} \tag{3}$$

Compute the SVD of $\mathbf{R}$

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}} \tag{4}$$

PLSC via the SVD

Compute the relationship between $\mathbf{X}$ and $\mathbf{Y}$

$$\mathbf{R} = \mathbf{X}^{\mathsf{T}}\mathbf{Y} \tag{3}$$

Compute the SVD of $\mathbf{R}$

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathsf{T}} \tag{4}$$

Compute the latent variables

$$\mathbf{L_X} = \mathbf{XU} \text{ and } \mathbf{L_Y} = \mathbf{YV} \tag{5}$$

PLSC via the SVD

Almost everything is the same:

- $\boldsymbol{\Delta}$ are singular values

PLSC via the SVD

Almost everything is the same:

- $\mathbf{\Delta}$ are singular values
- $\mathbf{F}_J = \mathbf{U}\mathbf{\Delta}$ (component scores for variables of $\mathbf{X}$)

PLSC via the SVD

Almost everything is the same:

- $\mathbf{\Delta}$ are singular values
- $\mathbf{F}_J = \mathbf{U\Delta}$ (component scores for variables of $\mathbf{X}$)
- $\mathbf{F}_K = \mathbf{V\Delta}$ (component scores for variables of $\mathbf{Y}$)

PLSC via the SVD

Almost everything is the same:

- $\mathbf{\Delta}$ are singular values
- $\mathbf{F}_J = \mathbf{U}\mathbf{\Delta}$ (component scores for variables of $\mathbf{X}$)
- $\mathbf{F}_K = \mathbf{V}\mathbf{\Delta}$ (component scores for variables of $\mathbf{Y}$)
- So our nomeclature will align with PCA

PLSC via the SVD

The new-ness

- $\mathbf{L_X} = \mathbf{XU}$ express the individuals w.r.t. $\mathbf{X}$

PLSC via the SVD

The new-ness

- $\mathbf{L_X} = \mathbf{XU}$ express the individuals w.r.t. $\mathbf{X}$
- $\mathbf{L_Y} = \mathbf{YV}$ express the individuals w.r.t. $\mathbf{Y}$

PLSC via the SVD

The new-ness

- $\mathbf{L_X} = \mathbf{XU}$ express the individuals w.r.t. $\mathbf{X}$
- $\mathbf{L_Y} = \mathbf{YV}$ express the individuals w.r.t. $\mathbf{Y}$
- Not in PCA

PLSC via the SVD

The new-ness

- $\mathbf{L_X} = \mathbf{XU}$ express the individuals w.r.t. $\mathbf{X}$
- $\mathbf{L_Y} = \mathbf{YV}$ express the individuals w.r.t. $\mathbf{Y}$
- Not in PCA
  - We'll call these "latent variable scores"

PLSC via the SVD

Maximizes the latent variables

PLSC via the SVD

Maximizes the latent variables

$$\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}} = \boldsymbol{\Delta} \tag{6}$$

PLSC via the SVD

Maximizes the latent variables

$$\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}} = \boldsymbol{\Delta} \tag{6}$$

When expanded

$(\mathbf{XU})^{\mathbf{T}}(\mathbf{YV}) = \boldsymbol{\Delta}$

PLSC via the SVD

Maximizes the latent variables

$$\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}} = \mathbf{\Delta} \tag{6}$$

When expanded

$(\mathbf{XU})^{\mathbf{T}}(\mathbf{YV}) = \mathbf{\Delta}$

$\mathbf{U}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{YV} = \mathbf{\Delta}$

PLSC via the SVD

Maximizes the latent variables

$$\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}} = \mathbf{\Delta} \tag{6}$$

When expanded

$(\mathbf{XU})^{\mathbf{T}}(\mathbf{YV}) = \mathbf{\Delta}$

$\mathbf{U}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{YV} = \mathbf{\Delta}$

$\mathbf{U}^{\mathbf{T}}\mathbf{RV} = \mathbf{\Delta}$

PLSC via the SVD

Maximizes the latent variables

$$\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}} = \mathbf{\Delta} \tag{6}$$

When expanded

$(\mathbf{XU})^{\mathbf{T}}(\mathbf{YV}) = \mathbf{\Delta}$

$\mathbf{U}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{YV} = \mathbf{\Delta}$

$\mathbf{U}^{\mathbf{T}}\mathbf{RV} = \mathbf{\Delta}$

$\mathbf{U}^{\mathbf{T}}\mathbf{U}\mathbf{\Delta}\mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{\Delta}$

PLSC via the SVD

Maximizes the latent variables

$$\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}} = \boldsymbol{\Delta} \tag{6}$$

When expanded

$(\mathbf{XU})^{\mathbf{T}}(\mathbf{YV}) = \boldsymbol{\Delta}$

$\mathbf{U}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{YV} = \boldsymbol{\Delta}$

$\mathbf{U}^{\mathbf{T}}\mathbf{RV} = \boldsymbol{\Delta}$

$\mathbf{U}^{\mathbf{T}}\mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathbf{T}}\mathbf{V} = \boldsymbol{\Delta}$

because

$\mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} = \mathbf{V}^{\mathbf{T}}\mathbf{V}$

PLSC via the SVD

It's effectively just PCA with some new-ness:

- $\mathbf{\Delta}$ are singular values

PLSC via the SVD

It's effectively just PCA with some new-ness:

- $\mathbf{\Delta}$ are singular values
- $\mathbf{U}$ akin to loadings (for variables of $\mathbf{X}$)

PLSC via the SVD

It's effectively just PCA with some new-ness:

- $\mathbf{\Delta}$ are singular values
- $\mathbf{U}$ akin to loadings (for variables of $\mathbf{X}$)
- $\mathbf{V}$ akin to loadings (for variables of $\mathbf{Y}$)

PLSC via the SVD

It's effectively just PCA with some new-ness:

- $\mathbf{\Delta}$ are singular values
- $\mathbf{U}$ akin to loadings (for variables of $\mathbf{X}$)
- $\mathbf{V}$ akin to loadings (for variables of $\mathbf{Y}$)
- $\mathbf{F}_J = \mathbf{U\Delta}$ (component scores for variables of $\mathbf{X}$)

PLSC via the SVD

It's effectively just PCA with some new-ness:

- $\mathbf{\Delta}$ are singular values
- $\mathbf{U}$ akin to loadings (for variables of $\mathbf{X}$)
- $\mathbf{V}$ akin to loadings (for variables of $\mathbf{Y}$)
- $\mathbf{F}_J = \mathbf{U\Delta}$ (component scores for variables of $\mathbf{X}$)
- $\mathbf{F}_K = \mathbf{V\Delta}$ (component scores for variables of $\mathbf{Y}$)

PLSC via the SVD

It's effectively just PCA with some new-ness:

- $\mathbf{\Delta}$ are singular values
- $\mathbf{U}$ akin to loadings (for variables of $\mathbf{X}$)
- $\mathbf{V}$ akin to loadings (for variables of $\mathbf{Y}$)
- $\mathbf{F}_J = \mathbf{U\Delta}$ (component scores for variables of $\mathbf{X}$)
- $\mathbf{F}_K = \mathbf{V\Delta}$ (component scores for variables of $\mathbf{Y}$)
- $\mathbf{L}_\mathbf{X} = \mathbf{XU}$ express the individuals w.r.t. $\mathbf{X}$

PLSC via the SVD

It's effectively just PCA with some new-ness:

- $\mathbf{\Delta}$ are singular values
- $\mathbf{U}$ akin to loadings (for variables of $\mathbf{X}$)
- $\mathbf{V}$ akin to loadings (for variables of $\mathbf{Y}$)
- $\mathbf{F}_J = \mathbf{U\Delta}$ (component scores for variables of $\mathbf{X}$)
- $\mathbf{F}_K = \mathbf{V\Delta}$ (component scores for variables of $\mathbf{Y}$)
- $\mathbf{L_X} = \mathbf{XU}$ express the individuals w.r.t. $\mathbf{X}$
- $\mathbf{L_Y} = \mathbf{YV}$ express the individuals w.r.t. $\mathbf{Y}$

Example

PLS Toy Dataset - Wine

- 36 different wines (e.g., USA red cab., CAN rose syrah)

PLS Toy Dataset - Wine

- 36 different wines (e.g., USA red cab., CAN rose syrah)
- 4 objective measures (e.g., "Alcohol", "Acidity")

PLS Toy Dataset - Wine

- 36 different wines (e.g., USA red cab., CAN rose syrah)
- 4 objective measures (e.g., "Alcohol", "Acidity")
- 9 subjective measures (e.g., "sweet", "acidic")

PLS Subjective Measures

	fruity	floral	vegetal	spicy
Chili_red_merlot	6	2	1	4
Chili_red_cabernet	5	3	2	3
Chili_red_shiraz	7	1	2	6
Canada_red_pinot	4	2	3	1
Canada_white_chardonnay	4	3	2	1
Canada_white_sauvignon	8	4	3	2
USA_rose_cabernet	8	3	3	3
USA_rose_pinot	6	1	1	2
USA_rose_syrah	9	3	2	5

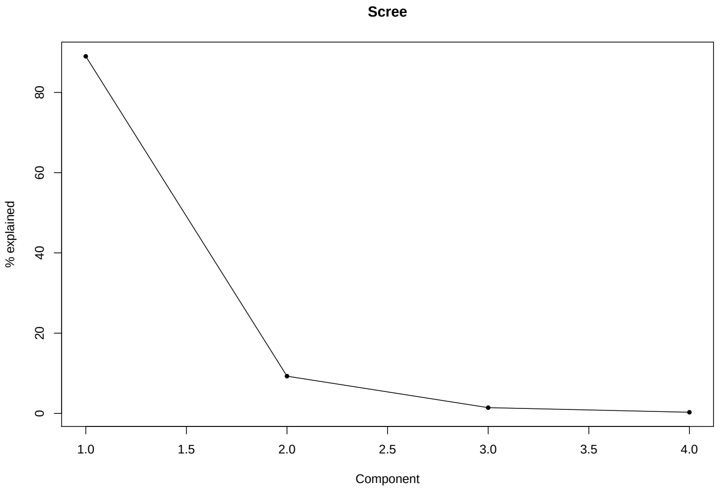
PLS Objective Measures

	Acidity	Alcohol	Sugar	Tanin
Chili_red_merlot	5.33	13.8	2.75	559
Chili_red_cabernet	5.14	13.9	2.41	672
Chili_red_shiraz	5.16	14.3	2.20	455
Canada_red_pinot	5.70	13.3	1.70	320
Canada_white_chardonnay	6.00	13.5	3.00	35
Canada_white_sauvignon	7.50	12.0	3.50	40
USA_rose_cabernet	5.71	12.5	4.30	93
USA_rose_pinot	5.40	13.0	3.10	79
USA_rose_syrah	6.50	13.5	3.00	89

PLS - Wine

```
pls.res <- tepPLS(
            DATA1 = wine.objective,
            center1 = T, scale1 = T,
            DATA2 = wine.subjective,
            center2 = T, scale2 = T,
            graphs=F
           )
```

PLS - Wine



PLS - Wine

While PLS is similar to PCA we should focus on what PLS maximizes:

PLS - Wine

While PLS is similar to PCA we should focus on what PLS maximizes:

$\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}}$

PLS - Wine

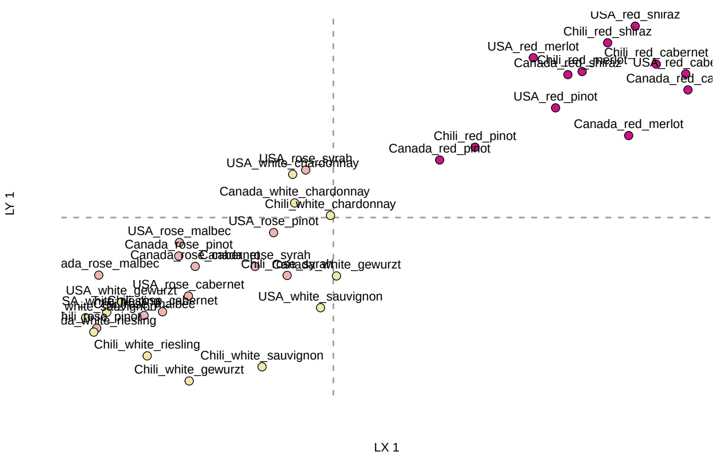
While PLS is similar to PCA we should focus on what PLS maximizes:

$\mathbf{L}_{\mathbf{X}}^{T}\mathbf{L}_{\mathbf{Y}}$

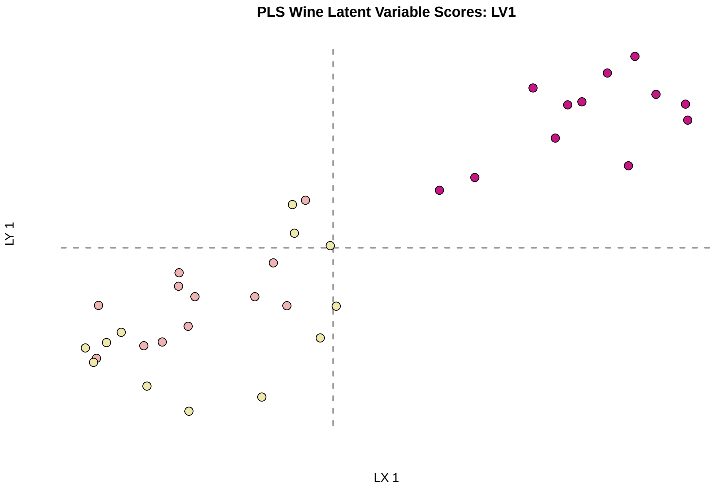
So we'll start with the latent variable scores

PLS - Wine

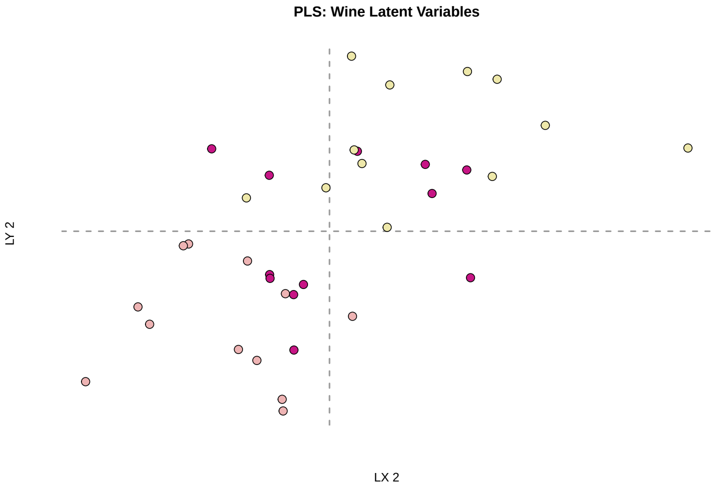
PLS Wine Latent Variable Scores: LV1



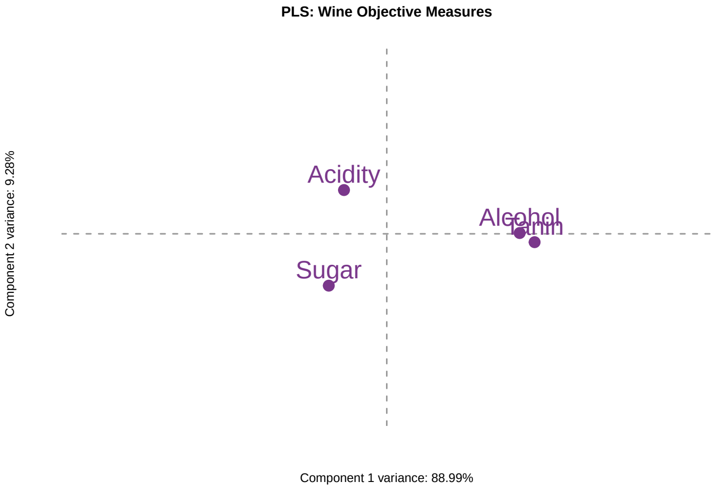
PLS - Wine



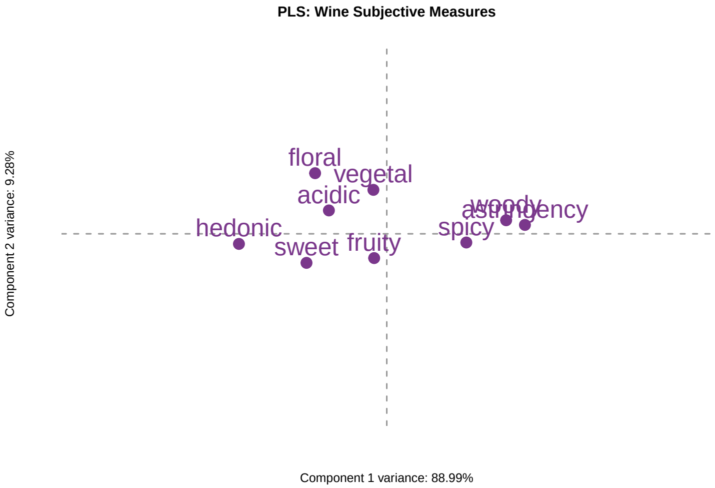
PLS - Wine



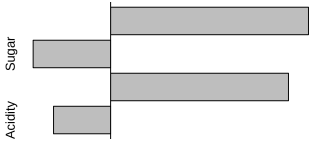
PLS - Wine

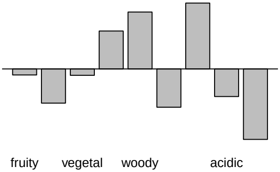

PLS: Wine Objective Measures
Acidity
Alcohol
Tanin
Sugar
Component 2 variance: 9.28%
Component 1 variance: 88.99%

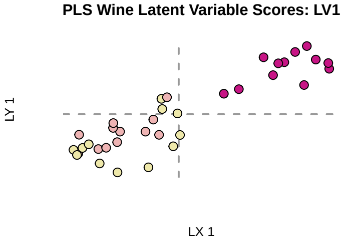
PLS - Wine


PLS: Wine Subjective Measures
floral
vegetal
acidic
hedonic
sweet
fruity
woody
astringency
spicy
Component 2 variance: 9.28%
Component 1 variance: 88.99%

PLS - Wine


Sugar
Acidity


fruity
vegetal
woody
acidic


PLS Wine Latent Variable Scores: LV1
LY 1
LX 1

Overview

We're experts now

- PCA is the core of PLS

We're experts now

- PCA is the core of PLS
- PLS generalizes PCA

We're experts now

- PCA is the core of PLS
- PLS generalizes PCA
    - If your two matrices are both the same, e.g., $\mathbf{X}$

We're experts now

- PCA is the core of PLS
- PLS generalizes PCA
  - If your two matrices are both the same, e.g., $\mathbf{X}$
  - PLS gives same results as PCA

We're experts now

- Visualize first

We're experts now

- Visualize first
- Use what you know to help construct the story from the numbers

We're experts now

- Visualize first
- Use what you know to help construct the story from the numbers
    - We'll see some additional helpers in next part today

Break!

We covered a lot!

- And we're not even at the good part yet!

We covered a lot!

- And we're not even at the good part yet!
- 7 ± 2 break

We covered a lot!

- And we're not even at the good part yet!
- 7 ± 2 break
    - Stretch, coffee

We covered a lot!

- And we're not even at the good part yet!
- 7 ± 2 break
    - Stretch, coffee
    - Pressing questions