Almost everything you need to know about PLS

Part 1: Background, Theory, and Examples

Jenny Rieck & Derek Beaton

October 24, 2017

PLSC Flavors

PLSC has many. . .

Flavors

- Standard
- Mean-Centered
- Seed
- And many more (not covered today)

Standard PLSC

A refresher

Standard PLSC

- computed as the SVD of the cross-product matrix of $\mathbf{X}$ and $\mathbf{Y}$
- $\mathbf{X}$ is a matrix of observations by variables
- $\mathbf{Y}$ is a matrix of observations by (some other) variables

as a reminder

- Today we will cover the most common variations in PLSC
- Terminology differs; we stick to PCA nomenclature
- Stay tuned for more in depth examples in R and Matlab

An example

PLSC dataset

ADNI ($N = 569$)

- 3 groups of participants
  - $N = 178$ healthy control
  - $N = 275$ late MCI
  - $N = 116$ AD
- 8 neuropsych measures
- 68 cortical thickness estimates (via Freesurfer)

Data matrices

$$\mathbf{X} = \begin{array}{c} \\ \text{Subj}_1 \\ \text{Subj}_2 \\ \text{Subj}_3 \\ \vdots \\ \text{Subj}_{N-1} \\ \text{Subj}_N \end{array} \begin{array}{c} \begin{array}{cccc} \text{BNT} & \text{Clock} & \ldots & \text{RAVLT} \end{array} \\ \begin{bmatrix} 10 & 5 & \ldots & 30 \\ 7 & 4 & \ldots & 26 \\ 3 & 0 & \ldots & 23 \\ \vdots & \vdots & & \vdots \\ 2 & 1 & \ldots & 18 \\ 8 & 4 & \ldots & 27 \end{bmatrix} \end{array} \qquad \mathbf{Y} = \begin{array}{c} \\ \text{Subj}_1 \\ \text{Subj}_2 \\ \text{Subj}_3 \\ \vdots \\ \text{Subj}_{N-1} \\ \text{Subj}_N \end{array} \begin{array}{c} \begin{array}{cccc} \text{R.IFG} & \text{L.IFG} & \ldots & \text{L.Fusi} \end{array} \\ \begin{bmatrix} 3.24 & 6.27 & \ldots & 2.32 \\ 5.89 & 0.26 & \ldots & 4.51 \\ 2.84 & 2.51 & \ldots & 1.17 \\ \vdots & \vdots & & \vdots \\ 1.96 & 8.9 & \ldots & 3.46 \\ 4.42 & 7.81 & \ldots & 1.96 \end{bmatrix} \end{array}$$

Figure 1: **X** and **Y** matrices in standard PLS

Standard PLSC scree

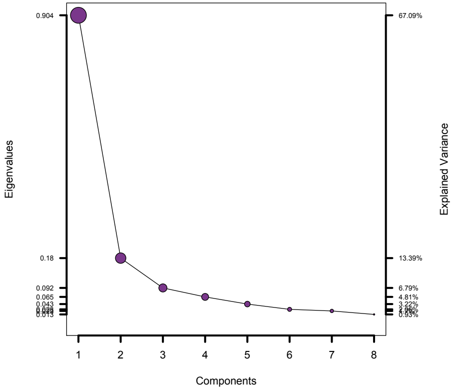


Figure 2

Beyond descriptive analyses

How many components are "significant"?

- Above expected (or average) contribution
- Permutation
    - scramble the rows of one matrix and recompute PLSC
    - create a null disribution of eigenvalues

For more details on component selection:

- Jackson (1993)
- Peres-Neto et al., (2005)
- Dray (2008)
- Josse and Husson (2011)

How many components to interpret?

- Again: a mix of art & science
- Use tests, effects sizes, and heuristics

How many components to interpret?

Component	p.value	eigenvalue	percent.variance
1	0.000	0.904	67.093
2	0.000	0.180	13.388
3	0.009	0.092	6.794
4	0.015	0.065	4.809
5	0.096	0.043	3.223
6	0.453	0.028	2.055
7	0.207	0.023	1.704
8	0.747	0.013	0.934

How many components to interpret?

- Note: there are no p-values of 0
- Zero means the observed value is outside the distribution
- But the distribution is from number of iterations
- Here: 1000
    - Thus $p = 0$ is actually $p < .001$

How many components to interpret?

Component	p.value	eigenvalue	percent.variance
1	0	0.904	67.093
2	0	0.18	13.388
3	0.009		
4	0.015		
5			
6			
7			
8			

Two reliable components

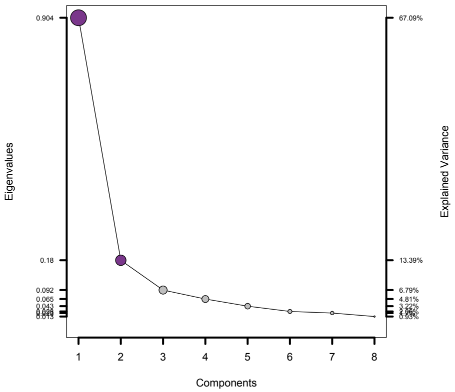


Figure 3

Latent Variables

Latent variables

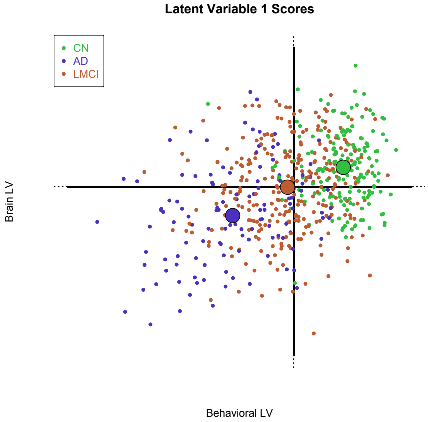


Figure 4

Latent variables

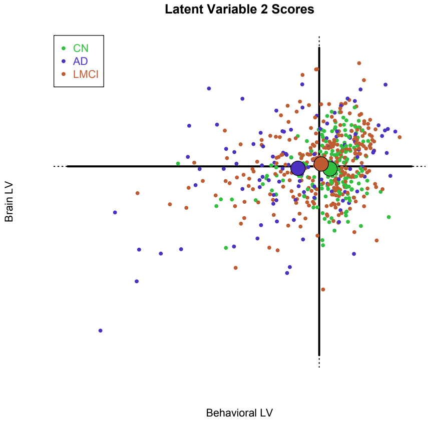


Figure 5

Component Scores

Neuropsych component scores

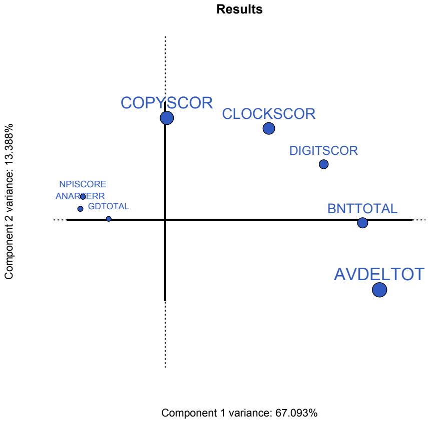


Figure 6

Structural thickness component scores

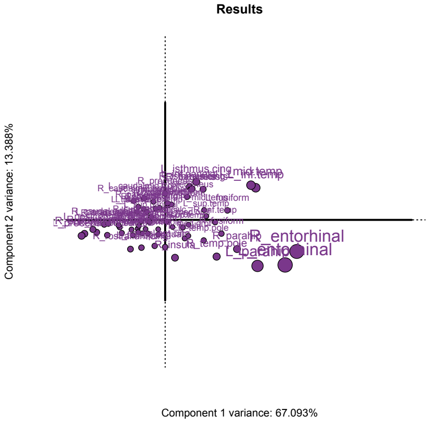


Figure 7

Beyond descriptive analyses

Which variables significantly contribute to each component?

- Bootstrap resampling (with replacement) builds distributions around our variables
  - Confidence intervals
  - Bootstrap ratios (mean divided by s.d. of distribution)

Which neuropsych variables significantly contribute?

Table 3: Bootstrap Ratios

	Component.1	Component.2
ANART.ERR	-2.85	0.57
COPY.SCOR	0.12	3.23
CLOCK.SCOR	3.53	3.72
DIGIT.SCOR	5.56	2.44
BNT.TOTAL	6.60	-0.10
AV.DEL.TOT	7.89	-3.50
GD.TOTAL	-1.85	0.02
NPI.SCORE	-2.58	0.80

Which neuropsych variables significantly contribute?

Table 4: Bootstrap Ratios $> +/- 2.5$

	Component.1	Component.2
ANART.ERR	-2.85	
COPY.SCOR		3.23
CLOCK.SCOR	3.53	3.72
DIGIT.SCOR	5.56	
BNT.TOTAL	6.6	
AV.DEL.TOT	7.89	-3.5
GD.TOTAL		
NPI.SCORE	-2.58	

Which neuropsych variables significantly contribute?

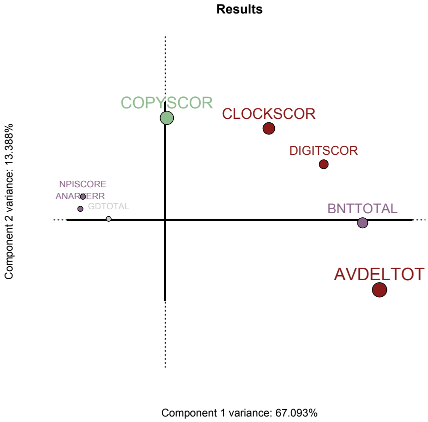


Figure 8

Which structural regions significantly contribute?

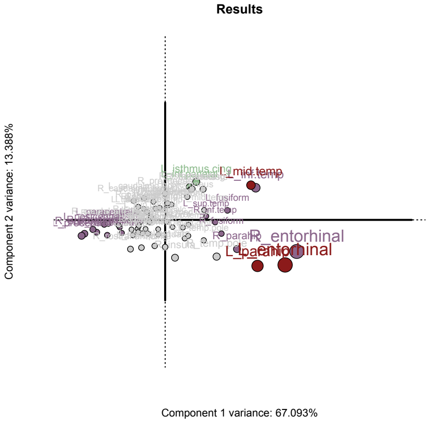


Figure 9

Putting all the pieces together

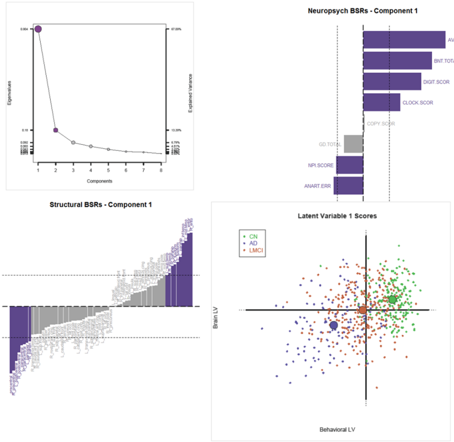


Figure 10

Mean-centered PLSC

Mean-centered PLSC

A.k.a.

- Barycentric discriminant analysis
- Between-groups PLSC
- Discriminant PLSC

Mean-centered PLSC

- used when observations are structured into groups or conditions
- **X** is a matrix of observations by variables
- **Y** is a dummy matrix that codes for experimental groups or conditions

An example

ADNI structural thickness data by diagnostic group

$$\mathbf{X} = \begin{array}{l} \\ \text{Subj}_1 \\ \text{Subj}_2 \\ \text{Subj}_3 \\ \vdots \\ \text{Subj}_{N\text{-}1} \\ \text{Subj}_N \end{array} \begin{array}{c} \begin{array}{ccccc} \text{R.IFG} & \text{L.IFG} & \text{R.STG} & \ldots & \text{L.Fusi} \end{array} \\ \left[\begin{array}{ccccc} 3.24 & 6.27 & 2.31 & \ldots & 2.32 \\ 5.89 & 0.26 & 3.99 & \ldots & 4.51 \\ 2.84 & 2.51 & 5.88 & \ldots & 1.17 \\ \vdots & \vdots & \vdots & & \vdots \\ 1.96 & 8.9 & 7.19 & \ldots & 3.46 \\ 4.42 & 7.81 & 5.99 & \ldots & 1.96 \end{array}\right] \end{array} \qquad \mathbf{Y} = \begin{array}{l} \\ \text{Subj}_1 \\ \text{Subj}_2 \\ \text{Subj}_3 \\ \vdots \\ \text{Subj}_{N\text{-}1} \\ \text{Subj}_N \end{array} \begin{array}{c} \begin{array}{ccc} \text{CN} & \text{MCI} & \text{AD} \end{array} \\ \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{array}\right] \end{array}$$

Figure 11: **X** and **Y** matrices in mean-centered PLSC

Group component scores

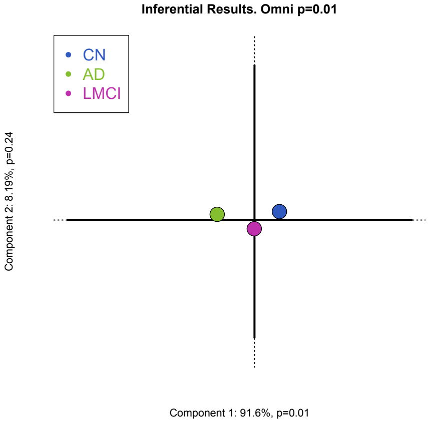


Figure 12

One set of LV scores

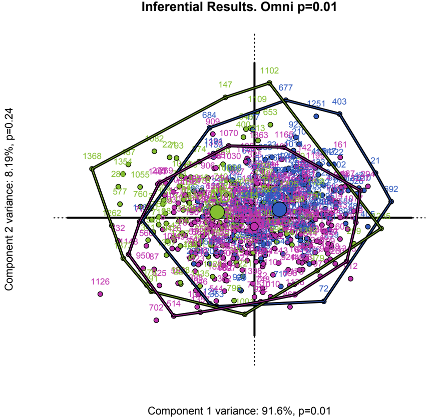


Figure 13

Fixed effects classification accuracy

Because mc-PLSC is about groups we can assess classification accuracy

Fixed effects classification accuracy

Table 5: 267 of 569 correctly classified. 46.92 % accuracy

	CN.actual	LMCI.actual	AD.actual
CN.predicted	117	108	25
LMCI.predicted	39	76	17
AD.predicted	22	91	74

Inference Testing

Bootstrapping group component scores

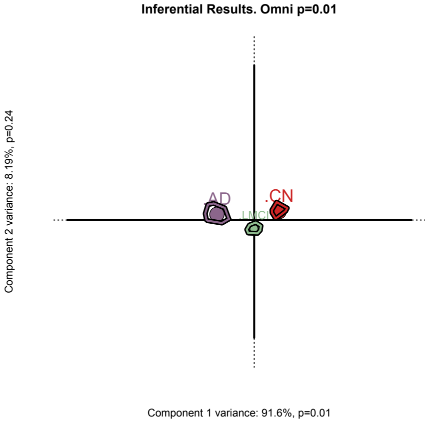


Figure 14

Bootstrapping structural thickness estimates

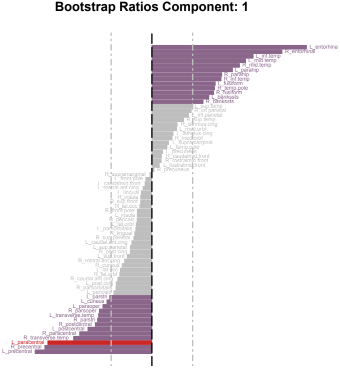


Figure 15

Leave-one-out cross validation

Table 6: 238 of 569 correctly classified. 41.83 % accuracy

	CN.actual	LMCI.actual	AD.actual
CN.predicted	107	111	26
LMCI.predicted	47	65	24
AD.predicted	24	99	66

“Seed” PLSC

Seed PLSC

- used to examine "connectivity" between two sets of varirables
- name comes from "connectivity"
- broadly it is correlation (or covariance) between
  - $\mathbf{X}_A$ for A set of variables
  - $\mathbf{X}_B$ for B set of variables
- Also referred to as "Burt bands"

Seed PLSC

- **X** is a matrix of observations by variables
- **Y** is a subset of variables (removed from **X**)
  - subset of voxels (i.e., ROI)

An example

Structural connectivity in ADNI

$$\mathbf{X} = \begin{array}{c} \\ \mathrm{Subj}_1 \\ \mathrm{Subj}_2 \\ \mathrm{Subj}_3 \\ \vdots \\ \mathrm{Subj}_{N\text{-}1} \\ \mathrm{Subj}_N \end{array} \begin{array}{c} \begin{array}{ccccccc} \mathrm{R.IFG} & \mathrm{L.IFG} & \mathrm{R.STG} & \mathrm{L.STG} & \mathrm{R.MTG} & \ldots & \mathrm{L.Fusi} \end{array} \\ \begin{bmatrix} 3.24 & 6.27 & 2.31 & 2.67 & 2.59 & \ldots & 2.32 \\ 5.89 & 0.26 & 3.99 & 2.71 & 1.02 & \ldots & 4.51 \\ 2.84 & 2.51 & 5.88 & 3.61 & 4.56 & \ldots & 1.17 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1.96 & 8.9 & 7.19 & 5.36 & 3.01 & \ldots & 3.46 \\ 4.42 & 7.81 & 5.99 & 6.3 & 0.23 & \ldots & 1.96 \end{bmatrix} \end{array}$$

Figure 16: **X** matrix of Freesurfer thickness estimates

Structural connectivity in ADNI

$$
\mathbf{X} = \begin{array}{c|ccccccc|}
 & \text{R.IFG} & \text{L.IFG} & \text{R.STG} & \text{L.STG} & \text{R.MTG} & \ldots & \text{L.Fusi} \\
\text{Subj}_1 & 3.24 & 6.27 & 2.31 & 2.67 & 2.59 & \ldots & 2.32 \\
\text{Subj}_2 & 5.89 & 0.26 & 3.99 & 2.71 & 1.02 & \ldots & 4.51 \\
\text{Subj}_3 & 2.84 & 2.51 & 5.88 & 3.61 & 4.56 & \ldots & 1.17 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\
\text{Subj}_{\text{N-1}} & 1.96 & 8.9 & 7.19 & 5.36 & 3.01 & \ldots & 3.46 \\
\text{Subj}_{\text{N}} & 4.42 & 7.81 & 5.99 & 6.3 & 0.23 & \ldots & 1.96
\end{array}
$$

Figure 17: Selecting lateral temporal regions as a seed

Structural connectivity in ADNI

$$
\mathbf{X} = \begin{array}{c} \\ \text{Subj}_1 \\ \text{Subj}_2 \\ \text{Subj}_3 \\ \vdots \\ \text{Subj}_{N\text{-}1} \\ \text{Subj}_N \end{array}
\begin{array}{c} \begin{array}{cccc} \text{R.IFG} & \text{L.IFG} & \dots & \text{L.Fusi} \end{array} \\
\begin{bmatrix} 3.24 & 6.27 & \dots & 2.32 \\ 5.89 & 0.26 & \dots & 4.51 \\ 2.84 & 2.51 & \dots & 1.17 \\ \vdots & \vdots & & \vdots \\ 1.96 & 8.9 & \dots & 3.46 \\ 4.42 & 7.81 & \dots & 1.96 \end{bmatrix} \end{array}
\qquad
\mathbf{Y} = \begin{array}{c} \\ \text{Subj}_1 \\ \text{Subj}_2 \\ \text{Subj}_3 \\ \vdots \\ \text{Subj}_{N\text{-}1} \\ \text{Subj}_N \end{array}
\begin{array}{c} \begin{array}{ccccc} \text{R.STG} & \text{L.STG} & \text{R.MTG} & \dots & \text{.TempPole} \end{array} \\
\begin{bmatrix} 2.31 & 2.67 & 2.59 & \dots & 2.59 \\ 3.99 & 2.71 & 1.02 & \dots & 1.02 \\ 5.88 & 3.61 & 4.56 & \dots & 4.56 \\ \vdots & \vdots & \vdots & & \vdots \\ 7.19 & 5.36 & 3.01 & \dots & 3.01 \\ 5.99 & 6.3 & 0.23 & \dots & 0.23 \end{bmatrix} \end{array}
$$

Figure 18: **X** & **Y** martices in seed PLS

Connectivity Matrices

Structural connectivity in ADNI

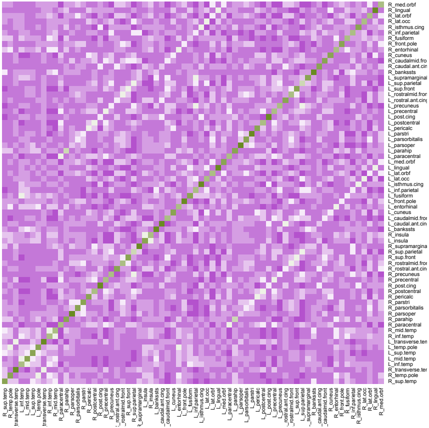


Figure 19

Structural connectivity in ADNI

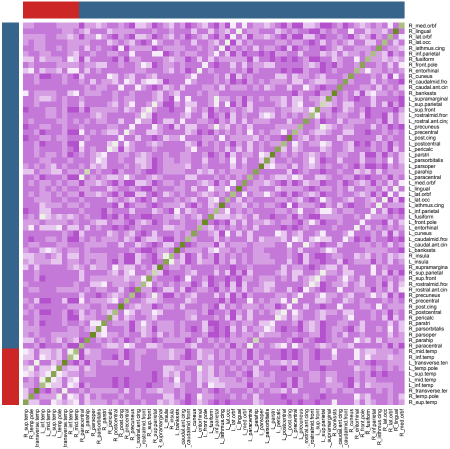


Figure 20

Structural connectivity in ADNI

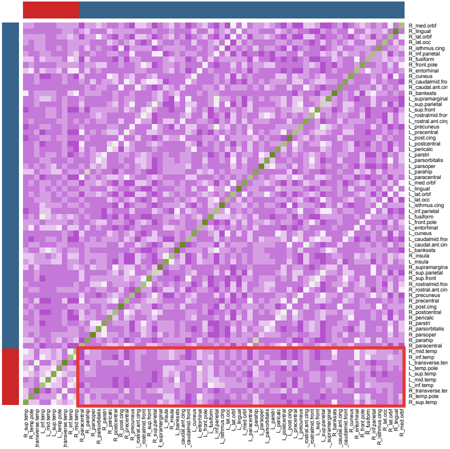


Figure 21

Seed PLSC results

Seed PLSC scree

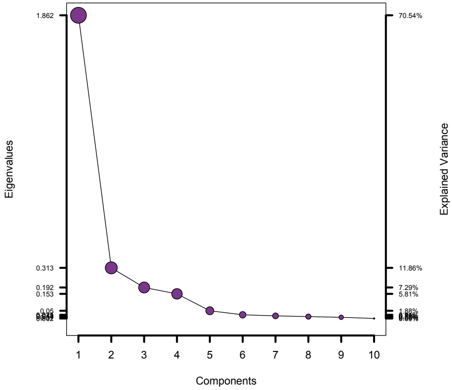


Figure 22

Structural connectivity latent variable

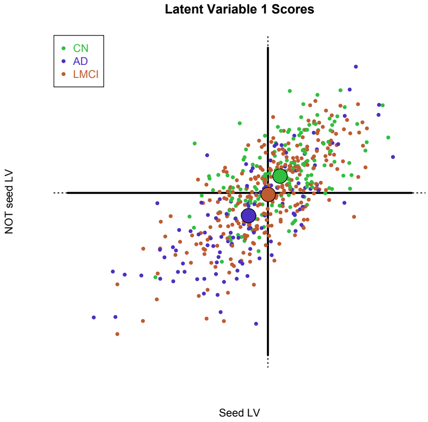
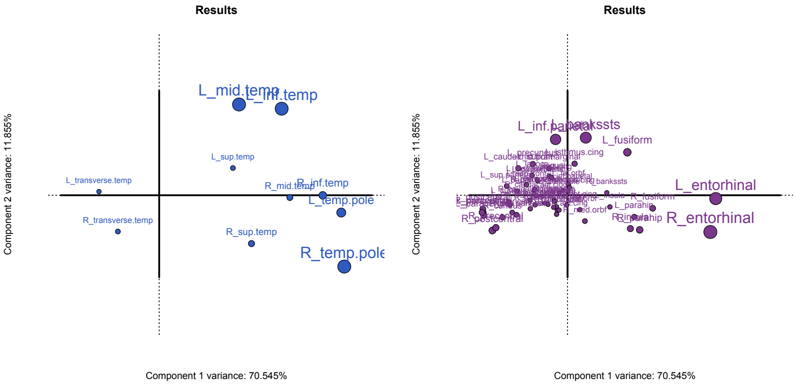


Figure 23

Seed and Not Seed component scores



PLSC wrap-up

PLSC

And Beyond

Friends of PLS

RRR & CCA

Reduced Rank Regression (RRR) and Canonical Correlation Analysis (CCA)

- PLS is about covariance: $\mathbf{X}^T\mathbf{Y}$
- RRR is like OLS:
  - $\mathbf{R}_{RRR} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T$
- CCA is about correlation:
  - $\mathbf{R}_{CCA} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T$
- These look very similar
  - But they really are not
  - See McIntosh & Misic (2013)

PLS-Correspondence analysis

- What if you have non-quantitative data?
  - For both **X** and **Y**
  - For just **X** or **Y**
  - Mixed within **X** or **Y**
- PLS-CA handles mixed data in a PLSC framework
  - Generalizes PLSC
  - See Beaton et al., (2016)

More inference & issues

Resampling & Inference

- See `https://github.com/derekbeaton/Workshops/tree/master/RTC/Apr2017` for prequel to today's workshop
- We've covered
  - Bootstrap
  - Permutation
  - Leave-one-out (a bit)
- We did not cover
  - Variations of the above
  - Split-half
- There are some known issues to be aware of
  - Kovacevic et al., (2013)
  - Churchill et al., (2013)

Conclusions

Major points

- If you know the SVD you know
  - "Almost everything you need to know about [blank]"
- PLS is a large family of techniques
  - In our fields we focus on PLSC
- PLSC has many flavors
  - We've covered a few
  - We'll cover a few more in November
- Today's material will be at:
  - `https://github.com/derekbeaton/Workshops/tree/master/RTC/Oct2017`

Before next time

- We will send out reading material
- Almost everything you need:
  - McIntosh & Lobaugh (2004)
  - Krishan et al., (2011)

Fin

- Questions?
- Comments?
- Complaints?