Principal Components & Multiple Correspondence Analyses

with resampling approaches for stability assessments

Derek Beaton

RRI RTC

May 03, 2019

Where to find everything

- Generally: https://github.com/derekbeaton/workshops

Where to find everything

- Generally: https://github.com/derekbeaton/workshops
- Today: https://github.com/derekbeaton/Workshops/tree/master/RTC/PCA_MCA_Resampling

Some set up

- Use RStudio (makes it easy)

Some set up

- Use RStudio (makes it easy)
- You can pull from the Git repo

Some set up

- Use RStudio (makes it easy)
- You can pull from the Git repo
  - Or copy individual files

Some set up

- Use RStudio (makes it easy)
- You can pull from the Git repo
  - Or copy individual files
- Make .Renviron file

Some set up

- Use RStudio (makes it easy)
- You can pull from the Git repo
  - Or copy individual files
- Make .Renviron file
  - Points to locations outside the repo

Some set up

- Use RStudio (makes it easy)
- You can pull from the Git repo
  - Or copy individual files
- Make .Renviron file
  - Points to locations outside the repo
- Run "/R/0_Create_ADNI_Dataset.R" first

Some set up

- Use RStudio (makes it easy)
- You can pull from the Git repo
  - Or copy individual files
- Make .Renviron file
  - Points to locations outside the repo
- Run "/R/0_Create_ADNI_Dataset.R" first
  - Then either run this .Rmd or

Some set up

- Use RStudio (makes it easy)
- You can pull from the Git repo
    - Or copy individual files
- Make .Renviron file
    - Points to locations outside the repo
- Run "/R/0_Create_ADNI_Dataset.R" first
    - Then either run this .Rmd or
    - Run scripts in order

Some set up

- Use RStudio (makes it easy)
- You can pull from the Git repo
  - Or copy individual files
- Make .Renviron file
  - Points to locations outside the repo
- Run "/R/0_Create_ADNI_Dataset.R" first
  - Then either run this .Rmd or
  - Run scripts in order
- Use of the ADNI data

Some set up

- Use RStudio (makes it easy)
- You can pull from the Git repo
  - Or copy individual files
- Make .Renviron file
  - Points to locations outside the repo
- Run "/R/0_Create_ADNI_Dataset.R" first
  - Then either run this .Rmd or
  - Run scripts in order
- Use of the ADNI data
  - Via the 'ADNIMERGE' package

An advertisement

- Lots of really cool R & RStudio stuff

An advertisement

- Lots of really cool R & RStudio stuff
- This presentation is 90% reproducible

An advertisement

- Lots of really cool R & RStudio stuff
- This presentation is 90% reproducible
  - Resampling is painful

An advertisement

- Lots of really cool R & RStudio stuff
- This presentation is 90% reproducible
  - Resampling is painful
- R & RStudio “Magic” BrainHackTO tutorial

An advertisement

- Lots of really cool R & RStudio stuff
- This presentation is 90% reproducible
  - Resampling is painful
- R & RStudio “Magic” BrainHackTO tutorial
  - Jenny Rieck & I

An advertisement

- Lots of really cool R & RStudio stuff
- This presentation is 90% reproducible
  - Resampling is painful
- R & RStudio "Magic" BrainHackTO tutorial
  - Jenny Rieck & I
  - May 21 or 22

An advertisement

- Lots of really cool R & RStudio stuff
- This presentation is 90% reproducible
  - Resampling is painful
- R & RStudio “Magic” BrainHackTO tutorial
  - Jenny Rieck & I
  - May 21 or 22
  - Possibly sold out?

An advertisement

- Lots of really cool R & RStudio stuff
- This presentation is 90% reproducible
  - Resampling is painful
- R & RStudio “Magic” BrainHackTO tutorial
  - Jenny Rieck & I
  - May 21 or 22
  - Possibly sold out?
  - We’ll make stuff available

Motivation for today

- Not everything is a number

Motivation for today

- Not everything is a number
- But with care, it can be turned into one

Overview

- Introduction

Overview

- Introduction
- PCA

Overview

- Introduction
- PCA
- CA

Overview

- Introduction
- PCA
- CA
- Resampling

Overview

- Introduction
- PCA
- CA
- Resampling
- Final notes

Introduction

Prehistory

- Basis:

Prehistory

- Basis:
  - Hotelling (1933)

Prehistory

- Basis:
  - Hotelling (1933)
  - Eckart & Yong (1936)

Prehistory

- Basis:
  - Hotelling (1933)
  - Eckart & Yong (1936)
- Traces back to

Prehistory

- Basis:
  - Hotelling (1933)
  - Eckart & Yong (1936)
- Traces back to
  - Cauchy (1829)

Prehistory

- Basis:
  - Hotelling (1933)
  - Eckart & Yong (1936)
- Traces back to
  - Cauchy (1829)
  - Galton (1859)

Prehistory

- Basis:
  - Hotelling (1933)
  - Eckart & Yong (1936)
- Traces back to
  - Cauchy (1829)
  - Galton (1859)
  - K. Pearson (1901)

Prehistory

- Basis:
  - Hotelling (1933)
  - Eckart & Yong (1936)
- Traces back to
  - Cauchy (1829)
  - Galton (1859)
  - K. Pearson (1901)
  - Spearman (1904)

History

- “Modern form” of PCA & factor analyses

History

- "Modern form" of PCA & factor analyses
  - Thurstone (1934)

History

- "Modern form" of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)

History

- “Modern form” of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)
  - Tucker (too many to list)

History

- "Modern form" of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)
  - Tucker (too many to list)
  - Many others in 1940s-1960s

History

- "Modern form" of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)
  - Tucker (too many to list)
  - Many others in 1940s-1960s
- CA

History

- "Modern form" of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)
  - Tucker (too many to list)
  - Many others in 1940s-1960s
- CA
  - Hirschfeld (1935)

History

- "Modern form" of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)
  - Tucker (too many to list)
  - Many others in 1940s-1960s
- CA
  - Hirschfeld (1935)
  - Guttman (1941)

History

- "Modern form" of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)
  - Tucker (too many to list)
  - Many others in 1940s-1960s
- CA
  - Hirschfeld (1935)
  - Guttman (1941)
  - Burt (1950)

History

- "Modern form" of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)
  - Tucker (too many to list)
  - Many others in 1940s-1960s
- CA
  - Hirschfeld (1935)
  - Guttman (1941)
  - Burt (1950)
  - Benzecri (1964)

History

- "Modern form" of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)
  - Tucker (too many to list)
  - Many others in 1940s-1960s
- CA
  - Hirschfeld (1935)
  - Guttman (1941)
  - Burt (1950)
  - Benzecri (1964)
  - Escofier (1965)

History

- “Modern form” of PCA & factor analyses
  - Thurstone (1934)
  - Fisher (1940)
  - Tucker (too many to list)
  - Many others in 1940s-1960s
- CA
  - Hirschfeld (1935)
  - Guttman (1941)
  - Burt (1950)
  - Benzecri (1964)
  - Escofier (1965)
- See Lebart's History & Prehistory of CA: http://www.dtmvic.com/doc/About_the_History_of_CA.pdf

Now & The Future

- PCA is always cool.

Now & The Future

- PCA is always cool.
- See the final slides for related methods

Now & The Future

- PCA is always cool.
- See the final slides for related methods
  - PCA makes you familiar with all of them

Now & The Future

- PCA is always cool.
- See the final slides for related methods
  - PCA makes you familiar with all of them
  - CA makes you an expert with all of them

PCA & CA

- Visualize multiple/high dimensions

PCA & CA

- Visualize multiple/high dimensions
- Dimensionality reduction

PCA & CA

- Visualize multiple/high dimensions
- Dimensionality reduction
- Matrix factorization

PCA & CA

- Visualize multiple/high dimensions
- Dimensionality reduction
- Matrix factorization
- Unsupervised learning

PCA & CA

- Find “components”

PCA & CA

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

PCA & CA

- Find “components”
  - Components are new variables that are combinations of the original variables
- Components explain maximal variance

PCA & CA

- Find "components"
  - Components are new variables that are combinations of the original variables
- Components explain maximal variance
  - Conditional to orthogonality

PCA & CA

- Find "components"
  - Components are new variables that are combinations of the original variables
- Components explain maximal variance
  - Conditional to orthogonality
- So what's the difference?

PCA vs CA

- PCA: For generally continuous (interval scale) data

PCA vs CA

- PCA: For generally continuous (interval scale) data
- CA: For (almost) everything else

PCA vs CA

- PCA: For generally continuous (interval scale) data
- CA: For (almost) everything else
  - And also for continuous data!

Under the hood

- The eigenvalue decomposition (EVD)

Under the hood

- The eigenvalue decomposition (EVD)
  - Requires squares, symmetric, and positive semi definite

Under the hood

- The eigenvalue decomposition (EVD)
  - Requires squares, symmetric, and positive semi definite
  - Generally correlation or covariance

Under the hood

- The eigenvalue decomposition (EVD)
  - Requires squares, symmetric, and positive semi definite
  - Generally correlation or covariance
- The singular value decomposition (SVD)

Under the hood

- The eigenvalue decomposition (EVD)
  - Requires squares, symmetric, and positive semi definite
  - Generally correlation or covariance
- The singular value decomposition (SVD)
  - Works with rectangular tables

Under the hood

- The eigenvalue decomposition (EVD)
  - Requires squares, symmetric, and positive semi definite
  - Generally correlation or covariance
- The singular value decomposition (SVD)
  - Works with rectangular tables
- A generalized SVD

Under the hood

- The eigenvalue decomposition (EVD)
  - Requires squares, symmetric, and positive semi definite
  - Generally correlation or covariance
- The singular value decomposition (SVD)
  - Works with rectangular tables
- A generalized SVD
  - Apply constraints (weights) to rows & columns of rectangular table

Under the hood

- The eigenvalue decomposition (EVD)
  - Requires squares, symmetric, and positive semi definite
  - Generally correlation or covariance
- The singular value decomposition (SVD)
  - Works with rectangular tables
- A generalized SVD
  - Apply constraints (weights) to rows & columns of rectangular table
  - Required for CA and fancier PCA-like techniques & extensions

Terms I will use today

- Component scores

Terms I will use today

- Component scores
  - Values assigned to rows (PCA & CA) or columns (CA) scaled by variance

Terms I will use today

- Component scores
  - Values assigned to rows (PCA & CA) or columns (CA) scaled by variance
- Correlation loadings (PCA)

Terms I will use today

- Component scores
  - Values assigned to rows (PCA & CA) or columns (CA) scaled by variance
- Correlation loadings (PCA)
  - Correlation of original data with row component scores (observations)

Terms I will use today

- Component scores
  - Values assigned to rows (PCA & CA) or columns (CA) scaled by variance
- Correlation loadings (PCA)
  - Correlation of original data with row component scores (observations)
- Explained variance

Terms I will use today

- Component scores
  - Values assigned to rows (PCA & CA) or columns (CA) scaled by variance
- Correlation loadings (PCA)
  - Correlation of original data with row component scores (observations)
- Explained variance
  - Eigenvalues

Terms I will use today

- Component scores
  - Values assigned to rows (PCA & CA) or columns (CA) scaled by variance
- Correlation loadings (PCA)
  - Correlation of original data with row component scores (observations)
- Explained variance
  - Eigenvalues
  - How much of the total variance per component

Terms I will use today

- Component scores
  - Values assigned to rows (PCA & CA) or columns (CA) scaled by variance
- Correlation loadings (PCA)
  - Correlation of original data with row component scores (observations)
- Explained variance
  - Eigenvalues
  - How much of the total variance per component
  - Variance = Sums of squares

Terms I will use today

- Component scores
  - Values assigned to rows (PCA & CA) or columns (CA) scaled by variance
- Correlation loadings (PCA)
  - Correlation of original data with row component scores (observations)
- Explained variance
  - Eigenvalues
  - How much of the total variance per component
  - Variance = Sums of squares
- Magic

Today

- ExPosition

Today

- ExPosition
  - Family of packages

Today

- ExPosition
  - Family of packages
  - Includes resampling

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra
  - Awesome ggplot2 visualizers for ExPosition

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra
  - Awesome ggplot2 visualizers for ExPosition
  - http://www.alboukadel.com/ & http://www.sthda.com/english/

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra
  - Awesome ggplot2 visualizers for ExPosition
  - http://www.alboukadel.com/ & http://www.sthda.com/english/
- ggplot2 & tidyverse

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra
  - Awesome ggplot2 visualizers for ExPosition
  - http://www.alboukadel.com/ & http://www.sthda.com/english/
- ggplot2 & tidyverse
- ours

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra
  - Awesome ggplot2 visualizers for ExPosition
  - http://www.alboukadel.com/ & http://www.sthda.com/english/
- ggplot2 & tidyverse
- ours
  - Developed here within ONDRI

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra
  - Awesome ggplot2 visualizers for ExPosition
  - http://www.alboukadel.com/ & http://www.sthda.com/english/
- ggplot2 & tidyverse
- ours
  - Developed here within ONDRI
  - New package for outliers

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra
  - Awesome ggplot2 visualizers for ExPosition
  - http://www.alboukadel.com/ & http://www.sthda.com/english/
- ggplot2 & tidyverse
- ours
  - Developed here within ONDRI
  - New package for outliers
  - Has some important bells-and-whistles

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra
  - Awesome ggplot2 visualizers for ExPosition
  - http://www.alboukadel.com/ & http://www.sthda.com/english/
- ggplot2 & tidyverse
- ours
  - Developed here within ONDRI
  - New package for outliers
  - Has some important bells-and-whistles
- Making things look fancy:

Today

- ExPosition
  - Family of packages
  - Includes resampling
  - Lots of PCA & CA techniques
- factoextra
  - Awesome ggplot2 visualizers for ExPosition
  - http://www.alboukadel.com/ & http://www.sthda.com/english/
- ggplot2 & tidyverse
- ours
  - Developed here within ONDRI
  - New package for outliers
  - Has some important bells-and-whistles
- Making things look fancy:
  - kable, kableExtra, gridExtra, ggcorrplot

Some alternatives

- FactoMineR

Some alternatives

- FactoMineR
- ade4

Some alternatives

- FactoMineR
- ade4
- ca

Some alternatives

- FactoMineR
- ade4
- ca
- MASS

Some alternatives

- FactoMineR
- ade4
- ca
- MASS
- psych

Some alternatives

- FactoMineR
- ade4
- ca
- MASS
- psych
- So many others

Typology

- SS Stevens

Typology

- SS Stevens
  - Not a boat!

Typology

- SS Stevens
  - Not a boat!
- Levels of measurement

Typology

- SS Stevens
  - Not a boat!
- Levels of measurement
  - Nominal (categorical)

Typology

- SS Stevens
  - Not a boat!
- Levels of measurement
  - Nominal (categorical)
  - Ordinal (ranked, discrete categories)

Typology

- SS Stevens
  - Not a boat!
- Levels of measurement
  - Nominal (categorical)
  - Ordinal (ranked, discrete categories)
  - Interval (continuous, arbitrary 0)

Typology

- SS Stevens
  - Not a boat!
- Levels of measurement
  - Nominal (categorical)
  - Ordinal (ranked, discrete categories)
  - Interval (continuous, arbitrary 0)
  - Ratio (continuous, non-arbitrary 0)

Typology

- SS Stevens
  - Not a boat!
- Levels of measurement
  - Nominal (categorical)
  - Ordinal (ranked, discrete categories)
  - Interval (continuous, arbitrary 0)
  - Ratio (continuous, non-arbitrary 0)
- Excellent examples: https://en.wikipedia.org/wiki/Level_of_measurement

Today's data

- Alzheimer's Disease Neuroimaging Initiative (ADNI)

Today's data

- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Data set:

Today's data

- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Data set:
  - 665 observations

Today's data

- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Data set:
  - 665 observations
  - 17 variables

Today's data

- Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Data set:
  - 665 observations
  - 17 variables
- Walk through this set to tell a whole story

Today's data

	Continuous	**Categorical**	**Ordinal**
DX		YES	
AGE	YES		
PTGENDER		YES	
PTEDUCAT			YES
PTETHCAT		YES	
PTRACCAT		YES	
APOE4		YES	YES
FDG	YES		
AV45	YES		
CDRSB			YES
ADAS13			YES
MOCA			YES
WholeBrain	YES		
Hippocampus	YES		
MidTemp	YES		
mPACCtrailsB	YES		
HMSCORE		YES	YES

Principal Components Analysis

Let's dive in

- We'll start with just two variables:

Let's dive in

- We'll start with just two variables:
- Trails

Let's dive in

- We'll start with just two variables:
- Trails
  - Neuropsych test

Let's dive in

- We'll start with just two variables:
- Trails
  - Neuropsych test
  - Executive function

Let's dive in

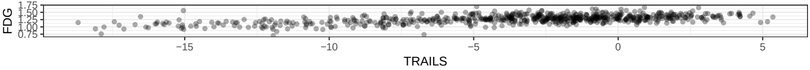
- We'll start with just two variables:
- Trails
  - Neuropsych test
  - Executive function
- FDG

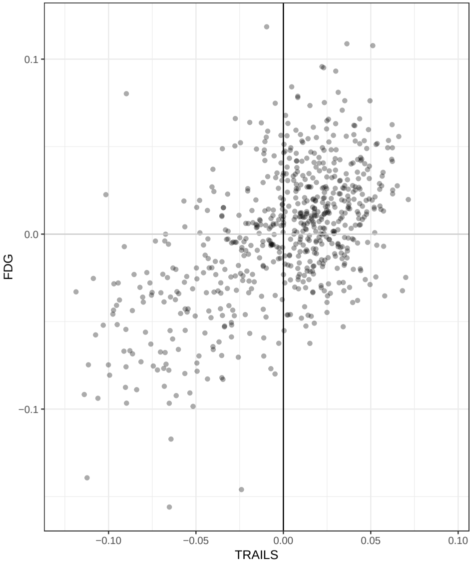
Let's dive in

- We'll start with just two variables:
- Trails
  - Neuropsych test
  - Executive function
- FDG
  - PET imaging; brain function

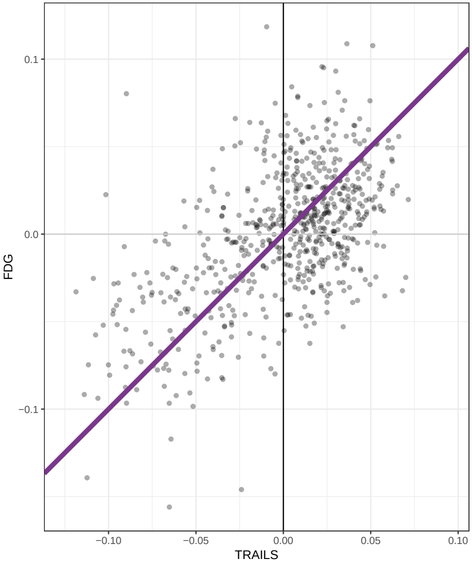
Let's dive in

- We'll start with just two variables:
- Trails
  - Neuropsych test
  - Executive function
- FDG
  - PET imaging; brain function
  - Average of several brain regions

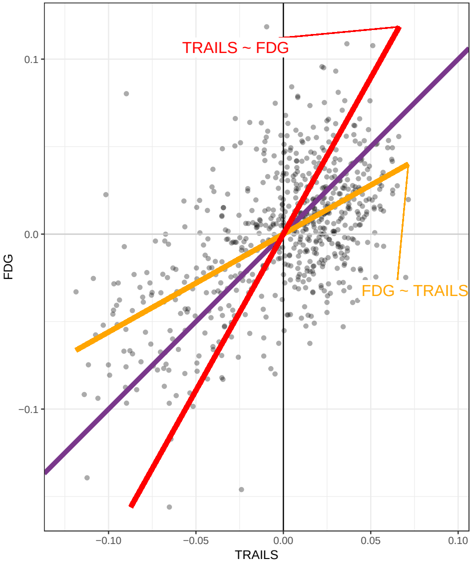

FDG
1.75
1.50
1.25
1.00
0.75
−15
−10
−5
0
5
TRAILS

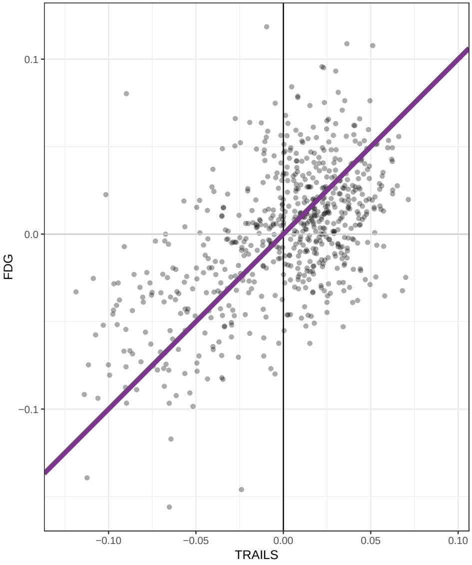


FDG
0.1
0.0
−0.1
−0.10
−0.05
0.00
0.05
0.10
TRAILS

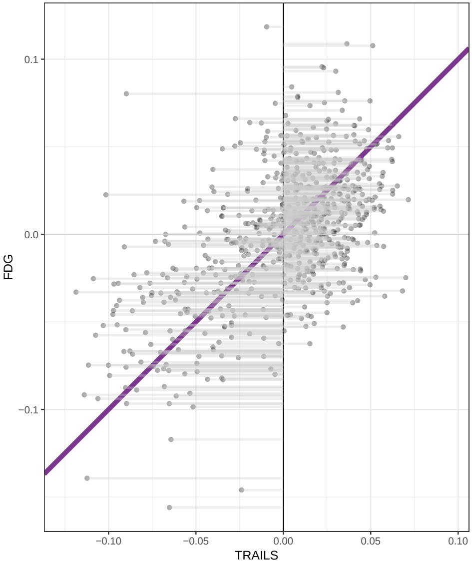


0.1
0.0
−0.1
FDG
−0.10
−0.05
0.00
0.05
0.10
TRAILS

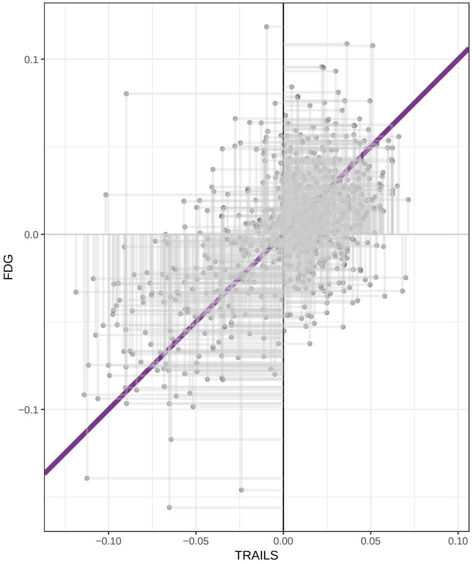

TRAILS ~ FDG
FDG ~ TRAILS
0.1
0.0
−0.1
FDG
−0.10
−0.05
0.00
0.05
0.10
TRAILS



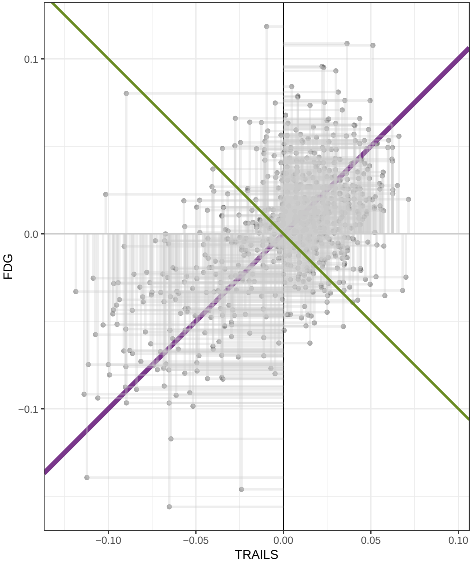
0.1
0.0
−0.1
FDG
−0.10
−0.05
0.00
0.05
0.10
TRAILS



0.1
0.0
−0.1
FDG
−0.10
−0.05
0.00
0.05
0.10
TRAILS

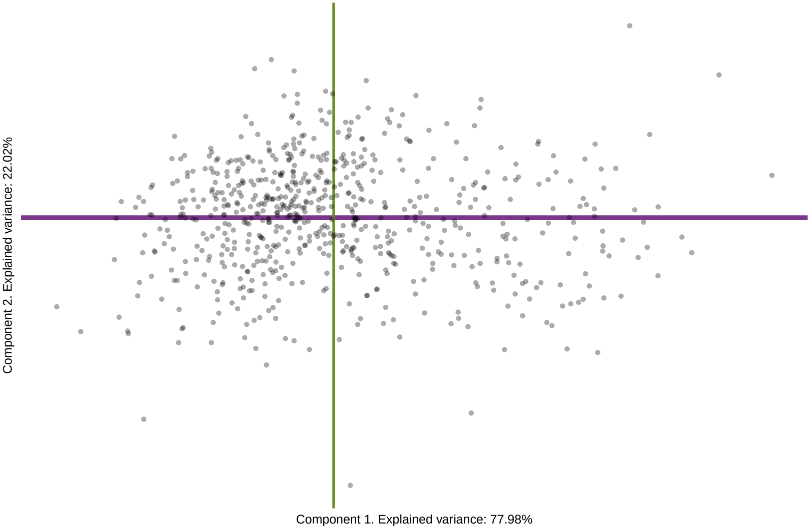


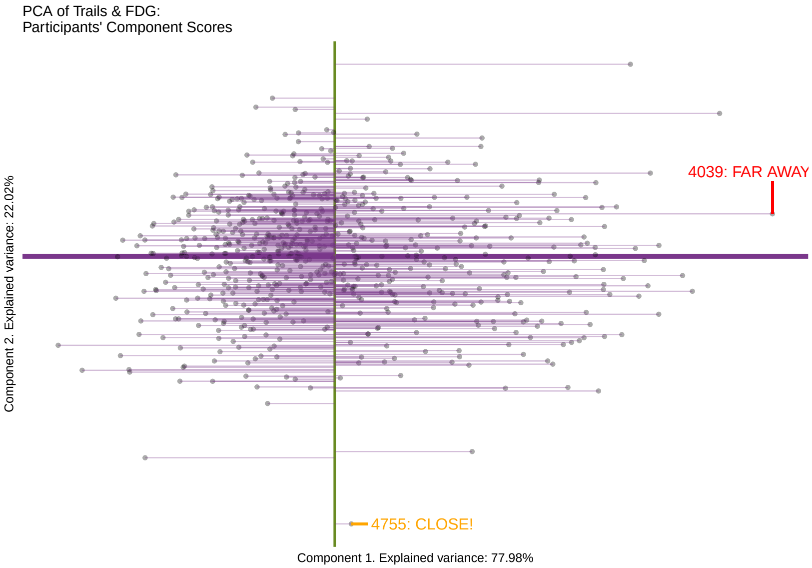
0.1
0.0
−0.1
FDG
−0.10
−0.05
0.00
0.05
0.10
TRAILS



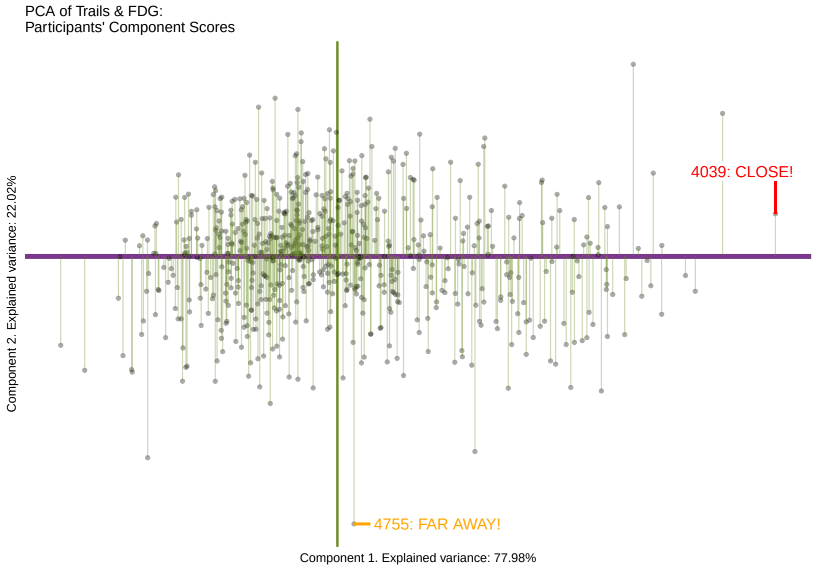
0.1
0.0
−0.1
FDG
−0.10
−0.05
0.00
0.05
0.10
TRAILS

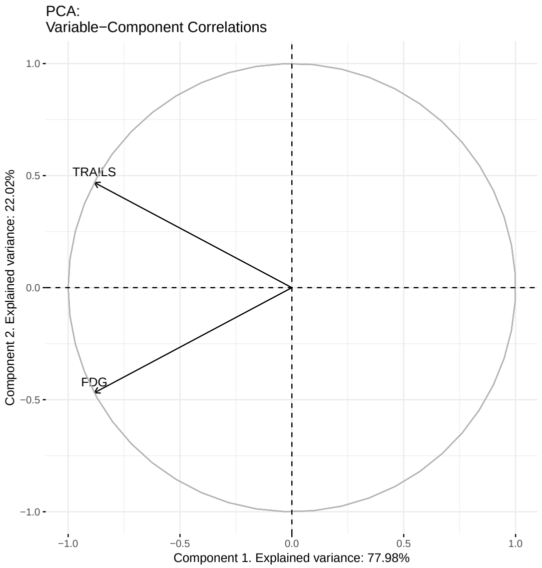
PCA of Trails & FDG:
Participants' Component Scores

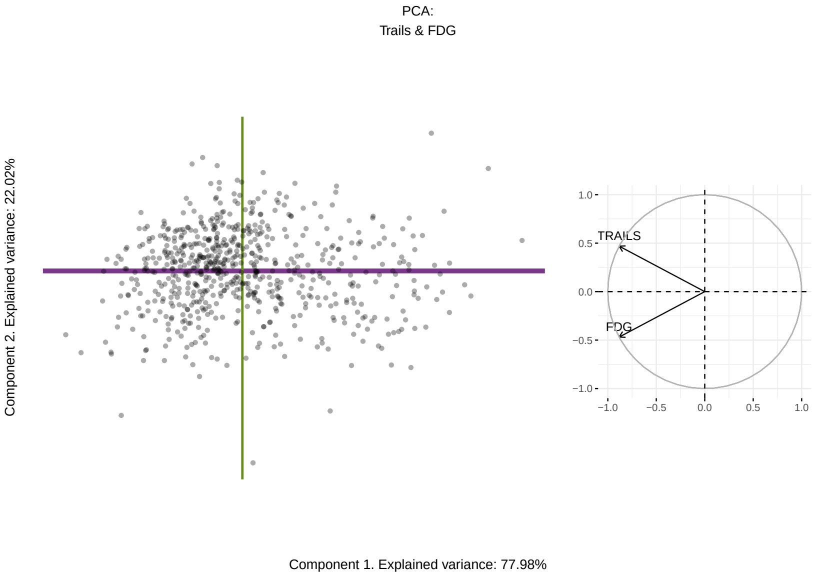


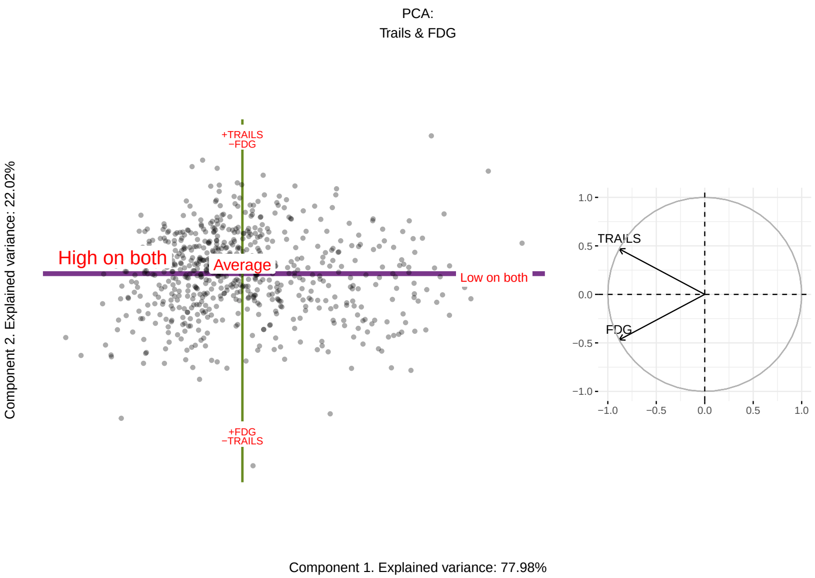


PCA of Trails & FDG:
Participants' Component Scores
Component 2. Explained variance: 22.02%
Component 1. Explained variance: 77.98%
4039: FAR AWAY
4755: CLOSE!


PCA of Trails & FDG:
Participants' Component Scores
4039: CLOSE!
4755: FAR AWAY!
Component 1. Explained variance: 77.98%
Component 2. Explained variance: 22.02%


PCA:
Variable−Component Correlations
TRAILS
FDG
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
Component 1. Explained variance: 77.98%
Component 2. Explained variance: 22.02%


PCA:
Trails & FDG
Component 2. Explained variance: 22.02%
Component 1. Explained variance: 77.98%
TRAILS
FDG
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0


PCA:
Trails & FDG
+TRAILS
−FDG
High on both
Average
Low on both
+FDG
−TRAILS
Component 2. Explained variance: 22.02%
Component 1. Explained variance: 77.98%
TRAILS
FDG
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

Scaling up

- Scale up: MORE DATA!

Scaling up

- Scale up: MORE DATA!
- All of the continuous variables

Scaling up

	AGE	FDG	AV45	WholeBrain	Hippocampus	MidTemp	mPACCtrailsB
5023	63.9	1.29	1.03	1057350.97	7904	21306	1.81
5026	70.5	1.08	1.44	1023057.28	8051	16501	−1.45
5027	75.5	1.06	1.44	986723.65	6534	17437	−17.27
5028	61.9	1.13	1.38	1182704.57	7481	20797	−11.5
5031	80.2	1.14	1.52	908133.86	5040	19032	−8.21
5037	67.3	0.98	1.21	1161499.61	5831	21428	−12.8
5040	75.9	1.24	1.01	943160.57	7994	16634	0.94
5047	68.8	1.7	1.48	1070406.07	7920	22043	−4.9
5054	74	1.12	1.43	1138040.06	6580	20836	−7.63
5058	61.8	0.97	1.54	1195549.29	7318	22757	−9.18
5063	71.5	0.92	1.61	817421.23	5364	12542	−15.03

A new plot

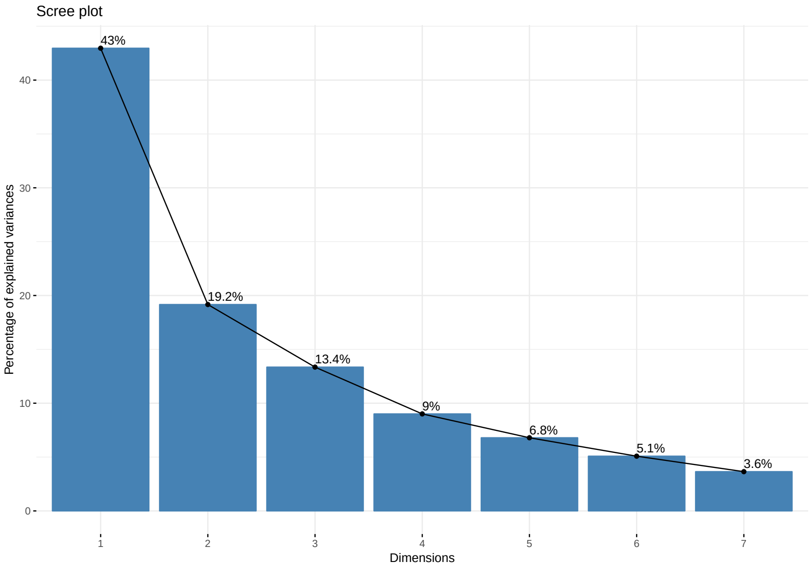
- Scree (Cattell)

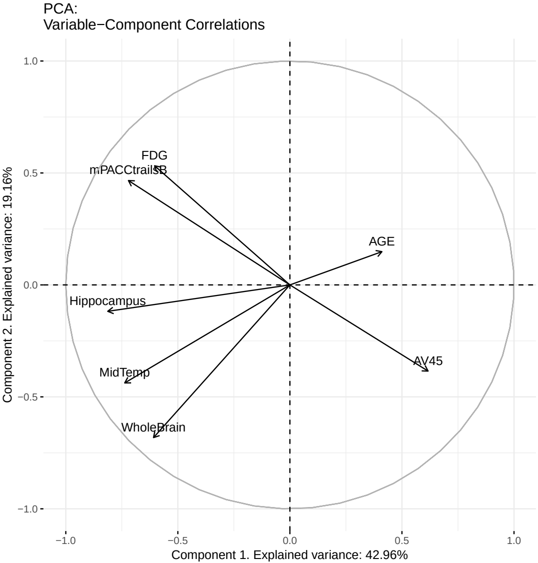
A new plot

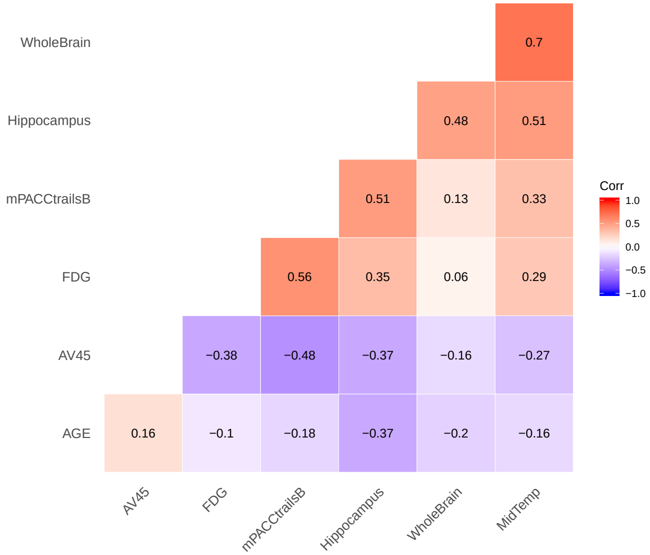
- Scree (Cattell)
- Junk at the bottom of a slope

A new plot

- Scree (Cattell)
- Junk at the bottom of a slope
- Shows us explained variance (%) per component

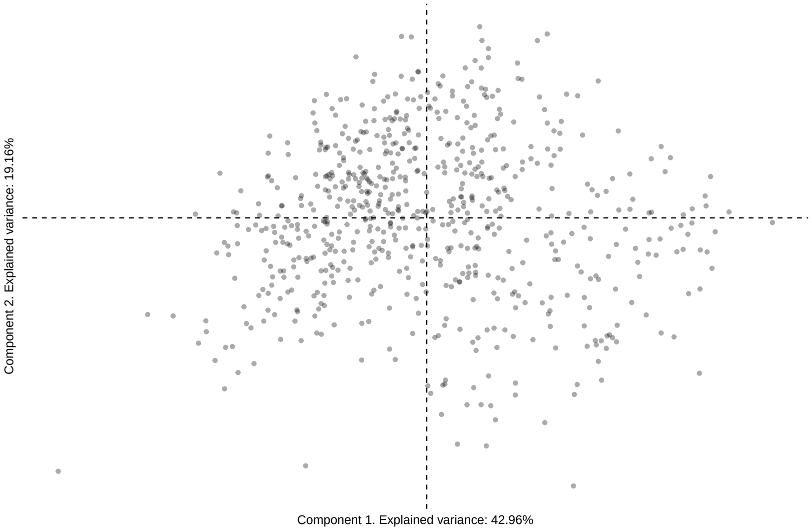

Scree plot
43%
19.2%
13.4%
9%
6.8%
5.1%
3.6%
Percentage of explained variances
0
10
20
30
40
1
2
3
4
5
6
7
Dimensions

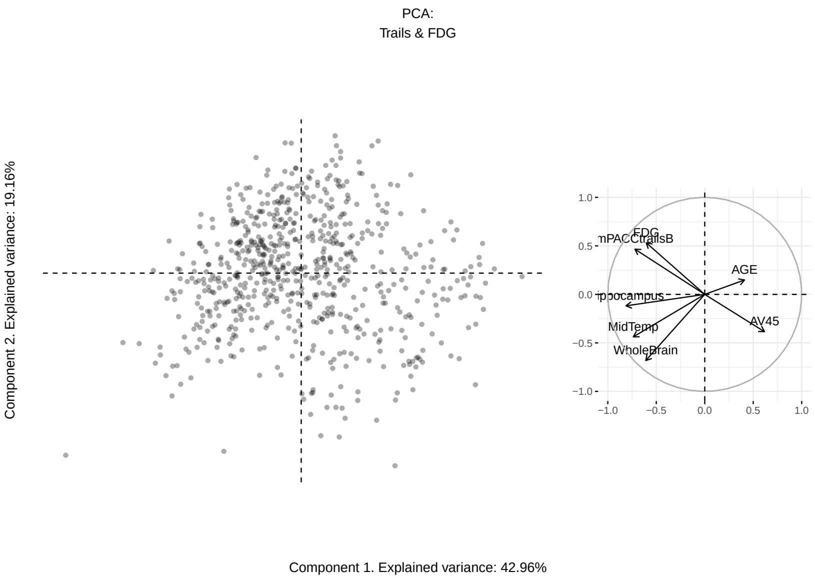

PCA:
Variable−Component Correlations
FDG
mPACCtrailsB
AGE
Hippocampus
AV45
MidTemp
WholeBrain
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
Component 2. Explained variance: 19.16%
Component 1. Explained variance: 42.96%



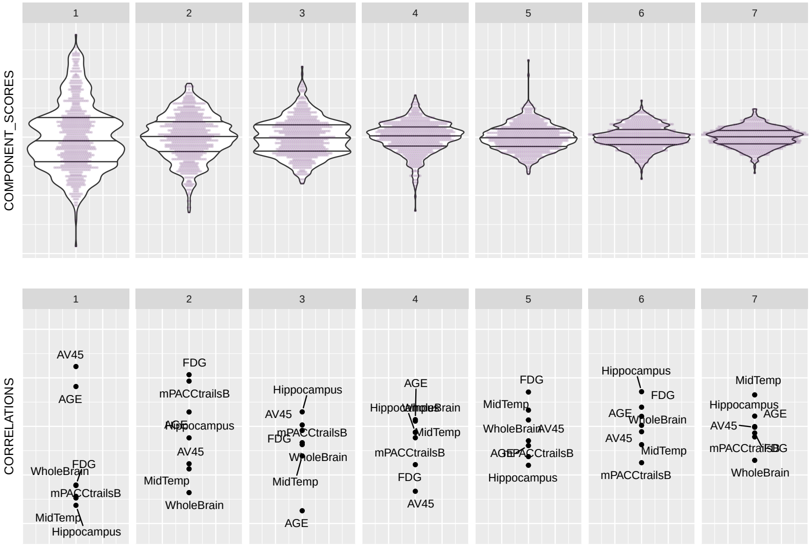
WholeBrain
Hippocampus
mPACCtrailsB
FDG
AV45
AGE
0.7
0.48
0.51
0.51
0.13
0.33
0.56
0.35
0.06
0.29
−0.38
−0.48
−0.37
−0.16
−0.27
0.16
−0.1
−0.18
−0.37
−0.2
−0.16
AV45
FDG
mPACCtrailsB
Hippocampus
WholeBrain
MidTemp
Corr
1.0
0.5
0.0
−0.5
−1.0

PCA:
Participants Component Scores





PCA:
Trails & FDG
Component 2. Explained variance: 19.16%
Component 1. Explained variance: 42.96%
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
FDG
mPACCtrailsB
AGE
ppocampus
AV45
MidTemp
WholeBrain



1
2
3
4
5
6
7
COMPONENT_SCORES
CORRELATIONS
AV45
AGE
FDG
WholeBrain
mPACCtrailsB
MidTemp
Hippocampus
FDG
mPACCtrailsB
AGE
Hippocampus
AV45
MidTemp
WholeBrain
Hippocampus
AV45
mPACCtrailsB
FDG
WholeBrain
MidTemp
AGE
AGE
Hippocampus
WholeBrain
MidTemp
mPACCtrailsB
FDG
AV45
FDG
MidTemp
WholeBrain
AV45
AGE
mPACCtrailsB
Hippocampus
Hippocampus
FDG
AGE
WholeBrain
AV45
MidTemp
mPACCtrailsB
MidTemp
Hippocampus
AGE
AV45
mPACCtrailsB
FDG
WholeBrain

Correspondence analyses

CA

- Like PCA in many ways

CA

- Like PCA in many ways
- Slightly different interpretations

CA

- Like PCA in many ways
- Slightly different interpretations
- So much cooler

CA

- Like PCA in many ways
- Slightly different interpretations
- So much cooler
  - Handles all types of data

Illustrative data

	DX	PTRACCAT
5023	CN	Asian
5026	MCI	White
5027	Dementia	White
5028	Dementia	White
5031	MCI	White
5037	Dementia	Black
5040	CN	Black
5047	MCI	Black
5054	Dementia	White
5058	Dementia	Asian
5063	Dementia	White

Disjunctive data

	DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
5023	0	1	0	0	0	0	1
5026	1	0	0	1	0	0	0
5027	0	0	1	1	0	0	0
5028	0	0	1	1	0	0	0
5031	1	0	0	1	0	0	0
5037	0	0	1	0	0	1	0
5040	0	1	0	0	0	1	0
5047	1	0	0	0	0	1	0
5054	0	0	1	1	0	0	0
5058	0	0	1	0	0	0	1
5063	0	0	1	1	0	0	0

Disjunctive data

	DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
5023	0	1	0	0	0	0	1
5026	1	0	0	1	0	0	0
5027	0	0	1	1	0	0	0
5028	0	0	1	1	0	0	0
5031	1	0	0	1	0	0	0
5037	0	0	1	0	0	1	0
5040	0	1	0	0	0	1	0
5047	1	0	0	0	0	1	0
5054	0	0	1	1	0	0	0
5058	0	0	1	0	0	0	1
5063	0	0	1	1	0	0	0

- Row sums are total number of *original* variables

Disjunctive data

	DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
5023	0	1	0	0	0	0	1
5026	1	0	0	1	0	0	0
5027	0	0	1	1	0	0	0
5028	0	0	1	1	0	0	0
5031	1	0	0	1	0	0	0
5037	0	0	1	0	0	1	0
5040	0	1	0	0	0	1	0
5047	1	0	0	0	0	1	0
5054	0	0	1	1	0	0	0
5058	0	0	1	0	0	0	1
5063	0	0	1	1	0	0	0

- Row sums are total number of *original* variables
- Sum within a variable (e.g. DX) is total number of rows

Disjunctive data

	DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
5023	0	1	0	0	0	0	1
5026	1	0	0	1	0	0	0
5027	0	0	1	1	0	0	0
5028	0	0	1	1	0	0	0
5031	1	0	0	1	0	0	0
5037	0	0	1	0	0	1	0
5040	0	1	0	0	0	1	0
5047	1	0	0	0	0	1	0
5054	0	0	1	1	0	0	0
5058	0	0	1	0	0	0	1
5063	0	0	1	1	0	0	0

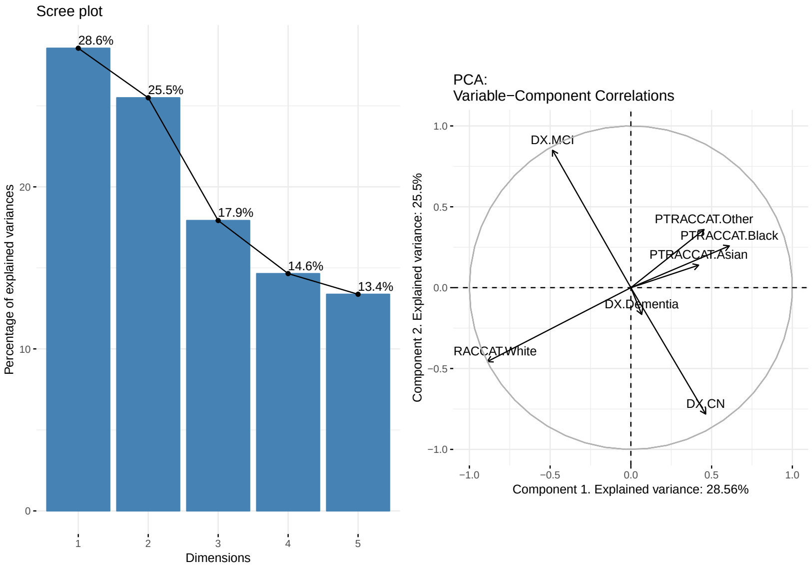
- Row sums are total number of *original* variables
- Sum within a variable (e.g. DX) is total number of rows
- Sum of the table is rows $\times$ columns

A bad idea: PCA

- "coding categorical variables with the indicator matrix of dummy variables and considering them as Gaussian, for instance, is almost a crime."

A bad idea: PCA

- “coding categorical variables with the indicator matrix of dummy variables and considering them as Gaussian, for instance, is almost a crime.”
    - “Jan de Leeuw and the French School of Data Analysis” (Husson, Josse, Saporta)



Scree plot
28.6%
25.5%
17.9%
14.6%
13.4%
Percentage of explained variances
Dimensions
PCA:
Variable−Component Correlations
DX.MCI
PTRACCAT.Other
PTRACCAT.Black
PTRACCAT.Asian
DX.Dementia
RACCAT.White
DX.CN
Component 2. Explained variance: 25.5%
Component 1. Explained variance: 28.56%

Why is that a bad idea?

	DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
DX.MCI	1	−0.815	−0.363	0.045	0.032	−0.043	−0.072
DX.CN	−0.815	1	−0.243	−0.047	0	0.067	0.003
DX.Dementia	−0.363	−0.243	1	0	−0.053	−0.035	0.116
PTRACCAT.White	0.045	−0.047	0	1	−0.562	−0.657	−0.45
PTRACCAT.Other	0.032	0	−0.053	−0.562	1	−0.031	−0.021
PTRACCAT.Black	−0.043	0.067	−0.035	−0.657	−0.031	1	−0.025
PTRACCAT.Asian	−0.072	0.003	0.116	−0.45	−0.021	−0.025	1

	DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
DX.MCI	365	0	0	341	11	10	3
DX.CN	0	235	0	213	6	12	4
DX.Dementia	0	0	65	60	0	1	4
PTRACCAT.White	341	213	60	614	0	0	0
PTRACCAT.Other	11	6	0	0	17	0	0
PTRACCAT.Black	10	12	1	0	0	23	0
PTRACCAT.Asian	3	4	4	0	0	0	11

A better idea

- Correspondence analysis (CA)

A better idea

- Correspondence analysis (CA)
  - Think of it as a χ^2 PCA

A better idea

- Correspondence analysis (CA)
  - Think of it as a χ^2 PCA
- Designed to handle things that look like counts

A better idea

- Correspondence analysis (CA)
  - Think of it as a χ^2 PCA
- Designed to handle things that look like counts
  - That includes categories

A better idea

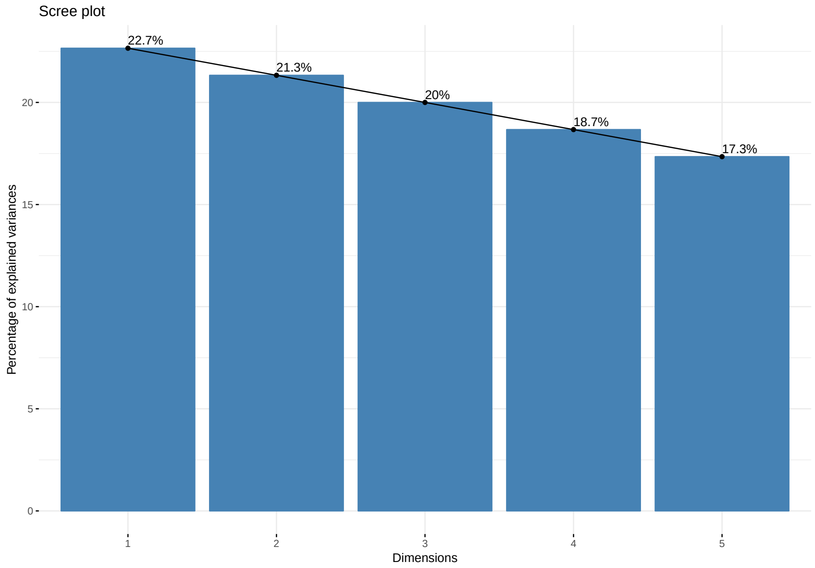
- Correspondence analysis (CA)
  - Think of it as a χ^2 PCA
- Designed to handle things that look like counts
  - That includes categories
  - And some other things

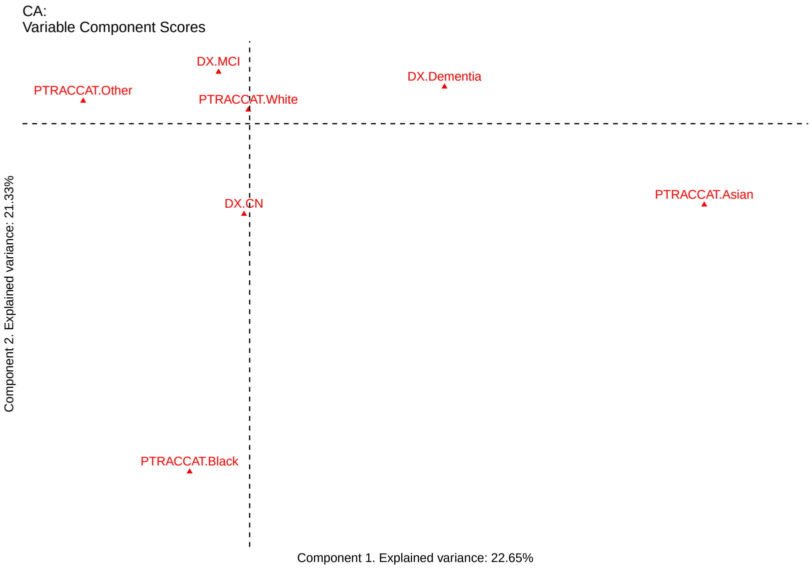
A better idea

- Correspondence analysis (CA)
  - Think of it as a χ^2 PCA
- Designed to handle things that look like counts
  - That includes categories
  - And some other things
- Row and column component scores exist on same scale

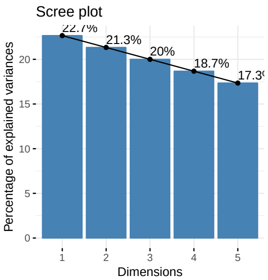
A better idea

- Correspondence analysis (CA)
  - Think of it as a χ^2 PCA
- Designed to handle things that look like counts
  - That includes categories
  - And some other things
- Row and column component scores exist on same scale
  - CA is a *bivariate* technique


Scree plot
22.7%
21.3%
20%
18.7%
17.3%
Percentage of explained variances
0
5
10
15
20
1
2
3
4
5
Dimensions


CA:
Variable Component Scores
DX.MCI
DX.Dementia
PTRACCAT.Other
PTRACCAT.White
DX.CN
PTRACCAT.Asian
PTRACCAT.Black
Component 2. Explained variance: 21.33%
Component 1. Explained variance: 22.65%

	DX	PTRACCAT
5023	CN	Asian
5026	MCI	White
5027	Dementia	White
5028	Dementia	White
5031	MCI	White
5037	Dementia	Black
5040	CN	Black
5047	MCI	Black
5054	Dementia	White
5058	Dementia	Asian
5063	Dementia	White



	DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
5023	0	1	0	0	0	0	1
5026	1	0	0	1	0	0	0
5027	0	0	1	1	0	0	0
5028	0	0	1	1	0	0	0
5031	1	0	0	1	0	0	0
5037	0	0	1	0	0	1	0
5040	0	1	0	0	0	1	0
5047	1	0	0	0	0	1	0
5054	0	0	1	1	0	0	0
5058	0	0	1	0	0	0	1
5063	0	0	1	1	0	0	0

Multiple correspondence analysis

- An extension of CA

Multiple correspondence analysis

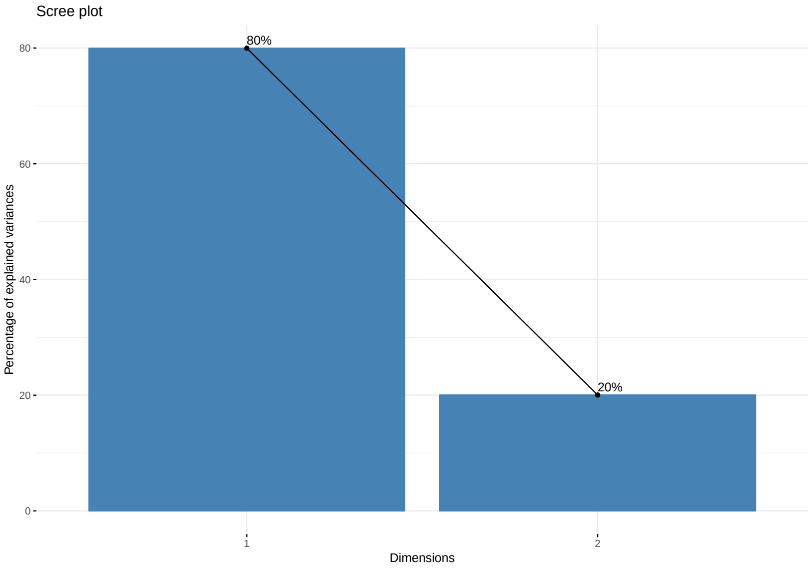
- An extension of CA
- Accomodates multiple categorical variables (CA only does 2)

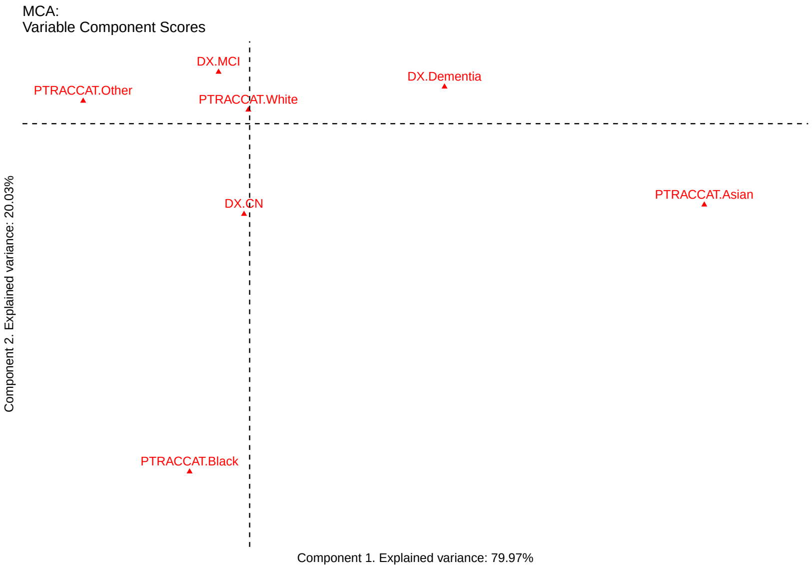
Multiple correspondence analysis

- An extension of CA
- Accomodates multiple categorical variables (CA only does 2)
- Corrects the dimensionality

Multiple correspondence analysis

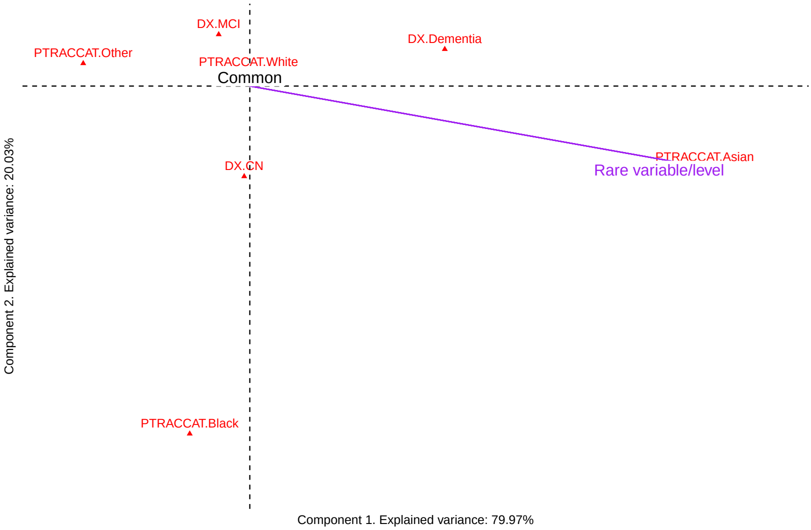
- An extension of CA
- Accomodates multiple categorical variables (CA only does 2)
- Corrects the dimensionality
- Has nearly magical properties (we'll see later)

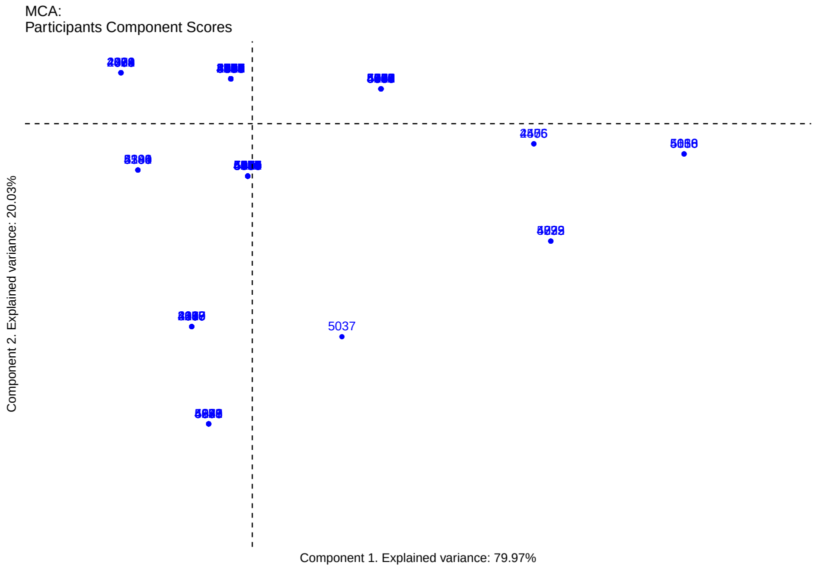

Scree plot
Percentage of explained variances
Dimensions
80%
20%
0
20
40
60
80
1
2


MCA:
Variable Component Scores
DX.MCI
DX.Dementia
PTRACCAT.Other
PTRACCAT.White
PTRACCAT.Asian
DX.CN
PTRACCAT.Black
Component 2. Explained variance: 20.03%
Component 1. Explained variance: 79.97%

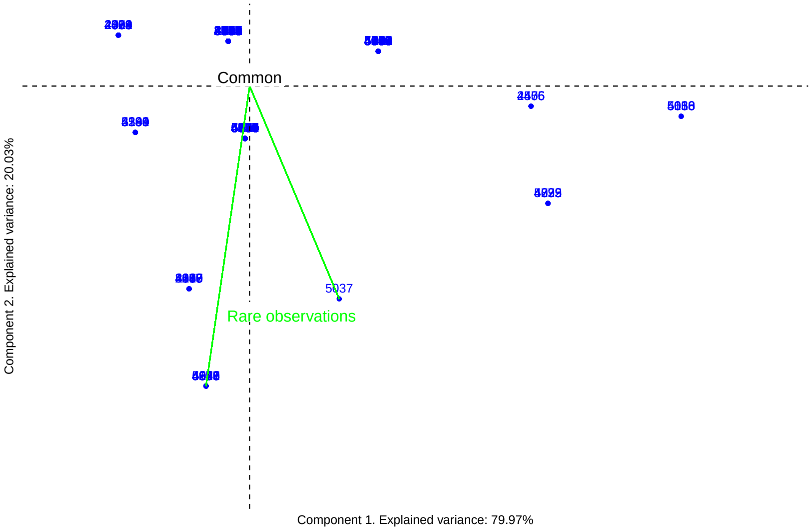
New interpretations

MCA:
Variable Component Scores




MCA:
Participants Component Scores
5037
Component 1. Explained variance: 79.97%
Component 2. Explained variance: 20.03%

MCA:
Participants Component Scores

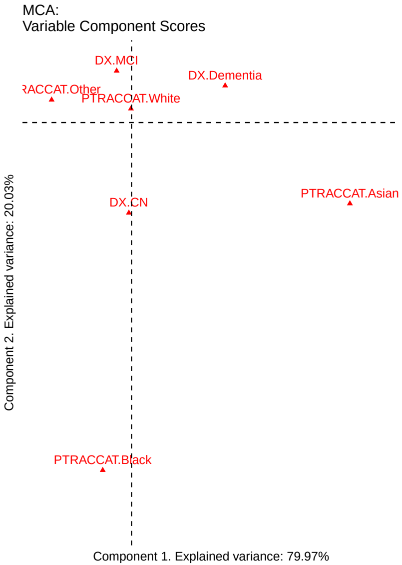


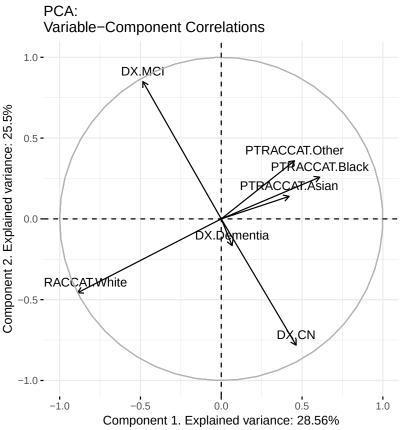
Why does it look like that?

DX.MCI	DX.CN	DX.Dementia	PTRACCAT.White	PTRACCAT.Other	PTRACCAT.Black	PTRACCAT.Asian
1	0	0	1	0	0	0
1	0	0	0	1	0	0
1	0	0	0	0	1	0
0	1	0	0	1	0	0
1	0	0	0	0	0	1
0	1	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	0	0	0
0	1	0	0	0	0	1
0	0	1	0	0	0	1
0	0	1	0	0	1	0

These are *all* the possible combinations from all 665

Compare the results





Compare the results

	PCA Comp. 1	PCA Comp. 2	PCA Comp. 3	PCA Comp. 4	PCA Comp. 5
MCA Comp. 1	0.17	-0.25	0.92	0.06	-0.26
MCA Comp. 2	-0.78	0.36	0.28	-0.42	0.03

- CA & MCA produce identical results, except MCA:

Compare the results

	PCA Comp. 1	PCA Comp. 2	PCA Comp. 3	PCA Comp. 4	PCA Comp. 5
MCA Comp. 1	0.17	-0.25	0.92	0.06	-0.26
MCA Comp. 2	-0.78	0.36	0.28	-0.42	0.03

- CA & MCA produce identical results, except MCA:
  - Drops components

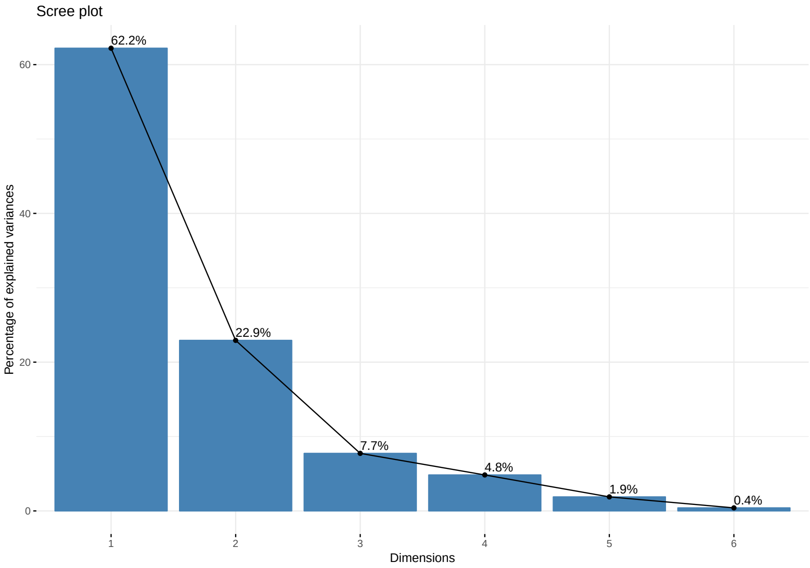
Compare the results

	PCA Comp. 1	PCA Comp. 2	PCA Comp. 3	PCA Comp. 4	PCA Comp. 5
MCA Comp. 1	0.17	-0.25	0.92	0.06	-0.26
MCA Comp. 2	-0.78	0.36	0.28	-0.42	0.03

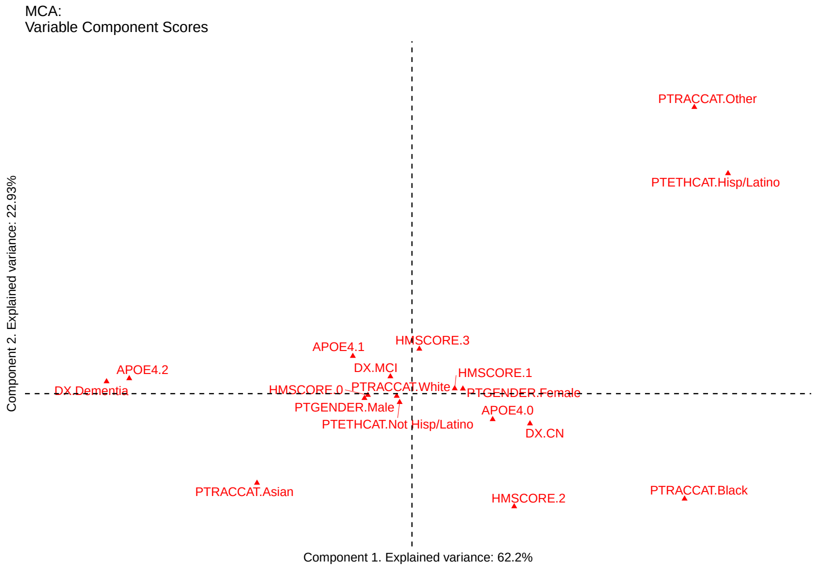
- CA & MCA produce identical results, except MCA:
  - Drops components
  - Corrects explained variance

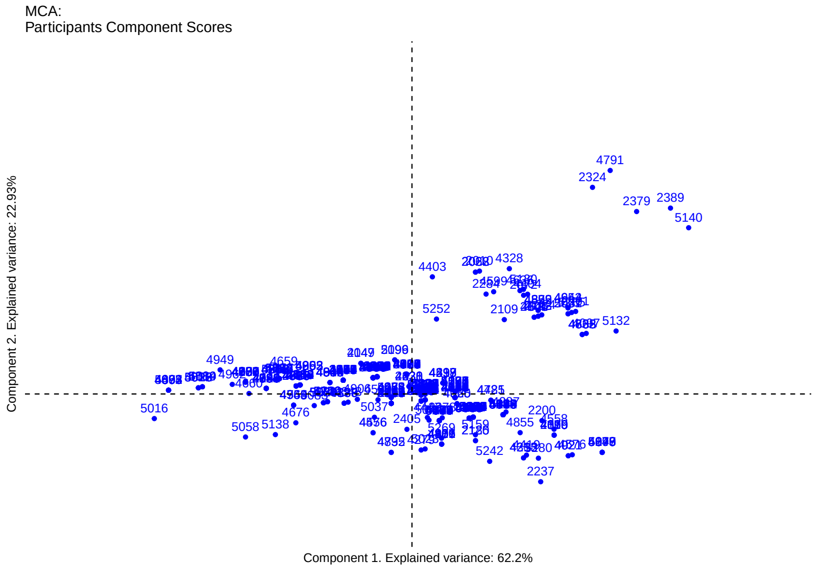
Scaling up

	DX	PTGENDER	PTETHCAT	PTRACCAT	APOE4	HMSCORE
5023	CN	Female	Not Hisp/Latino	Asian	0	0
5026	MCI	Female	Not Hisp/Latino	White	1	1
5027	Dementia	Male	Not Hisp/Latino	White	0	1
5028	Dementia	Male	Not Hisp/Latino	White	2	1
5031	MCI	Female	Hisp/Latino	White	0	1
5037	Dementia	Male	Not Hisp/Latino	Black	1	1
5040	CN	Female	Not Hisp/Latino	Black	0	1
5047	MCI	Female	Not Hisp/Latino	Black	2	1
5054	Dementia	Female	Not Hisp/Latino	White	1	0
5058	Dementia	Male	Not Hisp/Latino	Asian	0	0
5063	Dementia	Female	Not Hisp/Latino	White	1	1

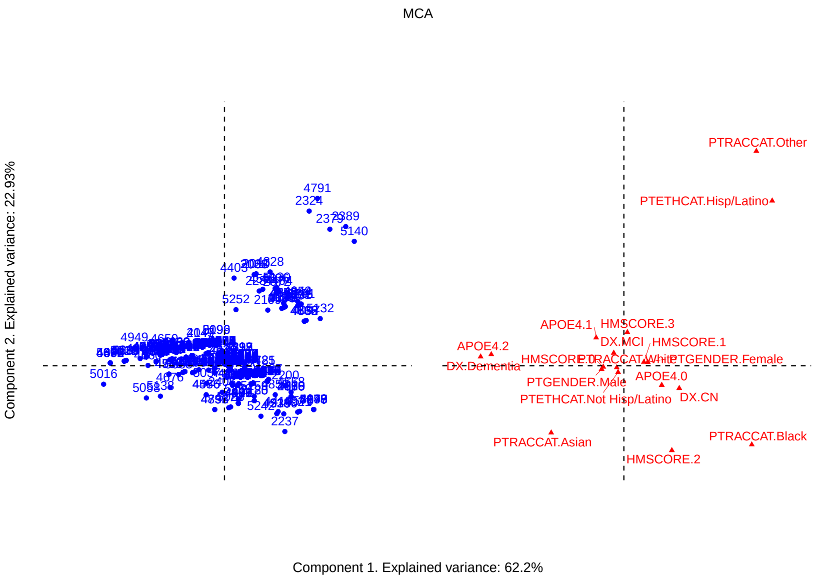


Scree plot
Percentage of explained variances
Dimensions
62.2%
22.9%
7.7%
4.8%
1.9%
0.4%
0
20
40
60
1
2
3
4
5
6


MCA:
Variable Component Scores
Component 2. Explained variance: 22.93%
Component 1. Explained variance: 62.2%
PTRACCAT.Other
PTETHCAT.Hisp/Latino
HMSCORE.3
APOE4.1
APOE4.2
DX.MCI
HMSCORE.1
DX.Dementia
HMSCORE.0
PTRACCAT.White
PTGENDER.Female
PTGENDER.Male
APOE4.0
PTETHCAT.Not Hisp/Latino
DX.CN
PTRACCAT.Asian
PTRACCAT.Black
HMSCORE.2



MCA:
Participants Component Scores
Component 2. Explained variance: 22.93%
Component 1. Explained variance: 62.2%
4791
2324
2379
2389
5140
4328
4403
5252
2109
5132
4949
5016
5058
5138
4676
5037
2405
4855
2200
5242
2237


MCA
Component 1. Explained variance: 62.2%
Component 2. Explained variance: 22.93%
PTRACCAT.Other
PTETHCAT.Hisp/Latino
APOE4.1
HMSCORE.3
APOE4.2
DX.MCI
HMSCORE.1
PTRACCAT.White
PTGENDER.Female
DX.Dementia
PTGENDER.Male
APOE4.0
PTETHCAT.Not Hisp/Latino
DX.CN
PTRACCAT.Asian
PTRACCAT.Black
HMSCORE.2
4791
2324
2389
5140
4403
5252
4949
5016
4676
2237

A very important detour

	PTGENDER	PTETHCAT
5023	Female	Not Hisp/Latino
5026	Female	Not Hisp/Latino
5027	Male	Not Hisp/Latino
5028	Male	Not Hisp/Latino
5031	Female	Hisp/Latino
5037	Male	Not Hisp/Latino
5040	Female	Not Hisp/Latino
5047	Female	Not Hisp/Latino
5054	Female	Not Hisp/Latino
5058	Male	Not Hisp/Latino
5063	Female	Not Hisp/Latino

Two variables with strictly two levels (i.e., binary data)

A very important detour

	PTGENDER.Male	PTGENDER.Female	PTETHCAT.Not Hisp/Latino	PTETHCAT.Hisp/Latino
5023	0	1	1	0
5026	0	1	1	0
5027	1	0	1	0
5028	1	0	1	0
5031	0	1	0	1
5037	1	0	1	0
5040	0	1	1	0
5047	0	1	1	0
5054	0	1	1	0
5058	1	0	1	0
5063	0	1	1	0

Disjunctive coding of two variables with strictly two levels (i.e., binary data) into four columns

A very important detour

	PTGENDER	PTETHCAT
5023	Female	Not Hisp/Latino
5026	Female	Not Hisp/Latino
5027	Male	Not Hisp/Latino
5028	Male	Not Hisp/Latino
5031	Female	Hisp/Latino
5037	Male	Not Hisp/Latino
5040	Female	Not Hisp/Latino
5047	Female	Not Hisp/Latino
5054	Female	Not Hisp/Latino
5058	Male	Not Hisp/Latino
5063	Female	Not Hisp/Latino

Two variables with strictly two levels (i.e., binary data)

A very important detour

	PTGENDER	PTETHCAT
5023	1	0
5026	1	0
5027	0	0
5028	0	0
5031	1	1
5037	0	0
5040	1	0
5047	1	0
5054	1	0
5058	0	0
5063	1	0

Binary coding of two variables with strictly two levels (i.e., binary data) in two columns

A very important detour

	PTGENDER	PTETHCAT
5023	0	1
5026	0	1
5027	1	1
5028	1	1
5031	0	0
5037	1	1
5040	0	1
5047	0	1
5054	0	1
5058	1	1
5063	0	1

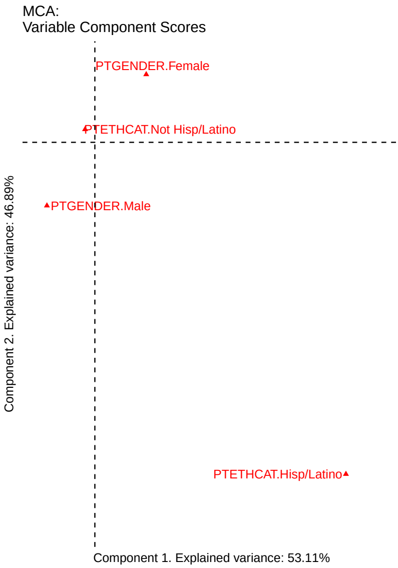
Alternate but equivalent binary coding of two variables with strictly two levels (i.e., binary data) in two columns

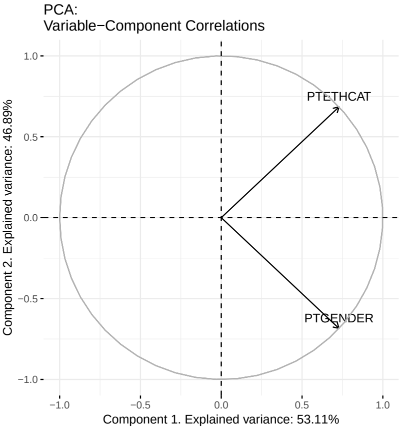
Always a bad idea?

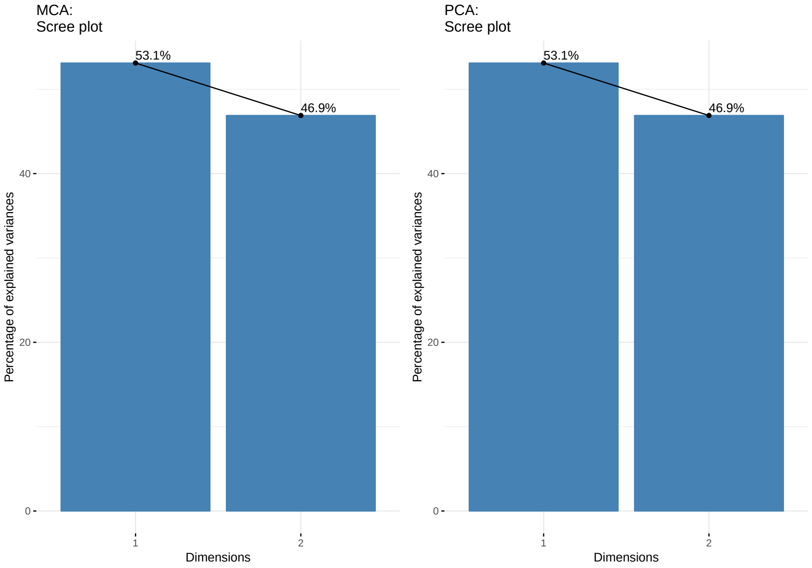
- MCA on the disjunctive coded data

Always a bad idea?

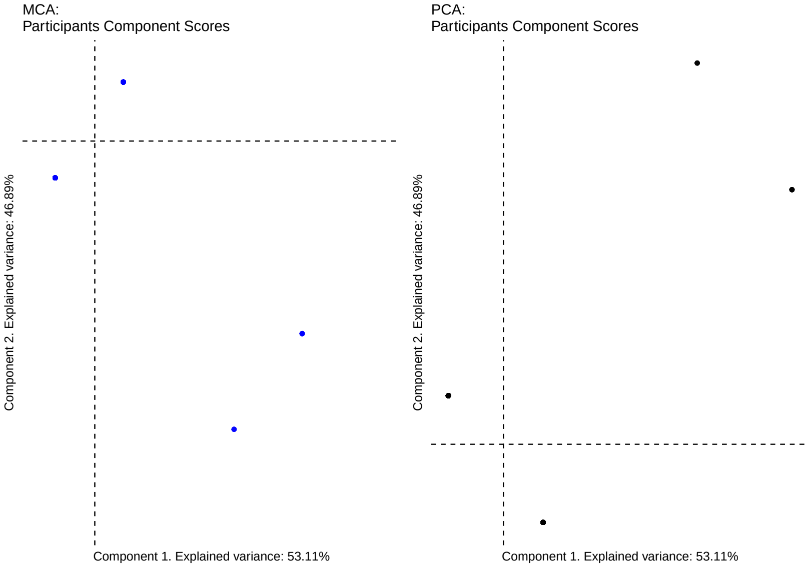
- MCA on the disjunctive coded data
- PCA on the binary coded data


MCA:
Variable Component Scores
PTGENDER.Female
PTETHCAT.Not Hisp/Latino
PTGENDER.Male
PTETHCAT.Hisp/Latino
Component 2. Explained variance: 46.89%
Component 1. Explained variance: 53.11%


PCA:
Variable−Component Correlations
PTETHCAT
PTGENDER
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
Component 2. Explained variance: 46.89%
Component 1. Explained variance: 53.11%


MCA:
Scree plot
53.1%
46.9%
40
20
0
Percentage of explained variances
1
2
Dimensions
PCA:
Scree plot
53.1%
46.9%
40
20
0
Percentage of explained variances
1
2
Dimensions

Oh, weird!



Component 2 is "flipped"

We will revisit this

	PCA Comp. 1	PCA Comp. 2
MCA Comp. 1	1	0
MCA Comp. 2	0	-1

Oh, double weird!

Let's get weird

	PTGENDER	PTETHCAT
PTGENDER	1.00	0.06
PTETHCAT	0.06	1.00

Let's get weird

	PTGENDER	PTETHCAT
PTGENDER	1.00	0.06
PTETHCAT	0.06	1.00

- $\phi = 0.06$

Let's get weird

	PTGENDER	PTETHCAT
PTGENDER	1.00	0.06
PTETHCAT	0.06	1.00

- $\phi = 0.06$
- Deep connections between χ^2, Normal, binomial (and others)

Let's get weird

	PTGENDER	PTETHCAT
PTGENDER	1.00	0.06
PTETHCAT	0.06	1.00

- $\phi = 0.06$
- Deep connections between χ^2, Normal, binomial (and others)
- We can expand the idea of "binary" or "binomial"

An old friend

	mPACCtrailsB	FDG
5023	1.12	0.13
5026	0.46	-1.31
5027	-2.77	-1.48
5028	-1.59	-0.97
5031	-0.92	-0.87
5037	-1.86	-2.00
5040	0.94	-0.21
5047	-0.25	3.05
5054	-0.80	-1.05
5058	-1.12	-2.13
5063	-2.31	-2.49

We perform(ed) PCA on these data

Escofier's Geometric Magic

- One of the “fuzzy” or “bipolar” coding schemes

Escofier's Geometric Magic

- One of the "fuzzy" or "bipolar" coding schemes
- Take each Z-scored continuous variable

Escofier's Geometric Magic

- One of the "fuzzy" or "bipolar" coding schemes
- Take each Z-scored continuous variable
- Duplicate it as $\left[\frac{1-Z}{2}\,\frac{1+Z}{2}\right]$

Escofier's Geometric Magic

	mPACCtrailsB-	mPACCtrailsB+	FDG-	FDG+
5023	-0.06	1.06	0.43	0.57
5026	0.27	0.73	1.16	-0.16
5027	1.88	-0.88	1.24	-0.24
5028	1.30	-0.30	0.98	0.02
5031	0.96	0.04	0.93	0.07
5037	1.43	-0.43	1.50	-0.50
5040	0.03	0.97	0.60	0.40
5047	0.62	0.38	-1.03	2.03
5054	0.90	0.10	1.03	-0.03
5058	1.06	-0.06	1.57	-0.57
5063	1.66	-0.66	1.74	-0.74

Escofier's Geometric Magic

	mPACCtrailsB-	mPACCtrailsB+	FDG-	FDG+
5023	-0.06	1.06	0.43	0.57
5026	0.27	0.73	1.16	-0.16
5027	1.88	-0.88	1.24	-0.24
5028	1.30	-0.30	0.98	0.02
5031	0.96	0.04	0.93	0.07
5037	1.43	-0.43	1.50	-0.50
5040	0.03	0.97	0.60	0.40
5047	0.62	0.38	-1.03	2.03
5054	0.90	0.10	1.03	-0.03
5058	1.06	-0.06	1.57	-0.57
5063	1.66	-0.66	1.74	-0.74

- Row sums are total number of *original* variables

Escofier's Geometric Magic

	mPACCtrailsB-	mPACCtrailsB+	FDG-	FDG+
5023	-0.06	1.06	0.43	0.57
5026	0.27	0.73	1.16	-0.16
5027	1.88	-0.88	1.24	-0.24
5028	1.30	-0.30	0.98	0.02
5031	0.96	0.04	0.93	0.07
5037	1.43	-0.43	1.50	-0.50
5040	0.03	0.97	0.60	0.40
5047	0.62	0.38	-1.03	2.03
5054	0.90	0.10	1.03	-0.03
5058	1.06	-0.06	1.57	-0.57
5063	1.66	-0.66	1.74	-0.74

- Row sums are total number of *original* variables
- Sum within a variable (e.g., FDG) is total number of rows

Escofier's Geometric Magic

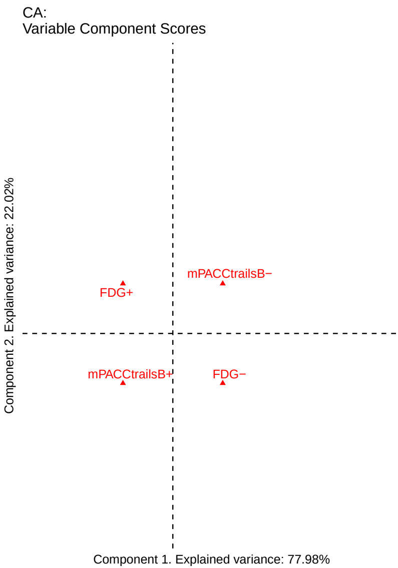
	mPACCtrailsB-	mPACCtrailsB+	FDG-	FDG+
5023	-0.06	1.06	0.43	0.57
5026	0.27	0.73	1.16	-0.16
5027	1.88	-0.88	1.24	-0.24
5028	1.30	-0.30	0.98	0.02
5031	0.96	0.04	0.93	0.07
5037	1.43	-0.43	1.50	-0.50
5040	0.03	0.97	0.60	0.40
5047	0.62	0.38	-1.03	2.03
5054	0.90	0.10	1.03	-0.03
5058	1.06	-0.06	1.57	-0.57
5063	1.66	-0.66	1.74	-0.74

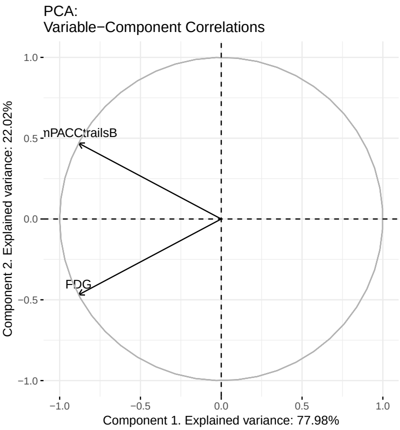
- Row sums are total number of *original* variables
- Sum within a variable (e.g., FDG) is total number of rows
- Sum of the table is rows $\times$ columns

Escofier's Geometric Magic

	mPACCtrailsB-	mPACCtrailsB+	FDG-	FDG+
5023	-0.06	1.06	0.43	0.57
5026	0.27	0.73	1.16	-0.16
5027	1.88	-0.88	1.24	-0.24
5028	1.30	-0.30	0.98	0.02
5031	0.96	0.04	0.93	0.07
5037	1.43	-0.43	1.50	-0.50
5040	0.03	0.97	0.60	0.40
5047	0.62	0.38	-1.03	2.03
5054	0.90	0.10	1.03	-0.03
5058	1.06	-0.06	1.57	-0.57
5063	1.66	-0.66	1.74	-0.74

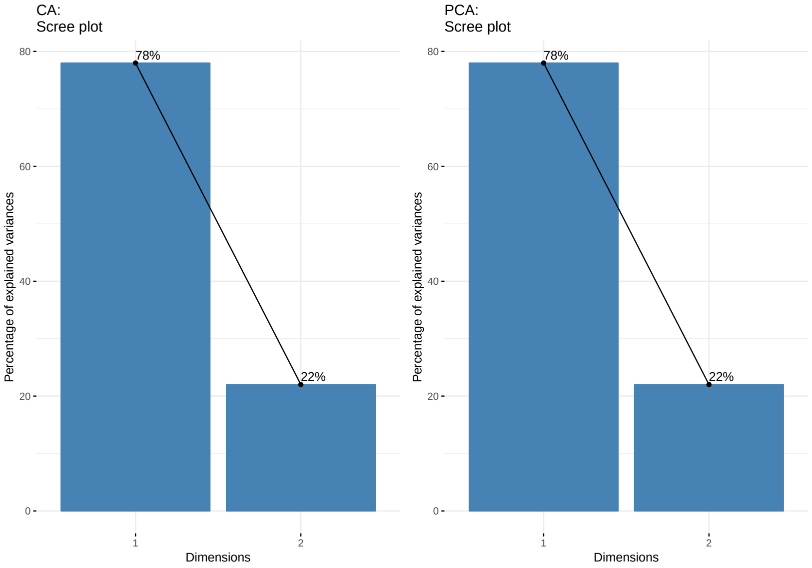
- Row sums are total number of *original* variables
- Sum within a variable (e.g., FDG) is total number of rows
- Sum of the table is rows $\times$ columns
- *These behave like disjunctive data!*



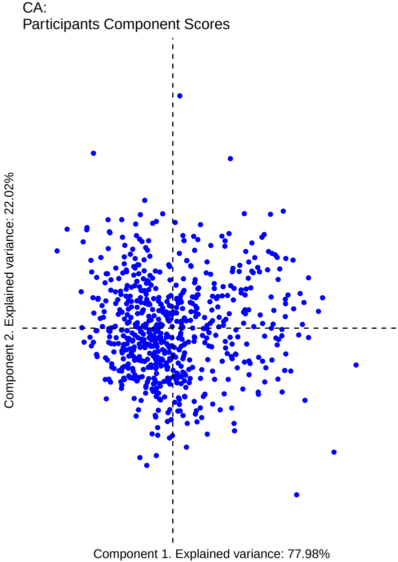


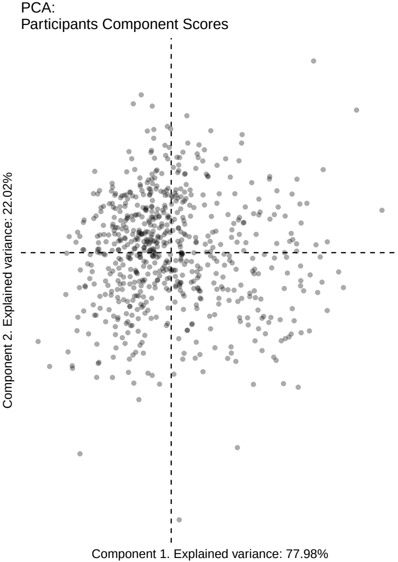
Oh, interesting!

Take note: each variable has two "poles"

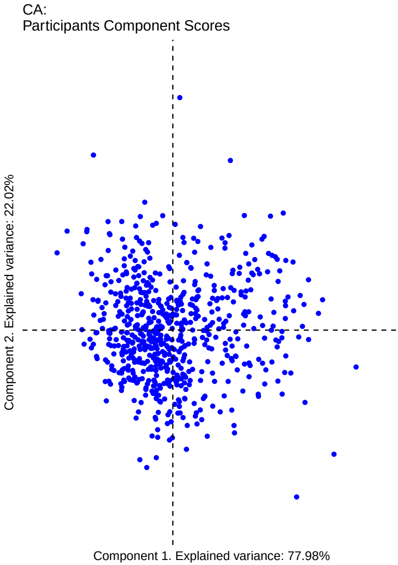

CA:
Scree plot
PCA:
Scree plot
Percentage of explained variances
78%
22%
0
20
40
60
80
1
2
Dimensions

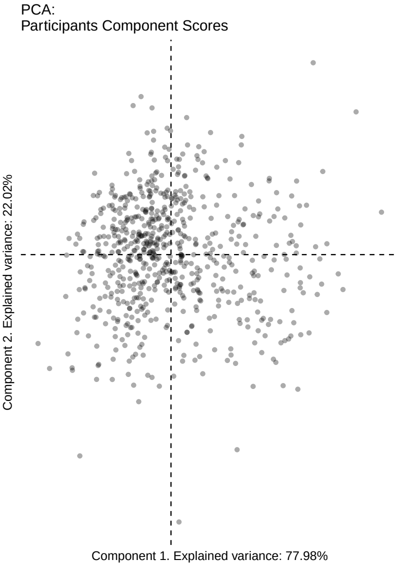
Oh, weird!





Oh, double weird!





Flips: They don't matter.

	PCA Comp. 1	PCA Comp. 2
CA Comp. 1	1	0
CA Comp. 2	0	-1

Flips: They don't matter.

Escofier's Geometric Trick

- Apply PCA to continuous data or

Escofier's Geometric Trick

- Apply PCA to continuous data or
- Apply CA to “Escofier transformed” data

Thermometer

- For ordinal data

Thermometer

- For ordinal data
- Another “fuzzy” or “bipolar” coding

Thermometer

- For ordinal data
- Another “fuzzy” or “bipolar” coding
- More Escofier Geometric Magic

Thermometer

- For ordinal data
- Another “fuzzy” or “bipolar” coding
- More Escofier Geometric Magic
  - Subtract the maximum (minimum is now 0)

Thermometer

- For ordinal data
- Another "fuzzy" or "bipolar" coding
- More Escofier Geometric Magic
  - Subtract the maximum (minimum is now 0)
  - $\left[\frac{\mathtt{max}(x)-x}{\mathtt{max}}\;\frac{x-\mathtt{min}(x)}{\mathtt{max}}\right]$

Thermometer

- For ordinal data
- Another “fuzzy” or “bipolar” coding
- More Escofier Geometric Magic
  - Subtract the maximum (minimum is now 0)
  - $\left[\frac{\mathtt{max}(x)-x}{\mathtt{max}}\ \frac{x-\mathtt{min}(x)}{\mathtt{max}}\right]$
- Apply CA

More Geometric Magic

	PTEDUCAT	CDRSB	ADAS13	MOCA
5023	18	0.0	6	30
5026	18	1.5	8	24
5027	18	4.0	27	19
5028	16	3.5	20	19
5031	14	2.0	16	20
5037	16	5.0	35	17
5040	18	0.0	8	20
5047	16	1.0	17	24
5054	18	3.5	22	21
5058	20	3.0	17	21
5063	14	2.5	38	16

More Geometric Magic

	PTEDUCAT+	PTEDUCAT-	CDRSB+	CDRSB-	ADAS13+	ADAS13-	MOCA+	MOCA-
5023	0.75	0.25	0.00	1.00	0.13	0.87	1.00	0.00
5026	0.75	0.25	0.27	0.73	0.17	0.83	0.57	0.43
5027	0.75	0.25	0.73	0.27	0.59	0.41	0.21	0.79
5028	0.50	0.50	0.64	0.36	0.43	0.57	0.21	0.79
5031	0.25	0.75	0.36	0.64	0.35	0.65	0.29	0.71
5037	0.50	0.50	0.91	0.09	0.76	0.24	0.07	0.93
5040	0.75	0.25	0.00	1.00	0.17	0.83	0.29	0.71
5047	0.50	0.50	0.18	0.82	0.37	0.63	0.57	0.43
5054	0.75	0.25	0.64	0.36	0.48	0.52	0.36	0.64
5058	1.00	0.00	0.55	0.45	0.37	0.63	0.36	0.64
5063	0.25	0.75	0.45	0.55	0.83	0.17	0.00	1.00

More Geometric Magic

	PTEDUCAT+	PTEDUCAT-	CDRSB+	CDRSB-	ADAS13+	ADAS13-	MOCA+	MOCA-
5023	0.75	0.25	0.00	1.00	0.13	0.87	1.00	0.00
5026	0.75	0.25	0.27	0.73	0.17	0.83	0.57	0.43
5027	0.75	0.25	0.73	0.27	0.59	0.41	0.21	0.79
5028	0.50	0.50	0.64	0.36	0.43	0.57	0.21	0.79
5031	0.25	0.75	0.36	0.64	0.35	0.65	0.29	0.71
5037	0.50	0.50	0.91	0.09	0.76	0.24	0.07	0.93
5040	0.75	0.25	0.00	1.00	0.17	0.83	0.29	0.71
5047	0.50	0.50	0.18	0.82	0.37	0.63	0.57	0.43
5054	0.75	0.25	0.64	0.36	0.48	0.52	0.36	0.64
5058	1.00	0.00	0.55	0.45	0.37	0.63	0.36	0.64
5063	0.25	0.75	0.45	0.55	0.83	0.17	0.00	1.00

- Row sums are total number of *original* variables

More Geometric Magic

	PTEDUCAT+	PTEDUCAT-	CDRSB+	CDRSB-	ADAS13+	ADAS13-	MOCA+	MOCA-
5023	0.75	0.25	0.00	1.00	0.13	0.87	1.00	0.00
5026	0.75	0.25	0.27	0.73	0.17	0.83	0.57	0.43
5027	0.75	0.25	0.73	0.27	0.59	0.41	0.21	0.79
5028	0.50	0.50	0.64	0.36	0.43	0.57	0.21	0.79
5031	0.25	0.75	0.36	0.64	0.35	0.65	0.29	0.71
5037	0.50	0.50	0.91	0.09	0.76	0.24	0.07	0.93
5040	0.75	0.25	0.00	1.00	0.17	0.83	0.29	0.71
5047	0.50	0.50	0.18	0.82	0.37	0.63	0.57	0.43
5054	0.75	0.25	0.64	0.36	0.48	0.52	0.36	0.64
5058	1.00	0.00	0.55	0.45	0.37	0.63	0.36	0.64
5063	0.25	0.75	0.45	0.55	0.83	0.17	0.00	1.00

- Row sums are total number of *original* variables
- Sum within a variable (e.g. EDU) is total number of rows

More Geometric Magic

	PTEDUCAT+	PTEDUCAT-	CDRSB+	CDRSB-	ADAS13+	ADAS13-	MOCA+	MOCA-
5023	0.75	0.25	0.00	1.00	0.13	0.87	1.00	0.00
5026	0.75	0.25	0.27	0.73	0.17	0.83	0.57	0.43
5027	0.75	0.25	0.73	0.27	0.59	0.41	0.21	0.79
5028	0.50	0.50	0.64	0.36	0.43	0.57	0.21	0.79
5031	0.25	0.75	0.36	0.64	0.35	0.65	0.29	0.71
5037	0.50	0.50	0.91	0.09	0.76	0.24	0.07	0.93
5040	0.75	0.25	0.00	1.00	0.17	0.83	0.29	0.71
5047	0.50	0.50	0.18	0.82	0.37	0.63	0.57	0.43
5054	0.75	0.25	0.64	0.36	0.48	0.52	0.36	0.64
5058	1.00	0.00	0.55	0.45	0.37	0.63	0.36	0.64
5063	0.25	0.75	0.45	0.55	0.83	0.17	0.00	1.00

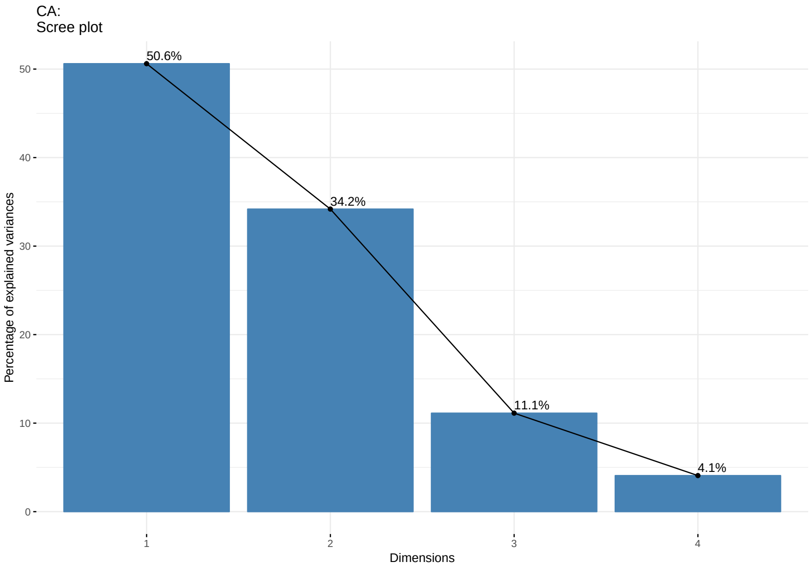
- Row sums are total number of *original* variables
- Sum within a variable (e.g. EDU) is total number of rows
- Sum of the table is rows $\times$ columns

More Geometric Magic

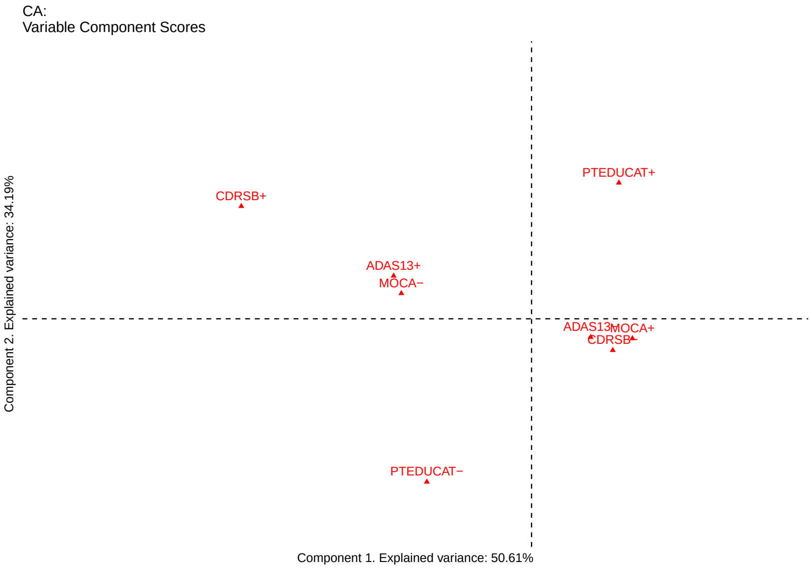
	PTEDUCAT+	PTEDUCAT-	CDRSB+	CDRSB-	ADAS13+	ADAS13-	MOCA+	MOCA-
5023	0.75	0.25	0.00	1.00	0.13	0.87	1.00	0.00
5026	0.75	0.25	0.27	0.73	0.17	0.83	0.57	0.43
5027	0.75	0.25	0.73	0.27	0.59	0.41	0.21	0.79
5028	0.50	0.50	0.64	0.36	0.43	0.57	0.21	0.79
5031	0.25	0.75	0.36	0.64	0.35	0.65	0.29	0.71
5037	0.50	0.50	0.91	0.09	0.76	0.24	0.07	0.93
5040	0.75	0.25	0.00	1.00	0.17	0.83	0.29	0.71
5047	0.50	0.50	0.18	0.82	0.37	0.63	0.57	0.43
5054	0.75	0.25	0.64	0.36	0.48	0.52	0.36	0.64
5058	1.00	0.00	0.55	0.45	0.37	0.63	0.36	0.64
5063	0.25	0.75	0.45	0.55	0.83	0.17	0.00	1.00

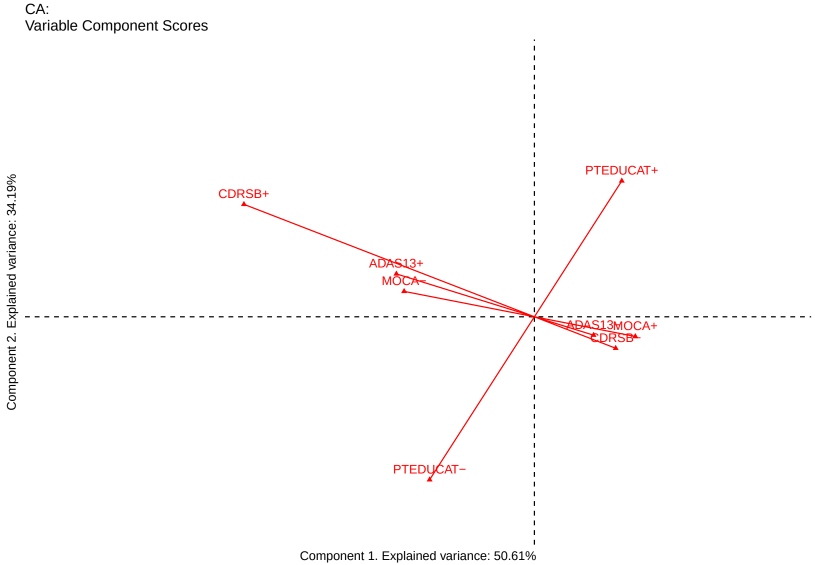
- Row sums are total number of *original* variables
- Sum within a variable (e.g. EDU) is total number of rows
- Sum of the table is rows $\times$ columns
- *These behave like disjunctive data!*

Let's take a look

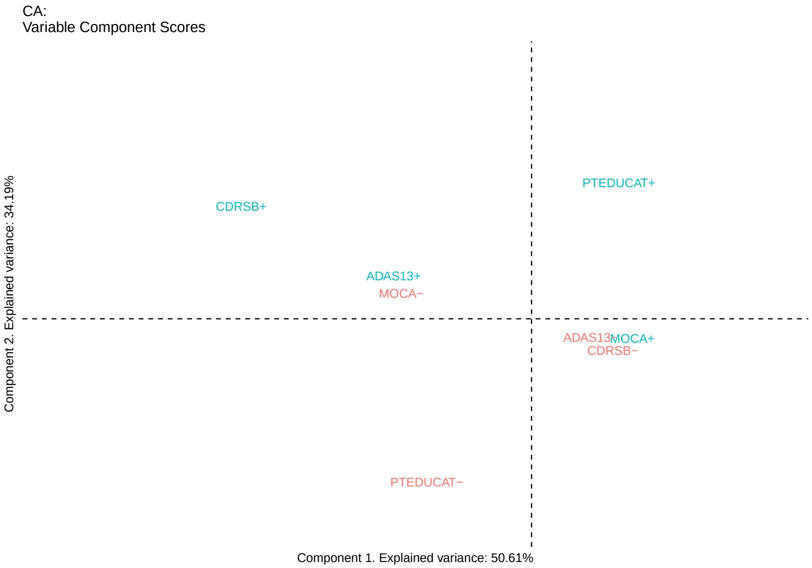


CA:
Scree plot
50.6%
34.2%
11.1%
4.1%
Percentage of explained variances
0
10
20
30
40
50
1
2
3
4
Dimensions

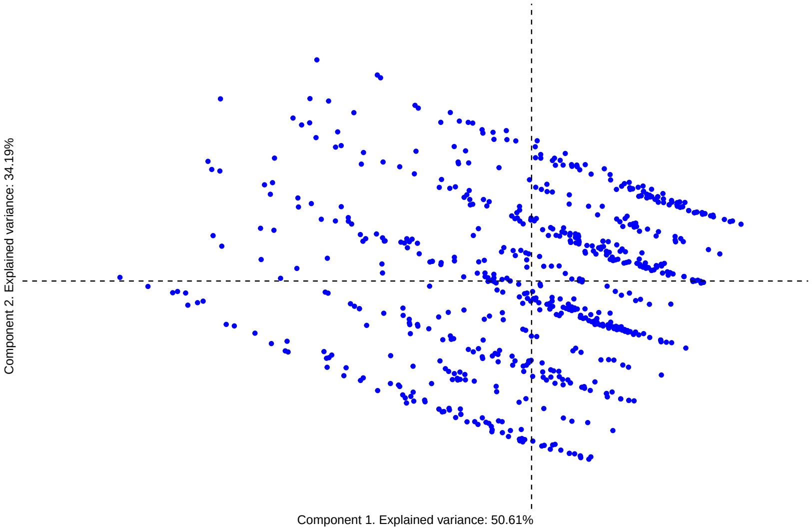

CA:
Variable Component Scores
Component 2. Explained variance: 34.19%
Component 1. Explained variance: 50.61%
PTEDUCAT+
CDRSB+
ADAS13+
MOCA−
ADAS13−
MOCA+
CDRSB−
PTEDUCAT−

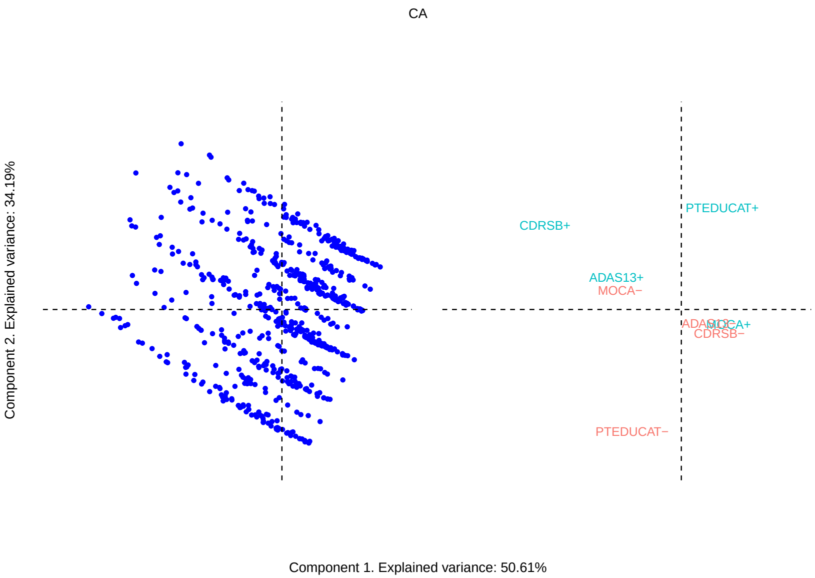


Special properties: Biploar coding passes through 0.


CA:
Variable Component Scores
PTEDUCAT+
CDRSB+
ADAS13+
MOCA−
ADAS13−
MOCA+
CDRSB−
PTEDUCAT−
Component 1. Explained variance: 50.61%
Component 2. Explained variance: 34.19%

CA:
Participants Component Scores




CA
PTEDUCAT+
CDRSB+
ADAS13+
MOCA−
ADAS13−
MOCA+
CDRSB−
PTEDUCAT−
Component 1. Explained variance: 50.61%
Component 2. Explained variance: 34.19%

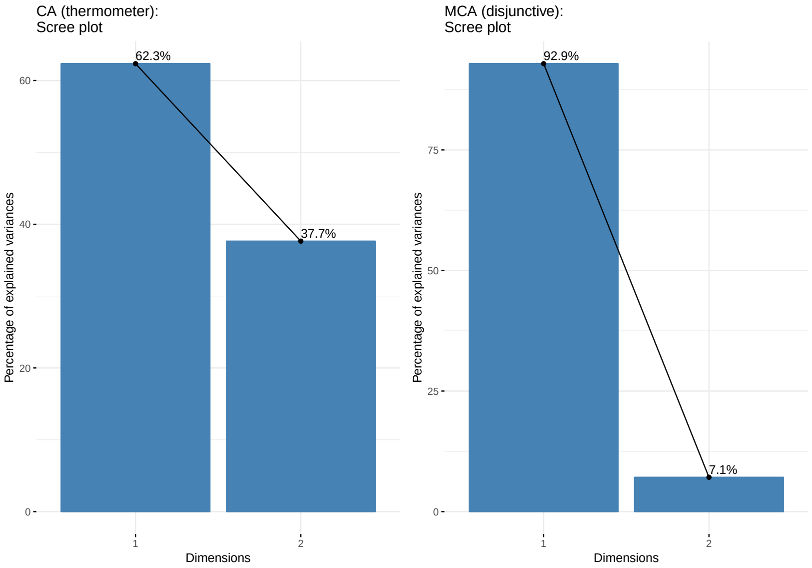
Thermometer vs. Disjunctive

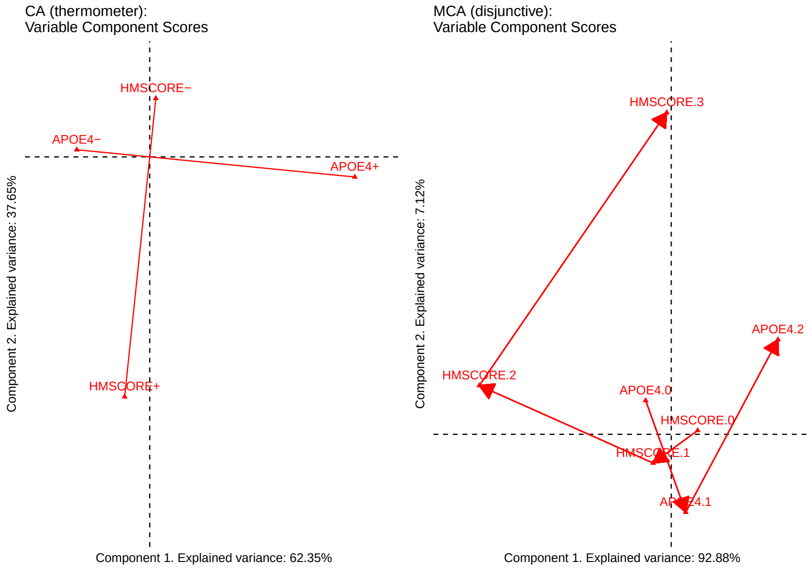
- Sometimes data could be either

Thermometer vs. Disjunctive

- Sometimes data could be either
- Let's analyze it both ways

	APOE4	HMSCORE
5023	0	0
5026	1	1
5027	0	1
5028	2	1
5031	0	1
5037	1	1
5040	0	1
5047	2	1
5054	1	0
5058	0	0
5063	1	1


CA (thermometer):
Scree plot
62.3%
37.7%
60
40
20
0
Percentage of explained variances
1
2
Dimensions
MCA (disjunctive):
Scree plot
92.9%
7.1%
75
50
25
0
Percentage of explained variances
1
2
Dimensions


CA (thermometer):
Variable Component Scores
HMSCORE−
APOE4−
APOE4+
HMSCORE+
Component 2. Explained variance: 37.65%
Component 1. Explained variance: 62.35%
MCA (disjunctive):
Variable Component Scores
HMSCORE.3
APOE4.2
HMSCORE.2
APOE4.0
HMSCORE.0
HMSCORE.1
APOE4.1
Component 2. Explained variance: 7.12%
Component 1. Explained variance: 92.88%

Thermometer vs. Disjunctive

- For a small (reasonable) number of levels: disjunctive

Thermometer vs. Disjunctive

- For a small (reasonable) number of levels: disjunctive
- Otherwise: thermometer

Thermometer vs. Disjunctive

- For a small (reasonable) number of levels: disjunctive
- Otherwise: thermometer
- Interpretation:

Thermometer vs. Disjunctive

- For a small (reasonable) number of levels: disjunctive
- Otherwise: thermometer
- Interpretation:
  - Thermometer is "easier"

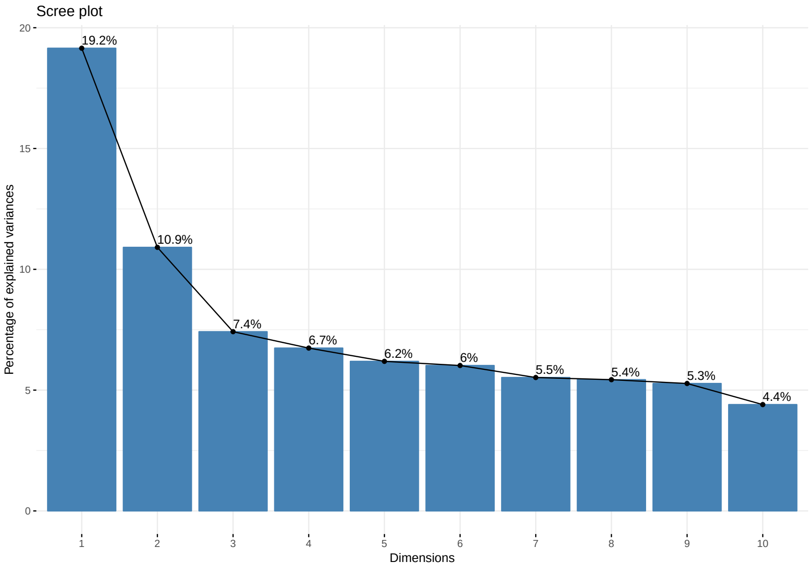
Thermometer vs. Disjunctive

- For a small (reasonable) number of levels: disjunctive
- Otherwise: thermometer
- Interpretation:
  - Thermometer is “easier”
  - Disjunctive is more informative

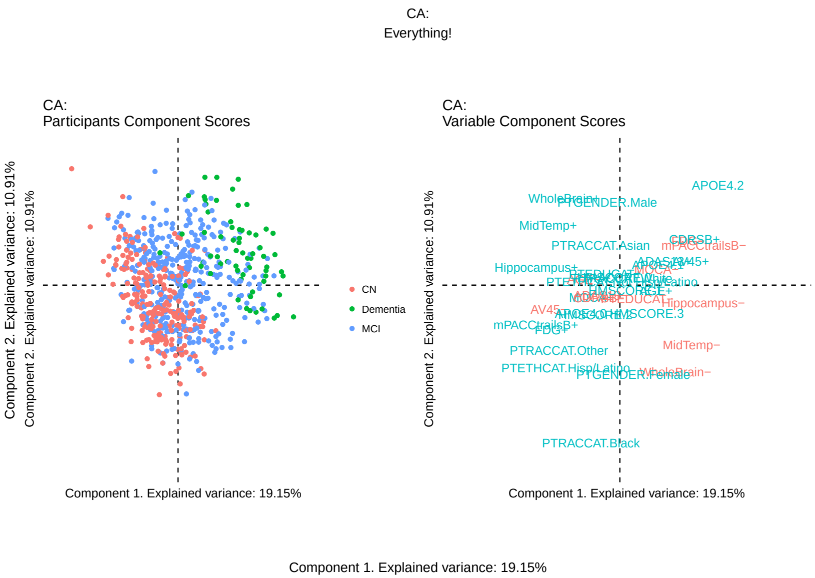
All of the data

All of the data

	DX	AGE	PTGENDER	PTEDUCAT	PTETHCAT	PTRACCAT	APOE4	FDG	AV45	CDRSB	ADAS13	MOCA	WholeBrain	Hippocampus	MidTemp	mPACCtrailsB	HMSCORE
5023	CN	63.9	Female	18	Not Hisp/Latino	Asian	0	1.29	1.03	0.0	6	30	1057351.0	7904	21306	1.81	0
5026	MCI	70.5	Female	18	Not Hisp/Latino	White	1	1.08	1.44	1.5	8	24	1023057.3	8051	16501	-1.45	1
5027	Dementia	75.5	Male	18	Not Hisp/Latino	White	0	1.06	1.44	4.0	27	19	986723.7	6534	17437	-17.27	1
5028	Dementia	61.9	Male	16	Not Hisp/Latino	White	2	1.13	1.38	3.5	20	19	1182704.6	7481	20797	-11.50	1
5031	MCI	80.2	Female	14	Hisp/Latino	White	0	1.14	1.52	2.0	16	20	908133.9	5040	19032	-8.21	1
5037	Dementia	67.3	Male	16	Not Hisp/Latino	Black	1	0.98	1.21	5.0	35	17	1161499.6	5831	21428	-12.80	1
5040	CN	75.9	Female	18	Not Hisp/Latino	Black	0	1.24	1.01	0.0	8	20	943160.6	7994	16634	0.94	1
5047	MCI	68.8	Female	16	Not Hisp/Latino	Black	2	1.70	1.48	1.0	17	24	1070406.1	7920	22043	-4.90	1
5054	Dementia	74.0	Female	18	Not Hisp/Latino	White	1	1.12	1.43	3.5	22	21	1138040.1	6580	20836	-7.63	0
5058	Dementia	61.8	Male	20	Not Hisp/Latino	Asian	0	0.97	1.54	3.0	17	21	1195549.3	7318	22757	-9.18	0
5063	Dementia	71.5	Female	14	Not Hisp/Latino	White	1	0.92	1.61	2.5	38	16	817421.2	5364	12542	-15.03	1



Scree plot
19.2%
10.9%
7.4%
6.7%
6.2%
6%
5.5%
5.4%
5.3%
4.4%
Percentage of explained variances
Dimensions
0
5
10
15
20
1
2
3
4
5
6
7
8
9
10


CA:
Everything!
CA:
Participants Component Scores
Component 2. Explained variance: 10.91%
Component 2. Explained variance: 10.91%
Component 1. Explained variance: 19.15%
CN
Dementia
MCI
CA:
Variable Component Scores
Component 2. Explained variance: 10.91%
Component 1. Explained variance: 19.15%
APOE4.2
WholeBrain+
PTGENDER.Male
MidTemp+
PTRACCAT.Asian
CDRSB+
mPACCtrailsB−
Hippocampus+
Hippocampus−
AV45
mPACCtrailsB+
FDG+
PTRACCAT.Other
MidTemp−
PTETHCAT.Hisp/Latino
PTGENDER.Female
WholeBrain−
PTRACCAT.Black
Component 1. Explained variance: 19.15%

Resampling

Background

- Generally the Gifi or Benzecri principles

Background

- Generally the Gifi or Benzecri principles
- Benzecri

Background

- Generally the Gifi or Benzecri principles
- Benzecri
  - “statistics was built a pompous discipline based on theoretical assumptions that are rarely met in practice”

Background

- Generally the Gifi or Benzecri principles
- Benzecri
  - “statistics was built a pompous discipline based on theoretical assumptions that are rarely met in practice”
  - “the models should follow the data, not vice versa”

Background

- Generally the Gifi or Benzecri principles
- Benzecri
  - “statistics was built a pompous discipline based on theoretical assumptions that are rarely met in practice”
  - “the models should follow the data, not vice versa”
  - “use the computer implies the abandonment of all the techniques designed before of computing”

Background

- Generally the Gifi or Benzecri principles
- Benzecri
  - "statistics was built a pompous discipline based on theoretical assumptions that are rarely met in practice"
  - "the models should follow the data, not vice versa"
  - "use the computer implies the abandonment of all the techniques designed before of computing"
- Gifi

Background

- Generally the Gifi or Benzecri principles
- Benzecri
  - “statistics was built a pompous discipline based on theoretical assumptions that are rarely met in practice”
  - “the models should follow the data, not vice versa”
  - “use the computer implies the abandonment of all the techniques designed before of computing”
- Gifi
  - Replication stability: new data, same techniques

Background

- Generally the Gifi or Benzecri principles
- Benzecri
  - “statistics was built a pompous discipline based on theoretical assumptions that are rarely met in practice”
  - “the models should follow the data, not vice versa”
  - “use the computer implies the abandonment of all the techniques designed before of computing”
- Gifi
  - Replication stability: new data, same techniques
  - Selection stability: Data variations

Background

- Generally the Gifi or Benzecri principles
- Benzecri
  - “statistics was built a pompous discipline based on theoretical assumptions that are rarely met in practice”
  - “the models should follow the data, not vice versa”
  - “use the computer implies the abandonment of all the techniques designed before of computing”
- Gifi
  - Replication stability: new data, same techniques
  - Selection stability: Data variations
  - Technique stability: Different technique, same data

Background

- Generally the Gifi or Benzecri principles
- Benzecri
  - “statistics was built a pompous discipline based on theoretical assumptions that are rarely met in practice”
  - “the models should follow the data, not vice versa”
  - “use the computer implies the abandonment of all the techniques designed before of computing”
- Gifi
  - Replication stability: new data, same techniques
  - Selection stability: Data variations
  - Technique stability: Different technique, same data
- Pause!

My beliefs

- Might give you inference/generalizability

My beliefs

- Might give you inference/generalizability
  - Depending on data, design, etc. . .

My beliefs

- Might give you inference/generalizability
  - Depending on data, design, etc...
- Practically

My beliefs

- Might give you inference/generalizability
  - Depending on data, design, etc. . .
- Practically
  - Assessing stability of *your* data

My beliefs

- Might give you inference/generalizability
  - Depending on data, design, etc...
- Practically
  - Assessing stability of *your* data
  - Provides critical diagnostics

Definitions

- Permutation: break relationships in the data

Definitions

- Permutation: break relationships in the data
- Split-half: mutually exclusive sets

Definitions

- Permutation: break relationships in the data
- Split-half: mutually exclusive sets
- Bootstrap: resample with reselection

Base data

	VARIABLE 1	VARIABLE 2	VARIABLE 3
OBS. 1	a	b	c
OBS. 2	d	e	f
...	...	...	...
OBS. N-1	u	v	w
OBS. N	x	y	z

Tiny illustrative data

Permutation

	VARIABLE 1	VARIABLE 2	VARIABLE 3
OBS. 1	x	e	c
OBS. 2	u	b	f
...	...	...	...
OBS. N-1	a	v	z
OBS. N	d	y	w

Tiny permuted illustrative data

Split-half

	VARIABLE 1	VARIABLE 2	VARIABLE 3
OBS. 1	a	b	c
OBS. 3	g	h	i
...	...	...	...

	VARIABLE 1	VARIABLE 2	VARIABLE 3
OBS. 42	α	π	ω
OBS. 2	d	e	f
...	...	...	...

Tiny split half illustrative data

Bootstrap

	VARIABLE 1	VARIABLE 2	VARIABLE 3
OBS. 42	α	π	ω
OBS. 42	α	π	ω
...	...	...	...
OBS. 1	a	b	c
OBS. N-1	u	v	w

Tiny bootstrap illustrative data

Uses in PCA & CA

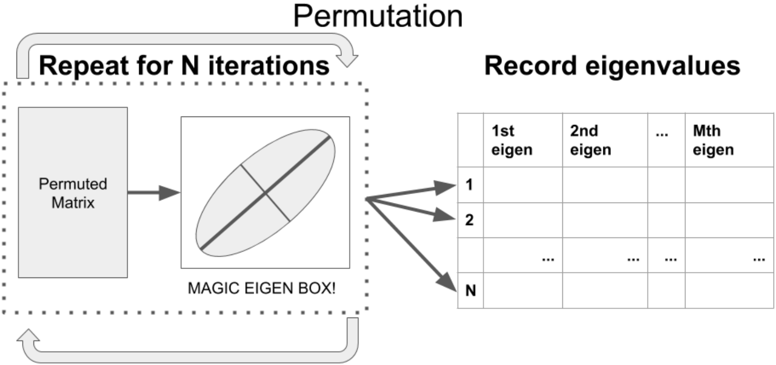
- Permutation: Effect size tests of components

Uses in PCA & CA

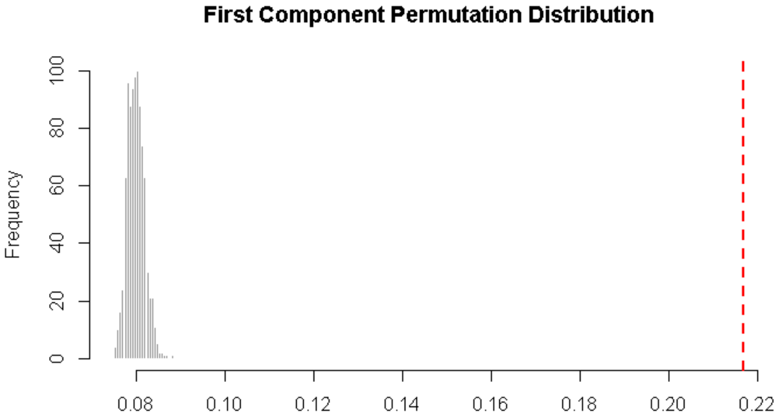
- Permutation: Effect size tests of components
- Split-half: Replication of components

Uses in PCA & CA

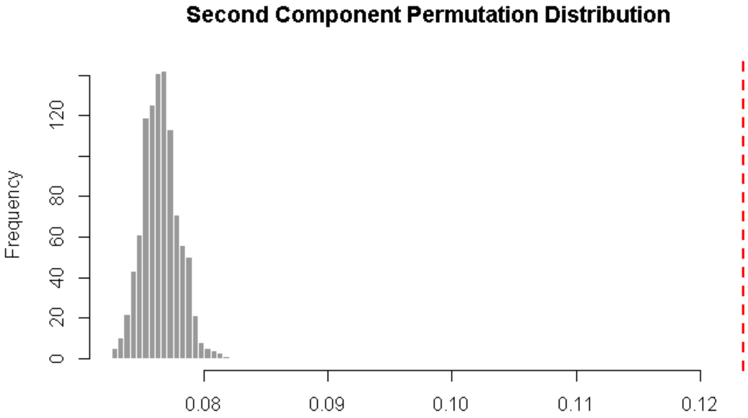
- Permutation: Effect size tests of components
- Split-half: Replication of components
- Bootstrap: Stability of variables



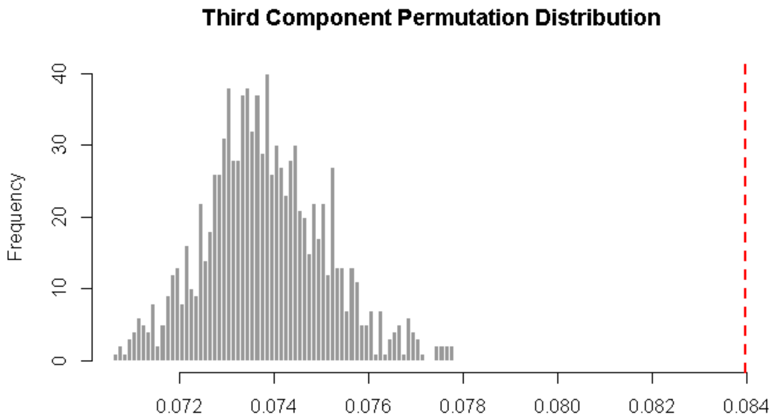
Permutation diagram



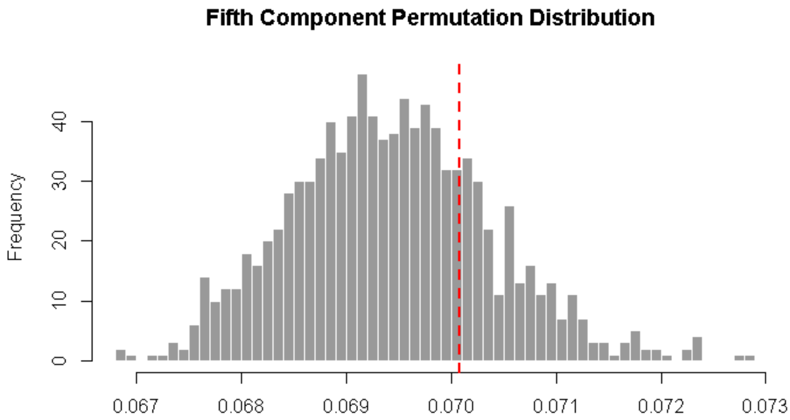
Permutation: First component



Permutation: Second component



Permutation: Third component



Permutation: Fifth component

Iterations: 1000

Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
0.001	0.001	0.001	0.001	0.236	0.255	1

Comp.8	Comp.9	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14
0.999	0.998	1	1	1	1	1

Comp.15	Comp.16	Comp.17	Comp.18	Comp.19	Comp.20	Comp.21
1	1	1	1	1	1	1

Permutation: p-values

Another detour

- p-values should (inversely) follow the scree

Another detour

- p-values should (inversely) follow the scree
- Diagnostic tests:

Another detour

- p-values should (inversely) follow the scree
- Diagnostic tests:
  - Large or erratic jumps

Another detour

- p-values should (inversely) follow the scree
- Diagnostic tests:
  - Large or erratic jumps
  - First or first few $ps \geq .5$

Conclusions

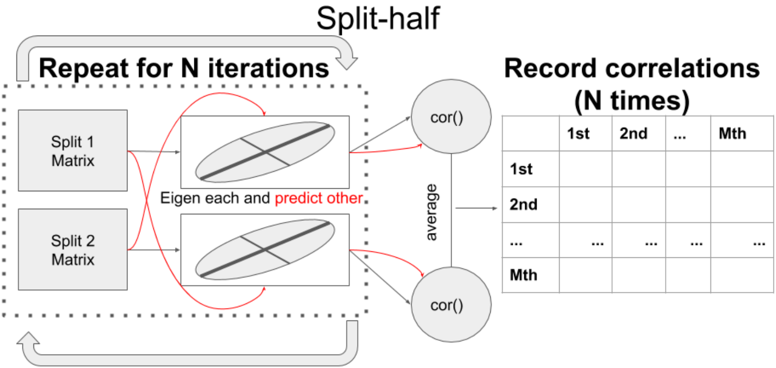
- First few components have larger than expected effect sizes

Conclusions

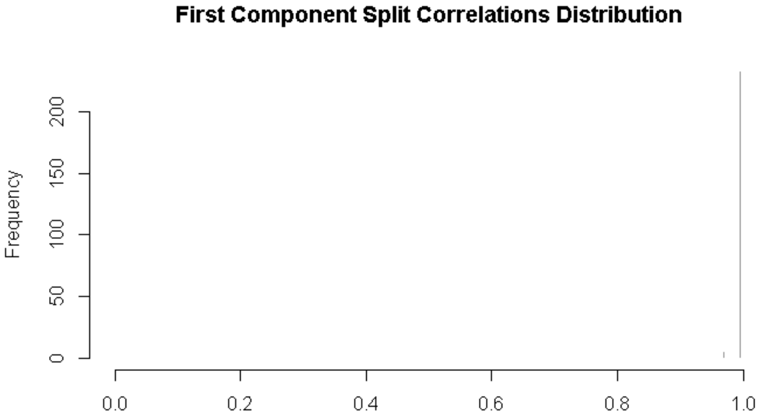
- First few components have larger than expected effect sizes
  - More variance than null

Conclusions

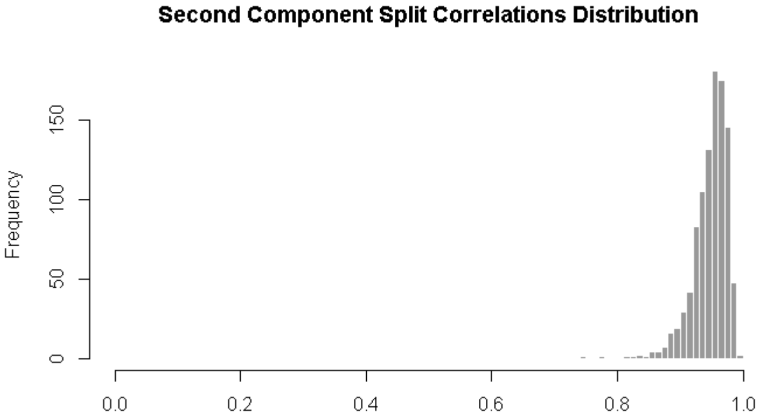
- First few components have larger than expected effect sizes
  - More variance than null
- We do not know if these generalize



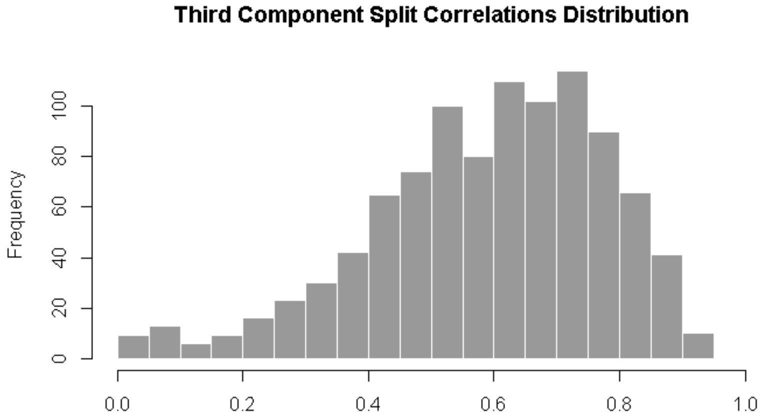
Split half diagram



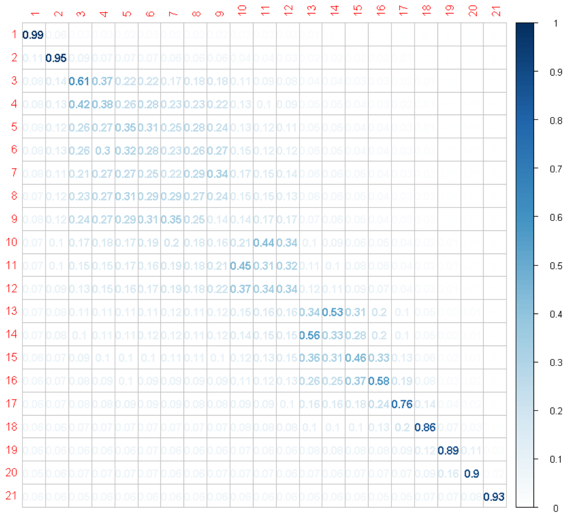
Split-half correlations: First component



Split-half correlations: Second component



Split-half correlations: Third component



Split-half correlations: All components

Conclusions

- We do *sort of* know if these generalize

Conclusions

- We do *sort of* know if these generalize
- First two really do

Conclusions

- We do *sort of* know if these generalize
- First two really do
- Next few: Maybe

Conclusions

- We do *sort of* know if these generalize
- First two really do
- Next few: Maybe
- Key observation:

Conclusions

- We do *sort of* know if these generalize
- First two really do
- Next few: Maybe
- Key observation:
  - Components *flip* order!

Conclusions

- We do *sort of* know if these generalize
- First two really do
- Next few: Maybe
- Key observation:
    - Components *flip* order!
    - We need to question the meaning of order of components in our data

Bootstrap goes last

- You need to know the number of components to interpret

Bootstrap goes last

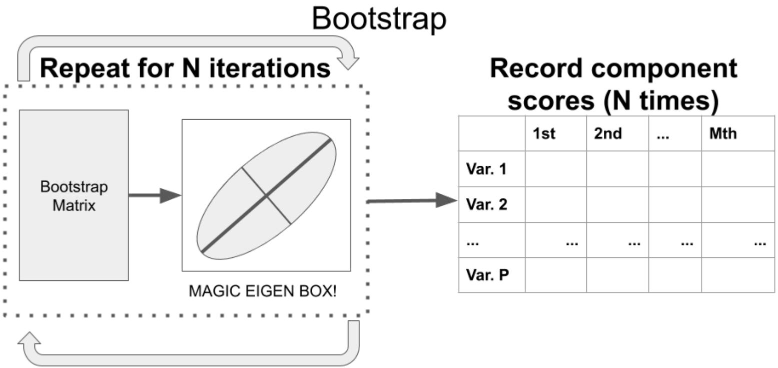
- You need to know the number of components to interpret
- We have 2

Bootstrap goes last

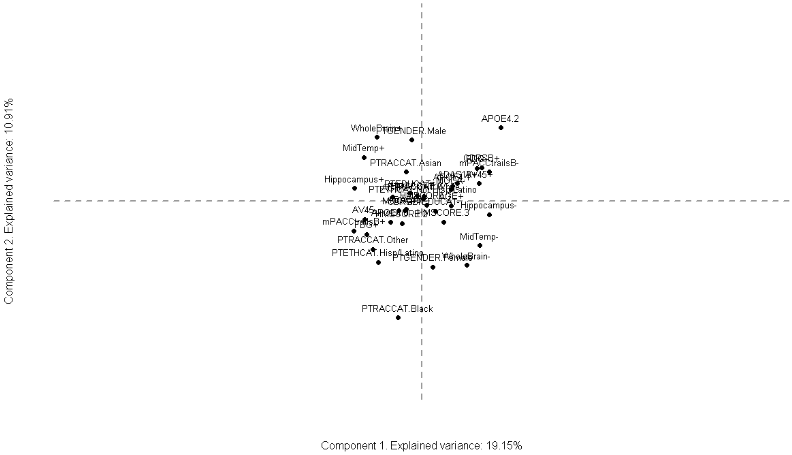
- You need to know the number of components to interpret
- We have 2
- Now you can interpret variables *per* component

Bootstrap goes last

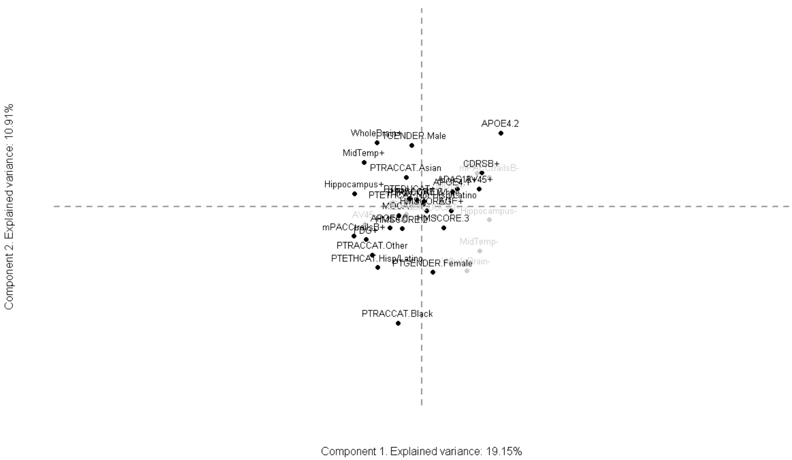
- You need to know the number of components to interpret
- We have 2
- Now you can interpret variables *per* component
  - Find the ones that are stable



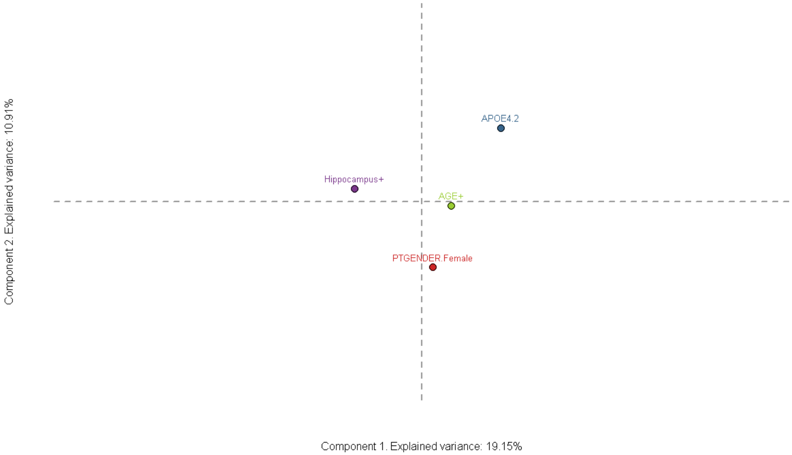
Bootstrap diagram



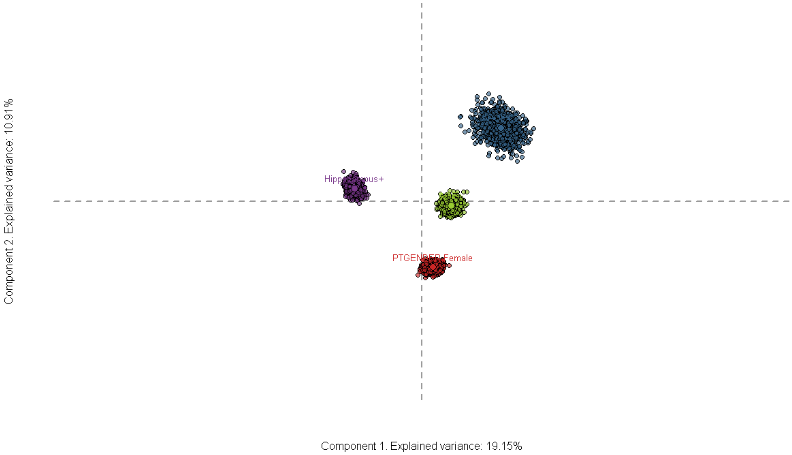
Bootstrap ratios



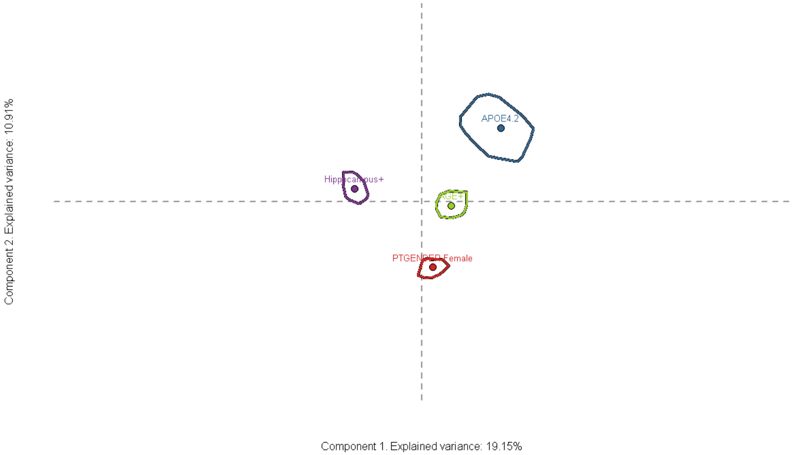
Bootstrap ratios



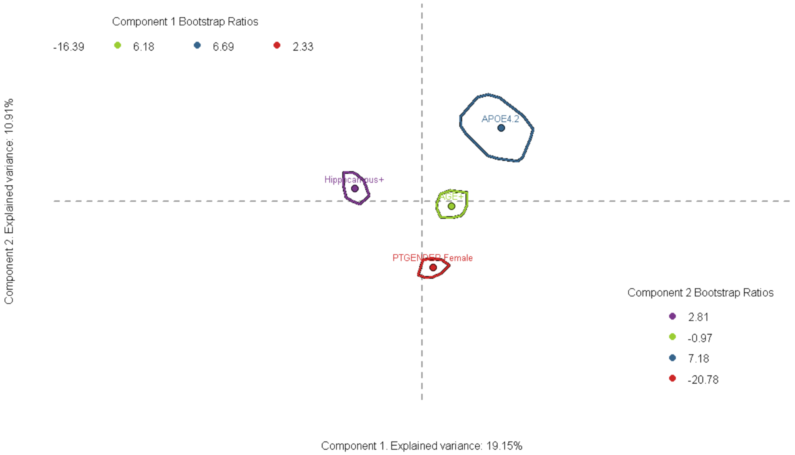
Bootstrap ratios



Bootstrap ratios



Bootstrap ratios



Bootstrap ratios

Conclusions

- Just a snapshot (there are more variables)

Conclusions

- Just a snapshot (there are more variables)
- APOE4 contributes to both

Conclusions

- Just a snapshot (there are more variables)
- APOE4 contributes to both
- The others generally contribute to one or the other

Final stretch

Things not discussed

- Corrections

Things not discussed

- Corrections
- Alternate preprocessing

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)
- Other resampling & Cross-validation loops.

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)
- Other resampling & Cross-validation loops.
  - Start at the "beginning"

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)
- Other resampling & Cross-validation loops.
  - Start at the "beginning"
- What about other data types?

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)
- Other resampling & Cross-validation loops.
  - Start at the "beginning"
- What about other data types?
  - I've actually misled you a bit

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)
- Other resampling & Cross-validation loops.
  - Start at the "beginning"
- What about other data types?
  - I've actually misled you a bit
  - Structural data

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)
- Other resampling & Cross-validation loops.
  - Start at the "beginning"
- What about other data types?
  - I've actually misled you a bit
  - Structural data
- Rotations

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)
- Other resampling & Cross-validation loops.
  - Start at the "beginning"
- What about other data types?
  - I've actually misled you a bit
  - Structural data
- Rotations
  - I don't rotate

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)
- Other resampling & Cross-validation loops.
  - Start at the "beginning"
- What about other data types?
  - I've actually misled you a bit
  - Structural data
- Rotations
  - I don't rotate
  - But I won't stop you from it

Things not discussed

- Corrections
- Alternate preprocessing
  - There's a lazy way (rank)
- Other resampling & Cross-validation loops.
  - Start at the "beginning"
- What about other data types?
  - I've actually misled you a bit
  - Structural data
- Rotations
  - I don't rotate
  - But I won't stop you from it
  - Report both

Rotation

- Two compelling examples rotation

Rotation

- Two compelling examples rotation
  - That weren't rotated

Rotation

- Two compelling examples rotation
  - That weren't rotated
  - Why?

Rotation

- Two compelling examples rotation
  - That weren't rotated
  - Why?
- CA of Mueller report

Rotation

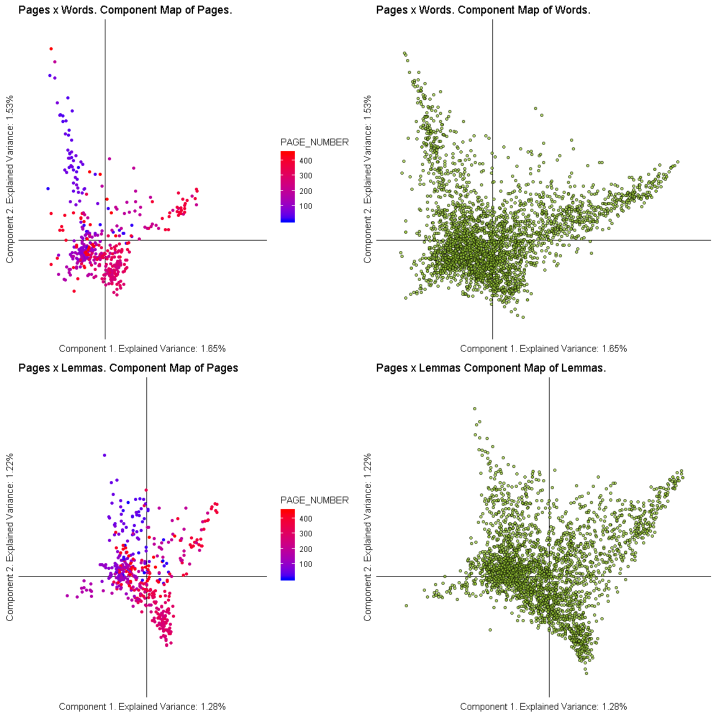
- Two compelling examples rotation
  - That weren't rotated
  - Why?
- CA of Mueller report
  - see http://github.com/derekbeaton/muellerreport_ca

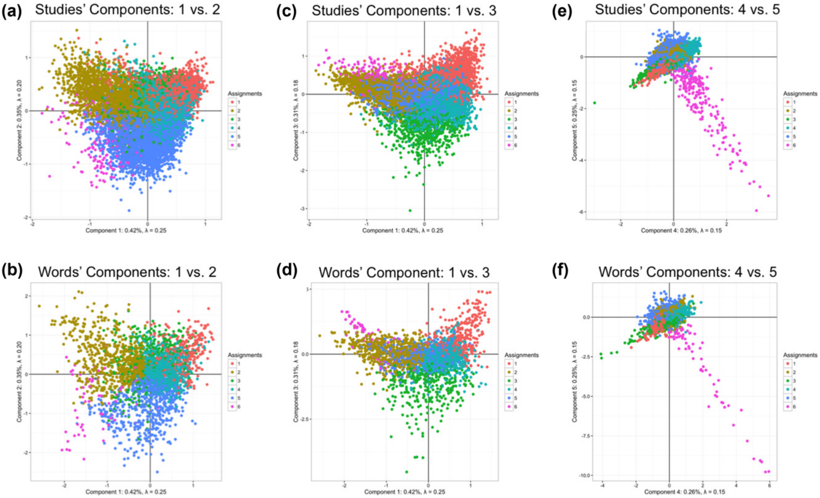
Rotation

- Two compelling examples rotation
  - That weren't rotated
  - Why?
- CA of Mueller report
  - see http://github.com/derekbeaton/muellerreport_ca
- CA of NeuroSynth (Alhazmi et al., 2018)

Rotation

- Two compelling examples rotation
  - That weren't rotated
  - Why?
- CA of Mueller report
  - see http://github.com/derekbeaton/muellerreport_ca
- CA of NeuroSynth (Alhazmi et al., 2018)
  - see http://github.com/fahd09/neurosynth_semantic_map


Pages x Words. Component Map of Pages.
Pages x Words. Component Map of Words.
Pages x Lemmas. Component Map of Pages
Pages x Lemmas Component Map of Lemmas.
Component 1. Explained Variance: 1.65%
Component 2. Explained Variance: 1.53%
Component 1. Explained Variance: 1.28%
Component 2. Explained Variance: 1.22%
PAGE_NUMBER
400
300
200
100


(a) Studies' Components: 1 vs. 2
Component 1: 0.42%, λ = 0.25
Component 2: 0.35%, λ = 0.20
Assignments
1
2
3
4
5
6
(c) Studies' Components: 1 vs. 3
Component 1: 0.42%, λ = 0.25
Component 3: 0.31%, λ = 0.18
Assignments
(e) Studies' Components: 4 vs. 5
Component 4: 0.26%, λ = 0.15
Component 5: 0.25%, λ = 0.15
Assignments
(b) Words' Components: 1 vs. 2
Component 1: 0.42%, λ = 0.25
Component 2: 0.35%, λ = 0.20
Assignments
(d) Words' Component: 1 vs. 3
Component 1: 0.42%, λ = 0.25
Component 3: 0.31%, λ = 0.18
Assignments
(f) Words' Components: 4 vs. 5
Component 4: 0.26%, λ = 0.15
Component 5: 0.25%, λ = 0.15
Assignments

One table

- Independent Components Analysis

One table

- Independent Components Analysis
  - Effectively a rotation

One table

- Independent Components Analysis
  - Effectively a rotation
- Factor analyses, mostly

One table

- Independent Components Analysis
  - Effectively a rotation
- Factor analyses, mostly
  - Different error terms + rotation

One table

- Independent Components Analysis
  - Effectively a rotation
- Factor analyses, mostly
  - Different error terms + rotation
- Non-negative matrix factorization

One table

- Independent Components Analysis
  - Effectively a rotation
- Factor analyses, mostly
  - Different error terms + rotation
- Non-negative matrix factorization
- Non-symmetric CA

One table

- Independent Components Analysis
  - Effectively a rotation
- Factor analyses, mostly
  - Different error terms + rotation
- Non-negative matrix factorization
- Non-symmetric CA
- Hellinger CA

One table

- Independent Components Analysis
  - Effectively a rotation
- Factor analyses, mostly
  - Different error terms + rotation
- Non-negative matrix factorization
- Non-symmetric CA
- Hellinger CA
- Compositional CA

One table

- Independent Components Analysis
  - Effectively a rotation
- Factor analyses, mostly
  - Different error terms + rotation
- Non-negative matrix factorization
- Non-symmetric CA
- Hellinger CA
- Compositional CA
- Multidimensional scaling (MDS)

Two tables: Part 1

- Partial least squares (correlation)

Two tables: Part 1

- Partial least squares (correlation)
- Partial least squares (regression)

Two tables: Part 1

- Partial least squares (correlation)
- Partial least squares (regression)
- Partial least squares (path modelling)

Two tables: Part 2

- Canonical Correlation Analysis

Two tables: Part 2

- Canonical Correlation Analysis
- Discriminant analyses

Two tables: Part 2

- Canonical Correlation Analysis
- Discriminant analyses
- Reduced rank regression/redundancy analysis

Two tables: Part 2

- Canonical Correlation Analysis
- Discriminant analyses
- Reduced rank regression/redundancy analysis
- Generalized PLS regression

Two tables: Part 2

- Canonical Correlation Analysis
- Discriminant analyses
- Reduced rank regression/redundancy analysis
- Generalized PLS regression
  - Beaton, Saporta, Abdi (2019)

Two tables: Part 2

- Canonical Correlation Analysis
- Discriminant analyses
- Reduced rank regression/redundancy analysis
- Generalized PLS regression
  - Beaton, Saporta, Abdi (2019)
  - Mixed data, most two table techniques

More than two tables

- STATIS

More than two tables

- STATIS
- Multiple factor analysis

Not as related

- tSNE

Not as related

- tSNE
- UMAP

Not as related

- tSNE
- UMAP
- More akin to non-metric multidimensional scaling

Not as related

- tSNE
- UMAP
- More akin to non-metric multidimensional scaling
  - Not always a fair comparison

For all types of data

- Distances (MDS, DiSTATIS, CovSTATIS)

For all types of data

- Distances (MDS, DiSTATIS, CovSTATIS)
- Networks (CA)

For all types of data

- Distances (MDS, DiSTATIS, CovSTATIS)
- Networks (CA)
  - More magic!

(Some) References

See the reference sections of these

- Beaton, D., Saporta, G., Abdi, H., & Alzheimer's Disease Neuroimaging Initiative. (2019). A generalization of partial least squares regression and correspondence analysis for categorical and mixed data: An application with the ADNI data. bioRxiv, 598888.

See the reference sections of these

- Beaton, D., Saporta, G., Abdi, H., & Alzheimer's Disease Neuroimaging Initiative. (2019). A generalization of partial least squares regression and correspondence analysis for categorical and mixed data: An application with the ADNI data. bioRxiv, 598888.
- Beaton, D., Sunderland, K. M., Levine, B., Mandzia, J., Masellis, M., Swartz, R. H., ... & Strother, S. C. (2019). Generalization of the minimum covariance determinant algorithm for categorical and mixed data types. bioRxiv, 333005.

And these

- Abdi, H., Guillemot, V., Eslami, A., & Beaton, D. (2017). Canonical correlation analysis. Encyclopedia of Social Network Analysis and Mining, 1-16.

And these

- Abdi, H., Guillemot, V., Eslami, A., & Beaton, D. (2017). Canonical correlation analysis. Encyclopedia of Social Network Analysis and Mining, 1-16.
- Beaton, D., Dunlop, J., & Abdi, H. (2016). Partial least squares correspondence analysis: A framework to simultaneously analyze behavioral and genetic data. Psychological methods, 21(4), 621.

Techniques

- Greenacre, M. (2017). Correspondence analysis in practice. CRC press.

Techniques

- Greenacre, M. (2017). Correspondence analysis in practice. CRC press.
- Greenacre, M. J. (1984). Theory and Applications of Correspondence Analysis. Retrieved from http://books.google.com/books?id=LsPaAAAAMAAJ

Techniques

- Greenacre, M. J. (2010). Correspondence analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(5), 613–619. https://doi.org/10.1002/wics.114

Techniques

- Greenacre, M. J. (2010). Correspondence analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(5), 613–619. https://doi.org/10.1002/wics.114
- Lebart, L., Morineau, A., & Warwick, K. M. (1984). Multivariate descriptive statistical analysis: correspondence analysis and related techniques for large matrices. Wiley.

Resampling

- Berry, K. J., Johnston, J. E., & Mielke, P. W. (2011). Permutation methods. Wiley Interdisciplinary Reviews: Computational Statistics, 3, 527–542. https://doi.org/10.1002/wics.177

Resampling

- Berry, K. J., Johnston, J. E., & Mielke, P. W. (2011). Permutation methods. Wiley Interdisciplinary Reviews: Computational Statistics, 3, 527–542. https://doi.org/10.1002/wics.177
- Strother, S. C., Anderson, J., Hansen, L. K., Kjems, U., Kustra, R., Sidtis, J., . . . Rottenberg, D. (2002). The Quantitative Evaluation of Functional Neuroimaging Experiments: The NPAIRS Data Analysis Framework. NeuroImage, 15(4), 747–771. https://doi.org/10.1006/nimg.2001.1034

Resampling

- Berry, K. J., Johnston, J. E., & Mielke, P. W. (2011). Permutation methods. Wiley Interdisciplinary Reviews: Computational Statistics, 3, 527–542. https://doi.org/10.1002/wics.177
- Strother, S. C., Anderson, J., Hansen, L. K., Kjems, U., Kustra, R., Sidtis, J., . . . Rottenberg, D. (2002). The Quantitative Evaluation of Functional Neuroimaging Experiments: The NPAIRS Data Analysis Framework. NeuroImage, 15(4), 747–771. https://doi.org/10.1006/nimg.2001.1034
- Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics, 7(1), 1–26.

Resampling

- Chernick, M. R. (2008). Bootstrap methods: A guide for practitioners and researchers (Vol. 619). Wiley-Interscience.

Resampling

- Chernick, M. R. (2008). Bootstrap methods: A guide for practitioners and researchers (Vol. 619). Wiley-Interscience.
- Hesterberg, T. (2011). Bootstrap. Wiley Interdisciplinary Reviews: Computational Statistics, 3, 497–526. https://doi.org/10.1002/wics.182

Resampling

- Chernick, M. R. (2008). Bootstrap methods: A guide for practitioners and researchers (Vol. 619). Wiley-Interscience.
- Hesterberg, T. (2011). Bootstrap. Wiley Interdisciplinary Reviews: Computational Statistics, 3, 497–526. https://doi.org/10.1002/wics.182
- McIntosh, A. R., & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: applications and advances. Neuroimage, 23, S250–S263.

Data

- Escofier, B. (1978). Analyse factorielle et distances répondant au principe d'équivalence distributionnelle. Revue de Statistique Appliquée, 26(4), 29–37.

Data

- Escofier, B. (1978). Analyse factorielle et distances répondant au principe d'équivalence distributionnelle. Revue de Statistique Appliquée, 26(4), 29–37.
- Escofier, B. (1979). Traitement simultané de variables qualitatives et quantitatives en analyse factorielle. Cahiers de l'Analyse Des Données, 4(2), 137–146.

Data

- Escofier, B. (1978). Analyse factorielle et distances répondant au principe d'équivalence distributionnelle. Revue de Statistique Appliquée, 26(4), 29–37.
- Escofier, B. (1979). Traitement simultané de variables qualitatives et quantitatives en analyse factorielle. Cahiers de l'Analyse Des Données, 4(2), 137–146.
- Greenacre, M. (2014). Data Doubling and Fuzzy Coding. In J. Blasius & M. Greenacre (Eds.), Visualization and Verbalization of Data (pp. 239–253). Philadelphia, PA, USA: CRC Press.