Principal Components & Multiple Correspondence Analyses

with resampling approaches for stability assessments

Derek Beaton

RRI RTC

May 01, 2019

Overview

- Something here

Overview

- Something here
- Where to find everything

Introduction

Slide A

Slide A

Slide B

Slide B

Taxonomy

The actual data

PCA

Stuff

PCA

How & When to use it

CA

Stuff

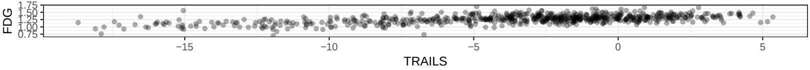
CA

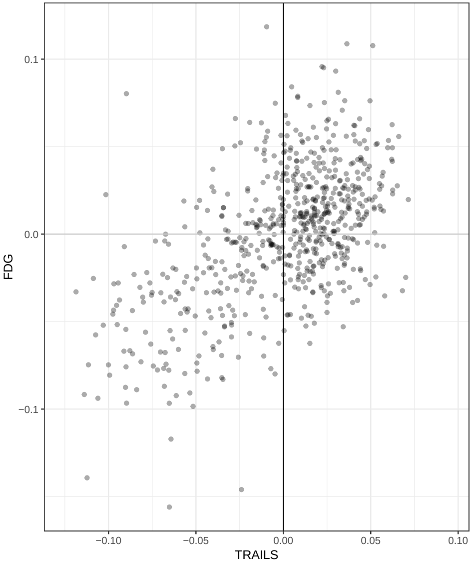
How & When to use it

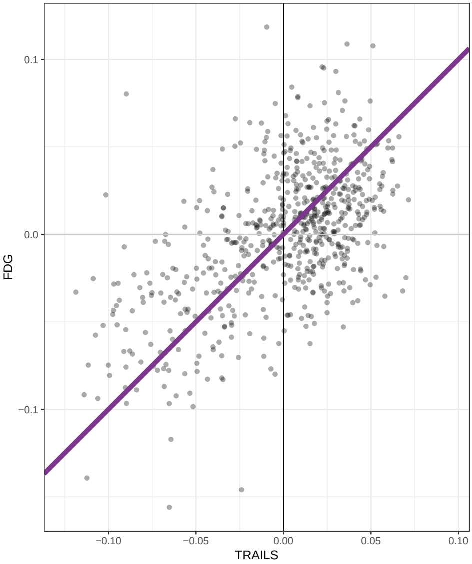
Today

Some alternatives

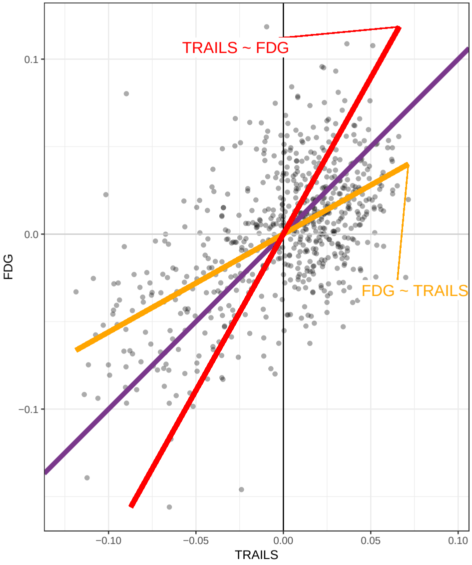
Principal Components Analysis


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

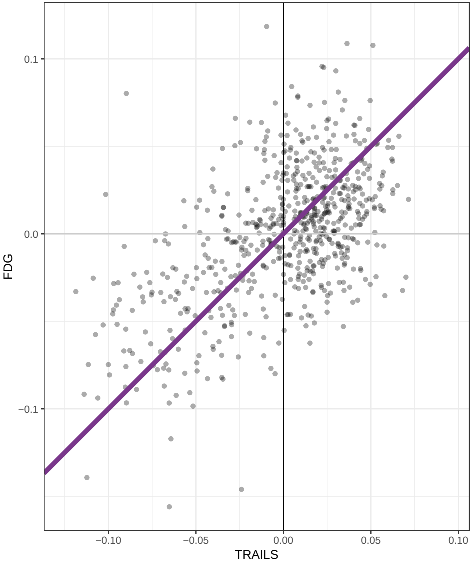

FDG
TRAILS
0.1
0.0
−0.1
−0.10
−0.05
0.00
0.05
0.10



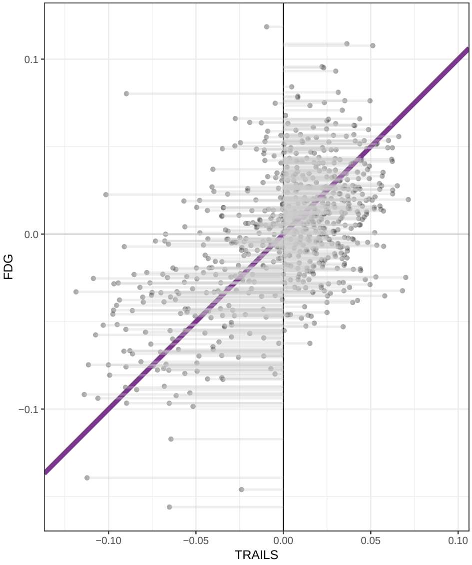
FDG
TRAILS
0.1
0.0
−0.1
−0.10
−0.05
0.00
0.05
0.10



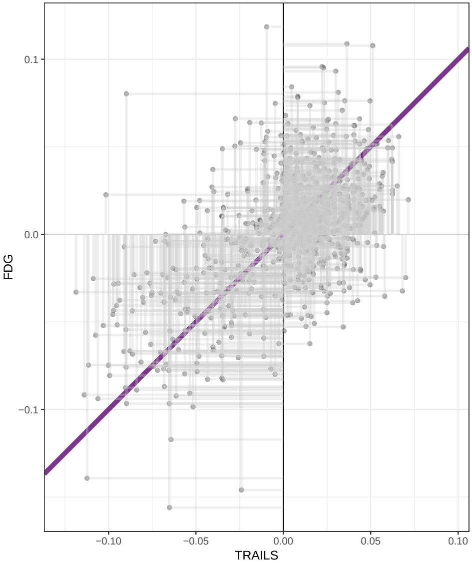
TRAILS ~ FDG
FDG ~ TRAILS
0.1
0.0
−0.1
FDG
−0.10
−0.05
0.00
0.05
0.10
TRAILS

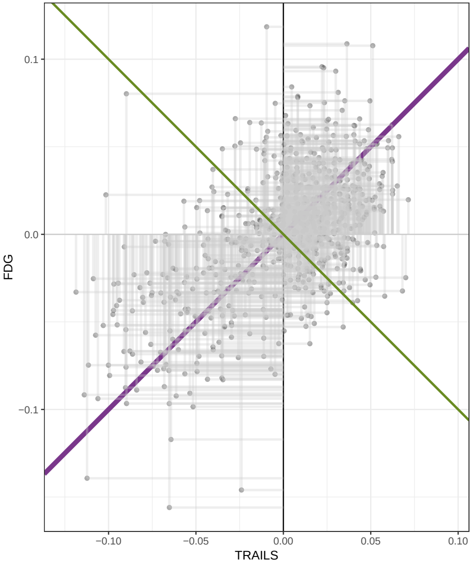


0.1
0.0
−0.1
FDG
−0.10
−0.05
0.00
0.05
0.10
TRAILS



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

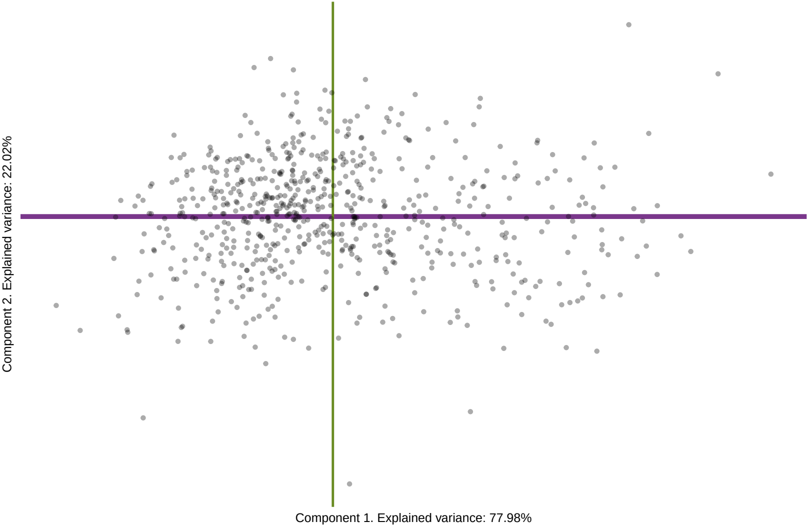

0.1
0.0
−0.1
FDG
−0.10
−0.05
0.00
0.05
0.10
TRAILS



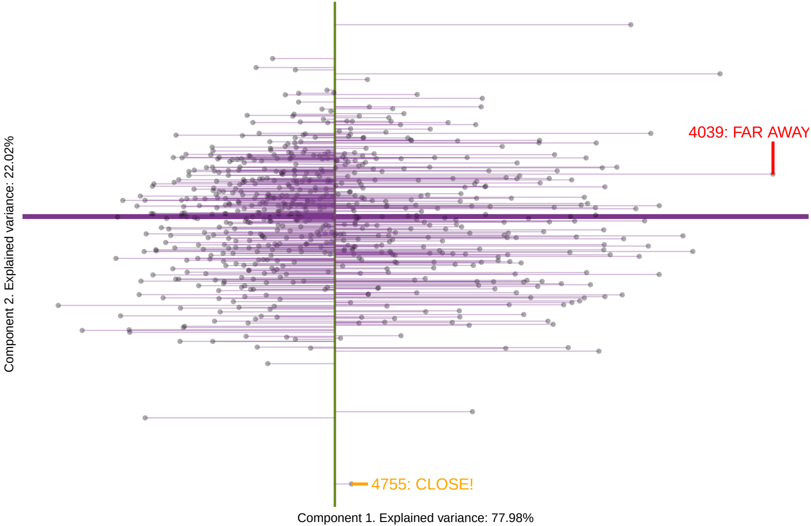
FDG
0.1
0.0
−0.1
−0.10
−0.05
0.00
0.05
0.10
TRAILS

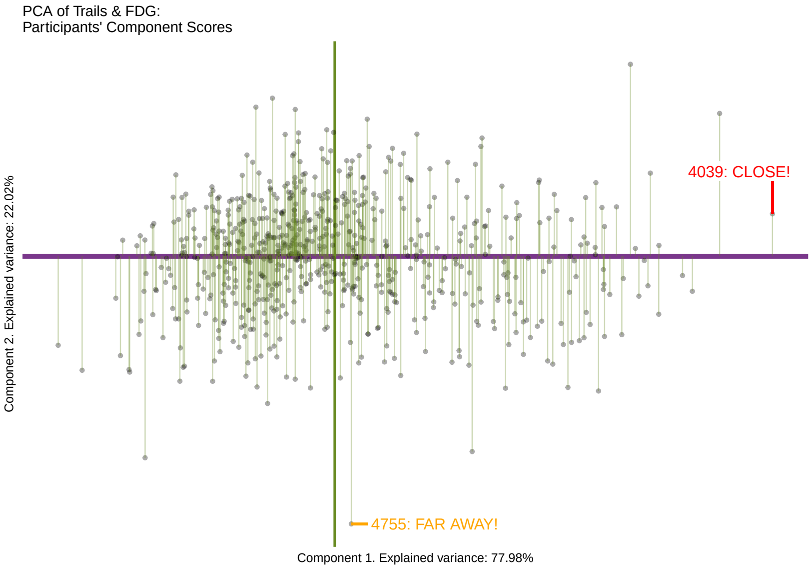
The SVD

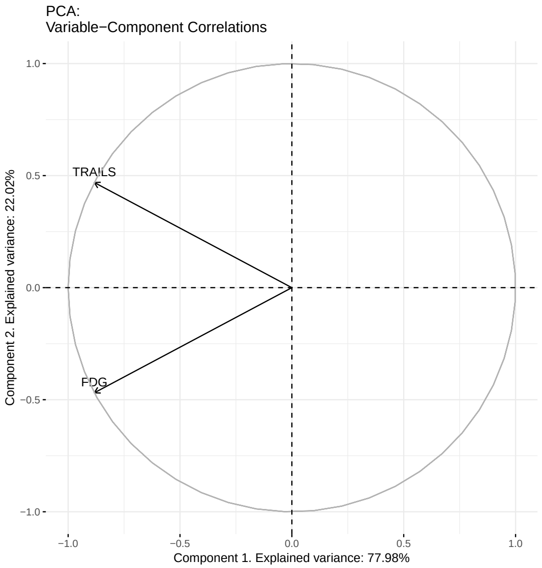
PCA of Trails & FDG:
Participants' Component Scores

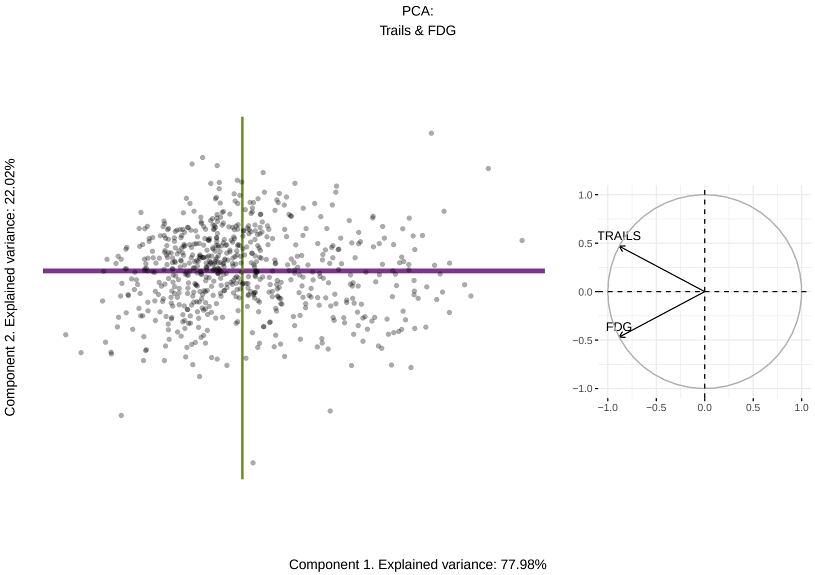


PCA of Trails & FDG:
Participants' Component Scores

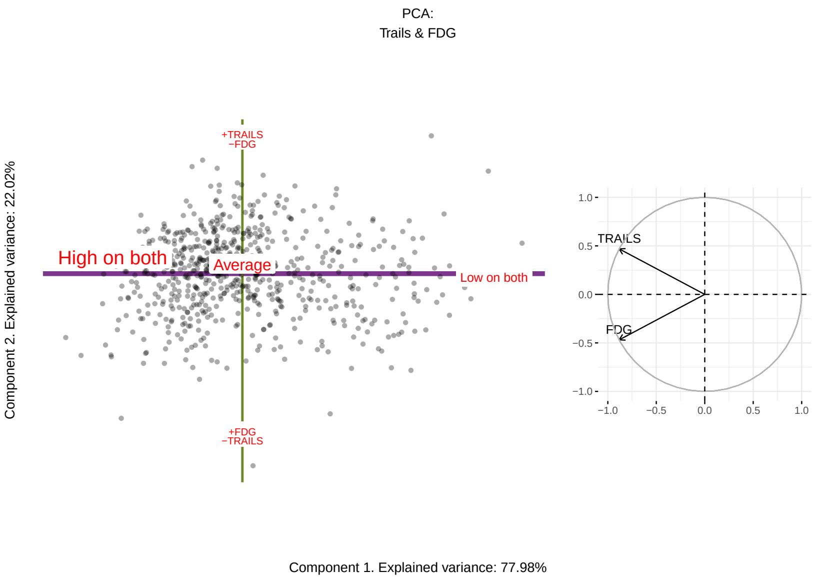



PCA of Trails & FDG:
Participants' Component Scores
4039: CLOSE!
4755: FAR AWAY!
Component 1. Explained variance: 77.98%
Component 2. Explained variance: 22.02%


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 2. Explained variance: 22.02%
Component 1. Explained variance: 77.98%

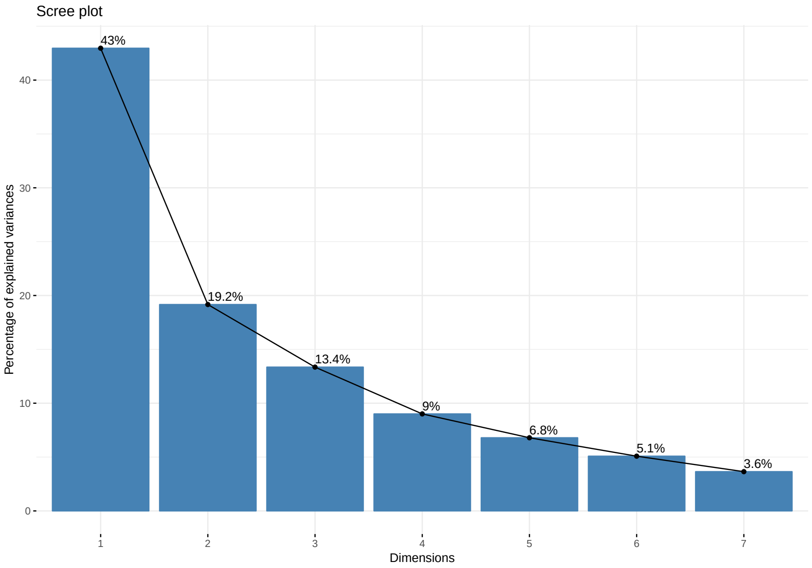


PCA:
Trails & FDG
Component 2. Explained variance: 22.02%
Component 1. Explained variance: 77.98%
TRAILS
FDG
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

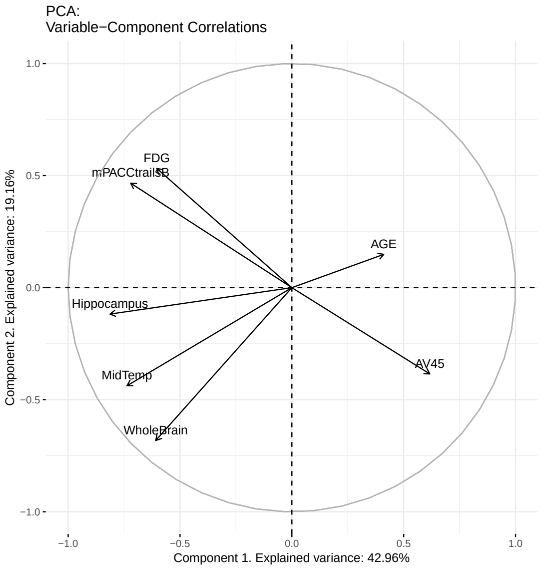

PCA:
Trails & FDG
+TRAILS
−FDG
High on both
Average
Low on both
+FDG
−TRAILS
Component 2. Explained variance: 22.02%
Component 1. Explained variance: 77.98%
TRAILS
FDG
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0

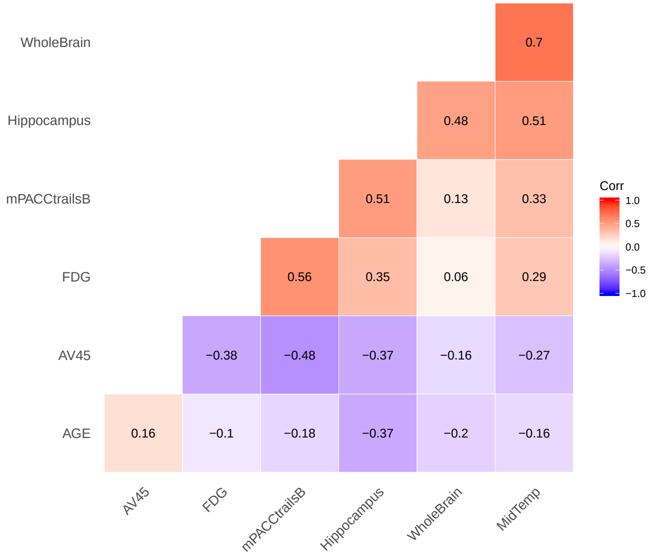
Scaling up

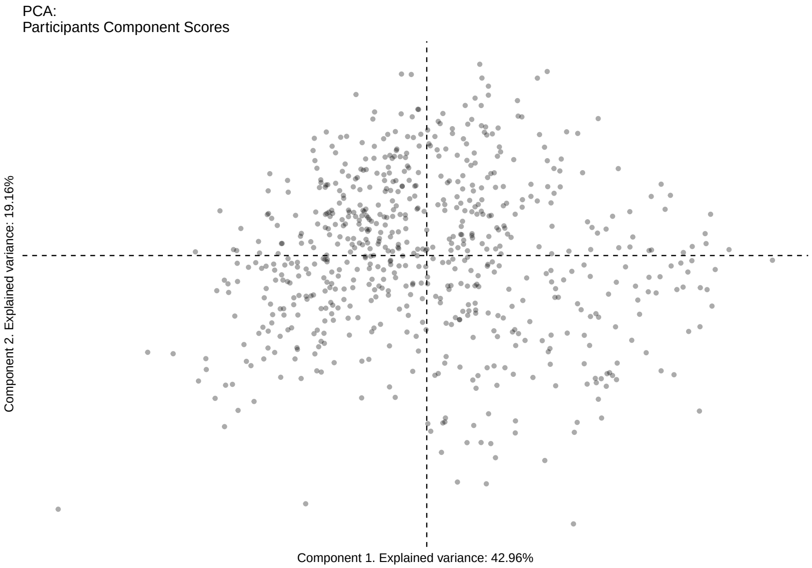
here: show the table and perhaps the code? maybe the code per slide for easy slides?



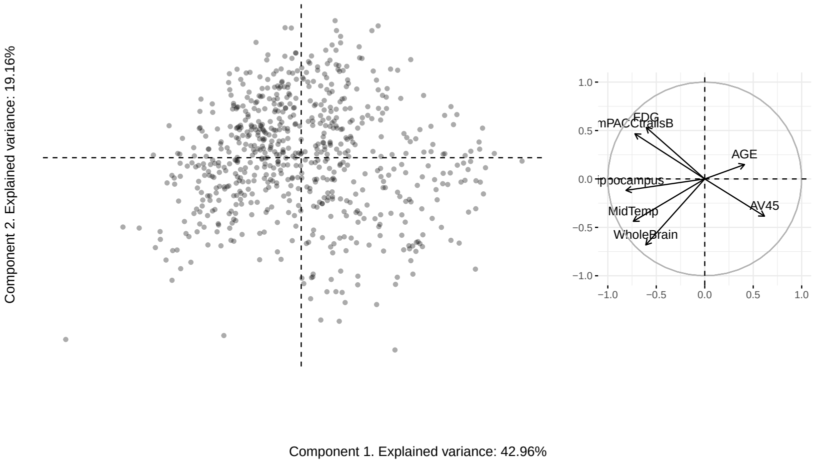
Scree plot
Percentage of explained variances
Dimensions
43%
19.2%
13.4%
9%
6.8%
5.1%
3.6%
0
10
20
30
40
1
2
3
4
5
6
7

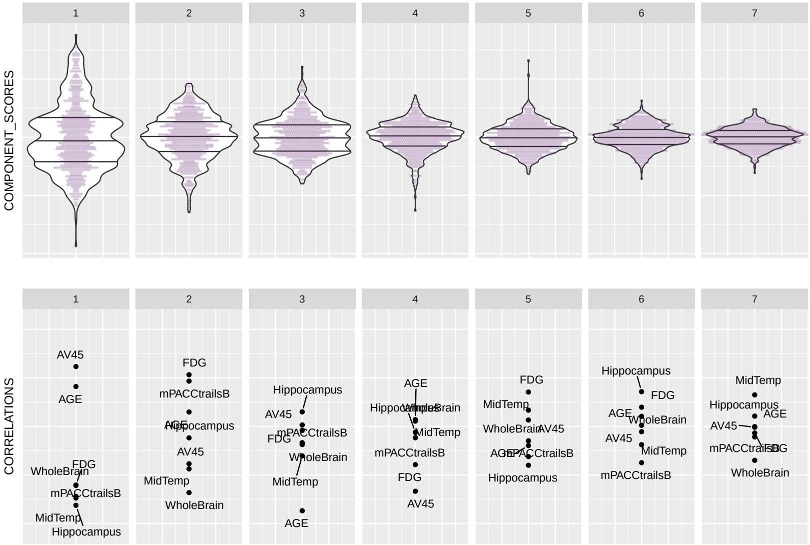

PCA:
Variable−Component Correlations
FDG
mPACCtrailsB
AGE
Hippocampus
AV45
MidTemp
WholeBrain
1.0
0.5
0.0
−0.5
−1.0
−1.0
−0.5
0.0
0.5
1.0
Component 2. Explained variance: 19.16%
Component 1. Explained variance: 42.96%


WholeBrain
Hippocampus
mPACCtrailsB
FDG
AV45
AGE
AV45
FDG
mPACCtrailsB
Hippocampus
WholeBrain
MidTemp
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
Corr
1.0
0.5
0.0
−0.5
−1.0


PCA:
Participants Component Scores
Component 1. Explained variance: 42.96%
Component 2. Explained variance: 19.16%

PCA:
Trails & FDG




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

Correspondence analyses

I don't know something.

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

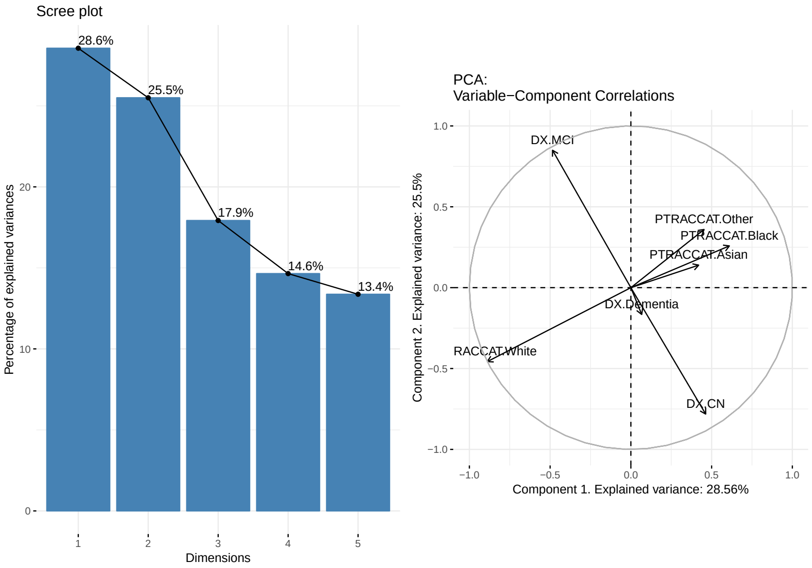
- Row sums are total number of *original* variables
- Sum within a variable (e.g. DX) is total number of rows

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


Scree plot
28.6%
25.5%
17.9%
14.6%
13.4%
Percentage of explained variances
0
10
20
1
2
3
4
5
Dimensions
PCA:
Variable−Component Correlations
DX.MCI
PTRACCAT.Other
PTRACCAT.Black
PTRACCAT.Asian
DX.Dementia
RACCAT.White
DX.CN
1.0
0.5
0.0
−0.5
−1.0
Component 2. Explained variance: 25.5%
−1.0
−0.5
0.0
0.5
1.0
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

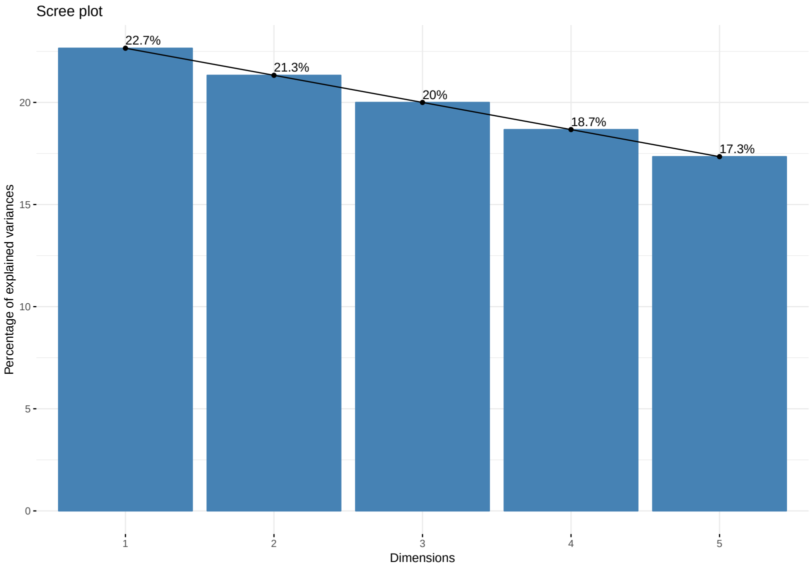
- Correspondence analysis (CA)
  - Think of it as a χ^2 PCA
- Deals with categories, counts (amongst others)

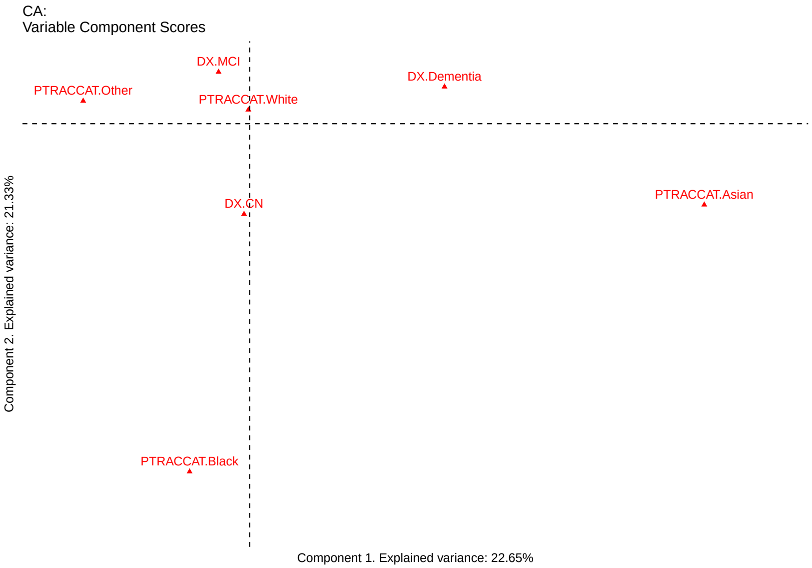
A better idea

- Correspondence analysis (CA)
  - Think of it as a χ^2 PCA
- Deals with categories, counts (amongst others)
- Row and column component scores exist on same scale

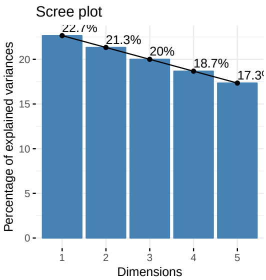
A better idea

- Correspondence analysis (CA)
  - Think of it as a χ^2 PCA
- Deals with categories, counts (amongst others)
- 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%
0
5
10
15
20
Percentage of explained variances
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 1. Explained variance: 22.65%
Component 2. Explained variance: 21.33%

	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

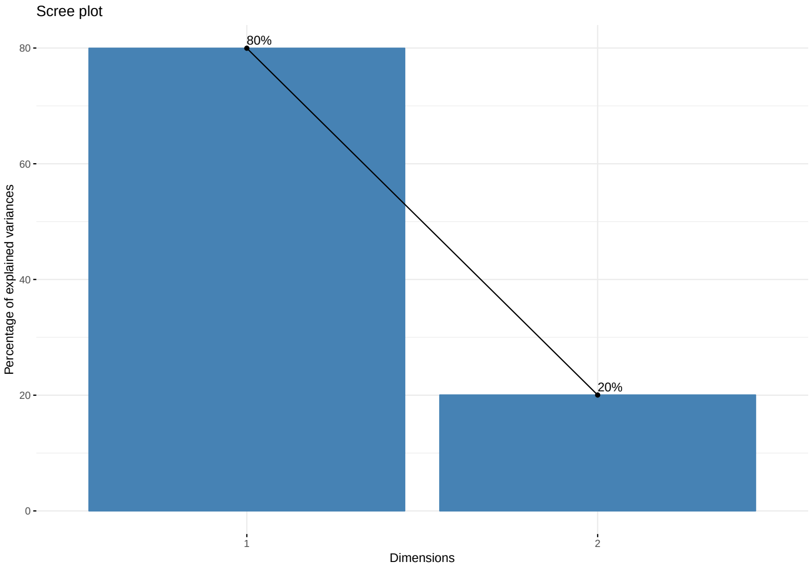
- An extension of CA
- Accomodates multiple categorical variables (CA only does 2)

Multiple correspondence analysis

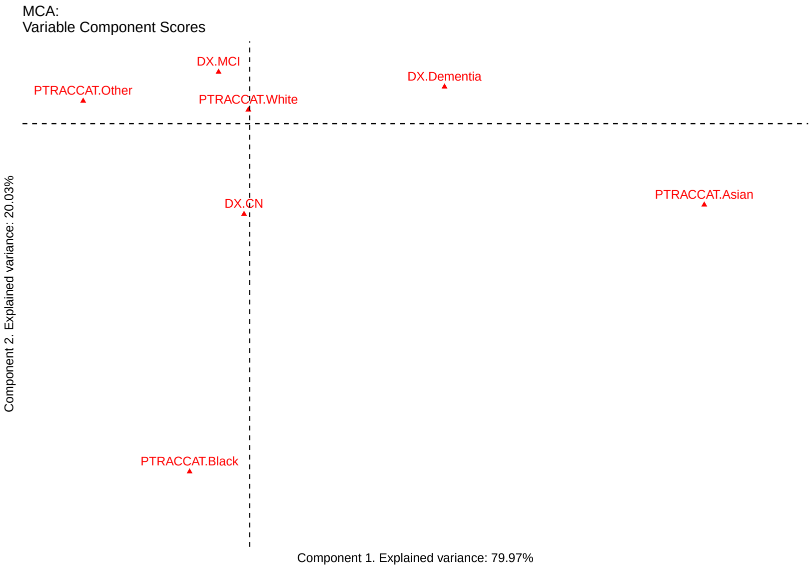
- An extension of CA
- Accomodates multiple categorical variables (CA only does 2)
- Corrects the dimensionality

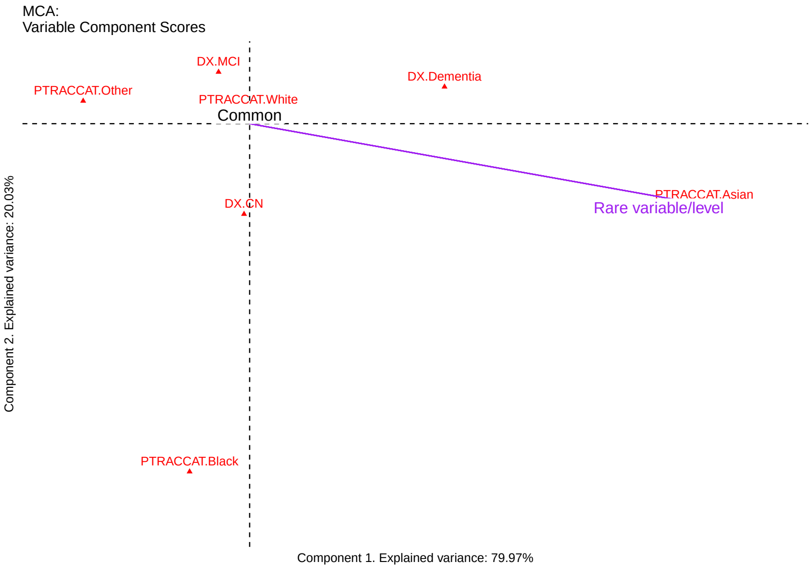
Multiple correspondence analysis

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

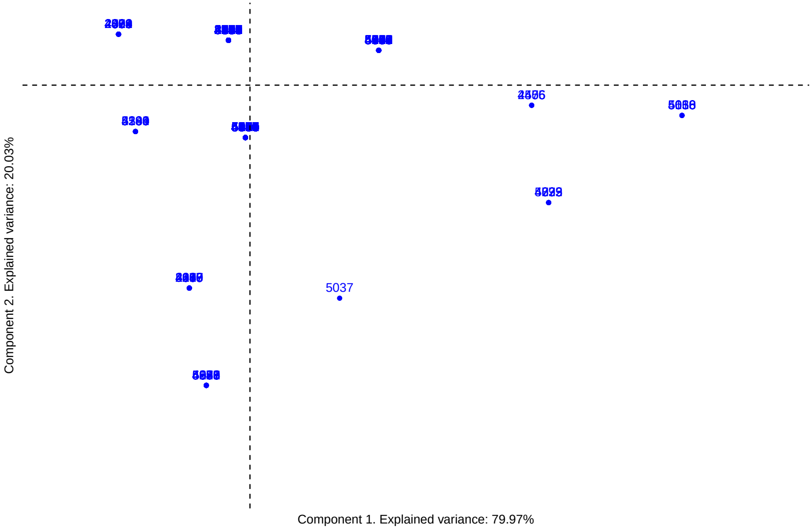


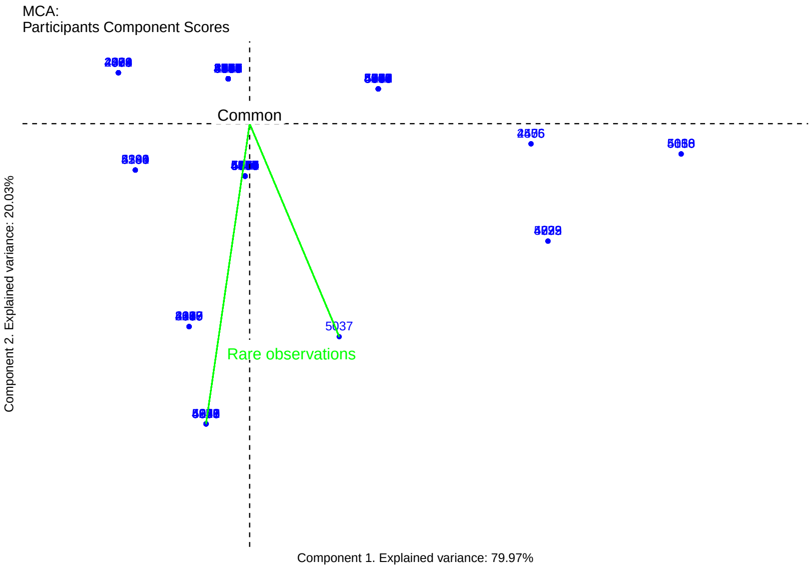
Scree plot
80%
20%
80
60
40
20
0
Percentage of explained variances
1
2
Dimensions


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


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

MCA:
Participants Component Scores




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

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

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

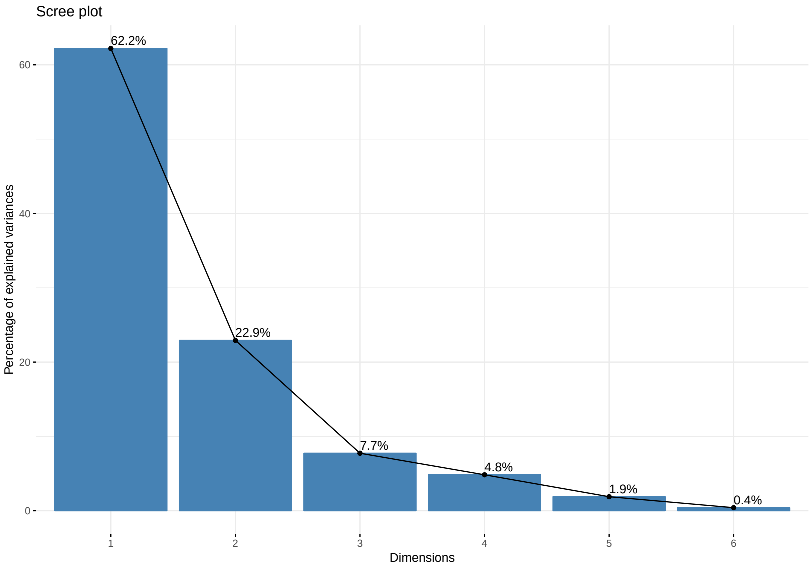
Compare the results

	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

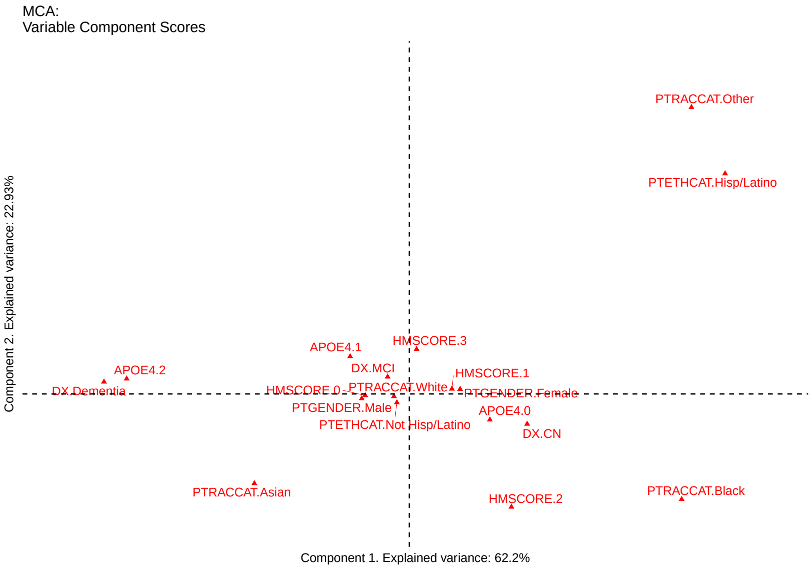
- CA & MCA produce identical results, except MCA:
  - Drops components
  - Corrects explained variance

Scaling up

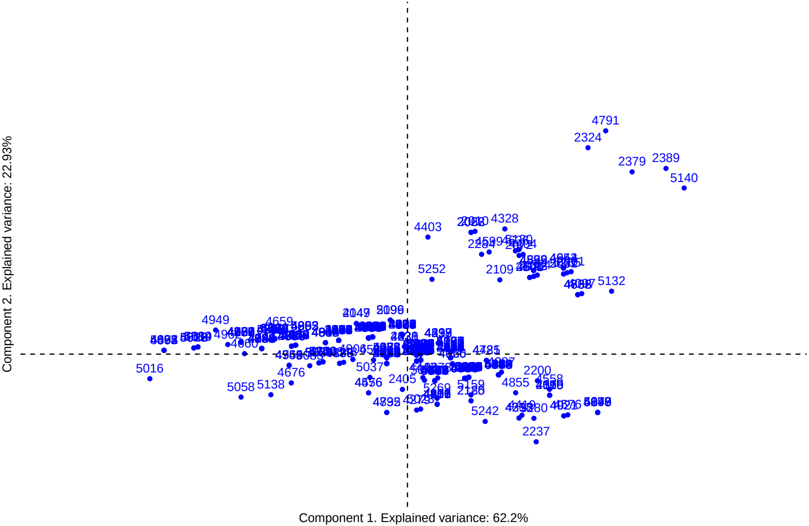
SHow the data here a bit

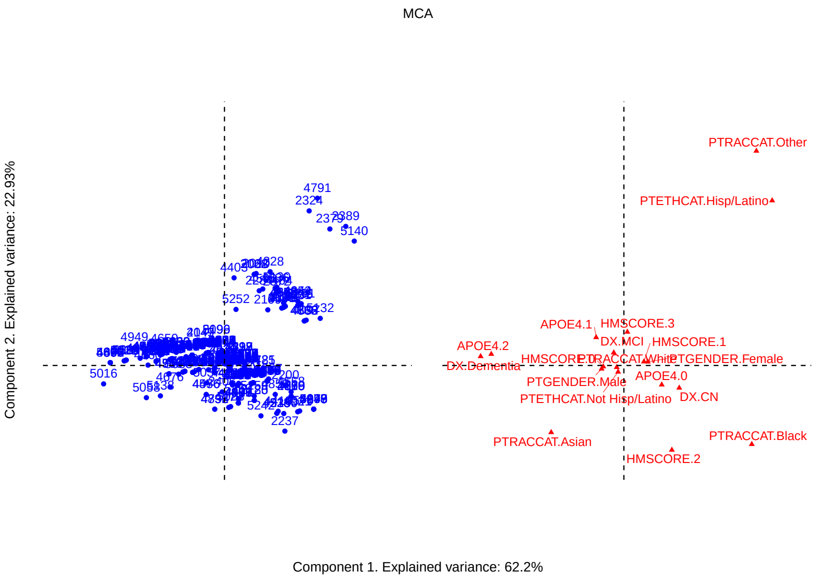


Scree plot
62.2%
22.9%
7.7%
4.8%
1.9%
0.4%
Percentage of explained variances
60
40
20
0
1
2
3
4
5
6
Dimensions


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

MCA:
Participants Component Scores




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

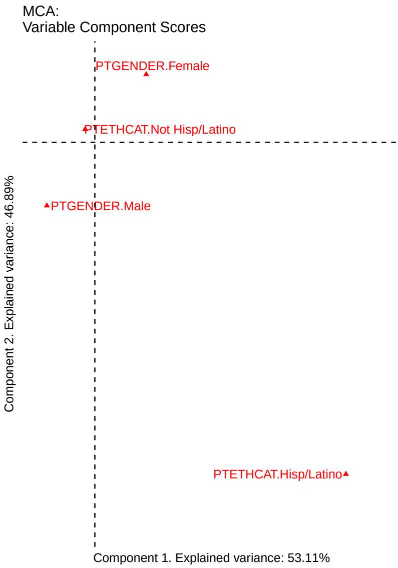
Alternate but equivalent binary coding of two variables with strictly two levels (i.e., binary data) in two columns

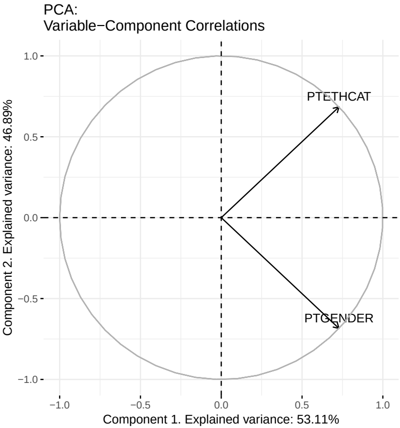
Always a bad idea?

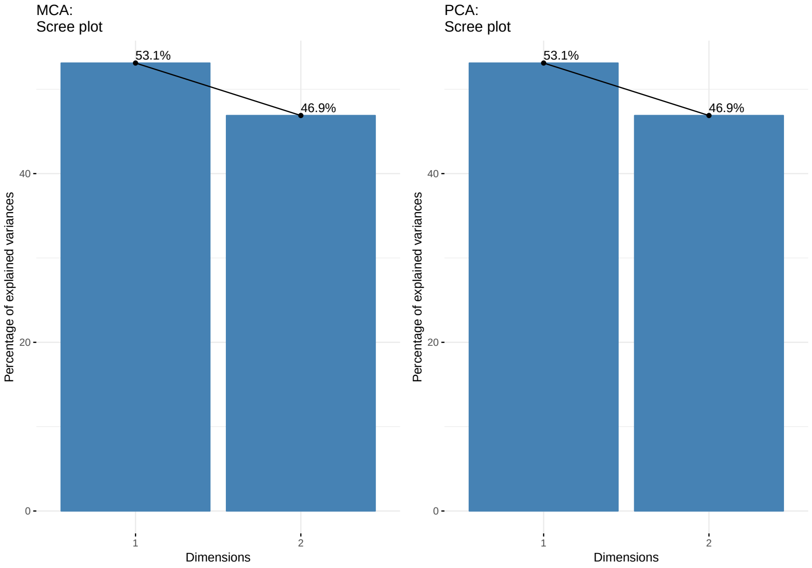
- MCA on the disjunctive coded data

Always a bad idea?

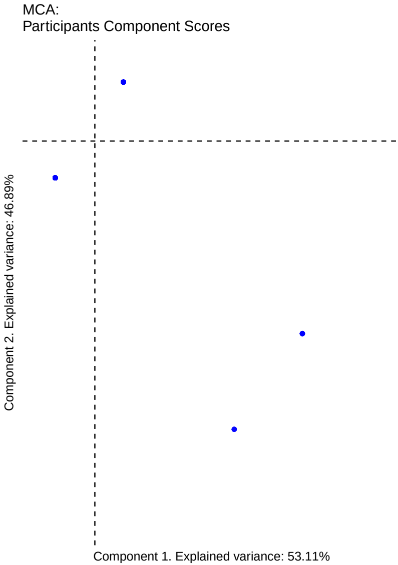
- MCA on the disjunctive coded data
- PCA on the binary coded data

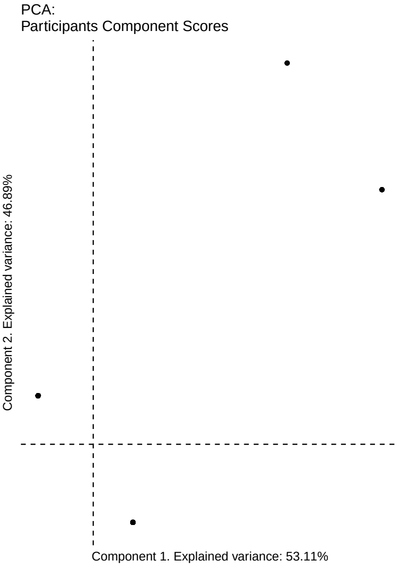

MCA:
Variable Component Scores
PTGENDER.Female
PTETHCAT.Not Hisp/Latino
PTGENDER.Male
PTETHCAT.Hisp/Latino
Component 2. Explained variance: 46.89%
Component 1. Explained variance: 53.11%


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%
Percentage of explained variances
Dimensions
PCA:
Scree plot
53.1%
46.9%
Percentage of explained variances
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 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

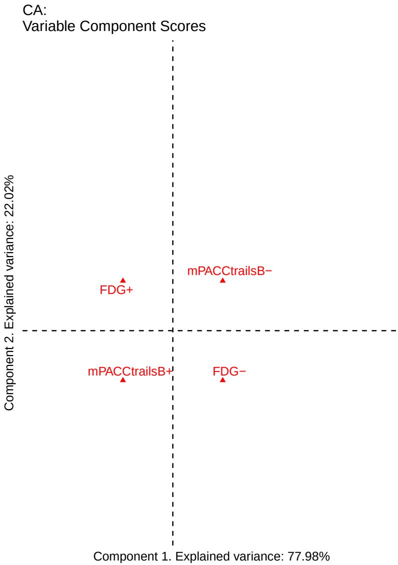
	mPACCtrailsB-	mPACCtrailsB+	FDG-	FDG+
5023	-0.06	1.06	0.43	0.57
5026	0.27	0.73	1.16	-0.16
5027	1.88	-0.88	1.24	-0.24
5028	1.30	-0.30	0.98	0.02
5031	0.96	0.04	0.93	0.07
5037	1.43	-0.43	1.50	-0.50
5040	0.03	0.97	0.60	0.40
5047	0.62	0.38	-1.03	2.03
5054	0.90	0.10	1.03	-0.03
5058	1.06	-0.06	1.57	-0.57
5063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

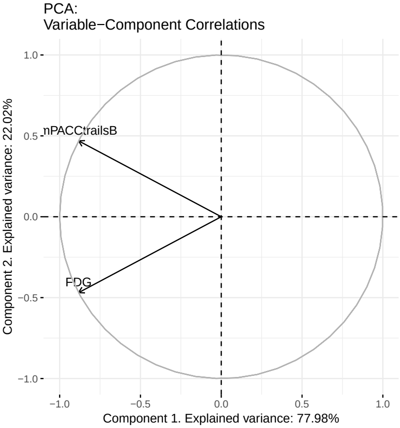
- Row sums are total number of *original* variables
- Sum within a variable (e.g., FDG) is total number of rows

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

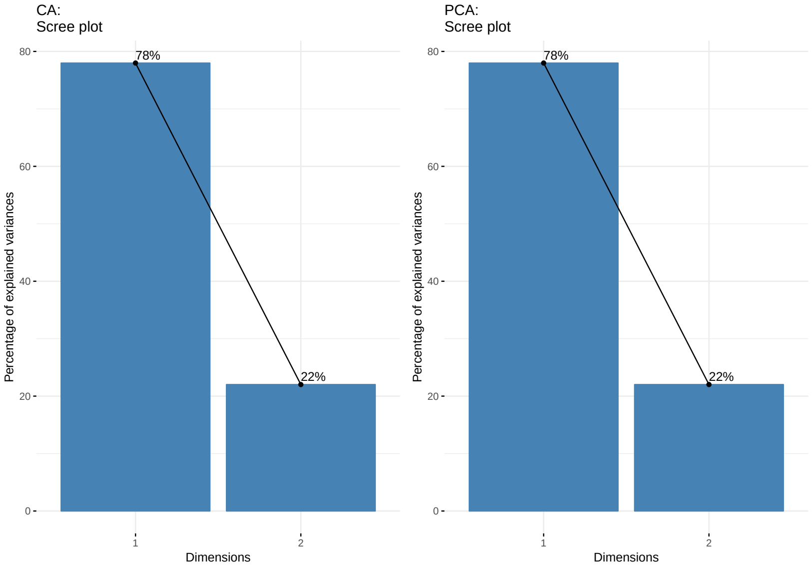
- Row sums are total number of *original* variables
- Sum within a variable (e.g., FDG) is total number of rows
- Sum of the table is rows $\times$ columns



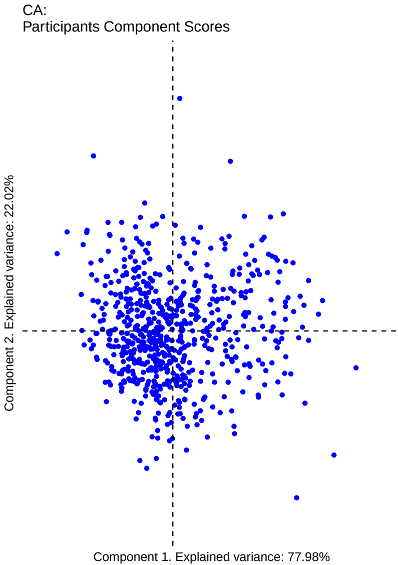


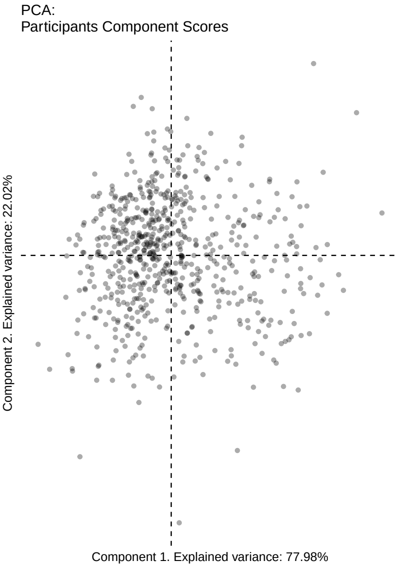
Oh, interesting!

Take note: each variable has two "poles"

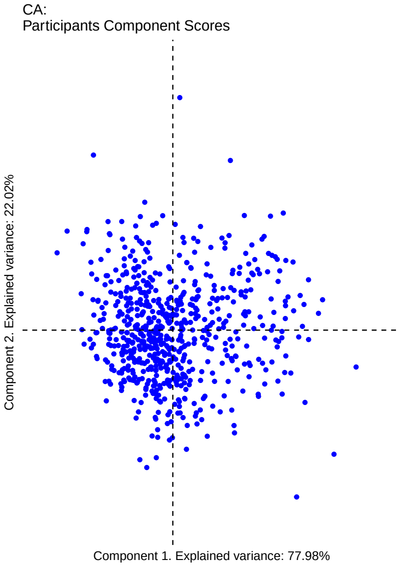

CA:
Scree plot
78%
22%
Percentage of explained variances
0
20
40
60
80
1
2
Dimensions
PCA:
Scree plot
78%
22%
Percentage of explained variances
0
20
40
60
80
1
2
Dimensions

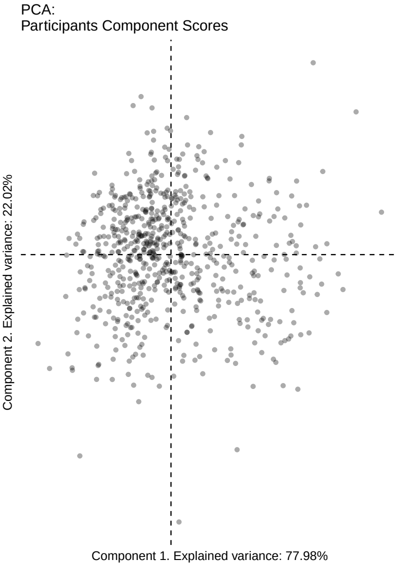
Oh, weird!





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)
  - $\left[\frac{\texttt{max}(x)-x}{\texttt{max}}\ \frac{x-\texttt{min}(x)}{\texttt{max}}\right]$

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

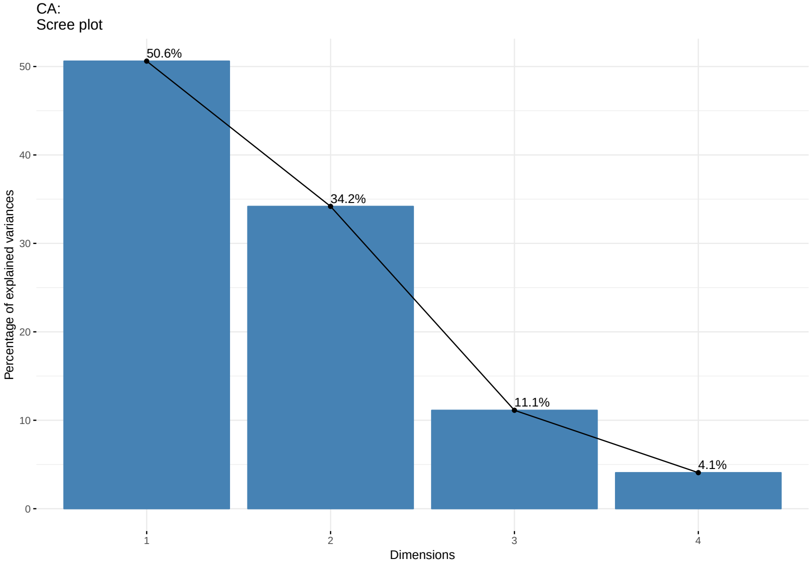
	PTEDUCAT+	PTEDUCAT-	CDRSB+	CDRSB-	ADAS13+	ADAS13-	MOCA+	MOCA-
5023	0.75	0.25	0.00	1.00	0.13	0.87	1.00	0.00
5026	0.75	0.25	0.27	0.73	0.17	0.83	0.57	0.43
5027	0.75	0.25	0.73	0.27	0.59	0.41	0.21	0.79
5028	0.50	0.50	0.64	0.36	0.43	0.57	0.21	0.79
5031	0.25	0.75	0.36	0.64	0.35	0.65	0.29	0.71
5037	0.50	0.50	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

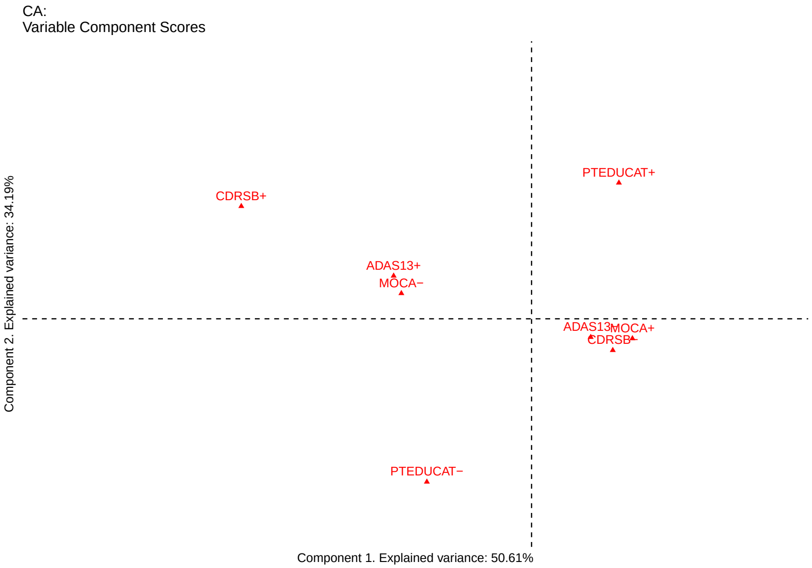
- Row sums are total number of *original* variables
- Sum within a variable (e.g. EDU) is total number of rows

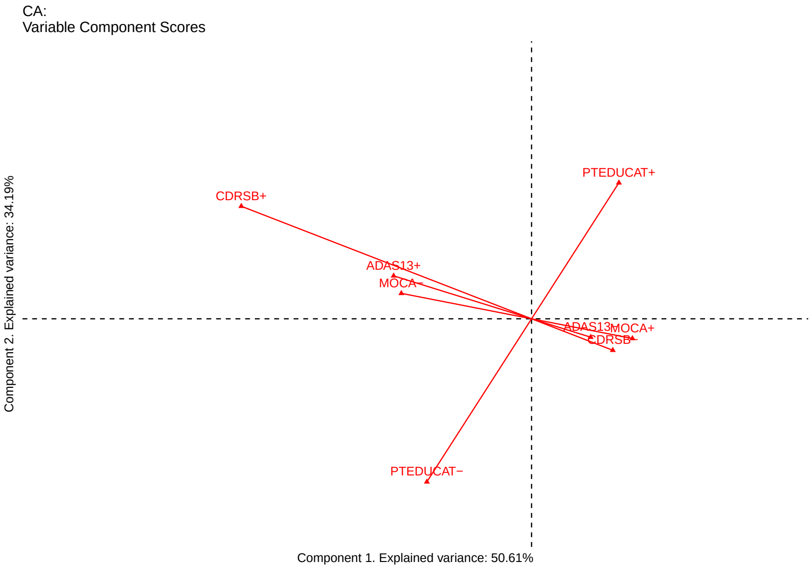
More Geometric Magic

	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

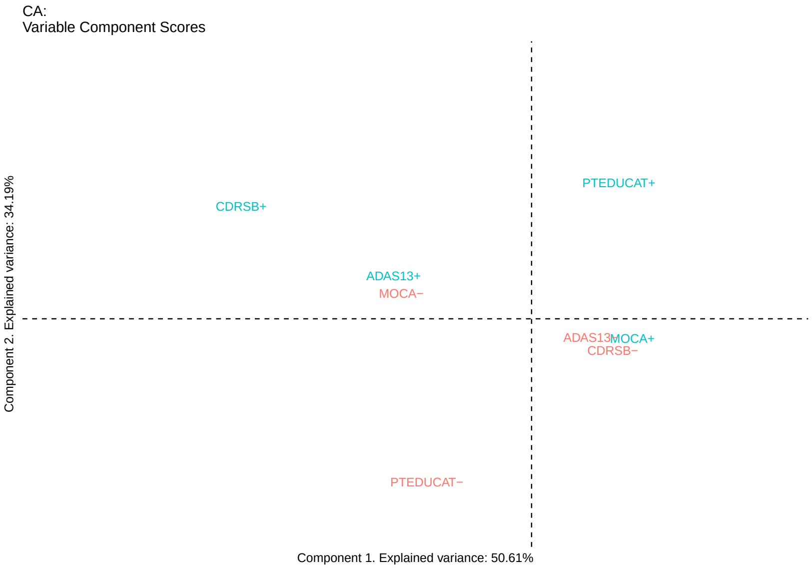
- Row sums are total number of *original* variables
- Sum within a variable (e.g. EDU) is total number of rows
- Sum of the table is rows $\times$ columns


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

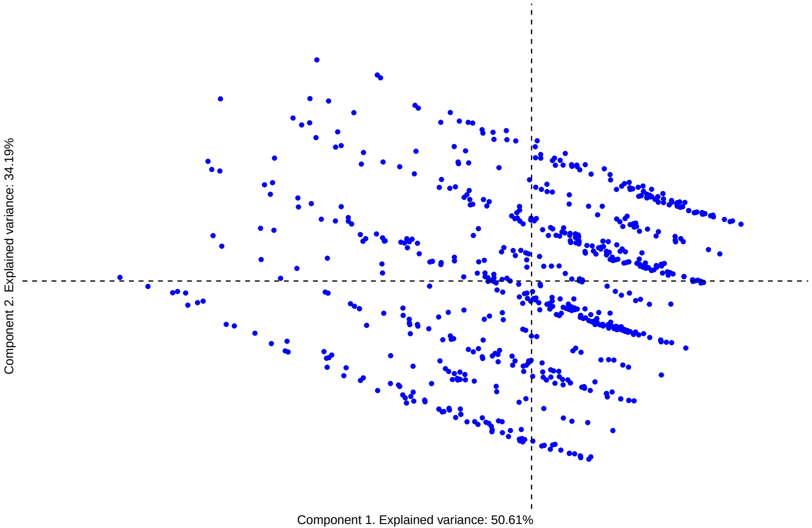

CA:
Variable Component Scores
PTEDUCAT+
CDRSB+
ADAS13+
MOCA−
ADAS13−
MOCA+
CDRSB−
PTEDUCAT−
Component 1. Explained variance: 50.61%
Component 2. Explained variance: 34.19%

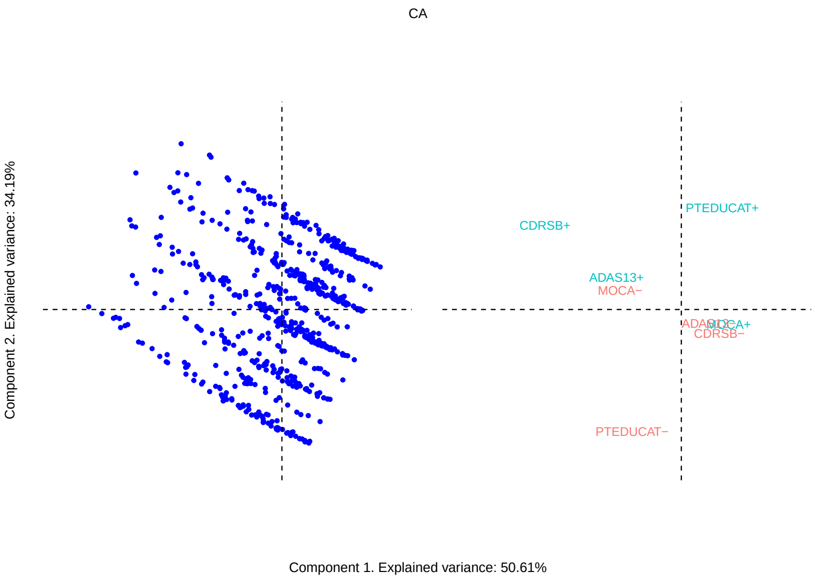


CA:
Variable Component Scores
PTEDUCAT+
CDRSB+
ADAS13+
MOCA−
ADAS13−
MOCA+
CDRSB−
PTEDUCAT−
Component 2. Explained variance: 34.19%
Component 1. Explained variance: 50.61%


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

CA:
Participants Component Scores




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

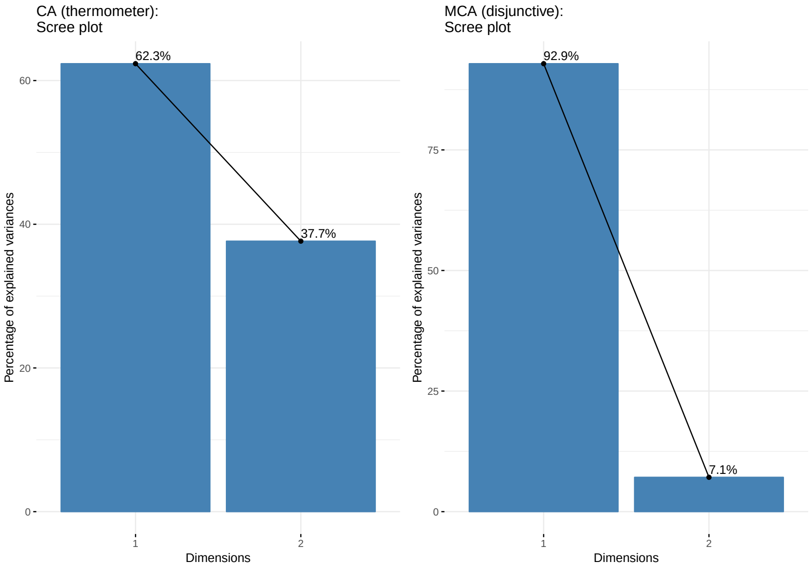
Thermometer vs. Disjunctive

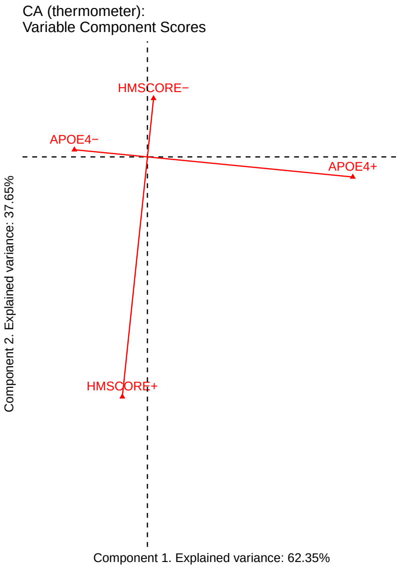
- Sometimes data could be either

Thermometer vs. Disjunctive

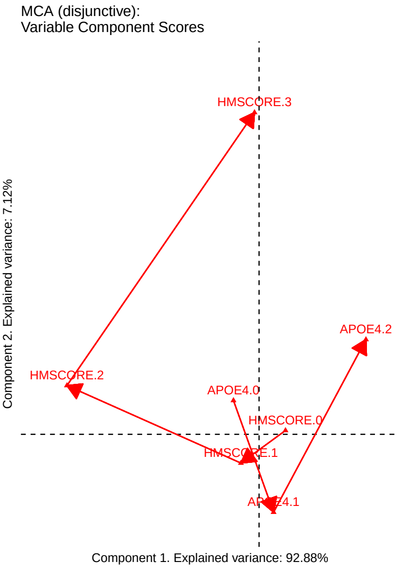
- Sometimes data could be either
- Let's analyze it both ways

	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

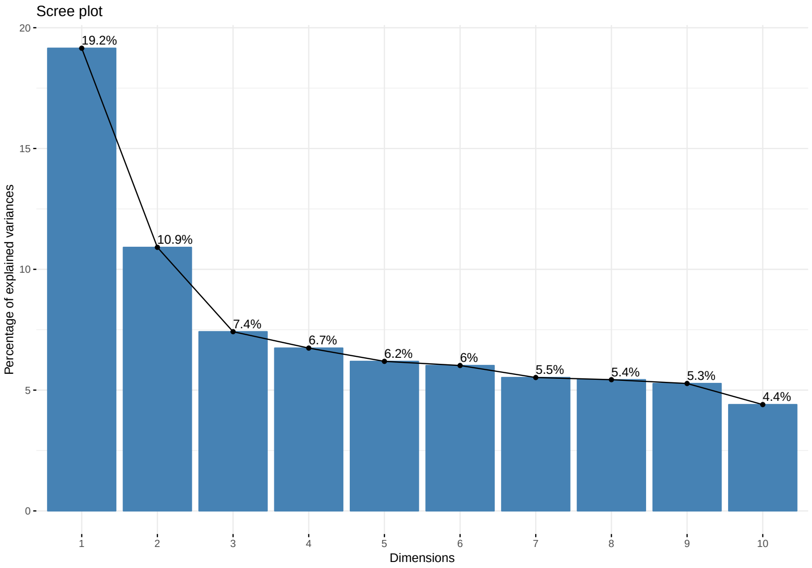
- For a small (reasonable) number of levels: disjunctive
- Otherwise: thermometer
- Interpretation:
  - Thermometer is "easier"

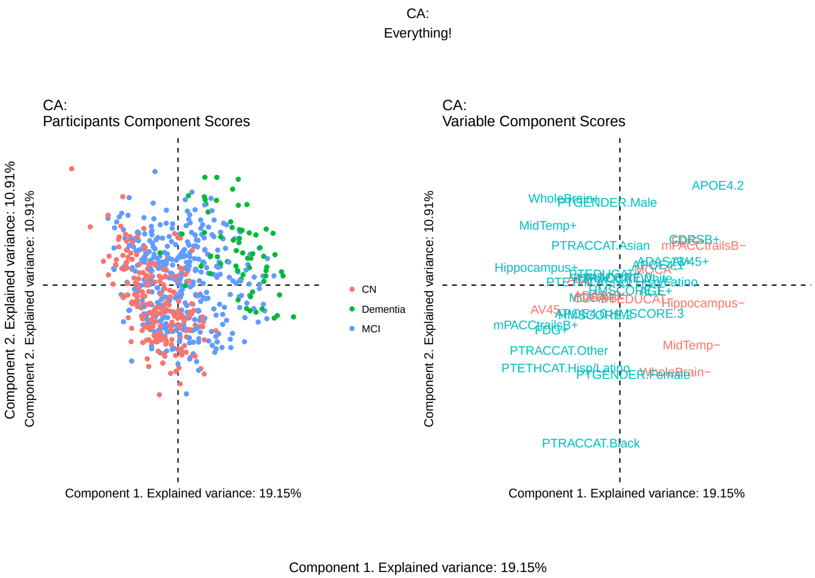
Thermometer vs. Disjunctive

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

All of the data

	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 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
PTGENDER.Male
MidTemp+
PTRACCAT.Asian
mPACCtrailsB−
Hippocampus+
Hippocampus−
AV45
FDG+
PTRACCAT.Other
MidTemp−
WholeBrain−
PTRACCAT.Black
Component 1. Explained variance: 19.15%

Resampling

FUCK