The Multivariate Flavors of JMP:

From Continuous to Categorical to Multiple-Source Data

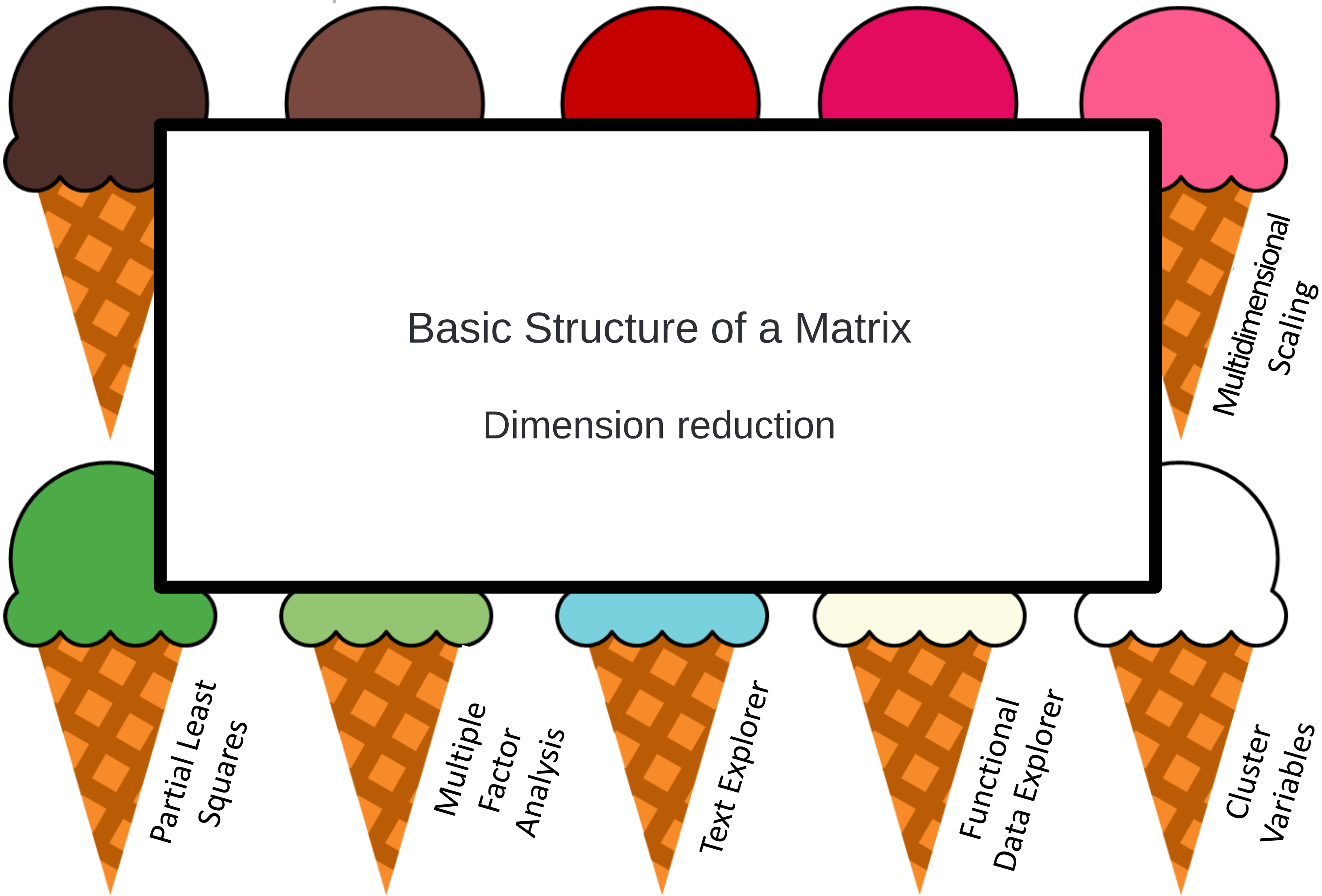
Laura Castro-Schilo, Ph.D.
Sr. Associate Research Statistician Developer
JMP Division, SAS Institute

MULTIVARIATE FLAVORS OF JMP

OVERVIEW

MULTIVARIATE FLAVORS OF JMP

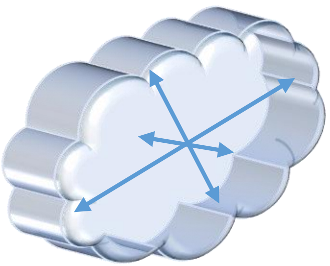
OVERVIEW



MULTIVARIATE FLAVORS OF JMP

OVERVIEW

BASIC STRUCTURE OF A MATRIX



- Decomposition of matrix into its characteristic components
  - Singular value decomposition (SVD)
  - Represent data as product of 3 matrices

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{T}}$$

BASIC STRUCTURE OF A MATRIX

Woody?

Fruity?

Spicy?

MULTIVARIATE FLAVORS OF JMP

SINGULAR VALUE DECOMPOSITION

$$\mathbf{X} = \mathbf{USV^T}$$

Woody	Fruity	Spicy
0.8	6.9	0.6
0.1	9.2	0.2
0.2	6.5	0.5
5	7.3	0.3
0.2	8.3	0.2
5.7	3.7	0.7
3.8	0.7	5
5.6	0.5	0.2
4.3	0.3	2.3
1.3	3.1	0.6
5.3	0	9.8
5.4	0	8.8
3.3	0.8	9.1

X

Data

=

Dim1	Dim2	Dim3
0.1989212	0.3202258	-0.126323
0.2208073	0.453361	-0.226535
0.1705592	0.3085778	-0.172721
0.3112392	0.3111744	0.3299106
0.2027492	0.4074437	-0.194208
0.2594041	0.1130174	0.4615508
0.2724333	-0.146537	0.0175289
0.1675449	-0.031407	0.5650599
0.1930972	-0.091533	0.2783023
0.1247567	0.1256511	0.018387
0.444887	-0.336456	-0.159485
0.4166369	-0.307906	-0.075486
0.3881544	-0.257781	-0.338794

U

Left Singular Vectors

Dimensions of row variables

Dim1	Dim2	Dim3
21.305092	0	0
0	17.023785	0
0	0	7.6370028

S

Singular Values

Importance of dimensions (ordered)

Woody	Fruity	Spicy
0.5700669	0.4908194	0.6588779
-0.153595	0.8514739	-0.501399
0.8071136	-0.184631	-0.560784

$\mathbf{V^T}$

Right Singular Vectors

Dimensions of column variables

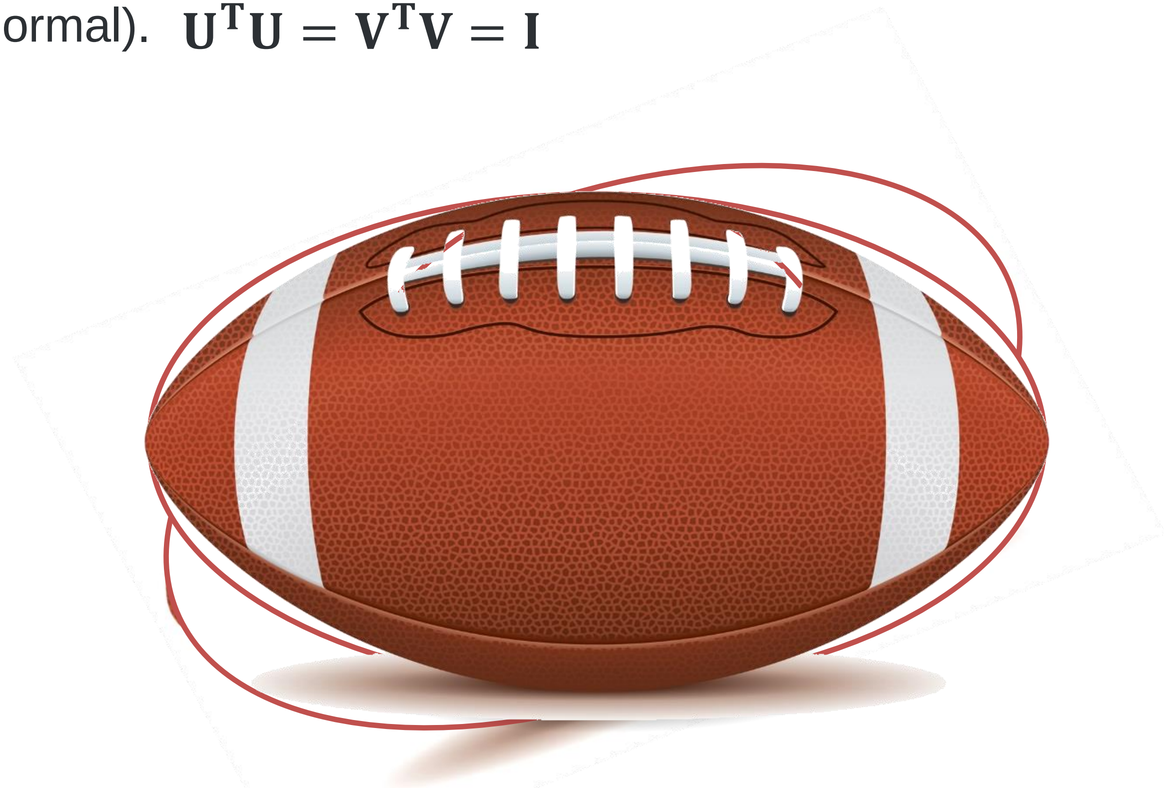
jmp

sas THE POWER TO KNOW.

MULTIVARIATE FLAVORS OF JMP

SINGULAR VALUE DECOMPOSITION

- Singular vectors (**U**, **V**) are orthogonal to each other and have unit length (orthonormal). $\mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{I}$



- Singular values, **S**, can be used to stretch out the vectors in **U** and **V** so they're no longer normalized but reflect the importance of each dimension.

SINGULAR VALUE DECOMPOSITION

- If **X** is symmetric, the singular vectors **U** and **V** will be identical.
  - Because pre- or post-multiplying a matrix by its transpose makes it symmetric, the basic structure matrices of **X**, $\mathbf{X}^\mathbf{T}\mathbf{X}$, and $\mathbf{X}\mathbf{X}^\mathbf{T}$, reveal the same basic structure.
  - Eigenvalue decomposition can also reveal the basic structure of **X**

$$\mathbf{X}^\mathbf{T}\mathbf{X} = \mathbf{V}\mathbf{S}\mathbf{U}^\mathbf{T}\mathbf{U}\mathbf{S}\mathbf{V}^\mathbf{T} = \mathbf{V}\mathbf{S}^2\mathbf{V}^\mathbf{T}$$
$$\mathbf{X}\mathbf{X}^\mathbf{T} = \mathbf{U}\mathbf{S}\mathbf{V}^\mathbf{T}\mathbf{V}\mathbf{S}\mathbf{U}^\mathbf{T} = \mathbf{U}\mathbf{S}^2\mathbf{U}^\mathbf{T}$$

SINGULAR VALUE DECOMPOSITION

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}}$$

"The basic structure of a matrix is like the layers of an onion; the components can be peeled off, one by one, and reassembled partially, or in whole"

Weller & Romney (1990)

DIMENSION REDUCTION

Woody	Fruity	Spicy
0.8	6.9	0.6
0.1	9.2	0.2
0.2	6.5	0.5
5	7.3	0.3
0.2	8.3	0.2
5.7	3.7	0.7
3.8	0.7	5
5.6	0.5	0.2
4.3	0.3	2.3
1.3	3.1	0.6
5.3	0	9.8
5.4	0	8.8
3.3	0.8	9.1

X

=

Dim1	Dim2	Dim3
0.1989212	0.3202258	-0.126323
0.2208073	0.453361	-0.226535
0.1705592	0.3085778	-0.172721
0.3112392	0.3111744	0.3299106
0.2027492	0.4074437	-0.194208
0.2594041	0.1130174	0.4615508
0.2724333	-0.146537	0.0175289
0.1675449	-0.031407	0.5650599
0.1930972	-0.091533	0.2783023
0.1247567	0.1256511	0.018387
0.444887	-0.336456	-0.159485
0.4166369	-0.307906	-0.075486
0.3881544	-0.257781	-0.338794

U

Dim1	Dim2	Dim3
21.305092	0	0
0	17.023785	0
0	0	7.6370028

S

Woody	Fruity	Spicy
0.5700669	0.4908194	0.6588779
-0.153595	0.8514739	-0.501399
0.8071136	-0.184631	-0.560784

V^T

*One-dimensional estimate of **X***

Woody	Fruity	Spicy
2.42	2.08	2.79
2.68	2.31	3.1
2.07	1.78	2.39
3.78	3.25	4.37
2.46	2.12	2.85
3.15	2.71	3.64
3.31	2.85	3.82
2.03	1.75	2.35
2.35	2.02	2.71
1.52	1.3	1.75
5.4	4.65	6.25
5.06	4.36	5.85
4.71	4.06	5.45

DIMENSION REDUCTION

Woody	Fruity	Spicy
0.8	6.9	0.6
0.1	9.2	0.2
0.2	6.5	0.5
5	7.3	0.3
0.2	8.3	0.2
5.7	3.7	0.7
3.8	0.7	5
5.6	0.5	0.2
4.3	0.3	2.3
1.3	3.1	0.6
5.3	0	9.8
5.4	0	8.8
3.3	0.8	9.1

X

=

Dim1	Dim2	Dim3
0.1989212	0.3202258	-0.126323
0.2208073	0.453361	-0.226535
0.1705592	0.3085778	-0.172721
0.3112392	0.3111744	0.3299106
0.2027492	0.4074437	-0.194208
0.2594041	0.1130174	0.4615508
0.2724333	-0.146537	0.0175289
0.1675449	-0.031407	0.5650599
0.1930972	-0.091533	0.2783023
0.1247567	0.1256511	0.018387
0.444887	-0.336456	-0.159485
0.4166369	-0.307906	-0.075486
0.3881544	-0.257781	-0.338794

U

Dim1	Dim2	Dim3
21.305092	0	0
0	17.023785	0
0	0	7.6370028

S

Woody	Fruity	Spicy
0.5700669	0.4908194	0.6588779
-0.153595	0.8514739	-0.501399
0.8071136	-0.184631	-0.560784

$\mathbf{V}^T$

*Two-dimensional estimate of **X***

Woody	Fruity	Spicy
1.58	6.72	0.06
1.5	8.88	-0.77
1.26	6.26	-0.24
2.97	7.77	1.71
1.4	8.03	-0.63
2.86	4.35	2.68
3.69	0.72	5.08
2.12	1.3	2.62
2.58	0.69	3.49
1.19	3.13	0.68
6.28	-0.22	9.12
5.87	-0.11	8.48
5.39	0.32	7.65

WHY THE LONG INTRO?

- All multivariate techniques in this session are based on:
  - Decompositions of transformed matrices:
    - Center, normalize, proportion, double-center, etc.*
  - Dimension reduction
- The techniques only differ in:
  - Pre-decomposition transformations of **X**
  - Post-decomposition transformations of **U** and **V**

* *Note:* transformations are sometimes implied (e.g., correlation matrices)

PRINCIPAL COMPONENTS ANALYSIS

PRINCIPAL COMPONENTS ANALYSIS

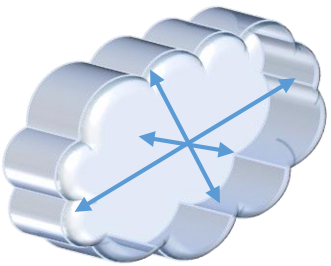
- Used with continuous data
- Goals of Analysis:
  - Identify underlying structure of data
  - Study inter-association of variables
  - Reduce dimensionality of data
    - Simplify ensuing analyses
  - Study inter-individual variability
    - Extract dimensions that distinguish individuals
    - Identify multivariate outliers
  - Measure latent variables (*but Factor Analysis can be better for this*)

PRINCIPAL COMPONENTS ANALYSIS

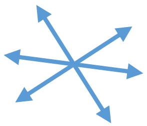
- Key output:
  - Eigenvalues (aka squared singular values)
  - Eigenvectors
  - Loadings
  - Percent of variance explained by each dimension
  - Principal component scores

- Most often known as the result of eigenvalue decomposition on a correlation (or covariance) matrix

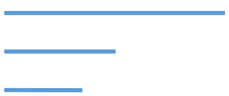
PRINCIPAL COMPONENTS ANALYSIS



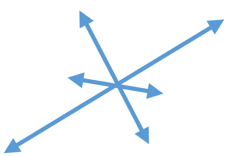
3-D Cloud for illustration



Unit length eigenvectors indicating main directions in data



Magnitude of each dimension from most to least important



Rescaled eigenvectors

PRINCIPAL COMPONENTS ANALYSIS

Eigenvalue decomposition of the covariance matrix of X:

$$\mathbf{S_{xx}}$$

BUT…

$$\mathbf{X_c} = \mathbf{X} - \mathbf{1}\hat{\boldsymbol{\mu}}^{\mathbf{T}}$$

Weights based on rows: *N*-1

$$\mathbf{S_{xx}} = \mathbf{X_c^T} \mathbf{D_r}^{-\frac{1}{2}\mathbf{T}} \mathbf{D_r}^{-\frac{1}{2}} \mathbf{X_c}$$

$$\mathbf{S_{xx}} = \frac{\mathbf{1}}{N-\mathbf{1}} (\mathbf{X_c^T X_c})$$

- Center X
- Sum of squares of centered X
- Divide all entries by *N*-1

PRINCIPAL COMPONENTS ANALYSIS

Alternatively, SVD of:

Weights: row (*N*-1)

$$\mathbf{D_r}^{-\frac{1}{2}}\left(\mathbf{X} - \mathbf{1}\hat{\boldsymbol{\mu}}^{\mathrm{T}}\right)$$

Raw data

Column means to center ***X***

- Center X
- Multiply each row by $\frac{1}{\sqrt{N-1}}$
- Multiply each column by the inverse of its corresponding standard deviation

PRINCIPAL COMPONENTS ANALYSIS

Eigenvalue decomposition of the correlation matrix of X:

$$\mathbf{R_{xx}}$$

BUT…

$$\mathbf{X_c} = \mathbf{X} - \mathbf{1}\hat{\boldsymbol{\mu}}^{\mathbf{T}}$$

Weights based on rows: *N*-1

$$\mathbf{R_{xx}} = \mathbf{D_c}^{-\frac{1}{2}} \left(\mathbf{X_c^T} \mathbf{D_r}^{-\frac{1}{2}\mathbf{T}} \mathbf{D_r}^{-\frac{1}{2}} \mathbf{X_c}\right) \mathbf{D_c}^{-\frac{1}{2}}$$

$$\mathbf{R_{xx}} = \frac{1}{N-1} \mathbf{D_c}^{-\frac{1}{2}} \left(\mathbf{X_c^T} \mathbf{X_c}\right) \mathbf{D_c}^{-\frac{1}{2}}$$

Weights based on columns: σ^2

- Center X
- Sum of squares of centered X
- Multiply each row and column by the inverse of the corresponding standard deviation
- Divide all entries by *N*-1

PRINCIPAL COMPONENTS ANALYSIS

SVD of:

Weights: row (N-1) and column (σ^2)

$$\mathbf{D_r}^{-\frac{1}{2}}\left(\mathbf{X} - \mathbf{1}\hat{\boldsymbol{\mu}}^{\mathbf{T}}\right)\mathbf{D_c}^{-\frac{1}{2}}$$

Raw data

Column means to center ***X***

- Center X
- Multiply each row by $\frac{1}{\sqrt{N-1}}$
- Multiply each column by the inverse of its corresponding standard deviation

PRINCIPAL COMPONENTS ANALYSIS

SVD of:

$$\mathbf{D_r}^{-\frac{1}{2}}\left(\mathbf{X} - \mathbf{1}\widehat{\boldsymbol{\mu}}^{\mathbf{T}}\right)\mathbf{D_c}^{-\frac{1}{2}} = \mathbf{USV^T}$$

Eigenvalues: $\mathbf{S}^2$
Eigenvectors: $\mathbf{V}$
Loadings: $\mathbf{VS}$
Scores: $\mathbf{US}$

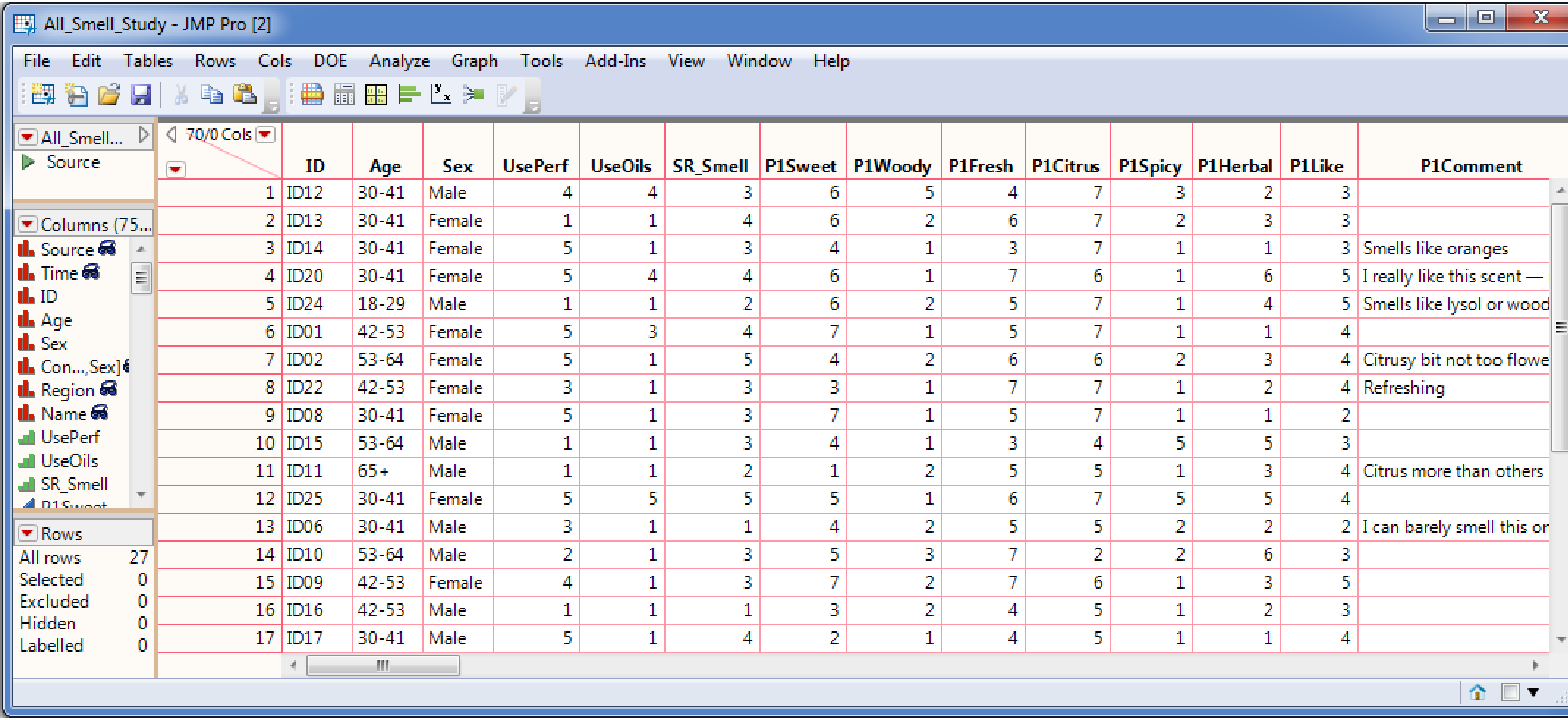
MOTIVATING EXAMPLE

Smell Study at SAS Headquarters

Rated 8 Scents:

1) Sweet Orange, Lavender
2) Peppermint, Lemon, Lavender
3) Tea Tree
4) Eucalyptus, Rosemary
5) Tea Tree, Eucalyptus, Lemon
6) Peppermint, Sweet Orange
7) Rosemary, Frankincense
8) ALL

PRINCIPAL COMPONENTS ANALYSIS

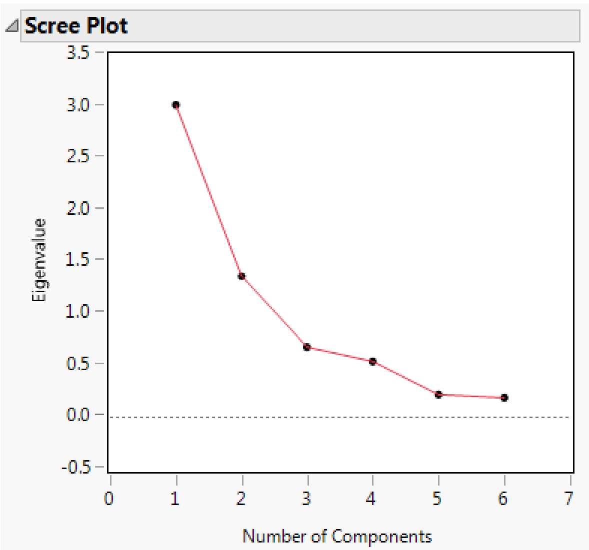


	ID	Age	Sex	UsePerf	UseOils	SR_Smell	P1Sweet	P1Woody	P1Fresh	P1Citrus	P1Spicy	P1Herbal	P1Like	P1Comment
1	ID12	30-41	Male	4	4	3	6	5	4	7	3	2	3	
2	ID13	30-41	Female	1	1	4	6	2	6	7	2	3	3	
3	ID14	30-41	Female	5	1	3	4	1	3	7	1	1	3	Smells like oranges
4	ID20	30-41	Female	5	4	4	6	1	7	6	1	6	5	I really like this scent —
5	ID24	18-29	Male	1	1	2	6	2	5	7	1	4	5	Smells like lysol or wood
6	ID01	42-53	Female	5	3	4	7	1	5	7	1	1	4	
7	ID02	53-64	Female	5	1	5	4	2	6	6	2	3	4	Citrusy bit not too flowe
8	ID22	42-53	Female	3	1	3	3	1	7	7	1	2	4	Refreshing
9	ID08	30-41	Female	5	1	3	7	1	5	7	1	1	2	
10	ID15	53-64	Male	1	1	3	4	1	3	4	5	5	3	
11	ID11	65+	Male	1	1	2	1	2	5	5	1	3	4	Citrus more than others
12	ID25	30-41	Female	5	5	5	5	1	6	7	5	5	4	
13	ID06	30-41	Male	3	1	1	4	2	5	5	2	2	2	I can barely smell this or
14	ID10	53-64	Male	2	1	3	5	3	7	2	2	6	3	
15	ID09	42-53	Female	4	1	3	7	2	7	6	1	3	5	
16	ID16	42-53	Male	1	1	1	3	2	4	5	1	2	3	
17	ID17	30-41	Male	5	1	4	2	1	4	5	1	1	4	

Data

MULTIVARIATE FLAVORS OF JMP

PRINCIPAL COMPONENTS ANALYSIS



Eigenvalues

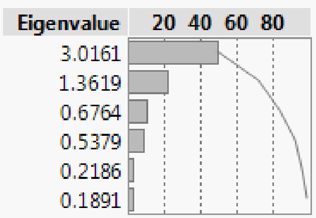
Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	3.0161	50.268		50.268
2	1.3619	22.698		72.966
3	0.6764	11.273		84.239
4	0.5379	8.965		93.205
5	0.2186	3.644		96.849
6	0.1891	3.151		100.000

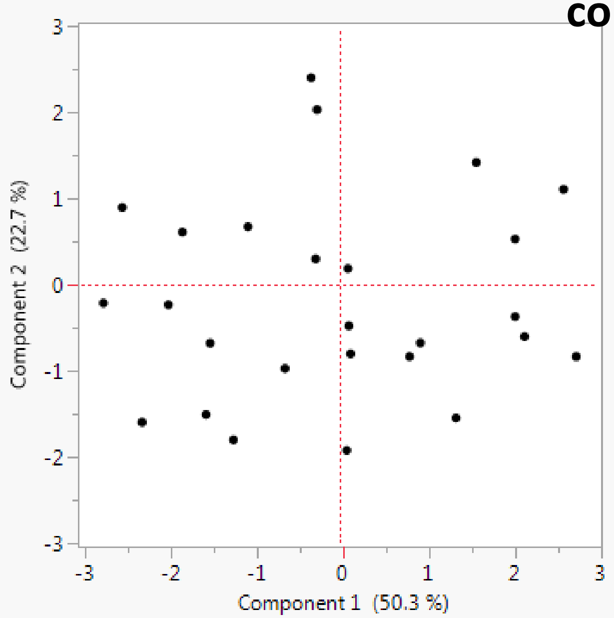
- Determine ideal number of dimensions (most popular):
  - Scree plot: Number of eigenvalues before the elbow
  - Number of eigenvalues larger than 1
  - Dimensions that sum up to ~80% of variance
  - All dimensions with coherent substantive meaning

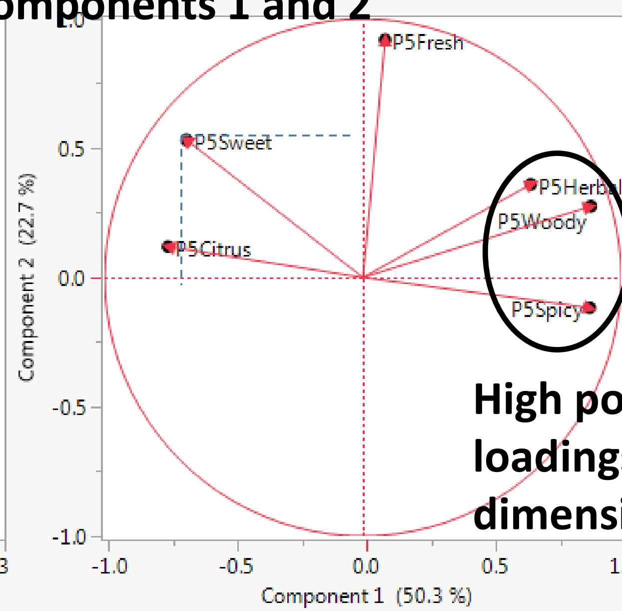
PRINCIPAL COMPONENTS ANALYSIS

Principal Components: on Correlations

Summary Plots







Sweet loads on components 1 and 2

High positive loadings on dimension 1

Select component Component 1 Component 2

Formatted Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
P5Woody	0.882336	0.275783	0.129051	-0.072320	0.049161	0.348032
P5Spicy	0.877634	-0.116755	0.293271	-0.076223	0.291578	-0.198218
P5Herbal	0.650136	0.360436	-0.471089	0.469285	0.034001	-0.064028
P5Fresh	0.085793	0.919051	0.309946	-0.072982	-0.171351	-0.131266
P5Sweet	-0.683711	0.532344	-0.308497	-0.277201	0.276317	0.028074
P5Citrus	-0.754859	0.119329	0.400728	0.473805	0.156001	0.080856

Suppress Absolute Loading Value Less Than 0.3

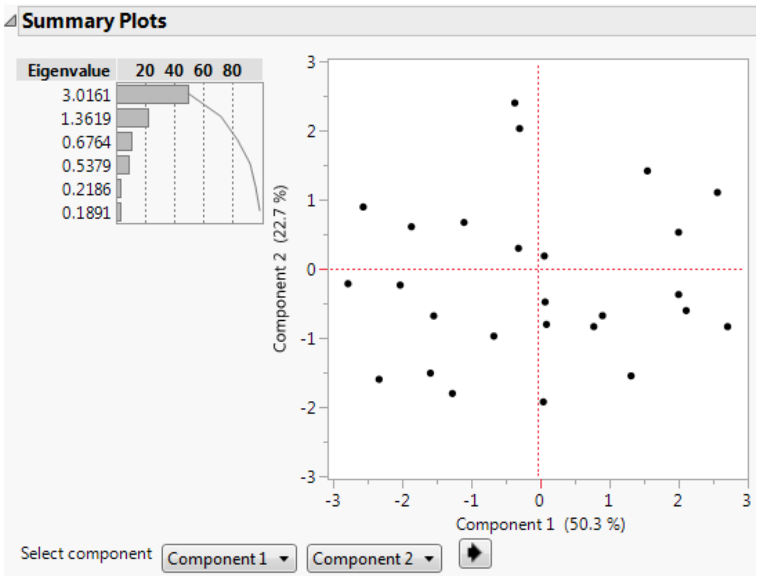
Dim Text 0.3

- **Loadings** help us interpret meaning of components
- Loadings are correlation coefficients between components/dimensions and variables

MULTIVARIATE FLAVORS OF JMP

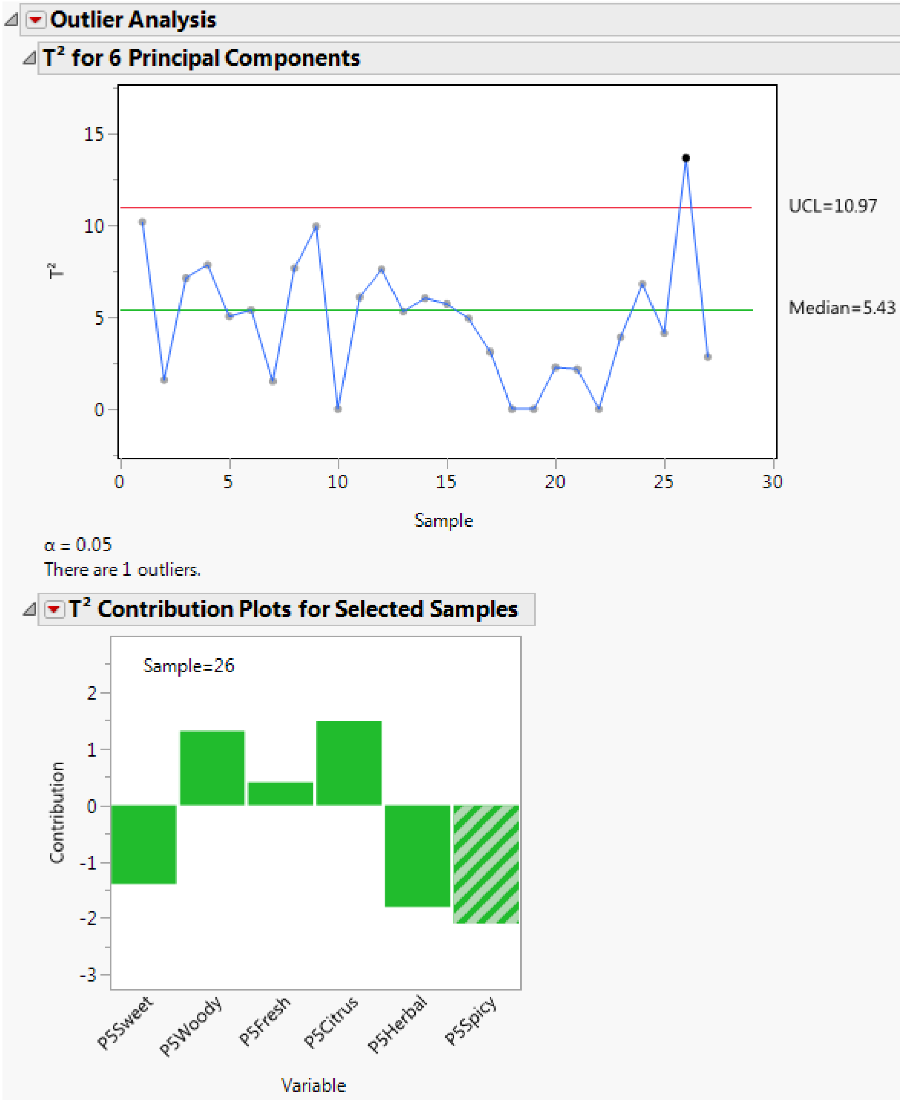
PRINCIPAL COMPONENTS ANALYSIS

83/2 Cols	P5Sweet	P5Woody	P5Fresh	P5Citrus	P5Herbal	P5Spicy	Prin1	Prin2
1	3	4	3	4	4	7	0.1271634586	-0.803718291
2	1	5	4	3	5	5	0.9373510535	-0.677519483
3	3	1	4	7	4	1	-2.744656595	-0.215015613
4	2	1	2	4	5	1	-1.551618211	-1.508841984
5	5	3	4	4	5	1	-1.826963841	0.6089708875
6	1	7	7	2	7	7	2.5998245145	1.1043223099
7	4	3	5	4	5	3	-1.06571194	0.6704409028
8	5	6	7	3	6	2	-0.331369201	2.3996353047
9	4	1	2	2	1	1	-2.29510529	-1.598492863



- **Component Scores** characterize the degree of endorsement of each dimension for every observation
- PCA Scores can be used in a variety of subsequent analyses (e.g., predictive models)
- **Score plot** facilitates identification of observations with very high/low scores and those close to the centroid

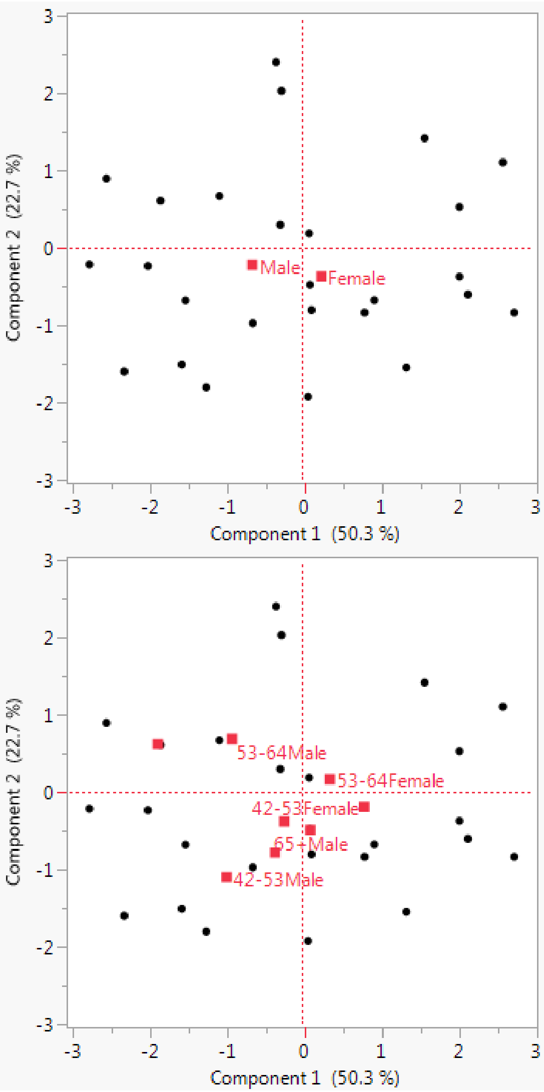
PRINCIPAL COMPONENTS ANALYSIS



- New to JMP 14: **Outlier Analysis**
- Enables identification of out-of-control points (multivariate outliers) through the T^2 statistic
- **Contribution plots** indicate exactly which variables are contributing most to the extreme observations

MULTIVARIATE FLAVORS OF JMP

PRINCIPAL COMPONENTS ANALYSIS



- **Supplementary variables** can be included to enrich interpretation of components
- Supplementary points are displayed at the averages of the scores of the corresponding respondents
  - E.g., Average of component scores for males and females results in coordinates for points in each dimension/component
- Creating “interaction” variables enables more nuanced interpretation of the plots

MULTIPLE CORRESPONDENCE ANALYSIS

MULTIPLE CORRESPONDENCE ANALYSIS

- Used with categorical data (ordinal or nominal)
- Goals of Analysis:
  - Goals are similar to PCA but there is much more emphasis on graphical displays
  - Identify underlying structure of data
  - Study inter-association of variable *categories*
  - Study inter-individual variability
    - Extract dimensions that distinguish individuals
    - Identify multivariate outliers

MULTIPLE CORRESPONDENCE ANALYSIS

- Key output:
  - MCA Map
  - Principal inertias (eigenvalues)
  - Principal coordinates (loadings)
  - Dimension contributions to column inertia (variance overlap between point and dimension)
  - Column contributions to total inertia
  - Column contributions to individual dimensions
  - Dimension contributions to total inertia (explained variance of each dimension)

MULTIPLE CORRESPONDENCE ANALYSIS

- Known as the decomposition of an Indicator matrix or a Burt matrix

Indicator	vs	Burt
$\mathbf{Z}$		$\mathbf{C} = \mathbf{Z}^{\mathbf{T}}\mathbf{Z}$

MULTIPLE CORRESPONDENCE ANALYSIS

The Data:

	P5Woody	P5Fresh	P5Citrus
1	Somewhat Woody	Somewhat Fresh	Somewhat Citrus
2	Somewhat Woody	Somewhat Fresh	Somewhat Citrus
3	Not Woody	Somewhat Fresh	Very Citrus
4	Not Woody	Not Fresh	Somewhat Citrus
5	Somewhat Woody	Somewhat Fresh	Somewhat Citrus
6	Very Woody	Very Fresh	Not Citrus

Raw Table
Categorical Variables

	Not Woody	Somewhat Woody	Very Woody	Not Fresh	Somewhat Fresh	Very Fresh	Not Citrus	Somewhat Citrus	Very Citrus
1	0	1	0	0	1	0	0	1	0
2	0	1	0	0	1	0	0	1	0
3	1	0	0	0	1	0	0	0	1
4	1	0	0	1	0	0	0	1	0
5	0	1	0	0	1	0	0	1	0
6	0	0	1	0	0	1	1	0	0

Indicator Table $\mathbf{Z}$
Concatenated Categories

	Not Woody	Somewhat Woody	Very Woody	Not Fresh	Somewhat Fresh	Very Fresh	Not Citrus	Somewhat Citrus	Very Citrus
1	2	0	0	1	1	0	0	1	1
2	0	3	0	0	3	0	0	3	0
3	0	0	1	0	0	1	1	0	0
4	1	0	0	1	0	0	0	1	0
5	1	3	0	0	4	0	0	3	1
6	0	0	1	0	0	1	1	0	0
7	0	0	1	0	0	1	1	0	0
8	1	3	0	1	3	0	0	4	0
9	1	0	0	0	1	0	0	0	1

Burt Table $\mathbf{C} = \mathbf{Z}^{\mathbf{T}}\mathbf{Z}$
Categories X Categories
Count on diagonal
Contingency on off-diagonal

MULTIPLE CORRESPONDENCE ANALYSIS

From PCA

$$\mathbf{D_r}^{-\frac{1}{2}}(\mathbf{X} - \mathbf{1}\hat{\boldsymbol{\mu}}^{\mathbf{T}})\mathbf{D_c}^{-\frac{1}{2}}$$

SVD of:

Weights: row marginals and column marginals (masses) of **P**

$$\mathbf{D_r}^{-\frac{1}{2}}(\mathbf{P} - \mathbf{r}\mathbf{c}^{\mathbf{T}})\mathbf{D_c}^{-\frac{1}{2}}$$

Correspondence Matrix

Row and column marginals (masses) to center **P**

$$\mathbf{P} = \frac{1}{NQ}\mathbf{Z}$$

N = number of rows

Q = number of variables (row sums of Z)

Z = indicator matrix

MULTIPLE CORRESPONDENCE ANALYSIS

SVD of:

$$\mathbf{D_r}^{-\frac{1}{2}}(\mathbf{P} - \mathbf{rc}^{\mathbf{T}})\mathbf{D_c}^{-\frac{1}{2}} = \mathbf{USV}^{\mathbf{T}}$$

Principal Inertias: $\mathbf{S}^2$

Standard Coordinates: $\mathbf{D_c}^{-\frac{1}{2}}\mathbf{V}$

Principal Coordinates for Columns: $\mathbf{D_c}^{-\frac{1}{2}}\mathbf{VS}$

Principal Coordinates for Rows: $\mathbf{D_r}^{-\frac{1}{2}}\mathbf{US}$

From PCA

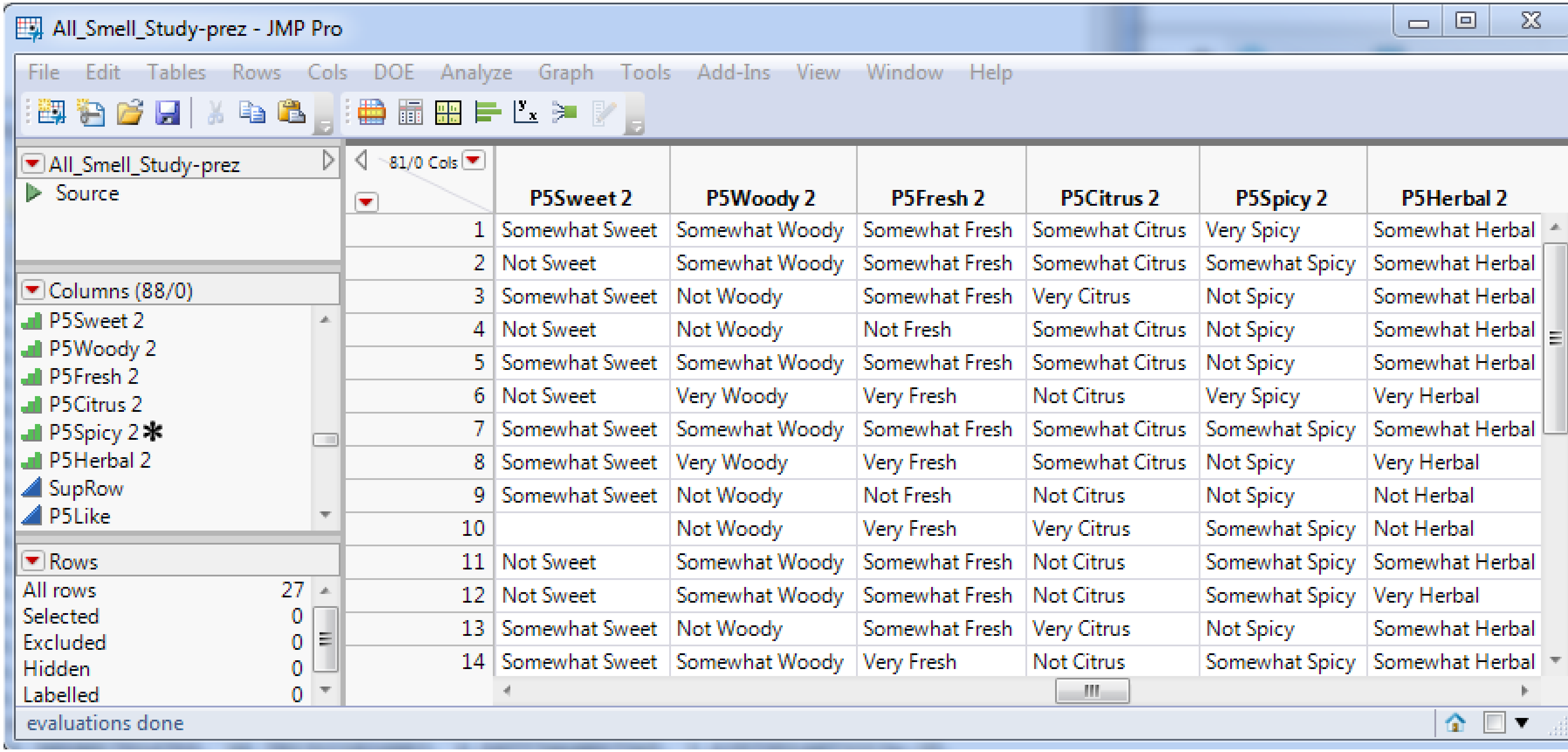
Eigenvalues: $\mathbf{S}^2$

Eigenvectors: $\mathbf{V}$

Loadings: $\mathbf{VS}$

Scores: $\mathbf{US}$

MULTIPLE CORRESPONDENCE ANALYSIS



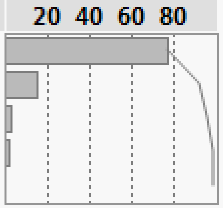
	P5Sweet 2	P5Woody 2	P5Fresh 2	P5Citrus 2	P5Spicy 2	P5Herbal 2
1	Somewhat Sweet	Somewhat Woody	Somewhat Fresh	Somewhat Citrus	Very Spicy	Somewhat Herbal
2	Not Sweet	Somewhat Woody	Somewhat Fresh	Somewhat Citrus	Somewhat Spicy	Somewhat Herbal
3	Somewhat Sweet	Not Woody	Somewhat Fresh	Very Citrus	Not Spicy	Somewhat Herbal
4	Not Sweet	Not Woody	Not Fresh	Somewhat Citrus	Not Spicy	Somewhat Herbal
5	Somewhat Sweet	Somewhat Woody	Somewhat Fresh	Somewhat Citrus	Not Spicy	Somewhat Herbal
6	Not Sweet	Very Woody	Very Fresh	Not Citrus	Very Spicy	Very Herbal
7	Somewhat Sweet	Somewhat Woody	Somewhat Fresh	Somewhat Citrus	Somewhat Spicy	Somewhat Herbal
8	Somewhat Sweet	Very Woody	Very Fresh	Somewhat Citrus	Not Spicy	Very Herbal
9	Somewhat Sweet	Not Woody	Not Fresh	Not Citrus	Not Spicy	Not Herbal
10		Not Woody	Very Fresh	Very Citrus	Somewhat Spicy	Not Herbal
11	Not Sweet	Somewhat Woody	Somewhat Fresh	Not Citrus	Somewhat Spicy	Somewhat Herbal
12	Not Sweet	Somewhat Woody	Somewhat Fresh	Not Citrus	Somewhat Spicy	Very Herbal
13	Somewhat Sweet	Not Woody	Somewhat Fresh	Very Citrus	Not Spicy	Somewhat Herbal
14	Somewhat Sweet	Somewhat Woody	Very Fresh	Not Citrus	Somewhat Spicy	Somewhat Herbal

Data can be ordinal or categorical

MULTIPLE CORRESPONDENCE ANALYSIS

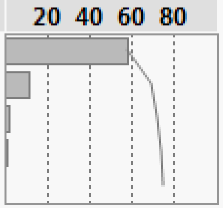
Benzecri Adjusted Inertia

Inertia	Adjusted Inertia	Percent	Cumulative Percent
0.49657	0.15672	77.78	77.78
0.31287	0.03078	15.28	93.06
0.23913	0.00756	3.75	96.81
0.22491	0.00488	2.42	99.24
0.19936	0.00154	0.76	100.00



Greenacre Adjusted Inertia

Inertia	Adjusted Inertia	Percent	Cumulative Percent
0.49657	0.15672	58.78	58.78
0.31287	0.03078	11.54	70.32
0.23913	0.00756	2.84	73.16
0.22491	0.00488	1.83	74.99
0.19936	0.00154	0.58	75.57

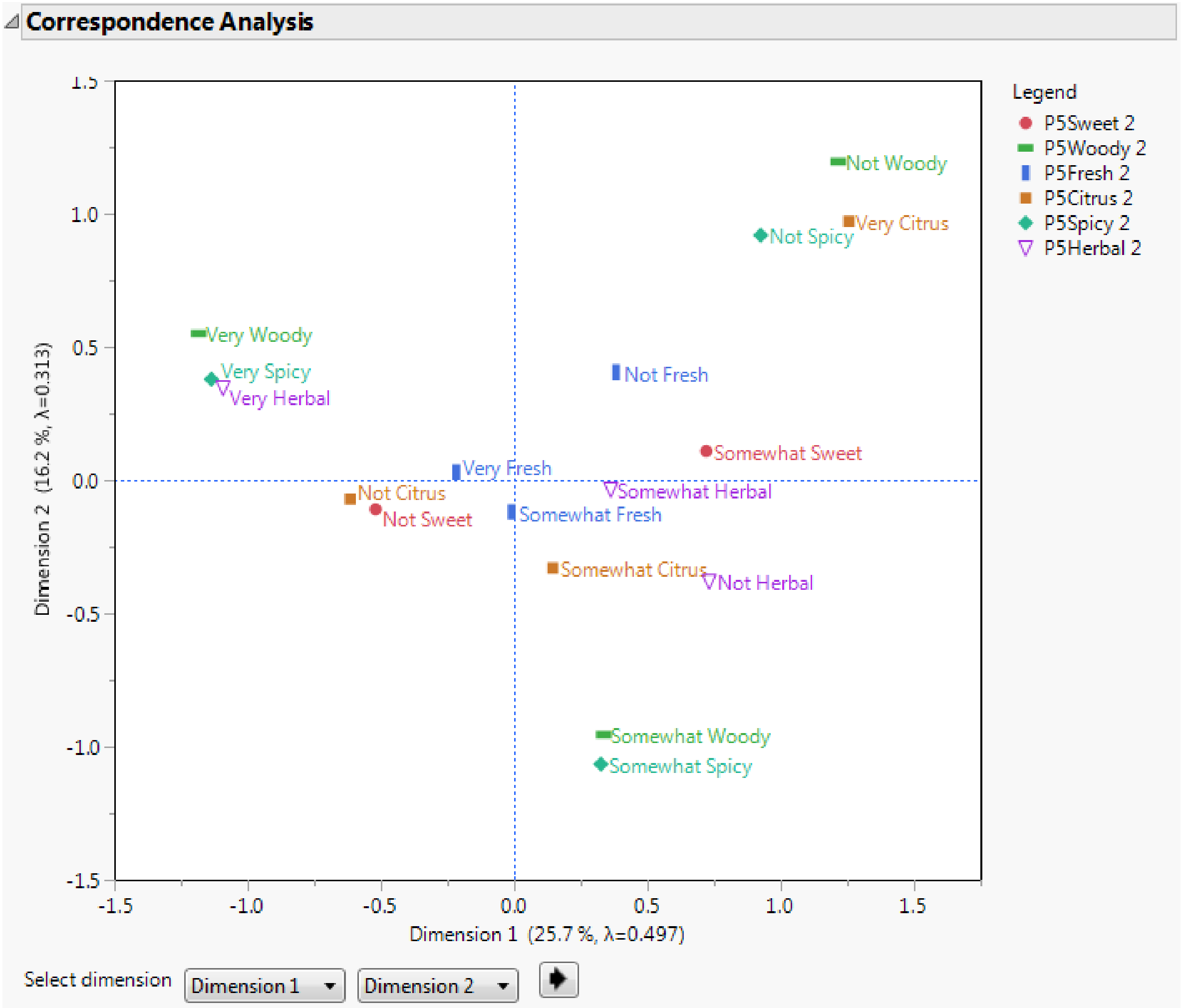


- Determine ideal number of dimensions:
  - Pareto plot: Use as scree plot. Number of eigenvalues before the elbow
  - Dimensions that sum up to ~80% of **adjusted** percent of inertia
  - All dimensions with coherent substantive meaning

- Adjusted inertias give a more accurate idea of the percentage of explained variance
  - Benzécri adjusted inertias are computed as percentages of the sum of eigenvalues that are greater or equal to 1/number of column variables
    - Inertias tend to be overestimated
  - Greenacre adjusted inertias are less optimistic than Benzécri's

MULTIVARIATE FLAVORS OF JMP

MULTIPLE CORRESPONDENCE ANALYSIS



- MCA Map is the key feature and contains huge amounts of information
- Points are plotted according to the column coordinates
- Project points onto each dimension to help interpret dimension's meaning
- Points close to each other are more strongly associated

MULTIPLE CORRESPONDENCE ANALYSIS

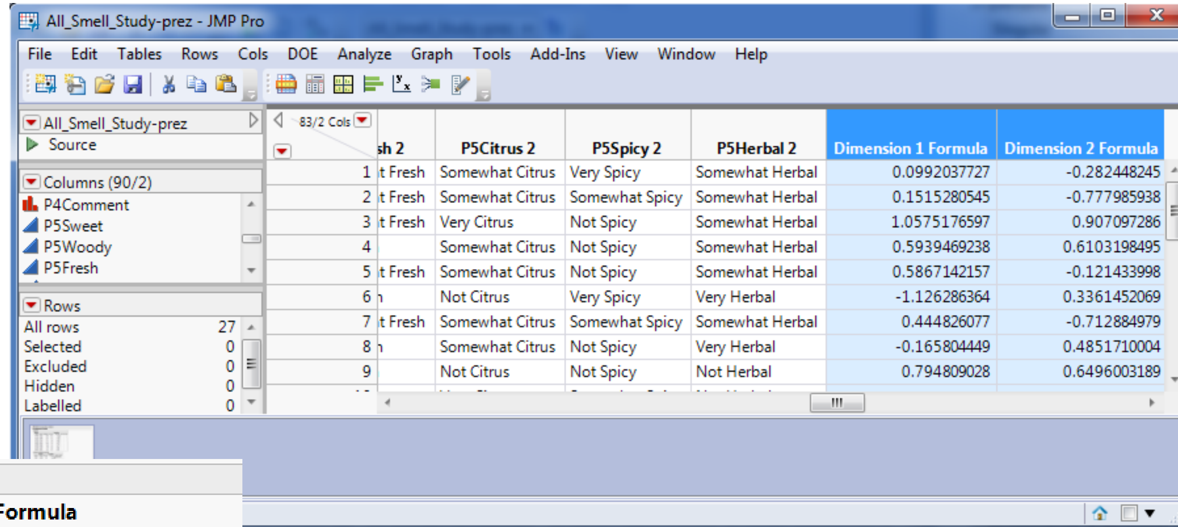
Column Coordinates

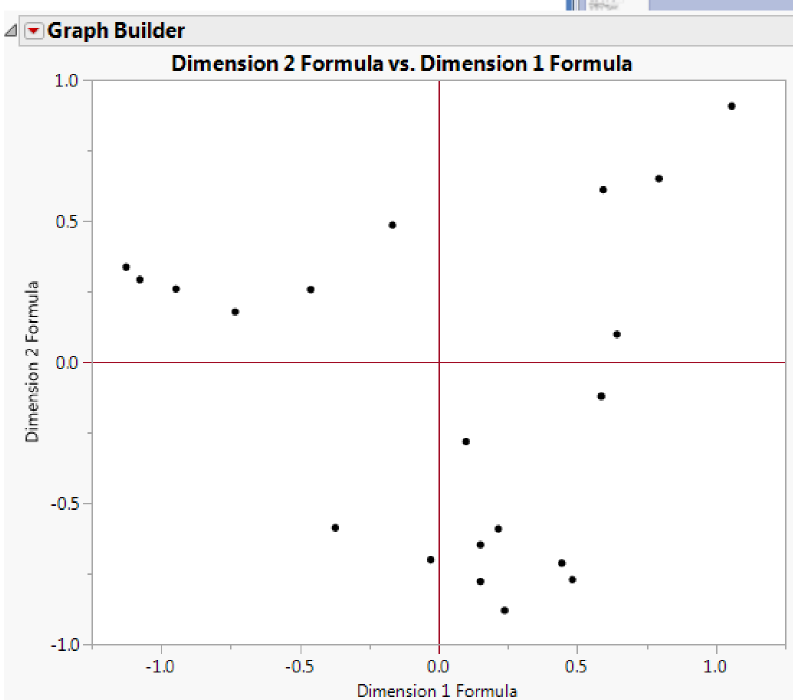
Y	Category	Dimension 1	Dimension 2	Dimension 3
P5Sweet 2	Not Sweet	-0.520	-0.109	0.184
P5Sweet 2	Somewhat Sweet	0.720	0.110	-0.477
P5Woody 2	Not Woody	1.216	1.196	0.422
P5Woody 2	Somewhat Woody	0.337	-0.955	-0.124
P5Woody 2	Very Woody	-1.183	0.551	-0.052
P5Fresh 2	Not Fresh	0.382	0.405	0.833
P5Fresh 2	Somewhat Fresh	-0.009	-0.119	-0.438
P5Fresh 2	Very Fresh	-0.217	0.030	0.536
P5Citrus 2	Not Citrus	-0.615	-0.071	0.393
P5Citrus 2	Somewhat Citrus	0.145	-0.330	-0.482
P5Citrus 2	Very Citrus	1.256	0.972	0.303
P5Spicy 2	Not Spicy	0.925	0.920	-0.178
P5Spicy 2	Somewhat Spicy	0.325	-1.065	0.376
P5Spicy 2	Very Spicy	-1.136	0.379	-0.228
P5Herbal 2	Not Herbal	0.731	-0.379	1.291
P5Herbal 2	Somewhat Herbal	0.362	-0.034	-0.740
P5Herbal 2	Very Herbal	-1.092	0.347	0.179

- **Column (principal) Coordinates** are like PCA loadings. They help us interpret meaning of components
- Column Coordinates are particularly helpful when MCA map is too crowded: we can sort them to identify which categories are at the extremes

MULTIVARIATE FLAVORS OF JMP

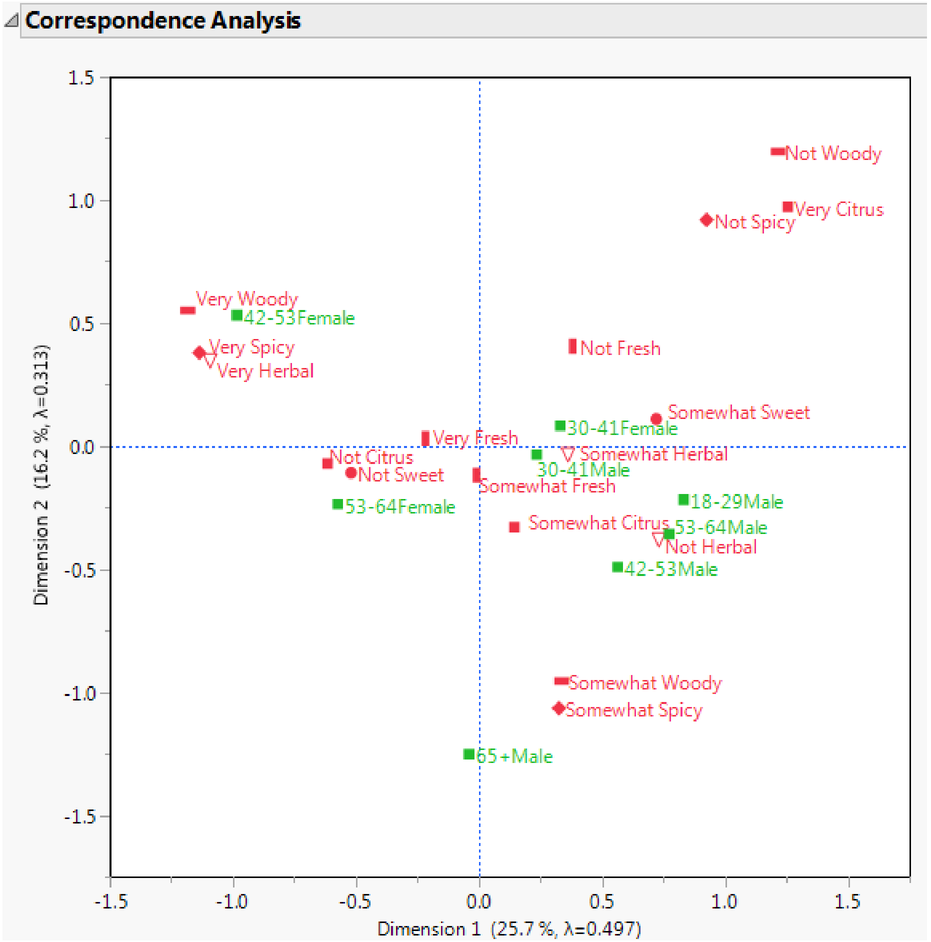
MULTIPLE CORRESPONDENCE ANALYSIS





- **Save Coordinate Formula** saves **Principal Row Coordinates** to the data table, which characterize the degree of endorsement of each dimension for every observation
- As with PCA Scores, these can be used in a variety of subsequent analyses (e.g., predictive models)
- We can plot row coordinates to identify observations with very high/low scores and those close to the centroid

MULTIPLE CORRESPONDENCE ANALYSIS



- As with PCA, **Supplementary variables** can be included to enrich interpretation of dimensions
- Supplementary points are displayed at the averages of the principal row coordinates of the corresponding respondents
  - E.g., Average for males and females results in coordinates for points in each dimension
- Creating “interaction” variables enables more nuanced interpretation of the plots

MULTIPLE FACTOR ANALYSIS

MULTIPLE FACTOR ANALYSIS

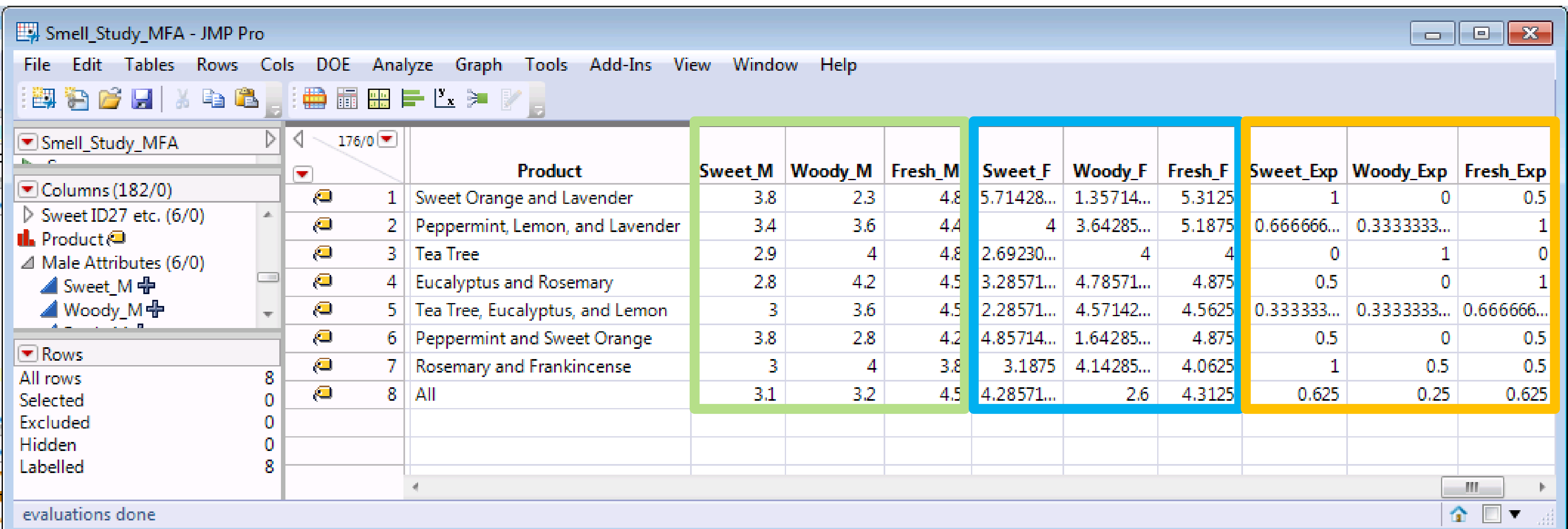
- Used with continuous data
- Goals of Analysis:
  - Identify underlying structure of data from **multiple sources**
  - Study inter-association of variables **across sources** of data
    - Compare information from multiple data tables
  - Reduce dimensionality of data accounting for multiple-source structure (analogous to PCA on Corr vs Cov)
  - Study inter-association of observations (**products** in CR)
    - Extract dimensions that distinguish observations
    - Identify multivariate outliers within and across sources of data
- Graphical displays are also emphasized

MULTIVARIATE FLAVORS OF JMP

MULTIPLE FACTOR ANALYSIS

Data:

Block *1* | Block *2* | ... | Block *g*



	Product	Sweet_M	Woody_M	Fresh_M	Sweet_F	Woody_F	Fresh_F	Sweet_Exp	Woody_Exp	Fresh_Exp
1	Sweet Orange and Lavender	3.8	2.3	4.8	5.71428...	1.35714...	5.3125	1	0	0.5
2	Peppermint, Lemon, and Lavender	3.4	3.6	4.4	4	3.64285...	5.1875	0.666666...	0.3333333...	1
3	Tea Tree	2.9	4	4.8	2.69230...	4	4	0	1	0
4	Eucalyptus and Rosemary	2.8	4.2	4.5	3.28571...	4.78571...	4.875	0.5	0	1
5	Tea Tree, Eucalyptus, and Lemon	3	3.6	4.5	2.28571...	4.57142...	4.5625	0.333333...	0.3333333...	0.666666...
6	Peppermint and Sweet Orange	3.8	2.8	4.2	4.85714...	1.64285...	4.875	0.5	0	0.5
7	Rosemary and Frankincense	3	4	3.8	3.1875	4.14285...	4.0625	1	0.5	0.5
8	All	3.1	3.2	4.5	4.28571...	2.6	4.3125	0.625	0.25	0.625

MULTIPLE FACTOR ANALYSIS

- Key output:
  - Consensus Map
  - Eigenvalues (aka squared singular values)
  - Eigenvectors
  - Loadings
  - Percent of variance explained by each dimension
  - Individual component scores
  - RV Correlations
  - Block Partial Contributions
  - Block Partial Scores

MULTIPLE FACTOR ANALYSIS

SVD of:

Row Weights | Centered and/or Normalized Matrix | Column Weights

$$\mathbf{X_c} = \mathbf{X} - \mathbf{1}\hat{\boldsymbol{\mu}}^{\mathbf{T}}$$

$$\mathbf{D_{ss}} = diag\left(\mathbf{X_c^T X_c}\right)$$

Row weights (N-1)

Column weights (S^2)

$$\mathbf{D_r}^{-\frac{1}{2}}\left(\mathbf{X_c}\mathbf{D_{ss}}^{-\frac{1}{2}}\right)\mathbf{D_c}^{-\frac{1}{2}}$$

Centered data

Divide each column by $\sqrt{SS}$ to normalize $\boldsymbol{X_c}$

$$\begin{bmatrix} \mathbf{D_{c1}}^{-\frac{1}{2}} & & \\ & \ddots & \\ & & \mathbf{D_{cg}}^{-\frac{1}{2}} \end{bmatrix}$$

MULTIPLE FACTOR ANALYSIS

SVD of:

$$\mathbf{D_r}^{-\frac{1}{2}}\left(\mathbf{X_c}\mathbf{D_{ss}}^{-\frac{1}{2}}\right)\mathbf{D_c}^{-\frac{1}{2}} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathbf{T}}$$

MFA Eigenvalues: $N\mathbf{S}^2$

MFA Eigenvectors: $\mathbf{D_c}^{-\frac{1}{2}}\mathbf{V}$

MFA Loadings: $\sqrt{N}\mathbf{D_c}^{-\frac{1}{2}}\mathbf{V}\mathbf{S}$

MFA Component Scores: $\sqrt{N}\mathbf{D_r}^{-\frac{1}{2}}\mathbf{U}\mathbf{S}$

From PCA

Eigenvalues: $\mathbf{S}^2$

Eigenvectors: $\mathbf{V}$

Loadings: $\mathbf{VS}$

Scores: $\mathbf{US}$

MULTIPLE FACTOR ANALYSIS

SVD of:

$$\mathbf{D_r}^{-\frac{1}{2}}\left(\mathbf{X_c}\mathbf{D_{ss}}^{-\frac{1}{2}}\right)\mathbf{D_c}^{-\frac{1}{2}} = \mathbf{USV}^{\mathbf{T}}$$

MFA Eigenvalues: $N\mathbf{S}^2$

MFA Eigenvectors: $\mathbf{D_c}^{-\frac{1}{2}}\mathbf{V}$

MFA Loadings: $\sqrt{N}\mathbf{D_c}^{-\frac{1}{2}}\mathbf{VS}$

MFA Component Scores: $\sqrt{N}\mathbf{D_r}^{-\frac{1}{2}}\mathbf{US}$

From MCA

Principal Inertias: $\mathbf{S}^2$

Standard Coordinates: $\mathbf{D_c}^{-\frac{1}{2}}\mathbf{V}$

Principal Coordinates for Columns: $\mathbf{D_c}^{-\frac{1}{2}}\mathbf{VS}$

Principal Coordinates for Rows: $\mathbf{D_r}^{-\frac{1}{2}}\mathbf{US}$

MULTIVARIATE FLAVORS OF JMP

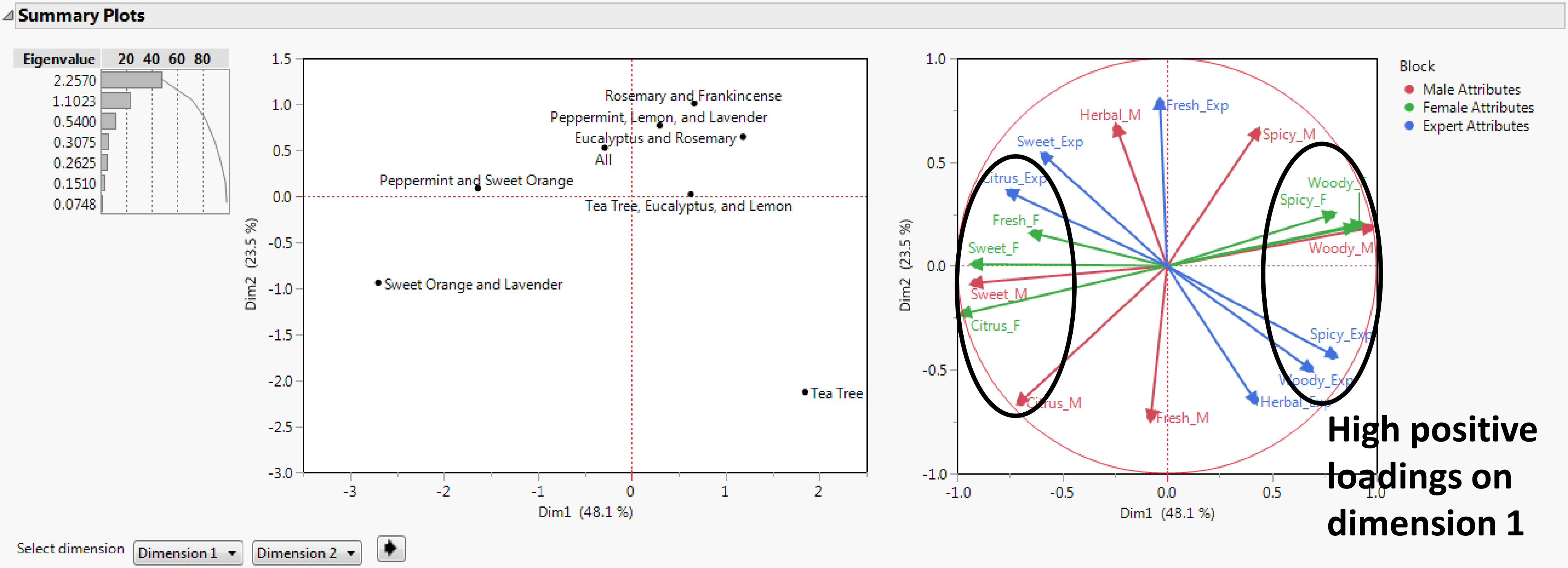
MULTIPLE FACTOR ANALYSIS

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	2.2570	48.072		48.072
2	1.1023	23.478		71.550
3	0.5400	11.501		83.052
4	0.3075	6.548		89.600
5	0.2625	5.592		95.192
6	0.1510	3.216		98.407
7	0.0748	1.593		100.000

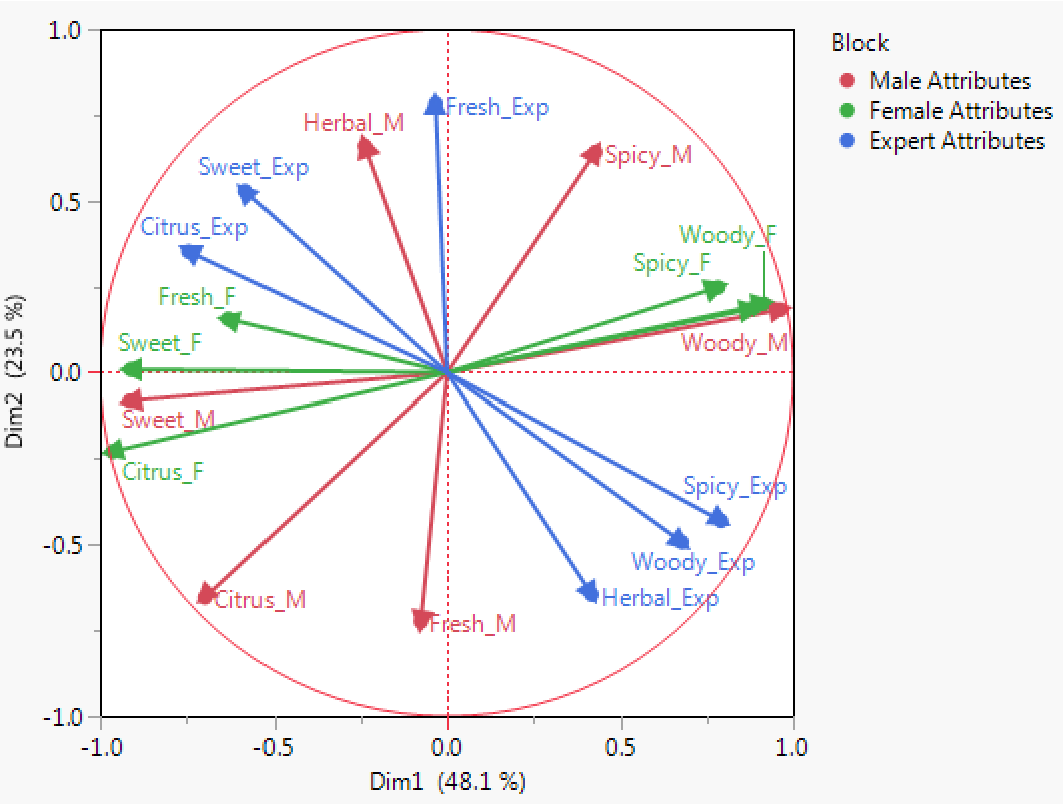
- Determine ideal number of dimensions (most popular):
  - Scree plot: Number of eigenvalues before the elbow
    - Can use Pareto plot or plot eigenvalues in GraphBuilder
  - ~~Number of eigenvalues larger than 1~~
    - Doesn’t apply anymore in MFA
  - Dimensions that sum up to ~80% of variance
  - All dimensions with coherent substantive meaning

MULTIPLE FACTOR ANALYSIS



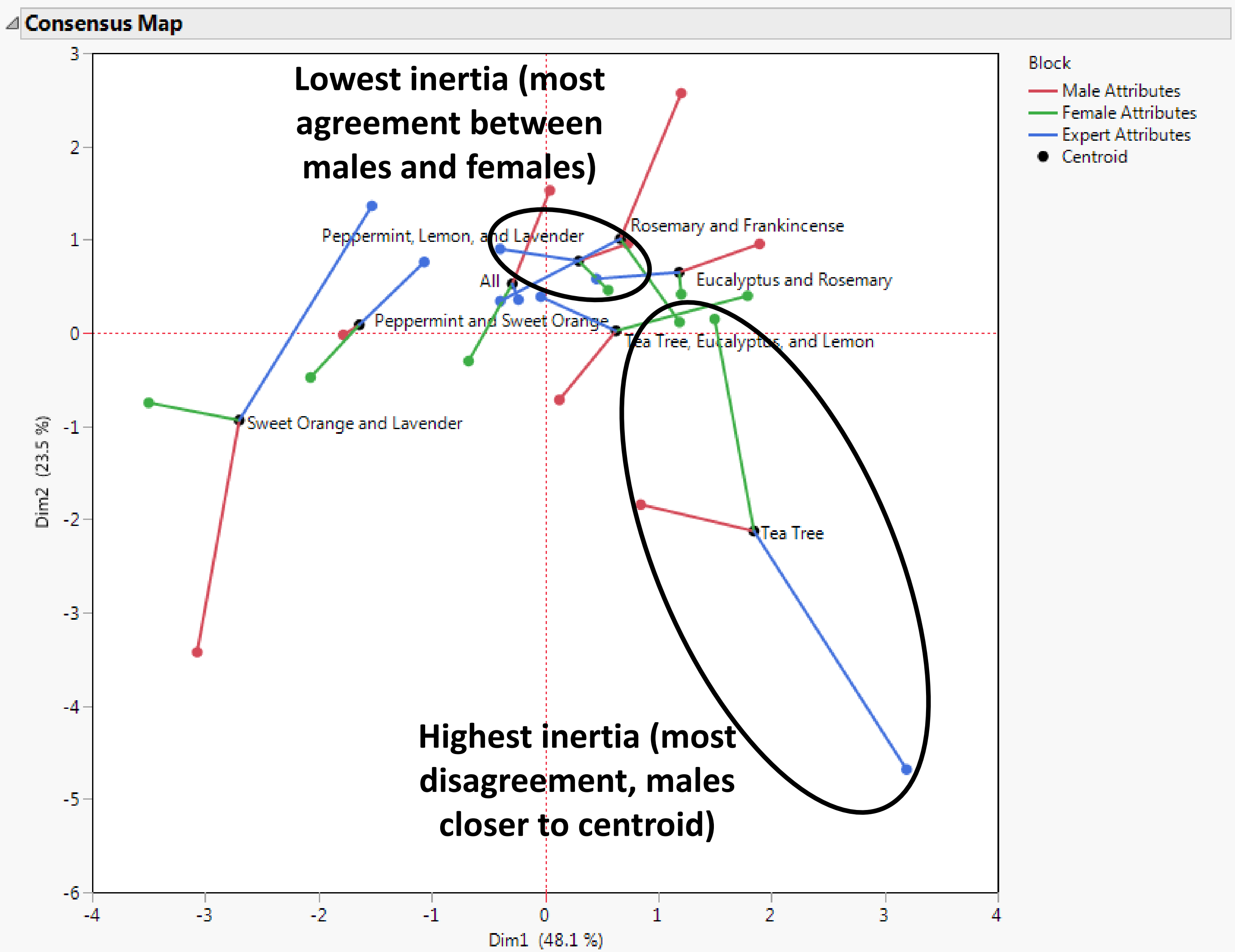
- Use **Score Plot** to identify how products “score” in each dimension
- E.g., Eucalyptus and Rosemary together with Tea Tree were rated as highly woody and spicy, whereas Sweet Orange and Lavender is correctly identified as high in sweet and citrus.
- Use **Loading Plot** to interpret meaning of consensus components
- Loadings are correlation coefficients between components/dimensions and variables

MULTIPLE FACTOR ANALYSIS



- Use **Loading Plot** to compare structure across sources
  - E.g., Males have higher dimensionality than females
- Vectors close to each other are more highly correlated
  - E.g., All sources mostly agree on perceptions of sweet and citrus
- Opposing vectors have opposite meaning
  - E.g., experts and males have opposite interpretation of freshness

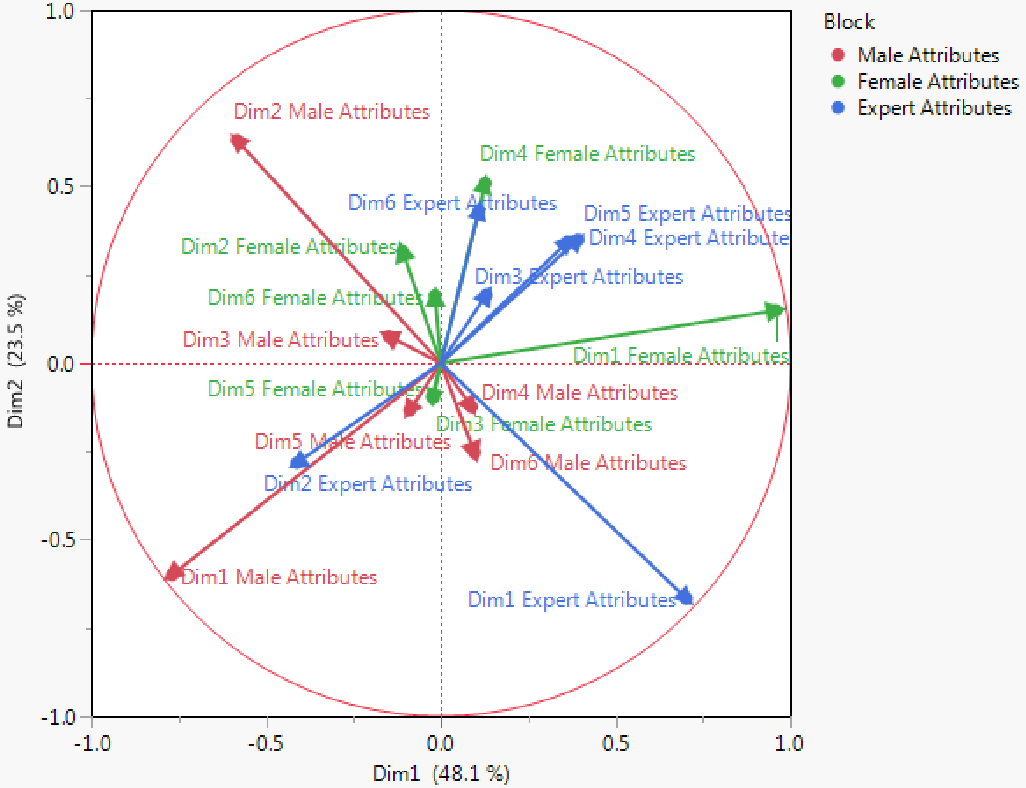
MULTIPLE FACTOR ANALYSIS



- Use **Consensus Map** to identify agreement or disagreement between sources.
  - "**Highlight Product**" slider facilitates this task by highlighting low/high inertia products
- Tea tree was experienced most differently across all
- Peppermint, lemon, and lavender was experienced most similarly across all
- Combination of "All" scents is closest to the origin

MULTIPLE FACTOR ANALYSIS

Partial Axes Plot



Select dimension Dimension 1 Dimension 2

RV Correlations

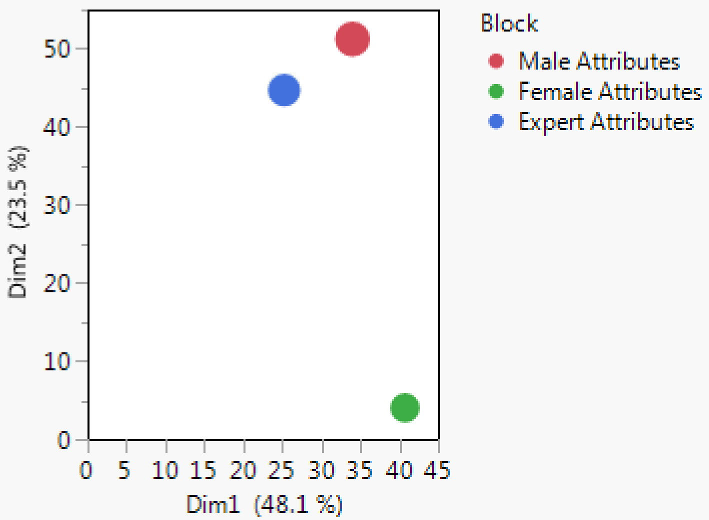
	Male Attributes	Female Attributes	Expert Attributes	Centroid
Male Attributes	1.00000	0.65948	0.38017	0.85059
Female Attributes	0.65948	1.00000	0.40677	0.84502
Expert Attributes	0.38017	0.40677	1.00000	0.73138
Centroid	0.85059	0.84502	0.73138	1.00000

- **Partial Axes Plot** displays correlations between separate PCA dimensions across sources with MFA (consensus) dimensions
  - 1st MFA dimension is very much like females' 1st dimension from their own separate PCA
  - 2nd MFA dimension is most like males' own 2nd PCA dimension
- **RV Correlations** quantify the level of shared variance across sources (squared correlation coefficient between matrices)
  - Experts have the least in common with males and females

MULTIPLE FACTOR ANALYSIS

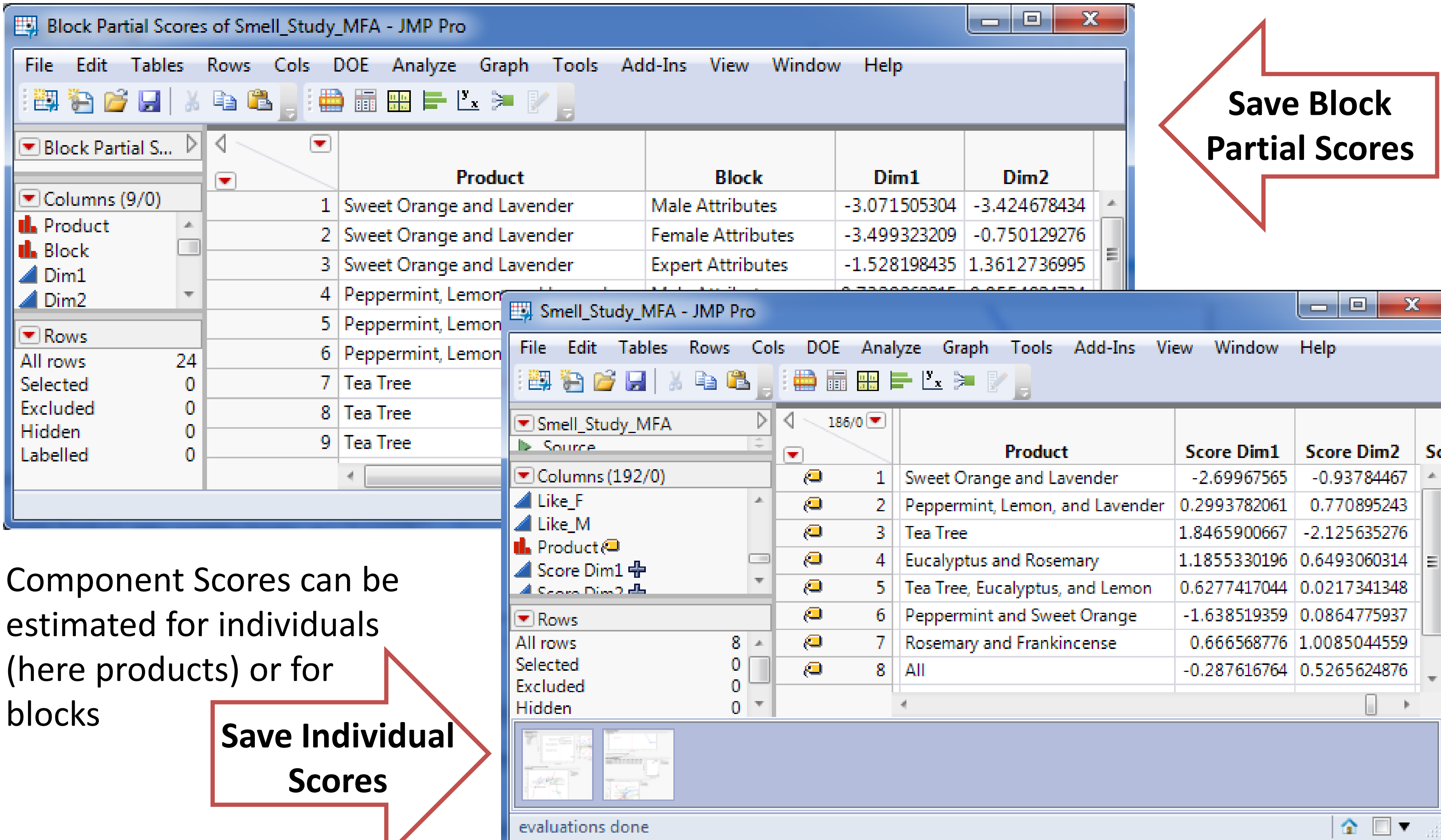
Block Partial Contributions

	Dim1	Dim2	Dim3	Dim4	Dim5	Dim6	Dim7
Male Attributes	33.98832	51.24859	33.95718	40.86565	51.29841	21.67875	49.27503
Female Attributes	40.74569	4.04093	26.97737	21.21832	22.15244	14.03899	7.88003
Expert Attributes	25.26599	44.71048	39.06545	37.91602	26.54915	64.28226	42.84494



- **Block Partial Contributions** quantify the percentage of contribution to each MFA dimension from each block (i.e., source)
  - E.g., 1st MFA dimension is mostly influenced by females' responses and least by experts' responses

MULTIPLE FACTOR ANALYSIS



Save Block Partial Scores

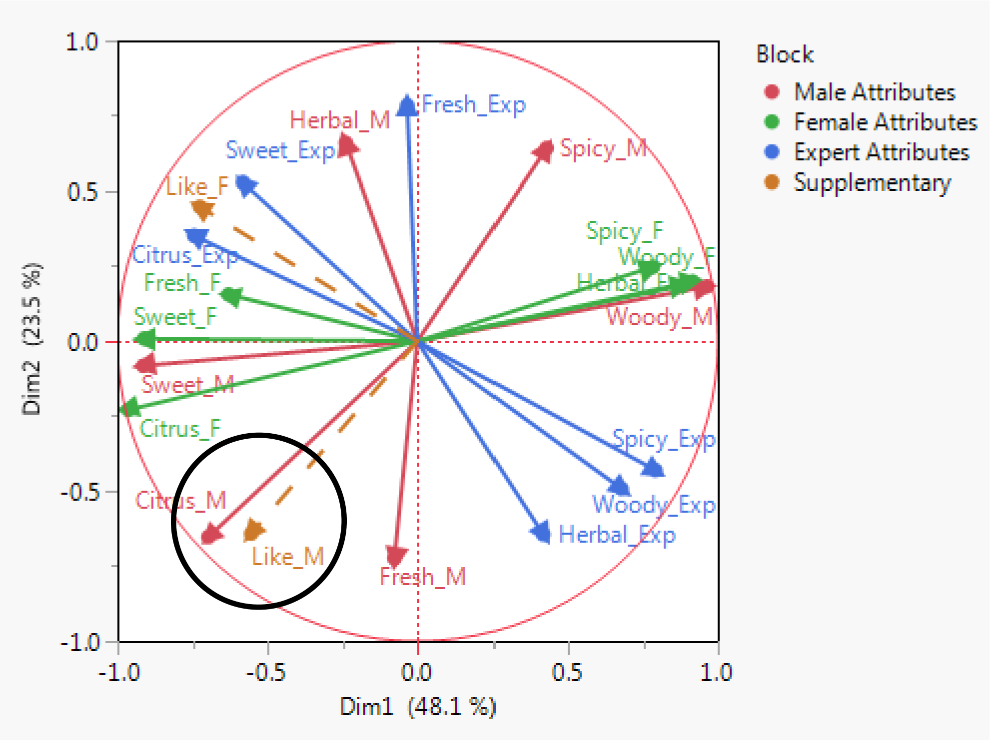
Component Scores can be estimated for individuals (here products) or for blocks

Save Individual Scores

MULTIPLE FACTOR ANALYSIS

Supplementary variables enrich the interpretation of our findings

Product	Like_F	Like_M
Sweet Orange and Lavender	3.8235294	3.4
Peppermint, Lemon, and Lavender	3.5882353	2.5
Tea Tree	2.6470588	3.1
Eucalyptus and Rosemary	3.4705882	2.9
Tea Tree, Eucalyptus, and Lemon	2.9411765	2.9
Peppermint and Sweet Orange	3.4705882	3.1
Rosemary and Frankincense	3.2941176	2.8
All	3.291176	3



Males liked best the scents they perceived as citrus and somewhat sweet and fresh, and didn’t like those scents they perceived as spicy.

Questions?

Improve understanding of techniques by drawing on their similarities