Canonical Correlation Analysis

Statistics 407, ISU

Definition

Find the linear combination of X, $\mathbf{a}'\mathbf{X}$ that is most correlated with a linear combination of Y, $\mathbf{b}'\mathbf{Y}$, ie maximize $r = cor(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y})$.

Then find the next most correlated pairs of linear combinations. ...

Determine how many linear combinations are needed to describe most of the association.

Definition

Find the linear combination of $\mathbf{X}$, $\mathbf{a}'\mathbf{X}$ that is most correlated with a linear combination of Y, $\mathbf{b}'\mathbf{Y}$, ie maximize $r = cor(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y})$.

Then find the next most correlated pairs of linear combinations. ...

Determine how many linear combinations are needed to describe most of the association.

Definition

n x p

Find the linear combination of $\mathbf{X}$, $\mathbf{a}'\mathbf{X}$ that is most correlated with a linear combination of Y, $\mathbf{b}'\mathbf{Y}$, ie maximize $r = cor(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y})$.

Then find the next most correlated pairs of linear combinations. ...

Determine how many linear combinations are needed to describe most of the association.

Definition

n x p

Find the linear combination of $\mathbf{X}$, $\mathbf{a}'\mathbf{X}$ that is most correlated with a linear combination of $\mathbf{Y}$, $\mathbf{b}'\mathbf{Y}$, ie maximize $r = cor(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y})$.

Then find the next most correlated pairs of linear combinations. ...

Determine how many linear combinations are needed to describe most of the association.

Definition

n x p

Find the linear combination of $\mathbf{X}$, $\mathbf{a}'\mathbf{X}$ that is most correlated with a linear combination of $\mathbf{Y}$, $\mathbf{b}'\mathbf{Y}$, ie maximize $r = cor(\mathbf{a}'\mathbf{X}, \mathbf{b}'\mathbf{Y})$.

n x q

Then find the next most correlated pairs of linear combinations. ...

Determine how many linear combinations are needed to describe most of the association.

Solution

Find the eigenvectors of

$$S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX}$$

to get $\mathbf{a}$ and the eigenvectors of

$$S_{YY}^{-1} S_{XY} S_{XX}^{-1} S_{YX}$$

give $\mathbf{b}$. The eigenvalues are the correlations.

Solution

Find the eigenvectors of

$$S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX}$$

to get $\mathbf{a}$ and the eigenvectors of

$$S_{YY}^{-1} S_{XY} S_{XX}^{-1} S_{YX}$$

give $\mathbf{b}$. The eigenvalues are the correlations.

Solution

Var-Cov of X

Find the eigenvectors of

$$S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX}$$

to get $\mathbf{a}$ and the eigenvectors of

$$S_{YY}^{-1} S_{XY} S_{XX}^{-1} S_{YX}$$

give $\mathbf{b}$. The eigenvalues are the correlations.

Solution

Var-Cov of X

Find the eigenvectors of

$$S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX}$$

to get $\mathbf{a}$ and the eigenvectors of

$$S_{YY}^{-1} S_{XY} S_{XX}^{-1} S_{YX}$$

give $\mathbf{b}$. The eigenvalues are the correlations.

Solution

Var-Cov of X

Find the eigenvectors of

$$S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX}$$

to get $\mathbf{a}$ and the eigenvectors of

$$S_{YY}^{-1} S_{XY} S_{XX}^{-1} S_{YX}$$

give $\mathbf{b}$. The eigenvalues are the correlations.

Var-Cov of Y

Solution

Var-Cov of X

Find the eigenvectors of

$$S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX}$$

to get $\mathbf{a}$ and the eigenvectors of

$$S_{YY}^{-1} S_{XY} S_{XX}^{-1} S_{YX}$$

give $\mathbf{b}$. The eigenvalues are the correlations.

Var-Cov of Y

Solution

Var-Cov of X

Find the eigenvectors of

$$S_{XX}^{-1} S_{XY} S_{YY}^{-1} S_{YX}$$

Covariance between X and Y

to get $\mathbf{a}$ and the eigenvectors of

$$S_{YY}^{-1} S_{XY} S_{XX}^{-1} S_{YX}$$

give $\mathbf{b}$. The eigenvalues are the correlations.

Var-Cov of Y

Example

- n=50, p=2, q=3

	pop15	pop75	sr	dpi	ddpi
Australia	29.35	2.87	11.43	2329.68	2.87
Austria	23.32	4.41	12.07	1507.99	3.93
Belgium	23.80	4.43	13.17	2108.47	3.82
Bolivia	41.89	1.67	5.75	189.13	0.22
Brazil	42.19	0.83	12.88	728.47	4.56
Canada	31.72	2.85	8.79	2982.88	2.43

Example

- n=50, p=2, q=3

	pop15	pop75	sr	dpi	ddpi
Australia	29.35	2.87	11.43	2329.68	2.87
Austria	23.32	4.41	12.07	1507.99	3.93
Belgium	23.80	4.43	13.17	2108.47	3.82
Bolivia	41.89	1.67	5.75	189.13	0.22
Brazil	42.19	0.83	12.88	728.47	4.56
Canada	31.72	2.85	8.79	2982.88	2.43

Example

- n=50, p=2, q=3

	pop15	pop75	sr	dpi	ddpi
Australia	29.35	2.87	11.43	2329.68	2.87
Austria	23.32	4.41	12.07	1507.99	3.93
Belgium	23.80	4.43	13.17	2108.47	3.82
Bolivia	41.89	1.67	5.75	189.13	0.22
Brazil	42.19	0.83	12.88	728.47	4.56
Canada	31.72	2.85	8.79	2982.88	2.43

Pct of pop under 15 older than 75 respectively

Example

- n=50, p=2, q=3

	pop15	pop75	sr	dpi	ddpi
Australia	29.35	2.87	11.43	2329.68	2.87
Austria	23.32	4.41	12.07	1507.99	3.93
Belgium	23.80	4.43	13.17	2108.47	3.82
Bolivia	41.89	1.67	5.75	189.13	0.22
Brazil	42.19	0.83	12.88	728.47	4.56
Canada	31.72	2.85	8.79	2982.88	2.43

Pct of pop under 15 older than 75 respectively

Example

- n=50, p=2, q=3

	pop15	pop75	sr	dpi	ddpi
Australia	29.35	2.87	11.43	2329.68	2.87
Austria	23.32	4.41	12.07	1507.99	3.93
Belgium	23.80	4.43	13.17	2108.47	3.82
Bolivia	41.89	1.67	5.75	189.13	0.22
Brazil	42.19	0.83	12.88	728.47	4.56
Canada	31.72	2.85	8.79	2982.88	2.43

Pct of pop under 15 older than 75 respectively

Savings ratio, per capita disposable income, % growth in dpi

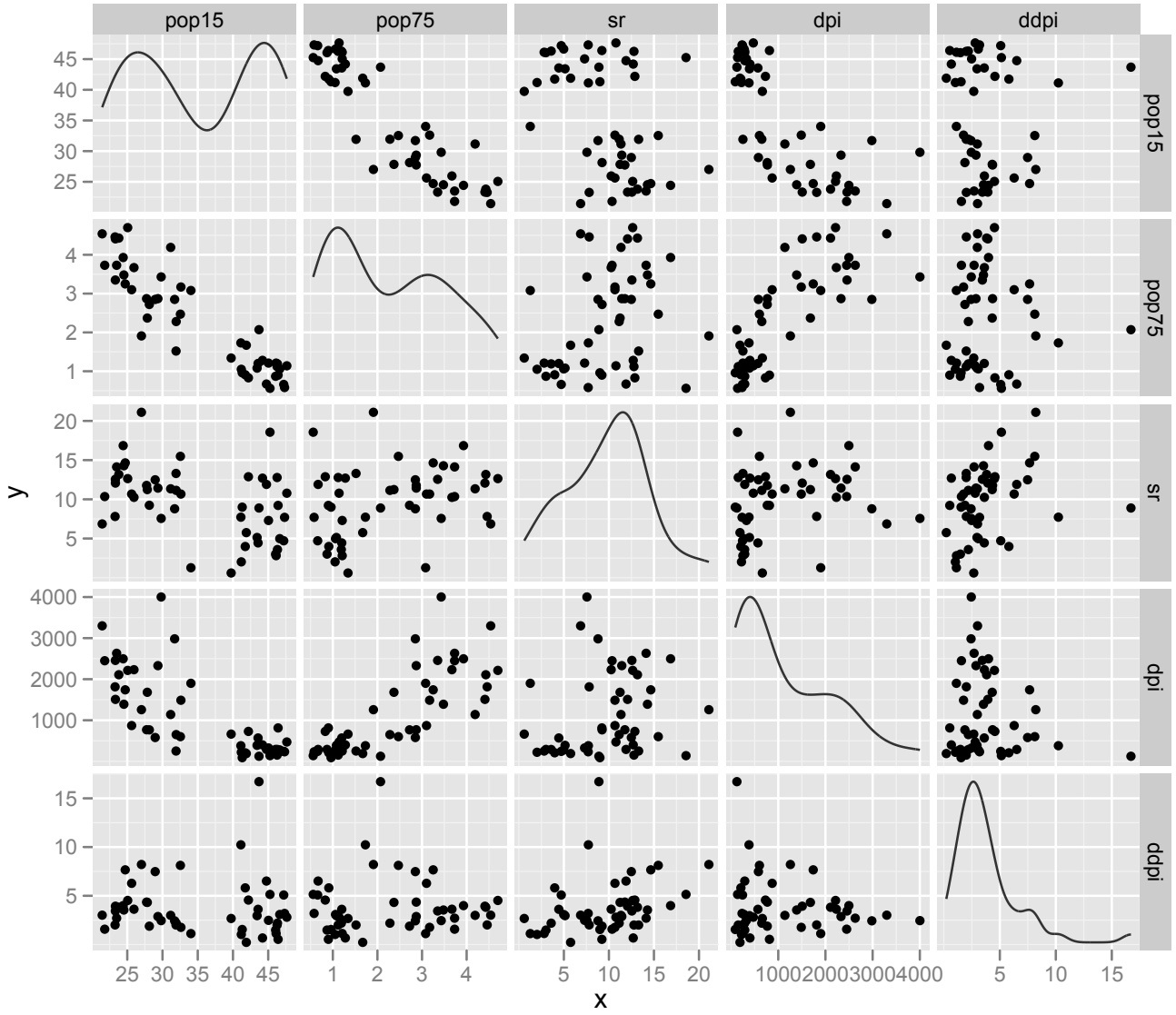
Summary Statistics

$$\bar{\mathbf{X}} = \begin{bmatrix} 35.1 \\ 2.29 \end{bmatrix} \quad \bar{\mathbf{Y}} = \begin{bmatrix} 9.67 \\ 1106.8 \\ 3.76 \end{bmatrix}$$

$$\mathbf{S}_{XX} = \begin{bmatrix} 1.00 & -0.91 \\ -0.91 & 1.00 \end{bmatrix} \quad \mathbf{S}_{YY} = \begin{bmatrix} 1.00 & 0.22 & 0.30 \\ 0.22 & 1.00 & -0.13 \\ 0.30 & -0.13 & 1.00 \end{bmatrix}$$

$$\mathbf{S}_{XY} = \left[\begin{array}{c|ccc} & sr & dpi & ddpi \\ \hline pop15 & -0.456 & -0.756 & -0.048 \\ pop75 & 0.317 & 0.787 & 0.025 \end{array}\right]$$

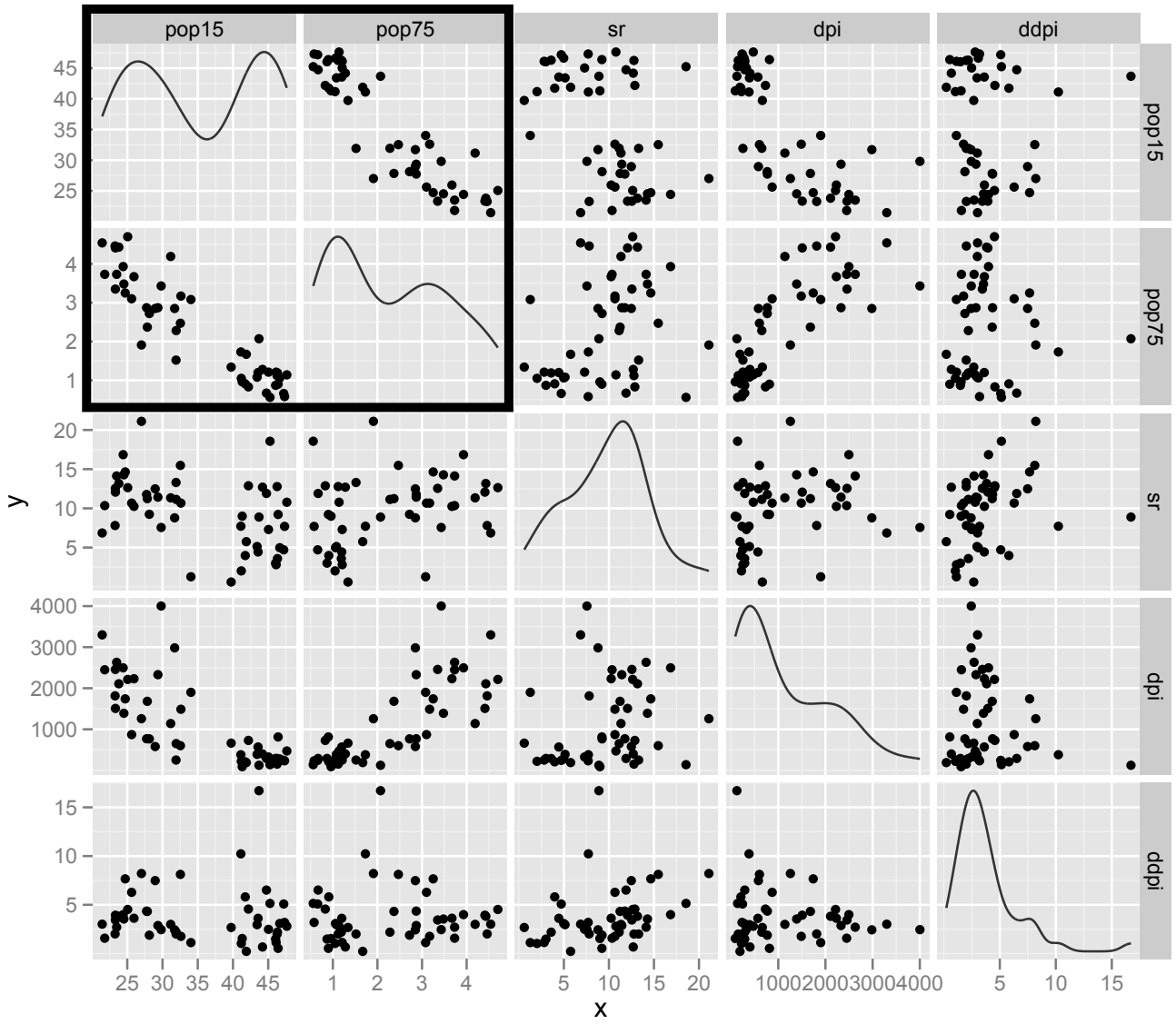
Example



Raw data: two groups.

Association between the groups

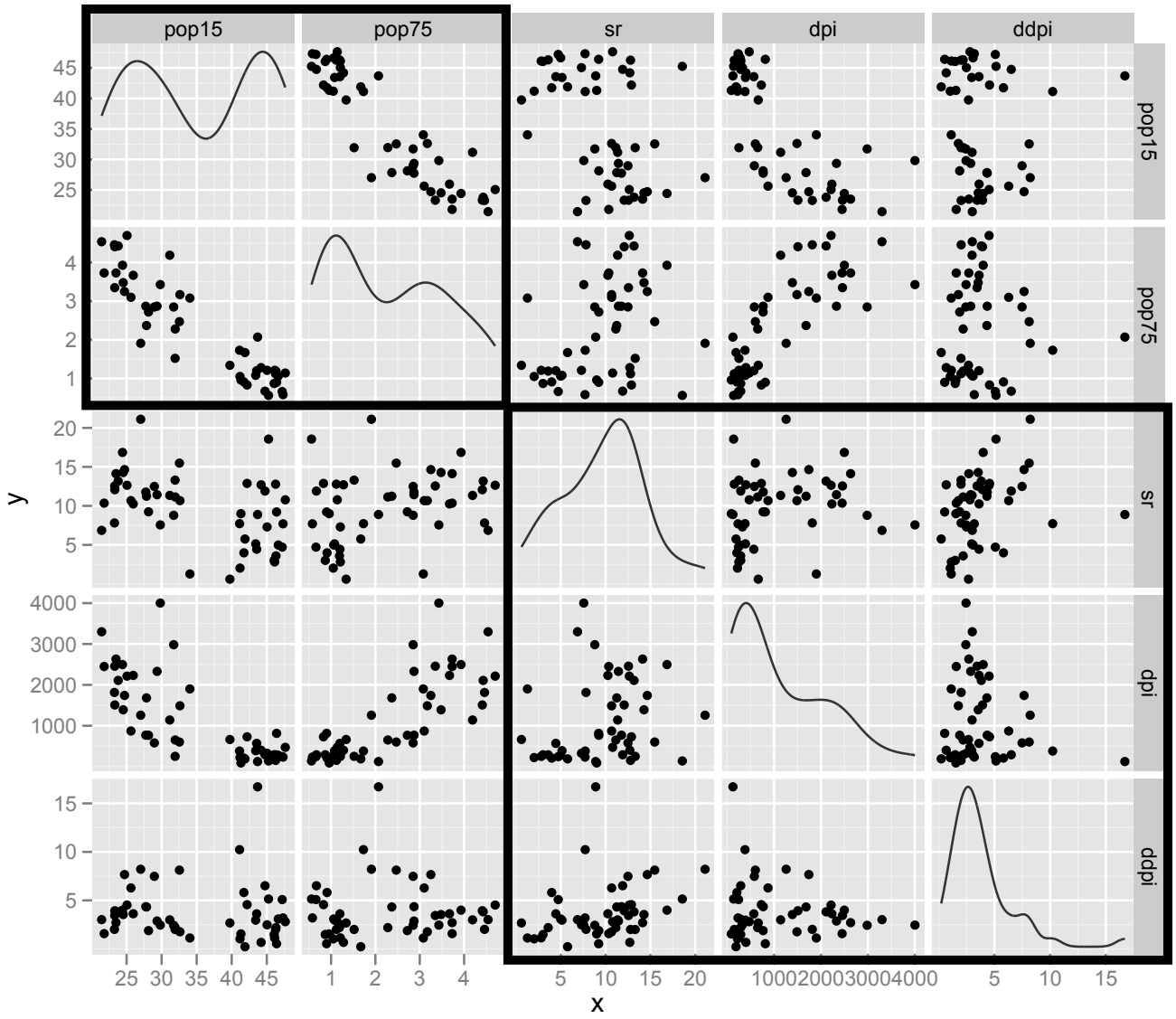
Example



Raw data: two groups.

Association between the groups

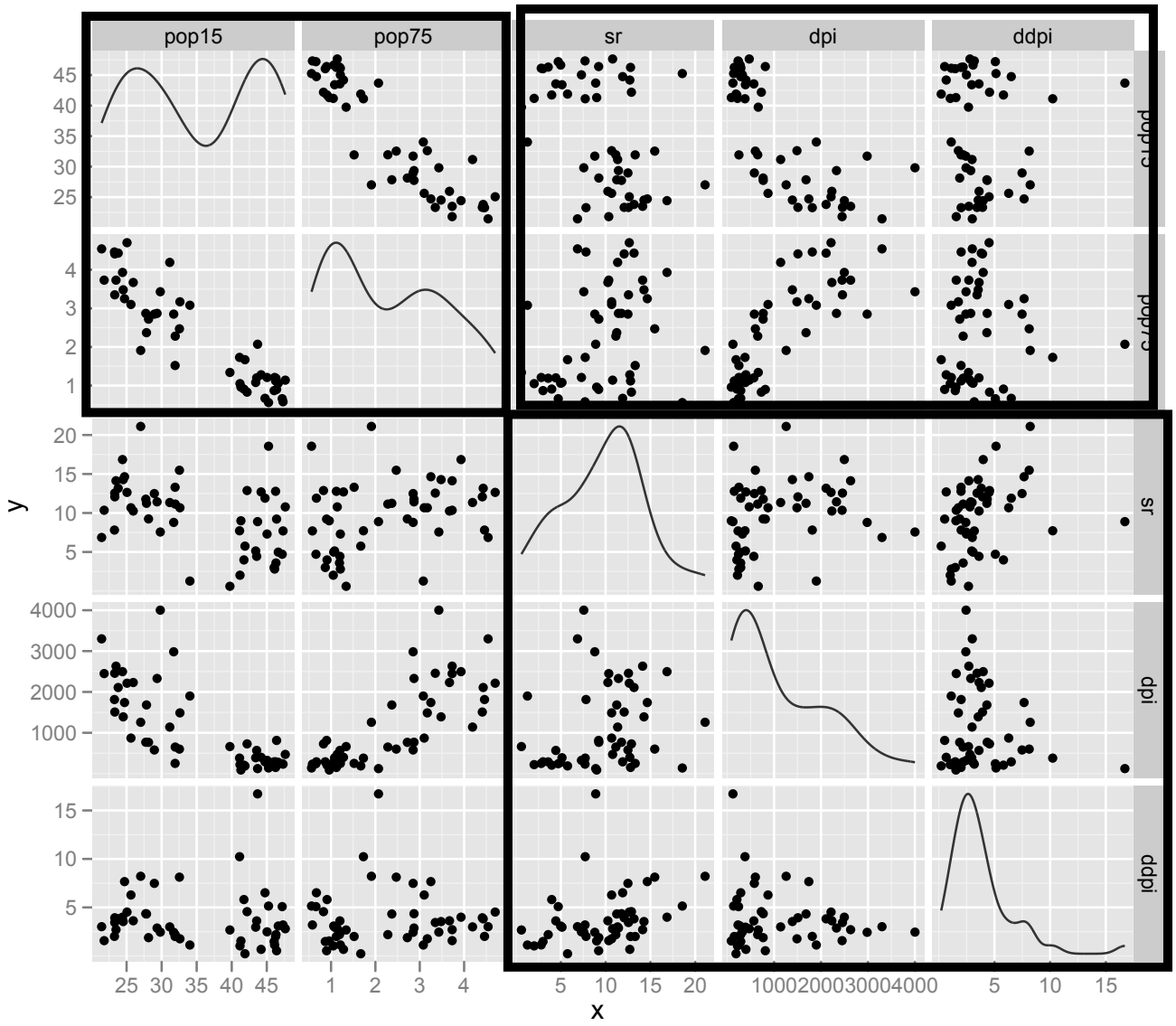
Example



Raw data: two groups.

Association between the groups

Example

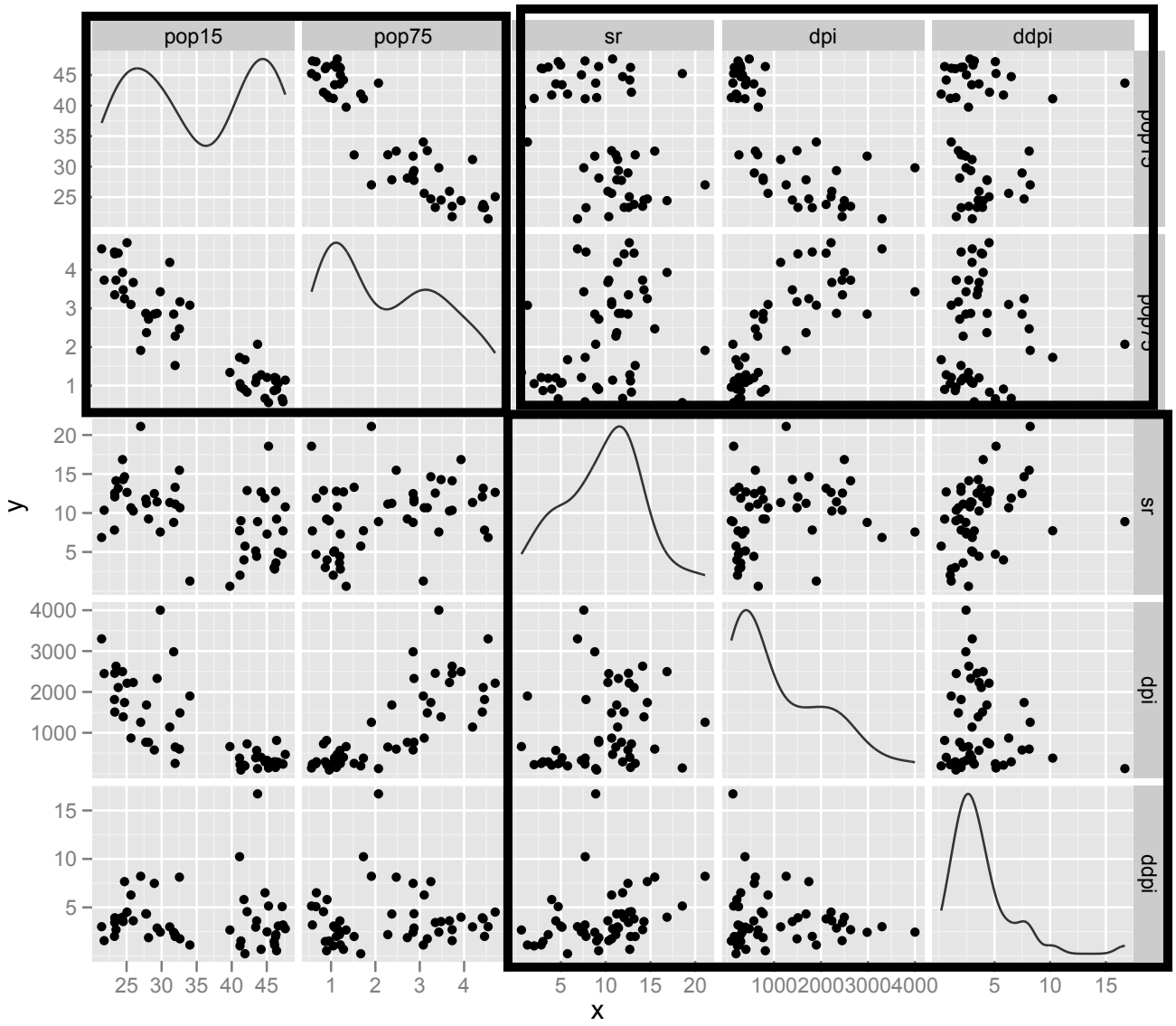


Raw data: two groups.

Association between the groups

Example

Var-Cov of X

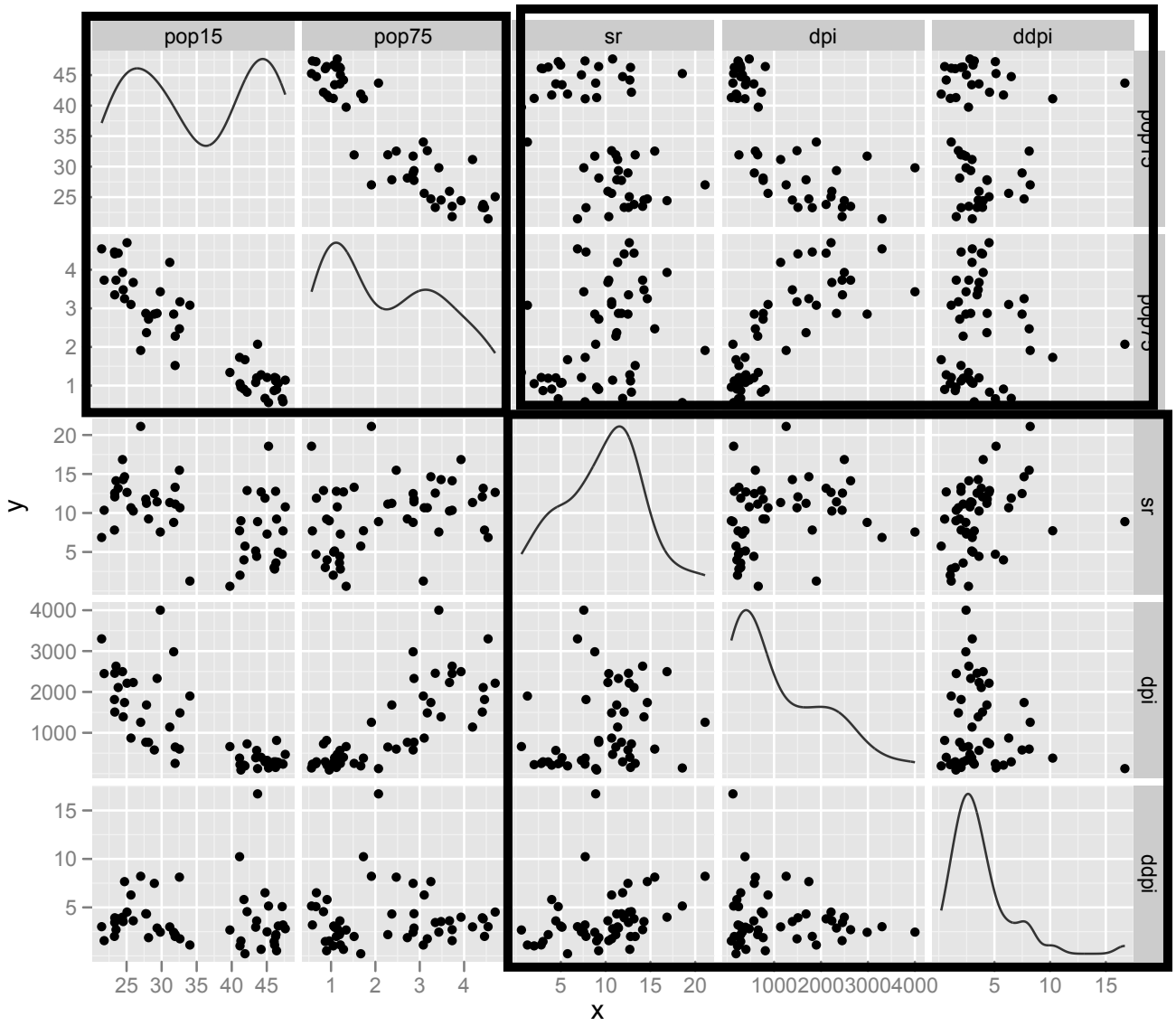


Raw data: two groups.

Association between the groups

Example

Var-Cov of X



Raw data: two groups.

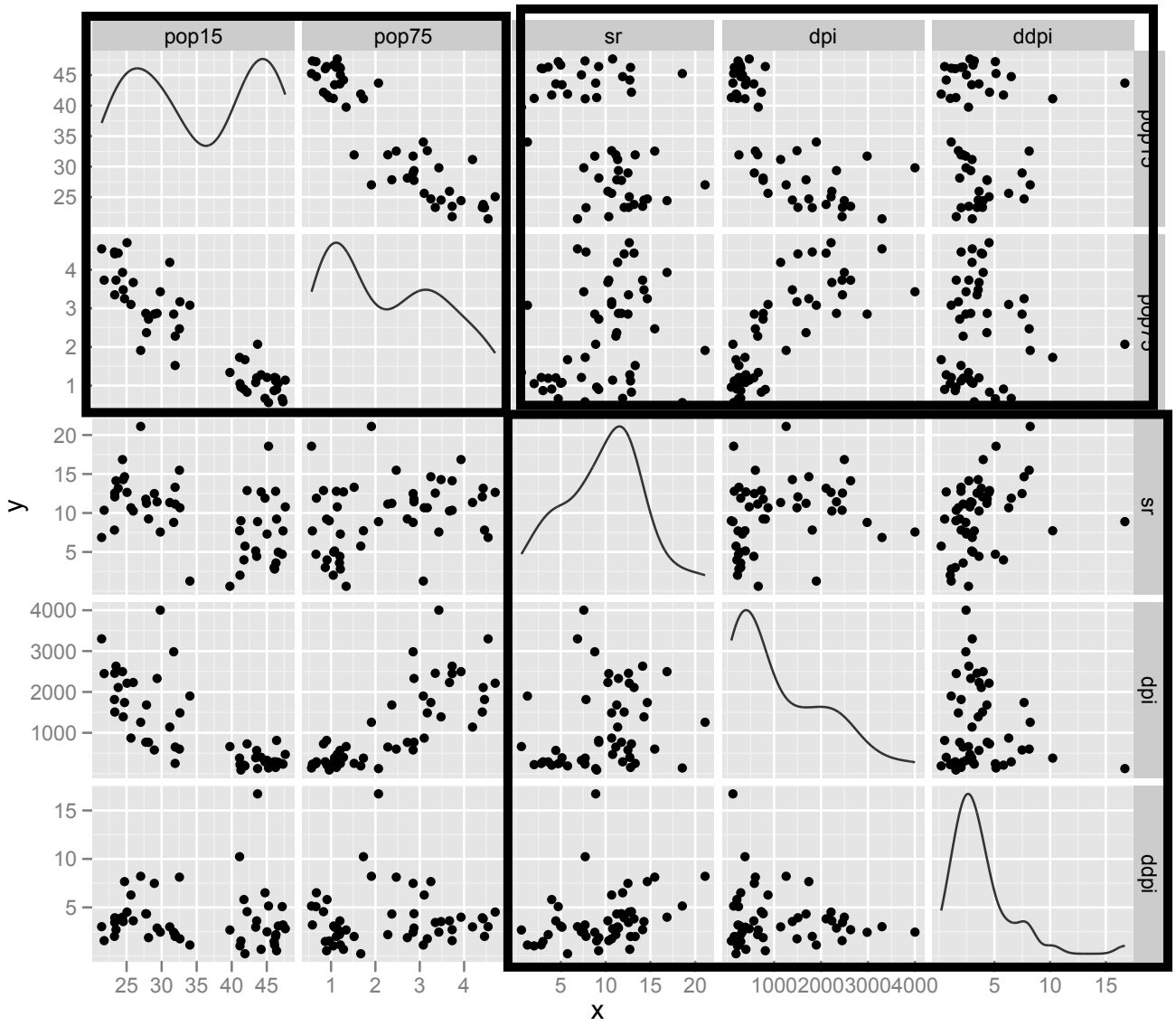
Association between the groups

Var-Cov of Y

Example

Var-Cov of X

Covariance between X, Y



Raw data: two groups.

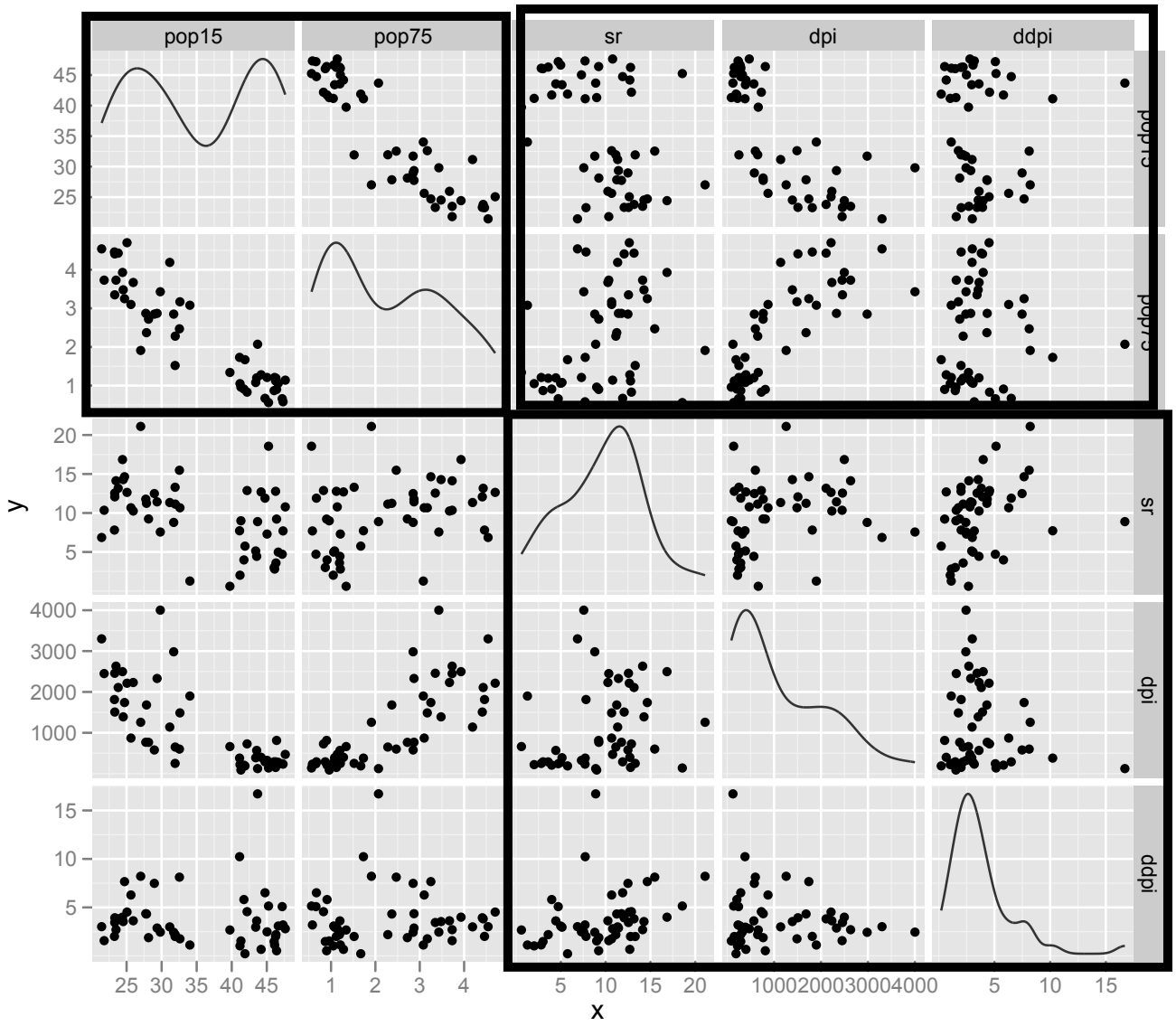
Association between the groups

Var-Cov of Y

Example

Var-Cov of X

Covariance between X, Y



Raw data: two groups.

Association between the groups

Data has some problems, outliers, nonlinear dependence.

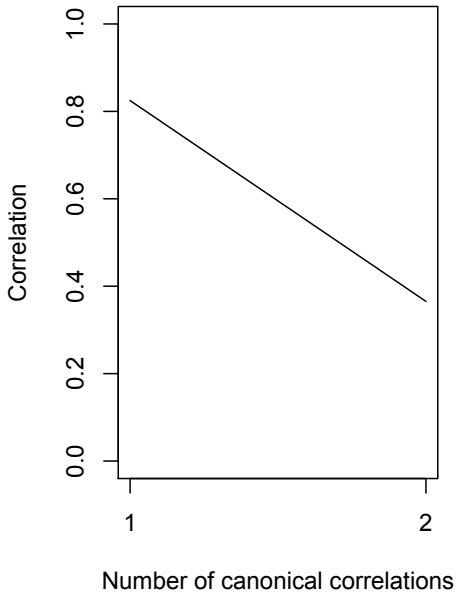
Var-Cov of Y

Canonical correlation

$$r_1 = 0.825, r_2 = 0.365$$

$$\mathbf{A} = \begin{bmatrix} -0.00911 & -0.0362 \\ 0.0486 & -0.260 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0.00847 & 0.0333 & -0.00516 \\ 0.000131 & -0.0000759 & 0.00000454 \\ 0.00417 & -0.00123 & 0.00519 \end{bmatrix}$$



Can't interpret coefficients because the variables had different scales

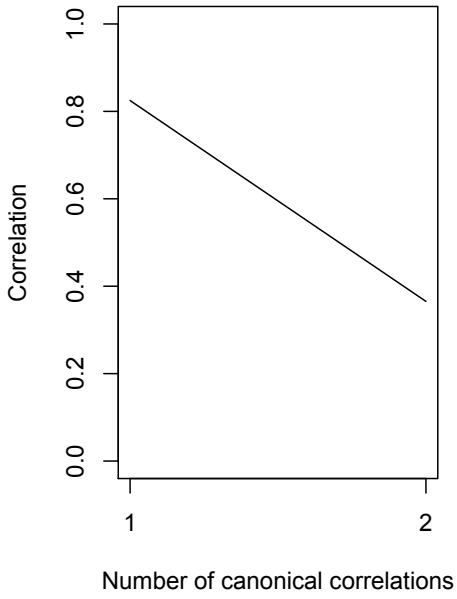
Canonical correlation

Results for standardized variables

$$r_1 = 0.825, r_2 = 0.365$$

$$\mathbf{A} = \begin{bmatrix} -0.083 & -0.331 \\ 0.062 & -0.336 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0.038 & 0.150 & -0.023 \\ 0.130 & -0.075 & 0.0045 \\ 0.012 & -0.035 & 0.149 \end{bmatrix}$$



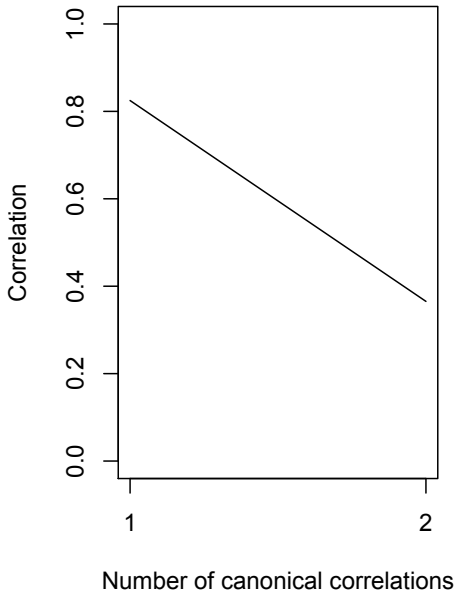
Canonical correlation

Results for standardized variables

$$r_1 = 0.825, r_2 = 0.365$$

$$\mathbf{A} = \begin{bmatrix} -0.083 & -0.331 \\ 0.062 & -0.336 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0.038 & 0.150 & -0.023 \\ 0.130 & -0.075 & 0.0045 \\ 0.012 & -0.035 & 0.149 \end{bmatrix}$$



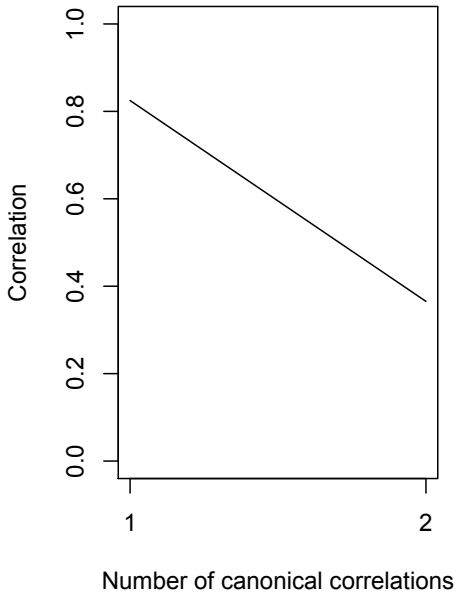
Canonical correlation

Results for standardized variables

$$r_1 = 0.825, r_2 = 0.365$$

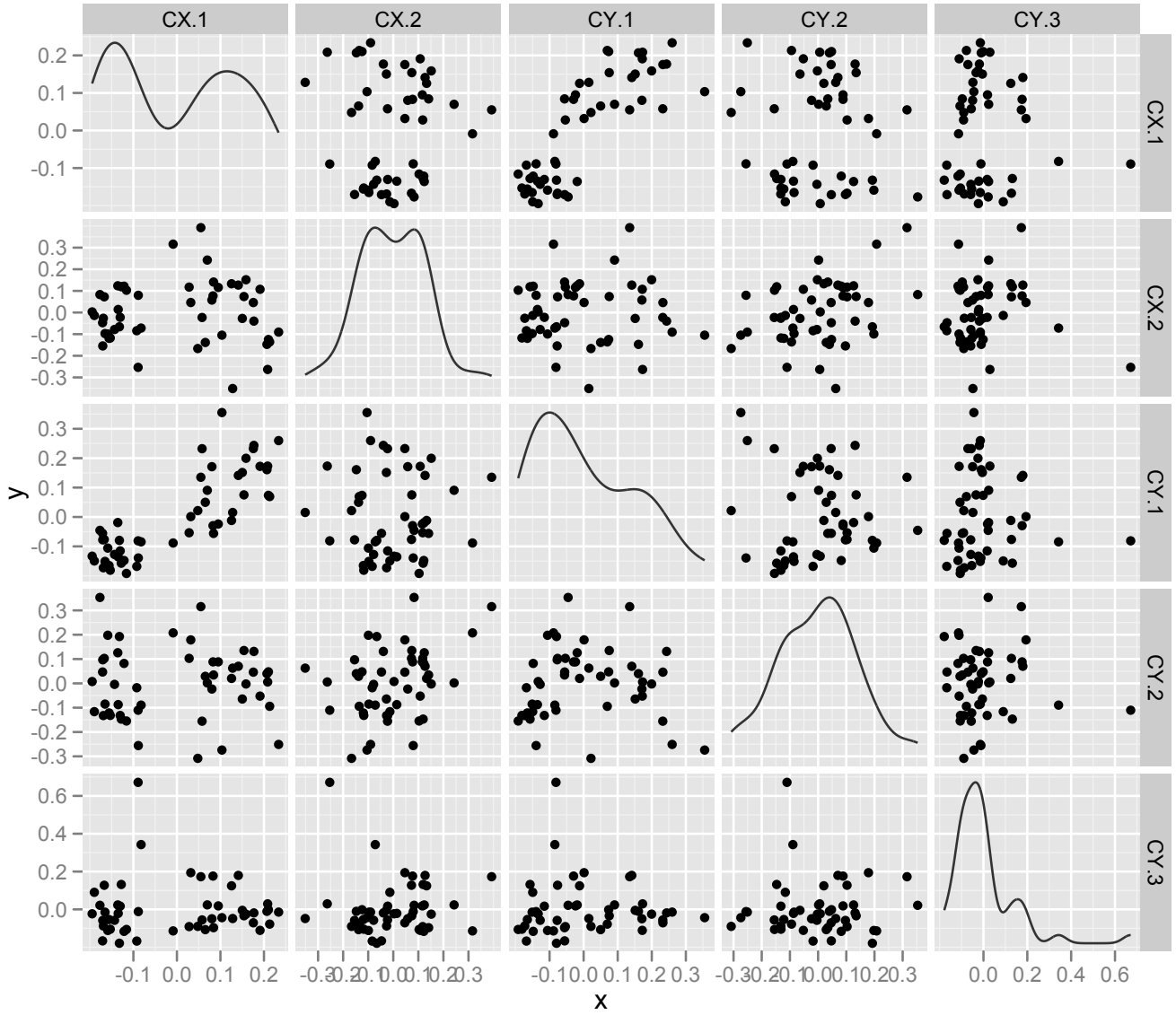
$$\mathbf{A} = \begin{bmatrix} -0.083 & -0.331 \\ 0.062 & -0.336 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0.038 & 0.150 & -0.023 \\ 0.130 & -0.075 & 0.0045 \\ 0.012 & -0.035 & 0.149 \end{bmatrix}$$



dpi most important variable in explaining contrast between age proportions

Example

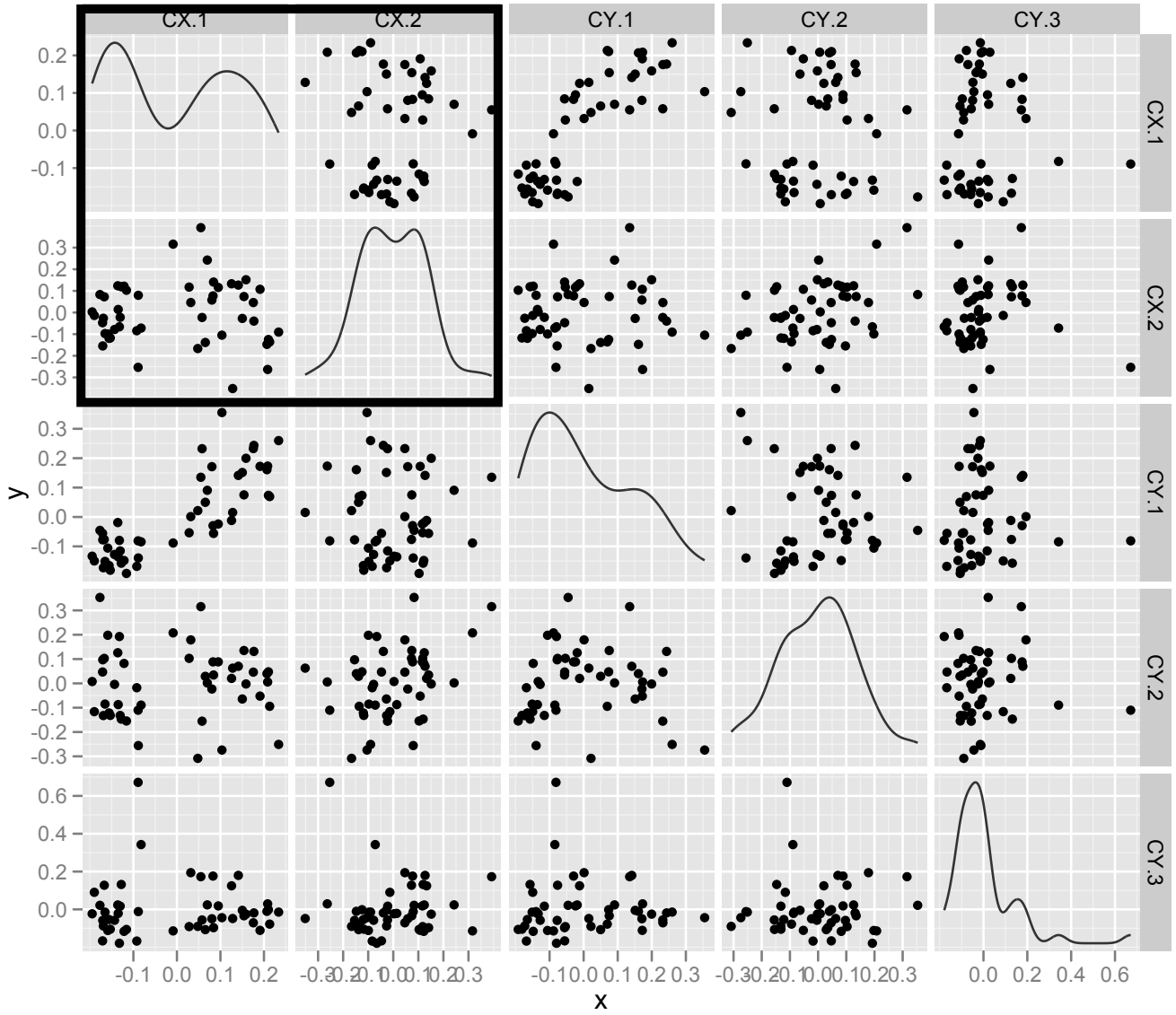


Canonical correlations.

First correlation stronger than individual vars, just.

No association between canonical variables.

Example

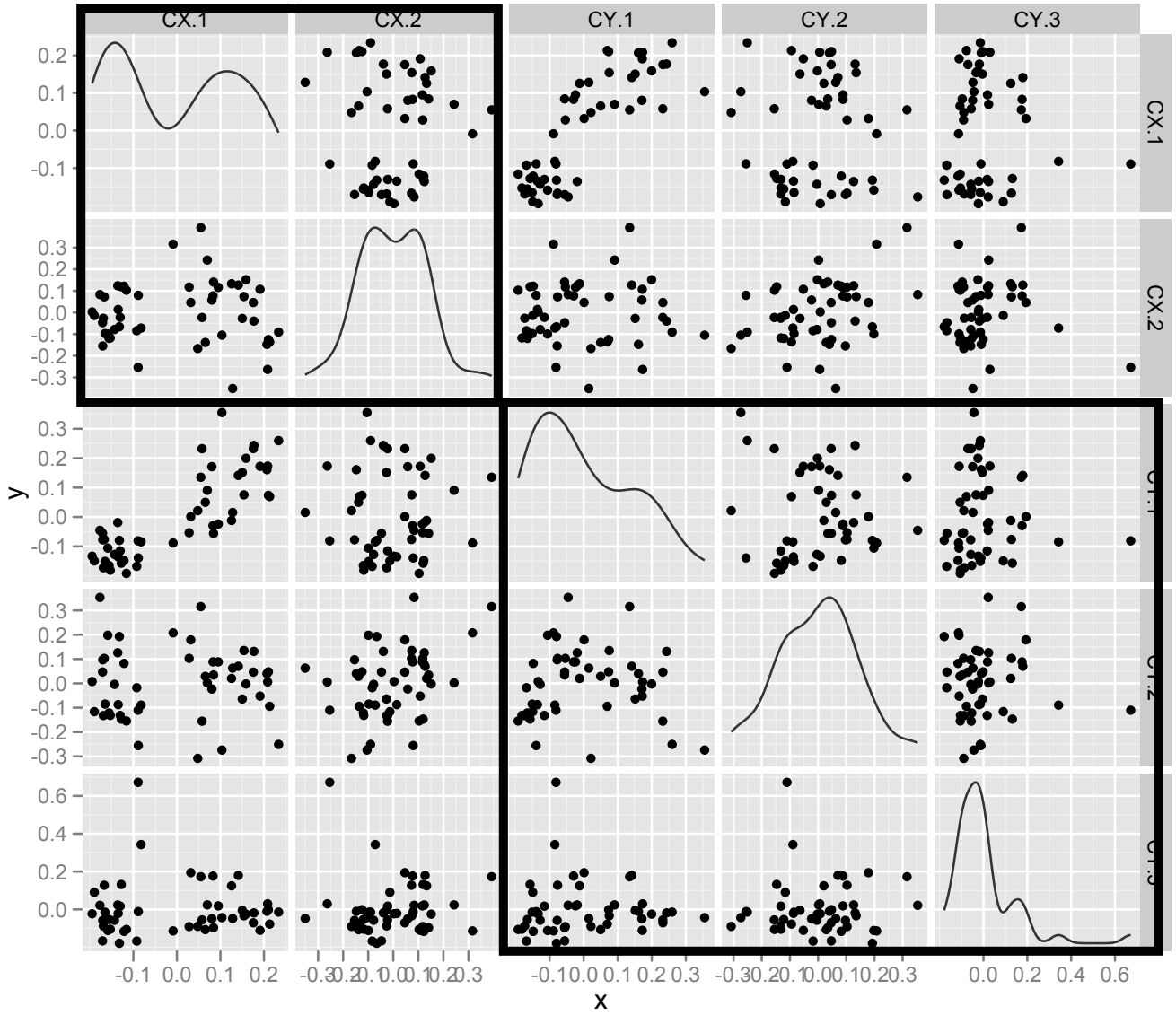


Canonical correlations.

First correlation stronger than individual vars, just.

No association between canonical variables.

Example

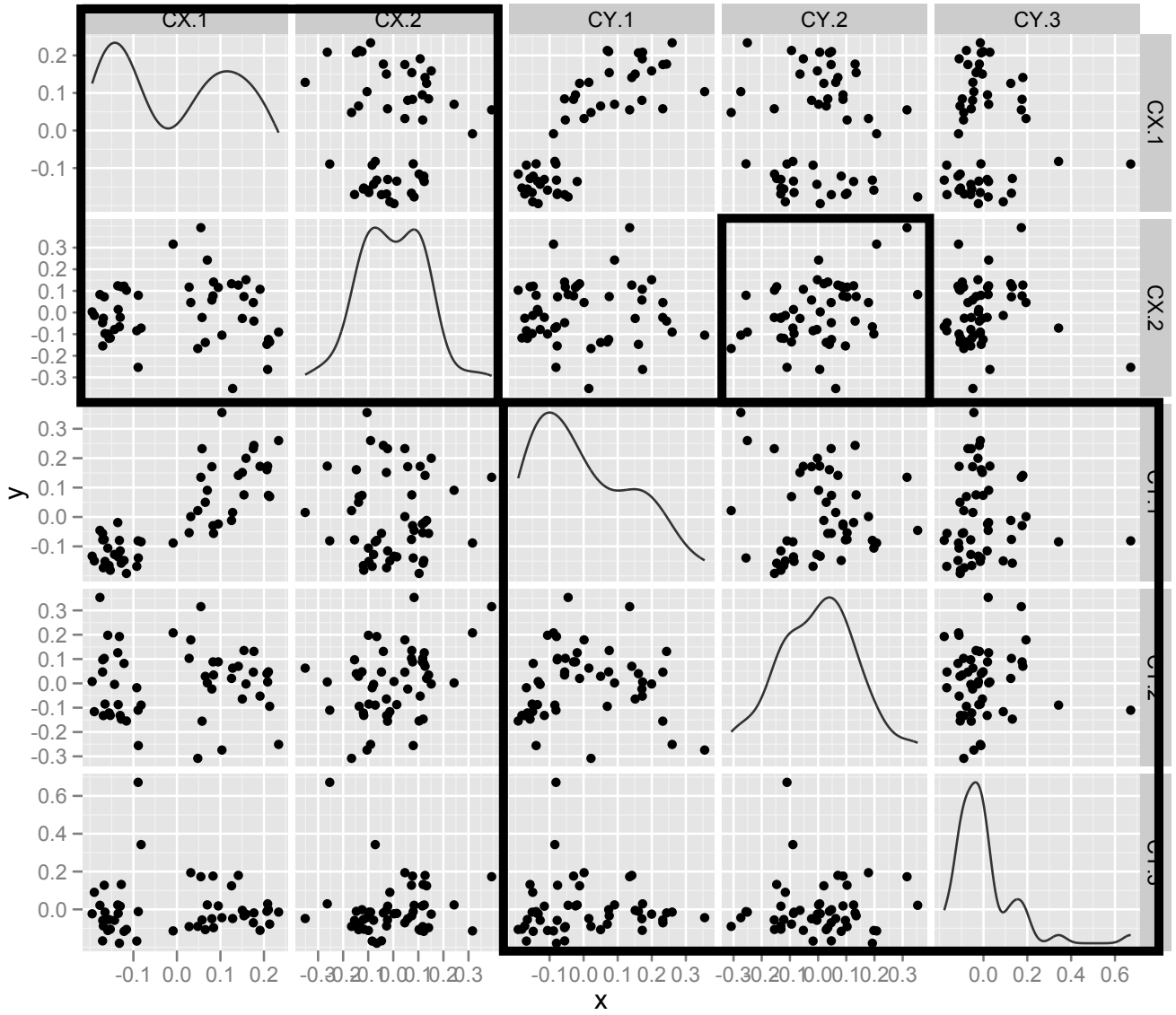


Canonical correlations.

First correlation stronger than individual vars, just.

No association between canonical variables.

Example

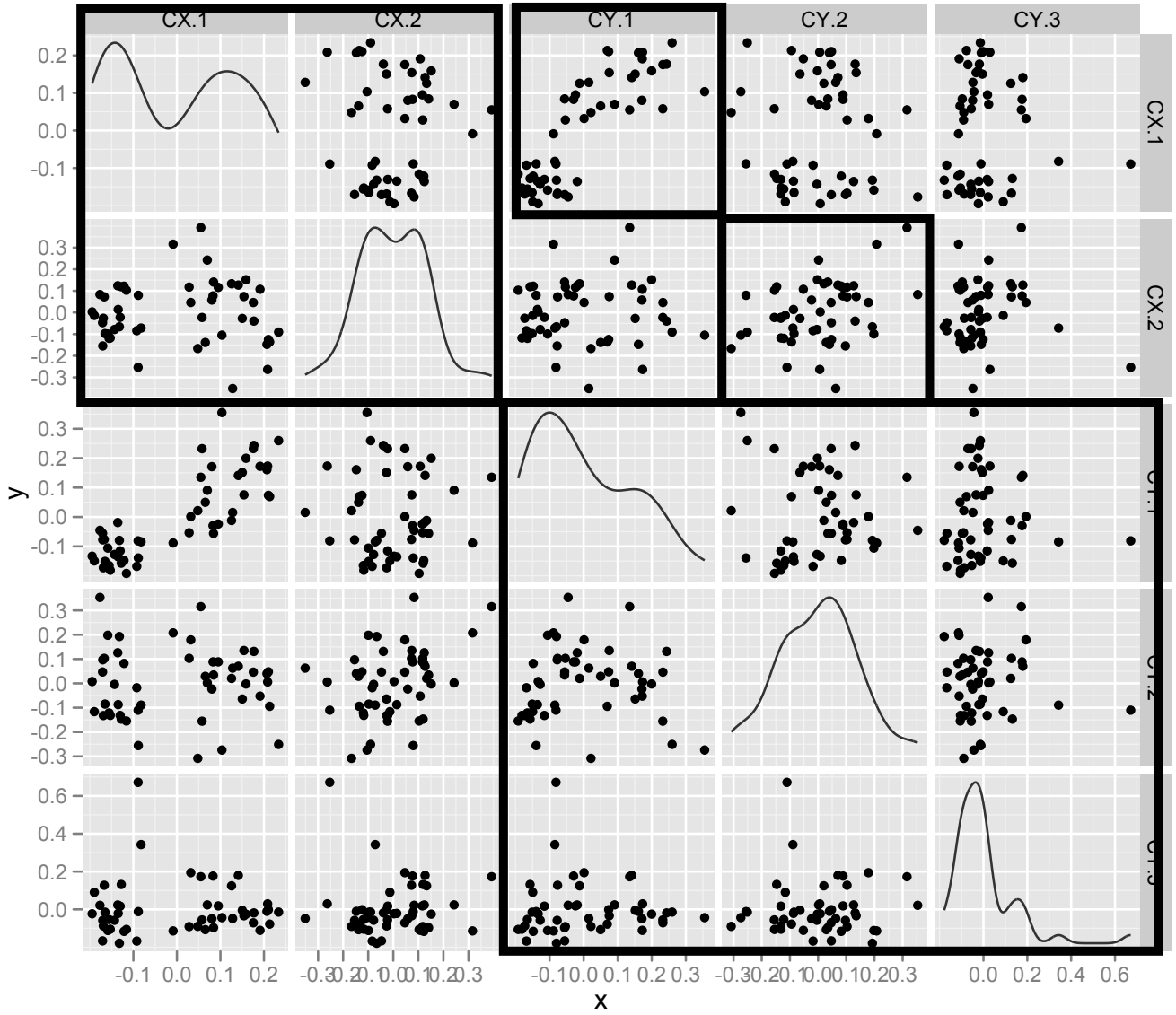


Canonical correlations.

First correlation stronger than individual vars, just.

No association between canonical variables.

Example



Canonical correlations.

First correlation stronger than individual vars, just.

No association between canonical variables.

Interpretation

- Highest correlation between dpi (with a little sr) and a contrast of age proportions.
- Low value on dpi (and sr) corresponds to high proportion of population under 15, and a low number over 75.