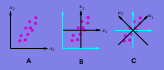
# (Simple) Correspondence Analysis (CA)

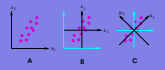
Hervé Abdi

www.utdallas.edu/∼herve herve@utdallas.edu



# 1. Refresher! and Menu

- Refresher: Weight and Masses
- $\chi^2$ "Chi-squared"
- Independence
- Start with a small example (punctuation)
- New: What is a profile?
- Masses and weights for CA
- Centroïd, barycenter, etc
- Distributional equivalence (big property)
- Row profile analysis
- What to do with the columns try 1
- What to do with the columns try 2
- Duality: (rows to columns) = (column to rows)



## 2. Example: Punctuation . . .

Data from Brunet (1989):
The punctuation marks of six great French writers, Rousseau, Chateaubriand, Hugo, Zola, Proust, and Giraudoux.

Data:
Number of times each writer uses three punctuation marks: the period, the comma, and all the other marks (*i.e.*, interrogation mark, exclamation mark, colon, and semicolon).

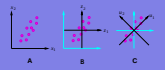
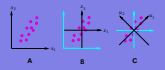


Table 1: The punctuation marks of six French writers (from Brunet, 1989).

| | Period | Comma | All the other marks |
|---|---|---|---|
| Rousseau | 7836 | 13112 | 6026 |
| Chateaubriand | 53655 | 102383 | 42413 |
| Hugo | 115615 | 184541 | 59226 |
| Zola | 161926 | 340479 | 62754 |
| Proust | 38177 | 105101 | 12670 |
| Giraudoux | 46371 | 58367 | 14299 |

**Table with total and frequencies**

| Author's name | Period | Comma | Other | $x_{i+}$ | $P(\text{Author}) \dfrac{x_{i+}}{x_{++}}$ |
|---|---|---|---|---|---|
| Rousseau | 7836 | 13112 | 6026 | **26974** | **.0189** |
| Chateaubriand | 53655 | 102383 | 42413 | **198451** | **.1393** |
| Hugo | 115615 | 184541 | 59226 | **359382** | **.2522** |
| Zola | 161926 | 340479 | 62754 | **565159** | **.3966** |
| Proust | 38177 | 105101 | 12670 | **155948** | **.1094** |
| Giraudoux | 46371 | 58367 | 14299 | **119037** | **.0835** |
| $x_{+j}$ | **423580** | **803983** | **197388** | **1424951** $x_{++}$ | 1.0000 |
| $P(\text{Mark}) \dfrac{x_{+j}}{x_{++}}$ | **.2973** | **.5642** | **.1385** | | |



## 3. $\chi^2$ Refresher: Independence Rows and Columns

- **Data:** A contingency table.
- **Question:** Do the writers differ in their punctuation?
- **Statistical answer:**
  test for independence Rows and Columns

### Chi-Square for Independence

The Data Table Under Independence:
Step one find the probabilities under independence.

The Product Rule: $A$ and $B$ independent events
$P(A\&B) = P(A) \times P(B)$.

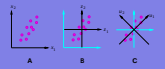
Probability of using a mark for an author:
Probability of the mark $\times$ Probability of the author

## Probability under Independence

| | Period | Comma | All the other marks |
|---|---|---|---|
| Rousseau | .0056 | .0107 | .0026 |
| Chateaubriand | .0414 | .0786 | .0193 |
| Hugo | .0750 | .1423 | .0349 |
| Zola | .1179 | .2238 | .0549 |
| Proust | .0325 | .0617 | .0152 |
| Giraudoux | .0248 | .0471 | .0116 |

## Real Probability

| | Period | Comma | All the other marks |
|---|---|---|---|
| Rousseau | .0055 | .0092 | .0042 |
| Chateaubriand | .0377 | .0719 | .0298 |
| Hugo | .0811 | .1295 | .0416 |
| Zola | .1136 | .2389 | .0440 |
| Proust | .0268 | .0738 | .0089 |
| Giraudoux | .0325 | .0410 | .0100 |



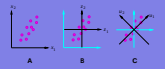
## Real Probability – Independence

| | Period | Comma | All the other marks |
|---|---|---|---|
| Rousseau | −.0001 | −.0015 | .0016 |
| Chateaubriand | −.0037 | −.0067 | .0105 |
| Hugo | .0062 | −.0128 | .0066 |
| Zola | −.0043 | .0152 | −.0109 |
| Proust | −.0057 | .0120 | −.0063 |
| Giraudoux | .0077 | −.0062 | −.0015 |

Square everything, sum and multiply by grand total:

$$\chi^2 = 1571.16$$

So it is awfully significant! (so what?)



# Official formula for $\chi^2$

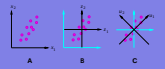
$$\chi^2 = \sum \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}$$

$$= \sum \frac{1}{E_{i,j}} (O_{i,j} - E_{i,j})^2$$

$$= \sum md^2$$

Remember a sum of square distances is called an *Inertia* (e.g. variance)

So: Chi-Square is an Inertia!



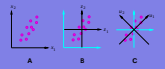
## 4. How to Analyze the Table

First idea: PCA right away!

Not a good idea!
To see why: Do it with the hidden alias of Zola: Aloz.

$$[2699 \quad 5675 \quad 1046]$$

(Just like Zola but wrote only a short novel)



# The PCA: Centered Data!


Figure 1: PCA analysis of the Punctuation. Centered Data.

# The PCA: Centered Data!


Figure 2: PCA analysis of the Punctuation. Centered Data. Aloz is a supplementary element.

# The PCA: Centered & Normalized Data!

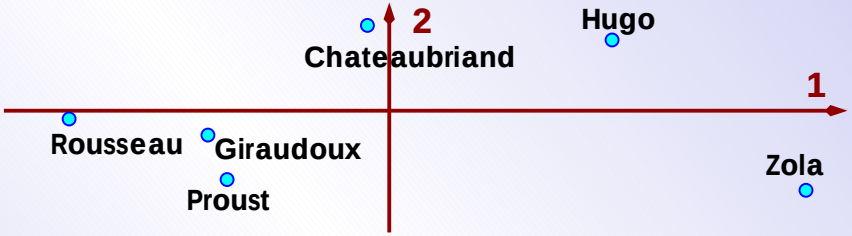



Figure 3: PCA analysis of the Punctuation. Centered & Normalized Data.

# The PCA: Centered & Normalized Data!

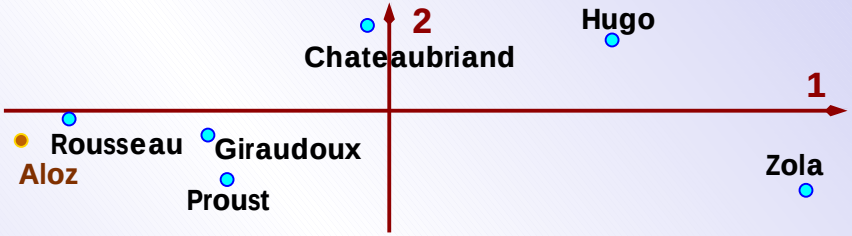
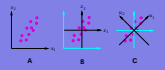



Figure 4: PCA analysis of the Punctuation. Centered & Normalized Data. Aloz is a supplementary element.

## The Profiles

| | Period | Comma | All the other marks |
|---|---|---|---|
| Rousseau | .2905 | .4861 | .2234 |
| Chateaubriand | .2704 | .5159 | .2137 |
| Hugo | .3217 | .5135 | .1648 |
| Zola | .2865 | .6024 | .1110 |
| Proust | .2448 | .6739 | .0812 |
| Giraudoux | .3896 | .4903 | .1201 |



# Plot them !

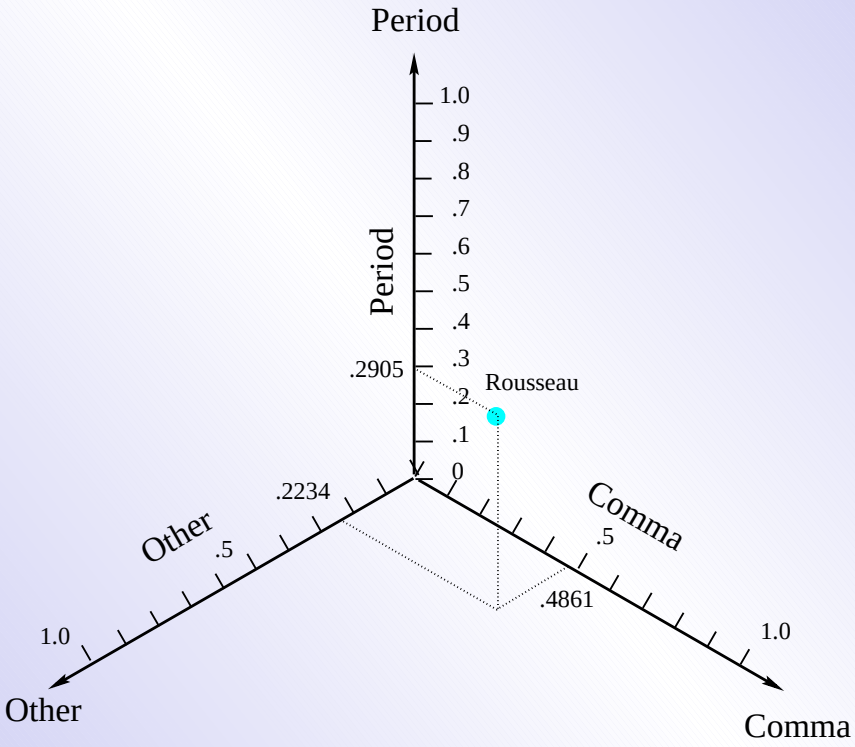



Figure 5: Plotting a Profile.

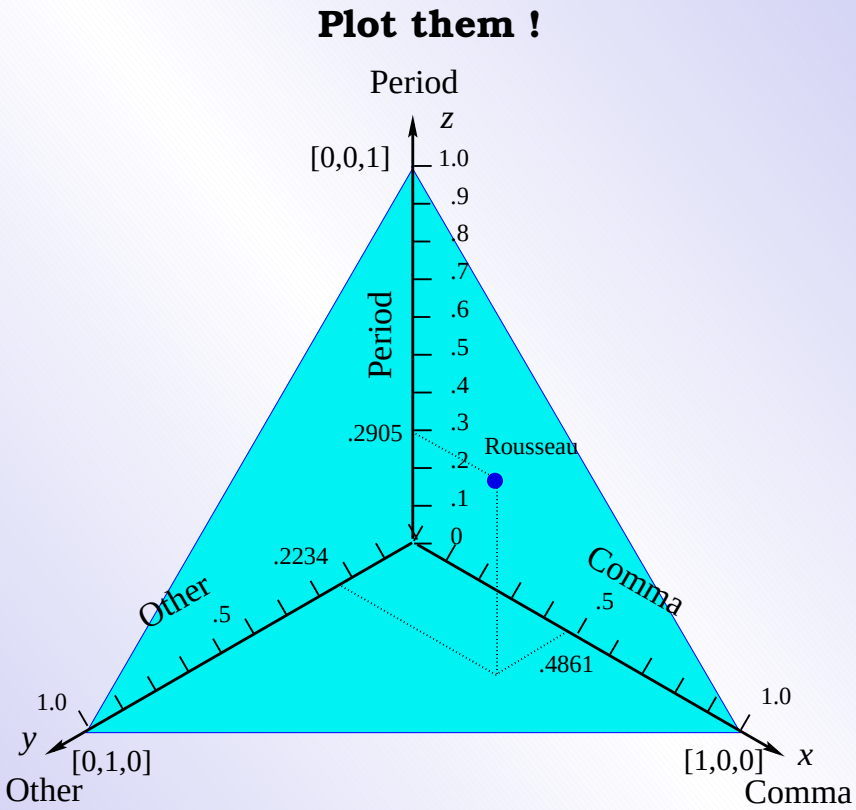



Figure 6: Plotting a Profile. The Simplex

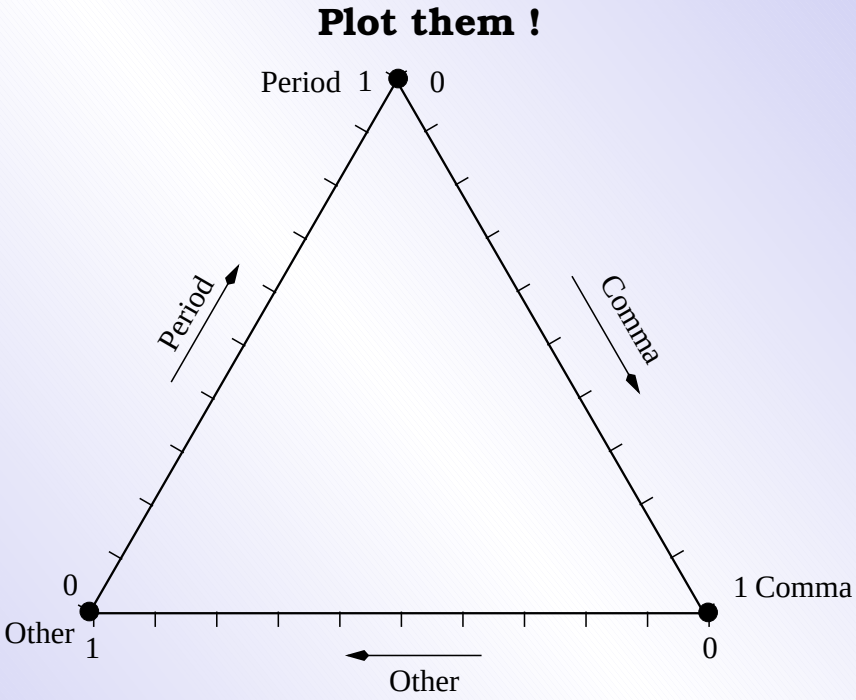



Figure 7: The Row Profile Space in $2D$.

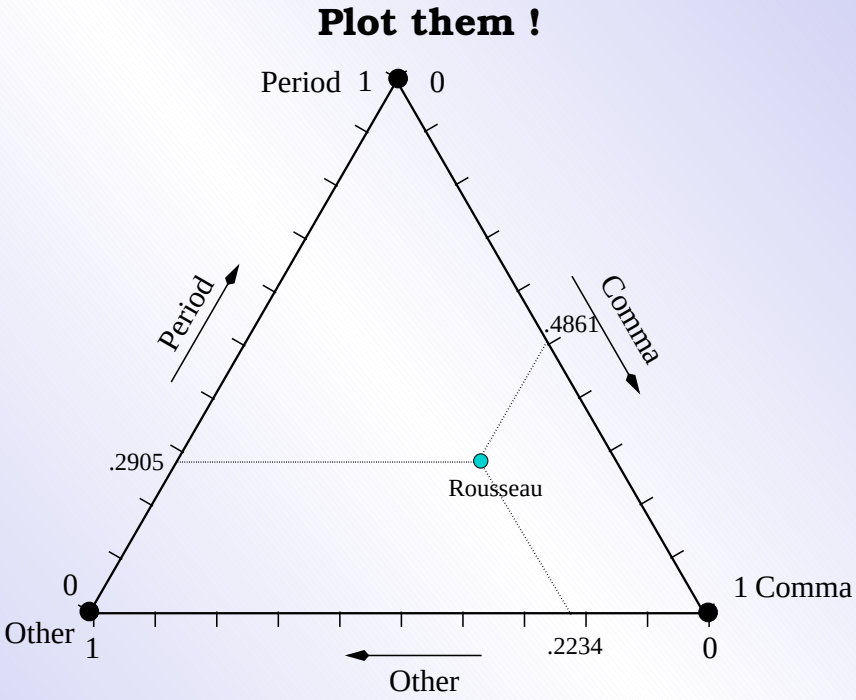



Figure 8: The Row Profile Space in $2D$. Plotting Rousseau

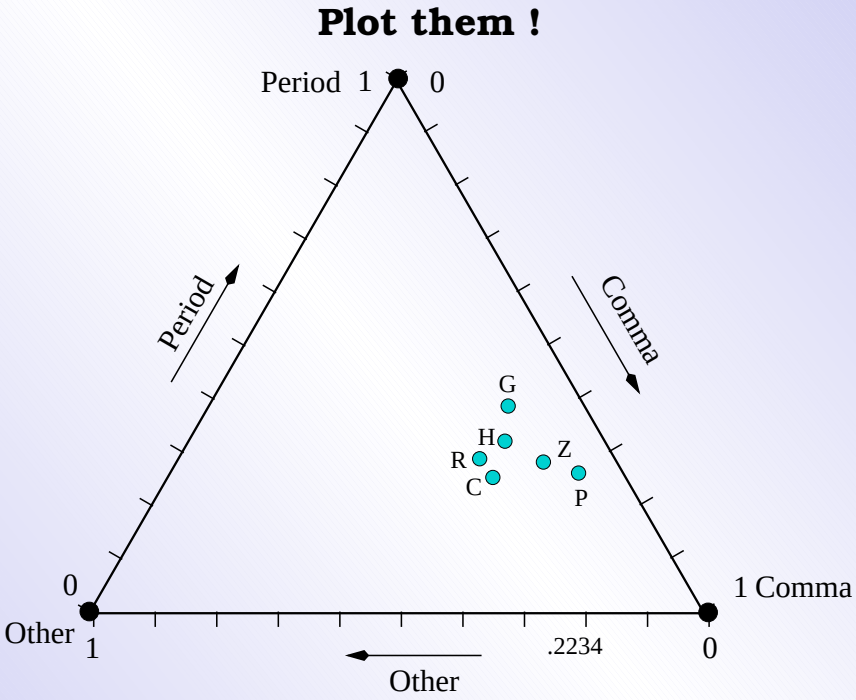



Figure 9: The Row Profile Space in $2D$. Plotting Them All

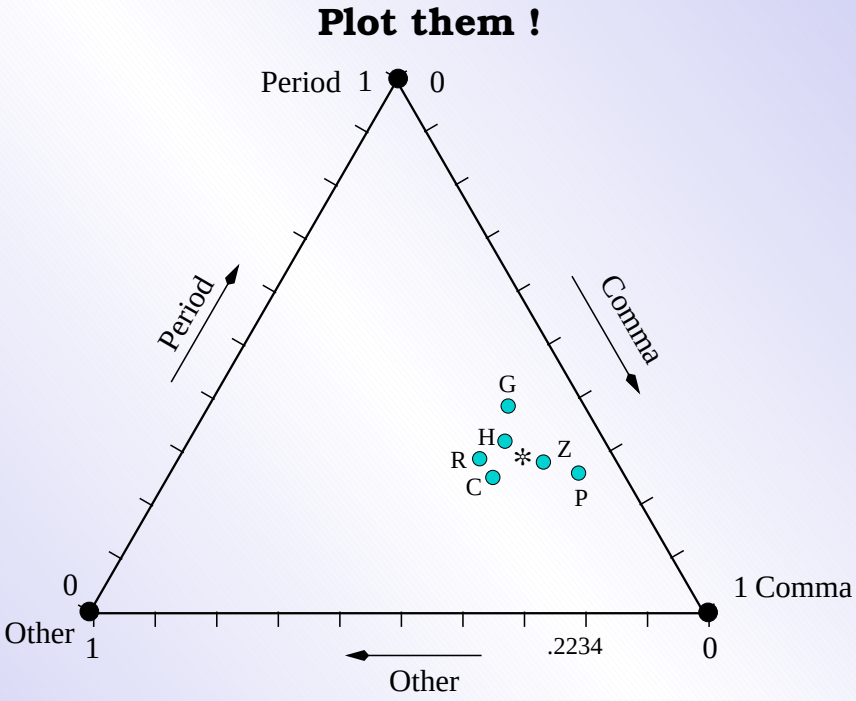



Figure 10: The Row Profile Space in $2D$. Plotting Them All. With Center of Gravity.

We want the space to express the *Information* of the variables

Information?

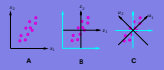
Rare variables provide a lot of information

Frequent variables provide little information

How to do that?

Stretch the space dimensions (variables).

with the *inverse* of the frequency



# Plot them !

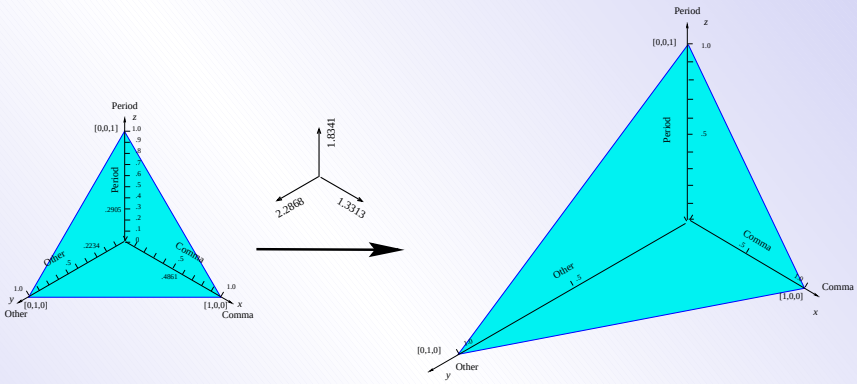



Figure 11: What the weights are doing to the row space.


Refresher! and Menu
Example: . . .
Chi-Square Refresher: . . .
How to Analyze the . . .
Home Page
Title Page
Page 24 of 85
Go Back
Full Screen
Close
Quit

# Plot them !

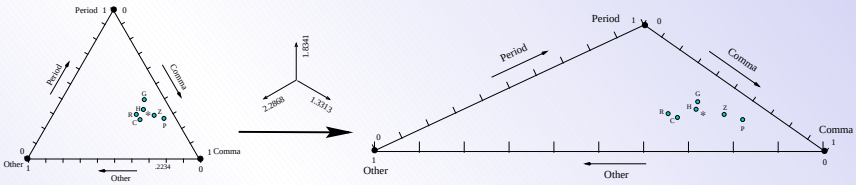



Figure 12: What the weights are doing to the row space.

# Plot them !

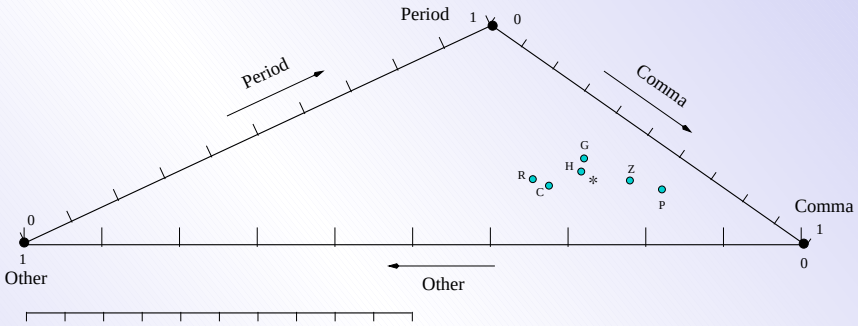



Figure 13: What the weights are doing to the row space. A closer view.

# Plot them !

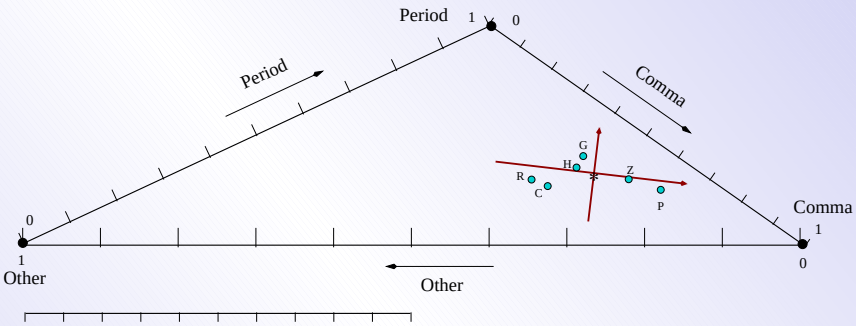



Figure 14: Zooming in.


Refresher! and Menu
Example: . . .
Chi-Square Refresher: . . .
How to Analyze the . . .
Home Page
Title Page
Page 27 of 85
Go Back
Full Screen
Close
Quit

# Plot them !

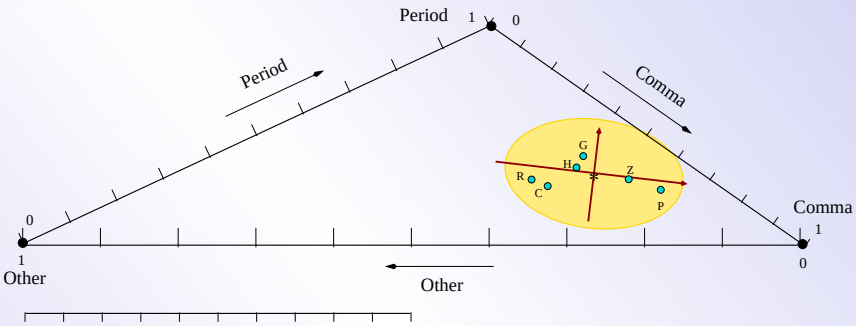



Figure 15: Zooming in.

## Plot them !

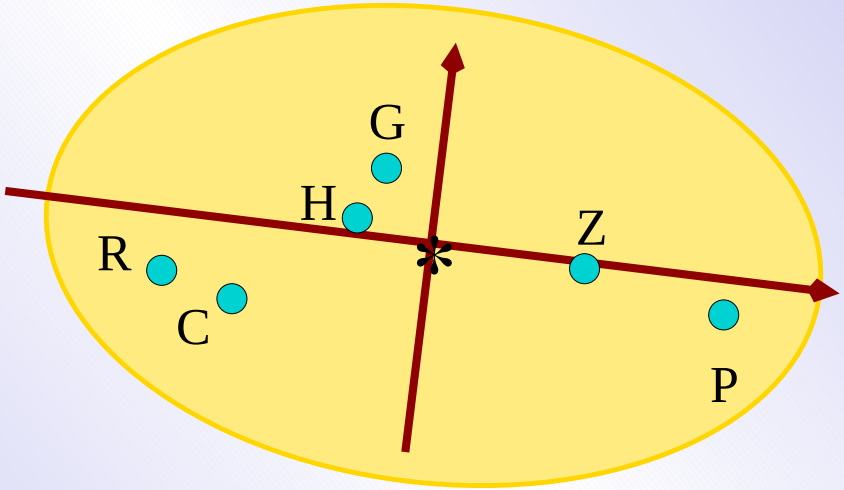



Figure 16: Zooming in.

## Plot them !

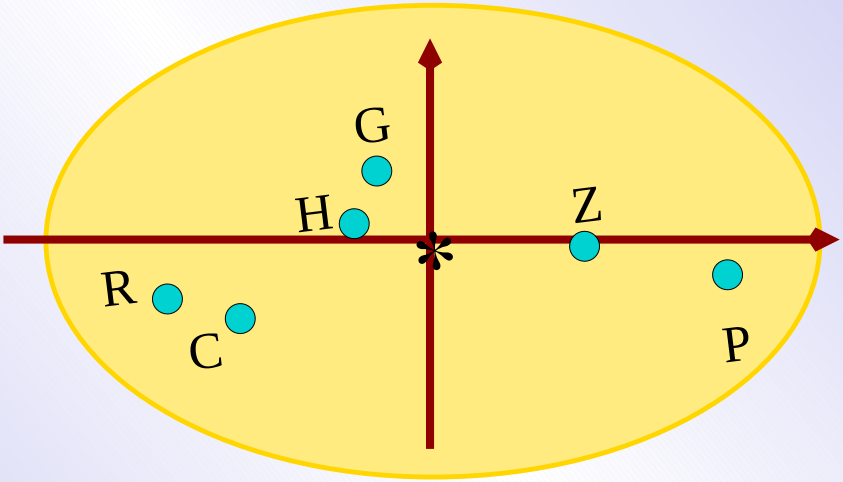



Figure 17: Zooming in.

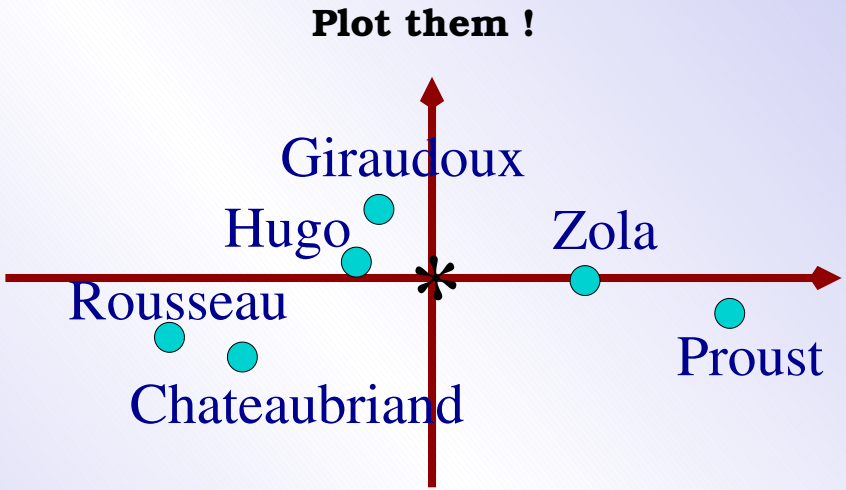



Figure 18: Zooming in.

Refresher! and Menu
Example: . . .
Chi-Square Refresher: . . .
How to Analyze the . . .
Home Page
Title Page
Page 31 of 85
Go Back
Full Screen
Close
Quit

# Pause Inertia

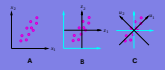
$$\mathcal{I}_R = \sum_{i,j} m_i \left[\sqrt{w_j}\,(r_{i,j} - c_j)\right]^2$$

$$= \sum_{i,j} m_i w_j \,(r_{i,j} - c_j)^2 \quad \text{but: } w_j = c_j^{-1}$$

$$= \sum_{i,j} m_i c_j^{-1} \,(r_{i,j} - c_j)^2$$

$$= \sum_{i,j} m_i \frac{(r_{i,j} - c_j)^2}{c_j}$$

$$= \frac{\chi^2}{x_{++}}$$



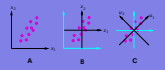
# Pause Inertia

So

$$\mathcal{I}_R = \frac{\chi^2}{x_{++}}$$

The inertia of the row profiles is proportional to $\chi^2$.

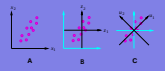
**Factors are $\chi^2$ orthogonal slices!**



# Pause Properties
# Distributional Equivalence

If two rows have the same profiles:
We can add them up

And nothing is changed!



**The Rows: Correspondence Analysis**

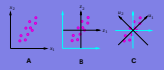
| | $F_1$ | $F_2$ | $\text{ctr}_1$ | $\text{ctr}_2$ | $m\times F_1^2$ | $m\times F_2^2$ | $m\times d^2$ | $\cos_1^2$ | $\cos_2^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Rousseau | −0.24 | −0.07 | 6 | 2 | .0011 | .0001 | .0012 | 91 | 9 |
| Chateaubriand | −0.19 | −0.11 | 28 | 29 | .0050 | .0016 | .0066 | 76 | 24 |
| Hugo | −0.10 | 0.03 | 15 | 4 | .0027 | .0002 | .0029 | 92 | 8 |
| Zola | 0.09 | −0.00 | 19 | 0 | .0033 | .0000 | .0033 | 100 | 0 |
| Proust | 0.22 | −0.06 | 31 | 8 | .0055 | .0004 | .0059 | 93 | 7 |
| Giraudoux | −0.05 | 0.20 | 1 | 58 | .0002 | .0032 | .0034 | 6 | 94 |
| $\sum$ | — | — | 100 | 100 | .0178 | .0056 | | | |
| | | | | | $\lambda_1$ | $\lambda_2$ | $\mathcal{I}_\mathcal{R}$ | | |
| | | | | | 76% | 24% | | | |
| | | | | | $\tau_1$ | $\tau_2$ | | | |

Table 2: The Projections, masse × Squared projections, Contributions, inertia to barycenter, and (Squared) Cosines for the punctuation example

## The columns point are too far away

### How to Keep Them ?

Idea: Compute their Inertia and Normalize



## The columns point are too far away

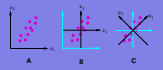
**A bit of Magic here**

The Inertia of the column points/vertices is equal to 1 for each dimension!

Gasp!!!!! Why ?

Intuition: Extreme points have maximum correlation!

For more, see end of slides: Mathematical digression



# Plot them !

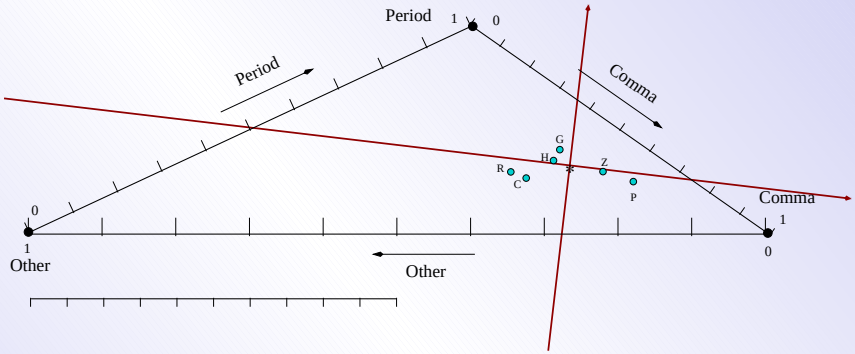



Figure 19: Zooming in.

# Plot them !

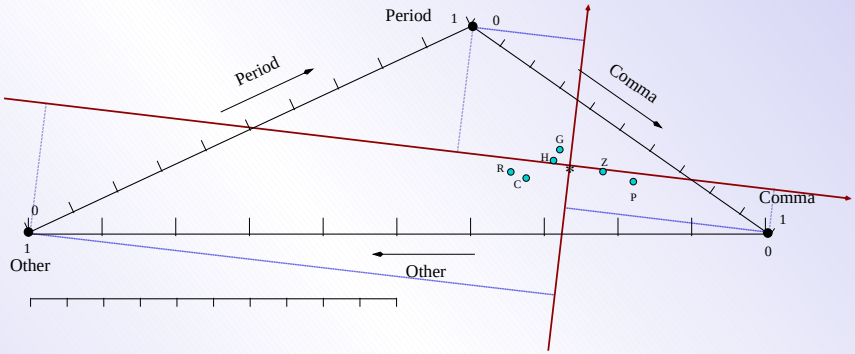



Figure 20: Zooming in.

# Plot them !

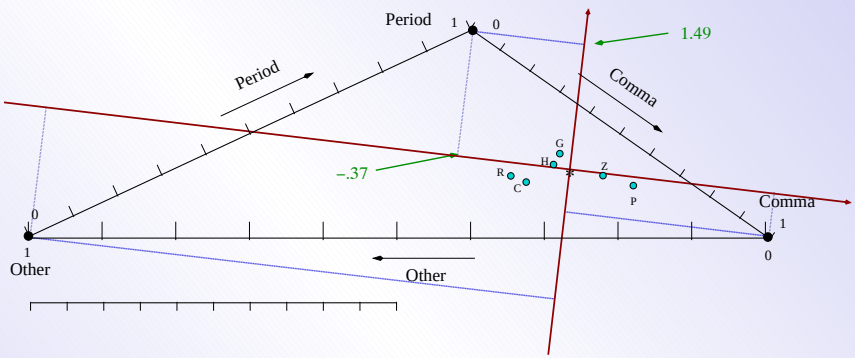



Figure 21: Zooming in.

# Plot them !

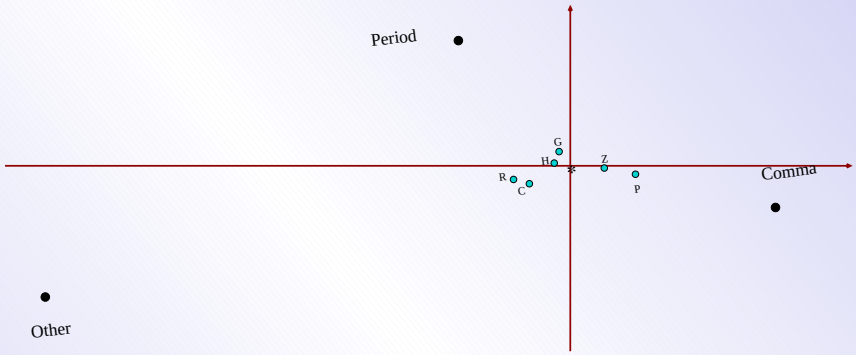



Figure 22: Zooming in.

Refresher! and Menu
Example: . . .
Chi-Square Refresher: . . .
How to Analyze the . . .
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 41 of 85
Go Back
Full Screen
Close
Quit

# Plot them !

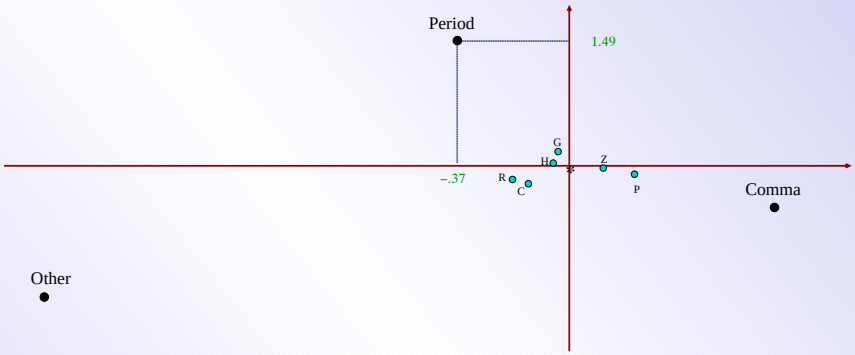



Figure 23: Zooming in.

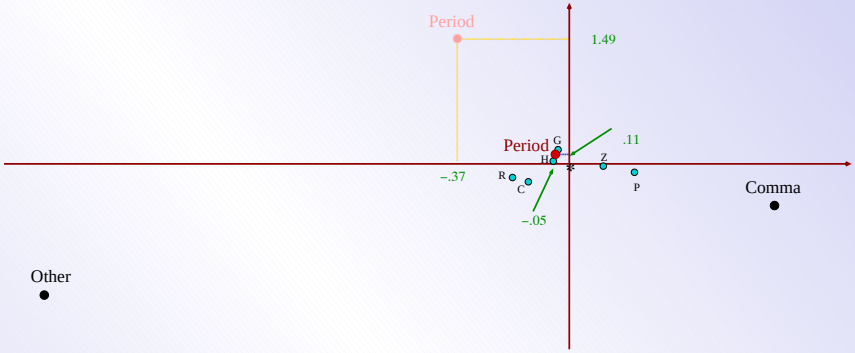



Figure 24: Zooming in.

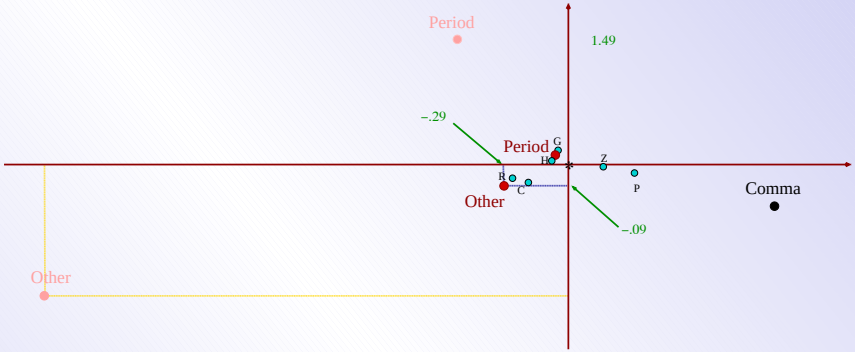



Figure 25: Zooming in.

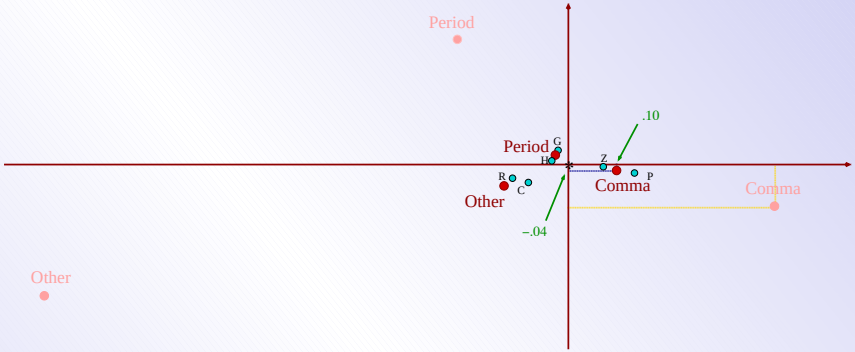



Figure 26: Zooming in.

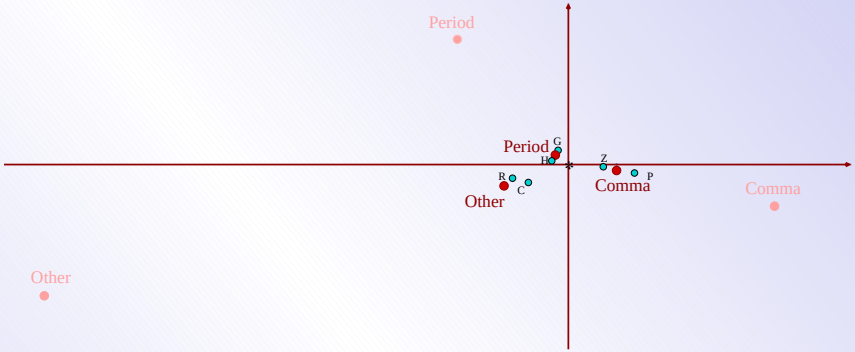



Figure 27: Zooming in.

Refresher! and Menu
Example: . . .
Chi-Square Refresher: . . .
How to Analyze the . . .
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 46 of 85
Go Back
Full Screen
Close
Quit

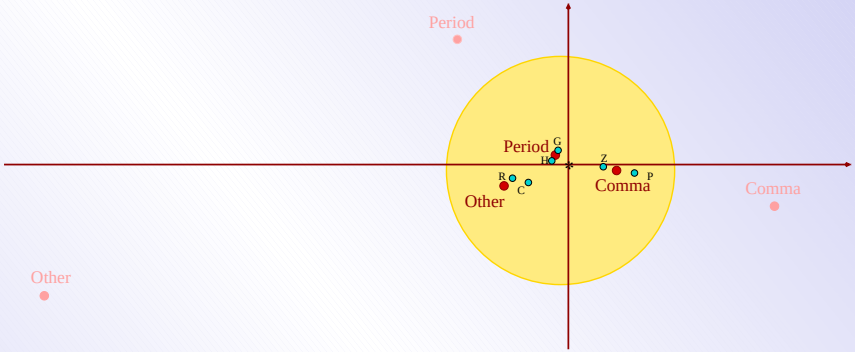



Figure 28: Zooming in.

Refresher! and Menu
Example: . . .
Chi-Square Refresher: . . .
How to Analyze the . . .
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 47 of 85
Go Back
Full Screen
Close
Quit

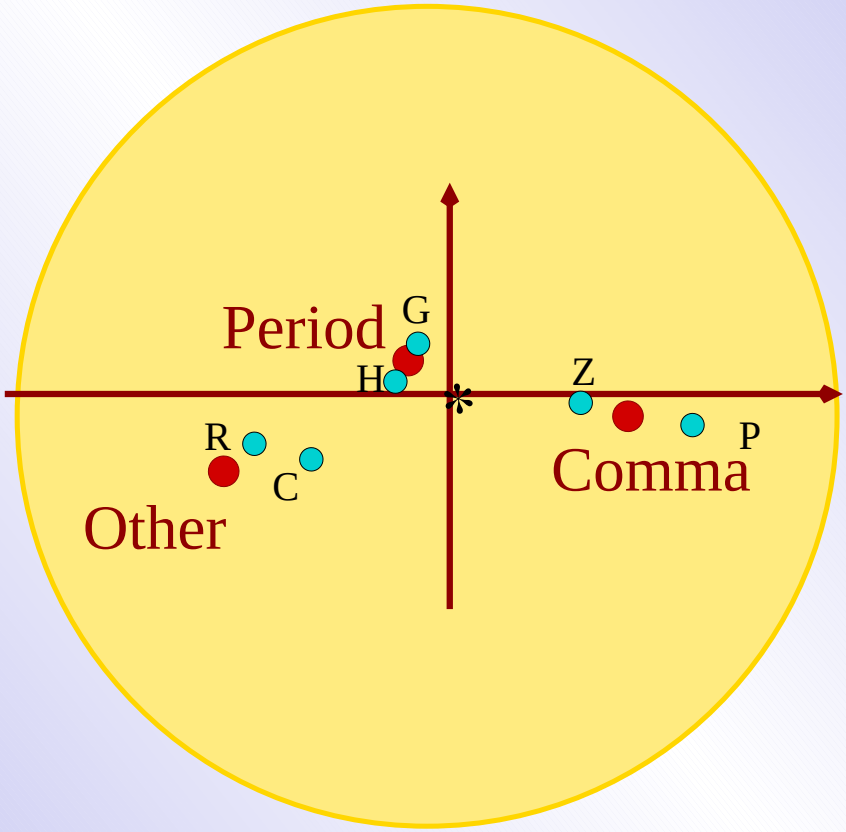



Figure 29: Zooming in.

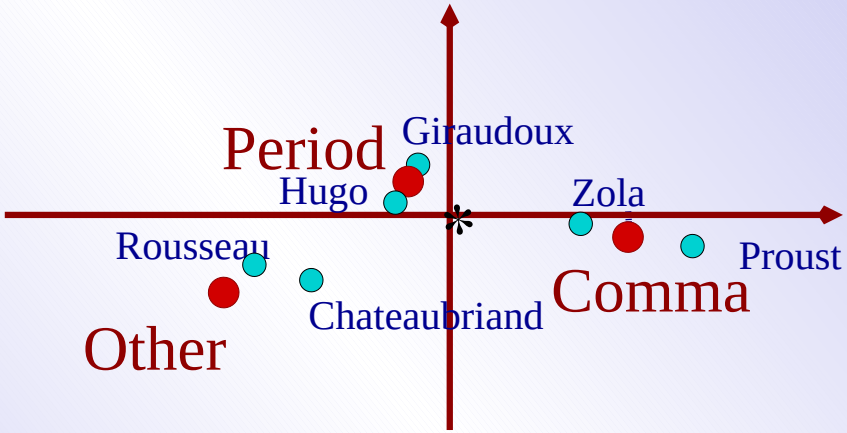
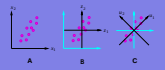



Figure 30: Zooming in.

| | $F_1$ | $F_2$ | $\text{ctr}_1$ | $\text{ctr}_2$ | $m\times F_1^2$ | $m\times F_2^2$ | $m\, d^2$ | $\cos_1^2$ | $\cos_2^2$ |
|---|---|---|---|---|---|---|---|---|---|
| PERIOD | −0.05 | 0.11 | 4 | 66 | .0007 | .0037 | .0044 | 16 | 84 |
| COMMA | 0.10 | −0.04 | 30 | 14 | .0053 | .0008 | .0061 | 88 | 12 |
| OTHER | −0.29 | −0.09 | 66 | 20 | .0118 | .0011 | .0129 | 91 | 9 |
| $\sum$ | — | — | 100 | 100 | .0178 | .0056 | | | |
| | | | | | $\lambda_1$ | $\lambda_2$ | $\mathcal{I}_\mathcal{C}$ | | |
| | | | | | 76% | 24% | | | |
| | | | | | $\tau_1$ | $\tau_2$ | | | |

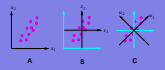
Table 3: Columns: The Projections, masse × Squared projections, Contributions, inertia to barycenter, and Cosines for the punctuation example

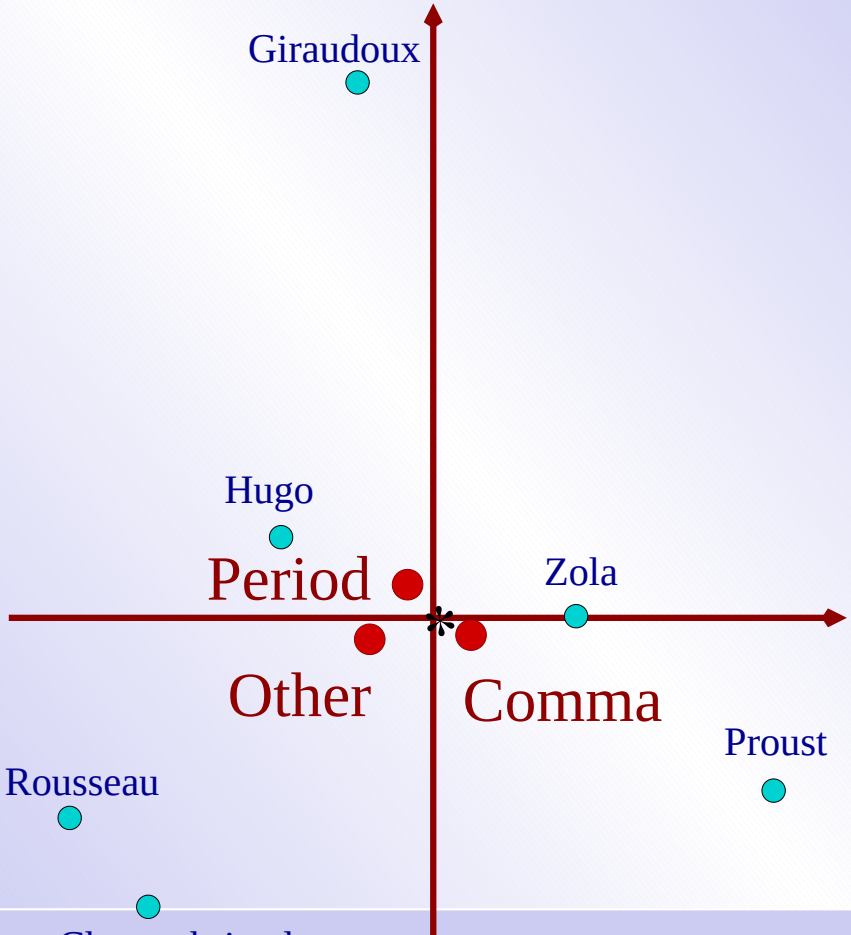


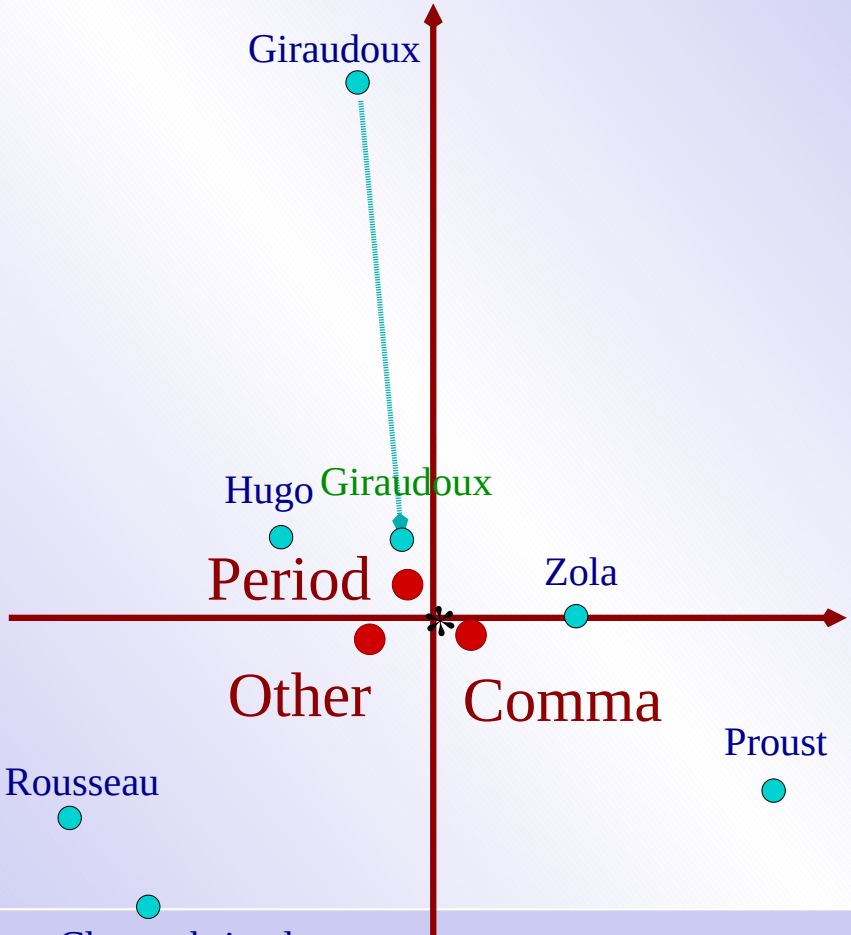
# The dual analysis: the Column Profiles

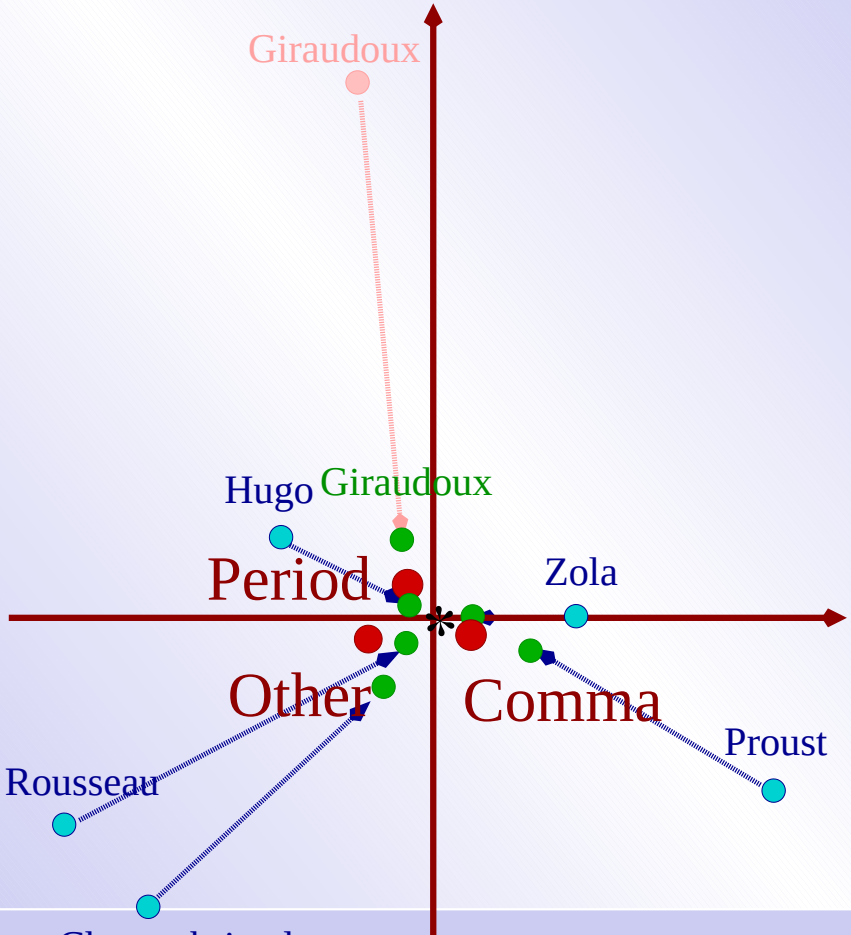
# The Column Profiles

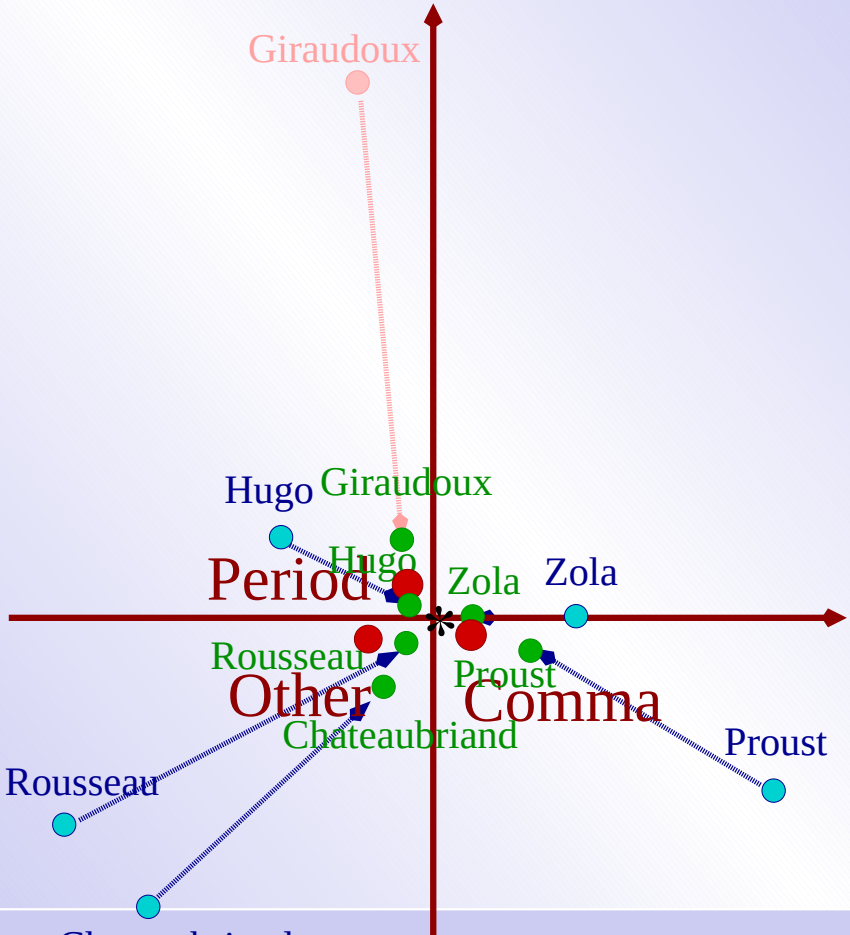
| | Period | Comma | All the other marks |
|---|---|---|---|
| Rousseau | .0185 | .0163 | .0305 |
| Chateaubriand | .1267 | .1273 | .21349 |
| Hugo | .2729 | .2295 | .3000 |
| Zola | .3823 | .4235 | .3179 |
| Proust | .0901 | .1307 | .0642 |
| Giraudoux | .1095 | .0726 | .0724 |

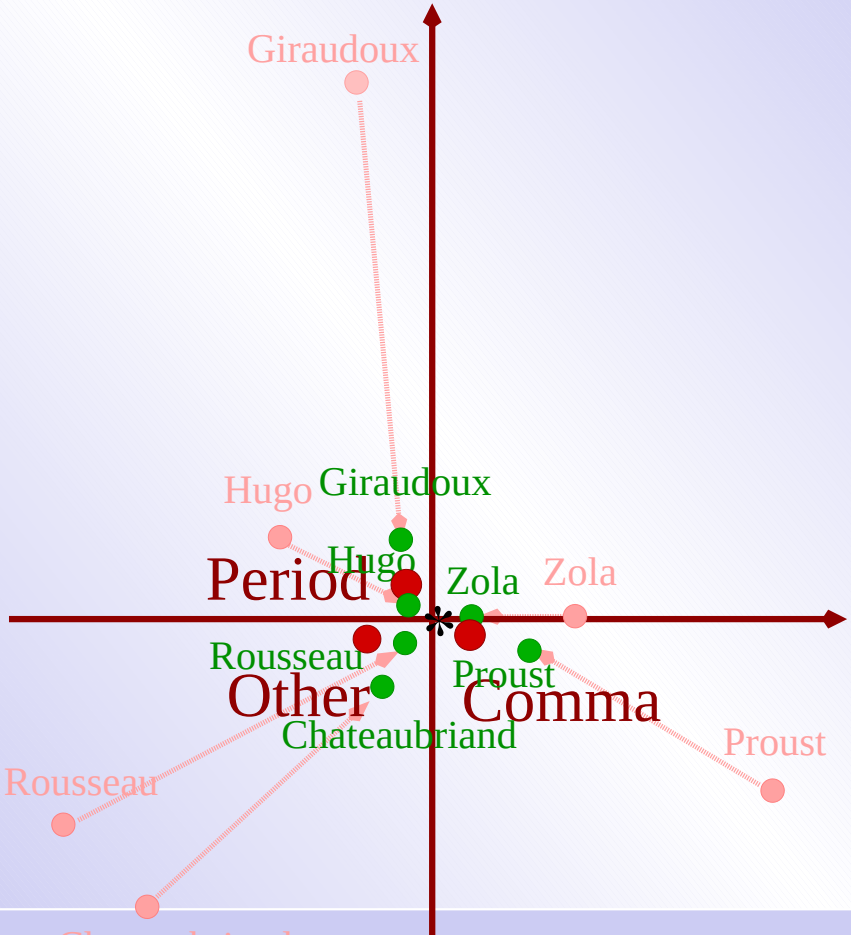


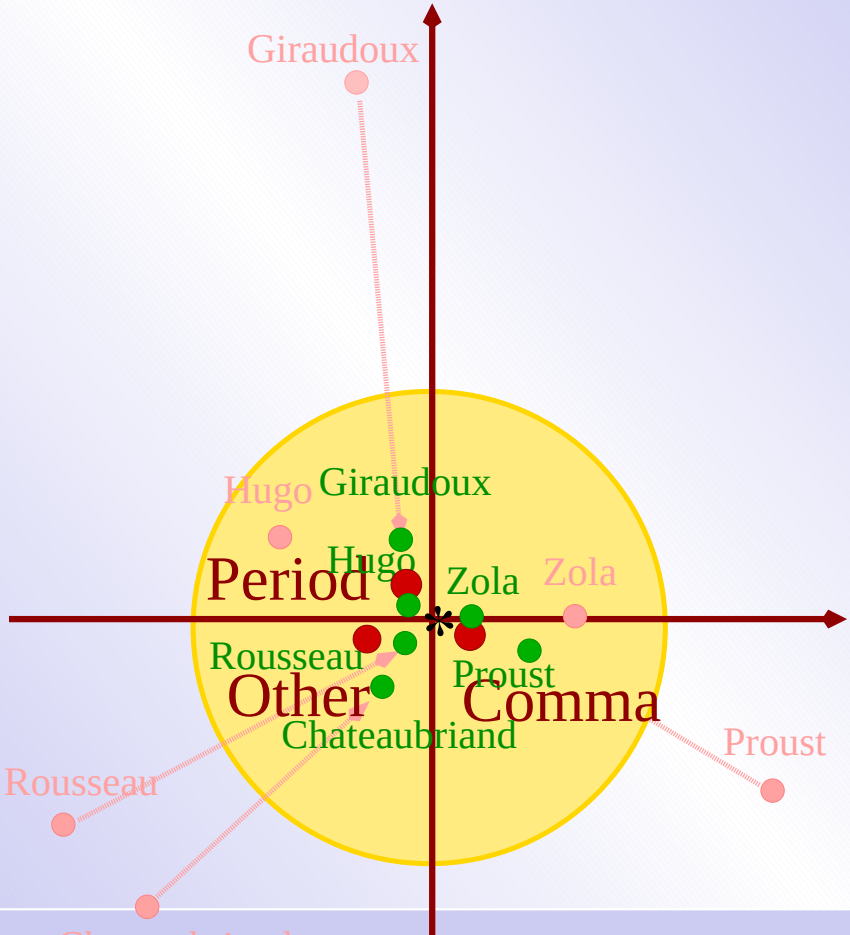

Giraudoux
Hugo
Period
Zola
Other
Comma
Proust
Rousseau


Giraudoux
Hugo
Giraudoux
Period
Zola
Other
Comma
Proust
Rousseau


Giraudoux
Hugo
Giraudoux
Period
Zola
Other
Comma
Proust
Rousseau


Giraudoux
Hugo
Giraudoux
Hugo
Period
Zola
Zola
Rousseau
Proust
Other
Comma
Chateaubriand
Proust
Rousseau


Giraudoux
Hugo
Giraudoux
Hugo
Period
Zola
Zola
Rousseau
Proust
Other
Comma
Chateaubriand
Proust
Rousseau


Giraudoux
Hugo
Giraudoux
Hugo
Period
Zola
Zola
Rousseau
Proust
Other
Comma
Chateaubriand
Proust
Rousseau

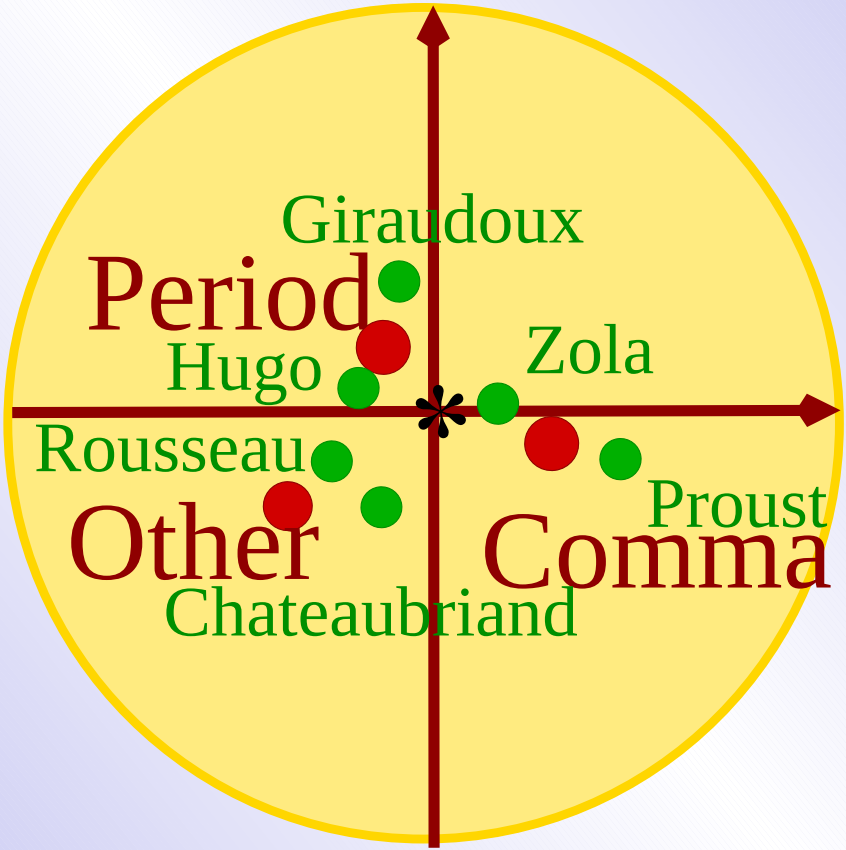



Figure 37: Zooming in.

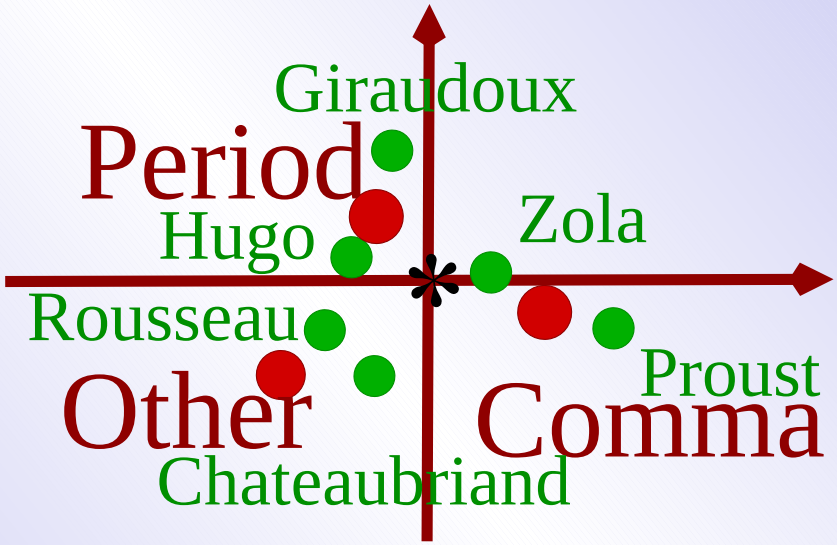
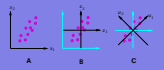
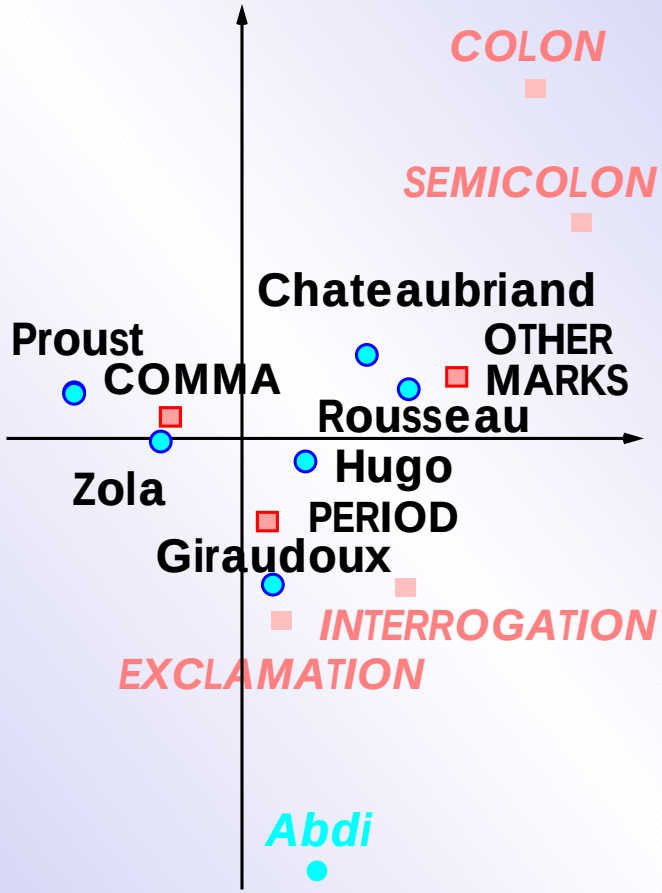



Figure 38: Zooming in.

# Supplementary Elements

- One new author: Abdi (1994) Chapter 1
- The other punctuation marks (colon, semicolon, interrogation, exclamation).




COLON
SEMICOLON
Chateaubriand
Proust
OTHER
COMMA
MARKS
Rousseau
Hugo
Zola
PERIOD
Giraudoux
INTERROGATION
EXCLAMATION
Abdi

# Another Example: The Sound of Colors

from: Abdi & Béra (2014). paper # C82

A science project (independent and dependent variables): Twenty-two participants (14 children and 10 adults) were presented with nine "pieces of music" and asked to associate one of ten colors to each piece of music. The pieces of music:

- 1) the music of a video game (video),
- 2) a Jazz song (jazz),
- 3) a country and western song (country),
- 4) a rap song (rap),
- 5) a pop song (pop),
- 6) an extract of the opera Carmen (opera),
- 7) the low F note played on a piano (low F),
- 8) the middle F note played on the same piano
- 9) the high F note still played on the same piano.

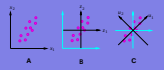
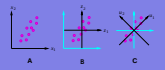


Table 4: Twenty-two participants associated one of ten colors to nine pieces of music. The column labeled $x_{i+}$ gives the total number of choices made for each color. $N$ is the grand total of the data table. The vector of mass for the rows, $\mathbf{r}$, is the proportion of choices made for each color ($r_i = x_{i+}/N$). The row labeled $x_{+j}$ gives the total number of times each piece of music was presented (*i.e.*, it is equal to the number of participants) The centroid row, gives these values as proportions.

| Color | Video | Jazz | Country | Rap | Pop | Opera | Low F | High F | Middle F | $x_{i+}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| red | 4 | 2 | 4 | 4 | 1 | 2 | 2 | 4 | 1 | 24 |
| orange | 3 | 4 | 2 | 2 | 1 | 1 | 0 | 3 | 2 | 18 |
| yellow | 6 | 4 | 5 | 2 | 3 | 1 | 1 | 3 | 0 | 25 |
| green | 2 | 0 | 5 | 1 | 3 | 3 | 3 | 1 | 5 | 23 |
| blue | 2 | 5 | 0 | 1 | 4 | 1 | 2 | 1 | 3 | 19 |
| purple | 3 | 3 | 1 | 0 | 0 | 3 | 0 | 2 | 1 | 13 |
| white | 0 | 0 | 0 | 0 | 1 | 4 | 1 | 5 | 3 | 14 |
| black | 0 | 2 | 0 | 11 | 1 | 3 | 10 | 1 | 1 | 29 |
| pink | 2 | 1 | 1 | 0 | 2 | 4 | 0 | 2 | 0 | 12 |
| brown | 0 | 1 | 4 | 1 | 6 | 0 | 3 | 0 | 6 | 21 |
| $x_{+j}$ | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | 22 | $N = 198$ |
| $\mathbf{c}^T$ | .11 | .11 | .11 | .11 | .11 | .11 | .11 | .11 | .11 | |



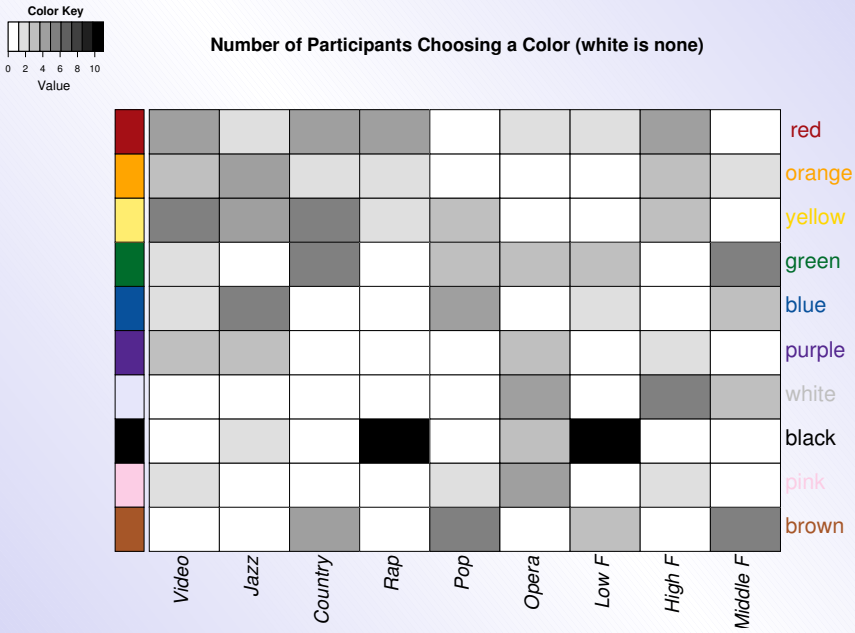
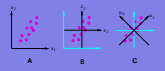



Figure 40: CA The Colors of Music. A nicer way to look at the data. A heat map of the data.



Refresher! and Menu
Example: . . .
Chi-Square Refresher: . . .
How to Analyze the . . .
Home Page
Title Page
Page 65 of 85
Go Back
Full Screen
Close
Quit

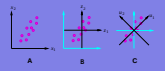


Table 5: CA The Color of Music. Factor scores, contributions, mass, mass $\times$ squared factor scores, inertia to barycenter, and squared cosines for the rows. For convenience, squared cosines and contributions have been multiplied by 1000 and rounded.

| | $F_1$ | $F_2$ | $\text{ctr}_1$ | $\text{ctr}_2$ | $r_i$ | $r_i \times F_1^2$ | $r_i \times F_2^2$ | $r_i \times d_{r,i}^2$ | $\cos_1^2$ | $\cos_2^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| red | −0.026 | 0.299 | 0 | 56 | .121 | .000 | .011 | .026 | 3 | 410 |
| orange | −0.314 | 0.232 | 31 | 25 | .091 | .009 | .005 | .030 | 295 | 161 |
| yellow | −0.348 | 0.202 | 53 | 27 | .126 | .015 | .005 | .057 | 267 | 89 |
| green | −0.044 | −0.490 | 1 | 144 | .116 | .000 | .028 | .048 | 5 | 583 |
| blue | −0.082 | −0.206 | 2 | 21 | .096 | .001 | .004 | .050 | 13 | 81 |
| purple | −0.619 | 0.475 | 87 | 77 | .066 | .025 | .015 | .050 | 505 | 298 |
| white | −0.328 | 0.057 | 26 | 1 | .071 | .008 | .000 | .099 | 77 | 2 |
| black | 1.195 | 0.315 | 726 | 75 | .146 | .208 | .014 | .224 | 929 | 65 |
| pink | −0.570 | 0.300 | 68 | 28 | .061 | .020 | .005 | .053 | 371 | 103 |
| brown | 0.113 | −0.997 | 5 | 545 | .106 | .001 | .105 | .108 | 12 | 973 |
| $\sum$ | — | — | 1000 | 1000 | — | .287 | .192 | .746 | | |
| | | | | | | $\lambda_1$ | $\lambda_2$ | $\mathcal{I}$ | | |
| | | | | | | 39% | 26% | | | |
| | | | | | | $\tau_1$ | $\tau_2$ | | | |

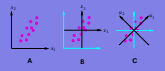




Table 6: CA The Colors of Music. Factor scores, contributions, mass, mass $\times$ squared factor scores, inertia to barycenter, and squared cosines for the columns. For convenience, squared cosines and contributions have been multiplied by 1000 and rounded.

| | $G_1$ | $G_2$ | $\widetilde{G}_1$ | $\widetilde{G}_2$ | $\text{ctr}_1$ | $\text{ctr}_2$ | $c_j$ | $c_j \times G_1^2$ | $c_j \times G_2^2$ | $c_j \times d_{c,j}^2$ | $\cos_1^2$ | $\cos_2^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Video | −0.54 | 0.39 | −1.01 | 0.88 | 113 | 86 | .111 | .03 | .017 | .071 | 454 | 232 |
| Jazz | −0.26 | 0.28 | −0.48 | 0.63 | 25 | 44 | .111 | .01 | .008 | .069 | 105 | 121 |
| Country | −0.29 | −0.31 | −0.54 | −0.70 | 33 | 55 | .111 | .01 | .011 | .066 | 142 | 161 |
| Rap | 0.99 | 0.40 | 1.85 | 0.90 | 379 | 91 | .111 | .11 | .017 | .133 | 822 | 132 |
| Pop | −0.12 | −0.64 | −0.23 | −1.45 | 6 | 234 | .111 | .00 | .045 | .064 | 26 | 709 |
| Opera | −0.24 | 0.33 | −0.44 | 0.74 | 22 | 61 | .111 | .01 | .012 | .079 | 78 | 149 |
| Low.F | 0.95 | −0.09 | 1.78 | −0.20 | 351 | 5 | .111 | .10 | .001 | .105 | 962 | 8 |
| High.F | −0.43 | 0.41 | −0.80 | 0.93 | 70 | 96 | .111 | .02 | .018 | .074 | 271 | 249 |
| Middle.F | −0.07 | −0.76 | −0.13 | −1.72 | 2 | 330 | .111 | .00 | .064 | .084 | 7 | 759 |
| $\sum$ | — | — | — | — | 1000 | 1000 | — | .287<br>$\lambda_1$<br>39%<br>$\tau_1$ | .192<br>$\lambda_2$<br>26%<br>$\tau_2$ | .746<br>$\mathcal{I}$ | | |

Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 67 of 85
Go Back
Full Screen
Close
Quit

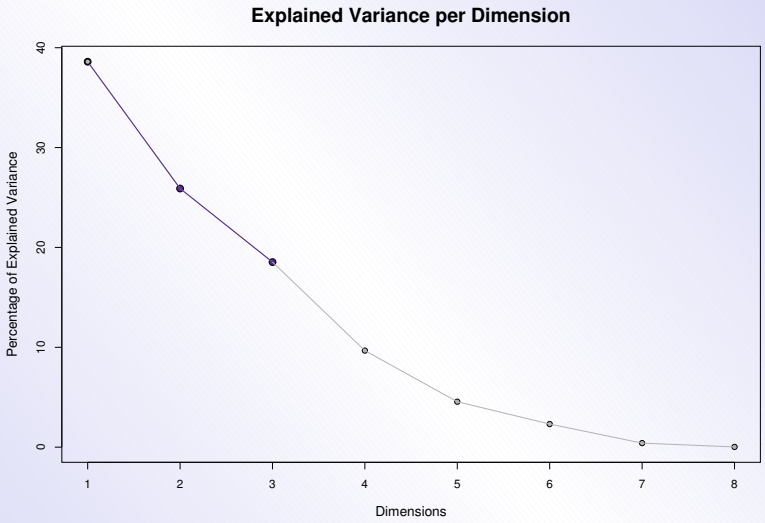



Figure 41: CA The Colors of Music. Scree Test of the percentage of explained variance. Three dimensions are larger than $\frac{1}{8}$.

Refresher! and Menu
Example: . . .
Chi-Square Refresher: . . .
How to Analyze the . . .
Home Page
Title Page
Page 68 of 85
Go Back
Full Screen
Close
Quit

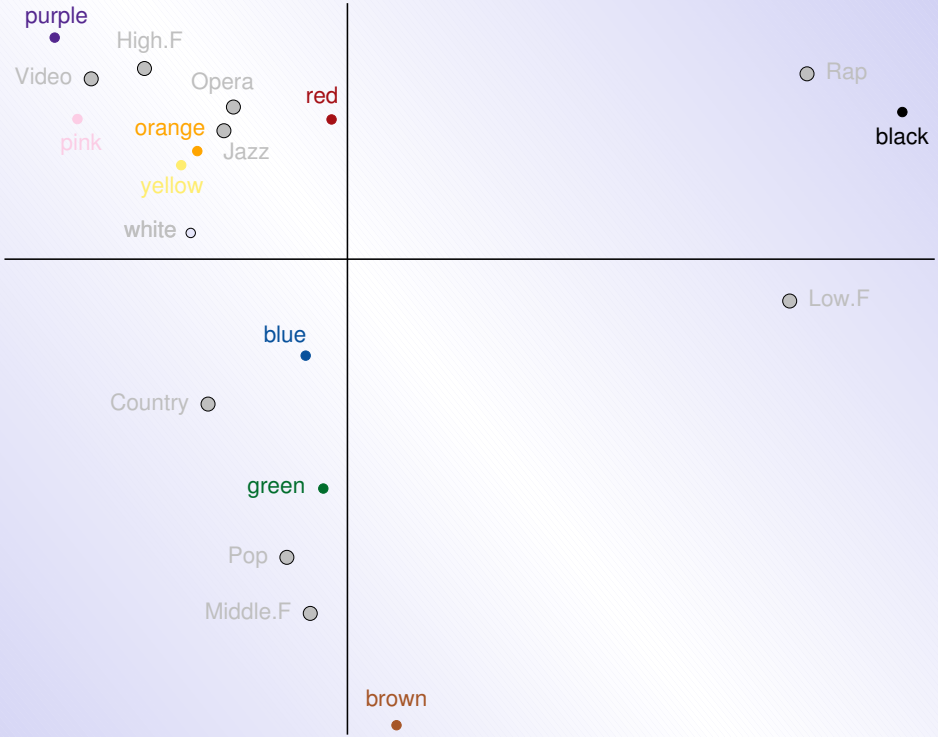
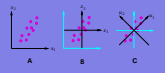



Figure 42: CA The Colors of Music. Symmetric Plot: The projections of the rows and the columns are displayed in the same map. $\lambda_1 = .287$, $\tau_1 = 39$; $\lambda_2 = .192$, $\tau_2 = 26$. In this plot the proximity between rows and columns cannot be directly interpreted.



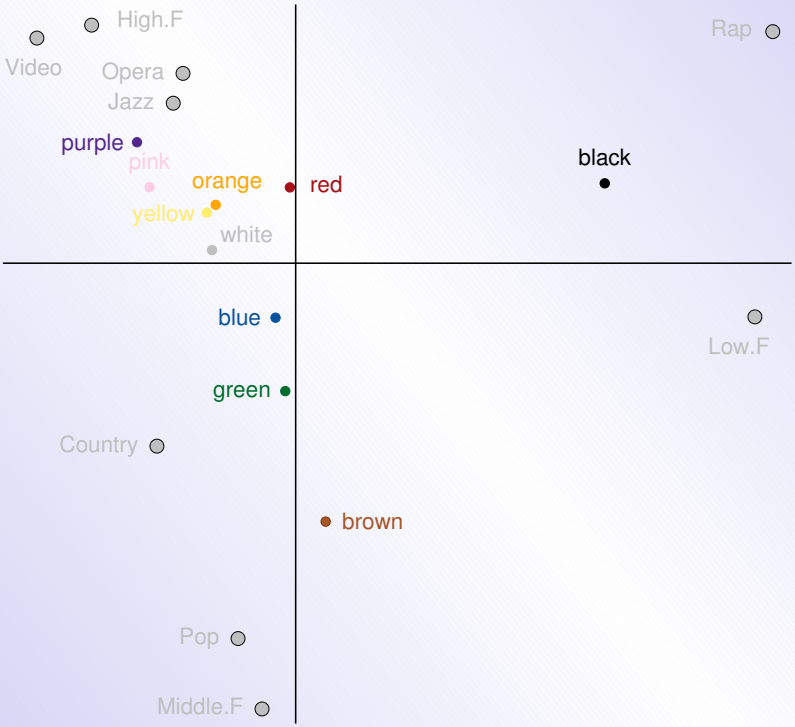
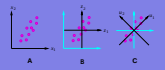



Figure 43: CA The Colors of Music. Asymmetric Plot: The projections of the rows and the columns are displayed in the same map. The inertia of the projections of the column factor scores is equal to one for each dimension and the inertia of the projections of the row factor scores are $\lambda_1 = .287$, $\tau_1 = 39$; $\lambda_2 = .192$, $\tau_2 = 26$.



# bonus Zola

# Bonus: Zola

Table 7: the punctuation of the 22 Novels of Zola. Original Data from Brunet (1996??).

| | . | ... | ! | ? | , | ; | : | – | — | “” | *It* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 Raquin | 3468 | 236 | 138 | 76 | 6195 | 691 | 168 | 285 | 543 | 56 | 83 |
| 2 Ferrat | 5131 | 362 | 236 | 245 | 8012 | 922 | 291 | 518 | 1115 | 174 | 83 |
| 3 Fortune | 6157 | 238 | 534 | 229 | 11346 | 936 | 362 | 711 | 1301 | 238 | 142 |
| 4 Curée | 4958 | 443 | 357 | 232 | 11164 | 738 | 364 | 679 | 1200 | 234 | 73 |
| 5 Ventre | 5538 | 534 | 426 | 232 | 13234 | 1015 | 318 | 734 | 1201 | 250 | 42 |
| 6 Conquête | 6292 | 934 | 756 | 512 | 11585 | 1285 | 402 | 1432 | 1916 | 162 | 74 |
| 7 Faute | 6364 | 679 | 859 | 462 | 13948 | 634 | 377 | 1067 | 1564 | 72 | 165 |
| 8 Excellence | 7258 | 728 | 1002 | 496 | 14295 | 889 | 543 | 1469 | 1907 | 247 | 116 |
| 9 Assommoir | 7820 | 769 | 1929 | 443 | 19244 | 1399 | 436 | 995 | 2272 | 76 | 311 |
| 10 Page | 6206 | 843 | 918 | 492 | 11953 | 647 | 347 | 1235 | 1409 | 107 | 80 |
| 11 Nana | 7821 | 1007 | 1796 | 611 | 17881 | 1087 | 509 | 1523 | 1797 | 185 | 80 |
| 12 Pot Bouille | 6875 | 1045 | 1873 | 651 | 17044 | 912 | 675 | 1669 | 1935 | 78 | 162 |
| 13 Bonheur | 6916 | 808 | 1313 | 651 | 18402 | 972 | 642 | 1531 | 2114 | 88 | 513 |
| 14 joie | 5803 | 710 | 972 | 623 | 13917 | 602 | 420 | 1142 | 1590 | 40 | 19 |
| 15 Germinal | 7944 | 606 | 1463 | 729 | 21388 | 908 | 621 | 1362 | 2083 | 46 | 117 |
| 16 Oeuvre | 5000 | 774 | 1692 | 668 | 18292 | 811 | 566 | 1107 | 1489 | 88 | 134 |
| 17 Terre | 6976 | 957 | 2307 | 796 | 23417 | 947 | 657 | 1681 | 2113 | 70 | 95 |
| 18 Rêve | 3052 | 292 | 385 | 237 | 9551 | 345 | 230 | 416 | 650 | 78 | 101 |
| 19 Bête | 5484 | 601 | 929 | 557 | 18264 | 673 | 467 | 957 | 1721 | 46 | 24 |
| 20 Argent | 5022 | 850 | 1235 | 569 | 19267 | 684 | 399 | 1049 | 1677 | 48 | 101 |
| 21 Débacle | 7440 | 860 | 1833 | 690 | 26482 | 832 | 564 | 1398 | 2197 | 68 | 64 |
| 22 Pascal | 4586 | 621 | 1072 | 464 | 15598 | 462 | 315 | 955 | 1218 | 50 | 44 |

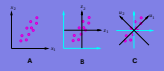


Table 8: The novels of Zola and their punctuation. The Projections, Contributions, inertia to barycenter, and Cosines for the punctuation example. First Analysis. All punctuation marks as active elements

| | $F_1$ | $F_2$ | $\text{ctr}_1$ | $\text{ctr}_2$ | $m \times d^2$ | $\cos_1^2$ | $\cos_2^2$ | Qual |
|---|---|---|---|---|---|---|---|---|
| 1 Raquin | −.24 | −.13 | 7 | 6 | .0018 | 62 | 18 | 81 |
| 2 Ferrat | −.30 | −.06 | 16 | 2 | .0027 | 94 | 3 | 97 |
| 3 Fortune | −.20 | −.09 | 10 | 5 | .0020 | 75 | 13 | 88 |
| 4 Curée | −.14 | −.10 | 4 | 6 | .0011 | 54 | 31 | 84 |
| 5 Ventre | −.12 | −.13 | 4 | 11 | .0014 | 39 | 44 | 83 |
| 6 Conquête | −.19 | .12 | 10 | 10 | .0025 | 58 | 22 | 79 |
| 7 Faute | −.05 | .02 | 1 | 0 | .0004 | 25 | 3 | 28 |
| 8 Excellence | −.12 | .05 | 5 | 3 | .0011 | 65 | 13 | 78 |
| 9 Assommoir | −.01 | .02 | 0 | 0 | .0012 | 0 | 2 | 2 |
| 10 Page | −.09 | .09 | 2 | 6 | .0010 | 32 | 37 | 68 |
| 11 Nana | −.02 | .07 | 0 | 5 | .0006 | 4 | 46 | 50 |
| 12 Pot Bouille | .03 | .13 | 0 | 17 | .0010 | 4 | 89 | 93 |
| 13 Bonheur | −.00 | .05 | 0 | 3 | .0017 | 0 | 8 | 8 |
| 14 Joie | .01 | .05 | 0 | 2 | .0005 | 1 | 19 | 20 |
| 15 Germinal | .05 | −.02 | 1 | 1 | .0004 | 31 | 9 | 39 |
| 16 L'œuvre | .15 | .02 | 7 | 0 | .0013 | 80 | 1 | 81 |
| 17 Terre | .14 | .04 | 8 | 2 | .0015 | 84 | 8 | 91 |
| 18 Rêve | .06 | −.13 | 1 | 8 | .0007 | 12 | 68 | 79 |
| 19 Bête | .10 | −.08 | 3 | 6 | .0010 | 48 | 31 | 79 |
| 20 Argent | .15 | −.04 | 7 | 2 | .0014 | 81 | 7 | 88 |
| 21 Débacle | .15 | −.06 | 10 | 5 | .0018 | 82 | 15 | 97 |
| 22 Pascal | .13 | −.04 | 5 | 1 | .0009 | 84 | 6 | 90 |

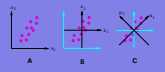
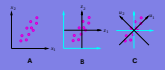


Table 9: Zola's punctuation. Columns: The Projections, Contributions, inertia to barycenter, and Cosines

| | $F_1$ | $F_2$ | $\text{ctr}_1$ | $\text{ctr}_2$ | $m \times d^2$ | $\cos_1$ | $\cos_2$ | Qual |
|---|---|---|---|---|---|---|---|---|
| . | −0.16 | 0.00 | 37 | 0 | .0061 | 93 | 0 | 93 |
| . . . | −0.01 | 0.17 | 0 | 13 | .0013 | 0 | 55 | 55 |
| ! | 0.22 | 0.20 | 12 | 28 | .0042 | 44 | 36 | 81 |
| ? | 0.11 | 0.14 | 1 | 6 | .0008 | 25 | 39 | 64 |
| ; | 0.07 | −0.05 | 20 | 21 | .0043 | 72 | 27 | 100 |
| : | −0.28 | −0.02 | 15 | 0 | .0032 | 73 | 0 | 73 |
| – | −0.02 | 0.06 | 0 | 1 | .0003 | 2 | 16 | 19 |
| — | −0.03 | 0.18 | 0 | 23 | .0017 | 2 | 74 | 76 |
| “” | −0.06 | 0.06 | 2 | 4 | .0008 | 28 | 24 | 52 |
| | −0.63 | −0.20 | 11 | 3 | .0025 | 63 | 7 | 70 |
| *It* | −0.17 | 0.12 | 1 | 1 | .0026 | 5 | 2 | 7 |



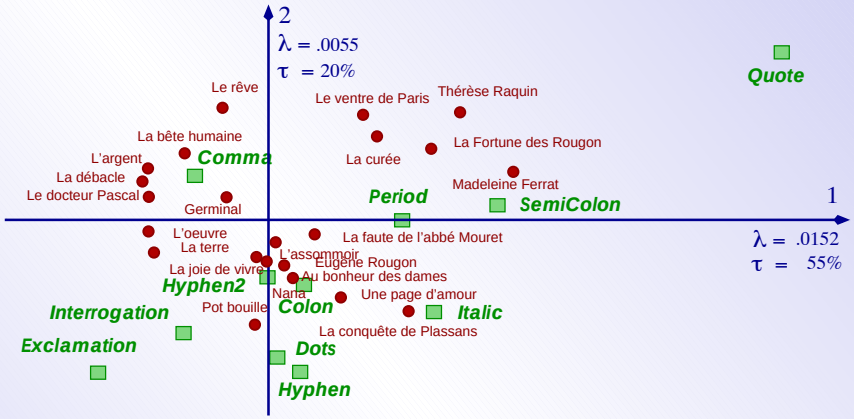



Figure 44: The punctuation of Zola. First Pass. Rows and Columns.

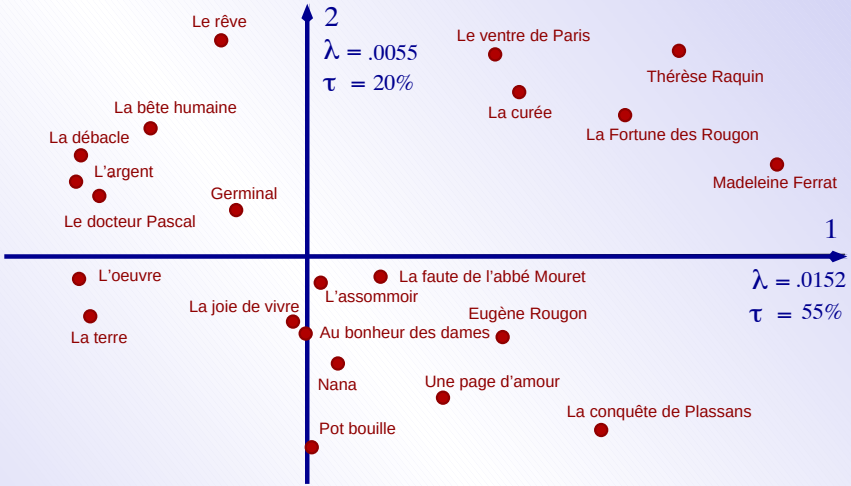



Figure 45: The punctuation of Zola. First Pass. Rows.

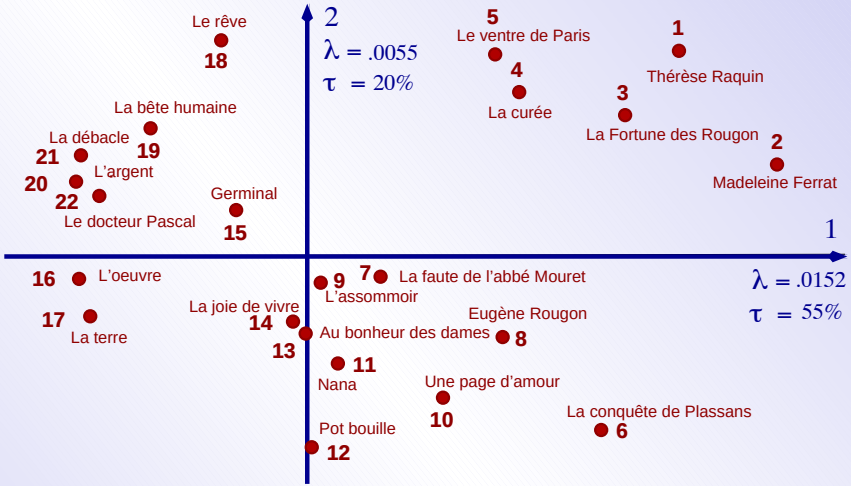



Figure 46: The punctuation of Zola. First Pass. Rows with # of the Novels.

Refresher! and Menu
Example: . . .
Chi-Square Refresher: . . .
How to Analyze the . . .
Home Page
Title Page
Page 77 of 85
Go Back
Full Screen
Close
Quit

# Recode

- Italic and Quote $\Longrightarrow$ supplementary
- Regroup ! and ?
- Regroup – and —

## Zola. Recoded Data

Table 10: the punctuation of the 22 Novels of Zola. Recoded Data: Italic and quote are supplementary elements, Exclamation and Interrogation are summed, Hyphen 1 and 2 are summed. Data from Brunet (1996??).

| | . | ... | !& ? | , | ; | : | – — |
|---|---|---|---|---|---|---|---|
| 1 Raquin | 3468 | 236 | 214 | 6195 | 691 | 168 | 828 |
| 2 Ferat | 5131 | 362 | 481 | 8012 | 922 | 291 | 1633 |
| 3 Fortune | 6157 | 238 | 763 | 11346 | 936 | 362 | 2012 |
| 4 Curée | 4958 | 443 | 589 | 11164 | 738 | 364 | 1879 |
| 5 Ventre | 5538 | 534 | 658 | 13234 | 1015 | 318 | 1935 |
| 6 Conquête | 6292 | 934 | 1268 | 11585 | 1285 | 402 | 3348 |
| 7 Faute | 6364 | 679 | 1321 | 13948 | 634 | 377 | 2631 |
| 8 Excellence | 7258 | 728 | 1498 | 14295 | 889 | 543 | 3376 |
| 9 Assommoir | 7820 | 769 | 2372 | 19244 | 1399 | 436 | 3267 |
| 10 Page | 6206 | 843 | 1410 | 11953 | 647 | 347 | 2644 |
| 11 Nana | 7821 | 1007 | 2407 | 17881 | 1087 | 509 | 3320 |
| 12 Pot Bouille | 6875 | 1045 | 2524 | 17044 | 912 | 675 | 3604 |
| 13 Bonheur | 6916 | 808 | 1964 | 18402 | 972 | 642 | 3645 |
| 14 joie | 5803 | 710 | 1595 | 13917 | 602 | 420 | 2732 |
| 15 Germinal | 7944 | 606 | 2192 | 21388 | 908 | 621 | 3445 |
| 16 Oeuvre | 5000 | 774 | 2360 | 18292 | 811 | 566 | 2596 |
| 17 Terre | 6976 | 957 | 3103 | 23417 | 947 | 657 | 3794 |
| 18 Rêve | 3052 | 292 | 622 | 9551 | 345 | 230 | 1066 |
| 19 Bête | 5484 | 601 | 1486 | 18264 | 673 | 467 | 2678 |
| 20 Argent | 5022 | 850 | 1804 | 19267 | 684 | 399 | 2726 |
| 21 Débacle | 7440 | 860 | 2523 | 26482 | 832 | 564 | 3595 |
| 22 Pascal | 4586 | 621 | 1536 | 15598 | 462 | 315 | 2173 |

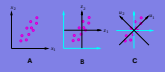
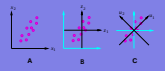


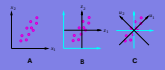
Table 11: Zola's punctuation. Columns: The Projections, Contributions, inertia to barycenter, and Cosines. Recoded Data: Italic and quote are supplementary elements, Exclamation and Interrogation are summed, Hyphen 1 and 2 are summed.

| | $F_1$ | $F_2$ | $\text{ctr}_1$ | $\text{ctr}_2$ | $m \times d^2$ | $\cos_1^2$ | $\cos_2^2$ | Qual |
|---|---|---|---|---|---|---|---|---|
| 1 Raquin | 0.24 | −0.16 | 8 | 10 | .0018 | 65 | 29 | 94 |
| 2 Ferat | 0.28 | −0.06 | 16 | 2 | .0023 | 95 | 4 | 99 |
| 3 Fortune | 0.18 | −0.09 | 8 | 6 | .0016 | 71 | 19 | 90 |
| 4 Curée | 0.10 | −0.09 | 2 | 5 | .0006 | 50 | 40 | 90 |
| 5 Ventre | 0.09 | −0.12 | 2 | 11 | .0010 | 33 | 57 | 90 |
| 6 Conquête | 0.19 | 0.11 | 11 | 10 | .0024 | 64 | 21 | 85 |
| 7 Faute | 0.05 | 0.01 | 1 | 0 | .0003 | 40 | 1 | 40 |
| 8 Excellence | 0.11 | 0.05 | 4 | 3 | .0009 | 68 | 16 | 84 |
| 9 Assommoir | 0.02 | 0.01 | 0 | 0 | .0004 | 6 | 3 | 9 |
| 10 Page | 0.10 | 0.08 | 3 | 5 | .0009 | 43 | 31 | 75 |
| 11 Nana | 0.03 | 0.07 | 0 | 5 | .0004 | 9 | 62 | 71 |
| 12 Pot Bouille | −0.01 | 0.12 | 0 | 16 | .0008 | 1 | 94 | 95 |
| 13 Bonheur | −0.01 | 0.03 | 0 | 1 | .0002 | 2 | 36 | 38 |
| 14 joie | 0.01 | 0.05 | 0 | 2 | .0002 | 1 | 57 | 58 |
| 15 Germinal | −0.04 | −0.03 | 1 | 1 | .0003 | 27 | 15 | 42 |
| 16 Oeuvre | −0.15 | 0.03 | 8 | 1 | .0013 | 83 | 3 | 86 |
| 17 Terre | −0.12 | 0.05 | 8 | 3 | .0012 | 84 | 11 | 95 |
| 18 Rêve | −0.08 | −0.13 | 1 | 9 | .0006 | 24 | 71 | 95 |
| 19 Bête | −0.09 | −0.06 | 3 | 4 | .0007 | 63 | 29 | 91 |
| 20 Argent | −0.15 | −0.03 | 8 | 1 | .0013 | 85 | 4 | 89 |
| 21 Débacle | −0.14 | −0.05 | 10 | 4 | .0015 | 87 | 12 | 99 |
| 22 Pascal | −0.13 | −0.03 | 5 | 1 | .0008 | 88 | 5 | 94 |



| | $F_1$ | $F_2$ | $\text{ctr}_1$ | $\text{ctr}_2$ | $m \times d^2$ | $\cos_1^2$ | $\cos_2^2$ | Qual |
|---|---|---|---|---|---|---|---|---|
| . | 0.17 | −0.01 | 45 | 0 | .0064 | 95 | 0 | 96 |
| . . . | 0.02 | 0.17 | 0 | 15 | .0013 | 1 | 57 | 58 |
| !& ? | −0.17 | 0.19 | 12 | 41 | .0040 | 40 | 51 | 91 |
| ; | −0.07 | −0.04 | 23 | 19 | .0041 | 76 | 23 | 100 |
| ; | 0.28 | −0.04 | 18 | 1 | .0033 | 73 | 1 | 74 |
| – | 0.02 | 0.06 | 0 | 1 | .0003 | 3 | 15 | 18 |
| — | 0.06 | 0.11 | 2 | 22 | .0018 | 18 | 62 | 80 |

Table 12: Zola's punctuation. Columns: The Projections, Contributions, inertia to barycenter, and Cosines. Recoded Data: Italic and quote are supplementary elements, Exclamation and Interrogation are summed, Hyphen 1 and 2 are summed.



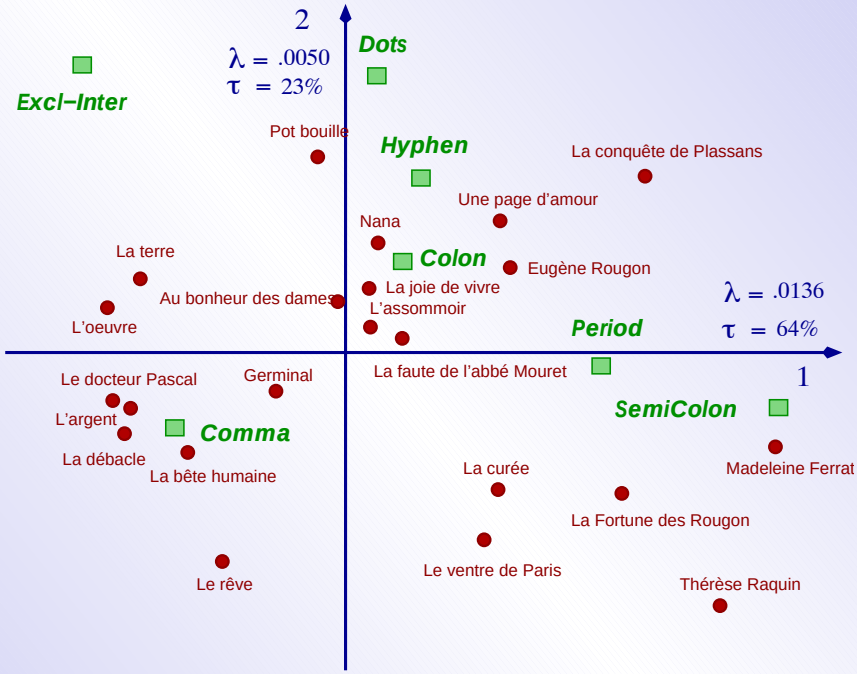



Figure 47: The punctuation of Zola. Second Pass. Rows and Columns.

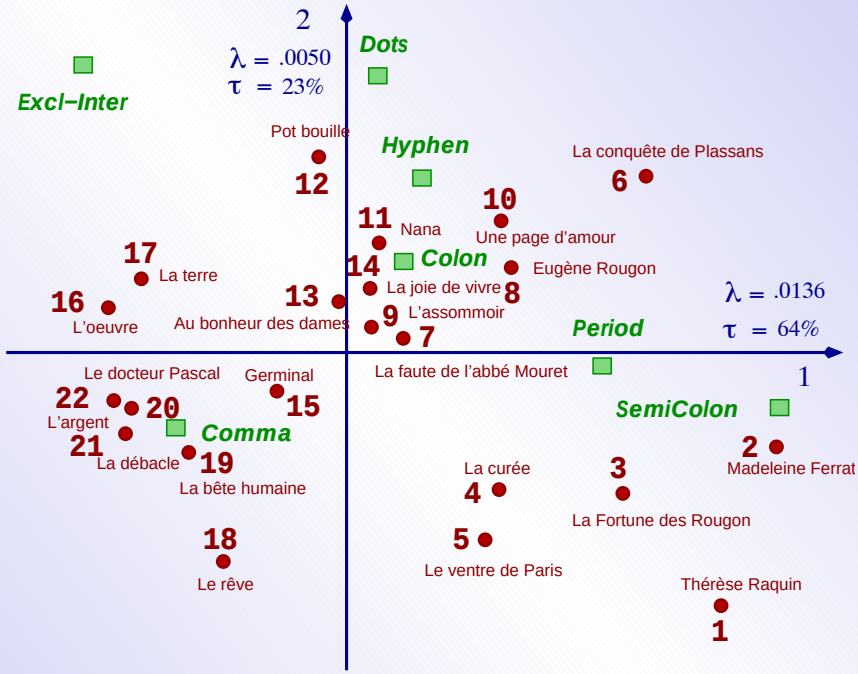



Figure 48: The punctuation of Zola. Second Pass. Rows.

# The Math you thought you had escaped!

**See Papers # C82 & #C69 for Details**


Refresher! and Menu

Example: . . .

Chi-Square Refresher: . . .

How to Analyze the . . .

Home Page

Title Page

◀◀ ▶▶

◀ ▶

Page 84 of 85

Go Back

Full Screen

Close

Quit

# With R!

Use the ExPosition package:
(the punctuation example comes with the package)

```
library('ExPosition')

data(authors)

# load the punctuation example

# to get the results of CA (with graphs):

ca.authors.res =  epCA(authors$ca$data)
```