

Introduction

A two-way **contingency table** is an $n \times m$ matrix that records the frequency of observations for each pair of categories from two categorical variables.

The **Contingency Wheel** is an interactive visual method for finding and analyzing associations in a large $n \times m$ table with $m < 100$ and n being 2 to 3 orders of magnitude larger than m .

Example: ca. 1 million ratings on 270,170 books by users in different countries

The Visual Metaphor

Columns \rightarrow Sectors

Cells \rightarrow Nodes

- angular coordinate by layout
- radial coordinate from $r_{i,j}$ the strength of association between row i and column j

$$r_{i,j} = \begin{cases} \frac{f_{i,j} - \hat{e}_{i,j}}{f_{i+} - \hat{e}_{i,j}} & f_{i,j} \geq \hat{e}_{i,j} \\ \frac{f_{i,j} - \hat{e}_{i,j}}{\hat{e}_{i,j}} & \text{otherwise} \end{cases}, \quad \hat{e}_{i,j} = \frac{f_{i+} \cdot f_{+j}}{f_{++}}$$

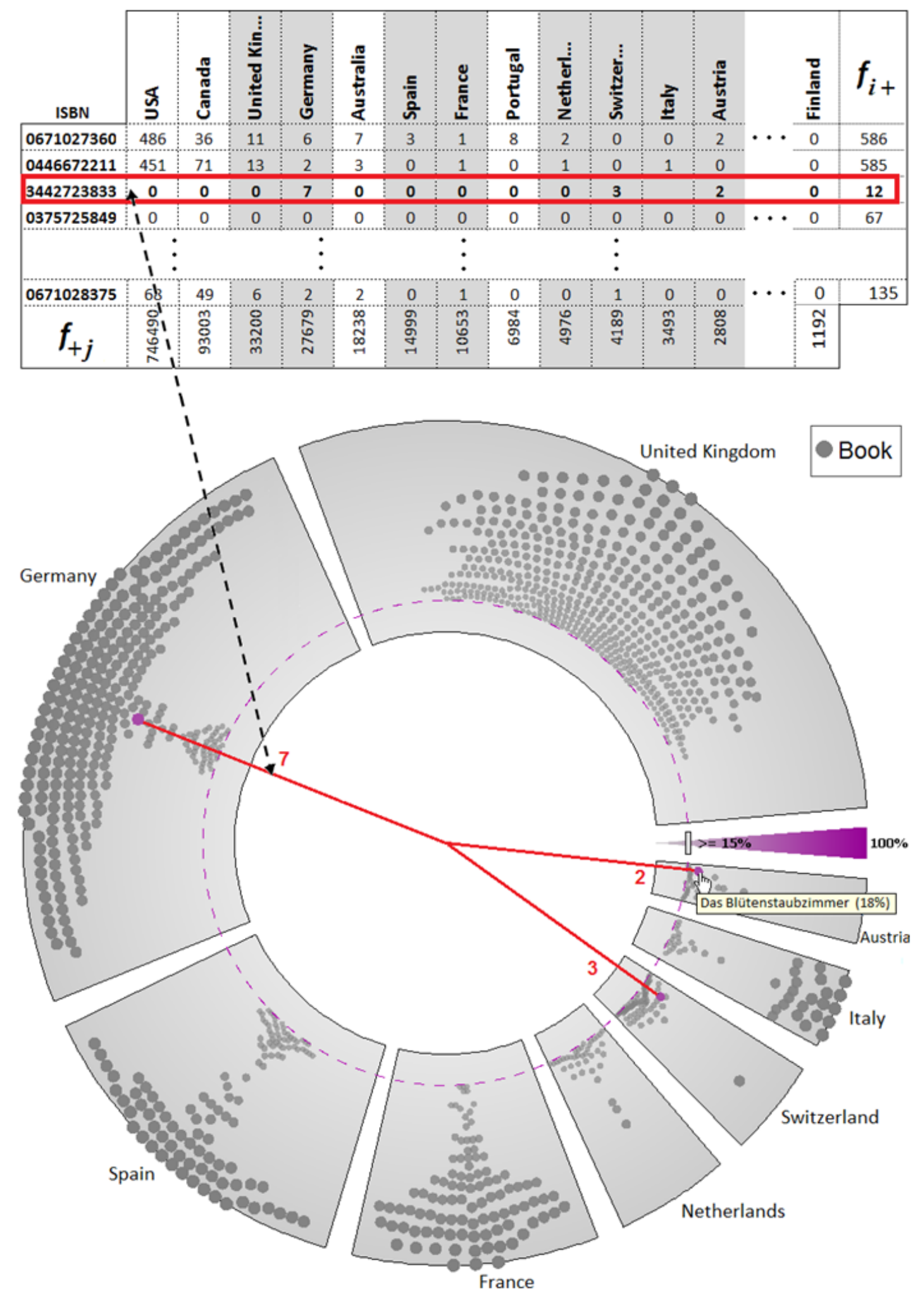
Thresholds

- $f_{i+} > T_s$ on row significance
- $r_{i,j} > T_r$ on association strength

Links

- column-column associations

$$rc_{j_1, j_2} = \frac{1}{f_{+j_1} + f_{+j_2}} \cdot \sum_{\substack{(0 < i \leq n) \wedge (T_s \leq f_{i+}) \\ \wedge T_r \leq r_{i,j_1} \wedge T_r \leq r_{i,j_2}}} (r_{i,j_1} + r_{i,j_2})$$



Conclusion

The Contingency Wheel enables analyzing and gaining insight into large tables (up to 500,000 x 100)

Effective visual and interaction metaphors for discovering and analyzing associations

Linked views effective in brushing and filtering data

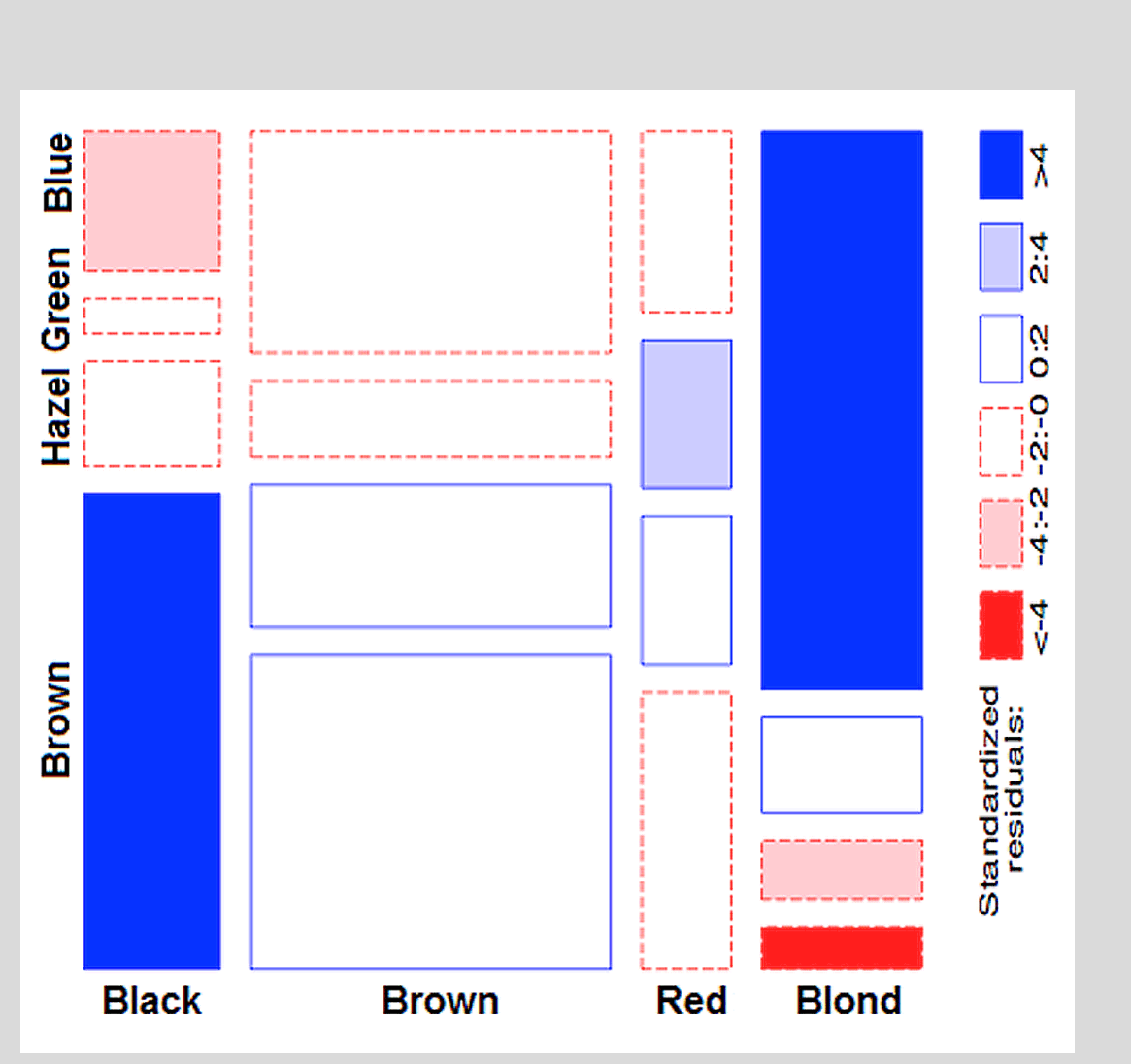
Future work

- Exploring different algorithms for node placement
- Using different association measures
- Hierarchical Clustering

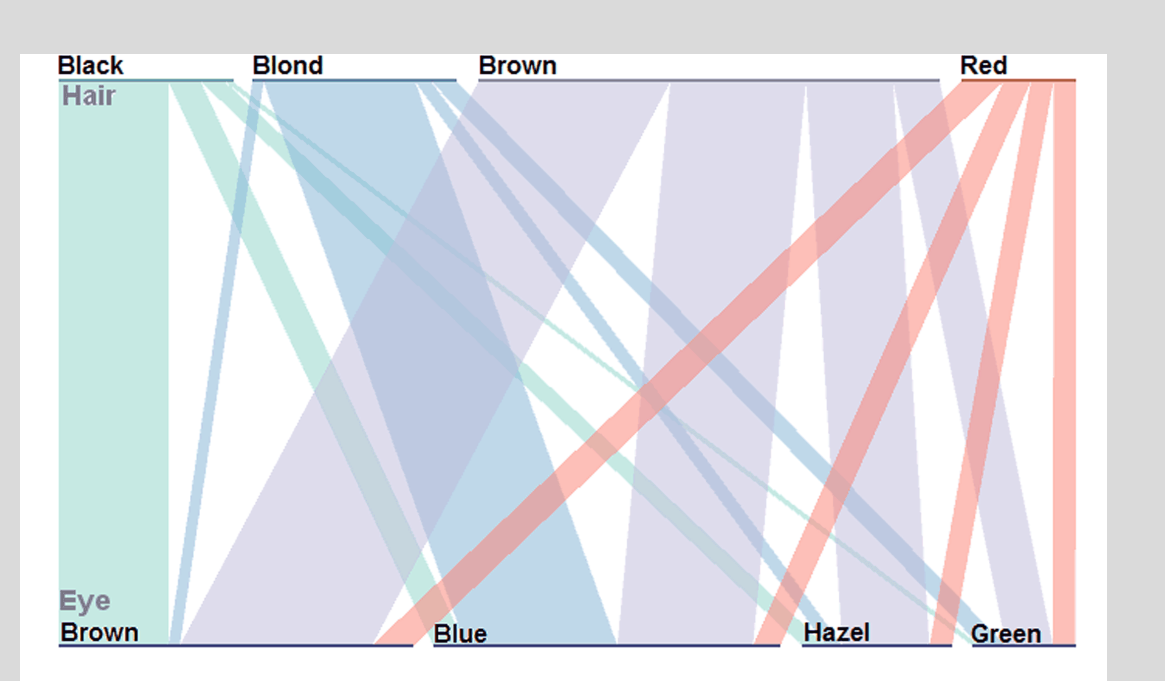
Related Work

| Eye \ Hair | Black | Brown | Red | Blond | Total |
|------------|-------|-------|-----|-------|-------|
| Brown | 68 | 119 | 29 | 7 | 220 |
| Blue | 20 | 84 | 17 | 94 | 215 |
| Hazel | 15 | 54 | 14 | 10 | 93 |
| Green | 5 | 29 | 14 | 14 | 16 |
| Total | 108 | 286 | 71 | 127 | 592 |

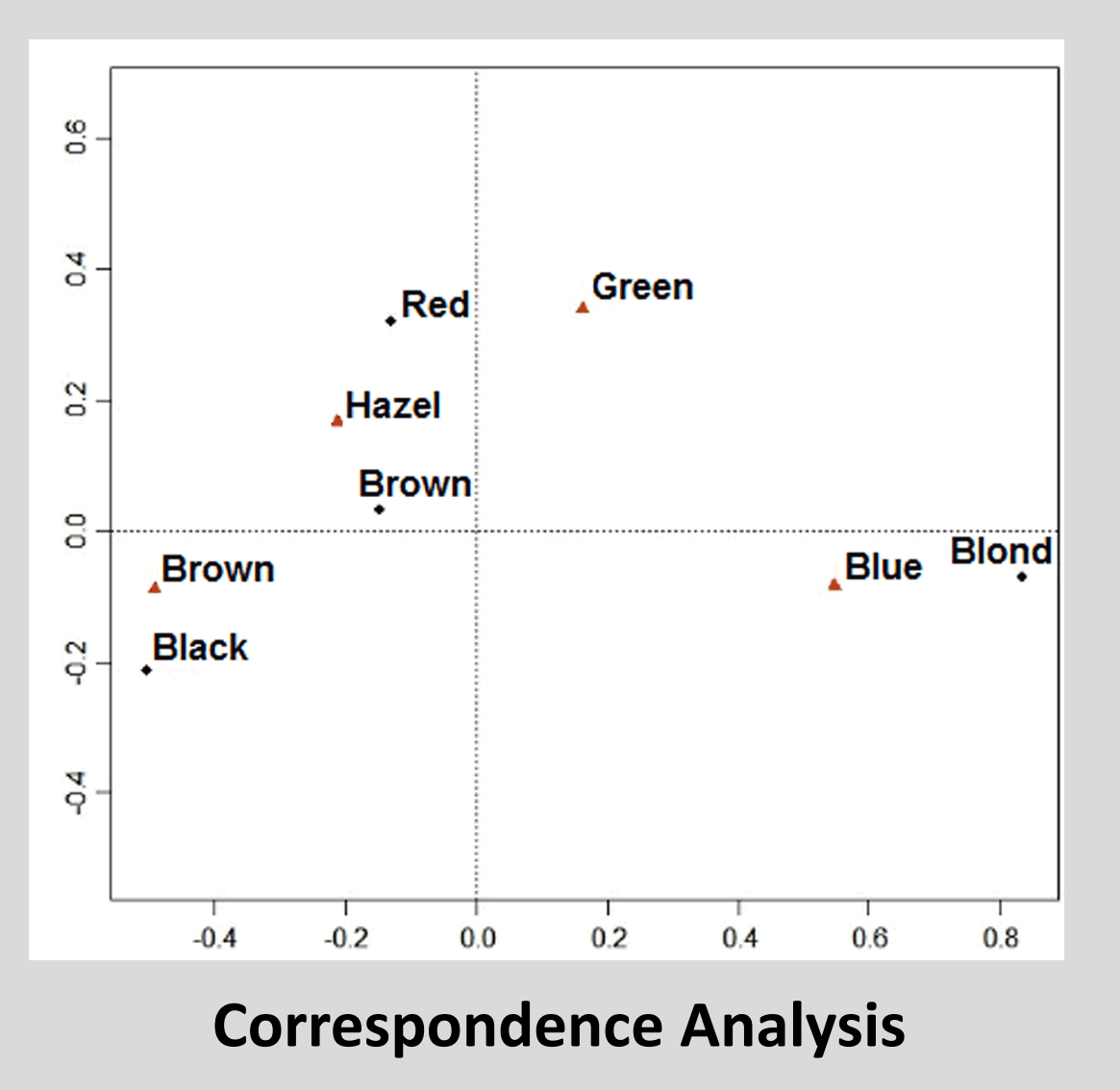
Snee's hair-and-eye-color dataset



Parallel Sets

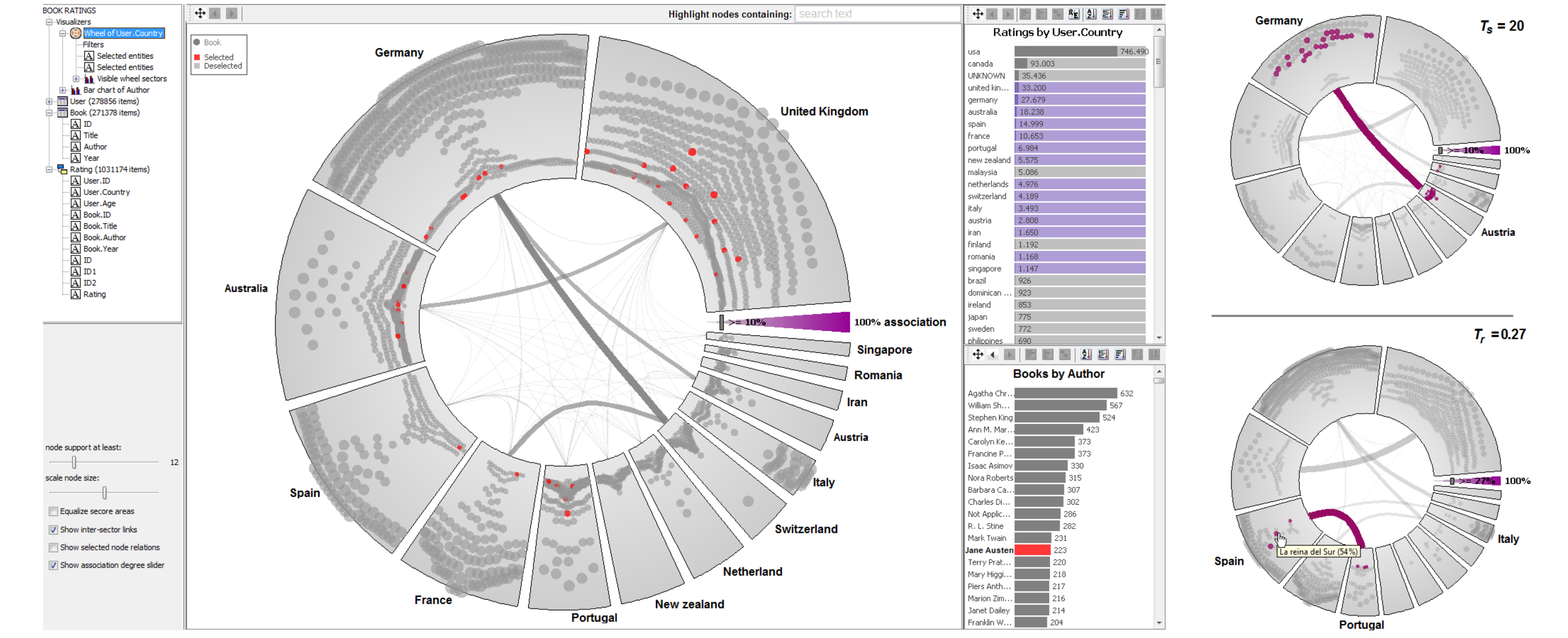


Correspondence Analysis

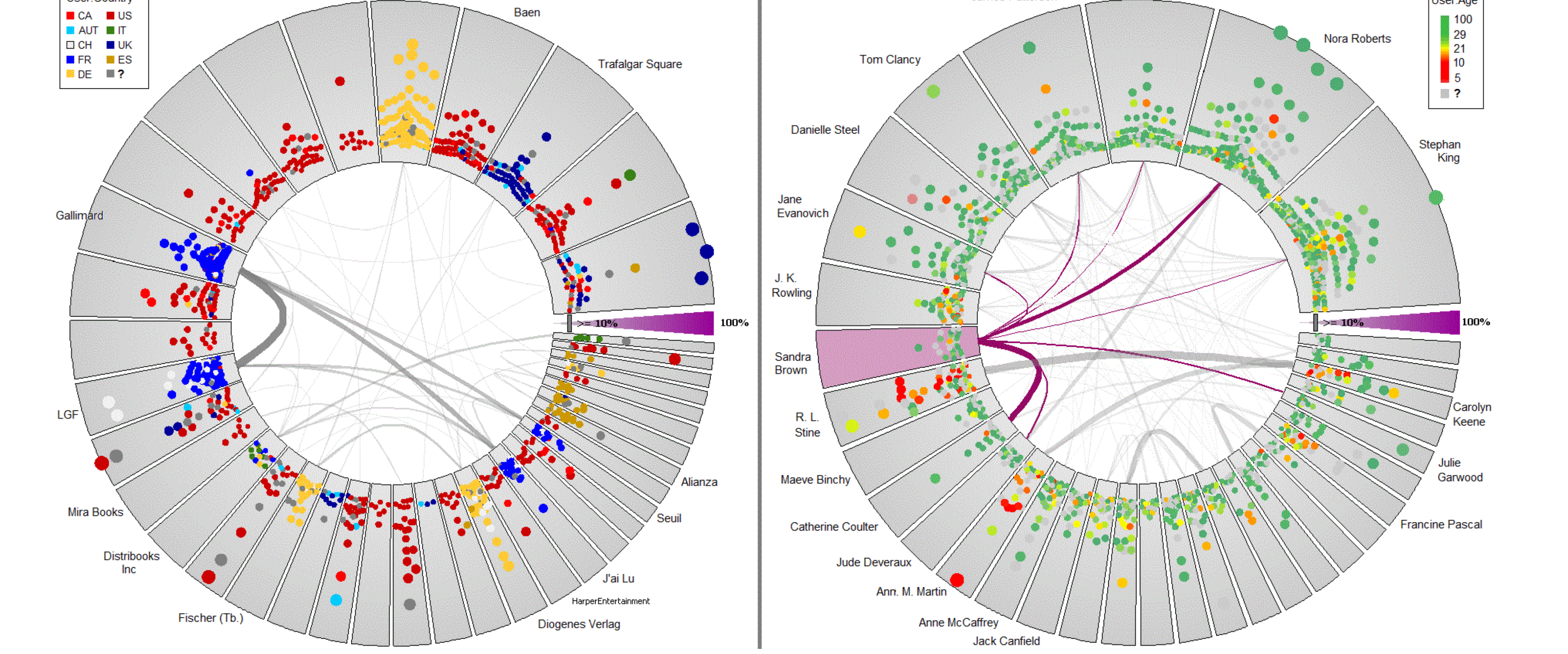


User Interaction

Selecting visible sectors from the bar chart
 Mapping attributes from linked views
 Selecting nodes/links
 Assigning thresholds (T_s and T_r)



Using Color for Finding Patterns



User ratings broken down by publisher. The nodes represent users and are colored by country. User ratings broken down by author. The nodes represent users and are colored by age.