情報可視化

LX6: 表の配置

数理・計算科学系

脇田 建

Arrange Tables

Express Values

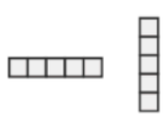
Separate, Order, Align Regions

- Separate

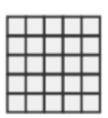
- Order

- Align

- *1 Key*
  List



- *2 Keys*
  Matrix



- *3 Keys*
  Volume

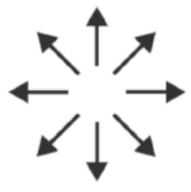
- *Many Keys*
  Recursive Subdivision

Axis Orientation

- Rectilinear

- Parallel

- Radial



Layout Density

- Dense

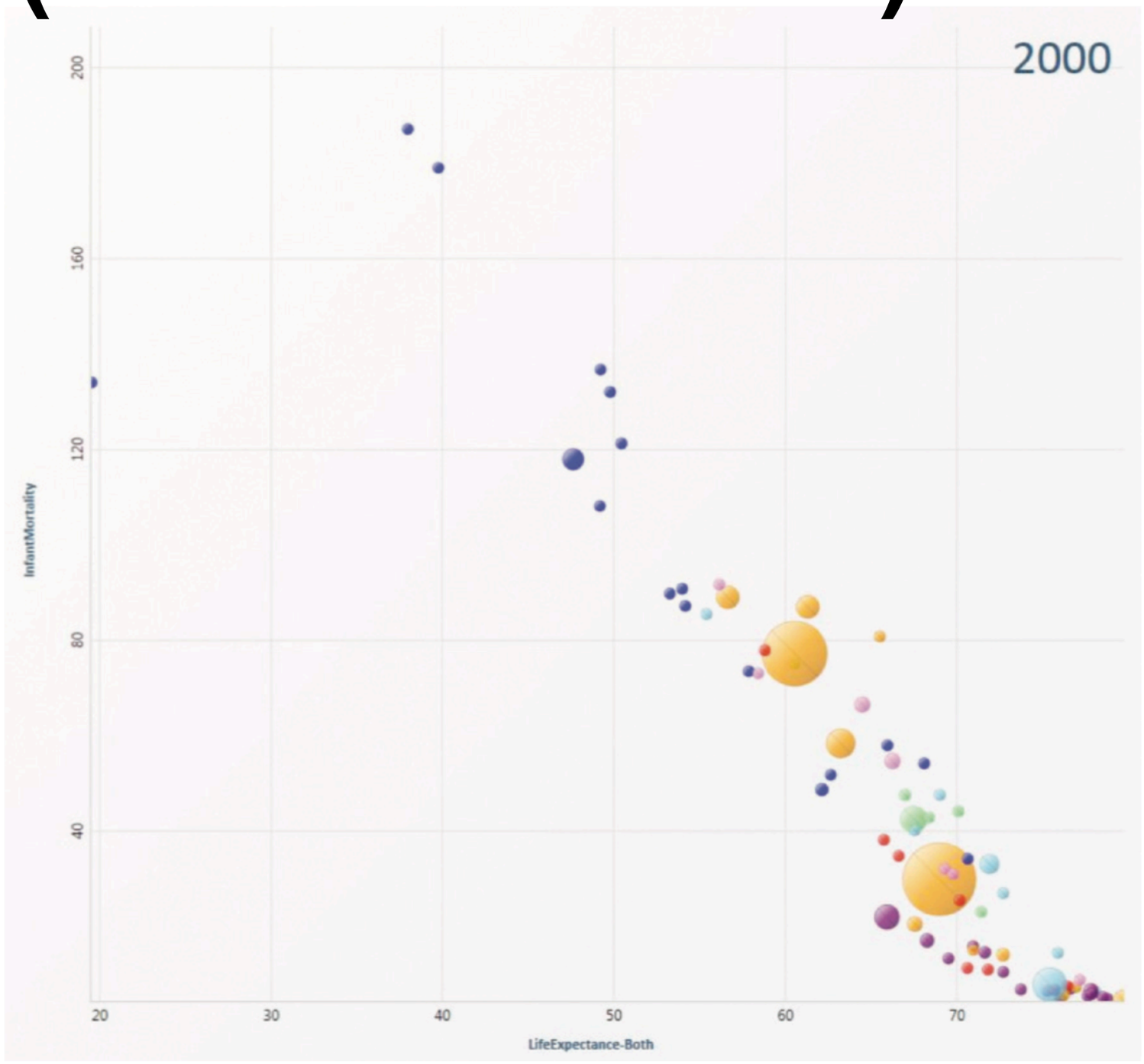
- Space-Filling

Figure 7.1. Design choices for arranging tables.

量的データの表現

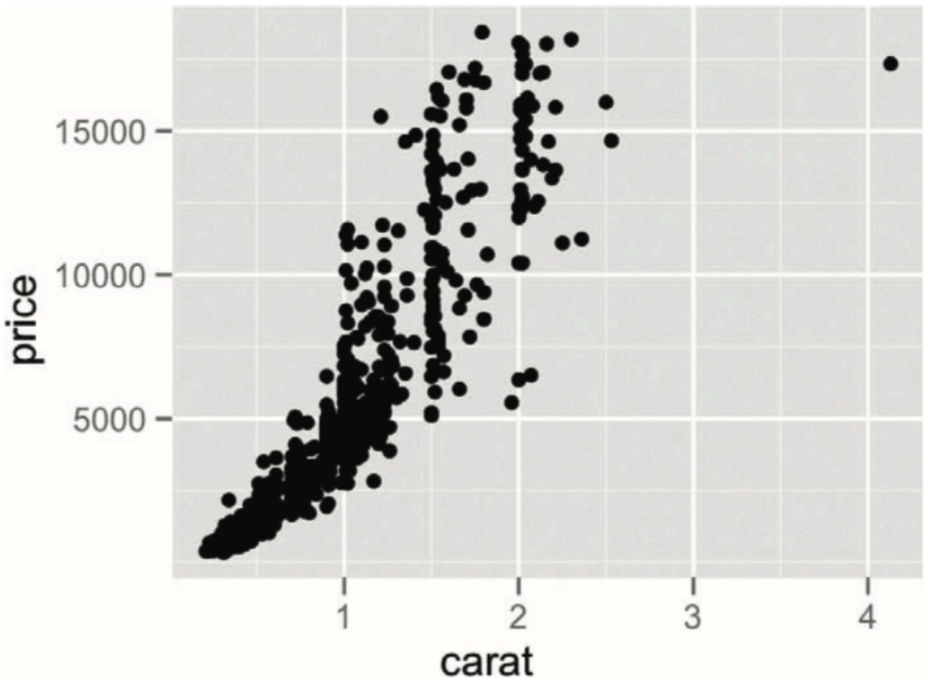
- 基本は点で表現: 位置チャネルを活用（最も強力だから）
    - 散布図: ほかのチャネル
        - Bubble chart: 色、大きさも活用した散布図
        - 複合グリフ (glyph) の活用: さらに複雑な場合
            - 散布図の点の位置に円グラフを配置

Bubble Chart (Robertson+08)

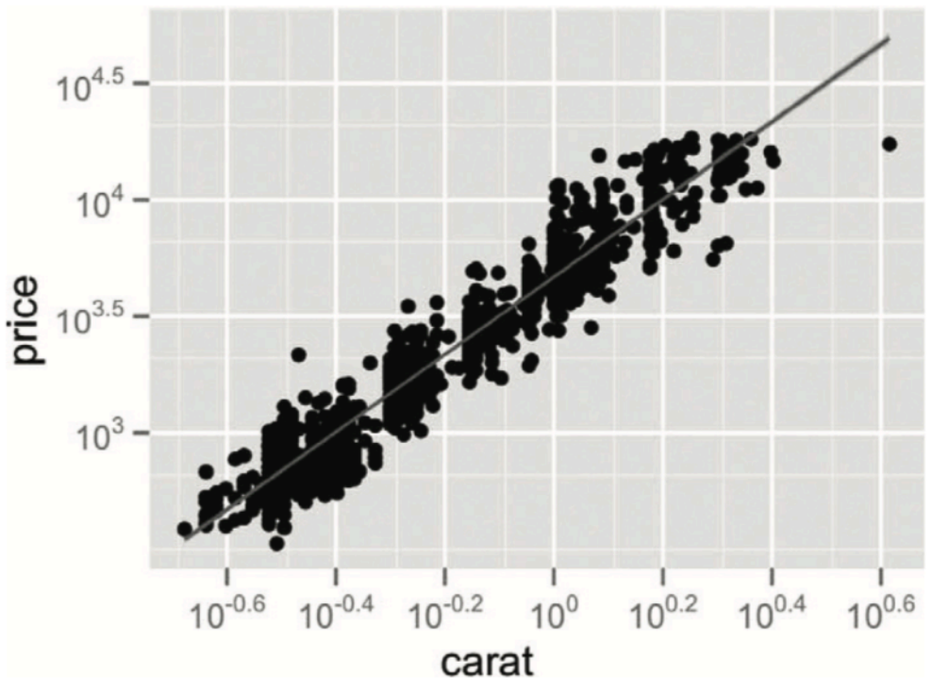


$$Q \times Q \times C \times Q$$

派生属性からのパターンの発見



(a)



(b)

Figure 7.3. Scatterplots. (a) Original diamond price/carat data. (b) Derived log-scale attributes are highly positively correlated. From [Wickham 10, Figure 10].

散布図の特徴

- What (data): 2つの量的データから構成される表
- How (encode): 縦・横方向の空間的位置と点
- Why (task): 傾向分析、分布、相関、クラスタ発見
- Scale: データ項目数 〜 数百

カテゴリーデータの表現

- 画面を領域に区分 (separate)
- 区分した領域に順序を導入 (order)
- データ項目の整列 (align)

棒グラフにおける順序

タスクに応じて順序を選択

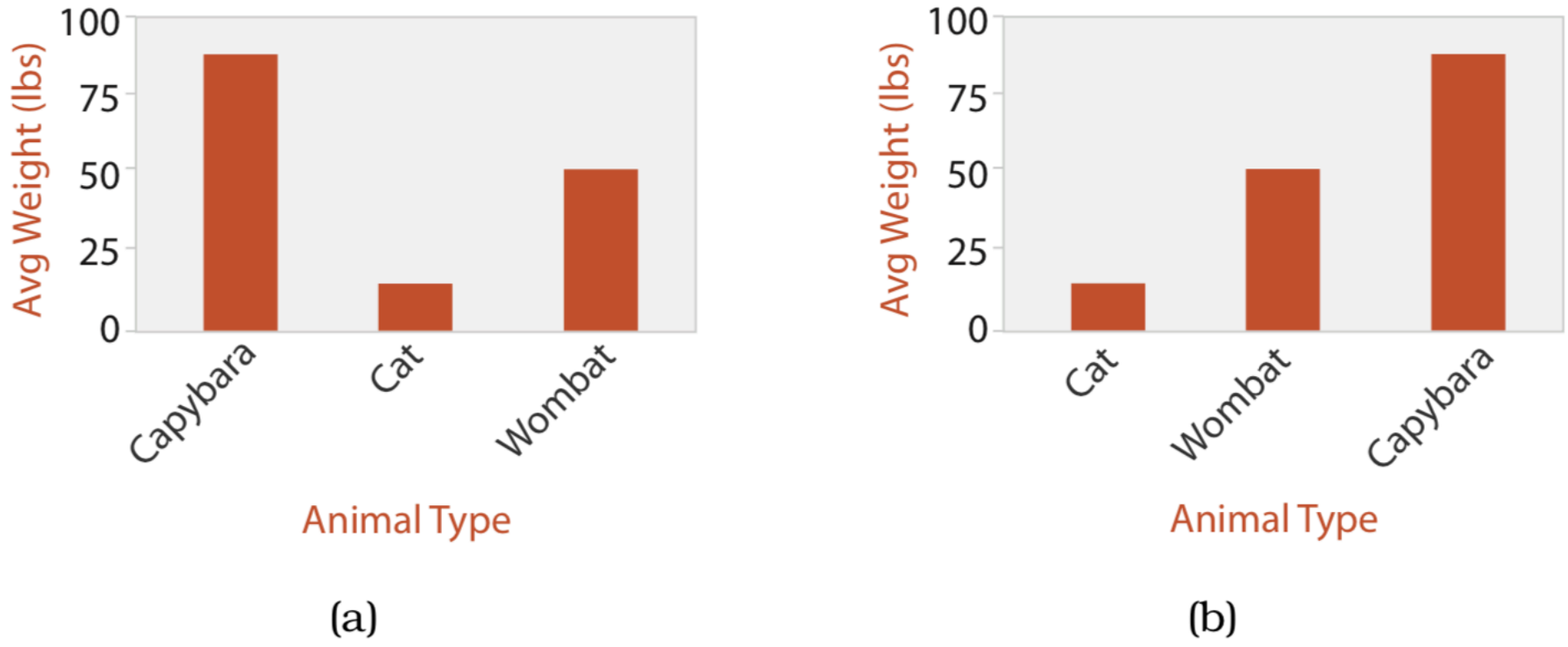


Figure 7.4. Bar chart. The key attribute, *species*, separates the marks along the horizontal spatial axis. The value attribute, *weight*, expresses the value with aligned vertical spatial position and line marks. (a) Marks ordered alphabetically according to species name. (b) Marks ordered by the weight attribute used for bar heights.

棒グラフの特徴

- What (data): キー：カテゴリデータ、値：量的データ
- How (encode): 線マーク、基準線からの位置、基準線に沿ってキーを間隔を置いて配置
- Why (task): キーの検索、値の比較
- Scale: キー数が数十〜数百

積み上げ棒グラフ

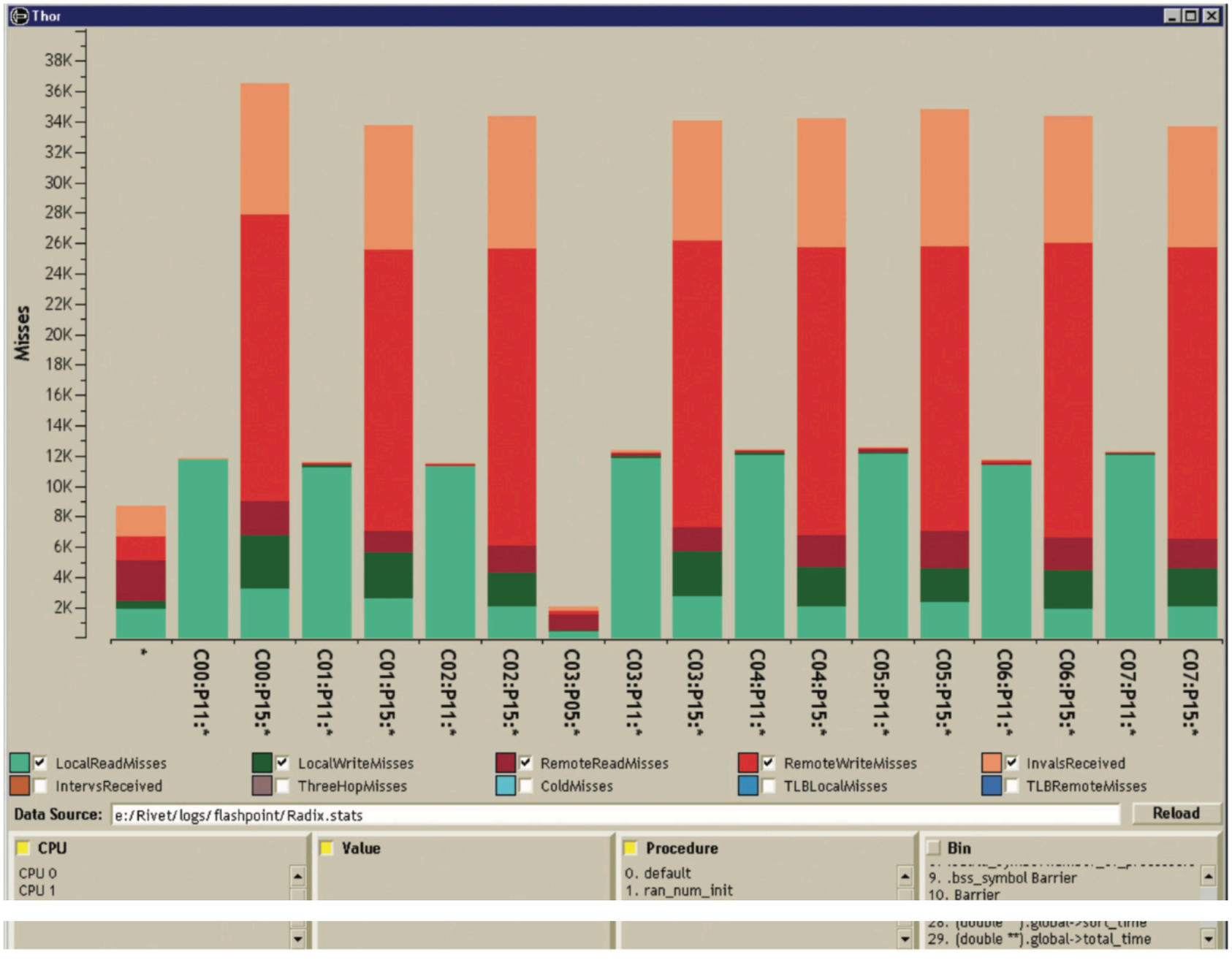


Figure 7.5. Stacked bar chart. The Thor memory profiler shows cache misses stacked and colored by miss type. From [Bosch 01, Figure 4.1].

多次元データ

色の利用は本質的

棒を積み上げる順序はとても大切

積み上げ棒グラフの特徴

- Stacked bar chart
- What (data): 多次元データ：（カテゴリーキー × 2）×（量的データ）
- How (encode): 棒形状に長さに二次キー種ごとの値を埋め込んだ部分構造、主キー種は基準線に沿って間隔を置いて配置
- Why (task): 全体に占める割合、値の検索、傾向分析
- Scale: 主キー < 数百、二次キー < 10程度

Streamgraph

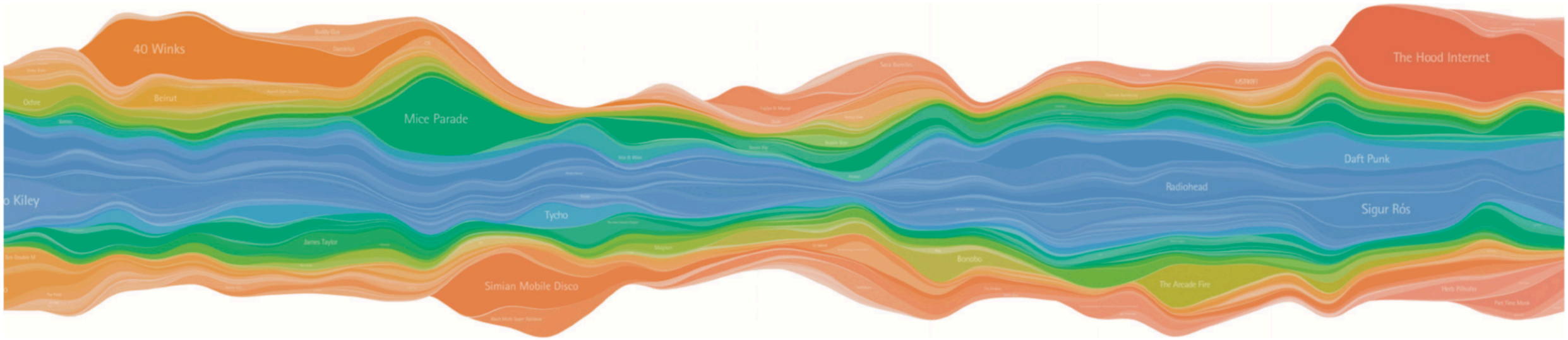
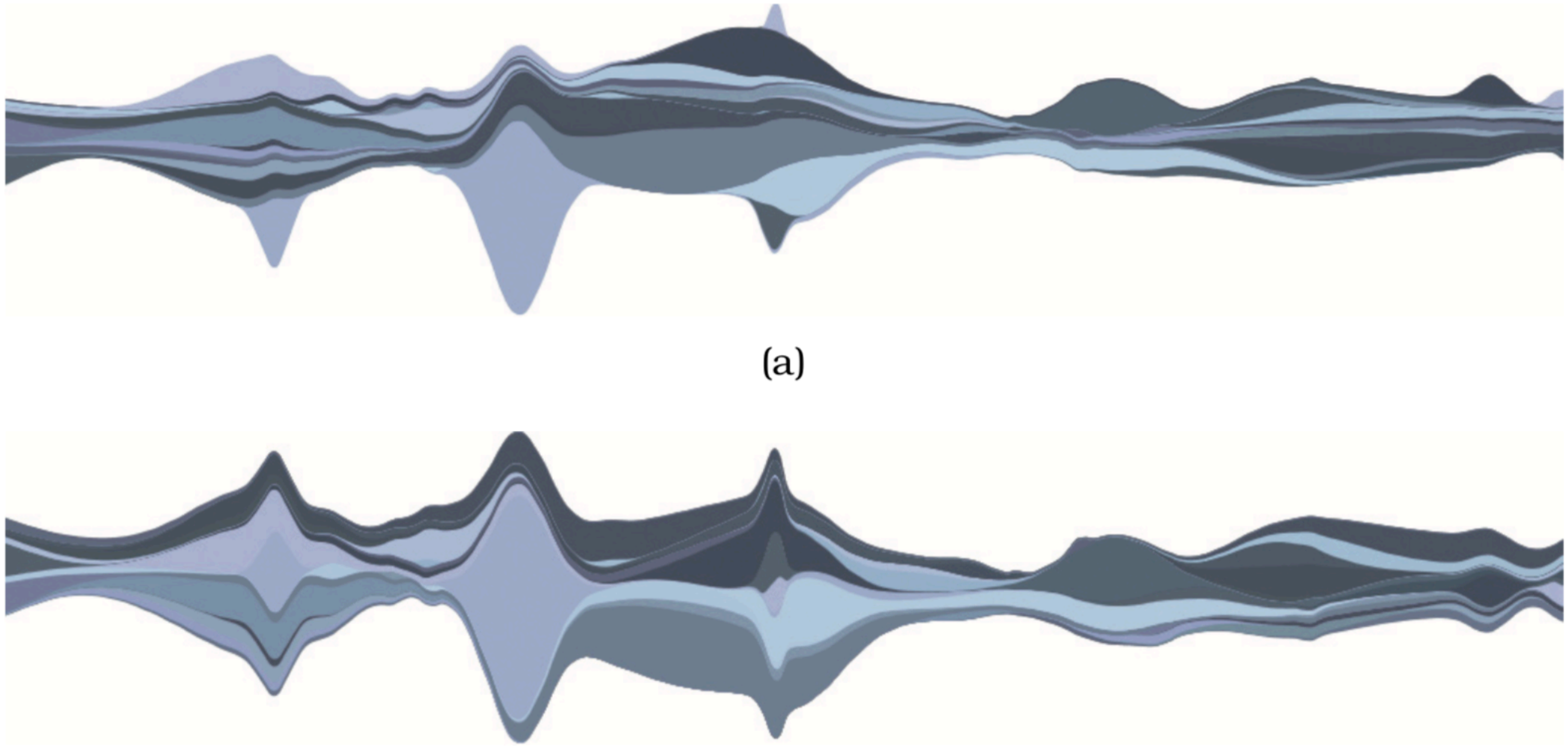


Figure 7.6. Streamgraph of music listening history. From [Byron and Wattenberg 08, Figure 0].

オリコンランキング的なデータの時系列可視化

積み上げ順序の違い



(a)

(b)

Figure 7.7. Streamgraphs with layers ordered by different derived attributes. (a) Volatility of artist's popularity. (b) Onset time when artist's music of first gained attention. From [Byron and Wattenberg 08, Figure 15].

Streamgraphの特徴

- What (data): 多次元データ、順序キー (時間)、カテゴリーキー(種別)、量的データ
- What (derived): 層の順序を定めるための量的派生データ
- How (encode): 派生データに沿ってカテゴリーキーを並べ替え、層の高さで量的データを表現
- Scale: 時間キー: 数百、カテゴリーキー: 数百

誤った表現が生む誤解

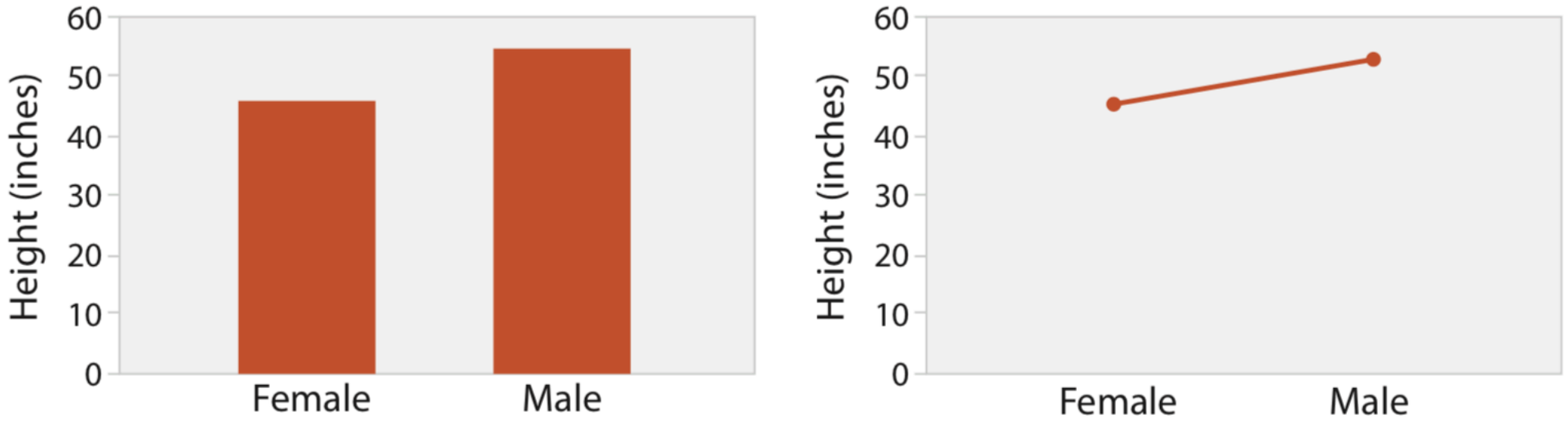
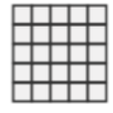


Figure 7.9. Bar charts and line charts both encode a single attribute. Bar charts encourage discrete comparisons, while line graphs encourage trend assessments. Line charts should not be used for categorical data, as in the upper right, because their implications are misleading. After [Zacks and Tversky 99, Figure 2].

2つのキー

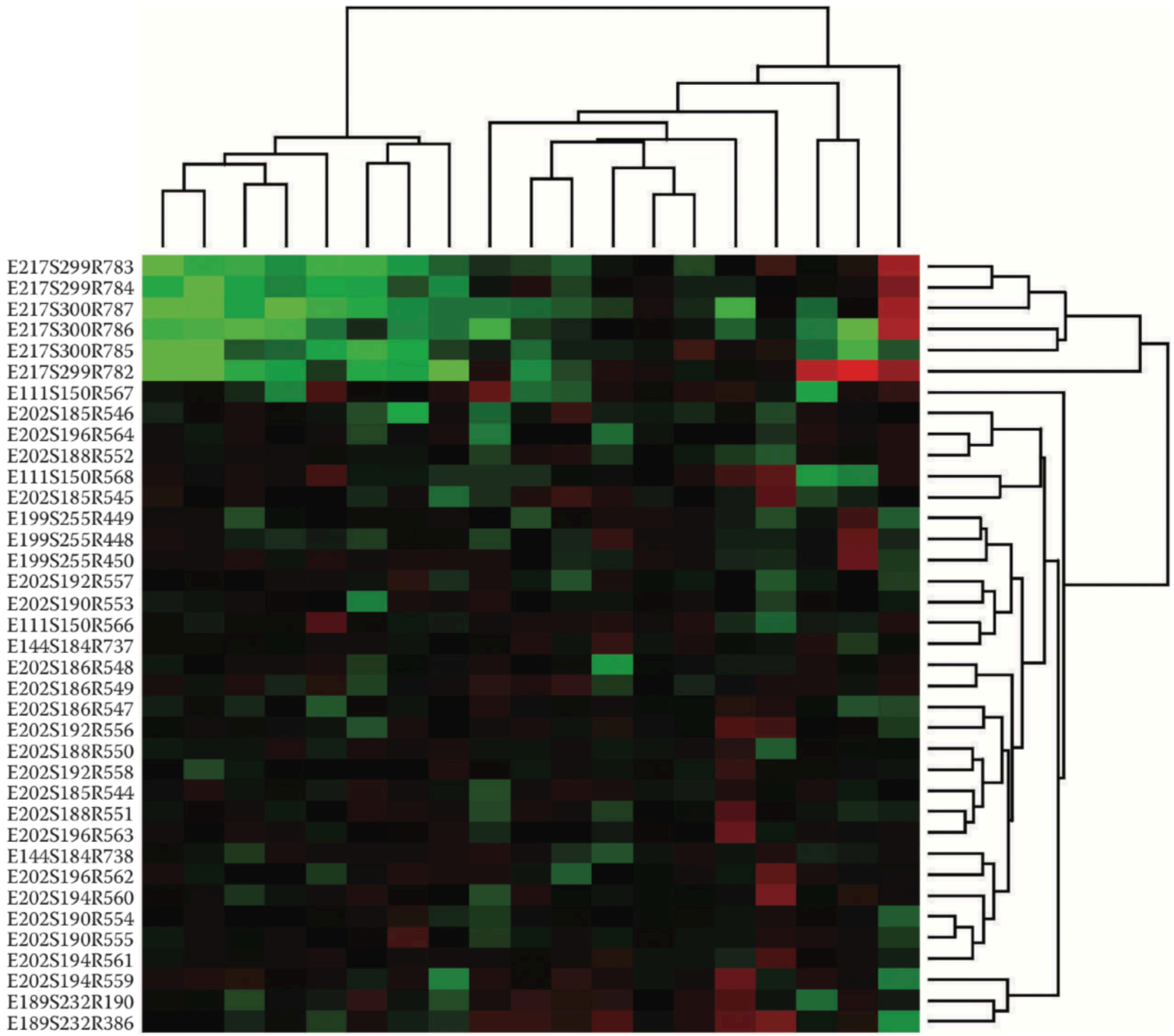
→ *2 Keys*
Matrix



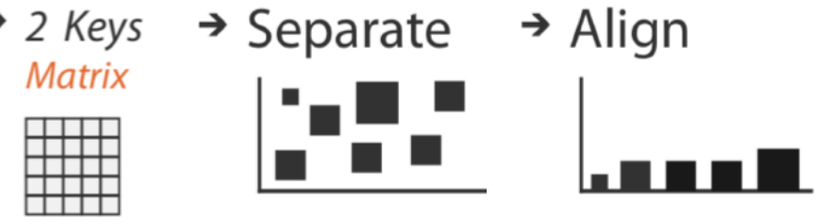
- 例：Microarrayセンサー（生命実験：遺伝子 x 実験条件 x 活性）

- 行列表現

Cluster Heatmap



Heatmapの特徴



- What (data): カテゴリーキー × 2(遺伝子、実験条件)、量的データ (活性)
- How (encode): 表 (領域マーク)、双方向カラーマップ
- Why (task): クラスター発見、外れ値、概要
- Scale: 値の数 < 100万、キーの数 < 数百、値のレベル: 3-11

Cluster heatmap の特徴

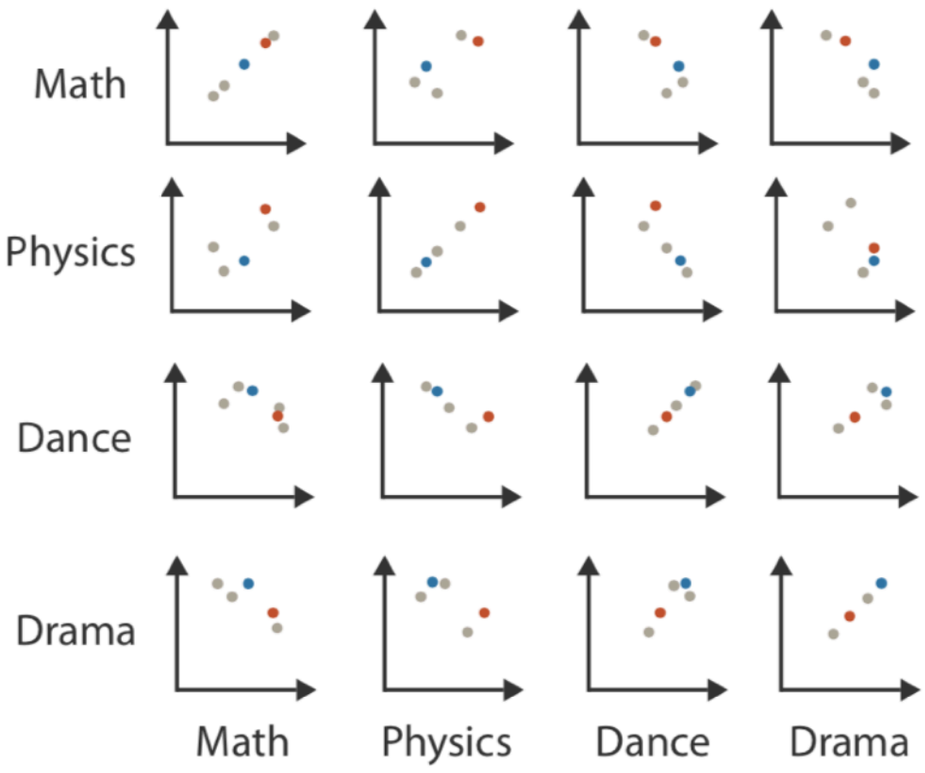
- Heatmapの特徴に加えて
- What (derived): ふたつのクラスター階層
- How (encode): 階層構造の順序にキーを配置

多数のキー

Table

Math	Physics	Dance	Drama
85	95	70	65
90	80	60	50
65	50	90	90
50	40	95	80
40	60	80	90

Scatterplot Matrix



Parallel Coordinates

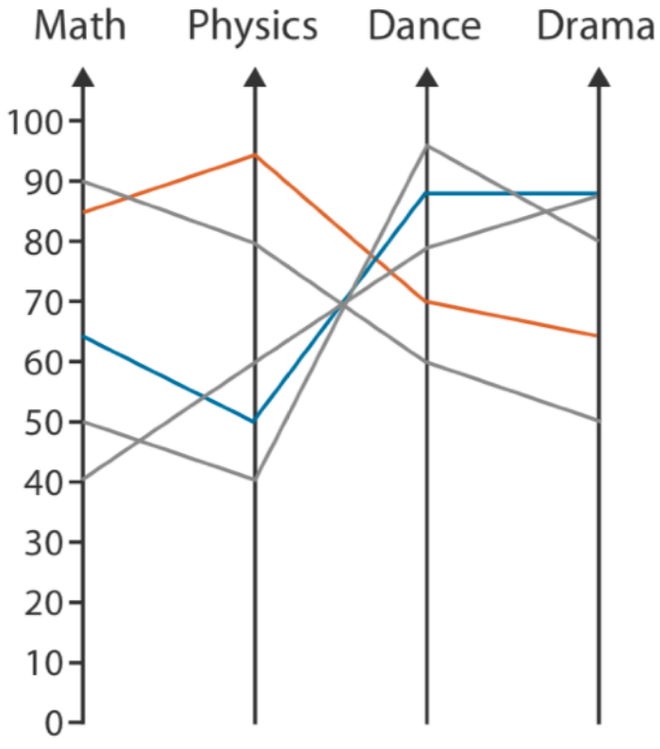


Figure 7.12. Comparison of scatterplot matrix and parallel coordinate idioms for a small data table. After [McGuffin 14].

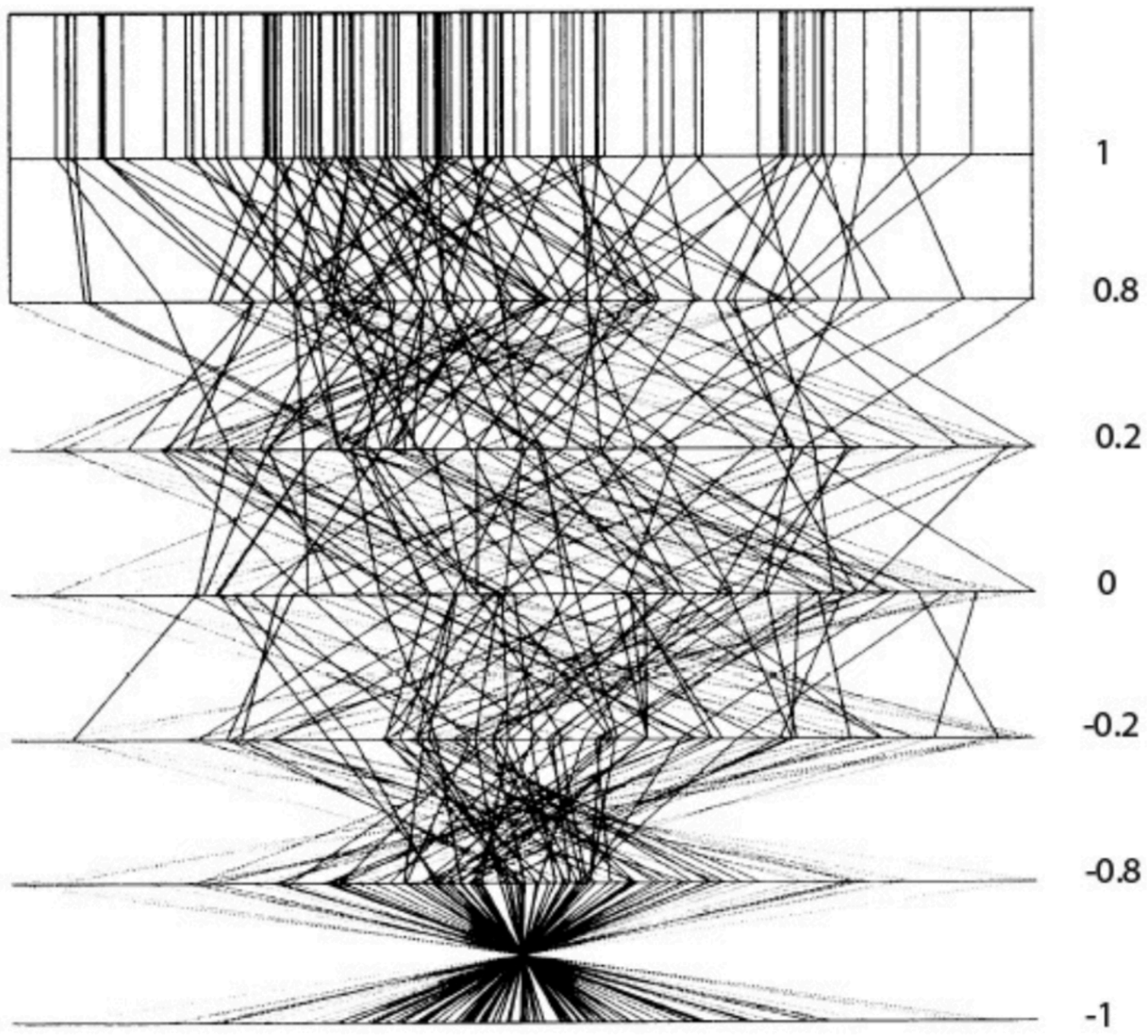
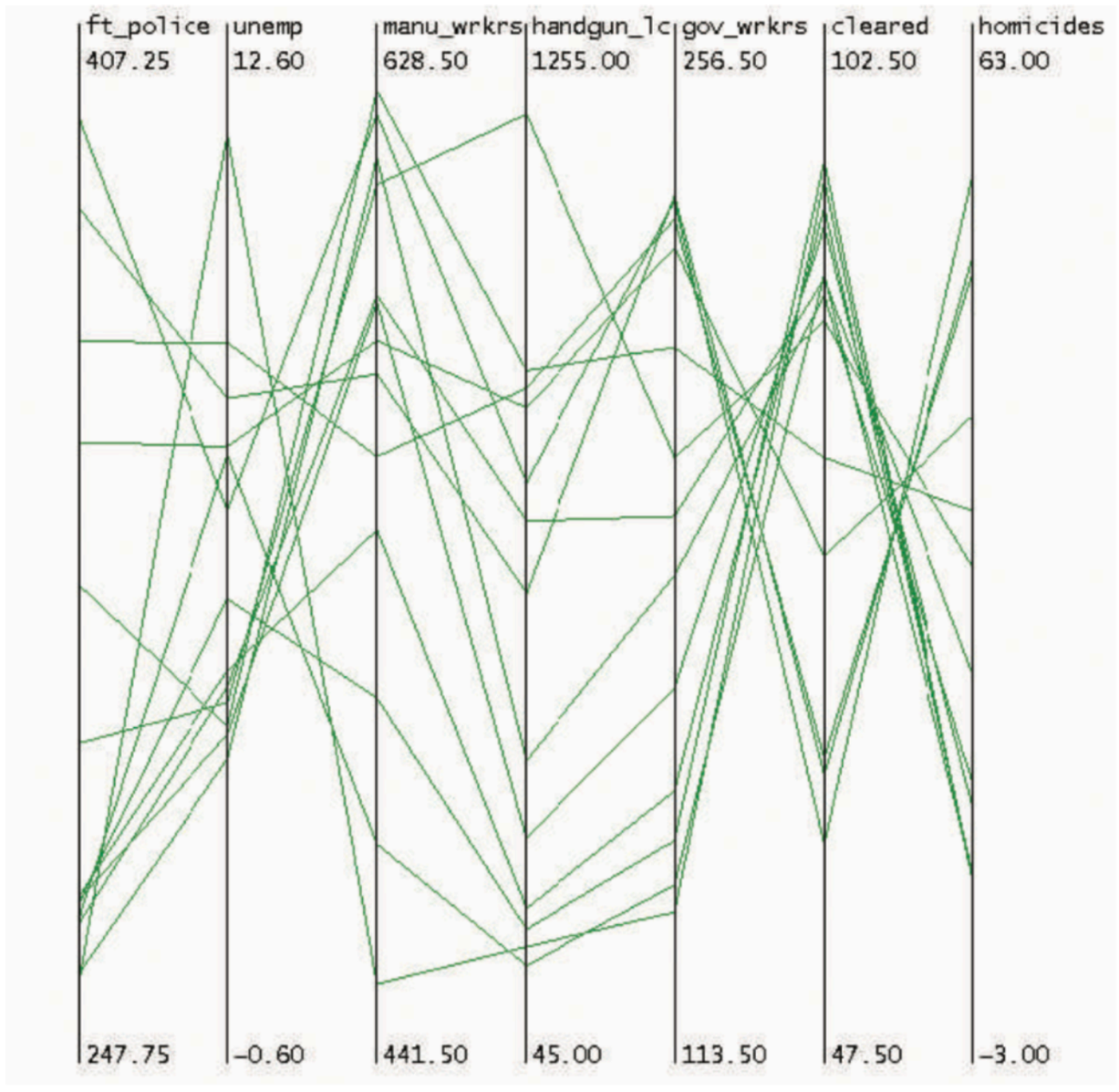
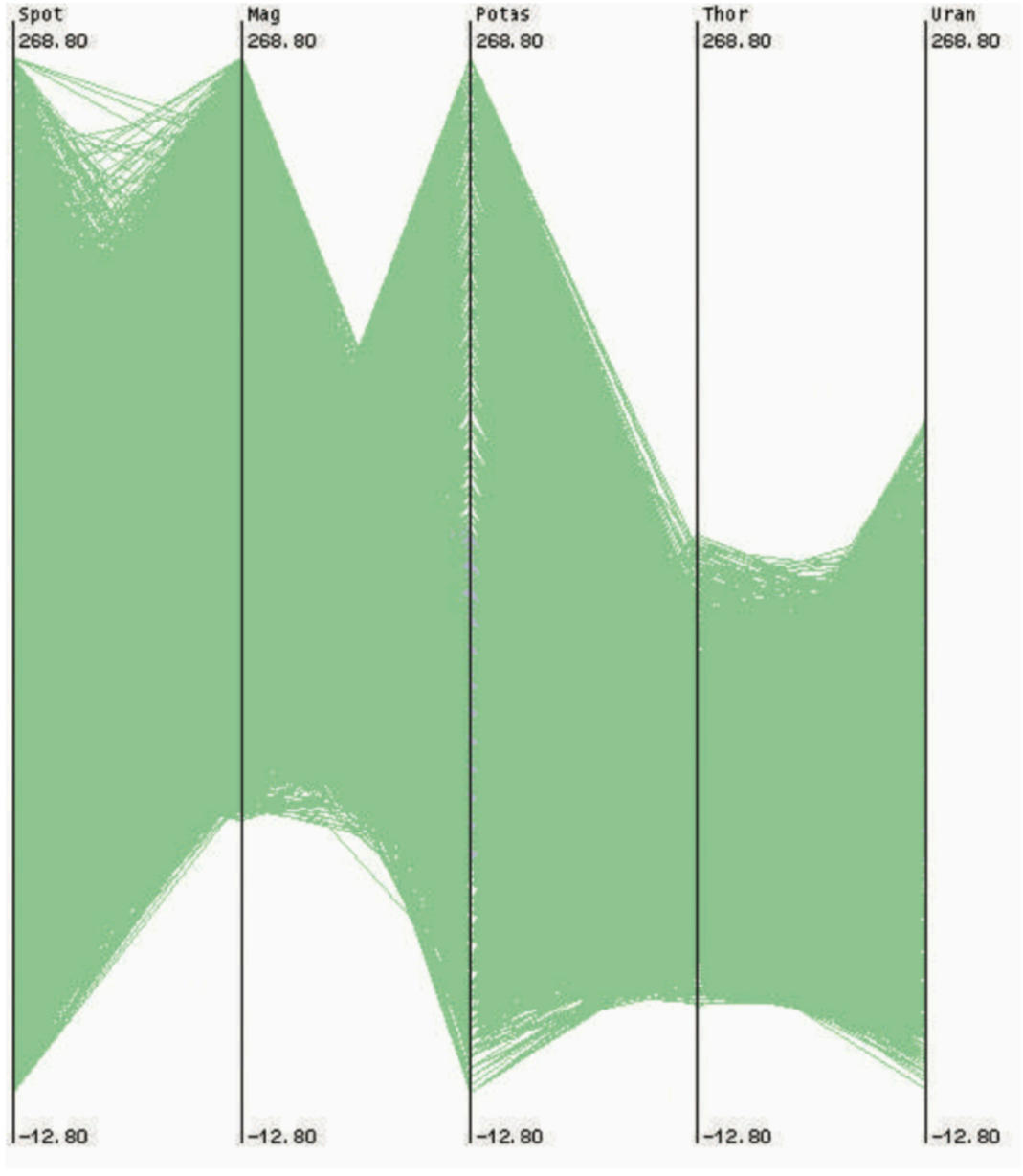


Figure 7.13. Parallel coordinates were designed to show correlation between neighboring axes. At the top, parallel lines show perfect positive correlation. At the bottom, all of the lines cross over each other at a single spot in between the two axes, showing perfect negative correlation. In the middle, the mix of crossings shows uncorrelated data. From [Wegman 90, Figure 3].



(a)



(b)

Figure 7.14. Parallel coordinates scale to dozens of attributes and hundreds of items, but not to thousands of items. (a) Effective use with 13 items and 7 attributes. (b) Ineffective use with over 16,000 items and 5 attributes. From [Fua et al. 99, Figures 1 and 2].