

## 第1章 データ可視化の本質

### 1.1 データを可視化すること ..... 14

- データを「見える」ようにする ..... 14
- 探索志向型データ可視化と説明志向型データ可視化 ..... 15
- 可視化が機能するとき ..... 16
- データ解釈は「ロジック」のあてはめ ..... 17
- 指標化によるロジックのあてはめ ..... 19
- 本書で扱わないこと ..... 22

### 1.2 可視化の効果を考える ..... 23

- 文字情報と視覚情報 ..... 23
- 数値把握が容易になる ..... 23
- パターンの発見が容易になる ..... 25
- 追加の情報と紐づける ..... 28

### 1.3 可視化で読み取れるロジック ..... 32

- 大小・遠近・多寡 ..... 32
- 変数間の関係性の存在 ..... 33
- 現象ダイナミクスの特徴 ..... 33
- 理論ロジックとの紐づけ ..... 35

### 第1章まとめ ..... 38

## 第2章 数量を把握するデータ可視化

### 2.1 数量と図形の大きさを紐づける ..... 40

- 二次元的な大きさ ..... 40
- 要素数が多い場合に全体像をとらえる ..... 41

### 2.2 大きさを比較する ..... 44

- 棒グラフについて考える ..... 44
- 棒グラフの機能と見せ方 ..... 45
- 必ずしも棒グラフにこだわる必要はない ..... 46

### 2.3 標本を視えるようにする ..... 49

- グループ同士の比較を行なう ..... 49
- データのまとまりを描画する方法 ..... 50

### 第2章まとめ ..... 55

## 第3章 メカニズムをとらえるデータ可視化

### 3.1 分布の特徴をとらえる ..... 58

- 分布の特徴・着目するポイント ..... 58
- 理論分布と比較する ..... 60
- 累積分布で当てはめる ..... 65

### 3.2 線で特徴をとらえる ..... 67

- 折れ線・エリアチャート ..... 67
- 沢山のものを比べる ..... 68
- スロープグラフ ..... 70
- 変化のパターンを考える ..... 71

### 3.3 2変数の関係をとらえる ..... 74

- 散布図・ペアプロット ..... 74
- 相関を見るフィッティング ..... 76
- 特定の関係性からの散らばり具合を見る ..... 78
- 情報の多いデータを散布図にする ..... 80

### 第3章まとめ ..... 84

## 第4章 多変数をとらえるデータ可視化

### 4.1 三つ以上の変数の可視化 ..... 86

- レーダーチャートは限定的な場面で使える ..... 86
- パラレルプロットで特徴をとらえる ..... 87
- ヒートマップ ..... 90

### 4.2 ネットワークをとらえる ..... 95

- ネットワーク・グラフ ..... 95
- 有向ネットワークと階層構造 ..... 98
- ネットワーク可視化のポイント ..... 100

### 4.3 「まとめる」可視化 ..... 106

- クラスタリング ..... 106
- 次元圧縮 ..... 110
- 時系列でクラスタリングを使う ..... 116

### 第4章まとめ ..... 119

## 第5章 データの分布をとらえる指標化

### 5.1 分布と統計量 ..... 122

- 統計量について考える ..... 122
- 統計量と分布の相性 ..... 124
- 統計量に残る情報・残らない情報 ..... 126

### 5.2 ばらつきをとらえる ..... 128

- 単位当たりの指標と規格化 ..... 128
- データのばらつきをとらえる指標 ..... 132
- ばらつきの度合いを比べる ..... 134
- 尺度と利用可能な指標 ..... 135

### 5.3 分布の形をとらえる ..... 139

- 歪みの度合いを測る ..... 139
- 情報エントロピー指標 ..... 141
- 不平等の指標 ..... 143

### 第5章まとめ ..... 148

## 第6章 関係性をとらえる指標化

### 6.1 「近いか遠いか」をとらえる ..... 150

- 距離を測る指標 ..... 150
- マハラビノス距離で集団からの逸脱度を測る ..... 151
- 配列の間の距離を測る ..... 153
- ネットワーク上における距離 ..... 154

## 6.2 分布同士の距離を測る ..... 156

- コルモゴロフ・スミルノフ統計量 ..... 156
- 全変動距離とワッサースタイン距離 ..... 157
- 情報量で分布の距離を見る (KL/JS 情報量) ..... 159
- 分布間距離の指標を比較する ..... 160

## 6.3 ペアになった分布間の距離指標 ..... 162

- 相関係数で類似度を測る ..... 162
- フィッシャー変換による相関係数の正規化 ..... 164
- 偏相関係数で実効的な相関を測る ..... 165
- コサイン類似度 ..... 167
- 誤差を測る指標 ..... 169

## 6.4 「つながり」をとらえる ..... 172

- 相互情報量で見る変数間のつながり ..... 172
- 移動エントロピー ..... 174
- つながりからネットワークへ ..... 177

第6章まとめ ..... 179

# 第7章 パターンをとらえる指標化

## 7.1 時間的なパターンをとらえる ..... 182

- 離散変数の時系列パターンの定量化 ..... 182
- 自己相関 ..... 184
- 周波数特性を測る ..... 186
- 指標としての応用：心拍データ分析 ..... 189

## 7.2 空間データのパターンをとらえる ..... 191

- ポロノイ分割による空間の定量化 ..... 191
- カーネル密度推定 ..... 193
- 画像のパターンを定量化する：グレイレベル共起行列 ..... 194

## 7.3 ネットワークのパターンをとらえる ..... 198

- リンクやノードを特徴付ける指標 ..... 198
- ネットワーク全体を特徴付ける指標 ..... 200
- ネットワーク指標の利用 ..... 202

第7章まとめ ..... 204

# 第8章 データ指標化・可視化のプロセス

## 8.1 効果的な可視化のテクニック ..... 206

- 図の見やすさを向上させるテクニック ..... 206
- 差を視るための図の拡大・縮小 ..... 208
- 情報を「探させない」可視化 ..... 210
- 何の可視化なのかが伝わるようにする ..... 214

## 8.2 指標化から可視化の戦略を考える ..... 217

- 可視化方針の決め方 ..... 217
- 「良い指標」とは ..... 219
- 指標の運用の落とし穴 ..... 221
- 指標化の恣意性に対処する ..... 223

## 8.3 可視化されたデータの解釈学 ..... 225

- 仮説を立てることの重要性 ..... 225
- 統計学的有意性と効果 ..... 227
- 分析プロセス全体として可視化をとらえる ..... 228

第8章まとめ ..... 231

**A.1 Python データ可視化の基本知識** ..... 234

- おすすめの環境構築 ..... 234
- Matplotlib の説明 ..... 236
- Matplotlib における図の各構成要素 ..... 240
- グラフの描画と他ライブラリとの連携 ..... 243

**A.2 可視化プログラムの開発・運用テクニック** ..... 248

- わかりやすいコーディングのポイント ..... 248
- 変数に何が入っているかを明示する ..... 249
- どこで何をしているのかわかりやすくする ..... 253

**A.3 付録コードの利用の仕方** ..... 256

- 付録コードについて ..... 256
- よく登場する基本ライブラリ ..... 257

あとながき ..... 263

索引 ..... 264

著者紹介 ..... 269