

練習問題で学ぶグラフデザイン



授業のゴールと授業内容

データ可視化の役割・考慮点・デザインパターンを理解した上で、色々なデータに対して適切にグラフをデザインできるようになることを目指します。

授業のゴール

- ① データ可視化の役割を理解する。
- ② データを可視化する際に考慮するポイントについて理解を深める。
- ③ 色々なグラフの種類を知り、可視化のレパートリーを増やす。
- ④ データを可視化する際に用いるデザインパターンについて理解を深める。



授業内容

理論

第1回

データ可視化とは？
データ可視化の目的
データ可視化の構成要素と文脈

実践

第2-4回

練習問題を通じたグラフづくり
量データの可視化
割合データの可視化
データの関係の可視化
データの分布の可視化

Outline

全4回の講座のアウトラインは下記ようになります。

第1回（理論）

I. データ可視化とは？

1. データ可視化とは
2. データ可視化の目的
3. デザインの要素と文脈



第2-4回（実践）

II. 量データの可視化

- Q1. 量データの可視化
- Q2. 差のある量データの可視化

III. 割合データの可視化

- Q3. 割合の可視化
- Q4. 割合の経時変化の可視化
- Q5. 階層的な割合の可視化

IV. データの関係の可視化

- Q6. 2つのデータの関係の可視化
- Q7. 3つのデータの関係の可視化

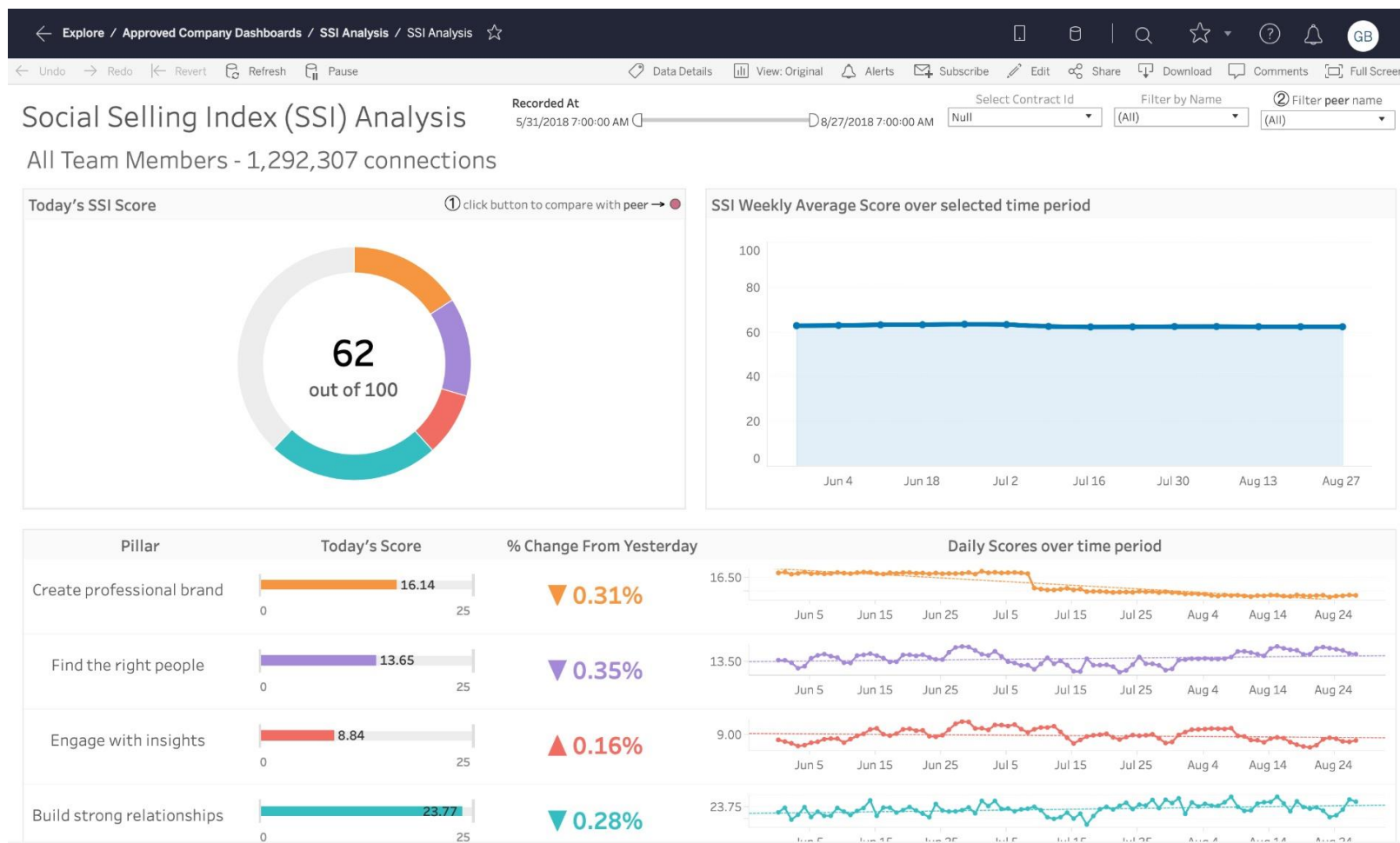
V. データの分布の可視化

- Q8. 1つの分布の可視化
- Q9. 2つ以上の分布の可視化

第1回

データ可視化とは？

データ可視化とは、データをグラフや表を使って表現することです。その際には、如何にわかりやすく、正確に表現できるかがポイントです。



<https://www.tableau.com/>

練習問題で学ぶグラフデザイン

データ可視化とは？

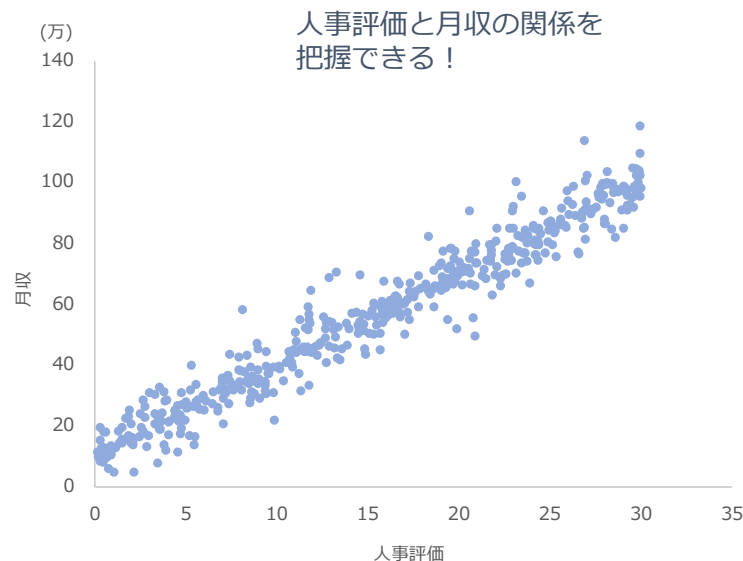
データを上手く可視化できれば、わかりやすく情報を伝えたり、新たな発見ができたりします。

例1

600人分のデータ

従業員ID	人事評価	月収
10010	10.12	39.7
10000	0.49	18.0
10026	25.95	94.0
10017	16.77	55.9
	●	
	●	
	●	
10596	18.89	66.0
10597	15.36	59.7
10598	22.52	74.8
10599	19.63	68.9
10600	17.03	60.4

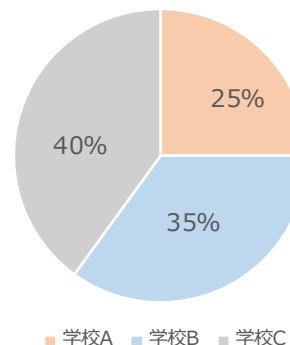
散布図による可視化



例2

学校名	人数	割合
学校A	125	25%
学校B	175	35%
学校C	200	40%

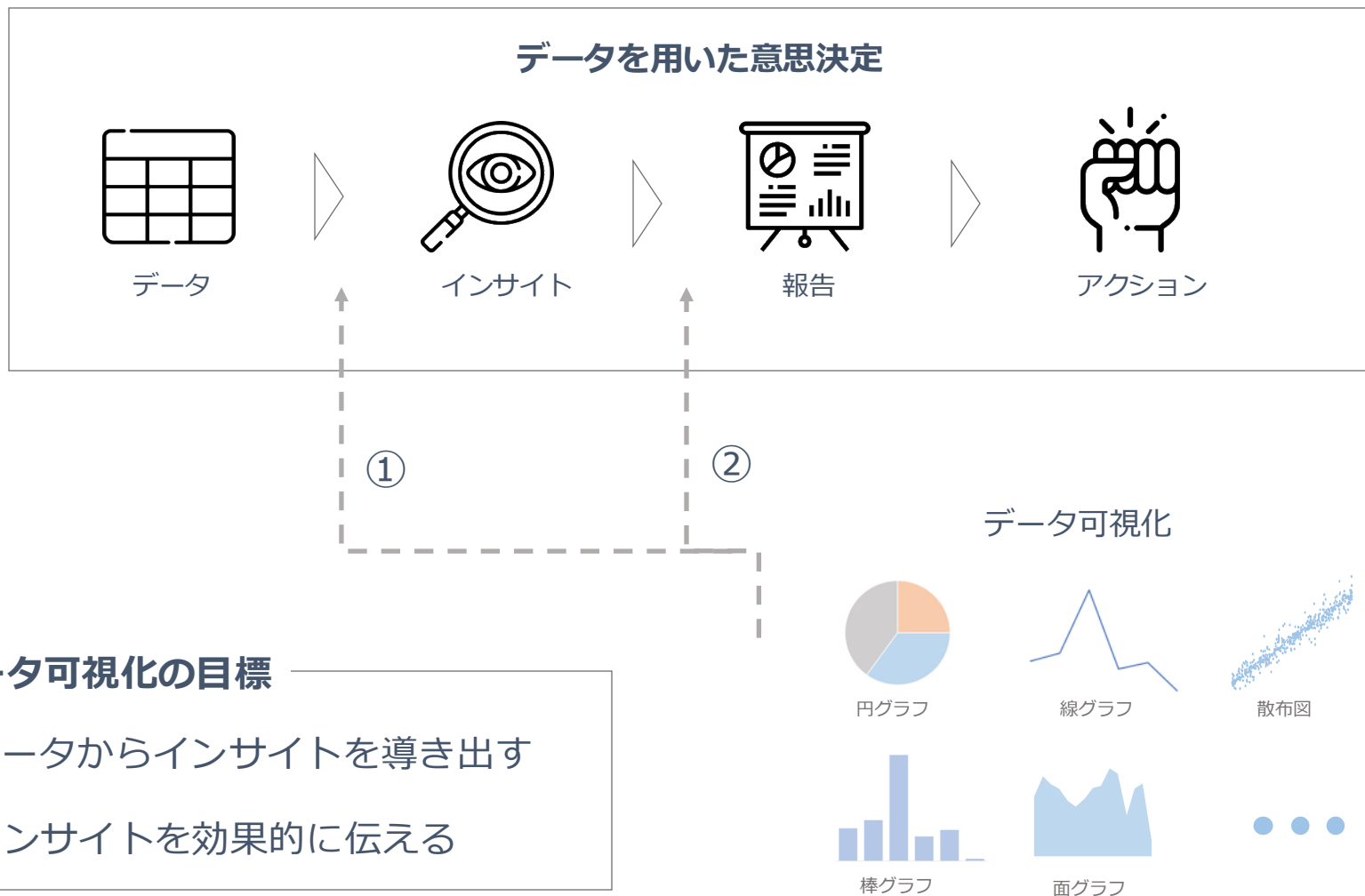
円グラフによる可視化



構成比が簡単にわかる。

データ可視化の目的

データ可視化は人間には理解が難しい複雑なデータから、①新たな知見を得ることや②データを効果的に伝えるために用いることができます。



データ可視化のデザイン

データ可視化では、以下のような特徴を持ったグラフデザインを目指します。

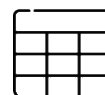
理想的なグラフの特徴

- ① データが持つ意味が簡単に理解できる。
- ② 正確にデータが可視化されており、誤解を招きにくい。
- ③ 気づかなかった知見が発見できる。
- ④ 美的であり、見ていて楽しい。



データ可視化の目標

- ① データからインサイトを導き出す



データ



インサイト

- ② インサイトを効果的に伝える



インサイト



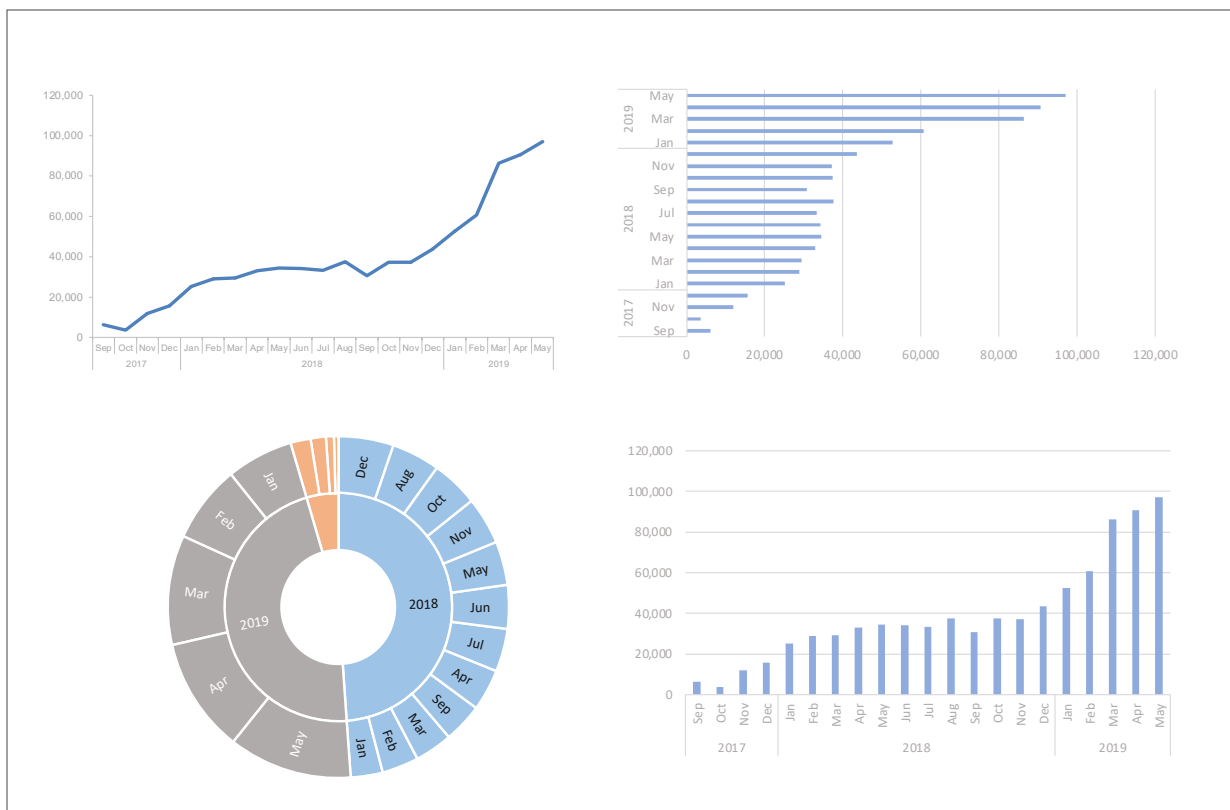
報告

データ可視化のデザイン

同じデータであっても、多様な可視化方法があります。これらを適切に選択/応用するために、デザインパターンについて学ぶ必要があります。

適切なグラフの選択

年度	月	ユーザ数
2017	Sep	6,212
2017	Oct	3,631
2017	Nov	11,864
2017	Dec	15,668
2018	Jan	25,151
2018	Feb	28,936
2018	Mar	29,405
2018	Apr	32,944
2018	May	34,375
2018	Jun	34,206
2018	Jul	33,357
2018	Aug	37,569
2018	Sep	30,693
2018	Oct	37,387
2018	Nov	37,257
2018	Dec	43,661
2019	Jan	52,661
2019	Feb	60,668
2019	Mar	86,407
2019	Apr	90,603
2019	May	97,043



データ可視化のデザイン

「なぜデータを可視化するのか」や「データ可視化どう使われるか」などデータ可視化を行う**文脈**を参考に適切なデザインを構成していく必要があります。

データ可視化のときに考慮する点

対象者は？ (to whom)

- 現場アナリスト？
- 社長/専務クラス？
- エンジニア？
- 社内/社外ステークホルダー？

必要なインサイトは？ (why)

- 長期的な傾向を得たい。
- 値ごとの比較をしたい。
- 新しい知見を発見したい。
- データごとの関係を知りたい。
- データのばらつきを把握したい。

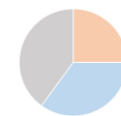
データの特徴は？ (what)

- データの次元数は？
- データのサイズ（量）は？
- データの種類は？（量データ / カテゴリデータ etc. ）
- データが表すものは？（e温度？ 個数？ etc. ）

コミュニケーションの手段は？ (how)

- スライド？
- ドキュメント？（ワードなど）
- ダッシュボード？
- WEBベース？
- モバイルアプリ？

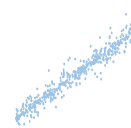
データ可視化方法の決定



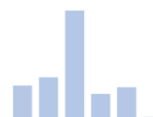
円グラフ



線グラフ



散布図



棒グラフ



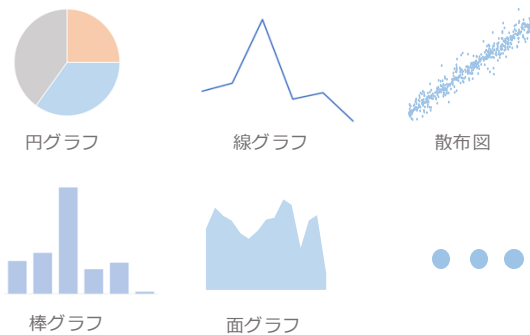
面グラフ



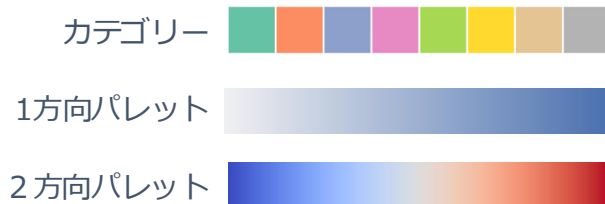
データ可視化のデザイン

文脈を踏まえた上で、以下のような様々な構成要素を取捨選択/応用などしていくのが、データ可視化のプロセスになります。構成要素の例として以下があります。

グラフの種類



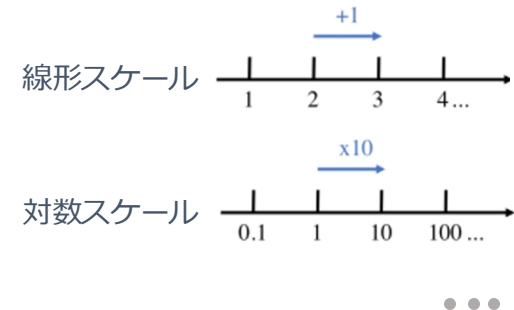
色の選択



グラフ要素

グラフ罫線	グラフタイトル
軸目盛り	軸タイトル
軸の範囲	データラベル
注釈	...

スケール



インタラクション

フィルタリング	ズーム
ドリルダウン	ソート
ホバリング	...



今回の講義外

データ可視化のデザイン

データ可視化のデザインを考える上で、データの種類について理解することが重要です。代表的なものとして以下のものがあります。

データタイプ	順序	連続/離散	例	説明
数値	あり	連続	1.0, 1.1, 1.2	中間値に意味がある数値
		離散	1個, 2個, 3個, 4個	中間値に意味のない数値
カテゴリ値	なし		A, B, C, D（成績）	順序があるカテゴリ値
		アメリカ、中国、日本	順序のないカテゴリ値	

データの種類ごとに適切な
デザインが異なる！

データ可視化のデザイン

また、データの種類だけでなく、データの次元数、サイズ、意味合いなどの要素も重要です。

データ次元数

- 1つのデータがもつ特徴の数（例えば、ある人のデータには色々な側面）のこと
- データの次元数が多い場合、事前に重要な特徴を選んだりなどの対処が必要な場合がある。

データ番号	特徴A	特徴B	特徴C		特徴Y	特徴Z
1	9.4	3.6	9.5		10.0	10.0
2	9.8	1.7	0.2		6.5	6.5
3	6.2	4.3	6.2		2.0	2.0
4	4.9	9.1	2.2		3.0	3.0
5	4.4	6.3	6.7	...	1.2	1.2
	...					
999	5.9	9.0	6.0		1.2	1.2
1000	2.2	2.5	0.8		1.7	1.7

データサイズ

- データの個数のこと
- データが多すぎる場合、データ点同士の重なりが生じがち。
- データが少なすぎる場合、グラフが疎になり、注目点がわかりにくく。

データの意味合い

- 特徴がどのような意味を持っているか？
- データの意味合いによってデザインが変わってくる。
例えば、30という数字があり、気温（℃）のデータだと高いと赤っぽいイメージ。月収（万円）だと黒や青。

データ可視化のデザイン

適切にデータを可視化するためには、デザインパターン以外にも、様々な知識やテクニックが関わってきます。多くの人が得意な分野を活かして関わられます。

IT系の人



① 数学や統計の知識

- 統計量（平均や割合など）の計算
- 予測分析やクラスタリングなど

② ITの知識

- データベースの構築運用
- ダッシュボードシステムの作製や運用
- データ処理プロセスの作製

ビジネス系の人



③ ビジネス知識

- データを活用したいテーマ選定スキル。
- データを用いた意思決定。

④ コミュニケーションスキル

- 資料作製スキル
- データを用いたプレゼンテーションスキル

アート系の人



⑤ アート・デザイン経験

- 印象に残るデザイン作成スキル
- 見ていて楽しいグラフデザイン

データ可視化のデザイン

次回以降は、例題を通じて、実際に良いデザインについて学んでいきます。実際に自分でデータを可視化してみることで、改善点などへの気づきを得ます。

第1回（理論）

I. データ可視化とは？

1. データ可視化とは
2. データ可視化の目的
3. デザインの要素と文脈



第2-4回（実践）

II. 量データの可視化

- Q1. 量データの可視化
- Q2. 差のある量データの可視化

III. 割合データの可視化

- Q3. 割合の可視化
- Q4. 割合の経時変化の可視化
- Q5. 階層的な割合の可視化

IV. データの関係の可視化

- Q6. 2つのデータの関係の可視化
- Q7. 3つのデータの関係の可視化

V. データの分布の可視化

- Q8. 1つの分布の可視化
- Q9. 2つ以上の分布の可視化

練習問題で学ぶグラフデザイン

データサイエンティスト

高比良 健太郎

第2回

Outline

全4回の講座のアウトラインは下記ようになります。

第1回（理論）

I. データ可視化とは？

1. データ可視化とは
2. データ可視化の目的
3. デザインの要素と文脈



第2-4回（実践）

II. 量データの可視化

- Q1. 量データの可視化
- Q2. 差のある量データの可視化

III. 割合データの可視化

- Q3. 割合の可視化
- Q4. 割合の経時変化の可視化
- Q5. 階層的な割合の可視化

IV. データの関係の可視化

- Q6. 2つのデータの関係の可視化
- Q7. 3つのデータの関係の可視化

V. データの分布の可視化

- Q8. 1つの分布の可視化
- Q9. 2つ以上の分布の可視化

Outline

全4回の講座のアウトラインは下記ようになります。

第1回（理論）

I. データ可視化とは？

1. データ可視化とは
2. データ可視化の目的
3. デザインの要素と文脈



第2-4回（実践）

II. 量データの可視化

- Q1. 量データの可視化
- Q2. 差のある量データの可視化

III. 割合データの可視化

- Q3. 割合の可視化
- Q4. 割合の経時変化の可視化
- Q5. 階層的な割合の可視化

IV. データの関係の可視化

- Q6. 2つのデータの関係の可視化
- Q7. 3つのデータの関係の可視化

V. データの分布の可視化

- Q8. 1つの分布の可視化
- Q9. 2つ以上の分布の可視化

データ可視化のデザイン

講義3回を通じて、練習問題を通じてグラフのデザインについて学んでいきます。

目標

- 色々なグラフの種類を知る。
- 練習問題ごと正解を学ぶのではなく、**デザインパターン**を学んでいく。

第1回 再掲

対象者は？

- 現場アナリスト？
- 社長/専務クラス？
- エンジニア？
- 社内/社外ステークホルダー？

必要なインサイトは？

- 長期的な傾向を得たい。
- 値ごとの比較をしたい。
- 新しい知見を発見したい。
- データごとの関係を知りたい。
- データのばらつきを把握したい。

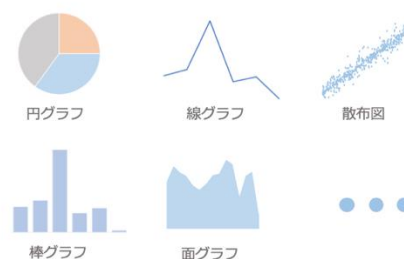
データの特徴は？

- データの次元数は？
- データのサイズ（量）は？
- データの種類は？（etc 量データ / カテゴリデータ）
- データが表すものは？（etc 温度？個数？）

コミュニケーションの手段は？

- スライド？
- ドキュメント？（ワードなど）
- プレゼンテーション？
- WEBベース？
- モバイルアプリ？

データ可視化方法の決定



練習問題 1

量データの可視化 – 量データの可視化 ①

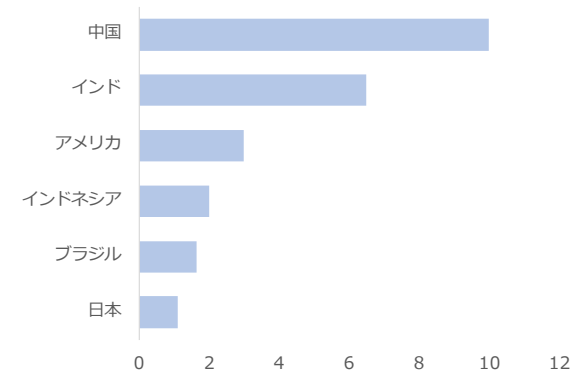
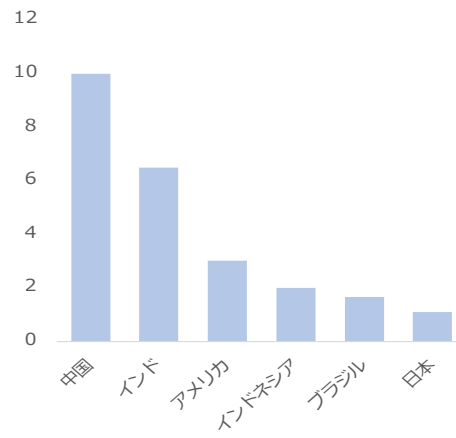
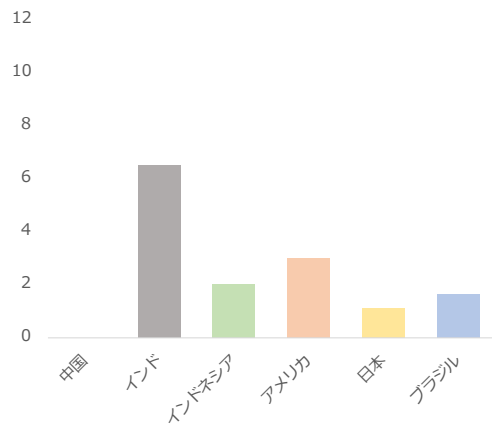
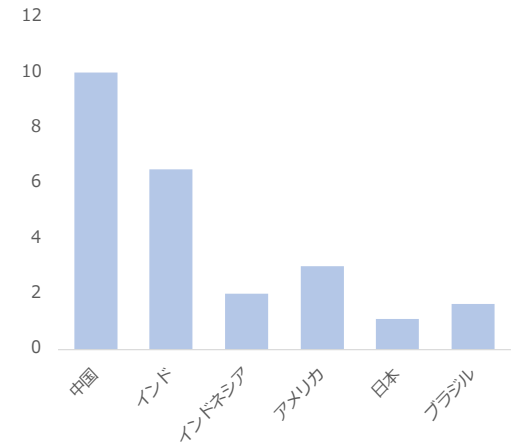
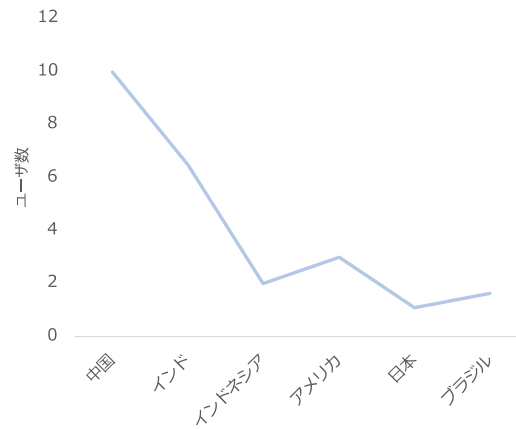
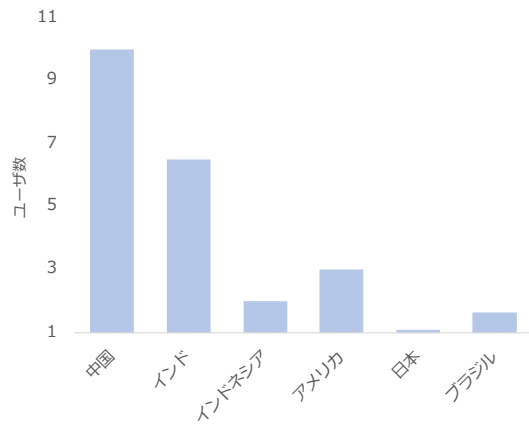
下記は主要国のインターネットユーザ数（2022年2月時点）になります。ユーザ数の比較がしやすいようなデータ可視化をデザインしてみてください。

国名	ユーザ数（億）
中国	10.0
インド	6.5
インドネシア	2.0
アメリカ	3.0
日本	1.2
ブラジル	1.7

<https://www.statista.com/statistics/262966/number-of-internet-users-in-selected-countries/>

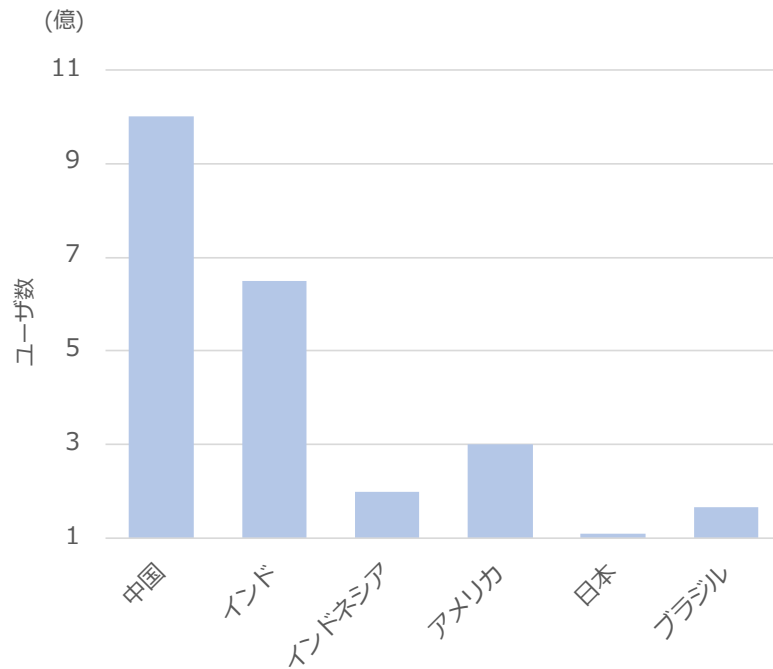
量データの可視化 – 量データの可視化 ①

回答例

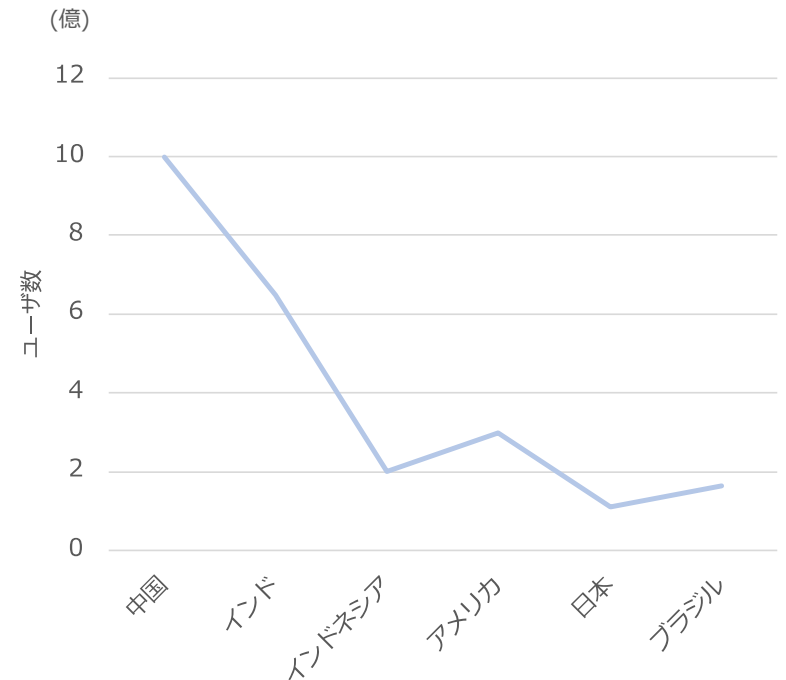


量データの可視化 – 量データの可視化 ①

棒グラフによる可視化



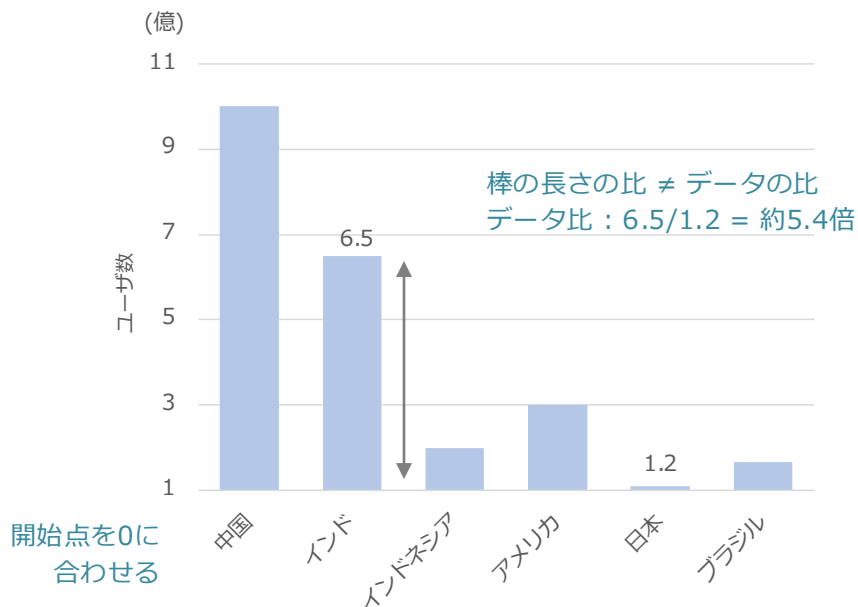
線グラフによる可視化



量データの可視化 – 量データの可視化 ①

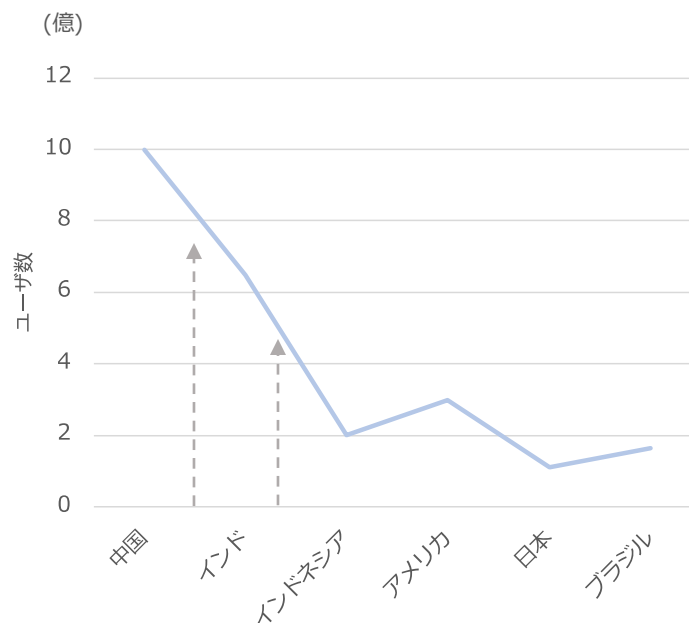
棒グラフによる可視化

- ⊗ 開始点が0基準ではないため、差が過度に強調されてしまっており、棒の長さの比がデータの比と一致しない。



線グラフによる可視化

- ⊗ 国と国の間の値に意味がない。
e.g.) 中国とインドの間の値に意味がない。
- ⊗ 線グラフで表現されている up/down に特に意味がない。



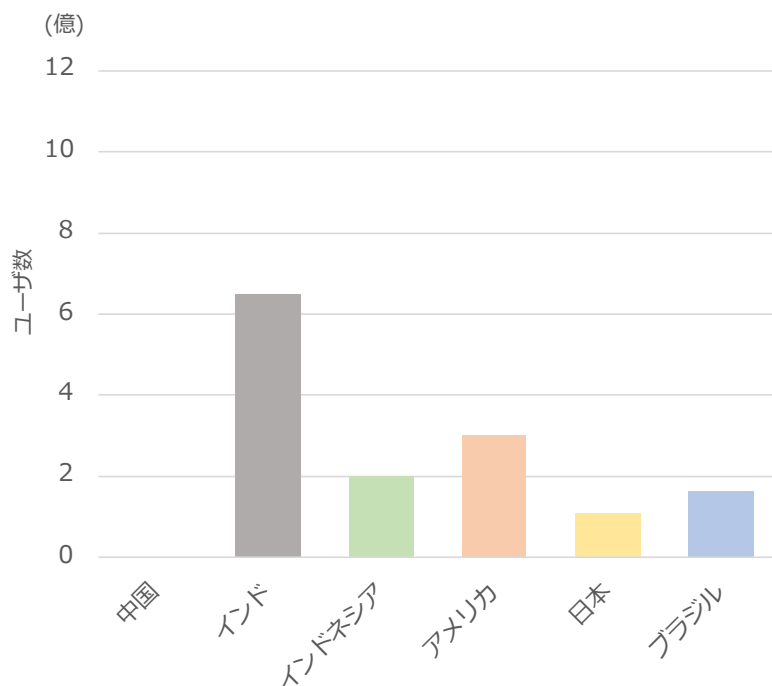
- 💡 軸の始点をカットしてしまうと、差が過度に強調されてしまい、棒の長さの比とデータの比が一致しなくなる。

- 💡 連続性がないデータには線グラフは使用しない。

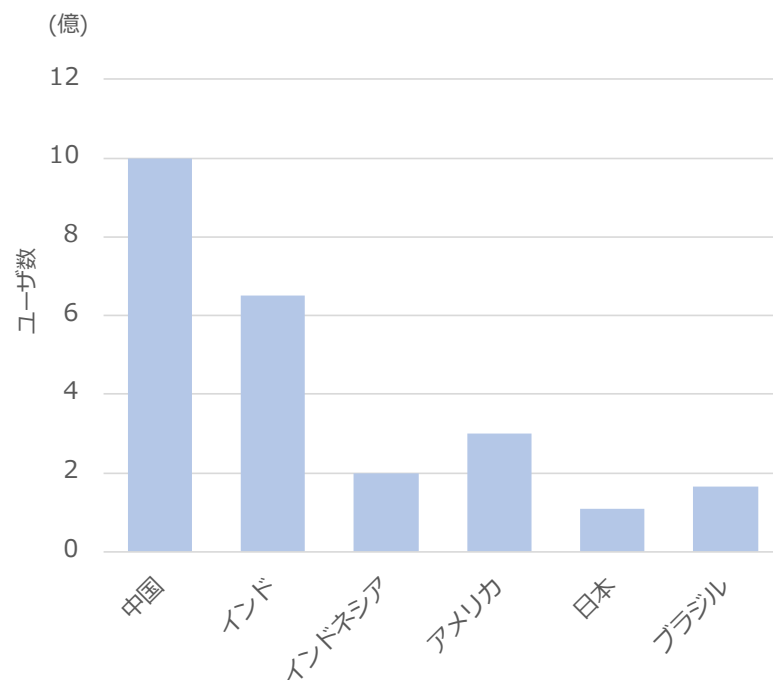
量データの可視化 – 量データの可視化 ①

棒グラフによる可視化

- 各棒ごとに異なる色を使用 -



棒グラフによる可視化



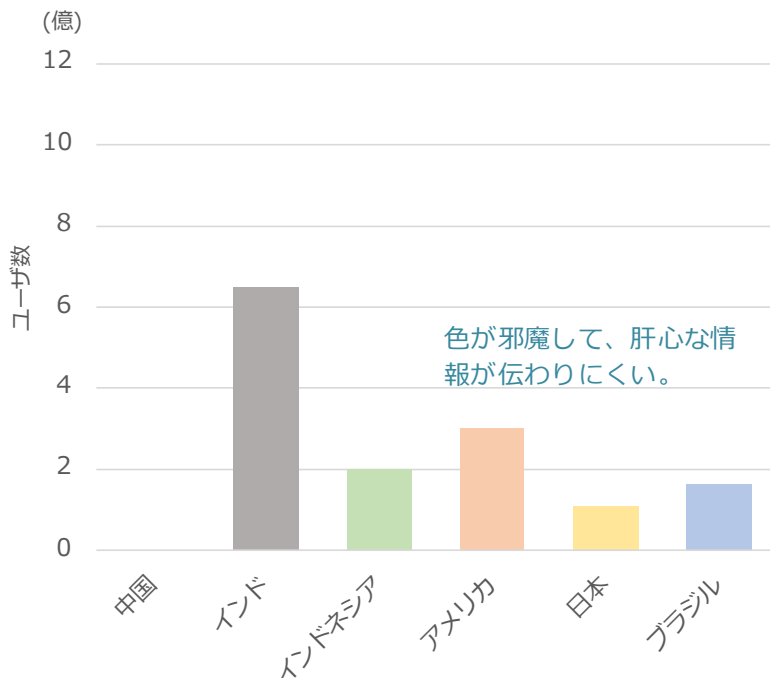
量データの可視化 – 量データの可視化 ①

棒グラフによる可視化

- 各棒ごとに異なる色を使用 -



国ごとに使用している色が意味を持っていない。

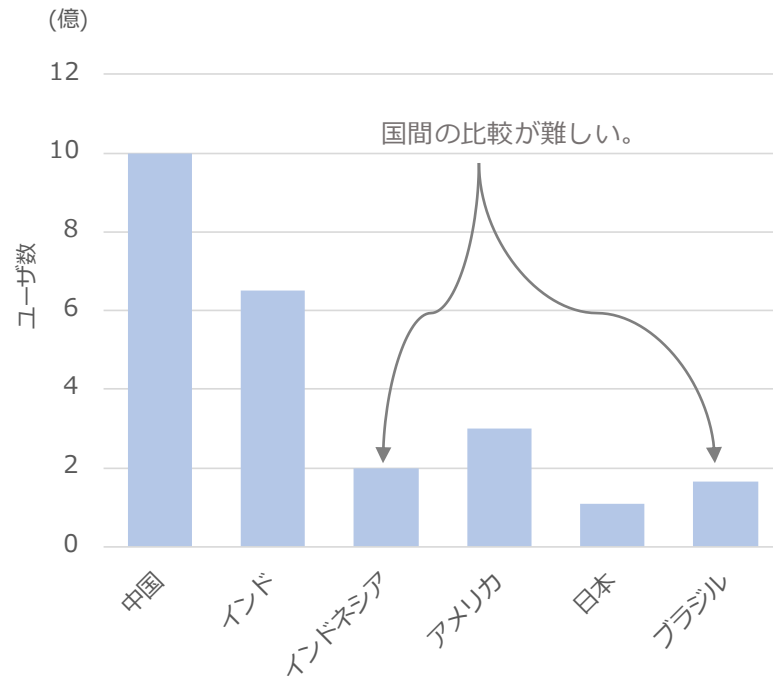


色を使えば、「情報がもっと豊かになるか？」考える。意味なく色を使わない。

棒グラフによる可視化



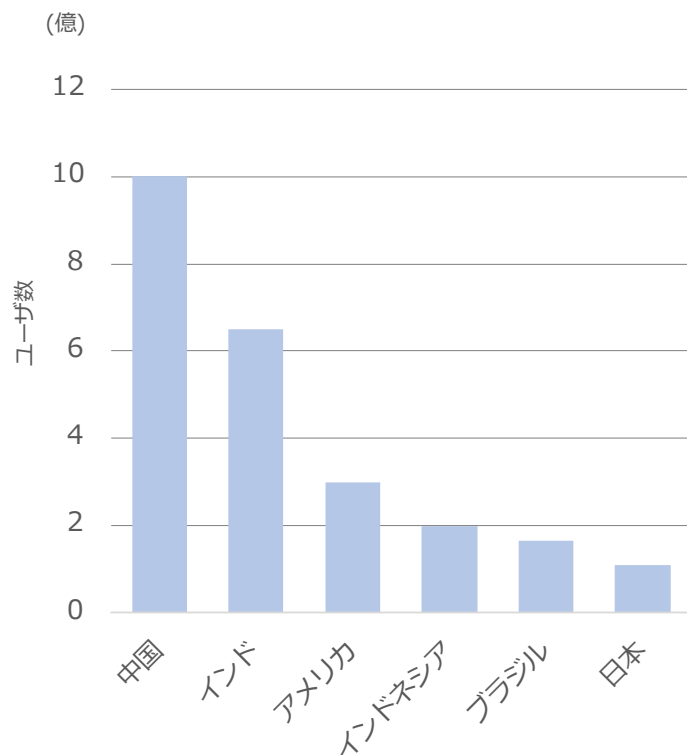
国の順番に特に意味がない。
隣り合わない棒の比較が難しい。



量データの可視化 – 量データの可視化 ①

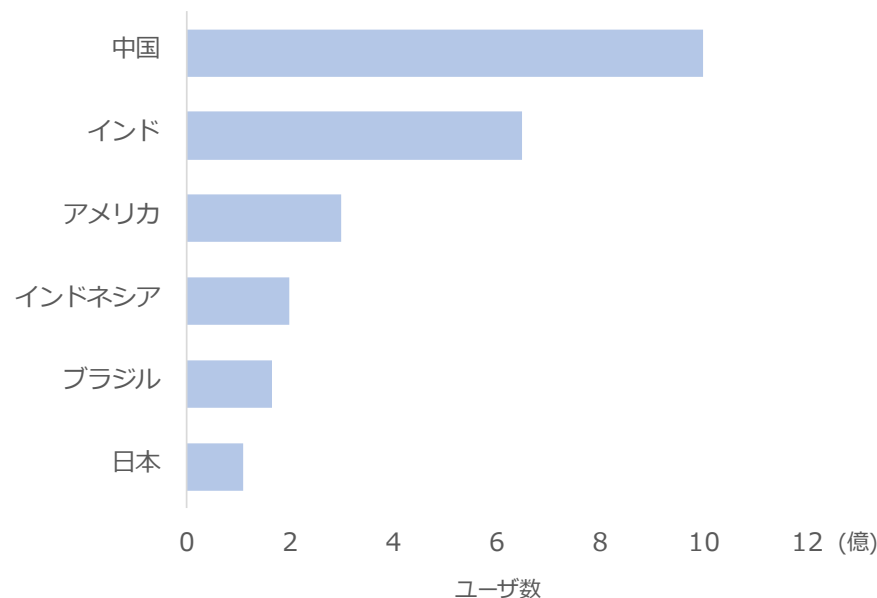
棒グラフによる可視化

- 棒の順序に規則性がある -



棒グラフによる可視化

- 棒の順序に規則 + 水平グラフ



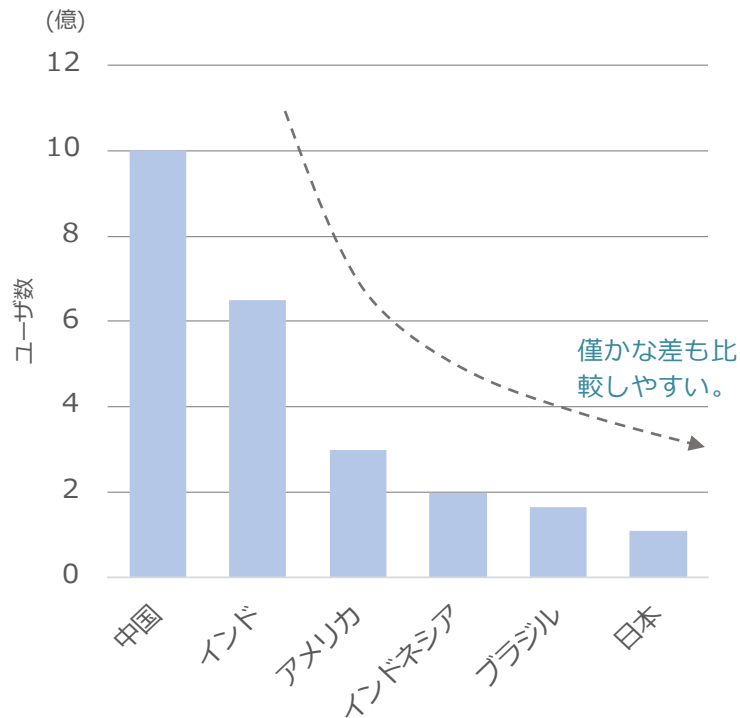
量データの可視化 – 量データの可視化 ①

棒グラフによる可視化

- 棒の順序に規則性がある -



棒グラフが昇順に並んでおり、微妙な差の比較もしやすい。



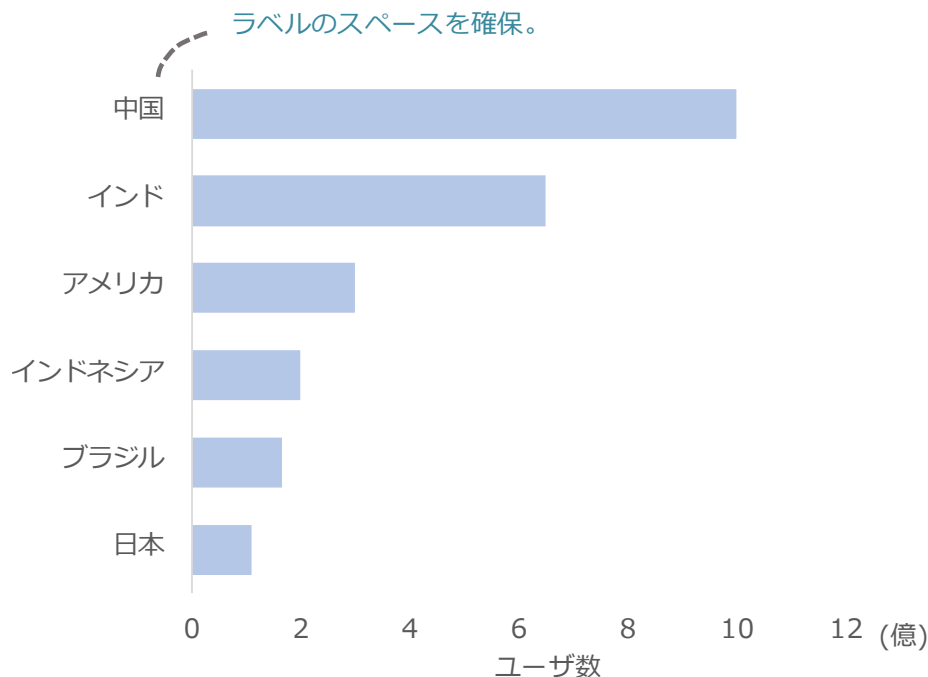
棒グラフを使うときは、棒の並び順にも気を配る。意味のある順番か比較がしやすいものを。

棒グラフによる可視化

- 棒の順序に規則 + 水平グラフ



ラベルの位置が十分に確保されており、ラベルがよみやすい。

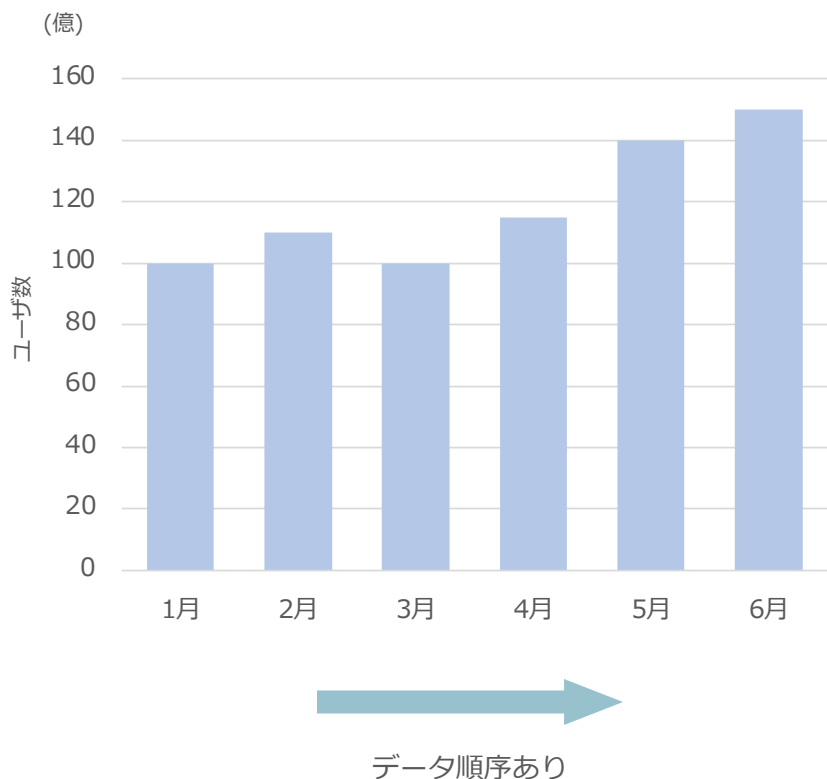


棒グラフの向きを水平にするとラベルのスペースを確保しやすい。

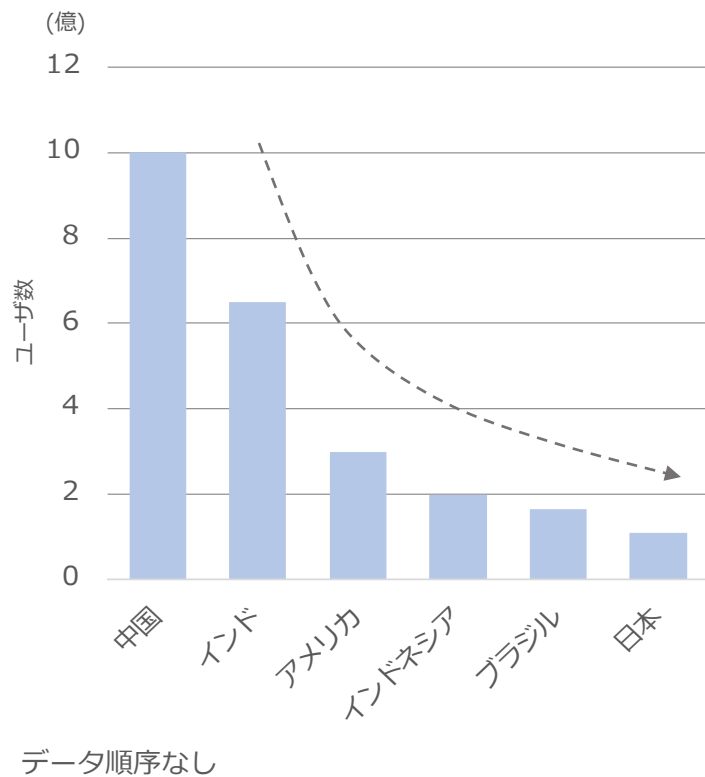
量データの可視化 – 量データの可視化 ①

棒グラフの並び順は、データに順序があれば、その順序で並び替える。
なければ、データの大きさで並び替えるのがおすすめ。

月ごとのユーザ数の推移



主要国のユーザ数



棒グラフを使うときは、棒の並び順にも気を配る。
意味のある順番があるならその順番で並び替え。なければ、データの大きさで並び替え。

量データの可視化 – 量データの可視化 ①



軸の始点をカットしてしまうと、差が過度に強調されてしまい、棒の長さの比とデータの比が一致しなくなる。



連続性がないデータには線グラフは使用しない。



色を使えば、「情報がもっと豊かになるか？」考える。意味なく色を使わない。



棒グラフを使うときは、棒の並び順にも気を配る。
意味のある順番があるならその順番で並び替え。なければ、データの大きさに並び替え。



棒グラフの向きを水平にするとラベルのスペースを確保しやすい。

練習問題 2

量データの可視化 – 差のある量の可視化②

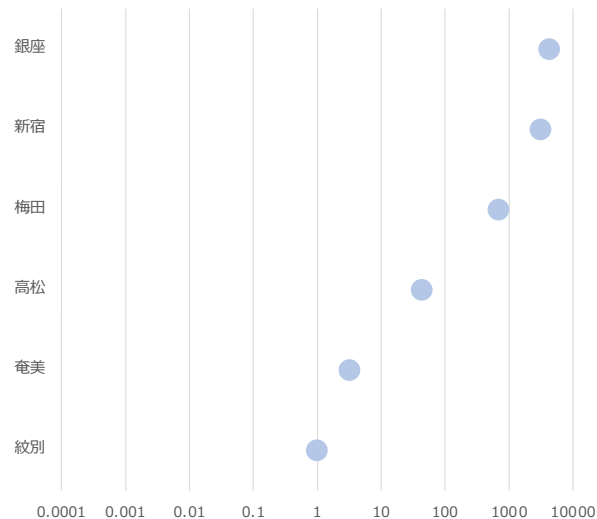
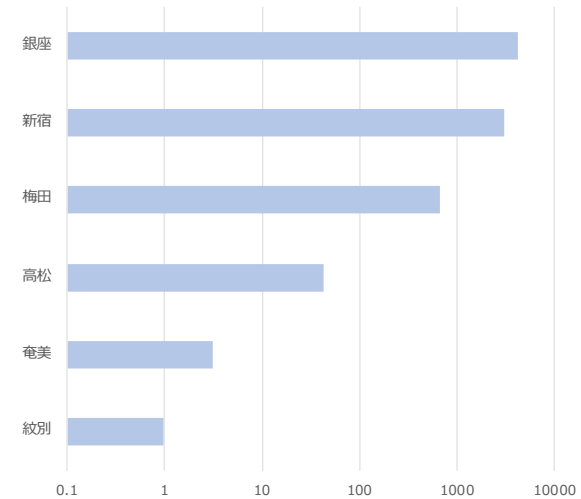
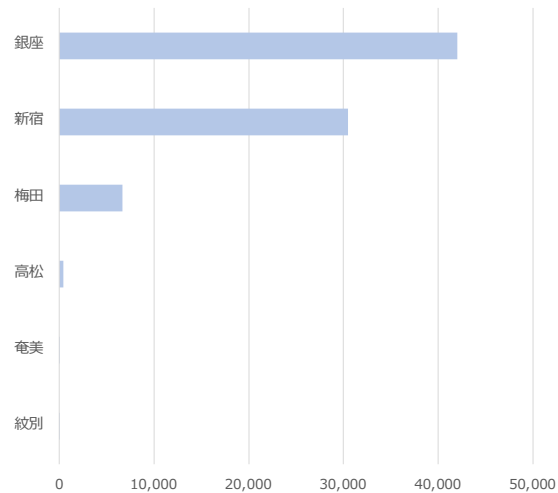
以下のような各地点ごとの公示地価のデータがあります。地点ごとの公示地価の比較を行うためには、どのようなグラフを作成すればよいでしょうか。

場所	地価 (円/m ²)
梅田	6,710,000
新宿	30,500,000
銀座	42,000,000
奄美	31,450
高松	430,000
紋別	9,725

<https://www.chikamap.jp/chikamap/Portal?mid=216>

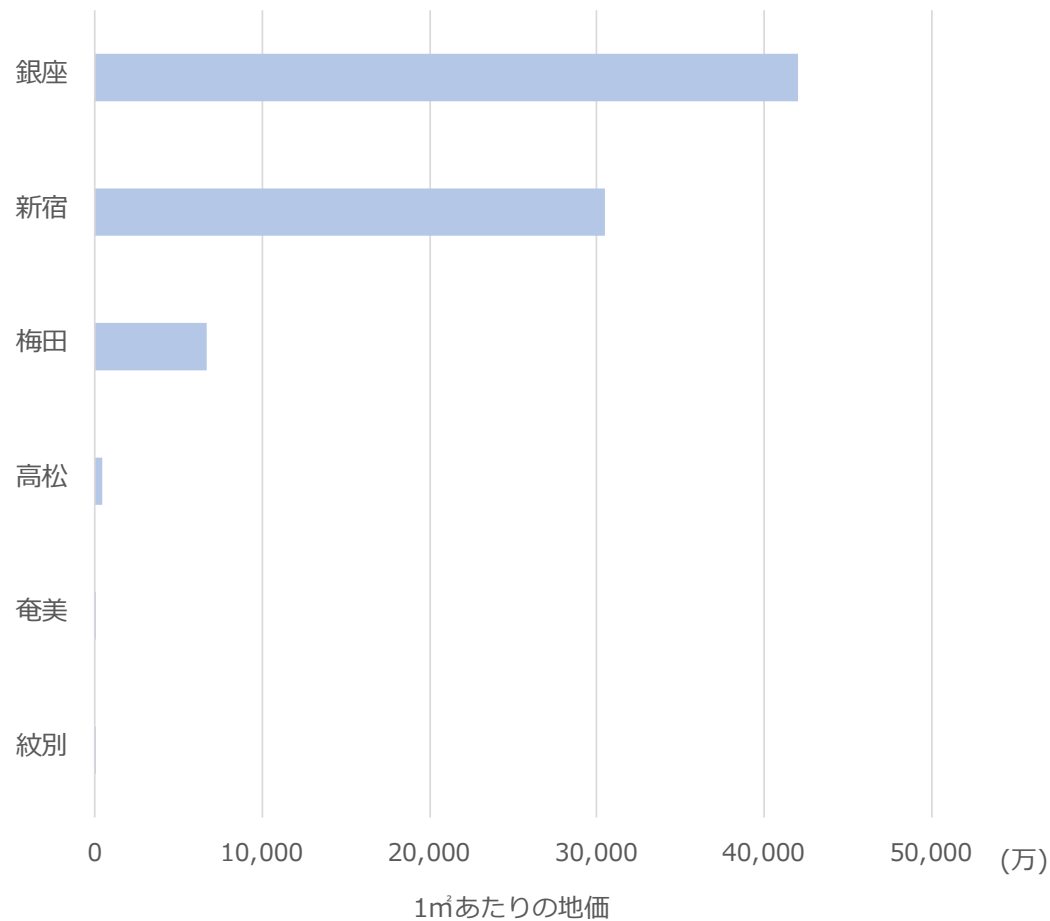
量データの可視化 – 差のある量の可視化②

回答例



量データの可視化 – 差のある量の可視化②

棒グラフによる可視化
- 線形スケールの使用 -

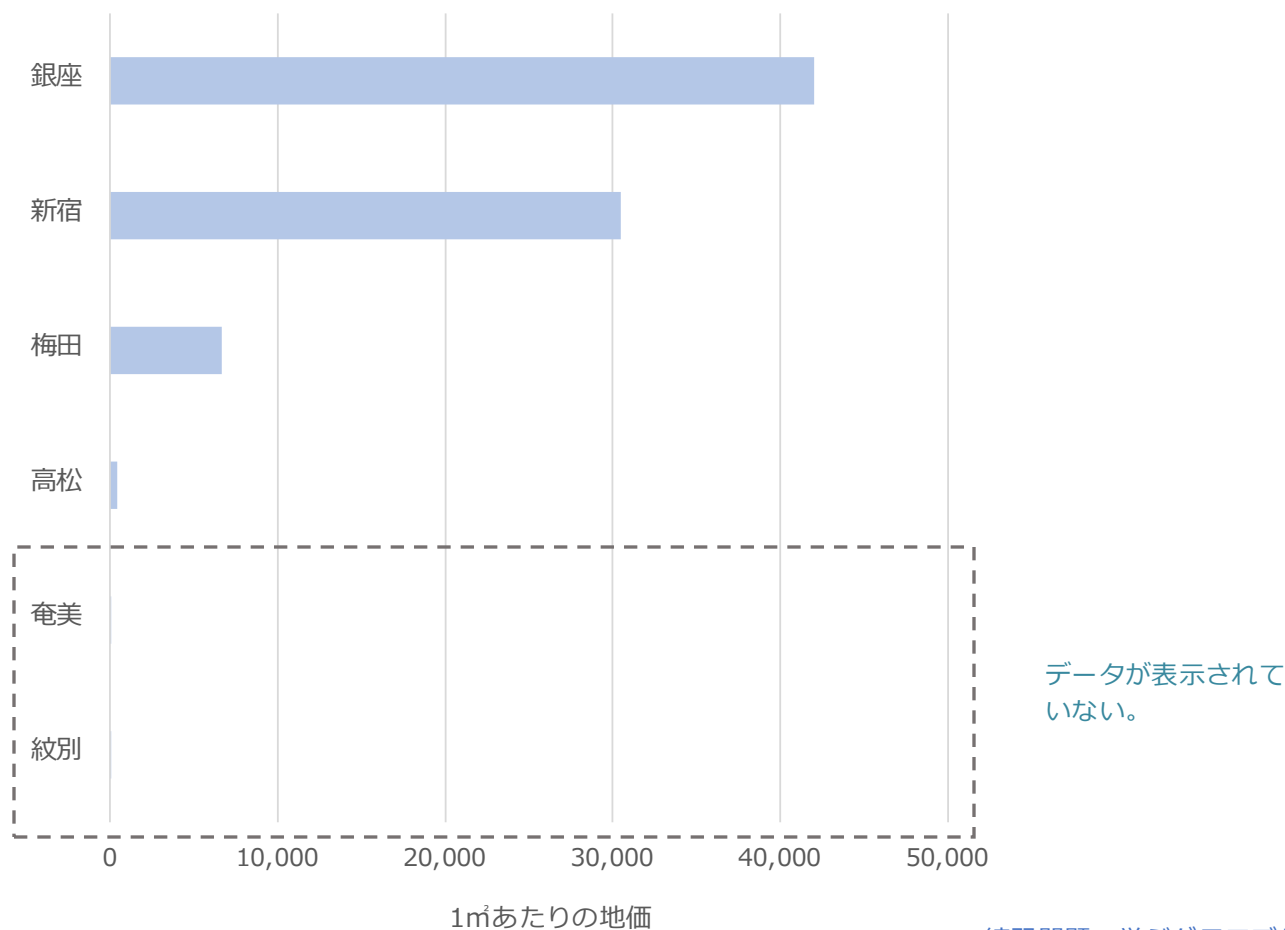


量データの可視化 – 差のある量の可視化②

棒グラフによる可視化

- 線形スケールの使用 -

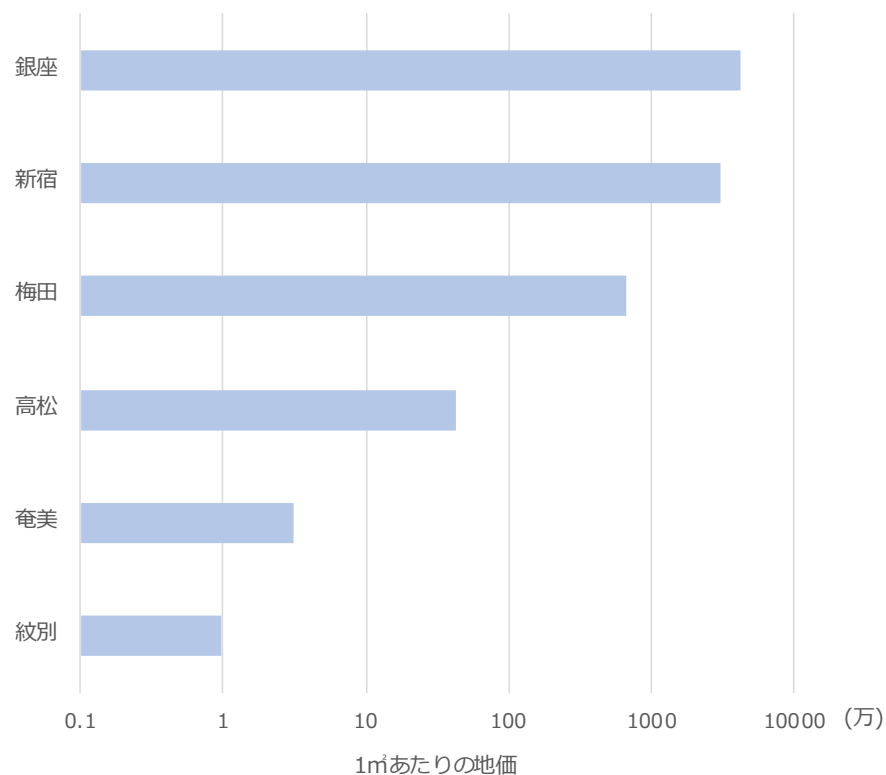
⊗ データ量に差があり、一部の棒グラフがつぶれてしまっている。



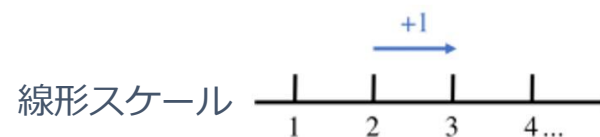
量データの可視化 – 差のある量の可視化②

棒グラフによる可視化

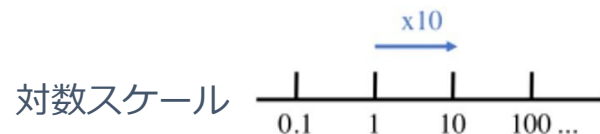
- 対数スケールの使用 -



スケール



等間隔の幅が一定量の変化を表す。

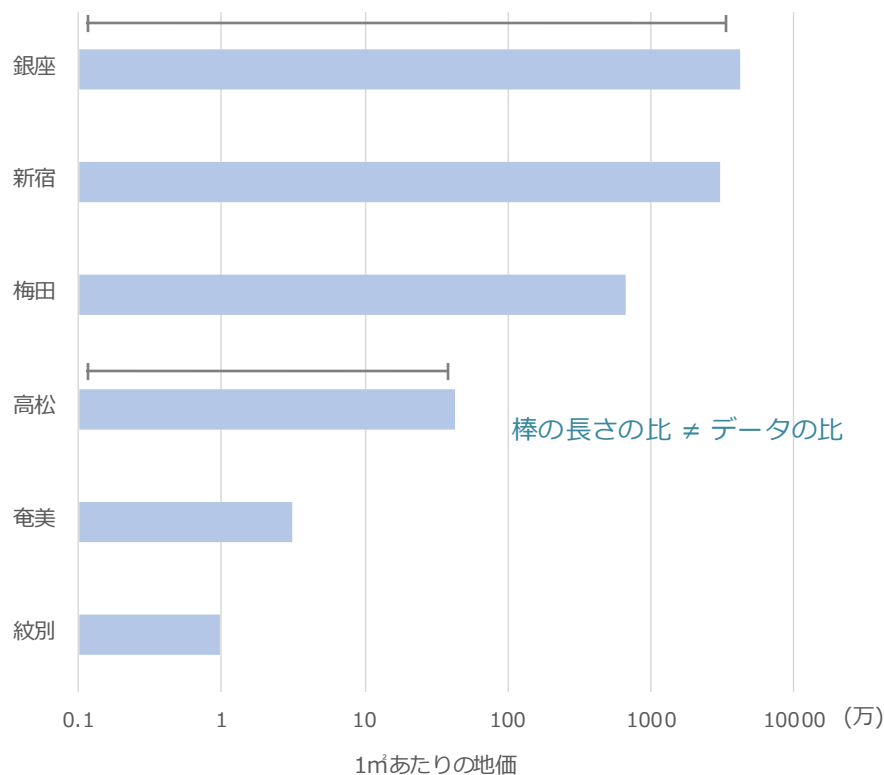


等間隔の幅が一定倍の変化を表す。
開始点が0ではない。

量データの可視化 – 差のある量の可視化②

棒グラフによる可視化

- 対数スケールの使用 -



広い範囲の正の値について表すことができる。



対数スケールであるため、棒グラフの長さが量の大きさを正確にあらわしていない。



データ量に差があるときは、スケールの変更を検討する。



対数スケールを用いることで、広い範囲の正の値について表すことができる。

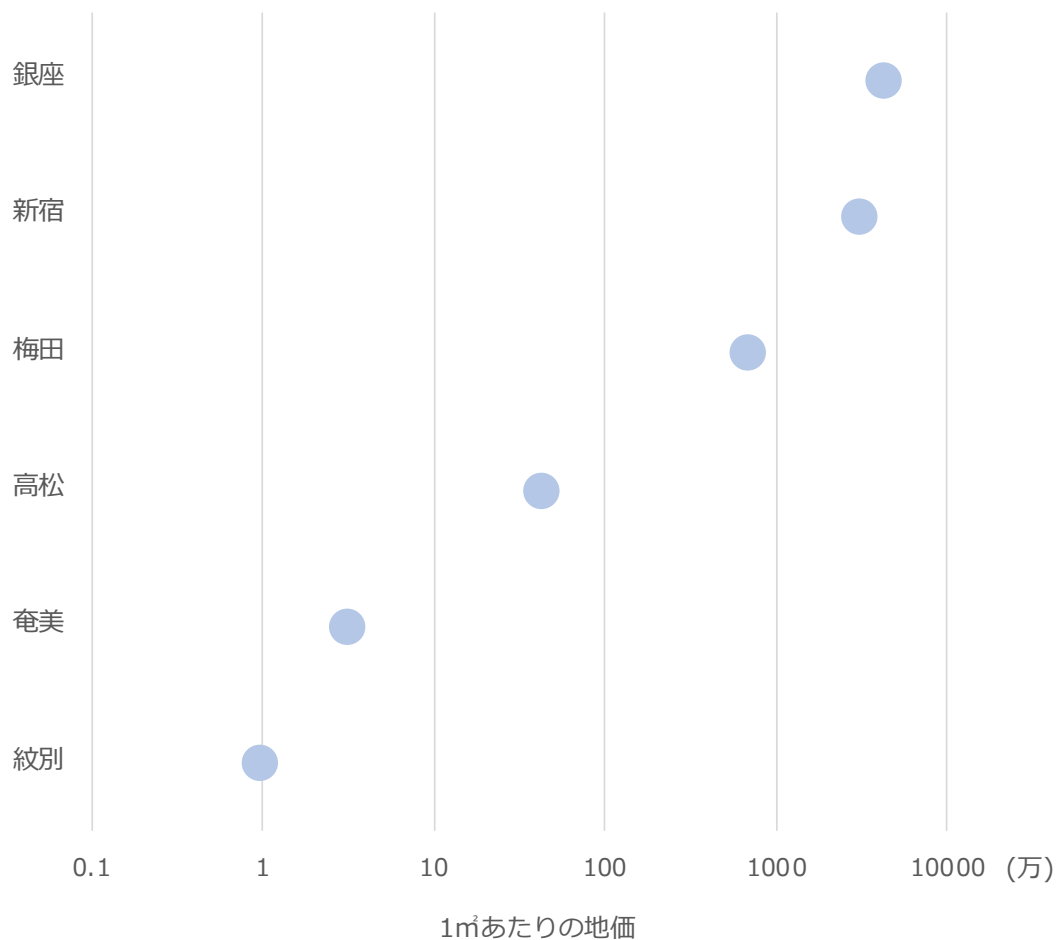


対数などのスケールを用いるときの棒グラフの使用は慎重に。

量データの可視化 – 差のある量の可視化②

ドットプロットによる可視化

- 対数スケールの使用 -

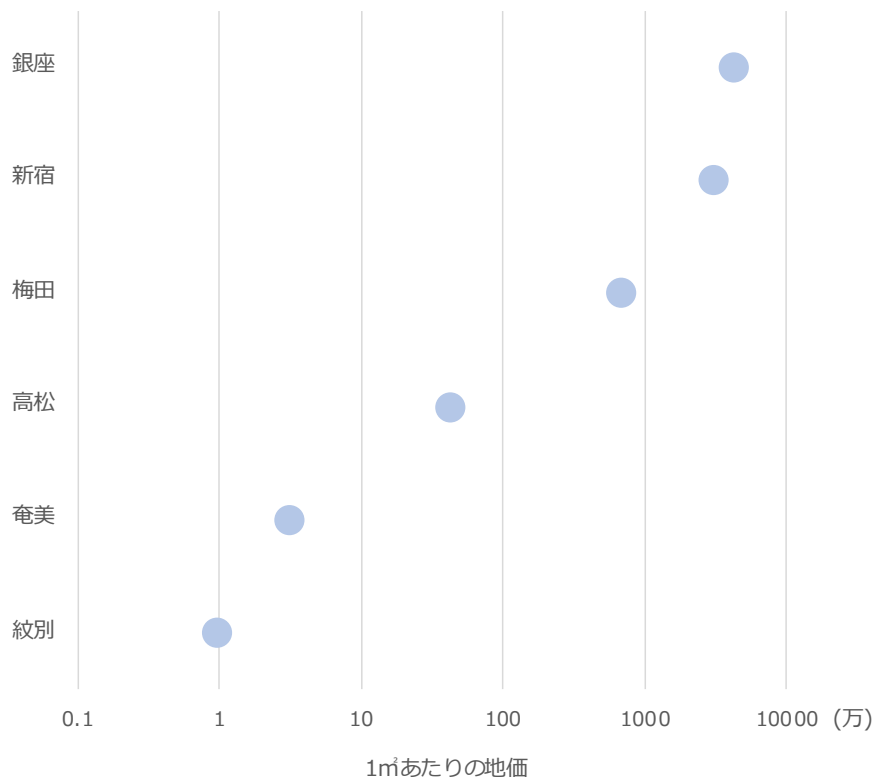


量データの可視化 – 差のある量の可視化②

ドットプロットによる可視化

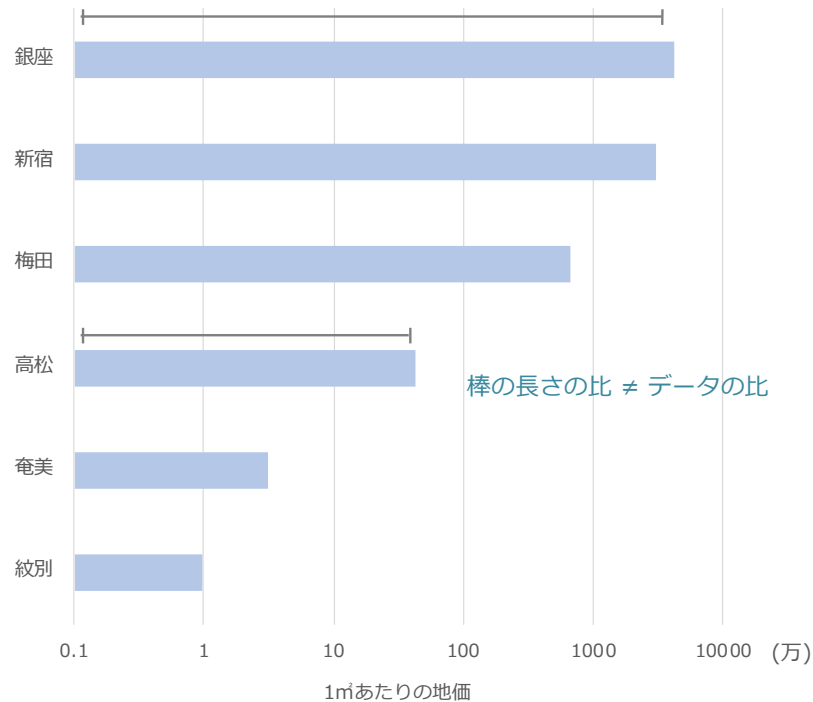


点の位置を意識した比較をすることで、棒グラフの長さ比のバイアスを除く。



軸が0始まりでないときは、ドットプロットの利用を検討

再掲 棒グラフによる可視化



量データの可視化 – 差のある量の可視化②



データ量に差があるときは、スケールの変更を検討する。



対数スケールを用いることで、広い範囲の正の値について表すことができる。



対数などの線形スケール以外のスケールを用いるときの棒グラフの使用は慎重に。



軸が0始まりでないときは、ドットプロットの利用を検討する。

練習問題 3

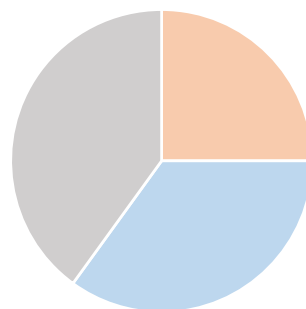
割合データの可視化 – 割合の可視化 ③

以下はある地区に居住する子どもたちの学校別の人数とそれぞれの全体に占める割合のデータです。各学校ごとの人数割合の差がよくわかるように可視化するにはどうすればいいでしょうか。

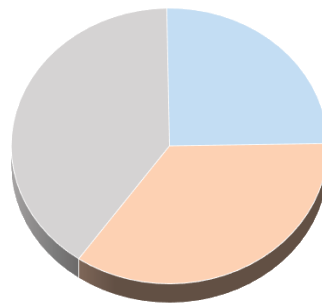
学校名	人数	割合
学校A	125	25%
学校B	175	35%
学校C	200	40%

割合データの可視化 – 割合の可視化 ③

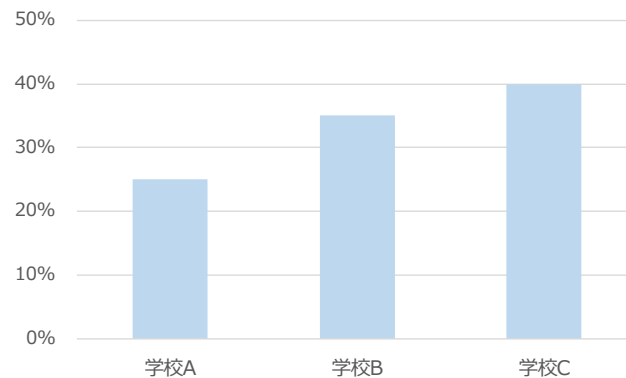
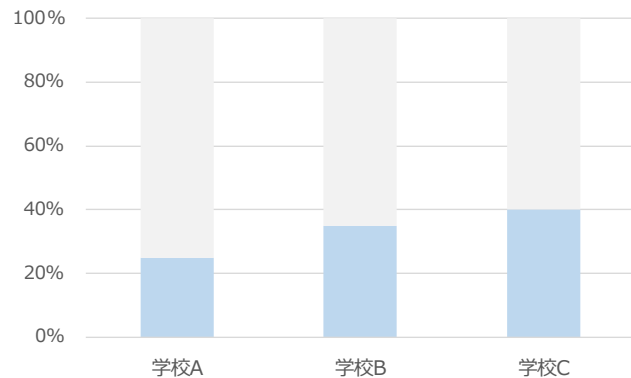
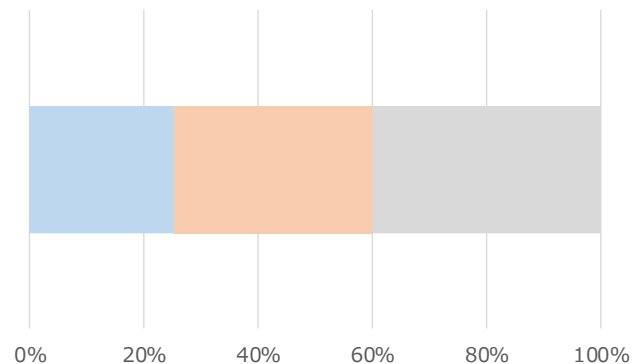
回答例



■ 学校A ■ 学校B ■ 学校C

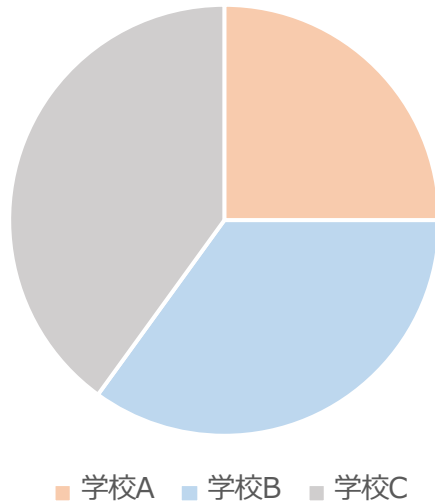


■ 学校A ■ 学校B ■ 学校C

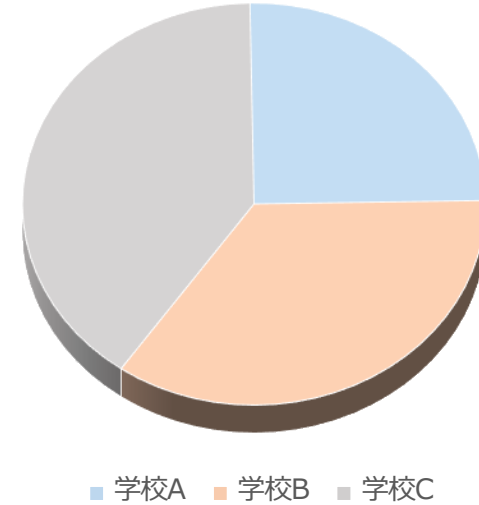


割合データの可視化 – 割合の可視化 ③

円グラフによる可視化



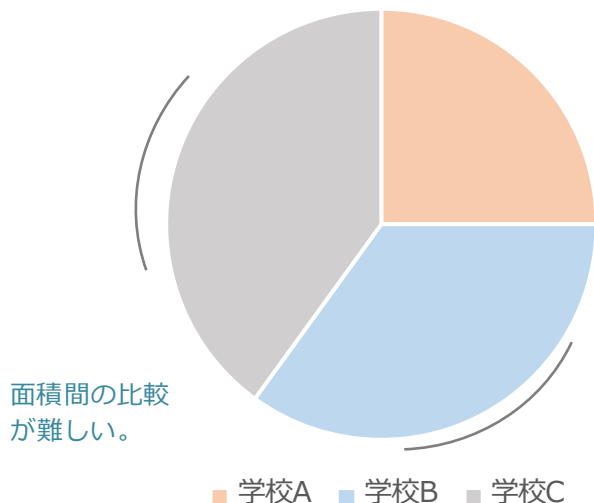
円グラフによる可視化（3D）



割合データの可視化 – 割合の可視化 ③

円グラフによる可視化

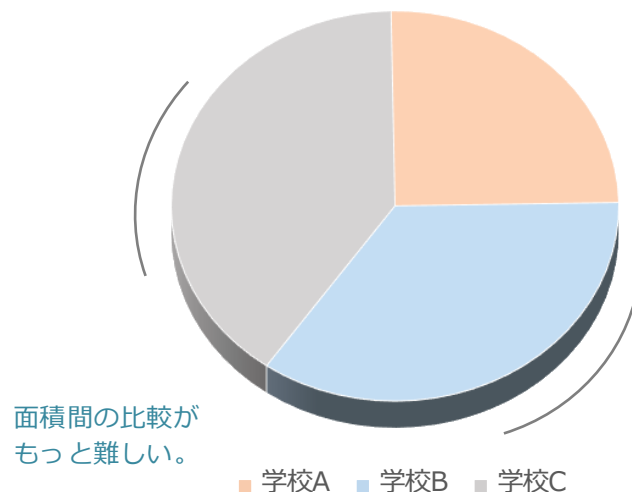
⚠ 扇形の面積の比較難しい。



💡 円グラフでの表現だと扇形の面積の比較が難しい。とくに割合ごとの差が小さい場合に注意。

円グラフによる可視化（3D）

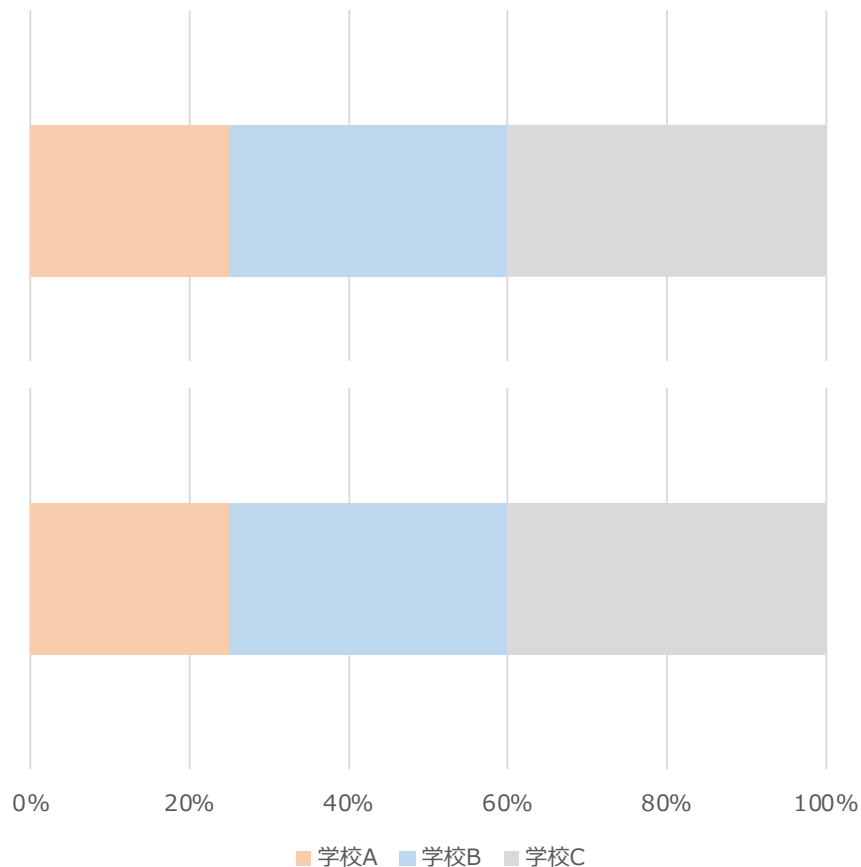
⊗ 3D表現は奥行きが出てしまう。奥行きがあると面積の比較が難しくなる。



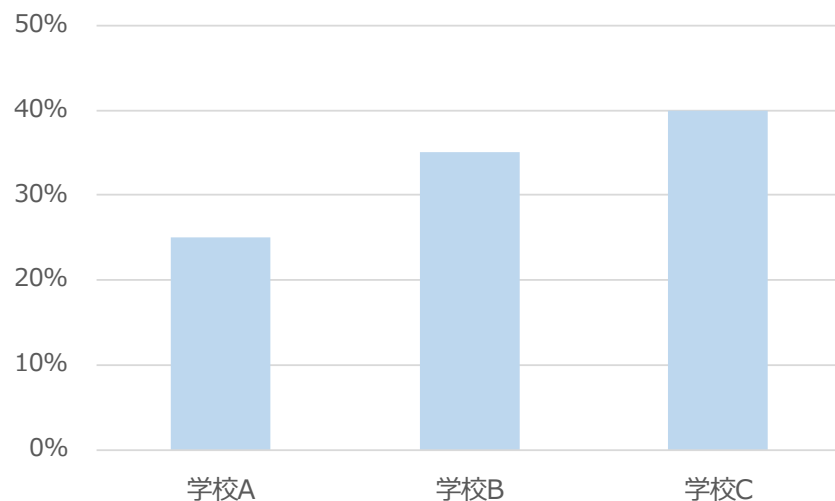
💡 基本的に3D表現は避けた方がいい。（もとのデータが3次元で無い限り）

割合データの可視化 – 割合の可視化 ③

積み上げ棒グラフによる可視化



棒グラフによる可視化

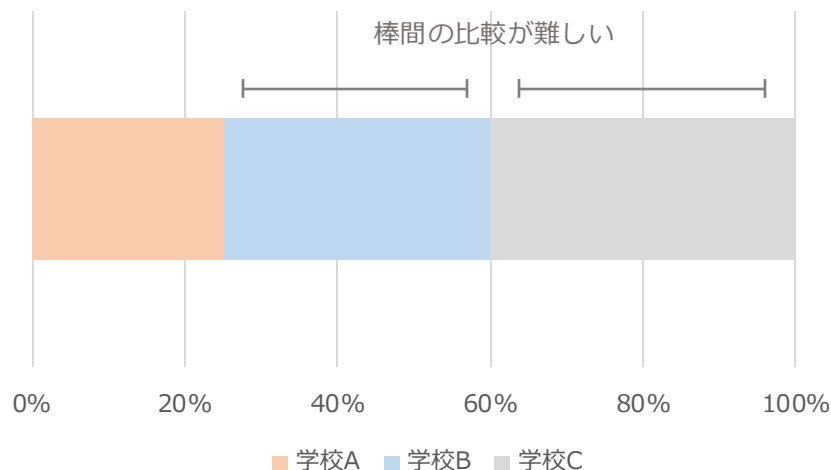


割合データの可視化 – 割合の可視化 ③

積み上げ棒グラフによる可視化



3つ以上の積み上げ棒グラフの場合、各幅ごとの比較は難しい。



積み上げ棒グラフだと差の比較が難しい場合がある。とくに割合ごとの差が小さい場合に注意。

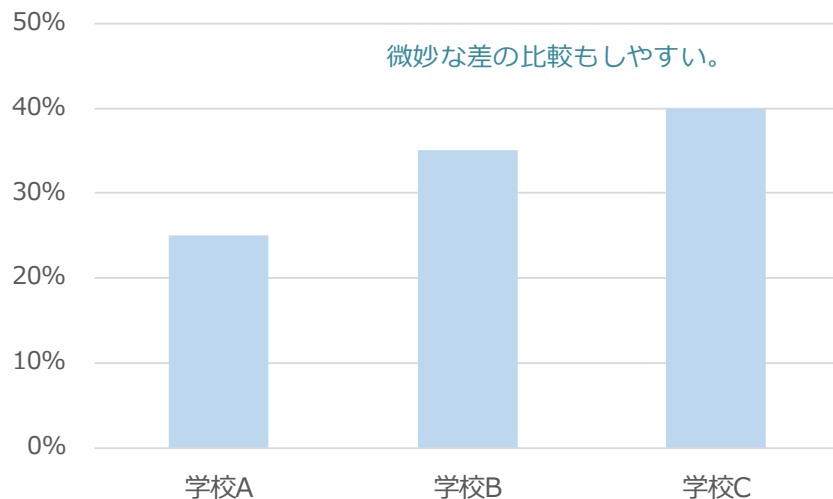
棒グラフによる可視化



棒グラフで表すと微妙な割合差の違いもわかりやすい。



全体の中に占める割合が表示されていない。

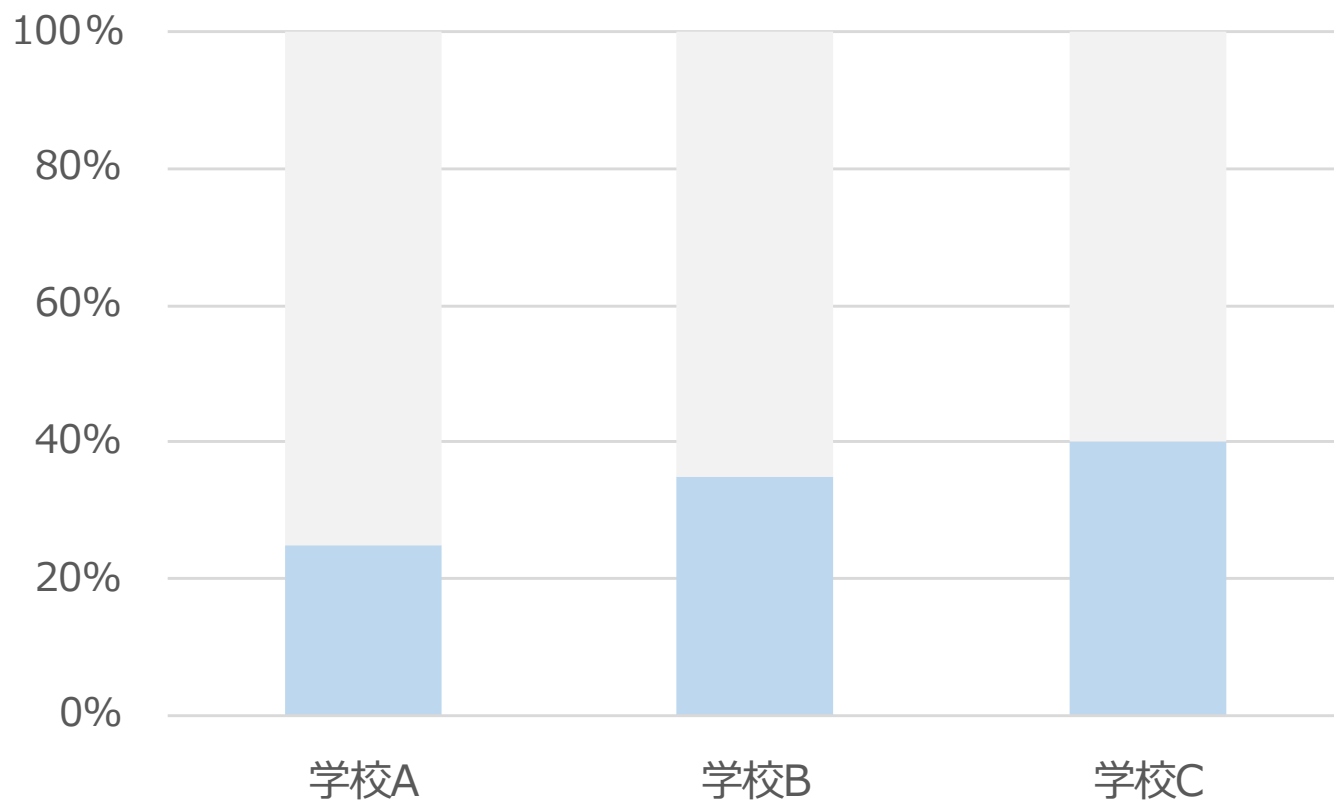


棒グラフを使えば、割合ごとの差が小さい場合でも、比較がしやすい。

割合データの可視化 – 割合の可視化 ③

棒グラフによる可視化

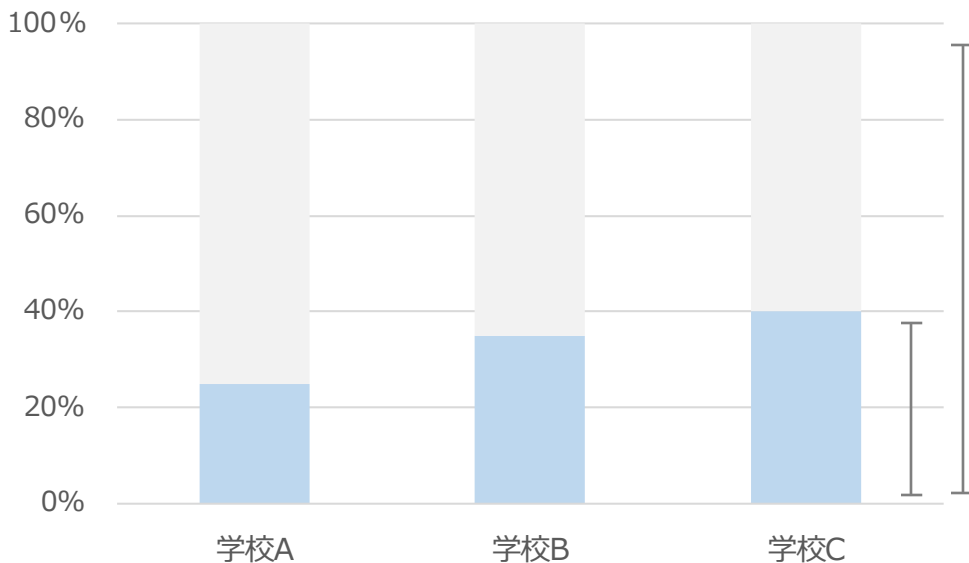
- 背景色の利用 -



割合データの可視化 – 割合の可視化 ③

棒グラフによる可視化

- 背景色の利用 -



棒グラフで表すと微妙な割合差の違いもわかりやすい。



背景色と対比することで、全体としての割合を把握できる。



棒グラフ+背景色で割合の差異と全体の中での割合を同時に表現できる。

割合データの可視化 – 割合の可視化 ③



円グラフでの表現だと扇形の面積の比較が難しい。とくに割合ごとの差が小さい場合に注意。



地形を図示したり、もとのデータが3次元で無い限り、基本的に3D表現は避ける。



割合ごとの差が小さい場合、積み上げ棒グラフだと差の比較が難しい。



一般に面積だと微妙な差の比較が難しい。



棒グラフを使えば、割合ごとの差が小さい場合でも、比較がしやすい。



棒グラフ+背景色で割合の差異と全体の中での割合を同時に表現できる。

練習問題で学ぶグラフデザイン

データサイエンティスト

高比良 健太郎

第3回

練習問題 4

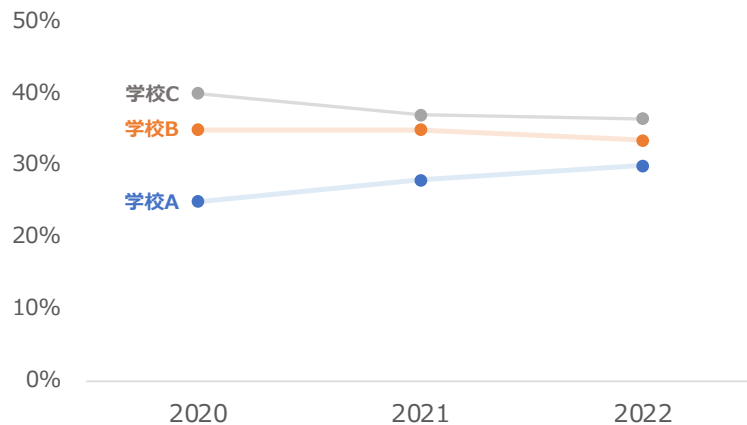
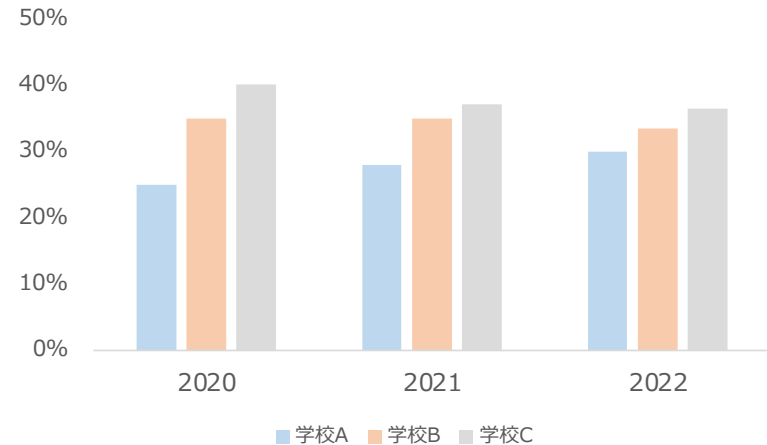
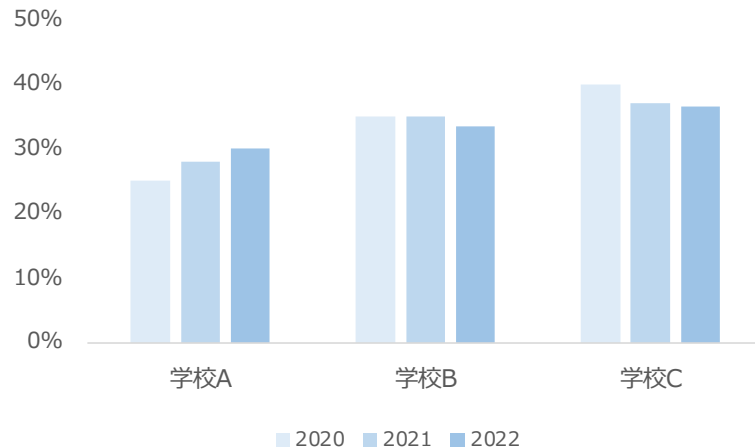
割合データの可視化 – 割合の経時変化の可視化 ④

以下はある地区に居住する子どもたちの学校別の人数とそれぞれの全体に占める割合のデータです。今回は3年分あります。各学校ごとの人数割合および、年度ごとの変化がよくわかるように可視化するにはどうすればいいのでしょうか。

学校名	2020		2021		2022	
	人数	割合	人数	割合	人数	割合
学校A	125	25.0%	146	28.0%	156	30.0%
学校B	175	35.0%	182	35.0%	174	33.5%
学校C	200	40.0%	192	37.0%	190	36.5%

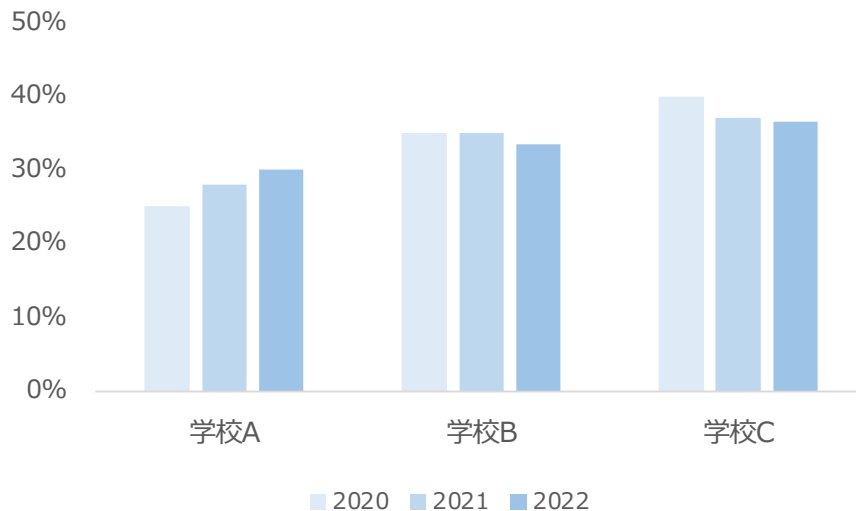
割合データの可視化 – 割合の経時変化の可視化 ④

回答例

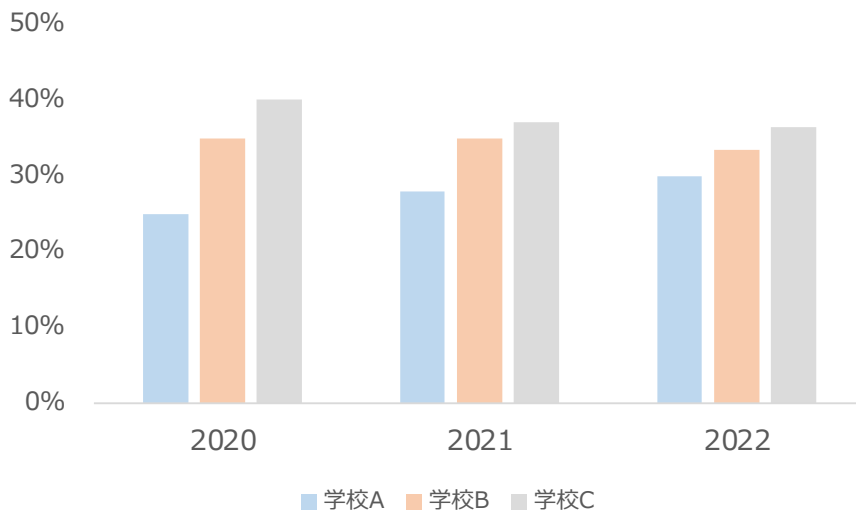


割合データの可視化 – 割合の経時変化の可視化 ④

棒グラフによる可視化
– ①学校 ②年度 –



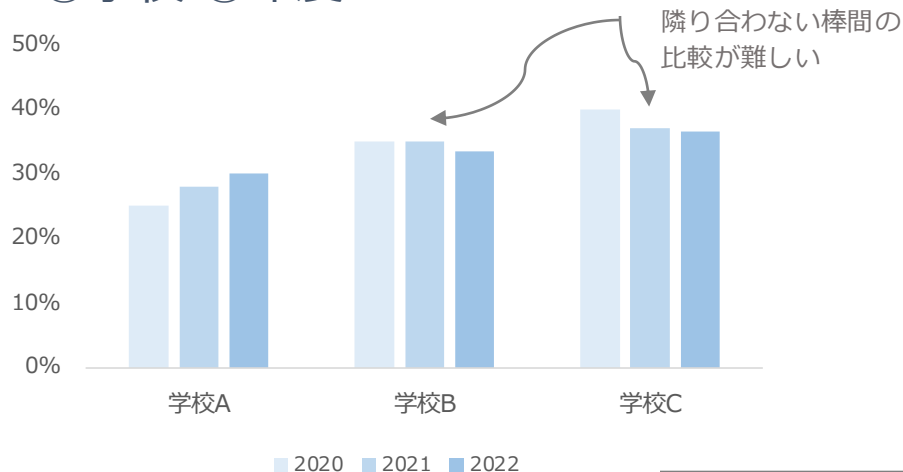
棒グラフによる可視化
– ①年度 ②学校 –



割合データの可視化 – 割合の経時変化の可視化 ④

棒グラフによる可視化

– ①学校 ②年度 –



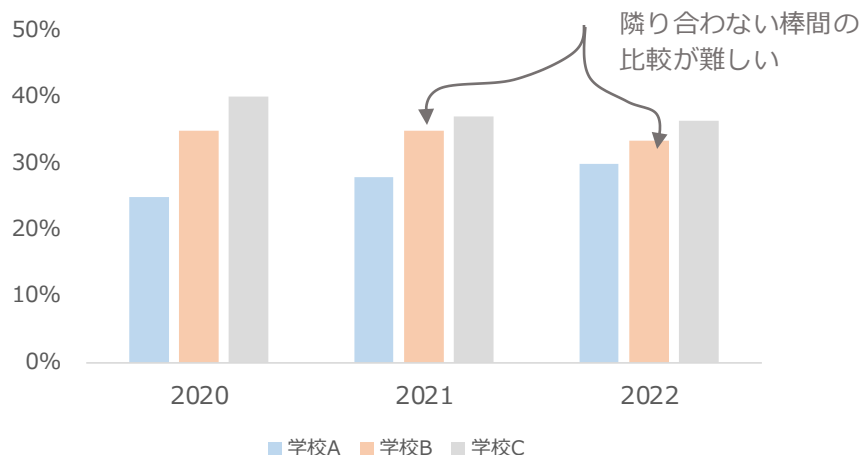
学校 → 年度の順番でデータを可視化しており。年度間の比較は行いやすい。



学校間の比較が難しい。
※同一年度の異なる学校が隣接していないため。

棒グラフによる可視化

– ①年度 ②学校 –



年度 → 学校の順番でデータを可視化している。学校間の比較は行いやすい。



年度間の比較が難しい。
※同一の学校の異なる年度が隣接していないため。

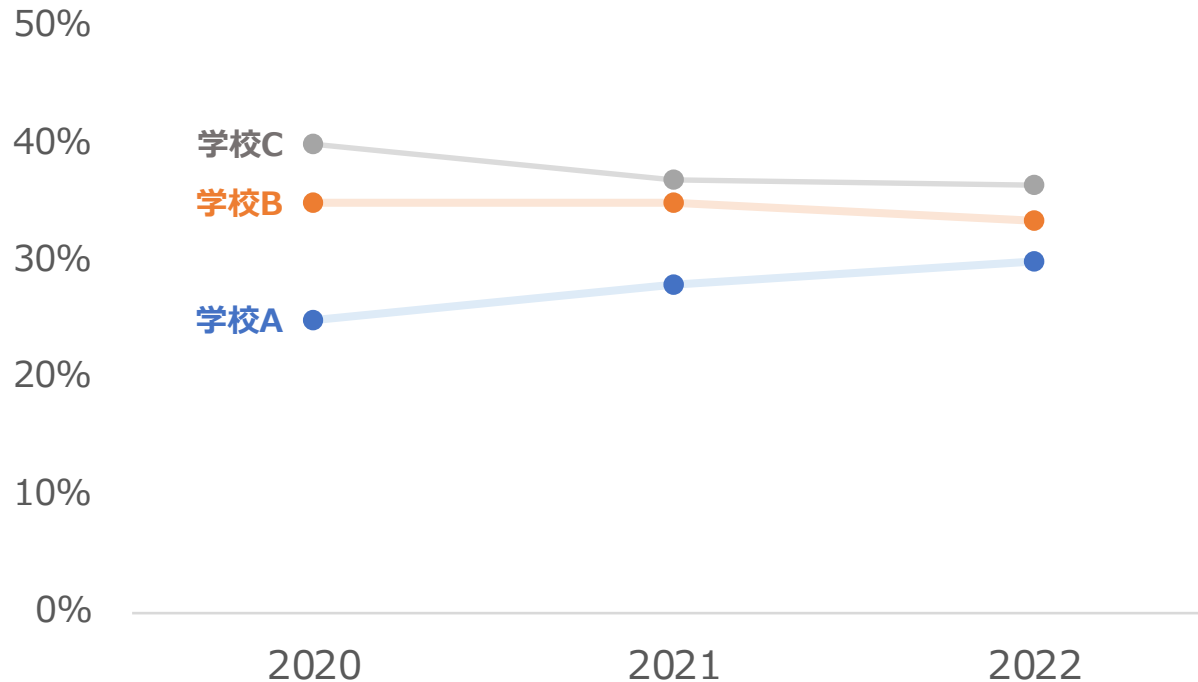


どのカテゴリに関する比較を行いたいかということに焦点を当てて軸を決定する。

割合データの可視化 – 割合の経時変化の可視化 ④

線グラフによる可視化

- 学校ごとの線グラフ -



割合データの可視化 – 割合の経時変化の可視化 ④

線グラフによる可視化

- 学校ごとの線グラフ -



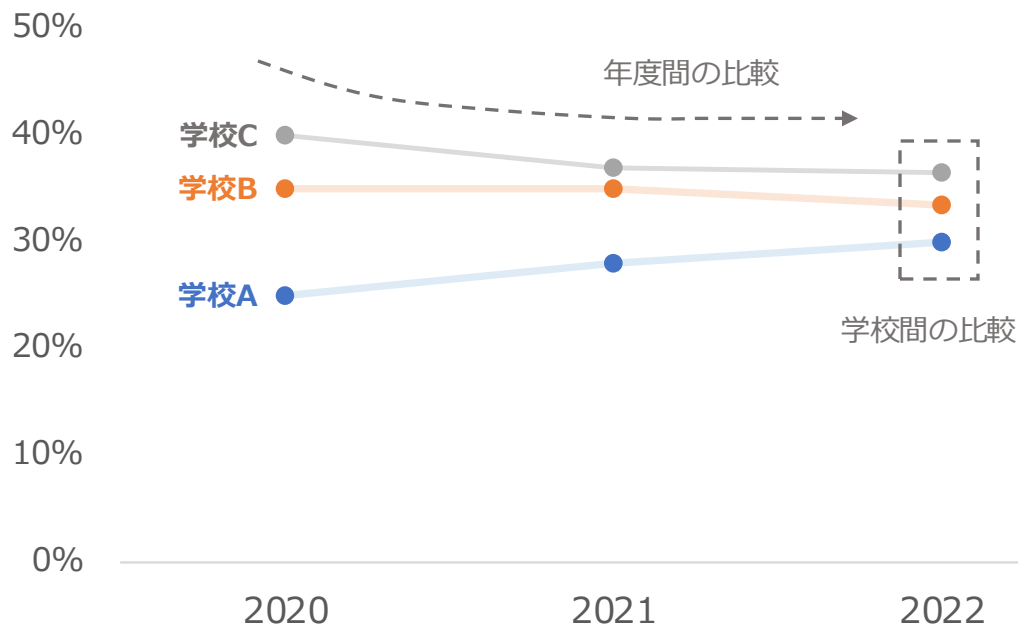
同じ年度の異なる学校間の比較 / 同じ学校の異なる年度間の比較のどちらも直感的に行うことができる。



複数のカテゴリのものの経時変化を表すときは、線グラフによる表現を試してみる。



年度をX軸に取ることで、連続グラフである線グラフの特徴を上手く活かしている。



割合データの可視化 – 割合の経時変化の可視化 ④



どのカテゴリに関する比較を行いたいかということに焦点を当てて軸を決定する。



複数のカテゴリのものの経時変化を表すときは、線グラフによる表現を試してみる。

練習問題 5

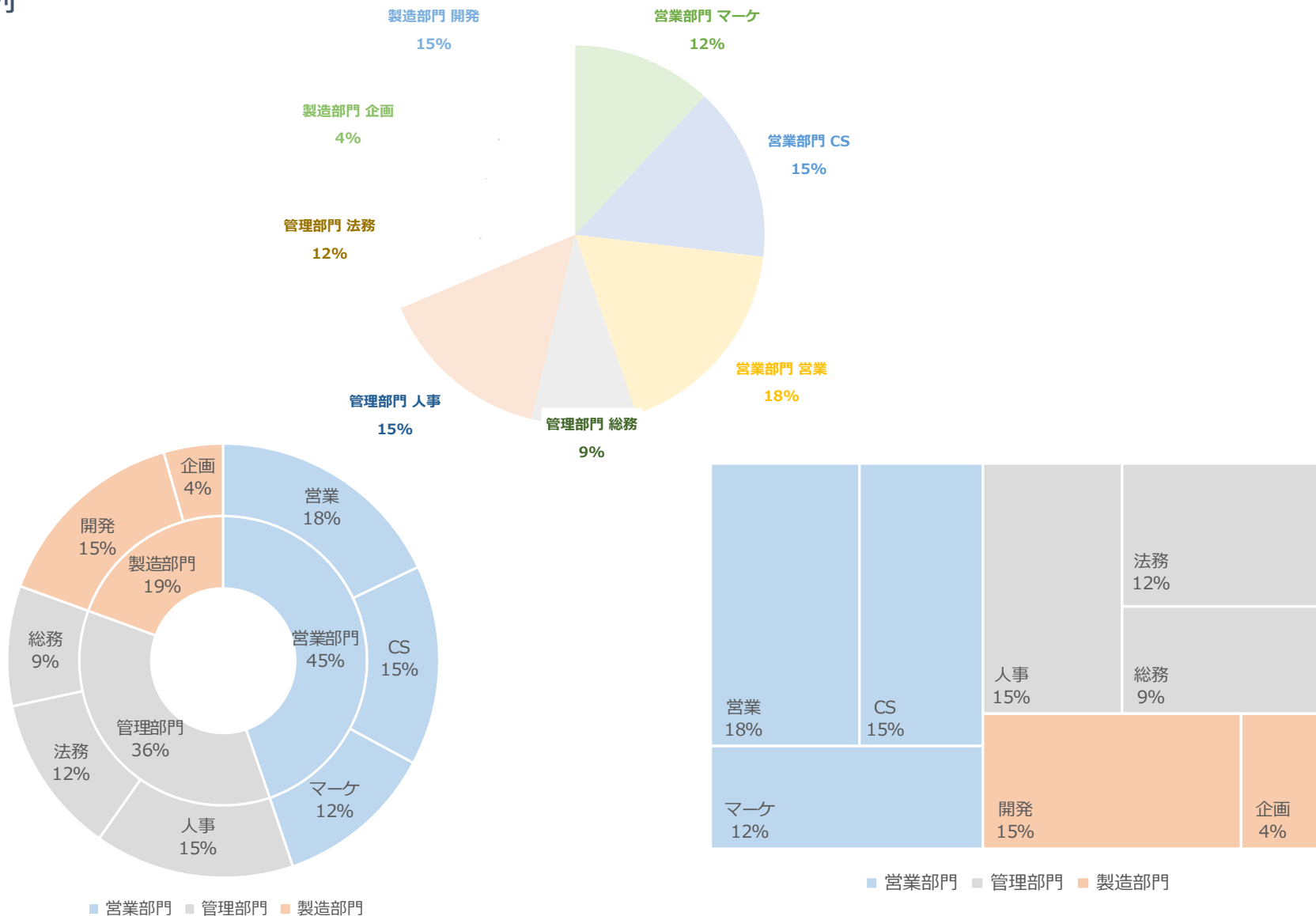
割合データの可視化 – 階層的な割合の可視化⑤

所属企業の部署ごとの従業員数は以下のものであったとします。各部署は大きな分類である部門に属しています。部署/部門の階層関係を踏まえ、所属企業の従業員割合を表すグラフを作成してみてください。

部門	部署	所属人数（人）	人数割合
営業部門	マーケ	8	12%
営業部門	CS	10	15%
営業部門	営業	12	18%
管理部門	総務	6	9%
管理部門	人事	10	15%
管理部門	法務	8	12%
製造部門	企画	3	4%
製造部門	開発	10	15%

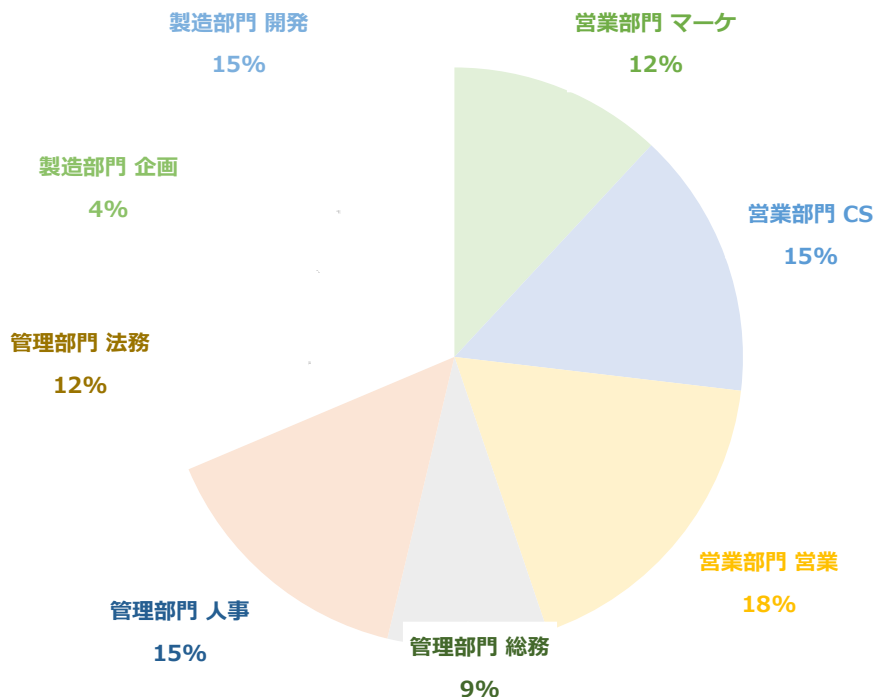
割合データの可視化 – 階層的な割合の可視化⑤

回答例

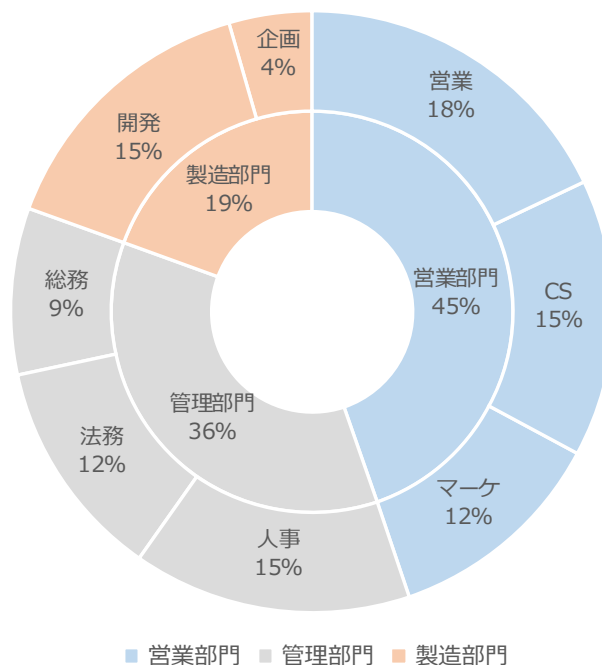


割合データの可視化 – 階層的な割合の可視化⑤

円グラフによる可視化



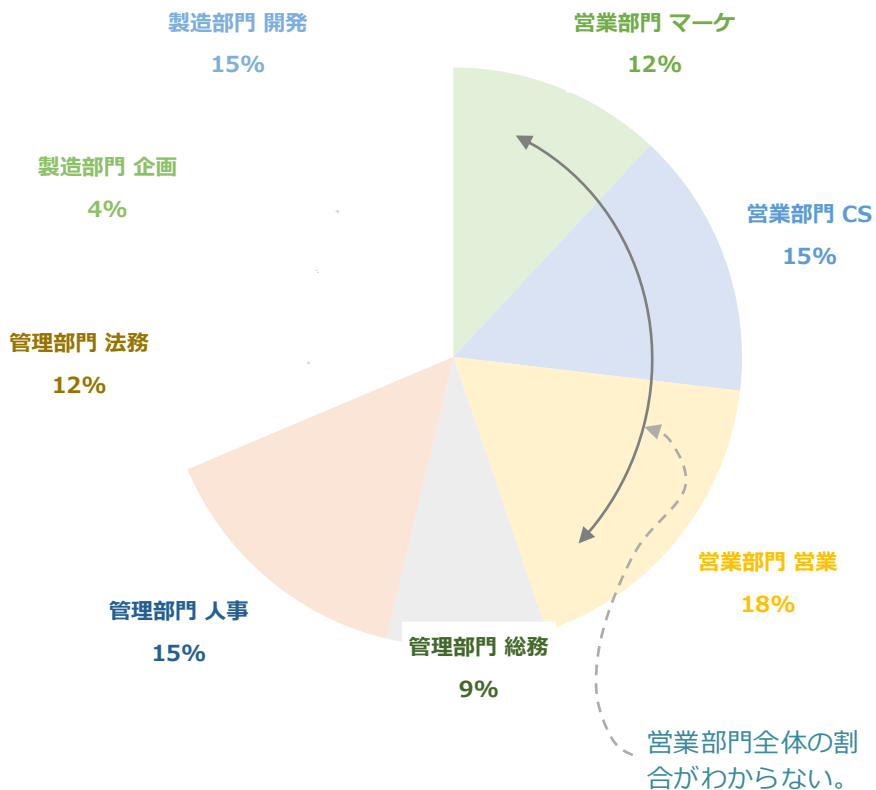
サンバースト図による可視化



割合データの可視化 – 階層的な割合の可視化⑤

円グラフによる可視化

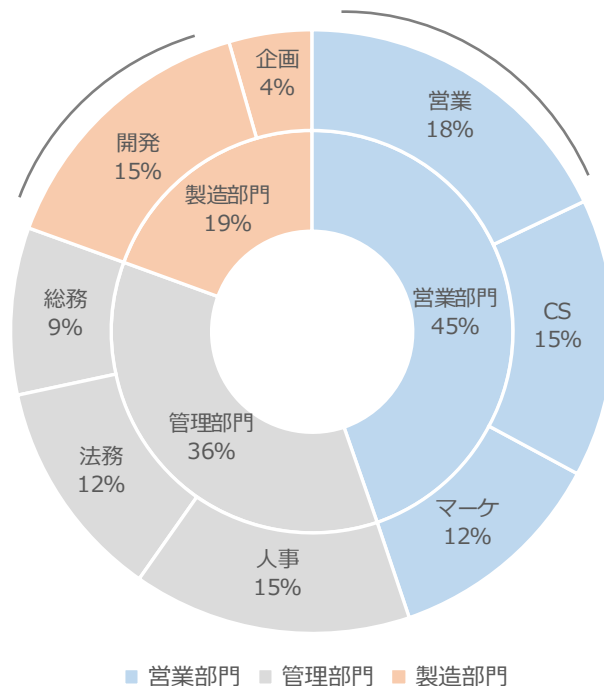
⚠ 部門（上位階層）ごとの割合がわかりにくい。



サンバースト図による可視化

☑ 各階層（部門および部署）の割合が理解しやすい。

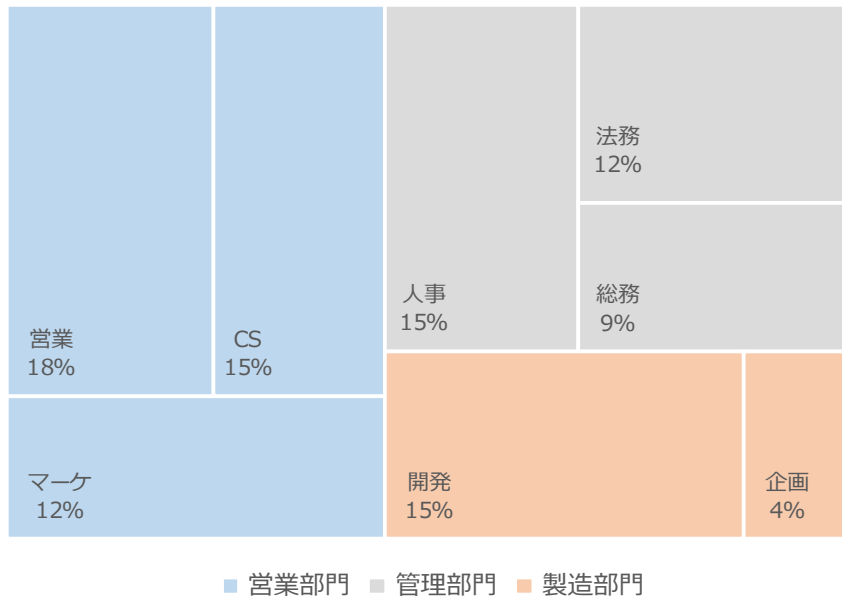
⚠ ドーナツ部分間の比較が難しい。



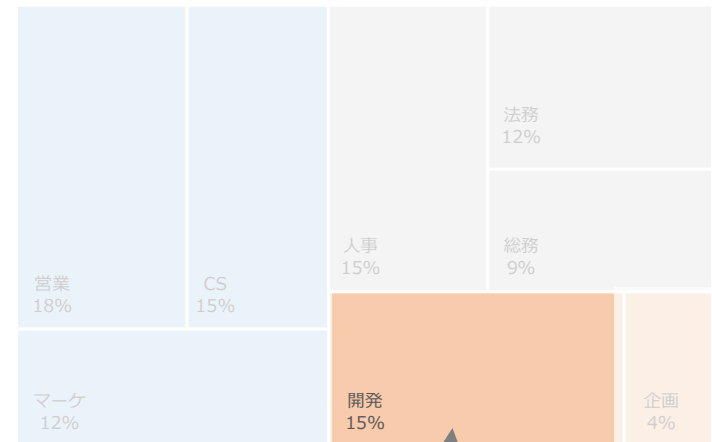
💡 階層的なデータに対しては、各階層の割合を上手く表現するためのグラフの使用を検討する。

割合データの可視化 – 階層的な割合の可視化⑤

ツリーマップによる可視化



ツリーマップ (Treemap)



■ 営業部門 ■ 管理部門 ■ 製造部門

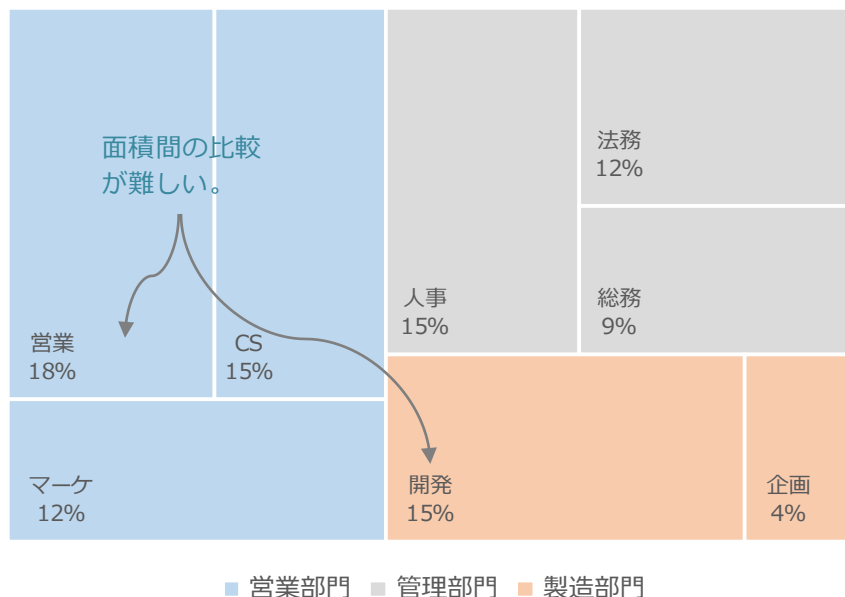
1つの長方形の面積が割合の大きさを表す。

割合データの可視化 – 階層的な割合の可視化⑤

ツリーマップによる可視化

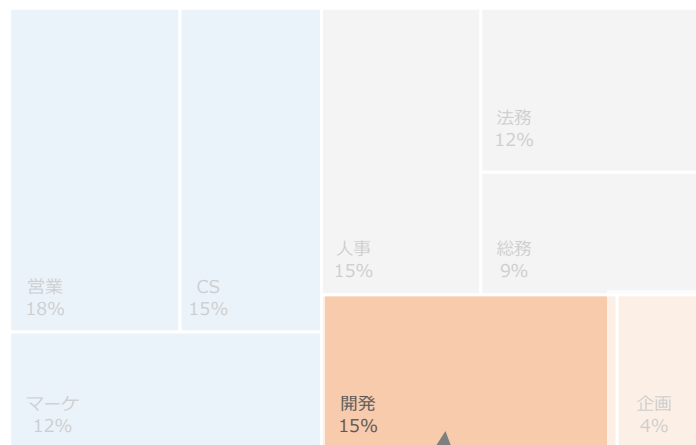
✓ 各階層（部門および部署）の割合が理解しやすい。

⚠ 長方形同士の比較が難しい。



💡 面積を用いた比較は、人間の目には難しい場合があることに注意！
(数値表示などを検討してみる)

ツリーマップ (Treemap)



■ 営業部門 ■ 管理部門 ■ 製造部門

1つの長方形の面積が割合の大きさを表す。

割合データの可視化 – 階層的な割合の可視化⑤



階層的なデータに対しては、各階層の割合を上手く表現するためのグラフの使用を検討する。



面積を用いた比較は、人間の目には難しい場合があることに注意！
(数値表示などを検討してみる)

練習問題 6

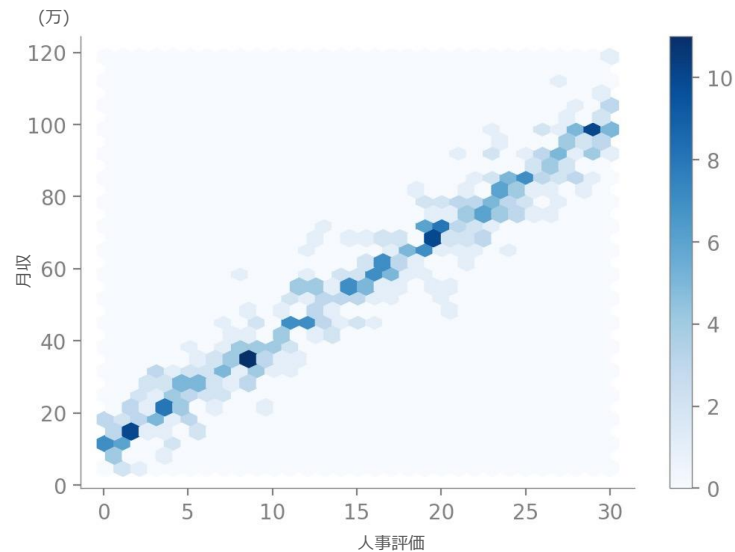
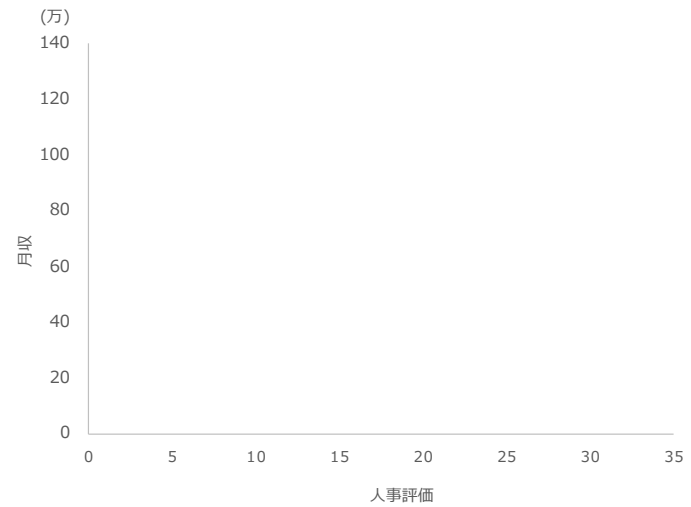
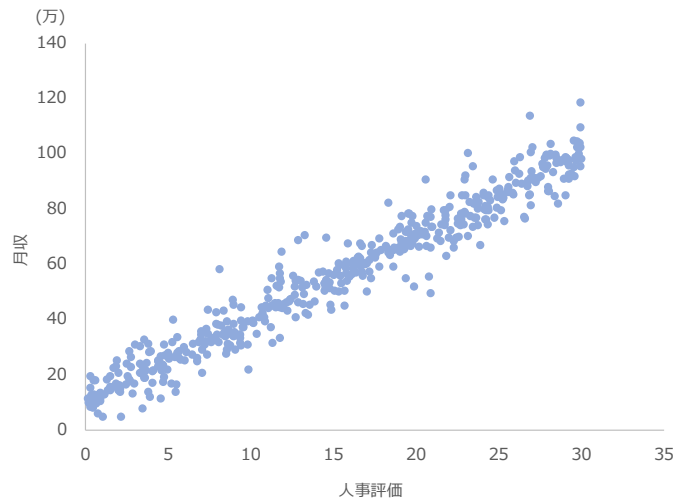
データの関係の可視化 - 2つのデータの関連の可視化 ⑥

ある会社での人事評価と月収に関する以下のようなデータ（600人分）があります。人事評価と月収の関係性を調べるためにどのようなグラフを作製すればよいでしょうか。

従業員ID	人事評価	月収	給与明細		
			基本給	残業代	手当
10010	10.12	39.7			
10000	0.49	18.0			
10026	25.95	94.0			
10017	16.77	55.9	10585	24.08	85.8
10011	11.49	44.5	10586	9.42	30.7
10022	22.29	68.6	10587	11.65	46.1
10023	22.94	79.1	10588	9.55	37.3
10006	5.56	33.8	10589	27.76	98.6
10002	1.69	22.5	10590	0.20	10.0
10003	2.55	19.5	10591	25.59	88.1
10019	18.93	72.8	10592	14.12	57.1
			10593	23.21	82.6
			10594	3.45	7.9
			10595	13.97	52.0
			10596	18.89	66.0
			10597	15.36	59.7
			10598	22.52	74.8
			10599	19.63	68.9
			10600	17.03	60.4

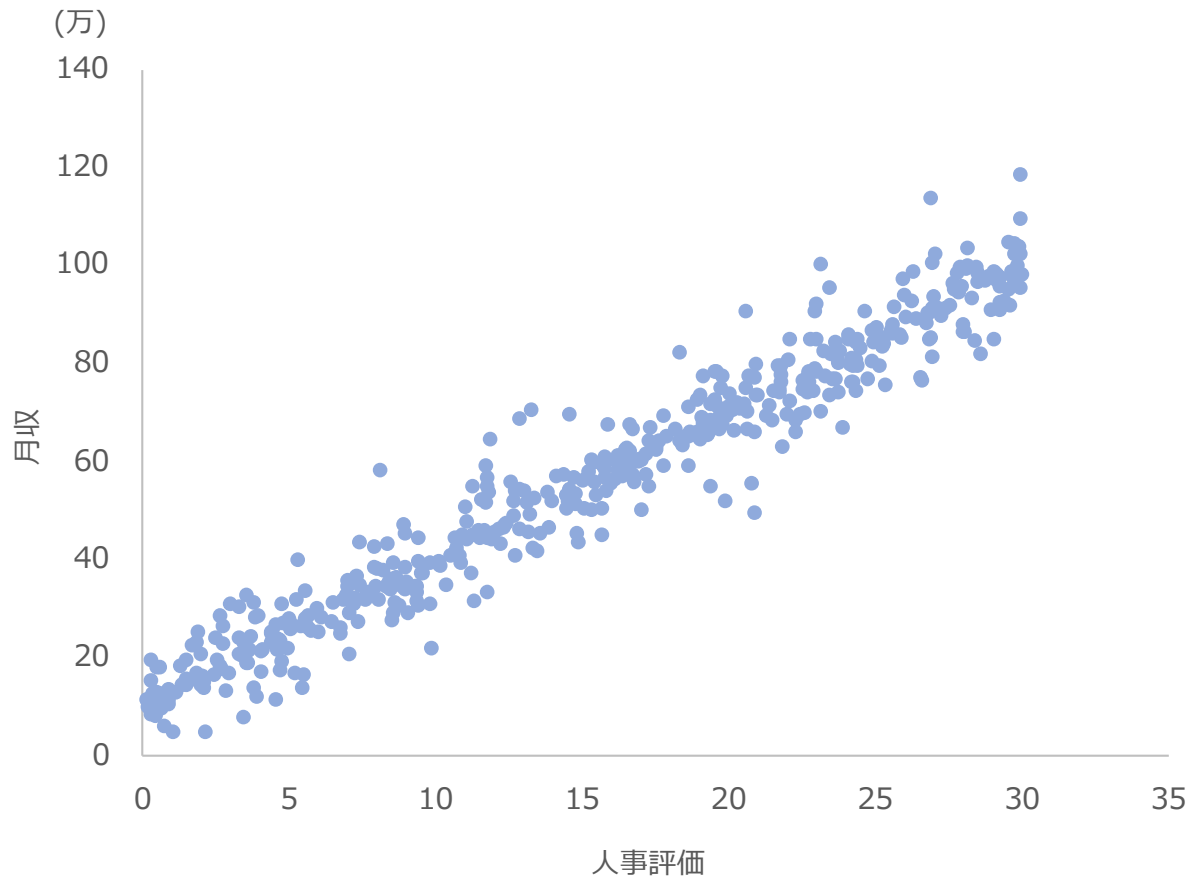
データの関係の可視化 – 2つのデータの関連の可視化 ⑥

回答例



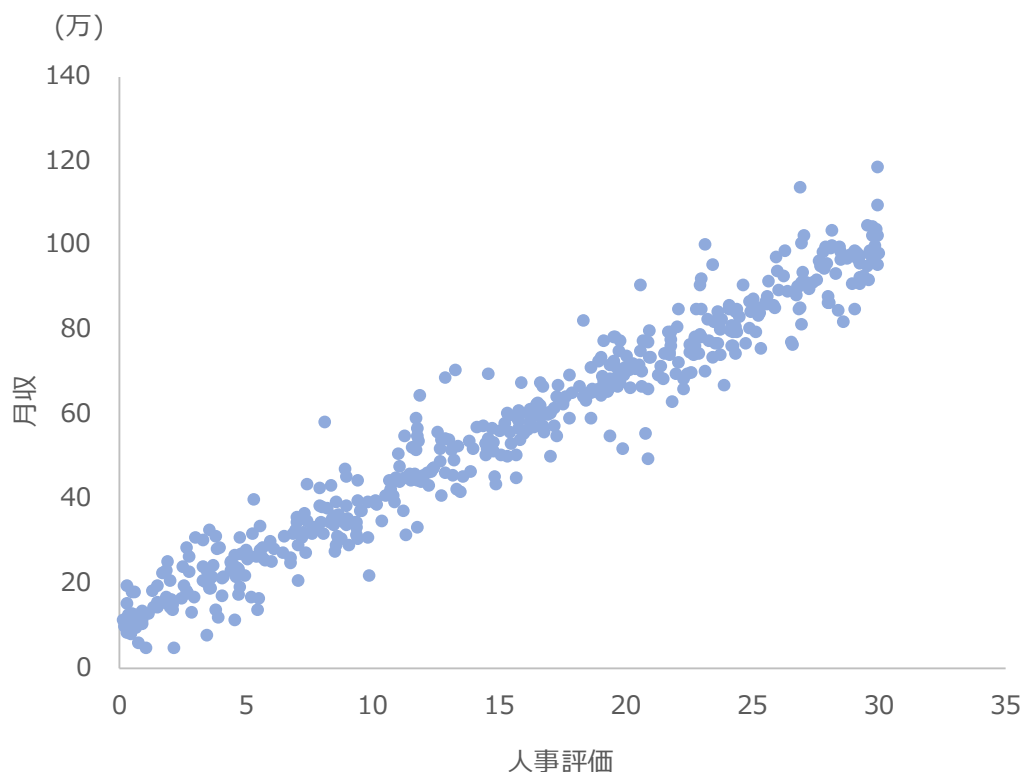
データの関係の可視化 – 2つのデータの関連の可視化 ⑥

散布図による可視化



データの関係の可視化 – 2つのデータの関連の可視化 ⑥

散布図による可視化



月収と人事評価の関係性がわかりやすい。



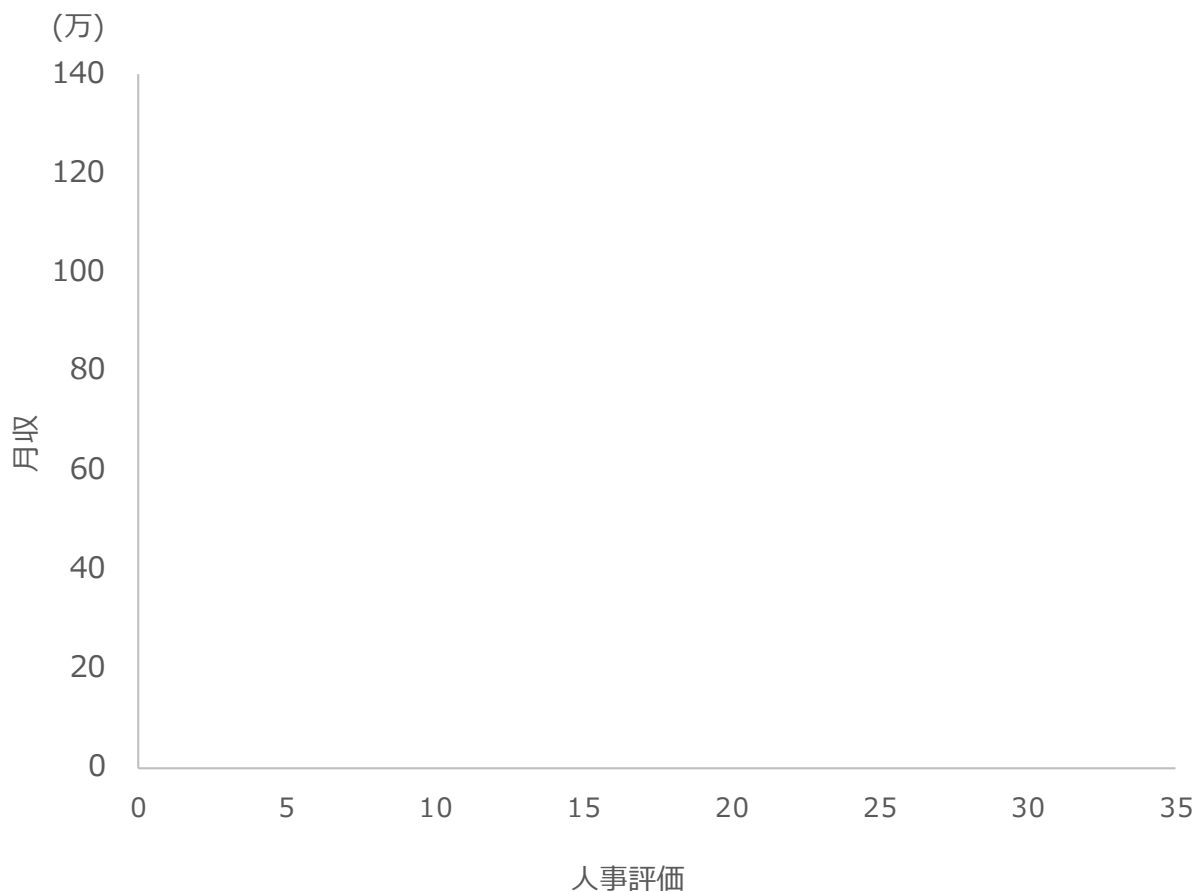
点同士が重なっているため、正確なばらつきを把握することができない。



描写するデータが多い場合は、点の重なりによってデータが見えなくなる場合があることに注意する。

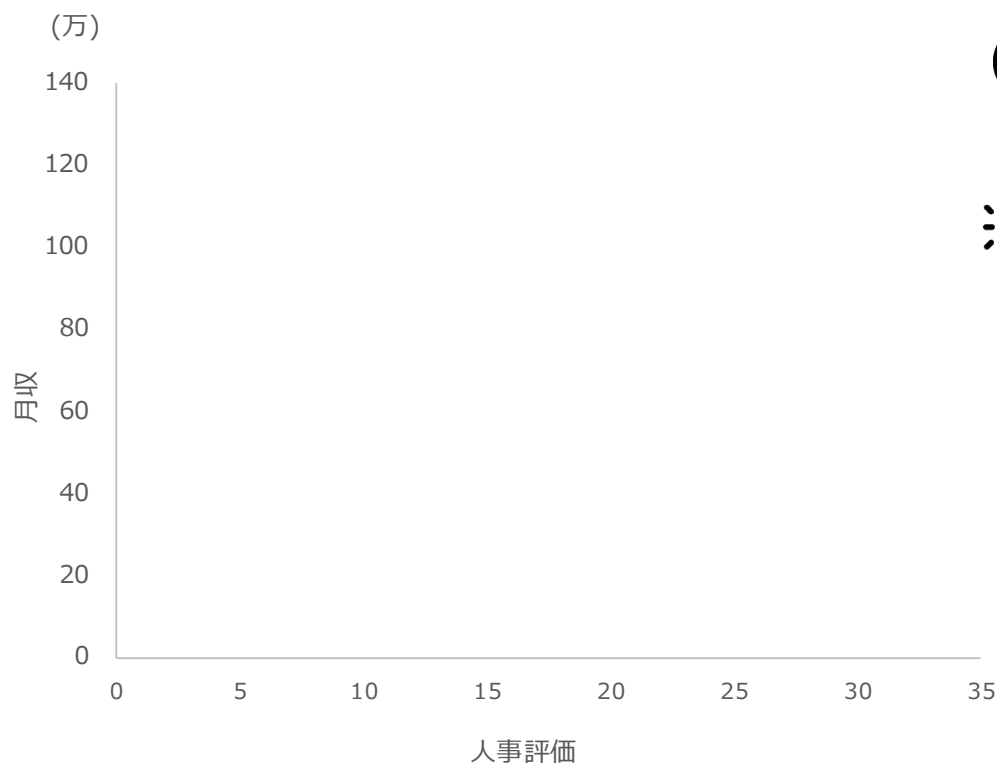
データの関係の可視化 – 2つのデータの関連の可視化 ⑥

散布図による可視化（半透明）



データの関係の可視化 – 2つのデータの関連の可視化 ⑥

散布図による可視化（半透明）



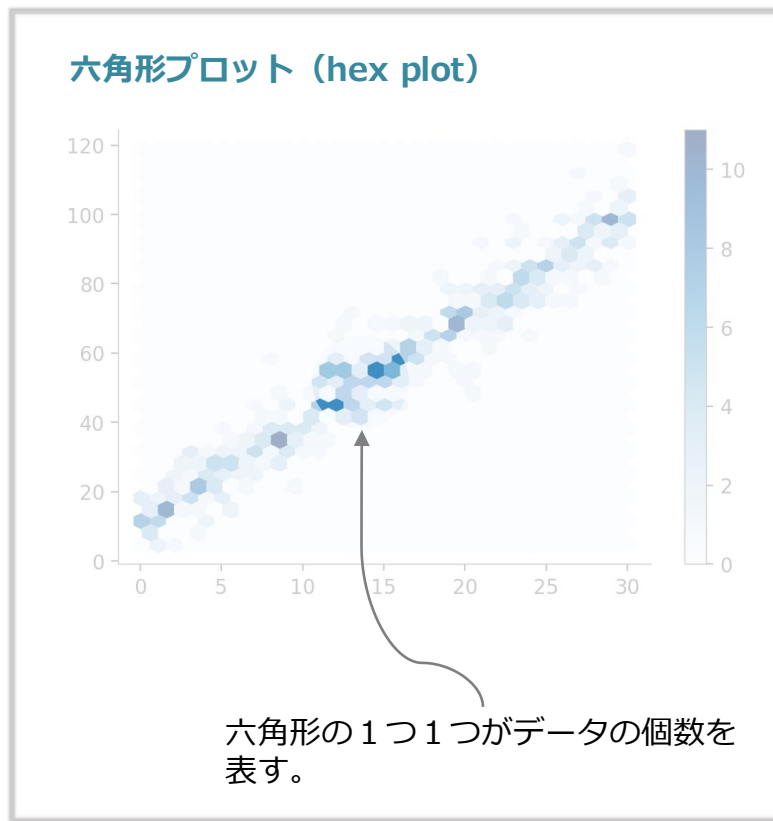
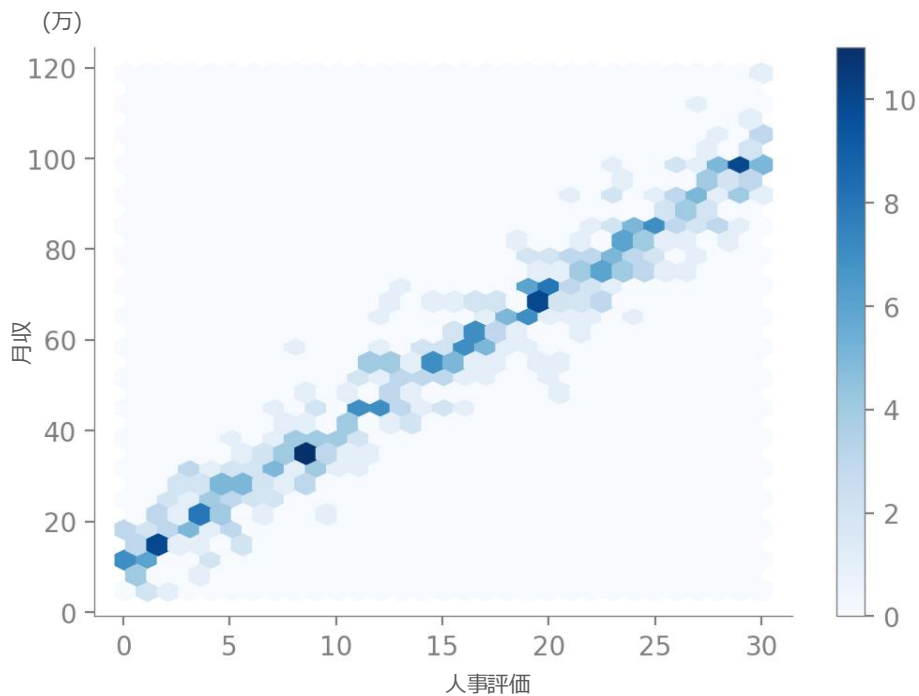
散布図に濃淡がつき、ばらつきをより正確に把握することができる。



データ量が多い場合は、マーカーの色を半透明にするとデータのばらつきが表現しやすい。

データの関係の可視化 – 2つのデータの関連の可視化 ⑥

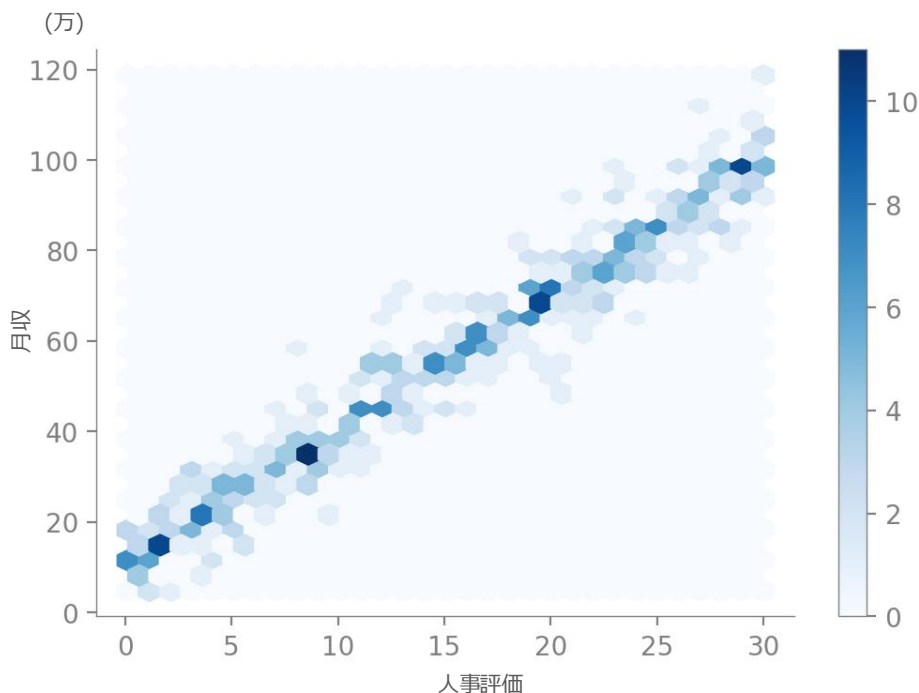
六角形プロットによる可視化



データの関係の可視化 – 2つのデータの関連の可視化 ⑥

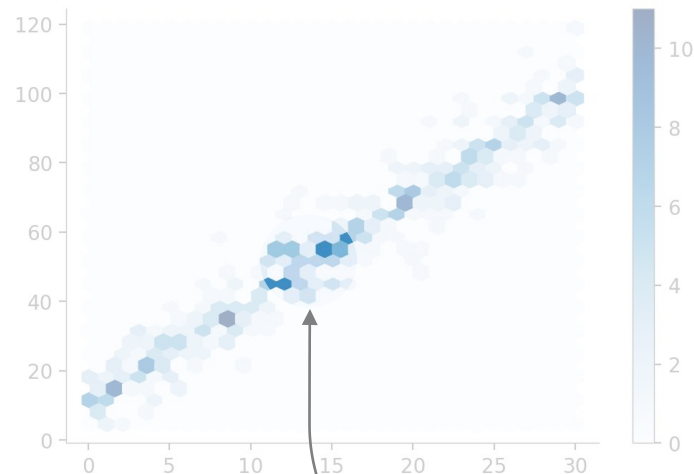
六角形プロットによる可視化

- ✓ 各格子ごとのデータ量が把握できる。ばらつきをより正確に捉えることができる。
- ✓ データ量が多い場合でも、色合いで分布を表現することができる。



- 💡 描写するデータ量が多いときは、六角形散布図の使用を検討する。

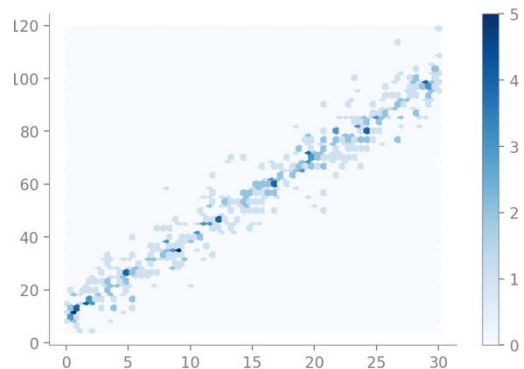
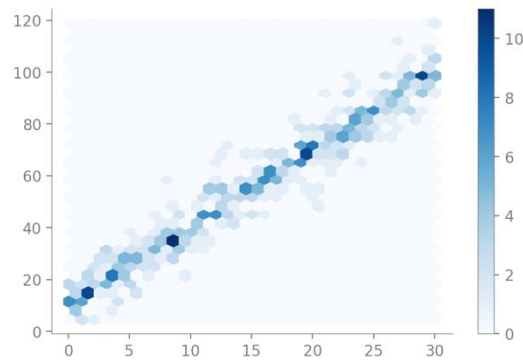
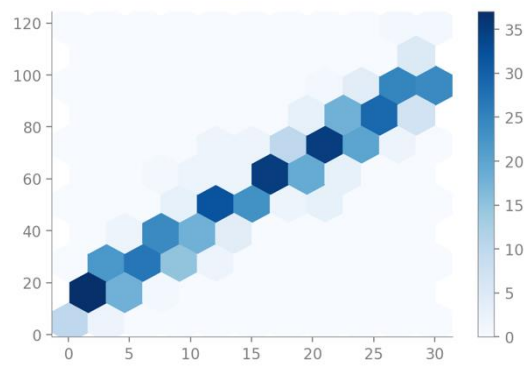
六角形プロット (hex plot)



六角形の1つ1つがデータの個数を表す。

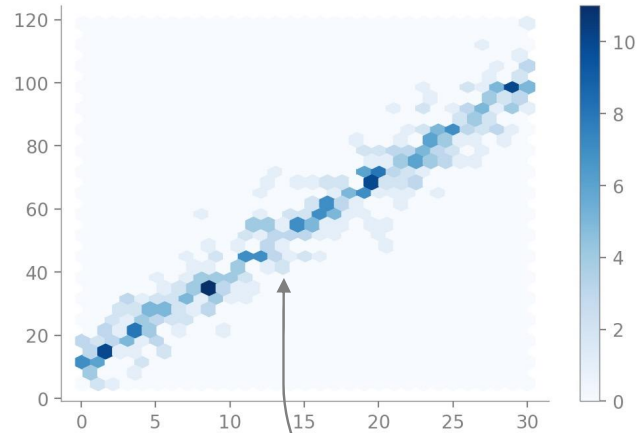
データの関係の可視化 – 2つのデータの関連の可視化 ⑥

六角形プロットによる可視化



データ点（六角形）の
大きさが異なる。

六角形プロット（hex plot）



六角形の1つ1つがデータの個数を
表す。



六角形プロットは、一つ一つのデータ点（六角形）の大きさによってばらつきの印象が異なってしまいます。適切なデータ幅を検討する。

データの関係の可視化 – 2つのデータの関連の可視化 ⑥



描写するデータが多い場合は、点の重なりによってデータが見えなくなる場合があることに注意する。



データ量が多い場合は、マーカーの色を半透明にするとデータのばらつきが表現しやすい。



描写するデータ量が多いときは、六角形散布図の使用を検討する。



六角形散布図は、一つ一つのデータ点（六角形）の大きさによってばらつきの印象が異なってしまう。適切なデータ幅を検討する。

練習問題で学ぶグラフデザイン

データサイエンティスト

高比良 健太郎

第4回

練習問題 7

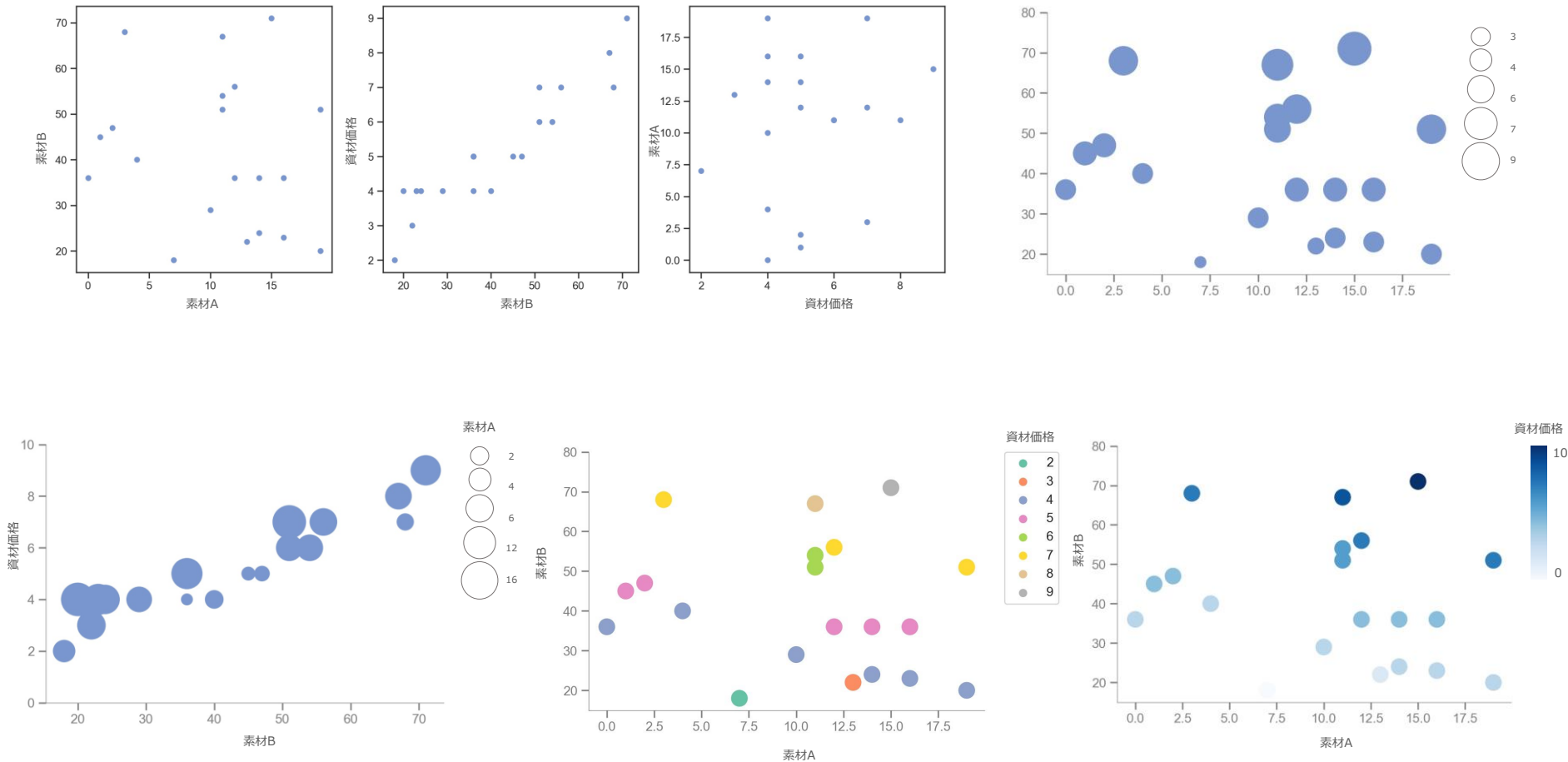
データの関係の可視化 – 3つのデータの関連の可視化 ⑦

ある工場において、下記のような資材を生産しています。主要な構成物として素材Aと素材Bがあり、それらを使ってできた資材価格が以下のようにになっています。素材A, 素材Bの関係および、それらを合わせたときの資材価格との関係を見るためにはどのような可視化を行えばよいでしょうか。

資材ID	素材A (mg/g)	素材B (mg/g)	資材価格 (\$/Kg)
101	4	40	4.5
102	3	68	7.1
103	2	47	4.9
104	15	71	8.6
105	13	22	3.5
106	11	67	7.8
107	11	54	6.5
108	16	23	3.9
109	16	36	5.3
110	10	29	3.9
111	0	36	3.7
112	14	36	5.0
113	1	45	4.6
114	11	51	6.2
115	14	24	3.7
116	12	56	6.8
117	7	18	2.5
118	19	20	3.9
119	12	36	4.8
120	19	51	7.0

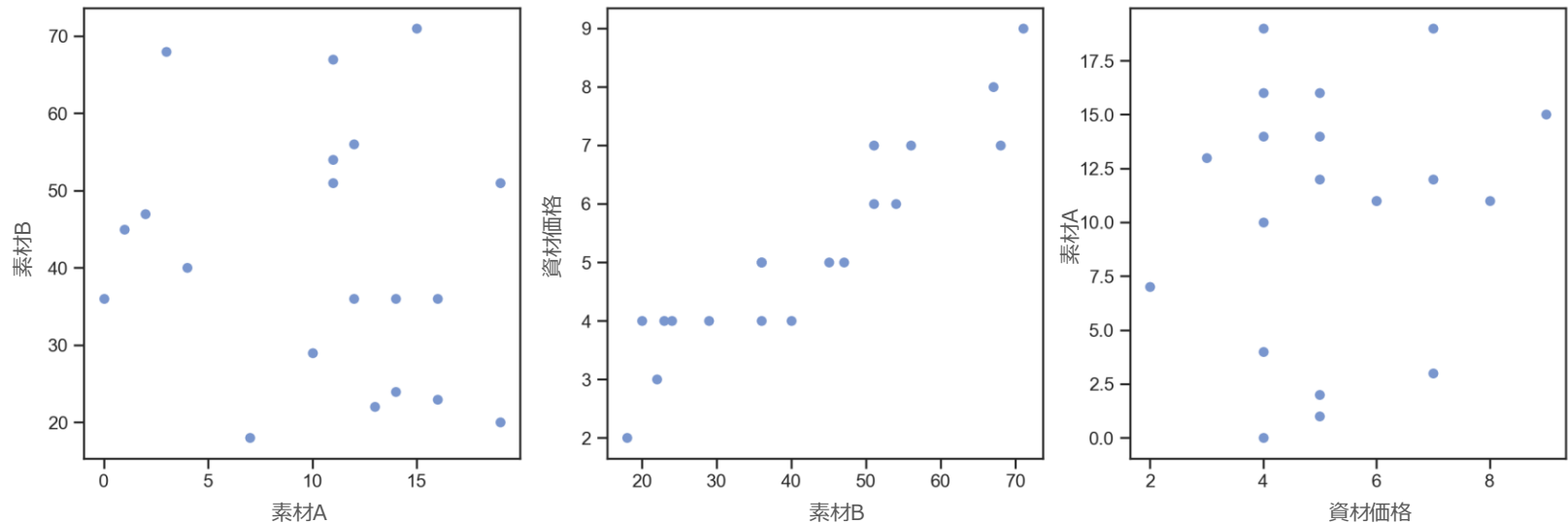
データの関係の可視化 – 3つのデータの関連の可視化 ⑦

回答例



データの関係の可視化 – 3つのデータの関連の可視化 ⑦

複数の散布図を用いた可視化

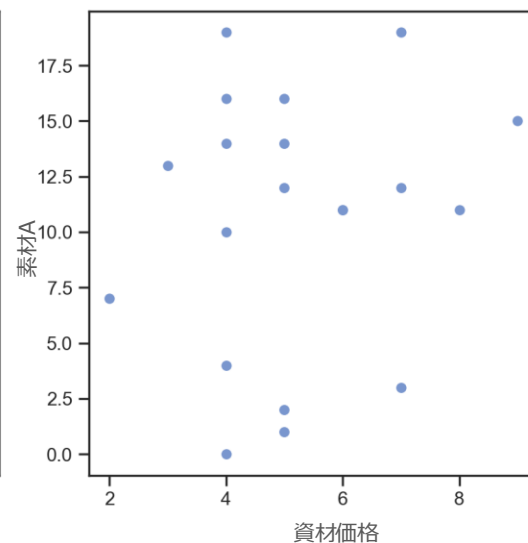
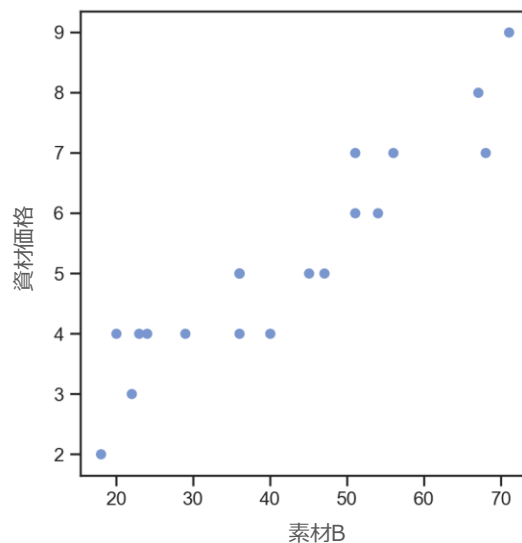
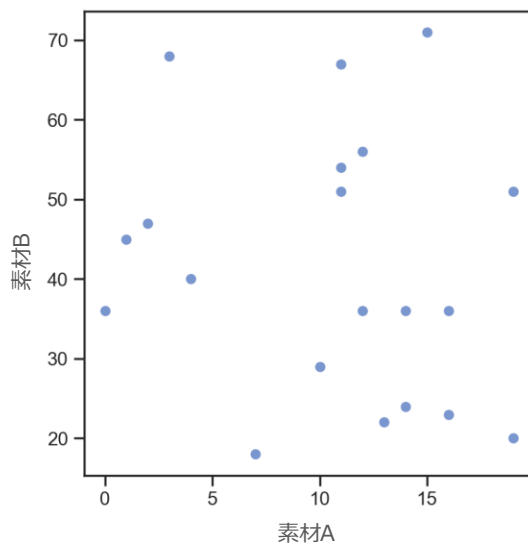


データの関係の可視化 – 3つのデータの関連の可視化 ⑦

複数の散布図を用いた可視化



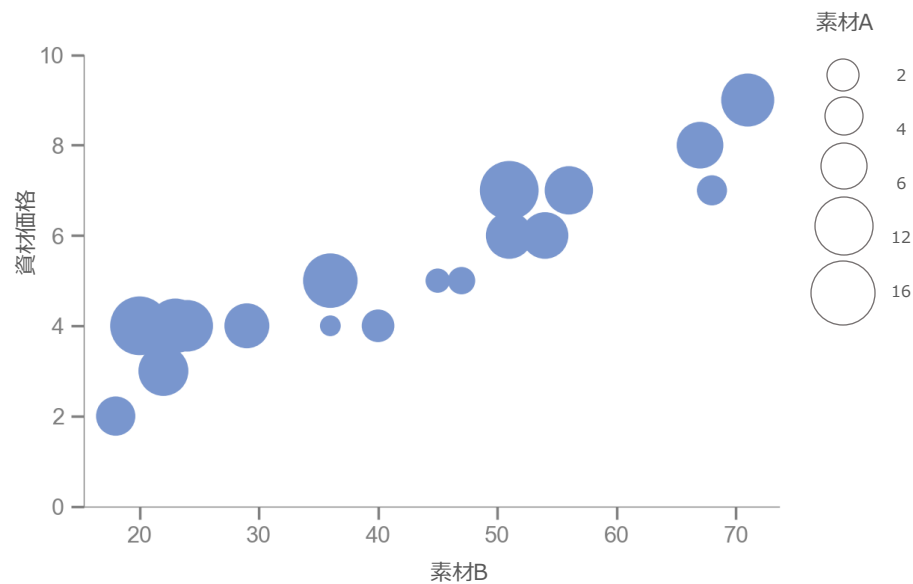
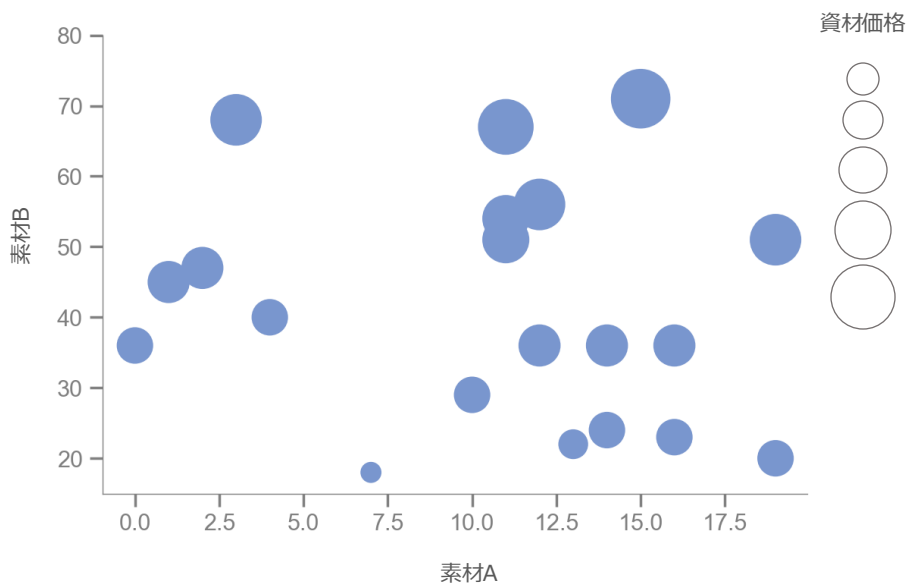
各ペア間の関係がよくわかる。



3つのデータ間の関連が見えない。

データの関係の可視化 – 3つのデータの関連の可視化 ⑦

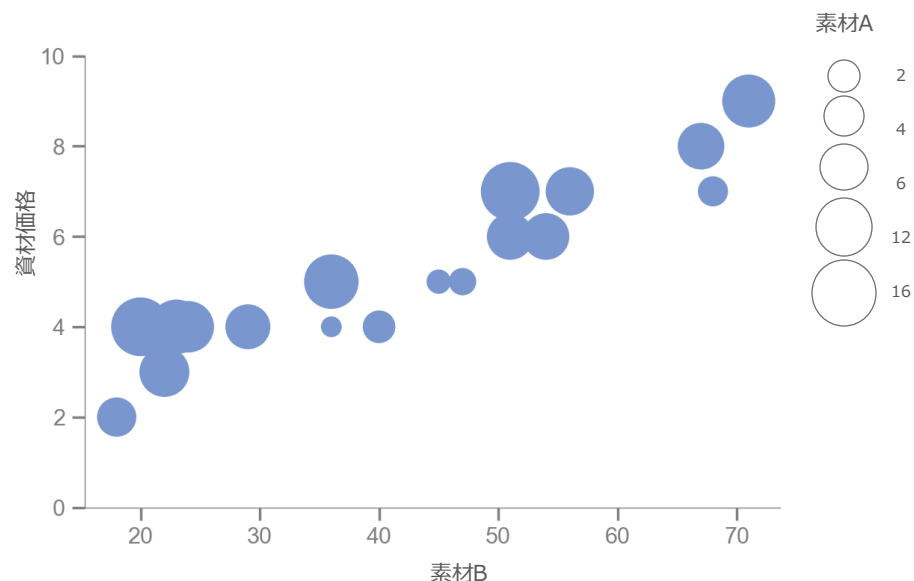
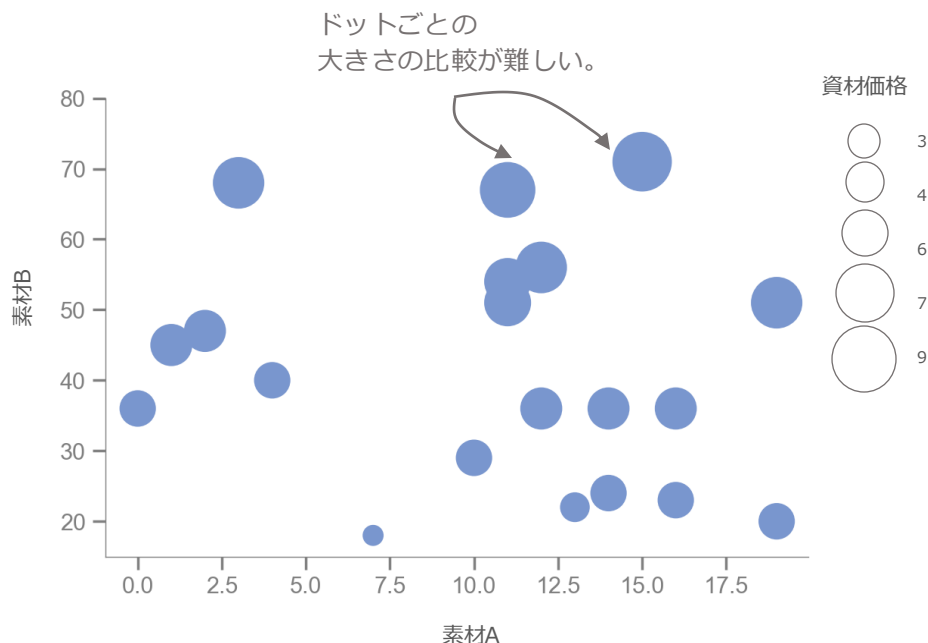
バブルチャートによる可視化



データの関係の可視化 – 3つのデータの関連の可視化 ⑦

バブルチャートによる可視化

✔ 3つのデータの関連を一度に可視化できている。



⚠ ドットごとの大きさの区別が難しい。

💡 バブルチャートは3つの特徴の可視化に有効。

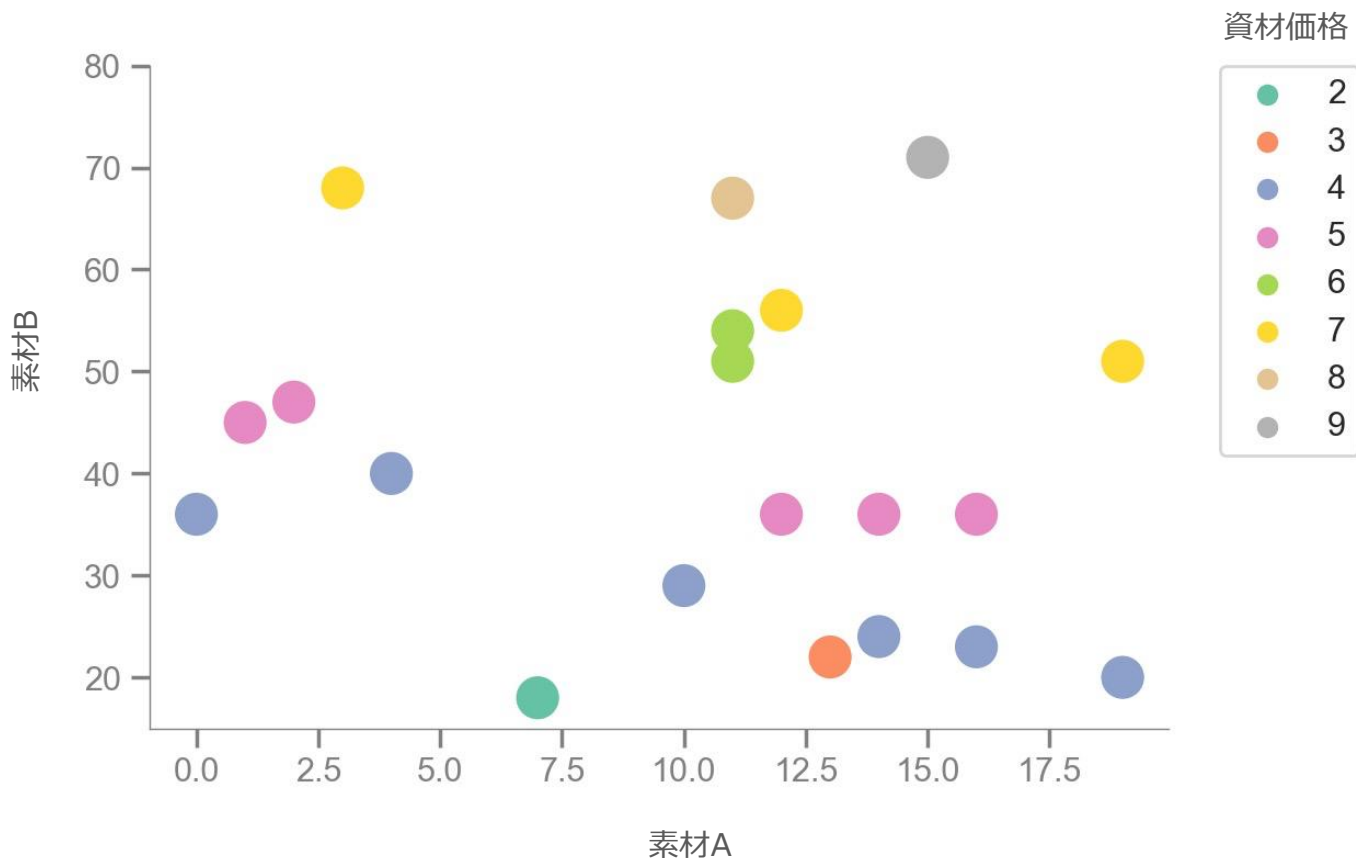
💡 3つ以上のデータを散布図により可視化するとき、軸に用いるデータにより大きく印象が異なる。
どのデータを軸に使用するか、要検討。

⚠ 軸に用いるデータによって大きく印象が
ことなる。

データの関係の可視化 – 3つのデータの関連の可視化 ⑦

色付き散布図による可視化

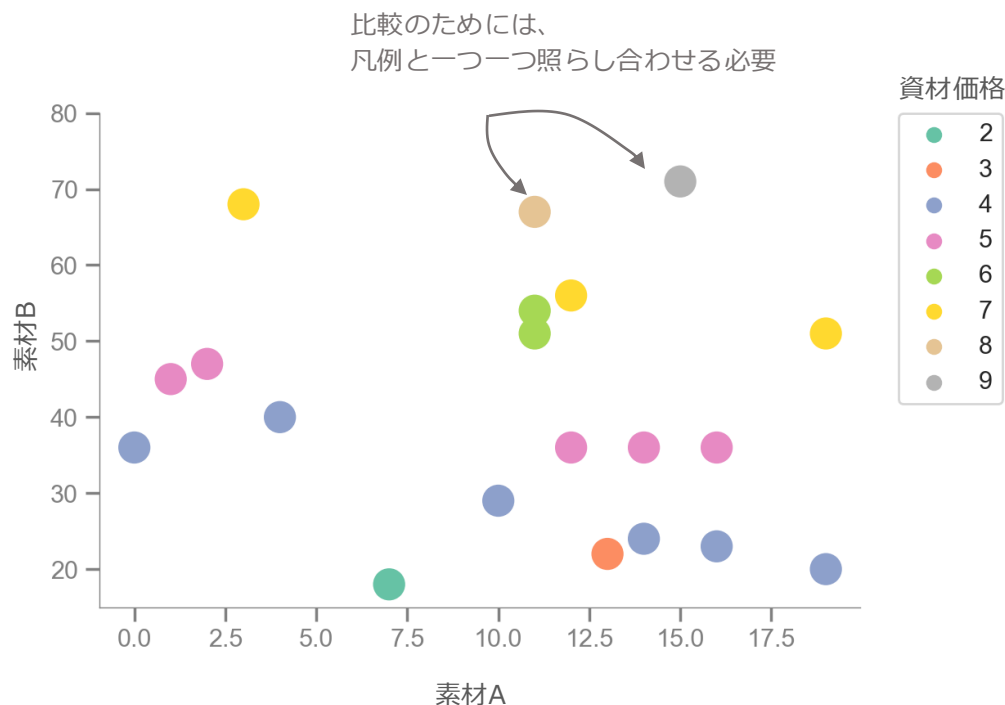
- 資材価格に応じて色付け -



データの関係の可視化 – 3つのデータの関連の可視化 ⑦

色付き散布図による可視化

- 資材価格に応じて色付け -



3つのデータの関連を一度に可視化できている。

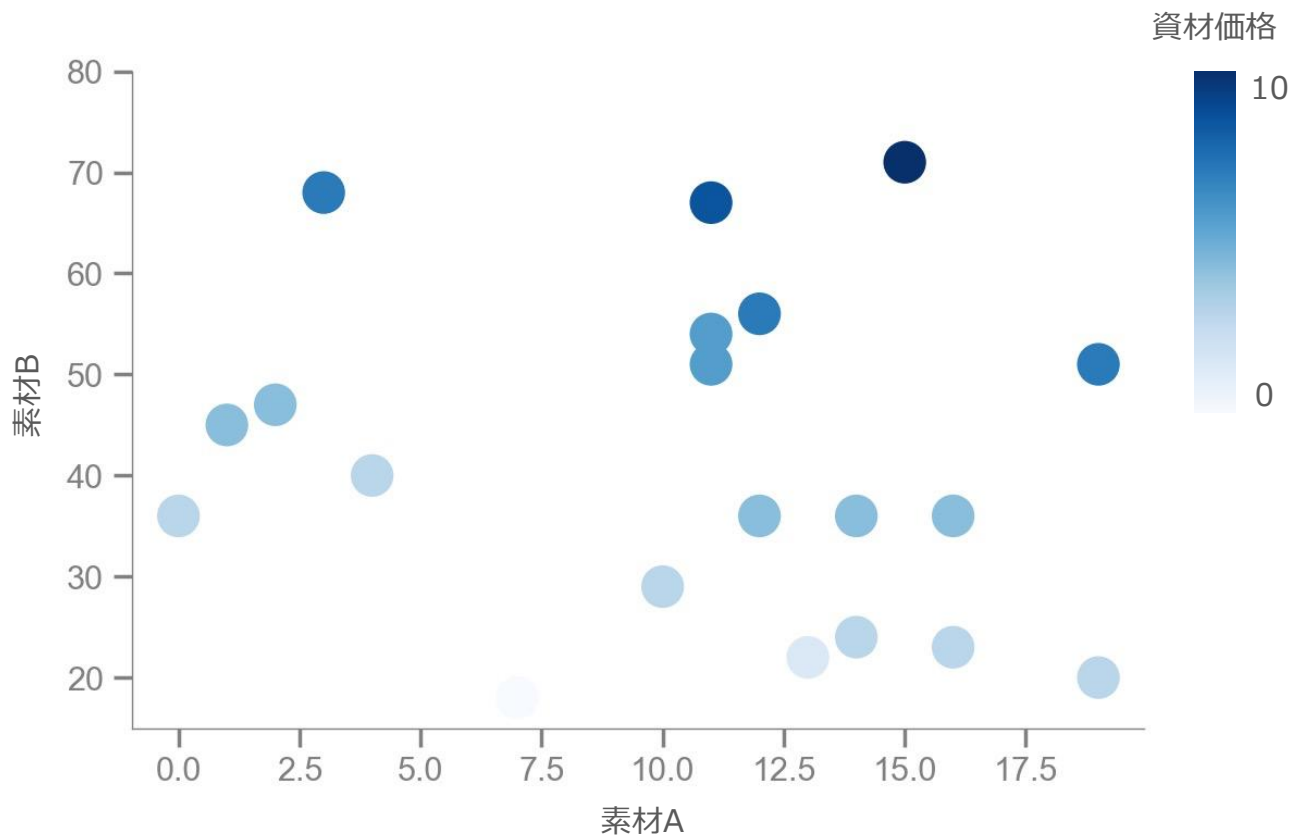


使用されている色と資材価格に関連がなく、資材価格の大小がわかりにくい。

データの関係の可視化 – 3つのデータの関連の可視化 ⑦

色付き散布図による可視化

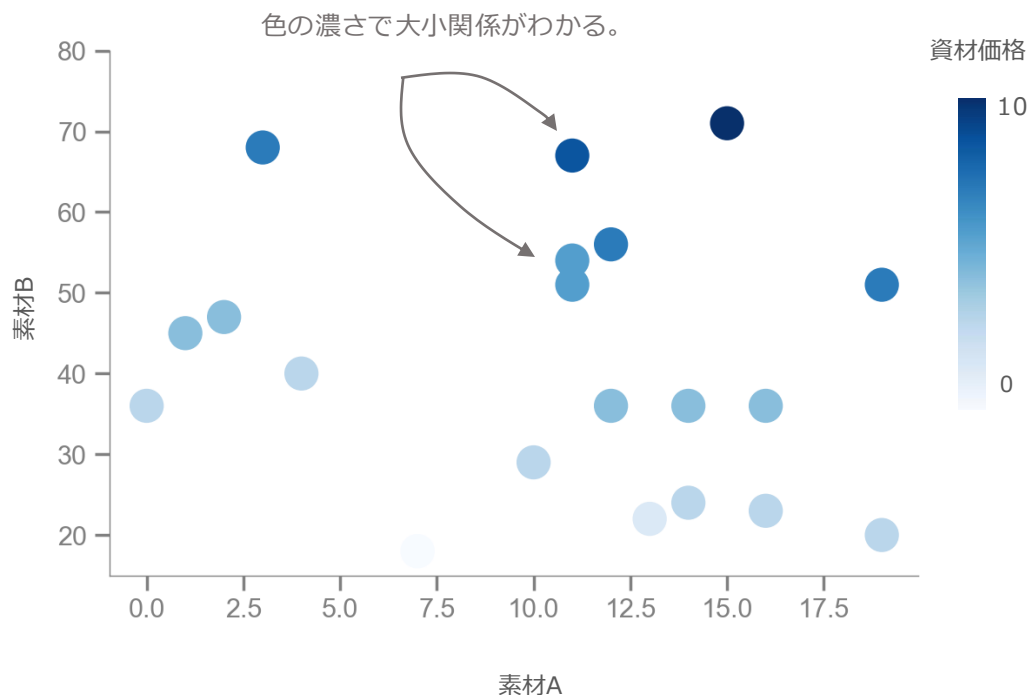
- 資材価格に応じてドットの色をグラデーション -



データの関係の可視化 – 3つのデータの関連の可視化 ⑦

色付き散布図による可視化

- 資材価格に応じてドットの色をグラデーション -



色の使われ方に順序性があり、値の大小が読み取りやすい。



散布図に用いる色のグラデーションにより、3つめのデータを表現できる。



色の使用時には、色のセット（パレット）が持つ意味を考える。

データの関係の可視化 – 3つのデータの関連の可視化 ⑦

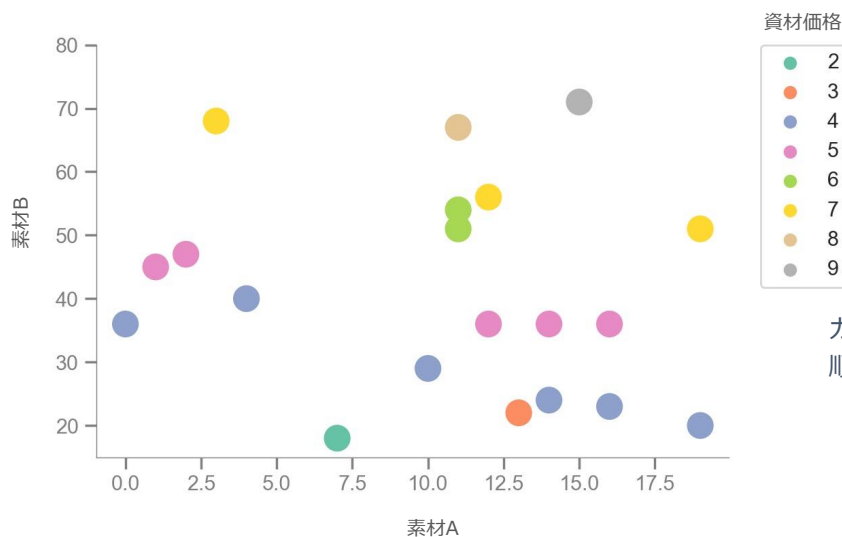
グラフのデザインを行うときは、色の構成（パレット）も重要です。代表的な色の構成として以下の3つがあり、それぞれ使用の際にポイントがあります。

	カテゴリーパレット	順序つきパレット	2方向パレット
例			
特徴	- 色の間に順序性がない。 - 順序のないカテゴリーデータを表すのに向いている。	- 色の間に順序性がある。 - 起点となる色に対して1方向の順序を持つ。 1方向に変化する量を表すときに向いている。	- 色の間に順序性がある。 - 起点となる色に対して2方向の順序を持つ。 2方向に変化する量を表すときに向いている。
使用例	- 男性と女性で色を変えたいとき。 - 都道府県ごとに別の色を使いたいとき。	- 年齢に沿って色を変えたいとき。 - 重さや音量などを表現したいとき。	- 温度によって色を変えたいとき。0度を原点にマイナスは青っぽく、プラスは赤っぽく。 - 利益によって色を変えたいとき。0円を原点に赤字を赤っぽく、黒字を青っぽく。

データの関係の可視化 – 3つのデータの関連の可視化 ⑦

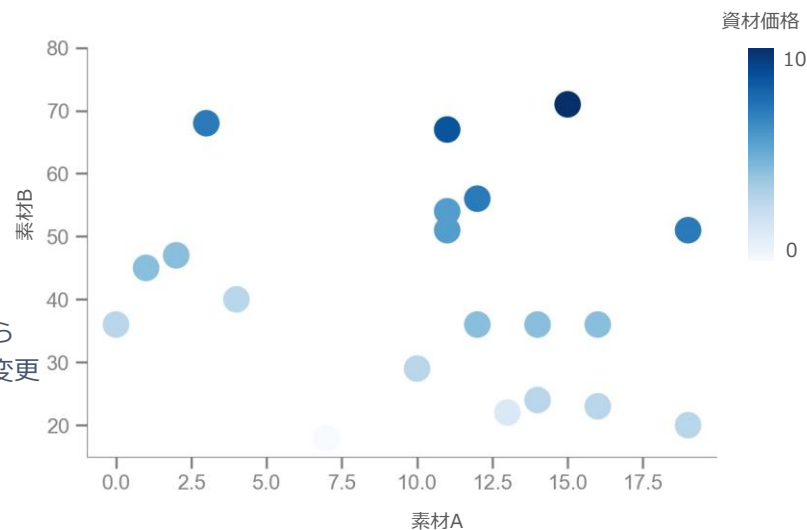
色付き散布図による可視化

カテゴリーパレット



カテゴリーから
順序つきへの変更

順序つきパレット



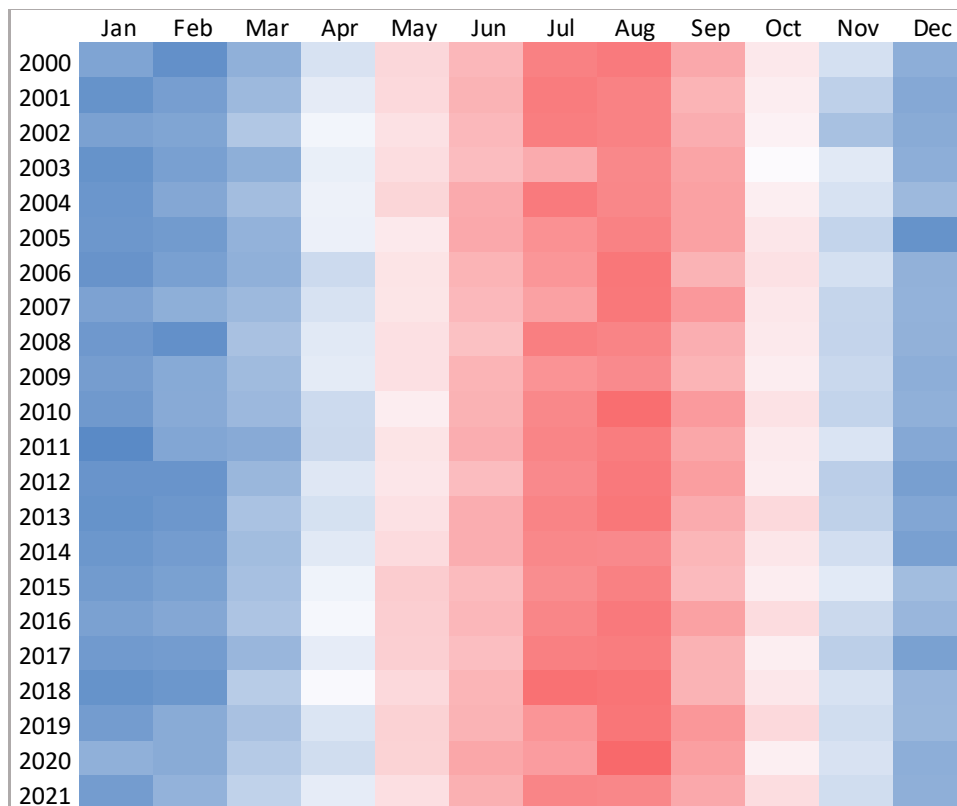
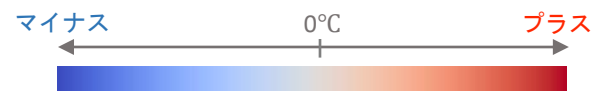
データの関係の可視化 – 3つのデータの関連の可視化 ⑦

3つのデータの関連の可視化は上記であげただけではありません。例えば、以下のようなデータ（カテゴリ値が2つあり、数値が1つ）の場合、ヒートマップなども使用できます。

ある地点の年月別平均「年・月・温度」の三要素可視化

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2000	-4	-6.4	-2.5	3.5	10.2	13.2	18.3	18.9	14.6	8.6	3.3	-2.8
2001	-6.1	-4.7	-1.4	4.7	10	13.6	18.8	18.2	13.5	8.1	1.4	-3.5
2002	-4.3	-3.9	0.3	5.8	9.2	13.1	18.6	18.2	14.1	7.7	-0.5	-3.1
2003	-6.1	-4.5	-2.7	5.1	9.6	12.7	14.3	17.6	15	6.9	4.4	-2.8
2004	-5.7	-3.6	-0.9	5.4	10.3	14.4	18.9	17.7	15.3	8	3.5	-1.4
2005	-5.5	-5.1	-2.3	5.3	8.5	14.6	16.8	18.2	15.3	8.8	1.8	-6.1
2006	-6	-4.5	-2.5	2.6	9	13.5	16.3	19.2	13.6	9.2	3.3	-2.4
2007	-4.2	-2.7	-1.4	3.5	8.9	13.1	15.3	19.1	16.1	8.7	2	-2.3
2008	-5.4	-6.4	-0.4	4.4	9.3	12.3	18.5	18	14	8.5	1.9	-2.4
2009	-4.8	-3.3	-1.2	4.6	9.3	13.5	16.6	17.4	13.5	8.1	2.3	-2.8
2010	-5.3	-3.2	-1.5	2.6	8.1	13.7	17.6	20.1	15.9	9.1	1.8	-2.5
2011	-7.2	-3.7	-3.2	2.5	9	14.1	17.9	18.7	14.7	8.4	3.8	-3.5
2012	-5.9	-5.9	-1.7	4.2	8.8	12.7	17.5	19	15.6	8.2	1.1	-4.6
2013	-6.1	-5.5	-0.3	3.4	9.2	14.1	18	19.2	14.3	10	1.5	-3.7
2014	-5.6	-4.9	-1	4.4	9.8	14.1	17.6	17.5	13.3	8.8	3.1	-4.5
2015	-5.1	-4.4	-0.6	5.6	11.2	12.8	17.2	18.3	12.9	8.1	4.5	-1
2016	-4.3	-3.6	-0.1	6.1	11	13.2	17.8	19	15.3	9.7	2.5	-1.8
2017	-5.2	-4.9	-1.8	4.8	10.9	12.5	18.4	18.7	13.7	8	1.2	-4.4
2018	-6.1	-5.6	0.9	6.4	10	13.4	19.8	19.5	13.6	8.7	3.5	-1.8
2019	-4.9	-3.1	-0.4	3.9	10.7	13.6	16.4	19.3	16.2	10	2.9	-1.7
2020	-2.5	-3.1	0.6	2.9	10.6	14.7	15.7	20.5	15.5	7.9	3.6	-2.8
2021	-4.9	-2.3	1.6	4.8	9.4	13.9	17.9	17.7	14.6	9.6	2.9	-2.6

2方向パレット



3つのデータの可視化でカテゴリ値が2個、数値（連続値）が1個の場合はヒートマップも有効。

データの関係の可視化 – 3つのデータの関連の可視化 ⑦



3つ以上のデータを散布図により可視化するときは、軸に用いるデータにより大きく印象が異なる。どのデータを軸に使用するか、要検討。



散布図に用いる色のグラデーションにより、3つめのデータを表現できる。



色の使用時には、色のセット（パレット）が持つ意味を考える。



「カテゴリーパレット」、「1方向パレット」、「2方向パレット」など、パレットには特徴があり、それぞれで適した使用タイミングがある。



3つのデータの可視化でカテゴリ値が2個、数値（連続値）が1個の場合はヒートマップも有効。

練習問題 8

データの分布の可視化 – 1つの分布の可視化 ⑧

ある町に居住する住民の年齢別人数が以下のものであったとします。年齢ごとの住民のばらつきをわかりやすく可視化するグラフをデザインしてみてください。

各年齢ごとの 人数データ	年齢（歳）	人数		
	1	4		
	2	4		
	3	9		
	4	0		
	5	7	61	7
	6	9	62	2
	7	9	63	5
	8	8	64	5
	9	4	65	5
	10	6	66	8
			67	4
			68	3
			69	1
			70	3
			71	3

データの分布の可視化 – 1つの分布の可視化 ⑧

分布の可視化は、まず任意の間隔ごとに集計したのちに可視化することが有効です。
このような集計結果をグラフにしたものをヒストグラムといいます。

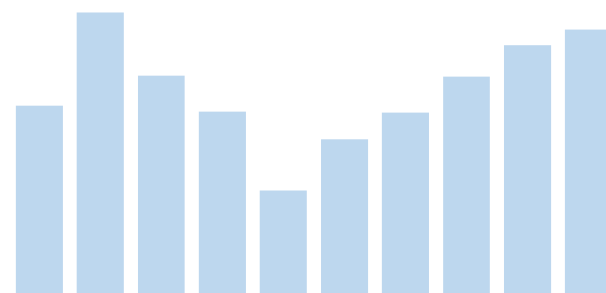
年齢（歳）	人数
1	4
2	4
3	9
4	0
5	7
6	9
...	
64	5
65	5
66	8
67	4
68	3
69	1
70	3
71	3

任意の間隔ごと
に人数を合計

年齢幅	人数
0 - 5	24
6 - 10	36
11 - 15	28
16 - 20	24
21 - 25	14
26 - 30	20
31 - 35	23
36 - 40	28
41 - 45	32
46 - 50	34
51 - 55	38
56 - 60	28
61 - 65	24
66 - 70	19
71 - 75	3

加工されたデー
タに対してグラ
フデザイン

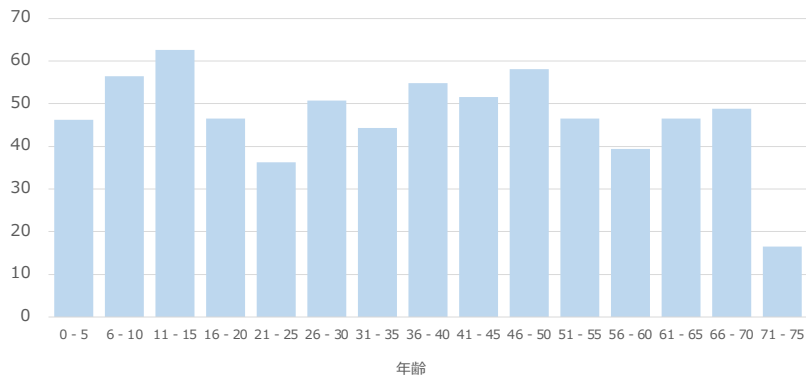
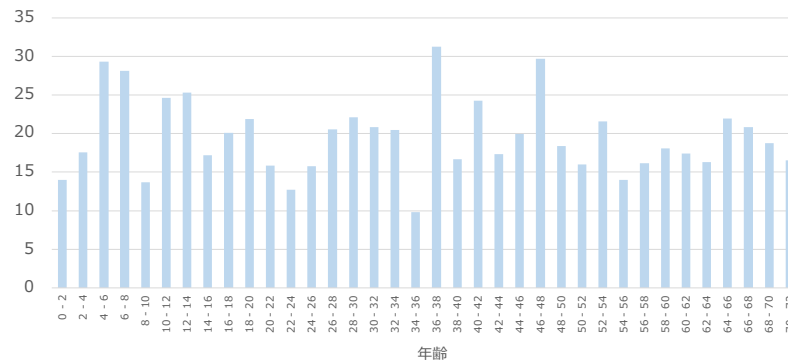
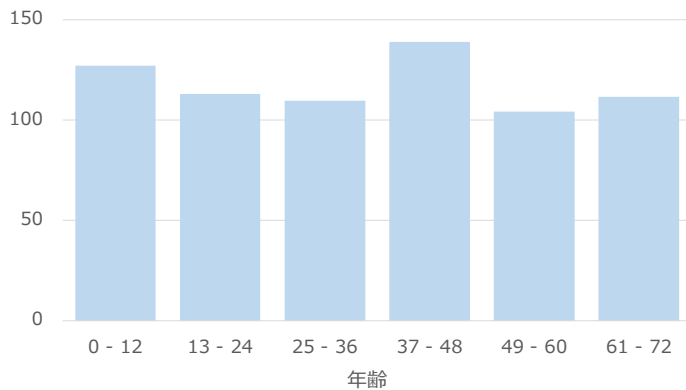
年齢別人数のグラフ



分布がみたいデータセットはデータ量が多いことが多い。その場合は、任意の間隔ごとに集計した後に可視化してみる。

データの分布の可視化 – 1つの分布の可視化 ⑧

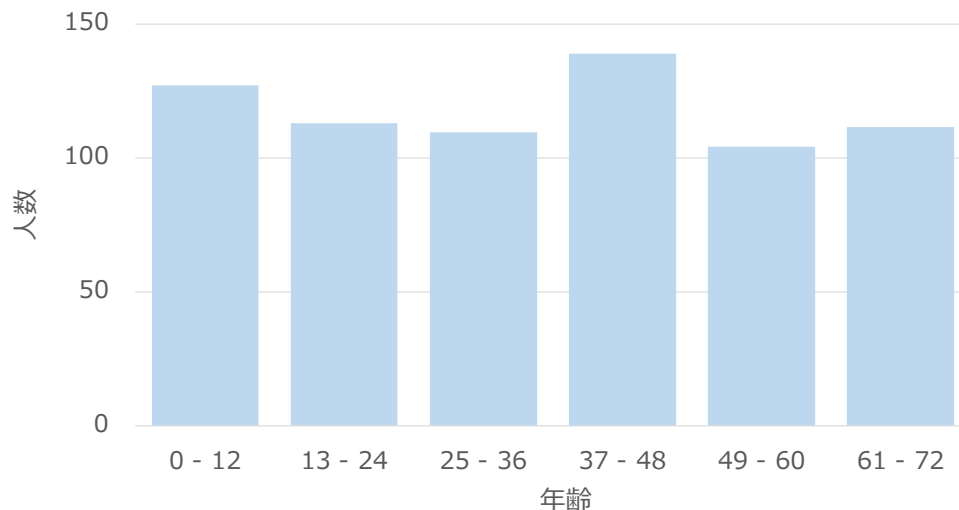
回答例



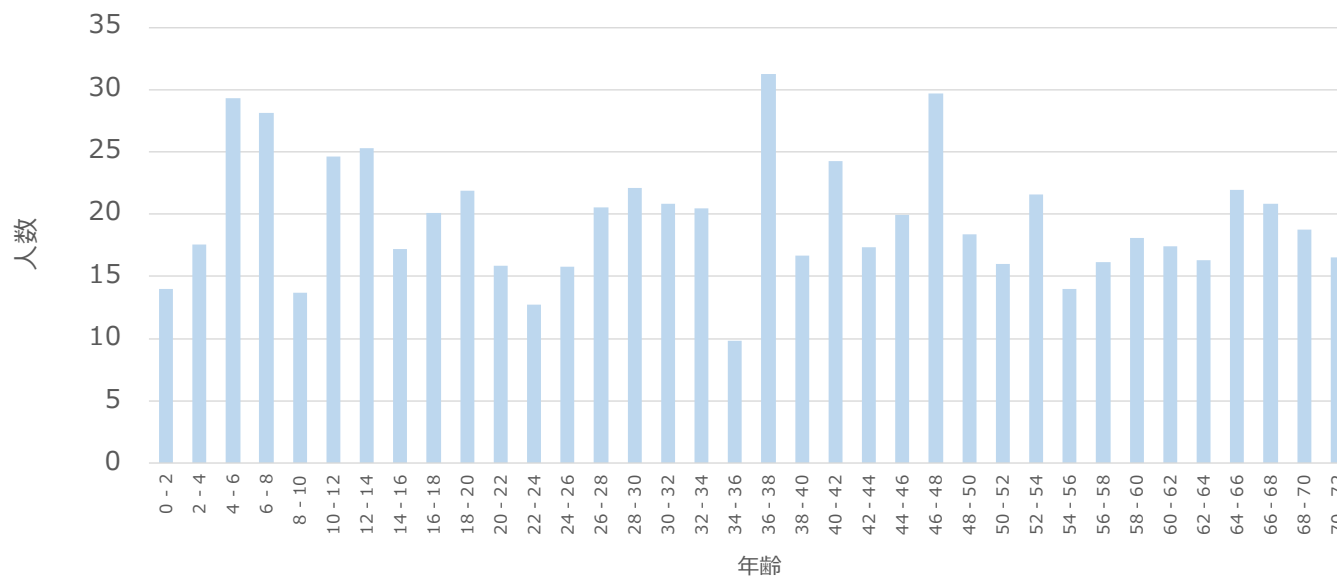
データの分布の可視化 – 1つの分布の可視化 ⑧

ヒストグラムによる可視化

12歳間隔で集計



2歳間隔で集計



データの分布の可視化 – 1つの分布の可視化 ⑧

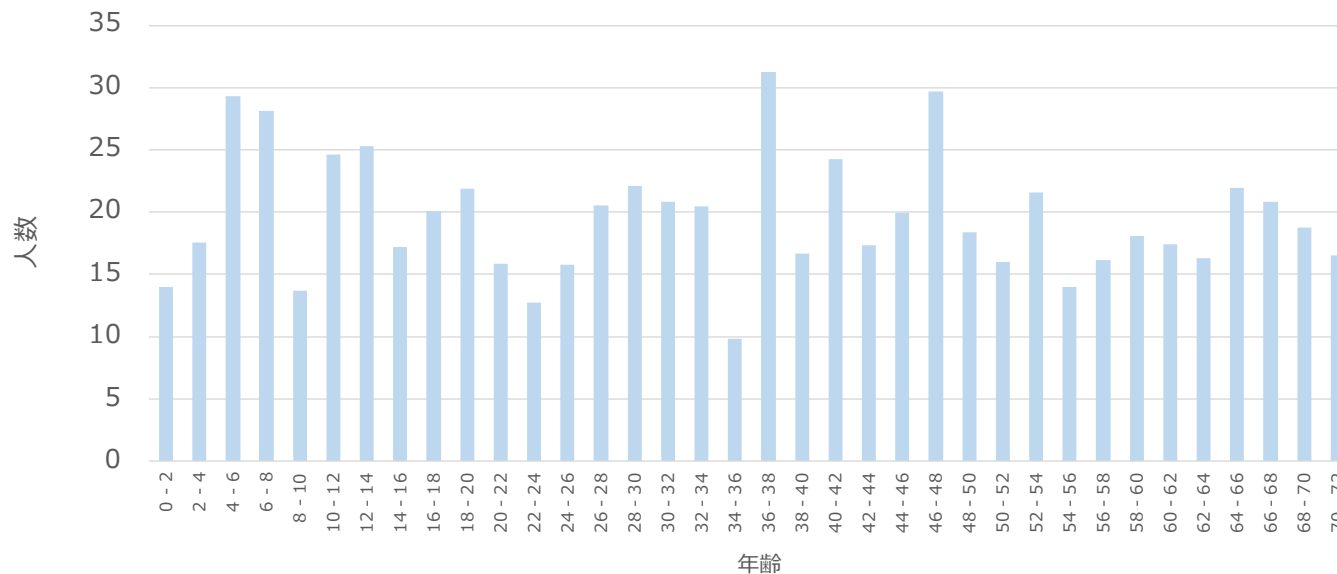
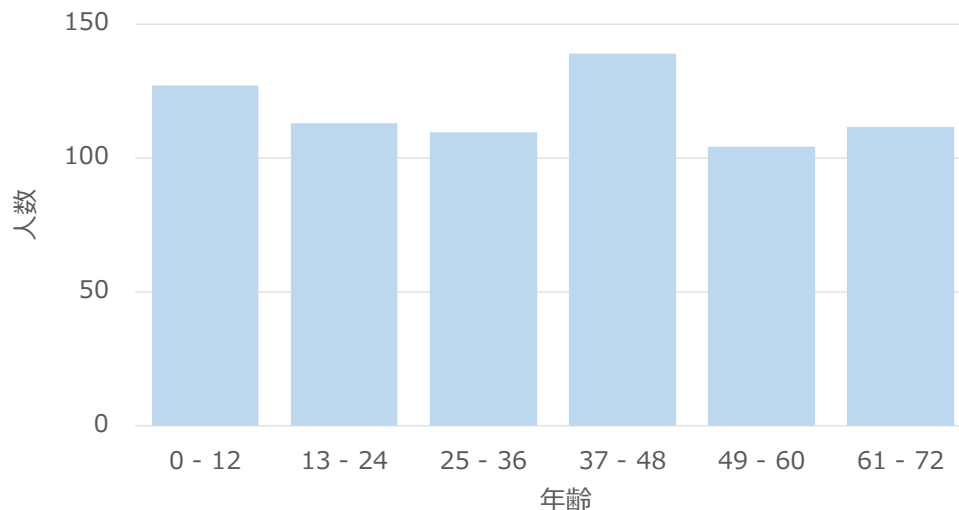
ヒストグラムによる可視化



ヒストグラムの間隔が太すぎ、年齢ごとのばらつきが埋もれてしまっている。

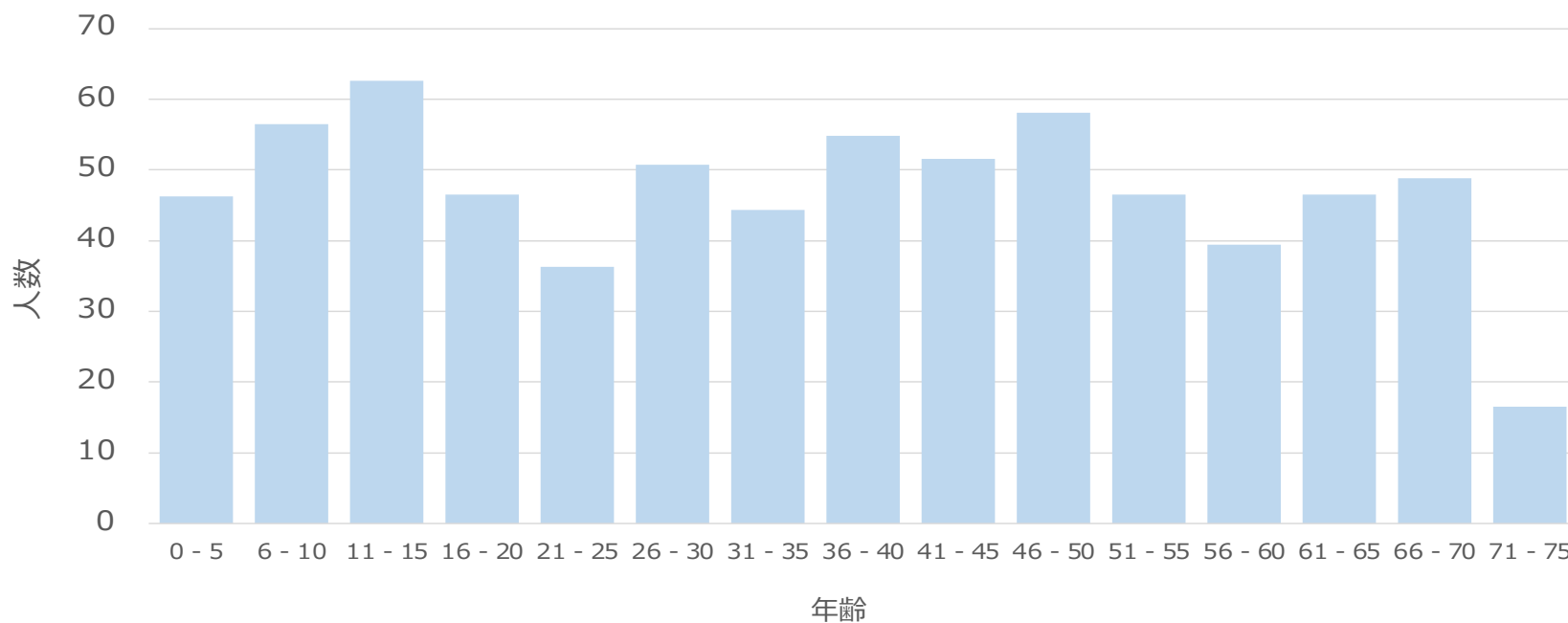


ヒストグラムの幅が狭すぎ、全体の傾向が見えづらい。



データの分布の可視化 – 1つの分布の可視化 ⑧

ヒストグラムによる可視化



データの分布の可視化 – 1つの分布の可視化 ⑧

ヒストグラムによる可視化



ヒストグラムの幅が、ばらつきが隠れておらず、アップダウンが激しくばらつきの傾向がわかりにくいということもない。



ヒストグラムを用いるときは、間隔の選び方に依存して見え方が大きく変わること

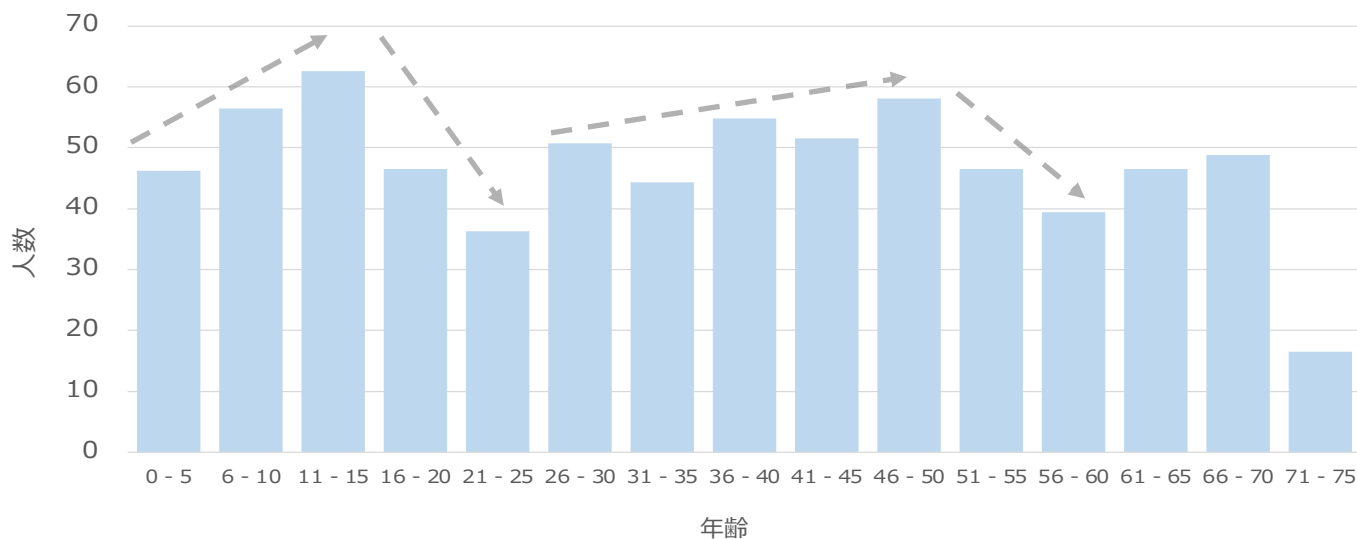


ヒストグラムの間隔は、①②を参考に。

① 意味のある特定の間隔はあるか？

e.g.) 日単位の程度を月で集計

② 棒が5-20程度になるくらいの幅で、データ量が多いほど、幅は細かく。



データの分布の可視化 – 1つの分布の可視化 ⑧



分布がみたいデータセットはデータ量が多いことが多い。その場合は、任意の間隔ごとに集計した後に可視化してみる。



ヒストグラムを用いるときは、間隔の選び方に依存して見え方が大きく変わることに注意。



ヒストグラムの間隔は、①②を参考に。

① 意味のある特定の間隔はあるか？

② 棒が5-20程度になるくらいの幅で、データ量が多いほど、幅は細かく。

練習問題 9

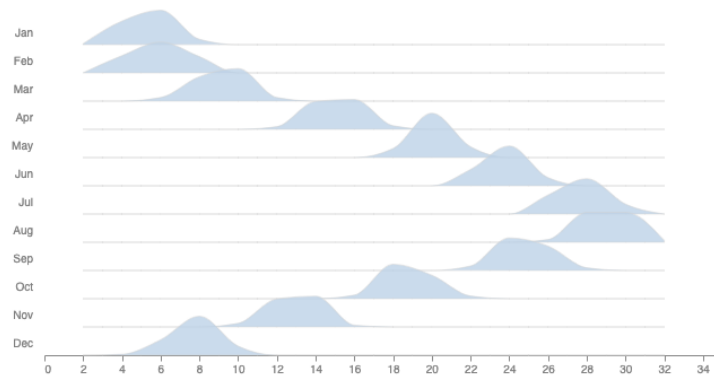
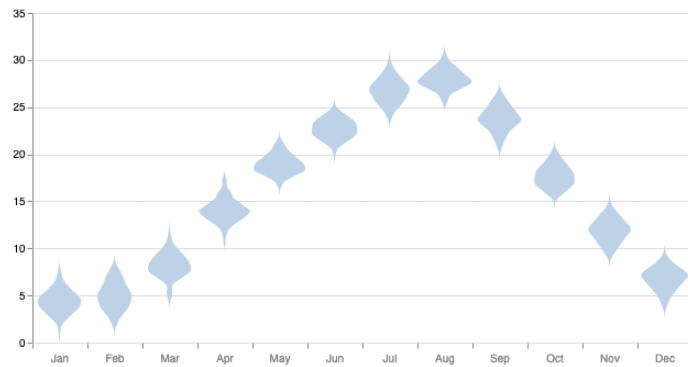
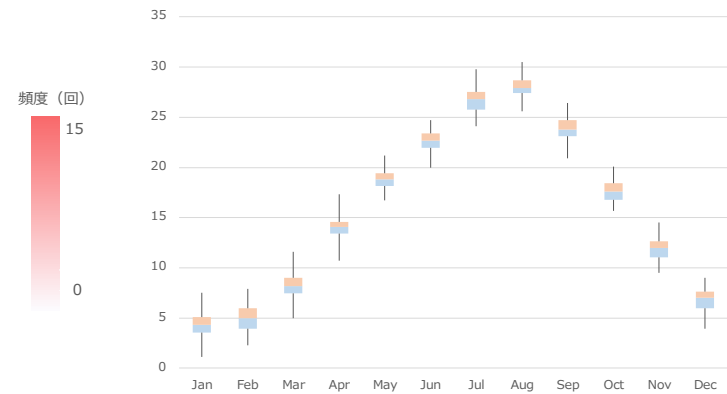
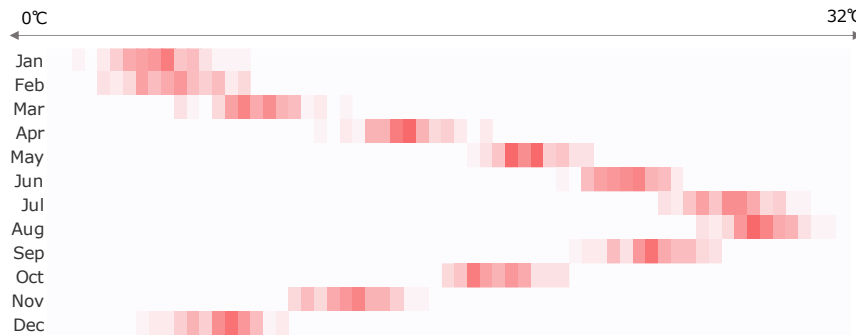
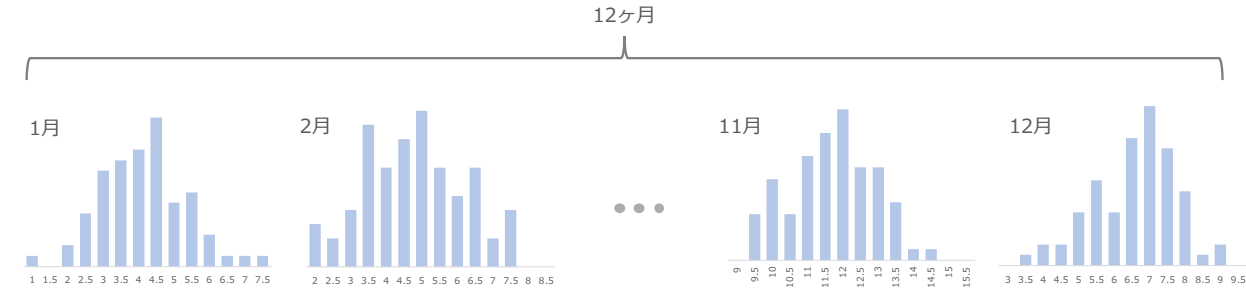
データの分布の可視化- 2つ以上の分布の可視化 ⑨

以下は、直近70年の京都市の月別平均気温です。各月ごとに平均気温がどのよう
にばらついているかがわかるようなグラフをデザインしてみてください。

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1951	3.5	4.8	7.4	12.5	18	21	25.2	28	21	18.1	10.8	7.2
1952	4.1	3.3	6.8	13.3	17.7	21.6	25.2	27.5	23.1	16.5	12.6	5.9
1953	2.9	4	8.6	11.6	17.8	22.1	25.7	27.2	23	17.4	10.2	7.1
1954	5.1	5.3	7.6	14.5	18	20	24.1	27.7	24.1	15.7	12.1	6.6
1955	3.2	5.6	9	14.1	18.2	23.3	27.7	27	23.3	17.2	9.6	6.8
1956	3.6	3.4	8.3	12.8	17.3	22.4	26.2	25.9	23.4	17.3	10.3	4.3
1957	4.3	3.6	5.2	14.1	16.7	21.2	25.2	27.3	20.9	16.5	12.1	7
1958	4	5.3	7.5	13.9	17.9	22.4	26.6	26.8	23.4	16.3	10.7	7.1
1959	3.4	7.5	9	14	18.4	21.8	26.7	27.3	24.3	17.5	12	6.7
1960	3.9	5.4	8.9	12.3	18	21.8	26.8	27.5	23.4	16.8	11.9	5.8
1961	2.6	3.3	8	14.1	19	22	27.4	28	25.7	19.3	12.3	6.6
1962	3.2	4.8	7.3	13.1	18.1	21	26.4	27.6	23.9	16.8	11	6.6
1963	1.1	2.3	7	14.2	19	23.5	27.1	26.7	21.5	16.2	11.2	6.7
1964	5.5	3.2	6.6	17.2	19.3	21.6	27.2	28.7	23.9	16.6	10.3	6.2
1965	3.3	3.8	5.5	10.7	17.9	22	25.8	27.7	21.7	15.9	11.8	5.8
1966	3.4	5.7	8.5	13.5	17.4	21.4	25.8	27.9	23.2	17.2	10.7	4.7
1967	3.2	4.1	7.7	13.8	19.9	23.1	26.6	28.4	23.7	16.4	12	4.6
1968	3.4	2.3	8	14.3	18.1	22.4	25.6	26.9	22.5	15.7	11.6	9
1969	4.8	5	6.9	13.8	19	21.4	26.1	28.1	24.7	16.7	11	5.7
1970	3.6	5.1	5	13.1	19.2	21	26.5	27.7	24.7	17.3	11.2	6.2
1971	4.3	5.1	7.4	13.6	18.5	22.5	26.7	27.3	22.9	16.3	11.8	7.4
1972	7.1	5.8	8.8	14.4	18.7	22.5	27.1	27.6	23.1	18.1	12	7.7
1973	6.2	6.9	7.8	15.8	18.4	21.9	28.4	28.5	22.3	16.8	10.1	4.3
1974	3.2	4.2	7	14.4	19.3	22.7	25.1	27.6	22.3	17	10.4	6.6
1975	4	3.7	7.4	14.1	18.5	22.5	26.9	27.2	25	17.6	12.2	6.3
...												
2014	4.4	5.1	9	14.4	19.8	24.1	27.6	27.5	23.3	18.8	13.1	5.5
2015	4.9	5.6	9.4	15.6	21.2	22.8	27.2	28.3	22.9	18.1	14.5	9
2016	5.7	6.4	9.9	16.1	21	23.2	27.8	29	25.3	19.7	12.5	8.2
2017	4.8	5.1	8.2	14.8	20.9	22.5	28.4	28.7	23.7	18	11.2	5.6
2018	3.9	4.4	10.9	16.4	20	23.4	29.8	29.5	23.6	18.7	13.5	8.2
2019	5.1	6.9	9.6	13.9	20.7	23.6	26.4	29.3	26.2	20	12.9	8.3
2020	7.5	6.9	10.6	12.9	20.6	24.7	25.7	30.5	25.5	17.9	13.6	7.2
2021	5.1	7.7	11.6	14.8	19.4	23.9	27.9	27.7	24.6	19.6	12.9	7.4

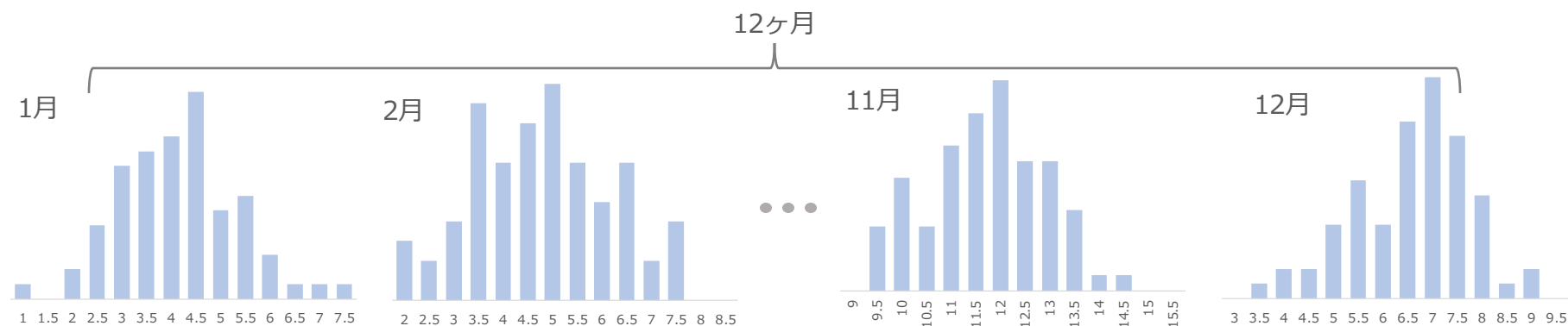
データの分布の可視化- 2つ以上の分布の可視化 ⑨

回答例



データの分布の可視化ー 2つ以上の分布の可視化 ⑨

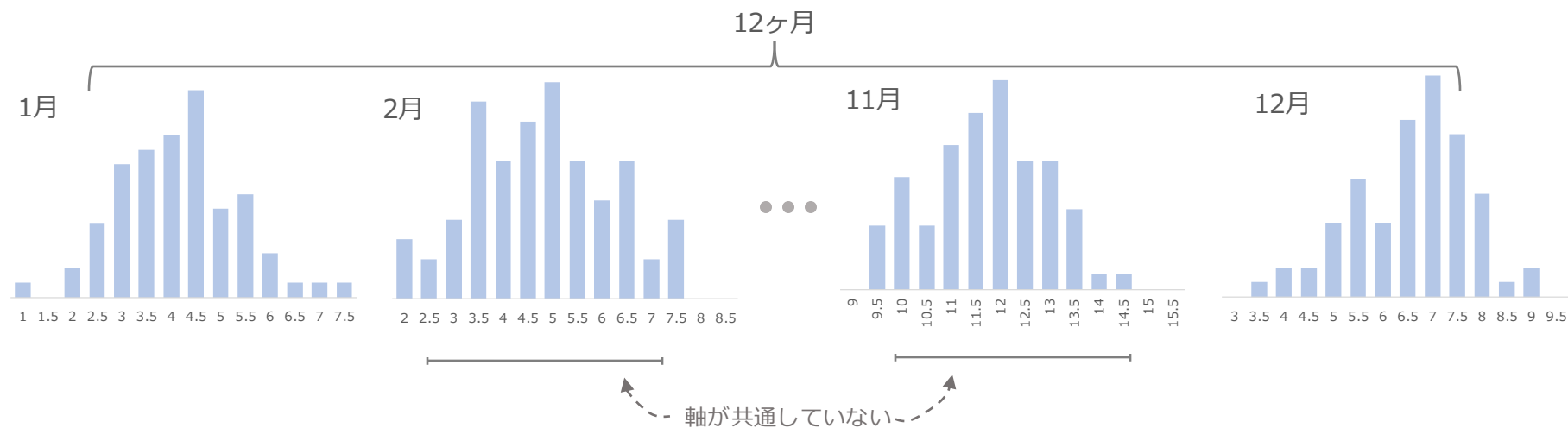
多重のヒストグラムによる可視化



月ごとにヒストグラムを作成
0.5℃の幅を使用

データの分布の可視化ー 2つ以上の分布の可視化 ⑨

多重のヒストグラムによる可視化



⚠ グラフの個数が多く、必要なスペースの確保が難しい。

⚠ グラフの個数が多く、グラフ全体を俯瞰的にみることが難しい。

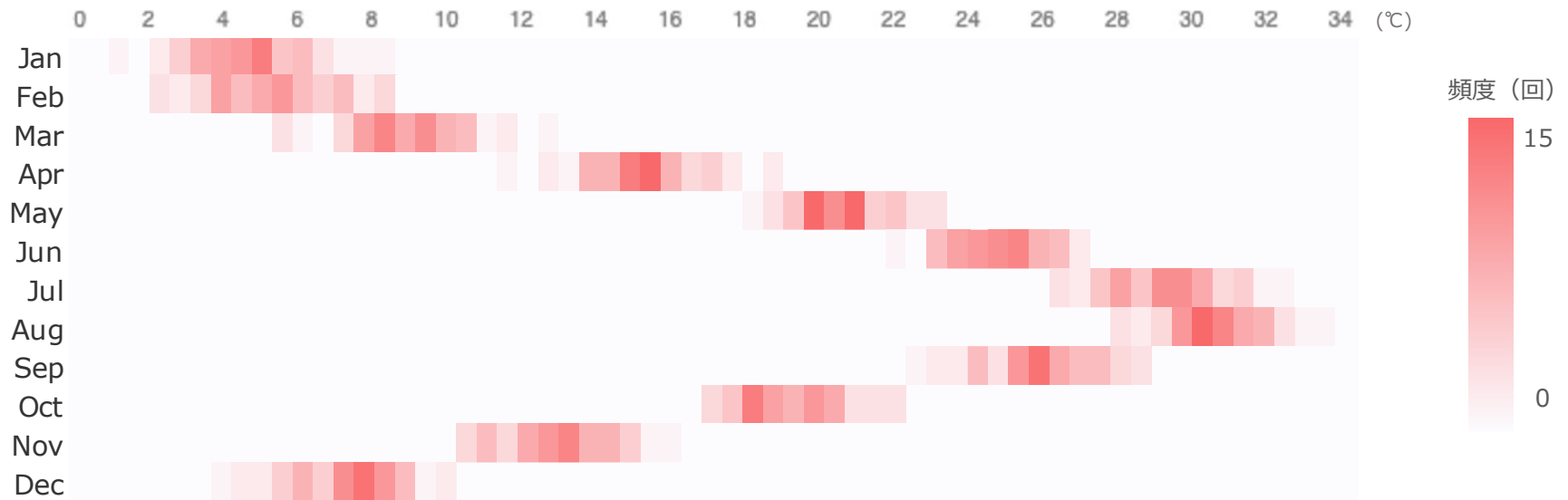
⚠ 複数のグラフ間で共通の軸をもとに比較することが難しい。



いくつかの要素を複数のヒストグラムを用いて可視化したいときは、分布の数を~4程度まで抑える。

データの分布の可視化- 2つ以上の分布の可視化 ⑨

気温幅で集計 + ヒートマップを用いた可視化



月ごとに0.5°Cの幅を使用して集計
回数を色で表す。

データの分布の可視化- 2つ以上の分布の可視化 ⑨

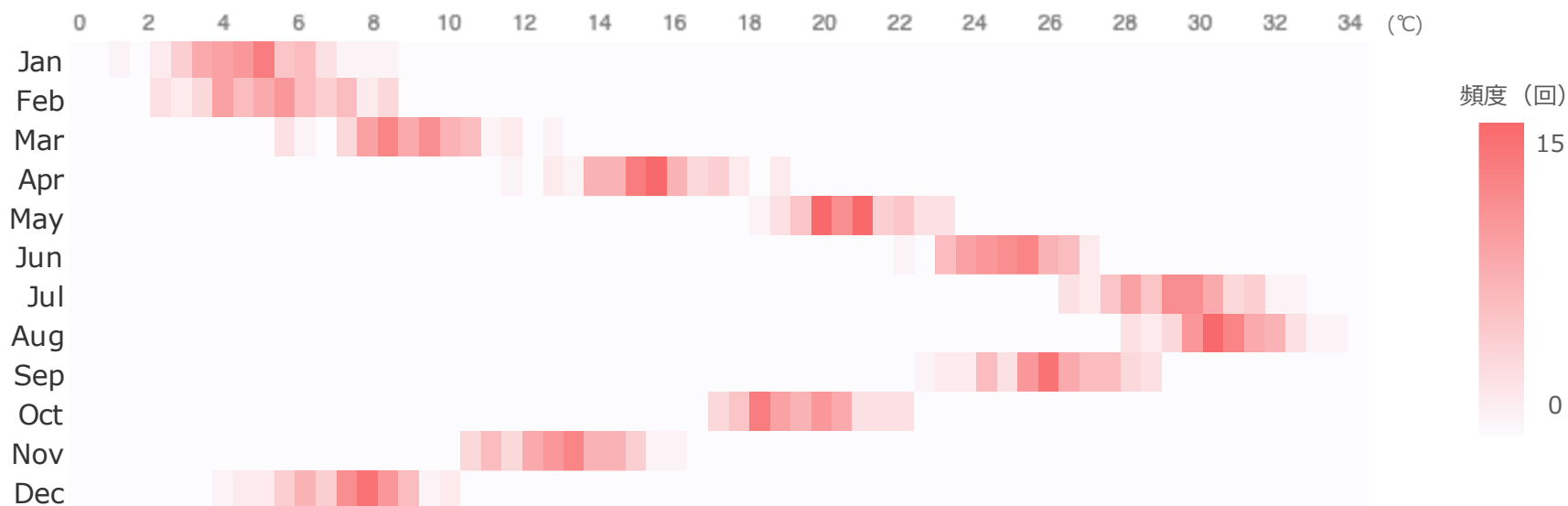
気温幅で集計 + ヒートマップを用いた可視化



必要なスペースが少なく、全体像を俯瞰的に見ることができる。



共通軸をもとに各ばらつきを比較することができる。



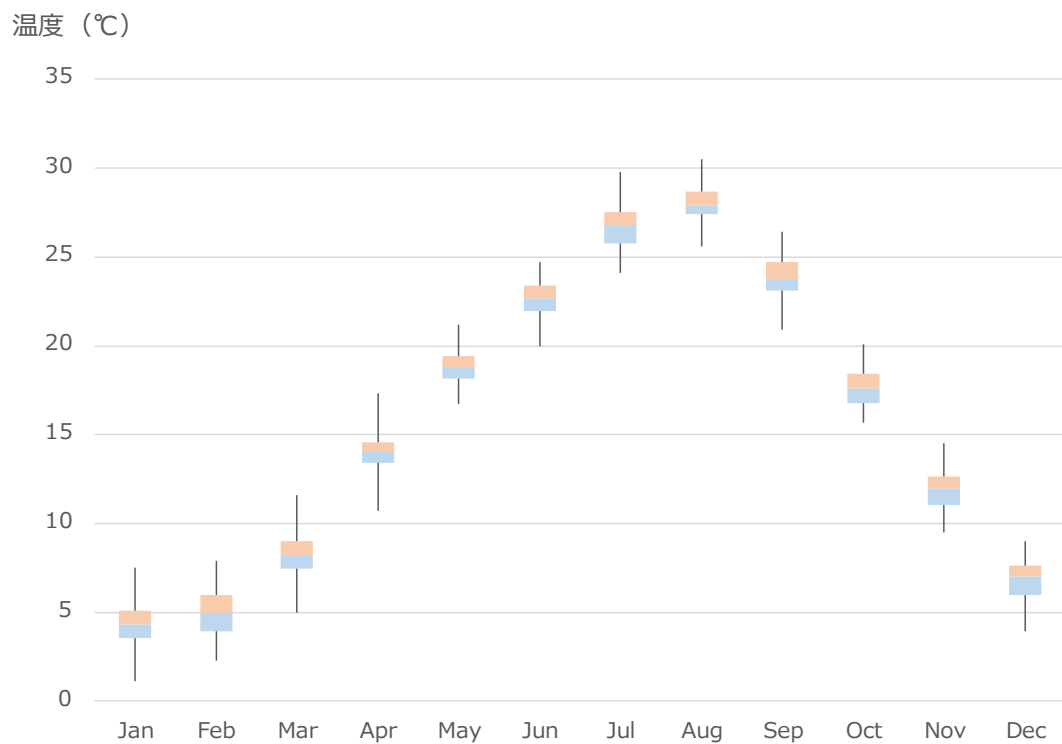
気温の幅のとり方に依存してしまう。



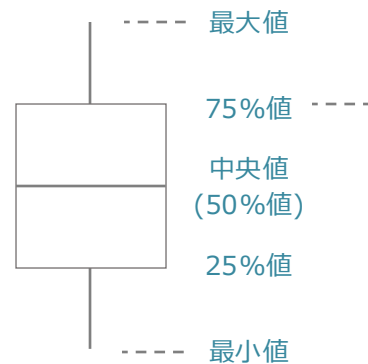
複数の分布を同時に可視化するには、一定の幅で集計 + ヒートマップを用いた可視化が有効。
省スペースで、共通軸を意識した比較が可能になる。

データの分布の可視化- 2つ以上の分布の可視化 ⑨

箱ひげ図を用いた可視化。



箱ひげ図 (box plot)



100個データがあるとき、小さい順に並べて75番目の数字

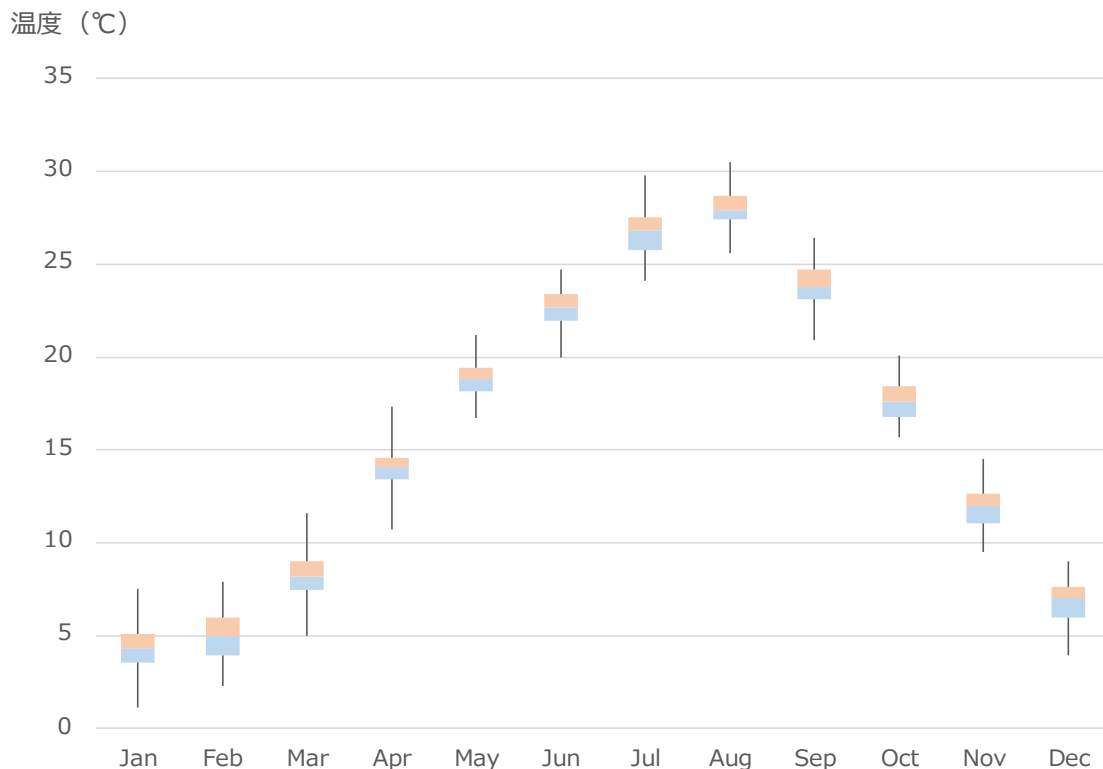
データの分布の可視化- 2つ以上の分布の可視化 ⑨

箱ひげ図を用いた可視化。

- ✓ 必要なスペースが少なく、全体像を俯瞰的に見ることができる。
- ✓ 共通軸をもとに各ばらつきを比較することができる。
- ✓ 中央値や25%値、75%値、最大値、最小値など関心のある値が明確である。
- ⚠ 詳細な分布がわからない。

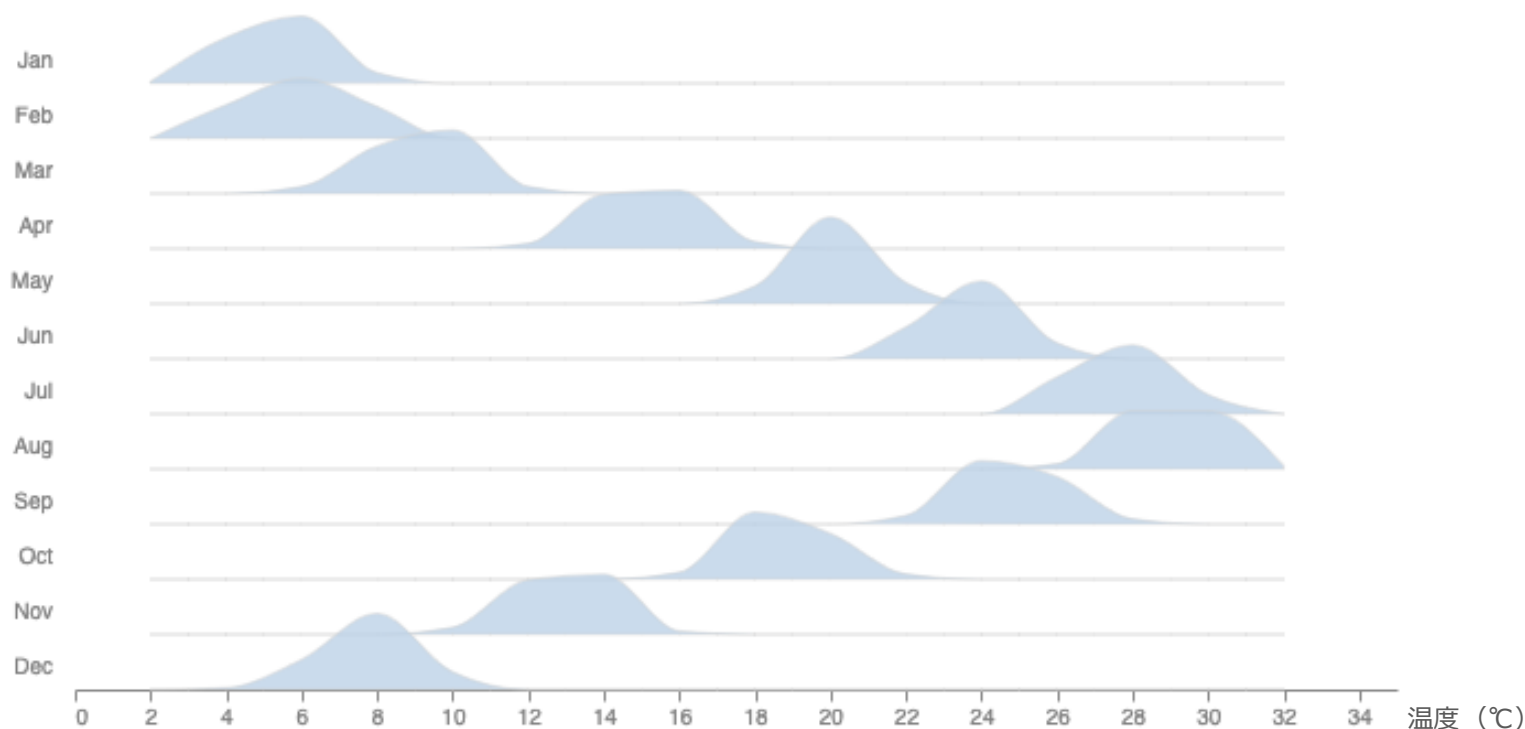


複数の分布の可視化には、箱ひげ図も有効。特に平均や四分位数に興味があるときに。



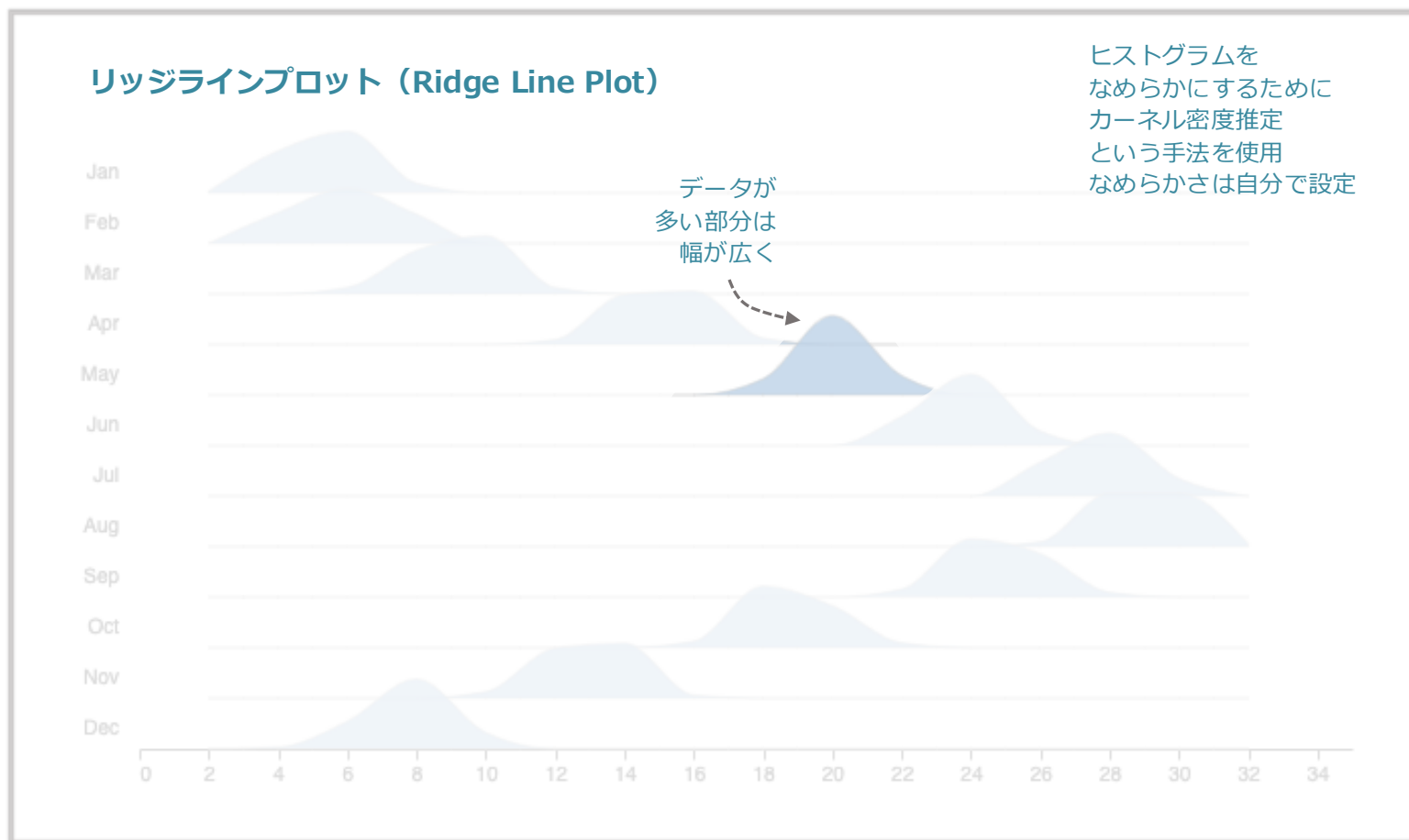
データの分布の可視化ー 2つ以上の分布の可視化 ⑨

発展) リッジラインプロットを用いた可視化。



データの分布の可視化ー 2つ以上の分布の可視化 ⑨

発展) リッジラインプロットを用いた可視化。



データの分布の可視化- 2つ以上の分布の可視化 ⑨

発展) リッジラインプロットを用いた可視化。

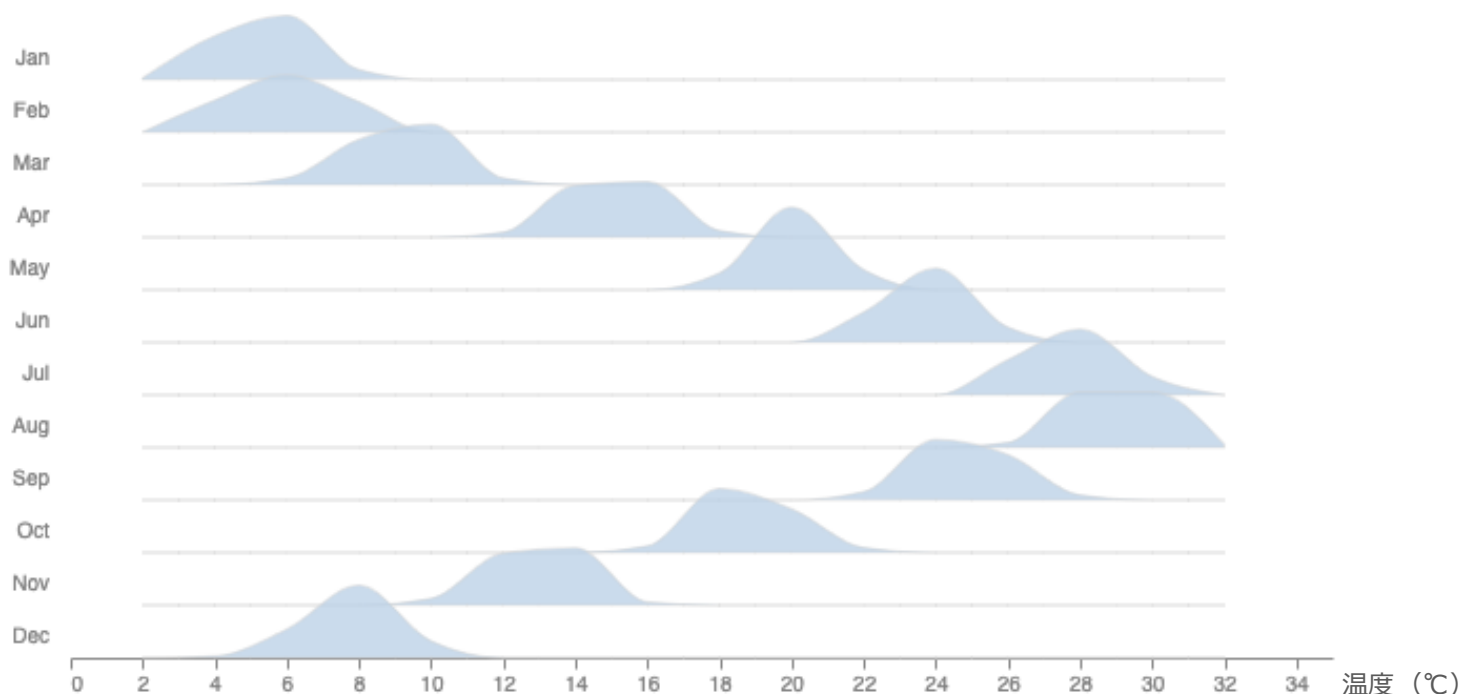
✓ 必要なスペースが少なく、共通軸を意識して、分布間の比較ができる。

✓ 詳細な分布を把握することができる。

⚠ カーネル密度推定の計算が必要。BIソフトのカスタムチャートの使用や、Python/JavaScriptなどをコーディングが必要。

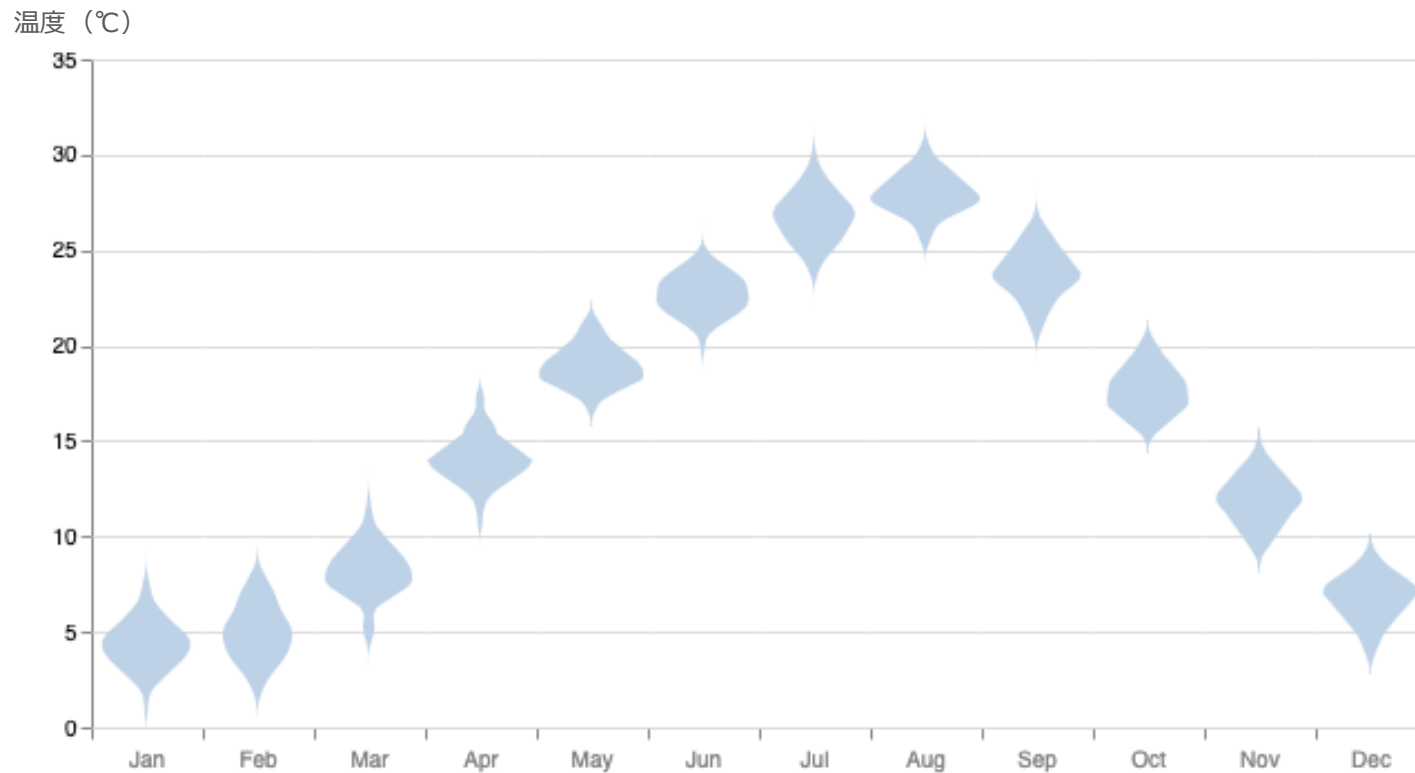


リッジラインプロットは、詳細な分布の可視化が可能。カーネル密度推定の設定に注意。



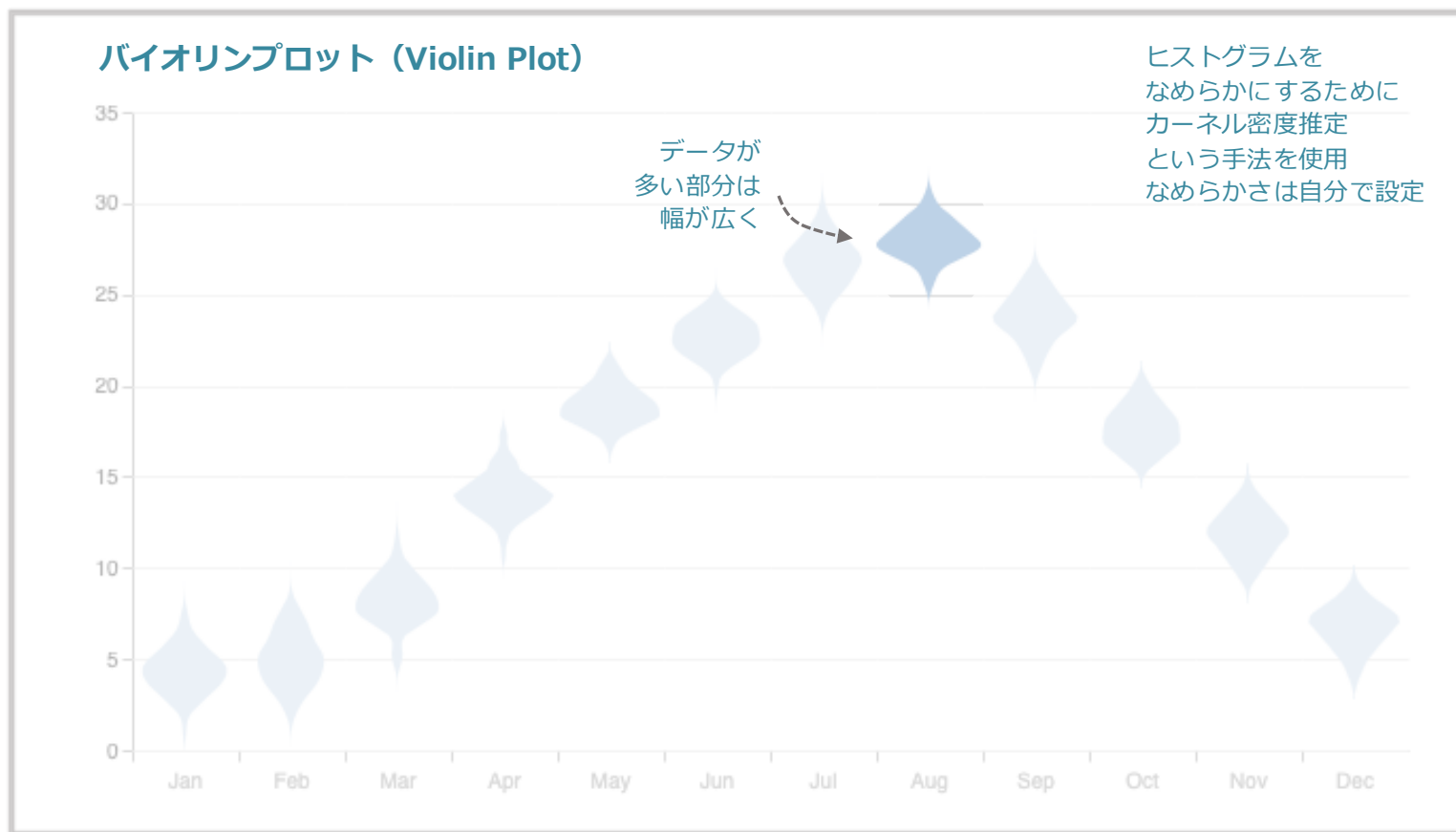
データの分布の可視化– 2つ以上の分布の可視化 ⑨

発展) バイオリンプロットを用いた可視化。



データの分布の可視化ー 2つ以上の分布の可視化 ⑨

発展) バイオリンプロットを用いた可視化。



データの分布の可視化– 2つ以上の分布の可視化 ⑨

発展) バイオリンプロットを用いた可視化。

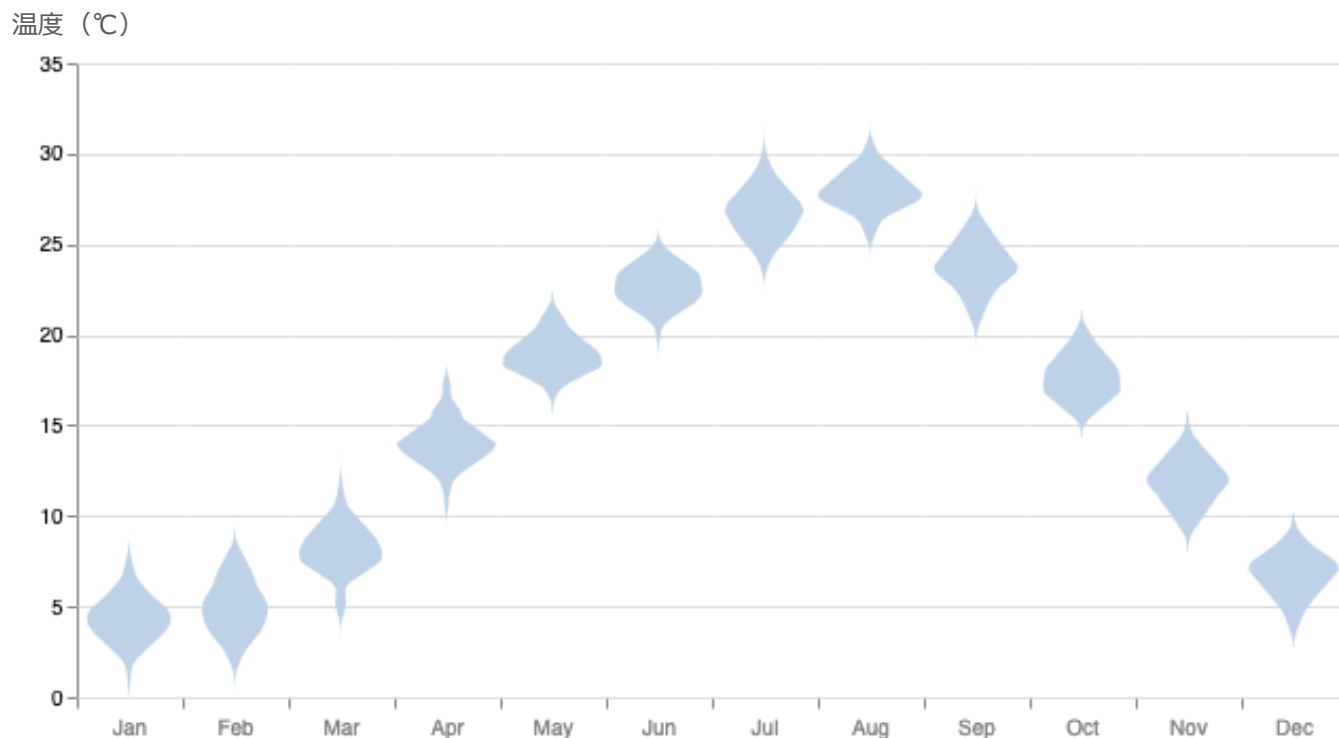
✓ 必要なスペースが少なく、共通軸を意識して、分布間の比較ができる。

✓ 詳細な分布を把握することができる。

⚠ カーネル密度推定の計算が必要。BIソフトのカスタムチャートの使用や、Python/JavaScriptなどをコーディングが必要。



バイオリンプロットは、詳細な分布の可視化が可能。カーネル密度推定の設定に注意。



データの分布の可視化ー 2つ以上の分布の可視化 ⑨



いくつかの要素を複数のヒストグラムを用いて可視化したいときは、分布の数を~4程度まで抑える。



複数の分布を同時に可視化するには、一定の幅で集計+ヒートマップを用いた可視化が有効。省スペースで、共通軸を意識した比較が可能になる。



複数の分布の可視化には、箱ひげ図も有効。特に平均や四分位数に興味があるときに。



バイオリンプロットは、詳細な分布の可視化が可能。カーネル密度推定の設定に注意。



リッジラインプロットは、詳細な分布の可視化が可能。カーネル密度推定の設定に注意。

授業のまとめ

データを可視化するときに気をつけることやデザインパターンを学びました。

授業のゴール

- ① データ可視化の役割を理解する。
- ② データを可視化する際に考慮するポイントについて理解を深める。
- ③ 色々なグラフの種類を知り、可視化のレパートリーを増やす。
- ④ データを可視化する際に用いるデザインパターンについて理解を深める。



授業内容

理論

第1回

データ可視化とは？
データ可視化の目的
データ可視化の構成要素と文脈

実践

第2-4回

練習問題を通じたグラフづくり
量データの可視化
割合データの可視化
データの関係の可視化
データの分布の可視化