

データリテラシー

第6回 データのグラフによる表現とデータ比較

メディア情報コース
平居 悠（ひらい ゆたか）

到達目標

- データサイエンス・AIが社会でどのように活用されているのかを理解する。
- データを適切に読み解き理解する力をつける。
- データを適切な可視化手法で他者に説明できるようになる。
- 表計算ソフトを使ってデータを集計・加工できるようになる。

前回の目標

代表値と散らばりの尺度から数量的にデータの分布の特徴を読み取れるようになる。

前回学んだこと

1. 変数
2. 代表値
3. 散らばりの尺度

変数の種類と尺度

変数の種類	変数の尺度
量的変数	比率尺度
	間隔尺度
質的変数	順序尺度
	名義尺度

質的変数

定量的な値では観測できないデータを扱う変数

例

- 性別：男, 女
- 血液型：A型, B型, O型, AB型
- GP：S, A, B, C, D

量的変数

長さや温度など定量的な
値（数値）で観測される
データを扱う変数

比率尺度（比例尺度、比尺度）

原点が決まっており、間隔と比率に意味がある。
比率尺度の変数同士では四則演算（足し算、引き算、掛け算、割り算）ができる。

例

- 長さ : 175 cm, 2236 m
- 重さ : 10 g, 1.5 kg
- 価格 : 750 円, \$150

間隔尺度

数値の間隔に意味がある。間隔尺度の変数同士では足し算と引き算ができる。

例

- 暦年：西暦2021年, 令和5年（和暦）
- 温度：36.5 °C

順序尺度

順序に意味がある。順序尺度の変数同士の演算には意味がない。

例

- 順位：1位, 2位, 3位
- 科目の評価：優, 良, 可, 不可

名義尺度

単に分類するためだけに使われる。便宜上、数値化はできるが、順序や演算には意味がない。

例

- 血液型：A型, B型, AB型, O型
- コース：経営, 政策, 地域経営, 観光まちづくり, メディア情報

代表値

分布の中心を表す値

例：平均値、中央値、最頻値

平均値（算術平均）

それぞれのデータの総和
をデータの個数で割った
もの

中央値（メディアン）

データを小さい順（もしくは
大きい順）に並べたときの
ちょうど中央の値

最頻値

観測値の中で最も多く現
れる観測値

パーセンタイル（分位数）

データを小さい順に並べたときの小さい方から $100p\%$ のところの値を $100p$ パーセンタイルという ($0 \leq p \leq 1$)。データを $p:1-p$ に分ける点ともいえる。

中央値 = 50パーセンタイル ($p = 0.5$)

四分位数

パーセンタイルのうち、特に25%, 50%, 75%パーセンタイルを**四分位数**という。これらはデータをちょうど四分割する点にあたるため、**四分位点**ともいう。

- 第1四分位数 Q_1 = 25パーセンタイル
- 第2四分位数 Q_2 = 50パーセンタイル (= 中央値)
- 第3四分位数 Q_3 = 75パーセンタイル

これに最小値 Q_0 と最大値 Q_4 を加えて Q_0, Q_1, Q_2, Q_3, Q_4 の5つの指標を**五数要約**という。

散らばりの尺度

分布の広がりを表す指標

分散

偏差の総和が0になるのは偏差に負の値があるせいなので、偏差の負の値を無くすために、**偏差を二乗**して平均をとったものを**分散 S^2** という。分散が分布の散らばりを表す尺度になる。

標準偏差

分散 S^2 の平方根（ルート）をとったもの

$$S = \sqrt{S^2}$$

データの単位と同じになるため、データの平均と直接比べられる。

前回の問い

- 量的変数と質的変数の違いは何か？
 - 定量的な値で観測されるデータを扱う変数を**量的変数**といい、定量的な値では観測できないデータを扱う変数を**質的変数**という。
- 代表値とは何か？
 - 平均値、中央値、最頻値など、分布の中心を表す値。
- 散らばりの尺度には何があるか？
 - 分散、標準偏差などがある。

タイピング練習スケジュール

第1回	ホームポジション
第2回	ホームポジション
第3回	ローマ字
第4回	英語初級
第5回	日本国憲法（trr試験）
第6回	ホームポジション
第7回	ローマ字
第8回	英語初級
第9回	日本国憲法（trr試験）
第10回	ホームポジション
第11回	ローマ字
第12回	英語初級
第13回	日本国憲法（trr試験）

タイピングの練習 (ホームポジション)

1. ブラウザを起動し、<https://www.koeki-prj.org/trr/>に繋ぐ。
2. 学籍番号（Cは大文字、省略なし8桁）を入力する。
3. Koeki MAILに届いたパスコードをPasscode: 欄に入力する。

ホームポジション

左手でタイプするキー

右手でタイプするキー



左手の人差指から小指までの
ホームポジション

両手の親指の
ホームポジション

右手の人差指から小指までの
ホームポジション

今回

第1回	社会におけるデータ・AI利活用
第2回	表計算ソフトの基本操作
第3回	セルの参照と集計
第4回	度数分布表とヒストグラム
第5回	代表値と散らばりの尺度
第6回	データのグラフによる表現とデータ比較
第7回	データの分布と標本抽出
第8回	推定
第9回	分割表とクロス集計
第10回	2変量データと相関
第11回	時系列データ
第12回	仮説検定
第13回	回帰分析

今回の目標

グラフを用いてデータを可視化できるようにする。

今回学ぶこと

1. グラフの種類
2. グラフの作成
3. データの分析と比較

今回の問い

- グラフにはどのような種類があるか？
- LibreOffice Calcでグラフを作成するには何を用いるのか？
- データの比較を行う際の注意点は何か？

今回学ぶこと

1. グラフの種類
2. グラフの作成
3. データの分析と比較

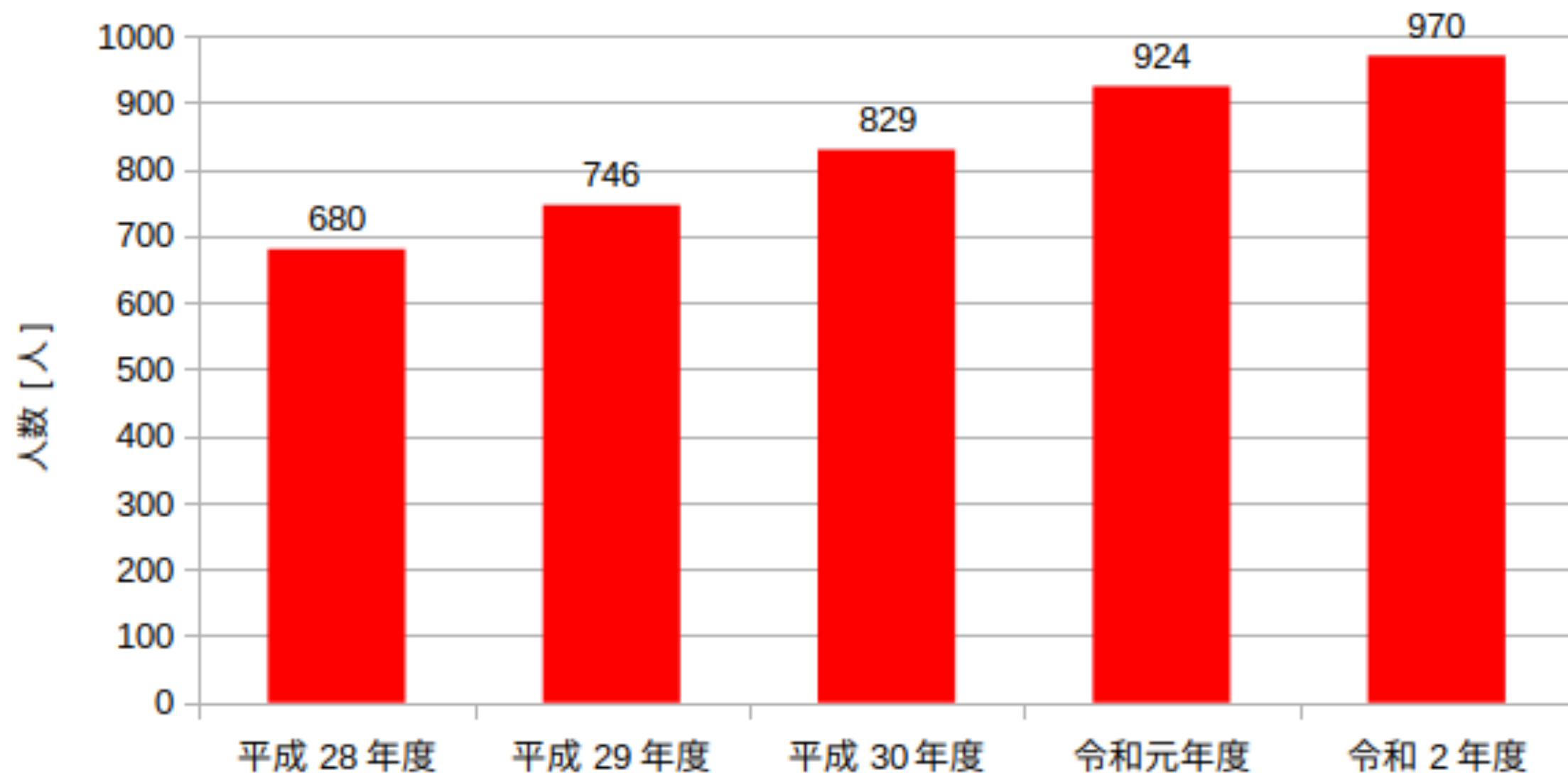
棒グラフ

長方形の棒の長さでデータの大小を表す
グラフ

- 棒は**縦棒**と**横棒**の二通りある。
- 複数の系列の棒を積み上げた**積み上げ棒グラフ**もある。
- 等幅の柱状グラフ（**ヒストグラム**）は棒グラフで描くことができる。

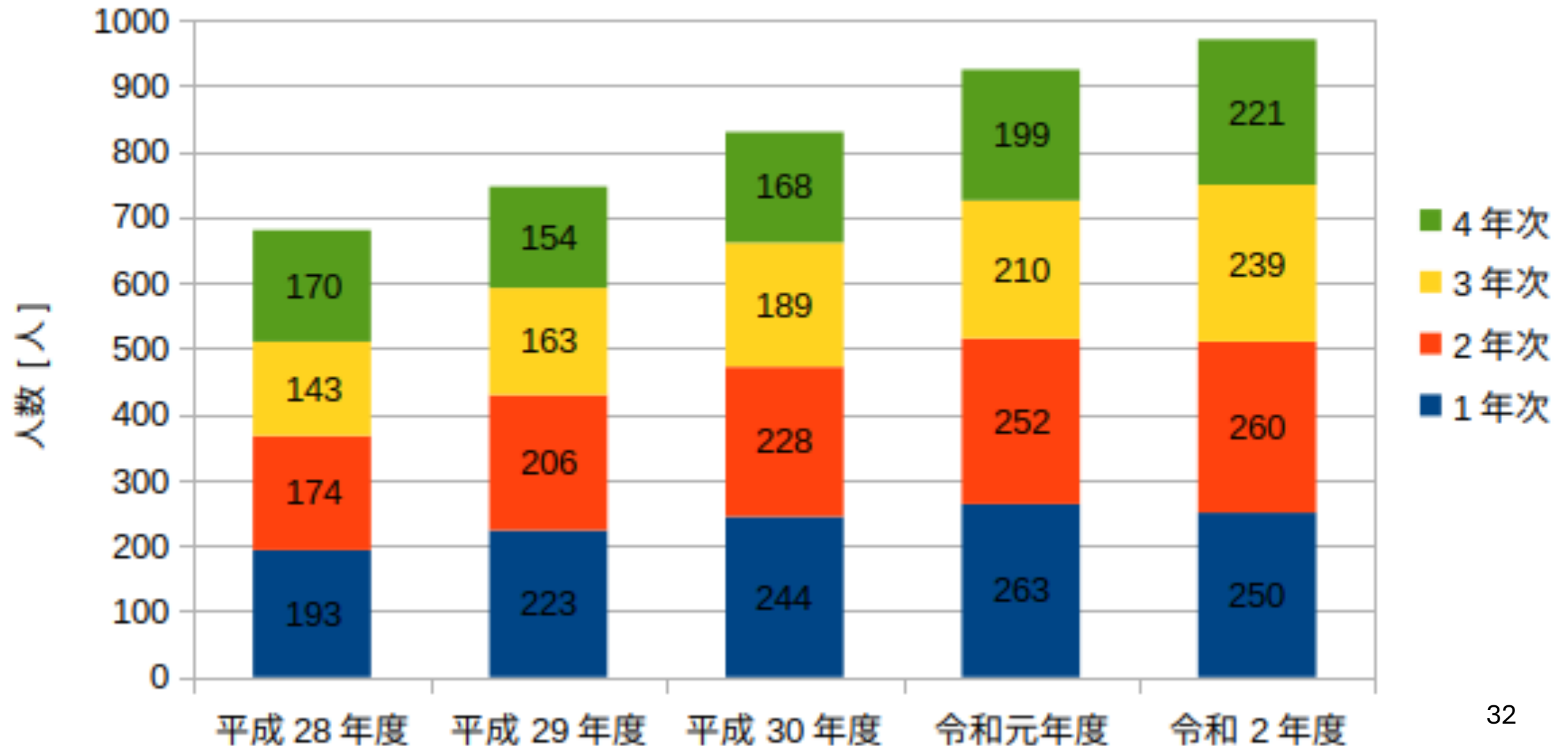
棒グラフの例

東北公益文科大学 在籍者数



棒グラフの例

東北公益文科大学 在籍者数



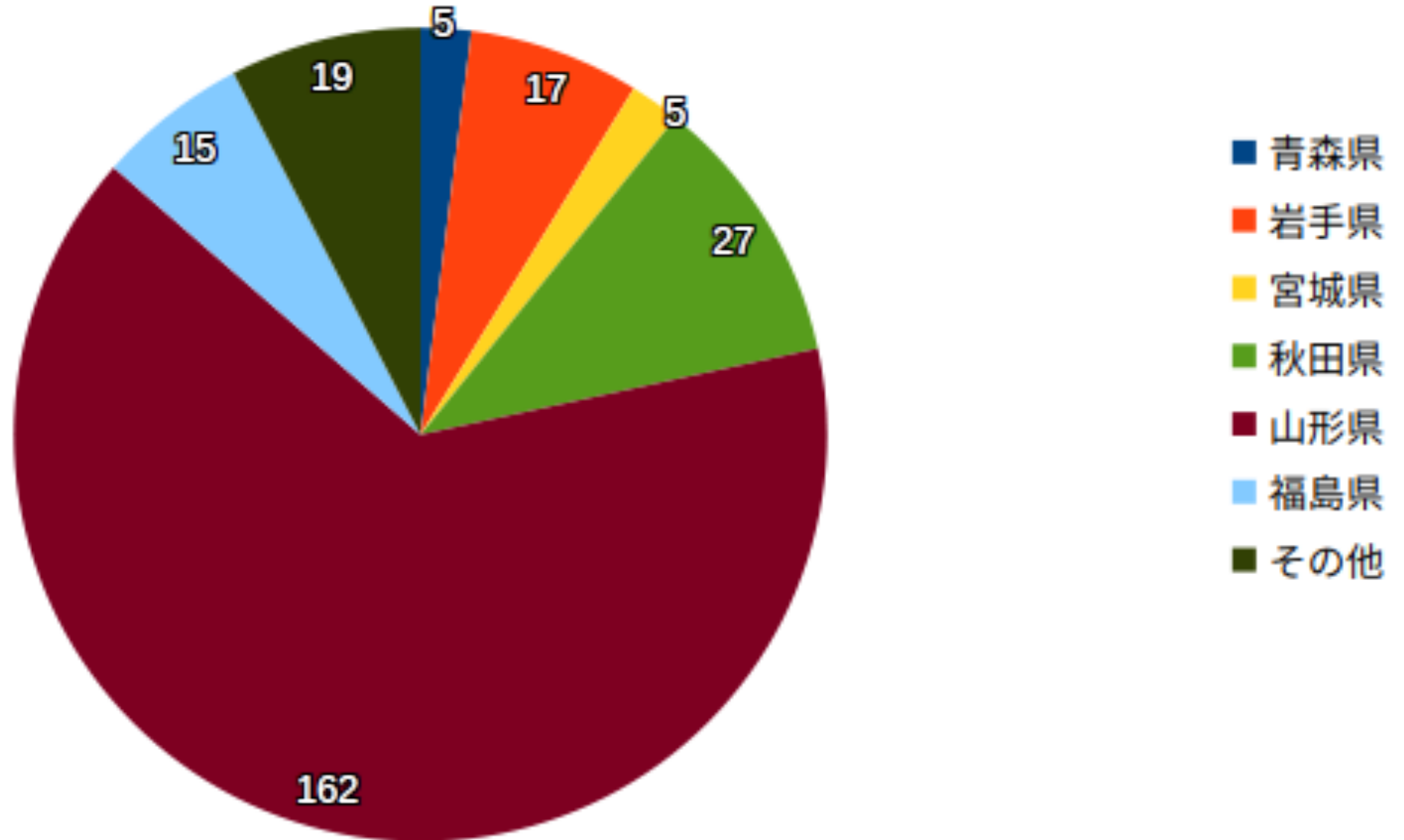
円グラフ

円を扇型に分割し、扇型の大きさ（角度）で割合を表すグラフ

- LibreOffice Calcではデフォルトでデータ系列は反時計回りの順に描かれるが、データ系列のオプションで時計回りの順に変更できる。

円グラフの例

令和 2 年度 東北公益文科大学 出身地別の入学者数



折れ線グラフ

データを点でプロットし、その点を線で結んだグラフ

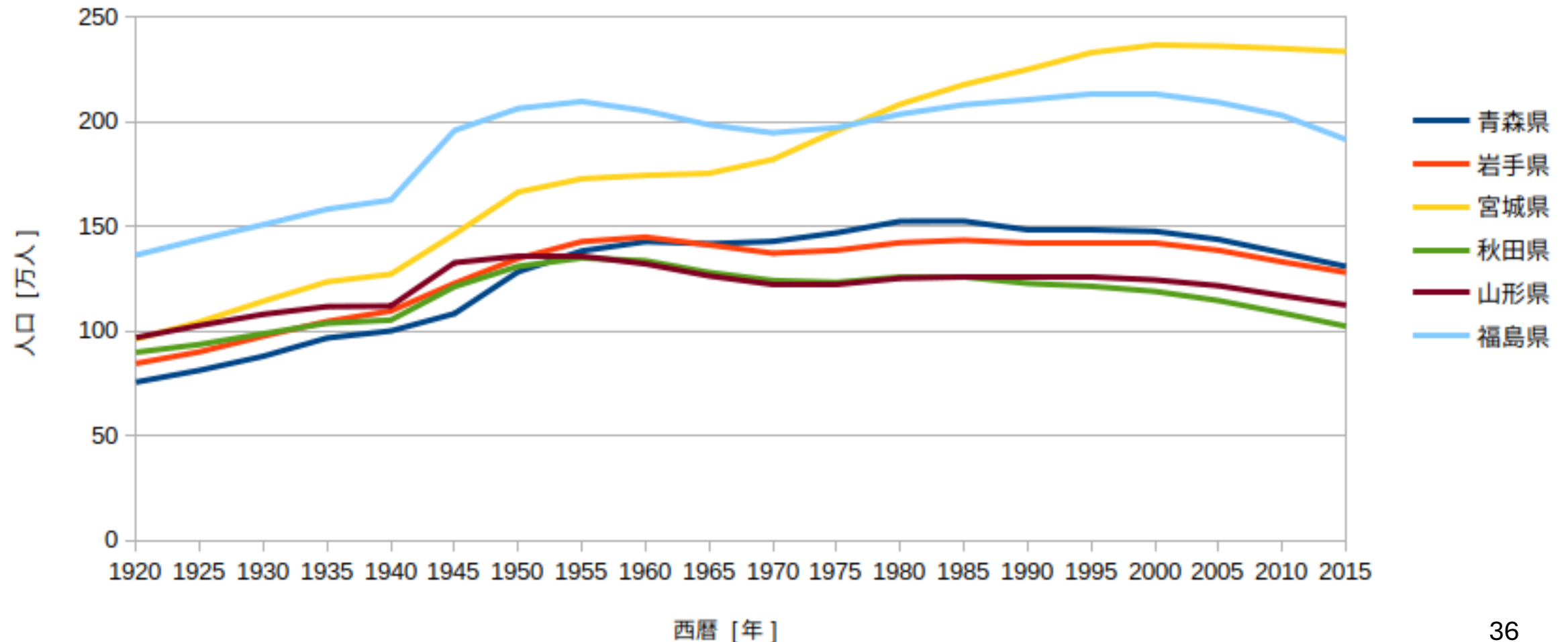
- 時系列データの場合、横軸を時間軸、縦軸を各時間における値とした折れ線グラフで時間変動を見ることが多い。

折れ線グラフの例

東北 6 県の総人口の推移

出典：総務省統計局「国勢調査：男女別人口及び人口性比－全国，都道府県（大正 9 年～平成 27 年）」

<https://www.stat.go.jp/data/kokusei/2015/kekka.html>, (参照 2020-10-01)



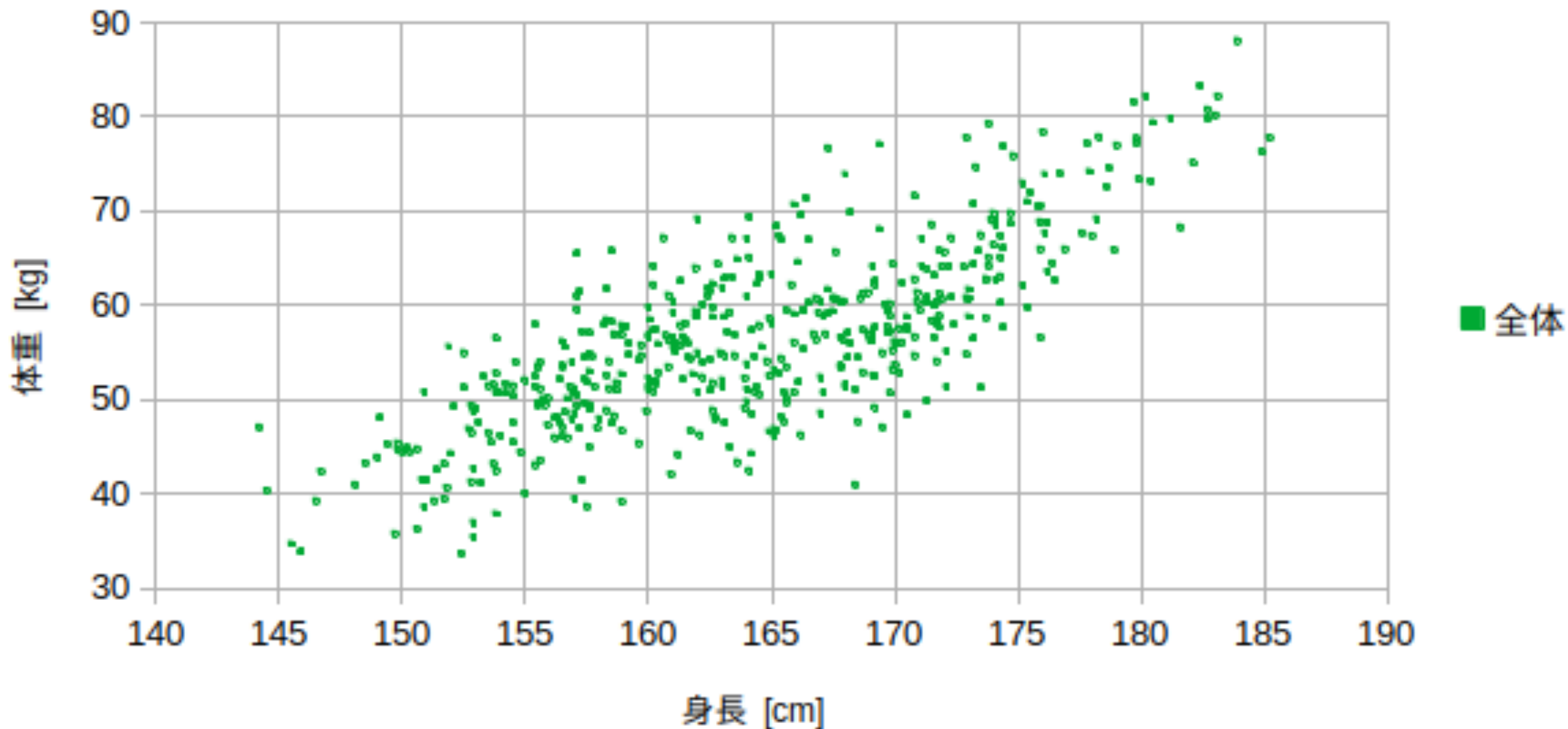
散布図

2変数のデータがあるときに縦軸（Y軸）と横軸（X軸）にそれぞれの変数を対応させ、各データを点などで描いたグラフ

- 同じ系列の点を線で結ぶと折れ線グラフになる。
- 2変数の（相関）関係を見るために使われることが多い。

散布図の例

データセット 1 の身長と体重

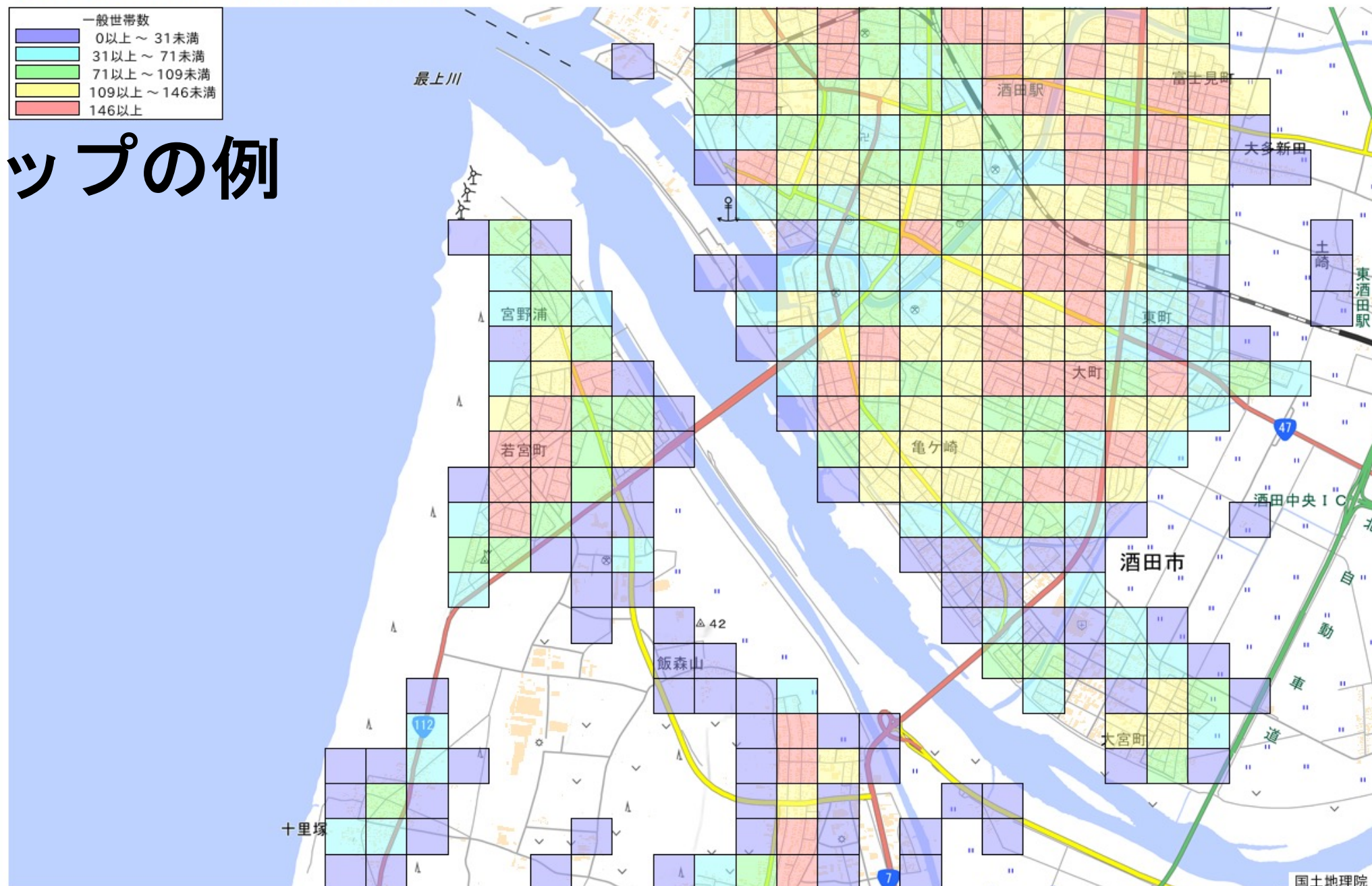


ヒートマップ

散布図でデータの密度を色で表現したグラフ

- 色の濃淡で値の大小を表す。
- GIS (Geographic Information System) のように地図と組み合わせてその地点の値を表すために使うこともある。

ヒートマップの例



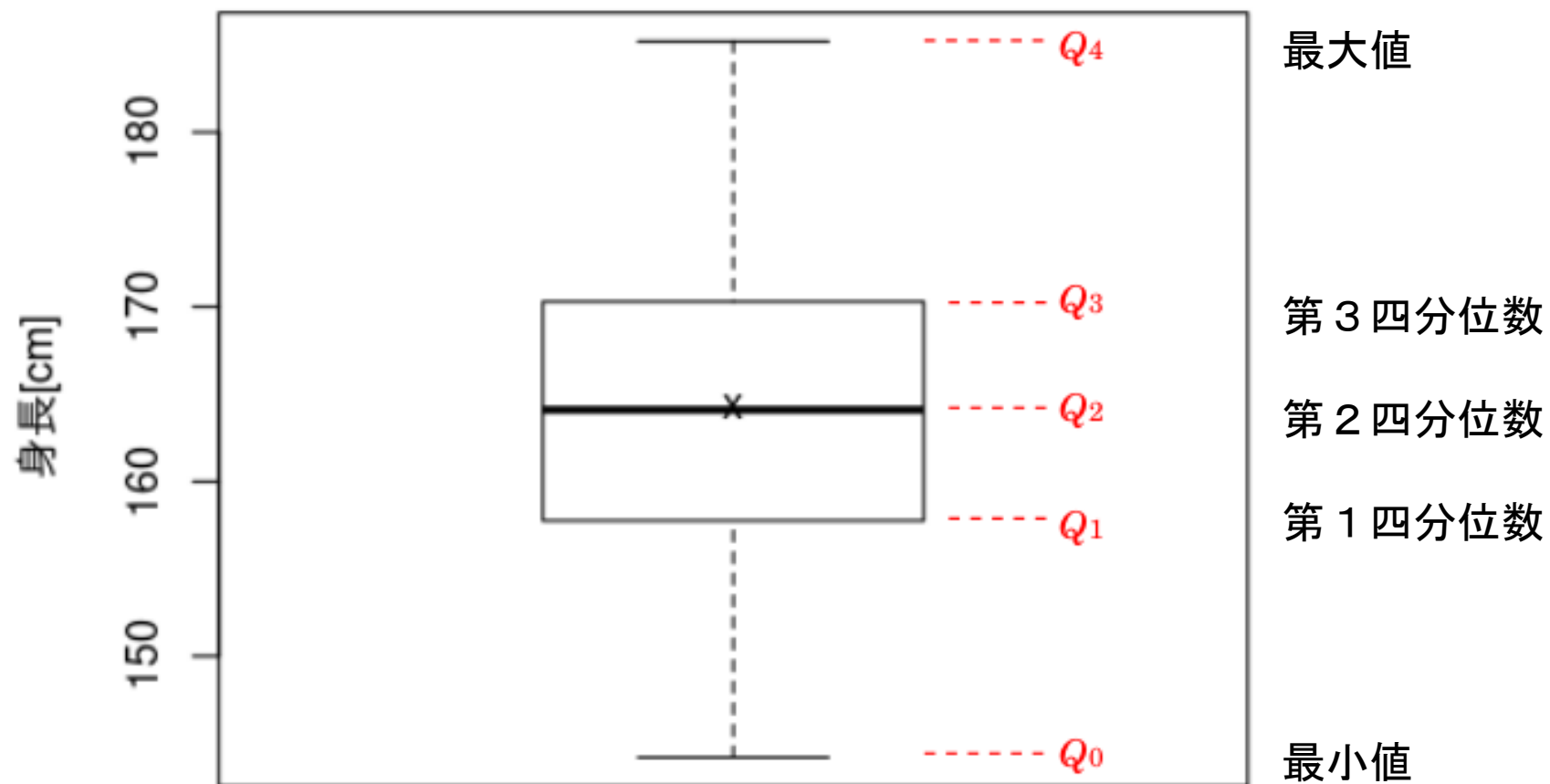
箱ひげ図

データを小さい順に並べたときの分布を四分位数（または五数要約）を使って「箱」と「ひげ」の長さで表したグラフ。

- データがどの付近に集中して分布しているのかが見やすく、特に複数の分布を比較するのに有効なグラフ。
- 峰が複数あるような分布では適切に表すことができないので注意。

箱ひげ図の例

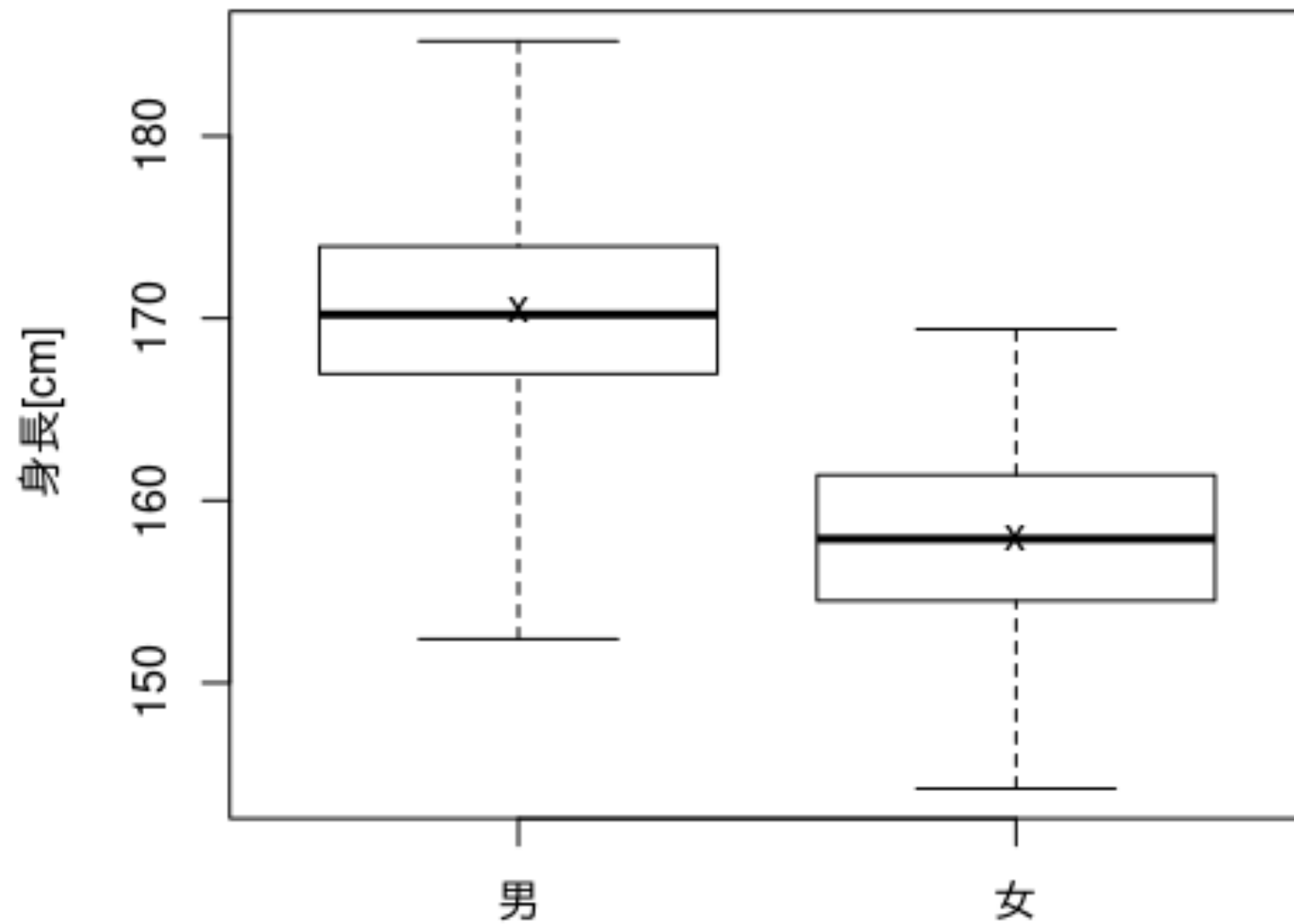
データセット1



全体

箱ひげ図の例

データセット1



今回の問い

- グラフにはどのような種類があるか？
 - 棒グラフ、円グラフ、折れ線グラフ、散布図、ヒートマップ、箱ひげ図などがある。
- LibreOffice Calcでグラフを作成するには何を用いるのか？
- データの比較を行う際の注意点は何か？

今回学ぶこと

1. グラフの種類
- 2. グラフの作成**
3. データの分析と比較

グラフウィザード

1. グラフにしたいデータの範囲を選択する。
2. 以下の操作で「グラフウィザード」のダイアログを表示する。

メニューバー	[挿入]→[グラフ] を選択
ツールバー	[グラフを挿入] ボタンをクリック

グラフウィザード設定手順

- ① グラフの種類
- ② データの範囲
- ③ データ系列
- ④ グラフ要素

各ステップは[次へ]で次のステップ、
[戻る]で前のステップに移動する。

①グラフの種類

グラフの種類を選択する。

グラフの種類を選択	グラフの種類
縦棒	棒グラフ（縦棒）
横棒	棒グラフ（横棒）
扇形	円グラフ
面	面グラフ
線	折れ線グラフ（横軸が質的変数）
散布図	散布図、折れ線グラフ

3Dルックはグラフからデータを読み取りにくくなるので使わない方がよい。

②データ範囲

グラフにするデータ範囲を選択する。グラフウィザードを表示する前にデータを選択していれば、ある程度自動的に設定してくれる。

設定項目	説明
データ範囲	グラフに表示するデータの範囲を指定する。 入力フォームの右の 【データ範囲の選択】 ボタンをクリックするとマウスで範囲が選択できる。
行内のデータ系列 列内のデータ系列	データ範囲において、一つの系列のデータが行方向もしくは列方向に並んでいるかいずれかを選択する。
最初の行を項目名に引用	データ範囲の最初の行をラベルとして使用するかを選択する。
最初の列を項目名に引用	データ範囲の最初の列をラベルとして使用するかを選択する。

③データ 系列

より細かい
データ系列の
設定を行える。

設定項目	説明	
データ系列	設定するデータ系列を選択する。 [追加] [削除] で系列を追加したり削除したりする。 [↑] [↓] で系列の順序を入れ替える。	
データ範囲	設定したいパラメータのデータ範囲を選択する。 選択したパラメータのデータ範囲はその下の入力フォームに直接入力するか、その右のボタンをクリックしてマウスで範囲選択する。	
	枠線の色	枠線の色のセルを選択する。
	塗りつぶしの色	塗りつぶしの色のセルを選択する。
	名前	系列名の書かれたセルを選択する。文字列を直接入力しても良いが少し制限がある。
	X値	X軸のデータ範囲を選択する。
	Y値	Y軸のデータ範囲を選択する。
範囲	ラベルの書かれたセル範囲を選択する。 棒グラフだとX軸のラベルであったり、円グラフだと系列名であったりとグラフの種類によって役割が異なる。 入力フォームに直接入力するか、その右のボタンをクリックしてマウスで範囲選択する。	
データラベル	散布図でデータラベルの書かれたセル範囲を選択する。 入力フォームに直接入力するか、その右のボタンをクリックしてマウスで範囲選択する。	

④ グラフ要素

設定項目	説明
タイトル	グラフのタイトルを入力する。
サブタイトル	グラフのサブタイトルを入力する。
X軸	X軸のラベルを入力する。
Y軸	Y軸のラベルを入力する。
目盛線を表示	どの軸に目盛線を表示するかを選択する。
凡例を表示	凡例（はんれい）を表示する場所を選択する。

演習：散布図

dataset1.odsの「身長」と「体重」のデータから散布図を描こう。

グラフの作成

1. セルD1からE501の範囲を選択する。
2. メニューバーの[挿入]→[グラフ]を選択して「グラフウィザード」のダイアログを表示する。

① グラフの種類

1. [散布図] を選択する。
2. 右側の 4 つある種類から [点のみ] を選択する（デフォルトで選択されている）。
3. [次へ] をクリックする。

②データ範囲

はじめにデータ範囲を指定したので、ここは何もしなくてよい。

[**次へ**]をクリックする。

③データ系列

データ範囲から自動的にX値とY値が設定される。

1. 「データ範囲」の[名前]をクリックし、「名前の範囲」に「全体」と入力する。
2. [次へ]をクリックする。

④ グラフ要素

1. 「タイトル」に「**データセット 1 の身長と体重の分布**」と入力する。
2. 「X軸」に「**身長 [cm]**」と入力する。
3. 「Y軸」に「**体重 [kg]**」と入力する。
4. 「目盛線を表示」の[**X軸**]と[**Y軸**]にチェックを入れる。
5. [**完了**]をクリックする。
これで散布図が描かれる。

グラフの編集

LibreOffice Calcでグラフを編集するには2つのモードを使い分ける。

- グラフオブジェクトのモード
- グラフ編集モード

グラフオブジェクトのモード

シート上のグラフの領域をシングルクリックするとグラフオブジェクトが選択されてグラフオブジェクトモードになる。

グラフィオブジェクトのモード

このモードでマウスを使ってグラフィオブジェクトの移動やサイズ変更ができる。

グラフオブジェクトのモード

右クリックしてコンテキストメニューから以下のことができる。

項目	説明
位置とサイズ	グラフオブジェクトの位置やサイズを数値で指定したり、回転などができる。
画像（イメージ）としてエクスポート	グラフオブジェクトを画像として保存する。
編集	グラフ編集モードに入る。

ツールバーのボタンを使っても同じようなことができる。

グラフ編集モード

以下のいずれかの操作でグラフ編集モードになる。

- シート上のグラフの領域を**ダブルクリック**
- グラフオブジェクトのモードで右クリックのコンテキストメニューから[**編集**]を選択

グラフ編集モード

- グラフ編集モードでは、グラフオブジェクトの様々な部品を選択して設定変更ができる
- 選択した部品はツールバーに表示される。
- 部品を選択した状態でツールバーや右クリックしてコンテキストメニューから設定変更を行う。

グラフオブジェクト

東北 6 県の総人口の推移 ← タイトル

サブタイトル

出典:総務省統計局「国勢調査:男女別人口及び人口性比ー全国,都道府県(大正9年~平成27年)」
<https://www.stat.go.jp/data/kokusei/2015/kekka.html>, (参照 2020-10-01)

y軸

グラフ

ラベル
↓
人口[万人]

データ系列

凡例

- 青森県
- 岩手県
- 宮城県
- 秋田県
- 山形県
- 福島県

西暦[年] ← ラベル

x軸

共通の項目

項目	説明
グラフの種類	グラフの種類を変更する。
データ範囲	データ範囲や系列を変更する。

「グラフ」の項目

項目	説明
位置とサイズ	グラフオブジェクト内でのグラフの位置やサイズを変更する。
タイトルを挿入	タイトル、サブタイトル、X軸とY軸のラベルを挿入する。
軸を挿入/削除	X軸やY軸を挿入/削除する。
凡例を挿入/削除	凡例を挿入/削除する。

「グラフ」の位置やサイズ変更はマウスのドラッグでもできる。

「X軸」「Y軸」の項目

項目	説明
軸の書式	目盛の最小値・最大値・間隔、軸の位置、目盛の書式などを変更する。
軸を挿入/削除	軸（の数字）を挿入/削除する。
主目盛線の書式	主目盛線の書式を変更する。
副目盛線の書式	副目盛線の書式を変更する。
主目盛線を挿入/削除	主目盛線を挿入/削除する。
副目盛線を挿入/削除	副目盛線を挿入/削除する。

「データ系列」の項目

項目	説明
データ系列の書式	プロットする点のアイコン・色、点を結ぶ線などを変更する。
データラベルを挿入/削除	プロットする点のデータラベル（数値）を挿入/削除する。
平均値線の挿入/削除	平均の位置を表す線を挿入/削除する。
X誤差範囲を挿入/削除	X軸方向のエラーバーを挿入/削除する。
Y誤差範囲を挿入/削除	Y軸方向のエラーバーを挿入/削除する。
近似曲線を挿入/削除	データに最適な近似曲線（直線）を挿入/削除する。

演習：散布図（点を小さくする）

1. グラフ編集モードにする。
2. グラフ上の青い四角の点のどれかをダブルクリックして「データ系列」のダイアログを表示する。
3. [線]タブをクリックする。
4. 「アイコン」の「幅」と「高さ」の値を「0.05 cm」にする。
5. [OK]をクリックする。

演習：散布図（Y軸の表示範囲を縮める）

X軸やY軸の表示範囲はデフォルトだと自動的に決められるが、縮めたり広げたりしたい場合は手動で設定する。

1. グラフ編集モードにする。
2. グラフのY軸の数字をダブルクリックして「Y軸」のダイアログを表示する。
3. **[目盛]**タブをクリックする。
4. 「最小値」の**[自動]**のチェックを外し、値を「30」にする。
5. 「最大値」の**[自動]**のチェックを外し、値を「90」にする。
6. **[完了]**をクリックする。

今回の問い

- グラフにはどのような種類があるか？
 - 棒グラフ、円グラフ、折れ線グラフ、散布図、ヒートマップ、箱ひげ図などがある。
- LibreOffice Calcでグラフを作成するには何を用いるのか？
 - グラフウィザードを用いる。
- データの比較を行う際の注意点は何か？

グラフを描くときの注意点

グラフはデータの全体像や傾向を見るためには便利だが、グラフの棒や点の位置からその値を読み取ろうとすると目分量になって読み間違いが生じやすい。

グラフを描くときの注意点

これを避けるためにはグラフにデータラベルを表示したり、グラフだけでなく表も付けるようにするとよい。

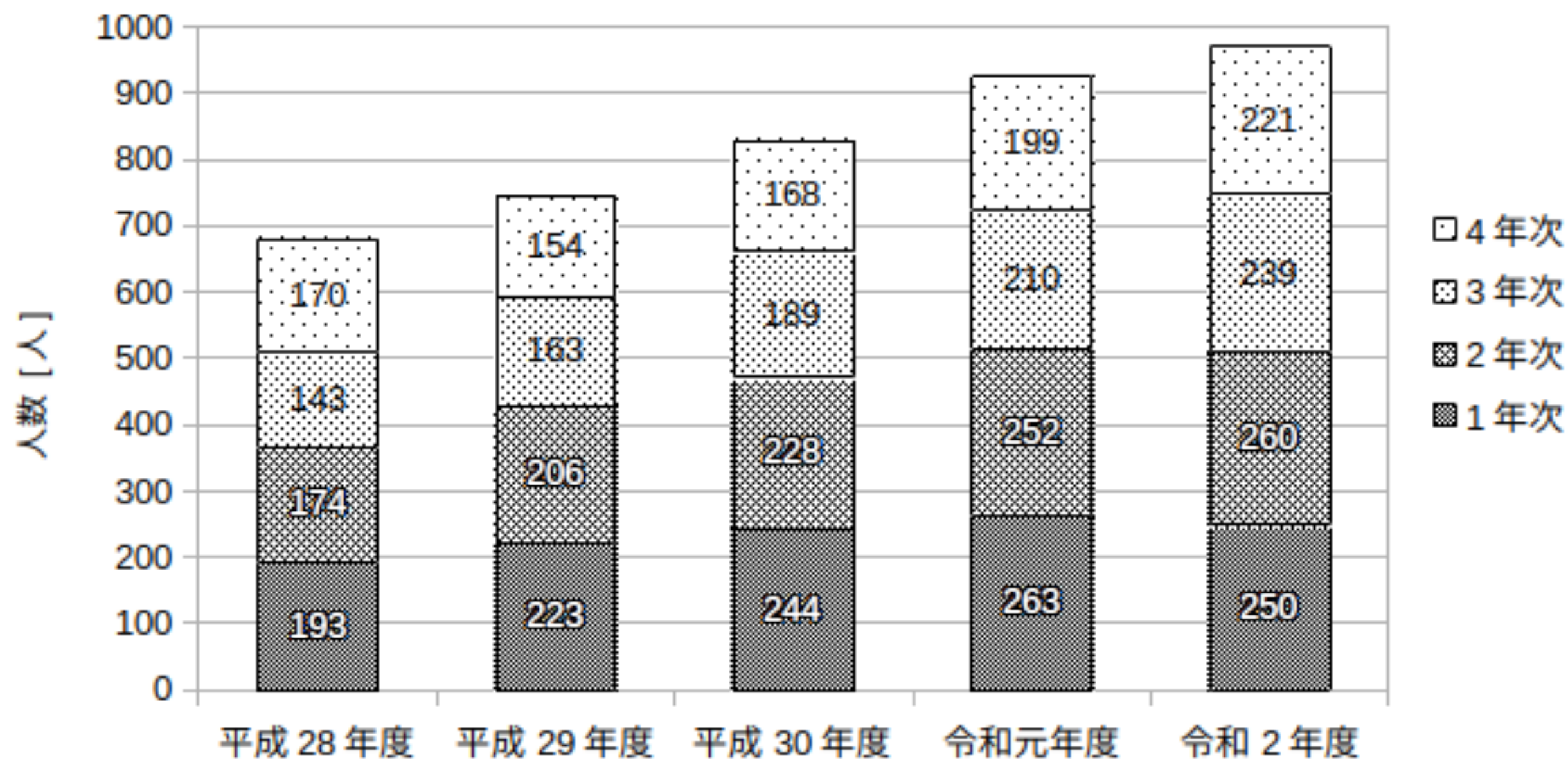
グラフを描くときの注意点

複数の系列を入れるときには、系列を区別できるように色を分けるが、似たような色だと区別がつきにくい。

グラフを描くときの注意点

グラフを印刷するときにモノクロ印刷しかできない場合は、**白黒の濃淡**や**斜線**や**網掛け**（パターン、ハッチング）で区別がつくようにする。

東北公益文科大学 在籍者数



グラフを描くときの注意点

グラフに対して3D化などの必要以上の装飾や加工を行うと読み間違いの元になりやすいので、**グラフはなるべくシンプルにする。**

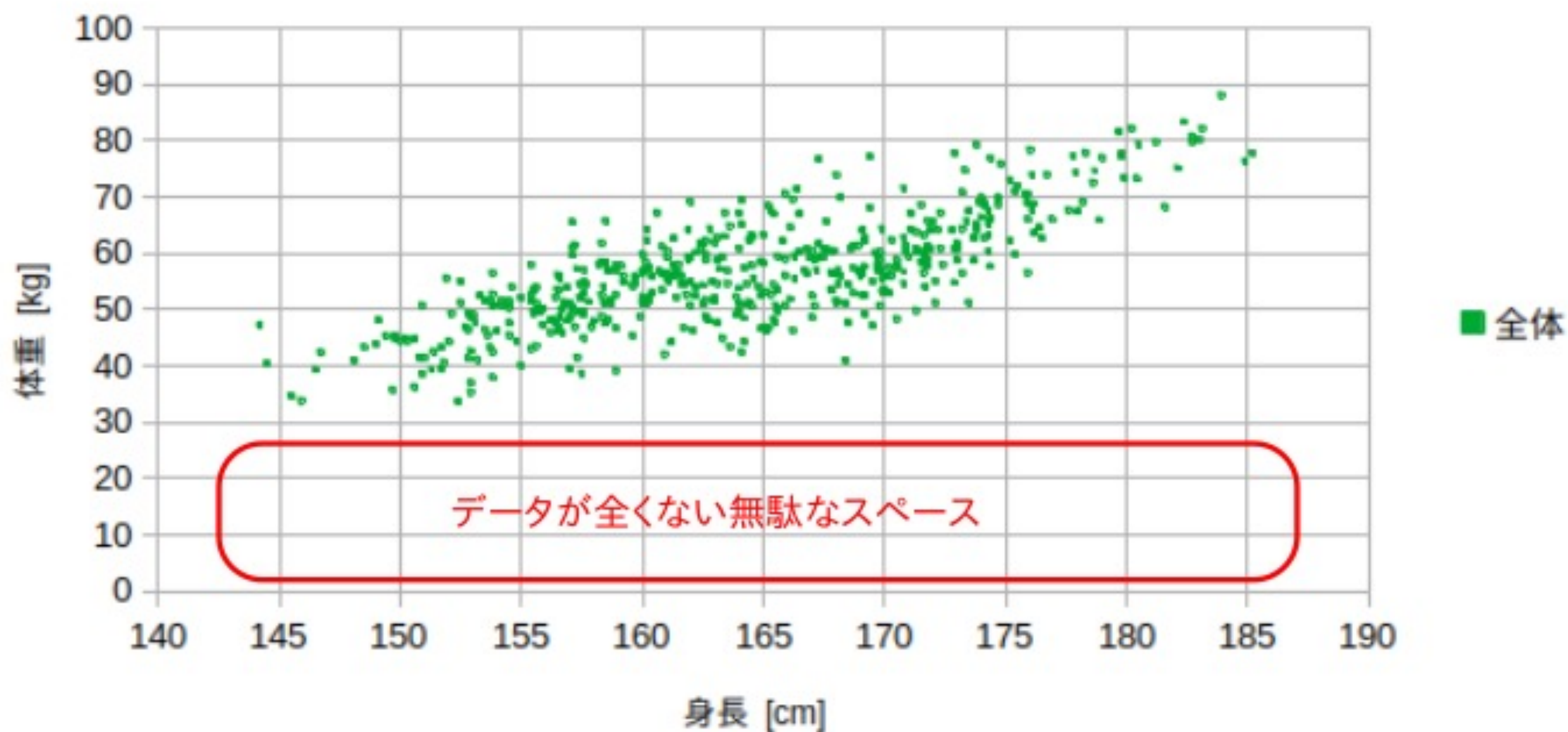
グラフを描くときの注意点

何のグラフなのか、縦軸と横軸は何なのか分かるように**タイトル**や**軸のラベル**を付ける。軸のラベルには**単位**も忘れないように付ける。

グラフを描くときの注意点

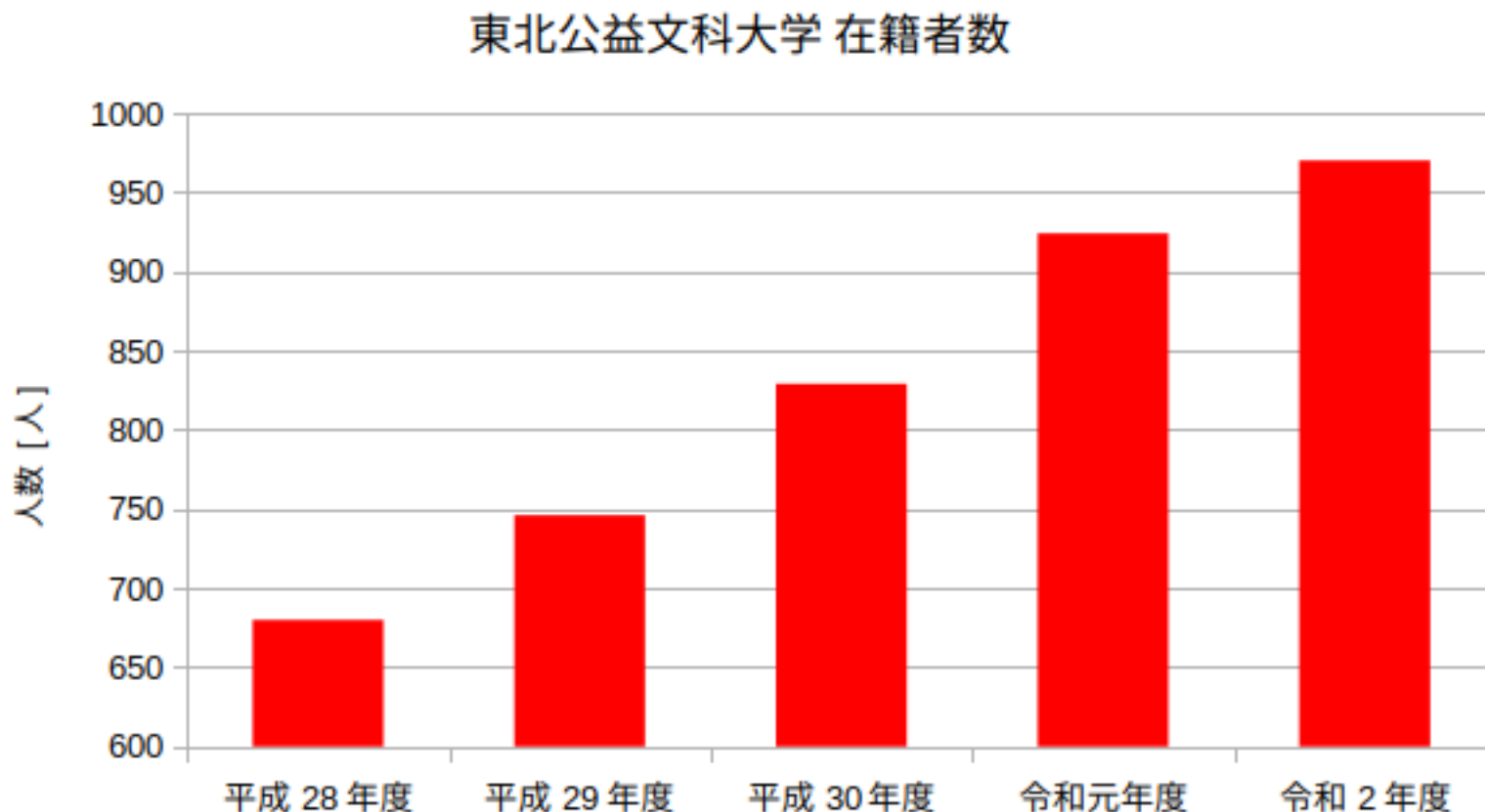
- 目盛の最小値と最大値はグラフ機能で自動的に決まる。
- データが全くない無駄なスペースが大きくなって、データが存在する範囲が狭くなってしまう場合は**最小値と最大値を手動で調整**する。

データセット 1 の身長と体重



グラフを描くときの注意点

グラフを見るときは、どこから値が始まっているのかに注意して、誇張表現に惑わされないように。



演習：散布図

dataset2.csvの「授業外学修時間[h]」と「経営学基礎」の点数のデータから散布図を描こう。タイトルは「経営学基礎の点数と授業外学修時間の関係」、X軸名は「授業外学修時間[h]」、Y軸名は「経営学基礎の点数」としよう。

今回学ぶこと

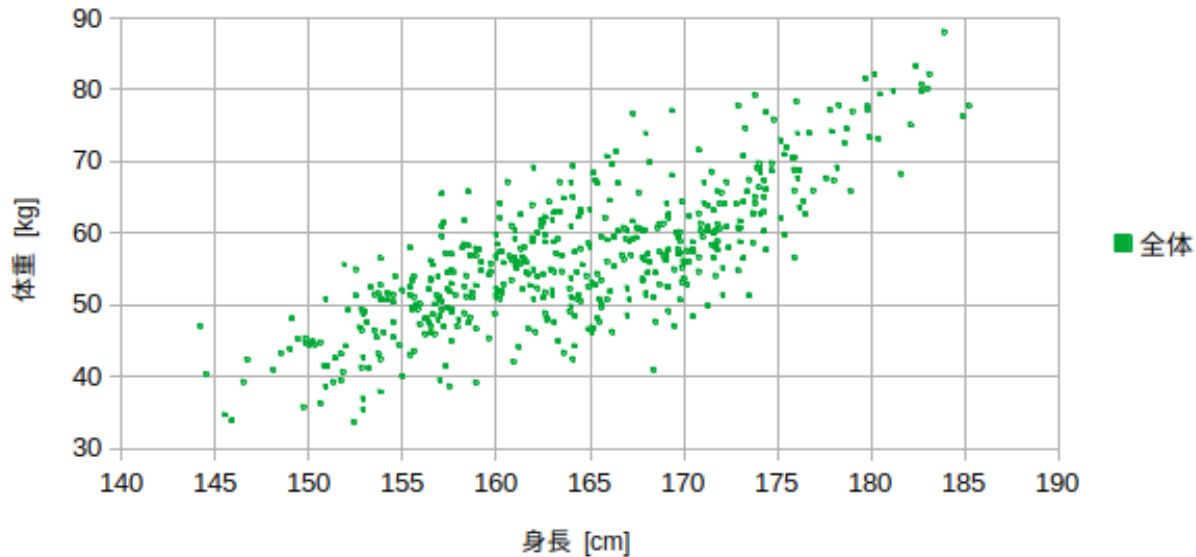
1. グラフの種類
2. グラフの作成
3. データの分析と比較

層別の分析が必要なデータ

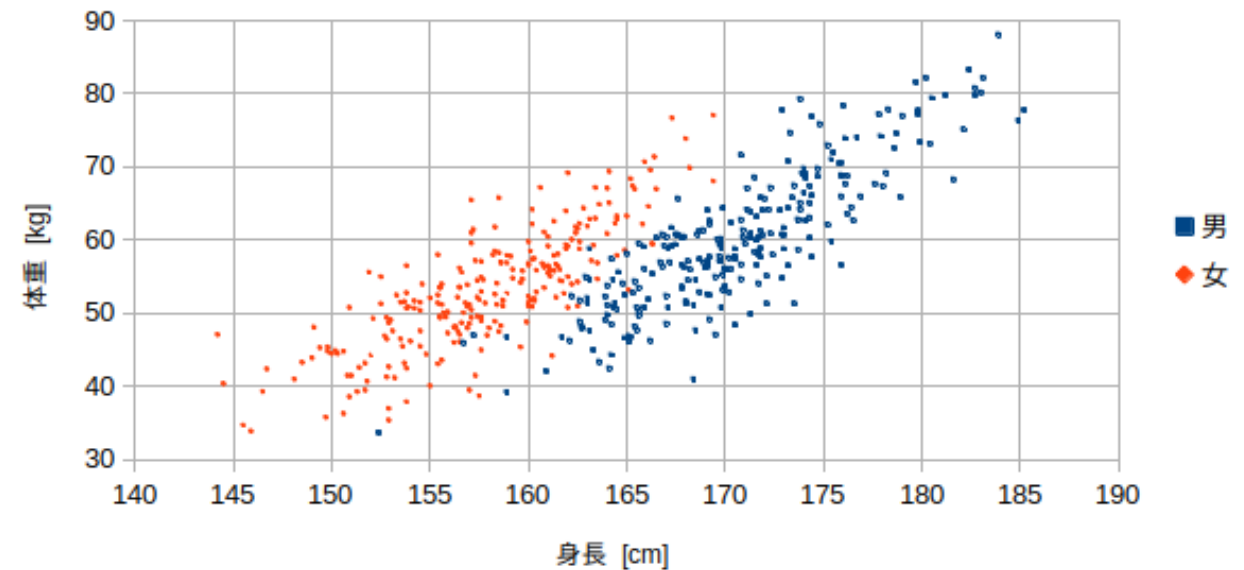
性別のような異なる属性が混ざっているデータは、**属性ごとに分けて**層別のデータにして分析する必要がある。

層別の分析が必要なデータ

データセット 1 の身長と体重



データセット 1 の身長と体重



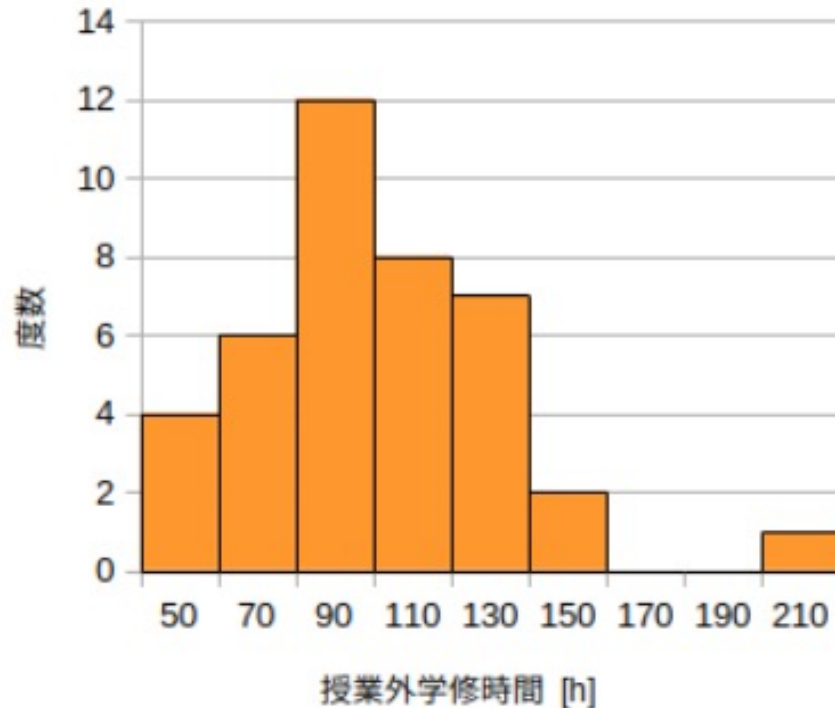
男女で関係性が異なるデータを混ぜてしまうと、散らばりが大きくなり関係性も変わってしまう。

条件をそろえた比較

データの比較をする際には、単にグループに分けてグラフにするだけでなく、条件をそろえた比較が必要

条件をそろえた比較

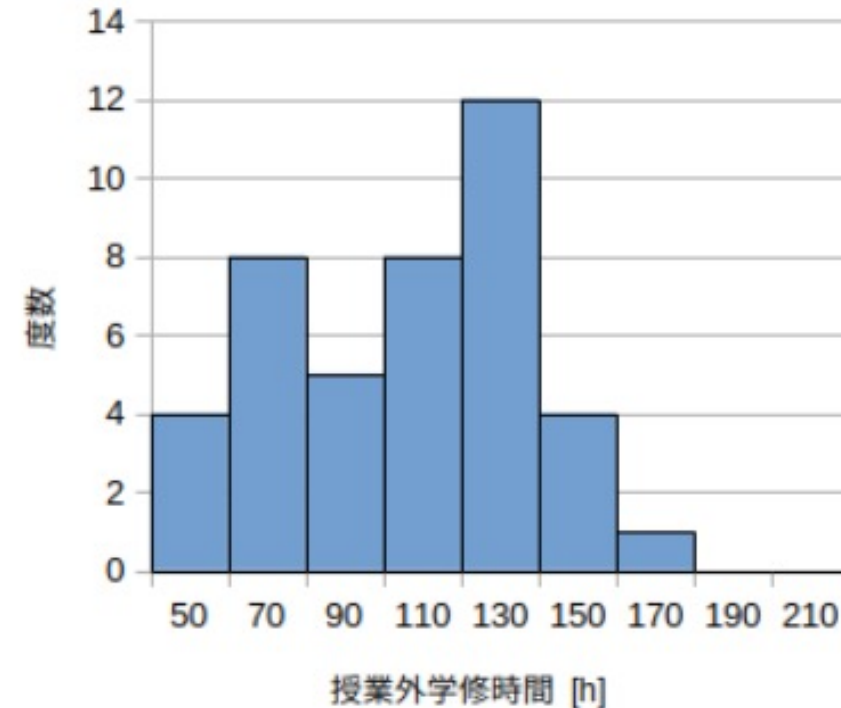
授業外学修時間（経営コース 2年）



比較？



授業外学修時間（政策コース 3年）



- コースによる授業外学修時間の違いを比較するなら、学年をそろえて比較した方がよい。
- 学年による授業外学修時間の違いを比較するなら、同じコースにするか、コースで分けずに比較した方がよい

今回の問い

- グラフにはどのような種類があるか？
 - 棒グラフ、円グラフ、折れ線グラフ、散布図、ヒートマップ、箱ひげ図などがある。
- LibreOffice Calcでグラフを作成するには何を用いるのか？
 - グラフウィザードを用いる。
- データの比較を行う際の注意点は何か？
 - データは条件をそろえて比較する。

今回学んだこと

1. グラフの種類
2. グラフの作成
3. データの分析と比較

今回の目標

グラフを用いてデータを可視化できるようにする。

課題

地図で見る統計 (jSTAT MAP)
(<https://jstatmap.e-stat.go.jp/trialstart.html>)で出身地付近のヒートマップを描きなさい。

課題手順

1. [ログインしないで始める]をクリックする。
2. 左上の検索フォームに地名を入力し、[Enter]キーを押して地図を表示する。
3. 右下の[統計地図作成]→[統計グラフ作成]をクリックして「統計グラフ作成」のダイアログを開く。
4. 「集計地域」で[5次メッシュ(250 mメッシュ)]を選択する。
5. 「統計名」で「国勢調査」を選択する。
6. 「年または年月」で[2020年]を選択する。
7. 「統計表（表題）」から[人口及び世帯]を選択する。
8. 「指標/データ」から[人口総数]にチェックを入れて、[指標選択]をクリックする。
9. [次へ]をクリックする。
10. 「集計地域」で[メッシュ]を選択する。
11. [集計開始]をクリックする。

課題保存方法

[統計地図作成]→[レポート作成]→[シンプルレポート]で「シンプルレポート」のダイアログを表示し、[2020年 国勢調査 5次メッシュ (250 mメッシュ)人口及び世帯]を選択して[Excelレポート作成]をクリックしてダウンロードする。

課題提出方法

保存したExcelファイル（拡張子.xlsx）を提出する。

提出先：yutaka.hirai@koeki-u.ac.jp

件名：データリテラシー第6回課題 [学籍番号]

締め切り：11月13日 (金)

次回

第 1 回	社会におけるデータ・AI利活用
第 2 回	表計算ソフトの基本操作
第 3 回	セルの参照と集計
第 4 回	度数分布表とヒストグラム
第 5 回	代表値と散らばりの尺度
第 6 回	データのグラフによる表現とデータ比較
第 7 回	データの分布と標本抽出
第 8 回	推定
第 9 回	分割表とクロス集計
第 1 0 回	2 変量データと相関
第 1 1 回	時系列データ
第 1 2 回	仮説検定
第 1 3 回	回帰分析