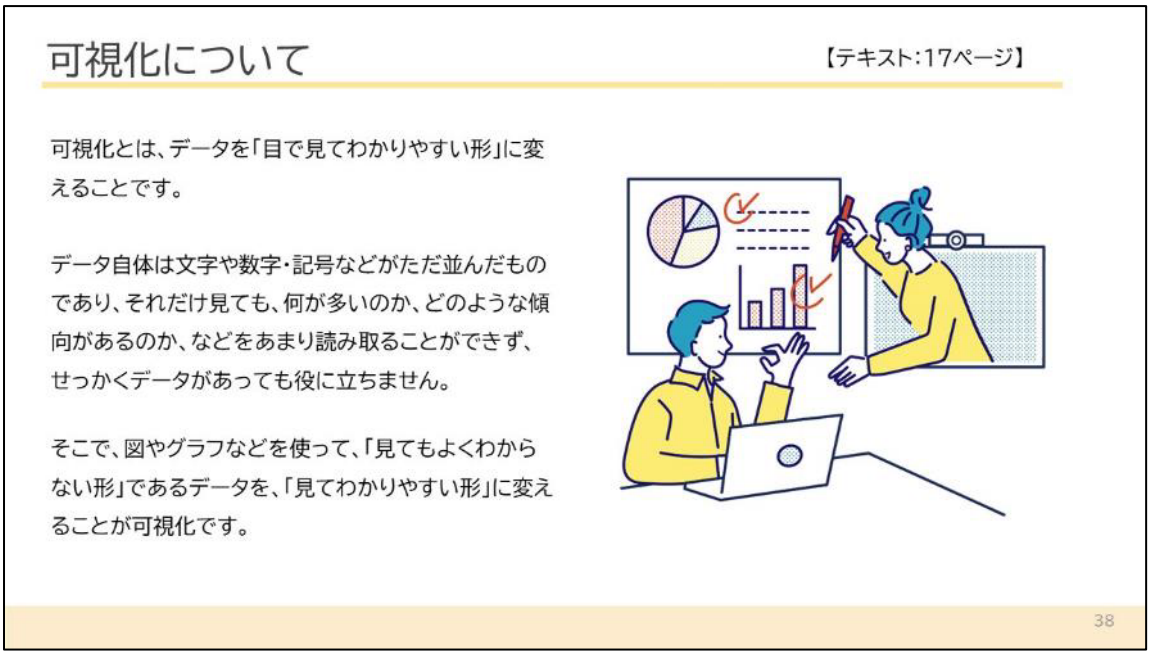
第二章
データの可視化



37



【スライドの解説】

■可視化とは

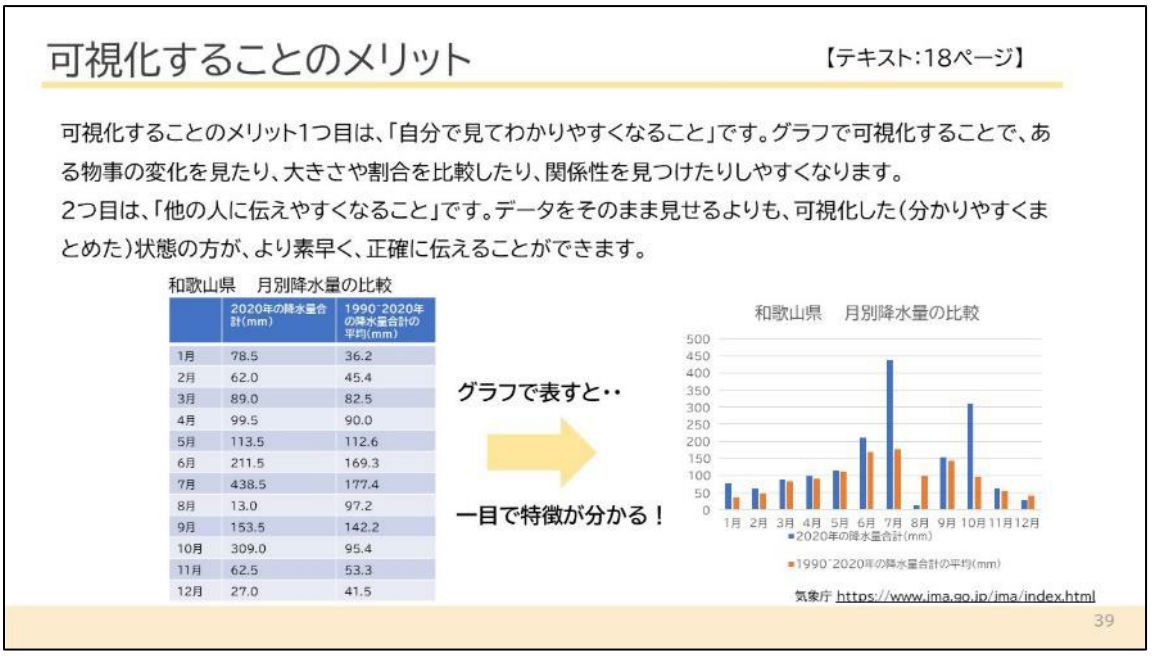
「見てもよくわからない」データを、図やグラフなどを使って、「目で見てわかりやすい形」に変えることです。

【このページのポイント】

図やグラフを用いると、データの全体像を分かりやすく伝えることができます。

【キーワード】

可視化



【スライドの解説】

■可視化（目で見える表現に変換）することのメリット

メリット①：ひと目見ただけではっきりとわかるようになることです。グラフで可視化することで、ある物事の変化を発見したり、大きさや割合を比較したり、関係性に気がついたりすることが容易になります。

メリット②：他の人に伝わりやすくなることです。データをそのまま見せるよりも、ひと目見ただけではっきりとわかるような表現の方が、正確に伝わります。

【このページのポイント】

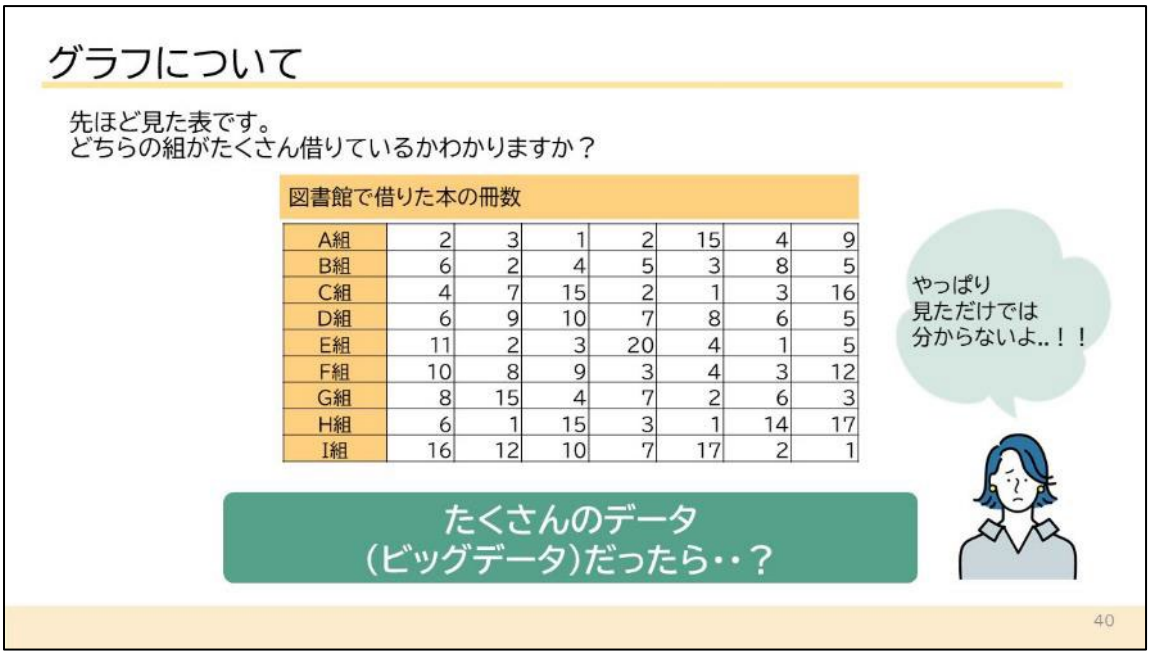
データを可視化することのメリットは、

・「自分で見てわかりやすくなること」

・「他の人に伝えやすくなること」

【キーワード】

可視化、グラフ



【スライドの解説】

■グラフについて

前にもお見せしたデータです。

A組からI組までの生徒たちが各月に図書館から借りた本の冊数を表にまとめた表です。

各クラスの借り出しにどんな特徴があるのかを考えてみましょう。

表ですと、特徴の前に、どのクラスが最も借りた冊数が多いのかが分かりません。

A組36冊　B組33冊　C組48冊　D組51冊　E組46冊

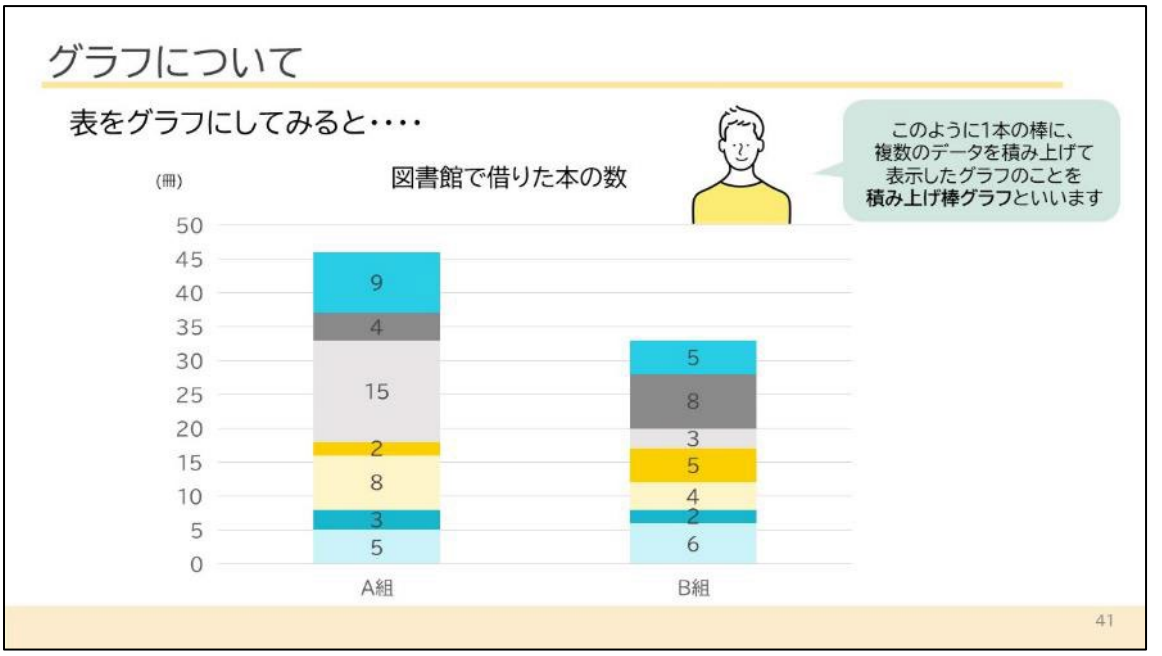
F組49冊　G組45冊　H組57冊　I組65冊

【このページのポイント】

データの量が多いと、表の数値だけを見ただけでは特徴が把握し難いです。

【キーワード】

可視化、グラフ



【スライドの解説】

■データの可視化についての具体例①

・積み上げ棒グラフ：A組とB組の生徒たちが各月に図書館から借りた本の冊数をまとめた表

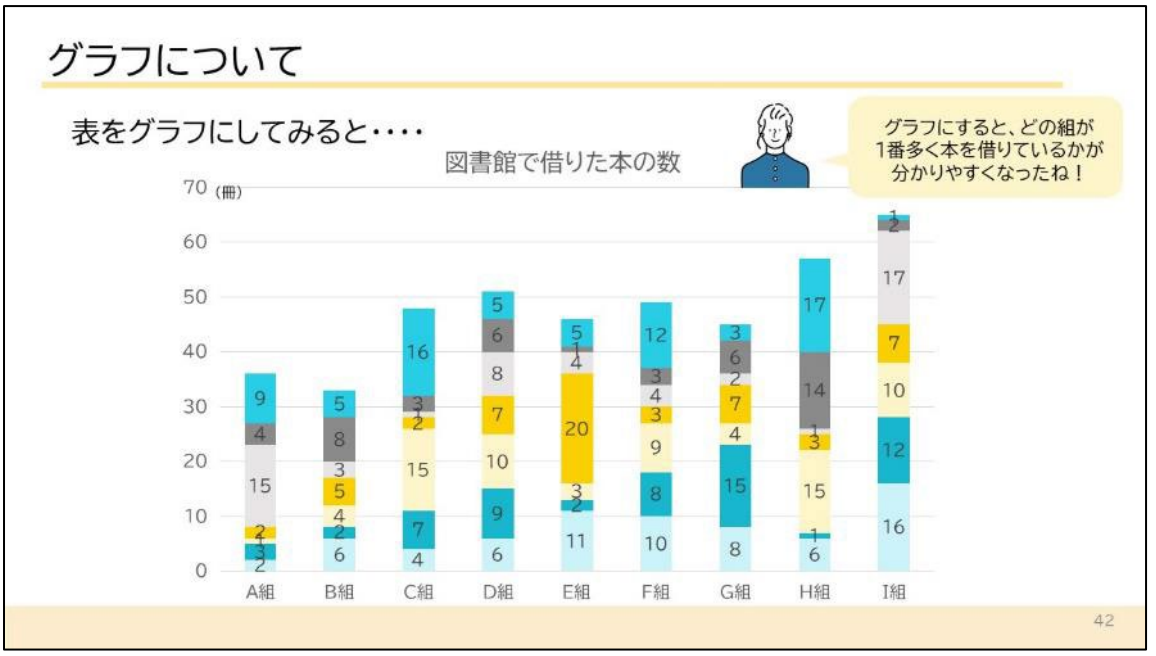
各月ごとの借りた冊数を色で分けて積み上げたグラフです。A組の借りた冊数は45冊と多いことが一目で分かります。B組の借りた冊数は33冊です。

【このページのポイント】

適切なグラフで可視化すると、得られたデータの特徴が目で把握できます。

【キーワード】

可視化、グラフ、積み上げ棒グラフ



【スライドの解説】

■データの可視化についての具体例②

・積み上げ棒グラフ：A組からI組の生徒たちが各月に図書館から借りた本の冊数をまとめた表

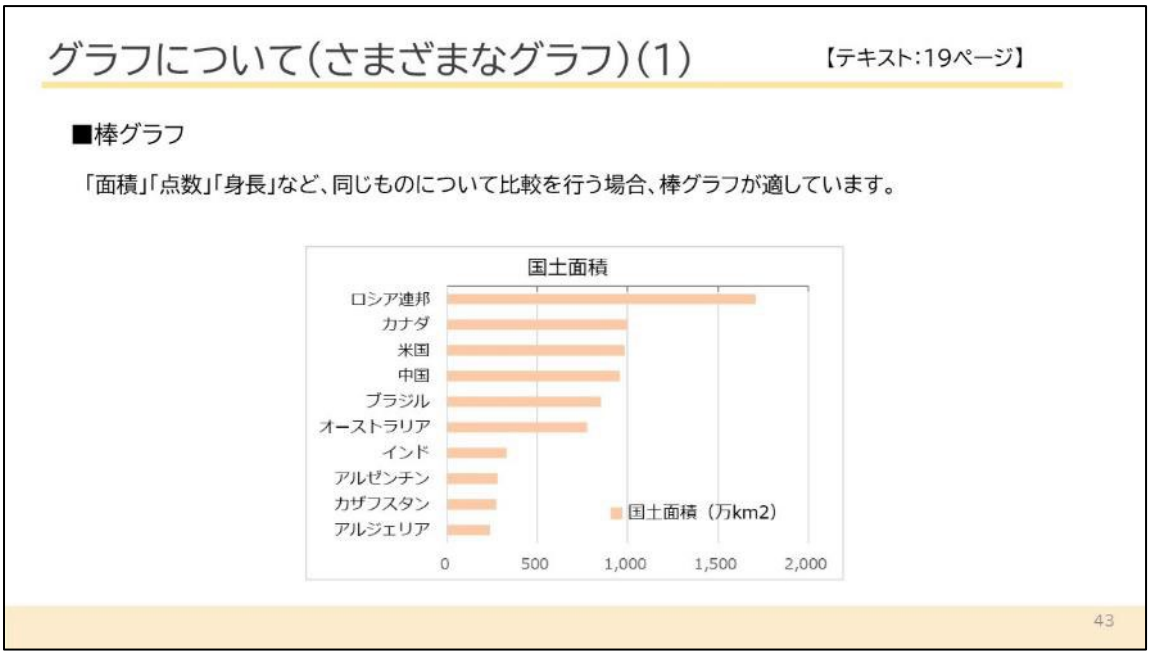
各月ごとの借りた冊数を色で分けて積み上げたグラフです。I組の借りた冊数は65冊と多いことが一目で分かります。一番少ないのがB組33冊です。

【このページのポイント】

適切なグラフで可視化すると、得られたデータの特徴が目で把握できます。

【キーワード】

可視化、グラフ、積み上げ棒グラフ



【スライドの解説】

グラフにはいくつかの種類があります。何を表したいかによって適しているグラフは異なります。ここでは代表的なグラフをいくつか紹介します。

■棒グラフ

棒グラフとは数量を棒の長さで示したグラフのことです。棒グラフは、棒と棒の間をあけます。

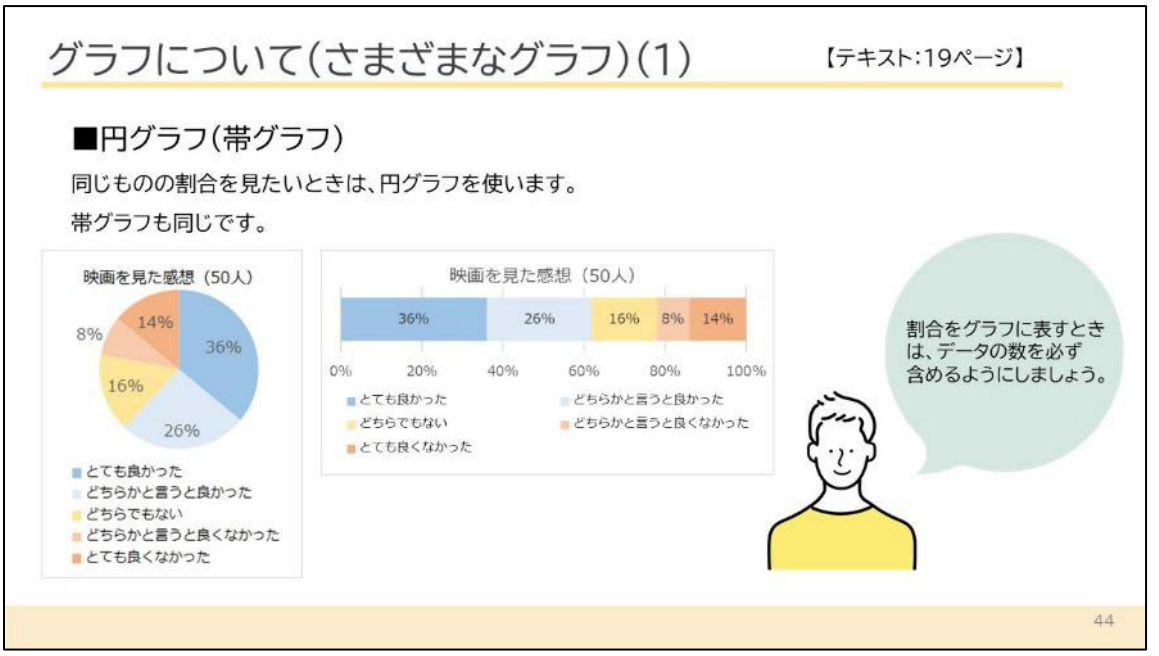
・例：国土面積を国別で表しています。国名は、名義尺度です。

【このページのポイント】

「面積」「点数」「身長」など一つの統計データの値を比較する場合は棒グラフが適しています。

【キーワード】

棒グラフ



【スライドの解説】

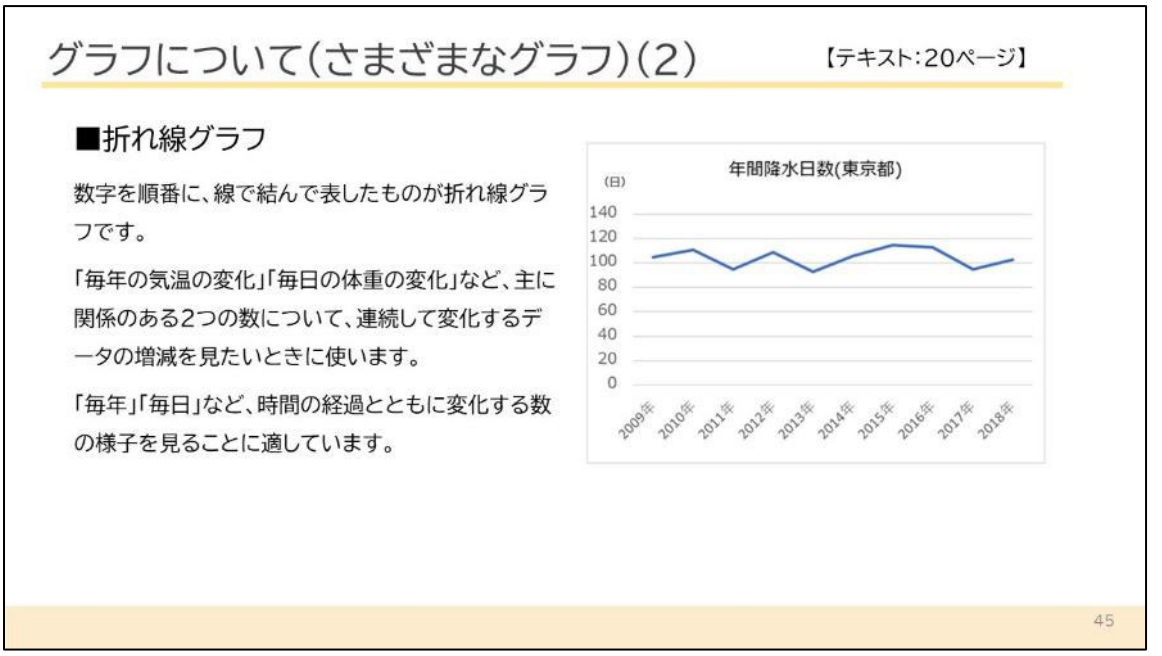
■円グラフ

「映画を見た感想」など１つのことについての回答の割合を比較したいときは、円グラフや帯グラフを使います。

「映画を見た感想」の回答は、「とてもよかった」、「どちらかというとよかった」などであり、大小関係にのみ意味がありますので順序尺度です。

【このページのポイント】

円グラフ、帯グラフ



【スライドの解説】

■折れ線グラフ

気温や降水量などデータを時間の経過順に集めることがあります。時間の経過順に並んだデータを「時系列データ」と言います。時系列のデータのように時間と気温や降水量などの変化を見たい場合やデータ数が多いときには、数字を線で結んで表わす折れ線グラフで表現します。

気温や降水量は、比例尺度です。

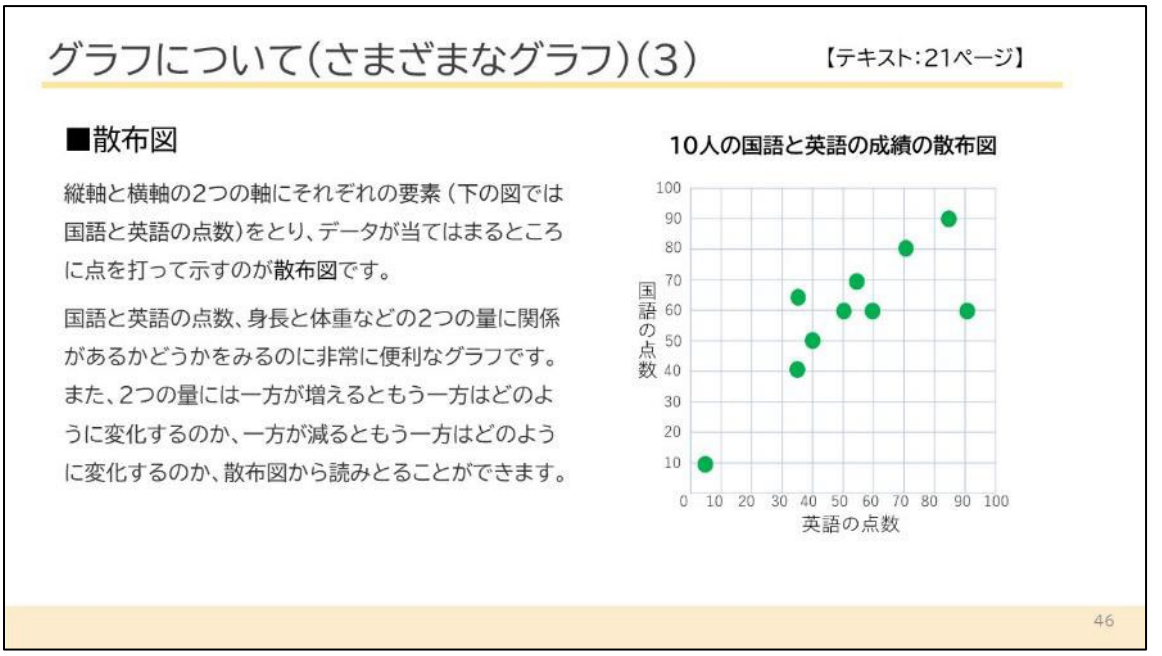
【このページのポイント】

時系列データのように、時間の経過とともに変化する「もの」の「変化」を見たいときに折れ線グラフで表わします。

時系列データを折れ線グラフで表す場合には、横軸に時間を、縦軸に時間の経過とともに変化する「もの」の値をとり、折れ線で結びます。

【キーワード】

時系列データ、折れ線グラフ



【スライドの解説】

■散布図

国語と英語の点数、身長と体重などの2つの量にどのような関係があるかを把握するのに非常に便利なグラフです。散布図で表すと、2つの量の一方が増えるともう一方はどのように変化するのか、一方が減るともう一方はどのように変化するのかを読みとることができます。

英語と国語という２つの得点データの散布図を描くには、横軸に英語の点数をとり、縦軸に国語の点数をとります。そして、生徒それぞれの国語と英語の点数が交わったところに点を打ちます。

【このページのポイント】

散布図は、2つの量に関係があるかどうかをみるのに非常に便利なグラフです。

【キーワード】

散布図、散布図の書き方

グラフについて(さまざまなグラフ)(3)　【テキスト:21ページ】

散布図の書き方

散布図を作成するには2ステップがあります。
1、関係を調べたいデータを縦軸と横軸にとる
2、プロット(打点)する

右に10人の国語と英語のテストの成績があります。この10人のデータを使った散布図を見てみましょう。そして国語と英語の点数にどのような関係があるのかを考えてみましょう。

10人の国語と英語の成績

	国語	英語
後藤	50点	40点
佐藤	80点	70点
山崎	10点	5点
鈴木	60点	50点
伊藤	60点	60点
清水	60点	90点
三上	90点	85点
遠藤	65点	35点
小川	70点	55点
戸田	40点	35点

47

【スライドの解説】

■散布図の書き方（2ステップ）

step1.　関係を調べたい、英語と国語の点数といったようなそれぞれの2つのデータを用意し、横軸に英語の点数をとり、縦軸に国語の点数をとります。

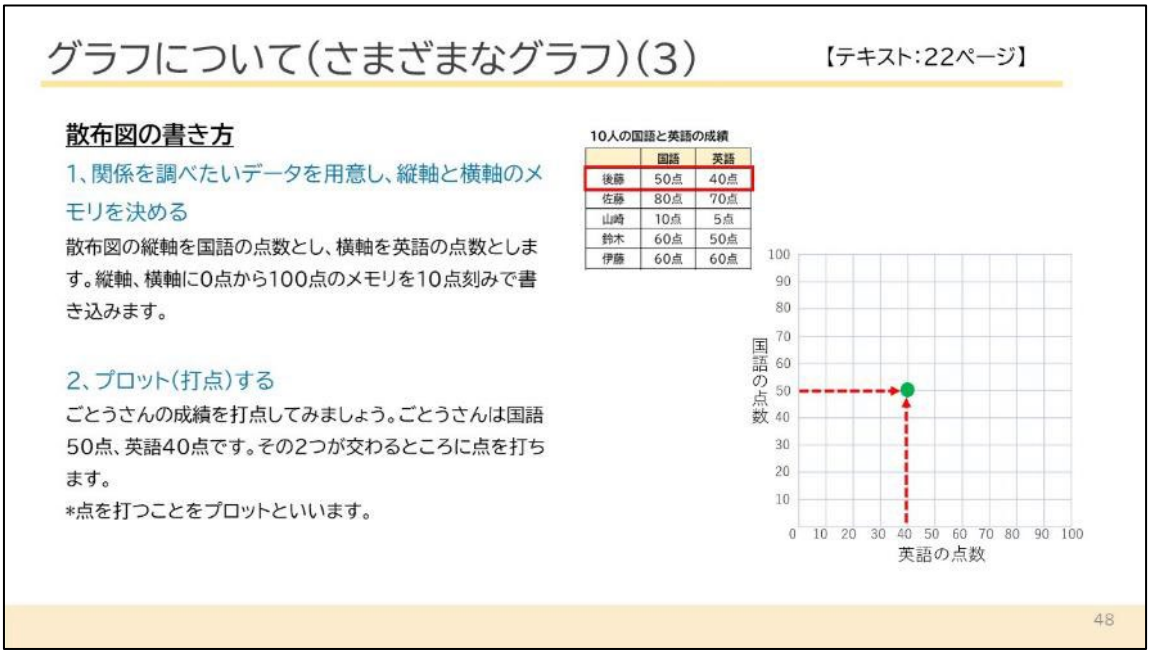
step2.　生徒それぞれの国語と英語の点数が交わったところに点を打ちます。

【このページのポイント】

散布図を書くには2つのステップがあります。

【キーワード】

散布図の書き方



【スライドの解説】

■散布図の書き方（詳細）

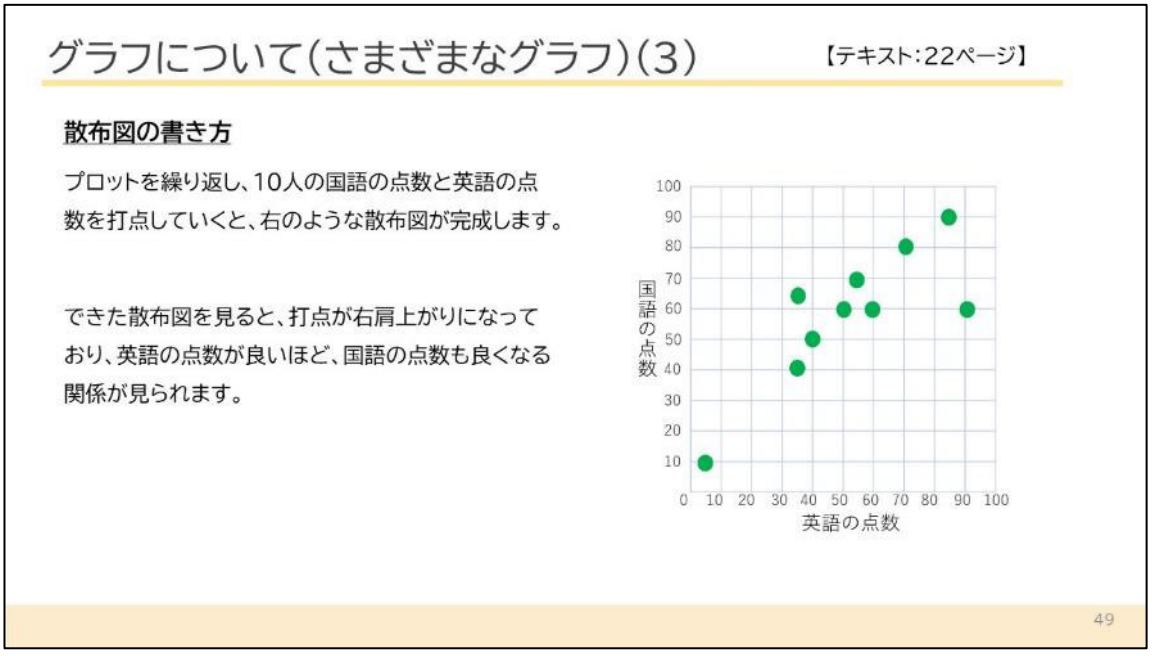
散布図を書くには、関係を調べたい、英語と国語の点数といった生徒それぞれの2つのデータが必要です。横軸に英語の点数の目盛りとして0点から100点を10点刻みにとり、縦軸に国語の点数の目盛りとして0点から100点を10点刻みにとります。そして、生徒それぞれの国語と英語の点数が交わったところに点を打ちます。この例では、後藤さんの国語と英語の点数を打っています。点を打つことを「**プロット**」と言います。

【このページのポイント】

点を打つことを「プロット」と言います。

【キーワード】

プロット



【スライドの解説】

■散布図の書き方（詳細・続き）

10人の生徒それぞれの国語の点数と英語の点数をプロットしていくと、上のような散布図が完成します。

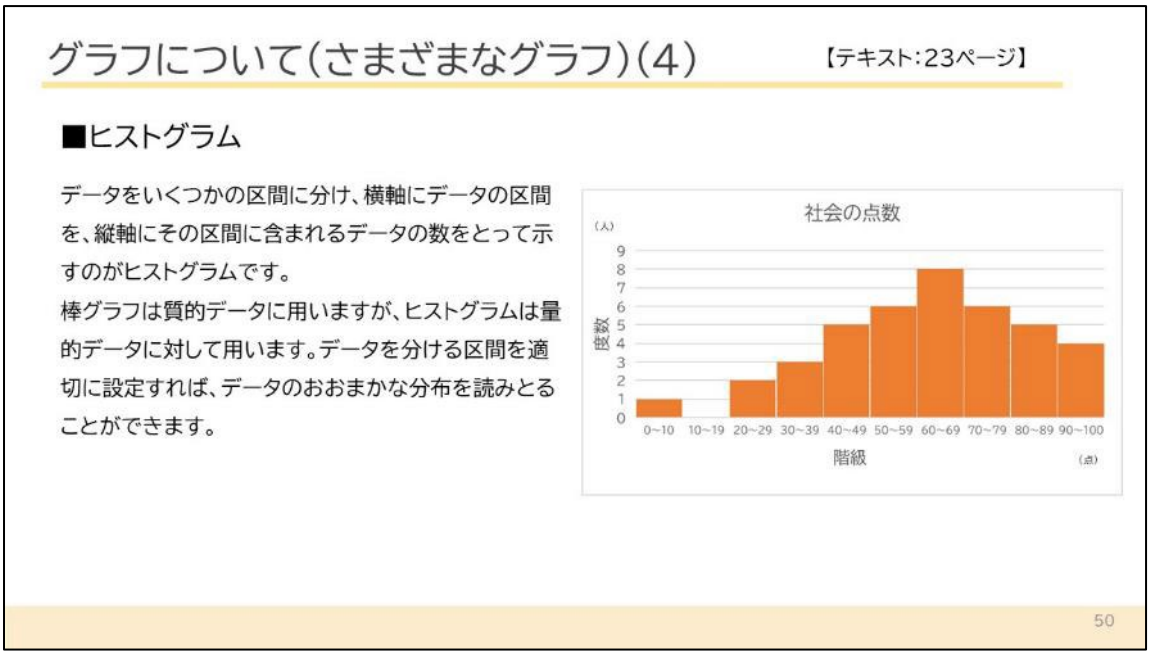
英語の点数が高いほど、国語の点数も高くなる関係が見られます。

【このページのポイント】

散布図は2つの量にどのような関係があるかをみるのに非常に便利なグラフです。

【キーワード】

散布図から読み取れること



【スライドの解説】

■ヒストグラム

ヒストグラムは、社会の点数のように、連続する数値のデータを表すグラフです。棒グラフと異なり、棒と棒の間が空いていません。

ヒストグラムは、データをいくつかの区間に分け、各区間に含まれるデータの数を表します。

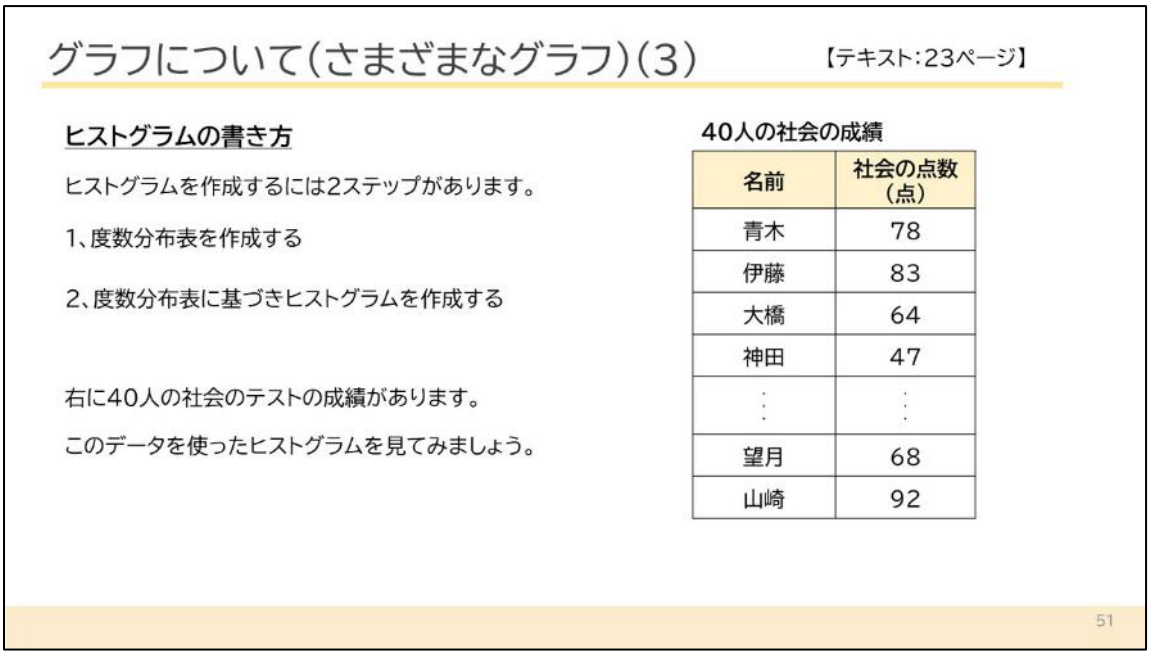
【このページのポイント】

ヒストグラムは、連続する数値のデータ(量的データ)を表すグラフです。

棒グラフは国名、スイーツの種類など数値ではないデータ(質的データ)に用います。

【キーワード】

ヒストグラム、棒グラフ、量的データ、質的データ



【スライドの解説】

■ヒストグラムの書き方(２ステップ)

step1.　度数分布表を作成する。

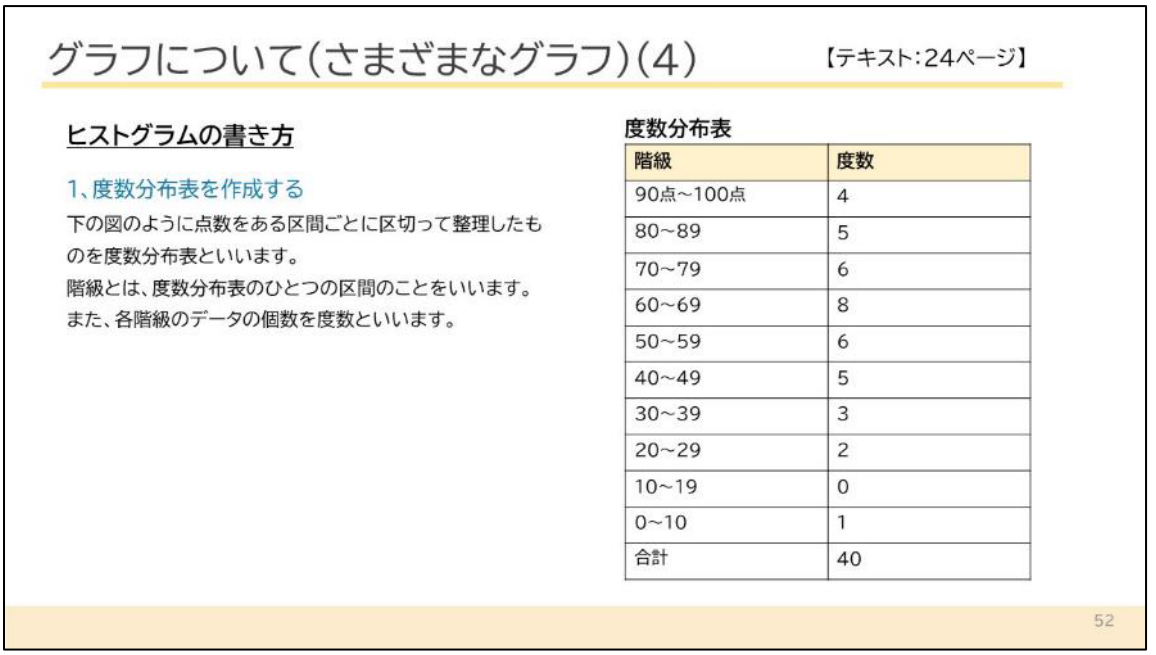
step2.　度数分布表に基づきヒストグラムを作成する。

【このページのポイント】

ヒストグラムを作成するには２ステップがあります。

【キーワード】

ヒストグラムの作成手順



【スライドの解説】

■ヒストグラムの書き方（詳細）

step1.　度数分布表を作成する

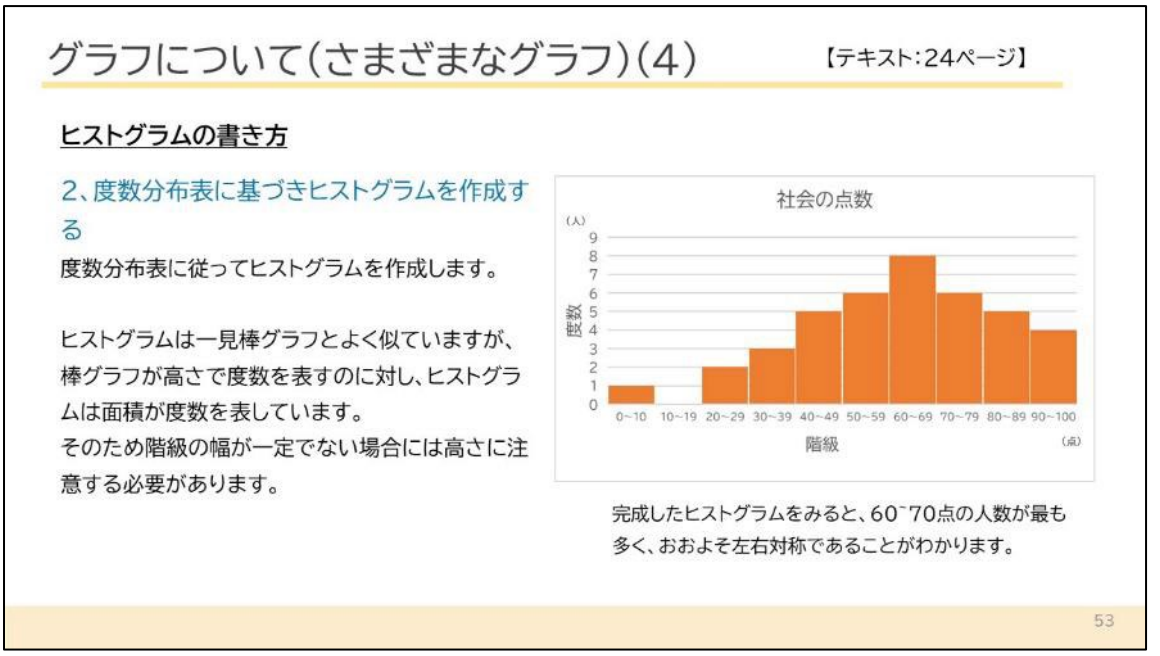
上の図のように社会科の点数0点から100点までを10等分します。10等分した各区間、0点以上から10点未満、10点以上から20点未満などを**階級**と言います。各階級に含まれる生徒の数(データ数)が**度数**です。**度数分布表**は階級と度数からなります。

【このページのポイント】

階級と度数からなる表が**度数分布表**です。

【キーワード】

度数分布表、階級、度数



【スライドの解説】

■ヒストグラムの書き方（詳細・続き）

step2.　度数分布表に基づきヒストグラムを作成する

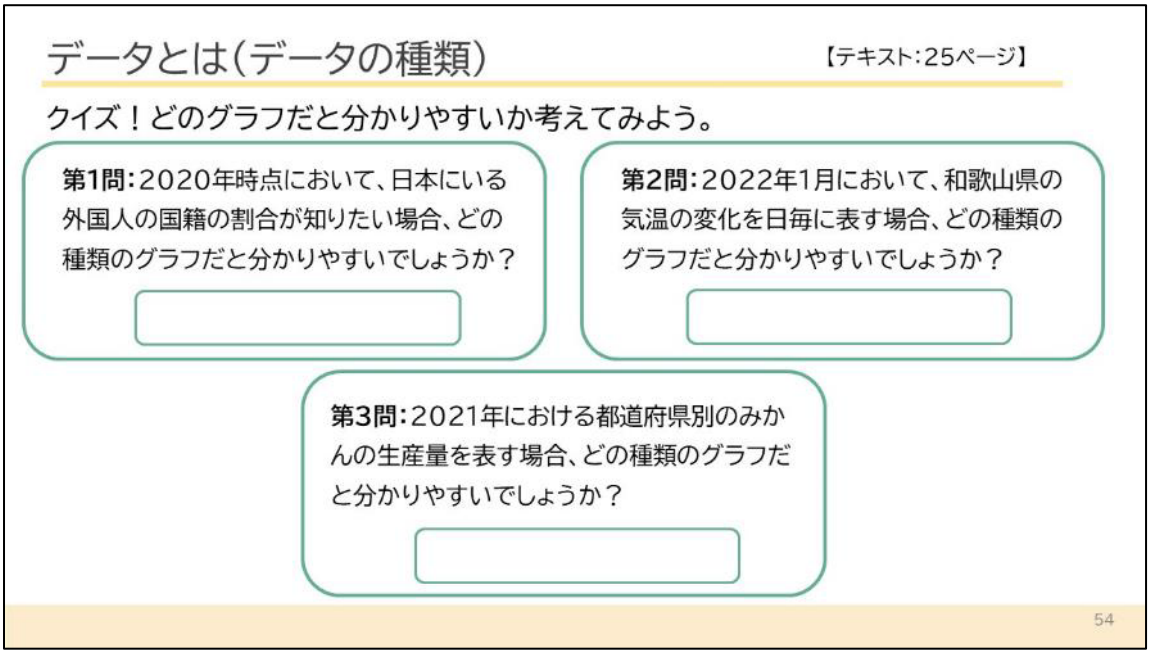
度数分布表に従ってヒストグラムを作成します。ヒストグラムは一見棒グラフとよく似ていますが、棒グラフが高さで度数を表すのに対し、ヒストグラムは面積が度数を表しています。そのため階級の幅が一定でない場合には高さに注意する必要があります。

【このページのポイント】

ヒストグラムは一見棒グラフとよく似ています。しかし、棒グラフが高さで度数を表すのに対し、ヒストグラムは面積が度数を表しています。

【キーワード】

度数分布表、階級、度数、ヒストグラム



【スライドの解説】

■データの種類とグラフについての問題

適したグラフを選んでみましょう。

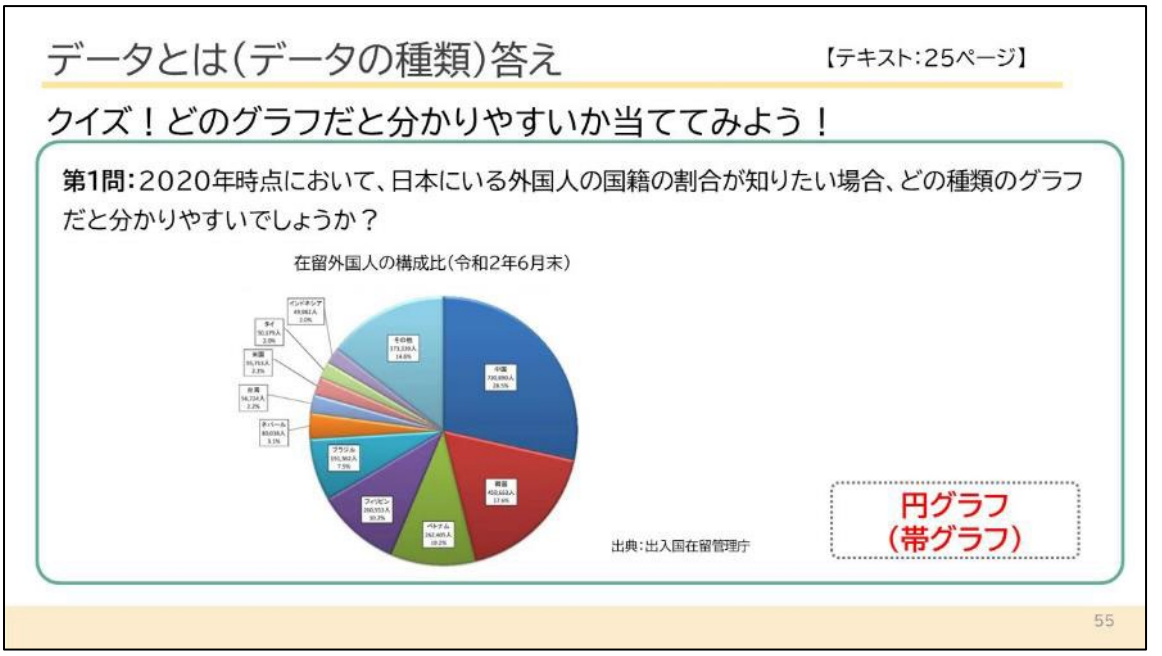
第1問　割合を示すにはどのようなグラフが適しているでしょうか。

第2問　日付や時間とともに変化することを示すにはどのようなグラフが適しているでしょうか。

第3問　都道府県別のみかんの生産量を示すにはどのようなグラフが適しているでしょうか。

【このページのポイント】

示したことに応じた適切なグラフがあります。



【スライドの解説】

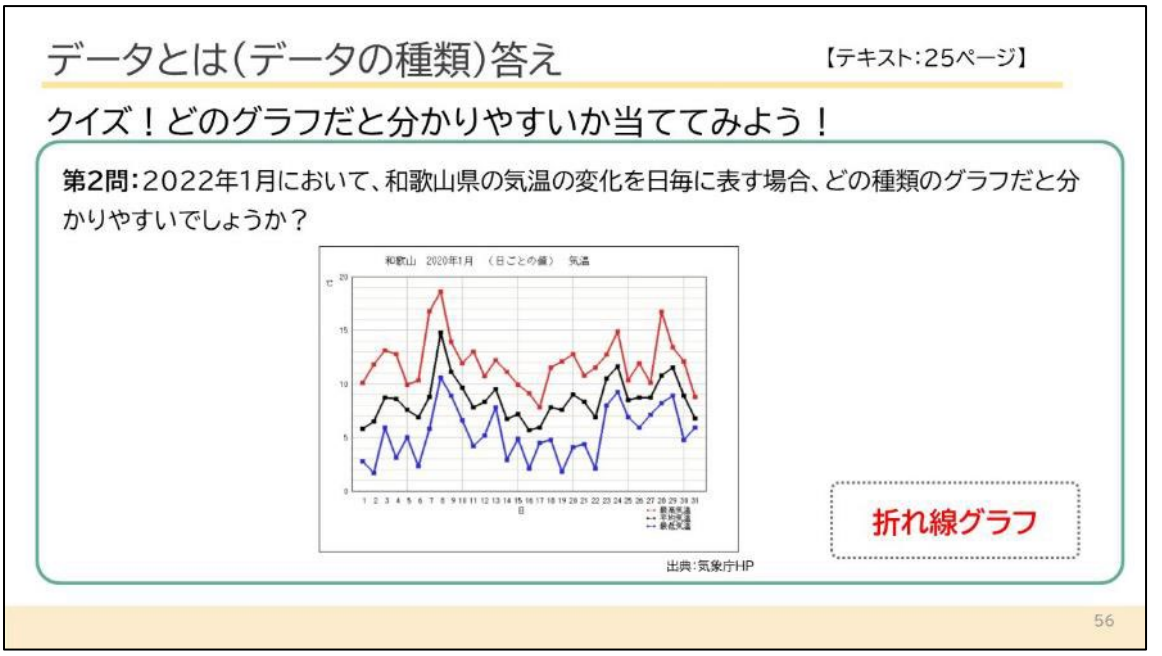
■答え

外国人の国籍という１つのもの「割合」を表すには、円グラフや帯グラフです。

国籍は名義尺度です。

【このページのポイント】

１つのもの「割合」を表すには、円グラフや帯グラフが適しています。



【スライドの解説】

■答え

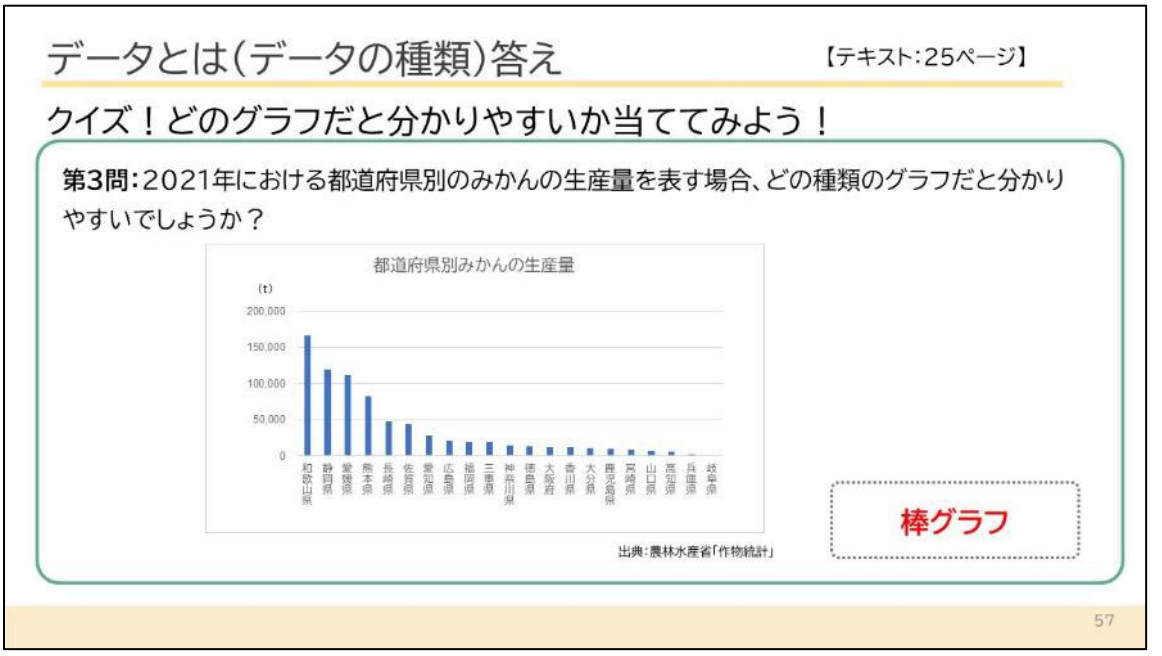
和歌山県の気温の変化のデータは、時系列データです。

時系列データには折れ線グラフが適しています。

折れ線グラフとは、数字を順番に、線で結んで表したグラフです。

【このページのポイント】

「毎年」「毎日」など、時間の経過とともに変化する数の様子を表すには折れ線グラフが便利です



【スライドの解説】

■答え

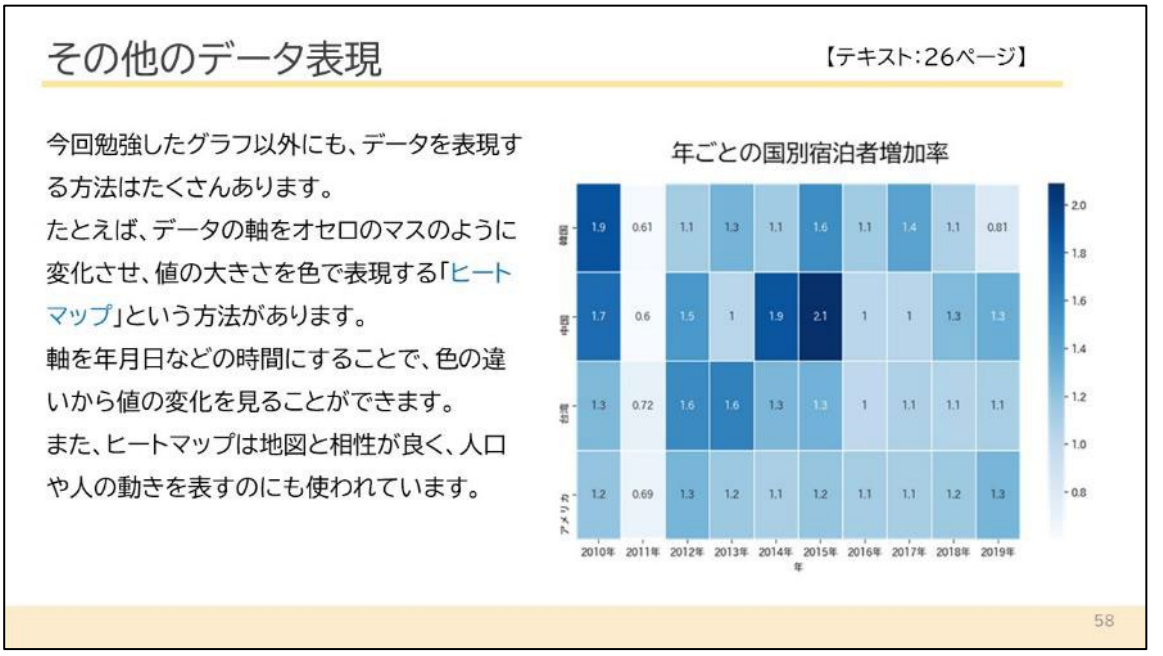
都道府県別のミカンの生産量を表すグラフは、棒グラフです。

横軸には都道府県、縦軸にはミカンの生産量をとります。

都道府県は名義尺度で質的データです。量的データではありませんので棒グラフとなります。

【このページのポイント】

横軸の都道府県は順序を入れ替えても問題ありません。上の棒グラフは、ミカンの生産量が多い都道府県で並んでいます。



【スライドの解説】

これまで学んだグラフ以外にも、データを表現する方法はたくさんあります。

■ヒートマップ

たとえば、データの値の大小を色の濃さで表現する「**ヒートマップ**」があります。

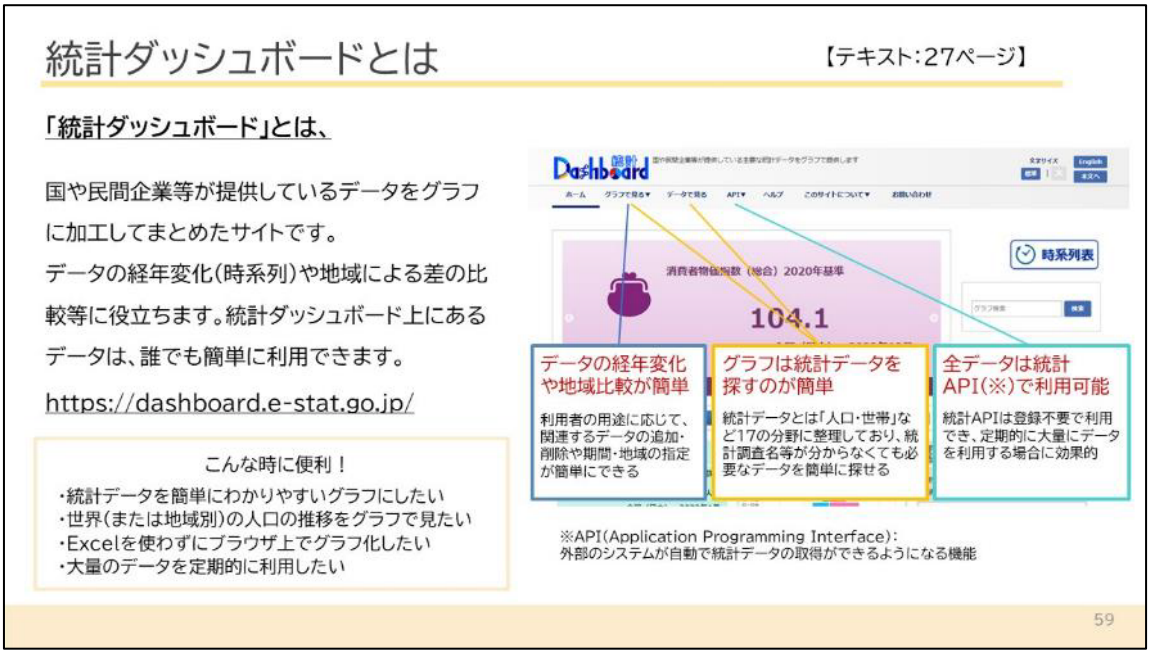
【このページのポイント】

「**ヒートマップ**」は値の大きさを色で表現します。

ヒートマップは地図と相性が良く、人口や人の動きを表すのにも使われています。

【キーワード】

ヒートマップ



【スライドの解説】

■統計ダッシュボード

「統計ダッシュボード」とは、国や民間企業等が提供しているデータをグラフに加工してまとめたサイトです。

データの経年変化（時系列）や地域による差の比較等に役立ちます。

【このページのポイント】

統計ダッシュボード上にあるデータは、誰でも簡単に利用できます。

【このページのポイント】

統計ダッシュボード：https://dashboard.e-stat.go.jp/



【スライドの解説】

■演習①

コロナ禍の入帰国者数を調べてみます。

【お題1】2020年4月～2022年9月までの日本への入(帰)国者数の推移を見てみましょう。

1. 「グラフ検索」の入力欄に「入国者数」と入力して検索をクリックします。
2. 「出入国者」と表記のあるデータをクリックします。2021年4月からの棒グラフが現れます。

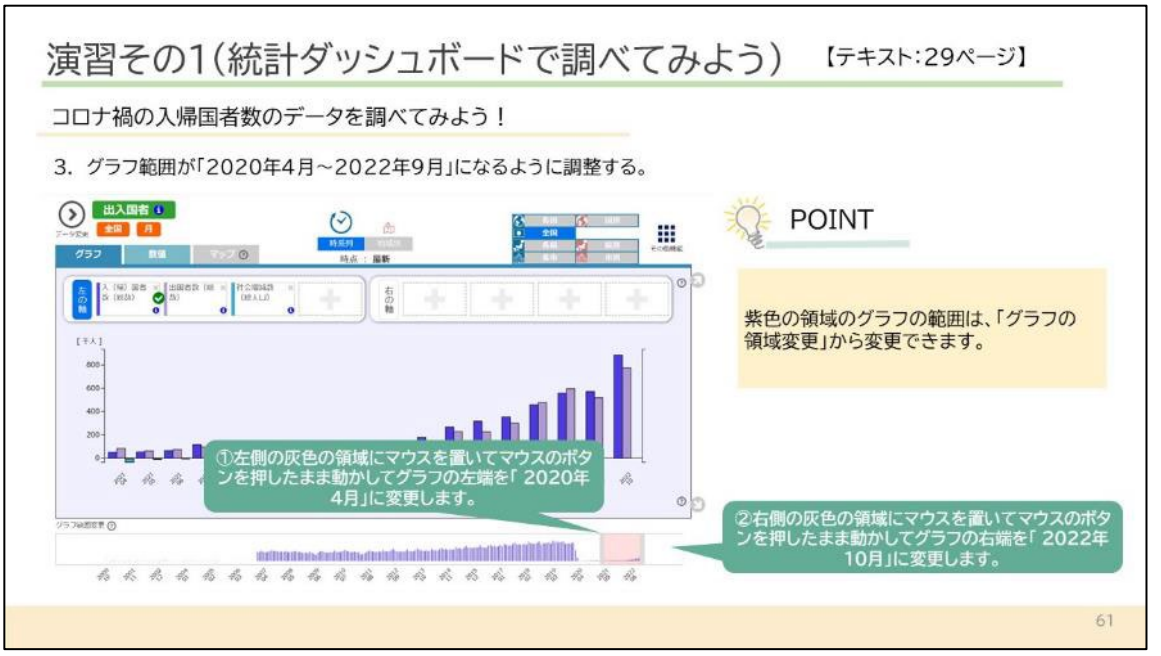
【このページのポイント】

実際に例題を使用しながら作成していきます。



【このページのポイント】

統計ダッシュボード、グラフ検索



【スライドの解説】

■演習①（続き）

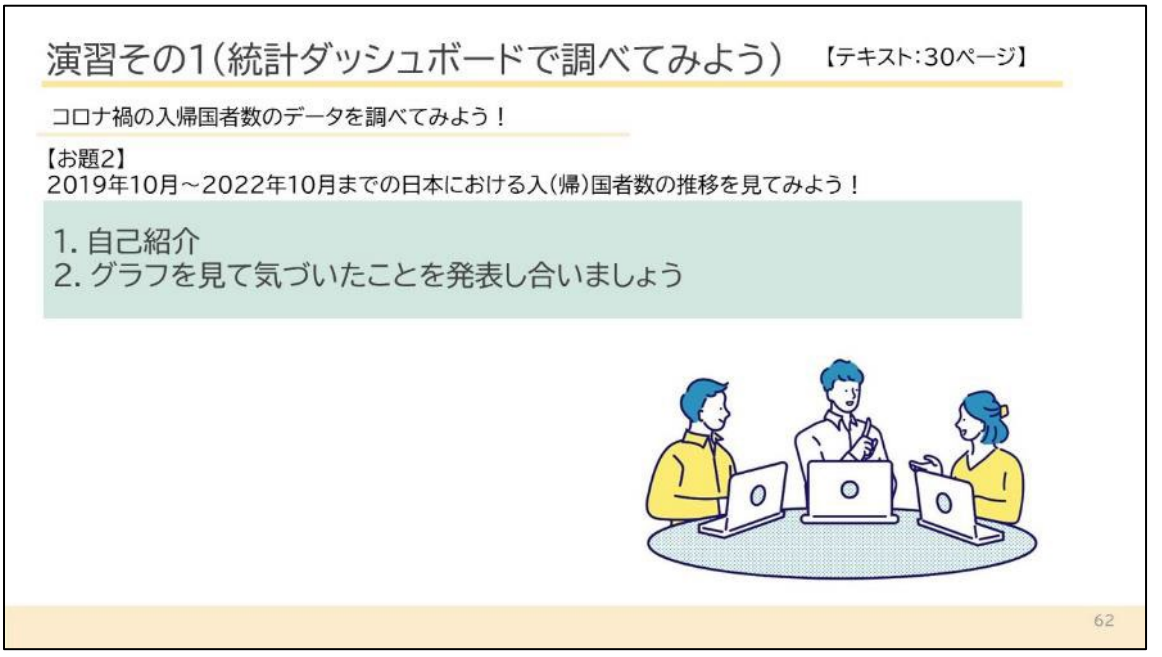
3.　グラフ範囲が「2020年4月～2022年9月」になるように調整します。

【このページのポイント】

グラフの範囲の変更は、「**グラフ範囲変更**」からマウスで操作できます。

【このページのポイント】

統計ダッシュボード、グラフの範囲の変更



【スライドの解説】

■演習①

【お題2】

2019年10月～2022年10月までの日本における入（帰）国者数の推移を見て、グループ内で気がついたことを話し合ってみましょう。

グループに分かれて自己紹介をしよう

こんなことを話してみましょう！

・お名前（ニックネームでOK）
・住んでいるところ
・ここまでのセミナーの感想など



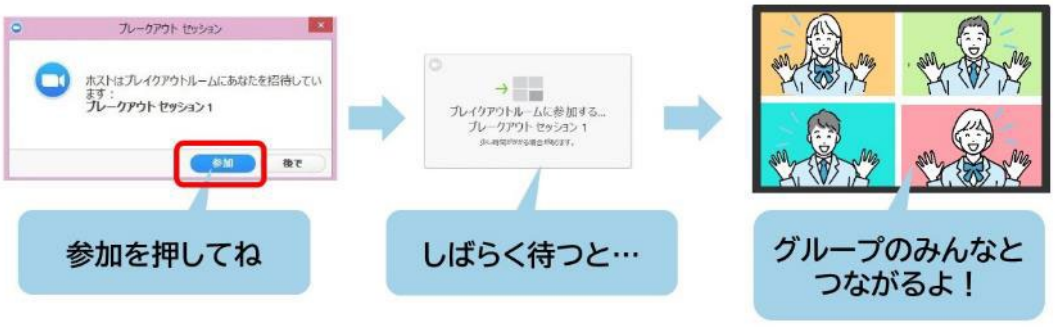
63

ブレイクアウトルームへの入り方



64

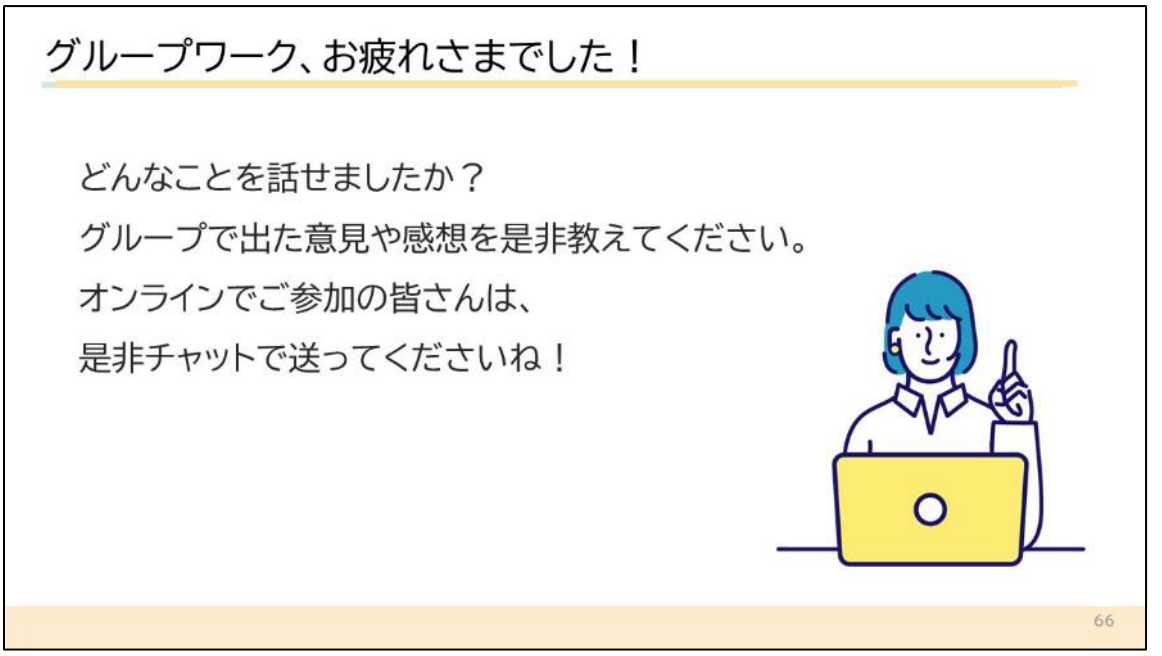
ブレイクアウトルームへの入り方



参加を押してね

しばらく待つと…

グループのみんなとつながるよ！



【スライドの解説】

2019年10月～2022年10月までの日本における入（帰）国者数の推移を見て、グループ内で話し合われた内容の一部を参考に載せておきますので、みなさんもぜひ考えてみてください。

■受講者からのご意見

・2019年10月～2022年10月までの日本における入（帰）国者数の推移グラフから、
日本への入（帰）国者数がコロナの影響を受けていることに気がつきました。

【このページのポイント】

日本における入（帰）国者数の推移のようなグラフのデータは、**時系列データ**と呼ばれます。時系列データを見たときは、コロナウイルス感染を防止するための水際対策(感染レベルが高い国からの入国を認めないなど)の政策や社会の動きを時系列にそってグラフに書き込んでみましょう。そこから気がつくことがあります。

【キーワード】

時系列データのグラフの見方



【スライドの解説】

■ダッシュボードの使い方

・「データ変更」からデータの周期や階級を変えることができます。

データ周期を「年」にしてみましょう。2019年まで入出国数が増えてきていましたが、2020年に大きく減少したことが見えます。2021年にはさらに減少しています。

・「数値」から選択している期間の数値データが見られます。

2019年から2021年度までの数値です。2021年の減少が大きいことが分かります。

時点	入(帰)国者数(総数)【人】	出国者数(総数)【人】	社会増減数【人】
2019	51,409,309	51,229,818	208,783
2020	8,045,450	7,910,567	41,907
2021	904,325	1,084,938	-35,188

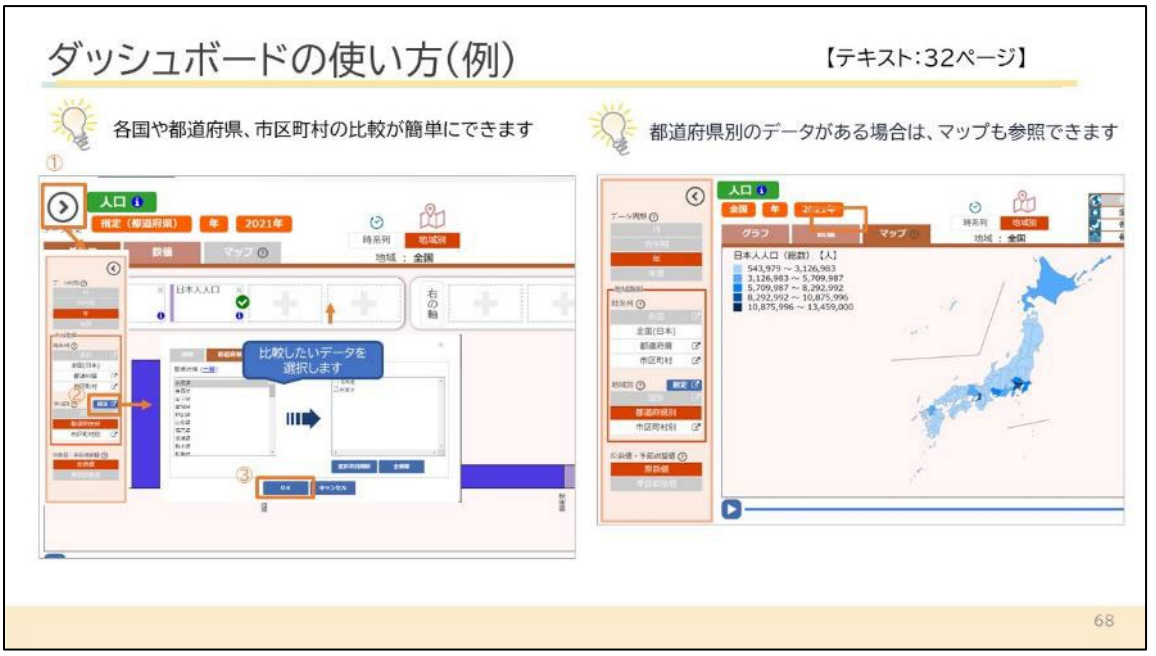
【このページのポイント】

「データ変更」からデータの周期や階級を変えることができます。

「数値」から選択している期間の数値データが見られます。

【キーワード】

ダッシュボードの使い方。



【スライドの解説】

■ダッシュボードの使い方の具体例

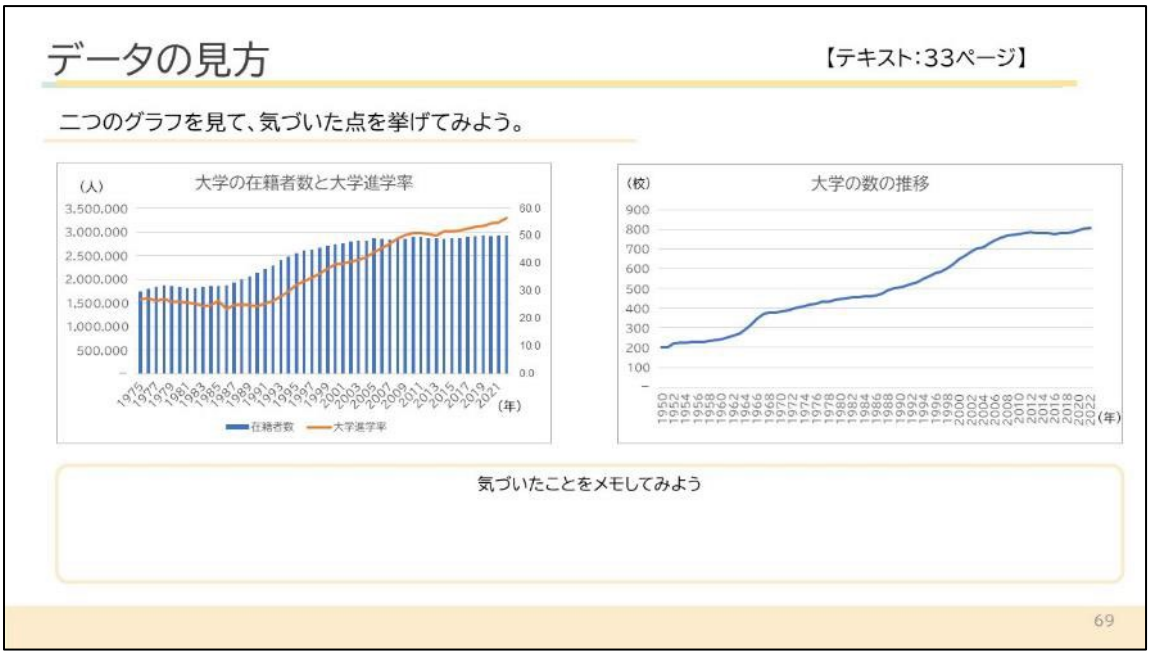
・各国や都道府県、市区町村の比較が簡単にできます

・都道府県別のデータがある場合は、マップも参照できます

【このページのポイント】

ダッシュボードは、各国や都道府県、市区町村の比較が簡単にできます。

年度を変えることで時系列の変化が見られます。



【スライドの解説】

■データの見方

データを見るときには、コツがあります。

大学の在籍者数を表した棒グラフと大学進学率を表した赤い色の折れ線グラフを同時に表わしたグラフと、大学の数の推移を表した折れ線グラフがあります。どちらも時系列データのグラフです。

大学の在籍者数の棒グラフは左側の縦軸が目盛(人)です。大学進学率の折れ線グラフは右側の縦軸が目盛(%)です。2021年の大学進学率は60%近いです。

この2つのグラフを見て、気づいたことを挙げましょう。

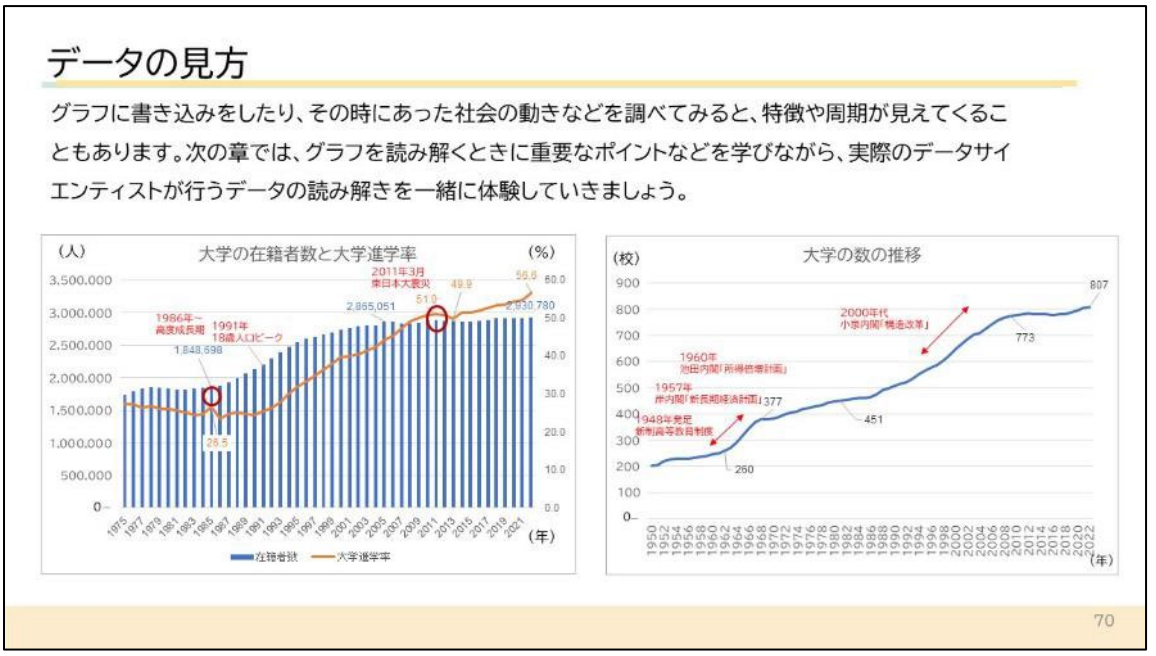
■受講者からのご意見

・大学進学率は上がっているのに、大学の在学者はそこまで増えていない。

・大学進学率と大学の数の増え方が似ている。

【このページのポイント】

二つのグラフを関連づけて、気がついたことを挙げてみましょう。



【スライドの解説】

■データの見方

社会の動きや震災などをグラフに書き加えていくと、データの特徴や周期が見えてくることもあります。

【このページのポイント】

■体験・データサイエンティストが時系列データをどのように読み解くか

・大学の在籍者数と大学進学率のグラフ（赤い字→その時の社会の動き）

①**高度成長期の始まり(1986年)**から大学の在籍者数も大学進学率も増加しています。大学の数のグラフを見ると、1962年から68年にかけて、大学の数の折れ線グラフの線の傾きが急になっています。

②**第二次世界大戦が1945年9月に終結**し、1947年から1949年に出産数が大きく増えました**(第1次ベビーブーム)**。その子たちが大学進学するのが1965年から1968年とすると、第1次ベビーブーム世代の進学先として、この時期に大学の数を100校以上増やしたと考えられます。

③**1991年に日本の18歳人口が最大**になっていますが、それ以降も大学の在籍者数は増え続け、2005年ごろから大きな変化がありません。しかし、大学進学率は上昇しています。**少子化の影響**という受講者からの意見がありましたが、その通りです。18歳人口が減少しているにもかかわらず大学在籍者にはあまり変化がないということは、18歳人口の大学進学率が増えていることになります。



次の章では、グラフを読み解くときに重要なポイントなどを学びながら、実際のデータサイエンティストが行うデータの読み解きを一緒に体験していきましょう。