

2023

— 困ったときのレポート作成講座 —

データの読み方編

総合情報基盤センター

沖野 浩二

今回はこんな方へ

〇〇授業期末レポート

参考文献

- ・なぜデータが必要なの？
- ・どんな風にデータを
探していけばいい？

「レポートって何？編」より

先行研究の活用で

研究課題の発見にも繋がる

今日取り上げる内容 / 目標

①データが必要の理由とは

②データの集め方

③例) 検索の流れ

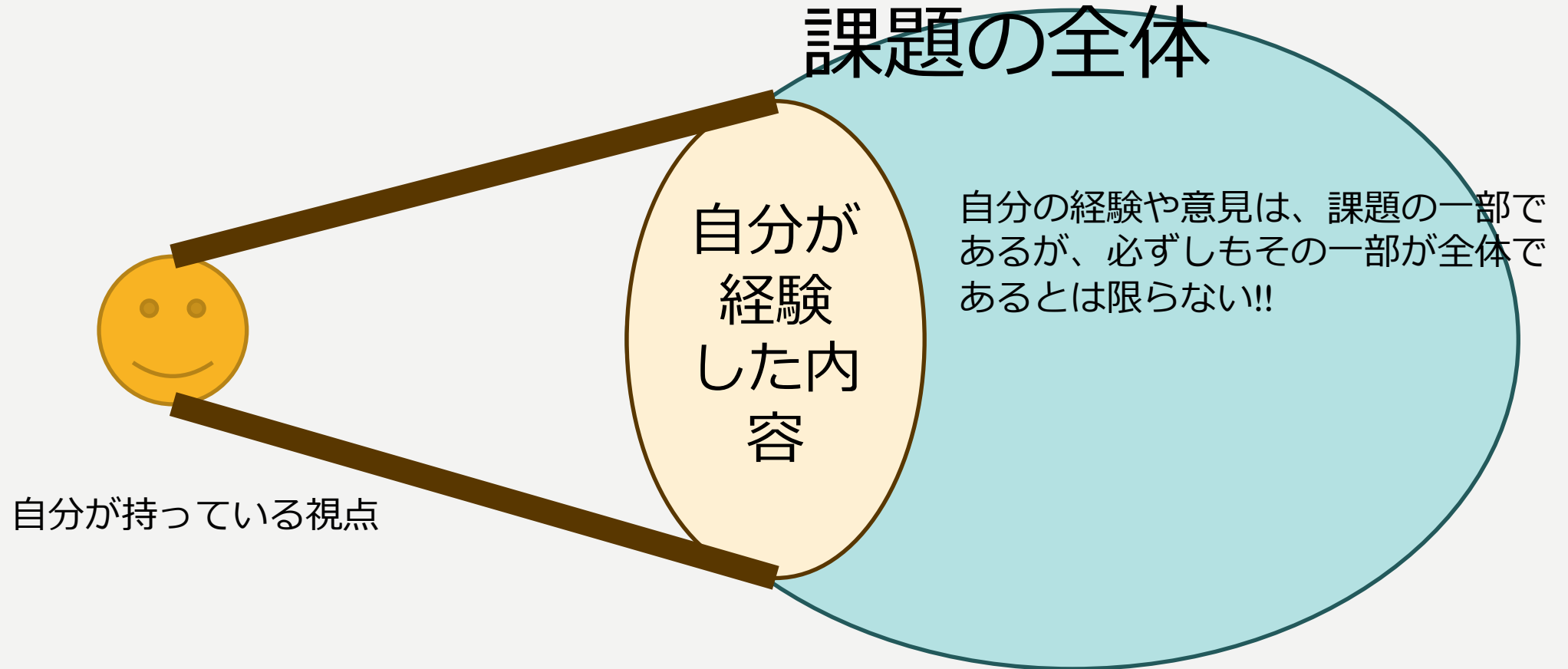
受講し終わっての目標

- データ取得の観点を得る
- データ検索の流れが分かる



① データが必要な理由

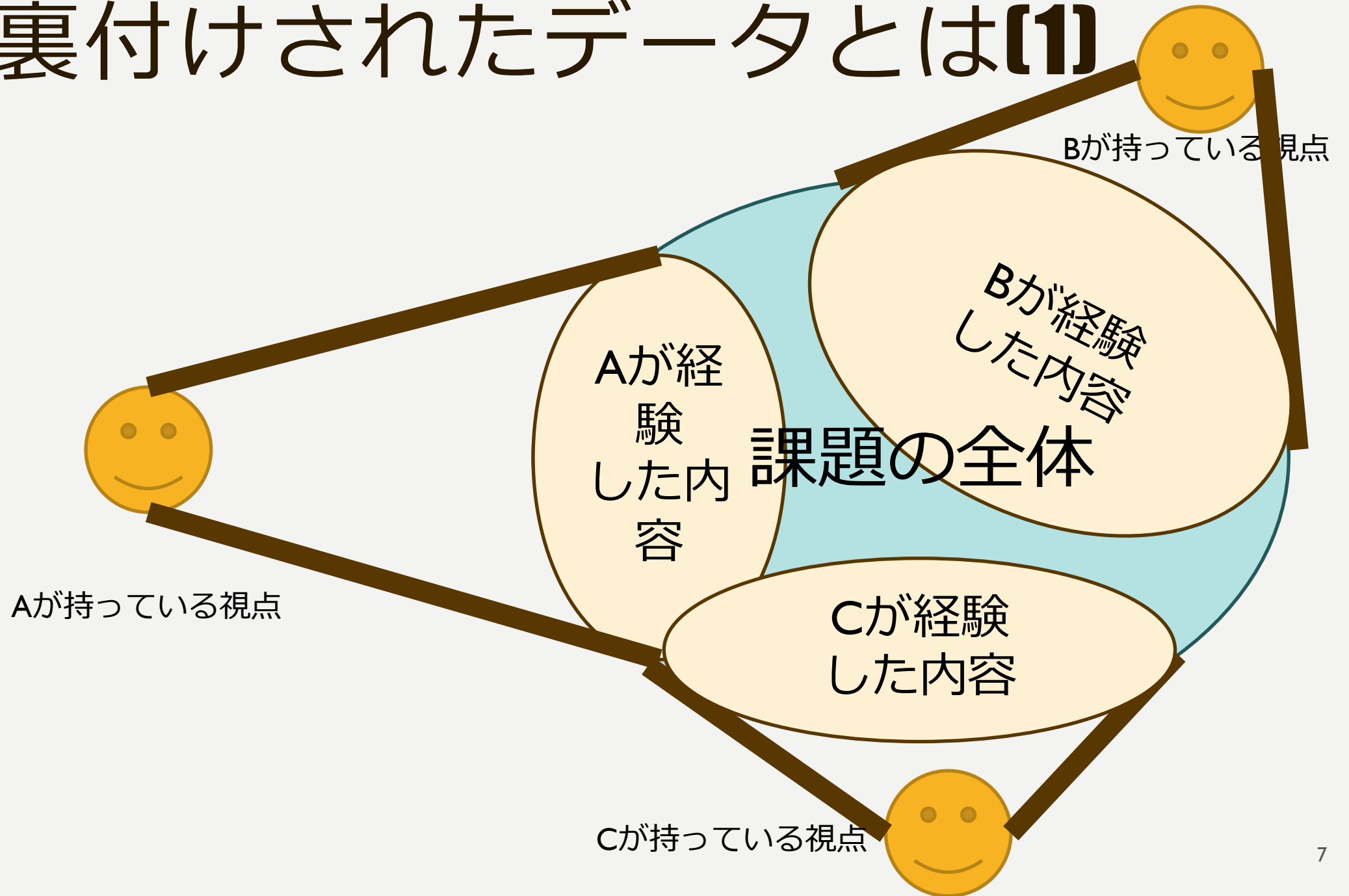
個人的な感想とは



個人の感想の問題点

- 個人の視点はもちろん重要
- エコーチェンバー
 - （ネットなどで）自分と似た興味・意見をもつ情報のみが帰ってくる状況
- 体験や情報が極端な例や見世物である可能性
 - ナラティブ
 - 語り手が紡ぐ「物語」「語り」などの意味で、語り手（見せ手）の意図によって物語が説明される
- レポートで求められていること
 - 個人の感想ではなく
 - 課題の全体を理解したうえでの意見が必要

裏付けされたデータとは【1】



裏付けされたデータとは**[2]**

例えば選挙

- 選挙に求めているもの
 - 自分が勝利してほしい候補や政党
 - 自分が行ってほしい政策
 - これは個人的な意見
- 課題の**全体**を示しているデータ
 - 投票先を決定する上で重要視する政策は？
- 課題の**変化**を示しているデータ
 - 有権者数の変化
 - 支持政党の変化
- 課題を取り巻く**環境**や**要素**を示しているデータ
 - 都道府県別の人口
 - 県別の議員数

求められている意見とは

- 個人的な意見ではない。（評価できない。）
- 裏付けのあるデータを元に考えられたものである必要がある。



②データの集め方

どのようにデータを集めるか

- 自分で収集
 - 実験
 - アンケート
 - etc
- 既存データを利用
 - 他の人などが集めたデータセット
 - 公的な公開されている資料（白書データや国連統計）

自分で収集

- 実験するときの注意点
 - 適切で安全な実験計画を立てる
 - 倫理的な問題がないかを検討する
 - 人に係る実験の場合には、倫理審査が必要である
- アンケートするときの注意点
 - 全数調査か標本調査か
 - (統計的に処理することができる) 最低限、必要な標本サイズ(*1)がある
 - 母集団からランダムに標本を選ぶ必要がある
 - 適切な設問設計となっているか (倫理的な問題を含む)
 - 実際には、倫理審査等が必要な場合もある
- レポートや初期調査では、既存のデータを利用することが重要!!

*1 松井博, "標本調査法入門", 財団法人日本統計協会, 2005

既存データの利用【1】

どのようなデータがあるかを知ろう

- データカタログサイト(デジタル庁)

<https://www.data.go.jp/>

- e-stat(総務省統計局)

<https://www.e-stat.go.jp/>

- 世界銀行OpenData

<https://data.worldbank.org/>

などなど (オープンデータ リストなどで検索)

既存データの利用【2】

他の人がどんなデータが利用しているか調べよう

- 本・新聞・雑誌などでの例
 - 特に新聞は有効!!
- Web検索
 - 政府系白書の解説
 - シンクタンク等の記事
 - ガベージニュース <http://www.garbagenews.net/>

切り口を探すには?

全体や詳細に関する情報収集時

- テーマを象徴しているデータは何があげられるか
 - 全体/変化/周辺/要素
- 資料中では、どのようなデータを用いて、何を議論していたか
 - これらは、図や表であらわれるだけでなく、文中にあることも

Point!!

- その問題を考えるうえで、どのようなデータがあれば、説明・考えることができるか。
- 実際に利用されているデータを探してみる

データの種類と可視化方法

他の人が作成しているものから読むのは、

- データの種類
 - どのように可視化（グラフ化）・分析している
 - いつからいつまでのデータを評価しているのか
-
- 基本
 - 割合：円グラフ
 - 変化：折れ線グラフ、棒グラフ
 - 見やすさ
 - 何を説明するために、データを可視化するのか
 - 沢山の項目をグラフ化しても見にくい

利用すべきデータの種類

- 公的な資料をベースに
 - 他人が内容を確認できる
 - データを集めた条件が示されている
 - 長年にわたるデータが収集されている
- (はじめは)有料のデータは基本的に利用しない
 - データの癖を知っておく必要がある
 - 対象や方法などから確認が必要

データの鮮度

- 古い資料・データを元に議論
 - 現在の状況とは異なっている可能性は
 - **古い議論**で行われていることを**最新のデータで再確認**することはレポートとして価値がある
- 短い間だけのデータを元に議論
 - そのデータが瞬間最高や最低の可能性は
 - ある期間だけ多く・少なくなっている可能性は
 - 話題のテーマの関するアンケートの特徴は
 - 直前のナラティブが影響



③ 例) 検索の流れ

実際のWEBからデータ探索 例) 若年層貧困

ガベージニュース 高卒男性は18.16万円...初任給の推移

<http://www.garbage news.net/archives/2564225.html>

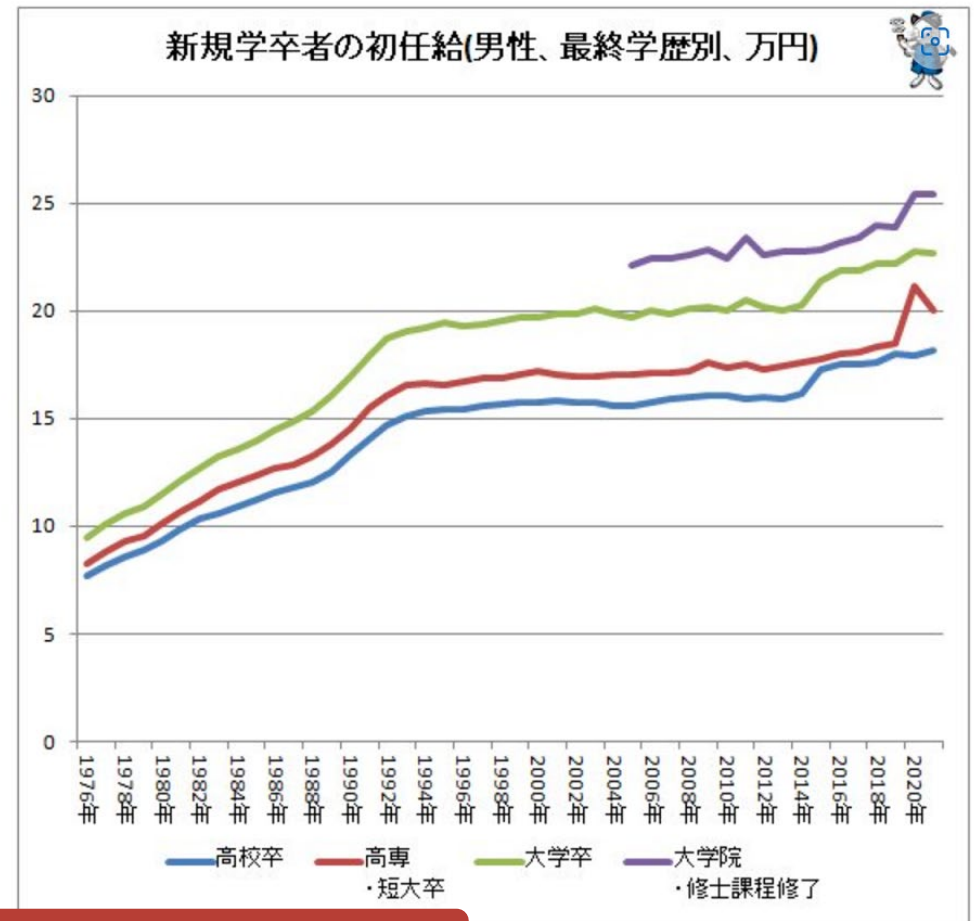
学歴・男女別の初任給動向

2022/5/30

今回検証する初任給とは、新規学卒者(各種類学校を卒業してそのまま就職
いわゆる就職浪人をした人を除く)がその企業で初めて取得した「賃金(所
額)」から通勤手当を除いた額。今回は企業規模の別は精査せず、すべて
案した平均値を用いる。また「所定内給与額」だが先行記事【フルタイム
は30万7400円・前年比でマイナス0.1%(最新)】で解説の通り、「基本
当などを足したもの」、つまり通常はほぼ固定して受け取れる額を意味す

ただし【令和2年賃金構造基本統計調査】の変更に伴う遡及集計について
2020年分調査から一部調査方法の変更が行われ、初任給については通勤
額となった(つまり他の労働者同様、所定内給与額そのもの)。2020年分
が跳ねているのはこれが原因である。

厚生労働省 賃金構造基本統計調査



新規学卒者の初任給(男性、最終学歴別、万円)

実際のデータの入手手順

どのような事象を述べることでその状況を見ることができるか。例えば若年層貧困の場合には、

- 初任給：進学がより良い給与へつながる
- 奨学金：進学への財政的な支援
- 家庭の資産：家庭環境と進学
- 学費：学ぶことへの負担

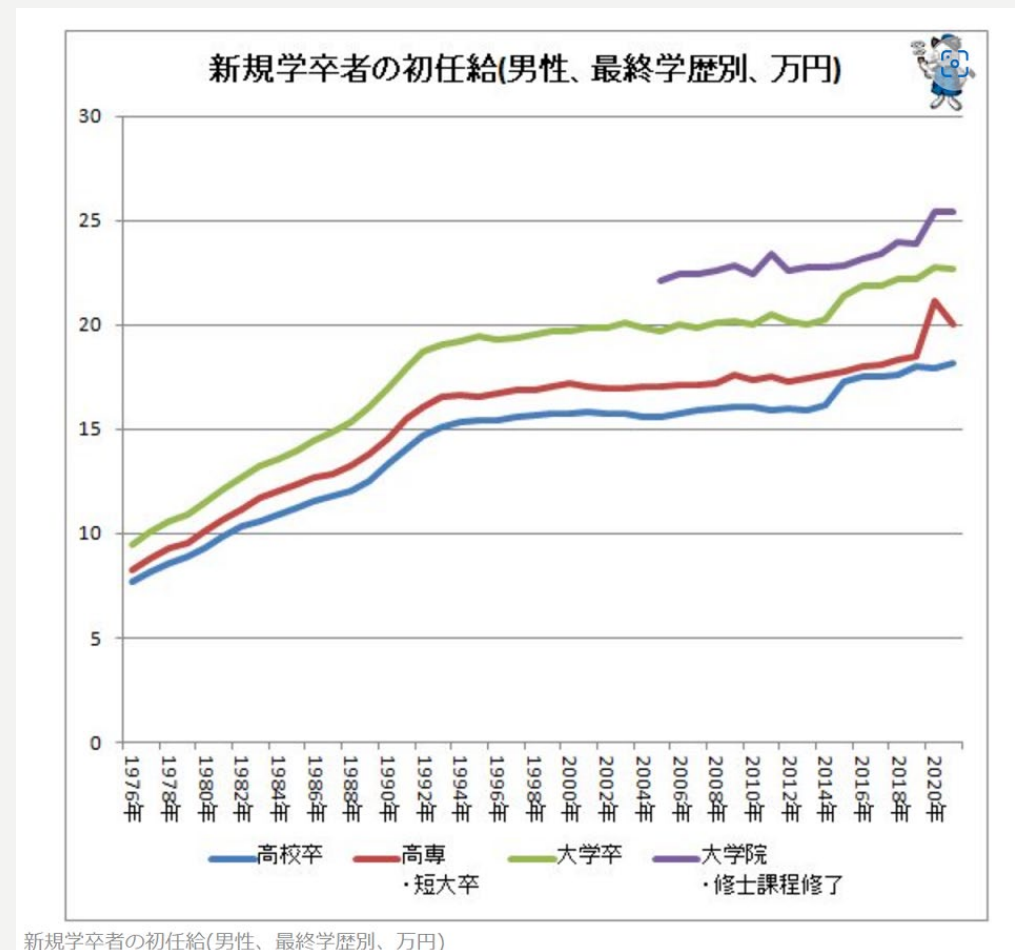
etc

難しいが、
ここを考えることが
レポート出題側の目的

実際のデータの入手手順

- 新卒者の初任給
本やネットのグラフをレポートで
利用することは（原則）NG
→孫引きになる

→自分で元データを手に入れて、
確認をして、グラフを作成して
初めて利用可能

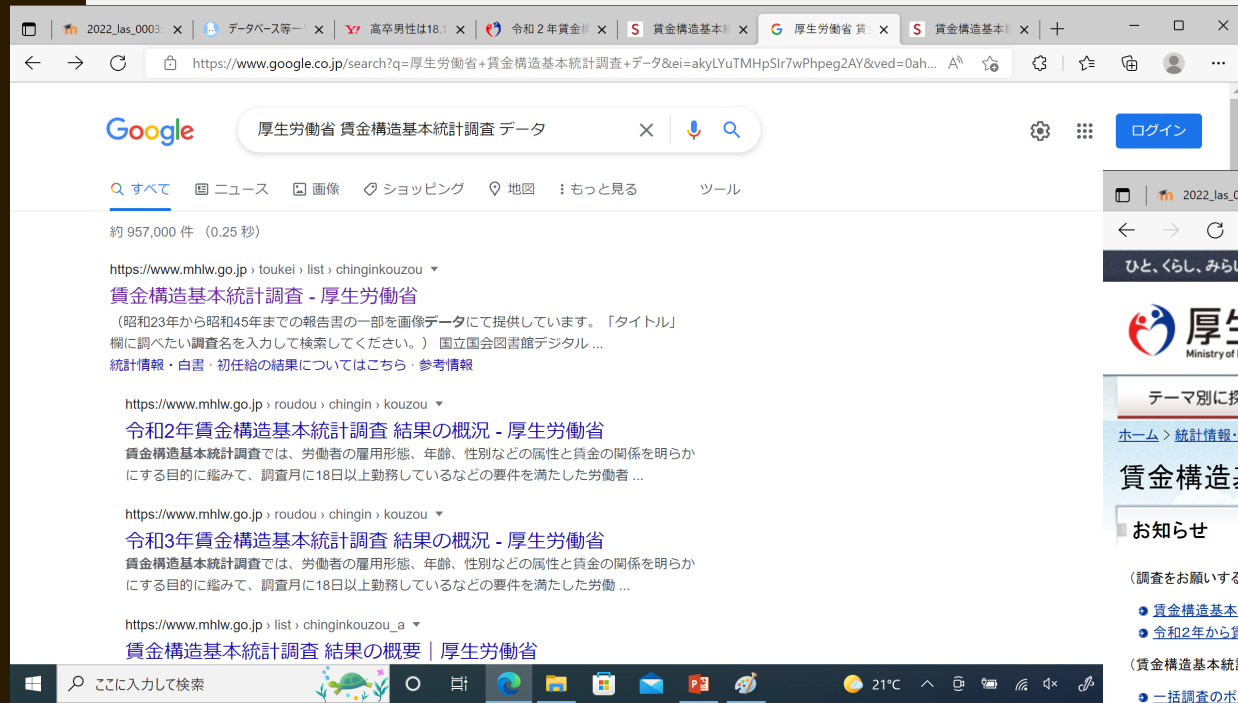


ガベージニュース 高卒男性は18.16万円...初任給の推移

<http://www.garbage news.net/archives/2564225.html>

2022/5/30

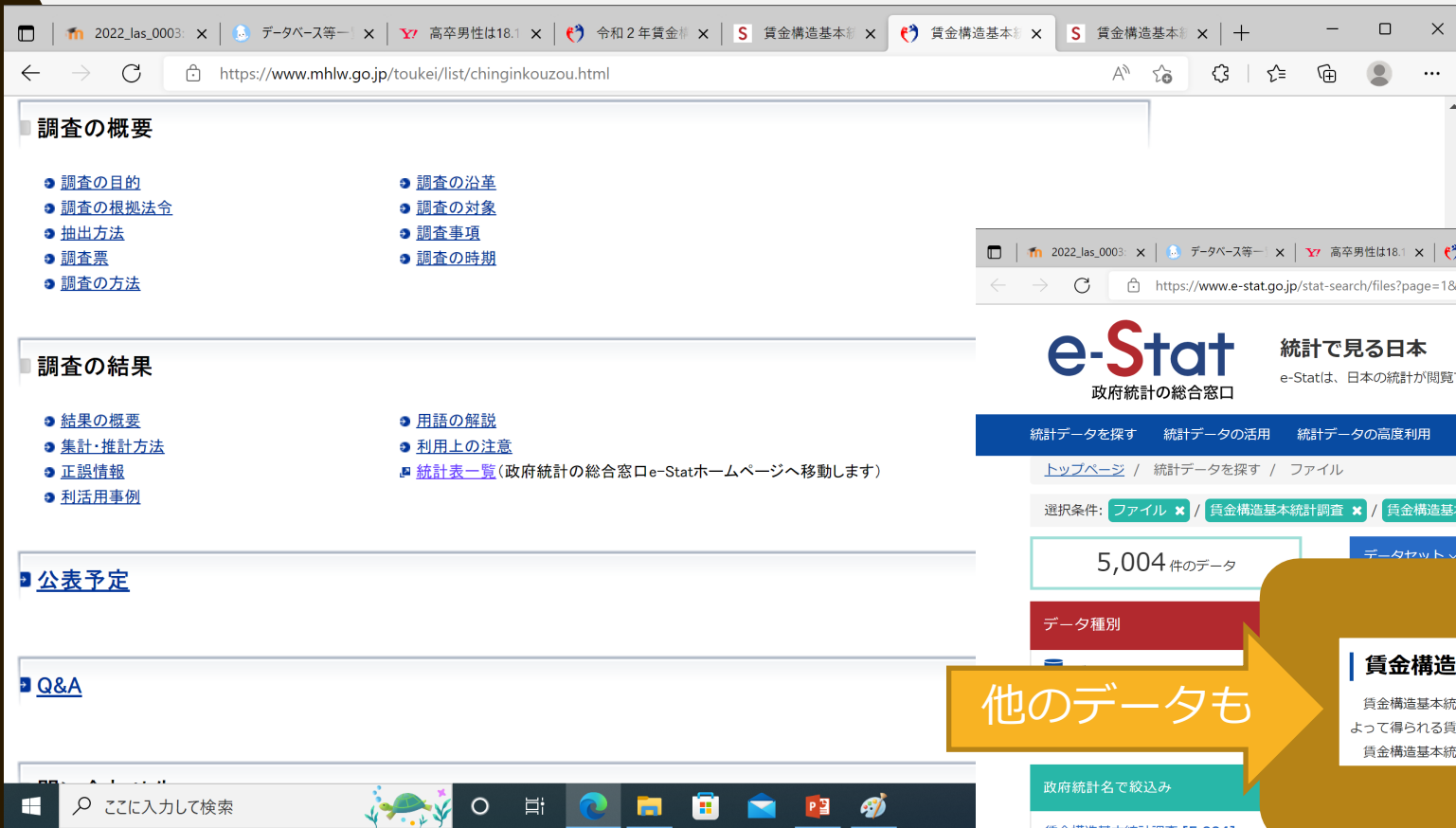
検索で実データを探す



Google 検索
厚生労働省 賃金構造基本統計調査 データ
結果表示

<https://www.google.co.jp/>

厚生労働省 賃金構造基本統計調査ページ
<https://www.mhlw.go.jp/toukei/list/chinginkouzou.html>



他のデータも



厚生労働省 賃金構造基本統計調査ページ
調査の概要 および 統計一覧

<https://www.mhlw.go.jp/toukei/list/chinginkouzou.html>

政府統計の総合窓口(e-Stat)
賃金構造基本統計調査

<https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00450091&tstat=000001011429>

データの範囲を変えてみると

- 新たなデータを追加
- 対象地域を変更（世界から日本など）
- 分類を変更してみる

データに何かを加えると

例えば、以下の変化と比較してみたら

- 経済成長率
- 賃金平均
- 内定率
- うどん、卵、マクドナルドなどの値段

今日の目標

受講し終わっての目標

- データ選択の観点を得る
- データ検索の流れが分かる

達成できましたか？