数理・データサイエンス・Al (応用基礎レベル)人文・社会科学系モデルシラバス 補足資料

2024年3月

特定分野校会議(人文・社会科学系)

動画教材の紹介

放送大学

「数理・データサイエンス・AI 応用基礎講座」

- <u>旧応用基礎モデルカリキュラム(R3.3版)準拠</u>
 - 3部構成 :データサイエンス基礎、データエンジニアリング基礎、AI基礎
 - それぞれ、9回、7回、9回の全25回で構成
 - 15分を1パートとし、3パートで1回分を構成
- 利用方法:放送を各大学・教員が録画して、対面授業の中で流すことについては 無料。オンライン・オンデマンド講義で利用するには、動画コンテンツ提供契約 (有料)やインターネット研修契約(有料)を結ぶ必要あり。

講座名	データサイエンス基礎								
モデルカリキュラム	1-1	1-6	1-7	1-2	1-3	1-4	1-5		
該当回	1	2,3,4	9	5	6	7	8		
講座名	データエンジニアリング基礎								
モデルカリキュラム	2-1	2-2	2-3	2-4	2-5	2-6	2-7		
該当回	1	2	3	4	5	6	7		
講座名	AI基礎							I.	
モデルカリキュラム	3-1	3-2	3-3	3-4	3-5	3-6	3-7	3-8	3-9
該当回	1	2	3	4	5	6	7	8	9

教科書の紹介

図解まるわかり『データサイエンスのしくみ』 (翔泳社)

- すべての項目を図解付きで見開きで紹介
- 各項目の概観を把握できる書籍
 https://www.shoeisha.co.jp/book/detail/9784798175805
- 1章 データサイエンスを支える技術 ~需要が高まる未来の必須科目~
- 2章 データの基本 ~ データの表現方法と読み方~
- 3章 データの処理と活用 ~データを分類し、予測する~
- 4章 知っておきたい統計学の知識 ~データから答えを推測する~

- 5章 知っておきたいAIの知識 ~よく使われる手法とそのしくみ~
- 6章 セキュリティとプライバシーの問題点 ~データ社会はどこに向かうのか?~



一生モノのビジネス共用『データサイエンス大全』 (ソシム)

- 現役データサイエンティストが執筆したテキスト
- 各部について重要用語編と実践スキル編に分かれ、多様な事例をベースに解説
- 各キーワードのイメージを直感的につかめる入門書 https://www.socym.co.jp/book/1402
- 一部 データサイエンスの基礎
- 二部 数理モデルと機械学習の基礎
- 三部 機械学習の応用
- 四部 ビジネス理解からデータの調理
- 五部 モデル構築から現場への導入



『文系のための統計学入門』データサイエンスの 基礎(日本評論社)

- 数式なしで統計学の概要を学べるテキスト
- 事例ベースで統計学の各手法の使い方が学べる
- 本書はデータサイエンス全般を扱っているわけではないが、<u>統計学の授業</u>の参考書として挙げておく https://www.nippyo.co.jp/shop/book/8597.html

- 2章 何でも平均値で大丈夫なのか
- 3章 確率的に生きるか確定的に生きるか
- 4章 学業成績の確率分布と偏差値
- 5章 該当アンケートはあてになるのか
- 6章 台風の予想円は信じてよいのか
- 7章 隠れた浮気を見破る方法
- 8章 薬品の含有率はきちんと守られて いるのか
- 9章 健康食品で血圧は下がるのか
- 10章 チョコレートを食べるとノーベル 賞が取れるのか

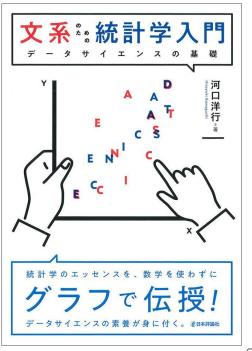
11章 広告費を増額すると売上高はどうな るか

12章 いろいろあるけれど、一番の原因は何なのか

13章 足したり掛けたりできない数字

14章 故障の有無を回帰分析する

15章 人工知能(AI)の母は統計学なのか



『応用基礎としてのデータサイエンス』 A I ×データ活用の実践 (講談社)

- ・旧「応用基礎レベルモデルカリキュラム」(R3.3版)に完全準拠した公式テキスト
- モデルカリキュラムとの対応:

http://www.mi.u-tokyo.ac.jp/consortium/e-learning_ouyoukiso.html

- 1章 データサイエンス基礎
- 1.1 データ駆動型社会とデータ分析の進め方
- 1.2 データの記述
- 1.3 データの可視化
- 1.4 データ分析の手法
- 1.5 数学基礎
- 2章 データエンジニアリング基礎
- 2.1 ビッグデータとデータエンジニアリング
- 2.2 データ表現、プログラミング基礎、アルゴリズム基礎
- 2.3 データ収集と加工、データベース
- 2.4 ITセキュリティ

- 3章 AI基礎
- 3.1 AIと社会
- 3.2 機械学習の基礎と予測手法
- 3.3 深層学習の基礎
- 3.4 ロボット、認識、言語



『データサイエンス応用基礎(データサイエンス大系)』(2024年5月出版予定)

- 数理・データサイエンス・AI 応用基礎レベル、 R6.2改訂版モデルカリキュラムの三分野(データサイエンス 基礎・データエンジニアリング基礎・AI基礎)の**すべての学修項目とほとんどのキーワード**を網羅
- 高度な内容は本文では扱わず出版社ウェブサイトに掲載のPDFで解説
- リテラシーレベルの『データサイエンス入門 第2版 (データサイエンス大系)』も既刊
- 応用基礎レベルのモデルカリキュラムのキーワードについて確認したい教員向け資料教材。

『データサイエンス応用基礎 (データサイエンス大系)』目次

1 データサイエンス基礎

- 1.1 データ駆動型社会
- 1.2 数学基礎
- 1.3 分析設計
- 1.4 データ観察
- 1.5 データ可視化
- 1.6 データ分析
- 1.7 アルゴリズム

2 データエンジニアリング基礎

- 2.1 ビッグデータとデータエンジニアリング
- 2.2 プログラミング基礎
- 2.3 データ表現
- 2.4 データ収集
- 2.5 データ加工
- 2.6 データベース
- 2.7 ITセキュリティ

3 AI基礎

- 3.1 AIの歴史と応用分野
- 3.2 AIと社会
- 3.3 機械学習の基礎と展望
- 3.4 予測·判断
- 3.5 深層学習の基礎と展望
- 3.6 生成AIの基礎と展望
- 3.7 認識
- 3.8 言語・知識
- 3.9 身体と運動
- 3.10 AIの構築・運用

PBL実践例の紹介

お茶の水女子大学のジェンダード・イノベーションに関するデータサイエンスセミナー (1/7)

- 内容:採用にひそむジェンダーギャップを分析できる2種類のアプリを開発し、 男女の採用基準のバランスが良いを思う採用基準を各自で設定し、その基準についてディスカッションを行った。
- データの詳細:採用選考過程における男女応募者の選考指標データ
- 分析できる内容:5種類のデータ(中学・高校・大学の成績、就職試験、MBAの成績)の中から2つのデータを選び、採用基準を設定する。その設定した基準に基づく決定境界の可視化とクロスバリデーションに基づく誤判別確率の計算を行う。
- セミナー開催のために気をつけたこと:
 - 専攻が異なる参加者が関心を持てる題材の選定
 - 学修レベルが異なる参加者に対応できる電子教材の検討

http://itolab.is.ocha.ac.jp/~itot/teaching/misc/IGI-DSseminar20240219.pdf

お茶の水女子大学のジェンダード・イノベーション に関するデータサイエンスセミナー (2/7)

ジェンダードイノベーションとは

性差や交差性に着目して研究や技術を開発することで イノベーションと発見を実現すること

O1 Problem Discovery 問題: 女性に副作用が多い?

製薬における ジェンダード・イノベーションの プロセスの一例

O2 Data Analysis 分析: 女性の副作用が本当に多い!

Decision Making 議論: 臨床試験の方法を変える?

04 Improvement 開発現場での改良

お茶の水女子大学のジェンダード・イノベーションに関するデータサイエンスセミナー(3/7)

セミナーの体験内容

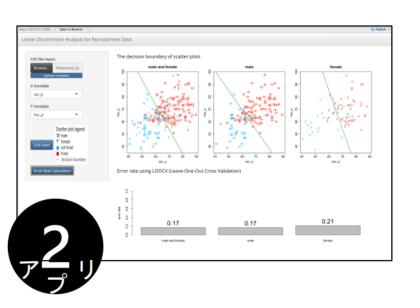
採用基準の探索

- 採用選考過程における男女 応募者の採用基準を調査
- 探索的なデータ分析の体験



採用可能性の判別

- 過去の採用結果をもとに男女 応募者の採用される可能性を判別
- 線形判別分析の体験



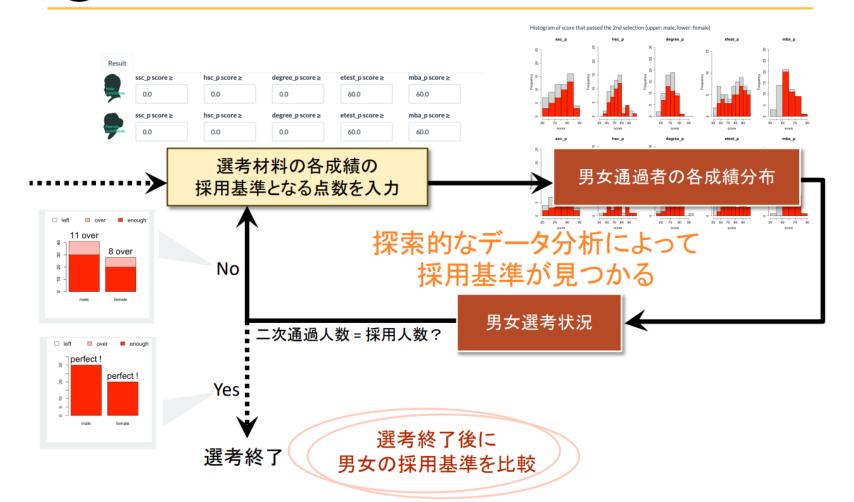
お茶の水女子大学のジェンダード・イノベーションに関するデータサイエンスセミナー(4/7)

1 実施できる採用選考のルール

- 選考回数:2回(一次選考と二次選考)
- 各選考材料:5種類の成績からユーザが指定
 - 5種類の成績:中学,高校,大学,就職試験,MBA
- 各選考指標
- 一次選考: 一次選考材料の成績の合計点
- 二次選考:二次選考材料の各成績の点数
- 各選考方法
 - 一次選考:2種類の順位(全体順位,男女別順位)からユーザが指定
- 二次選考:男女別順位(男女内訳:男性・女性が多い,男女同数の3種類)
- 各選考人数:ユーザが指定
- 各選考結果:通過者の情報

お茶の水女子大学のジェンダード・イノベーションに関するデータサイエンスセミナー (5/7)

11二次選考におけるデータ分析の操作工程

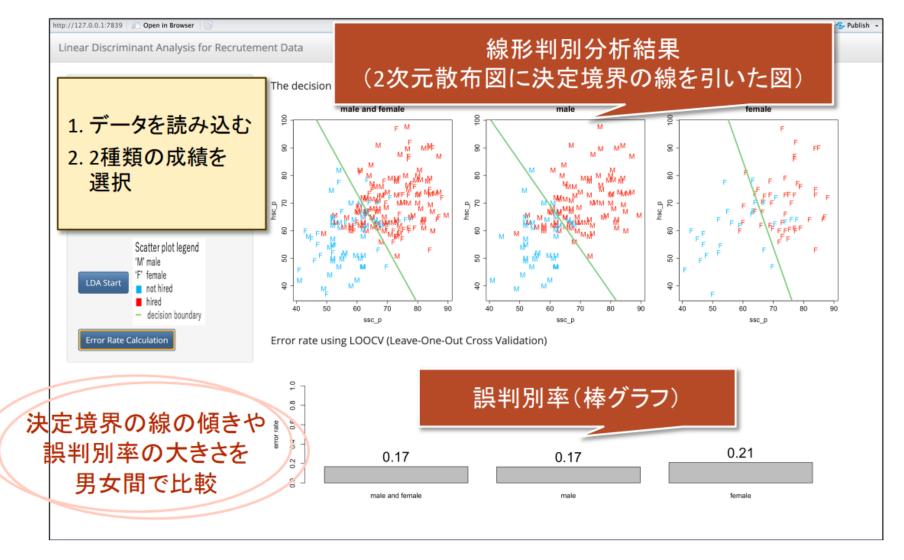


お茶の水女子大学のジェンダード・イノベーションに関するデータサイエンスセミナー(6/7)

2分析できる内容

- ・ 割り当てる2変数:5種類の成績からユーザが指定
 - 5種類の成績:中学,高校,大学,就職試験,MBA
- 計算できる内容
 - 過去の応募者の採用・不採用を判別する直線(決定境界)
 - Leave-One-Out Cross Validation (LOOCV)を用いた誤判別率
- 得られる結果
 - 応募者全員,男性応募者,女性応募者の決定境界
 - 応募者全員,男性応募者,女性応募者の誤判別率の値

お茶の水女子大学のジェンダード・イノベーションに関するデータサイエンスセミナー (7/7)



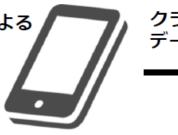
金沢大学の観光DXに関するPBL演習(1/7)

- ・観光DX:デジタル技術・ICTを活用して取得されたデータを分析し、既存の観光サービスの質・価値を向上したり、社会課題を解決することを目的とした取り組み。
- 内容:クルーズ船の観光客の行動履歴を収集し、人気の観光地 を見つけ、観光客が求めるデータの提供を行う。
- データの詳細:観光客に携帯してもらったスマートウォッチから収集される各種データおよび移動履歴データ
- 分析できる内容:観光客の移動履歴の可視化や、魅力的な観光 地の抽出などを行い、魅力的な街づくりに貢献する。

金沢大学の観光DXに関するPBL演習(2/7)

現在行っている取り組みースマートウォッチからデータ取得一





スマートフォン



クラウドor金沢大学サーバ

スマートウォッチ

❖ 位置情報

- ❖ 心拍数
- ❖ 呼吸数
- ❖ 血圧
- ❖ 加速度
- ❖ ストレス指標 etc.¦

隠れ観光スポットは?

待ち時間でイライラ しやすい場所は?

どんな交通手段で観光 している?



- ❖観光をしているだけで様々な データ群が一括取得できる 基盤があれば…
- →スマートウォッチ!
- ❖スマートウォッチから取得 可能なデータ群を一括で蓄積 可能なアプリの開発.

金沢大学の観光DXに関するPBL演習(3/7)

観光行動 Tourist Behavior

■ GPSロガー GPS logger



2014年7月30日 北國新聞朝刊



金沢大学の観光DXに関するPBL演習(4/7)

クルーズ観光客の「感動」見える化 Visualization of cruise tourists' impression

観光客の「感動」見える化 金沢大、心拍データで

<u>ネット・II</u> フォローする

2020年3月6日 2:00

保存

金沢大学は4月から、金沢市内で観光客の行動や生体データを取得して観光振興に生かす取り組みを始める。全地球測位システム(GPS)と心拍センサーが付いた端観光客の「感動」を可視化するほか、小型カメラで視覚情報を取得する。民間企連携し、データを活用したアプリの開発にも取り組む。年間で1000人の観光デー取得を目指す。



観光客に端末やカメラを装着してもらい、データを取得する(金沢市)



金沢大学の観光DXに関するPBL演習(5/7)

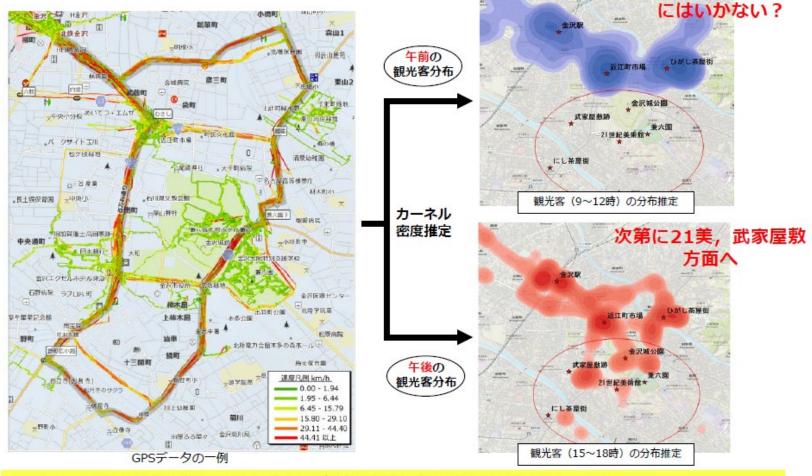
心拍データを用いた魅力的な観光地の発掘・評価



金沢大学の観光DXに関するPBL演習(6/7)

スマートウォッチから取得したデータの分析例 Behavior Analysis by GPS

ーGPSによる移動奇跡・時間ー



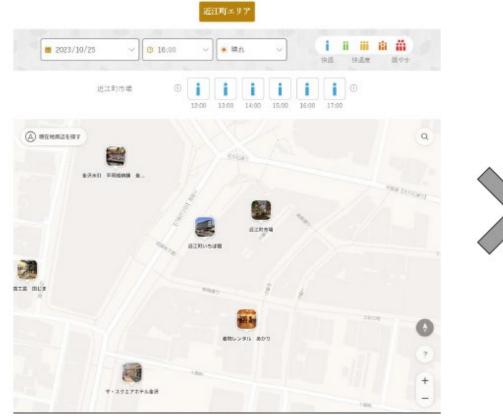
■ いつ, どこを観光している?が統計的・視覚的に把握可能

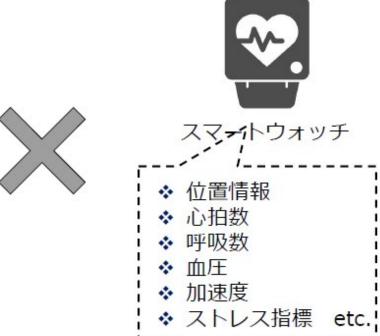
午前中は金沢城付近

金沢大学の観光DXに関するPBL演習(7/7)

観光DX推進を目指したデータ取得基盤の構築 Behavior Analysis using smart watch

- スマートウォッチを用いたリアルタイムデータ取得基盤
- どこが混んでいる?どこが空いている?を観光客に向けてリアルタイムに情報提供
- 快適な観光ルートのリアルタイム提案





琉球大学の企業インターンを活用したPBL (1/4) (データサイエンス実践演習I)

目的:企業の抱える課題を解決するための実習を通して、 データサイエンスの基礎を学んだ学生が取得した知識を活 用して取り組むことにより、実践的な力を身につける。

内容:2023年度は工学部、農学部、国際地域創造学部、大学院理工学研究科の計11名の学生が参加し農業従事者支援AIアプリの開発を行なった。

琉球大学の企業インターンを活用したPBL (2/4)

演習での取り組み

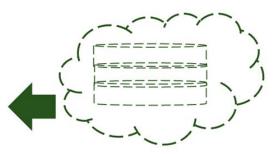
学部生毎に主たる作業をアサインせず、学部内で獲得して知識を活用するものの学部に依存せずに共通した課題にチームで取り組む

▶ 今既に世の中にあるものから有効データ・必要データを活用することを優先し、AIに栽培物の育て方としての有用情報を還元させる











Artzsリー機能活用とプロンプトエンジニアリ



インターンシップ参加学生の 主たるロール:各学部学生が 学部内で獲得する知識に依存 せず、枠を越えて、タッグを 組んだ活動でゴールを目指す



- 1. 回答還元UI開発
- 2. 世の中にある有用文献 (データ=今回はWEB上で公開されているサイト) の取り込み
- 3. プロンプト実装プロトタイプ
- 4. 還元情報(AIからの回答)の整形(利用者が理解しやすくするためプログラミング技術を活用して編集)



琉球大学の企業インターンを活用したPBL (3/4)

産学連携イノベーション・ビジネス体験インターンシップ・カリキュラム

3週間 (30時間程度) 8/21~9/8 (内2日程度終日) 【オンライン@沖縄】 e-Learning AI入門、プログラミング講座、モバイルアプリ検証環境構築

受講状況の確認と不明点などの質疑対応やアドバイスなどを企業の方が行うメンターリング・セッションを期間中4-5回設けます。e-Learningは、都合のつく時間帯に受講できます。



3日間 (終日) 9月:今後日程調整 【オフライン@沖縄】モバイルアプリサービスをアジャイル開発プロジェクトで実現することを体験

如何にAIを活用して高度化を実現するか、属人的ではない働き方の実現に近づけるかを、その苦労も含めて、体験し学ぶ機会を目指します。



1日

【オンライン@沖縄】Wrap Up 経験事例報告作成と発表 インターンシップ体験を経験事例としてまとめていただき、発表

琉球大学の企業インターンを活用したPBL(4/4)

産学連携イノベーション@沖縄 インターンシップ2023 達成度総括

本インターンシップで目指したこと

DX人財としてのコミュニケーションカ

目標・目的

体験値

プログラミング



コンピュータとのコミュニケーショ ンカ



コンピュータとの共創の仕方

プロンプトエンジニアリング



AIとのコミュニケーションカ



AIとの共創の仕方

チームワーク



人とのコミュニケーションカ



グループワーク



チームビルディング:夕食懇親会

ビジネスパーソン必須スキル

論理的思考



分析理解力・説明力・問題解決力



うまくいったケース、いかないケー スの理由を分析し説明



プログラム処理の問題を論理的に原 因究明し解決

お客様・生活者志向



お客様・生活者(利用者)の立場に 立つ



利用者視点で自身のモノづくりの成 果を評価

エンジニアの醍醐味

モノづくりの達成感



動くものを見える化した達成感



動くものを見えるもの、触れるもの としてAI連携UIを実現

和歌山大学のID-POSを用いたデータ分析 演習 (1/5)

- 内容:地元大手企業から提供を受けた本物のID-POSデータを用いて、 数名程度で編成するグループワークを実施し、企業が実際に抱える課 題の解決に実践的に取り組む
- データの詳細:匿名加工された、レジでの支払時等に記録される商品の購買データ
- 主な手法:基本的な統計やグラフ作成がメイン
- 特に力をいれたこと:難しい手法(機械学習)などの利用ではなく、 グループ内で協調して議論し、仮説を考え検証することを行っている
- 注意点:
 - Pandasなどの頻出のデータ抽出手法は授業内で練習している
 - プログラミングがメインにはならないようにしており、教員やTAがサポートしている
 - ID-POSデータのため、匿名加工をおこなっている
 - データが漏洩しないように、受講学生から誓約書をもらっている

和歌山大学のID-POSを用いたデータ分析 演習 (2/5)

• 講義概要

- PythonとExcel を使い、実データを用いたデータマイニングを中心とした講義と演習を実施
- 実際のPOSデータを題材として、データマイニングに取り組み、実践的なデータマイニングと可視化の手法の修得
- 業務把握、仮説立案、データ分析・検証、提案、プレゼンテーションなど、データサイエンティストの一連の仕事の流れを体験

プログラミングなどの技術的なサポートができるような体制をとっている(プログラミングができる学生に協力をもらっており、苦手なチームはサポートしている)

和歌山大学のID-POSを用いたデータ分析 演習 (3/5)

- 分析テーマとデータについて(2023年度)
 - スーパーマーケットから提示されたテーマ
 - 環境変化による購買変化
 - 温暖化等災害の増加
 - コロナ禍から現在までの購買変化
 - 値上げ
 - 次世代食生活への対応
 - 次世代ミート
 - 会員施策への提案
 - 会員カードを発行する企業は、常に会員数を増やしたいと考えている
- 提供データ(直近のデータ、複数年度、複数店舗のデータ)
 - 2017年~2022年 9月度~12月度×全国4店舗

テーマは、より現実の問題解決に取り組む ため、企業から出してもらっており、その 時の状況に応じて異なっている。

コロナ禍は、コロナ禍に関するテーマだっ たが、以前は食品ロス削減がテーマだった こともある。

> データは毎年、直近 のデータをもらって いる。

全て匿名加工された データを作成しても らっている。

和歌山大学のID-POSを用いたデータ分析

• 授業実施中の様子

ボードはかなり 効果的 学生には、データを漏洩しないように丁寧に説明しており、説明後に、誓約書もとっている。



企業の情報課の方の 説明中の様子

企業の方には期間中 常駐してもらい、質疑に 回答してもらっている。



議論中の様子

データは学内の特定の計 算機からしかアクセスで きないようにしている。



最終発表会の様子

最終発表会では、企業の方5~6名に 審査してもらい、順位をつけ、表彰 している。評価指標は、企業と相談 し、大学で作成している。

和歌山大学のID-POSを用いたデータ分析 演習 (5/5)

• 参加証明書と講義担当者、授業担当者との記念撮影



通常、PBL(授業)にでても参加証明書などはもらえないが、イベント的に実施することで、かなり満足度が高い。

ペットボトル緑茶に関する会場調査データ (滋賀大学調べ) 担当者になりきって分析してみよう!

- 調査背景・目的(仮想例だが、これに基づき実際に調査・分析した)
 - ある飲料メーカーは、長年、ペットボトル入りの緑茶を販売しており、今後さらに売上を拡大するために、緑茶の味を生活者の嗜好に合わせる必要があると考えている。しかし、生活者は味の違いだけでなくブランドに対する印象などを含めて考慮し、購買するブランドを選択していると考えられる。
 そこで、各社の緑茶の味覚評価を得ることで、生活者の嗜好にあった味を提供できているかを確認する。また、ブランド名を提示しない時と提示した時の評価を比較し、各ブランドのブランド力を確認する。
- 調査対象者は以下の条件にすべて合致し、会場調査に参加できる方
 - 20歳~59歳の男女
 - 1都3県に在住
 - 市販の緑茶飲料 (250~700ml 程度のペットボトル) の自購入・自飲用者 (週1本以上)
 - 「お~いお茶」「綾鷹」「伊右衛門」「生茶」をすべて知っている人
- 調査手法
 - 会場調査 ※ 調査対象者はインターネット調査による事前調査を用いて呼集
- 調査時期
 - 2020年2月29日(土) ~ 3月2日(月)
- 有効回答数
 - 258人
- サンプル構成
 - 性別・年代・最も好きな緑茶ブランドで均等割付
- データや分析例など
 - https://www.ds.shiga-u.ac.jp/education/sampledata/



SSDSE (教育用標準データベース) を 用いたデータ分析

- SSDSEは都道府県単位または市区町村単位の公的データが分析しやすい形に集計されたデータである。
- Excel形式とcsv形式の両方があり、様々な分析環境で扱いやすい形で提供されている。
- 身近な地域のデータについて、簡単な集計、可視化(棒グラフや統計地図など) を行うことでデータの特徴をつかむ練習に使える。
- 地方ごとに再集計して地方の特徴を比較するためにデータの加工を行ったり、少し高度な多変量解析を行ったりなど、受講者のレベルに応じた使い方が可能。
- 過去の統計データ分析コンペティションの分析事例を含む、様々な活用例も紹介 されている。

https://www.nstac.go.jp/use/literacy/ssdse/utility/?doing_wp_cron=1705336 767.0556979179382324218750

e-Stat (政府統計の総合窓口)、RESASを用いたデータ分析

- e-Statは公的統計を公開しているサイトで、Excelファイルやcsvファイルでデータを提供しているだけでなく、データフレーム形式で公開されているデータについてはサイト上で様々なグラフを表示させることもできる。
- e-Statで提供されているデータの一部はRESASにも組み込まれている。RESAS の方が直感的なグラフ操作が可能である。
- e-Stat、RESASともにサイト上でグラフを表示できるので、自分で集計、可視化を行わなくても、様々な地域の特徴を捉えることができる。
- e-Stat、RESASを用いたデータ分析、政策提案の事例は「地方創生☆政策アイデアコンテスト」のサイトで確認できる。ここで紹介されている内容のうち、序盤のデータ集計、可視化についてはデータの見方、比較の観点として参考になると思われる。

https://contest.resas-portal.go.jp/2023/index.html

国土数値情報ダウンロードサイト(都道府県 地価調査データ)の活用事例

都道府県地価調査に基づくポイントデータの地図表示 https://github.com/TakaakiAokiWork/DSPracticeExampleForHumanities/tree/main/LandPrices-PointsOnMap

都道府県地価調査データを取得する

「国土数値情報ダウンロードサービス」 by 国土交通省を使って,「国土数値情報 都道府県地価調査(ポイント)」を取得する

ここでは、滋賀県のデータ(2023年、世界測地系) L02-23_25_GML.zipをダウンロードし、ファイルを解凍する。

都道府県地価調査データ(geojson file)を読み込む

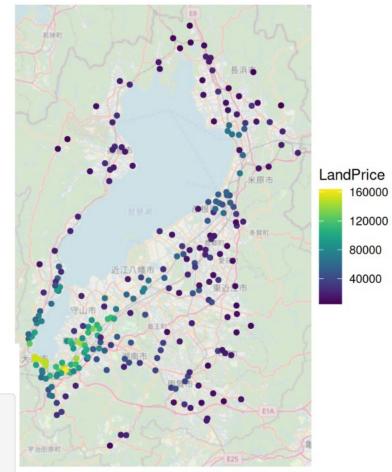
解凍したファイルのうち、GeoJsonフォーマットのデータを読み込む

https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L02-v2 3.html

```
data = st_read("L02-23_25.geojson")
```

地図上にplotしてみる

```
ggplot(data) +
  annotation_map_tile( zoomin = 0, alpha = 0.5) + labs(caption = "\frac{\text{YU00a9 OpenStreetMap contributors"}}{\text{background}} + # use OpenStreetMap data as
  background
  geom_sf(aes(color = LandPrice)) + scale_color_viridis() + theme_void()
```



© OpenStreetMap contributors

コロラド大学ネットワークデータ(文学作品 レ・ミゼラブルの人物相関図)の分析事例

Rを使ったレ・ミゼラブルの人物相関図の可視化とグループ分け

https://github.com/TakaakiAokiWork/DSPracticeExampleForHumanities/tree/main/LesMiserables-SocialNetworkAnalysys

Read network data file

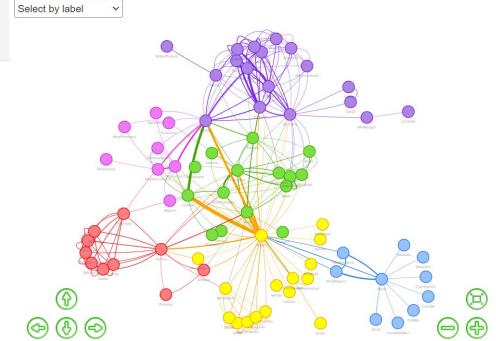
1 Download Les Miserable network data file (GML format) from https://github.com/gephi/gephi/wiki/Datasets. 2 Extract the downloaded, zipped file.

```
g = read.graph("lesmiserables.gml", format="gml")
# show summary of the graph
summary(g)
```

Community detection (グループ分け)

Modulariy (Louvain method)

```
comm = cluster_louvain(g)
data = toVisNetworkData(g, idToLabel = FALSE)
data$nodes$group = membership(comm) # Community detectionの結果で色分け
visNetwork(nodes = data$nodes, edges = data$edges) |>
visOptions(selectedBy = "label") |> # 特定人物名に注目
visInteraction(hideEdgesOnDrag=TRUE, dragNodes=TRUE, zoomView=TRUE, navigationButtons=TRUE)
```



参照データベース:コロラド大学ネットワークデータ集 https://nlftp.mlit.go.jp/