

ビジネスに役立つデータ分析 （入門編）

瓜生真也（徳島大学デザイン型AI教育研究センター）

自己紹介

うりゅう しんや

瓜生真也

職歴 徳島大学 デザイン型AI教育研究センター 助教 (2021年10月～)

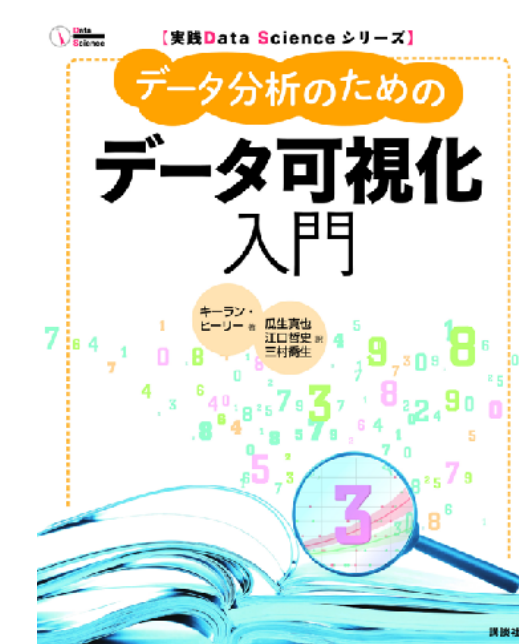
国立研究開発法人 国立環境研究所 (2017年11月~2021年9月)

株式会社ナイトレイ (2016年4月~2017年10月)

専門 データサイエンス、地理空間解析、データ可視化

プログラミング: R、Python

生態学 (植物、森林)



Rユーザのための
tidymodels
実践入門

2023年1月
技術評論社より
出版予定

講座の内容

第一週

情報化社会におけるデータの利活用
ビジネスにおけるデータ分析の重要性
データ分析に取りかかる前に…
Rプログラミング入門（速習）

第四週

データサイエンス入門 (2)
主成分分析
クラスタリング
時系列データ

第二週

データを扱うためのリテラシー
データの要約
グラフの作り方、読み方

第五週（最終週）

ツールの使い方
振り返り

第三週

データサイエンス入門 (1)
回帰
分類

講座番号

R104

ビジネスに役立つデータ分析
(入門編)

定員
10名

日時

木曜日
10:30~12:00

11/1011/1711/2412/1
12/8

全5回

受講料

6,000 円

講師

瓜生真也

講座の概要

身近なデータを題材に、データの見方を学びます。データを要約し、グラフで表現することで、意外な特徴を発見し、意思決定につなぐことができるようになります。また必修化された高校教科情報Ⅰの内容を学ぶこともできます。

各自で、テキストの購入
必須
別紙☆テキスト一覧表参照

この講座の意義

情報化社会におけるビジネスでのデータ分析の手法を身につける

目的

ビジネスで利用できるデータ分析の手法を学ぶ

プログラミング言語Rを通してデータ分析の実践を行う

期待されるもの

必修科目となった高等学校情報科「情報I」の内容のフォローアップ

ビジネスで活用可能なデータ分析の手法や可視化表現

R言語への理解、技能

諸注意

資料置き場: https://github.com/uribo/cue2022aw_r104

投影するプレゼンテーション、ソースコードを置いています
(来週分は来週更新)

実行環境はこちらで用意したものを利用します

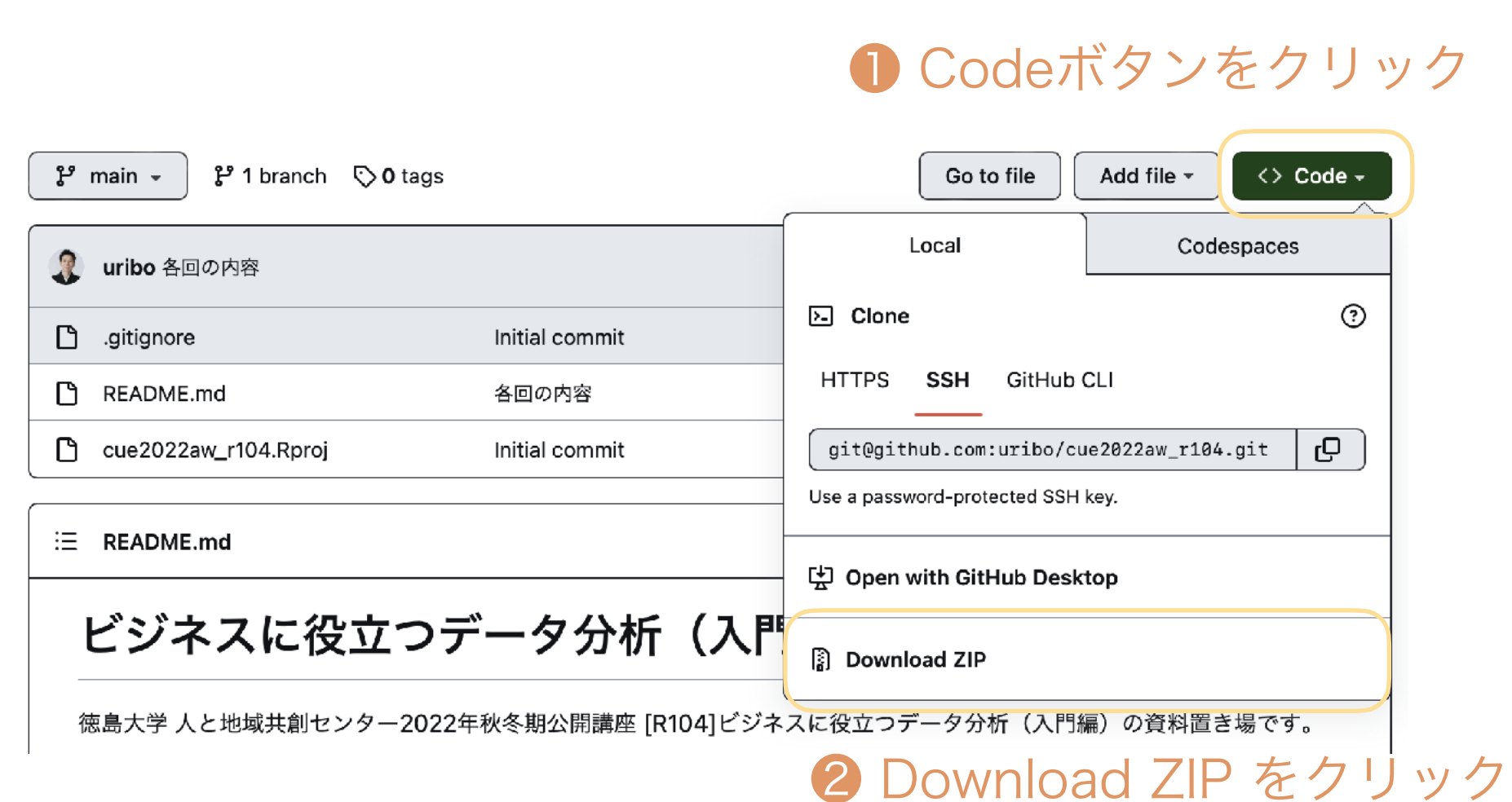
手元の環境で試されても問題ありませんが、
サポートしきれない場合があります

質問等

いただいた質問等には極力返答できるように努めます

各回または最終回終了後、メール (to: uryu.shinya@tokushima-u.ac.jp) にて問い合わせください

講座に対するご意見も歓迎です



第一週: 情報化社会におけるデータの利活用

瓜生真也（徳島大学デザイン型AI教育研究センター）

講座の内容

第一週

情報化社会におけるデータの利活用

ビジネスにおけるデータ分析の重要性

データ分析に取りかかる前に…

Rプログラミング入門（速習）

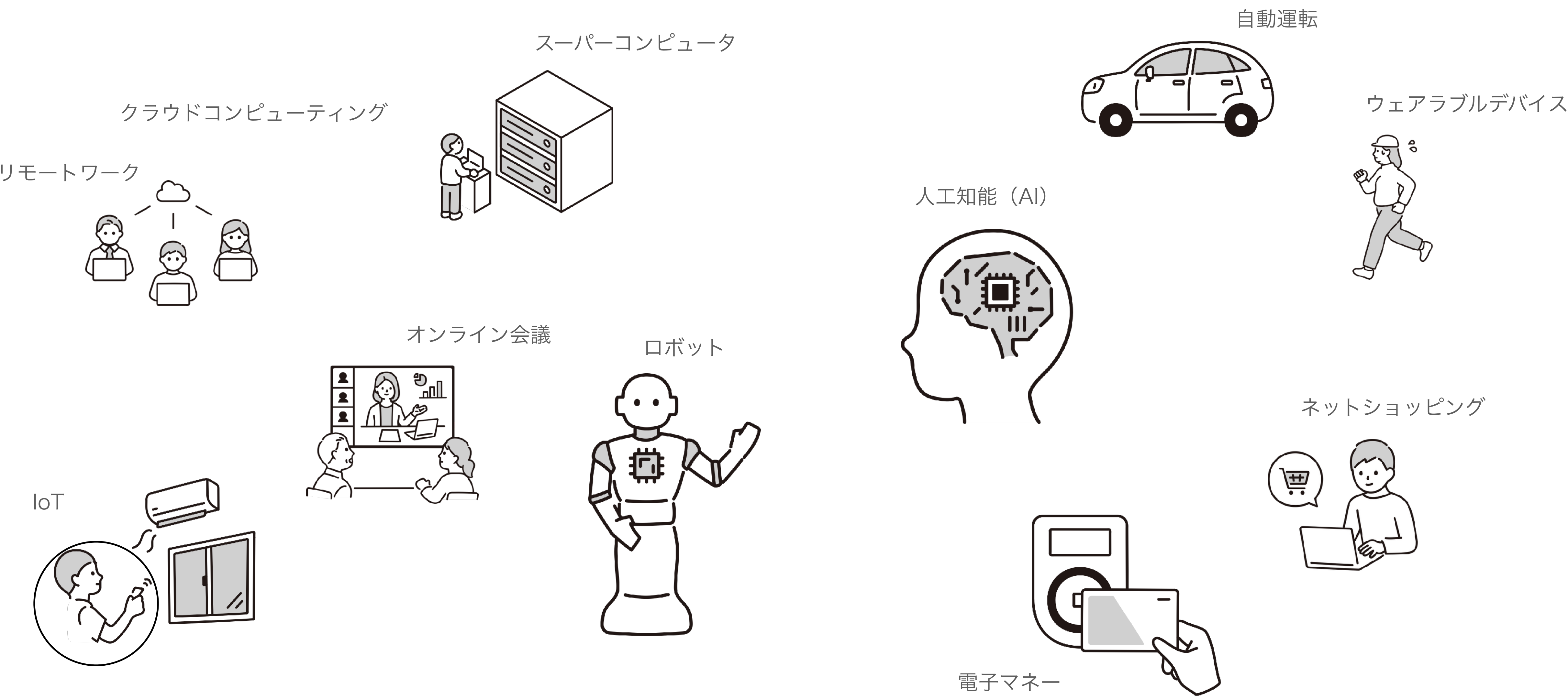
第二週

第三週

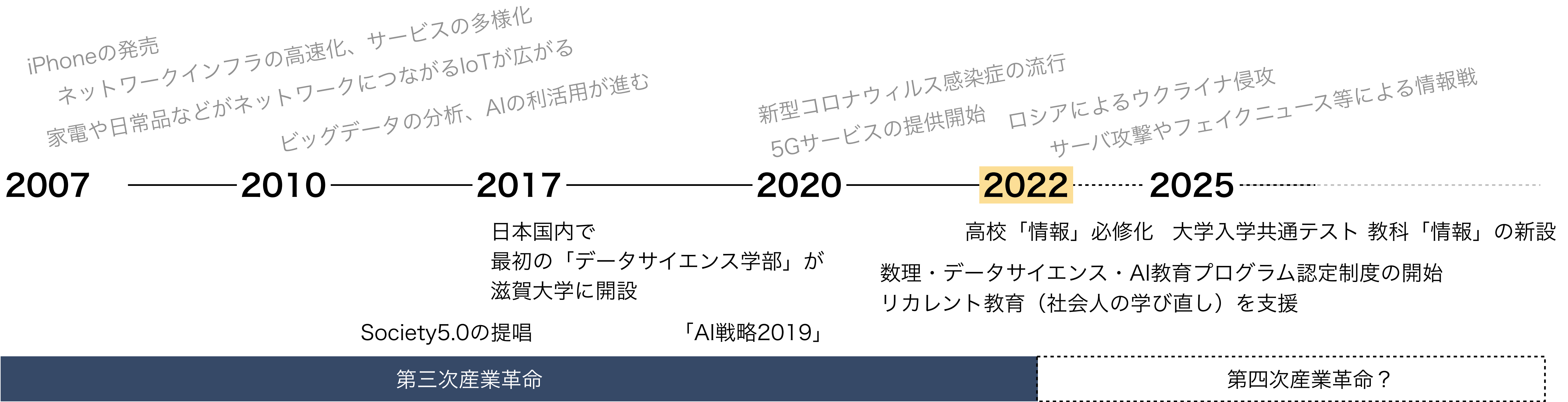
第四週

第五週

情報化社会を支えるサービス、技術



情報化社会（現代）とこれから（未来）



ビッグデータの登場

特徴

英語の頭文字を取って5Vと呼ばれる※

容量(Volume)

扱うデータそのものの数

テラバイト、ペタバイト、エクサバイト（ 10^{18} バイト）

速度(Velocity)

入出力のためのデータ通信の速度、リアルタイムデータの反映

ICTの高度化 第5世代移動通信システム(5G)

多様性(Variety)

データの種類の増加。音声、動画、プローブ、センサー、決済 etc.

構造化データと非構造化データ オルタナティブデータ（金融）

正確性(Veracity)

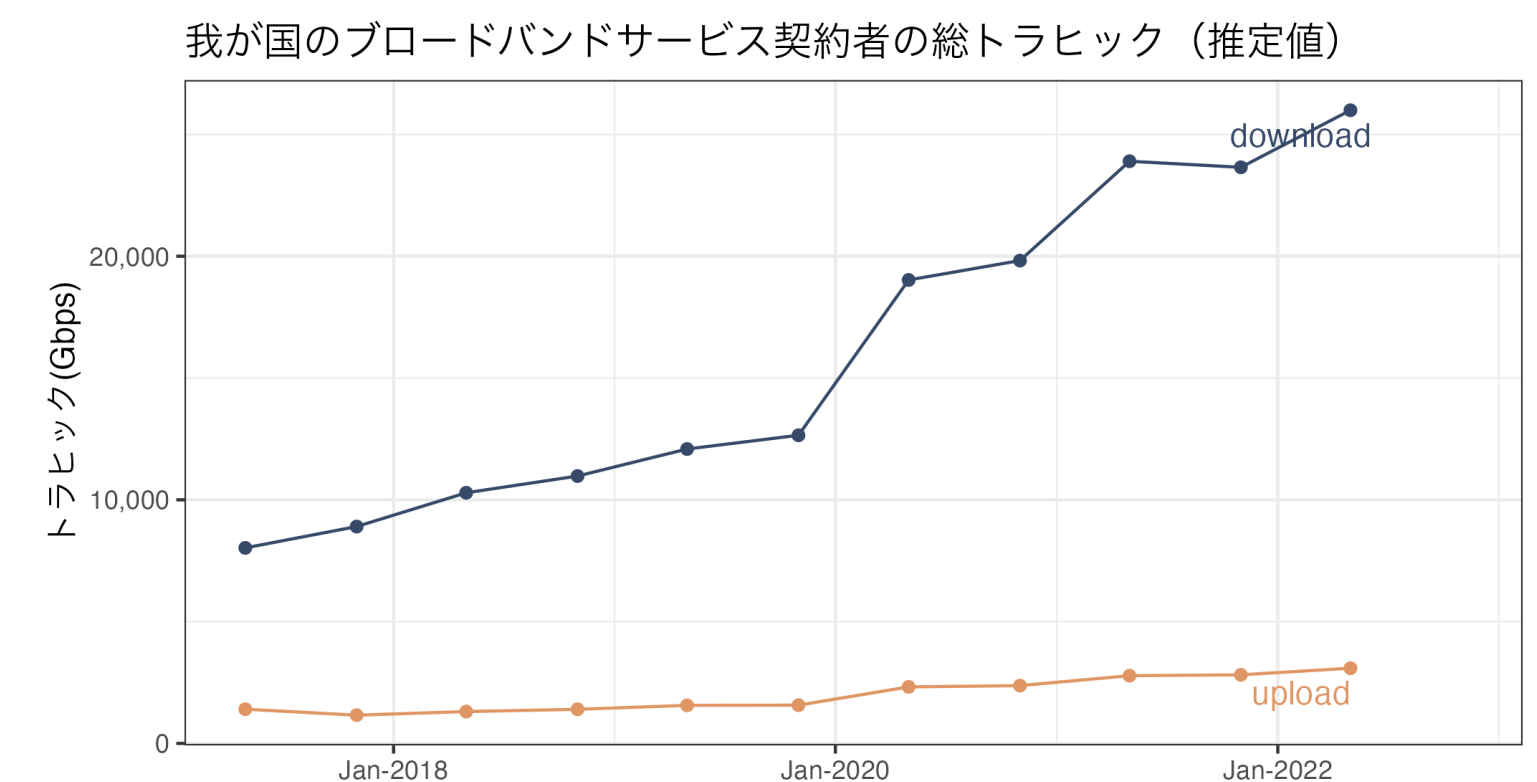
データ自身の信憑性

フェイクニュース ノイズ

価値(Value)

産業や人間社会における富、豊かさ、

満足感をもたらすもの



総務省 報道資料より作成
https://www.soumu.go.jp/joho_tsusin/eidsystem/market01_05_03.html

※Volume, Velocity, Varietyの3つで3Vとすることもあります

ビジネスにおける データ分析の重要性

ビッグテック: ビッグデータを活用した企業の成功

主な事業

代表的な製品、サービス

Alphabet

検索エンジン、インターネット広告

 Google  YouTube

Amazon

Eコマース、クラウドコンピューティング

 Amazon Web Services Kindle

Meta Platforms

ソーシャルネットワーク

 Facebook  Instagram

Apple

デジタル家電、メディアストリーミング

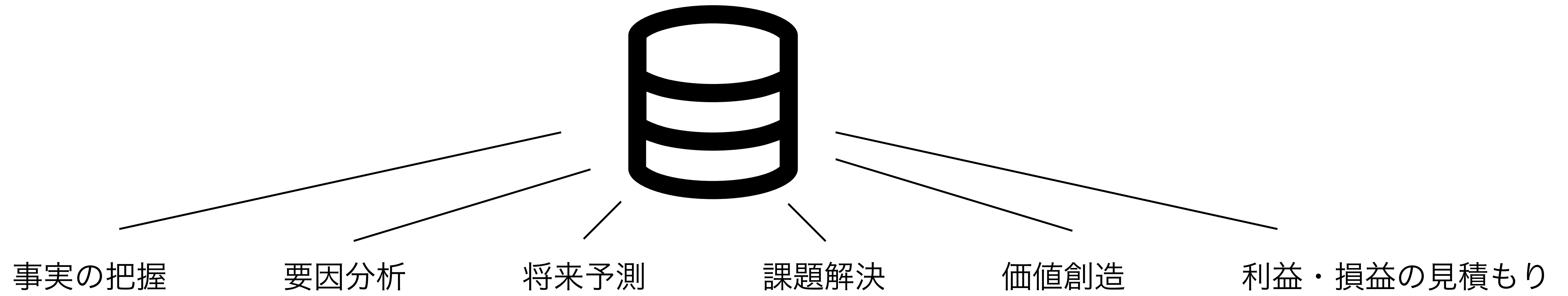
 Apple Music iPhone

これらの企業・サービスに共通した特徴→**データ駆動型**

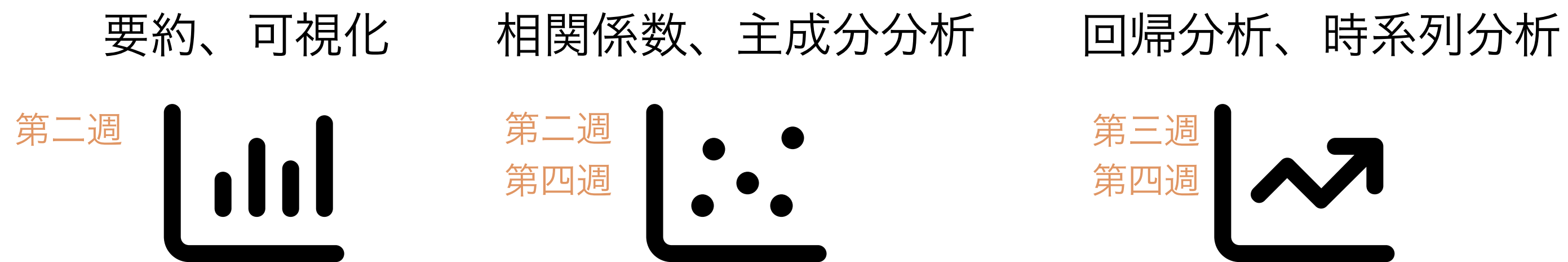
収集されたビッグデータを活用し、課題解決や価値創造をもたらす意思決定を行う

データに基づく物事の考え方

データを分析することで何が見えてくる？



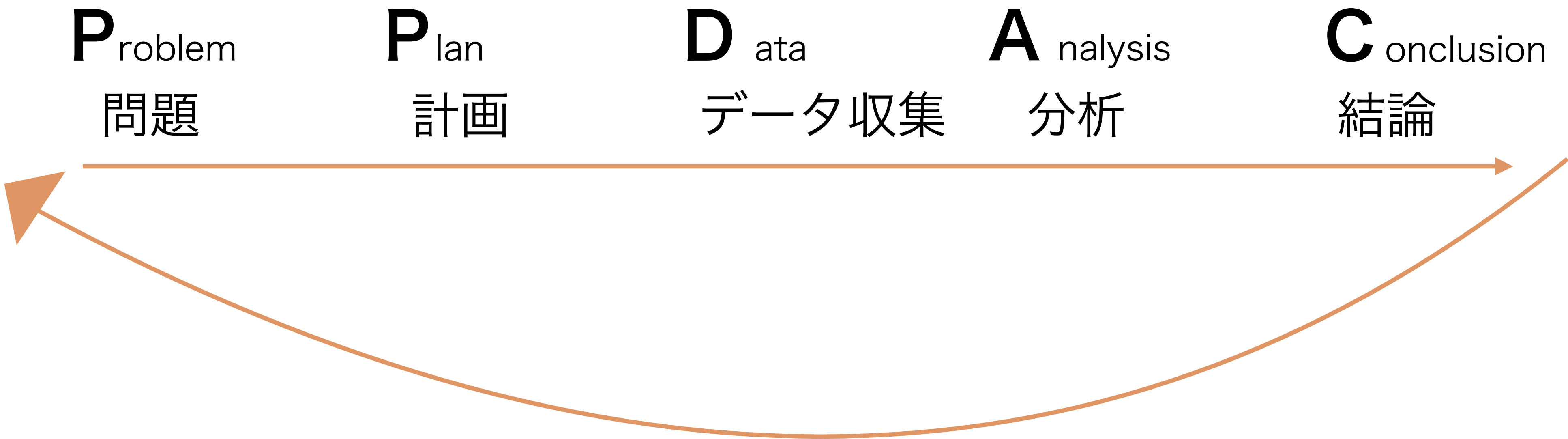
データ分析の手法の例



データから重要な情報を引き出し、ビジネスにつなげていくプロセス

「直感」をデータから検証する、妥当性を検証 ※バイアス（考え方や意見の偏り）に注意

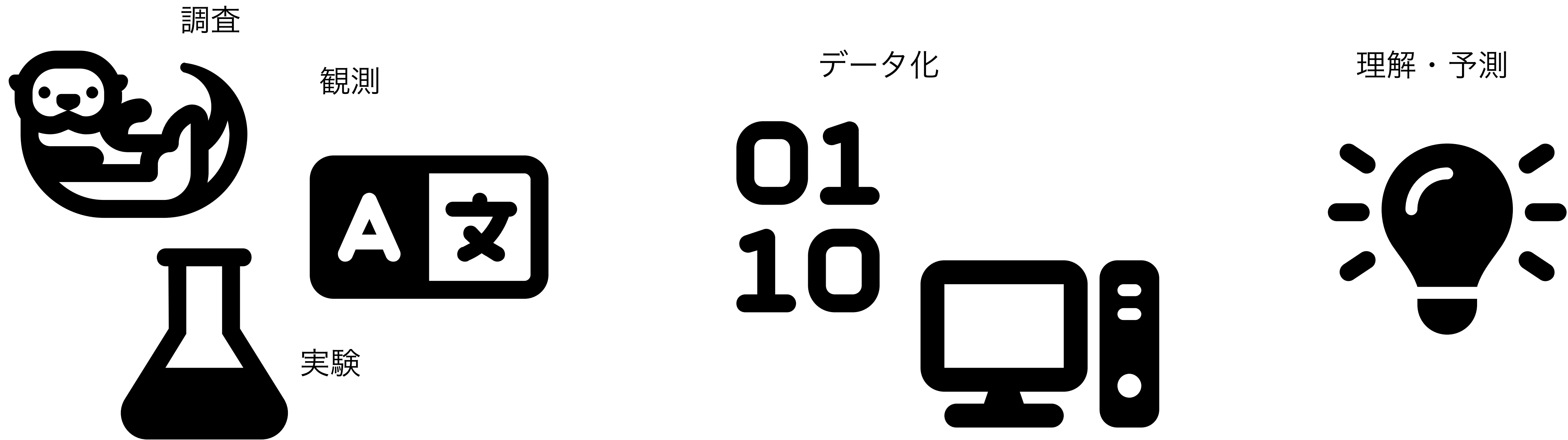
PPDACサイクル: 統計的探究の国際的枠組み



データ分析に取りかかる前に…

データ分析とは何か

データを人間が利用できる形に変換・処理を行うことで、対象についての**理解**や**予測**を目指す手続き



data
データ

判断や立論のもとになる資料・情報・事実—『スーパー大辞林』
実在から情報を抽出し、符号化する

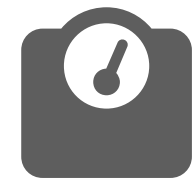
身の回りにあるデータ



データの記述

変数 … 共通の手法によって得られた値。対象によって数値が変化する値を意味する

例えば、



動物の体重、動物の分類群、動物園の来園者数

6

食肉類

320

3.5

鳥類

615

5.4

食肉類

1024

量的変数

質的変数

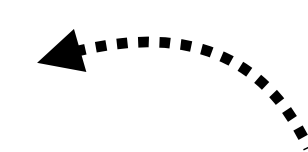
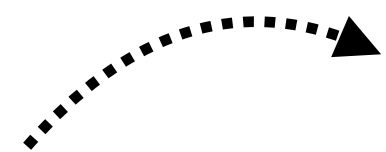
量的変数



連続変数



離散変数



データを記録する精度によって小数点以下の値が変わる

とり得る値が一定の間隔によりバラバラ

量的変数は足したり割ったりという演算ができるが、質的変数ではそれができない

表形式のデータ

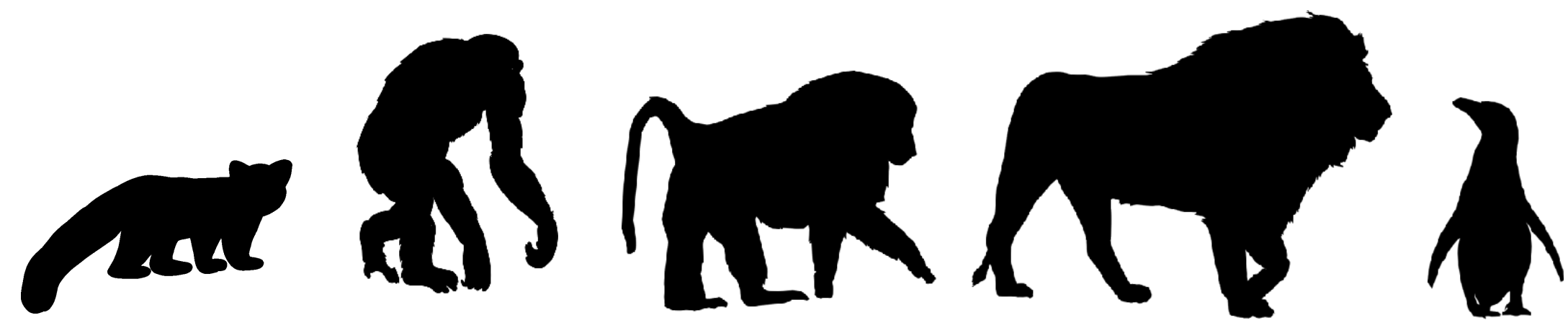
データ分析では一般的なデータの表現形式

構造化データ（取得・収集段階で並びや形式が厳密に決められるデータ）として提供されることが多い。

代表的なもの…csvファイル、エクセルファイル、データベース

構造化データに対し、文章、画像、音声、動画といったデータの記録形式に細かなルールがないものを**非構造化データ**と呼ぶ。

データフレーム (data frame)



動物についての分類群と名称（種名）、体長と体重の4つの変数を記録

食肉類	レッサーパンダ	63.5	6
霊長類	チンパンジー	85.0	60
霊長類	マントヒヒ	80.0	20
食肉類	ライオン	250.0	225
鳥類	フンボルトペンギン	69.0	6

データフレームの見方

分類群

種名

体長(cm)

体重(km)

列の名前として変数名が記録される

食肉類	レッサーパンダ	63.5	6
霊長類	チンパンジー	85.0	60
霊長類	マントヒヒ	80.0	20
食肉類	ライオン	250.0	225
鳥類	フンボルトペンギン	69.0	6

行

食肉類	レッサーパンダ	63.5	6
-----	---------	------	---

観測対象についてのすべての変数の値を含む

列

分類群
食肉類
霊長類
霊長類
食肉類
鳥類

変数の中に全データの値を含む

なぜデータ分析でプログラミング言語を利用するか

表計算ソフトでの分析… 不可能ではない

→プログラミング言語の利用により、表計算ソフトでの分析の欠点を補うことができる

表計算ソフトでの分析の欠点

特殊なソフトウェア・アプリケーションが必要となることがある

（基本的に）GUIでの操作のため、繰り返しの処理が手動化される

ワークシートで扱える範囲に制限がある

プログラミング言語を使用した場合の利点

誰もが無償で利用可能なものがある（ただし補償があるわけではない。利用は自己責任）

スクリプトに処理を記述、保存することで繰り返しの処理が自動化できる

データが変わった際に修正を加えやすい

プログラミング言語技能の獲得により、データ分析以外でも役立つ機会がある

参考文献・URL

滋賀大学データサイエンス学部・長崎大学情報データ科学部 共編 (2022). データサイエンスの歩き方 (学術図書出版社)
ISBN: 978-4-7806-0936-3

石田基広、大薮進喜 監修・著 上田哲史、掛井秀一、金西計英、谷岡広樹、中山慎一、芳賀昭弘 著 (2021).
情報科学入門: 統計・データサイエンス・AI (技術評論社)
ISBN: 978-4-297-12040-5

北川源四郎、竹村彰通 編 内田誠一、川崎能典、孝忠大輔、佐久間淳、椎名洋、中川裕志、樋口知之、丸山宏 著 (2021).
教養としてのデータサイエンス (講談社サイエンティフィク)
ISBN: 978-4-06-523809-7

「令和4年版情報通信白書」(総務省) <https://www.soumu.go.jp/johotsusintokei/whitepaper/r04.html>