

# 序論

## Data visualization

川田恵介  
東京大学

keisukekawata@iss.u-tokyo.ac.jp

2025-07-30

## 1 概論

### 1.1 本講義

- 「Data Visualization を活用したデータ分析」について、包括的な紹介を行う
  - ▶ 企業の戦略や政策等の意思決定への活用を念頭
  - ▶ Visualization + データ分析の基礎概念の学習 + 実習

### 1.2 多様化

- 学際的に発展しており、(伝統的)統計学、計量経済学、医療/生物統計、奇怪学習(AI)など、(互いに重複する)分野が存在し、さまざまな分析方法を提案
  - ▶ 本講義では極力包括的にこれらの手法を紹介し、実装する

### 1.3 データ例: 中古マンション取引データ

District	Price	Size	year_2024
千代田区	94	40	1
千代田区	100	65	0
千代田区	130	65	1
千代田区	98	65	0
千代田区	58	40	0
千代田区	330	95	1

- 事例と変数(属性)からなる

## 1.4 データ分析

- 業務の電子化等に伴い、数多くのデータが蓄積されている
  - ▶ そのままでは膨大な情報が乱雑に保存されているだけであり、人間が活用できない
- データ分析の大目標: データを人間が理解でき、かつ、信頼できる”情報”に変換する
  - ▶ 人間による意思決定の支援に繋げたい

## 1.5 Data Visualization

- 情報は、数値や数式で表現されてきた
  - ▶ 人間の認知能力との相性が悪く、情報を直感的に理解するためには、ある程度の訓練が必要
- PC の処理能力向上に伴い、情報を視覚的に表現しやすくなった
  - ▶ より多くの情報を伝えやすくなる
- 実務においても注目されている (例: Uber Eats)

## 1.6 例: 平均取引価格表

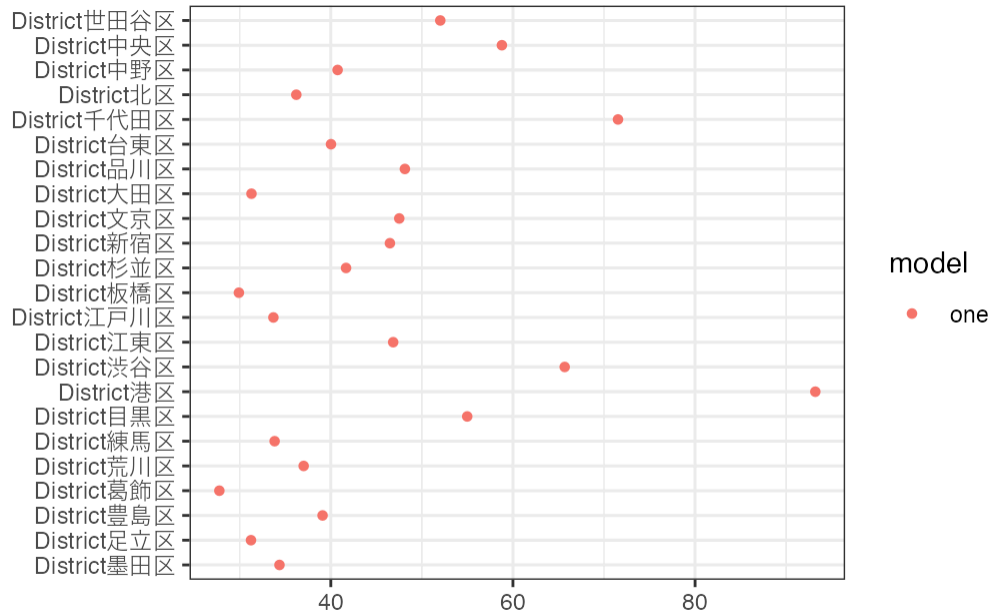
Call:

```
lm(formula = Price ~ 0 + District, data = data)
```

Coefficients:

District世田谷区	District中央区	District中野区	District北区
52.02	58.79	40.75	36.21
District千代田区	District台東区	District品川区	District大田区
71.54	40.02	48.13	31.29
District文京区	District新宿区	District杉並区	District板橋区
47.51	46.49	41.67	29.91
District江戸川区	District江東区	District渋谷区	District港区
33.69	46.85	65.68	93.20
District目黒区	District練馬区	District荒川区	District葛飾区
54.98	33.83	37.02	27.76
District豊島区	District足立区	District墨田区	
39.09	31.23	34.36	

## 1.7 例: 平均取引価格の図



## 1.8 Dashboard

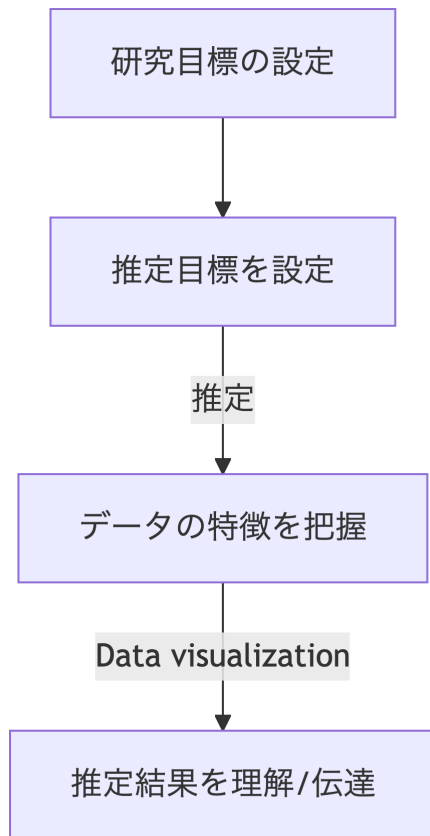
- 日常的に確認すべき指標を一覧として表示し、日常的に更新する
  - ▶ 例: NTT データ, デジタル省
- 社会や組織の動きを包括的に捉えることを目標
  - ▶ 「特異な動きをした数字」のみに注目する(“センセーショナルリズム”)のではなく、普通の動きも合わせて認識する必要がある

## 1.9 まとめ

- 実務組織からの関心も高いデータ分析の学習は強く推奨
  - ▶ 例: Amazon, Cyber Agent, Microsoft, Mizuho, Netflix, Uber
  - ▶ 講師自身の経験: 日本経済研究センター、日本銀行、厚生労働省、内閣府
- 成績評価: 授業参加度 (20%)、毎回の課題 (40%)、最終レポート (40%)
- 参考章: Statistical Inference via Data Science

## 2 データ分析の流れ

### 2.1 Work flow



## 2.2 データの特徴推定

- データが持つ大量の情報を、適切に整理/要約する
- 注意点: データの特徴そのものは、研究関心ではない
  - ▶ データの背後にある社会(母集団)の特徴が関心 (推定目標)
  - ▶ 社会の特徴を推論できるように、情報を抽出する (推定) 必要がある

## 2.3 研究/推定目標

- 推定目標の設定: データから推定可能、かつ、研究目標に関連する社会の特徴を設定する
- 研究目標の設定: データ分析から明らかにしたい社会やその仕組みの特徴を設定する
- 講義は、データからの推定 → 推定目標の設定 → 研究目標の設定 の順番で議論する

## 2.4 研究目標の例

- 時系列比較: 2019-2024 年で、東京 23 区の中古マンション価格はどのように変化したのか?

- 因果効果: ある介入(例:改築)によって、ある結果(例:取引価格)がどのように**変化**するのか推定する
  - ▶ 社会の”仕組み”に関する研究課題

## 2.5 到達目標

- エントリーシート等で、“大学で学んだこと”を書く際に、
  - ▶ “機械学習や因果推論などのデータ分析の手法を用いて、東京の不動産市場の分析を行い、結果を dashboard 形式でレポートにまとめた経験がある”と書くことができる

## Bibliography