

# TRADのいま

データサイエンスコンソーシアム，慶應義塾大学

柴田 里程



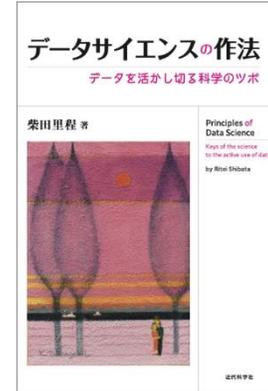
柴田里程著『データサイエンスの作法』近代科学社, 2020

カミヤシロ

5つ星のうち3.0 TRADとは

2021年6月18日に日本でレビュー済み

Amazonで購入



TRADというソフトで出来ることはパラレル・コーディネート・プロット  
(並行座標プロット)です。少しRに慣れていれば自分でも描けます。  
それより、簡潔に書かれた各章の「まとめ」が役立ちます。  
厚紙の「下敷き」のような形式でおまけにして欲しかったです。

ソフトウェア: できる, できない

ソフトウェア環境: (必ずしもはっきりしていない) 目的に向かって進むための快適な環境

# データサイエンス環境

データの理解を先導する環境

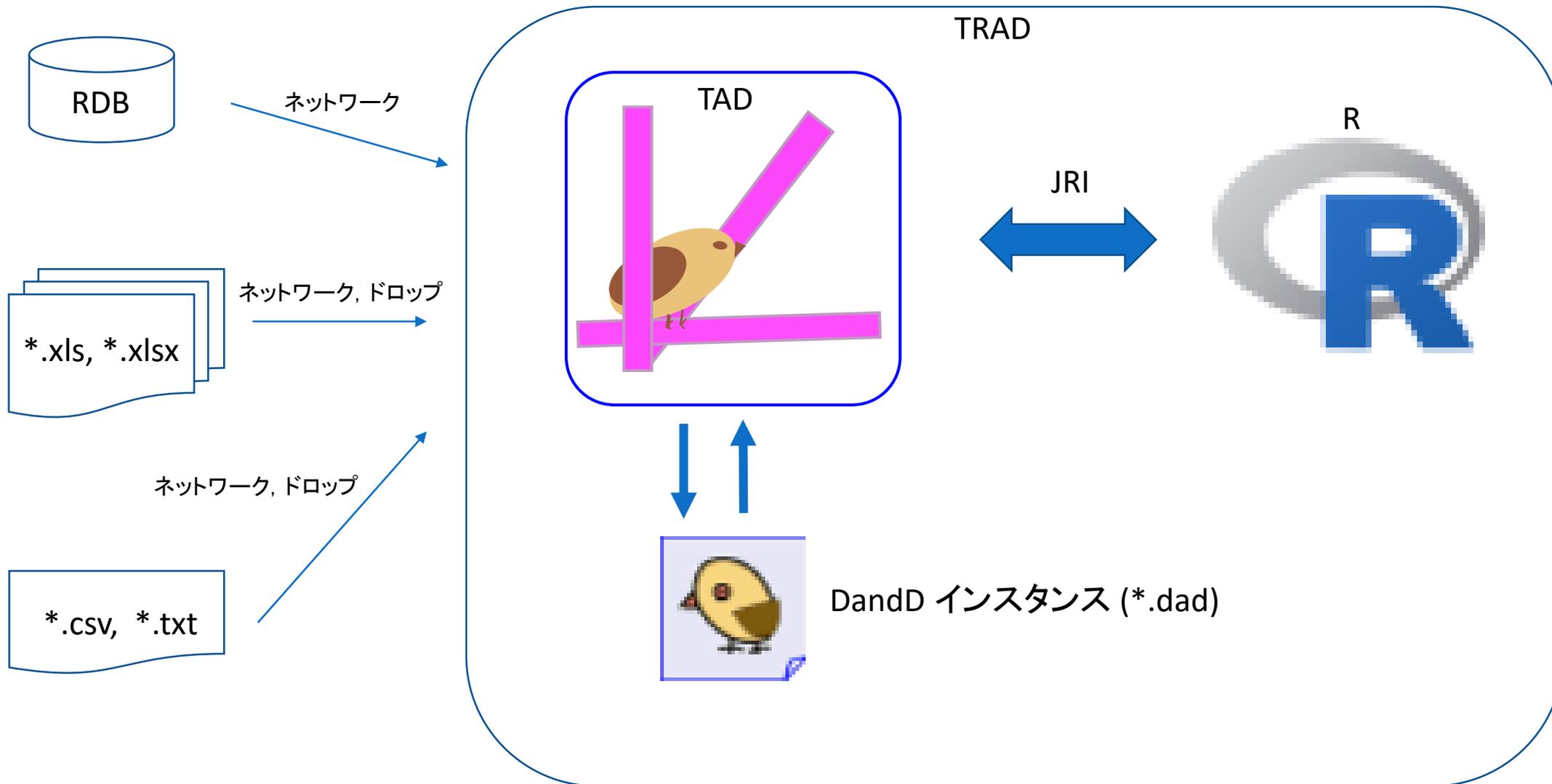
人間がデータを理解するとは？

人間がデータを理解する手順は？

人間がデータを理解するための快適な環境は？



データサイエンス研究



# SDGs data science (Sustainable Development Goals for Data Science)

## 持続可能なデータサイエンス

労力の無駄遣いを廃し, 得られた理解のシステムティックな蓄積と活用

つまり, 人間の有効活用

### Goals

1. 人間
2. 実践
3. 伝達
4. 蓄積
5. 活用
6. 研究

# データサイエンス研究

データベース理論 (RDBMS) から学ぶこと

1. アトミック
2. 正規化
3. 記録の集合としてのデータテーブル

データを処理する

データベース理論と異なる点

1. 基本はデータテーブルではなくデータベクトル
2. 変数の概念
3. 変数の属性
4. 視点, 視野, 視覚
5. 意味

データを理解する

# 変量の多重類別

## 1. 型 (type)

- a. 計測値
- b. 計数
- c. 序数
- d. 日時
- e. 時間
- f. 記録度数
- g. マーク
- h. 順マーク
- i. 論理

## 2. 系 (system)

- a. 基数系
- b. 座標系
- c. 階層
- .
- .
- .

## 3. 役 (role)

- a. ID
- b. 補助
- c. 主

## 4. モード (mode)

- a. 説明変量
- b. 被説明変量
- .
- .
- .

多重類別は固定したものではなく、独立しているとも限らない

# 型: 値の類別

型 (type)	原型 (primitive type)
a. 計測値	実数
b. 計数	非負整数
c. 序数	自然数
d. 日時	ロング整数(経過ミリ秒)
e. 時間	ロング整数(経過ミリ秒)
f. 記録度数	正数
g. マーク	マーク値
h. 順マーク	マーク値 (+ 順序)
i. 論理	0 または 1

分割表: The Lady Tasting Tea

	ミルクが先 だと思う	ミルクが後 だと思う
ミルクが先	29	21
ミルクは後	19	31

ミルク	判定	件数
先 先 後 後	先 後 先 後	29 21 19 31
マーク	マーク	記録度数

## 2次データ

ミルク	判定	件数
先 先 後 後	先 後 先 後	29 21 19 31
マーク	マーク	記録度数



## 1次データ

ミルク	判定
先 先 先 先 . . . 後 後 後 後 後	先 先 先 後 . . . 先 後 後 先 先
マーク	マーク

# 記録度数と計数

記録度数: その変数以外の(主)変数が度数の対象を規定

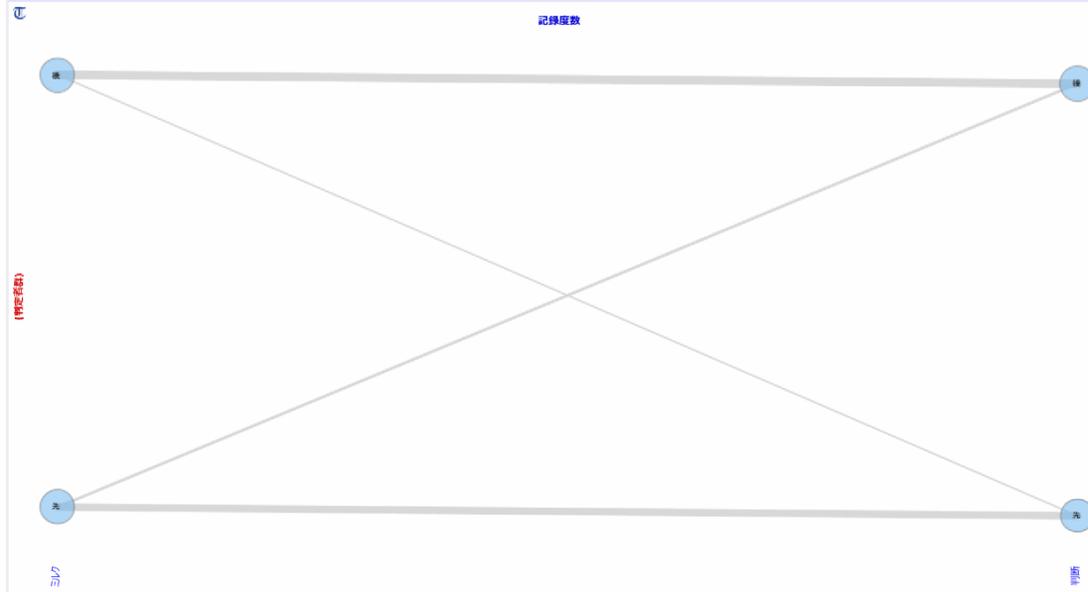
計数: 対象は独立に定まる

例:

ミルク 判定 件数

VS

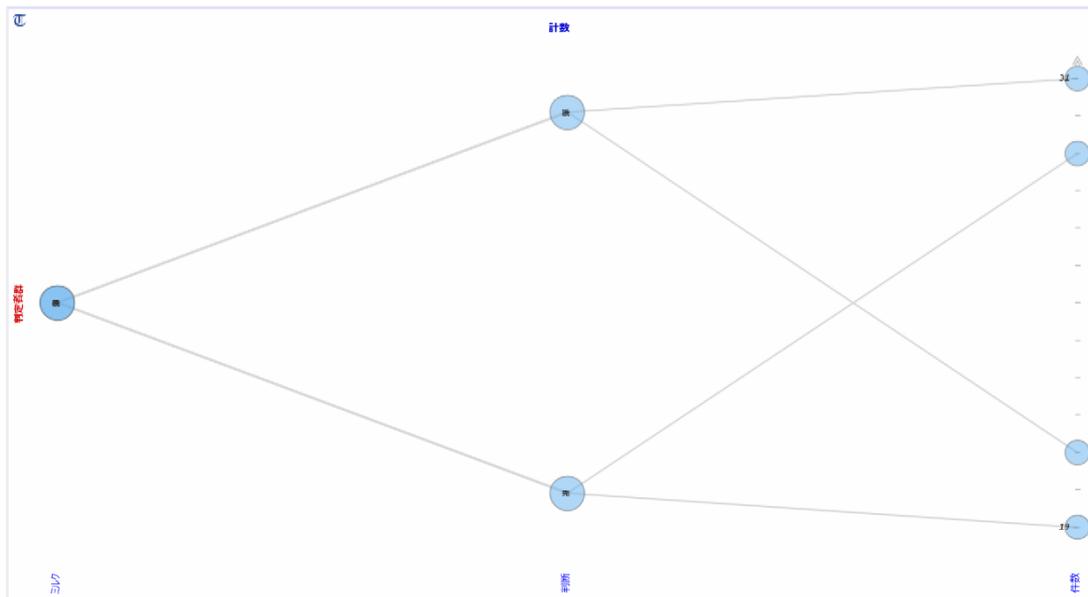
時刻 市町村 自動車事故数



線の太さが記録度数を反映

記録度数

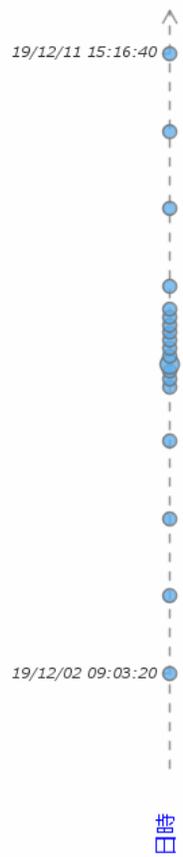
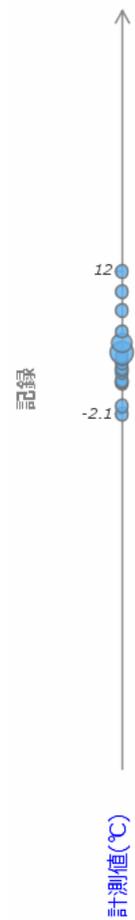
計数



ミルクが先: 50杯  
ミルクが後: 50杯

度数にミルクが先の杯かミルクが後の杯かは関係しない

# TRADでの型の表示



# 系：複数の変量が構成する構造

基数系： 組で基数を構成する変量群  
例：年月日, 時分秒

座標系： 組で座標系を構成する変量群  
例：緯度・経度, 空間座標

階層： 木構造を構成する変量群  
例： . . . に関する変量  
. . . の . . . に関する変量

TRAD

部分関係  
(HypoRelation)

/A/I/a のような長名

型が日時,  
時間の単  
一変量に

正規化

# 役

TRAD

**ID:** 値が記録を一意に定める変量

**補助変量:** 参考にはするが, とりあえずは除外する変量

Id の色で区別

**主変量:** 注目する変量でデータに対する「視点」を定める

型とこの役は独立  
型が ID だからといって役も ID である必要はない

# 1次データ

ミルク	判定
先 先 先 先 . . . 後 後 後 後 後	先 先 先 後 . . . 先 後 後 先 先
マーク	マーク



# 0次データ

判定者	性別	ミルク	判定
1 2 3 4 . . . 96 97 98 99 100	男 女 男 男 . . . 女 女 男 男 男	先 先 先 先 . . . 後 後 後 後 後	先 先 先 後 . . . 先 後 後 先 先
ID	マーク	マーク	マーク
ID	補助	主	主

# モード:モデルでの役割

データに対するモデルを導入した時点で発生する変量の類別

- |          |         |
|----------|---------|
|          | TRAD    |
| 1. 説明変量  | 変量名を黒色  |
| 2. 被説明変量 | 変量名を灰色に |
- 
1. 回帰モデルなどの当てはめ結果の視覚表示
  2. モデル当てはめに必要な指定

# データサイエンス研究

