

# RとQuartoではじめるデータサイエンス《2026》

## #4 Rの基本的な操作方法(2)

苅谷千尋

Thu, 30, Apr, 2026

### 目次

1. 前回の振り返り
2. データを読み込む
3. 必要なデータを取り出す (基本編)
4. 必要なデータを取り出す (発展編)
5. 必要なデータを取り出す (応用編)

### 0. 本日の目標

1. フォルダ・ファイル関係・階層を意識し、データを読み込めるようになる
2. 元データを加工して (選択、結合)、必要なデータを作る **流れ**を理解する
3. `mutate()` と `summarise()` の違いを理解し、目的に応じて新しい変数を作成・集計できるようにする
4. 代入の考え方を理解して、使いこなせるようになる
5. 水準 (factor) を設定することで、並び順や可視化が変わることを **体験的に**理解する
6. 縦持ちデータの便利さがどこにあるのかを **実感**してみる

### 2. データを読み込む

- SSDSE (教育用標準データセット) : **エクセル集計が前提**。まだまだRでは使いづらい
- CSVファイル : Comma-Separated Values
  - 書式 (色・セル結合) や数式などの情報はもたないが、様々なアプリで読み書きできる汎用性の高いデータ形式
  - CSVファイルとエクセルファイルを選択できる場合は、**CSVを推奨**
- 文字コード
  - 世界基準 : UTF-8 ; 日本 : ローカル文字コードShift\_JIS中心 (文字化け原因)
- データの読み込み : `read_csv()` 関数
  - 例 : `row_data <- read_csv("data/SSDSE-D-2023.csv")`
  - Cf. Excelファイルの読み込み : readxlパッケージを追加
- ファイルパス : ファイルがどのフォルダに保存されているかを階層構造で表した文字列
  - 絶対パス : 一番上の階層 (ルート) から、ファイルまでの経路をすべて書いたもの
    - Windows: `C:/Users/user/Documents/大学/レポート.docx`
  - 相対パス : 基準となる場所から見たファイルの位置まで書いたもの
    - Windows: `大学/レポート.docx`
    - Rの相対パスの基準 : プロジェクト (.Rproj) があるフォルダ。ここを基準に階層関係を考える

### 3. 必要なデータを取り出す (基本編)

- `mutate()` 関数
  - データフレームに**新しい列を追加**したり、**既存の列を変形**したりする関数
  - 各行ごとに処理を行いながら、**計算・水準の設定・文字の整形**など多様に使える
- `mutate()` と `summarise()` の相違点
  - どちらも計算に使う関数だが、データのまとめ方が異なる
    - `mutate()` : 行はそのまま、列が増える
    - `summarise()` : 行が減る (集約される)
- 代入
  - パイプは「処理の流れ」であり、結果を自動保存する仕組みではない
  - 再利用したい場合は `<-` (代入演算子) でオブジェクトに保存する
- 水準の設定
  1. 文字ベクトルを作り、`factor()` で水準を指定
  2. `reorder()` : 昇順 (小さい → 大きい)

### 5. 必要なデータを取り出す (発展編)

- 縦持ち変換・横持ち変換
  - `pivot_longer()` : 横持ちデータを縦持ちデータに変換
  - `pivot_wider()` : 縦持ちデータを横持ちデータに変換
- 横持ちデータの特徴
  - 同じ種類のデータなのに列がバラバラ
  - → 一度に計算しにくい; グラフにしにくい
- 縦持ちデータの特徴
  - 同じ種類のデータは1つの列 (変数) に整理されている
  - カテゴリごとに扱えるようになる
  - → 同じ処理を1回でまとめて書けるので、コードがシンプルになる
  - → 関数を自然に使えるようになる (例: `group_by()` や `ggplot()`)
  - → 同じ種類のデータをまとめて扱えるようになる

### 宿題

#### 授業の感想:

- 授業の感想:
  - 回答先: Google Forms
  - 締め切り: 5月1日 (金) 23時59分

#### 演習:

- 内容: 演習⑦⑨
- 回答先: dropbox
- 締め切り: 5月11日 (月) 23時59分