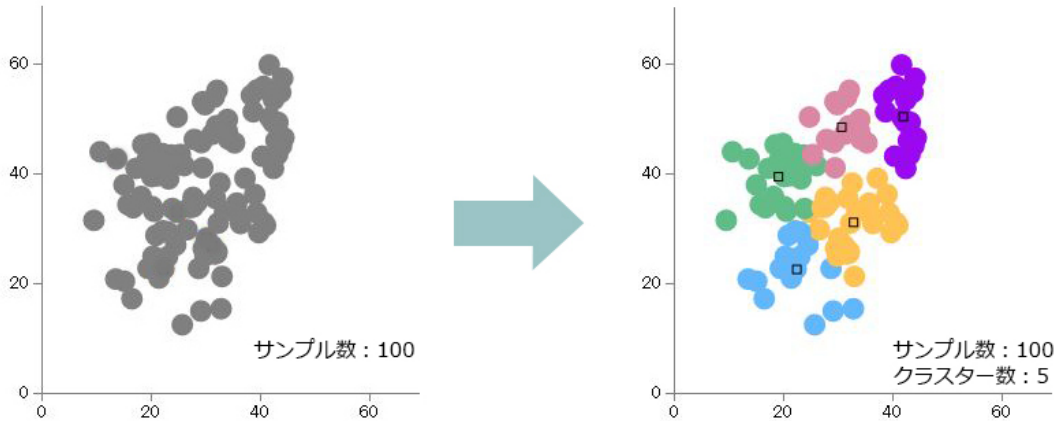


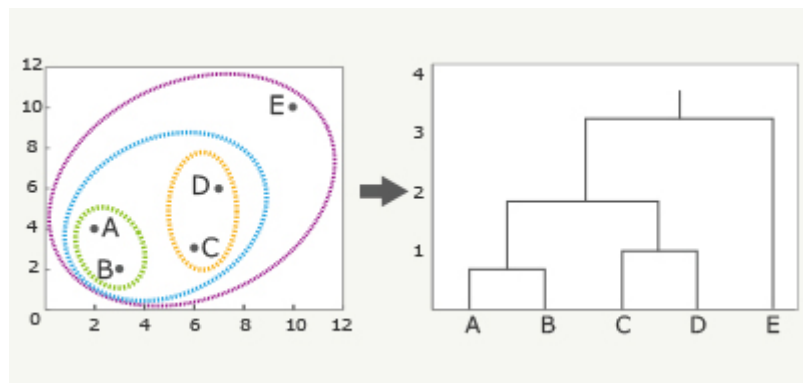
01. クラスター解析とは

クラスター解析とは、**データを統計的にグループ化（クラスタリング）する方法**のこと。

プロットのまとまりでグループを視覚化する非階層クラスター解析。



樹状にグループを視覚化する階層クラスター解析がある。



Rでクラスター解析を行うために、以下の手順が必要。

1. **データを読み込む**
2. **データを距離行列に変換**
3. **距離行列をクラスタリング**
4. **解析結果を視覚化**

02-01. 関数の探し方 & 引数の調べ方

◇ 探し方

Rでやってみたいことがあったら、英語を使って『**OO in R**』とググる。

（例）表の読み込みをしたいから、『read table in r』とググる。

◇ 引数の調べ方

英語圏エンジニアのブログ等で自分のやりたいことをしているfunctionを探す。

その後、`function` の具体的な引数を <https://www.rdocumentation.org> で調べる。

(例) `read.table` で表の読み込みができることを見つけ、`read.table` とググる。

関数リンク：<https://www.rdocumentation.org/packages/utils/versions/3.5.1/topics/read.table>

以下の画面では、`read.table` の**引数の一覧**が示されている。

`read.table` の後ろを丸括弧で囲い、その中に `file`、`header`、`sep` などの引数を書き込む必要がある。

The screenshot shows the RDocumentation page for the `read.table` function. The page title is "read.table". It indicates it is from `utils v3.5.1` by `R-core R-core@R-project.org` and has a "17th Percentile" badge. The "Data Input" section states: "Reads a file in table format and creates a data frame from it, with cases corresponding to lines and variables to fields in the file." The "Keywords" section lists `file` and `connection`. The "Usage" section contains the following code snippets:

```
read.table(file, header = FALSE, sep = "", quote = "\"",
  dec = ".", numerals = c("allow.loss", "warn.loss", "no.loss"),
  row.names = NULL, col.names = as.is = I(stringsAsFactors = FALSE),
  na.strings = "NA", colClasses = NA, nrows = -1,
  skip = 0, check.names = TRUE, fill = FALSE, lines.skip = TRUE,
  strip.white = FALSE, blank.lines.skip = TRUE,
  comment.char = "#",
  allowEscapes = FALSE, flush = FALSE,
  stringsAsFactors = default.stringsAsFactors(),
  fileEncoding = "", encoding = "unknown", text, skipNul = FALSE)

read.csv(file, header = TRUE, sep = ",", quote = "\"",
  dec = ".", fill = TRUE, comment.char = "#", ...)

read.csv2(file, header = TRUE, sep = ";", quote = "\"",
  dec = ".", fill = TRUE, comment.char = "#", ...)
```

以下の画面では、`Arguments` (引数)の説明がされている。

`file` : 読み込むデータ名を入力。

`header` : `TRUE` の場合、ヘッダーが有るものとして読み込む。`FALSE` は無い場合。

`sep` : `" "` の場合、スペース区切り形式として読み込む。ここに書かれてはいないが、カンマ形式の場合 `,`、タブ切り形式の場合 `\t` に変更可能。

The screenshot shows the "Arguments" section of the `read.table` function documentation. It lists the following arguments with their descriptions:

- file**: the name of the file which the data are to be read from. Each row of the table appears as one line of the file. If it does not contain an *absolute* path, the file name is *relative* to the current working directory, `getwd()`. Tilde-expansion is performed where supported. This can be a compressed file (see `file`). Alternatively, `file` can be a readable text-mode `connection` (which will be opened for reading if necessary, and if so `close` d (and hence destroyed) at the end of the function call). (If `stdin()` is used, the prompts for lines may be somewhat confusing. Terminate input with a blank line or an EOF signal, `Ctrl-D` on Unix and `Ctrl-Z` on Windows. Any pushback on `stdin()` will be cleared before return.) `file` can also be a complete URL. (For the supported URL schemes, see the 'URLs' section of the help for `url` .)
- header**: a logical value indicating whether the file contains the names of the variables as its first line. If missing, the value is determined from the file format: `header` is set to `TRUE` if and only if the first row contains one fewer field than the number of columns.
- sep**: the field separator character. Values on each line of the file are separated by this character. If `sep = ""` (the default for `read.table`) the separator is 'white space', that is one or more spaces, tabs, newlines or carriage returns.
- quote**: the set of quoting characters. To disable quoting altogether, use `quote = ""` . See `scan` for the behaviour on quotes embedded in quotes. Quoting is only considered for columns read as character, which is all of them unless `colClasses` is specified.
- dec**: the character used in the file for decimal points.
- numerals**: string indicating how to convert numbers whose conversion to double precision would lose accuracy, see `type.convert` . Can be abbreviated. (Applies also to complex-number inputs.)
- row.names**: a vector of row names. This can be a vector giving the actual row names, or a single number giving the column of the table which contains the row names, or character string giving the name of the table column containing the row names.

◇ (1) `install.packages()` : パッケージをインストール

(例) `ape` パッケージ

```
install.packages("ape")
```

パッケージをインストール可能。

インストール時に、どのミラーサイトでインストールするか聞かれるので、『Japan(Tokyo)』を選ぶ。

リンク : <https://www.rdocumentation.org/packages/utils/versions/3.5.1/topics/install.packages>

◇ (2) `library()` : パッケージを展開

(例) `ape` パッケージ

```
library(ape)
```

パッケージを展開（アクティブ化）する関数。

パッケージはRを閉じるたびに非アクティブ化されてしまうので、**起動するたびに `library()` が必要。**

リンク : <https://www.rdocumentation.org/packages/base/versions/3.5.1/topics/library>

◇ (3) ディレクトリの指定

読み込みたいファイルは指定したディレクトリ(=作業フォルダ)内にはないといけないので、まず指定。

Rの上タブにある『**ファイル⇒ディレクトリの変更**』から、ディレクトリを指定可能。

◇ (4) `object <- "〇〇"`

(例) `test.txt` を `table` と名づけたオブジェクトに格納。

```
table <- "test.txt"
```

Rで中核となる命令。

データなどを文字に割り当てる。

割り当てを『**格納**』と言う。(受験数学の『〇〇を文字Xと置く』のイメージ)

データをオブジェクトに格納してから扱っていく。

◇ (5) `remove(list=ls())` : 全てのオブジェクトを削除

```
remove(list=ls())
```

今までに作成されたオブジェクトを全て削除可能。

試行錯誤していると間違った格納をしていることがあるので、時々リセットしておく。

リンク : <https://www.rdocumentation.org/packages/base/versions/3.5.1/topics/remove>

◇ (6) `summary()` : データの中身を確認

(例) オブジェクトである `table` の中身を確認。

```
summary(table)
```

データの中身を見ることが可能。

リンク : <https://www.rdocumentation.org/packages/base/versions/3.5.1/topics/summary>.

◇ (7) `read.table()` : 表データの読み込み

(例) `test.txt` をタブ有り・スペース区切りとして読み込み。

```
table <- read.table("test.txt",  
header=TRUE,  
sep="")
```

拡張子も忘れずに、ファイル名を `" "` で囲って指定。(オブジェクトを引数として選ぶ場合 `" "` は不要)

`header` : ヘッダーの有無を指定。

`sep` : 表データ形式を指定。(`/t` : タブ切、 `,` : カンマ区切り)

処理の出力先として、`table` に格納。

リンク : <https://www.rdocumentation.org/packages/utils/versions/3.5.1/topics/read.table>

※ `table` として、次のようなペアワイズ比較の有意差の対応表を扱うこともできる。

(ペアワイズ : 総当たりに比較された状態)

	A	B	C	D	E
A					
B	0.003				
C	0.001	0.0015			
D	0.002	0.007	0.08		
E	0.0001	0.03	0.034	0.02	

◇ (8) `class()` : データフレームを確認

(例) `table` のデータフレームを確認。

```
class(table)
```

データの `class` を確認。

`function`によって読み込みのできるデータフレームが異なるので、事前に確認しておく。

例として、`table` の `class` を確認。

リンク : <https://www.rdocumentation.org/packages/base/versions/3.5.1/topics/class>

◇ (9) `data.matrix()` : データフレームを距離行列に変換

(例) `table` を距離行列に変換。

```
lower.matrix <- data.matrix(table)
```

```
class(lower.matrix)
```

クラスター解析を行う `function` は、データフレームが距離行列でないと認識できないため、`data.matrix` で事前に変換しておく必要がある。

変換後、ために、`class` でデータフレームを確認すると、`matrix` と返されるはず。

有意差表は下半分にしかデータがないので、`lower.matrix` として格納。