

# 数量的アプローチ1 第1回 2024年4月5日

- 本日の内容

1. 数量的アプローチで何を学ぶか

回帰分析とは？

2. 統計ソフトRのインストール

この授業を通して使う統計ソフトの導入

3. データの作成と読み込み

データ分析の準備作業



# RとRStudio

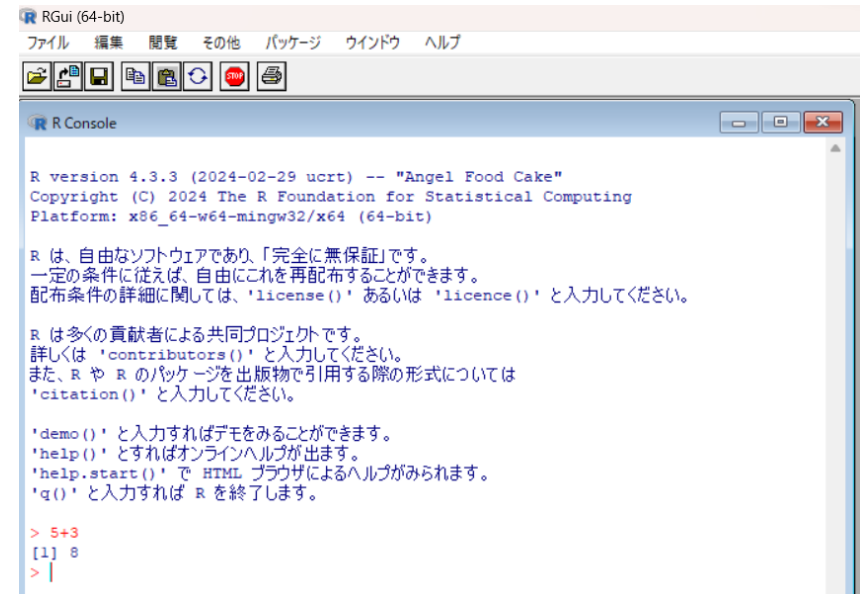
- R

高機能統計分析ソフト（フリー）

R言語を使用

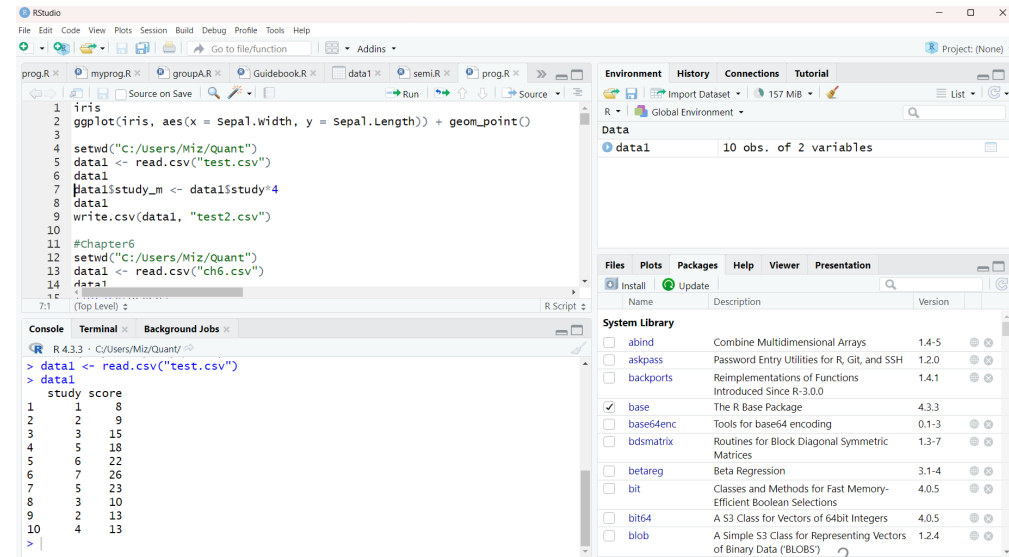
世界中の研究者が使用

Rだけでも分析可能



- RStudio


Rを使いやすくするための統合開発環境（IDE: Integrated Development Environment）



# R（とRStudio）のインストール前に

- ユーザー名の確認

Win：スタートメニュー  から  の横に見える。

Mac：Finder  から「移動」→「ホーム」とすると  
/Users/**ユーザー名** のように表示される。

- ここが**日本語**だと大変かも。

とりあえず今日はインストールして、うまくいかなければ対処。

# Rのインストール Winの場合

- <https://cran.r-project.org/bin/windows/base/> に移動し、  
ここをクリックしてDLする。

## R-4.3.3 for Windows

[Download R-4.3.3 for Windows](#) (79 megabytes, 64 bit)

[README on the Windows binary distribution](#)

[New features in this version](#)

This build requires UCRT, which is part of Windows since Windows 10 and Windows Server 2016. On older systems, UCRT has to be installed manually from [here](#).

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server.

# Macの場合

- <https://cran.r-project.org/bin/macosx/> に移動し、

- プロセッサで  
どちらかを

## R for macOS

This directory contains binaries for the base distribution and of R and packages to run on macOS. R and package binaries for R versions older than 4.0.0 are only available from the [CRAN archive](https://cran.r-project.org) so users of such versions should adjust the CRAN mirror setting (<https://cran.r-project.org>) accordingly.

Note: Although we take precautions when assembling binaries, please use the normal precautions with downloaded executables.

### R 4.3.3 "Angel Food Cake" released on 2024/02/29

Please check the integrity of the downloaded package by checking the signature:

```
pkgutil --check-signature R-4.3.3-arm64.pkg
```

in the *Terminal* application. If Apple tools are not available you can check the SHA1 checksum of the downloaded image:

```
openssl sha1 R-4.3.3-arm64.pkg
```

### Latest release:

For Apple silicon (M1-3) Macs:

[R-4.3.3-arm64.pkg](#)

SHA1-  
hash: 37dc051e0a241eeef8e8207b2710067767781e6c  
(ca. 92MB, notarized and signed)

For older Intel Macs:

[R-4.3.3-x86\\_64.pkg](#)

SHA1-hash: cbff7e4657eb67d3a35f4c74772680aa3c6b8d4b  
(ca. 94MB, notarized and signed)

**R 4.3.3** binary for macOS 11 (**Big Sur**) and higher, signed and notarized packages.

Contains R 4.3.3 framework, R.app GUI 1.80, Tcl/Tk 8.6.12 X11 libraries and Texinfo 6.8. The latter two components are optional and can be omitted when choosing "custom install", they are only needed if you want to use the `tc1tk` R package or build package documentation from sources.

macOS Ventura users: there is a known bug in Ventura preventing installations from some locations without a prompt. If the installation fails, move the downloaded file away from the *Downloads* folder (e.g., to your home or Desktop).

# インストールの手順

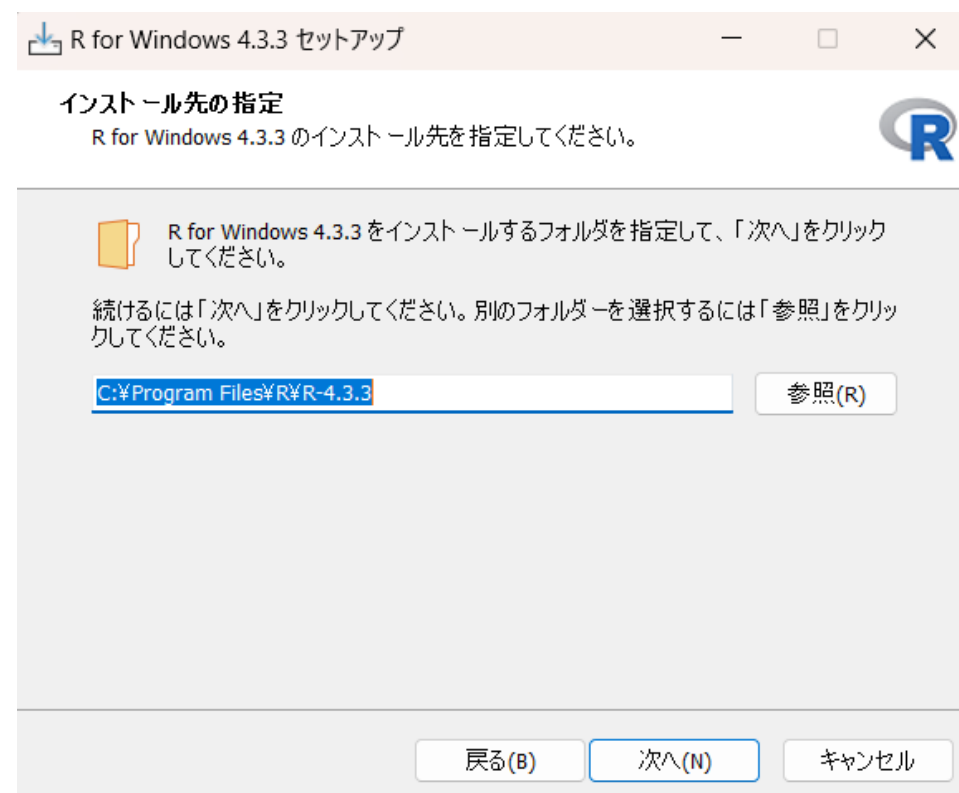
「R-4.3.3-win.exe」という実行ファイル（MacはR-4.3.3-arm64.pkg）を入手したらクリック。Winでは以下の表示がでるので、ボタンを押しつつ進める。（Macは「続ける」→「インストール」）

1. 「このアプリがデバイスに変更を加えることを許可しますか？」には「はい」。
2. セットアップに資料する言語は「日本語」のまま「OK」。
3. 情報をスクロールして確認したら「次へ」。

# インストール先について注意

## 4. インストール先の指定

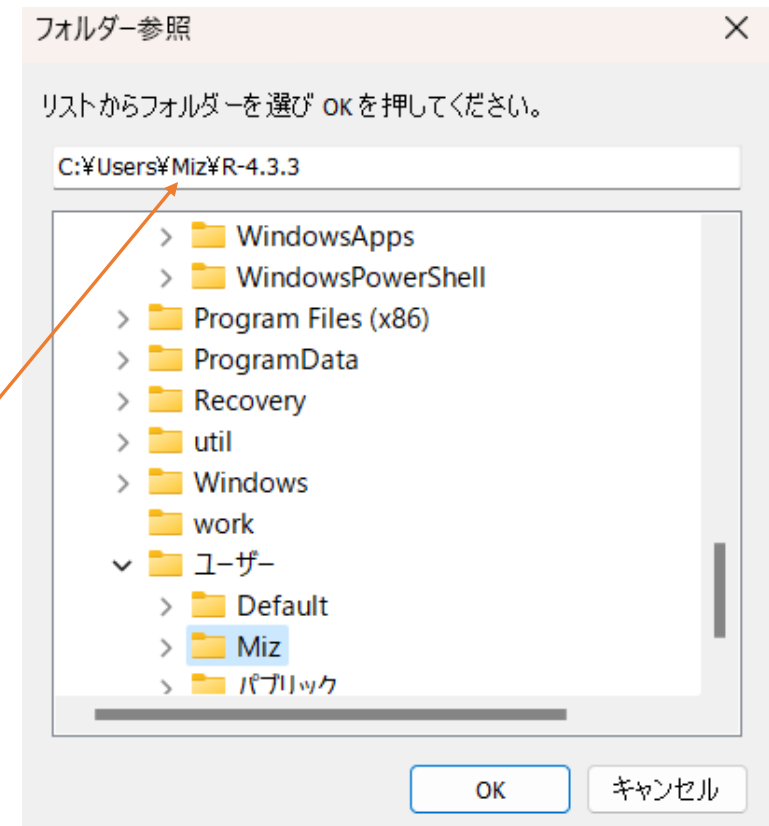
- デフォルトでは右のように表示されるので、「参照」からCドライブ直下のユーザーフォルダにインストールする。
- インストール先が以下のように表示されたらOK  
C:\Users\ユーザー名\R-4.3.3



# インストール先について注意

## 4. インストール先の指定

- 少し下に行くと「ユーザー」があるのでクリックするとユーザー名のフォルダがある。
- ユーザー名をクリックして、インストール先が以下のように表示されていたら「OK」を押す  
C:¥Users¥ユーザー名¥R-4.3.3
- そして「次へ」を押す






# インストールの手順

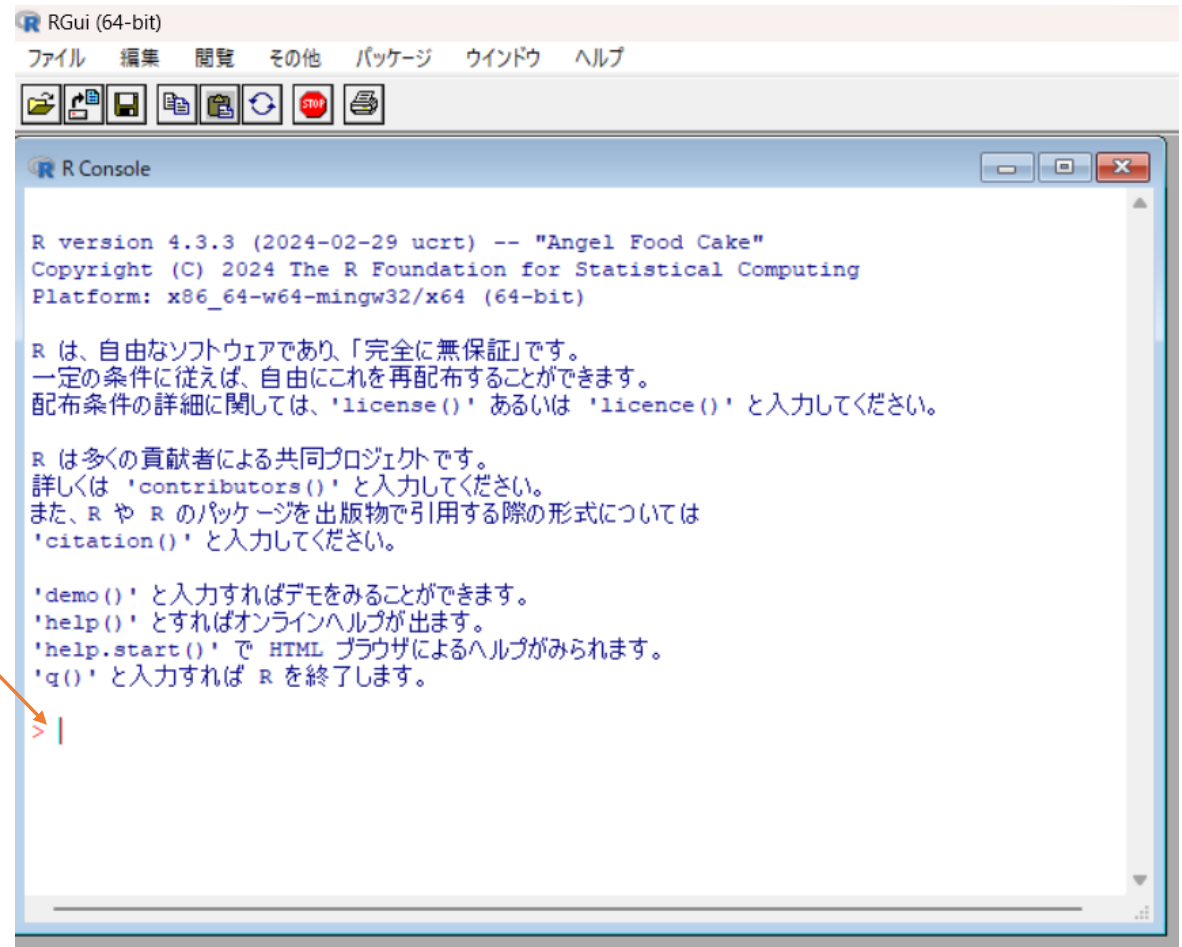
5. コンポーネントの選択は「利用者向けインストール」を選択して「次へ」。
6. 起動時オプションはそのまま「いいえ」で「次へ」。
7. スタートメニューフォルダーの指定は、そのまま「次へ」。
8. 追加タスクの選択は、そのまま「次へ」。  
ここでインストールが始まるので少し待つ。
9. セットアップウィザードの完了画面では「完了」。

# Rを操作

- 今後のため、スタートメニュー  からRアイコン上で右クリックで「スタートにピン留め」しておくて便利。
- スタートメニューからクリックして起動。
- Rが立ち上がり、RGuiの中にR Consoleが表示される。

# Rコンソール

- コマンド・プロンプト
- ここにコマンドを打ち込んで実行する。



```
RGui (64-bit)
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ
R Console
R version 4.3.3 (2024-02-29 ucrt) -- "Angel Food Cake"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R は、自由なソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()' あるいは 'licence()' と入力してください。

R は多くの貢献者による共同プロジェクトです。
詳しくは 'contributors()' と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
'citation()' と入力してください。

'demo()' と入力すればデモをみることができます。
'help()' とすればオンラインヘルプが出ます。
'help.start()' で HTML ブラウザによるヘルプがみられます。
'q()' と入力すれば R を終了します。

> |
```

# Rの操作

- R Consoleのコマンド・プロンプトに

5+3

と入力して「Enter」を押す。

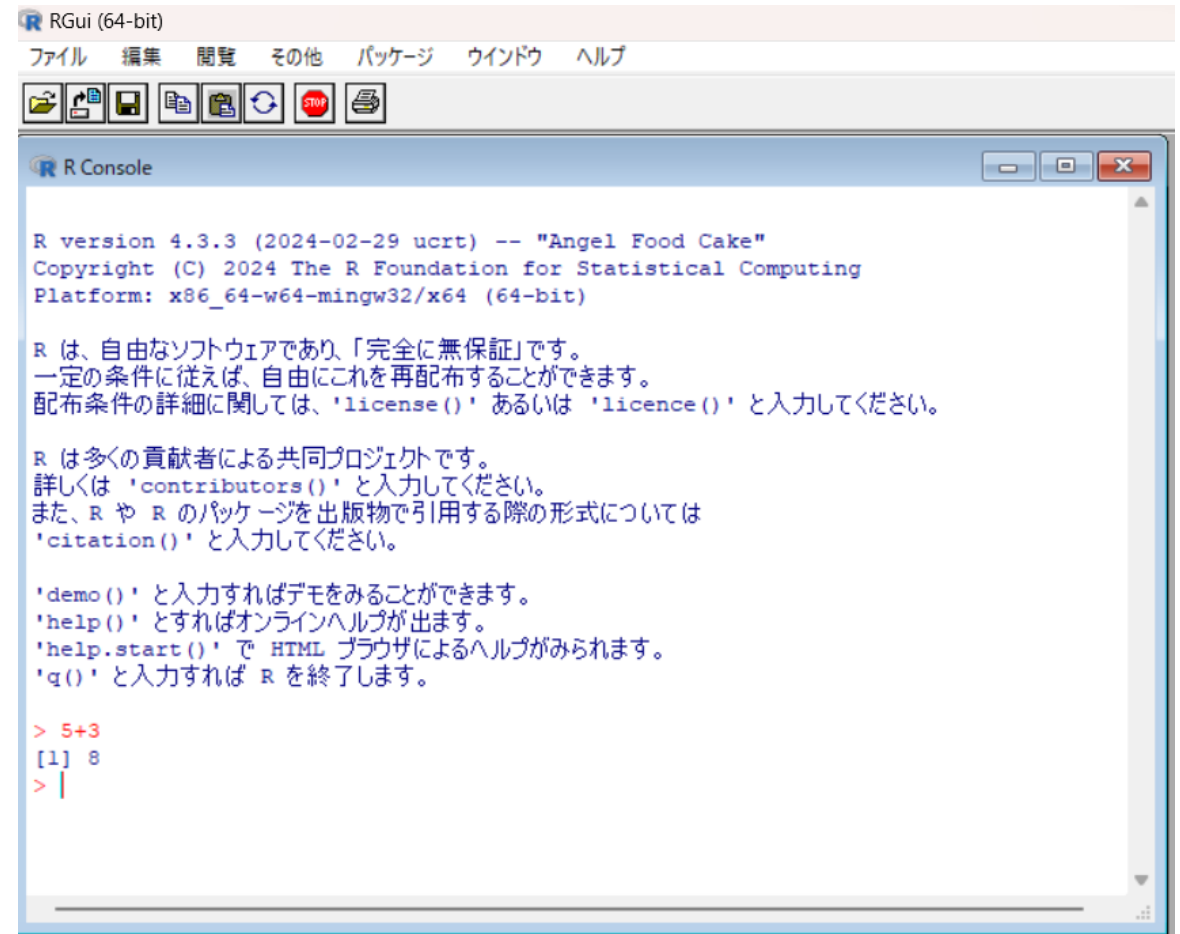
- その他の計算も試す。

21 - 36

3 × 14

12 ÷ 3

3<sup>5</sup>



```
RGui (64-bit)
ファイル 編集 閲覧 その他 パッケージ ウィンドウ ヘルプ

R Console

R version 4.3.3 (2024-02-29 ucrt) -- "Angel Food Cake"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R は、自由なソフトウェアであり、「完全に無保証」です。
一定の条件に従えば、自由にこれを再配布することができます。
配布条件の詳細に関しては、'license()' あるいは 'licence()' と入力してください。

R は多くの貢献者による共同プロジェクトです。
詳しくは 'contributors()' と入力してください。
また、R や R のパッケージを出版物で引用する際の形式については
'citation()' と入力してください。

'demo()' と入力すればデモをみることができます。
'help()' とすればオンラインヘルプが出ます。
'help.start()' で HTML ブラウザによるヘルプがみられます。
'q()' と入力すれば R を終了します。

> 5+3
[1] 8
> |
```

# Rの操作

- 自由度4の  $\chi^2$ 分布のグラフを描いてみる。
- 以下のように打ち込んで実行 (Enter) 。


```
curve(dchisq(x, df = 4), xlim = c(0, 15))
```

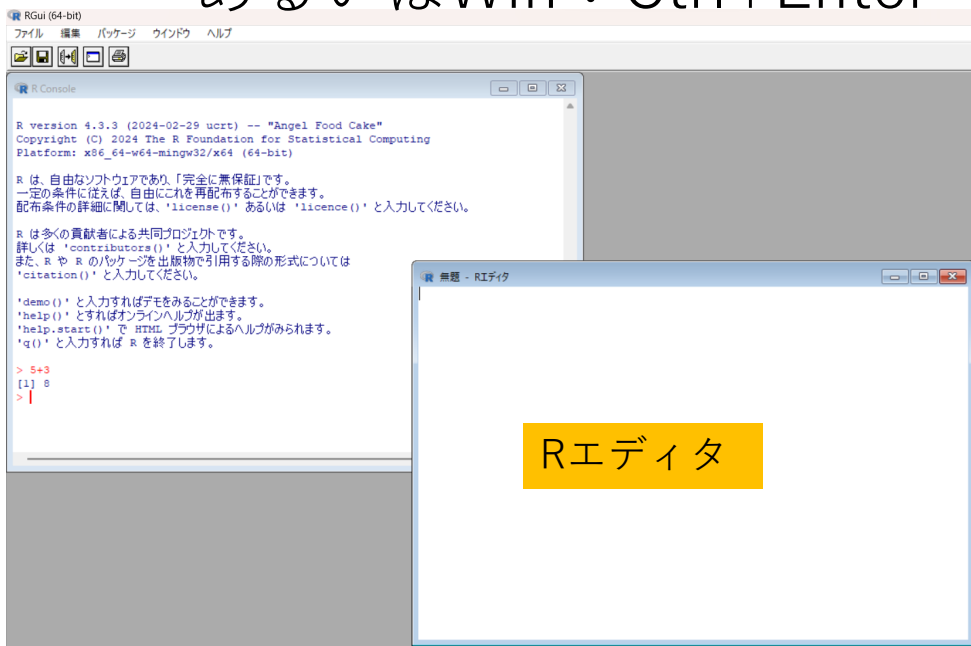
- 演算子 (= など) の前後や 「,」 の後は半角スペースを入れると見やすい。  
ただし、このスペースはなくても問題ない。

# Rの操作

- コンソールに入力して実行すると、プログラムを保存することができず、次回の作業時に初めから打ち直すことになる。
- そこで、Rエディタにプログラムを書きながら分析する。  
Win「ファイル」→「新しいスクリプト」  
Macの場合「ファイル」→「新規文書」  
でRエディタを開く。

# Rエディタにプログラムを書く

- 実行したいプログラムのある行にプロンプトを移動して  
実行ボタン  を押す (Win)  
あるいは Win : Ctrl+Enter



Mac : command+Enter



# Rエディタの利用

- Rエディタに以下のように打ち込む。
- 打ち込んだら範囲指定して、あるいは1行ずつ実行。

```
x <- c(2, 5, 6, 4, 1)
```

```
y <- c(10, 13, 16, 12, 9)
```

```
plot(x, y)
```

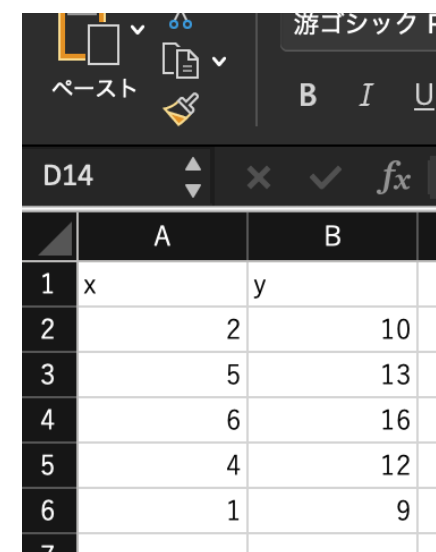


# Rの文法

- 「<」と「-」は別の文字として入力。「=」でもできる。
- ここでは複数の数値を変数x、yというベクトルに保存。
- plotは散布図を描く命令。plotのカッコ内は横軸（x軸）、縦軸（y軸）の順に記述するのがポイント。

# Excelとデータの関係

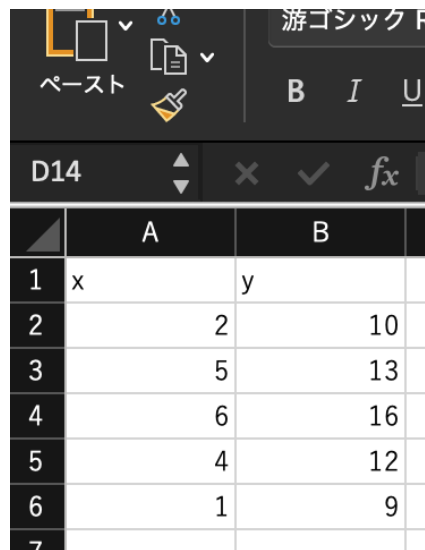
- Excelでのデータ分析の場合  
直接データを操作する
- Excelでの分析の問題点
  - 単純に面倒（細かい分析ができない）
  - データを加工した時に記録が残らない
  - 次回の作業時に、同じ分析手順を実行するのは難しい
  - 他の人に同じ作業手順を伝えるのは困難



	A	B
1	x	y
2	2	10
3	5	13
4	6	16
5	4	12
6	1	9
7		

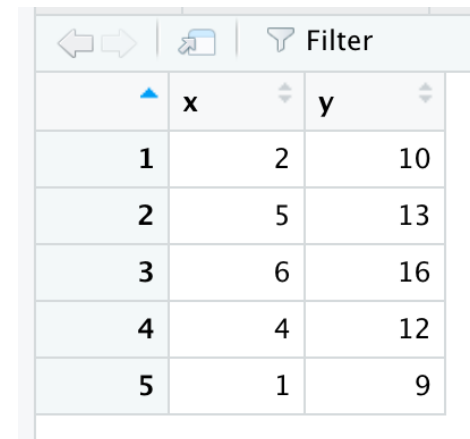
# Rとデータの関係

- Excelデータ（csvファイル）をRに読み込み、R上で操作する。
- Rでデータを加工してもExcelデータに影響なし。



	A	B
1	x	y
2	2	10
3	5	13
4	6	16
5	4	12
6	1	9
7		

CSVファイル



	x	y
1	2	10
2	5	13
3	6	16
4	4	12
5	1	9

データフレーム

## CSVファイルについて

- Rではcsvファイルが使われることが多い。

comma separated values/variables

余分な情報が入っていないので軽い。

テキストデータなので様々なソフトで開ける。

- Excelの見た目ではセルに分かれているが、実際には

2,10

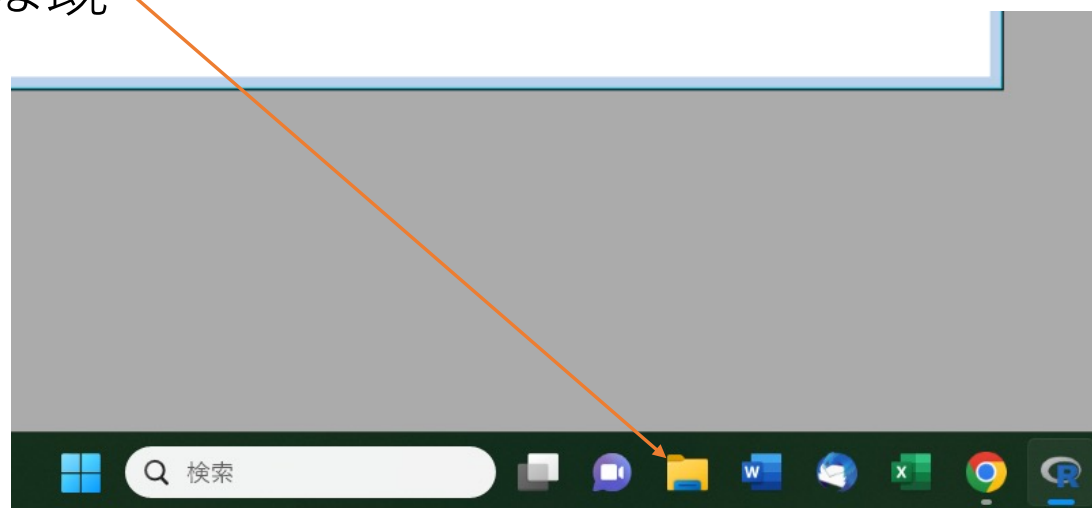
5,13

のようにデータが並んでいる。データにカンマがある時は注意。

# データ作成の前に

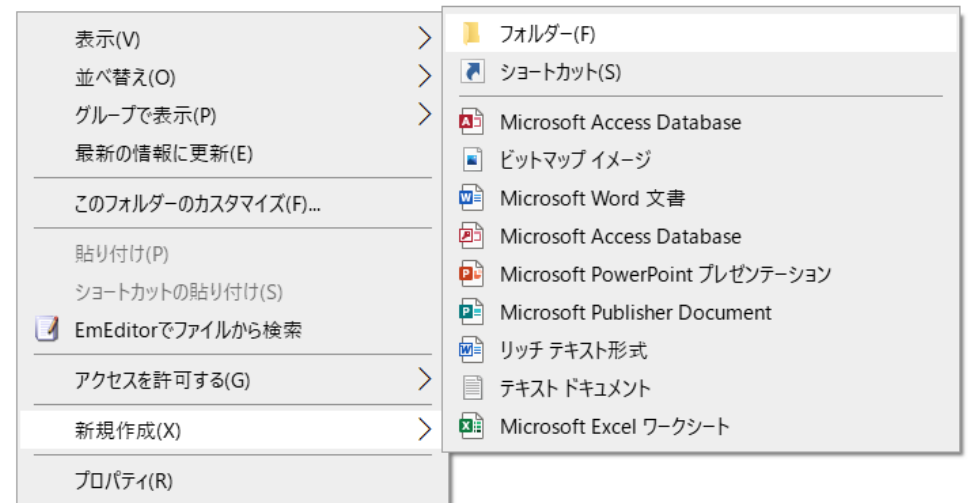
- エクスプローラーからユーザーフォルダを探す。

ユーザーフォルダ内には既にR-4.3.3があるはず。



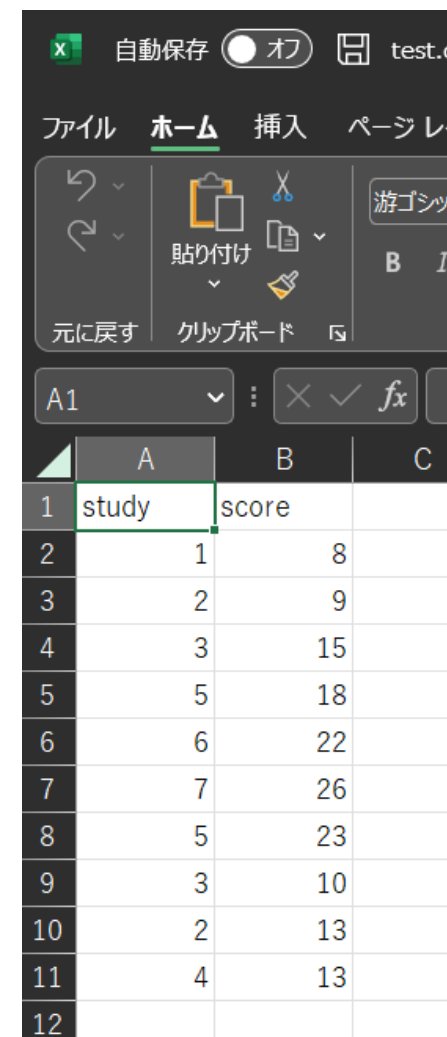
# フォルダの作成

- ユーザーフォルダ内で右クリックして「新規作成」から「フォルダ」を選択してフォルダ名を「Quant」とする。
- Macの人はデスクトップに「Quant」フォルダをつくっておく。




# 読み込みデータの作成

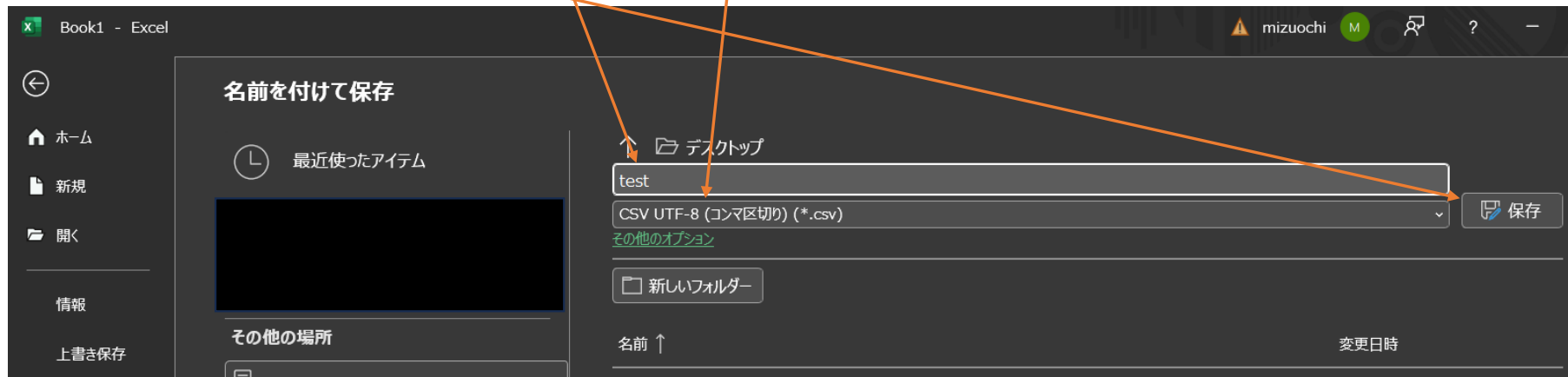
- Rエディタに打ち込んでデータを作成できるが、大きなデータになるとさすがに難しい。
- そこで外部ファイルとしてデータを作成し、読み込む。
- 新規にExcelを開き、右のように打ち込む。1行目が変数名、2～11行目がデータ。



	A	B	C
1	study	score	
2	1	8	
3	2	9	
4	3	15	
5	5	18	
6	6	22	
7	7	26	
8	5	23	
9	3	10	
10	2	13	
11	4	13	
12			

# データの保存 (Win)

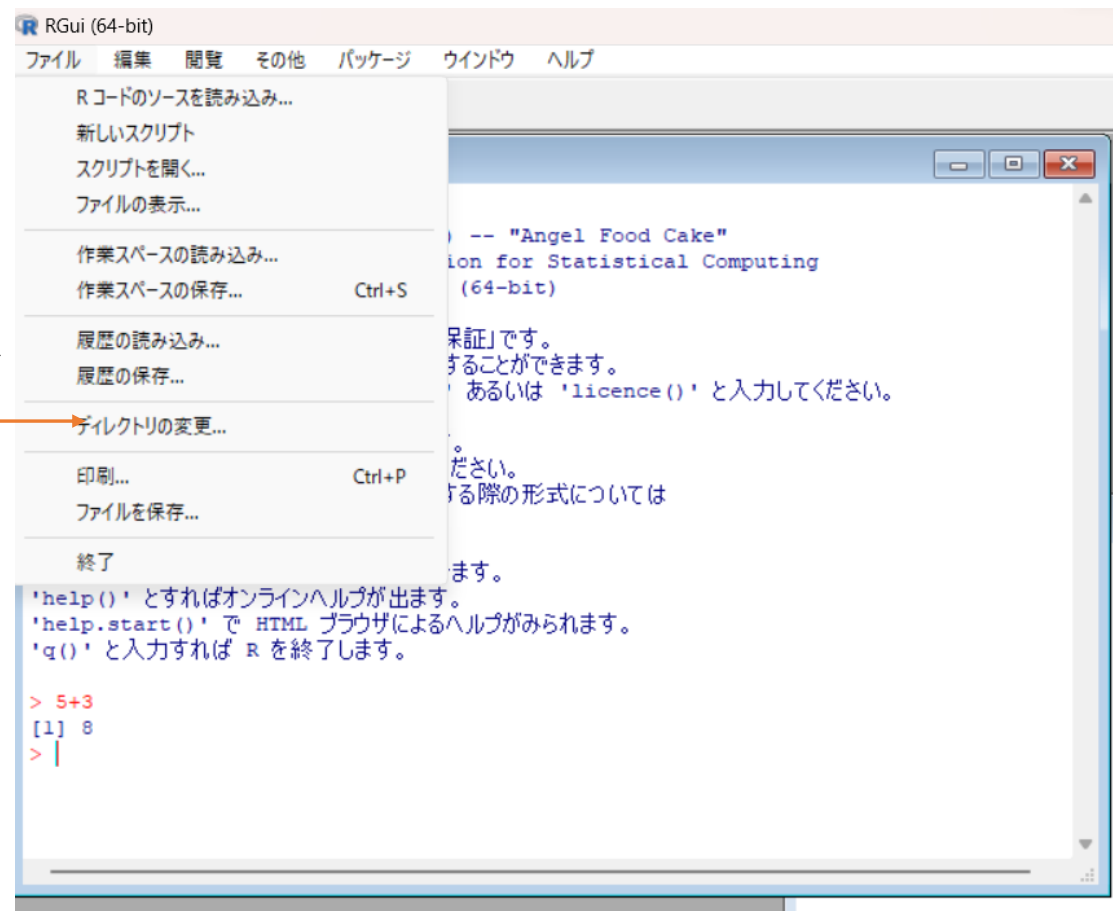
- とりあえずデスクトップに保存し、後でQnantフォルダに移す。
1. Excelで  ボタンを押し、以下の画面に。
  2. ファイル名を「test」に。
  3. 「ファイルの種類」からCSV UTF-8 (コンマ区切り) を指定。
  4. 「保存」を押す。





# 作業場所の指定

- R に作業場所が「Quant」フォルダであると指示する。起動のたびに指定する必要あり。
- R Console を **アクティブ** にして、「ファイル」→「ディレクトリの変更」→「Quant」フォルダを指定。
- これでデータの読み込みや保存等が「Quant」フォルダで行われることになる。
- Macは「その他」から。



# コマンドの実行

- 作業場所の指定が終わったら、R エディタに以下のようにコマンドし実行。

```
data1 <- read.csv("test.csv")
```

```
data1
```

- 1行目がデータ読み込みのコマンド。
- 2行目が読み込んだデータを表示するコマンド。

## コマンドの実行

- R Consoleに、読み込まれたデータが表示されている。
- csvファイルとしてはtestというファイル名だが、R上ではdata1という名前で扱われることに注意。
- このR上でのデータファイルのことをデータフレームと呼ぶ。

# 散布図の作成

- 散布図のコマンド

```
plot (data1$study, data1$score, xlab = "勉強時間", ylab = "得点")
```

xlab : x軸のラベル指定

ylab : y軸のラベル指定

# 作業の保存

- RエディタをアクティブにしてWin、Macとも「ファイル」→「保存」
- ファイル名をつけて「保存」を押す。ここではファイル名「prog」とする。
- 保存が終わったらRを終了（右上の×、左上の赤を押す）。
- 作業スペースの保存は「いいえ」を押す。

# 作業の再開

- Rを再び立ち上げる。
- ディレクトリの変更。 (毎回！)
- Win 「ファイル」 → 「スクリプトを開く」
- Mac 「ファイル」 → 「文書を開く」  
でprog.Rを指定。
- Rエディタが開くので、前回の作業を再開することができる。