

消費者心理学のための統計学 ～R入門

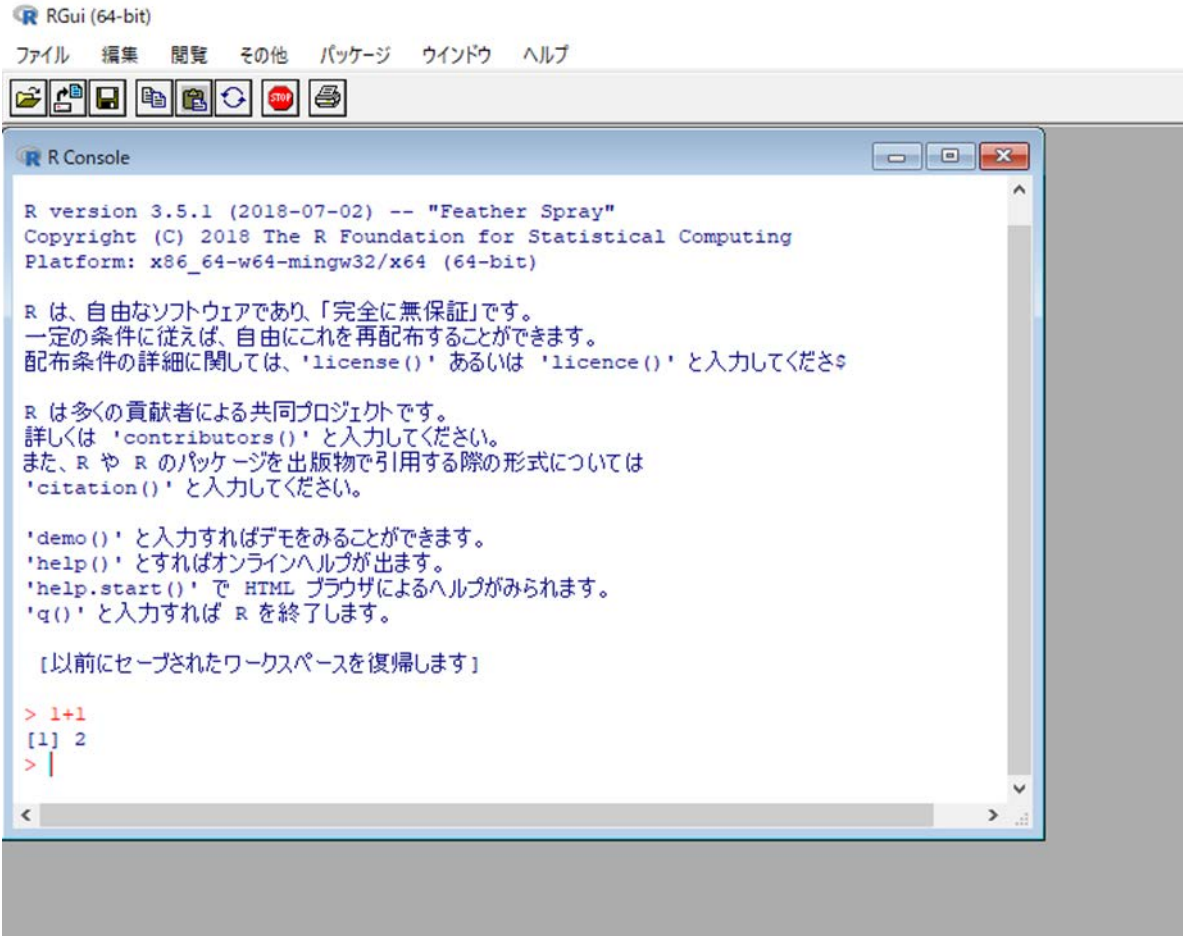
北九州市立大学経済学部

齋藤朗宏

大学入試センター研究開発部

荘島宏二郎

Rとは



```
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 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 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。

[以前にセーブされたワークスペースを復帰します]

> 1+1
[1] 2
> |
```

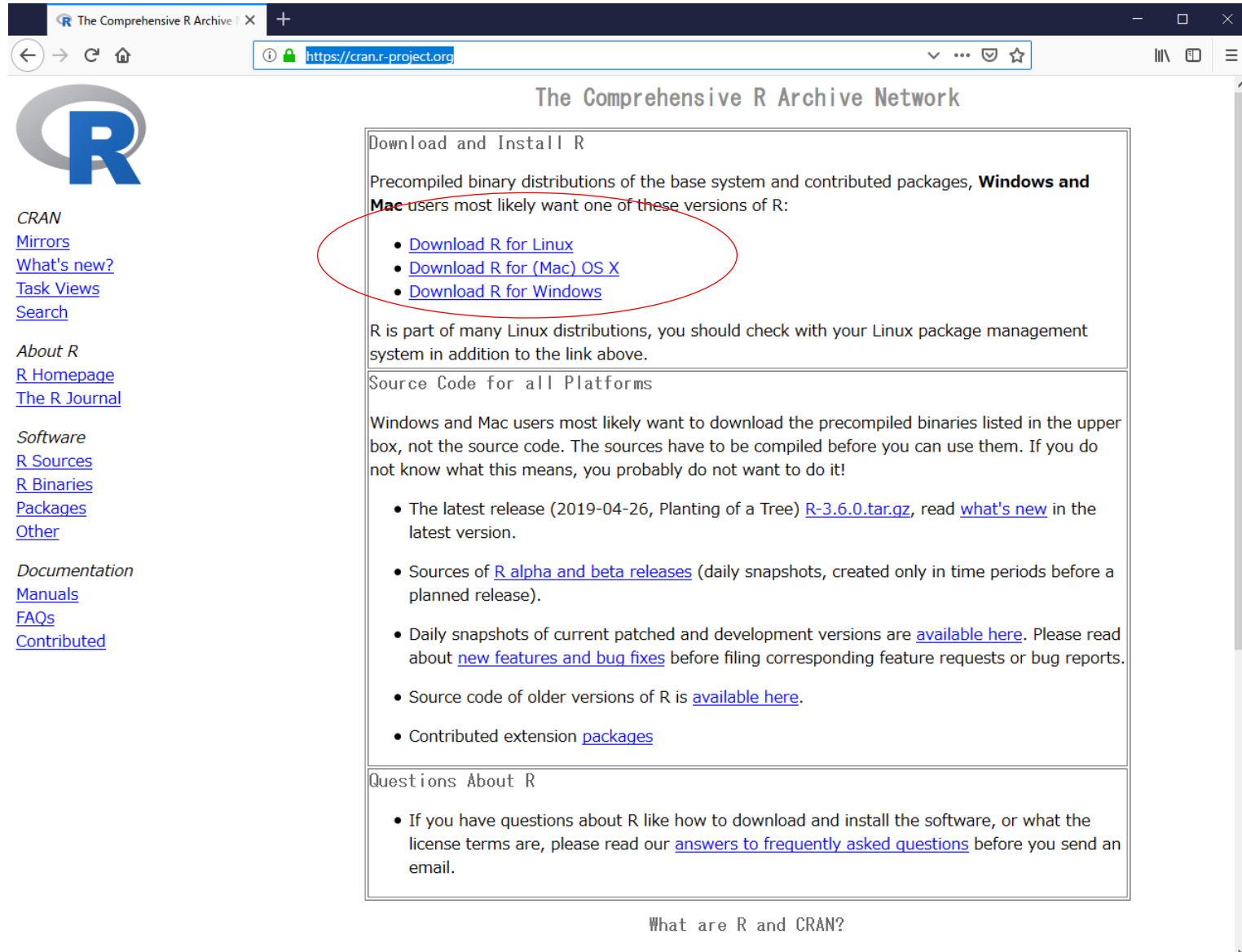
- プログラミング言語である「R言語」を扱うことができる、フリーかつオープンソースの統計解析向けソフトウェア。
- 1990年代後半にオークランド大学の Robert GentlemanとRoss Ihakaにより初期のソフトウェアが開発された。
- 現在では、R Projectという形を取り、共同で開発が進められている。

Rの利点

- Rというソフトウェア単体として考えても、関数電卓としての使用からプログラミングに至るまで、極めて多くの機能を持つ。
- 不足している機能は、パッケージをインストールすることで拡張できる(勿論これもフリー)。
- 近年使用人口が爆発的に増加し、書籍、情報を公開している書籍、ウェブページも多い。
- バージョンアップが頻繁で、バグの修正も早い。
- Windows, Mac OSX, Linuxで公開されており、プラットフォームを問わず使える。

- 基本的には「プログラム」であり, SPSS, Microsoft Excel等のスプレッドシート型の, GUIベースのソフトウェアに慣れ親しんだ人には, やや敷居が高い.
- 表出力はSPSS程綺麗で(見やすく)はない.

- <https://cran.r-project.org/>



The screenshot shows the CRAN website with the following content:

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

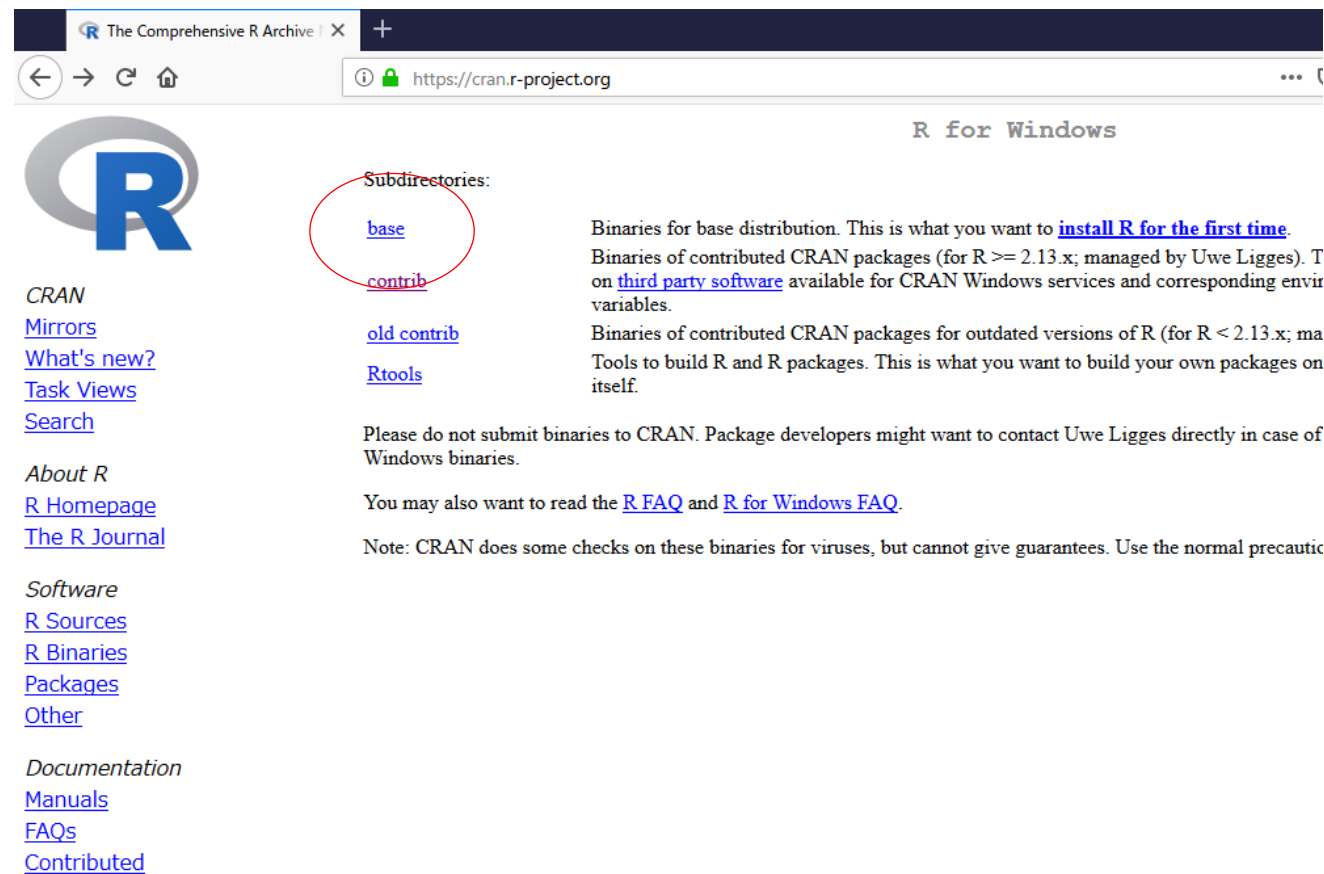
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2019-04-26, Planting of a Tree) [R-3.6.0.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.


What are R and CRAN?



The Comprehensive R Archive | X +

← → ↻ 🏠 ⓘ 🔒 https://cran.r-project.org ...

R for Windows



Subdirectories:

- [base](#) Binaries for base distribution. This is what you want to [install R for the first time](#).
- [contrib](#) Binaries of contributed CRAN packages (for R \geq 2.13.x; managed by Uwe Ligges). T on [third party software](#) available for CRAN Windows services and corresponding envir variables.
- [old contrib](#) Binaries of contributed CRAN packages for outdated versions of R (for R $<$ 2.13.x; ma
- [Rtools](#) Tools to build R and R packages. This is what you want to build your own packages on itself.

Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautic

CRAN

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

About R

- [R Homepage](#)
- [The R Journal](#)

Software

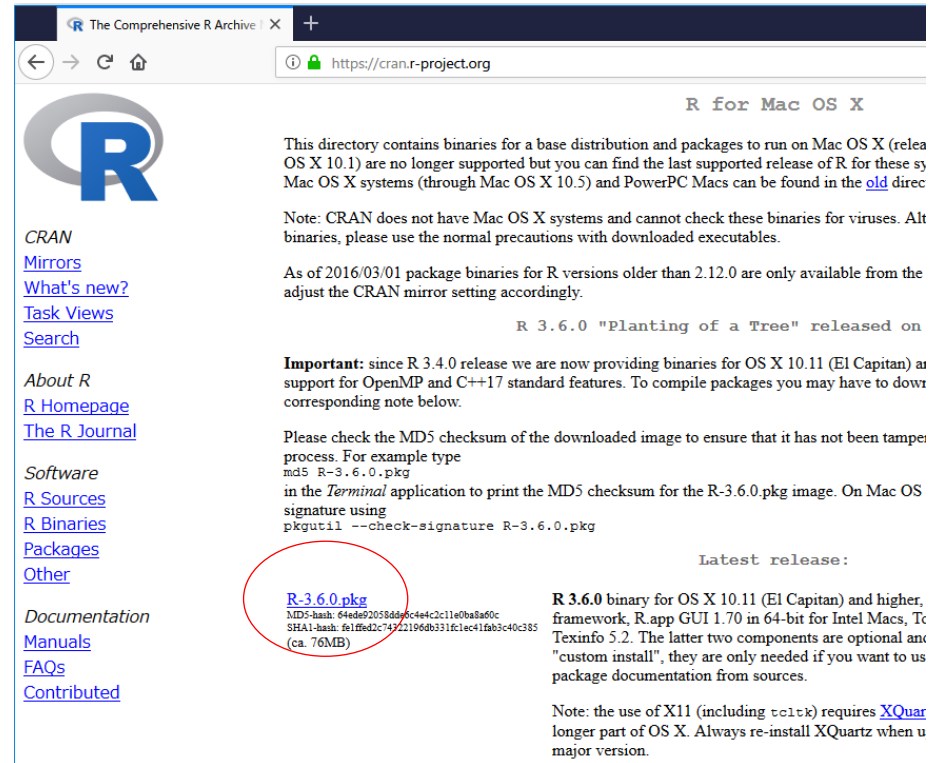
- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

Documentation

- [Manuals](#)
- [FAQs](#)
- [Contributed](#)

Rのインストール(Mac)

- わかる人はHomebrewなどからインストールすればいい。わからないのであれば、以下を使用。



The screenshot shows a web browser window with the address bar displaying `https://cran.r-project.org`. The page title is "R for Mac OS X". The main content area contains the following text:

This directory contains binaries for a base distribution and packages to run on Mac OS X (relea OS X 10.1) are no longer supported but you can find the last supported release of R for these sy Mac OS X systems (through Mac OS X 10.5) and PowerPC Macs can be found in the [old direc](#)

Note: CRAN does not have Mac OS X systems and cannot check these binaries for viruses. Alt binaries, please use the normal precautions with downloaded executables.

As of 2016/03/01 package binaries for R versions older than 2.12.0 are only available from the adjust the CRAN mirror setting accordingly.

R 3.6.0 "Planting of a Tree" released on

Important: since R 3.4.0 release we are now providing binaries for OS X 10.11 (El Capitan) a support for OpenMP and C++17 standard features. To compile packages you may have to down corresponding note below.

Please check the MD5 checksum of the downloaded image to ensure that it has not been tamper process. For example type

```
md5 R-3.6.0.pkg
```

in the *Terminal* application to print the MD5 checksum for the R-3.6.0.pkg image. On Mac OS signature using

```
pkgutil --check-signature R-3.6.0.pkg
```

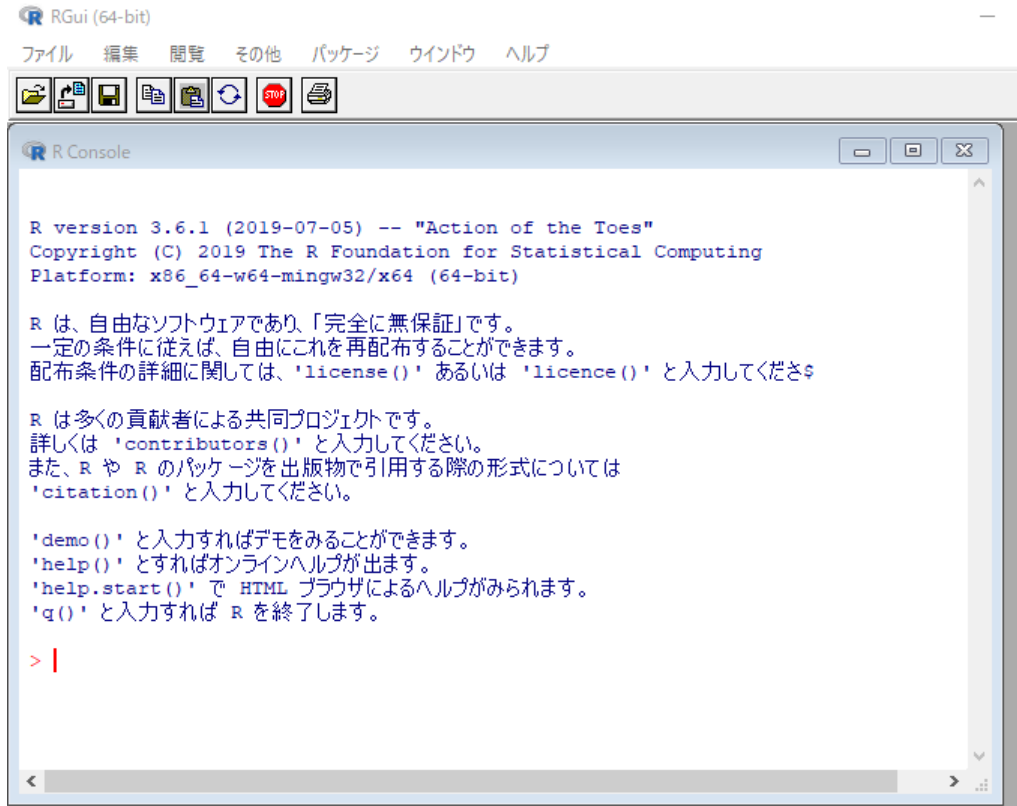
Latest release:

R 3.6.0 binary for OS X 10.11 (El Capitan) and higher, framework, R.app GUI 1.70 in 64-bit for Intel Macs, Te Texinfo 5.2. The latter two components are optional and "custom install", they are only needed if you want to us package documentation from sources.

Note: the use of X11 (including `tccltk`) requires [XQuar](#) longer part of OS X. Always re-install XQuartz when u major version.

On the left side of the page, there is a navigation menu with links: CRAN, Mirrors, What's new?, Task Views, Search, About R, R Homepage, The R Journal, Software, R Sources, R Binaries, Packages, Other, Documentation, Manuals, FAQs, Contributed.

The link "R-3.6.0.pkg" is circled in red in the original image.



```
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 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 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をインストールすると、Windowsの場合i386とx64と2種類の起動ファイルがある。取り敢えずx64で起動してみて、問題があるようならi386を実行すれば良い。
- >の部分にコマンドを入力することでプログラムを実行できる。

+	加算
-	減算
*	乗算
/	除算
^	べき乗

1+2

4-1

3*2

10/5

2^3

- 基本的な演算子は Excelと同じ.
- 式を入力してEnterで 次の行に結果が表示

```
R Console

R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 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

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。

> 1+2
[1] 3
> |
```

変数

- プログラミング言語における変数(Variable)とは, 数値や文字などが入った箱を指す(正確には, 変数そのものは箱に付けた名札を意味する. 統計学における変数と多少意味が違うので注意).
- 変数には, 数値, 文字などや, それらを組合せたものを保存できる.
- 名前を付けておくことで, 後で箱から値を取り出して計算などに使うことができる.

変数への値の代入

- 文字列や数値, 計算式の結果などの値を箱に入れる操作.
- `<-`もしくは`=`で代入操作を表し、`<-`などの右側に保存する値, 左に変数名を書く.

```
a<- 1  
a = 1
```

- `a`という変数に1という値を代入した. この変数は計算式の中でも用いることができる.

```
a + 2
```

- 代入した値にもう一度値を代入し直すこともできる.
- 以下を入力して実行.

```
a <- 2
a
a <- a + 1
a
```

- 3行目でaの値が置き換わっているため, 1回目と2回目で異なる値が表示されている.
- 「`a <- a + 1`」というのはプログラミングでは非常によく出てくる表記. それまでのaの値に1を足して, 新しいaの値にするという操作.

- A店の15日分の売り上げ.
- 75,82,63,52,89,73,64,75,85,70,49,85,69,68,70
- この合計値を求めるには？

```
a1<-75
a2<-82
a3<-63
(中略)
a15<-70
a1+a2+a3+(中略)+a15
```

- 1つの変数に1つの値を当てはめる方法では大変.
- ベクトルという一連の値を使う.
- 変数代入と同じ方法で、ベクトル名<-c(カンマで区切られた一連の値)とすればよい.

```
shopa <- c(75,82,63,52,89,73,64,75,85,70,49,85,69,68,70)
shopa
```

- 関数名(引数)で関数を実行可能.
- 例:関数sqrt()は, 括弧の中身の平方根を求める関数.

```
sqrt(2)
a <- 2
sqrt(a)
```

- Excelで用いたような関数にも概ね対応する関数が存在する.

max()	最大値	max(shopa)
min()	最小値	min(shopa)
sum()	合計値	sum(shopa)
mean()	平均値	mean(shopa)
median()	中央値	median(shopa)
var()	分散	var(shopa)
sd()	標準偏差	sd(shopa)

- 2変数に対して用いられる関数も同様である.

```
shopb <- c(77,87,59,56,52,56,76,90,87,79,81,70,76,70,81)
var(shopa, shopb)
cor(shopa , shopb)
```

var()
cor()

分散(ここに2つの変数を与えると共分散)
相関係数

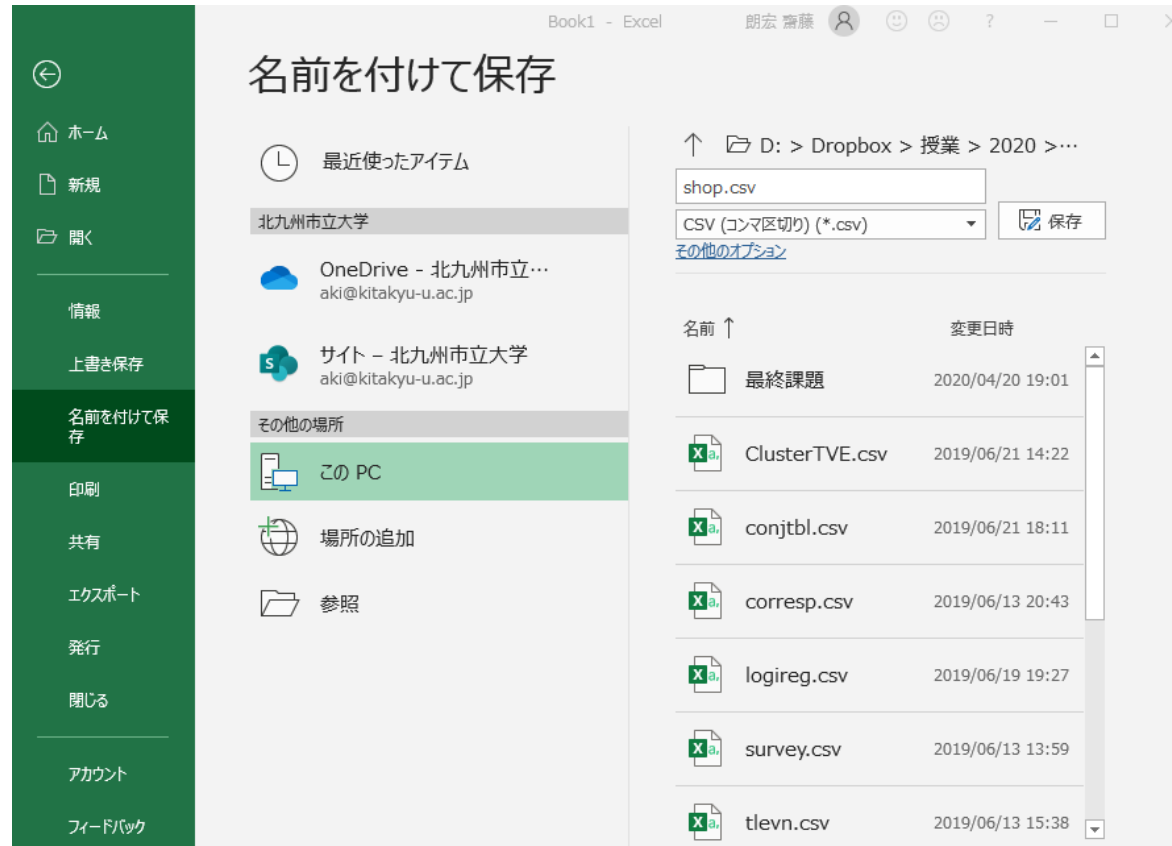
- Rでデータを読み込ませる方法はいくつかあるが, ワーキングディレクトリ内(ソフトウェア上で指定されている作業用フォルダ)にファイルを置いて読み込ませるのが簡単.
- 作業用フォルダの指定は, `getwd()`の表示を参考に.
- 何も指定しなければ, ドキュメントフォルダが指定されている.

`getwd()`
`setwd()`

作業用フォルダの位置を確認
作業用フォルダを指定

```
> var(shopa, shopb)
[1] 14.74762
> cor(shopa, shopb)
[1] 0.104602
> getwd()
[1] "C:/Users/aki/Documents"
> |
```

<



- csvファイルの保存.
- MacOSなどでは UTF-8が一般的だが, UTF-8ではないコンマ区切りで保存することを推奨する.
- このcsvファイルをワーキングディレクトリ内に置く.

```
v08c0.dat <- read.csv("v08c0.csv", fileEncoding="Shift_JIS")
v08c0.dat
```

- `read.csv("ファイル名")`でワーキングディレクトリ内にある当該ファイル名のファイルを読み込む。
- `fileEncoding=`以降はWindowsでは不要だが、MacOSなどの場合必要。
- Excelのcsvファイル書き出しでUTF-8を用いると、再度Excelで開く際に日本語部分が文字化けを起こす可能性があるため、Shift_JISで統一している(そのためファイル保存時にUTF-8を避けている)。

	A店の売り上げ	B店の売り上げ
1	75	77
2	82	87
3	63	59
4	52	56
5	89	52
6	73	56
7	64	76
8	75	90
9	85	87
10	70	79
11	49	81
12	85	70
13	69	76
14	68	70
15	70	81

```
v08c0.dat  
v08c0.dat[8,2]  
v08c0.dat[1,]  
v08c0.dat[,1]  
v08c0.dat[,]
```

- read.csv()ではリスト形式でファイルを読み込む.
- リストではリスト名[行番号,列番号]のようにすることで特定の位置の値を取り出せる.
- 行番号も列番号も省略できる. 省略した場合, 行全体, もしくは列全体が取り出される.

	A店の売り上げ	B店の売り上げ
1	75	77
2	82	87
3	63	59
4	52	56
5	89	52
6	73	56
7	64	76
8	75	90
9	85	87
10	70	79
11	49	81
12	85	70
13	69	76
14	68	70
15	70	81

```
sum(v08c0.dat)
sum(v08c0.dat[,1])
mean(v08c0.dat[,1])
median(v08c0.dat[,1])
var(v08c0.dat[,1])
sd(v08c0.dat[,1])
var(v08c0.dat)
cor(v08c0.dat)
```

- `sum(リスト)`ではリスト全体が合計される。たとえばA店の売り上げの合計が出したいのであれば、`[,1]`のように指定することで、1列目全体の合計とできる。`mean()`なども同じように指定できる。
- `var()`と`cor()`はリストを指定すると、リスト内にあるすべての変数の間での共分散、相関を求めることができる。

- グラフには様々な作成の方法があるが, `plot()`を用いるのが基本. `plot()`では, 与えられたデータからある程度自然に判断してグラフを作成してくれる.

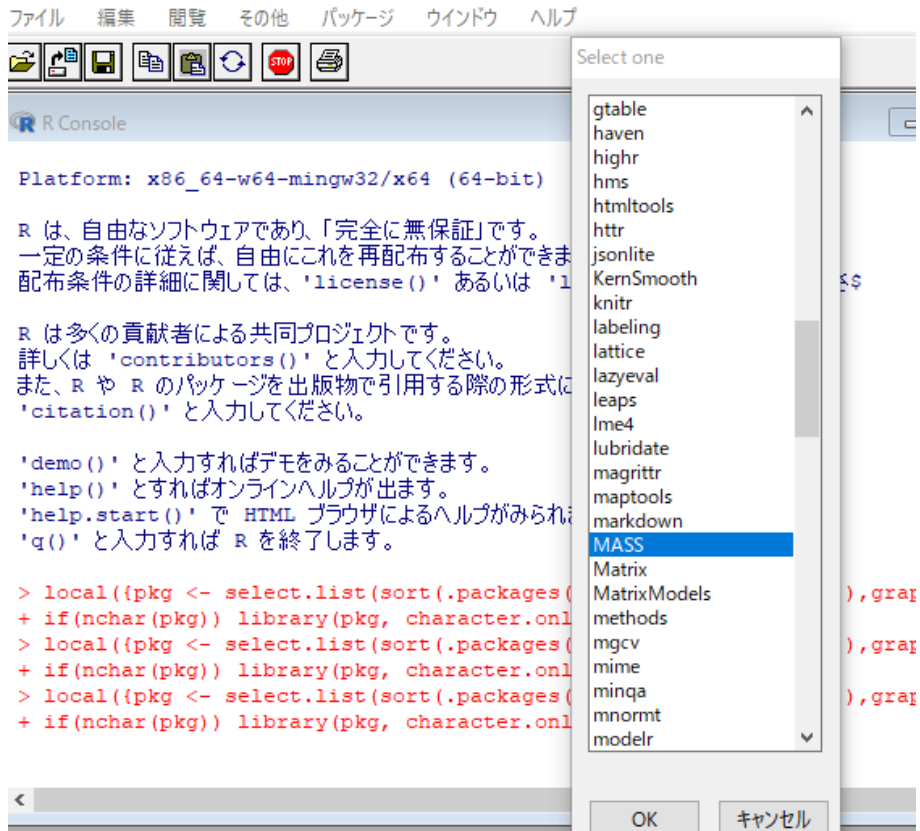
```
plot(v08c0.dat)
```

- ここでは, 2つの変数が与えられているので, 散布図が描画される.
- `plot()`の中で, `type=`を指定することで形式を指定することもできる.

```
plot(v08c0.dat[,1], type = "l")
```

- MacOSでグラフの日本語が表示できない場合. 以下を実行した後で改めてグラフを描くとよい.

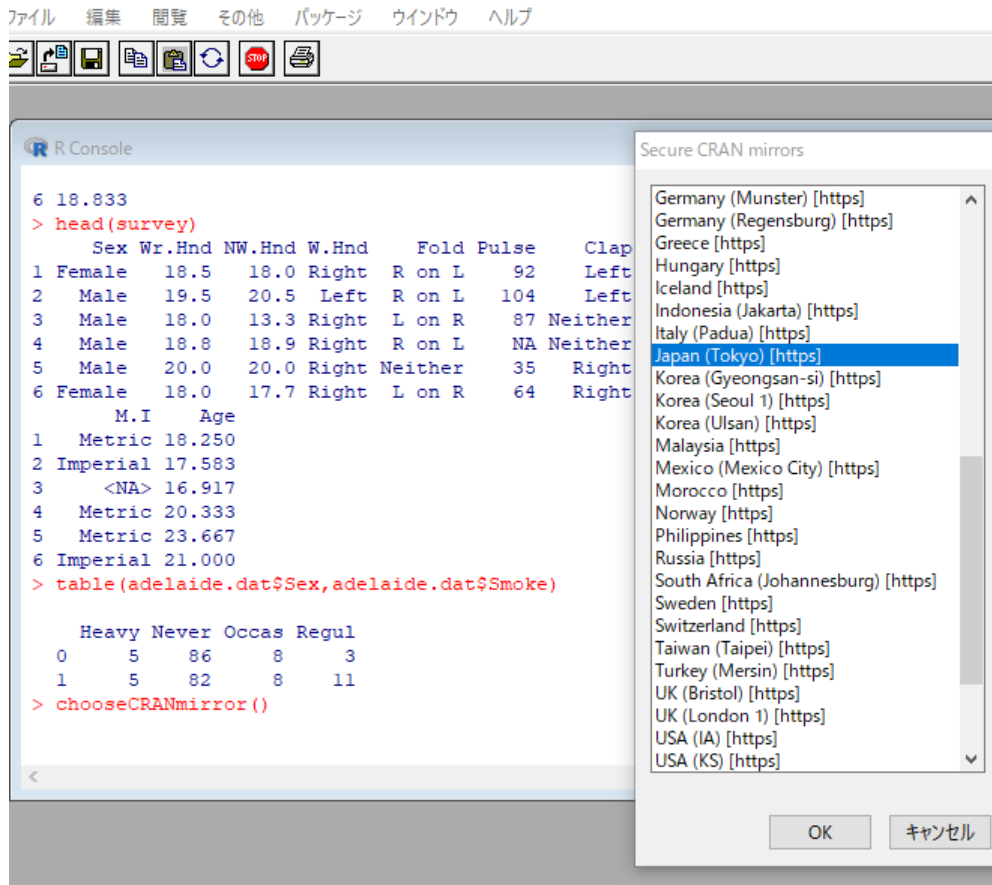
```
par(family = "HiraKakuProN-W3")
```



- Rでは、追加機能を「**パッケージ**」として読み込むことで、ソフトに様々な分析手法や機能を追加できる。
- パッケージを読み込む方法は以下の2通り。
 - メニューから**パッケージ**→**パッケージの読み込み**で読み込むパッケージを選択する。
 - コマンド**library(“パッケージ名”)**を実行する。
- まず、最初からインストールされている「**MASS**」パッケージを読み込む。

library("MASS")
でも同じ

パッケージのインストール



The screenshot shows the R Studio interface. The R Console on the left displays the following commands and output:

```
6 18.833
> head(survey)
  Sex Wr.Hnd NW.Hnd W.Hnd  Fold Pulse  Clap
1 Female 18.5 18.0 Right R on L  92  Left
2 Male 19.5 20.5 Left R on L 104  Left
3 Male 18.0 13.3 Right L on R  87 Neither
4 Male 18.8 18.9 Right R on L  NA Neither
5 Male 20.0 20.0 Right Neither 35  Right
6 Female 18.0 17.7 Right L on R  64  Right

  M.I  Age
1 Metric 18.250
2 Imperial 17.583
3 <NA> 16.917
4 Metric 20.333
5 Metric 23.667
6 Imperial 21.000

> table(adelaide.dat$Sex, adelaide.dat$Smoke)

  Heavy Never Occas Regul
0     5     86     8     3
1     5     82     8    11

> chooseCRANmirror()
```

The 'Secure CRAN mirrors' dialog box is open, showing a list of mirrors. 'Japan (Tokyo) [https]' is selected.

Country	URL
Germany (Munster)	[https]
Germany (Regensburg)	[https]
Greece	[https]
Hungary	[https]
Iceland	[https]
Indonesia (Jakarta)	[https]
Italy (Padua)	[https]
Japan (Tokyo)	[https]
Korea (Gyeongsan-si)	[https]
Korea (Seoul 1)	[https]
Korea (Ulsan)	[https]
Malaysia	[https]
Mexico (Mexico City)	[https]
Morocco	[https]
Norway	[https]
Philippines	[https]
Russia	[https]
South Africa (Johannesburg)	[https]
Sweden	[https]
Switzerland	[https]
Taiwan (Taipei)	[https]
Turkey (Mersin)	[https]
UK (Bristol)	[https]
UK (London 1)	[https]
USA (IA)	[https]
USA (KS)	[https]

- 最初からRにインストールされているパッケージは限られており, それ以外はインターネットからダウンロードするなどしてインストールする必要がある。
- まず, **パッケージ**→**CRANミラーサイトの設定**で日本国内のサイトを指定した上で, **パッケージ**→**パッケージのインストール**もしくは `install.packages("パッケージ名")`をコマンド実行でインストール出来る。

アデレード大学の調査データ

A	B	C	D	E	F	G	H	I	J	K	L
Sex	Wr.Hnd	NW.Hnd	W.Hnd	Fold	Pulse	Clap	Exer	Smoke	Height	M.I	Age
0	18.5	18	Right	R on L	92	Left	Some	Never	173	Metric	18.25
1	19.5	20.5	Left	R on L	104	Left	None	Regul	177.8	Imperial	17.583
1	18.8	18.9	Right	R on L	NA	Neither	None	Never	160	Metric	20.333
1	20	20	Right	Neither	35	Right	Some	Never	165	Metric	23.667
0	18	17.7	Right	L on R	64	Right	Some	Never	172.72	Imperial	21
1	17.7	17.7	Right	L on R	83	Right	Freq	Never	182.88	Imperial	18.833
0	17	17.3	Right	R on L	74	Right	Freq	Never	157	Metric	35.833
1	20	19.5	Right	R on L	72	Right	Some	Never	175	Metric	19
1	18.5	18.5	Right	R on L	90	Right	Some	Never	167	Metric	22.333
0	17	17.2	Right	L on R	80	Right	Freq	Never	156.2	Imperial	28.5
0	16	16	Right	L on R	NA	Right	Some	Never	155	Metric	18.75
0	19.5	20.2	Right	L on R	66	Neither	Some	Never	155	Metric	17.5
0	17.5	17	Right	R on L	NA	Right	Freq	Never	156	Metric	17.167
0	18	18	Right	L on R	89	Neither	Freq	Never	157	Metric	19.333
1	19.4	19.2	Left	R on L	74	Right	Some	Never	182.88	Imperial	18.333
1	20.5	20.5	Right	L on R	NA	Left	Some	Never	190.5	Imperial	19.75

- アデレード大学で実施された各種調査のデータ。
- ここから、各種集計などを試みる。
- MASSパッケージを読み込んだ状態だと、以下のコマンドでデータのヘルプが見られる。

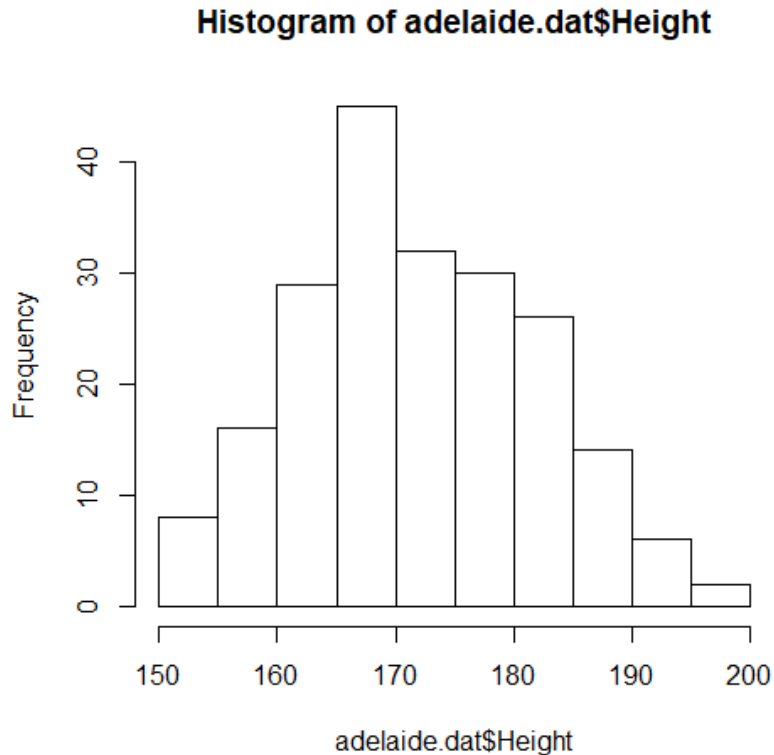
?survey


```
head(survey)
# 平均値を求める(方法1)
mean(survey[,10], na.rm = TRUE)
# 平均値を求める(方法2)
mean(survey $Height, na.rm = TRUE)
```

#はコメントアウトと言い、その行の#以降はコメントとして無視される。よって、この#平均値を求める(方法1)という部分もまとめて実行しても問題ない。

- `head(データセット名)`で、そのデータの最初6行を表示する。`tail(データセット名)`なら最後の6行を確認出来る。
- このデータからたとえば身長のみを取り出して平均値を求めたい場合。
- `mean(survey[,10], na.rm = TRUE)`
 - データセット「survey」から、10列目を取り出して平均値を求める。
- `mean(survey$Height, na.rm = TRUE)`
 - データセット「adelaide.dat」の中にある変数Heightを取り出して平均値を求める。
- `na.rm = TRUE`とすることで、欠測値(NA)を取り除いて平均値を求める。これをTRUEに設定しないと計算できない。

ヒストグラムの作成



- この図のように、ある範囲に、そこに該当する人数が何人いるかまとめてグラフ化したものをヒストグラムと呼ぶ。
- たとえば、150～155の間には9人くらい該当する学生がいることがわかる。
- Rでは、`hist(変数名)`とすることで、自動的に最適に調整したヒストグラムを作成してくれる。

```
hist(survey$Height)
```

- そのほかの細かい調整法などが知りたければ、`?hist`でヘルプを確認すること。
- 一例としては、以下のように`breaks = c()`でヒストグラムの区切り位置を設定できる。

```
hist(survey$Height, breaks = c(150,160,170,180,190,200))
```

```
> table(survey$Sex, survey$Smoke)
```

```
      Heavy Never Occas Regul  
Female    5    99     9     5  
Male     6    89    10    12
```

```
table(adelaide.dat$Sex,adelaide.dat$Smoke)
```

- 男性の喫煙者, 女性の喫煙者, 男性の非喫煙者, 女性の非喫煙者のように分類して集計したい場合には, **クロス集計表(分割表)**を作成する.
- Rでは**table(変数1,変数2)**とすれば作成できる.
- 男女ほぼ同数なので, ヘビースモーカーなどはあまり差がないが, 日常的くらいの吸い方の方は男性の方が多いことが確認できる.