R Commander を用いた統計解析の基礎(1)

岡田　昌史

1.　R のインストールと環境設定

フリーソフトウェアである統計解析環境，「R」をまずは皆さんがお使いのパソコンにインストールしましょう。以下のインストール方法については Windows バージョンで説明しています。Macintosh をお使いの場合は，インストールはもっと簡単なのでとくに説明は不要でしょう;-)

The Comprehensive R Archive Network

Frequently used pages

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:

- Linux
- MacOS X
- Windows (95 and later)

Source Code for all Platforms

R は基本的な機能が搭載された「本体」と、追加機能を収録した「パッケージ」からなりますから、まずは「本体」をインストールしましょう。R が配布されている Web サイト，http://cran.md.tsukuba.ac.jp/ に行くと上のようなページが出てくると思います．ここから，「Windows (95 and later) 」のリンクをクリックします．

R for Windows

This directory contains binaries for a base distribution and packages to run on Windows (NT, 95 and later) on Intel and
NT on Alpha and other platforms).

Note: CRAN does not have Windows systems and cannot check these binaries for viruses. Use the normal precautions
executables.

Subdirectories:

base	Binaries for base distribution (managed by Duncan Murdoch)
contrib	Binaries of contributed packages (managed by Uwe Ligges)

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges dire

すると，「base」「contrib」というリンクがあります．R 本体は「base」のリンクの先にあります．

R-2.5.1 for Windows

This directory contains a Windows binary distribution of R-2.5.1.

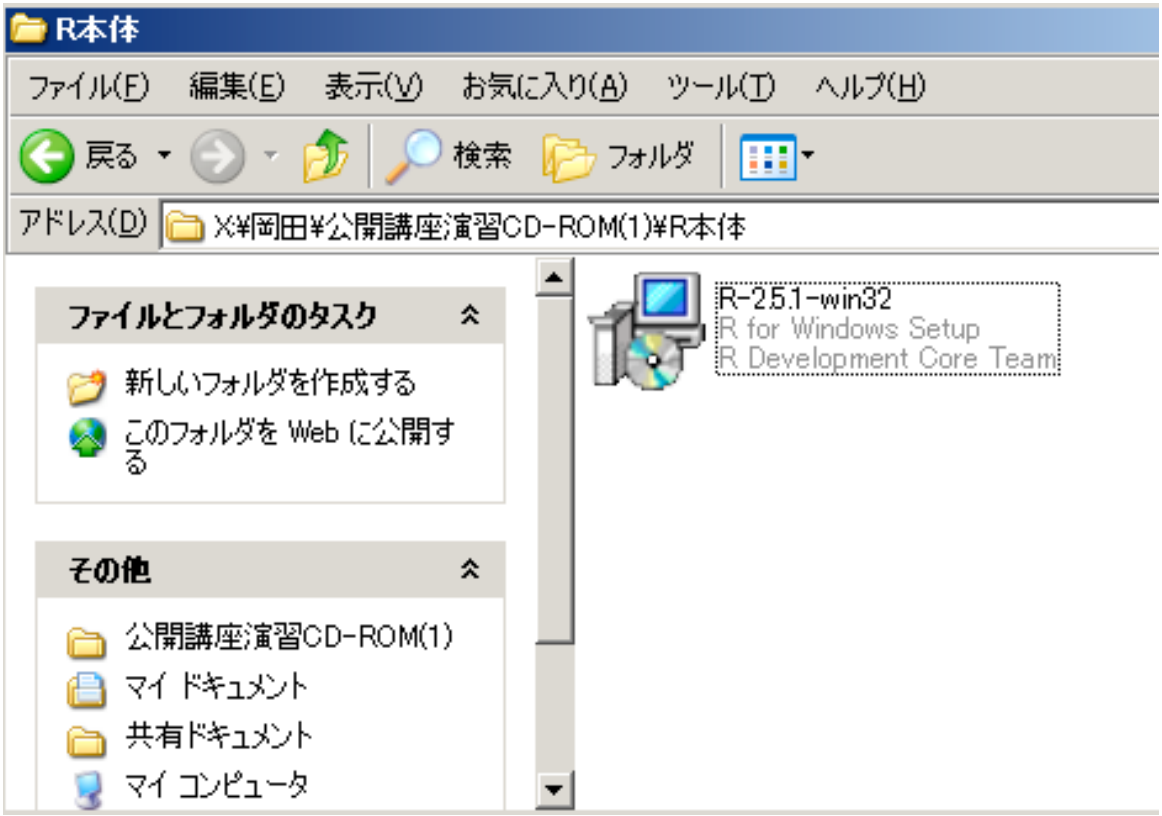
Patches to this release are incorporated in the r-patched snapshot build.

A build of the development version (which will eventually become the next major release of R) is available in the r-devel s

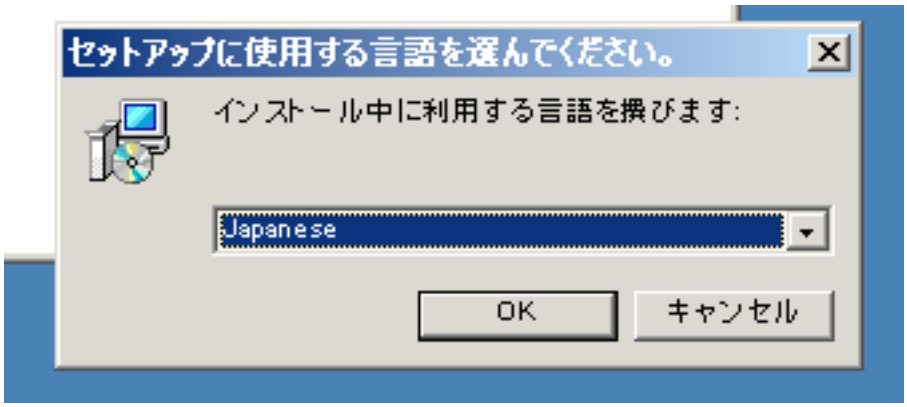
In this directory:

README.R-2.5.1	Installation and other instructions.
CHANGES	New features of this Windows version.
NEWS	New features of all versions.
R-2.5.1-win32.exe	Setup program (about 29 megabytes). Please download this from a mirror near you.
old	Previous releases.
md5sum.txt	md5sum output for the setup program. A Windows GUI version of md5sum is available at http://www.md5summer.org/; a Windows command line version is available at http://www.etree.org/md5com.html.

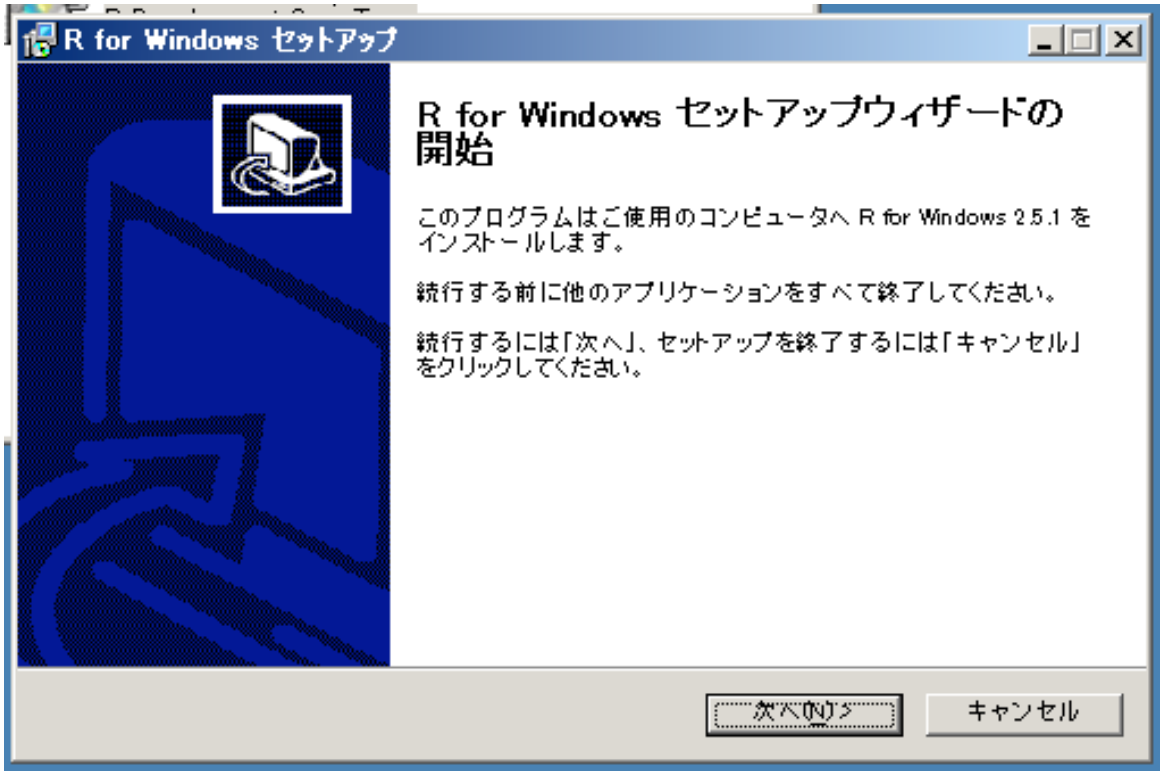
アクセスした時点での最新版によって少々名前は異なるかもしれませんが，「R-2.5.1-win32.exe」のようなリンクがＲ本体です．こちらをダウンロードします．ダウンロードされたファイルは以下のようなアイコンになっているはずです．



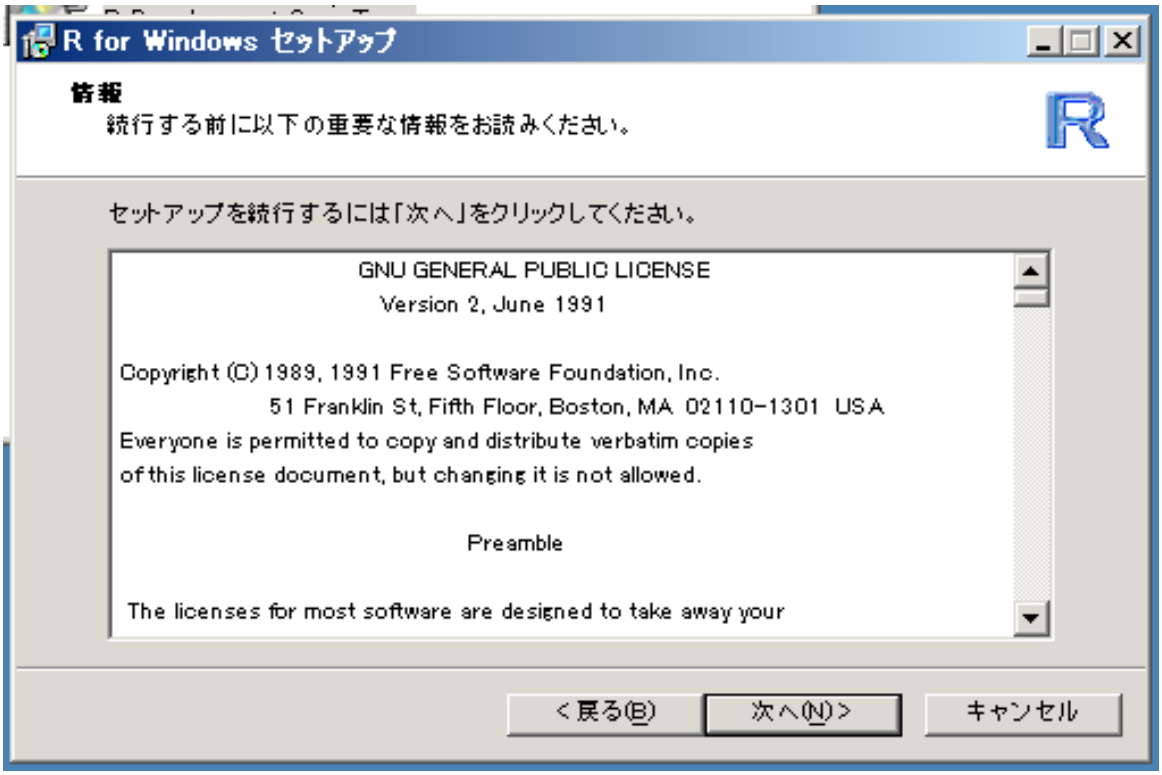
この R-2.5.1-win32 をダブルクリックして実行して下さい。



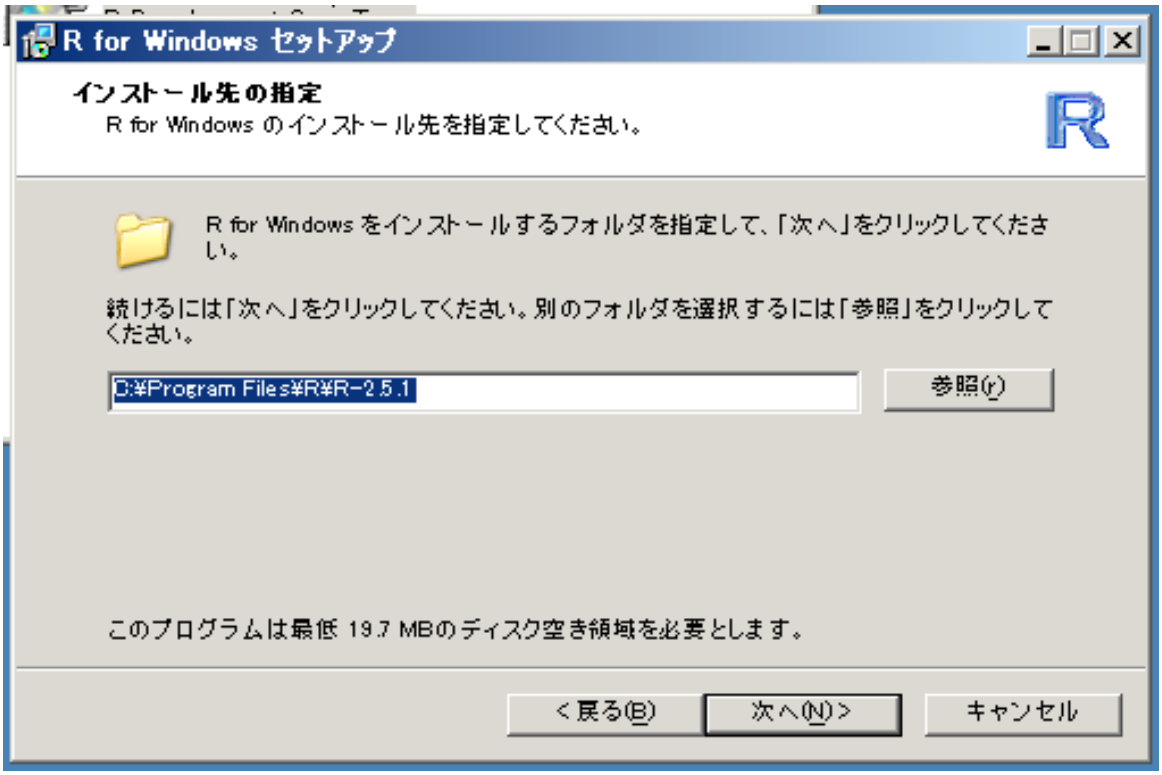
まずは言語の選択になります。最初から「Japanese」が選択されていますので、そのまま Ok ボタンを押して下さい。



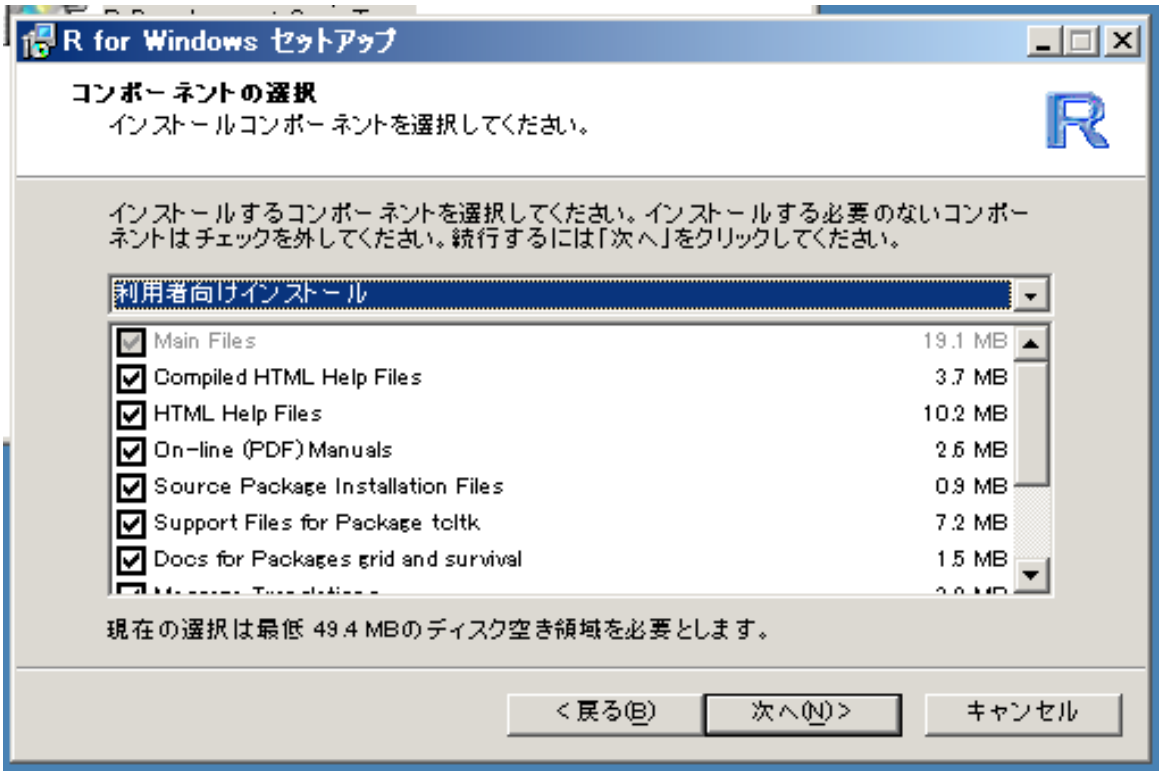
セットアップウィザードが開始されますので、「次へ」を押して下さい。



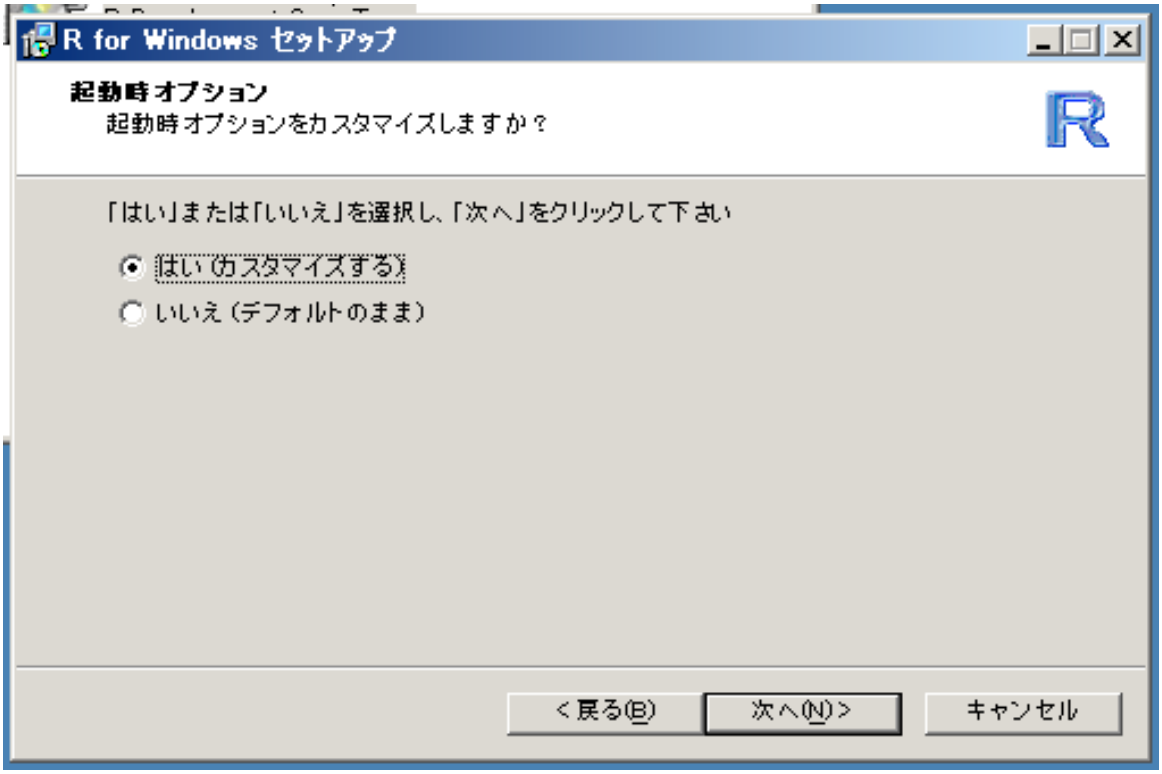
使用許諾条件（ライセンス）が表示されます。説明した通り、再配布や改変の自由を保証する、フリーソフトウェアライセンス(GNU General Public License, GPL) です。「次へ」を押してインストールへ進みます。



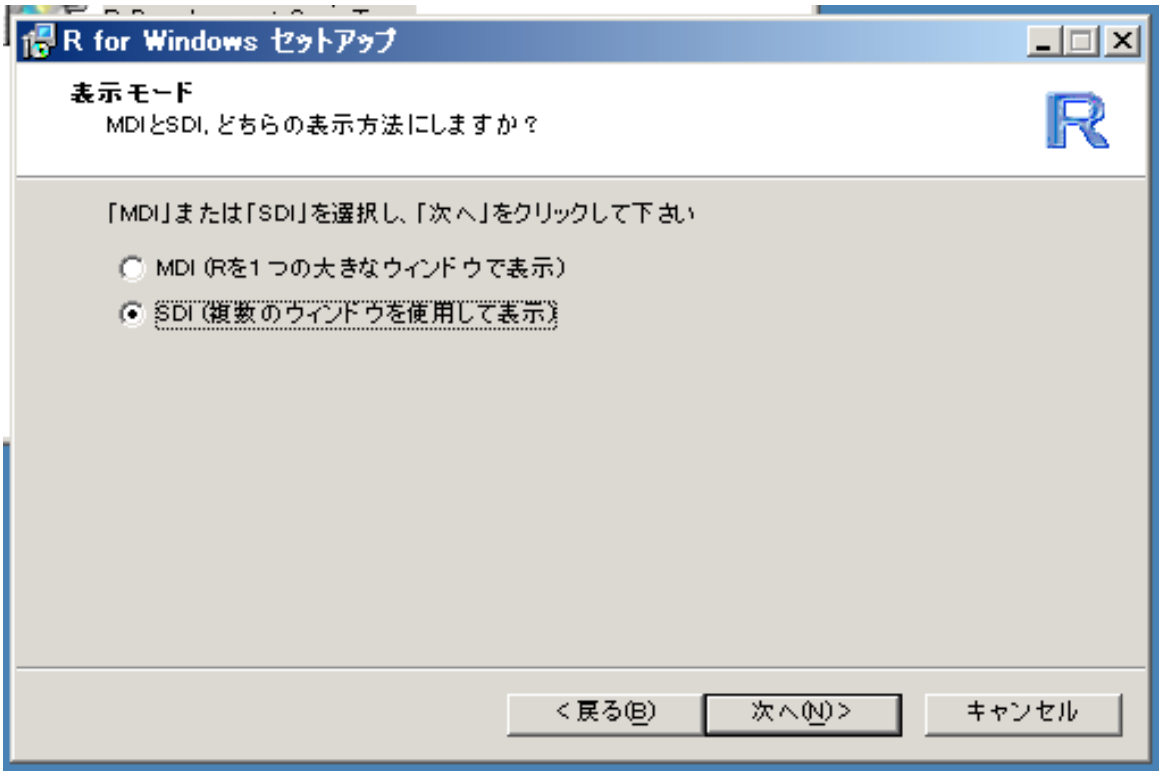
インストールする場所を指定します。通常は変更の必要はありませんので、「次へ」を押して下さい。



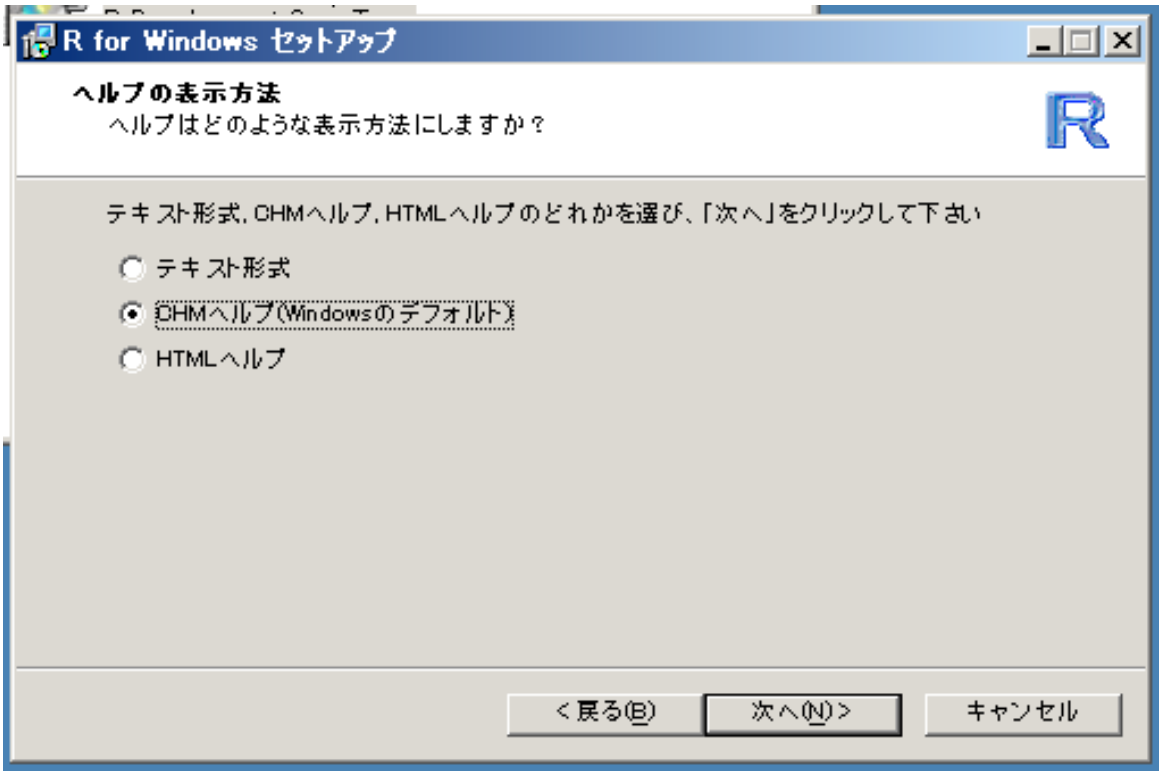
インストール内容を選択できます。通常はこの「利用者向けインストール」で問題ありませんので、「次へ」をクリックして進んでください。



起動時オプションについて聞かれます。これは主に、インターネットへの接続にプロキシサーバが必要な環境で利用する場合に設定します。カスタマイズしておくほうをオススメしますので、「いいえ」がチェックされていますが、**「はい」にチェックしなおしてから**「次へ」をクリックしてください。



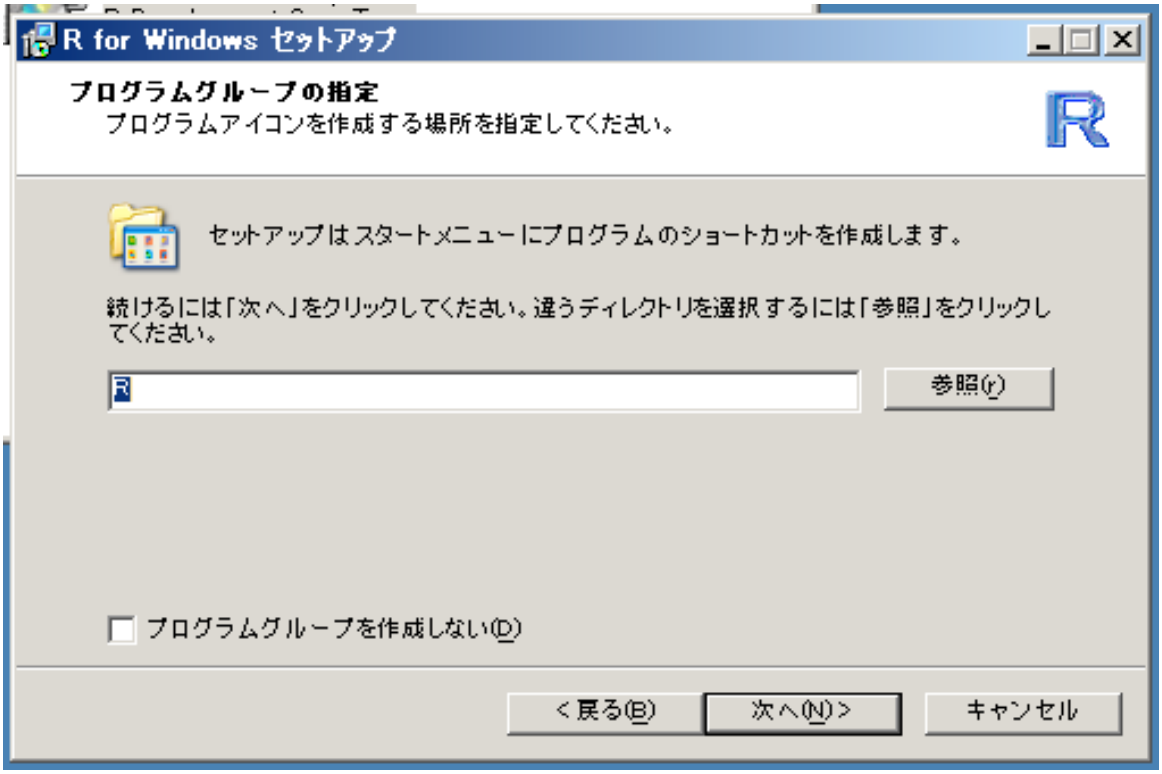
R を 1 つの大きなウィンドウで表示して、ほぼ常に全画面表示で利用するか、あるいは他のソフトのウィンドウと共存させるかを選択できます。Excel などを併用する場合は、全画面表示だと若干わかりにくくなりますので、**「SDI」にチェックしなおしてから**「次へ」をクリックしてください。



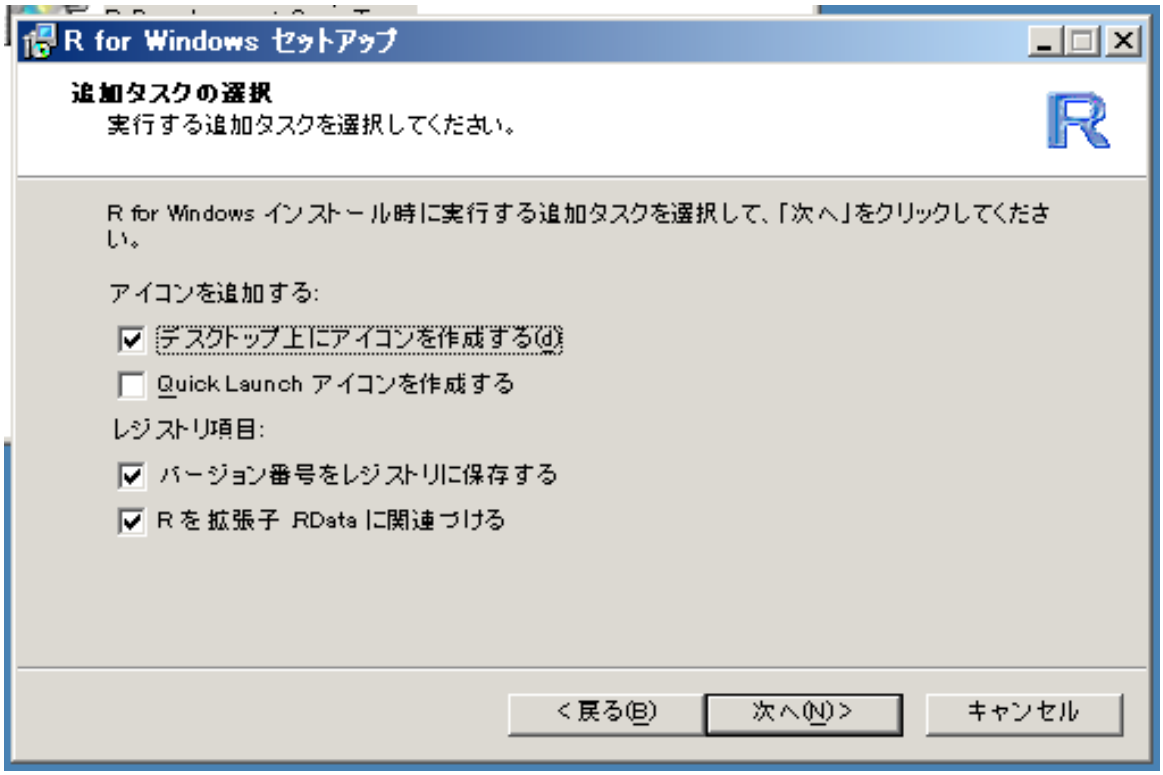
R のオンラインヘルプの表示方法を選択できます。Windows で使う場合は CHM ヘルプが使いやすいので、このまま「次へ」を押して下さい。

インターネットへの接続時に、プロキシサーバの設定を参照するかどうかを設定しています。

基本的には「Internet2」のほうが便利ですので、**「Internet2」に変更してから**「次へ」を押して下さい。

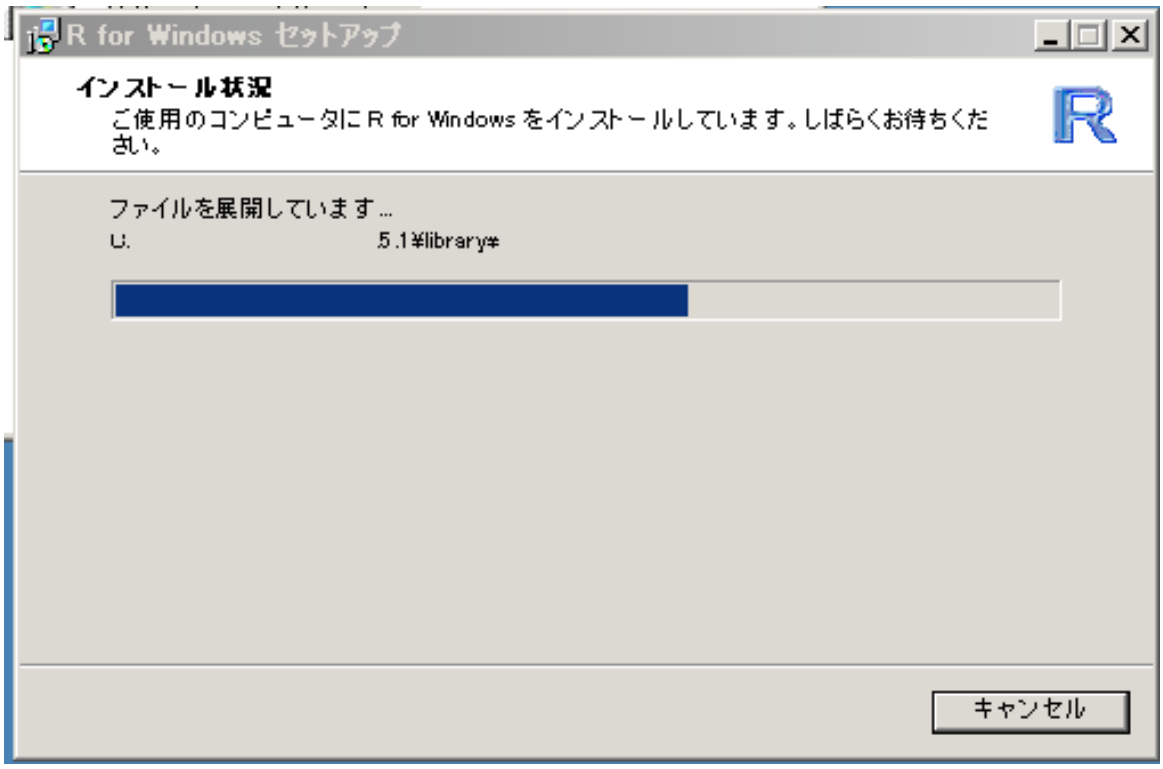


スタートメニューへの登録名を選択できます。デフォルトのままでかまいませんので、このまま「次へ」を押して下さい。



起動を便利にするために、デスクトップ上にアイコンを作成するかどうかを選択できます。基本的にはこのままでもかまいませんが、スタートボタンの隣のクイック起動エリア（下図）からプログラムを起動するのがお好みの場合は、「デスクトップ上にアイコンを作成する」のチェックをはずし、「Quick Launch アイコンを作成する」のほうをチェックされると便利でしょう。

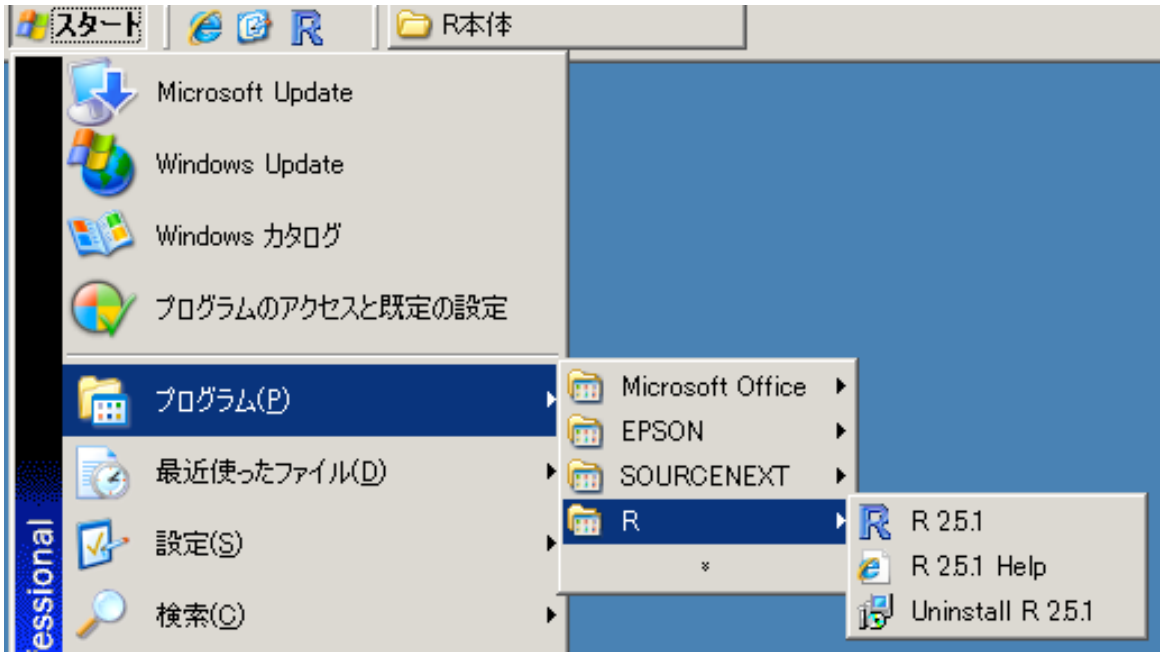
あとの項目はそのままでかまいません。「次へ」を押すと、インストールが開始されます。



しばらく待つと、インストールが完了しますので、「完了」ボタンを押せば R のインストールは完了です。

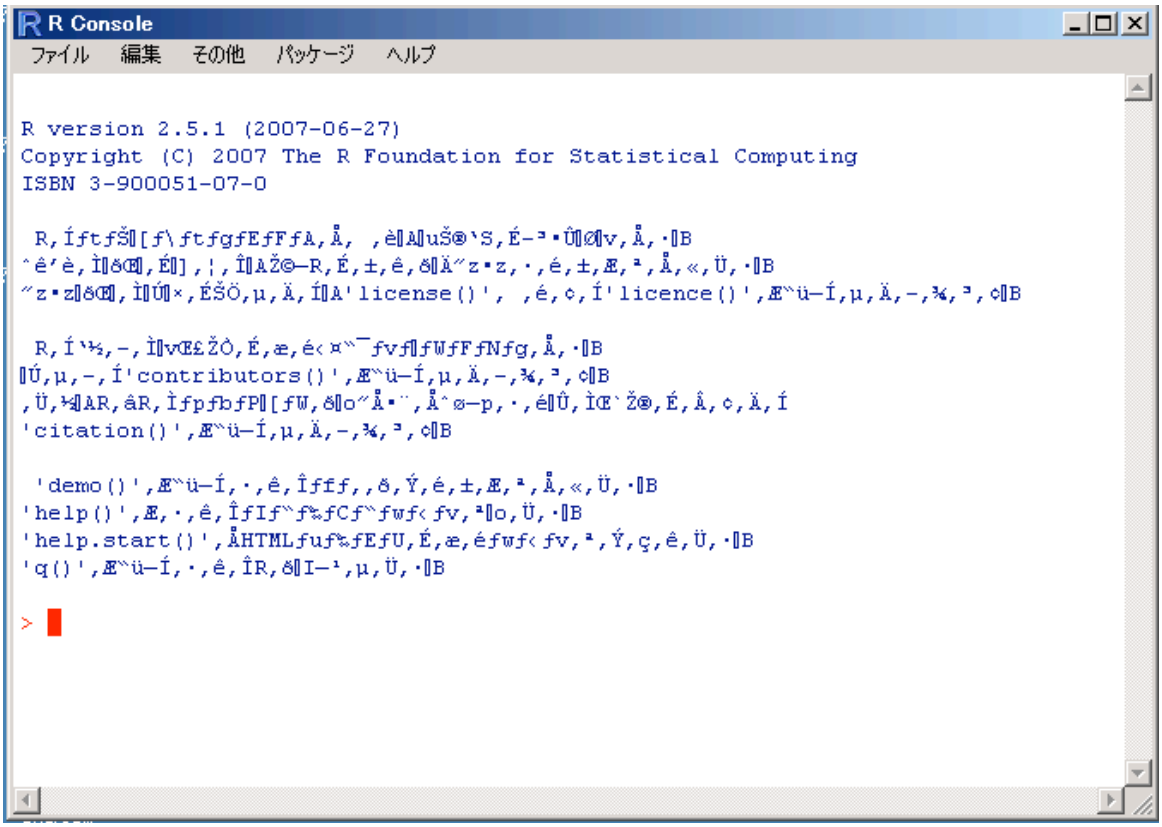
2. 起動と終了

それでは、早速インストールした「R」を起動してみましょう。起動にはいろいろな方法があります。最も基本的には、スタートメニューから起動できます。「(すべての) プログラム」-「R」-「R 2.5.1」を選択すれば起動できます。

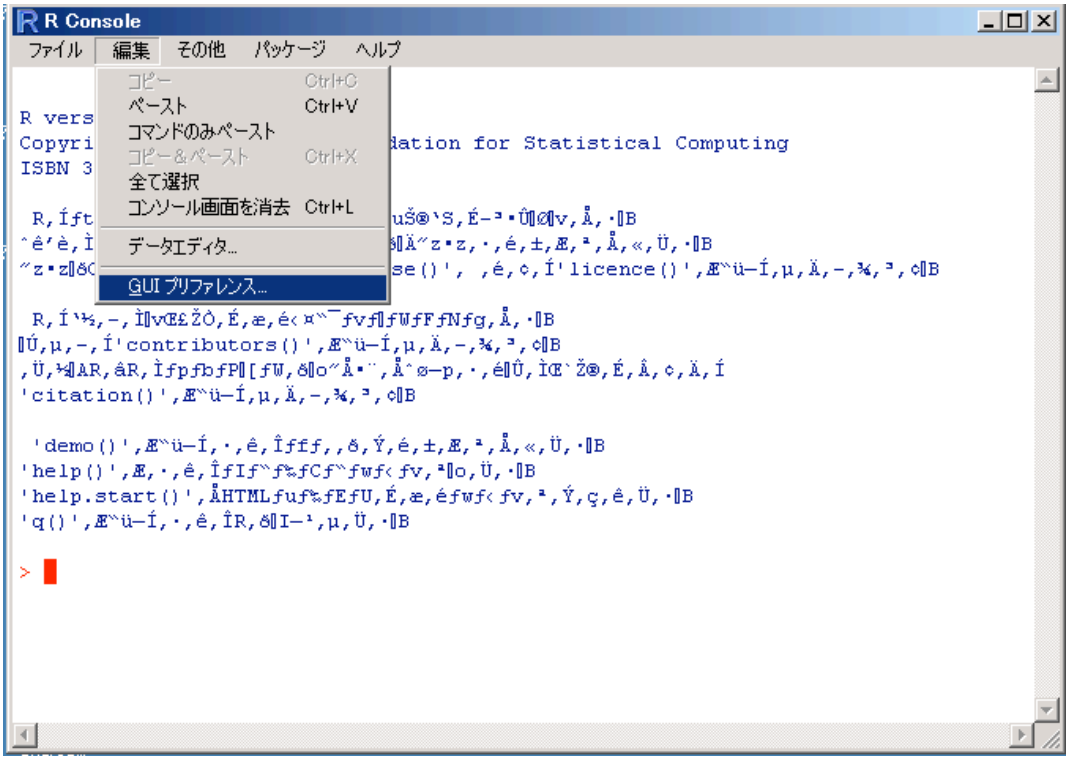


あるいは、デスクトップ上にアイコンが出ている場合は、これをダブルクリックしてもよいですね。

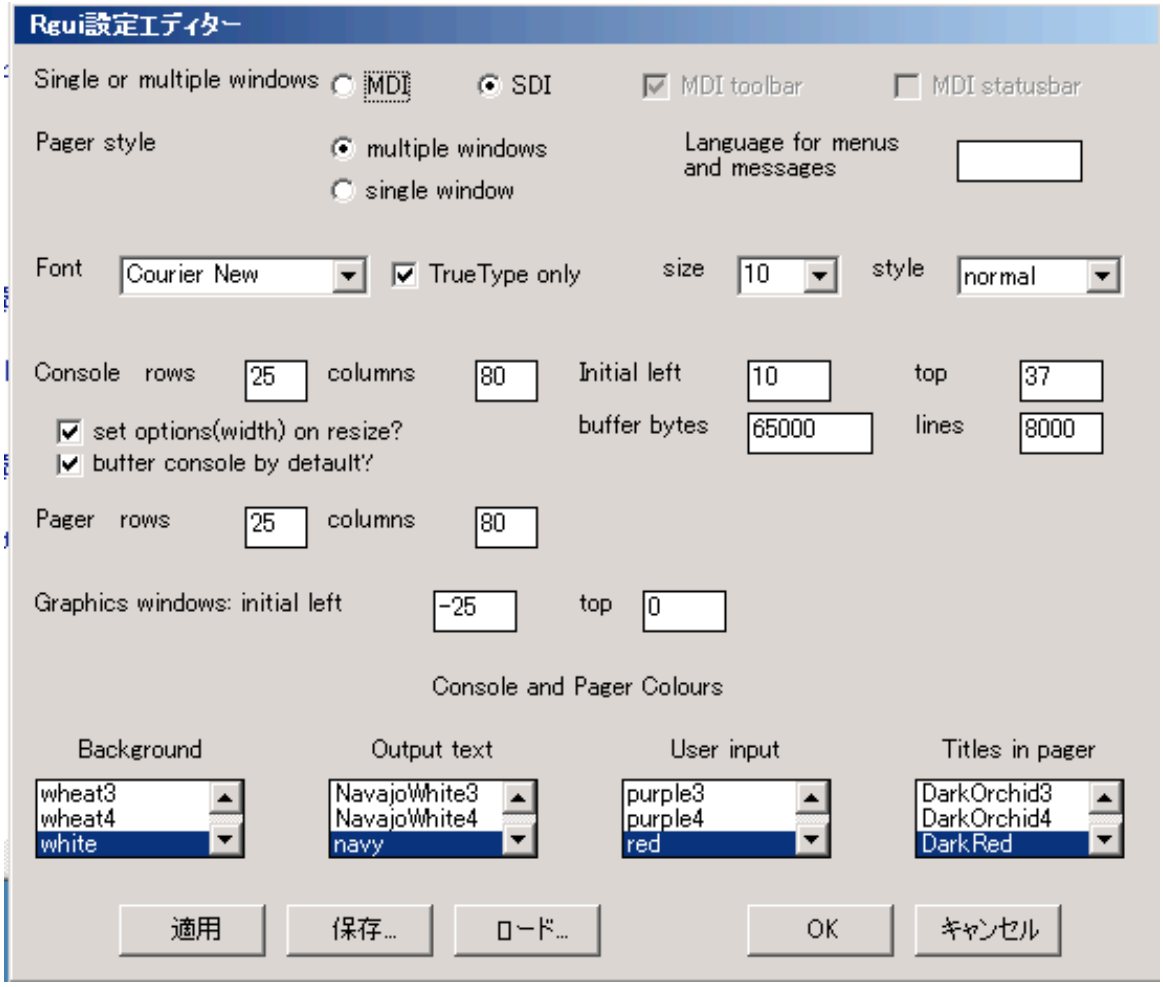
Quick Launch アイコンを作成した場合は、スタートボタンのとなりのクイック起動エリアに同じようなアイコンができています。それをクリックしても起動できます。



起動すると、無事、「ファイル」「編集」「パッケージ」などの日本語メニューを持った R が起動しますが、残念なことにメッセージが文字化けしているようです。世界中、いろいろな国の人が使うので、最初から日本語に最適なよう設定しておくことができません。日本語で快適に使うためには、１つだけ最初に設定をする必要があります。なお、Macintosh 版を利用している場合は、Mac OS の先進的な多国語対応機能のおかげで、R では何も設定しなくても文字化けは起こりませんから、R を終了してみるところまで飛ばして進んで下さい。

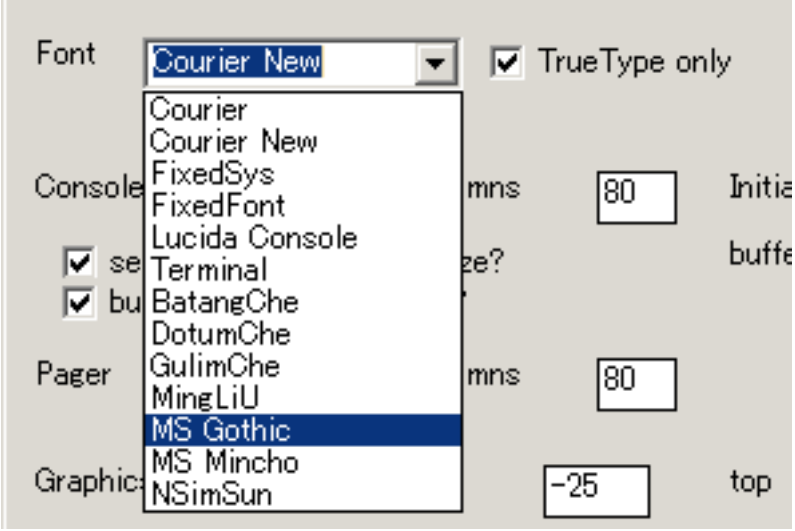


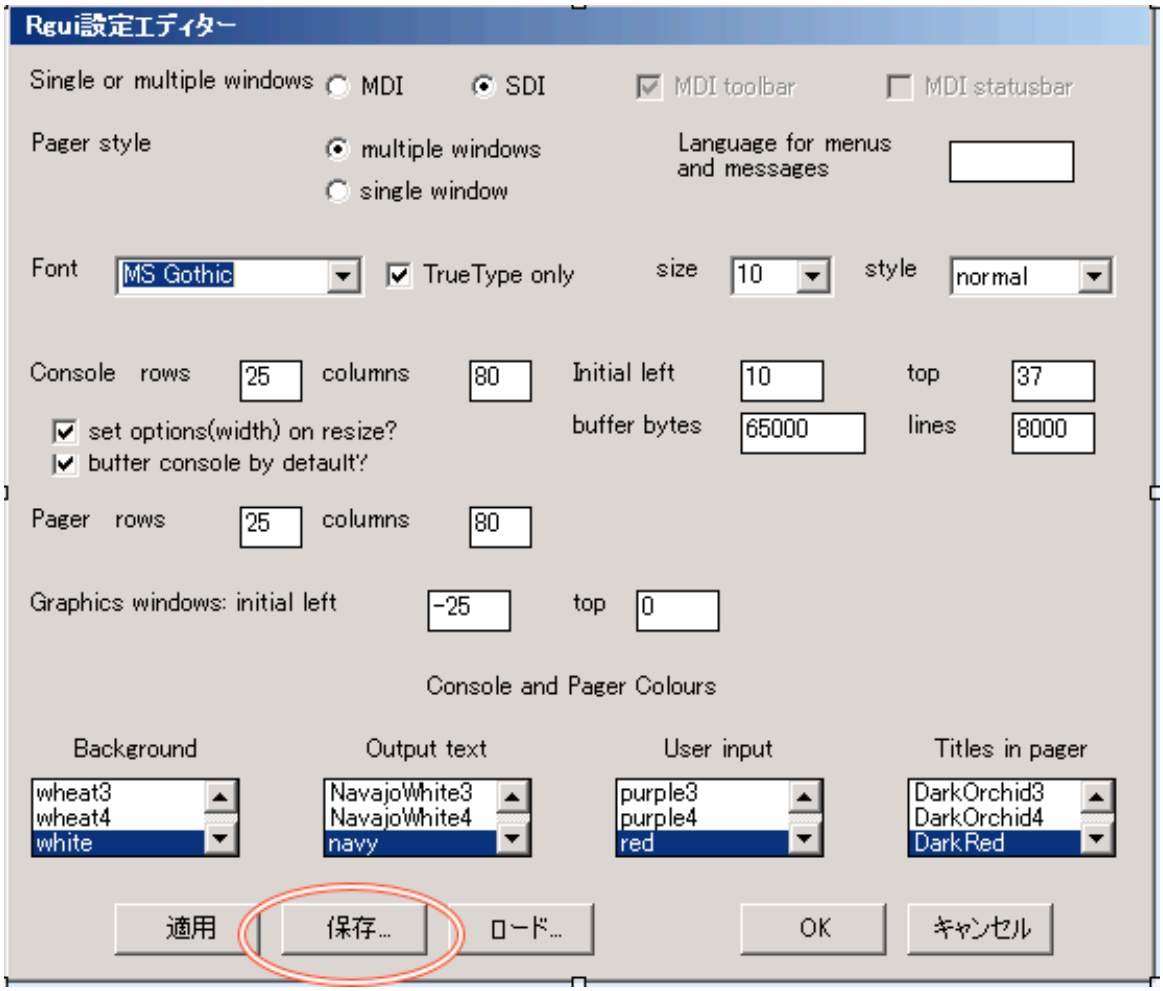
「編集」メニューから「GUI プリファレンス...」を選んで下さい。



「Rgui 設定エディター」が表示されます。ここで、R の表示する文字の色やサイズなどの調整ができますが、上から三段目の「Font」の部分に注目して下さい。

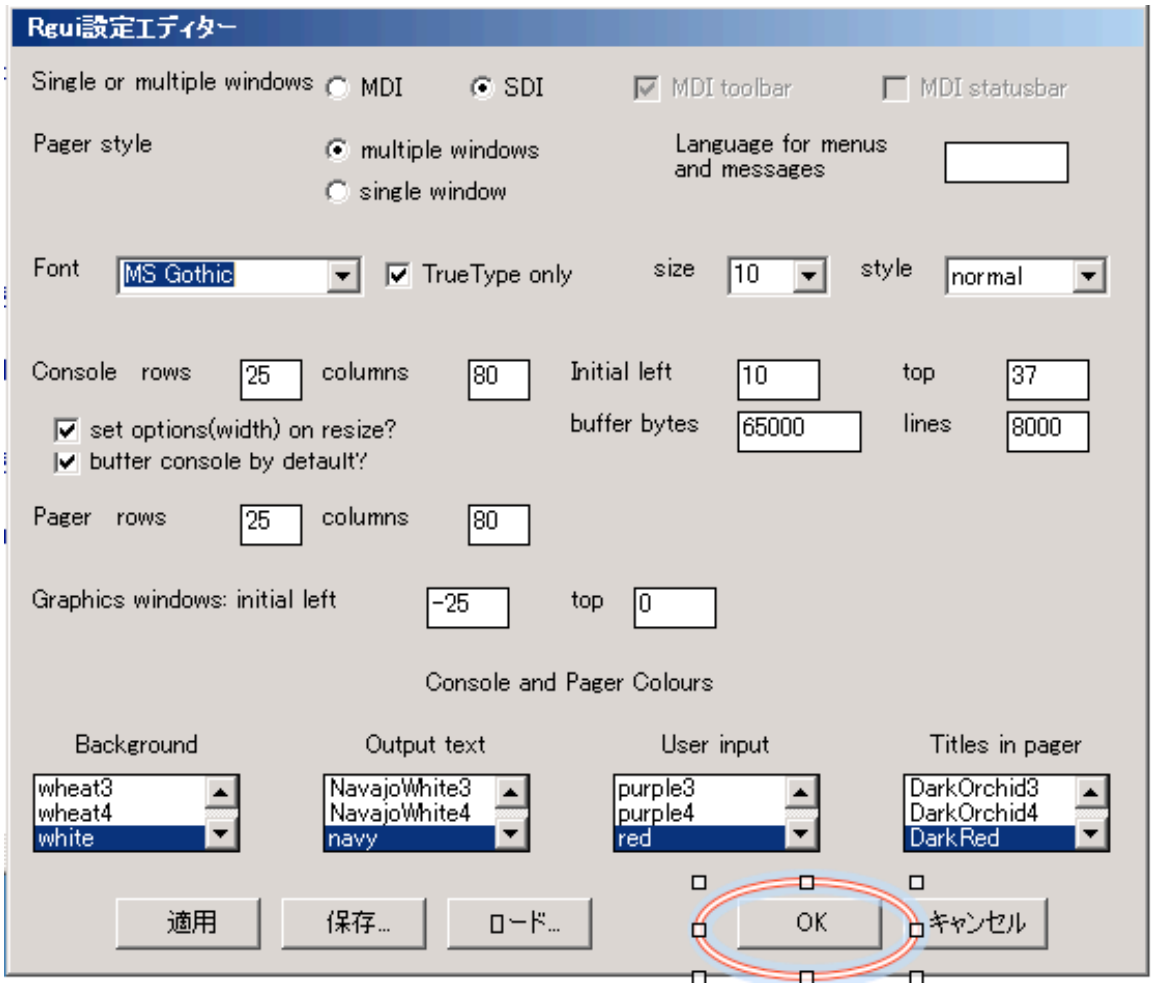
「Courier New」というのは、英文のみに使えるフォントです。これが文字化けしてしまった原因になっています。ここをクリックして、「MS Gothic」に変更してみましょう。



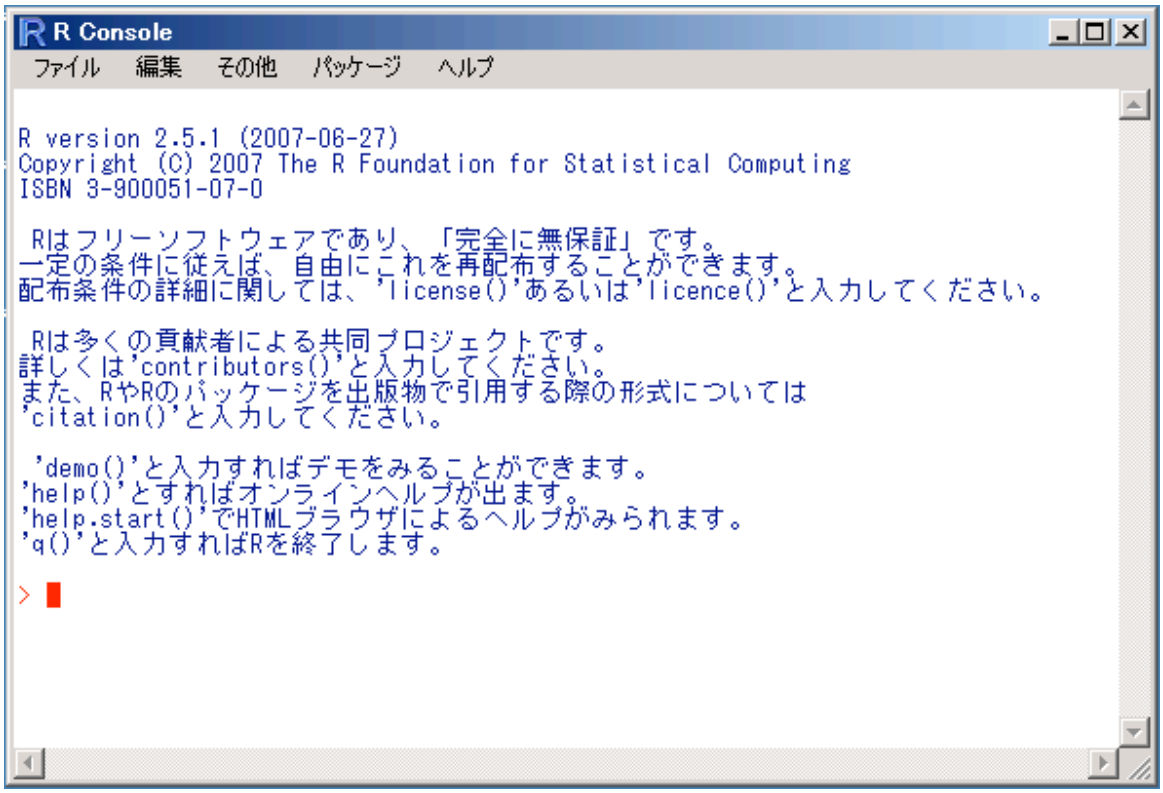


変更できたら、毎回設定しなおす必要がないように、設定を保存しましょう。「保存」ボタンを押して下さい。

設定ファイルを保存する場所を聞かれますが、通常はこのままでよいので、そのまま「保存」ボタンを押して下さい。

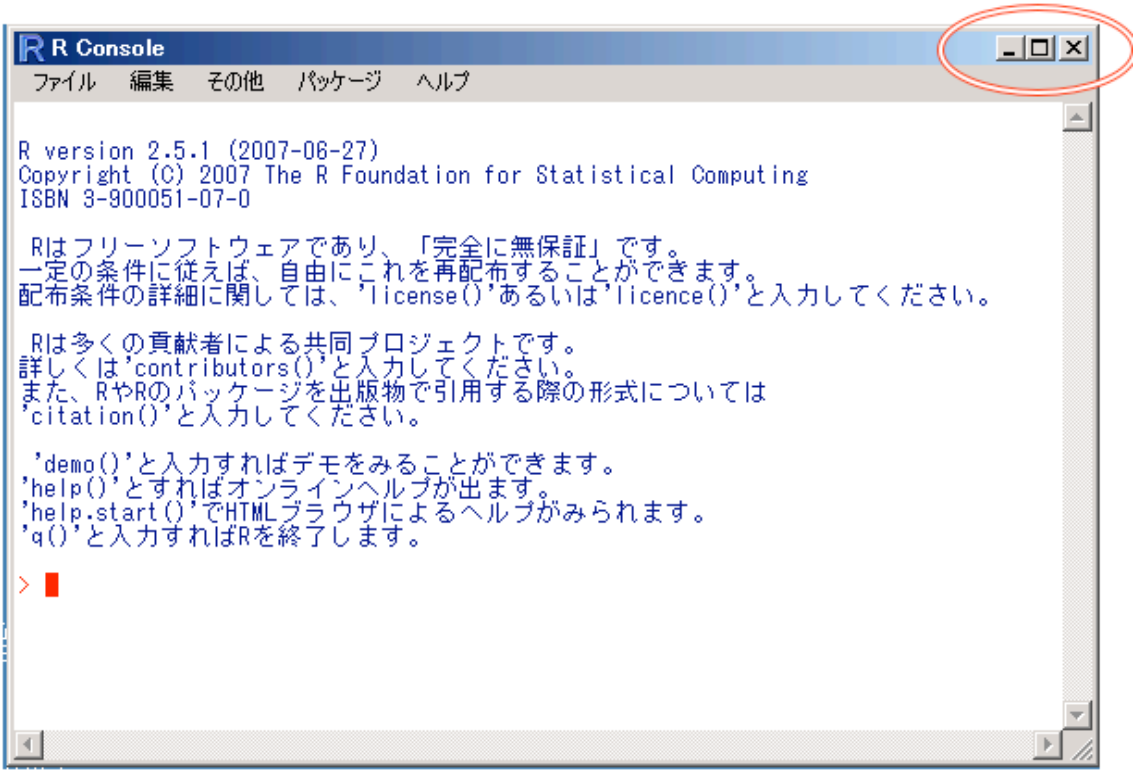


保存が完了すると、「Rgui 設定エディター」に戻りますので、「Ok」ボタンを押せば、日本語環境のための設定は終了です。

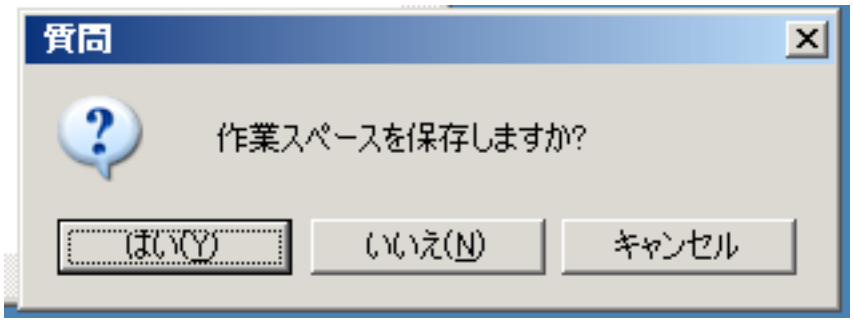


R が表示しているメッセージが文字化けなく表示されるようになりました。

それでは、一度 R を終了してみましょう。



「R Console」と出ているウィンドウの右上の「×」ボタンをクリックすれば、終了できます。Macintosh の場合は、ウィンドウ左上の赤ボタンです。さっそく、押してみましょう。



おっと、ここで、「作業スペースを保存しますか？」と聞かれます。Macintosh では、「ワークスペースのイメージファイルを保存しますか？」と聞かれますが、同じ意味です。

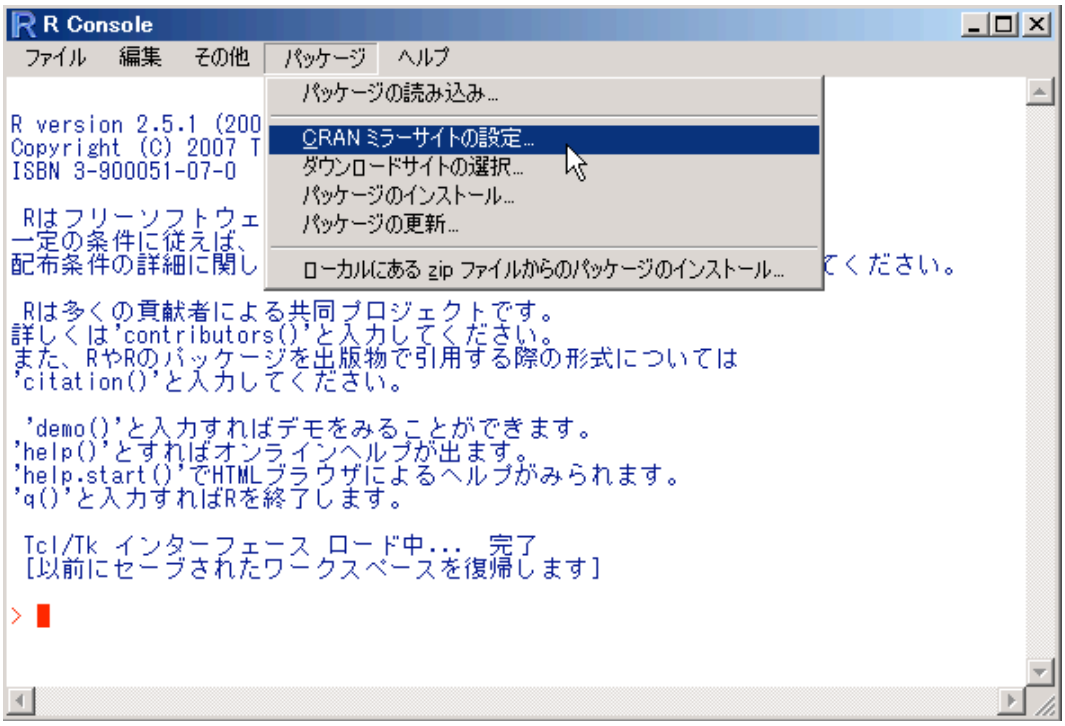
ともあれ、この「作業スペース or ワークスペース」とは、皆さんが R を使ってデータを入力, 解析をすすめていく際に、そのデータや解析結果、数理モデルなどは、すべてこのスペースに自動的に保存されます。つまり、作業スペースが保存してあれば、一度 R を終了し、もう一度起動したときにも、前回入力したデータや、前回の解析で作成したモデルなどはそのまま続けて利用することができる、ということです。なので、基本的には作業スペースは保存して終了するようにしましょう。ただし、とくに新しい作業を何も行わなかったときや、間違えて重要なデータを消去してしまった場合は、保存せずに終了すれば起動中に行った作業をなかったことにできます。

さて、無事に終了することができたら、再度 R を起動し、R Commander のインストールに進みましょう。

3. R Commander ほか、パッケージのインストール

R は基本機能部分だけでも十分に高機能になってきていますが、もともとプロフェッショナル用のツールという色合いが濃く、あまり初心者にやさしくありません。具体的には、他のソフトのようにメニューから解析方法を選択して実行する、ということができるようになっていません。基本的には、赤く表示されている「＞」記号、これを「プロンプト」と呼びますが、これのあとにたとえば「t.test(data$height, data$weight)」などと、いわば「R 言語」による指示を書き込んでいかなくてはならないのです。これは大変面倒に見えますが、実は、慣れてくると頭の中で考えている解析プランをそのままの形で書くことができ、そのまま正確に記録に残して、たとえばデータに変更があった場合に同じ手順をすぐに繰り返して実行してみることができるので、長く使っているとこちらのほうが圧倒的に便利だったりします。

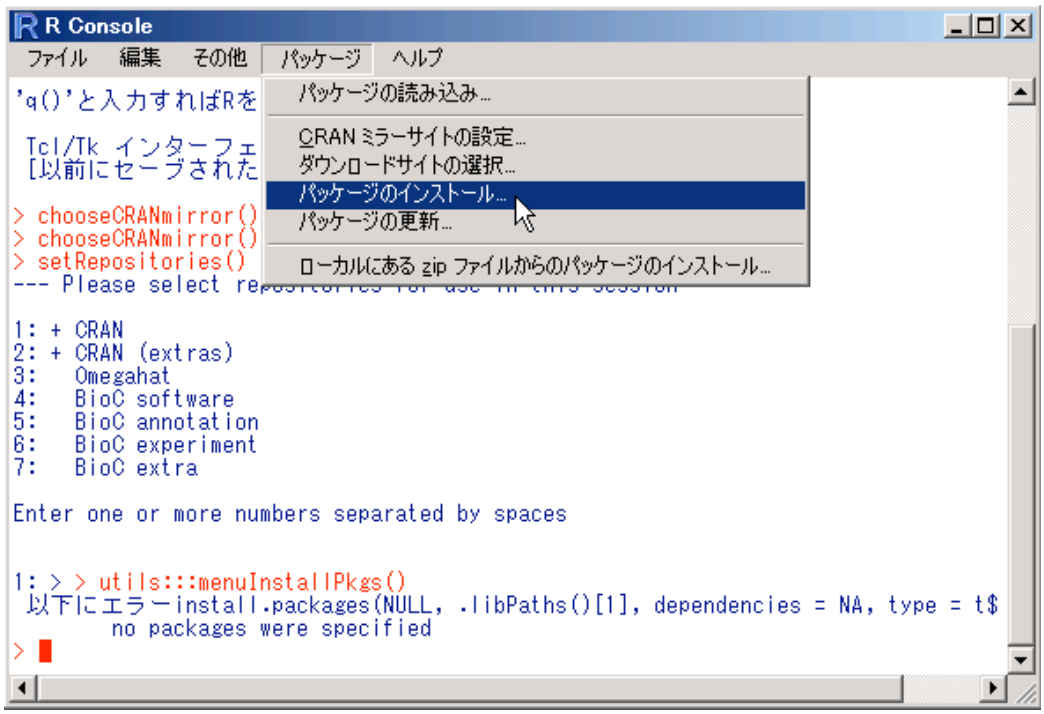
ですが、やはり最初に学ぶには敷居が高い、というところがありますので、初心者向けに比較的簡単な操作で R の機能を使ってみることができるように作られた「パッケージ」があります。それが「R Commander」、略して「Rcmdr」です。これは、R の基本機能ではないので、追加としてインターネット経由でダウンロードし、インストールする必要があります。パッケージのインストールは、R を起動してから、R の中で行います。

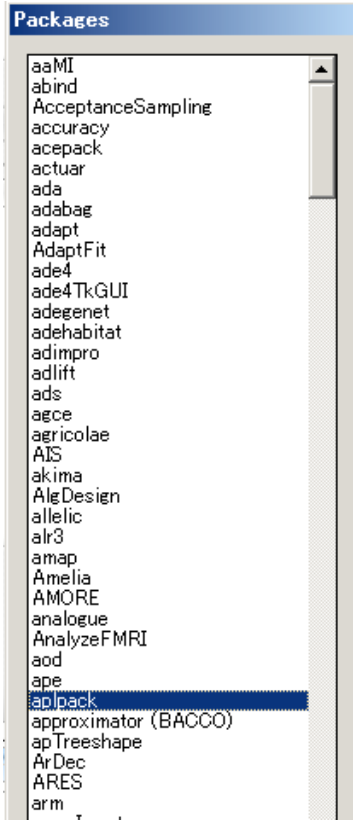


R Console ウィンドウ上で、「パッケージ」メニューを選びます。さらに「CRAN ミラーサイトの設定...」を選択してください。パッケージをダウンロードできる Web サイトは世界中にあるのですが、せっかくですので国内のものを使ってみたいと思います。どの Web サイトからダウンロードするかをここで指定します。

縦に長い、「CRAN mirror」というウィンドウが出ますから、「Japan (Tsukuba)」を選択して、Ok ボタンを押してください。これで、筑波大学にあるダウンロード用 Web サイトからパッケージをインストールする、と指定したことになります。

次に、同じく「パッケージ」メニューから、「パッケージのインストール...」メニューを選びましょう。





すると、また大変縦長な「Packages」というウィンドウが開きます。これがすべて R の拡張機能、「パッケージ」です。ここでは、R Commander の本体、「Rcmdr」をはじめとして、この演習で必要ないろいろなパッケージをインストールします。Ctrl キーを押しながらクリックすることで複数のパッケージを同時に選択できますから、以下にあげるパッケージをすべて選択して、Ok ボタンを押してください。

インストールするパッケージ:

abind

car

effects

ISwR

leaps

lmtest

mapproj

maps

maptools

multcomp

mvtnorm

Rcmdr

RcmdrPlugin.HH

RcmdrPlugin.TeachingDemos

relimp

rgl

RODBC

sgeostat

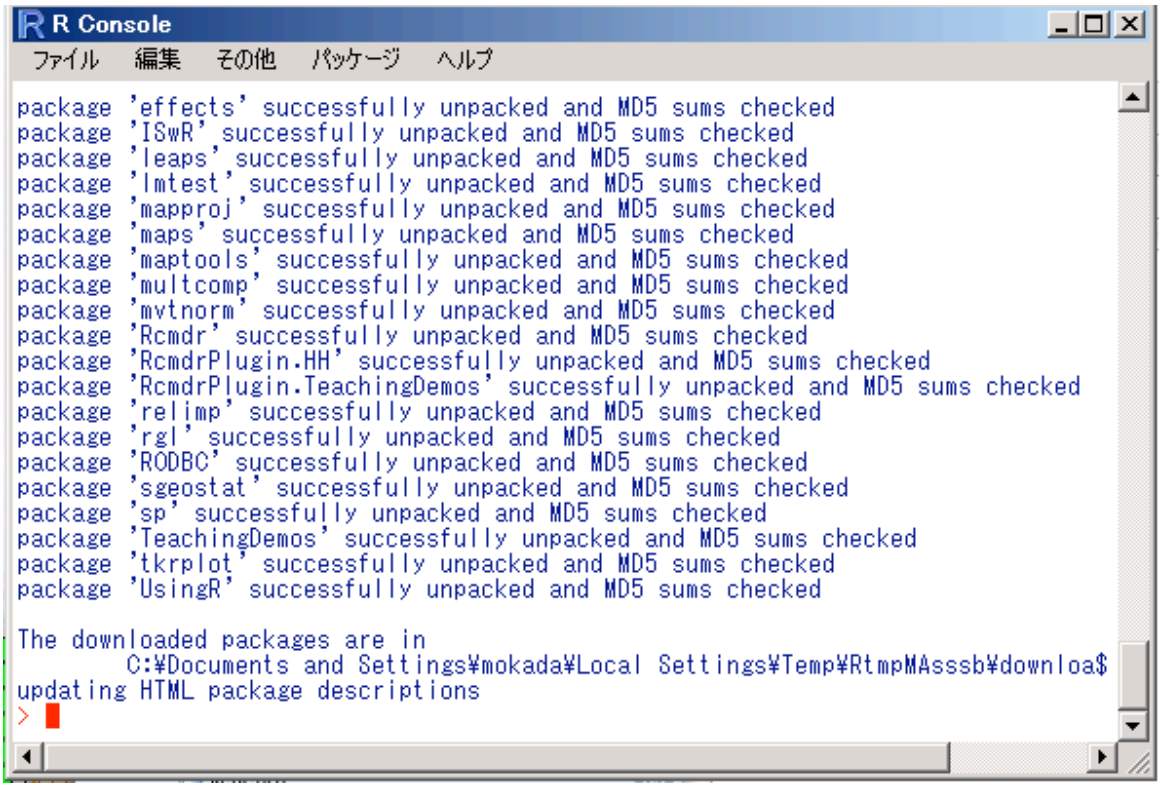
sp

TeachingDemos

tkrplot

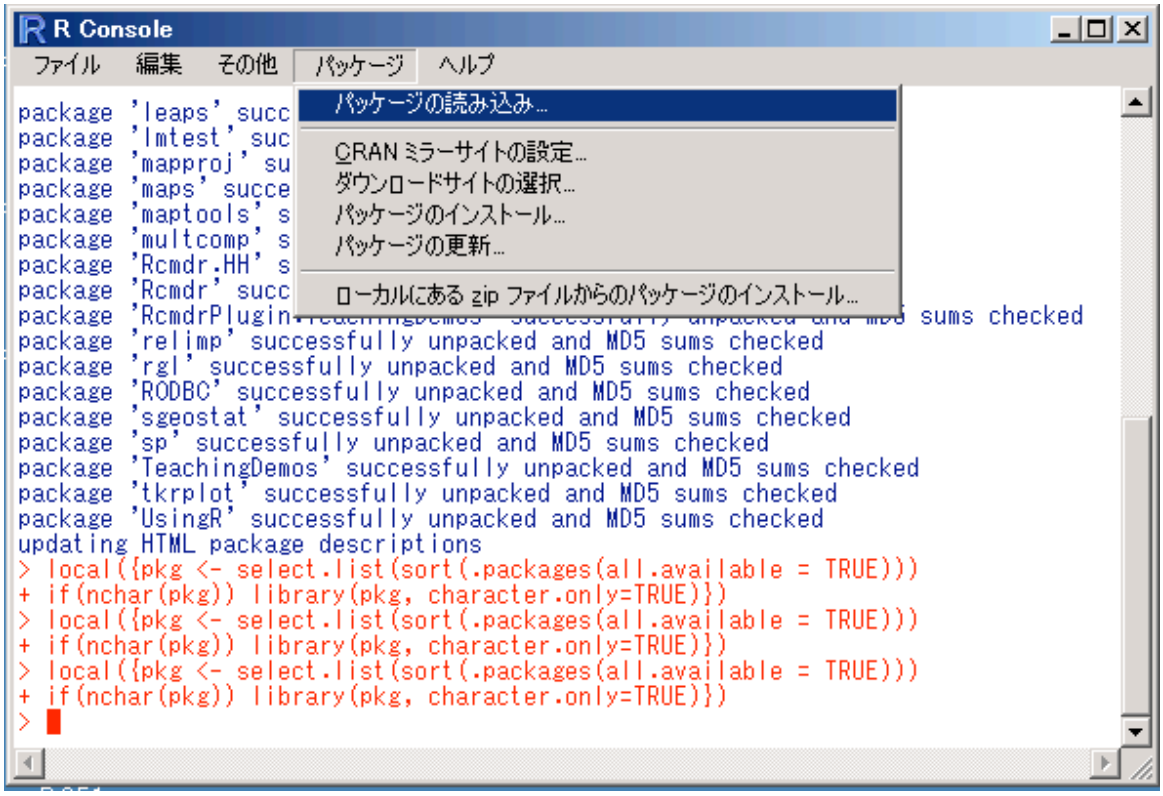
UsingR

…お疲れさまでした。これで、R Commander を使う準備ができました。



すると、パッケージを解凍し、内容のチェックをした上でインストールしたというメッセージが表示されて、インストールが完了します。インストールが完了したパッケージは、すぐに使い始めることができます。早速、R Commander を利用してみましょう。

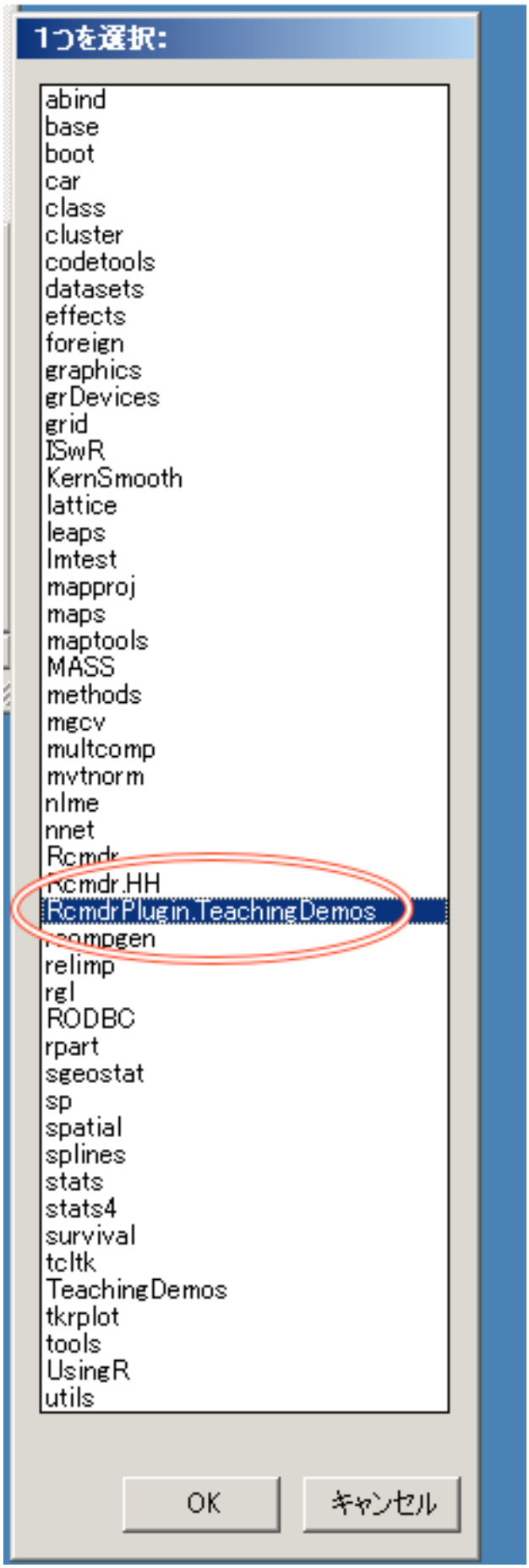
4. R Commander の起動と終了



メニューの「パッケージ」から、「パッケージの読み込み...」をクリックします。なお、Macintosh 版では、このメニューがないので、プロンプト(">" 記号)の直後に

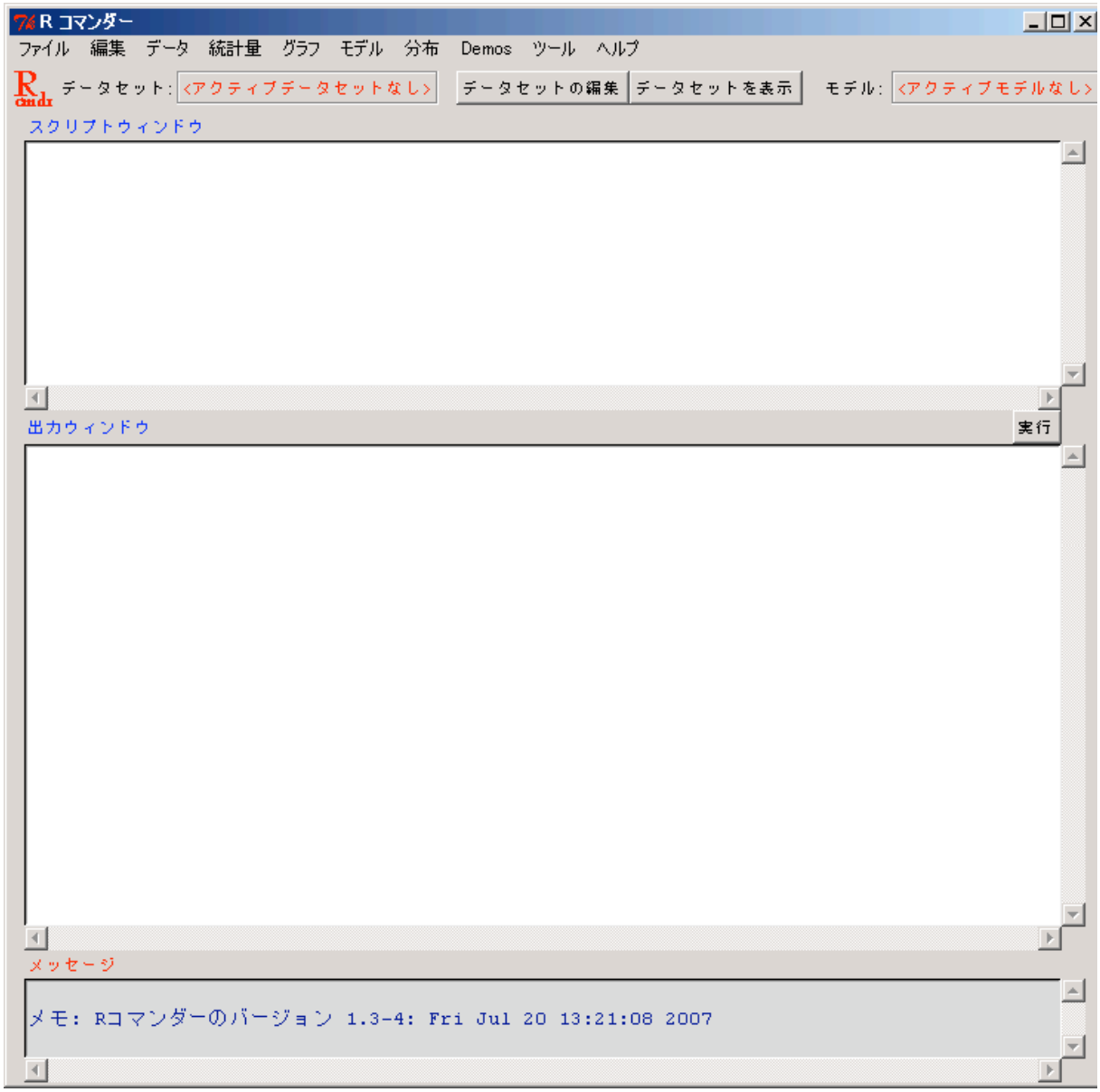
library(RcmdrPlugin.TeachingDemos)

と打ち込んで下さい。ちなみに、Windows 版でもこの方法は利用できます。



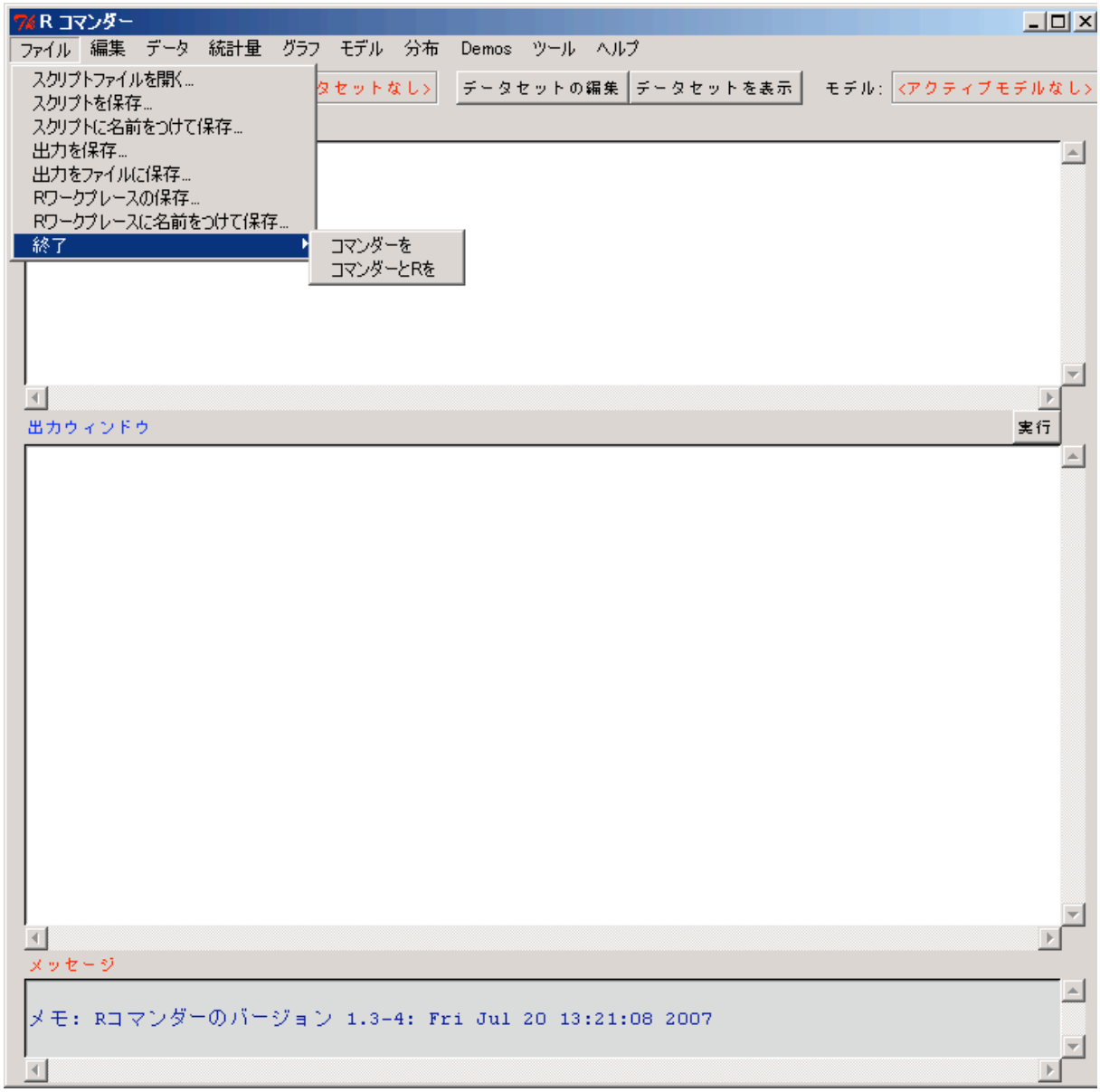
さて、パッケージを読み込もうとすると、インストール済みのパッケージのうち、どれを読み込むかを選択する画面になります。R Commander のパッケージ名は「Rcmdr」ですが、今回はとくに教育用に改変された別のバージョンを使いますので、「RcmdrPlugin. TeachingDemos」となります。これを選択して「Ok」をクリックしてください。

すると、以下のような「R コマンダー」というウィンドウが出現します。

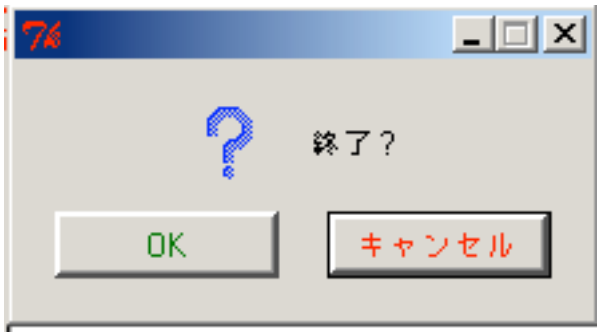


以後の作業は、このウィンドウの上で行っていくことになります。

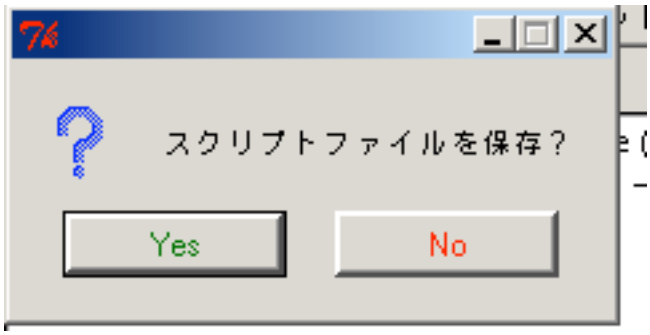
それではまず、コマンダーを終了してみます。



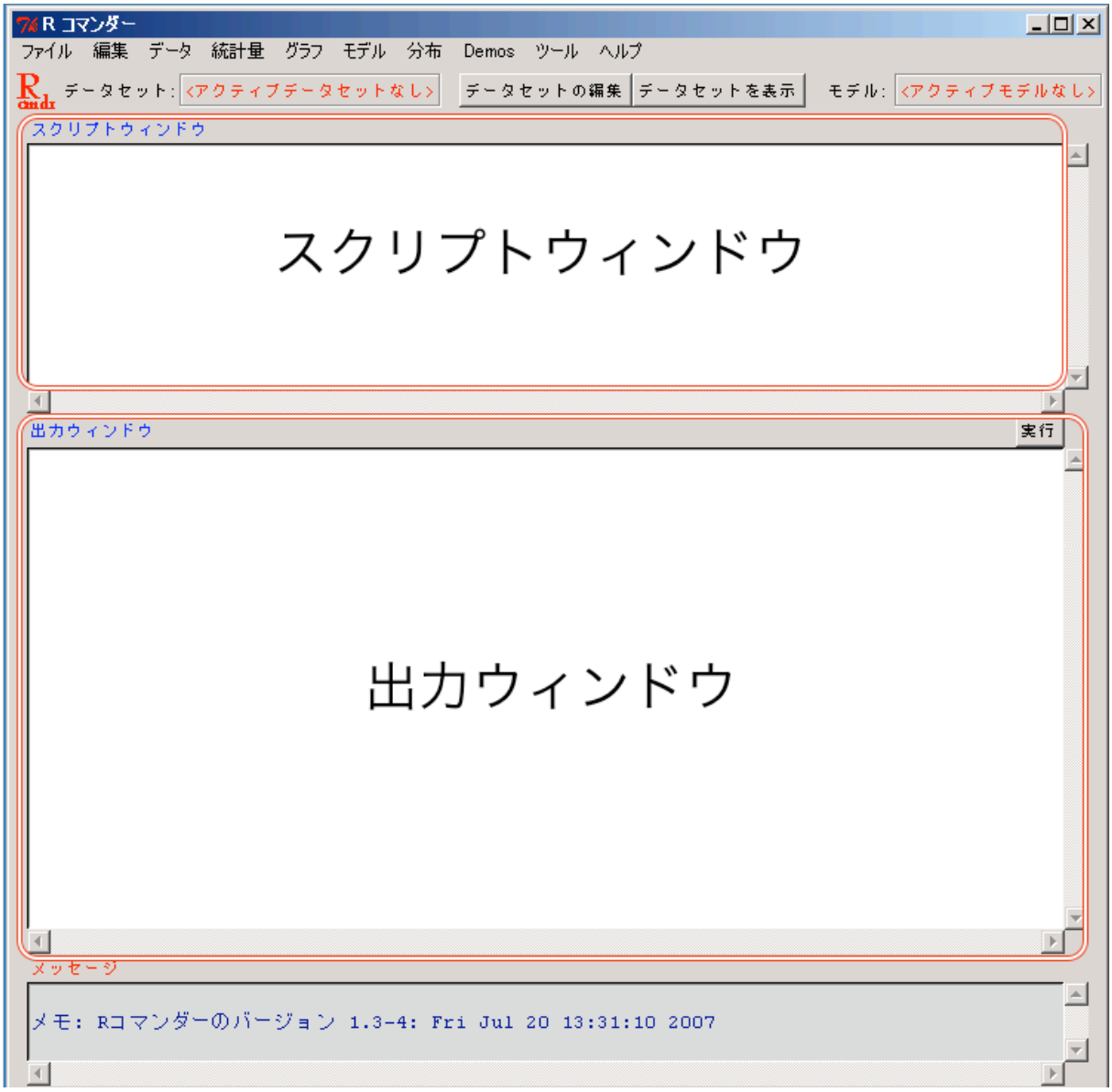
「ファイル」メニューから「終了」を選ぶと、「コマンダーを」「コマンダーと R を」の２つが選べます。コマンダーと R を一緒に終了できると便利そうですが、そうすると作業スペースが自動的に保存されないという問題がありますので、**基本的には「コマンダーを」を選択して、**コマンダーのみ終了しましょう。その後は普通の R に戻りますので、R を終了すれば Ok です。
なお、コマンダーの終了を選択すると、それを確認するウィンドウが出ます。



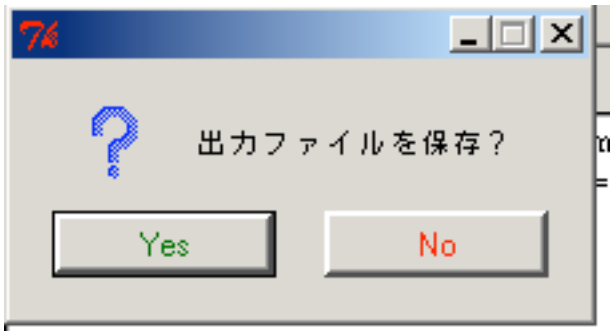
「Ok」を押せば、コマンダーを終了できます。今回はこのまま終了できますが、今後、コマンダー上で何か作業をしたあとに終了しようとすると、さらに次のような確認が出ます。



これは、コマンダーの画面の上半分、「スクリプトウィンドウ」に出ている内容をテキストファイルに保存するかどうかをきいています。



このウィンドウには、メニューから実行した解析内容を、プロンプト（">" 記号）からの入力で行うとしたら、どのような命令になるかが随時表示されています。つまり…もし、コマンダーから実行した内容をあとから繰り返して実行してみたい、と思ったときには、このスクリプトウィンドウの内容を保存しておき、それをあとから普通の R にコピー＆ペーストすれば Ok なのです。
さらに、「出力ウィンドウ」の内容も保存するかどうかきいてきます。



これは、下半分の「出力ウィンドウ」の内容をテキストファイルに保存するかどうかをきいています。初歩的な解析の結果は、この出力ウィンドウに表示されますので、ここを保存しておけば R を終了したあとでも見直すことができます。**スクリプトウィンドウ、出力ウィンドウとも、原則として保存しておくようにしましょう。**

さて、これで、R コマンダーの利用の準備が整いました。一度、コマンダーを終了し、続けて R 自体も終了して、再度 R を起動して、さらにコマンダーを起動するところまでやってみて下さい。いよいよ、実際の利用に入っていきましょう。

5. データセットの読み込み

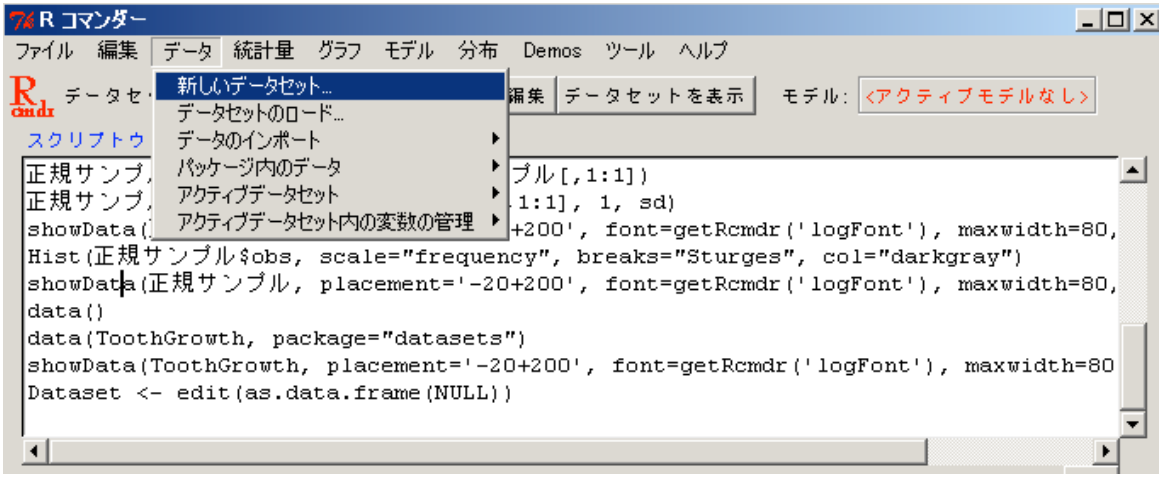
R コマンダーでは、解析の対象とする一連のデータを「データセット」と呼びます。データセットは自分で Excel などで作成するだけではなく、R の本体や、パッケージにもサンプルがたくさん含まれています。また、特定の確率分布に基づいたデータを「発生」させることも簡単にできます。ここでは、まずは Excel などで作成したデータの読み込みをやってみましょう。

まず最初に、Excel で作ったデータを R コマンダーに読み込んでみましょう。小さなデータですが、以下のものを Excel に入力してみてください。

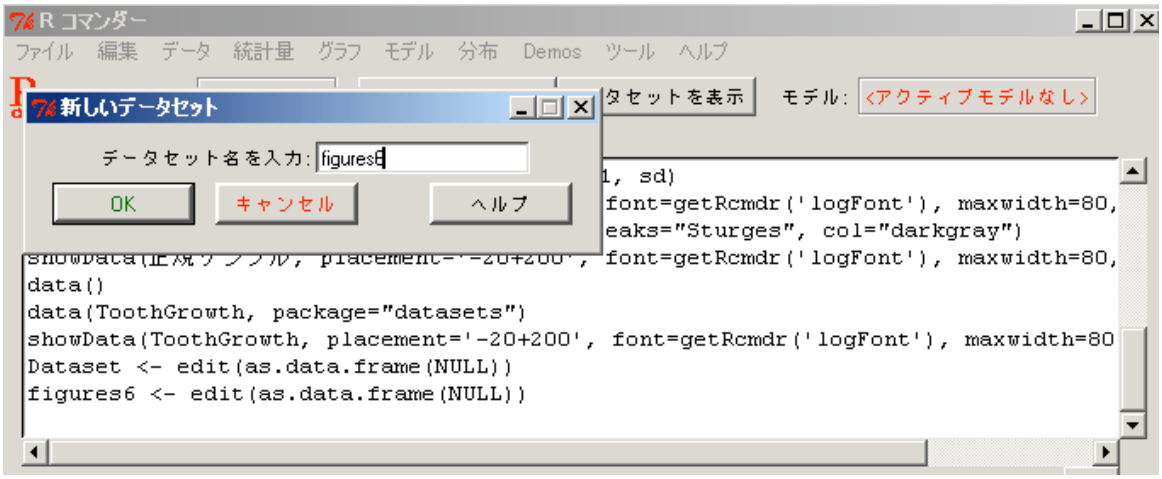
height	weight
1.75	60
1.80	72
1.65	57
1.90	90
1.74	95
1.91	72

これは、6 人の身長(height)と体重(weight)をあらわすデータです。

Excel などが PC に入っていない場合でも大丈夫。R には、まだまだ十分とはいえませんが、表形式のデータを入力する機能も備えられています。

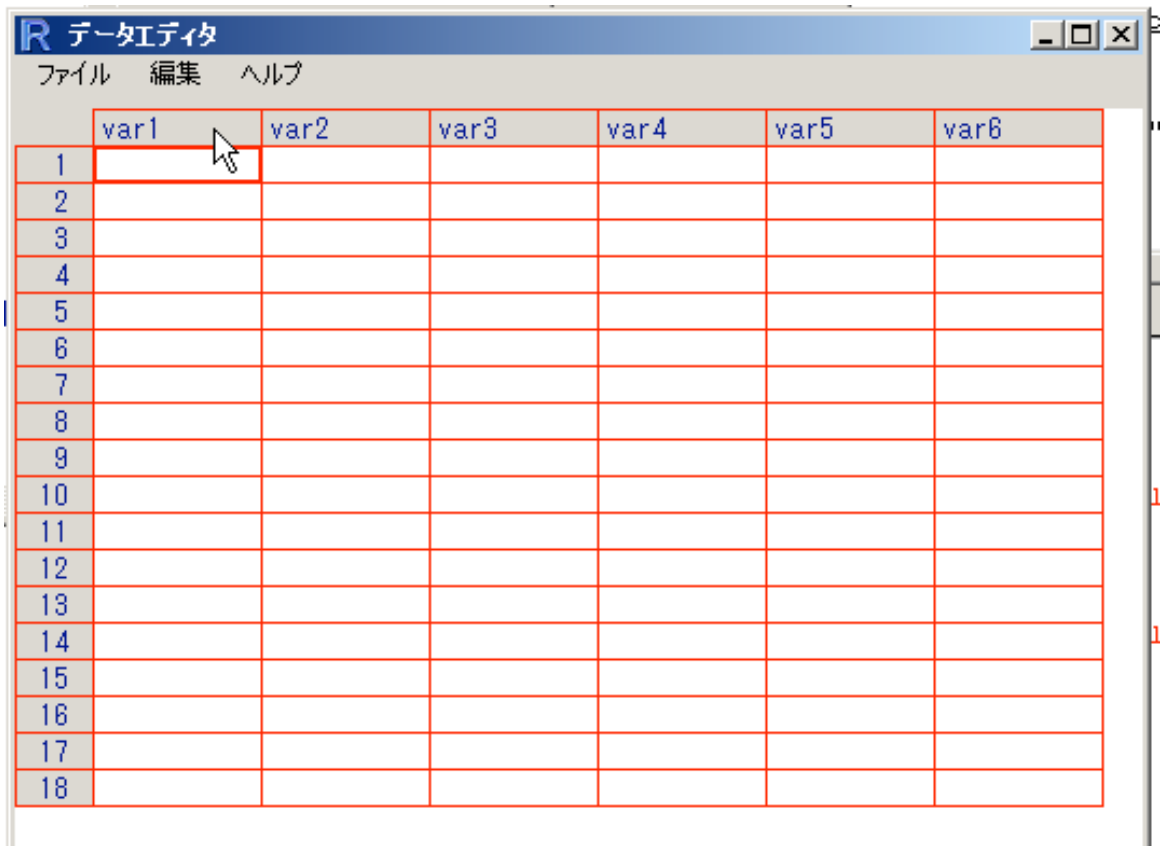


コマンダーの「データ」メニューから、「新しいデータセット...」を選んでください。

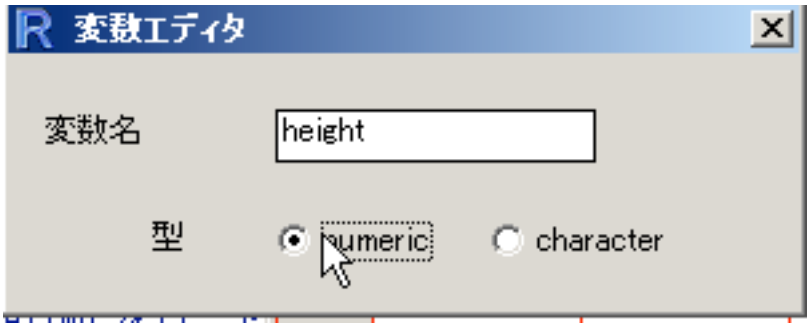


作成するデータセットに名前をつけるように言われますから、そうですね、6 人の体型ですから、「figures6」としましょうか。入力して Ok ボタンを押してください。

すると、表形式の「データエディタ」があらわれます。

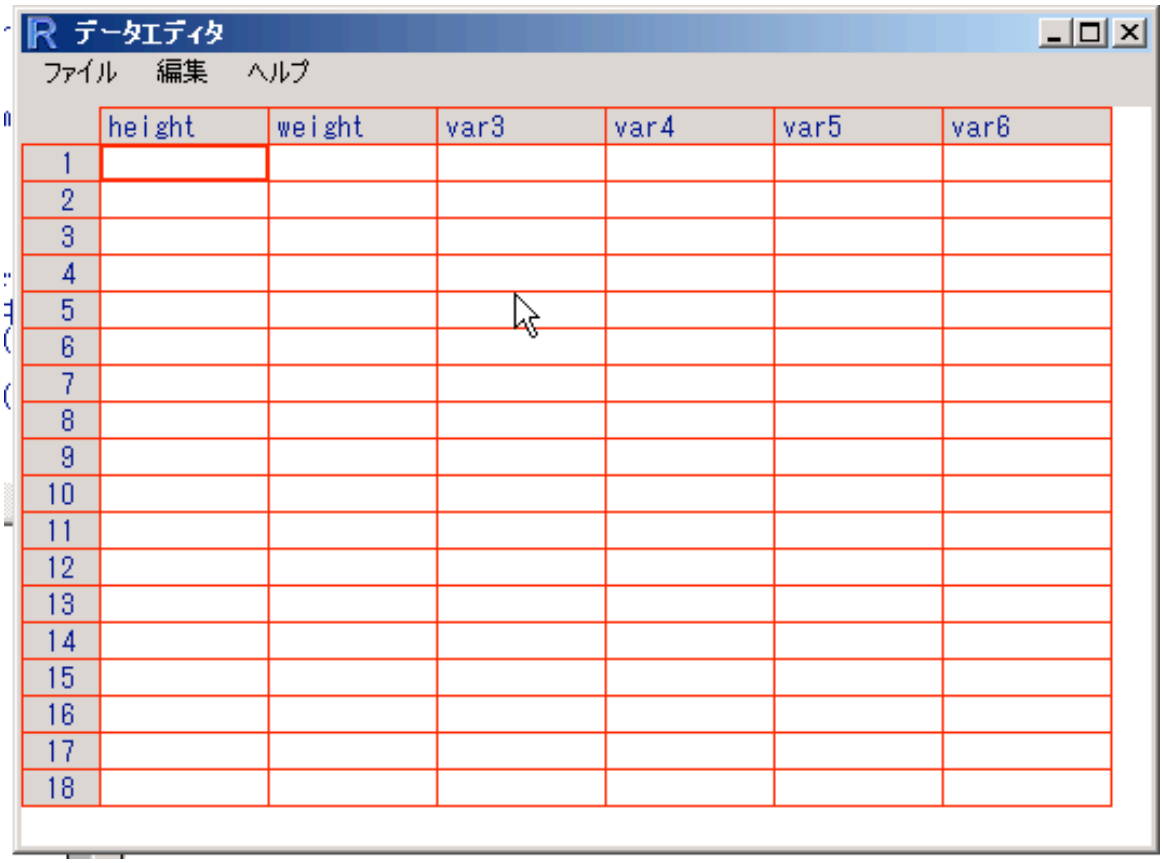


ここに入力をしていきます。Excel と違い、列は A,B,C,...といったものではなく、var1, var2,... となっていますが、ここには任意の名前をつけることができます。ここでは height と weight なので、そのように名付けましょう。var1 のところをクリックしてください。

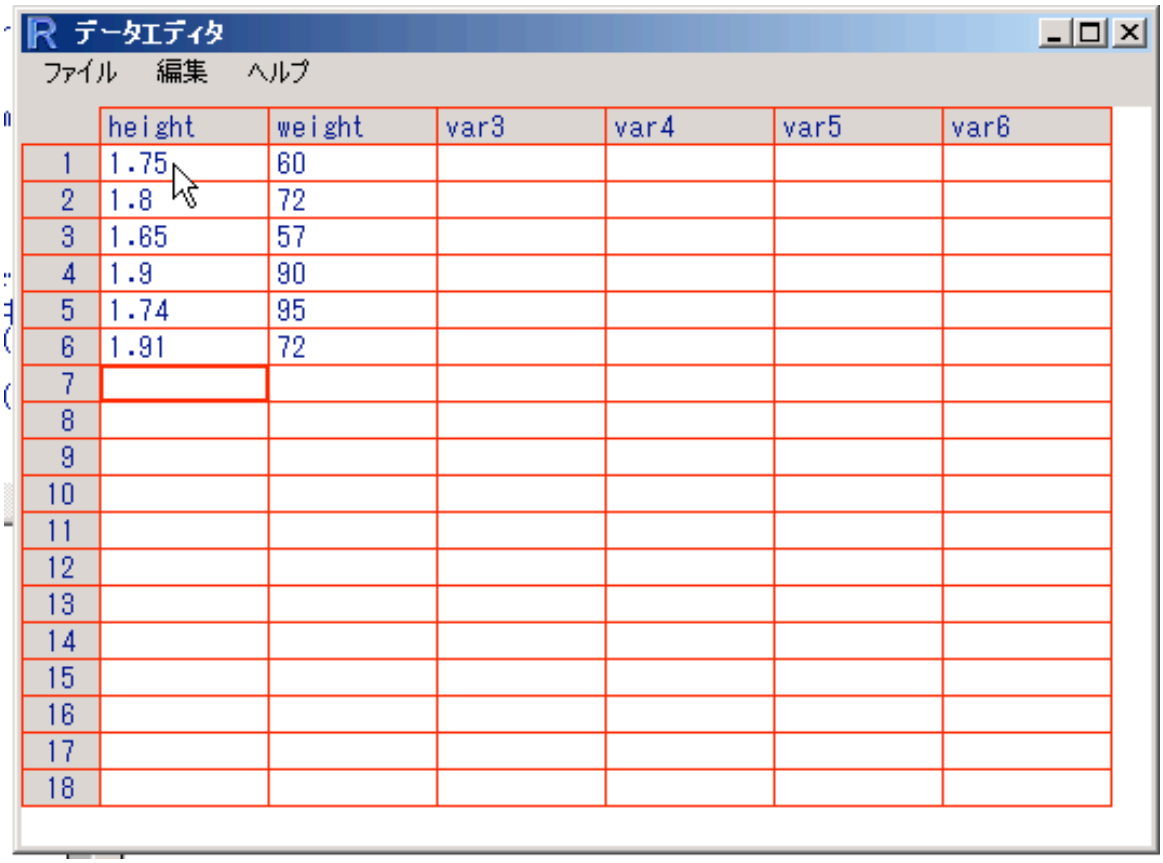


「変数エディタ」があらわれます。**変数名、height を入力し、「型」は「numeric」に変更しましょう。**型の numeric と character は、つまり数値か、文字列か、ということですが、おおむね、連続変数か、カテゴリー値か、ということにそれぞれ対応します。変数エディタは右上の × ボタンで閉じることができます。

同様に、var2 を numeric の「weight」に変更してください。以下のようになりましたか？

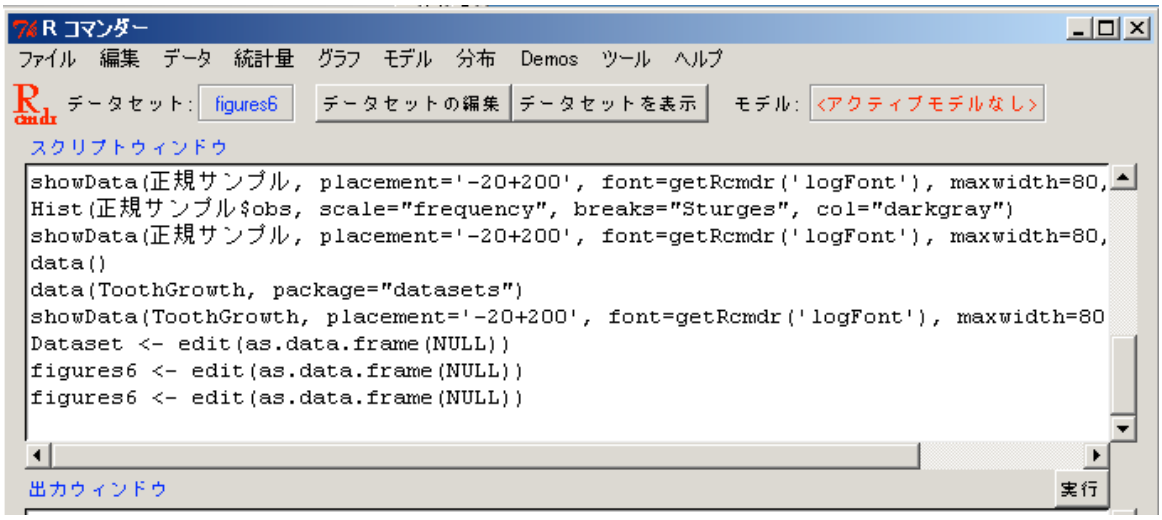


このようになったら、あとは簡単です。次々とデータを入力していきましょう。

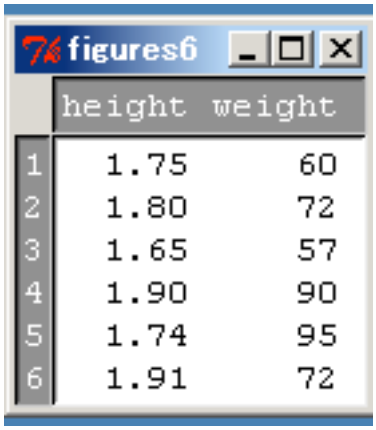


	height	weight	var3	var4	var5	var6
1	1.75	60				
2	1.8	72				
3	1.65	57				
4	1.9	90				
5	1.74	95				
6	1.91	72				
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						

これでデータが完成です。右上の × ボタンでデータエディタを閉じると、コマンダーのデータセットとして「figures6」が入っていることがわかります。



せっかくなので、中身をチェックしてみましょう。「データセットを表示」ボタンを押してみてください。

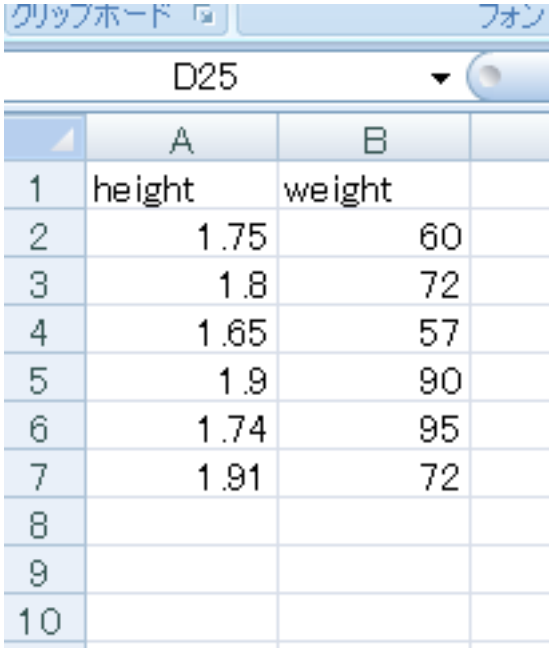


figures6

	height	weight
1	1.75	60
2	1.80	72
3	1.65	57
4	1.90	90
5	1.74	95
6	1.91	72

ちゃんと入力できていますか？

さて、Excel の方も入力ができたでしょうか。R は Windows 版では、Excel の.xls 書類もそのまま開くことができるのですが、Macintosh 版では残念ながらそれができないので、今回はカンマ区切りファイル(.csv ファイル)を経由してファイルを読み込む方法で説明します。これなら Office2007 でも安心です。

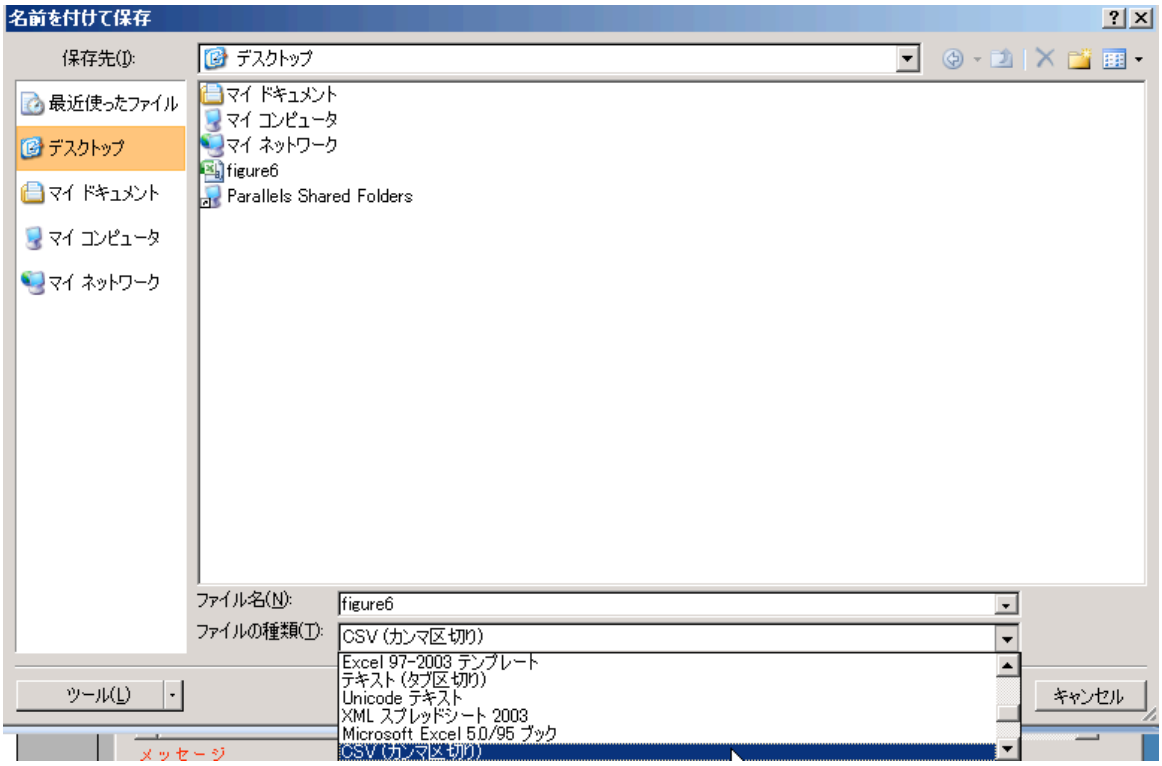


クリップボード　フォン

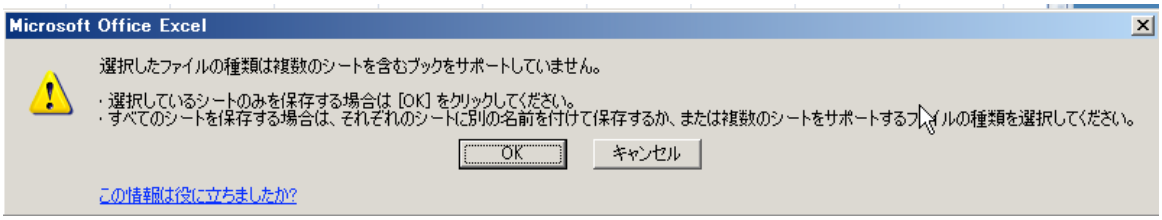
D25

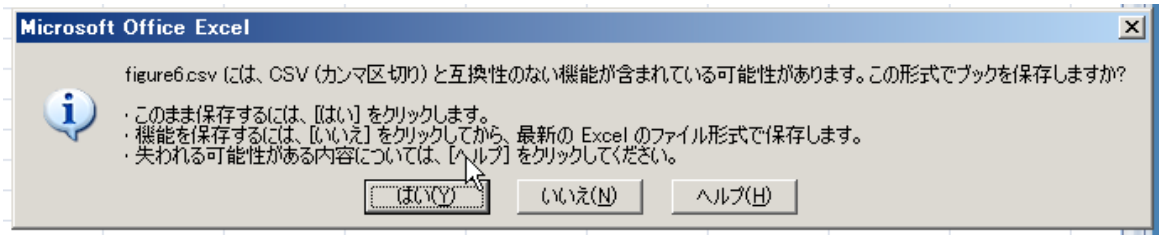
	A	B
1	height	weight
2	1.75	60
3	1.8	72
4	1.65	57
5	1.9	90
6	1.74	95
7	1.91	72
8		
9		
10		

以下のように、１行目に列名をいれて、２行目からデータをいれていきましょう。いれおわったら、保存しますが、このときにちょっと工夫が必要です。



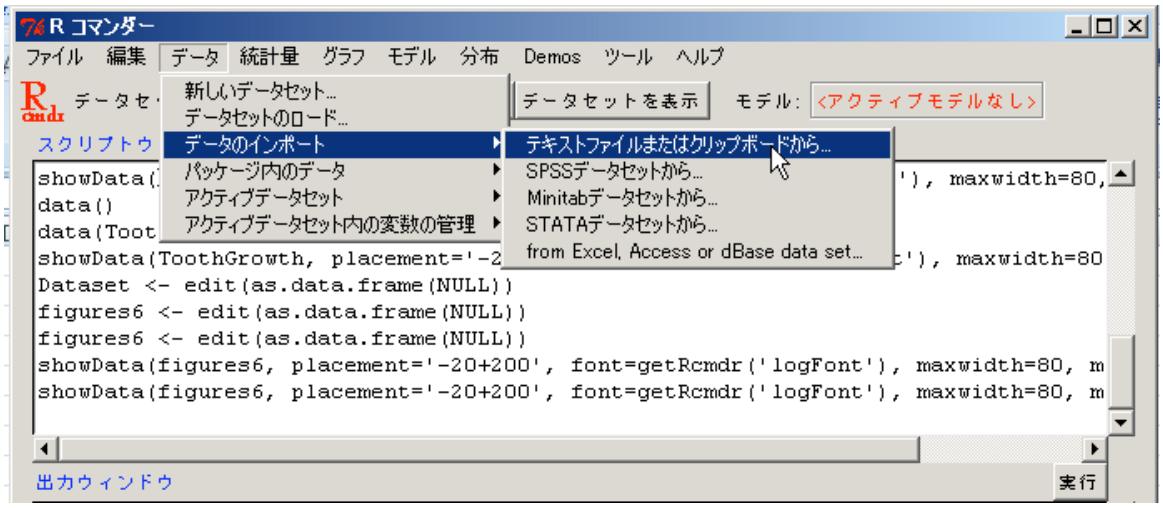
まず、ファイル名は「figure6」としましょう。そして、「ファイルの種類」を選びます。「CSV (カンマ区切り)」としてください。保存する場所はデスクトップでも、マイドキュメントでも、お好きな場所でかまいません。「保存」ボタンを押すと、「選択したファイルの種類は複数のシートを含むブックをサポートしていません」とか、「figure6.csv には CSV と互換性のない機能が...」とか、いろいろと文句をいってきますが、気にせず「はい」をクリックしてしまいましょう。



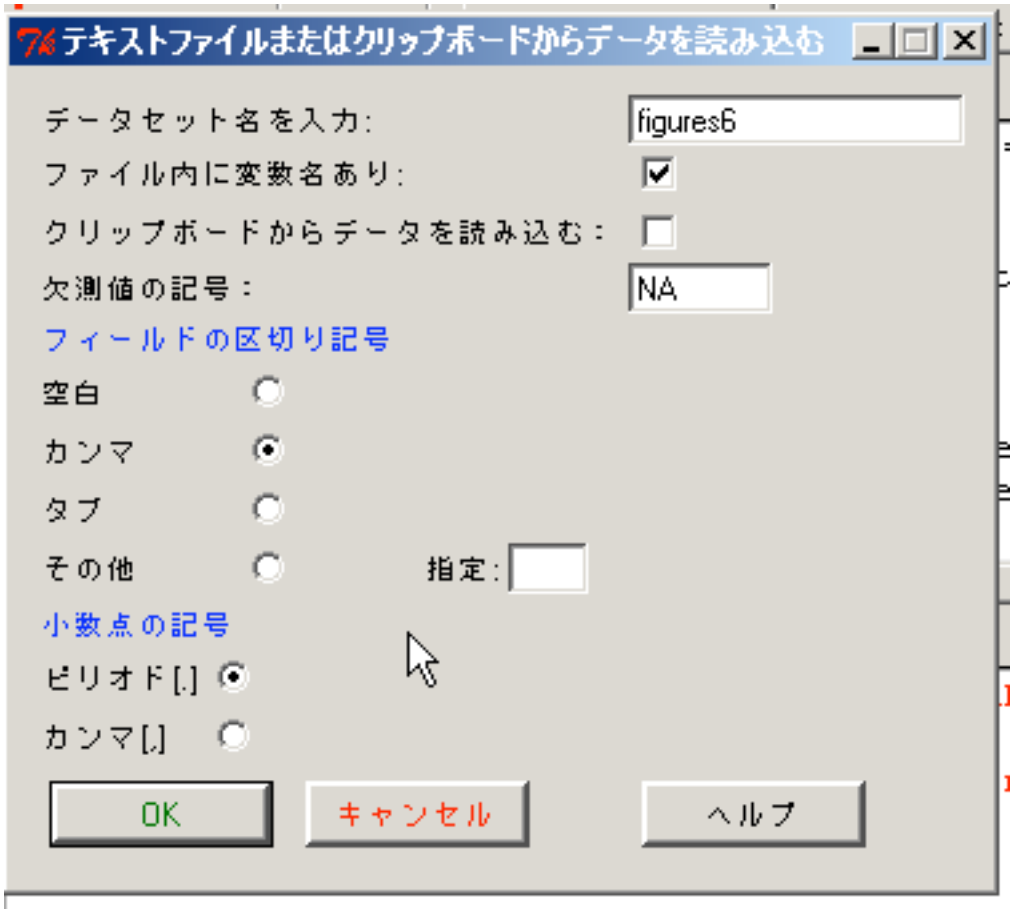


そうすると、無事に保存され、「figure6.csv」というファイルができあがっているはずです。

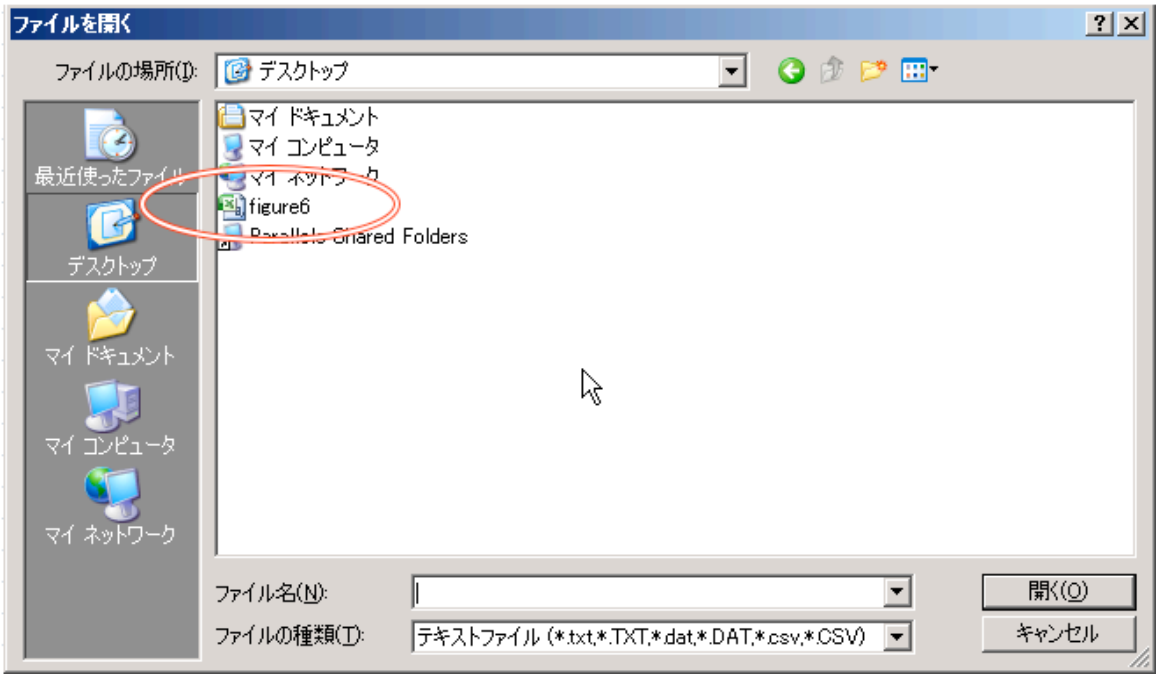
さて、コマンダーに戻りましょう。



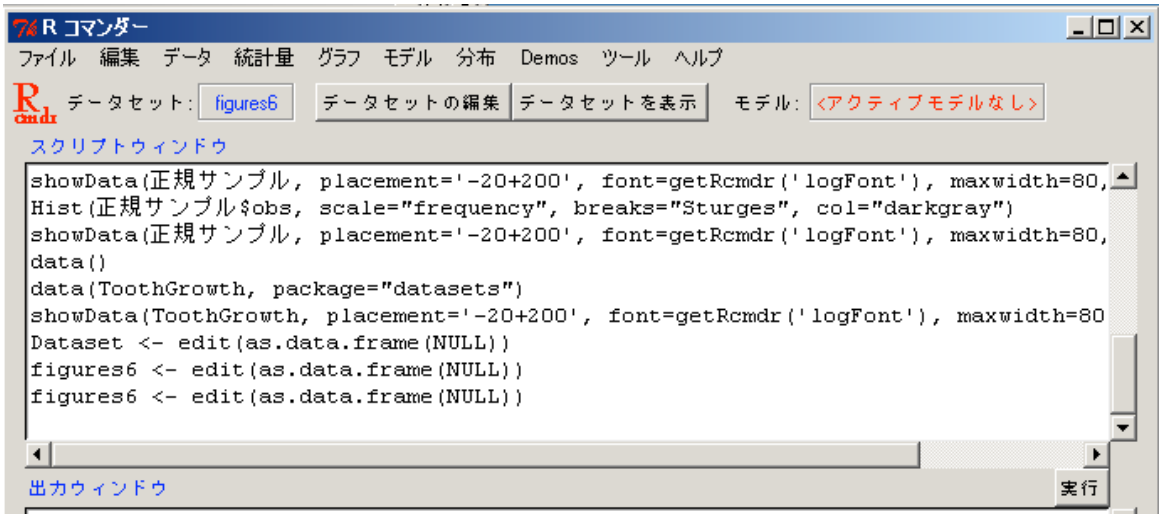
「データ」メニューから「データのインポート」を選び、さらに「テキストファイルまたはクリップボードから...」を選びます。カンマ区切りファイル、つまり Comma-Separated Values ファイルは、カンマでデータを区切ったテキストファイルの一種なのです。



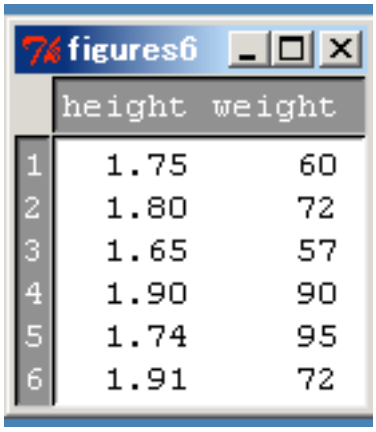
すると、「テキストファイルまたはクリップボードからデータを読み込む」ウィンドウが出ます。ここで指定するべきは、「データセット名」「フィールドの区切り記号」です。ここで指定したデータセット名が、コマンダー内でのデータセットの名前になります。今回は「figures6」としておきましょう。フィールドの区切り記号は、まさにデータを区切っている文字です。カンマ区切りファイルを利用したので、「カンマ」を選びましょう。変数名(height, weight)を含めて入力したので、「ファイル内に変数名あり」がチェックされていることも確認しましょう。そして、「Ok」ボタンです。



「ファイルを開く」ウィンドウが出ます。さきほど保存した「figure6」を選び、「開く」を押しましょう。



コマンダーウィンドウの左上の「データセット」欄に、「figure6」が表示されましたね。「データセットを表示」ボタンを押して、表示してみてください。

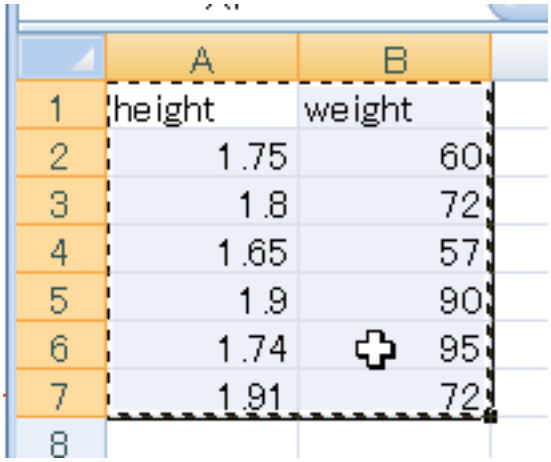


figures6

	height	weight
1	1.75	60
2	1.80	72
3	1.65	57
4	1.90	90
5	1.74	95
6	1.91	72

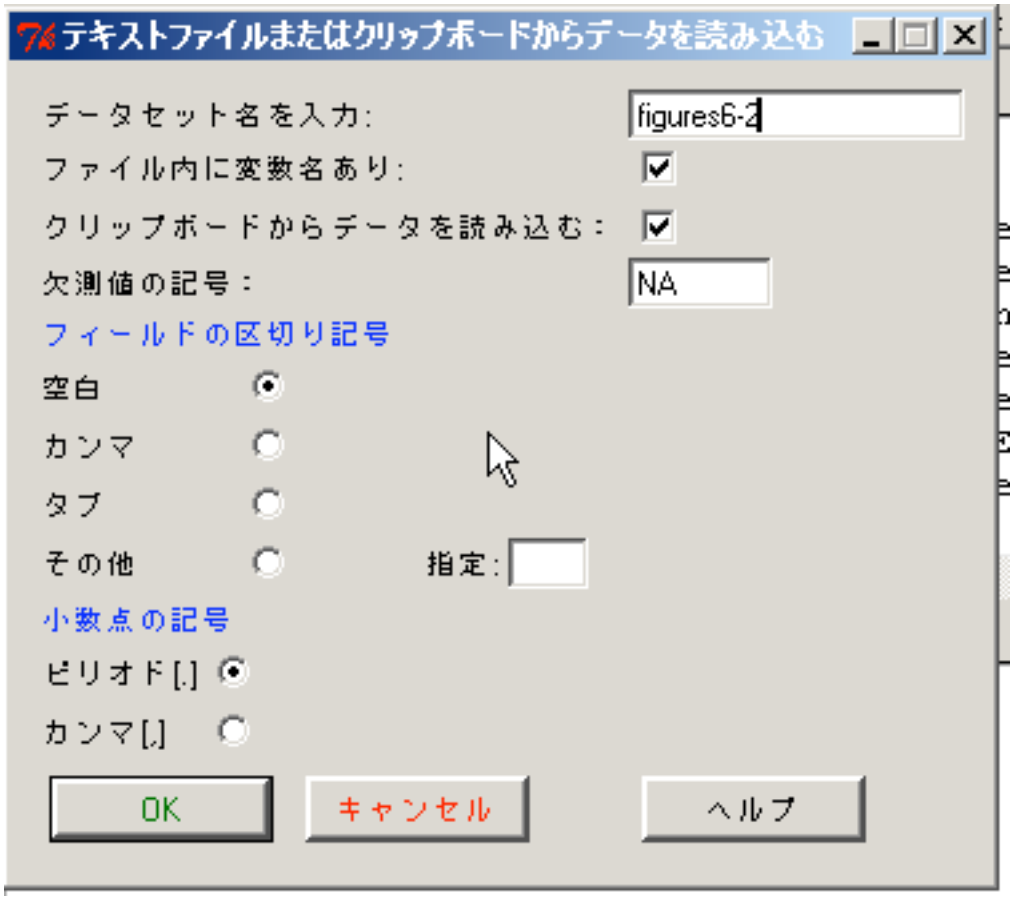
左図のように、ちゃんと読み込まれているでしょうか。

さらに、この応用として、一度 Excel からカンマ区切りファイルで保存をしなくても、Excel 上で「コピー」したデータを読み込んでくることも可能です。



	A	B
1	height	weight
2	1.75	60
3	1.8	72
4	1.65	57
5	1.9	90
6	1.74	95
7	1.91	72
8		

データのインポートを選び、「テキストファイルまたはクリップボードからデータを読み込む」ウィンドウを出すところまでは一緒ですが、ここで



「クリップボードからデータを読み込む」をチェックして、「フィールドの区切り記号」を「空白」にすれば Ok です。試してみてください。

さて、これでデータを読み込むことができました。このデータを用いてさまざまな解析を行っていきましょう。