

R 入門

Rとは

- データ解析のためのプログラミング言語
- S言語をもとに開発された
- フリーソフト
- Windows, Mac OS, Linux で動作
- RStudioと組み合わせると便利
 - RのIDE(統合開発環境)
 - RstudioもWindows版、Mac OS版、Linux版がある

Rのインストール (1)

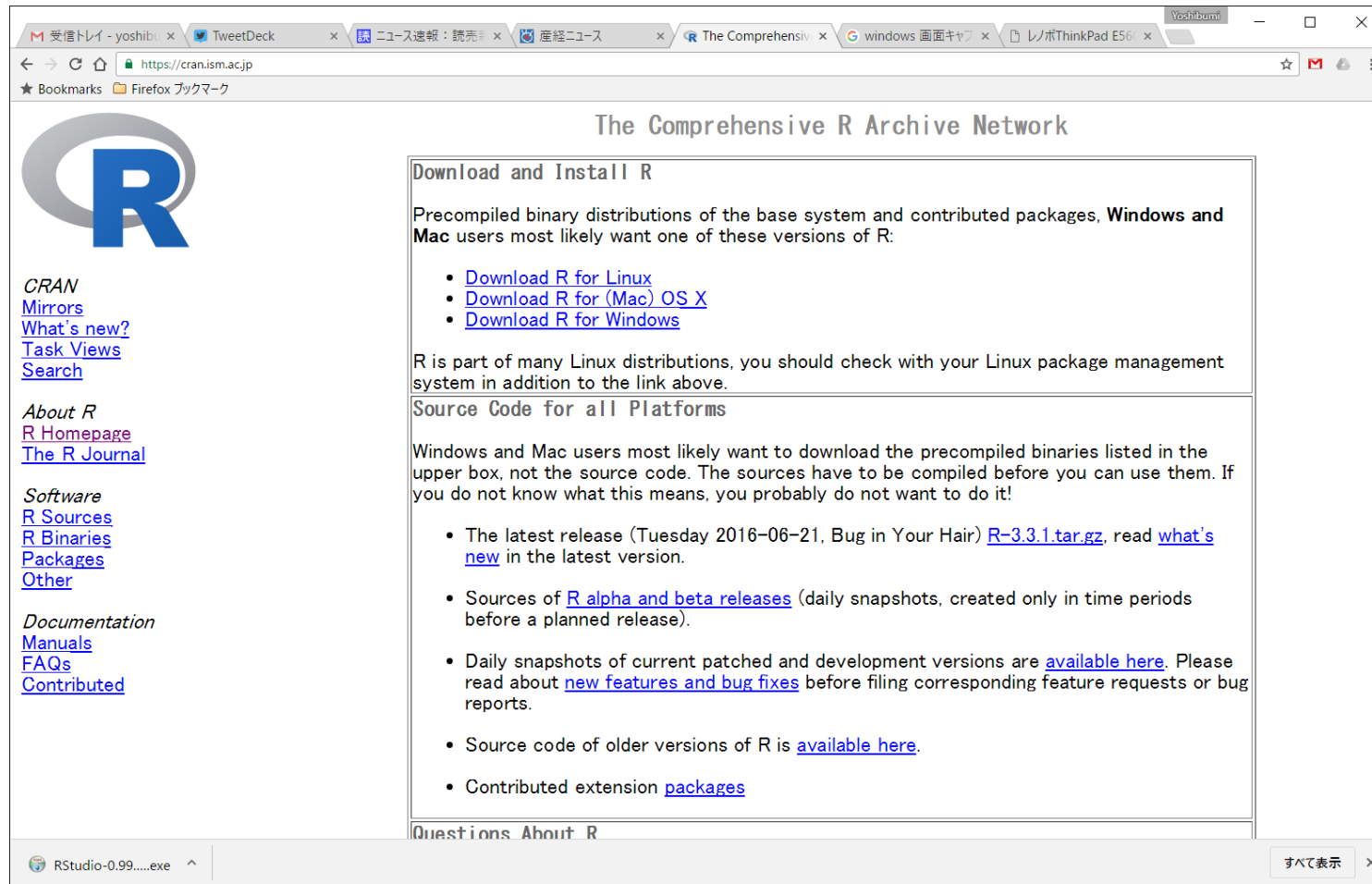
CRANからダウンロード
<https://cran.ism.ac.jp/>

(サイトが見つからない場合はCRANで検索すること)

自分の環境にあったRをダウンロードする

(Windows版, Mac版, Linux版)

Rは頻繁にアップデートされています。新しい機能が付け加わったり、バグが修正されている場合があるので、新しいバージョンが公開されているか注意が必要。



The screenshot shows a web browser window displaying the CRAN website. The browser's address bar shows the URL <https://cran.ism.ac.jp/>. The page title is "The Comprehensive R Archive Network". The main content area is titled "Download and Install R" and contains the following text:

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.

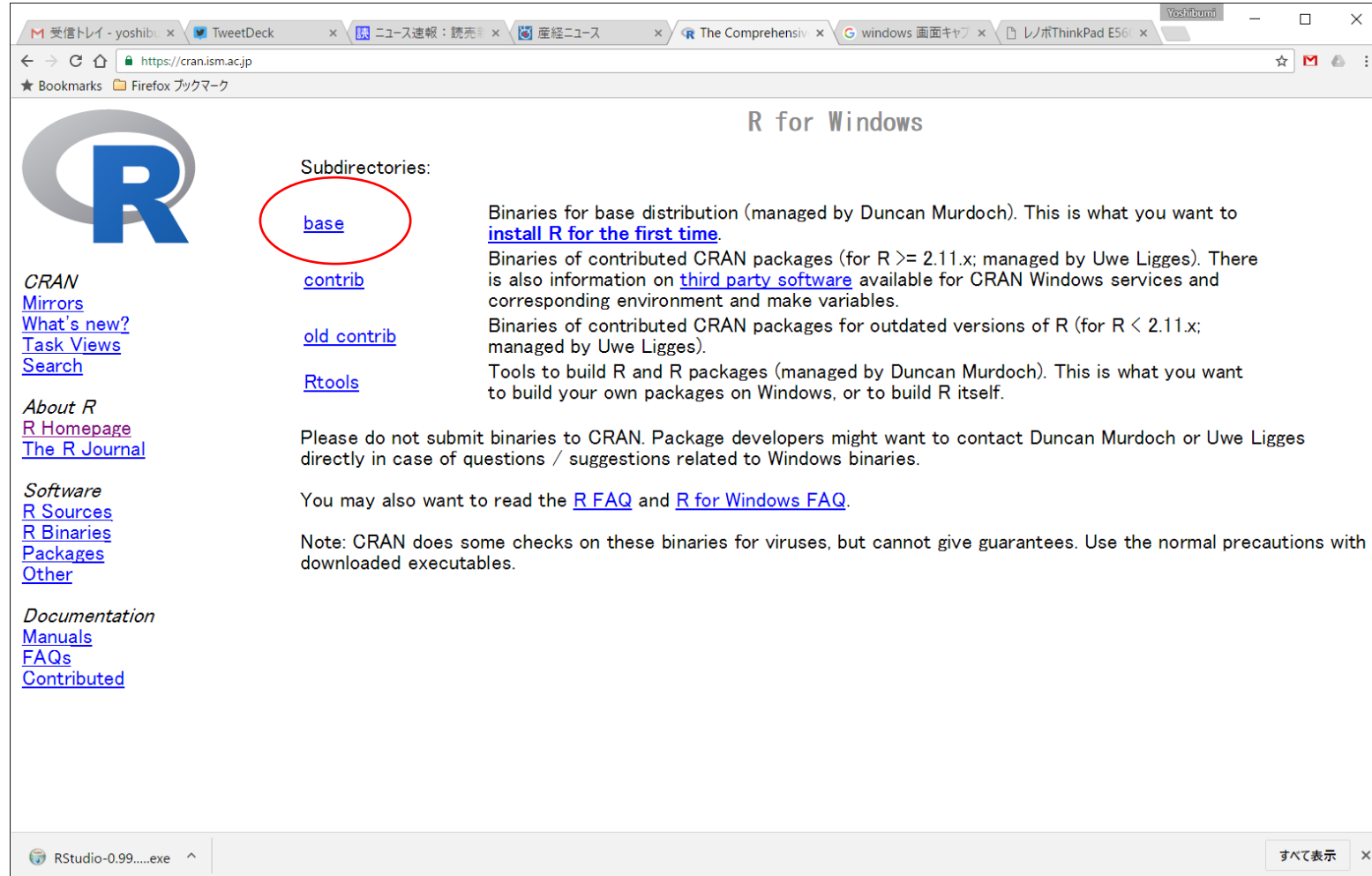
Source Code for all Platforms

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 (Tuesday 2016-06-21, Bug in Your Hair) [R-3.3.1.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](#)

At the bottom of the page, there is a link for "Questions About R". The browser's taskbar at the bottom shows "RStudio-0.99.....exe" and a "すべて表示" button.

Rのインストール (2)

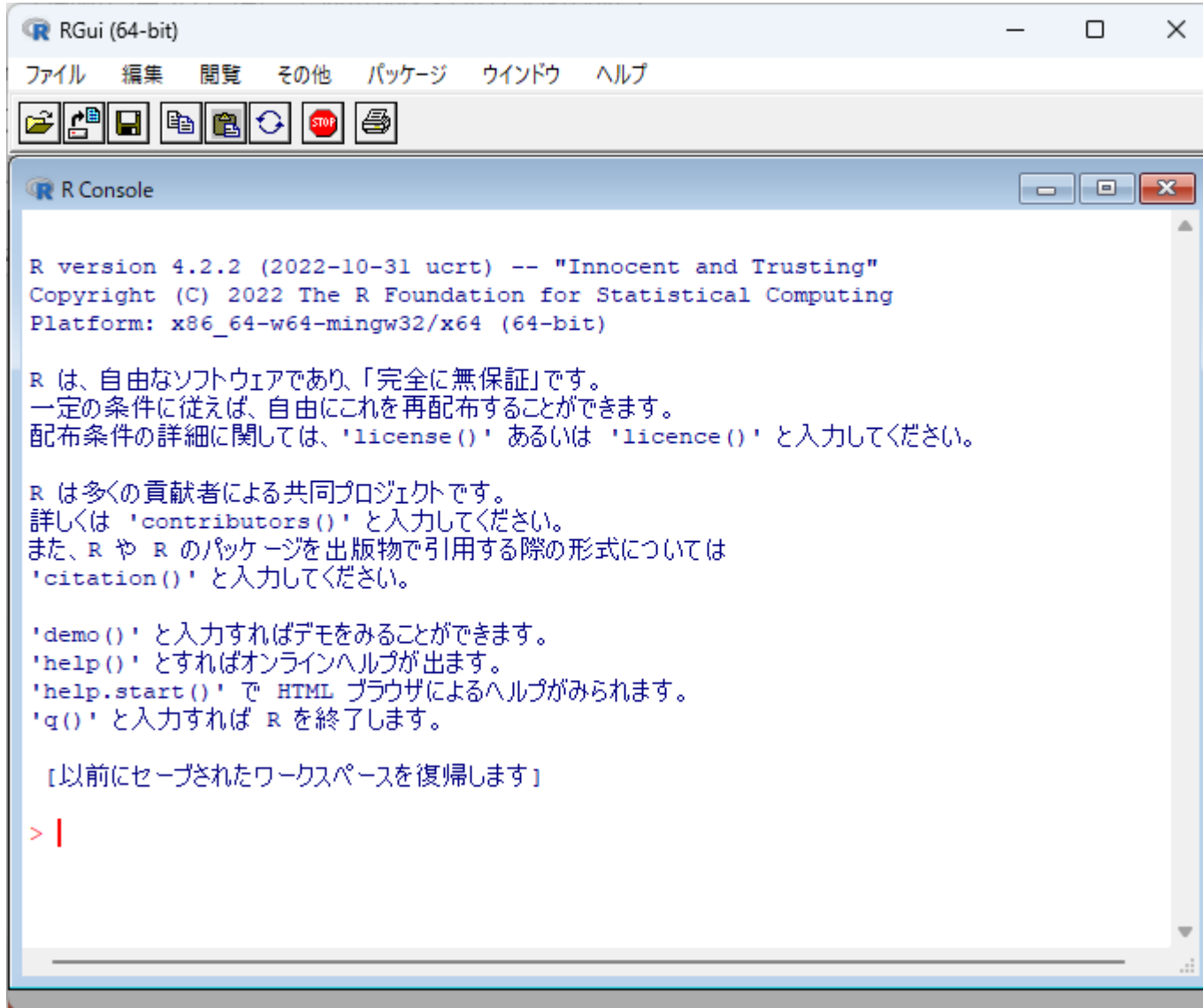


The screenshot shows the CRAN website for Windows. The page title is "R for Windows". On the left side, there is a navigation menu with links for "CRAN", "Mirrors", "What's new?", "Task Views", "Search", "About R", "R Homepage", "The R Journal", "Software", "R Sources", "R Binaries", "Packages", "Other", "Documentation", "Manuals", "FAQs", and "Contributed". In the center, under the heading "Subdirectories:", there are four links: "base", "contrib", "old contrib", and "Rtools". The "base" link is circled in red. To the right of these links, there are descriptions for each subdirectory. The "base" description states: "Binaries for base distribution (managed by Duncan Murdoch). This is what you want to **install R for the first time**." The "contrib" description states: "Binaries of contributed CRAN packages (for R >= 2.11.x; managed by Uwe Ligges). There is also information on **third party software** available for CRAN Windows services and corresponding environment and make variables." The "old contrib" description states: "Binaries of contributed CRAN packages for outdated versions of R (for R < 2.11.x; managed by Uwe Ligges)." The "Rtools" description states: "Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or to build R itself." Below the subdirectories, there is a note: "Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries." and another note: "You may also want to read the [R FAQ](#) and [R for Windows FAQ](#)." At the bottom of the page, there is a note: "Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables." The browser's address bar shows "https://cran.ism.ac.jp". The browser's taskbar at the bottom shows "RStudio-0.99.....exe" and a "すべて表示" button.

これはWindows版
Baseを選んでダウンロード

Mac OSの場合は
Rx.x.x.pkgというファイル(x.x.xはバージョン番号)
なお、これとは別に
Xquartzのインストールが必要(詳しくはRのダウンロードサイトで)

Rのインストール (3)



ダウンロードしたファイルをクリック
してインストール
左がRの起動画面

Rのインストール 注意

- ベースのRに加え，分析手法に応じた拡張パッケージのインストールが必要になります。
- Windowsの場合，ログインユーザー名を漢字やカナの2byte文字にしているとパッケージがうまくインストールできない可能性あり。また，ログインユーザーのDocumentsフォルダーがクラウド(OneDrive)になっている場合もうまくインストールできない可能性あり。
- → ログインユーザー名は1byte文字（英数字）で
- R4.2.0以降，ライブラリーのインストール先は下記の通り（Windowsの場合）

C:¥Users¥xxx¥AppData¥Local¥R¥win-library

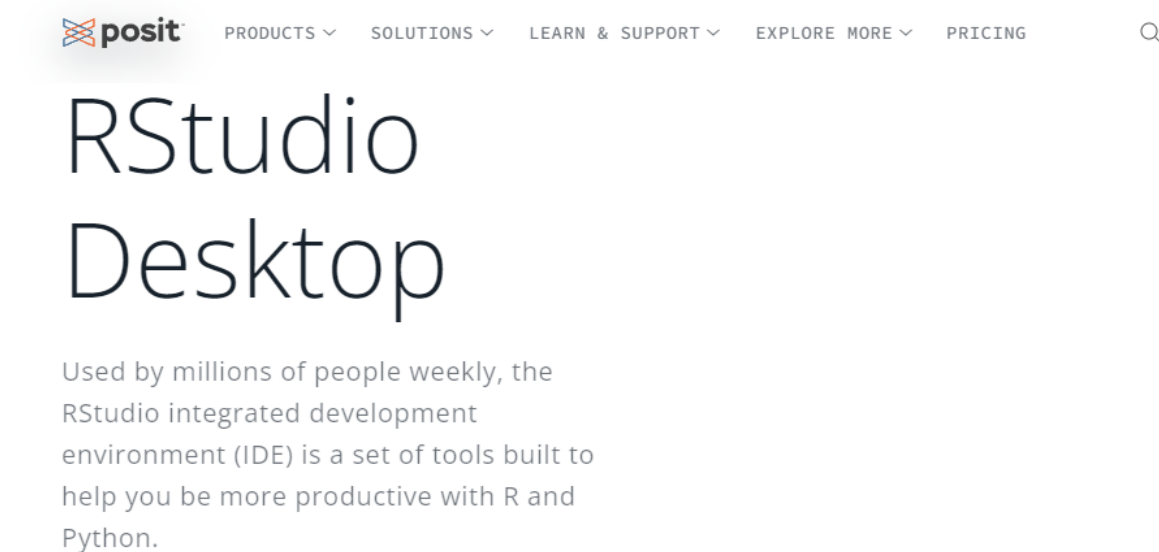
以前はドキュメントフォルダー以下にRのパッケージをインストールするフォルダーが作られていて，そのフォルダー名に2byte文字が使われているとエラーがでていた（Rのwindows版が2byte文字を処理できないため）。

- R4.2.0以降，OneDriveがホームディレクトリーでもトラブルは無いと言われてますが，ローカルアカウントでログインした方が安全だと思われれます。次のページを参照してください

<https://okumuralab.org/~okumura/stat/R-win.html>

- macの場合，パッケージは /Library/Frameworks/ 以下のフォルダーにインストールされます（隠しファイルになっているかもしれません）

Rstudio のインストール



The screenshot shows the top navigation bar of the RStudio website with links for PRODUCTS, SOLUTIONS, LEARN & SUPPORT, EXPLORE MORE, and PRICING. Below the navigation is the RStudio logo and the text 'RStudio Desktop'. A paragraph describes RStudio as an integrated development environment (IDE) used by millions of people weekly, built to help users be more productive with R and Python.

次のページからRstudioをダウンロードしてインストールする

<https://posit.co/download/rstudio-desktop/>

Windows版, Mac版, Linux版がある

1: Install R

RStudio requires R 3.3.0+. Choose a version of R that matches your computer's operating system.

[DOWNLOAD AND INSTALL R](#)

2: Install RStudio

[DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS](#)

Rstudio

開いているデータファイルのビューアー

コンソール画面

The screenshot shows the RStudio interface with the following components:

- Data Viewer:** A table with columns: wage, education, experience, ethnicity, smsa, region, parttime. It shows 16 rows of data.
- Environment Pane:** Shows the loaded data object 'CPS1988' with 28155 observations and 7 variables.
- Console:** Displays the R startup message, including the license and help information.

	wage	education	experience	ethnicity	smsa	region	parttime
1	364.94	7	45	cauc	yes	northeast	no
2	123.46	12	1	cauc	yes	northeast	yes
3	370.37	9	9	cauc	yes	northeast	no
4	754.94	11	46	cauc	yes	northeast	no
5	593.54	12	36	cauc	yes	northeast	no
6	377.23	16	22	cauc	yes	northeast	no
7	284.90	8	51	cauc	yes	northeast	no
8	561.13	12	34	cauc	yes	northeast	no
9	264.06	12	0	cauc	yes	northeast	no
10	1643.83	14	18	cauc	yes	northeast	no
11	474.83	12	17	cauc	yes	northeast	no
12	209.15	8	42	cauc	yes	northeast	no
13	244.88	10	10	cauc	yes	northeast	no
14	474.83	14	19	cauc	no	northeast	no
15	213.68	12	40	cauc	no	northeast	no
16	864.20	16	42	cauc	no	northeast	no

```
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
[Workspace loaded from ~/.RData]  
> |
```

素のRのコンソール画面と同じ

ここにコードをタイプする

結果はこの画面に (テキストファイルのもの)

グラフ等の出力は右下の windowに出力される

フォント等の設定は Tools/Global Options で

Packagesのダウンロードサイトも Tools/Global Optionsの Packgaesで指定する (日本のサイトを指定した方が良い)

パッケージのインストール

AERというパッケージをインストールする
(AER は C.Kleiber and A.Zeileis,
“Applied Econometrics with R” , Springer,
2008 より)

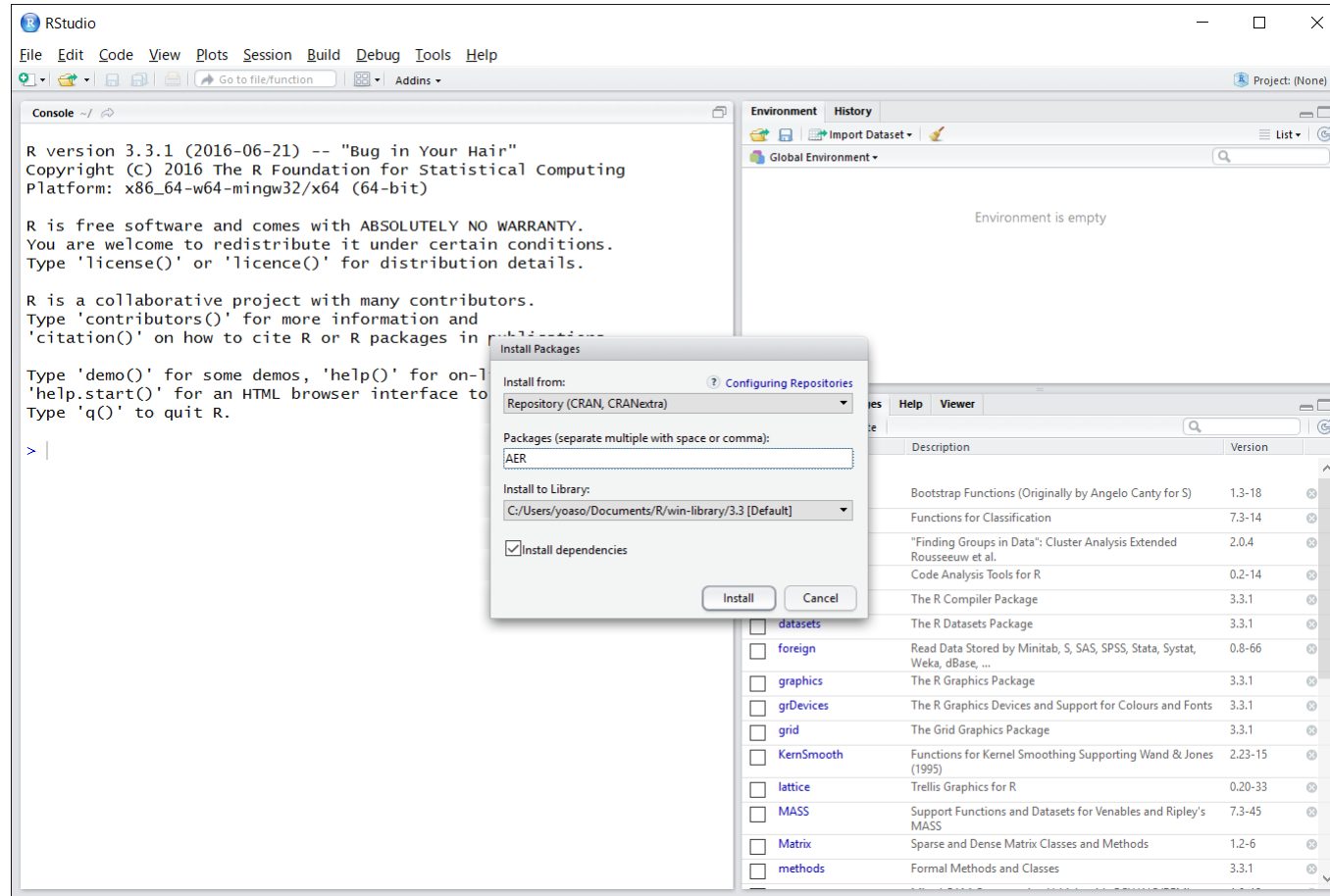
コンソール画面で

Install.packages(“AER”)

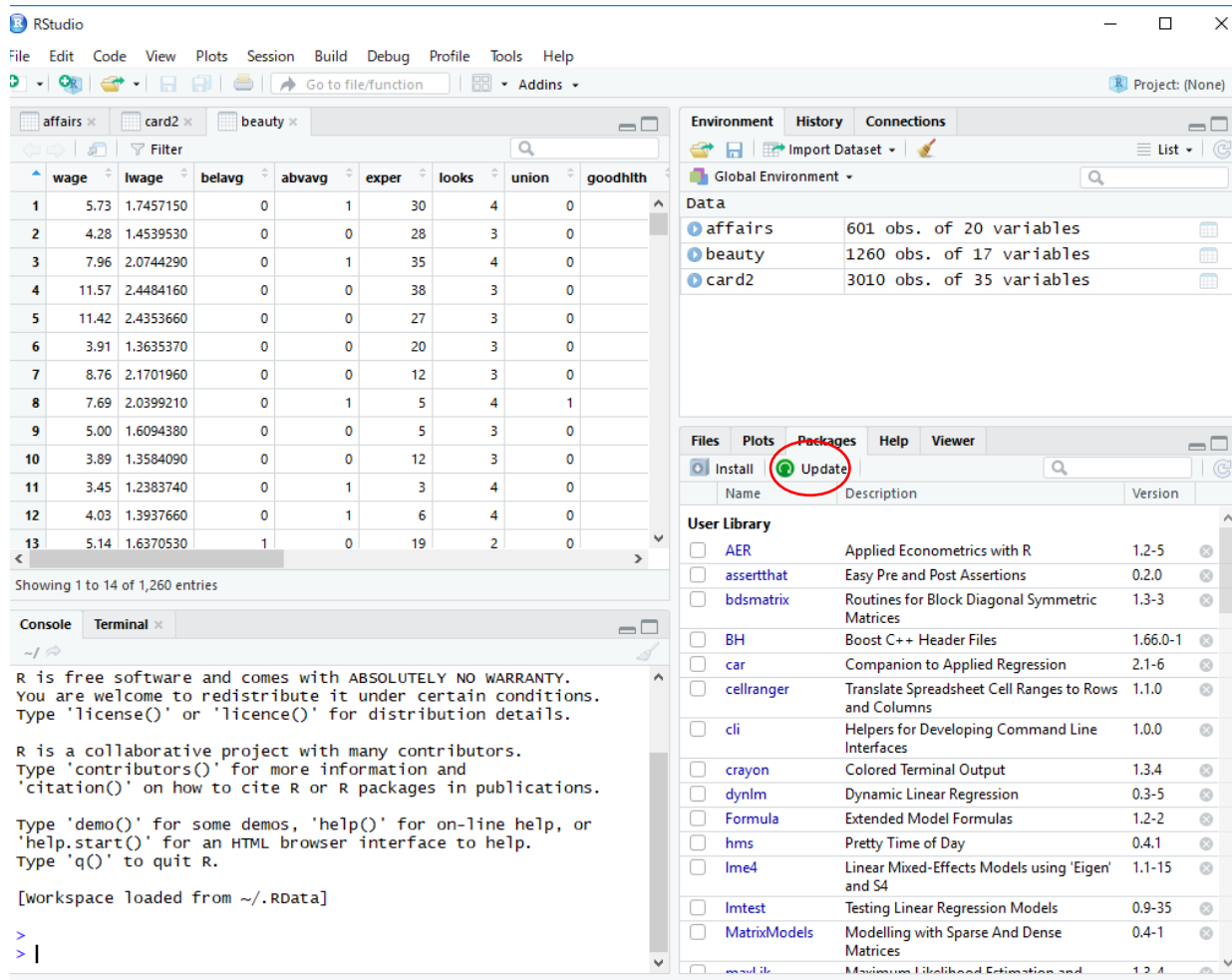
とタイプするか、Rstudioの右下window
のタブでpackagesを選び、さらにinstall
を選択すると左のような画面がポップ
アップする→ Packagesの欄にAERとタイ
プする

インストールしたパッケージがどのフォル
ダーにあるかを調べる。コマンドライ
ンで

library()
とタイプする



パッケージの更新



パッケージの更新（パッケージのinstall直後は不要）

右下のwindowのPackageタブからUpdateを選択すると、Packageの更新ができる

Rstudio 自体の更新

メニューから

Help/Check for Updates

を選択。新しいversionが出ていればRstudioのホームページからダウンロードしてインストール

R本体→CRANで確認

パッケージの更新(2)

- パッケージのインストールに失敗する場合
 - パッケージをソースコードからコンパイルする必要がある場合、windows環境ではRtoolsのインストールが必要になる。
 - → R studio のエラーメッセージにしたがってRtoolsをダウンロードしてインストール
 - その際、パッケージのインストールされているフォルダーを一旦削除してから、新しいパッケージをインストールしないとうまくインストールできない場合あり
- パッケージのインストールされているフォルダー
 - windows C:¥Users¥xxx¥AppData¥Local¥R¥win-library
 - (R4.2.0以前) C:¥Users¥xxx¥Documents¥R
 - mac /Library/Frameworks/

Rの作業用フォルダー

- 作業用フォルダー(working directory)の確認方法
RまたはRstudioのコマンドラインで次のようにタイプすればわかります
getwd()
帰ってきたフォルダー名に漢字や仮名が含まれていれば、フォルダーに2byte文字が使われています
英数字(半角の英数字) なら問題ありません
- 作業用フォルダに2byte文字が使われている場合、いったんRとRstudioをアンインストールし、1byte文字の新しいログインアカウントを作成し、そのアカウントでRとRstudioをインストールするのがもっとも簡単です。
 - なお、旧アカウントのDocumentsフォルダー内のファイルをバックアップしておいて、新アカウントに移した方がよいでしょう
 - 新しいアカウントの作成については、ネットで調べてください

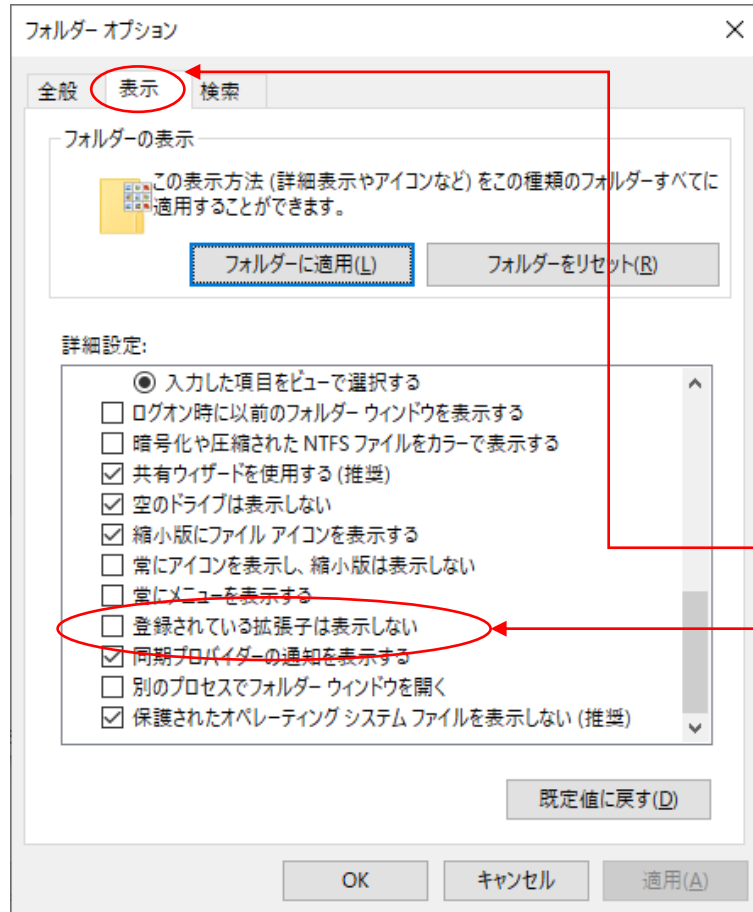
データのimport

- テキストデータ
 - CSVファイル
 - comma separated value: 最もよく使われる形式
 - , でデータの区切り; 改行でオブザベーションの区切り
 - 第1行目に変数名があると, 多くのソフトで変数名も読み取ってくれる
 - データの区切りがカンマではなく, スペースだったり, タブだったりする場合もあり
 - 固定長ファイル
 - 第xカラムから第yカラムまでに変数zの値が格納されているという形式のファイル
- Excel データ
 - 最近のソフトでは, Excelファイルを読み込んでくれるものも多くなってきた
 - Excel → CSVの変換は容易

データのimport (2)

- CSVデータ等を扱う際は，改行やタブコードが画面に表示されるエディタを使った方が便利
 - サクラエディタ，秀丸 (Windows)
 - CotEditor (Mac)
 - いずれもOS付属の標準ソフトではないのでインストールが必要
 - Excelでも扱える
- エクスプローラまたはファイnderで，ファイルの拡張子が表示されるように設定しておく
 - Windowsの場合：次ページ参照
 - Macの場合：ファイnderの環境設定から詳細タブを選択し，「すべてのファイル名拡張子を表示」にチェックを入れる

ファイルの拡張子を表示する方法(windows)



- エクスプローラのメニューから「表示」を選択
- 「オプション」リボンを選択
- 「フォルダーオプション」という左のような画面がポップアップ
- 「表示」タブを選択
- 「登録されている拡張子は表示しない」のチェックをはずす

データのimportの実際

- wooldridge のデータセット
 - wage1.des データセットの説明：テキストファイル
 - wage1.raw データセット本体：テキストファイル（本体に変数名は含まれていない）
 - wage1.xls データセット本体：excelファイル
- テキストエディタは、タブや改行等の文字コードが表示され、折り返し桁数の指定ができるものがベター（windows付属のメモ帳ではやや不便）
- 注意
 - エクスプローラで拡張子を表示するようにしておく
 - エクスプローラのメニューから 表示/オプション/ →フォルダーオプション →表示 タブを選択し、詳細設定の項目で「登録されている拡張子は表示しない」のチェックをはずす（Windows10の場合）

wage1.des の内容

```
1 WAGE1.DES
2
3 wage      educ      exper      tenure     nonwhite   female     married    numdep
4 smsa      northcen  south      west       construc  ndurman   trcompu    trade
5 services  profserv  profocc    clerocc    servocc    lwage     expersq    tenursq
6
7 Obs:      526
8
9 1. wage      average hourly earnings
10 2. educ      years of education
11 3. exper     years potential experience
12 4. tenure    years with current employer
13 5. nonwhite  =1 if nonwhite
14 6. female    =1 if female
15 7. married   =1 if married
16 8. numdep    number of dependents
17 9. smsa      =1 if live in SMSA
18 10. northcen =1 if live in north central U.S.
19 11. south    =1 if live in southern region
```

5行 78桁 CRLF 20 SJIS REC 挿入

wage1.raw の内容

先頭行に変数名が含まれていない

0	1	2	3	4	5	6	7	8
1	3.10	11	2	0	0	1	0	2
2	3.24	12	22	2	0	1	1	3
3	3.00	11	2	0	0	0	0	2
4	6.00	8	44	28	0	0	1	0
5	5.30	12	7	2	0	0	1	1
6	8.75	16	9	8	0	0	1	0
7	11.25	18	15	7	0	0	0	0
8	5.00	12	5	3	0	1	0	0
9	3.60	12	26	4	0	1	0	2
10	18.18	17	22	21	0	0	1	0
11	6.25	16	8	2	0	1	0	0
12	8.13	13	3	0	0	1	0	0
13	8.77	12	15	0	0	0	1	2
14	5.50	12	18	3	0	0	0	0
15	22.20	12	31	15	0	0	1	1
16	17.33	16	14	0	0	0	1	1
17	7.50	12	10	0	0	1	1	0
18	10.63	13	16	10	0	1	0	0
19	3.60	12	13	0	0	1	1	3

この画面は、1行を折り返さないように表示している。
●秀丸で同様にするには、メニューから表示/折り返し/最大とする

wage1.xlsの内容

先頭行に変数名は含まれていない

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	3.1	11	2	0	0	1	0	2	1	0	0	1	0	0	
2	3.24	12	22	2	0	1	1	3	1	0	0	1	0	0	
3	3	11	2	0	0	0	0	2	0	0	0	1	0	0	
4	6	8	44	28	0	0	1	0	1	0	0	1	0	0	
5	5.3	12	7	2	0	0	1	1	0	0	0	1	0	0	
6	8.75	16	9	8	0	0	1	0	1	0	0	1	0	0	
7	11.25	18	15	7	0	0	0	0	1	0	0	1	0	0	
8	5	12	5	3	0	1	0	0	1	0	0	1	0	0	
9	3.6	12	26	4	0	1	0	2	1	0	0	1	0	0	
10	18.18	17	22	21	0	0	1	0	1	0	0	1	0	0	
11	6.25	16	8	2	0	1	0	0	1	0	0	1	0	0	
12	8.13	13	3	0	0	1	0	0	1	0	0	1	0	0	
13	8.77	12	15	0	0	0	1	2	1	0	0	1	0	0	
14	5.5	12	18	3	0	0	0	0	1	0	0	1	0	0	
15	22.2	12	31	15	0	0	1	1	1	0	0	1	0	0	
16	17.33	16	14	0	0	0	1	1	1	0	0	1	0	0	
17	7.5	12	10	0	0	1	1	0	1	0	0	1	0	0	
18	10.63	13	16	10	0	1	0	0	1	0	0	1	0	0	

- 以下では次の方法でデータを読み込む

1. excel ファイルからのimport

2. テキストファイルからのimport

以下で用いるファイルは、先頭行に変数名が入っていない。Importの後に変数名を指定することもできるが面倒 → あらかじめ、先頭行に変数名を入れてからimportする

一般的には、データセットをexcelで管理し、先頭行に変数名を含めておく
変数の説明は別のシートに記入しておく

excelファイルが読めないソフトでも、CSVファイルに変換して読むことができる

Excel ファイルのimport

wage1.xls

(変数名は入っていない)

wage1.desの変数名をexcelの別のシートにコピー&ペーストで貼り付ける(この段階ではA列に変数名が入っていて、B列以降には何も入っていない)。これを空白行でフィールドに分解する

→メニューから データ/区切り位置で分解する

数値データのみのシートに戻り、1行目に空白行の挿入
その空白行に変数名の行をコピーして貼り付ける

The top screenshot shows an Excel spreadsheet with the following data in column A:

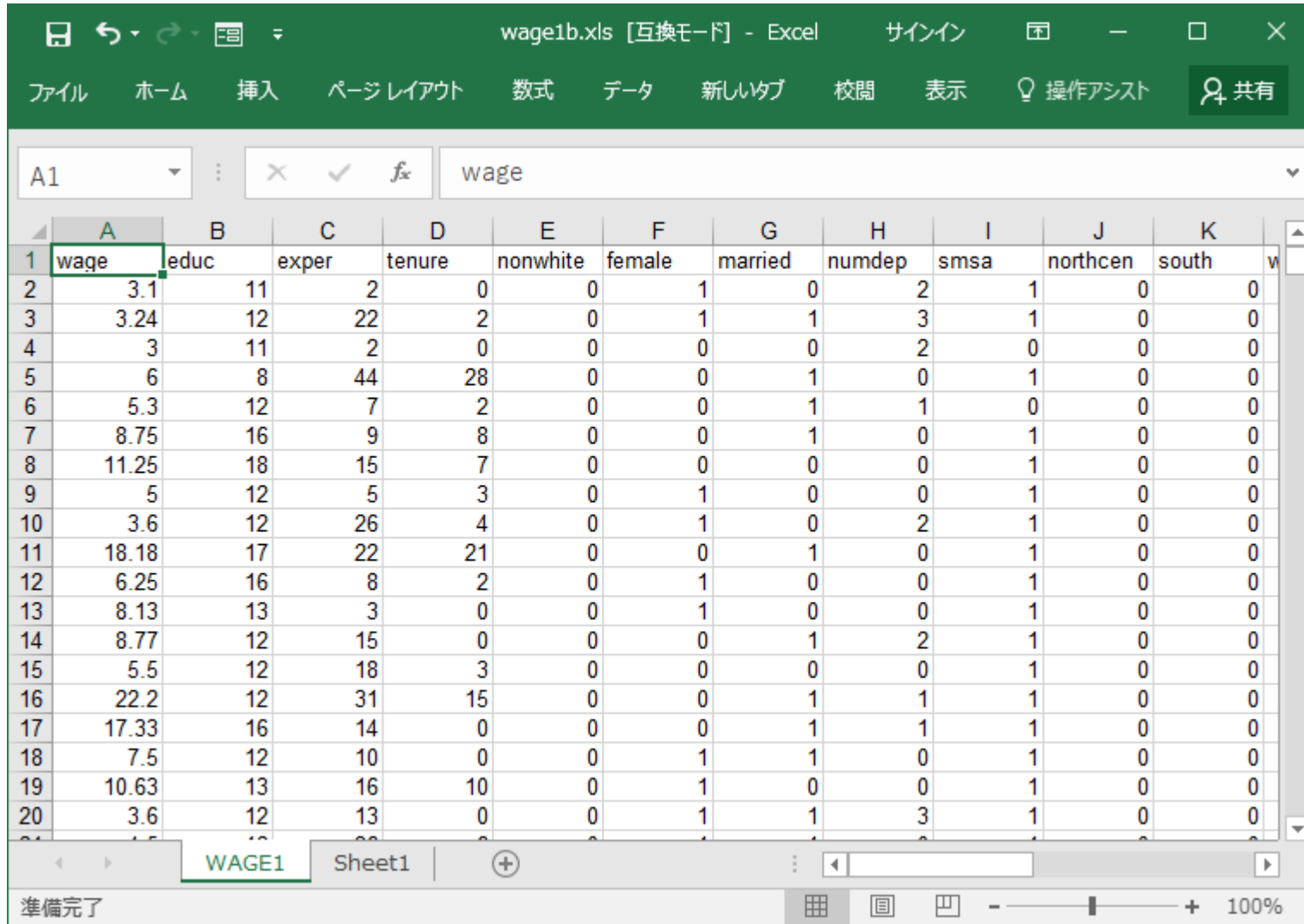
Row	Column A
1	3.1
2	3.24
3	3
4	6
5	5.3
6	8.75
7	11.25
8	5
9	3.6
10	18.18
11	6.25
12	8.13
13	8.77
14	5.5
15	22.2
16	17.33
17	7.5
18	10.63
19	3.6
20	4.5

The bottom screenshot shows the same spreadsheet with variable names imported into column A:

Row	Column A	Column B	Column C	Column D	Column E	Column F	Column G	Column H	Column I	Column J	Column K
1	wage	educ	exper	tenure	nonwhite	female	married	numdep			
2	smsa	northcen	south	west	construc	ndurman	trcommpu	trade			
3	services	profserv	profocc	clerocc	servocc	lwage	expersq	tenursq			

A dialog box with the text '準備完了' (Completed) is shown at the bottom left of the spreadsheet.

Excelファイルのimport(2) ヘッダー行の完成



The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L
1	wage	educ	exper	tenure	nonwhite	female	married	numdep	smsa	northcen	south	w
2	3.1	11	2	0	0	1	0	2	1	0	0	
3	3.24	12	22	2	0	1	1	3	1	0	0	
4	3	11	2	0	0	0	0	2	0	0	0	
5	6	8	44	28	0	0	1	0	1	0	0	
6	5.3	12	7	2	0	0	1	1	0	0	0	
7	8.75	16	9	8	0	0	1	0	1	0	0	
8	11.25	18	15	7	0	0	0	0	1	0	0	
9	5	12	5	3	0	1	0	0	1	0	0	
10	3.6	12	26	4	0	1	0	2	1	0	0	
11	18.18	17	22	21	0	0	1	0	1	0	0	
12	6.25	16	8	2	0	1	0	0	1	0	0	
13	8.13	13	3	0	0	1	0	0	1	0	0	
14	8.77	12	15	0	0	0	1	2	1	0	0	
15	5.5	12	18	3	0	0	0	0	1	0	0	
16	22.2	12	31	15	0	0	1	1	1	0	0	
17	17.33	16	14	0	0	0	1	1	1	0	0	
18	7.5	12	10	0	0	1	1	0	1	0	0	
19	10.63	13	16	10	0	1	0	0	1	0	0	
20	3.6	12	13	0	0	1	1	3	1	0	0	

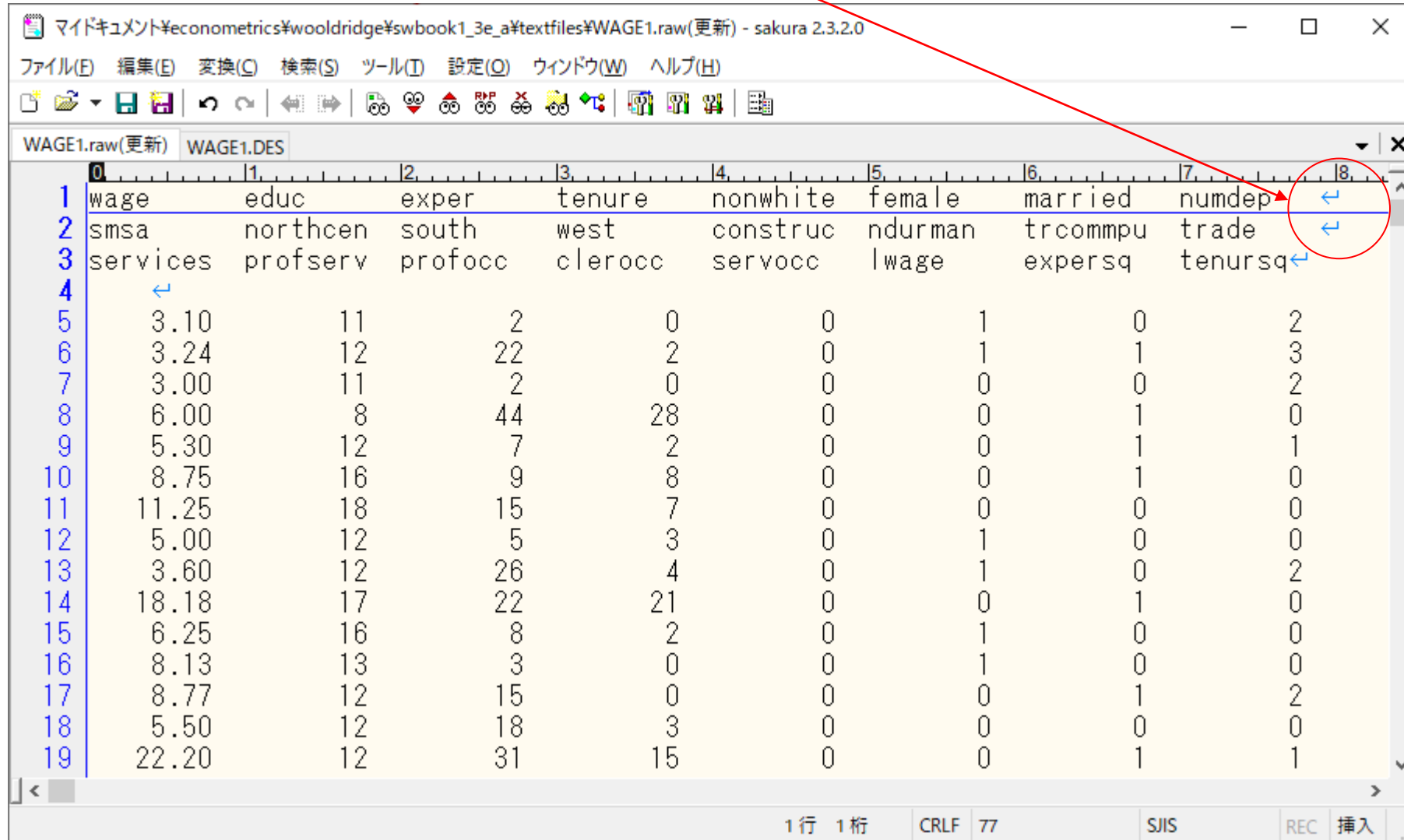
- 先頭行に変数名が入ったファイルが完成したら、適当な名前をつけて保存する
- 元のexcelファイルとは違うファイル名で違うフォルダーに保存しておく（元のファイルを壊さないため）
- このファイルをCSV形式で保存することもできる
(Excelで「名前をつけて保存」を選択し、保存するときにファイル形式としてCSVを選択する)

テキストファイルのimport(1)

wage1.desの変数名をコピーして、wage1.rawの先頭行に挿入

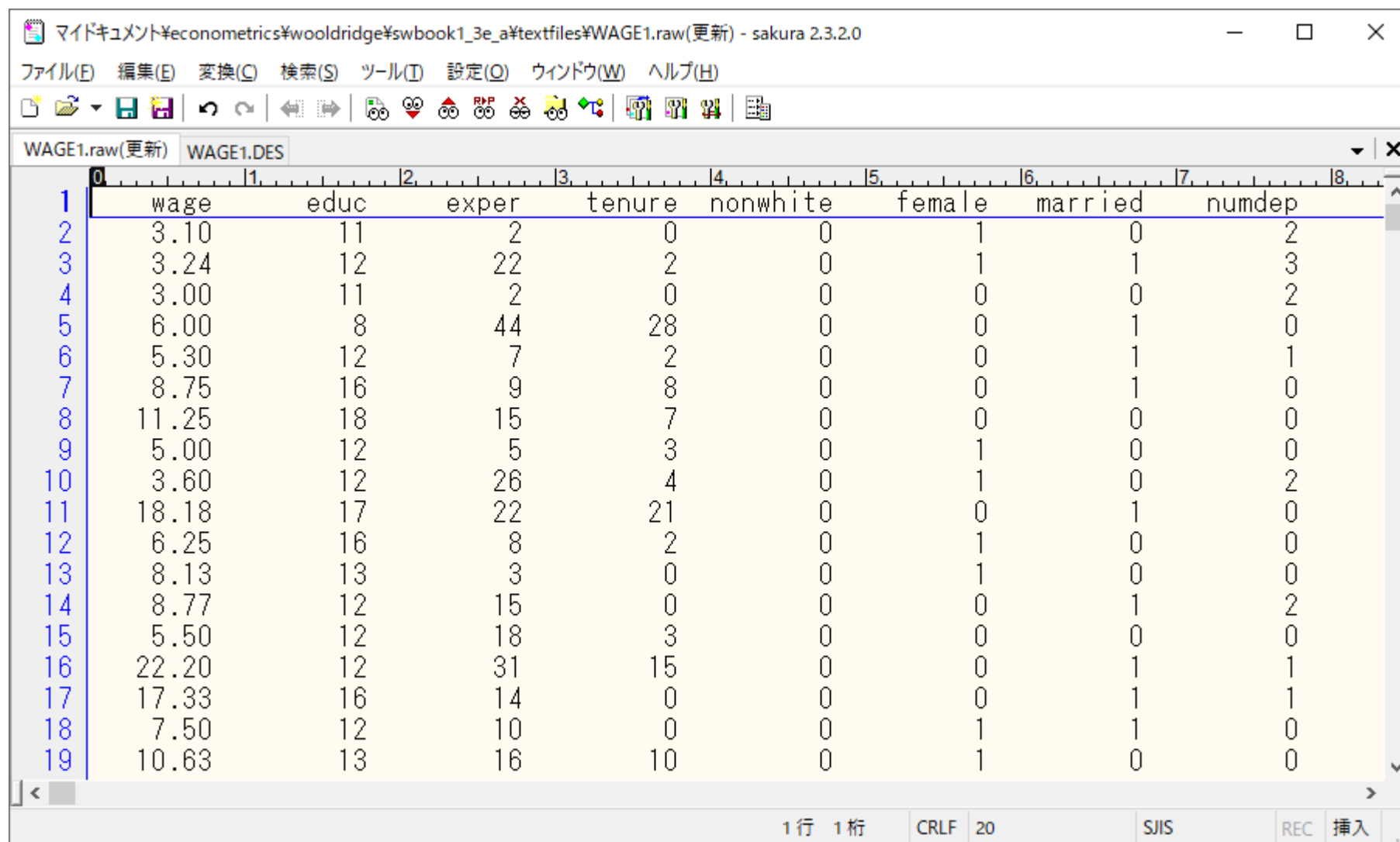
行の途中で表示を折り返さないように設定

この段階では変数名の途中で改行が入っている



改行を取り除いて変数名を1行に直す。適当な名前をつけて保存する（拡張子はcsv、元のフォルダーとルと違うフォルダーに保存した方がよい）。

変数名とデータの桁をそろえておくと読み込みのときに失敗が少ない。



	0	1	2	3	4	5	6	7	8
1	wage	educ	exper	tenure	nonwhite	female	married	numdep	
2	3.10	11	2	0	0	1	0	2	
3	3.24	12	22	2	0	1	1	3	
4	3.00	11	2	0	0	0	0	2	
5	6.00	8	44	28	0	0	1	0	
6	5.30	12	7	2	0	0	1	1	
7	8.75	16	9	8	0	0	1	0	
8	11.25	18	15	7	0	0	0	0	
9	5.00	12	5	3	0	1	0	0	
10	3.60	12	26	4	0	1	0	2	
11	18.18	17	22	21	0	0	1	0	
12	6.25	16	8	2	0	1	0	0	
13	8.13	13	3	0	0	1	0	0	
14	8.77	12	15	0	0	0	1	2	
15	5.50	12	18	3	0	0	0	0	
16	22.20	12	31	15	0	0	1	1	
17	17.33	16	14	0	0	0	1	1	
18	7.50	12	10	0	0	1	1	0	
19	10.63	13	16	10	0	1	0	0	

CSVファイルのimport

Rstudio右上のwindow (Environment) の Import Datasets タブを選択し, From Text (base) .. を選択 →importするファイルを指定

Menuから File/Import Datasets/From Text (base)とたどってもよい

左の画面では, ヘッダーの有無, データの区切り等の指定ができる

問題がなければimportボタンを押してデータをインポートする

The screenshot shows the RStudio interface with the 'Import Dataset' dialog box open. The dialog is configured for importing a CSV file. The 'Name' field contains 'wage1'. The 'Input File' field contains the file path 'wage,educ,exper,tenure,nonwhite,female,married,numdep,s'. The 'Data Frame' section shows a preview of the data with columns: wage, educ, exper, tenure, nonwhite, female, married. The 'Import' button is circled in red.

wage	educ	exper	tenure	nonwhite	female	married
3.10	11	2	0	0	1	0
3.24	12	22	2	0	1	1
3.00	11	2	0	0	0	0
6.00	8	44	28	0	0	1
5.30	12	7	2	0	0	1
8.75	16	9	8	0	0	1
11.25	18	15	7	0	0	0
5.00	12	5	3	0	1	0
3.60	12	26	4	0	1	0
18.18	17	22	21	0	0	1
6.25	16	8	2	0	1	0
8.13	13	3	0	0	1	0
8.77	12	15	0	0	0	1
5.50	12	18	3	0	0	0
22.20	12	31	15	0	0	1
17.33	16	14	0	0	0	1

Excel, CSVファイルのimport

- Excel ファイル, CSVファイル

Import Datasets タブ/From Excel
または
Import Datasets タブ/From Text (readr) ..
からファイルを選択

Menuから
File/Import Datasets/From Excel
File/Import Datasets/From Text (readr)
で選択することもできる

ファイルを選択すると左のような画面が表れる。問題がなければimportボタンをクリックしてimport完了

ヘッダー（先頭行が変数名か）の有無,
データの区切りなどを指定できる

The screenshot shows the RStudio interface with the 'Import Text Data' dialog box open. The dialog is titled 'Import Text Data' and has a 'File/Url' field containing '~/.econometrics/wage1.csv'. Below this is a 'Data Preview' section showing a table with columns: wage, educ, exper, tenure, nonwhite, female, married, numdep, smsa, northcen, south. The 'Import Options' section is circled in red and includes: Name: wage1, Skip: 0, First Row as Names (checked), Trim Spaces (checked), Open Data Viewer (checked), Delimiter: Comma, Quotes: Default, Escape: None, Comment: Default, NA: Default, and Locale: Configure... The 'Code Preview' section shows the following R code: library(readr), wage1 <- read_csv('~/.econometrics/wage1.csv'), view(wage1). The 'Import' and 'Cancel' buttons are at the bottom right.

空白行で区切られたテキストデータのインポート

空白行で区切られているデータ（1個以上のwhite spaceで区切り）

From Text(base)の場合

import画面で, separator を “white space” とする

From Text(readr)の場合

import画面で, delimiter を “white space” とする

あるいは, コマンドラインで

```
data1 <- read.table(“ファイル名“, header=TRUE)
```

とするとdata1にデータがインポートされる

header=TRUE は先頭行に変数名がある場合（ない場合は指定しない）

ファイル名はフォルダーの区切りを/で指定する

例) **read.table(“~/econometrics/cps78_85.txt“, header=TRUE);** ~/ : working directory,

あるいはフルパスで指定する。または,

choose.files()を使うと, ファイル選択の画面があらわれる（これが一番簡単）

```
data1 <- read.table(choose.files(), header=TRUE)
```

インポートが成功すると左上のwindowに data viewer, コンソール画面にはlogが表示される。

The screenshot shows the RStudio interface. The top-left pane displays a data viewer for 'wage1', showing a table with 15 rows and 12 columns. The top-right pane shows the Environment window with 'wage1' listed as a data object with 526 observations and 24 variables. The bottom-left pane shows the Console window with the following output:

```
package 'MatrixModels' successfully unpacked and MD5 sums checked
package 'pbkrtest' successfully unpacked and MD5 sums checked
package 'quantreg' successfully unpacked and MD5 sums checked
package 'car' successfully unpacked and MD5 sums checked
package 'lmtest' successfully unpacked and MD5 sums checked
package 'sandwich' successfully unpacked and MD5 sums checked
package 'zoo' successfully unpacked and MD5 sums checked
package 'Formula' successfully unpacked and MD5 sums checked
package 'AER' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\yoaso\AppData\Local\Temp\RtmpKcpw5r\downloaded_packages

> wage1 <- read.csv("~/econometrics/wage1.csv")
> View(wage1)
```

The bottom-right pane shows the Packages window with a list of installed and available packages.

```
> wage1 <- read.csv("~/econometrics/wage1.csv")
> View(wage1)
```

慣れてくれば、直接、上のようにタイプしても良い。

(最初の行：wage1.csvをインポートして wage1というデータセット（データフレーム）に格納というコマンド， <- は代入せよ というコマンド)

注意：read.csv()はCSVデータをインポートするための関数。

詳しい説明については右下windowで Helpのタブを選択し，検索ボックスに read.csvとタイプせよ。

~/econometrics/ はフォルダーの場所（ファイルを保存しているフォルダーによって異なります）

データ分析

- データフレーム
 - Importされたデータセット wage1は**データフレーム**という構造のデータ
 - クロスセクションデータを便利に扱える
 - その他のデータ構造： 数値，文字列，ベクトル，行列，リストなどのデータ構造がRでは扱える。また，時系列データやパネルデータを扱うような構造もあり
- Rでは[データフレーム\$変数名]でデータフレームの中の変数にアクセスできる。
- 変数名の指定だけでアクセスするには，データフレームをattachさせておく。
 - コンソール画面でattach(データファイル名) とタイプ
 - 複数のデータフレームを扱う場合，異なるデータフレームで同じ変数名が使われていることがある
 - この場合，複数のデータフレームをattachさせたままだとエラーが起こるので，必ず，一つのデータフレームの分析が終了した場合（あるいはRを終了させる前に）データフレームを切り離しておく
- データフレームの切り離し
 - コンソール画面で，detach(データフレーム名) とタイプ

データ分析(2)

wage1がimportされているとして
コンソール画面で

summary(wage1)とタイプ → wage1というデータフレーム中の変数全部の要約統計量

wage1\$wage とタイプ → データフレームwage1の中のwageのデータを表示

以下のコマンドをタイプしてみてください (# 以下は説明なのでタイプ不要)

```
attach(wage1)      # データフレームwage1内の変数に直接アクセスできるようにする
summary(wage)     # wage(賃金) の要約統計量
table(educ)       # educ(教育年数)の度数分布
hist(educ)        # wage のhistogram
plot(educ, wage)  # educとwageの散布図
cor(educ,wage)    # educとwageの相関係数
```

他の変数についても試してみてください

```
detach(wage1)     #データフレームwage1を切り離す
```

作業終了

ファイルのロード, 保存, Rの終了

- パッケージAER内の CPS1988.rdaをロードする

```
library("AER")
```

```
data("CPS1988", package="AER")
```

library("パッケージ名")はパッケージを読み込むコマンド (データセットの利用だけなら不要)

- read.table(), write.table()
 - データの区切りがスペースまたはタブで区切られている場合のインポート, エクスポートに使う

上記のコマンドはしばらくは使いません

- 終了
コンソール画面で q() とタイプするか, Rstudioのメニューから File/Quit Session を選択

ファイルのロード, 保存, 終了(2)

このページのコマンドも当分の間は使いません

- ファイルの保存

作成したデータファイル (データ・フレーム) をcsvファイルで保存→write.csv() 関数

データ・フレーム名がwage1を, Rのworking directoryの下に, wage01.csvで保存したい場合, RStudioのコンソール画面で次のようにタイプする。

```
write.csv(wage1, file=~ /wage01.csv")
```

- Rのバイナリーデータとして保存するには

```
save(wage1, file=~ /wage01.rda")
```

- バイナリーデータのロード

```
load(~ /wage01.rda")
```

rda : Rのバイナリーデータの拡張子

注意: / はフォルダーの区切り。通常はフォルダの区切りは¥ (バックスラッシュ)

~/ はRのworking directory。Windows10の場合, ユーザーフォルダーの下のDocumentsというフォルダー。

ファイルのロード，保存，終了(3)

- Rで一連の作業を終了→ コマンドの履歴，作成したオブジェクト（保存した変数や回帰分析の結果など）は .Rdata, .Rhistory というファイルに保存されます。これらのファイルはRのworking directory（デフォルトではDocumentsフォルダ）にあります。
- 自宅のPC，学校のPCという風に，複数のPCで作業を続けたい場合， .Rdata, .RhistoryをそれぞれのPCのworking directory にコピーすれば，作業を引き継げます。
- これらのファイルの場所がよくわからない場合（隠しファイルになっていたり，working directoryの場所がよくわからない場合など）→ Rstudioのコンソール画面をテキストファイルで保存して，次の作業時にそのファイルを利用する（初心者にはこれが一番簡単でしょう）。

変数の作成・保存

- コマンドラインで次のようにタイプする

新変数名 <- 計算式

例) `lnwage <- log(wage)`
`exper2 <- exper * exper`
`exper2 <- exper^2`

- 注意) `wage`がデータフレーム`wage1`内の変数である場合、上で作成した`lnwage`, `exper2`はデータフレーム`wage1`内にはないことに注意。この場合(`wage1`がattachされているとして)

```
wage2 <- wage1          #データフレームwage1をコピーしてデータフレームwage2を作る
wage2$lnwage <- log(wage)
detach(wage1)
write.csv(wage2, file="ファイル名")
```

とすると`lnwage`がデータフレーム`wage2`内に保存される。最後に後で繰り返し利用できるように、`write.csv`でcsvファイルとして保存しておく

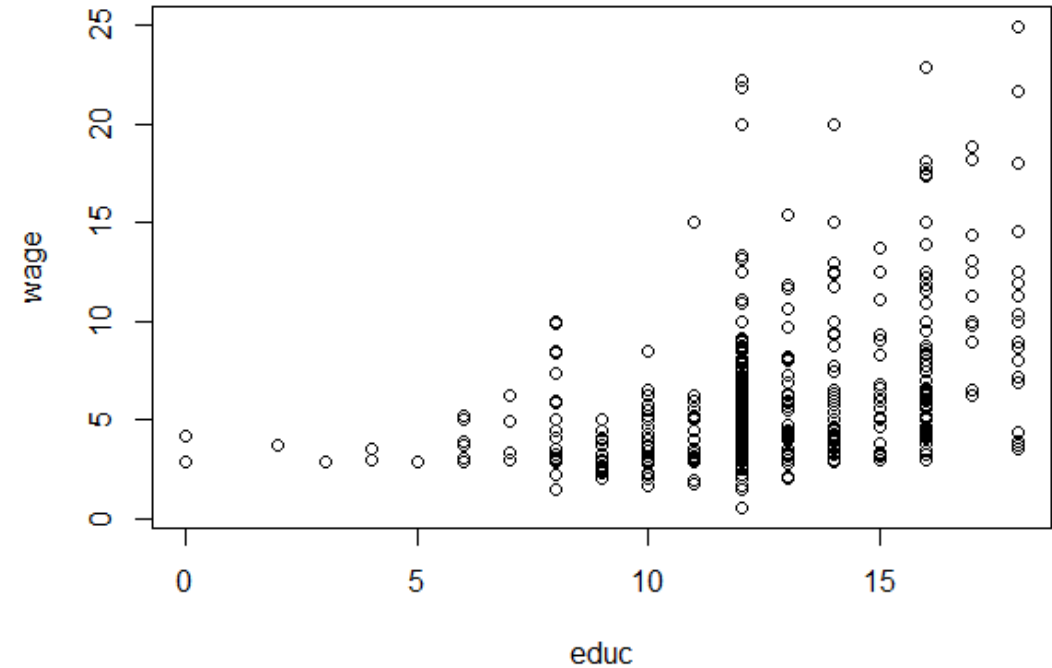
新しいデータフレームを作成したのは、もとのデータフレームをそのままにしたかったから（元のデータフレームに上書きしたい場合は、2行目を`wage1$lnwage`とすればよい）

グラフの出力例

図はRstudioの右下のwindowに表示されます。
そのwindowの上部にexportタブを押す→ファイルとして保存 or clipboardにコピーできる→Word等にpasteできる



コマンド
`hist(wage)`



コマンド
`plot(educ, wage)`

回帰分析

`lm(y ~ x1 + x2 + x3)`

$y = a + b_1x_1 + b_2x_2 + b_3x_3 + e$ という式を当てはめる場合

`lm()`関数を用いて回帰分析を行う

Rでの回帰分析の手順

(回帰の結果を保存する) オブジェクト名 <- `lm(y ~ x1 + x2 + x3)`

`summary(オブジェクト名)` で結果を取り出す

例)

```
wage1_lm <- lm(wage ~ educ + tenure + exper)
```

```
summary(wage1_lm)
```

```
Call:
lm(formula = wage ~ educ + tenure + exper)
```

回帰分析の出力結果
(コンソール画面に出てきます)

```
Residuals:
```

Min	1Q	Median	3Q	Max
-7.6068	-1.7747	-0.6279	1.1969	14.6536

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.87273	0.72896	-3.941	9.22e-05 ***
educ	0.59897	0.05128	11.679	< 2e-16 ***
tenure	0.16927	0.02164	7.820	2.93e-14 ***
exper	0.02234	0.01206	1.853	0.0645 .

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.084 on 522 degrees of freedom
```

```
Multiple R-squared:  0.3064, Adjusted R-squared:  0.3024
```

```
F-statistic: 76.87 on 3 and 522 DF, p-value: < 2.2e-16
```

相関行列, 分散共分散行列, 散布図行列

- 3つ以上の変数の相関係数等を一度に出力する方法
- 複数の変数を`cbind()`でまとめ, 適当な変数名をつける。
 - `cbind()`: column bind 列ベクトルをまとめる関数
- その変数名で, `cor()`, `var()`を実行
- 散布図行列の場合は, `pairs()`

例) `wage1.csv` のデータで, `wage,educ,exper,tenure`の相関行列等を求める (下の例は、3つの変数を変数`x`にまとめ、4変数間の相関行列等を求めるコマンド)

```
x <- cbind(wage, educ, exper, tenure)
```

```
cor(x)
```

```
var(x)
```

```
pairs(x)
```

出力結果

```
> cor(x)
```

	wage	educ	exper	tenure
wage	1.0000000	0.40590333	0.1129034	0.34688957
educ	0.4059033	1.00000000	-0.2995418	-0.05617257
exper	0.1129034	-0.29954184	1.0000000	0.49929145
tenure	0.3468896	-0.05617257	0.4992914	1.00000000

相関行列

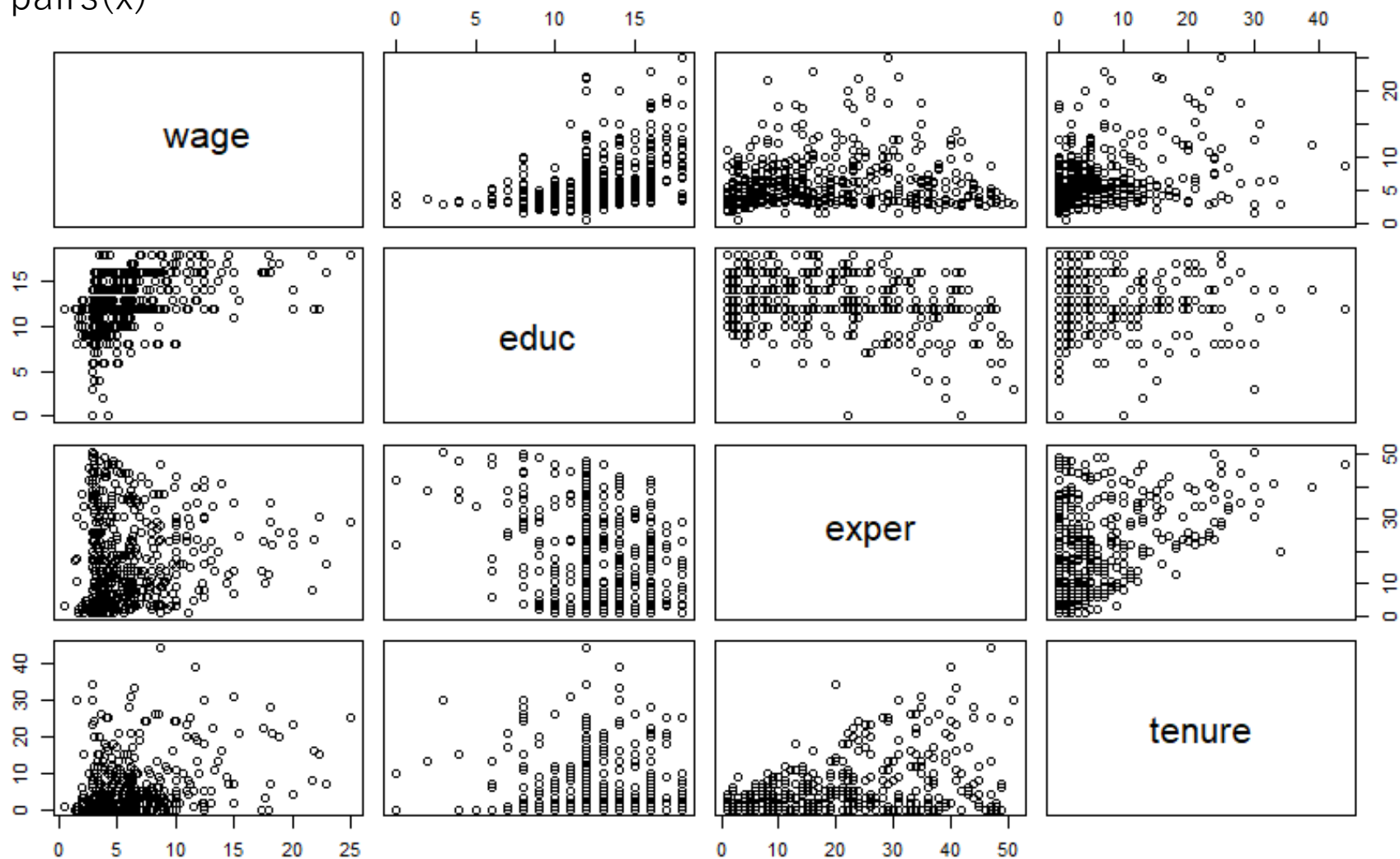
```
> var(x)
```

	wage	educ	exper	tenure
wage	13.638884	4.150864	5.659076	9.255208
educ	4.150864	7.667485	-11.257266	-1.123715
exper	5.659076	-11.257266	184.203516	48.956303
tenure	9.255208	-1.123715	48.956303	52.192855

分散共分散行列

散布図行列の出力結果

コマンド `pairs(x)`



時系列データ

以下のページは、時系列データを扱う際に参照すればよい。当分は不要。

• データフレーム

- クロスセクションデータの分析に適したデータ構造
- 時系列データの場合、変数名とそのデータに加えて、時系列の情報（開始時点、終了時点、データの周期（年次、4半期、月次）、…）の情報が必要
- 時系列データ構造に変換すると、ラグ付きの変数（1期前の値、2期前の値、…）が作成できた李、差分（1期前から変化、…）の変換ができるようになる
- データフレームで同じような操作をするためには、データのimport前のcsvまたはxlsxファイルの段階でラグ付き変数や変数の差分をあらかじめ作成しておく必要がある。

• 時系列データへの変換

- `ts()` でデータフレームや個々の変数を時系列構造に変換
- `tsdata <- ts(data, ……)`
- `tsdata`: 新オブジェクト名, `data`: データフレームまたは変数名,
- `…`: optionの指定 ; `start`=開始時点, `end`=終了時点, `frequency`=周期の指定 (1 : 年次, 4: 四半期, 12 : 月次)
- `start=1948` (1948年が開始時点), `start=c(1983,2)` (開始時点が1983年2月または1983年第2四半期) `c(n, m)`は数字n,mの組という意味
- optionを省略するとデータは1から順番に番号が付けられる

時系列データ(2)

例) データフレームintdefを時系列構造をもったintdeftsに変換

```
intdefts <- ts(intdef, start=1948, frequency=1)
```

- intdefts内の変数へのアクセスはやや面倒
 - intdefts\$変数名 ではアクセスできない
 - attach() も使えない
 - intdefts[, 列の番号]またはintdefts[, "変数名"]でアクセス
(intdefts[, 2]は行列intdefの第2列を取り出すという意味)
 - 例) intdefts[,2] , intdefts[, "i3"]
 - plot(intdefts[, "i3"])でi3という変数の時系列プロット図が得られる
- zoo()という関数による変換の方が便利かもしれない(package zooをインストールする必要あり)

時系列データ(3)

- 時系列データの回帰分析では`dynlm()`という関数が便利
 - 回帰モデルの指定時に、ラグ付き変数が見える
 - `reg1 <- dynlm(inf ~ i3 + L(i3) + L(i3,2) + def + L(def) + L(def,2), data= intdefts)`
 - `L(i3)`: `i3`の1期前の値, `L(i3, 2)`: `i3`の2期前の値
 - `data=` データセット名 で時系列データセット名の指定が必要
 - `intdefts`内の`inf`, `i3`, `def`を用いた回帰
 - 結果は`summary(reg1)`で取り出す
 - 係数の制約の検定は通常と同様 (`lmtest()` を用いる)
- `dynlm`というpackageをインストールして、`library(dynlm)`でロードしておく必要
 - Rstudioの右下のwindowでpackage一覧の中から`dynlm`を探しチェックを入れる (なければインストールしてから)

問題

- まず、wage1.xlsまたはwage1.rawの先頭行に変数名を加え、手順にしたがってRにimportし、wage1というデータフレームを作る
- 1. wage1中のwage, educ, exper, tenureの要約統計量を求めよ。
- 2. wage, educ, exper, tenure のヒストグラムを作れ
- 3. educとwageの散布図を描け
- 4. educとwageの相関係数を求めよ
- 5. 被説明変数をwage, 説明変数をeducとした単回帰を行い、結果を報告せよ
(解釈はいらない)

以上をWord等にまとめて提出（コンソール画面のコマンドとその結果は、コピー&ペイストでWordに張り付ける。図はplotを表示しているwindowのexportタブから、ファイル経由またはclipboard経由でWordに貼り付ける）

文献

次の最初の2冊は全くの初心者向け

- G.Grolemund 『RstudioではじめるRプログラミング入門』 オライリー・ジャパン, オーム社, 2015年, 2600円+税
- A.ジュール,E.イノウエ,E.ミースターズ 『R初心者のためのABC』 シュプリンガー・ジャパン, 2010年, 3500円+税

-
- Norman Matloff 『アート・オブ・Rプログラミング』 オライリー・ジャパン, 2012年, 3200円+税
 - C.Kleiber and A.Zeileis, *Applied Econometrics with R*, Springer, 2008
 - Florian Heiss, *Using R for Introductory Econometrics*, 2016
<http://www.urfie.net/> (左のページを参照 Amazo.co.jpでも購入できる)

Internet上の資料

- An Introduction to R

Comprehensive R Archive Network(<https://cran.r-project.org/index.htm>)にあるマニュアル（英語：日本語訳もあります）

CRANにはさまざまな入門的文書があるので、自分にあったものをさがしてみてください

計量経済学用にはCRANに “Econometrics in R” by Grant Farnsworth があります

- 日本語ページ

RjpWiki: <http://www.okada.jp/RWiki/?RjpWiki>

R Tips <http://cse.naro.affrc.go.jp/takezawa/r-tips/r2.html>

Rの場合、わからないことがあればとりあえずネットで検索してみてください