

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にもWindows版、Mac OS版、Linux版がある

Rのインストール(1)

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 (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](#)

Questions About R

すべて表示

CRAN(The Comprehensive R Archive Network) からダウンロード

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

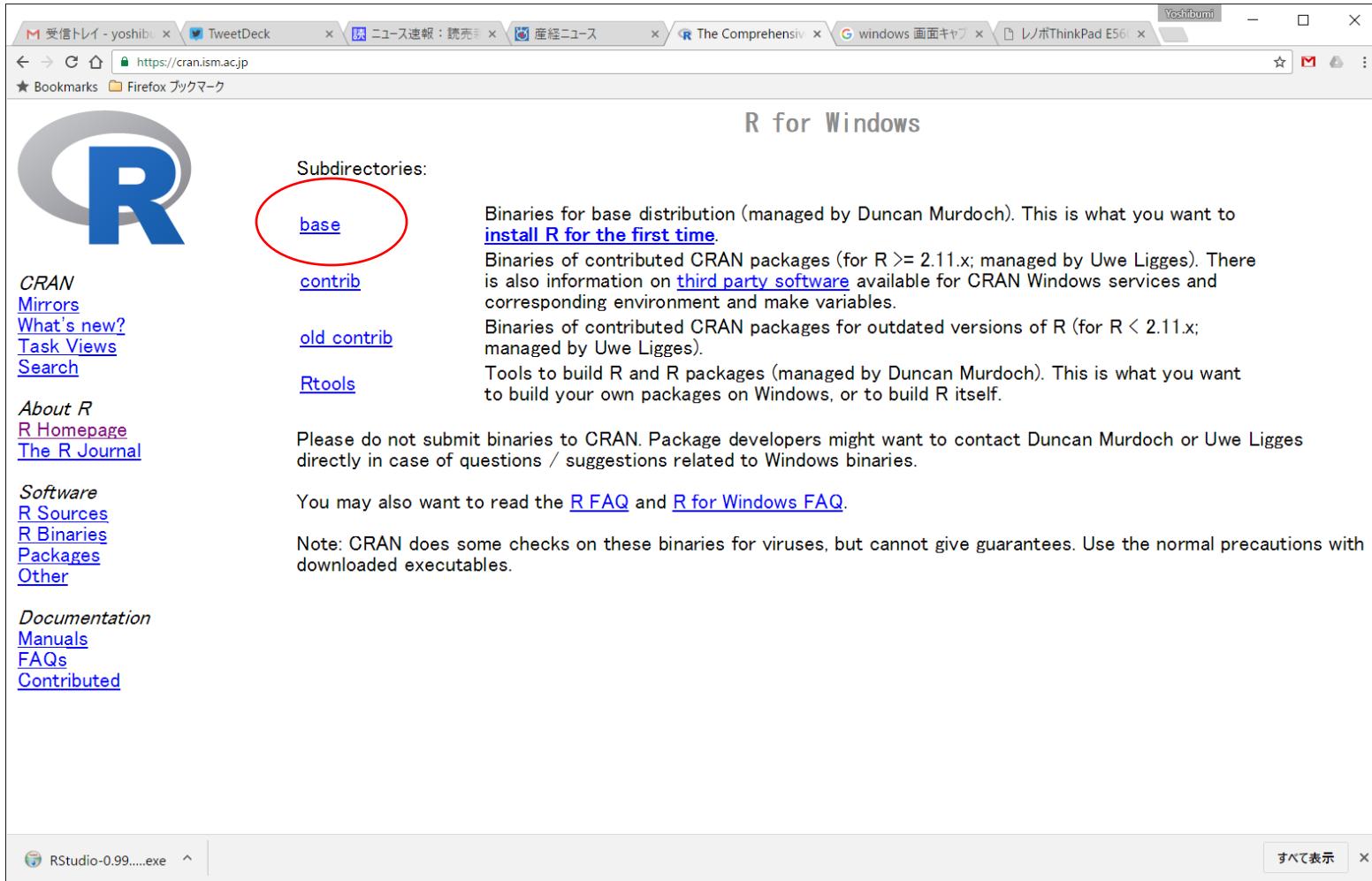
(サイトが見つからない場合は“CRAN”を検索)

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。

Rのインストール(2)



The screenshot shows a Firefox browser window with multiple tabs open. The active tab is titled "R for Windows". The page content includes:

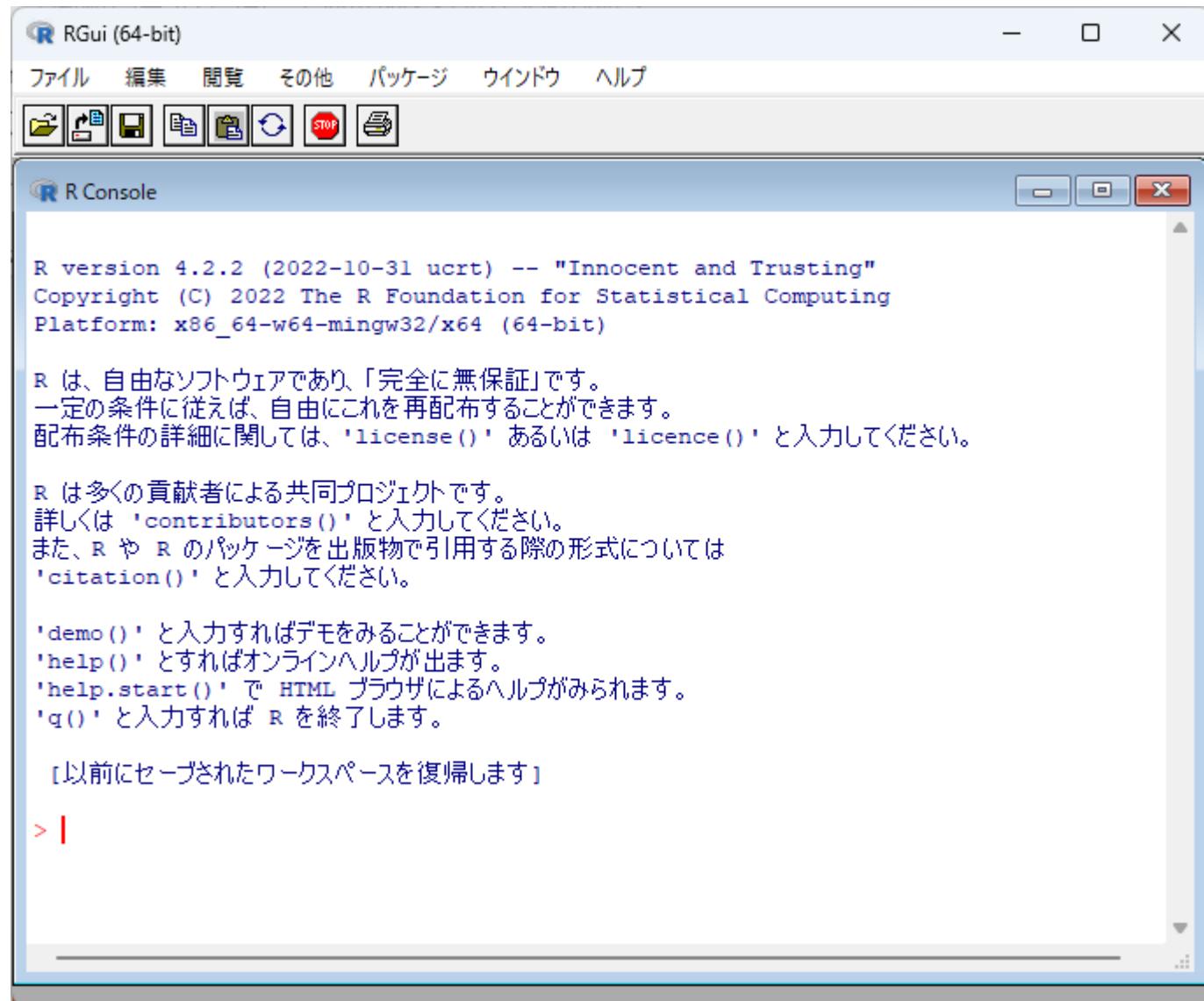
- Subdirectories:**
 - [base](#) (circled in red)
 - [contrib](#)
 - [old_contrib](#)
 - [Rtools](#)
- Binaries for base distribution (managed by Duncan Murdoch). This is what you want to [install R for the first time](#).
- Binaries of contributed CRAN packages (for R \geq 2.11.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.
- Binaries of contributed CRAN packages for outdated versions of R (for R $<$ 2.11.x; managed by Uwe Ligges).
- 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.
- 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.
- 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 precautions with downloaded executables.

At the bottom of the browser window, there is a taskbar with an icon for "RStudio-0.99....exe".

これはWindows版
Baseを選んでダウンロード

Mac OSの場合は
Apple silicon 版とintel版が
あるので、自分の使用して
いるCPUに注意
なお、これとは別に
Xquartzのインストールが
必要かもしれない（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（後述）。
- 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。→ 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

アカウント名（ログインユーザー名）に2バイト文字を用いると， WindowのドキュメントフォルダーがC:¥Users¥xxx以下に作られる（xxxがアカウント名）。Rはこのフォルダーの下の階層に作業用フォルダーを作る。RのWindows版は2byte文字を処理できないため， エラーが出る。

- また， WindowsのホームディレクトリーをOneDriveにしていると（現在は， 新規にwindowsをセットアップするとき， OneDriveがホームディレクトリーになる模様）， トラブルが生じる場合がある。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かもしれません。次のページを参照してください

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

Rstudio のインストール



PRODUCTS ▾ SOLUTIONS ▾ LEARN & SUPPORT ▾ EXPLORE MORE ▾ PRICING



RStudio Desktop

Used by millions of people weekly, the RStudio integrated development environment (IDE) is a set of tools built to help you be more productive with R and Python.

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

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

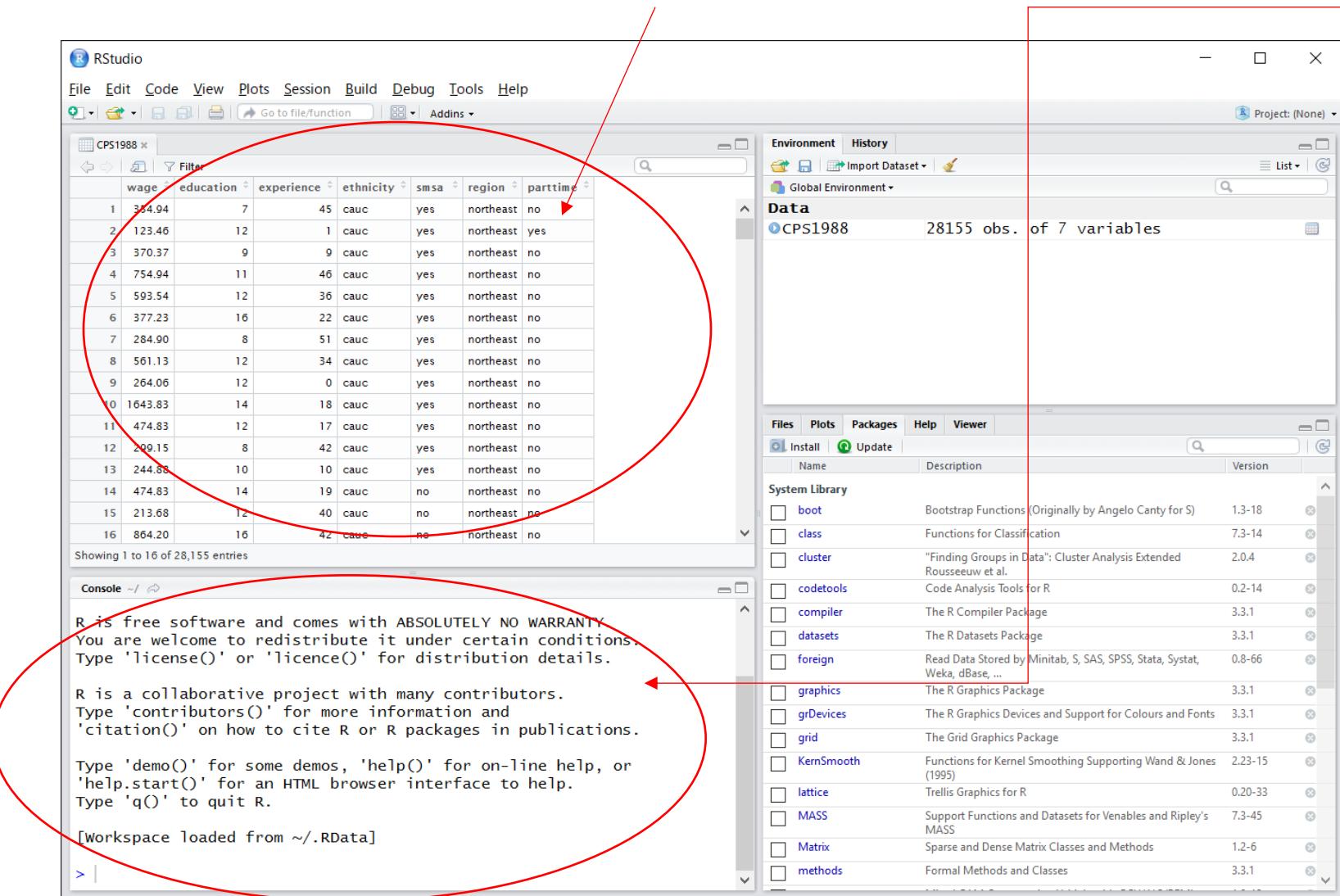
<https://posit.co/download/rstudio-desktop/>

Windows版, Mac版, Linux版がある

みつからない場合は, Rstudio で検索してください

Rstudio

データファイル・ビュアー



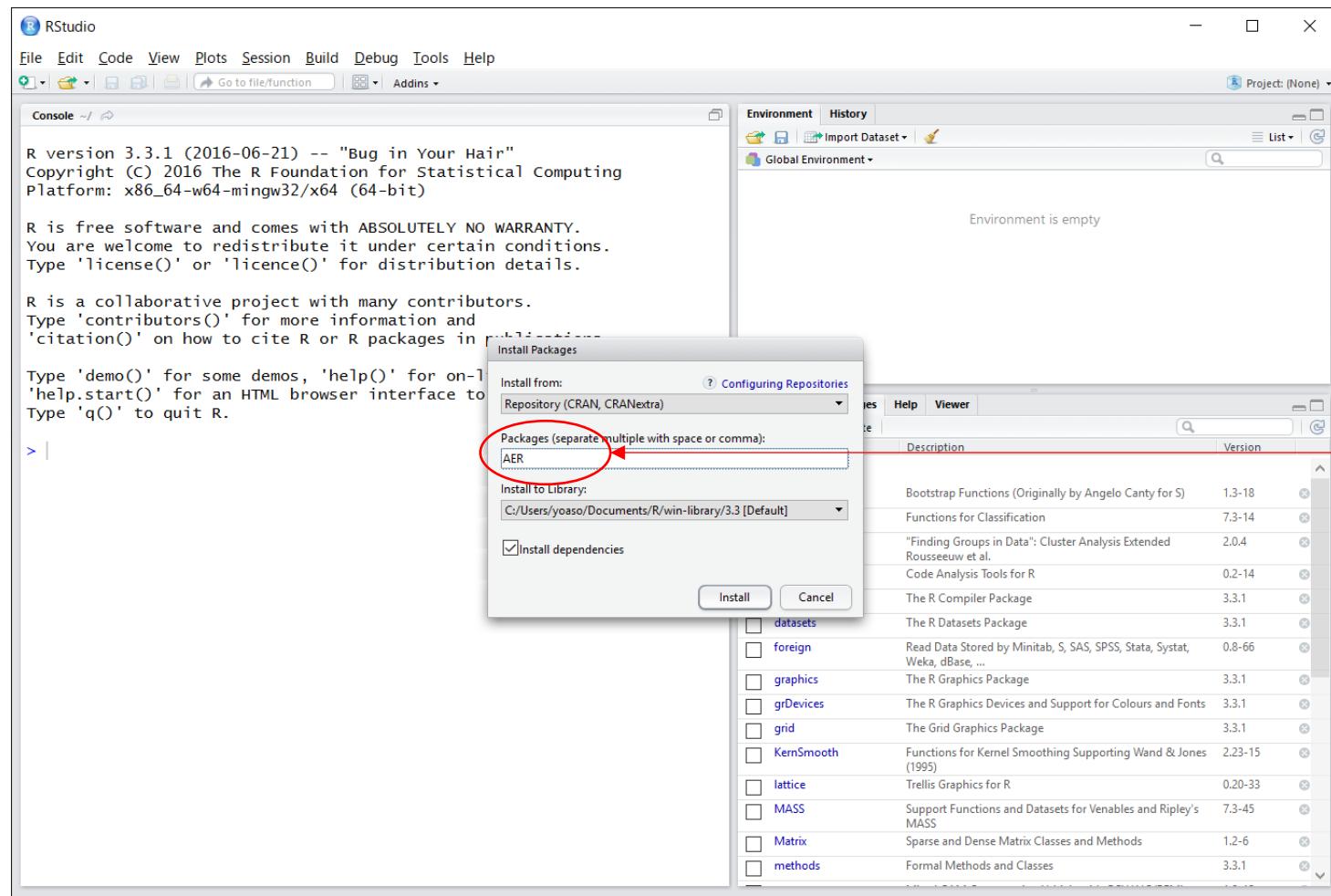
コンソール画面

- 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
- 結果はこの画面に表示される
グラフ等の出力は右下
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
Packages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



- 他にも分析に必要なパッケージはあります
が、まず、AERというパッケージをインス
トールします。
- AER : C.Kleiber and A.Zeileis, “Applied
Econometrics with R”, Springer, 2008 よ
り
- Rstudioの右下windowのタブでpackagesを
選び、さらにinstallを選択する→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
- (別の方法) コンソール画面で
`install.packages("AER")`
とタイプ

インストールしたパッケージがどのフォル
ダにあるかを調べたい場合は、コマンドラ
インで `library()` とタイプする

パッケージの更新

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The left sidebar contains tabs for 'affairs', 'card2', and 'beauty'. The main area has tabs for Environment, History, and Connections. Below these are sections for Data, Global Environment, and a search bar. The bottom pane is titled 'Packages' with tabs for Files, Plots, Packages, Help, and Viewer. The 'Update' tab is highlighted with a red circle. The 'User Library' section lists various R packages with their descriptions and versions.

Name	Description	Version
AER	Applied Econometrics with R	1.2-5
assertthat	Easy Pre and Post Assertions	0.2.0
bdsmatrix	Routines for Block Diagonal Symmetric Matrices	1.3-3
BH	Boost C++ Header Files	1.66.0-1
car	Companion to Applied Regression	2.1-6
cellranger	Translate Spreadsheet Cell Ranges to Rows and Columns	1.1.0
cli	Helpers for Developing Command Line Interfaces	1.0.0
crayon	Colored Terminal Output	1.3.4
dynlm	Dynamic Linear Regression	0.3-5
Formula	Extended Model Formulas	1.2-2
hms	Pretty Time of Day	0.4.1
lme4	Linear Mixed-Effects Models using 'Eigen' and S4	1.1-15
lmtest	Testing Linear Regression Models	0.9-35
MatrixModels	Modelling with Sparse And Dense Matrices	0.4-1
minqa	Minimization Using Quasi-Newton Methods	1.2-4

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
 - RtoolsはRのバージョンにあったものをインストールしてください
 - RtoolsはCRANのホームページのsearchで探すか, google等で検索してください。
- パッケージが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 `getwd()`とタイプする
 - 帰ってきたフォルダ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
 - Windowsの場合、2バイト文字がフォルダ名に使われているとRがうまく動作しない（現時点で）
 - 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
 - working directoryを1byte文字のフォルダーに変更する
 - 親ディレクトリに1バイト文字が使われていないか注意
 - あるいは、1byte文字の新しいログインアカウントを作成し、そのアカウントにRとRstudioをインストールしなおす
 - 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

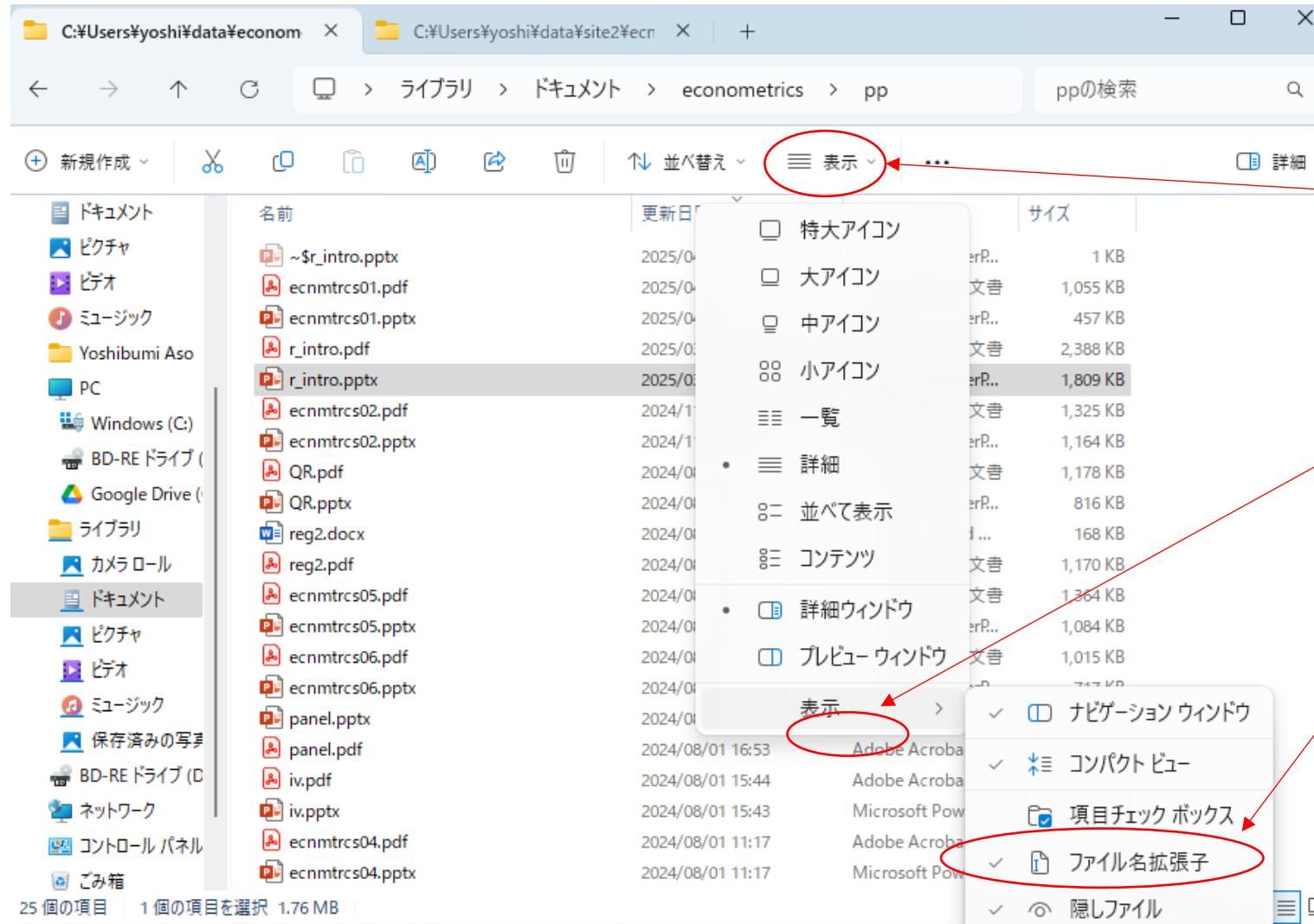
データのimport

- 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、VSCode (Windows)
 - CotEditor、VSCode (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ンダーで、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：ファインダー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)



- エクスプローラーのメニューから「表示」を選択
- 「表示」を選択
- 「ファイル名拡張子」にチェックを入れる

CSVファイルの例

CSVファイル等の処理には、タブや改行等の文字コードが表示されるエディタが便利

```
wage,educ,exper,tenure,nonwhite,female,married,numdep,smsa,northcen,sout
3.1,11,2,0,0,1,0,2,1,0,0,1,0,0,0,0,0,0,0,0,1.131402,4,0
3.24,12,22,2,0,1,1,3,1,0,0,1,0,0,0,0,1,0,0,0,1,1.175573,484,4
3,11,2,0,0,0,0,2,0,0,0,1,0,0,0,1,0,0,0,0,0,1.098612,4,0
6,8,44,28,0,0,1,0,1,0,0,1,0,0,0,0,0,0,0,1,0,1.791759,1936,784
5,3,12,7,2,0,0,1,1,0,0,0,1,0,0,0,0,0,0,0,0,1.667707,49,4
8,75,16,9,8,0,0,1,0,1,0,0,1,0,0,0,0,0,1,1,0,0,2.169054,81,64
11,25,18,15,7,0,0,0,0,1,0,0,1,0,0,0,1,0,0,1,0,0,2.420368,225,49
5,12,5,3,0,1,0,0,1,0,0,1,0,0,0,0,0,0,1,0,0,1.609438,25,9
3,6,12,26,4,0,1,0,2,1,0,0,1,0,0,0,1,0,0,1,0,0,1.280934,676,16
18,17,22,21,0,0,1,0,1,0,0,1,0,0,0,0,0,0,1,0,0,2.900322,484,441
6,25,16,8,2,0,1,0,0,1,0,0,0,1,0,0,0,1,0,0,1,0,0,1.832582,64,4
8,13,13,3,0,0,1,0,0,1,0,0,1,0,0,0,0,0,1,0,1,0,2.095561,9,0
8,77,12,15,0,0,0,1,2,1,0,0,1,0,0,0,0,0,0,0,0,2.171337,225,0
5,5,12,18,3,0,0,0,0,1,0,0,1,0,0,0,0,0,0,0,0,1.704748,324,9
22,2,12,31,15,0,0,1,1,1,0,0,1,0,0,0,0,0,0,1,0,0,3.100092,961,225
17,33,16,14,0,0,0,1,1,1,0,0,1,0,0,0,0,0,1,1,0,0,2.852439,196,0
7,5,12,10,0,0,1,1,0,1,0,0,1,0,0,0,0,0,0,1,0,0,2.014903,100,0
10,63,13,16,10,0,1,0,0,1,0,0,0,1,0,0,0,0,0,1,1,0,0,2.36368,256,100
3,6,12,13,0,0,1,1,3,1,0,0,1,0,0,0,0,0,0,0,0,1.280934,169,0
```

これは、学歴や職歴と賃金の関係を調べたクロスセクションデータ

1行目がヘッダー行で、変数名
2行目からデータが入っている

データの区切りが“,” (comma)
オブザベーションの区切りが改行

改行マーク

csvファイルはExcelでも開ける。
また、Excelシートをcsvファイルとして保存できる

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、折り返し行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

The screenshot shows a text editor window with the title "マイドキュメント¥econometrics¥wooldridge¥swbook1_3e_a¥textfiles¥WAGE1.DES - sakura 2.3.2.0". The menu bar includes ファイル(F)、編集(E)、変換(C)、検索(S)、ツール(T)、設定(O)、ウィンドウ(W)、ヘルプ(H). The toolbar contains various icons for file operations. The main window displays the WAGE1.DES file content, which includes variable definitions and their meanings.

行番号	変数名	説明
1	WAGE1.DES	
2		
3	wage	average hourly earnings
4	educ	years of education
5	exper	years potential experience
6	tenure	years with current employer
7	nonwhite	=1 if nonwhite
8	female	=1 if female
9	married	=1 if married
10	numdep	number of dependents
11	smsa	=1 if live in SMSA
12	northcen	=1 if live in north central U.S
13	south	=1 if live in southern region

At the bottom, status bars show "5行 78行 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

The screenshot shows a software window titled "マイドキュメント¥econometrics¥wooldridge¥swbook1_3e_a¥textfiles¥WAGE1.raw - sakura 2.3.2.0". The menu bar includes "ファイル(F)", "編集(E)", "変換(O)", "検索(S)", "ツール(T)", "設定(O)", "ウインドウ(W)", and "ヘルプ(H)". Below the menu is a toolbar with various icons. The main area displays a table with 12 columns and 20 rows. The first column contains row numbers from 1 to 11. The second column contains numerical values. The subsequent columns contain mostly zeros, with some non-zero values appearing in the third and fourth columns. The table is titled "WAGE1.DES" and "WAGE1.raw". At the bottom of the window, there are status indicators: "1行 1桁", "CRLF 20", "SJIS", "REC", and "挿入".

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
	8.13	13	3	0	0	1	0	0
	8.77	12	15	0	0	0	1	2
	5.50	12	18	3	0	0	0	0
	22.20	12	31	15	0	0	1	1
	17.33	16	14	0	0	0	1	1
	7.50	12	10	0	0	1	1	0
	10.63	13	16	10	0	1	0	0
	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

1	3.1	11	2	0	0	1	0	2	1	0	0	1	0	0	0	0	C
2	3.24	12	22	2	0	1	1	3	1	0	0	1	0	0	1	0	
3	3	11	2	0	0	0	0	2	0	0	0	0	1	0	0	0	
4	6	8	44	28	0	0	1	0	1	0	0	0	1	0	0	0	
5	5.3	12	7	2	0	0	1	1	0	0	0	0	1	0	0	0	
6	8.75	16	9	8	0	0	1	0	1	0	0	0	1	0	0	0	
7	11.25	18	15	7	0	0	0	0	1	0	0	0	1	0	0	0	
8	5	12	5	3	0	1	0	0	1	0	0	0	1	0	0	0	
9	3.6	12	26	4	0	1	0	2	1	0	0	1	0	0	0	0	
10	18.18	17	22	21	0	0	1	0	1	0	0	0	1	0	0	0	
11	6.25	16	8	2	0	1	0	0	1	0	0	0	1	0	0	0	
12	8.13	13	3	0	0	1	0	0	1	0	0	0	1	0	0	0	
13	8.77	12	15	0	0	0	1	2	1	0	0	1	0	0	0	0	
14	5.5	12	18	3	0	0	0	0	1	0	0	1	0	0	0	0	
15	22.2	12	31	15	0	0	1	1	1	0	0	1	0	0	0	0	
16	17.33	16	14	0	0	0	1	1	1	0	0	1	0	0	0	0	
17	7.5	12	10	0	0	1	1	0	1	0	0	1	0	0	0	0	
18	10.63	13	16	10	0	1	0	0	1	0	0	1	0	0	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の後で変数名を指定することができるが面倒 → 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と便利
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

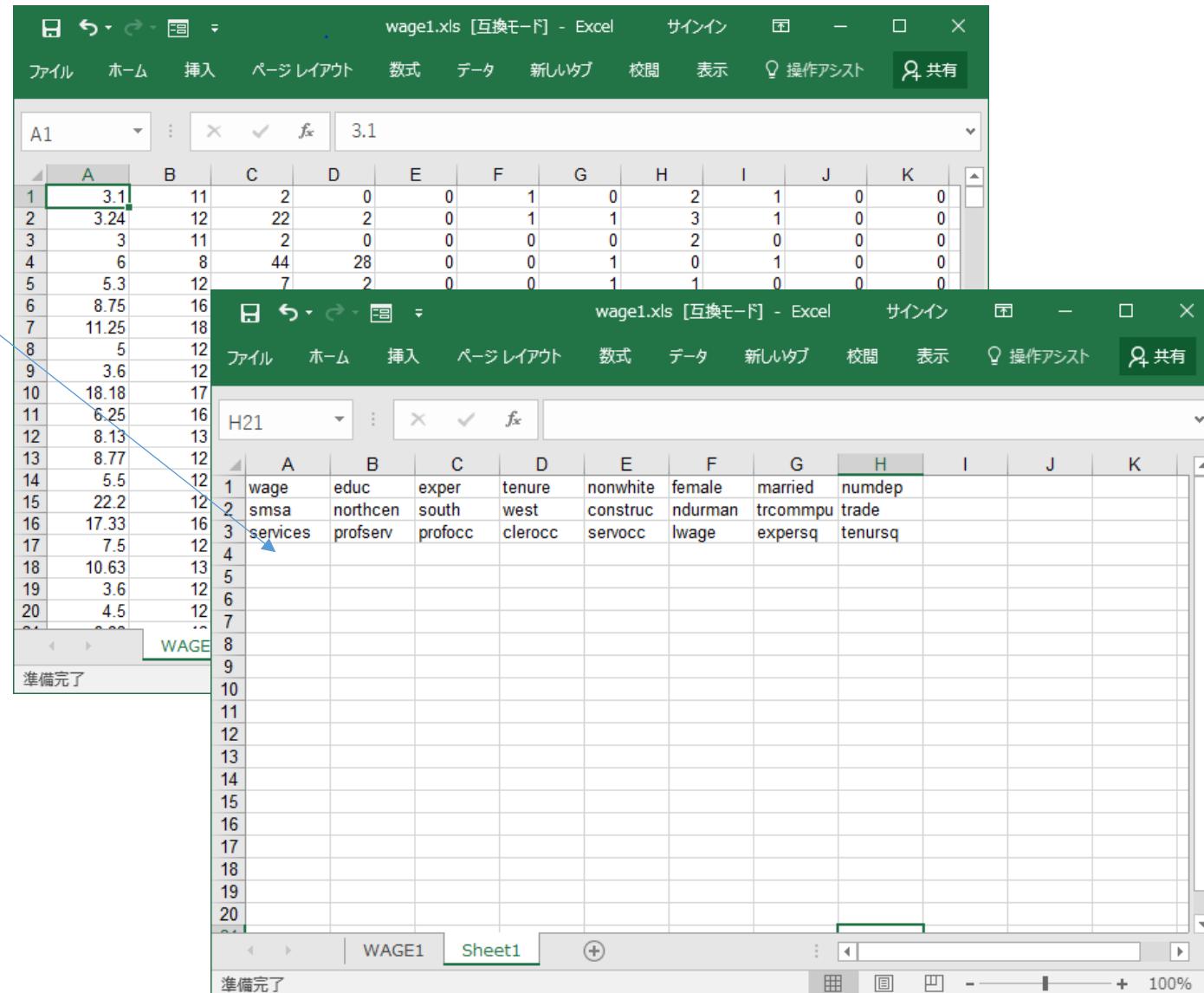
Excel ファイルのimport

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&ペイストで貼り付ける（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



Excelファイルのimport(2) ヘッダー行の完成

	wage	educ	exper	tenure	nonwhite	female	married	numdep	smsa	northcen	south	...
2	3.1	11	2	0	0	1	0	2	1	0	0	0
3	3.24	12	22	2	0	1	1	3	1	0	0	0
4	3	11	2	0	0	0	0	2	0	0	0	0
5	6	8	44	28	0	0	1	0	1	0	0	0
6	5.3	12	7	2	0	0	1	1	0	0	0	0
7	8.75	16	9	8	0	0	1	0	1	0	0	0
8	11.25	18	15	7	0	0	0	0	1	0	0	0
9	5	12	5	3	0	1	0	0	1	0	0	0
10	3.6	12	26	4	0	1	0	2	1	0	0	0
11	18.18	17	22	21	0	0	1	0	1	0	0	0
12	6.25	16	8	2	0	1	0	0	1	0	0	0
13	8.13	13	3	0	0	1	0	0	1	0	0	0
14	8.77	12	15	0	0	0	1	2	1	0	0	0
15	5.5	12	18	3	0	0	0	0	1	0	0	0
16	22.2	12	31	15	0	0	1	1	1	0	0	0
17	17.33	16	14	0	0	0	1	1	1	0	0	0
18	7.5	12	10	0	0	1	1	0	1	0	0	0
19	10.63	13	16	10	0	1	0	0	1	0	0	0
20	3.6	12	13	0	0	1	1	3	1	0	0	0

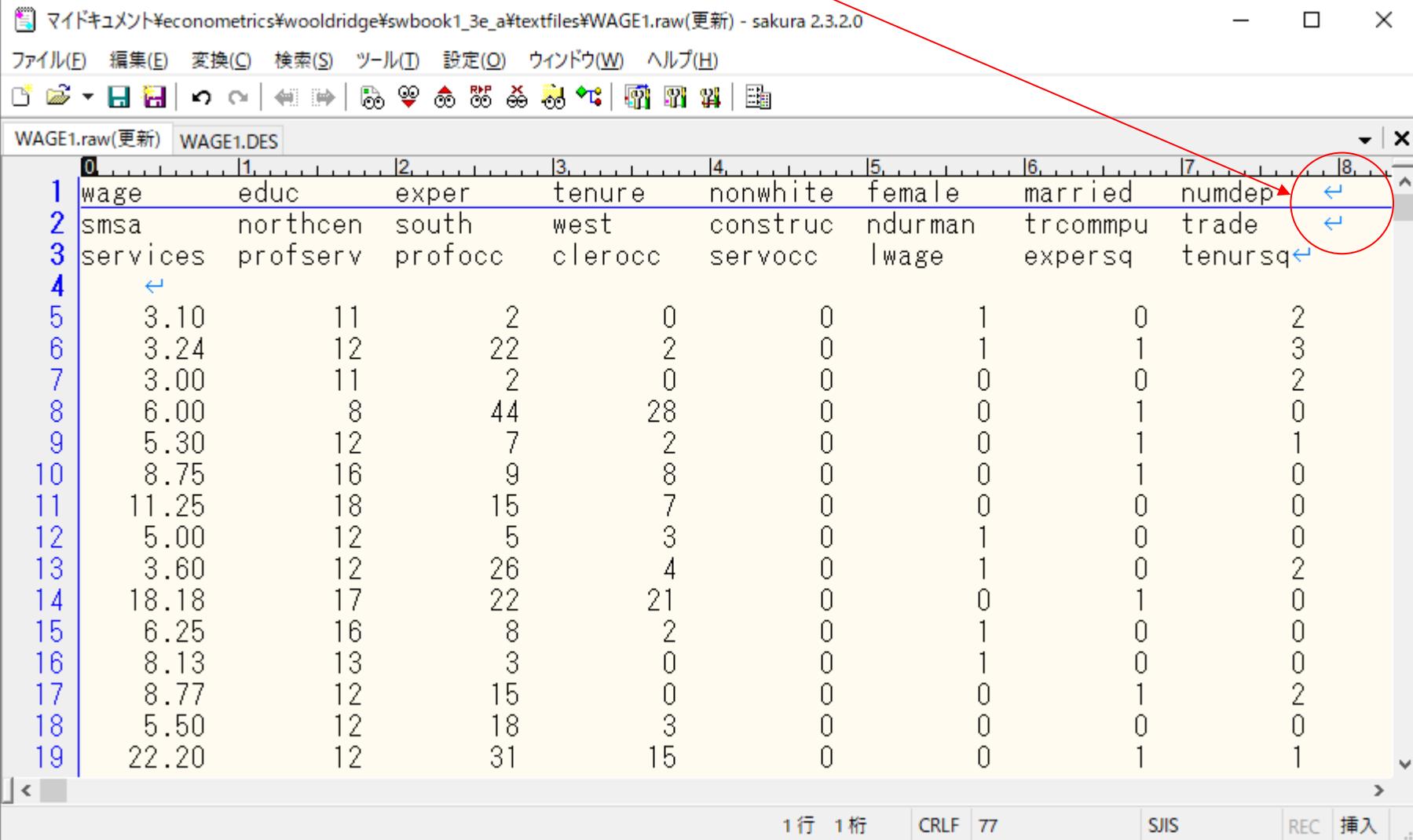
- 先頭行に変数名が入ったファイルが完成したら、適当な名前をつけて保存する
- 元のexcelファイルとは違うファイル名で違うフォルダに保存しておく（元のファイルを壊さないため）
- このファイルをCSV形式で保存することもできる（Excelで「名前をつけて保存」を選択し、保存するときにファイル形式としてCSVを選択する）

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に改行が入っている



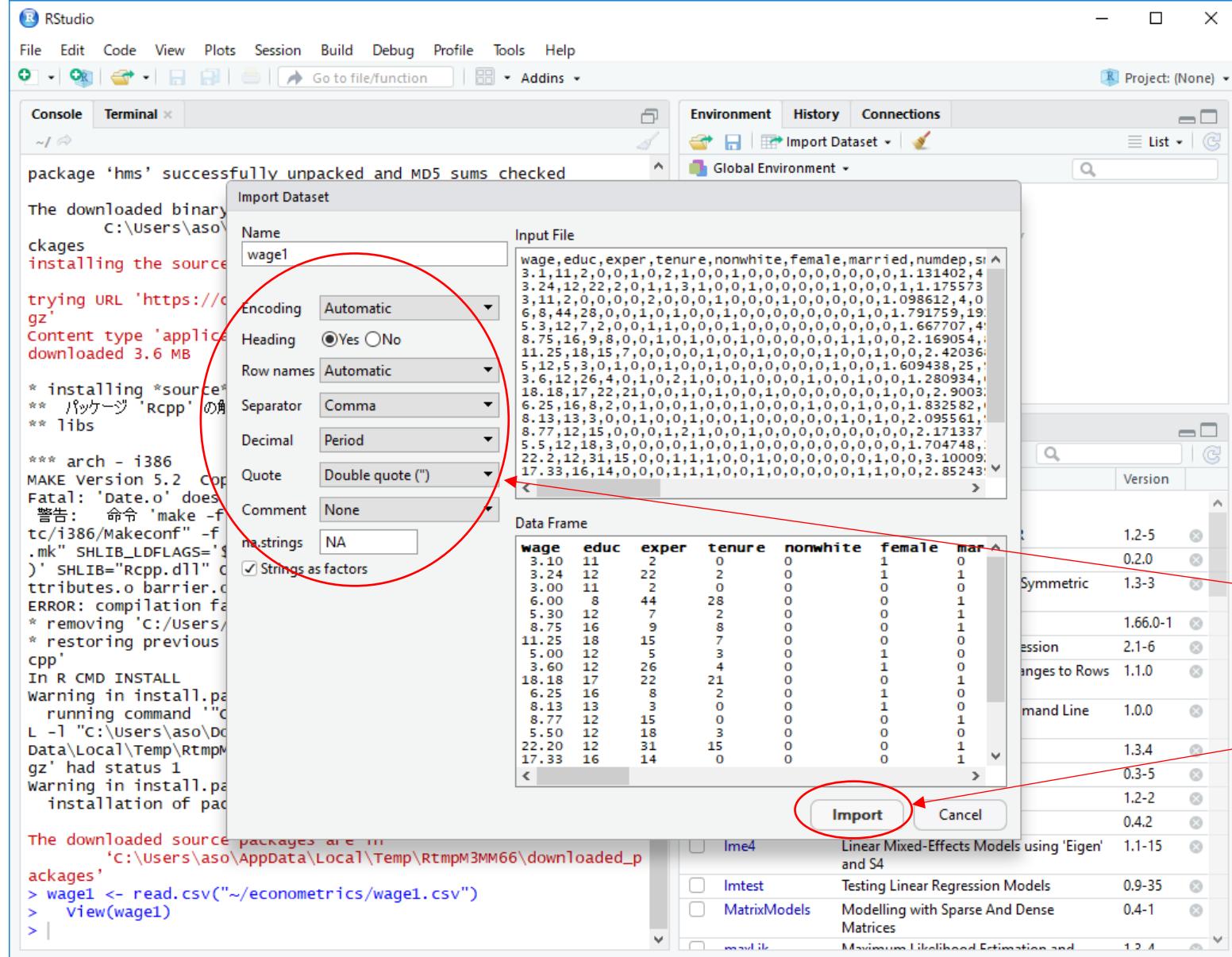
	0	1	2	3	4	5	6	7	8
1	wage	educ	exper	tenure	nonwhite	female	married	numdep	←
2	smsa	northcen	south	west	construc	ndurman	trcommu	trade	←
3	services	profserv	profocc	clerocc	servocc	lwage	expersq	tenursq	←
4	←								
5	3.10	11	2	0	0	1	0	2	
6	3.24	12	22	2	0	1	1	3	
7	3.00	11	2	0	0	0	0	2	
8	6.00	8	44	28	0	0	1	0	
9	5.30	12	7	2	0	0	1	1	
10	8.75	16	9	8	0	0	1	0	
11	11.25	18	15	7	0	0	0	0	
12	5.00	12	5	3	0	1	0	0	
13	3.60	12	26	4	0	1	0	2	
14	18.18	17	22	21	0	0	1	0	
15	6.25	16	8	2	0	1	0	0	
16	8.13	13	3	0	0	1	0	0	
17	8.77	12	15	0	0	0	1	2	
18	5.50	12	18	3	0	0	0	0	
19	22.20	12	31	15	0	0	1	1	

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	wage	educ	exper	tenure	nonwhite	female	married	numdep
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

CSVファイルのimport



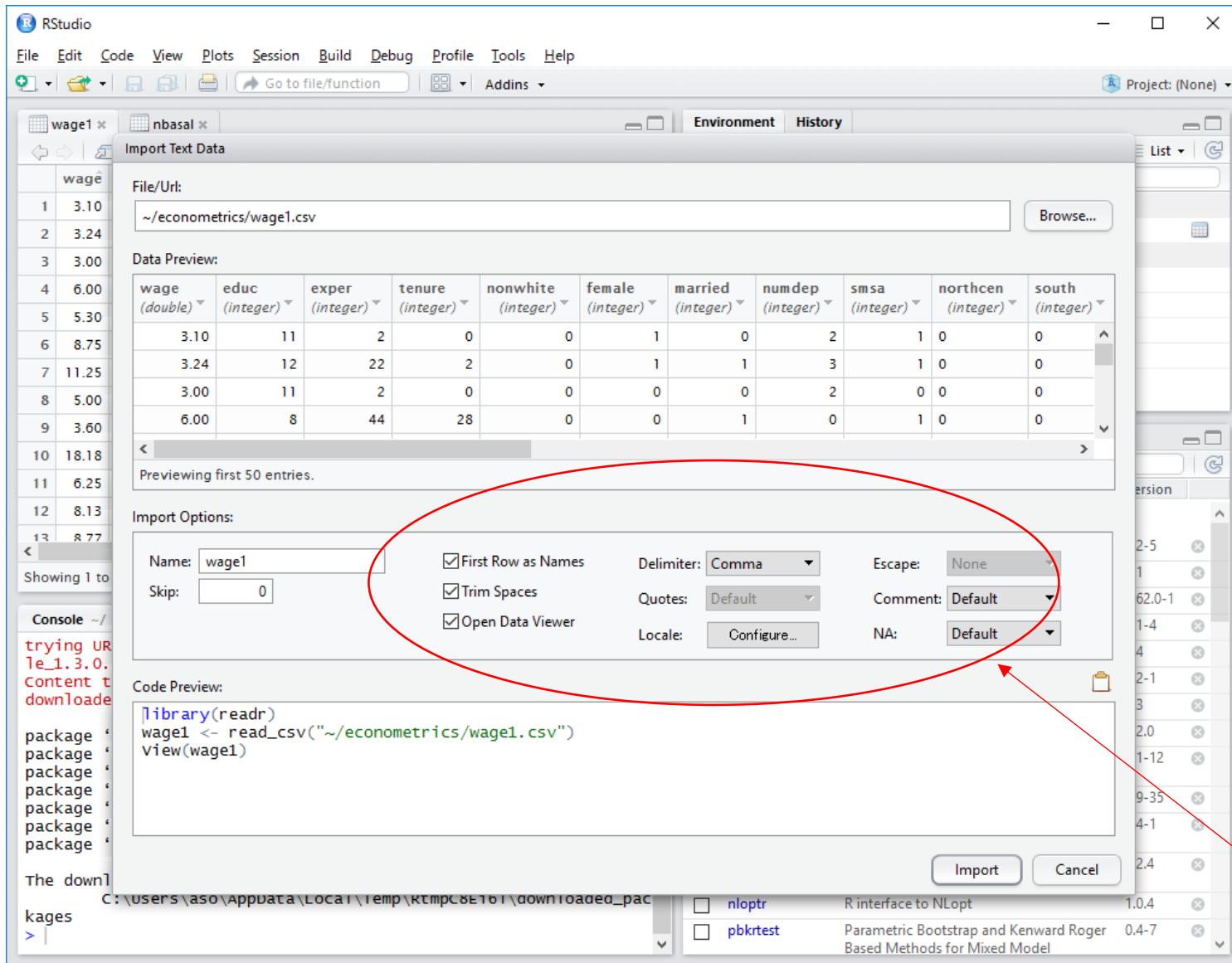
Rstudio右上のwindow
(Environment) の Import Datasets
タブを選択し,
From Text (base) .. を選択
→importするファイルを指定

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

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

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. On the left, the 'Console' window displays the log of package installations:

```
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
```

Below this, the user has run the commands:

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

A red oval highlights the last two lines of the console output.

In the center, the 'Import Dataset' window is open, showing the 'Data' tab with the dataset 'wage1' selected. It indicates 526 obs. of 24 variables.

On the right, the 'Environment' window shows the global environment with the 'Data' tab selected, displaying the same information about the 'wage1' dataset.

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

慣れてくれば、上のように、直接コマンドをタイプしても良い。
(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えるような構造もあり
- 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（異なるデータフレームの中に変数名が同じ別の変数の存在等の理由で）， 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しまったり, 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, 次回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)
```

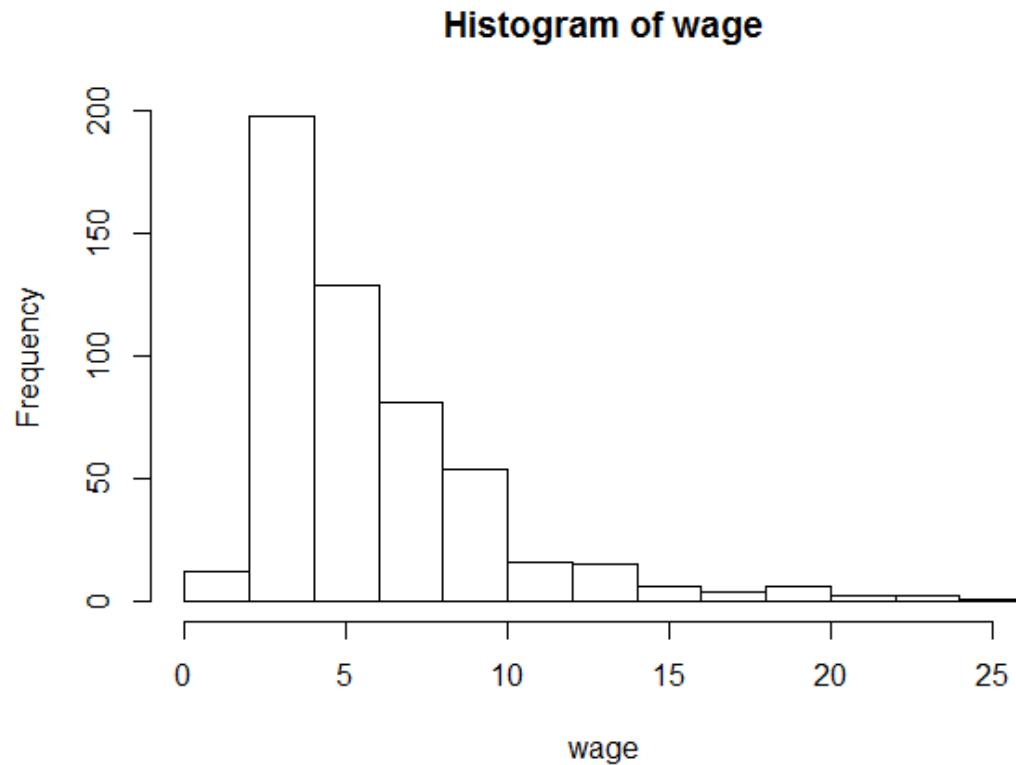
```
write.csv(wage2, file="ファイル名")
```

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

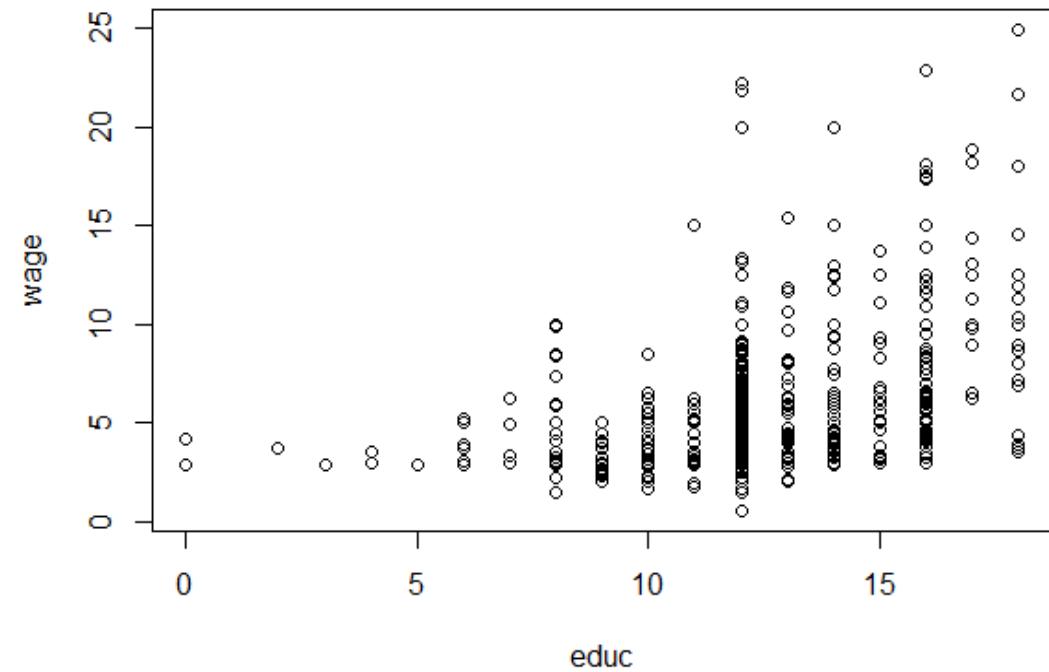
新しいデータフレームを作成したのは、もとのデータフレームをそのままにしたかったから（元の
データフレームに上書きしたい場合は、2行目をwage1\$lnwageとすればよい）

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	0.4059033	0.1129034	0.34688957
educ	0.4059033	1.0000000	-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

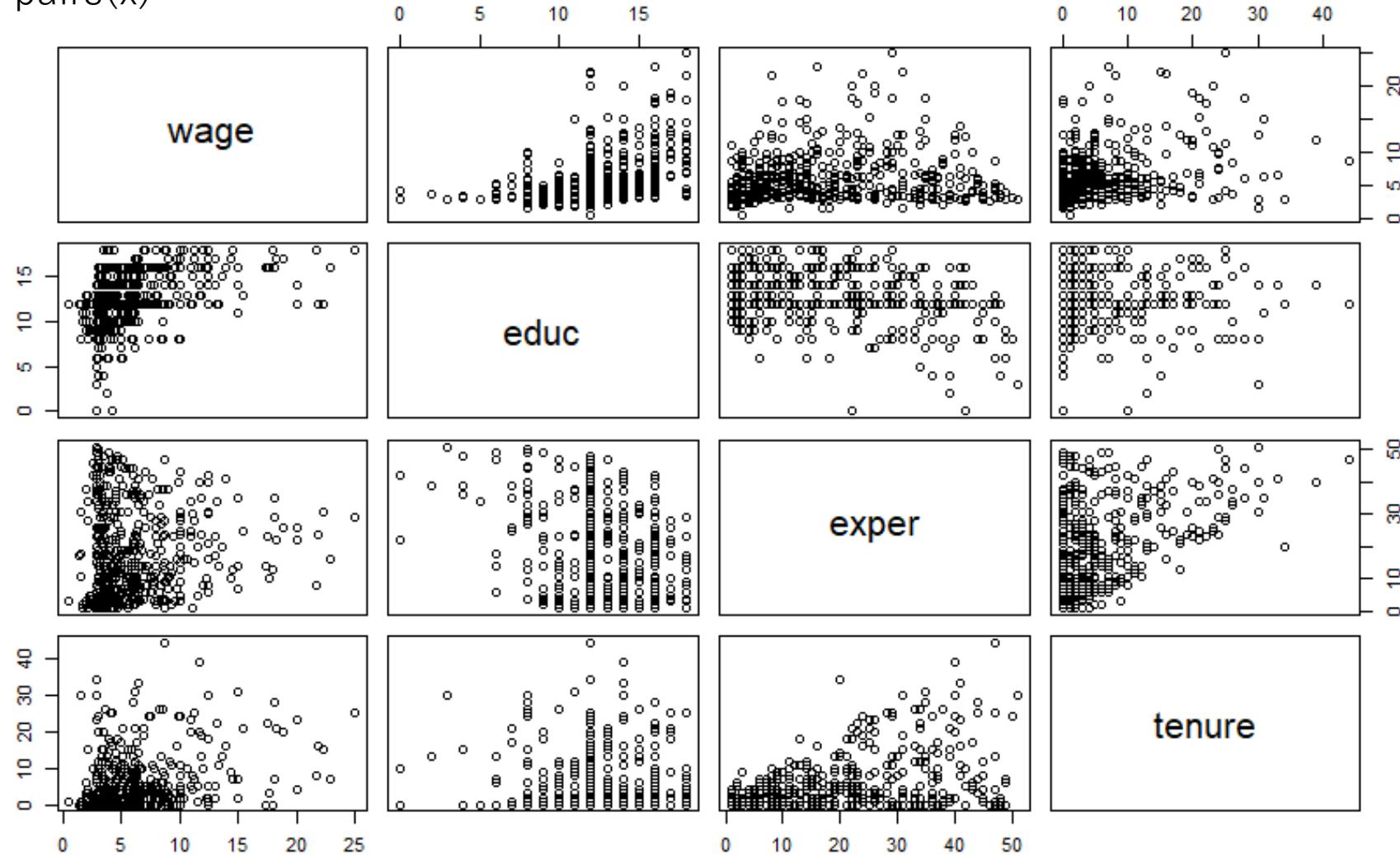
```
> var(x)
```

	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考すればよい。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、…）が作成できた季、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、ラグ付き変数が使える
 - ・ `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の回帰分析と同じ (`Imtest()` を用いる)
- ・ `dynlm`というpackageをインストールして, `library(dynlm)`でロードしておくる必要
 - ・ 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 Amazon.co.jpでも購入できる)

Internet上の資料

- An Inroduction 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）
 - Rをインストールすればこの文書もインストールされます。Windowsの場合、
C:\Prgram Files\R\R-x.y.z\doc\manual
にあります(x.y.zはRのバージョンを表す)
 - 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：G.V.Fransworth, “Econometrics in R”など
- 日本語ページ
 - RjpWiki: <http://www.okadajp.org/RWiki/?RjpWiki>

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