

バイオスタティスティクス基礎論 ガイダンス

岩田洋佳 hiroiwata@g.ecc.u-tokyo.ac.jp

2020/4/9

講義の内容

バイオスタティスティクス基礎論は、R と RStudio を用いて、実習形式で統計の初歩を学びます。R は統計解析のためのコンピュータ言語で、RStudio はそれを利用するためのソフトウェアの一つです。R をインストールするだけでも、R 言語を使うことができますが、本講義では RStudio を用いて授業を行います。

講義で扱う内容は、以下のとおりです。

1. R でデータを視覚化する (4 月 17 日)
2. 回帰分析、分散分析 (4 月 24 日)
3. 主成分分析、多次元尺度構成法 (5 月 1 日)
4. 階層的クラスタ解析、非階層的クラスタ解析 (5 月 8 日)

授業の進め方

ファイルは事前にダウンロードしてください

各回の授業は、いずれも、テキスト (PDF、HTML、Google Docs)、R コード (R)、データ (CSV) ファイルを用いて実習形式で進めていきます。全てのファイルについて、遅くとも講義の前日までには、アグリバイオのホームページおよび ITC-LMS からダウンロードできるようにしておきます。講義の前に必ずダウンロードを行っておいて下さい。

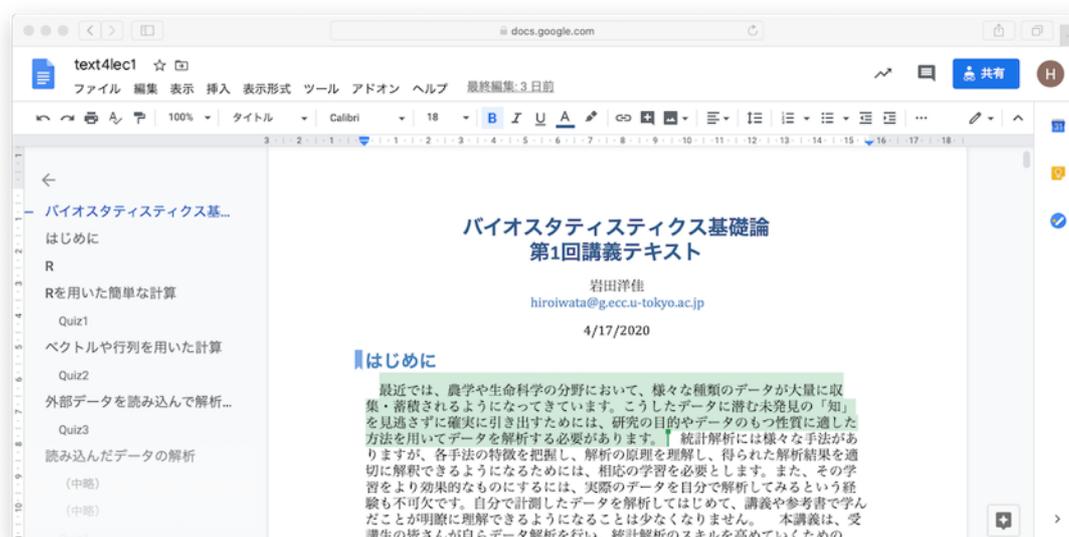
授業は音声による説明を基本にして進めていきます

本来であれば、映像も使いながら説明をすると分かりやすいかもしれませんが、本授業では、通信量を抑えるために、音声による説明をもとに進めていきます。zoom にはチャット機能もありますので、それで皆さんとのコミュニケーションをとりながら授業を進めていきたいと考えています。感じとしては、ラジオ番組の DJ のような感じで授業ができれば理想的だと考えています。なお、顔の見えない受講生に向かって授業をするのは結構辛いところもあります。チャットでの質問、コメントよろしくお願いします。

なお、テキストは、PDF も、HTML も、Google Docs も全て同じ内容ですが、Google Docs の場合は、私がどこに焦点をあてて話を進めているかが分かるのが利点です。来週の授業の Google docs (準備中) は、次のリンクをクリックすると閲覧できます。

<https://docs.google.com/document/d/1Ozj5aN-AyqXizBKHPSQIDM4--pX61jGemoQw2L6eK0M>

Google Docs では、下図のように、私がどの部分に焦点をあてているか（選択しているか）を見ることができます。

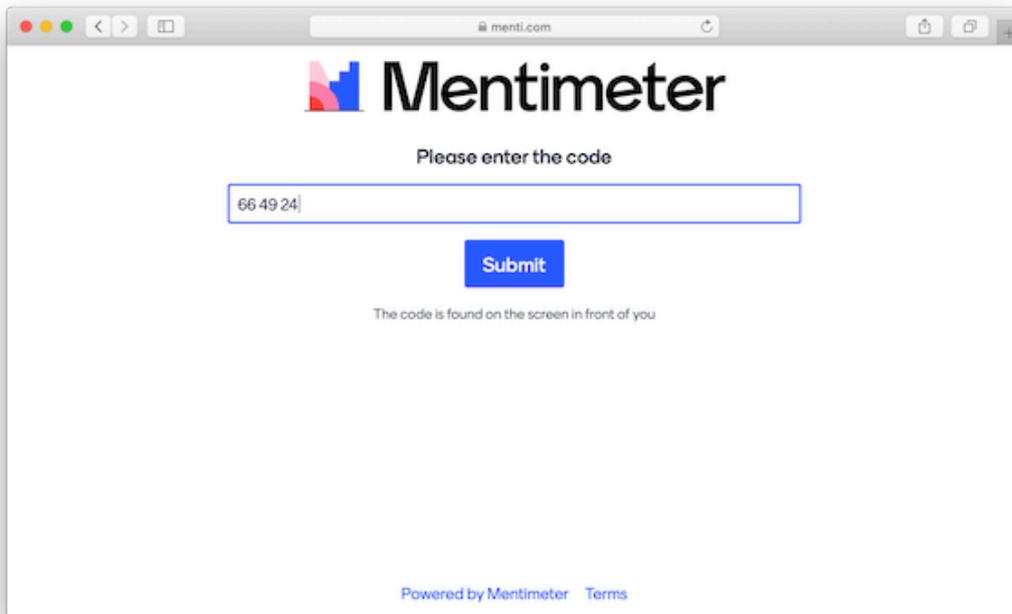


授業では、私がテキストの特定の場所に焦点をあてながら進めていきます。皆さんは、テキストと同時に配布される R コードを用いて、私の説明に従って R コードを実行して下さい。なお、実行しているコードの説明、統計的な意味については、テキストにしたがって、私が音声で説明していきます。

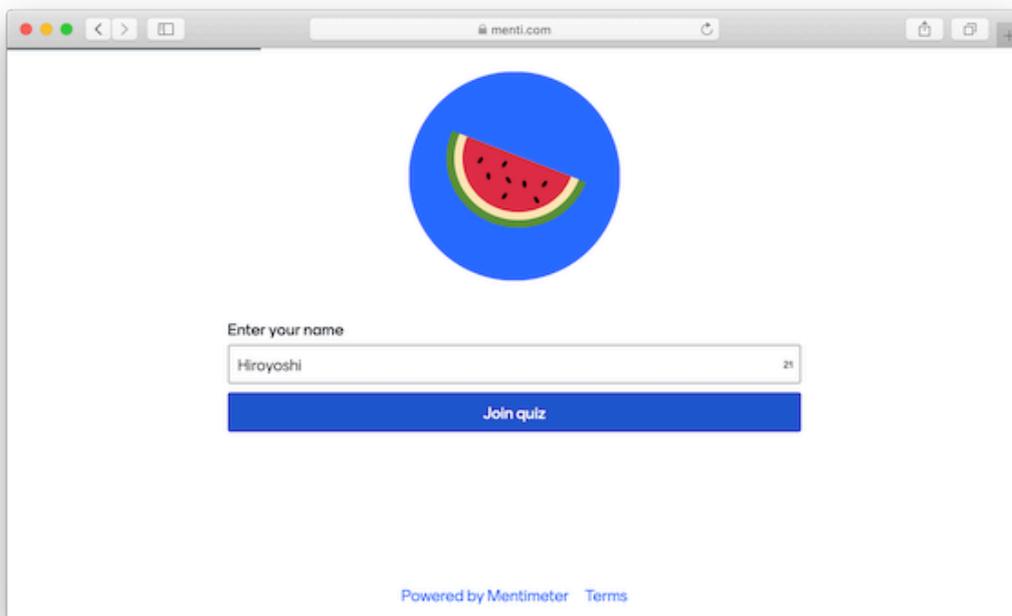
授業の途中でクイズを行う予定です

授業に、少しメリハリやゲーム性をつけるために、途中でクイズを実施します。ここでも、少し練習をしてみましょう。

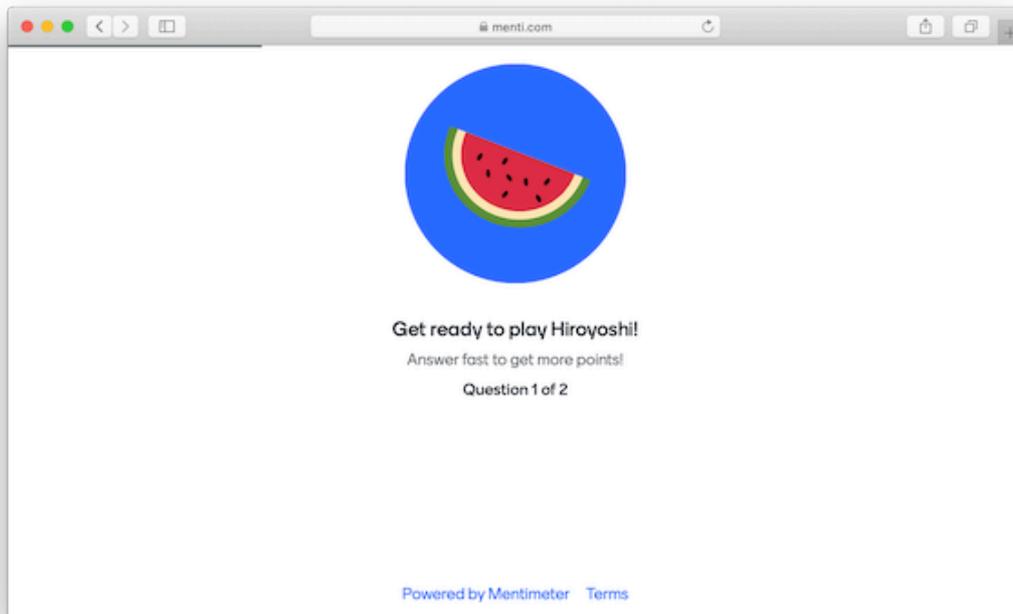
1. まず、<http://www.menti.com/> に行き、「66 49 24」というコードを入力して下さい。



- 次に、ニックネームを入れて、“Join quiz”をクリックして下さい。



- すると待機状態になります。しばらくするとクイズが始まります。皆さんの健闘を祈ります。



なお、このクイズは一種のゲームで、成績評価には一切関係しません。楽しんで取り組んで下さい。

成績の評価

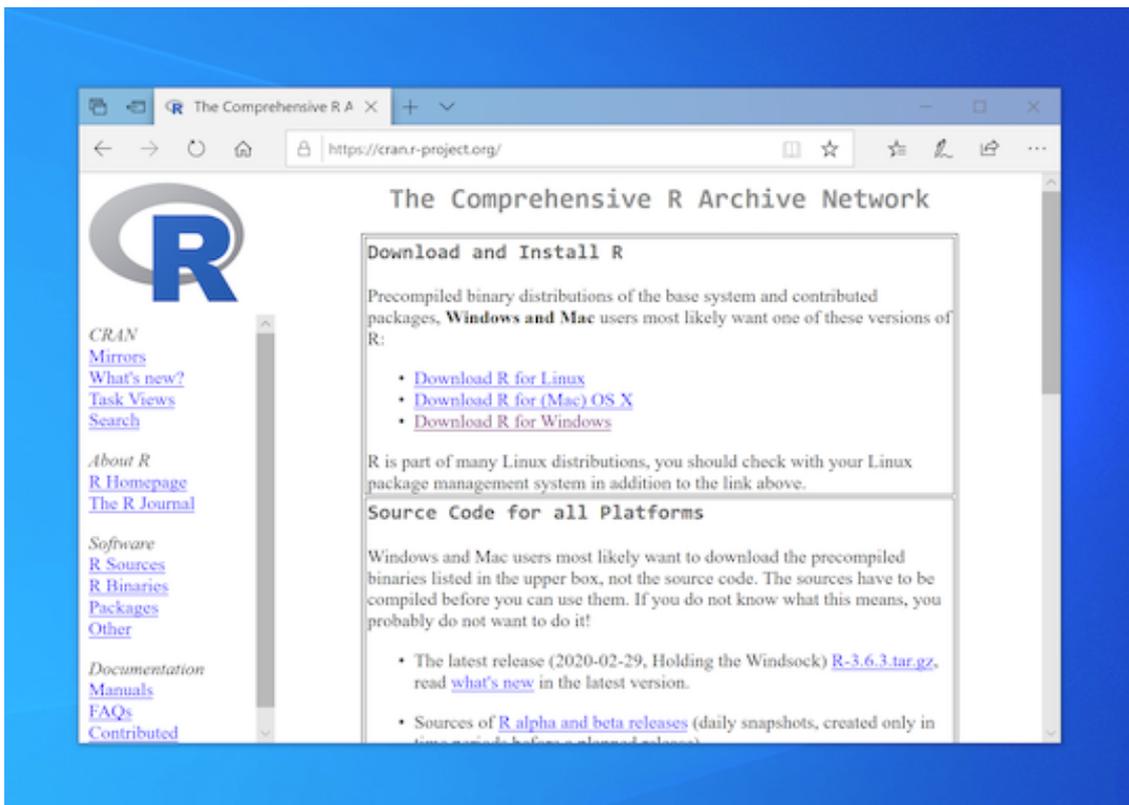
成績の評価は、出席（ITC-LMS を用いて出席をとる予定です）と毎回出されるレポート（ITC-LMS を用いて提出してもらう予定です）をもとに評価します。

R と RStudio のインストール

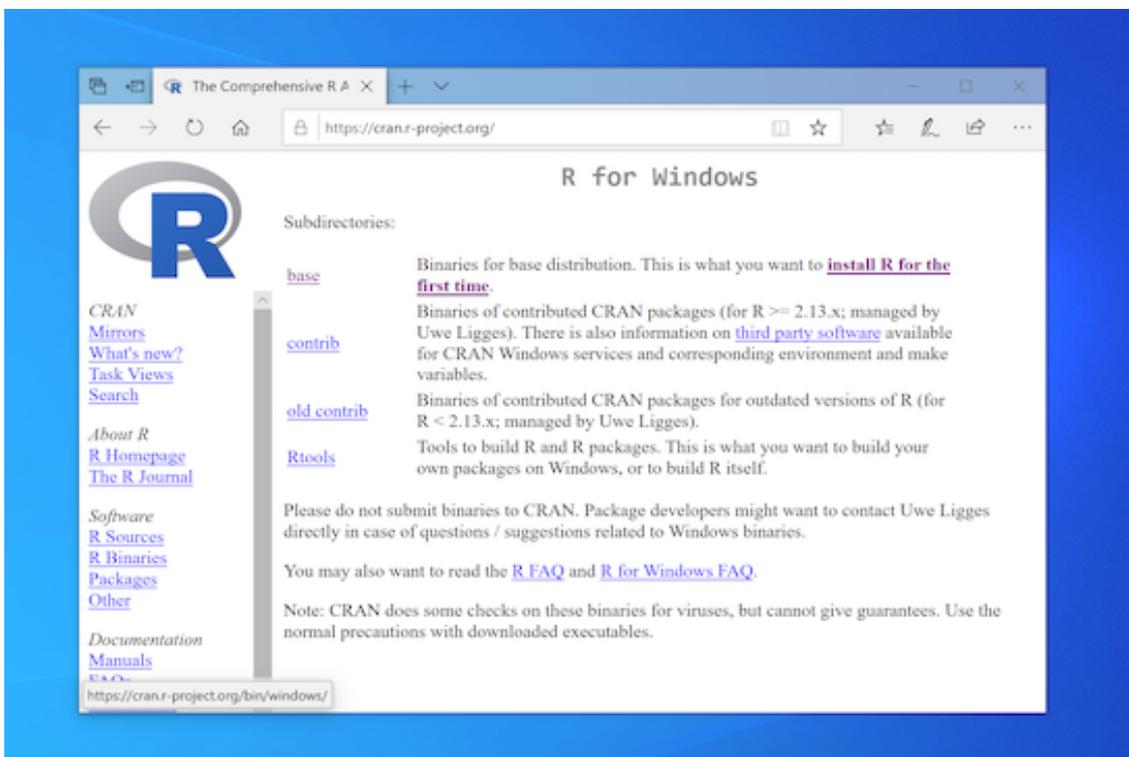
Windows の場合

R のインストール

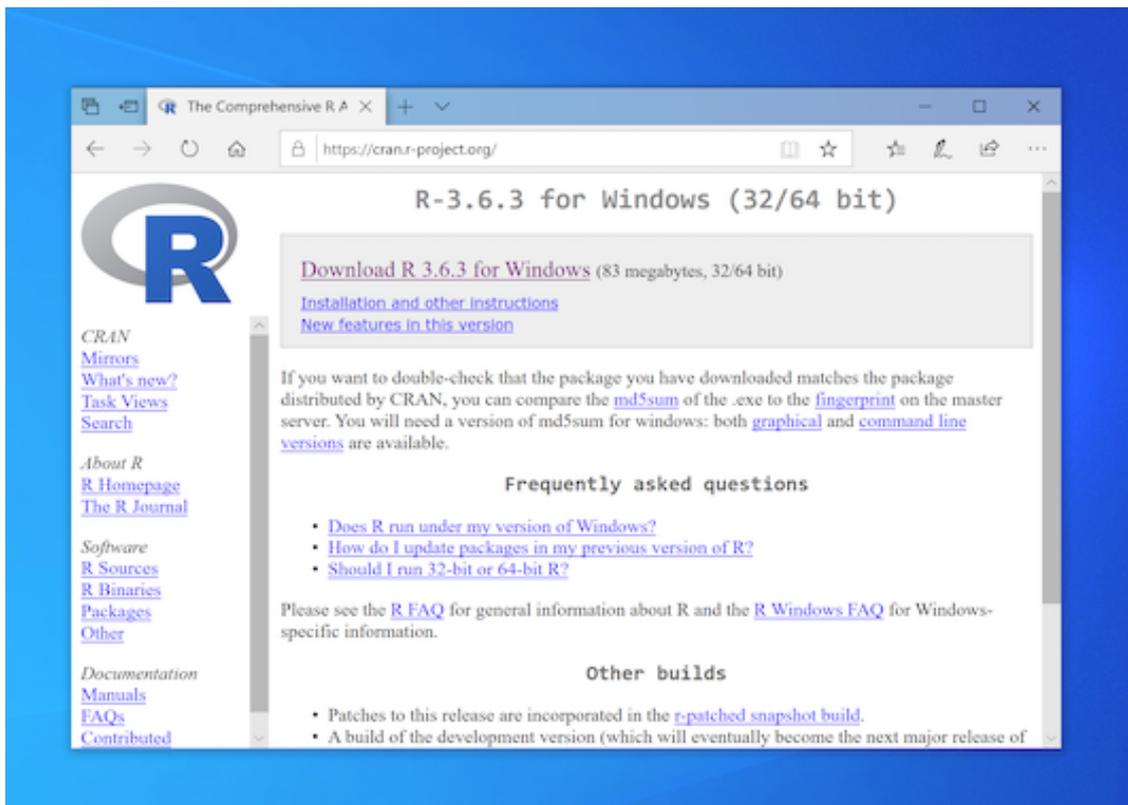
1. まず、CRAN のサイト <https://cran.ism.ac.jp/> に移動します。
2. “Download R for Windows”をクリックします。



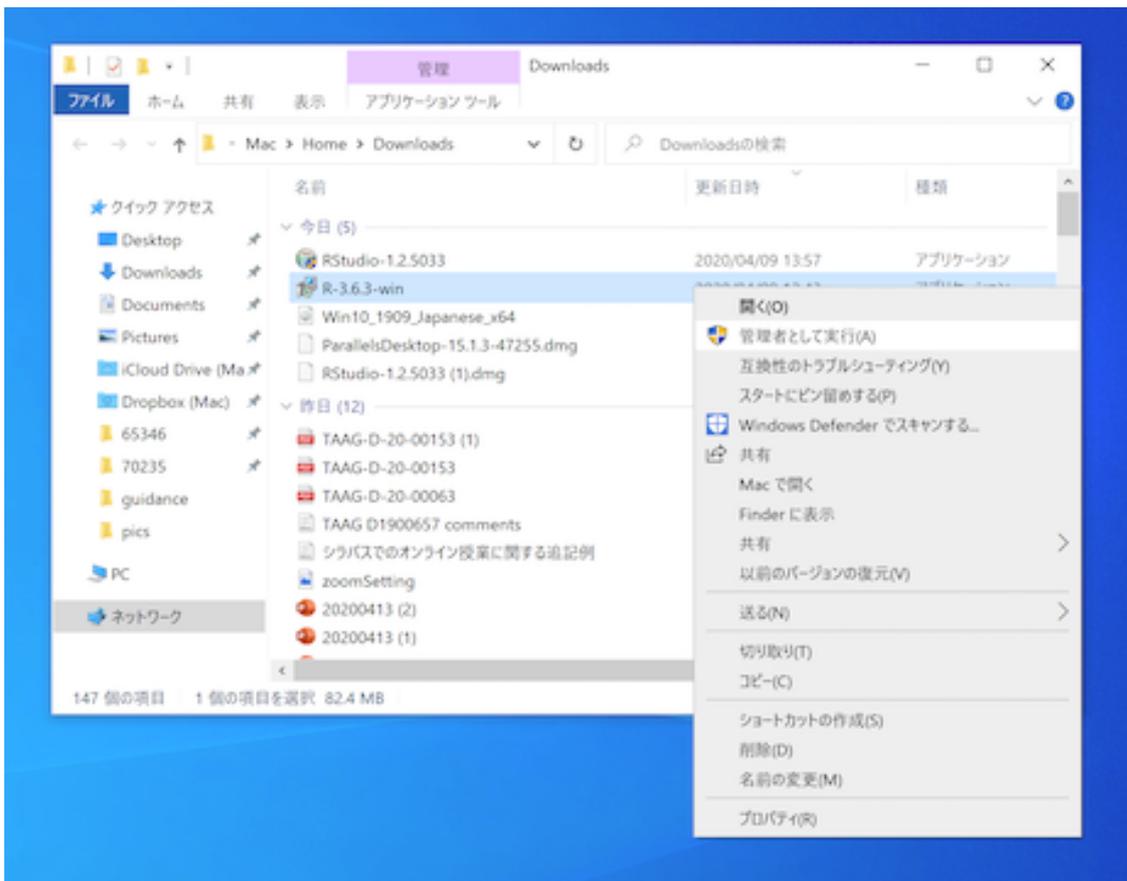
3. “base”を選択します。



4. “Download R 3.6.3 for Windows”をクリックします。最新バージョンはそのときによって異なります。



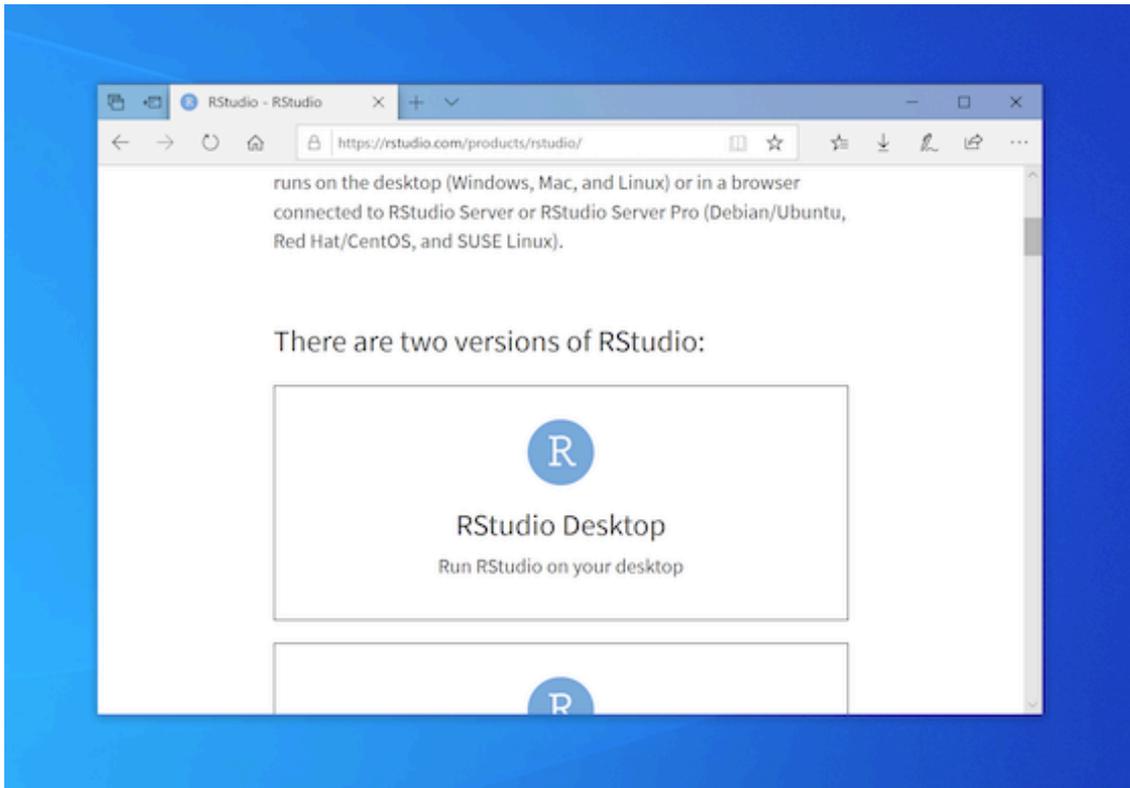
5. ダウンロードをするときは、「実行」ではなく、「保存」をして下さい。
6. ダウンロードして保存されたファイルをクリックして、R をインストールします。なお、必ず「管理者として実行」（管理者権限でインストール）してください。



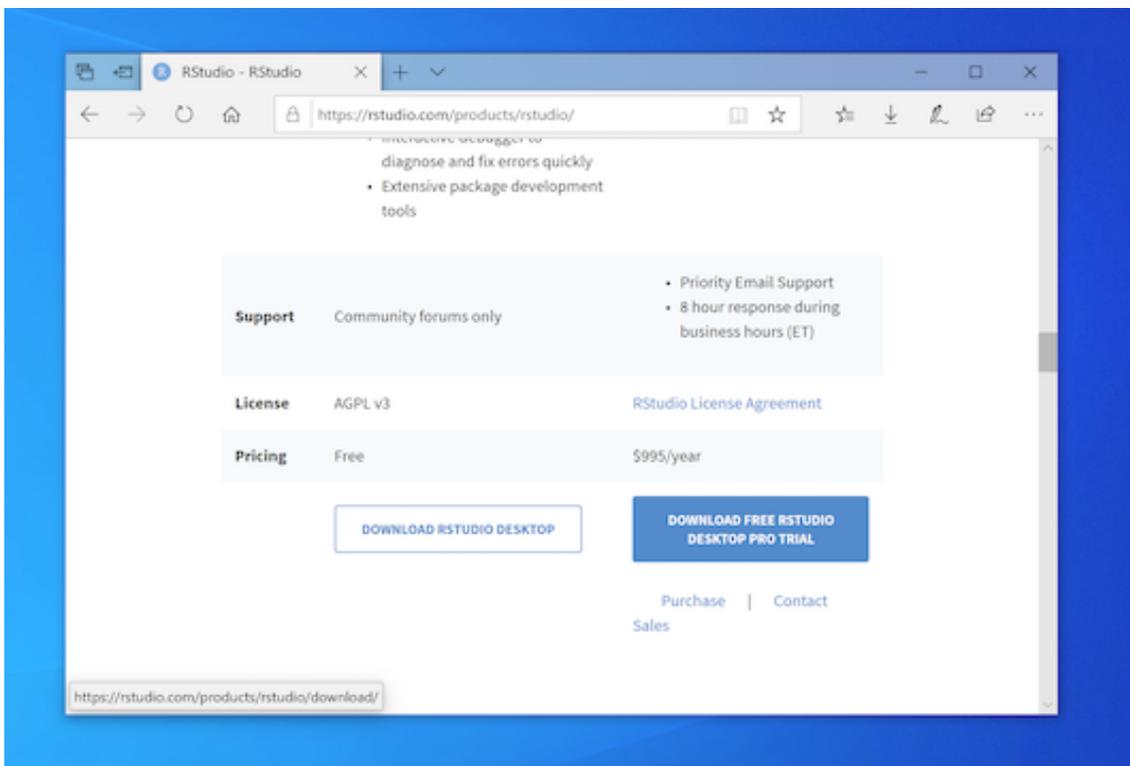
7. インストール時のオプションは、特に変更をする必要はありません。デフォルトの設定のままでインストールを進めていきます。

R Studio のインストール

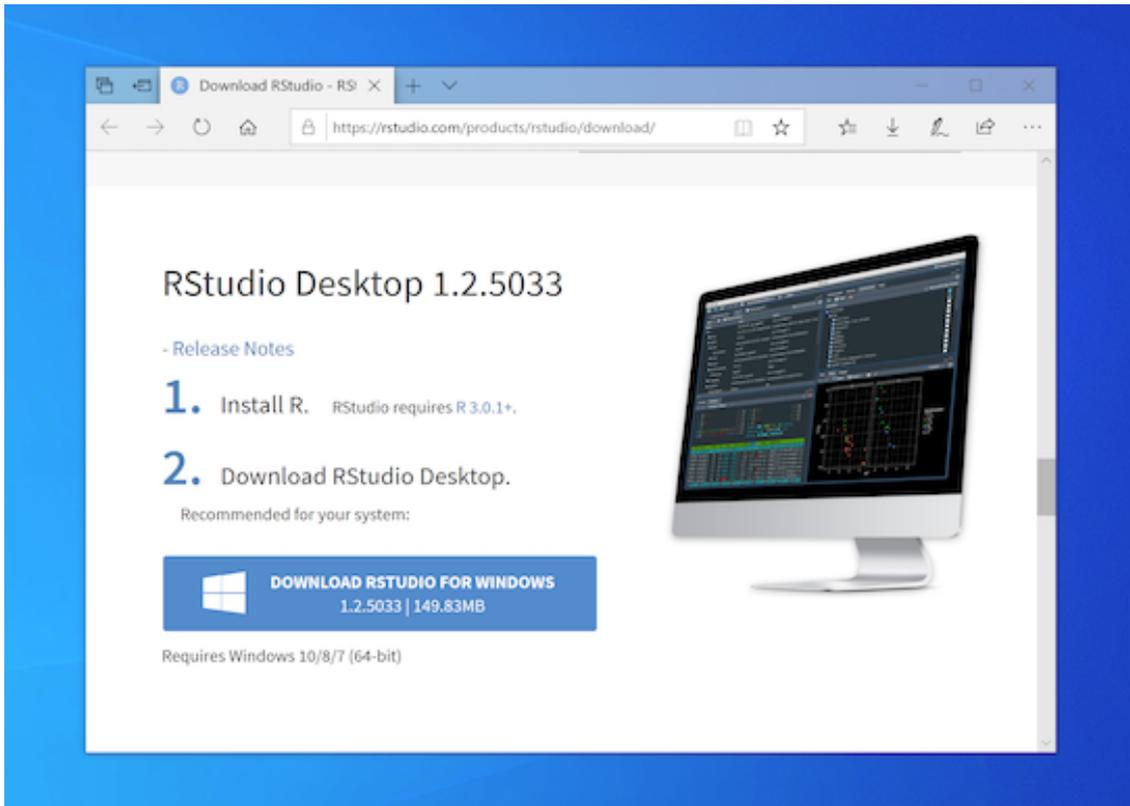
1. RStudio のダウンロードサイト <https://rstudio.com/products/rstudio/> に移動します。
2. 次に、“RStudio Desktop”をクリックします。



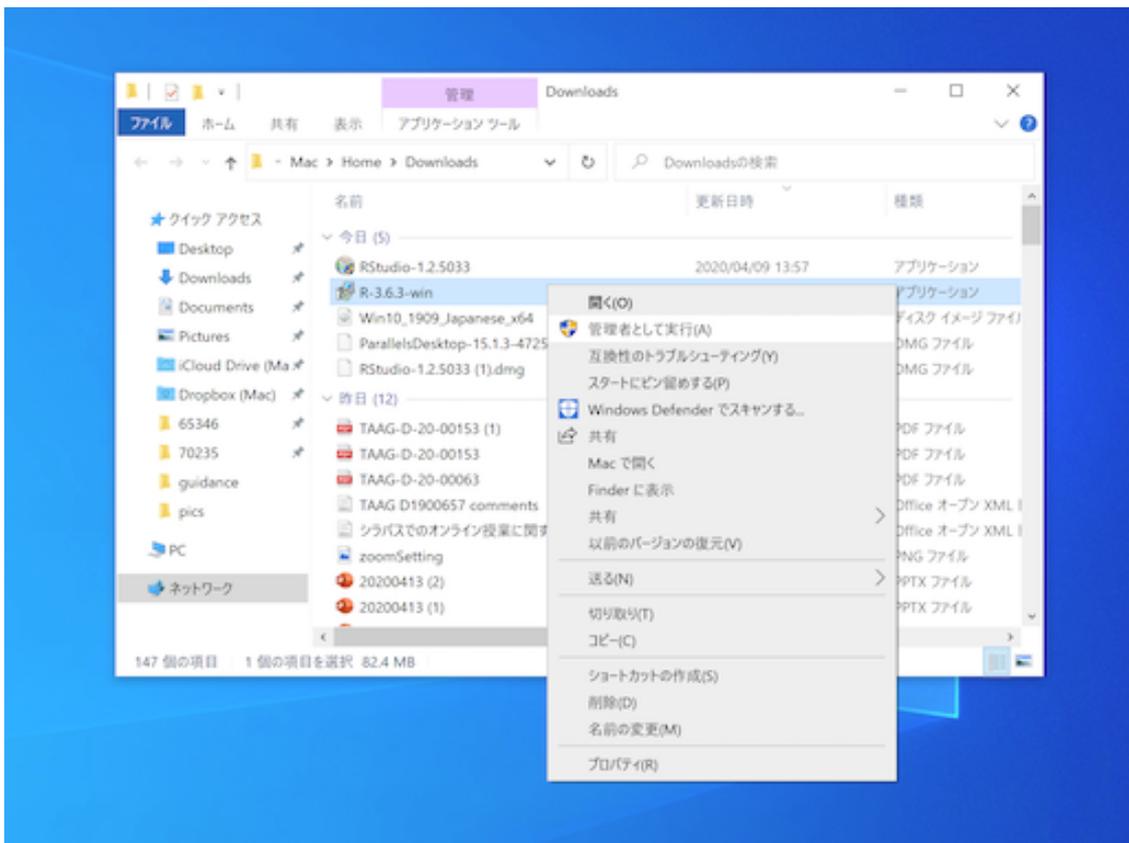
3. “DOWNLOAD RSTUDIO DESKTOP”をクリックします。



4. 一番左（フリー版の RStudio Desktop）の“DOWNLOAD”をクリックします。



5. ダウンロードをするときは、「実行」ではなく、「保存」をして下さい。
6. ダウンロードして保存されたファイルをクリックして、**RStudio** をインストールします。なお、必ず「管理者として実行」（管理者権限でインストール）してください。

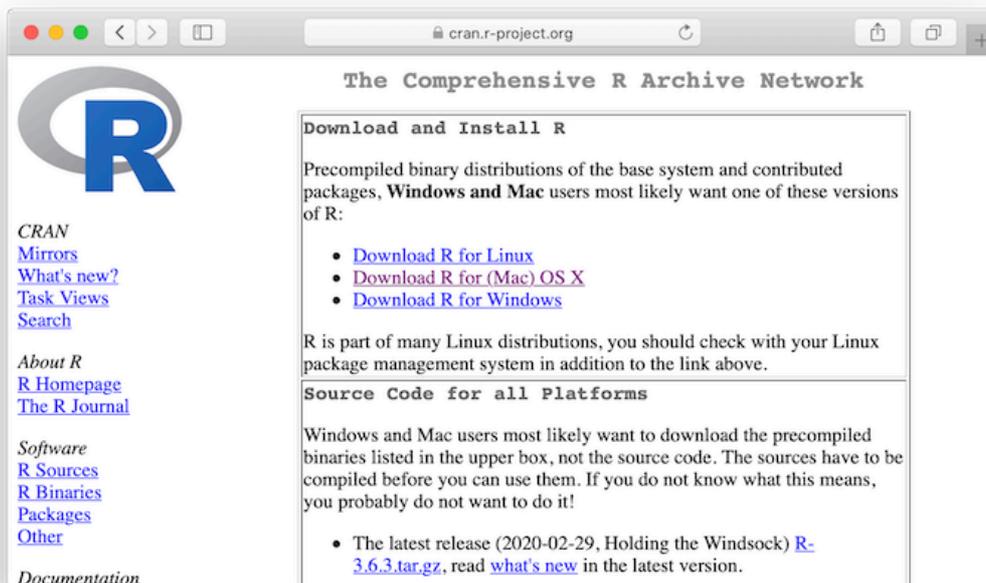


7. インストール時のオプションは、特に変更をする必要はありません。デフォルトの設定のままでインストールを進めていきます。

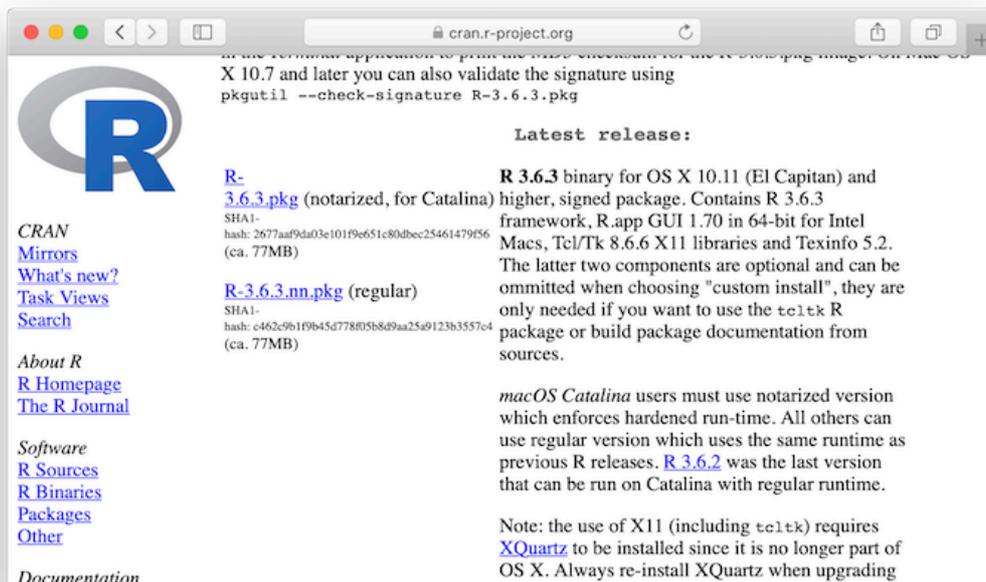
Mac OS の場合

R のインストール

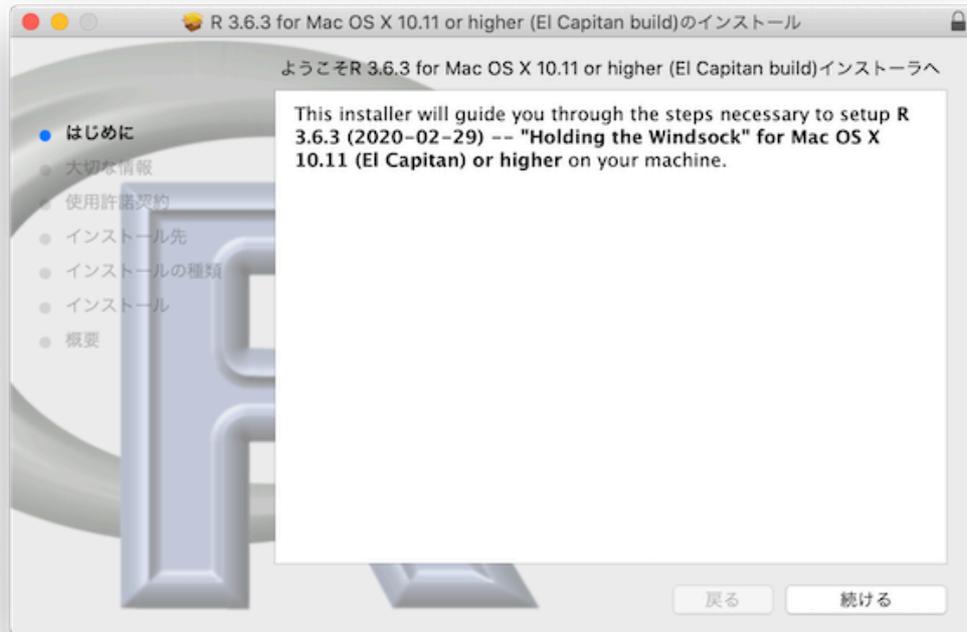
1. まず、CRAN のサイト <https://cran.ism.ac.jp/> に移動します。
2. “Download R for (MAC) OS X”をクリックします。



3. 次の図の左下にある“R-3.6.3.pkg”をクリックします。最新バージョンはそのときによって、異なると思います。



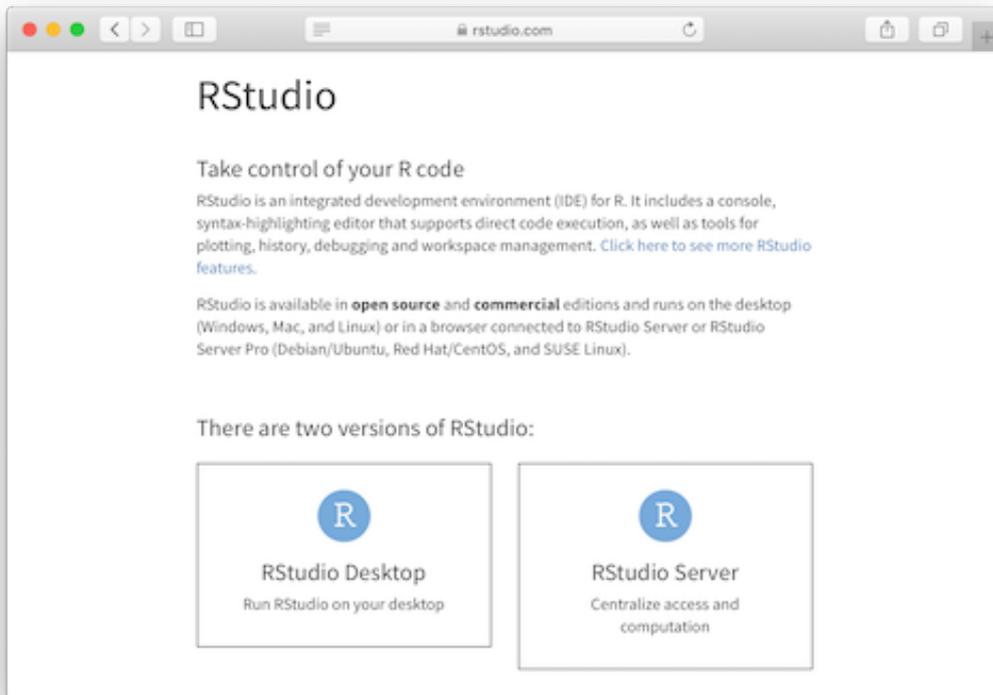
4. ダウンロードして保存されたファイルをクリックして、R をインストールします。



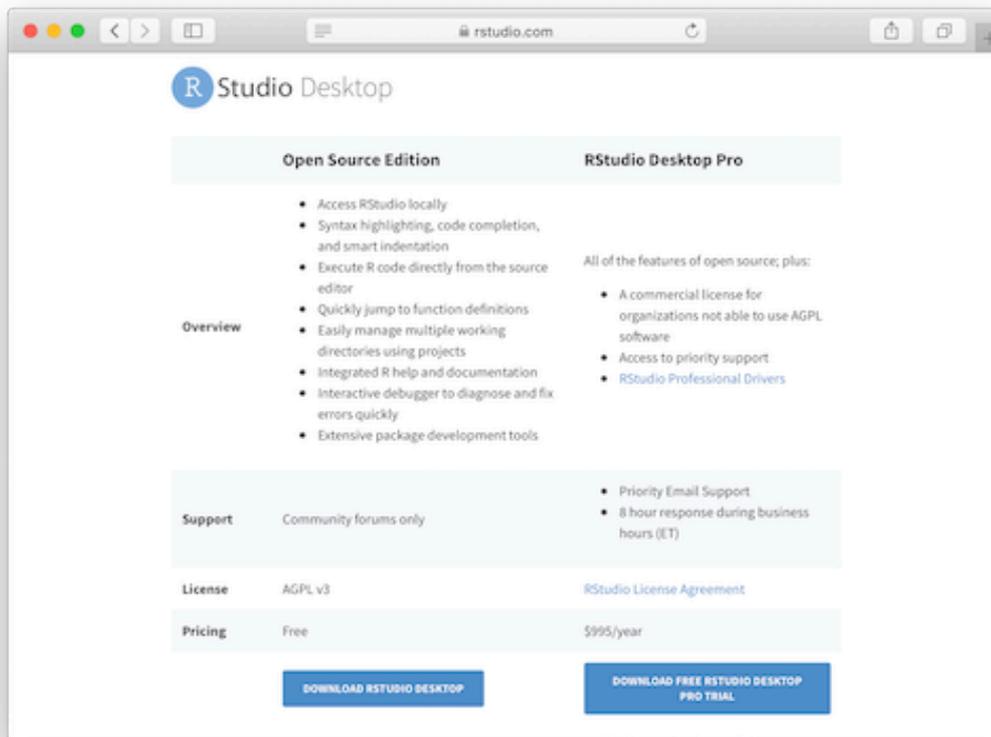
5. インストール時のオプションは、特に変更をする必要はありません。デフォルトの設定のままでインストールを進めていきます。

R Studio のインストール

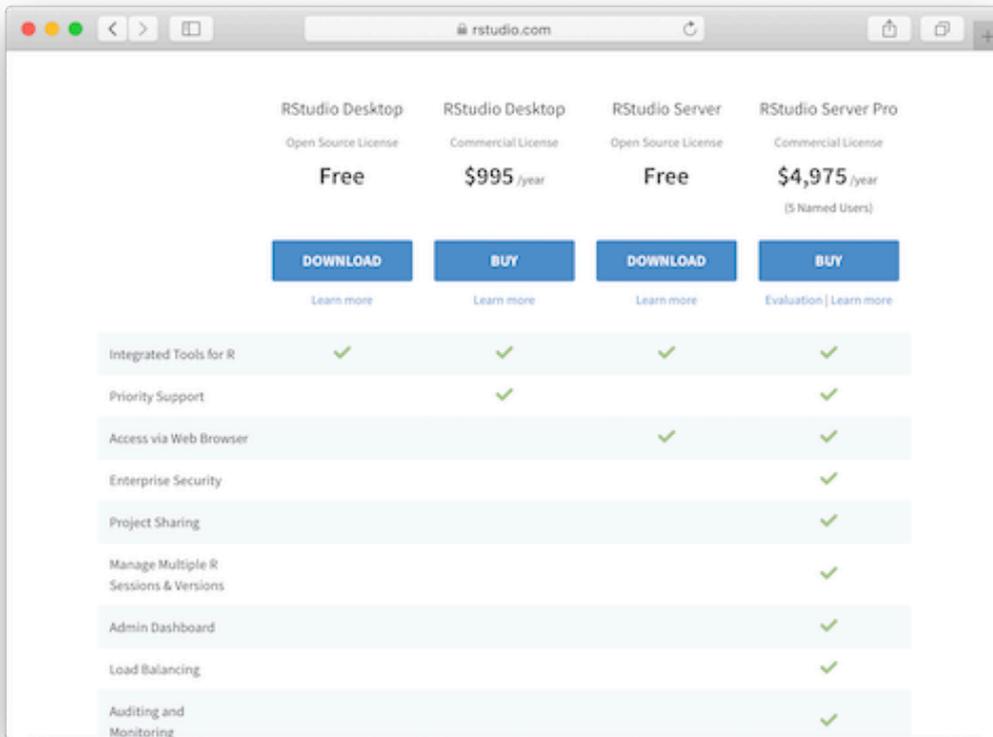
1. RStudio のダウンロードサイト <https://rstudio.com/products/rstudio/> に移動します。
2. 次に、“RStudio Desktop”をクリックします。



3. “DOWNLOAD RSTUDIO DESKTOP”をクリックします。



4. 一番左（フリー版の RStudio Desktop）の“DOWNLOAD”をクリックします。



5. ダウンロードして保存されたファイルをクリックすると以下のように表示されます。あとは、RStudio をドラックして、Applications の上にドロップします。



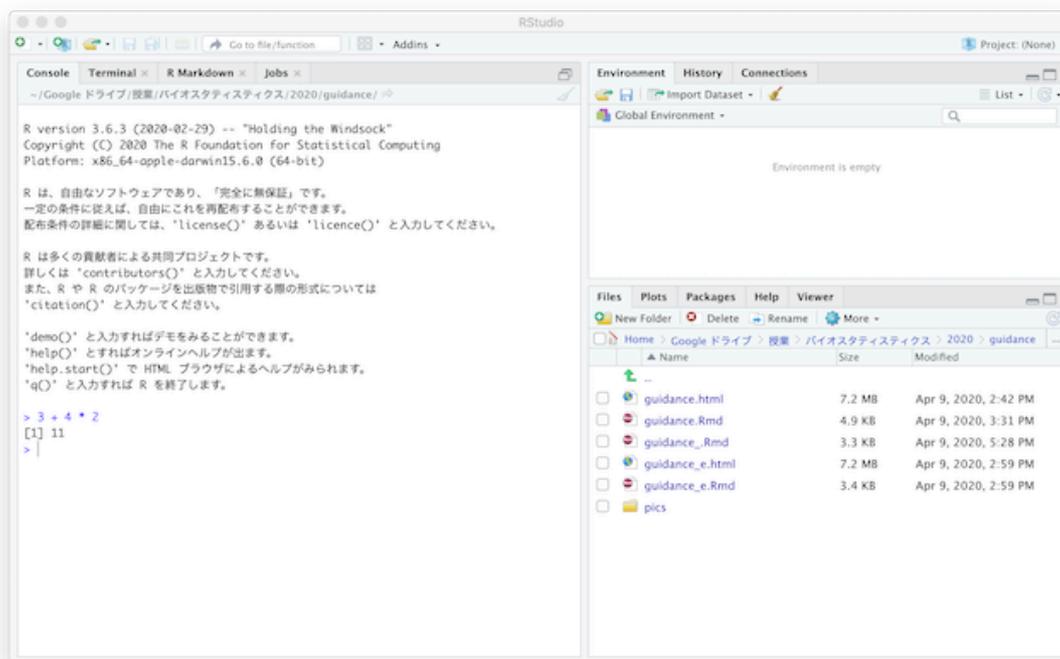
R と RStudio を試してみる

R と RStudio が動くことを確認する

ここからは、Windows と Mac OS を一緒にして説明していきます。それは、RStudio のインターフェースが Windows と Mac OS で一緒だからです。

まずは、左側の Console に、次のように打ち込んでみましょう。

```
3 + 4 * 2
## [1] 11
```



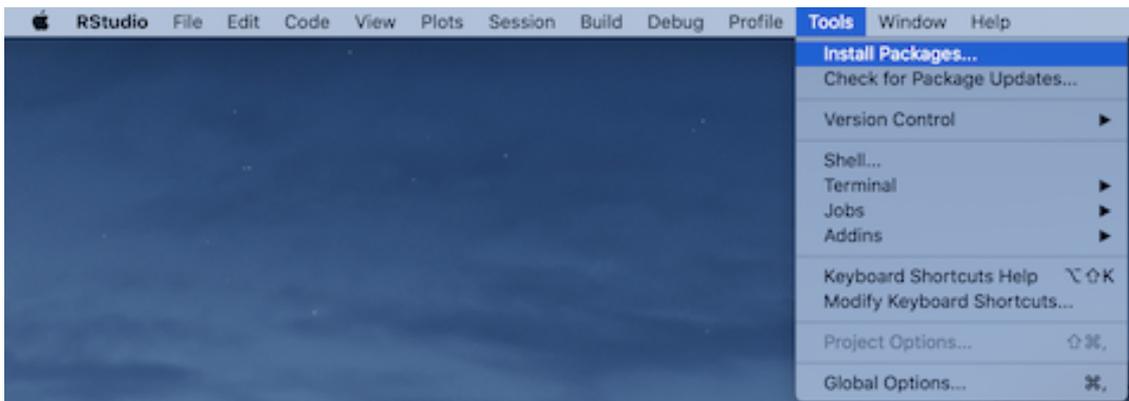
計算が実行され、結果が返ってきます。R の使い方については、来週からの授業で詳しく説明するのでこのくらいにしておきましょう。

R と RStudio を講義で使うための準備をする

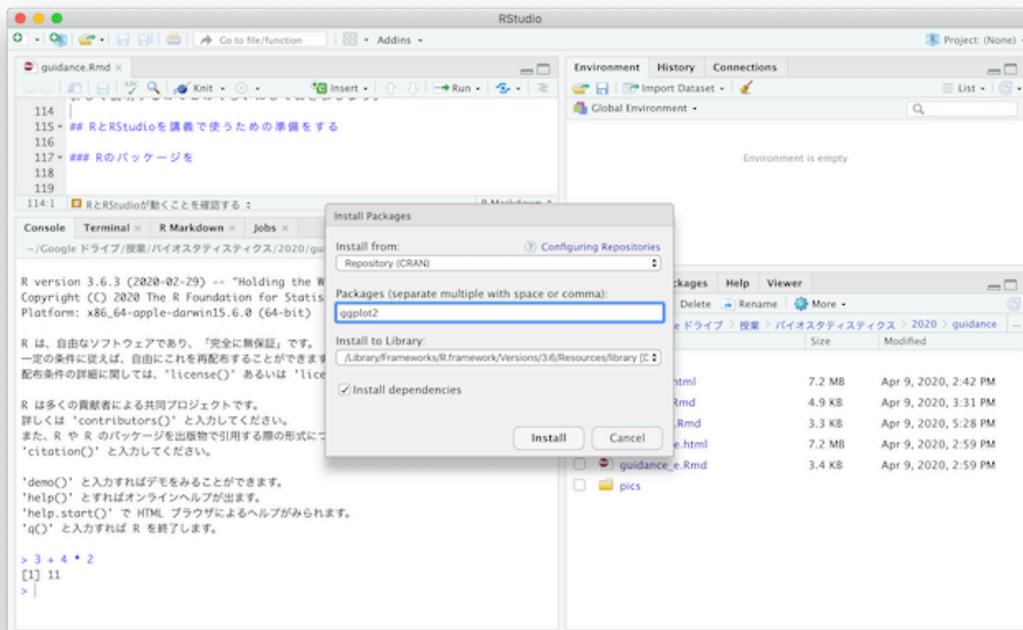
R のパッケージをインストールする

R は、そのままでもいろいろな統計解析を実施することができますが、パッケージとよばれる「拡張セット」のようなものをインストールすると、利用できる解析手法が大幅に広がります。ここでは、パッケージのインストールの仕方について説明します。なお、インストールを行うためには、パソコンがインターネットに接続されていなければなりません。

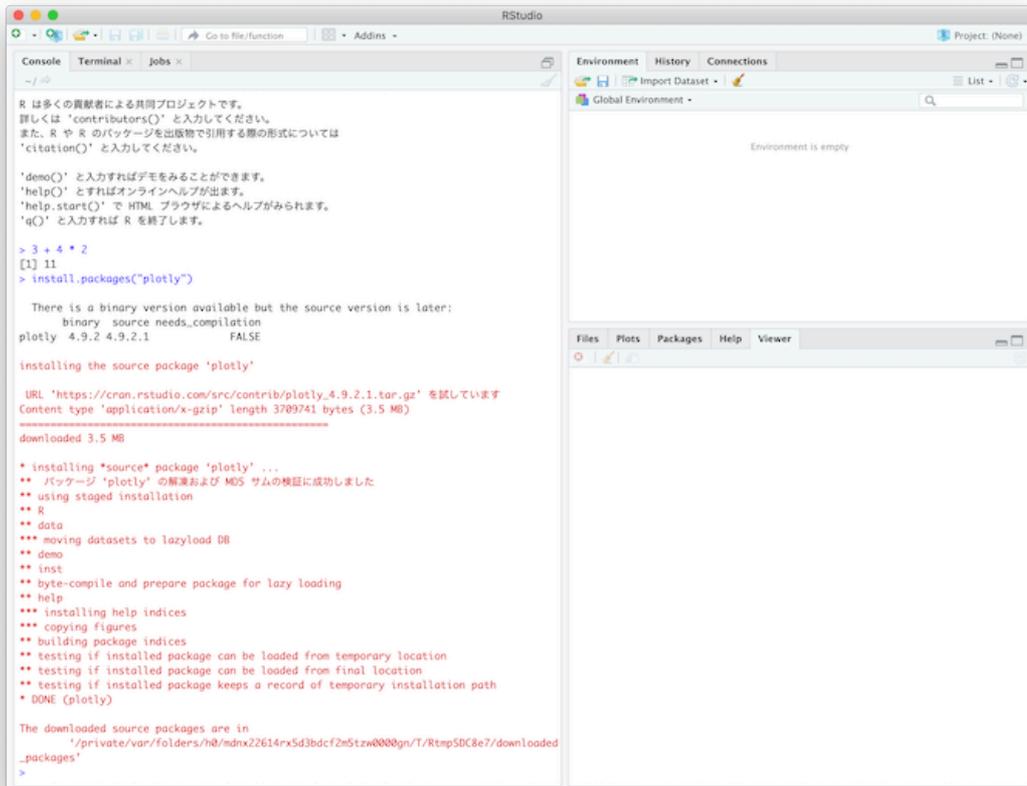
1. RStudio のメニューの Tools から、“Install Packages”を選びます。



2. そのあと示されるダイアログボックス（下図中央）の **Packages** に、インストールしたいパッケージ名を入力します。ここでは、**plotly** というパッケージをインストールしてみます。



3. **plotly** パッケージがダウンロードされてインストールされます。

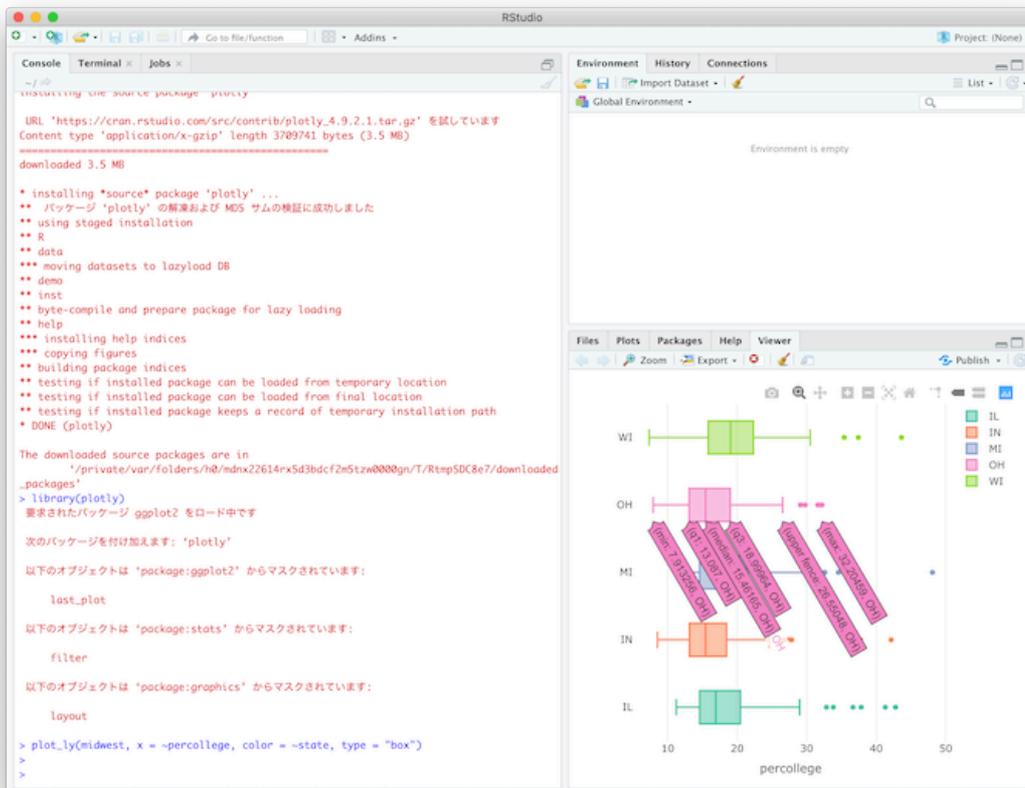


4. 以下のコマンドを Console に打ち込んで、図を描いてみます。

```
library(plotly)
```

```
plot_ly(midwest, x = ~percollege, color = ~state, type = "box")
```

RStudio で実行すると、このように表示されます。



R の Working Directory を設定する

R では、R を実行する「作業場」“Working Directory”を指定しなければならない場合が少なくありません。これは、授業で使う R コードやデータを R が見つけられるようにするためです。ここでは、RStudio を使って、これを指定してみましょう。

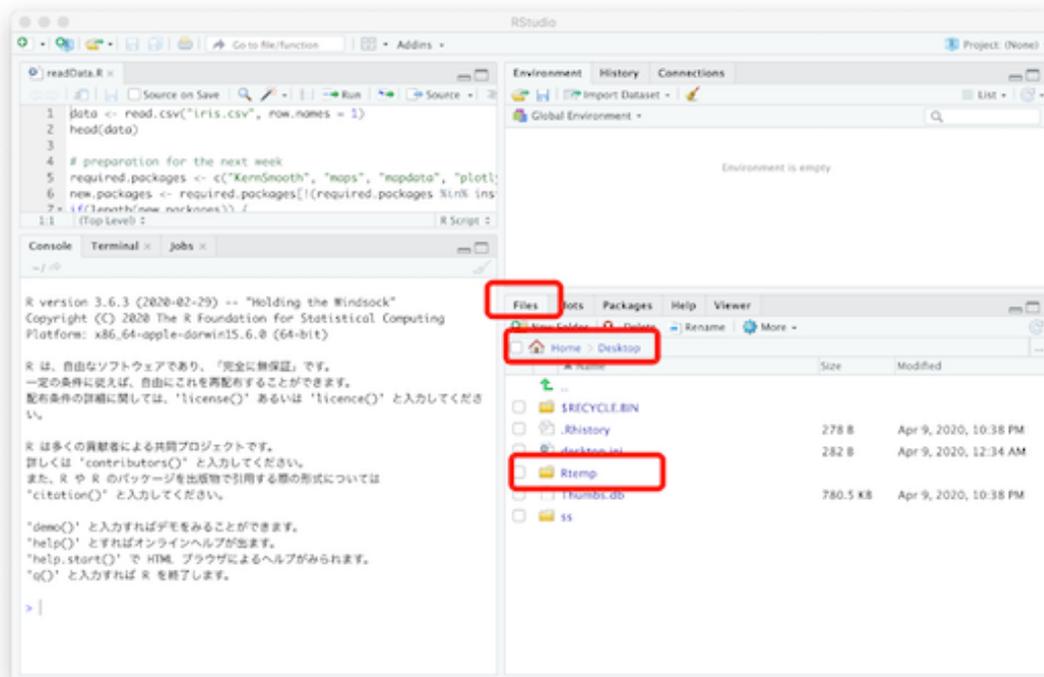
まず最初に、以下の Google Drive から、使用する R コード (readData.R) とデータ (iris.csv) をダウンロードします。

<https://drive.google.com/drive/u/2/folders/160TTfRf0Lz9qxm8vLdciaYZmSOZr9BL>

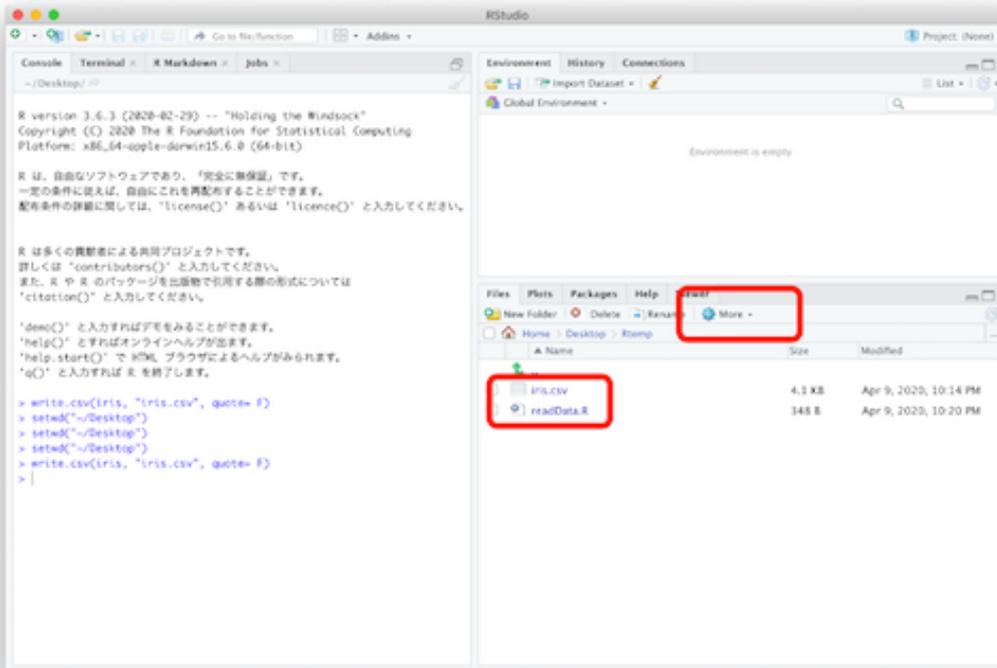
次に、これらのファイルをどこか自分の作業しやすいフォルダ (ディレクトリに移動してみましょう)。ここでは、デスクトップに Rtemp というフォルダを作って保存したとします。

RStudio で、このフォルダを“Working Directory”として指定するには、以下のようにします。

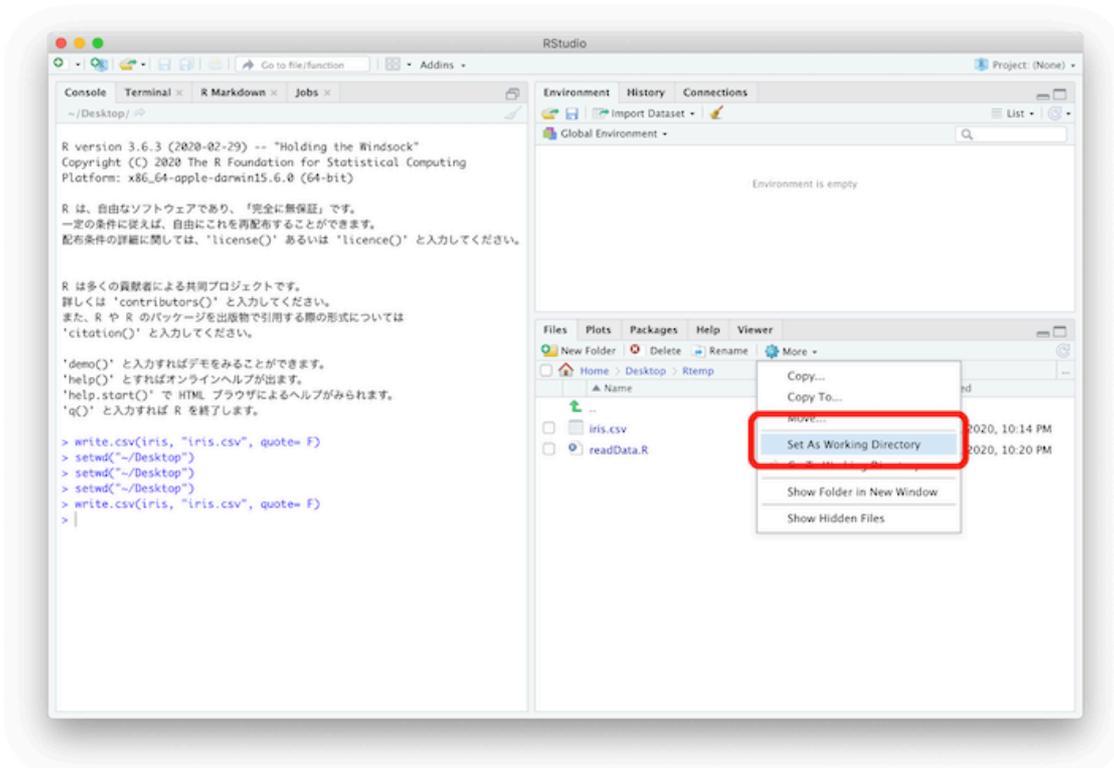
1. RStudio の右側下にある「Files」というタブをクリックします。



2. 「Home > Desktop」となっているところや、その下に示されているフォルダをクリックすることで、パソコン内にある様々なフォルダに移動できます。
3. ここでは、「Rtemp」に移動したいので、Rtemp をクリックします。保存しておいた R コードやデータのファイルがあることが分かります。
4. R コードやデータがある場所まで移動できたら、歯車型のアイコンの横に「More」と書いてあるところをクリックします。



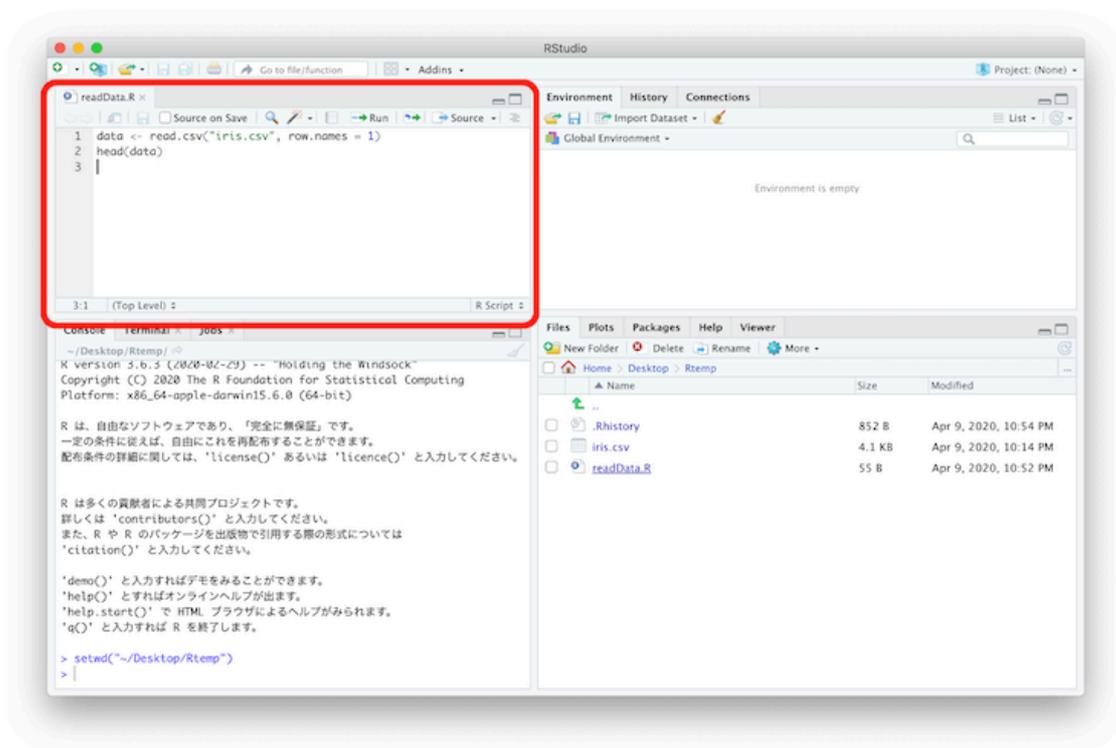
5. するとドロップダウンメニューが示されますので、「Set As Working Directory」を選びます。



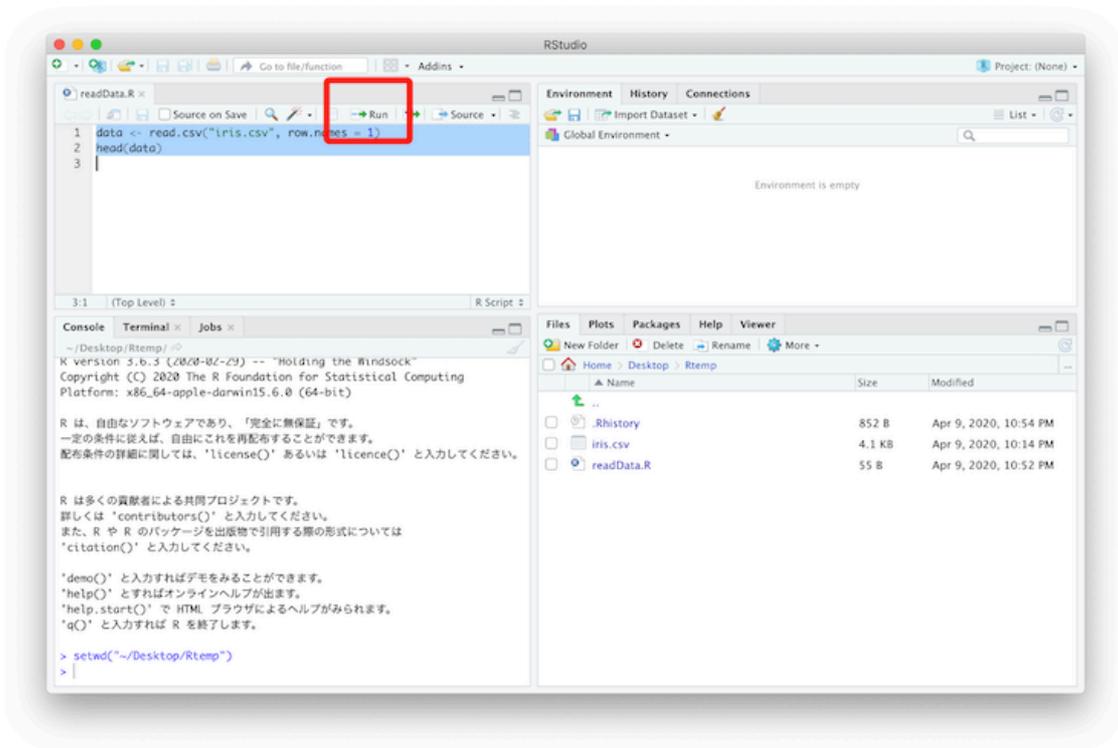
6. これで、“Working Directory”をデスクトップ上にある Rtemp というフォルダにすることができました。

こうすることで、R からデータファイルを簡単に読み込むことができますようになります。

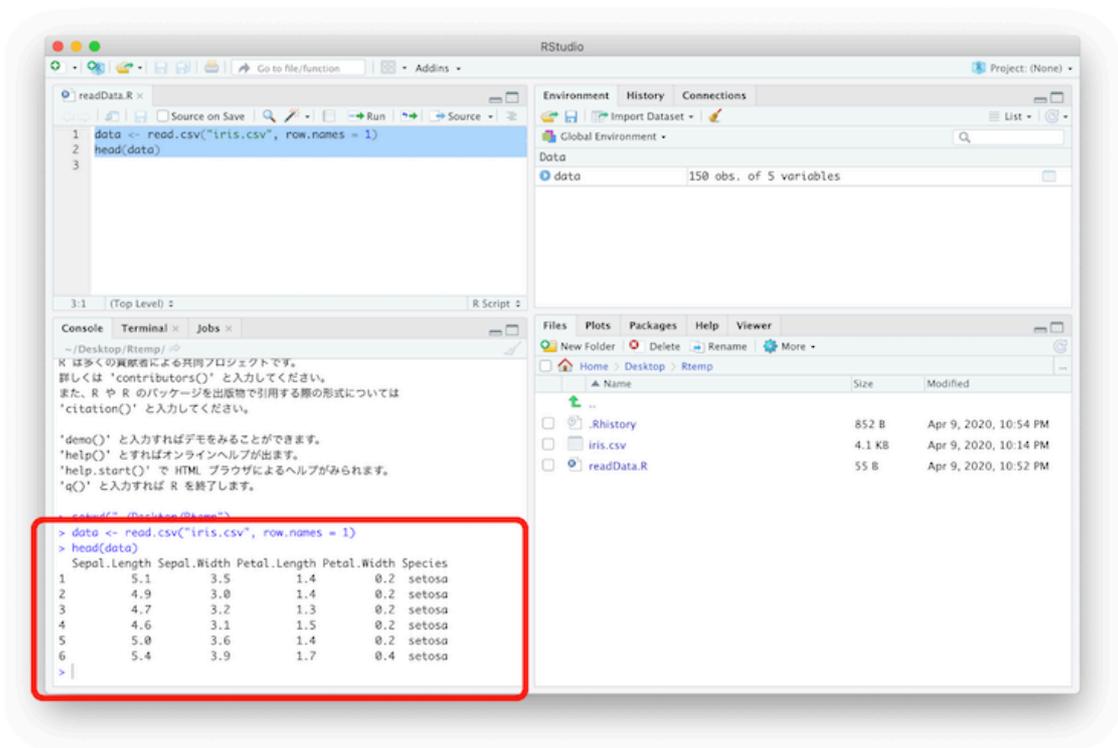
まずは、R コードを開いてみましょう。右下にある「readData.R」というファイルをクリックすると、左上に読み込んだコードが示されます。



このうち、最初の2行を選択して、実行してみましょう。実行は、Run というアイコンをクリックします。あるいは、Mac OS では「command + return (enter)」、Windows では「ctrl + enter (return)」でも実行できます。



うまく実行できると、以下のように表示されます。



ここまでできれば、来週の授業で躓くところは少ないでしょう。