

# R による探索的データ分析入門 - 可視化 演習

発電基盤開発課 高津一誠

2018 年 10 月 19 日

## 1 RStudio の使い方

R を使うときは開発環境ツールの RStudio を起動してください。RStudio はデータ分析作業をサポートする便利なツールです。

### 1.1 画面構成

RStudio の画面は、4 つのワクから構成されています。

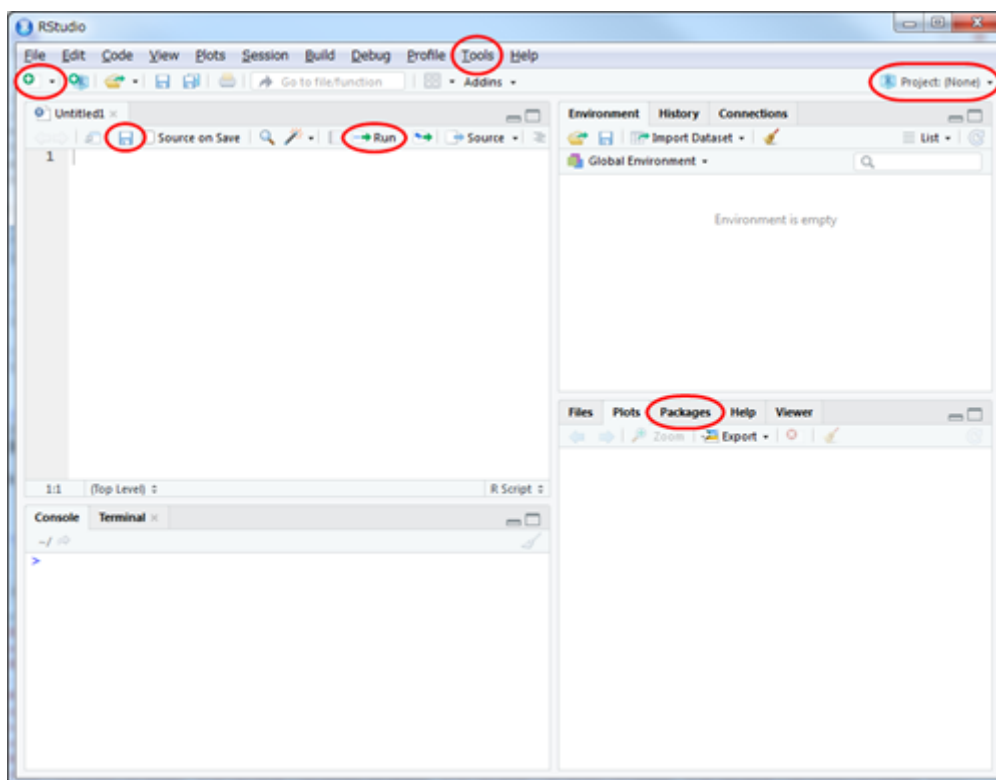


図1 RStudio の画面構成

印をつけてあるのが、ここで使用する項目ですが、参考までにそれぞれのワクについて説明します。

左上のワクは、後でスクリプトファイルを作成すると表示されます。ここに R の処理を書いていきます。

左下のワクはコンソールで、R のコマンドを実行するところです。右上のワクにはタブが並んでいます。[Environment] にはデータの一覧が表示されます。項目をクリックすると左上のワクにタブが新しくタブができて、データの内容が表示されます。[History] には入力したコマンドの履歴が表示されます。右下のワクには、プロジェクトフォルダに含まれるファイル进行操作する [Files]、グラフなどの出力を表示する [Plots] と [Viewer]、R の機能を拡張するパッケージを管理する [Packages]、英語ですがコマンドの使い方を表示する [Help] があります。

## 1.2 初期設定を行う

RStudio をインストールした後に、以下の設定を行ってください。

- 一時データを保存しないようにする。
  - メニューから [Tools -> Global Options] を選択する。
  - Restore .RData into workspace at startup: チェックを外す
  - Save workspace to .RData on exit: Never
- ライブラリをインストールする。
  - 右下のワクの [Packages] タブを選択する。
  - [Install] ボタンを押す。
  - [Packages] の欄に `tidyverse` と入力し [Install] を押す。

## 1.3 データ分析をはじめる

新しくデータ分析をはじめるときには、以下のようにしてください。

- プロジェクトを作成する。
  - あらかじめ、プロジェクトを保存するフォルダを用意する。
  - ツールバー右端の [Project] ボタンを押す。
  - [New Project] を選び、表示された画面で [Existing Directory] を選ぶ。
  - [Browse] ボタンを押して、用意したフォルダを選択し、[Open] ボタンを押す。
  - [Create Project] ボタンを押す。
- R の処理を記録するファイル（スクリプトファイル）を作成する。
  - ツールバー左端の [+] ボタンを押し、[R Script] を選ぶ。
  - 左上のワクに作られた [Untitled1] タブの下にある保存ボタン（フロッピーディスクのアイコン）を押す。
  - ファイル名を入力して [Save] ボタンを押す。（拡張子は自動で追加される。）
- ライブラリをインポートする。
  - スクリプトファイルの先頭に、以下を記述する。

```
library(tidyverse)
```

- カーソルが先頭行にある状態で、タブ内の右上の [Run] ボタンを押す。
- 処理を書く。スクリプトファイルに、実行させたいコマンドを記述する。

## 1.4 コマンドを実行する

スクリプトファイルに書いたコマンドを実行するには、実行したい行にカーソルがある状態で [Run] ボタンを押します。すると、コンソールにコマンドが転送されて実行されます。

複数行を実行したいときや、行内の一部分だけを実行したいときは、範囲選択した状態で [Run] ボタンを押します。スクリプトファイルが完成した後なら、ファイルの内容を一括して実行することも可能です。その場合は [Source] ボタンのドロップダウンメニューから [Source with Echo] を選んでください。

## 1.5 ショートカットキー

よく使う演算子や操作は、ショートカットキーで入力すると便利です。以下を使ってみてください。

表1: ショートカットキー

ショートカットキー	説明
Alt + -(ハイフン)	代入演算子 (<-)
Ctrl + Shift + m	パイプ演算子 (%>%)
Ctrl + Enter	実行
Ctrl + z	Undo
Ctrl + Shift + z	Redo

## 1.6 入力の補完機能を活用する

RStudio ではキーボードから文字を入力すると候補が表示されます。(表示されないときは Tab キーを押すと表示されます。) 矢印キー (↑↓) で候補を選択して Enter キーや Tab キーを押すと入力の手間を省くことができます。

## 2 可視化の演習

前回勉強した R による可視化を使って、データを分析してみましょう。

R に組み込まれている、**diamonds** データセットを使ってデータ分析をしてください。**diamonds** はダイヤモンドの品質と価格を格納したデータで、前回の **iris** と同じように **tidyverse** パッケージをロードすると使えるようになります。

```
library(tidyverse)
```

```
diamonds
```

```
## # A tibble: 53,940 x 10
##   carat cut      color clarity depth table price      x      y      z
##   <dbl> <ord>    <ord> <ord>    <dbl> <dbl> <int> <dbl> <dbl> <dbl>
## 1 0.23 Ideal     E     SI2     61.5   55   326  3.95  3.98  2.43
## 2 0.21 Premium  E     SI1     59.8   61   326  3.89  3.84  2.31
## 3 0.23 Good     E     VS1     56.9   65   327  4.05  4.07  2.31
## 4 0.290 Premium I     VS2     62.4   58   334  4.2   4.23  2.63
## 5 0.31 Good     J     SI2     63.3   58   335  4.34  4.35  2.75
## 6 0.24 Very Good J     VVS2    62.8   57   336  3.94  3.96  2.48
## 7 0.24 Very Good I     VVS1    62.3   57   336  3.95  3.98  2.47
## 8 0.26 Very Good H     SI1     61.9   55   337  4.07  4.11  2.53
## 9 0.22 Fair     E     VS2     65.1   61   337  3.87  3.78  2.49
## 10 0.23 Very Good H     VS1     59.4   61   338  4     4.05  2.39
## # ... with 53,930 more rows
```

変数の説明は以下のとおりです。

表2: diamond データセット

データ列 (変数)	説明
carat	重さ (カラット)
cut	カット等級
color	色
clarity	透明度
x	長さ (mm)
y	幅 (mm)
z	深さ (mm)
depth	深さ比 ( $z/(x+y)/2$ )
table	上面幅/最大幅

## 2.1 演習 1

ダイヤモンドのカラットと価格には相関があるでしょうか？グラフを描いて調べてください。

## 2.2 演習 2

カラットと価格の関係は、カット等級によって変化するでしょうか？カット等級ごとのサブグラフを描いて調べてください。