

R ではじめるデータ・サイエンス

氏名 (Name)

Last Updated: 2023-03-23

右上の Code マークから、ソースファイルもダウンロードすることができます。

Posit (RStudio) Cloud での実習

RNotebook ファイルを新規作成し、このページを、全選択して、コピー・ペーストします。(rmarkdown のインストールが必要です) Save As ... または、フロッピーディスクマークをおして、ds-edu などとして保存。説明を読みながら、コードチャンクと言われる、色が変わっている、の右上にある三角印を押すと、実行されます。

このあとで、RNotebook についても、少し説明します。

準備 : Posit Cloud ログイン

Posit Cloud (<https://posit.cloud/>) のアカウントを作成して、ログインした状態から始めます。

無論 RStudio で R を利用できる方は、ほとんど同じですが、Posit Cloud には、ファイルや、データも入っていますので、そのまま活用することが可能です。

RStudio での、動作不良の場合の対応は省略させていただきます。できれば、上のメニューの File から、New Project を選択して、Project を作って、そこで作業をされることをお勧めします。

問題がある時は、下のサイトを参照してください。Posit Cloud についても、多少の説明が書いてあります。

<https://icu-hsuzuki.github.io/ds4aj/ronrstudio.html>

右上の New Project を選択して、始めることが普通ですが、本日は、

このリンクにアクセスしていただき、Save a Permanent Copy を選択していただくと、そのなかで、自分のプロジェクトとして、実行することが可能になります。

<https://posit.cloud/content/5539763>

はじめに

基本コマンド - in Console

下のコマンドを左下の窓枠 (Pane) に入れてみてください。

- `head(cars)`
- `str(cars)`
- `summary(cars)`
- `df <- cars`

– `<-` と `_` と `%>%` と ``` back tick のキーボード上での位置の確認

- head(df)
- plot(df)
- View(cars) または、右上の Environment から、df_cars をクリック
- ?cars または Help 検索窓で cars, head など

実際の出力

```
head(cars)
```

```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10
```

```
str(cars)
```

```
## 'data.frame':   50 obs. of  2 variables:
##  $ speed: num  4 4 7 7 8 9 10 10 10 11 ...
##  $ dist : num  2 10 4 22 16 10 18 26 34 17 ...
```

```
summary(cars)
```

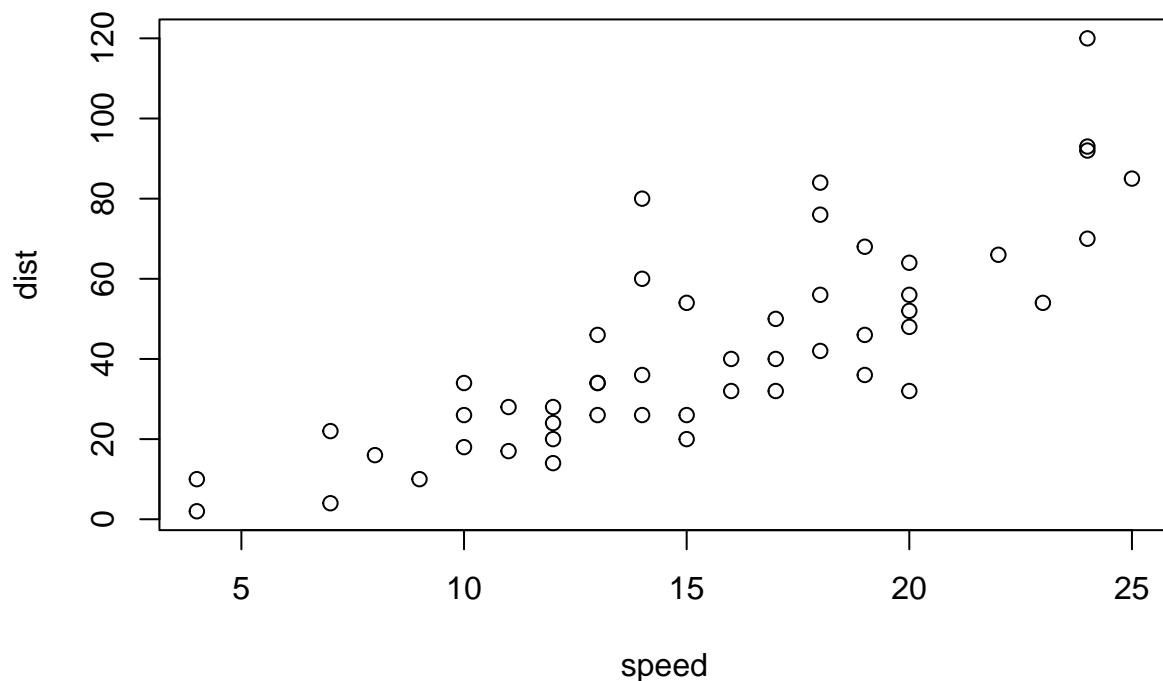
```
##      speed      dist
##  Min.   : 4.0   Min.   : 2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean    : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.    :120.00
```

```
df <- cars
```

```
head(df)
```

```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10
```

```
plot(df)
```



おすすめ: `Sys.setenv(LANG = "en")`

システム言語を英語に設定すると、不明のエラーメッセージがでたときにも、検査で解決できることが多い。

参考

データサイエンスをはじめましょう の中コンソールで実行 に、簡単な説明があります。

Package の確認とインストール

R パッケージは、R の拡張機能で、コード、データ、ドキュメントを標準化されたコレクション形式で含んでおり、標準的なものは、R Studio の Top Bar の Tool > Install Packages からインストールできます。

- 次の三つをインストールします: tidyverse, rmarkdown, WDI
- Tools から Install packages を選び、一つ一つ入れていくと、install できます。少し時間がかかります。

備考

R は、多くの方が、パッケージを開発して、それぞれの分野で、使いやすくしていった統計解析のためのコンピュータ言語です。ただ、基本をのぞいて、それぞれの好みにあわせて、開発された面もあり、統一性にはかける言語です。あとで説明する、探索的データ分析 (Exploratory Data Analysis) のために必要な基本的なものを、整合性のある形で、再構築したものが、tidyverse パッケージで、中心となった、H. Wickham は、RStudio (現在の Posit) に加わりました。そのメンバーの、Y. Xie (口益口) たちが開発したものが、rmarkdown に関連したパッケージです。

R Markdown 入門

R Notebook

R Markdown はデータサイエンスのためのオーサリングフレームワーク。

コード（プログラム）とその実行結果、を記録・表示し、高品質のレポートの作成を可能にします。

R Notebook は、独立してインタラクティブに実行できるチャンクを持つ R Markdown ドキュメントの一つの形式で、入力のすぐ下に出力が表示することができます。

1. File > New File > R Notebook
2. Save with a file name, say, test-notebook
3. Preview by [Preview] button
4. Run Code Chunk `plot(cars)` and then Preview again.

- テンプレートファイルの利用

下のリンクを開き、全部を選択し、R Notebook ファイルも全選択して貼り付ければ、日本版も見ることができます。

- 日本語ノートブック : <https://ds-sl.github.io/intro2r/RNotebook-J.html>
- いろいろな形式の出力 : <https://ds-sl.github.io/intro2r/Rmarkdown-J.html>

`intro2rj.Rmd`

右下の File タブから、`intro2rj.Rmd` を開きます。Preview の右のギアマークの Preview in Viewer Pane を選択してから、Preview ボタンを押すと、内容を確認することができます。

あとから使うので、ロードしておきます。最初に次のようなコードを実行します。右の三角を押します。

Posit Cloud のサイトには、data ディレクトリ（フォルダ）が作成されていますが、

参考

データサイエンスをはじめましょう 中の R Markdown に、“Reproducible and Literate Programming” (Reproducible（再現可能）かつ、Literate な（理解できるように記述した）Program（プログラム・コード）を共有すること）をたいせつにする RMarkdown の重要性について簡単に書いてあります。

R Notebook での実習

準備 Setup

データディレクトリを作成しておきます。今後、データは、このディレクトリに保存します。すでに、存在する場合は、その旨メッセージが出ますが、中身が消去されることはありません。

```
dir.create("./data")
```

パッケージはインストールしてあっても、利用するときには、次のように使える状態に（ロード）します。

一つ目は、tidyverse ですが、もう一つは、あとから活用する、世界開発指標に関するデータを利用するためのパッケージです。