

## R の使いかたとアソシエーション分析（第 5 章補足資料）

第 5 章のアソシエーション分析では、<sup>アール</sup>R を使って分析を行いました。R は統計解析向けのプログラミング言語であり、さまざまな統計解析手法を実行可能です。

R を使うときは、<sup>アールスタジオ</sup>RStudio という開発環境が非常に便利です。また、ちょっと試してみたい場合には、<sup>ポジットクラウド</sup>Posit Cloud（旧名称 RStudio Cloud）というブラウザ上で RStudio の環境を使えるサービスも便利です。ここでは、RStudio と Posit Cloud それぞれについて、インストールなどの準備の方法と使い方を説明していきます。

### R を使うための準備

#### 1. RStudio を使う場合：R と RStudio のインストール

まず、R 自体をインストールしましょう。R は、公式サイト（<https://cran.r-project.org/>）から自分の OS に合うものをダウンロードします。ダウンロードしたらファイルを実行するとインストーラーが起動するので、インストーラーの指示に従ってインストールを行います。

RStudio は、公式サイト（<https://posit.co/downloads/>）から「RStudio Desktop」をダウンロードします。「RStudio Desktop Pro」という有料版もありますが、通常は無料版で問題ありません。2024 年 9 月時点では、無料版は <https://posit.co/download/rstudio-desktop/> からダウンロードできます。なお、このページに移動すると、最初のステップとして「Install R」と書かれていますが、前述の方法で R のインストールが済んでいる場合は、2 ステップめの「Install RStudio」を行ってください。インストールは、R のときと同様にインストーラーの指示に従えば問題なく実行できます。

インストールが終わったら、RStudio を起動します。図 1 のような画面が出れば OK です。

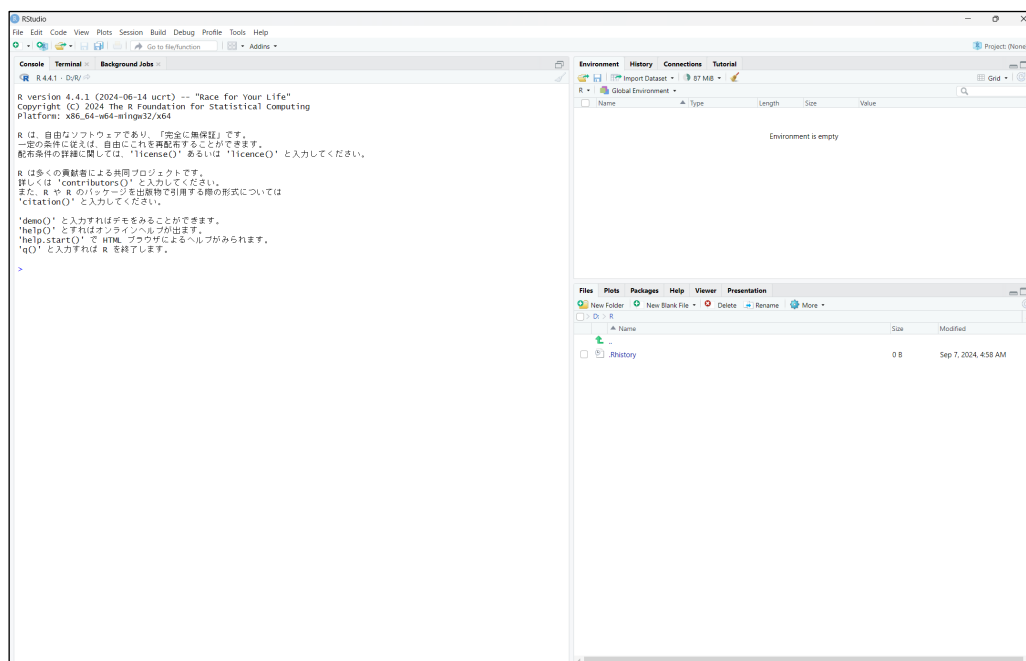


図 1 RStudio を起動した画面

## 2. Posit Cloud を使う場合：アカウント登録からプロジェクト作成

続いて、ブラウザ上で RStudio を使えるサービスである Posit Cloud を見てみましょう。まず、Posit Cloud の公式サイト (<https://posit.cloud/>) にアクセスし、右上の「Sign Up」を押します。すると、Posit Cloud の料金プラン画面が出てきます。ここでは簡単に試すだけなので、「Free」を選んでください。なお、2024 年 9 月時点では、Free プランの場合「1 か月あたり 25 時間まで」「作成できるプロジェクト（プロジェクトに関しては後述）は 25 個まで」などの制約があり、使用できるメモリや CPU にも制限がありますが、今回は十分です。

「Learn more」を押すと次の画面で Sign Up を促されるので、アカウント登録をします。Google のアカウントをもっている場合は、「Sign up with Google」を選んでログインすると便利です。もちろん、それ以外の方法でアカウントを作っても問題ありません。

アカウント登録をしてログインをすると、図 2 のような画面が出てくるので、「Posit Cloud」を選択します。

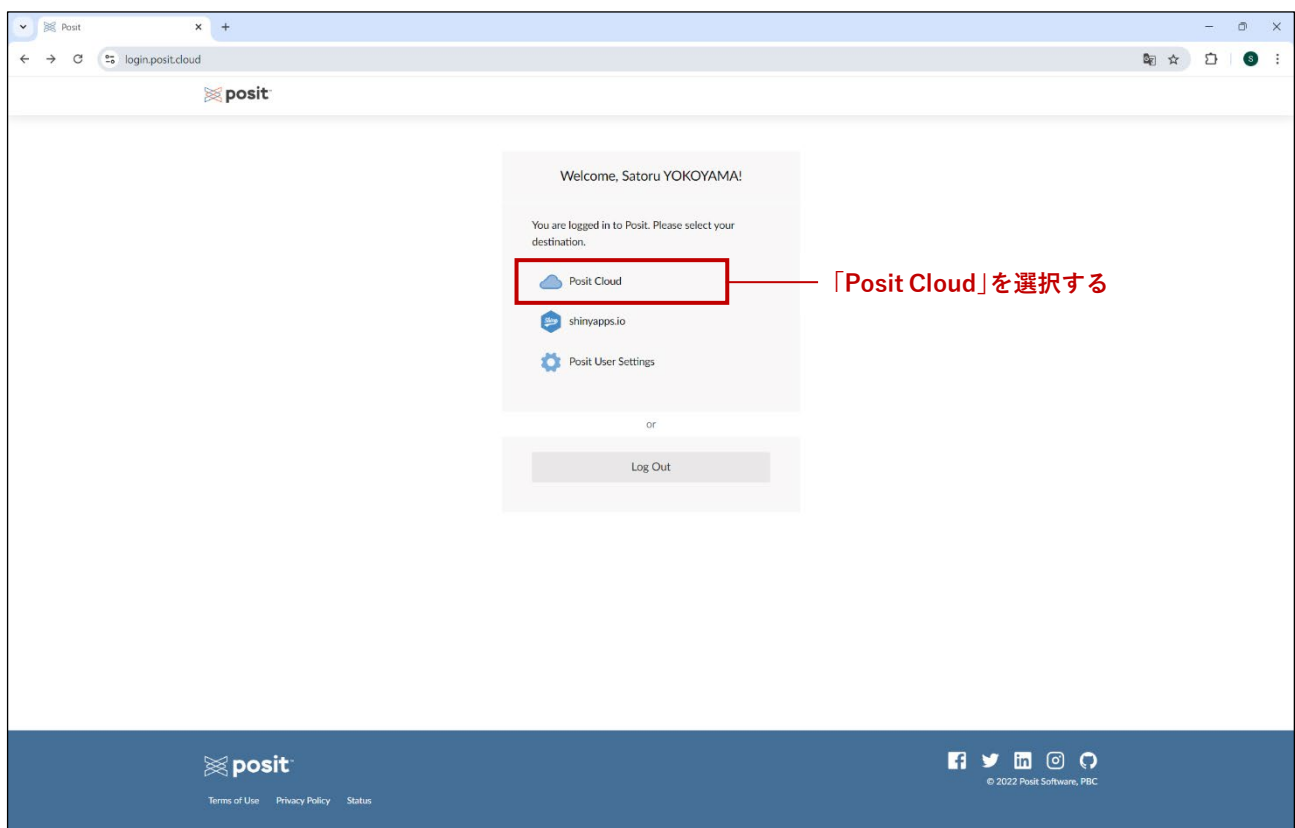


図 2 Posit Cloud での分析ツールの選択画面

すると図 3 のような画面が出てくるので、右上の「New Project」をクリックして表示されたメニューから「New Project」を選択し、新しいプロジェクトを作成します（図 4）。

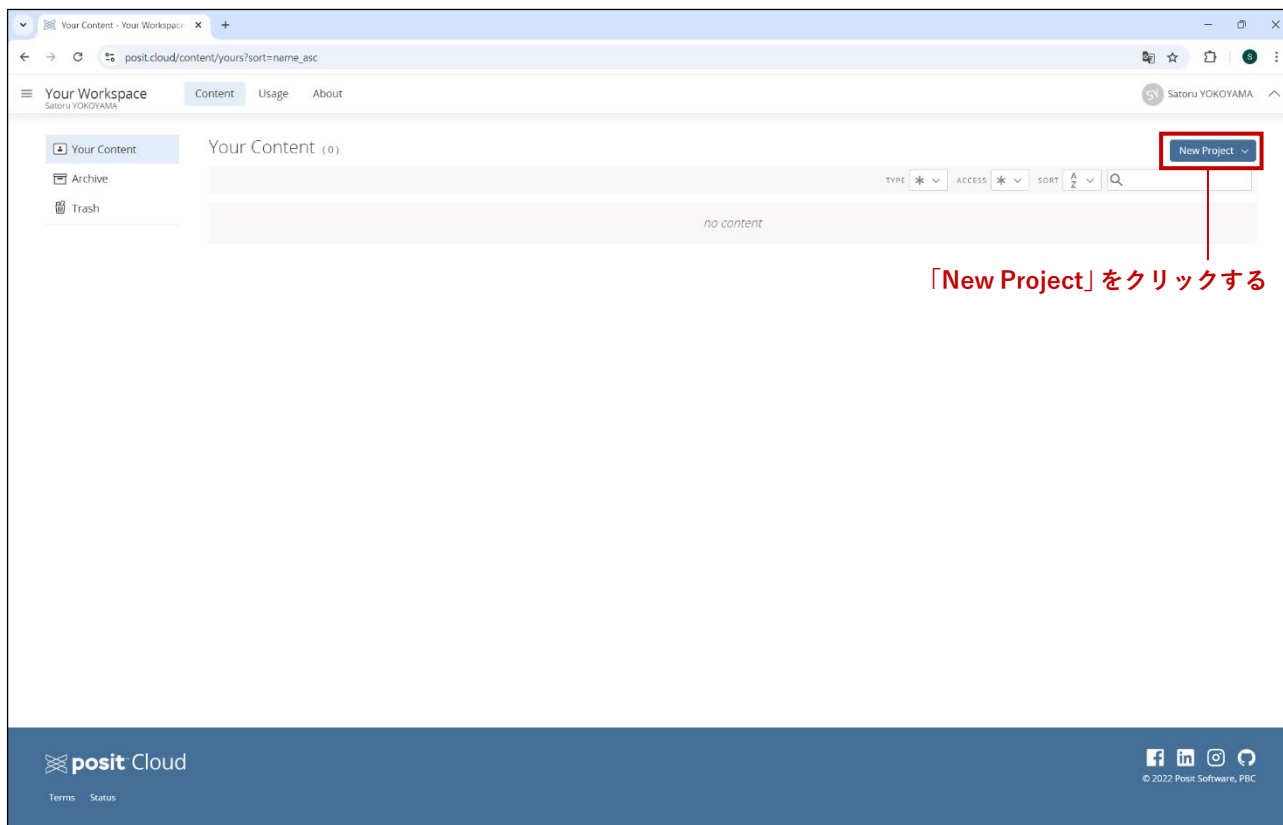


図3 Posit Cloud のプロジェクト作成画面

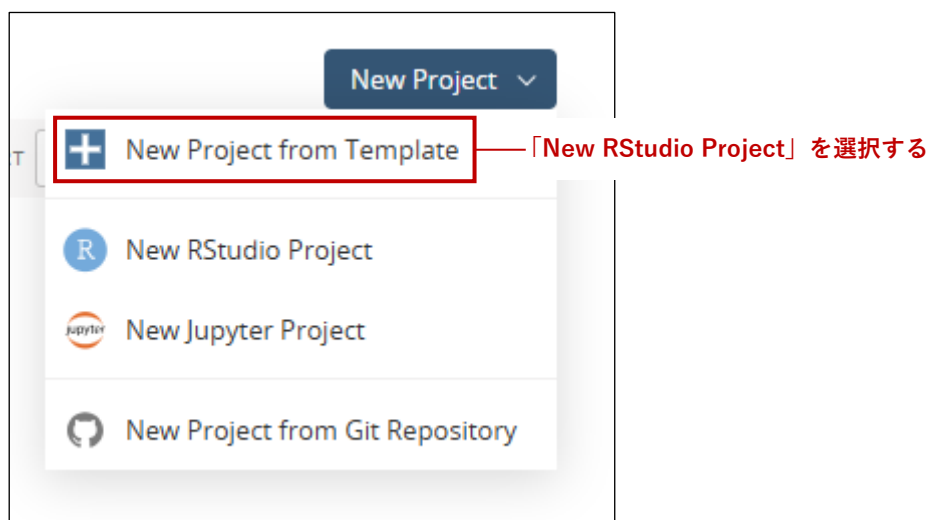


図4 プロジェクトの新規作成

しばらく待って、図5のような画面になればOKです。左上に表示されているプロジェクト名が「Untitled Project」となっているので、適当に名前を変えておくといいでしょう。日本語は入らないので、今回は「Association Analytics」としておきます。

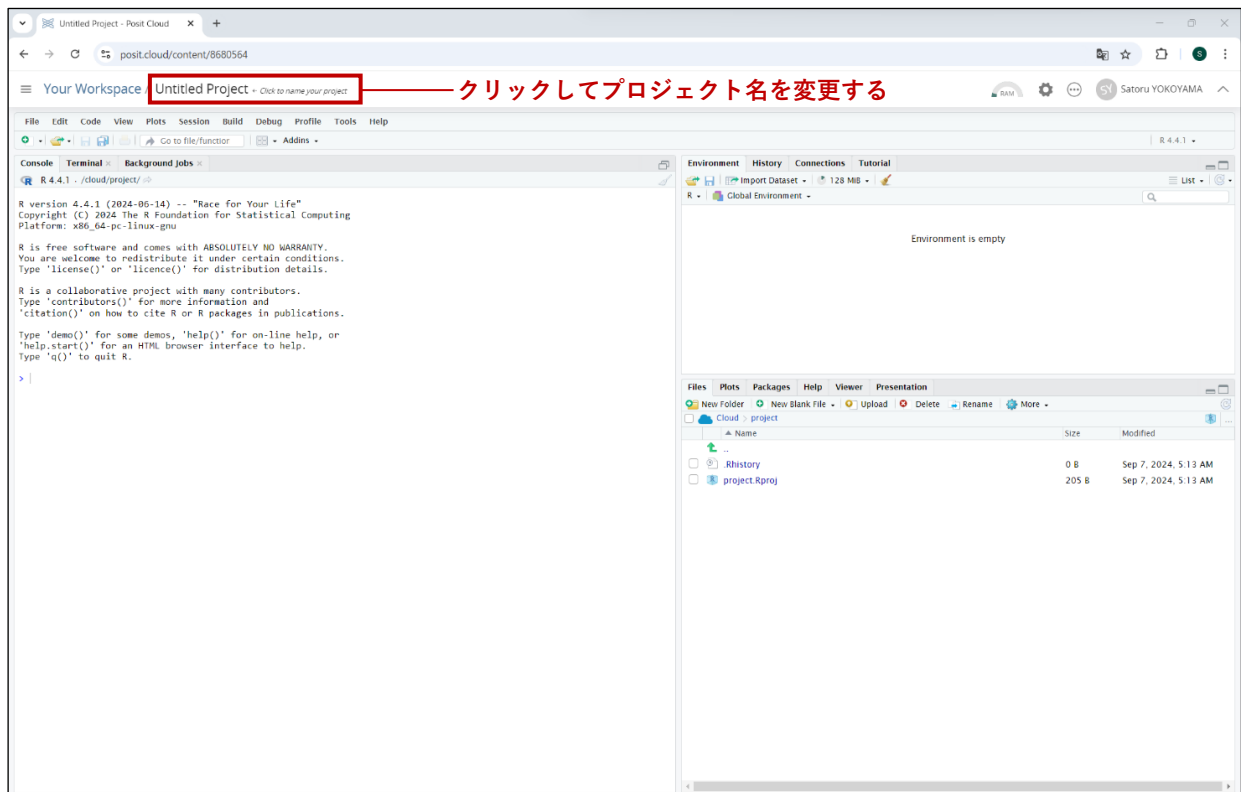


図5 Posit Cloud の起動画面

## RStudio の「プロジェクト」とはなにか

ここで、RStudio のプロジェクトについて簡単に説明します。RStudio では、作業をプロジェクトという単位で扱います。作業（たとえば、筆者の場合は「研究用」「授業用」「この本の執筆用」など）ごとにプロジェクトを作成すると、プロジェクトを切り替えて使うことができます。これによって、複数の作業でのデータや結果をそれぞれ別々に管理することができます。

Posit Cloud の場合は、最初にプロジェクトを作成してから RStudio の画面に進みました。R と RStudio をパソコンにインストールした場合は、RStudio を起動すると右上にある「Project: (None)」を押して「New Project」を選択すると、図 6 のような画面が出てきます。「New Directory」を選択すると、新しいプロジェクトを作成できます。

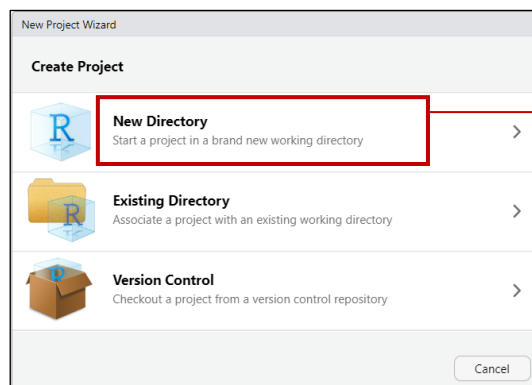


図6 RStudio での Create Project の選択画面

## R によるアソシエーション分析の実行

最後に、実際にアソシエーション分析を行ってみます。ここでは、Posit Cloud で実行してみます。基本的には、パソコンにインストールした RStudio でも手順は同様です。

### 1. 分析するデータの準備

まず、分析用のデータを使えるようにしましょう。Posit Cloud の場合、右下のエリアの「Upload」から分析するデータ（併売データ.xlsx）をアップロードします。パソコンにインストールした RStudio の場合は、作成したプロジェクトの作業フォルダーにデータをコピーまたは移動しておきます。

### 2. 分析するデータの読み込み

準備したデータを読み込みます。R などのプログラミング言語は、アルファベットで**コマンド**を打ち込み実行することで、計算結果などの出力を得ることができます。ここでは、「データを読み込む」というコマンドを入力してみましょう。分析する「併売データ.xlsx」は Excel ファイルで、1 枚目のシートにデータがあります。まずは、左側の画面に次のコマンドを入力し、Enter キーを押して実行します。

```
install.packages("openxlsx")  
library(openxlsx)
```

これは、R で Excel ファイルを読み込むための機能をインストールして（1 行目）、その機能呼び出す（2 行目）というコマンドです。これでデータの読み込みができるようになりました。

続いて、以下のコマンドでデータを読み込みます。

```
data = read.xlsx("併売データ.xlsx")
```

ここで以下のコマンドを入力して実行すると、図 7 のように左上にデータが表示されます。

```
View(data)
```

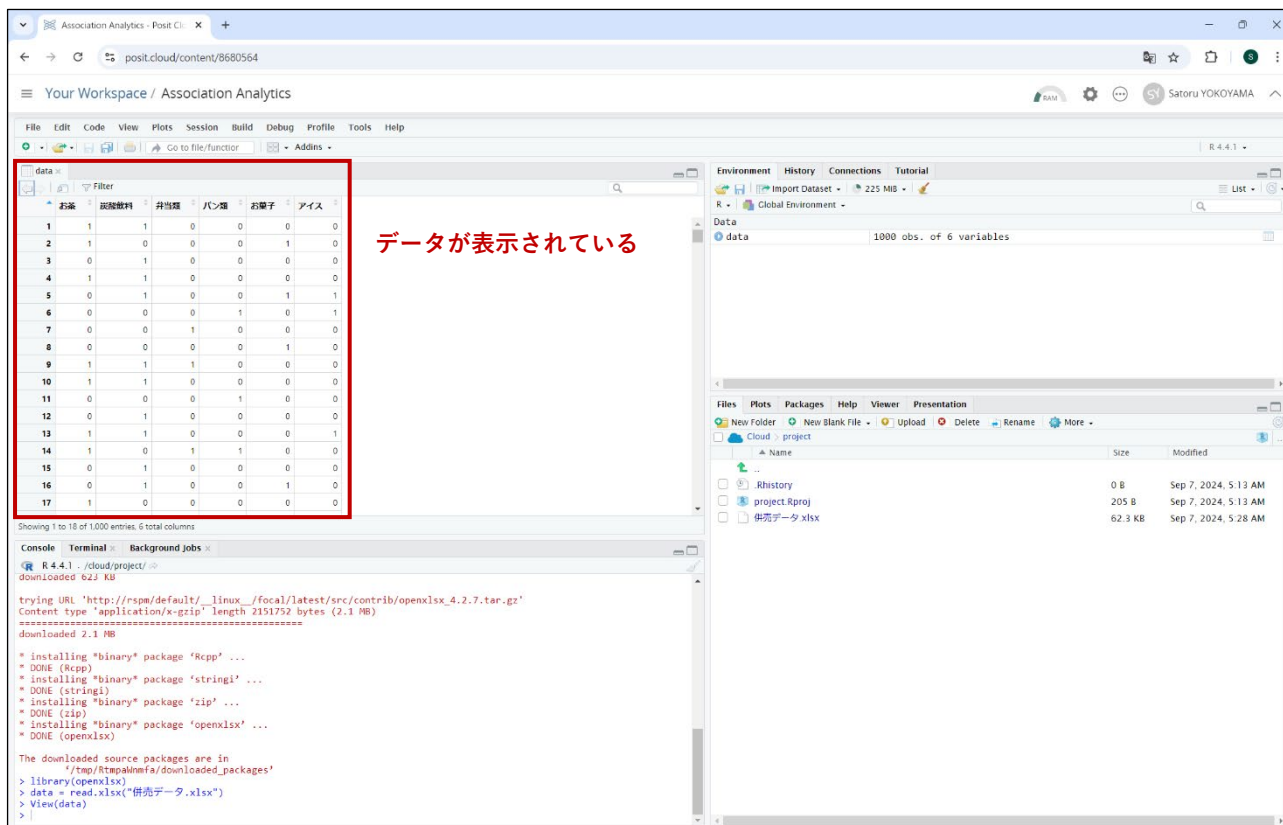


図7 RStudioでのデータの表示

### 3. アソシエーション分析の実行

最後に、アソシエーション分析を行きましょう。以下のコマンドを入力して、アソシエーション分析を実行するために必要な機能をインストールして、その機能呼び出します。

```
install.packages("arules")
```

```
library(arules)
```

また、データは読み込んだままの形式だとアソシエーション分析では分析できないので、以下のようデータを変換します。

```
data.tran = as(as.matrix(data), "transactions")
```

準備ができたので、実際にコマンドを入力してアソシエーション分析を実行します。

```
data.ap = apriori(data.tran)
```

上のコマンドの場合、支持度 0.1、確信度 0.8 として分析を行った結果が得られます。本文中では条件を緩めて、支持度 0.05、確信度 0.4 と指定して分析を行いました。そういったさまざまな要素（**パラメーター**といいます）の指定は、以下のようになります。

```
data.ap = apriori(data.tran, parameter=list(support=0.05, confidence=0.4))
```

分析で得られたルールを表示させるときには、以下のコマンドを実行します。

```
inspect(data.ap)
```

ただし、このままだとリフト値が順に並んでおらず見づらいので、以下のコマンドを使ってリフト値を降順に並べ替えて出力しましょう。

```
inspect(sort(data.ap, by="lift"))
```

これにより、本文 172 ページのような結果が得られます。

同様の手順で、必要に応じて支持度や確信度の値を変更して分析してみてください。