

【データサイエンス基礎講座】:2014年12月度

データサイエンス基礎講座(超初級・実践編)
2014年11月26日～12月17日<全5回>用演習資料

R入門

インストールから活用まで

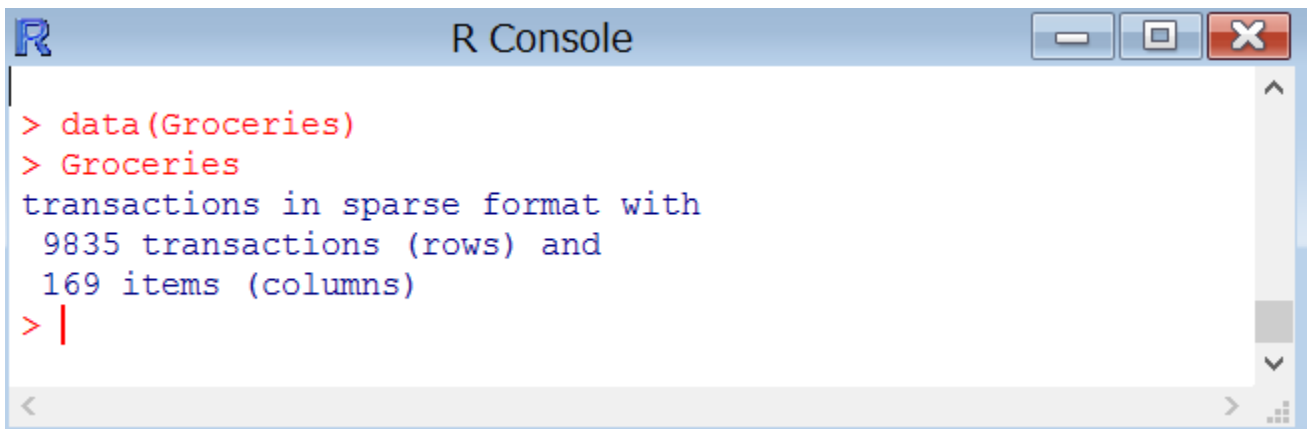
- ・主催:株式会社インプレス
- ・企画/製作:フューチャーブリッジパートナーズ株式会社



3 限目

アソシエーション分析

- アソシエーション(関連を分析)
- 標準ではサポートされていないので、パッケージarulesをインストール
- Rでのコマンド
 - install.packages (“arules”)
 - library (arules)
- うまくインストールできない場合
 - 書き込み設定がされていない可能性
 - ⇒Rのインストールディレクトリ(一般的には、“C:¥Program Files¥R¥R-3.1.1¥library”)の書き込み設定を変更、もしくは、管理者としてRを起動
- インストールに成功すると
 - data (Groceries)
 - Groceries



```
R Console
> data(Groceries)
> Groceries
transactions in sparse format with
 9835 transactions (rows) and
 169 items (columns)
> |
```

ある食料雑貨店で収集した30日分のPOSデータ
牛乳などの169品目、9835件の購買履歴

アソシエーション分析

- summaryで確認

```
R Console
> summary(Groceries)
transactions as itemMatrix in sparse format with
9835 rows (elements/itemsets/transactions) and
169 columns (items) and a density of 0.02609146

most frequent items:
  whole milk other vegetables    rolls/buns          soda          yogurt          (Other)
      2513          1903          1809          1715          1372          34055

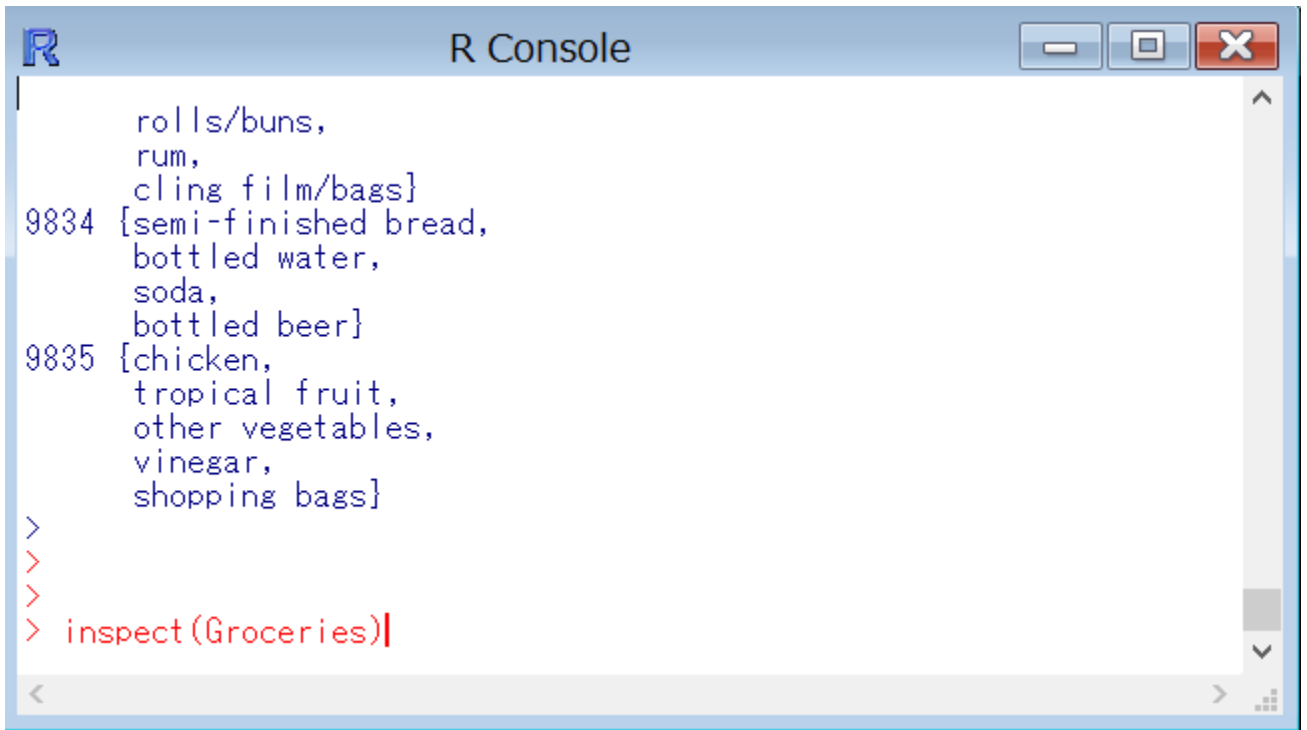
element (itemset/transaction) length distribution:
sizes
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26
2159 1643 1299 1005 855 645 545 438 350 246 182 117 78 77 55 46 29 14 14 9 11 4 6 1 1
 27 28 29 32
 1  1  3  1

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  2.000   3.000  4.409  6.000 32.000

includes extended item information - examples:
  labels level2 level1
1 frankfurter sausage meet and sausage
2  sausage sausage meet and sausage
3  liver loaf sausage meet and sausage
> |
```

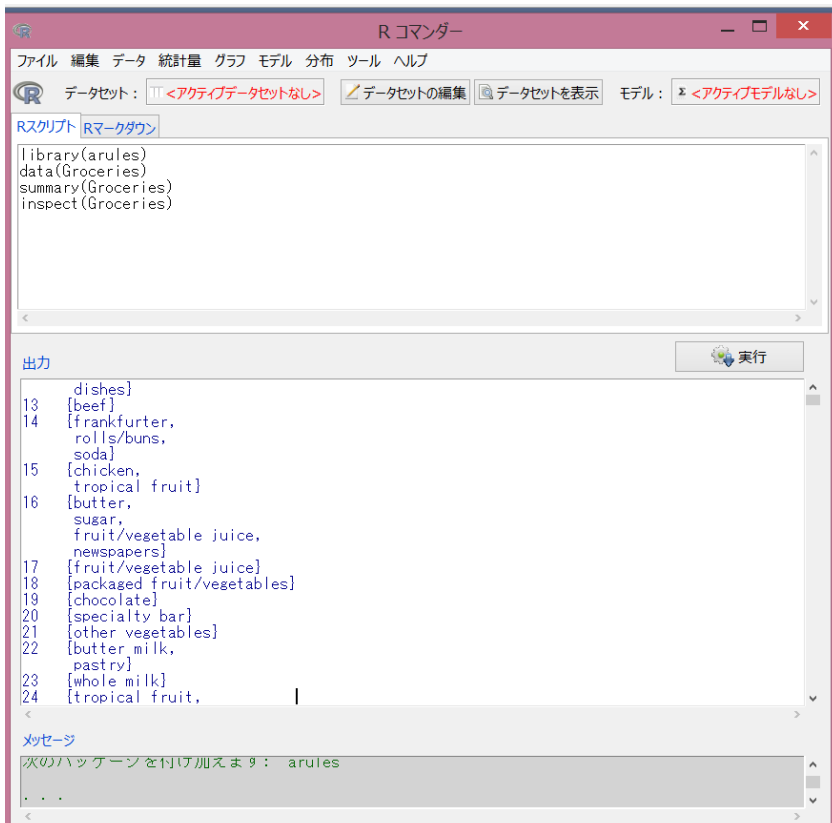
アソシエーション分析

- 組み合わせを表示
 - inspect(Groceries)



```
R Console  
  
    rolls/buns,  
    rum,  
    cling film/bags}  
9834 {semi-finished bread,  
    bottled water,  
    soda,  
    bottled beer}  
9835 {chicken,  
    tropical fruit,  
    other vegetables,  
    vinegar,  
    shopping bags}  
>  
>  
>  
> inspect(Groceries)|
```

すべての出力結果を確認したい場合は、Rコマンドーを利用



```
R コマンドー  
ファイル 編集 データ 統計量 グラフ モデル 分布 ツール ヘルプ  
データセット: <<アクティブデータセットなし>> データセットの編集 データセットを表示 モデル: <<アクティブモデルなし>>  
Rスクリプト Rマークダウン  
library(arules)  
data(Groceries)  
summary(Groceries)  
inspect(Groceries)  
  
出力  
13 dishes  
14 {beef}  
15 {frankfurter,  
    rolls/buns,  
    soda}  
16 {chicken,  
    tropical fruit}  
17 {butter,  
    sugar,  
    fruit/vegetable juice,  
    newspapers}  
18 {fruit/vegetable juice}  
19 {packaged fruit/vegetables}  
20 {chocolate}  
21 {specialty bar}  
22 {other vegetables}  
23 {butter milk,  
    pastry}  
24 {whole milk,  
    tropical fruit,  
    |  
メッセージ  
次のハッシュマークを付け加えよう: arules
```


アソシエーション分析

- インタラクティブグラフでビジュアル化
 - `install.packages("arulesViz")`
 - `library(arulesViz)`
 - `gruleX = apriori(Groceries, p=list(support=0.03, confidence=0.05, ext=TRUE))`
 - `gruleX2 = subset(gruleX, subset=(lift>=1.5))`
 - `plot(gruleX2, method="graph", interactive=TRUE)`

