

1.3 Accessing information in data frames の後半

値の特定によるデータの引き出し

行の番号がわからなくても、列の値を特定することによりそれにマッチする行のみを取り出すことができる。

例) AnimacyOfTheme が animate の行のみを取り出す

行に対する制約なので、コンマの前に条件を書く。

```
verbs [verbs$AnimacyOfTheme == "animate",]
```

以下と同じこと。

```
subset (verbs, AnimacyOfTheme == "animate")
```

=ではなく==を用いる。前者だと値の代入となってしまう。

上記を応用すると以下のようなことが可能となる(&はその前と後の両方の条件を満たす必要があり|はどちらかを満たしていればよい)

```
verbs [verbs$AnimacyOfTheme == "animate" & verbs$LengthOfTheme > 2,]
```

これは AnimacyOfTheme の値が animate で、LengthOfTheme の値が 2 を超えている行を取り出すコマンド。

行名と列名

rownames で行名を、colnames で列名を取り出すことができる

- `head (rownames (verbs))`
- `colnames (verbs)`

文字列ベクトル

列名 (RealizationOfRec、Verb など) は文字列ベクトルであり、行名 (1、2、3 など) も同様に文字列ベクトルである。後者は値が数字であるため文字列ベクトルであることは意外だが、有益なことも多い。以下に例を示す。

verbs.rs というデータフレームを verbs の一部を用いて新たに作成する

```
verbs.rs = verbs [rs,]
```

以下のコマンドは元の verbs 内の 638 番目の行を返す。

```
verbs.rs [1,]
```

出力結果

```
RealizationOfRec Verb AnimacyOfRec AnimacyOfTheme LengthOfTheme  
638             PP pay      animate      inanimate      0.6931472
```

これは 638 を文字列ベクトルとして扱うことによっても同一の出力が得られる。

```
verbs.rs ["638",]
```

以下も同様の結果を生む

```
verbs [638,]
```

文字列ベクトルと因子

`verbs.rs$AnimacyOfRec`

は以下の結果を出力する

```
[1] animate  animate  animate  animate  inanimate
```

```
Levels: animate inanimate
```

まず、出力の最初の行で、`animate` や `inanimate` はクオテーションマークで括られていない。これは、行名や列名とは違い、数値ではない列は R によって自動的に「factor」に変換されるからである。factor の値は level と呼ばれる。ここでは `AnimacyOfRec` が取り得る値が `animate` と `inanimate` しかないため、その二つが出力の二行目に表示されている。

文字列ベクトルに対しては統計処理を行うことができないが、factor に対してはできる。これが R が自動的に factor に変換する理由である。

文字列ベクトルとして扱いたいのであれば、`as.character()`関数を用いれば良い。

```
verbs.rs$AnimacyOfRec = as.character(verbs.rs$AnimacyOfRec)
```

`verbs.rs$AnimacyOfRec` を出力すると、クオテーションマークが付いているのが確認できる。因子に戻すには、`as.factor()`関数を用いる。

```
verbs.rs$AnimacyOfRec = as.factor(verbs.rs$AnimacyOfRec)
```

これまでのデータフレームを使って、より小さなデータフレームを作る

```
verbs.rs2 = verbs[c(638, 390),]
```

`verbs.rs2` の出力は以下ようになる。

	RealizationOfRec	Verb	AnimacyOfRec	AnimacyOfTheme	LengthOfTheme
638	PP	pay	animate	inanimate	0.6931472
390	NP	lend	animate	animate	0.6931472

データフレームの中に `AnimacyOfRec` の値は `animate` しかないにも関わらず、

```
verbs.rs2$AnimacyOfRec を実行すると
```

```
[1] animate animate
```

```
Levels: animate inanimate
```

と二種類の値があるように返してしまう。これは元の値を R が覚えているためで、避けるには一度文字列ベクトルに直してから factor に再変換する

```
as.factor(as.character(verbs.rs2$AnimacyOfRec))
```

か、あるいは以下のコマンドを使う

```
verbs.rs2$AnimacyOfRec [drop = TRUE]
```