

stringr

- `stringi` のラッパー
- `tidyverse` に含まれている

正規表現のオプション

- `ignore_case=T`
- `multiline=T`

`str_c()` 文字列の結合

- 例： `ID = str_c(Lang, Year, PID, SID, sep="_")`
 - `Lang, Year, PID, SID` をアンダースコアでつないで、新しく `ID` という文字列にする

`str_which()` 文字列がある行番号を調べる

```
str_which( カラム名, "文字列")
```

`str_detect()` 該当する文字列があるかどうか調べる

```
str_detect( データ, "正規表現")
```

- `subset()` と合わせて使うと便利
 - データフレーム中の特定の列に「ある種の文字列」があるかどうかを調べて、その文字列を含む行だけを選び出す。
 - 「ある種の文字列」の例：小文字の連続で書かれている「単語」が複数あるもの

`str_extract()` 指定したパターンが該当した文字列を抽出する

- 正規表現で複数のパターンの文字列が該当する場合、個々に該当したパターンを出力

`str_replace()` (文字列, 置換対象表現, 置換後表現)

`str_remove()` (文字列, 削除表現)

- 上の置換で、何もなしで置換と同等

`str_remove_all()` (文字列, 削除表現)

- 文字列中出现しているすべての該当パターンを削除
 - `_all` なしだと、最初のものだけしか処理しない

`str_count()` (文字列, `"\w+"`) 単語数のカウント

- 一文字以上の英字の連続 (`"\w+"`) の数を数える

<https://www.statology.org/word-count-in-r/>

参考サイト

<https://heavywatal.github.io/rstats/stringr.html>