

Data frames - leitura e manipulação.

André Moura Gomes da Costa - andmgdc@gmail.com
Maria das Gracas Braga Ceccato

04/09/2019

Manipulação de *characters*

Manipulação de *characters*

Podemos fazer manipulação de *characters* com alguns pactoes diferentes.

Aqui será apresentado o pacote **stringr**, que apresenta funções equivalentes às do pacote **base**, mas com nomes mais fáceis.

O *cheatsheet* relativo a este pacote é bastante explicativo.

Para a manipulação apropriada de *characters* devemos aprender, também, expressões regulares, *regex*.

Algumas possibilidades

Identificar um padrão no início ou no fim de um *character*.

Identificar padrões com repetição de um dígito.

Identificar um padrão de uma URL válida.

Identificar um número dentro de um texto.

Caracteres especiais úteis:

- `^` - Início do character
- `$` - Fim do character
- `[]` - Classe de character
 - `[a-z]`; `[A-Z]`; `[0-9]` - Caracteres em um intervalo
 - `[^]` - Negação
- `|` - Ou lógico
- `()` - Restrição de escopo
- `?! - Ignorar diferença entre maiúscula e minúscula`
- `?` - Existência ou não de dado dígito (ou expressão)
- `.` - Qualquer character
- `*` - Ocorrência de 0 ou mais vezes do último dígito
- `+` - Ocorrência de 1 ou mais vezes do último dígito
- `{ }` - Ocorrência de um dado número de vezes de um padrão

Usando na prática

Vamos ver o uso destas funcionalidades em pequenos exemplos.

Vamos abrir o arquivo `Aula3_stringr.R` e explorar algumas funções do *stringr*.

Há situações em que deve-se utilizar as expressões regulares de maneiras criativas. É muito útil buscar ajuda na internet, pois, provavelmente, você não é o primeiro a ter esse tipo de problema.

Lidando com dados temporais no R

lubridate

O R possui classes especiais para se lidar com tempo.

O pacote **lubridate** permite uma fácil manipulação desses dados temporais.

Algumas funções úteis:

- `ymd`, `mdy`, `dmy`
- `ymd_hms`, `mdy_hs`, `dmy_h`
- `year`, `month`, `mday`, `yday`, `hour`, `minute`, `second`
- `with_tz`, `force_tz`

Sumarizando dados

Sumarizando dados

Vamos abrir o arquivo `Aula3_Descritivas.R` e explorar algumas funções:

- `mean()`, `sd()`, `var()`, `min()`, `max()`, `median()`, `quantile()`, `sum()`
- `summary()`
- `str()`

Gráficos

Algumas funções do R para visualizar dados graficamente são:

- `plot()`
- `hist()`
- `boxplot()`
- `stem()`

Na próxima aula exploraremos o uso do pacote `ggplot2` para fazer gráficos muito interessantes e de visual agradável.

Gráficos de pares

Há diversas funções de pacotes que permitem a visualização de gráficos de pares facilmente.

- `pairs()` do **graphics** (gráfico padrão do R)
- `ggpairs()` do **GGally**, baseado no `ggplot`
- `pairs.panel()` do pacote **psych**

Exercícios

Obrigado!