



# Работа с текстом



## Package stringr

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

<https://stringr.tidyverse.org/>

<https://github.com/rstudio/cheatsheets/blob/main/strings.pdf>



Длина

`str_length("abc")`





# Объединение

`str_c("x", "y")`

`str_c("x", "y", sep = ", ")`

`str_c("prefix-", c("a", "b", "c"), "-suffix")`

`str_c(c("x", "y", "z"), collapse = ", ")`



## Выбор части

```
x <- c("Apple", "Banana", "Pear")
```

```
str_sub(x, 1, 3)
```

```
str_sub(x, -3, -1)
```



# Регистр

`str_to_upper(c("i", "l"))`

`str_to_lower("L")`



# Регулярные выражения

```
x <- c("apple", "banana", "pear")
```

```
str_view(x, "an")
```

• – ЛЮБОЙ СИМВОЛ

```
str_view(x, ".a.")
```

## Специальные символы

```
str_view(c("abc", "a.c", "bef"), "a.c")
```

```
str_view(c("abc", "a.c", "bef"), "a\\.c")
```

```
\\. \\. \\. \\. \\. *
```





# Якорь

`x <- c("apple", "banana", "pear")`

`str_view(x, "^a")` - **НАЧАЛО**

`str_view(x, "a$")` - **КОНЕЦ**



# Классы

- \d: любые числа
- \s: любые пробелы – пробел, табуляция, новая строка
- [abc]: a, b или c
- [^abc]: любой символ, кроме a, b или c



# Повторения

- $?$ : 0 или 1
- $+$ : 1 или больше
- $*$ : 0 или больше
- $\{n\}$ : точно  $n$
- $\{n, \}$ :  $n$  или больше
- $\{, m\}$ : как минимум  $m$
- $\{n, m\}$ : между  $n$  и  $m$



# Нахождение паттернов

```
x <- c("apple", "banana", "pear")  
str_detect(x, "e")
```

```
sum(str_detect(words, "^t"))  
words[str_detect(words, "x$")]
```



# Разделение

`sentences %>%`

`head(5) %>%`

`str_split(" ")`