Getting / cleaning data 2

Working with strings

String operations

For these examples, we'll use some data on passengers of the Titanic. You can load this data using:

```
# install.packages("titanic")
library("titanic")
data("titanic_train")
```

We will be using the `stringr` package:

```
library("stringr")
```

String operations

This data includes a column called "Name" with passenger names. This column is somewhat messy and includes several elements that we might want to separate (last name, first name, title).

Here are the first few values of "Name":

```
titanic_train %>% select(Name) %>% slice(1:3)
```

```
##                                                 Name
## 1                             Braund, Mr. Owen Harris
## 2 Cumings, Mrs. John Bradley (Florence Briggs Thayer)
## 3                              Heikkinen, Miss. Laina
```

String operations

The `str_trim` function from the `stringr` package allows you to trim leading and trailing whitespace:

```
with_spaces <- c("    a ", "  bob", " gamma")
with_spaces
```

```
## [1] "    a " "  bob"  " gamma"
```

```
str_trim(with_spaces)
```

```
## [1] "a"     "bob"   "gamma"
```

This is rarer, but if you ever want to, you can add leading and / or trailing whitespace to elements of a character vector with `str_pad` from the `stringr` package.

String operations

There are also functions to change a full character string to uppercase, lowercase, or title case:

```
str_to_upper("Braund, Mr. Owen Harris")
```

```
## [1] "BRAUND, MR. OWEN HARRIS"
```

```
str_to_lower("Braund, Mr. Owen Harris")
```

```
## [1] "braund, mr. owen harris"
```

```
str_to_title("braund, mr. owen harris")
```

```
## [1] "Braund, Mr. Owen Harris"
```