Titanic exercise

Van Gestel Axel, Jetze Luyten, David Silva Troya

Install required packages

```r
# install.packages("readr")
# install.packages("dplyr")
# install.packages("ggplot2")
```

Load required libraries

```r
library(readr)
library(dplyr)
library(ggplot2)
```

Read the data set (uses readr)

```r
column_types <- cols(
  Survived = col_factor(),
  Pclass = col_factor(ordered = TRUE),
  Sex = col_factor(),
  Embarked = col_factor()
)
train <- read_csv("./kaggle/titanic/train.csv", col_types = column_types)
```

Rename the factors to be human readable (uses dplyr)

```r
train$Survived <- recode_factor(train$Survived,
                                "0" = "No",
                                "1" = "Yes")

train$Pclass <- recode_factor(train$Pclass,
                              "1" = "1st",
                              "2" = "2nd",
                              "3" = "3rd")

train$Embarked <- recode_factor(train$Embarked,
                                "C" = "Cherbourg",
                                "S" = "Southampton",
                                "Q" = "Queenstown")
```

View train tibble

```r
train
```

```
## # A tibble: 891 x 12
```

```
##      PassengerId Survived Pclass Name   Sex     Age SibSp Parch Ticket   Fare Cabin
##            <dbl> <fct>    <fct>  <chr>  <fct> <dbl> <dbl> <dbl> <chr>   <dbl> <chr>
## 1             1 No       3rd    Braun~ male     22     1     0 A/5 2~   7.25 <NA>
## 2             2 Yes      1st    Cumin~ fema~    38     1     0 PC 17~  71.3  C85
## 3             3 Yes      3rd    Heikk~ fema~    26     0     0 STON/~   7.92 <NA>
## 4             4 Yes      1st    Futre~ fema~    35     1     0 113803 53.1  C123
## 5             5 No       3rd    Allen~ male     35     0     0 373450  8.05 <NA>
## 6             6 No       3rd    Moran~ male     NA     0     0 330877  8.46 <NA>
## 7             7 No       1st    McCar~ male     54     0     0 17463  51.9  E46
## 8             8 No       3rd    Palss~ male      2     3     1 349909 21.1  <NA>
## 9             9 Yes      3rd    Johns~ fema~    27     0     2 347742 11.1  <NA>
## 10           10 Yes      2nd    Nasse~ fema~    14     1     0 237736 30.1  <NA>
## # ... with 881 more rows, and 1 more variable: Embarked <fct>
```

Plots and stuff (uses ggplot2)

```
ggplot(data = train) + geom_point(mapping = aes(
  x = Age,
  y = Fare,
  shape = Sex,
  color = Survived
))
```