lesson02d_3458_561973_titanic.R

mgastner

2022-01-21

```
# MTG: When you use `read_csv()` instead of `read.csv()` you must load the
# **readr** package. I must add the next line to run your script.
library(readr)

titanic <- read_csv("titanic.csv")
```

```
## Rows: 2208 Columns: 11
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): fam_name, given_name, gender, class, ticket
## dbl (5): age, pax_on_tckt, pnd, shl, pnc
## lgl (1): survived
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#A comma-separated values file is a delimited text file that uses a comma to separate values.
#Each line of the file is a data record. Each record consists of one or more fields,
#separated by commas. The use of the comma as a field separator is the source of the name for this
#file format.


#dimensions of spreadsheet

nrow(titanic)
```

```
## [1] 2208
```

```
#there are 2208 rows

ncol(titanic)
```

```
## [1] 11
```

```
#there are 11 columns

#unique values in the column class:
unique(titanic$class)
```

```
## [1] "3rd"  "Crew" "2nd"  "1st"
```

```
#there are 4 unique values. 3rd, crew, 2nd, 1st

#how many people are in each class
table(titanic$class)
```

```
##
##  1st  2nd  3rd Crew
##  286  271  709  942
```

```
#there are 286 people in the 1st class
#271 people in 2nd class
#709 people in 3rd class
#942 people among the crew

#create a subset called sec_class that only contains second class passengers
sec_class <- titanic[titanic$class == "2nd",]

median(sec_class$age)
```

```
## [1] 29
```

```
sum(sec_class$age == 30)
```

```
## [1] 12
```

```
sum(sec_class$gender == "Female" & sec_class$age == 30)
```

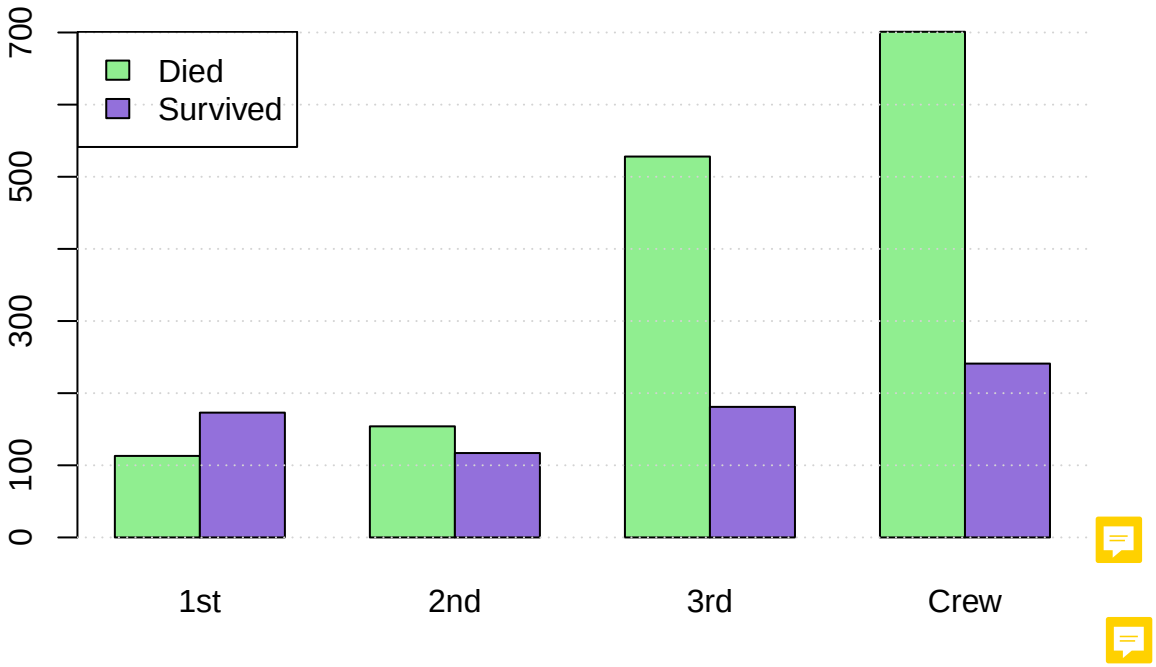
```
## [1] 8
```

```
sec_class$price <- sec_class$pnd + sec_class$shl/20 + sec_class$pnc/240

class_survived <- table(titanic$survived, titanic$class)
class_survived
```

```
##
##         1st 2nd 3rd Crew
##   FALSE 113 154 528  701
##   TRUE  173 117 181  241
```

```
barplot(class_survived,
        beside = TRUE,
        col = c("lightgreen", "mediumpurple"))
grid(nx = NA, ny = NULL)
legend("topleft", legend = c("Died", "Survived"),
       fill = c("lightgreen", "mediumpurple"))
```


Died
Survived
0
100
300
500
700
1st
2nd
3rd
Crew