Godambe.R

Maya S. Godambe

2022-10-17

```
# Loading Packages necessary for analyzing data
library(readr)
library(package = "tidyverse")
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v dplyr   1.0.10
## v tibble  3.1.8      v stringr 1.4.1
## v tidyr   1.2.1      v forcats 0.5.2
## v purrr   0.3.5
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(package = "table1")
```

```
##
## Attaching package: 'table1'
##
## The following objects are masked from 'package:base':
##
##     units, units<-
```

```
library(package = "descr")
library(package = "lsr")
require(SASxport)
```

```
## Loading required package: SASxport
```

```
## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,
## logical.return = TRUE, : there is no package called 'SASxport'
```

```
library(foreign)
library(dplyr)
```

```
#Loading data from git up repository use "read_csv" command. Data titled "titanic"
titanic<-read_csv("https://raw.githubusercontent.com/kijohnson/Data/main/titanic.csv")
```

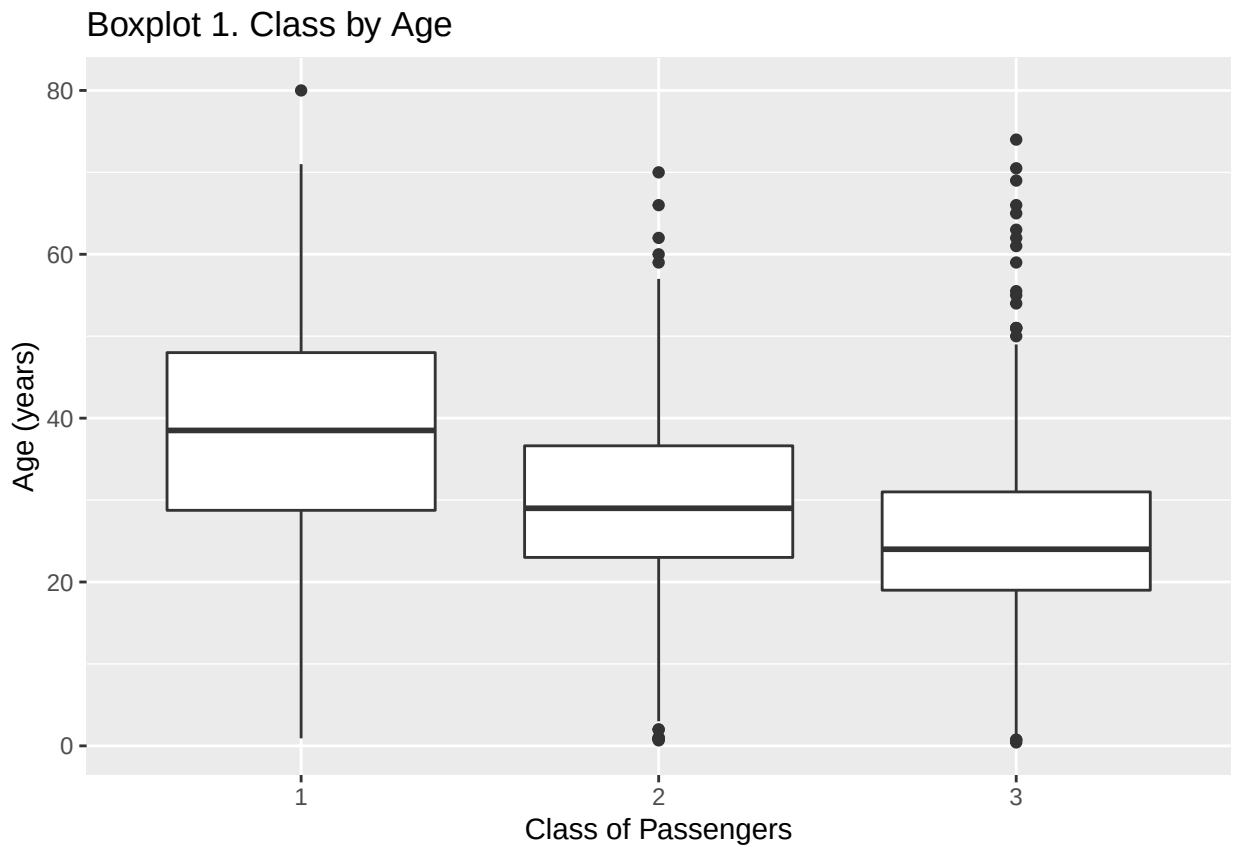
```
## Rows: 887 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): Name, Sex
## dbl (6): Survived, Pclass, Age, Siblings/Spouses Aboard, Parents/Children Ab...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#RECODE CLASS AS A FACTOR with levels, 1,2, and 3
titanic1 <- titanic %>%
  mutate(Pclass = recode_factor(.x = Pclass,
                                '1' = '1',
                                '2' = '2',
                                '3' = '3'))
```

```
# Class 1 has a median age of 38.5 years, Class 2 has a median age of 29 years,
#and Class 3 has a median age of 24 years
aggregate(titanic1$Age,
          list(titanic1$Pclass),
          median)
```

```
##   Group.1    x
## 1       1 38.5
## 2       2 29.0
## 3       3 24.0
```

```
#boxplot of class by Age (years)
ggplot(data = titanic1, aes(x = Pclass, y = Age),
       group= Age) +
  geom_boxplot() +
  labs(x = "Class of Passengers", y = "Age (years)",  title = "Boxplot 1. Class by Age ")
```




```
# using the dplyr package, finding percentage of males and females by class
titanic1 %>%
    group_by(Pclass, Sex) %>%
    summarise( percent = 100 * n() / nrow(titanic))
```

```
## `summarise()` has grouped output by 'Pclass'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 6 x 3
## # Groups:   Pclass [3]
##   Pclass Sex    percent
##   <fct>  <chr>    <dbl>
## 1 1      female   10.6
## 2 1      male     13.8
## 3 2      female    8.57
## 4 2      male     12.2
## 5 3      female   16.2
## 6 3      male     38.7
```