Diamond sizes

Rajas Purohit

05-22-2023

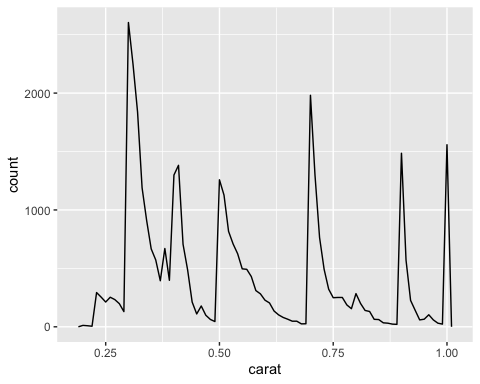
print("Hello in Markdown!")

## [1] "Hello in Markdown!"

head(Titanic)

## , , Age = Child, Survived = No  
##   
## Sex  
## Class Male Female  
## 1st 0 0  
## 2nd 0 0  
## 3rd 35 17  
## Crew 0 0  
##   
## , , Age = Adult, Survived = No  
##   
## Sex  
## Class Male Female  
## 1st 118 4  
## 2nd 154 13  
## 3rd 387 89  
## Crew 670 3  
##   
## , , Age = Child, Survived = Yes  
##   
## Sex  
## Class Male Female  
## 1st 5 1  
## 2nd 11 13  
## 3rd 13 14  
## Crew 0 0  
##   
## , , Age = Adult, Survived = Yes  
##   
## Sex  
## Class Male Female  
## 1st 57 140  
## 2nd 14 80  
## 3rd 75 76  
## Crew 192 20

We have data about 53940 diamonds. Only 17502 are larger than 1 carat. The distribution of the remainder is shown below:



# Week 4: dplyr package

#Task: Write the function to get a dataset from Base R: Titanic  
#and give the dataframe a new name of your choice  
#(hint: you will want your data to be a dataframe. Use the function: as.data.frame(Titanic))  
titanic <- as.data.frame(Titanic)  
  
#See the top rows of the data  
#TASK: Write the function to see the top rows of the data  
head(titanic)

## Class Sex Age Survived Freq  
## 1 1st Male Child No 0  
## 2 2nd Male Child No 0  
## 3 3rd Male Child No 35  
## 4 Crew Male Child No 0  
## 5 1st Female Child No 0  
## 6 2nd Female Child No 0

#Install and call the package dplyr  
#TASK: Write the functions to install and call dplyr  
#install.packages("dplyr")  
library(dplyr)

#Let's just see the Survived and Sex columns  
#Task: Write the function to 'select' the Survived and Sex columns   
#(hint: use the 'select' function)  
#titanic\_select <- as\_tibble(titanic)  
#titanic\_select %>% select(Survived, Sex)  
select(titanic, c(Survived, Sex))

## Survived Sex  
## 1 No Male  
## 2 No Male  
## 3 No Male  
## 4 No Male  
## 5 No Female  
## 6 No Female  
## 7 No Female  
## 8 No Female  
## 9 No Male  
## 10 No Male  
## 11 No Male  
## 12 No Male  
## 13 No Female  
## 14 No Female  
## 15 No Female  
## 16 No Female  
## 17 Yes Male  
## 18 Yes Male  
## 19 Yes Male  
## 20 Yes Male  
## 21 Yes Female  
## 22 Yes Female  
## 23 Yes Female  
## 24 Yes Female  
## 25 Yes Male  
## 26 Yes Male  
## 27 Yes Male  
## 28 Yes Male  
## 29 Yes Female  
## 30 Yes Female  
## 31 Yes Female  
## 32 Yes Female

#Let's name the dataset with just the two columns, Survived and Sex  
#TASK: Write the function to save the two columns as one new dataset  
#and give it a name  
titanic\_survived\_and\_sex <- select(titanic, c(Survived, Sex))

#Let's get rid of the Sex column in the new dataset created above  
#TASK: Write the function that deselects the sex column  
#(hint: use the 'select' function to not select a -column)  
titanic\_survived\_and\_sex %>% select(Survived)

## Survived  
## 1 No  
## 2 No  
## 3 No  
## 4 No  
## 5 No  
## 6 No  
## 7 No  
## 8 No  
## 9 No  
## 10 No  
## 11 No  
## 12 No  
## 13 No  
## 14 No  
## 15 No  
## 16 No  
## 17 Yes  
## 18 Yes  
## 19 Yes  
## 20 Yes  
## 21 Yes  
## 22 Yes  
## 23 Yes  
## 24 Yes  
## 25 Yes  
## 26 Yes  
## 27 Yes  
## 28 Yes  
## 29 Yes  
## 30 Yes  
## 31 Yes  
## 32 Yes

#Let's rename a column name  
#TASK: Write the function that renames 'Sex' to 'Gender'  
colnames(titanic)[2] ="Gender"

#Let's make a new dataframe with the new column name  
#TASK: Write the function that names a new dataset that includes the 'gender' column  
titanic\_gender <- select(titanic, Gender)

#Let's 'filter' just the males from our dataset  
#TASK: Write the function that includes only rows that are 'male'  
titanic\_gender\_male <- subset(titanic, Gender = "Male")

#Let's 'arrange' our data by gender (not the data you just filtered)  
#TASK: Write the function to group the data by gender (hint: arrange())  
arrange(titanic, Gender)

## Class Gender Age Survived Freq  
## 1 1st Male Child No 0  
## 2 2nd Male Child No 0  
## 3 3rd Male Child No 35  
## 4 Crew Male Child No 0  
## 5 1st Male Adult No 118  
## 6 2nd Male Adult No 154  
## 7 3rd Male Adult No 387  
## 8 Crew Male Adult No 670  
## 9 1st Male Child Yes 5  
## 10 2nd Male Child Yes 11  
## 11 3rd Male Child Yes 13  
## 12 Crew Male Child Yes 0  
## 13 1st Male Adult Yes 57  
## 14 2nd Male Adult Yes 14  
## 15 3rd Male Adult Yes 75  
## 16 Crew Male Adult Yes 192  
## 17 1st Female Child No 0  
## 18 2nd Female Child No 0  
## 19 3rd Female Child No 17  
## 20 Crew Female Child No 0  
## 21 1st Female Adult No 4  
## 22 2nd Female Adult No 13  
## 23 3rd Female Adult No 89  
## 24 Crew Female Adult No 3  
## 25 1st Female Child Yes 1  
## 26 2nd Female Child Yes 13  
## 27 3rd Female Child Yes 14  
## 28 Crew Female Child Yes 0  
## 29 1st Female Adult Yes 140  
## 30 2nd Female Adult Yes 80  
## 31 3rd Female Adult Yes 76  
## 32 Crew Female Adult Yes 20