DS\_01.R

anany

2024-06-15

#loading necessary libraries  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.3

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.3.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(readr)

## Warning: package 'readr' was built under R version 4.3.3

#loading the dataset  
Titanic\_Dataset <- read\_csv("~/Titanic-Dataset.csv")

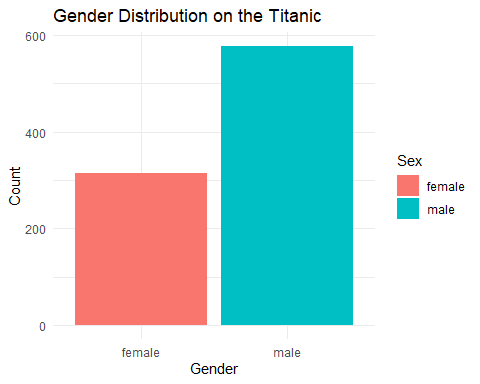
## Rows: 891 Columns: 12

## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (5): Name, Sex, Ticket, Cabin, Embarked  
## dbl (7): PassengerId, Survived, Pclass, Age, SibSp, Parch, Fare  
##   
## 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.

print(Titanic\_Dataset)

## # A tibble: 891 × 12  
## PassengerId Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin  
## <dbl> <dbl> <dbl> <chr> <chr> <dbl> <dbl> <dbl> <chr> <dbl> <chr>  
## 1 1 0 3 Braun… male 22 1 0 A/5 2… 7.25 <NA>   
## 2 2 1 1 Cumin… fema… 38 1 0 PC 17… 71.3 C85   
## 3 3 1 3 Heikk… fema… 26 0 0 STON/… 7.92 <NA>   
## 4 4 1 1 Futre… fema… 35 1 0 113803 53.1 C123   
## 5 5 0 3 Allen… male 35 0 0 373450 8.05 <NA>   
## 6 6 0 3 Moran… male NA 0 0 330877 8.46 <NA>   
## 7 7 0 1 McCar… male 54 0 0 17463 51.9 E46   
## 8 8 0 3 Palss… male 2 3 1 349909 21.1 <NA>   
## 9 9 1 3 Johns… fema… 27 0 2 347742 11.1 <NA>   
## 10 10 1 2 Nasse… fema… 14 1 0 237736 30.1 <NA>   
## # ℹ 881 more rows  
## # ℹ 1 more variable: Embarked <chr>

#bar chart for gender distribution  
gender\_distribution <- Titanic\_Dataset %>%  
 filter(!is.na(Sex)) %>%  
 count(Sex)  
  
ggplot(gender\_distribution, aes(x = Sex, y = n, fill = Sex)) +  
 geom\_bar(stat = "identity") +  
 labs(title = "Gender Distribution on the Titanic", x = "Gender", y = "Count") +  
 theme\_minimal()



# Histogram for Age Distribution  
ggplot(Titanic\_Dataset, aes(x = Age)) +  
 geom\_histogram(binwidth = 5, fill = "blue", color = "black", na.rm = TRUE) +  
 labs(title = "Age Distribution on the Titanic", x = "Age", y = "Frequency") +  
 theme\_minimal()

