R Notebook

#Load Librarys

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.0 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.1 ✔ tibble 3.1.8  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

library(patchwork)

## Warning: package 'patchwork' was built under R version 4.2.3

#Import Dataset

titanic <- read.csv("titanic(1).csv",header=TRUE)

#EDA

str(titanic)

## 'data.frame': 1313 obs. of 3 variables:  
## $ Name : chr "Allen, Miss Elisabeth Walton" "Allison, Miss Helen Loraine" "Allison, Mr Hudson Joshua Creighton" "Allison, Mrs Hudson JC (Bessie Waldo Daniels)" ...  
## $ PClass : chr "1st" "1st" "1st" "1st" ...  
## $ Survived: int 1 0 0 0 1 1 1 0 1 0 ...

summary(titanic)

## Name PClass Survived   
## Length:1313 Length:1313 Min. :0.0000   
## Class :character Class :character 1st Qu.:0.0000   
## Mode :character Mode :character Median :0.0000   
## Mean :0.3427   
## 3rd Qu.:1.0000   
## Max. :1.0000

#Wrangle

titanic$Name <- as.character(titanic$Name)  
titanic$Survived <- as.factor(titanic$Survived)  
titanic$PClass <- as.factor(titanic$PClass)  
titanic$Survived <- fct\_recode(titanic$Survived,"Died"="0","Lived"="1")  
str(titanic)

## 'data.frame': 1313 obs. of 3 variables:  
## $ Name : chr "Allen, Miss Elisabeth Walton" "Allison, Miss Helen Loraine" "Allison, Mr Hudson Joshua Creighton" "Allison, Mrs Hudson JC (Bessie Waldo Daniels)" ...  
## $ PClass : Factor w/ 4 levels "\*","1st","2nd",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ Survived: Factor w/ 2 levels "Died","Lived": 2 1 1 1 2 2 2 1 2 1 ...

#Create variable title, extract the second word of each name.

titanic$Title <- sapply(titanic$Name, FUN=function(x)   
{regmatches(x,regexec(',[[:space:]](.\*?)[[:space:]]',x))[[1]][2]})

#Create variable gender, create list of female titles and if they are in the list then the gender #is set to female if not it is set to male.

Female\_titles <- c("Miss","Mrs","Lady","Madame","Ms")  
titanic$Gender <- ifelse(titanic$Title %in% Female\_titles, "Female","Male")

#Create new variable for maritial status

titanic$martial\_status <- ifelse(titanic$Title %in% c("Mrs", "Lady",   
"Madame"), "Married", "Not married")  
titanic$martial\_status <- as.factor(titanic$martial\_status)

#Change new variables into factors so we can plot them.

titanic$martial\_status <- as.factor(titanic$martial\_status)  
titanic$Gender <- as.factor(titanic$Gender)

Is there a relationship between survival and gender? - Yes.

plot1 <- titanic %>%   
ggplot(data=titanic,mapping=aes(x=Survived,fill=Gender)) +  
 geom\_bar(stat="count") +  
 xlab("Survival") +  
 ylab("Count") +  
 ggtitle("Relationship between Survival and Gender")+  
 scale\_fill\_manual(values = c("gray", "black"))

#pass titanic to a new dataset to get rid of annoying class \*

titanic\_right <- titanic %>%   
 filter(PClass != "\*")

Is there a relationship between survial and class? -Yes

plot2 <- titanic\_right %>%   
ggplot(data=titanic\_right,mapping=aes(x=Survived,fill=PClass)) +  
 geom\_bar(stat = "count") +  
 xlab("Survival") +  
 ylab("Passenger Class") +  
 ggtitle("Relationship between survival and class") +  
 scale\_fill\_manual(values = c("grey", "black", "orange"))

Is there a relationship between survival and marital status? #Create new dataset and filter for only females, since we cannot do males.

titanic\_females <- titanic %>%   
 filter(Gender == "Female")

#Reorder it.

titanic\_females$Survived <- fct\_rev(titanic\_females$Survived)

Is there a relationship between survival and marital status? - yes it seems like less married women died then non-women. We dont have enough data for males.

plot3 <- titanic\_females %>%   
ggplot(data = titanic\_females,mapping = aes(x=Survived,fill=martial\_status)) +  
geom\_bar(stat = "count") +  
 xlab("survival") +  
 ylab("martial status") +  
 ggtitle("Relationship between survival and martial status for females") +  
 scale\_fill\_manual(values = c("gray", "black"))

((plot1 / plot2 / plot3)) +  
plot\_annotation(title = "Titanic Relationship Plots",  
theme = theme(plot.title = element\_text(size = 20,  
colour = "purple"))) +  
theme(text = element\_text('mono'))

