Summarize and Visualize More Than Two Categorical Variables

eCornell

5/3/2021

Step 1: Load the data set and define the colors.

Q: How does male/female survivorship vary across passenger class?

Step 2: Examine how male and female survivorship varied across passenger class.

To examine the differences in how passenger class affected males and females, transform the survival status to 1 (yes) and 0 (no) with the ifelse() command. Then, use the aggregate() function to calculate proportions in each group.

```
titanic$SurvBin=ifelse(titanic$Survived=="Yes", 1, 0) # Use the ifelse() function to turn "Yes" into 1 and "No" i
nto 0

# The aggregate() function takes the formula for a question then performs the specified function on each group cr
eated by the question:
prop=aggregate(
               SurvBin ~ PClass + Sex, # This argument tells R how we want to separate our data: We'll calculate
the proportions based on SurvBin, and separate groups based on both Passenger Class and Sex, for a total of six g
roups
               FUN=mean, # Calculate the mean for each group (mean could be switched for another function, such a
s median)
               data=titanic)  # Perform this function on the titanic data
prop # Display the table of means by group.
```

```
##   PClass    Sex   SurvBin
## 1    1st female 0.9370629
## 2    2nd female 0.8785047
## 3    3rd female 0.3773585
## 4    1st   male 0.3296089
## 5    2nd   male 0.1445087
## 6    3rd   male 0.1162325
```

Step 3: Visualize the differences in survival by passenger class and sex with a barplot