Summarize and Visualize Two Categorical Variables

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

Q: Was the survival rate among male passengers higher than the survival rate among female passengers?

We need to know four numbers: 1. Number of male passengers who survived 2. Number of female passengers who survived 3. Number of male passengers who did not survive 4. Number of female passengers who did not survive

Step 2: Create a contingency table and calculate proportions of survivors grouped by sex.

Create a contingency table with the table() command that shows the number of individuals in each of the four groups of interest.

Use the table() command on the whole data set.

```
table(titanic$Sex, titanic$Survived)
```

```
##
##           No Yes
##   female 154 308
##   male   709 142
```

Use the prop.table() command to calculate proportions of survivors.

```
prop.table(table(titanic$Sex, titanic$Survived), 1)
```

```
##
##                 No       Yes
##   female 0.3333333 0.6666667
##   male   0.8331375 0.1668625
```

```
# Setting second argument to 1 will give us proportions
# within each row (% survivors among female, % survivors among male)

# Change this to 2 to calculate proportions
# within each column (e.g., % males among survivors)
```

Step 3: Create a stacked barplot to show number of survivors grouped by sex.