Homework2

Miles Tweed

1/24/2021

3. Practicing with the Titanic data

3.1 Find the overall proportion of males and females who survived the sinking of the Titanic.

(Ignore whether child or adult.)

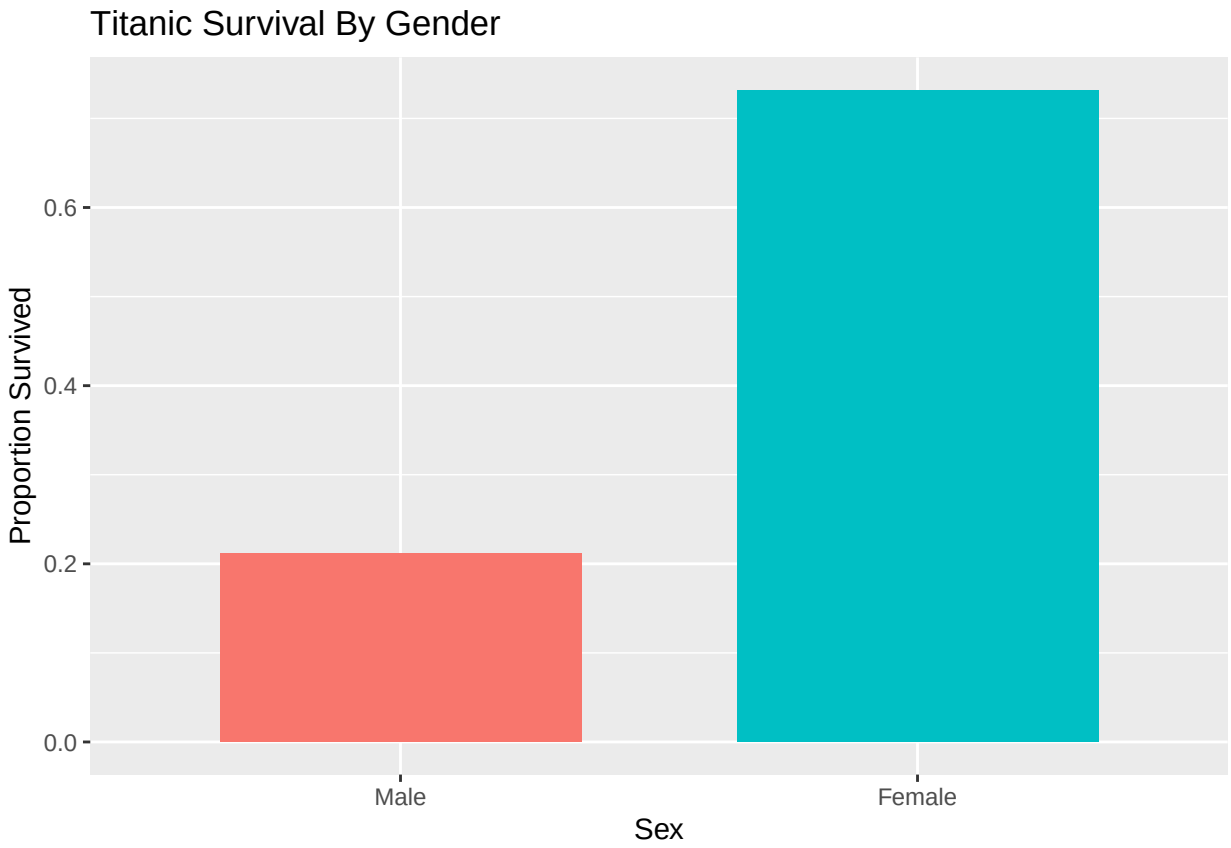
```
mydata <- as.data.frame(Titanic)

mydata %>%
  count(Sex, Survived, wt = Freq) %>%
  group_by(Sex) %>%
  mutate(Total=sum(n), Proportion = n/Total) %>%
  filter(Survived == 'Yes') %>%
  rename(Count = 'n')
```

```
## # A tibble: 2 x 5
## # Groups:   Sex [2]
##   Sex    Survived Count Total Proportion
##   <fct>  <fct>    <dbl> <dbl>      <dbl>
## 1 Male   Yes        367  1731      0.212
## 2 Female Yes        344   470      0.732
```

3.2 Plot the overall proportion of survival in a bar graph, by gender

```
mydata %>%
  count(Sex, Survived, wt = Freq) %>%
  group_by(Sex) %>%
  mutate(tot=sum(n), prop = n/tot) %>%
  filter(Survived == 'Yes') %>%
  ggplot(aes(x = Sex, y = prop, fill = Sex)) +
  geom_bar(stat = 'identity', width = 0.7) +
  labs(title = 'Titanic Survival By Gender') +
  ylab('Proportion Survived') +
  theme(legend.position = 'none')
```



3.3 Find the overall proportion of males and females who survived the sinking of the Titanic, by class booked.

(Ignore whether child or adult.)

```
# because we want props of survival by sex and class, we count by sex, class and survival

mydata %>%
  count(Sex, Class, Survived, wt=Freq) %>%
  group_by(Sex, Class) %>%
  mutate(Total = sum(n), Proportion = n/Total) %>%
  filter(Survived=='Yes') %>%
  rename(Count = 'n')
```

```
## # A tibble: 8 x 6
## # Groups:   Sex, Class [8]
##   Sex    Class Survived Count Total Proportion
##   <fct>  <fct> <fct>    <dbl> <dbl>      <dbl>
## 1 Male   1st   Yes         62   180      0.344
## 2 Male   2nd   Yes         25   179      0.140
## 3 Male   3rd   Yes         88   510      0.173
## 4 Male   Crew  Yes        192   862      0.223
## 5 Female 1st   Yes        141   145      0.972
## 6 Female 2nd   Yes         93   106      0.877
## 7 Female 3rd   Yes         90   196      0.459
## 8 Female Crew  Yes         20    23      0.870
```

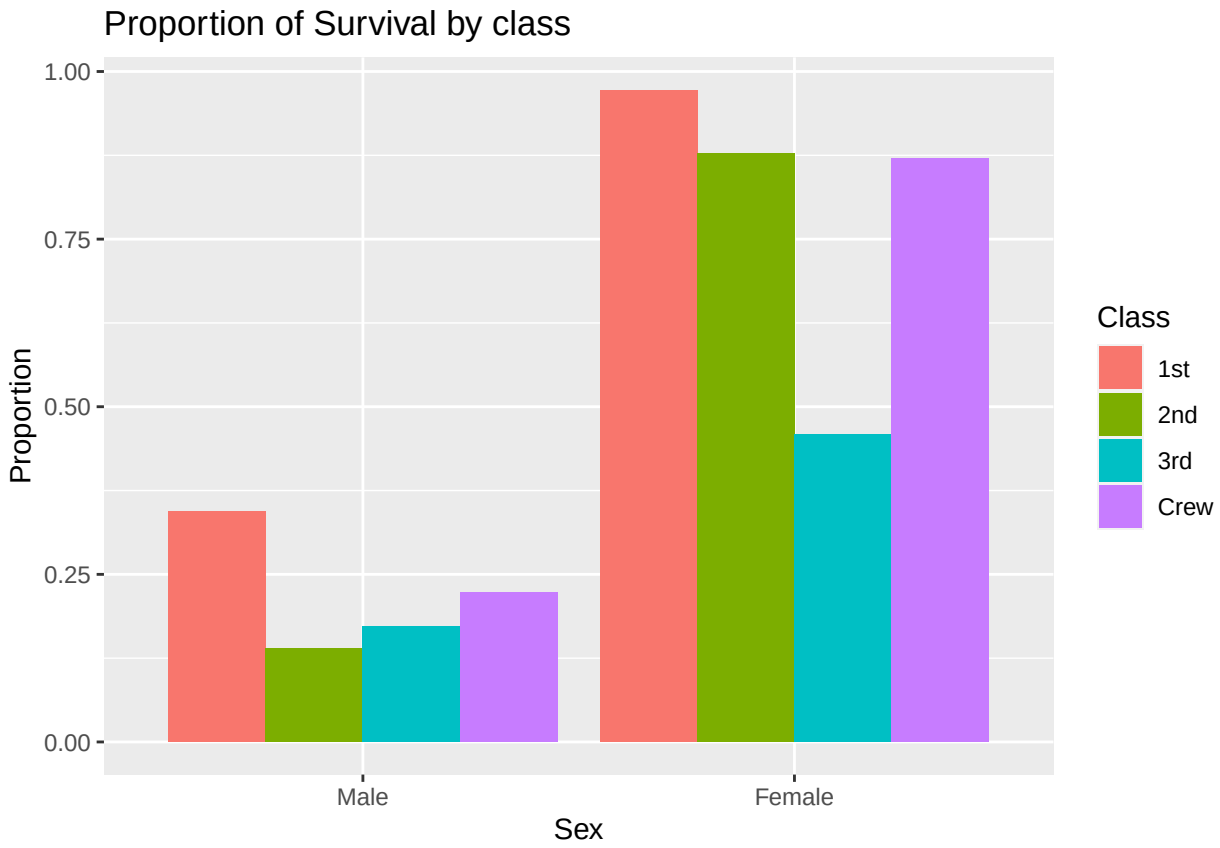
3.4 Plot the survival proportions by class

There are a couple of options: For each class, we could compare males vs. females:

```
mydata %>%
  count(Sex, Class, Survived, wt=Freq) %>%
  group_by(Sex, Class) %>%
  mutate(Total = sum(n), Proportion = n/Total) %>%
  filter(Survived=='Yes') %>%
  rename(Count = 'n') %>%
  ggplot(aes(x = Class, y = Proportion, fill = Sex)) +
  geom_bar(stat = 'identity', position = 'dodge') +
  labs(title = 'Proportion of Survival by class')
```

Or, we could compare the proportions surviving in each class, separately for males and females:

```
mydata %>%
  count(Sex, Class, Survived, wt=Freq) %>%
  group_by(Sex, Class) %>%
  mutate(Total = sum(n), Proportion = n/Total) %>%
  filter(Survived=='Yes') %>%
  rename(Count = 'n') %>%
  ggplot(aes(x = Sex, y = Proportion, fill = Class)) +
  geom_bar(stat = 'identity', position = 'dodge') +
  labs(title = 'Proportion of Survival by class')
```



3.5 Same as 3.3, but now also adjust for whether child or adult (the age variable).

So, find the proportion of passengers who survived, by sex, class and whether child or adult.

```
mydata %>%
  count(Sex, Class, Survived, Age, wt=Freq) %>%
  group_by(Sex, Class, Age) %>%
  mutate(Total = sum(n), Proportion = n/Total) %>%
  filter(Survived == 'Yes')
```

```
## # A tibble: 16 x 7
## # Groups:   Sex, Class, Age [16]
##    Sex    Class Survived Age       n Total Proportion
##    <fct>  <fct> <fct>    <fct> <dbl> <dbl>      <dbl>
##  1 Male   1st   Yes      Child     5     5      1
##  2 Male   1st   Yes      Adult    57   175      0.326
##  3 Male   2nd   Yes      Child    11    11      1
##  4 Male   2nd   Yes      Adult    14   168      0.0833
##  5 Male   3rd   Yes      Child    13    48      0.271
##  6 Male   3rd   Yes      Adult    75   462      0.162
##  7 Male   Crew  Yes      Child     0     0    NaN
##  8 Male   Crew  Yes      Adult   192   862      0.223
##  9 Female 1st   Yes      Child     1     1      1
## 10 Female 1st   Yes      Adult   140   144      0.972
## 11 Female 2nd   Yes      Child    13    13      1
## 12 Female 2nd   Yes      Adult    80    93      0.860
## 13 Female 3rd   Yes      Child    14    31      0.452
## 14 Female 3rd   Yes      Adult    76   165      0.461
```

```
## 15 Female Crew  Yes      Child     0     0   NaN
## 16 Female Crew  Yes      Adult    20    23     0.870
```

Or, more succinctly, using `add_count`:

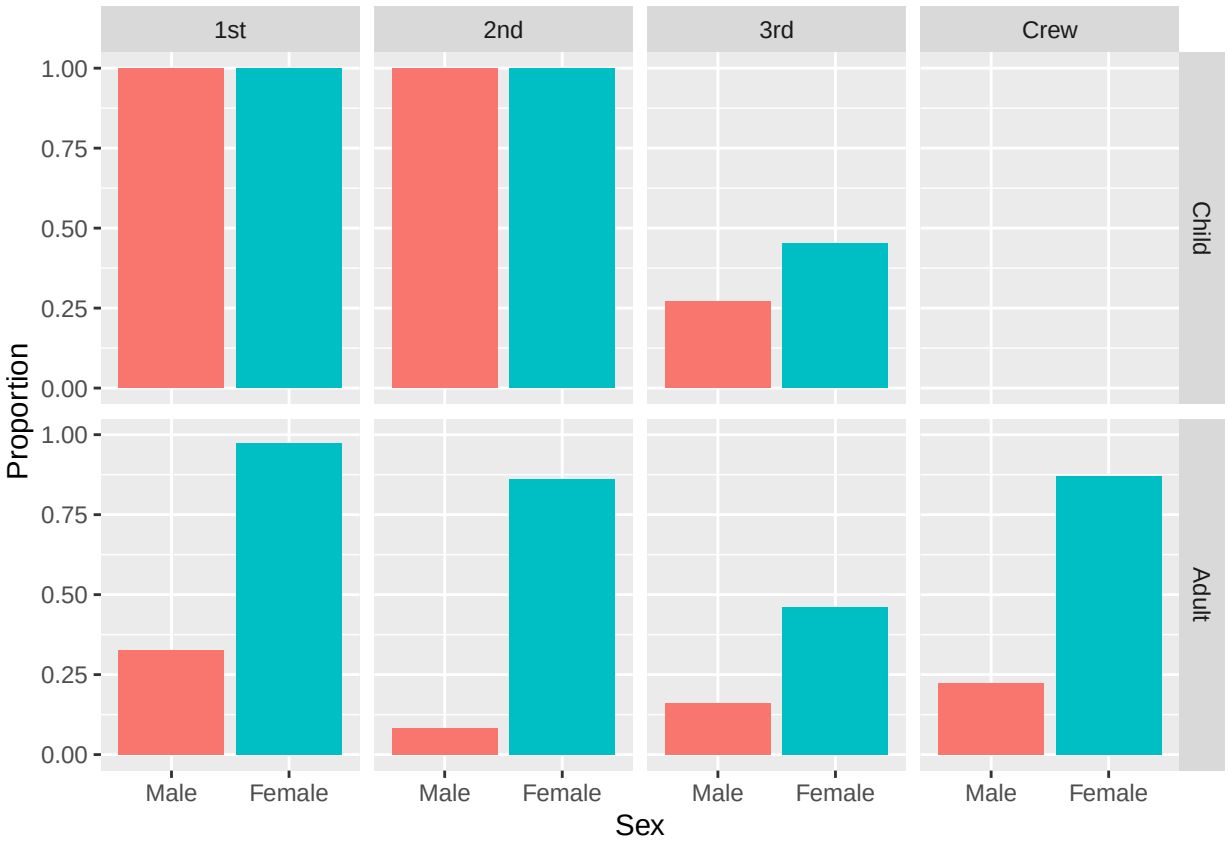
```
mydata %>%
  add_count(Class,Sex,Age, wt = Freq) %>%
  group_by(Class, Sex, Age) %>%
  mutate(Proportion = Freq/n) %>%
  filter(Survived == 'Yes')
```

```
## # A tibble: 16 x 7
## # Groups:   Class, Sex, Age [16]
##    Class Sex    Age   Survived  Freq     n Proportion
##    <fct> <fct>  <fct> <fct>    <dbl> <dbl>      <dbl>
##  1 1st   Male   Child Yes          5     5     1
##  2 2nd   Male   Child Yes         11    11     1
##  3 3rd   Male   Child Yes         13    48     0.271
##  4 Crew  Male   Child Yes          0     0   NaN
##  5 1st   Female Child Yes          1     1     1
##  6 2nd   Female Child Yes         13    13     1
##  7 3rd   Female Child Yes         14    31     0.452
##  8 Crew  Female Child Yes          0     0   NaN
##  9 1st   Male   Adult Yes         57   175     0.326
## 10 2nd   Male   Adult Yes         14   168     0.0833
## 11 3rd   Male   Adult Yes         75   462     0.162
## 12 Crew  Male   Adult Yes        192   862     0.223
## 13 1st   Female Adult Yes        140   144     0.972
## 14 2nd   Female Adult Yes         80    93     0.860
## 15 3rd   Female Adult Yes         76   165     0.461
## 16 Crew  Female Adult Yes         20    23     0.870
```

3.6 Plot the proportions when adjusting for all three variables (Sex, Class, Age): `facet_grid`

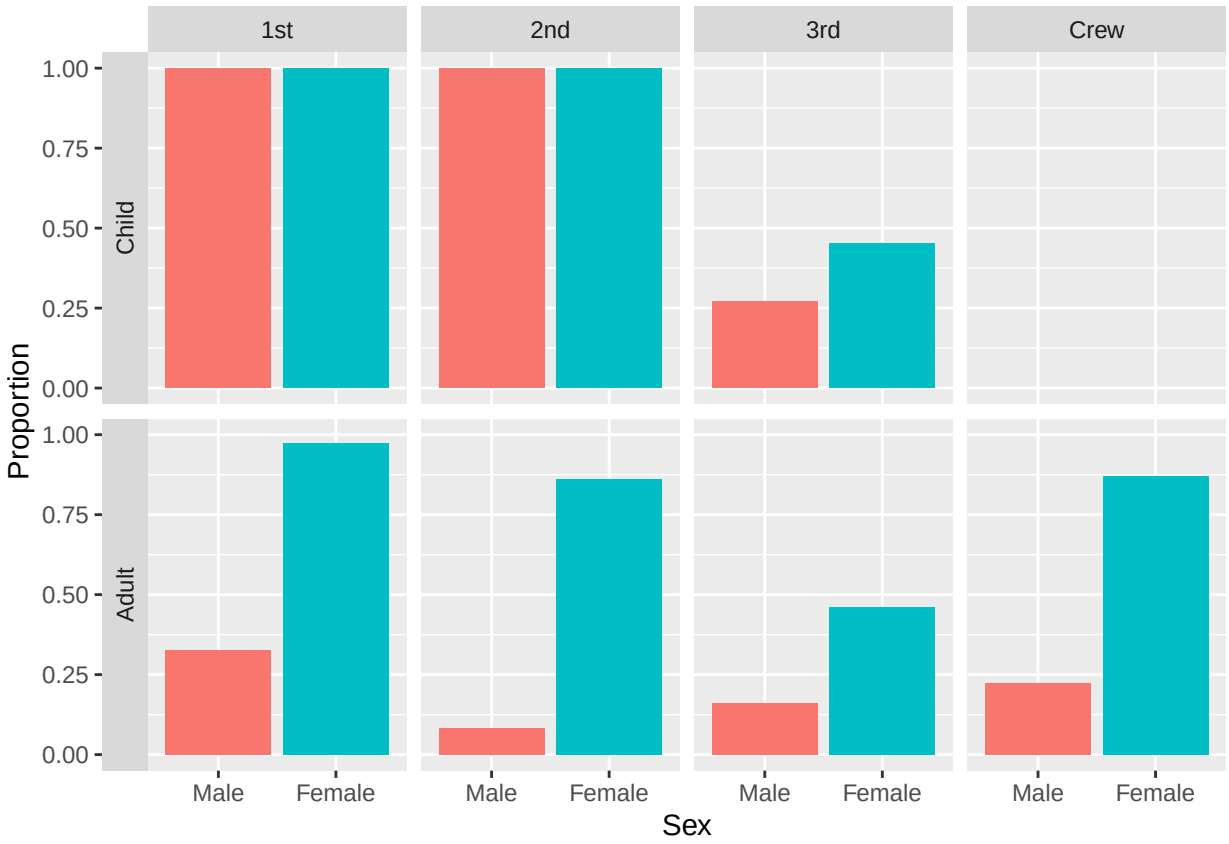
When comparing the proportions, we are now adusting for three variables: Sex, Class and Age. To get all these in one plot, we will assign one variable to be a row variable, one variable to be a column variable, and a third to be the variable that is displayed on the x-axis. Naturally, there are many possibilities. Let's first construct a graph that, similar to the second to last above, has sex on the x-axis, class as the column variable and age as the row variable. We assign the row and column variable in `facet_grid`, which we now use instead of `facet_wrap`:

```
mydata %>%
  add_count(Class,Sex,Age, wt = Freq) %>%
  group_by(Class, Sex, Age) %>%
  mutate(Proportion = Freq/n) %>%
  filter(Survived == 'Yes', !is.na(Proportion)) %>%
  ggplot(aes(x = Sex, y = Proportion, fill = Sex)) +
  geom_bar(stat = 'identity') +
  facet_grid(rows = vars(Age), cols = vars(Class)) +
  theme(legend.position = 'none')
```



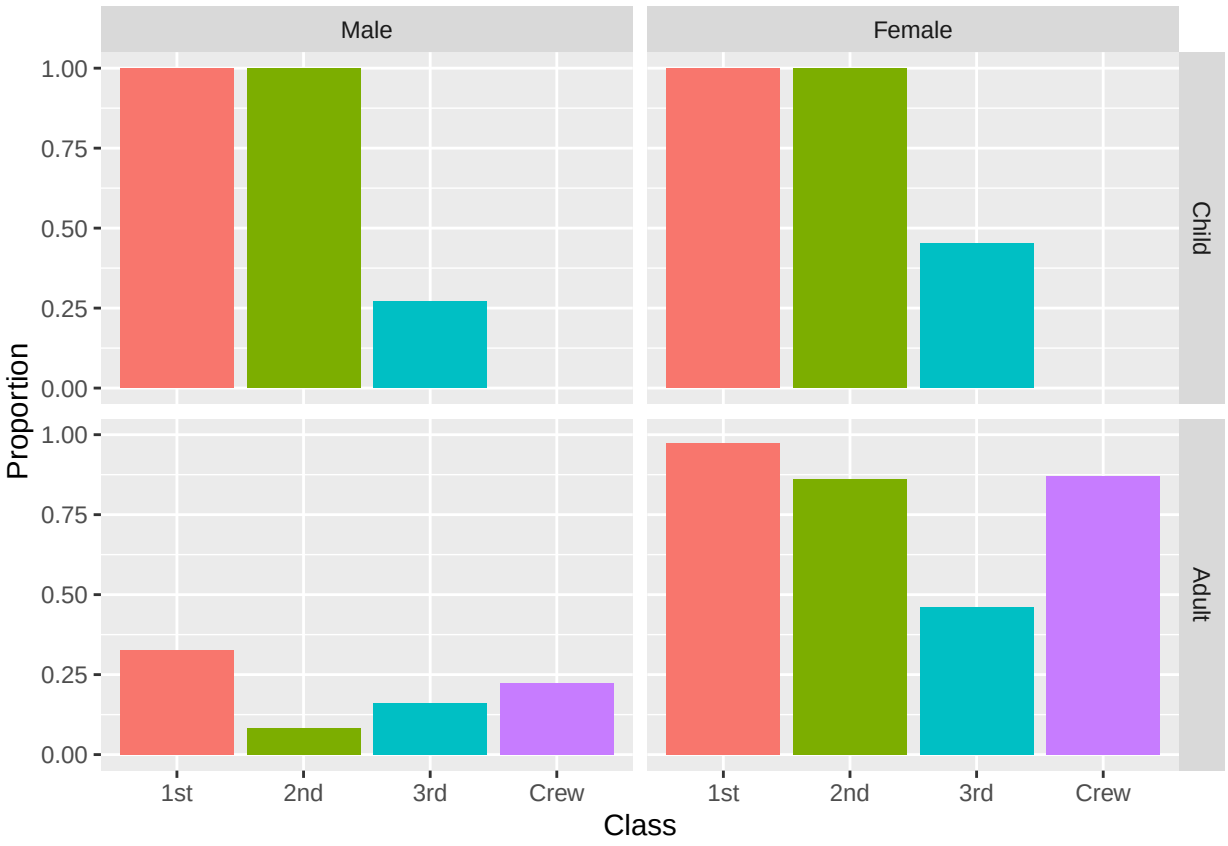
If you want the label of the column variable appear on the left side instead of on the right-side, use the `switch='y'` options:

```
mydata %>%
  add_count(Class,Sex,Age, wt = Freq) %>%
  group_by(Class, Sex, Age) %>%
  mutate(Proportion = Freq/n) %>%
  filter(Survived == 'Yes', !is.na(Proportion)) %>%
  ggplot(aes(x = Sex, y = Proportion, fill = Sex)) +
  geom_bar(stat = 'identity') +
  facet_grid(rows = vars(Age), cols = vars(Class), switch = 'y') +
  theme(legend.position = 'none')
```



Try other versions of this plot, by changing what you assign for row, column and x-axis variable!

```
mydata %>%
  add_count(Class,Sex,Age, wt = Freq) %>%
  group_by(Class, Sex, Age) %>%
  mutate(Proportion = Freq/n) %>%
  filter(Survived == 'Yes', !is.na(Proportion)) %>%
  ggplot(aes(x = Class, y = Proportion, fill = Class)) +
  geom_bar(stat = 'identity') +
  facet_grid(rows = vars(Age), cols = vars(Sex)) +
  theme(legend.position = 'none')
```


Male
Female
Child
Adult
Proportion
1.00
0.75
0.50
0.25
0.00
1st
2nd
3rd
Crew
Class