

### Questions 1-3

Here is a contingency table of college students with their **Favorite Color** (Red or Blue) down the columns and their **School** (Berkeley or Stanford) across the rows.

|          | Red | Blue |
|----------|-----|------|
| Berkeley | 10  | 90   |
| Stanford | 60  | 40   |

#### Question 1

Find the proportion of all students who attend Berkeley. What type of proportion is this?

#### Question 2

Find the proportion of all students who attend Berkeley *and* like red best. What type of proportion is this?

#### Question 3

*Of the students who attend Berkeley*, find the proportion that like red best. What type of proportion is this?

### Questions 4-5

RMS Titanic was a British passenger liner with 2,224 people aboard (including both passengers and crew). The Titanic sank in the North Atlantic Ocean on 15 April 1912 after striking an iceberg on her way to New York City. It was the deadliest sinking of a single ship at the time with almost 70% of the 2,224 passengers and crew dying.

Below are the first five rows of 2,224 for a data frame called `titanic`.

#### Question 4

What is the unit of observation for `titanic`?

#### Question 5

The three following graphs were generated using `titanic`.

Which graph is **best** used to answer the following questions? *Write the letter associated with the graph to the right of each question*

*How many first class passengers survived the Titanic's sinking?* \_\_\_\_\_

*Is there an association between a passenger's class and their survival?* \_\_\_\_\_

*What is the mode of the Class variable?* \_\_\_\_\_

## Questions 6-7

Open up RStudio and write down the code you used to complete each of the following questions in the space below. *Make sure you load in any libraries where necessary!*

### Question 6

Consider the vector q6 which was made as follows:

```
q6 <- c(1,2,3,4,5,6,NA)
```

Load q6 as is into your session. Then, write R code to calculate the mean of q6 so that the result is 3.5 (this is the mean of the numbers 1 through 6). *Hint: how can you learn more about the function mean()?*

### Question 7

The promote dataset can be found in the stat20data package. Using this data, write ggplot2 code to make a stacked, normalized bar chart having identified gender on the x axis, with the bars being filled in by promotion decision. Write the code you used below. Then, make a claim about the association between the two variables (we will revisit this study in more detail later in the course)!