

**The Scene:** The University of California at Berkeley was concerned that they may have discriminated against women in their graduate admissions process for the fall quarter of 1973. In this activity we explore the admission data for that term.

Begin this worksheet by completing the following steps:

1. Open a new script in RStudio.
2. In line 1 of your script enter and run `library(tidyverse)`
3. In line 2 of your script enter and run

```
berk = read.csv("https://mphitchman.com/stats/data/berk23.csv")
```

4. Use subsequent lines of your script for code to answer questions below.

- 
1. List the variable (column) names in this data set.
  2. How many total applicants are there in this data set?
  3. How many applicants were male? female? Hint: Try `table(berk$sex)`
  4. What proportion of applicants were male? female?
  5. How many applicants were admitted? denied admission?
  6. What proportion of applicants were admitted? denied admission?
  7. If you randomly selected 100 applicants from this data set, about how many would you expect to be male? About how many would you expect to have been accepted?
  8. Gather the data into a two-way table of sex by admission status. To do this, run this code:  
`table(berk$sex,berk$status)`

Fill in the table below:

	admitted	denied	total
female			
male			
total			

You may want to **visualize the two-way table with a stacked bar plot**

```
ggplot(berk)+  
  geom_bar(aes(x=sex,fill=status))
```

9. Consider for the moment just the male applicants. Of the males who applied to one of these programs, what proportion were admitted? What proportion of the females were admitted?
10. Do these proportions seem to support the claim that males were given preferential treatment in admission decisions?
11. To try to isolate the program or programs responsible for this mistreatment of female applicants, we can calculate the proportion of males and the proportion of females within each program who were admitted. The table below shows the admission numbers separated by program for males and females. In the last two columns, record the proportion of male and female applicants, respectively, admitted to each program.

Program	m admit	m denied	f admit	f denied	prop'n m admit	prop'n f admit
A	512	313	89	19		
B	353	207	17	8		
C	120	205	202	391		
D	138	279	131	244		
E	53	138	94	299		
F	22	351	24	317		

FYI, the following code would find these counts: `table(berk$sex,berk$status,berk$program)`

You may want to **visualize the different admission rates by sex in each program**:

```
ggplot(berk)+
  geom_bar(aes(x=sex,fill=status),position="fill")+
  facet_grid(. ~ program)+
  ylab("admission rate")+
  ggtitle("Admission rates by Sex and Program")
```

12. Does it seem as if any program is responsible for the large discrepancy between men and women in the overall proportions admitted?
13. Reason from the data given to explain how it happened that men had a much higher rate of admission overall even though women had higher rates in most programs and no program favored men very strongly.