# Well-Known Example of Confounding

A famous example of an unaccounted for confounding variable having serious repercussions was when the UC Berkeley was sued in 1973 for bias against women who had applied for admission to graduate schools.

Let's consider the $n = 4526$ people who applied to the 6 largest departments.
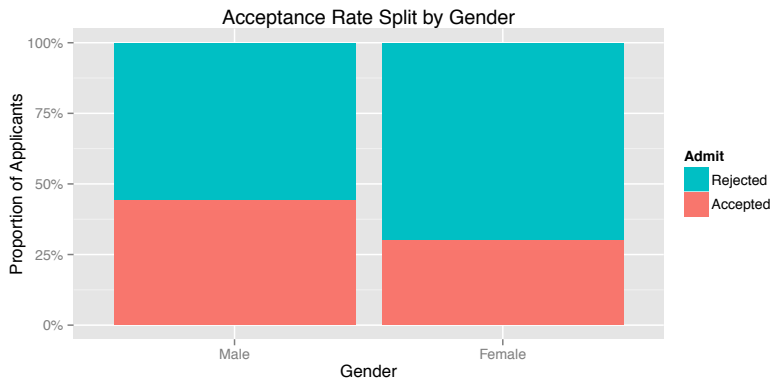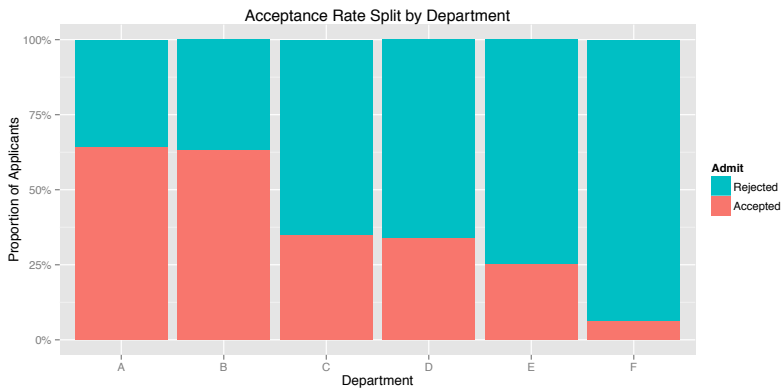
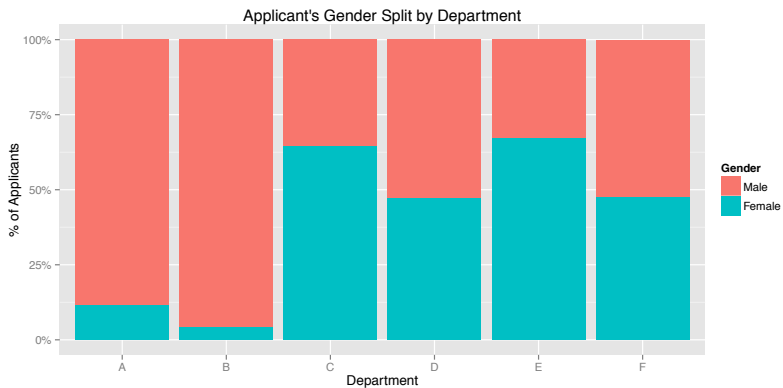Of the $n = 4526$ applicants:

Overall Acceptance Rate
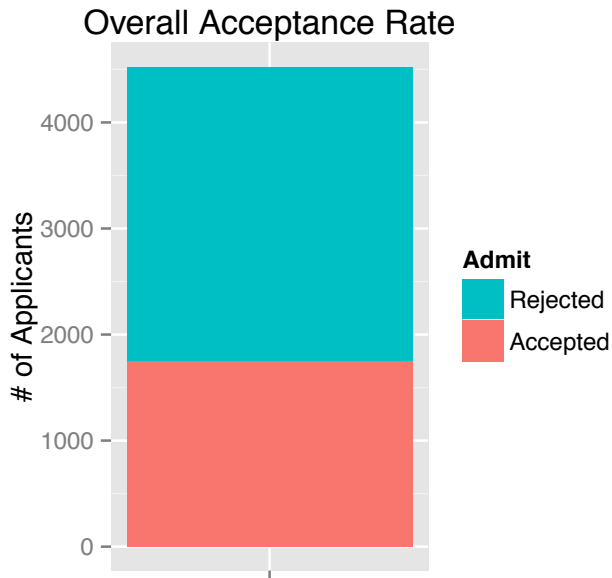
Split the counts by gender:
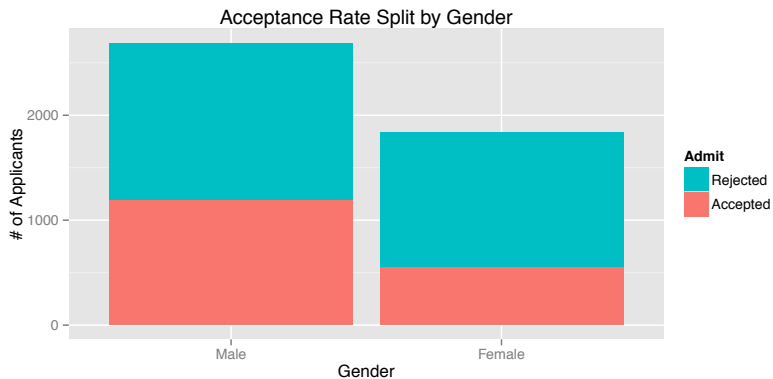
Look at proportions instead of counts:

What was the "competitiveness" of departments?

Where were the women applying?
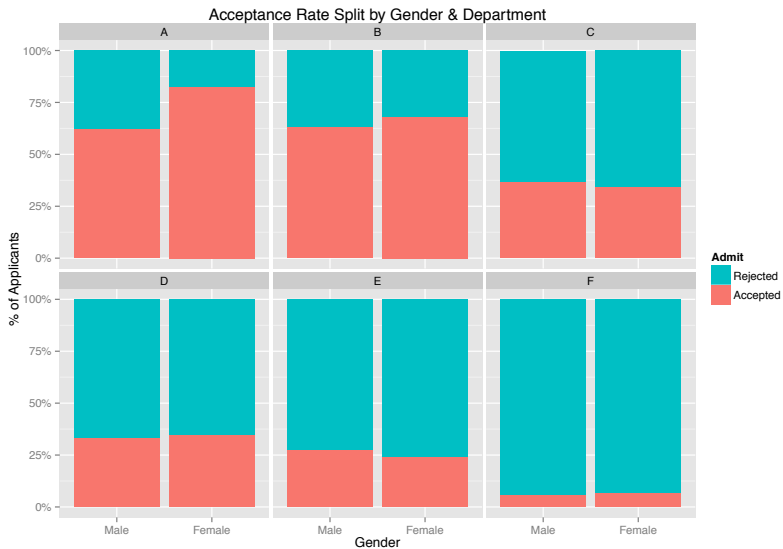

Applicant's Gender Split by Department

So while in aggregate things looked like this:

Acceptance Rate Split by Gender
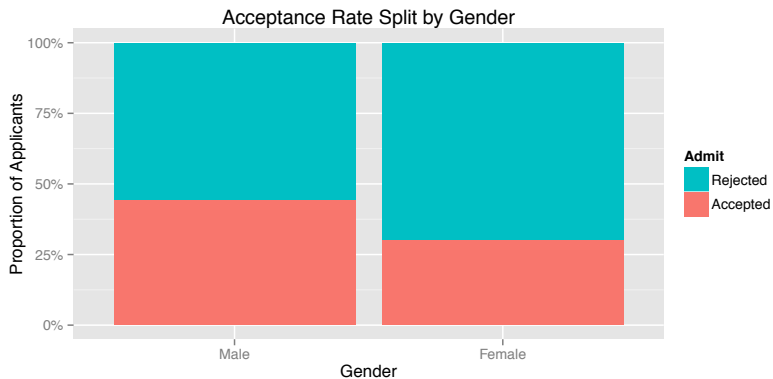
You need to account for department!

Acceptance Rate Split by Gender & Department

Bickel et al.'s (1975) Explanation

There was the presence of a confounding variable: competitiveness of applying to the department, which is a function

- number of applicants
- number of available slots

So it wasn't that departments were discriminating against women, rather:

- women tended to apply to departments with high competition and hence lower admission rates, primarily the humanities.
- men tended to apply to departments with low competition and hence higher admission rates, primarily the sciences.

Bickel et al.'s (1975) Explanation

In fact, Bickel et al. found that "If the data are properly pooled...there is a small but statistically significant bias in favor of women."

This was the exact opposite claim of the lawsuit. This is known as Simpson's Paradox.

Simpson's Paradox

(From Wikipedia) Simpson's paradox occurs when a trend that appears in different groups of data disappears when these groups are combined, and the reverse trend appears for the aggregate data.

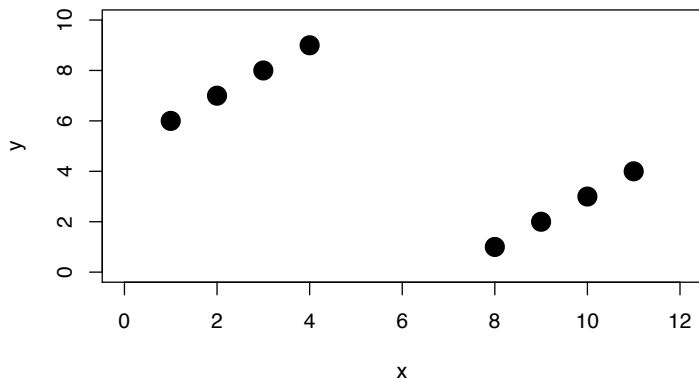This is due to a confounding variable.

Bickel et al.'s (1975) Conclusion

"The bias in the aggregated data stems not from any pattern of discrimination on the part of admissions committees, which seem quite fair on the whole, but apparently from prior screening at earlier levels of the educational system."

"Women are shunted by their socialization and education toward fields of graduate study that are generally more crowded, less productive of completed degrees, and less well funded, and that frequently offer poorer professional employment prospects."

The original paper can be found here.
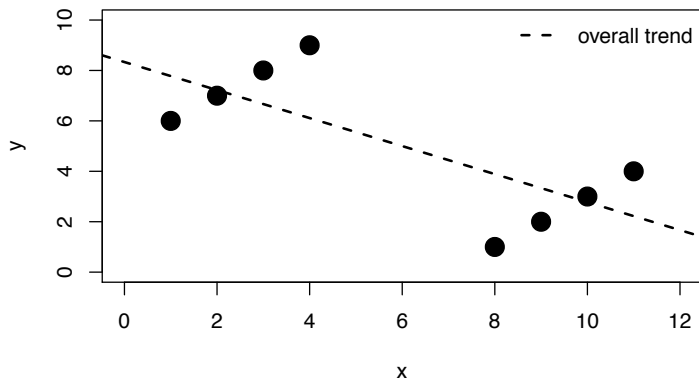
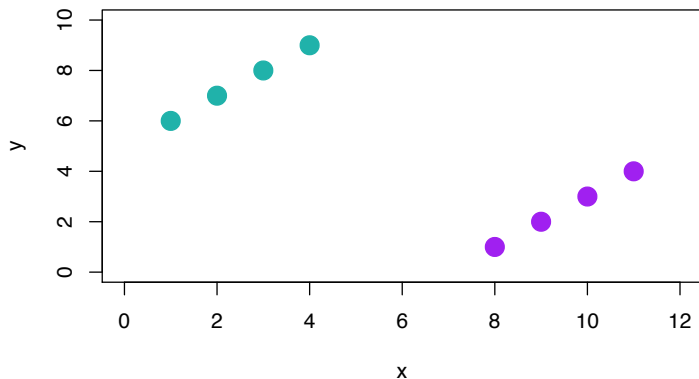# A Graphical Illustration of Simpson's Paradox

Say we have the following points:

A Graphical Illustration of Simpson's Paradox

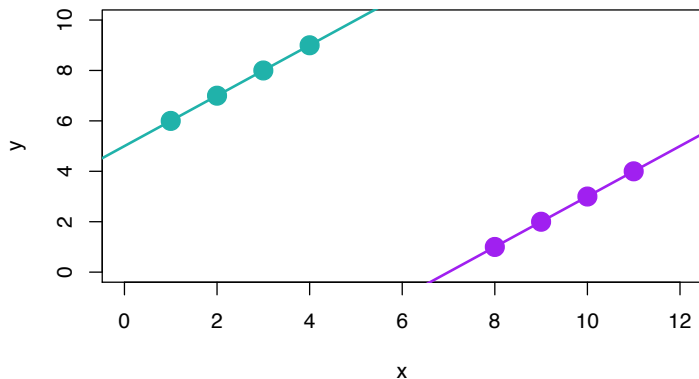Overall, if we fit a single line, the explanatory variable x is negatively related with the outcome variable y:

A Graphical Illustration of Simpson's Paradox

But say we consider a confounding variable, in this case color, and fit two separate lines for each group:

A Graphical Illustration of Simpson's Paradox

The subgroups now exhibit a positive relationship!

A Graphical Illustration of Simpson's Paradox

i.e. the trend in aggregate is the reverse of the trend in the subgroups (teal & purple).