Homework 1 - UC Berkeley Admission Trends

AUTHOR
Liyans Mathews (24201531)

Introduction

The **UC Berkely Admissions** dataset is a R built-in dataset that summarizes the admission figures at the University of California, Berkeley, for six departments. It includes *admission status*, *gender*, *department* and *frequencies of each combination*.

This dataset was originally analyzed in a 1975 article, which explored allegations of gender bias in graduate school admissions. At first glance, the overall admission rates appeared to favor male applicants. However, a closer examination by department revealed a more complex outcome a phenomenon known as Simpson's Paradox, where trends in aggregated data disappear when examined in subgroups.

A brief discussion of gender-wise admission distribution across departments has been done below.

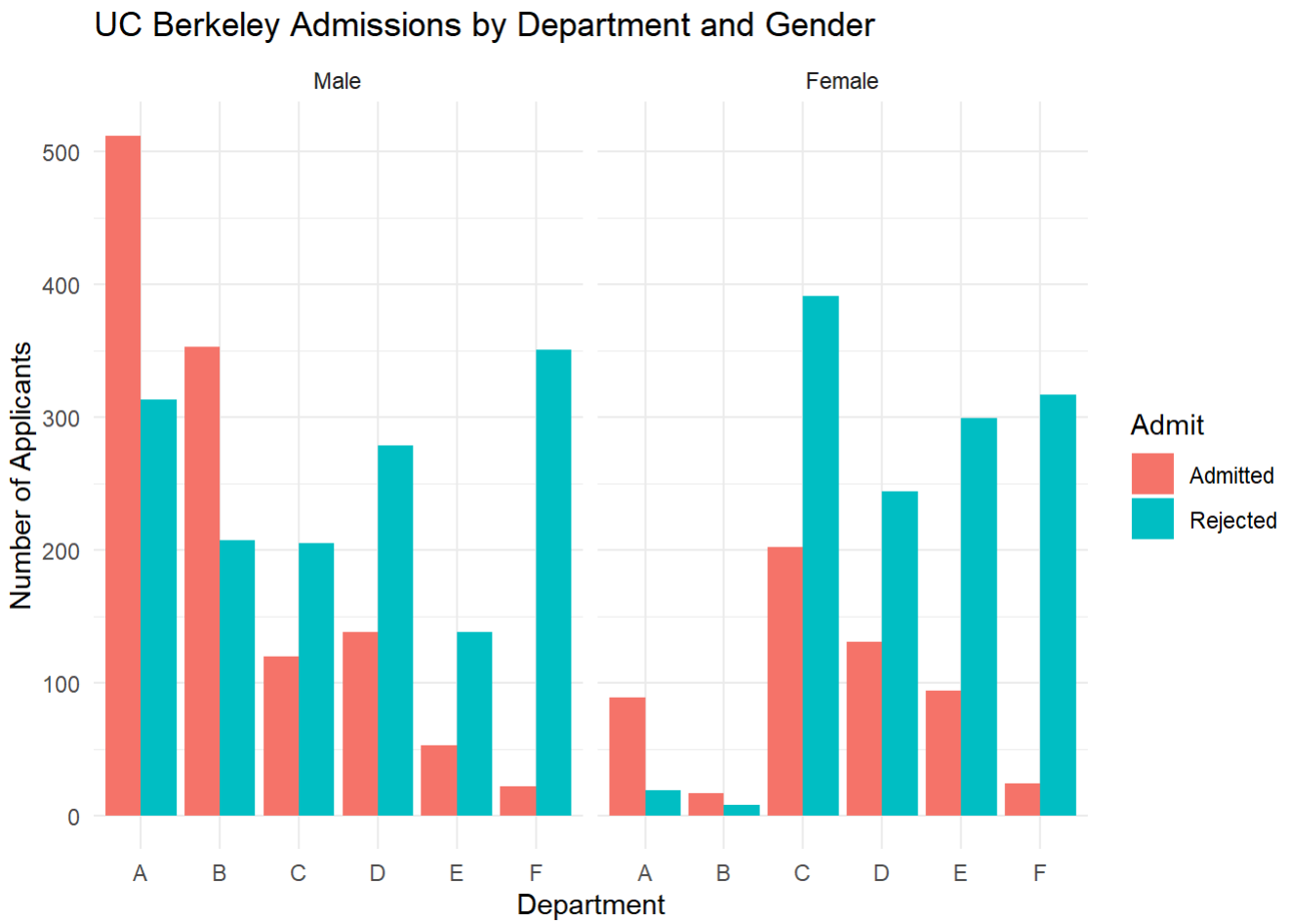
Descriptive Summary

The dataset has a **3-dimensional table**,resulting from cross-tabulating 4526 observations on 3 variables. The variables and their description are as follows:

- **Department**: A–F
- **Gender**: Male or Female
- **Admit Status**: Admitted or Rejected

```
     Admit Gender Dept Freq
1 Admitted   Male    A  512
2 Rejected   Male    A  313
3 Admitted Female    A   89
4 Rejected Female    A   19
5 Admitted   Male    B  353
6 Rejected   Male    B  207
```

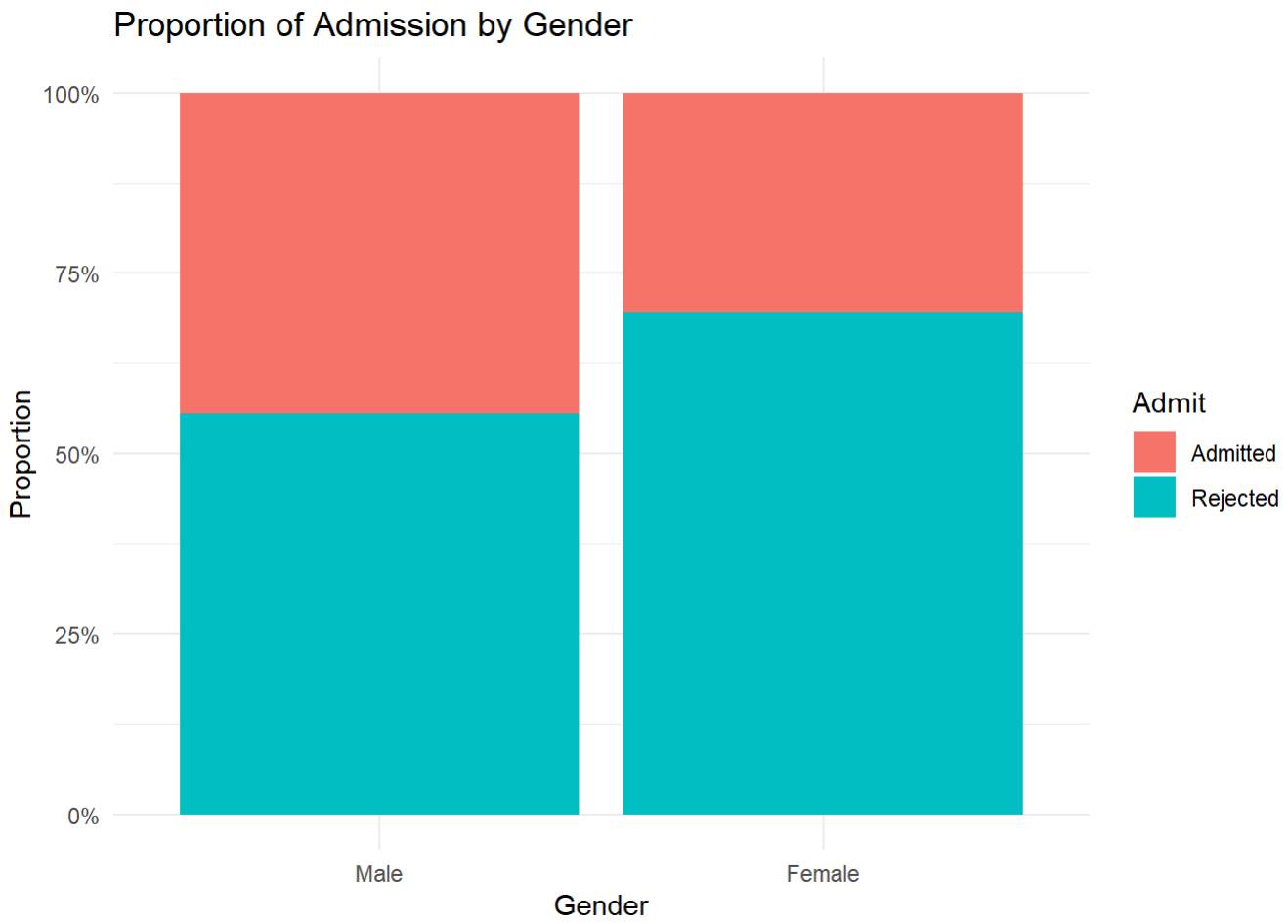
Admission by Department and Gender



- The **bar chart** above shows the *distribution of admitted and rejected applicants across departments, categorised by gender.*
- Initially it appears that males were admitted more often, this breakdown reveals departmental variation in admission rates and gender ratios.
- For instance, in Departments A and B where most male applicants applied,the acceptance rates were relatively high. In contrast, female applicants were more likely to apply to departments with lower overall acceptance rates, such as Departments C, D, and F.
- This reflects a key example of Simpson's Paradox, where aggregated data hides important subgroup trends.
- Without disaggregating the data, it is easy to conclude that *gender bias existed at the university level but differences in application patterns by department played a major role in shaping the overall numbers.*

The above analysis highlights the possibility of aggregate data being misleading at times, and how visual analysis can help uncover the actual information.

Admission Outcomes Across Departments by Gender



- This **proportional bar plot** shows that the *overall admission rate was higher for males than for females.*
- There are various factors that could have influenced this, one such possibility is that the *the lower overall admission rate for females may not be due to direct discrimination, but rather due to differences in application behavior across departments.*
- The above phenomenon is referred to as Simpson's Paradox. It highlights how aggregate statistics can disrupt important subgroup-level insights.