

Lecture 6: Examining Categorical Data

Contingency Tables

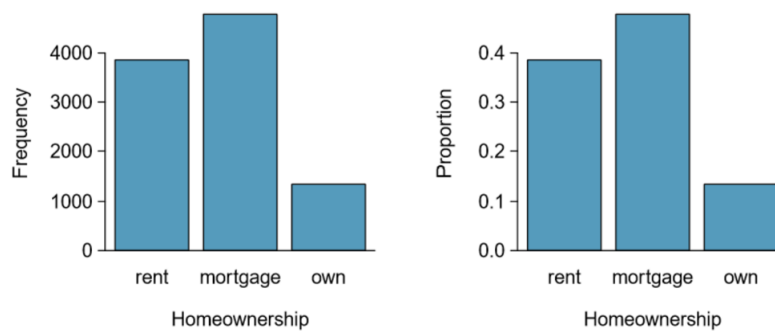
We can use a **contingency table** to break down data with respect to two categorical variables.

| | | homeownership | | | Total |
|----------|------------|---------------|----------|------|-------|
| | | rent | mortgage | own | |
| app_type | individual | 3496 | 3839 | 1170 | 8505 |
| | joint | 362 | 950 | 183 | 1495 |
| | Total | 3858 | 4789 | 1353 | 10000 |

Question of interest: Does there appear to be a relationship between loan application type and homeownership status? We can start to answer this question by examining row proportions:

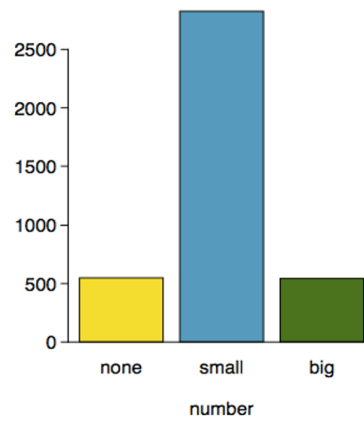
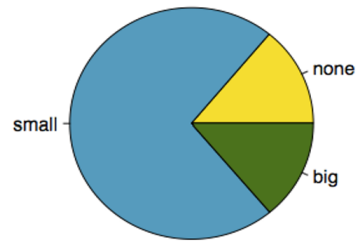
- % of all applicants who rent:
- % of individual applicants who rent:
- % of joint applicants who rent:

Bar Plots



Differences between bar plots and histograms?

Pie Charts



Side-by-Side Plots

