Considering categorical data

Contingency tables

A table that summarizes data for two categorical variables is called a *contingency table*.

Contingency tables

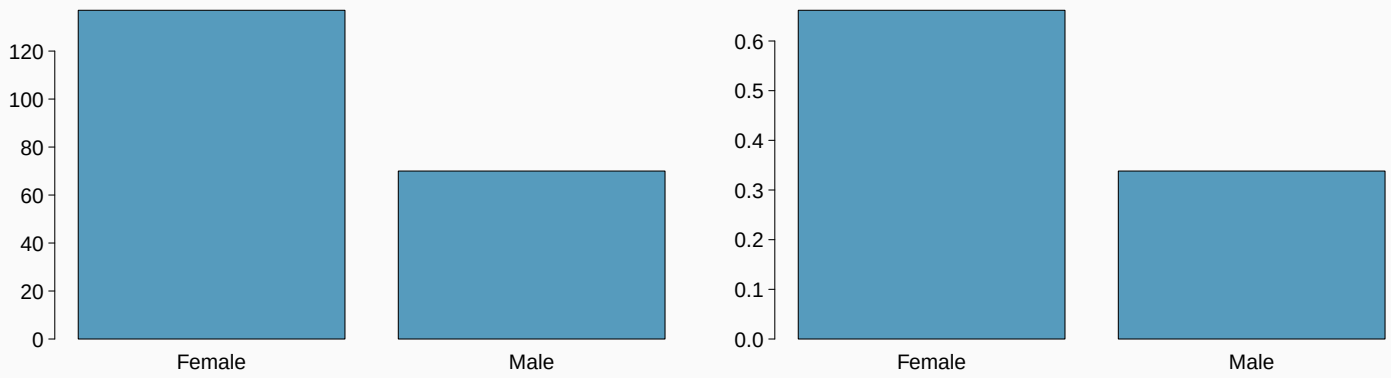
A table that summarizes data for two categorical variables is called a *contingency table*.

The contingency table below shows the distribution of students' genders and whether or not they are looking for a spouse while in college.

		looking for spouse		
		No	Yes	Total
gender	Female	86	51	137
	Male	52	18	70
	Total	138	69	207

Bar plots

A *bar plot* is a common way to display a single categorical variable. A bar plot where proportions instead of frequencies are shown is called a *relative frequency bar plot*.



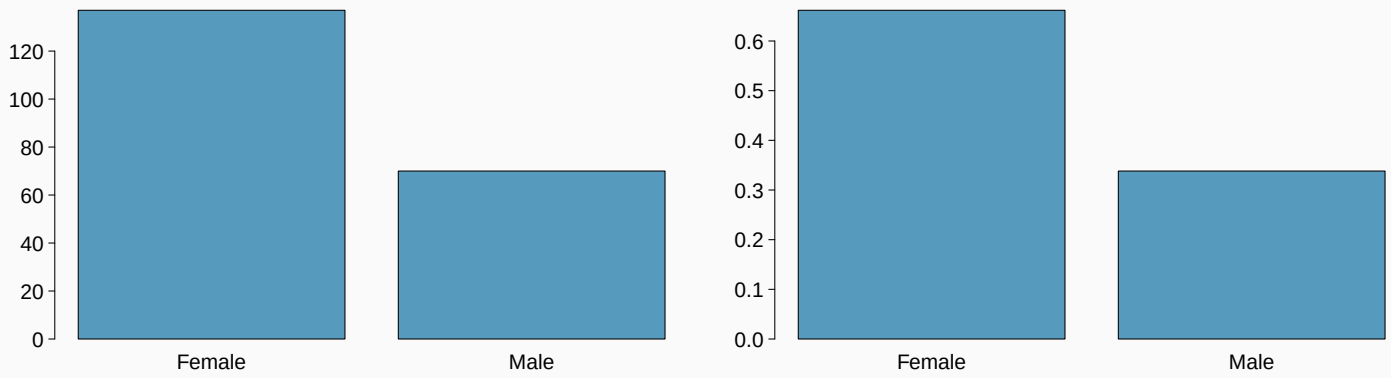
Bar plots

A *bar plot* is a common way to display a single categorical variable. A bar plot where proportions instead of frequencies are shown is called a *relative frequency bar plot*.

How are bar plots different than histograms?

Bar plots

A *bar plot* is a common way to display a single categorical variable. A bar plot where proportions instead of frequencies are shown is called a *relative frequency bar plot*.



How are bar plots different than histograms?

Bar plots are used for displaying distributions of categorical variables, while histograms are used for numerical variables. The x-axis in a histogram is a number line, hence the order of the bars cannot be changed, while in a bar plot the categories can be listed in any order (though some orderings make more sense than others, especially for ordinal variables.)

Choosing the appropriate proportion

Does there appear to be a relationship between gender and whether the student is looking for a spouse in college?

		looking for spouse		
		No	Yes	Total
gender	Female	86	51	137
	Male	52	18	70
	Total	138	69	207

Choosing the appropriate proportion

Does there appear to be a relationship between gender and whether the student is looking for a spouse in college?

		looking for spouse		
		No	Yes	Total
gender	Female	86	51	137
	Male	52	18	70
	Total	138	69	207

To answer this question we examine the row proportions:

Choosing the appropriate proportion

Does there appear to be a relationship between gender and whether the student is looking for a spouse in college?

		looking for spouse		
		No	Yes	Total
gender	Female	86	51	137
	Male	52	18	70
	Total	138	69	207

To answer this question we examine the row proportions:

- % Females looking for a spouse: $51/137 \approx 0.37$

Choosing the appropriate proportion

Does there appear to be a relationship between gender and whether the student is looking for a spouse in college?

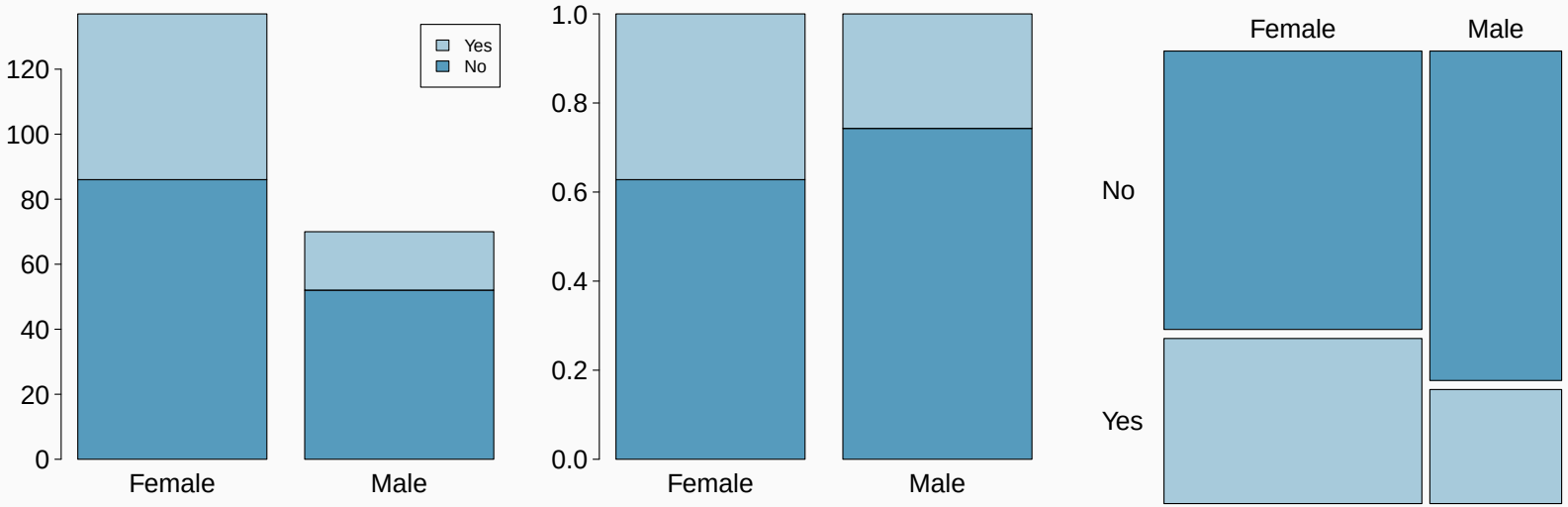
		looking for spouse		
		No	Yes	Total
gender	Female	86	51	137
	Male	52	18	70
	Total	138	69	207

To answer this question we examine the row proportions:

- % Females looking for a spouse: $51/137 \approx 0.37$
- % Males looking for a spouse: $18/70 \approx 0.26$

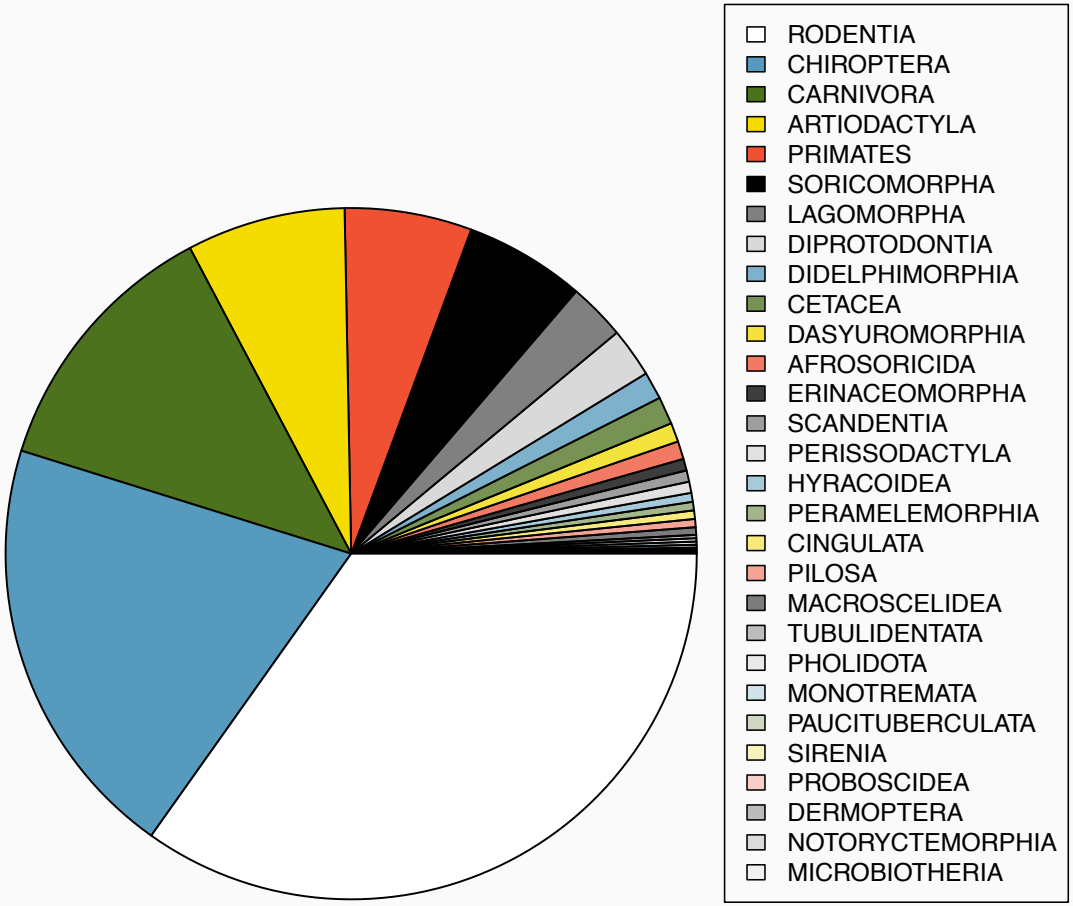
Segmented bar and mosaic plots

What are the differences between the three visualizations shown below?



Pie charts

Can you tell which order encompasses the lowest percentage of mammal species?



Data from *http://www.bucknell.edu/msw3*.

Side-by-side box plots

Does there appear to be a relationship between class year and number of clubs students are in?