Contingency tables

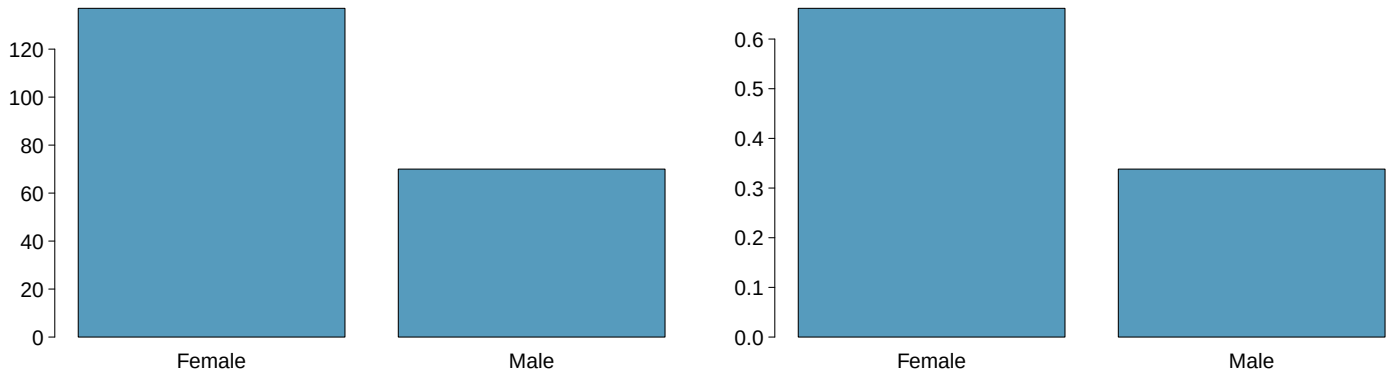
A table that summarizes data for two categorical variables is called a *contingency table*.

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*.



How are bar plots different than histograms?

Choosing the appropriate proportion

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

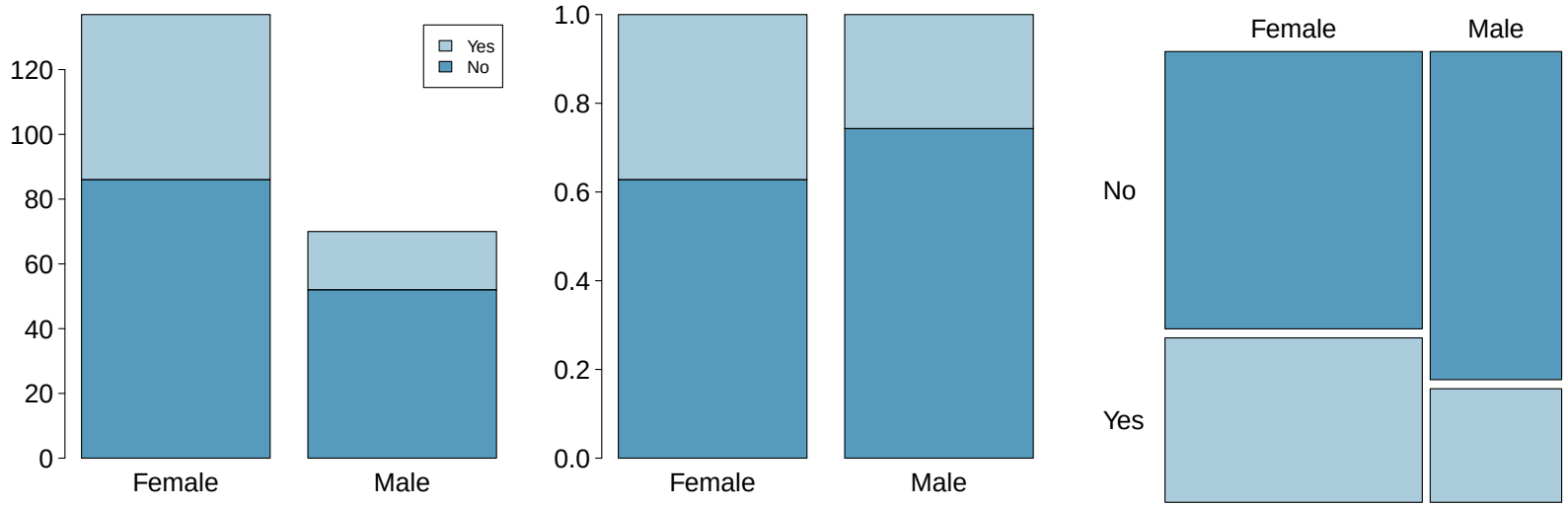
		looking for spouse		
		No	Yes	Total
gender	Female	86	51	137
	Male	52	18	70
	Total	138	69	207

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?



Example: Marbles

Imagine an urn contains 100 marbles, each having a color and a pattern. The colors are green, violet, and blue. The patterns are checkered, dotted, and striped. The frequencies (counts) are shown in the contingency table.

	checkered	dotted	striped	total
green	20	18	21	59
violet	8	9	3	20
blue	12	3	6	21
total	40	30	30	100

There are a variety of interesting proportions: simple part of whole, compound part of whole, and part of part.

Marbles - Simple (Marginal) Proportions

	checkered	dotted	striped	total
green	20	18	21	59
violet	8	9	3	20
blue	12	3	6	21
total	40	30	30	100

Simple part of whole:

- What proportion of marbles are checkered?
- What proportion of marbles are green?
- What proportion of marbles are striped?
- What proportion of marbles are blue?

Marbles - Compound Proportions

	checkered	dotted	striped	total
green	20	18	21	59
violet	8	9	3	20
blue	12	3	6	21
total	40	30	30	100

Compound part of whole:

- What proportion of marbles are dotted and violet?
- What proportion of marbles are green and striped?
- What proportion of marbles are green or striped?
- What proportion of marbles are dotted or violet?

Marbles - Conditional Proportions

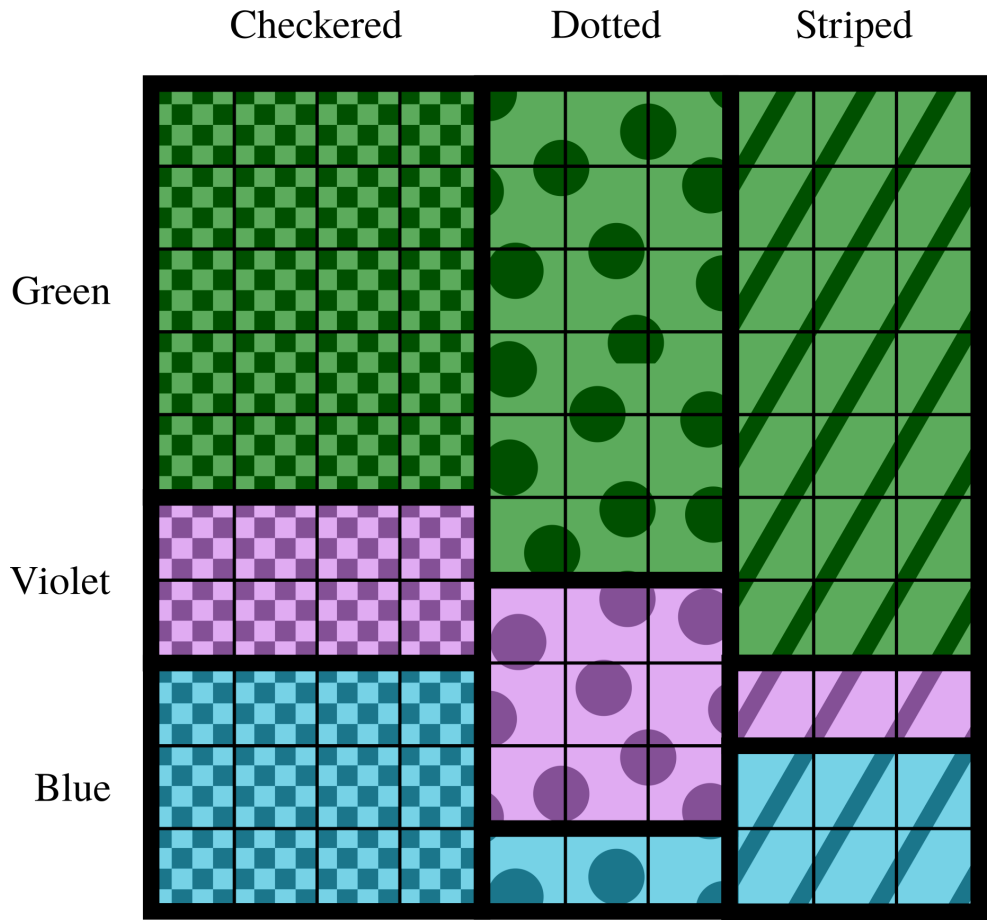
	checkered	dotted	striped	total
green	20	18	21	59
violet	8	9	3	20
blue	12	3	6	21
total	40	30	30	100

Part of part:

- What proportion of violet marbles are dotted?
- What proportion of dotted marbles are violet?
- What proportion of green marbles are striped?
- What proportion of striped marbles are green?

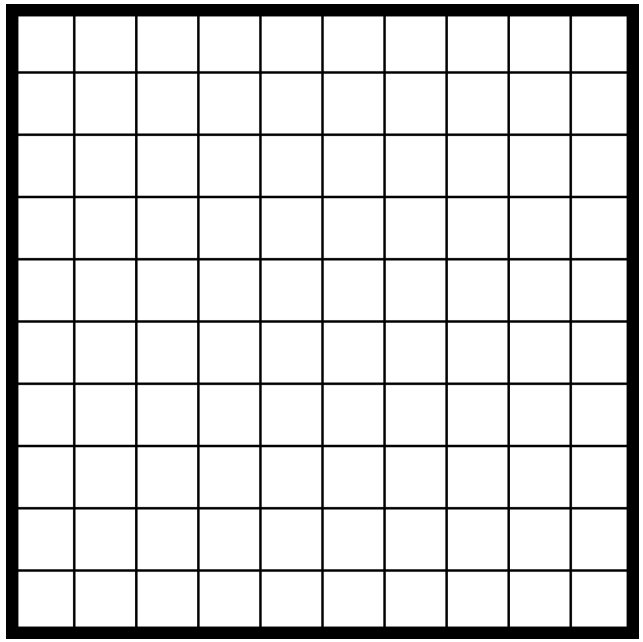
Marbles - Mosaic Plot

	checkered	dotted	striped	total
green	20	18	21	59
violet	8	9	3	20
blue	12	3	6	21
total	40	30	30	100



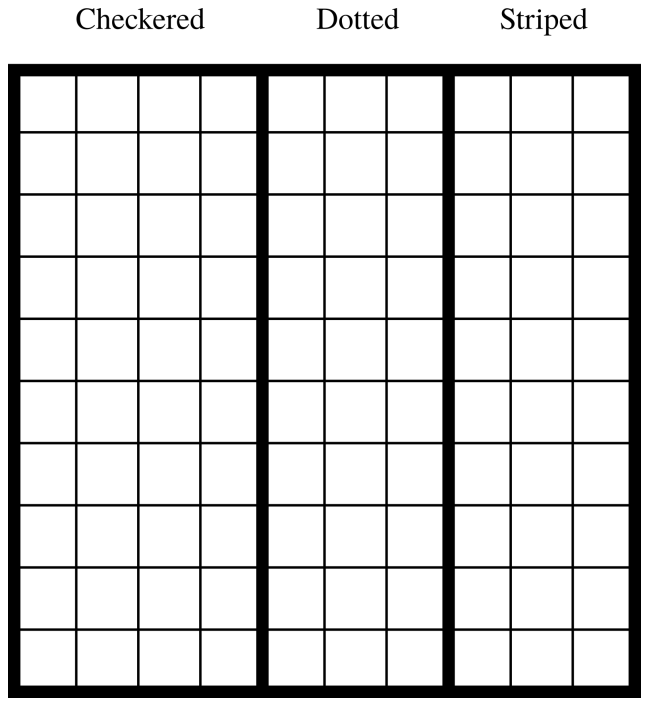
Marbles - Making a Mosaic Plot

1. Start with a rectangle (a square is good).



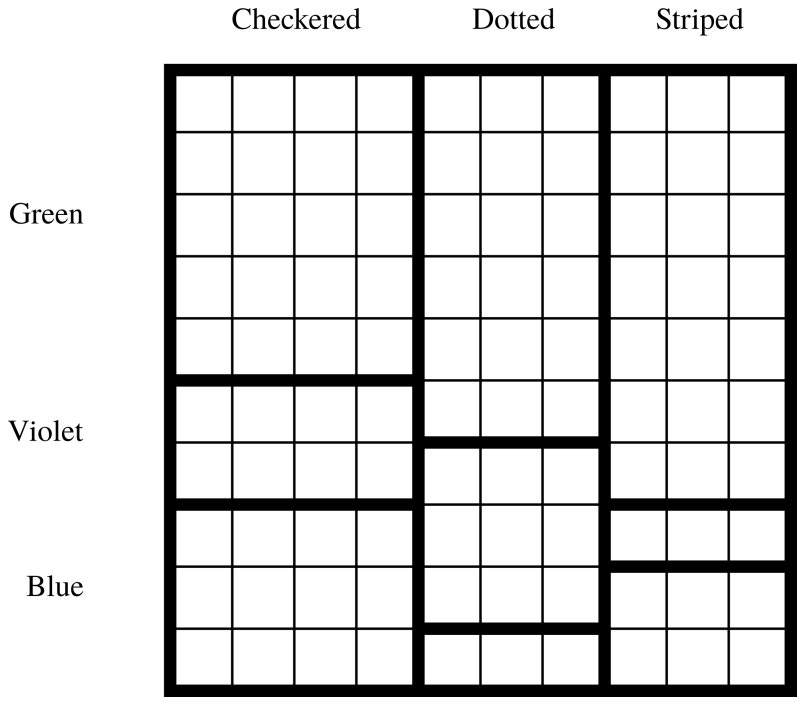
1. Start with a rectangle (a square is good).
2. Find marginal proportions of first variable for widths of columns.

$$\frac{40}{100} = 0.4,\ \frac{30}{100} = 0.3,\ \frac{30}{100} = 0.3$$

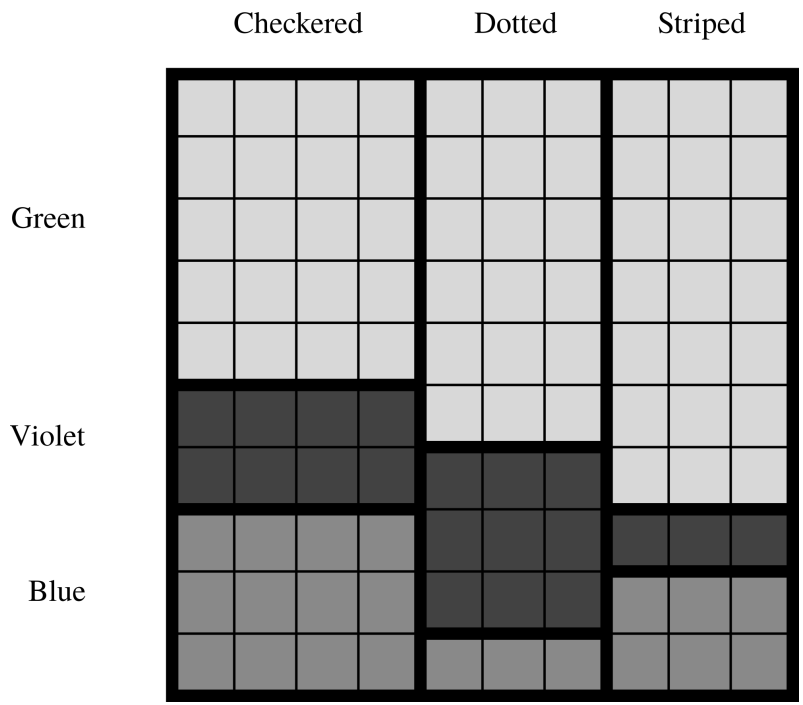


1. Start with a rectangle (a square is good).
2. Find marginal proportions of first variable for widths of columns.
3. For heights of rectangles, determine part-of-part proportions, e.g. the proportion of checkered marbles that are green.

$$\frac{20}{40} = 0.5 \quad \frac{18}{30} = 0.6 \quad \frac{21}{30} = 0.7$$
$$\frac{8}{40} = 0.2 \quad \frac{9}{30} = 0.3 \quad \frac{3}{30} = 0.1$$
$$\frac{12}{40} = 0.3 \quad \frac{3}{30} = 0.1 \quad \frac{6}{30} = 0.2$$



1. Start with a rectangle (a square is good).
2. Find marginal proportions of first variable for widths of columns.
3. For heights of rectangles, determine part-of-part proportions, e.g. the proportion of checkered marbles that are green.
4. Shade according to second variable.



Independence in Mosaic Plots

If the two variables are independent, then a mosaic plot's horizontal lines will be continuous (unbroken).