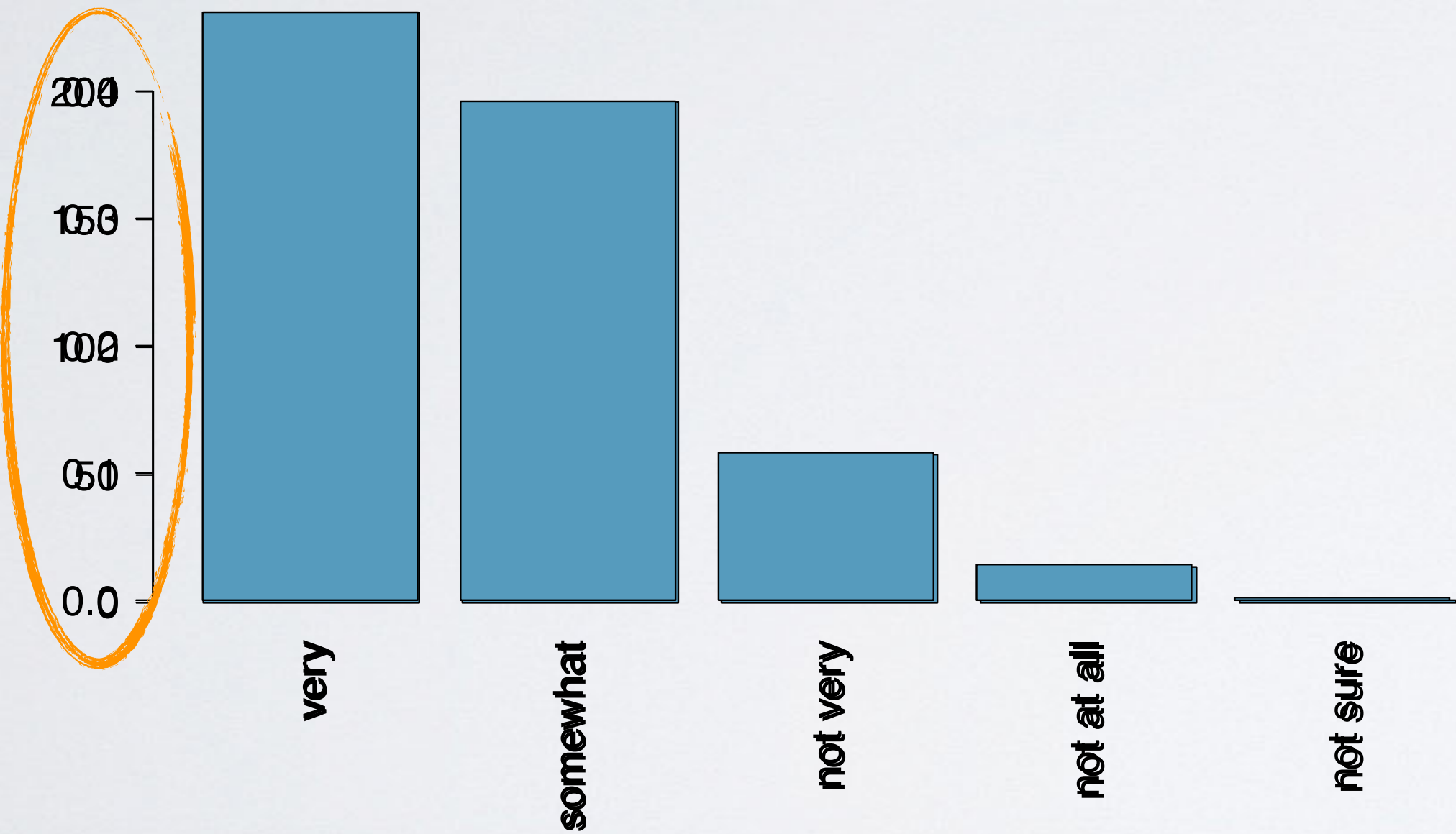


exploring categorical variables

- ▶ describe distribution of a single categorical variable
- ▶ evaluate relationship between two categorical variables
- ▶ evaluate relationship between a categorical and a numerical variable

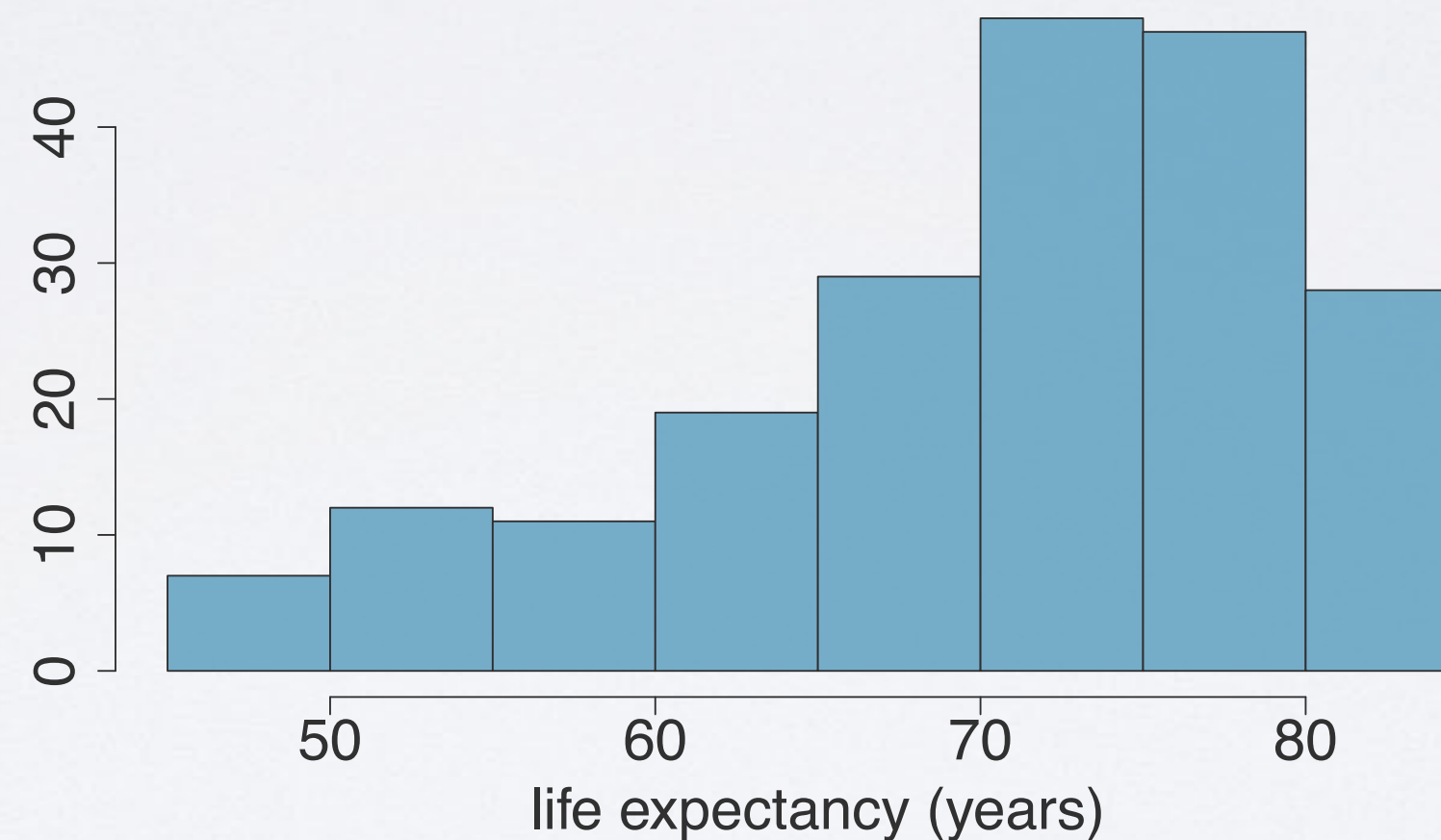
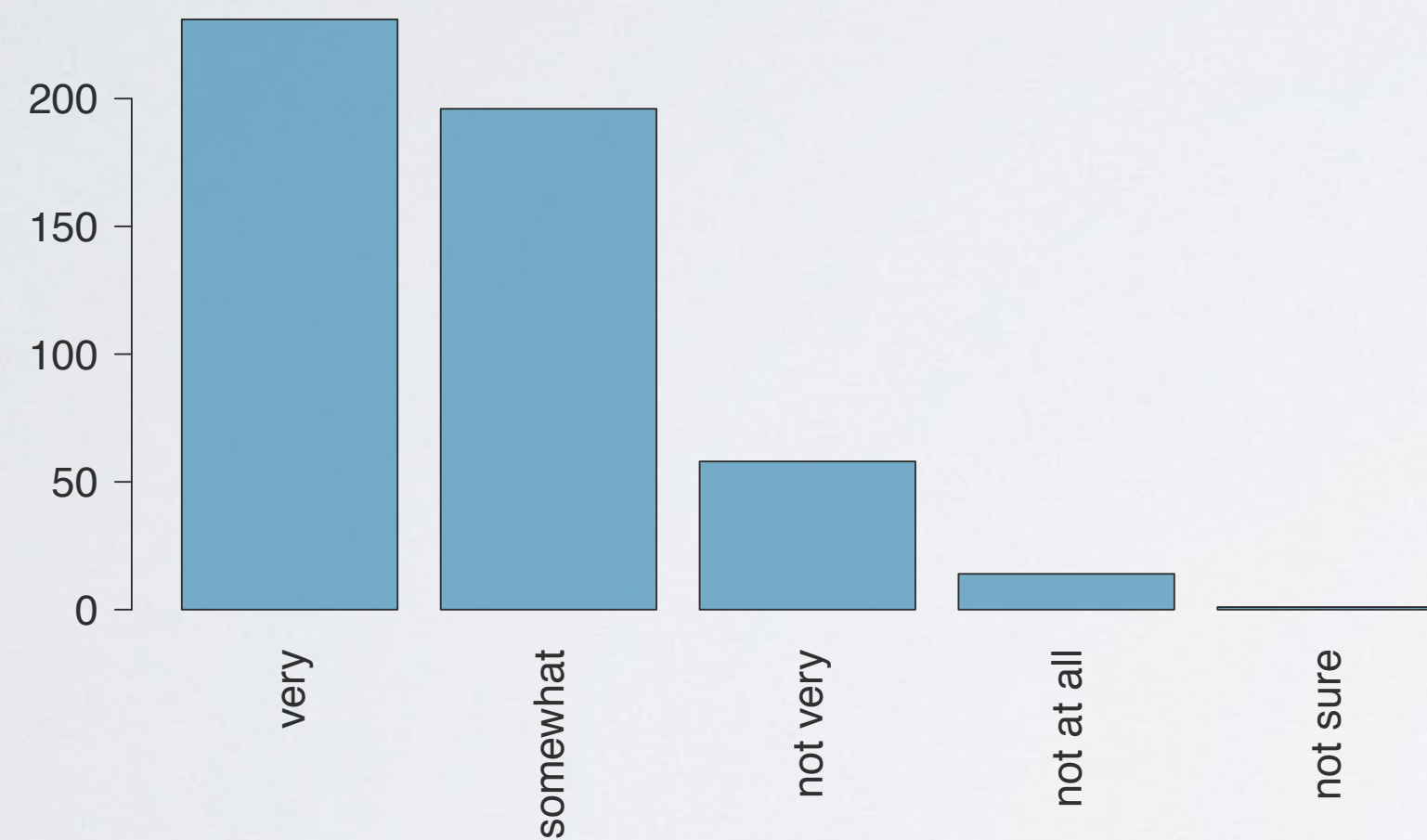
frequency table & bar plot

Difficulty saving money	Counts	Frequencies
Very	231	46%
Somewhat	196	39%
Not very	58	12%
Not at all	14	3%
Not sure	1	~0%
Total	500	100%



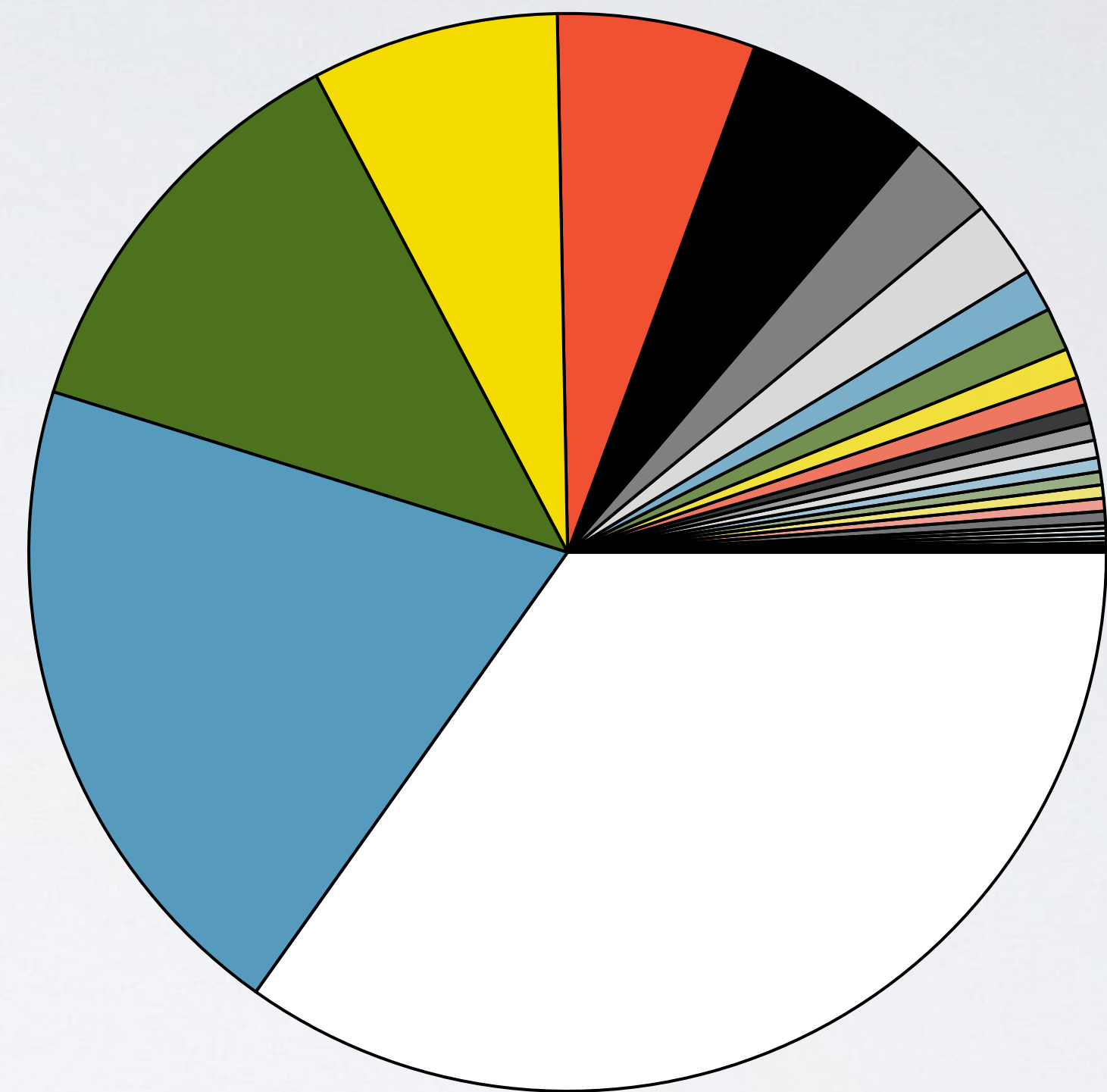
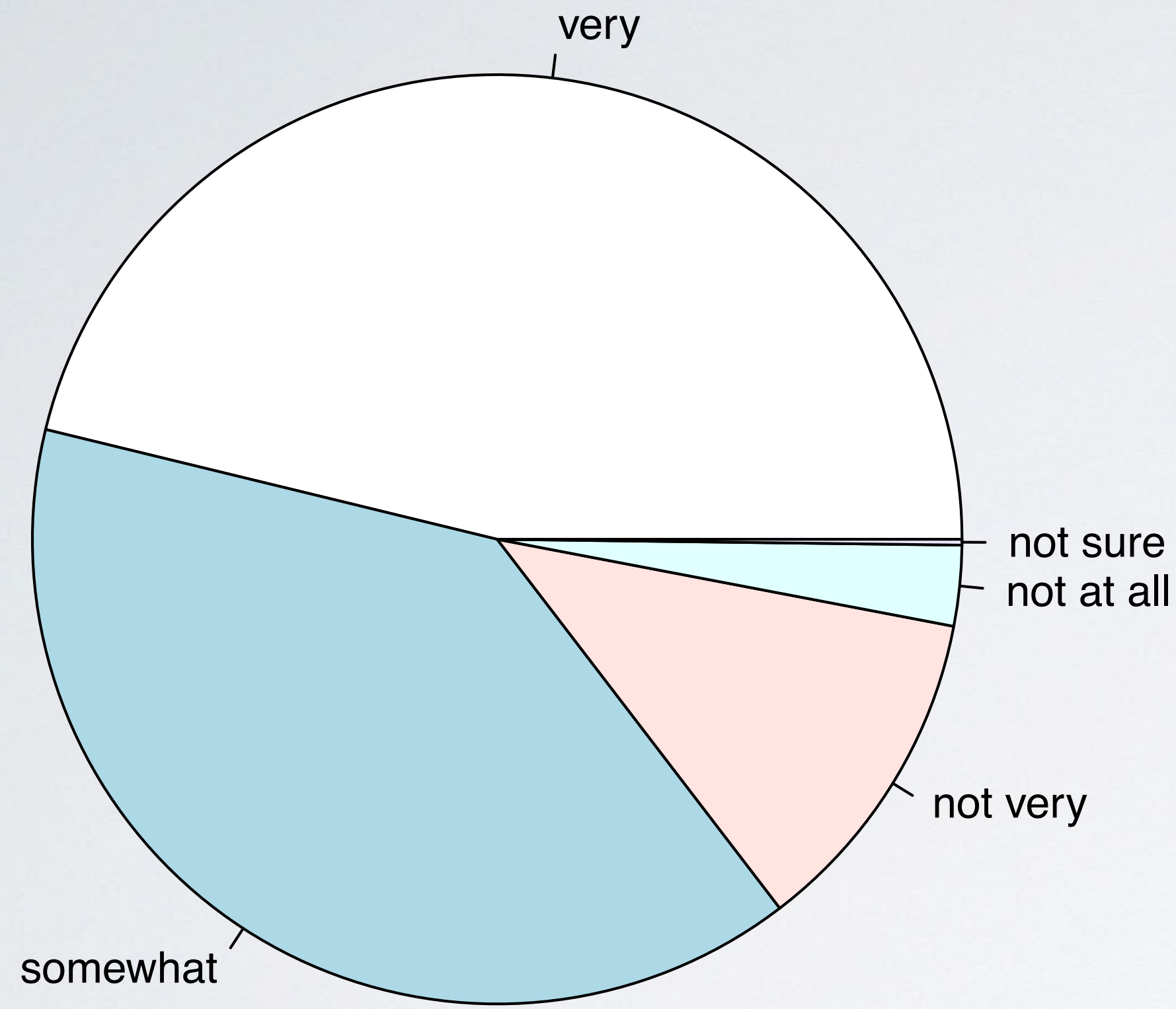
How are bar plots different than histograms?

- ▶ barplots for categorical variables, histograms for numerical variables
- ▶ x-axis on a histogram is a number line, and the ordering of the bars are not interchangeable



~~pie chart?~~

no!



- RODENTIA
- CHIROPTERA
- CARNIVORA
- ARTIODACTYLA
- PRIMATES
- SORICOMORPHA
- LAGOMORPHA
- DIPROTODONTIA
- DIDELPHIMORPHIA
- CETACEA
- DASYUROMORPHIA
- AFROSORICIDA
- ERINACEOMORPHA
- SCANDENTIA
- PERISSODACTYLA
- HYRACOIDEA
- PERAMELEMORPHIA
- CINGULATA
- PILOSA
- MACROSCELIDEA
- TUBULIDENTATA
- PHOLIDOTA
- MONOTREMATA
- PAUCITUBERCULATA
- SIRENIA
- PROBOSCIDEA
- DERMOPTERA
- NOTORYCTEMORPHIA
- MICROBIOTHERIA

contingency table

		Income				
		< \$40K	\$40-80K	> \$80K	Refused	Total
Difficulty saving	Very	128	63	31	9	231
	Somewhat	54	71	61	10	196
	Not very	17	7	27	7	58
	Not at all	3	6	5	0	14
	Not sure	0	1	0	0	1
	Total	202	148	124	26	500

relative frequencies

		Income				
		< \$40K	\$40K - \$80K	> \$80K	Refused	Total
Difficulty saving	Very	128	63	31	9	231
	Somewhat	54	71	61	10	196
	Not very	17	7	27	7	58
	Not at all	3	6	5	0	14
	Not sure	0	1	0	0	1
Total		202	148	124	26	500

< \$40K: $128 / 202 = 63\%$ find it very difficult to save

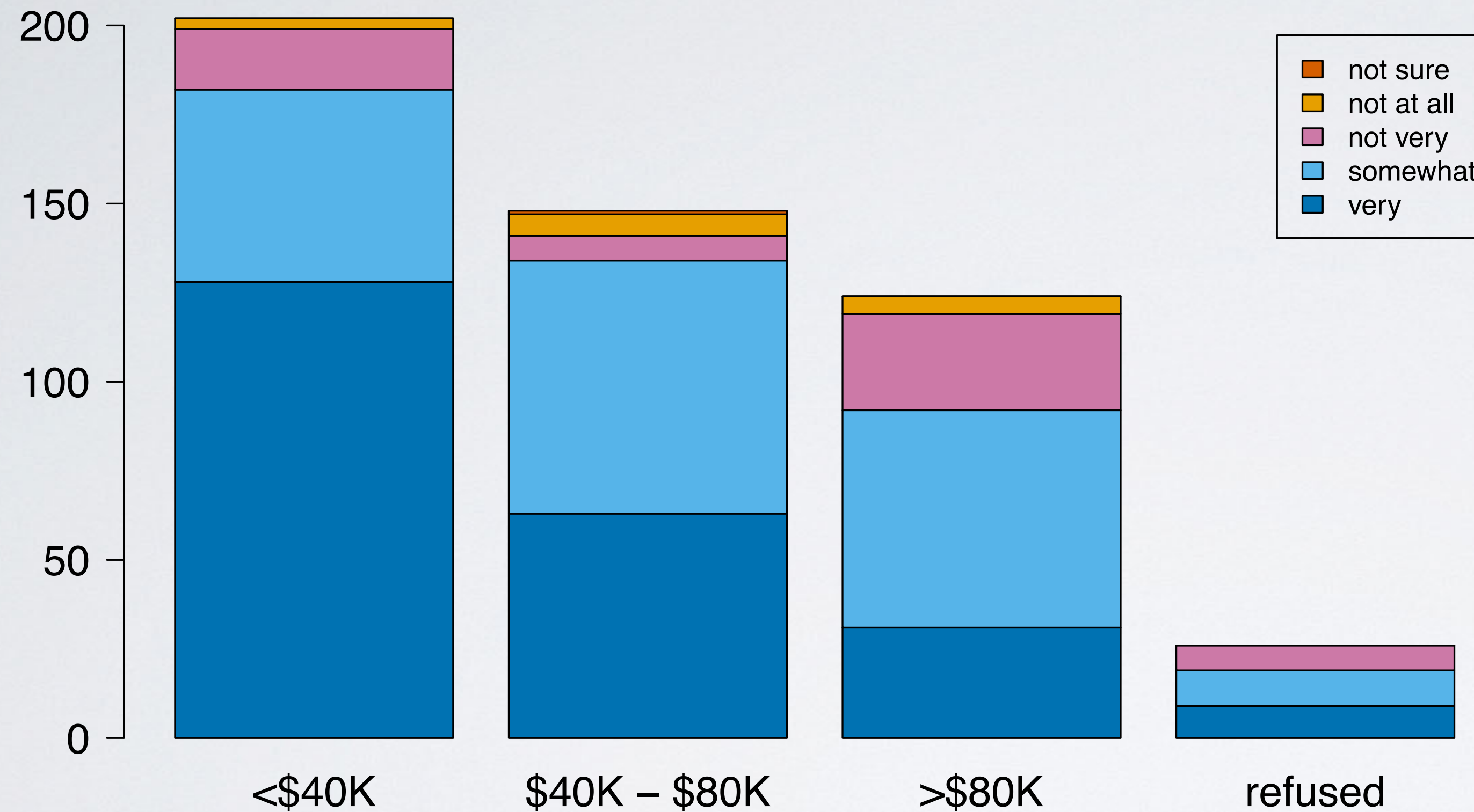
\$40K-\$80K: $63 / 148 = 43\%$

> \$80K: $31 / 124 = 25\%$

Refused: $9 / 26 = 35\%$

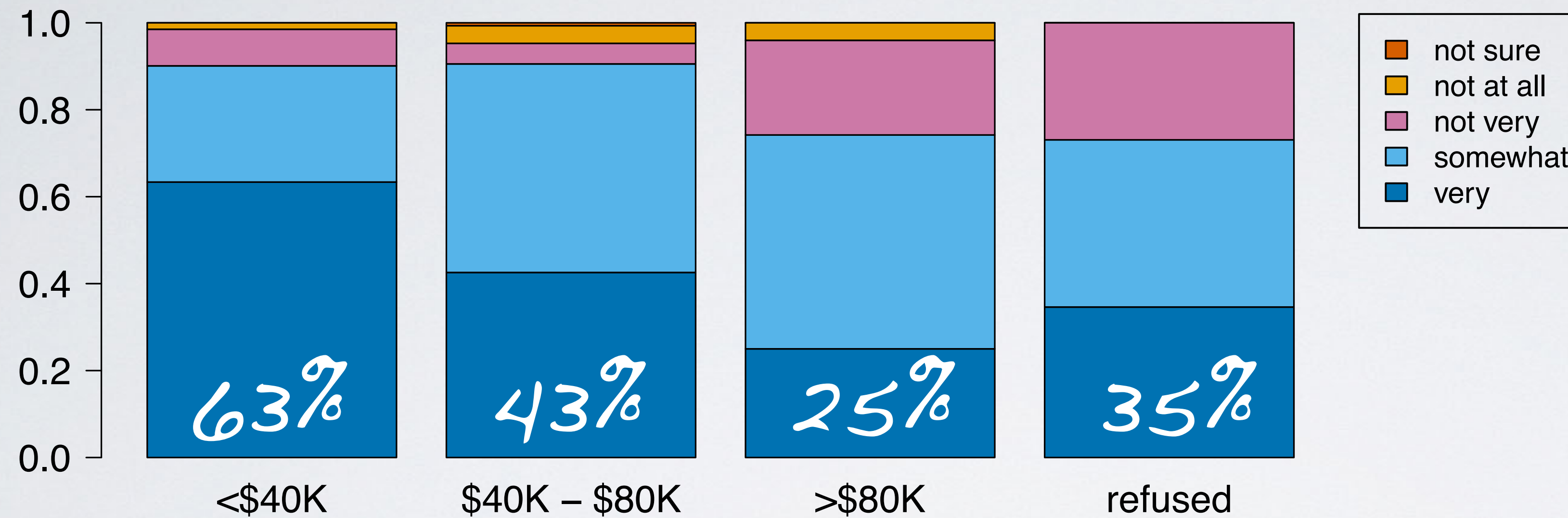
feelings about difficulty of saving money and income are associated (dependent)

segmented bar plot

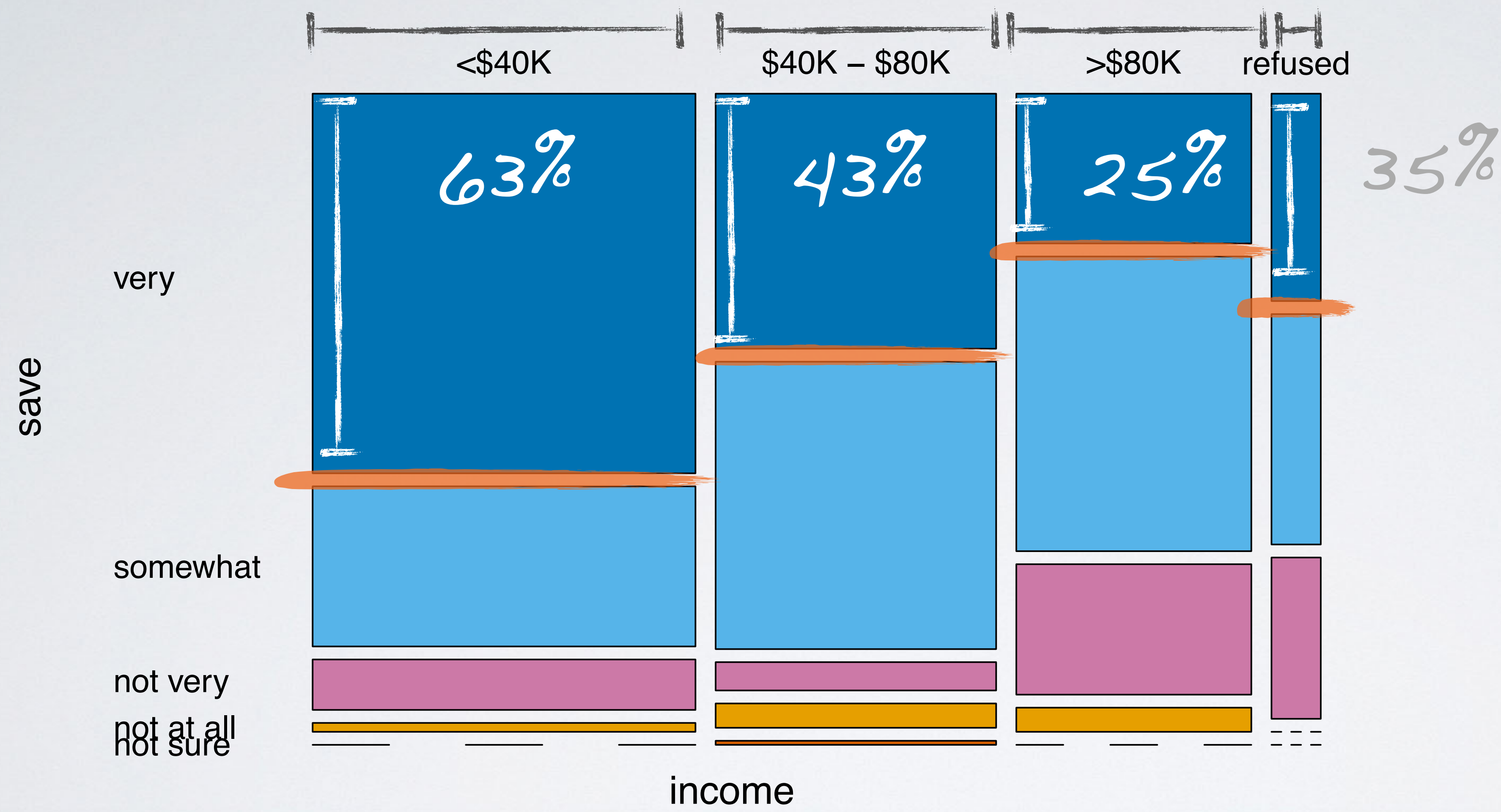


- ▶ useful for visualizing conditional frequency distributions
- ▶ compare relative frequencies to explore the relationship between the variables

relative frequency segmented bar plot



mosaicplot



side-by-side box plots

