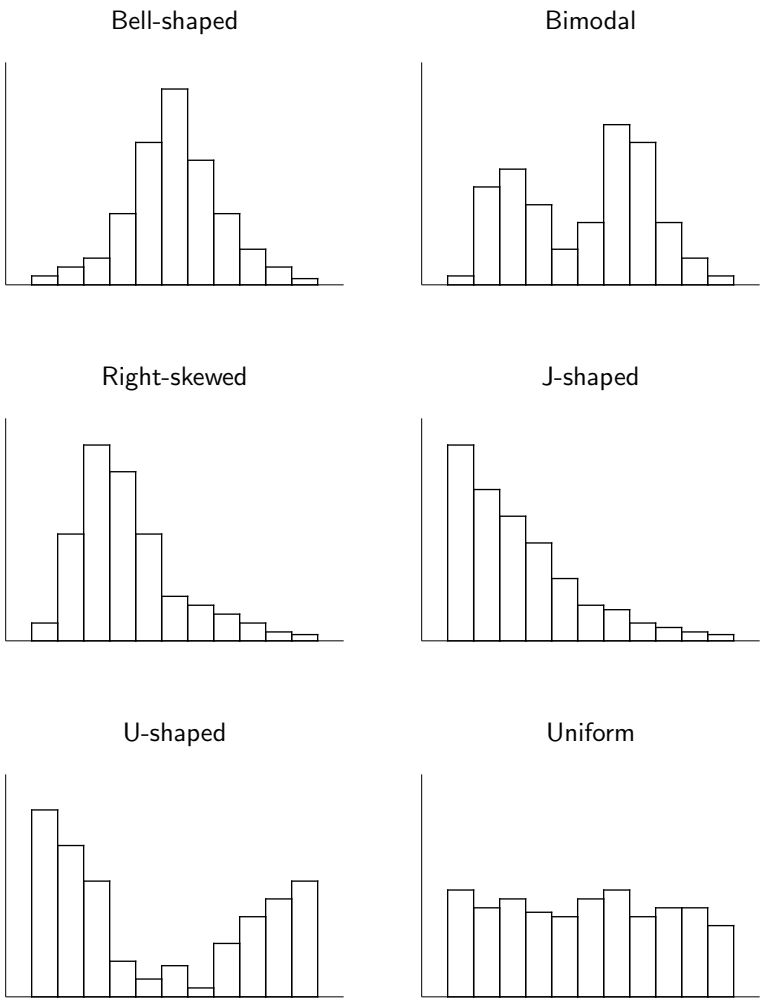
DESCRIBING VARIABILITY USING HISTOGRAMS

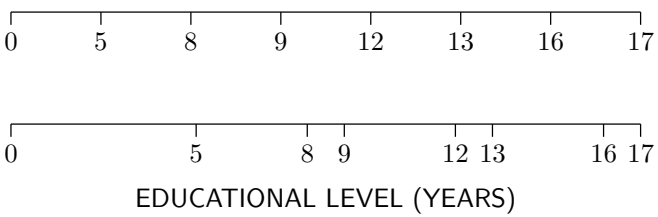


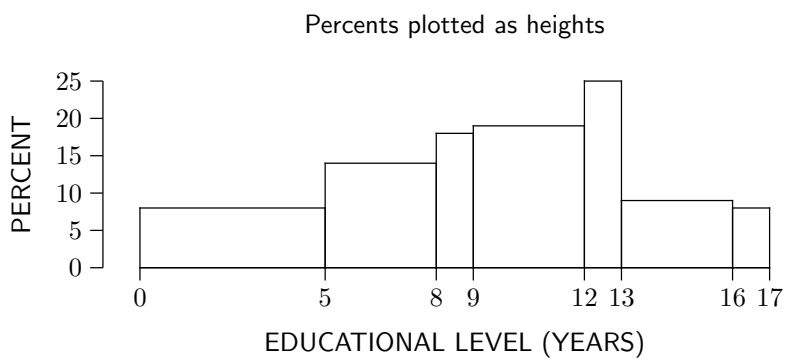
DRAWING A HISTOGRAM

- Years of schooling completed for persons age 25 and over in the U.S. in 1960. The class intervals include the left endpoints, but not the right.

Educational Level	*Percent*
0 – 5	8
5 – 8	14
8 – 9	18
9 – 12	19
12 – 13	25
13 – 16	9
16 or more	8

- Let's try drawing a histogram for the distribution of educational level by year:

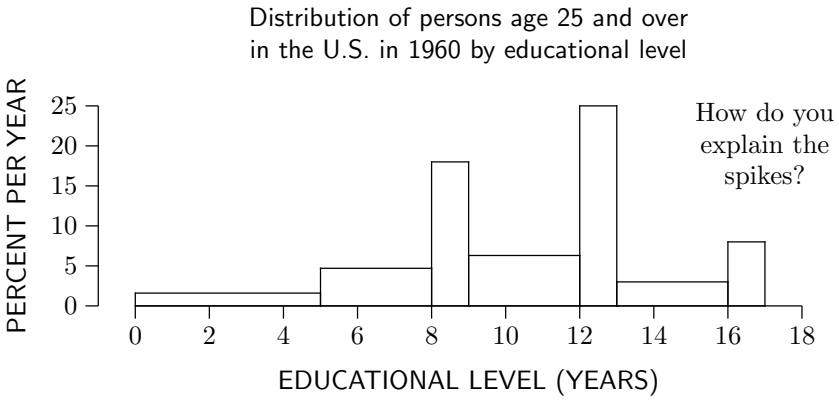




- Are there more people with 9 to 12 years of schooling, or with 12 to 13 years?

• What should the height of the block over the interval from 9 to 12 years be?

$$\frac{19\%}{3 \text{ years}} = 6\tfrac{1}{3}\% \text{ per year.}$$



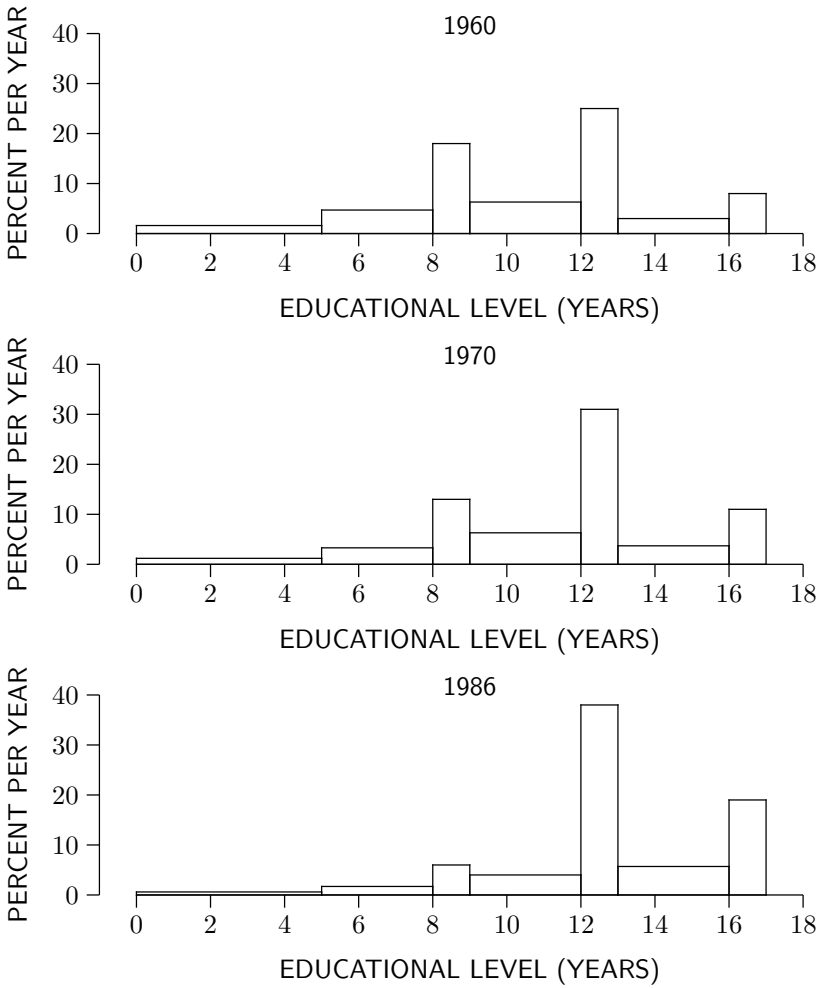
• What is the area of the block over the interval from 9 to 12 years?

$$\begin{aligned}\text{Area} &= \text{Width} \times \text{Height}\\ &= (3 \text{ years}) \times 6\tfrac{1}{3}\% \text{ per year}\\ &= 19\%.\end{aligned}$$

• Remember:

- A histogram represents percents by _____. The area of a block represents the percentage of cases in the corresponding class interval.
- The height of each block equals the percentage of cases in the corresponding class interval, divided by the length of that interval.

USING HISTOGRAMS FOR COMPARISONS



• Which block represents the percentage of the population which completed college? How has that percentage changed over time?

SMOKING

• In a Public Health Service study, a histogram was plotted showing the number of cigarettes smoked per day by each subject (male current smokers), as shown below. The density is marked in parentheses. The class intervals include the right endpoint, but not the left.

• The percentage who smoked 10 cigarettes or less per day is around

1.5% 15% 30% 50%

• The percentage who smoked more than a pack a day, but not more than 2 packs, is around

1.5% 15% 30% 50%

(There are 20 cigarettes in a pack.)

• The percent who smoked more than a pack a day is around

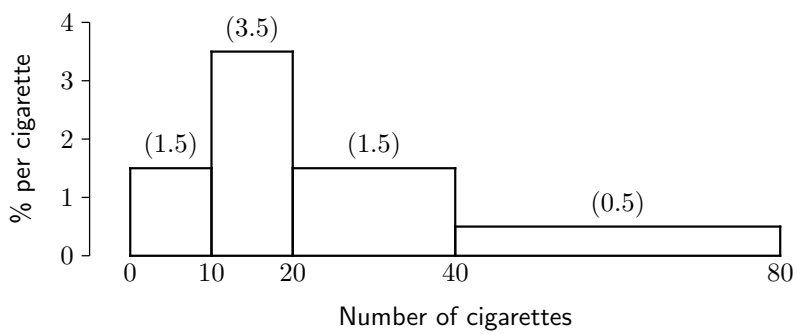
1.5% 15% 30% 50%

• The percent who smoked more than 3 packs a day is around

0.25 of 1% 0.5 of 1% 10%

• The percent who smoked 20 cigarettes a day is around

0.35 of 1% 0.5 of 1% 1.5% 3.5% 10%



WAGES

• A histogram of monthly wages for part-time employees is shown below. Nobody earned more than $1,000 a month. The block over the class interval from $200 to $500 is missing. How tall must it be?

• Area of the missing block = _____.

$$\begin{aligned}\text{Missing Area} &= \text{Total Area} - \text{Other Areas}\\ &= 100\% - (10\% + 20\% + 25\%)\\ &= 100\% - 55\% = 45\%.\end{aligned}$$

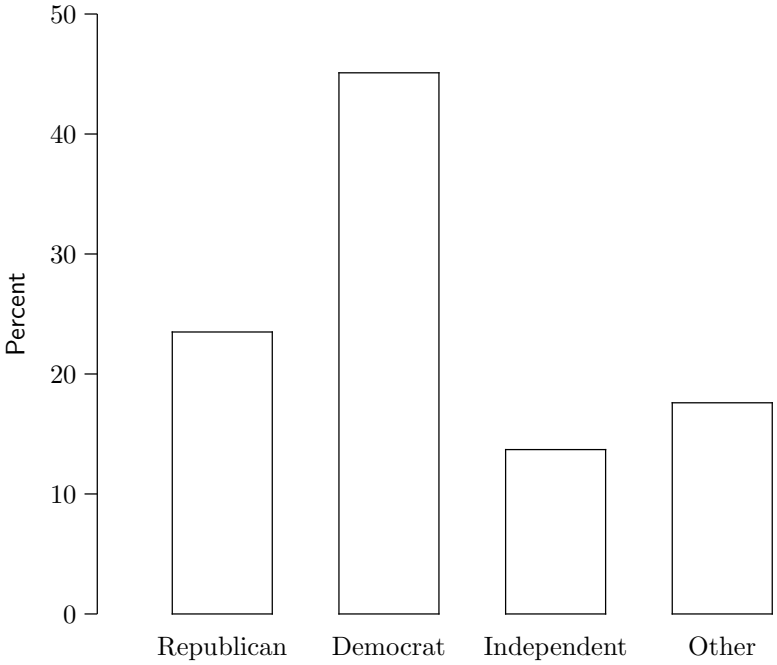
• Height of the missing block = _____.

$$\text{Height} = \frac{\text{Area}}{\text{Width}} = \frac{45\%}{\$300} = 15\% \text{ per } \$100.$$

• Remember: the total area of a histogram is _____.

BAR CHARTS

• The following chart shows the percentage breakdown of the students in Statistics 200 for Spring quarter 1997 by Political Affiliation:



• When is it appropriate to use a histogram to display the distribution of a variable, and when a bar chart?

Type of variable	*Method of display*
Quantitative	Histogram
Qualitative	Bar chart

AN ALTERNATIVE TO HISTOGRAMS: THE STEM-AND-LEAF PLOT

• The following display is a "stem-and-leaf" plot of the list of differences 200, 80, 0, −50, −120, . . . , 190, 50

Verbal SAT score − Math SAT score

for students in this course.

```
 −1∗ | 22
   • | 97555
 −0∗ | 2211
  0∗ | 000111122234
   • | 55556678
  1∗ | 114
   • | 7999
  2∗ | 0

Leaf unit = 10
  1∗ | 4 = 140
```

Numbers to the left of the | are "stems"

Digits to the right of the | are "leaves"

Diff	*Stem*	*Leaf*
200	2	0
80	0	8
0	0	0
−50	−0	5
−120	−1	2
⋮	⋮	⋮
190		
50		

• Two differences, 190 and −40, have not yet been plotted; where do they go?

• Do the students tend to have higher Verbal scores, or higher Math scores?

• How are stem-and-leaf plots and histograms similar/dissimilar? What are the advantages/disadvantages of stem-and-leaf plots relative to histograms?

SAT SCORES

- The following data set lists the Verbal and Math SAT scores for the students in this course. (*You do not need to copy these lists.)*

Verbal SAT scores

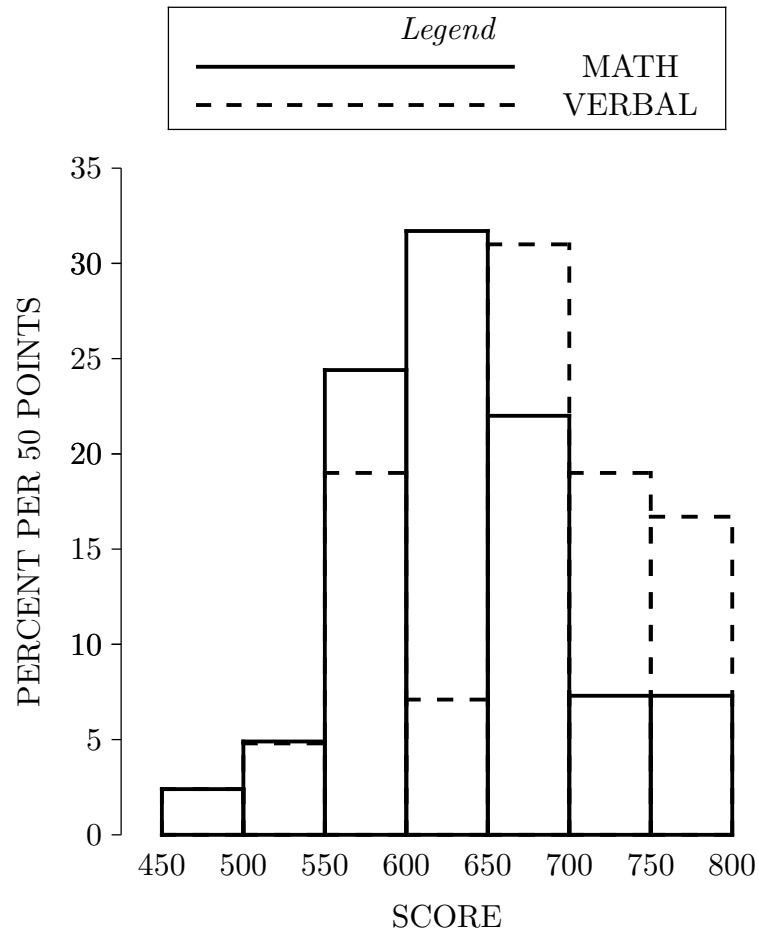
800 760 800 550 680 600 660 730 740 600
700 500 680 710 600 710 700 700 560 650
680 530 660 590 780 700 800 570 600 600
800 710 690 630 720 800 670 710 660 750
680

Math SAT scores

600 680 800 600 800 650 780 670 650 650
560 500 700 700 600 600 680 710 580 720
620 515 640 600 590 650 630 610 590 550
610 640 660 720 680 690 650 700 610 560
630

- How do the two sets of scores compare?

MATH AND VERBAL SAT SCORES
STATISTICS 200, SPRING QUARTER, 1997
Class intervals contain the right endpoint



SKETCHES

• Six sketches are shown below. Four of them are histograms for the following variables in a small town:

(a) heights of all members of households where both parents are less than 24 years old;
(b) heights of married couples;
(c) heights of all people;
(d) heights of all automobiles.

Match the variables with their histograms. Explain your reasoning.