Mix and Match

The following items describe a column in a data table. On the basis of the description, assign a one- or two-word name to the variable, indicate its type (categorical, ordinal, or numerical), and describe the cases. If the data are a time series, identify the frequency of the data. The first item is done as an example.

	Variable Name	Type			Cases
		Nom	Ord	Num	
Weights of people, measured in pounds	Weight			✓	People
1. Brand of car owned by drivers					
2. Income of households, measured in dollars					
3. Color preference of consumers in a focus group					
4. Counts of customers who shop at outlets of achain of retail stores					
5. Size of clothing items given as Small, Medium, Large, and Extra large					
6. Shipping costs in dollars for deliveries from acatalog warehouse					
7. Prices of various stocks, in dollars per share					
8. Number of employees absent from work each day					
9. Sex (male or female) of respondents in a survey					
10. Education of customer, recorded as High school,Some college, or College Graduate					

Mix and Match

Match the brief description to the item or value.

|		
1. Position of a peak in the histogram	(a)	median
2. Half of the cases are smaller than this value	(b)	mean
3. Length of the box in the boxplot	(c)	standard deviation
4. Proportion of cases lying within the box of the boxplot	(d)	variance
5. A histogram with a long right tail	(e)	z-score
6. The average of the values of a numerical variable	(f)	two-thirds
7. The average squared deviation from the average	(g)	mode
8. The square root of the variance	(h)	interquartile range
9. The number of standard deviations from the mean	(i)	one half
10. Proportion of a bell-shaped distribution within 1 SD of $\bar{y}$	(j)	skewed

Think About It

23. If the median size used by 500 songs is 3.5 MB, will these all fit on a Shuffle that has 2 GB of storage? Can you tell?

24. (a) If the average size of 500 songs is 5.1 MB with standard deviation 0.5 MB, will these fit on a Shuffle that has 2 GB of storage?
(b) Does the size of the standard deviation matter in answering part (a)?

25. Suppose you were looking at the histogram of the incomes of all of the households in the United States. Do you think that the histogram would be bell shaped? Skewed to the left or right?

26. For the incomes of all of the households in the United States, which do you expect to be larger, the mean or the median?

27. An adjustable rate mortgage allows the rate of interest to fluctuate over the term of the loan, depending on economic conditions. A fixed rate mortgage holds the rate of interest constant. Which sequence of monthly payments has smaller variance, those on an adjustable rate mortgage or those on a fixed rate mortgage?

32. Which has larger variance, the prices of items in a grocery or the prices of cars at a new car dealer? Which has the larger coefficient of variation?

33. In a study of how the environment affects spending, researchers varied the presence of music and scents in a shopping mall. Interviews with shoppers who made unplanned purchases revealed the following average purchase amounts. Each group has about 100 shoppers.[18]

Condition	Mean Spending ($)
None	64.00
Scent	55.34
Music	96.89
Scent and music	36.70

(a) If each group includes 100 shoppers, which group spent the most money in toto?

(b) If medians summarized each group's spending rather than means, would you be able to identify the group with the largest spending?

(c) Did every shopper who was exposed to scent and music spend less than every shopper who was only exposed to music?

42. The following display summarizes the percentage of household income that goes to rent for 1,000 families that live in Denver, Colorado.
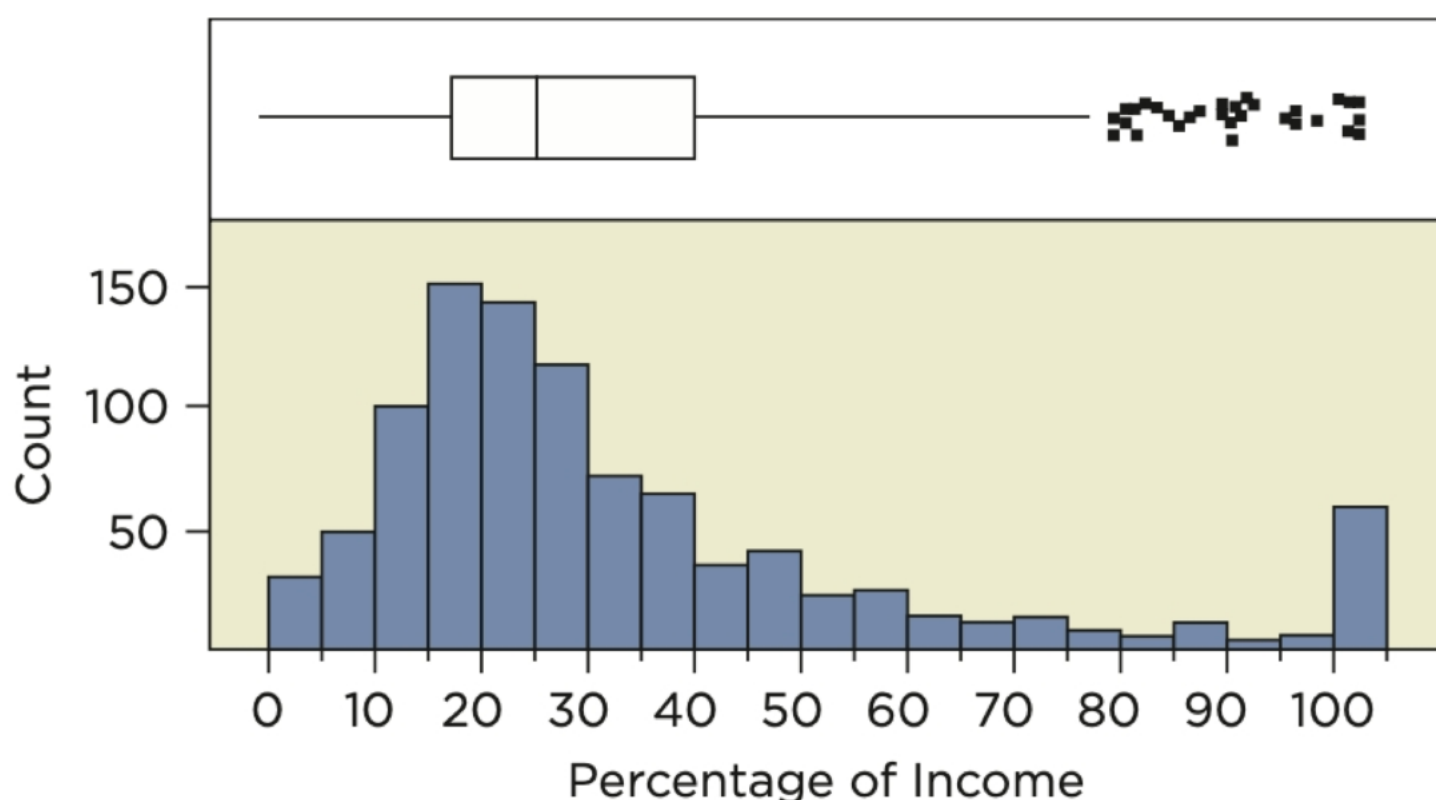

(a) Estimate the mean and median from the figure.
(b) Estimate the IQR from the figure.
(c) What's the effect of the cluster of outliers on the right of the SD?
(d) What's your explanation for the tall bin at the right side of the histogram?

Mix and Match

1. Match the description to the scatterplot.
- (a) Negative direction, linear, moderate variation around line
- (b) Positive direction, linear, small variation around line
- (c) No pattern. A plot of scrambled pairs would look the same
- (d) Negative then positive direction, valley-shaped curve, moderate variation around curve

I.

II.

III.

IV.
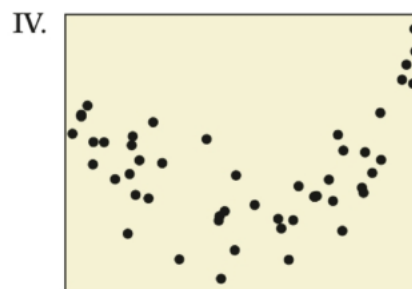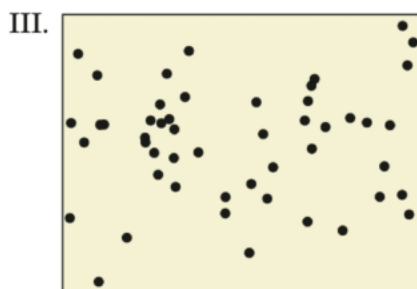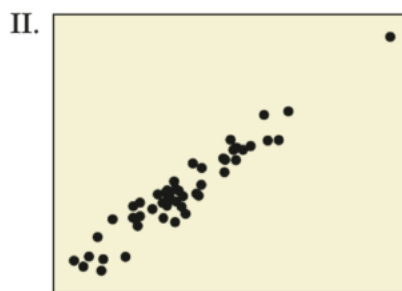
4. Match the value of the correlation to the data in the scatterplot.
   (a) $r = 0$ (b) $r = -0.9$ (c) $r = 1$ (d) $r = 0.4$

I.

II.

III.

IV.
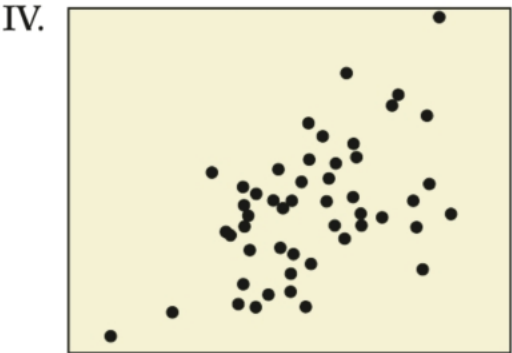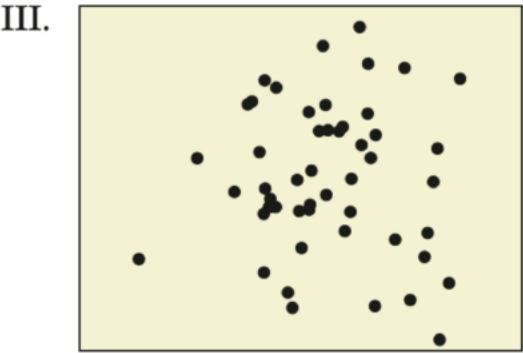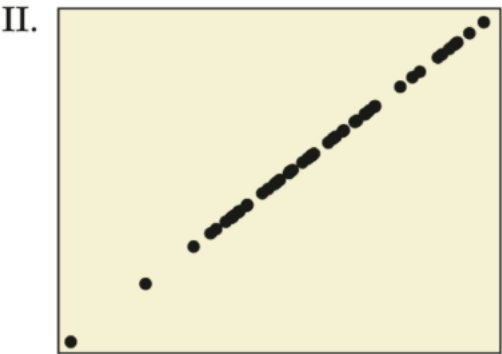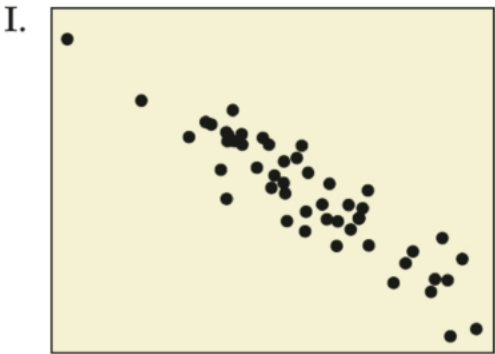
Think About It

21. Suppose you collect data for each of the following pairs of variables. You want to make a scatterplot. Identify the response and explanatory variables. What would you expect to see in the scatterplot? Discuss the direction, curvature, and variation.
   (a) Sales receipts: number of items, total cost
   (b) Productivity: hours worked, items produced
   (c) Football players: weight, time to run 40 yards
   (d) Fuel: number of miles since filling up, gallons left in your tank
   (e) Investing: number of analysts recommending a stock, its subsequent price change

22. Suppose you collect data for each of the following pairs of variables. You want to make a scatterplot. Identify the response and explanatory variables. What would you expect to see in the scatterplot? Discuss the likely direction, curvature, and variation.
   (a) Winter electricity: kilowatts, temperature
   (b) Long-distance calls: time (minutes), cost
   (c) Air freight: weight of load, fuel consumption
   (d) Advertising: length of commercial, number who remember ad
   (e) Health: amount of exercise per week (hours), percentage body fat

True/False

Mark each statement True or False. If you believe that a statement is false, briefly say why you think it is false.

5. The explanatory variable defines the x-axis in a scatterplot.

6. In a plot of income (the response) versus education (the explanatory variable) for managers, managers with the lowest levels of education are at the right-hand side of the figure.

7. The presence of a pattern in a scatterplot means that values of the response increase with the values of the explanatory variable.

8. The visual test for association is used to decide whether a plot has a real or imagined pattern.

9. A line with positive slope describes a linear pattern with a positive direction.

10. A company has learned that the relationship between its advertising and sales shows diminishing marginal returns. That is, as it saturates consumers with ads, the benefits of increased advertising diminish. The company should expect to find linear association between its advertising and sales.

11. If net revenue at a firm is about 10% of gross sales, then a scatterplot of net revenue versus gross sales would show a nonlinear pattern rather than a line.

Find the matching item from the second column.

1. Independent events	(a) $P(\mathbf{A} \text{ and } \mathbf{B}) = P(\mathbf{A}) \times P(\mathbf{B})$
2. Disjoint events	(b) $\mathbf{A}^c$
3. Union	(c) $\mathbf{S}$
4. Intersection	(d) $P(\mathbf{A} \text{ or } \mathbf{B}) + P(\mathbf{A} \text{ and } \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B})$
5. Complement of $\mathbf{A}$	(e) $P(\mathbf{A} \text{ or } \mathbf{B}) \le P(\mathbf{A}) + P(\mathbf{B})$
6. Sample space	(f) $\mathbf{A} \text{ and } \mathbf{B}$
7. Addition Rule	(g) $P(\mathbf{A} \text{ and } \mathbf{B}) = 0$
8. Complement Rule	(h) $\mathbf{A} \text{ or } \mathbf{B}$
9. Boole's inequality	(i) $P(\mathbf{A}^c) = 1 - P(\mathbf{A})$
10. Dependent events	(j) $P(\mathbf{A} \text{ and } \mathbf{B}) = P(\mathbf{A}) \times P(\mathbf{B})$

True/False

Mark each statement True or False. If you believe that a statement is false, briefly say why you think it is false.

Exercises 11–16. A market research assistant watches the next five customers as they leave the store. He records whether the customer is carrying a store bag that indicates the customer made a purchase. He writes down a yes or a no for each. Define the events.

$\mathbf{A} = \{$first two shoppers have a bag$\}$
$\mathbf{B} = \{$last two shoppers have a bag$\}$
$\mathbf{C} = \{$last three shoppers have a bag$\}$

11. The sample space $\mathbf{S}$ for this experiment has 10 elements.

12. The assumption of independence implies that each shopper has the same probability for carrying a bag.

13. $P(\mathbf{A}) + P(\mathbf{B}) = P(\mathbf{A}\,or\,\mathbf{B})$.

14. The probability that both events $\mathbf{B}$ and $\mathbf{C}$ occur is equal to $P(\mathbf{B})$.

15. The probability that a randomly chosen customer purchases with a credit card **or** spends more than $50 is the same as or larger than the probability that the customer purchases with a credit card **and** spends more than $50.

16. If each shopper has the same chance of making a purchase and shoppers behave independently of one another, then $P(\mathbf{A}\,and\,\mathbf{C}) = P(\mathbf{A}) \times P(\mathbf{C})$.

Exercises 17–22. The Human Resources (HR) group at a large accounting firm interviews prospective candidates for new hires. After each interview, the firm rates the candidate on a 10-point scale, with the rating 10 denoting exceptionally good candidates and 1 denoting those that the firm rates poor. The HR group rated 6 candidates on Monday and 6 candidates on Tuesday. The outcomes of these 12 ratings form the sample space.

17. The events $\mathbf{A} = \{$3 candidates on Monday rate above 7$\}$ and $\mathbf{B} = \{$two candidates on Tuesday rate above 7$\}$ are disjoint events.

18. If 8 of the 12 candidates on Monday and Tuesday rate above 6, then the probability of a candidate rating above 6 on Wednesday is 8/12.

19. The HR group has monitored the outcome of these interviews for several years. The Law of Large Numbers assures us that HR personnel can use these data to learn the probability of a candidate scoring above 8 during an interview.

25. A shopper in a convenience store can make a food selection from frozen items, refrigerated packages, fresh foods, or deli items. Let the event **A** = {frozen, refrigerated, fresh} and **B** = {fresh, deli}.
  (a) Find the intersection **A** *and* **B**.
  (b) Find the union of **A** and **B**.
  (c) Find the event **A**^{c}.

26. A credit-rating agency assigns ratings to corporate bonds. The agency rates bonds offered to companies that are most likely to honor their liabilities AAA. The ratings fall as the company becomes more likely to default, dropping from AAA to AA, A, down to BBB, BB, B, CCC, CC, R, and then D (for in default). Let the event **W** = {AAA, AA, A, BBB, BB, B} and **V** = {BBB, BB, B, CCC, CC}.
  (a) Find the intersection **W** *and* **V**.
  (b) Describe the union **W** *or* **V**.
  (c) Find the complement (**W** *or* **V**)^c.

27. A brand of men's pants offered for sale at a clothing store comes in various sizes. The possible waist sizes are

Waist: {24 inches, 26 inches, . . . , 46 inches}

with inseams (length of the pant leg)

Inseam: {28 inches, 29 inches, . . . , 40 inches}

Define the event **B** = {waist 40 inches or larger} and **T** = {inseam 36 inches or larger}.
  (a) Describe the choice of a customer that is in the event (**B** *and* **T**).
  (b) What would it mean if $P(\mathbf{B} \text{ } and \text{ } \mathbf{T}) = P(\mathbf{B}) \times P(\mathbf{T})$?
  (c) Does the choice of a tall, thin customer lie in the event (**B** *and* **T**) or the event (**B** *or* **T**)?