

Name \_\_\_\_\_

**(1) The data at the right** show annual rainfall in a selection of 30 U.S. cities.

(a) In the space below sketch a stem-and-leaf plot of the annual rainfall data. Feel free to select the stems to show the data best. [Do a practice plot on the back of one of these tests pages; put a pretty picture here.]

City	Annual rainfall (inches)
Akron, OH	36
Albany-Schenectady-Troy, NY	35
Allentown, Bethlehem,PA-NJ	44
Atlanta, GA	47
Baltimore, MD	43
Birmingham, AL	53
Boston, MA	43
Bridgeport-Milford, CT	45
Buffalo, NY	36
Canton, OH	36
Chattanooga, TN-GA	52
Chicago, IL	33
Cincinnati, OH-KY-IN	40
Cleveland, OH	35
Columbus, OH	37
Dallas, TX	35
Dayton-Springfield, OH	36
Denver, CO	15
Detroit, MI	31
Flint, MI	30
Grand Rapids,MI	31
Greensboro-Winston-Salem,NC	42
Hartford, CT	43
Houston, TX	46
Indianapolis, IN	39
Kansas City,MO	35
Lancaster, PA	43
Los Angeles,Long Beech, CA	11
Louisville, KY-IN	30
Memphis, TN-AR-MS	50

(b) Use your stem-and-leaf plot to find:

the median: \_\_\_\_\_

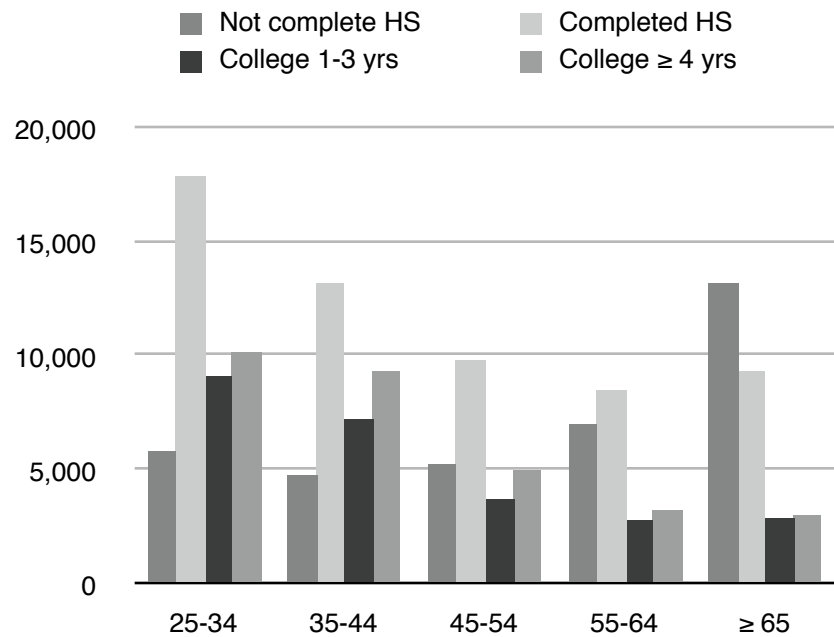
the upper quartile: \_\_\_\_\_

the lower quartile: \_\_\_\_\_

(c) Sketch a box-and-whiskers plot for this data showing the median, the quartiles, and whiskers extending to the quartiles  $\pm 1.5$  times the interquartile range. [If you couldn't find the quartiles, just put guesses in (b) and use those in (c).] Show outliers as dots.

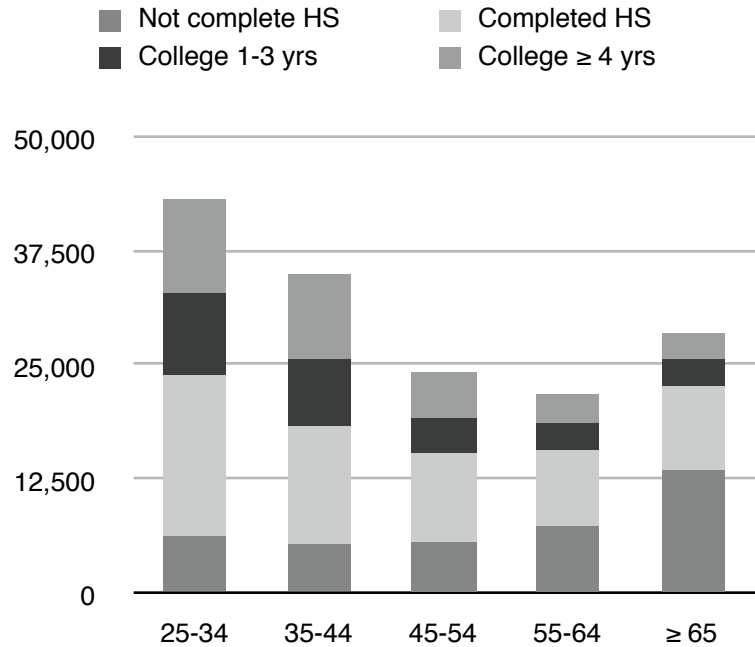
(d) What cities are outliers?

**(2) The bar chart below** shows the adult population of the U.S. (from about ten years ago) broken down into five age categories and four education categories. ("HS" refers to high school.) [Note if you have trouble reading which grey goes with which category, note that the greys go *in order* of the amount of education.]

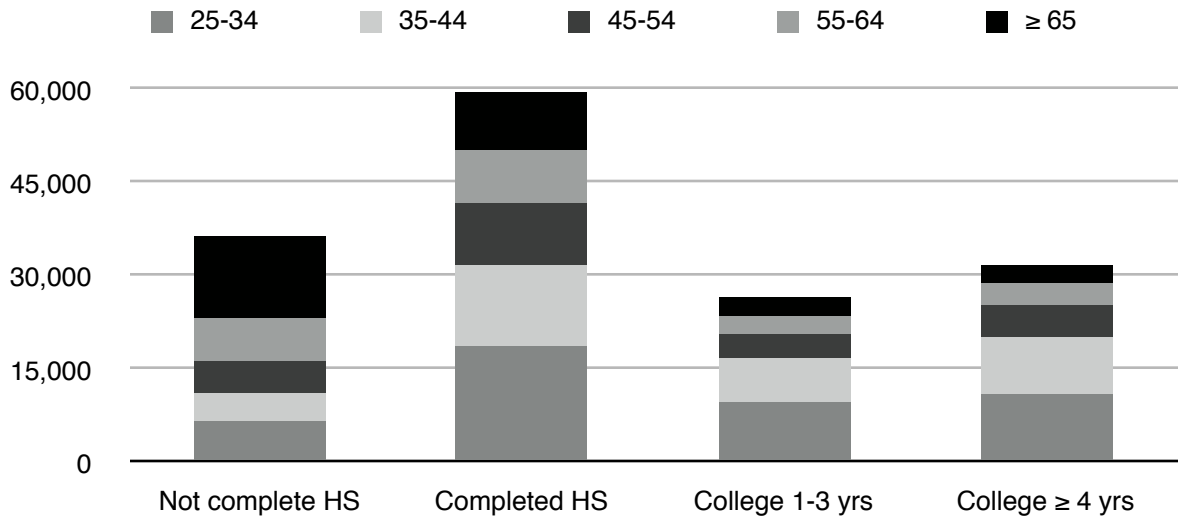


(a) What interesting stories can you tell about age and education in this data from this plot?

(b) At the right is the same data shown in a stacked bar chart. What can you observe in this plot that may have been difficult to see in the chart in (a)?  
Answer here:



(c) Below is another stacked bar chart, looking at the data in still a different way. What do you observe here?



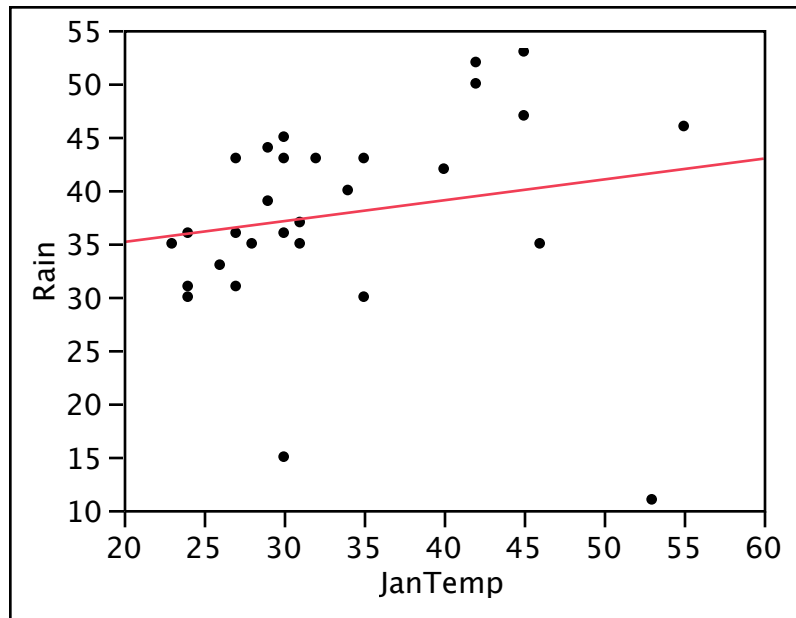
Answer:

(d) The data used for the charts above is shown below. The total of all the cells is 151,617.

	25-34	35-44	45-54	55-64	≥ 65
Not complete HS	5836	4841	5230	7024	13183
Completed HS	17889	13200	9860	8580	9412
College 1-3 yrs	9069	7309	3698	2793	2915
College ≥ 4 yrs	10174	9332	5008	3246	3018

What is the marginal distribution for the age categories?

**(3) At the right is a scatter plot** from the same data set shown partially in (1), plotting annual rainfall in 30 U.S. metropolitan areas versus the average temperature in January in degrees Fahrenheit. The regression line best fitting these points is also shown.



(a) The equation of the line is  $\text{Rain} = 21.0 + 0.54 * \text{JanTemp}$ .

The slope of this line is \_\_\_\_\_. The y-intercept of this line is \_\_\_\_\_.

(b) What would the regression line predict for the annual rainfall of a U.S. metropolitan area that has an average January temperature of 50°?

\_\_\_\_\_

(c) What are the units of 0.54 in the regression equation?

(d) R-squared for this regression is 0.303. What does that tell you about the variance of Rain?

(e) What is the correlation between January temperatures and rainfall in this data?\_\_\_\_\_

(f) There is a point in the scatter plot representing a metropolitan area with average January temperature of  $53^{\circ}$  and annual rainfall of 11 inches (happens to be the Los Angeles area). Describe the effect of that point on the slope of the regression line. What would happen to the slope of the line if that point were excluded from the analysis?

(g) What would happen to R-squared if the point in (f) were excluded from the analysis?

**(4) A recent New York Times/CBS poll** showed that 59% of Americans say the government should provide national health insurance. Suppose that's the true population proportion, and we ask a random sample of 100 people if they favor national health insurance.

(a) What is the probability that the proportion in this sample who favor national health insurance is less than 50%?

(b) If we asked 200 people the same question, would the probability the sample proportion is less than 50% be greater or less than in (a)?

**(5) Understandings in your own words**

(a) Explain to a mildly confused statistics student how we know the mean is the balancing point of a distribution.

(b) Explain to a mildly confused statistics student how we know that the distribution of means of samples of size  $n$  from any population is bell-shaped.

(c) Explain how we know without computing that the mean of the distribution of per capita incomes in Brazil (for example) is sure to be greater than the median. [Continue on back if necessary.]