Chapter 5
Summarizing Bivariate Data

Source: TPS

5.1 Bivariate Relationships

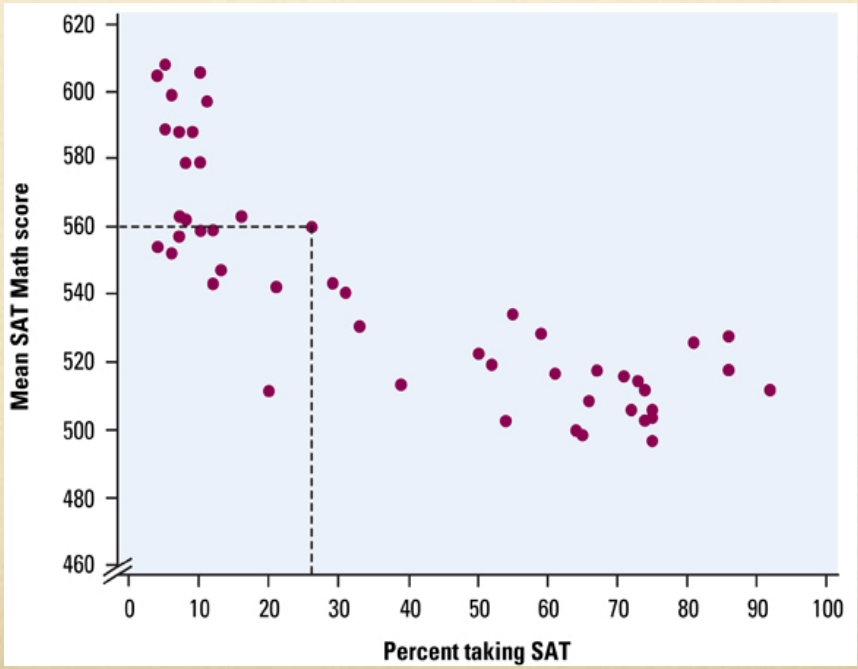
- What is Bivariate data?
- When exploring/describing a bivariate (x,y) relationship:
  - ☑ Determine the Explanatory and Response variables
  - ☑ Plot the data in a scatterplot
  - ☑ Note the Strength, Direction, and Form
  - ☑ Note the mean and standard deviation of x and the mean and standard deviation of y
  - ☑ Calculate and Interpret the Correlation, r
  - ☑ Calculate and Interpret the Least Squares Regression Line in context.
  - ☑ Assess the appropriateness of the LSRL by constructing a Residual Plot.

Response Vs. Explanatory Variables

- Response variable measures an outcome of a study, explanatory variable helps explain or influences changes in a response variable (like independent vs. dependent).
- Calling one variable explanatory and the other response doesn't necessarily mean that changes in one CAUSE changes in the other.
- Ex: Alcohol and Body temp: One effect of Alcohol is a drop in body temp. To test this, researches give several amounts of alcohol to mice and measure each mouse's body temp change. What are the explanatory and response variables?

Scatterplots

- Scatterplot shows the relationship between two quantitative variables measured on the same individuals.
  - Explanatory variables along X axis, Response variables along Y.
  - Each individual in data appears as the point in the plot fixed by the values of both variables for that individual.
  - Example:



Interpreting Scatterplots

- Direction: in previous example, the overall pattern moves from upper left to lower right. We call this a negative association.
- Form: The form is slightly curved and there are two distinct clusters. What explains the clusters? (ACT States)
- Strength: The strength is determined by how closely the points follow a clear form. The example is only moderately strong.
- Outliers: Do we see any deviations from the pattern? (Yes, West Virginia, where 20% of HS seniors take the SAT but the mean math score is only 511).

Association

Interpreting a Scatterplot

In any graph of data, look for the **overall pattern** and for **striking deviations** from that pattern.

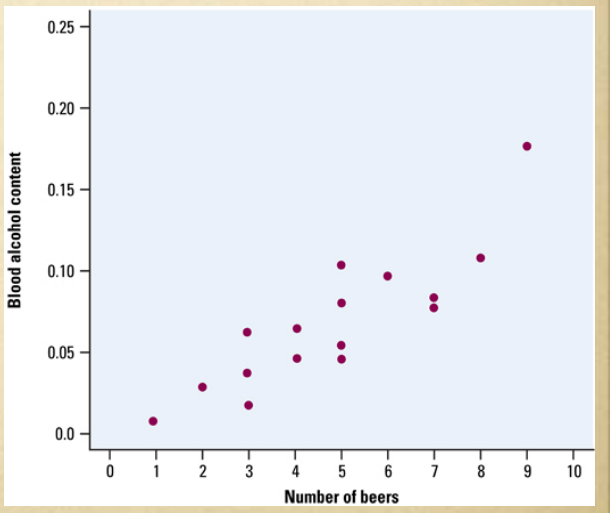
You can describe the overall pattern of a scatterplot by the **direction, form,** and **strength** of the relationship.

An important kind of deviation is an **outlier**, an individual value that falls outside the overall pattern of the relationship.

Positive Association, Negative Association

Two variables are **positively associated** when above-average values of one tend to accompany above-average values of the other, and below-average values also tend to occur together.

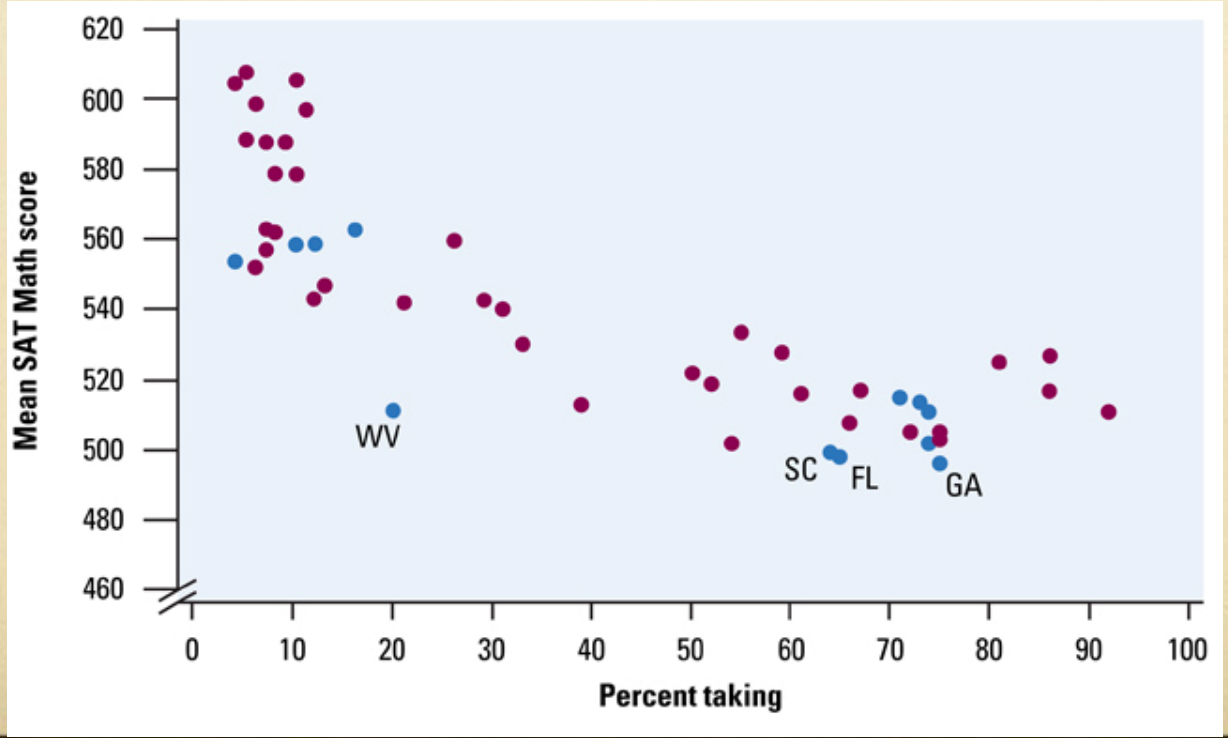
Two variables are **negatively associated** when above-average values of one tend to accompany below-average values of the other, and vice versa.



Introducing Categorical Variables

Categorical Variables in Scatterplots

To add a categorical variable to a scatterplot, use a different plotting color or symbol for each category.



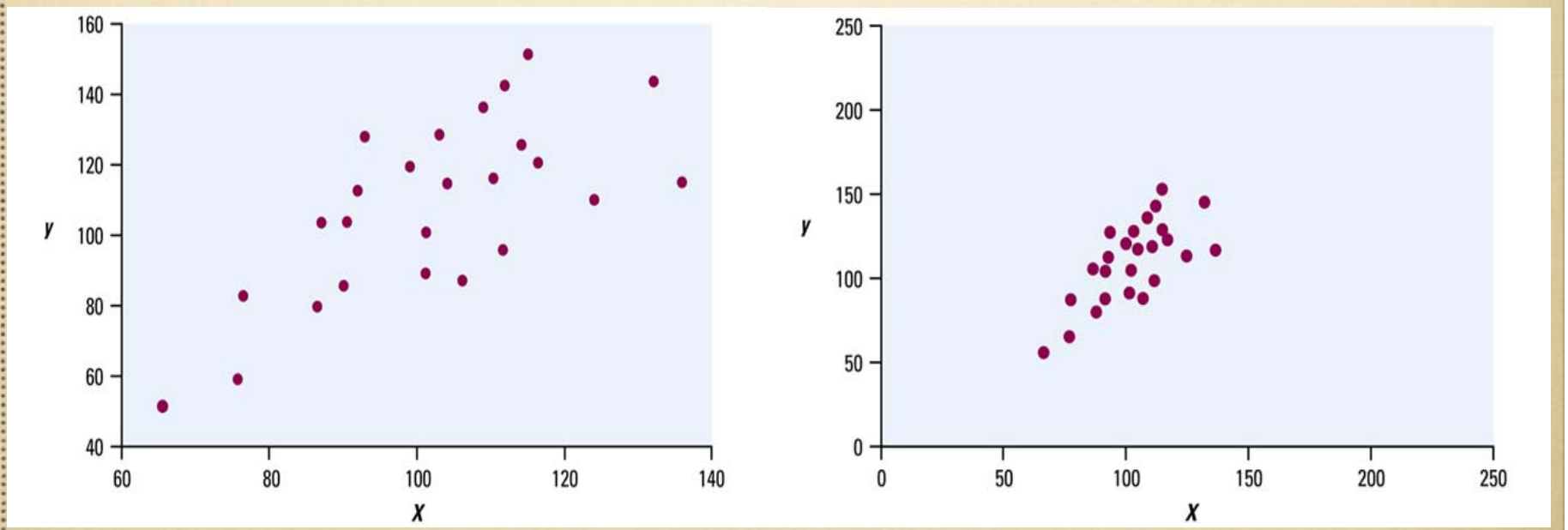
Calculator Scatterplot

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Beers	5	2	9	8	3	7	3	5	3	5	4	6	5	7	1	4
BAC	0.1	0.03	0.19	0.12	0.04	0.0950	0.07	0.06	0.02	0.05	0.07	0.1	0.085	0.09	0.01	0.05

- Enter the Beer consumption in L1 and the BAC values in L2
- Next specify scatterplot in Statplot menu (first graph). X list L1 Y List L2 (explanatory and response)
  - Use ZoomStat.
    - Notice that their are no scales on the axes and they aren't labeled. If you are copying your graph to your paper, make sure you scale and label the Axis (use Trace)

Correlation

- Caution- our eyes can be fooled! Our eyes are not good judges of how strong a linear relationship is. The 2 scatterplots depict the same data but drawn with a different scale. Because of this we need a numerical measure to supplement the graph.



r

- The Correlation measures the direction and strength of the linear relationship between 2 variables.
- Formula- (don't need to memorize or use): $r = \frac{\sum Z_x Z_y}{n-1}$
- In Calc: Go to Catalog (2nd, zero button), go to DiagnosticOn, enter, enter. You only have to do this ONCE! Once this is done:
- Enter data in L1 and L2 (you can do calc-2 var stats if you want the mean and sd of each)
    - Calc, LinReg (A + Bx) enter

Interpreting r

- The absolute value of r tells you the strength of the association (0 means no association, 1 is a strong association)
- The sign tells you whether it's a positive or a negative association. So r ranges from -1 to +1
  - Note- it makes no difference which variable you call x and which you call y when calculating correlation, but stay consistent!
  - Because r uses standardized values of the observations, r does not change when we change the units of measurement of x, y, or both. (Ex: Measuring height in inches vs. ft. won't change correlation with weight)
  - values of -1 and +1 occur ONLY in the case of a perfect linear relationship , when the variables lie exactly along a straight line.

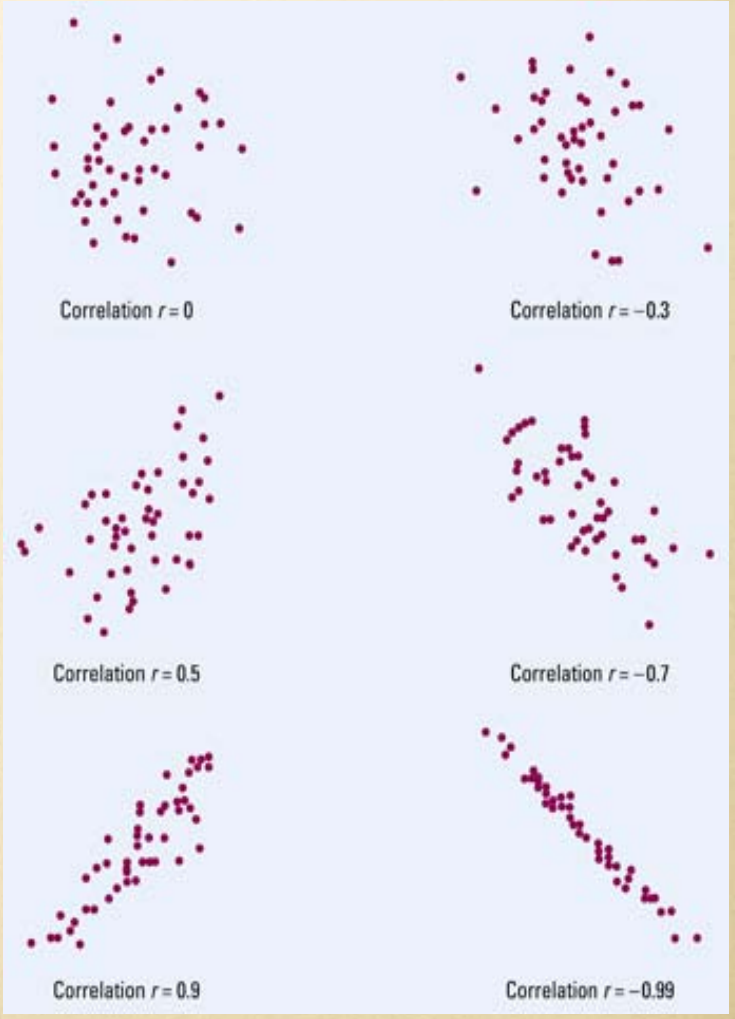
Examples

1. Correlation requires that both variables be quantitative

2. Correlation measures the strength of only LINEAR relationships, not curved...no matter how strong they are!

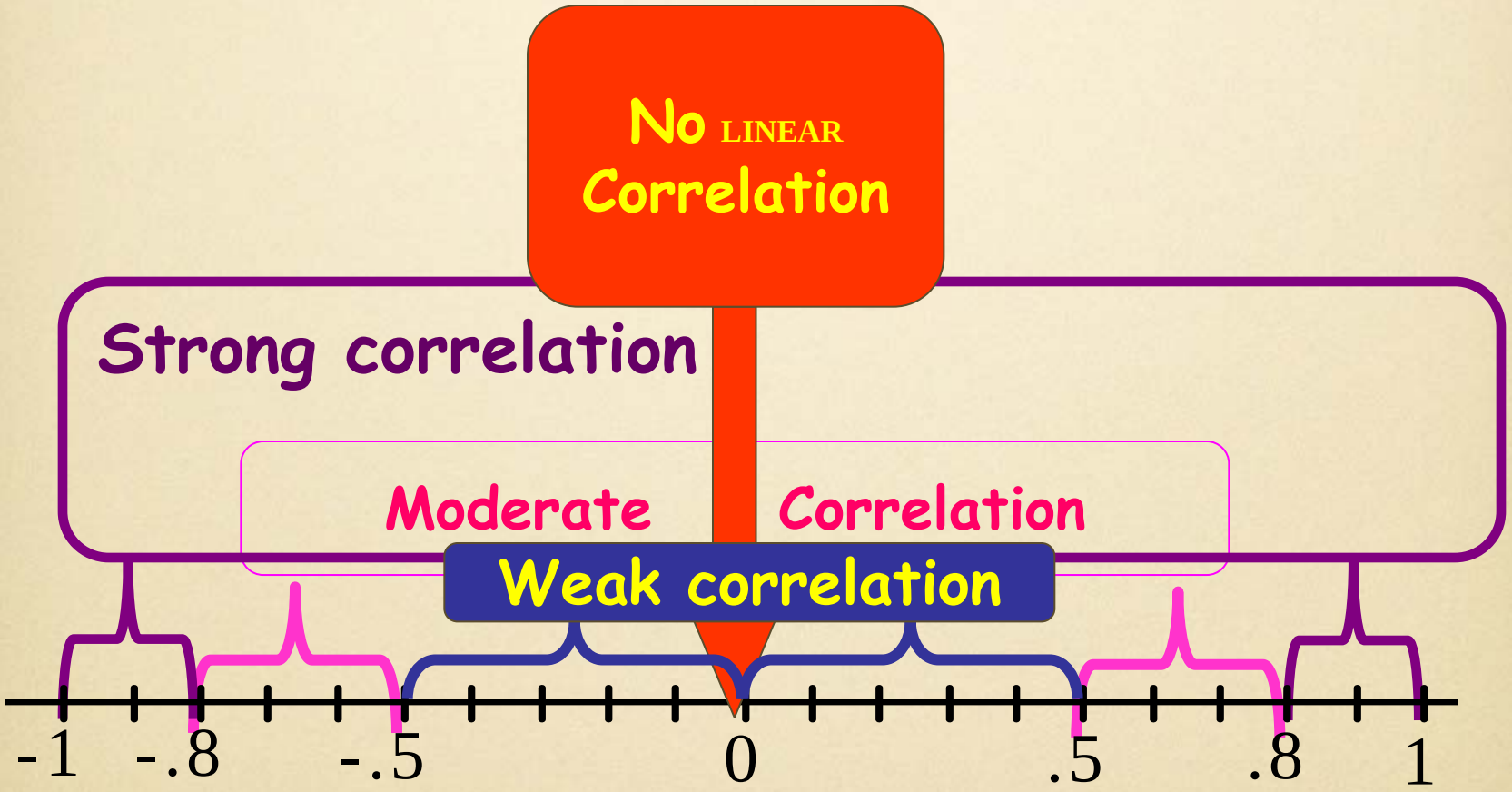
3. Like the mean and standard deviation, the correlation is not resistant: r is strongly affected by a few outlying observations. Use r with caution when outliers appear in the scatterplot

4. Correlation is not a complete summary of two-variable data, even when the relationship is linear- always give the means and standard deviations of both x and y along with the correlation.



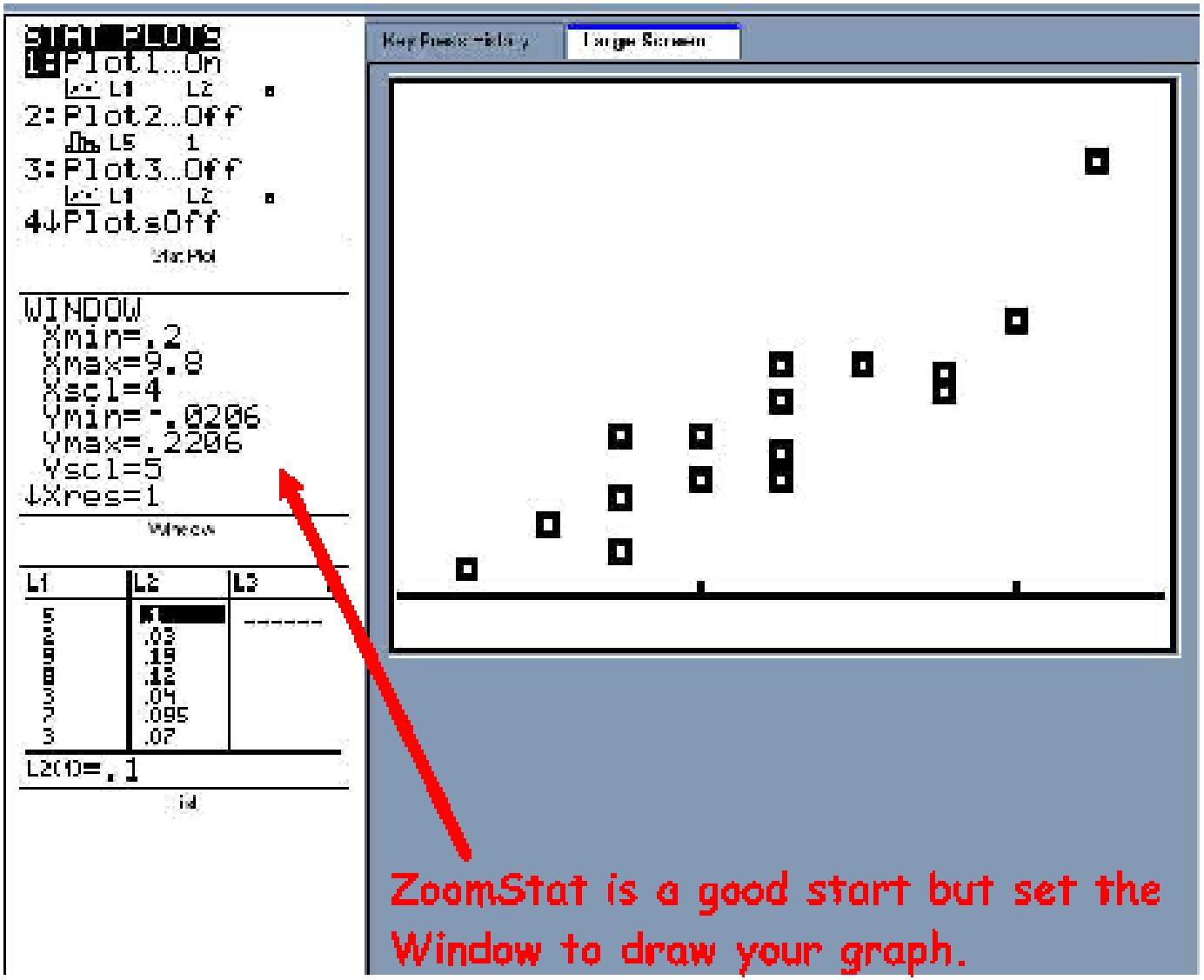
Properties of *r* (correlation coefficient)

- legitimate values are $-1 \leq r \leq 1$



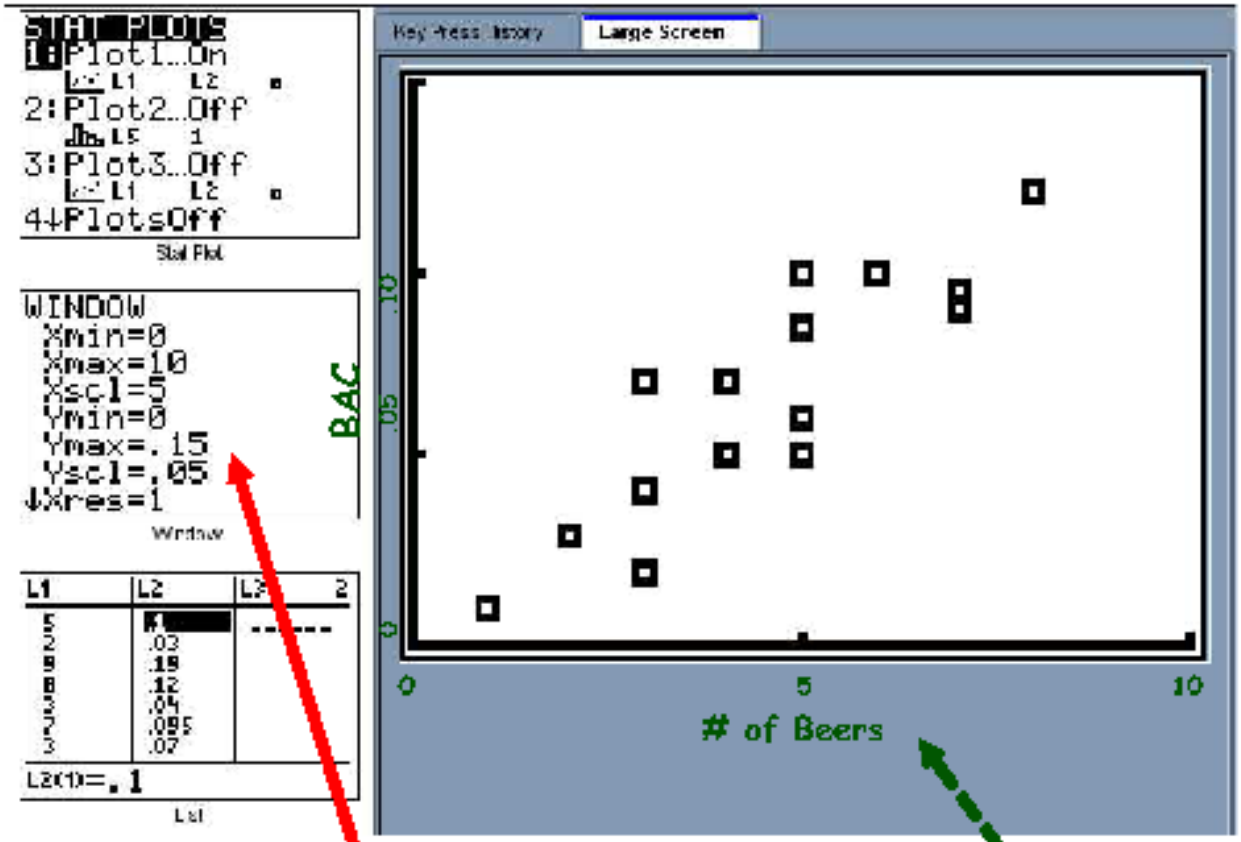
Appendix

5.1 APPENDIX: Beer vs BAC Scatterplot



"Beer vs BAC Scatterplot"

SET WINDOW to sketch your graph

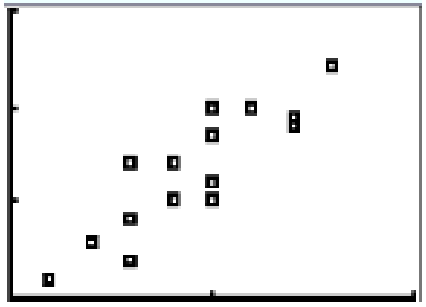


WINDOW: I typically plot the minimum and maximum values; set the SCL in easy to graph units like 5's or powers of 10.

Scatterplots need to label axes; and mark scales

Do Beer and BAC have an association?

From our scatterplot, we see Beer and BAC have a positive, linear, & moderately strong association.



But how strong is it?

- Find r
- r=.894
-
- Actually it is a strong association between Beer and BAC because "r" is between .8 and 1.