

Chapter 3: Scatterplots, Association & Correlation

Skills

- Construct a scatterplot
- Describe a scatterplot
- Identify the explanatory and response variables
- Calculate and interpret the correlation coefficient

Two Quantitative Variables

Univariate = one variable (what we've been doing for most of the year)

Bivariate = two variables

We've done two qualitative variables...two quantitative variables is new!

Step 1: Graph the data!
(with a scatterplot)

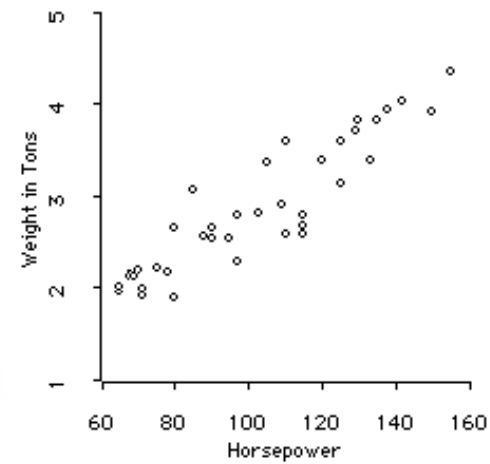
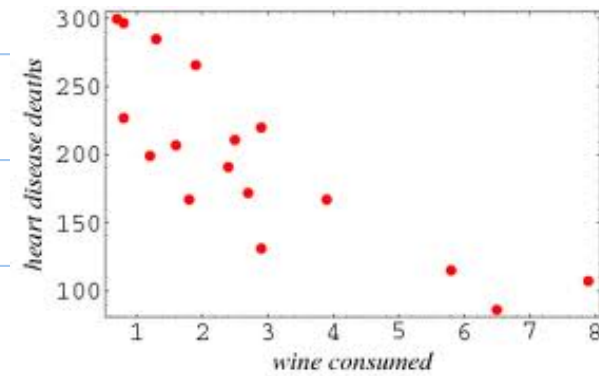
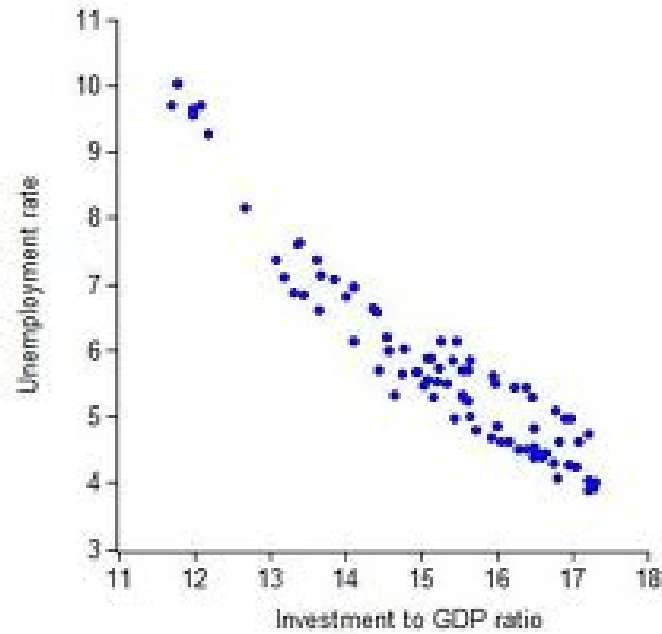
Describing a Scatterplot

We describe most graphs with center, shape and spread.

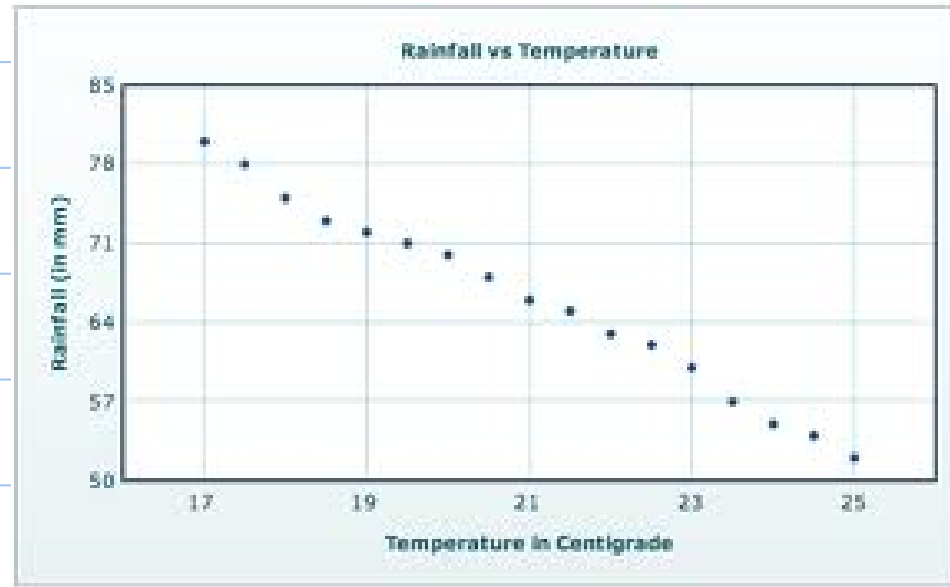
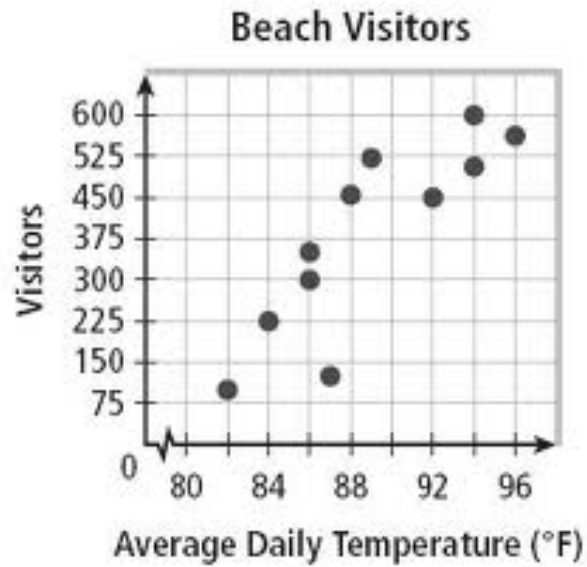
For scatterplots, we instead describe

- Association
 - Direction
 - Strength
 - Unusual Features
-
-
-
-
-
-
-
-
-
-

Association (Linearity)



Direction



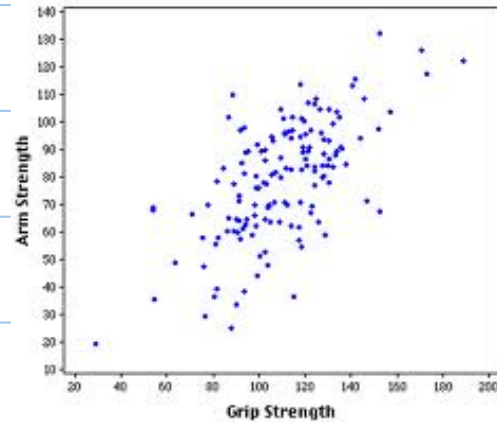
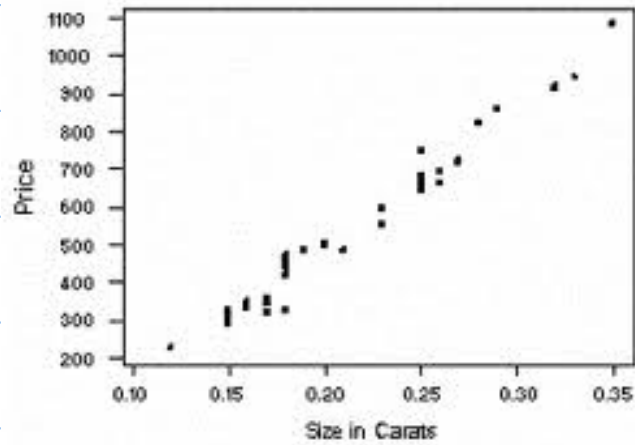
Strength

How wide of a brush do you need to paint over the points?

Wide brush = weak/low strength

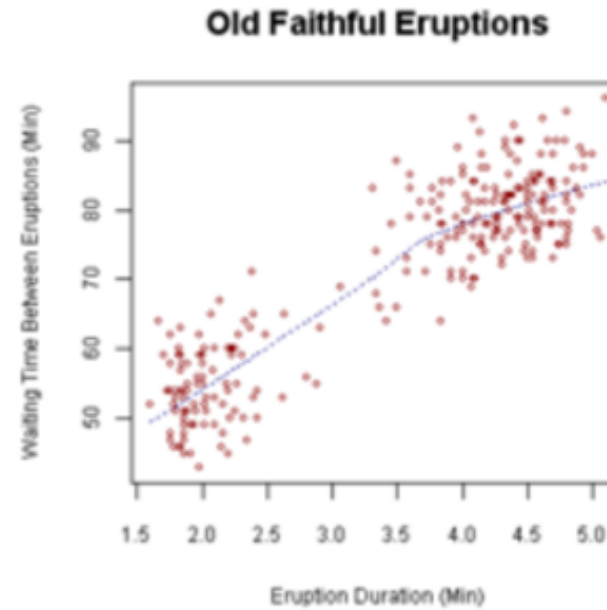
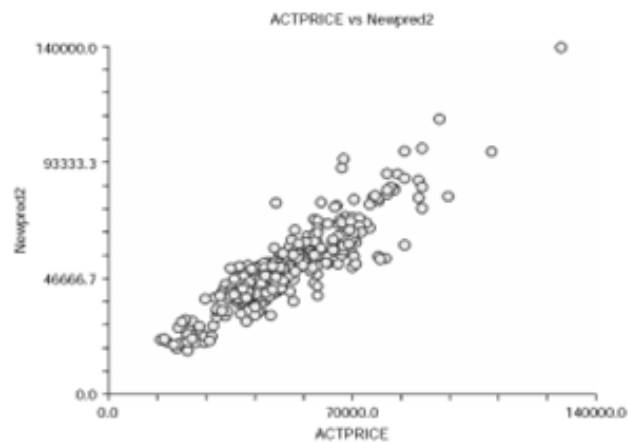
Thin brush = strong/high strength

Strength



Unusual Features

- Clusters
- Gaps
- Outliers



Example

Six male gray kangaroos were sampled. The nasal length and width were measured (in millimeters). Researchers would like to predict the nasal width from the nasal length.

Example

Length	Width
--------	-------

609	241
-----	-----

629	222
-----	-----

620	233
-----	-----

564	207
-----	-----

645	247
-----	-----

493	189
-----	-----

Example

Is there a relationship between the amount of fat and number of calories in fast food hamburgers?

Example

Fat	Calories
-----	----------

19	410
----	-----

31	580
----	-----

34	590
----	-----

35	570
----	-----

39	640
----	-----

39	680
----	-----

43	660
----	-----

More on Strength

Describing correlation as strong, moderate or weak is quite subjective...if only there were an objective measurement of correlation!

Pearson's Product-Moment Correlation Coefficient

$$r = \frac{\sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)}{n - 1}$$

Correlation Coefficient

r measures the strength and direction of a linear relationship.

r is meaningless if the relationship isn't linear.

Thus, the value of r CANNOT be used to determine if a relationship is linear!

It CAN be used to support arguments about relationships that appear to be linear.

Fun facts about r :

- r has a value between -1 and 1 (inclusive).
 - the sign of r is also the direction of the data (positive or negative).
 - r has no units, and is unaffected by changing units in the original data.
-
-
-
-
-
-
-
-
-
-

Sometimes you'll need this:

CATALOG

D

Diagnostic On

(ENTER)

In General:

LinReg ($a + bx$) L_1, L_2