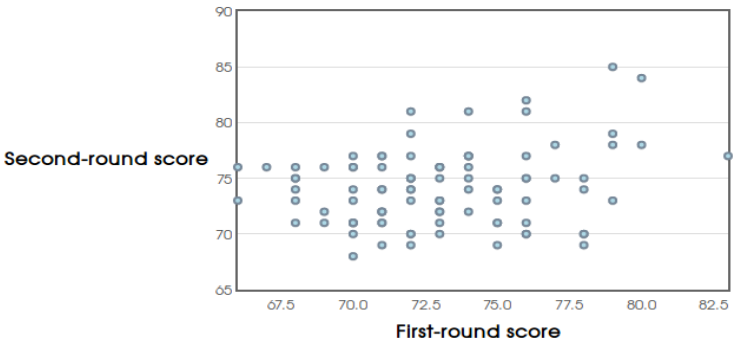
1. The Masters is one of the four major golf tournaments. is a scatterplot of the scores for the first two rounds of the 2010 Masters for all the golfers entered. Only the 60 golfers with the lowest two-round total advance to the final two rounds. The plot has a grid pattern because golf scores must be whole numbers.

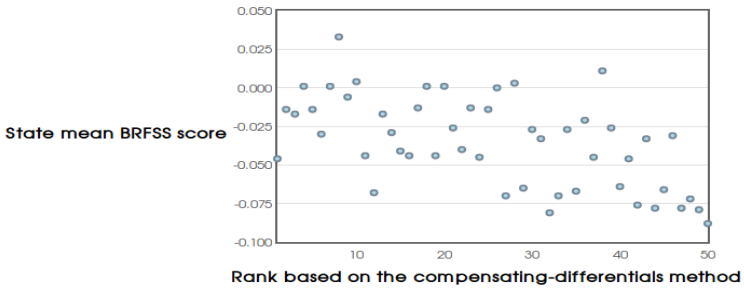


(a) Read the graph: What was the lowest score in the first round of play? How many golfers had this low score? What were their scores in the second round?

(b) Read the graph: Sandy Lyle had the highest score in the second round. What was this score? What was Lyle's score in the first round?

(c) Is the correlation between first-round scores and second- round scores closest to $r = 0.1, r = 0.5$, or $r = 0.9$? Explain your choice. Does the graph suggest that knowing a professional golfer's score for one round is much help in predicting his score for another round on the same course?

2. Human happiness or well-being can be assessed either subjectively or objectively. Subjective assessment can be accomplished by listening to what people say. Objective assessment can be made from data related to well-being such as income, climate, availability of entertainment, housing prices, lack of traffic congestion, etc. Do subjective and objective assessments agree? To study this, investigators made both subjective and objective assessments of happiness for each of the 50 states. The subjective measurement was the mean score on a life-satisfaction question found on the Behavioral Risk Factor Surveillance System (BRFSS), which is a state-based system of health surveys. Lower scores indicate a greater degree of happiness. To objectively assess happiness, the investigators computed a mean well-being score (called the compensating- differentials score) for each state, based on objective measures that have been found to be related to happiness or well-being. The states were then ranked according to this score (Rank 1 being the happiest). The figure below is a scatterplot of mean BRFSS scores (response) against the rank based on the compensating-differentials scores (explanatory).



(a) Is there an overall positive association or an overall negative association between mean BRFSS score and rank based on the compensating-differentials method?

(b) Does the overall association indicate agreement or disagreement between the mean subjective BRFSS score and the ranking based on objective data used in the compensating-differentials method?

(c) Are there any outliers? If so, what are the BRFSS scores corresponding to these outliers?

3. Some studies have suggested that a nightly glass of wine may not only take the edge off a day but also improve health. Is wine good for your health? A study of nearly 1.3 million middle-aged British women examined wine consumption and the risk of breast cancer. The researchers were interested in how risk changed as wine consumption increased. Risk is based on breast cancer rates in drinkers relative to breast cancer rates in nondrinkers in the study, with higher values indicating greater risk. In particular, a value greater than 1 indicates a greater breast cancer rate than that of nondrinkers. Wine intake is the mean wine intake, in grams per day, of all women in the study who drank approximately the same amount of wine per week. Here are the data (for drinkers only):

Wine intake: x grams per day	2.5	8.5	15.5	26.5
Relative Risk y	1.00	1.08	1.15	1.22

(a) Make a scatterplot of these data. Based on the scatterplot, do you expect the correlation to be positive or negative? Near ± 1 or not?

(b) Find the correlation r between wine intake and relative risk. Do the data show that women who consume more wine tend to have higher relative risks of breast cancer?

4. The deadly Ebola virus is a threat to both people and gorillas in Central Africa. An outbreak in 2002 and 2003 killed 91 of the 95 gorillas in 7 home ranges in the Congo. To study the spread of the virus, measure "distance" by the number of home ranges separating a group of gorillas from the first group infected. Here are data on distance and time in number of days until deaths began in each later group

Distance	1	3	4	4	4	5
Time	4	21	33	41	43	46

(a) Make a scatterplot. Which is the explanatory variable? What kind of pattern does your plot show?

(b) Find the correlation r between distance and time.

(c) If time in days were replaced by time in number of weeks until death began in each later group (fractions allowed so that 4 days becomes 4/7 weeks), would the correlation between distance and time change? Explain your answer.

5. **Sparrowhawk colonies.** One of nature's patterns connects the percent of adult birds in a colony that return from the previous year and the number of new adults that join the colony. Here are data for 13 colonies of sparrowhawks:

Percent return x	74	66	81	52	73	62	52	45	62	46	60	46	38
New adults y	5	6	8	11	12	15	16	17	18	18	19	20	20

(a) Plot the count of new adults (response) against the percent of returning birds (explanatory). Describe the direction and form of the relationship. Is the correlation r an appropriate measure of the strength of this relationship? If so, find r.

(b) For short-lived birds, the association between these variables is positive: changes in weather and food supply drive the populations of new and returning birds up or down together. For long-lived territorial birds, on the other hand, the association is negative because returning birds claim their territories in the colony and don't leave room for new recruits. Which type of species is the sparrowhawk?

6. **Our brains don't like losses.** Most people dislike losses more than they like gains. In money terms, people are about as sensitive to a loss of \$10 as to a gain of \$20. To discover what parts of the brain are active in decisions about gain and loss, psychologists presented subjects with a series of gambles with different odds and different amounts of winnings and losses. From a subject's choices, they constructed a measure of "behavioral loss aversion." Higher scores show greater sensitivity to losses. Observing brain activity while subjects made their decisions pointed to specific brain regions. Here are data for 16 subjects on behavioral loss aversion and "neural loss aversion," a measure of activity in one region of the brain

Neural	−50.0	−39.1	−25.9	−26.7	−28.6	−19.8	−17.6	5.5
Behavioral	0.08	0.81	0.01	0.12	0.68	0.11	0.36	0.34
Neural	2.6	20.7	12.1	15.5	28.8	41.7	55.3	155.2
Behavior	0.53	0.68	0.99	1.04	0.66	0.86	1.29	1.94

(a) Make a scatterplot that shows how behavior responds to brain activity.

(b) Describe the overall pattern of the data. There is one clear outlier. What is the behavioral score associated with this outlier?

(c) Find the correlation r between neural and behavioral loss aversion both with and without the outlier. Does the outlier have a strong influence on the value of r? By looking at your plot, explain why adding the outlier to the other data points causes r to increase.

7. **Sulfur, the ocean, and the sun.** Sulfur in the atmosphere affects climate by influencing formation of clouds. The main natural source of sulfur is dimethyl sulfide source of sulfur is dimethyl sulfide (DMS) produced by small organisms in the upper layers of the oceans. DMS production is in turn influenced by the amount of energy the upper ocean receives from sunlight. Here are monthly data on solar radiation dose (SRD, in watts per square meter) and surface DMS concentration (in nanomolars) for a region in the Mediterranean.

SRD	12.55	12.91	14.34	19.72	21.52	22.41	37.65	48.41
DMS	0.796	0.692	1.744	1.062	0.682	1.517	0.736	0.720
SRD	74.41	94.14	109.38	157.79	262.67	268.96	289.23	
DMS	1.820	1.099	2.692	5.134	8.038	7.280	8.872	

(a) Make a scatterplot that shows how DMS responds to SRD.

(b) Describe the overall pattern of the data. Find the correlation r between DMS and SRD. Because SRD changes with the seasons of the year, the close relationship between SRD and DMS helps explain other seasonal patterns.

8. **Sloppy writing about correlation.** Each of the following statements contains a blunder. Explain in each case what is wrong.

(a) "There is a high correlation between the sex of American workers and their income."

(b) "We found a high correlation ($r = 1.09$) between students' ratings of faculty teaching and ratings made by other faculty members."

(c) "The correlation between height and weight of the subjects was $r = 0.63$ centimeter."