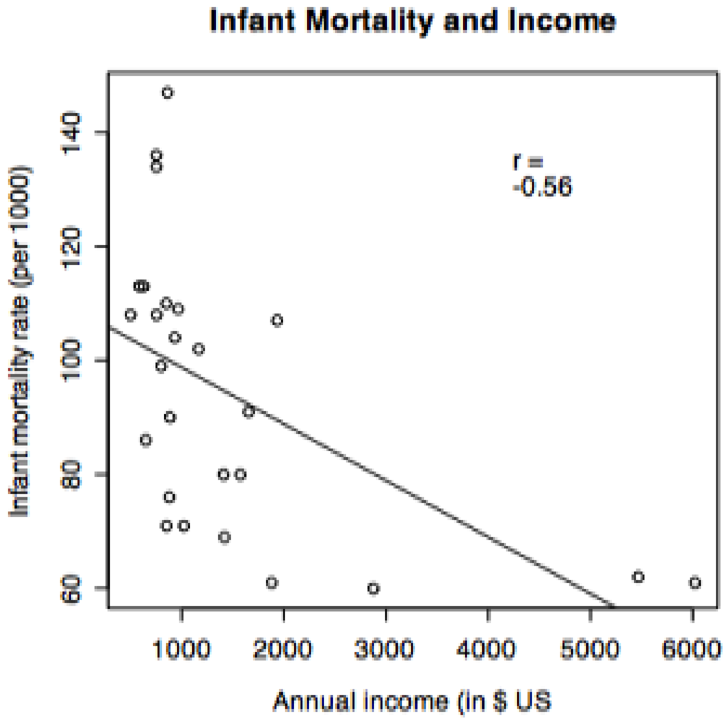
Chapter 9-Correlation

9.1 Infant Mortality and Income:



The outliers would appear to have a distorting effect on the correlation coefficient. However, if you replot the data without those points the relationship is still apparent and the correlation only drops to -.54.

9.3 With 24 degrees of freedom, and two-tailed test at $\alpha = .05$ would require $r > \pm .388$.

9.5 We can conclude that infant mortality is closely tied to both income and the availability of contraception. Infants born to people living in poverty are much more likely to die before their first birthday, and the availability of contraception significantly reduces the number of infants put at risk in the first place.

9.7 Because both income and contraception are related to mortality, we might expect that using them together would lead to a substantial increase in predictability. But note that they are correlated with each other, and therefore share some of the same variance.

9.9 Psychologists have a professional interest in infant mortality because some of the variables that contribute to infant mortality are behavioral ones, and we care about understanding, and often controlling, behavior. Psychologist have an important role to play in world health that has little to do with pills and irrigation systems.

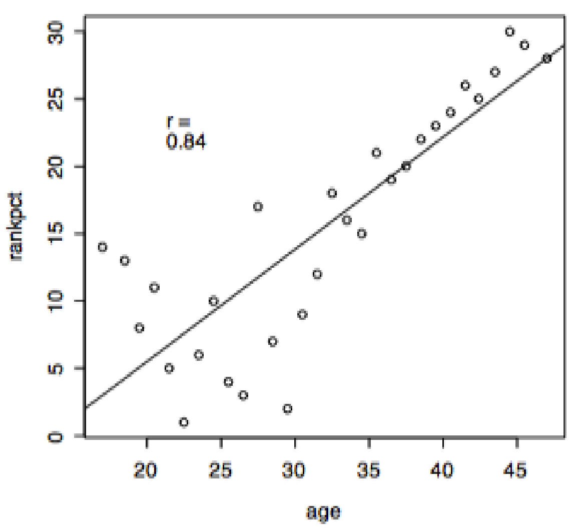
There is a great deal of data available on these issues, and students can easily find it on the Internet. If they are interested in this question, they might also be interested in searching for similar literature on HIV/AIDS.

This question was partly intended to make students think about the fact that all sorts of things are of interest to psychologists. We don't just run animals in a maze or inquire into people's dirty minds. In addition, low birthweight is a risk factor for all sorts of infant outcomes.

9.11 *R* Code Taken from chapter

```
DownData <-
read.table("http://www.uvm.edu/~dhowell/fundamentals8/DataFiles/Ex9-10.dat",
header = TRUE)
attach(DownData)
pctDown <- Down/Births
plot(Age,Births)
plot(Age, pctDown)
ranks <- rank(pctDown)
plot(Age, ranks)
cor(Age, ranks)
```

9.13 If you convert the dependent variable to ranks (or if you convert it to logs), the relationship becomes much more linear because you have reduced the influence of the higher incidence rates. The relationship has gone from exponential to nearly linear, although I don't know what you have really gained in explanatory power over what you see in the figure above. The plot follows.



This is not a Spearman correlation because we have only ranked one of the variables.

9.15 The relationship between test scores in Katz' study and SAT scores for application purposes is a relevant question because we would not be satisfied with a set of data that used SAT questions and yet gave answers that were not in line with SAT performance. We want to know that the tests are measuring at least roughly the same thing. In addition, by knowing the correlation between SATs and performance without seeing the questions, we get a better understanding of some of what the SAT is measuring.

9.17 Correlation for the data in Exercise 9.14:

SAT:	mean = 598.57	$\sum X = 16760$	St. Dev. = 61.57
Test:	mean = 46.21	$\sum Y = 1294$	St. Dev. = 6.73

$$\text{cov}_{YX} = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{N-1} = \frac{780500 - \frac{16760 * 1294}{28}}{27} = 220.3175$$

$$r = \frac{\text{cov}_{YX_1}}{s_Y s_{X_1}} = \frac{220.3175}{61.57 * 6.73} = .53$$

With 26 *df* we would need a correlation of .374 to be significant. Since our value exceeds that, we report that there is strong evidence that relationship between test scores and the SAT is reliably different from 0.

9.19 When we say that two correlations are not significantly different, we mean that they are sufficiently close that they could both have come from samples from populations with exactly the same population correlation coefficient.

9.21 The answer to this question depends on the students' expectations.

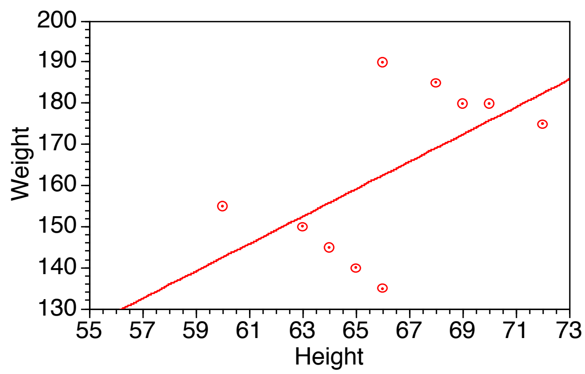
9.23 It is sometimes appropriate to find the correlation between two variables even if you know that the relationship is slightly curvilinear. A straight line often does a remarkably good job of fitting a curved function, provided that it is not too curved.

9.25 The amount of money that a country spends on health care may have little to do with life expectancy because to change a country's life expectancy you have to change the health of a great many individuals. Spending a great deal of money on one person, even if it were to extend her life by dozens of years, would not change the average life expectancy in any noticeable way. Often the things that make a major change in life expectancy, like inoculations, really cost very little money.

> The African Red Cross estimates that there are 300-500 million cases of malaria each year, resulting in 1.5 to 2.5 million deaths. In particular, more than 90% of the deaths are in children under 5 years of age, and they occur predominantly in sub-Saharan Africa. Malaria cases could be cut by up to a third with insecticide treated bednets, which are very cheap by U.S. healthcare standards.

9.27 Extremely exaggerated data on male and female weight and height to show a negative slope within gender but a positive slope across gender:

Height	68	72	66	69	70	66	60	64	65	63
Weight	185	175	190	180	180	135	155	145	140	150
Gender	Male	Male	Male	Male	Male	Fem.	Fem.	Fem.	Fem.	Fem.



It is helpful to get the students to realize that what they are effectively plotting here is the relationship between *Gender* and Weight, more than between Height and Weight.

9.29 We have confounding effects here. If we want to claim that red wine consumption lowers the incidence of heart disease, we have a problem because the consumption of red wine is highest in those areas with the greatest solar radiation, which is another potential cause of the effect. We would have to look at the relationship between red wine and heart disease controlling for the effects of solar radiation.

9.31 This is an Internet search with no fixed answer.

9.33 Using *R* is tricky because you need to install the MBESS package and the gsl package. Ci.R The easier way is to computer the correlation (.3805) and use the package that I refer to in the question. The result will be

r = .3805 **n =** 50	Reset Calculate

0.95 and 0.99 Confidence Intervals of rho

	Lower Limit	Upper Limit
0.95	0.115	0.595
0.99	0.025	0.65