The following plot shows the distribution of the average SAT I scores at 1158 US colleges.

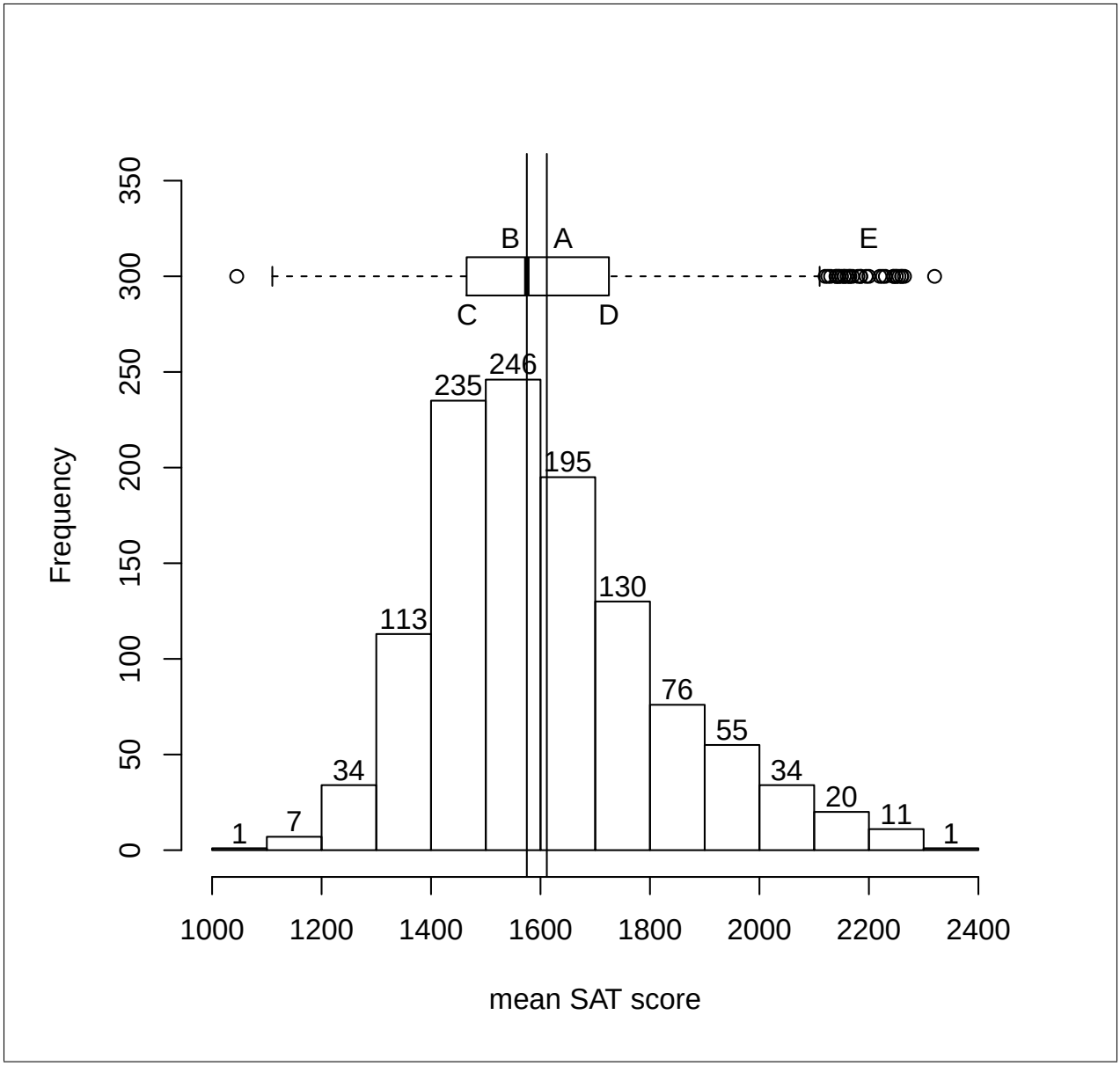

Figure 1: Mean SAT scores across US colleges

sum: 1866165 **variance:** 44380.36 **num observations:** 1158

Answer questions 1-11 using Fig. 1 and information provided above. Show all work for questions requiring calculation. The value of each question is shown in parentheses.

(1) What is the mean of this data? (5)

(2) What is the standard deviation of this data? (5)

(3) The letters A through E mark features of interest on the plot. What is the name of each? (10)

A:

B:

C:

D:

E:

What is...

(4) ...the name of the distance between points C and D? (2)

(5) ...the bin size of the histogram in Fig. 1 (2)

(6) Points A and B mark two measures of central tendency. In a few sentences, explain why they are not equal. (6)

Assume fig. 1 uses a representative sample of US colleges (there are actually around 4500). If I were to randomly name a US college, what is the ***probability*** *that the mean SAT score at the named school is...* (you do *not* need to use the PDF or CDF)

(7) ... greater than 1800? (6)

(8) ... less than 1400? (6)

(9) ... between 1400 and 1800? (6)

Let's say that we wish to model our sample as normally distributed.

(10) If we wanted to fit a theoretical density curve to these data, what 2 parameters must we calculate? (just name them – this is not a question about R code) (6)

(11) Are these data well approximated by a normal distribution? Explain why or why not in a few sentences. (6)

Now, we plot the same data against admission rate...

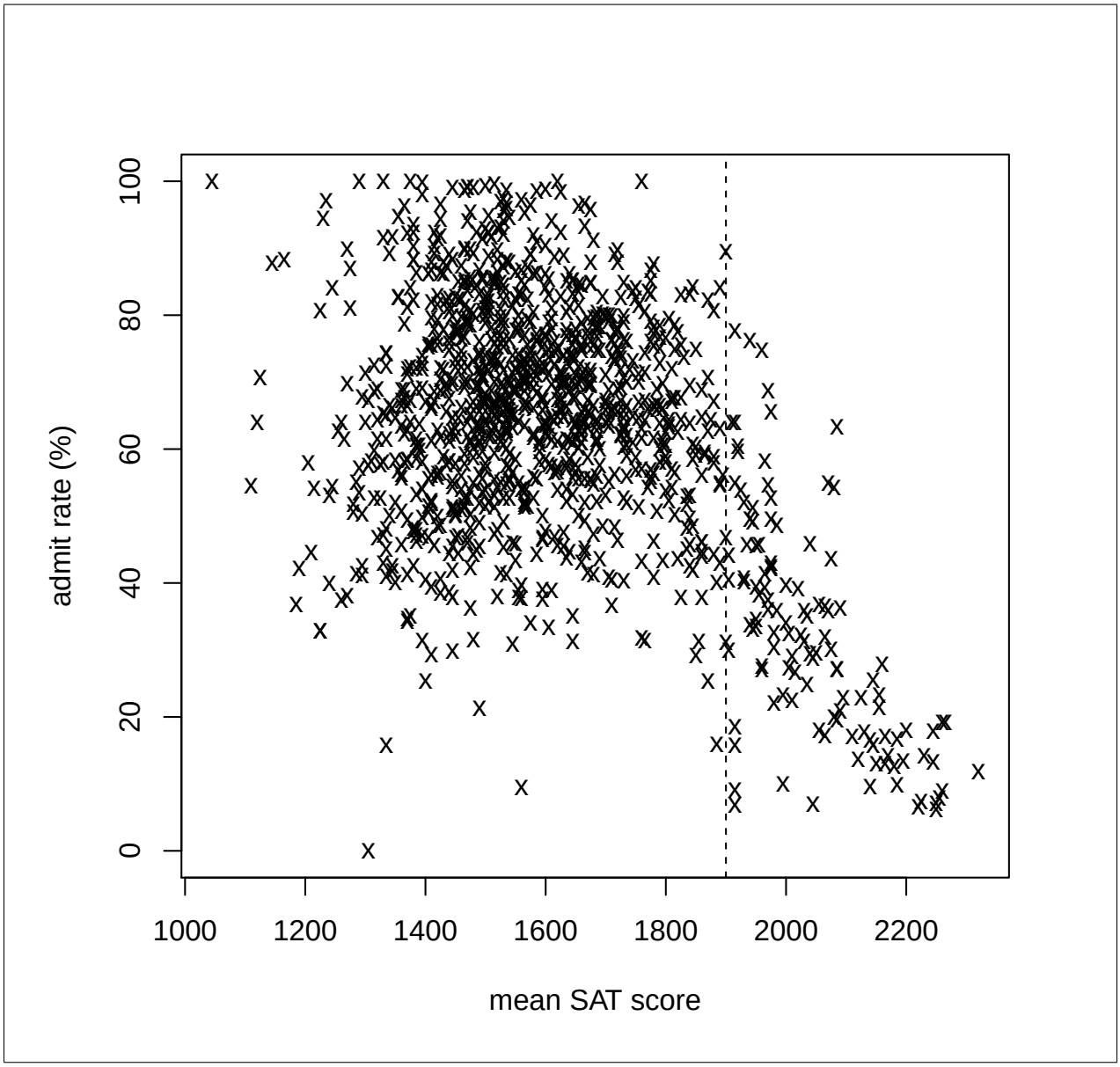

Figure 2: Admit rate versus mean SAT I score

variance of SAT scores:	**variance of admit rate:**	**mean of SAT scores:**
44380.36	360.93	1611.54

sum of SAT scores:	**covariance of admit rate and SAT score:**
1866165	-1756.05

Answer questions 12-15 using Fig. 2

(12) What is the overall correlation between SAT score and admit rate? (6)

Without doing any further math, *describe* what you think the correlation will be like for the following. Negative, close-to-zero, or positive? Big or small? How do you know?

(13) . . . schools with mean SAT $<$ 1900 (6)

(14) . . . schools with mean SAT $\geq$ 1900 (6)

(15) In a few sentences, explain why correlation/covariance do *not* provide a good summary of the *overall* bivariate relationship. (7)

Here's one possible explanation for Fig. 2

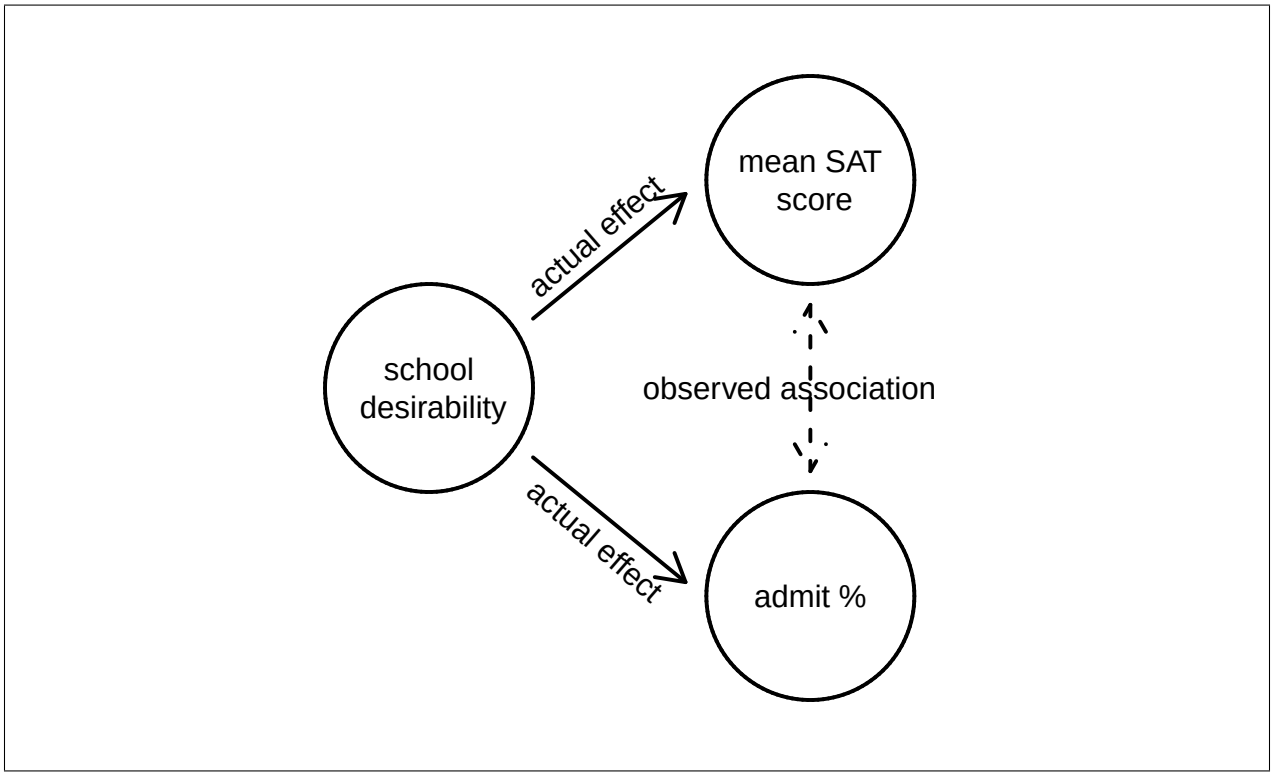

Figure 3: One possible explanation for observed association

(16) What is the general name for the situation in figure 3? In a few sentences, explain what happens in such a scenario. (15)