Fox Module 4 Bivariate displays practice problems on ozone scatter-plot matrices

(The attached PDF file has better formatting.)

** Exercise 4.1: Scatterplot matrix

The data are from a study of ozone levels in the atmosphere. The four variables in the scatterplot matrix are

- *rad:* solar radiation
- *temp:* daily temperature
- *wind:* wind speed
- *ozone:* ozone level in atmosphere

A. What are the ranges of the four variables?
B. What is the thin red line in each plot?
C. Two plots show the relation of temperature and solar radian. One plot shows a humped curve and the other plot shows a positively sloped curve. Explain what each curve implies.
D. Which variables are negatively correlated over their entire ranges?

Part A: The exact ranges are hard to read from the axes of the scatterplot matrix. They are

- Daily temperature ranges from 57 to 97 degrees.
- Wind speed ranges from 2.3 to 20.7 miles per hour.
- Solar radiation ranges from 7 to about 334 units.
- Ozone levels range from 1 to 168 units (particles in a given volume of air).

Jacob: Do we read the ranges along the horizontal axis or the vertical axis?

Rachel: For the panel with the name of the variable, the scales are the same along the two axes.

Take heed: Final exam problems about the ranges of the variables test if you look at the proper scales. They do not test the exact values.

Part B: The thin red line is a *loess* curve.

Jacob: The *loess* curve uses linear regression, but it is curved, not straight. Why is this?

Rachel: The linear regression differs at each point. For wind speed of 5, it may usespoints of wind speed from 0 to 10; at wind speed 10, it may use points of wind speed from 5 to 15.

Jacob: Random fluctuations might distort the loess curve. Suppose a wind speed = 11, ozone levels were high (by random fluctuation). For a regression from 1 to 11, β is high; for a regression from 0 to 10, β is lower.

Rachel: The loess curve uses weighted regressions. For a regression from 1 to 11, the point 11 receives little weight.

Part C: The humped curve in the first column and second row has solar radian as the explanatory variable and temperature as the response variable. For low solar radian, as solar radian increases, temperature increases; for high solar radian, as solar radian increases, temperature decreases.

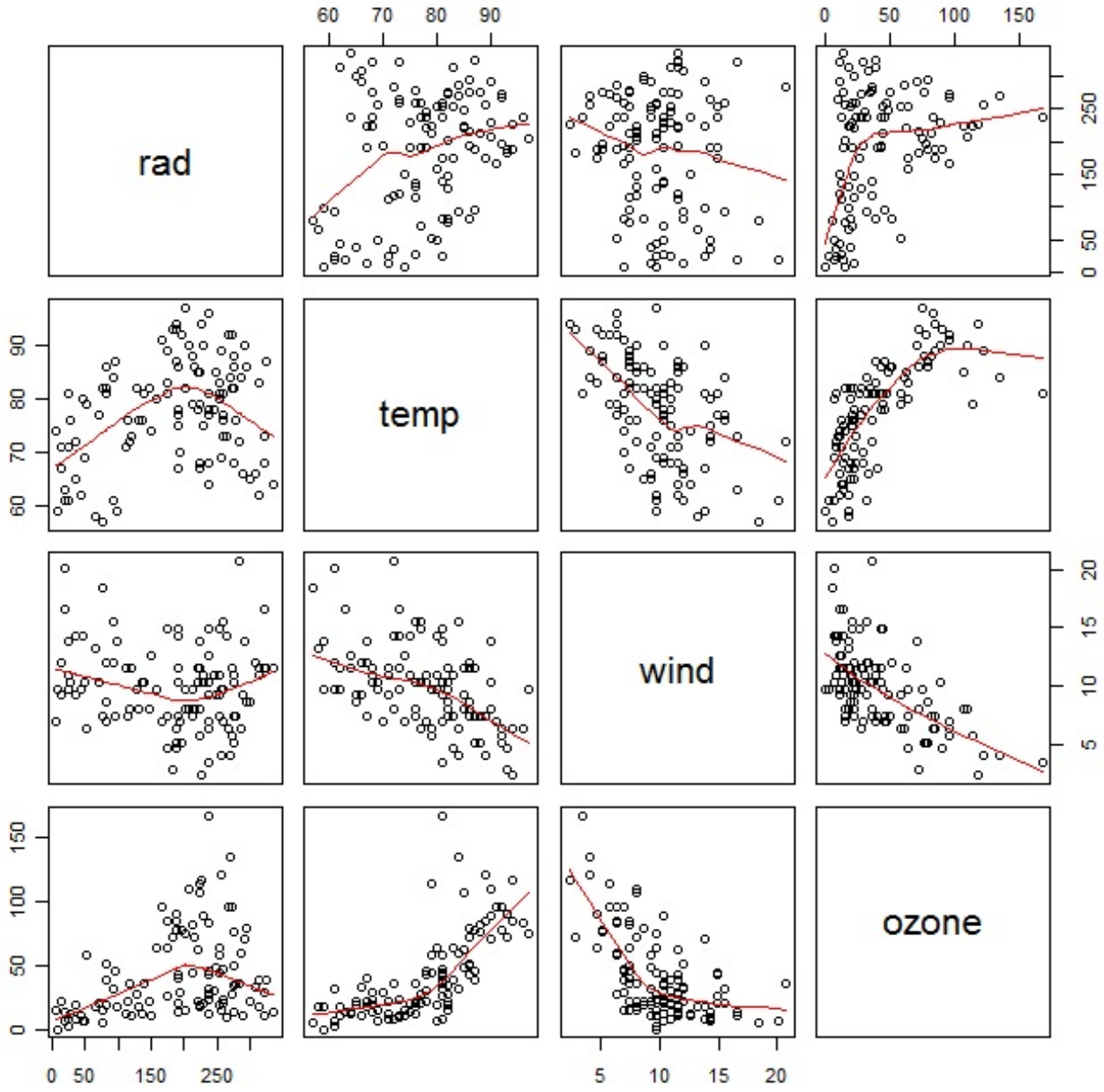
The positively sloped curve in the first row and second column has temperature as the explanatory variable and solar radian as the response variable. As temperature increases, solar radian increases.

Jacob: Does temperature cause the change in solar radiation, or does solar radiation cause the change in temperature?

Rachel: The graph does not imply causation. Neither directly affects the other.

Part D: Temperature and wind speed are negatively correlated. In California, when wind speeds are higher, the days are cooler.

Wind speed and ozone levels are also negatively correlated, at least at lower wind speeds. Higher wind disperses the ozone.


rad
temp
wind
ozone
60 70 80 90
0 50 100 150
0 50 150 250
0 50 100 150
60 70 80 90
5 10 15 20
0 50 150 250
5 10 15 20