Quality Control Case

Everybody seems to disagree about just why so many parts have to be fixed or thrown away after they are produced. Some say that its the temperature of the production process, which needs to be held constant (within a reasonable range). Others claim that its clearly the density of the product, and that if we could only produce a heavier material, the problems would disappear. Then there is Ole, who has been warning everyone forever to take care not to push the equipment beyond its limits. This problem would be the easiest to fix, simply by slowing down the production rate; however, this would increase costs. Interestingly, many of the workers on the morning shift think that the problem is those inexperienced workers in the afternoon, who, curiously, feel the same way about the morning workers.

Ever since the factory was automated, with computer network communication and bar code readers at each station, data have been piling up. You have finally decided to have a look. After your assistant aggregated the data by 4-hour blocks and then typed in the AM/PM variable, you found the following description of the variables:

- `temperature`: measures the temperature variability as a standard deviation during the time of measurement
- `density`: indicates the density of the final product
- `rate`: rate of production
- `am`: 1 indicates morning and 0 afternoon
- `defect`: average number of defects per 1000 produced

```
quality = data.frame(
temp=c(.97,2.85,2.95,2.84,1.84,2.05,1.5,2.48,2.23,3.02,2.69,2.63,1.58,2.48,2.25,
2.76,2.36,1.09,2.15,2.12,2.27,2.73,1.46,1.55,2.92,2.44,1.87,1.45,2.82,1.74),
density=c(32.08,21.14,20.65,22.53,27.43,25.42,27.89,23.24,23.97,19.45,23.17,
22.7,27.49,24.07,24.38,21.58,26.3,32.19,25.73,25.18,23.74,24.85,30.01,
29.42,22.5,23.47,26.51,30.7,22.3,28.47),
rate=c(177.7,254.1,272.6,273.4,210.8,236.1,219.1,238.9,251.9,281.9,254.5,265.7,
213.3,252.2,238.1,244.7,222.1,181.4,241,226,256,251.9,192.8,223.9,260.0,236,
237.0,221,253.2,207.9),
am=c(0,1,1,1,0,1,0,0,0,1,1,1,0,0,0,1,1,0,0,0,1,1,0,1,1,0,0,1,1,0),
defect=c(.2,47.9,50.9,49.7,11,15.6,5.5,37.4,27.8,58.7,34.5,45,6.6,31.5,23.4,
42.2,13.4,0,20.6,15.9,44.4,37.6,2.2,1.5,55.4,36.7,24.5,2.8,60.8,10.5)
)
```

Questions:

1. Generate a scatterplot matrix and a correlation matrix. Interpret the correlations. What obvious conclusions can you draw?

2. Run a multiple regression predicting defect rate from the other four variables. Is the overall model significant? Which predictors, if any, are significant? Compute and interpret variance inflation factors. What obvious conclusions can you draw?

3. Predict defect from each of the predictor variables separately, e.g., defect from temp, defect from density, defect from rate, etc. Which of the predictors are significant in the simple linear regressions?

4. Would it be appropriate to regress defect on the first component?

5. Perform further analysis as needed. What action do you recommend? Why?

6. How would you present your findings to a client?