

# STAT 330

## Homework 7.2

### 1) Return to College Ranking Data:

- a. Let's use the college ranking data from Homework #2 again. I am giving you a SAS version of the data (*colleges.sas7bdat*).

Read in the data and create 2 new variables: a state variable and a dummy variable for whether the school is public or private. Both of these variables can be computed from the name variable. Colleges with names that have asterisks at the end are public schools.

- b. Generate a frequency distribution for the state variable. Save the frequencies to a SAS data set.
- c. I want a SAS data set that for all 51 states (the usual 50 and DC) has a rank variable. That is, NY has the most colleges in the list; its rank would be 1. WY has the least, and its rank would be 51 (also because alphabetically its toward the bottom). Use the data set created in (b) to create this new data set using PROC and DATA steps. My version of the resulting data set is *outStates.sas7bdat*.
- d. Sensibly combine the data set from (c) with the original data. I called this *use.sas7bdat*.
- e. Replicate the graphs found in *HW7 Output.pdf*.
- i. Horizontal bar chart for public/private variable.
  - ii. Distribution of public/private variable by state, but only for the top 10 ranking states, according to our ranking variable created in (c). Limit the data processing in the PROC to the top 10 states. Do not use another DATA step to subset the data.
  - iii. Average college ranking score (*score*) by school type.
  - iv. A no-frills scatter plot of graduation rate (*predGradRate*) by score. It looks like there could be a bit of an upward trend relationship, though there's a lot of random variation.
  - v. Redo the scatter plot and add a LOESS curve, a smooth spline, which can indicate whether there is a relationship of sorts.
- f. Edit your code so that all output gets saved to a PDF file. That's what I did to create your *HW7 Output.pdf*.