Applied and Mathematical Statistics
Homework Five

Figure 1: A tractor pulling a chisel plow in Slovenia. Figure source: Petar Milošević `http://en.wikipedia.org/wiki/File:Traktor_na_oranju.JPG`.

Please do these explorations before class on Thursday.
These problems concern the dataset found in the file `bank_data.csv` on the course web page. This is data from the world bank for the year 2000. The columns contain:

- `GNI`: Gross national income per capita
- `tractors`: number of tractors per 100 sq. km of arable land
- `male_lit`: male literacy rate
- `female_lit`: female literacy rate
- `total_lit`: total literacy rate

Note: I removed a few outliers: Iceland, Lichtenstein, Monaco, and a few other little counties. Iceland has an insane number of tractors—over 12,000 per 100 km^2—which seems impossible.[1] Lichtenstein, Monaco, etc., have a really large GNI so I excluded them.

We will use a linear regression to investigate the following questions:

1. How does GNI depend on female literacy rate?
2. How does GNI depend on male literacy rate?
3. How does GNI depend on tractors?
4. How does female literacy rate depend on male literacy rate?

For each question, do the following:

1. Plot the data
2. Form a linear model
3. Plot the regression line on the same plot as the data
4. Look at a summary of `lm` and interpret: How good is the line? Is there a statistically significant relation between the two variables? What does the relationship mean?

[1] Is there even *any* arable land in Iceland? What do Icelanders do with so many tractors?