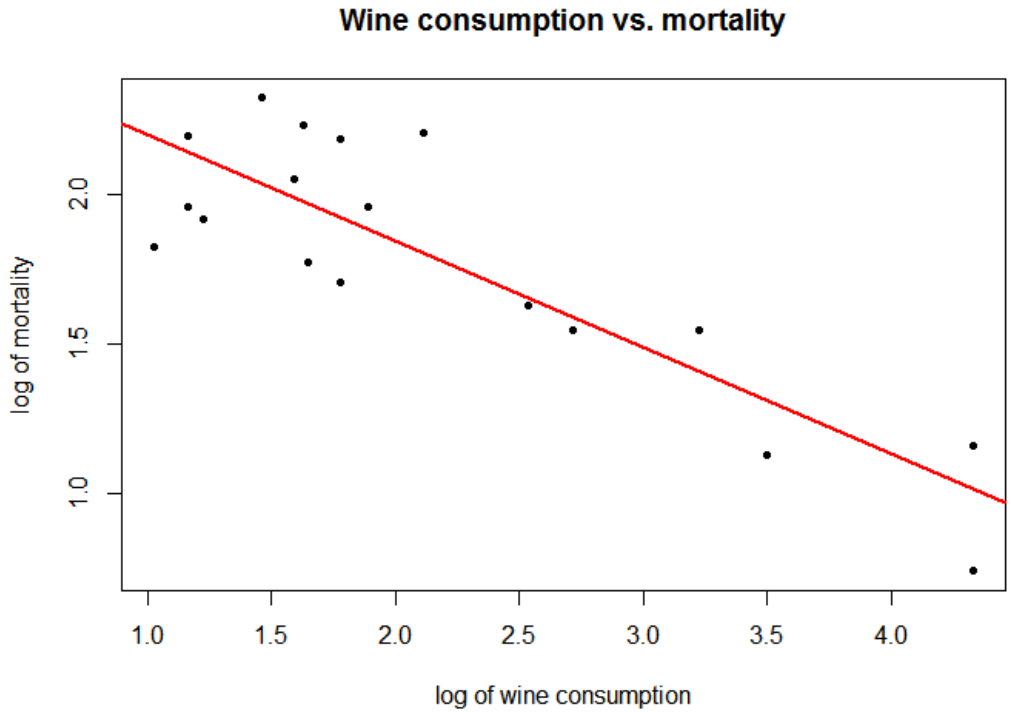
Correlation does NOT imply causation!

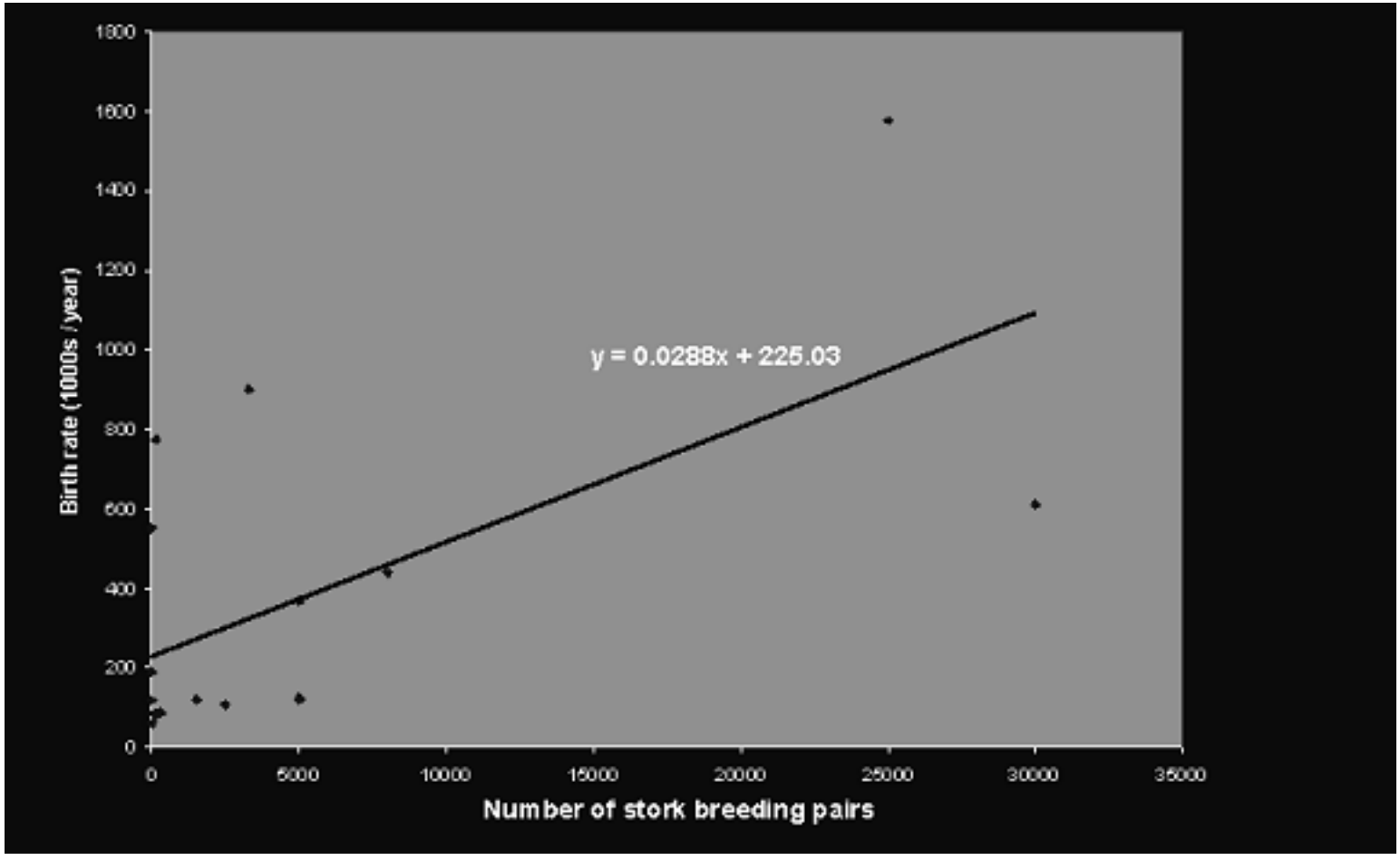
Subho Majumdar

March 4, 2014

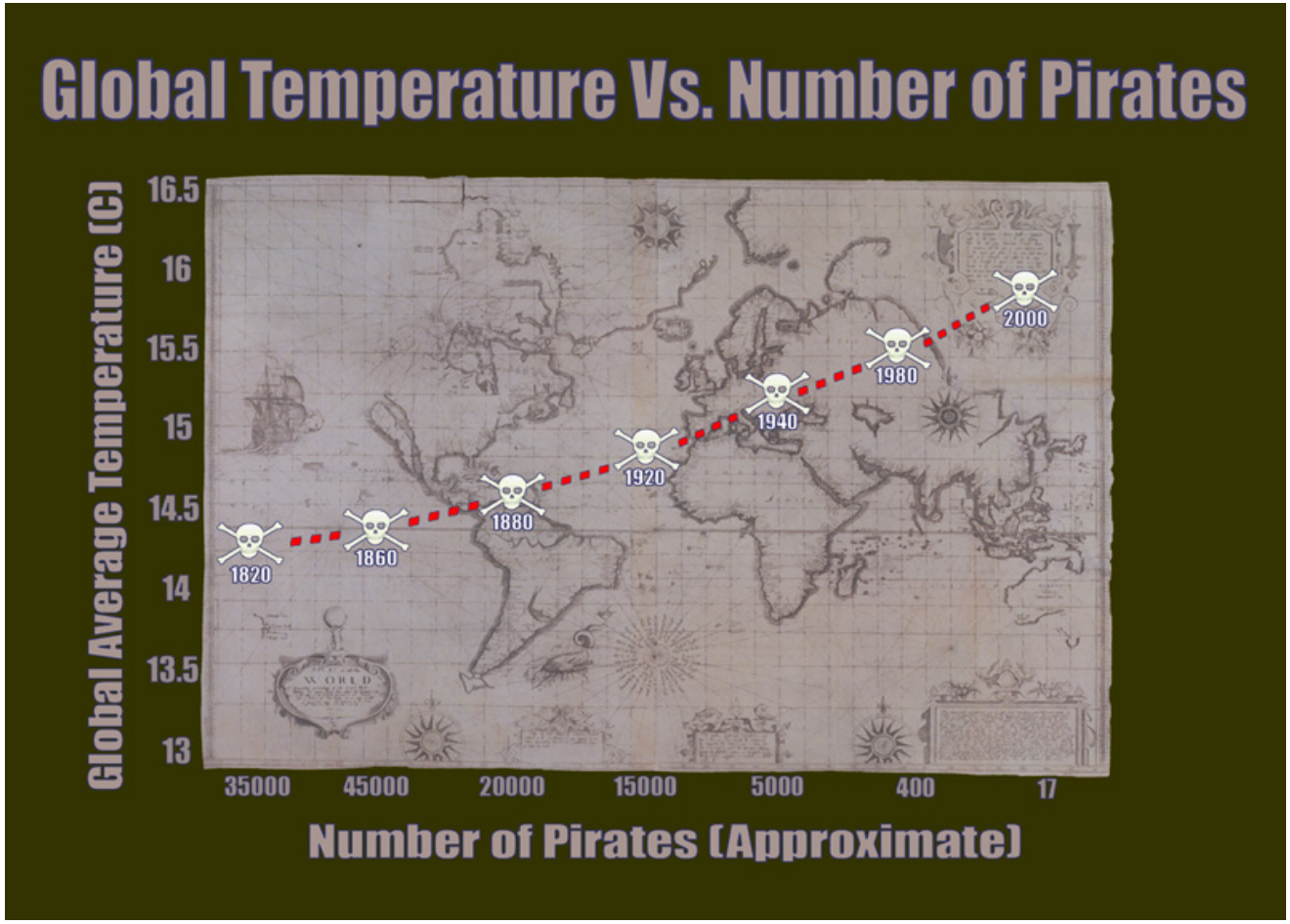
From 8.23 in your hw3 we have a nice negative correlation between the two variables, and it basically says that the more wine you drink, to longer you're likely to live. So drink wine to live long, right?



Because we also know that more storks = more babies!



Or because of the fact that number of pirates are coming down because of global warming!



Or more chocolate consumption means more Nobel laureates!!

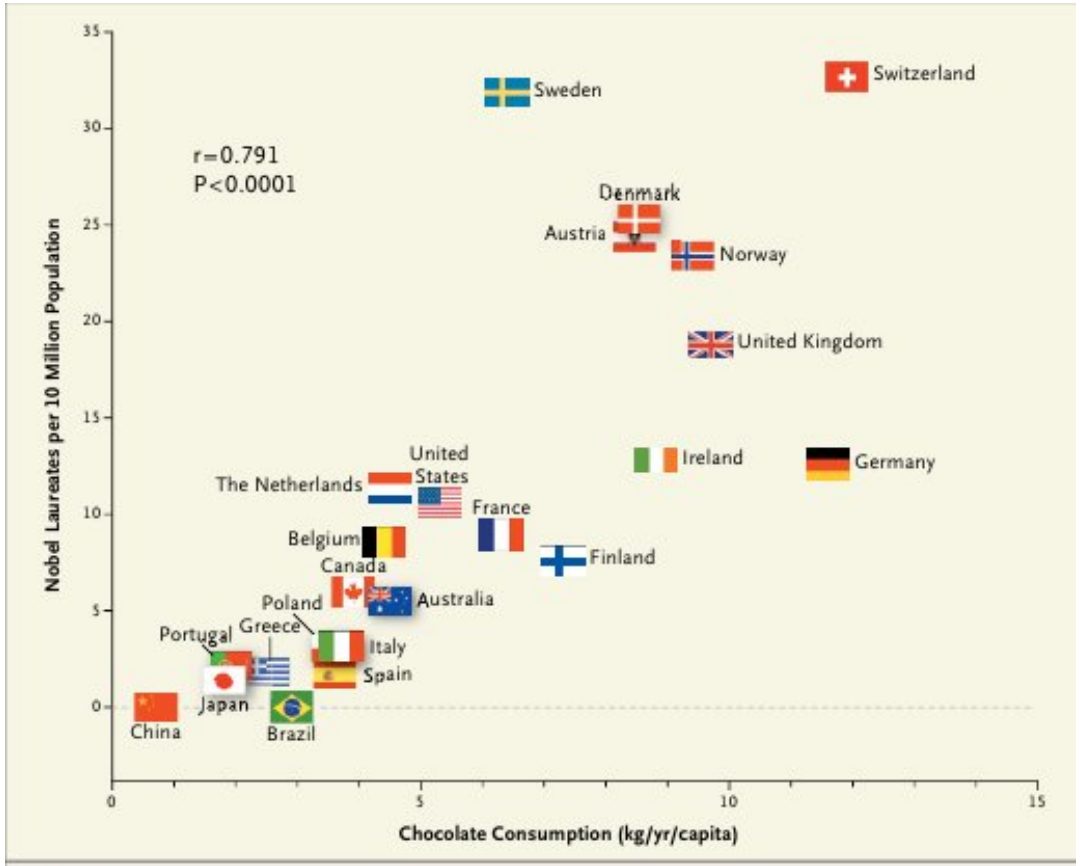


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

So what's going wrong in all these examples?

The problem is that all of these are observational data, and we have no control over the factors affecting the variables in the data we have.

To establish a cause-effect relationship we need **experiments**, where we can control for external factors.

Definitions

Experimental study

In an **experiment** investigators apply treatments to experimental units (people, animals, plots of land, etc.) and then proceed to observe the effect of the treatments on the experimental units.

Observational study

In an **observational study** investigators observe subjects and measure variables of interest without assigning treatments to the subjects. The treatment that each subject receives is determined beyond the control of the investigator.

Example

To establish the effect of smoking on lung capacity, you need to an actual experiment (left), not just take observations from random people(right).

Experiment

- Find 100 women age 20 who do not currently smoke.
- Randomly assign 50 of the 100 women to the smoking treatment and the other 50 to the no smoking treatment.
- Those in the smoking group smoke a pack a day for 10 years while those in the control group remain smoke free for 10 years.
- Measure lung capacity for each of the 100 women.

Observational study

- Find 100 women age 30 of which 50 have been smoking a pack a day for 10 years while the other 50 have been smoke free for 10 years.
- Measure lung capacity for each of the 100 women.

Sources

- Google: 'correlation does not imply causation'
- http://www.public.iastate.edu/~dnett/S401/nexpvsobs.pdf