DS Visualisation and Analysis

Abbey Waldron

This Week - Correlations and Functional Forms

1. Show last week's work to the class
2. What is a correlation?
3. Some maths…describing (functional) relationships
4. Start work on problems

Random Student Generator

Week 2 Problems: West Africa Ebola

Make one plot to answer each question.

1. Which country has the most rapid initial rise in ebola cases?
2. How does the total number of ebola deaths change over time?
3. When do new country infections occur?
4. What are the per country confirmed case totals per month in 2014?
5. Make one more plot to show something interesting.

Continuing your R education...

Sometimes the easiest way to perform these tasks is with looping:

- http://blog.datacamp.com/tutorial-on-loops-in-r/

Correlation

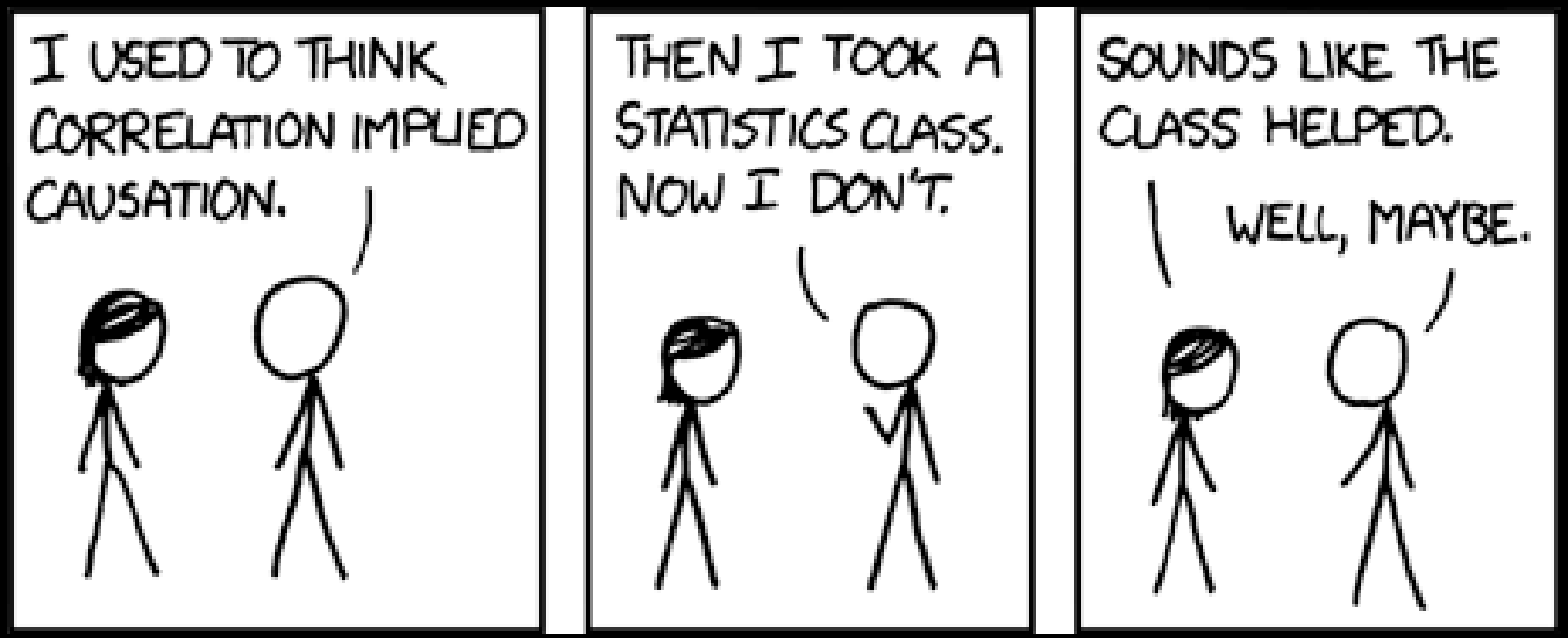
Correlation

Strong/Weak?

Direction?

Linear/Non-linear…?

Correlation

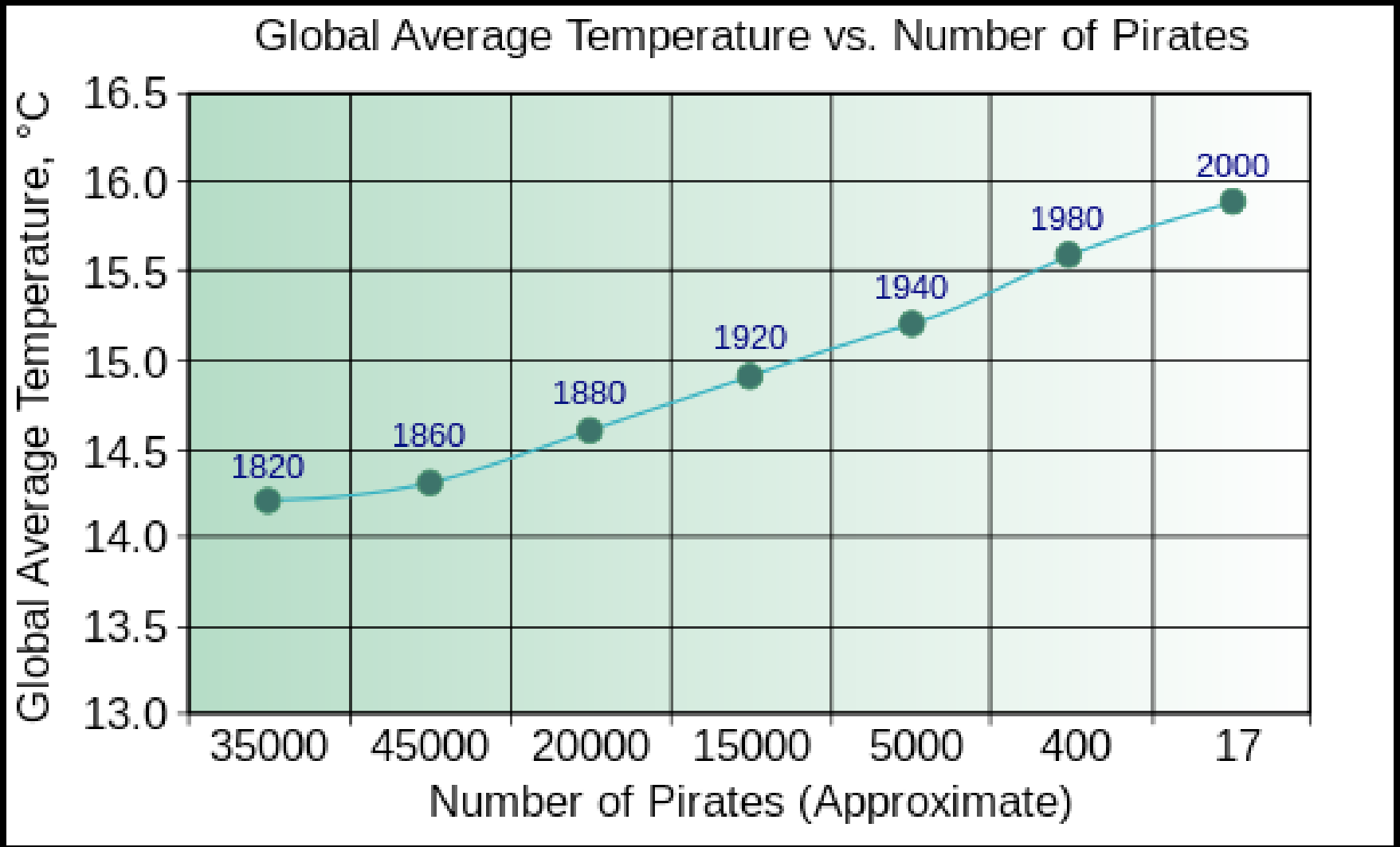


xkcd.com/552

Causation

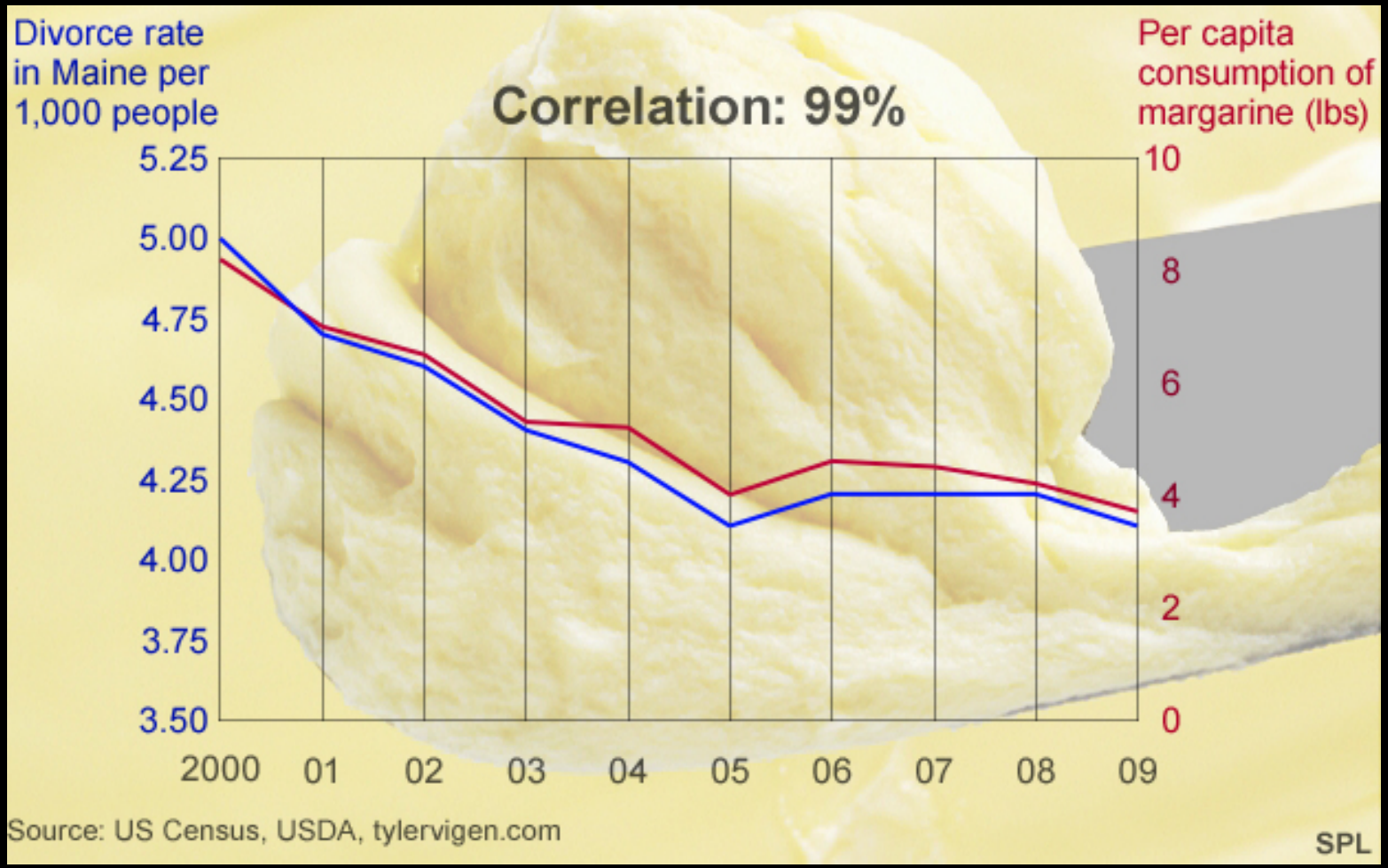
CORRELATION DOES NOT IMPLY CAUSATION

Causation?



(Thank you Wikipedia!)

Causation?


Divorce rate in Maine per 1,000 people
Correlation: 99%
Per capita consumption of margarine (lbs)
5.25
5.00
4.75
4.50
4.25
4.00
3.75
3.50
10
8
6
4
2
0
2000
01
02
03
04
05
06
07
08
09
Source: US Census, USDA, tylervigen.com
SPL

Types of Causation

If A and B are correlated then there are different possibilities of causation:

- A causes B
- B causes A
- C causes A and B (lurking factor)
- A causes C which causes B (or vice versa)
- A causes B and B causes A (cyclic or bidirectional)
- there is no connection between A and B only coincidence

How to discover correlations?

How to discover correlations?

Make some plots!! Of course…

- Typically make scatterplots of each pair of variables
- Can you see a relationship? If not then not strong.
- Describe relationships using functions (not only linear!)

R makes this easy…

Look at `pairs()`

Functions

Functions

In general, plot the dependent values on the y (vertical) axis and the independent values on the x (horizontal) axis, but the difference is not always there.

$$y = f(x) = ...$$

Function Examples

- linear, quadratic, cubic…polynomial
- exponential, logarithmic
- Gaussian
- etc.

Why we need to be quantative…

Later on we are going to try to use some variables to predict others, this requires fitting a sensible function to the available data.

These problems come in two main categories:

1. You have a theoretical model for how the variables should be related
2. You have no theoretical model for the relationship and you have to guess somehting from the data
3. Some combination of the two due to *e. g.* unexpected noise.

The second case: guessing functions

Often there is not a single right answer (theme of this course…)

Which function is good enough?

- Needs to describe the major features of the data
- Should be minimal - as simple as will work
- May well not be unique, you can try fitting different functional forms and see which one works best
- We will start the actual fitting next week

What is a feature?

Things to look for and check match:

- behaviour as x $\rightarrow \pm\infty$
- turning points (gradient = 0)
- crossing points with the axes

Week 3 Problems

1. Finding and describing correlations
2. Functional forms

Backup Slides