There are three types of lies
   — lies, damned lies and statistics
Benjamin Disraeli

- British prime minister (Tory).
William Gladstone

- Defeated Disraeli in the general election of 1868.
- President of the Royal Statistical Society 1867-1869.
That question is this: Is man an ape or an angel? I, my lord, I am on the side of the angels. I repudiate with indignation and abhorrence those new fangled theories.

(Oxford Diocesan Conference 25/11/1864)
Eugenics is the self direction of human evolution.

Like a tree, eugenics draws its materials from many sources and organizes them into a harmonious entity.
This monument was erected in remembrance of the destruction of the South Sea in 1720.
A rational approach to uncertainty?
Absorption spectra
Is abstraction the problem?
Baker & Bellis, 1993, Animal Behaviour

- **count**: Graph showing data points across different values.
- **prop.partner**: Graph showing data points across different values.
- **time.ipc**: Graph showing data points across different values.
The Baker and Bellis Analysis
At the end of the process they asked whether the apparent straight line relationships were stronger than could plausibly have arisen by chance. On this basis they concluded that there is evidence for count declining with proportion of time spent together. Time since last copulation seemed not to play a detectable role. But they also collected another dataset...
More conclusions. . .

- Going through the same process as with the first data set, leads to the conclusion that only female weight is linearly related to count.
- But a careful look at the residuals shows that this conclusion is completely dependent on a single data point with very low sperm count.
- Re-do the analysis without this datum, and only volume matters.
- Actually it’s the same subjects in both datasets, and we can match up the volumes with the first dataset.
- Repeating the first analysis with volume added, leads to the dull conclusion that there is only any evidence for a linear relationship between count and volume.
- This result has limited marketing potential.
But why straight lines anyway?
Smoothing

1. What if the relationship between the residuals and a variable does not look like a straight line?
2. Why not let it be a smooth curve, instead?
How to choose the best fit curve?

- Take a bendy strip of wood.
- Hook it up to the data points with springs.
- The result is a **spline**
Splines are controllable

- Changing the flexibility of the spline changes the curve.

- Splines can be described mathematically, in a way that is easy to work with.
Smooth surfaces: thin plate splines

- For smooth surfaces there are several options
- We can replace the bendy strip, with a bendy sheet...
More smooth surfaces: tensor product splines

- Or we can make a surface from a lattice of bendy strips.
- The strips should usually have different degrees of flexibility in the two directions.
Yet more smooth surfaces: soap films

- For smoothing within oddly shaped areas, it can help to replace bendy sheets/strips, with a soap film.
- This avoids smoothing across the area boundary.
How flexible should the spline be?

- Mathematically, all these ways of describing a surface, have the degree of smoothness controlled by just one or two numbers . . .
- . . . which must be chosen. How?

![Graphs showing the effect of λ too high, about right, and too low on the smoothness of a spline fitting.](attachment:image.png)
Cleaning up a brain scan

Model log FPQ as a smooth surface, represented using a thin plate spline.

Springs attaching the plate to the data have strength dependent on the height of the plate.
Is Cairo getting hotter?

A model . . .

- The temperature varies smoothly with day of year.
- There *might* be an additional smooth long term trend in temperature.
- The small scale day to day fluctuations are probably correlated between one day and the next.
Yes it is.
How can we predict the octane rating from the spectrum?
Octane prediction model

Model: octane rating is a constant plus the average value of the red curve multiplied by the spectrum (blue).

Need to estimate the red curve.
Octane prediction fit

Estimated function

octane

measured

fitted
Diabetic Retinopathy Study

Model is that probability of retinopathy is related to a sum of smooth curves depending on \( \text{bmi}, \text{gly} \) and \( \text{dur} \) plus smooth surfaces depending on \( \text{bmi} \& \text{gly}, \text{gly} \& \text{dur} \) ...
Diabetic Retinopathy Results

\[ s(\text{dur}, 3.26) \]
\[ s(\text{gly}, 1) \]
\[ s(\text{bmi}, 2.67) \]

\[ \text{te}(\text{dur}, \text{gly}, 0) \]
\[ \text{te}(\text{dur}, \text{bmi}, 0) \]
\[ \text{te}(\text{gly}, \text{bmi}, 2.5) \]
Diabetic Retinopathy Results II

\[ \text{bmi} \]
\[ \text{gly} \]
\[ \text{linear predictor} \]

\[ 15 \quad 20 \quad 25 \quad 30 \quad 35 \quad 40 \quad 45 \quad 50 \]

\[ 10 \quad 15 \quad 20 \]

\[ \text{bmi} \]
\[ \text{gly} \]
\[ \text{linear predictor} \]

\[ \text{red/green are } +/\pm \text{ TRUE s.e.} \]
R is a collaborative project with many contributors. Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[Previously saved workspace restored]

> library(mgcv)
This is mgcv 1.3-17
> model <- gam(y ~ s(x0) + s(x1) + s(x2) + s(x3))
> plot(model, pages=1)
> X11()
> plot(fitted(model), residuals(model))
> model

Family: gaussian
Link function: identity

Formula:
  y ~ s(x0) + s(x1) + s(x2) + s(x3)

Estimated degrees of freedom:

  s(x0) s(x1) s(x2) s(x3) Total

  5.172972 2.35135 8.88174 2.14572 12.25528

GCU score: 4.610979

>
Gladstone and Disraeli are from the House of Commons web site.
The 1921 Eugenics conference logo is from en.wikipedia.org/wiki/File:Eugenics_congress_logo.png
The Gates of Auschwitz are from oncampus.richmond.edu/academics/education/projects/webquests/holocaust/images/arbeit_macht_frei.jpg
Hogarth’s South Sea Bubble can be found at www.library.hbs.edu/hc/ssb/images/using-top.jpg, but I’ve lost where I found the one shown.
The absorption spectrum figure is from www.te-software.co.nz/blog/augie-auer.htm
Reproductions of Picasso’s Les Demoiselles d’Avignon are available from many sites. The one shown is possibly from www.enjoyart.com/library/featured_artists/pablopicasso/large/Bmcgaw-P591.jpg
The cover of Sperm Wars was taken from www.amazon.co.uk.
Data Credits

- The Global CO2 and temperature data are from www.cru.uea.ac.uk/cru/data/temperature/ and the Scripps Institute CO2 research group.
- The Aral Sea CO2 data are from the SeaWifs satellite.
- For full credits for the Cairo and Brain Scan data, see R package `gamair`.
- The octane data are available in R package `pls`.
- The Retinopathy data are available in R package `gss`.