



BUILDING SCIENCE
IN AFRICA

AIMS VIDEO COURSES
SUPPORTING BOOKLET

PROBABILITY & STATISTICS

WITH
PROF DAVID SPIEGELHALTER

AIMS
SOUTH AFRICA



African Institute for Mathematical Sciences

6 MELROSE ROAD | MUIZENBERG | CAPE TOWN 7945 | SOUTH AFRICA

TEL: +27 (0)21 787 9320 | FAX: +27 (0)21 787 9321

EMAIL: info@aims.ac.za | WEB: www.aims.ac.za

AIMS Online Courses

The mission of the AIMS academic programme is to provide an excellent, advanced education in the mathematical sciences to talented African students in order to develop independent thinkers, researchers and problem solvers who will contribute to Africa's scientific development.

Teaching at AIMS is based on the principle of learning and understanding, rather than simply listening and writing, during classes, and on creating an atmosphere of increasing our knowledge through class discussions, through small group discussions, by formulating conjectures and assessing the evidence for them, and sometimes going down wrong paths and learning from the mistakes that led us there. The essential features of the classes at AIMS are that, in contrast to formal lecture courses, they are highly interactive, where the students engage with the lecturer throughout the class time, are encouraged to learn together in a journey of questioning and discovery, and where lecturers respond to the needs of the class rather than to a pre-determined syllabus. AIMS teaching philosophy is to promote critical and creative thinking, to experience the excitement of learning from true understanding, and to avoid rote learning directed only towards assessment.

Leading international and local experts offer the courses at AIMS, which are three weeks long (each module consisting of 30 hrs) and collectively form the coursework for a structured masters degree which also includes a research component. The advertised content is a guide, and the lecturers are encouraged, and indeed expected, to adapt daily to meet the current needs of the students.

Over the past ten years AIMS has achieved international recognition for this innovative and flexible approach. It has been the starting point for the remarkable success of our students and alumni and we all benefit from the support of many who have "witnessed the AIMS-magic and keep coming back for more."

This year we have decided to film selected courses and to make them available to a larger audience as an online facility. African universities may choose to use these courses to supplement and enhance their own postgraduate programmes. We believe this would be best achieved through engagement with AIMS. One way for this to happen, would be for AIMS to suggest or nominate a specialist tutor to spend time at the university, guiding students who follow the online programme. Where possible expert lecturers who have taught at AIMS may visit the university to give a short introduction to the course. We would welcome this interaction as well as the contribution our online courses will make to the growth of the mathematical sciences ecosystem in Africa.

Barry Green
Director & Professor of Mathematics
African Institute for Mathematical Sciences
January 2013

AIMS Council

Ramesh Bharuthram (University of the Western Cape) Hendrik Geyer (Stellenbosch University) Barry Green (AIMS) Grae Worster (Cambridge University) Daya Reddy (University of Cape Town)
Graham Richards (Oxford University) Stephané Ouvry (Université de Paris Sud XI) Tsou Sheung Tsun (Oxford University) Neil Turok (Perimeter Institute)

PROBABILITY & STATISTICS
2012

PROF DAVID SPIEGELHALTER
DAY 6



AIMS
SOUTH AFRICA

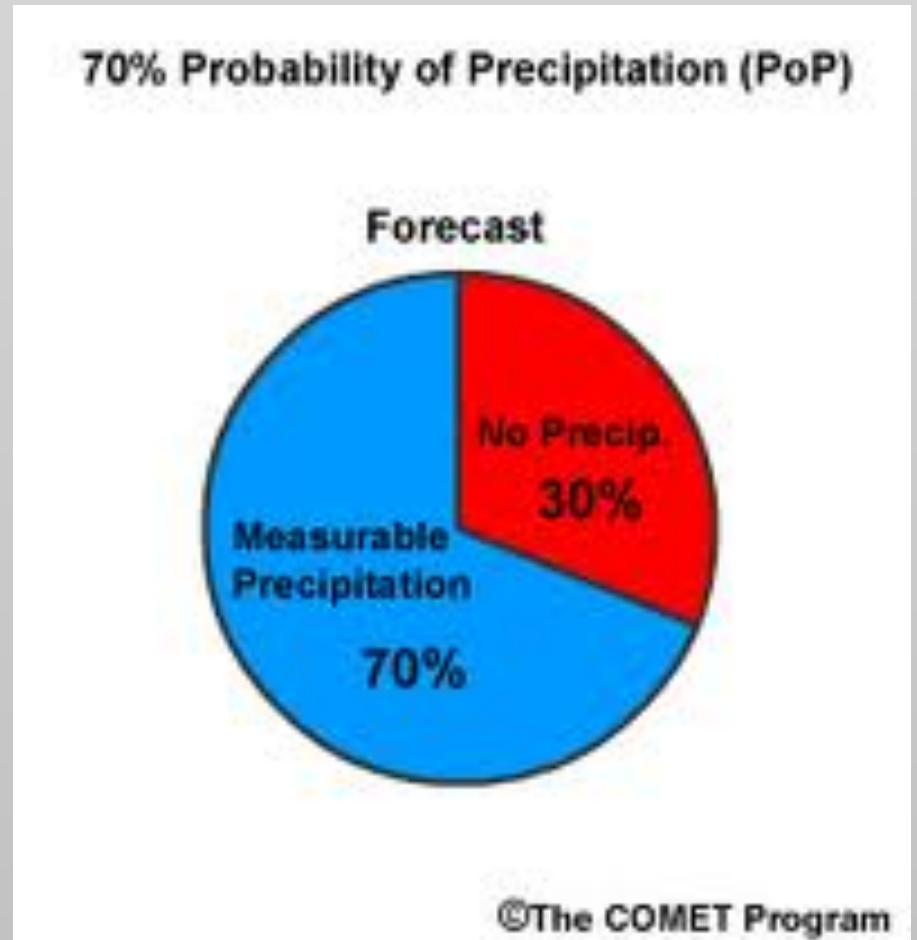
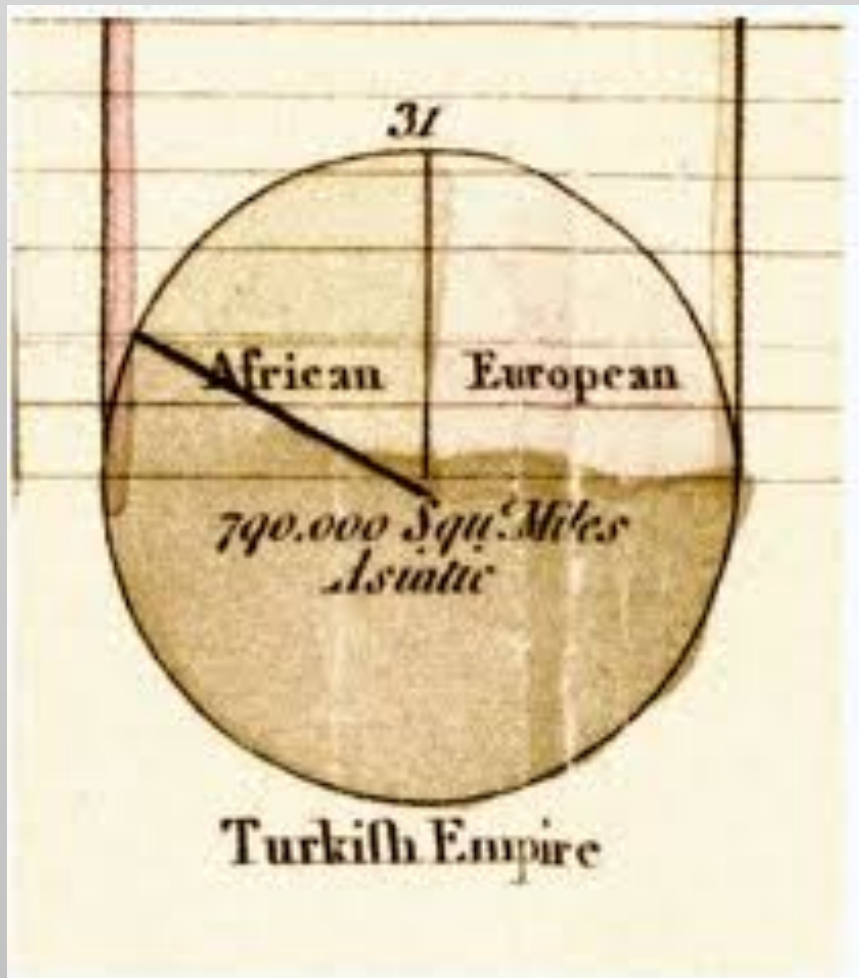
Stages in statistical discovery

- . Important problem
- . Good survey / experimental design
- . Good quality data
- . **Data exploration and visualisation**
- . Modelling data as random variables arising from some distribution
- . Formal statistical inference about the 'truth'

Pie-charts

(William Playfair 1801)

- good for 'part-to-whole' comparisons



2012 PRESIDENTIAL RUN

GOP CANDIDATES

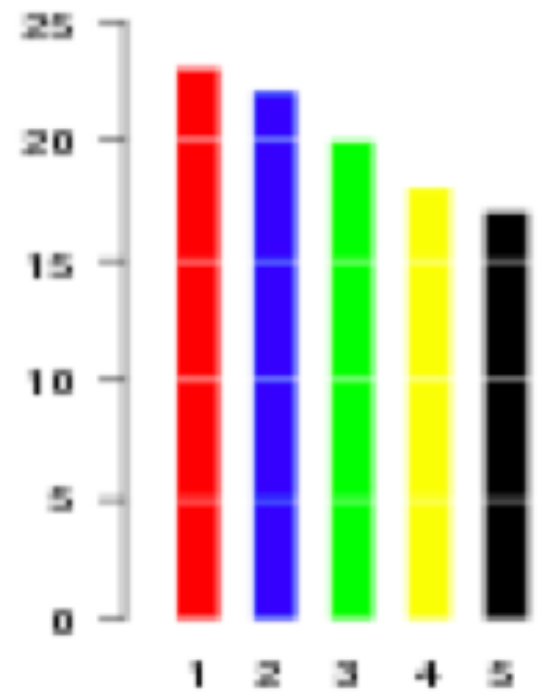
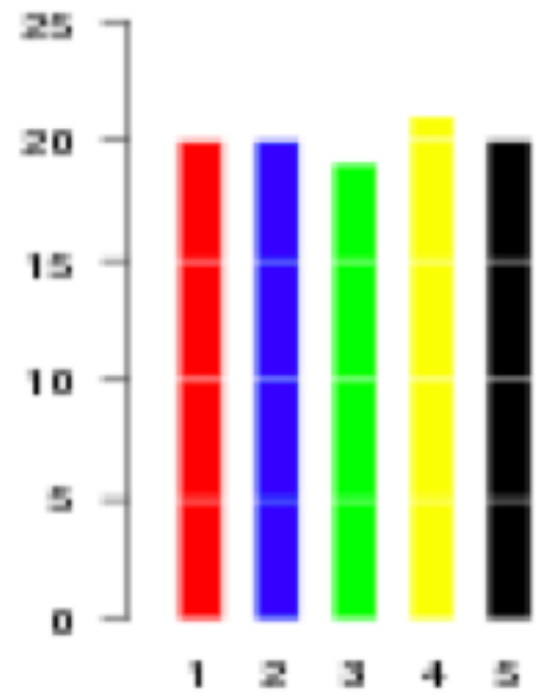
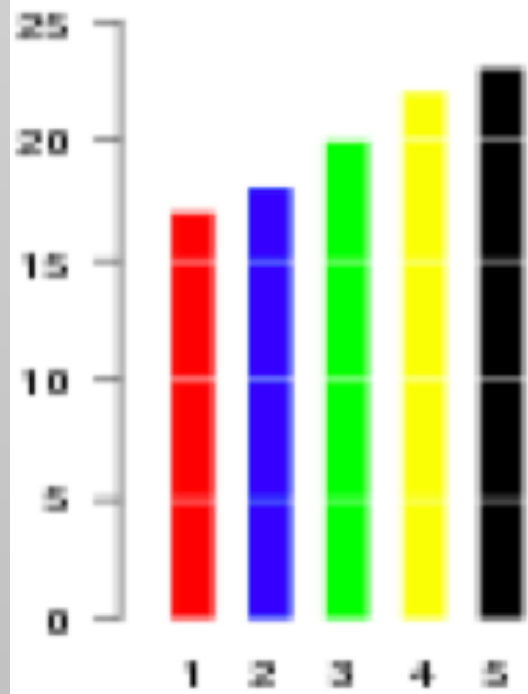


FOX

47'

SOURCE: OPINIONS

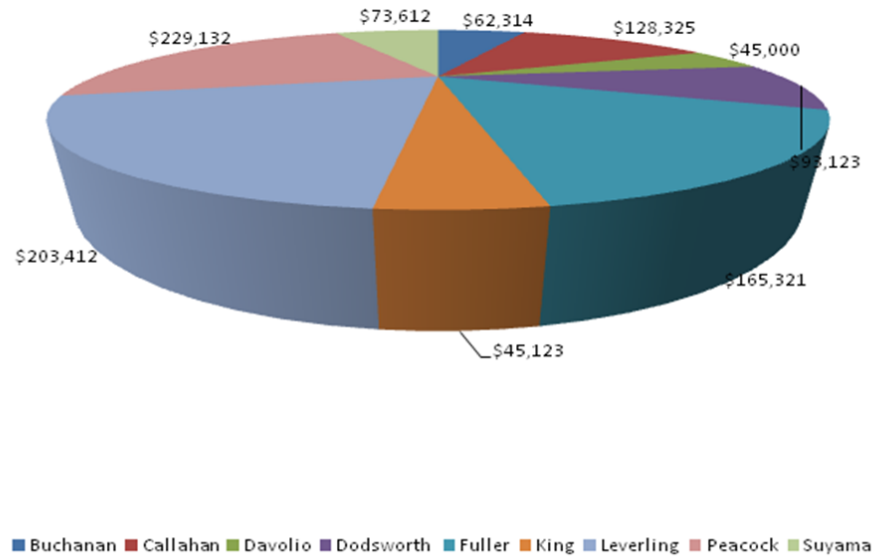
DYNAMIC

A**B****C**

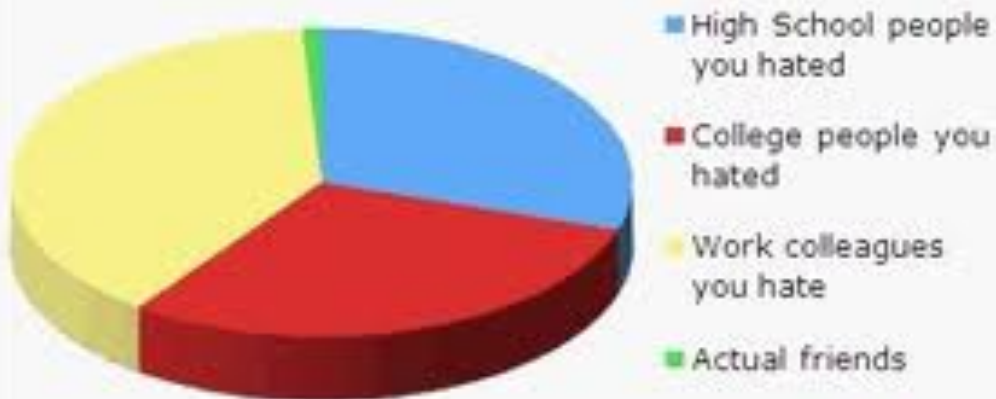
Pie charts bad for comparing areas – these look very similar

3-d pie charts?

Order Amounts by Salesperson



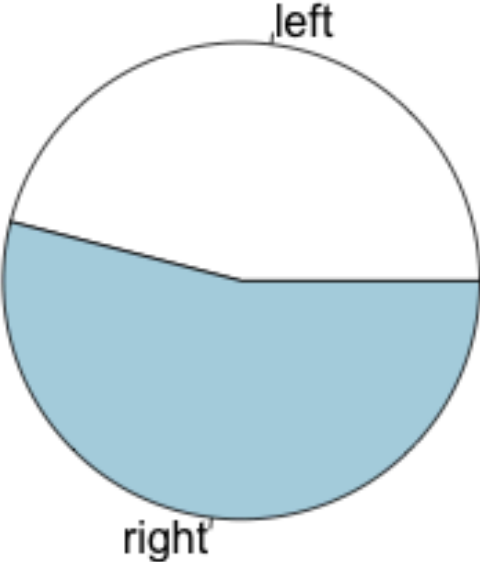
People who find you on Facebook





```
tabArmCross = table(ArmCross)
pie(tabArmCross,main="Arm on top when crossing")
```

Arm on top when crossing



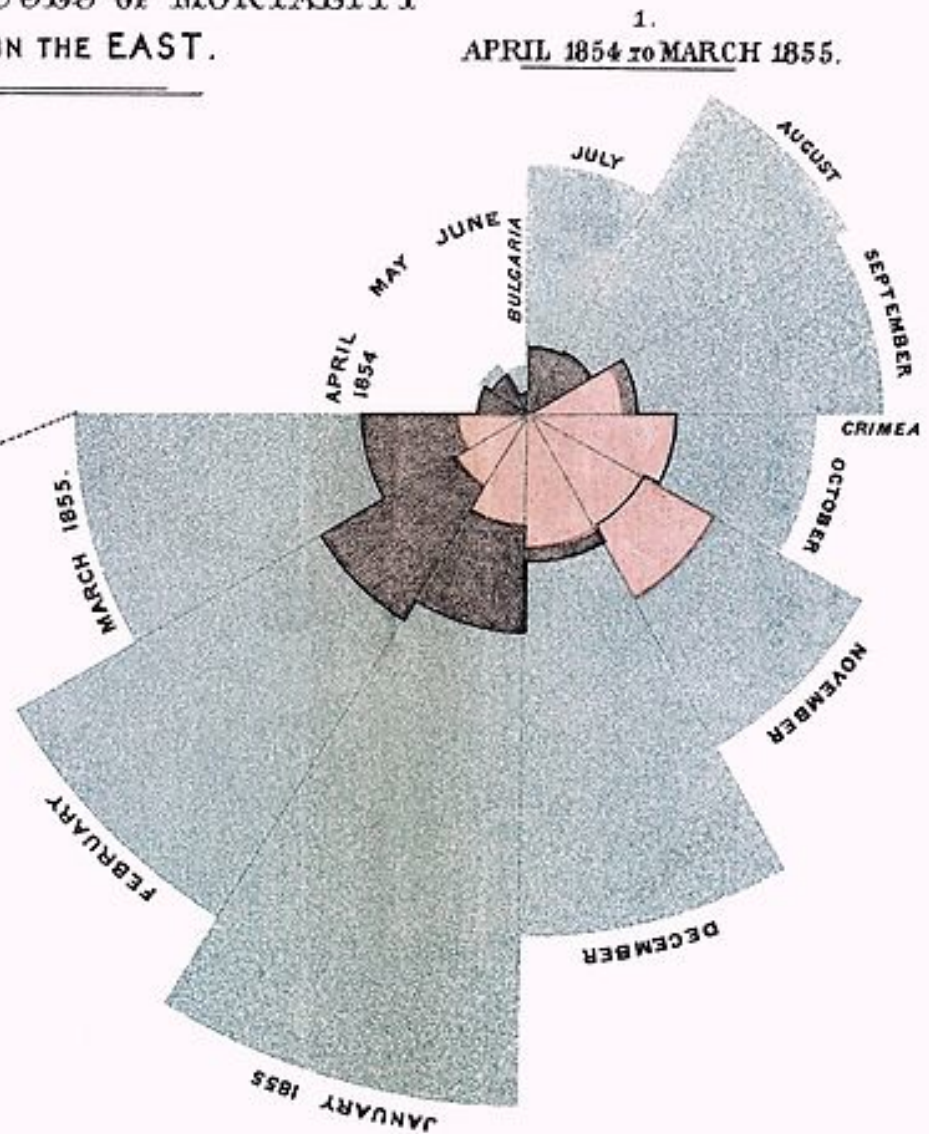
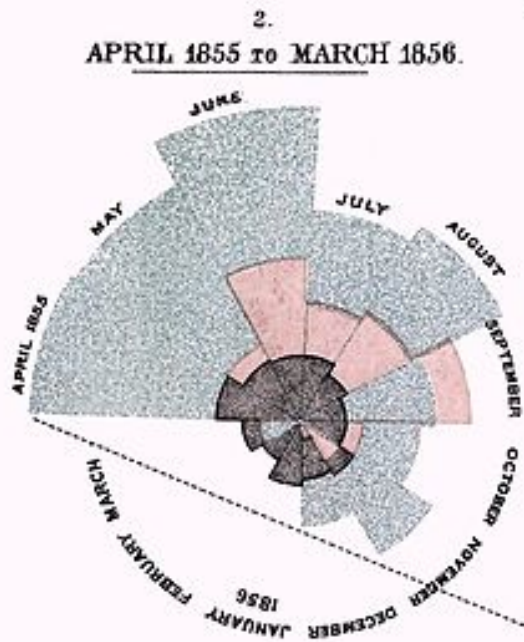
Florence Nightingale in the Crimean War 1854-1855



'Polar-area charts'

Segment area proportional to number

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

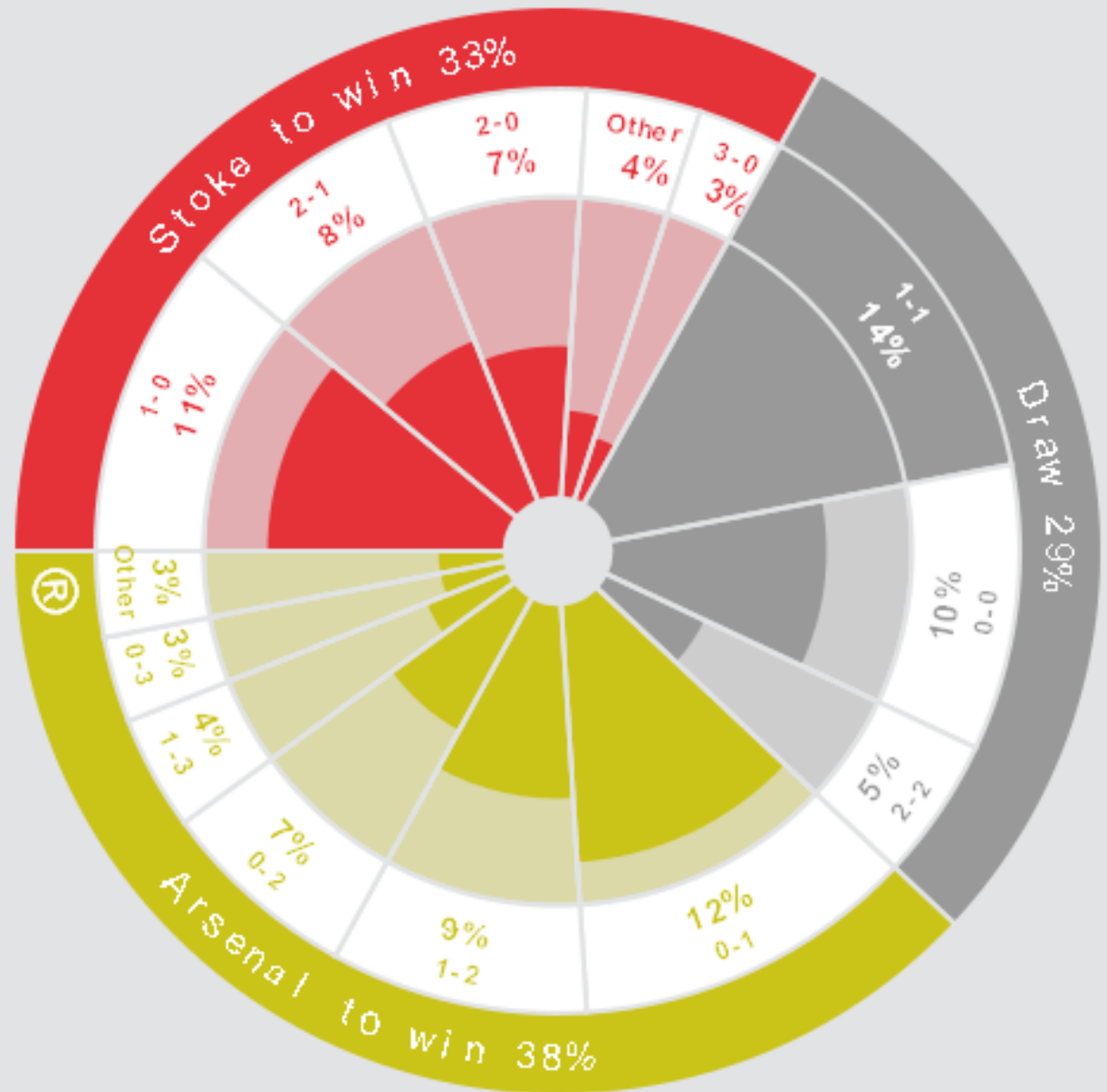
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during the month.

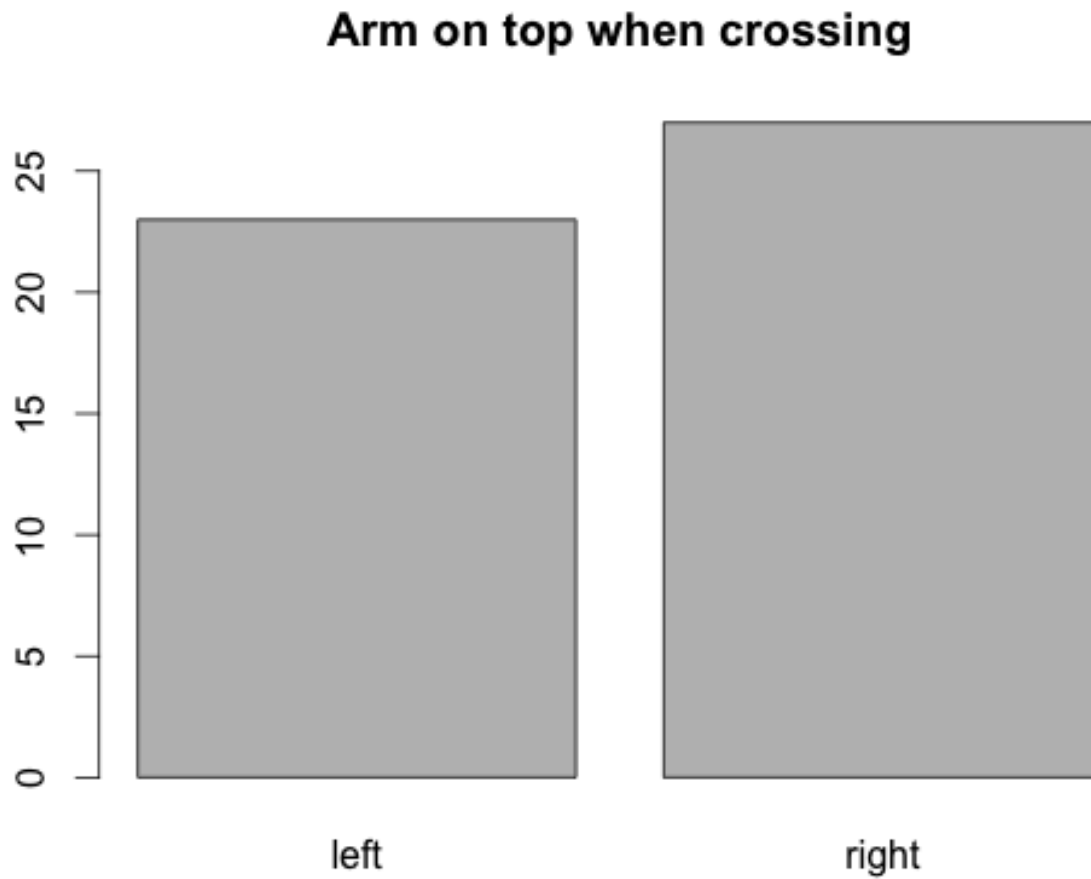
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

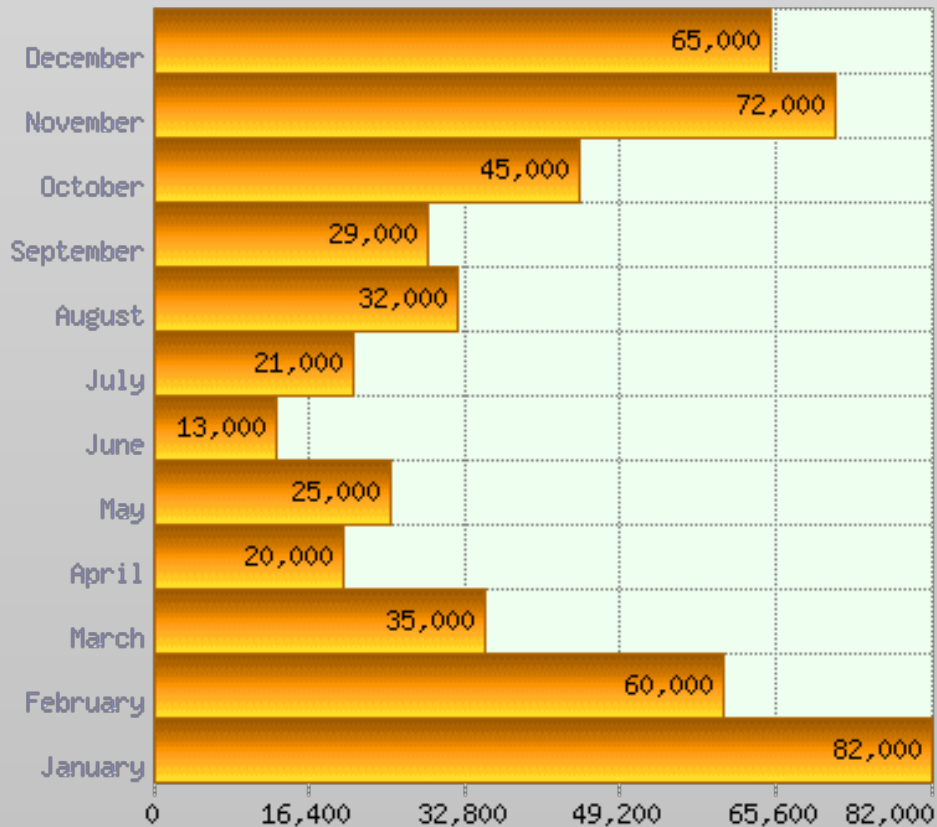
Kickoff.co.uk



Bar charts

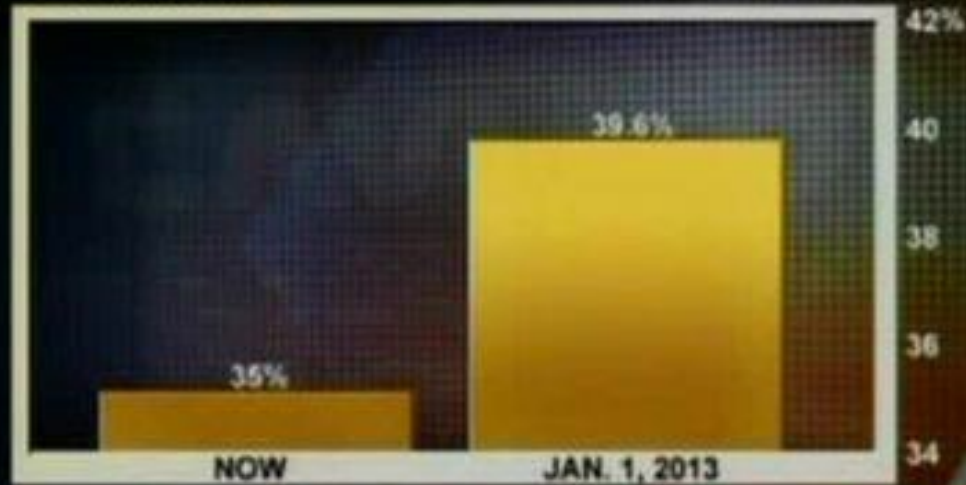


Shading is OK - 3d isn't!



IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



8:01 p ET

FOX
BUSINESS

TOP STORIES

TECHNOLOGY

CONSUMER

WITH THE JUSTICE DEPARTMENT AND ACQUIRES FULL T

DOW 13008.68 ▼ 64.33




S&P 1379.32 ▼ 5.98

NASDAQ 2939.52 ▼ 6.32

Stacked bar charts





Decision: No Additional Therapy



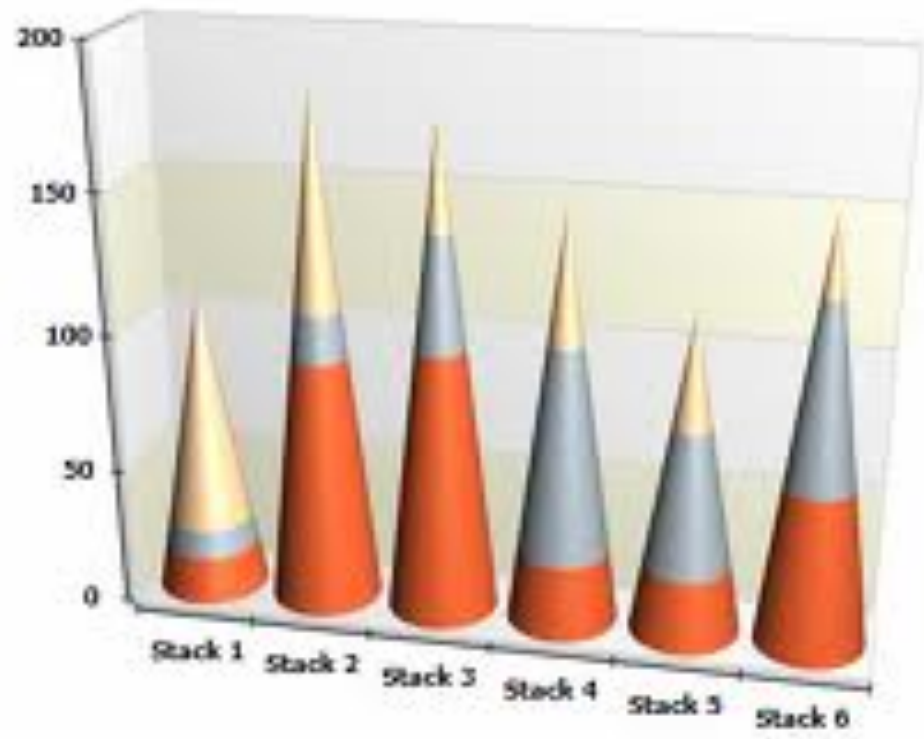
-  78 out of 100 people are alive in 5 years.
-  12 out of 100 people die because of cancer.
-  10 out of 100 people die of other causes.

Decision: Chemotherapy



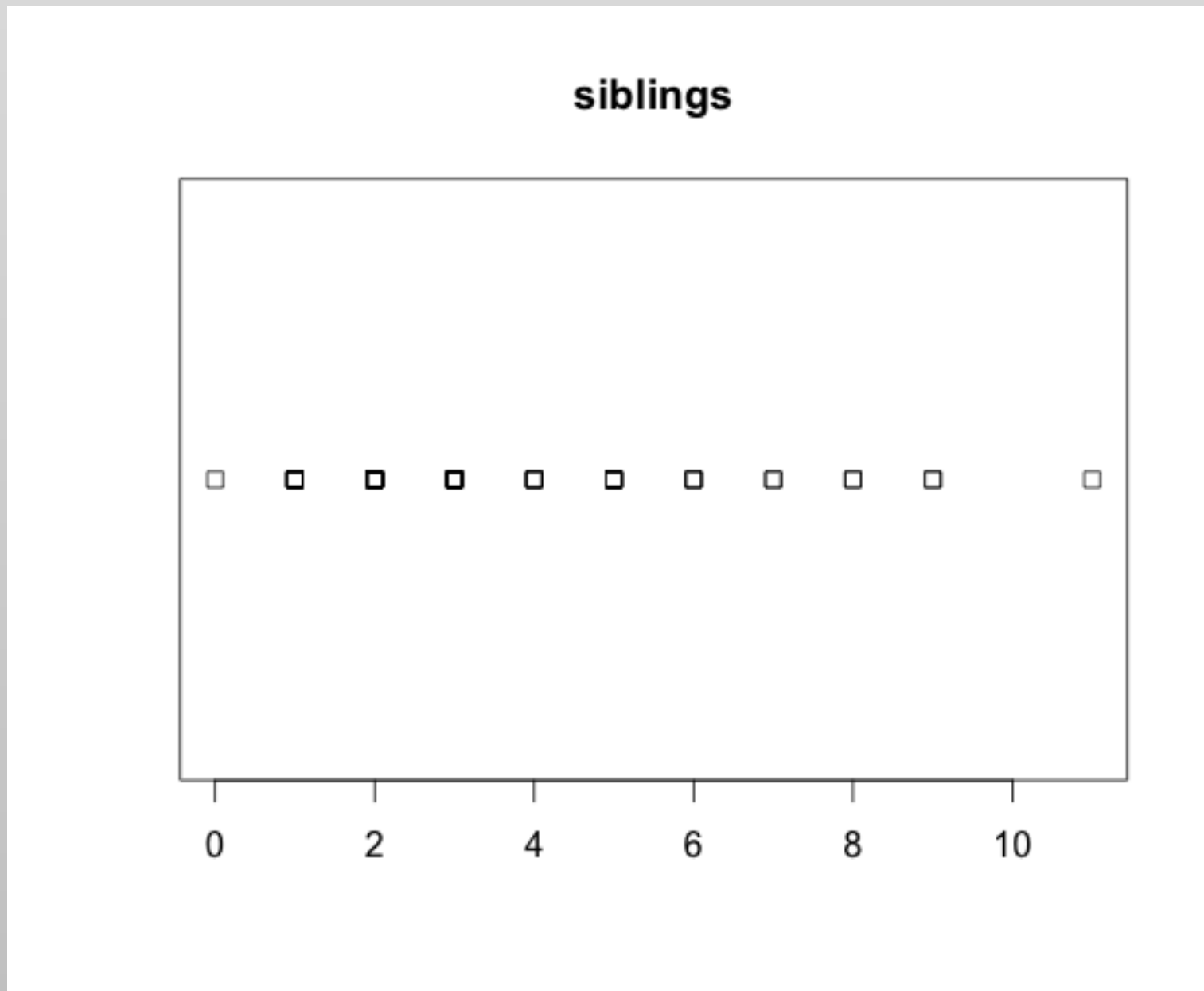
-  78 out of 100 people are alive in 5 years. Plus...
-  5 out of 100 people are alive because of therapy.
-  7 out of 100 people die because of cancer.
-  10 out of 100 people die of other causes.

Stacked Bar Chart

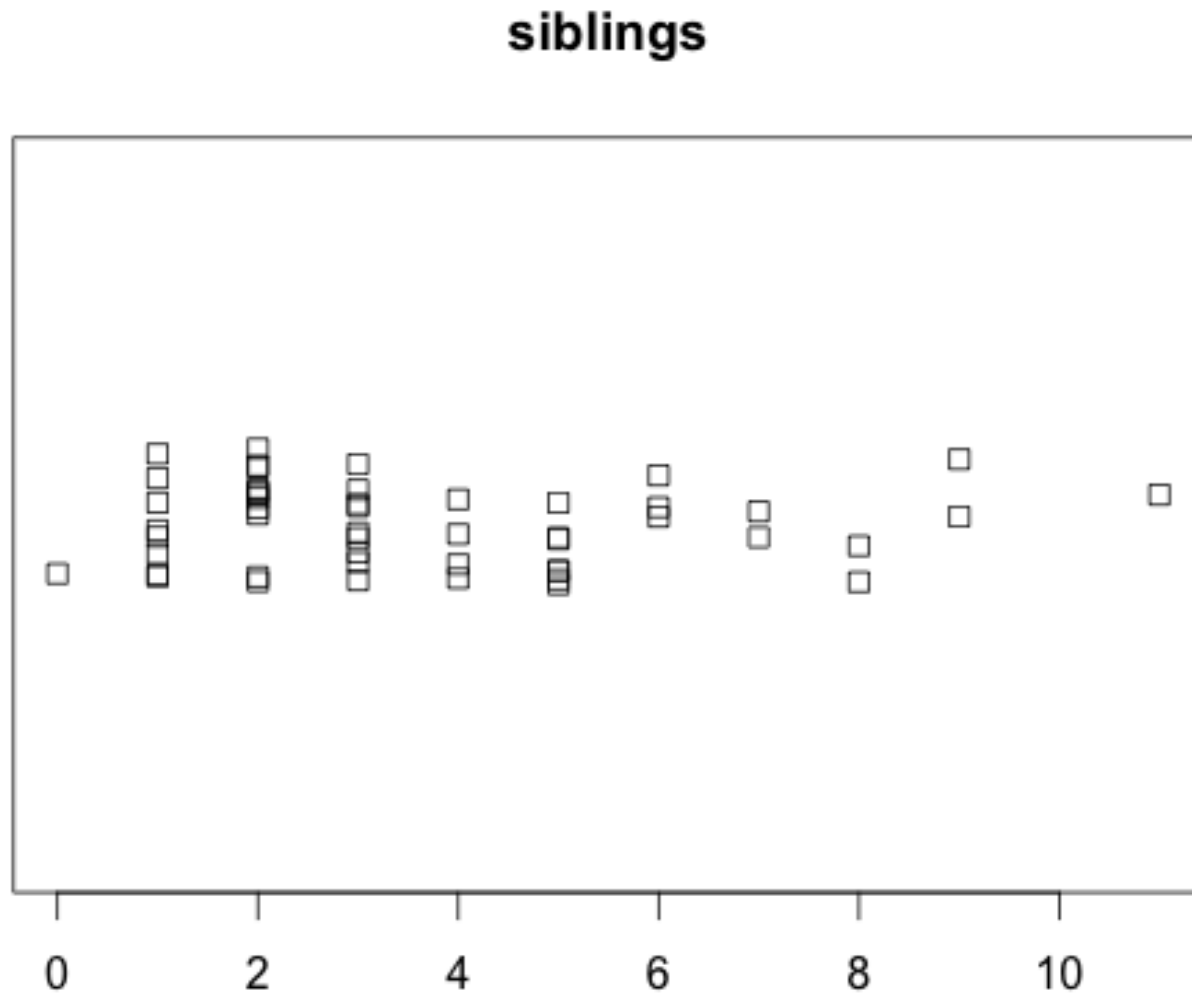


Dot plots

```
stripchart(siblings,main="siblings") # does not show numbers well
```

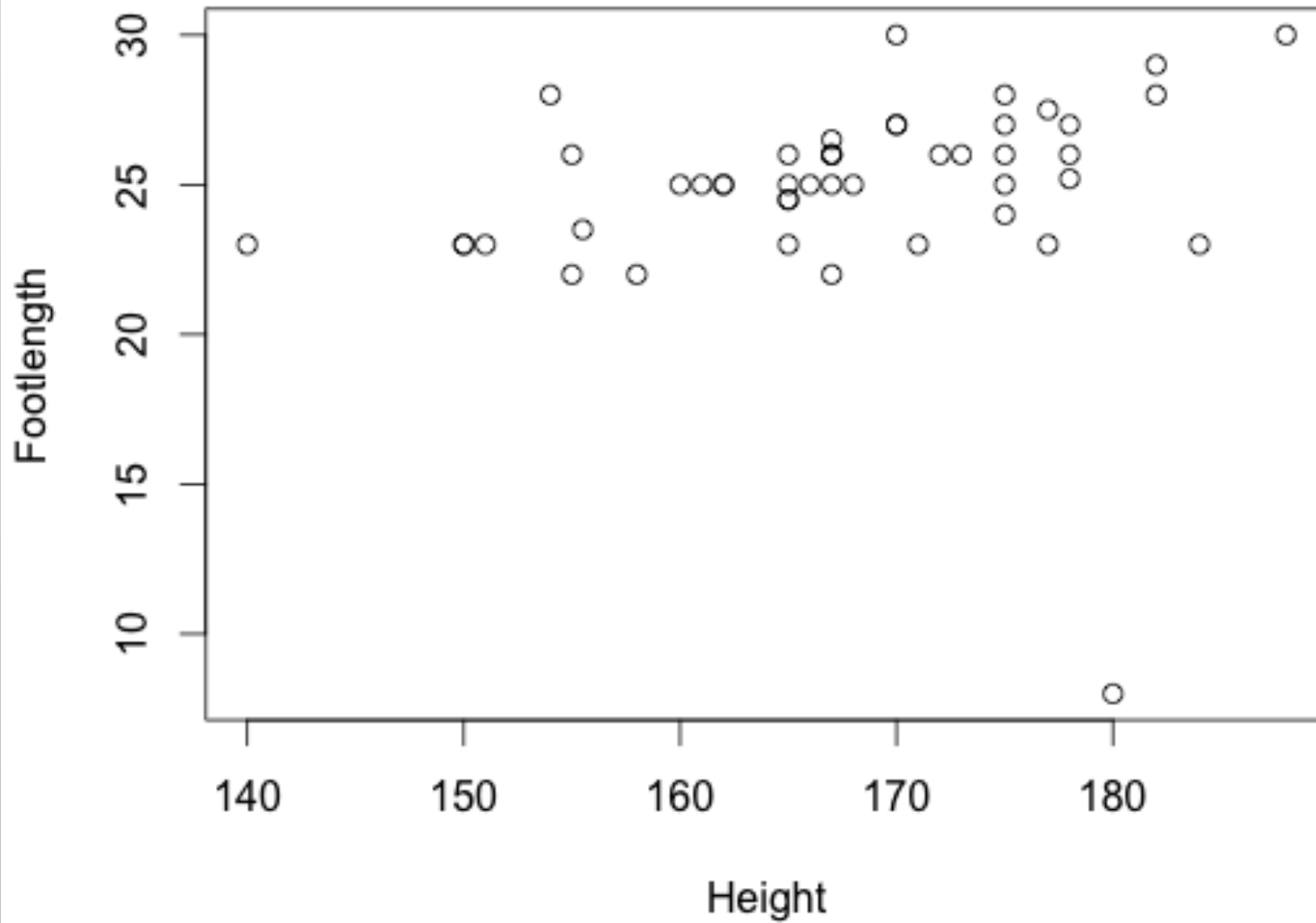


```
stripchart(siblings, method="jitter",main="siblings") # add a jitter
```

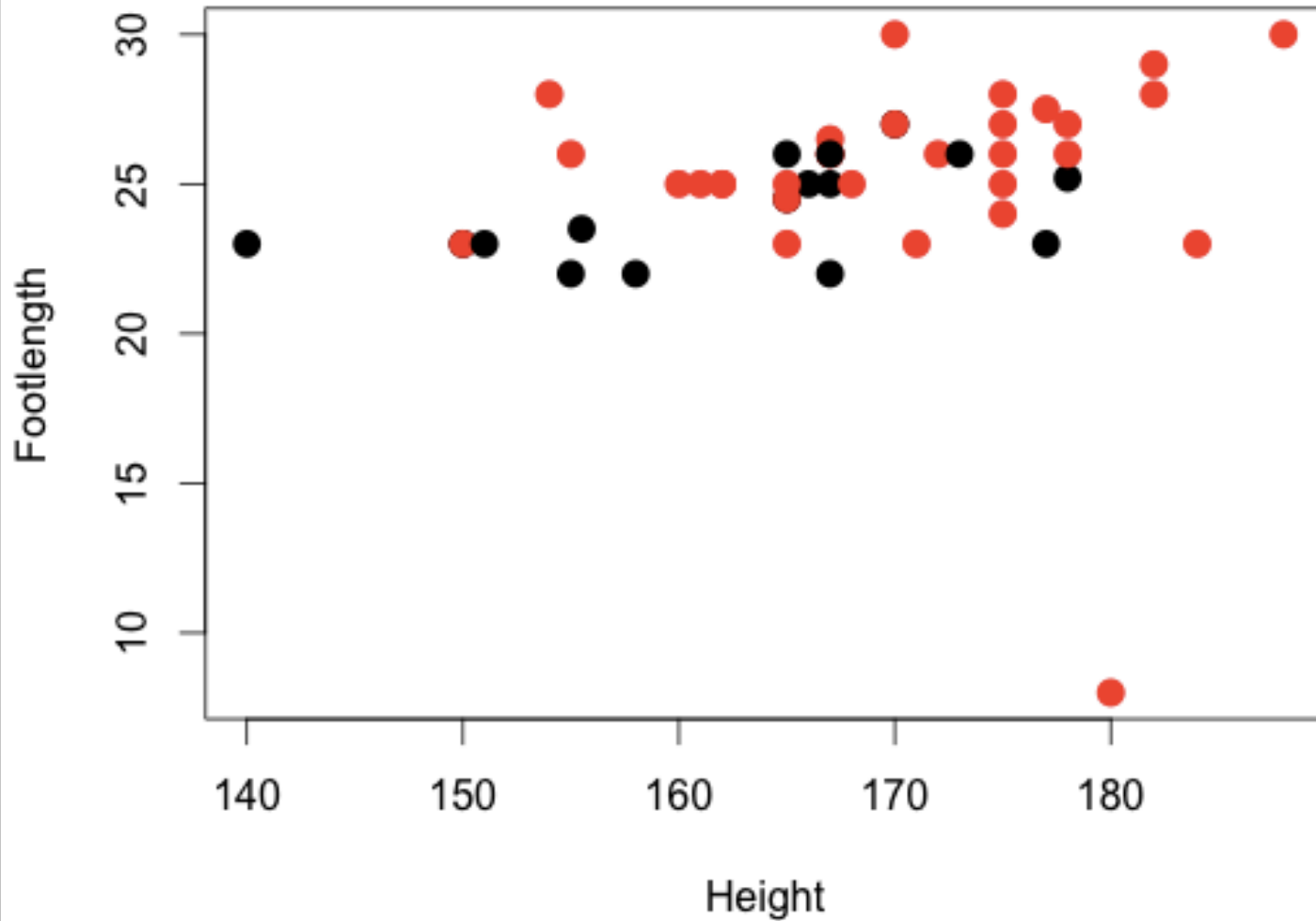


Scatterplots

```
plot(Height, Footlength)
```

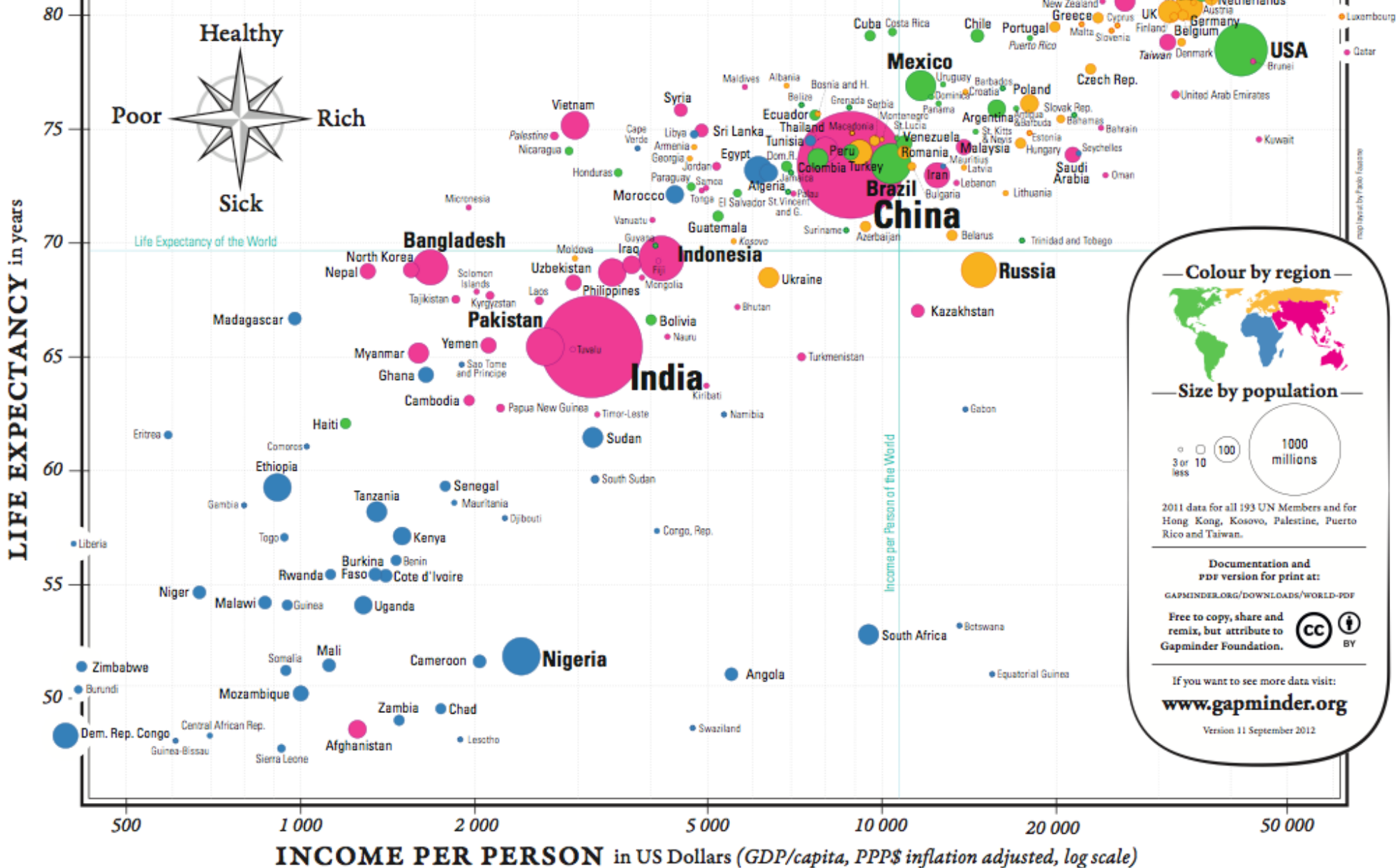


```
plotcolours=c(1,2)[Gender] # chooses 1 or 2 according to factor Gender  
plot(Height,Footlength,pch=16,col=plotcolours,cex=1.5) # big coloured blobs
```



GAPMINDER WORLD 2012

Mapping the Wealth and Health of Nations



Line graphs

UNEMPLOYMENT RATE UNDER PRESIDENT OBAMA



2011

SOURCE: BUREAU OF LABOR STATISTICS

AR WAR AND A TROOP WITHDRAWAL AT THE E/ NAS FUT 2,292.50

UNEMPLOYMENT RATE UNDER PRESIDENT OBAMA



2011

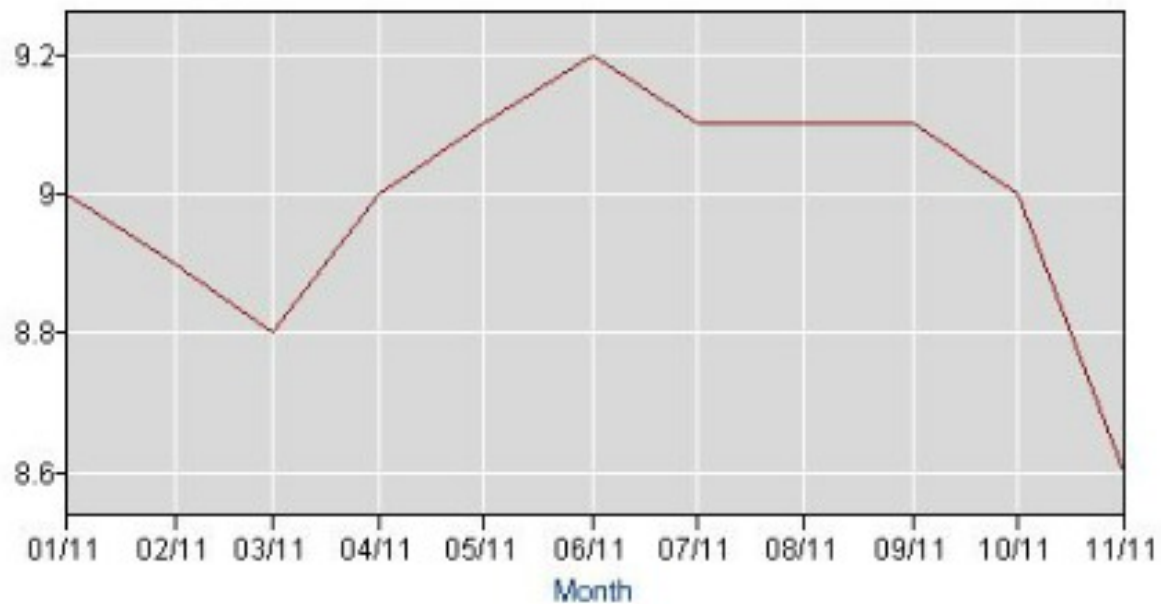
SOURCE: BUREAU OF LABOR STATISTICS

AR WAR AND A TROOP WITHDRAWAL AT THE E/ NAS FUT 2,292.50

Data extracted on: December 12, 2011 (9:50:59 AM)

Labor Force Statistics from the Current Population Survey

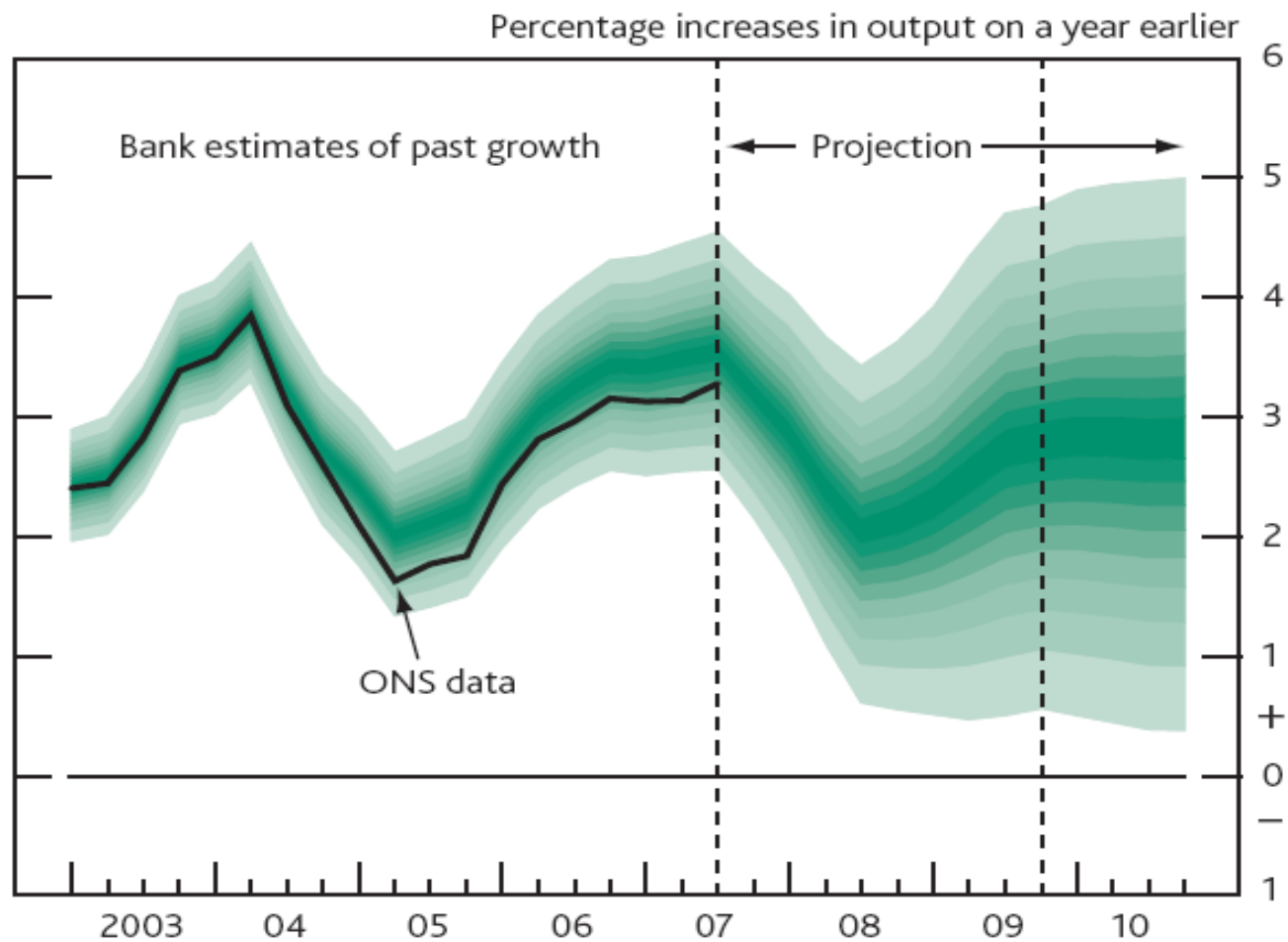
Series Id: LNS14000000
Seasonally Adjusted
Series title: (Seas) Unemployment Rate
Labor force status: Unemployment rate
Type of data: Percent or rate
Age: 16 years and over



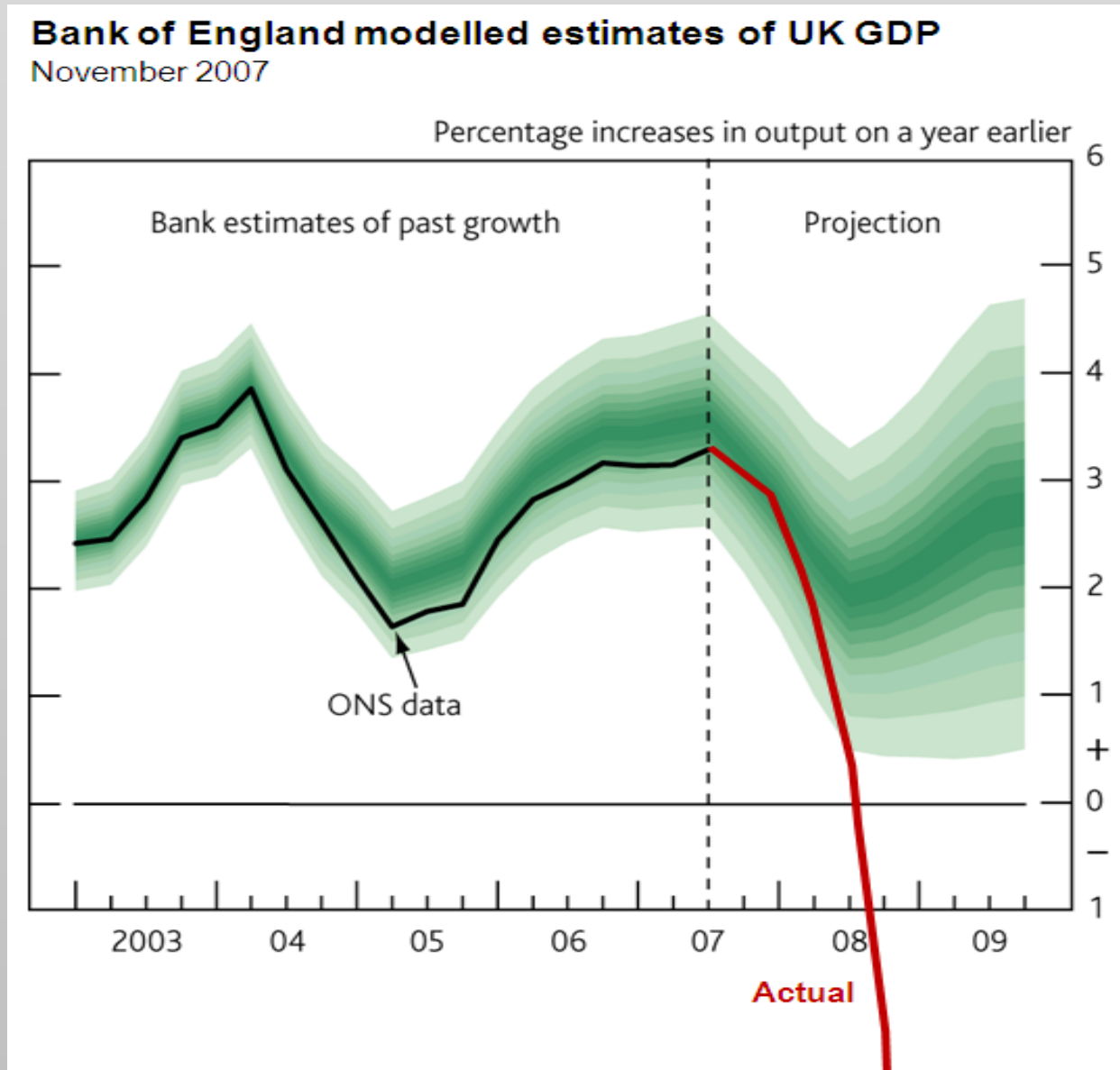
Bank of England Fan Charts

- *If economic circumstances identical to today's were to prevail on 100 occasions ...*
- *Consequently, GDP growth is expected to lie somewhere within the entire fan on 90 out of 100 occasions*

Chart 5.1 GDP projection based on market interest rate expectations



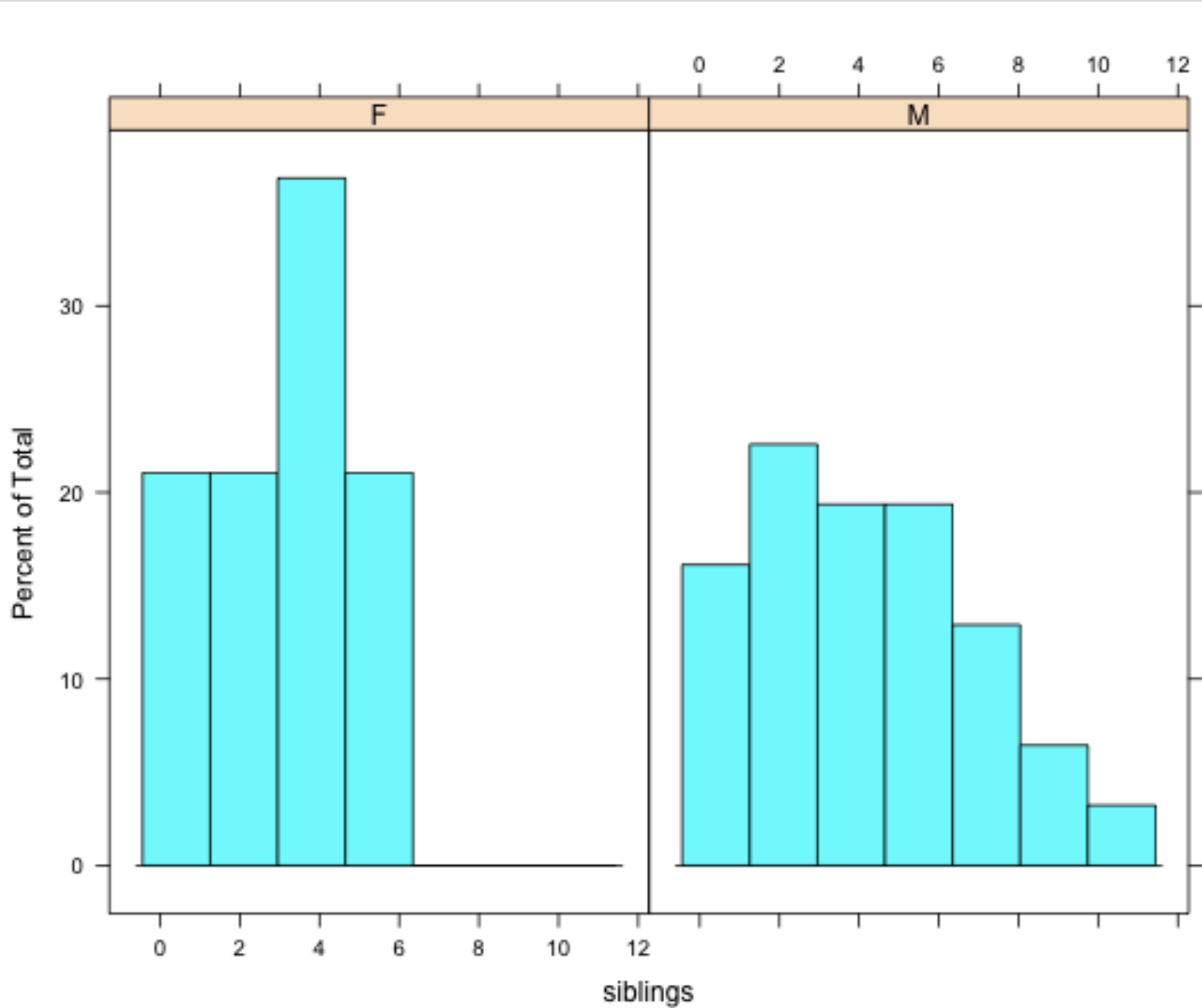
But must be ready for surprises



'Small multiples'

The human eye is good at spotting patterns

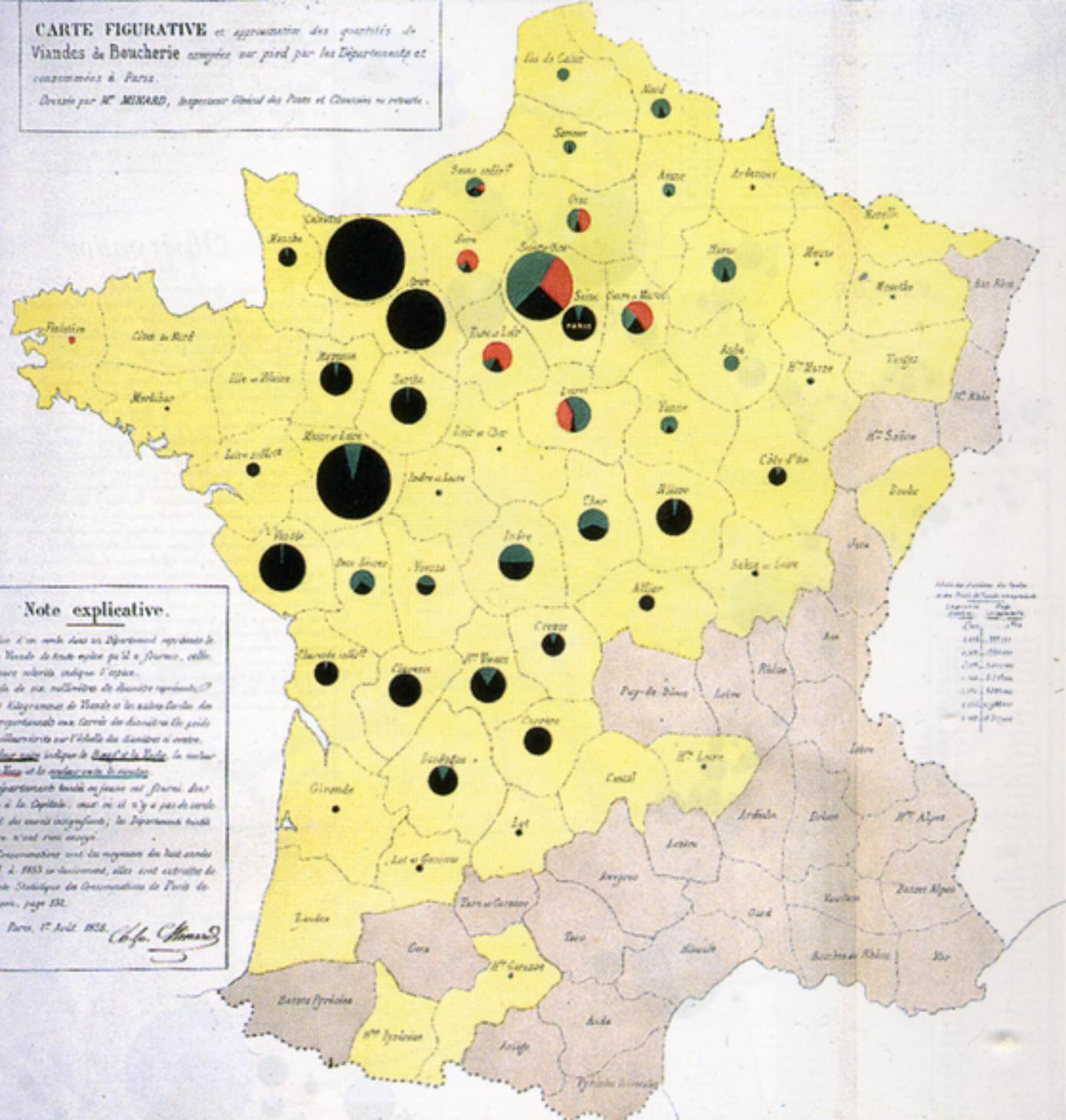
histogram(~siblings|gender) # Need to include ~, to make a 'formula'



Spatial

CARTE FIGURATIVE et approximation des quantités de
Viandes de Boucherie consommées sur pied par les Départements et
consommées à Paris.

Dressée par M. MINARD, Inspecteur Général des Poids et Mesures de Paris.



Note explicative.

La surface d'un cercle dans un Département représente le poids de Viande de toute espèce qu'il a fournie, celle de Secteurs colorés indique l'espèce.

Tous cercles de six millimètres de diamètre représentent 500 000 kilogrammes de Viande et les autres cercles des poids proportionnels aux cercles de diamètre ou les poids sont indiqués sur l'échelle des diamètres à gauche.

Les couleurs indiquent le Boeuf et le Porc, la couleur rouge le Bœuf et le Porc consommés à Paris.

Les Départements hachés en jaune ont fourni des Secteurs à la Capitale, mais si il n'y a pas de cercles est fait de cercles insignifiants, les Départements hachés en brun n'ont rien envoyé.

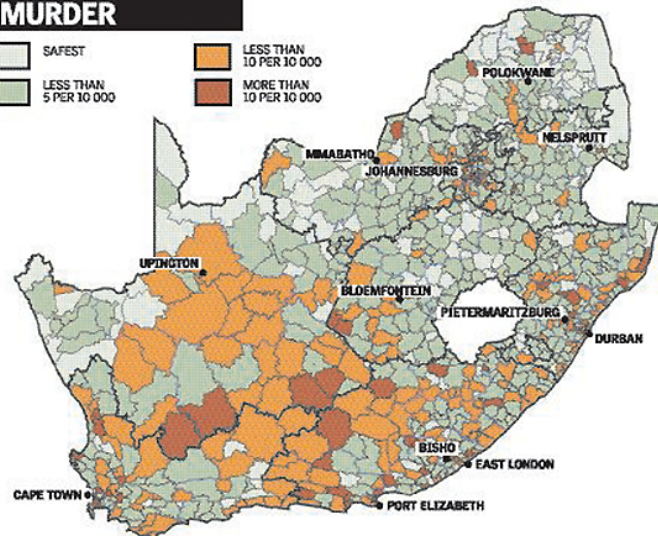
Ces Consommations sont les moyennes des huit années de 1845 à 1853 inclusivement, elles ont été tirées de l'Annuaire Statistique des Consommations de Paris de M. Buisson, page 131.

Paris, le 17 Avril 1853. M. Minard

Echelle des diamètres des cercles
et des Secteurs colorés

Diamètre	Poids
6 mm	500 000 kg
12 mm	1 000 000 kg
18 mm	1 500 000 kg
24 mm	2 000 000 kg
30 mm	2 500 000 kg
36 mm	3 000 000 kg
42 mm	3 500 000 kg
48 mm	4 000 000 kg
54 mm	4 500 000 kg
60 mm	5 000 000 kg

MURDER



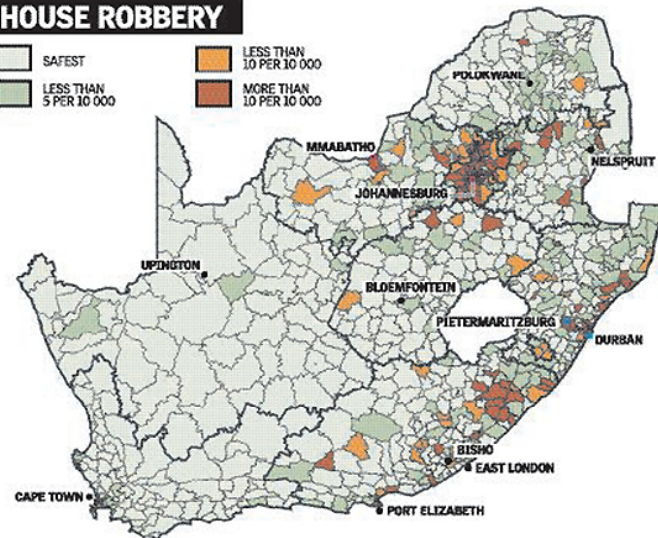
HOTSPOTS

- The Northern Cape and the Western Cape are the worst, followed by the Eastern Cape. Frasersburg, followed by places such as Kalamas, Kalmoes and Grikwastad in the Northern Cape are high risk areas.
- In the Western Cape, Vredenburg near Paternoster, Graaffwater near Clanwilliam and Klawer in the interior are hotspots. Closer to Cape Town, the danger zones are Nyanga, Crossroads, Langa, Khayelitsha, Klapmuts, Franschoek. In the Eastern Cape: Graaff Reinet, Loubertina, Patensie, Kippiaat and Jeffrey's Bay are high risk zones.

SAFER ZONES:

- LIMPOPO, except for Mara in the north, and MPUMALANGA, except for Kinross on the eastern highveld, are among the safer zones for murder.
- The FREE STATE has a low risk, except Wanda and Luckhoff in the southern Free State.

HOUSE ROBBERY



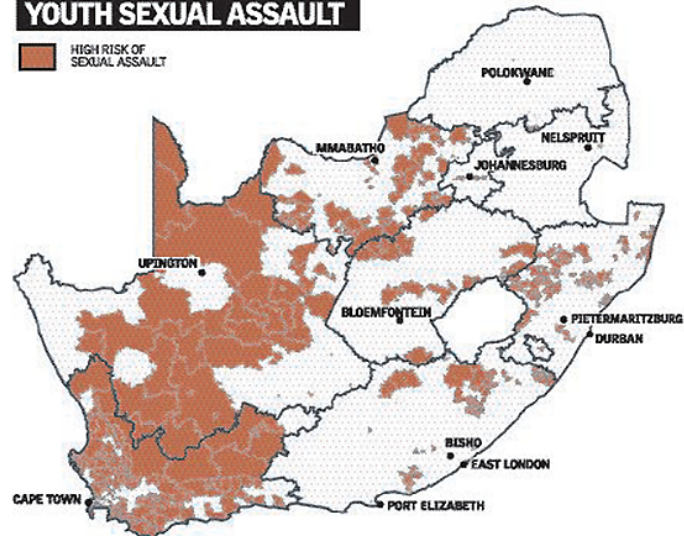
HOTSPOTS

- The entire Gauteng is at high risk of house robberies, except for a few tiny spots on the map. Areas close to the coast of Kwa-Zulu Natal and the Eastern Cape are also at high risk.
- In the Eastern Cape, the high risk areas include the suburbs around Pietermaritzburg, suburbs northwest of Durban city and Richmond and Thornville in the Midlands.

SAFER ZONES

- The Northern Cape, Western Cape, Limpopo and the Free State. The whole of the Western Cape is at low risk, except Hout Bay, the southern suburbs such as Rondebosch and Nyanga and Crossroads.

YOUTH SEXUAL ASSAULT



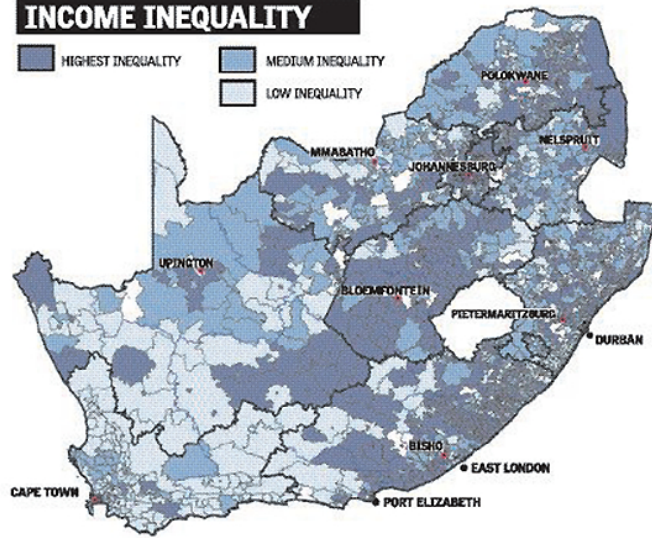
HOTSPOTS

- Children are at great risk in most areas of the Western and Northern Cape.

SAFER ZONES:

- There are few high risk areas in Mpumalanga or Limpopo. Gauteng, the Free State, the Eastern Cape and the KZN coastal region are also not at high risk.

INCOME INEQUALITY



INEQUALITY:

HIGH INEQUALITY:

- Gauteng, the Free State.

HIGH EQUALITY:

- Large parts of the Western and Northern Cape.

Animations

Florence Nightingale
Google Public Data Explorer
Hans Rosling

Google Public Data Explorer

<http://www.google.com/publicdata/directory>

Infographics

'Journalistic'

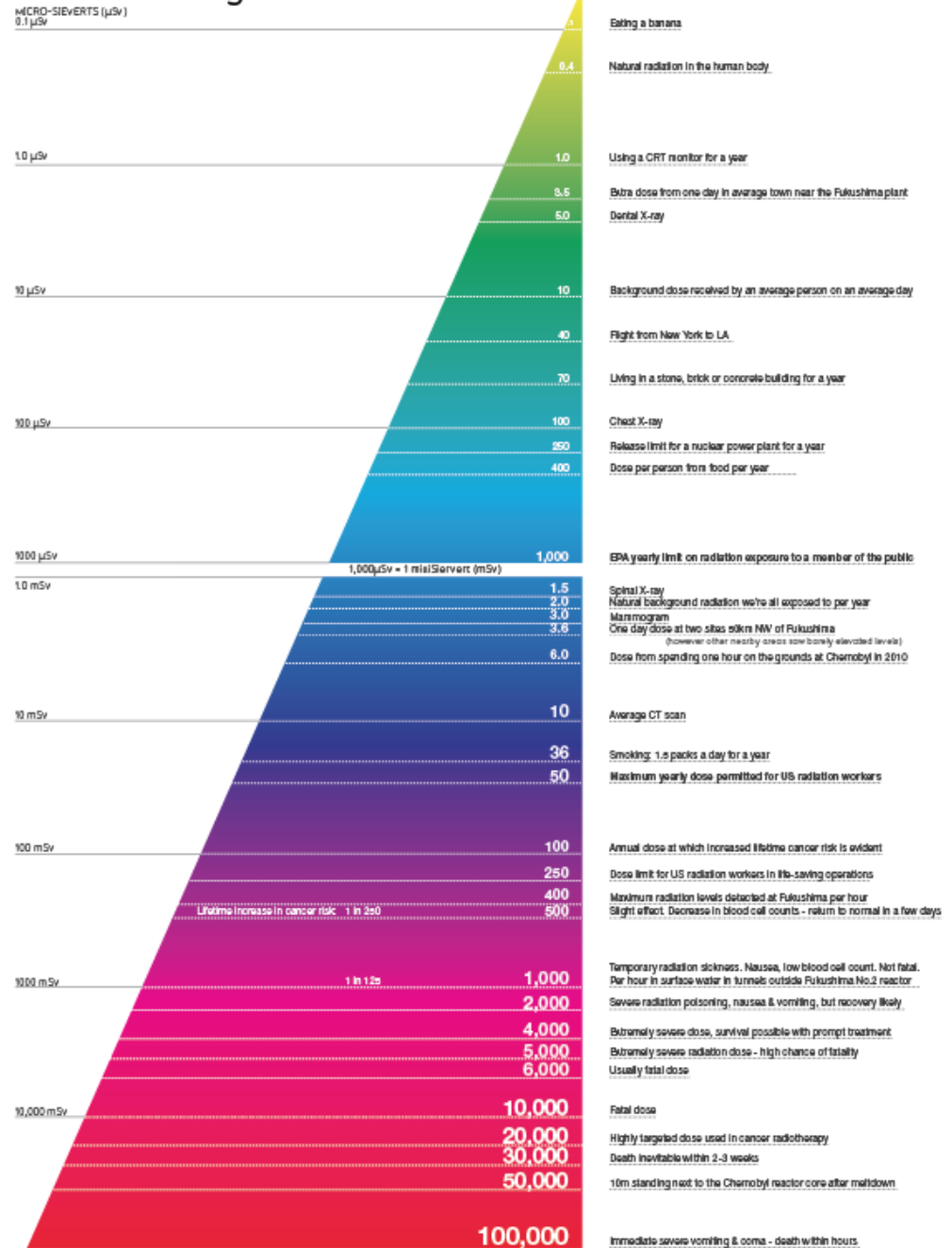
Use images, colour, layout to increase impact

Almost works of art

Increasingly popular

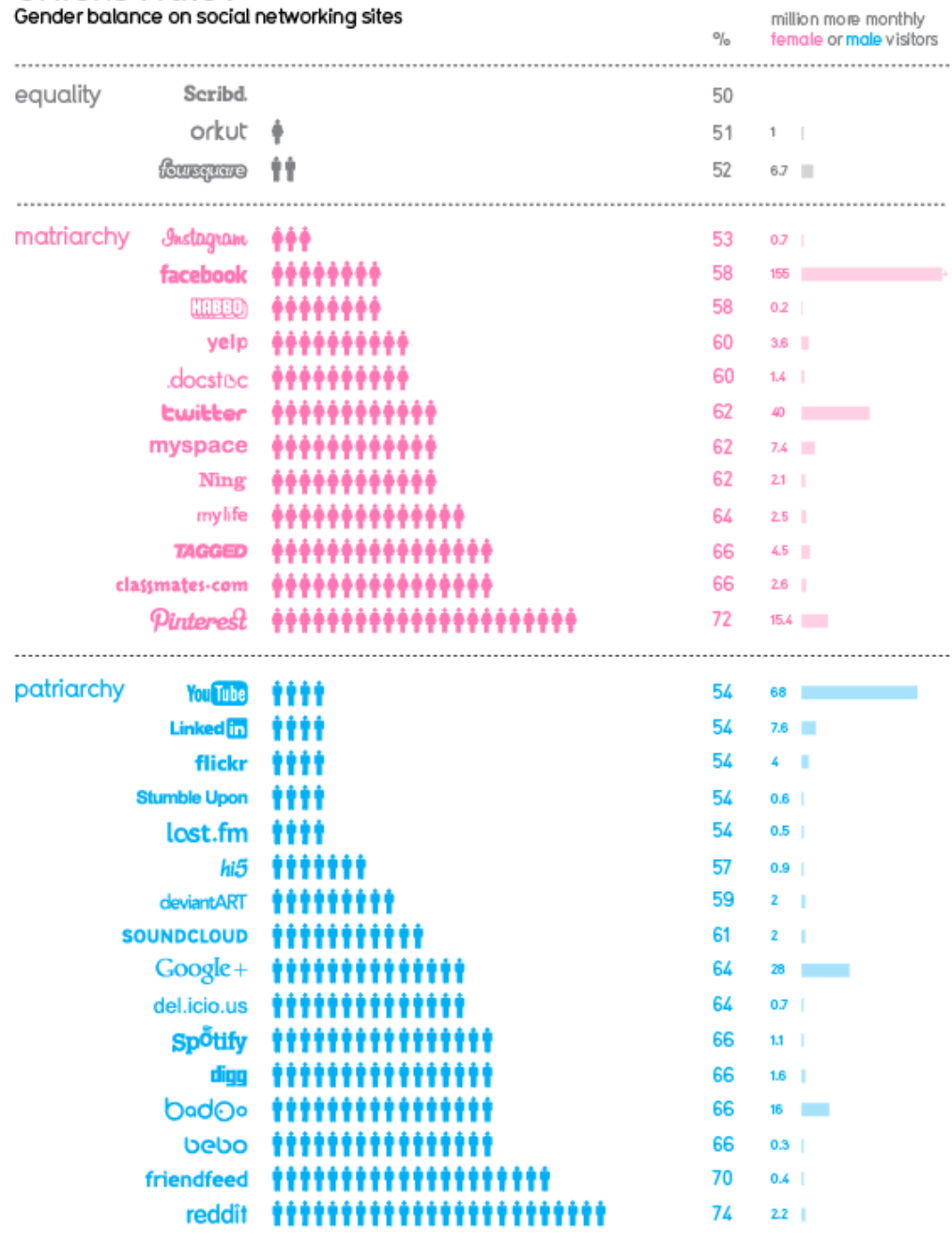
David McCandless *Information is Beautiful*

Radiation Dosage Chart

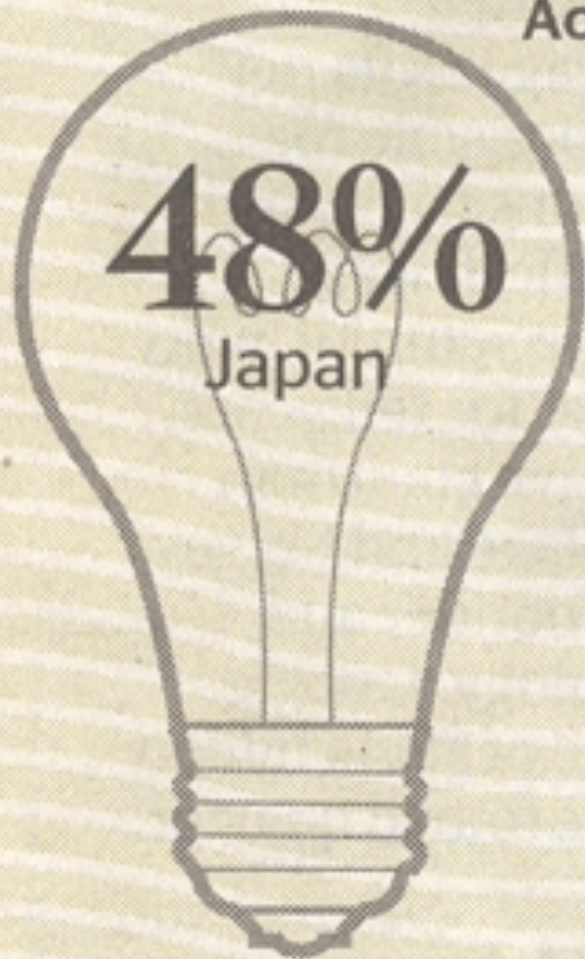


Chicks Rule?

Gender balance on social networking sites



Acceptance rate of patent applications



The best statistical graphic ever?

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. MINARD, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Léger, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent vers Orscha et Witebsk, avaient toujours marché avec l'armée.

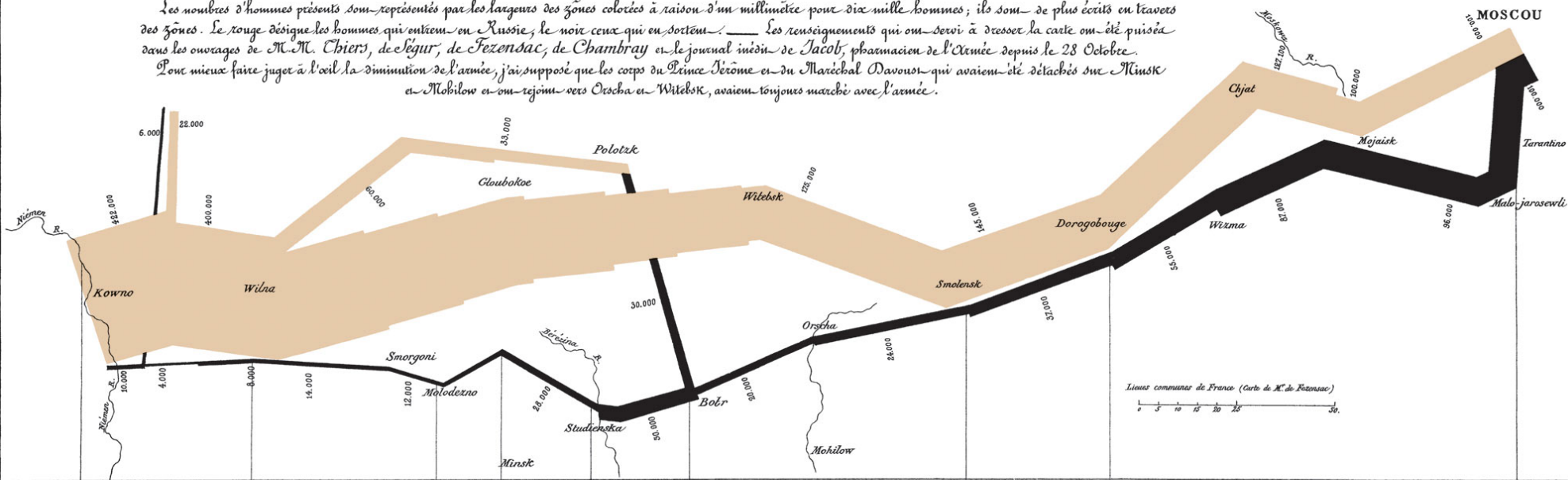
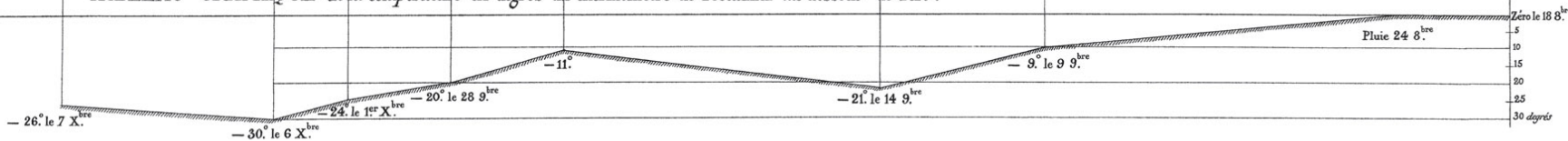


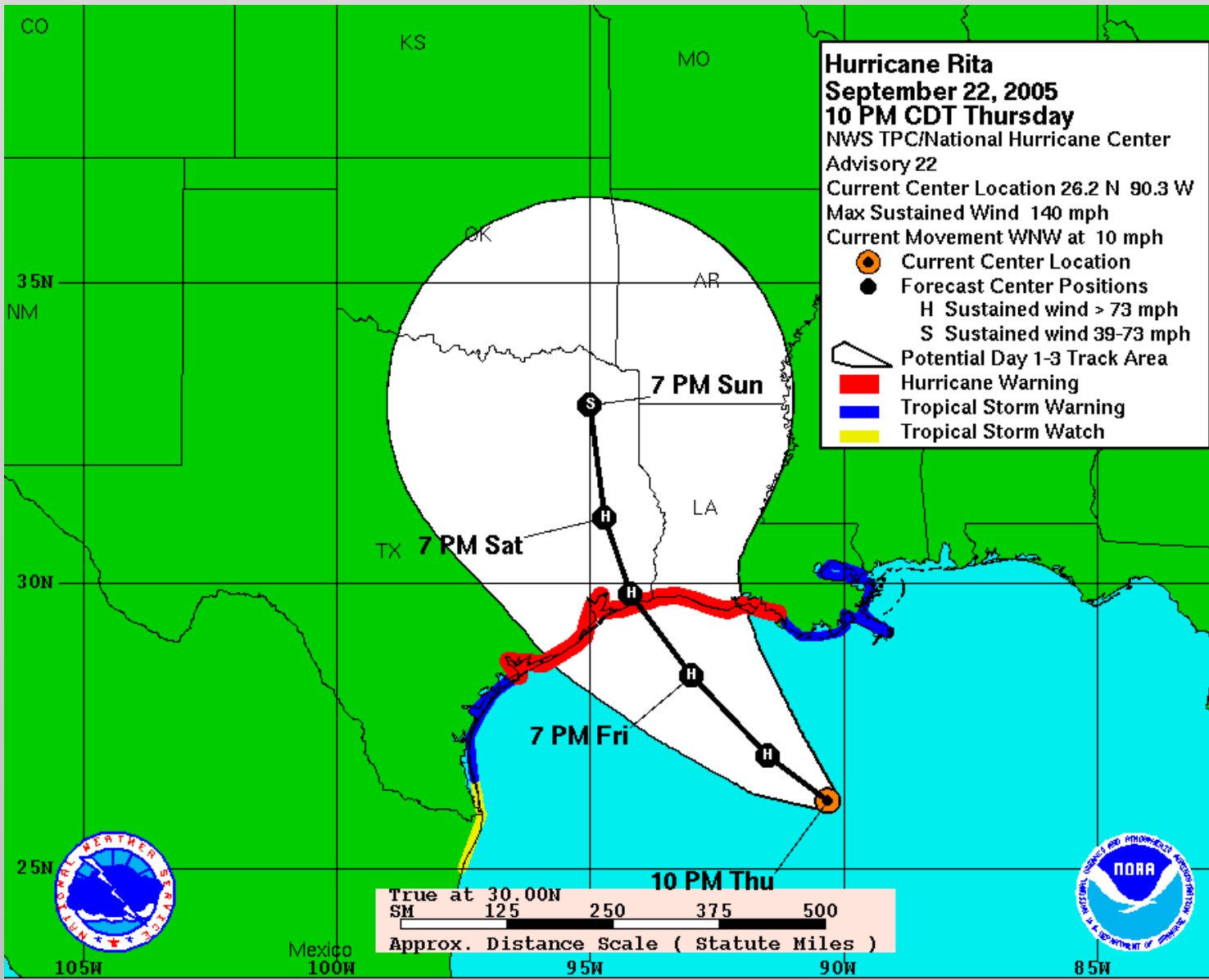
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

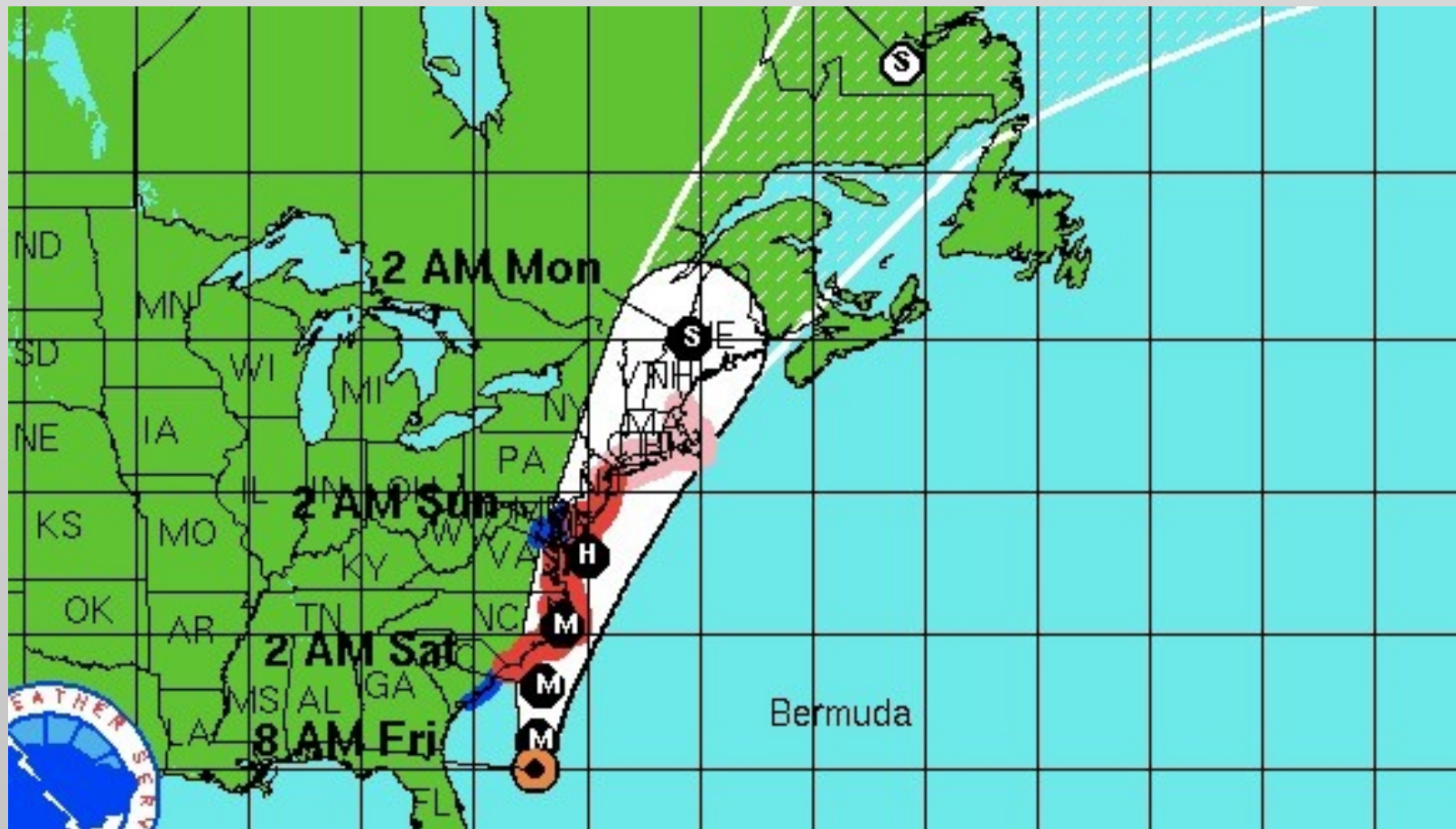


Les Cosaques passent au galop le Niémen gelé.

Spatial – weather uncertainty

"Cone of Uncertainty" for hurricane warnings





2011: NBC News for Hurricane Irene



'Possible futures'
metaphor

18th January 2007


 LIVE

BBC NEWS CHANNEL



Last Updated: Thursday, 18 January 2007, 20:46 GMT

 [E-mail this to a friend](#)

 [Printable version](#)

Nine dead as storms hit England

Nine people, including a two-year-old boy, have died as gales and heavy downpours hit England.

The boy died when a wall fell on him in London and a woman died in Stockport in a similar incident. A man in Cheshire died when he was hit by a tree.

Another man died when he was



A bus left the road during high winds on the A10 in Cambridgeshire

How well was this predicted 5 days beforehand?

24 'possible futures' from Met Office

MOGREPS (THORPEX)

PMSL (hPa)

T+120h

DT 12Z on 13/01/2007
VT 12Z on 18/01/2007

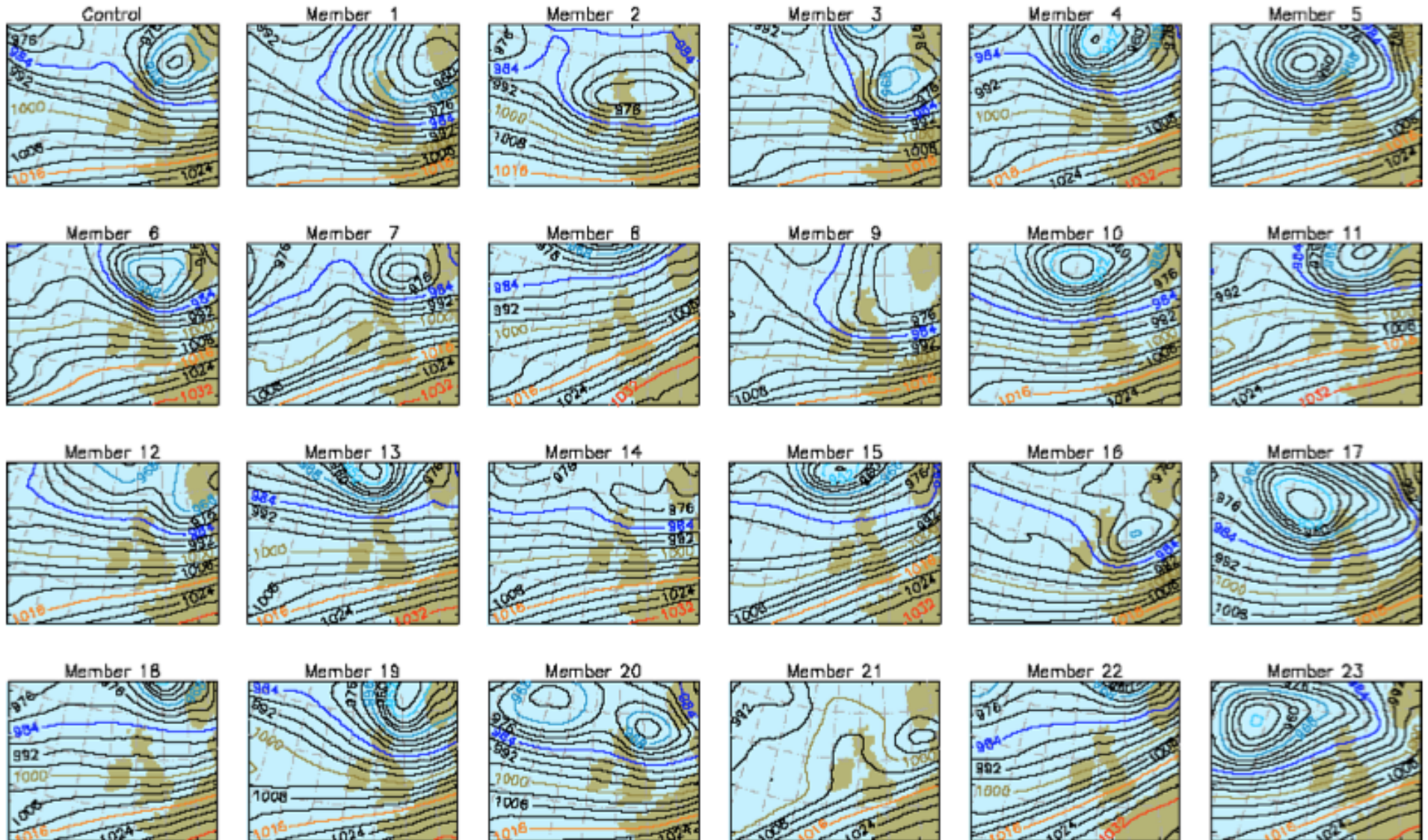
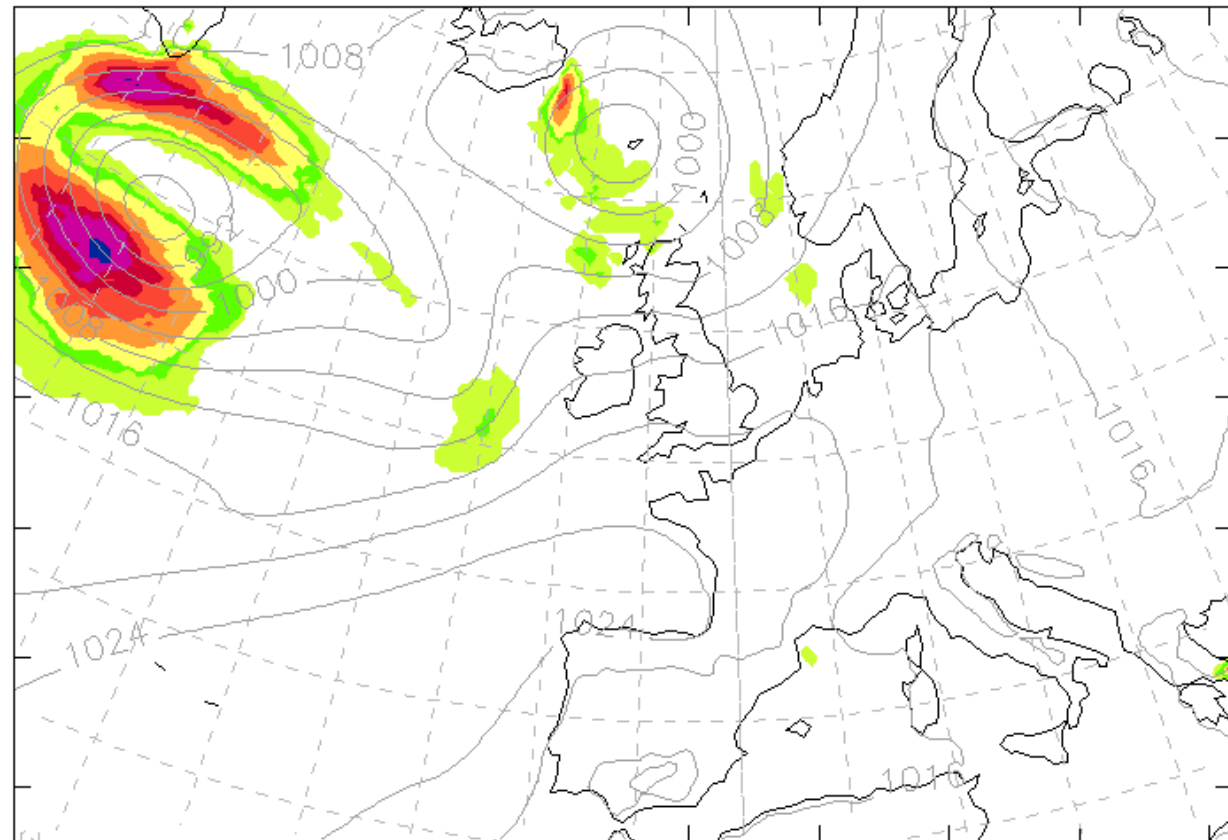


Figure 4.1.7: T+120 MOGREPS-15 postage stamps of PMSL for VT 12Z 18/01/2007

Met Office: Prob of Gusts > 40kt

MOGREPS (Regional) Probability map for GustSpeed > 40.0knots
DT 06Z on Thu 15/07/2010 VT 12Z on Sat 17/07/2010 lead time 54h
(Ensemble Mean PMSL plotted as faint background)



William Hill 1st June 2011

Nadal

Gulbis Hewitt
Berdych

Roddick

Karlovic

Davydenko

Isner

Montis

Soderling

Djokovic

Melzer

del Potro

Tsonga

Bakker

Kohlhreiber

Raonic

Querrey

Murray

Istomin

Lopez

Federer

Yuzhny

Cilic

Nadal

Federer

Murray

Djokovic

del Potro

Roddick

Raonic

Tsonga

Cilic

Querrey

Berdych

Karlovic

Hewitt

Monfils

Melzer

Davydenko

Bakker

Istomin

Soderling

Isner

Gulbis

Lopez

Kohlschreiber

Youzhny

Some sources

- Books by Edward Tufte, eg *The Visual Display of Quantitative Information* QA276.3.T83 in library

Websites

IBM's Many-Eyes lets you upload data and create your own visualisations

<http://www-958.ibm.com/software/data/cognos/manyeyes/visualizations>

Google provides html templates for many graphics

https://developers.google.com/chart/interactive/docs/quick_start

Some R sources

Nice tutorial on R Graphics

<http://www.statmethods.net/index.html>

Good tutorial by author of lattice

http://www.isid.ac.in/~deepayan/R-tutorials/labs/04_lattice_lab.pdf

Graphics tutorial

<http://gbi.agrsci.dk/statistics/courses/phd07/material/Day1-R-Intro/basic-graphic.pdf>

<http://csg.sph.umich.edu/docs/R/graphics-1.pdf>

<http://www.math.csi.cuny.edu/Statistics/R/simpleR/>

R Graphics for enthusiasts

<http://gallery.r-enthusiasts.com/>