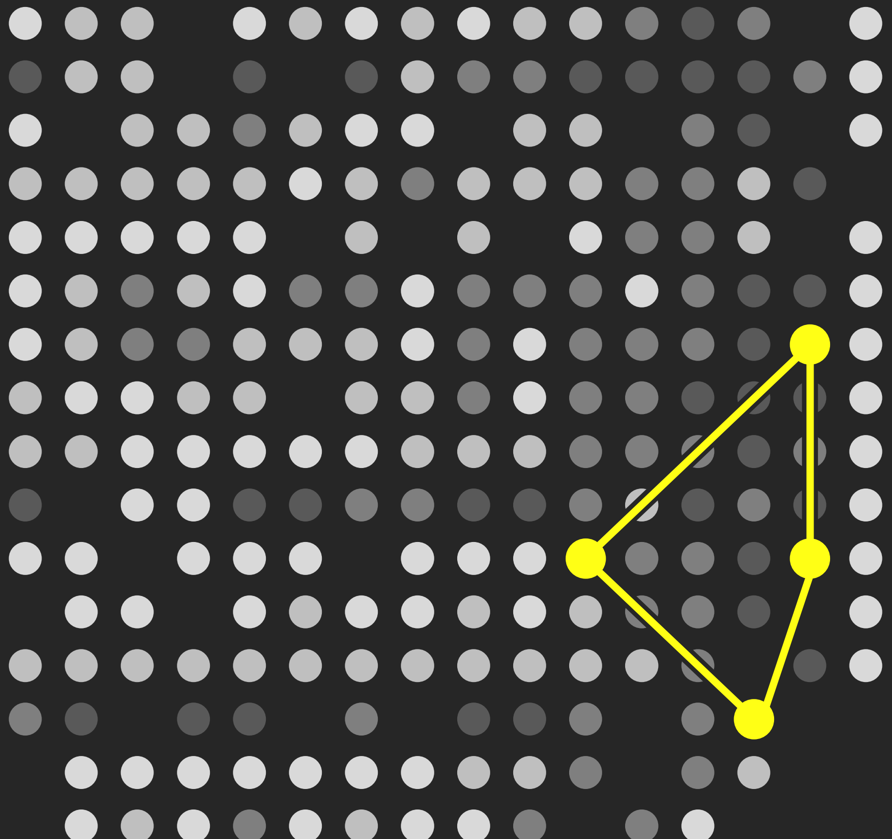
CENTRE FOR INTERDISCIPLINARY METHODOLOGIES

@gregmci Greg McInerny



ME

Ecology
BSc
UEA

Ecology &
Evolution
PhD, Leeds

Postdoc
Microsoft Research
Cambridge

Res. Fellow
Oxford
2020science

CIM
Warwick

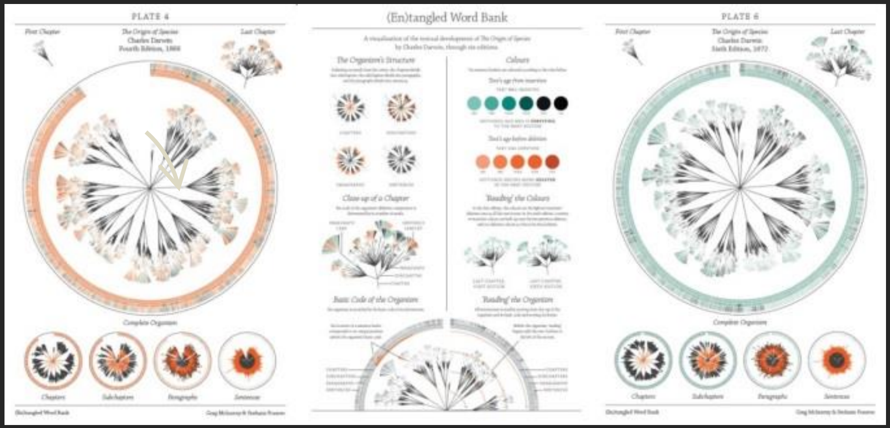
SCIENCE

SOFTWARE

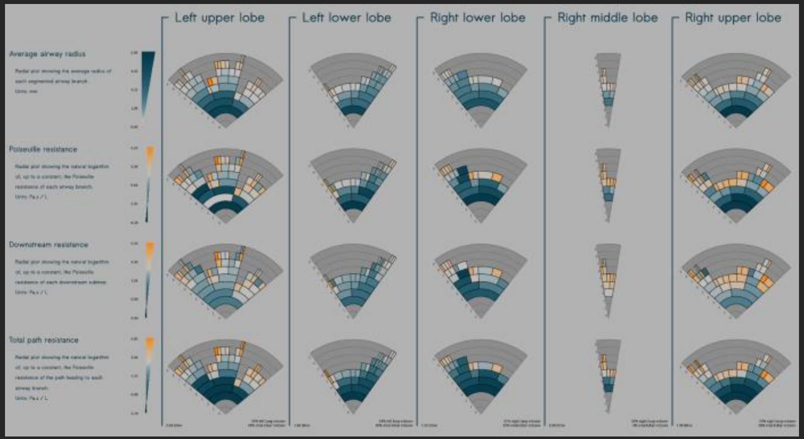
GRAPHS

VISUALISATION

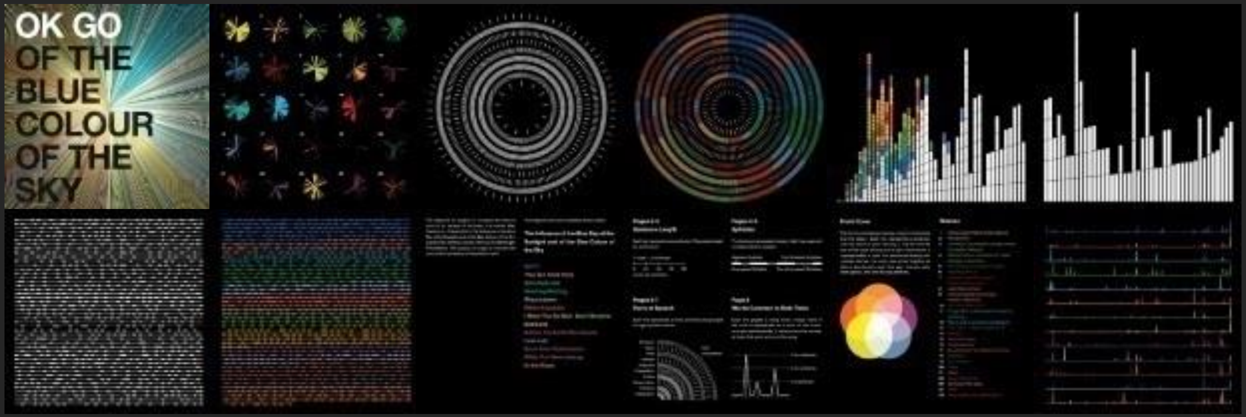
ART- DESIGN



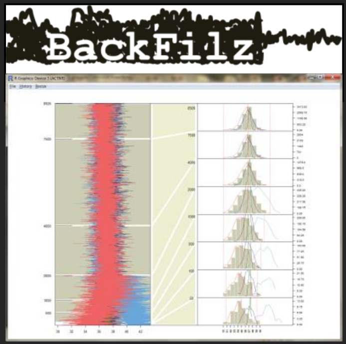
SCIENCE



ALBUM COVER



STATISTICAL GRAPHICS



What does it mean to ‘learn’ about…

a) visualisation?
b) visualisation in R?

Lecture >

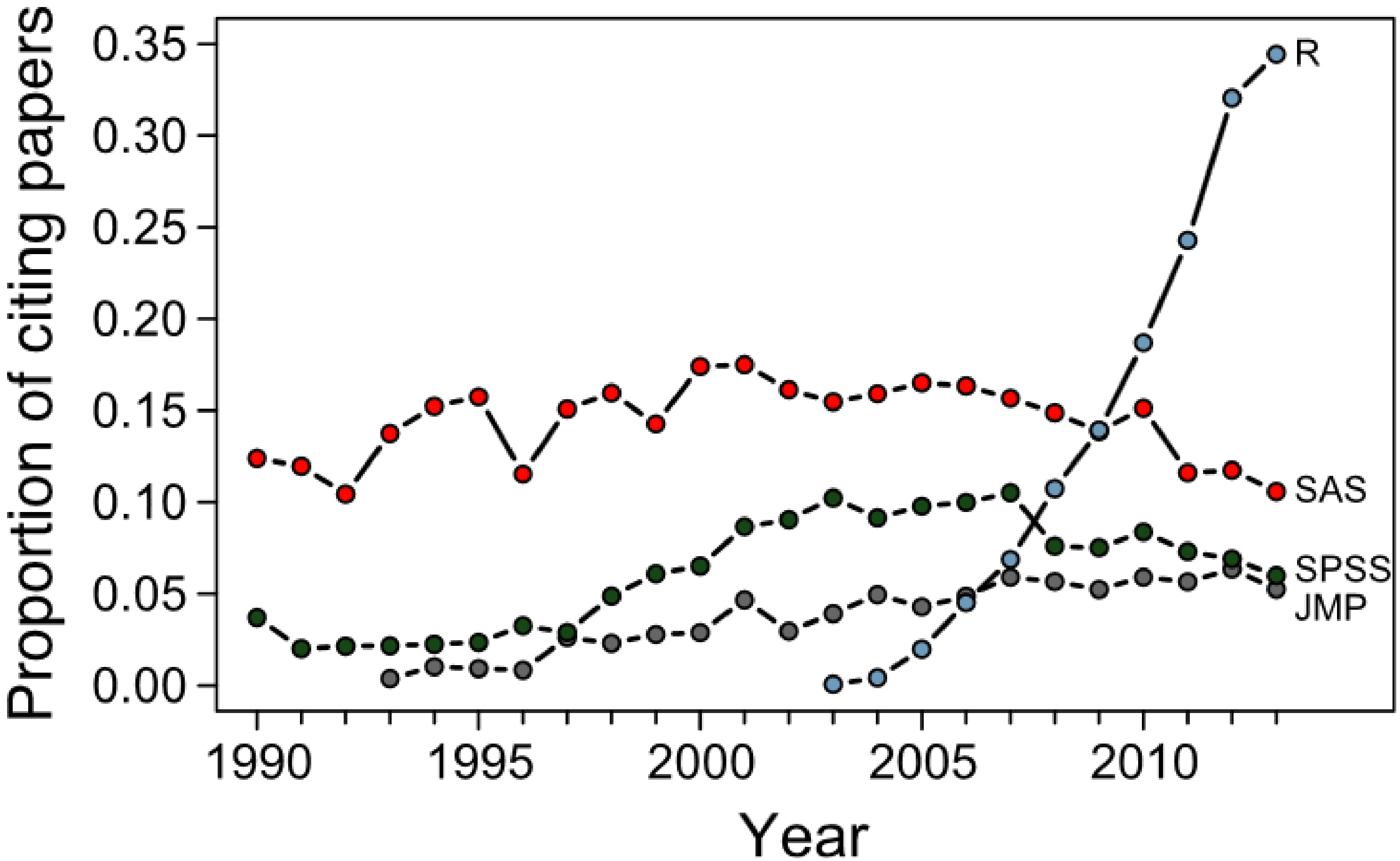
Design

Perceptual Biases

Software

Lab >

1. Base graphics
2. Ggplot2
3. Grid graphics



http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1394/epdf

Fig. 3. Changes in the usage of four leading statistical programs from 1990 to 2013. Gray circles indicate the program JMP, blue circles indicate the program R, red circles indicated the program SAS, and green circles indicate the program SPSS. Data are the proportion of total papers in seven top ecology journals utilizing each technique.

How we organise and present information matters a lot!

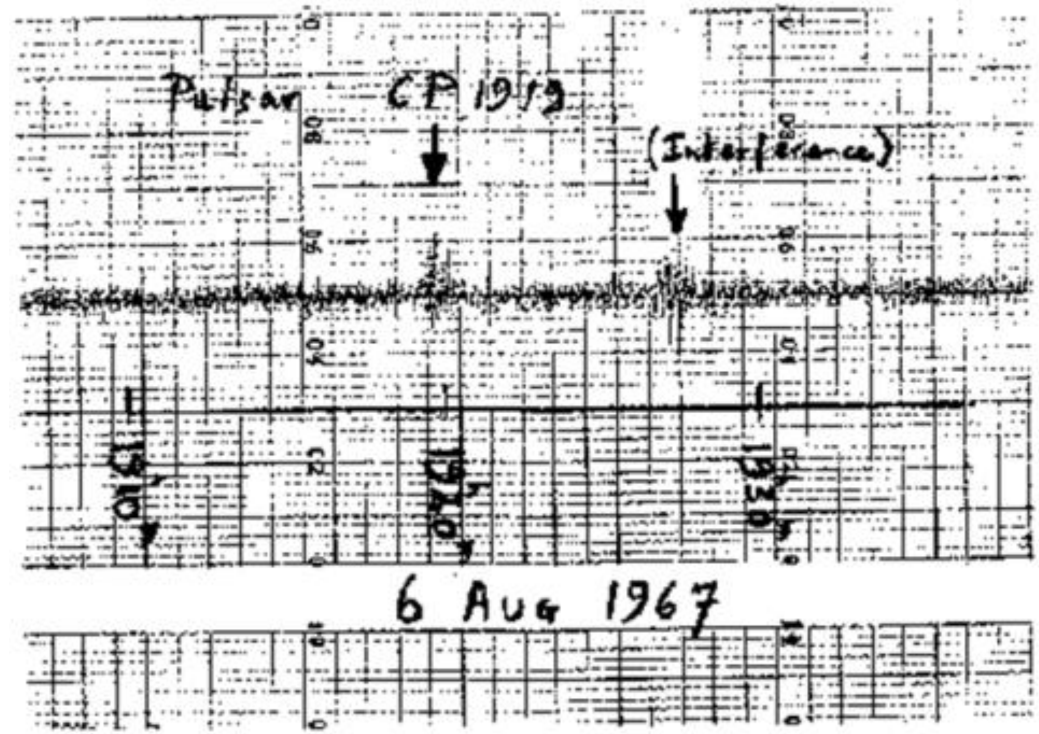
http://vrf.wpengine.netdna-cdn.com/wp-content/uploads/2014/08/organized-02.jpg

http://vrf.wpengine.netdna-cdn.com/wp-content/uploads/2014/08/organized-02.jpg

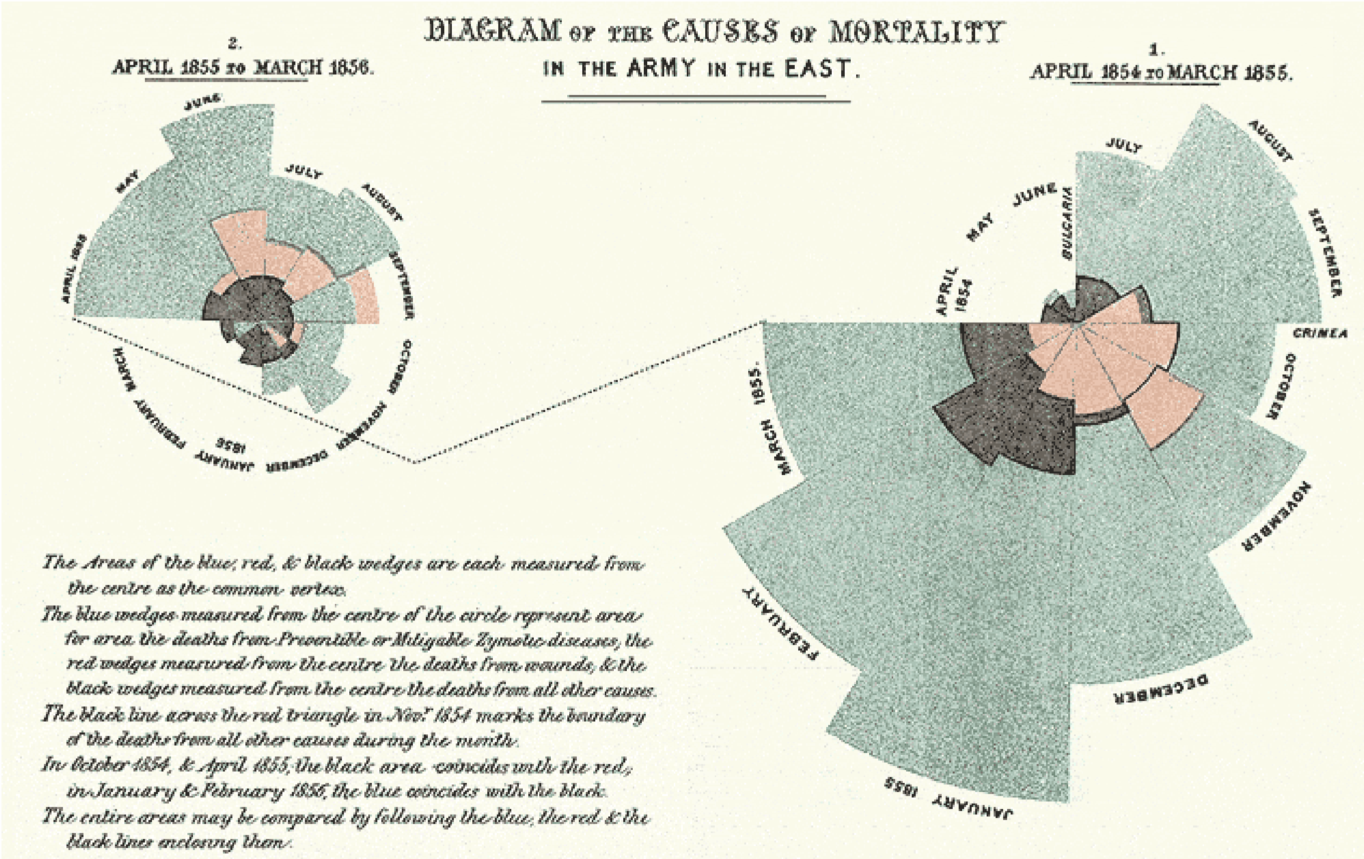
Jocelyn Bell Burnell

Discovery of pulsars

http://www.bbc.co.uk/programmes/b016812j



Lyne, AG & Smith, FG. (1990) *Pulsar Astronomy*. Cambridge University Press.



preventable diseases like typhus killed ten times more troops than battle wounds

http://communicatingdata.org/read-white-paper/

For my wife Melinda and me, the problem of global health inequity became visible 15 years ago, when we saw a simple pie chart in the newspaper breaking down the major causes of death among children.

One of the bigger slices of the pie, representing 500,000 dead children annually, was labelled: rotavirus.

Our reaction was somewhere between disbelief and disgust. How could we not have seen even the barest outlines of this tragedy?

That rotavirus slice in the pie chart set us on fire.

… all of a sudden it didn't seem like there was any time to waste

We decided to do everything we could to get the vaccine out to every child who needed it.

http://www.gatesfoundation.org/media-center/speeches/2013/01/bill-gates-dimbleby-lecture

~~Pretty~~

Design *visualisation* systems

that maximise

cognitive & scientific productivity

[after Ware 2013, G11.1]

The best example of visualisation…

Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

http://en.wikipedia.org/wiki/Anscombe's_quartet#cite_note-Anscombe-1
Anscombe, F. J. (1973). *American Statistician* **27** (1): 17–21.

Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

mean(x) = 9

var(x) = 11

mean(y) = 7.5

var(y) = 4.1

cor(x,y) = 0.816

Linear regression line
→ Y = 0.5x+3

http://en.wikipedia.org/wiki/Anscombe's_quartet#cite_note-Anscombe-1
Anscombe, F. J. (1973). *American Statistician* **27** (1): 17–21.

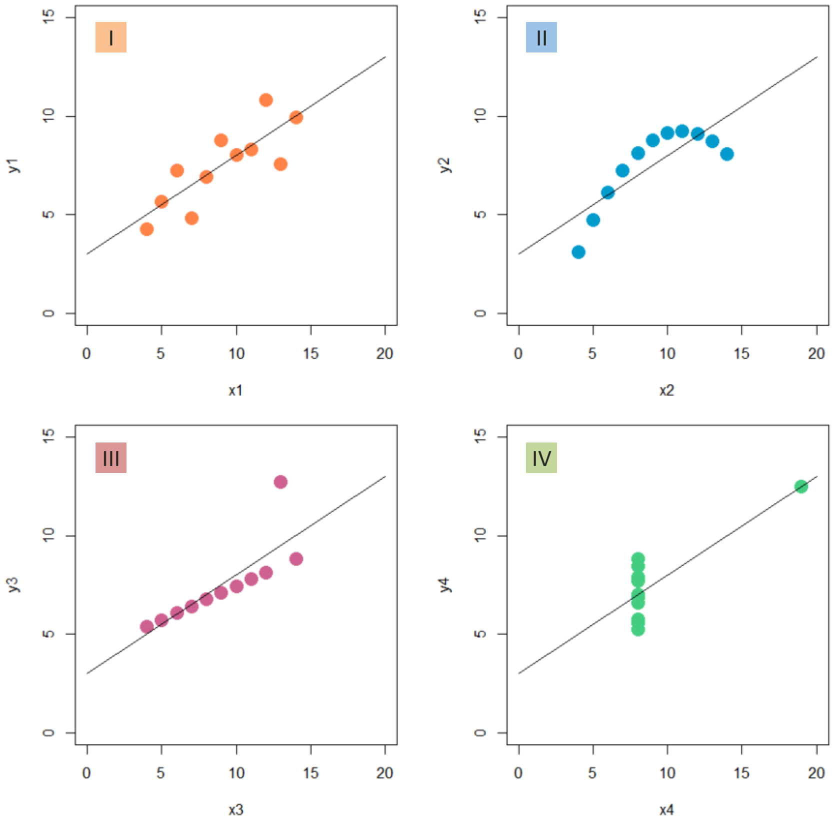
Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

http://en.wikipedia.org/wiki/Anscombe's_quartet#cite_note-Anscombe-1
Anscombe, F. J. (1973). *American Statistician* **27** (1): 17–21.

Anscombe's Quartet

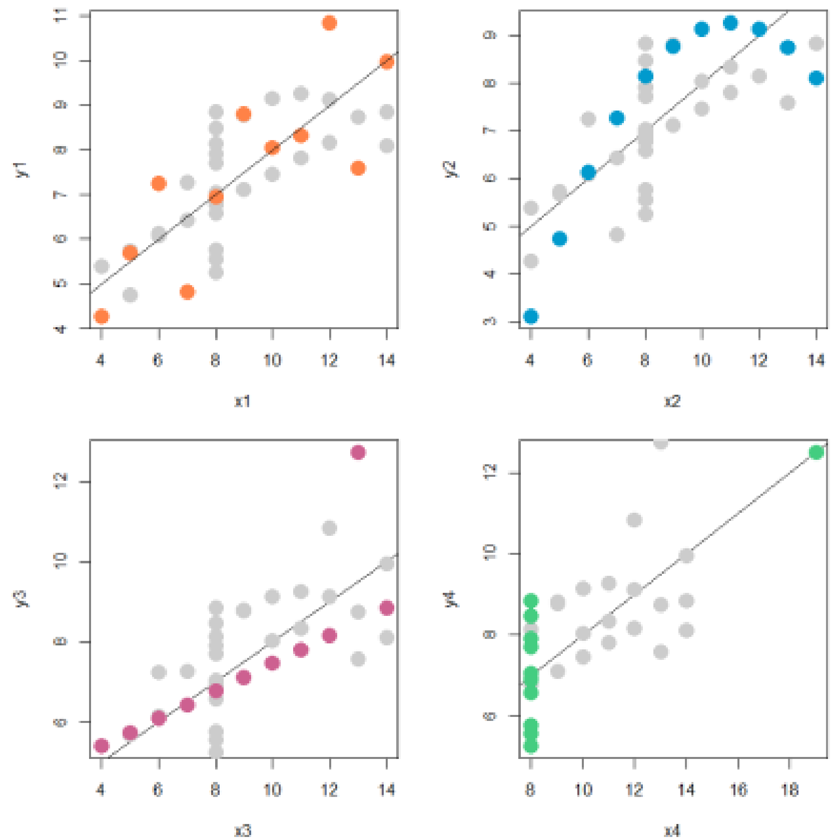
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



http://en.wikipedia.org/wiki/Anscombe's_quartet#cite_note-Anscombe-1
Anscombe, F. J. (1973). *American Statistician* **27** (1): 17–21.

Anscombe's Quartet

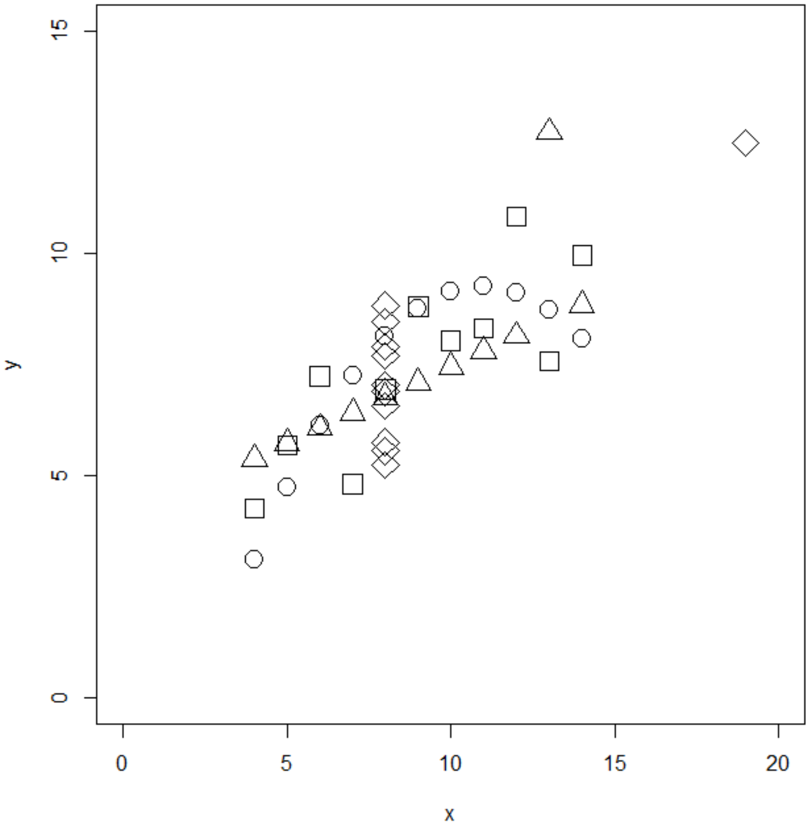
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



http://en.wikipedia.org/wiki/Anscombe's_quartet#cite_note-Anscombe-1
Anscombe, F. J. (1973). *American Statistician* **27** (1): 17–21.

Anscombe's Quartet

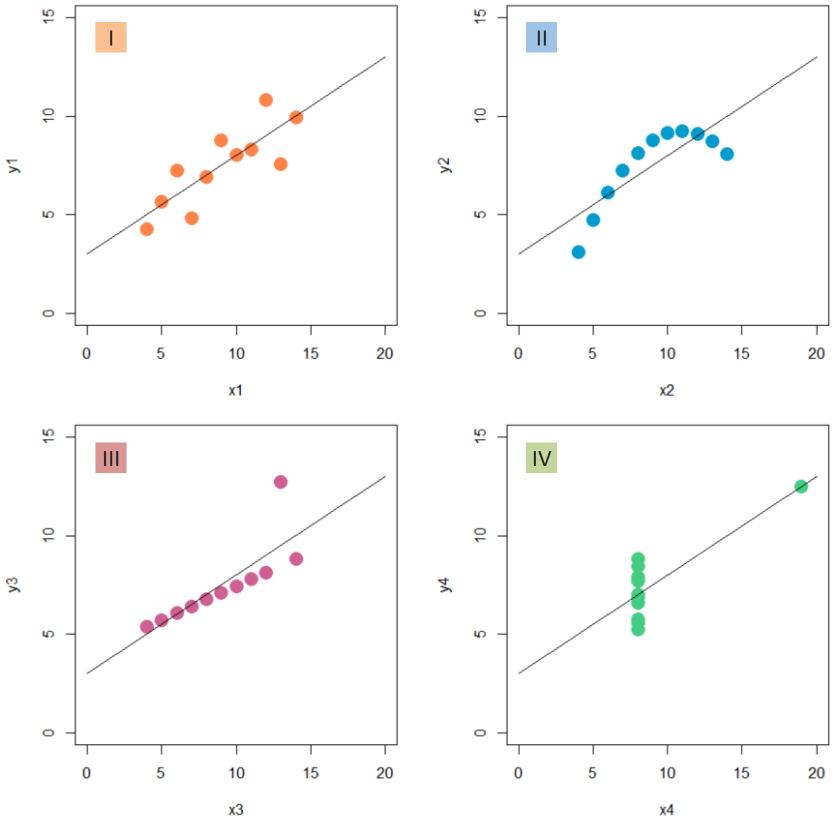
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



http://en.wikipedia.org/wiki/Anscombe's_quartet#cite_note-Anscombe-1
Anscombe, F. J. (1973). *American Statistician* **27** (1): 17–21.

Anscombe’s Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

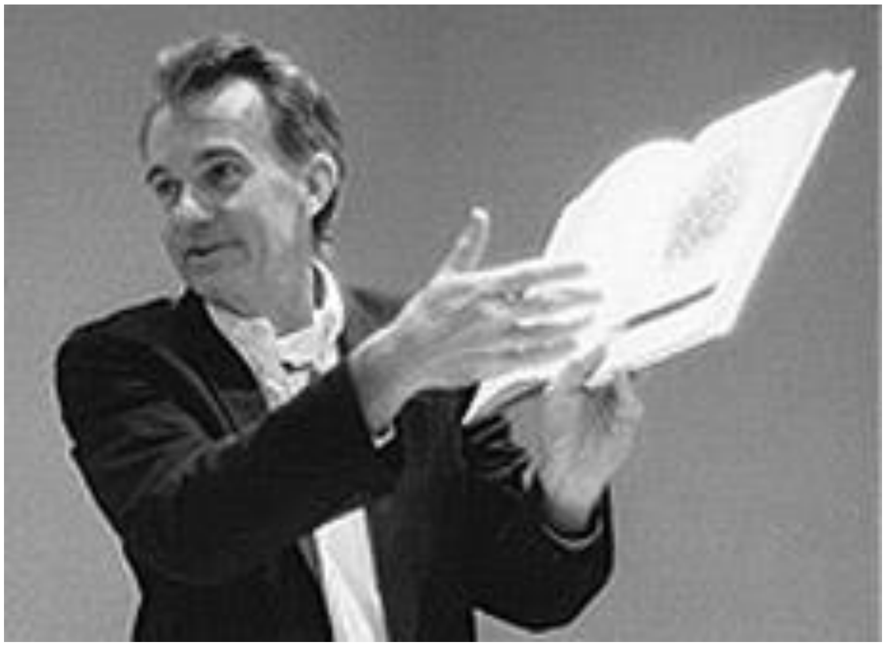


http://en.wikipedia.org/wiki/Anscombe's_quartet#cite_note-Anscombe-1
Anscombe, F. J. (1973). *American Statistician* **27** (1): 17–21.

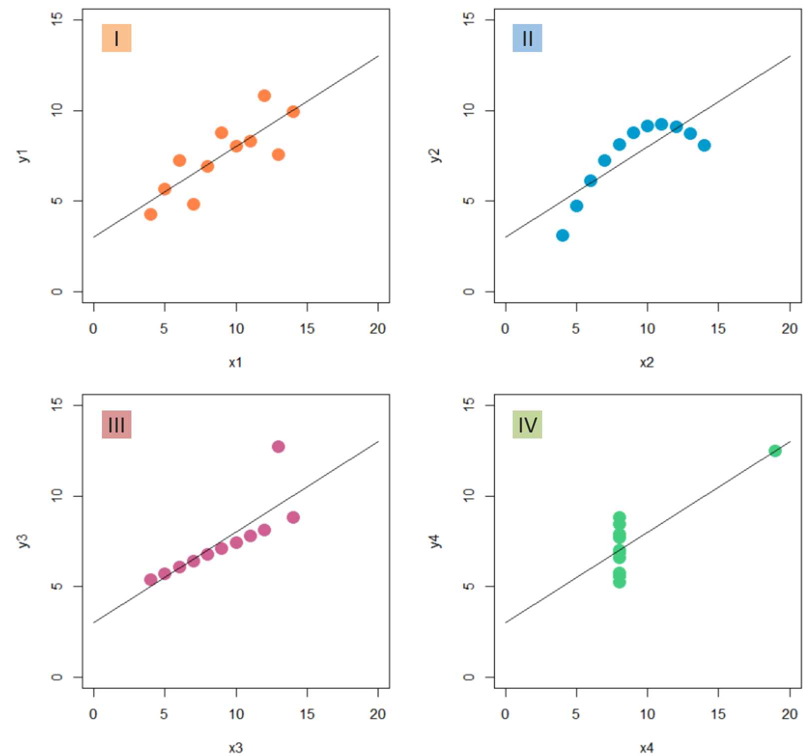
'Small Multiples'…

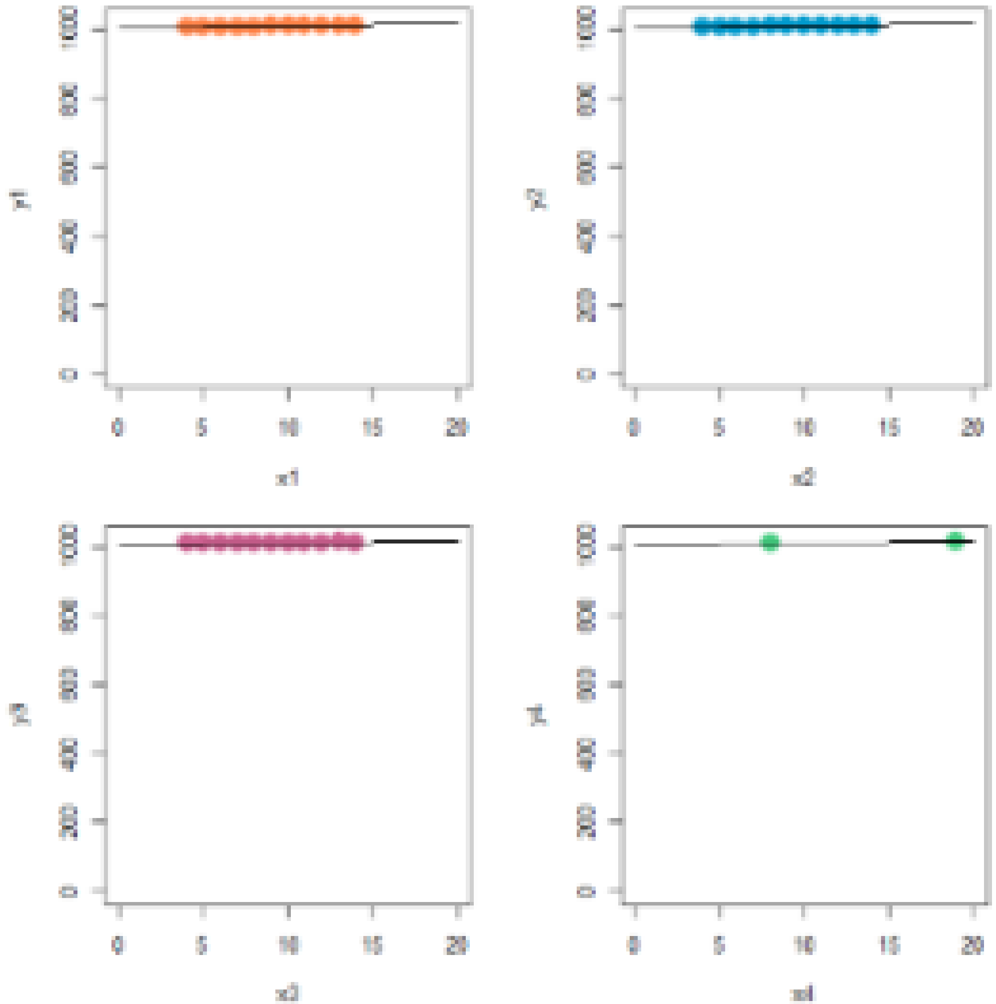
Data are partitioned into a series of plots rather than a single plot.
Reduces 'confusion'… Increases 'salience'.

"visually enforcing comparisons…"

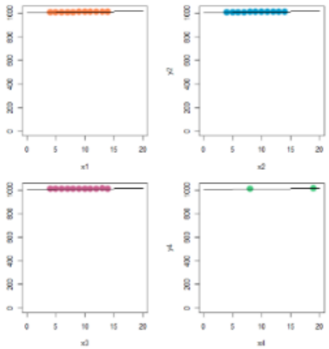


Tufte, Edward (1990). *Envisioning Information*. Graphics Press.

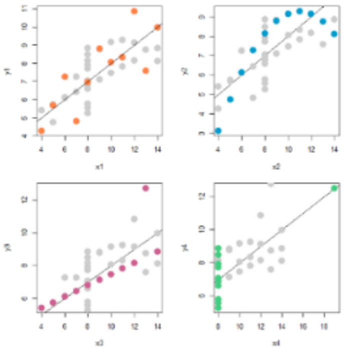




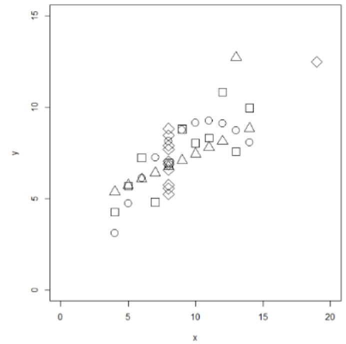
This design doesn’t necessarily work for all data…



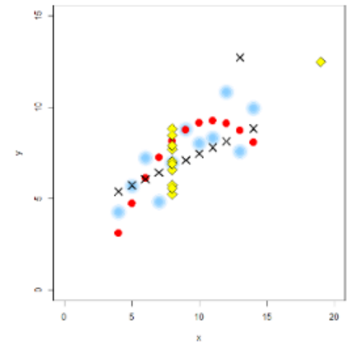
a) Same design but the data values are increased by 100



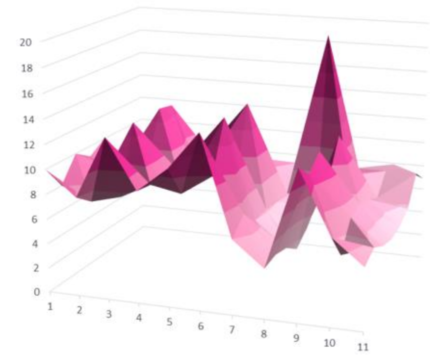
b) Scales fitted to the range of each data set



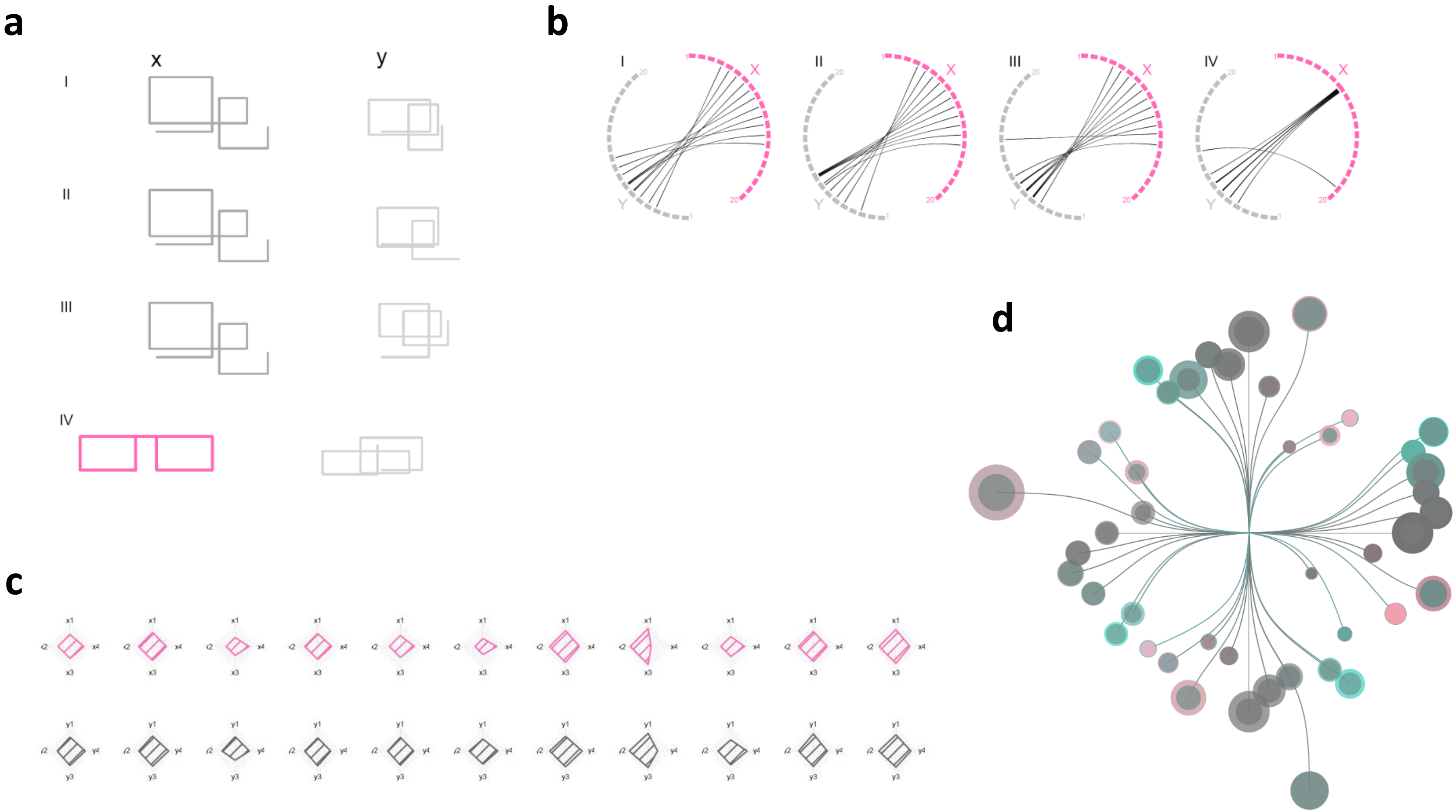
c) Single plot with symbols in black and white



d) Single plot with symbols designed based on pre-attentive processing

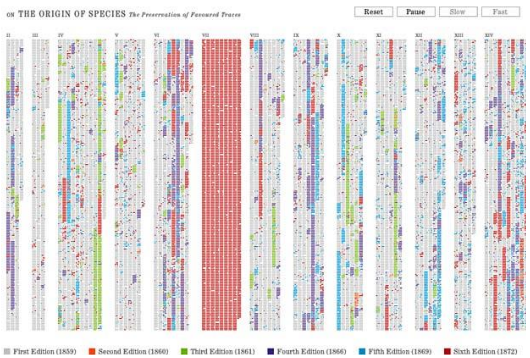


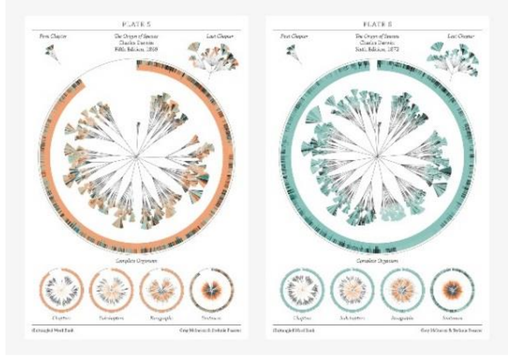
e) Surface plot using Excel default formatting options


a
x
y
I
II
III
IV
b
I
II
III
IV
X
Y
c
x1
x2
x3
x4
y1
y2
y3
y4
d

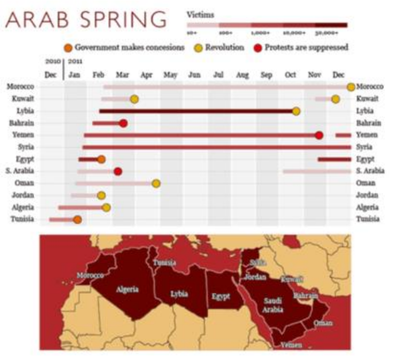
1. Visualisations can reveal.
2. Design is data-dependent
3. There are >1 possibilities

a.i **a.ii**

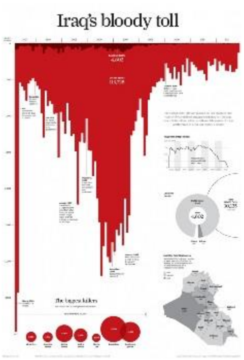


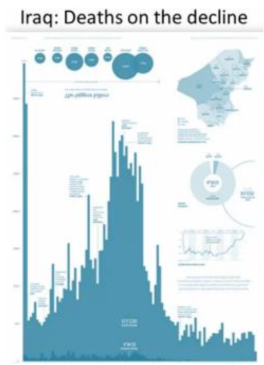


b.i **b.ii**

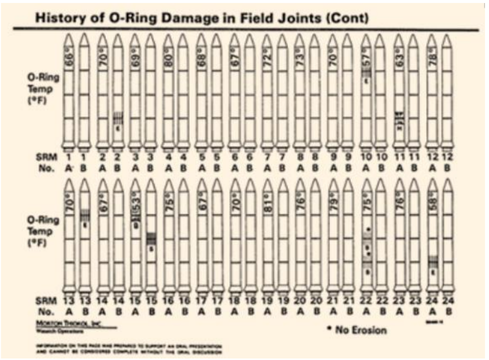


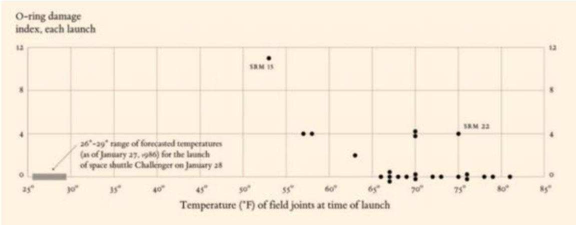
c.i **c.ii**





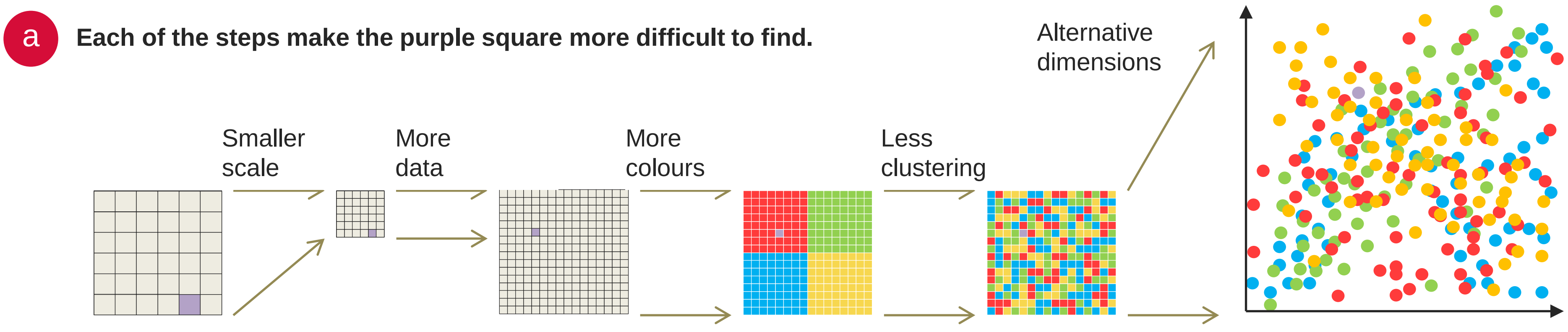
d.i **d.ii**





[1] https://fathom.info/traces/
[2] http://www.stefanieposavec.co.uk/-everything-in-between/#/entangled-word-bank/
[3-6] see Viegas & Wattenberg (2015)
[3] Tufte1990
[4] Tufte 1990
[5] http://www.informationisbeautifulawards.com/showcase/113-arab-spring
[6] http://www.thefunctionalart.com/2015/02/redesigning-circular-timeline.html
[7] http://www.scmp.com/infographics/article/1284683/iraqs-bloody-toll
[8] https://www.youtube.com/watch?v=Ybwh4lejYO4

An example of how science leads design…

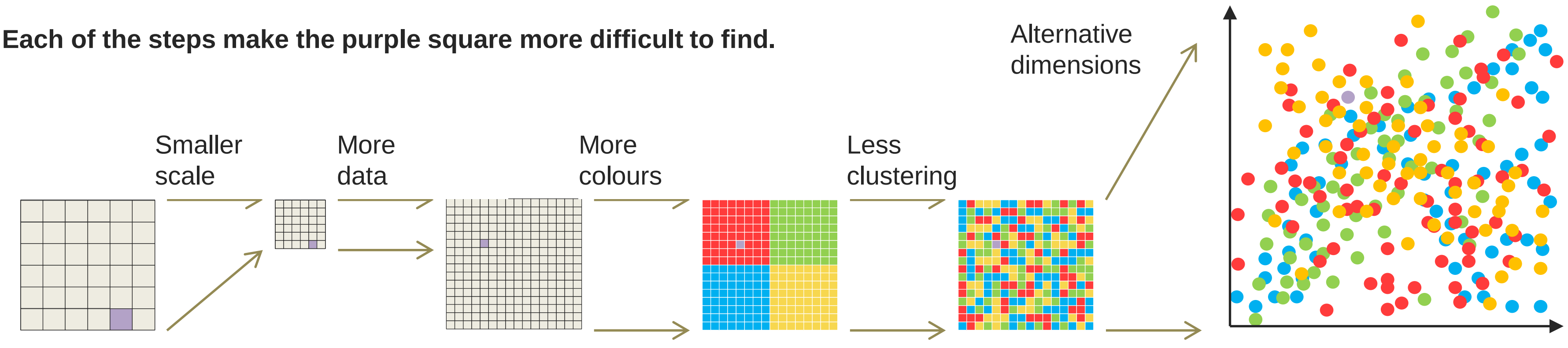


Random data to grouped data set = 1.2 to 2 time quicker

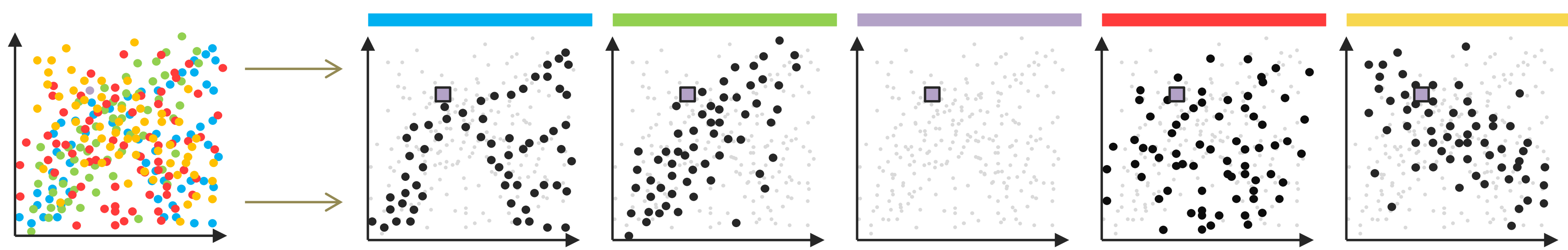
Random 'large' data set to 'smaller' grouped data set = 8 times quicker

Gramazio C, Schloss, K, Laidlaw D. The relation between visualization size, grouping, and user performance. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis). 2014; 20(12): 1953-1962.

a **Each of the steps make the purple square more difficult to find.**

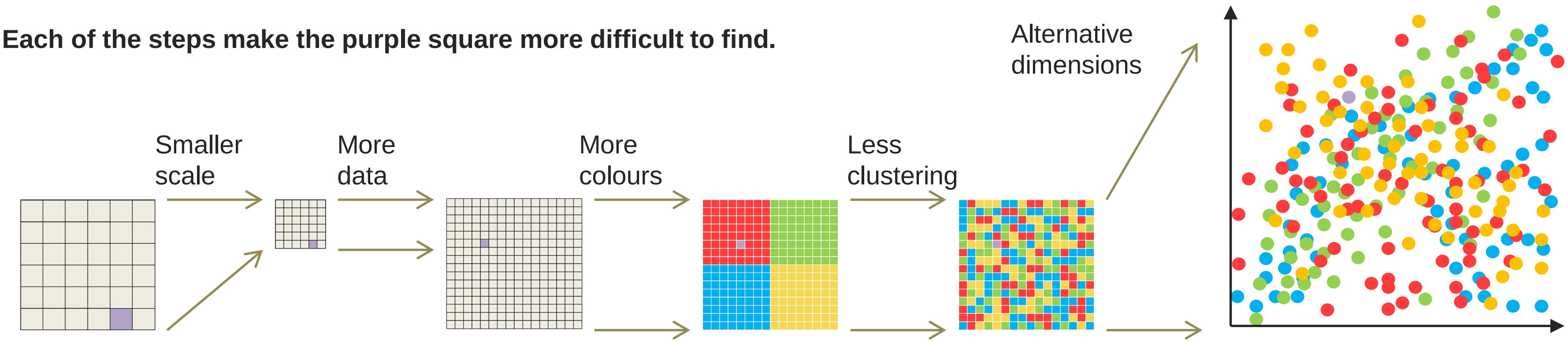


b **‘Small multiples’ negates the need for complicated colour and symbol schemes, allowing the patterns to be set in context.**

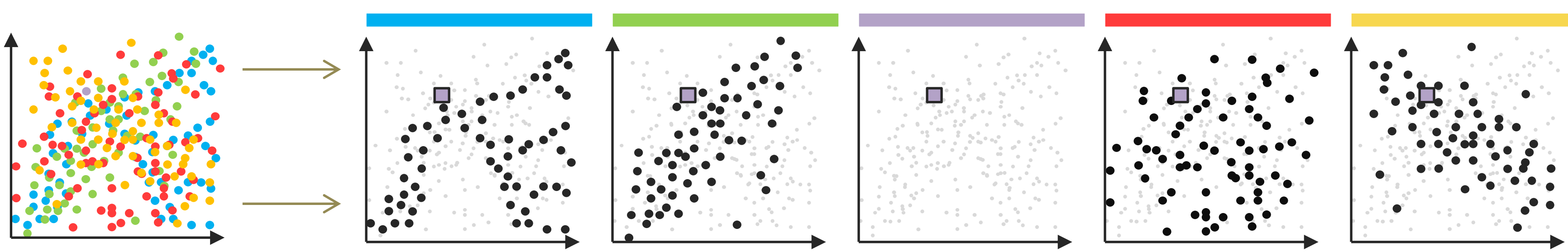


Gramazio C, Schloss, K, Laidlaw D. The relation between visualization size, grouping, and user performance.
IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis). 2014; 20(12): 1953-1962.

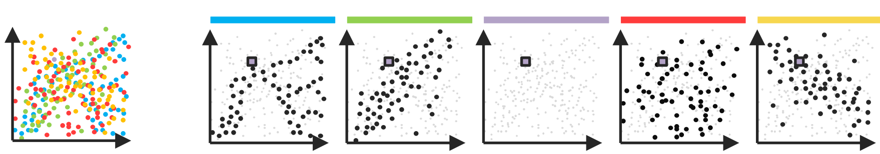
a **Each of the steps make the purple square more difficult to find.**



b **‘Small multiples’ negates the need for complicated colour and symbol schemes, allowing the patterns to be set in context.**



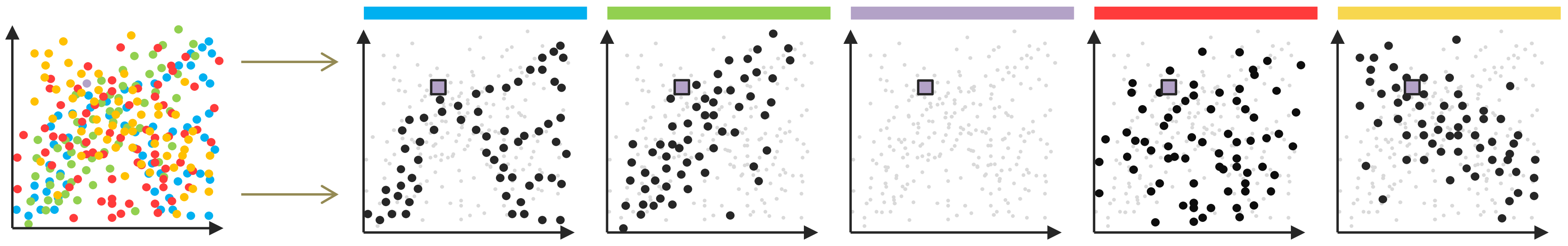
c **Important patterns are still observable in small graphics.**



Gramazio C, Schloss, K, Laidlaw D. The relation between visualization size, grouping, and user performance. IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis). 2014; 20(12): 1953-1962.

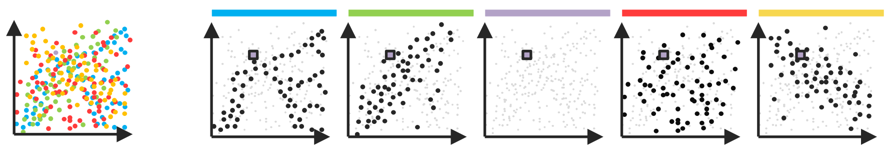
b

'Small multiples' negates the need for complicated colour and symbol schemes, allowing the patterns to be set in context.



c

Important patterns are still observable in small graphics.



a

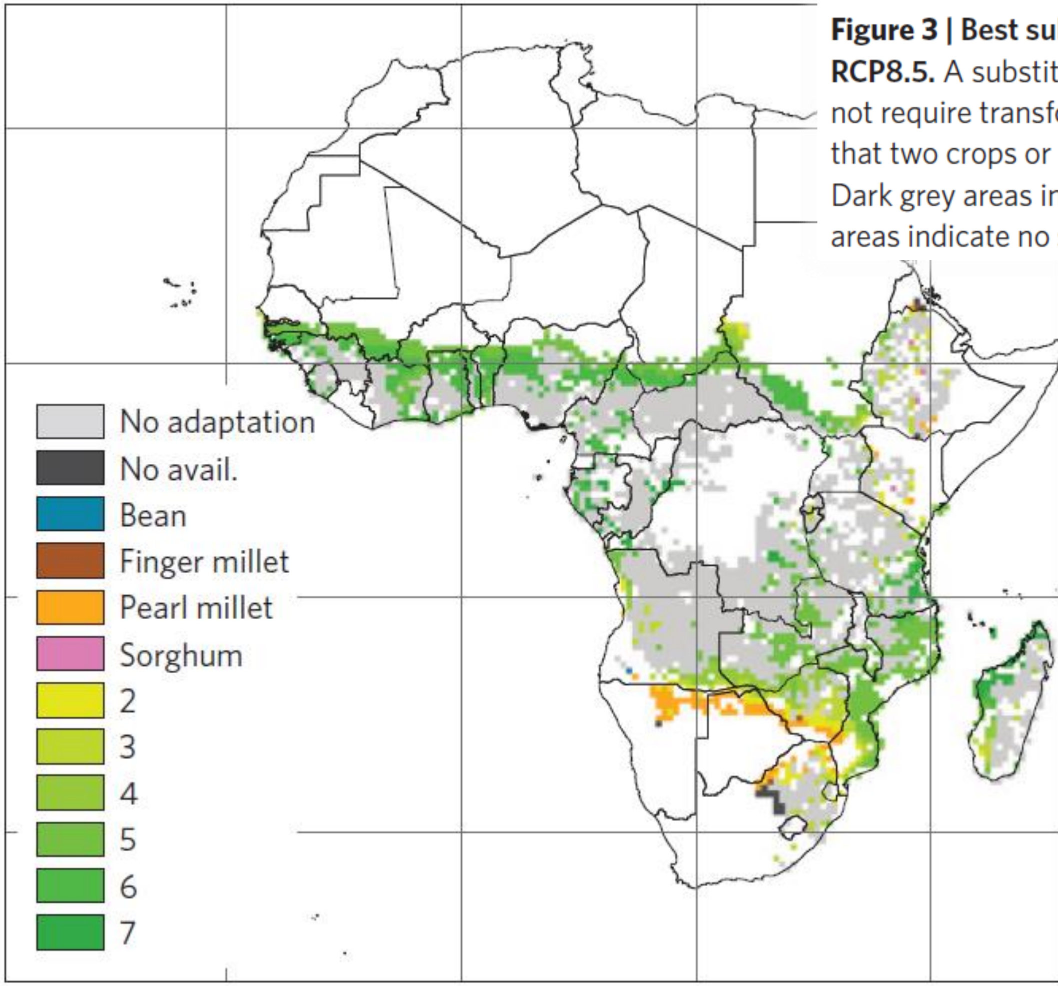


Figure 3 | Best substitute crops at mean time of crossing for maize for RCP8.5. A substitute is defined in a given pixel as a crop that by 2100 does not require transformation. **a**, Map of best substitutes. Green areas indicate that two crops or more can be potential substitutes on a continuous scale. Dark grey areas indicate that no substitution is possible, whereas light grey areas indicate no substitution is needed. **b**, Percentage area (from total area

http://0-www.nature.com.pugwash.lib.warwick.ac.uk/nclimate/journal/vaop/ncurrent/pdf/nclimate2947.pdf

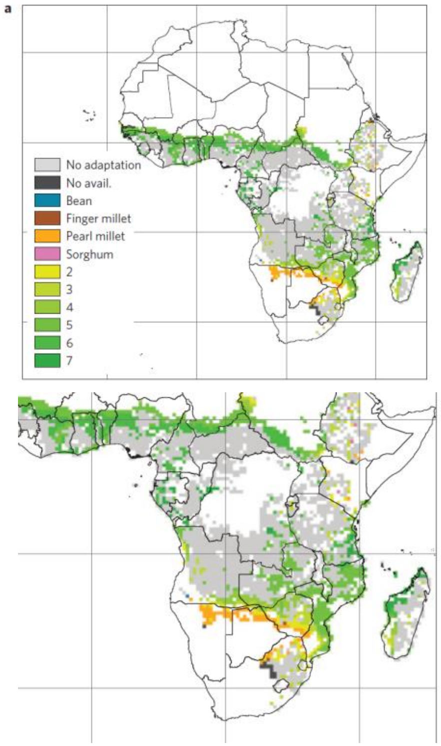


Figure 3 | Best substitute crops at mean time of crossing for maize for RCP8.5. A substitute is defined in a given pixel as a crop that by 2100 does not require transformation. **a**, Map of best substitutes. Green areas indicate that two crops or more can be potential substitutes on a continuous scale. Dark grey areas indicate that no substitution is possible, whereas light grey areas indicate no substitution is needed. **b**, Percentage area (from total area

http://0-www.nature.com.pugwash.lib.warwick.ac.uk/nclimate/journal/vaop/ncurrent/pdf/nclimate2947.pdf

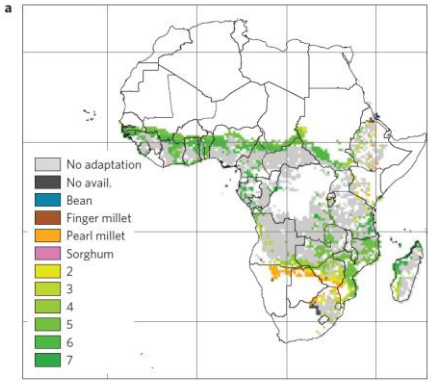
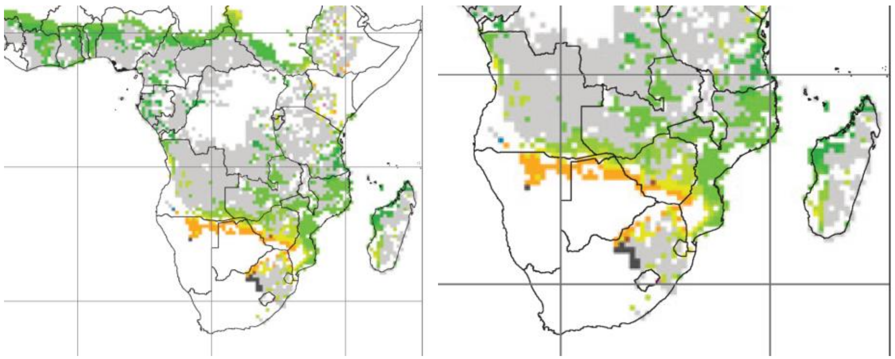
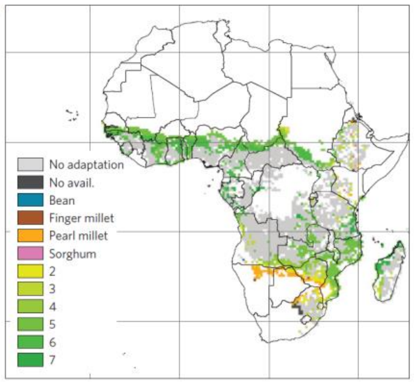


Figure 3 | Best substitute crops at mean time of crossing for maize for RCP8.5. A substitute is defined in a given pixel as a crop that by 2100 does not require transformation. **a**, Map of best substitutes. Green areas indicate that two crops or more can be potential substitutes on a continuous scale. Dark grey areas indicate that no substitution is possible, whereas light grey areas indicate no substitution is needed. **b**, Percentage area (from total area



http://0-www.nature.com.pugwash.lib.warwick.ac.uk/nclimate/journal/vaop/ncurrent/pdf/nclimate2947.pdf

a



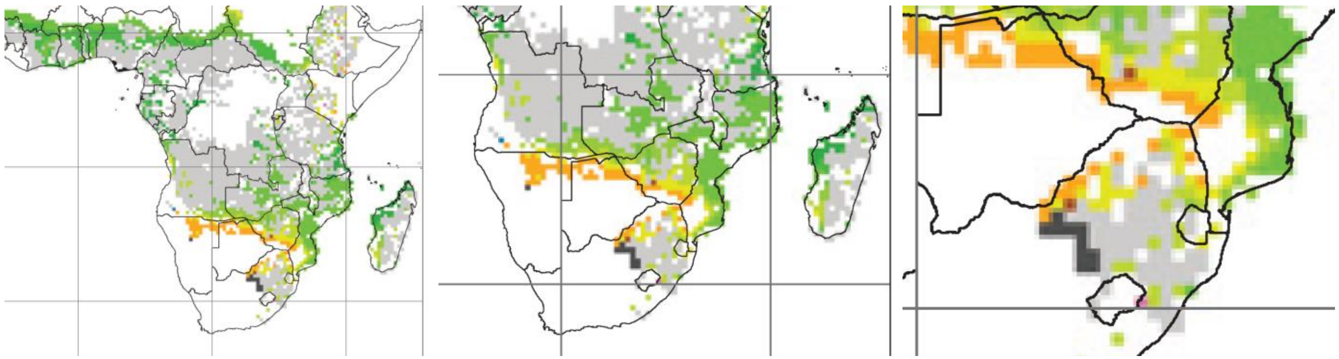


Figure 3 | Best substitute crops at mean time of crossing for maize for RCP8.5. A substitute is defined in a given pixel as a crop that by 2100 does not require transformation. **a**, Map of best substitutes. Green areas indicate that two crops or more can be potential substitutes on a continuous scale. Dark grey areas indicate that no substitution is possible, whereas light grey areas indicate no substitution is needed. **b**, Percentage area (from total area

http://0-www.nature.com.pugwash.lib.warwick.ac.uk/nclimate/journal/vaop/ncurrent/pdf/nclimate2947.pdf

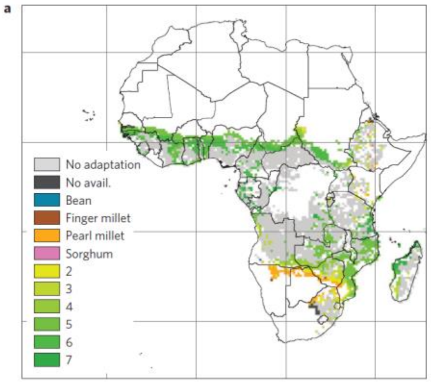
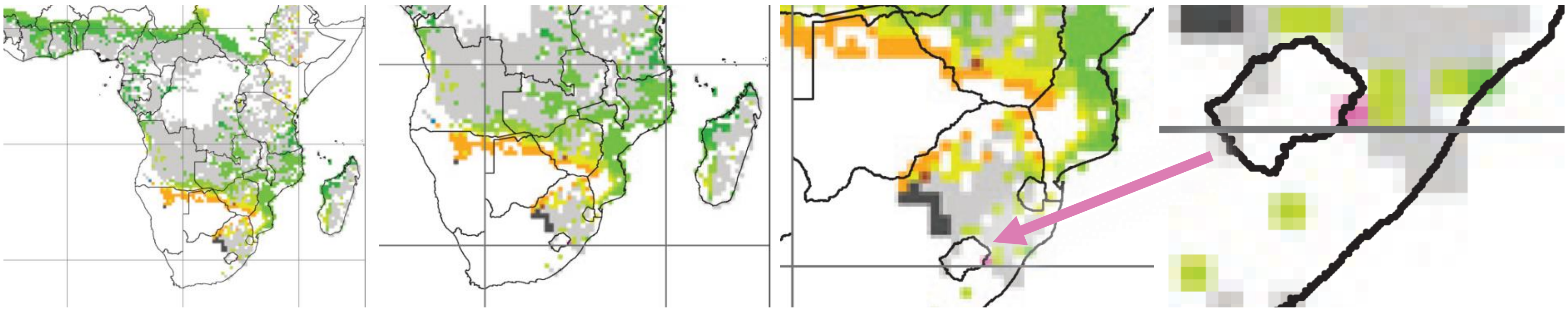
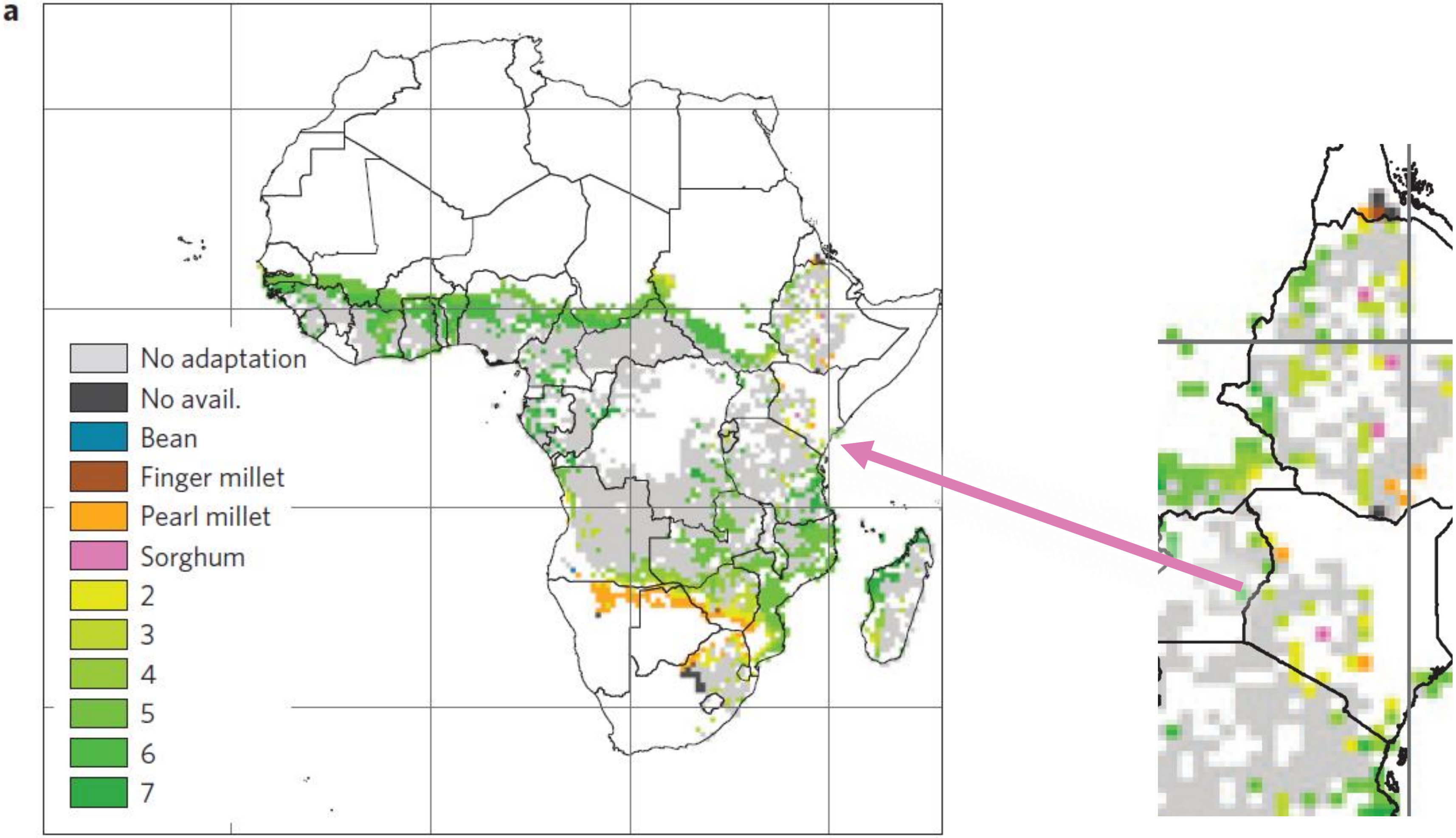


Figure 3 | Best substitute crops at mean time of crossing for maize for RCP8.5. A substitute is defined in a given pixel as a crop that by 2100 does not require transformation. **a**, Map of best substitutes. Green areas indicate that two crops or more can be potential substitutes on a continuous scale. Dark grey areas indicate that no substitution is possible, whereas light grey areas indicate no substitution is needed. **b**, Percentage area (from total area

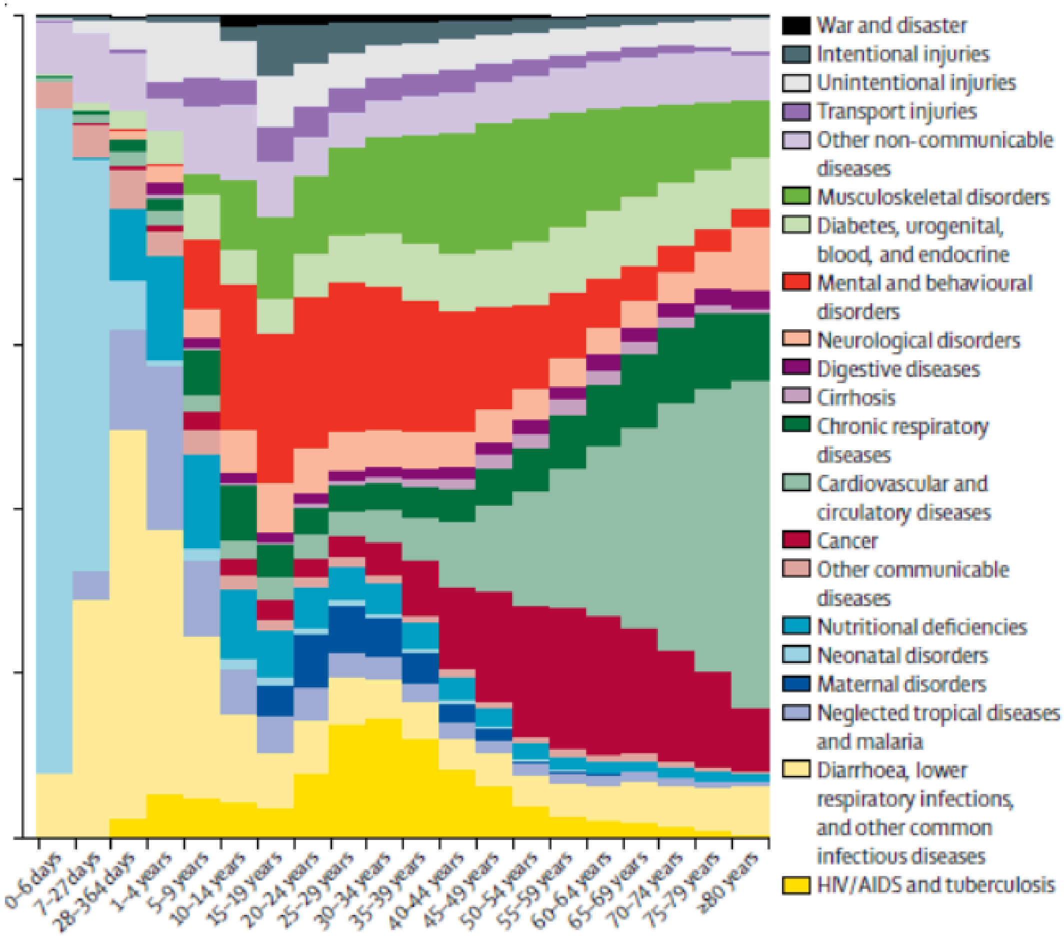


http://0-www.nature.com.pugwash.lib.warwick.ac.uk/nclimate/journal/vaop/ncurrent/pdf/nclimate2947.pdf



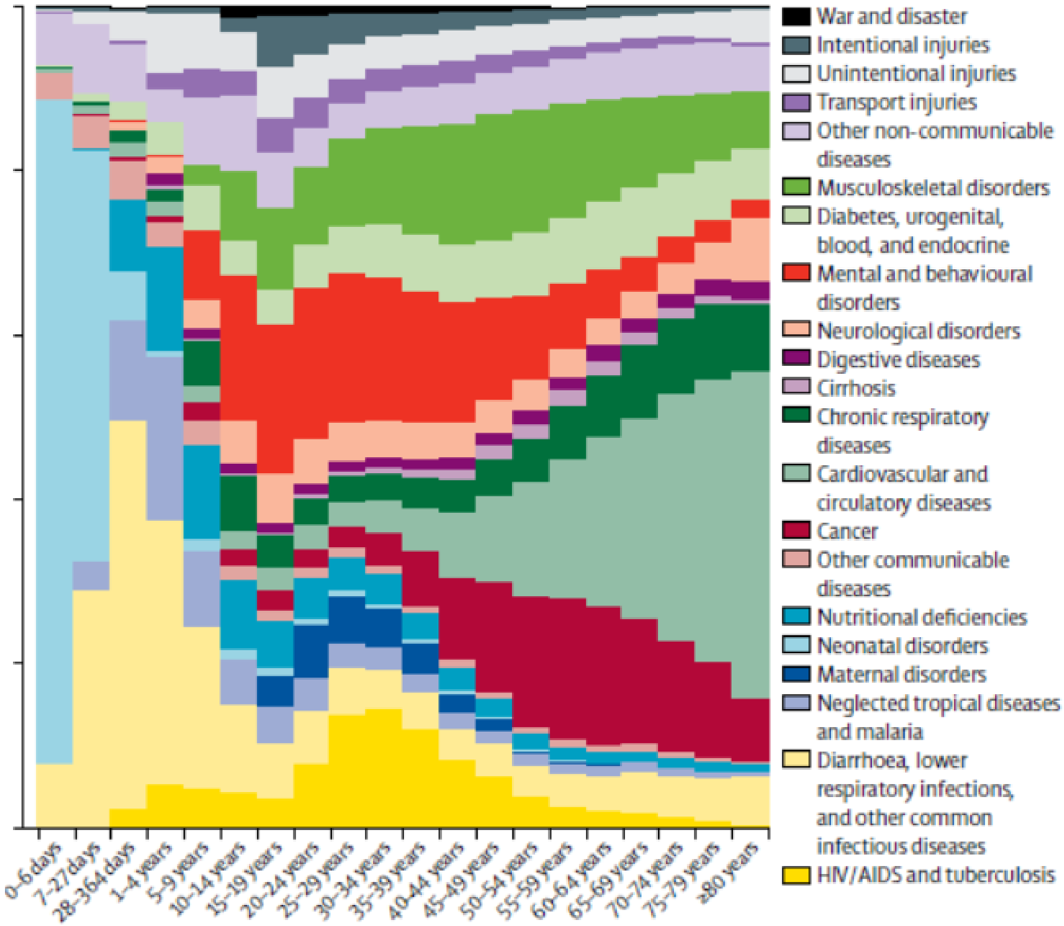
http://0-www.nature.com.pugwash.lib.warwick.ac.uk/nclimate/journal/vaop/ncurrent/pdf/nclimate2947.pdf

All visualisations are inevitably biased…



Nick Golding @_NickGolding_ · Dec 2

This figure in a recent paper made my eyes hurt - any suggestions to better visualise these proportions? pic.twitter.com/uMix0Sv9E1



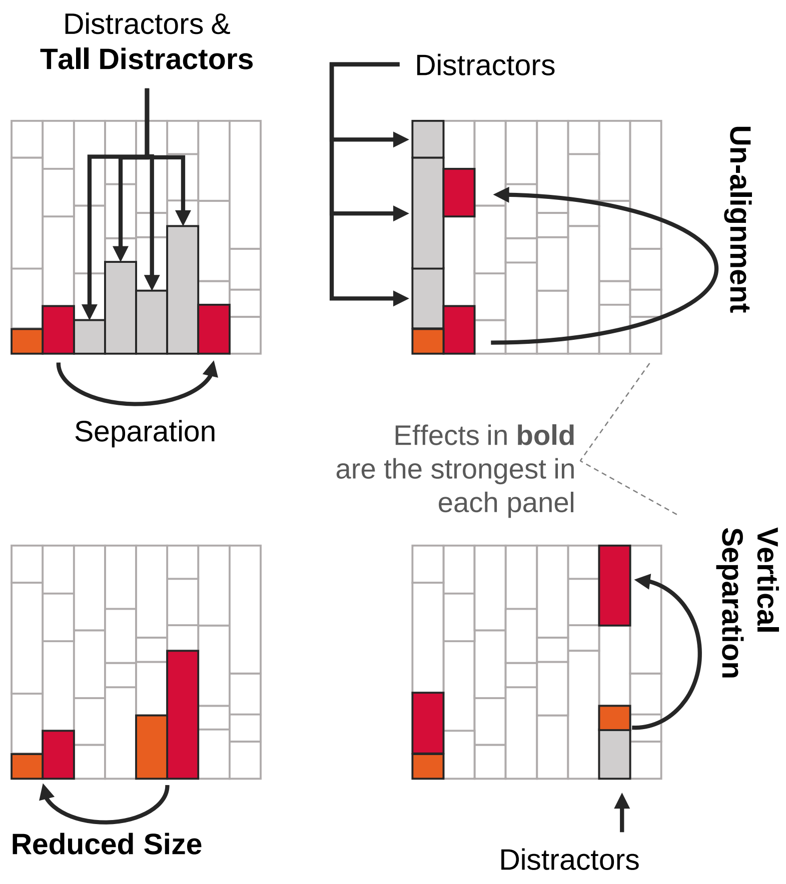
"It looks like a fruit salad with lots of watermelon…"

Esther, aged 10.

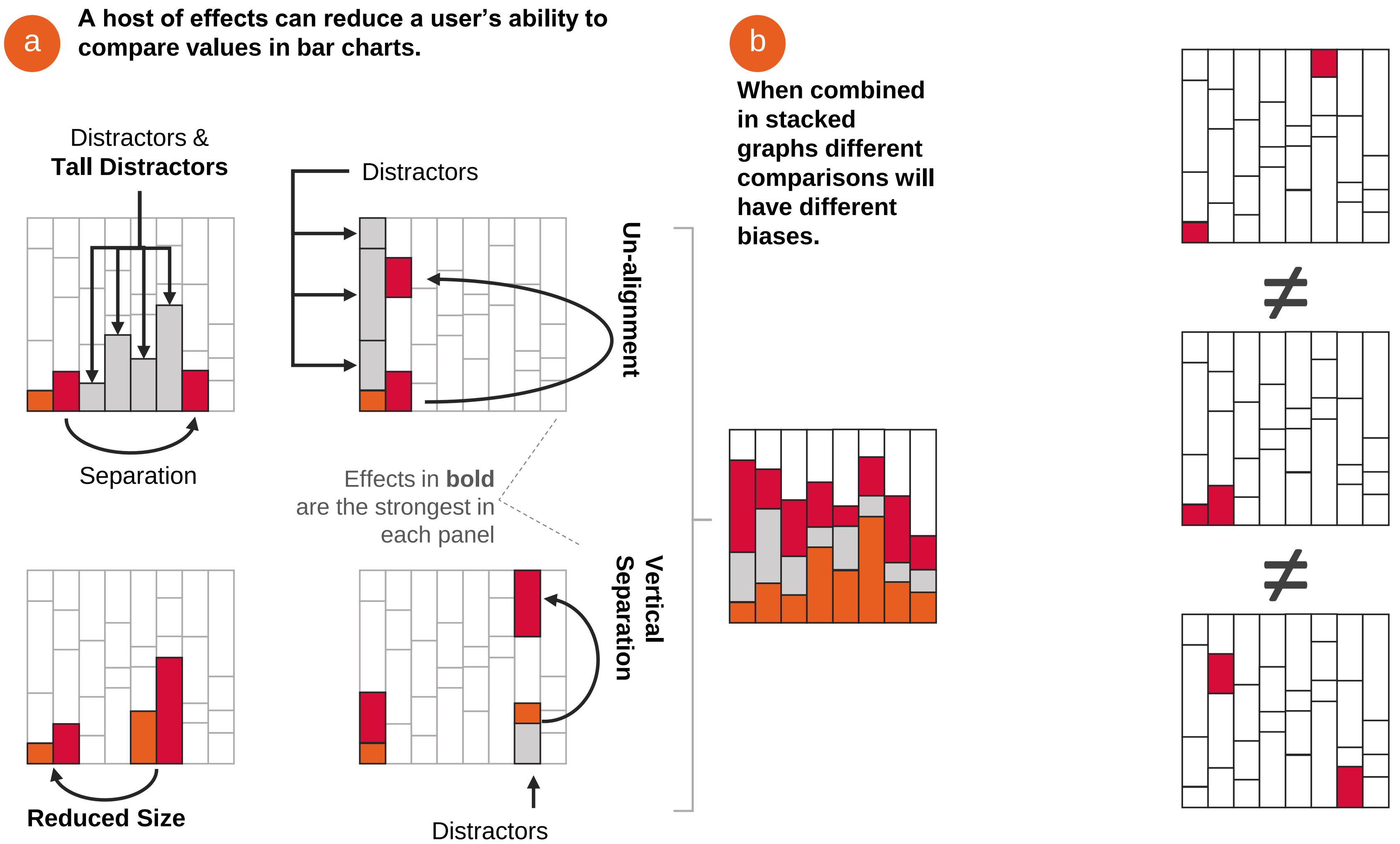
Nick Golding @_NickGolding_ · Dec 2
This figure in a recent paper made my eyes hurt - any suggestions to better visualise these proportions? pic.twitter.com/uMix0Sv9E1

a

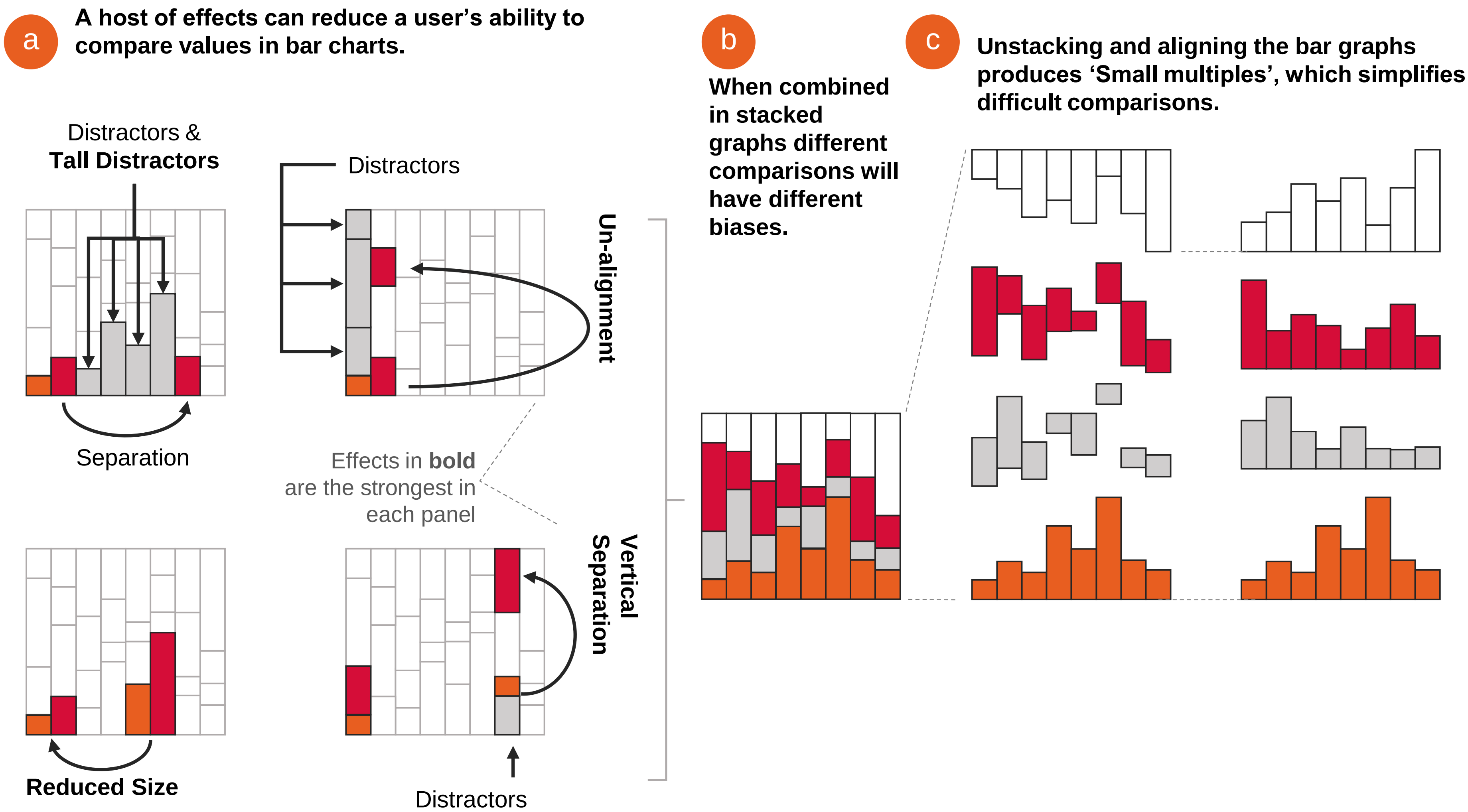
A host of effects can reduce a user's ability to compare values in bar charts.



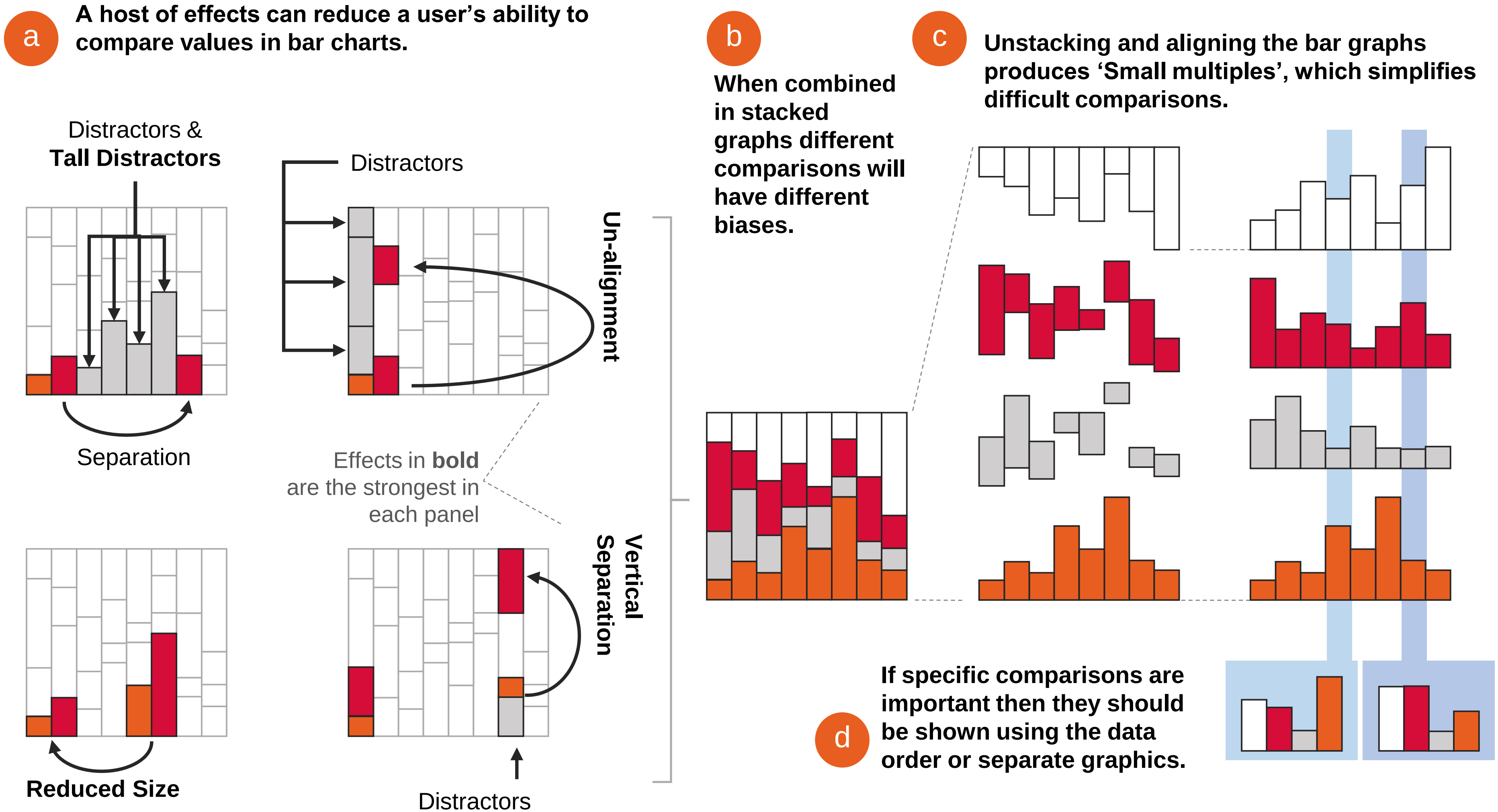
Talbot J, Setlur V, Anand A. Four Experiments on the Perception of Bar Charts.
IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis). 2014; 20(12): 2152 - 2160.



Talbot J, Setlur V, Anand A. Four Experiments on the Perception of Bar Charts.
IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis). 2014; 20(12): 2152 - 2160.



Talbot J, Setlur V, Anand A. Four Experiments on the Perception of Bar Charts.
IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis). 2014; 20(12): 2152 - 2160.



Talbot J, Setlur V, Anand A. Four Experiments on the Perception of Bar Charts.
IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis). 2014; 20(12): 2152 - 2160.

Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010

Christopher J L Murray†‡, Theo Vos, Rafael Lozano, Mohsen Naghavi, Abraham D Flaxman, Catherine Michaud, Majid Ezzati, Kenji Shibuya, Joshua A Salomon, Safa Abdalla*, Victor Aboyans*, Jerry Abraham*, Ilana Ackerman*, Rakesh Aggarwal*, Stephanie Y Ahn*, Mohammed K Ali*, Mohammad A AlMazroa*, Miriam Alvarado*, H Ross Anderson*, Laurie M Anderson*, Kathryn G Andrews*, Charles Atkinson*, Larry M Baddour*, Adil N Bahalim*, Suzanne Barker-Collo*, Lope H Barrero*, David H Bartels*, Maria-Gloria Basáñez*, Amanda Baxter*, Michelle L Bell*, Emelia J Benjamin*, Derrick Bennett*, Eduardo Bernabé*, Kavi Bhalla*, Bishal Bhandari*, Boris Bikbov*, Aref Bin Abdulhak*, Gretchen Birbeck*, James A Black*, Hannah Blencowe*, Jed D Blore*, Fiona Blyth*, Ian Bolliger*, Audrey Bonaventure*, Soufiane Boufous*, Rupert Bourne*, Michel Boussinesq*, Tasanee Braithwaite*, Carol Brayne*, Lisa Bridgett*, Simon Brooker*, Peter Brooks*, Traolach S Brugha*, Claire Bryan-Hancock*, Chiara Bucello*, Rachelle Buchbinder*, Geoffrey Buckle*, Christine M Budke*, Michael Burch*, Peter Burney*, Roy Burstein*, Bianca Calabria*, Benjamin Campbell*, Charles E Canter*, Hélène Carabin*, Jonathan Carapetis*, Loreto Carmona*, Claudia Cella*, Fiona Charlson*, Honglei Chen*, Andrew Tai-Ann Cheng*, David Chou*, Sumeet S Chugh*, Luc E Coffeng*, Steven D Colan*, Samantha Colquhoun*, K Ellicott Colson*, John Condon*, Myles D Connor*, Leslie T Cooper*, Matthew Corriere*, Monica Cortinovis*, Karen Courville de Vaccaro*, William Couser*, Benjamin C Cowie*, Michael H Criqui*, Marita Cross*, Kaustubh C Dabhadkar*, Manu Dahiya*, Nabila Dahodwala*, James Damsere-Derry*, Goodarz Danaei*, Adrian Davis*, Diego De Leo*, Louisa Degenhardt*, Robert Dellavalle*, Allyne Delossantos*, Julie Denenberg*, Sarah Derrett*, Don C Des Jarlais*, Samath D Dharmaratne*, Mukesh Dherani*, Cesar Diaz-Torne*, Helen Dolk*, E Ray Dorsey*, Tim Driscoll*, Herbert Duber*, Beth Ebel*, Karen Edmond*, Alexis Elbaz*, Suad Eltahir Ali*, Holly Erskine*, Patricia J Erwin*, Patricia Espindola*, Stalin E Ewoigbokhan*, Farshad Farzadfar*, Valery Feigin*, David T Felson*, Alize Ferrari*, Cleusa P Ferri*, Eric M Fèvre*, Mariel M Finucane*, Seth Flaxman*, Louise Flood*, Kyle Foreman*, Mohammad H Forouzanfar*, Francis Gerry R Fowkes*, Marlene Fransen*, Michael K Freeman*, Belinda J Gabbe*, Sherine E Gabriel*, Emmanuela Gakidou*, Hammad A Ganatra*, Bianca Garcia*, Flavio Gaspari*, Richard F Gillum*, Gerhard Gmel*, Diego Gonzalez-Medina*, Richard Gosselin*, Rebecca Grainger*, Bridget Grant*, Justina Groeger*, Francis Guillemin*, David Gunnell*, Ramyani Gupta*, Juanita Haagsma*, Holly Hagan*, Yara A Halasa*, Wayne Hall*, Diana Haring*, Josep Maria Haro*, James E Harrison*, Rasmus Havmoeller*, Roderick J Hay*, Hideki Higashi*, Catherine Hill*, Bruno Hoen*, Howard Hoffman*, Peter J Hotez*, Damian Hoy*, John J Huang*, Sydney E Ibeanusi*, Kathryn H Jacobsen*, Spencer L James*, Deborah Jarvis*, Rashmi Jasrasaria*, Sudha Jayaraman*, Nicole Johns*, Jost B Jonas*, Ganesan Karthikeyan*, Nicholas Kassebaum*, Norito Kawakami*, Andre Keren*, Jon-Paul Khoo*, Charles H King*, Lisa Marie Knowlton*, Olive Kobusingye*, Adofo Koranteng*, Rita Krishnamurthi*, Francine Laden*, Ratilal Lalloo*, Laura L Laslett*, Tim Lathlean*, Janet L Leasher*, Yong Yi Lee*, James Leigh*, Daphna Levinson*, Stephen S Lim*, Elizabeth Limb*, John Kent Lin*, Michael Lipnick*, Steven E Lipshultz*, Wei Liu*, Maria Loane*, Summer Lockett Ohno*, Ronan Lyons*, Jacqueline Mabweijano*, Michael F MacIntyre*, Reza Malekzadeh*, Leslie Mallinger*, Sivabalan Manivannan*, Wagner Marcenes*, Lyn March*, David J Margolis*, Guy B Marks*, Robin Marks*, Akira Matsumori*, Richard Matzopoulos*, Bongani M Mayosi*, John H McAnulty*, Mary M McDermott*, Neil McGill*, John McGrath*, Maria Elena Medina-Mora*, Michele Meltzer*, Ziad A Memish*, George A Mensah*, Tony R Merriman*, Ana-Claire Meyer*, Valeria Miglioli*, Matthew Miller*, Ted R Miller*, Philip B Mitchell*, Charles Mock*, Ana Olga Mocumbi*, Terrie E Moffitt*, Ali A Mokdad*, Lorenzo Monasta*, Marcella Montico*, Maziar Moradi-Lakeh*, Andrew Moran*, Lidia Morawska*, Rintaro Mori*, Michele E Murdoch*, Michael K Mwaniki*, Kovin Naidoo*, M Nathan Nair*, Luigi Naldi*, K M Venkat Narayan*, Paul K Nelson*, Robert G Nelson*, Michael C Nevitt*, Charles R Newton*, Sandra Nolte*, Paul Norman*, Rosana Norman*, Martin O'Donnell*, Simon O'Hanlon*, Casey Olives*, Saad B Omer*, Katrina Ortblad*, Richard Osborne*, Doruk Ozgediz*, Andrew Page*, Bishnu Pahari*, Jeyaraj Durai Pandian*, Andrea Panozo Rivero*, Scott B Patten*, Neil Pearce*, Rogelio Perez Padilla*, Fernando Perez-Ruiz*, Norberto Perico*, Konrad Pesudovs*, David Phillips*, Michael R Phillips*, Kelsey Pierce*, Sébastien Pion*, Guilherme V Polanczyk*, Suzanne Polinder*, C Arden Pope III*, Svetlana Popova*, Esteban Porrini*, Farshad Pourmalek*, Martin Prince*, Rachel L Pullan*, Kapa D Ramaiah*, Dharani Ranganathan*, Homie Razavi*, Mathilda Regan*, Jürgen T Rehm*, David B Rein*, Guiseppe Remuzzi*, Kathryn Richardson*, Frederick P Rivara*, Thomas Roberts*, Carolyn Robinson*, Felipe Rodriguez De Leòn*, Luca Ronfani*, Robin Room*, Lisa C Rosenfeld*, Lesley Rushton*, Ralph L Sacco*, Sukanta Saha*, Uchechukwu Sampson*, Lidia Sanchez-Riera*, Ella Sanman*, David C Schwebel*, James Graham Scott*, Maria Segui-Gomez*, Saeid Shahraz*, Donald S Shepard*, Hwashin Shin*, Rupak Shivakoti*, David Singh*, Gitanjali M Singh*, Jasvinder A Singh*, Jessica Singleton*, David A Sleet*, Karen Sliwa*, Emma Smith*, Jennifer L Smith*, Nicolas J C Stapelberg*, Andrew Steer*, Timothy Steiner*, Wilma A Stolk*, Lars Jacob Stovner*, Christopher Sudfeld*, Sana Syed*, Giorgio Tamburlini*, Mohammad Tavakkoli*, Hugh R Taylor*, Jennifer A Taylor*, William J Taylor*, Bernadette Thomas*, W Murray Thomson*, George D Thurston*, Imad M Tleyjeh*, Marcello Tonelli*, Jeffrey A Towbin*, Thomas Truelsen*, Miltiadis K Tsilimbaris*, Clotilde Ubeda*, Eduardo A Undurraga*, Marieke J van der Werf*, Jim van Os*, Monica S Vavilala*, N Venketasubramanian*, Mengru Wang*, Wenzhi Wang*, Kerrianne Watt*, David J Weatherall*, Martin A Weinstock*, Robert Weintraub*, Marc G Weisskopf*, Myrna M Weissman*, Richard A White*, Harvey Whiteford*, Natasha Wiebe*, Steven T Wiersma*, James D Wilkinson*, Hywel C Williams*, Sean R M Williams*, Emma Witt*, Frederick Wolfe*, Anthony D Woolf*, Sarah Wulf*, Pon-Hsiu Yeh*, Anita K M Zaidi*, Zhi-Jie Zheng*, David Zonies*, Alan D Lopez†

Summary

Background Measuring disease and injury burden in populations requires a composite metric that captures both premature mortality and the prevalence and severity of ill-health. The 1990 Global Burden of Disease study proposed disability-adjusted life years (DALYs) to measure disease burden. No comprehensive update of disease burden worldwide incorporating a systematic reassessment of disease and injury-specific epidemiology has been done since the 1990 study. We aimed to calculate disease burden worldwide and for 21 regions for 1990, 2005, and 2010 with methods to enable meaningful comparisons over time.

Methods We calculated DALYs as the sum of years of life lost (YLLs) and years lived with disability (YLDs). DALYs were calculated for 291 causes, 20 age groups, both sexes, and for 187 countries, and aggregated to regional and global estimates of disease burden for three points in time with strictly comparable definitions and methods. YLLs were calculated from age-sex-country-time-specific estimates of mortality by cause, with death by standardised lost

Lancet 2012; 380: 2197–223

This online publication has been corrected. The corrected version first appeared at thelancet.com on February 22, 2013

See Comment pages 2053, 2054, 2055, 2058, 2060, 2062, and 2063

See Special Report page 2067

See Articles pages 2071, 2095, 2129, 2144, 2163, and 2224

*Authors listed alphabetically

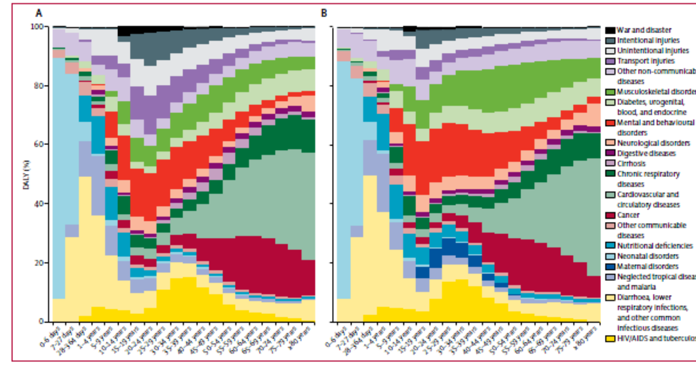


Figure 2: Percentage of global disability-adjusted life years by age, sex, and cause in 2010
Distribution of DALYs for male individuals (A) and female individuals (B). DALY=disability-adjusted life years. An interactive version of this figure is available online at http://healthmetricsandevaluation.org/gbd/visualizations/regional.

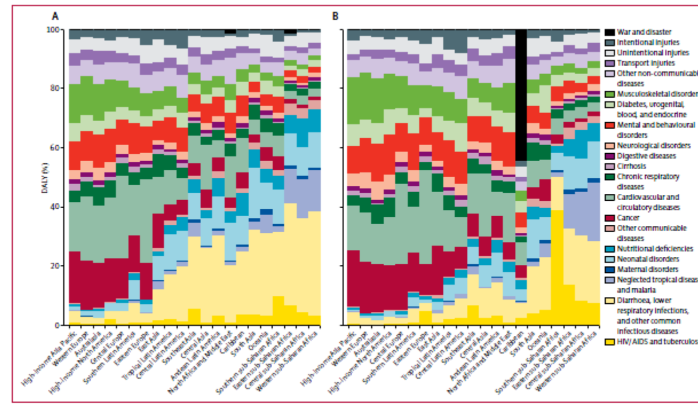
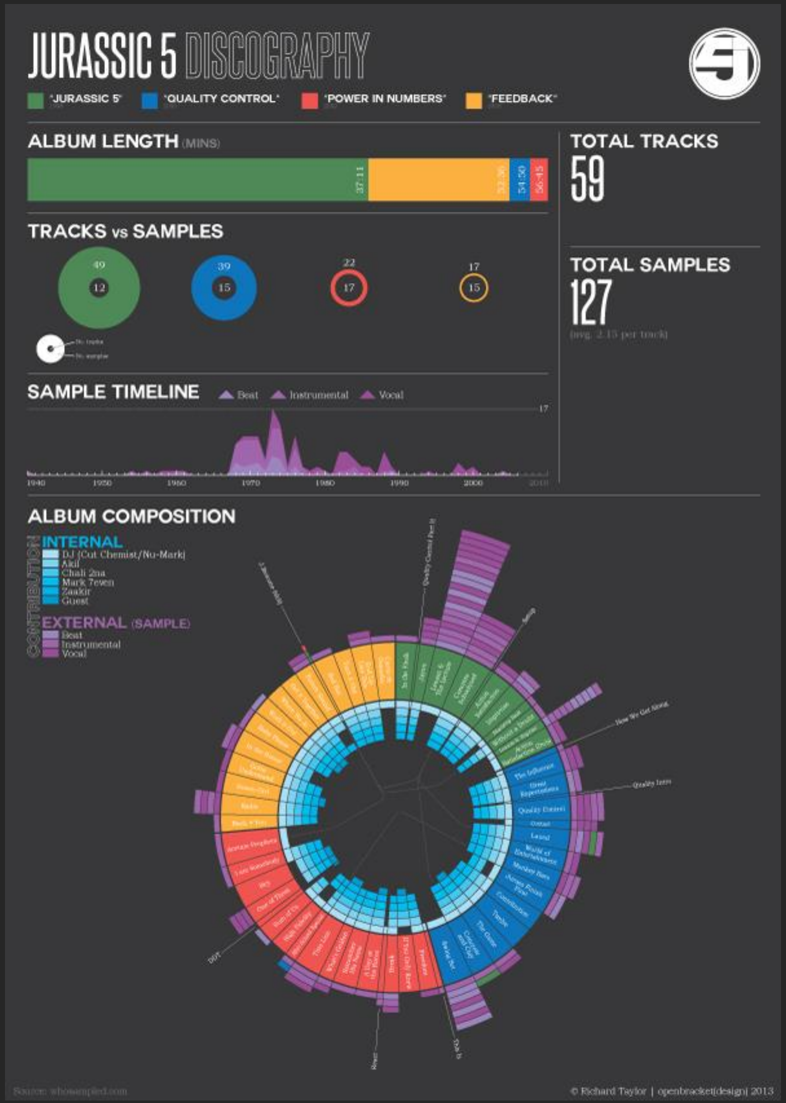
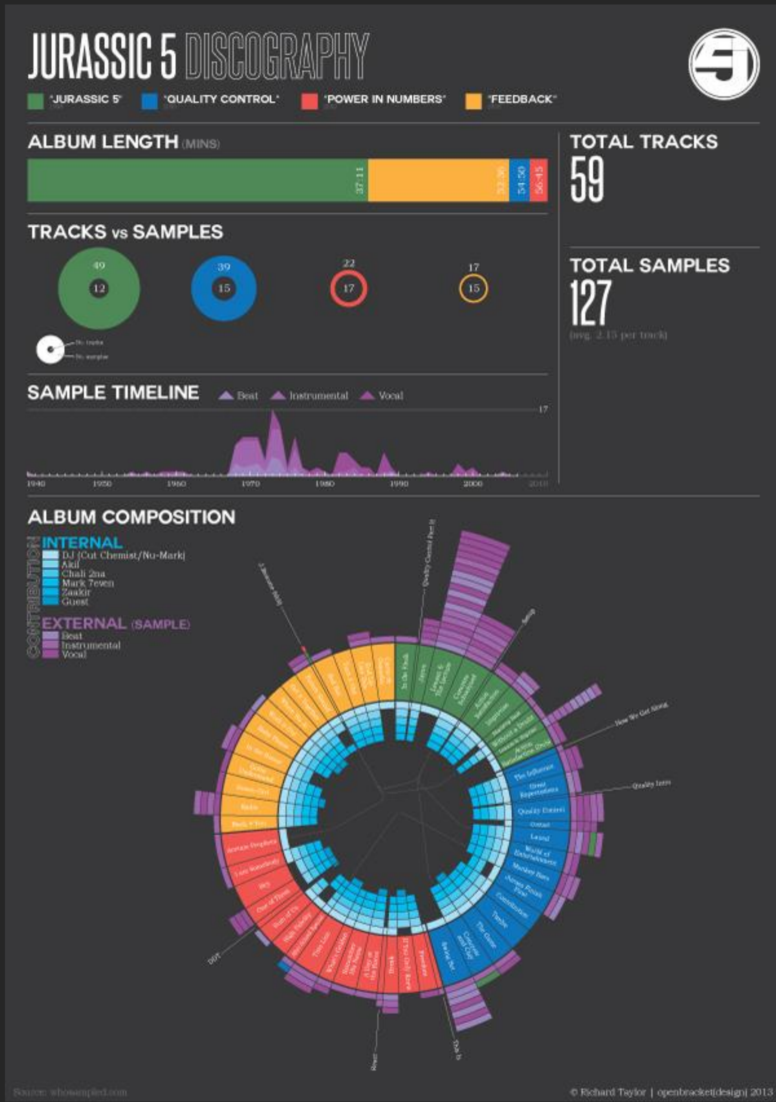


Figure 7: Percentage of disability-adjusted life years by 21 main cause groupings and region, 1990 and 2010
Proportion in 1990 (A) and 2010 (B). An interactive version of this figure is available online at http://healthmetricsandevaluation.org/gbd/visualizations/regional.

ANATOMY OF A VISUALISATION



http://openbracketdesign.co.uk/wp-content/uploads/2013/03/Jurassic-5-Data-Vis-600.jpg



http://openbracketdesign.co.uk/wp-content/uploads/2013/03/Jurassic-5-Data-Vis-600.jpg

Lines

Boxes

Squiggly lines

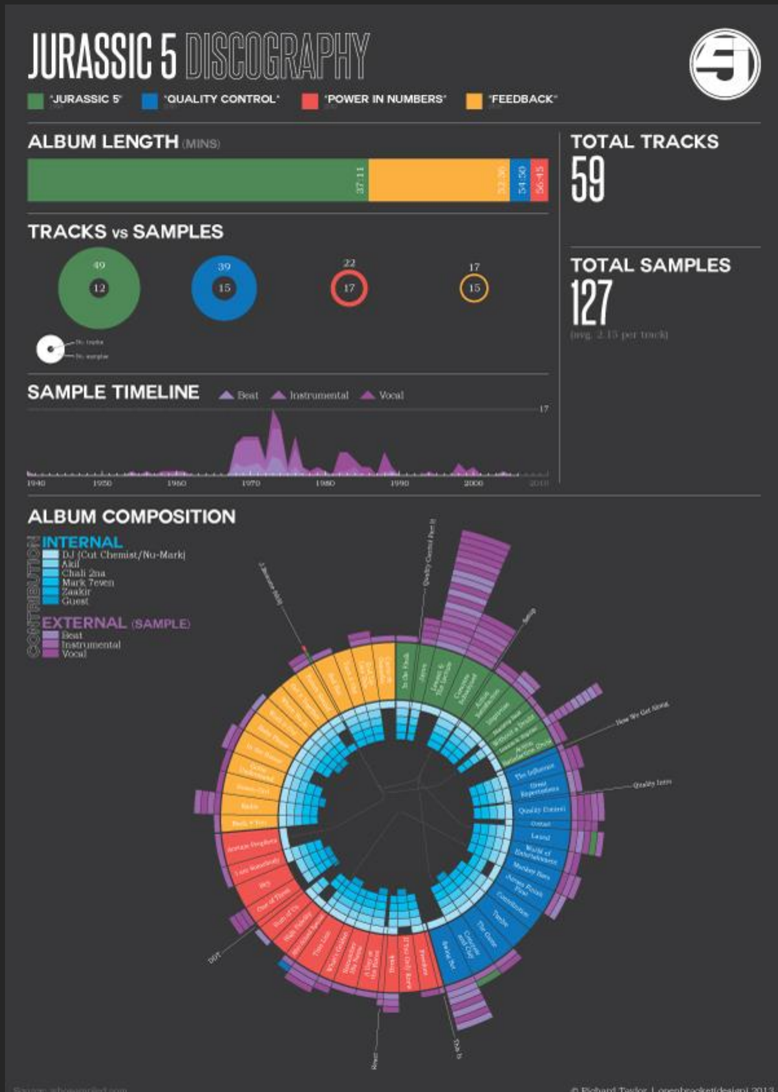
Text

Circles

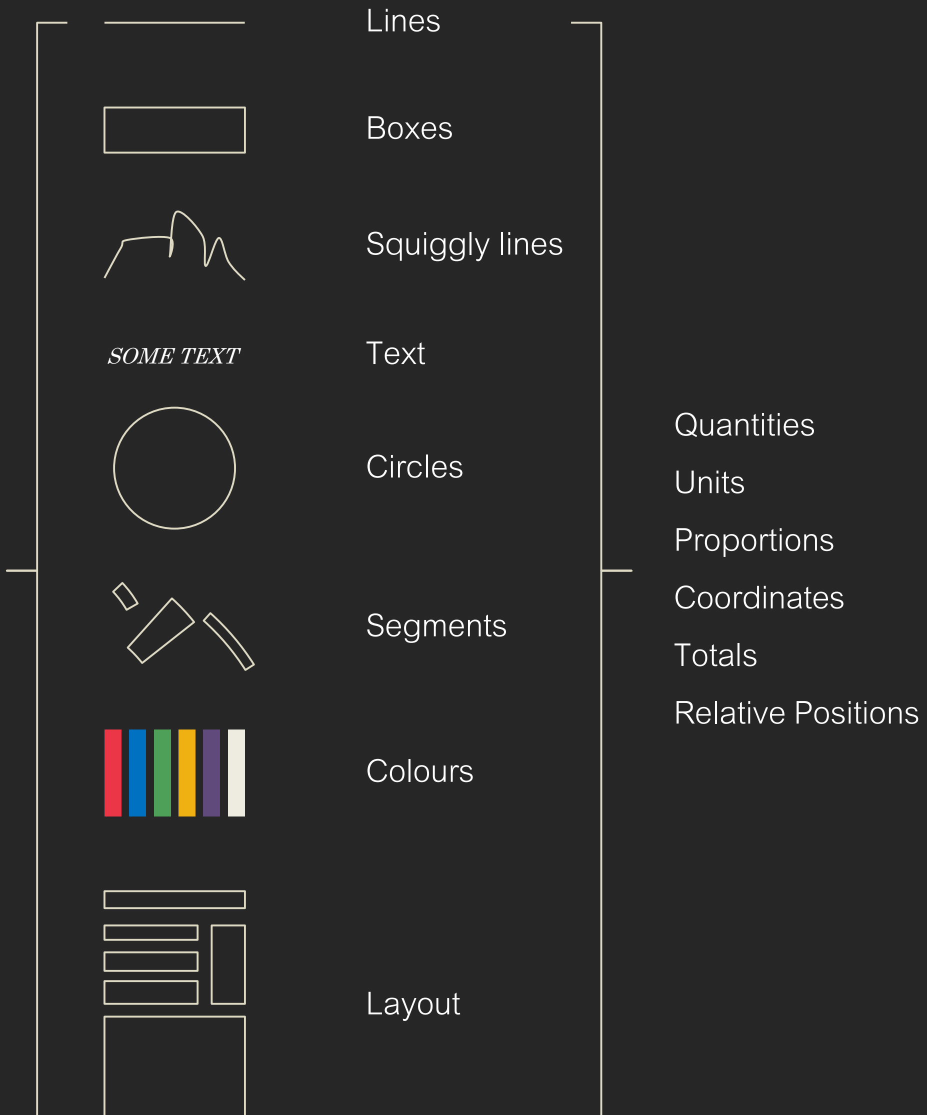
Segments

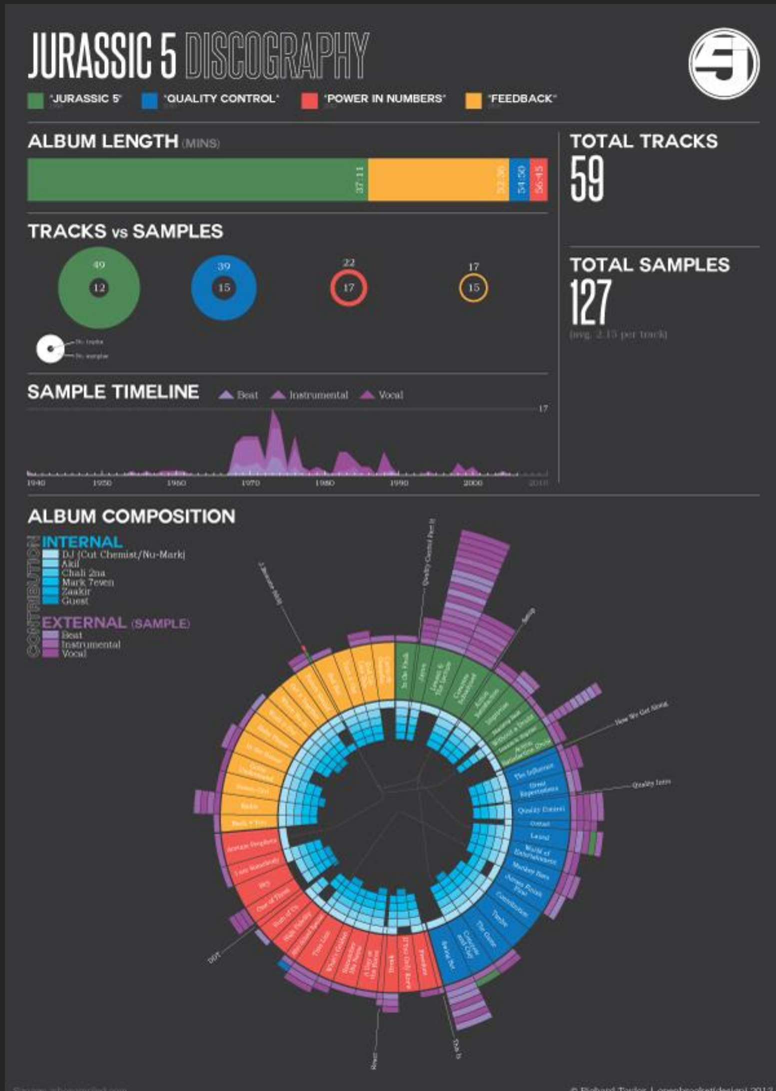
Colours

Layout



http://openbracketdesign.co.uk/wp-content/uploads/2013/03/Jurassic-5-Data-Vis-600.jpg

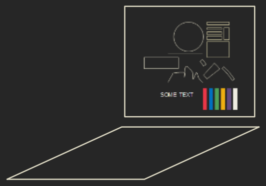




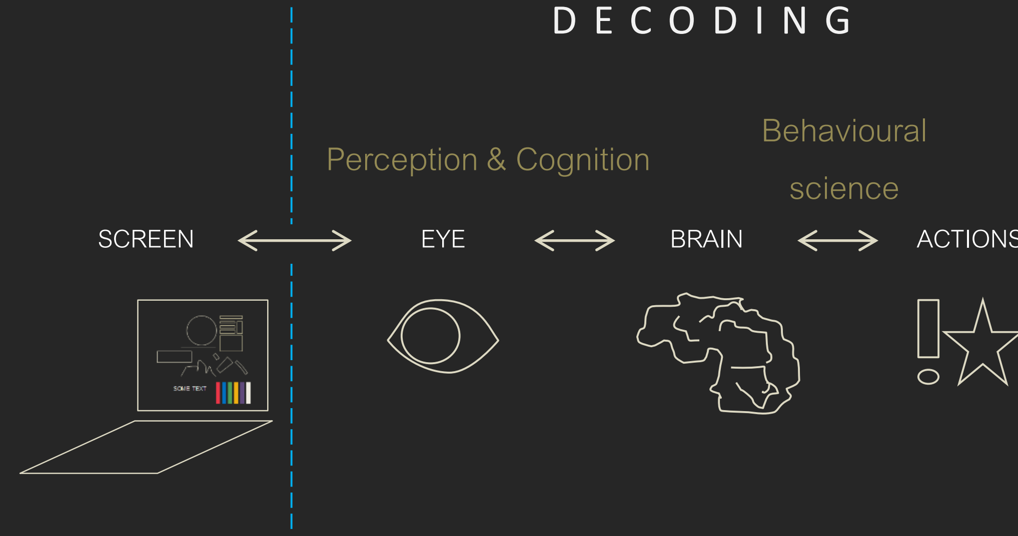
http://openbracketdesign.co.uk/wp-content/uploads/2013/03/Jurassic-5-Data-Vis-600.jpg

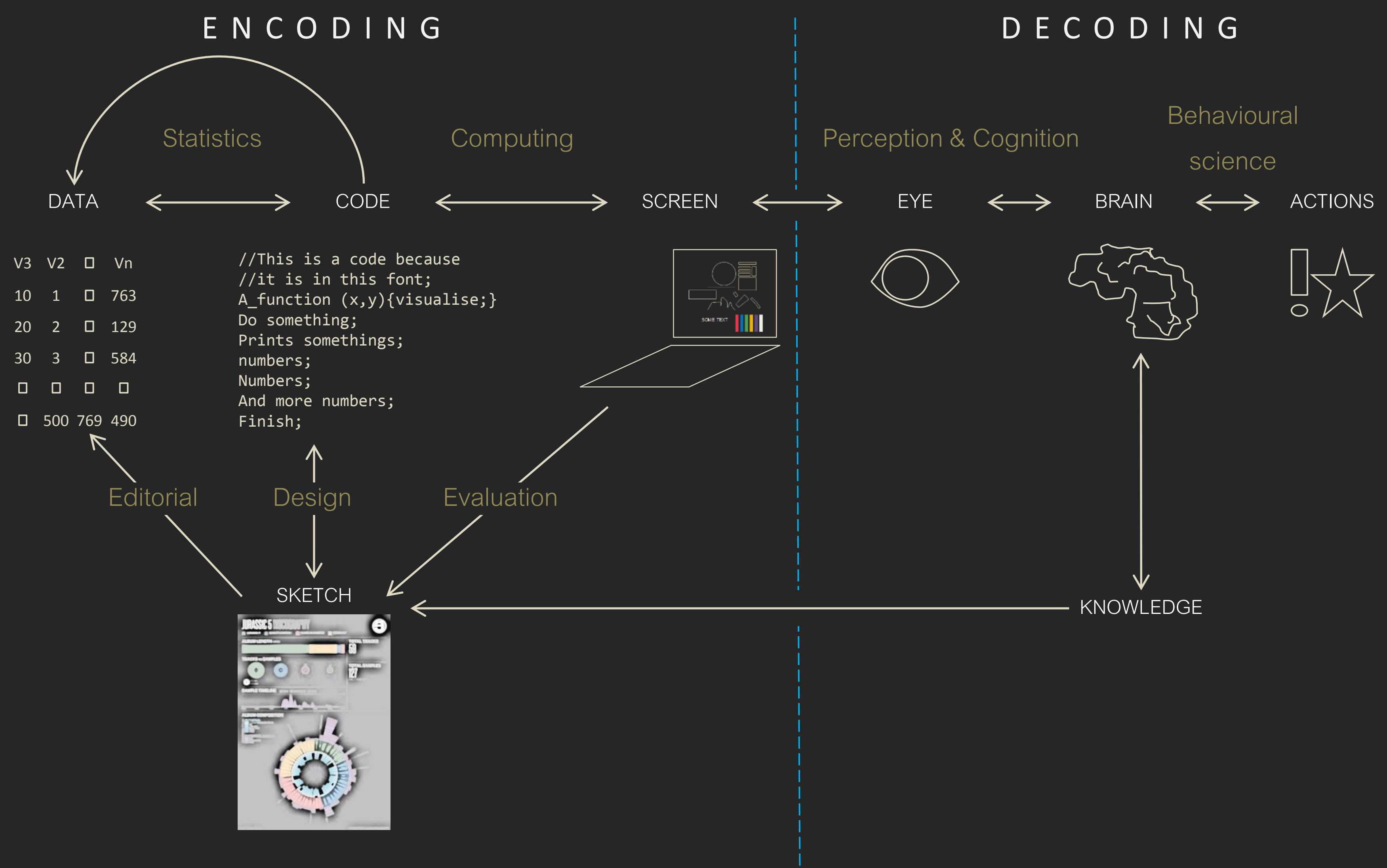
Lots of computational, editorial and design choices….

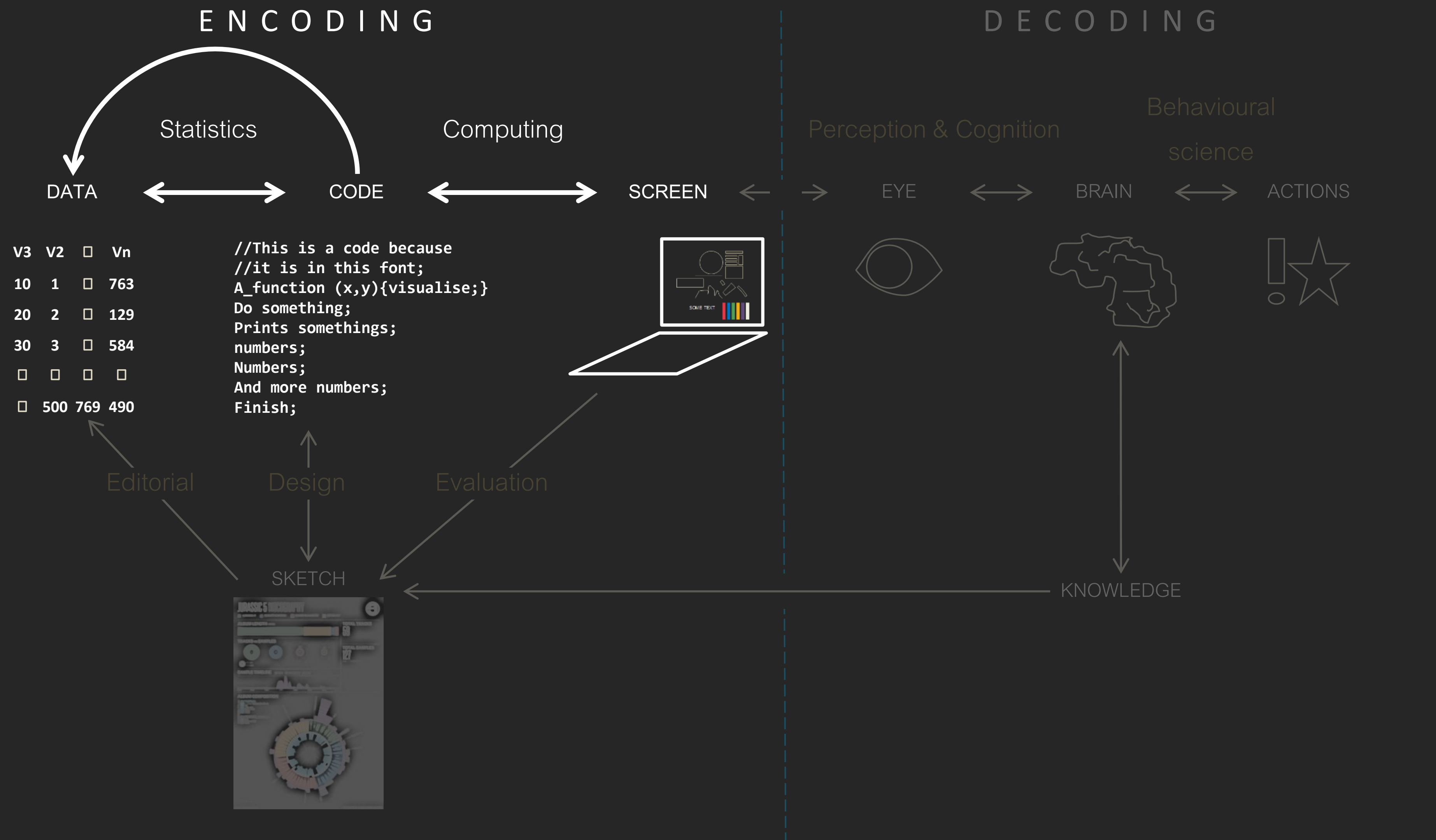
SCREEN



D E C O D I N G


Perception & Cognition
Behavioural
science
SCREEN
EYE
BRAIN
ACTIONS
SOME TEXT


ENCODING
DECODING
Statistics
Computing
Perception & Cognition
Behavioural science
DATA
CODE
SCREEN
EYE
BRAIN
ACTIONS
V3 V2 □ Vn
10 1 □ 763
20 2 □ 129
30 3 □ 584
□ □ □ □
□ 500 769 490
//This is a code because
//it is in this font;
A_function (x,y){visualise;}
Do something;
Prints somethings;
numbers;
Numbers;
And more numbers;
Finish;
SOME TEXT
Editorial
Design
Evaluation
SKETCH
KNOWLEDGE


ENCODING
DECODING
Statistics
Computing
Perception & Cognition
Behavioural science
DATA
CODE
SCREEN
EYE
BRAIN
ACTIONS
V3 V2 □ Vn
10 1 □ 763
20 2 □ 129
30 3 □ 584
□ □ □ □
□ 500 769 490
//This is a code because
//it is in this font;
A_function (x,y){visualise;}
Do something;
Prints somethings;
numbers;
Numbers;
And more numbers;
Finish;
Editorial
Design
Evaluation
SKETCH
KNOWLEDGE

Statistics and Computing

Leland Wilkinson
The Grammar of Graphics
Second Edition

This book was written for statisticians, computer scientists, geographers, researchers, and others interested in visualizing data. It presents a unique foundation for producing almost every quantitative graphic found in scientific journals, newspapers, statistical packages, and data visualization systems. While the tangible results of this work have been several visualization software libraries, this book focuses on the deep structures involved in producing quantitative graphics from data. What are the rules that underlie the production of pie charts, bar charts, scatterplots, function plots, maps, mosaics, and radar charts? Those less interested in the theoretical and mathematical foundations can still get a sense of the richness and structure of the system by examining the numerous and often unique color graphics it can produce. The second edition is almost twice the size of the original, with six new chapters and substantial revision. Much of the added material makes this book suitable for survey courses in visualization and statistical graphics.

From reviews of the first edition:

"Destined to become a landmark in statistical graphics, this book provides a formal description of graphics, particularly static graphics, playing much the same role for graphics as probability theory played for statistics." —*Journal of the American Statistical Association*

"Wilkinson's careful scholarship shows around every corner. This is a *tour de force* of the highest order." —*Psychometrika*

"All geography and map libraries should add this book to their collections; the serious scholar of quantitative data graphics will place this book on the same shelf with those by Edward Tufte, and volumes by Cleveland, Bertin, Monmonier, MacEachren, among others, and continue the unending task of proselytizing for the best in statistical data presentation by example and through scholarship like that of Leland Wilkinson." —*Cartographic Perspectives*

"In summary, this is certainly a remarkable book and a new ambitious step for the development and application of statistical graphics." —Computational Statistics and Data Analysis

About the author:

Leland Wilkinson is Senior VP, SPSS Inc. and Adjunct Professor of Statistics at Northwestern University. He is also affiliated with the Computer Science department at The University of Illinois at Chicago. He wrote the SYSTAT statistical package and founded SYSTAT Inc. in 1984. Wilkinson joined SPSS in a 1994 acquisition and now works on research and development of visual analytics and statistics. He is a Fellow of the ASA. In addition to journal articles and the original SYSTAT computer program and manuals, Wilkinson is the author (with Grant Blank and Chris Gruber) of *Desktop Data Analysis with SYSTAT*.

ISBN 0-387-24544-8
EAN
9 780387 245447 >

springeronline.com

Wilkinson The Grammar of Graphics Second Edition

Springer

Statistics and Computing

Leland Wilkinson

The Grammar of Graphics

Second Edition

Springer

Grammar makes language expressive. A language consisting of words and no grammar (statement = word) expresses only as many ideas as there are words. By specifying how words are combined in statements, a grammar expands a language's scope…

Grammar makes language expressive. A language consisting of words and no grammar (statement = word) expresses only as many ideas as there are words. By specifying how words are combined in statements, a grammar expands a language's scope…

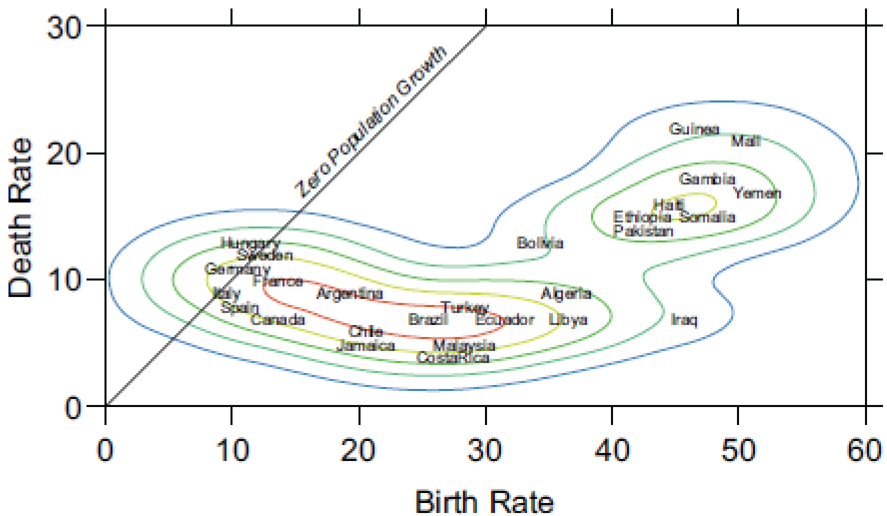


***Figure 1.1** Plot of death rates against birth rates for selected countries*

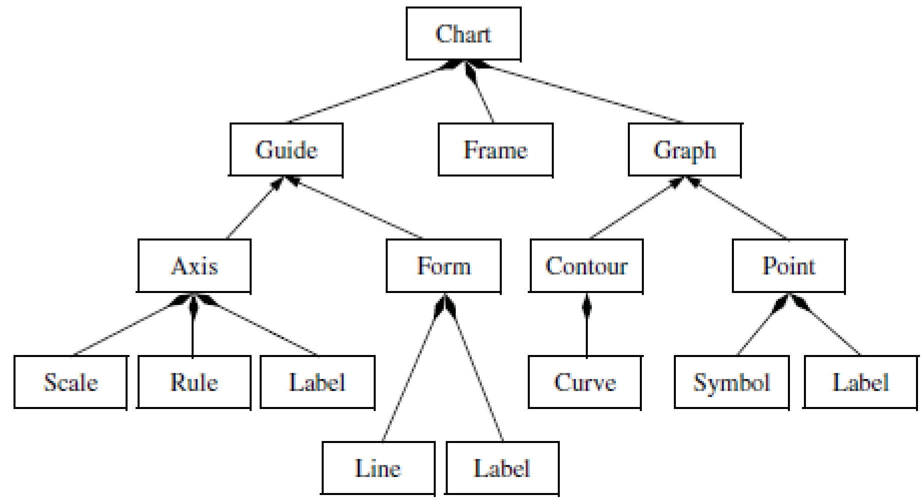
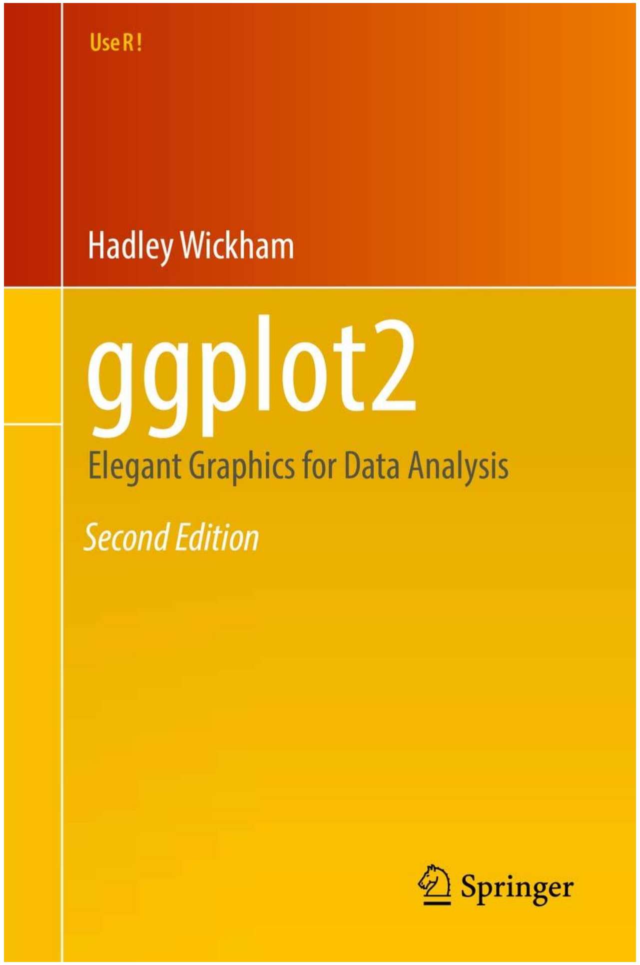
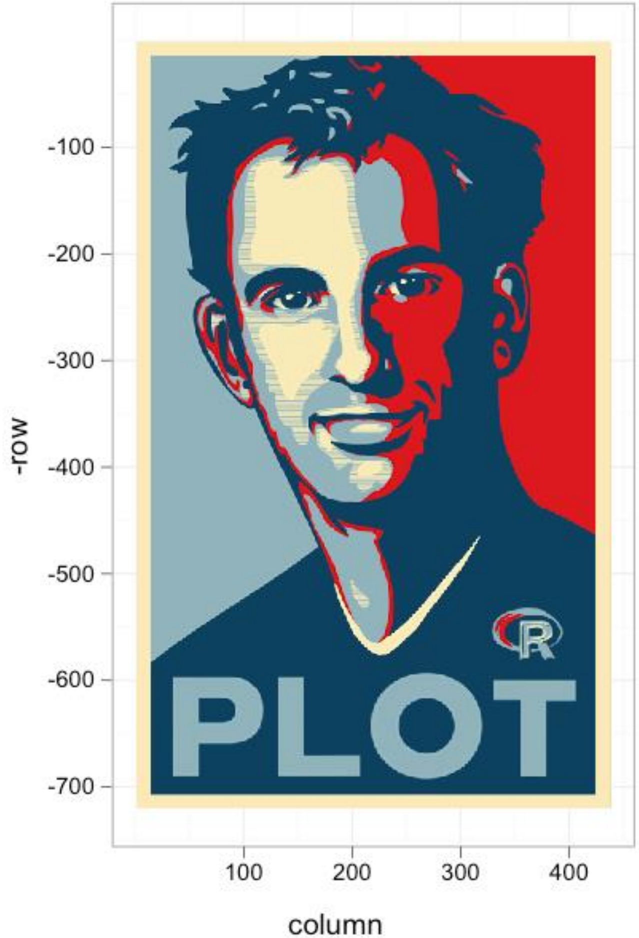


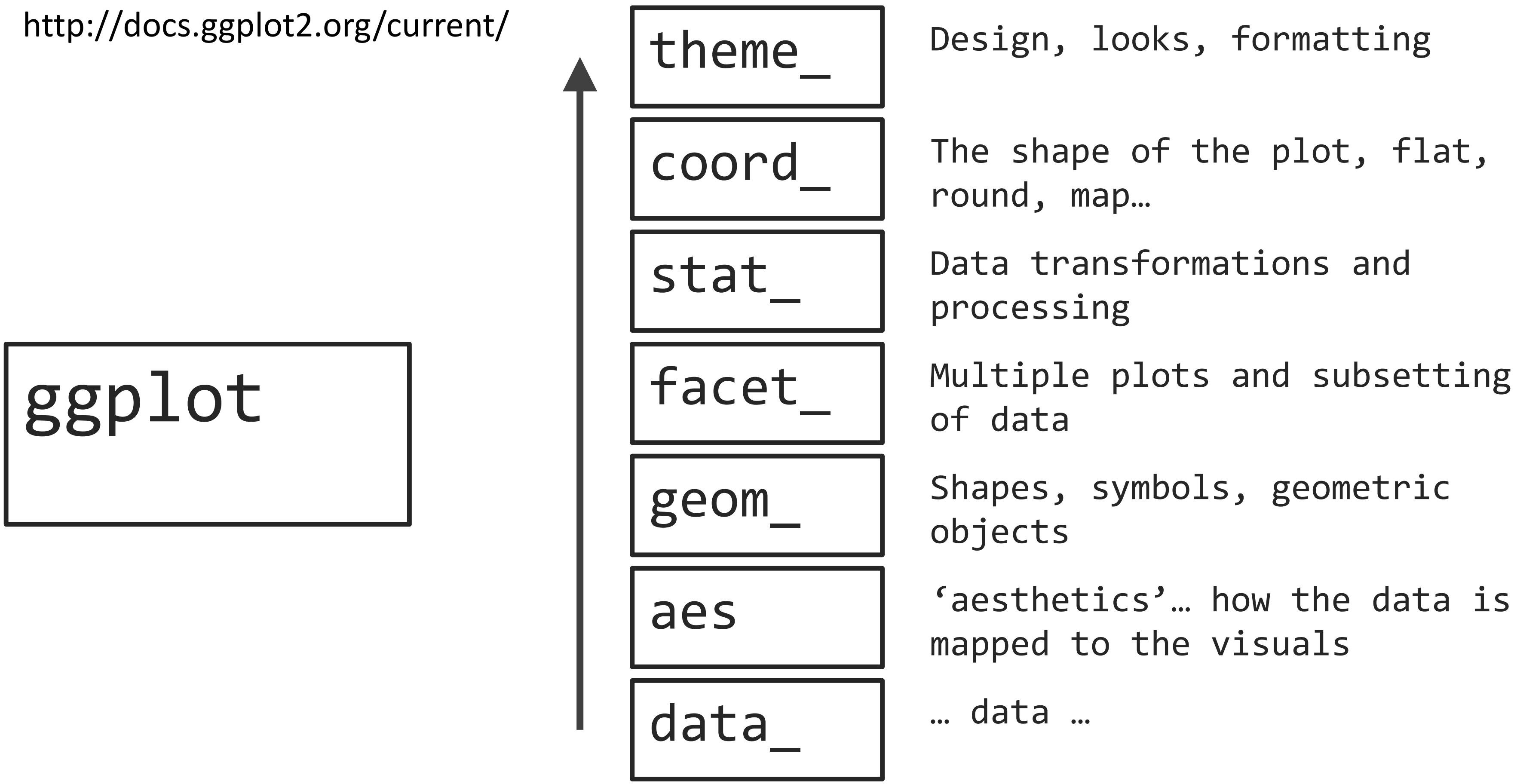
***Figure 1.2** Design tree for chart in Figure 1.1*



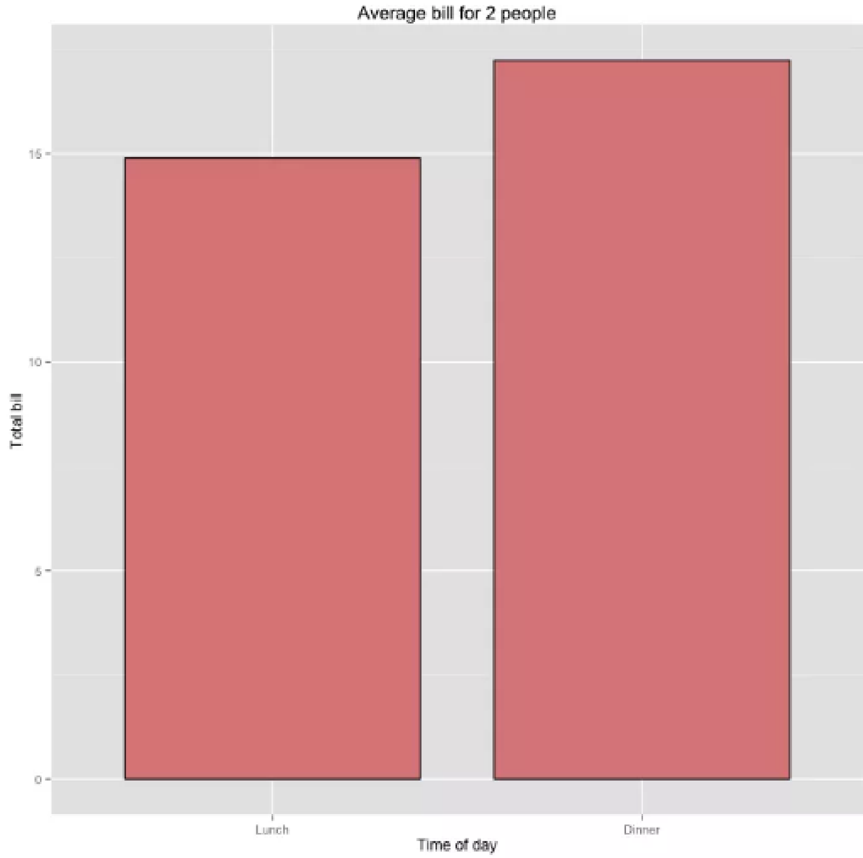


https://priceonomics.com/hadley-wickham-the-man-who-revolutionized-r/

http://docs.ggplot2.org/current/

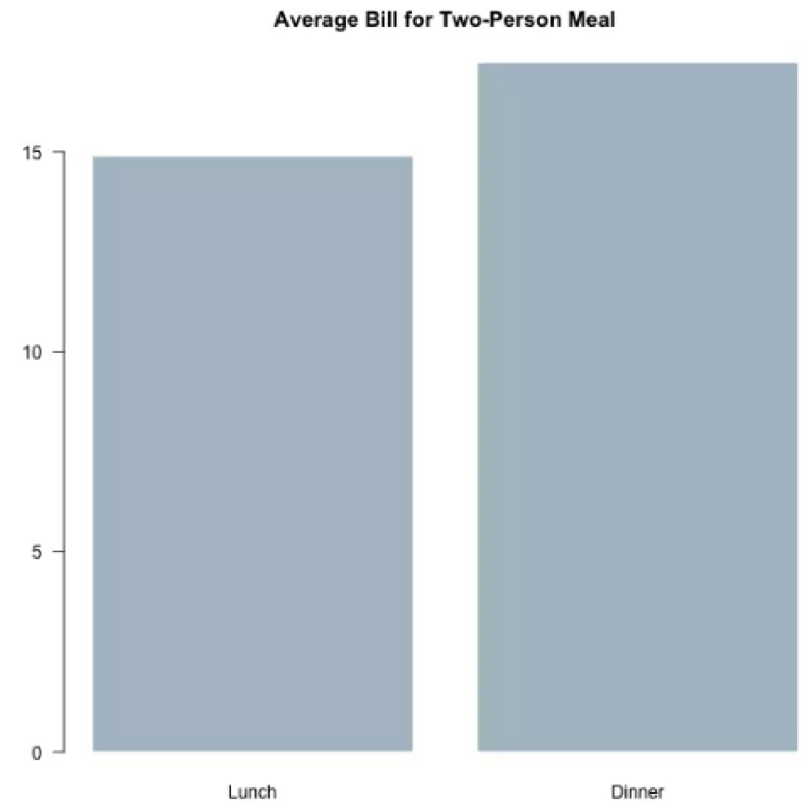


ggplot2




```
ggplot(data=dat, aes(x=time, y=total_bill, fill=time)) +
  geom_bar(colour="black", fill="#DD8888", width=.8,
stat="identity") +
  guides(fill=FALSE) +
  xlab("Time of day") + ylab("Total bill") +
  ggtitle("Average bill for 2 people")
```

Base Graphics




```
par(las=1)
barplot(dat$total_bill,
        names.arg=dat$time,
        col="#AFC0CB",
        border=FALSE,
        main="Average Bill for Two-Person Meal")
```

http://flowingdata.com/2016/03/22/comparing-ggplot2-and-r-base-graphics/

```
ggplot(data=dat, aes(x=time, y=total_bill, fill=time)) +
  geom_bar(colour="black", fill="#DD8888", width=.8,
stat="identity") +
  guides(fill=FALSE) +
  xlab("Time of day") + ylab("Total bill") +
  ggtitle("Average bill for 2 people")
```

```r
dat <- data.frame(
  time = factor(c("Lunch","Dinner"), levels=c("Lunch","Dinner")),
  total_bill = c(14.89, 17.23)
)
```

```r
ggplot(data=dat, aes(x=time, y=total_bill, fill=time)) +
  geom_bar(colour="black", fill="#DD8888", width=.8, stat="identity") +
  guides(fill=FALSE) +
  xlab("Time of day") + ylab("Total bill") +
  ggtitle("Average bill for 2 people") +
  facet_wrap(~time, ncol=2)
```

```r
ggplot(data=dat, aes(x=time, y=total_bill, fill=time)) +
  geom_bar(colour="black", fill="#DD8888", width=.8,
stat="identity") +
  guides(fill=FALSE) +
  xlab("Time of day") + ylab("Total bill") +
  ggtitle("Average bill for 2 people")

par(las=1)
barplot(dat$total_bill,
        names.arg=dat$time,
        col="#AFC0CB",
        border=FALSE,
        main="Average Bill for Two-Person Meal")
```

1.5.4 Not a Book of Virtues

This system is capable of producing some hideous graphics. There is nothing in its design to prevent its misuse. We will occasionally point out some of these instances (*e.g.*, Figure 9.25). That the system *can* produce such graphics is simply a consequence of its basis on the mathematical rules that determine the meaning of graphs, rather than on the *ad hoc* rules we sometimes use to produce graphics. These rules are not based on personal preferences but rather on the mathematics and perceptual dimensions underlying the graphics we draw in practice. These rules are just as capable of producing graphics for *USA Today* as for *Scientific American*.

Today's Lab

In this lab you will be introduced to the three main ways of creating graphic in R – using 'base' graphics, the 'ggplot' package and 'grid' graphics. There are six scripts which have the instructions, directions and questions as comments. You will not complete them all! The goals are:

- to orientate you with the structure of the different methods (a broad, but shallow overview),
- encourage you to defy the defaults and show you how to make visuals your own,
- introduce some design and computational thinking,
- and give you a spring board to become an independent learner.

Remember (as I should have said in the lecture) not all the visuals we will produce make sense! Some examples are just show you alternatives, or signposts things you may consider later on. Except for the jpeg, all the data is contained in the scripts. Do not linger too long on looking at the data. That is what the visualisations are for. Please use the scripts in this order…

1. anscombe.R
2. anscombelayouts.R
3. truncated.R
4. piecharts.R
5. anscombeGGplot.R
6. ukko.R (also using ukko5.jpg, download this and save it)

Don't feel like you have to learn every command, every argument and every method. We all look everything up all the time. The key is to know enough that you can articulate your question. Most questions are already answered on the internet.

QUESTIONS?

CENTRE FOR INTERDISCIPLINARY METHODOLOGIES

This book does *not* contain discussions about which sort of plot is most appropriate for a particular sort of data, nor does it contain guidelines for correct graphical presentation. In fact, instructions are provided for producing types of plots that are generally disapproved of…

Paul Murrell (2006) R graphics. Chapman & Hall.

correlation

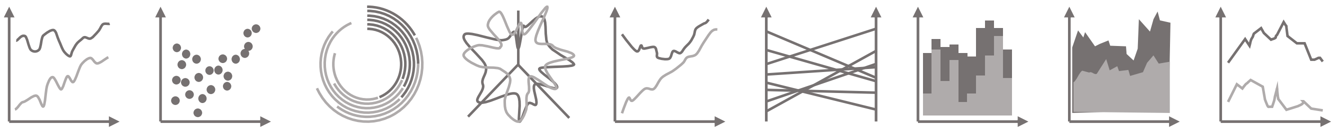
Ranking Visualizations of Correlation Using Weber's Law

Lane Harrison, Fumeng Yang, Steven Franconeri, Remco Chang

Abstract— Despite years of research yielding systems and guidelines to aid visualization design, practitioners still face the challenge of identifying the best visualization for a given dataset and task. One promising approach to circumvent this problem is to leverage perceptual laws to quantitatively evaluate the effectiveness of a visualization design. Following previously established methodologies, we conduct a large scale (n=1687) crowdsourced experiment to investigate whether the perception of correlation in nine commonly used visualizations can be modeled using Weber's law. The results of this experiment contribute to our understanding of information visualization by establishing that: 1) for all tested visualizations, the precision of correlation judgment could be modeled by Weber's law, 2) correlation judgment precision showed striking variation between negatively and positively correlated data, and 3) Weber models provide a concise means to quantify, compare, and rank the perceptual precision afforded by a visualization.

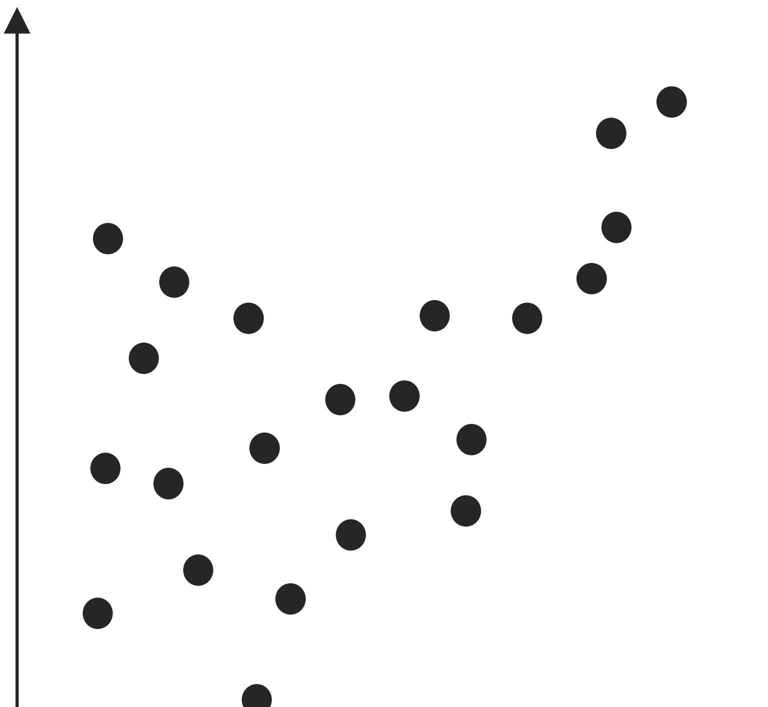
Index Terms—Perception, Visualization, Evaluation

Which of these graphs is best for noticing correlations between variables?

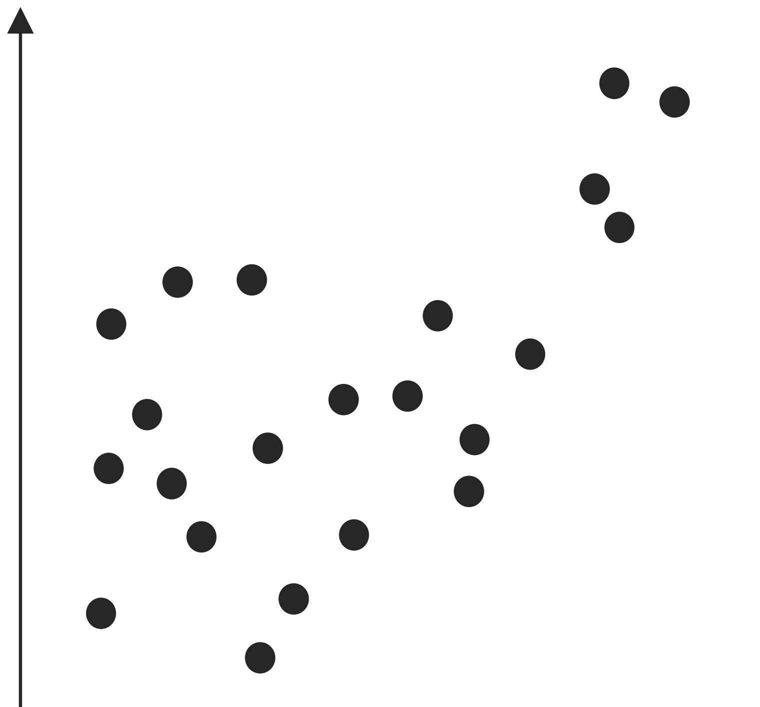


Which has the higher level of correlation?

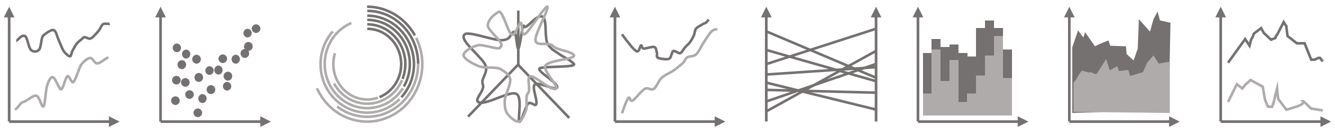
A



B

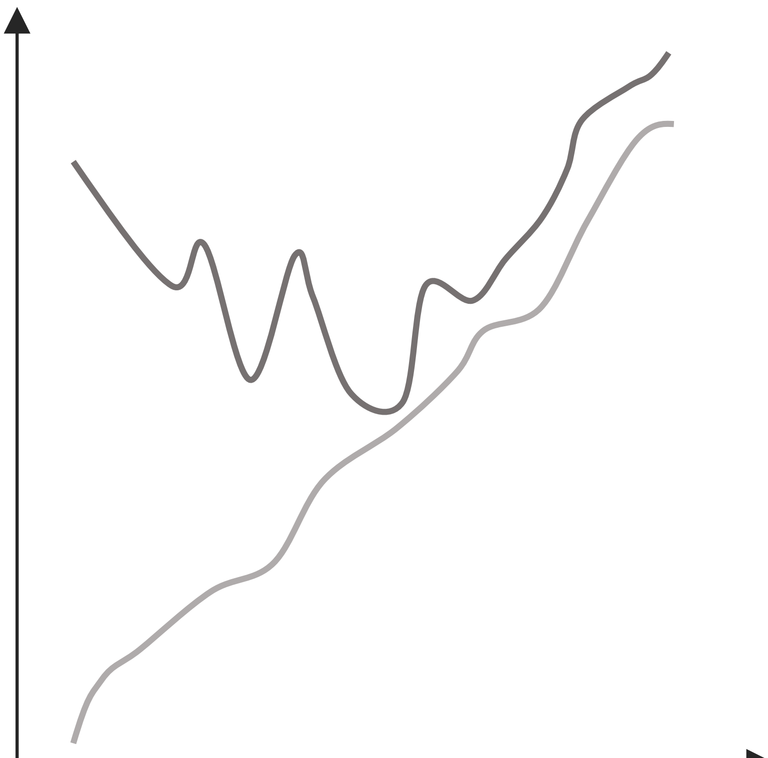


Which of these graphs is best for noticing correlations between variables?

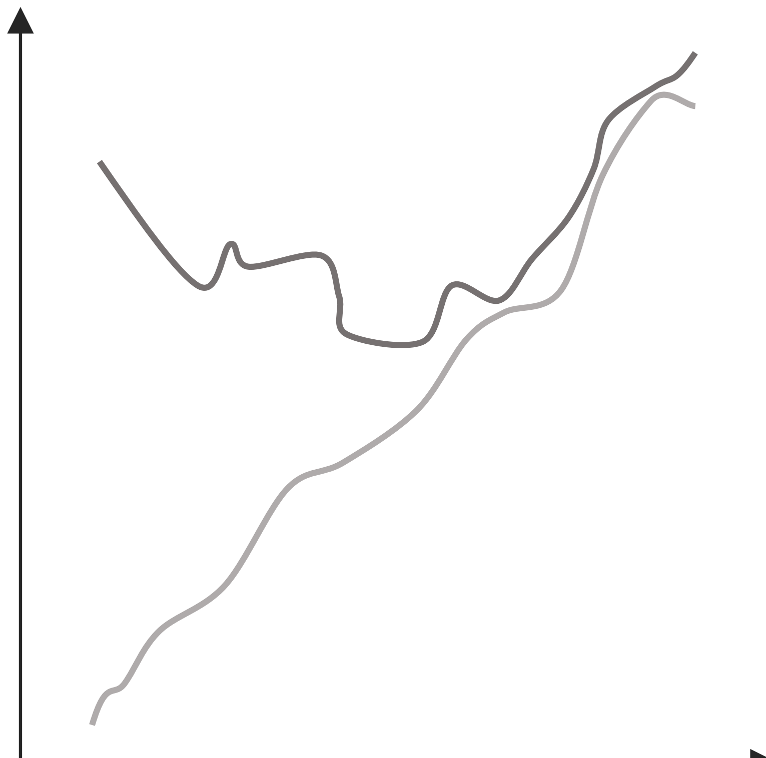


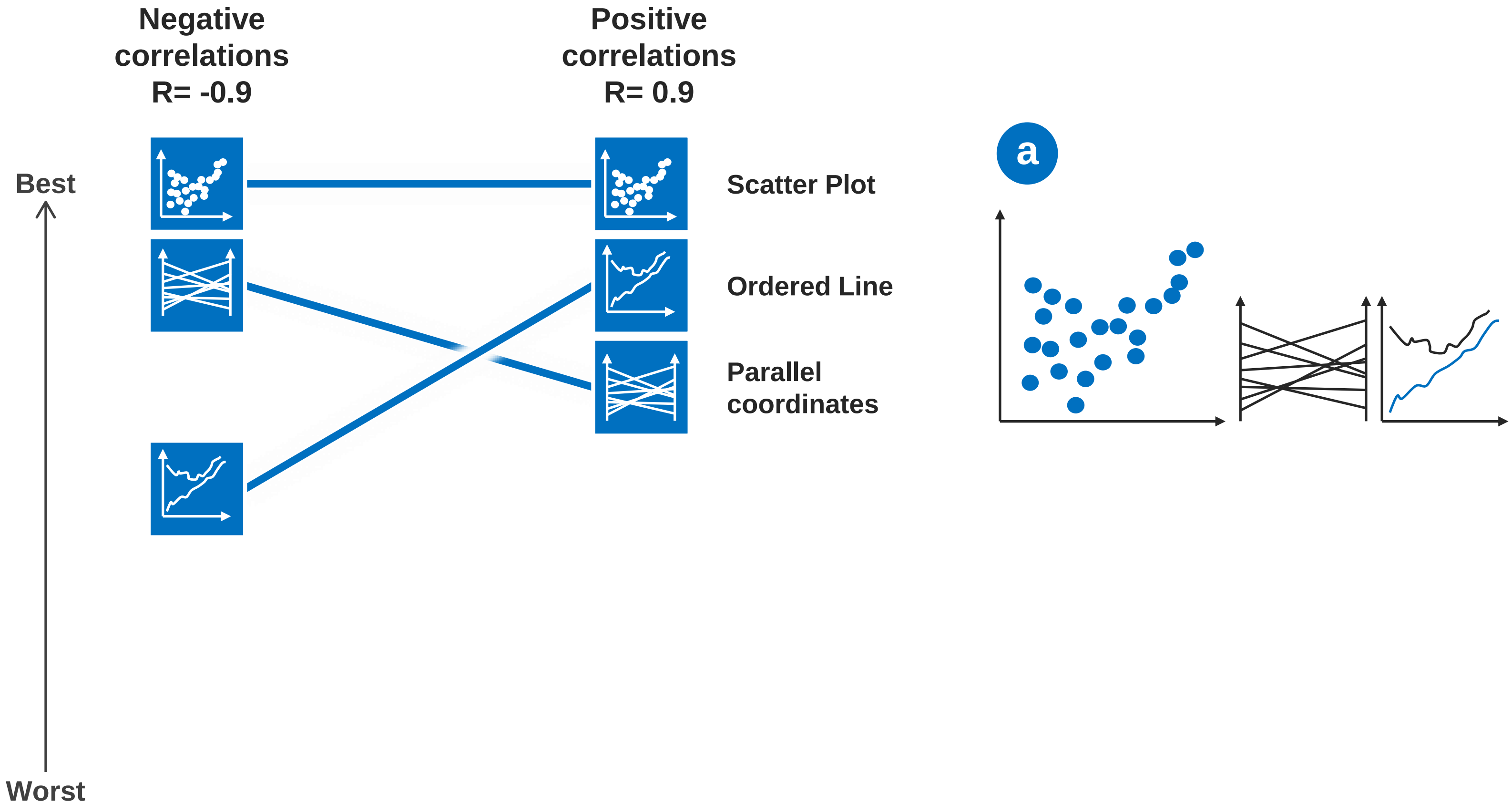
Which has the higher level of correlation?

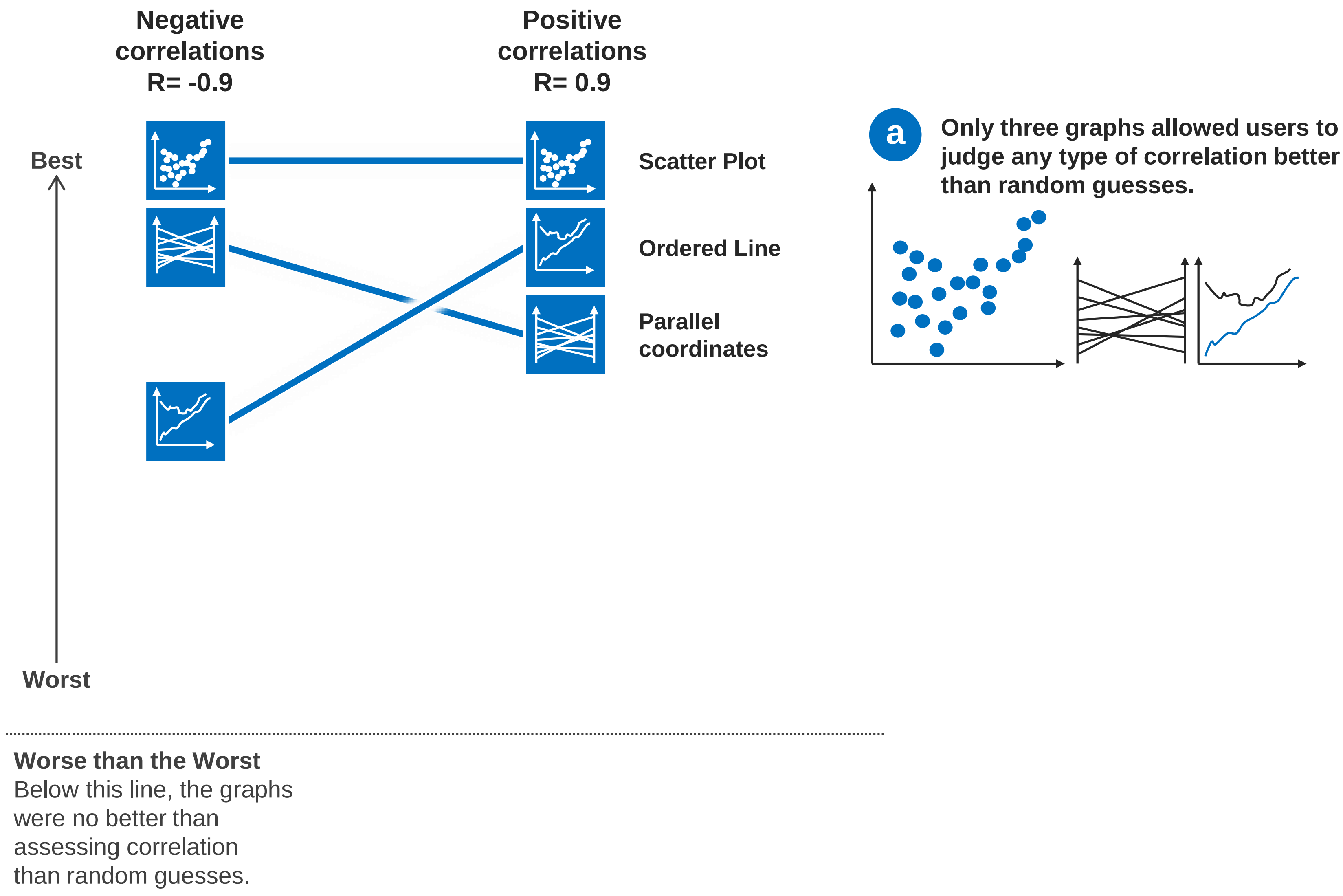
A

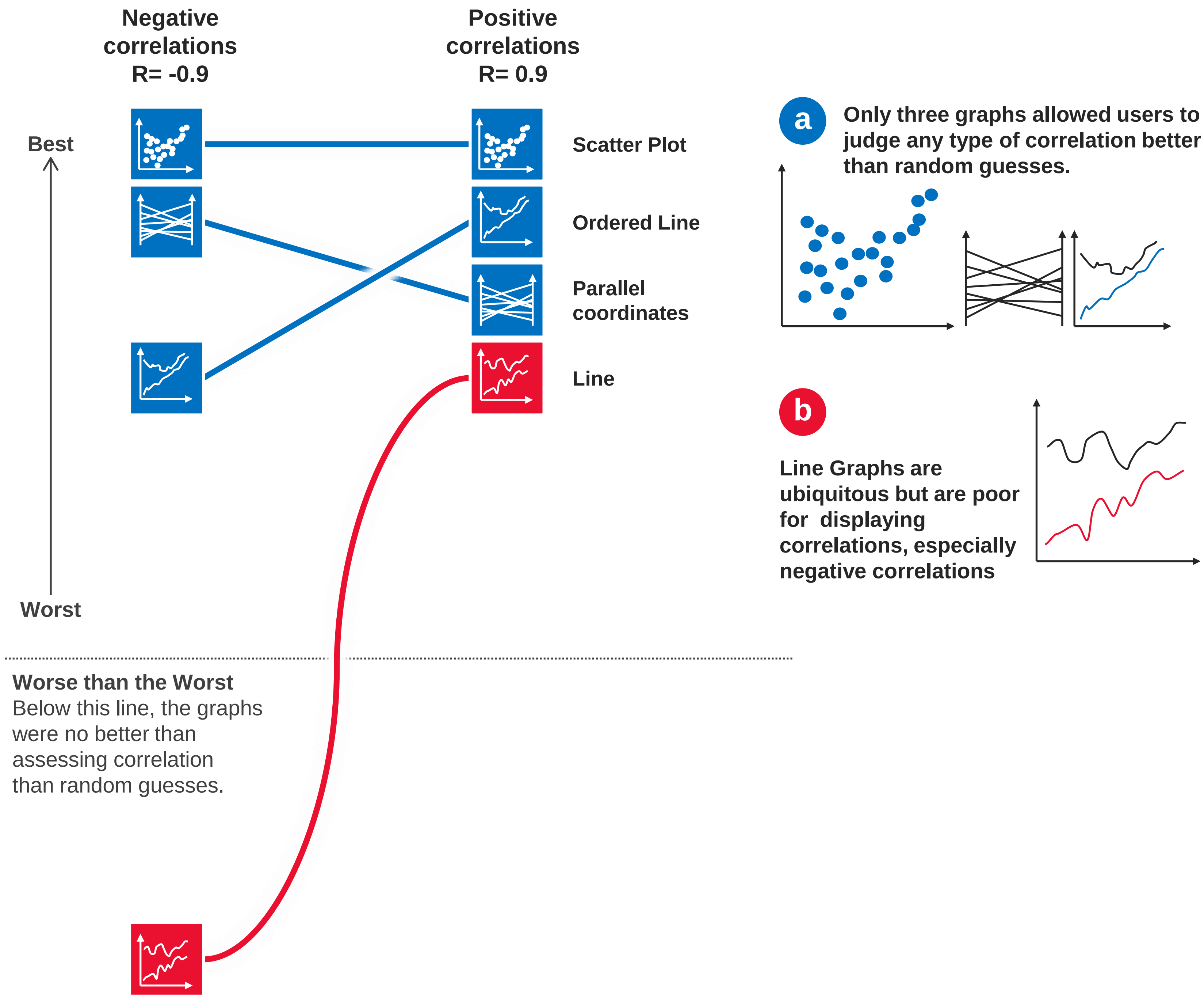


B

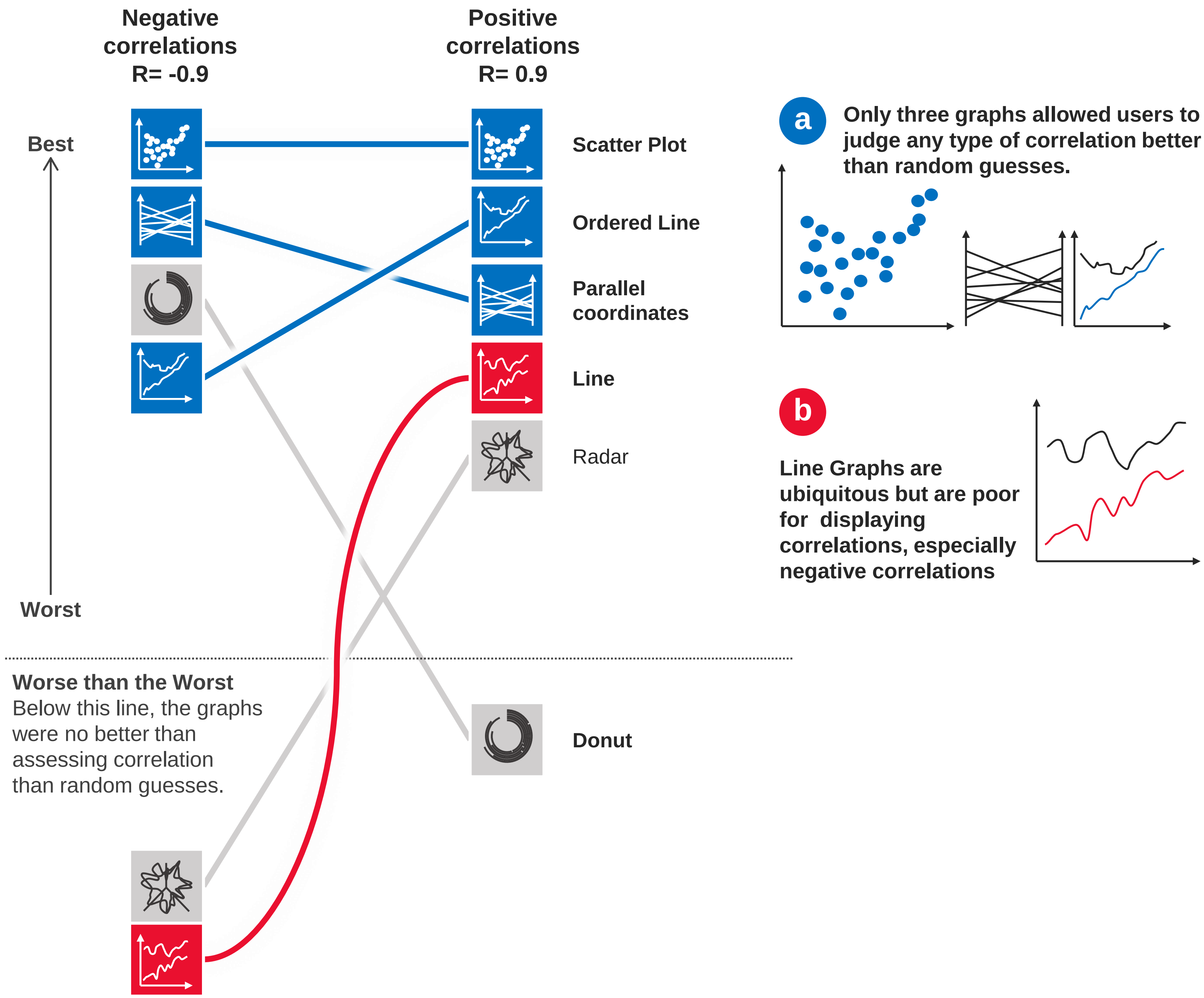


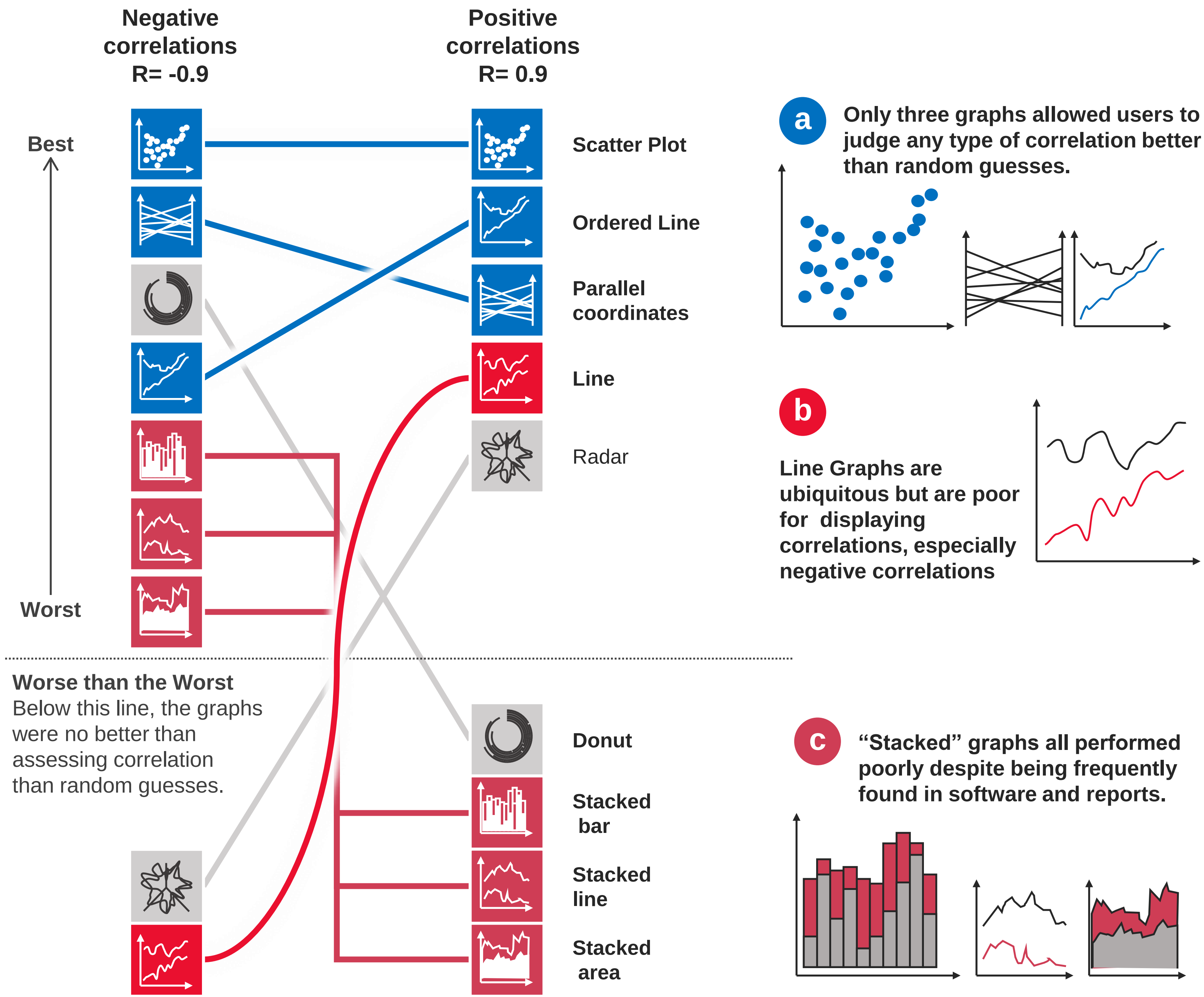

Negative correlations R= -0.9
Positive correlations R= 0.9
Best
Worst
Scatter Plot
Ordered Line
Parallel coordinates
a


Negative correlations R= -0.9
Positive correlations R= 0.9
Best
Worst
Scatter Plot
Ordered Line
Parallel coordinates
a
Only three graphs allowed users to judge any type of correlation better than random guesses.
Worse than the Worst
Below this line, the graphs were no better than assessing correlation than random guesses.

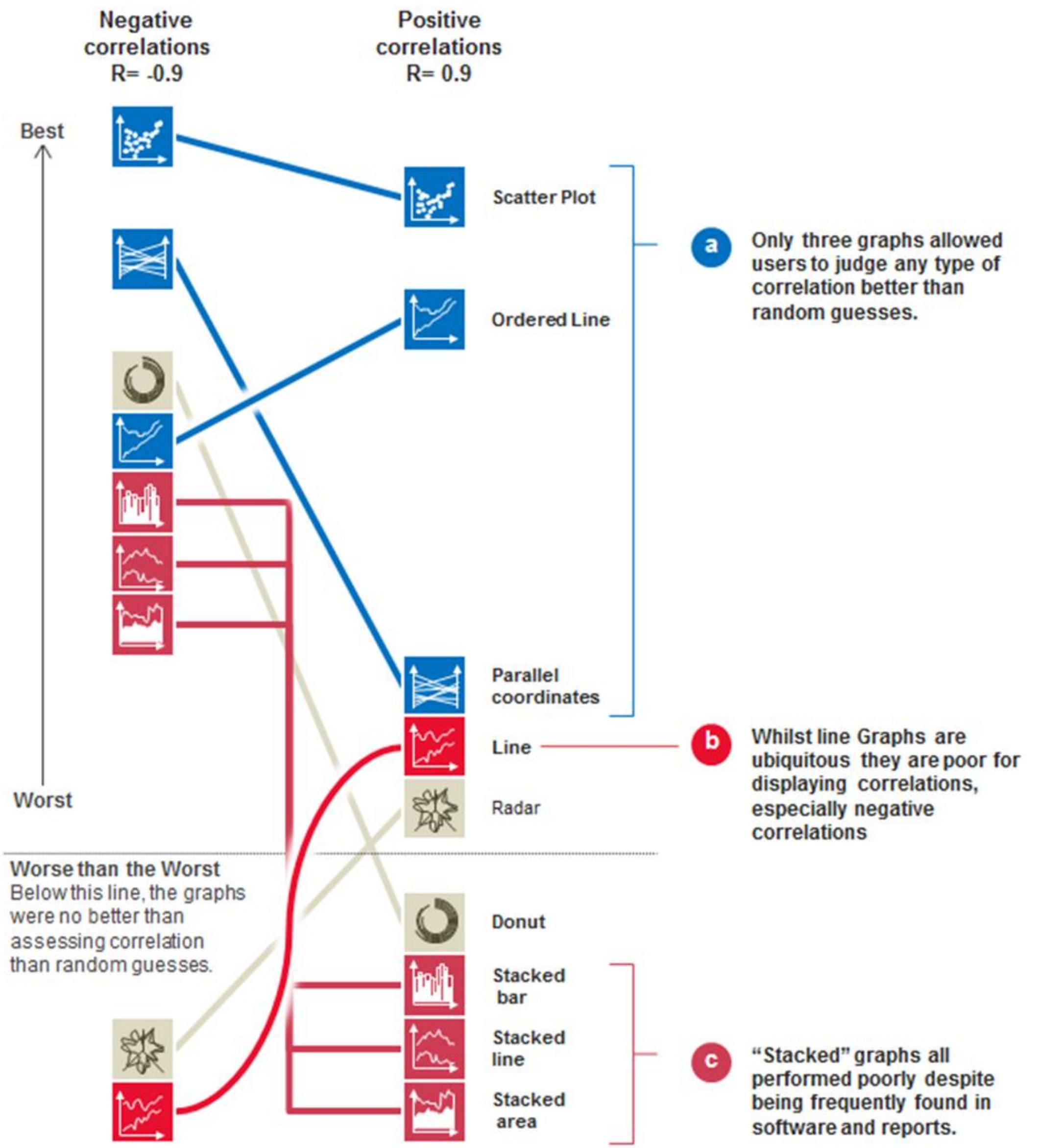


Negative correlations R= -0.9
Positive correlations R= 0.9
Best
Worst
Scatter Plot
Ordered Line
Parallel coordinates
Line
a
Only three graphs allowed users to judge any type of correlation better than random guesses.
b
Line Graphs are ubiquitous but are poor for displaying correlations, especially negative correlations
Worse than the Worst
Below this line, the graphs were no better than assessing correlation than random guesses.

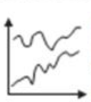

Negative correlations R= -0.9
Positive correlations R= 0.9
Best
Worst
Scatter Plot
Ordered Line
Parallel coordinates
Line
Radar
Donut
Worse than the Worst
Below this line, the graphs were no better than assessing correlation than random guesses.
a
Only three graphs allowed users to judge any type of correlation better than random guesses.
b
Line Graphs are ubiquitous but are poor for displaying correlations, especially negative correlations


Negative correlations R= -0.9
Positive correlations R= 0.9
Best
Worst
Scatter Plot
Ordered Line
Parallel coordinates
Line
Radar
Worse than the Worst
Below this line, the graphs were no better than assessing correlation than random guesses.
Donut
Stacked bar
Stacked line
Stacked area
a
Only three graphs allowed users to judge any type of correlation better than random guesses.
b
Line Graphs are ubiquitous but are poor for displaying correlations, especially negative correlations
c
"Stacked" graphs all performed poorly despite being frequently found in software and reports.

Which graph is best for noticing differences in correlation?

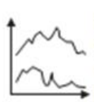


What is an ordered line graph?

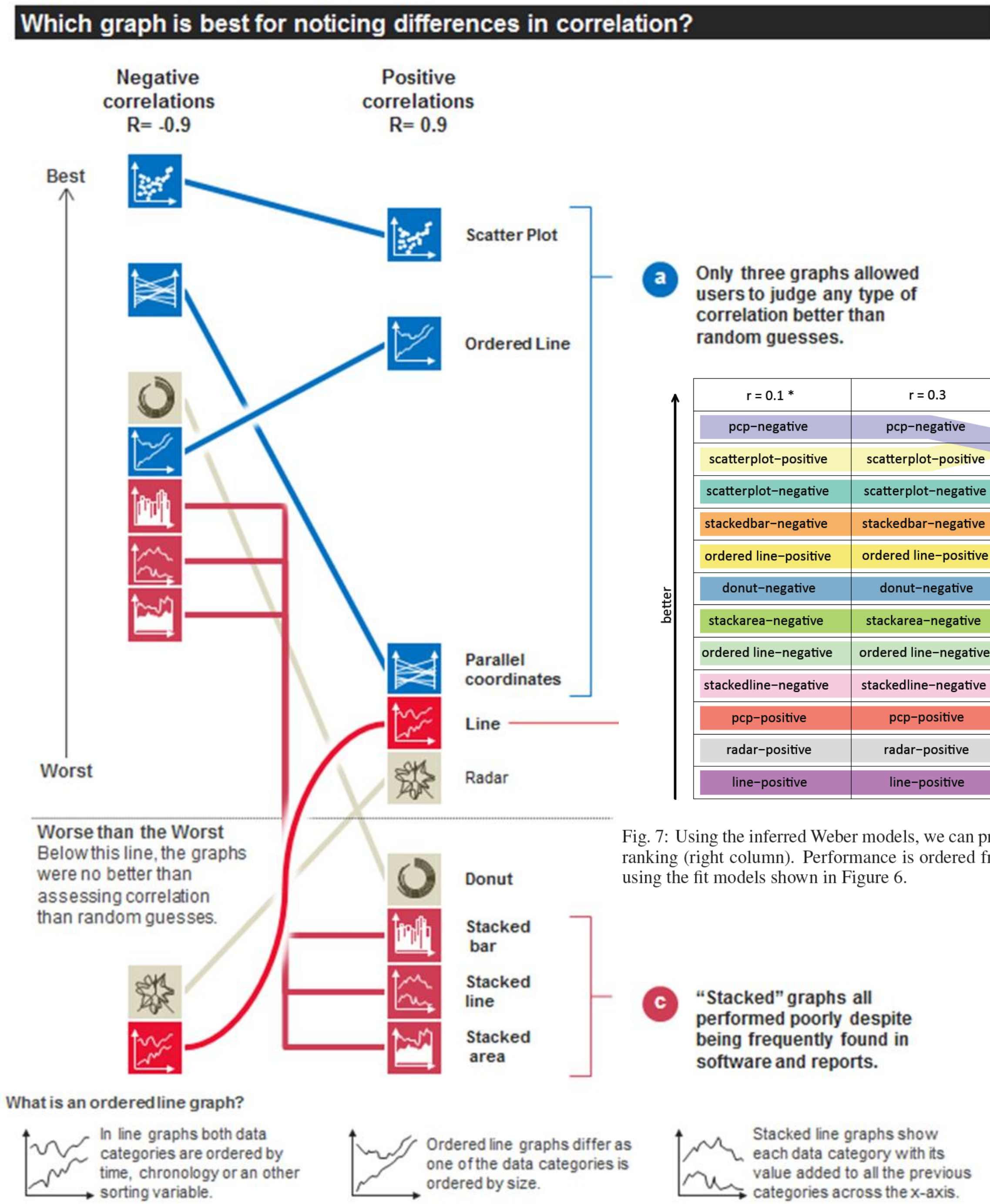


In line graphs both data categories are ordered by time, chronology or an other sorting variable.

Ordered line graphs differ as one of the data categories is ordered by size.



Stacked line graphs show each data category with its value added to all the previous categories across the x-axis.

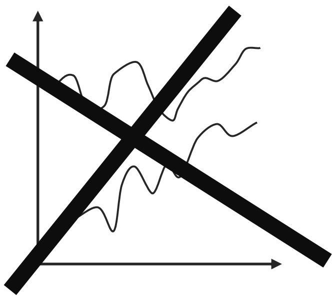


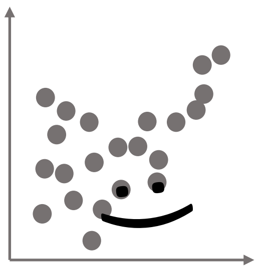
r = 0.1 *	r = 0.3	r = 0.5	r = 0.7	r = 0.9 *	overall
pcp–negative	pcp–negative	scatterplot–positive	scatterplot–negative	scatterplot–negative	scatterplot–positive
scatterplot–positive	scatterplot–positive	pcp–negative	scatterplot–positive	scatterplot–positive	pcp–negative
scatterplot–negative	scatterplot–negative	scatterplot–negative	pcp–negative	pcp–negative	scatterplot–negative
stackedbar–negative	stackedbar–negative	stackedbar–negative	stackedbar–negative	ordered line–positive	stackedbar–negative
ordered line–positive	ordered line–positive	ordered line–positive	ordered line–positive	donut–negative	ordered line–positive
donut–negative	donut–negative	donut–negative	donut–negative	ordered line–negative	donut–negative
stackarea–negative	stackarea–negative	stackarea–negative	ordered line–negative	stackedbar–negative	stackarea–negative
ordered line–negative	ordered line–negative	ordered line–negative	stackarea–negative	stackedline–negative	ordered line–negative
stackedline–negative	stackedline–negative	stackedline–negative	stackedline–negative	stackarea–negative	stackedline–negative
pcp–positive	pcp–positive	pcp–positive	pcp–positive	radar–positive	pcp–positive
radar–positive	radar–positive	radar–positive	radar–positive	pcp–positive	radar–positive
line–positive	line–positive	line–positive	line–positive	line–positive	line–positive

better

Fig. 7: Using the inferred Weber models, we can produce a perceptually-driven ranking for individual correlation (r) values, as well as an overall ranking (right column). Performance is ordered from the best (top) to the worst (bottom). The columns denoted by * are predicted responses using the fit models shown in Figure 6.

- Line graphs are surprisingly ineffective!
- Scatter plots are simple but precise.





- Design for the task(s) and the data (small multiples?)