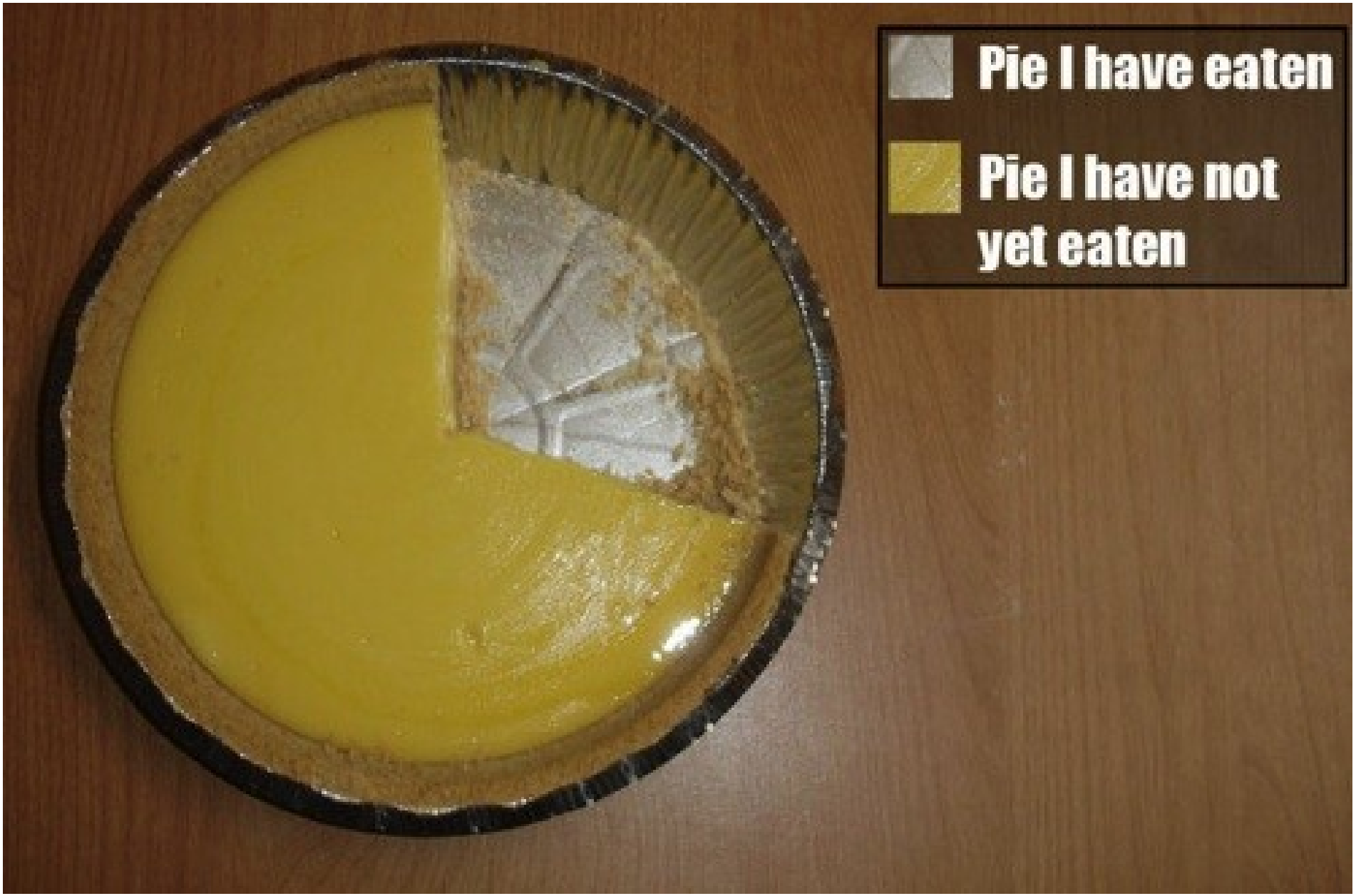
Stop Making Pie Charts!

An opinionated guide to the craft of data visualisation

Robin Gower
Data Visualisation Berlin
29.03.16

infonomics.ltd.uk
@robsteranium


Pie I have eaten
Pie I have not yet eaten

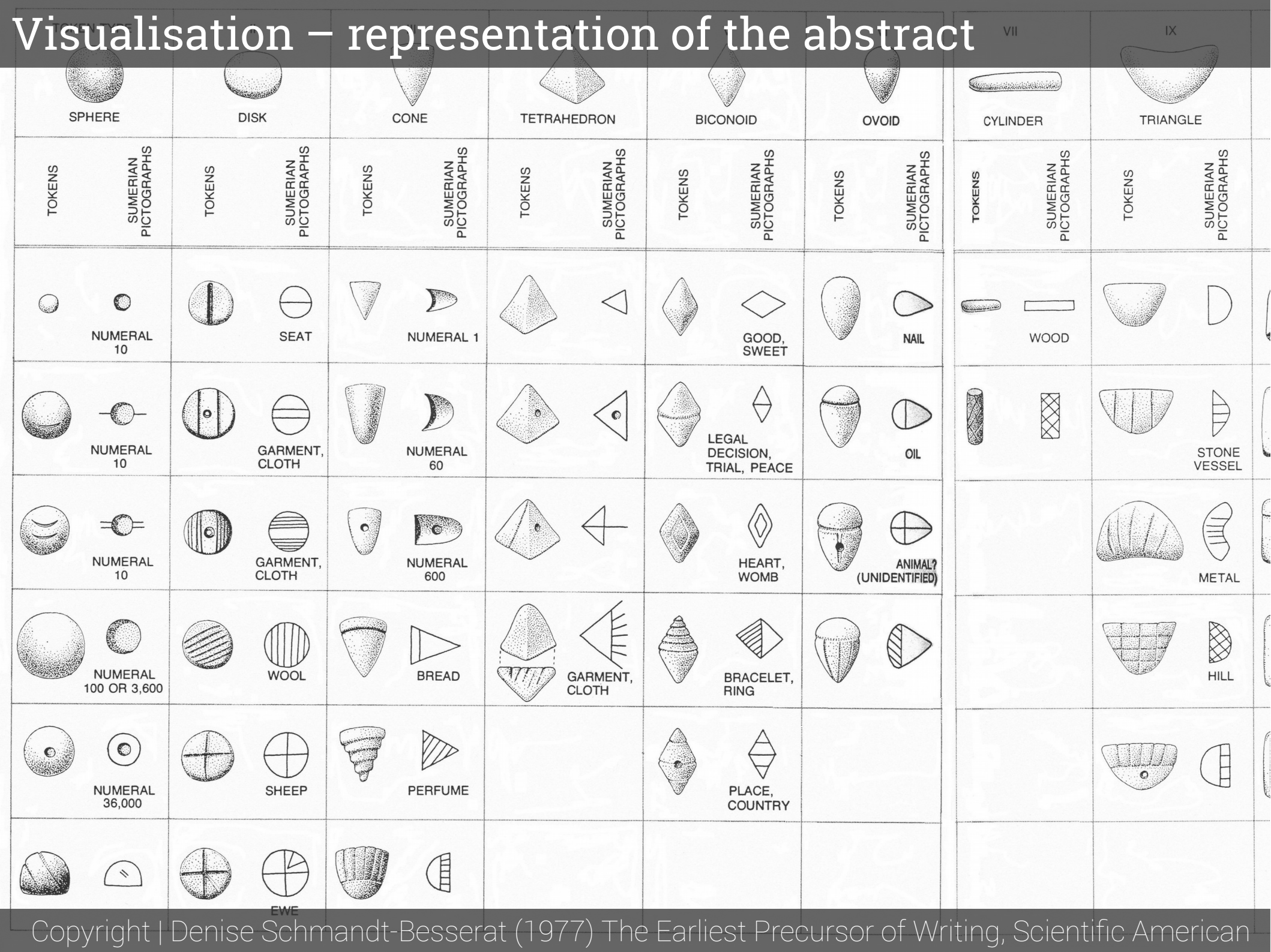
Motivation
Components
Perception
Guidance

Motivation

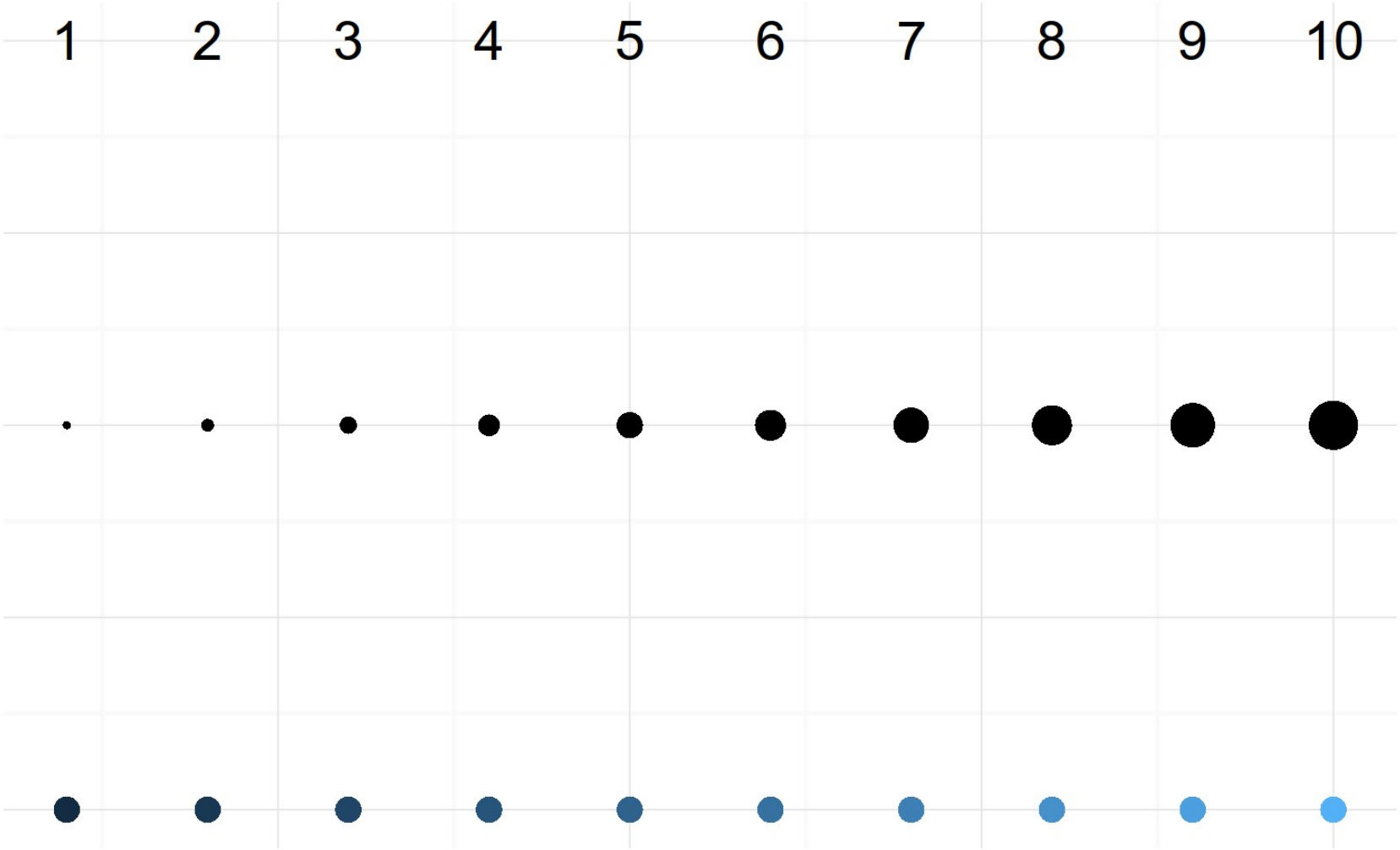
Data – raw symbols

Information – meaning from context

Visualisation – representation of the abstract



Encoding – data → aesthetics



How similar are these sets?

Anscombe's Quartet

Set 1		Set 2		Set 3		Set 4	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Anscombe (1973) Graphs in Statistical Analysis reproduced in Anthony (2012) Qlik Design Blog

How similar are these sets?

Anscombe's Quartet

Set 1		Set 2		Set 3		Set 4	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Property	Value
Mean of X in each case:	9 (exact)
Variance of X in each case:	11 (exact)
Mean of Y in each case:	7.50
Variance of Y in each case:	4.122 or 4.127
Correlation between X & Y in each case:	0.816
Linear regression line in each case:	y=3.00 + 0.500x

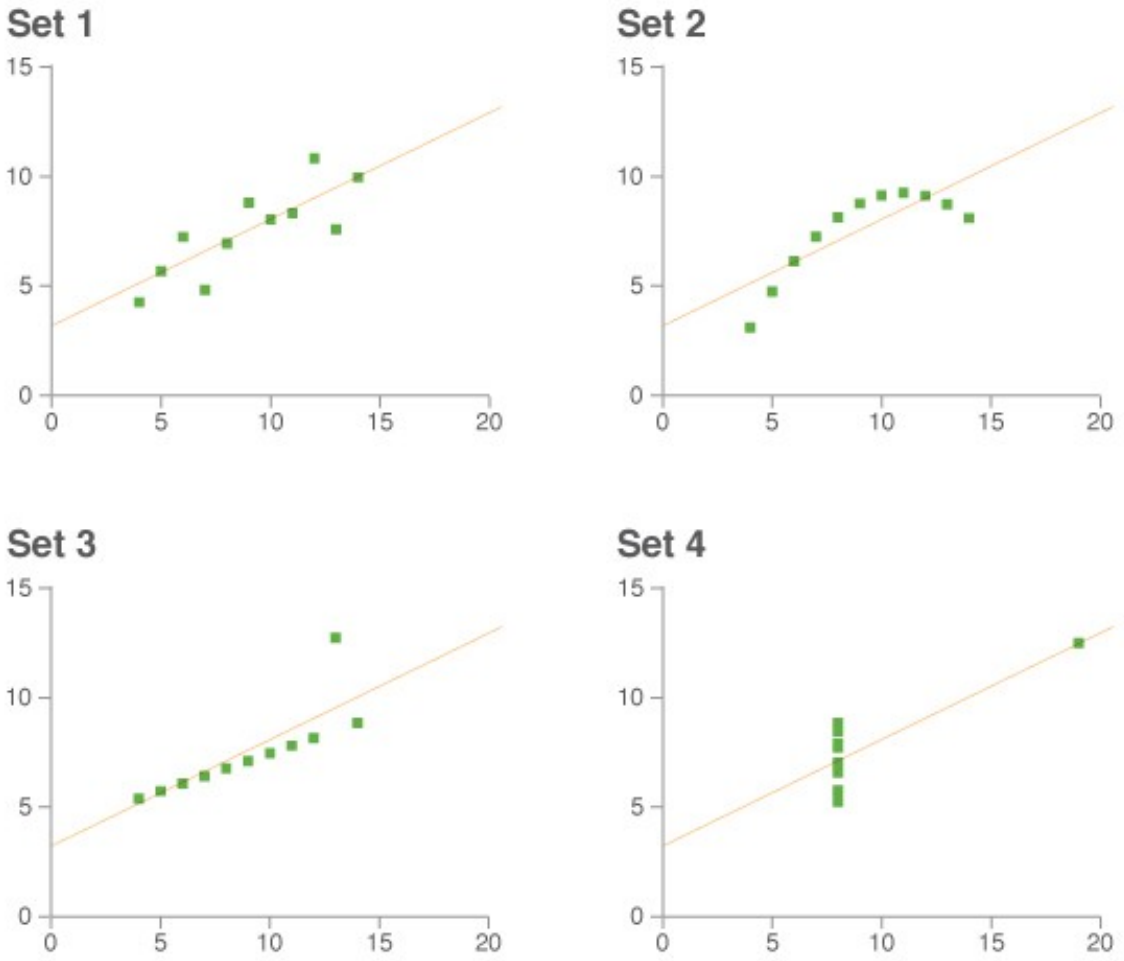
Anscombe (1973) Graphs in Statistical Analysis reproduced in Anthony (2012) Qlik Design Blog

How similar are these sets?

Anscombe's Quartet

Set 1		Set 2		Set 3		Set 4	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Property	Value
Mean of X in each case:	9 (exact)
Variance of X in each case:	11 (exact)
Mean of Y in each case:	7.50
Variance of Y in each case:	4.122 or 4.127
Correlation between X & Y in each case:	0.816
Linear regression line in each case:	y=3.00 + 0.500x



Anscombe (1973) Graphs in Statistical Analysis reproduced in Anthony (2012) Qlik Design Blog

Components

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

9 780387 245447

springeronline.com

Wilkinson

The Grammar of Graphics Second Edition

Statistics and Computing

Leland Wilkinson

The Grammar of Graphics

Second Edition

Springer

Variables

lib

Filter

	http...data.gmdsp.org.uk.data.trafford.libraries.stats.prop.area	Area	Adult.Fiction	Adult.Non.Fiction	Junior.Easy.Reader	Junior.Fiction	Junior.Non.Fiction	Talking.Books	Young.Adul
1	http://opendatacommunities.org/doc/geography/admi...	Altrincham	4988	1993	1391	5812	1084	207	
2	http://opendatacommunities.org/doc/geography/admi...	Ashton upon Mersey	5342	2207	884	3682	693	237	
3	http://opendatacommunities.org/doc/geography/admi...	Bowdon	4879	1825	1202	4234	502	295	
4	http://opendatacommunities.org/doc/geography/admi...	Broadheath	5203	2417	1418	6389	826	314	
5	http://opendatacommunities.org/doc/geography/admi...	Brooklands	5709	2112	1317	5534	878	344	
6	http://opendatacommunities.org/doc/geography/admi...	Bucklow-St Martins	3021	732	301	1026	227	159	
7	http://opendatacommunities.org/doc/geography/admi...	Clifford	1329	1126	493	1684	558	38	
8	http://opendatacommunities.org/doc/geography/admi...	Davyhulme East	6390	2097	789	3488	321	257	
9	http://opendatacommunities.org/doc/geography/admi...	Davyhulme West	6925	1866	733	2467	471	143	
10	http://opendatacommunities.org/doc/geography/admi...	Flixton	6784	2619	1356	4352	489	486	
11	http://opendatacommunities.org/doc/geography/admi...	Gorse Hill	3089	1237	777	2569	344	64	
12	http://opendatacommunities.org/doc/geography/admi...	Hale Barns	6281	2038	1169	3724	520	518	
13	http://opendatacommunities.org/doc/geography/admi...	Hale Central	5590	2601	1555	6166	1103	534	
14	http://opendatacommunities.org/doc/geography/admi...	Longford	2307	1526	329	1444	326	164	
15	http://data.gmdsp.org.uk/data/trafford/area/out-of-area	OutOfBorough / Unmatched	13394	6840	2472	11141	1752	1398	
16	http://opendatacommunities.org/doc/geography/admi...	Priory	4997	2484	1507	5197	1026	367	
17	http://opendatacommunities.org/doc/geography/admi...	Sale Moor	4132	1295	606	2509	433	173	
18	http://opendatacommunities.org/doc/geography/admi...	St Mary's	8686	2973	1094	3433	842	1106	
19	http://opendatacommunities.org/doc/geography/admi...	Stretford	5076	2514	725	3527	706	456	
20	http://opendatacommunities.org/doc/geography/admi...	Timperley	7849	2461	1482	5603	991	320	
21	http://opendatacommunities.org/doc/geography/admi...	Urmston	5651	2532	1036	4276	732	377	
22	http://opendatacommunities.org/doc/geography/admi...	Village	5363	2058	680	4020	545	275	

Showing 1 to 22 of 22 entries

Transformations

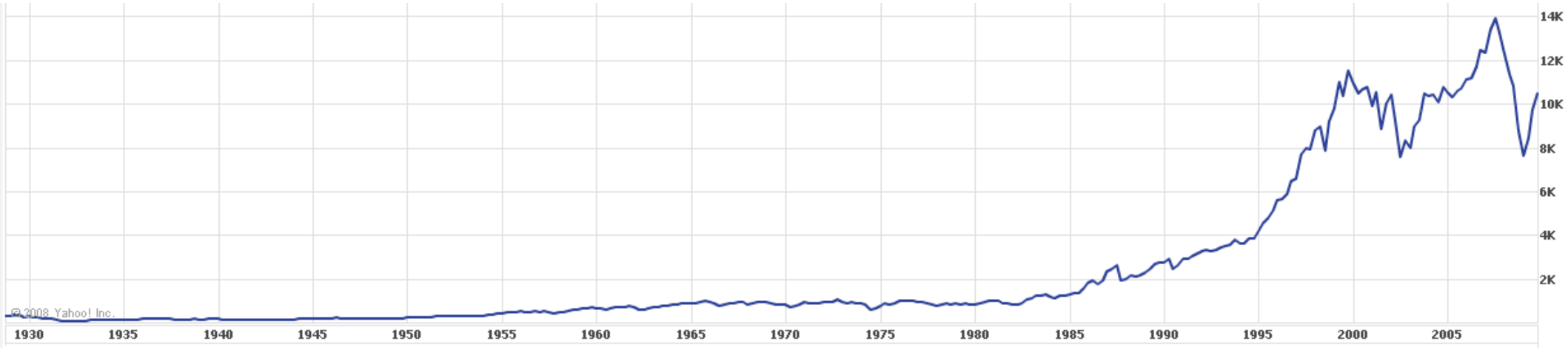
lib[, c("Area", "Adult.Fiction", "Adult.Non.Fiction", "Adult.Total",

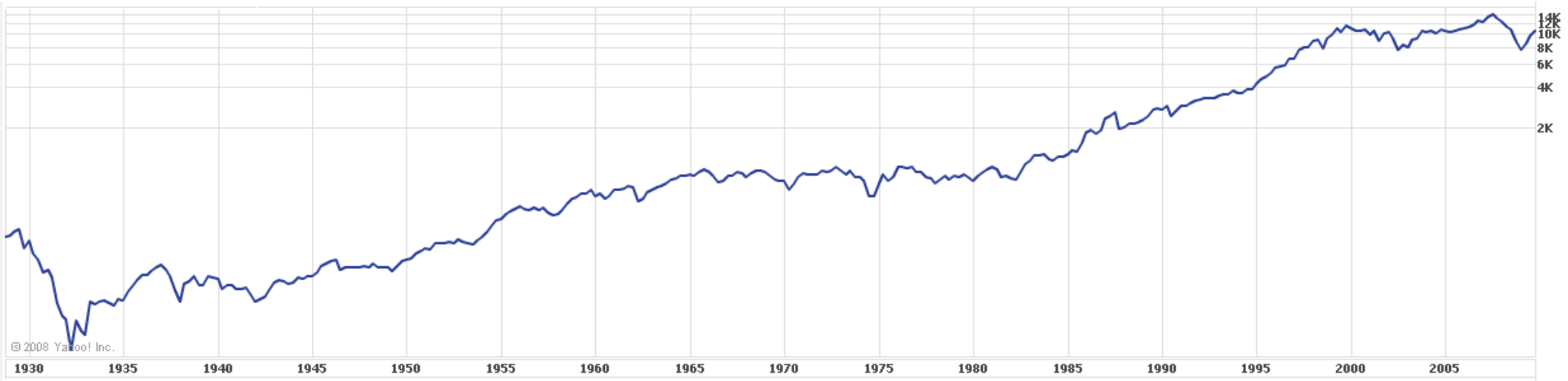
Filter

	Area	Adult.Fiction	Adult.Non.Fiction	Adult.Total	Adult.Fiction.Ratio	Adult.Fiction.Ratio.Rank
7	Clifford	1329	1126	2455	0.5413442	1
14	Longford	2307	1526	3833	0.6018784	2
15	OutOfBorough / Unmatched	13394	6840	20234	0.6619551	3
16	Priory	4997	2484	7481	0.6679588	4
19	Stretford	5076	2514	7590	0.6687747	5
13	Hale Central	5590	2601	8191	0.6824564	6
4	Broadheath	5203	2417	7620	0.6828084	7
21	Urmston	5651	2532	8183	0.6905780	8
2	Ashton upon Mersey	5342	2207	7549	0.7076434	9
11	Gorse Hill	3089	1237	4326	0.7140546	10
1	Altrincham	4988	1993	6981	0.7145108	11
10	Flixton	6784	2619	9403	0.7214719	12
22	Village	5363	2058	7421	0.7226789	13
3	Bowdon	4879	1825	6704	0.7277745	14
5	Brooklands	5709	2112	7821	0.7299578	15
18	St Mary's	8686	2973	11659	0.7450039	16
8	Davyhulme East	6390	2097	8487	0.7529162	17
12	Hale Barns	6281	2038	8319	0.7550186	18
20	Timperley	7849	2461	10310	0.7612997	19
17	Sale Moor	4132	1295	5427	0.7613783	20
9	Davyhulme West	6925	1866	8791	0.7877375	21
6	Bucklow-St Martins	3021	732	3753	0.8049560	22

Scales – mapping to a common unit

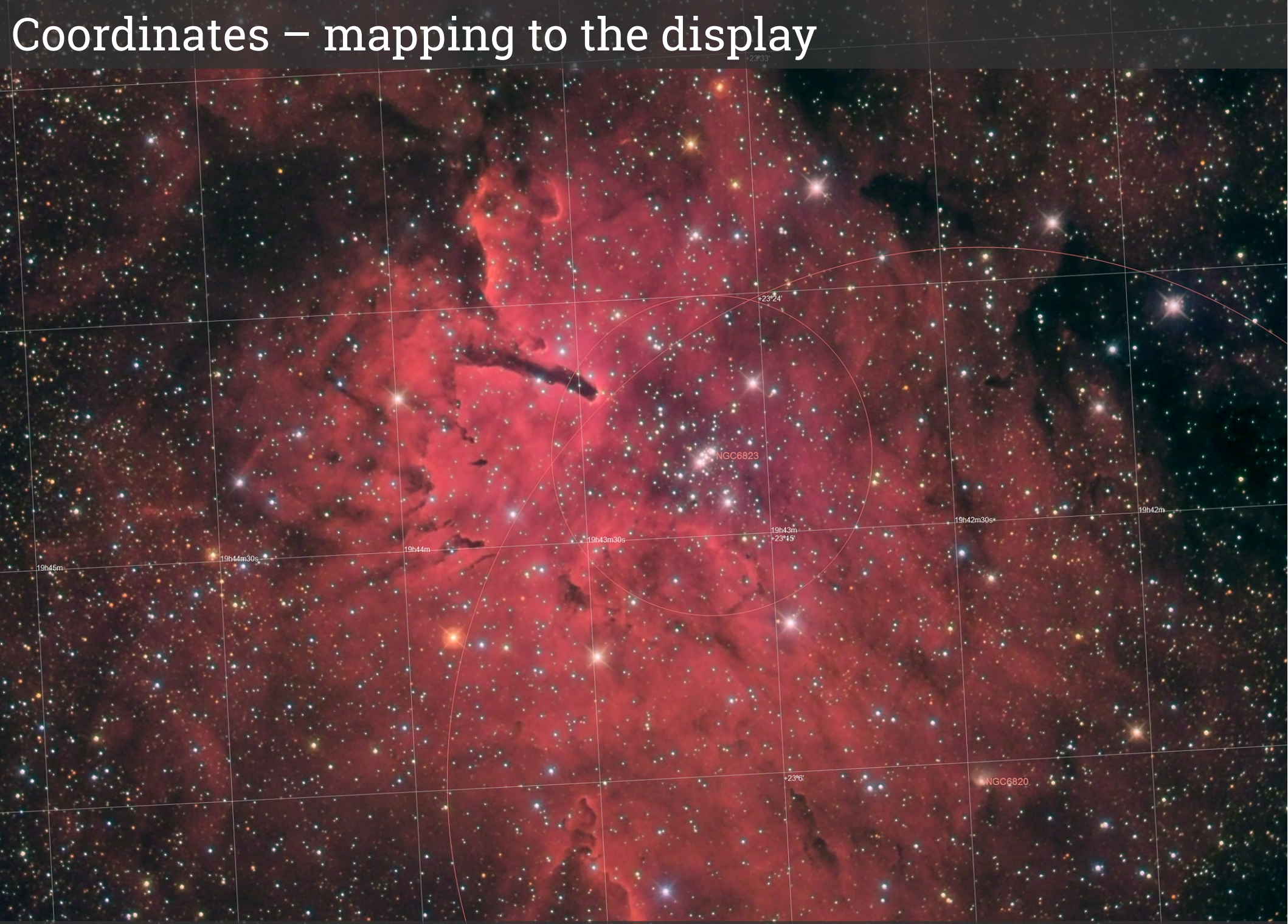
Scales – mapping to a common unit



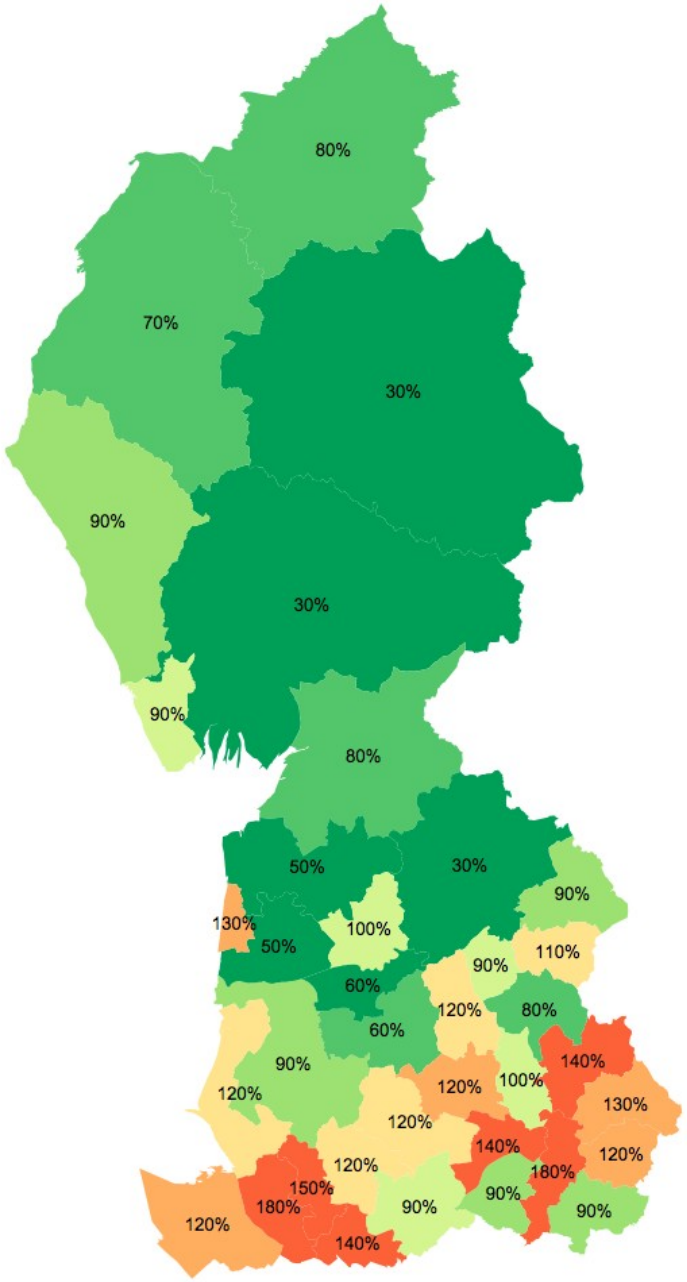


Yahoo Finance via the Generalist (http://legeneraliste.perso.sfr.fr/?p=echelle_log_eng)

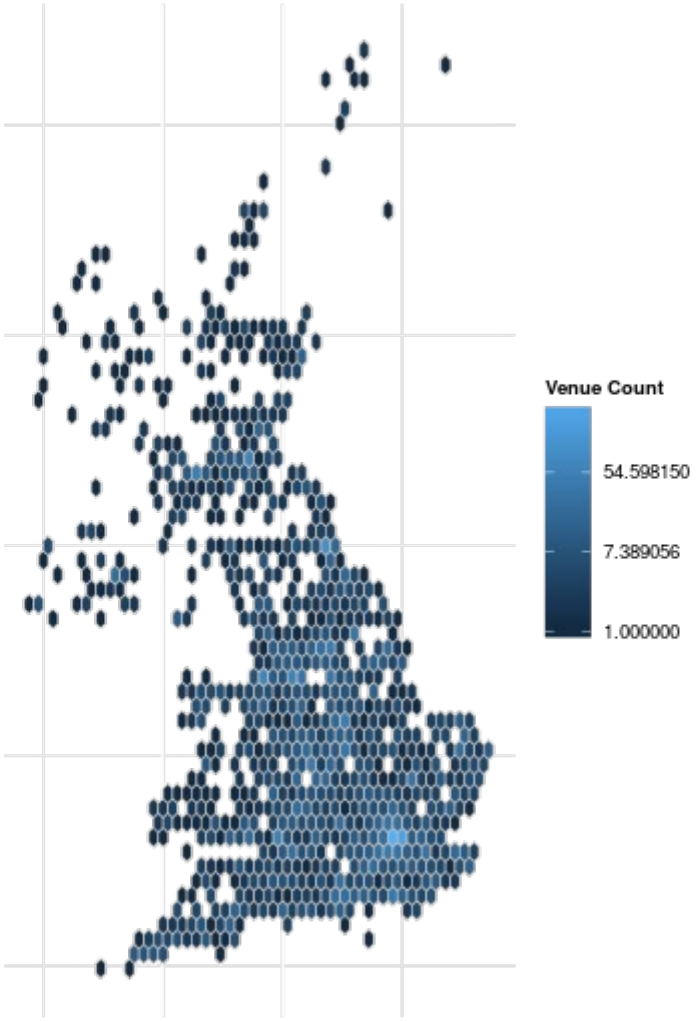
Coordinates – mapping to the display

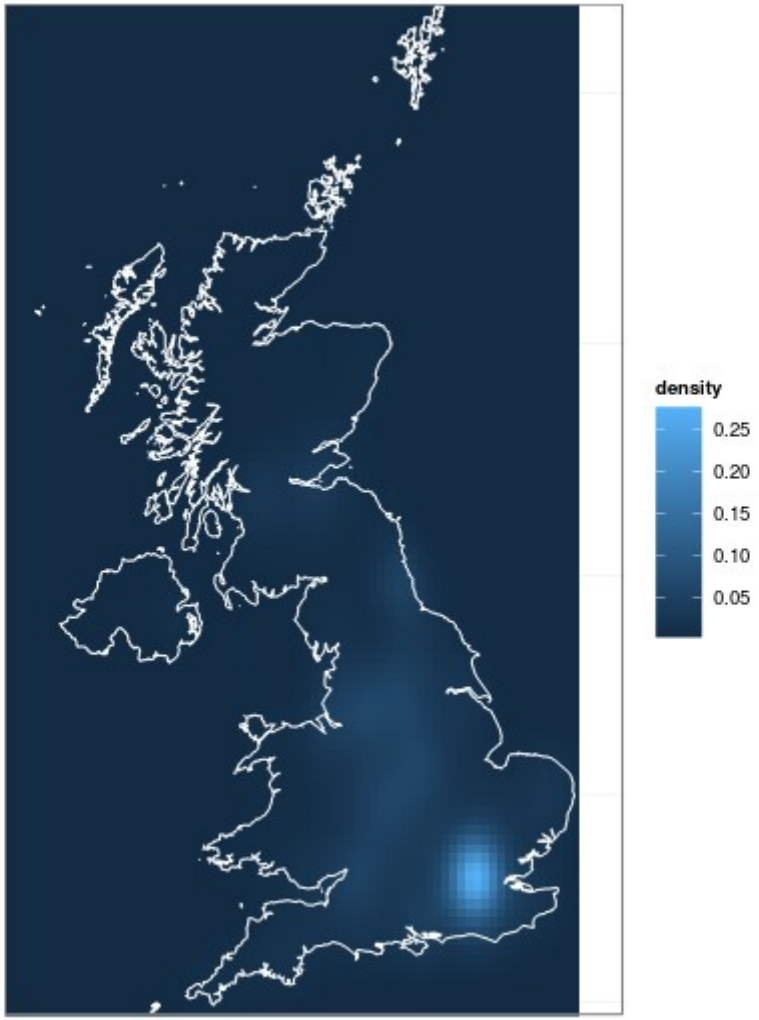


Coordinates – mapping to the display



Coordinates – mapping to the display





Elements – aesthetic attributes

this page is intentionally left blank

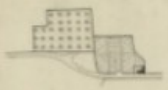
ORDNANCE SURVEY

CHARACTERISTIC SHEET

for the Engraved Six Inch Maps

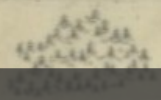
OF GREAT BRITAIN.

Parks and Ornamental Ground.

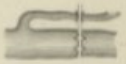


Orchards and Gardens.

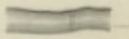
Woods. (Deciduous Trees)



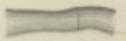
Fir Plantations



Bridges.



Ford.

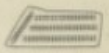


Ferry.

Marl Pit.

Gravel Pit.

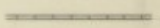
Quarry.



Tenter Ground.

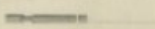
Well.

Trigonometrical Point.



Double Line of Railway.

Single d° d°



Canals.

Main Roads

Other Roads

Contours & Altitudes.

Instrumental Contours

Interpolated d°

Boundaries

Counties

Divisions of Counties

Parliamentary Div^ns of Counties

Hundreds or Wapentakes or Wards

Parishes (Mother or Ancient)

Civil Parishes or Townships

Counties & Hundreds

Counties & Mother Parishes

Counties & Civil Parishes or Townships

Counties Hundreds Mother Parishes & Townships

Parliamentary Borough

Municipal Borough

County Boroughs

Parliamentary Div^ns of C^ty Borough

Wards of Corporate Towns

Perception

3.14159265358979323846264338327950288419716939937510582097494459230781640628620899862803482534211706798214808651328230664

Pre-attentive Processing

3.14159265358979323846264338327950288419716939937510582097494459230781640628620899862803482534211706798214808651328230664

Decoding accuracy

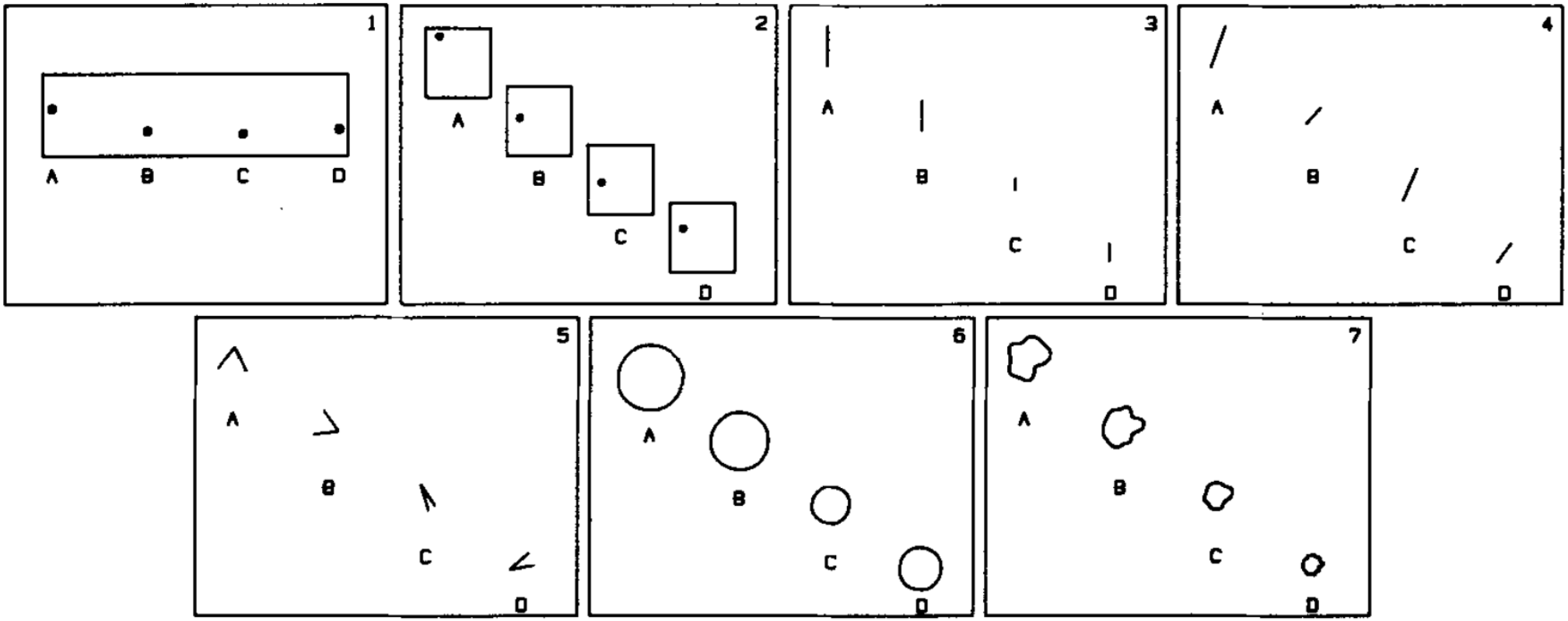


FIG. 2. Stimuli from experiment. An experiment was run to investigate the relative accuracy of basic graphical judgments. The seven types of displays in this figure were judged by subjects. The displays required the following judgments (proceeding from left to right and top to bottom). (1) position along a common scale; (2) position along identical, non-aligned scales; (3) length, (4) slope; (5) angle; (6) area; (7) area.

Decoding accuracy

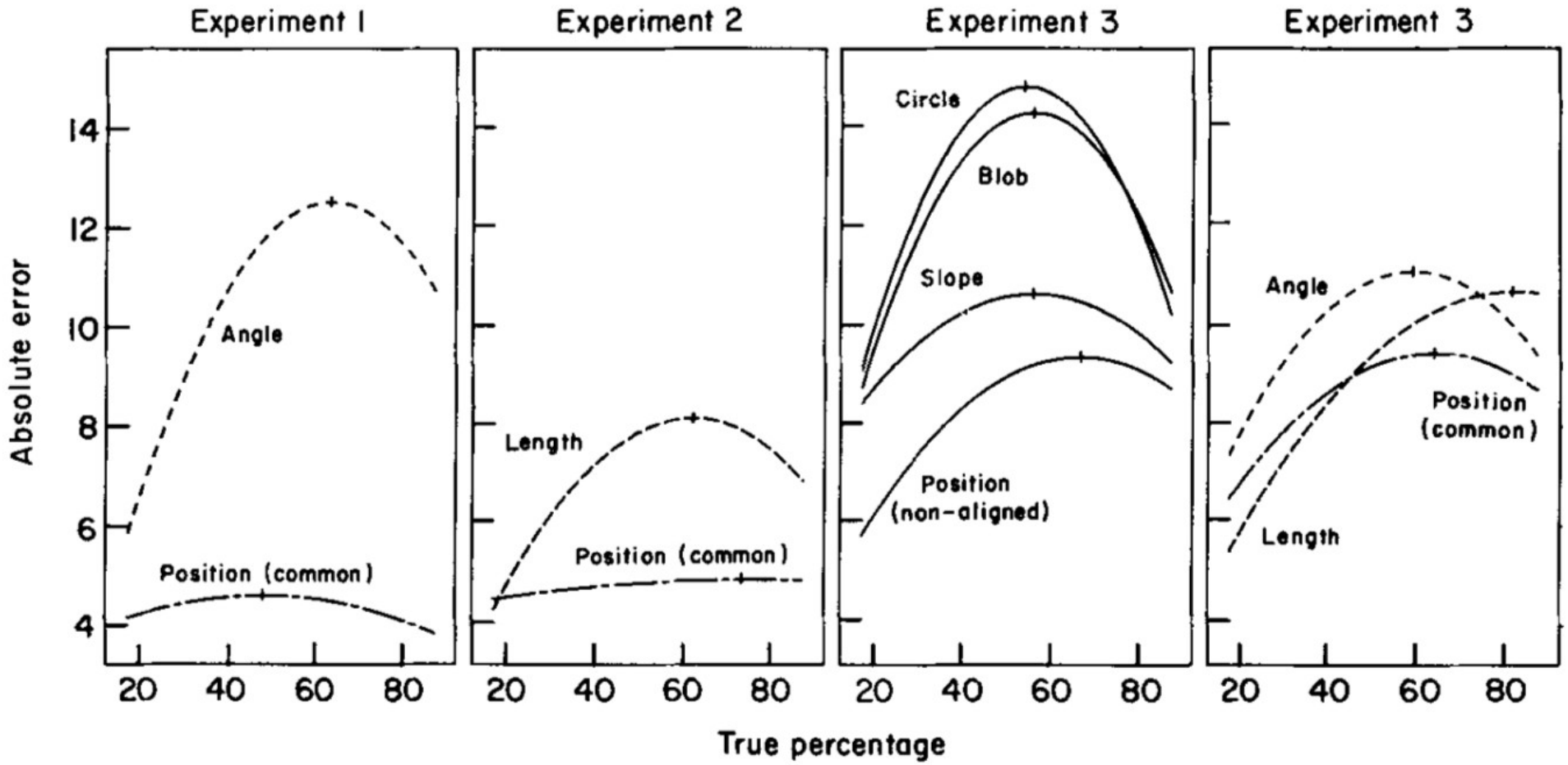


FIG. 5. Absolute judgment error as a function of true percentage. The dependence of the absolute error of judgments was modeled as a quadratic function of the true percentage for each of the basic graphical judgments in three experiments. The coefficients of the polynomials were estimated by least squares. The figure shows the fitted quadratics with the maxima marked by the short vertical lines. Position judgments are the most accurate, length judgments are second, angle and slope judgments are third, and area judgments are last.

Ranking of Perceptual Tasks

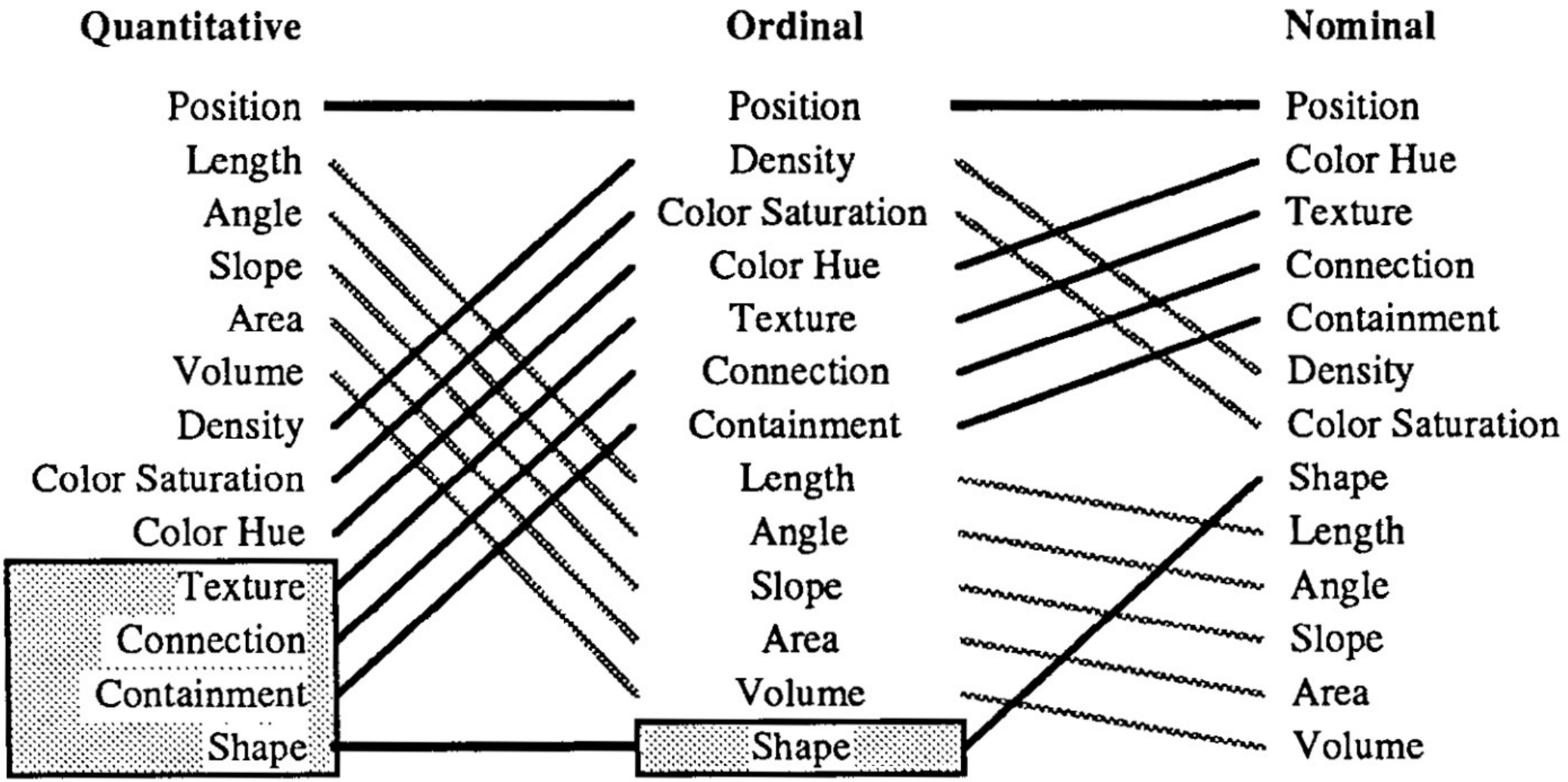
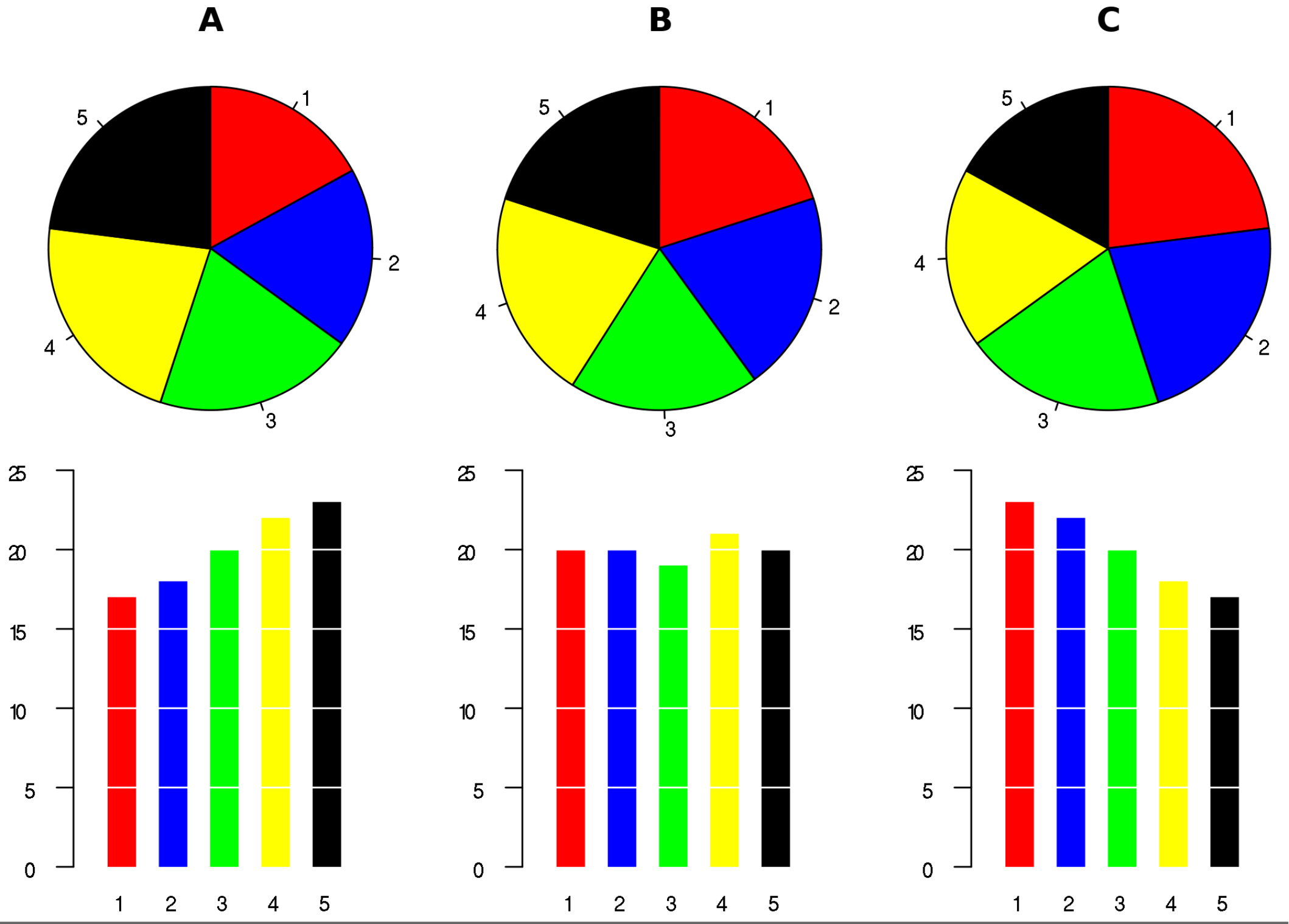
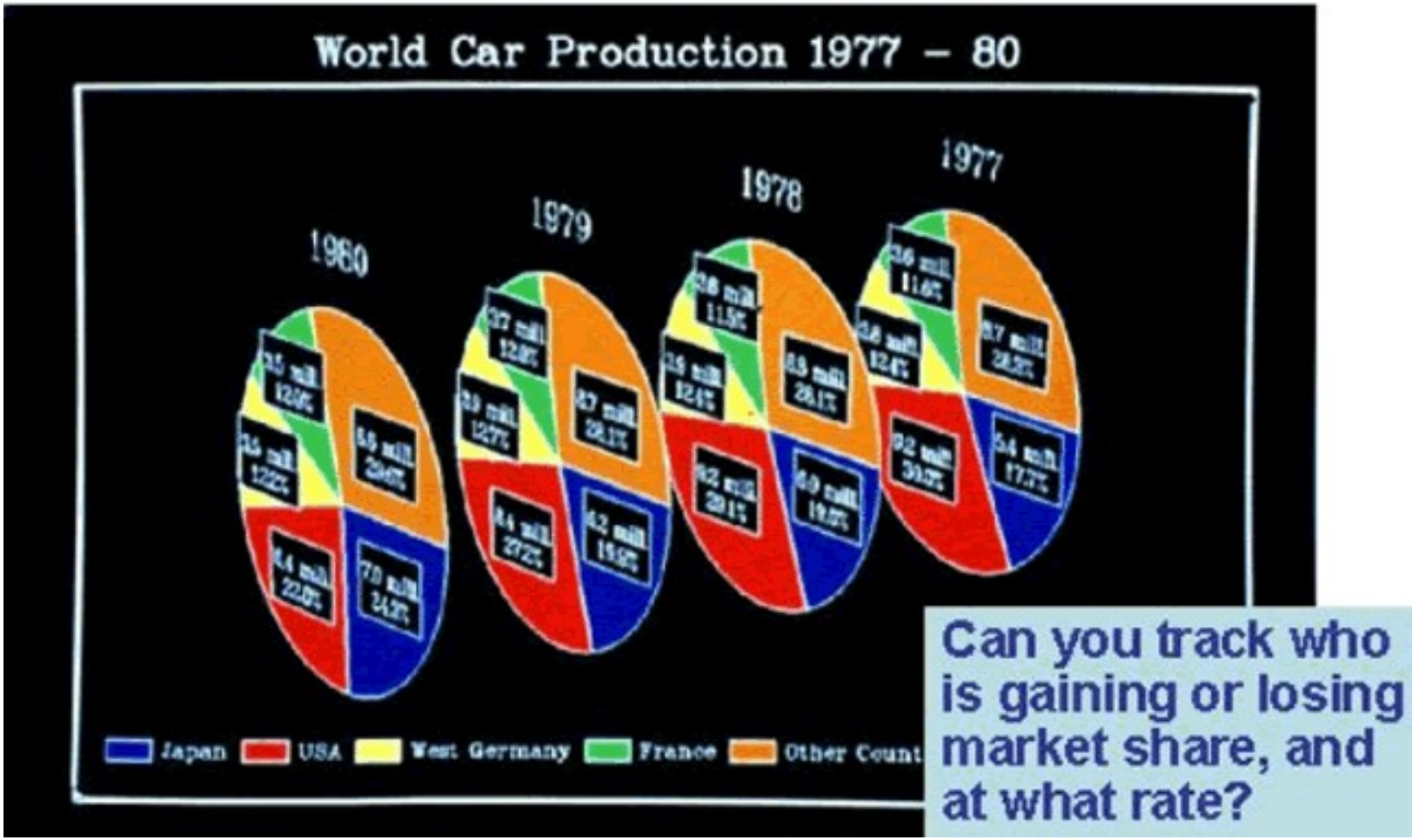
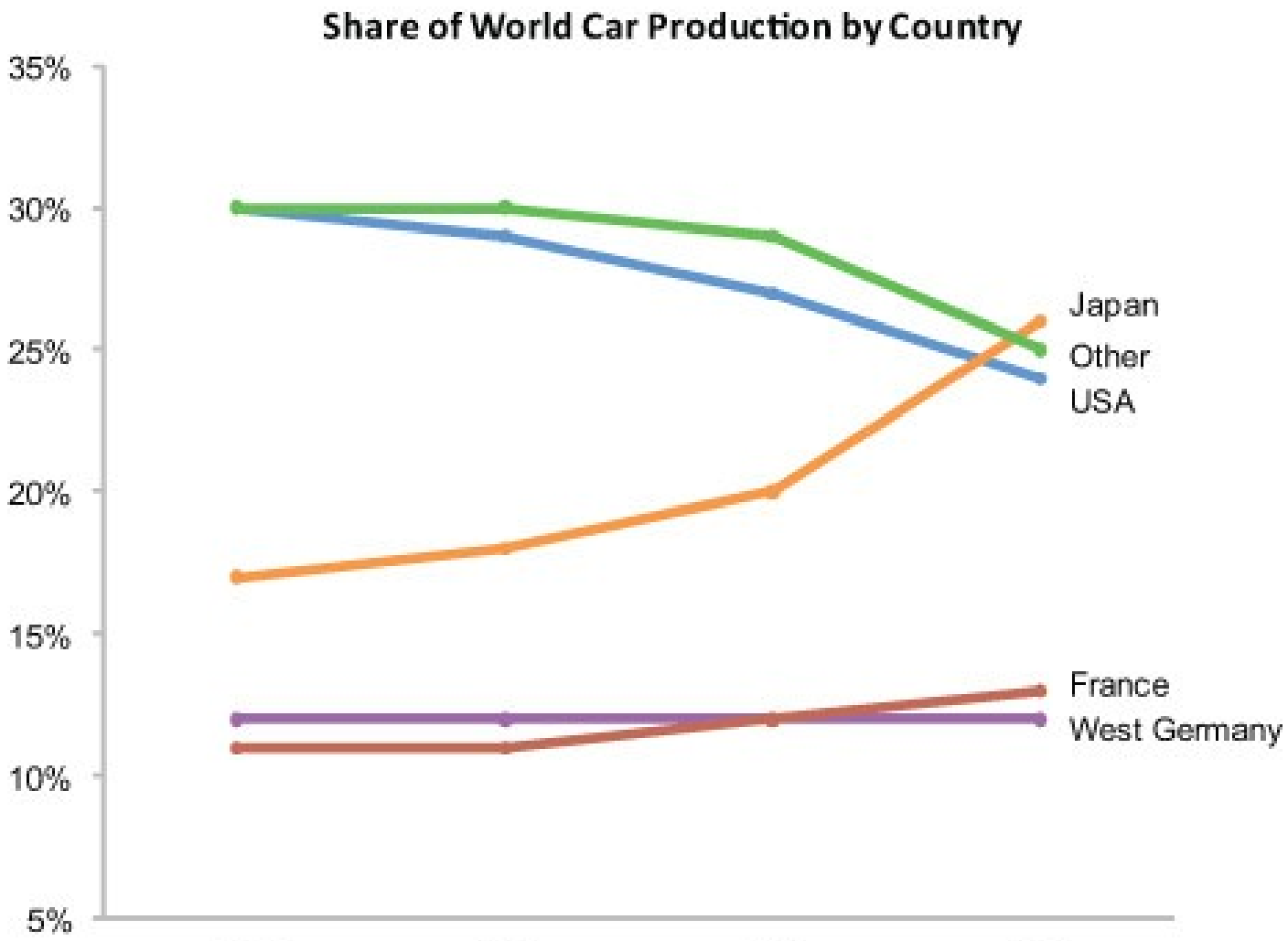


Fig. 15. Ranking of perceptual tasks. The tasks shown in the gray boxes are not relevant to these types of data.

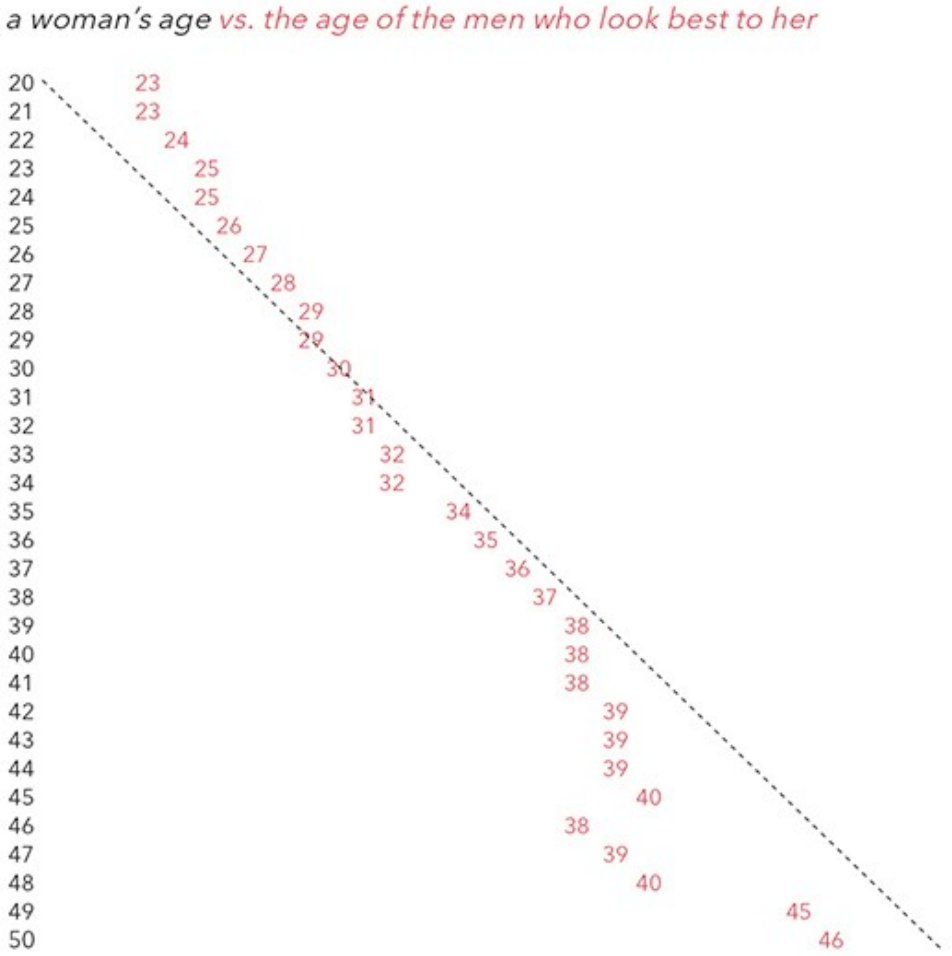


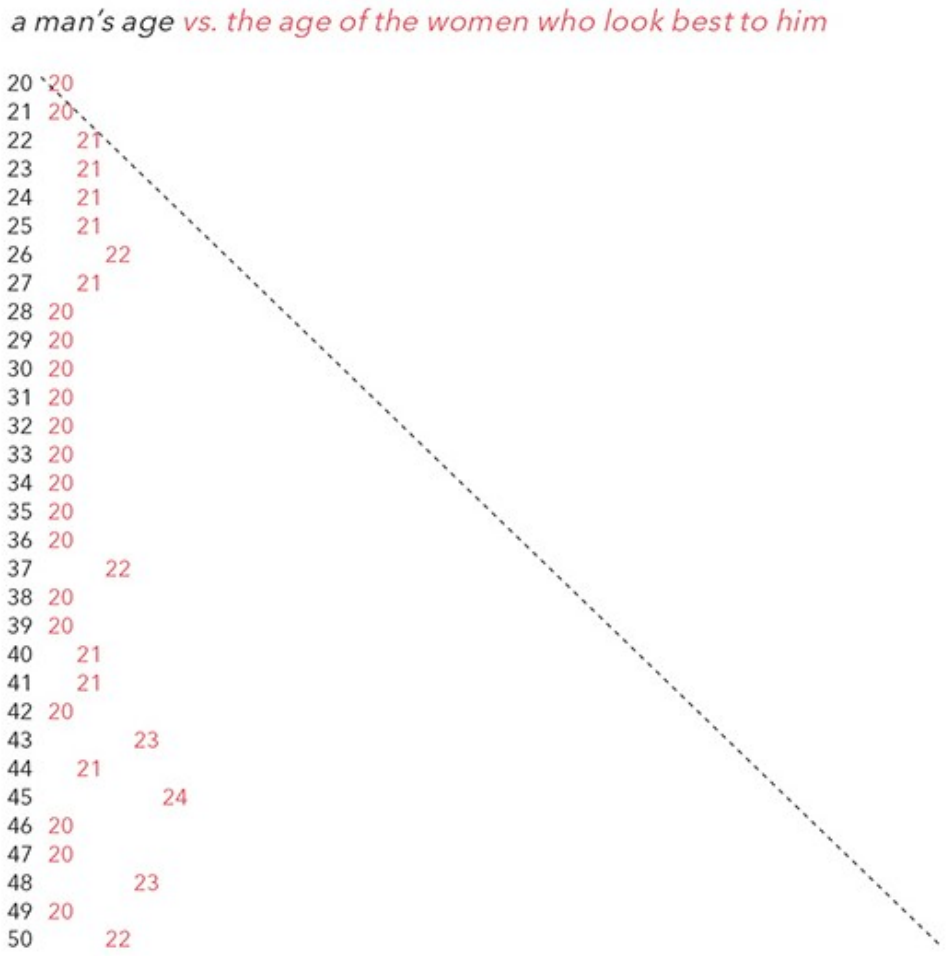

World Car Production 1977 - 80
1980
1979
1978
1977
Japan
USA
West Germany
France
Other Count
Can you track who is gaining or losing market share, and at what rate?



	1977	1978	1979	1980
Japan	17%	18%	20%	26%
Other	30%	30%	29%	25%
USA	30%	29%	27%	24%
France	11%	11%	12%	13%
West Germany	12%	12%	12%	12%
Total	100%	100%	100%	100%

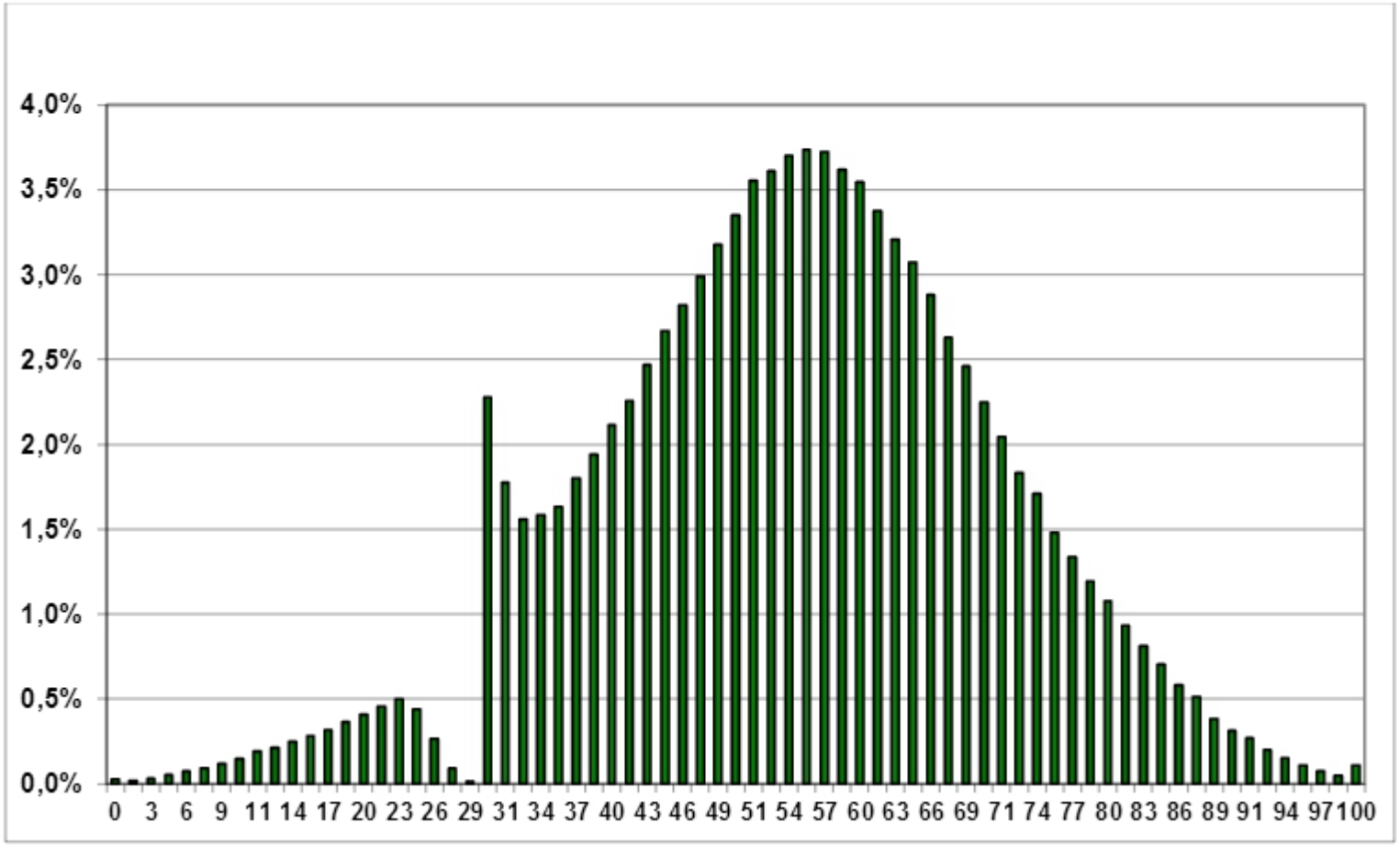
Aesthetics – Position



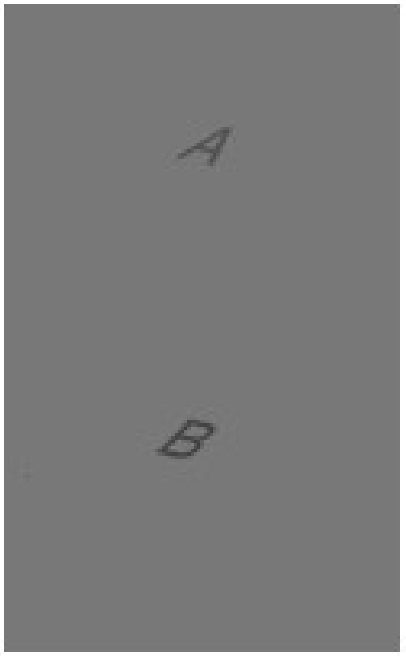


Aesthetics – Position

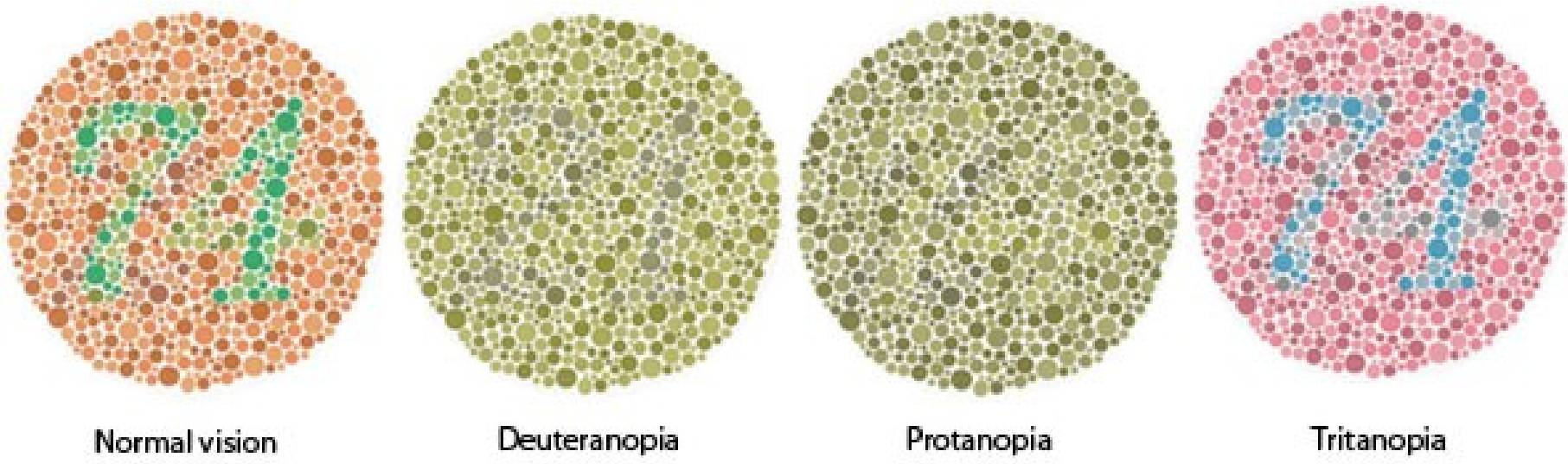
2.1. Poziom podstawowy



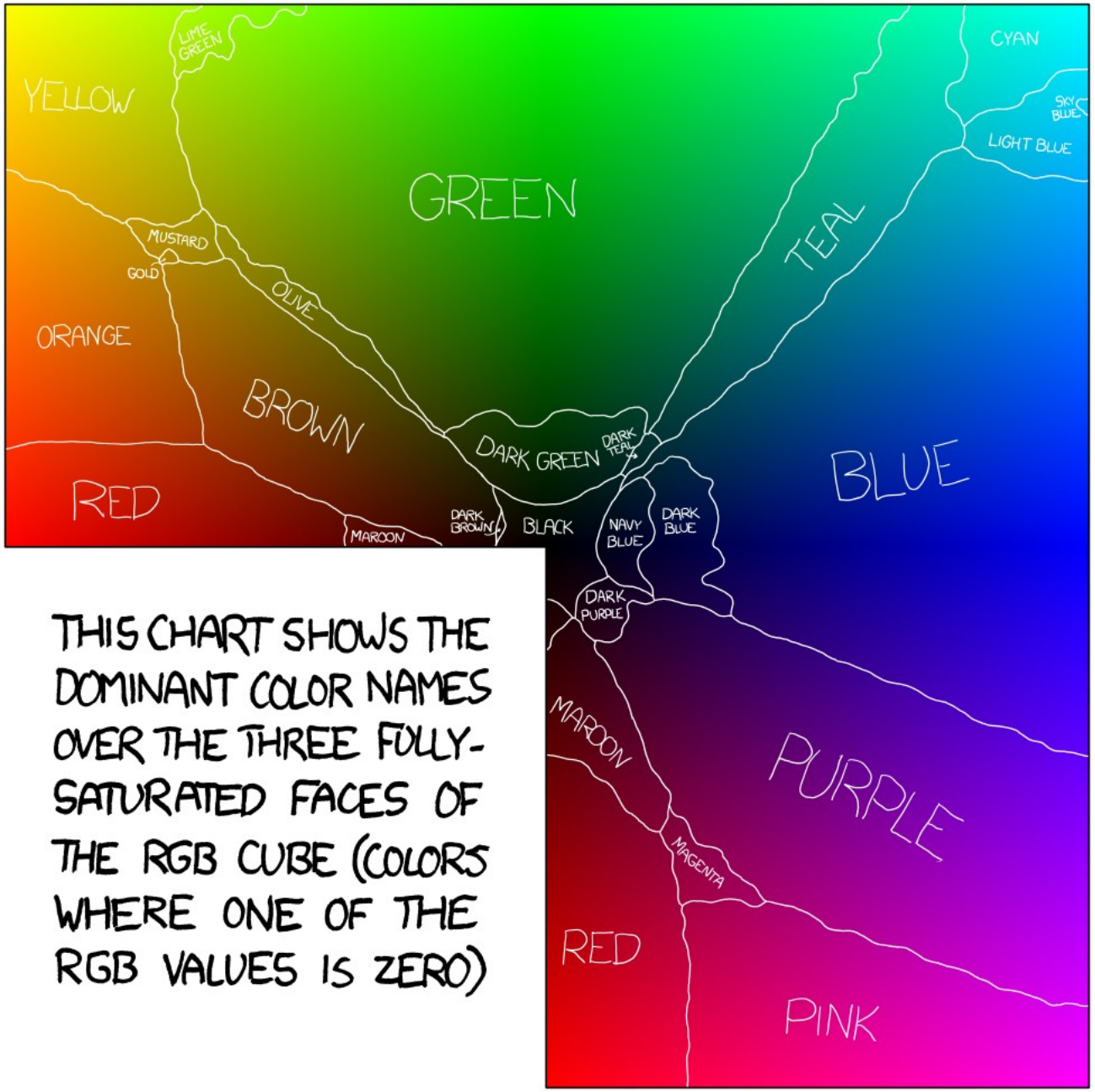
Aesthetics – Colour – depends on context



Aesthetics – Colour – not the same to everyone

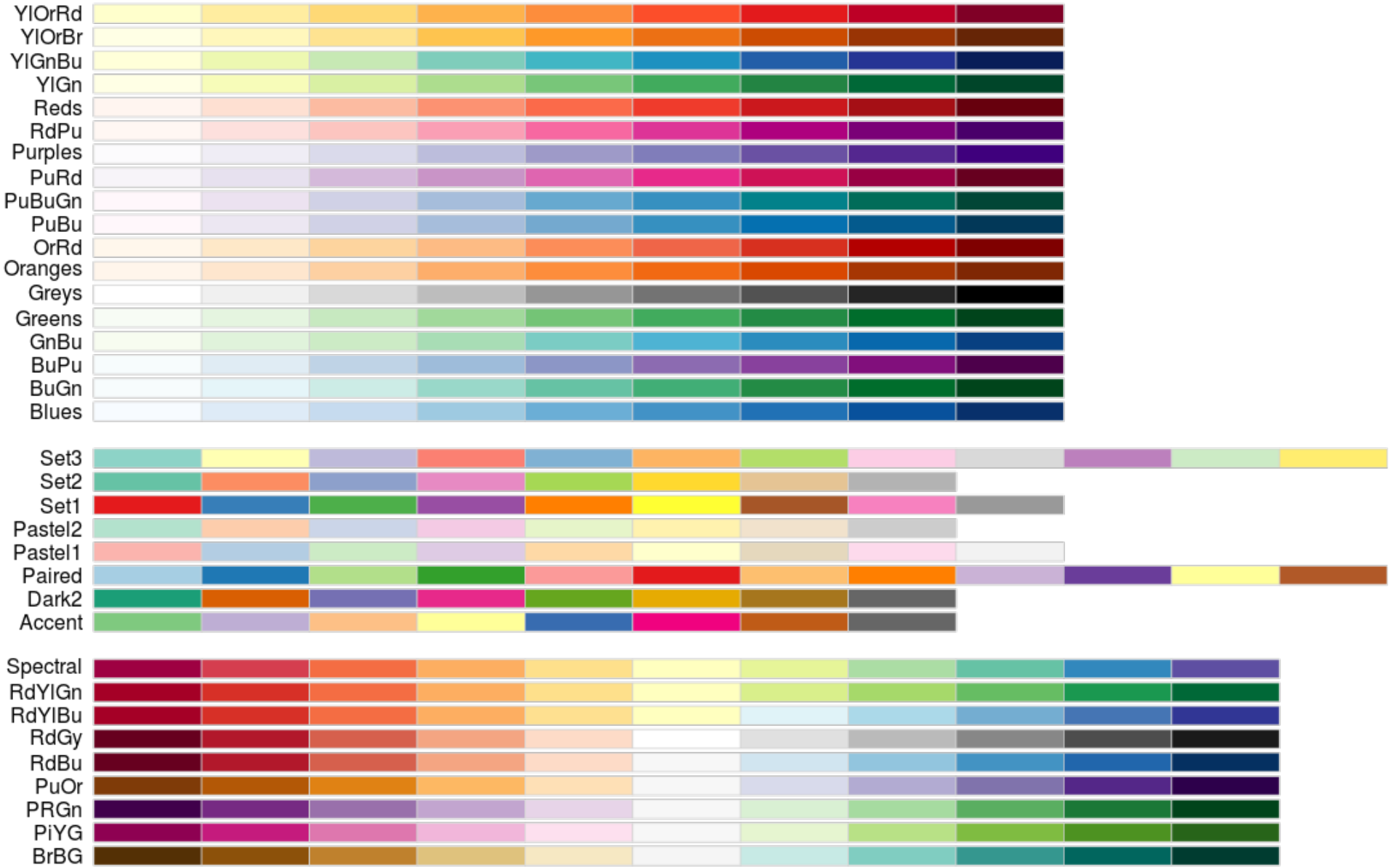


Aesthetics – Colour – limits to perception

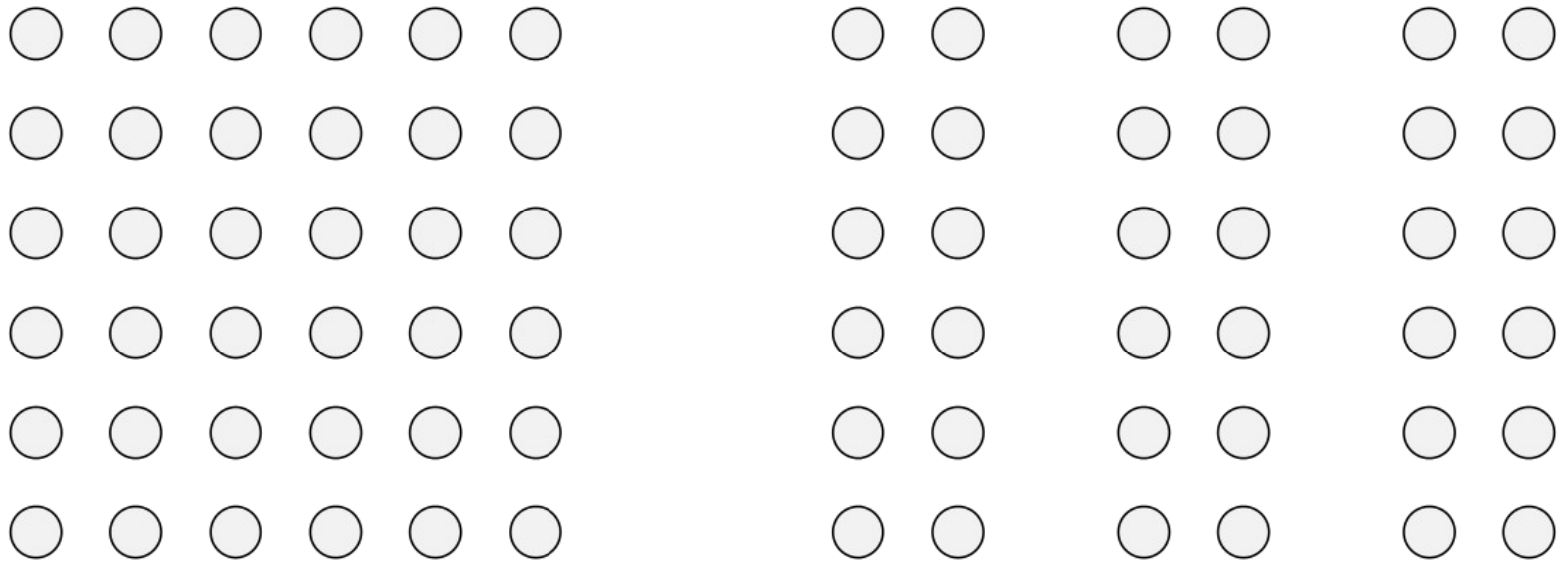


XKCD Colour Survey 2010 http://blog.xkcd.com/2010/05/03/color-survey-results/

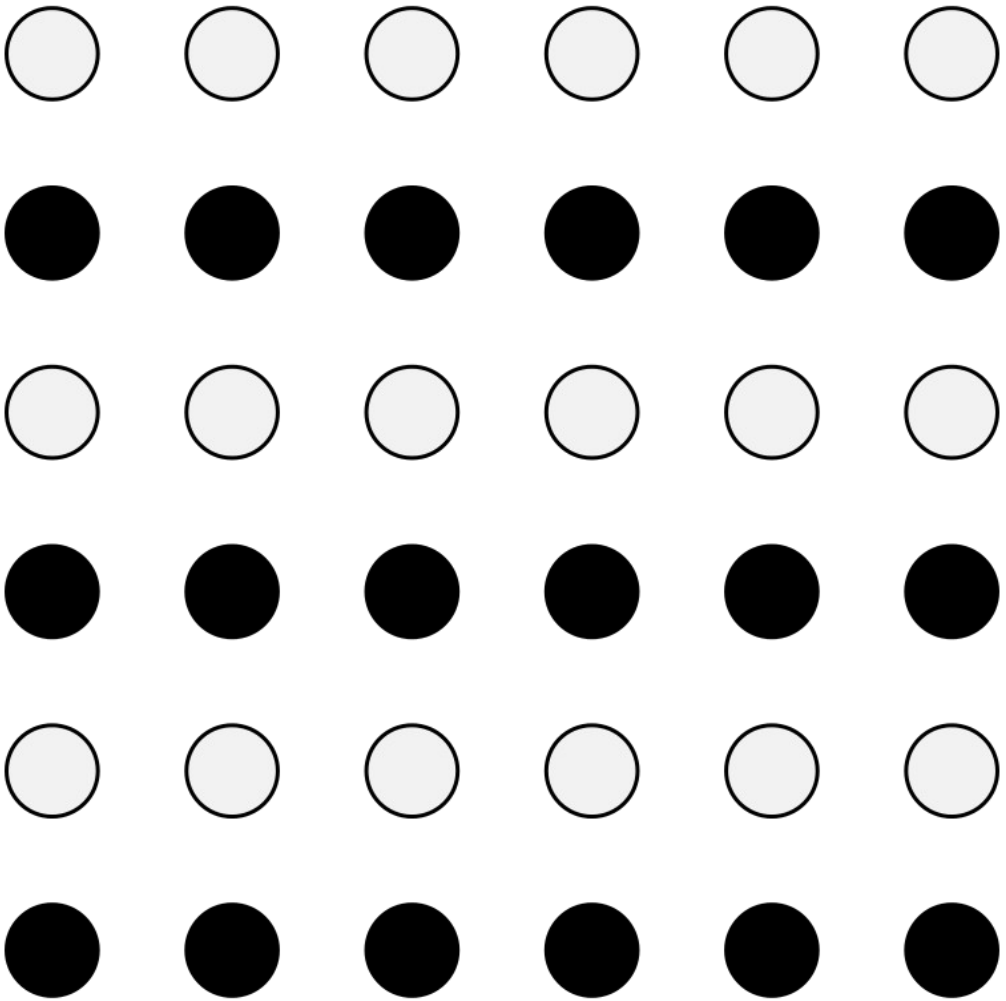
Aesthetics – Colour – reach for a palette



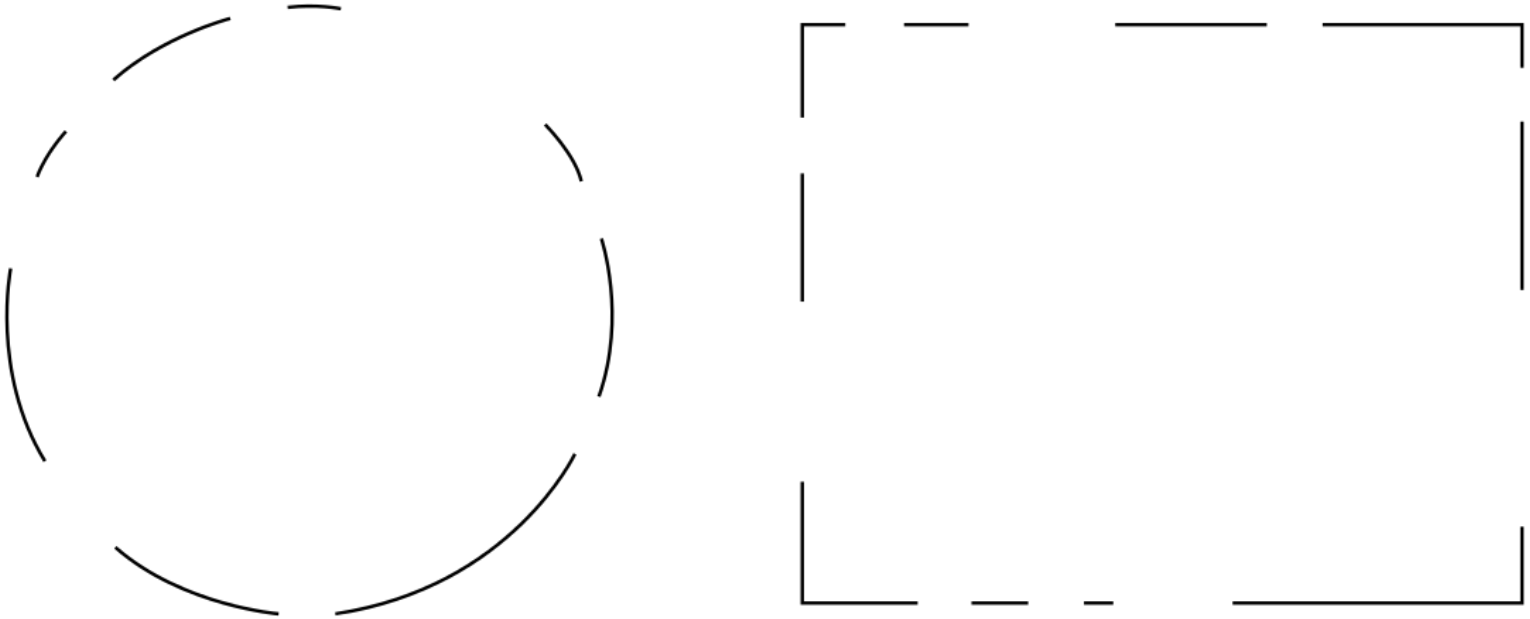
Gestalt laws of grouping – proximity



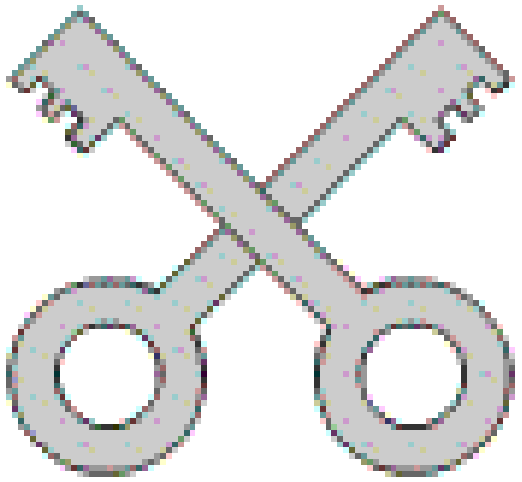
Gestalt laws of grouping – similarity



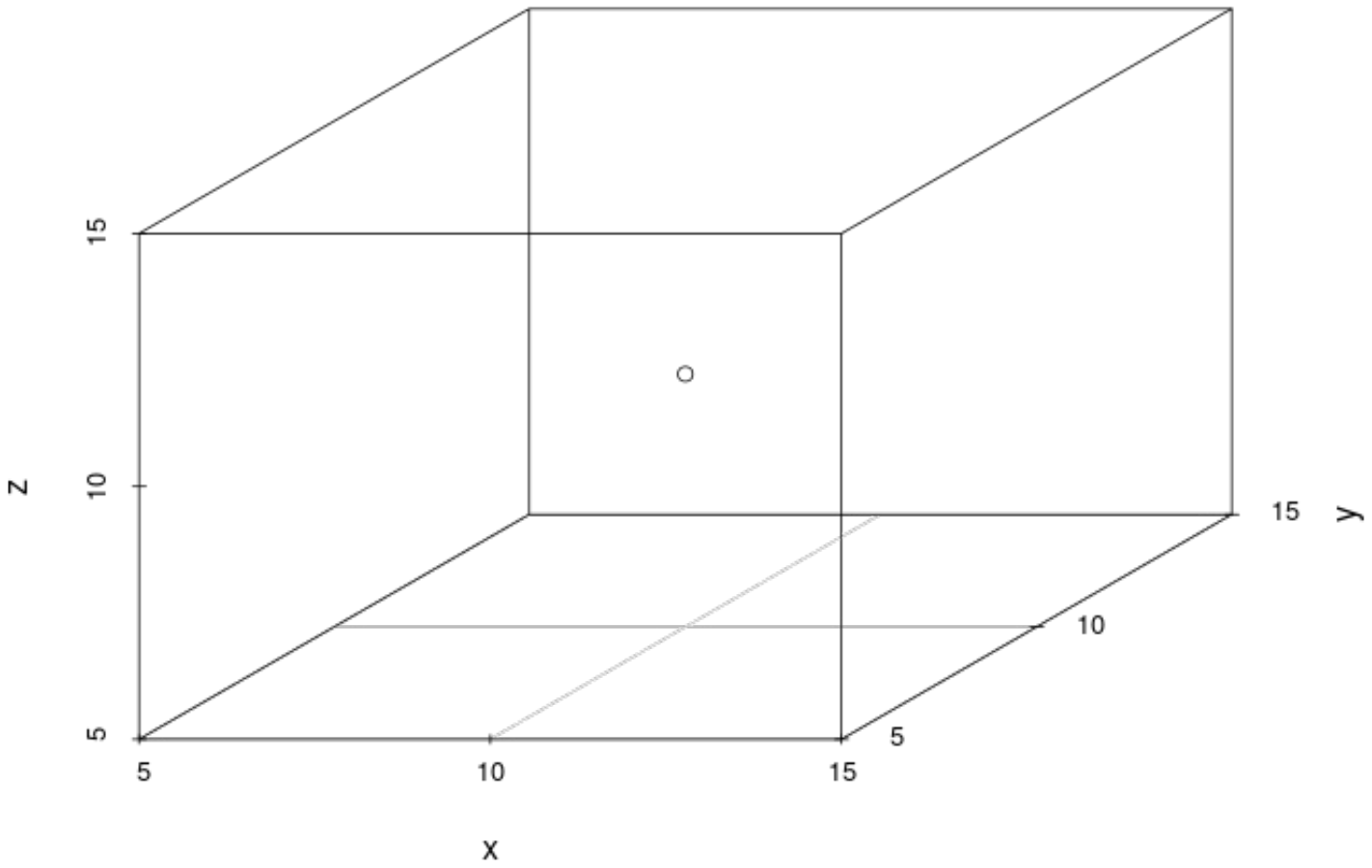
Gestalt laws of grouping – closure

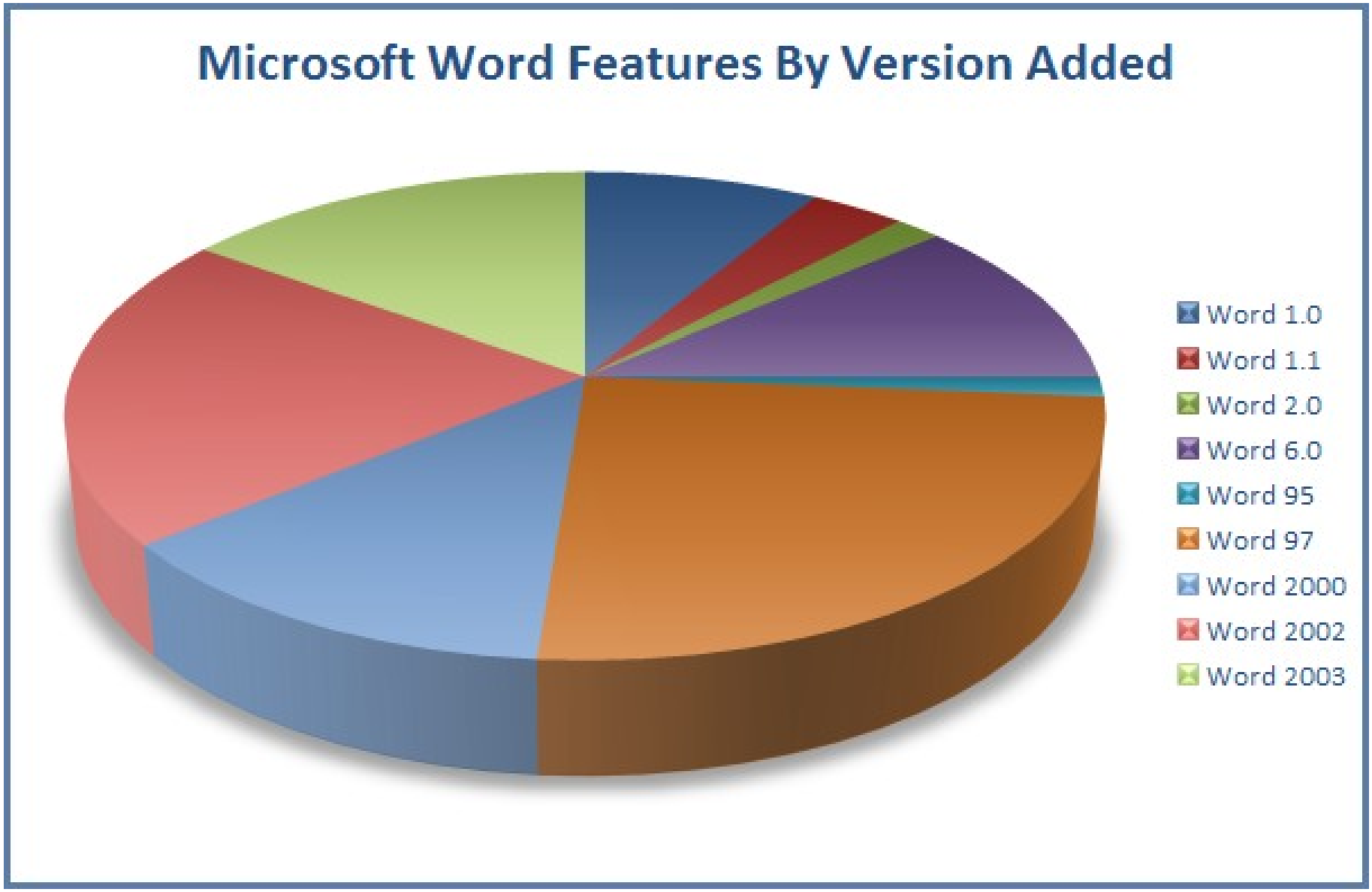


Gestalt laws of grouping – continuation

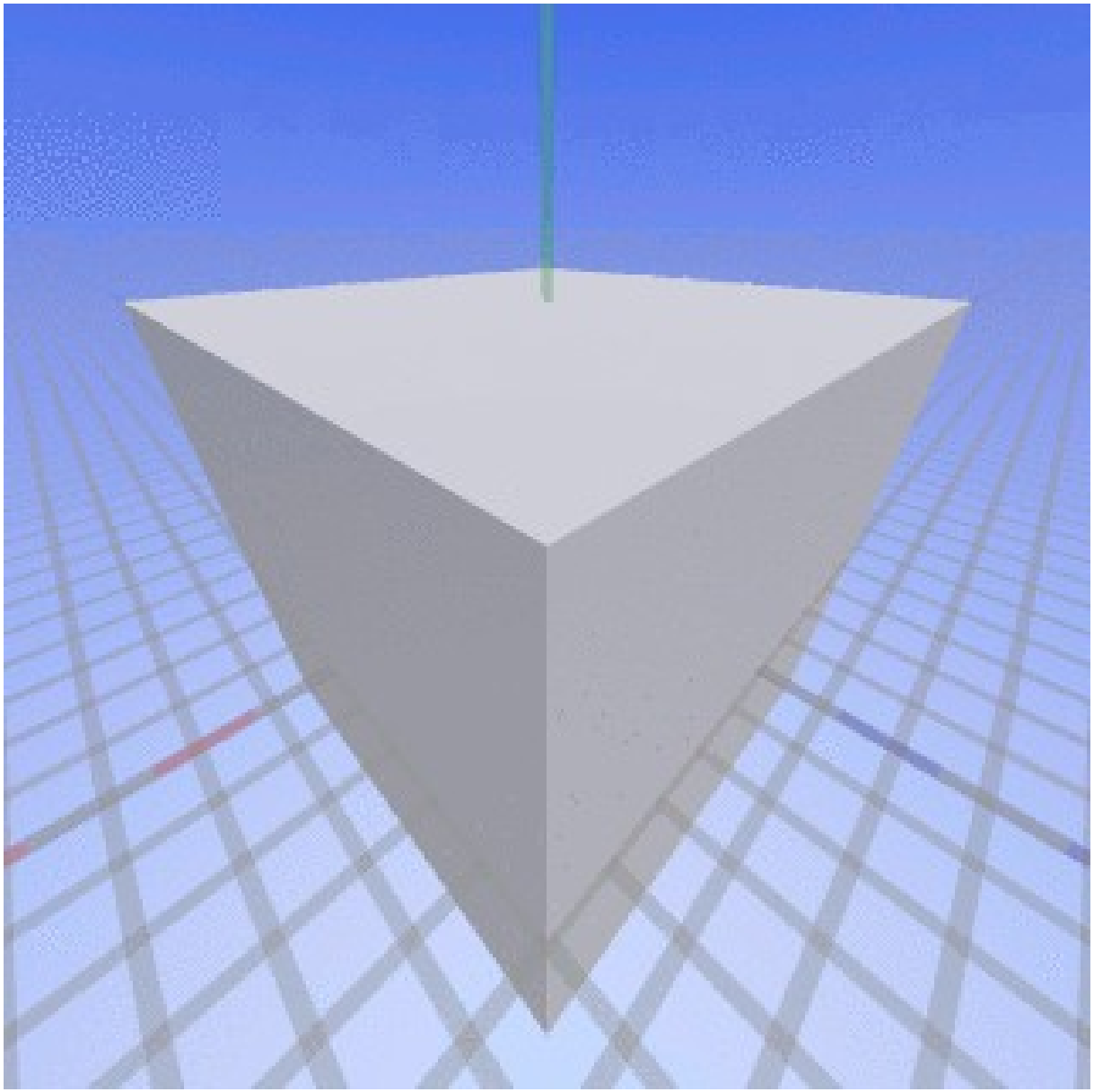


3D is bad (on 2D displays)

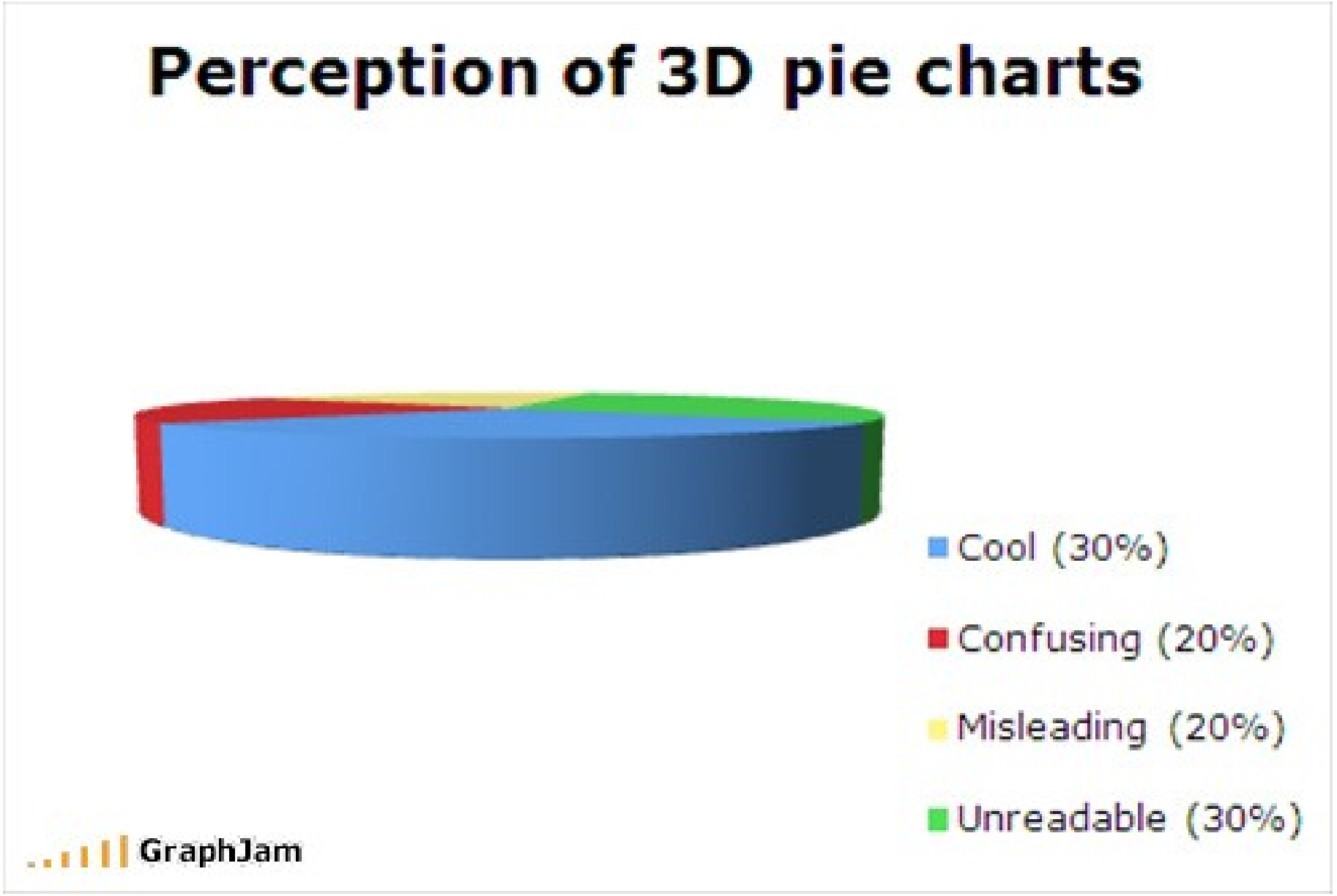



Microsoft Word Features By Version Added
Word 1.0
Word 1.1
Word 2.0
Word 6.0
Word 95
Word 97
Word 2000
Word 2002
Word 2003

Perspective Distortion

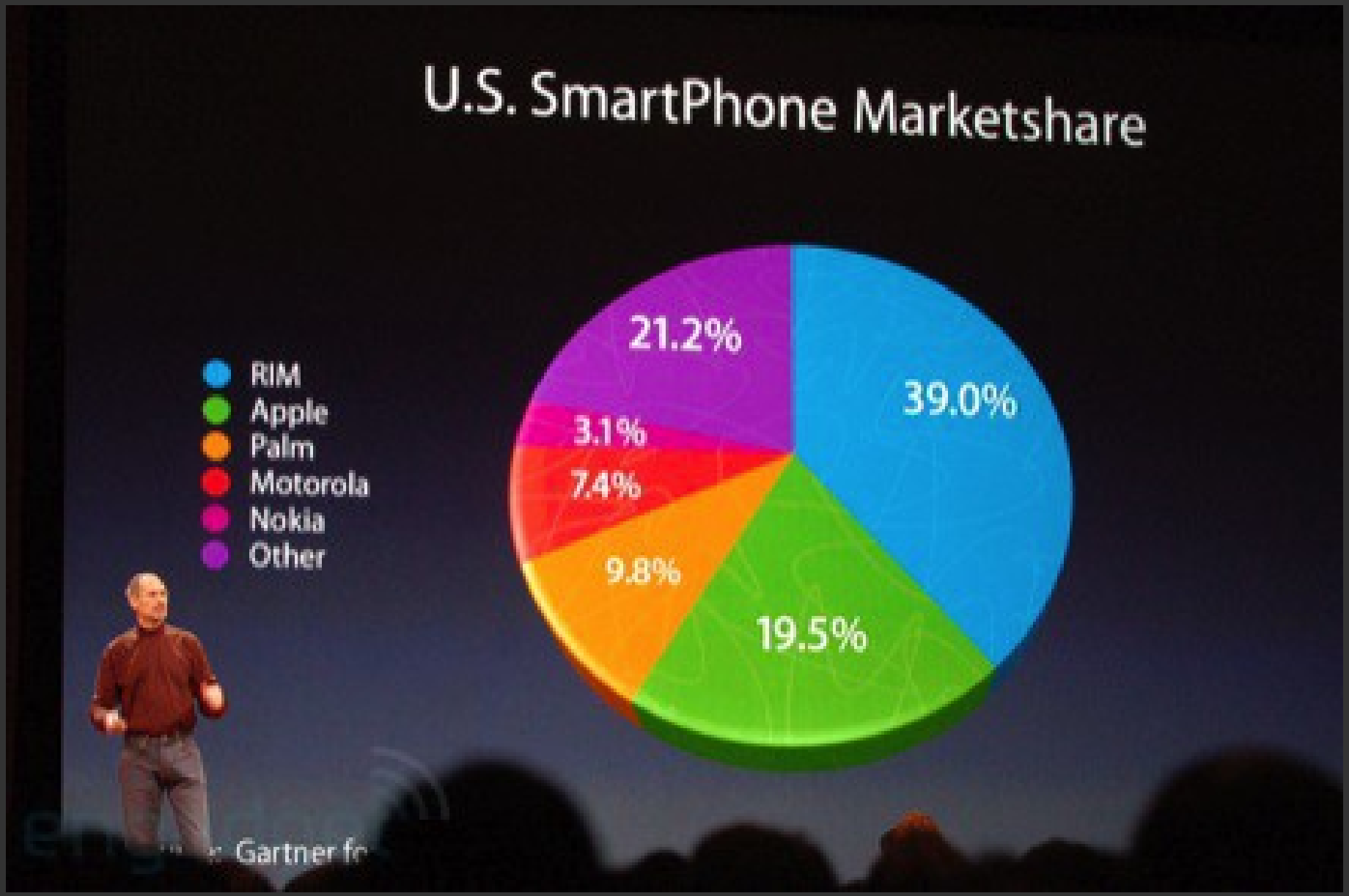


Perception vs Perspective



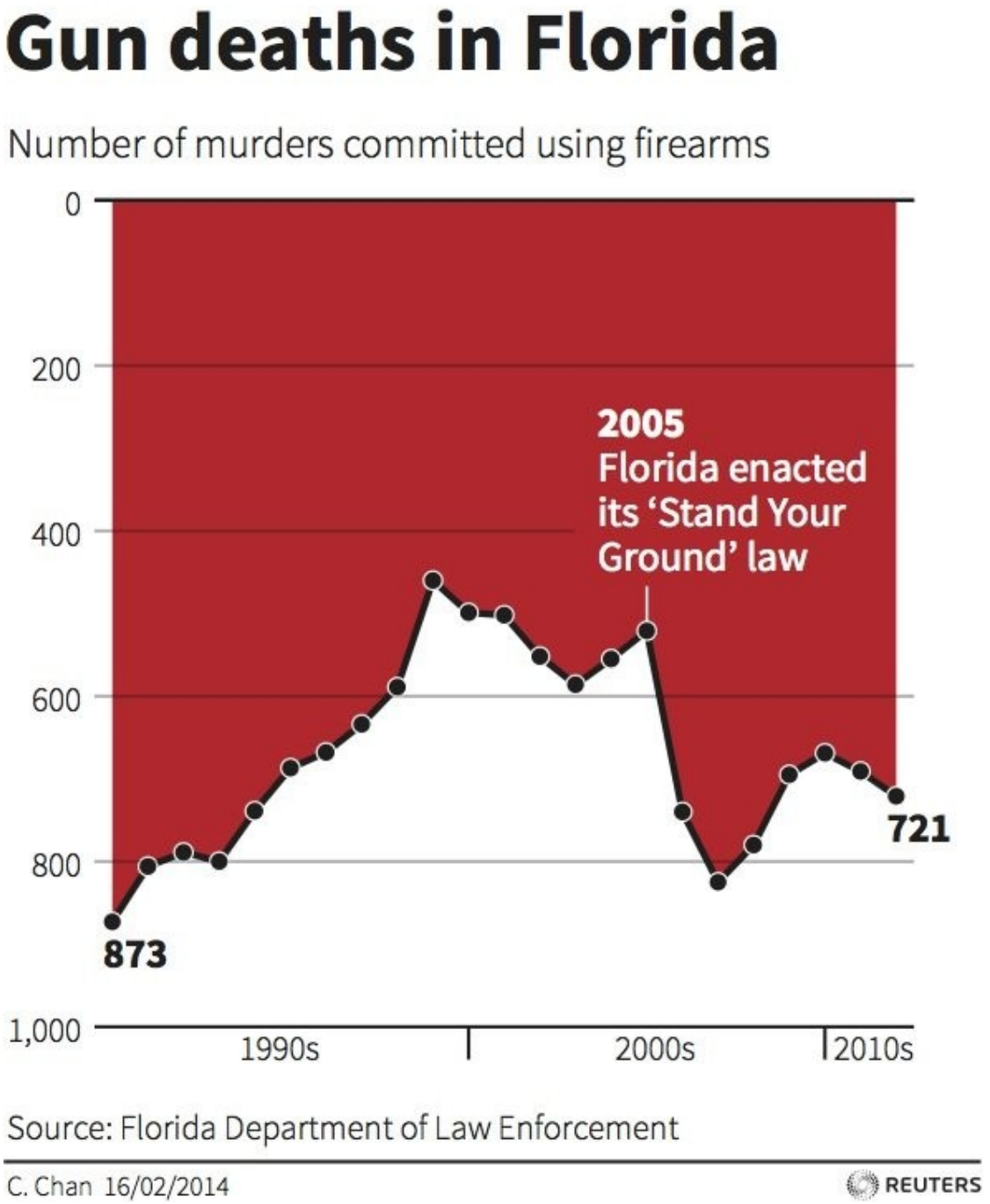
Guidance

Chart Junk – 3d pies are a great way to deceive



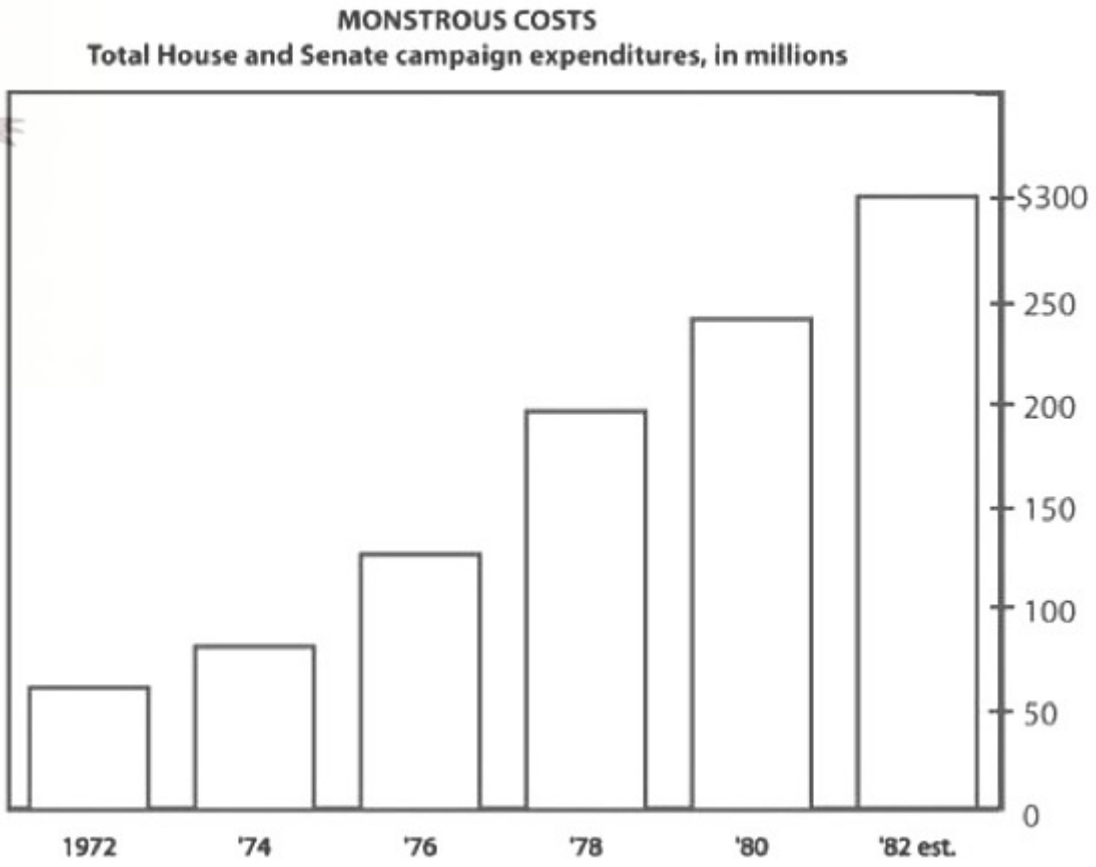
2008 Macworld Expo via Engadget

Chart Junk – you can lie with line charts too



Florida Dept of Law Enforcement via Reuters

Chart Junk – improves memorability



Bateman et al (2010) Useful Junk? via eagereyes.org

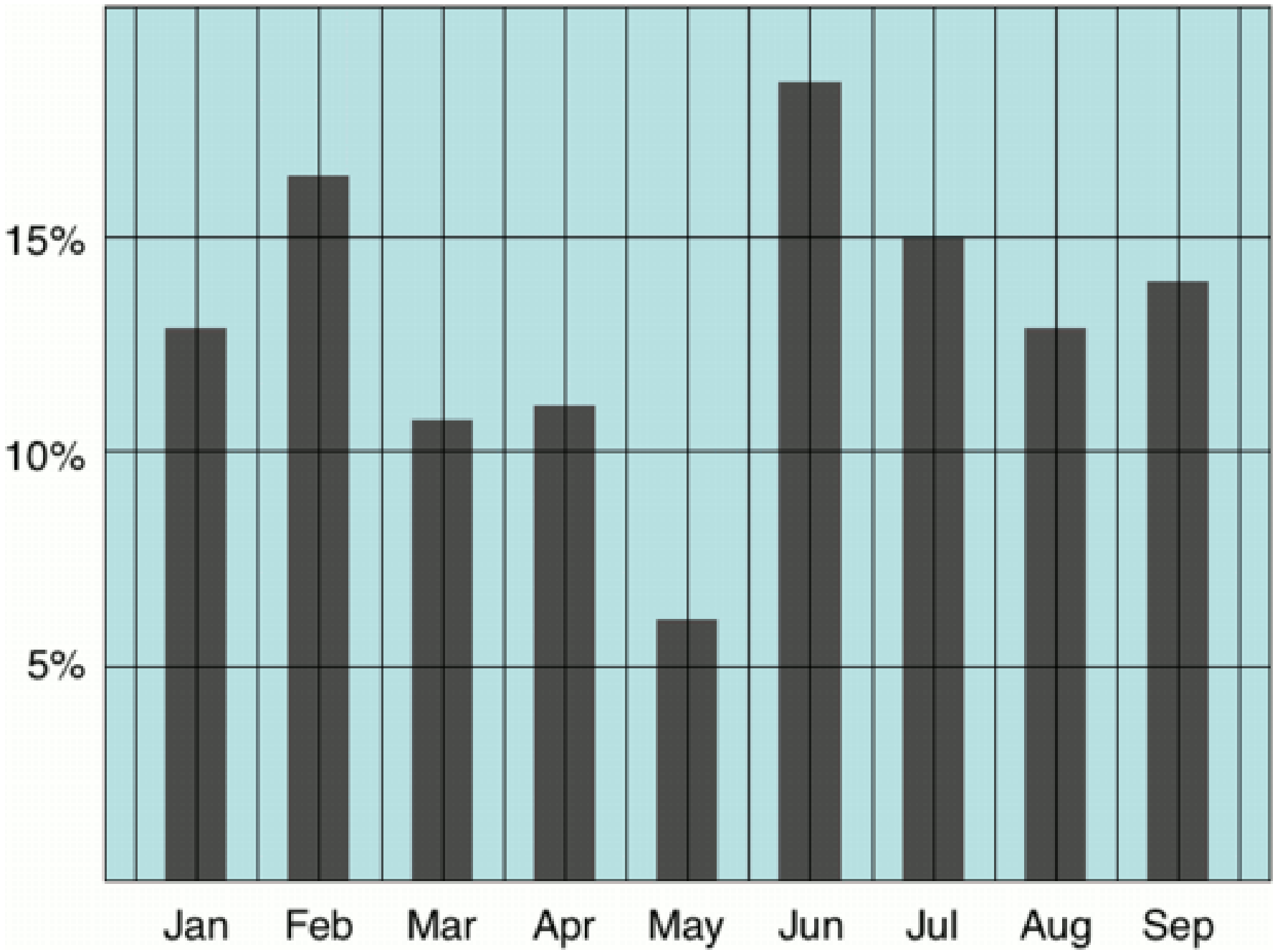
Data-Ink Ratio

$$\text{Data-ink Ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

$$= \text{proportion of a graphic's ink devoted to the non-redundant display of data-information}$$

$$= 1 - \text{proportion of graphic that can be erased}$$

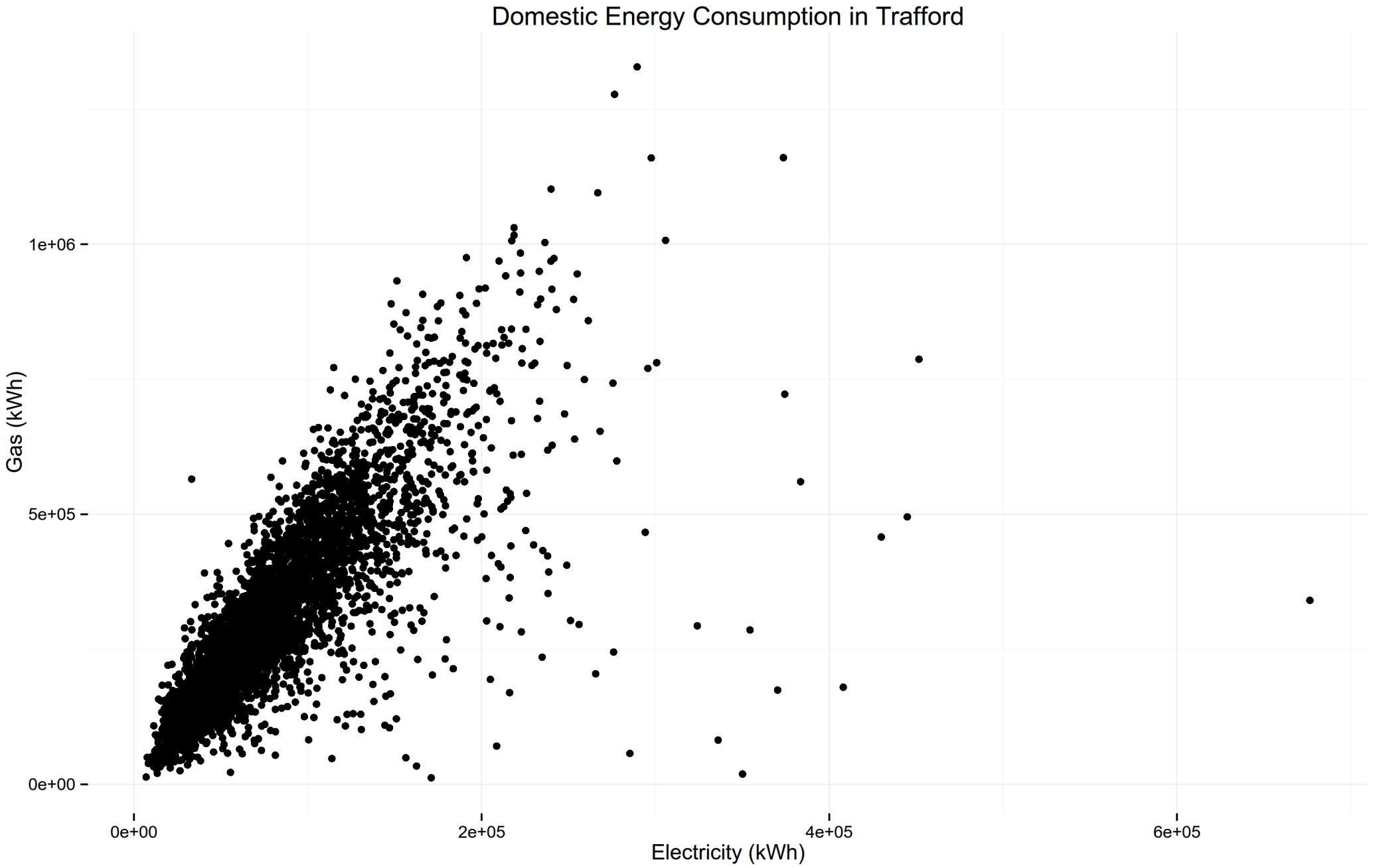
Data-Ink Ratio - Example



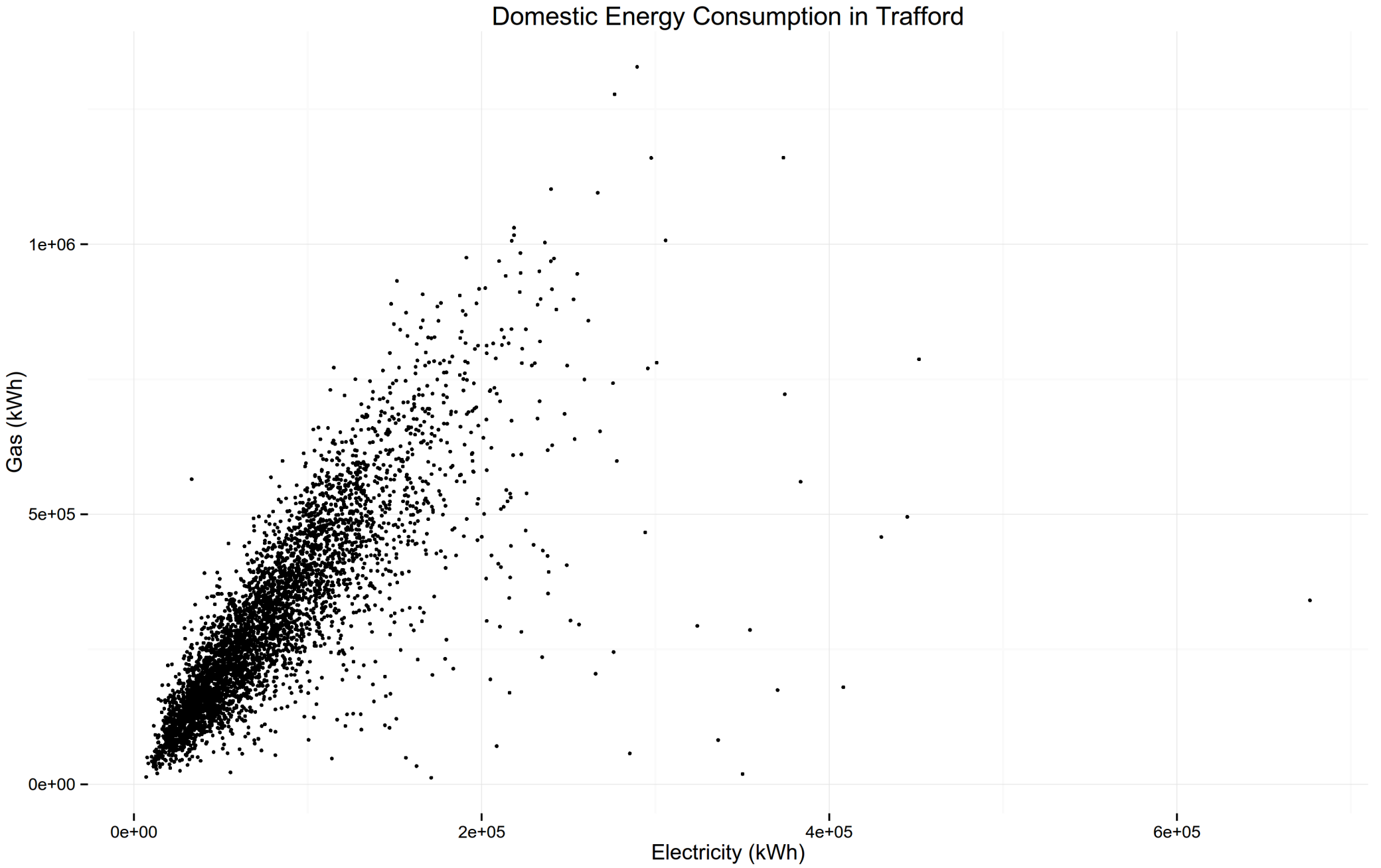
Data-Ink Ratio

http://darkhorseanalytics.com/blog/data-looks-better-naked/

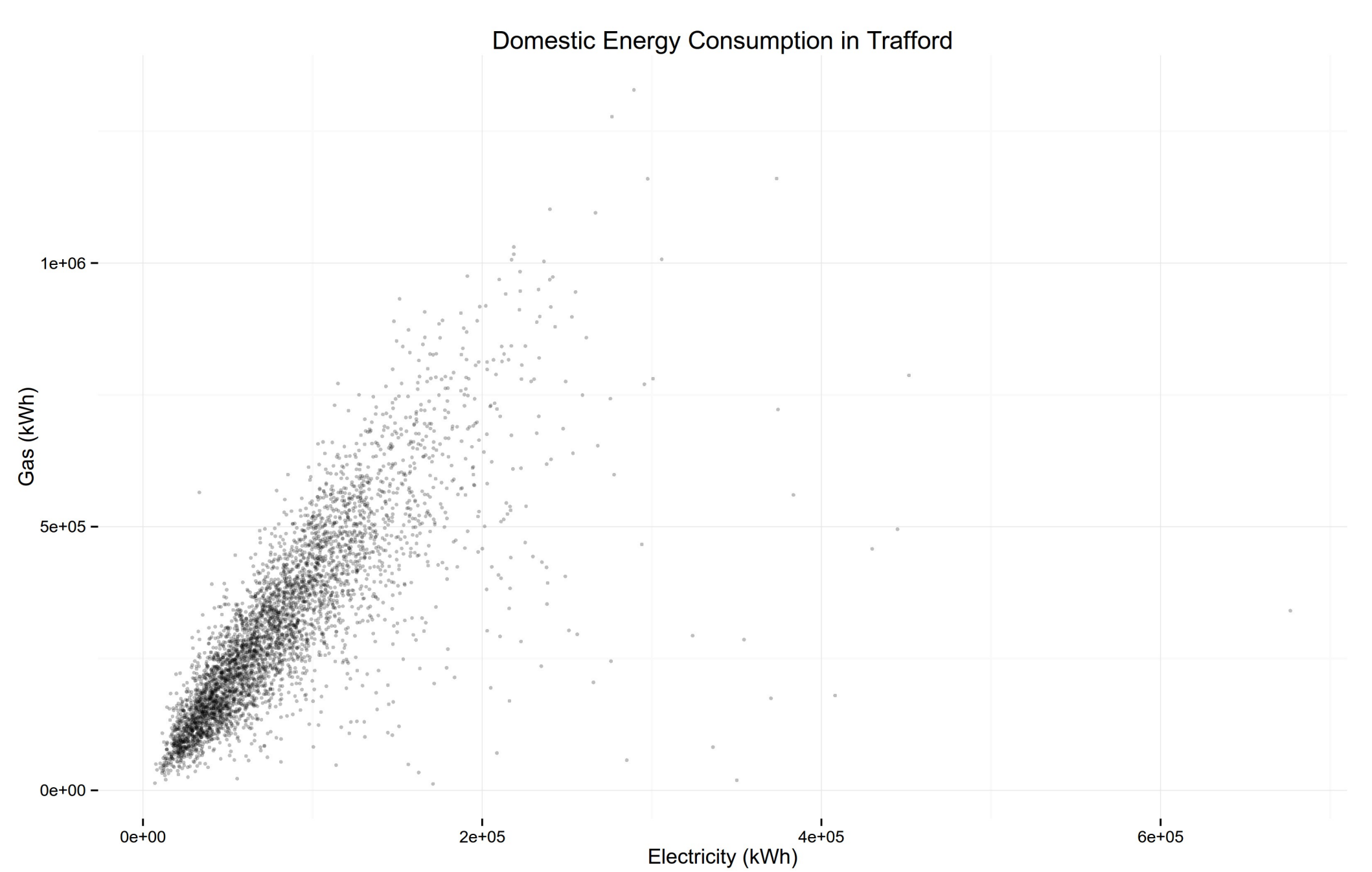
Over-plotting



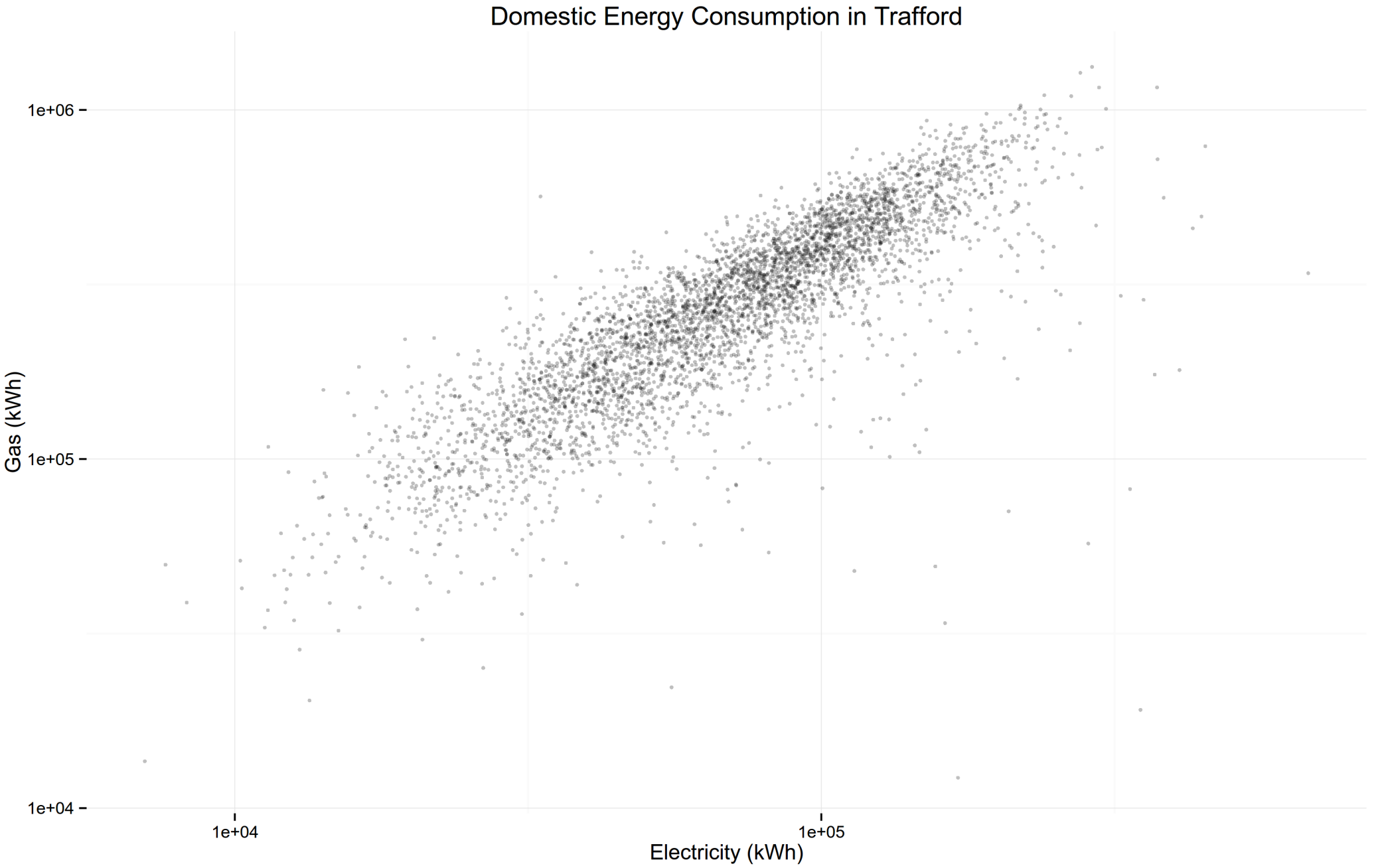
Over-plotting - smaller



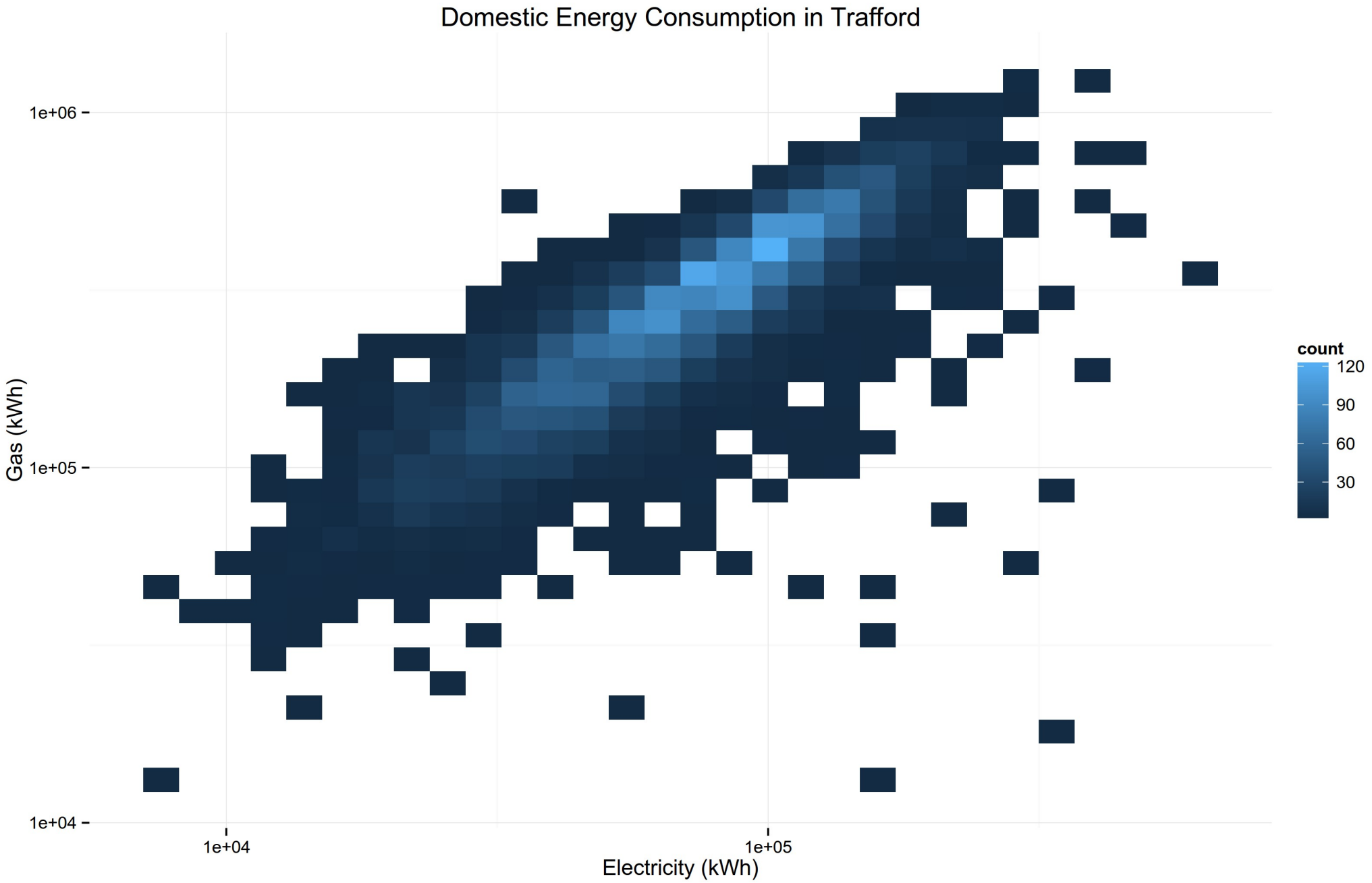
Over-plotting - transparency



Over-plotting – logarithmic scale



Over-plotting – binning



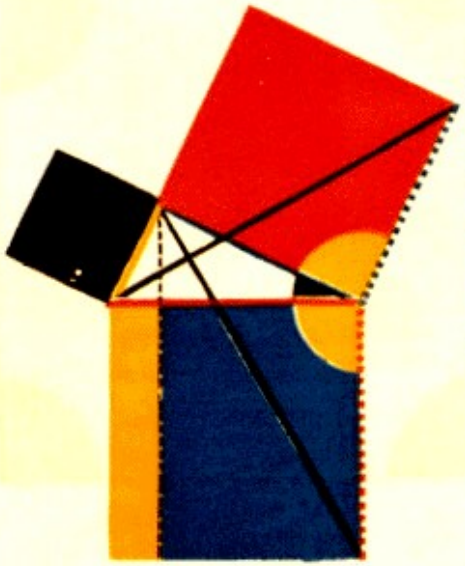
1 National Sector Trends

The global financial crisis of 2008 has gradually lead to recession in the real economy. The affect has varied by sector. The Annual Population Survey shows that total employment has grown over the past five years, peaking in 2007/8  . Growth has continued in Agriculture  , Construction  , Distribution & Hospitality  , Energy & Water  and Public Sector  . Employment has begun to fall in Business Services  , Manufacturing  , Other  and Transport & Communications .

2 Local Sector Mix

In comparison with the National economy, Manchester has more employment in Manufacturing, Distribution & Hospitality, Transport & Communications, Business Services and Other and less in Agriculture, Energy & Water, Construction and Public Sector.

Figure 1: Location Quotients



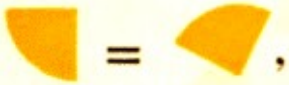
IN *a right angled triangle*  *the ſquare on the hypotenuſe* —— *is equal to the ſum of the ſquares of the ſides,* (—— *and* ——).

On ——, —— and ——

deſcribe ſquares, (pr. 46.)

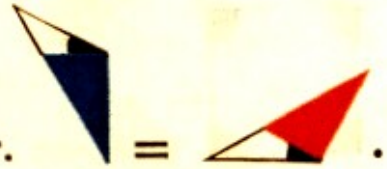
Draw ····· || - - - - - (pr. 31.)

alſo draw —— and ——.

 ,

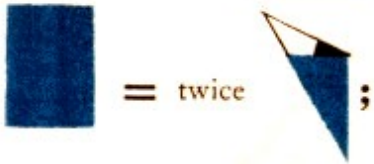
To each add  ∴  ,

—— = - - - - - and —— = ·····;

∴  .

Again, becauſe —— || ·····

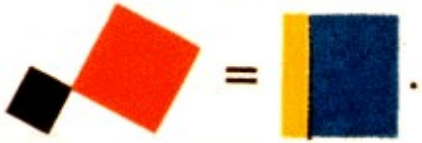
 ,

and  ;

∴  .

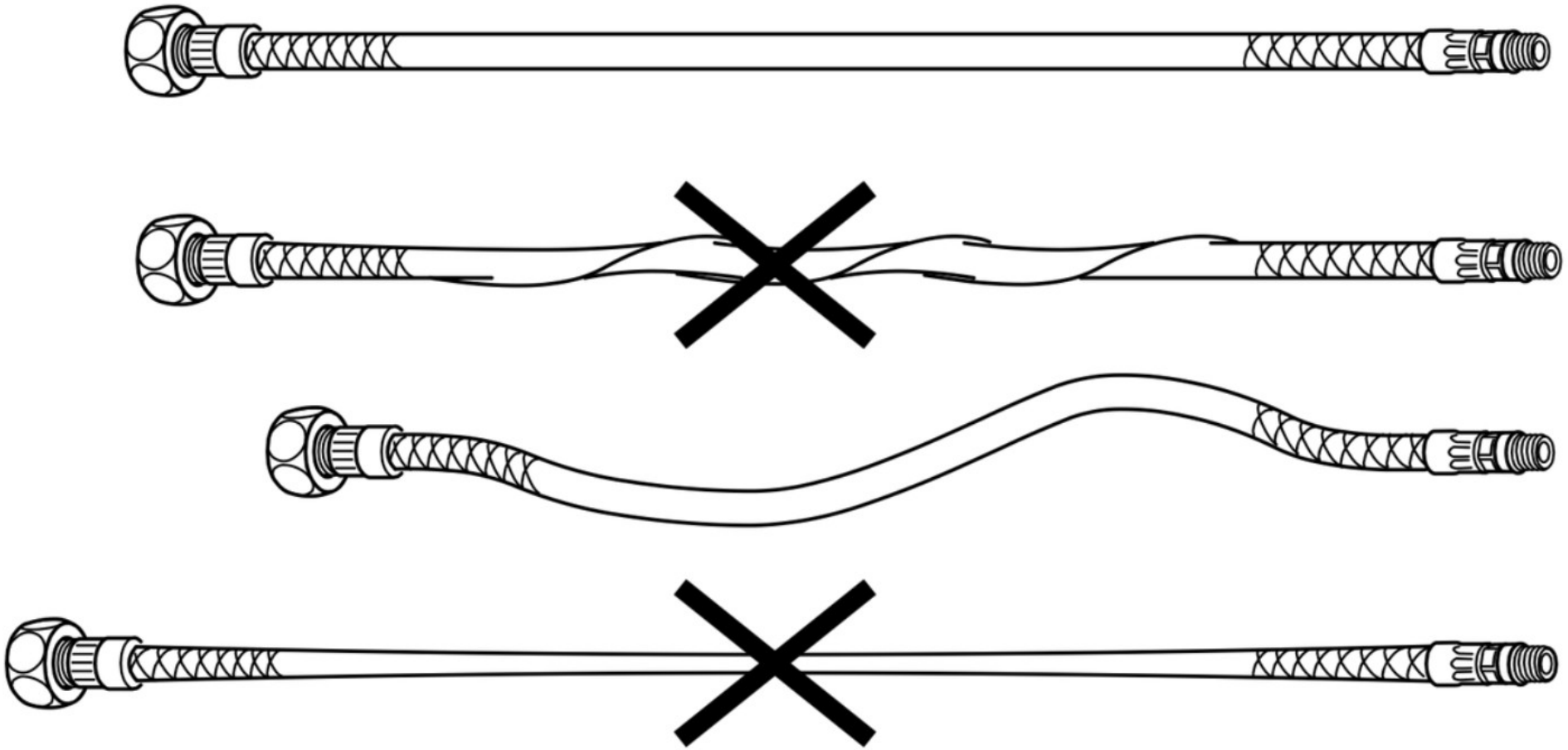
In the ſame manner it may be ſhown

that  ;

hence  .

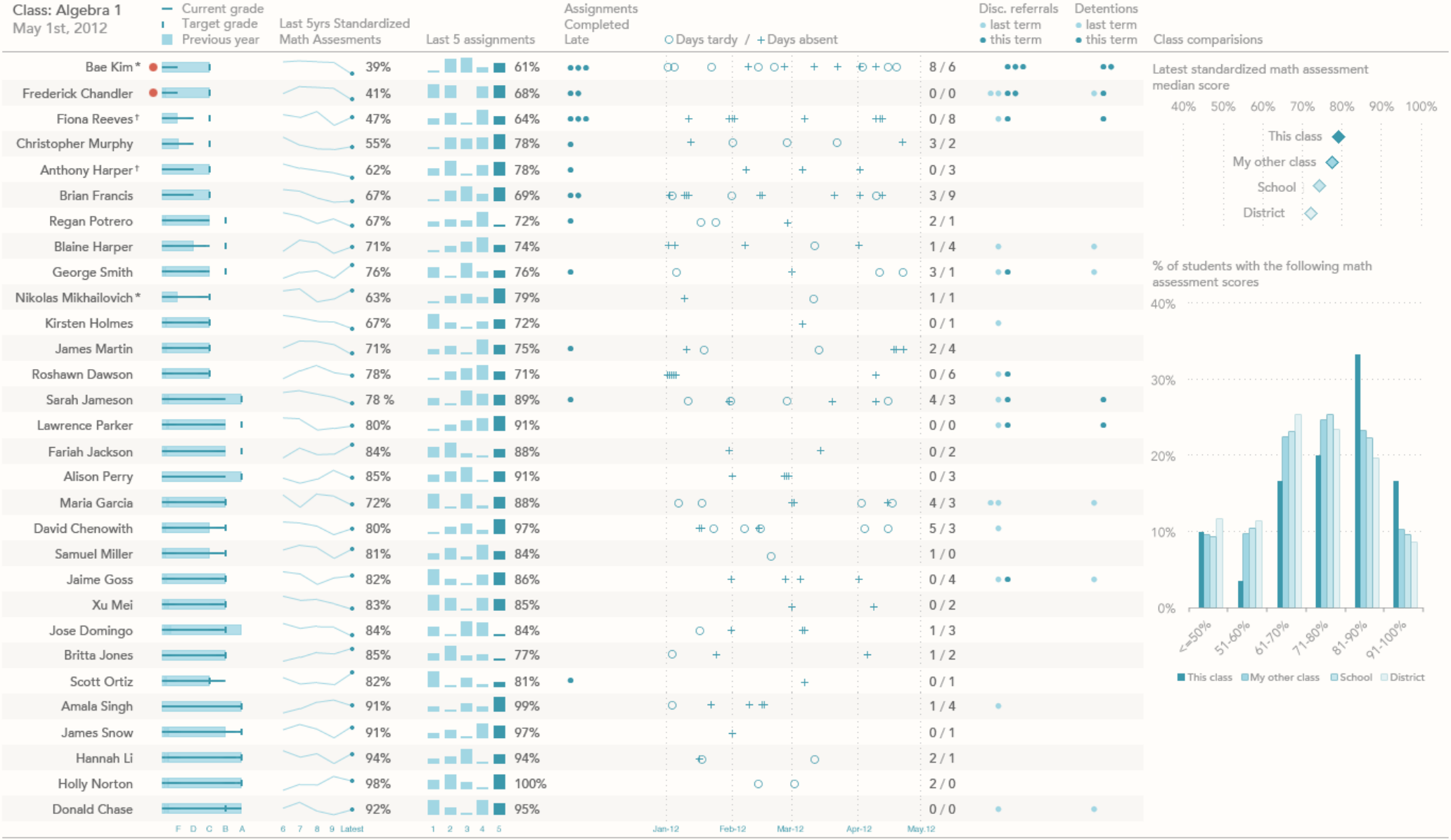
Q. E. D.

Small Multiples



IKEA discovered via Tufte twitter

Small Multiples



* No english language proficiency † Special education

Note: Assessment and assignment scores are being expressed as the percentage of points that were earned out of the total points possible.

Stop Making Pie Charts!

An opinionated guide to the craft of data visualisation

Robin Gower
Data Visualisation Berlin
29.03.16

infonomics.ltd.uk
@robsteranium