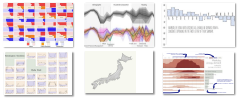
Data Visualization

GEOG 3091 Geocomputation

Roger Beecham
www.roger-beecham.com



www.gicentre.net

world-leading research group in geographic data visualization

MULTIPLE VIEWS

blog series by top infovis academics and practitioners

VIS

***the** conference at which data visualization software, guidelines and best practice is published*

dataset 1		dataset 2		dataset 3		dataset 4	
x1	y1	x2	y2	x3	y3	x4	y4
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

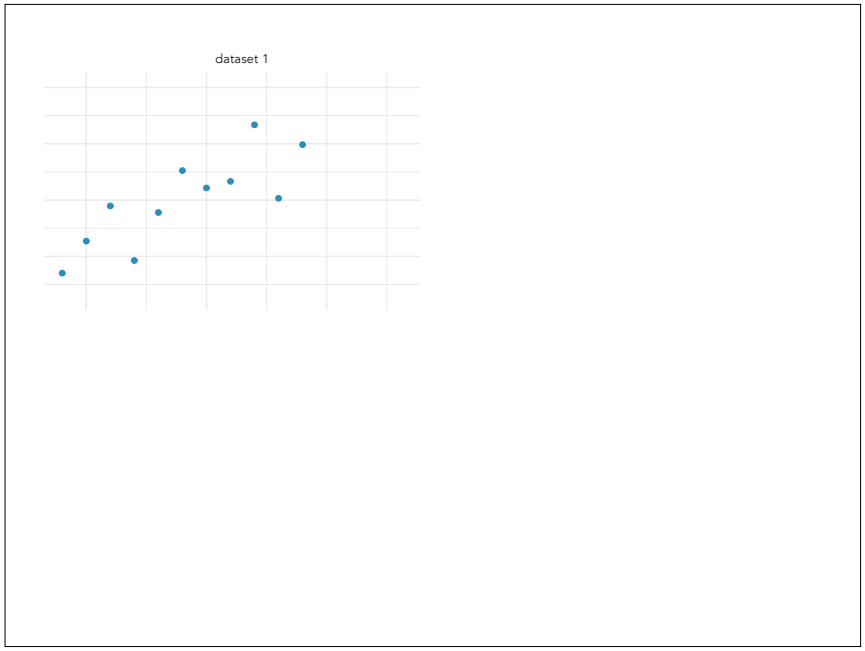
dataset 1		dataset 2		dataset 3		dataset 4	
x1	y1	x2	y2	x3	y3	x4	y4
10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89

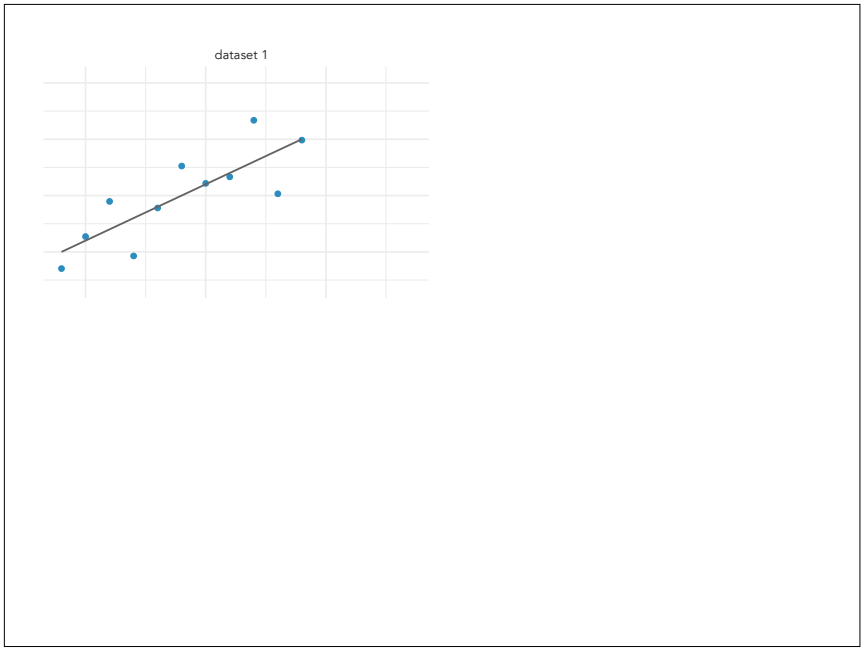
	dataset 1		dataset 2		dataset 3		dataset 4	
	x1	y1	x2	y2	x3	y3	x4	y4
	10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
	8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
	13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
	9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
	11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
	14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
	6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
	4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
	12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
	7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
	5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50

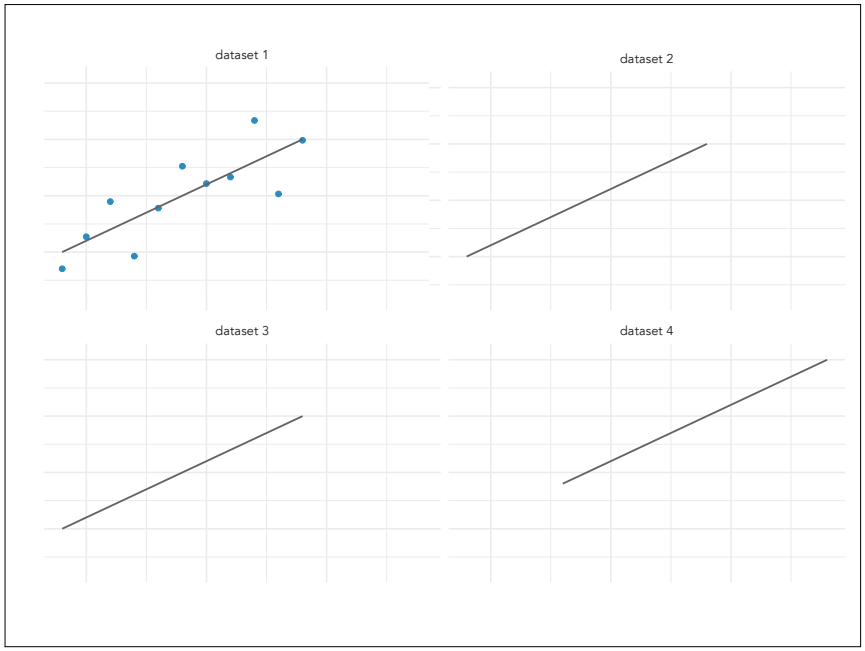
	dataset 1		dataset 2		dataset 3		dataset 4	
	x1	y1	x2	y2	x3	y3	x4	y4
	10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
	8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
	13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
	9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
	11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
	14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
	6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
	4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
	12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
	7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
	5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
Variance	11.00	4.13	11.00	4.13	11.00	4.12	11.00	4.12

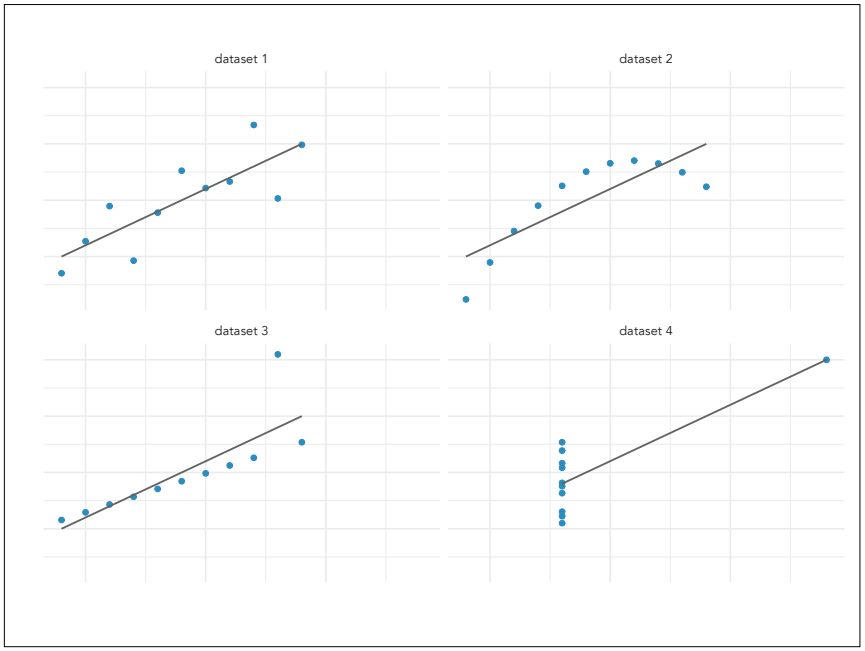
	dataset 1		dataset 2		dataset 3		dataset 4	
	x1	y1	x2	y2	x3	y3	x4	y4
	10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
	8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
	13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
	9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
	11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
	14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
	6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
	4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
	12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
	7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
	5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
Variance	11.00	4.13	11.00	4.13	11.00	4.12	11.00	4.12
Correlation	0.82		0.82		0.82		0.82	

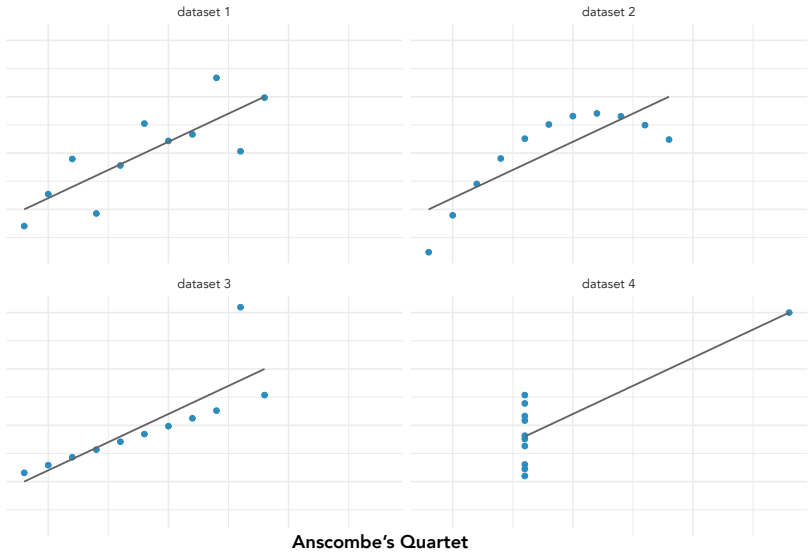
	dataset 1		dataset 2		dataset 3		dataset 4	
	x1	y1	x2	y2	x3	y3	x4	y4
	10.00	8.04	10.00	9.14	10.00	7.46	8.00	6.58
	8.00	6.95	8.00	8.14	8.00	6.77	8.00	5.76
	13.00	7.58	13.00	8.74	13.00	12.74	8.00	7.71
	9.00	8.81	9.00	8.77	9.00	7.11	8.00	8.84
	11.00	8.33	11.00	9.26	11.00	7.81	8.00	8.47
	14.00	9.96	14.00	8.10	14.00	8.84	8.00	7.04
	6.00	7.24	6.00	6.13	6.00	6.08	8.00	5.25
	4.00	4.26	4.00	3.10	4.00	5.39	19.00	12.50
	12.00	10.84	12.00	9.13	12.00	8.15	8.00	5.56
	7.00	4.82	7.00	7.26	7.00	6.42	8.00	7.91
	5.00	5.68	5.00	4.74	5.00	5.73	8.00	6.89
Mean	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
Variance	11.00	4.13	11.00	4.13	11.00	4.12	11.00	4.12
Correlation	0.82		0.82		0.82		0.82	


dataset 1

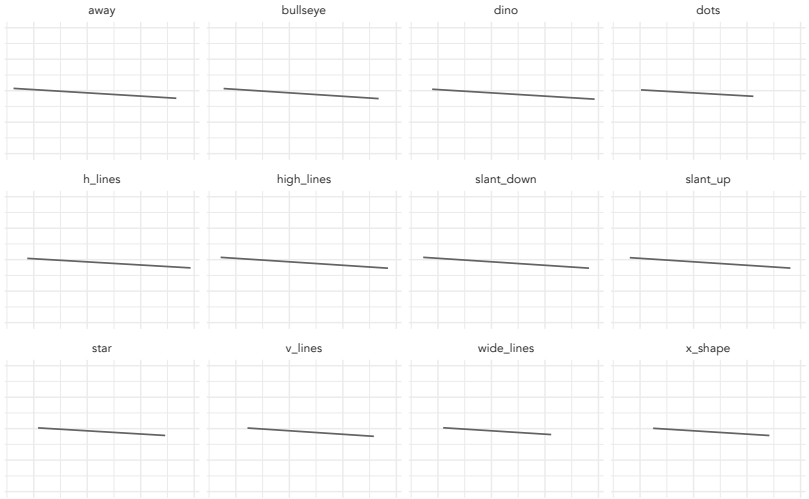

dataset 1

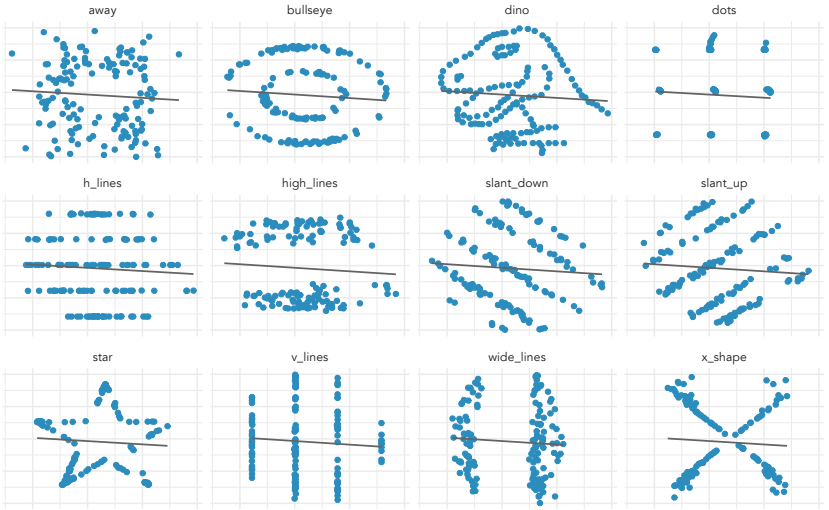

dataset 1
dataset 2
dataset 3
dataset 4


dataset 1
dataset 2
dataset 3
dataset 4

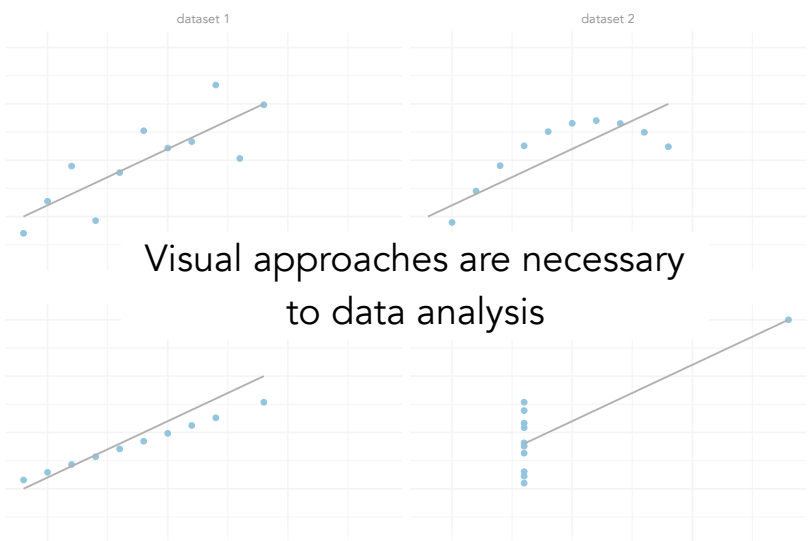


Anscombe's Quartet
Francis Anscombe, 1973





Matejka and Fitzmaurice, 2017



Anscombe's Quartet
Francis Anscombe, 1973

Visual approaches are necessary to **modern** data analysis

Visual approaches are necessary to **modern** data analysis

Visualization is fundamental to meeting the unprecedented challenges and exploiting the wonderful opportunities of the ever-expanding deluge of data confronting virtually every field

Jim Hollan, UC San Diego

Visual approaches are necessary to **modern** data analysis

Visualization is fundamental to meeting the unprecedented challenges and exploiting the wonderful opportunities of the ever-expanding deluge of data confronting virtually every field

Jim Hollan, UC San Diego

Nowadays visualization is one of the key ways, in which scientific discoveries and advances in collective understanding are made.

Therefore, those scientists, researchers, data explorers, and analysts who rely in their everyday work on producing visualizations and generating images, should have some acquaintance with the most rudimentary ABC of good design practice.

Marek Kultys, Independent Designer

Journeys

memberID	isMember	oTime	dTime	oStati	dStati	bikeID
INTEGER	BOOLEAN	DATETIME	DATETIME	INTEGE	INTEGE	INTEGER
#####	1	30/07/10 03:35	30/07/10 03:41	308	308	1396
#####	1	30/07/10 04:05	30/07/10 04:21	290	286	2017
#####	1	30/07/10 04:22	30/07/10 04:42	81	174	738
#####	1	30/07/10 04:37	30/07/10 04:47	14	14	2992
#####	1	30/07/10 04:39	01/08/10 04:10	169	169	2786
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.

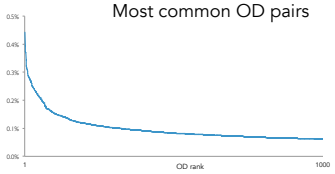
~ 60 million by 2017

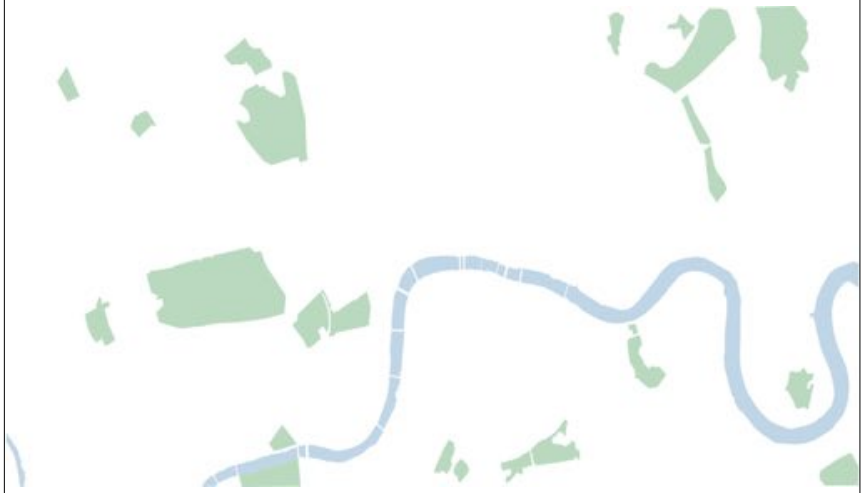
Summary statistics

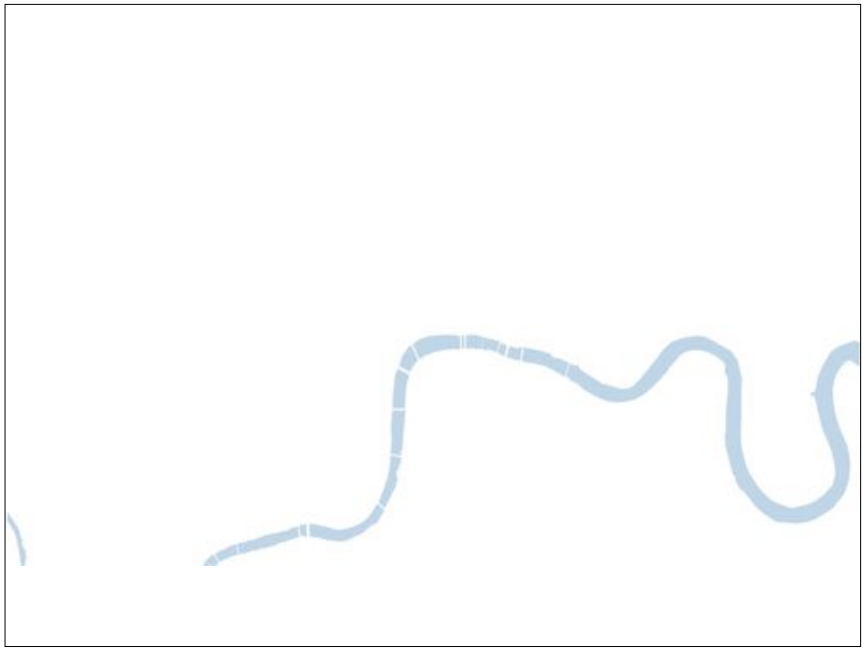
Summary statistics

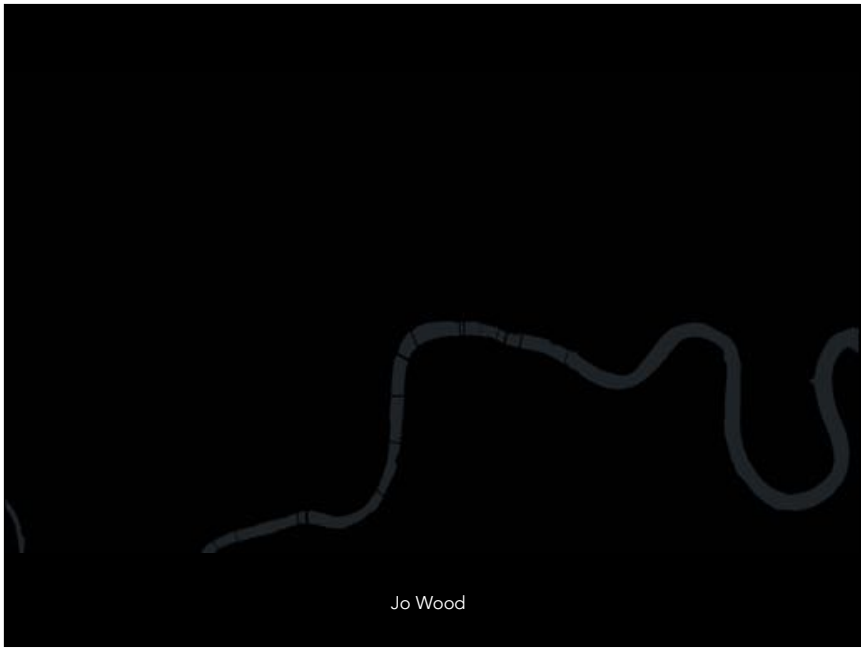
~ 10 million in 2017
Modal day of week : Wed-Thu
Modal month of year : Jun-Jul
Average Travel Time : 8 mins

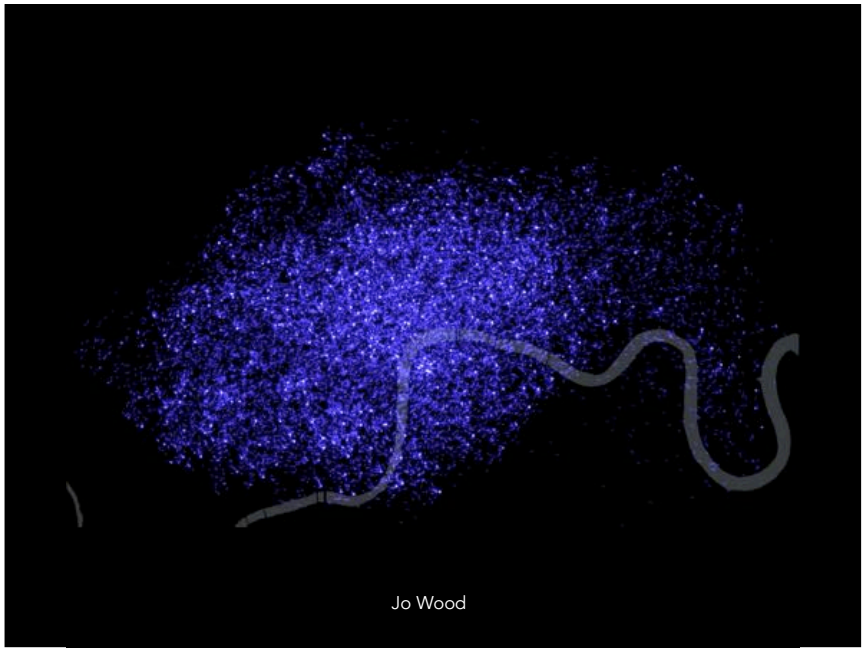
Most common OD pairs

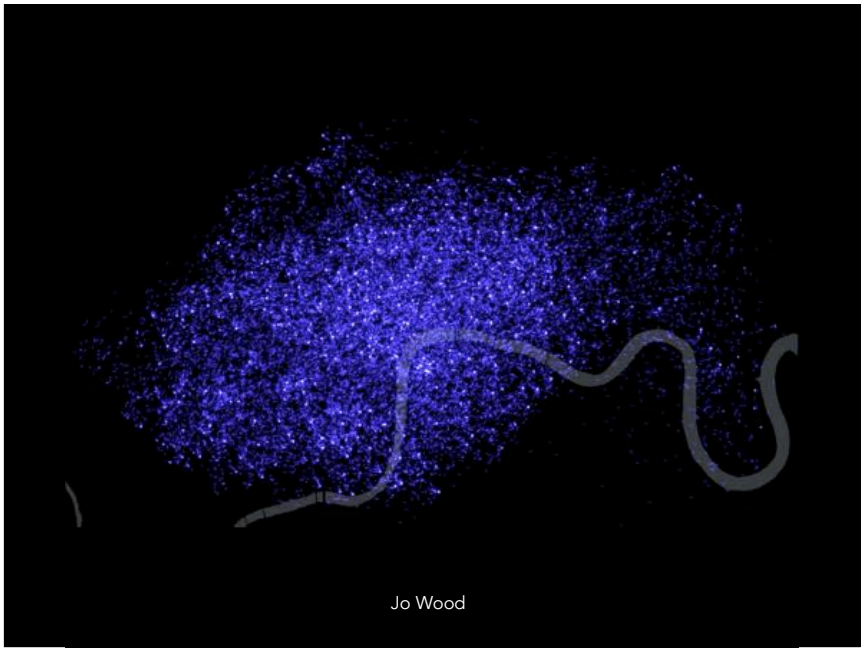







Jo Wood


Jo Wood


Jo Wood

Beecham and Wood, 2014

Session Outcomes

Appreciate core principles of data visualization design

Employ these principles when critiquing data graphics

Create effective data visualizations using software built on these principles

Characteristics of effective data graphics

Tufte, Edward R. (2001), The Visual Display of Quantitative Information, Graphics Press, Cheshire, Connecticut

Effective data graphics should

1. Show the data

2. Induce the viewer to think about the substance of the data

3. Avoid distorting what the data have to say

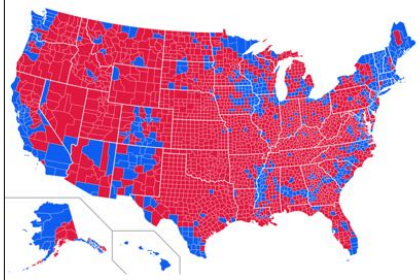
4. Present many numbers in a small space

5. Make large data sets coherent

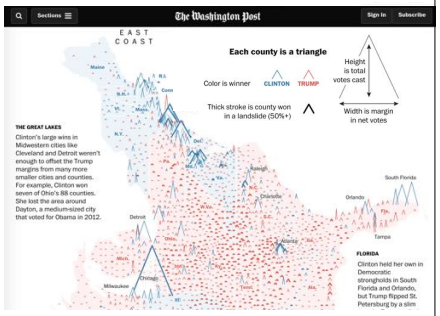
6. Encourage the eye to compare different pieces of data

7. Reveal the data at several levels of detail
from a broad overview to a fine structure

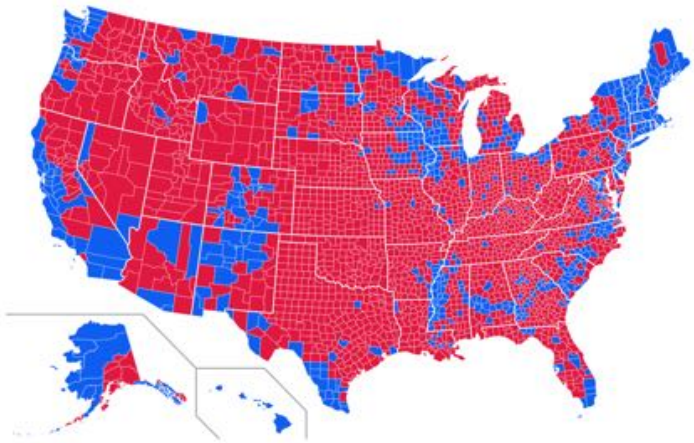
Tufte (2001: 13)



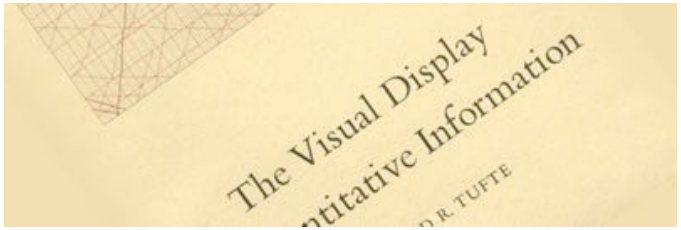
Natalie Schmidt, on Medium



Lazaro Gamio and Dan Keating, Washington Post



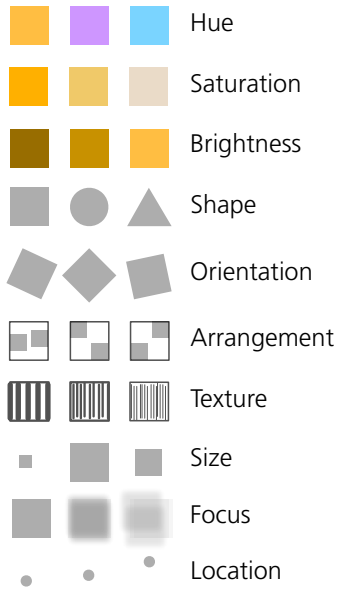
Natalie Schmidt, on Medium



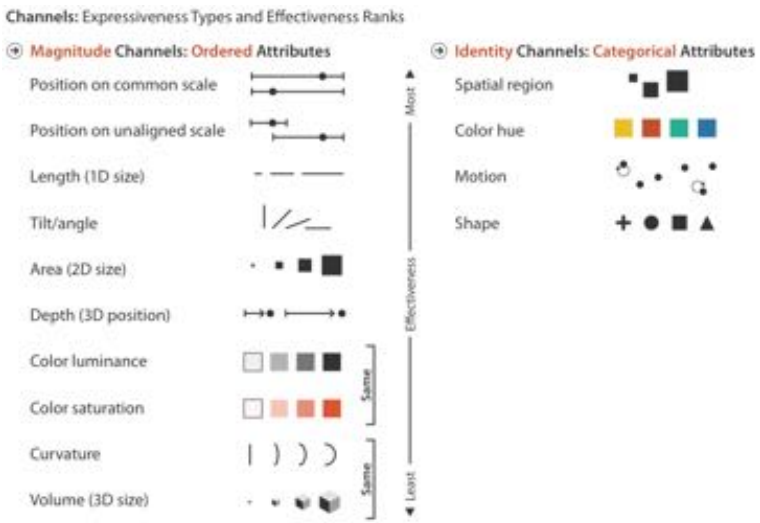
Data graphics visually display measured quantities by means of the combined use of points, lines, a coordinate system, numbers, symbols, words, shading, and color.

Tufte, 2001

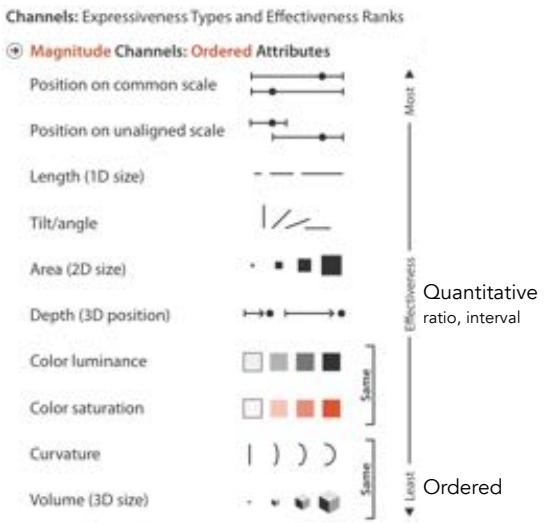
Visual variables | channels



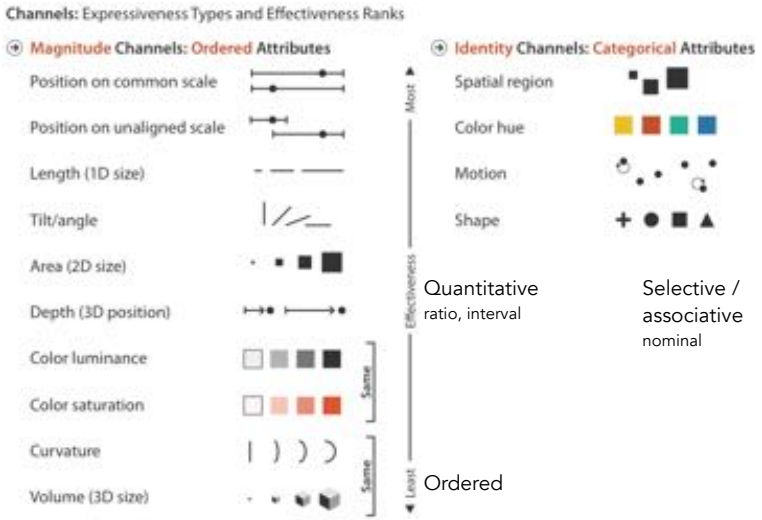
Bertin, 1983



Munzner, 2014
Visualization , Analysis & Design



Munzner, 2014
Visualization , Analysis & Design



Munzner, 2014
Visualization , Analysis & Design

Selective | Associative

Selective | Associative

How many **3**s do you see?

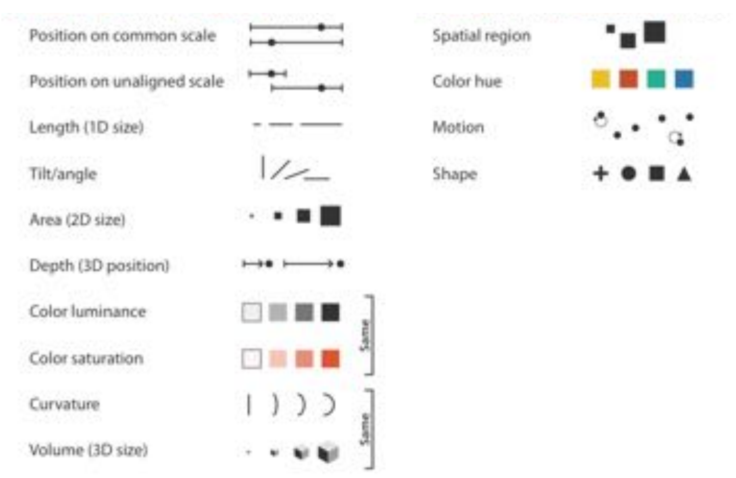
visual channel : none

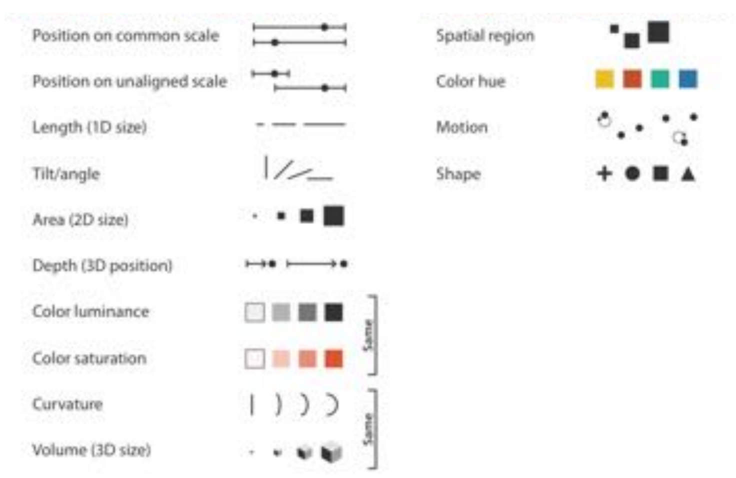
Selective | Associative

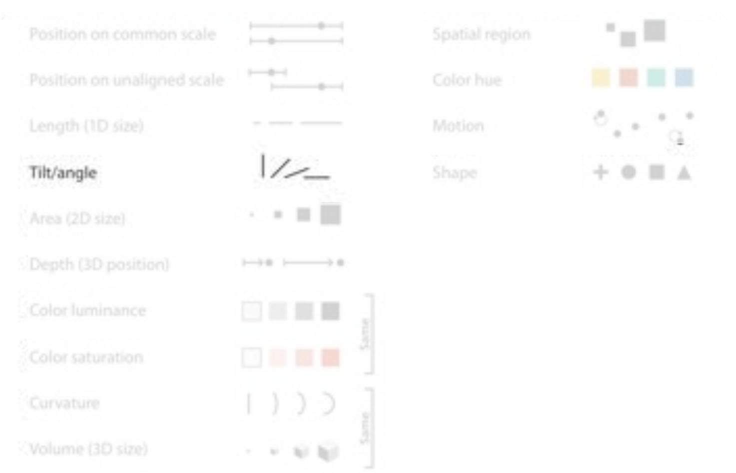
How many **3**s do you see?

37655418
44812212
82453219

visual channel : none





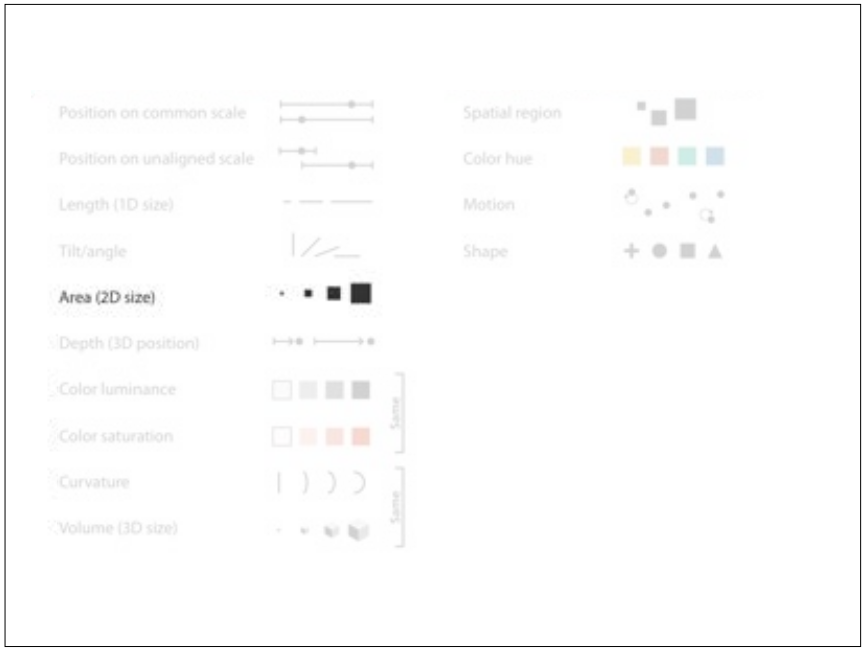


Selective I Associative

How many **3**s do you see?

513 755202122814720404449109462911907883 6
8624417462505921412 389796215700900478196
10790780115475862193 902153 10196179771281
2966015694126969760908126411855467014540

visual channel : tilt



Selective | Associative

How many **3**s do you see?

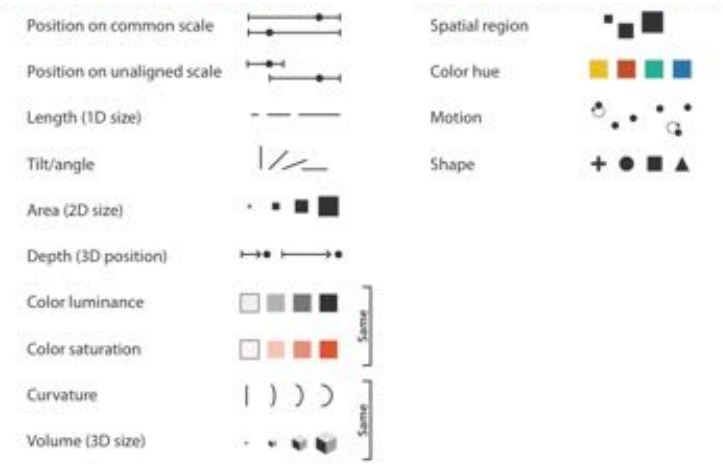
visual channel : 2d area

Selective | Associative

How many **3**s do you see?

1810717433286272787460422968772440752976
5481114296407078115079852599446262850159
1734526230789702074898629792692879216328
7921452697625142910280425596600312888064

visual channel : 2d area



Selective | Associative

How many **3**s do you see?

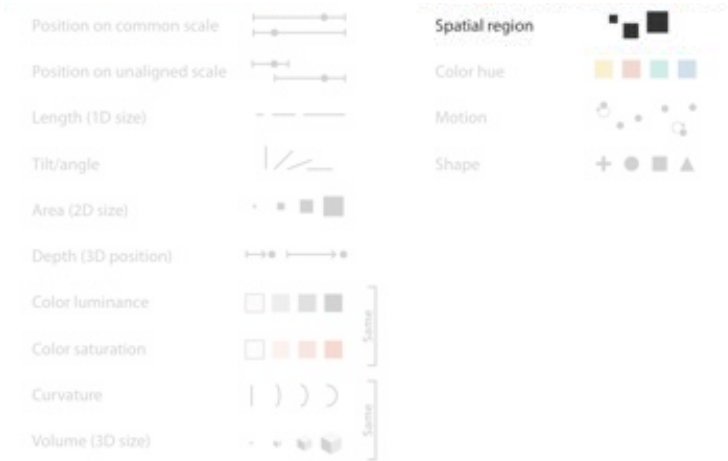
visual channel : colour hue

Selective | Associative

How many **3**s do you see?

610769284896869017670265017617578745634 5
978062632988776446990198470542475234479 8
066449646739720275018242270260444586990 5
882351082925985068197014167114902476009 2

visual channel : colour hue



Selective | Associative

How many **3**s do you see?

visual channel : spatial region

Selective | Associative

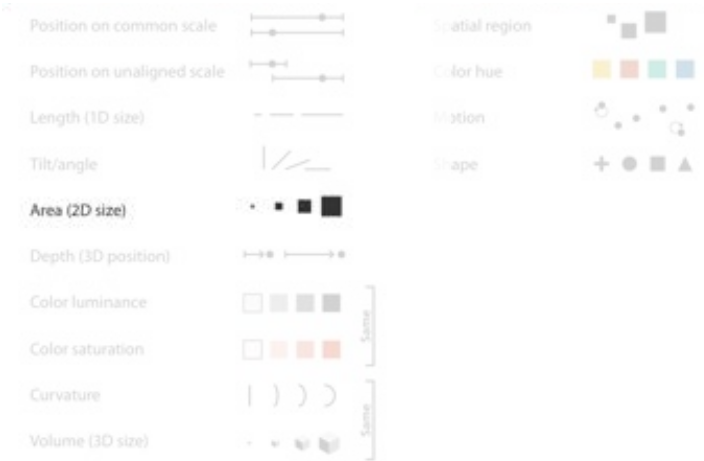
How many **3**s do you see?

33333

1992449260184678298245575745669544099480
5222862564278801424772108748670942271244
5907108580647580921567267712806201969450
2265619091605667699221779518552055 6

visual channel : spatial region

Quantity | Magnitude



Quantity | Magnitude

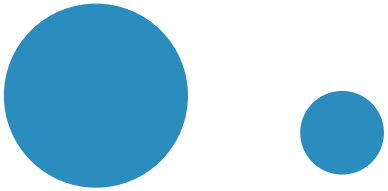
What proportion of the area of the larger circle does the smaller circle take up?

visual channel : 2d area

via Jason Dykes

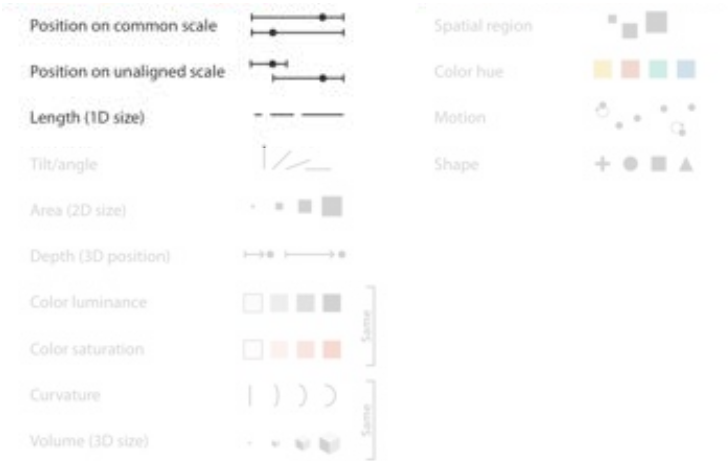
Quantity | Magnitude

What proportion of the area of the larger circle does the smaller circle take up?



visual channel : 2d area

via Jason Dykes



Quantity | Magnitude

What proportion of the longer line does the smaller line account for?

visual channel : 2d area

via Jason Dykes

Quantity | Magnitude

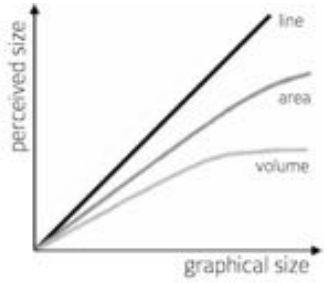
What proportion of the longer line does the smaller line account for?

visual channel : 2d area

via Jason Dykes

Quantity | Magnitude

What proportion of the area of the larger circle does the smaller circle take up?



visual channel : 2d area

Quantity | Magnitude

What proportion of the area of the larger circle does the smaller circle take up?

visual channel : 2d area

via Jason Dykes

Quantity | Magnitude

What proportion of the area of the larger circle does the smaller circle take up?

visual channel : 2d area

via Jason Dykes

Quantity | Magnitude

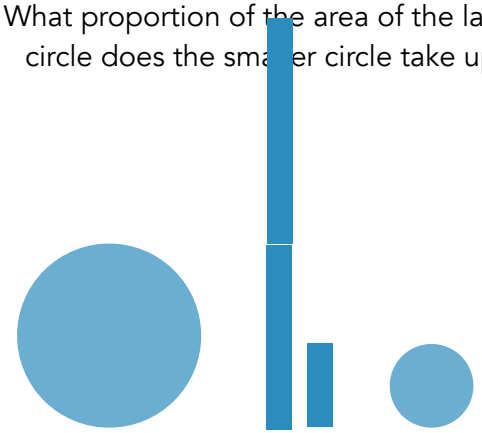
What proportion of the area of the larger circle does the smaller circle take up?

visual channel : 2d area

via Jason Dykes

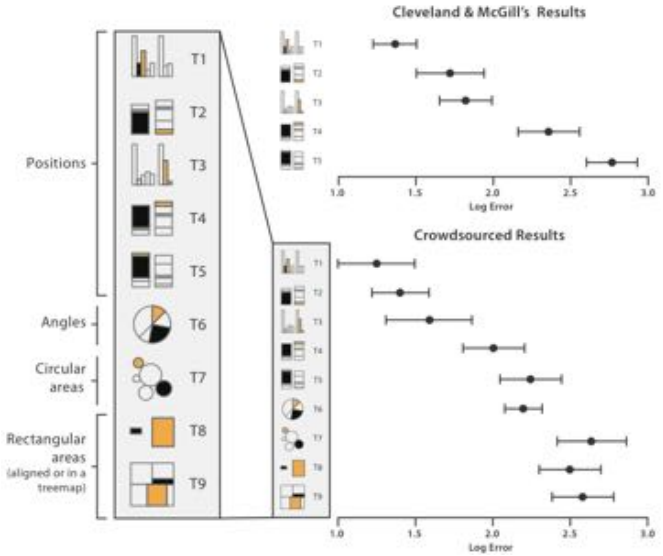
Quantity | Magnitude

What proportion of the area of the larger circle does the smaller circle take up?



visual channel : 2d area

via Jason Dykes

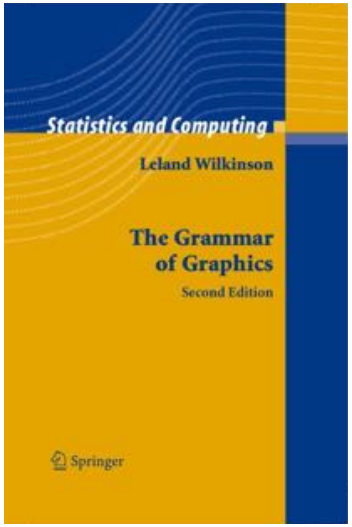


Heer & Bostock 2010


Vega-Lite

Vega





Data

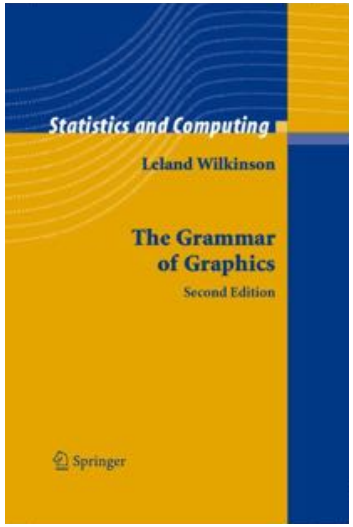
Transformation

Element

Scale

Guide

Coord



Vega-Lite

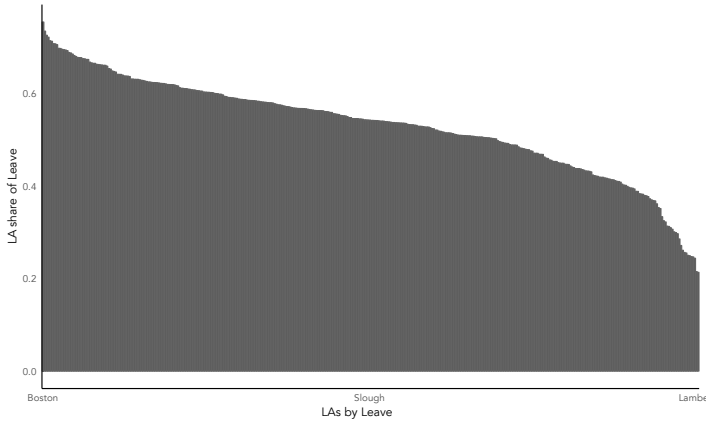
Data	variables you want to represent
Aesthetics	mapping of data to visual channels
Geom	shapes to represent data (point, line, bar)
Facets	split on a (nominal/ordinal) variable to generate small multiples
Statistics	aggregates using statistical models
Coordinates	plotting space you are using
Themes	non-data ink: design with a particular visual fonts, colours and other design elements.

Brexit data: share of leave vote by Local Authority

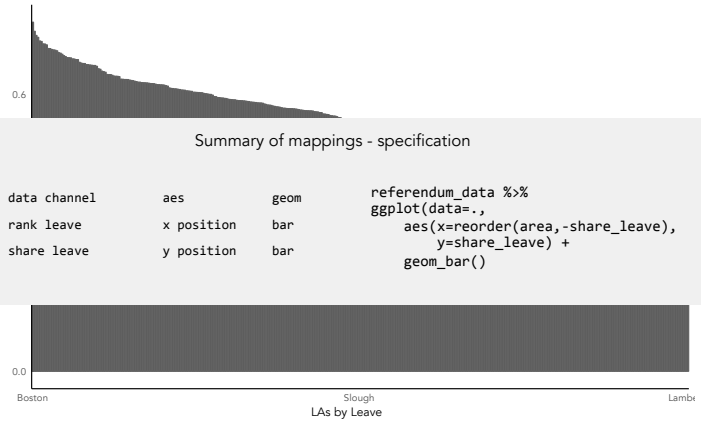
Demographics data: skills levels, occupation and diversity by Local Authority

LAs ordered by share of Leave

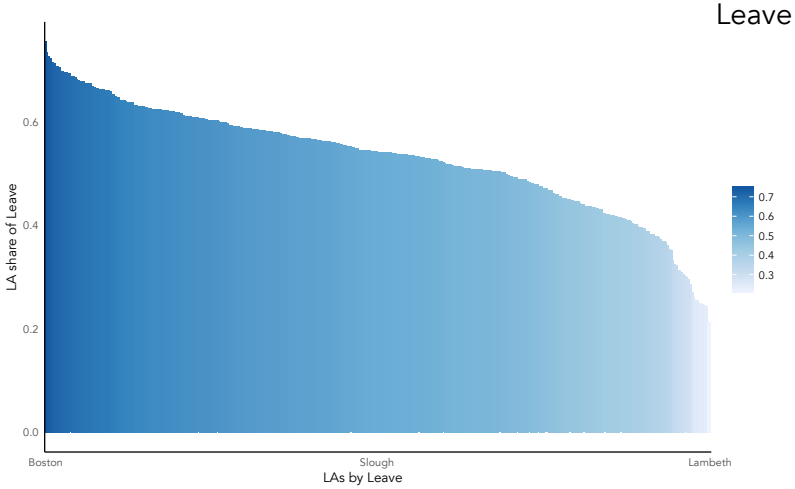
LAs ordered by share of Leave



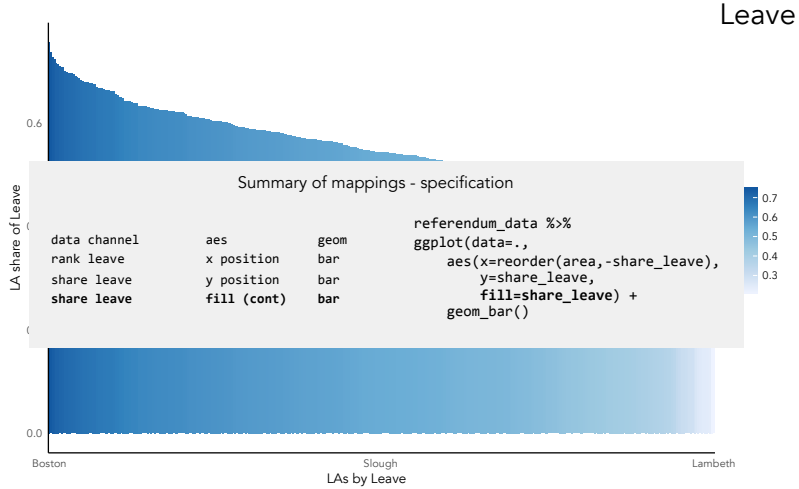
LAs ordered by share of Leave



LAs ordered by share of Leave



LAs ordered by share of Leave



Summary of mappings - specification

data channel	aes	geom
rank leave	x position	bar
share leave	y position	bar
share leave	**fill (cont)**	**bar**

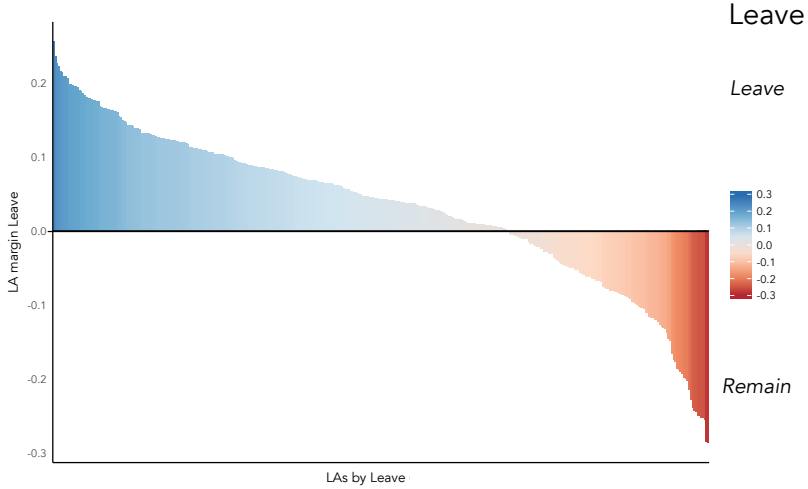
```
referendum_data %>%
ggplot(data=.,
    aes(x=reorder(area,-share_leave),
        y=share_leave,
        fill=share_leave) +
    geom_bar()
```

LAs ordered by share of Leave

Leave

Remain

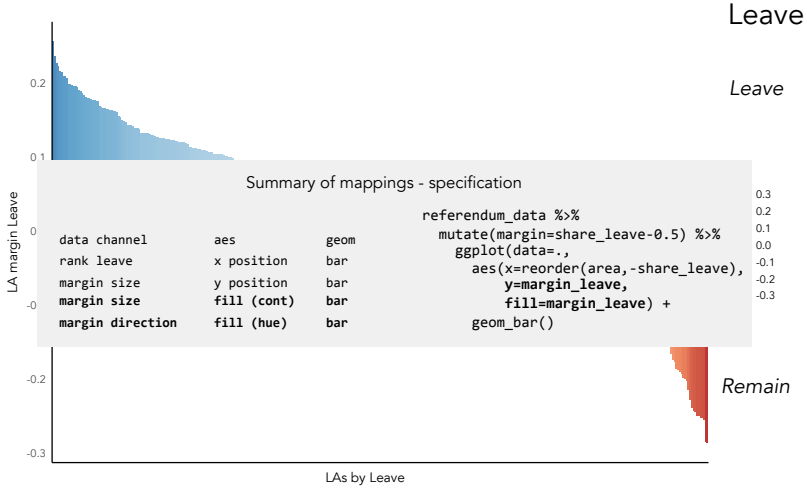
LAs ordered by share of Leave



Leave

Remain

LAs ordered by share of Leave



Leave

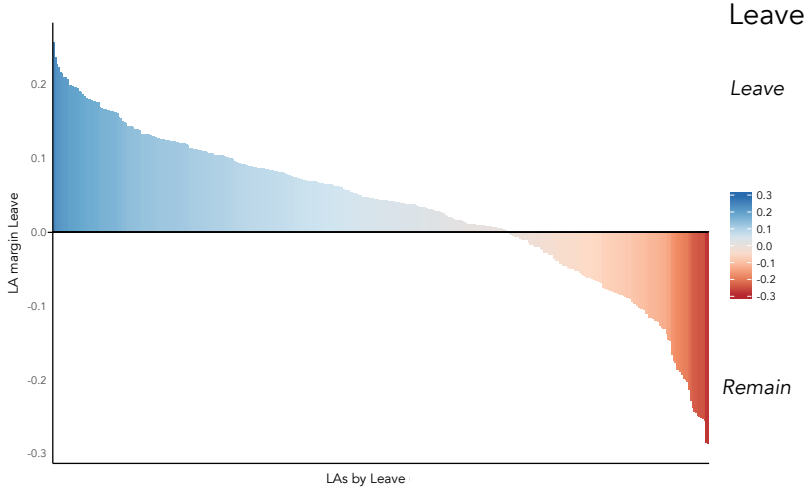
Summary of mappings - specification

data channel	aes	geom
rank leave	x position	bar
margin size	y position	bar
margin size	**fill (cont)**	**bar**
margin direction	**fill (hue)**	**bar**

```
referendum_data %>%
  mutate(margin=share_leave-0.5) %>%
    ggplot(data=.,
      aes(x=reorder(area,-share_leave),
          y=margin_leave,
          fill=margin_leave) +
      geom_bar()
```

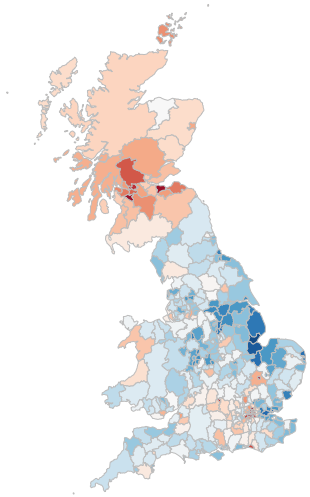
Remain

LAs ordered by share of Leave

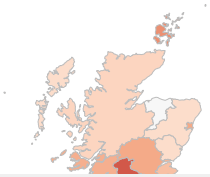


LAs ordered by geospatial position

LAs ordered by geospatial position



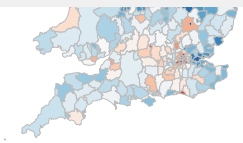
LAs ordered by geospatial position



Summary of mappings - specification

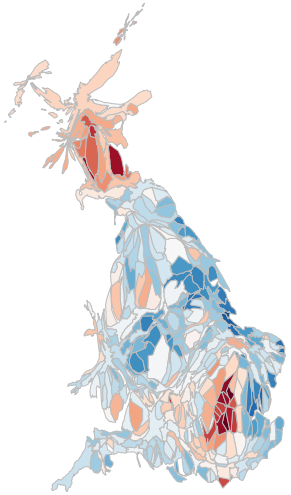
data channel	aes	geom
la position	**x,y position**	**poly**
margin size	fill (cont)	poly
margin direction	fill (hue)	poly

```
referendum_data %>%
ggplot(data=.,
    aes(x=easting,
        y=northing,
        fill=share_leave) +
    geom_polygon()
```



LAs ordered by geospatial position

LAs ordered by geospatial position

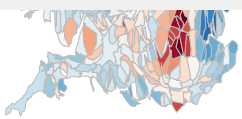


LAs ordered by geospatial position

Summary of mappings - specification

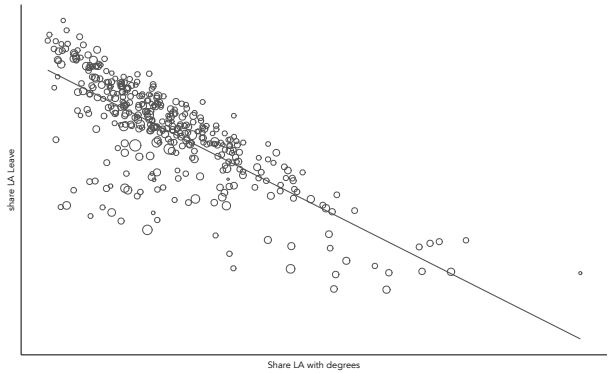
data channel	aes	geom
la position	x,y position	poly
la area	**size**	**poly**
margin size	fill (cont)	poly
margin direction	fill (hue)	poly

```
referendum_data %>%
ggplot(data=.,
    aes(x=easting,
        y=northing,
        fill=share_leave,
        size=area) +
    geom_polygon()
```

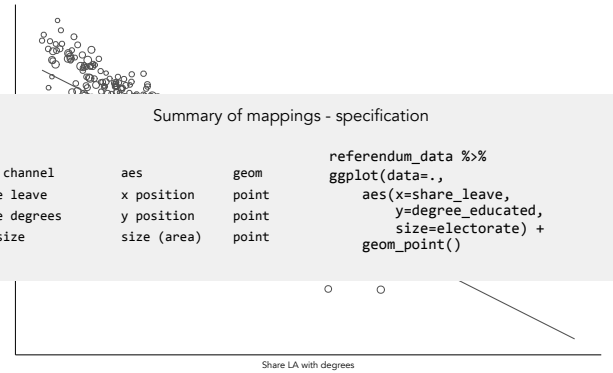


Leave vote by degree-level education

Leave vote by degree-level education



Leave vote by degree-level education

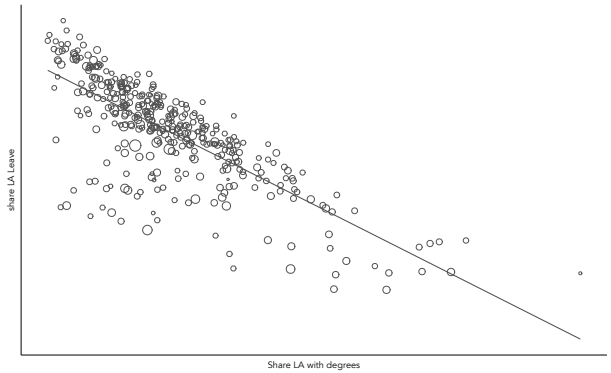


Summary of mappings - specification

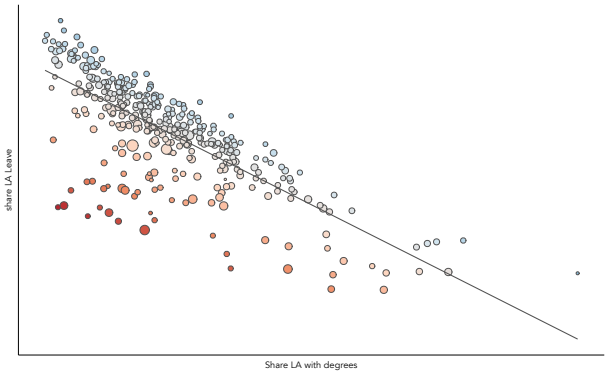
data channel	aes	geom
share leave	x position	point
share degrees	y position	point
pop size	size (area)	point

```
referendum_data %>%
ggplot(data=.,
    aes(x=share_leave,
        y=degree_educated,
        size=electorate) +
    geom_point()
```

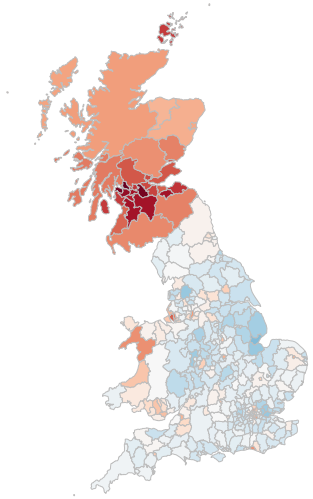
Leave vote by degree-level education



Leave vote by degree-level education

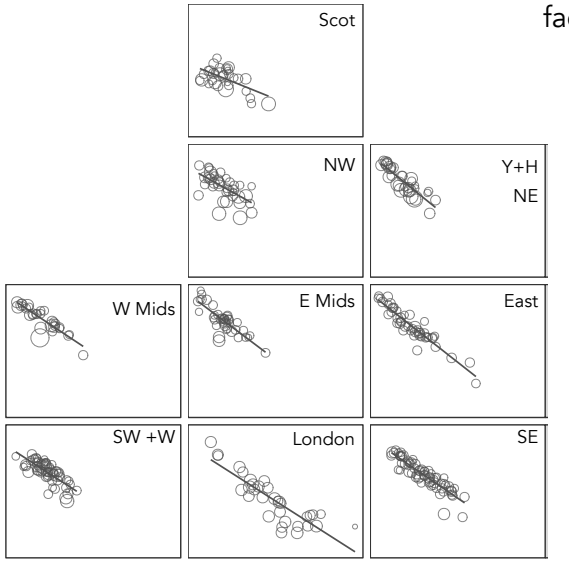


Leave vote by degree-level education

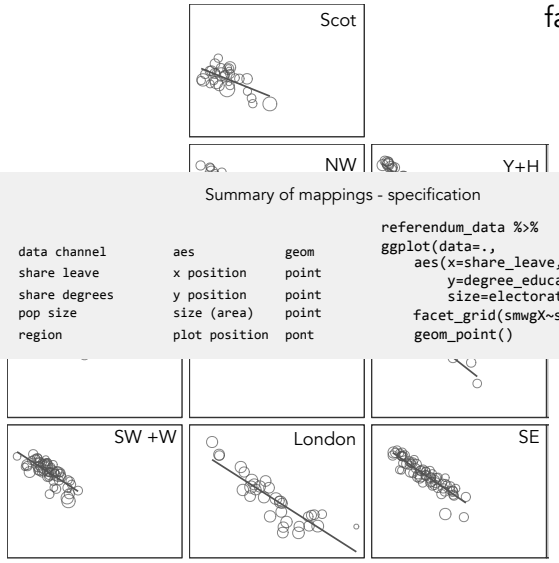


Leave vote by degree-level education faceted by region

Leave vote by degree-level education faceted by region



Leave vote by degree-level education faceted by region



Summary of mappings - specification

data channel	aes	geom
share leave	x position	point
share degrees	y position	point
pop size	size (area)	point
region	plot position	pont

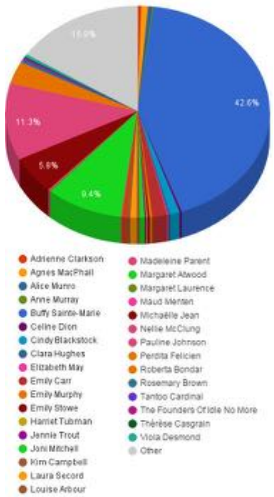
```
referendum_data %>%
ggplot(data=.,
    aes(x=share_leave,
        y=degree_educated,
        size=electorate) +
    facet_grid(smwgX~smwgY) +
    geom_point()
```

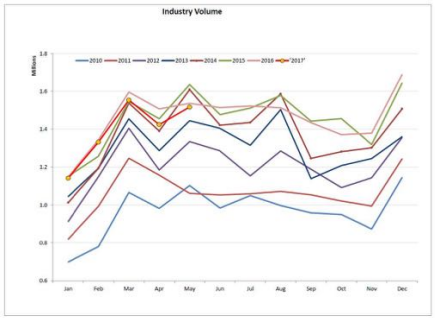
Break

Part 2 : visualization guidelines

WTF Visualizations

Visualizations that make no sense.

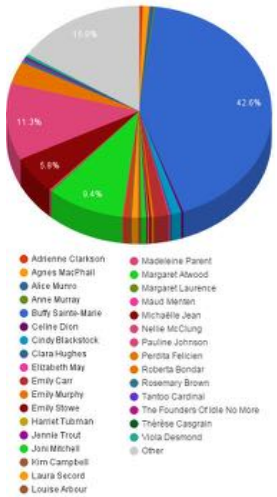




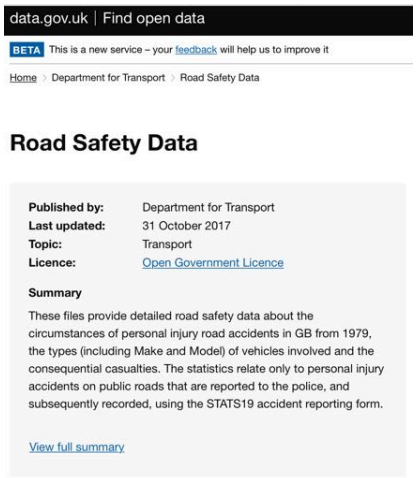
http://viz.wtf

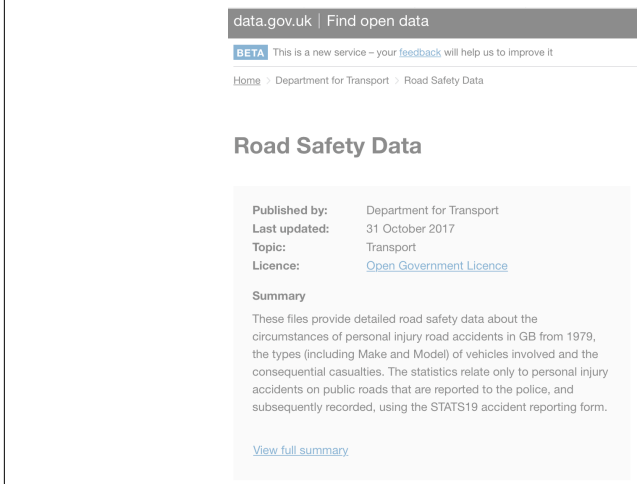
WTF Visualizations

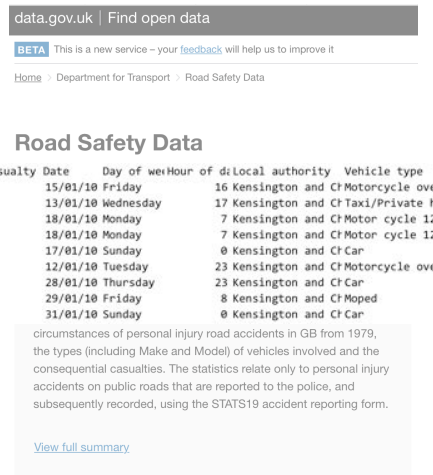
Visualizations that make no sense.



http://viz.wtf

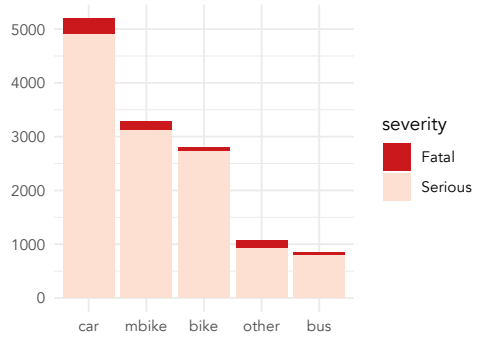




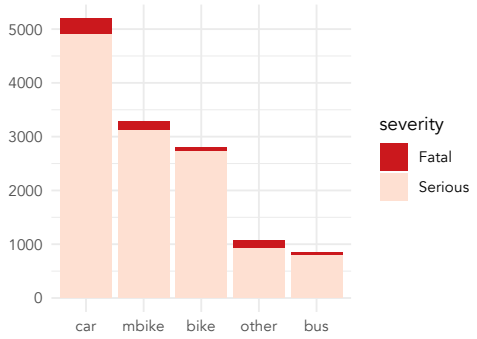


Guideline 1 : match visual channel to data type

Guideline 1 : match visual channel to data type

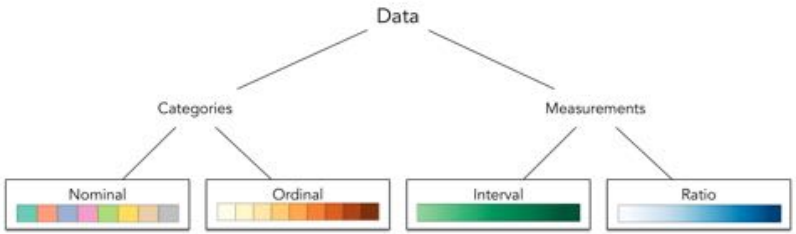


Guideline 1 : match visual channel to data type

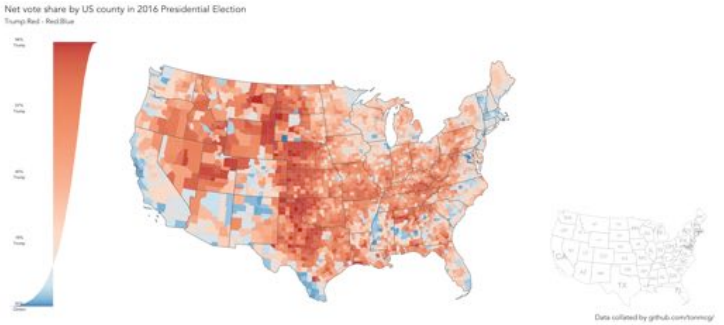


dimension	type	visual channel
casualty count	ratio	length (bar height)
casualty count by mode	ordinal	position (x-axis)
casualty severity	ordinal	colour lightness

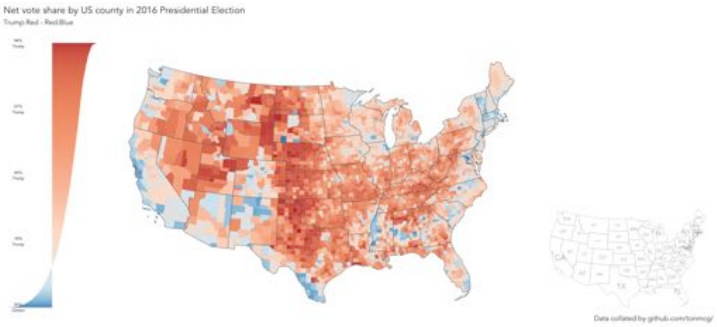
Guideline 1 : match visual channel to data type



Guideline 1 : match visual channel to data type

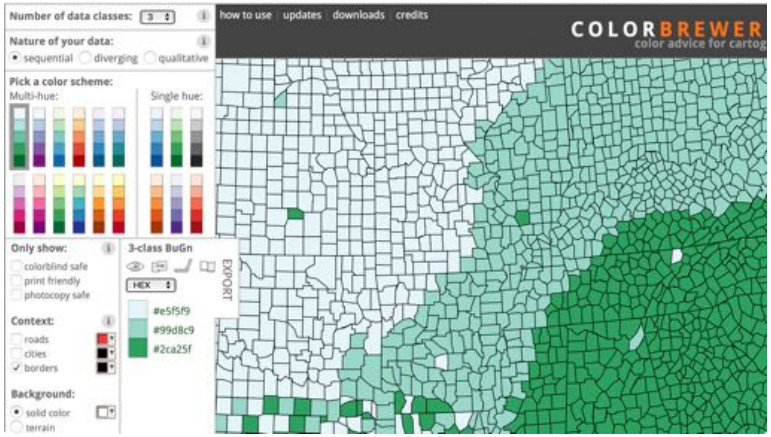


Guideline 1 : match visual channel to data type



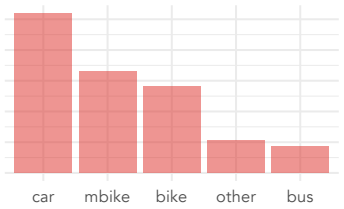
dimension	type	visual channel
majority Clinton|Trump	nominal	colour hue (red|blue)
majority size	ratio	colour lightness

Guideline 1 : match visual channel to data type

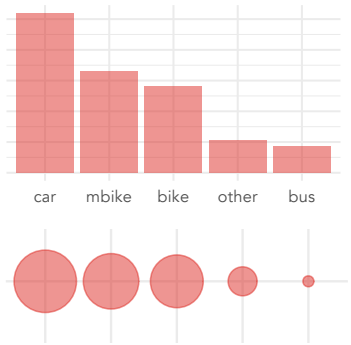


Guideline 2 : exploit our cognitive abilities

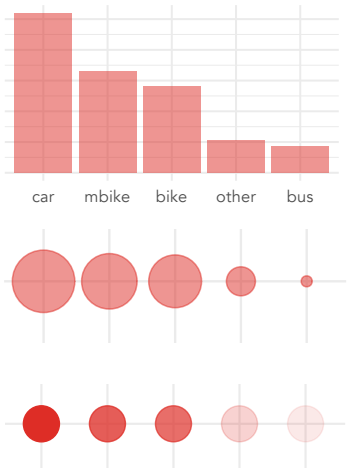
Guideline 2 : exploit our cognitive abilities



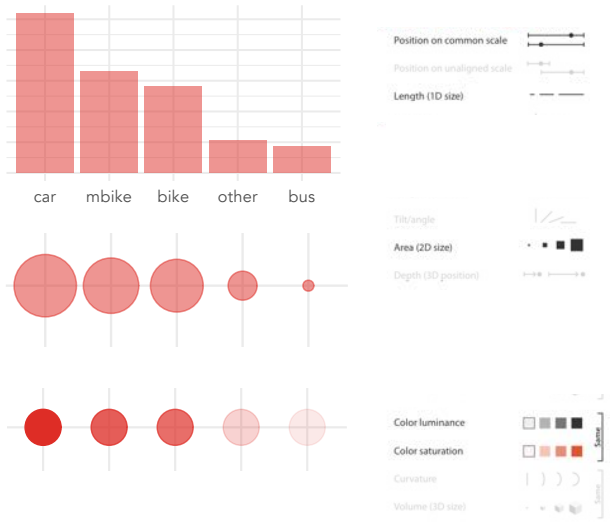
Guideline 2 : exploit our cognitive abilities



Guideline 2 : exploit our cognitive abilities

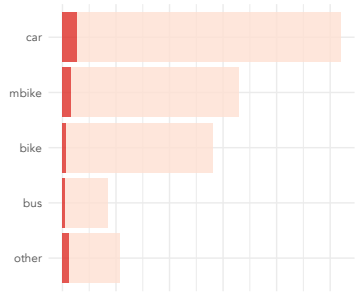


Guideline 2 : exploit our cognitive abilities

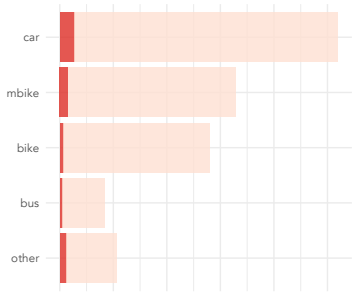


Guideline 3: use layout to encourage comparison

Guideline 3: use layout to encourage comparison

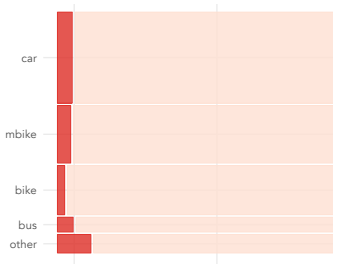


Guideline 3: use layout to encourage comparison



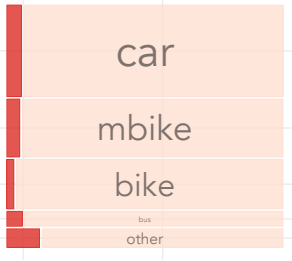
width : fatal+serious freq
height : not encoded
order : travel mode (freq)

Guideline 3: use layout to encourage comparison



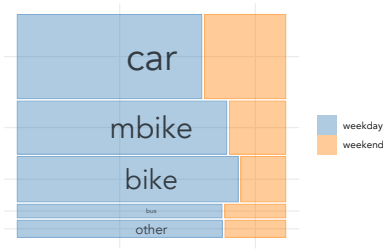
width : proportion serious | fatal
height : travel mode freq
order : travel mode (freq)

Guideline 3: use layout to encourage comparison



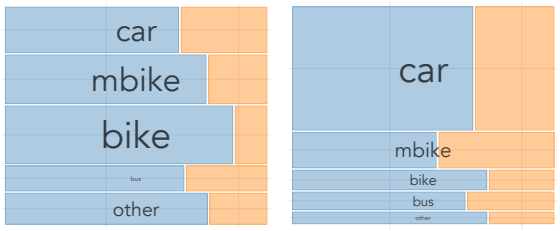
width : proportion serious | fatal
height : travel mode freq
order : travel mode (freq)

Guideline 3: use layout to encourage comparison



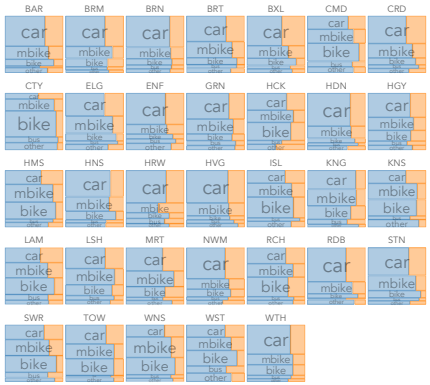
width : **proportion weekday | weekend**
height : travel mode freq
order : travel mode (freq)

Guideline 3: use layout to encourage comparison



width : proportion weekday | weekend
height : travel mode freq
order : travel mode (freq)
facet : borough (select)

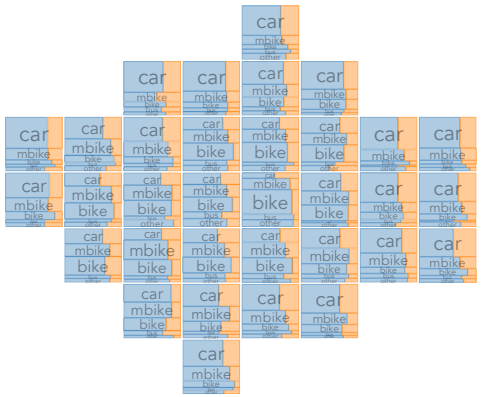
Guideline 3: use layout to encourage comparison



width : proportion weekday | weekend
height : travel mode freq
order : travel mode (freq)
facet : borough alphabetical

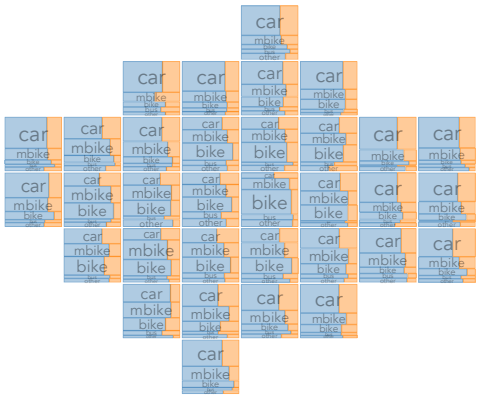
Guideline 3: use layout to encourage comparison

Guideline 3: use layout to encourage comparison

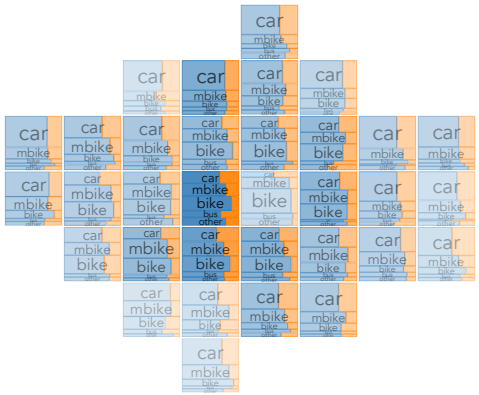


width : proportion weekday | weekend
height : travel mode freq
order : travel mode (freq)
facet : borough semi-geo position

Guideline 3: use layout to encourage comparison

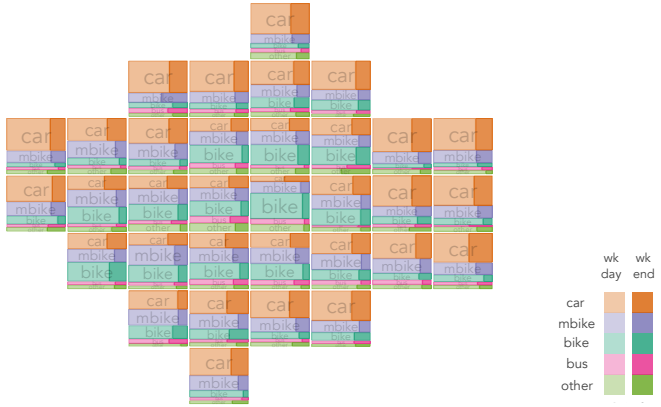


Guideline 3: use layout to encourage comparison



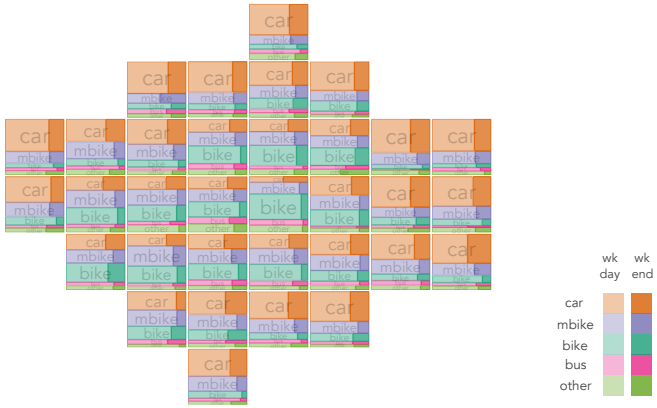
width : proportion weekday | weekend
height : travel mode freq
order : travel mode (freq)
facet : borough semi-geo position
lightness : total accidents

Guideline 3: use layout to encourage comparison



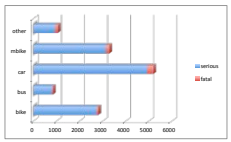
width : proportion weekday | weekend
height : travel mode freq
order : travel mode (freq)
facet : borough semi-geo position
lightness : weekday / weekend
hue : vehicle

Guideline 3: use layout to encourage comparison



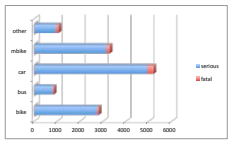
Guideline 4 : emphasise the important de-emphasize the unimportant

Guideline 4 : emphasise the important de-emphasize the unimportant

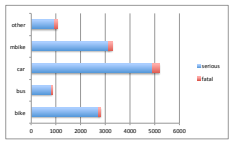


a bad excel default

Guideline 4 : emphasise the important de-emphasize the unimportant

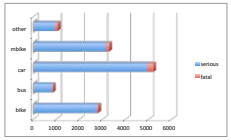


a bad excel default

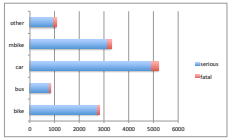


remove pointless 3d

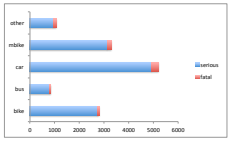
Guideline 4 : emphasise the important de-emphasize the unimportant



a bad excel default

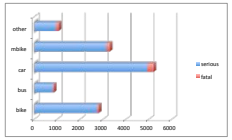


remove pointless 3d

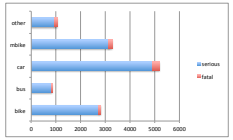


remove reference lines

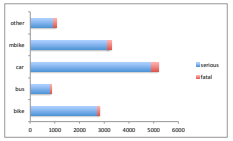
Guideline 4 : emphasise the important de-emphasize the unimportant



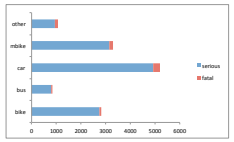
a bad excel default



remove pointless 3d

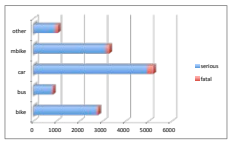


remove reference lines

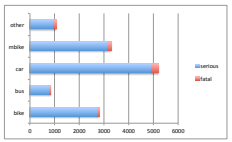


remove bar shadow and gradient

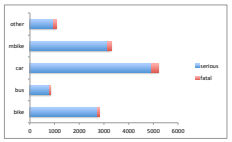
Guideline 4 : emphasise the important de-emphasize the unimportant



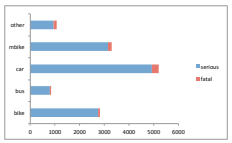
a bad excel default



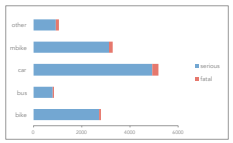
remove pointless 3d



remove reference lines

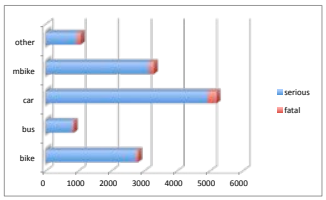


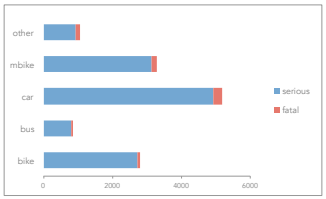
remove bar shadow and gradient



emphasise data, de-emphasise axes

Guideline 4 : emphasise the important de-emphasize the unimportant





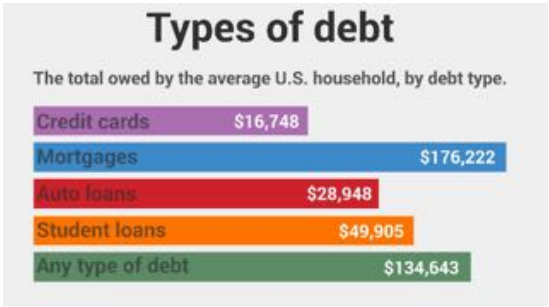
Data-Ink ratio = data ink : total ink in graphic

Tufte 2001

Guideline 5 : don't lie

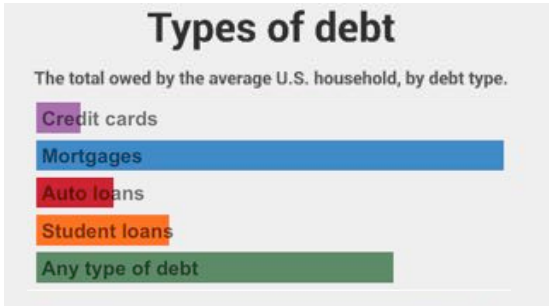
http://viz.wtf

Guideline 5 : don't lie



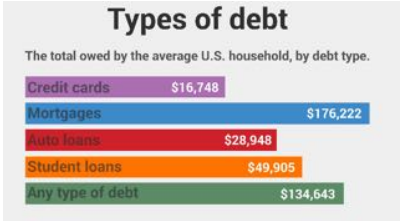
http://viz.wtf

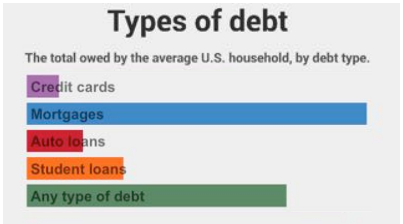
Guideline 5 : don't lie



http://viz.wtf

Guideline 5 : don't lie





http://viz.wtf

Guideline 5 : don't lie

Graphical integrity and the lie factor

The representation of numbers as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

Tufte 2001 : 77

Guideline 5 : don't lie

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

via Jason Dykes

Guideline 5 : don't lie

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

via Jason Dykes

Guideline 5 : don't lie

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

gap = 20 pixels
red = €4.8 / px
blu = €6.1 / px

gap should be …
1.6 pixels
gap is …
20 pixels
20 / 1.6 =

x13!

via Jason Dykes