

Visual Analytics

Communicating data-driven insights
through data visualization techniques
and useful dashboards

Mikel Madina &
Miren Berasategi
miren.berasategi@deusto.es

Data Communication



0. Introduction

Key points

- **Data driven**: as seen in previous sessions with Professors Onieva, Gutiérrez and Lorenzo
 - **Insights**: the capacity to gain an accurate and deep understanding of something *through*
 - **Data visualization techniques**: that take the user from data to insight
 - **Dashboards**: as *situation awareness* tools
- + Tableau Desktop to practice

Section outline

0. Introduction: the *why* and the *what for* of visualization
1. Graphs: some reminders, idioms to map variables to graphs
2. Promote insight: by adding meaningful modifications to graphs
3. Dashboards: situation awareness, dos and don'ts
4. Epilogue

Practice: build a simple dashboard with online marketing campaign data

Why use visualization

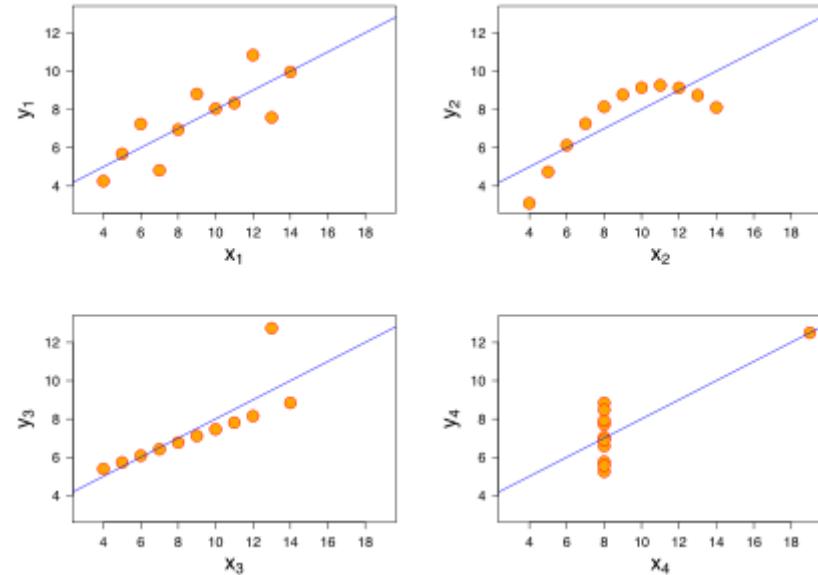
- Sight is our most developed sense
- The visual system provides a very high-bandwidth channel to our brains
- A significant amount of visual information processing occurs in parallel at the preconscious level
- The human brain is *trained* to identify visual patterns
- Summary statistics have the intrinsic limitation of data loss

Why use visualization

Anscombe's quartet

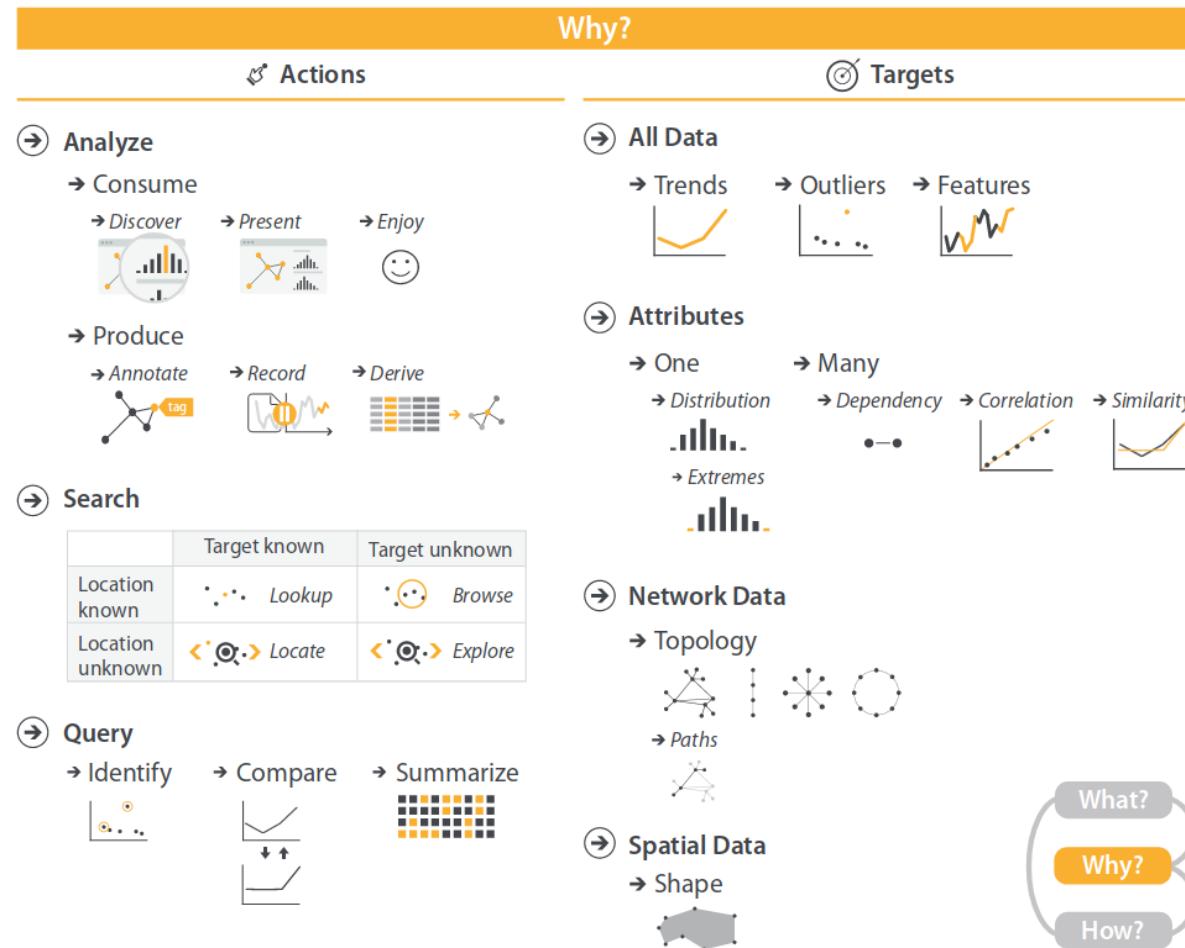
| I | | II | | III | | IV | |
|------|-------|------|------|------|-------|------|-------|
| x | y | x | y | x | y | x | y |
| 10.0 | 8.04 | 10.0 | 9.14 | 10.0 | 7.46 | 8.0 | 6.58 |
| 8.0 | 6.95 | 8.0 | 8.14 | 8.0 | 6.77 | 8.0 | 5.76 |
| 13.0 | 7.58 | 13.0 | 8.74 | 13.0 | 12.74 | 8.0 | 7.71 |
| 9.0 | 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 |

| Property | Value | Accuracy |
|-----------------------------|---------------------|---|
| Mean of x | 9 | exact |
| Sample variance of x | 11 | exact |
| Mean of y | 7.50 | to 2 decimal places |
| Sample variance of y | 4.125 | plus/minus 0.003 |
| Correlation between x and y | 0.816 | to 3 decimal places |
| Linear regression line | $y = 3.00 + 0.500x$ | to 2 and 3 decimal places, respectively |



Anscombe's Quartet

What to use visualization for



What to use visualization for

Targets

→ All Data

→ Trends



→ Outliers



→ Features



→ Network Data

→ Topology



→ Paths



→ Attributes

→ One

→ Distribution

→ Extremes

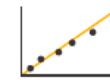


→ Many

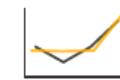
→ Dependency

...

→ Correlation



→ Similarity



→ Spatial Data

→ Shape



Munzner 2014, p.56

What to use visualization for

Actions

>Analyze

→ Consume

→ Discover



→ Present



→ Enjoy



→ Produce

→ Annotate



→ Record



→ Derive



Search

| | Target known | Target unknown |
|------------------|--------------|----------------|
| Location known | | |
| Location unknown | | |

Query

→ Identify



→ Compare



→ Summarize



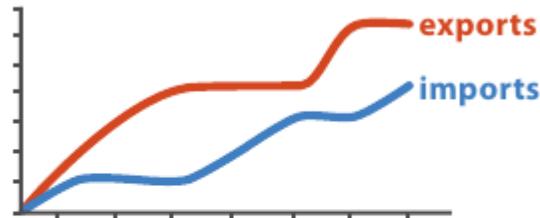
Munzner 2014, p.46

What to use visualization for

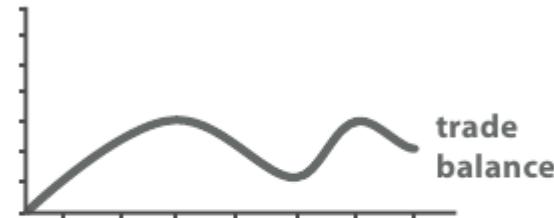
There is a strong relationship between the form of the data (the attribute/variable and dataset types) and what kinds of vis[ualization] idioms are effective at displaying it. (...) Don't just draw what you are given; decide what the right thing to show is, create it with a series of transformations from the original database, and draw that!

Munzner 2014, p.50

What to use visualization for



Original Data



Derived Data

$$\text{trade balance} = \text{exports} - \text{imports}$$

Derived attributes can be directly visually encoded. Munzner 2014, p.52

Practice: meet our sample data

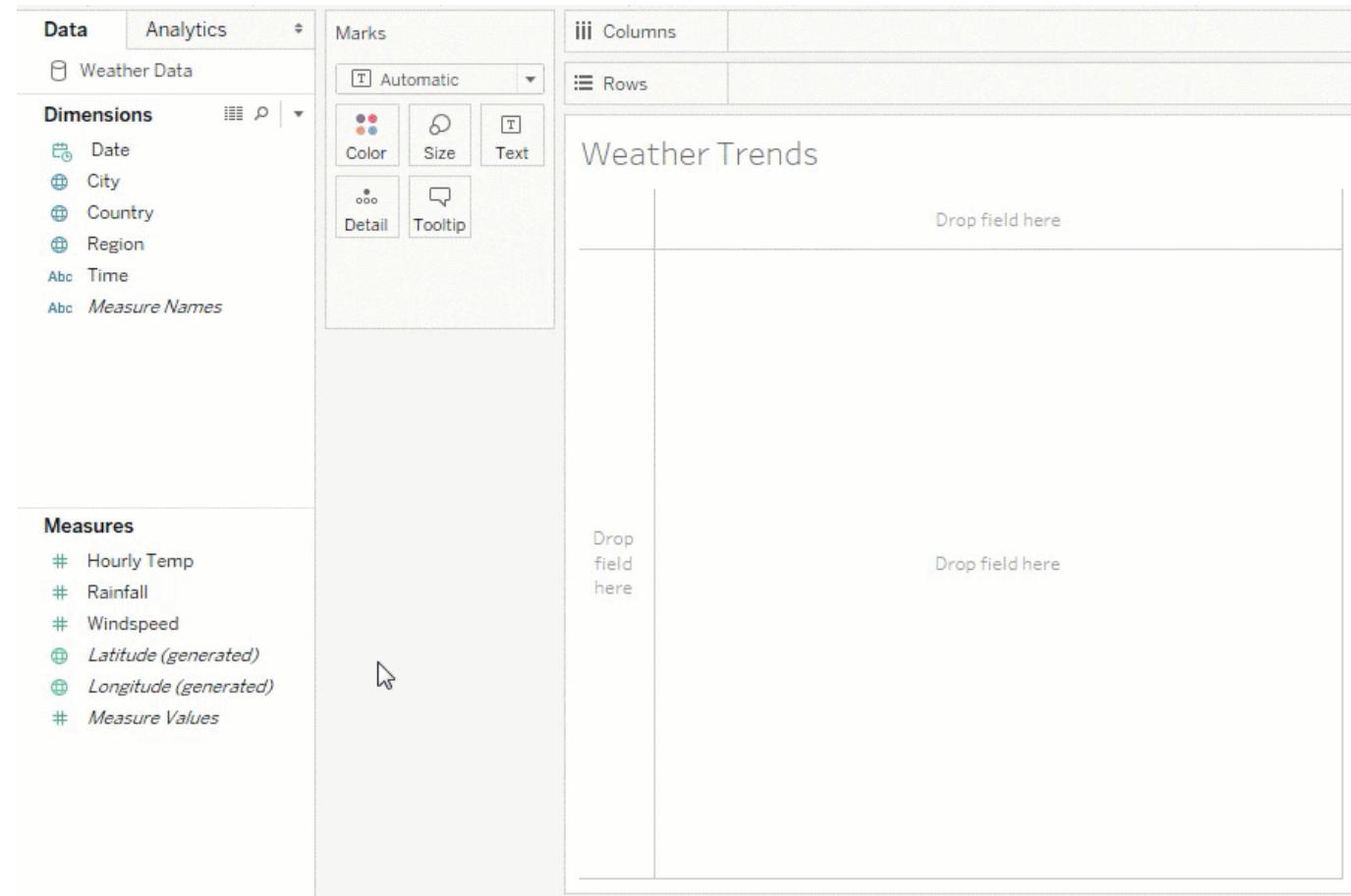
Download and open
data.xls : fake data for
online marketing goals
and tools

from ALUD or

<http://mrn.bz/MUMA2019data>

| | A | B | C | D | E | F |
|----|----------|----------|-------|--------|--------|------|
| 1 | source | quarter | spent | visits | income | goal |
| 2 | Adwords | 20160101 | 1000 | 50000 | 900 | 1500 |
| 3 | Twitter | 20160101 | 200 | 8500 | 1300 | 1000 |
| 4 | Facebook | 20160101 | 500 | 20000 | 800 | 1500 |
| 5 | Adwords | 20160401 | 1000 | 48000 | 1200 | 1500 |
| 6 | Twitter | 20160401 | 300 | 9000 | 1400 | 1000 |
| 7 | Facebook | 20160401 | 750 | 21500 | 1400 | 1500 |
| 8 | Adwords | 20160701 | 1000 | 50000 | 1500 | 1500 |
| 9 | Twitter | 20160701 | 400 | 10000 | 1000 | 1000 |
| 10 | Facebook | 20160701 | 750 | 23000 | 200 | 1500 |
| 11 | Adwords | 20161001 | 1000 | 45000 | 1250 | 1500 |
| 12 | Twitter | 20161001 | 500 | 11000 | 1000 | 1000 |
| 13 | Facebook | 20161001 | 1000 | 25000 | 2000 | 1500 |
| 14 | Adwords | 20170101 | 1000 | 50000 | 1100 | 1500 |
| 15 | Twitter | 20170101 | 500 | 8500 | 1300 | 1000 |
| 16 | Facebook | 20170101 | 1000 | 20000 | 800 | 1500 |
| 17 | Adwords | 20170401 | 1000 | 48000 | 1500 | 1500 |
| 18 | Twitter | 20170401 | 500 | 9000 | 1400 | 1000 |
| 19 | Facebook | 20170401 | 1000 | 21500 | 1400 | 1500 |
| 20 | Adwords | 20170701 | 1000 | 50000 | 1500 | 1500 |
| 21 | Twitter | 20170701 | 500 | 10000 | 1000 | 1000 |
| 22 | Facebook | 20170701 | 1000 | 23000 | 200 | 1500 |
| 23 | Adwords | 20171001 | 1000 | 45000 | 1250 | 1500 |
| 24 | Twitter | 20171001 | 400 | 11000 | 1000 | 1000 |
| 25 | Facebook | 20171001 | 1000 | 25000 | 2000 | 1500 |
| 26 | | | | | | |

Tableau Desktop



The screenshot shows the Tableau Desktop interface with the following components:

- Data pane:** Shows a connection to "Weather Data" and lists dimensions: Date, City, Country, Region, and Time; measures: Hourly Temp, Rainfall, Windspeed, Latitude (generated), Longitude (generated), and Measure Values.
- Marks card:** Set to "Automatic" with options for Color, Size, Text, Detail, and Tooltip.
- Columns and Rows sections:** Both are empty with placeholder text "Drop field here".
- View area:** Titled "Weather Trends", also with a "Drop field here" placeholder.

Tableau Software

Tableau Desktop

1. Load data
2. Explore the Data Source tab

See the subtle blue/green colour of the variable type icon? Take notice, it is important:

Understanding the difference between the blue and green items in Tableau is (IMHO) the single most important piece of understanding necessary to make Tableau function well.

Tom Brown, [Blue things and Green things](#)

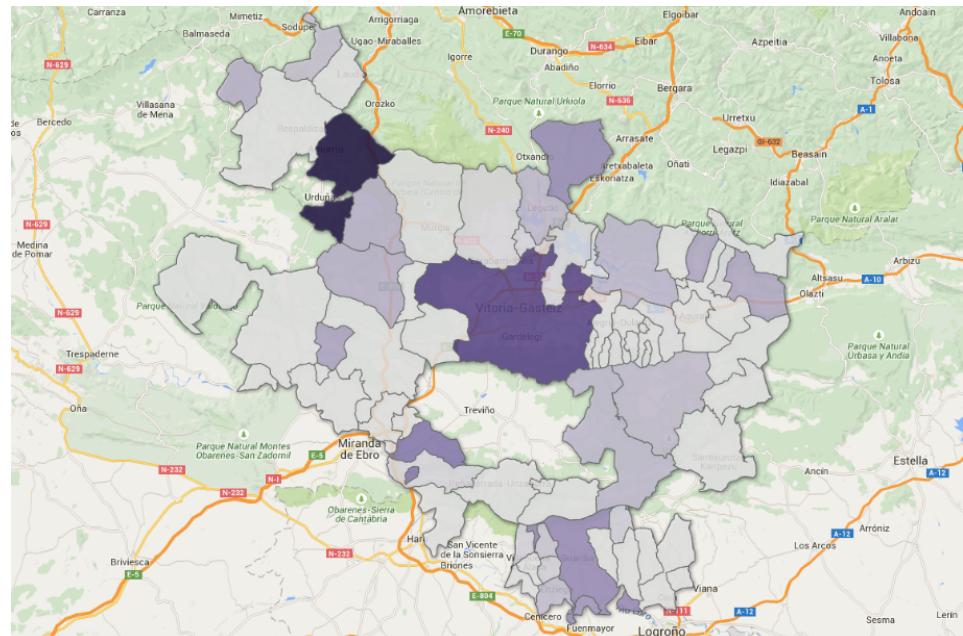
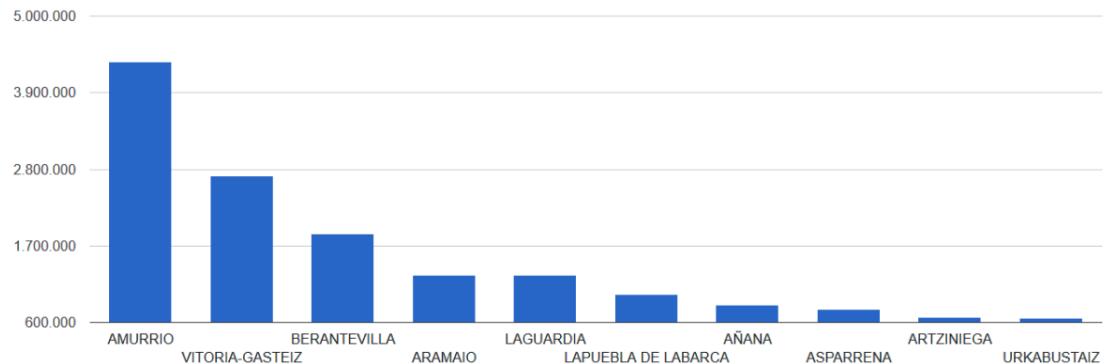
1. Graphs

Section outline

1. Reminder: variable types
2. Mapping variables to graphs (\approx *translating*)
 - Marks
 - Channels, channel types
 - Using marks and channels
3. So, which graph?

Practice: explore dimensions, measures and graph types
in Tableau

ABERASTURI;1025880,88
 ABETXUKO;1626,20
 ABEZIA;183184,81
 ABORNIKANO;54530,28
 ACEBEDO;13519,09
 ACOSTA;64930,00
 ADANA;53139,42
 AGIÑAGA;314344,94
 ALAITZA;75534,95
 ALBENIZ;61152,16
 ALCEDO;21313,54
 ALDA;27922,86
 ALEGRIA-DULANTZI;142607
 3,93
 ALORIA;5211,22
 AMARITA;174311,53
 AMETZAGA
 ASPARRENA;41376,28
 AMETZAGA ZUIA;233555,14
 AMURRIO;4348908,77
 ANDA;477,65
 ANDOIN;21622,76
 ANDOLLU;6532,92
 ANGOSTINA;65719,04
 ANTEZANA;23135,10
 ...



?

1.1 Reminder: variable types

- Quantitative
 - Continuous
 - Discrete
 - Qualitative
 - Categorical
 - Ordinal
 - Special types
 - time
 - space
 - ...
- A question of time**
- Spatial and time/hour variables are special variable types. **Time variables** are specially complex:
- are there 365 days in every year? 30 days in every month? 24 hours in every day?
 - *timezones* make it even more complex to use hours or time of day
- Time may be used as a continuous or as a qualitative variable.
- as a qualitative variable, it has a hierarchy: year > month > (week >) day > hour > minute
 - but different hierarchies may be necessary: bimonthly publications, multiple work shifts in a day...

1.2 Mapping variables to graphs

Understanding **marks** and **channels** provides the building blocks for analyzing visual encodings (Munzner 2014, p.95)

1.2.1 Marks

A **mark** is a basic graphical element in an image

→ Points



→ Lines



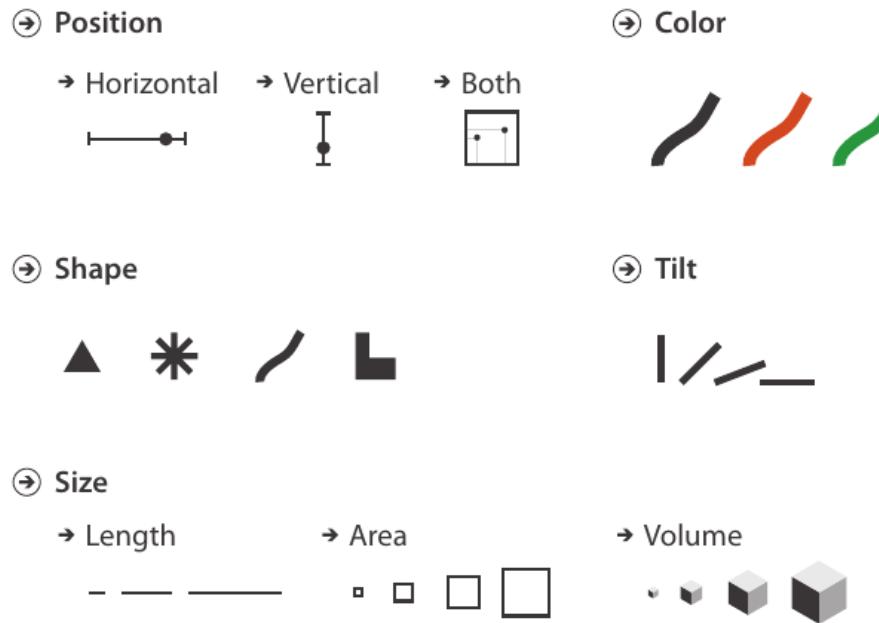
→ Areas



Marks are geometric primitives (Munzner 2014, p.96)

1.2.2 Channels

A visual **channel** is a way to control the appearance of marks



Visual channels control the appearance of marks (Munzner 2014, p.96)

1.2.2 Channels

One and only one attribute/variable should be used per channel.

Multiple channels per attribute are possible (**redundant encoding**), but this approach has limitations.

1.2.2 Channels

The **size** and **shape** channels cannot be used on all types of marks, but most combinations are still possible:

- lines have two *size channels*: length + width
- points refer to location but can be *size* and *shape* coded

1.2.3 Channel types

Two kinds of sensory modalities:

1. **Identity**: what, where
2. **Magnitude**: how much

It does not make sense to ask magnitude questions for shape, color hue. We can ask about magnitudes with length, area or volume; color luminance or saturation; and angle/tilt/slope.

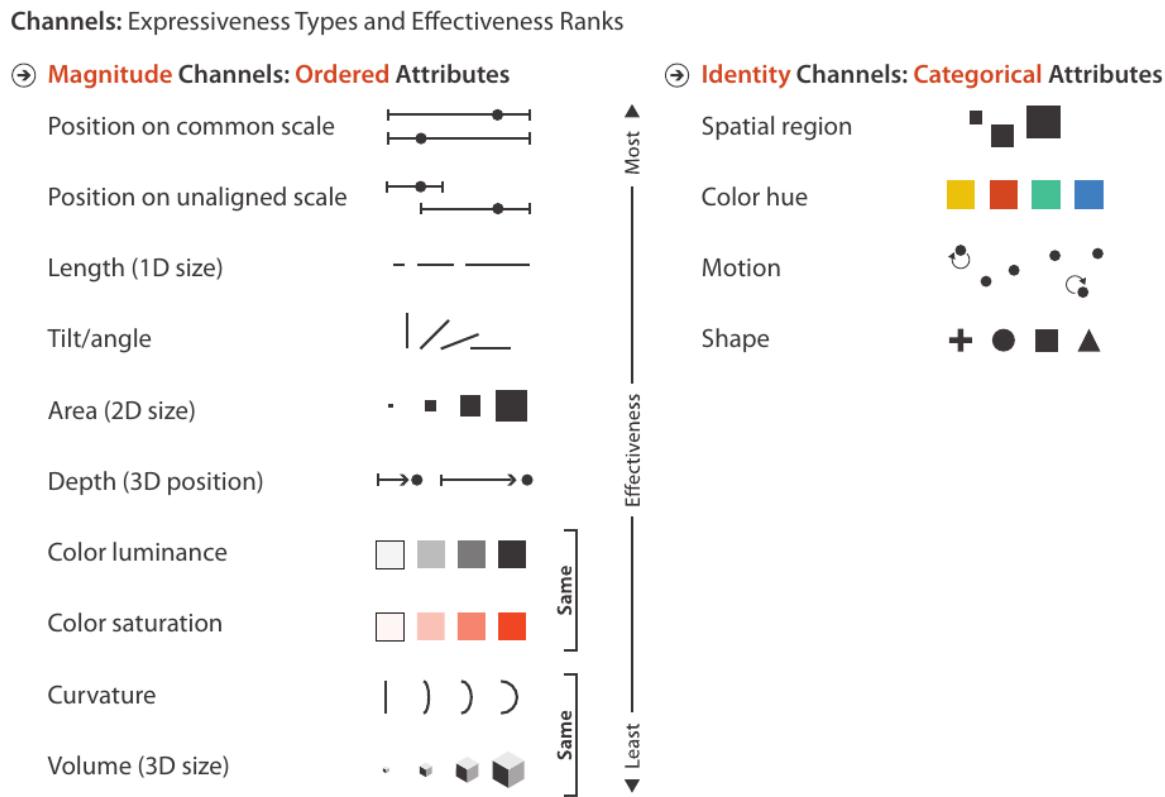
1.2.4 Using marks and channels

All channels are not equal.

The selection of marks and channels should be guided by the principles of **expressiveness** and **effectiveness**.

Once the most important attributes/variables for the desired insight have been identified, the selection of marks and channels should ensure that they are **encoded with the highest ranked**.

1.2.4 Using marks and channels

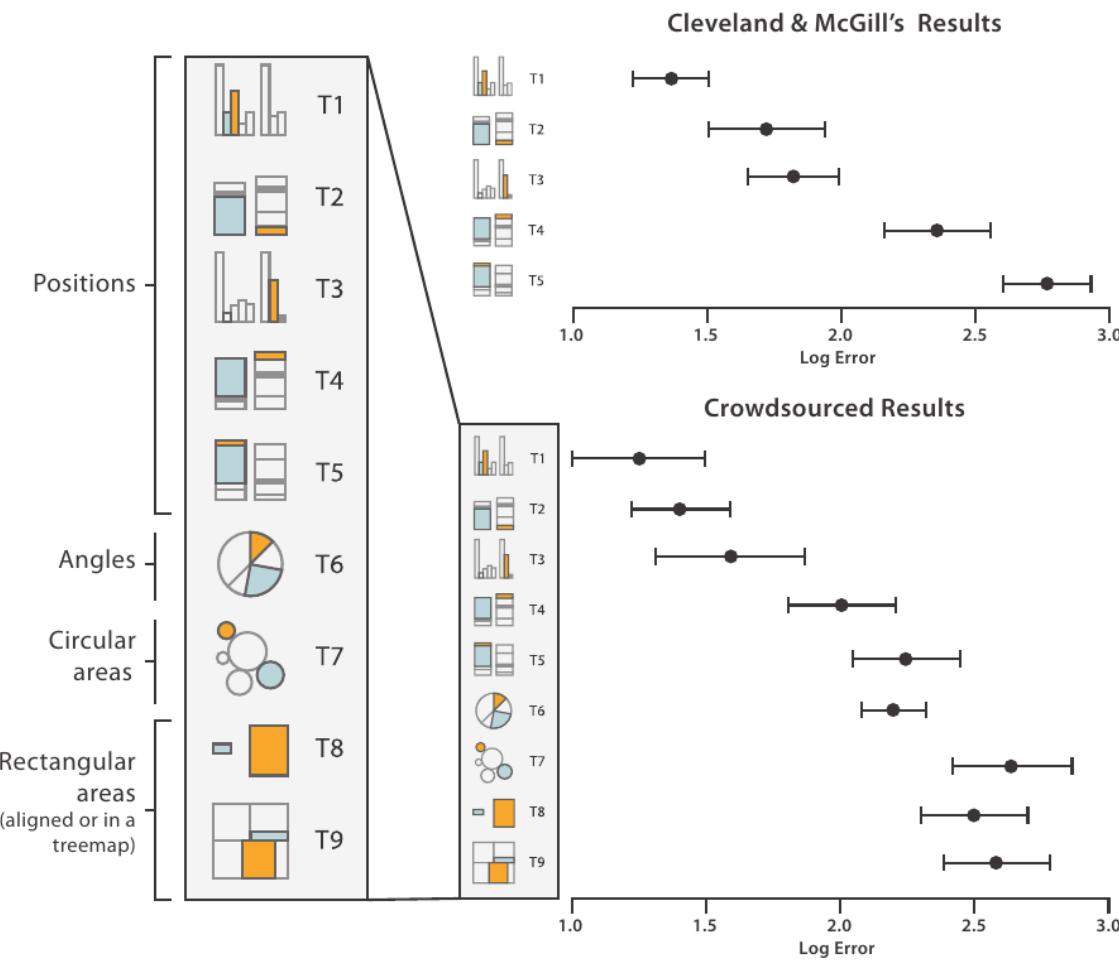


Channels ranked by effectiveness according to data and channel type. Ordered data should be shown with the magnitude channels, and categorical data with the identity channels (Munzner 2014, p.102)

1.2.4 Using marks and channels

The choice of **which attributes/variables to encode with position** is the most central choice in visual encoding.

1.2.4 Using marks and channels



Error rates across visual channels (Munzner 2014, p.105)

1.3 So, which graph?

Chart Suggestions—A Thought-Starter

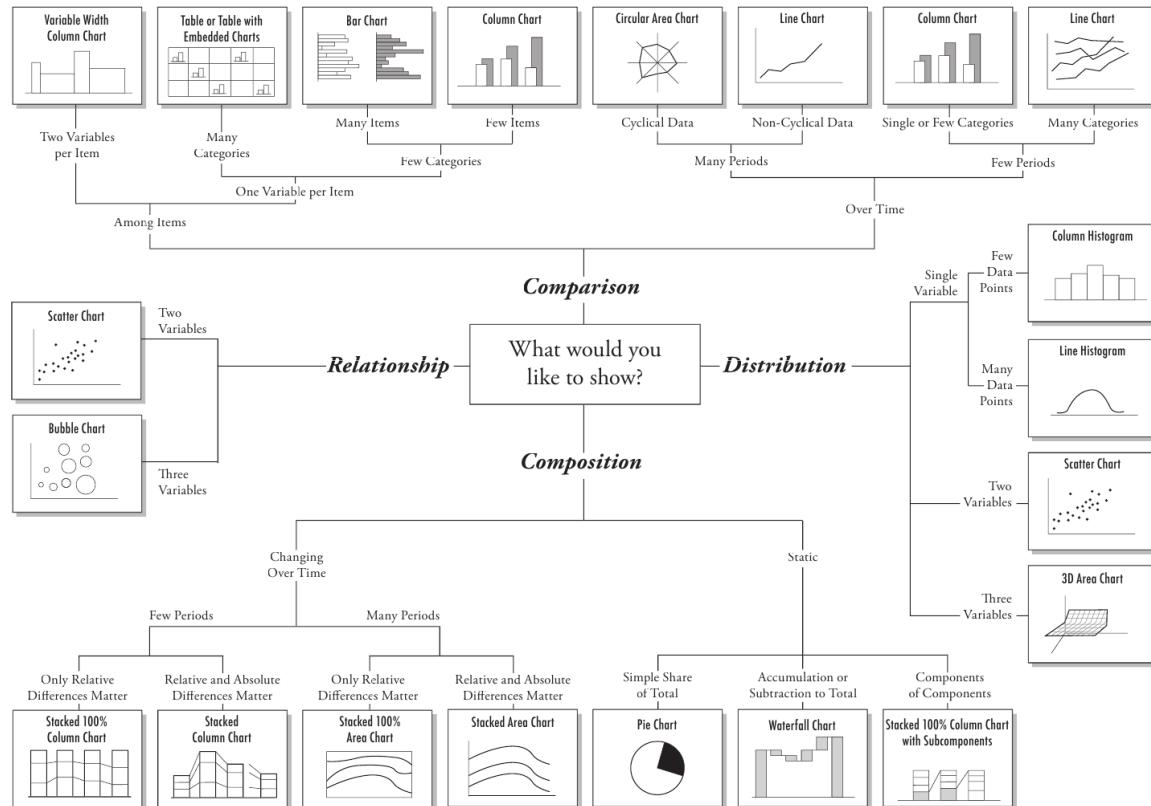


Tableau: let's explore

- Dimensions and measures (remember also blue *vs.* green)
- Encode = drag
- Show me tab

2. Promote insight

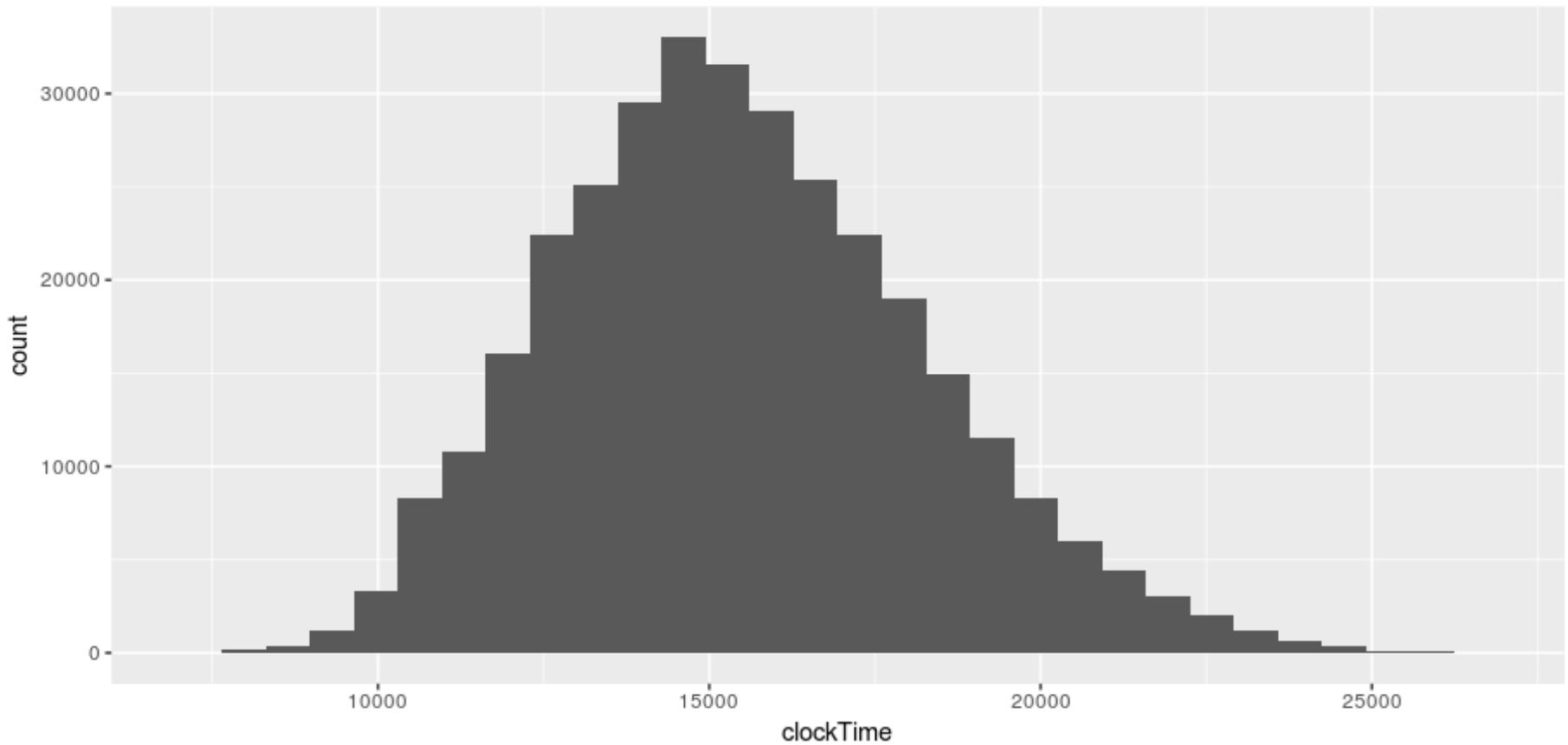
Section outline

How can we enable easier insight through data visualization?

1. Change default settings
2. Make simpler graphs
3. Highlight observations
4. Add attributes as context
5. Add statistical information

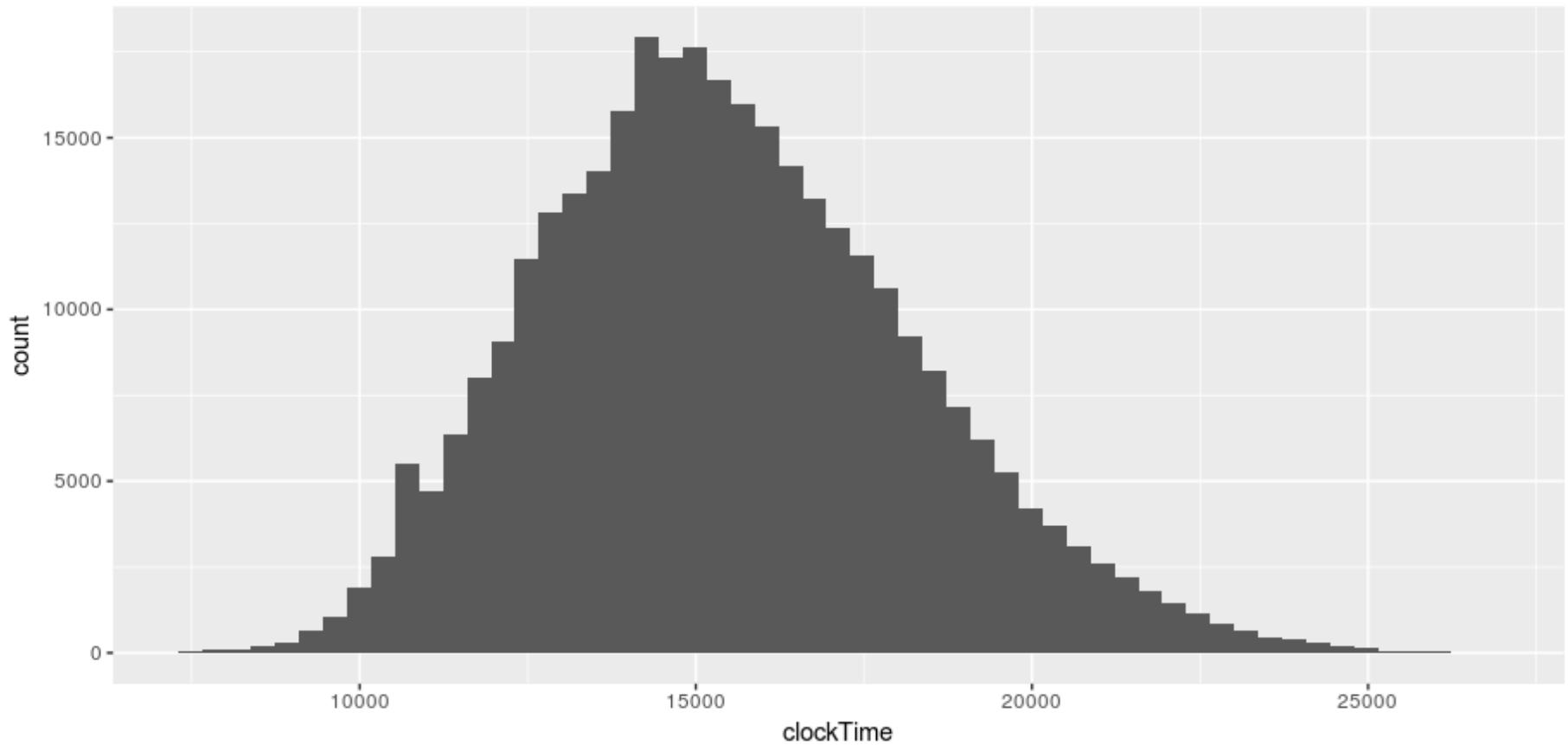
Practice: build (not so) basic graphs

2.1 Change default settings



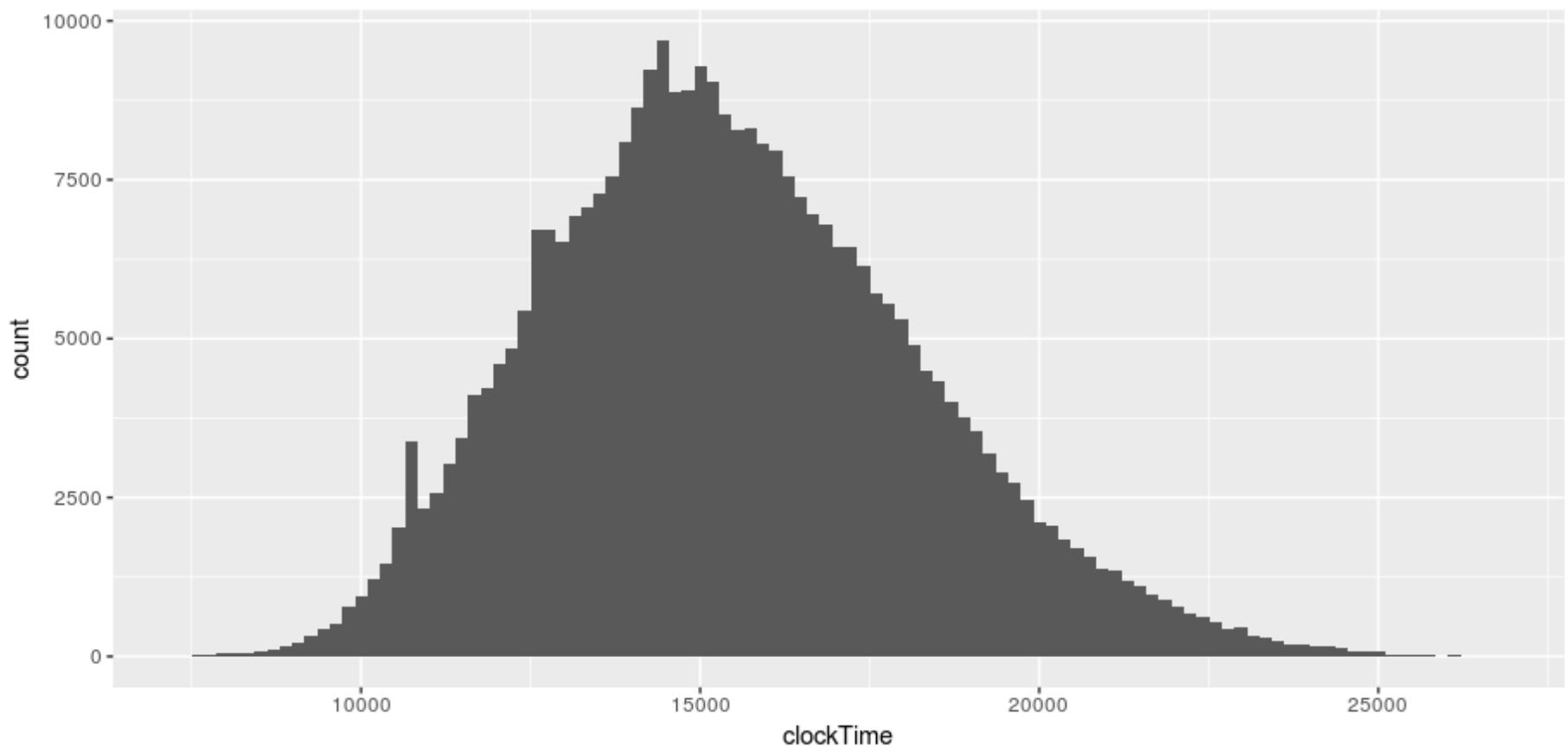
Data source: [Berlin marathon times](#)

2.1 Change default settings



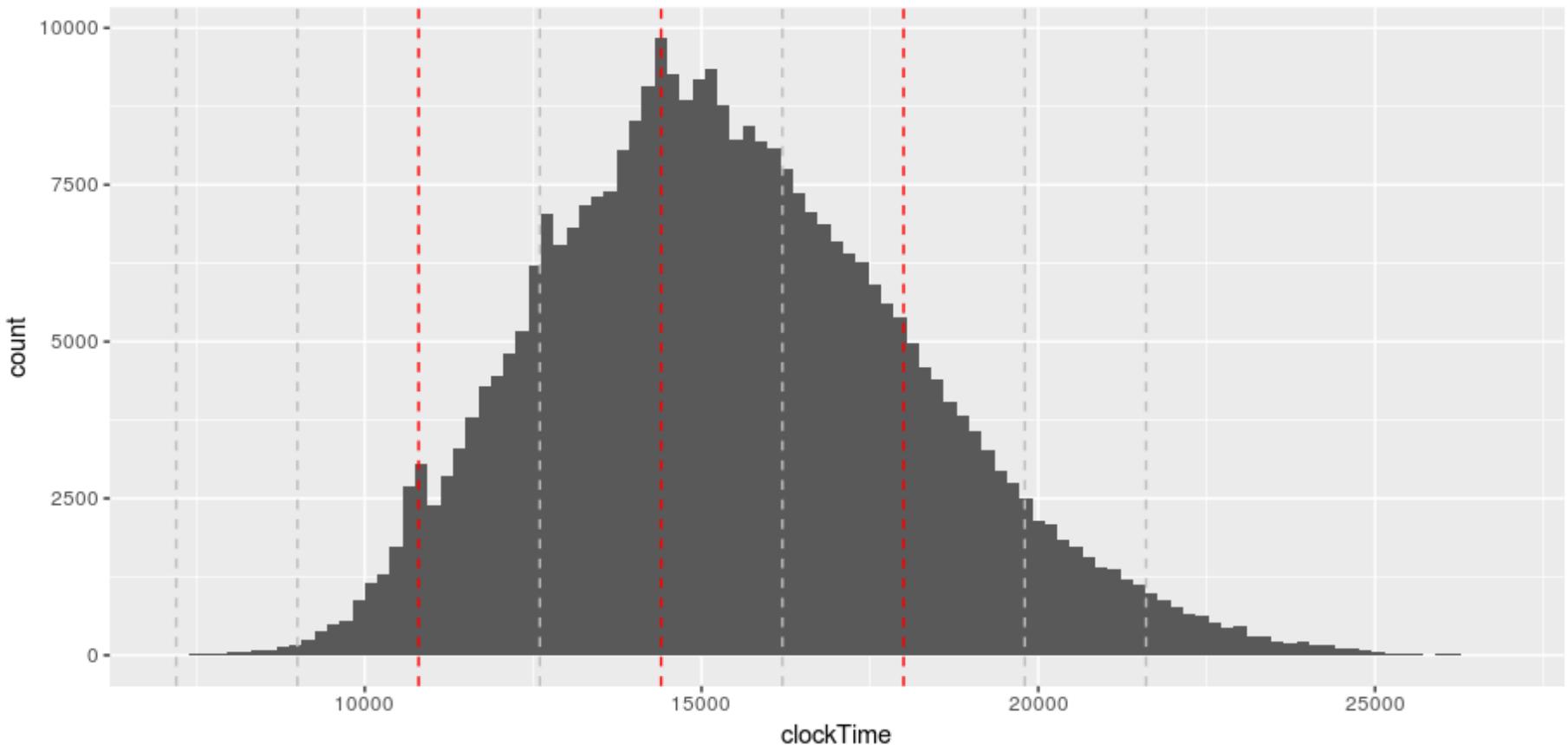
Data source: [Berlin marathon times](#)

2.1 Change default settings



Data source: [Berlin marathon times](#)

2.1 Change default settings



Data source: Berlin marathon times

2.2 Make simpler graphs

Data-ink is the non-erasable core of the graphic, the non-redundant ink arranged in response to variation in the numbers represented.

Tufte 1983

2.2 Make simpler graphs



[Remove to Improve \(the data-ink ratio\)](#)

by [Joey Cherdarchuk](#)



A step-by-step example: [Data looks better naked](#)

2.2 Make simpler graphs

More on decluttering:

Declutter Your Data Visualizations



Nussbaumer, [Declutter Your Data Visualizations](#)

2.3 Highlight observations

Through preattentive attributes:

- they are processed in spatial memory without our conscious action
- make it easier to understand what is represented through a design: saves from consciously processing data

2.3 Highlight observations

756395068473
658663037576
860372658602
846589107830

FIGURE 4.2 Count the 3s example

Nussbaumer 2015, p.103

2.3 Highlight observations

756**3**95068473

65866**3**037576

860**3**72658602

846589107**830**

FIGURE 4.3 Count the 3s example with preattentive attributes

Nussbaumer 2015, p.104

2.3 Highlight observations

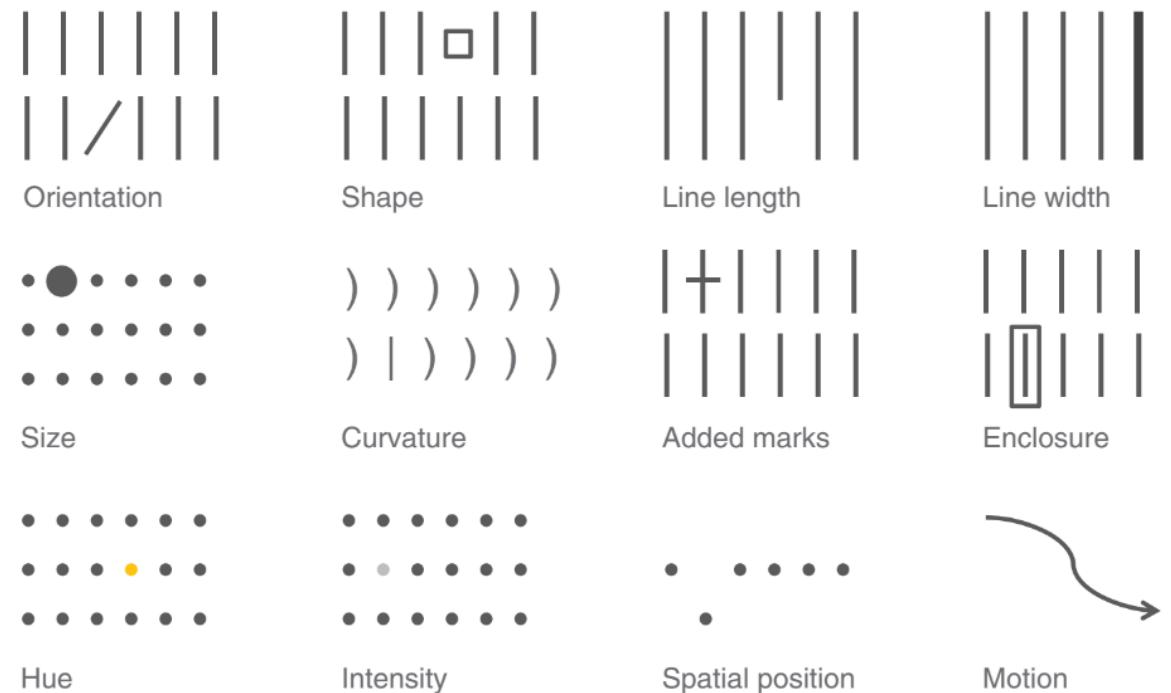


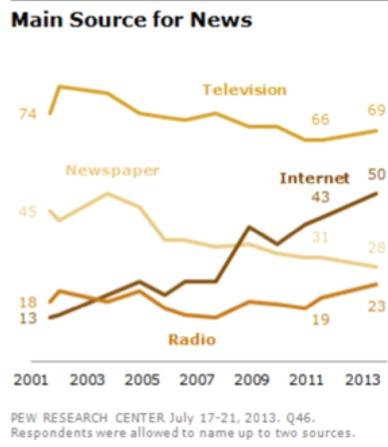
FIGURE 4.4 Preattentive attributes

Source: Adapted from Stephen Few's *Show Me the Numbers*, 2004.

Nussbaumber 2015, p.105

2.3 Highlight observations

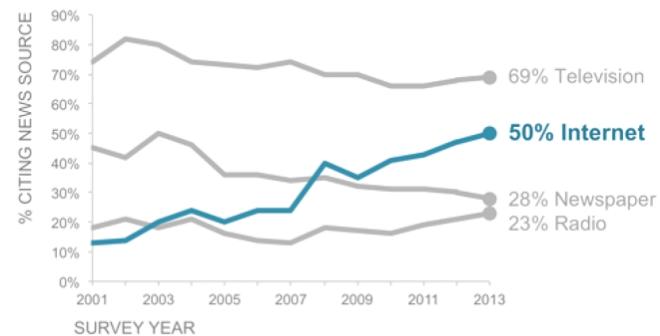
1. More Americans get news online... 50% of the public now cite the internet as a main source for national and international news, still below television, newspapers and radio. (Report)



More Americans get news online

50% of the public cite the **internet** as a main source for national & international news. This remains below television, but is far above newspapers and radio.

Main source for news



PEW RESEARCH CENTER July 17-21, 2014 Q46.
Respondents were allowed to name up to two sources.

Source: <http://www.pewresearch.org/fact-tank/2013/10/16/12-trends-shaping-digital-news/>
© 2010 - 2016 Cole Nussbaumer Knaflic. All rights reserved.

storytelling with data

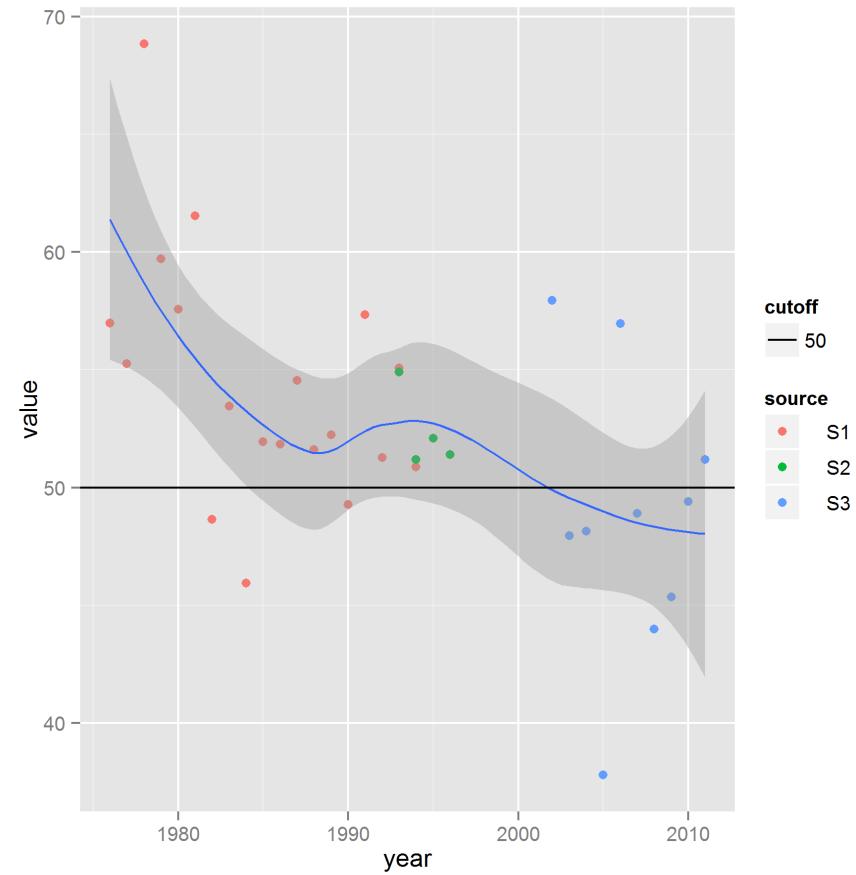
Nussbaumber, Do you see it? The importance of contrast when communicating with data [video]

2.4 Add variables (as context)

- Adding preexisting variables (in moderation)
- Creating conditional variables from preexisting variables
 - binaries or with few levels are best
 - example of calculated field or variable: weekend date

2.5 Add statistical information

- statistical summaries
(mean, variance)
- models



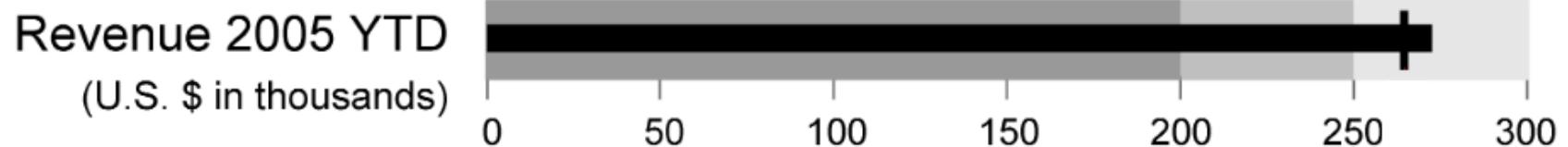
source

Tableau: (not so) basic graphs



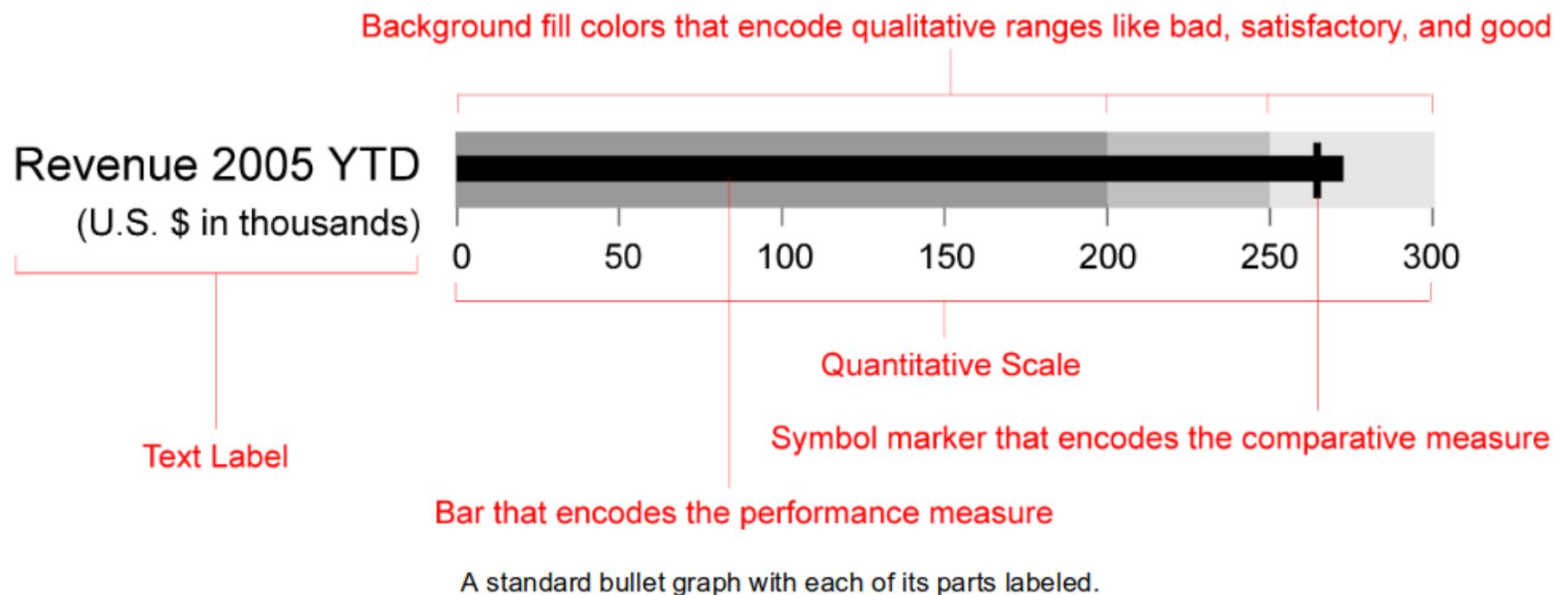
Sparklines ([Tufte 2006](#))

Tableau: (not so) basic graphs



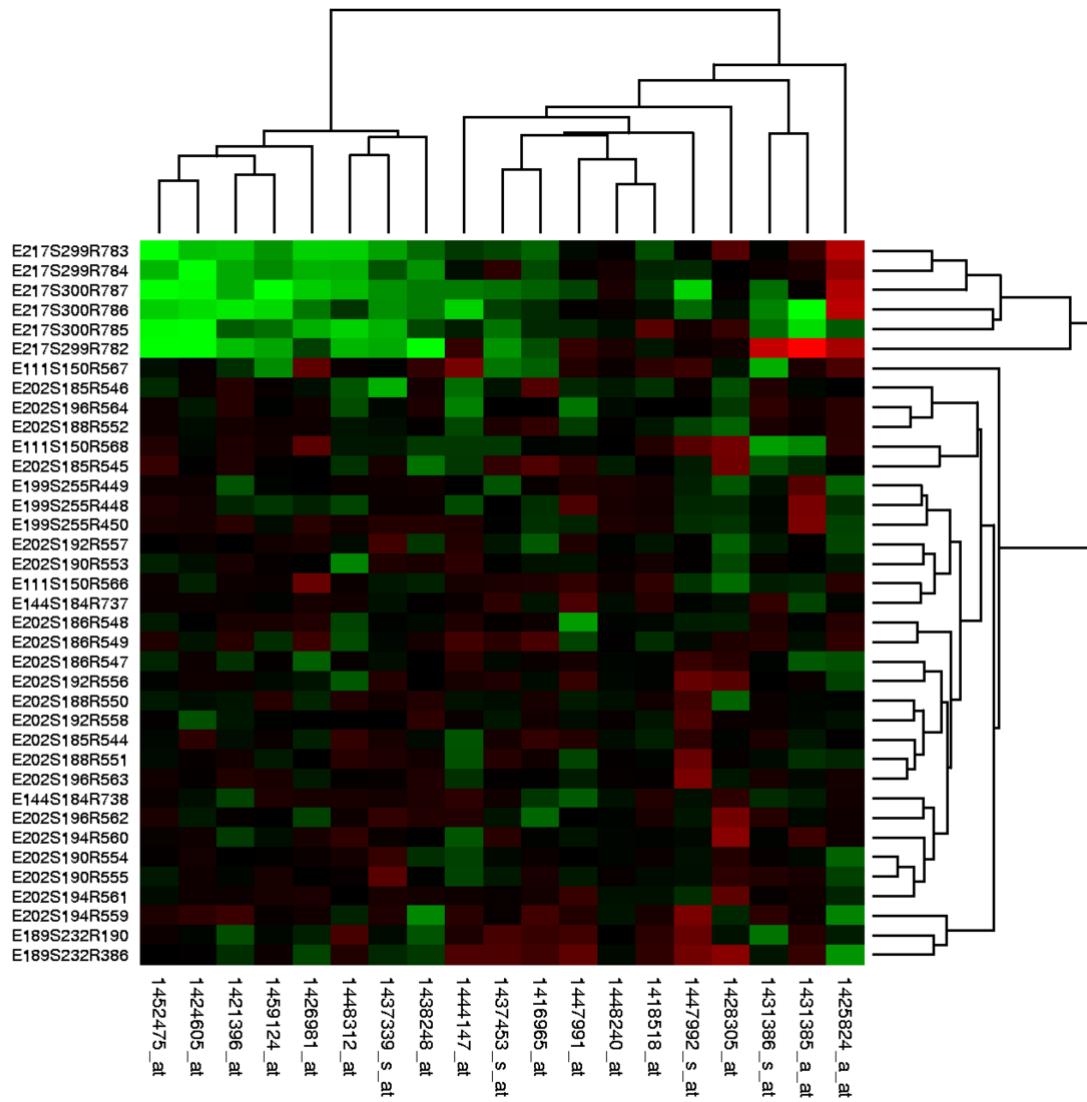
Bulletgraphs ([Few 2007](#))

Tableau: (not so) basic graphs



Bulletgraphs (Few 2007)

Tableau: (not so) basic graphs



Heatmaps (Few 2006)

3. Dashboards

Section outline

1. What is a dashboard?
2. Common design mistakes
3. Key goals in the visual design process
4. Example

Practice: layout and format graphs into a dashboard

3.1 What is a dashboard?

Visual display of the most information needed to achieve one or more objectives which fits entirely on a single computer screen so it can be monitored at a glance.

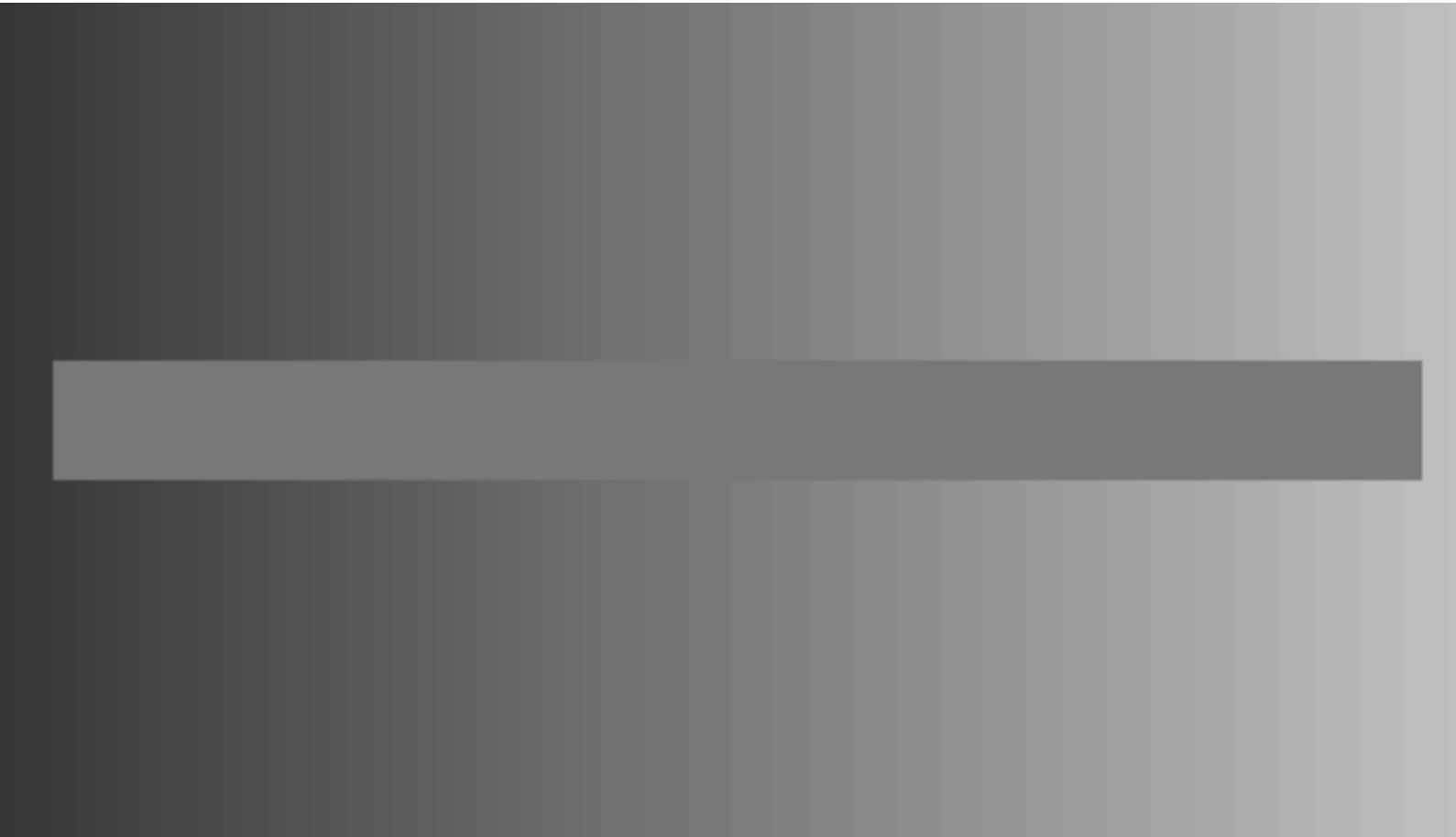
Few 2013

3.1 What is a dashboard?

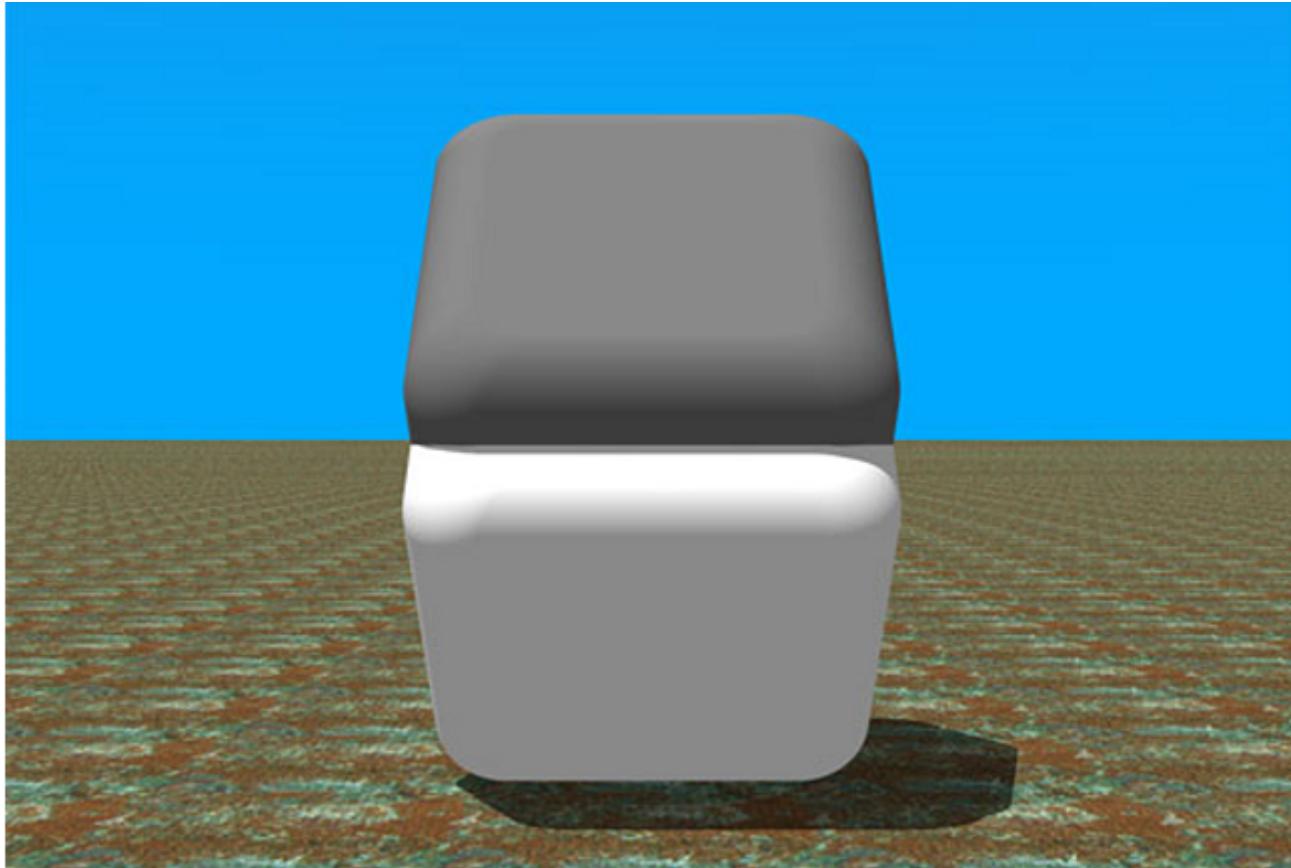
- **Visual display:** *I see = I understand* → **insight to achieve specific objectives:** may require gathering information that is otherwise unrelated or disperse
- **fits in a single computer screen:** it must all be seen at once (short-term memory effect)
- **monitored at a glance:** doesn't need to provide all the details, but if it doesn't, it should make it as easy and seamless as possible to get to that information



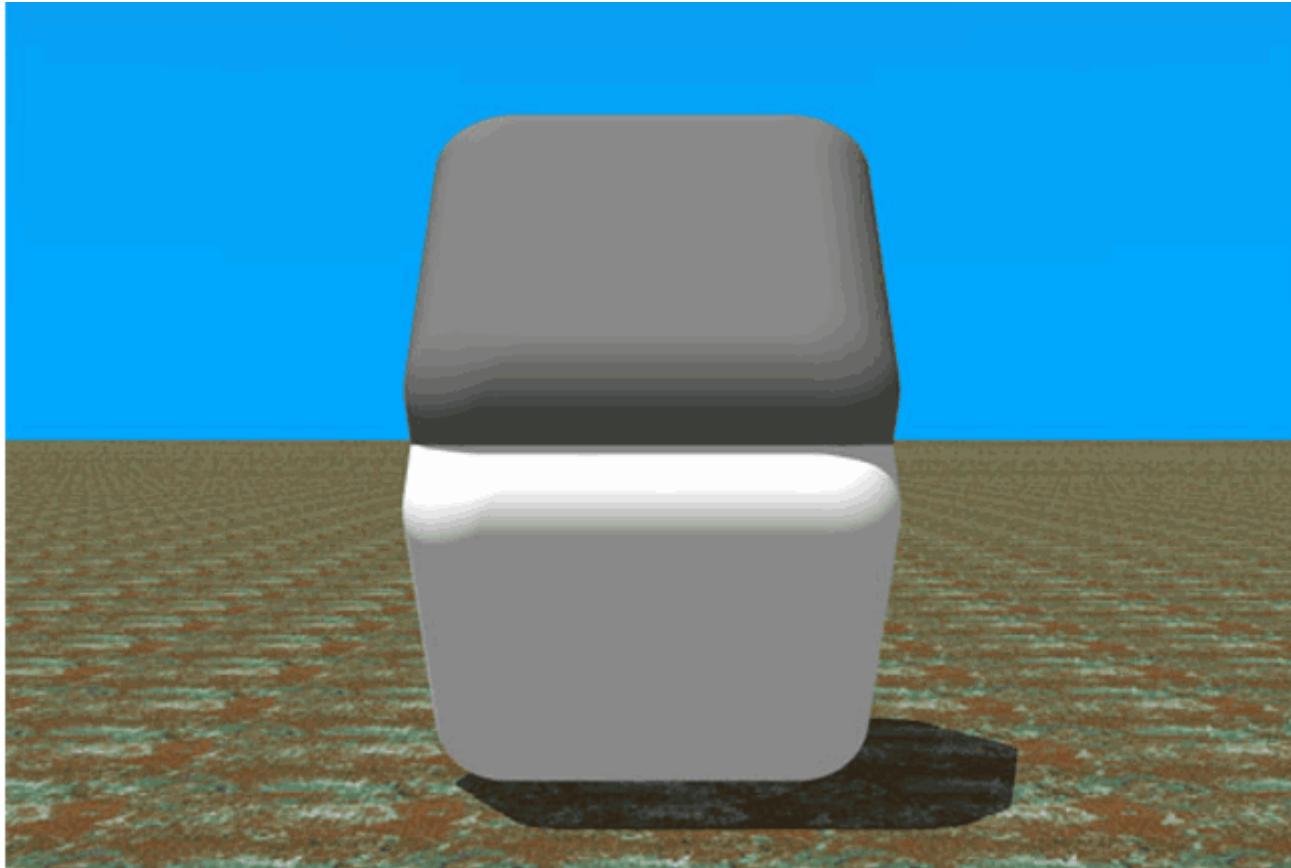
What colour is the bar in the middle?



What colour is the bar in the middle?



Are these two boxes the same colour?



Are these two boxes the same colour?



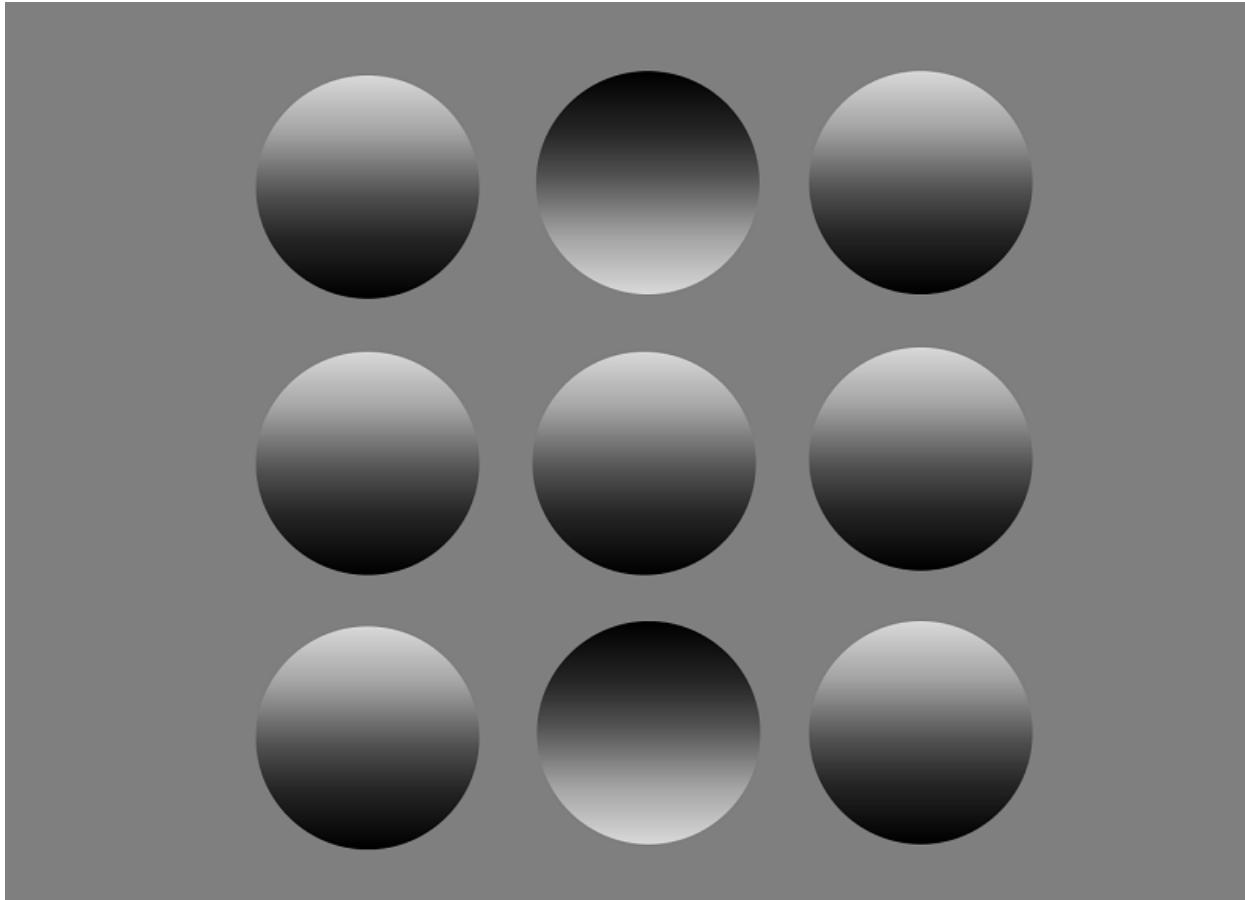
Source: Wikipedia (from the Lunar and Planetary Institute: <http://www.lpi.usra.edu>)

This is a picture of a crater

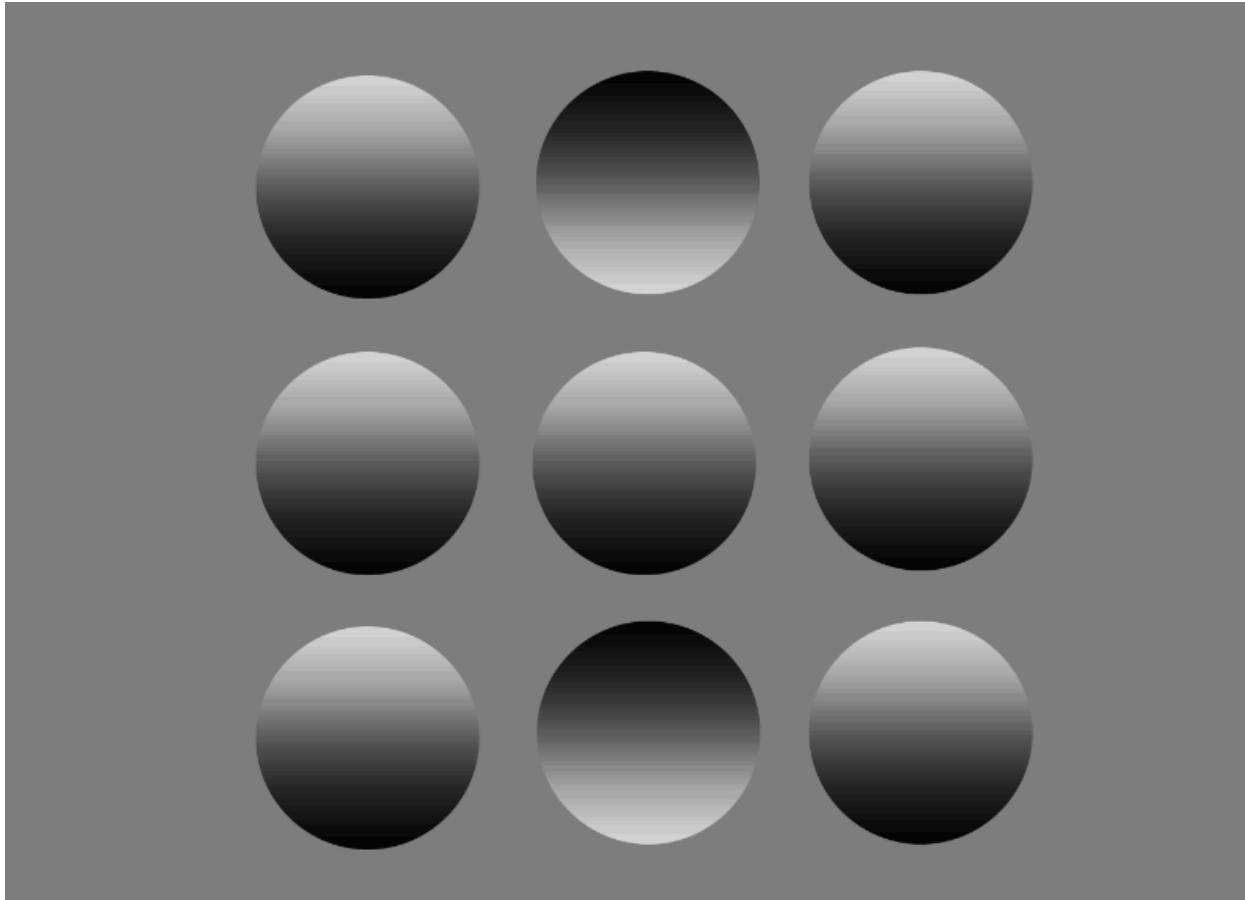


Source: Wikipedia (from the Lunar and Planetary Institute: <http://www.lpi.usra.edu>)

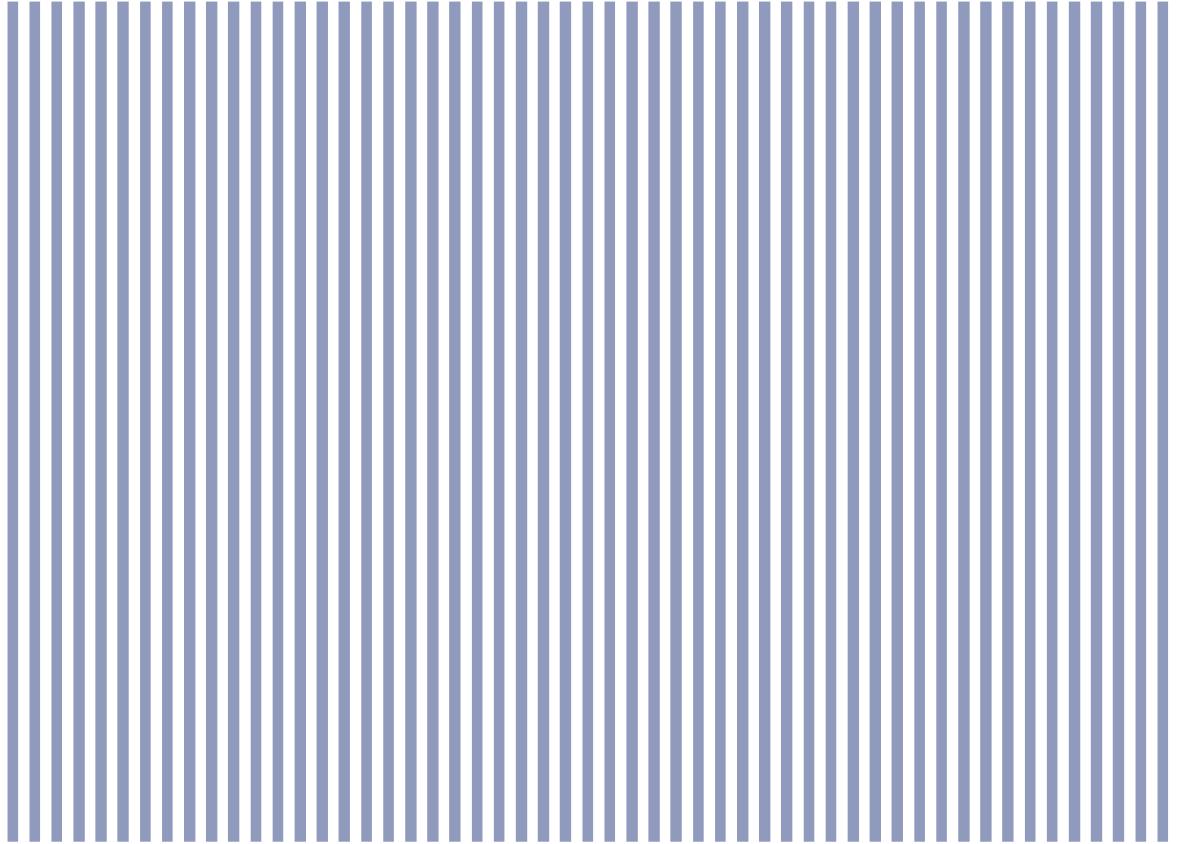
This is a picture of a crater



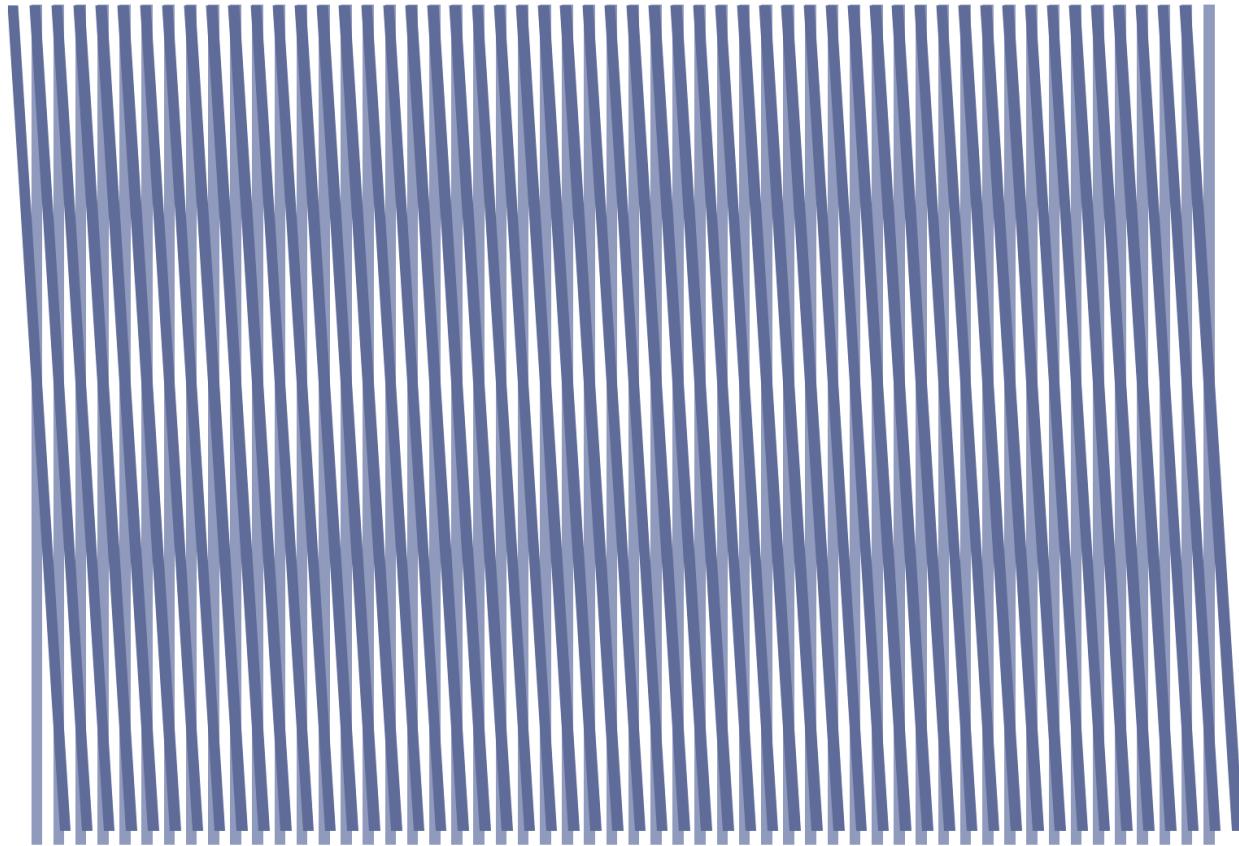
Dots with shadows



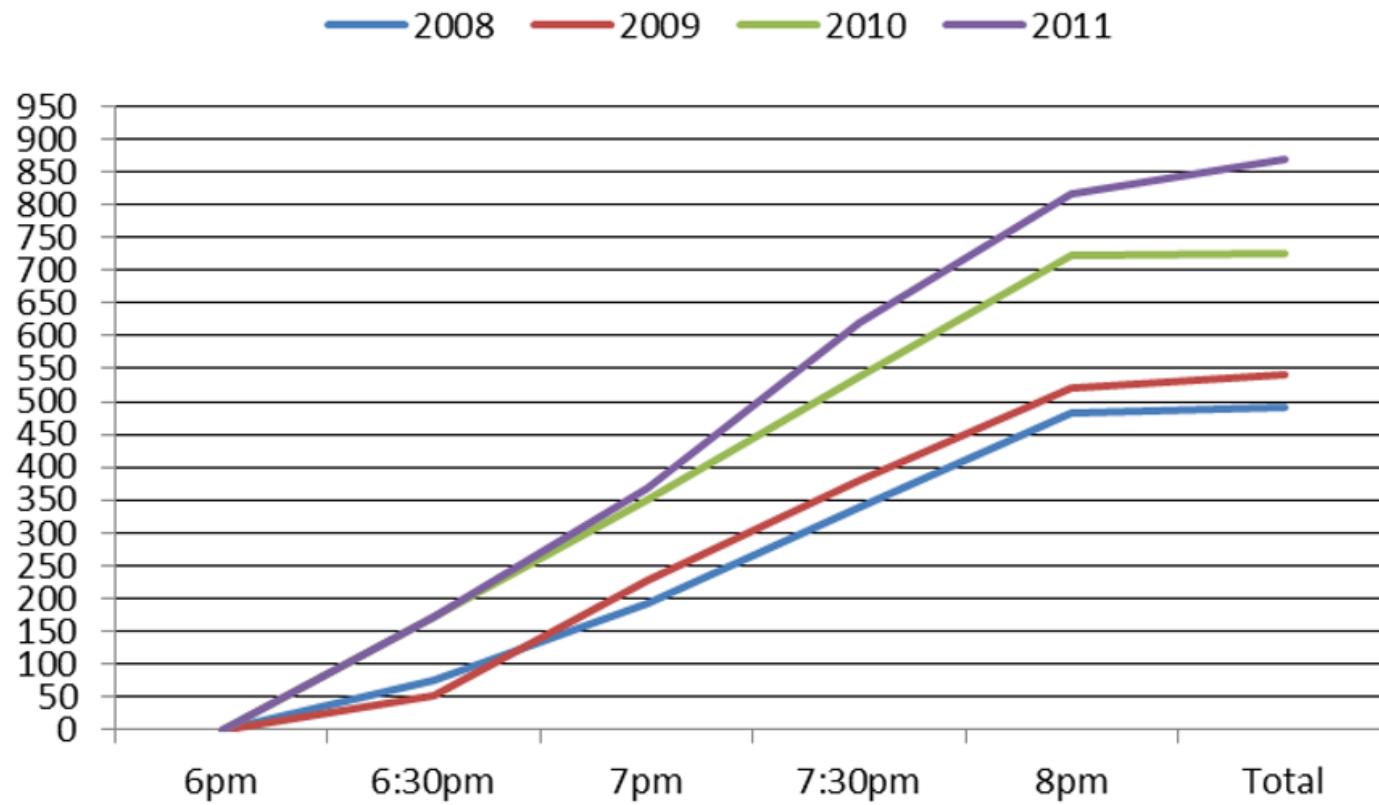
Dots with shadows



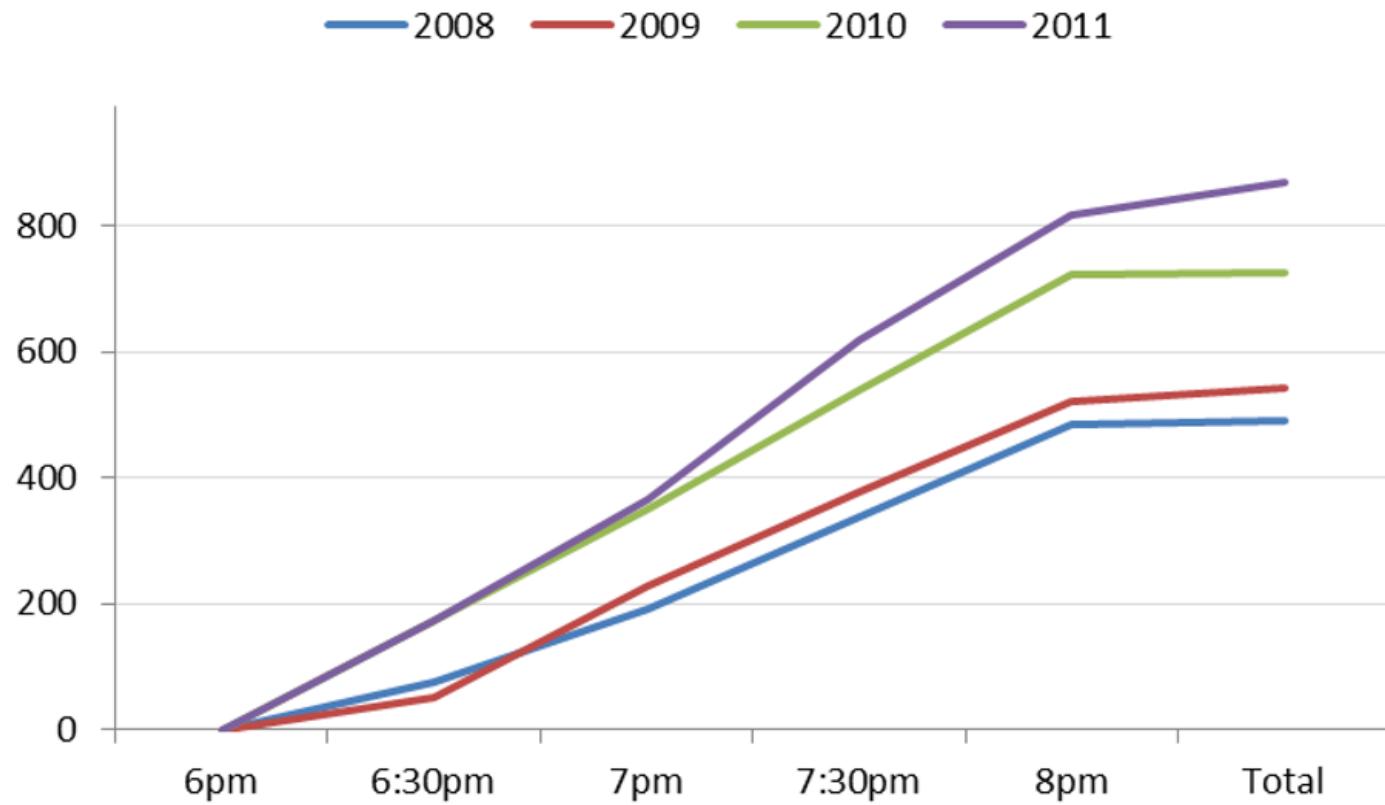
The Moiré effect



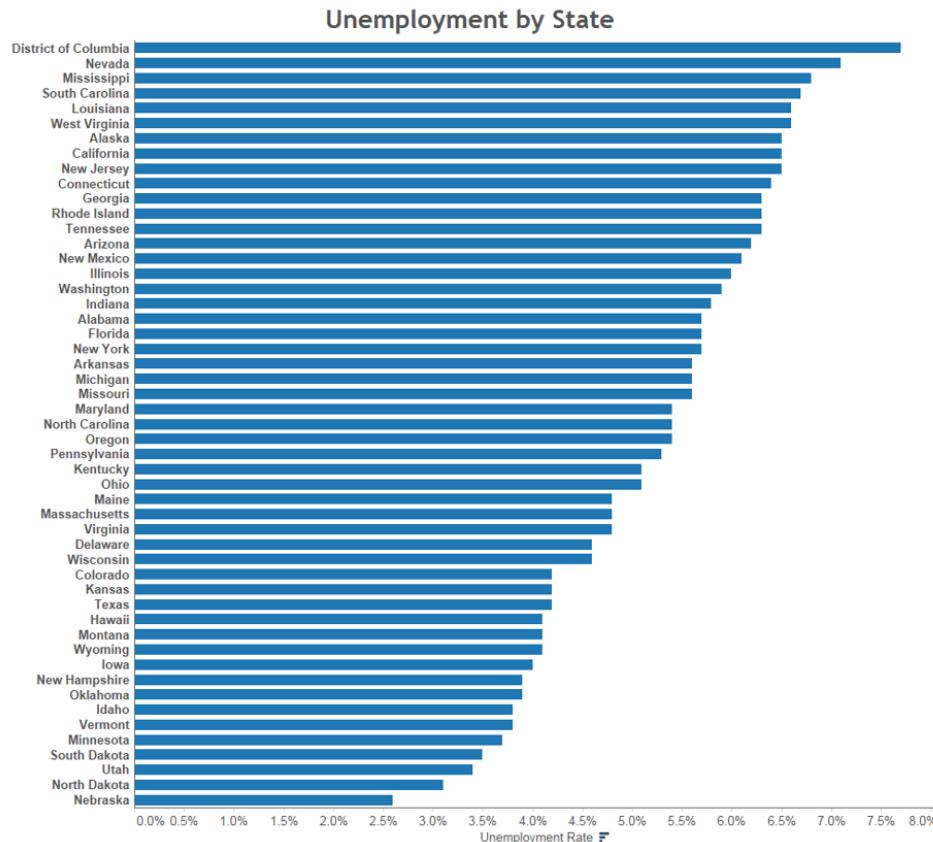
The Moiré effect



The Moiré effect in a timeline



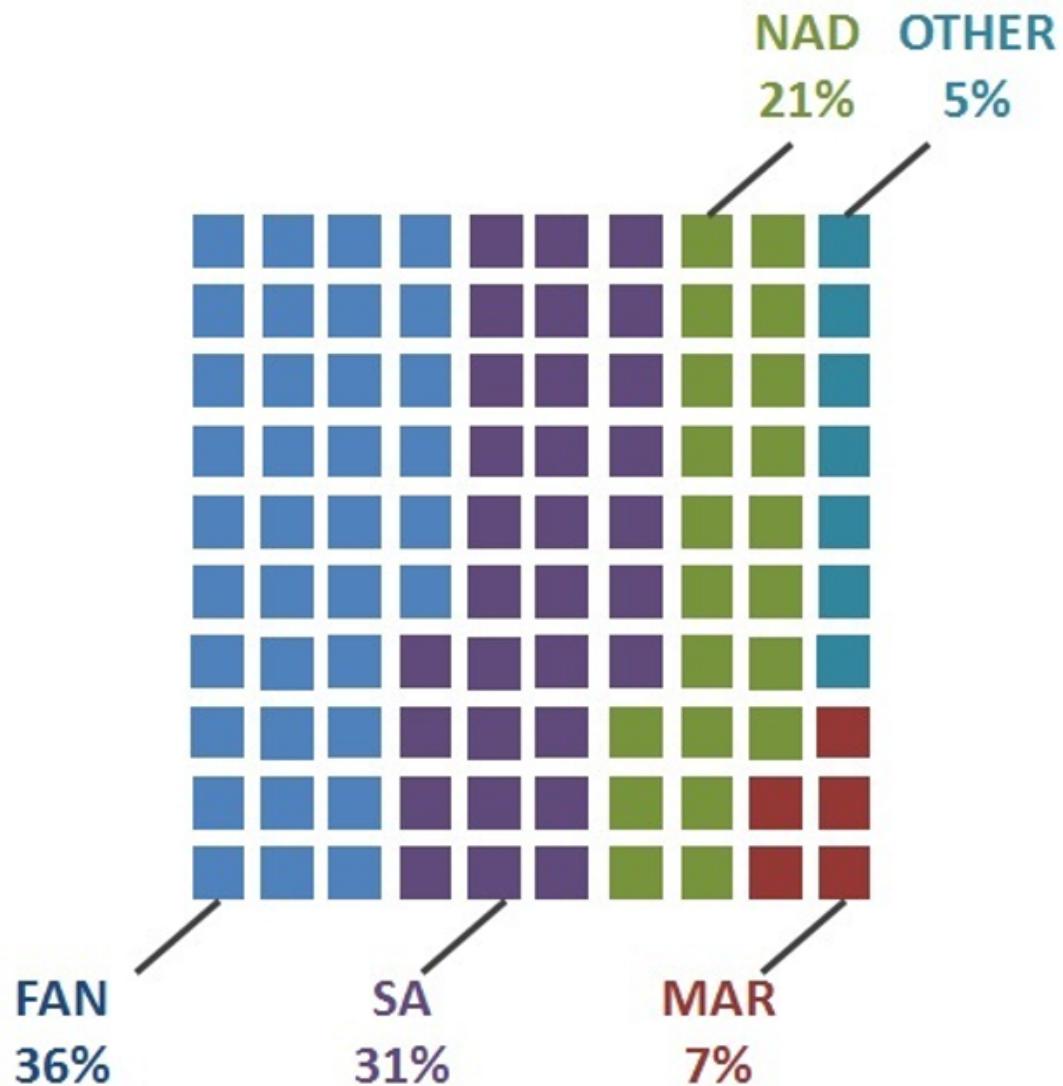
The Moiré effect in a timeline



The Moiré effect in a barchart



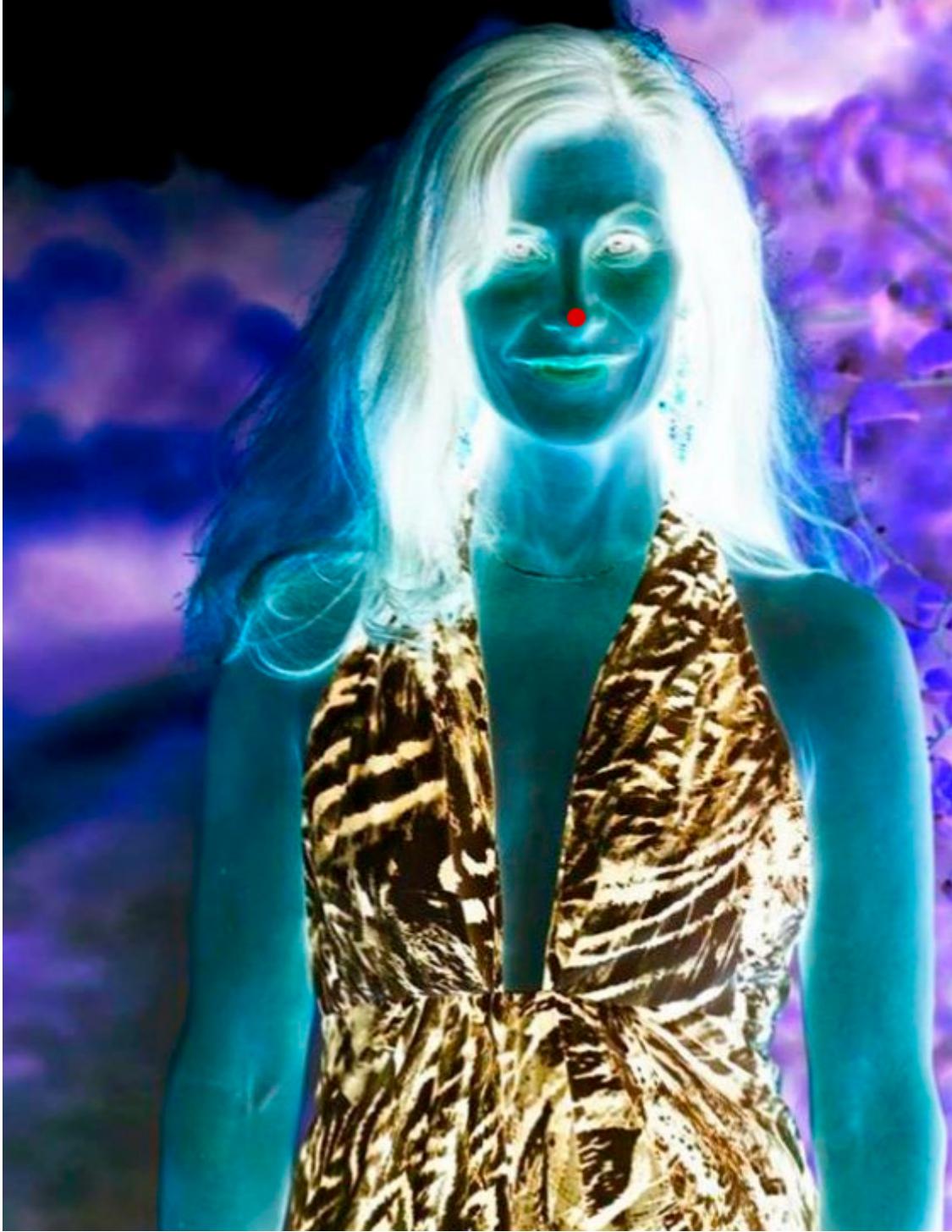
The Hermann effect



The Hermann effect in a unit chart







3.2 Common design mistakes

- Exceeding the boundaries of a single screen
- Supplying inadequate context for the data
- Displaying excessive detail or precision
- Choosing a deficient measure
- Choosing inappropriate display media
- Introducing meaningless variety
- Using poorly designed display media

continues...

3.2 Common design mistakes

...continued

- Encoding quantitative data inaccurately
- Arranging the data poorly
- Highlighting important data ineffectively or not at all
- Cluttering the display with useless decoration
- Misusing or overusing color
- Designing an unattractive visual display

3.3 Key goals in the visual design process

From previous section:

- make simpler graphs (declutter)

3.3 Key goals in the visual design process

From previous section:

- make simpler graphs (declutter)

Perfection is achieved, not when there is nothing more to add, but when there is nothing left to take away.

Antoine de Saint-Exupery

3.3 Key goals in the visual design process

From previous section:

- make simpler graphs (declutter)
- highlight observations
- add attributes/variables as context or statistical information

3.3 Key goals in the visual design process

In other words:

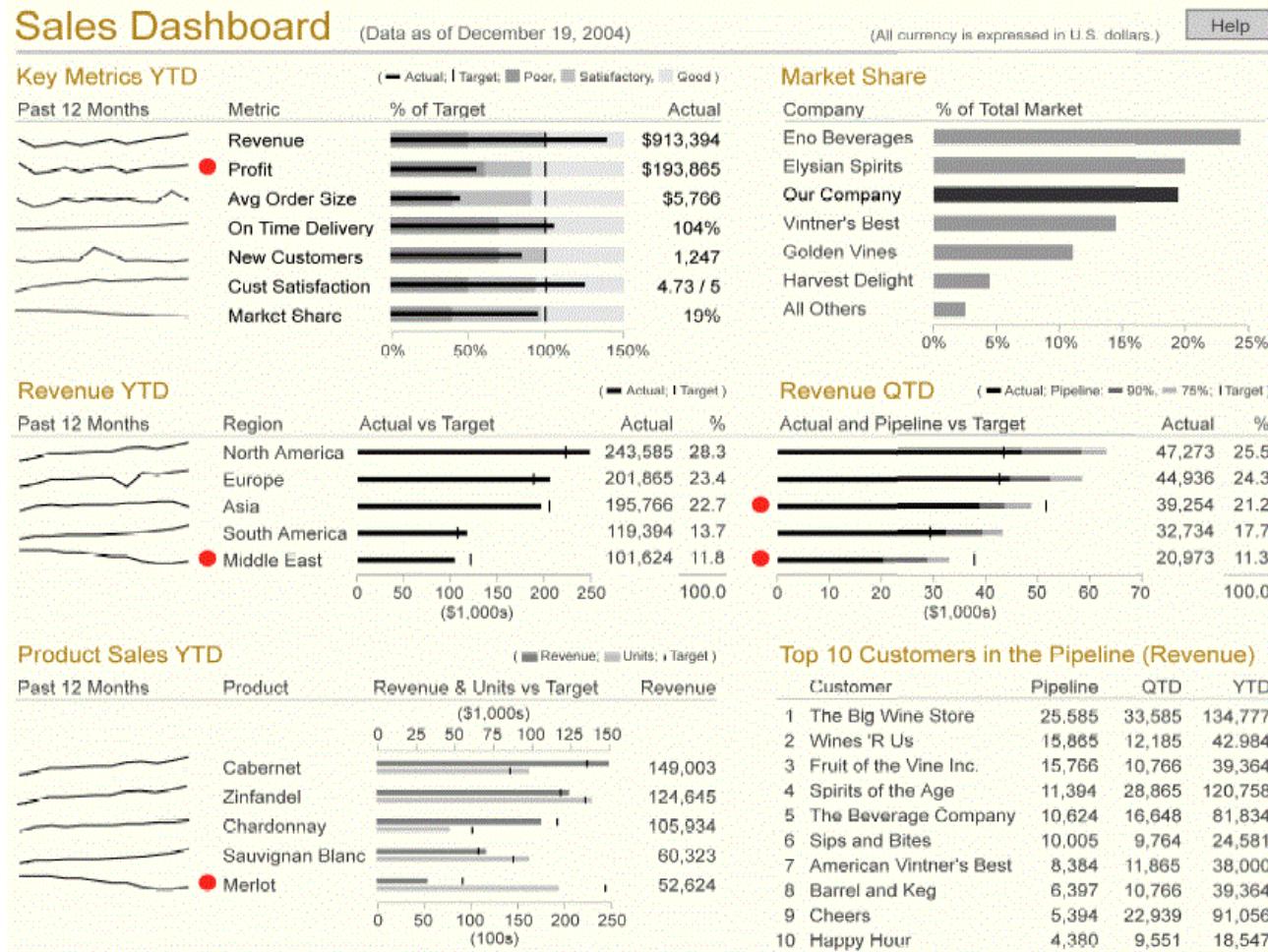
1. Reduce non-data pixels

- eliminate all unnecessary non-data pixels
- de-emphasize and regularize the non-data pixels that remain

2. Enhance data pixels

- eliminate all unnecessary data pixels
- highlight the most important data-pixels that remain

3.4 Example



Few 2013

Dashboards in Tableau

Dashboards in Tableau are containers of *sheets* of graphs.
Allow for quite basic but functional formatting.

Tableau: Actions

Some degree of interactivity with Actions : highlight and filter

Tableau: What else?

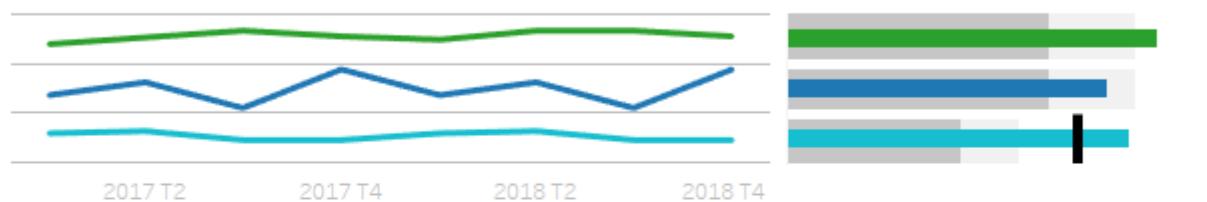
Calculated fields

...

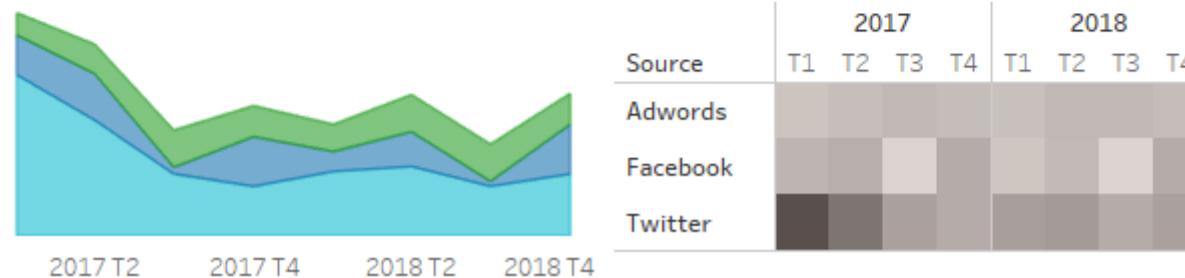
Tableau: Final styling

Marketing campaigns results Last 2 years

Income trend



Returns on investment



Final look of our dashboard

Epilogue

What? I want more!

- Tableau, [Free Training Videos](#)
 - Tableau, [Análisis visual: como aprovechar las características cognitivas humanas para comprender tus datos](#)
- Coursera, [Visual Analytics with Tableau](#)
- Udemy, (Ranked) [Data Visualization Courses](#) or [Tableau Courses](#)

References

Abela, Andrew (2006). [Choosing a good chart](#).

Few, Stephen (2009). [Dashboard Design for Real-Time Situation Awareness](#) [White Paper]

— (2013). *Information Dashboard Design*. Analytics Press: [316.763 F 44 s](#)

Kirk, Andy (2016). *Data Visualisation: A Handbook for Data Driven Design*. SAGE: London [316.763 K 63 a](#)

Munzner, Tamara (2015). *Visualization Analysis and Design*. CRC Press: Boca Raton, Florida [316.763 M 92 t](#)

Tufte, Edward R. (1983). *The Visual Display of Quantitative Information*. Graphics Press: California [316.763 T 87 e](#)

Thank you!

This presentation is available at

<http://mrn.bz/MUMA2019>

Miren Berasategi
miren.berasategi@deusto.es

License



Copyright © 2018 University of Deusto

This work (except for the quoted images, whose rights are reserved to their owners) is licensed under the Creative Commons “Attribution-ShareAlike” License. To view a copy of this license, visit

<http://creativecommons.org/licenses/by-sa/3.0/>