

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

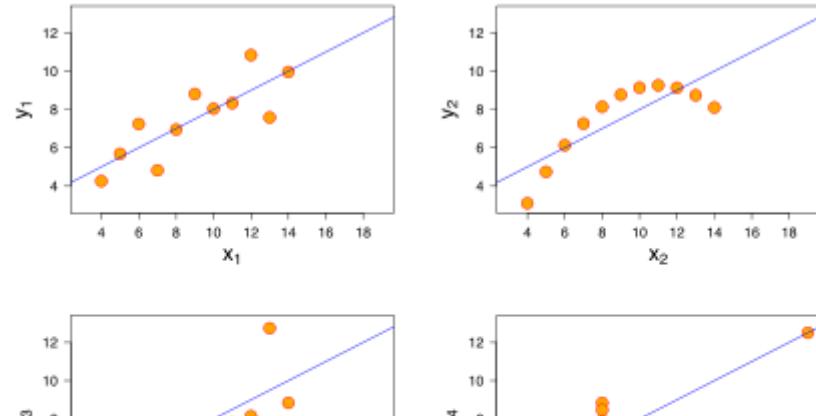
- Sight is our most developed sense
- The visual system provides a very high-bandwidth channel to our brains
- A significant amount of visual information processing occurs in parallel at the preconscious level
- The human brain is *trained* to identify visual patterns
- Summary statistics have the intrinsic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

Speaker notes

Implications of visual perception relevant to visualization design on next section

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		



Speaker notes

Traditional summary statistics can be misleading. These datasets share almost identical mean, variance, correlation and linear regression lines.

1. Shows *normal* distribution
2. There is correlation, but it's not linear
3. There is correlation, it IS LINEAR, but different from what emerged from the data
4. No relationship whatsoever. Outlier is enough for *apparent* correlation

Bottom line: summary statistics, although very sensitive to outliers, are important, but plotting the data (visualization) is also a necessary step during the first stages of the data analytics process, before making any assumptions.

What to use visualization for



Speaker notes

What are we using visualization for?

Important to answer to this question in an **abstract form**, instead of domain-specific.

We are going to divide into

- *targets* = nouns
- *actions* = verbs

What to use visualization for



→ All Data

- Trends
- Outliers
- Features

→ Network Data

- Topology

Speaker notes

What is going to be the OBJECT of visualization

- Very broadly relevant for all kinds of data:
 - trends: patterns. Increase, decrease, peaks, troughs, plateaus
 - outliers: some data just doesn't fit well with the backdrop / the rest
 - features: any particular structures of interest
- One attribute/variable. Frequent targets of interest:
 - find an individual value, distribution of all values, extremes
- Many attributes:
 - dependency: values for one attribute directly depend on those of a second
 - correlation: there is a tendency for the values of a second to be tied to those of the first
 - similarity
- The rest (network/spatial) are specific to certain types of datasets

The abstract task of understanding **trends, outliers, distributions and correlations** are extremely common reasons to use data visualization.

What to use visualization for

Actions

Analyze

→ Consume

→ Discover

→ Present

→ Enjoy

Search

	Target known	Target unknown
--	--------------	----------------

Speaker notes

WHAT is done to that object

- Analyze data:
 - consume (most common use case):
 - discover: find new knowledge that was not previously known (**generate** a new hypothesis or **verify** an existing one)
 - present: known info, i.e. insights, a story... communicate something specific and already understood to an audience
 - enjoy: for the fun of it
 - produce: annotate (add graphical or textual notes associated with one or more visualization element), record (capture steps), derive (produce new data elements)
- Search for an element of interest (based on knowledge of identity & location)
 - lookup: humans in tree vis of species of mammals
 - locate: where are rabbits? lagomorphs – not rodents
 - browse: when users don't know exactly what they're looking for but have an idea of characteristics
 - explore: not sure of either one
- Once a target or set of targets is identified, query those at one of 3 scopes (progression from one to many to all targets)
 - identify (US election map example)
 - compare, more difficult, requires more sophisticated idioms
 - summarize = overview (extremely common)

What to use visualization for

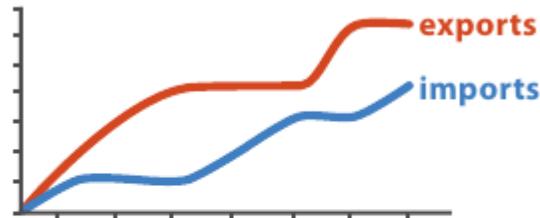
There is a strong relationship between the form of the data (the attribute/variable and dataset types) and what kinds of vis[ualization] idioms are effective at displaying it. (...) Don't just draw what you are given; decide what the right thing to show is, create it with a series of transformations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

Munzner 2014, p.50

Speaker notes

Transformation (=derive) may be required depending on the desired type of insight – derived attributes/variables extend the original dataset

What to use visualization for



Original Data



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

Derived Data

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

Speaker notes

A dataset often needs to be transformed beyond its original state in order to create a visual encoding that can solve the desired problem. **Derived attributes** extend the dataset beyond the original set of attributes.

- in some cases, the same data with a change of type (temperature vs. “hot” or “cold”)
- in other cases, access to additional info is required (geo: city name to lat/lon)
- created through arithmetic, logarithmic or statistical operations (i.e. diff field) **which can then be directly visually encoded**

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>

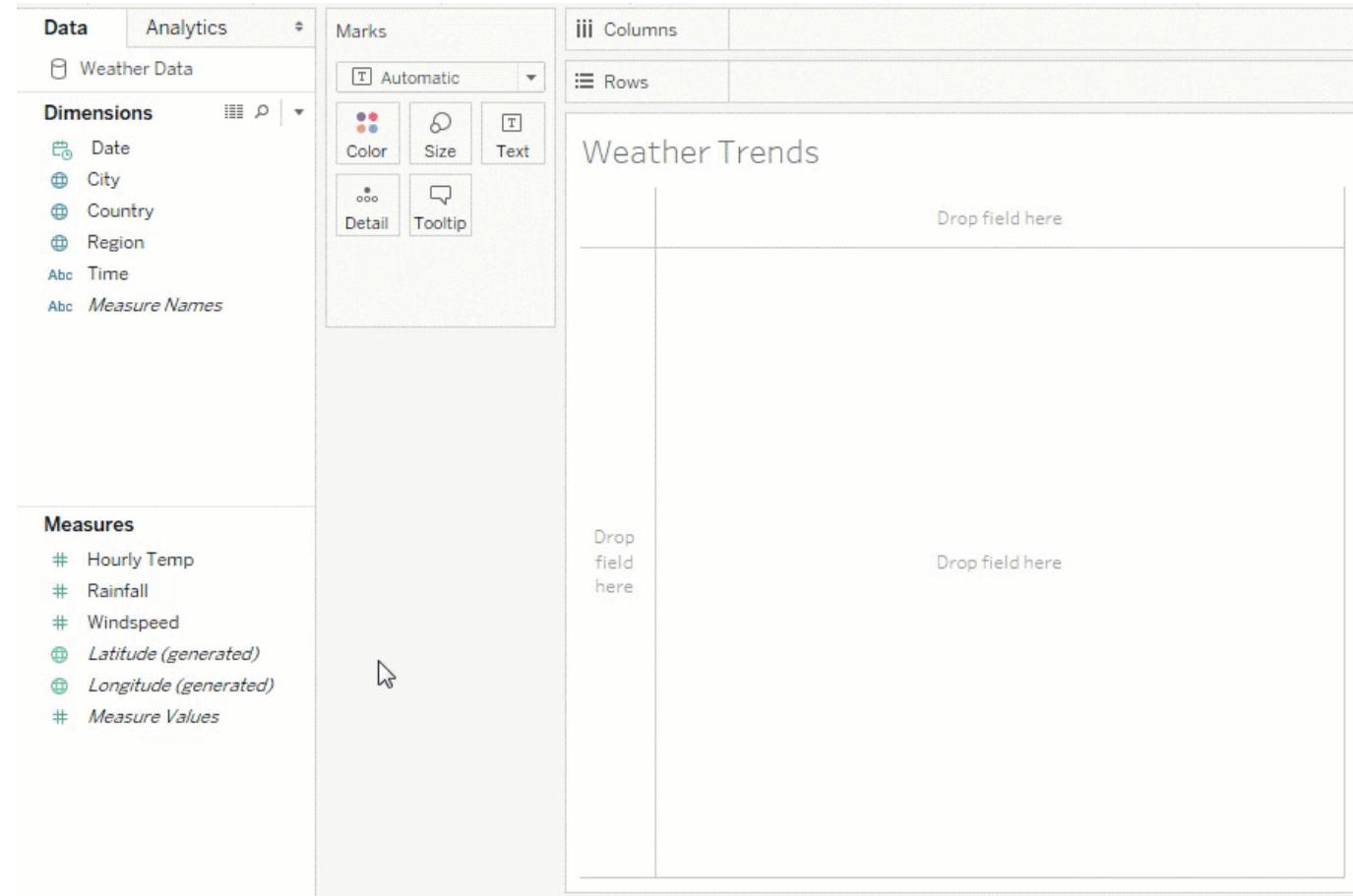
	A	B	C	D	E	F
1	source	quarter	spent	visits	income	goal
2	Adwords	20160101	1000	50000	900	1500
3	Twitter	20160101	200	8500	1300	1000
4	Facebook	20160101	500	20000	800	1500
5	Adwords	20160401	1000	48000	1200	1500
6	Twitter	20160401	300	9000	1400	1000
7	Facebook	20160401	750	21500	1400	1500
8	Adwords	20160701	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
...						

Speaker notes

Shows results of campaigns in three different sources, per quarters (trimester). Money spent in each source, number of visits got from that campaign, income generated, and goal income for each source.

Takes for granted many things, such as validity/reasonability of goals, of spent money per goal...

Tableau Desktop



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

- Data Panel:** Shows the "Weather Data" source and lists dimensions (Date, City, Country, Region, Time) and measures (Hourly Temp, Rainfall, Windspeed, Latitude (generated), Longitude (generated), Measure Values).
- Marks Panel:** Set to "Automatic" and includes options for Color, Size, Text, Detail, and Tooltip.
- Columns Panel:** A header section with a "Drop field here" placeholder.
- Rows Panel:** A header section with a "Drop field here" placeholder.
- Dashboard Area:** Titled "Weather Trends" and contains a single blank white square with a "Drop field here" placeholder.

Tableau Software

Tableau Desktop

1. Load data

Speaker notes

- Possible to *merge* more sources (or **Connections**), only one for our practice – might require to define `New Union` to determine how merge takes place
- Shows the **only sheet** in the file, may be more
- Connection > Live/Extract , first one allows to modify source file and update work accordingly

Interprets the `source` field as `Abc` , rest as `#` (well done). Only, we would rather improve the `quarter` interpretation as `Date` instead of `Number`

Green & blue: **blue fields** are discrete, **green fields** are continuous. We'll see implications of this later on. For now, it's all good

Rename, hide... fields

The Excel > Tableau convert is quite straightforward (does a good job), more complex sources may require more manipulation at this point

This screen allows to preview how Tableau is interpreting our dataset: we are not allowed to change values, only variable interpretations

The source file (`data.xls`) is never modified

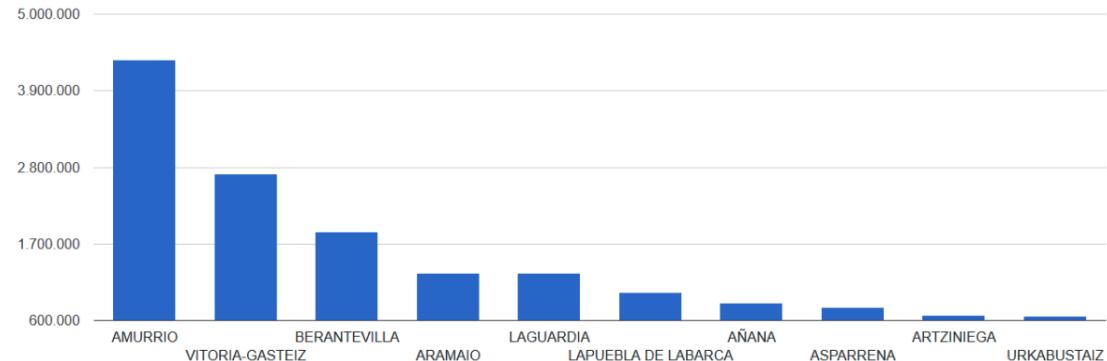
1. Graphs

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



Speaker notes

Dataset showing town name and amount of subsidies (funds) received in a given year. How do you think we could visualize this?

- visualization requires *mapping* or *translation* from *data* to *visual language* (idioms)
 - this can be done in many ways
1. subsidy amount = bar height, color irrelevant
 2. to show in map **transformation is required** (not possible directly)
 - color IS relevant (subsidy amount = color saturation)
 - comparison is more difficult: not *ordered* (althoug not impossible)
 - allows for more direct insight (if context - knowledge of the area): Vitoria-Gasteiz not getting the higher amount
 3. anything else? (i.e. pie chart, treemap...)

1.1 Reminder: variable types

- Quantitative
 - Continuous
 - Discrete
 - Qualitative
 - Categorical
 - Ordinal
- 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

Speaker notes

- continuous variables can take on infinite/uncountable values (time, real numbers)
- discrete variables can take a finite/countable number of values (periods of time, integers)
- categorical/nominal: two or more categories, any order (gender, political party...). Cannot be quantified, do not allow arithmetic operations, cannot be assigned any order
- ordinal: allows for rank order (1st, 2nd, 3rd...). Can be dichotomous (yes/no, right/wrong, sick/healthy...) or non-dichotomous (spectrum of values: completely agree, mostly agree, mostly disagree, completely disagree)

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



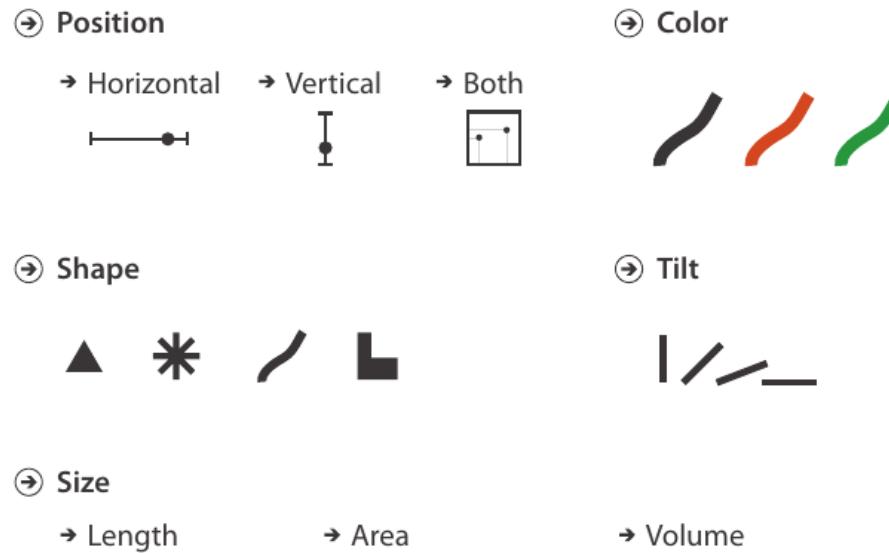
Marks are geometric primitives (Munzner 2014, p.96)

Speaker notes

- a zero-dimensional (0D) mark is a point
- a one-dimensional (1D) mark is a line
- a two-dimensional (2D) mark is an area
- a three-dimensional (3D) mark is possible but not frequently used (will see why)

1.2.2 Channels

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



Speaker notes

- spatial position (_where_)
- color, with three distinct aspects: hue, saturation, luminance (more in the next slide)
- size, one for each added dimension: length 1D, area 2D, volume 3D
- angle/tilt/slope
- others are shape, curvature

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.

Speaker notes

more channels are “used up”, so less attributes can be encoded in total,

BUT the attributes that are shown will be very easily perceived (i.e. color of political party for bar charts)

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

Speaker notes

Area marks cannot typically be size or shape coded: state or country already has a certain size and shape

- lines: if length is *taken* by a variable, it can't be used for another one, but width can be used to size code. They can be made wider on an individual basis to encode an additional attribute, or an entire set of bars can simply be made wider in a uniform way to be more visible
- points: intrinsically convey information only about position and are exactly the vehicle for conveying additional information through area and shape (and color!)

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.

Speaker notes

*The human perceptual system has two fundamentally different kinds of sensory modalities. The **identity** channels tell us information about what something is or where it is. In contrast, the **magnitude** channels tell us how much of something there is (Munzner 2014, p.99).*

1.2.4 Using marks and channels

All channels are not equal.

The selection of marks and channels should be guided by

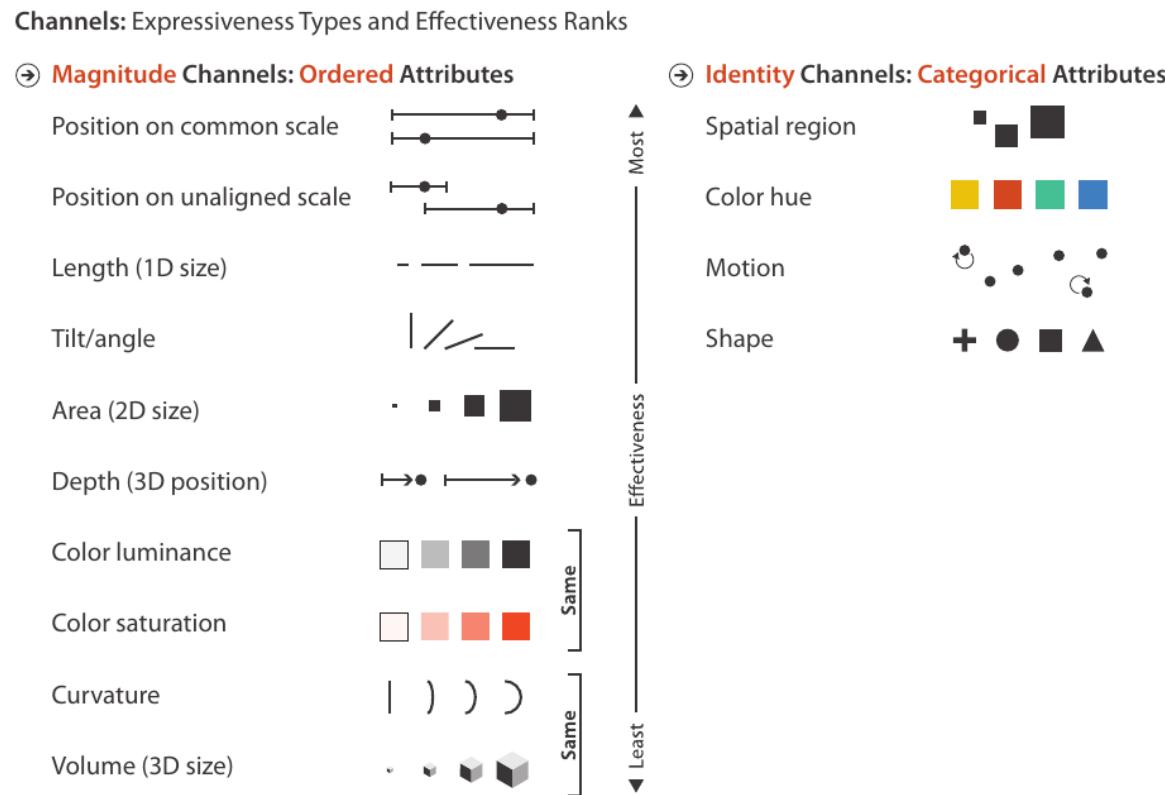
Speaker notes

*the same data attribute encoded with two different visual channels will result in different information content in our heads after it has passed through the perceptual and cognitive processing pathways of the human visual system
(Munzner 2014, p.100)*

- expressiveness: the visual encoding should express all of, and only, the information in the dataset attributes. Ordered data should be shown in a way that we perceive as ordered; unordered data SHOULD NOT be shown in a way that implies an ordering that does not exist.
- effectiveness: the importance of the channel (task abstraction, targets and actions) should match its salience, how noticeable it is. The most important attributes/variables should be encoded with the most effective channels, and less important attributes can be matched with less effective channels.

These can be combined to create a ranking of channels according to the type of data that is being visually encoded.

1.2.4 Using marks and channels



Speaker notes

Channels related to spatial position are at the top of both lists, and they are the only ones appearing on both lists (none of the others are effective for both data types). This primacy applies only to 2D positions, 3D depth is a much lower-ranked channel

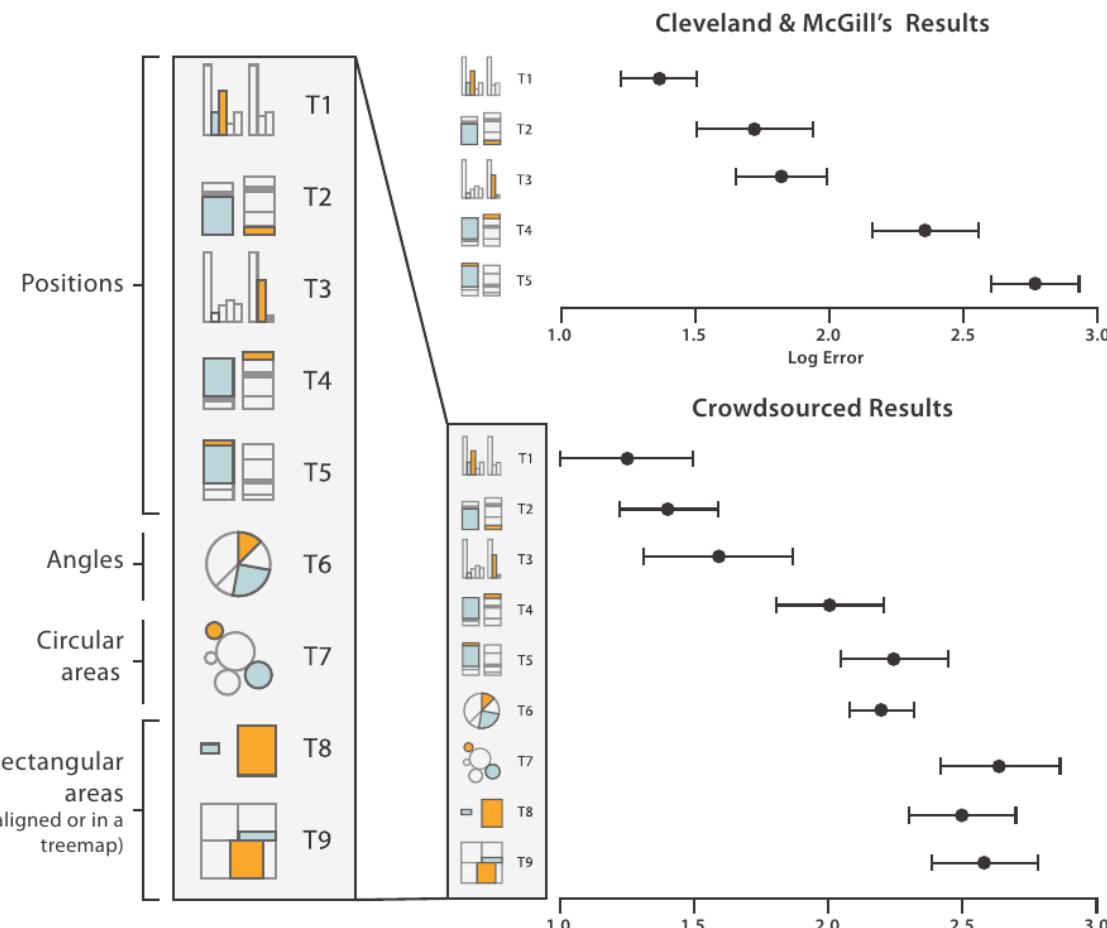
1.2.4 Using marks and channels

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

Speaker notes

the attributes encoded with position will dominate the user's mental model –their internal mental representation used for thinking and reasoning– compared with those encoded with any other visual channel.

1.2.4 Using marks and channels

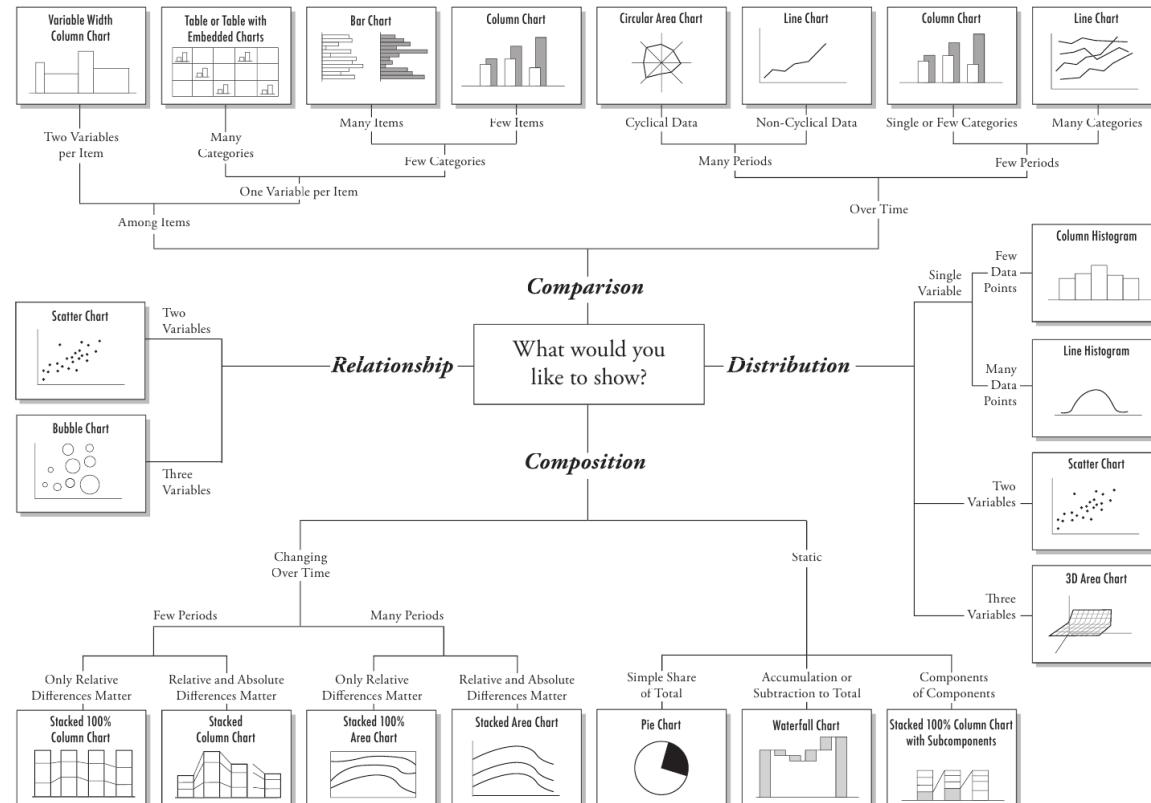


Speaker notes

Perceptual accuracy of each channel type: how accurate is the interpretation made by users depending on the type of channel used to represent certain data

1.3 So, which graph?

Chart Suggestions—A Thought-Starter



Speaker notes

Variable types and insights (Munzner's "targets") as *ingredients*, which *recipes* can be used to cook data visualization?

There are many tools for this, represented only one.

Speaker notes

Left column, Data tab:

- dimensions contain qualitative values
- measures contain numeric, quantitative values (you can apply calculations to them and aggregate them)

Reminder (blue vs. green, [learn more](#)):

- green = continuous. Its values are treated as an infinite range. Generally, continuous fields add axes to the view.
- blue = discrete. Its values are treated as finite. Generally, discrete fields add headers to the view.

Mapping variables to graphs is done, in Tableau, by dragging items from the left column to either: `Columns / Rows` , or `Marks` .

`Show me` tab displays the most common graphs and the minimum requirement of data for each graph. If `graphs = recipes`, lets us know which ingredients are required to cook a certain recipe.

Also works the other way round: if I select my ingredients from the left bar, the `Show me` tab highlights the recipes available for those ingredients.

(test some graphs) - Source only > table - add spent > table, pie chart etc

Fields can be used more than once (redundant coding) for easier identification, by dragging to more than one place. barchart income (rows) + source (columns), source to colour

Options in `Show me` are different from the `Marks` box. `Show me` only allows graphs that *make sense*, very basic and tested graphs. `Marks` allows to *force* graphs in case we need more complex displays. USE WITH CAUTION, possible to create graphs that make no sense or are misleading (change to “line” in `Marks` box)

2. Promote insight

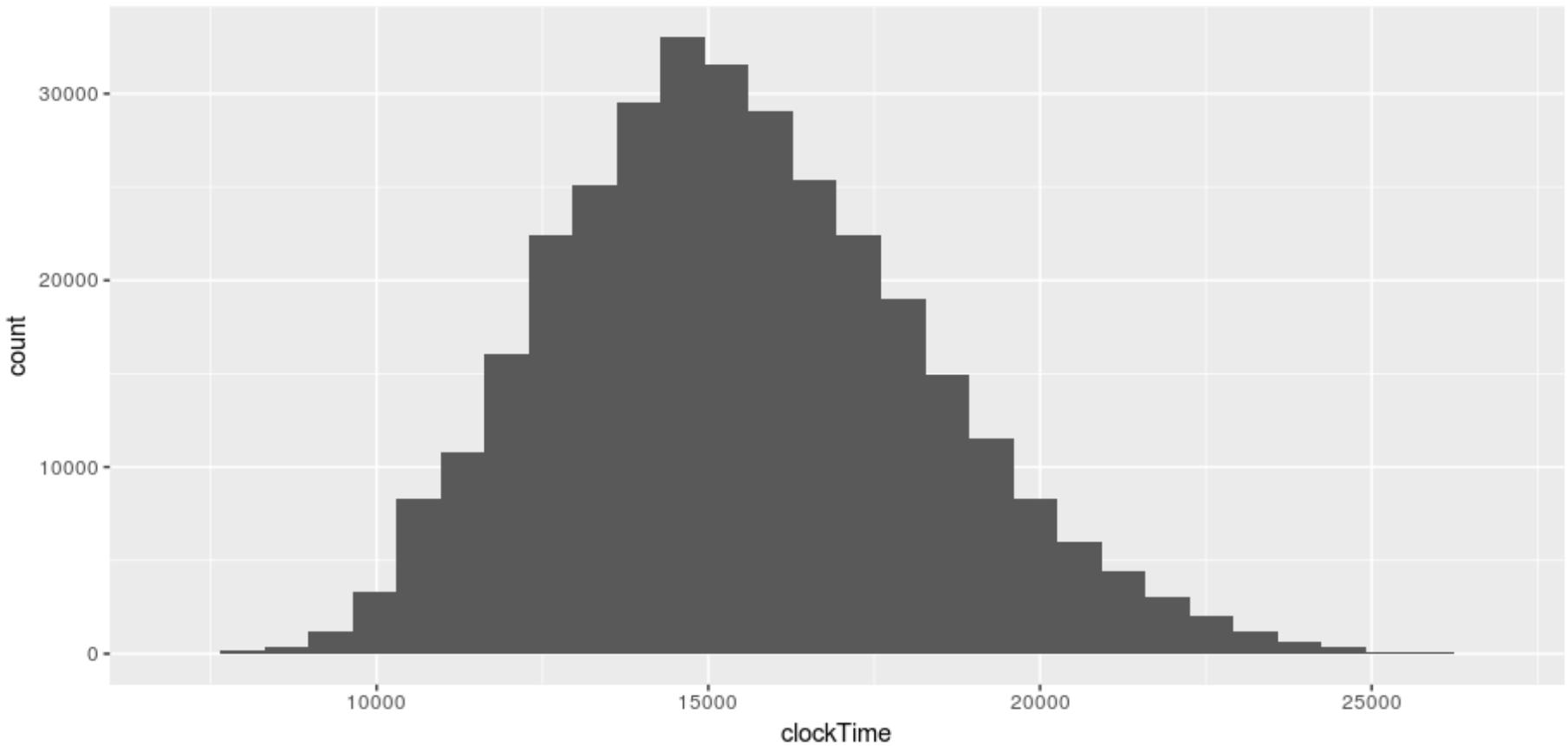
Section outline

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

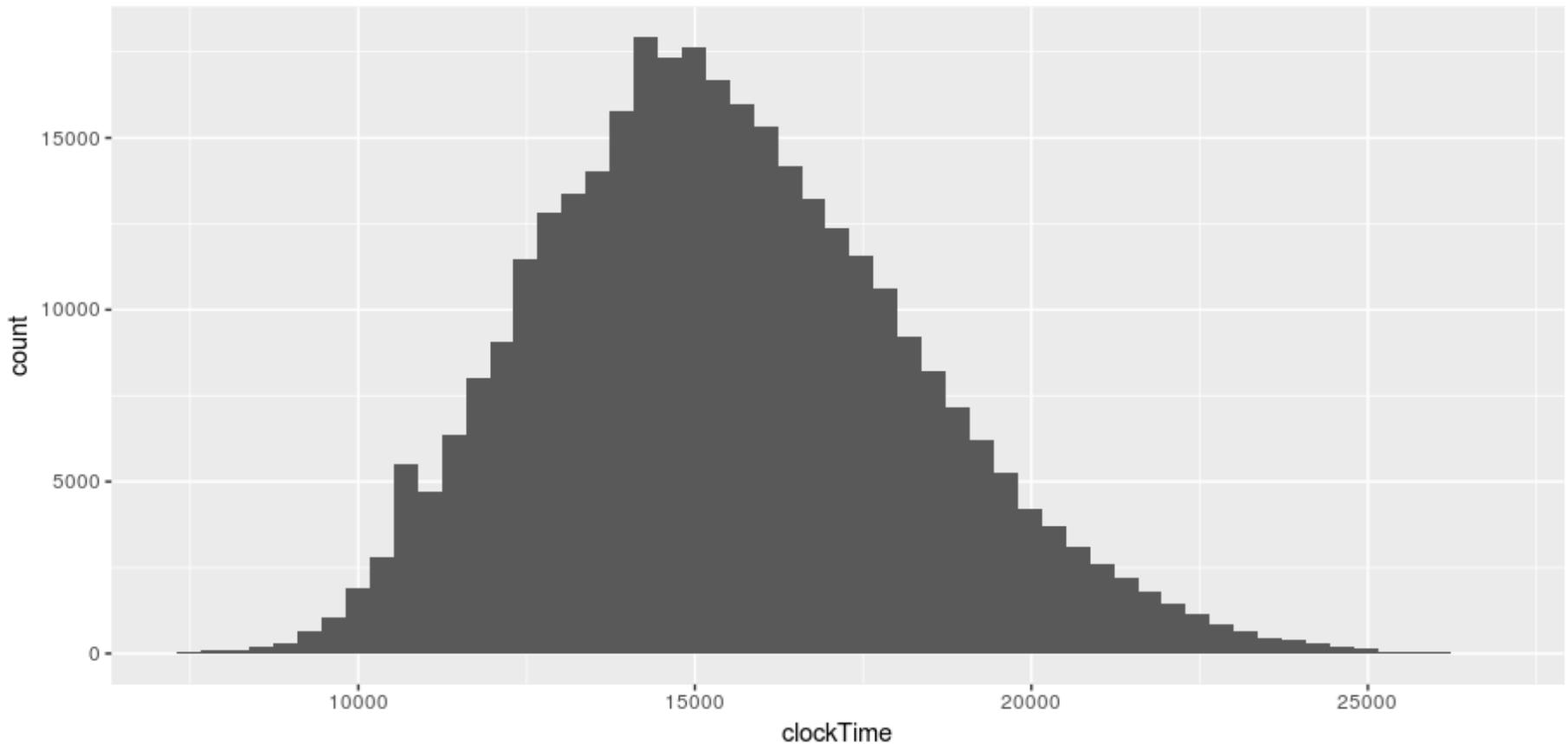
2.1 Change default settings



Speaker notes

Official results from the Berlin marathons. The default aggregation from the tool used to create this visualization presents an almost perfect normal distribution.

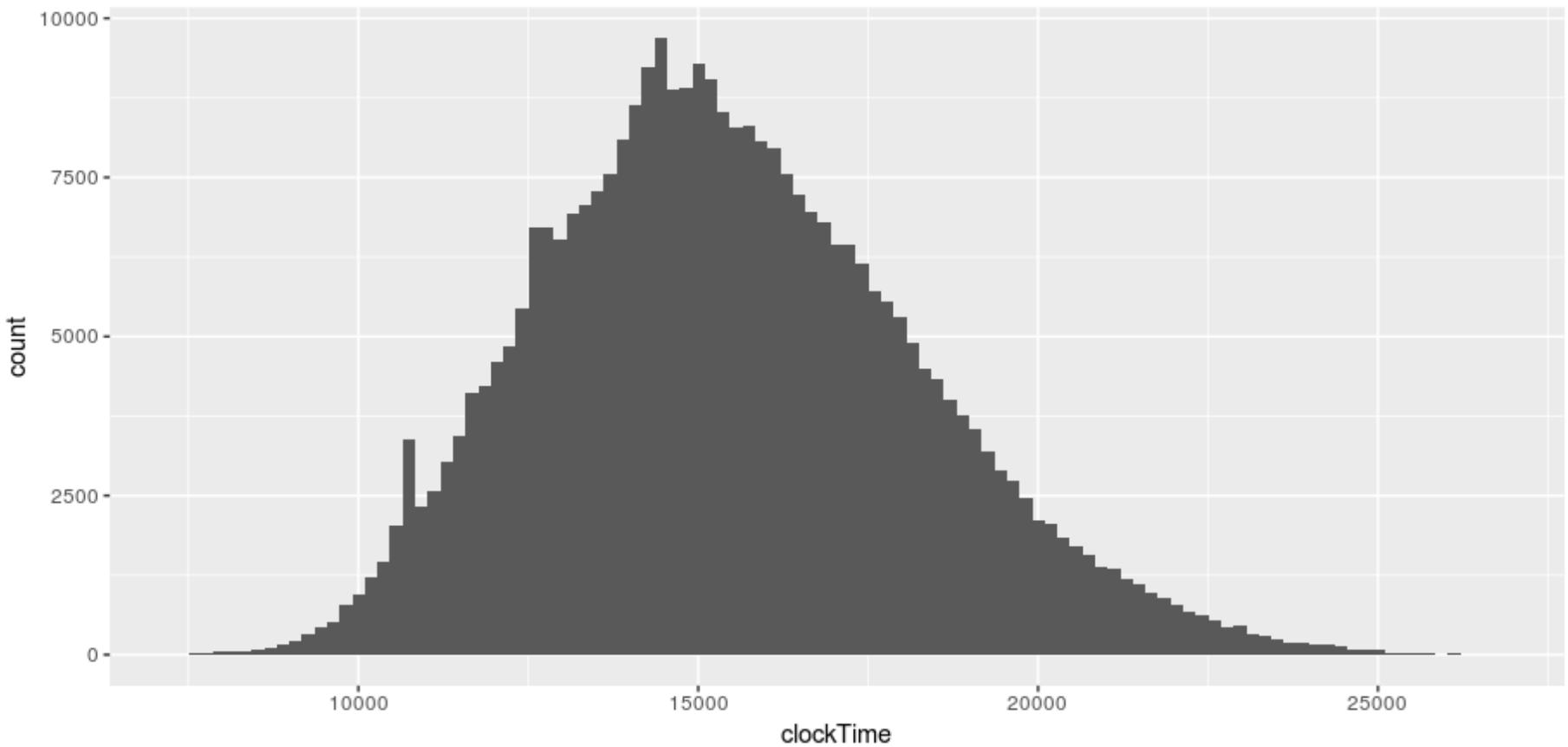
2.1 Change default settings



Speaker notes

More detail (less aggregation) allows to see some highlighted bars

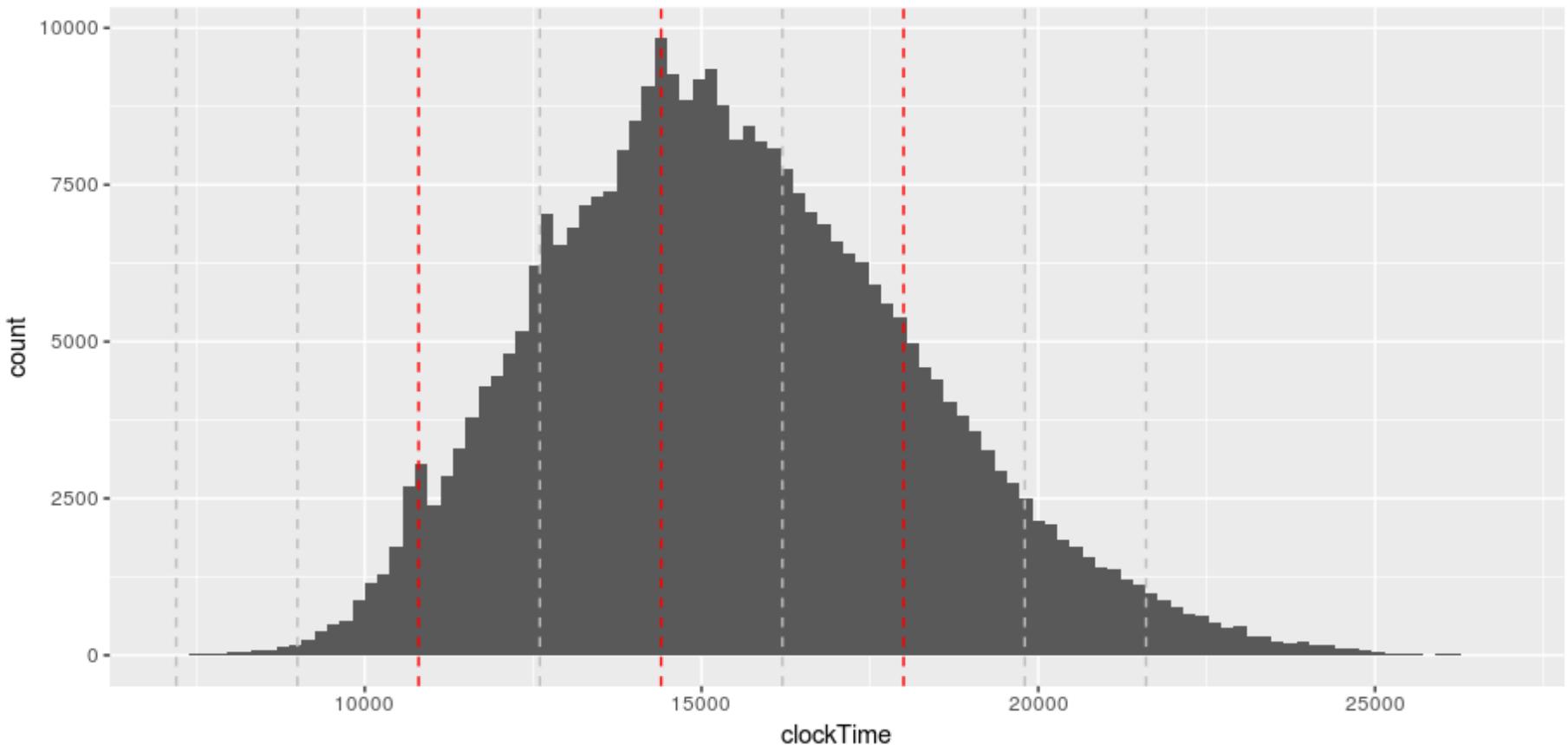
2.1 Change default settings



Speaker notes

Even more detail (less aggregation) makes those bars stand out more and new ones appear: what is happening here?

2.1 Change default settings



Speaker notes

vertical dashed lines show hour marks in red (3h, 4h, 5h, main goal times for marathon runners), half-hour marks in grey. People try to fit into their closest *rounded time*...

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

Speaker notes

we should remove all non-data-ink and redundant data-ink, within reason, to increase the data-ink-ratio and create a sound graphical design.

some redundancy is often more effective, however, most graphics don't struggle with understatement. In fact, most contain a stunning amount of excess ink (or pixels). Rather than dressing our data up we should be stripping it down.

2.2 Make simpler graphs



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

by [Joey Cherdarchuk](#)



Speaker notes

In it we start with a chart, similar to what we've seen in many presentations, and vastly improve it with progressive deletions and no additions.

The next time you are trying to improve a chart, consider what you can take away rather than what you can add.

2.2 Make simpler graphs

More on decluttering:



Nussbaumer, [Declutter Your Data Visualizations](#)

Speaker notes

6min duration video, example

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

Speaker notes

A preattentive visual property is one which is processed in spatial memory without our conscious action. In essence it takes less than 500 milliseconds for the eye and the brain to process a preattentive property of any image. These properties can be harnessed to make it easier for a user to understand what is presented through the design and save them from consciously processing all the data presented in short-term memory which requires more effort.

2.3 Highlight observations

756395068473
658663037576
860372658602
846589107830

FIGURE 4.2 Count the 3s example

Macmillan 2015 - 100

Speaker notes

There are 6. No visual cues to help count. It makes this a challenging exercise, you have to hunt through the lines of text, looking for the number 3.

What happens when we make a single change to the block of numbers. Repeat the exercise of counting the number 3s.

2.3 Highlight observations

756395068473
658663037576
860372658602
846589107830

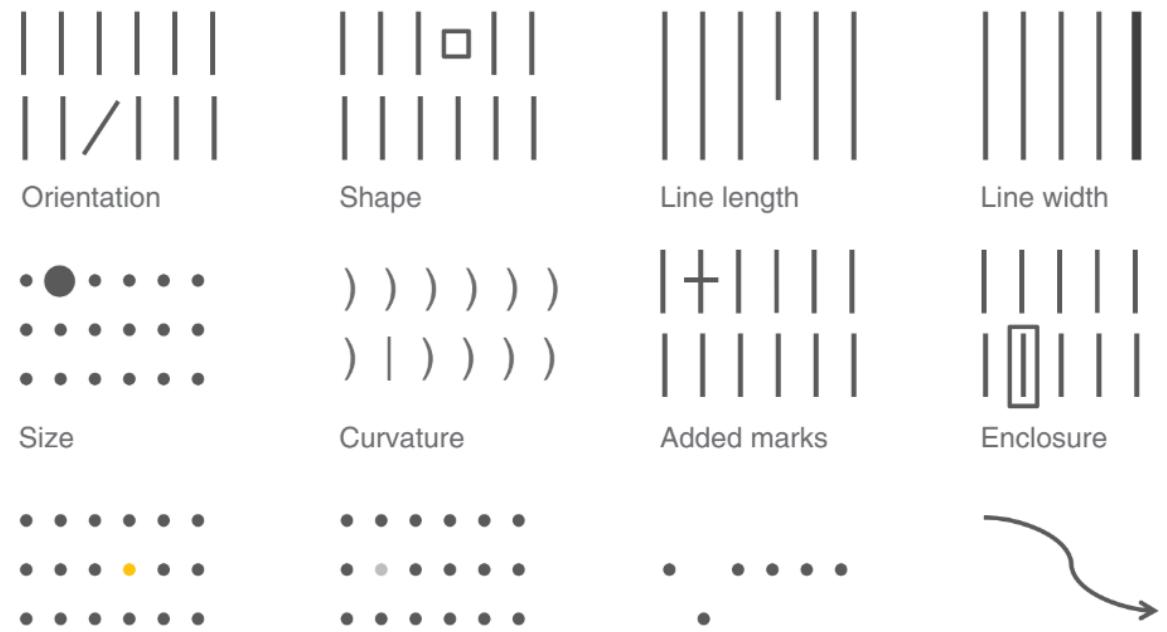
FIGURE 4.3 Count the 3s example with preattentive attributes

Nussbaumer 2015, p.104

Speaker notes

Note how much easier and faster the same exercise is. There is no time to blink, no time to think, no NEED to think. The preattentive attribute of intensity of color in this case, makes the number 3s stand out from the rest. Our brain is quick to pick them up without having to dedicate any conscious thought to it.

2.3 Highlight observations



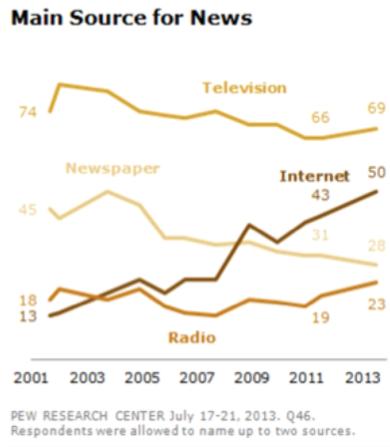
Speaker notes

your eye is drawn to the one element within each group that is different from the rest, you don't have to look for it. Our brain is hardwired to quickly pick up differences we see in the environment.

NOTE that people tend to associate quantitative values to some of the preattentive attributes (length, width, size; not colour). This tells us which of the attributes can be used to code quantitative information (line length, spatial position, or to a more limited extent, line width, size, and intensity can be used to reflect relative value), and which should be used as categorical differentiators.

2.3 Highlight observations

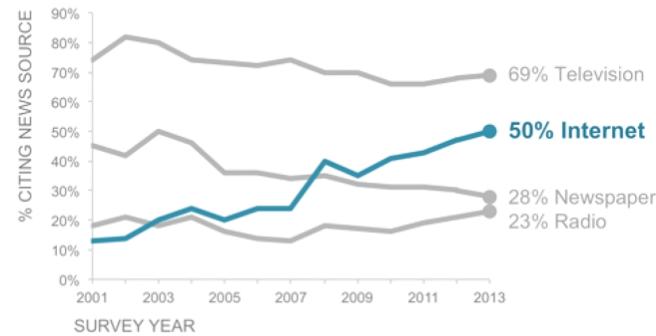
1. More Americans get news online... 50% of the public now cite the internet as a main source for national and international news, still below television, newspapers and radio. (Report)



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 

Speaker notes

After decluttering, more *channels* are available for highlight. This allows to

1. draw your audience's attention quickly to where you want them to look, and
2. create a visual hierarchy of information

2.4 Add variables (as context)

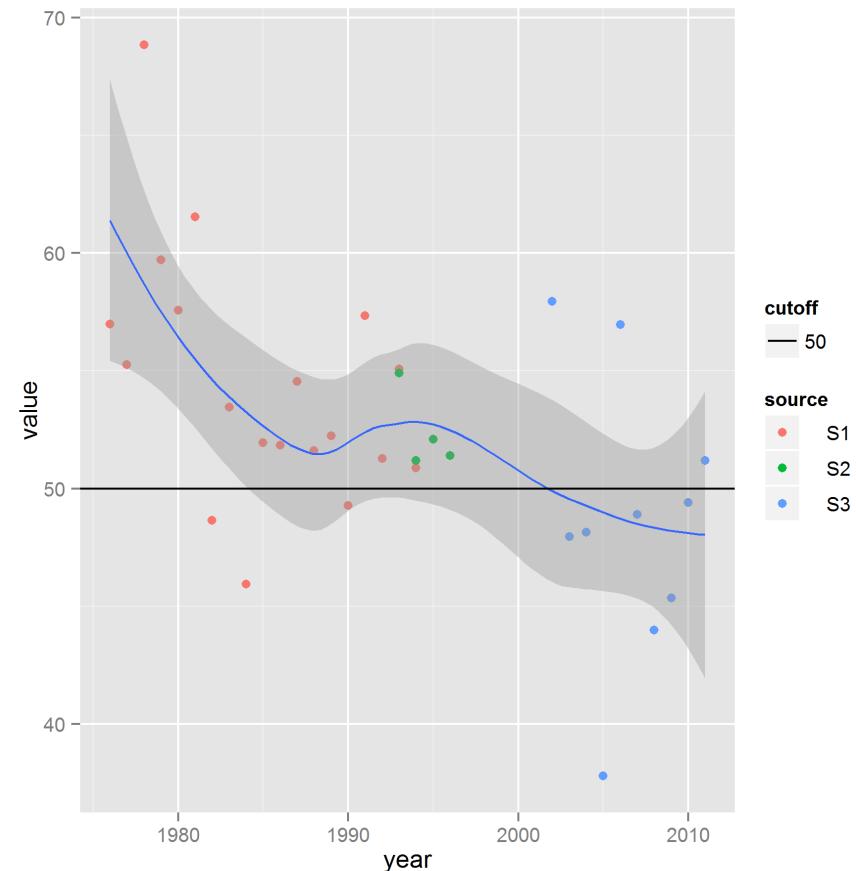
- Adding preexisting variables (in moderation)
- Creating conditional variables from preexisting variables
 - binaries or with few levels are best
 - example of calculated field or variable: weekend date

Speaker notes

useful only if addition conveys meaning / enables insight. In the graph, the colour coding is explanatory to the left, but it is not to the right.

2.5 Add statistical information

- statistical summaries
(mean, variance)
- models



source

Speaker notes

Select Quarter and Income, line graph from Show me.

Quarter: will default to higher level of aggregation (YEAR). Possible to select different aggregations (when and if attribute is correctly interpreted): year > q > month...

When selecting aggregate levels with dates (by clicking on arrow):

- top half considers date as discrete (will change to blue if selected). Useful to compare the same period of time in different years.
- bottom half considers date as continuous.

Source to rows will draw one graph per source.

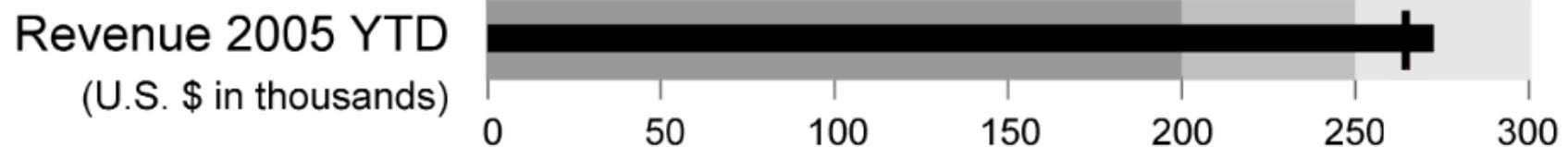
Rename sheet to Sparkline , Save.

TWEAK MORE:

- Remove axis data: right click > uncheck “Show header”
- Remove grid lines: right click > Format... Select borders, everything in Rows/Columns to “None”. Select lines, “Grid lines” to “None”
- Add label to line itself: drag Source to Detail (in Marks)
- Label to lighter gray: right click on source in Marks > Format... Pane, select color
- Color code lines by source: drag source to Color in Marks. Change defaults: down arrow > Edit colors (Tableau Classic 20)

What is the problem with this? No reference to bottom line, they may all be in the same axes. We need Row dividers.

Tableau: (not so) basic graphs



Bulletgraphs ([Few 2007](#))

Speaker notes

Source to rows, income to columns > draws bars.

Where is goal? Drag to columns > two different graphs (not what we want).

Select source + income + goal > bullet graph from Show me.

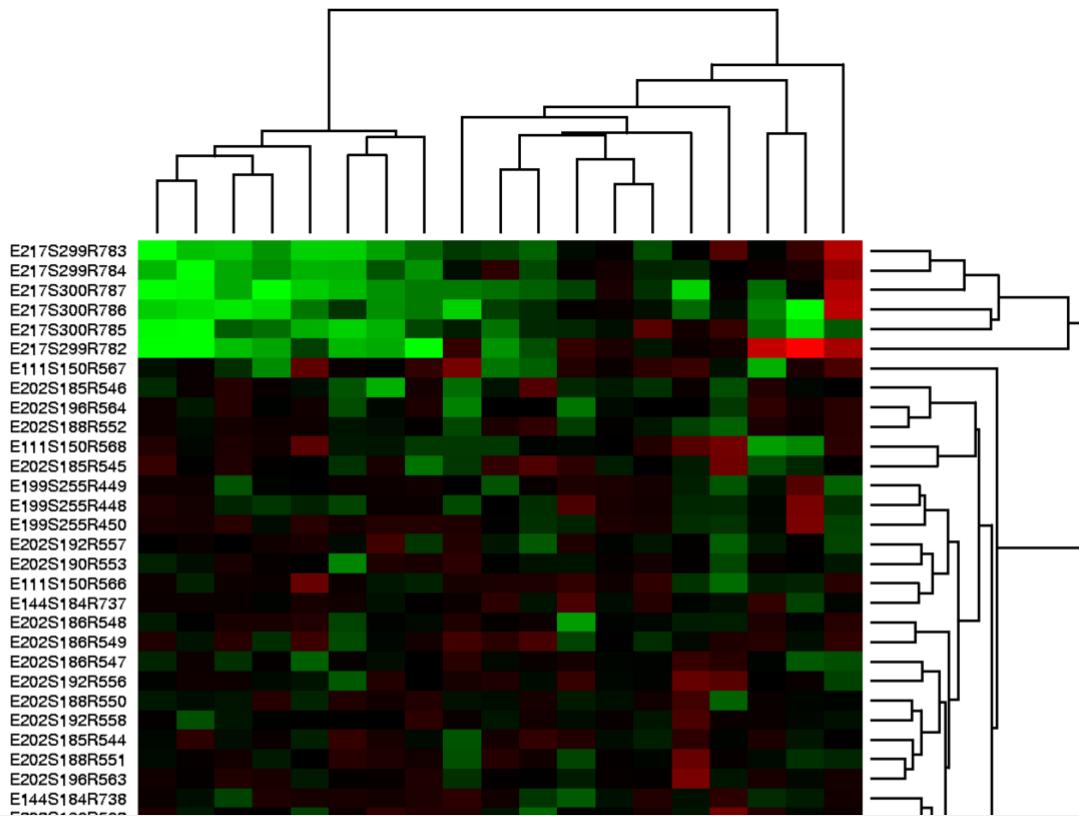
Shows totals, may not be what we want. Anyway, what did this do? Now manually:

1. source to rows, income to columns
2. now we want goal as black bar (reference line), goal valortations as background (distribution band)
3. Double click on ref. line, adds a line of average income
 - scope: default per pane : avg per each *combination*. per table , whole table (same as per pane when there is only one dimension). per cell avg. per each mark.
 - line: only shows income for now. How to add goal? Drag to Marks > detail, not going to display but is available to use. Now line: sum(goal), label (none).
 - formatting: bolder, black
4. Double click on distribution band.
 - scope: per cell
 - computation: value: 60%/80%, percent of sum(goal) instead of income. label: none
 - formatting: check fill below (used normally in bullet graphs because *below* is usually bad)
5. Style:
 - Marks > size, make thinner bars to make background more visible
 - Format > lines: rows: grid lines, solid white / columns: none

Rename sheet to bullet graph, save.

- Drag source to color (will take colors from previous selection)
- Hide headers

Tableau: (not so) basic graphs



Speaker notes

Source + quarter + income > Show me, heatmap. Columns YEAR, click on + (adds column Q)

In Marks, sum(income) is redundant as label and color. If remove label, pure heatmap.

Change palette to Grey warm.

Interpretation:

- both years show low levels for FB on Q3, why?
- both years are *hotter* on Q2, why?

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

Speaker notes

When they work, they provide a powerful means to tame the beast of data overload. Most dashboard live however to a fraction of their potential, not because of poor technology, but because of poor design.

They must grab your attention when it's needed, make it easy to spot what's most important in a screen full of data, and give you the means to understand what's happening (insight) and respond without delay.

Not designed to wow people upon first sight, but to inform people with precisely what they need in the way they need it day in and day out.

3.1 What is a dashboard?

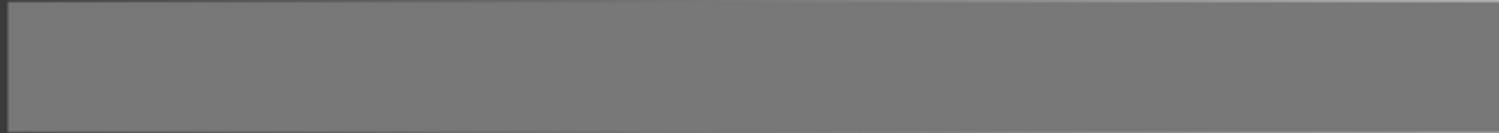
- **Visual display:** *I see = I understand* → **insight to achieve specific objectives:** may require gathering information that is otherwise unrelated or disperse
- **fits in a single computer screen:** it must all be seen at

6 1 2 3 4 5 6 7 8 9 10

Speaker notes

- combine text and graphs, with emphasis on graphics. There is such an intimate connection between what we see and how we think, we actually use the expression “I see” as a substitute for “I understand”. when we make sense of something, we refer to what we’ve learned as “insight”. the more you can rely on images to tell the story, the faster it can be perceived. We perceive images in a parallel manner, which is **much faster than the serial/sequential perception of language**. We have very limited short-term memory: dashboards address this issue in two ways:
 - making the content in each slot more “information rich” (only 4-5 slots; numbers/graphs?)
 - reducing the need to rely on it, everything relevant is available simultaneously. External form of memory.

Must the information be constantly refreshed in real time? Only if the objectives that it serves require real-time information. If you are monitoring air traffic using a dashboard, you must immediately be informed when something is wrong. On the other hand, if you are making strategic decisions about how to boost sales, a snapshot of information as of last night, or perhaps even the end of last month, should work fine.



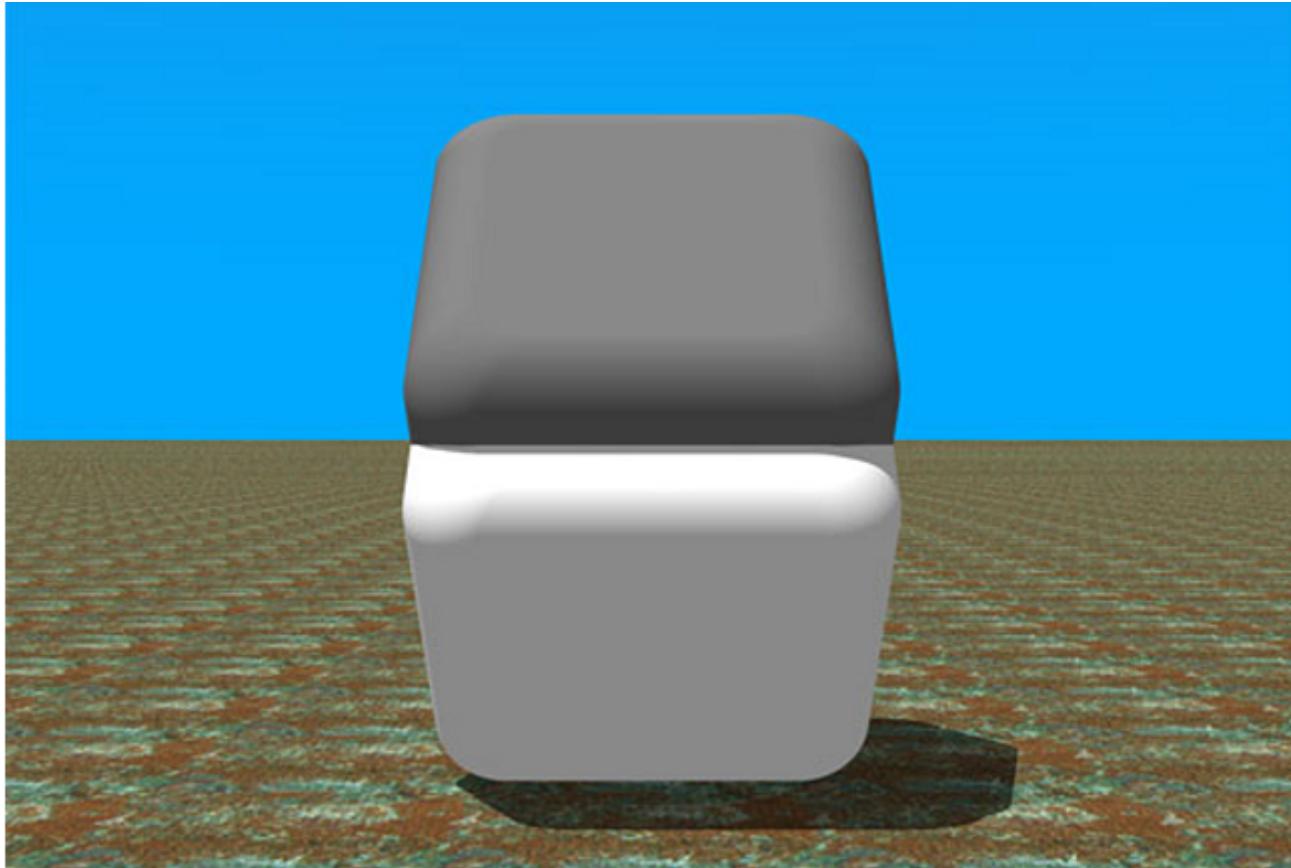
What colour is the bar in the middle?



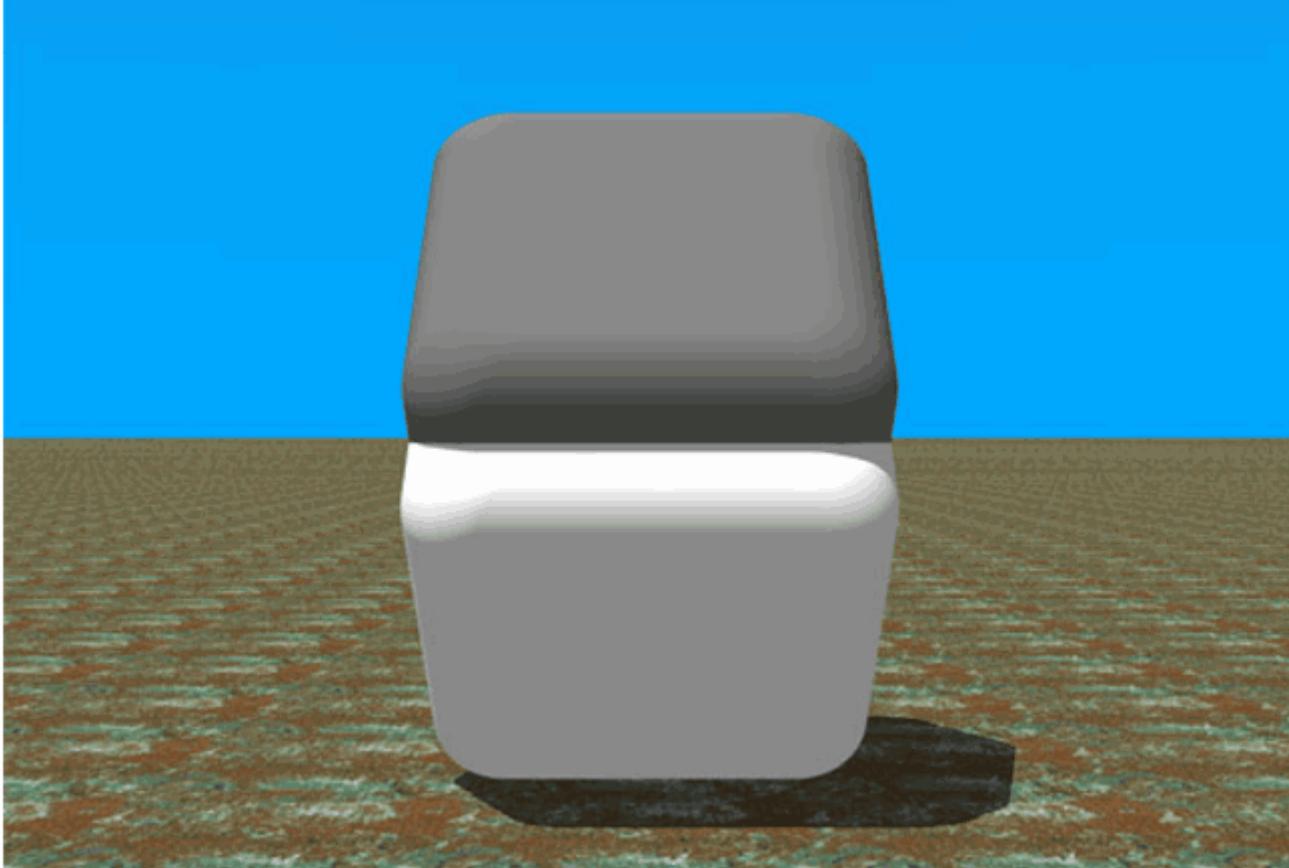
What colour is the bar in the middle?

Speaker notes

In this example, the bar appears to go from light gray on the left to dark gray on the right, but the bar is exactly the same color. The background is gradient, going from dark to light and this affects the way that we perceive the color of the bar. Once I remove the background color you can see that the bar is exactly the same color across.



Are these two boxes the same colour?



Are these two boxes the same colour?

Speaker notes

this is an object that appears gray on the top and white on the bottom. This is exactly the same color. You will prove it to them with the slide animation. As Dr. Lotto explains, we live in a world of a single light source from above. The perceive the top portion of this object to be a well-lite gray surface and the bottom part of the image to be a poor-lite white surface in shadow. There is no shadow in this image.



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

This is a picture of a crater

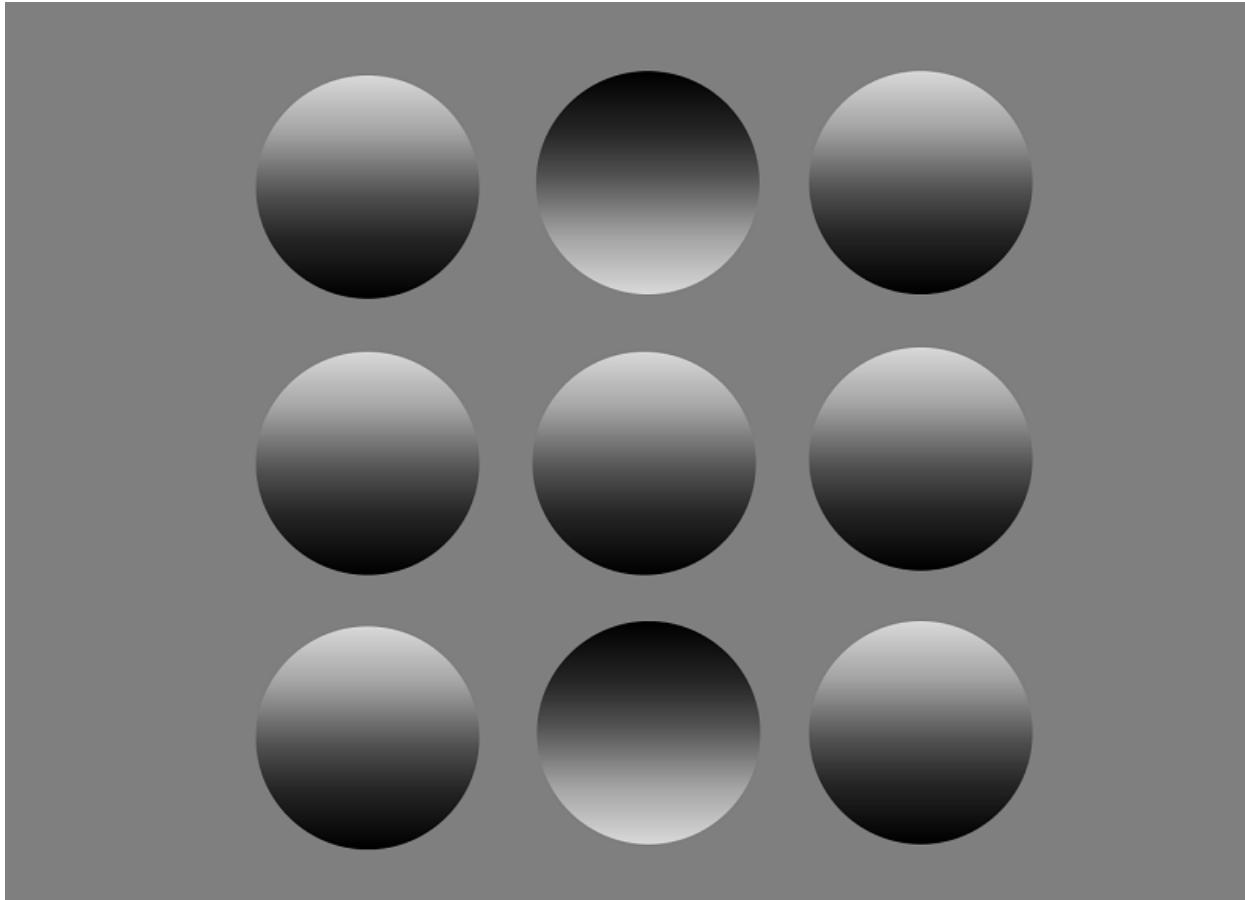


Source: Wikipedia (from the Lunar and Planetary Institute: <http://www.lpi.usra.edu>)

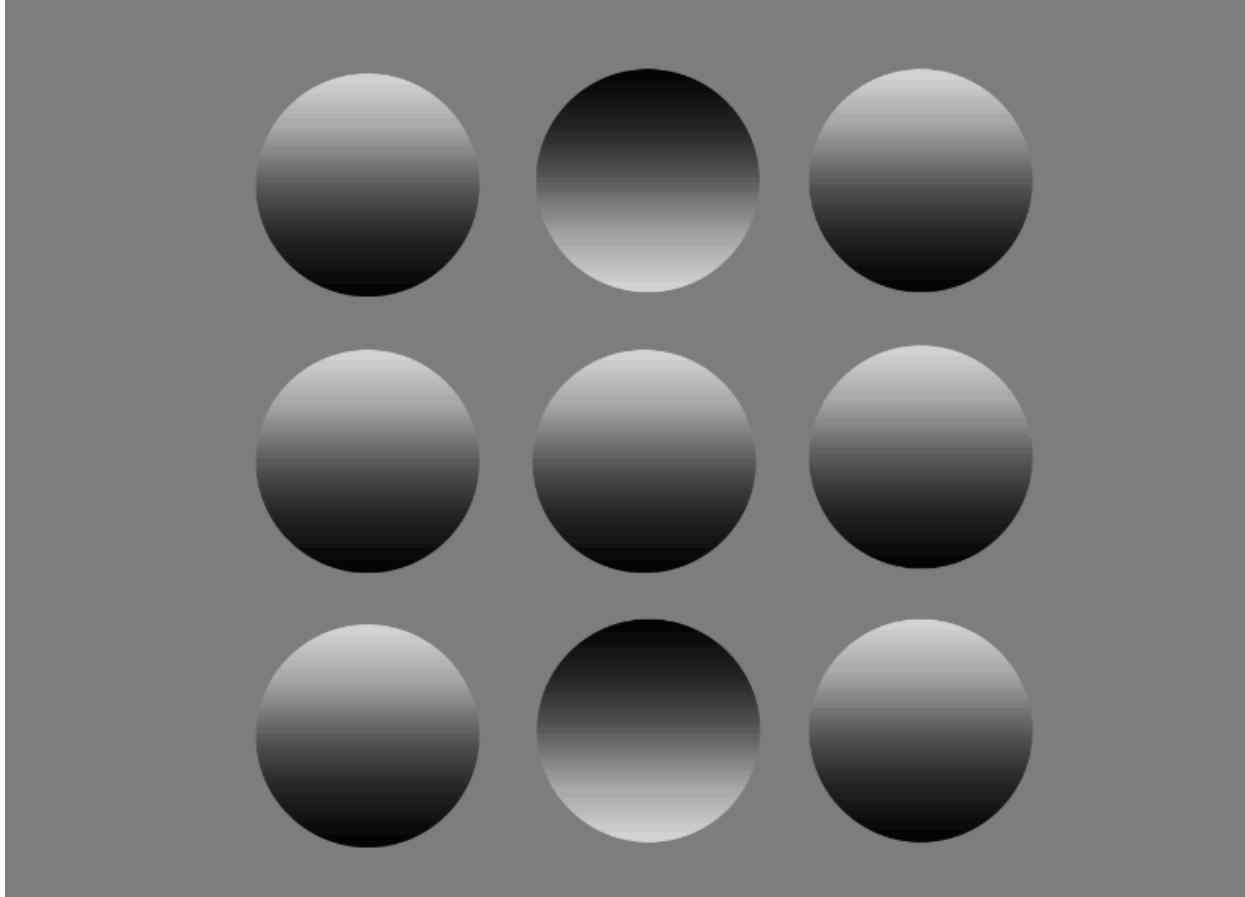
Speaker notes

Watch what happens when we rotate this slide. Now we have to climb the hill.

Our minds expect light coming from above, putting the top of the buttons in light and bottom parts in shadow.



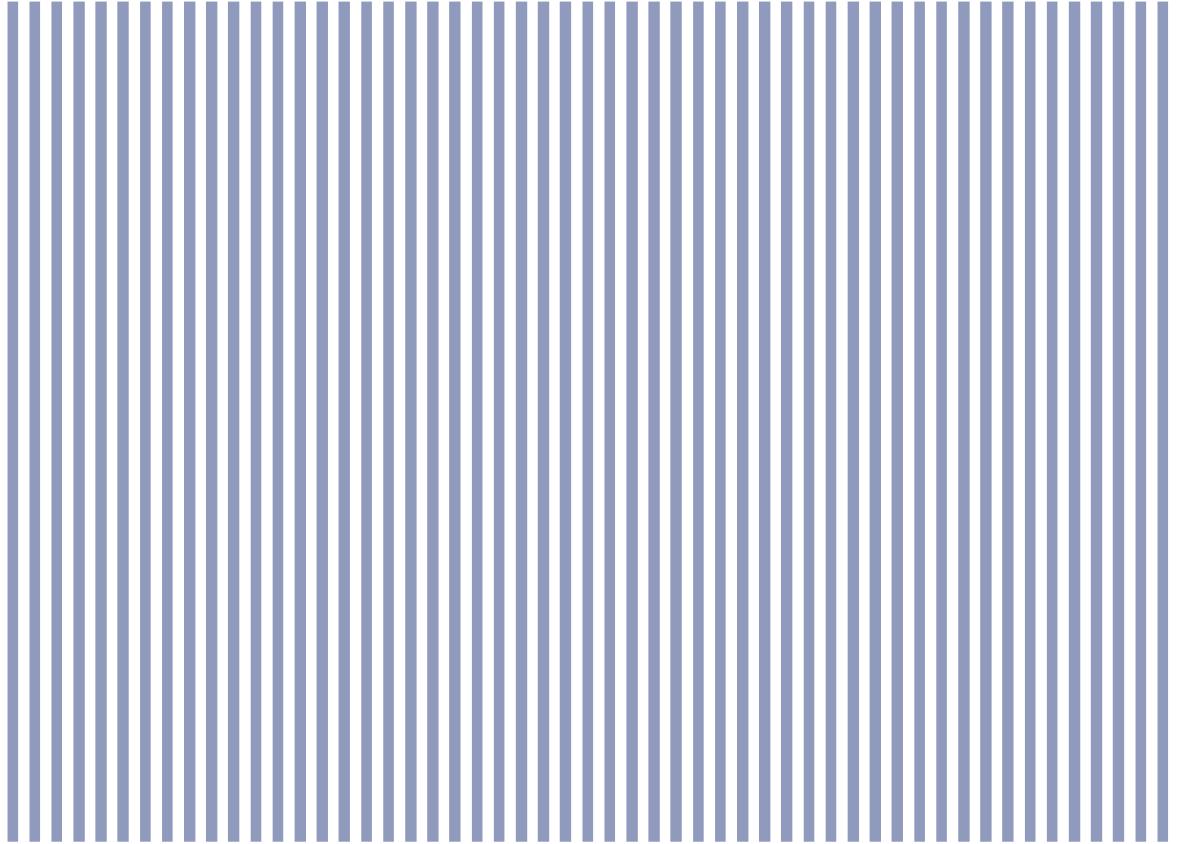
Dots with shadows



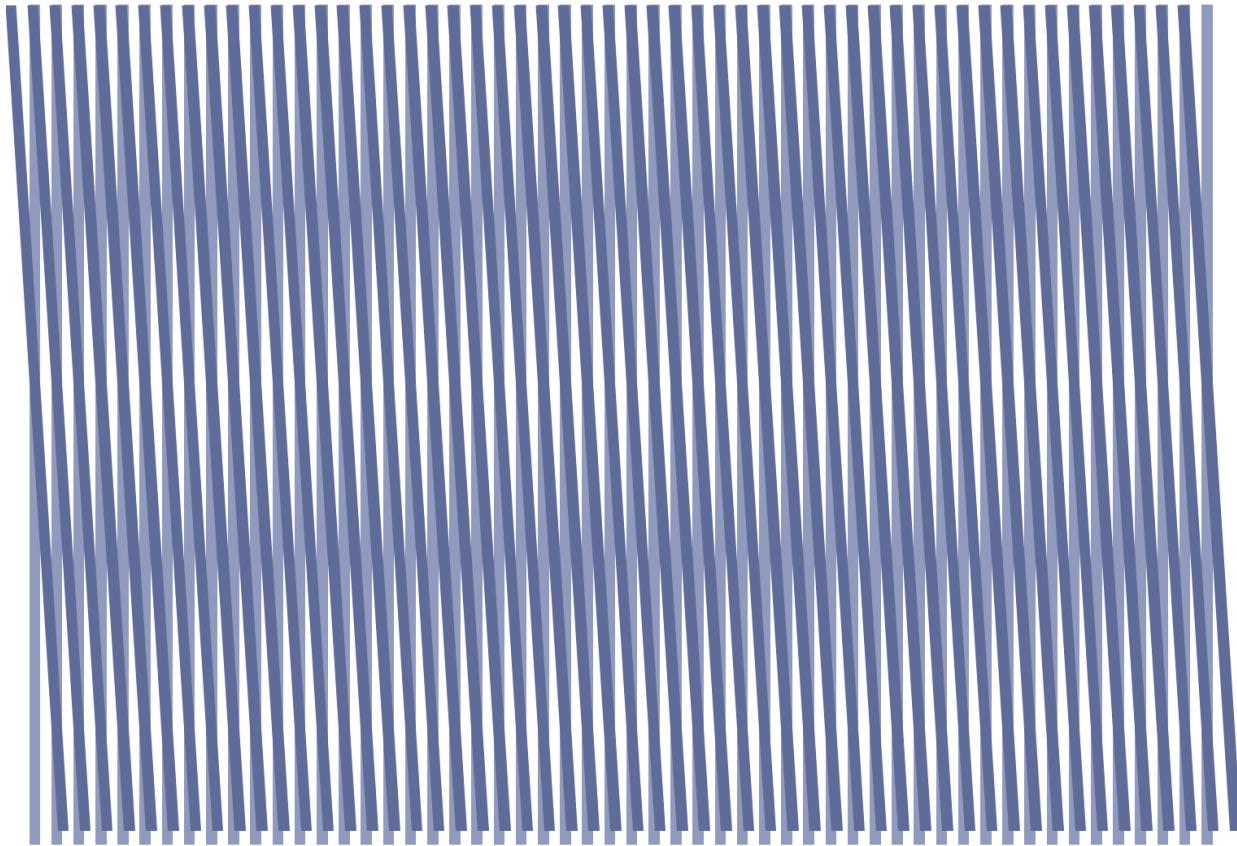
Dots with shadows

Speaker notes

This is another example of shadow. We see some of these buttons (the letter H) as stuck out and the other two pressed in. The ones pressed in are the opposite. Slide animation will rotate this slide to show the opposite effect.



The Moiré effect

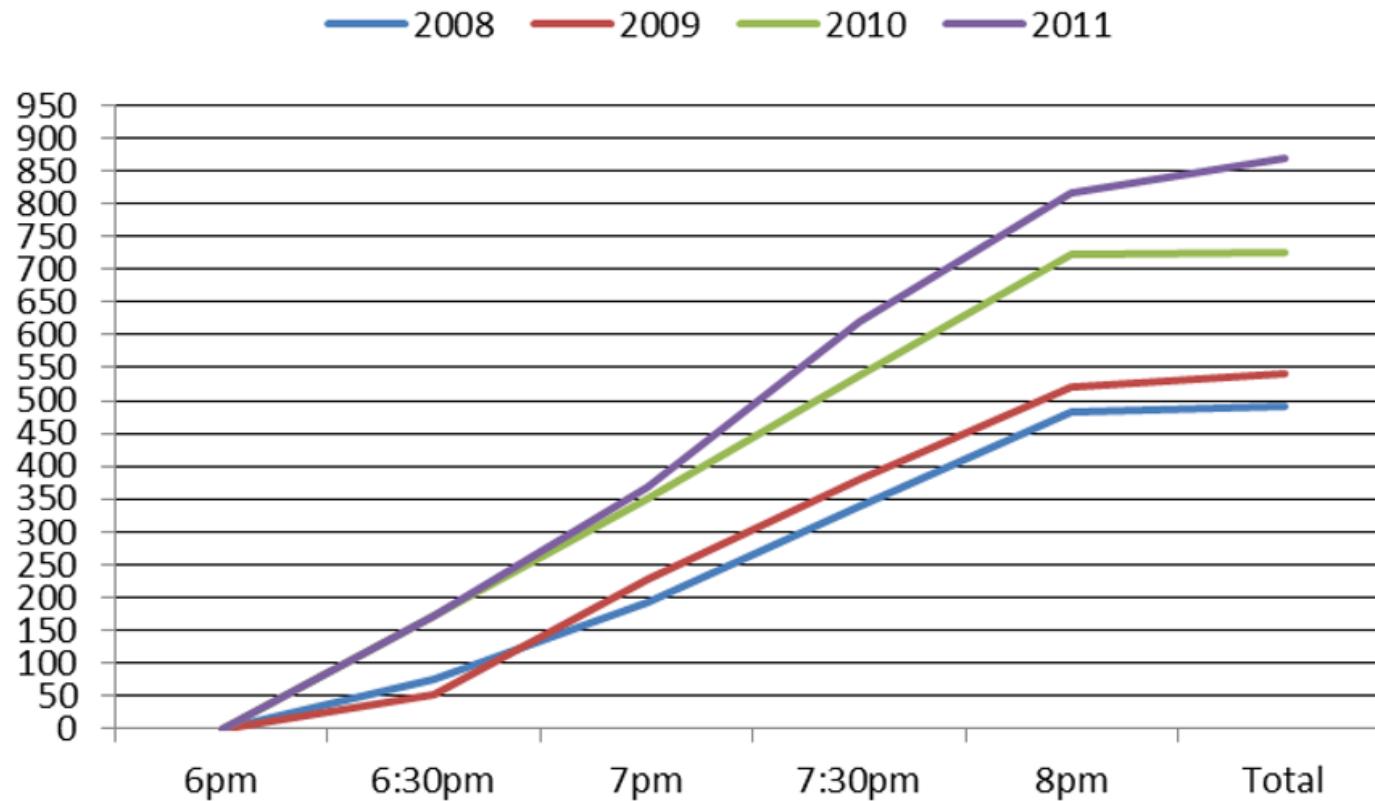


The Moiré effect

Speaker notes

This is called the Moire Effect. Notice the 'shimmer' that occurs when you stare at this image. There is no animation here.

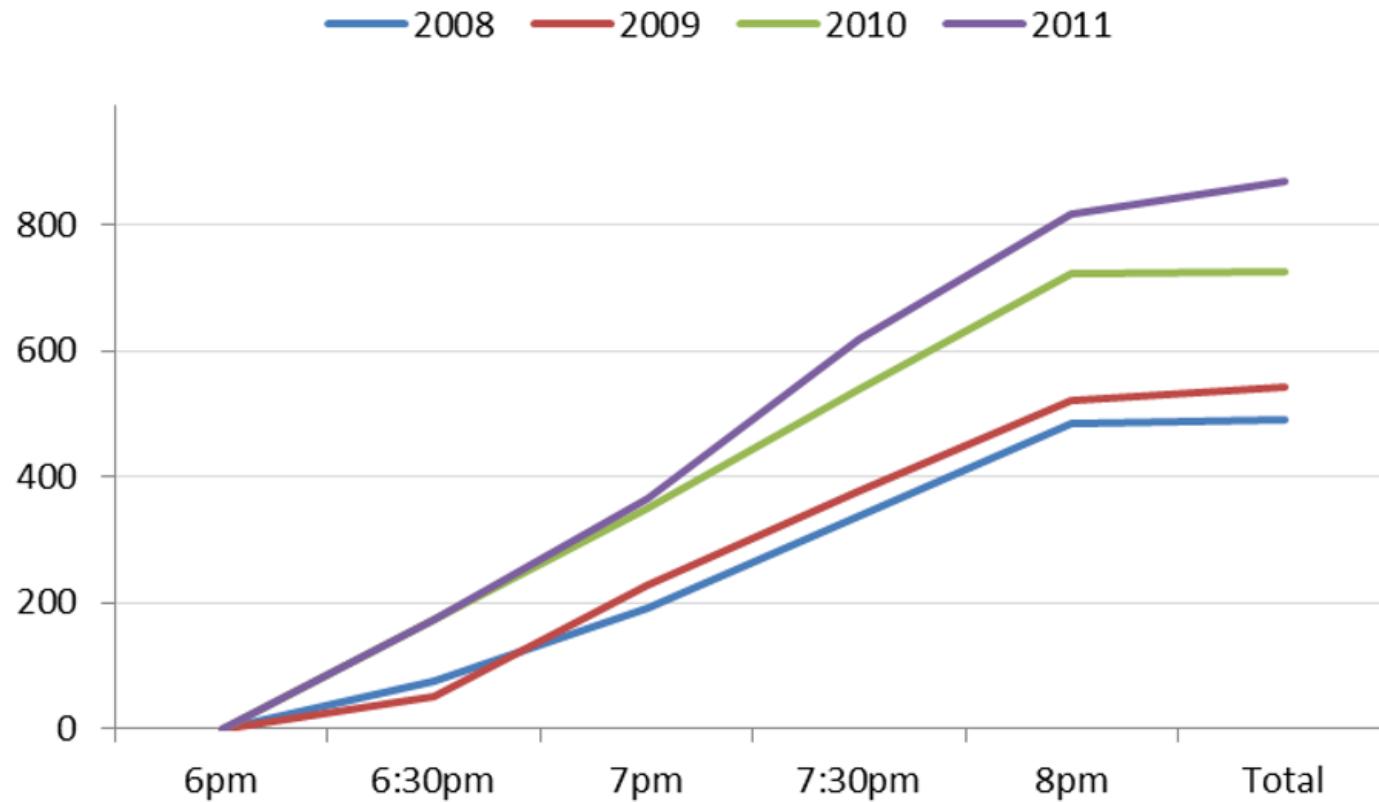
This 'shimmering effect' is even worse when one grid is one top of another.



The Moiré effect in a timeline

Speaker notes

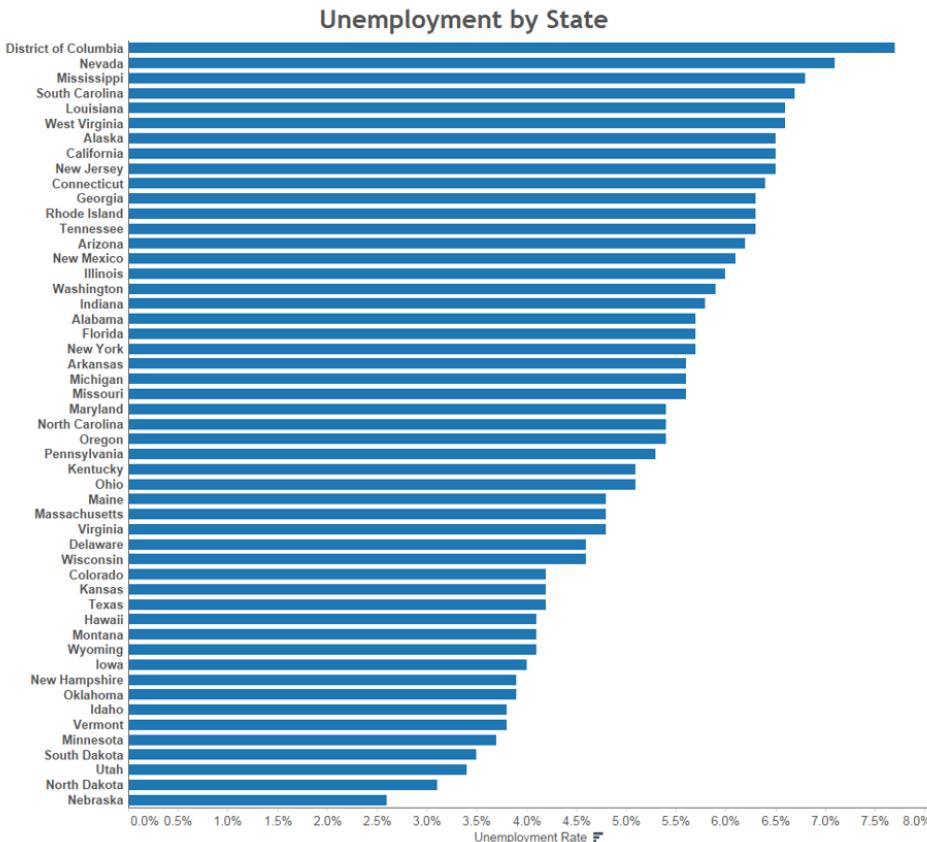
We can see this in data visualization with gridlines and bar charts. In this example, the gridlines are dark and narrowly spaced in increments of 50. Notice that the gridlines get in the way of visualizing the trend in the lines.



The Moiré effect in a timeline

Speaker notes

On the bottom, the gridlines are gray spaced wider in increments of 200. No strange effect



The Moiré effect in a barchart

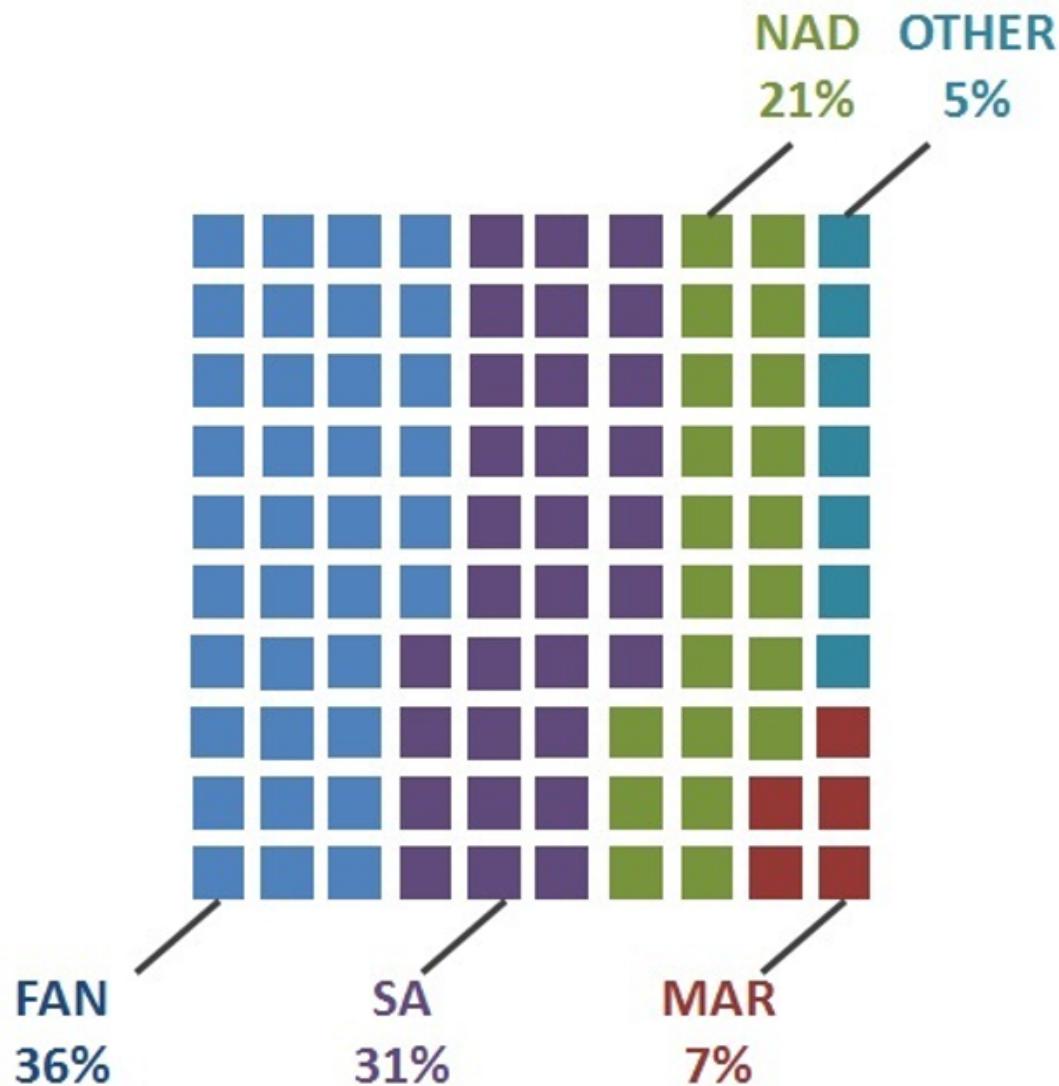
Speaker notes

Here is an example where the Moiré effect occurs when there are so many bar charts.



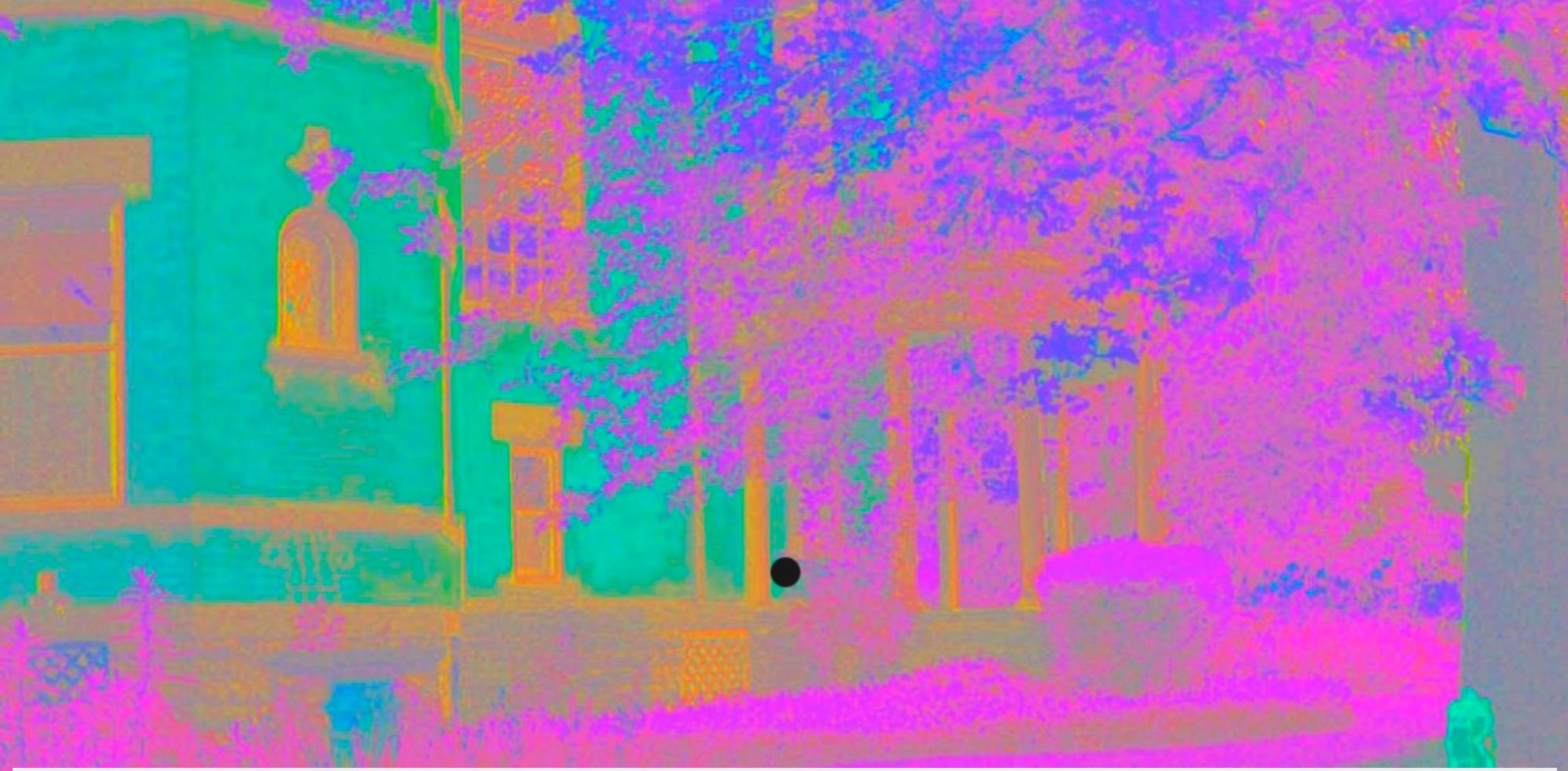
Speaker notes

The Hermann effect, also known as the scintillating grid, shows dark dots in between the squares. This distracting effect is visible as an after effect when our eyes move around the grid.



Speaker notes

This can sometimes be seen in unit charts or waffle charts.



Speaker notes

[Ask the students to stare at the black dot for about 20 seconds. Then change to the next slide. During the 20 seconds you can explain what is going to happen.]

Stare at the black dot for about 15 to 20 seconds. The brain is processing this information, but after a short period of time the image is embedded in our brains. The brain operates on a very small amount of energy and so it has to work efficiently. It basically says, 'I got it. No need to keep processing the same information.'



Speaker notes

This is a black and white image that will appear in color at first, because the brain has the previous slide in memory with the color.



Speaker notes

[Ask the students to stare at the red dot on the woman's nose for about 20 seconds.]

Let's do another one. Stare at the red dot on the woman's nose for about 15 to 20 seconds. This time I'm not going to show you any picture, just a blank screen."

Speaker notes

[The image of the woman will project on the blank screen]

Now you can tell everyone that you saw a ghost in the data visualization class.

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

Speaker notes

- more than one screen (or scrolling) requires to rely on short-term memory, far more processing effort from the mind.
- values compared to what? good or bad? how good or bad? are we on track?
- too much detail slows users down without any benefit, makes her filter out what is important
- spent & income instead of benefit, two values instead of ratios for example
- choosing the right graph (previous section)
-
- channels without attribute/reference variable, specially colour

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

Speaker notes

- non-zero (vertical) axis for example
- top-left is most prominent, bottom-right the less prominent area. use accordingly
- should draw eyes to most relevant information regardless of location. too much highlight means no highlight at all = if everything is prominent, nothing is prominent
- blank space is always better than meaningless decoration. If logos must be used, make them small and visually subtle, and place them out of the way
- REMEMBER colour is a channel that should almost always refer to an attribute/variable
- not excessively decorated does not mean unattractive. no need to add touches to make the dashboard pretty, but rather attractively display the data itself

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

Speaker notes

1. reduce

- decoration graphics, variations in color without meaning, borders or background colours as separation instead of whitespace, gradients instead of solid colour, grid lines (graphs or tables), fill colours for alternating rows, unjustified additional dimensions (specially 3D)... (channels without attribute)
- occasionally useful non-data pixels to de-emphasize: axis lines, borders or backgrounds as separation, grid lines, legends

2. enhance

- show only what is really needed: eliminate unnecessary data, condensate and summarize, level of detail should not exceed what's necessary
- depends on desired type of insight. Some are always important (static means of emphasis), some are only important when certain conditions are met (alerts, require dynamic means of emphasis)

3.4 Example



Speaker notes

- colour is used sparingly. headings separate sections, the only other colour different from gray is on red alerts. colours do not compete with each other for attention
- the prime real estate on the screen has been used for the most important data: upper-left corner of the screen
- small, concise display media have been used to support the display of a dense set of data in a small amount of space. sparklines and bullet graphs convey a lot of information without looking cluttered
- legends are unobtrusive
- white space alone has been used to delineate and group data (no borders, grid lines, background fills)
- not cluttered with instructions and descriptions that are only occasionally needed (those in “Help”)

Dashboards in Tableau

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

Speaker notes

New dashboard. Dashboards in Tableau are container of *sheets* of graphs. Dashboard size is selected depending on destination format (screen or paper, size...). Choose small, 640x420

- Sheets area shows graphs we have created.
- Objects displays additional layout elements that could be used

Drag elements to window: Sparklines top-left, bullet graphs top-right, heatmap bottom

Hide headers

Tableau: Actions

Some degree of interactivity with Actions : highlight and filter

Speaker notes

We can add a minimum level of interactivity

Dashboard > Actions, Highlight.

- source: sparkline
- target: all

Dashboard > Actions, Filter.

- source: sparkline, run on select
- target: all

Tableau: What else?

Calculated fields

...

Speaker notes

Add a ROI column:

Right click > Create Calculated Field. Name ROI, $[\text{Income}]/[\text{Spent}]$

What can we do with this?

- Replace the heatmap (different insight)
- Trend area chart:
 - Select Quarter, Source, ROI, click Area chart from “Show me”. Click on Quarter ‘+’ sign.

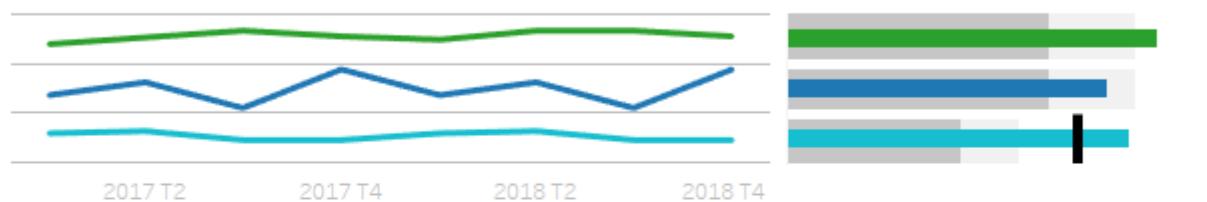
What would you like to see in this dashboard?

spent/income ratio, spent/visits ratio, income/visits ratio...

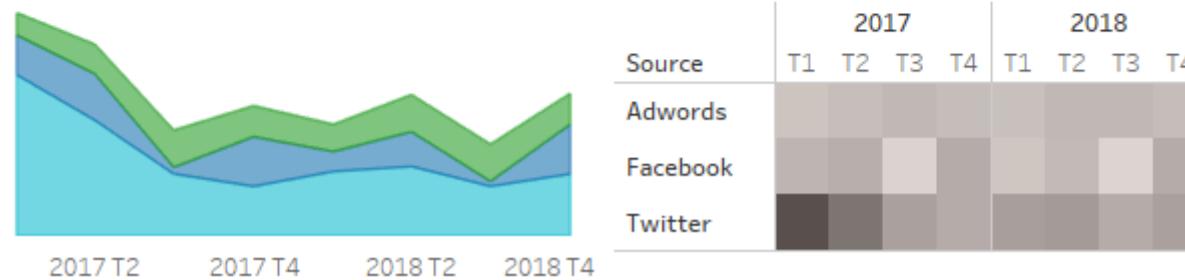
Tableau: Final styling

Marketing campaigns results Last 2 years

Income trend



Returns on investment



Speaker notes

Title: Marketing campaigns result Last 2 years White space Tooltips! “ brought during ”

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/MUMA2018>

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/>