

FIT3179 Data Visualisation

Week 04: Designing Effective Visualisations



Presentations

Lecture Overview

- Data-ink ratio
- Chartjunk
- Gestalt principles
- Storytelling with visualisation
- Macro and micro reading
- How to lie with data visualisation

Data-ink ratio

- *Data-ink ratio* by Edward Tufte

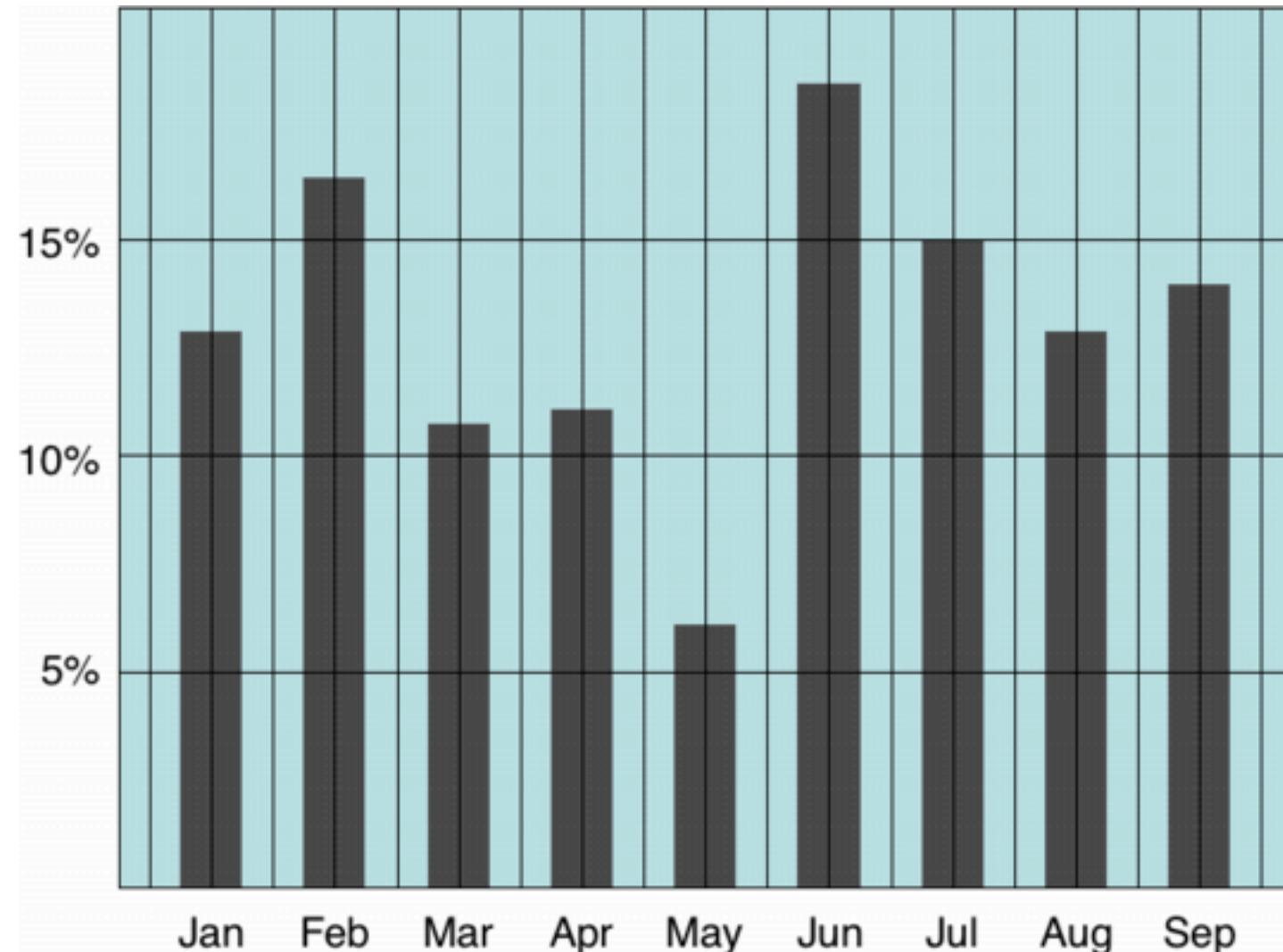
- A measurement of the amount of ink used to represent data in a ratio.

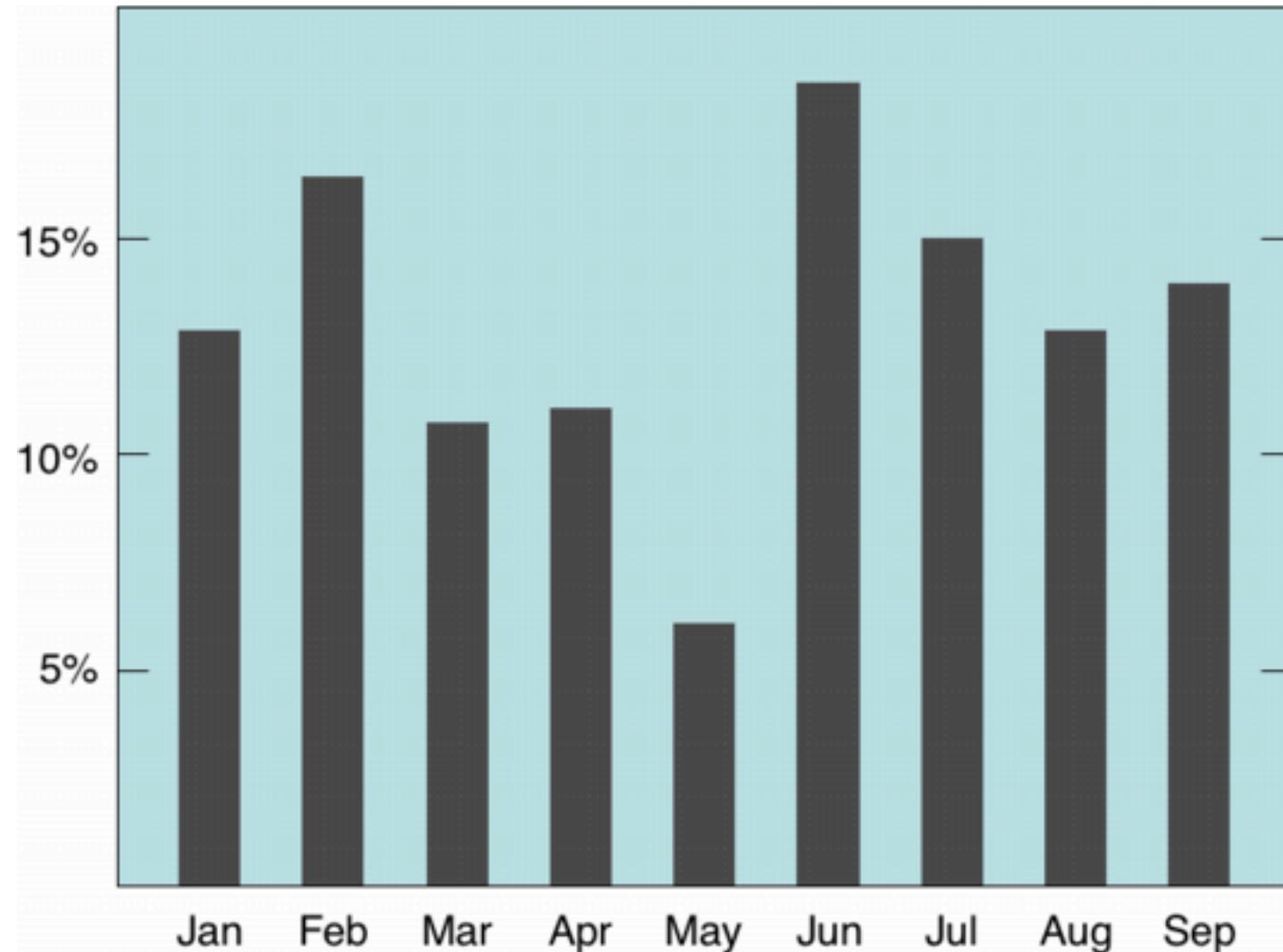
$$\text{data-ink ratio} = \frac{\text{ink for elements that encode data}}{\text{total ink for all elements}}$$

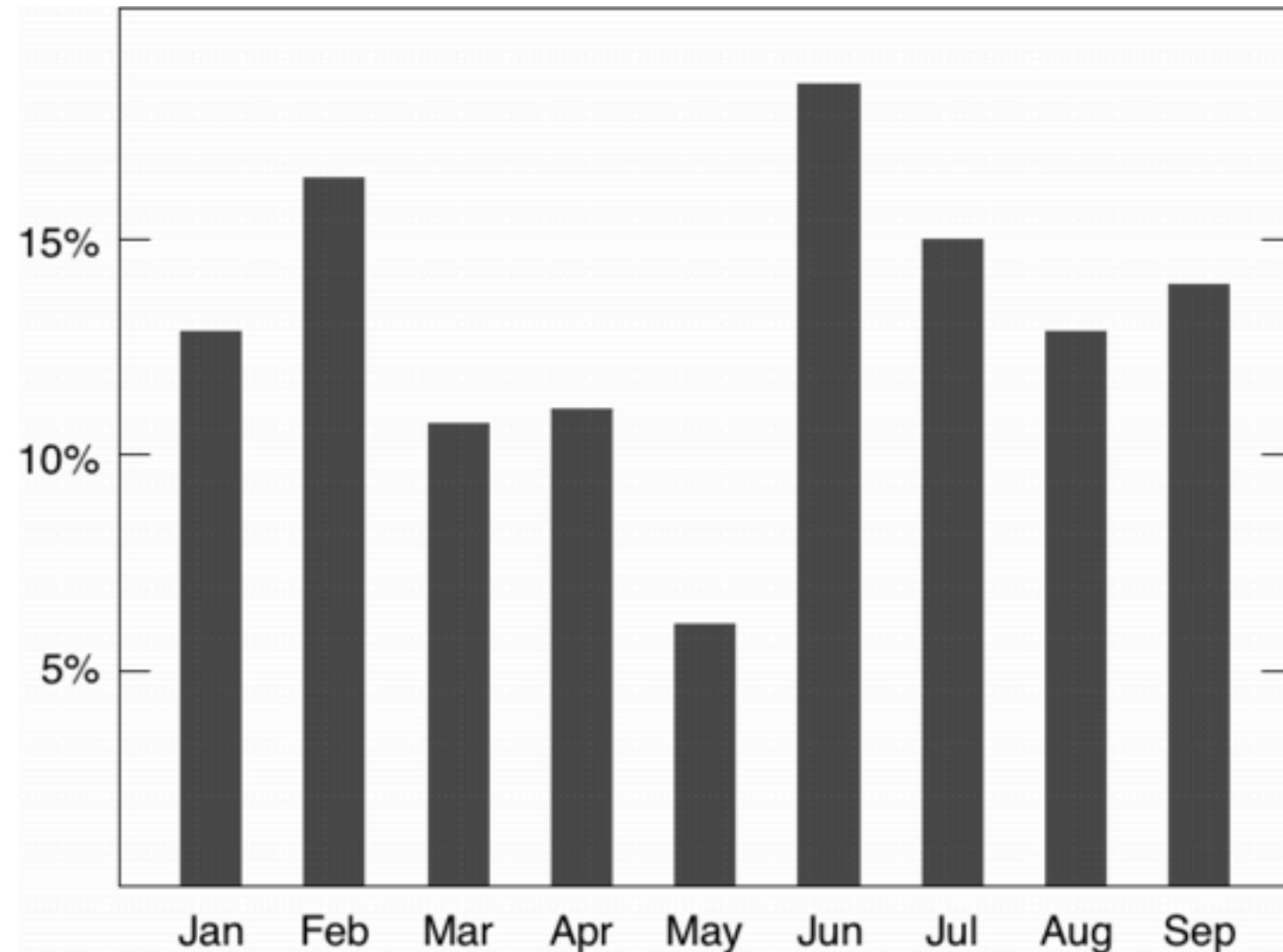
- Tufte argues that

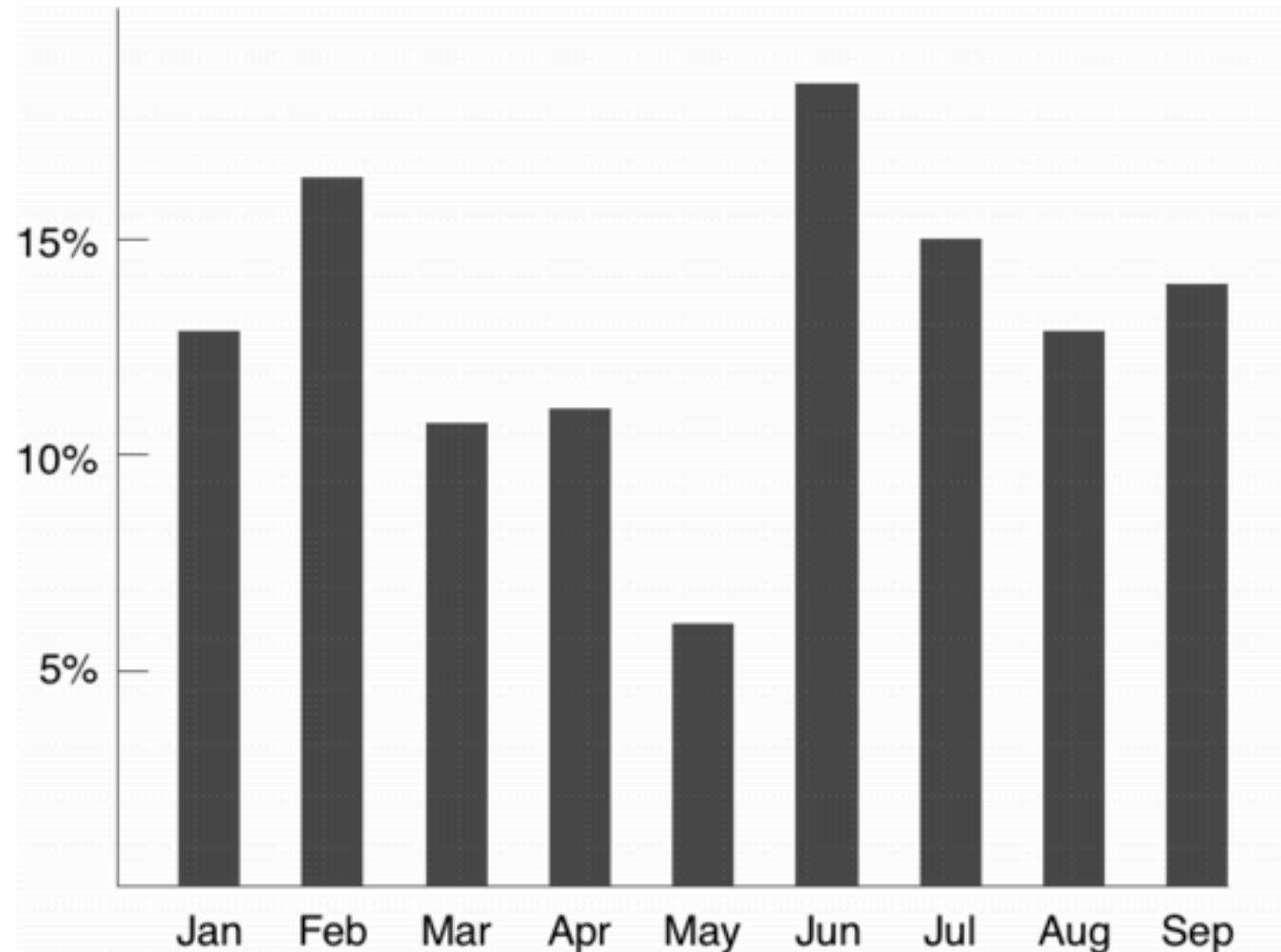
- data-ink ratio should be close to 1
 - ornamental elements should be removed (= chartchunk)
 - strip away ink not dedicated to data

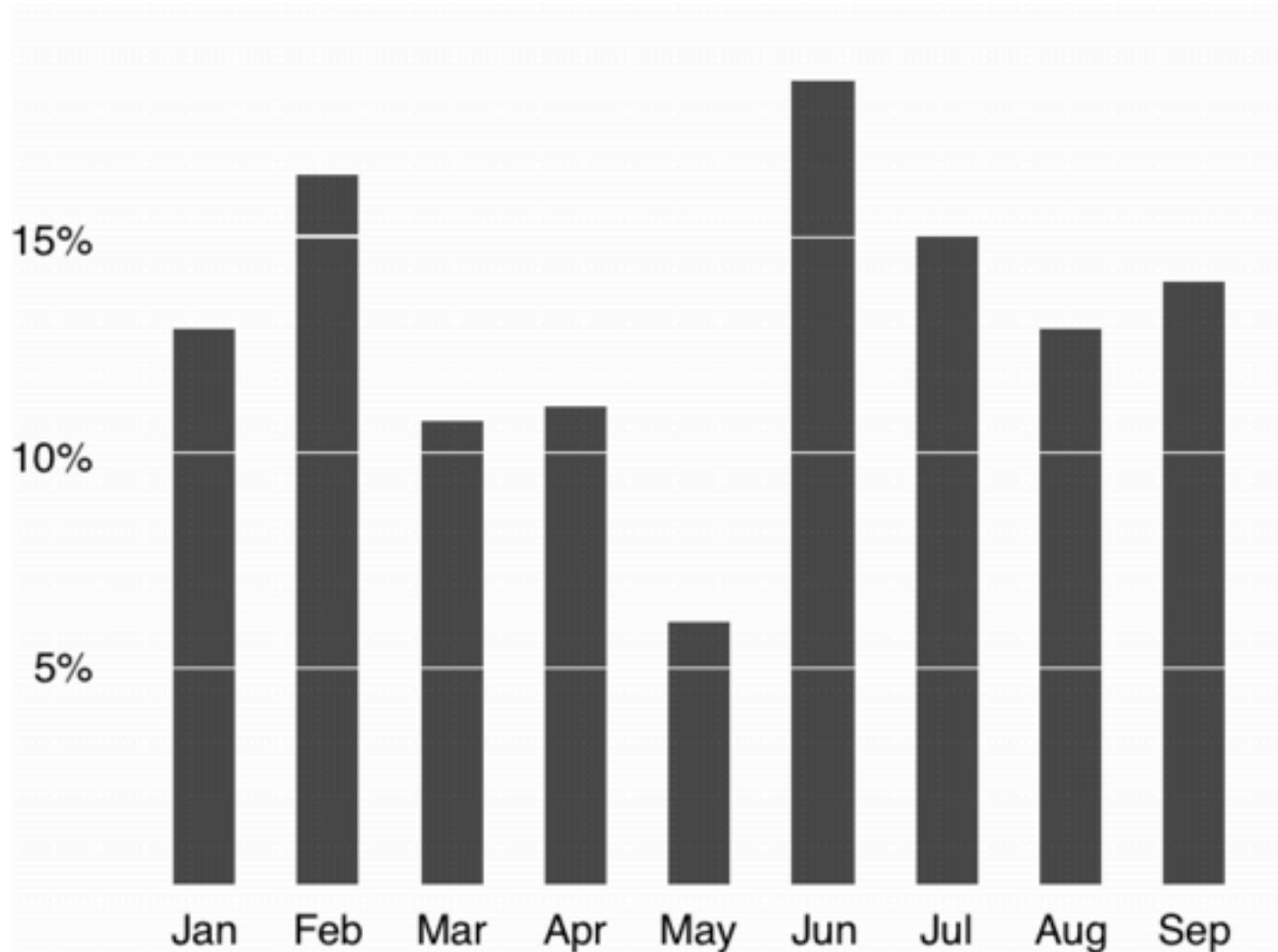
For more information, see Edward Tufte, [The visual display of quantitative information](#).









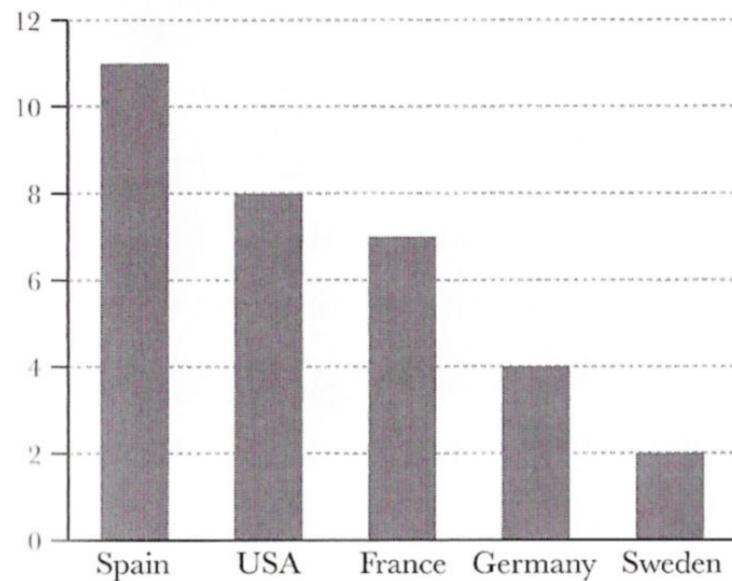


Lower data-ink ratio

more ink used for elements
that don't encode data

Unemployment rate

(As a percentage of total workers)

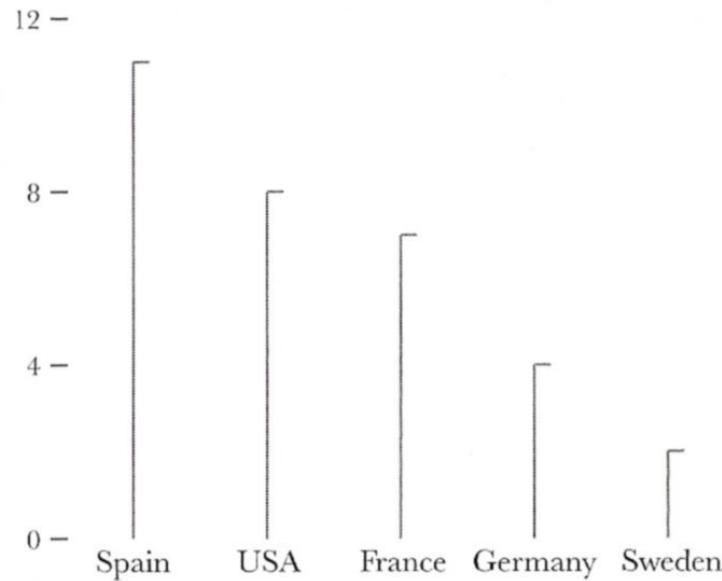


Higher data-ink ratio

less ink used for elements
that don't encode data

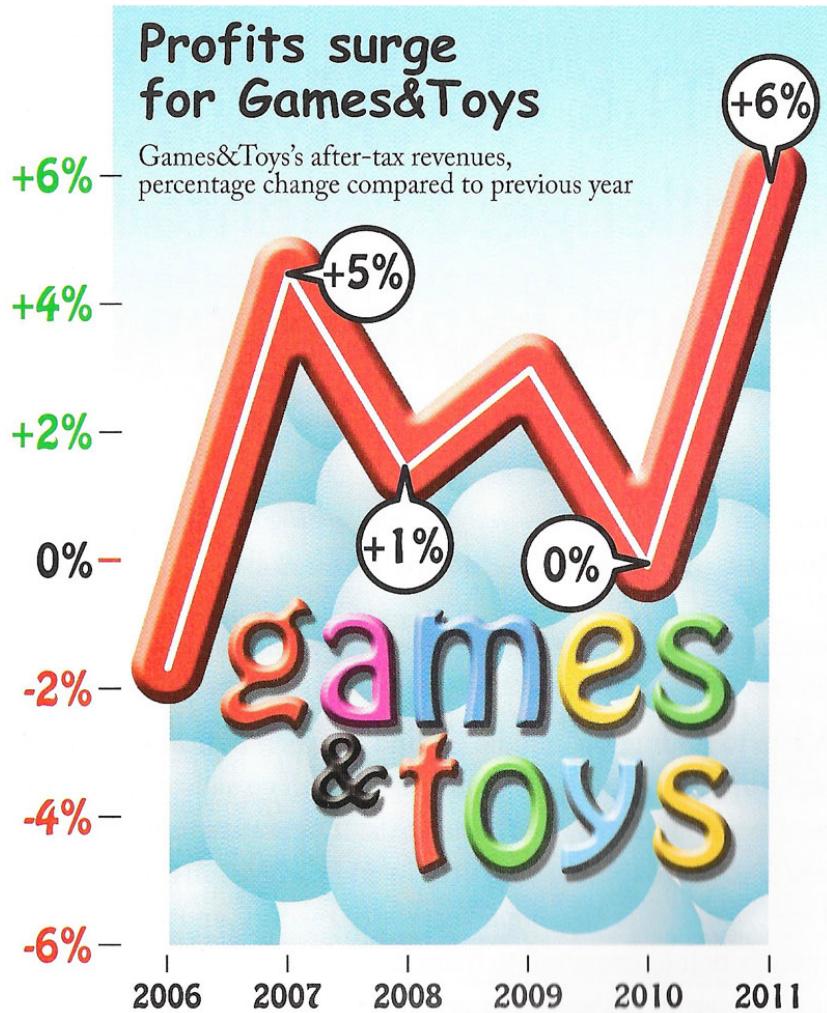
Unemployment rate

(As a percentage of total workers)

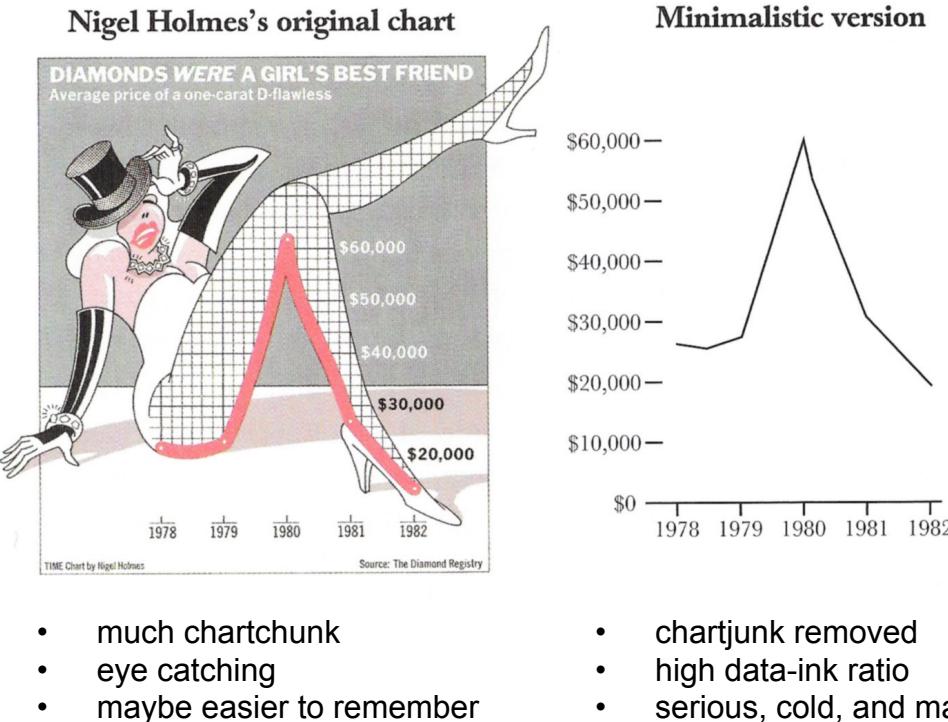


Chartchunk

- Edward Tufte: Chartjunk is unnecessary and redundant.
- It is the ornamental ink that does not add to the understanding of data.
- It is commonly seen when the visualisation doesn't actually have much data.
 - So the creator tries to fill the space with graphics!



Chartjunk vs. data-ink ratio



- Need to find a balance between decorative graphical design and increasing data-ink ratio.
- Well-designed ornamental elements (i.e. chartjunk) can increase memorability, and create good feeling about an artifact, which may make us better at using it to accomplish a goal.
- We are not graphic artists, so better focus on high data-ink ratio.

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks

Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT

Guidelines for designing information charts often state that the presentation should reduce ‘chart junk’ – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people’s accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords

Charts, information visualization, imagery, memorability.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI):
Miscellaneous.

General Terms

Design, Human Factors

INTRODUCTION

Many experts in the area of chart design, such as Edward Tufte, criticize the inclusion of visual embellishment in charts and graphs; their guidelines for good chart design often suggest that the addition of *chart junk*, decorations and other kinds of non-essential imagery, to a chart can make interpretation more difficult and can distract readers from the data [22]. This *minimalist* perspective advocates plain and simple charts that maximize the proportion of

data-ink – or the ink in the chart used to represent data.

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their charts, from small decorations to large images and visual backgrounds. One well-known proponent of visual embellishment in charts is the graphic artist Nigel Holmes, whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).

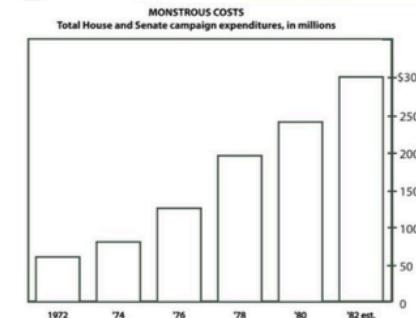
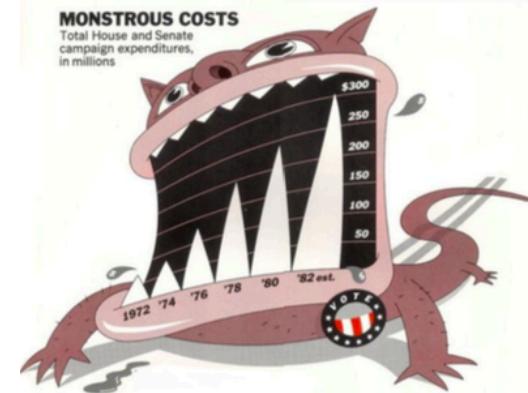
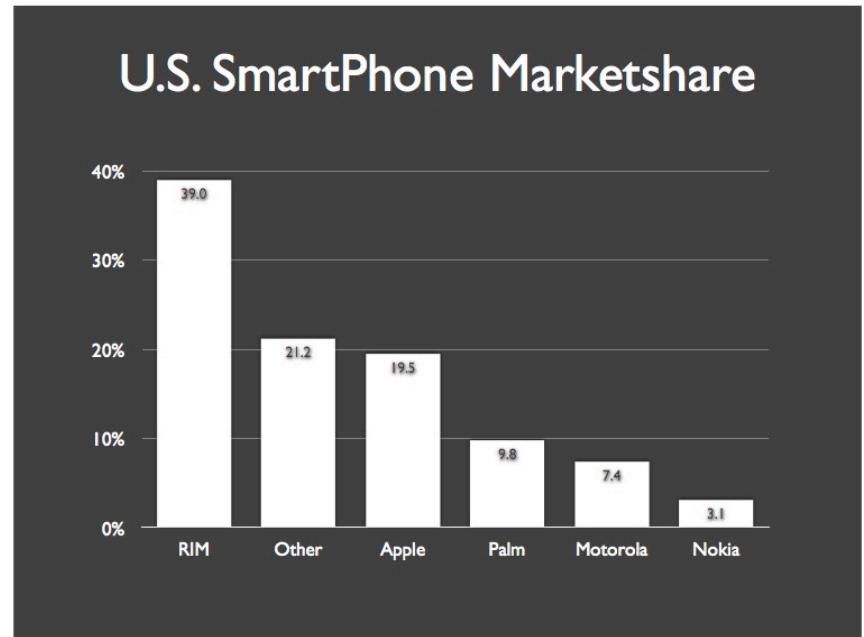
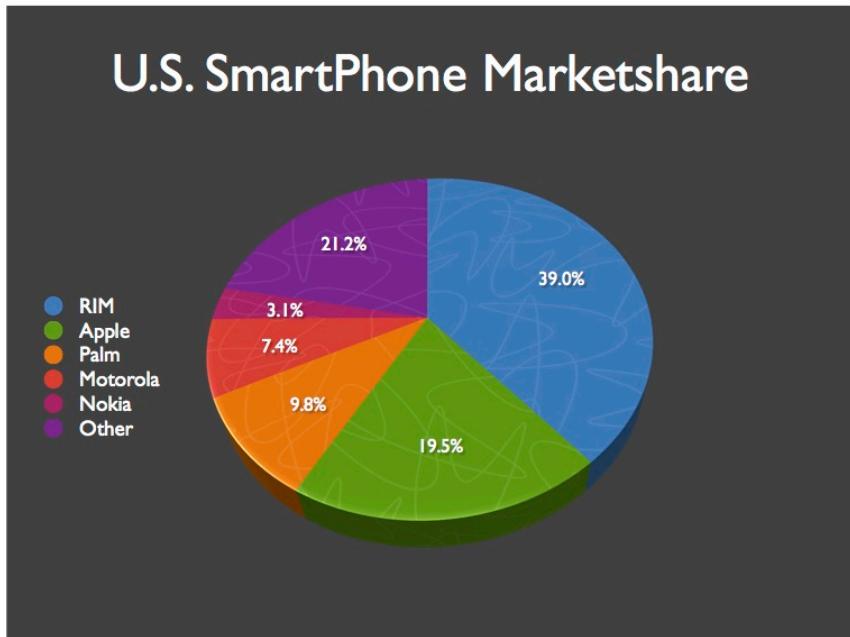


Figure 1. A chart by Holmes [7] (above), and a ‘plain’ version.

3D chartjunk



①

An Unchanging Slice of the Pie

Total R&D as Percent of Discretionary Spending



SOURCE: OMB

②

Who Spends Federal Dollars

Federal R&D by Performer, FY 2009



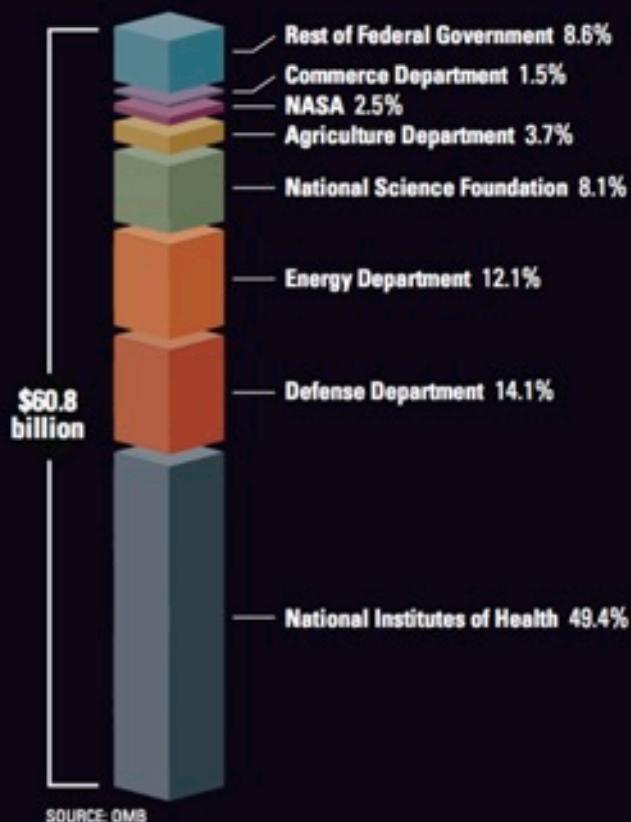
\$116.6 billion

Industry \$46.6 billion	Intramural \$26.6 billion	Universities/Colleges \$26.0 billion
National Labs \$10.4 billion	Other \$6.9 billion	

④

The Biggest Research Agencies

Spending on basic and applied research in FY 2010



28

SOURCE: OMB

Who Funds Academic Research

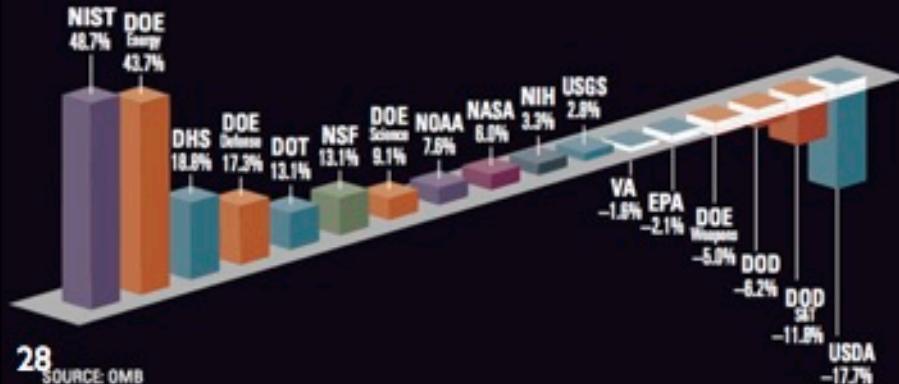
R&D Expenditures at Colleges and Universities, FY 2009

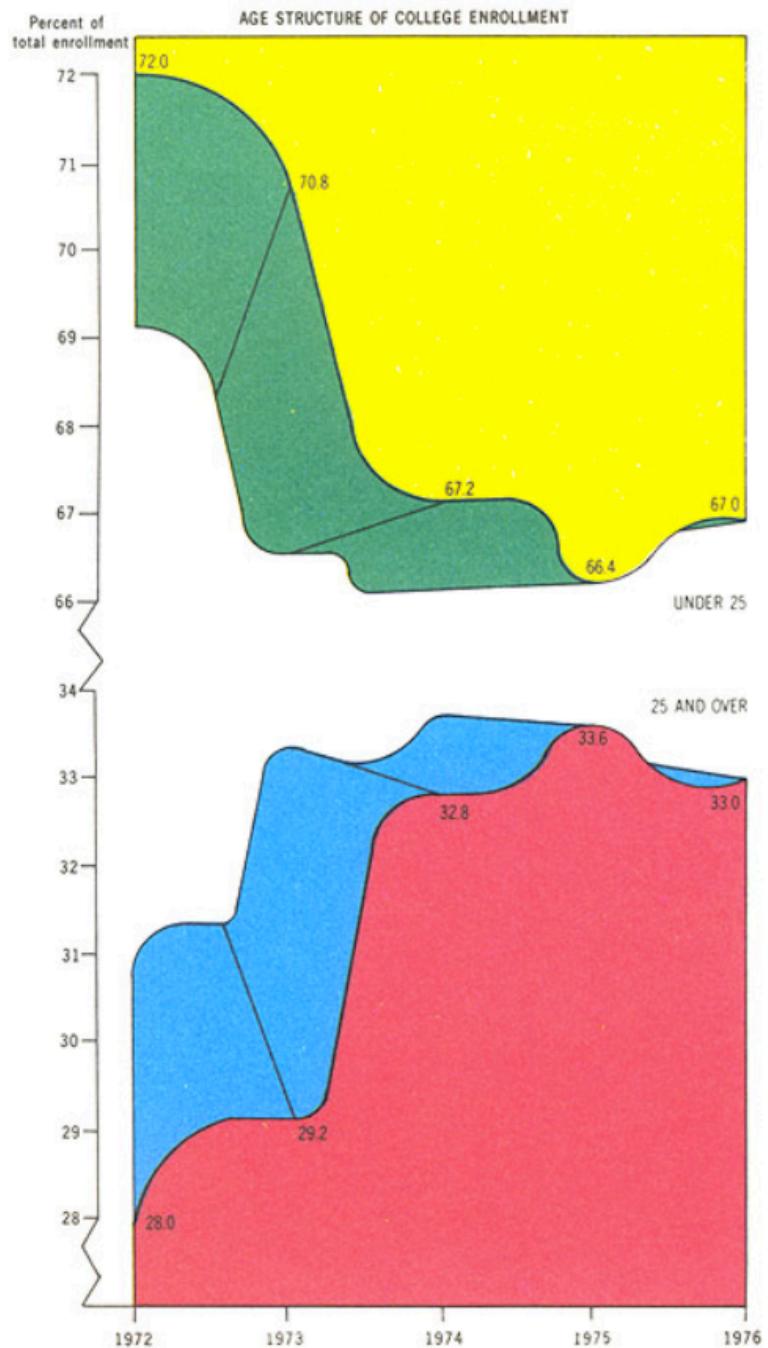


Federal Government 59.3%	Institutional Funds 20.4%	All Other Sources 7.8%
State and Local Government 6.6%	Industry 5.8%	

SOURCE: NSF

③ Winners/Losers in the President's FY 2012 Request



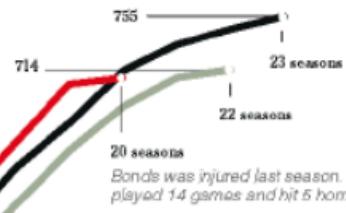
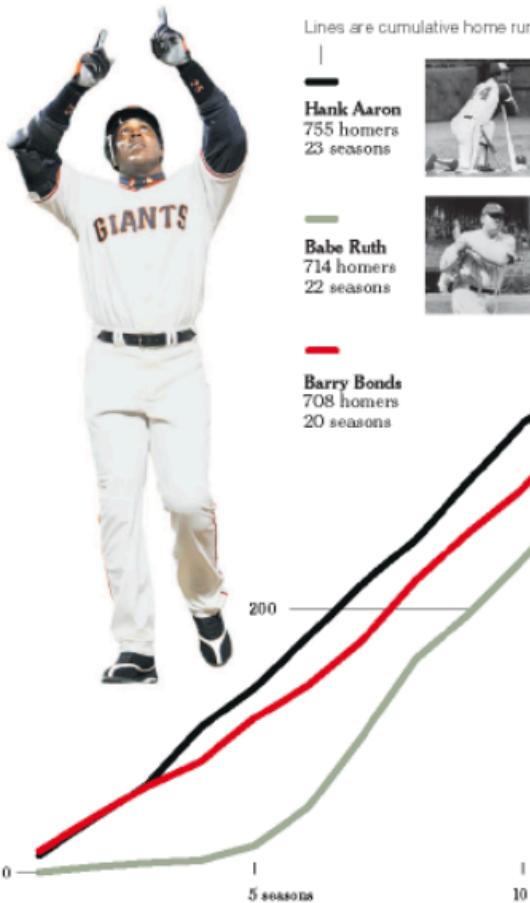


755

Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.

Lines are cumulative home runs.



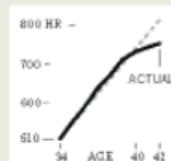
Homer Pace After Age 34

If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

PROJECTION BASED ON AVERAGE OF PREVIOUS FIVE SEASONS

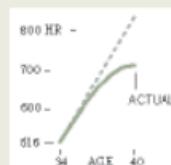
Aaron

Actual homers slightly outpace projected homers for five seasons.



Ruth

Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.



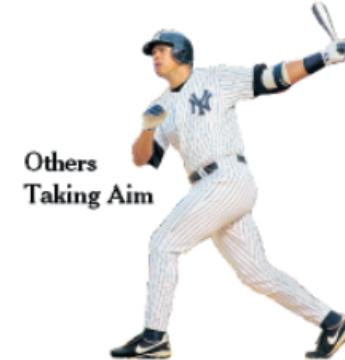
Bonds

From age 35 to 39, he averaged 14 more homers a season than projected.



Note: Ages as of July 1 of each season.

Others Taking Aim



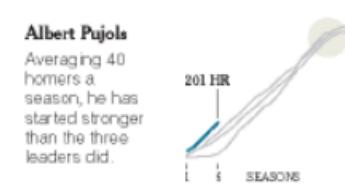
Alex Rodriguez

Is ahead of the pace set by all three home run leaders.



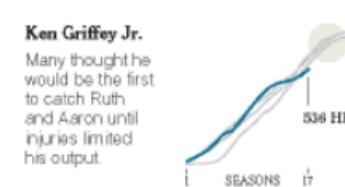
Albert Pujols

Averaging 40 homers a season, he has started stronger than the three leaders did.



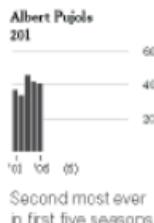
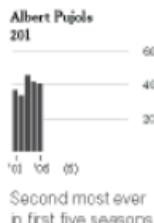
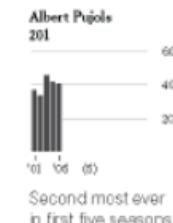
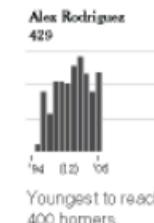
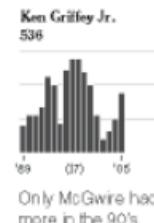
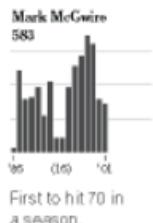
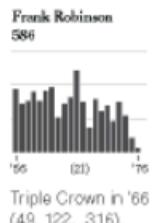
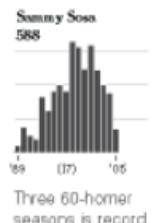
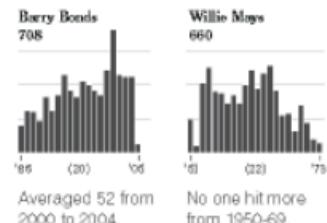
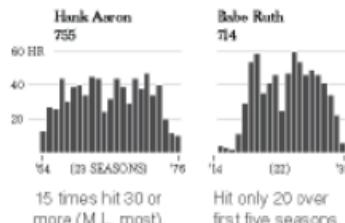
Ken Griffey Jr.

Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.



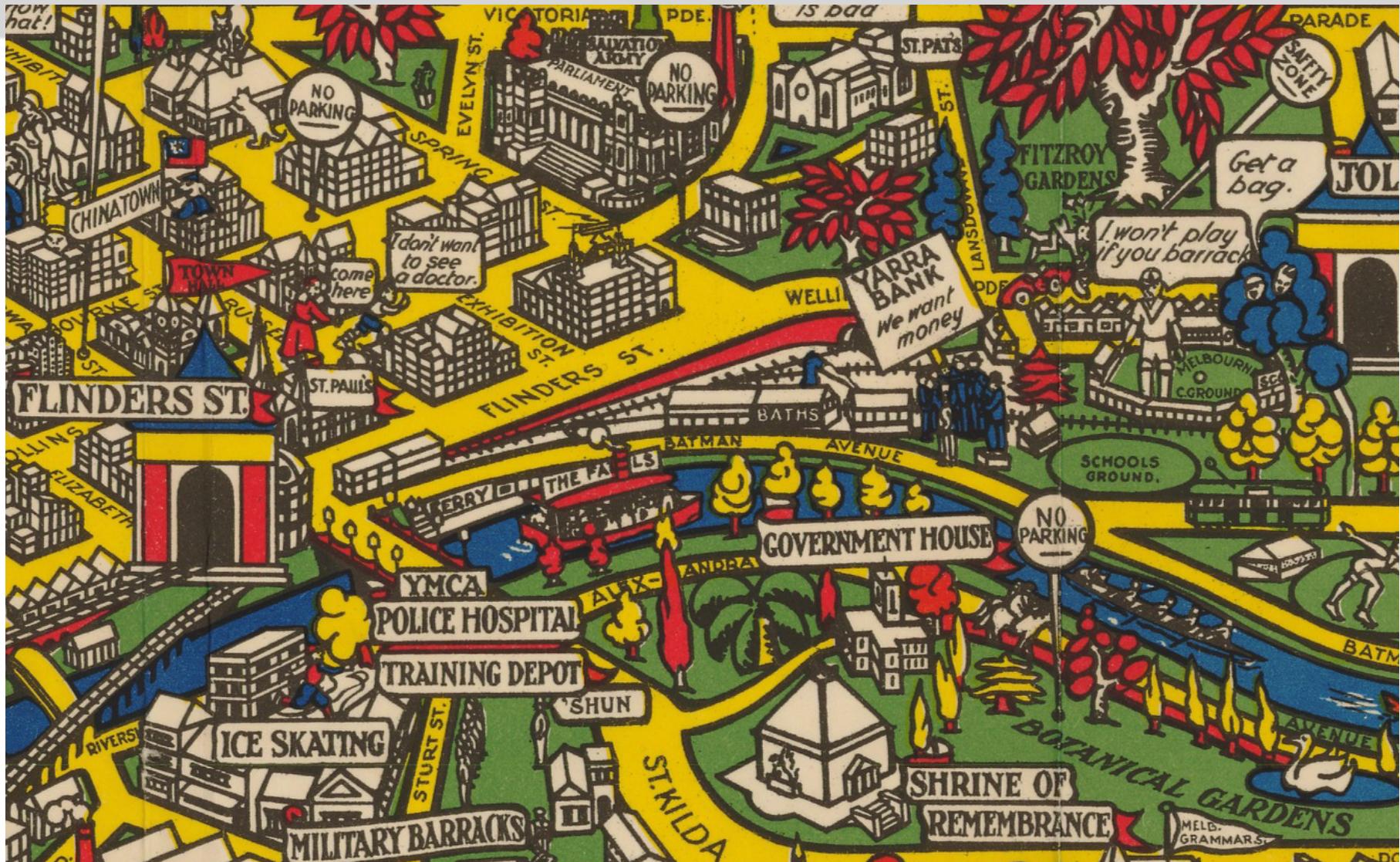
Differing Paths to the Top of the Charts

The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).





Wonder map of Melbourne by John Power Studios, 1934



Wonder map of Melbourne by John Power Studios, 1934

Storytelling: guiding the reader

- Storytelling is about guiding the viewer through the visualisation
 - We can have the viewer see our visualisation in a specific way.
 - We can control in what order the data is seen.
- The eye moves from left to right and from top to bottom
 - In western society we read left to right and up to down
 - Because of this we tend to go clockwise with our eyes first
- Use gestalt principles
 - Guiding lines (we will follow them instinctively)
 - Similar images, patterns or colours (we instinctively relate them together)

Proximity



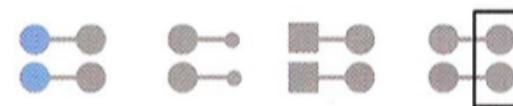
Enclosure and Closure



Similarity



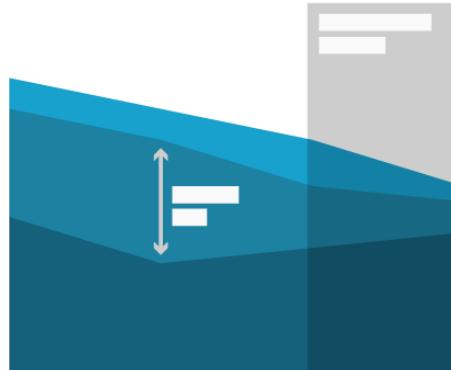
Connection



Mandatory reading: pages 71–80 of *Storytelling with data*
by C. Nussbaumer Knaflc

Gestalt principles for an annotated chart

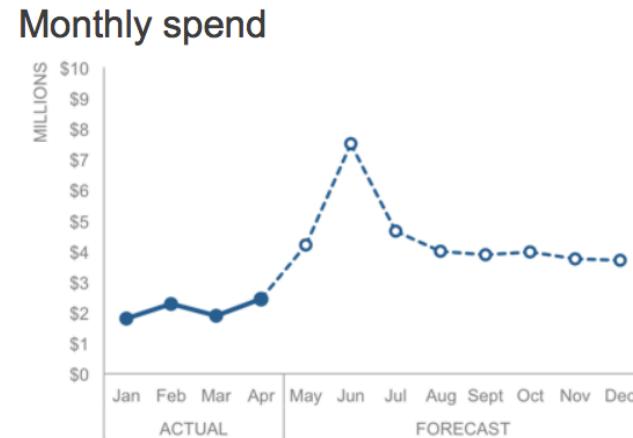
		Monthly spend (M)
ACTUAL	Jan	\$1.80
	Feb	\$2.28
	Mar	\$1.89
	Apr	\$2.44
FORECAST	May	\$4.20
	Jun	\$7.50
	Jul	\$4.65
	Aug	\$3.99
	Sept	\$3.88
	Oct	\$3.97
	Nov	\$3.75
	Dec	\$3.70



What are the gestalt principles for creating an annotated chart?

Gestalt principles for an annotated chart

Monthly spend (M)	
ACTUAL	Jan \$1.80
	Feb \$2.28
	Mar \$1.89
	Apr \$2.44
FORECAST	May \$4.20
	Jun \$7.50
	Jul \$4.65
	Aug \$3.99
	Sept \$3.88
	Oct \$3.97
	Nov \$3.75
	Dec \$3.70



Similarity: vary line patterns and point symbols to separate actual and forecast groups.

Connectedness: lines between dots

Closure: boxes in legend for grouping actual and forecast groups.

Gestalt principles for an annotated chart

Monthly spend

MILLIONS



Monthly spend

MILLIONS



Label data points if values are important.

Forecast values illustrate expectations and are unsure, so you may prefer to not show them.

Gestalt principles for an annotated chart

Monthly spend

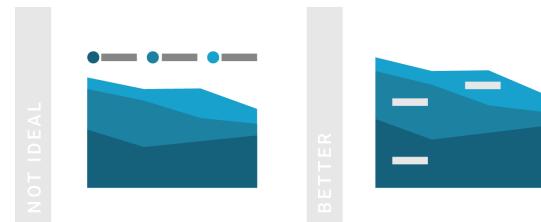


Similarity: vary colour to direct attention to most important data point

Monthly spend

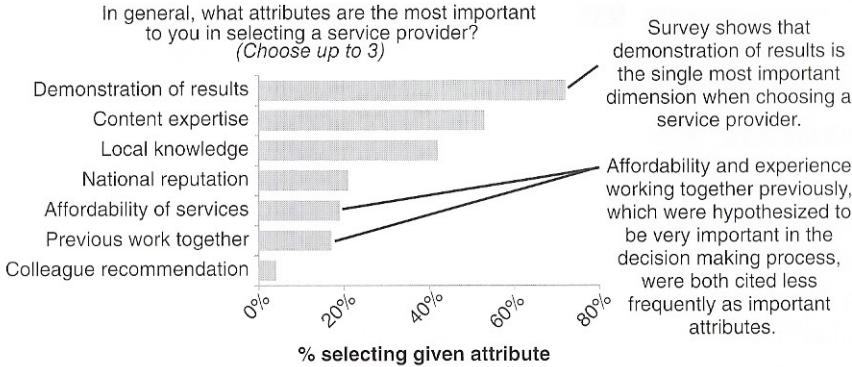


Explanatory text on the diagram.
Annotations on maps and diagrams are very powerful.



Gestalt principles for visual order (annotated chart)

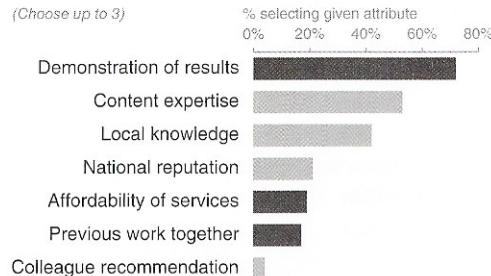
Demonstrating effectiveness is most important consideration when selecting a provider



Data source: xyz; includes N number of survey respondents. Note that respondents were able to choose up to 3 options.

Demonstrating effectiveness is most important consideration when selecting a provider

In general, **what attributes are the most important** to you in selecting a service provider?



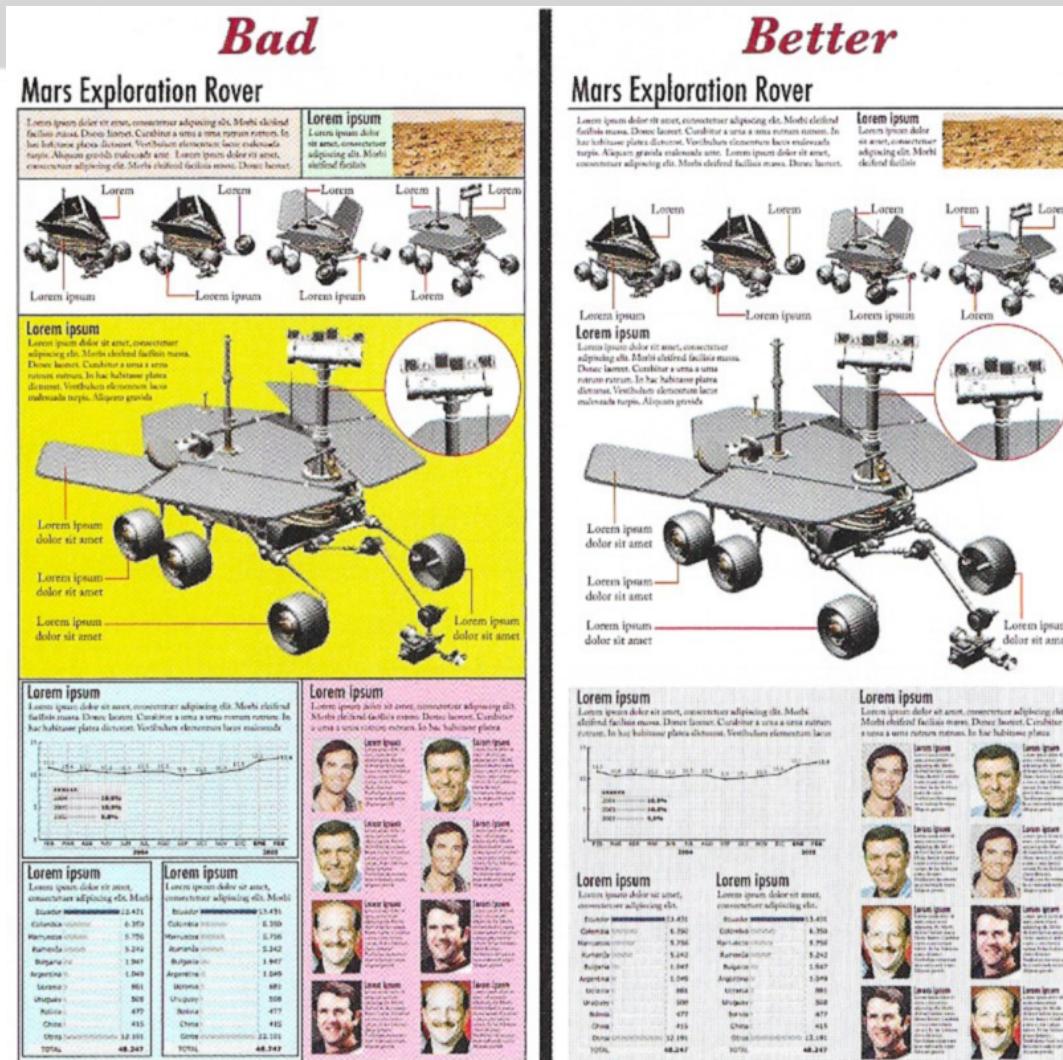
Survey shows that **demonstration of results** is the single most important dimension when choosing a service provider.

Affordability and experience working together previously, which were hypothesized to be very important in the decision making process, were both cited less frequently as important attributes.

- Poor placement of elements
- Poor visual hierarchy
- Little guidance to important elements

- Proximity to create groups
- Similarity (bright and dark bars) to highlight important elements
- Hierarchy by variation in typography (bold and regular type, black and gray type)
- Order by alignment and white space.

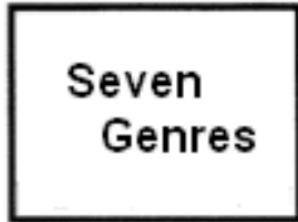
Gestalt principles for visual order (partitioned poster)



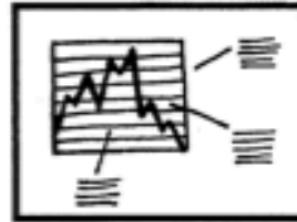
- Closure (i.e. boxes) for grouping creates sloppy look
 - Overloaded with boxes
 - Proximity for grouping
 - Clean look, less distracting background colour

- Clearly, the order that we place frames and arrange our visual space can have a time implication
 - Knowing this, we can weave a story with our visualisations
- The purpose of a narrative is to capture the *cognitive thread* for our viewers
 - The cognitive thread is the sequence of concepts we hold in working memory and how they are linked together
 - 7 narrative visualisation genres for linking visualisation frames

7 Genres of Narrative Visualisations



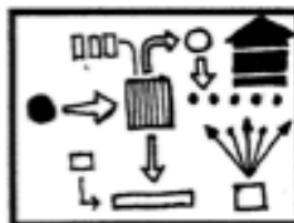
Magazine Style



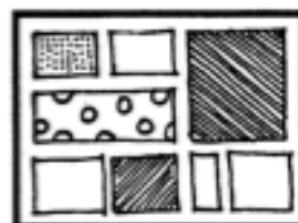
Annotated Chart



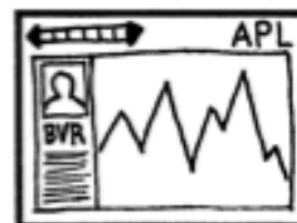
Partitioned Poster



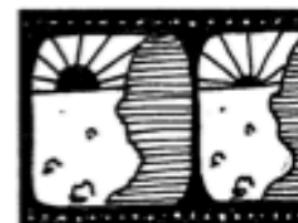
Flow Chart



Comic Strip



Slide Show



Film/Video/Animation

- Difference in number and order of frames
- Genres can be combined

Segel and Heer (2010), Narrative Visualization: Telling Stories with Data

Pulse of the Nation:

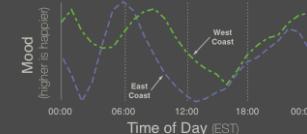
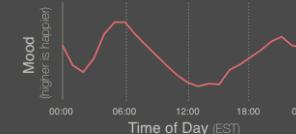
U.S. Mood Throughout the Day inferred from Twitter

All times are Eastern Standard Time (EST)



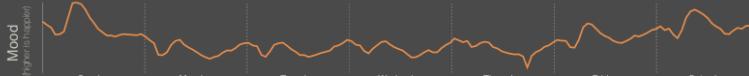
Mood Variations

A number of interesting trends can be observed in the data. First, overall daily variations can be seen (first graph), with the early morning and late evening having the highest level of happiness. Second, geographic variations can be observed (second graph), with a significantly happier west coast that is consistently three hours behind the east coast.



Weekly Variations

Weekly trends can be observed as well, with weekends much happier than weekdays.

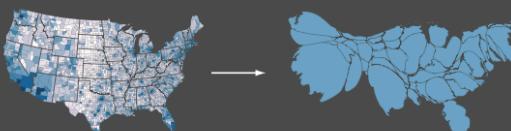


About the Data and Visualization

The plots were calculated using over 300 million tweets (Sep 2006 – Aug 2009) collected by MPI-SWS researchers, represented as density-preserving cartograms. The mood of each tweet was inferred using ANEW word list (Bradley, M.M., & Lang, P.J. *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings*. Technical report C-1, The Center for Research in Psychophysiology, University of Florida). County area data was taken from the U.S. Census Bureau at <http://factfinder.census.gov>, and the base U.S. map was taken from Wikimedia Commons. User locations were inferred using the Google Maps API, and mapped into counties using PostGIS and U.S. county maps from the U.S. National Atlas. Mood colors were selected using Color Brewer 2.

About Cartograms

A cartogram is a map in which the mapping variable (in this case, the number of tweets) is substituted for the true land area. Thus, the geometry of the actual map is altered so that the shape of each region is maintained as much as possible, but the area is scaled in order to be proportional to the number of tweets that originate in that region. The result is a density-equalizing map. The cartograms in this work were generated using the cart software by Mark E. J. Newman.



Northeastern University

College of Computer and Information Science [†]
Center for Complex Network Research [‡]

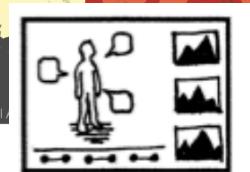
HARVARD UNIVERSITY [§]

16:00

15:00

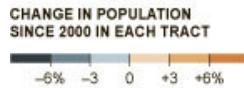
14:00

13:00

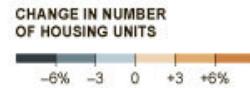
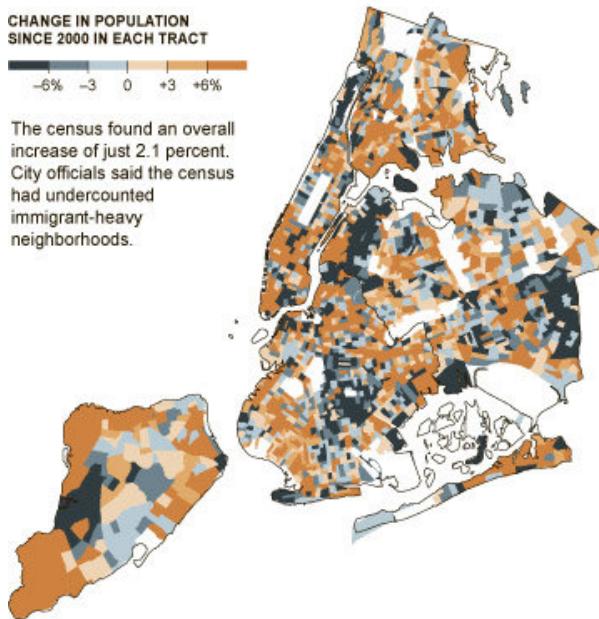


City Grew, but Less Than Expected

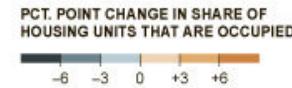
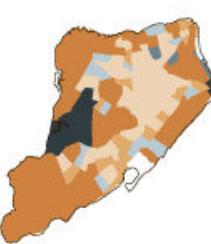
Over the last 10 years, New York housing expanded substantially. But the bust left many units vacant, and the city's population grew less than many had expected.



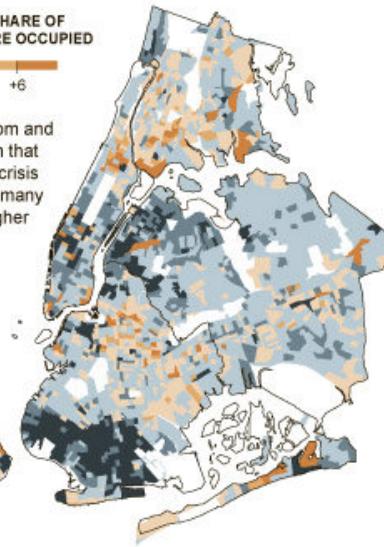
The census found an overall increase of just 2.1 percent. City officials said the census had undercounted immigrant-heavy neighborhoods.



A residential building boom early in the decade saw thousands of new units built across the city in places like Long Island City, Harlem, South Bronx, the Far West Side, Williamsburg and Bedford-Stuyvesant.

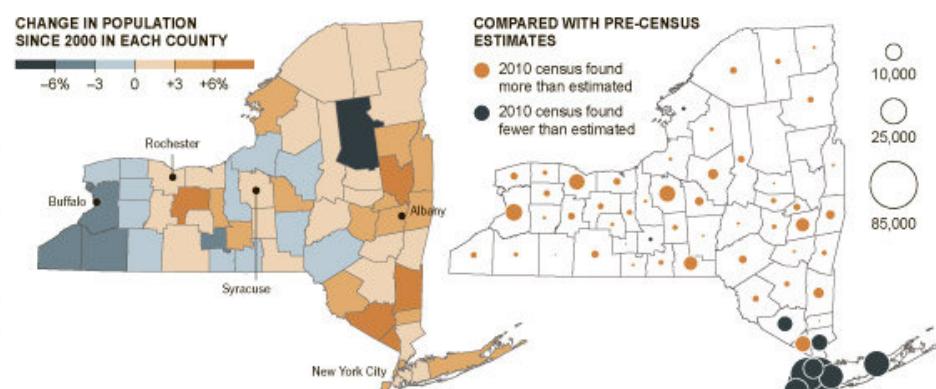


But the foreclosure boom and the economic downturn that followed the subprime crisis later in the decade left many neighborhoods with higher vacancy rates.



What had been expected Official 2010 census counts were lower than recent census estimates in the Democratic areas downstate, but higher than estimates in the Republican upstate.

BOROUGH	2000 CENSUS COUNT	2010 CENSUS COUNT	CHANGE SINCE 2000	PRE-CENSUS 2010 POPULATION ESTIMATE	DIFFERENCE FROM ESTIMATE	CHANGE IN POPULATION SINCE 2000 IN EACH COUNTY
Brooklyn	2,465,326	2,504,700	+1.6%	2,581,109	-76,409 -3.0%	
Queens	2,229,379	2,230,722	+0.1%	2,316,711	-85,989 -3.7%	
Manhattan	1,537,195	1,585,873	+3.2%	1,630,494	-44,621 -2.7%	
Bronx	1,332,650	1,385,108	+3.9%	1,400,761	-15,653 -1.1%	
Staten Island	443,728	468,730	+5.6%	492,714	-23,984 -4.9%	
New York City	8,008,278	8,175,133	+2.1%	8,421,789	-246,656 -2.9%	
NEARBY COUNTIES						
Suffolk	1,419,369	1,493,350	+5.2%	1,520,140	-26,790 -1.8%	
Nassau	1,334,544	1,339,532	+0.4%	1,363,061	-23,529 -1.7%	
Westchester	923,459	949,113	+2.8%	960,102	-10,989 -1.1%	



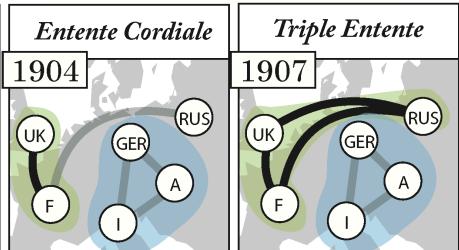
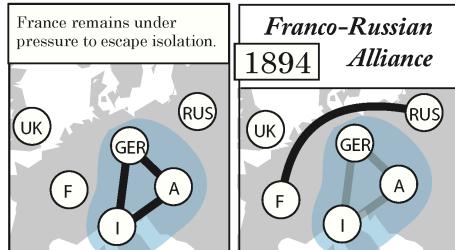
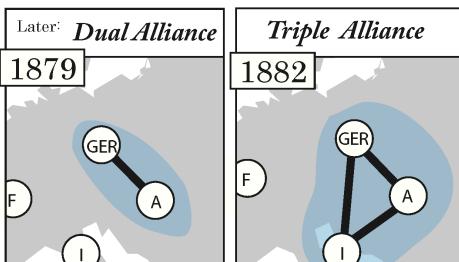
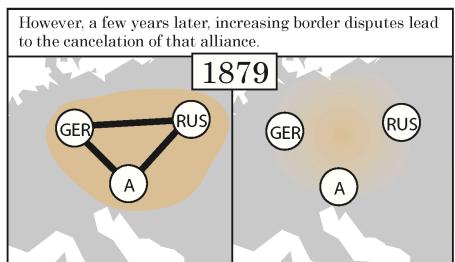
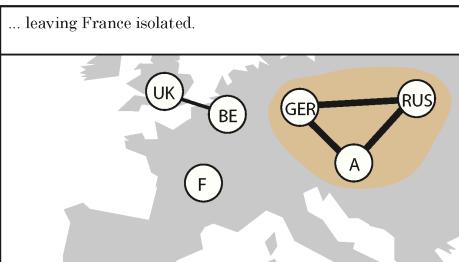
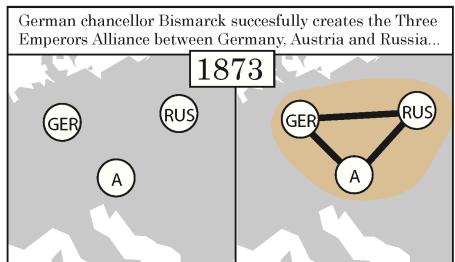
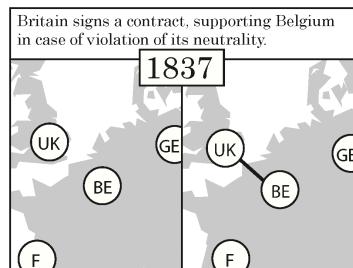
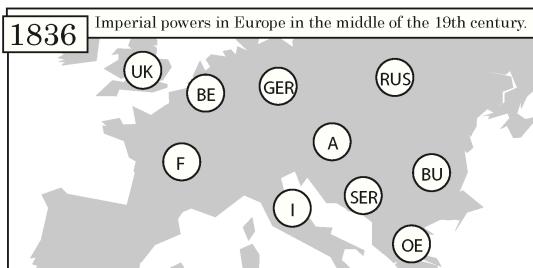
Sources: Census Bureau; Andrew A. Beveridge and Susan Weber, Social Explorer; Martin M. Dilan, New York state senator



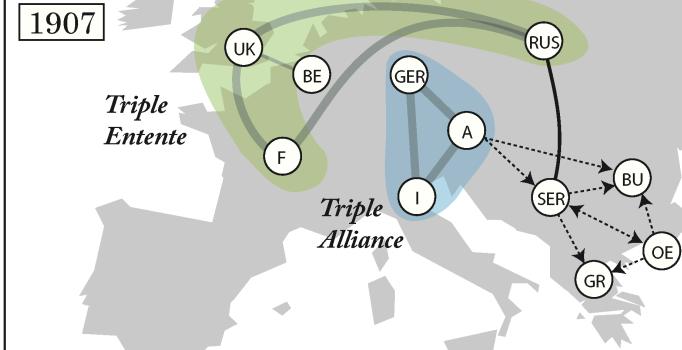
Partitioned Poster

MATTHEW ERICKSON AND FORD FESSenden/THE NEW YORK TIMES

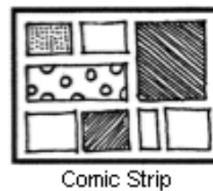
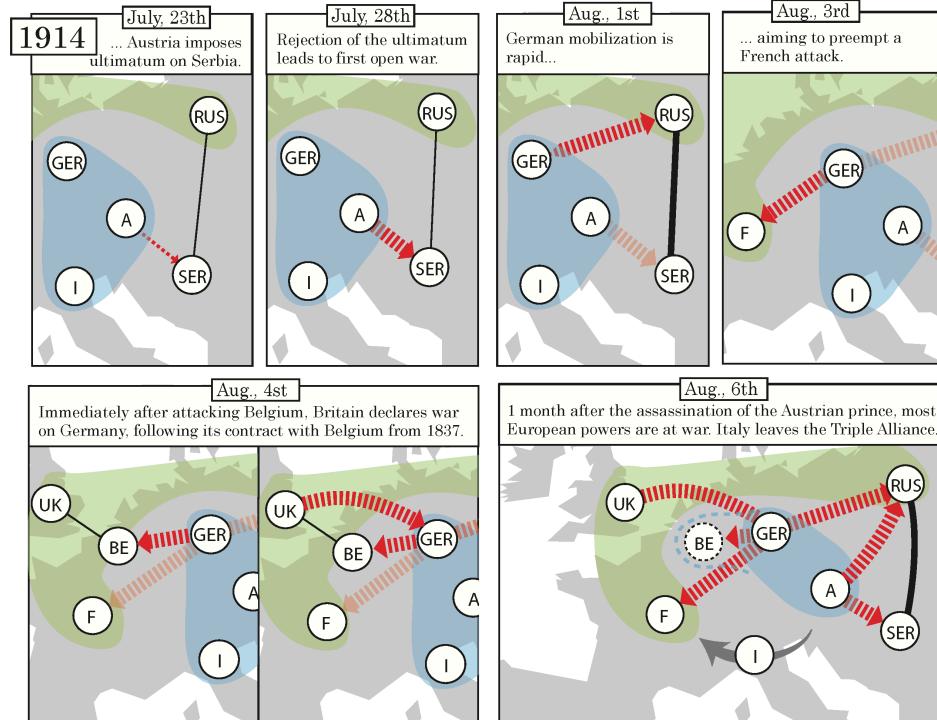
European Alliances before World War I (1836-1914)

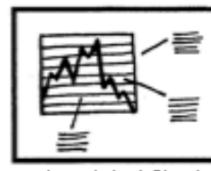
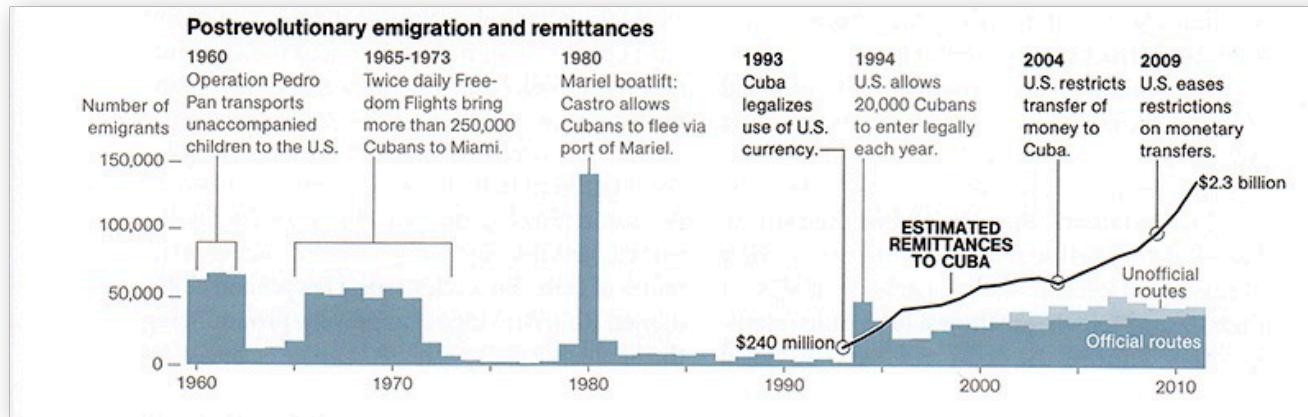


In 1907 the most important European powers are divided into two major alliances. On the Balkan, Serbia, Austria Greece and the Ottoman empire gamble for influence.



On June 28th, Archduke Franz Ferdinand of Austria is assassinated in Serbia while visiting Sarajevo. Austria accuses Serbian officials of being involved in the incident, aiming to annex Serbia. Russia wants Serbia to remain independent from Austria. During July 1914, European powers try to find a diplomatic solution, called the "July Crisis". However, ...





Annotated Chart

How America stopped prosecuting white-collar crime and public corruption, in charts

by Catherine Rampell • August 7

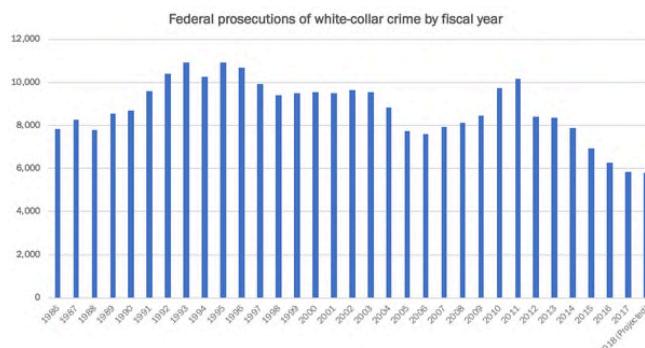


Paul Manafort arrives for a hearing at U.S. District Court in Washington in June. (Mandel Ngan/AFP/Getty Images)

Memo to all those supposed law-and-order fans out there, including [in the White House](#): The United States is currently on track to notch the fewest prosecutions of white-collar and public corruption crimes on record.

And presumably *not* because erstwhile white-collar criminals and corrupt public officials have suddenly started behaving themselves.

I mentioned this in my Tuesday [column](#), which was pegged to the glitzy Paul Manafort trial (and various other financial/corporate/political scandals in Trump World). The column included the following chart, showing the decline in federal white-collar crime prosecutions.



Source: Transactional Records Access Clearinghouse

These data come from Syracuse University's Transactional Records Access Clearinghouse. TRAC found that at the Justice Department's current pace, this fiscal year will see the fewest white-collar prosecutions on record (1986 is the



Magazine Style

Macro and micro reading

- Graphics should be readable in whole images and in parts
- *Macro reading* of the visualisation.
 - This is usually the **simplest and broadest reading** and often the one the designer has intended.
- *Micro readings* of the visualisation.
 - This is the **details** or subtle story and is often personal and unique to each viewer.
- Tufte:
 - “to clarify, add detail”: provide broad overview and include detail
 - "Such [micro/macro] designs can report immense detail, organizing complexity through multiple and (often) hierarchical layers of contextual reading"

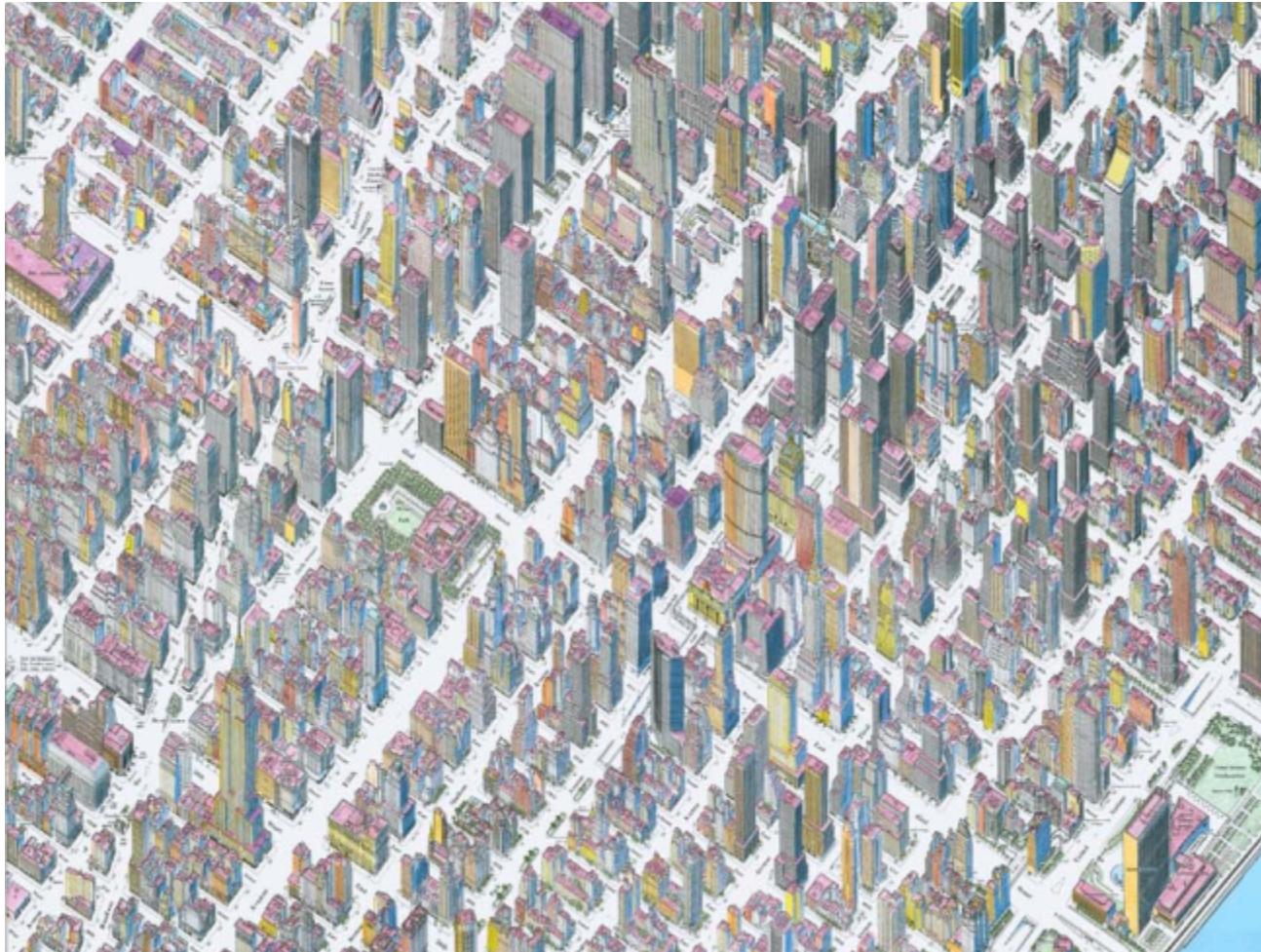
The question we have asked each time when designing our visualisations is “what do we want to say?”

- That broad question enables us to come up with *a macro reading...* the main point of the visualisation
- It is the general idea we get from viewing a visualisation

Micro readings are the details we can determine from the visualisation

- “Individual stories” about the data and information being presented
 - Finding these stories may be deliberate on the part of the designer, or things that the viewer discovers
- Let’s take an example...

Constantine Anderson's NYC Map



Constantine Anderson's NYC Map



This is a really famous example, one that *Tufte* discusses

- At a *macro level*, you get a sense of the whole of Manhattan
 - The scale of the buildings, the streets, the layout, and where main things are
- A *micro level* reading reveals the details...
 - A micro reading shows over 2000 pieces of text, including the names of buildings, parks, streets subway stations and the like
 - But this doesn't get in the way of a macro reading
 - The reader brings their own view to it.



Macro:

- Large landforms
 - Road network
 - Rivers and lake

Micro

- Buildings
 - Creeks
 - Paths
 - Orchards
 - Swamps
 - etc.

Macro

- Path and loss of the army
- Effect of temperature

Micro

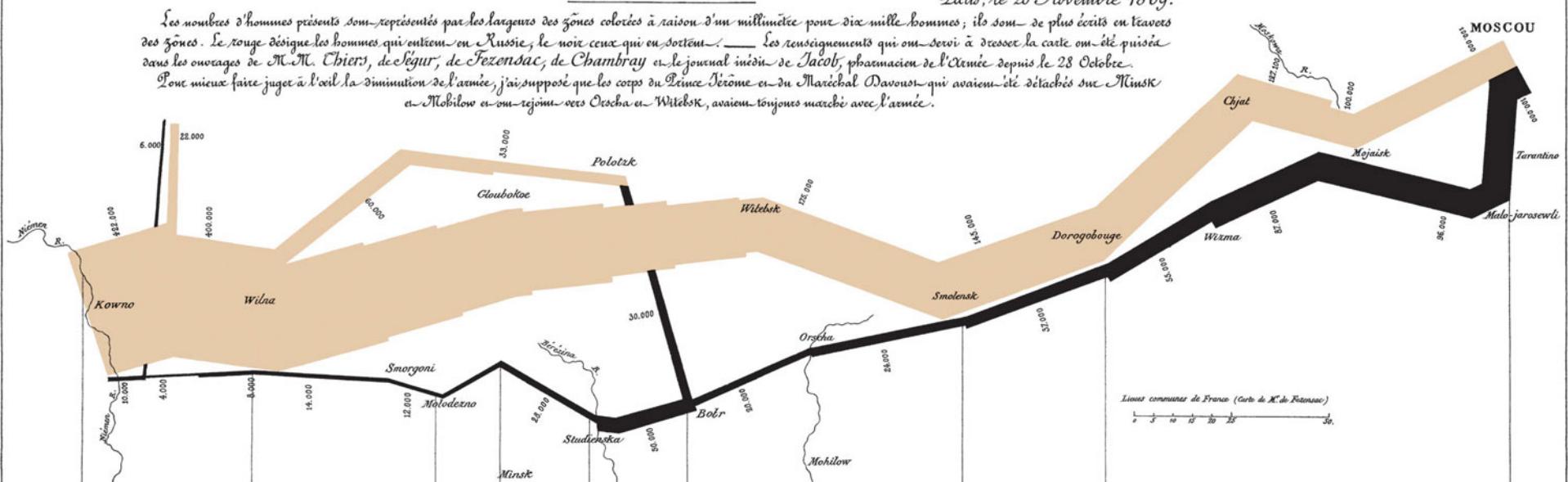
- Fluctuation of temperature over space and time
- Place names and river names

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à desser la carte ont été puisé dans les ouvrages de M. M. Chier, de Cléger, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk en Mohilow et se rejoignent vers Orsha en Witelsk, avaient toujours marché avec l'armée.


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les cosaques passent au galop
le Niemen gelé.



Macro/Micro



Creating visualisations that work on multiple levels is not easy

- Large datasets are needed
- However always ask yourself how you can make your visualisations provide these multiple layers of meaning
 - A thoughtful consideration of different narratives means that you're thinking of different types of WHY
 - This leads to better visualisations, better knowledge and richer wisdom being generated in the users!
- Consider interactive zoomable interfaces

How to lie with data visualisation

Our visualisation is like a news story in some ways. It is quite easy to slant a story in a visualisation, create bias or even lie. We need to know how to do it and how to avoid doing it if we prefer honesty. We want to avoid telling the wrong story and lying without meaning to.

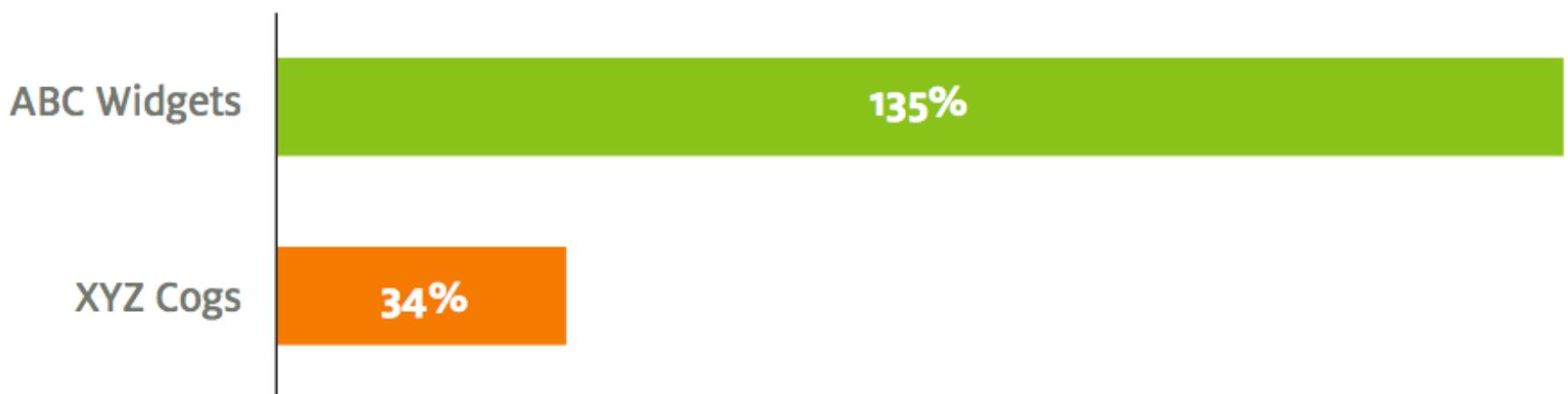
We can use graphical techniques to misinterpret or misrepresent data (without being strictly wrong)

- Things we might do
 - Comparing two data sets that have no real meaning together (we can impose wrong meaning)
 - Play with colour or images to associate things that shouldn't be linked together
 - Mess with things like perspective
 - Fudge numbers and scales

- Absolute figures can mean one thing, while relative comparisons can say something else
- Consider absolute...



- Now relative...
- What tells the better story? What story do you want to tell?

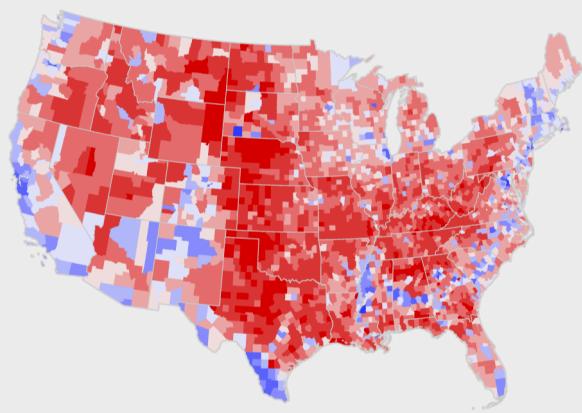


Lies with proportional vs non-proportional

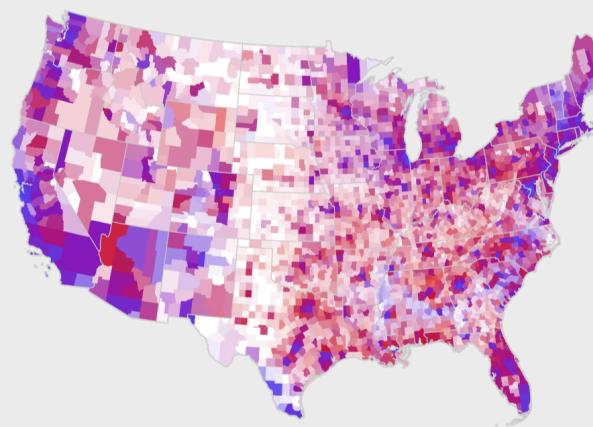
- What party won?



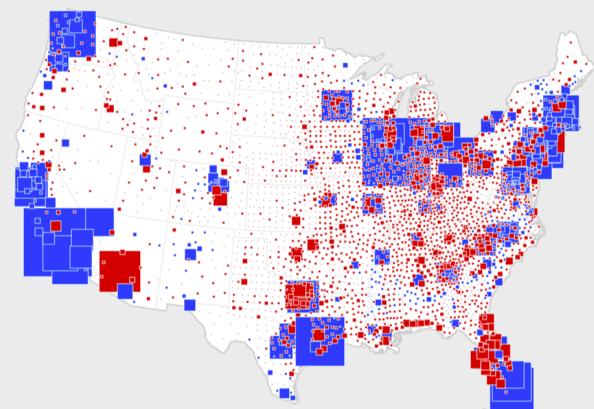
Lies with proportional vs non-proportional



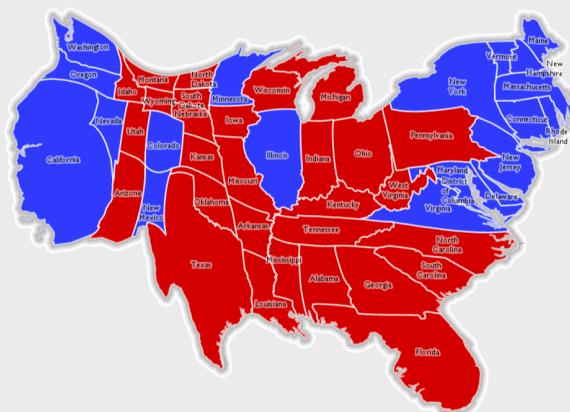
Choropleth
10 equally spaced classes



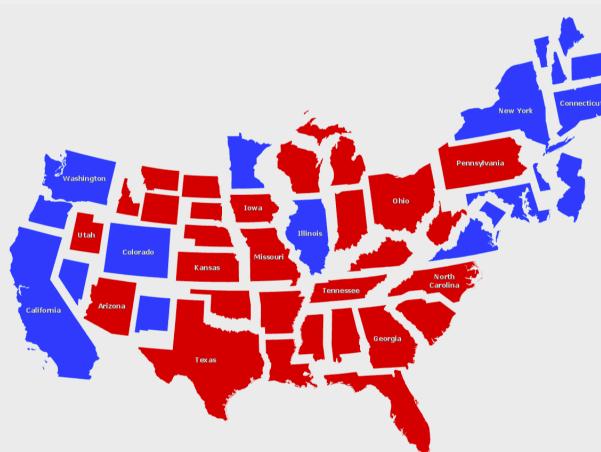
Value-by-alpha
Transparency changes with
number of votes



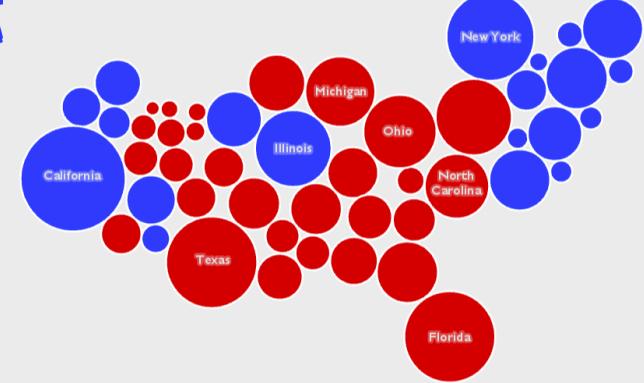
Area-proportional symbols



Contiguous cartogram



Non-contiguous cartogram



Dorling cartogram

Source: Ken Field. Check out his gallery of US election maps:

<http://carto.maps.arcgis.com/apps/MinimalGallery/index.html?appid=b3d1fe0e8814480993ff5ad8d0c62c32>

Lies with proportional vs non-proportional

- What about population in geography?

Total Population of 2000 Census Block Groups Population Density of 2000 Census Block Groups



- What about population in geography?

Most dangerous cities

Total murders in 2014

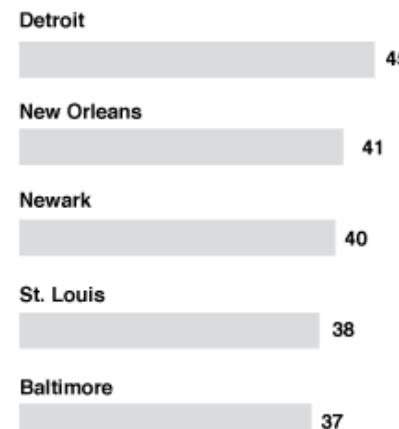
WRONG



Most dangerous cities

Murder rate in major US cities in 2014,
per 100,000 people

RIGHT

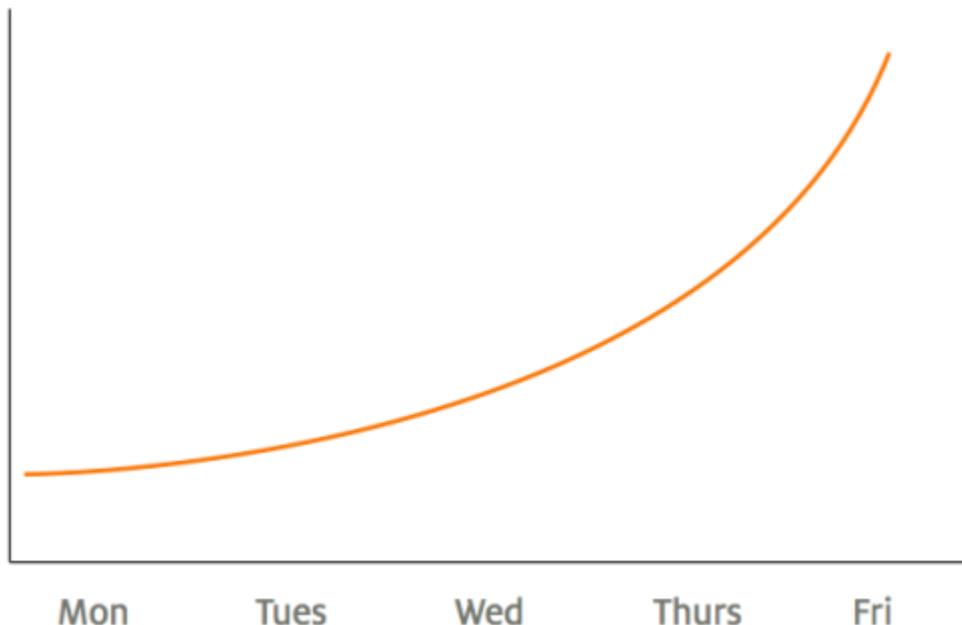


Lies with Playing with Scales

- These may be the same thing... which has more impact?

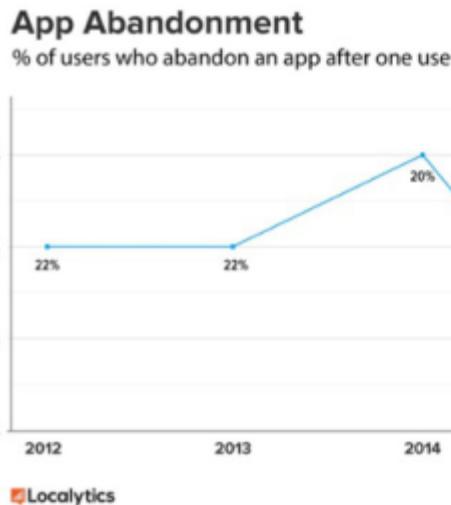


- What does this say if we don't put a scale on?
- But what does it really mean?



Lies with Playing with Axes

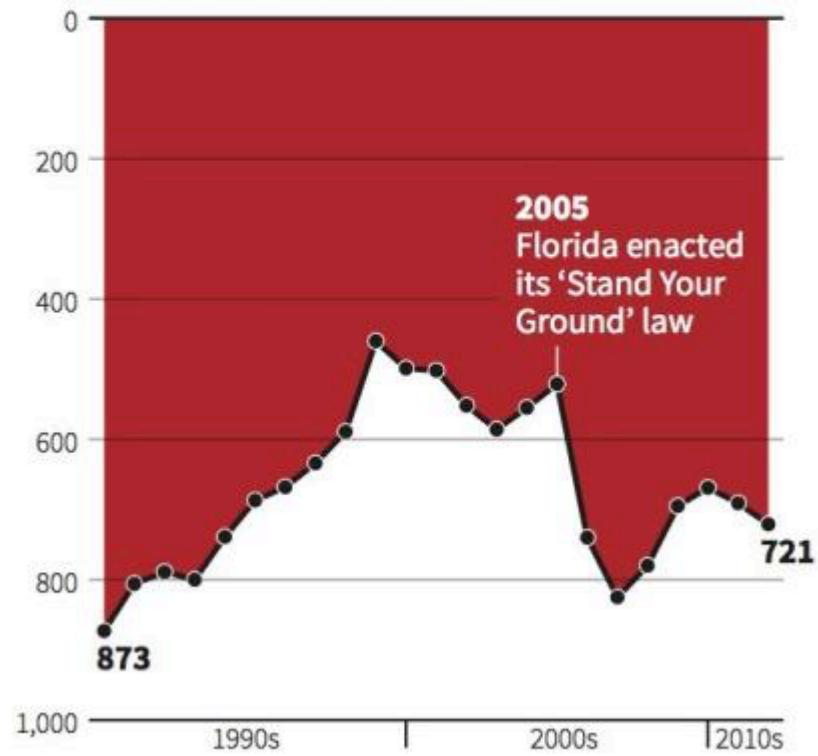
- What do you notice about the y-axis?



Source: Yunke Hu in the Moodle forums FIT2079, 2016

Gun deaths in Florida

Number of murders committed using firearms



C. Chan 16/02/2014

REUTERS

Lies with Tricking the Eye

- Look at these comparisons



Hippo #1
2 units



Hippo #2
4 units



Hippo #1
2 units



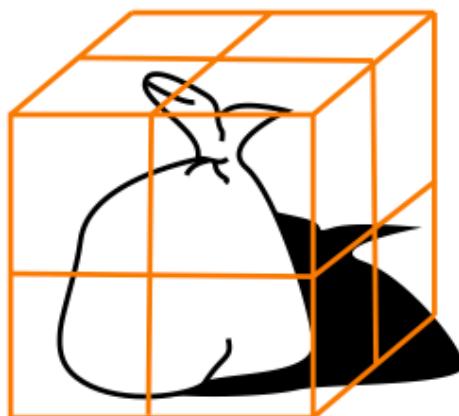
Hippo #2
4 units

Lies with Tricking the Eye

- Sure the height is right, but the bigger bag looks more than twice the size of the smaller one!
- We may mean the second one (but it lacks impact)



Hippo #1
2 units



Hippo #2
4 units



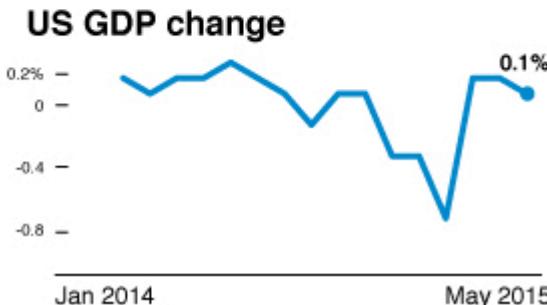
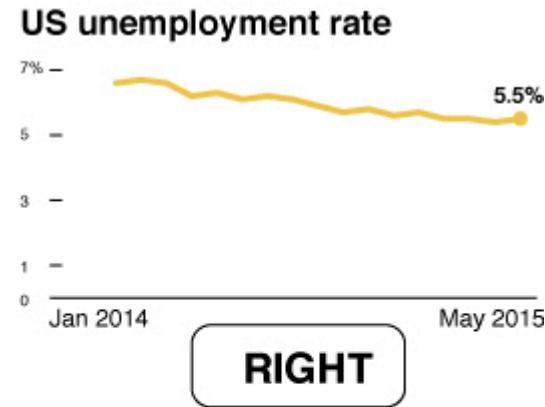
Hippo #1
2 units



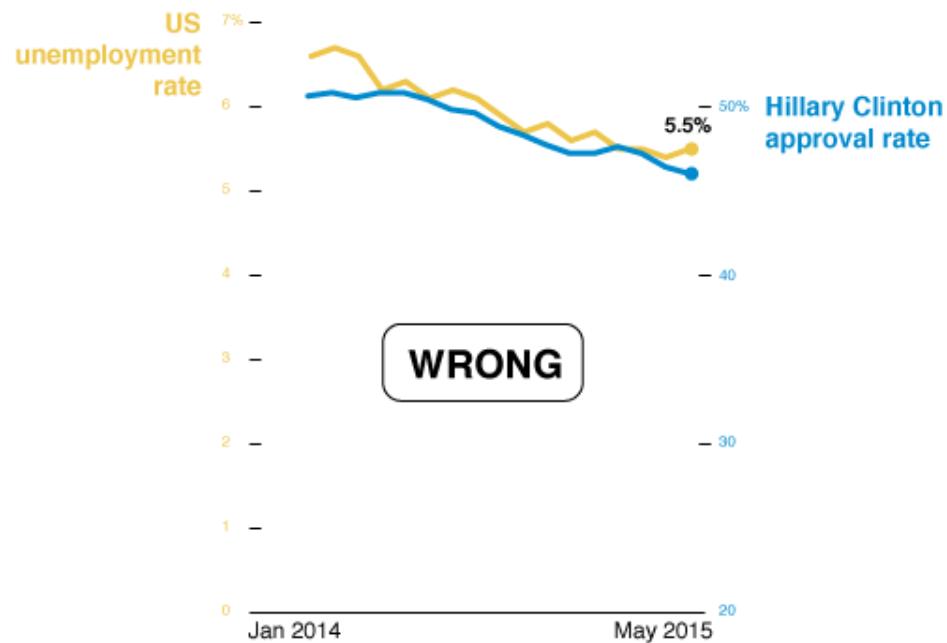
Hippo #2
4 units

Lies with Improper Comparisons

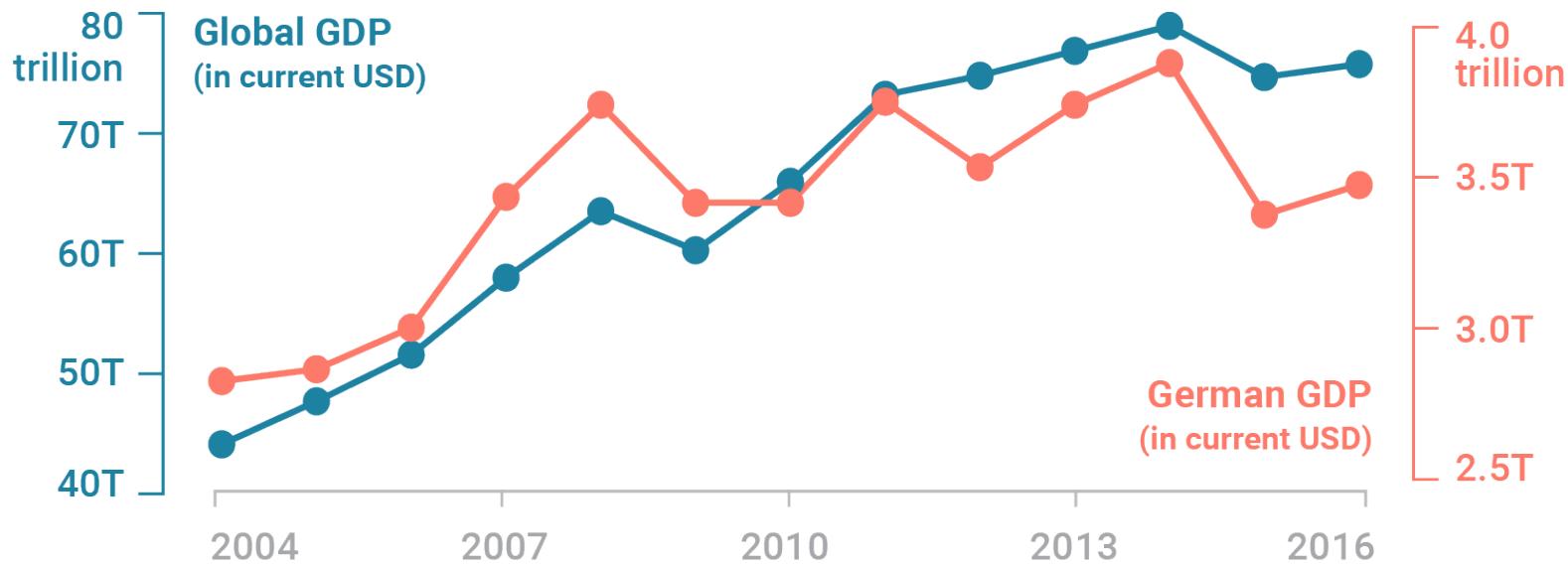
- What if I put two different items on a visualisation... would we associate them together?



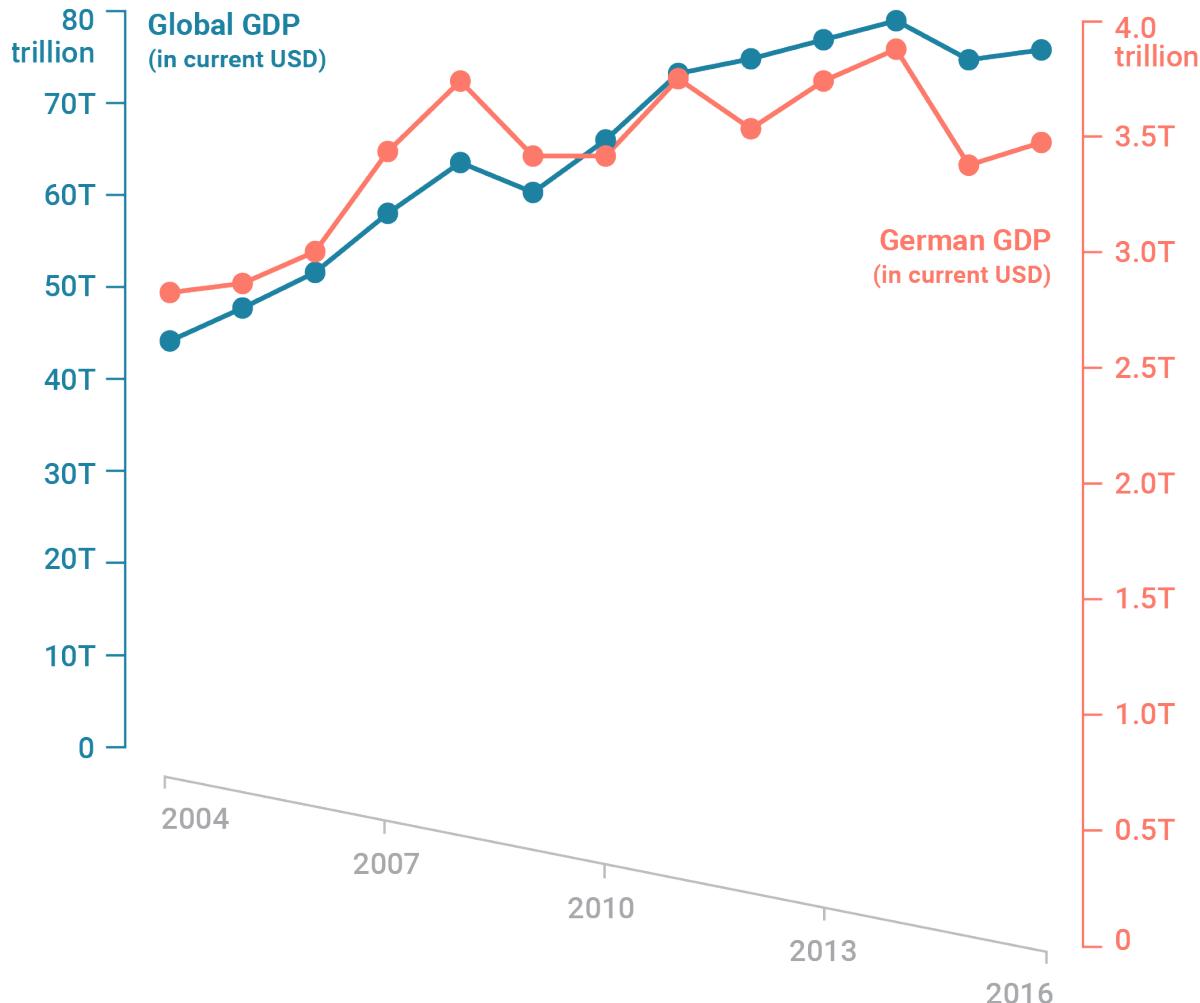
Just because two sets of numbers follow a similar path doesn't mean there is a correlation.



Problems of charts with two axes



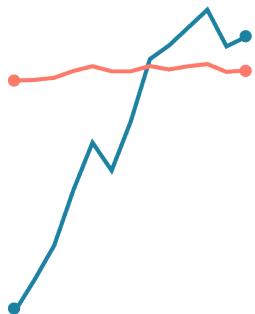
Problems of charts with two axes



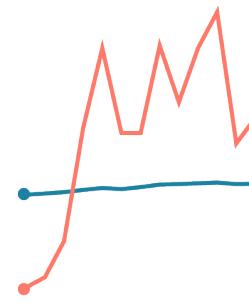
Source: Lisa Charlotte Rost, <https://blog.datawrapper.de/dualaxis/>

Problems of charts with two axes

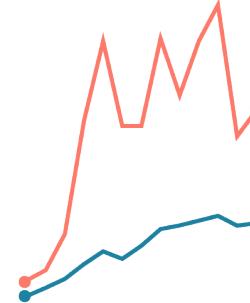
Just tweak the scales a little bit and you can express whatever you like!



Orange steady,
Blue massively increasing.



Blue steady,
Orange increasing.



Both started at the same
level, but Orange increased
far more than Blue.



Both started at the same
level, but Blue increased far
more than Orange.



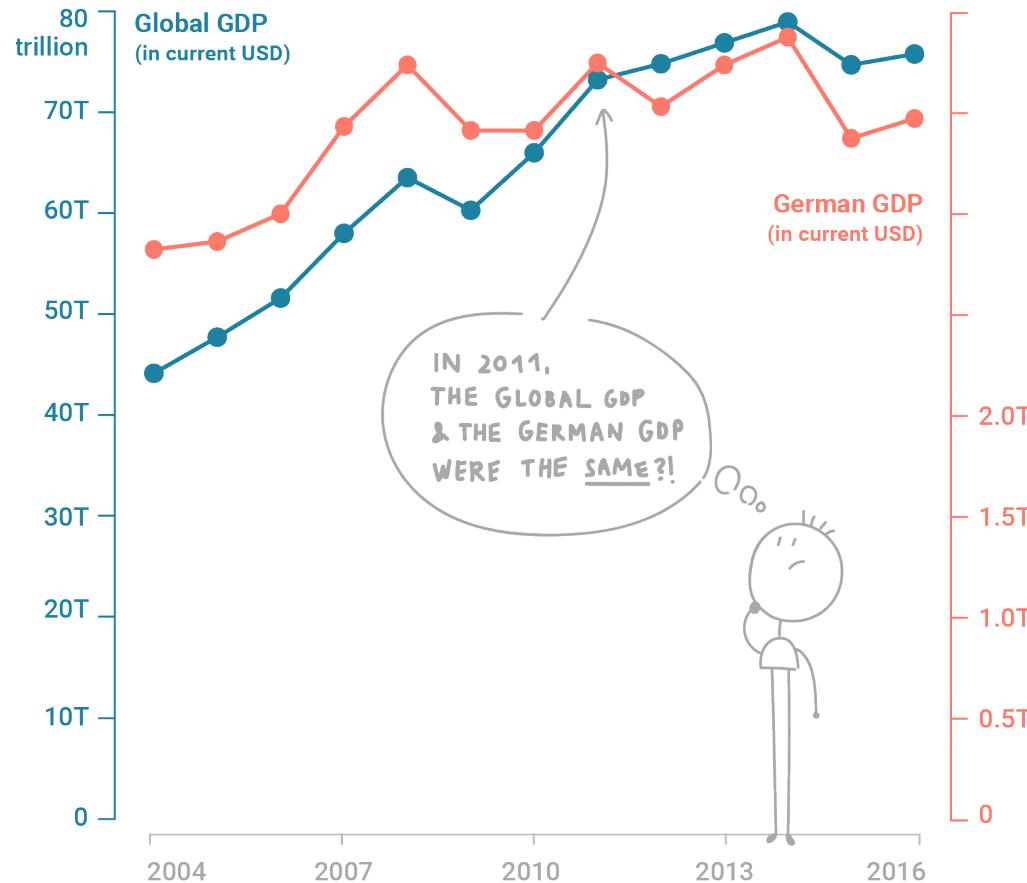
Both started with the
same increase, then Blue
raced to the top.



Both steady.

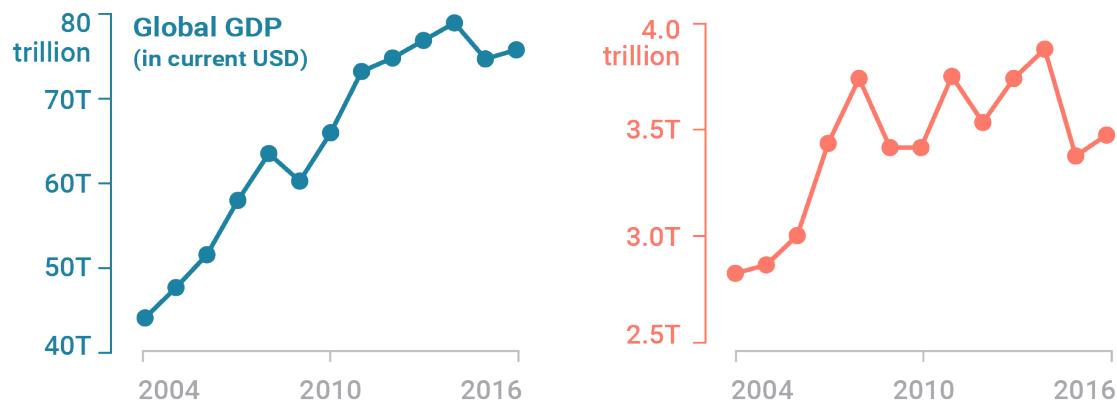
Problems of charts with two axes

Zero baselines at the same height can still be misleading.

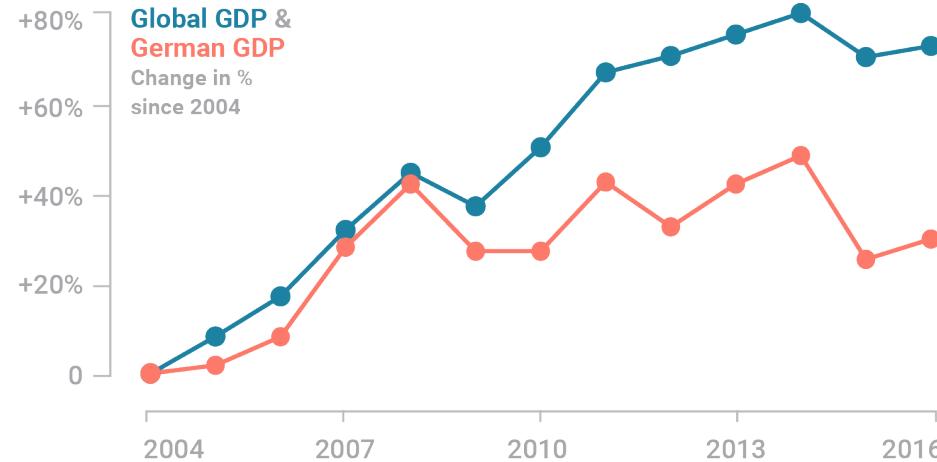


Problems of charts with two axes

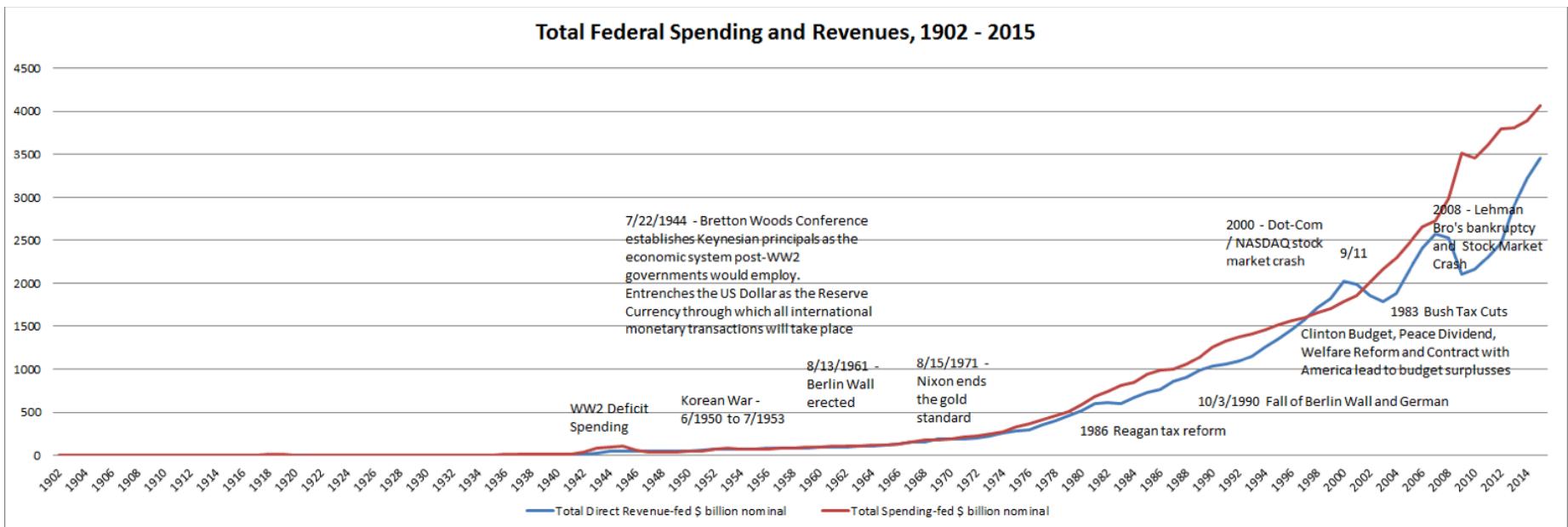
Alternative: 2 charts



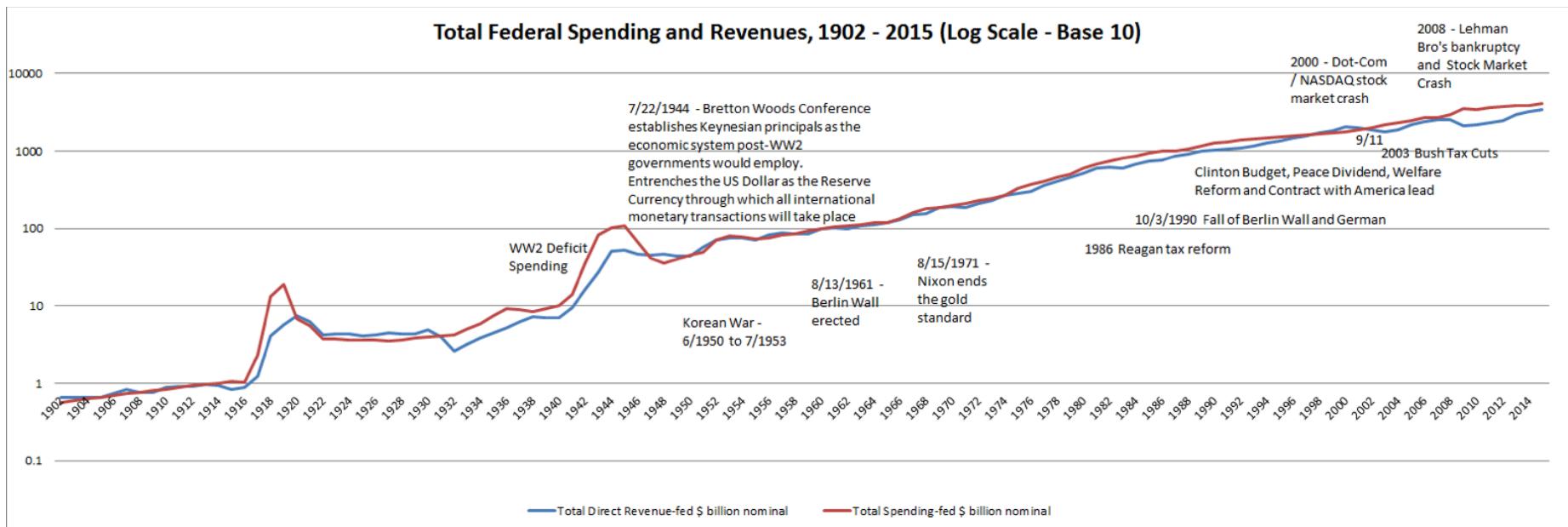
Alternative: Relative change



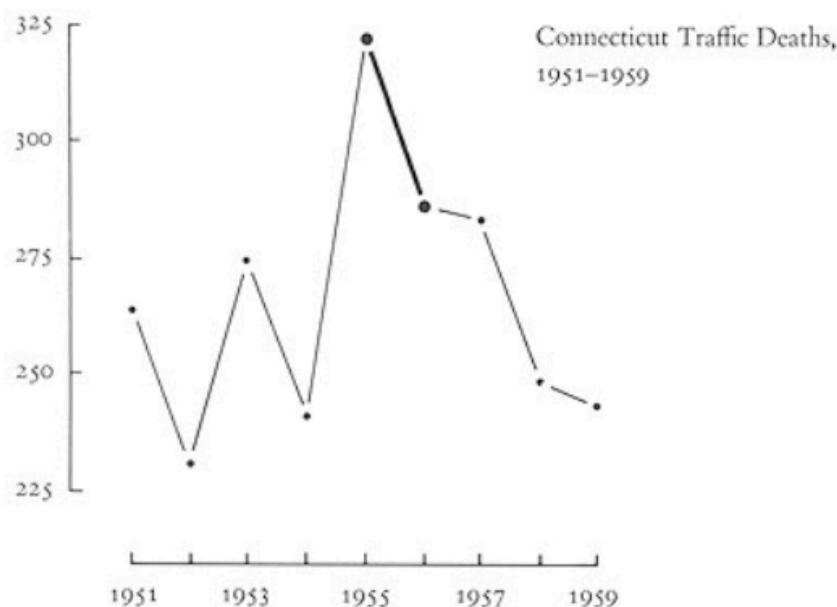
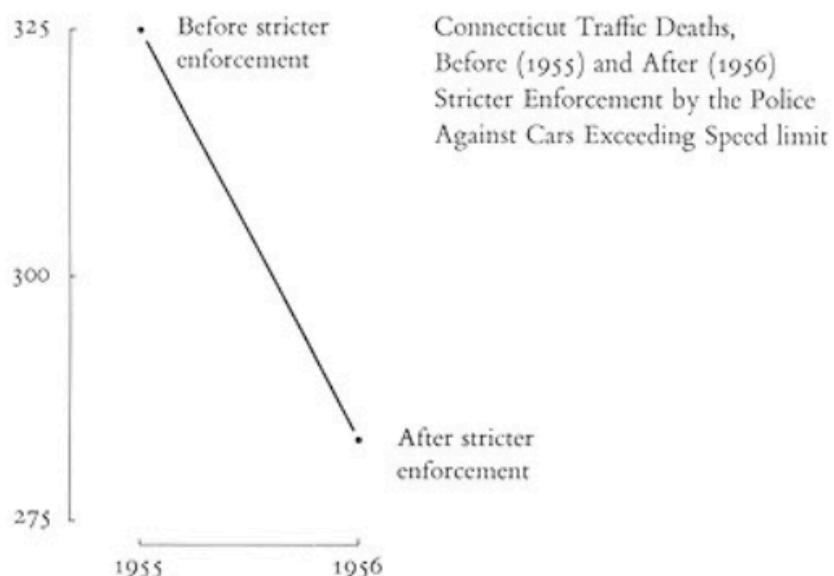
Lies with Linear vs Log Scale



Lies with Linear vs Log Scale



Lies with Selective Truth



You must consider all these things when creating your visualisations.

- Too often these things are done without meaning.
- Sometimes they are done on purpose.

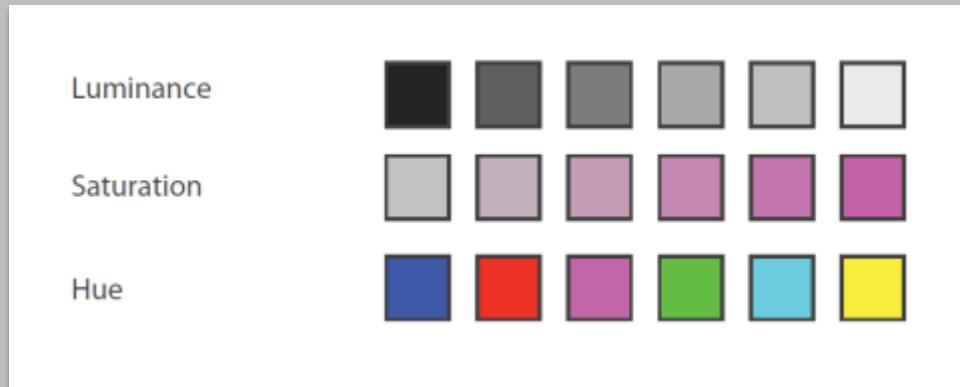
Remember that our visualisations are intended to create *wisdom* in the viewer and help them make decisions and discoveries

- Just make sure you are saying what you mean to say!

Effective use of colour for data visualisation

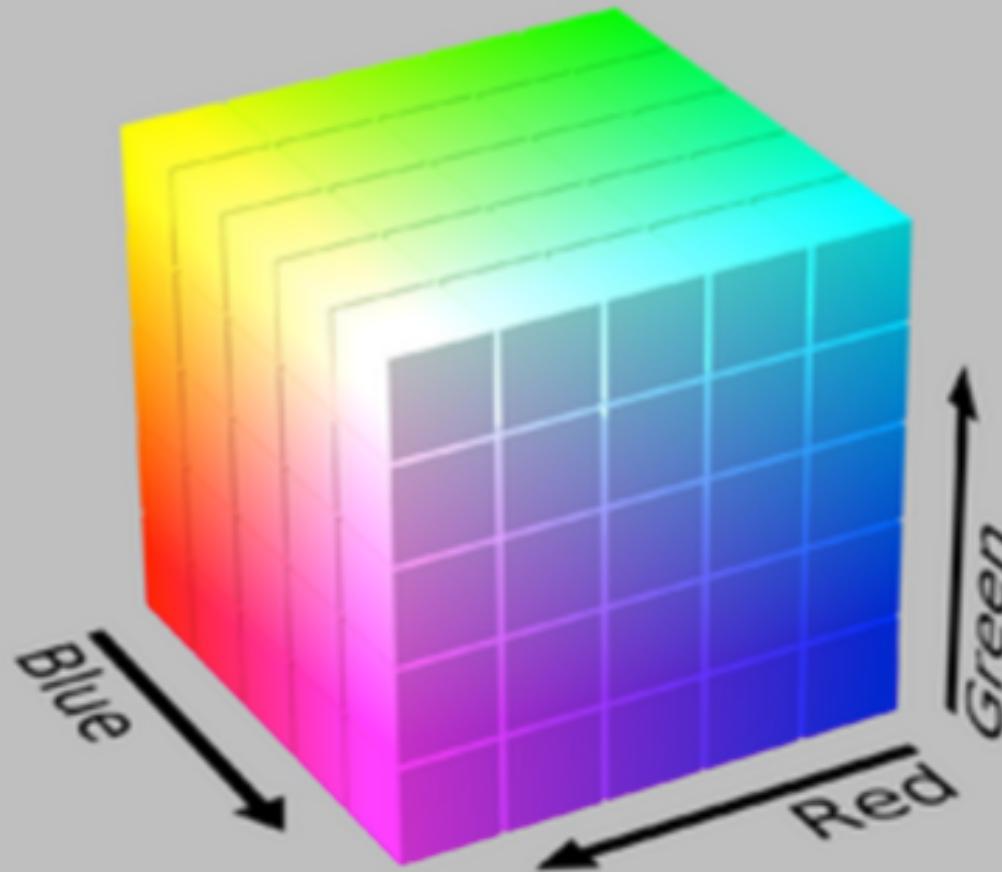
VAD chapter 10

Colour Spaces



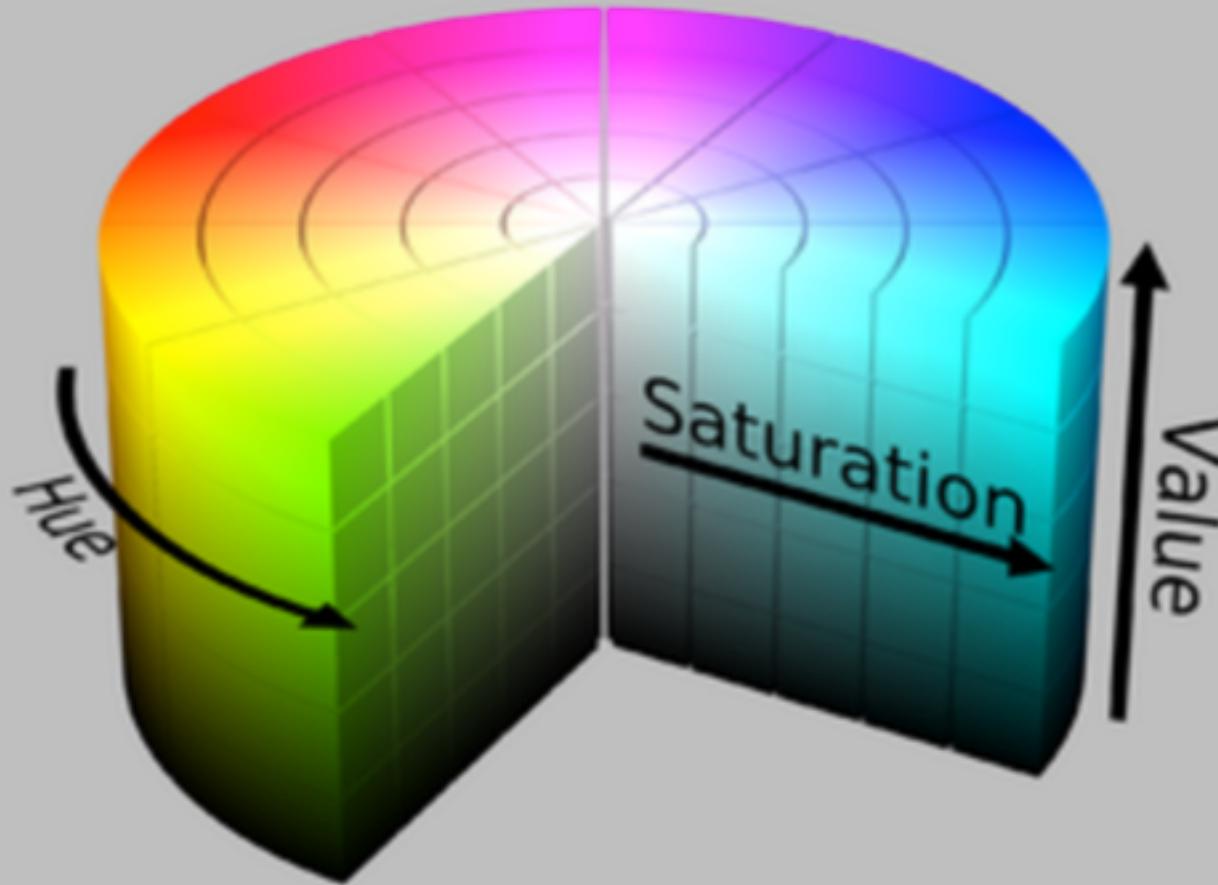
Colour Spaces

RGB



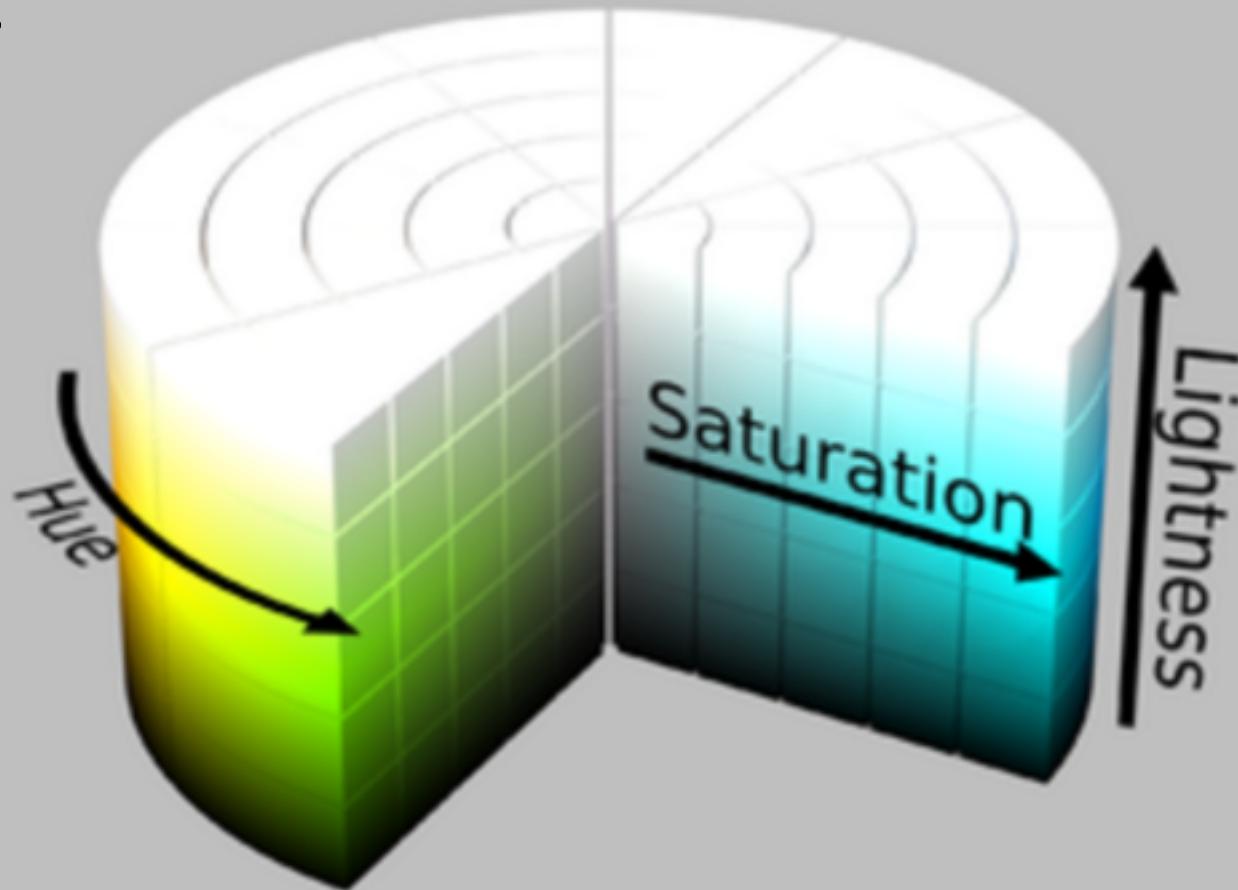
Colour Spaces

HSV



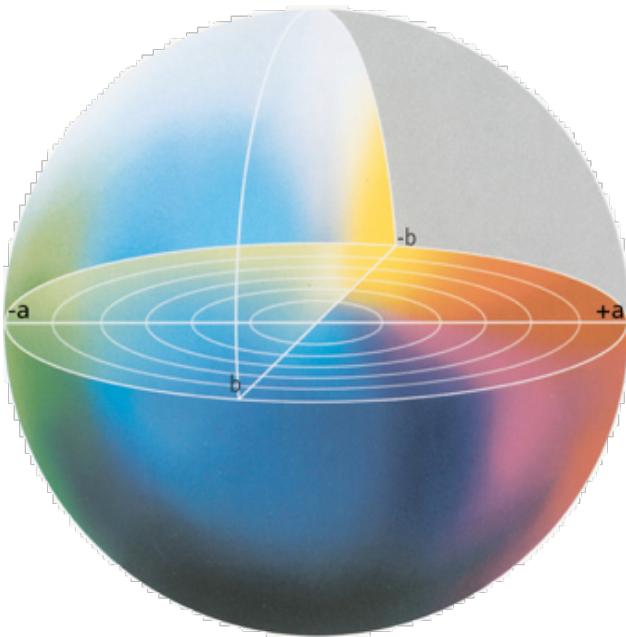
Colour Spaces

HSL



$L^*a^*b^*$: Perceptually Linear Colour space

- L^* : luminance axis
- a and b: red-green and blue-yellow axes



Corners of the RGB color cube



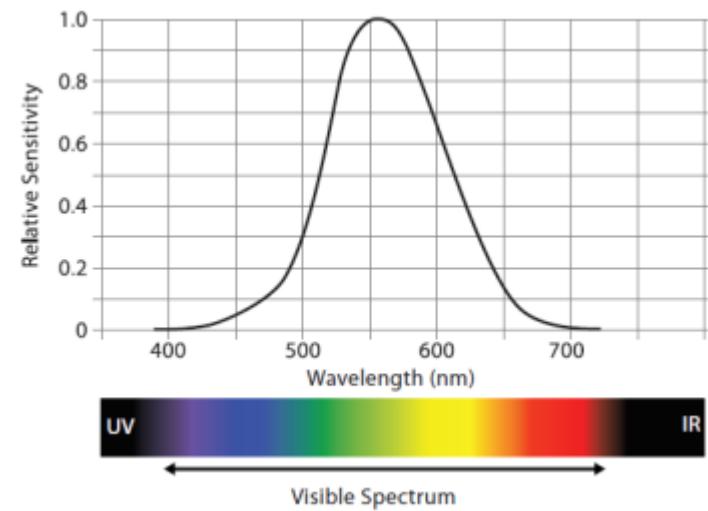
L from HSL
All the same



Luminance



L^*



Colour Spaces

- RGB, HSV, HSL, XYZ, L*a*b* have 3 axes
- RGB, HSV, HSL, XYZ are not visually equidistant, that is, the numerical mean of two colours is not the visual mean of these colours.
- L*a*b* is visually equidistant.
- HSV and HSL used for colour pickers, because of their intuitive axes.
- XYZ and L*a*b* used for colour management, that is, matching colours between different devices.

The Evil Rainbow Colourmap

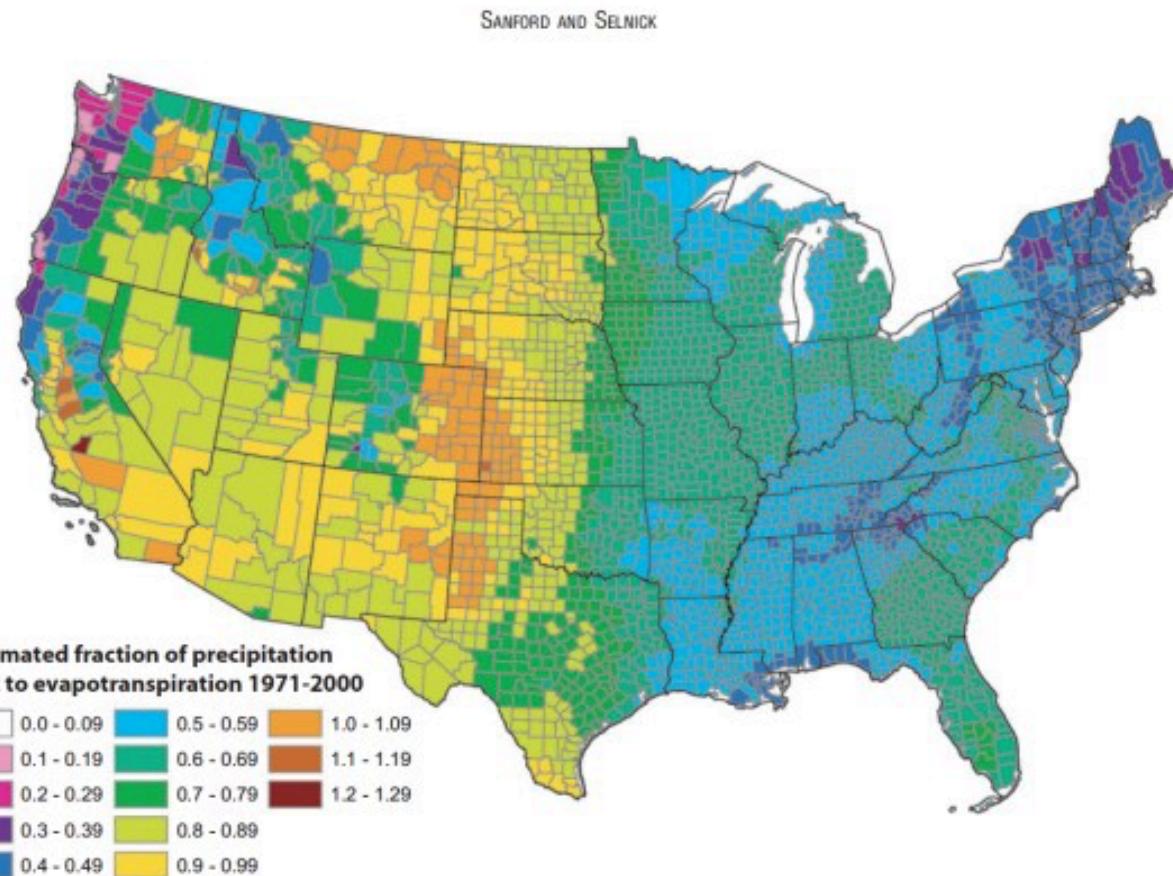


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

The Evil Rainbow Colourmap

SANFORD AND SELNICK

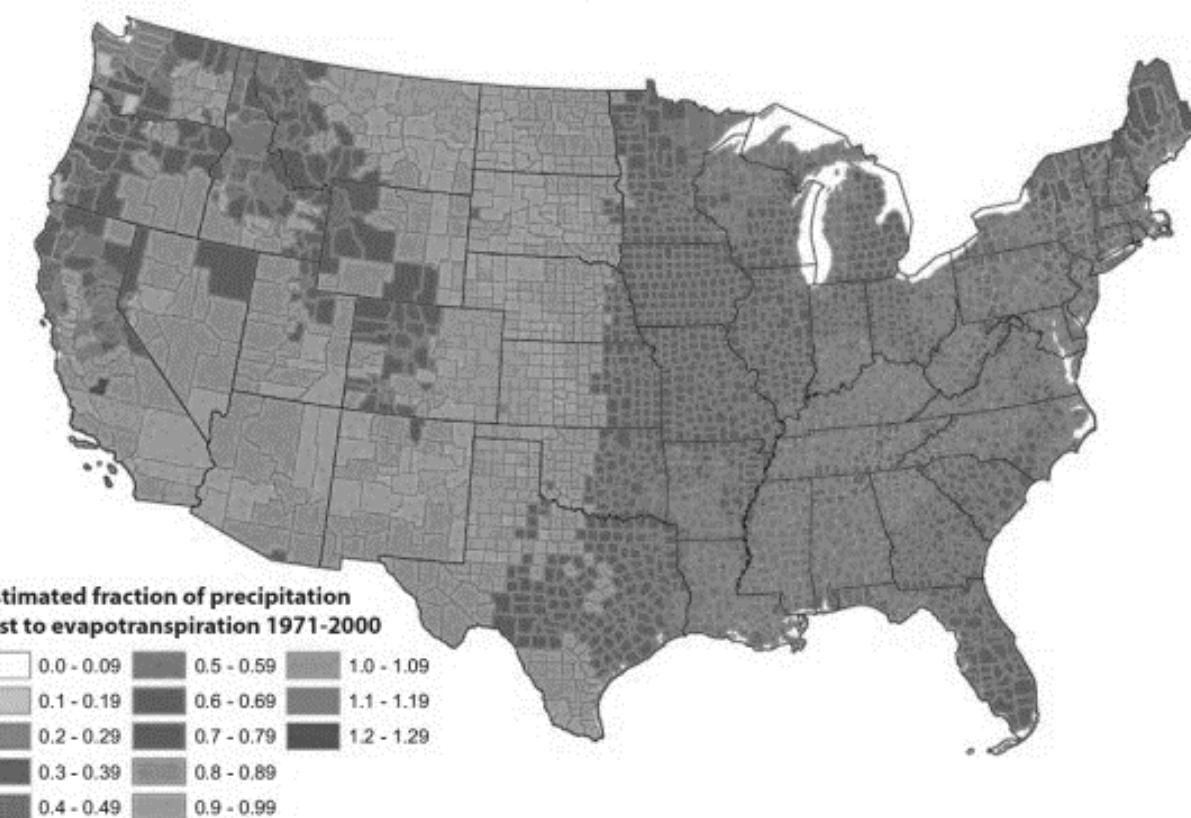


FIGURE 13. Estimated Mean Annual Ratio of Actual Evapotranspiration (ET) to Precipitation (P) for the Conterminous U.S. for the Period 1971-2000. Estimates are based on the regression equation in Table 1 that includes land cover. Calculations of ET/P were made first at the 800-m resolution of the PRISM climate data. The mean values for the counties (shown) were then calculated by averaging the 800-m values within each county. Areas with fractions >1 are agricultural counties that either import surface water or mine deep groundwater.

May 29, 2018

Thoughts & How To's

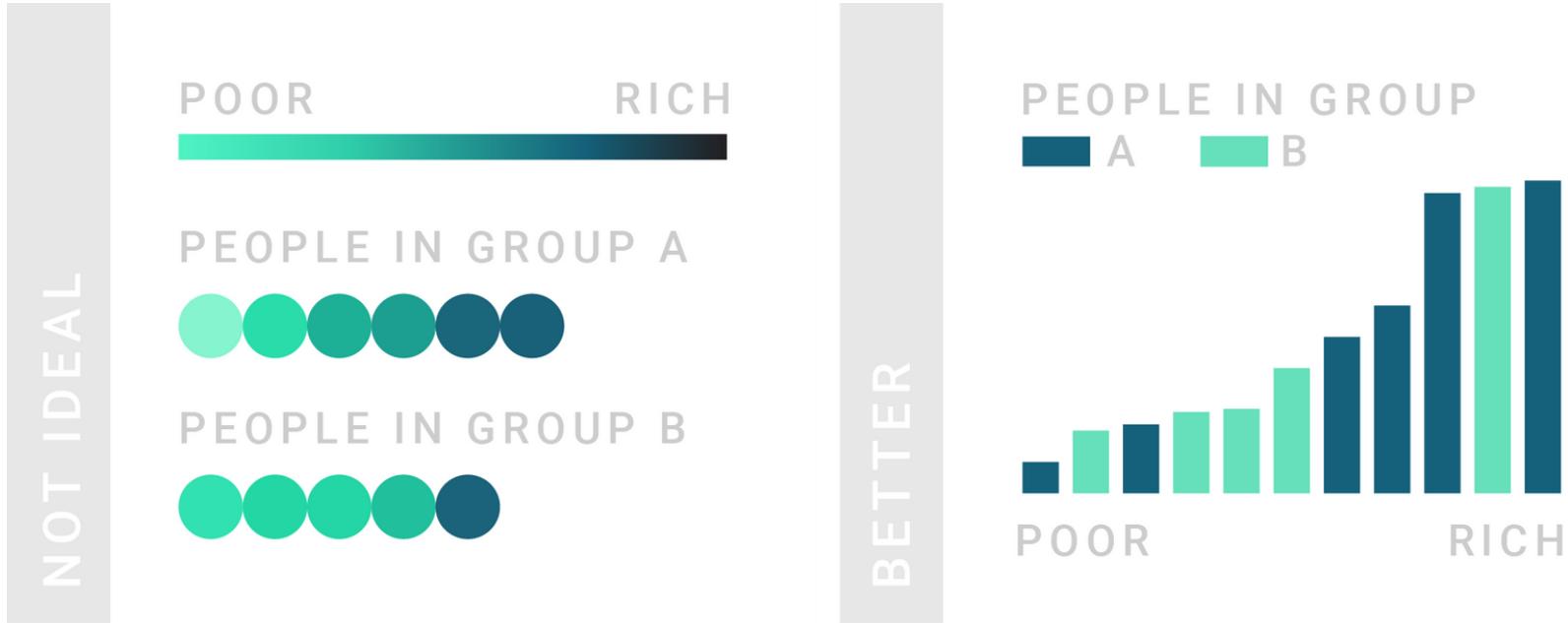
by Lisa Charlotte
Rost

What to consider when choosing colors for data visualization

Data Visualisation can be defined as representing numbers with shapes – and no matter what these shapes look like (areas, lines, dots), they need to have a color. Sometimes colors just make the shapes visible, sometimes they encode data or categories themselves. We'll focus mostly on the latter in this article. But we'll also take a general look at colors and what to consider when choosing them:

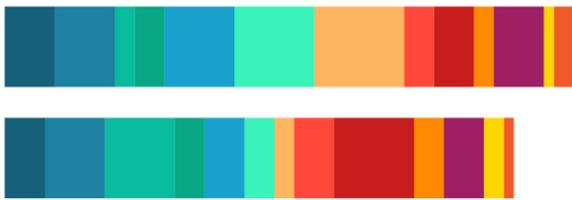
<https://blog.datawrapper.de/colors/>

Choosing Colour for Data Visualisation



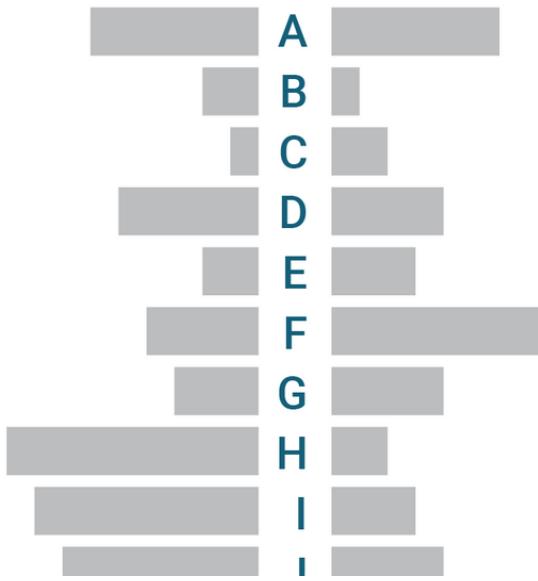
Choosing Colour for Data Visualisation

NOT IDEAL



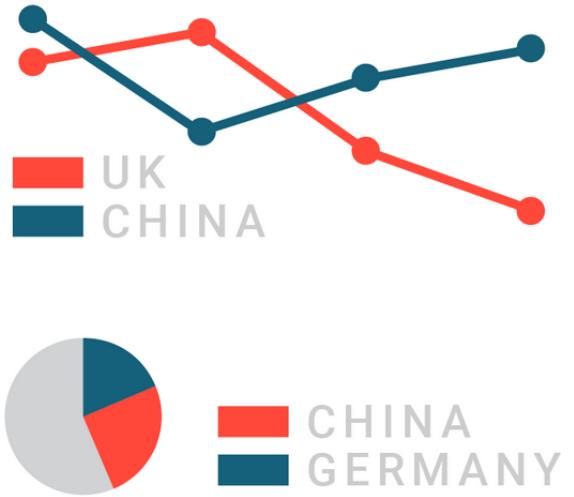
A	B	C
D	E	F
G	H	I
J	K	L
M		

BETTER



Choosing Colour for Data Visualisation

NOT IDEAL



BETTER



Choosing Colour for Data Visualisation

COLOR KEY



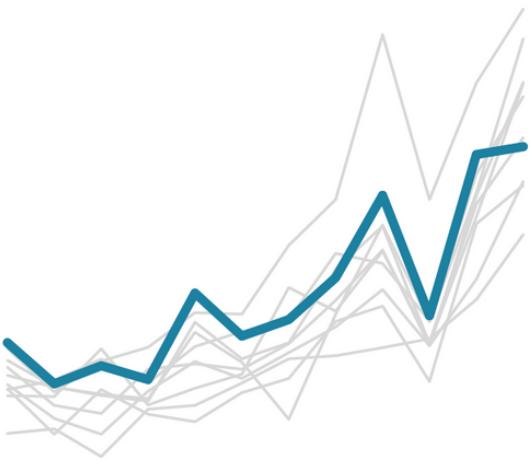
SHARE OF
PEOPLE IN
CHINA AND
GERMANY

Choosing Colour for Data Visualisation

NOT IDEAL



BETTER



Choosing Colour for Data Visualisation

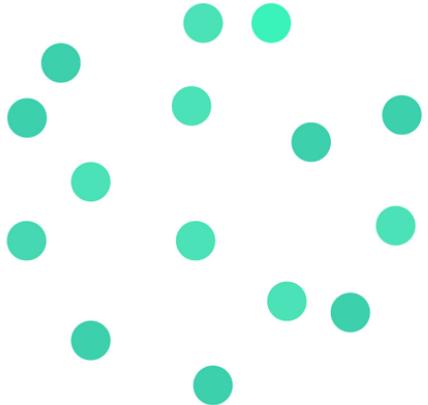
CONTRAST

CONTRAST RATIOS

1.0	A light grey horizontal bar.	A medium grey horizontal bar.	A bright green horizontal bar.	A bright green horizontal bar.
1.1	Choose if you dislike readers.	That's bad.	That's bad.	Horrible.
1.5	Ok in 1% of the cases.	Not ideal.	That's bad.	My eyes!
2.5	Can be a good choice.	Ok.	Not ideal.	That's bad.
4.5	Safe choice.	Great.	Ok.	Not ideal.

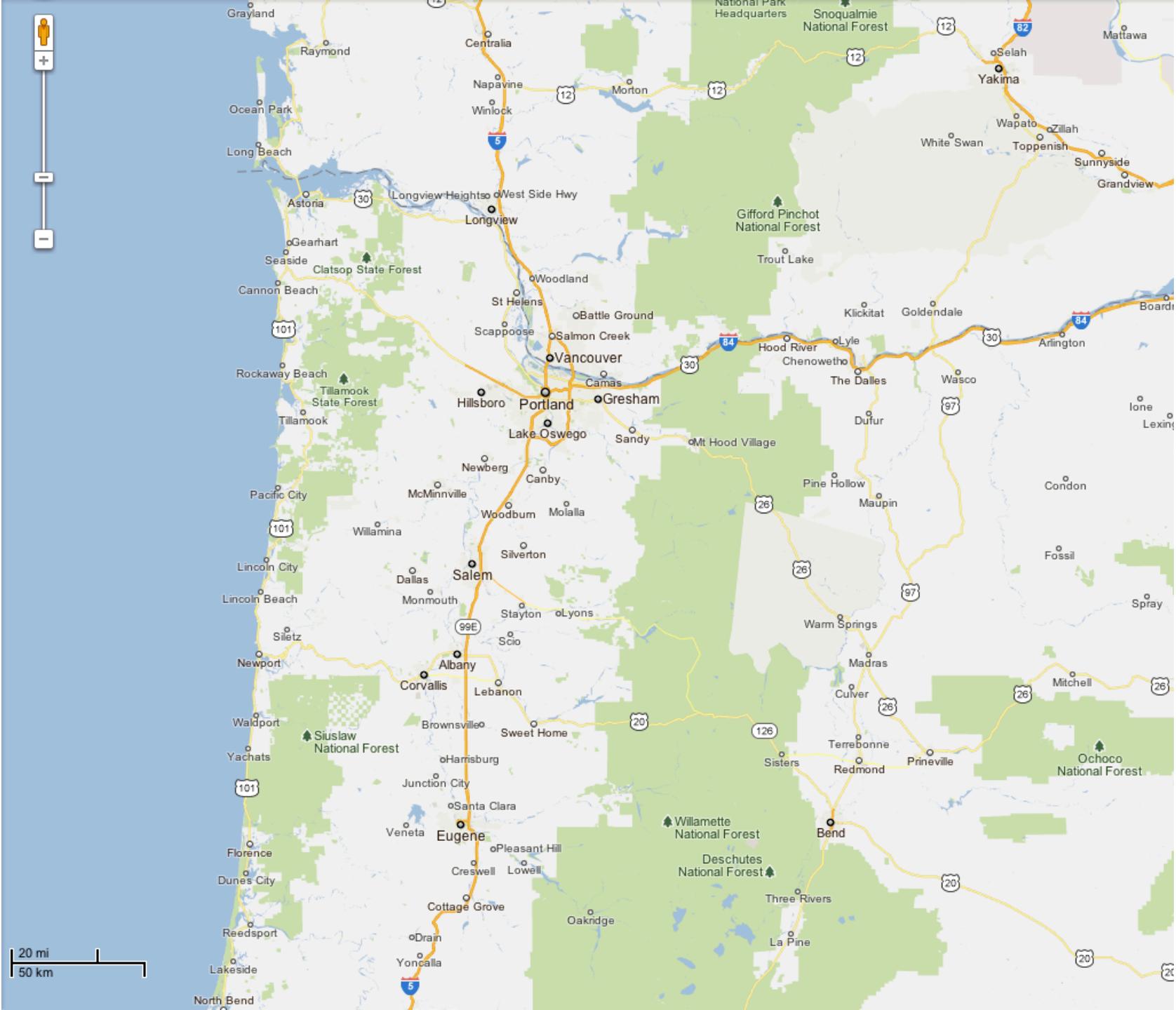
Choosing Colour for Data Visualisation

NOT IDEAL



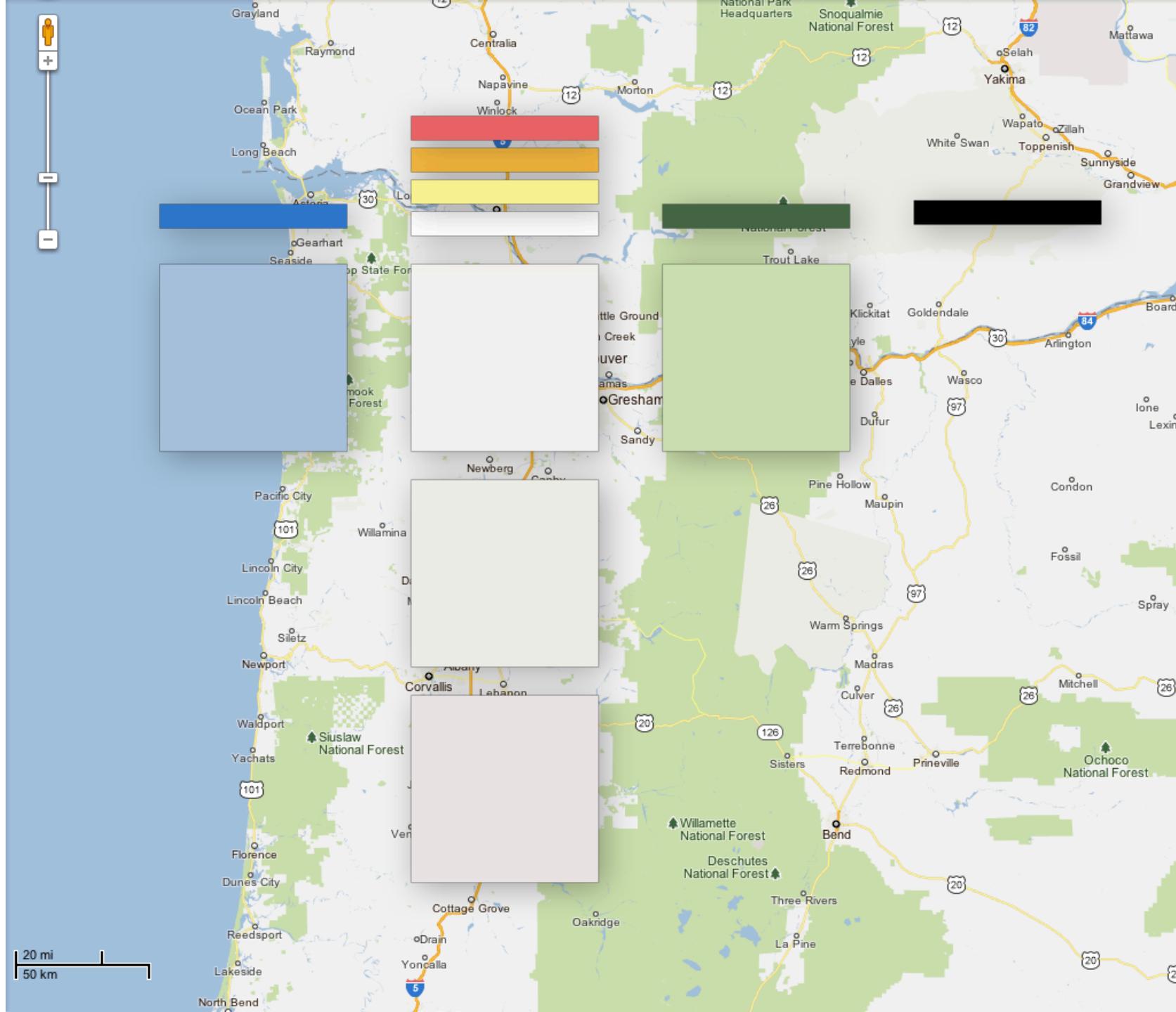
BETTER





20 mi
50 km

90



Gulf of Mexico

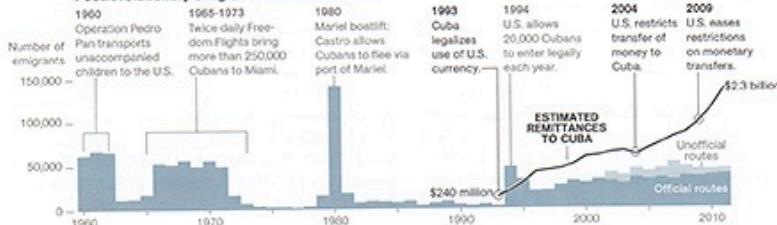


CUBANS OUT, DOLLARS IN

Among the most reliable sources of hard currency are the billions in remittances flowing in from the Cuban diaspora and the more than two billion dollars spent annually by tourists. The number of people leaving the island each year fluctuates with the twists and turns of politics and policy.

Cuban native and NGS cartographer Juan Valdés narrates an emigration timeline.

Postrevolutionary emigration and remittances



CAN CUBA STAY AFLOAT?

Hampered by the U.S. embargo and the collapse of the U.S.S.R. in 1991, Cuba's economy has floundered. Today it is precariously propped up by Venezuelan aid. Production of sugar—and most agricultural and industrial products—has plummeted since the 1959 revolution. Its best hope for growth: oil, nickel—and people. Tourism is up, along with cash sent to residents from abroad.



Cuban nationals abroad and tourists by country of origin



MARTIN GAMACHE, NMG STAFF; ALEXANDER STEGMAYER AND JUAN JOSE VALDÉS, NGS MAPS
SOURCES: CUBAN NATIONAL STATISTICS OFFICE AND U.S. CENSUS; U.S. CUSTOMS AND BORDER PROTECTION; UN STATISTICS DIVISION; DEVELOPMENT RESEARCH CENTRE ON MIGRATION; GLOBAL CITATION AND POVERTY; MIGRATION POLICY INSTITUTE; EMILIO MORALES AND JOE SCARPA/CIAA; CONSULTING GROUP; ARCH RETTER, CARLETON UNIVERSITY; JORGE R. PISON, UNIVERSITY OF TEXAS AT AUSTIN; MODIS LAND COVER, NASA LP DAAC

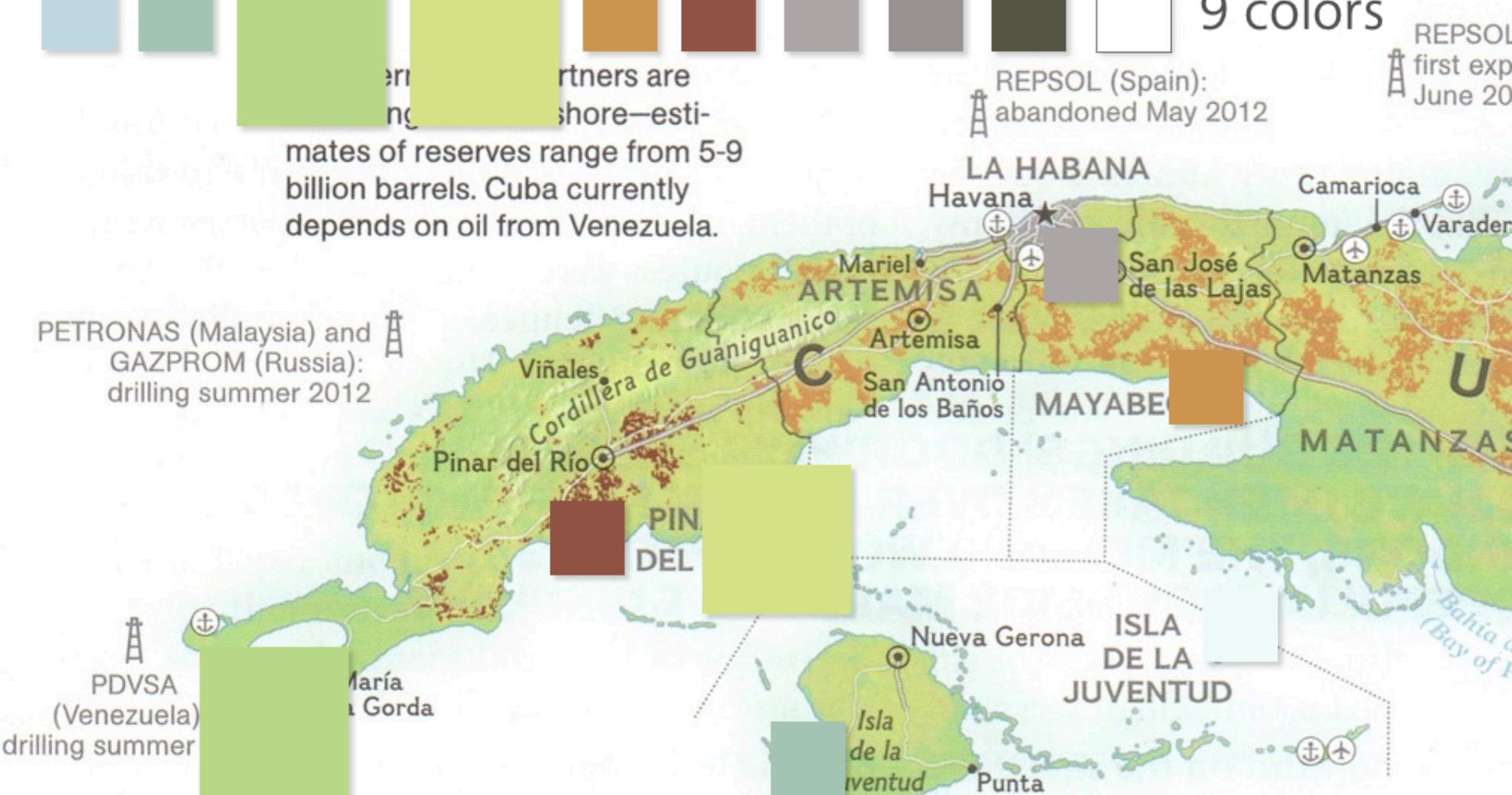


9 colors

er
ng
rners are
shore—esti-
mates of reserves range from 5-9
billion barrels. Cuba currently
depends on oil from Venezuela.

REPSOL
first exp.
June 2012

PETRONAS (Malaysia) and
GAZPROM (Russia):
drilling summer 2012



Exploratory oil
well



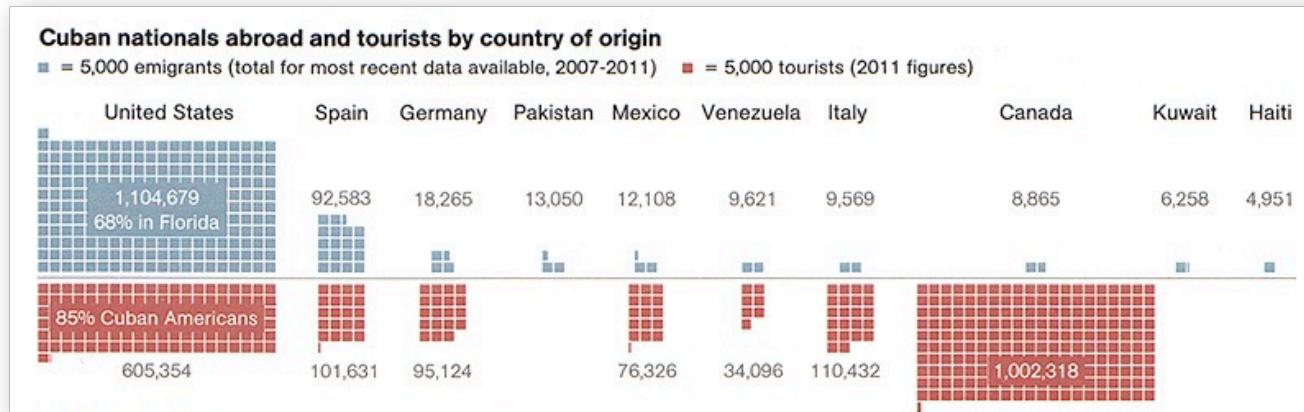
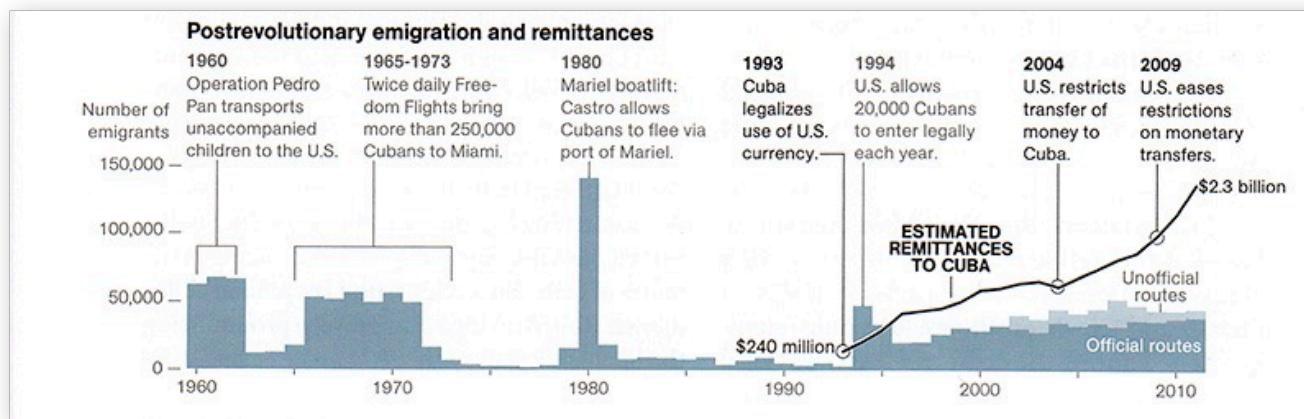
International ports of entry



Tobacco-
growing area

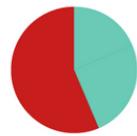


Caribean
Sea



Choosing Colour for Data Visualisation

NOT IDEAL



GOOD
BAD

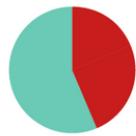


FOREST
LAKE



FEMALE
MALE

BETTER



GOOD
BAD



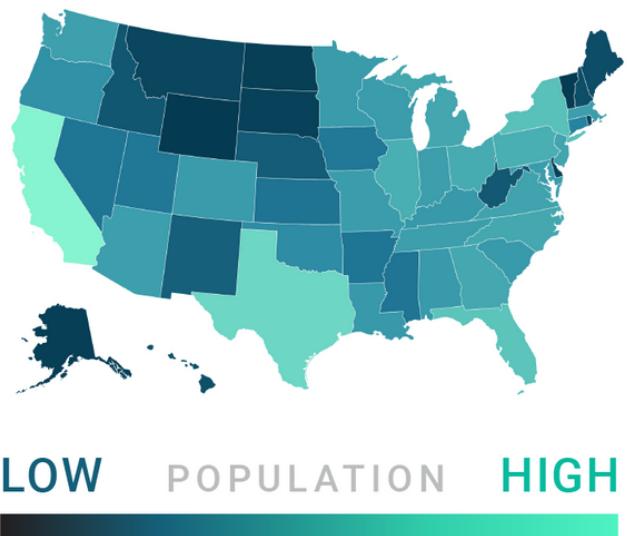
FOREST
LAKE



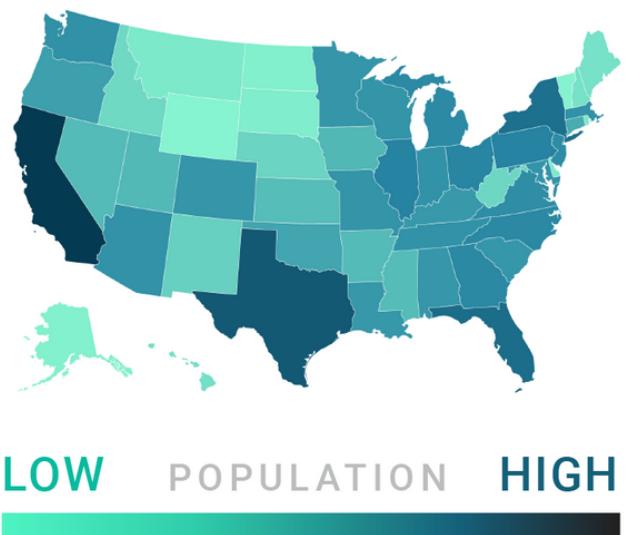
FEMALE
MALE

Choosing Colour for Data Visualisation

NOT IDEAL



BETTER

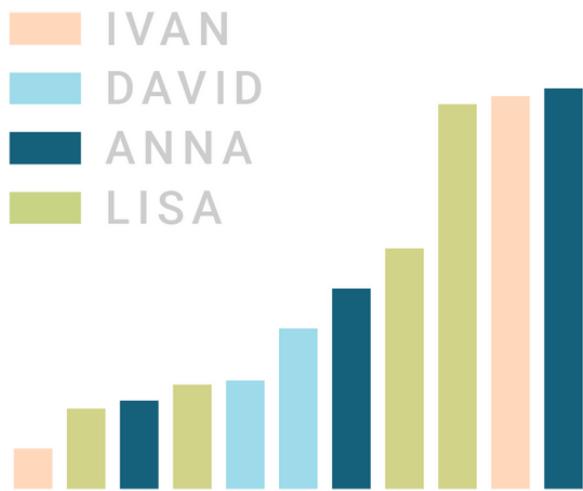


Choosing Colour for Data Visualisation

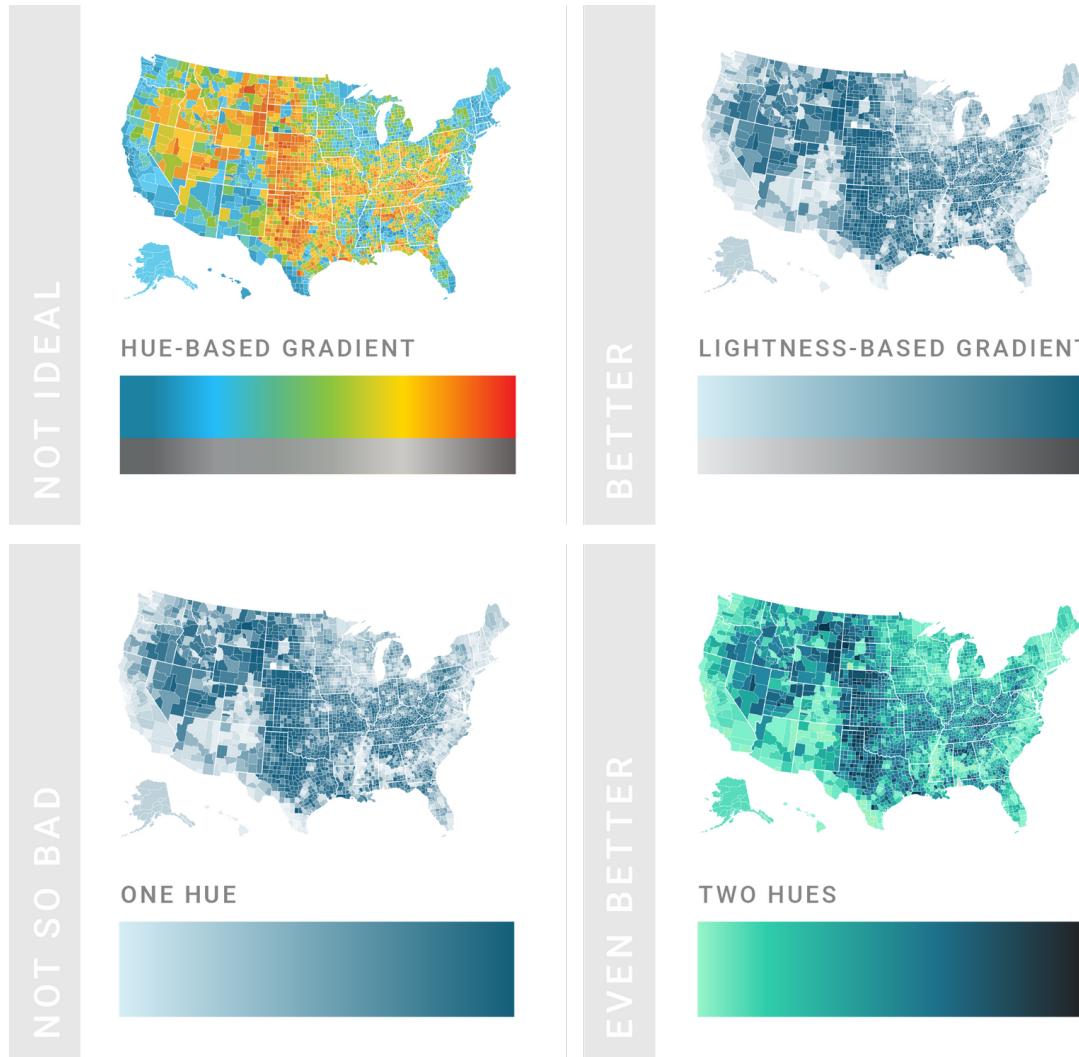
NOT IDEAL



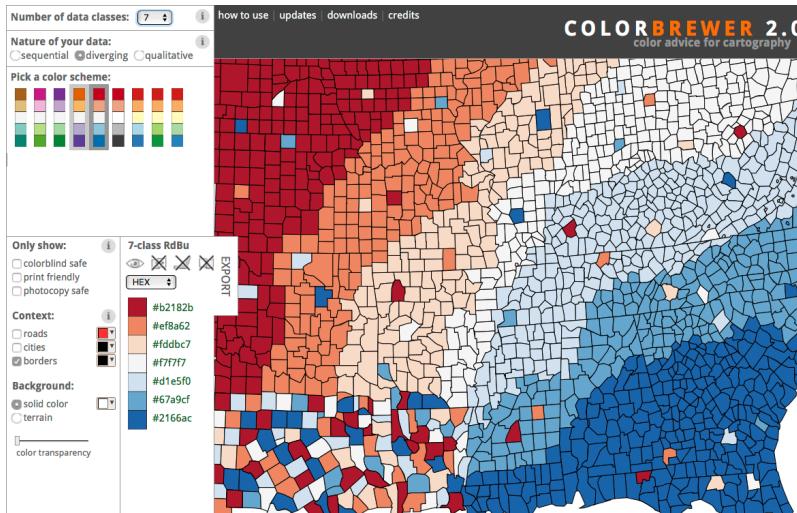
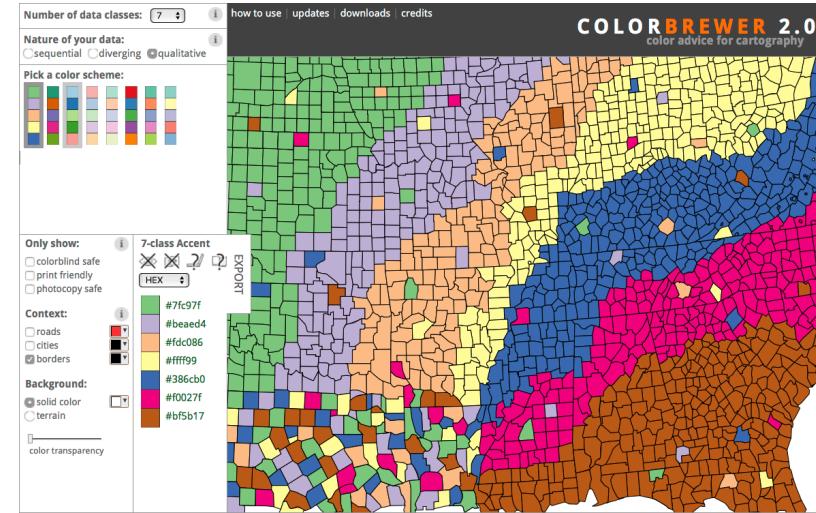
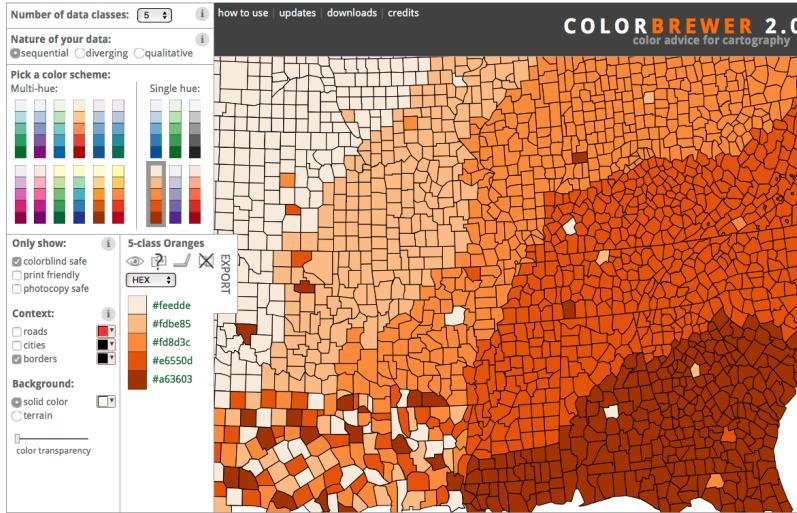
BETTER



Choosing Colour for Data Visualisation



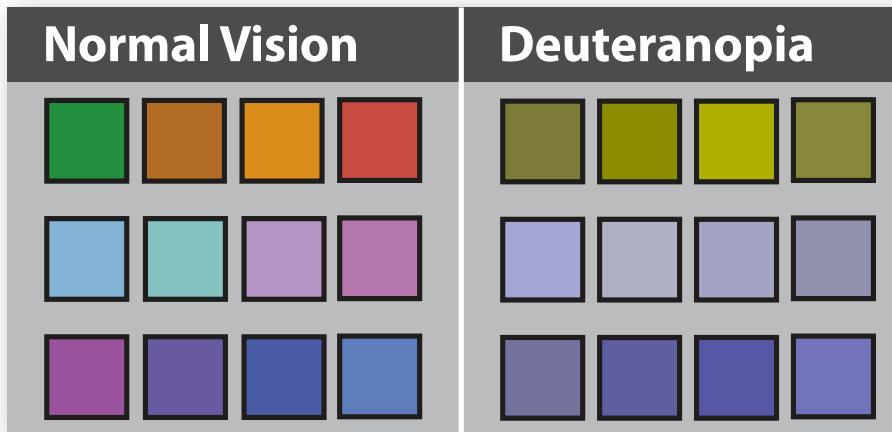
Sequential, diverging and qualitative colour schemes



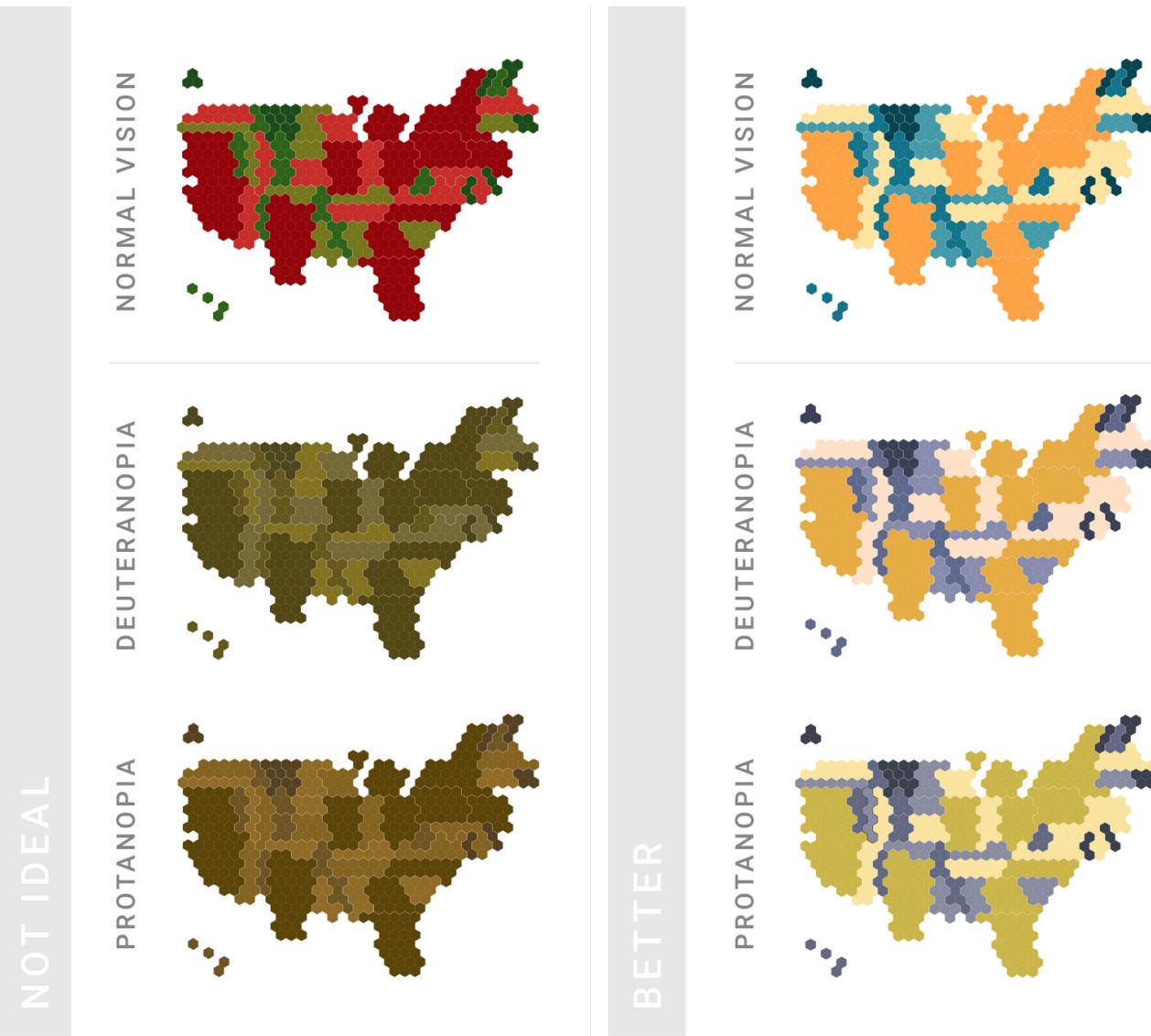
<http://colorbrewer2.org>

Colour for the colour vision impaired

- 8% of all men have a color vision deficiency
- They mainly confuse red and green (protanopia and deutanopia)



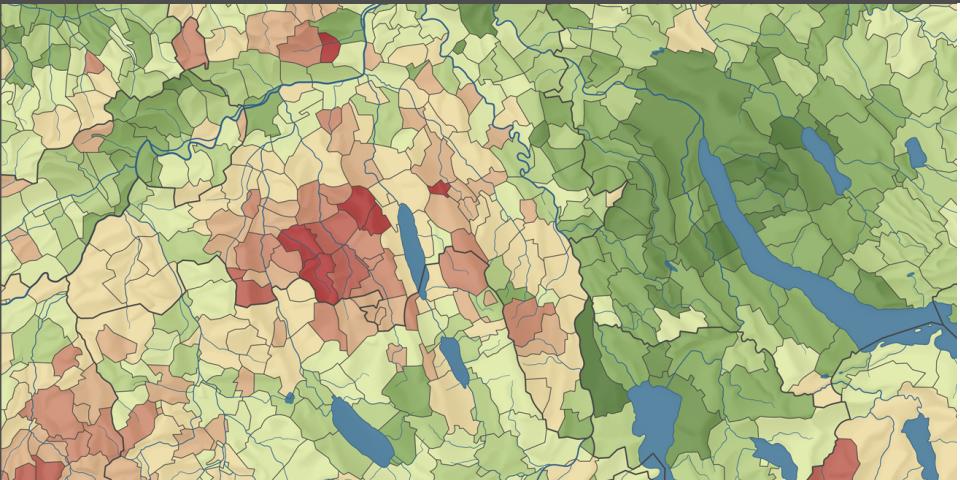
Choosing Colour for Data Visualisation



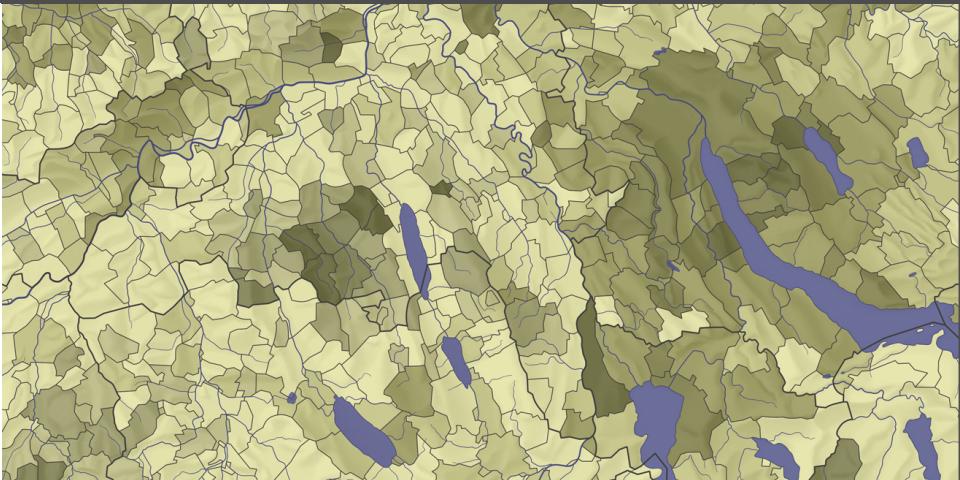
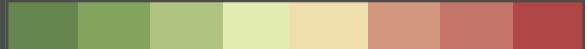
Source: Lisa Charlotte Rost, <https://blog.datawrapper.de/colors/>

Colour for the colour vision impaired

Diverging Red-Green Color Scheme

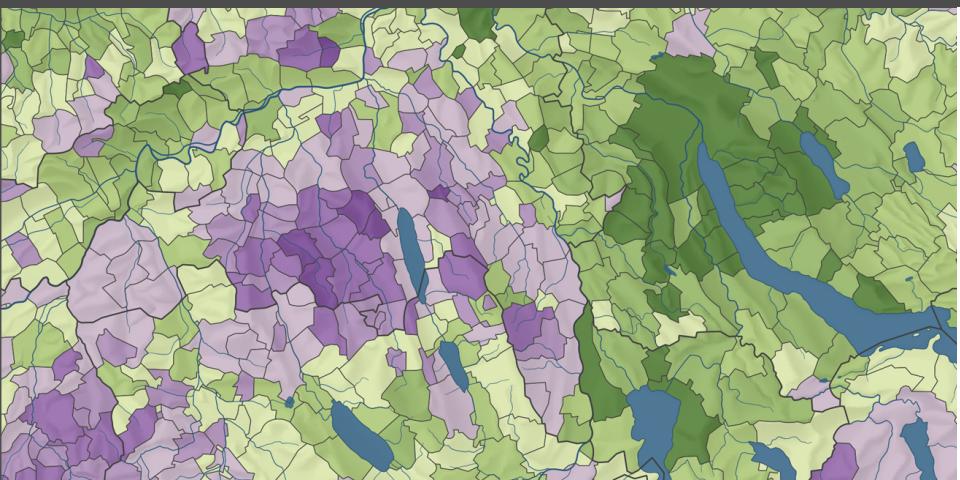


Normal Vision

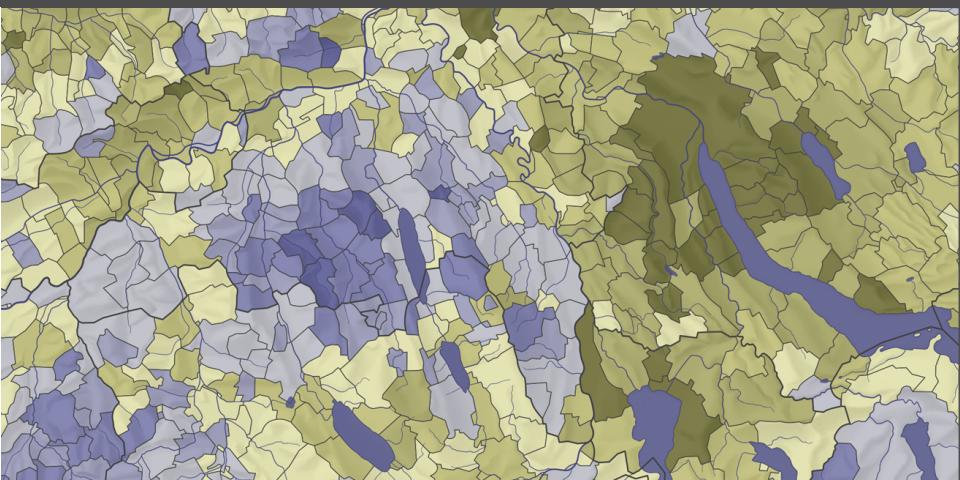
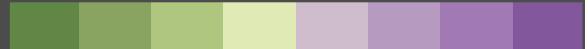


Deuteranopia

Diverging Purple-Green Color Scheme



Normal Vision



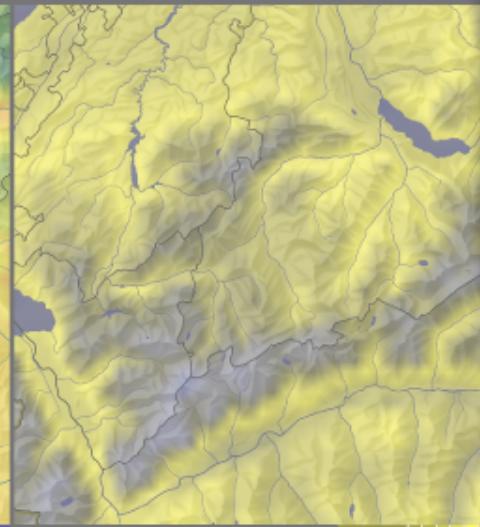
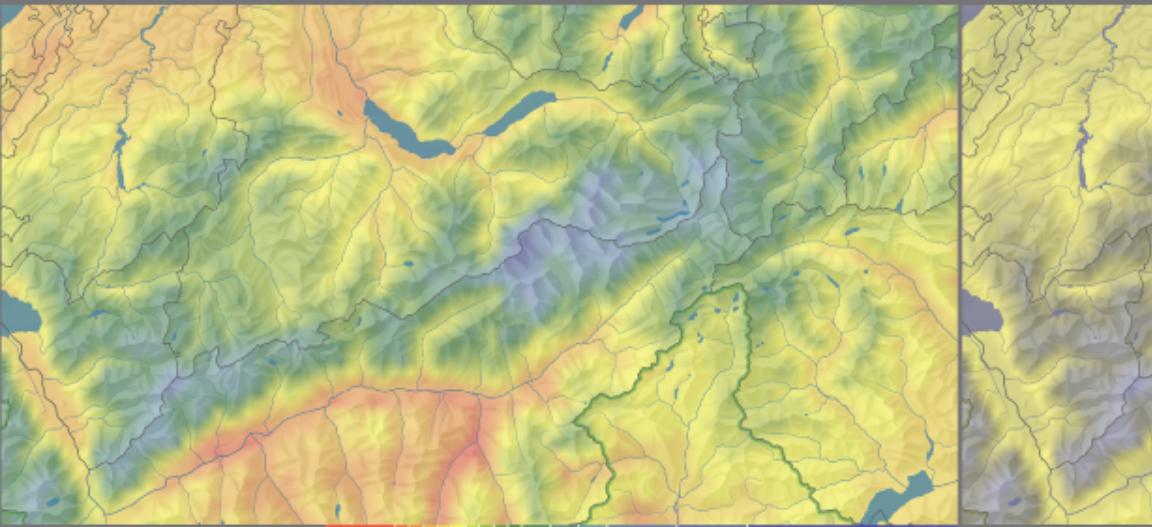
Deuteranopia



Normal Vision

- Deutanopia (Common) F5
- Protanopia (Rare) F6
- Tritanopia (Very Rare)
- [Save Filtered Screen Image...](#)
- [Preferences...](#)
- [About...](#)
- [Quit Color Oracle](#)

Spectral Color Scheme



- Required reading:
 - *A Quick Guide to Spotting Graphics That Lie*
<http://news.nationalgeographic.com/2015/06/150619-data-points-five-ways-to-lie-with-charts/>
 - *Gestalt principles of visual perception* by Cole Nussbaumer Knaflic from Storytelling with Data, p. 81–80 (on Moodle)

Presentations next Week!

Based on the Allocate+ tutorial class list, the students in the CLAYTON Tuesday 12PM–2PM lab class will be presenting at the start of the class this coming week.

All students post to the forum

Please read the instructions on the Weekly forums and post your viz. research and analysis there!

- Find a **visualisation that lies** or a **visualisation that makes interesting use of colour**. Discuss in a few words:
 - Why it is lying and what should be done to fix this,
 - or why colour is used in an interesting way,
 - What / Why / How,
 - data-ink ratio,
 - chartjunk,
 - applied gestalt principles,
 - the storytelling genre,
 - the macro and micro reading,