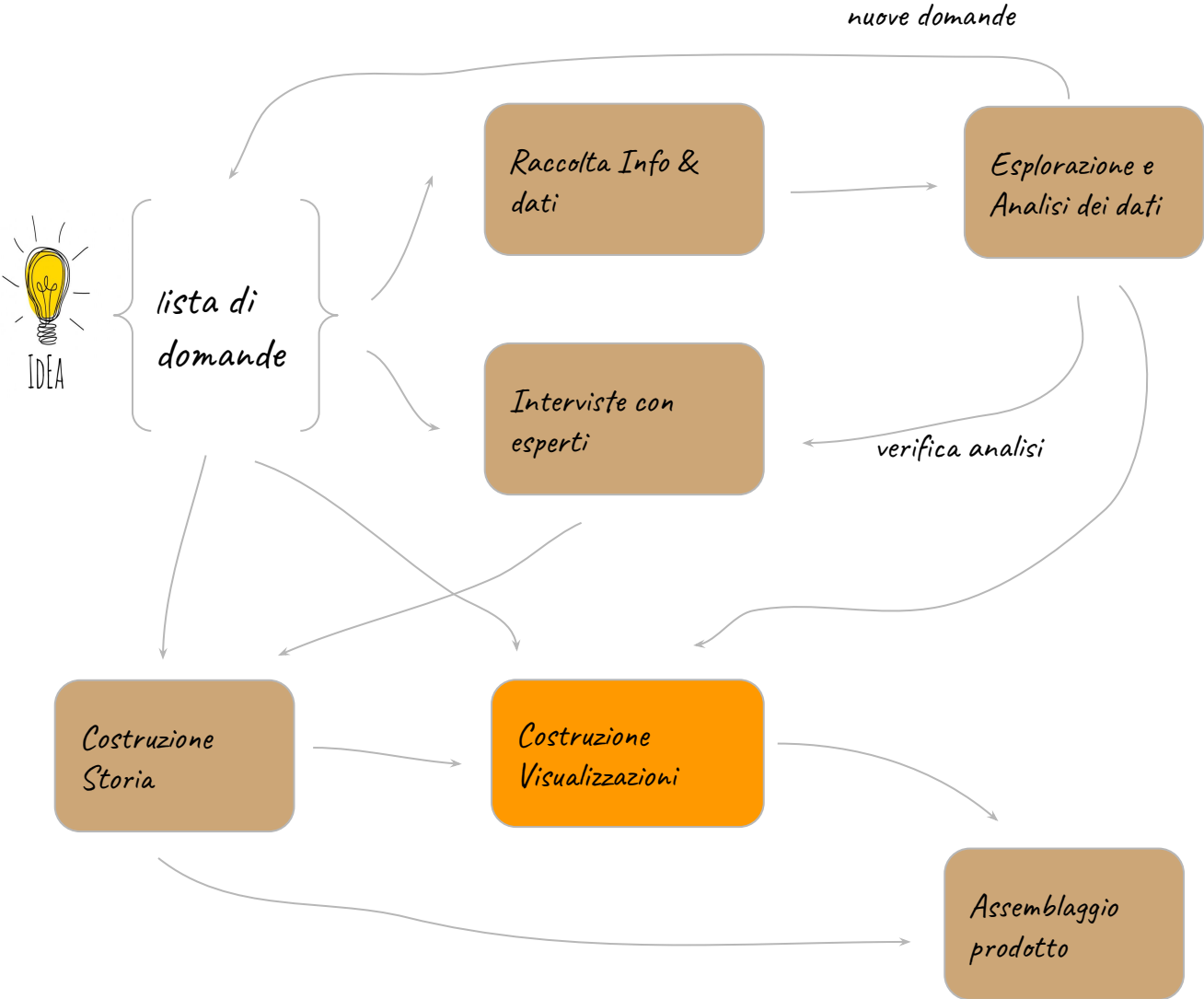




# Visual Languages to Telling Data Stories

# Data Journalism Workflow



# Overview

---

- The main principles of Data Visualization
- Attractive and Functional
- The Art of Insight

# Alberto Cairo

---

**The functional art: An Introduction to Information Graphics and Visualization** (Alberto Cairo - 2013)

**The Truthful Art: Data, Charts, and Maps for Communication** (Alberto Cairo - 2016)



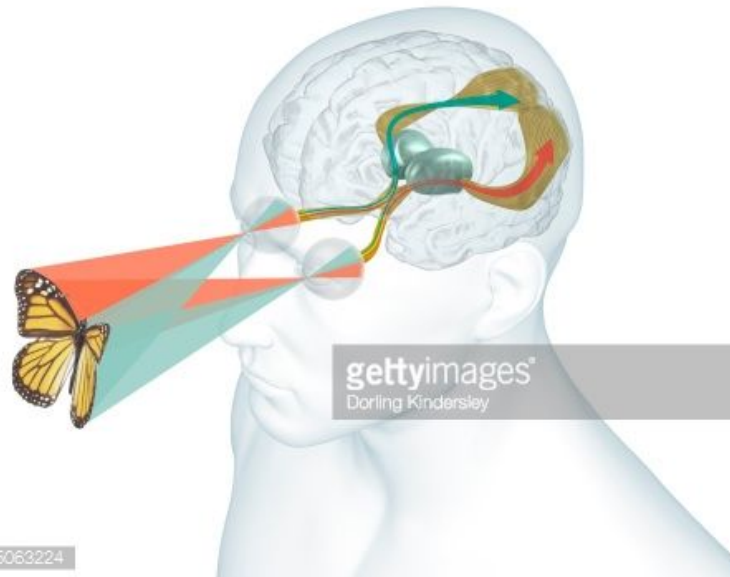
---

# **The Main Principles of Data Visualization**

# What visualisation is

---

- A graphical representation of abstract data (numbers, texts, geographic, time, ...)
- An instrument for analysis, communication, and understanding exploiting the **human vision system**



# Visualization System

---

Our **vision system** is faster at decoding **images** than **texts**.



# Visualization System

---

Exploiting the large band channel of HVS it's possible

- to gain large amount of information at once
- to extract **meaning** and see **patterns** and **trends** in a **data set**

**Find the  
hidden  
leopard**





# Visualization System

---

So publishing information with graphics makes it possible

- to gain information in a more **accessible**, simple and comprehensive way
- to obtain both a **wider view** and a more **detailed view** of the information
- to gain **insight**

# Visual encoding

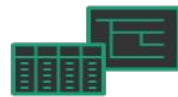


```
110100100101
011100110101
110110101010
110101011100
101010010101
000111000111
```

Raw data



*Data analysis*



Analytical abstraction



*Visual encoding*



Visualization



*Visual decoding*



Target user's insights

# 2012 Ukrainian Election -

---

An important Ukrainian Parliamentary Election took place on 28 October 2012

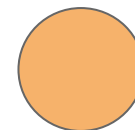
I guess none of you understand Ukrainian but ...

Party of Regions (pro Russia)



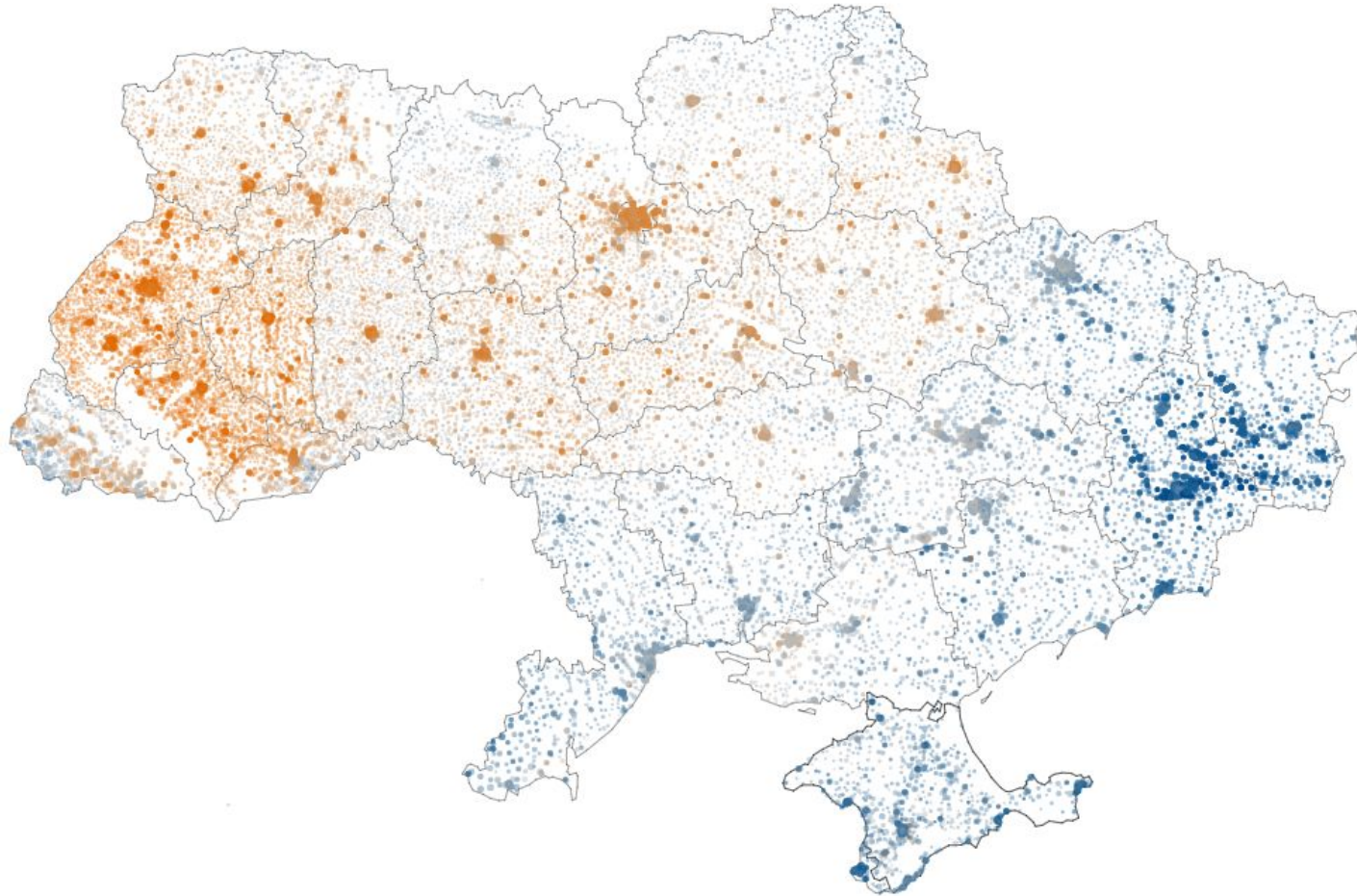
Vs

Fatherland Party (pro Europe)



# Map of 2012 Ukrainian Election Results

---

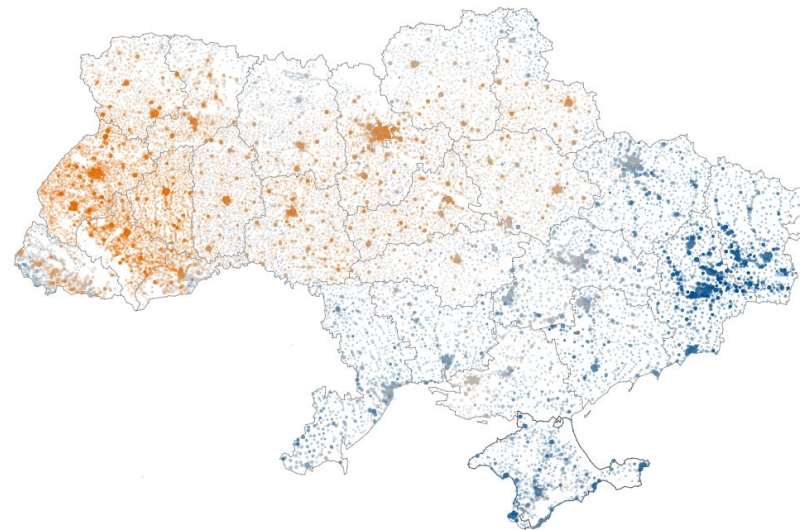


# What do you understand?

---

The Ukraine is completely divided:

- The Western part votes for the pro-**Europe** party.
- The Eastern part votes for the pro-**Russian** party.



**This pattern would not be easily visible just by looking at the dataset.**

# Misconception in visualising Data

---

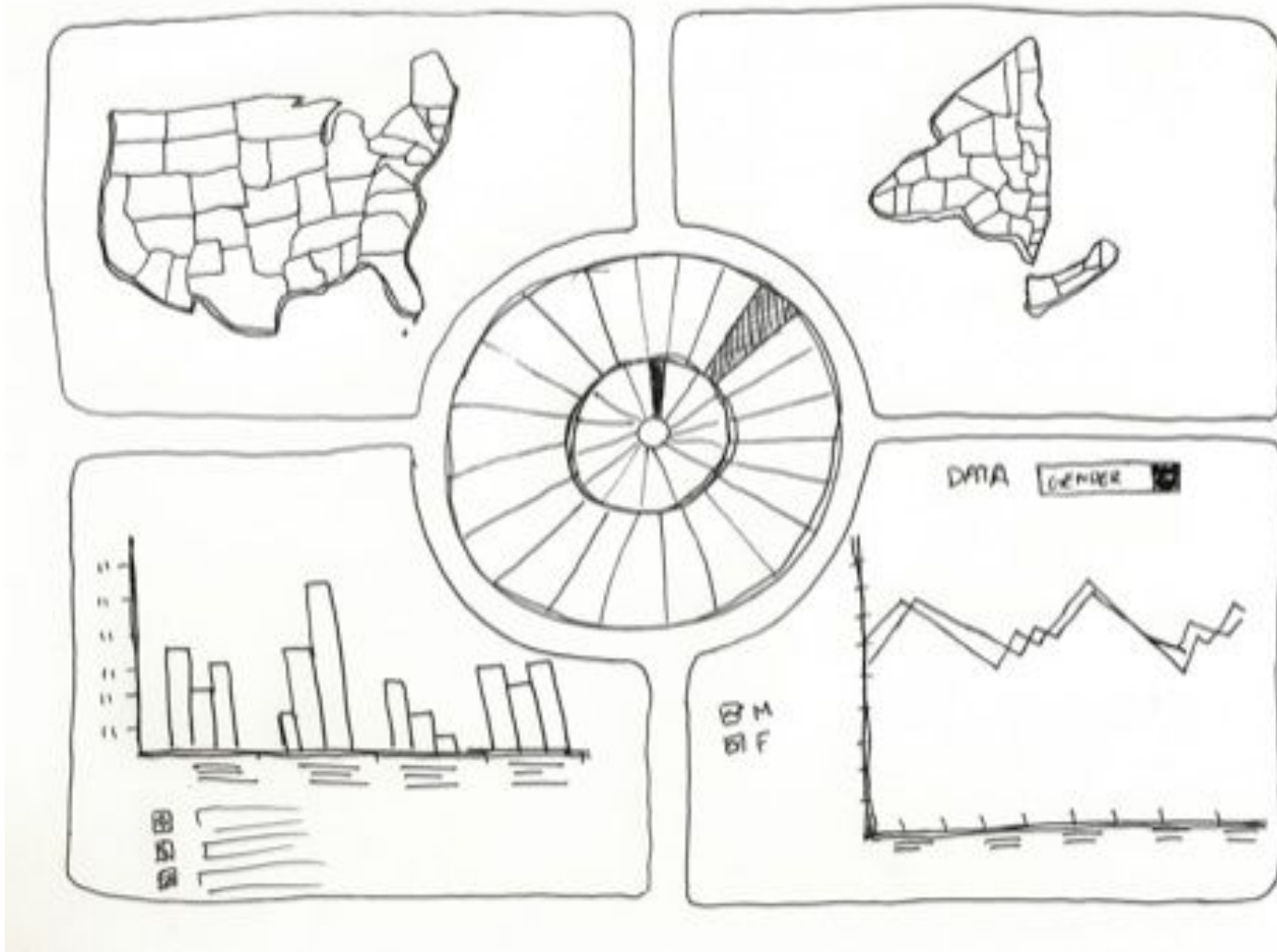
**‘Only great designers can create visualisations’** such as NyT Team, Washington Post, Boston Globe (naturally they are a big team, are the reference)

Making visualisations depends on thinking visually, not on a computer, so:

- Start with a **sketch of an idea** for visualising a story (don't start with a tool such as Tableau, Excel, Google Fusion, ...)
- If the visualisation does not work on paper, you are easily able to start a new one.

# Sketch out your idea

---



# Misconception in visualising Data

---

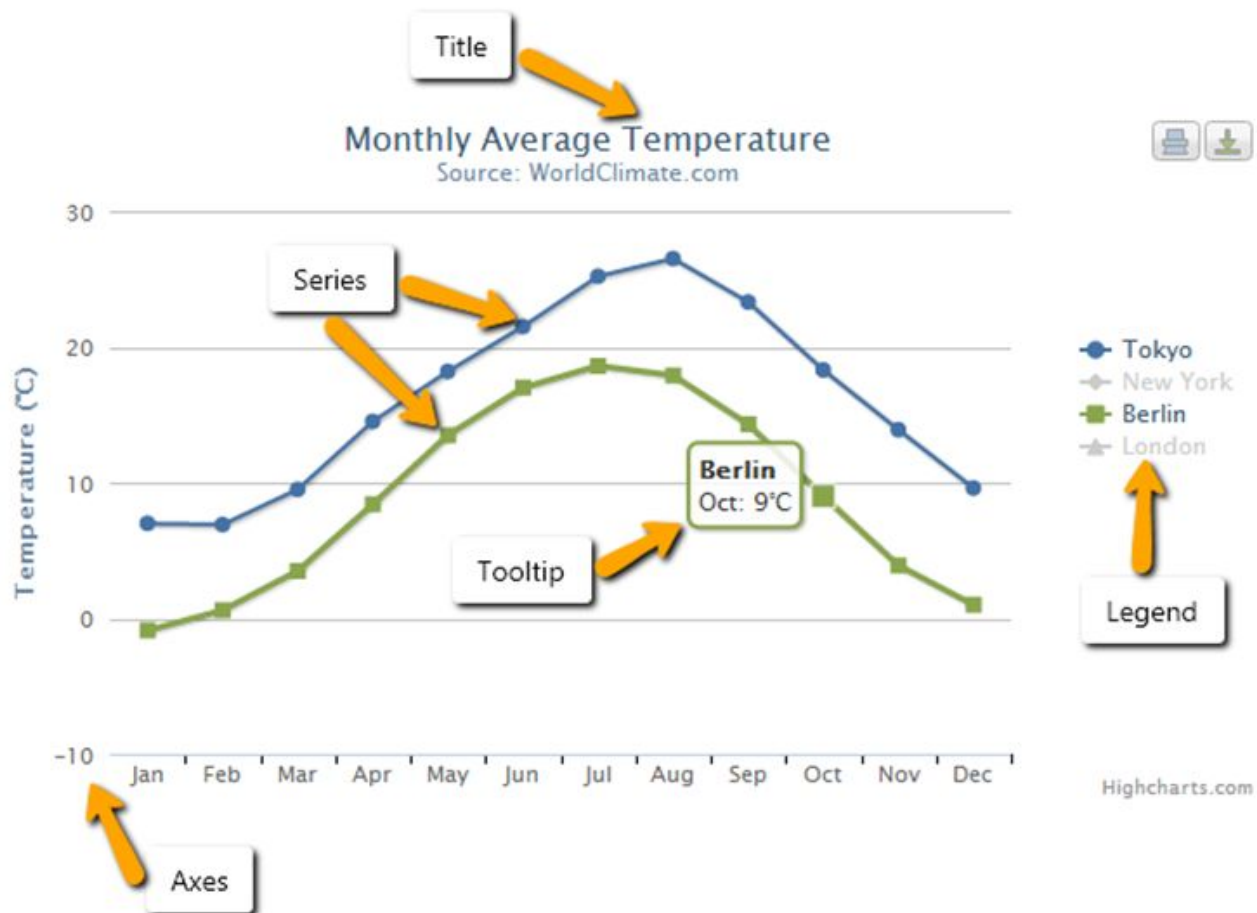
**‘Infographics is about designing cool pictures’**

- Creating something attractive will attract readers.
- **After attracting readers, you also need to deliver something useful**
- Give a context to understand

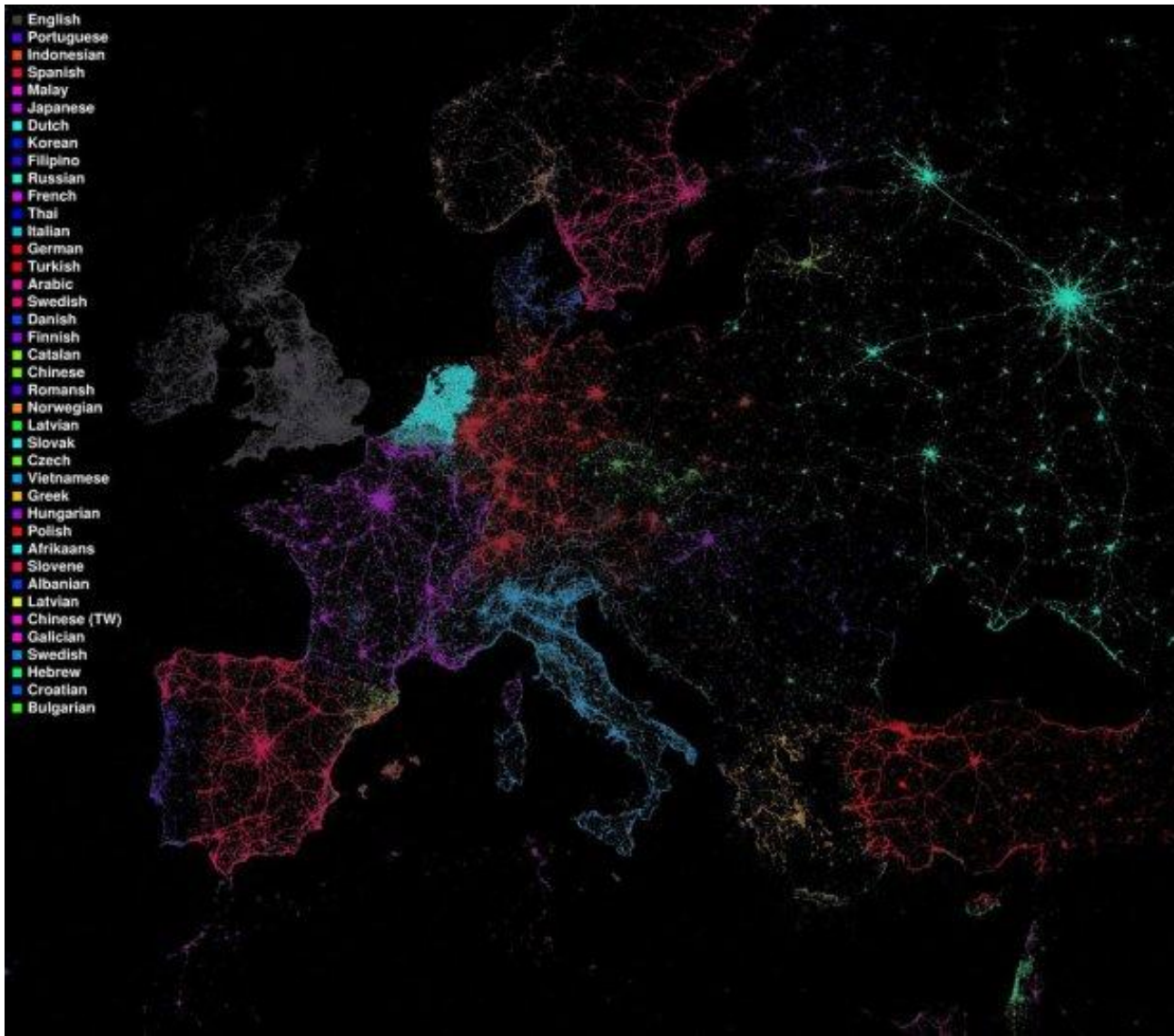


# Give the context

It's really frustrating don't understand



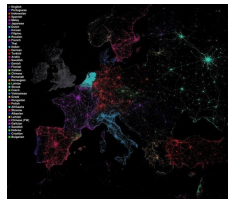
# Language communities on twitter



Language communities of Twitter – Eric Fischer – FlowingData:

# Language communities on twitter

---



When looking at a visualisation, try to read it and extract meaning from it.

- Most countries show that the main language used in their country is their own.
- It shows expected information, which is not insightful.

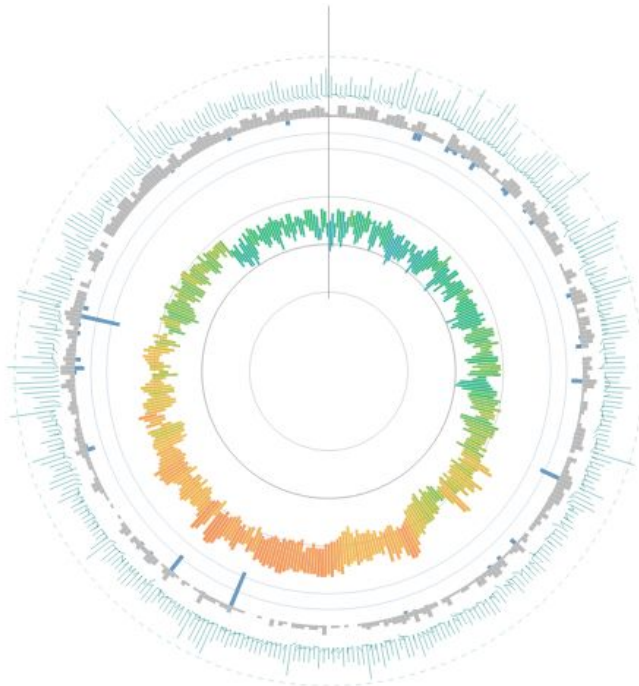
**Rather than the general trends, the exceptions are more useful.**

# Shneiderman's Mantra

---

How to understand and visualise large quantities of data?

“Overview first, zoom and filter, then details-on-demand”



The Weather  
Wheel  
by [WebVisLab](#)

# Summary of main principles of DV

---

DV creation doesn't require software

- Think visually
- Sketch first

DV can't be only attractive

- make readable with the context
- give overview then details

Provides insight into the data

- It's not just transforming data into visual forms
- The visual form has to reveal hidden facts

---

**A great visualization**

# The 4 features that define a great visualisation

---

1. **Functional** The shape of the graphic is adapted to the questions the visualisation should help answer.
2. **Beautiful/Attractive/Aesthetic level** If it is not attractive, readers won't stop to read and interact with it. Attention-Grabbing is 50% of the work.
3. **Insightful** Revealing hidden facts that are unexpected or surprising or extremely important.
4. **Enlightening** The information that the visualisation reveals changes the perception of the reader.

---

# Attractive & Functional

choosing the best graphic form



# Try to objectively guess the best graphic form

---

1. Think about the AUDIENCE and the publication that you are working for
  - a. teenagers or serious people
2. Think of the questions your graphic should help answer
  - a. ~~I love maps so I will use maps forever~~
3. Can you understand the graphic without reading every single number? *The goal of visualising data is to provide rapid access to trends and patterns in the data without having to read the data.*

# Data Vis. Classification

---

[Read the data visualization catalogue](#)

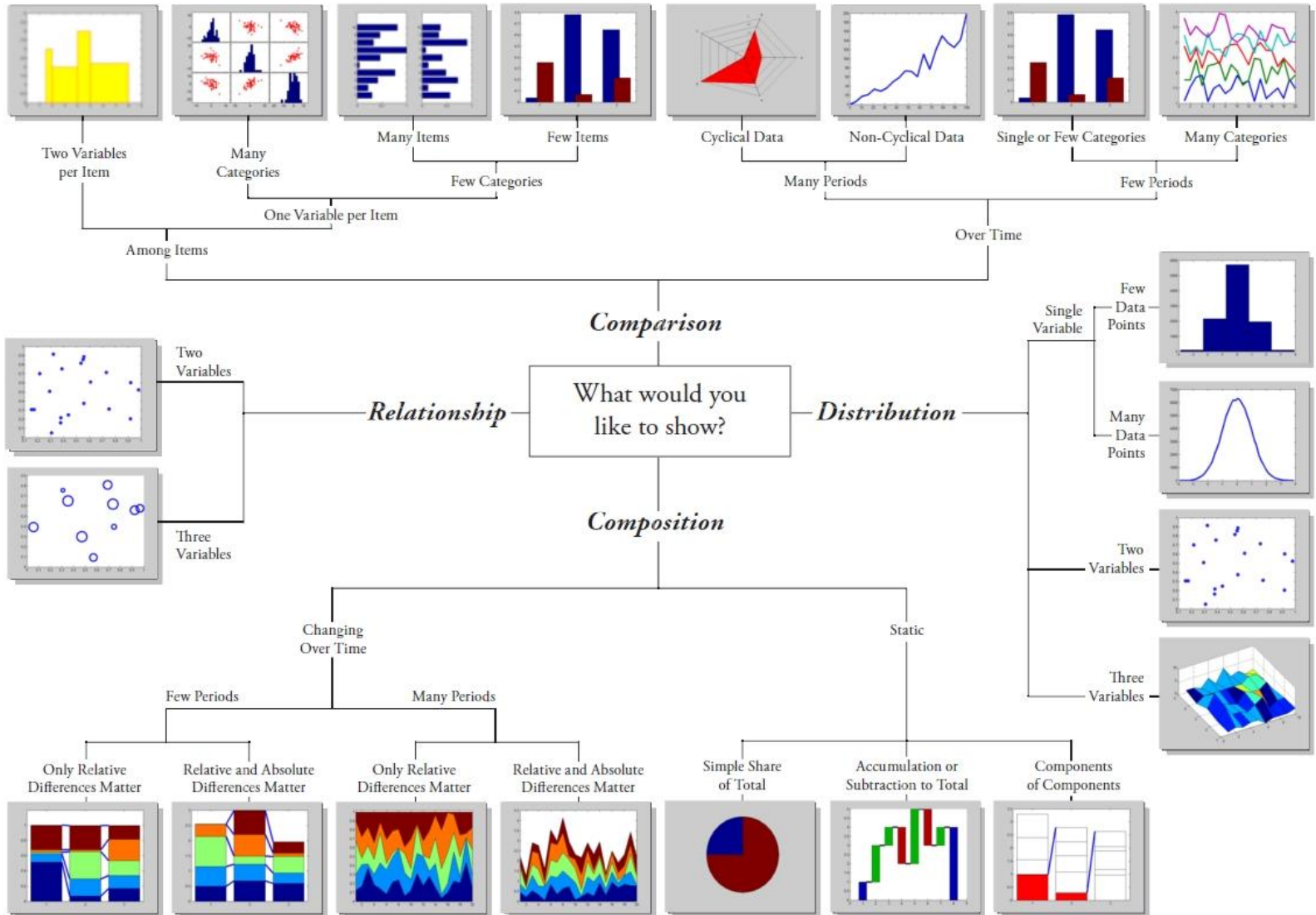
What is the question?

- comparison
- relationship
- distribution
- composition

Then select the chart

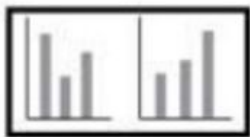


# Chart Suggestions—A Thought-Starter



Permette valutazioni più accurate

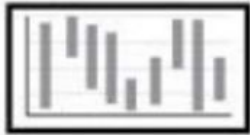
Permette valutazioni meno accurate



Posizione lungo una scala comune



Posizione lungo scale non allineate



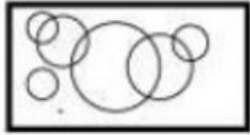
Lunghezza



Direzione



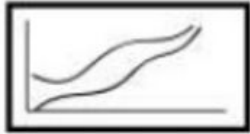
Angolazione



Area



Volume



Curvatura



Tonalità



Intensità del colore

Allows more accurate comparisons

Allows more generic comparisons

2D position along common, aligned scale



2D position along common, but unaligned scales



Length



Slope



Angle



Area



Colour intensity



Volume

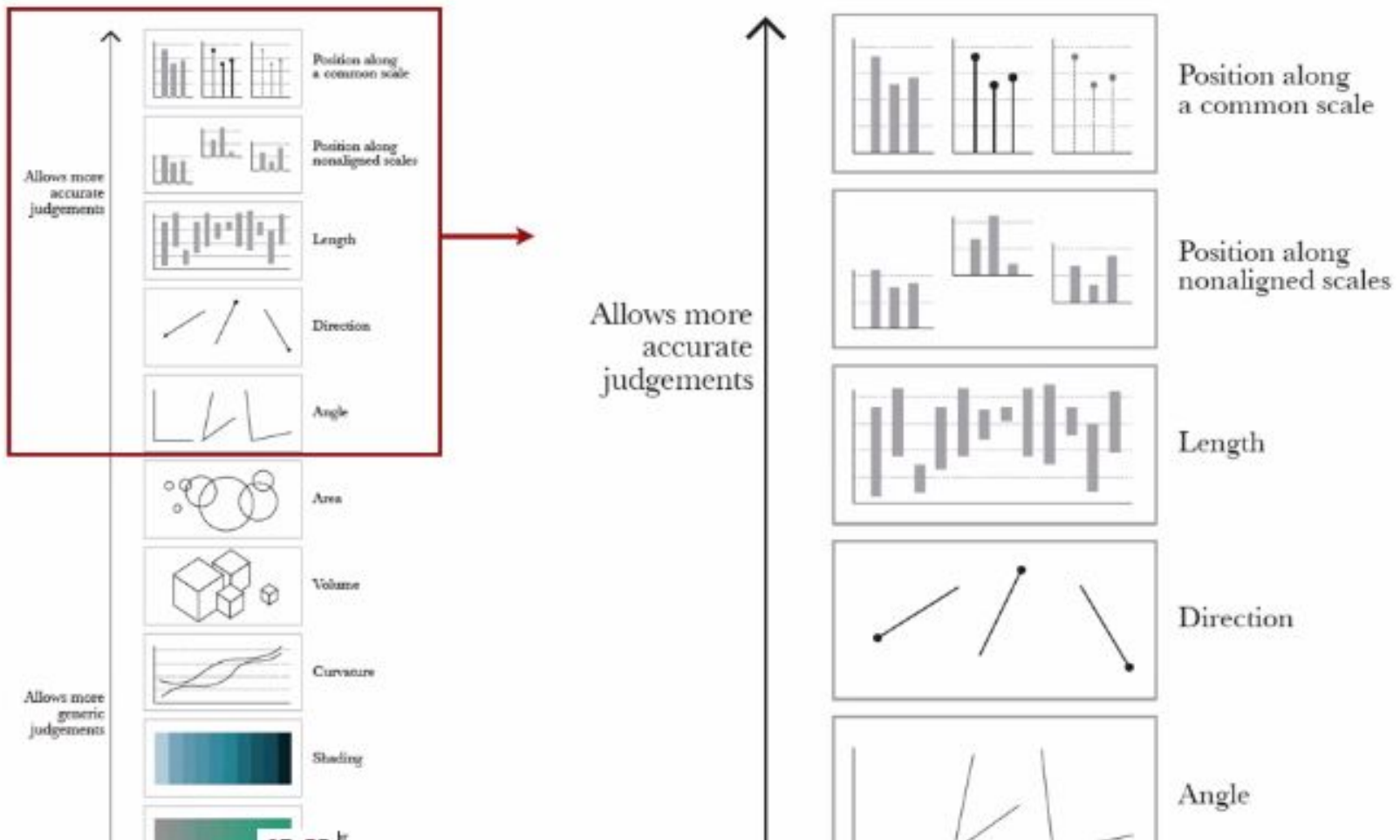


Colour hue



# Target: accurate comparison between some numbers

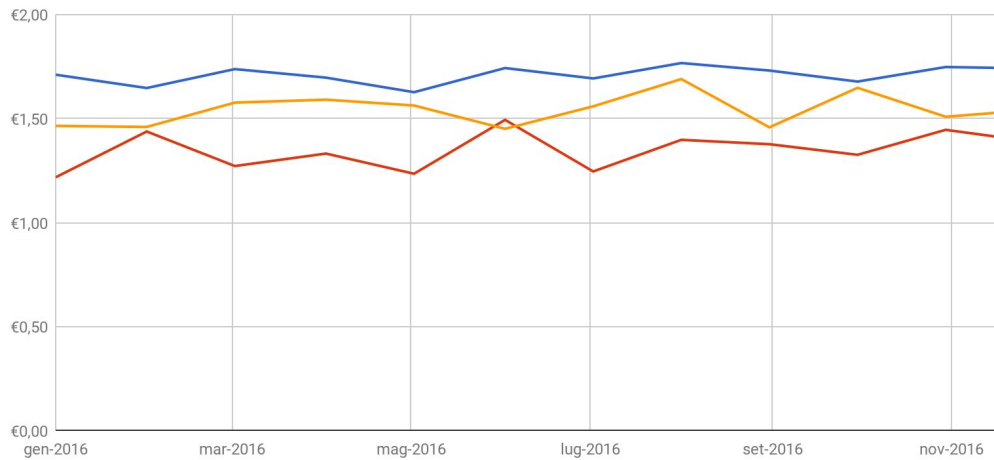
William Cleveland and Robert McGill



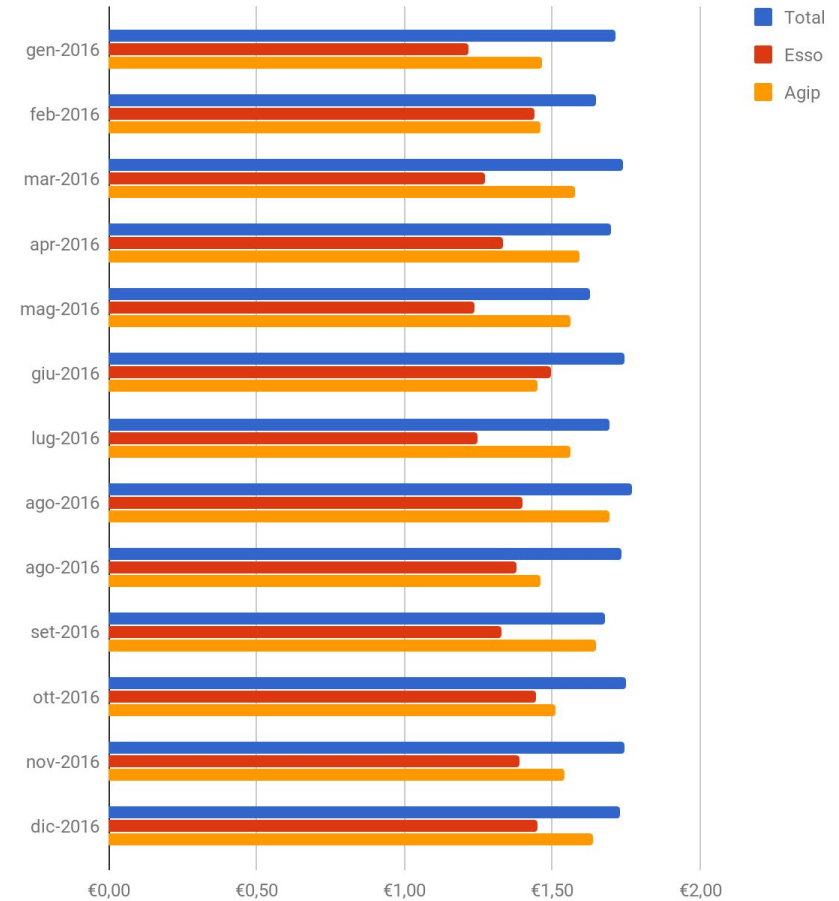
# Sample: accurate comparison

random numbers (google spreadsheet)

Prezzo medio benzina

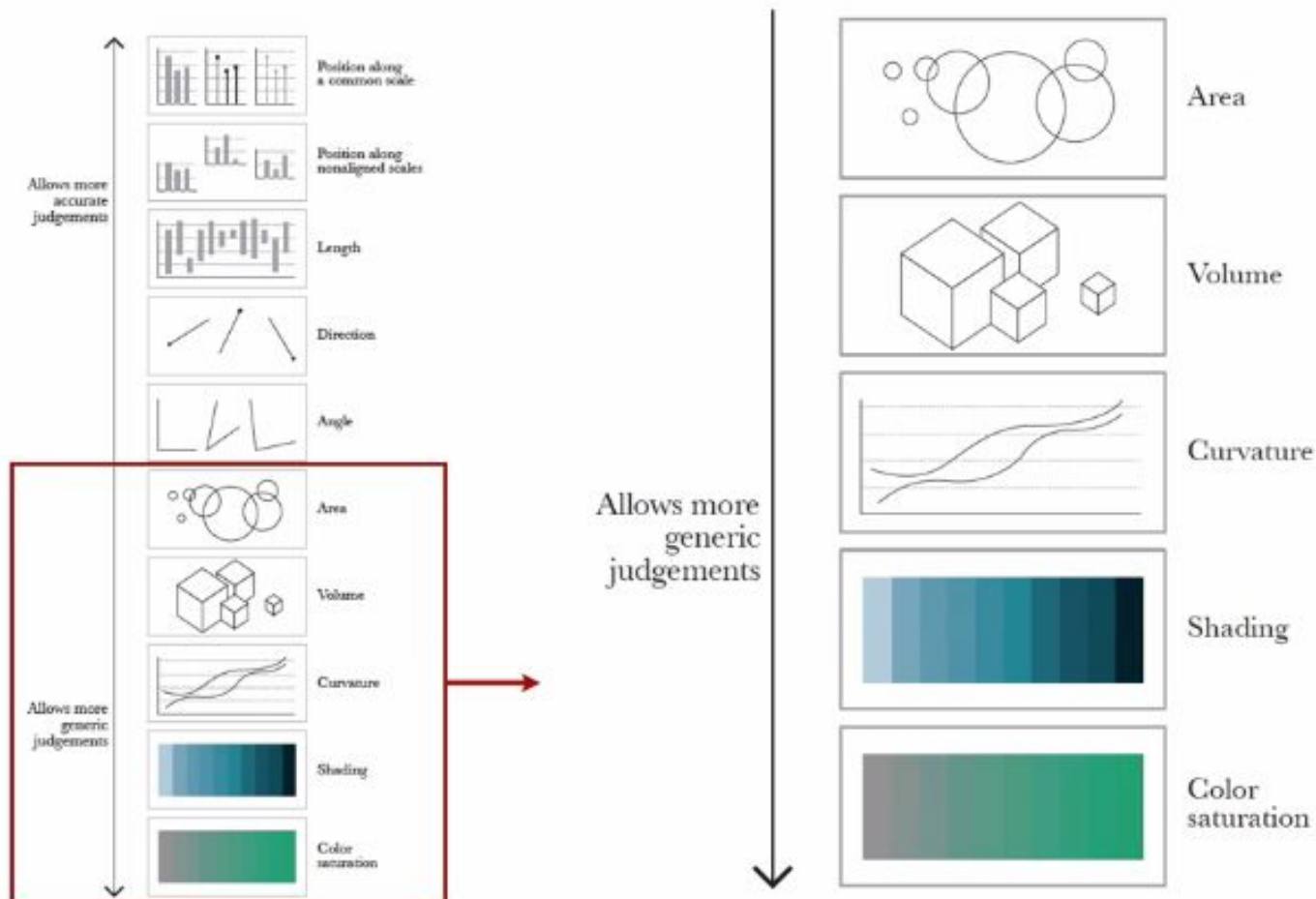


Prezzo medio benzina



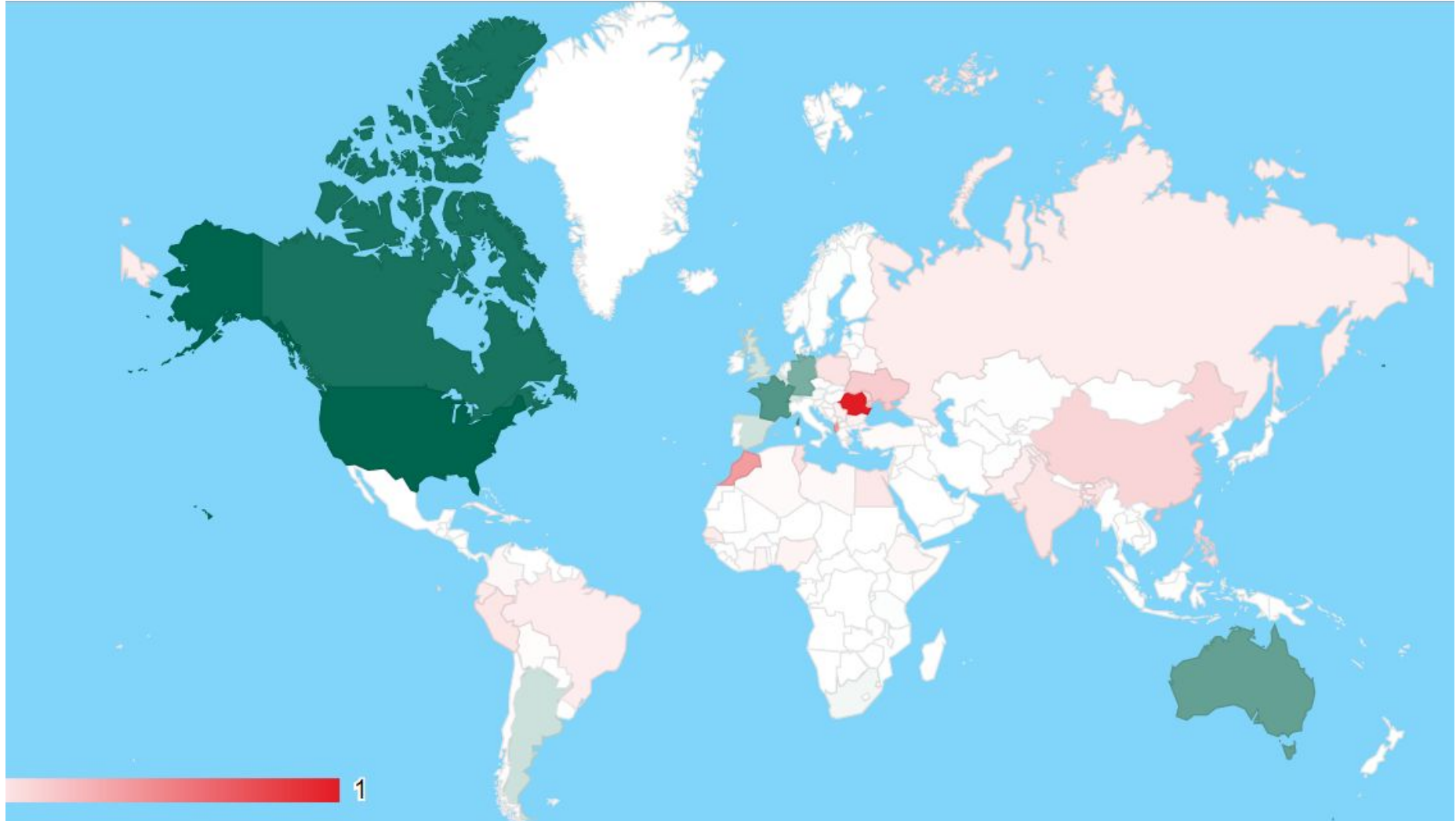
# Target: overview of trend

William Cleveland and Robert McGill



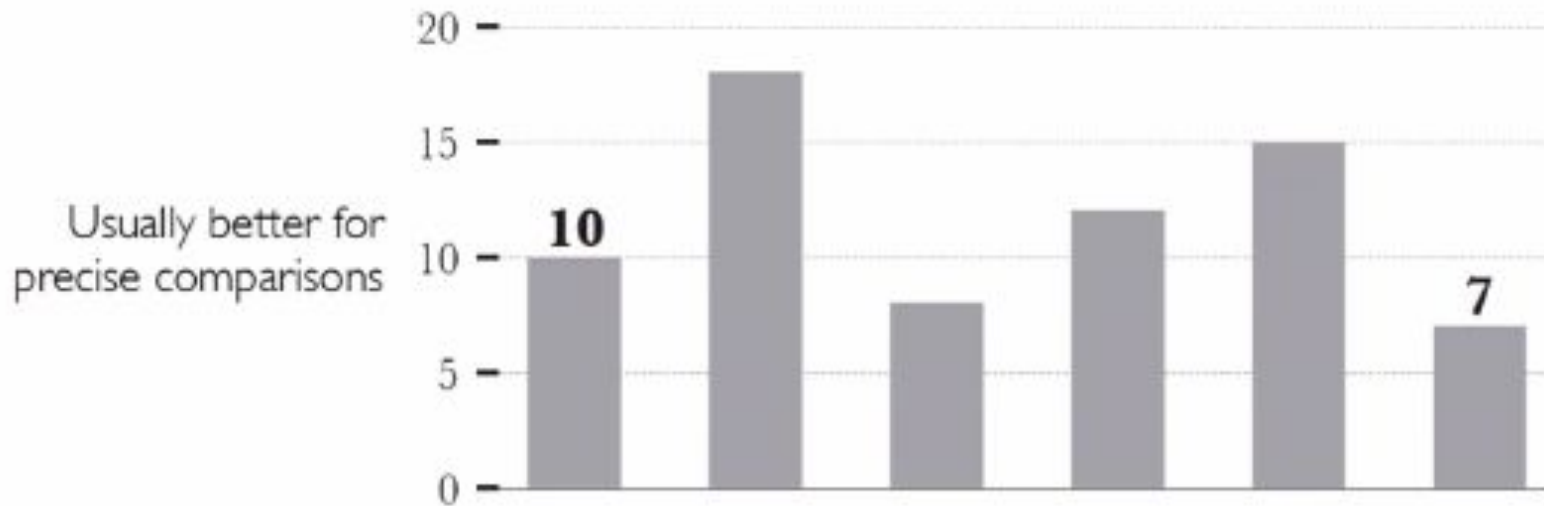
# Sample: overview of italian migration

---





# Accurate comparison with numbers



It may work for overviews of large datasets (choropleth maps, for instance)



# Analyse graphic published

---

After making the graphic attractive, you need to make it functional

- What are you supposed to do with the graphic?
- What questions are supposed to be answered?

Example - **Banking on the World Bank** – extract from GOOD Magazine - **The Donating countries**

# Top Donating Countries



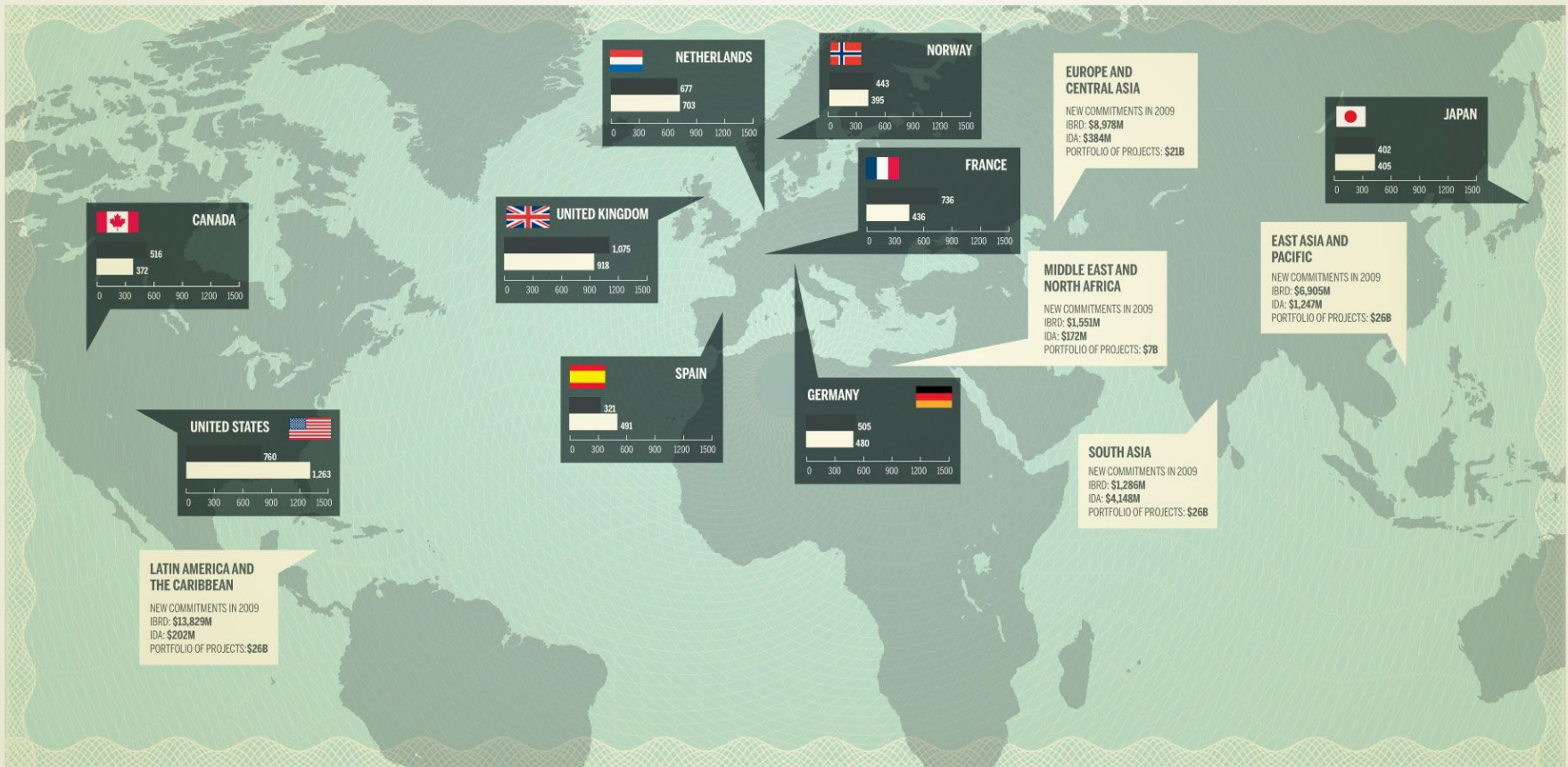
## BANKING ON THE WORLD BANK

Developing countries around the world depend on assistance from the World Bank to help support their often fragile economies. We have highlighted the regional recipients and respective outlays, as well as the nine largest donor countries contributing to the Fund's efforts to fight poverty around the globe.

IBRD: THE INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT  
IDA: THE INTERNATIONAL DEVELOPMENT ASSOCIATION



SOURCE: WORLD BANK.ORG



# Comments

---

User question: **What is the ranking of countries by aid given? I want to see who gave more money**

If you want to know the answer you need to memorize the numbers and put in a rank.

User question: **Who has been the most generous?**

- This overview is incomplete, a variable is missing to make the comparison between the countries complete.
- You need to provide the amount given per capita per country.

The eventual goal of the graphic seems to be to locate the different countries. Is it important?

# Rule

---

**A designer needs to anticipate the readers' questions and facilitate what their needs are.**

# Improve an existing graphic

---

1. Extract the numbers from the graphic
2. Think of the questions do you want to answer
3. Redesign the graphic
4. Make more than one visualisation for your dataset

# Extract the data

---

You can use a spreadsheet

- Excel spreadsheet
- Google Fusion Tables
- OpenOffice Calc
- Google spreadsheet

# Think about the question

---

- Add the population of the countries to get relative values for comparison.
- Divide the total amount of money by the population to get the amount per capita per country.
- Compare the money given to GDP (Gross domestic product), to get a percentage of the total GDP given.
- In which country did the amount of money given change the most between 2008 and 2009?
- Calculate the difference in percentage in 2008 and 2009.



# Redesign the graphic

---

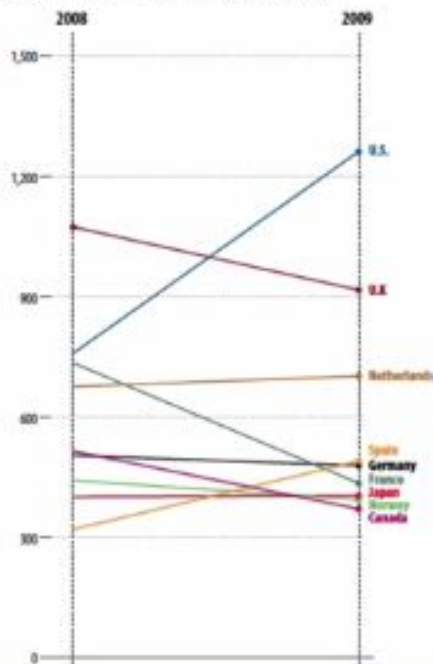
- The original graphic is attractive but needs to be made functional:
- Make a sketch of how to redesign the graphic.
  - The first slope graph shows the amount given in absolute terms.
    - Slope graphs are used to represent change over time.
  - The second slope graph shows the amount in relative terms.
- The original graphic can be used if it has a purpose.
  - A flow map of where the money given goes to.

## THE 9 LARGEST CONTRIBUTORS

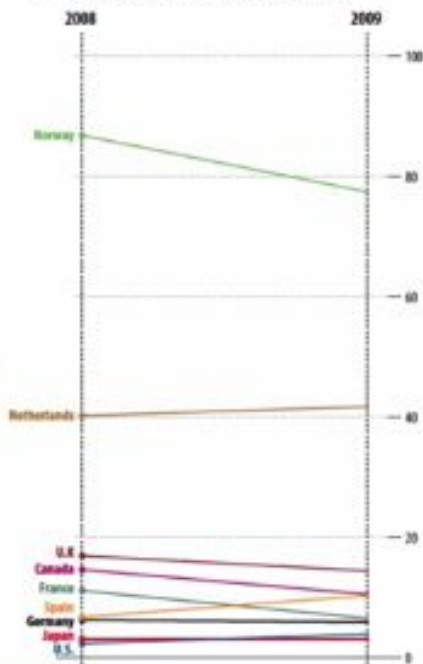
Write a nice introduction here, perhaps just 1-line long will be enough

### THE U.S. LEADS THE WORLD IN ABSOLUTE TERMS... ...BUT NOT IN RELATIVE TERMS

Total annual contributions in millions of U.S. dollars



Annual per capita contributions, in U.S. dollars

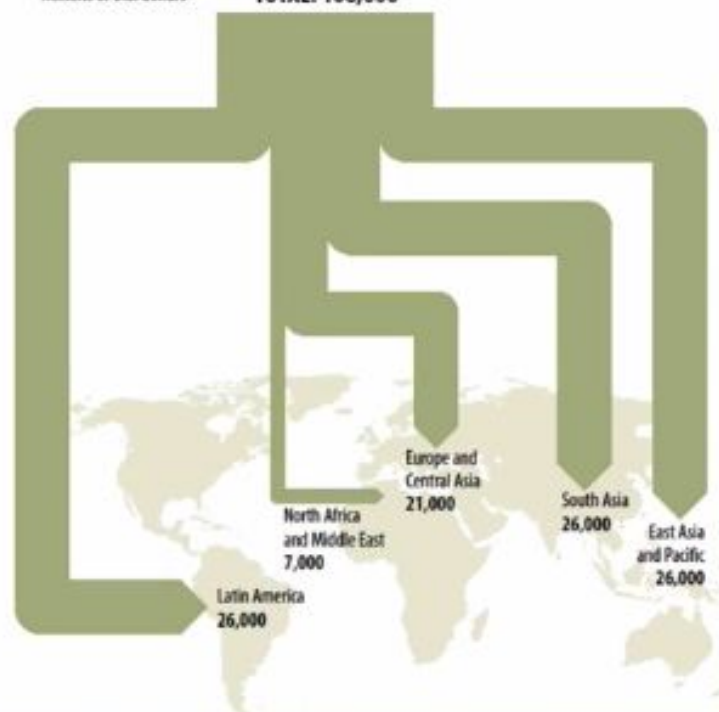


## WHERE THE MONEY GOES

Write a nice introduction here, perhaps just 1-line long will be enough

Millions of U.S. dollars

TOTAL: 106,000



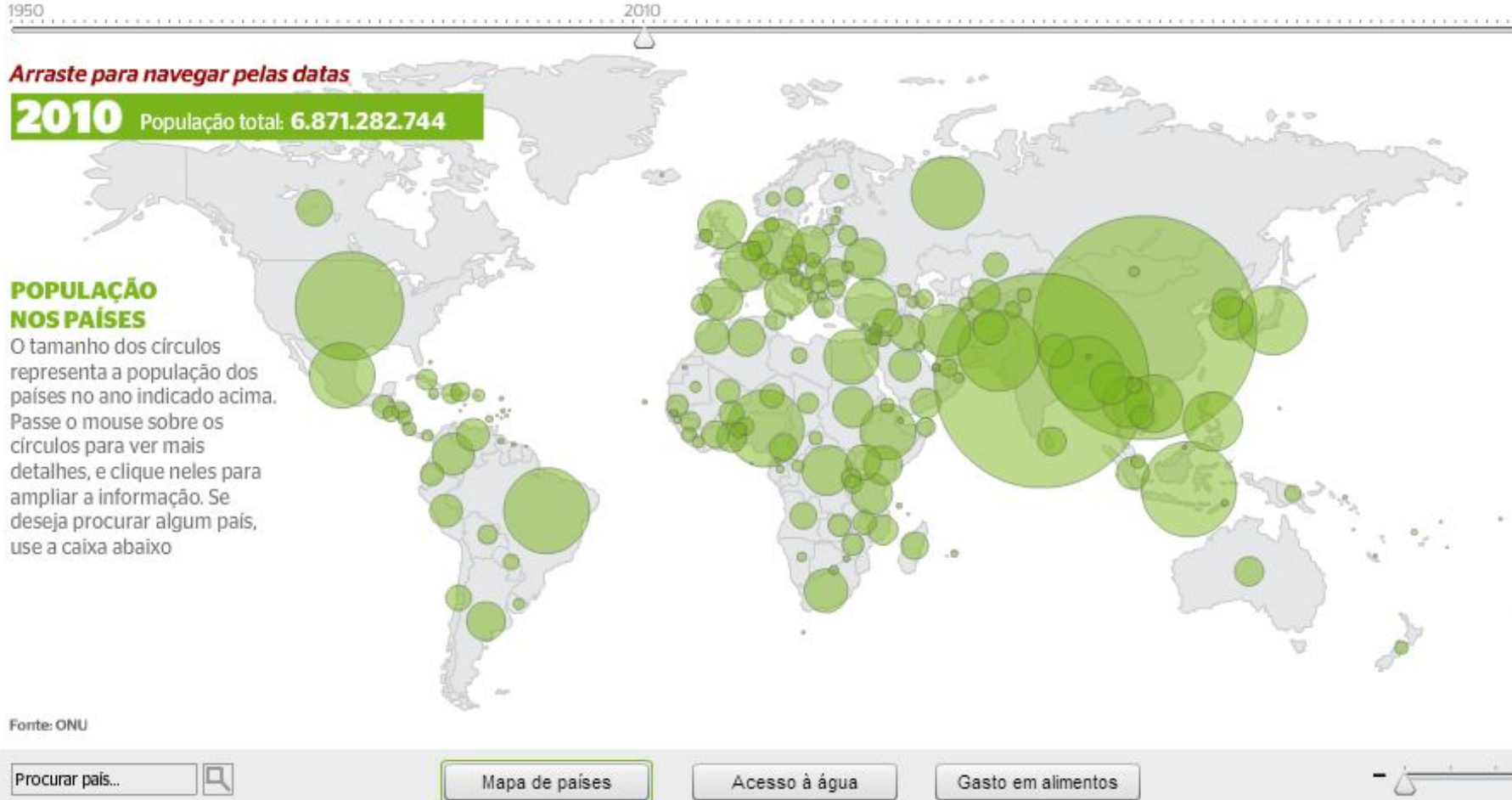
# Make more than one graphic

---

Topic: World population

- World Population Clock
- World Population by Country - The Guardian (Simon Rogers)
- World population since 1950 – Época Magazine (Alberto Cairo)

# Bubble Map with time slider



# Ranking Table with filters

→  
- Arraste para navegar pelas datas

**2016** População total: 7.335.772.141

Selecione segundo população ▼

Selecione por continente ▼

Posição ranking	País	2015	2016	2017
1	China	1.369.742.518	1.374.247.493	1.378.293.629
2	Índia	1.308.220.695	1.324.435.274	1.340.420.482
3	Estados Unidos	323.885.140	326.559.708	329.221.547
4	Indonésia	251.880.386	254.134.577	256.328.783
5	Brasil	203.293.957	204.827.785	206.307.431
6	Paquistão	189.648.128	192.835.890	195.993.631
7	Nigéria	179.790.814	184.410.727	189.143.615
8	Bangladesh	158.316.614	160.191.646	162.018.577
9	Rússia	142.229.066	142.040.701	141.829.117
10	Japão	126.071.988	125.884.727	125.663.983
11	México	120.057.647	121.298.238	122.503.851
12	Filipinas	101.420.912	103.078.273	104.739.511
13	Etiópia	91.999.509	93.815.451	95.627.442
14	Vietnã	92.442.560	93.292.501	94.111.051
15	Egito	88.178.706	89.548.180	90.895.716

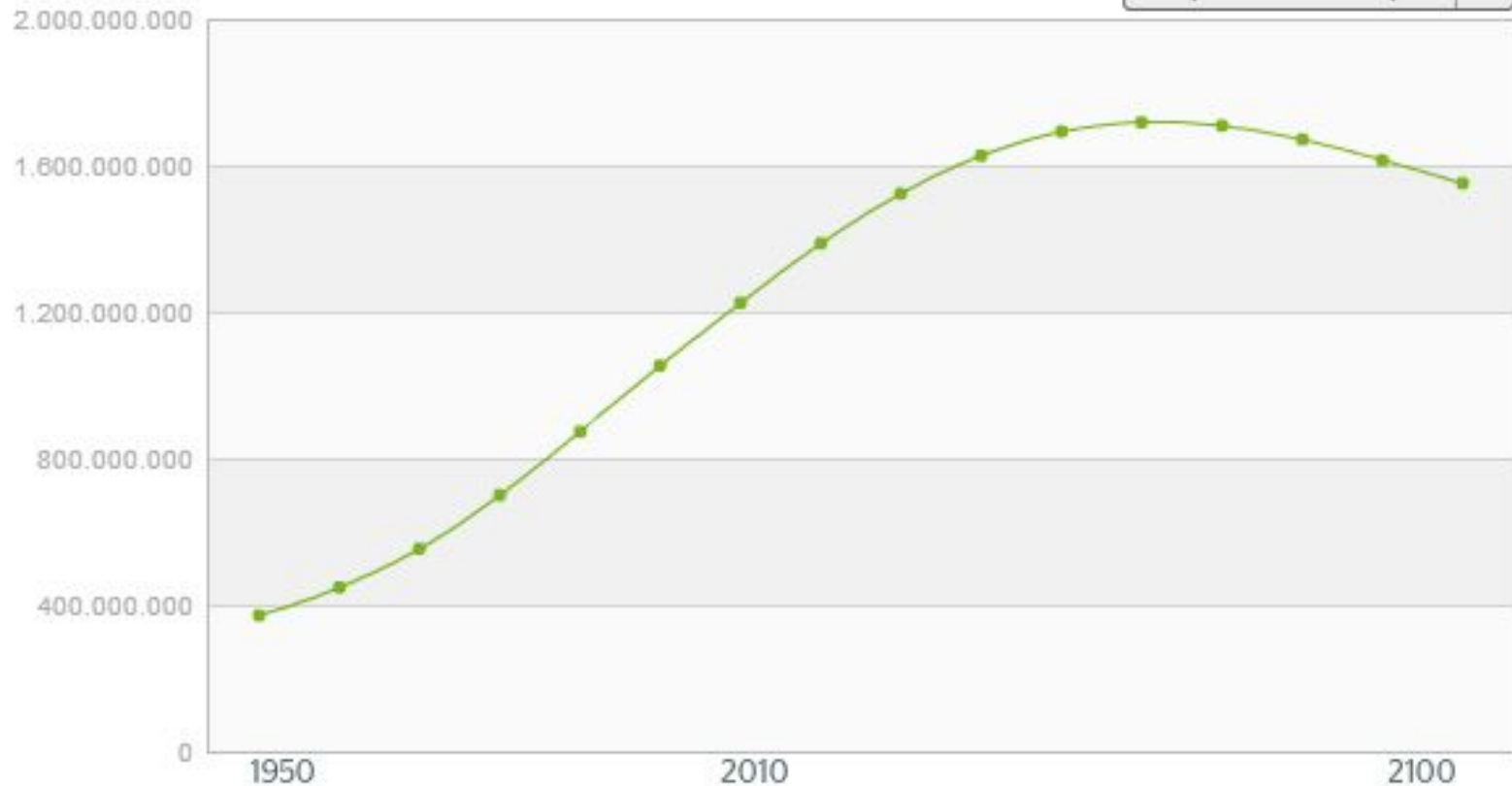
# Line Chart



ÍNDIA

POPULAÇÃO ENTRE 1950 E 2100

Compare com outro país ▼



PASSE O MOUSE SOBRE A LINHA PARA OBTER DETALHES DA POPULAÇÃO

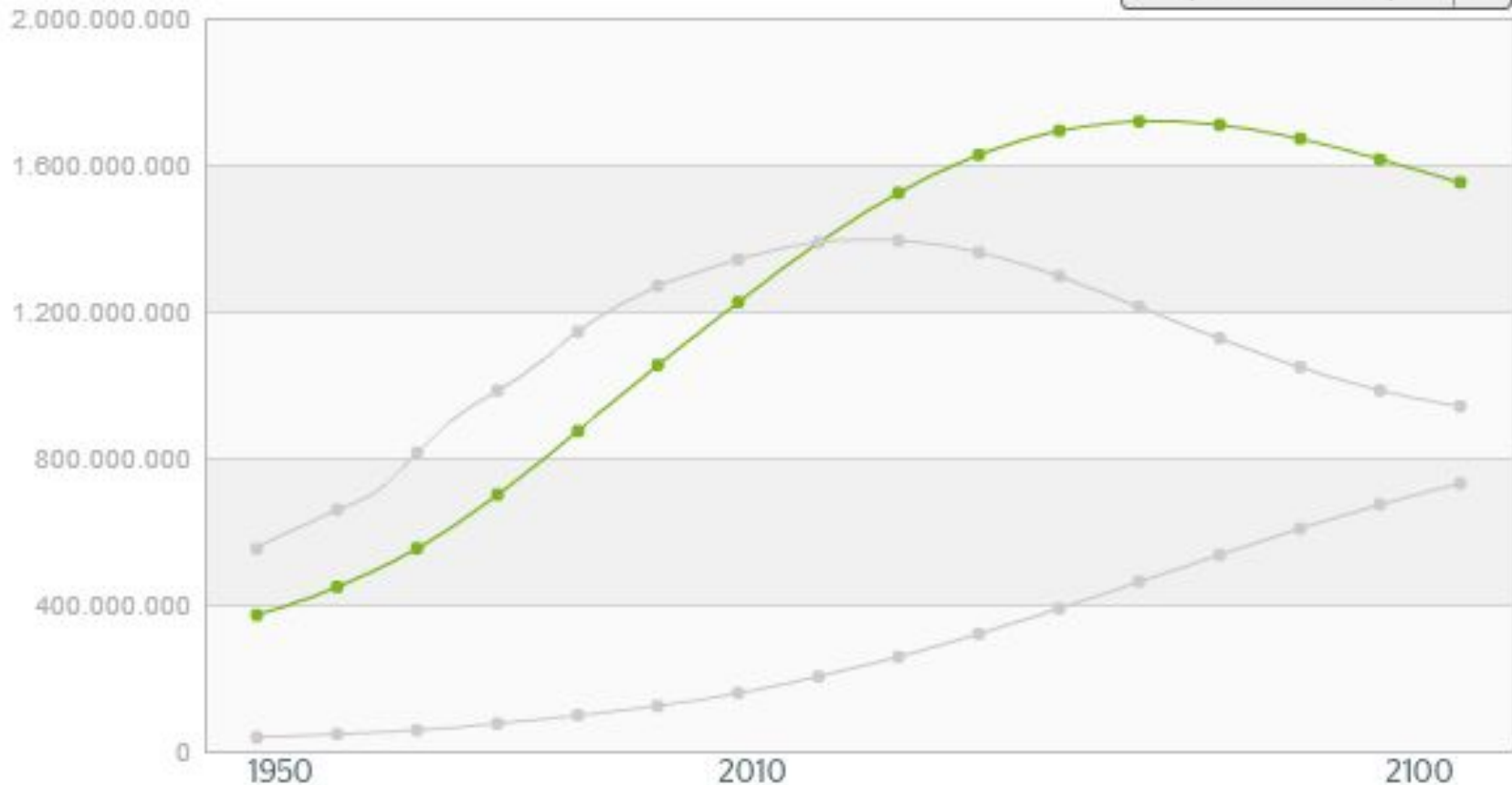
# Line Chart - India, China, Nigeria



ÍNDIA

POPULAÇÃO ENTRE 1950 E 2100

Compare com outro país ▼



# Summary for choosing the best graphic form

---

- **To represent your data, think about what questions you want readers to answer.** In base alle domande puoi scegliere il tipo di grafo più adatto. Se ha molteplici domande puoi pensare anche ad avere più di un grafo
- In large or interactive visualisations, it is appropriate or necessary to represent your data more than once.



# Samples

---

<http://truth-and-beauty.net/>

[Better Life Index](#)

---

# **The Art of Insight**

insightful & enlightening

# Insight (intuition, discernment, penetration,...)

---

- Insight is the discovery of unexpected or relevant information in any data set.
- A visualisation is created to give readers access and insight to data that they would otherwise not have.
- Many infographics **lack context**, meaning that they are not very insightful or relevant.
  - (e.g. [2000 mobile phones stolen every day](#))
- A number on its own is meaningless, it becomes relevant in context.

# Insight (intuition, discernment, penetration,...)

---

- Many infographics **lack context**, meaning that they are not very insightful or relevant.
- A number on its own is meaningless, it becomes relevant in context.

◦ **Call for alarm: 2,000 mobiles stolen every day**



# Create the context

---

There were 742,000 victims of mobile phone theft in England and Wales during 2012 and 2013, according to the Crime Survey for England and Wales

$742000/2=371000$  yearly

$371000/365 = 1016$  daily

[Uk Population = 65m](#)

[UK Mobile Phones = 83m](#)

4,4% mobile phones stolen

# What is the point of your graphic? Of your Story?

---

If something stands out, **highlight that bit of data.**

Example - **Homicides in the District** – Washington Post

- The graphic is an interactive graphic of homicides in Washington D.C.
- It also offers an extra feature to explore some of the main trends found in the data.
- *This serves as an entry point into the data.*

# Homicides in the District

## SEARCH

See homicides and victims in your neighborhood:

e.g. Mt. Pleasant, Capitol Heights

The Post tracked all homicides in D.C. between 2000 and 2011 to learn what happened to each case. It found that fewer than a third of the homicides have led to a conviction for murder or manslaughter. Find homicides in your neighborhood and see citywide trends here.

## EXPLORE

See the main trends we've discovered in this map.



Homicide rate falling in the District of Columbia »



About 30 percent of homicides have led to convictions »



Drug killings down 84 percent since 2000 »



Most dangerous age: 24 percent of those killed were in their early 20s »

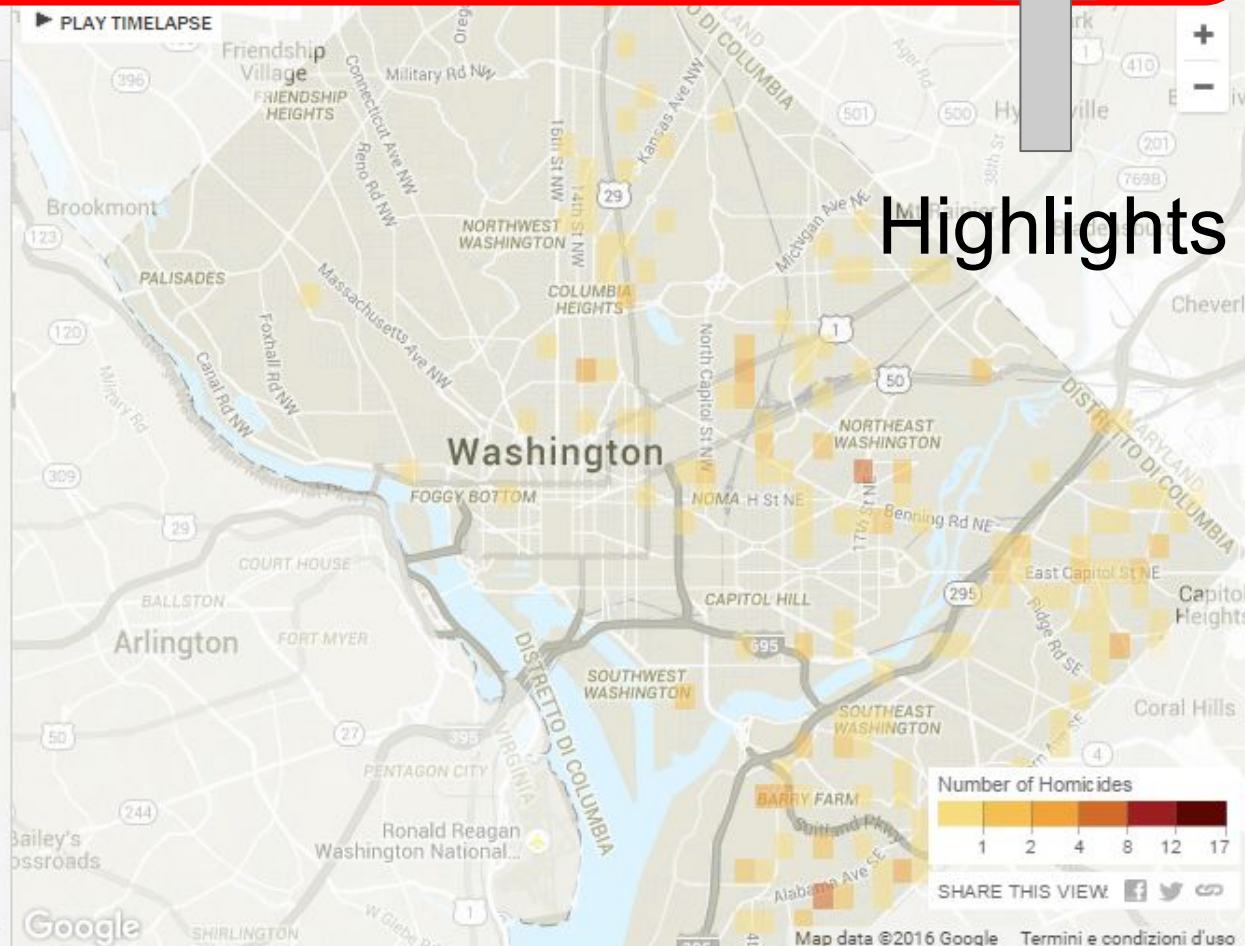
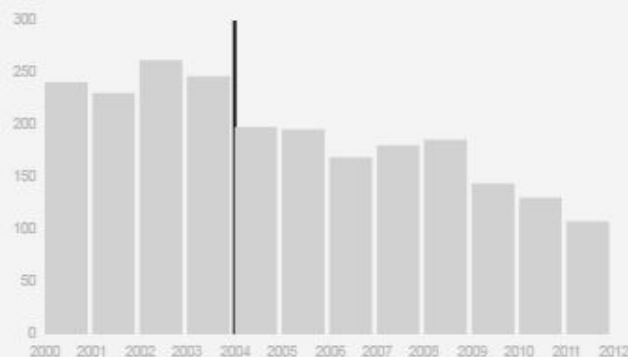
## DISTRICT OF COLUMBIA

BACK TO MENU

Victims: 20-to-24 year-olds accounted for 24 percent of homicide victims in 2011

More than half of the District's homicide victims between 2000 and 2011 were between the ages of 15 and 29. About 93 percent of those victims were male, and 94 percent were African American. Females accounted for 11 percent of the city's homicide victims, including 23 who were under the age of 14.

Homicides in D.C., 2000-2011



# Annotation

---

Interesting highlights in the data that you can explore are called 'Annotations'. The feature used to access these highlights is called the 'Annotation layer'.

You have three layers:

1. The **visual representation of the data**.
2. The **navigation layer** where readers can explore the data in interactive graphics.
3. The **annotation layer** where you highlight the revealing or interesting data points.

Write a proper headline, a short introduction, and explain the context of the data for the graphic.



# Example

---

## A Chicago Divided by Killings – The New York Times

- The first two layers consist of a map of homicides in the city of Chicago, with an overlay of the racial majority in each area.
- The third layer highlights some of the districts where exceptions or extremes can be found.
- *An introduction to the graphic helps the reader to understand the data.*

# Structuring your infographic as a story

---

If a visualisation can be structured as a story, the information is often more easily understood.

Example - [Brazil's Demographic Opportunity \(PDF\)](#) – The Functional Art – Alberto Cairo

- The graphic is based on the 2010 census data on Brazil.
- The first part shows the increase and decrease in population in the different regions.
- The second layer added context: comparing census data with fertility rates.

**DIAGRAM**  
NEWS IN PERSPECTIVE

**Brazil's Demographic Opportunity**

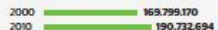
How Brazil can take advantage of a future with fewer children per couple.

Alberto Cairo, Francine Lima, Marco Vergotti

PRELIMINARY DATA FROM THE 2010 CENSUS create an interesting picture of the changes that the Brazilian population has gone through in the past ten years. Brazil's population grew, on average, 10% between 2000 and 2010, but the fertility rate is below 2.1 children per woman, the minimum to keep a population from shrinking. According to César Marques, a demographer from the University of Campinas, the main challenge Brazil will face in the future is how to maintain a healthy Social Security system if the number of older and retired people will likely be much larger than it is today. Read on to learn about all the variables at play in this story.

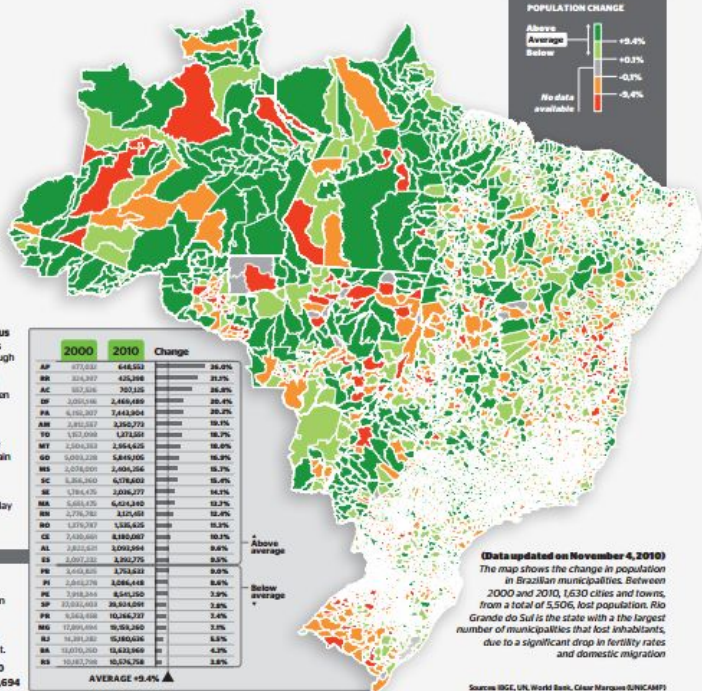
**1. BRAZIL'S POPULATION IS BIGGER**

The 2010 Census has revealed a 9.4% population increase between 2000 and 2010. The differences between states, as you can see on the chart on the right, are noticeable. Most rich states, such as São Paulo and Rio, didn't grow as fast as the ones in the north-east.



	2000	2010	Change
AP	877,032	648,553	-28.4%
MS	268,287	426,196	58.1%
AC	803,528	303,125	-62.4%
DF	2,050,146	2,468,689	20.4%
PA	6,162,267	3,444,954	-44.2%
AM	3,810,287	3,369,771	-11.6%
TO	1,833,068	1,871,681	2.1%
MT	2,504,263	2,944,425	17.6%
GO	6,076,228	5,849,195	-3.8%
MG	20,748,000	2,064,156	10.1%
SC	3,364,360	6,778,603	101.4%
RS	10,846,420	10,998,277	0.1%
MA	5,633,425	6,414,340	14.2%
RR	2,776,762	3,121,681	12.4%
RO	1,176,247	1,348,426	14.5%
CE	7,449,460	8,182,061	9.8%
AL	3,823,421	3,093,864	-19.1%
SE	3,097,232	3,962,275	28.0%
ES	6,646,628	5,763,523	-13.3%
PI	3,843,276	3,084,448	-19.7%
PE	3,918,244	4,541,150	15.4%
SP	33,022,000	30,543,000	-7.5%
PR	3,364,428	3,264,221	-3.0%
RJ	17,891,684	16,546,160	-7.5%
RN	4,381,180	4,386,426	0.1%
BA	10,011,760	10,638,069	6.3%
MT	10,823,768	10,624,768	-1.8%

AVERAGE +9.4%

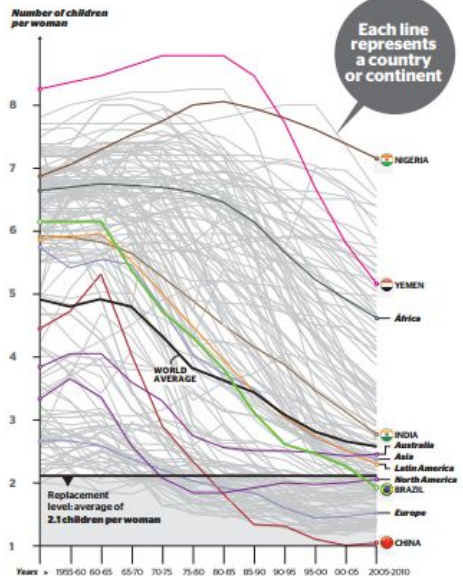


(Data updated on November 4, 2010)  
The map shows the change in population in Brazilian municipalities. Between 2000 and 2010, 1,630 cities and towns, from a total of 5,506, lost population. Rio Grande do Sul is the state with the largest number of municipalities that lost inhabitants, due to a significant drop in fertility rates and domestic migration.

Source: IBGE, UN, World Bank, César Marques @UNICAMP

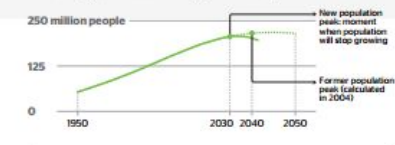
**2. —BUT THE FERTILITY RATE IS MUCH LOWER THAN EXPECTED**

A study in 2004 estimated that in 2010, the fertility rate would be 2.4 children per woman, on average. But new data collected by the IBGE prove that the fertility rate is already 1.9, below the threshold called "replacement rate". When the fertility rate drops below this number, the population of a country will eventually start to shrink and grow older.



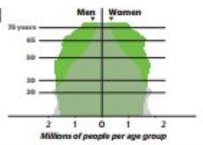
**3. AS A CONSEQUENCE, POPULATION WILL STOP GROWING—**

Forecasts made in 2004 anticipated that Brazil's population would stop growing in 2040. But the most recent data from the IBGE suggests that this could happen much earlier, in 2030.



**4. —AND IT WILL BECOME OLDER**

Comparing the current population pyramid with the one predicted for 2050



**How Brazil can transform the population challenge into an opportunity**

- As the population ages, the proportion of people of working age increases. The country will therefore have more people producing wealth (if the labor market can absorb them) and fewer children to consume investments. It is a window of opportunity, because in some cases the number of people of working age to fall back when older people are leaving the market.
- The population under 15 years of age is falling today. A smaller number of student in public schools will facilitate the quality of teaching, if the amount invested in education stays the same.
- Educational policy focused on low-income youth favors the formation of more skilled workforce and greater social mobility.
- In the future, Brazil will reach the stage of Europe and Japan, which struggle to support their elders. This is why it's so important to prepare a more balanced retirement system, which will include retirement at a later age.

1

2

3

# Example

---

The story: The consequences of the current fertility rate in Brazil.

- Main graphic: Brazil's population has grown from 2000 to 2010
- Line graph: Fertility rates per country
- Third and fourth graph: Brazil's population growth will decrease and the population will grow older
- Final part: How Brazil can transform the population challenge into an opportunity

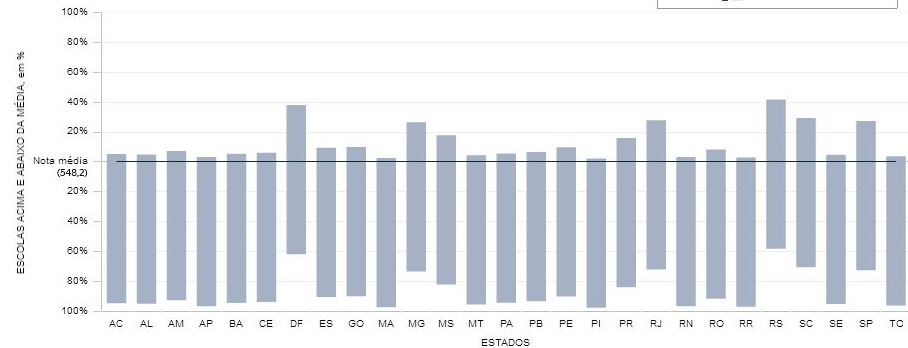
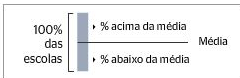
# Enlightening (illuminate, open someone's eyes)

---

- Graphics can change a reader's mind in relation to certain important issues.
- Preconceived ideas and how visualisations can confirm the preconceived idea.
  - *“Private schools are better than public schools”.*
  - *“The immigrants are invading Italy”*

## Escolas acima e abaixo da média nacional no Enem 2010

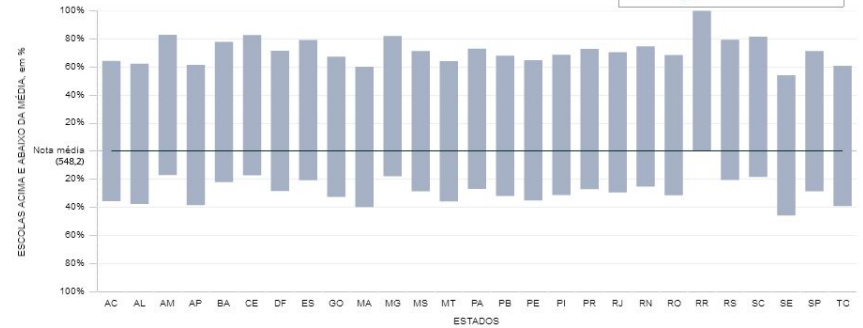
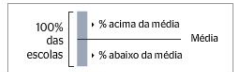
Com qualquer taxa de participação



**Public**

## Escolas acima e abaixo da média nacional no Enem 2010

Com qualquer taxa de participação



**Private**

**The Enem school ranking** – Época magazine

Bars representing 27 states, representing 100% of the schools in that state, but positioned according to the percentage of schools scoring above and below average