Bokeh for data visualization

...

Claudia Guirao Fernández

PyDay Madrid

17 sept 2016

Who am I?

Claudia Guirao

@claudiaguirao

Data Scientist at Kernel Analytics

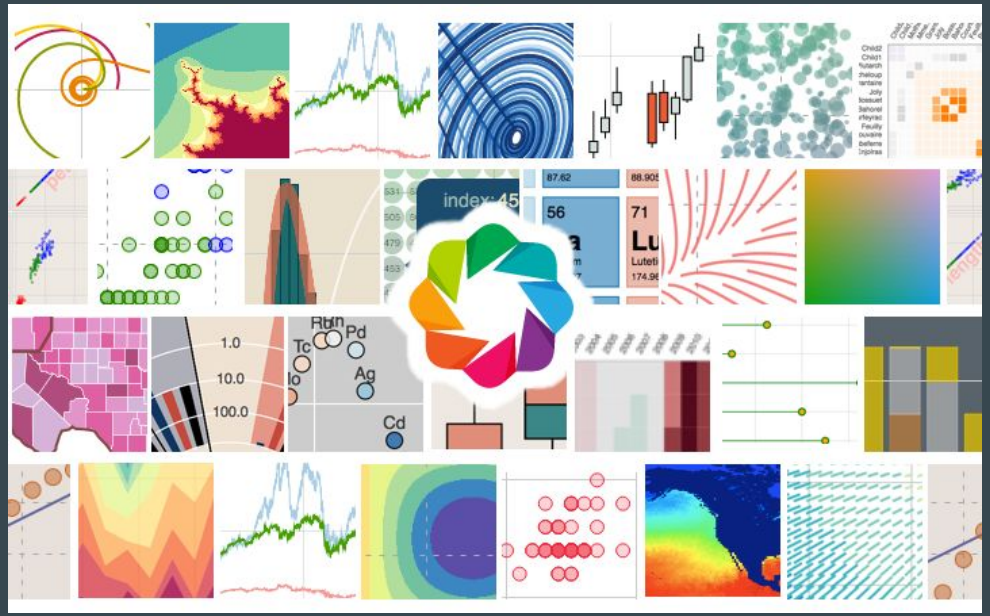
Proud member of Python Spain

Background:

- Double Degree in Law and Business Administration
- Master Degree in Visual Analytics

Today’s workshop

- Data visualization introduction
- Bokeh core concepts
  - Overview Bokeh
  - Navigating Bokeh's interfaces
  - Sharing your work



Workshop materials

- https://github.com/bokeh/bokeh-notebooks.git
- También puede descargarse en: https://github.com/bokeh/bokeh-notebooks/archive/master.zip
- Dependencies:
  - bokeh=0.12.0
  - pandas=0.18
  - ipython-notebook=4.0.4
  - ipywidgets=4.1.1

Data Viz history and theory

~~“Information is power”~~

“Knowledge is power”

“We are infoxicated” - Alfons Cornella

Data Viz history and theory

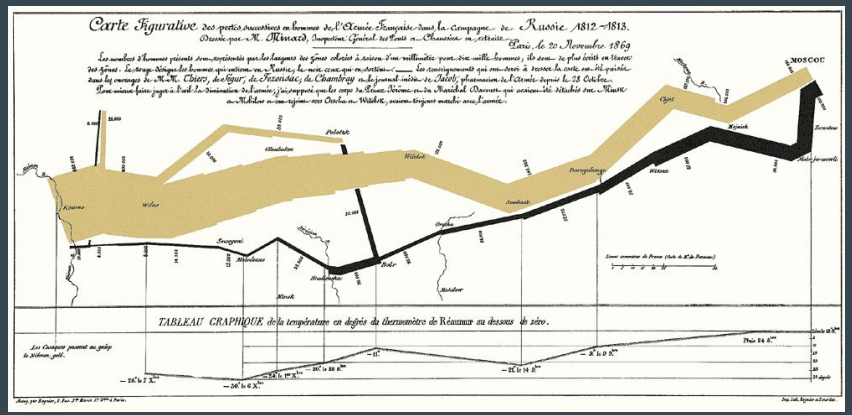
2 main objectives of data viz:

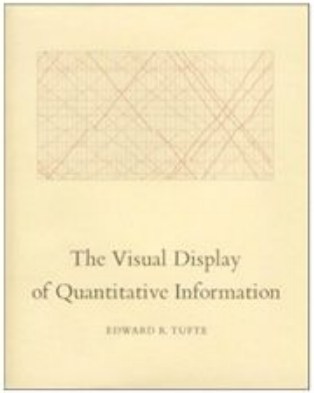
- explore amounts of data
- communicate data effectively

The goal of any good visualization has to focus the recipient of the information on what really is relevant and important: it leads to what in English is called ***insight***

Data visualization history

- Cartography
- Edward Tufte - The Visual Display of Quantitative Information
- Data visualization minimalism - Data-Ink Ratio



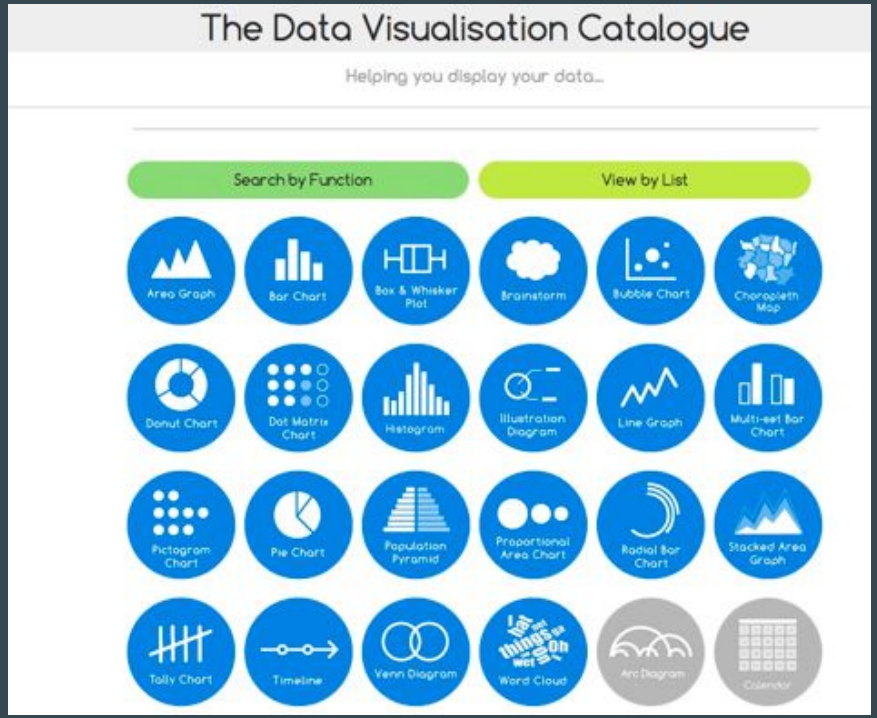


“Above all else show the data”

Tufte, 1983

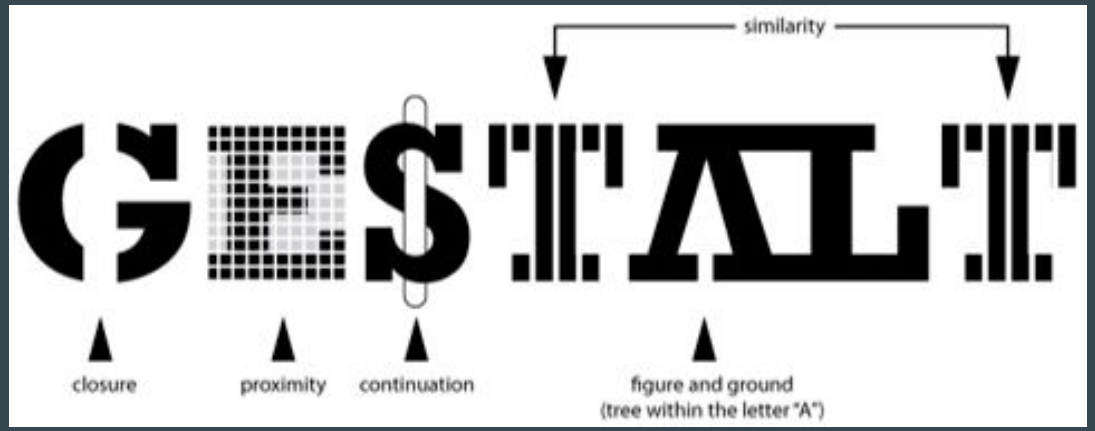
Data visualization categories

- Trends
- Patterns
- Anomalies
- Connections
- Correlations
- Comparison
- Hierarchy
- Locations
- Processes



http://www.datavizcatalogue.com/

Visualization components



YouTube

- Line
- Form
- Direction

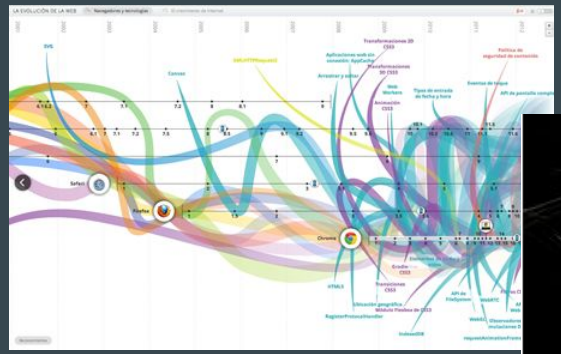
- Size
- Texture
- Color

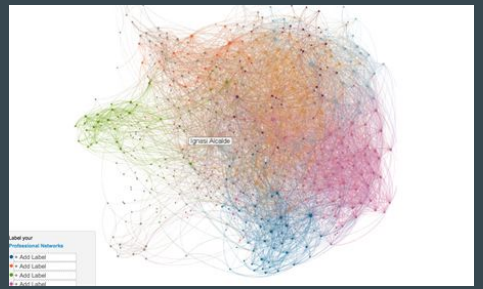
- Balance
- Proximity
- Alignment

- Repetition
- Contrast
- Space

Data viz is for everybody

http://evolutionofweb.appspot.com/





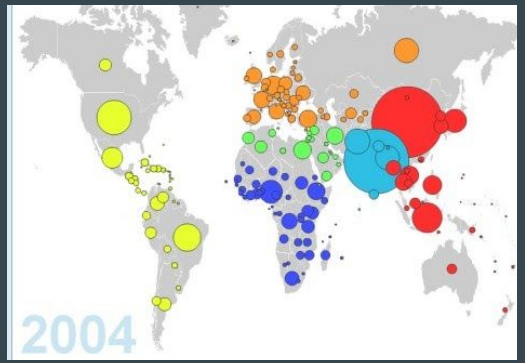
http://2010.tnsdigitallife.com/

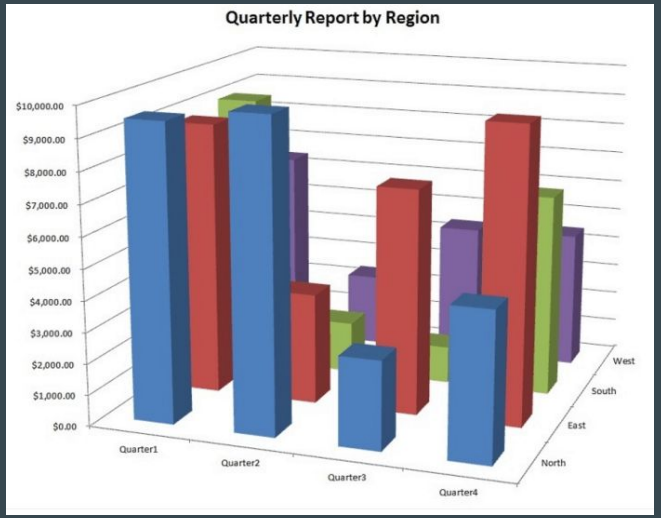
Data visualization tools

• Excel • PowerBI • Libreoffice • Tableau • QlikSense/View • ...	• ggplot • Matplotlib • Seaborn • Bokeh • Plotly	• D3 • custom solutions: javascript + SVG + HTML5 • javascript libraries (highcharts)

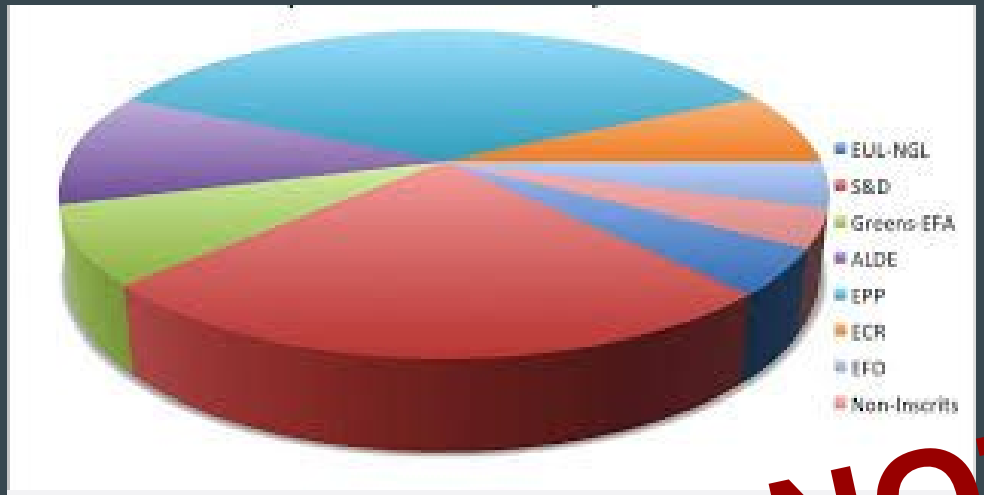

Style
ALONE IS MERELY
an elegant
look


NewTest


2004

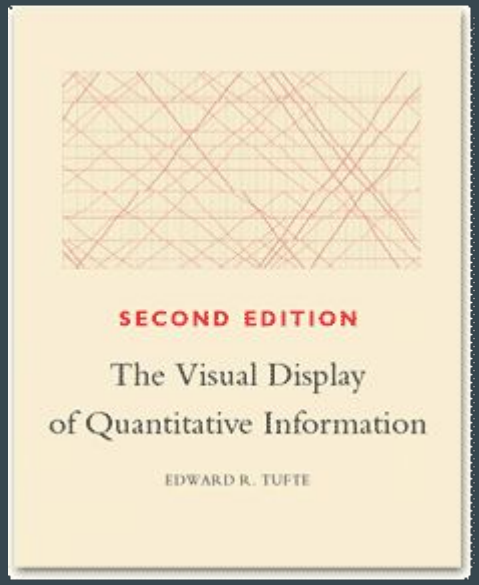

Quarterly Report by Region
$10,000.00
$9,000.00
$8,000.00
$7,000.00
$6,000.00
$5,000.00
$4,000.00
$3,000.00
$2,000.00
$1,000.00
$0.00
Quarter1
Quarter2
Quarter3
Quarter4
West
South
East
North

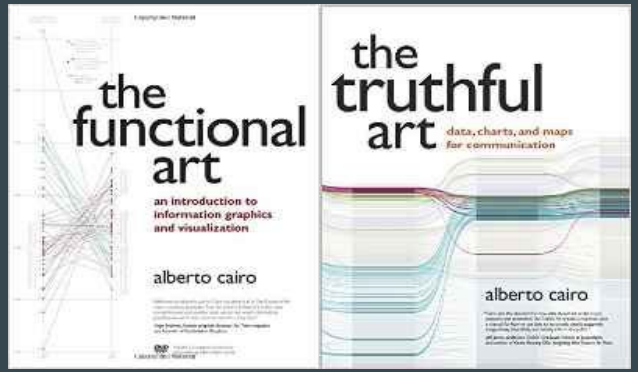
GAPMINDER

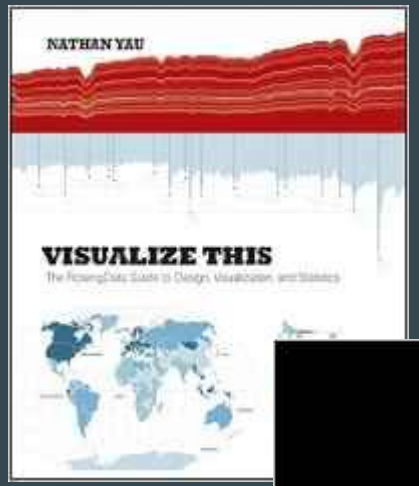


Please do NOT do this !!!

Some good books



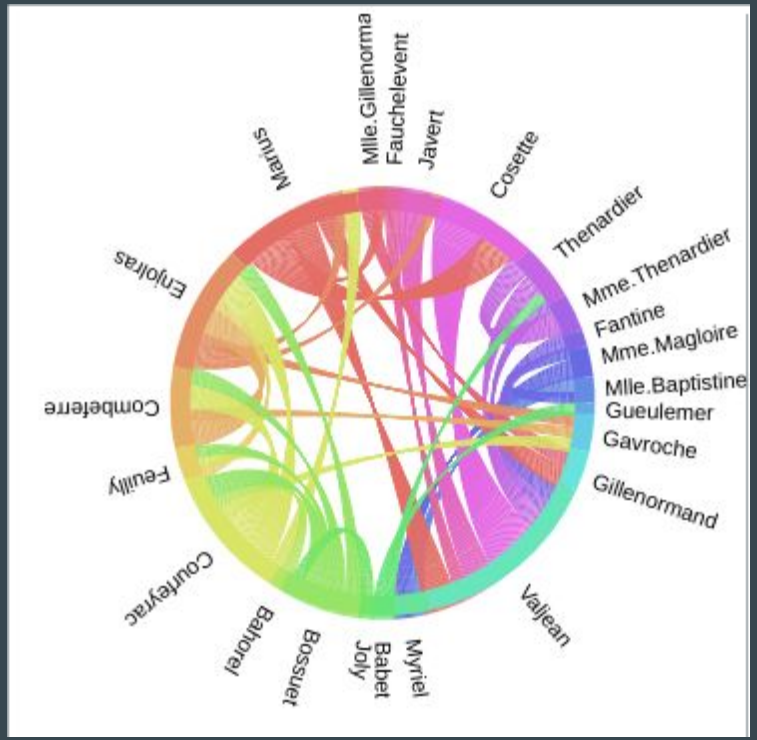




What is Bokeh?

- Data visualization library
- Web-based & interactive
- NO javascript required (yupiiiiiiiii)
- Large data set visualization
- Real time data visualization

http://bokeh.pydata.org/en/latest/docs/gallery.html




```
import pandas as pd
from bokeh.charts import output_file, Chord
from bokeh.io import show
from bokeh.sampledata.les_mis import data

nodes = data['nodes']
links = data['links']

nodes_df = pd.DataFrame(nodes)
links_df = pd.DataFrame(links)

source_data = links_df.merge(nodes_df, how='left', left_on='source', right_index=True)
source_data = source_data.merge(nodes_df, how='left', left_on='target', right_index=True)
source_data = source_data[source_data["value"] > 5]

chord_from_df = Chord(source_data, source="name_x", target="name_y", value="value")
output_file('chord_from_df.html', mode="inline")
show(chord_from_df)
```

Interfaces

- Ways to use Bokeh - pick the one that's right for you:
    - **`bokeh.charts`:** high speed. One-line charts. Processes your data & spits out a chart
    - **`bokeh.plotting`:** sensible defaults. Tries to pick sensible defaults You organize your data, it organizes your plot
    - **`bokeh.models`:** high customization. The lowest level. Offers you the most control. Do all the work yourself

Bokeh charts

Area, Bar, Box, Donut, Dot, Heatmap, Histogram, Horizon, Line, Scatter, Step, Timeseries

Wherever possible, the interface is designed to be simple to use with pandas, by accepting a DataFrame and names of columns directly to specify data.
reference

Pandas friendly