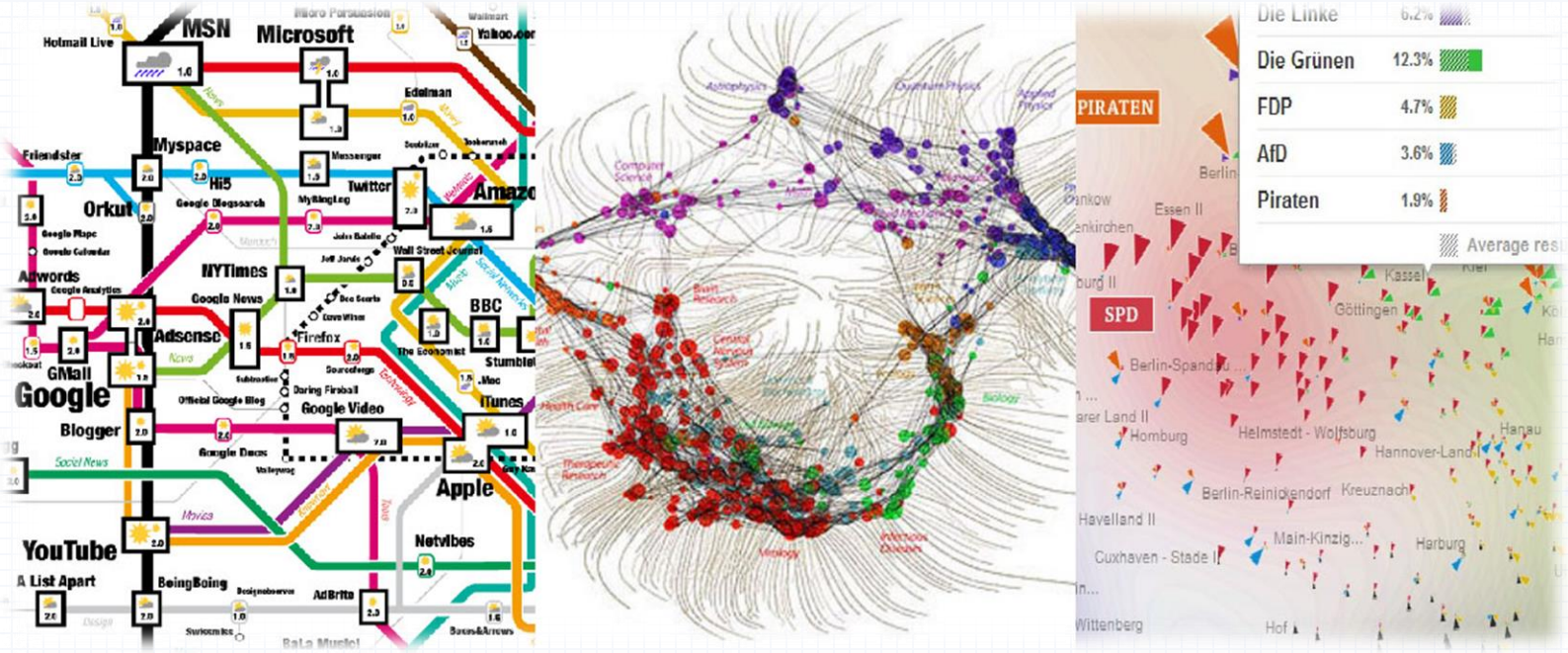


Data Visualization

HAW HAMBURG

Vinh Phan

Seminar on 28/10/2014

Agenda

1. Objective
2. Motivation
3. Approach & Risk Assessment

Objective

1. Objective

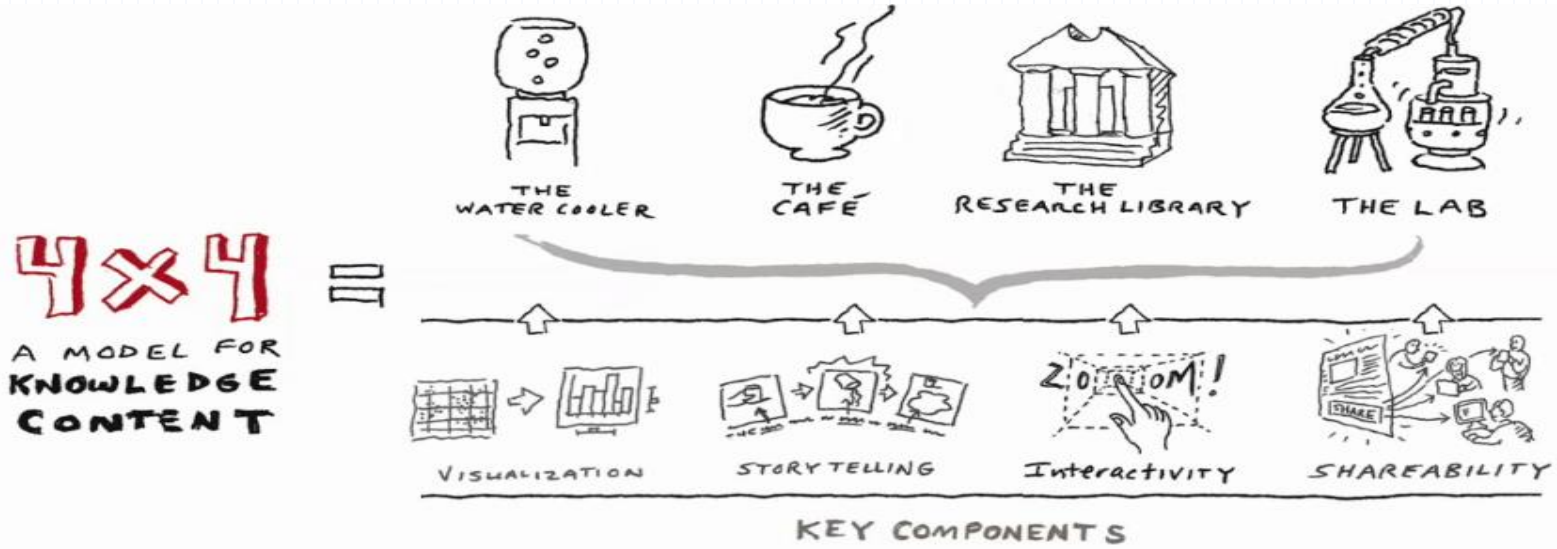
Is it possible to effectively make compelling stories from large amount of (possibly complex) data by means of visualization?

Motivation

2. Motivation

- Get the right people to the right content at the right time
  - 4x4 model for knowledge content
  - Good visualization is a key aspect
- What makes a good visualization
  - Accuracy
  - Story
  - Knowledge

http://bit.ly/1DkUUiH



Approach & Risk Assessment

3. Approach & Risk Assessment

3.1 Getting the Big Idea

- Get some ideas (about what to tell/explain/prove)

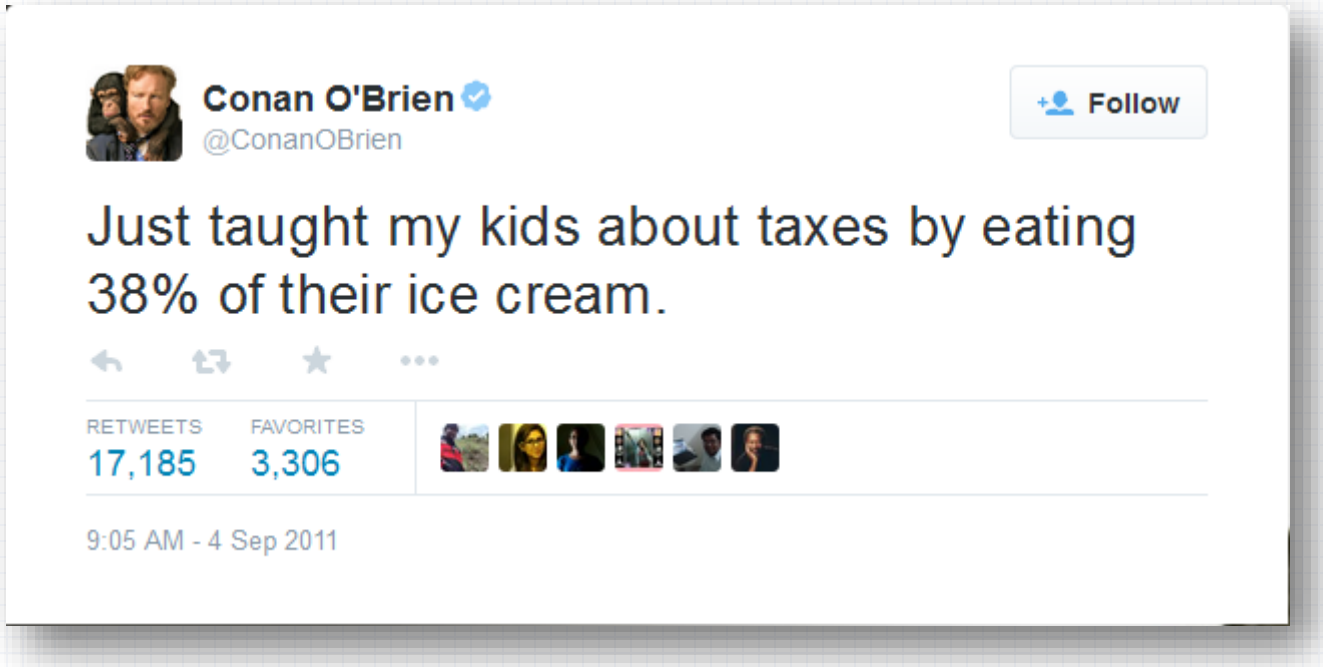
©NYTimes

©NYPost

3. Approach & Risk Assessment

3.1 Getting the Big Idea

- Get some ideas (about what to tell/explain/prove)



http://bit.ly/1tyBSWg

3. Approach & Risk Assessment

3.2 Channeling the Audience

- Culture
- Level of expertise
- Accessibility
- Consumption context/channel
  - affects approach/level of standard
- True believers or skeptics?
  - affects acceptance, may lead to bias
- Action
  - What kind of conversation & interaction/reaction should this visualization evoke?

3. Approach & Risk Assessment

3.2 Channeling the Audience

- Culture
  - affects language, perspective, color, narrative context…

3. Approach & Risk Assessment

3.2 Channeling the Audience

- Level of expertise
  - affects language, context information, approach…

vs.

3. Approach & Risk Assessment

3.2 Channeling the Audience

- Accessibility
  - Big issue: visual impairment (color blindness…)
  - affects visual element design (color, font size, contrast…)
  - http://www.color-blindness.com/coblis-color-blindness-simulator

3. Approach & Risk Assessment

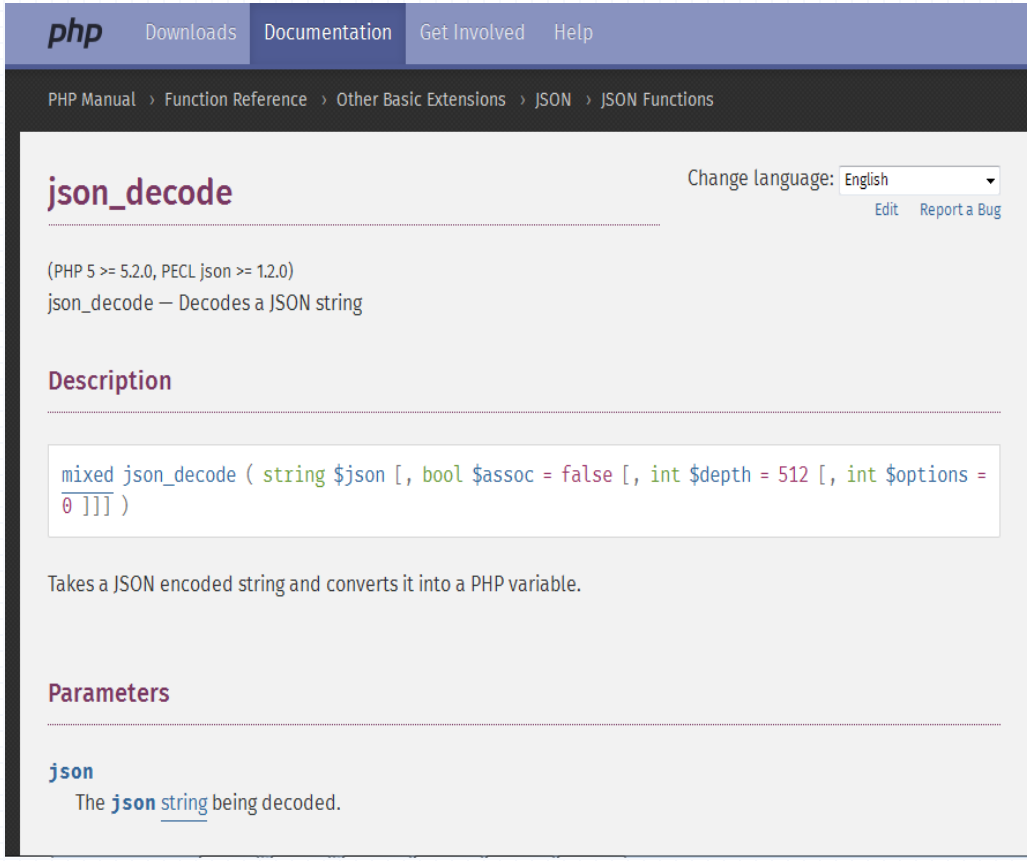
3.3 Sourcing & Preparing The Data

- Find, evaluate and select reliable data. Possible sources include:
  - Government & political data (Data.gov, Destatis.de, Socrata, Census.gov…)
  - Data aggregators (Programmable Web, Infochimps, Junar, Buzzdata…)
  - Social data (Twitter, Foursquare, Facebook…)
- Understand the data/data parameters
  - Mean, median, rank index, percentile, correlation, causation…
  - Sample size & methodology
  - Helps: sport mistakes, reduce bias
- Explore the data
  - Spreadsheet, visual analytic tools/softwares
  - Add more contextual data if necessary

3. Approach & Risk Assessment

3.3 Sourcing & Preparing The Data

- Clean and get data into the right format
  - Data adjustments: calculate indexes, ratios, percentiles, aggregate/regroup data…
  - ➢ Spreadsheet, Data Wrangler, Google Refine…
  - Data conversion: Excel/CSV/TSV… to JSON/XML/MySQL…
  - ➢ Online tools, programming language built-in functions…
- Associated risks include:
  - Data cleaning: complexity, time consuming.
  - Inability to find reliable data source.
  - Data requires extensive, deep knowledge to comprehend

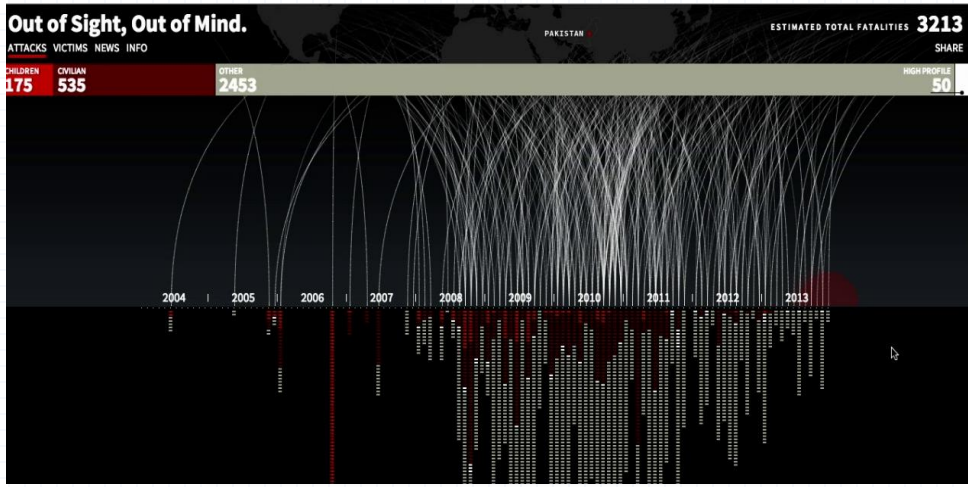


3. Approach & Risk Assessment

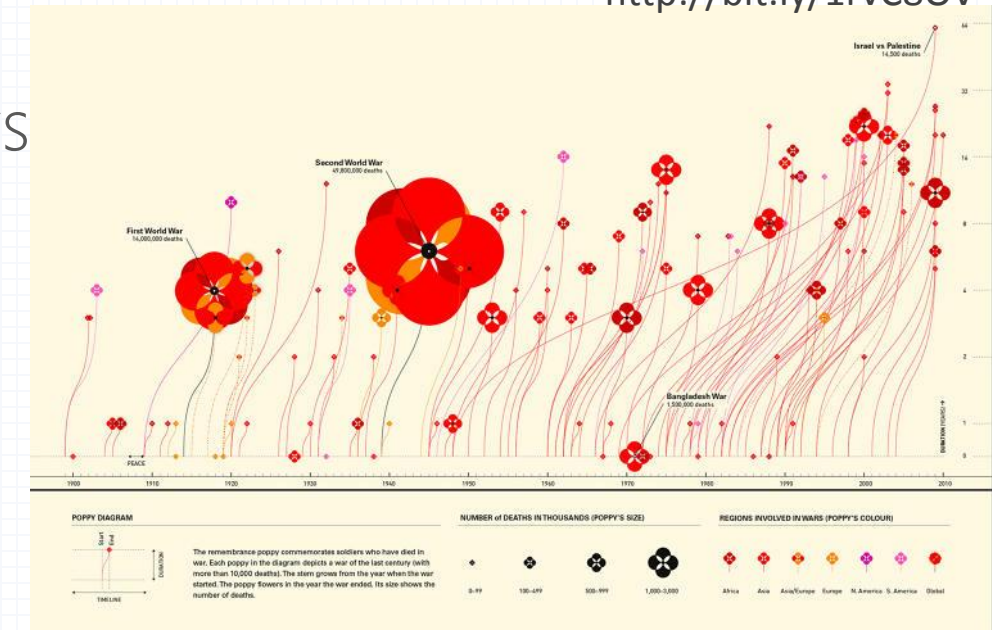
3.4 Experimenting with Visual Designs/Visual Elements

- Sketches & Wireframes
- Visual Elements
  - Illustration, iconography, typography
  - Position, size, shape, color, contrast…
- Finding the right paradigm
  - Basic forms (charts, diagrams, maps…) vs
  - Creativity & innovation
- Risks include
  - Lack of good design skill/artistry/time
  - Overuse of visual elements

http://bit.ly/1fhDw3U



http://bit.ly/1rvC8OV



3. Approach & Risk Assessment

3.5 Finding the Right Technology & Start Visualizing

Things to consider

- Platform vision

One-off
Short-term
Static

Reusable
Permanent
Evolving

3. Approach & Risk Assessment

3.5 Finding the Right Technology & Start Visualizing

- Audience

Technophobic
Old-fashioned
Under-gadgetized

Techno-driven
Modern
Gadget-laden

3. Approach & Risk Assessment

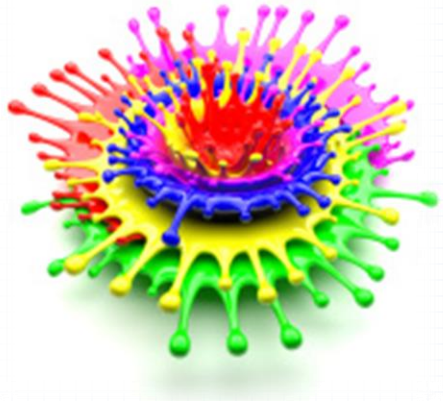
3.5 Finding the Right Technology & Start Visualizing

- Visual/Conceptual goals

Simple
Standard

©DepositPhotos

Complex
Custom



©iStockPhotos

3. Approach & Risk Assessment

3.5 Finding the Right Technology & Start Visualizing

- Software packages/Off-the-shelf solutions
  - Tableau, QlikView, Gephi, Highcharts, Zingchart, Pentaho…
  - \+ Quickly build complex, interactive visualizations
  - \+ Shorter learning curve
  - \- Few customizations, less flexibility
  - \- Cost

http://www.highcharts.com

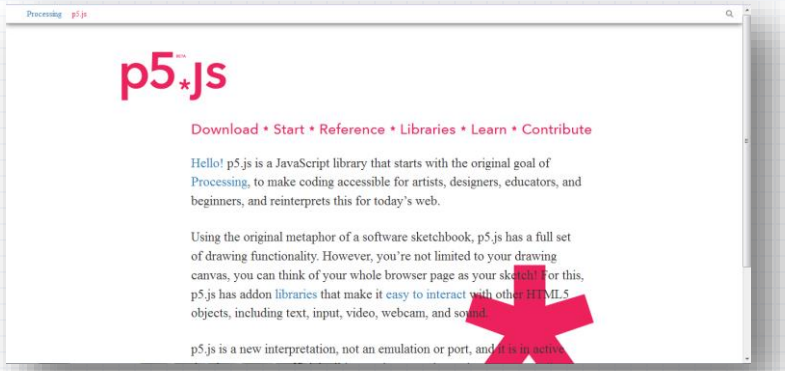
http://www.zingchart.com

3. Approach & Risk Assessment

3.5 Finding the Right Technology & Start Visualizing

- DIY: HTML5 vs. SVG libraries
- HTML5: raster-based, loss of details when resizing. Cannot add event-listeners to canvas shapes. Great for raster images & sprites…
  - Chart.js, Fabric.js, p5.js…
- SVG: XML-based, minimum loss of quality when resizing/porting to different devices. Not scale well in charts with large amounts of elements…
  - D3.js, Raphael.js, Snap.svg…

http://www.p5js.org



http://www.d3js.org

3. Approach & Risk Assessment

3.5 Finding the Right Technology & Start Visualizing

Associated risks include

- Technical abilities

Got it

©Photos.com

Needs help

©Masterfiles.com

3. Approach & Risk Assessment

3.6 Sharing, Getting Feedbacks & Assessing Outcome

Survey methodology:

- Small poll (~ 5-10 questions, ~ 10-20 participants)
- Compare between visualization vs. visualization / visualization vs. raw data
- What to assess:
  - How fast a fact/an attribute of the data can be recognized
  - How compelling/credible the stories are, according to users' opinion
  - Overall user experience/usability

3. Approach & Risk Assessment

3.6 Sharing, Getting Feedbacks & Assessing Outcome

Risks include

- Not enough survey respondents
- Not enough meaningful questions
- Feedbacks' quality
- Accuracy/Objectivity of result

Thank you for your attention.