Data visualization and journalism
Alexandre Gonçalves
computer science + journalism

- Computer science as a tool to improve news media
- Data visualization as a tool to tell a story
- Computer science as a tool to understand news media
- Data visualization as a tool to find patterns in media behavior
data visualization to tell a story
data visualization to tell a story

- visualization for the masses
- visualization to build communion
- visualization for social change
- visualization + narrative
- visualization and the duty of beauty
visualization for the masses
Communication is linked to terms such as “sharing,” “participation,” “association,” “fellowship,” and “the possession of a common faith.” This definition exploits the ancient identity and common roots of the terms “commonness,” “communion,” “community,” and “communication.” (James Carey, Communication as Culture)
visualização para mudar o mundo

Fies: o governo empresta dinheiro

- Estudantes que não pagam: problema do governo, não da instituição;
- Eles recebem dinheiro do governo, não dos estudantes;
- As instituições com o Fies é de "capitalismo sem risco";
- Os contratos de matrículas e ações são conexos;
- O Fies é um sistema de socialização do risco.

Projects.nytimes.com/guantanamo
Malofiej 2012

Fies: the government lends money to students to fund their college education in private institutions (low interest rates and long grace periods).

Private institutions instruct their students to apply for the loan: they prefer to receive money from the government than from the students.

If the students do not pay for the loan, it is a government’s problem, not a college owner’s problem. Private profit, socialized risk.
visualization + narrative

Statistics don’t bleed.
(Arthur Koestler)

bit.ly/nsa_decoded
Malofiej 2014
visualization and the duty of beauty

Beauty is mysterious as well as terrible. God and devil are fighting there, and the battlefield is the heart of man. (Dostoevsky, The Brothers Karamazov)

why the NYT bias?
<table>
<thead>
<tr>
<th>challenges</th>
<th>news stories are expensive</th>
<th>most people agree that they are important</th>
<th>nobody wants to pay for it</th>
</tr>
</thead>
<tbody>
<tr>
<td>data journalism is even more expensive</td>
<td>non-profit organizations or public funding</td>
<td>data literacy among journalists</td>
<td></td>
</tr>
</tbody>
</table>
data visualization to find patterns in media behavior

- SOPA/PIPA debate
- Russian blogosphere
- Flashmobs in Brazil
What is Media Cloud?

Web crawler → Database ← Search platform

Perl, PostgreSQL, Apache Solr
**SOPA/PIPA**

- Defeated in January 2012.
SOPA/PIPA is anti-piracy legislation

Jan 18

SOPA/PIPA is about Internet censorship

Source: ProPublica
Benkler et al.

- the emergence of the networked public sphere

bit.ly/harvard_sopa
Controversy Mapper

Week X:

- NYT
- WSJ
- wired
- reddit
- blog A

The diagram shows the connections between these entities for Week X.
Controversy Mapper

Week X:

- NYT
- WSJ
- wired
- reddit
- blog A
Qualitative analysis

- What did WIRED publish that week?
- Who wrote the articles?
- Why were those articles quoted by other media sources?
Many continued to link to the list of SOPA supporters at the House Judiciary site and to the Wikipedia pages describing SOPA and PIPA. (...)

An often-repeated meme from the earliest stages of the debate had been that the cost of piracy to the United States is $58 billion each year. This meme was repeated across many stories, including across mainstream media outlets. On January 3, 2012, Julian Sanchez published a piece in Cato @ Liberty in which he dismantled this $58 billion figure. Although few actors linked directly to Cato, and the Cato site appeared as a small node in this week, the story itself received attention due to an attention backbone, as Techdirt and EFF directed readers to [the article].
Rolezinhos are "...gatherings organised via social networks in which tens, hundreds, and sometimes thousands of youngsters from São Paulo's poor periphery meet up in the city’s shopping malls..."

The Economist, 01/20/2014

Shopping Itaquera, 01/11/2014, 3,000 youngsters
Aim: framing analysis of the rolezinhos issue

To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation. (Entman, 1993)
<table>
<thead>
<tr>
<th>Frame element</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>problem definition</td>
<td>main actor: youth, state and society</td>
</tr>
<tr>
<td></td>
<td>causal interpretation:</td>
</tr>
<tr>
<td></td>
<td>economic: new middle class</td>
</tr>
<tr>
<td></td>
<td>social: social criticism, lack of civility, social or racial prejudice</td>
</tr>
<tr>
<td></td>
<td>cultural: <em>funk ostentação</em> and social networks, few options for recreation</td>
</tr>
<tr>
<td>moral evaluation</td>
<td>main actors: youth, malls, police</td>
</tr>
<tr>
<td></td>
<td>government: city level, state level, federal level, courts</td>
</tr>
<tr>
<td></td>
<td>society: media, progressives, conservatives</td>
</tr>
<tr>
<td>treatment recommendation</td>
<td>against: fines and detentions, <em>rolezinhos</em> in public parks</td>
</tr>
<tr>
<td></td>
<td>pro: more <em>rolezinhos</em> and protests, dialogue and reflection</td>
</tr>
</tbody>
</table>

Table 2.1: Variables for hierarchical cluster analysis.
Figure 3-1: Root of the tree and the three main frames.

Figure 3-5: Arrastão cluster. A collapsed and annotated version of Figure 3-3.

Middle ground cluster (Figure 3-4)

Arrastão cluster (Figure 3-3)

Apartheid cluster (Figure 3-2)

YouTube subcluster (Figure 3-8)

Traditional media subcluster (Figure 3-7)

[175, Folha de S. Paulo: The malls’ perspective]

[174, Video: Girls in rolezinhos]

[173, Rolezinho in a neighboring town]

Neocons subcluster (Figure 3-9)
Arab Spring in Russia and the US

bit.ly/russian_blogosphere
thank you!

a.goncalves@columbia.edu

bit.ly/pres_alex