Data-Driven Documents

Introdução ao D3

Diego Barros
diego.fariabarros@gmail.com

Quem sou eu?

D3 - Documento Dirigido a Dados

- Biblioteca JavaScript para manipular documentos baseados em dados
- Ênfase nos padrões Web sem nos forçar a usar um *framework* proprietário

https://github.com/diegobarros

Michael Bostock [2011]

$\mathbb{D}^3$: Data-Driven Documents

Michael Bostock, Vadim Ogievetsky and Jeffrey Heer

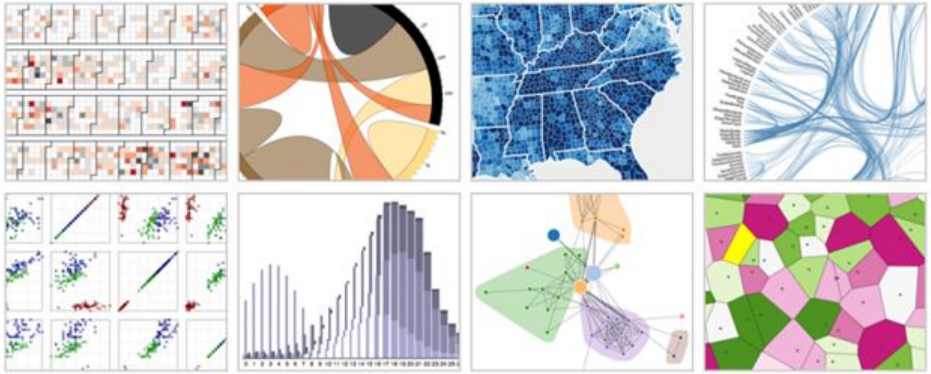


Fig. 1. Interactive visualizations built with D3, running inside Google Chrome. From left to right: calendar view, chord diagram, choropleth map, hierarchical edge bundling, scatterplot matrix, grouped & stacked bars, force-directed graph clusters, Voronoi tessellation.

Abstract—Data-Driven Documents (D3) is a novel representation-transparent approach to visualization for the web. Rather than hide the underlying scenegraph within a toolkit-specific abstraction, D3 enables direct inspection and manipulation of a native representation: the standard *document object model* (DOM). With D3, designers selectively bind input data to arbitrary document elements, applying dynamic transforms to both generate and modify content. We show how representational transparency improves expressiveness and better integrates with developer tools than prior approaches, while offering comparable notational efficiency and retaining powerful declarative components. Immediate evaluation of operators further simplifies debugging and allows iterative development. Additionally, we demonstrate how D3 transforms naturally enable animation and interaction with dramatic performance improvements over intermediate representations.

Index Terms—Information visualization, user interfaces, toolkits, 2D graphics.

1 Introduction

When building visualizations, designers often employ multiple tools simultaneously. This is particularly true on the web, where interactive visualizations combine varied technologies: HTML for page content, CSS for aesthetics, JavaScript for interaction, SVG for vector graphics, and so on. One of the great successes of the web as a platform is the (mostly) seamless cooperation of such technologies, enabled by it incurs a high opportunity cost: it ignores developers' knowledge of standards, and the tools and resources that augment these standards.

The resulting cost to *accessibility*—the difficulty of learning the representation—may trump efficiency gains, at least for new users. Scarcity of documentation and ineffectual debugging exacerbate the problem, impeding users from gaining deeper understanding of toolkit

Porque aprendemos d3?

- Possibilita implementar a maioria das técnicas de visualização de dados
- Te dá o poder de criar suas próprias técnicas, algo que bibliotecas de gráficos não podem fazer

Bibliotecas de Gráficos X Bibliotecas de Visualização

1. Highcharts JS
2. Fusioncharts
3. plotly.js
4. Google Charts
5. ChartJS
6. Chartlist.js
7. n3-charts
8. Ember Charts
9. ZingChart
10. uvCharts

1. **D3js**
2. **HTML 5 Canvas**
3. **WebGL (Three.js)**
4. Processing

Visualização?!

O termo visualização se aplica a qualquer **representação visual da informação** que permita **análise, exploração** e **comunicação dos dados**

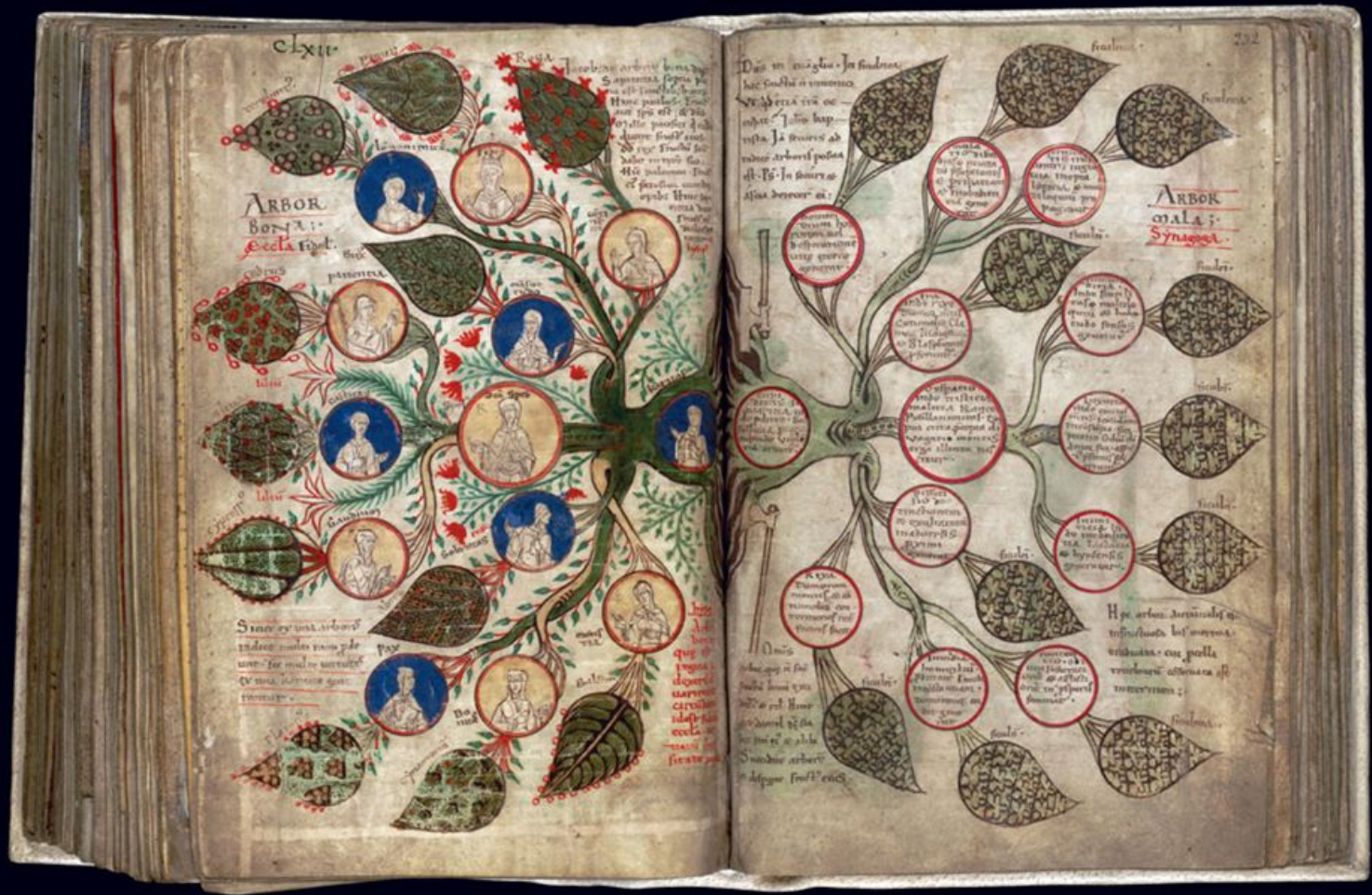
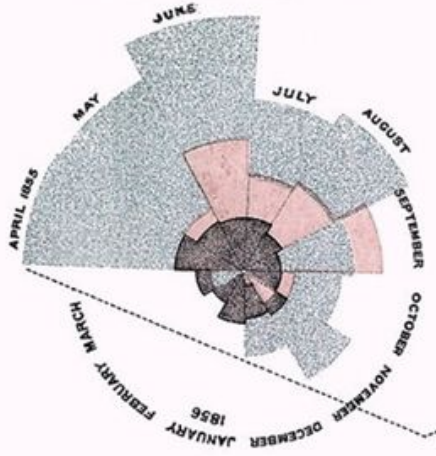
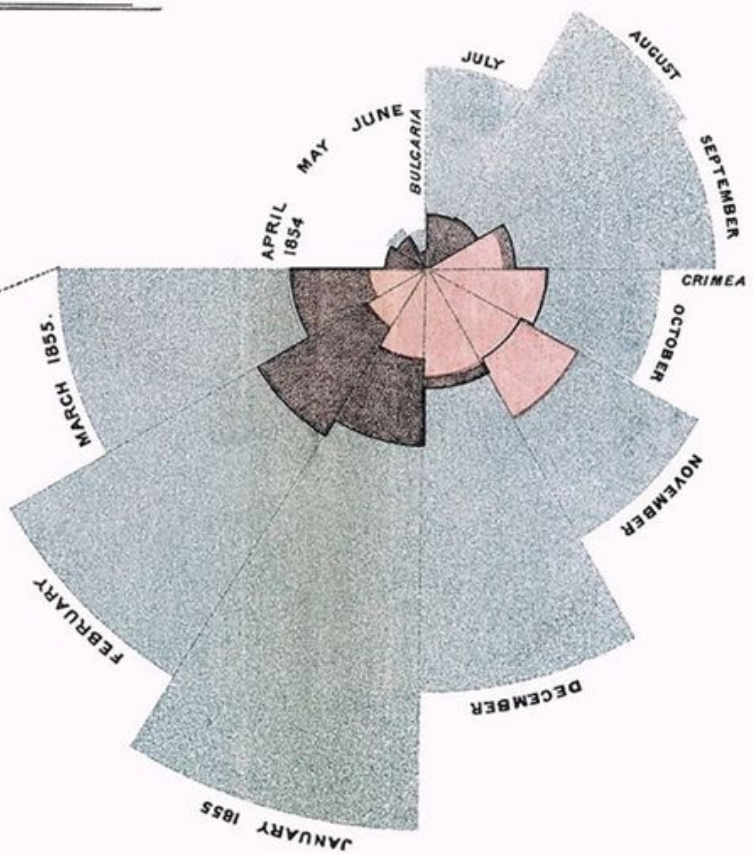

CLXII
ARBOR
BONA
ARBOR
MALA
Synagoga

DIAGRAM of the CAUSES of MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 to MARCH 1856.



1.
APRIL 1854 to MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

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

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par des largeurs des zônes colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit 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 Davoust qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

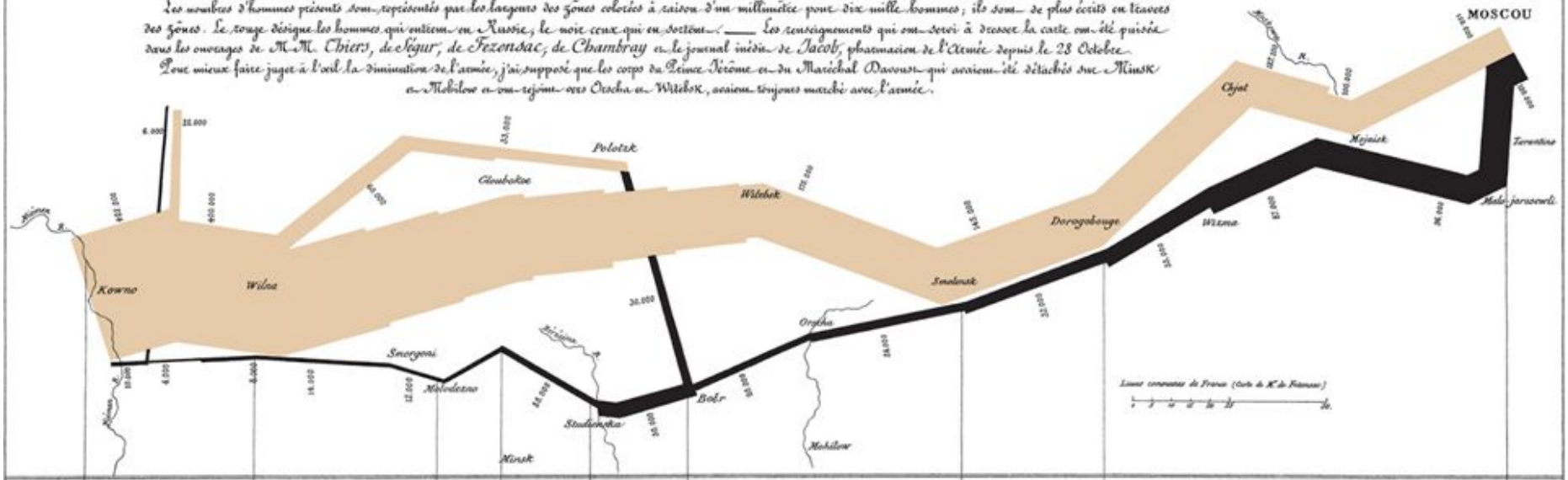
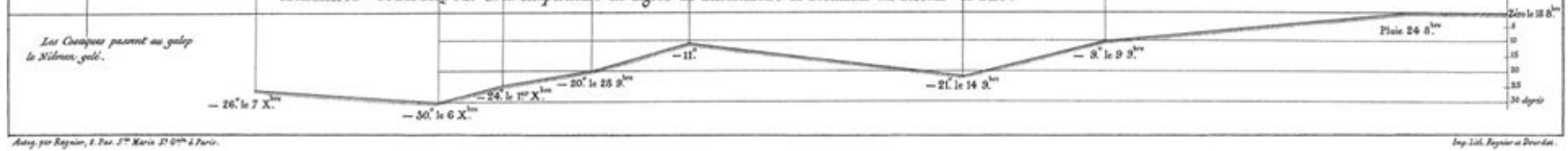
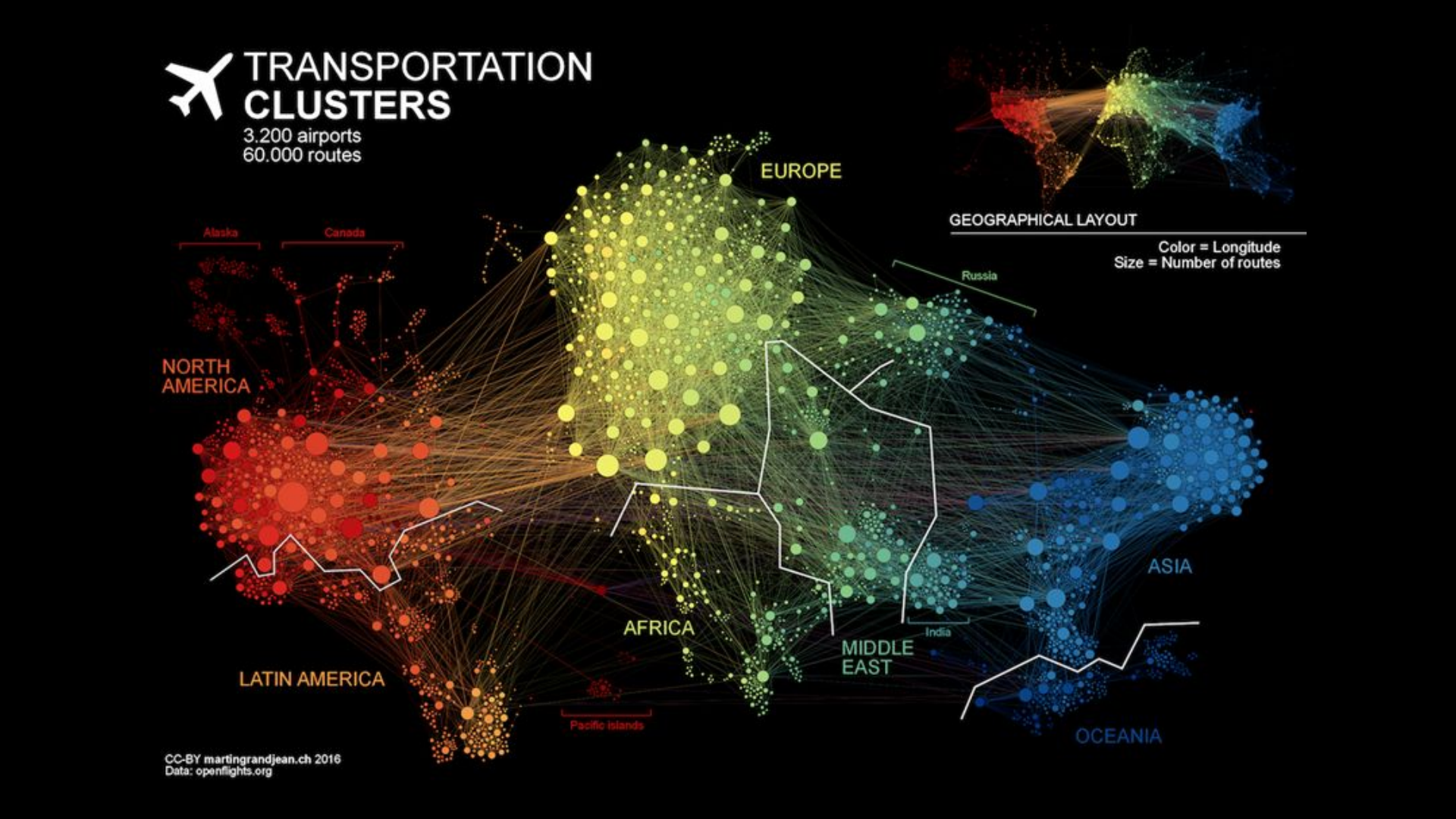


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

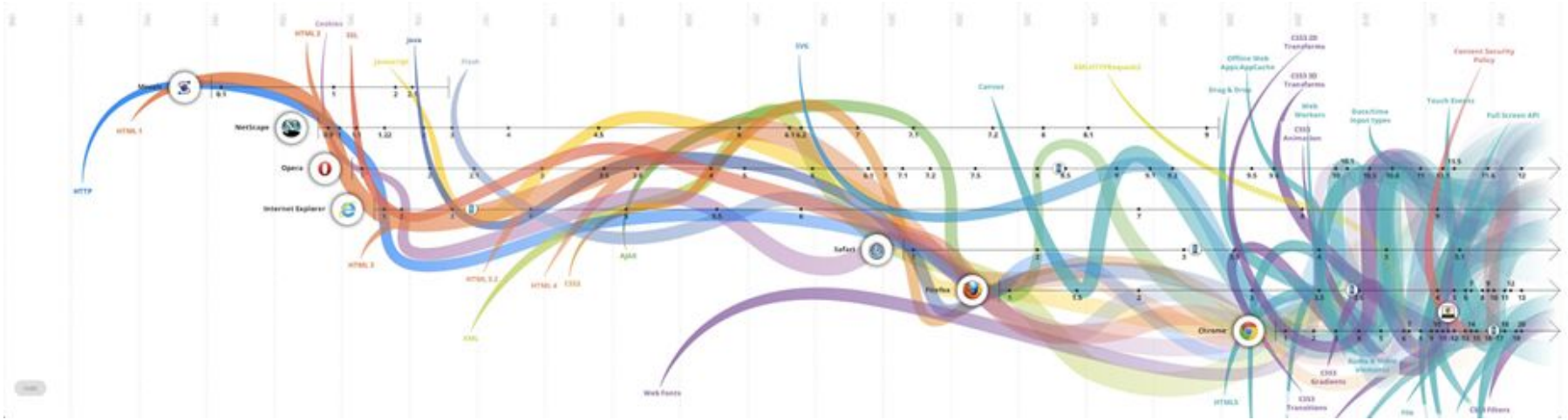


Autog. par Regnier, 8 Pas St.te Marie St Germ.n à Paris.

Imp. Lith. Regnier et Dourdan.


TRANSPORTATION
CLUSTERS
3.200 airports
60.000 routes
EUROPE
GEOGRAPHICAL LAYOUT
Color = Longitude
Size = Number of routes
Alaska
Canada
Russia
NORTH
AMERICA
ASIA
AFRICA
India
MIDDLE
EAST
LATIN AMERICA
Pacific Islands
OCEANIA
CC-BY martingrandjean.ch 2016
Data: openflights.org

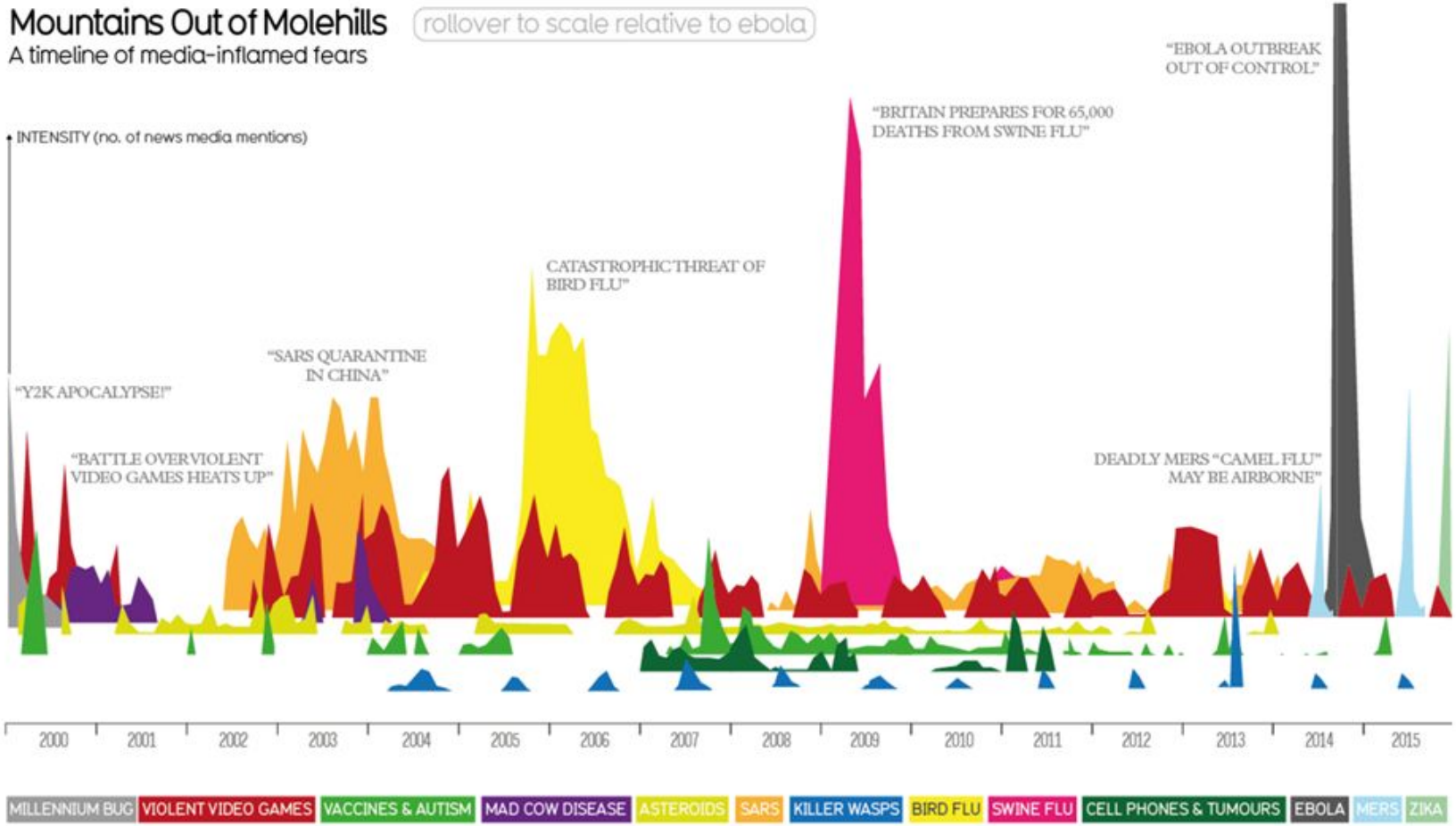
A Evolução da Web



Mountains Out of Molehills

rollover to scale relative to ebola

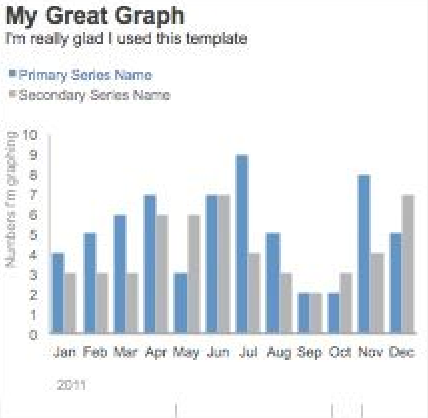
A timeline of media-inflamed fears

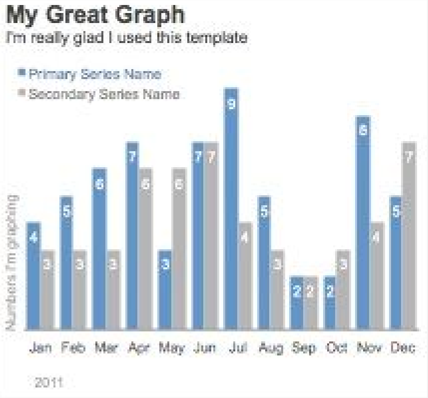


design & concept: David McCandless informationisbeautiful.net

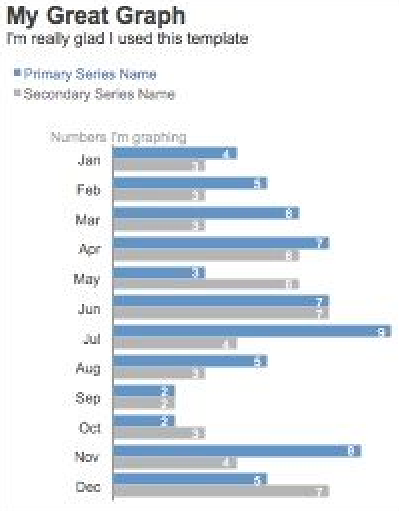
source: Google Trends, Google News Timeline // data retrieved 31st Dec 2015

STANDARD BAR CHART
I like bar charts because they are easy to read and people are familiar with them. Note that I've given you two versions here, one with a y-axis and one where the y-axis is ommitted and points are labeled directly. Choose the one that best fits your data and needs.

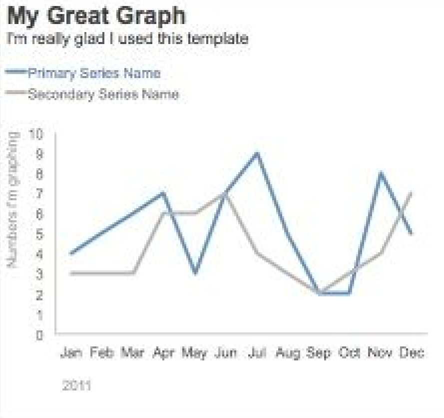


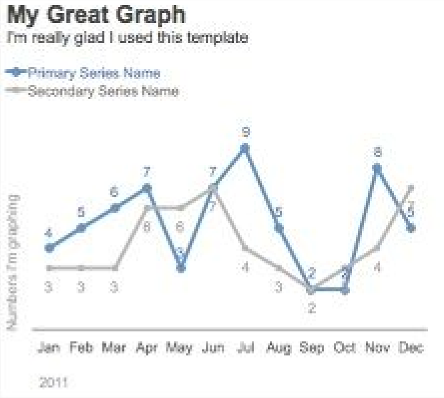


HORIZONTAL BAR CHART
Horizontal bar charts are great if your category names are long, because they are displayed from left to right, as most people read. Note that you will likely need to adjust the graph below for longer axis labels. You can do this by condensing the chart area to make more room for axis labels.



LINE CHART
Is your data continuous? It should be if you're using a line graph. I've provided two options below: the first has a y-axis; the second omits the y-axis and labels the points directly (this latter can get kind of messy, so consider labeling just important points, like the first and last or min and max, depending on your needs).





Já dizia Ben Shneiderman . . .

O propósito da visualização é insight, não imagens

(Ben Shneiderman)

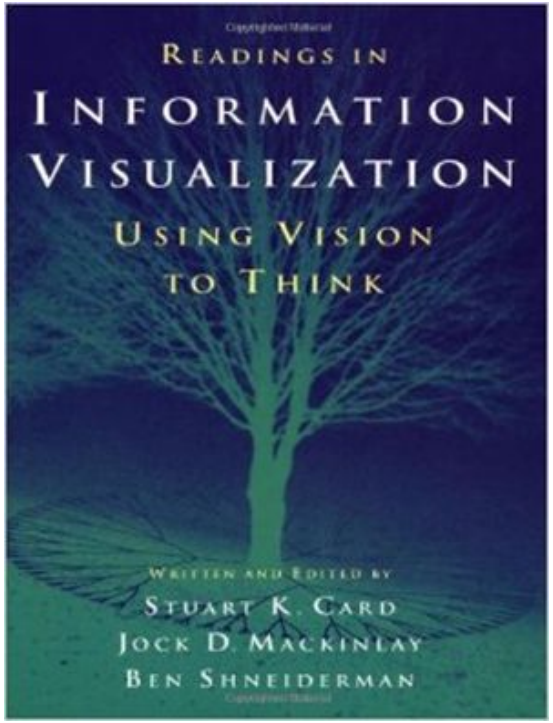
Visualização de Dados

- A área de Visualização de Dados envolve o estudo e a criação de representações visuais
- **Objetivo**: comunicar e permitir que padrões, tendências e anomalias sejam revelados em grandes conjuntos de dados

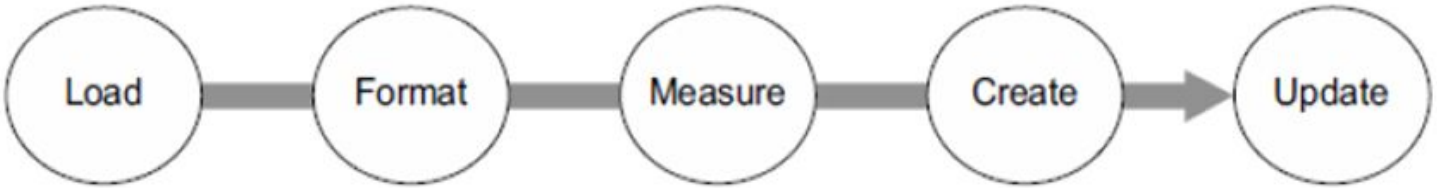
Visualização de Dados

- **Visualização de Informação**: dados abstratos
- **Visualização Científica:** dados físicos geralmente provenientes de uma medição

Visualização de Informação é o uso de representações visuais interativas apoiado por computador de dados abstratos para ampliar a cognição



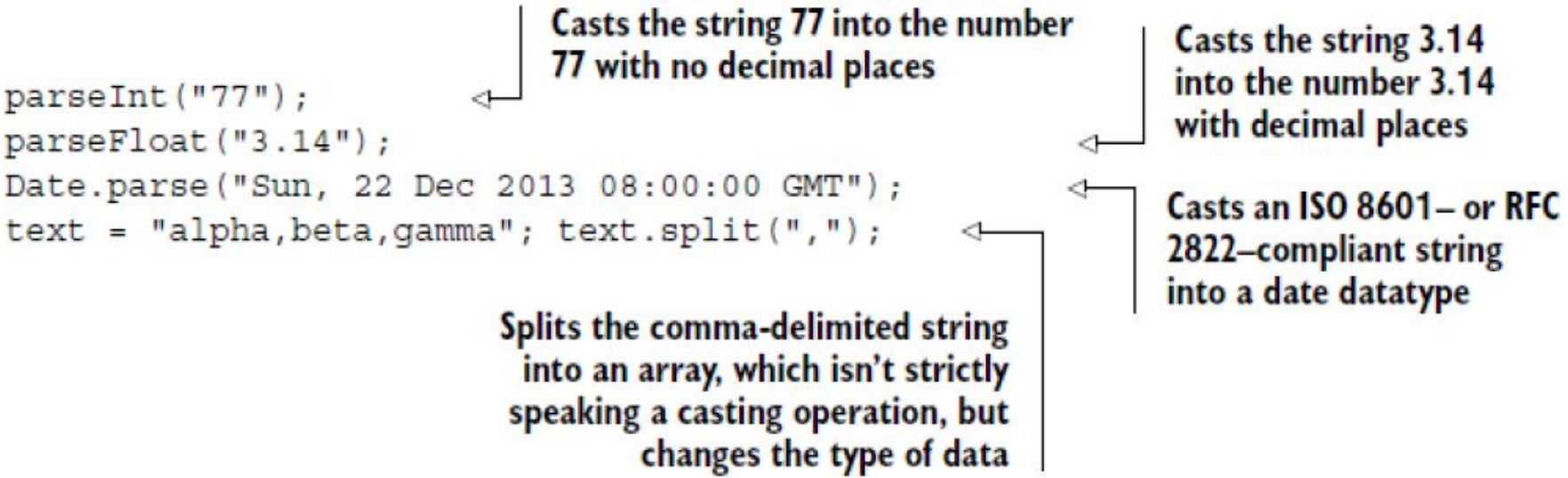
Processo de Visualização



Carregar

```
d3.text();

d3.xml();

d3.csv();

d3.json();

d3.html();
```

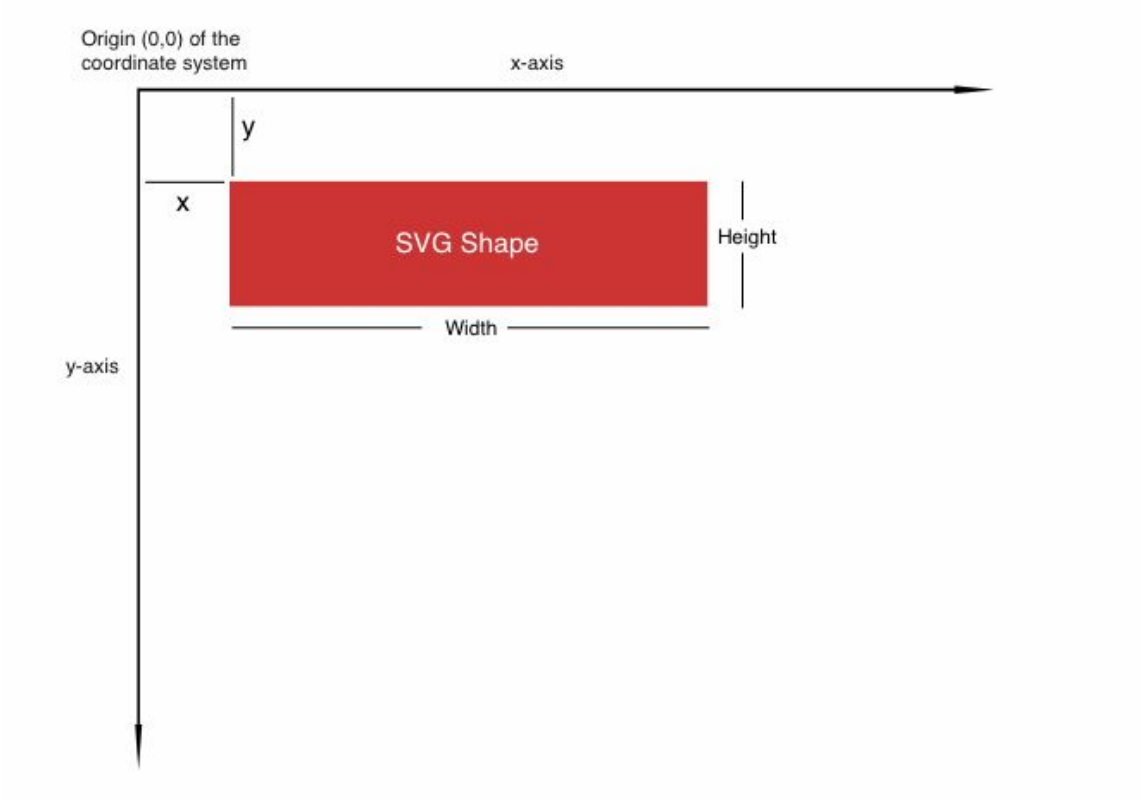
Formatar


parseInt("77");
parseFloat("3.14");
Date.parse("Sun, 22 Dec 2013 08:00:00 GMT");
text = "alpha,beta,gamma"; text.split(",");
Casts the string 77 into the number 77 with no decimal places
Casts the string 3.14 into the number 3.14 with decimal places
Casts an ISO 8601– or RFC 2822–compliant string into a date datatype
Splits the comma-delimited string into an array, which isn't strictly speaking a casting operation, but changes the type of data

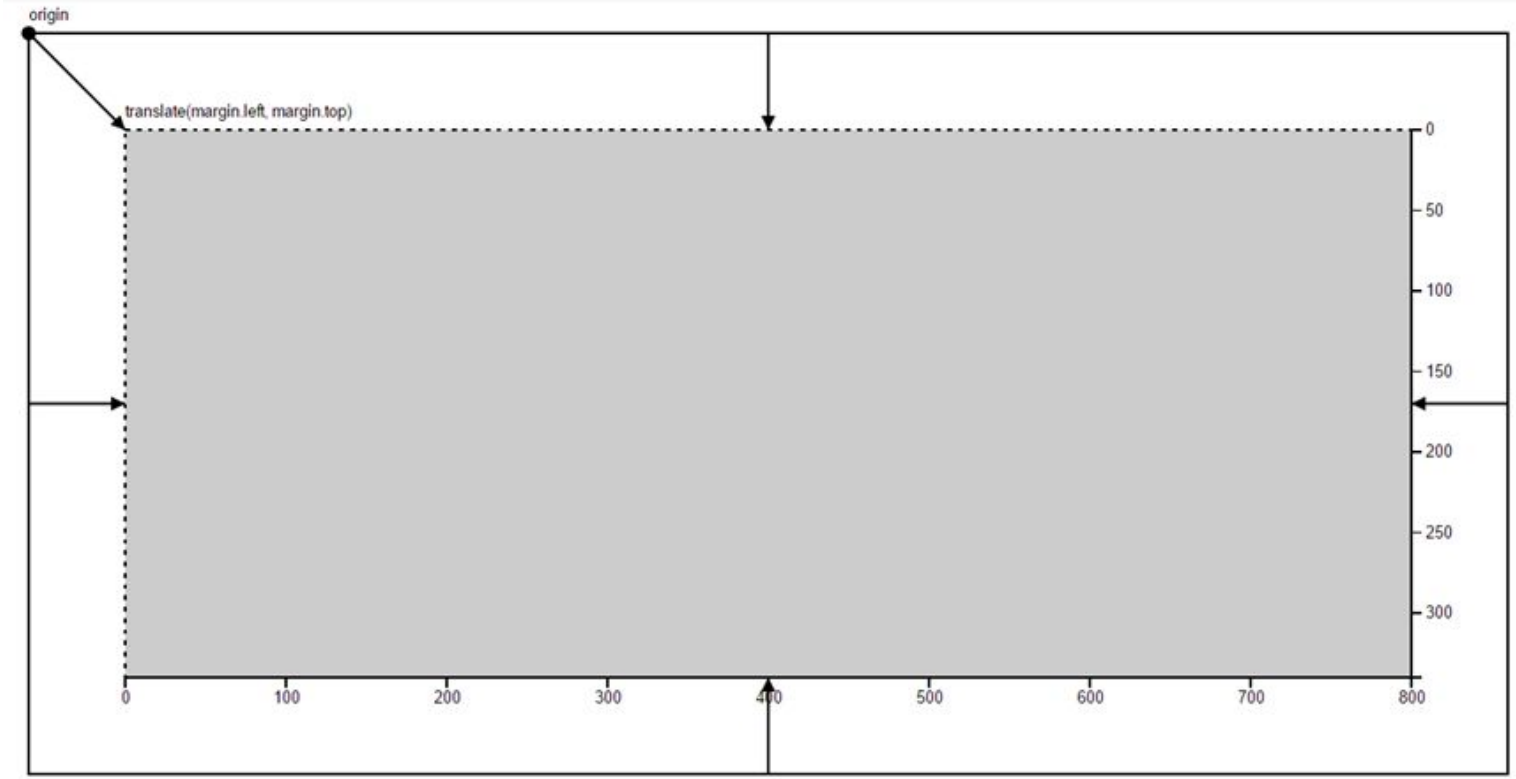
Medir

```
d3.min();

d3.max();

d3.mean();

d3.median();

d3.deviation();
```

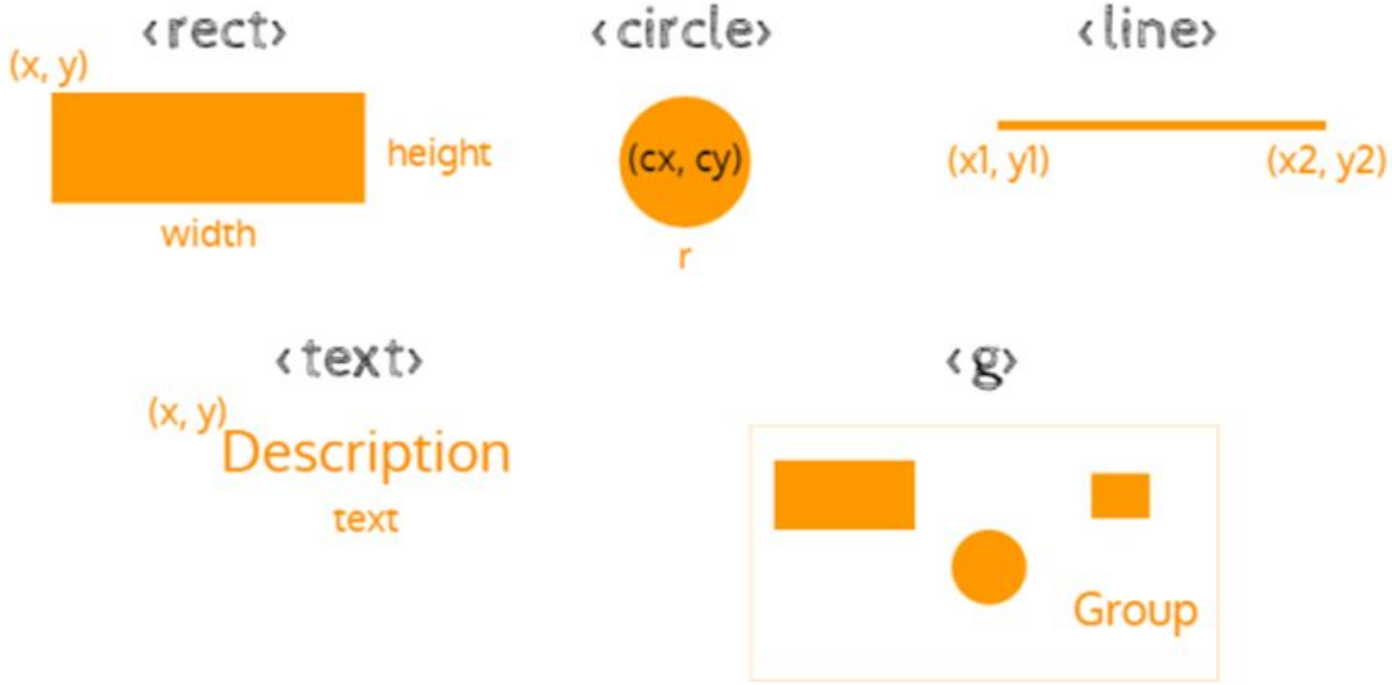
Coordenadas SVG



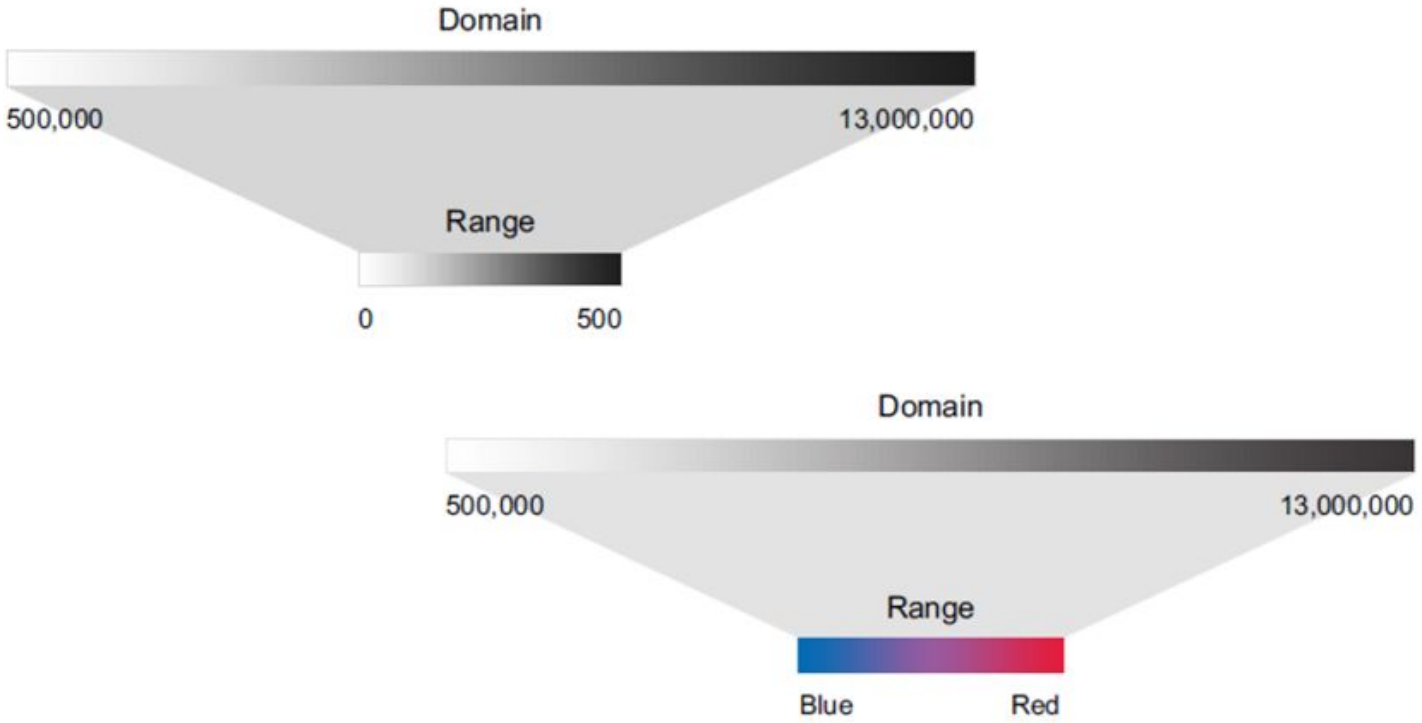
Convenção de Margens



Alguns Elementos Básicos



Escalas


Domain
500,000
13,000,000
Range
0
500
Domain
500,000
13,000,000
Range
Blue
Red

Círculos Representando Dados

- Quando usar círculos para representar dados, dimensione pela área ao invés do diâmetro, raio ou circunferência
- Área do Círculo = $\pi\ r^2$
- Isolando o raio:

$$r = \sqrt{(area\ \ do\ \ círculo\ /\ \pi)}$$

Padrão de Atualização