Tenets of good data visualisation + introduction to Observable

Gurman Bhatia
@chartwaali
Founder, Revisual Labs

Barnard College · Columbia University
October 6, 2023

revisual labs

Hello!

I am an ***information designer, data visualisation developer and former journalist*** based in New Delhi, India.

Currently, I am building India's first information design agency - Revisual Labs.

rv.l

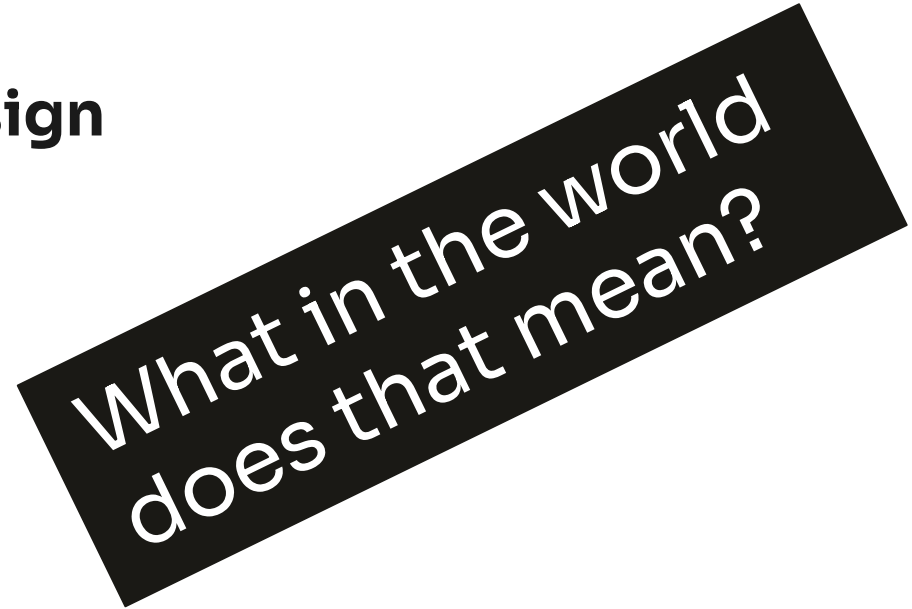
I tell stories

with **data, code** and **design**

rv.l

I tell stories

with **data, code** and **design**



Sometimes I take a lackluster chart in a research paper

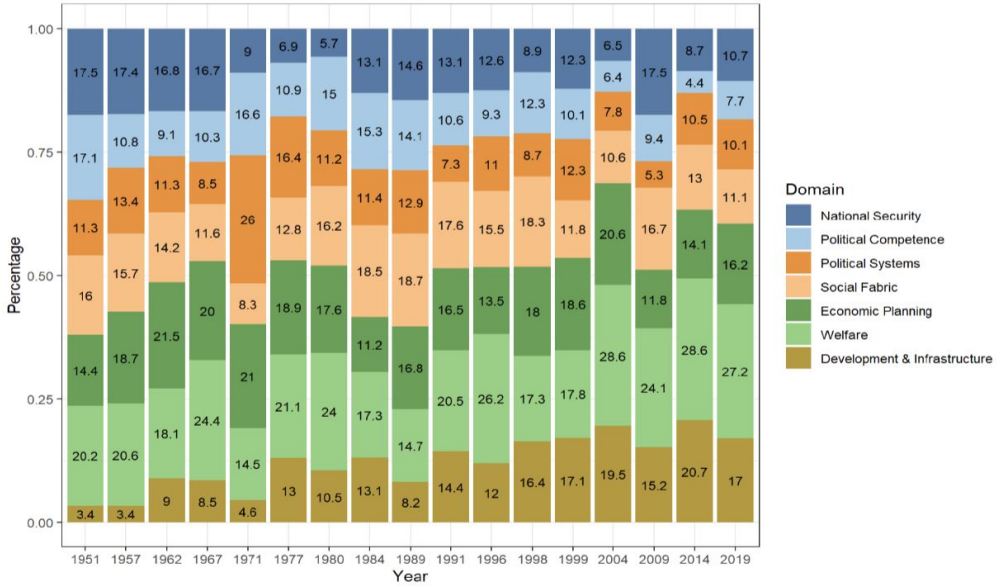
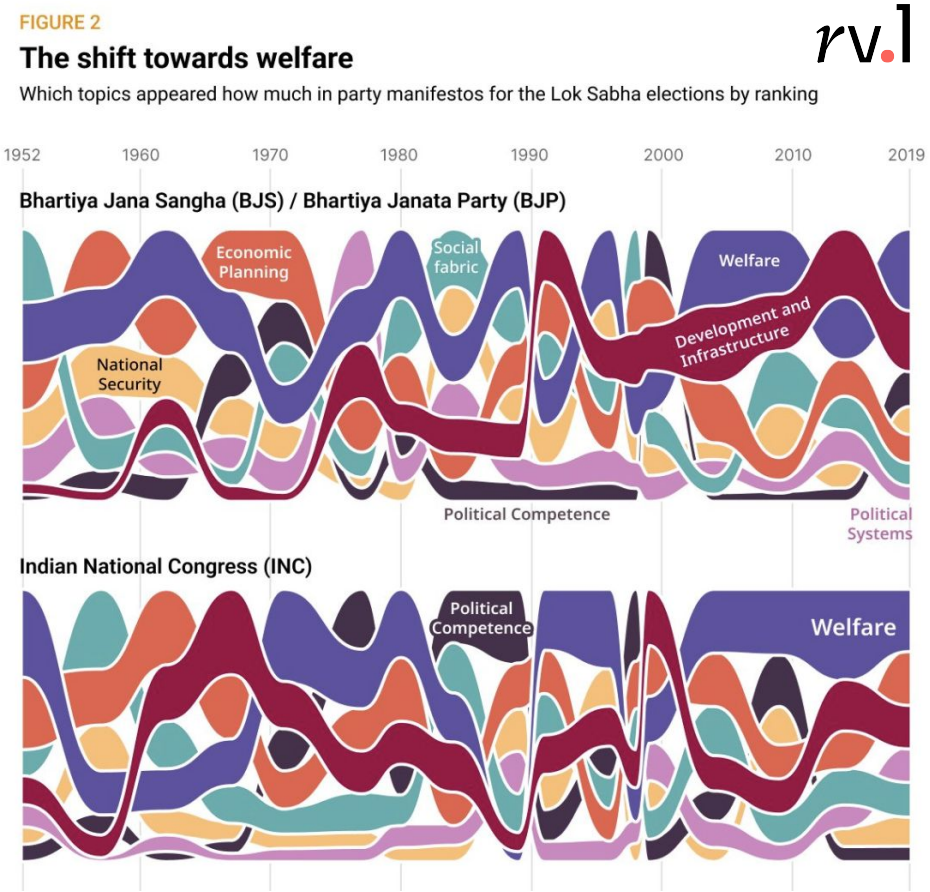


Figure 2: Distribution of Various Domains Across Manifestos

And turn it into something with an insight



Or sometimes I take a press release...

Press Release

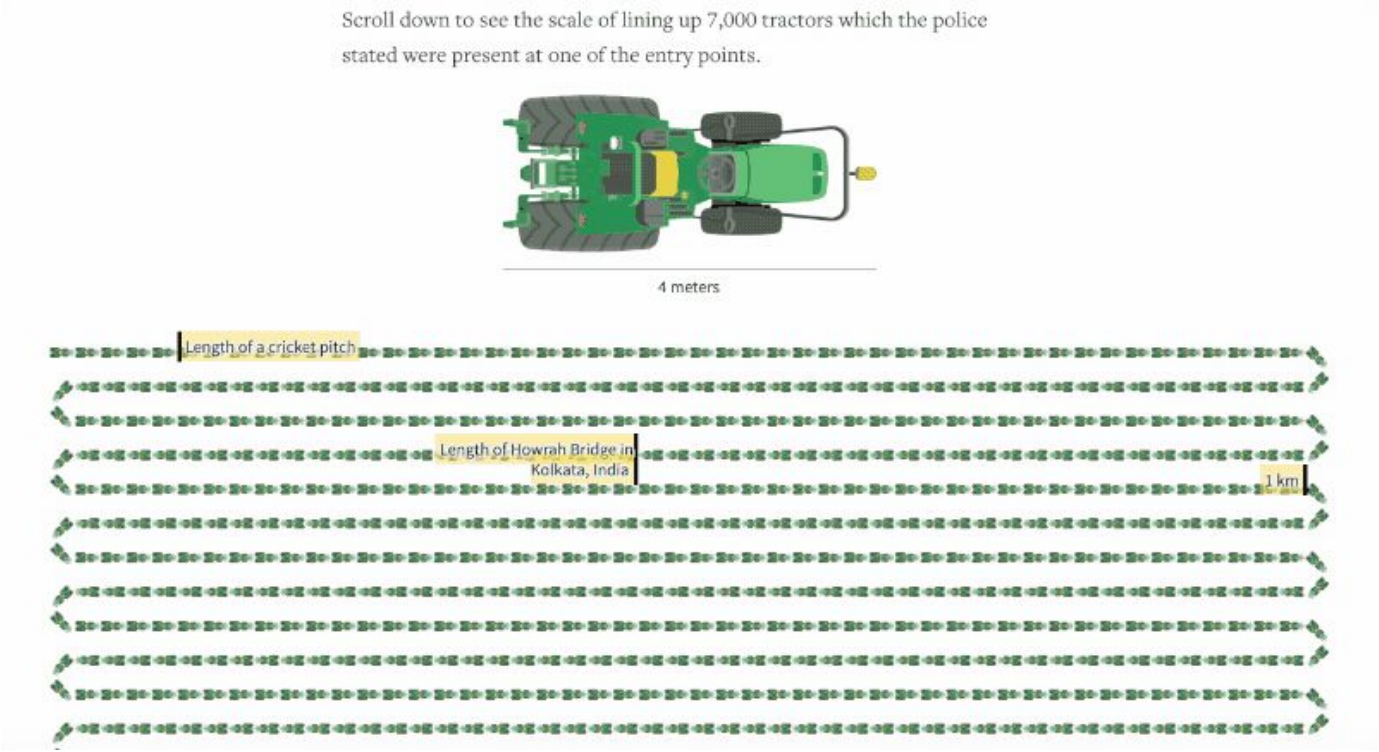
Sayunkt Kisan Morcha comprising of various Kisan organisations proposed a Kisan Tractor Rally on the occasion of the Republic Day on 26.01.2021. The Sayunkt Kisan Morcha had conducted several rounds of meetings with the Delhi Police in connection with their proposed tractor rally. They had given an undertaking to the Delhi Police to take out a peaceful rally as per the proposed plans on 04 agreed upon routes.

On 26.01.2021 at about 08.30 AM about 6000-7000 tractors had assembled at Singhu borders and they were to reach upto Sanjay Gandhi Transport Nagar and take right turn. Instead of the agreed upon route, they insisted upon going to the Central Delhi and despite persuasion by Delhi Police, the farmers led by Nihangs on their horses fully equipped with deadly weapons like swords, kripans and fursas charged the police and broke the several layers of barricades, which were erected between Mukarba Chowk and Transport Nagar.

And write some code...

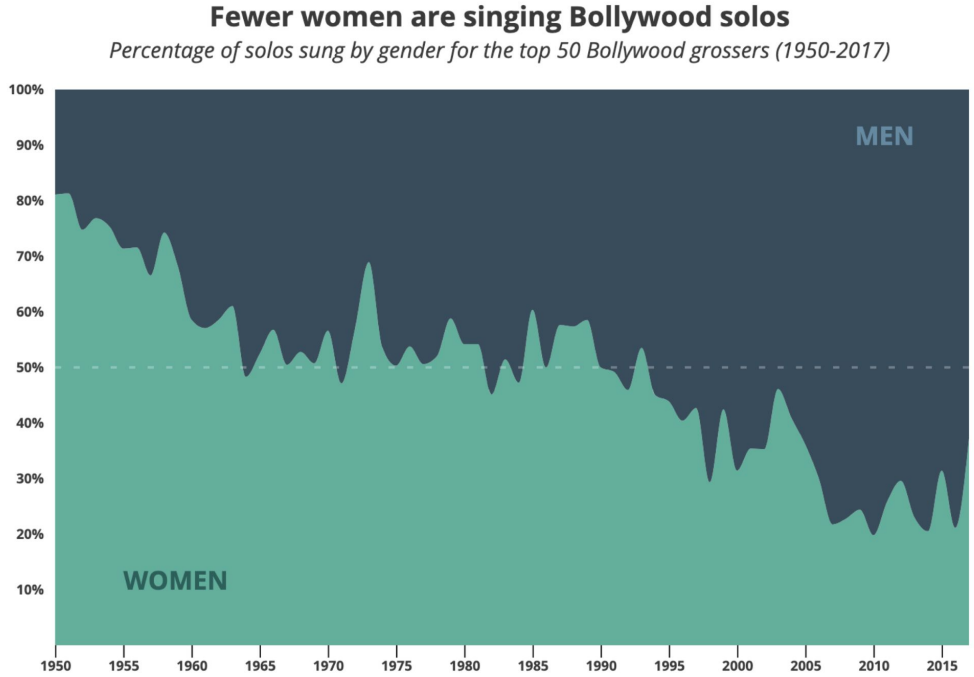
```
d3.select('#app')
  .selectAll('.tractor')
  .data(dataAppend)
  .enter()
  .append('img')
  .attr('class', (d, i) => {
    let iNum;
    if (i % double == double-1 && Math.floor(i / perRow) % 2 === 1) {
      iNum = i - (perRow - 1);
    } else {
      if (i >= perRow && Math.floor(i / perRow) % 2 === 1) {
        const inHun = (i + 1) % perRow;
        if (inHun <= half && inHun >= 0) {
          const o = perRow * 2 - ((i + 1) % double);
          iNum = i + 1 + o - (i % half);
        } else {
          const o = ((i + 1) % double) - perRow * 2;
          iNum = i + 1 - o - (i % half) - half;
        }
      } else {
        iNum = i + 1;
      }
    }
    // iNum = i + 1;
    if (i % perRow === 0 && i !== 0) {
      if ((i / perRow) % 2 == 1) {
        return 'tractor down-left-angle top-margin i-' + iNum;
      } else {
        return 'tractor down-right-angle bottom-margin i-' + iNum;
      }
```

To make a looooong line of tractors



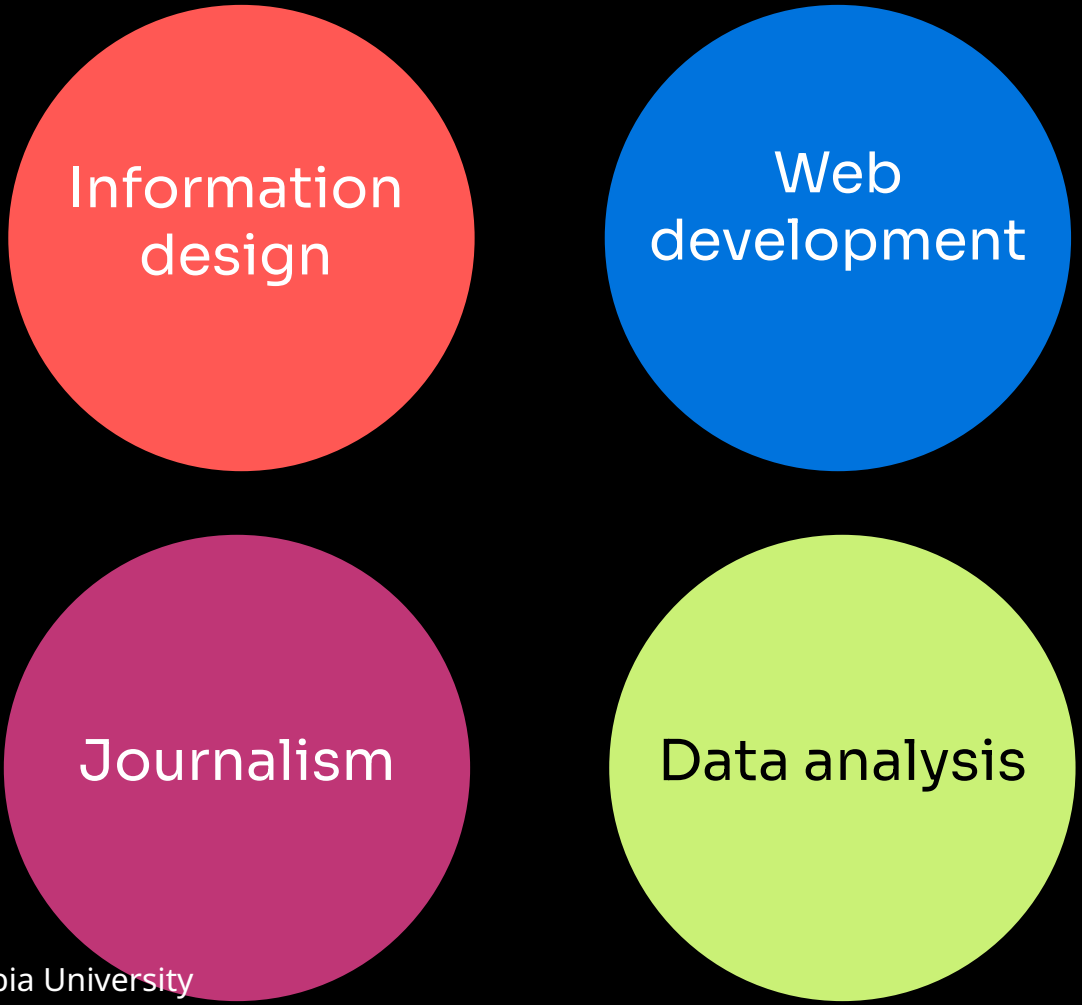
Or sometimes I analyse a database of songs...

To show how the female solo has declined in Bollywood...



The core skills of what I do

rv.1

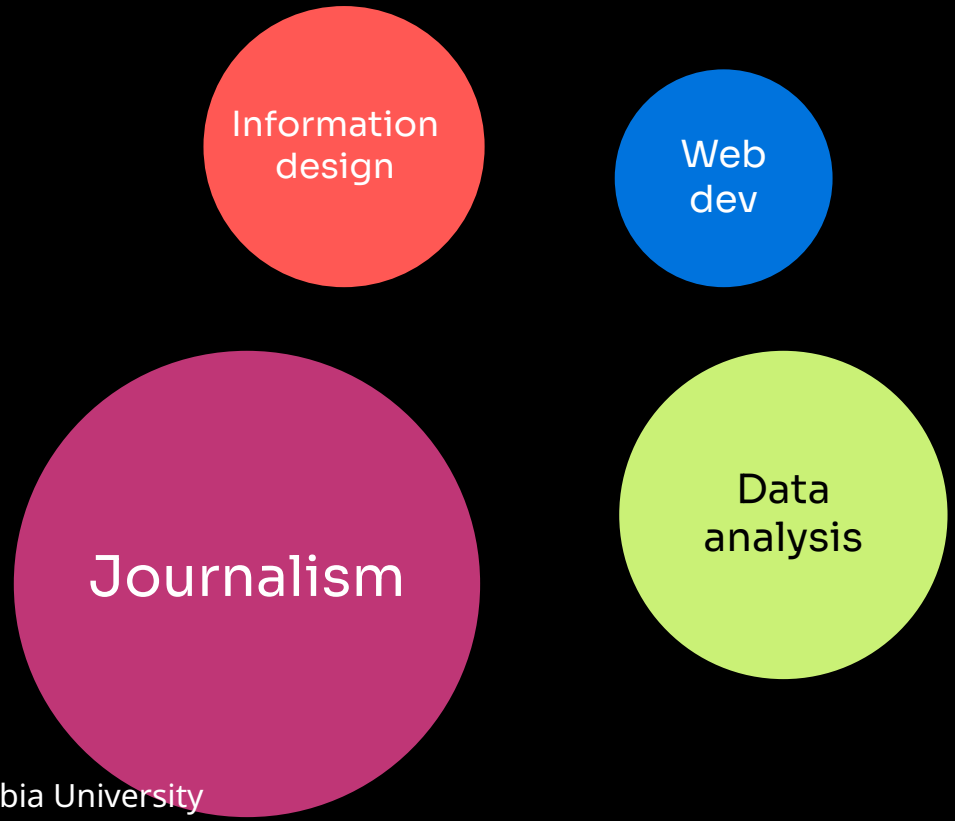


Barnard College · Columbia University

www.revisual.co

My work
8 years ago

rv.l

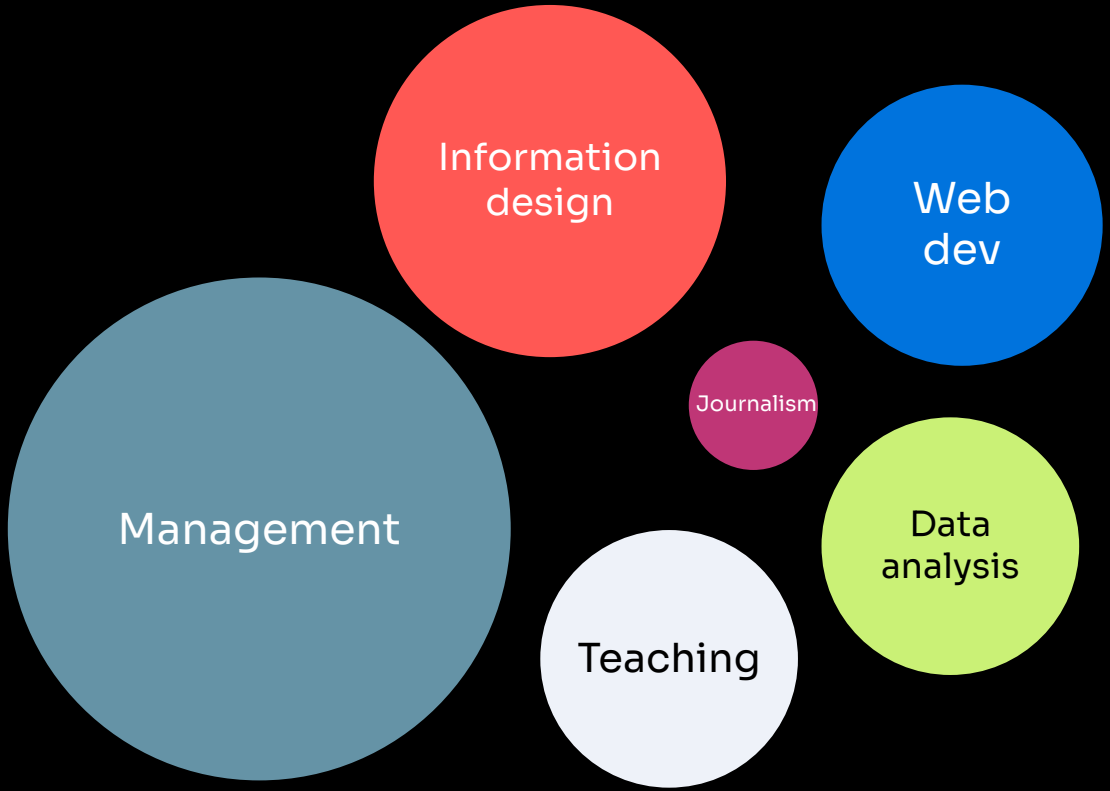


Barnard College · Columbia University

www.revisual.co

My work today

rv.l



Barnard College · Columbia University

www.revisual.co

Thinking about data visualisation

Good tools, properly used, can complement and enhance our thinking skills ***but never replace them.***

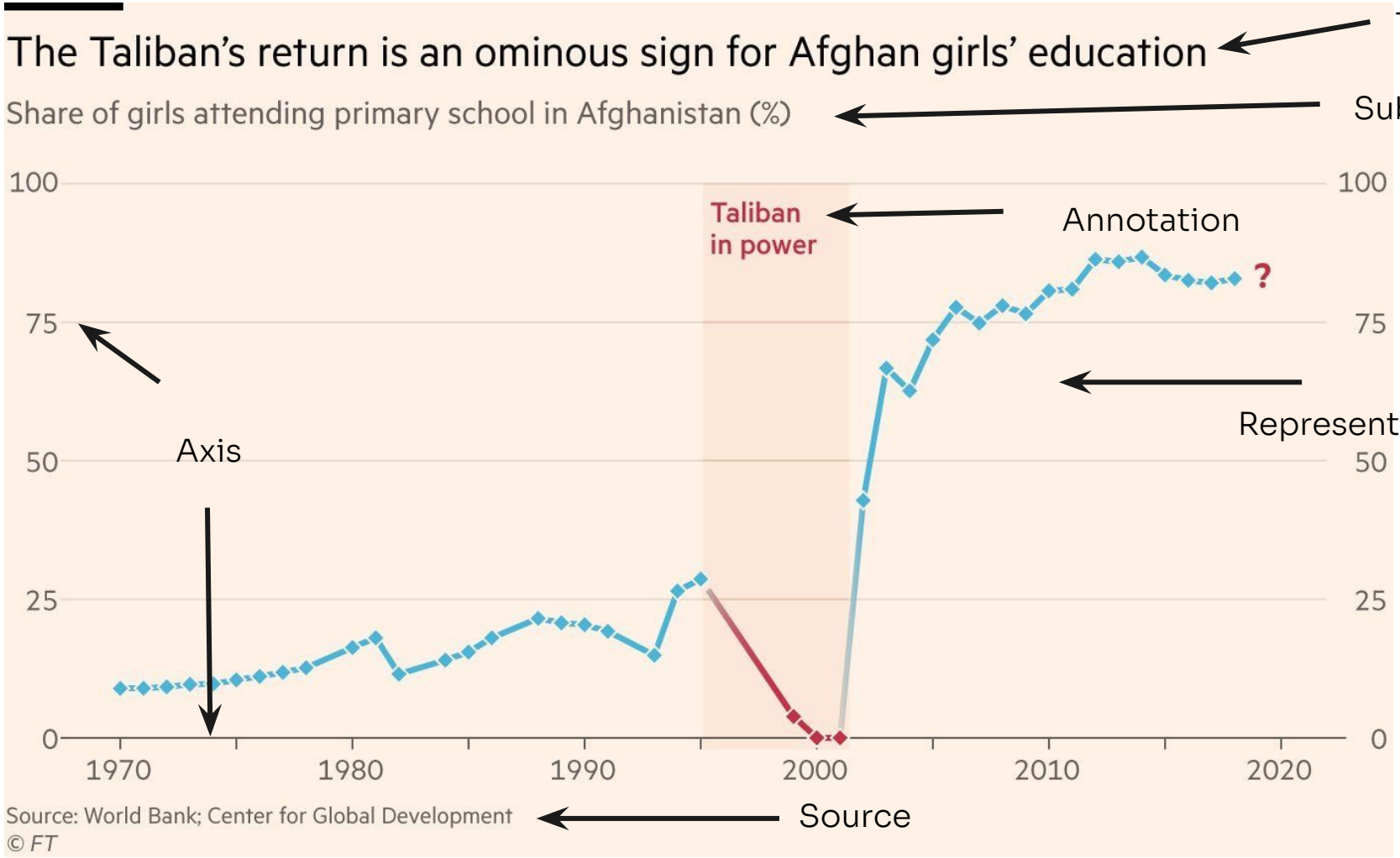
\- Stephen Few

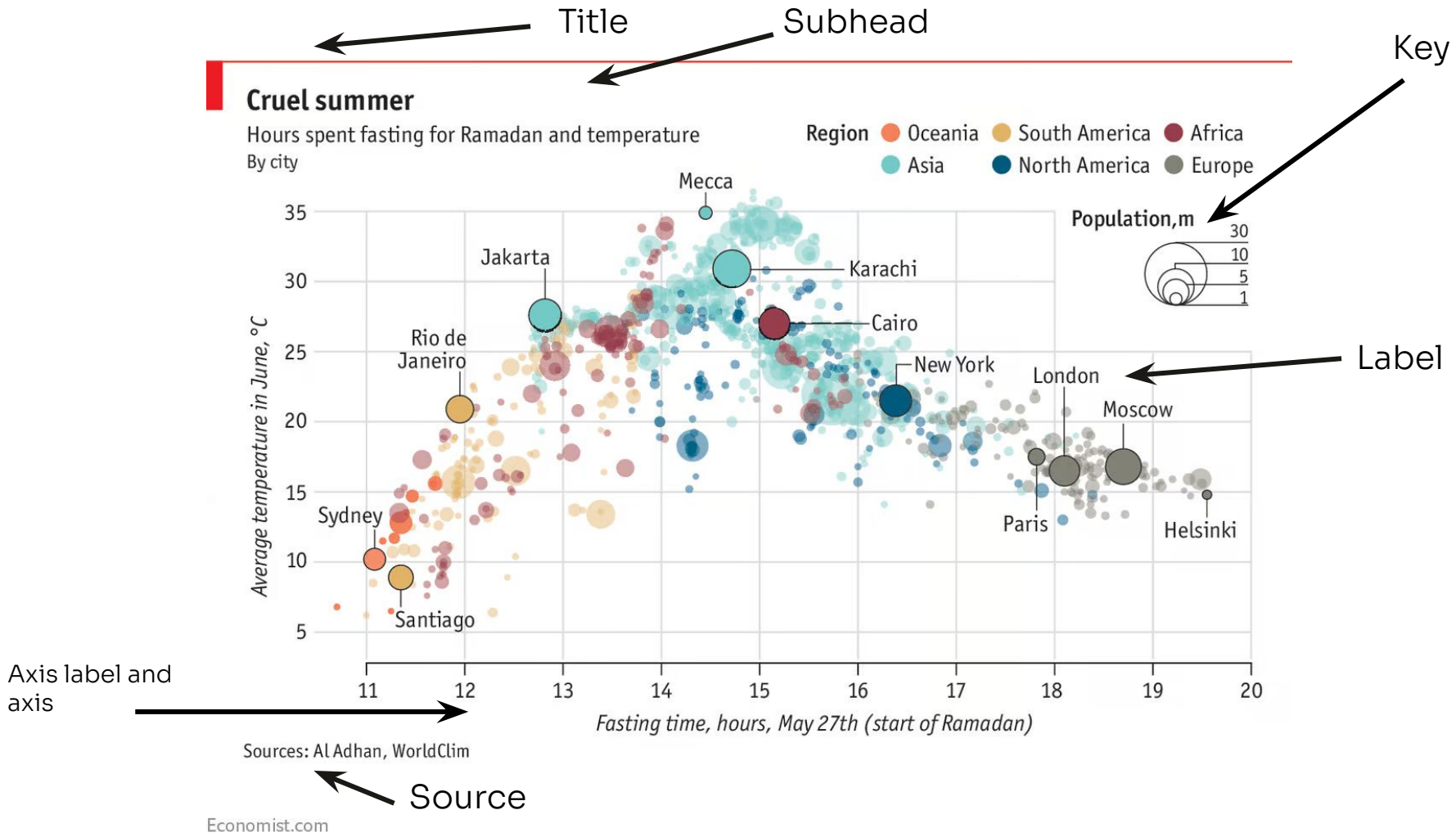
What makes a good chart?

Four questions to think when making a chart:

1. How can I make my chart **trustworthy**?
2. How can I make my chart **accessible**?
3. How can I make my chart **elegant**?
4. How can my viz **tell a story**?

Adapted from Andy Kirk's "Data Visualization"


The Taliban's return is an ominous sign for Afghan girls' education
Share of girls attending primary school in Afghanistan (%)
100
75
50
25
0
Taliban
in power
?
1970
1980
1990
2000
2010
2020
Source: World Bank; Center for Global Development
© FT
Title
Subhead
Annotation
Axis
Data
Representation
Source


Title
Subhead
Key
Label
Axis label and axis
Source
Cruel summer
Hours spent fasting for Ramadan and temperature
By city
Region
Oceania
South America
Africa
Asia
North America
Europe
Population, m
30
10
5
1
Mecca
Jakarta
Karachi
Cairo
Rio de Janeiro
New York
London
Moscow
Sydney
Paris
Helsinki
Santiago
Average temperature in June, °C
35
30
25
20
15
10
5
11
12
13
14
15
16
17
18
19
20
Fasting time, hours, May 27th (start of Ramadan)
Sources: Al Adhan, WorldClim
Economist.com

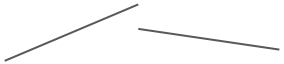
The chart should, on its own, tell the story

But we do it with Marks and Channels

Marks

Items or links

Points



Lines

Areas

Connections

From Tamara Munzner's Visualization Analysis and Design

But we do it with Marks and Channels

Channels

How marks look based on the data

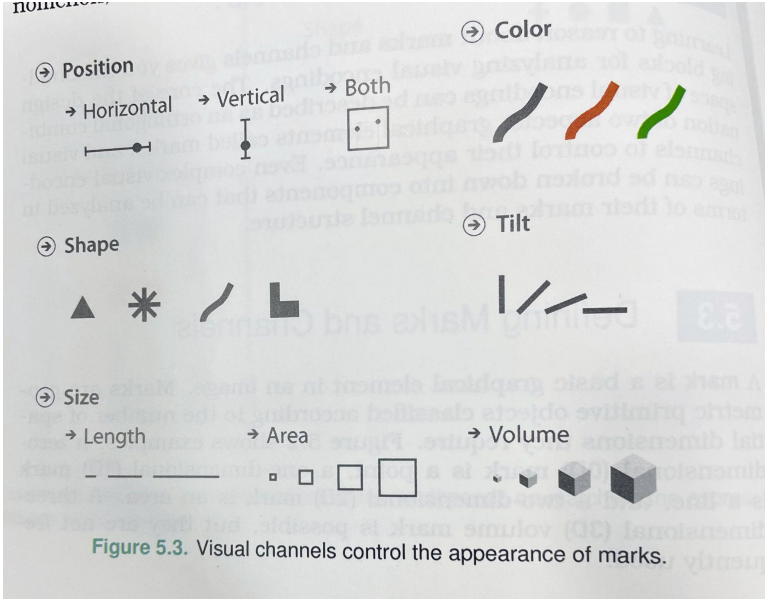


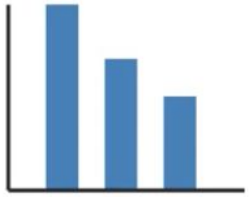
Figure 5.3. Visual channels control the appearance of marks.

From Tamara Munzner's Visualization Analysis and Design

But we do it with Marks and Channels

Visual encoding

- analyze idiom structure as combination of marks and channels

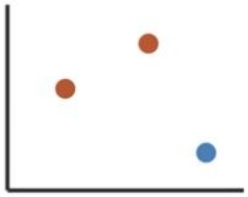


1:
vertical position

mark: line

2:
vertical position
horizontal position

mark: point



3:
vertical position
horizontal position
color hue

mark: point

4:
vertical position
horizontal position
color hue
size (area)

mark: point

What is data encoding?

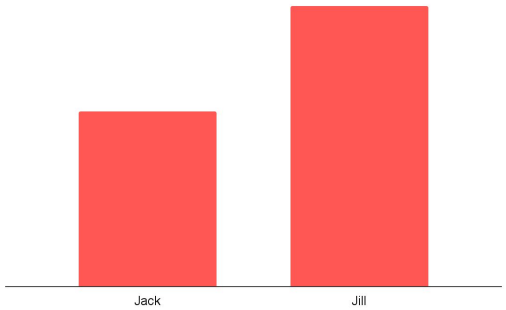
Taking

Data

Name	Age
Jack	25
Jill	50

And making it

Visual



Some types of visual encodings

Length

Position

Slope

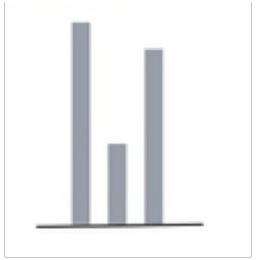
Angle

Colour

Area

Some types of visual encodings

Length



Position

Slope

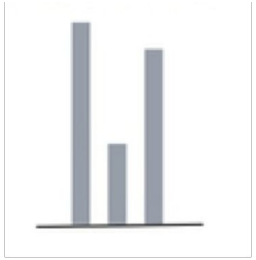
Angle

Colour

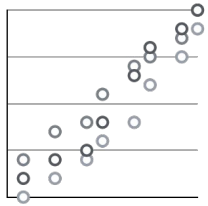
Area

Some types of visual encodings

Length



Position



Slope

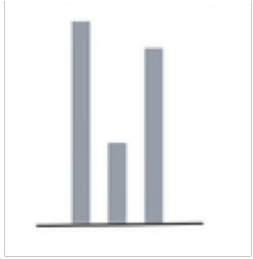
Angle

Colour

Area

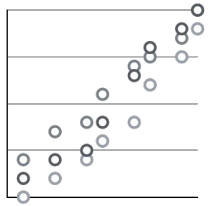
Some types of visual encodings

Length



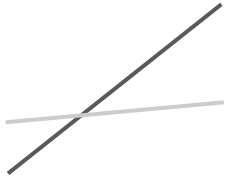
Angle

Position



Colour

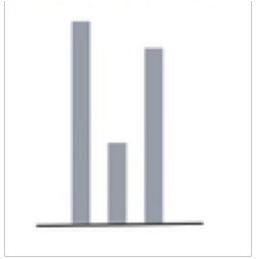
Slope



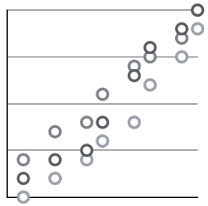
Area

Some types of visual encodings

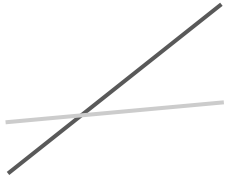
Length



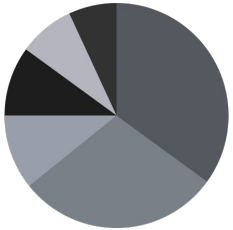
Position



Slope



Angle

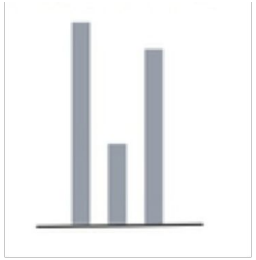


Colour

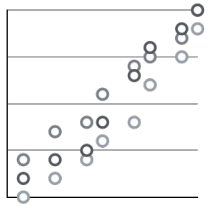
Area

Some types of visual encodings

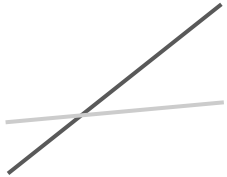
Length



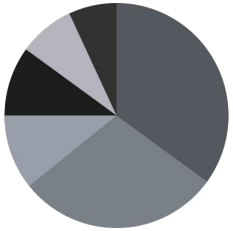
Position



Slope



Angle

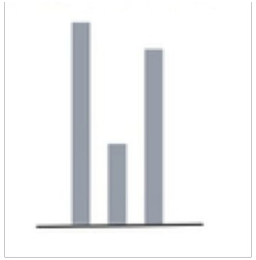


Colour

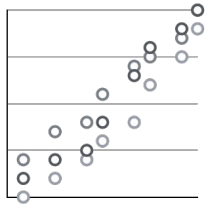
Area

Some types of visual encodings

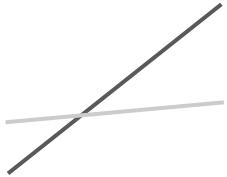
Length



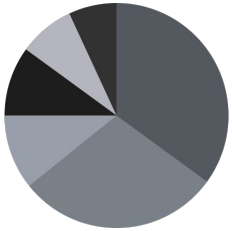
Position



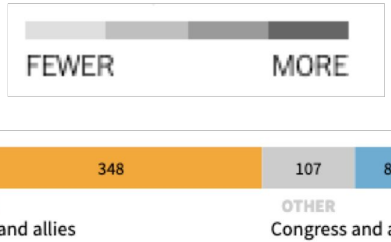
Slope



Angle

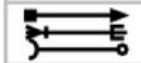


Colour

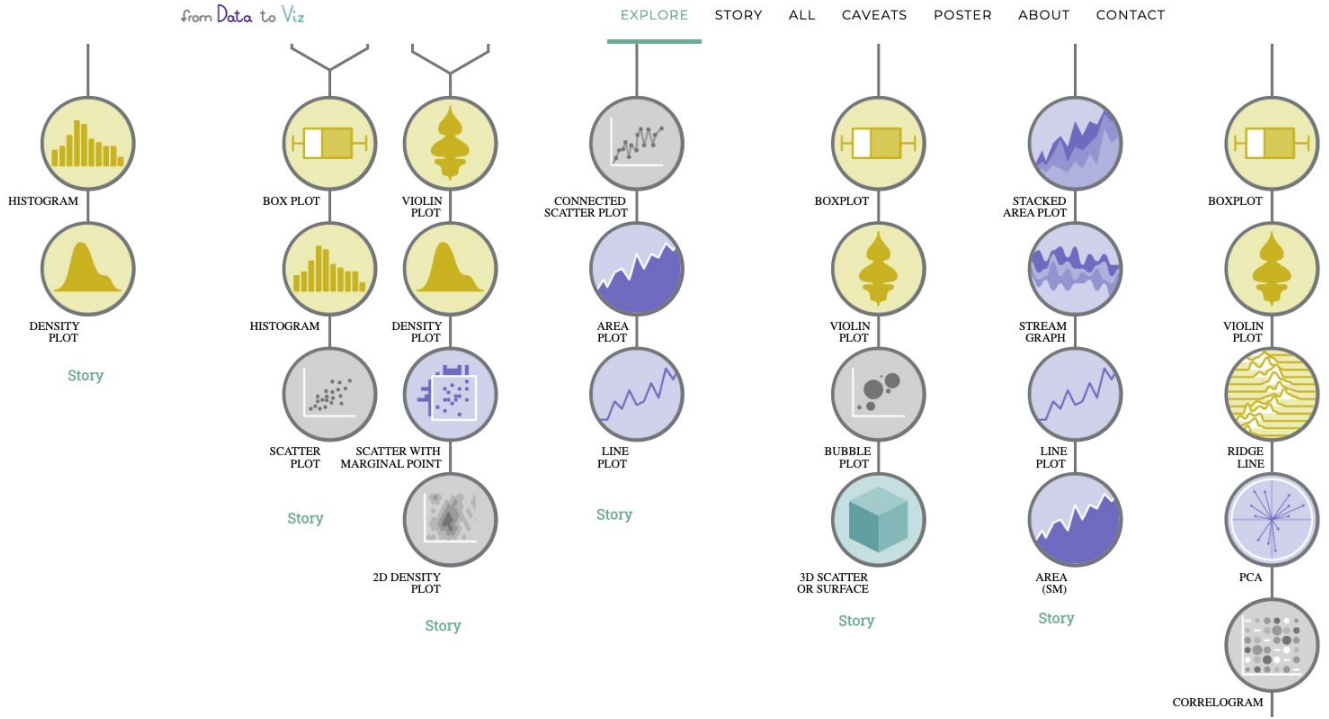


Area

Many more

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

Different chart types

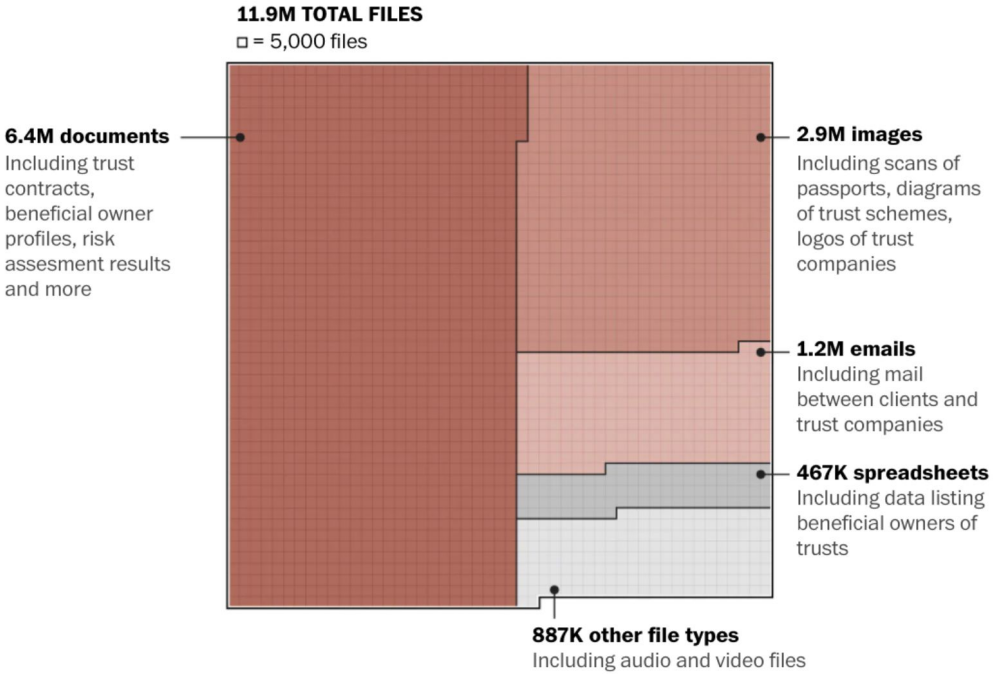


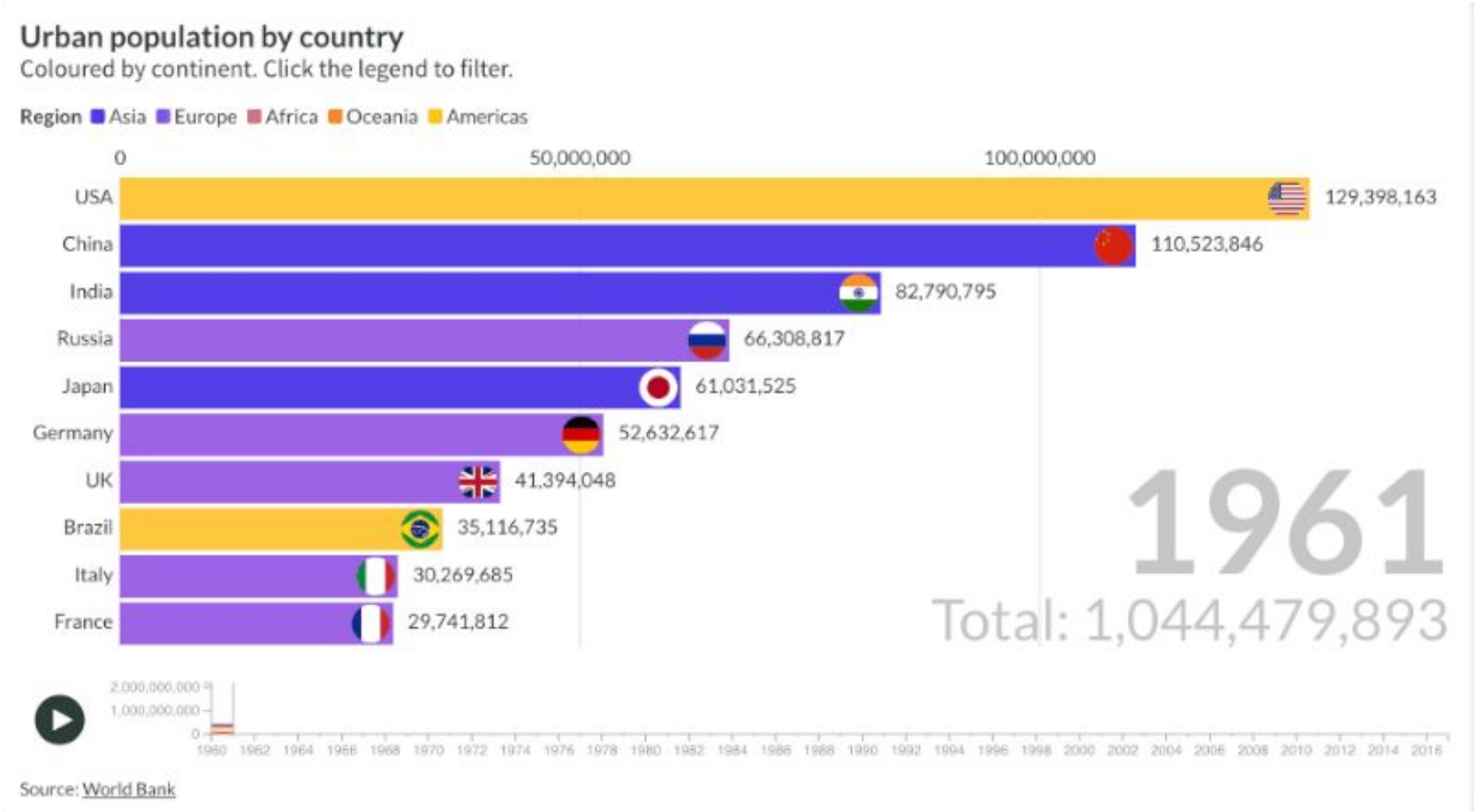
Exercise #1

Identify what ***variables*** are being encoded using which ***encoding***

Types of files included in the Pandora Papers investigation

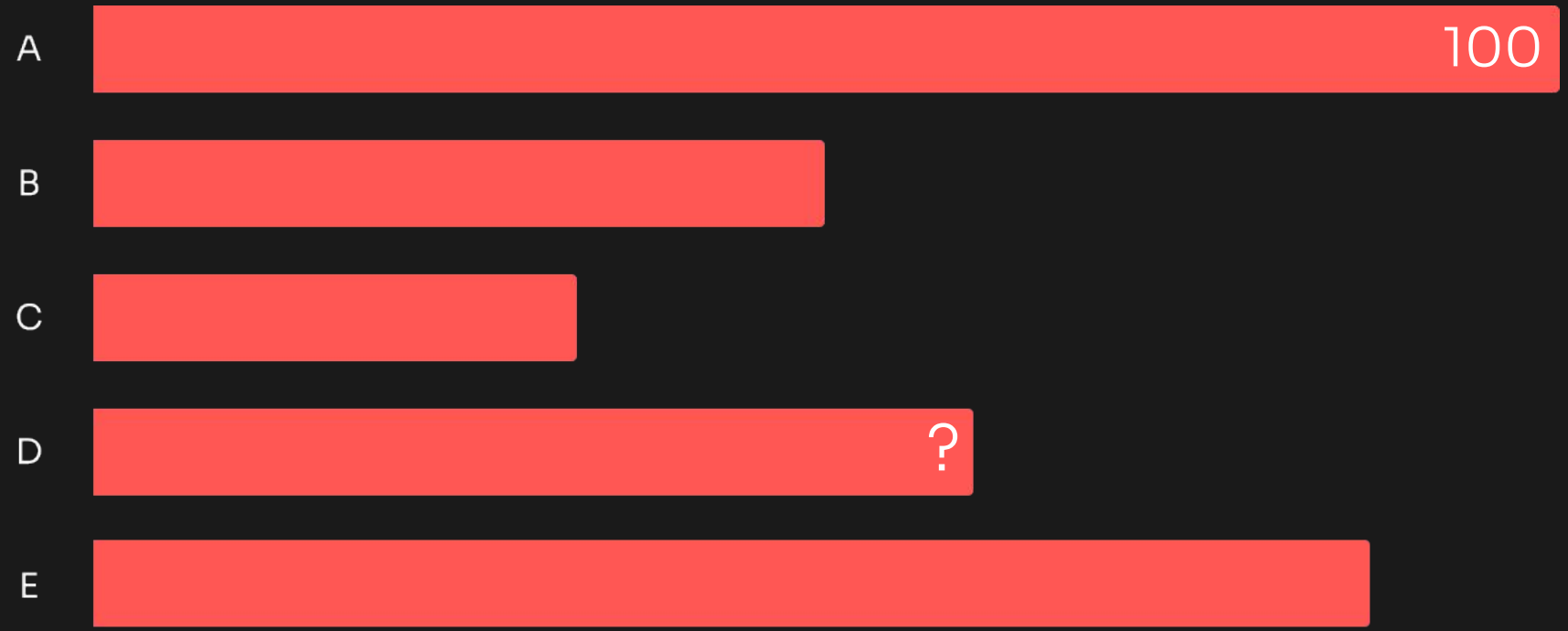
The trove of financial records unlocks otherwise impenetrable financial schemes and identifies the individuals behind them.



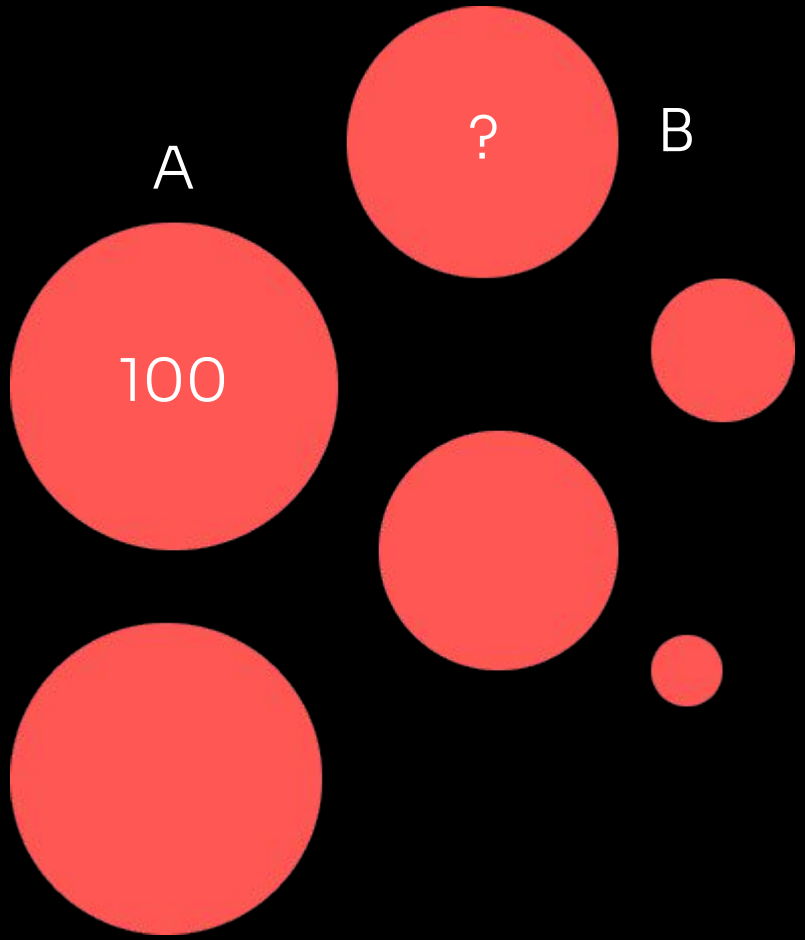

Urban population by country
Coloured by continent. Click the legend to filter.
Region Asia Europe Africa Oceania Americas
0
50,000,000
100,000,000
USA
129,398,163
China
110,523,846
India
82,790,795
Russia
66,308,817
Japan
61,031,525
Germany
52,632,617
UK
41,394,048
Brazil
35,116,735
Italy
30,269,685
France
29,741,812
1961
Total: 1,044,479,893
2,000,000,000
1,000,000,000
0
1960 1962 1964 1966 1968 1970 1972 1974 1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016
Source: World Bank

BBC FOUR
lifespan
75
50
25
927
income
$400
$4 000
$40 000

What is the value of D?



What is the value of B?



Channels: Rankings

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Color luminance, Color saturation: Same

Curvature, Volume (3D size): Same

Best

Effectiveness

Least

Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

- **expressiveness**
  - match channel and data characteristics
- **effectiveness**
  - channels differ in accuracy of perception

1 dataset 100 visualizations

Can we come up with 100 visualizations from one simple dataset?

As an information design agency working with data visualization every day, we challenged ourselves to accomplish this using insightful and visually appealing visualizations.

We wanted to show the diversity and complexity of data visualization and how we can tell different stories using limited visual properties and assets.

Number of World Heritage Sites

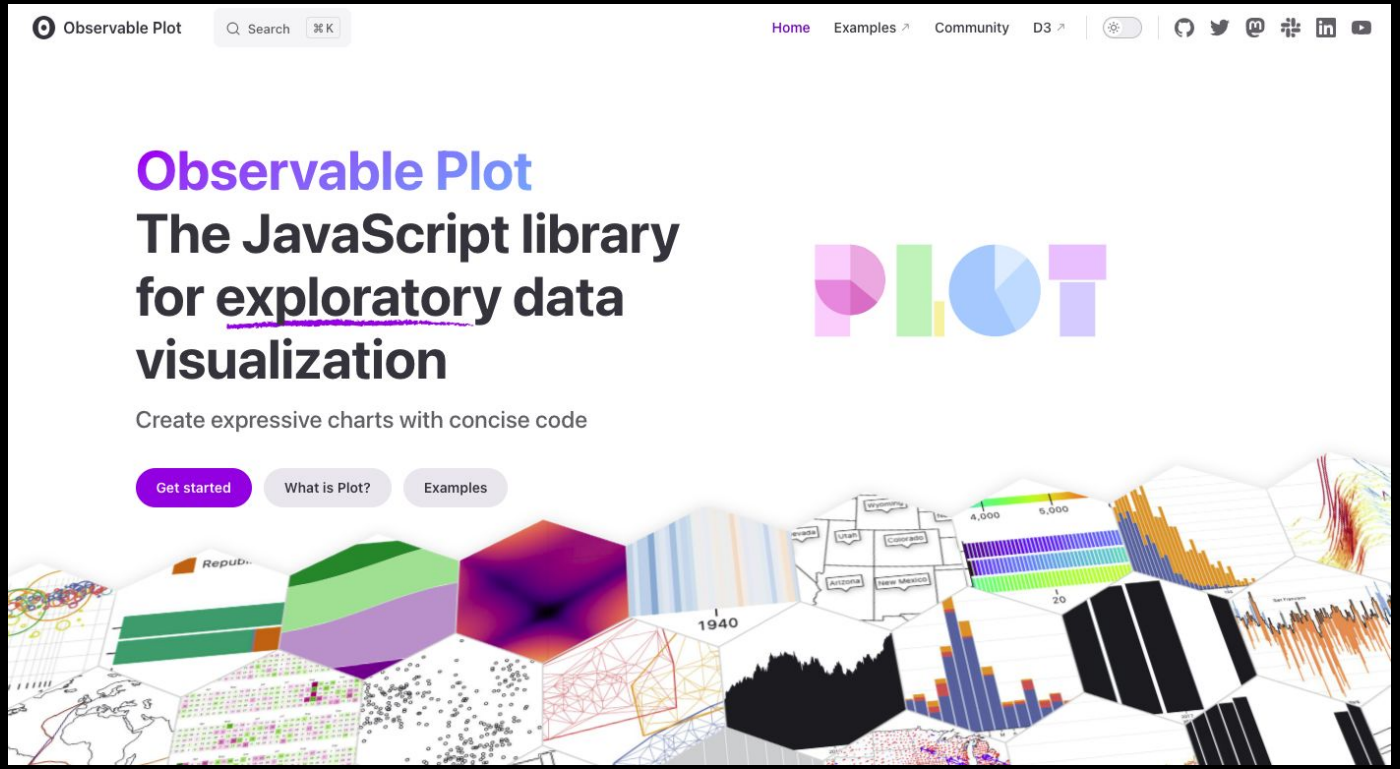
	Norway	Denmark	Sweden
2004	5	4	13
2022	8	10	15

Learn more

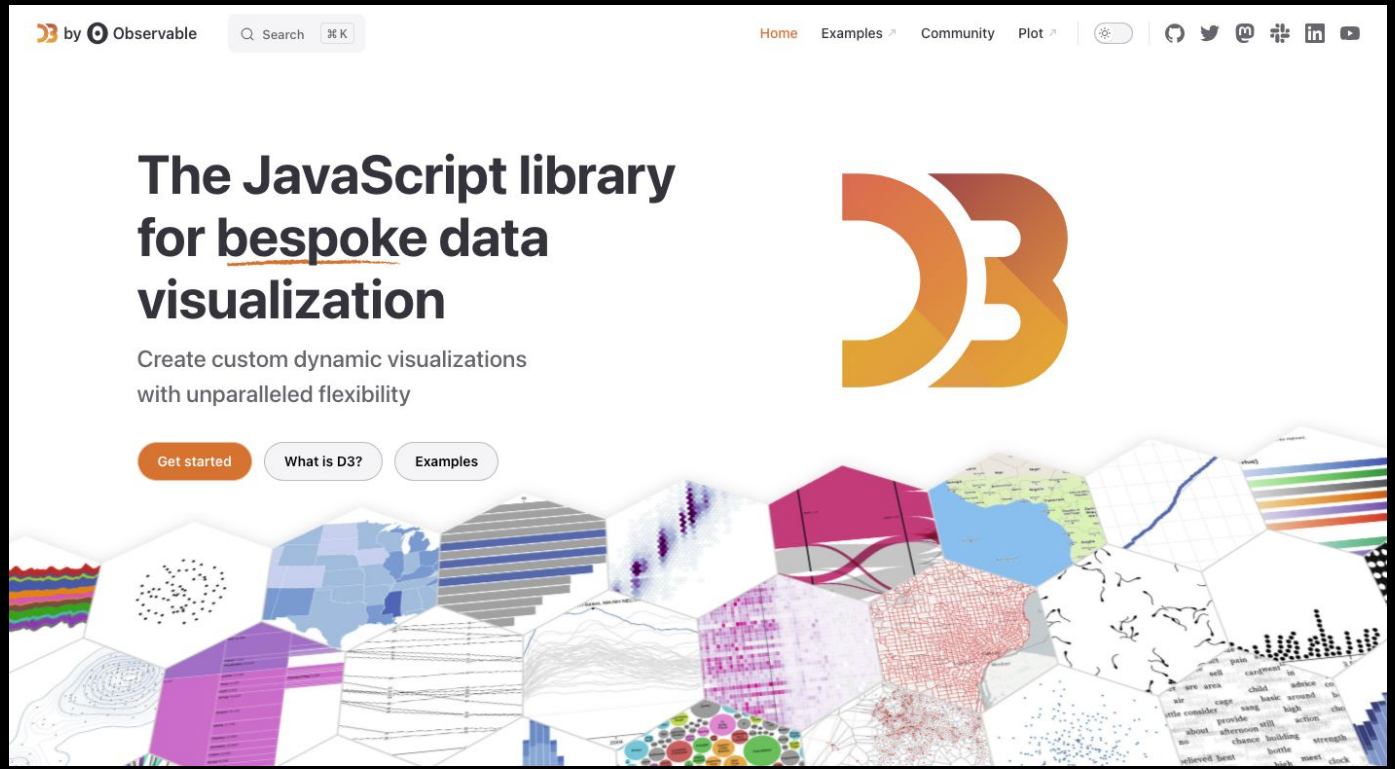
Good tools, properly used, can complement and enhance our thinking skills ***but never replace them.***

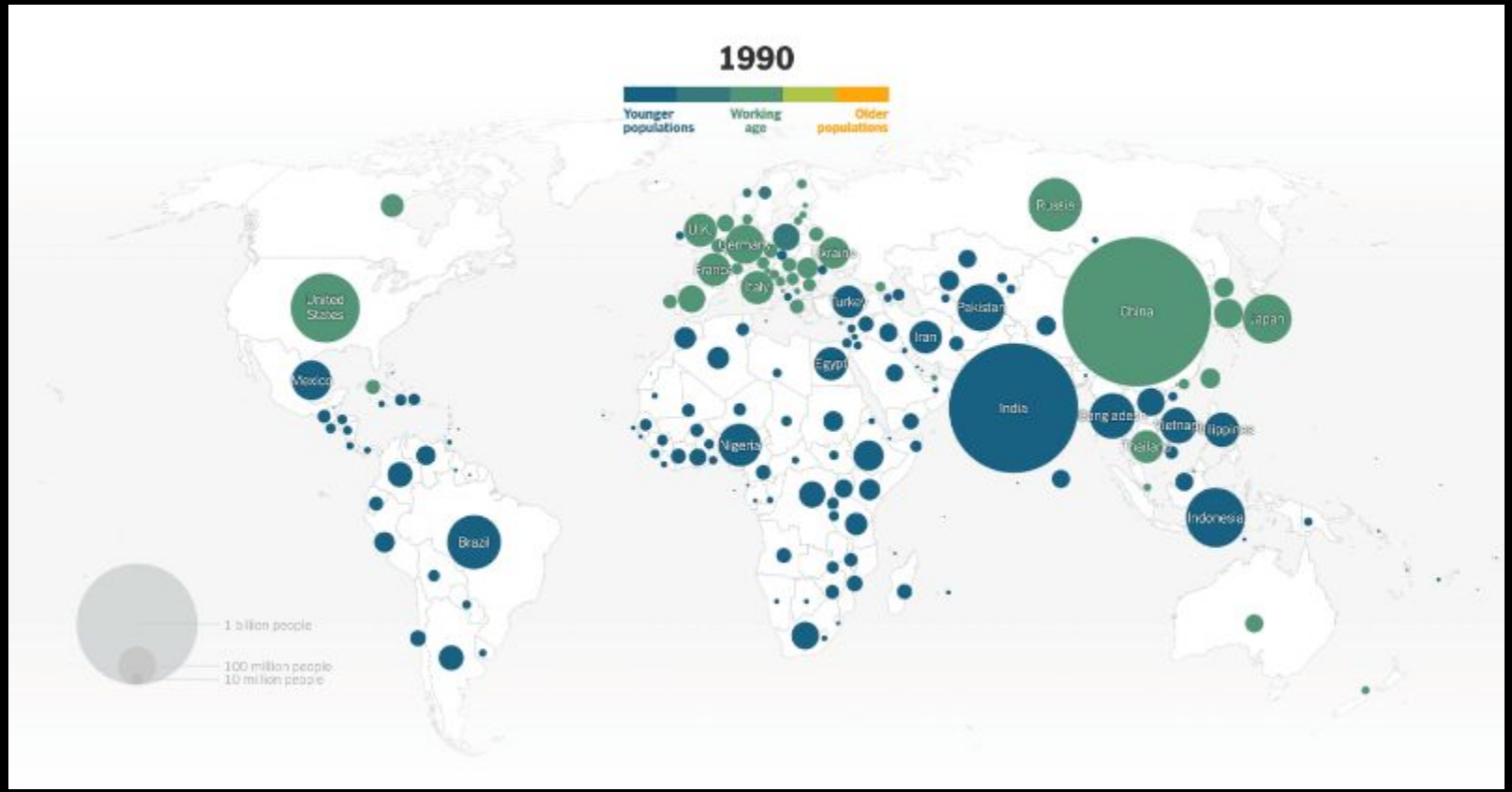
\- Stephen Few

But I am still going to show you a tool!



From the makers of D3




1990
Younger populations
Working age
Older populations
United States
Mexico
Brazil
Russia
China
India
Japan
Pakistan
Iran
Nigeria
Indonesia
1 billion people
100 million people
10 million people

Every Hindi film song by Lata Mangeshkar, in one graphic

73 years. At least 5,401 songs. In other words, Lata Mangeshkar has been at it for a long time.

By Gurman Bhatia and Aparna Alluri

September 28, 2016

Scroll to continue

What is D3?

The web is

HTML + CSS + Javascript

The web is

HTML + CSS + Javascript

and **d3.js is a javascript library**

For customised interactive charts on the web - the current go-to is D3.js + Svelte.

What is Observable then?

Web-based exploratory data analysis tool with the powers of D3.js

Observable Follow
Use data to think, together.

Fork...

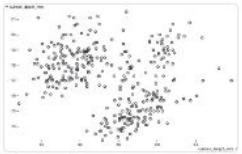
Public Edited Sep 26 Paused 11 forks 80 Likes

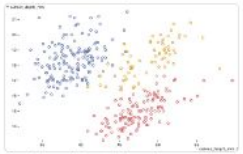
Plot Gallery

Here are a selection of Observable Plot examples.

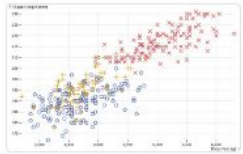
Dot, Image, Text

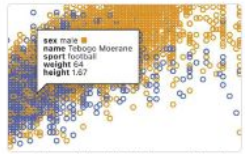
Plot individual data points.


Scatterplot

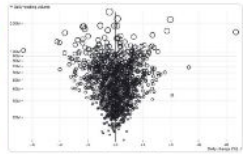

Scatterplot with color

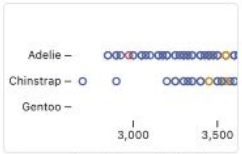
Diverging colored scatterplot


Symbol scatterplot

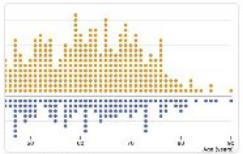


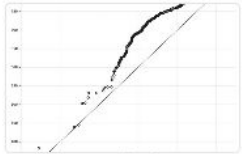
Scatterplot with interactive t...

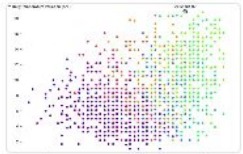

Proportional symbol scatter...

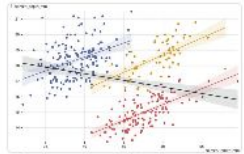


Scatterplot with ordinal dim...

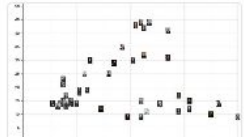

Stacked dots

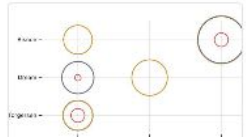

Quantile-quantile plot


Seattle temperature amplitu...


Simpson's paradox

Dot plot





https://bit.ly/csc-plot