Storytelling with Data

Nam Wook Kim

Mini-Courses — January @ GSAS
2018

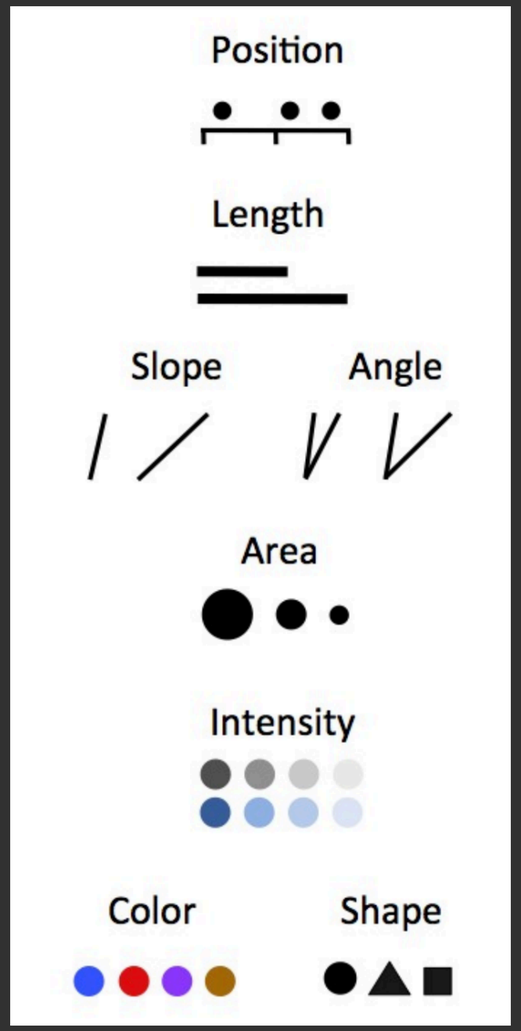
Goal

Beyond exploratory analysis: visualization for communication

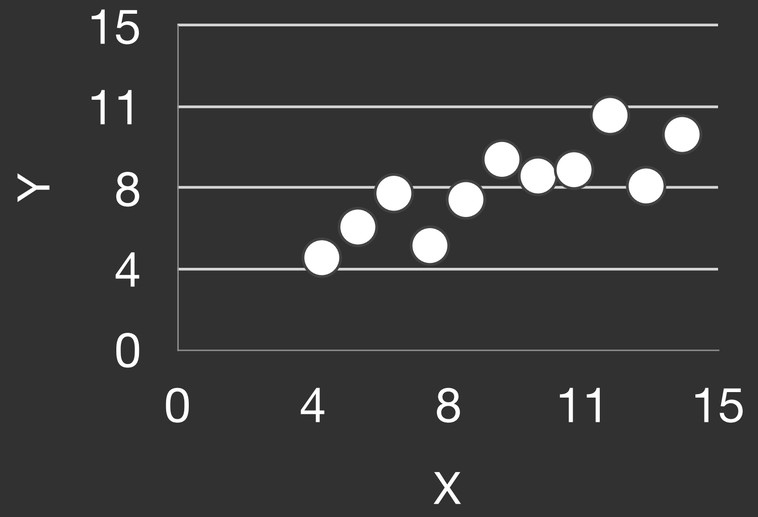
Data

X	Y
10.0	8.04
8.0	6.95
13.0	7.58
9.0	8.81
11.0	8.33
14.0	9.96
6.0	7.24
4.0	4.26
12.0	10.84
7.0	4.82
5.0	5.68

Perceptual Variables



Visualization



Position (x, y)

Design Criteria

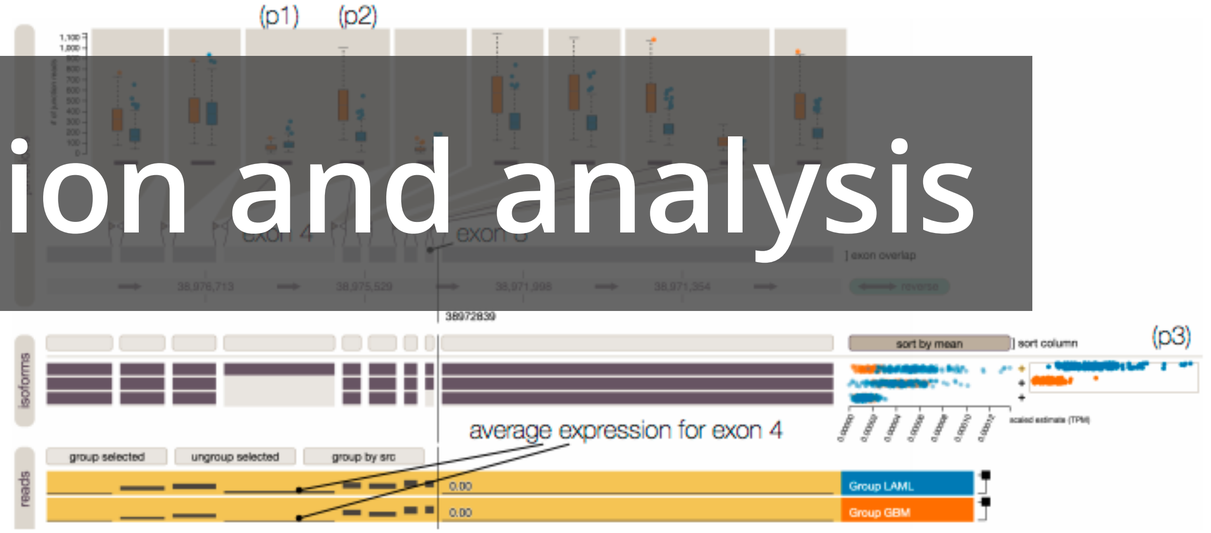
Expressiveness

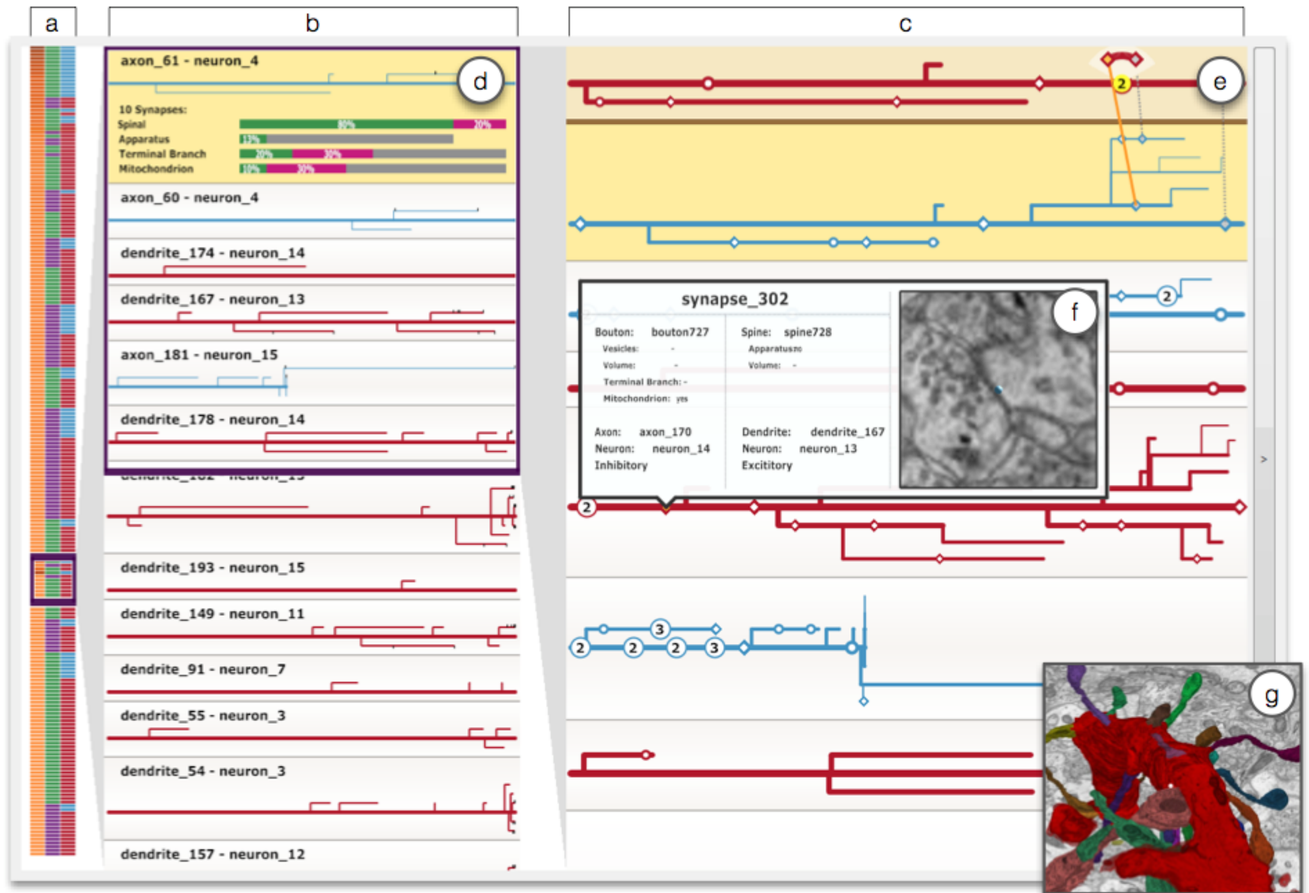
A set of facts is expressible in a visualization if it expresses all the facts and only the facts in the data.

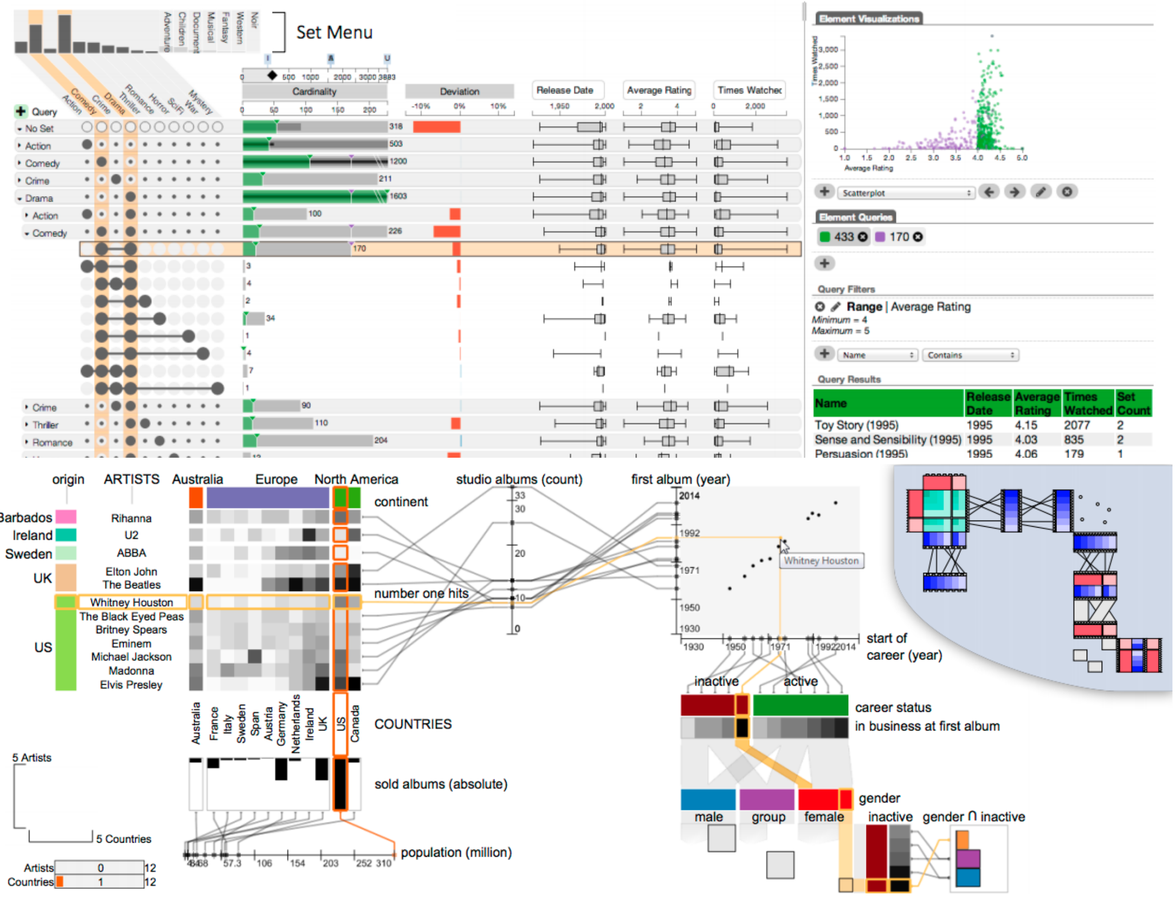
Effectiveness

A visualization is more effective than another one if the information conveyed is more readily perceived.

Rapid data exploration and analysis







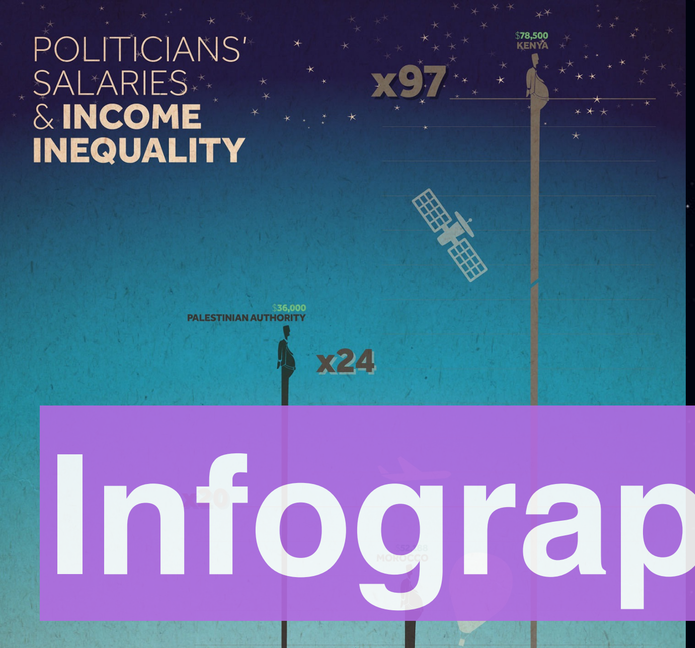
Data ⟷ Insights ⟷ ?

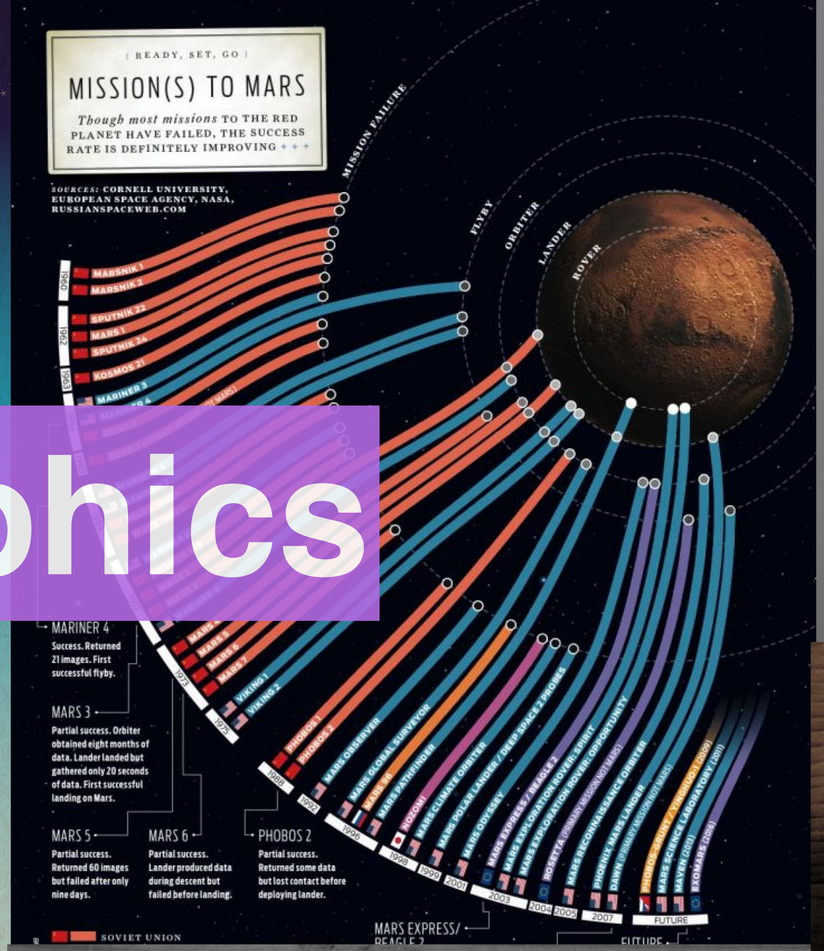
Exploratory

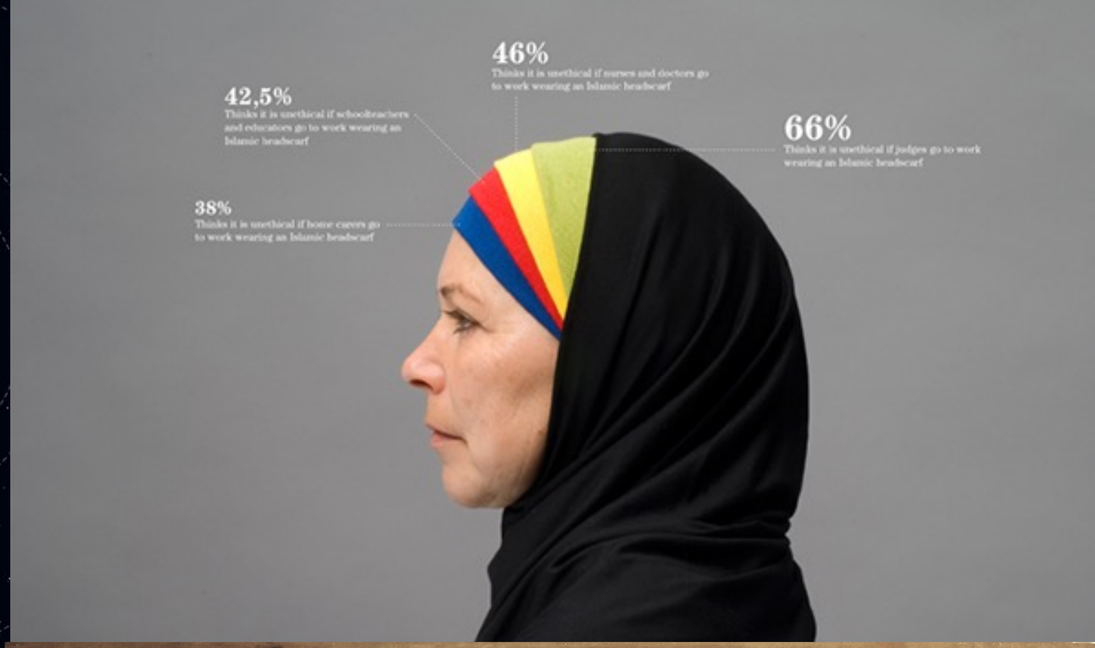
- Data Centered
- Domain Experts
- Analysis
- Desktop
- In-Lab

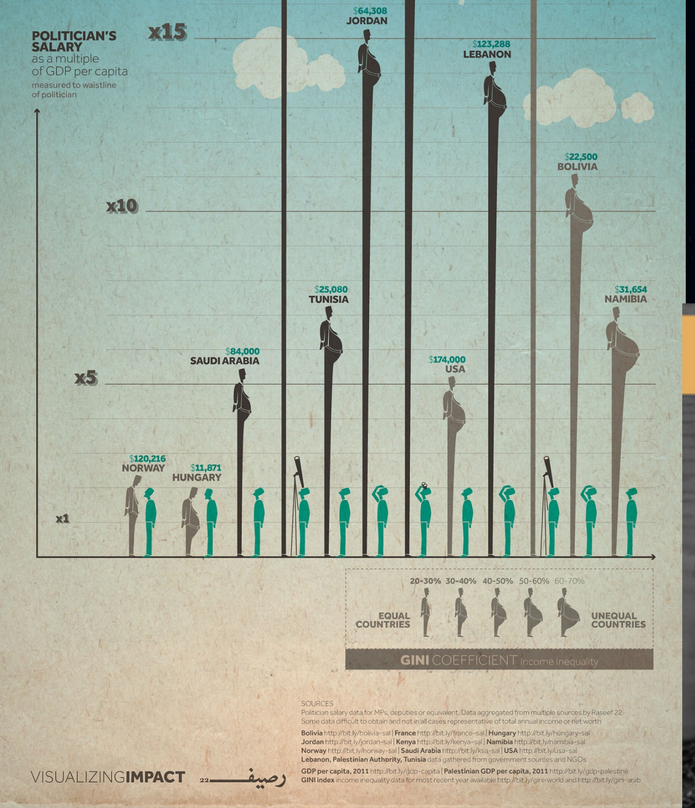
Need to communicate insights found from data

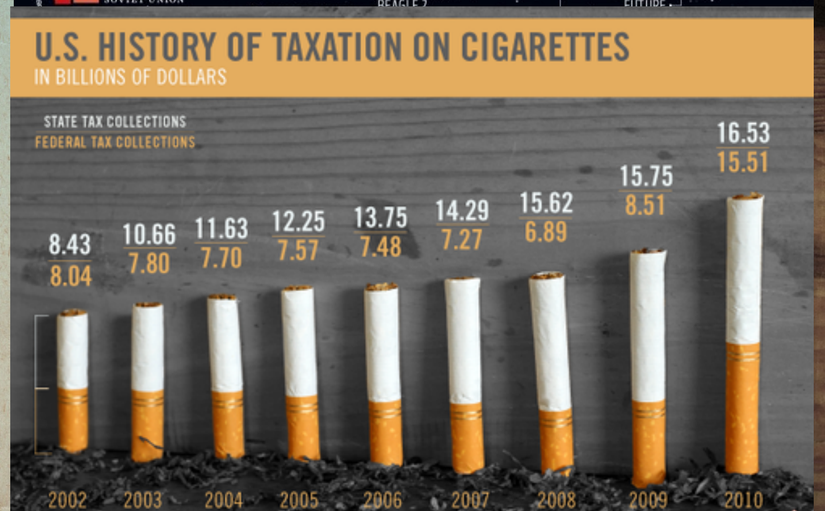
Infographics

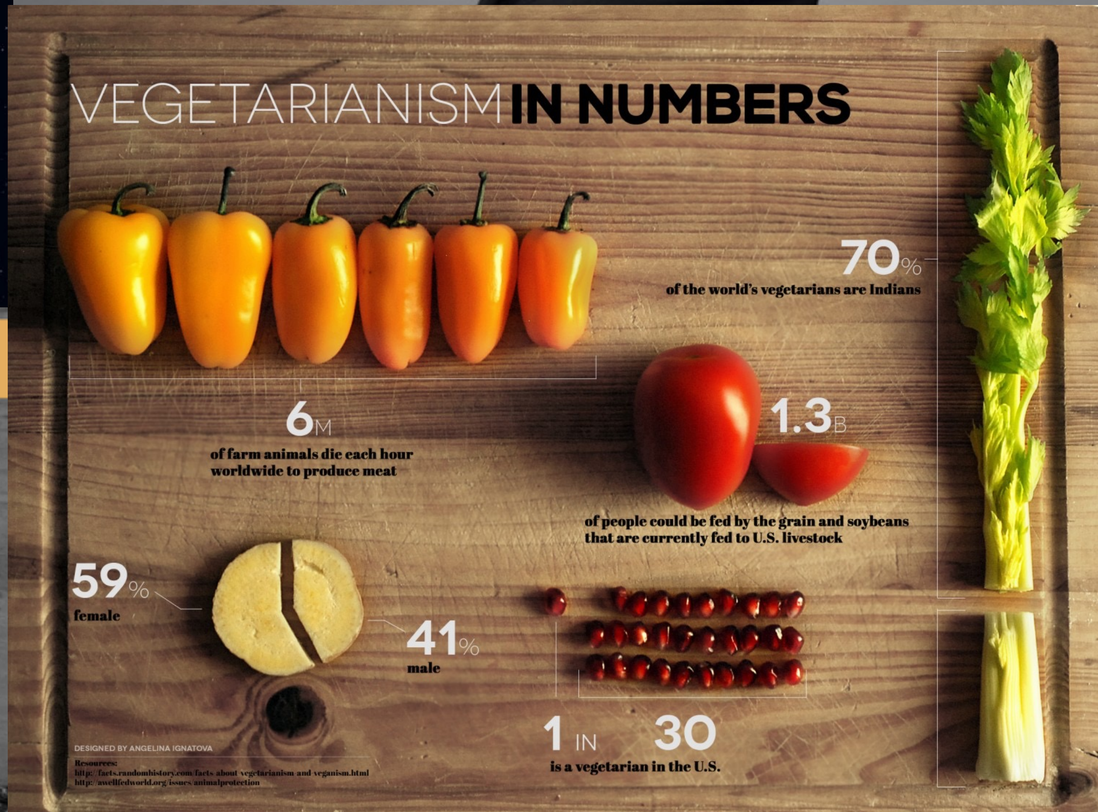












Storytelling

Hans Rosling

tapestry

Storytelling is the most powerful way to put ideas into the world today.

Robert McKee

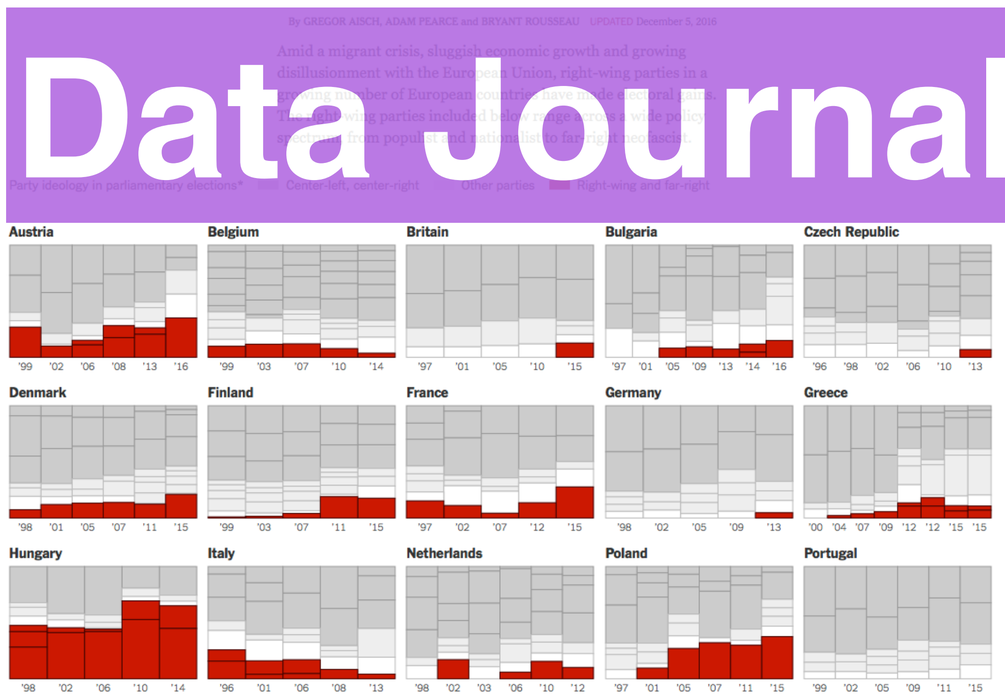
cole nussbaumer knaflic

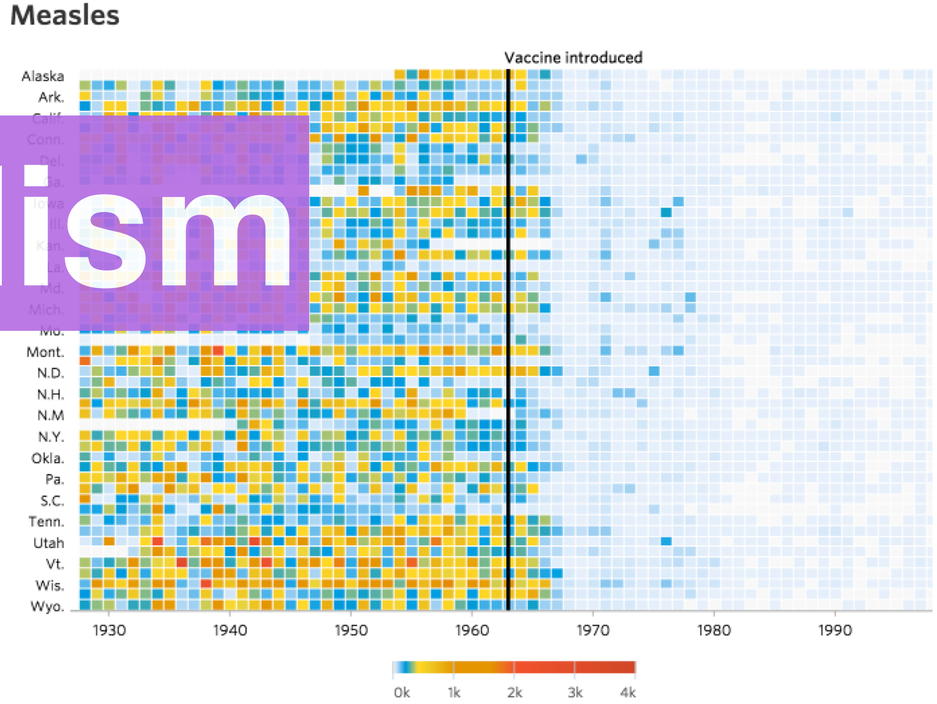
storytelling with data

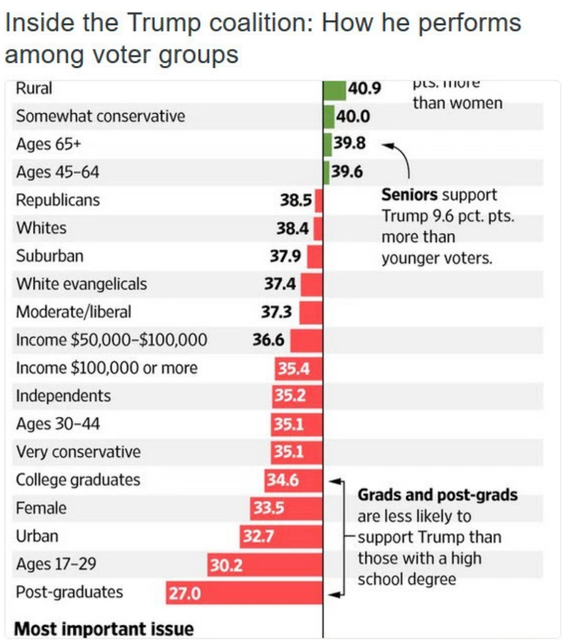
a data visualization guide for business professionals

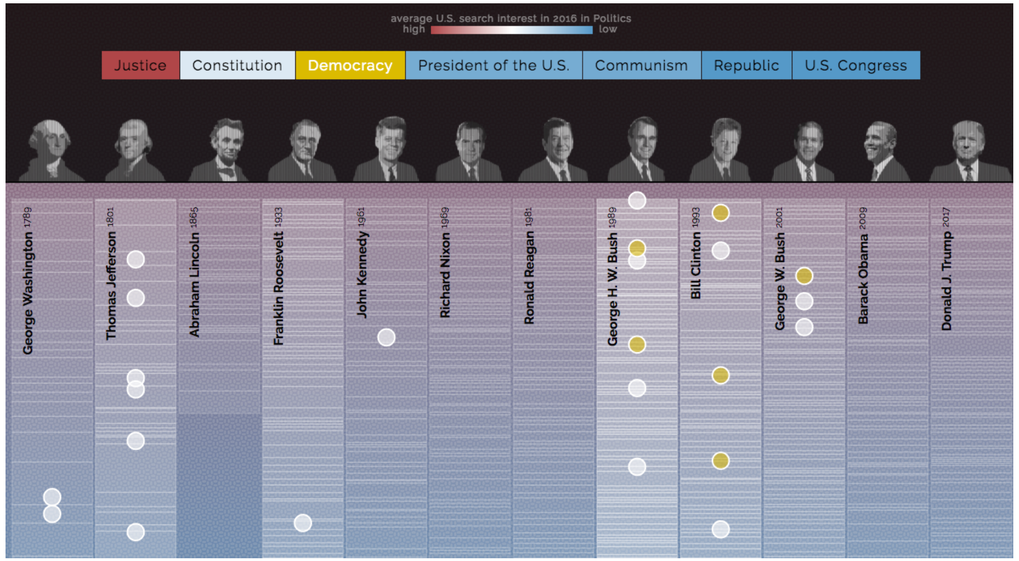
WILEY

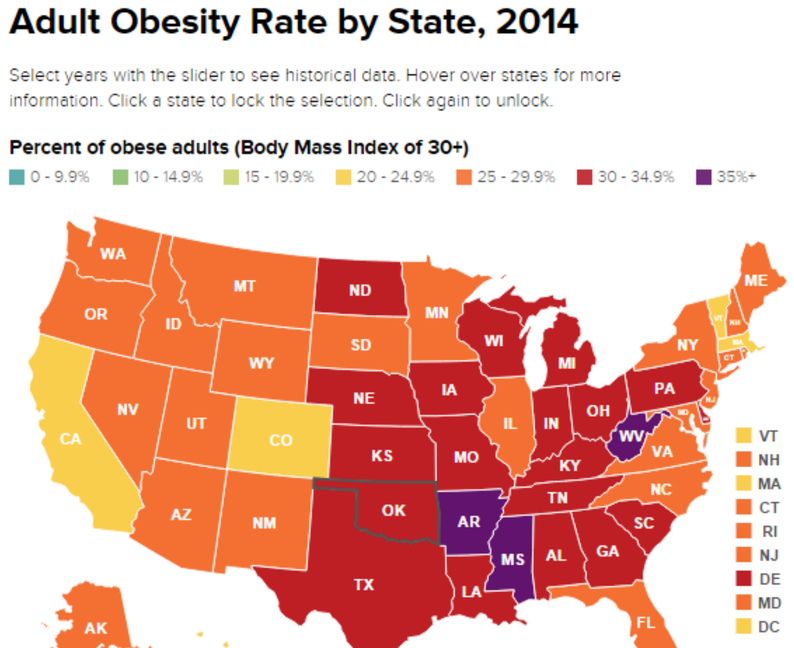
Data Journalism

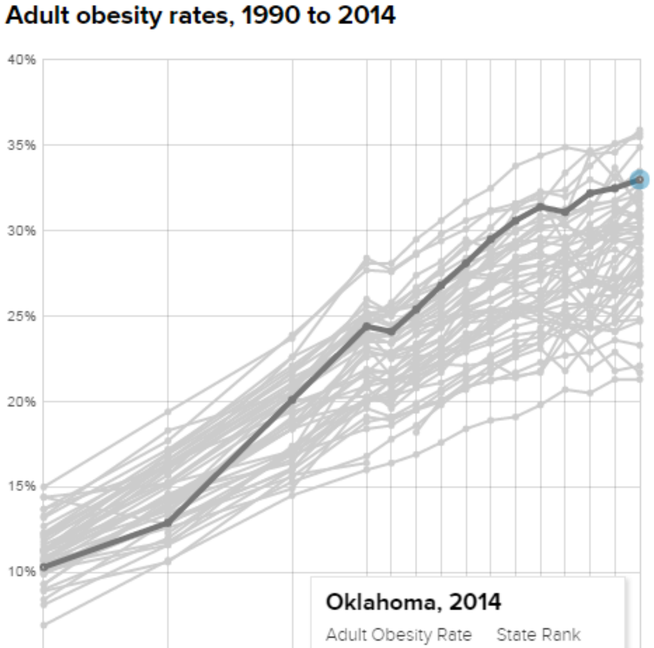




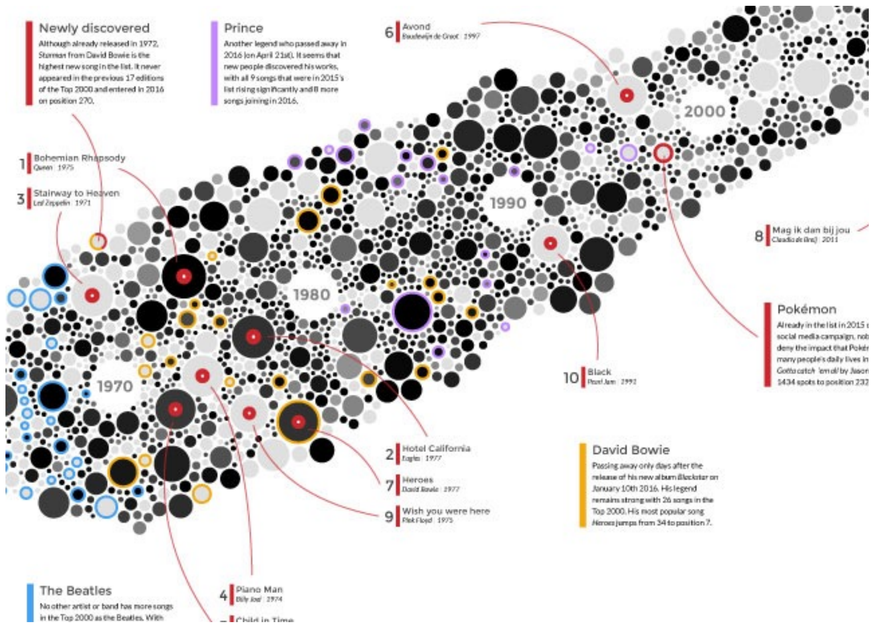


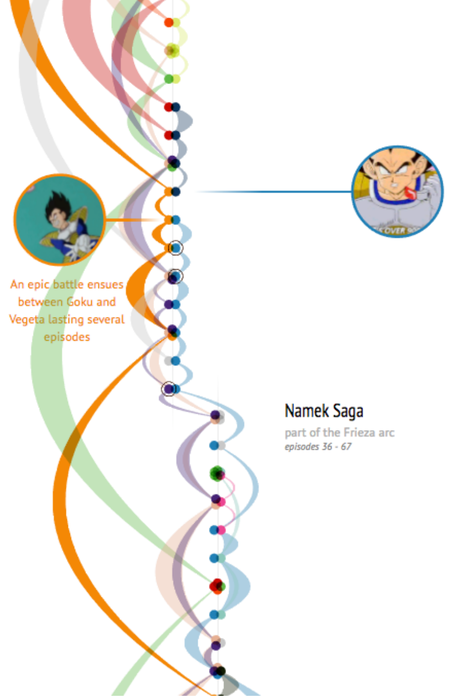


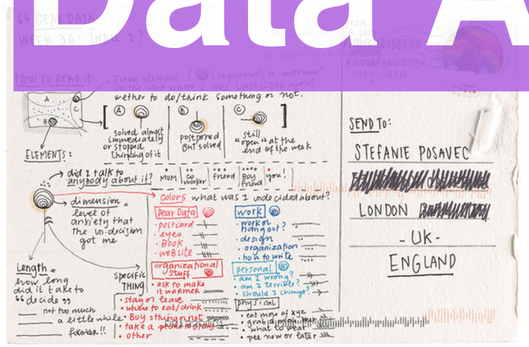


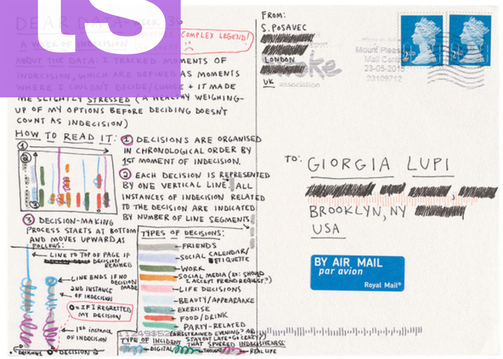


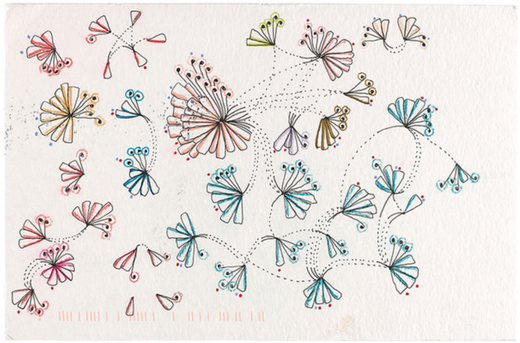
Data Arts

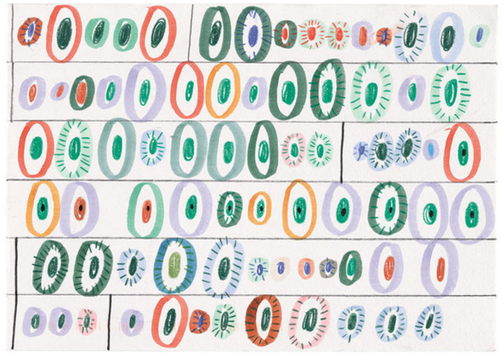


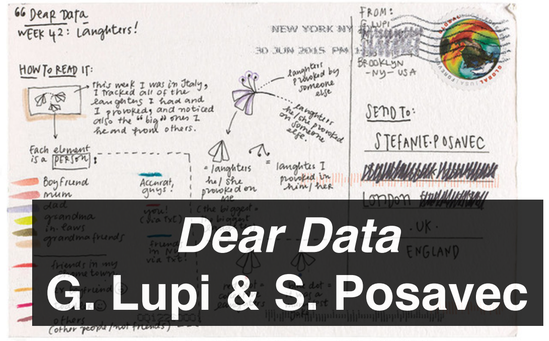


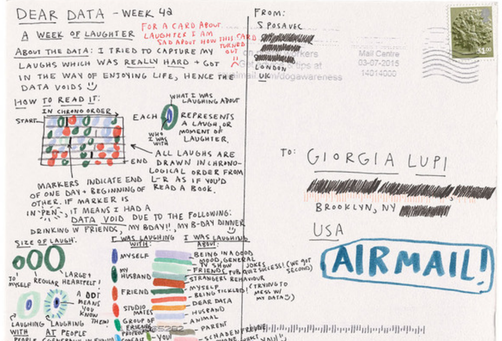


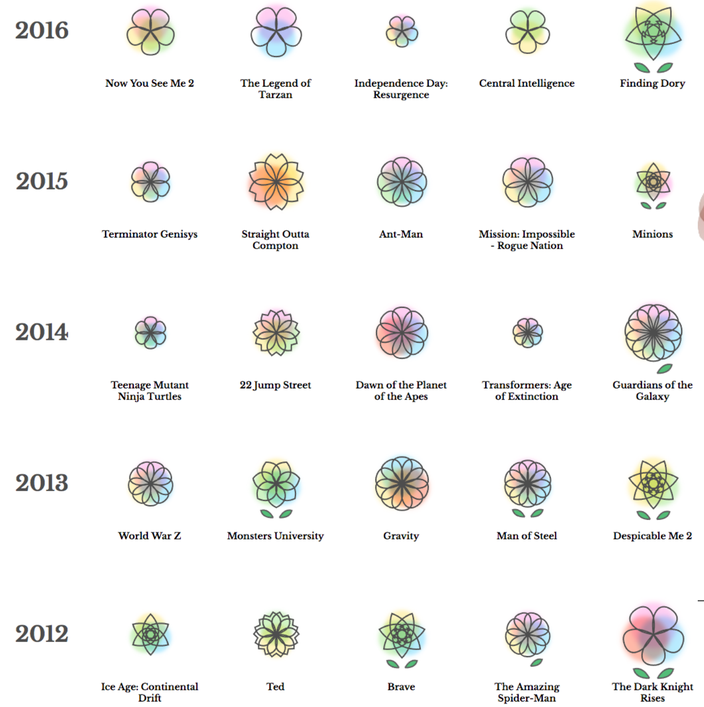


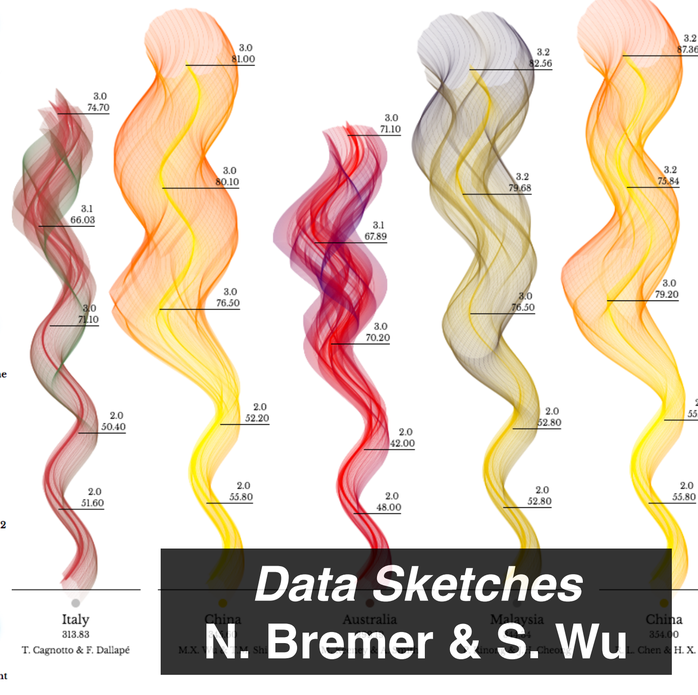








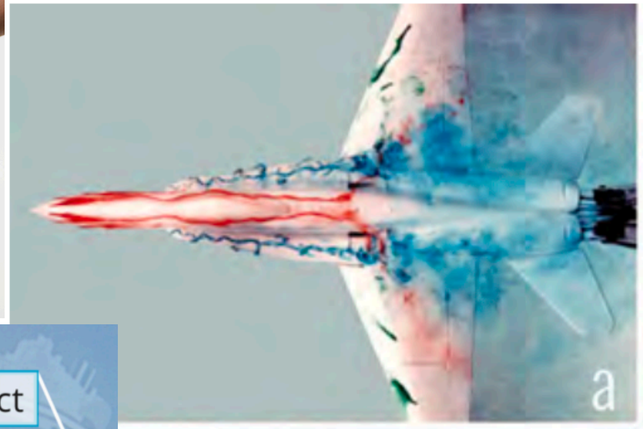


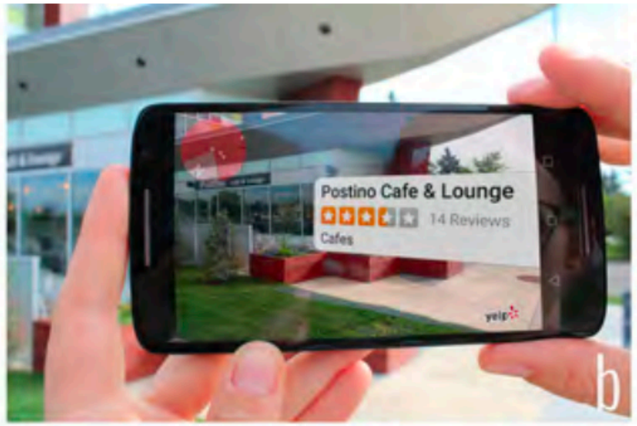


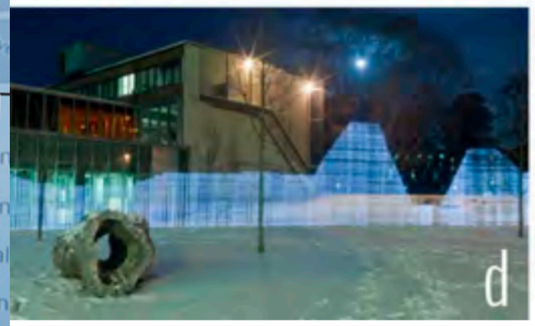
Dear Data
G. Lupi & S. Posavec

Data Sketches
N. Bremer & S. Wu

New Environments







Data ⟷ Insights ⟷ Messages

Exploratory

- Data Centered
- Domain Experts
- Analysis
- Desktop
- In-Lab

Beyond Exploratory

- Human Centered
- General Audience
- Communication
- Off-Desktop
- In-the-Wild

Topics

- Visualization for communication
- Telling compelling stories with data
- Story Points and Dashboards in Tableau

Visualization for Communication

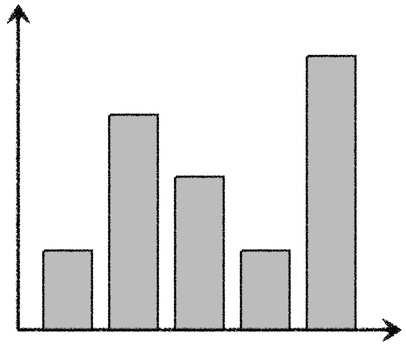
Visualization for Exploration

- Faster and accurate reading
- Rapidly generate many visualizations
- No titles, annotations, embellishments

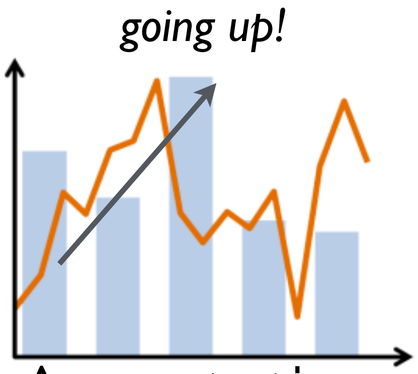
Visualization for Communication

- Good headlines
- Annotations and highlights
- Redundant encodings
- Pictograms and useful embellishments
- Explanations (e.g., legend, source)

Examples of Redundant Encoding

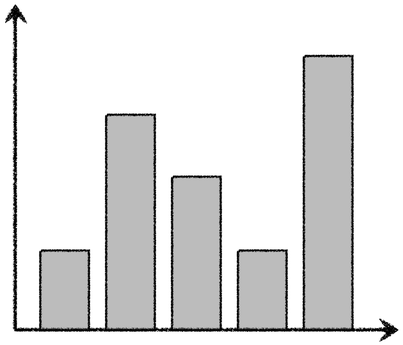


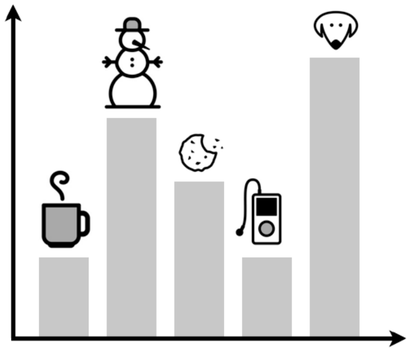
Examples of Redundant Encoding



Annotation

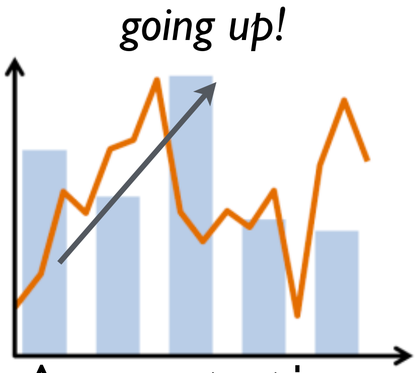
Message Redundancy





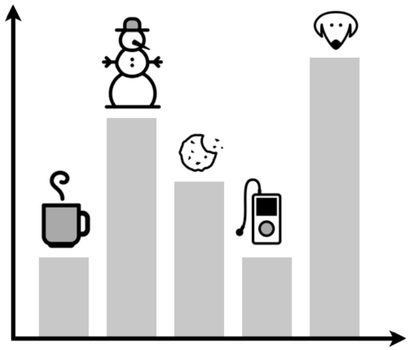
Human recognizable objects

Examples of Redundant Encoding

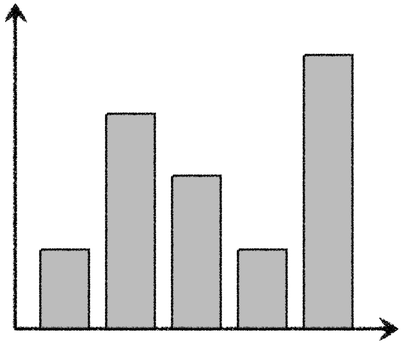


Annotation

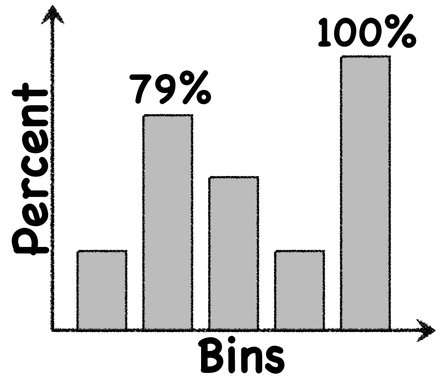
Message Redundancy



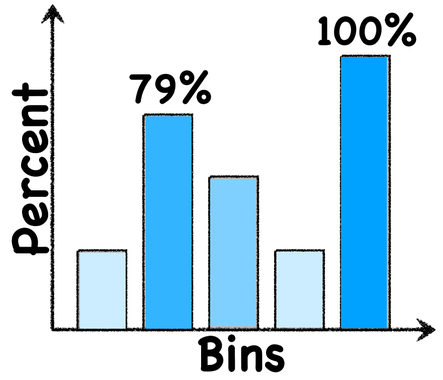
Human recognizable objects



Data Redundancy

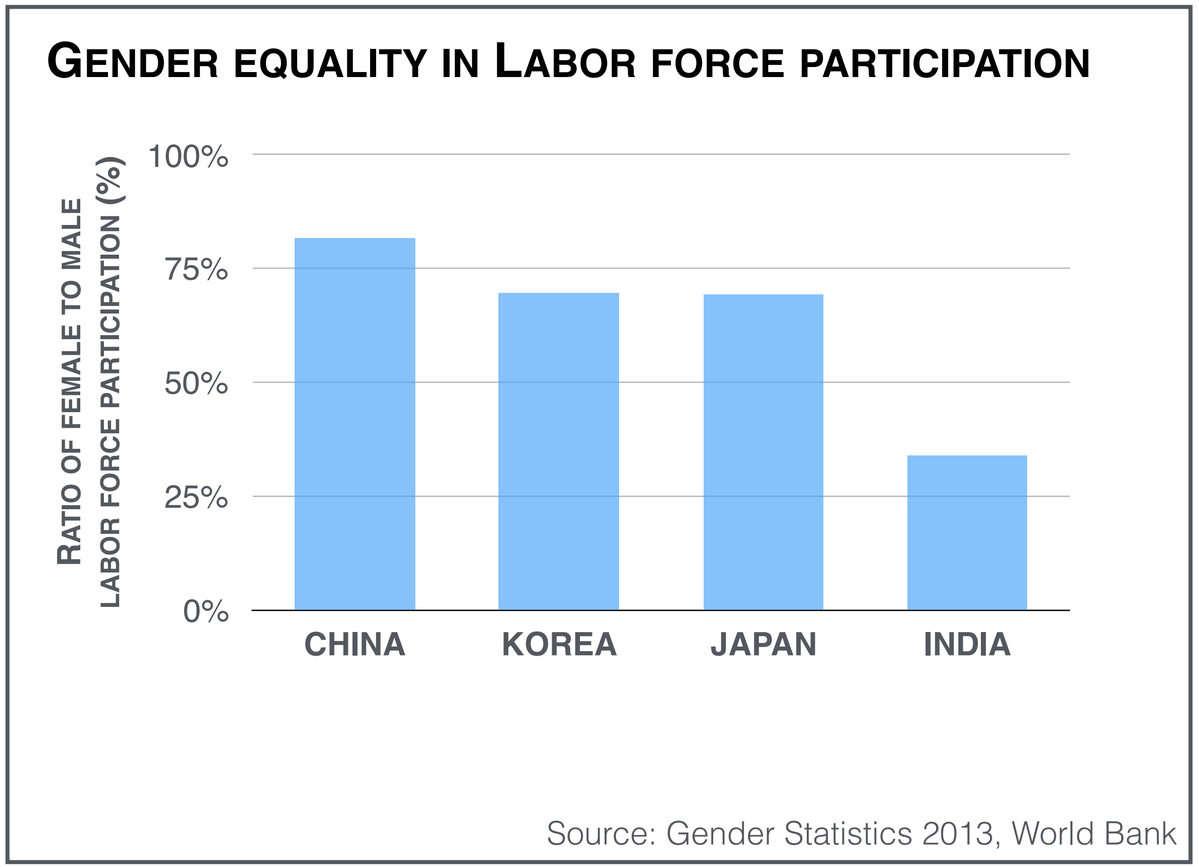


Data Labels

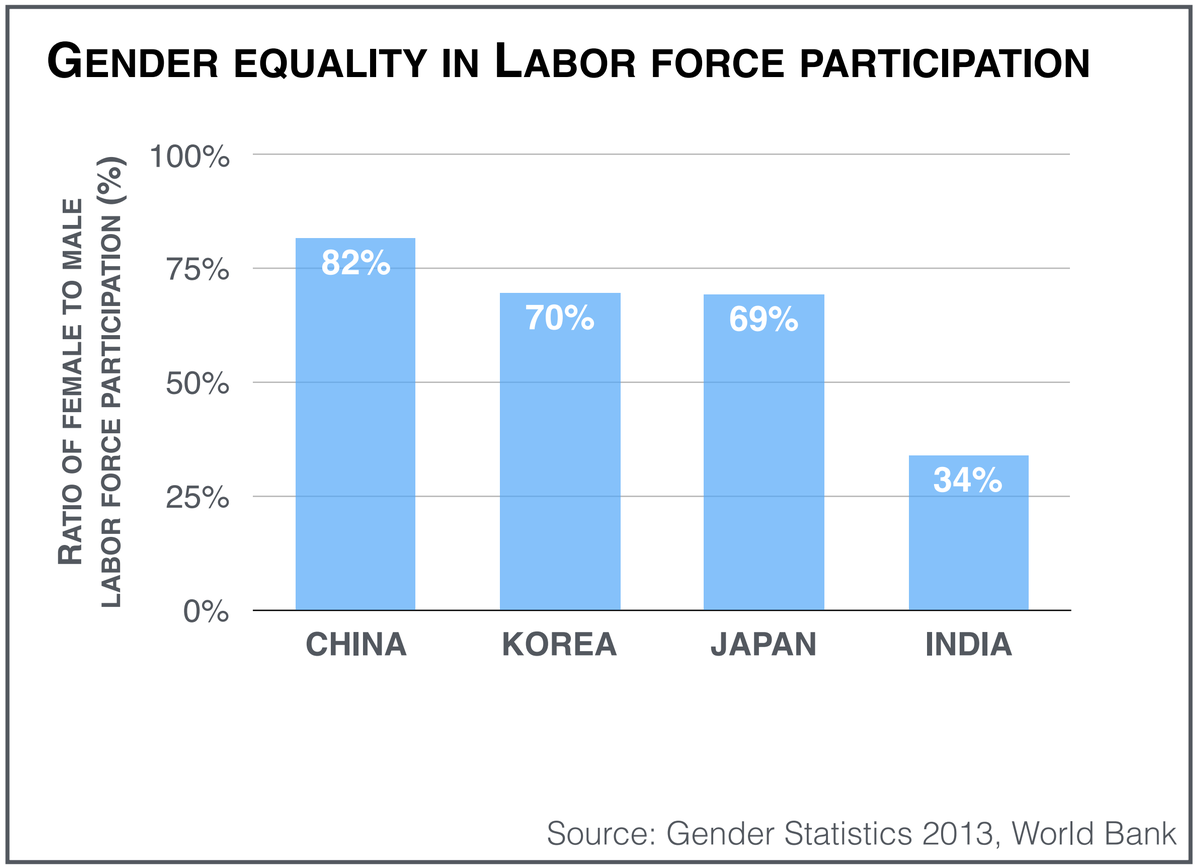


Lightness

Examples


GENDER EQUALITY IN LABOR FORCE PARTICIPATION
RATIO OF FEMALE TO MALE LABOR FORCE PARTICIPATION (%)
100%
75%
50%
25%
0%
CHINA
KOREA
JAPAN
INDIA
Source: Gender Statistics 2013, World Bank

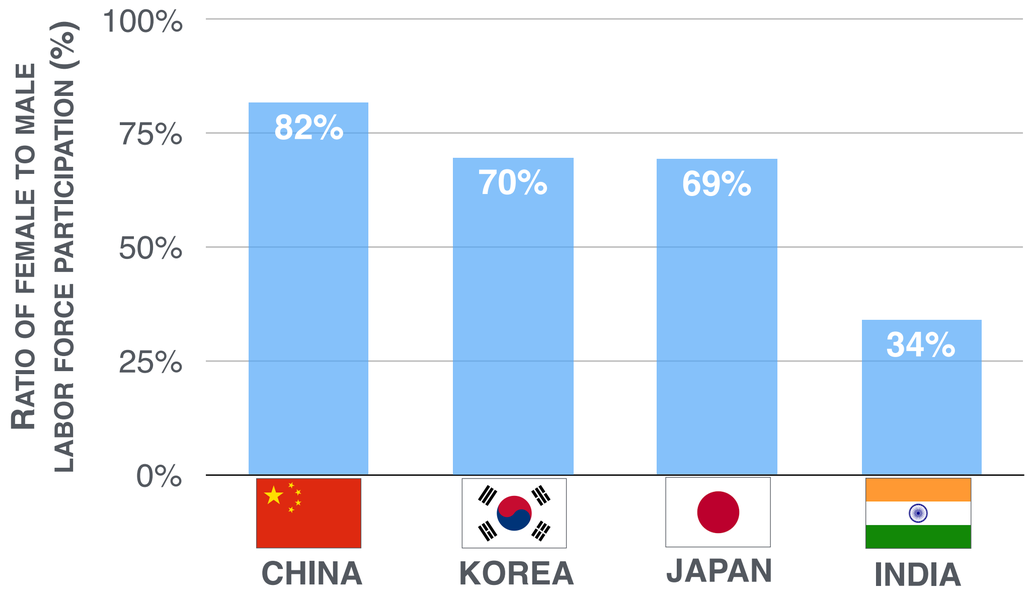
ORIGINAL



DATA REDUNDANCY

GENDER EQUALITY IN LABOR FORCE PARTICIPATION

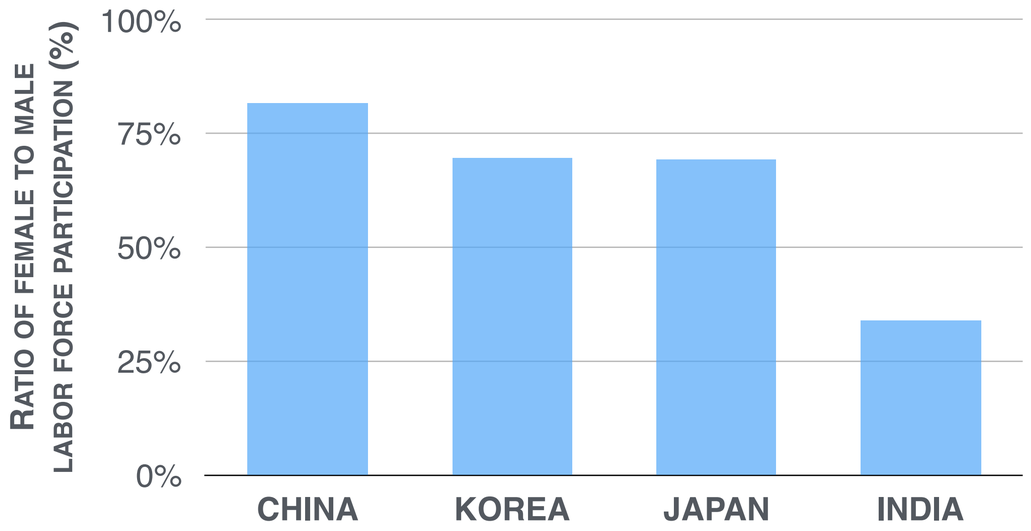
CHINA LEADS IN FEMALE LABOR FORCE PARTICIPATION WHEREAS INDIA LAGS SIGNIFICANTLY BEHIND AMONG OTHER ASIAN-PACIFIC COUNTRIES IN 2013.



Source: Gender Statistics 2013, World Bank

DATA & MESSAGE REDUNDANCY

GENDER EQUALITY IN LABOR FORCE PARTICIPATION

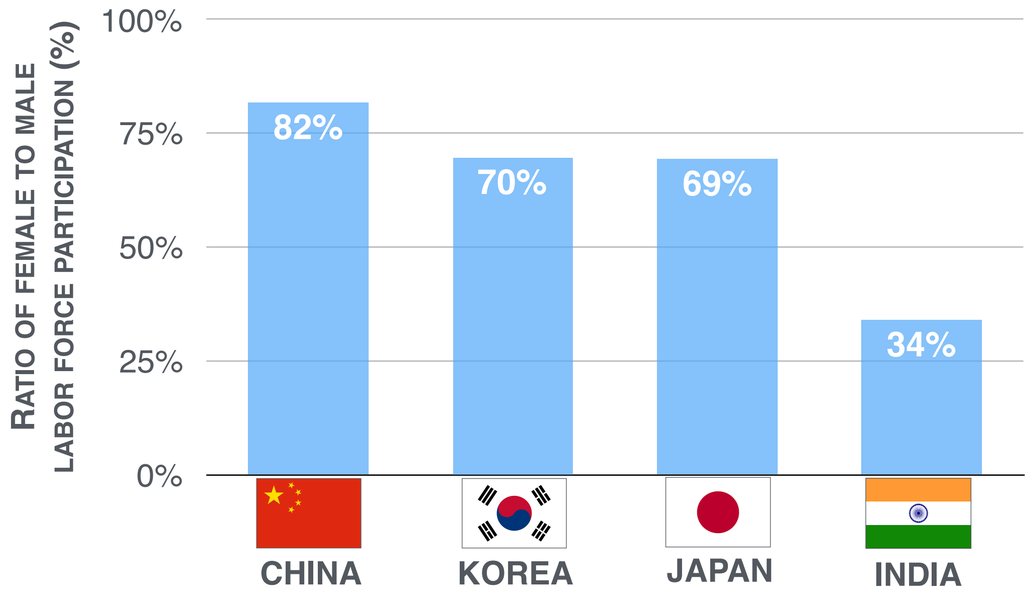


Source: Gender Statistics 2013, World Bank

ORIGINAL

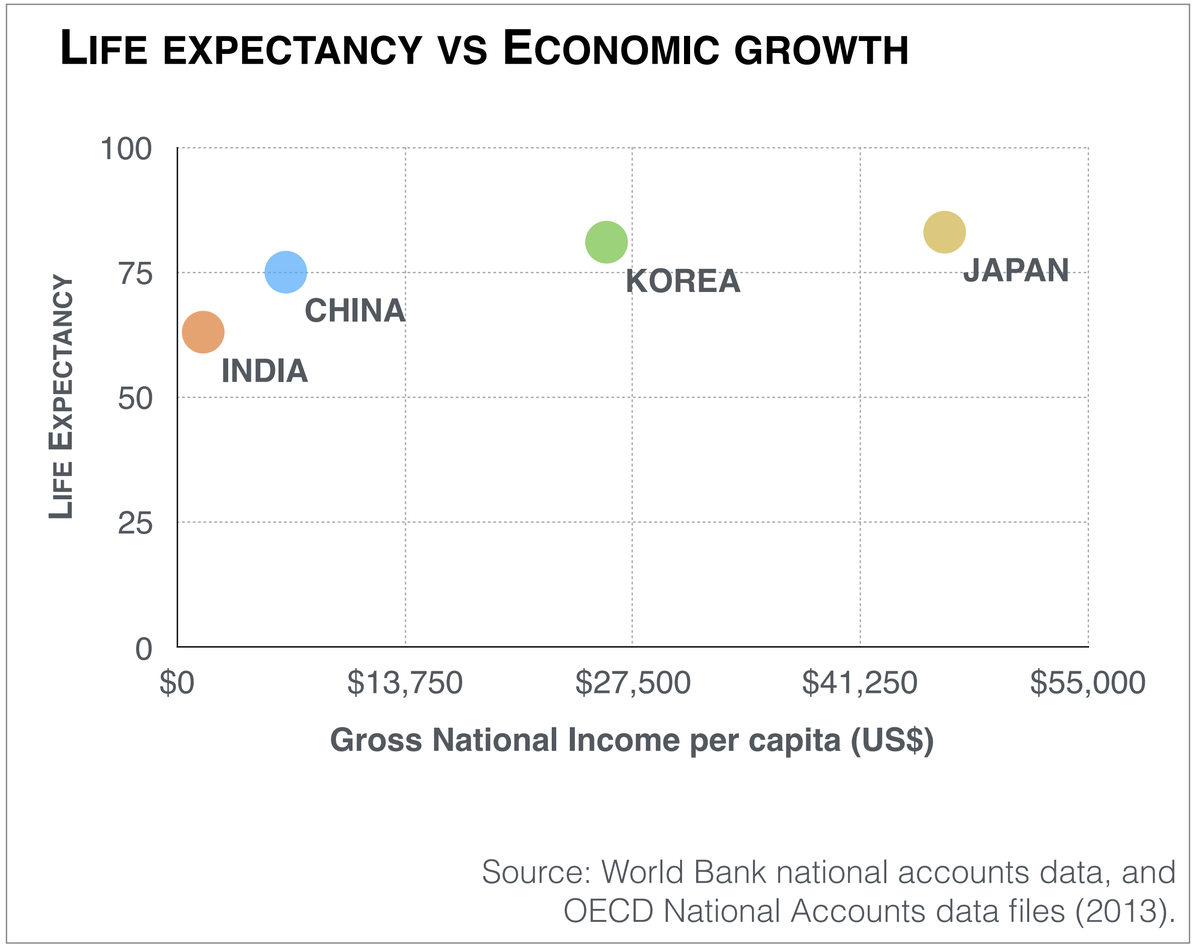
GENDER EQUALITY IN LABOR FORCE PARTICIPATION

CHINA LEADS IN FEMALE LABOR FORCE PARTICIPATION WHEREAS INDIA LAGS SIGNIFICANTLY BEHIND AMONG OTHER ASIAN-PACIFIC COUNTRIES IN 2013.

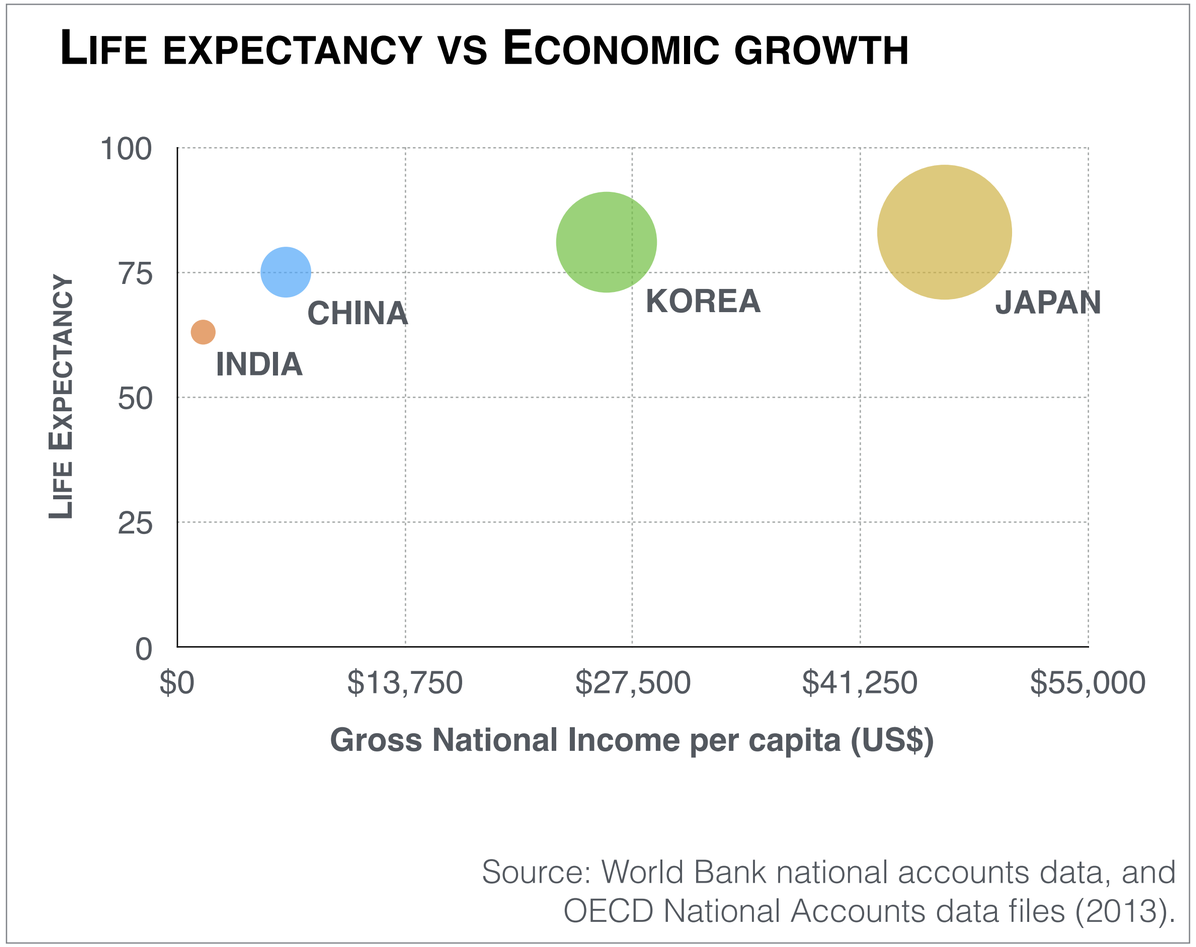


Source: Gender Statistics 2013, World Bank

DATA & MESSAGE REDUNDANCY


LIFE EXPECTANCY VS ECONOMIC GROWTH
100
75
50
25
0
LIFE EXPECTANCY
CHINA
INDIA
KOREA
JAPAN
$0
$13,750
$27,500
$41,250
$55,000
Gross National Income per capita (US$)
Source: World Bank national accounts data, and
OECD National Accounts data files (2013).

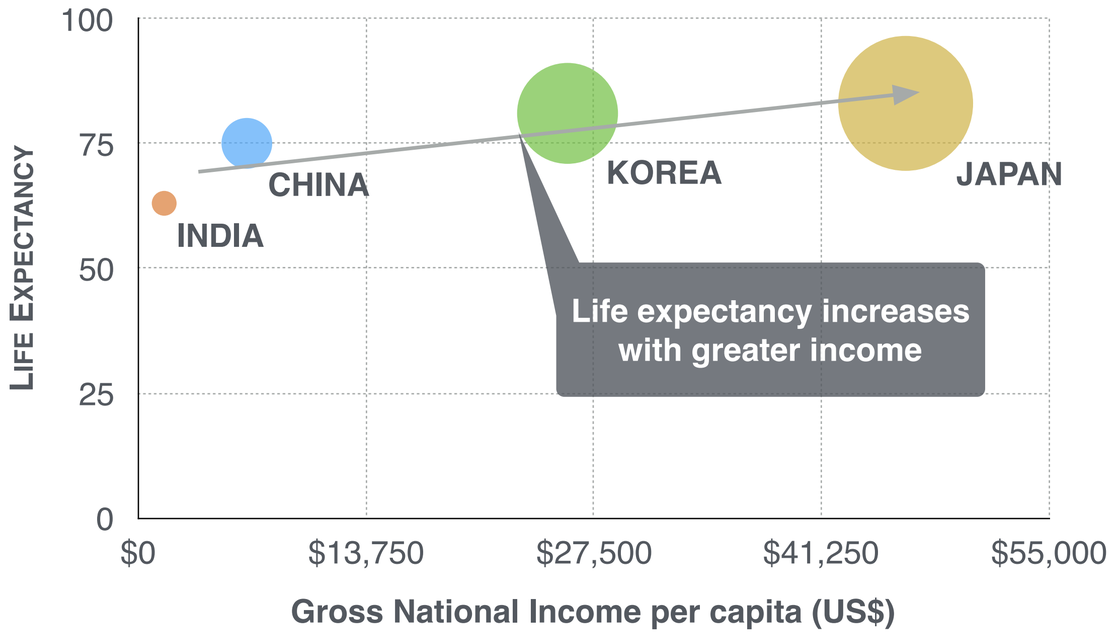
ORIGINAL



DATA REDUNDANCY

LIFE EXPECTANCY VS ECONOMIC GROWTH

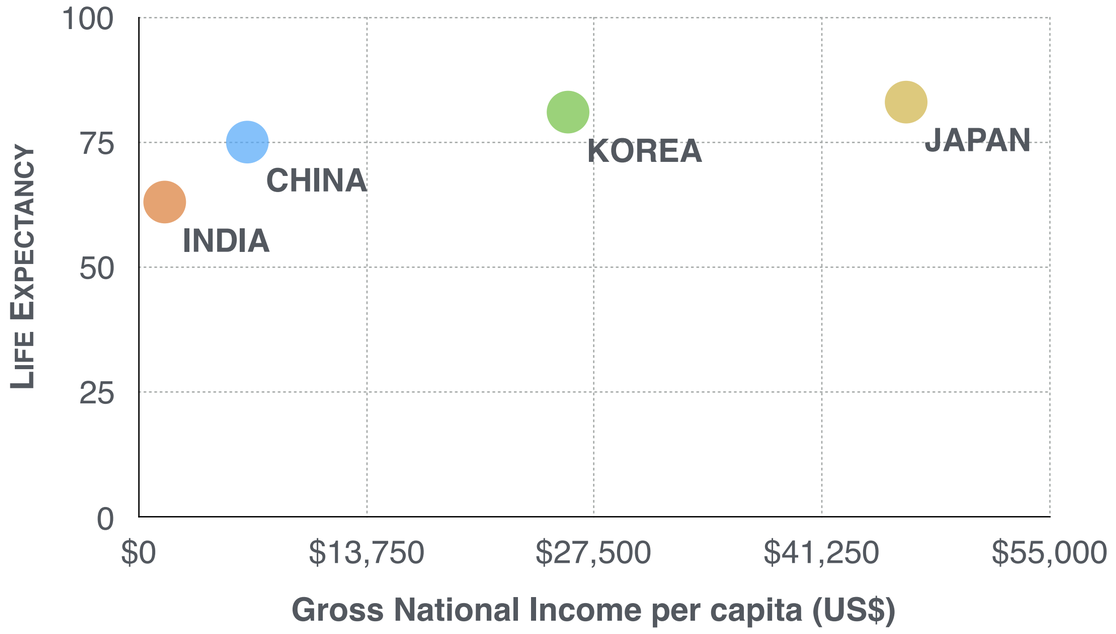
RELATION BETWEEN LIFE EXPECTANCY AND GROSS NATIONAL INCOME OF ASIAN-PACIFIC COUNTRIES IN 2013



Source: World Bank national accounts data, and OECD National Accounts data files (2013).

DATA & MESSAGE REDUNDANCY

LIFE EXPECTANCY VS ECONOMIC GROWTH

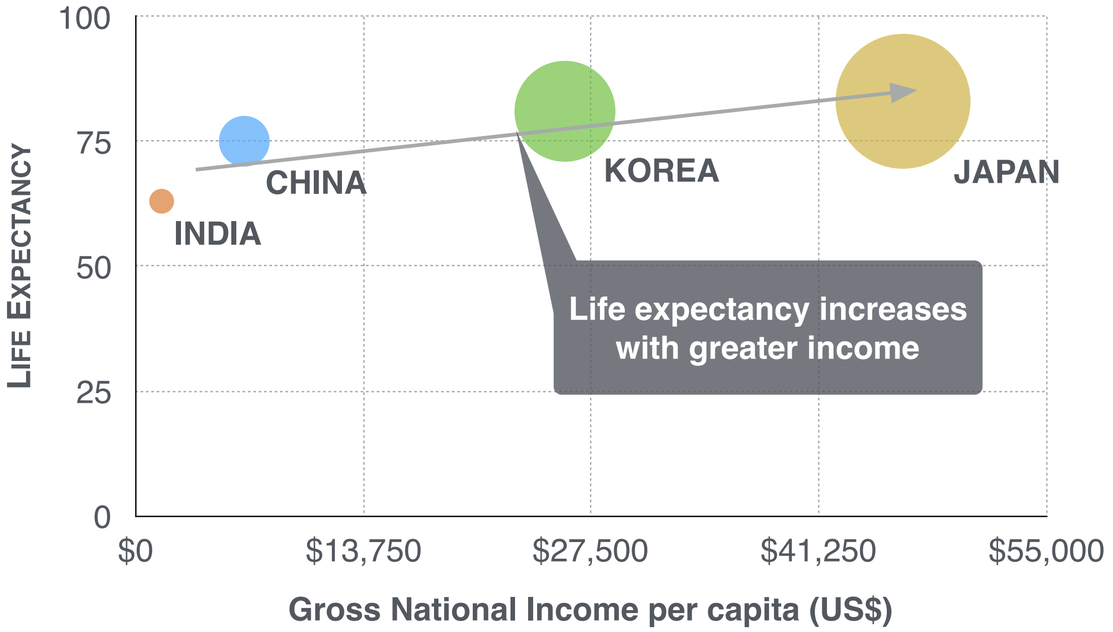


Source: World Bank national accounts data, and OECD National Accounts data files (2013).

ORIGINAL

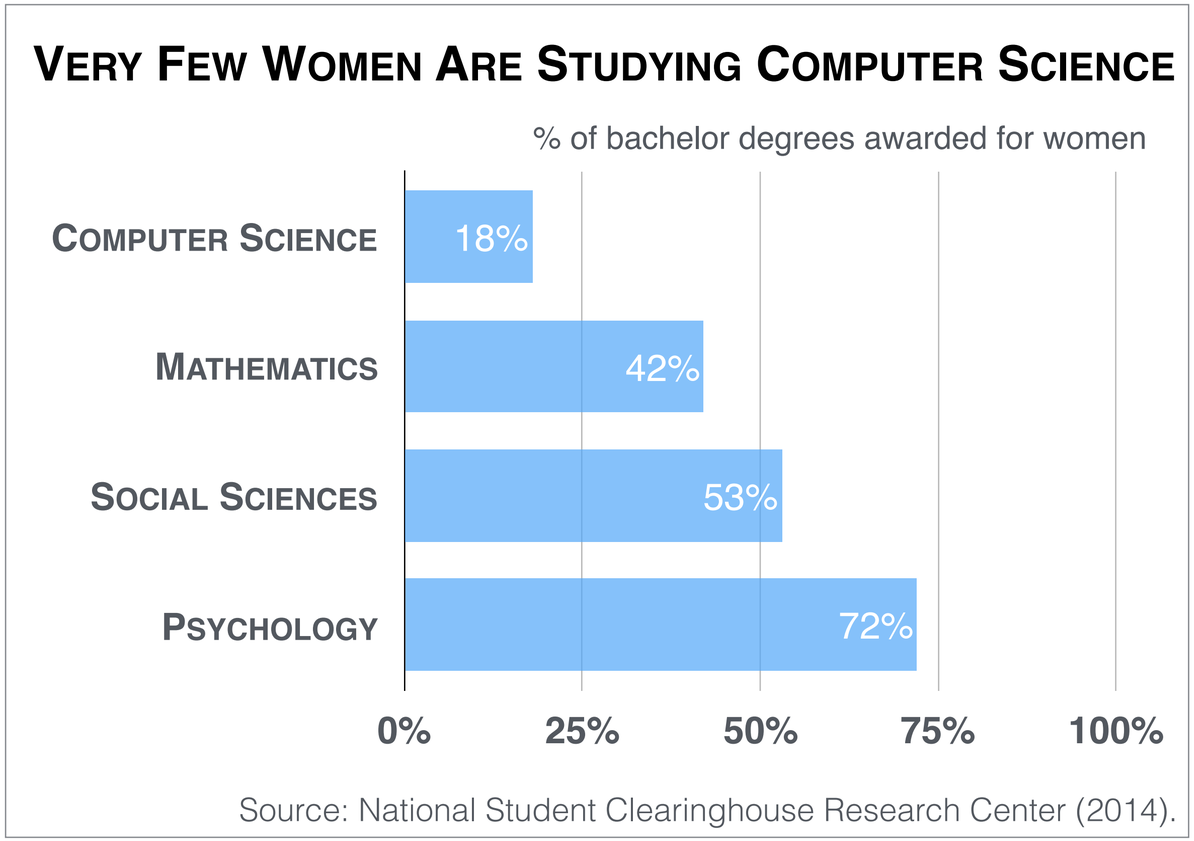
LIFE EXPECTANCY VS ECONOMIC GROWTH

RELATION BETWEEN LIFE EXPECTANCY AND GROSS NATIONAL INCOME OF ASIAN-PACIFIC COUNTRIES IN 2013

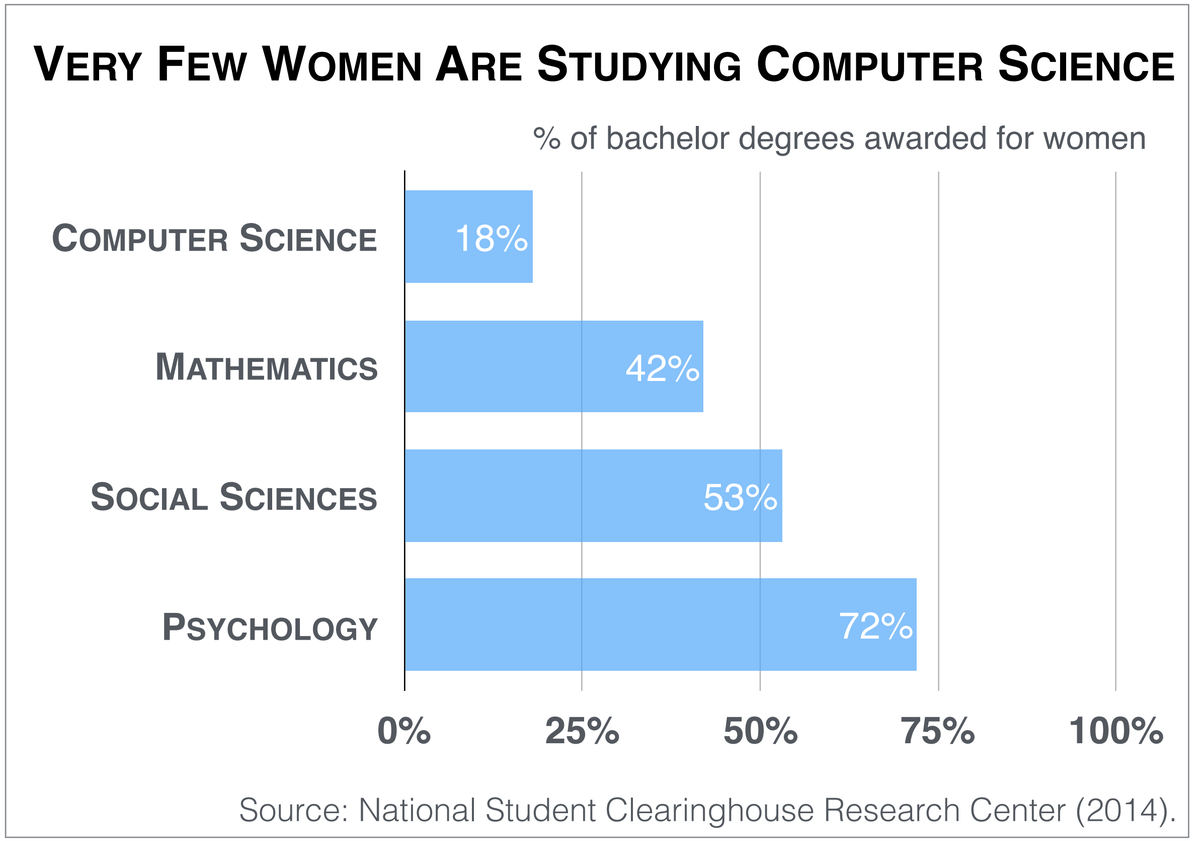


Source: World Bank national accounts data, and OECD National Accounts data files (2013).

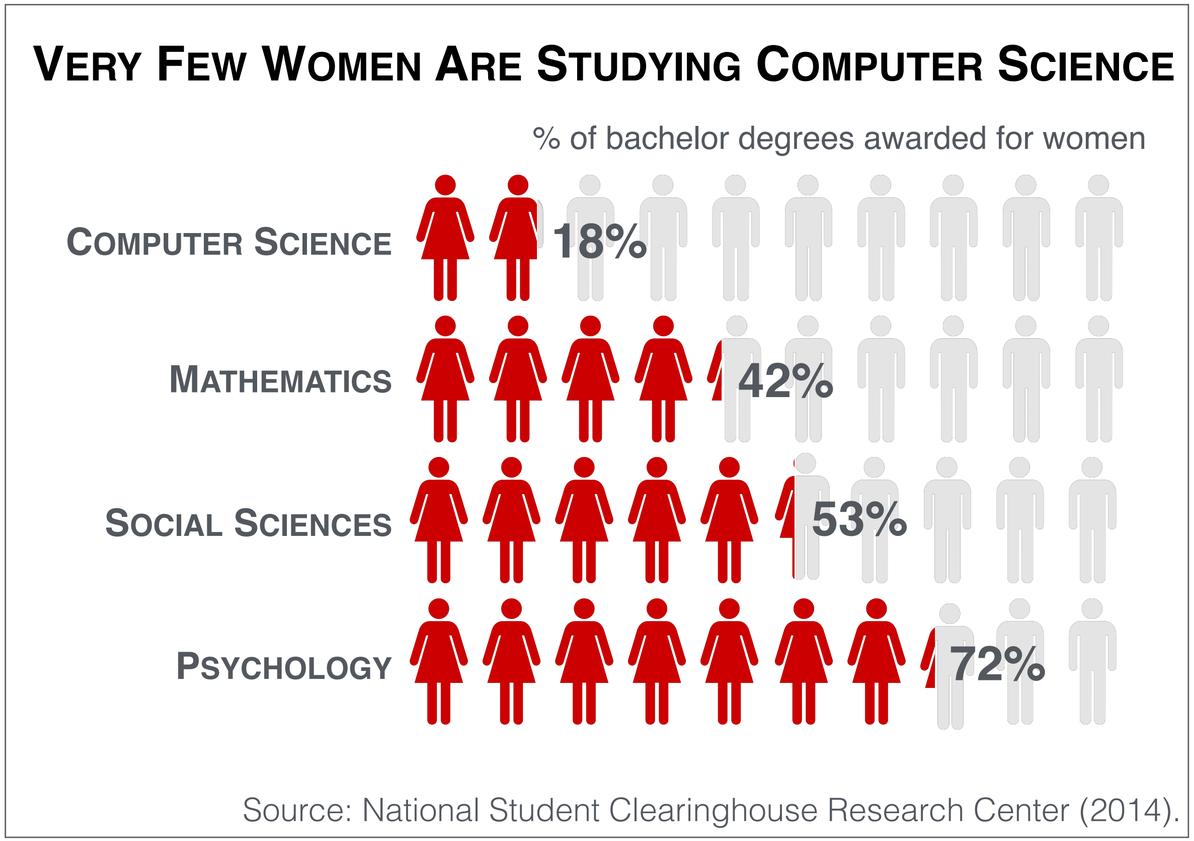
DATA & MESSAGE REDUNDANCY



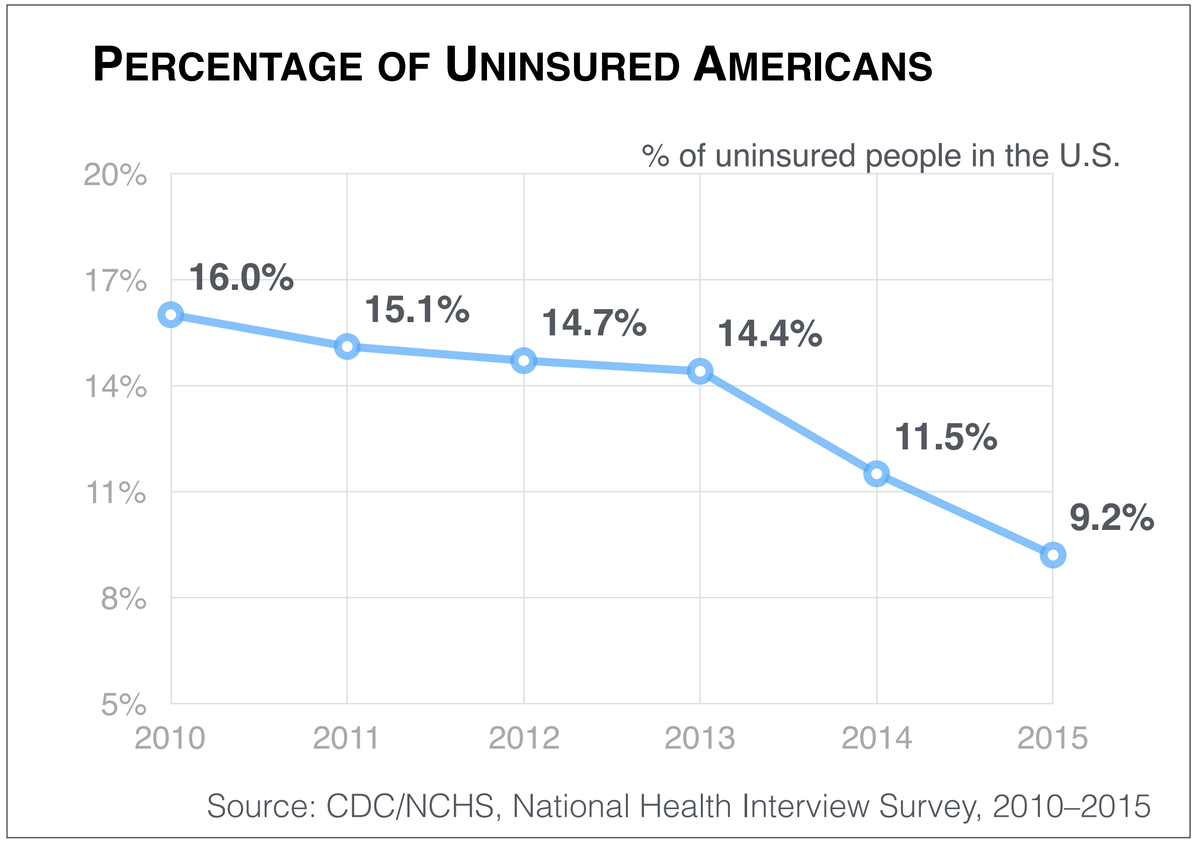
WITHOUT PICTOGRAM



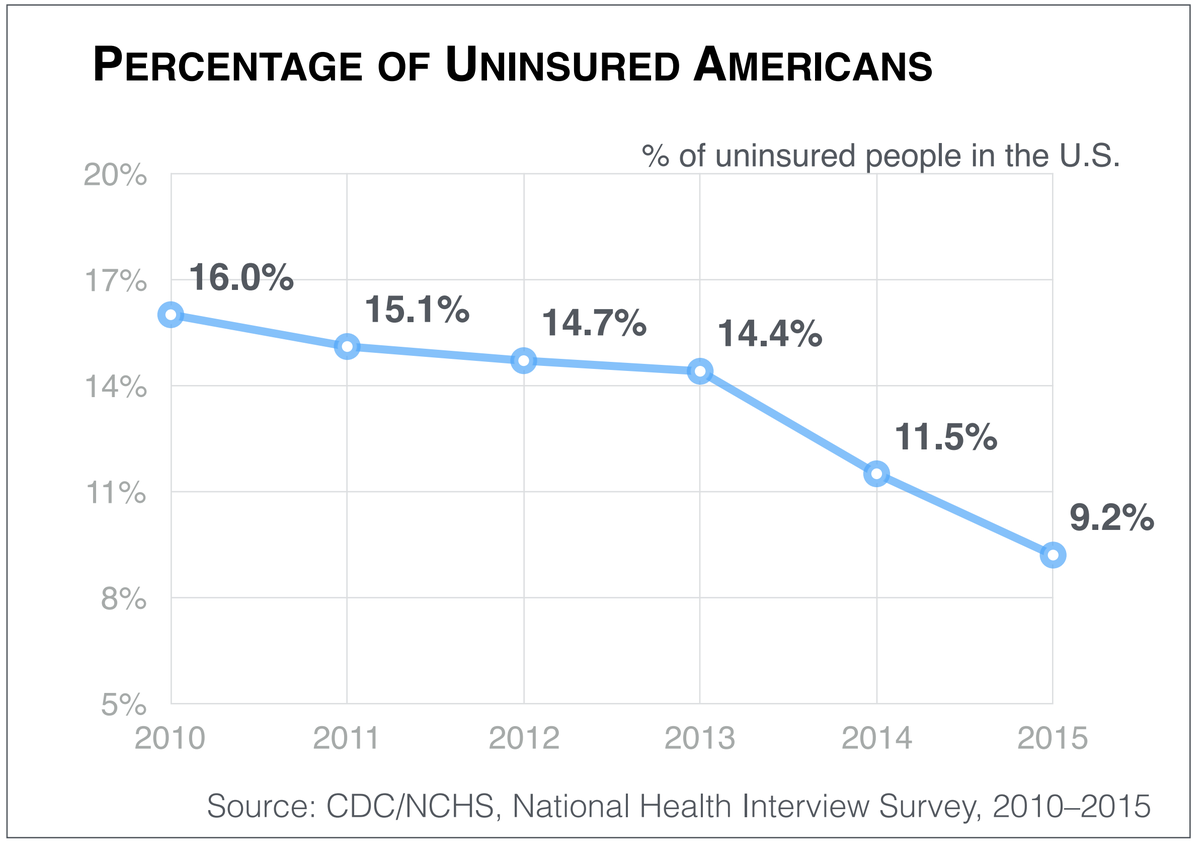
WITHOUT PICTOGRAM



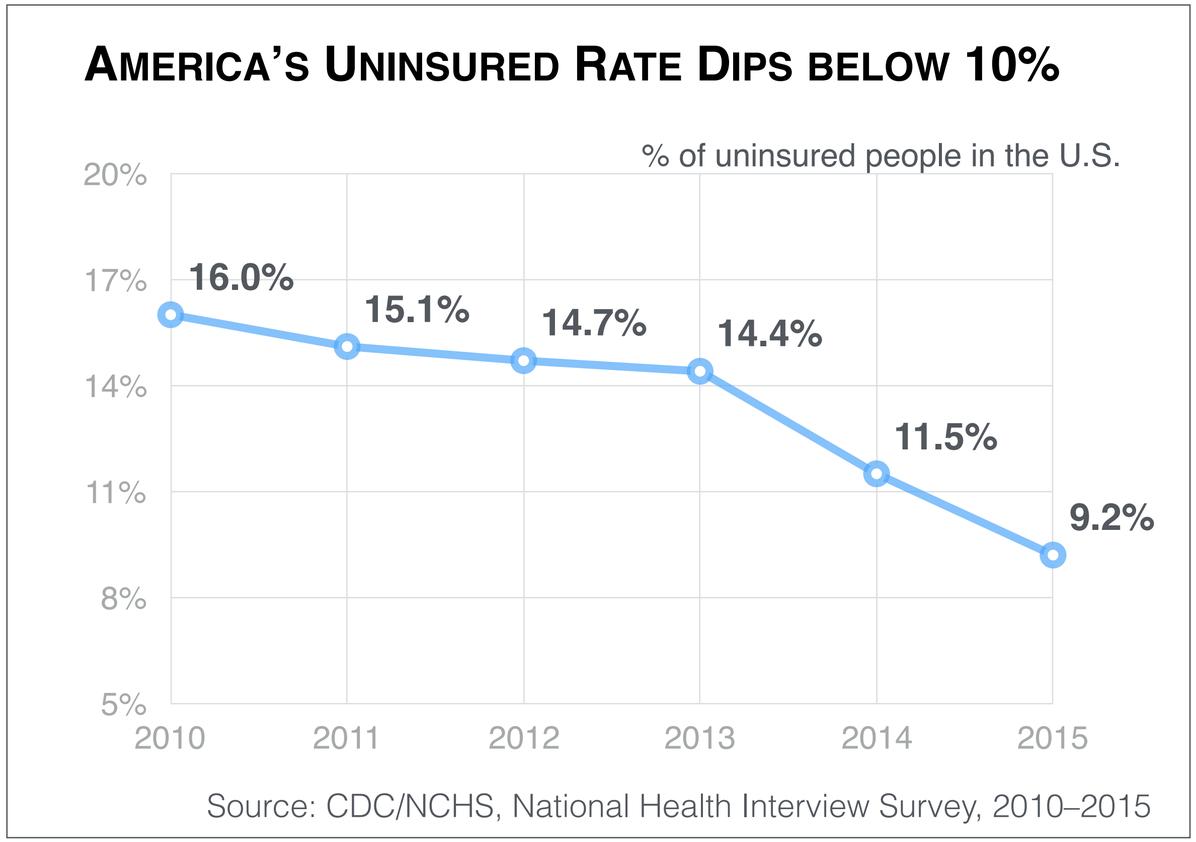
WITH PICTOGRAM


PERCENTAGE OF UNINSURED AMERICANS
% of uninsured people in the U.S.
20%
17%
14%
11%
8%
5%
16.0%
15.1%
14.7%
14.4%
11.5%
9.2%
2010
2011
2012
2013
2014
2015
Source: CDC/NCHS, National Health Interview Survey, 2010–2015

BAD TITLE



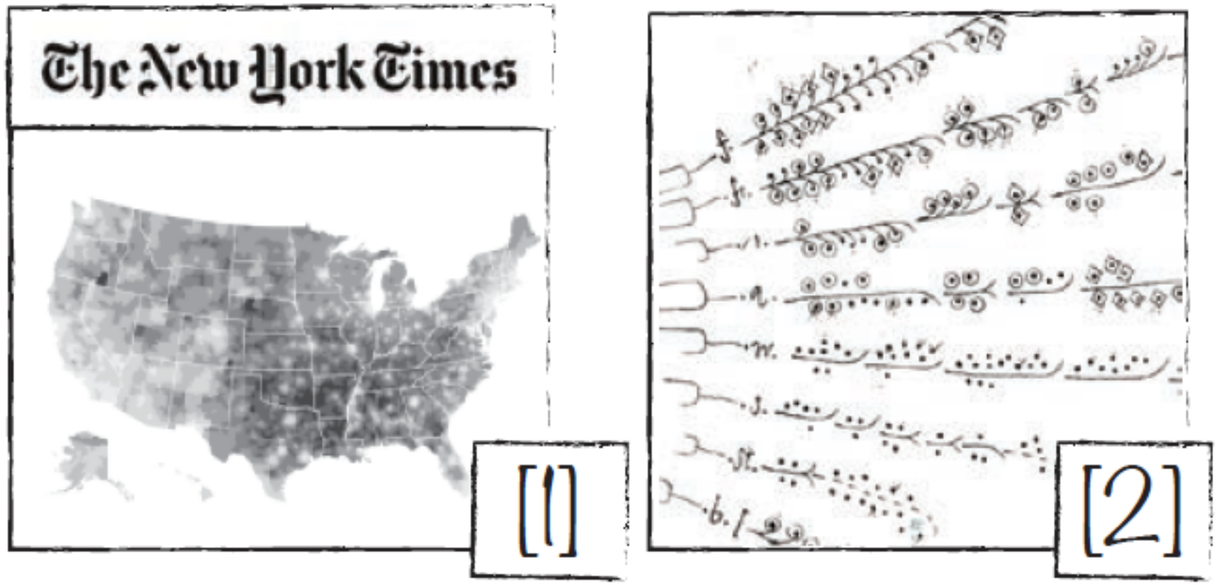
BAD TITLE



GOOD TITLE

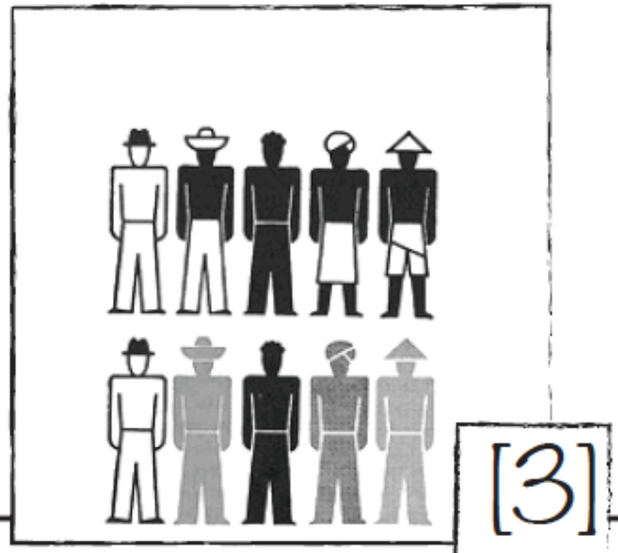
Telling Stories with Data

Visualizations can be effective in showing data,

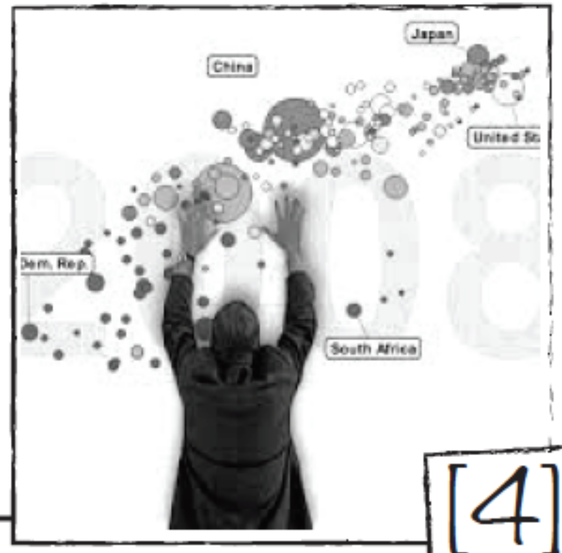


[1]

[2]



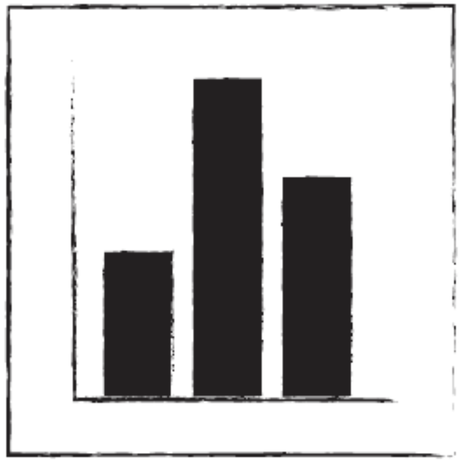
[3]



[4]

Bach et al. CG&A'17

but a single picture alone may not be able to explain nor to engage an audience in decoding a message.



How do we engage an audience?

How do we break down complexity?

How do we guide the audience?

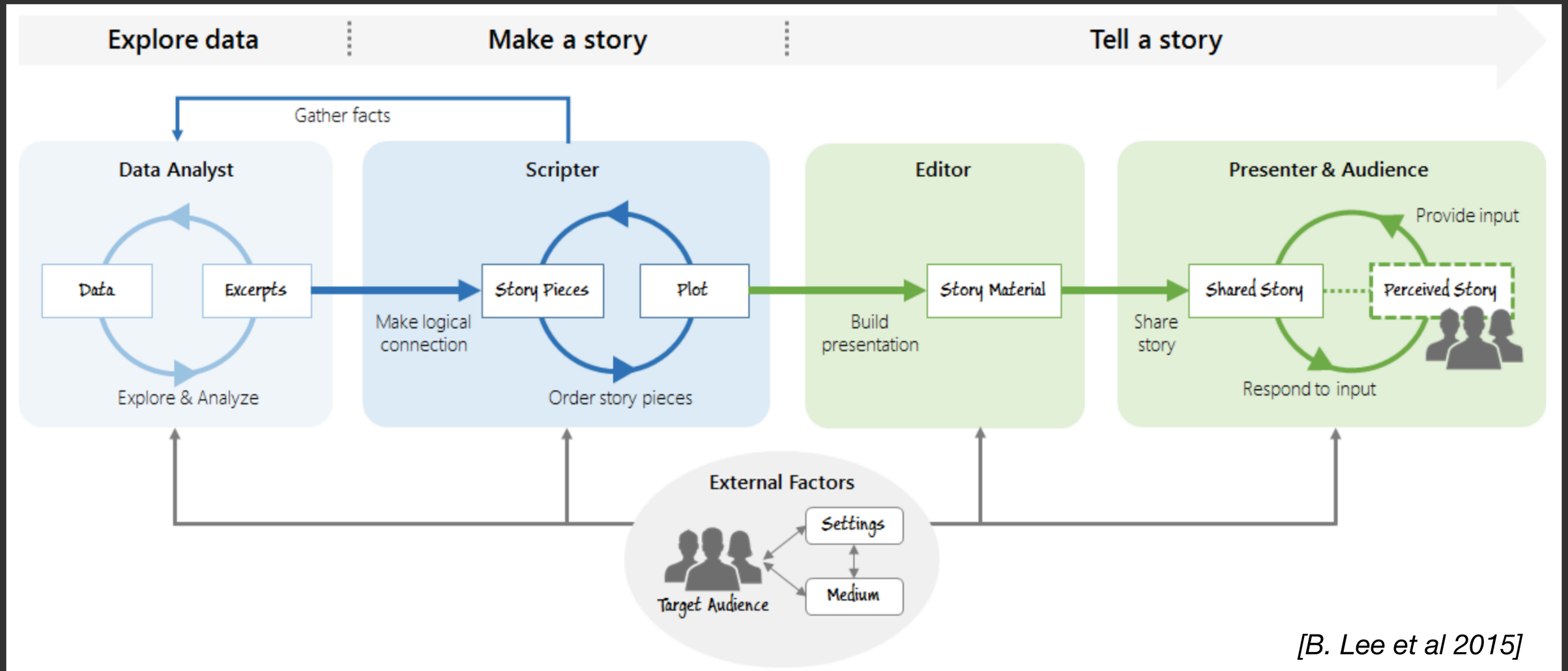
Bach et al. CG&A'17

Stories

Stories are the most powerful delivery tool for information, more powerful and enduring than any other art form


Use of elements from storytelling
to convey compelling data stories

Data Storytelling Process: transforming data into visual stories.



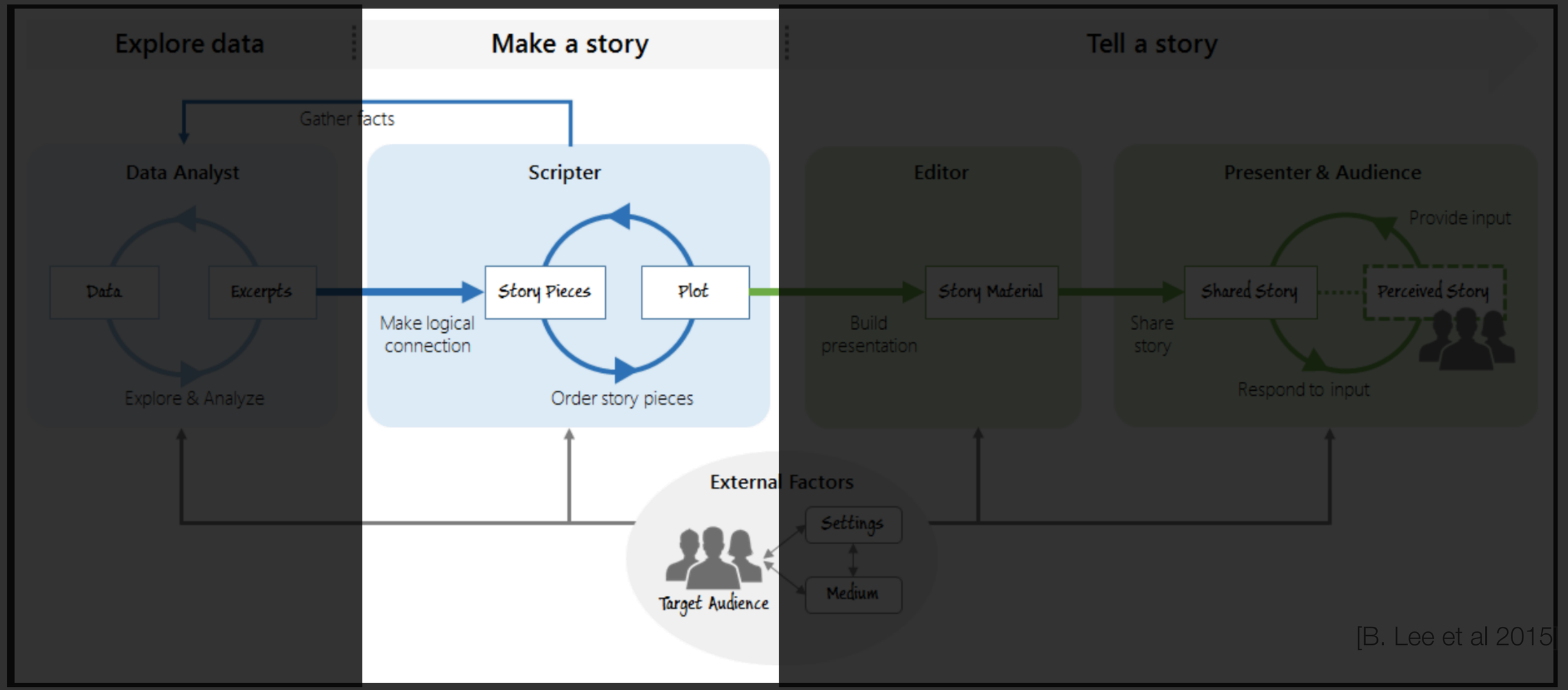
[B. Lee et al 2015]

Data Storytelling Process: transforming data into visual stories.

Iterative exploration:
Traditional visualization process

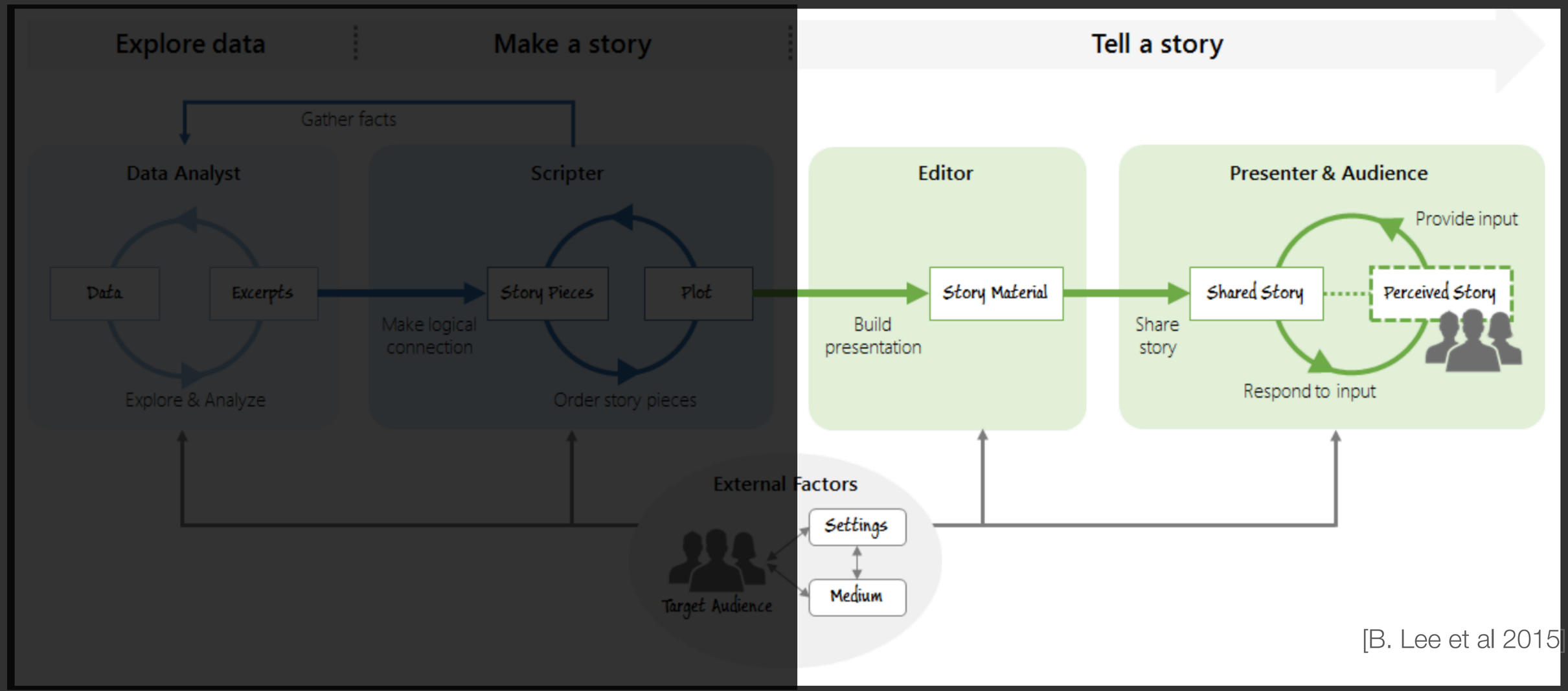
[B. Lee et al 2015]

Data Storytelling Process: transforming data into visual stories.



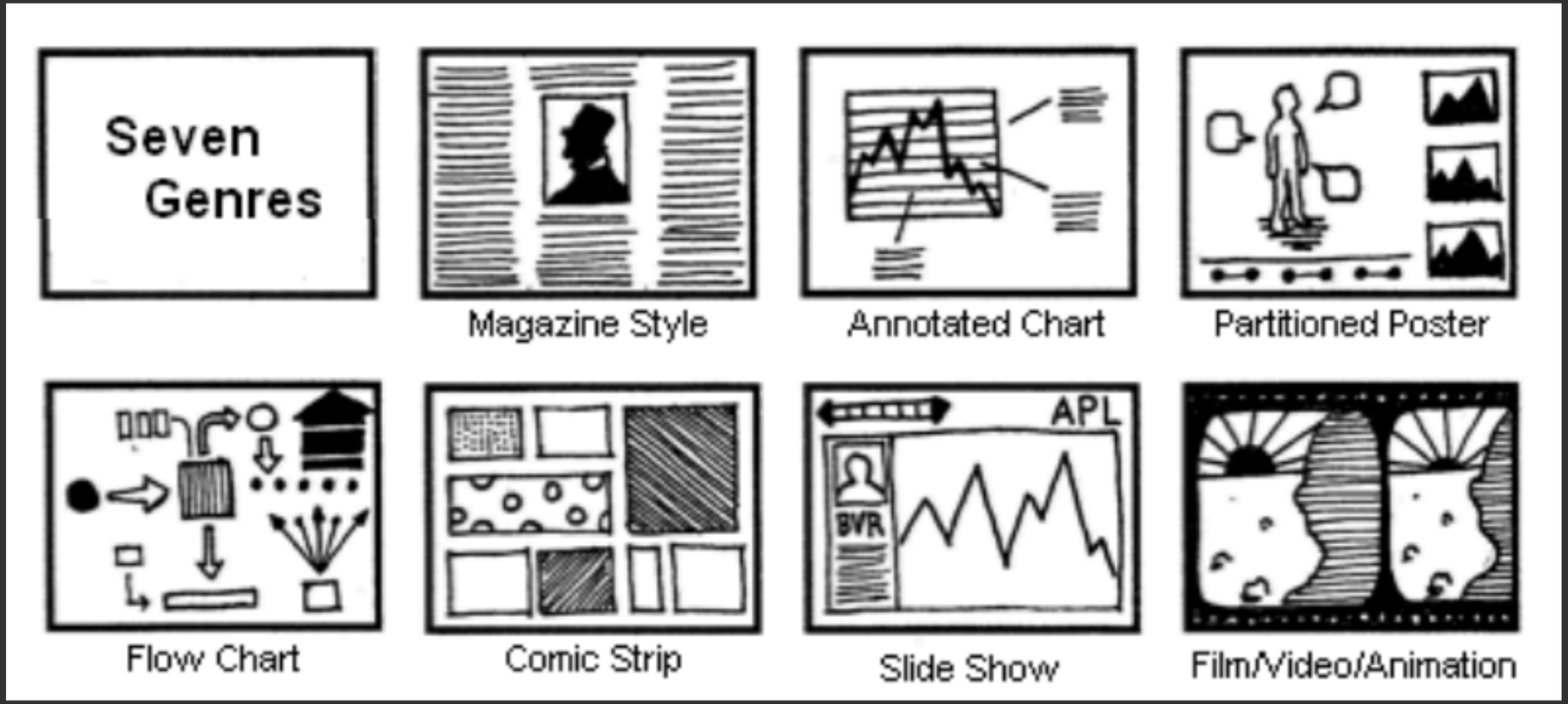
[B. Lee et al 2015]

Data Storytelling Process: transforming data into visual stories.



[B. Lee et al 2015]

Genres of Data Stories



Narration Styles

Author-Driven

- Strict ordering of scenes
- Heavy messaging
- No interactivity

Reader-Driven

- No prescribed ordering
- No messaging
- Free interactivity

Narration Styles

Author-Driven

- Linear ordering of scenes
- Heavy messaging
- No interactivity

+

Reader-Driven

- No prescribed ordering
- No messaging
- Free interactivity

Author-Driven


1967
2.5
3
3.5
6.5
7

Reader-Driven

Epoch
000,000

Learning rate: 0.03

Activation: Tanh

Regularization: None

Regularization rate: 0

Problem type: Classification

DATA

Which dataset do you want to use?

Ratio of training to test data: 50%

Noise: 0

Batch size: 10

REGENERATE

FEATURES

Which properties do you want to feed in?

X_1

X_2

X_1^2

X_2^2

X_1X_2

$\sin(X_1)$

$\sin(X_2)$

2 HIDDEN LAYERS

4 neurons

2 neurons

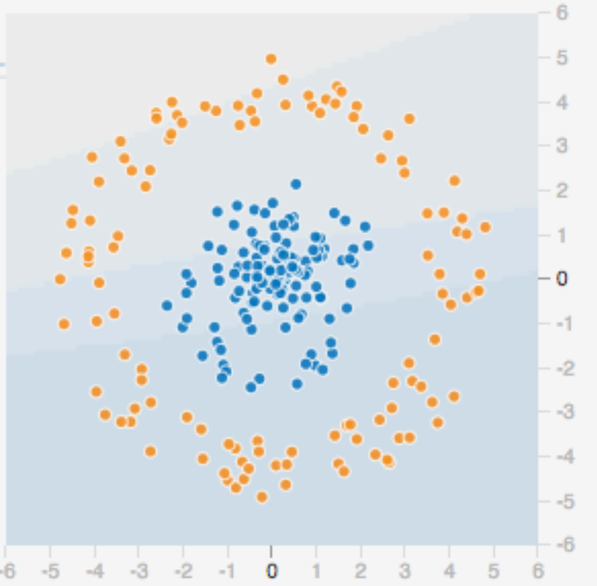
*The outputs are mixed with varying **weights**, shown by the thickness of the lines.*

*This is the output from one **neuron**. Hover to see it larger.*

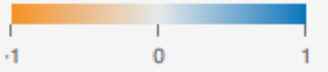
OUTPUT

Test loss 0.513

Training loss 0.500



Colors shows data, neuron and weight values.



Show test data

Discretize output

Author-Driven + Reader-Driven

Published: February 2, 2010

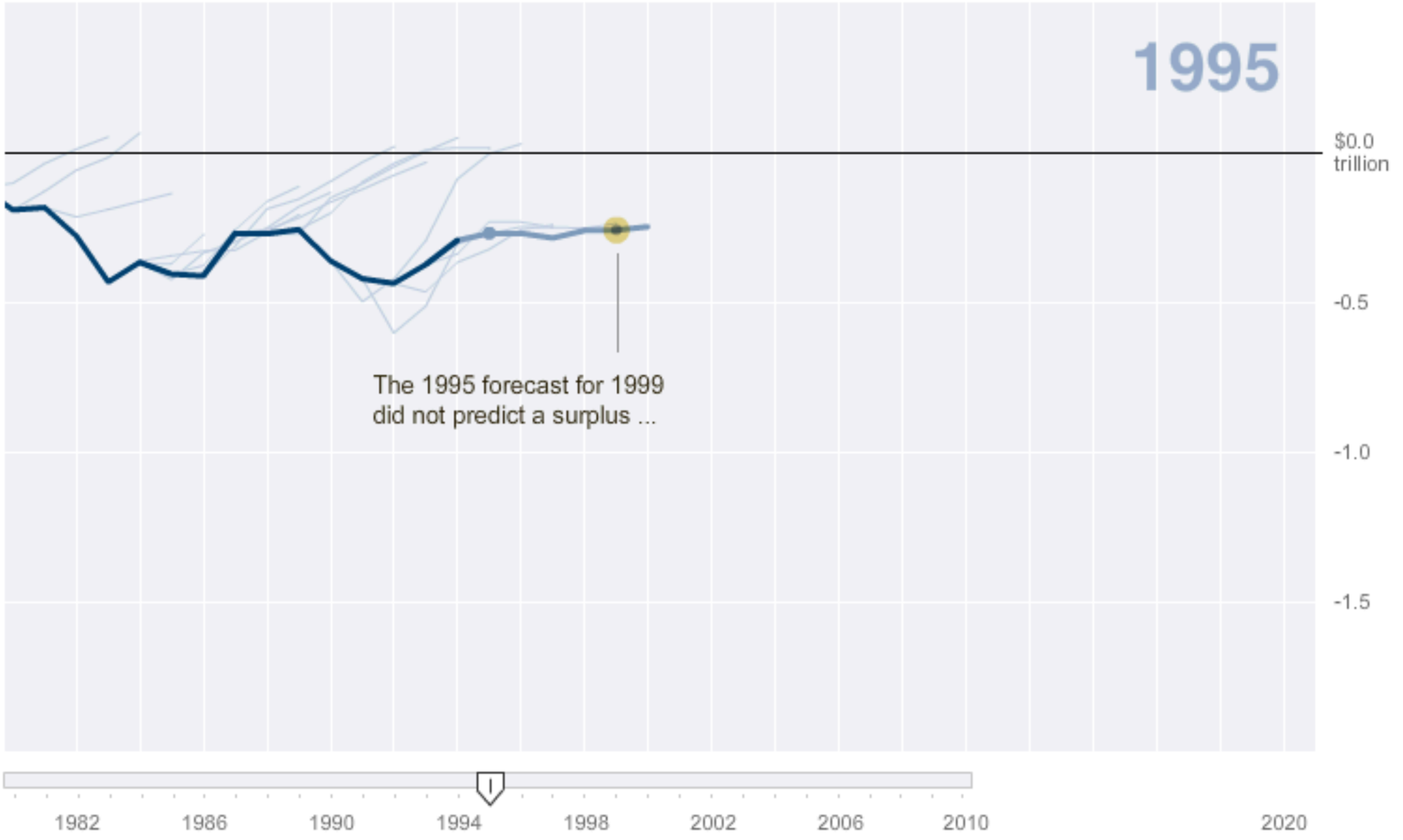
Budget Forecasts, Compared With Reality

Just two years ago, surpluses were predicted by 2012. How accurate have past White House budget forecasts been?

Past forecasts

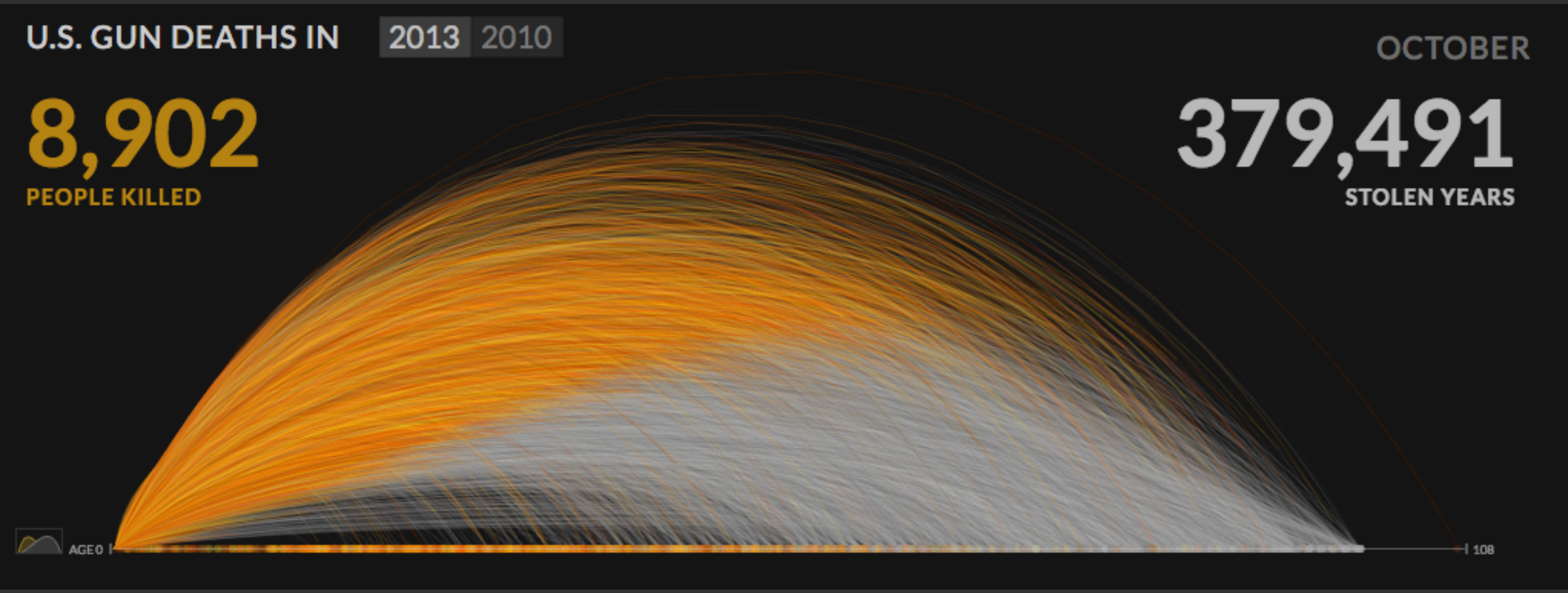
Even that may be an understatement. In the last 30 years, about 80 percent of four-year deficit forecasts have been too optimistic.

The early Clinton budgets — which failed to predict the surpluses that were generated, in part, by a stock market bubble — are the only major exception.



By AMANDA COX | Send Feedback

Source: Office of Management and Budget


U.S. GUN DEATHS IN
2013
2010
OCTOBER
8,902
PEOPLE KILLED
379,491
STOLEN YEARS
AGE 0
108

THE Fallen OF World War II

Visual Narrative Design

Employ Narrative Structure

Bach et al 2017

755

Steroids or Not, the Pursuit Is On

Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.

Lines are cumulative home runs.

Hank Aaron
755 homers
23 seasons

Babe Ruth
714 homers
22 seasons

Barry Bonds
708 homers
20 seasons

Bonds takes lead
Home runs after 16 seasons

Bonds	567
Aaron	554
Ruth	516

755
714
23 seasons
22 seasons
20 seasons
Bonds was injured last season. He played 14 games and hit 8 homers.

600
400
200
0
14th season

According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.

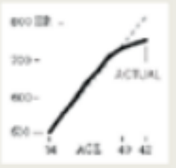
5 seasons
10
15

Homer Pace After Age 34

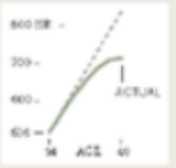
If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.

PROJECTED PACE BASED ON AVERAGE OF PREVIOUS FIVE SEASONS

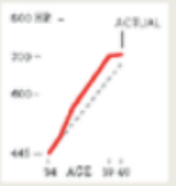
Aaron
Actual homers slightly outpace projected homers for five seasons.



Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.

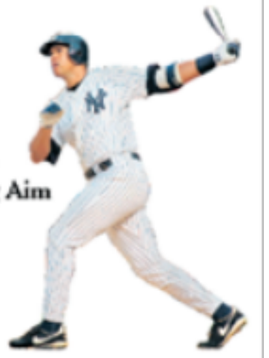


Bonds
From age 35 to 39, he averaged 14 more homers a season than projected.

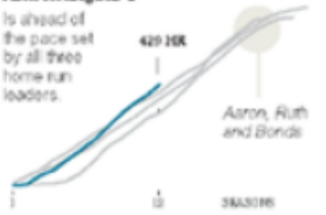


Note: Ages as of July 1 of each season.

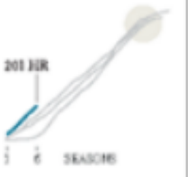
Others Taking Aim



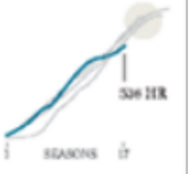
Alex Rodriguez ▲
Is ahead of the pace set by all three home run leaders.



Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.

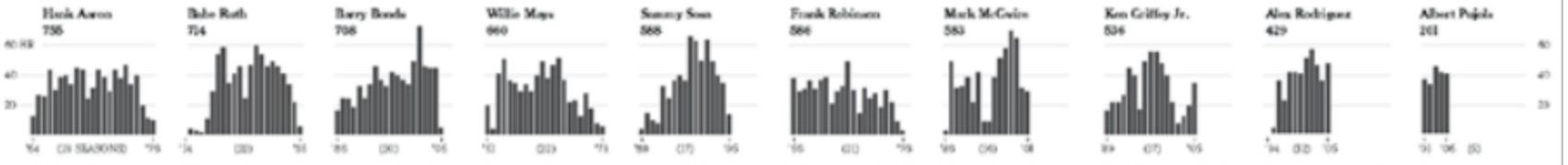


Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.



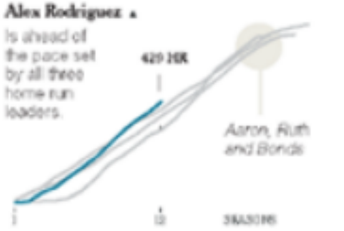
Differing Paths to the Top of the Charts

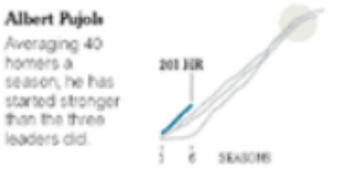
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).

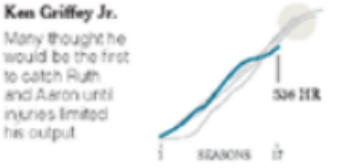


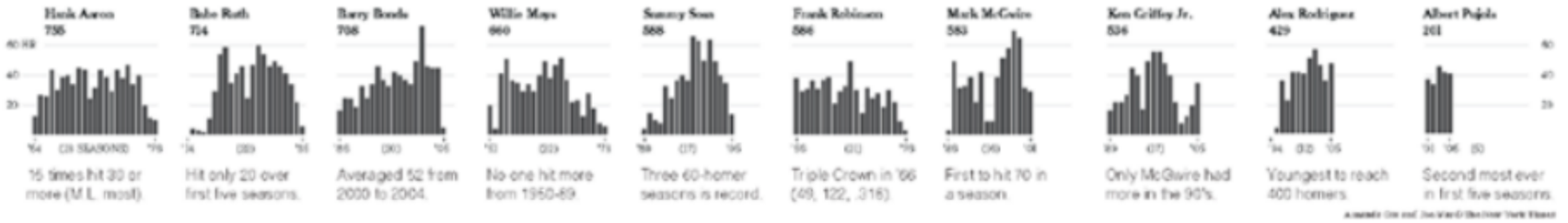

755
Steroids or Not, the Pursuit Is On
Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.
Lines are cumulative home runs.
Hank Aaron
755 homers
23 seasons
Babe Ruth
714 homers
22 seasons
Barry Bonds
708 homers
20 seasons
Bonds takes lead
Home runs after 16 seasons
Bonds 587
Aaron 554
Ruth 516
BEGINNING
According to allegations in a book about Bonds, he began taking steroids before the 1999 season, his 14th in the league. Two seasons later, he hit 73 home runs, surpassing Aaron's career pace.
Homer Pace After Age 34
If the accusations are correct, Bonds was 34 in his first season on steroids. Here are projected home run paces for each player after age 34.
PROJECTED PACE BASED ON AVERAGE OF PREVIOUS FIVE SEASONS
Aaron
Actual homers slightly outpace projected homers for five seasons.
Ruth
Averaged 46.4 homers a season from age 30 to 34. Averaged 42.5 for next four seasons.
Bonds
From age 35 to 39, he averaged 16 more homers a season than projected.
Note: Ages as of July 1 of each season.


Others Taking Aim

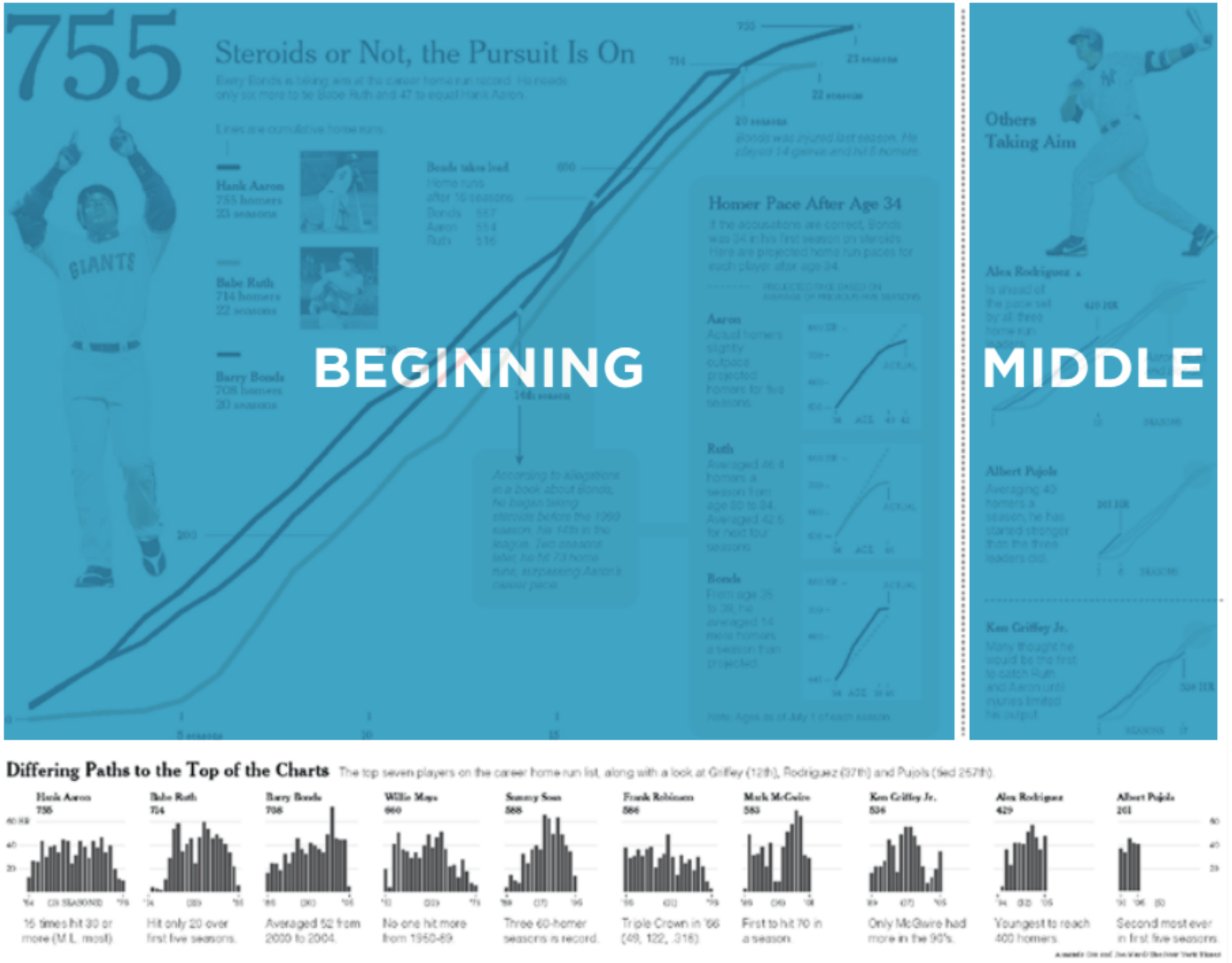

Alex Rodriguez
Is ahead of the pace set by all three home run leaders.
429 HR
Aaron, Ruth and Bonds
SEASONS


Albert Pujols
Averaging 40 homers a season, he has started stronger than the three leaders did.
201 HR
SEASONS

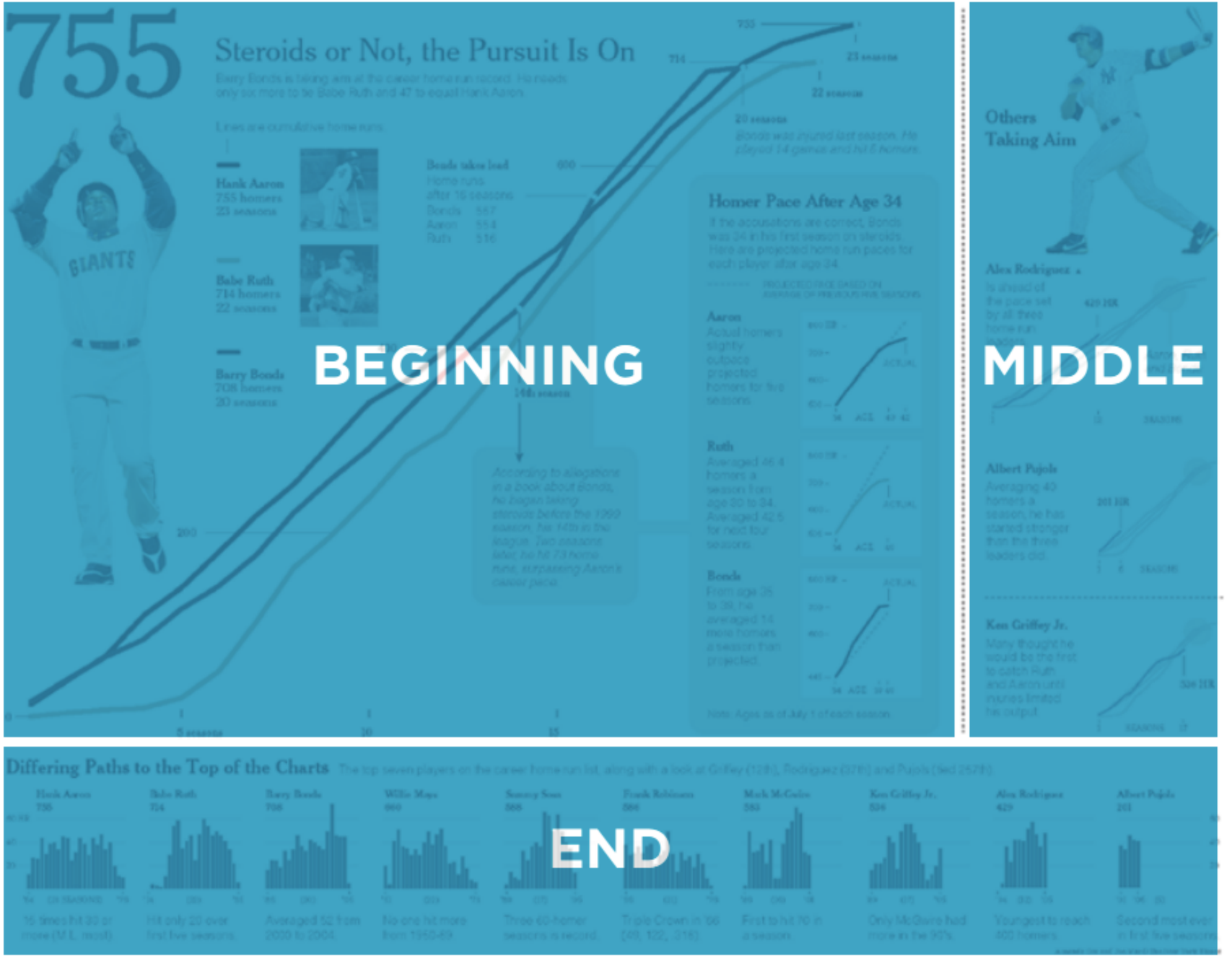

Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.
536 HR
SEASONS


Differing Paths to the Top of the Charts
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).
Hank Aaron
755
16 times hit 30 or more (M.L. most).
Babe Ruth
714
Hit only 20 over first five seasons.
Barry Bonds
708
Averaged 52 from 2000 to 2004.
Willie Mays
660
No one hit more from 1950-69.
Sammy Sosa
588
Three 60-homer seasons is record.
Frank Robinson
586
Triple Crown in '66 (49, 122, .316).
Mark McGwire
583
First to hit 70 in a season.
Ken Griffey Jr.
536
Only McGwire had more in the 90's.
Alex Rodriguez
429
Youngest to reach 400 homers.
Albert Pujols
201
Second most ever in first five seasons.

E. Segel


755
Steroids or Not, the Pursuit Is On
Barry Bonds is taking aim at the career home run record. He needs only six more to tie Babe Ruth and 47 to equal Hank Aaron.
Lines are cumulative home runs.
Hank Aaron
755 homers
23 seasons
Babe Ruth
714 homers
22 seasons
Barry Bonds
708 homers
20 seasons
Bonds takes lead
Home runs after 16 seasons
Bonds 587
Aaron 554
Ruth 516
BEGINNING
MIDDLE
Homer Pace After Age 34
Aaron
Ruth
Bonds
Others Taking Aim
Alex Rodriguez
Albert Pujols
Ken Griffey Jr.
Differing Paths to the Top of the Charts
The top seven players on the career home run list, along with a look at Griffey (12th), Rodriguez (37th) and Pujols (tied 257th).
Hank Aaron 755
Babe Ruth 714
Barry Bonds 708
Willie Mays 660
Sammy Sosa 588
Frank Robinson 586
Mark McGwire 583
Ken Griffey Jr. 534
Alex Rodriguez 429
Albert Pujols 201
16 times hit 30 or more (M.L. record)
Hit only 20 over first five seasons.
Averaged 52 from 2000 to 2004.
No one hit more from 1950-69.
Three 60-homer seasons is record.
Triple Crown in '66 (49, 122, .316).
First to hit 70 in a season.
Only McGwire had more in the 90's.
Youngest to reach 400 homers.
Second most ever in first five seasons.

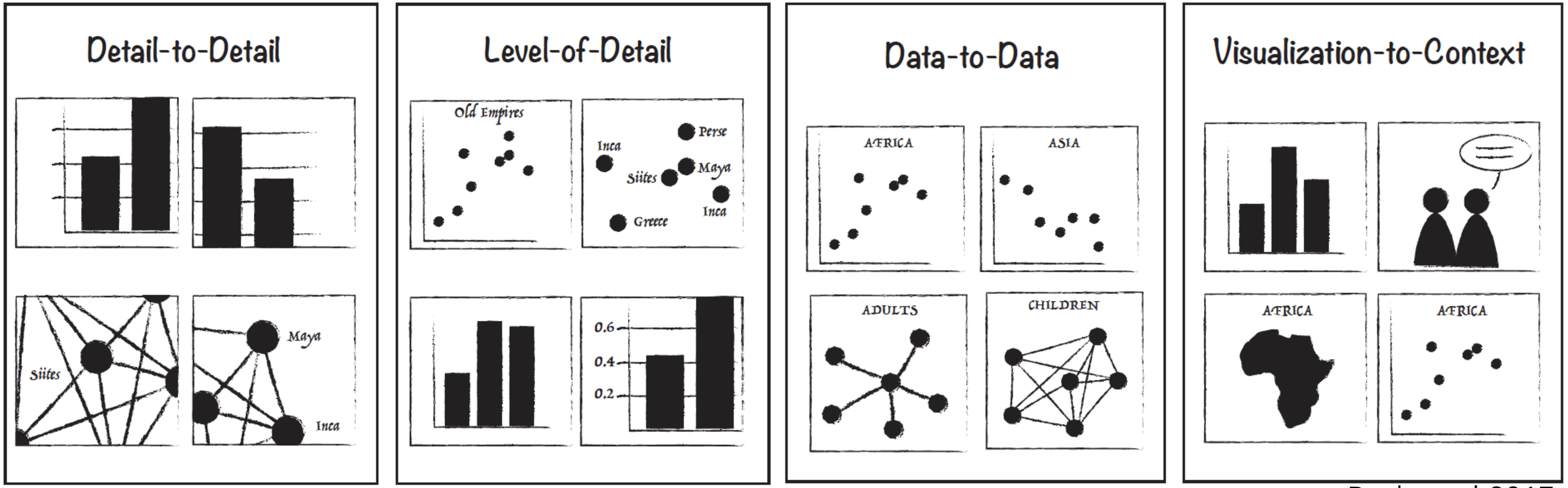
E. Segel


755
Steroids or Not, the Pursuit Is On
Hank Aaron
755 homers
23 seasons
Babe Ruth
714 homers
22 seasons
Barry Bonds
708 homers
20 seasons
Homer Pace After Age 34
Others
Taking Aim
Alex Rodriguez
Albert Pujols
Ken Griffey Jr.
BEGINNING
MIDDLE
END
Differing Paths to the Top of the Charts

E. Segel

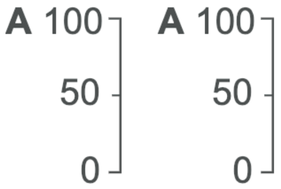
Use Multiple Charts

To break down the complexity of the story and progressively reveal different facets of the data.

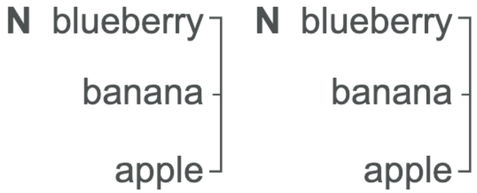


Bach et al 2017

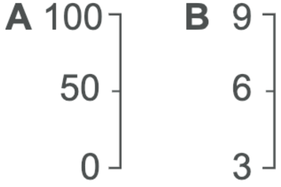
C1.1 Same Field, Same XY Scale



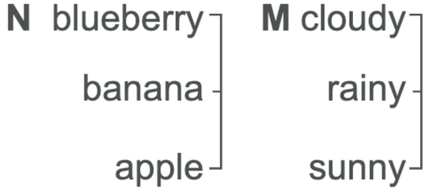
C1.2 Same Field, Same Values in Same Order



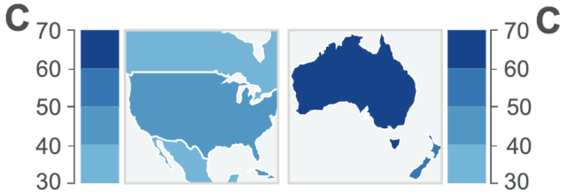
C2.1 Different Fields, Different XY Domains



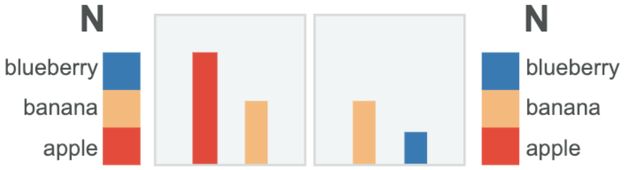
C2.2 Different Fields, Different Nominal Values



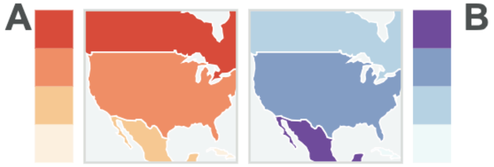
C1.3 Same Field, Same Quantitative Color Scale



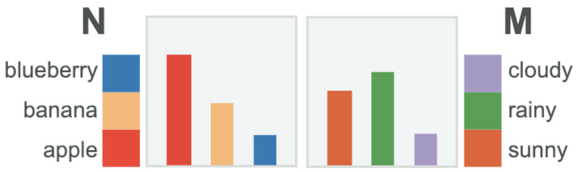
C1.4 Same Field, Same Value-Color Mapping



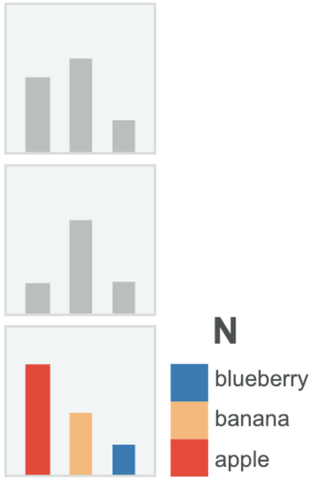
C2.3 Different Fields, Non-Overlapping Hues



C2.4 Different Fields, Non-overlapping Palettes



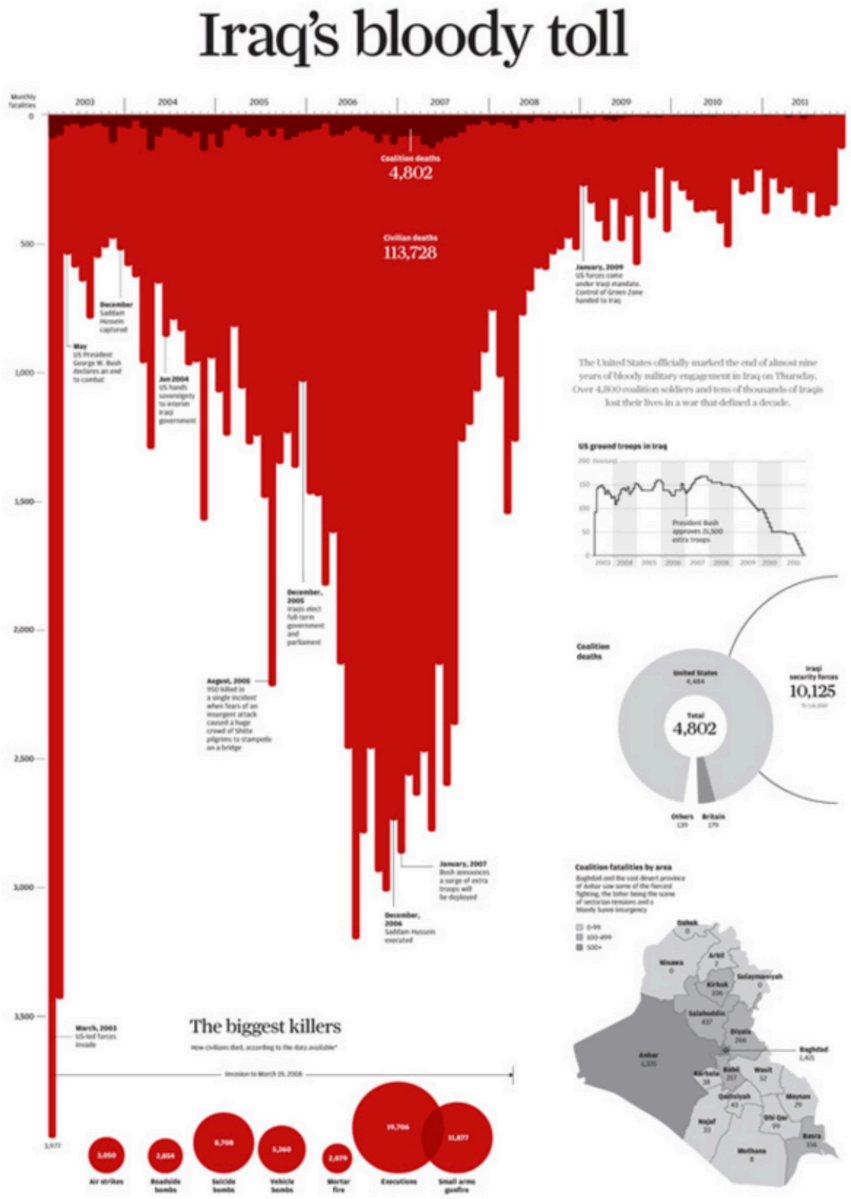
C2.5 Non-overlapping Nominal and Constant Colors



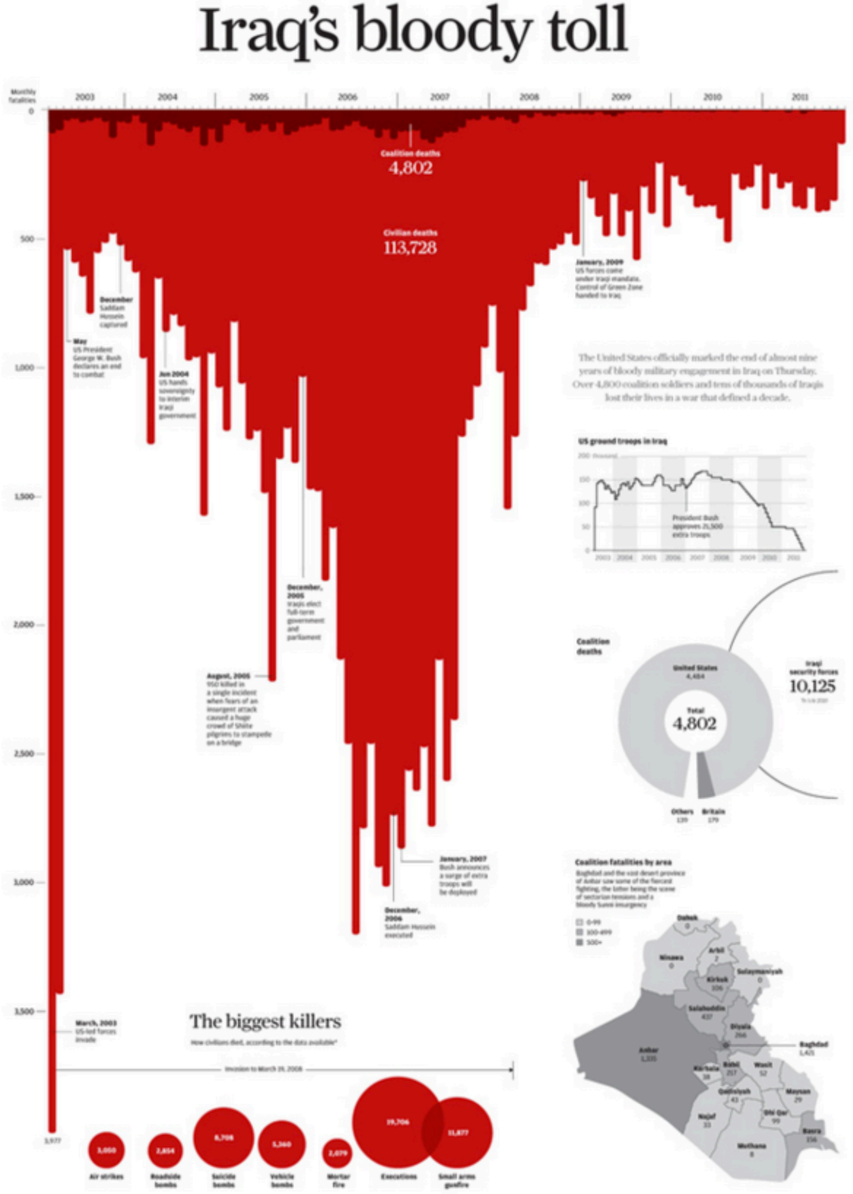
Keeping Multiple Views Consistent

Qu et al 2018

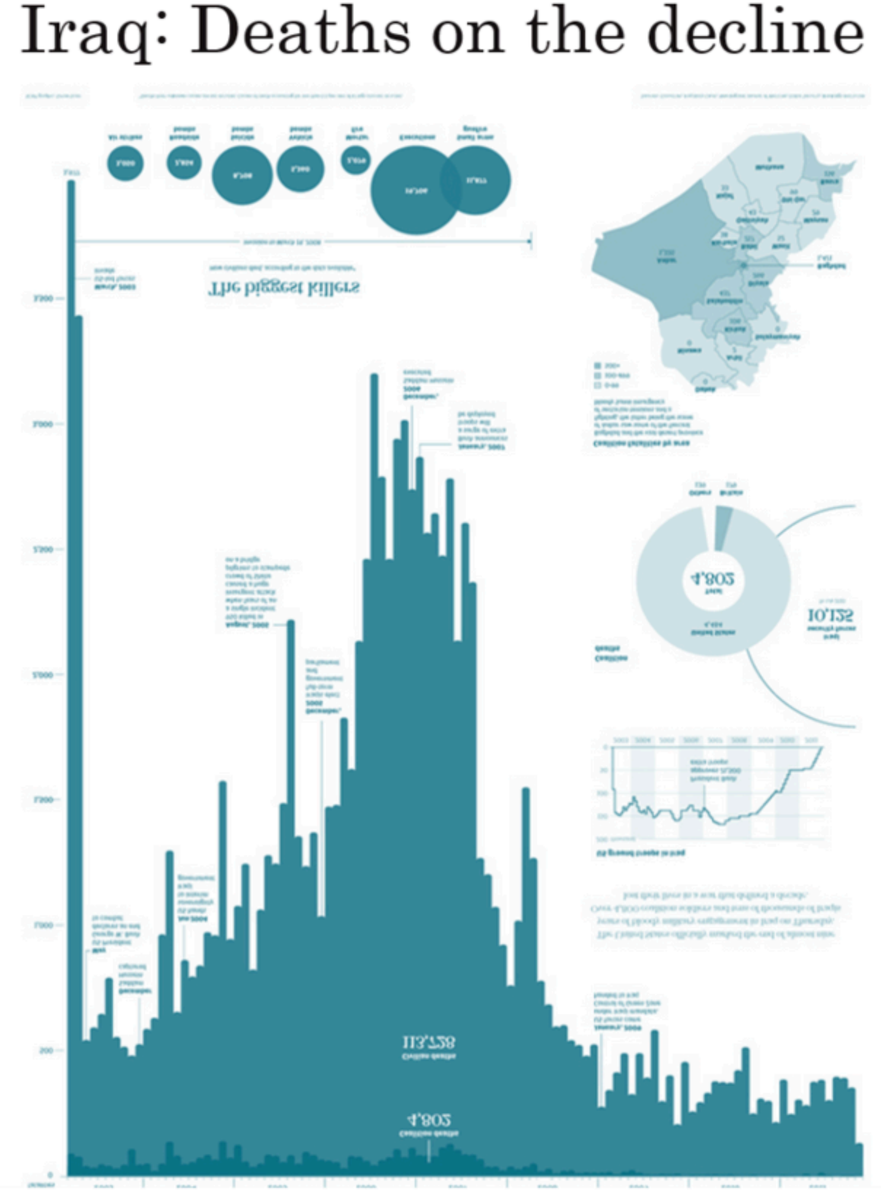
Storytelling, Double-edged Sword?



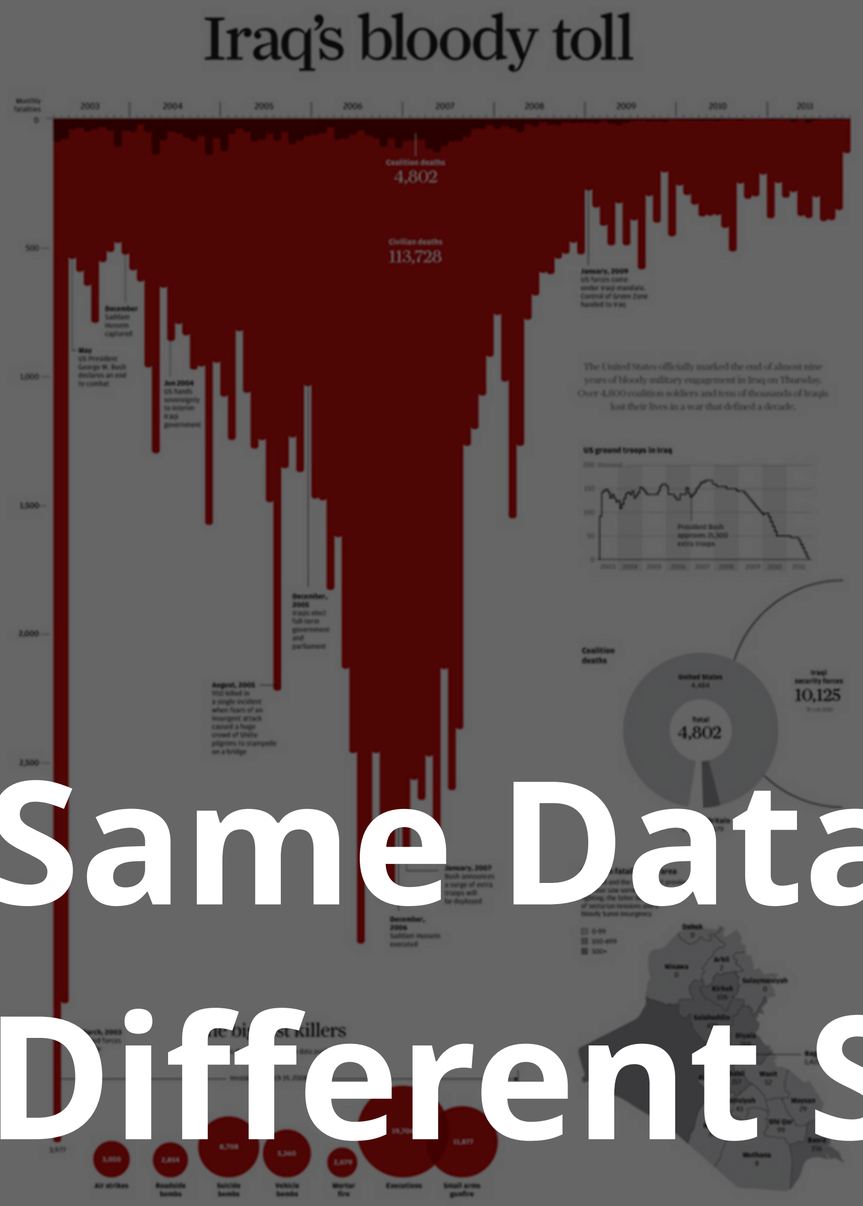
[South China Morning Post 2011]



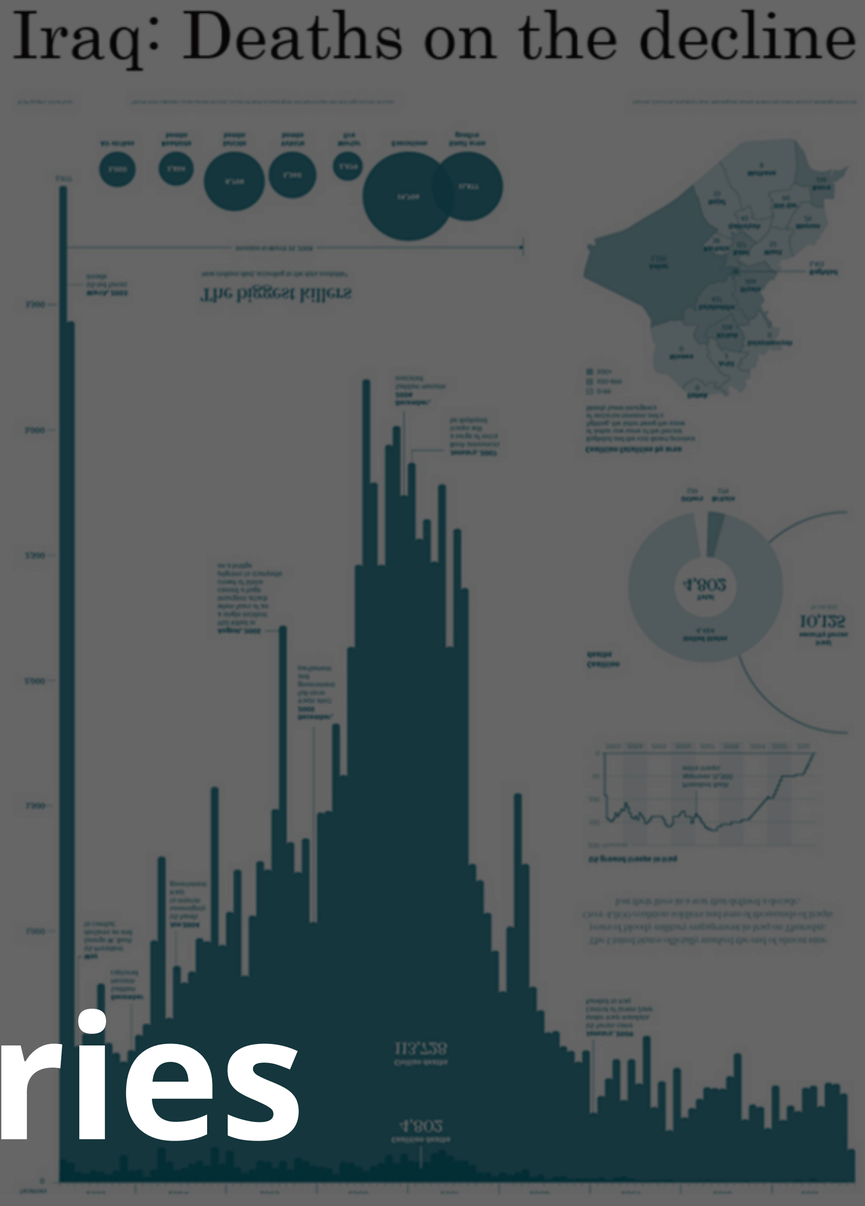
[South China Morning Post 2011]



Flipped upside down….



[South China Morning Post 2011]



Flipped upside down….

Same Data, Different Stories

Presentation & Storytelling in Tableau

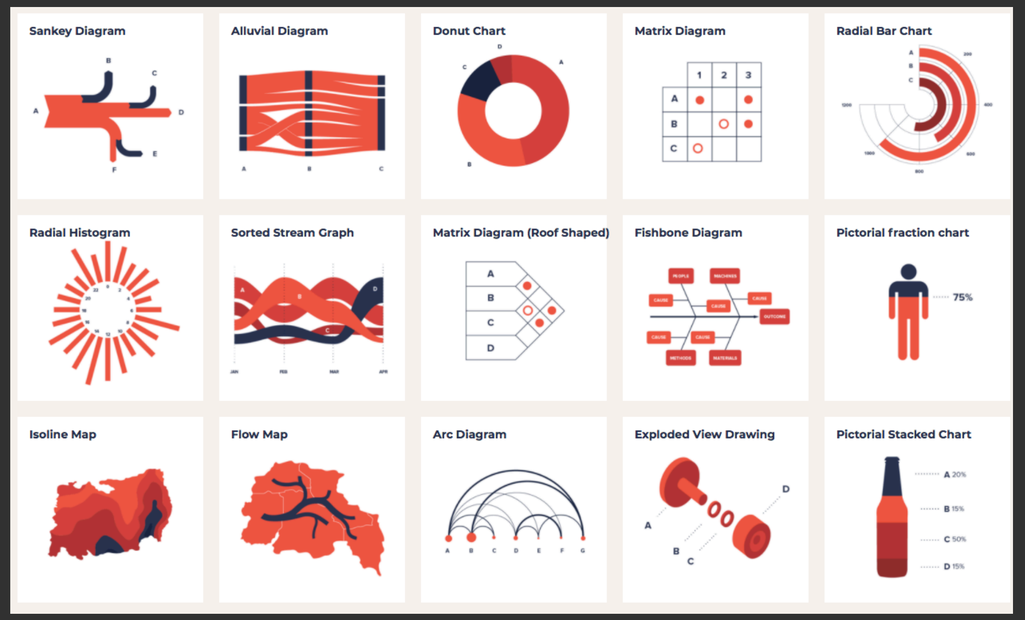
Dashboard

Annotation

Story Points

Next

Advanced visualizations



5 min break