

Storytelling with Data

ONE LOVE. ONE FUTURE.

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

Position



Length



Slope



Angle



Area



Intensity



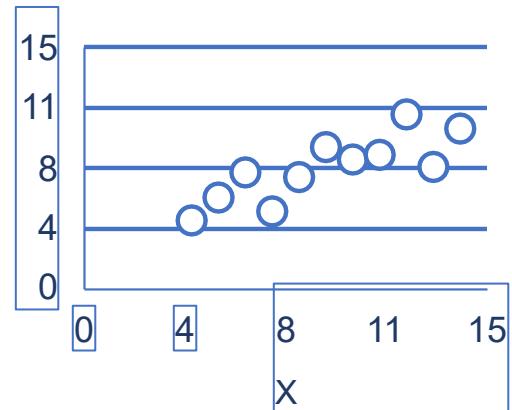
Color



Shape



Visualization



Position (x, y)

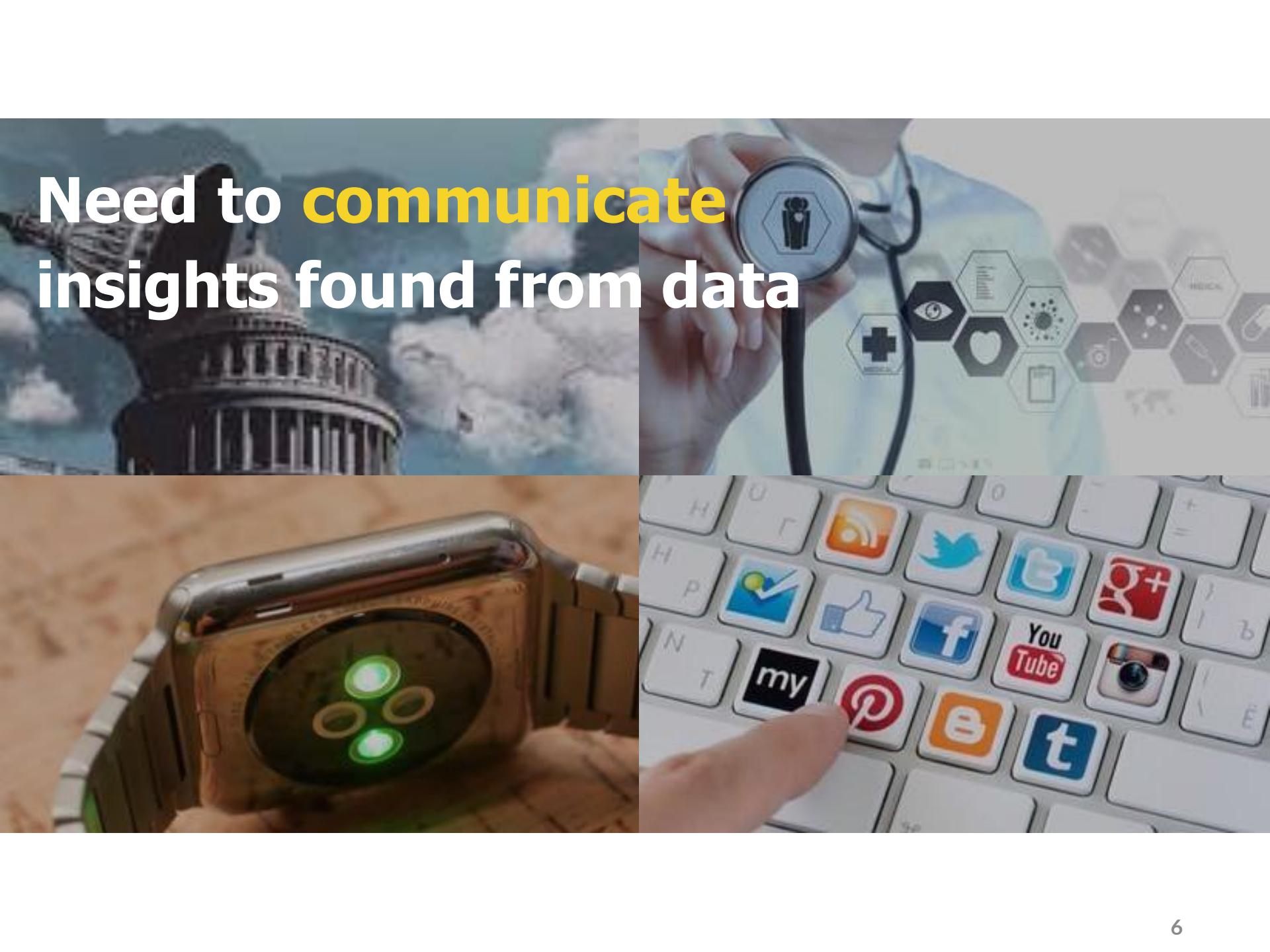
- Expressiveness
 - A set of facts is expressible in a visualization if it expresses all the facts and only the facts in the data.
- Effectiveness
 - A visualization is more effective than another one if the information conveyed is more readily perceived.

Data Insights



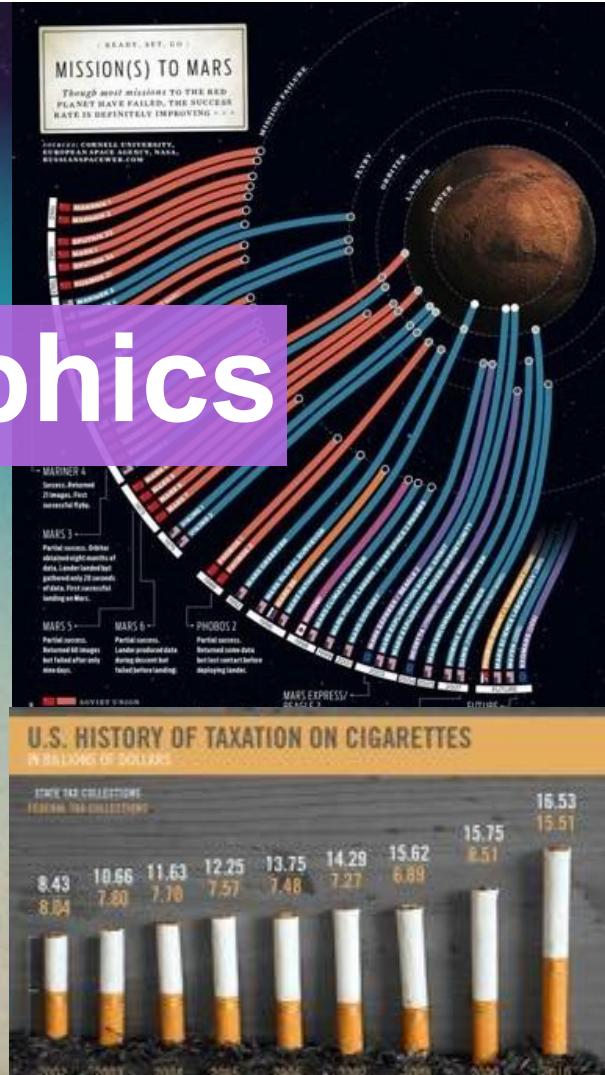
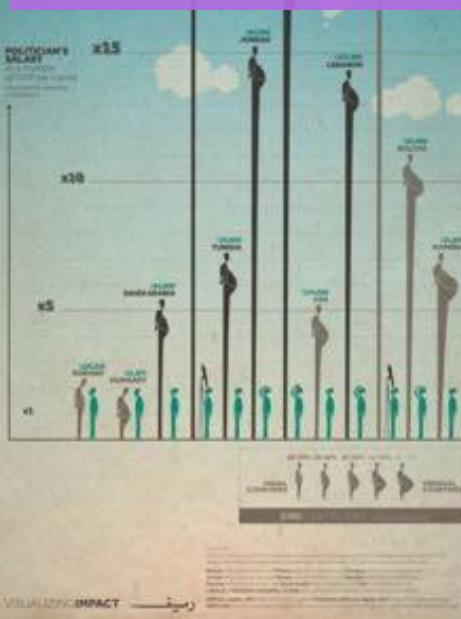
Exploratory

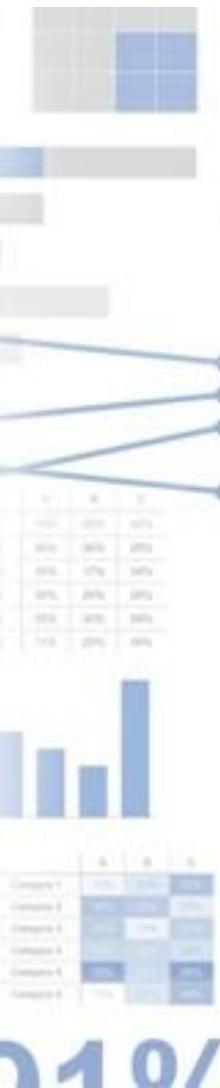
- Data Centered
- Domain Experts
- Analysis
- Desktop
- In-Lab



Need to **communicate** insights found from data

Infographics





cole nussbaumer knaflic

storytelling with **data**

a data
visualization
guide for
business
professionals

WILEY

Data Insights Messages

Exploratory

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

Beyond Exploratory

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

1. Understand the context
2. Choose an appropriate visual display
3. Use redundant encoding
4. Eliminate clutter
5. Focus your audience's attention
6. Add explanations
7. Tell a story

Understand the context

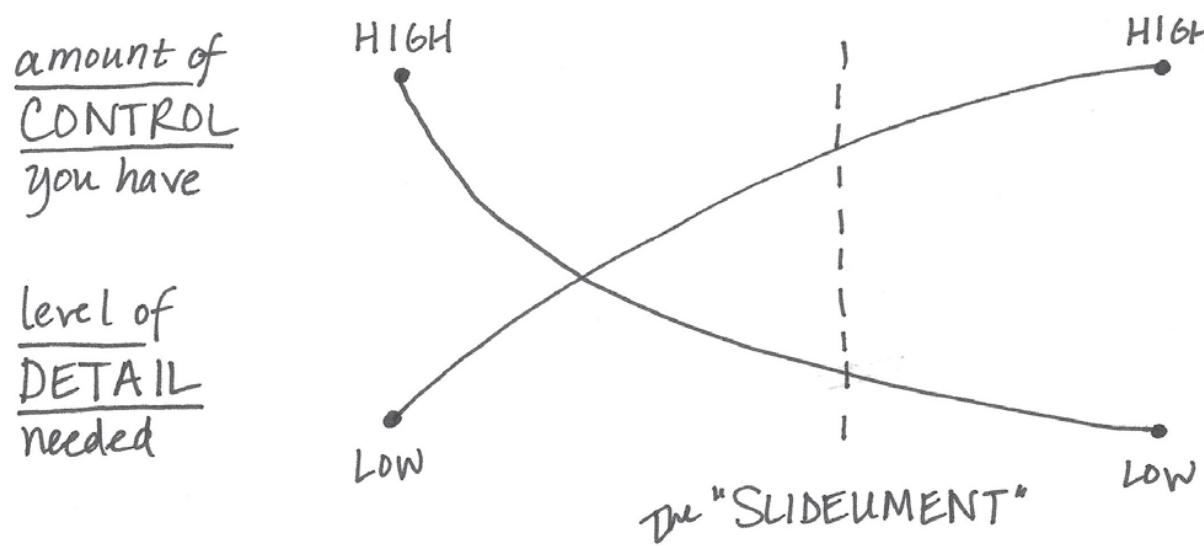
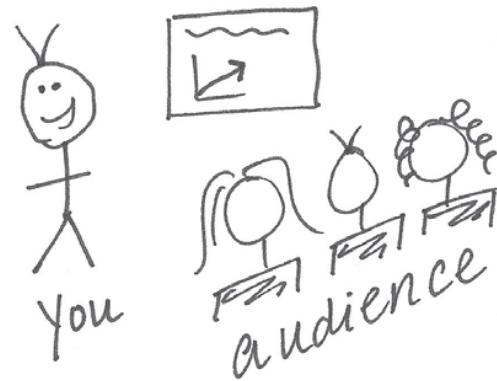
Explanatory analysis

- The point of communicating our analysis to our audience
 - Be in the explanatory space
 - You have a specific thing you want to explain, a specific story you want to tell

Explanatory analysis

- A few things to think about and be extremely clear on before visualizing any data or creating content.
 - Who
 - First, To whom are you communicating?
 - What
 - Second, What do you want your audience to know or do?
 - Mechanism
 - How will you communicate to your audience?
 - Tone
 - What tone do you want your communication to set
 - How
 - What data is available that will help make my point?

LIVE PRESENTATION WRITTEN DOC OR EMAIL



Choosing an effective visual

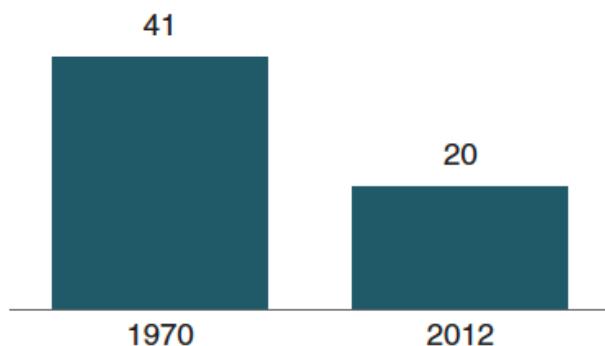
Simple text

- When you have just a number or two to share, simple text can be a great way to communicate
 - Think about solely using the number—making it as prominent as possible—and a few supporting words to clearly make your point.

Example: April 2014 Pew Research Center report on stay-at-home moms

Children with a "Traditional" Stay-at-Home Mother

% of children with a married stay-at-home mother with a working husband



Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Source: Pew Research Center analysis of March Current Population Surveys Integrated Public Use Microdata Series (IPUMS-CPS), 1971 and 2013

Adapted from PEW RESEARCH CENTER

20%

of children had a **traditional stay-at-home mom** in 2012, compared to 41% in 1970

Heatmap

- A way to visualize data in tabular format
 - Leverage colored cells that convey the relative magnitude of the numbers
 - Use color saturation to provide visual cues, helping our eyes and brains more quickly target the potential points of interest

Table

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

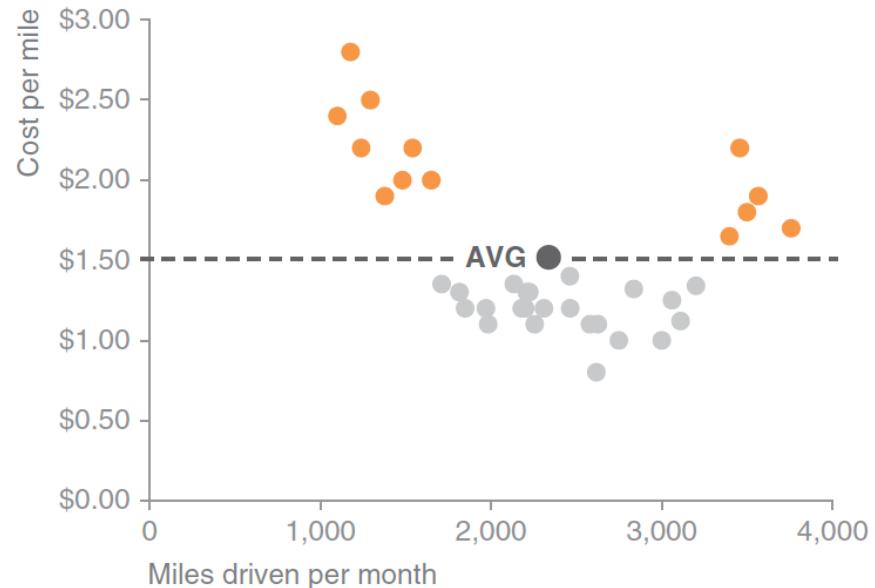
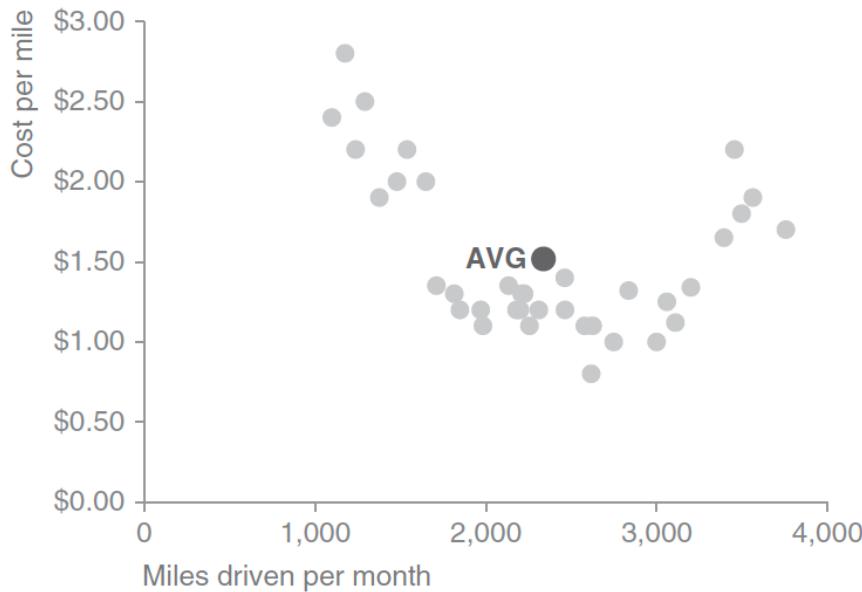
Heatmap

LOW-HIGH

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Use color to highlight

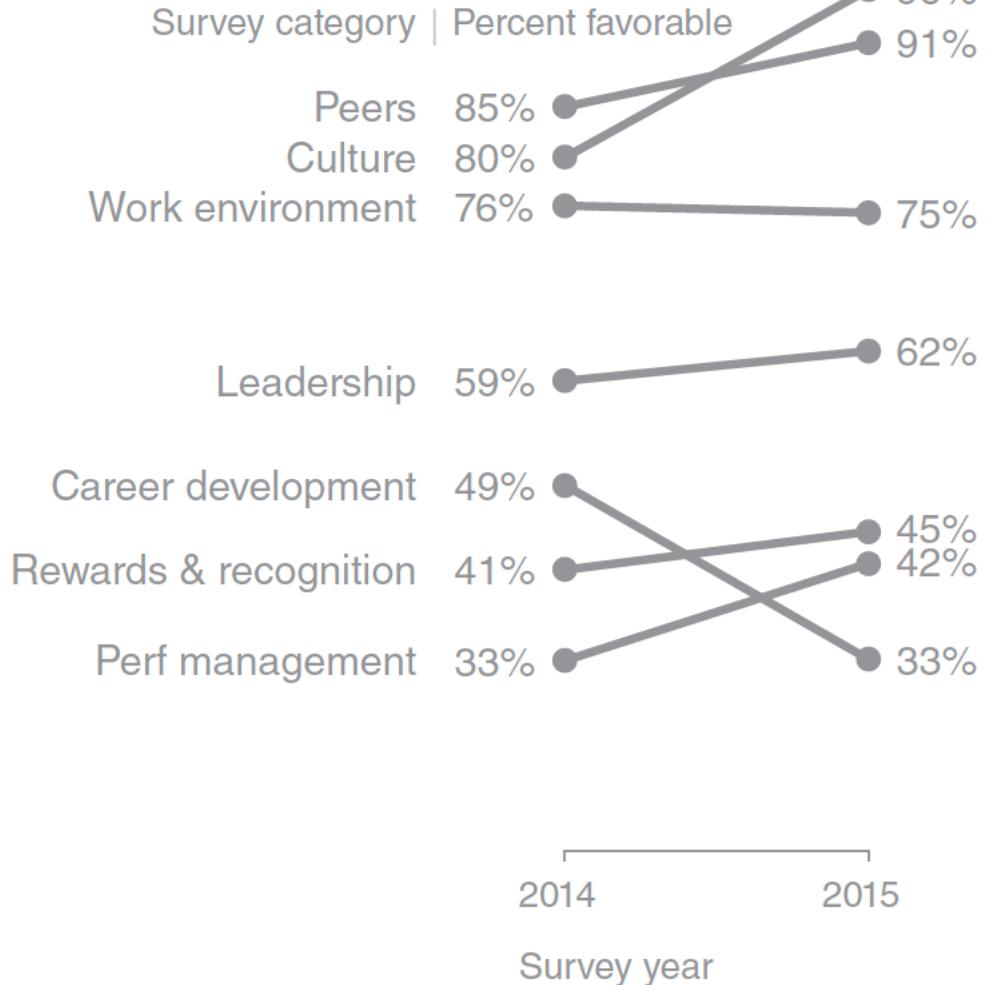
- Focus primarily on those cases where cost per mile is above average



Slopegraph

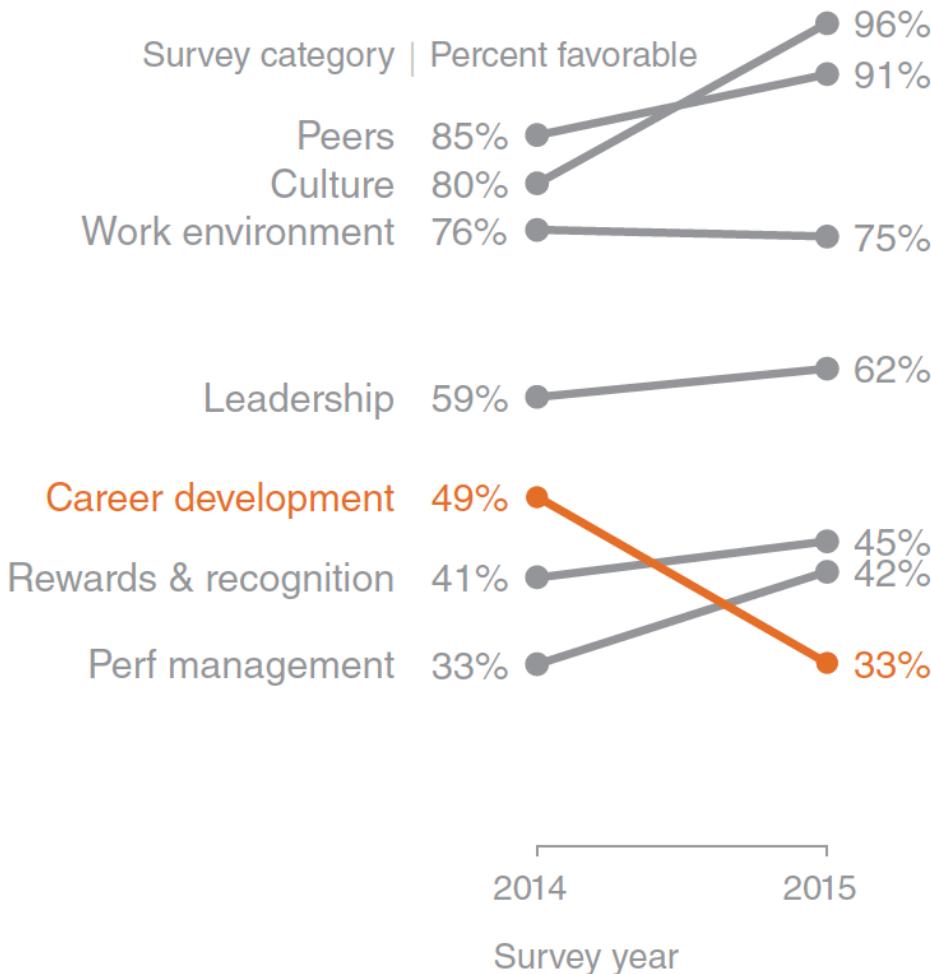
- Show relative increases and decreases or differences across various categories between the two data point.
- Slopegraph shows
 - Absolute values
 - The visual increase or decrease in rate of change.

Employee feedback over time



Draw attention to the single category that decreased over time

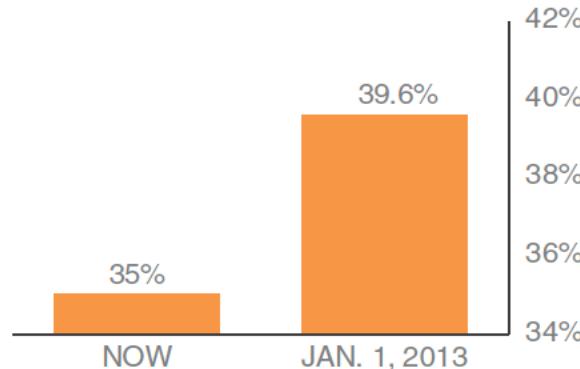
Employee feedback over time



Bar charts

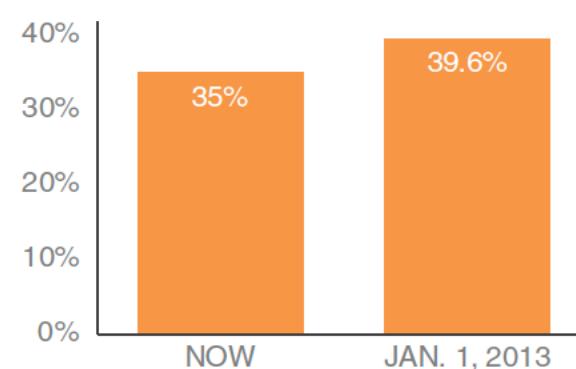
Non-zero baseline: as originally graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE

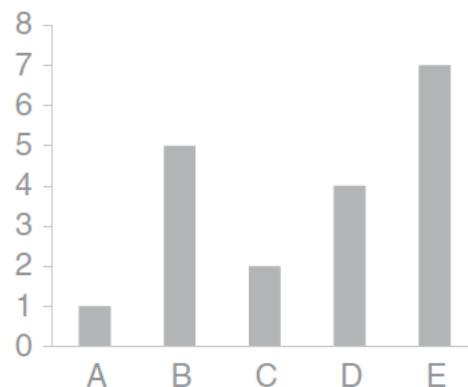


Zero baseline: as it should be graphed

IF BUSH TAX CUTS EXPIRE
TOP TAX RATE



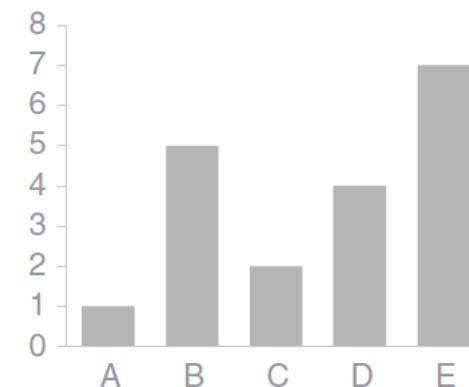
Too thin



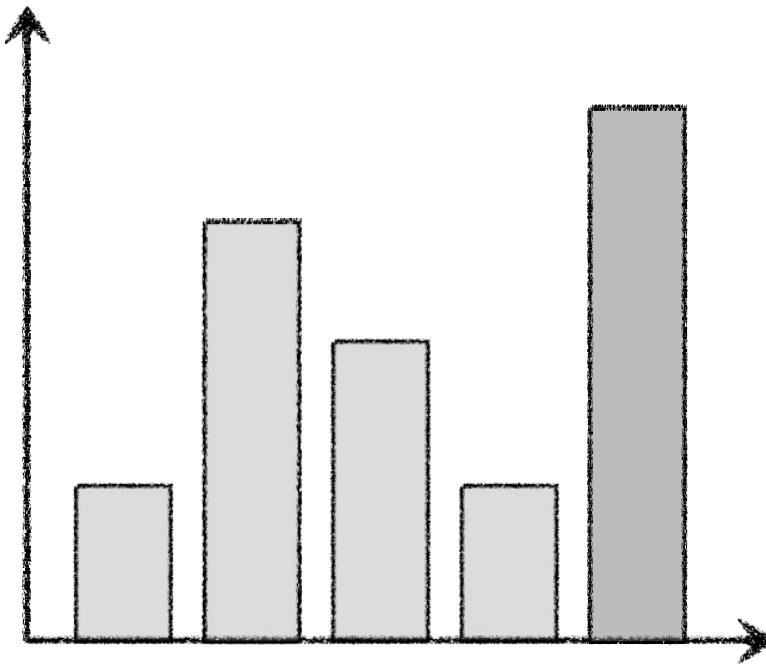
Too thick

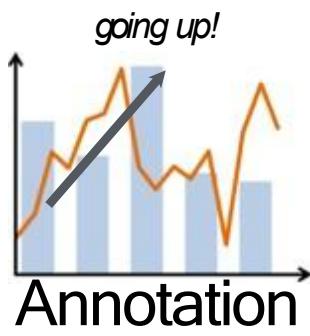


Just right

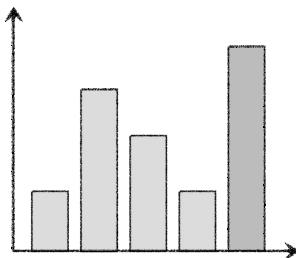


Redundant encoding



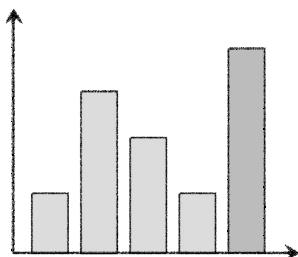


Message
Redundancy

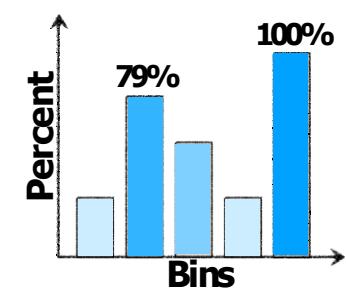
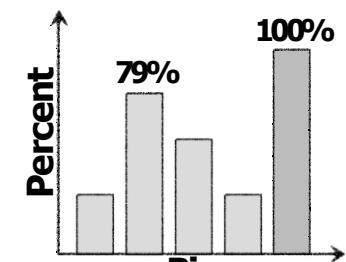




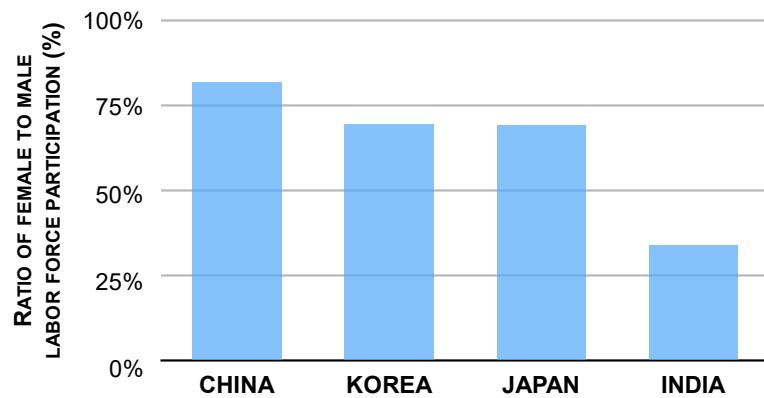
Message Redundancy



Data Redundancy



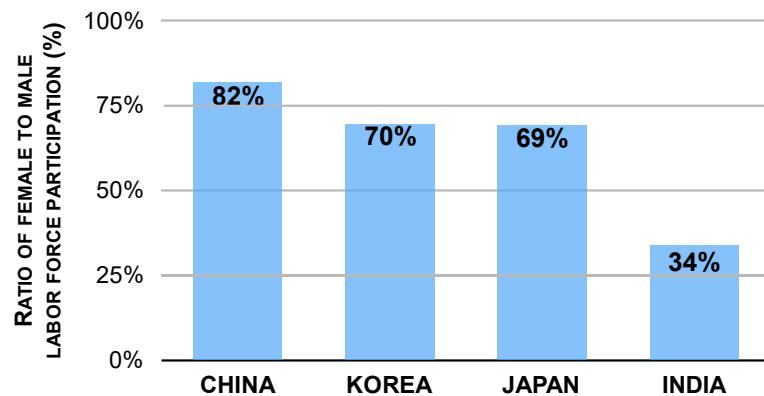
GENDER EQUALITY IN LABOR FORCE PARTICIPATION



Source: Gender Statistics 2013, World Bank

ORIGINAL

GENDER EQUALITY IN LABOR FORCE PARTICIPATION

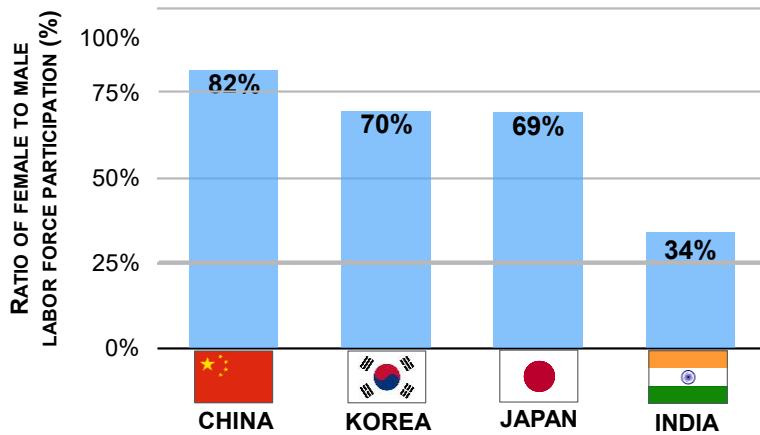


Source: Gender Statistics 2013, World Bank

DATA REDUNDANCY

GENDER EQUALITY IN LABOR FORCE PARTICIPATION

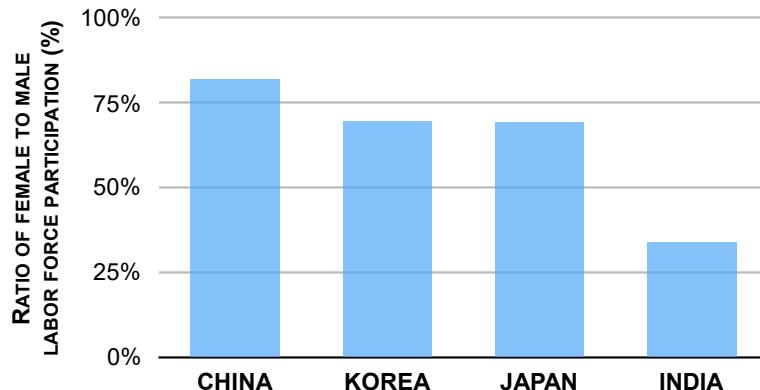
CHINA LEADS IN FEMALE LABOR FORCE PARTICIPATION WHEREAS INDIA LAGS SIGNIFICANTLY BEHIND AMONG OTHER ASIAN-PACIFIC COUNTRIES IN 2013.



Source: Gender Statistics 2013, World Bank

DATA & MESSAGE REDUNDANCY

GENDER EQUALITY IN LABOR FORCE PARTICIPATION

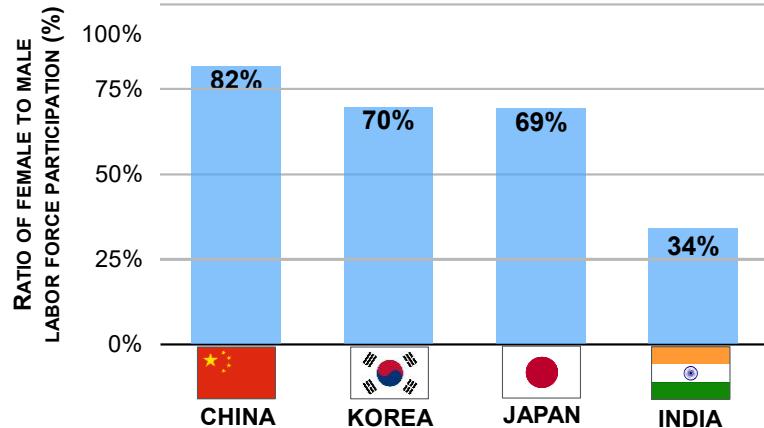


Source: Gender Statistics 2013, World Bank

ORIGINAL

GENDER EQUALITY IN LABOR FORCE PARTICIPATION

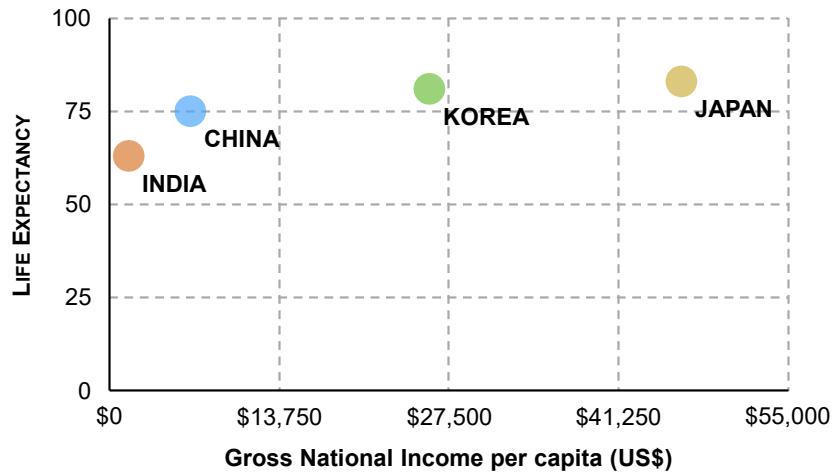
CHINA LEADS IN FEMALE LABOR FORCE PARTICIPATION WHEREAS INDIA LAGS SIGNIFICANTLY BEHIND AMONG OTHER ASIAN-PACIFIC COUNTRIES IN 2013.



Source: Gender Statistics 2013, World Bank

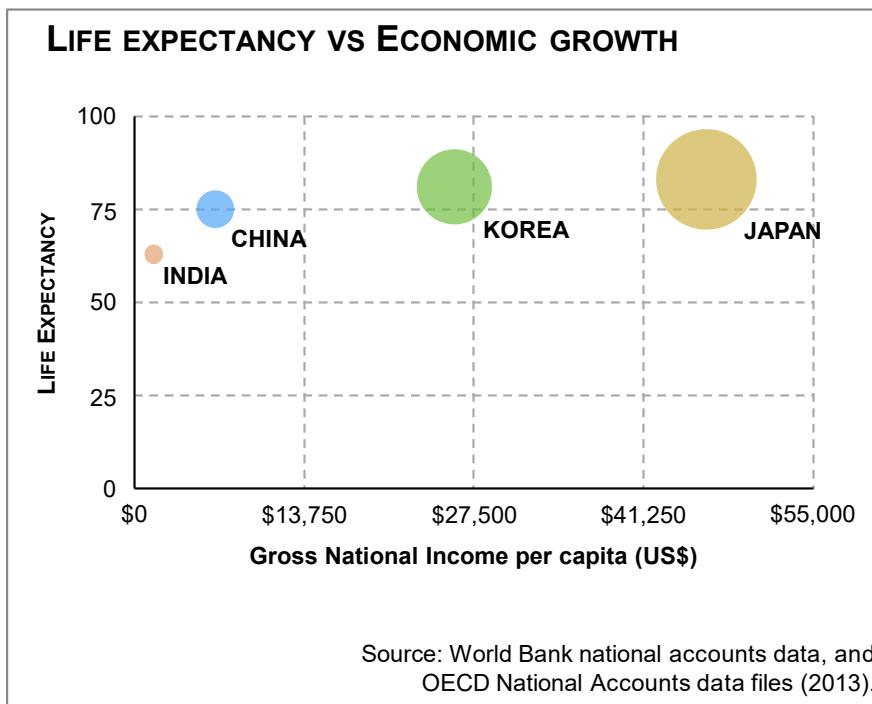
DATA & MESSAGE REDUNDANCY

LIFE EXPECTANCY VS ECONOMIC GROWTH

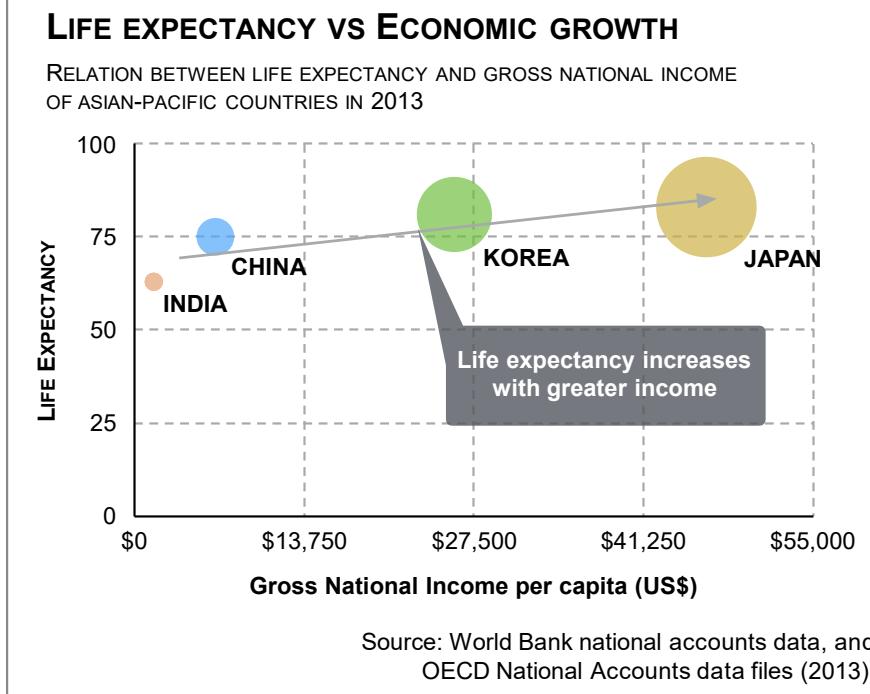


Source: World Bank national accounts data, and
OECD National Accounts data files (2013).

ORIGINAL

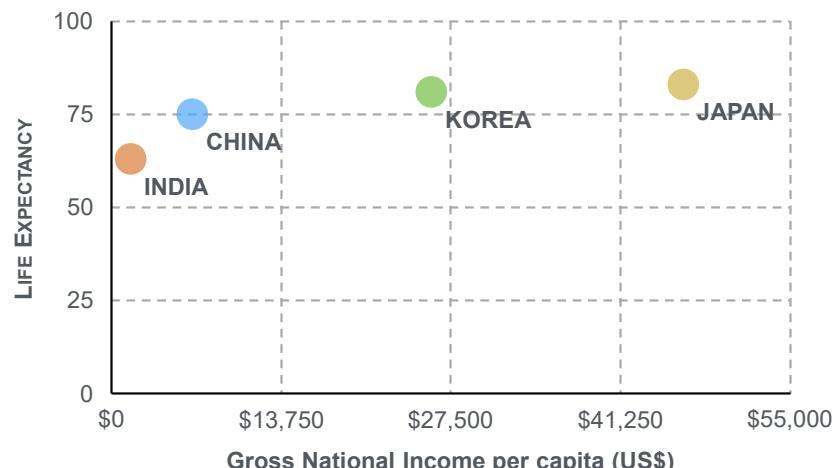


DATA REDUNDANCY



DATA & MESSAGE REDUNDANCY

LIFE EXPECTANCY VS ECONOMIC GROWTH

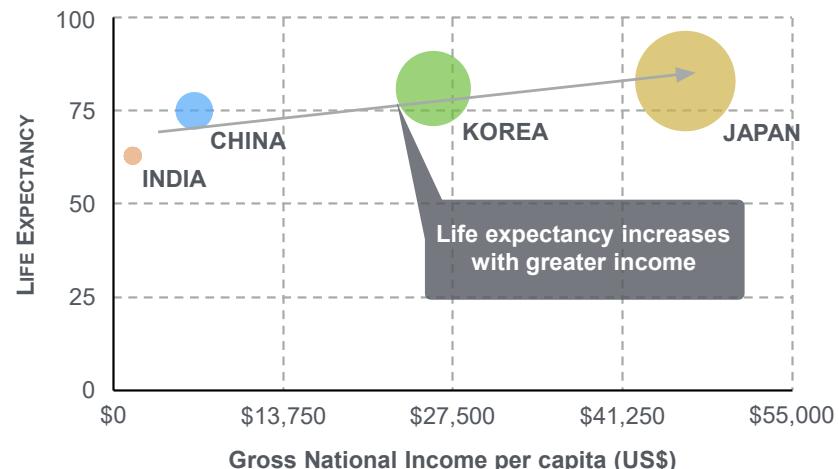


Source: World Bank national accounts data, and
OECD National Accounts data files (2013).

ORIGINAL

LIFE EXPECTANCY VS ECONOMIC GROWTH

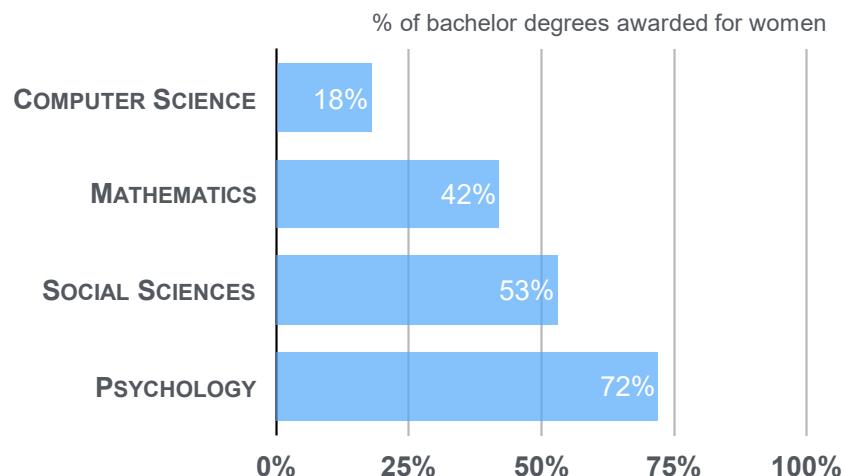
RELATION BETWEEN LIFE EXPECTANCY AND GROSS NATIONAL INCOME OF ASIAN-PACIFIC COUNTRIES IN 2013



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

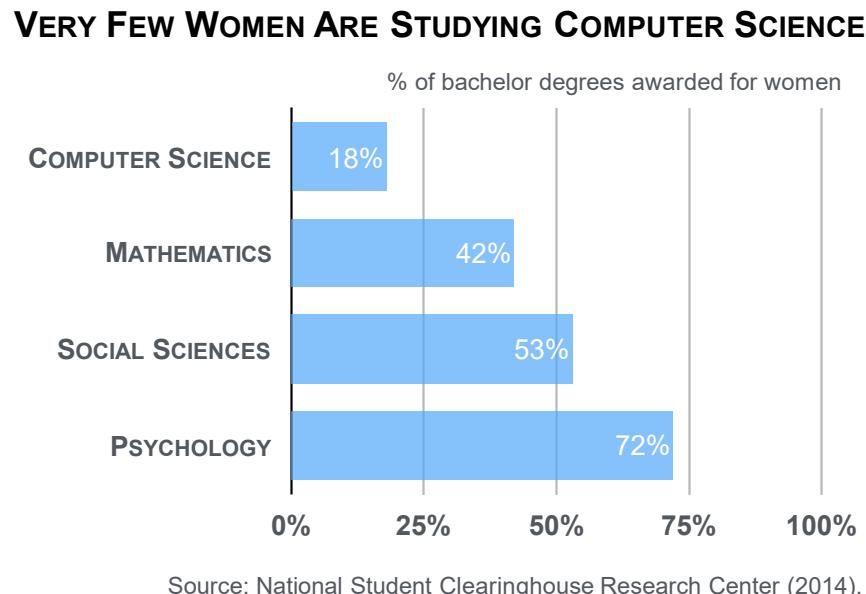
DATA & MESSAGE REDUNDANCY

VERY FEW WOMEN ARE STUDYING COMPUTER SCIENCE

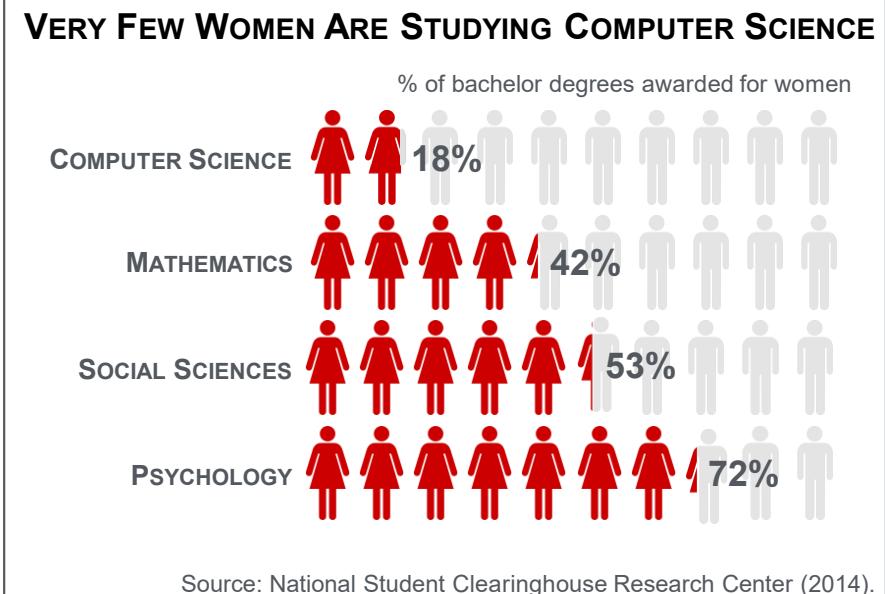


WITHOUT PICTOGRAM

Example 3



WITHOUT PICTOGRAM



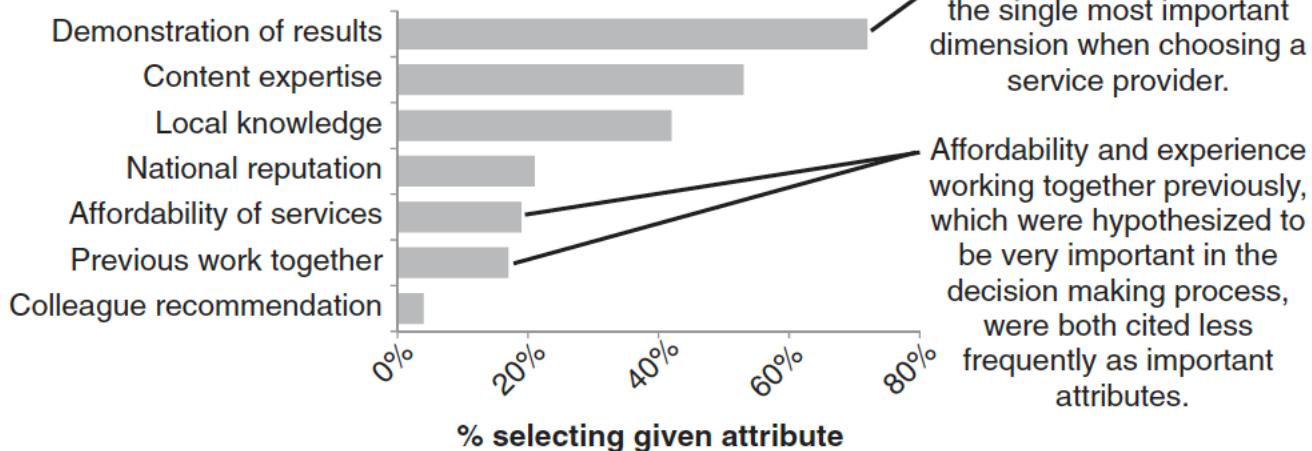
WITH PICTOGRAM

Eliminate clutter

Example: Lack of visual order

Demonstrating effectiveness is most important consideration when selecting a provider

In general, what attributes are the most important to you in selecting a service provider?
(Choose up to 3)



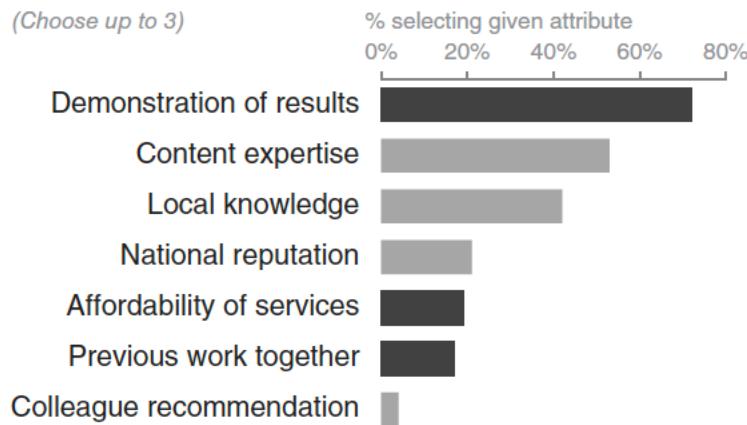
Data source: xyz; includes N number of survey respondents. Note that respondents were able to choose up to 3 options.

Improve the placement and formatting of elements

Demonstrating effectiveness is most important consideration when selecting a provider

In general, **what attributes are the most important** to you in selecting a service provider?

(Choose up to 3)



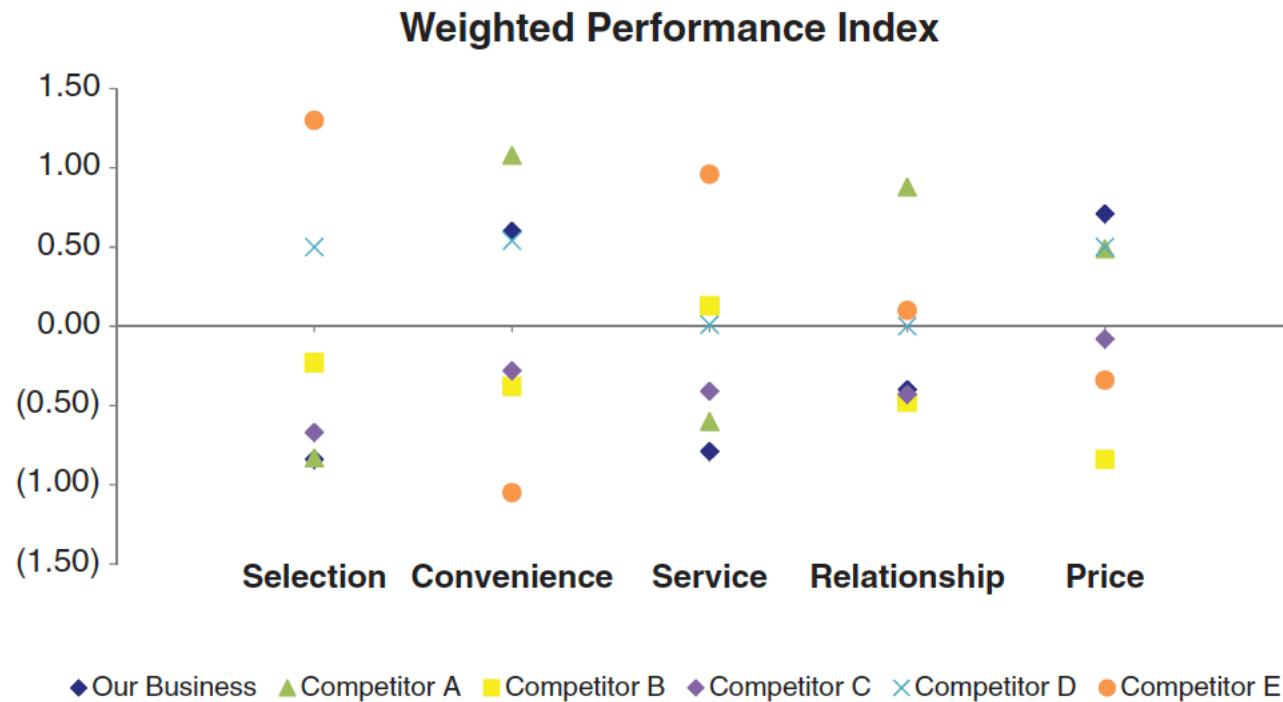
Survey shows that **demonstration of results** is the single most important dimension when choosing a service provider.

Affordability and experience working together previously, which were hypothesized to be very important in the decision making process, were both cited less frequently as important attributes.

Data source: xyz; includes N number of survey respondents.
Note that respondents were able to choose up to 3 options.

Example: Weighted performance index across categories

- Shows the weighted performance index across categories for your company and five competitors



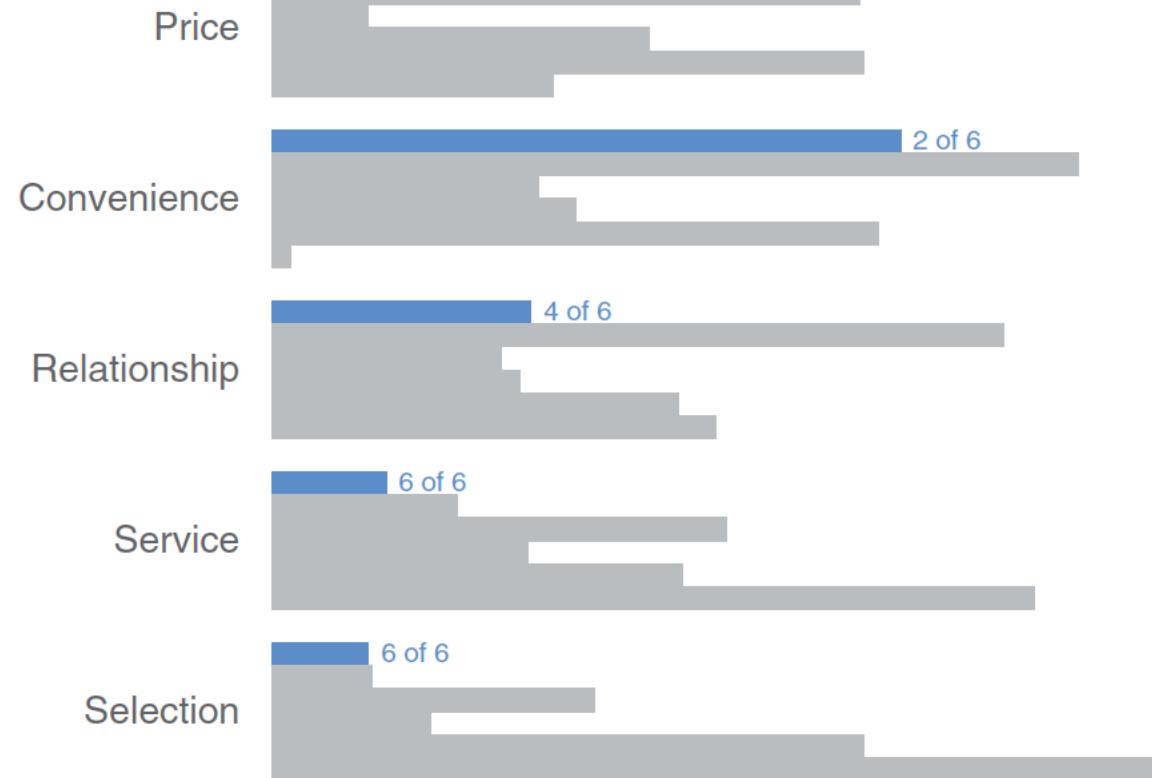
Use contrast more strategically

Performance overview

■ Our business

- Competitor A
- Competitor B
- Competitor C
- Competitor D
- Competitor E

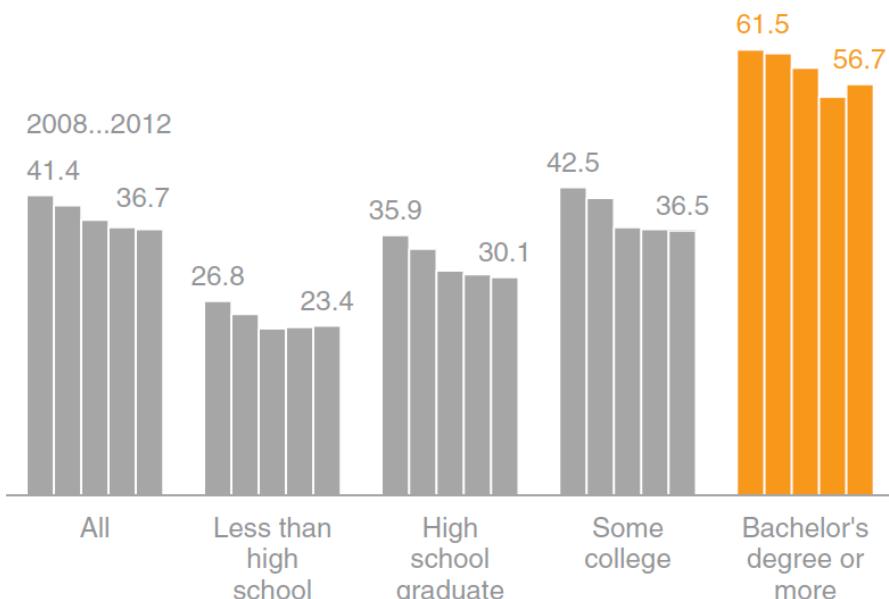
Weighted performance index | relative rank



Eliminate distractions

New Marriage Rate by Education

Number of newly married adults per 1,000 marriage eligible adults

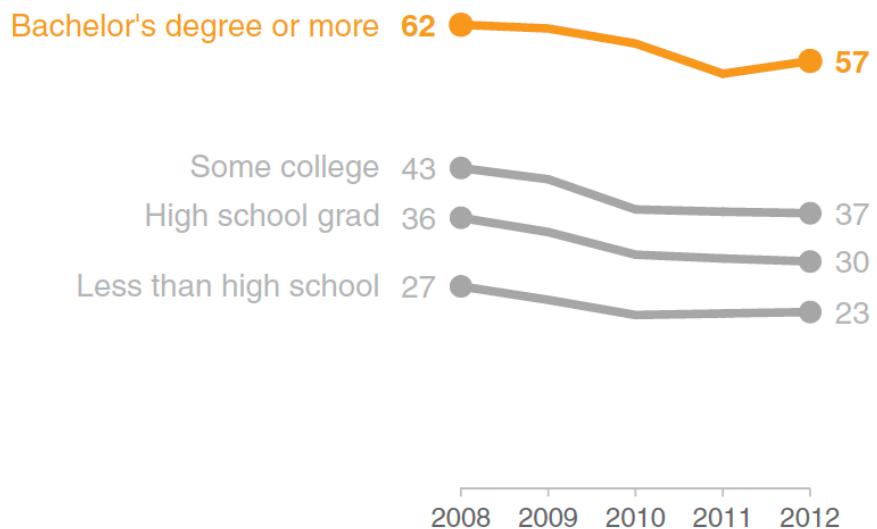


Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

New marriage rate by education

Number of newly married adults per 1,000 marriage eligible adults



Note: Marriage eligible includes the newly married plus those widowed, divorced, or never married at interview.

Source: U.S. Census

Focus your audience's attention



Recall preattentive attributes

- Preattentive attributes are **visual properties that we notice without using conscious effort to do so**.
- use preattentive attributes strategically, they can help us enable our audience to see what we want them to see before they even know they're seeing it!



Orientation



Shape



Line length



Line width



Size



Curvature



Added marks



Enclosure



Hue



Intensity



Spatial position



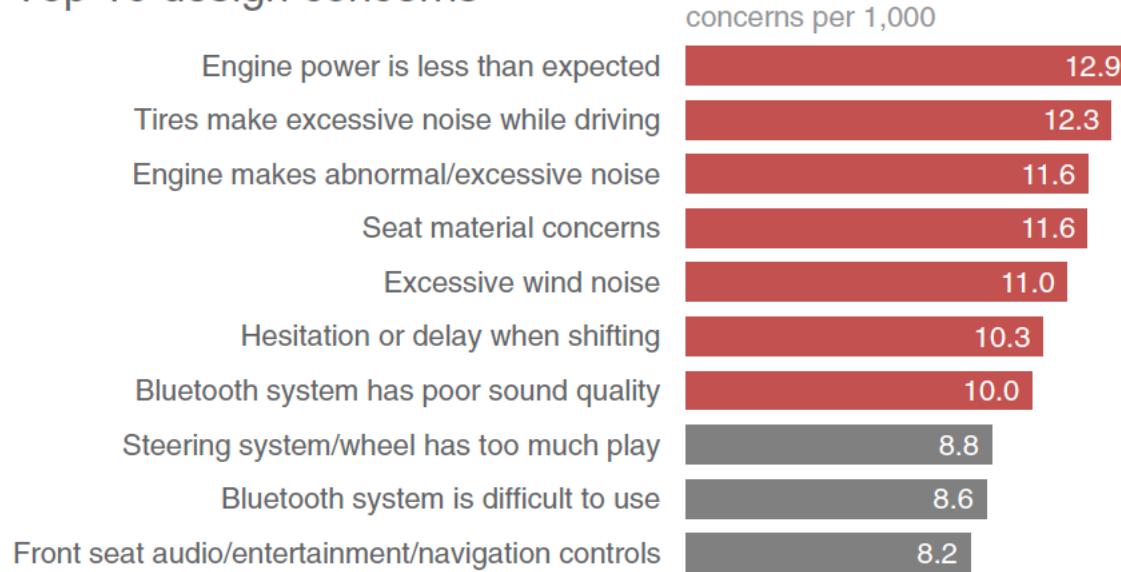
Motion

Example: Leverage color to draw attention

7 of the top 10 design concerns have 10 or more concerns per 1,000.

Discussion: is this an acceptable default rate?

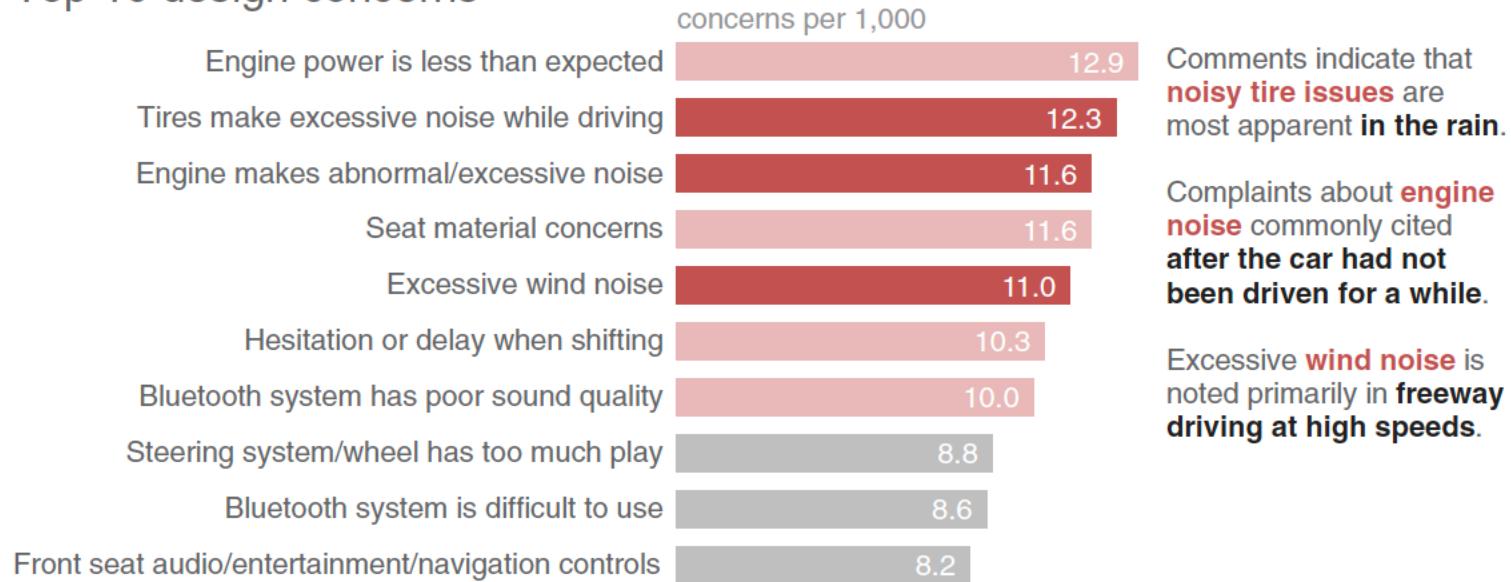
Top 10 design concerns



Example: Create a visual hierarchy of information

- Modified focus and text to lead our audience from the macro to the micro parts of the story.

Top 10 design concerns



Use color sparingly

Country Level Sales Rank Top 5 Drugs

Rainbow distribution in color indicates sales rank in given country from #1 (red) to #10 or higher (dark purple)

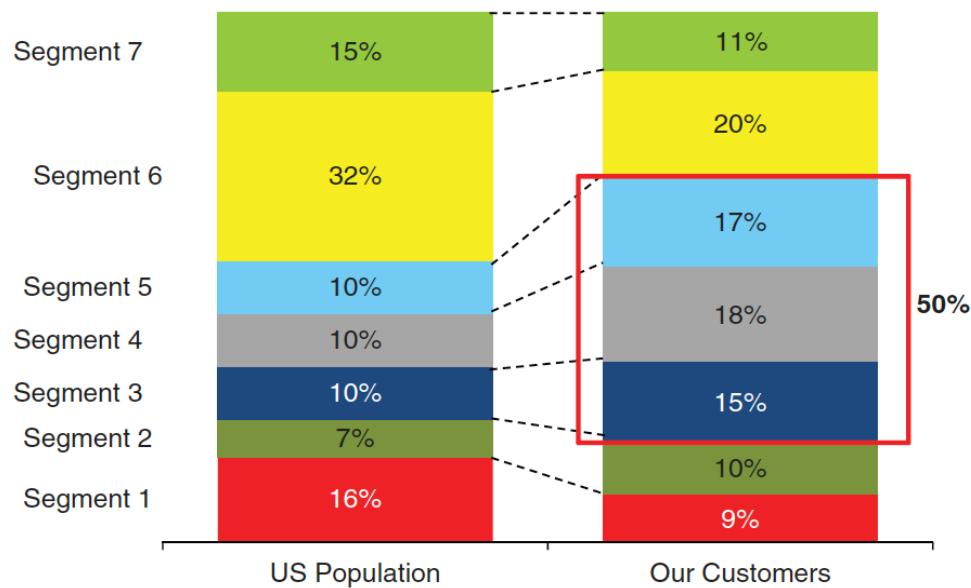
Country	A	B	C	D	E
AUS	1	2	3	6	7
BRA	1	3	4	5	6
CAN	2	3	6	12	8
CHI	1	2	8	4	7
FRA	3	2	4	8	10
GER	3	1	6	5	4
IND	4	1	8	10	5
ITA	2	4	10	9	8
MEX	1	5	4	6	3
RUS	4	3	7	9	12
SPA	2	3	4	5	11
TUR	7	2	3	4	8
UK	1	2	3	6	7
US	1	2	4	3	5

Top 5 drugs: country-level sales rank

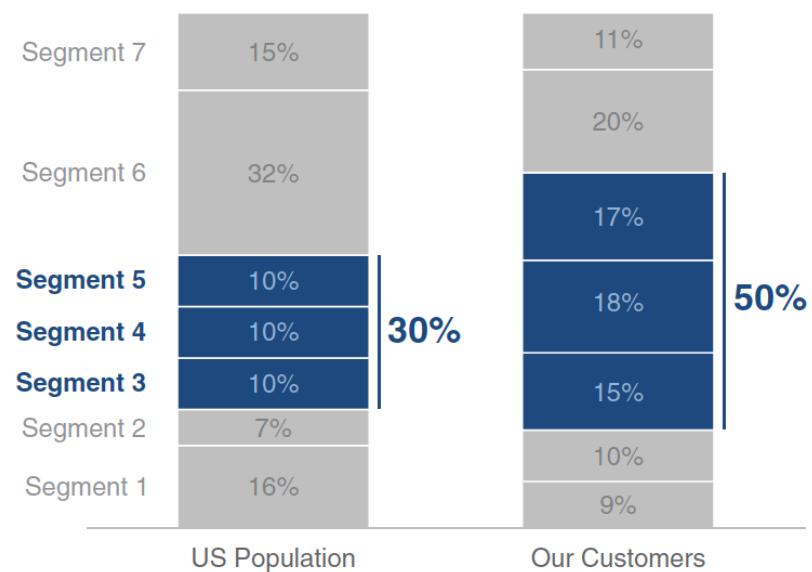
RANK	1	2	3	4	5+
COUNTRY DRUG	A	B	C	D	E
Australia	1	2	3	6	7
Brazil	1	3	4	5	6
Canada	2	3	6	12	8
China	1	2	8	4	7
France	3	2	4	8	10
Germany	3	1	6	5	4
India	4	1	8	10	5
Italy	2	4	10	9	8
Mexico	1	5	4	6	3
Russia	4	3	7	9	12
Spain	2	3	4	5	11
Turkey	7	2	3	4	8
United Kingdom	1	2	3	6	7
United States	1	2	4	3	5

Aesthetic design

Distribution by customer segment



Distribution by customer segment

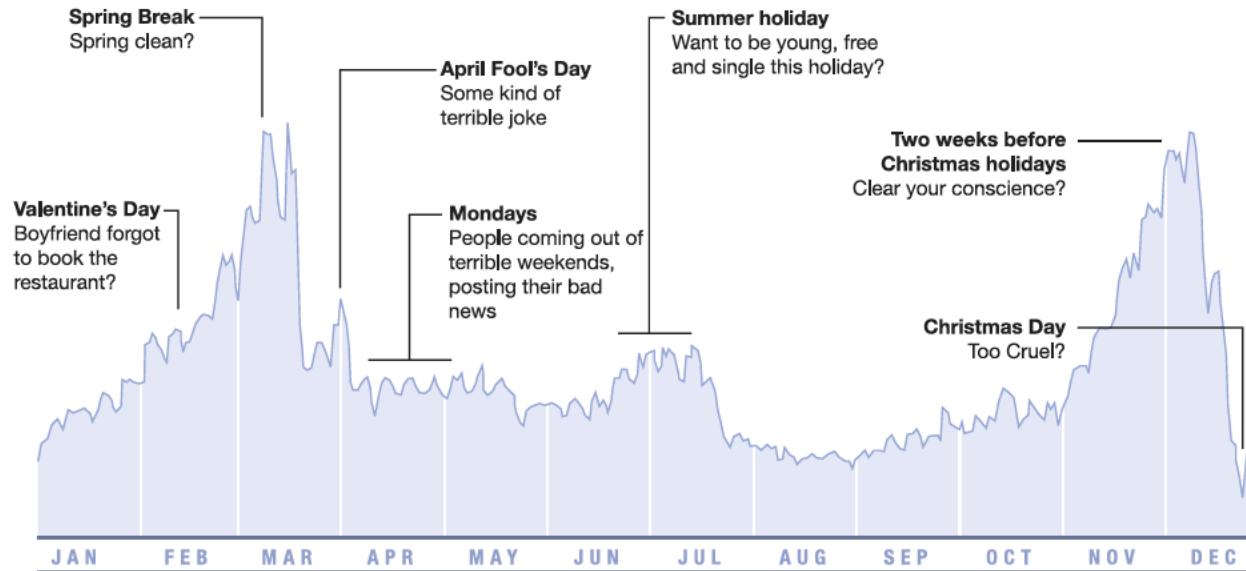


Add explanations

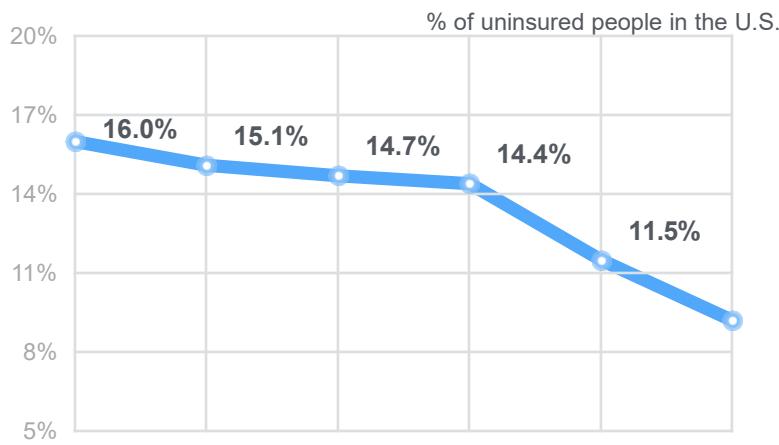
Text in data visualization

- Useful to annotate important or interesting points directly on a graph.

Peak Break-up Times
According to Facebook status updates



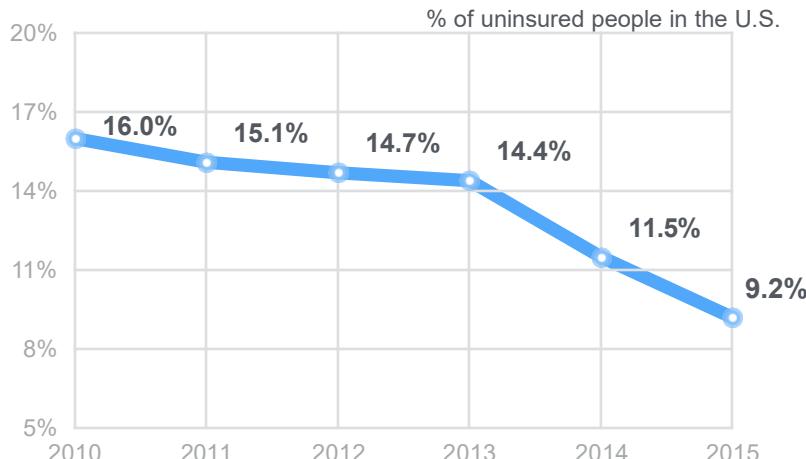
PERCENTAGE OF UNINSURED AMERICANS



BAD TITLE

Good title

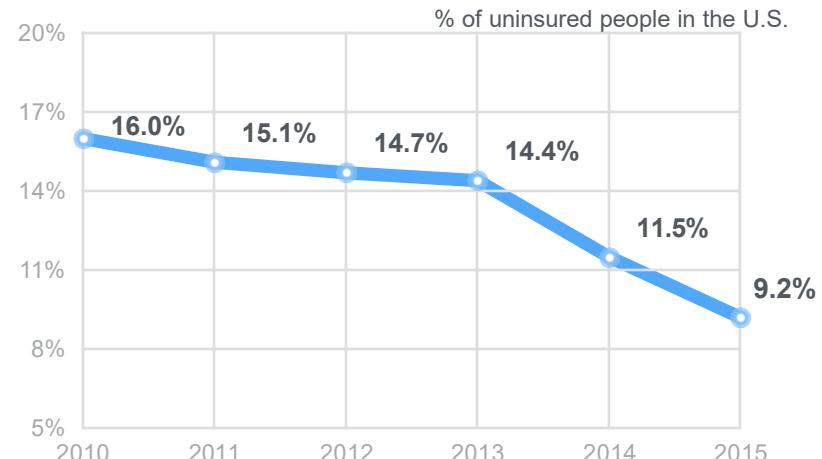
PERCENTAGE OF UNINSURED AMERICANS



Source: CDC/NCHS, National Health Interview Survey, 2010–2015

BAD TITLE

AMERICA'S UNINSURED RATE DIPS BELOW 10%



Source: CDC/NCHS, National Health Interview Survey, 2010–2015

GOOD TITLE

Tell stories with data

Why telling a story with data

- A single chart may not be enough
 - To explain
 - 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?

A painting depicting a man with a beard and a cap, wearing a brown coat, sitting on a stool and gesturing with his hands as if telling a story. Three children are seated in front of him, listening attentively. The setting appears to be a rustic, dimly lit room or cabin.

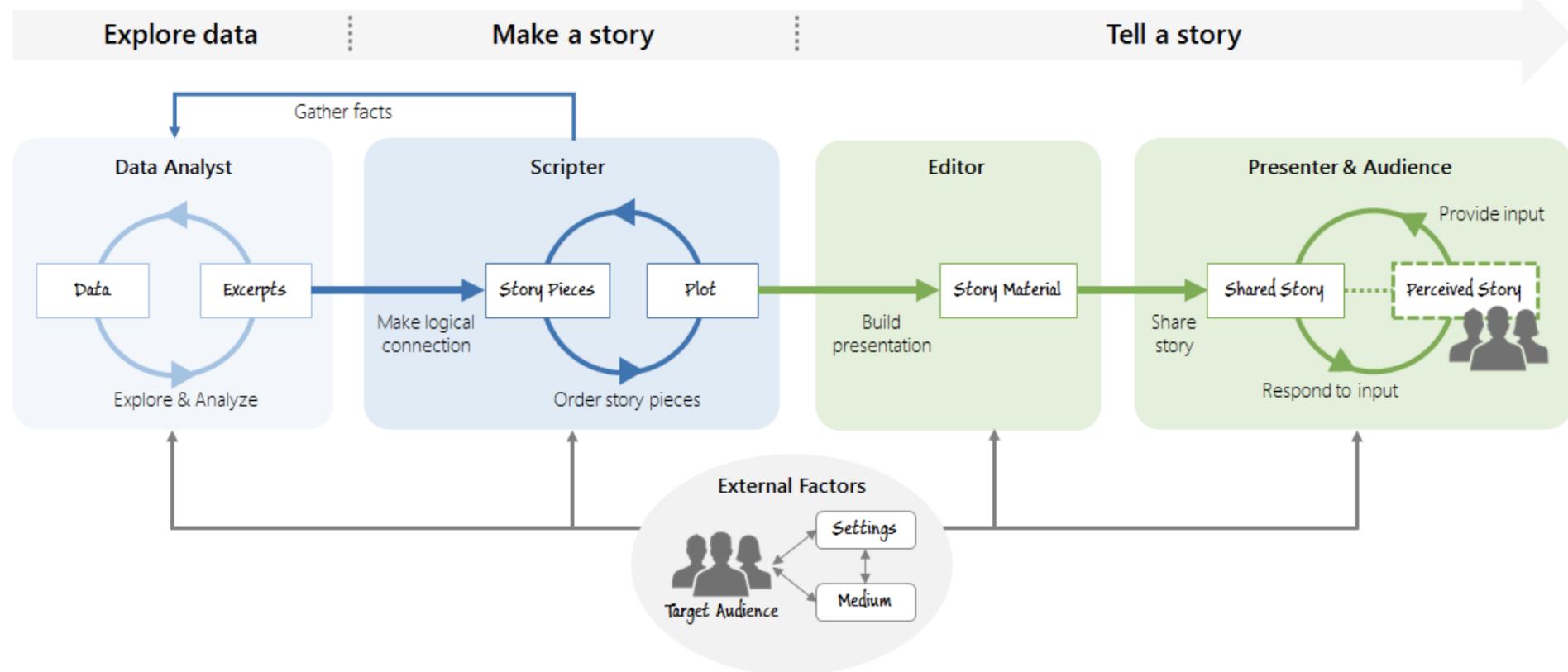
Stories

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

- <https://www.youtube.com/watch?v=hVimVzgtD6w>

Use of elements from storytelling
to convey compelling data stories

Data Storytelling Process: transforming data into visual stories



[B. Lee et al 2015]

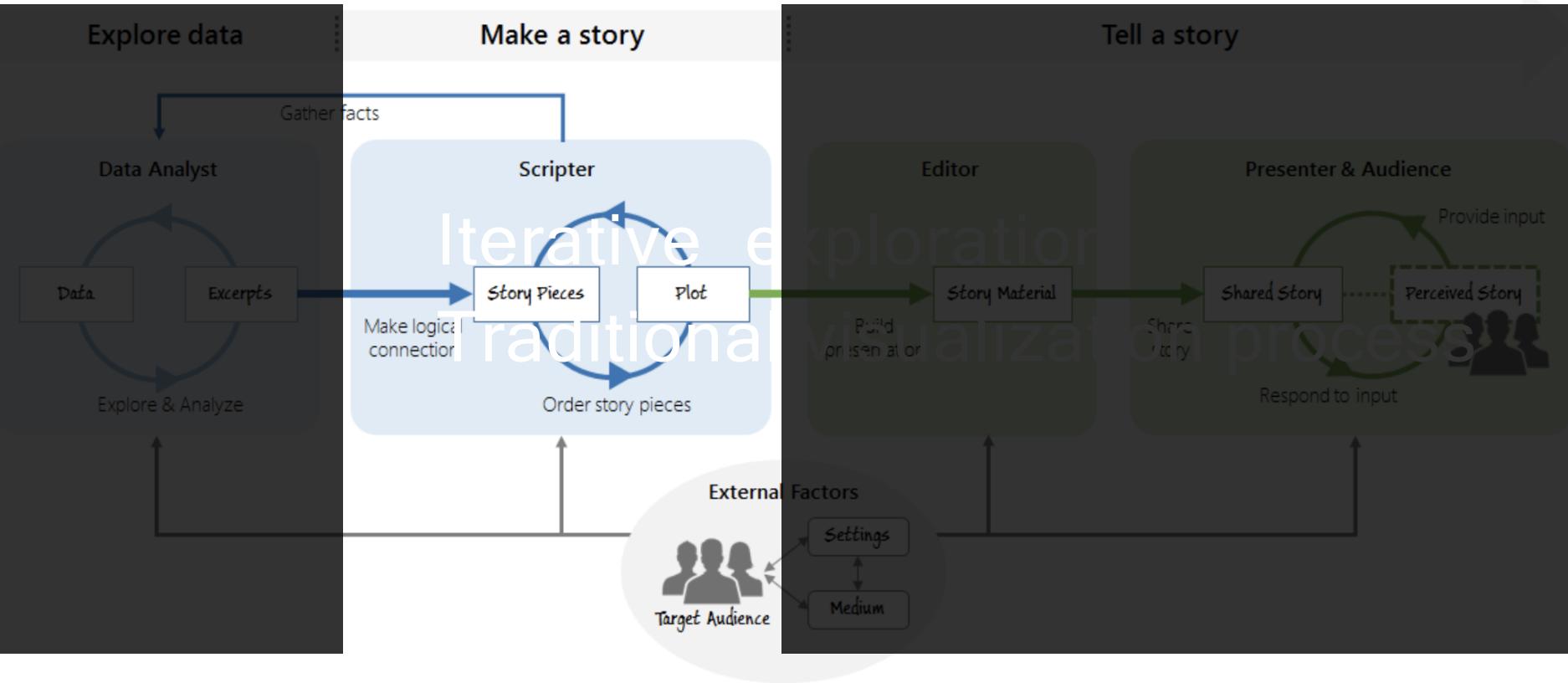
Data Storytelling Process: transforming data into visual stories

Explore data



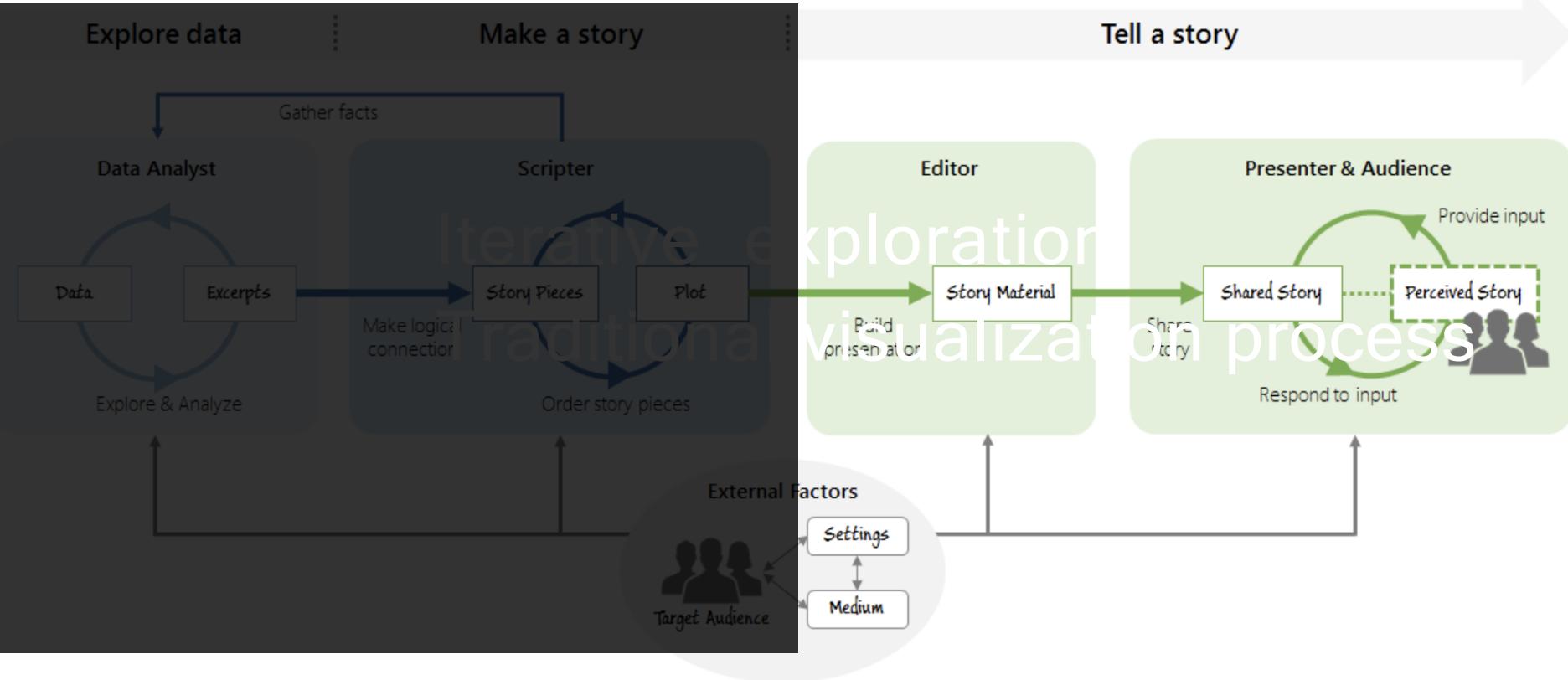
[B. Lee et al 2015]

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]

7 Key Elements of Compelling Data Storytelling

- Clear and coherent narrative structure
 - Captivating introduction, logical flow, connecting data points and insights, compelling climax, memorable ending
- Characterisation of data points
 - Relevance, impact, and alignment with the narrative
- Engaging visuals and data representations
- Contextualisation and relatability
- Emotional resonance
- Simplicity and clarity
- Adaptation to the audience

Narration Styles

- Author-Driven
 - Strict ordering of scenes
 - Heavy messaging
 - No interactivity
- Reader-Driven
 - No prescribed ordering
 - No messaging
 - Free interactivity

https://www.youtube.com/watch?feature=player_detailpage&v=hVimVzgtD6w#t=159s



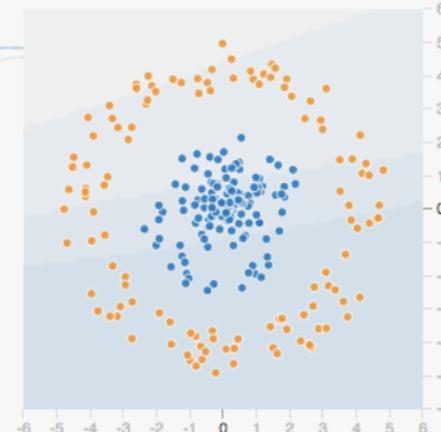
Epoch 000,000 Learning rate 0.03 Activation Tanh Regularization None Regularization rate 0 Problem type Classification

DATA
Which dataset do you want to use?

Ratio of training to test data: 50%
Noise: 0
Batch size: 10

FEATURES
Which properties do you want to feed in?


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

Published: February 2, 2010

Budget Forecasts, Compared With Reality

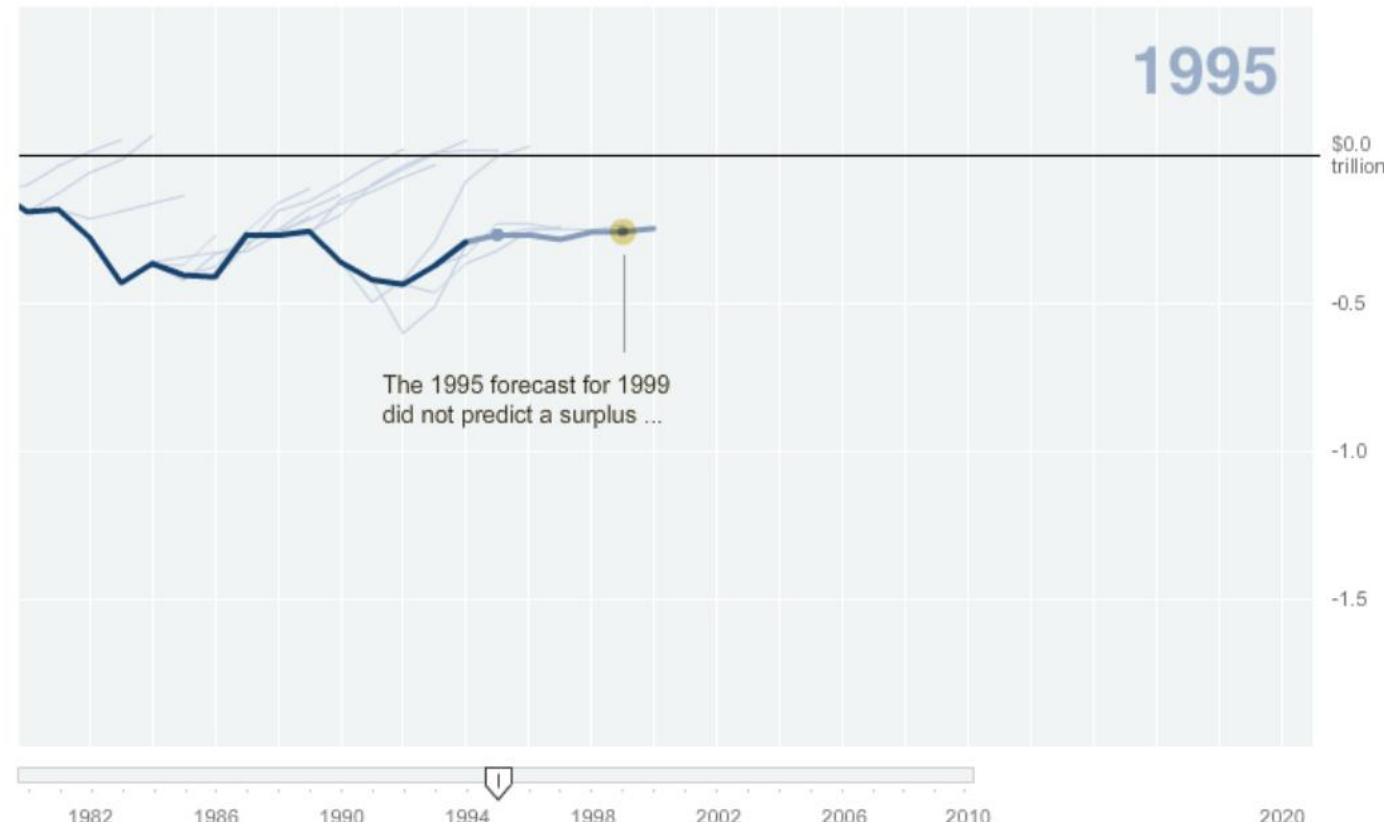
Just two years ago, surpluses were predicted by 2012. How accurate have past White House budget forecasts been?

1 2 3 4 5 6 NEXT ►

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.

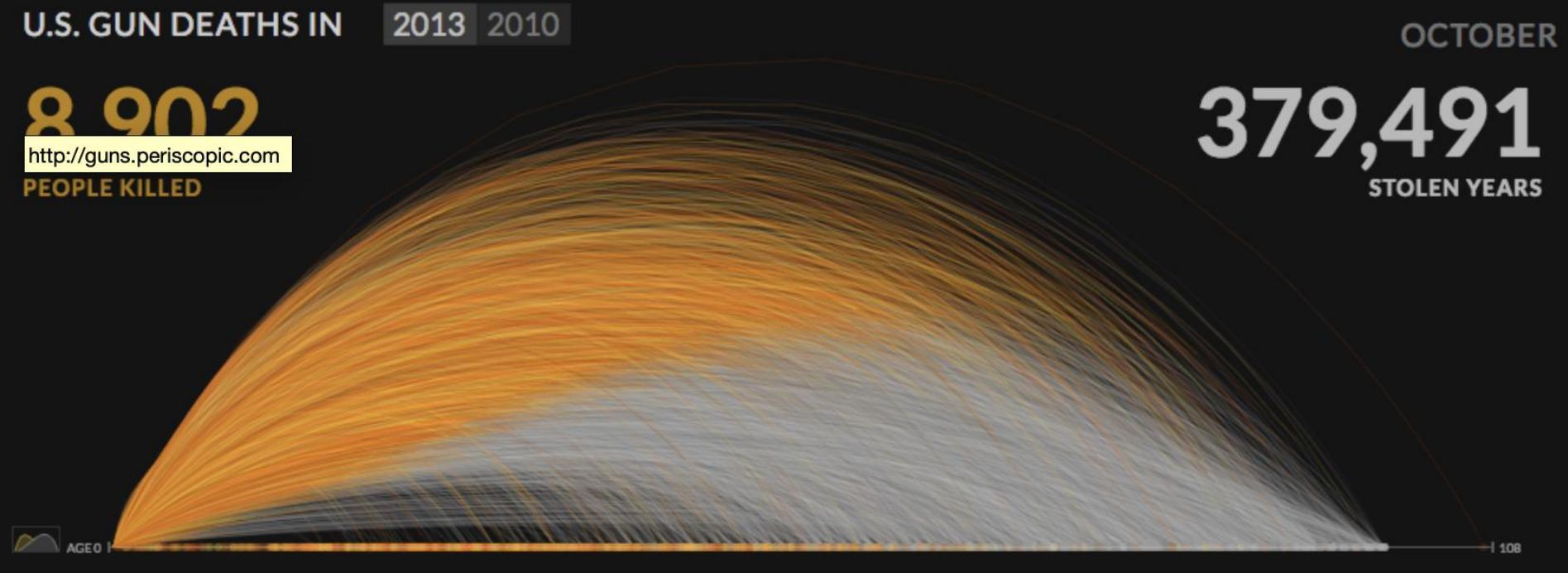


By AMANDA COX | Send Feedback

Source: Office of Management and Budget

[TWITTER](#) [LINKEDIN](#) [SHARE](#)

<https://guns.periscopic.com/>



<http://www.fallen.io/>

THE
Fallen
OF
World War II



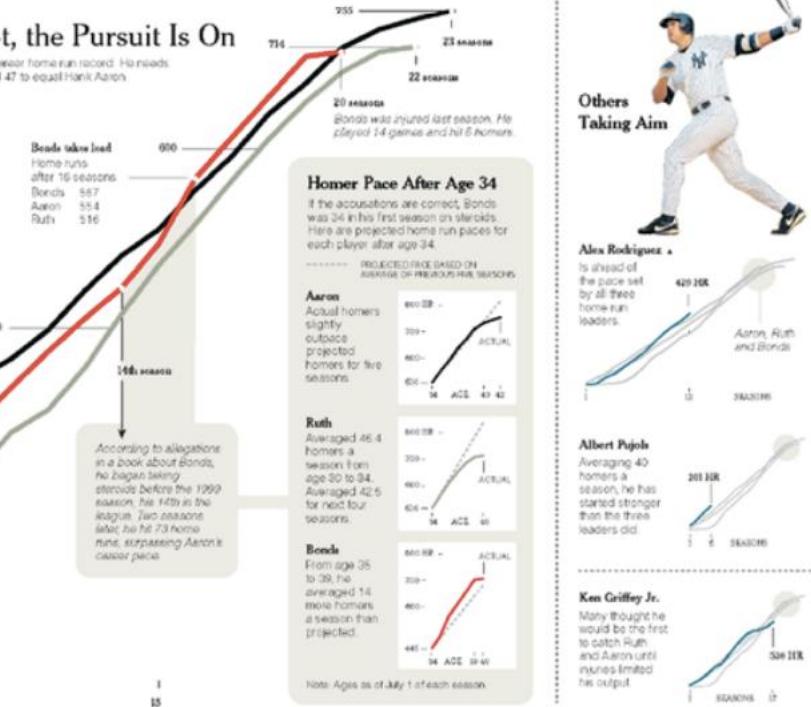
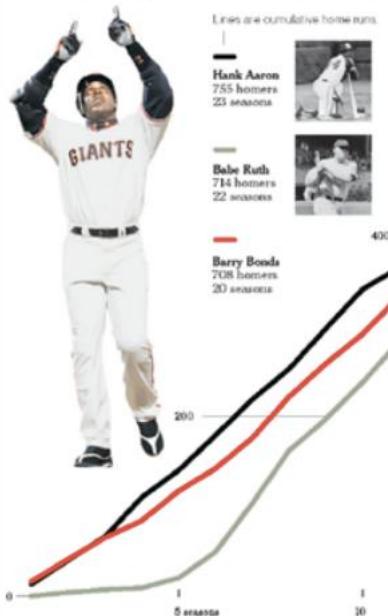
Visual narrative design

Employ narrative structure

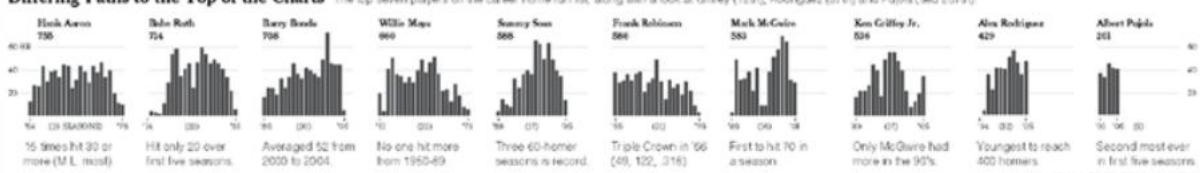


Bach et al 2017

755



Differing Paths to the Top of the Charts



755

Steroids or Not, the Pursuit Is On

Every player taking aim at the career home run record, no heads only six more to be above Ruth and 47 to equal Hank Aaron.

Lines are cumulative home runs.

Hank Aaron
755 home runs
23 seasons



Bonds also had
home runs after 10 seasons.
Bonds: 617
Aaron: 514
Ruth: 514

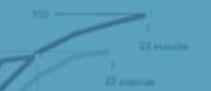
baseball
714 home runs
22 seasons



Berry Bonds
716 home runs
20 seasons

BEGINNING

Averaging 34 home runs in a book about steroids, he began hitting 40-plus before the 1989 season. He had 144 HRs 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 24 in his 12th season as the 1993 team prepared him to play his last game after age 34.

DATA COURTESY OF THE JOURNAL OF RESEARCH IN BASEBALL

Aaron:
Averaged 34 home runs slightly exceeding projected home runs for his seasons.

Bonds:
Averaged 40.4 home runs a season from age 32 to 34. Averaged 42.5 home runs for his final two seasons.

Bonds:
From age 35 to 39, averaged 21.6 home runs, unknown than 1993-97.

HOME AGES AS OF MAY 7, 2005 SEASONS

Others Taking Aim

Alex Rodriguez

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



Albert Pujols
Averaging 40 home runs 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.



Alex Rodriguez
Youngest to reach 400 home runs.

Albert Pujols
Second most ever in first five 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 home runs
41 seasons
15 times hit 30 or more (M.L. most)

Babe Ruth
714 home runs
34 seasons
Hit only 20 over first five seasons.

Berry Bonds
716 home runs
38 seasons
Averaged 52 from 2000 to 2004.

Willie Mays
660 home runs
30 seasons
No one hit more than 1950-69.

Sandy Koufax
588 home runs
16 seasons
Three 60-home run seasons is record.

Frank Robinson
584 home runs
20 seasons
Triple Crown in '66 (49, 122, 318)

Mark McGwire
583 home runs
18 seasons
First to hit 70 in a season.

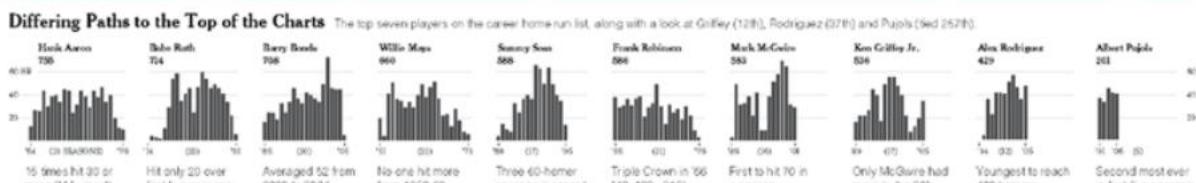
Ken Griffey Jr.
586 home runs
18 seasons
Only McGwire had more in the 90's.

Alex Rodriguez
429 home runs
14 seasons
Youngest to reach 400 home runs.

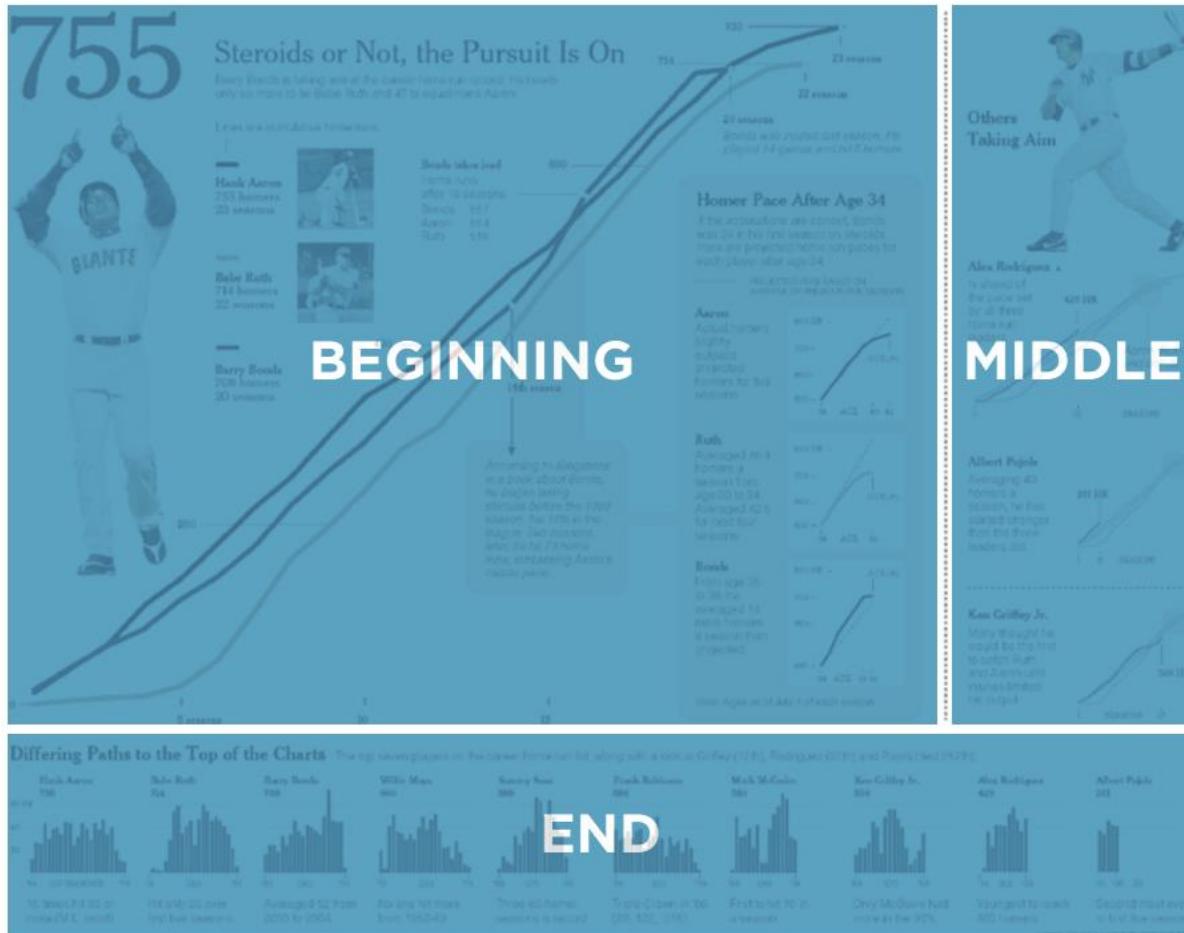
Albert Pujols
261 home runs
9 seasons
Second most ever in first five seasons.

According to the Journal of Research in Baseball

E. Segel



E. Segel



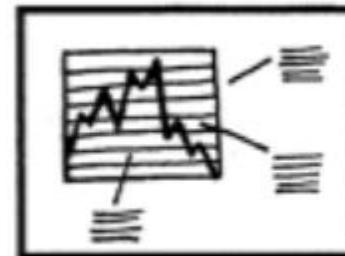
E. Segel

Genres of data stories

Seven
Genres



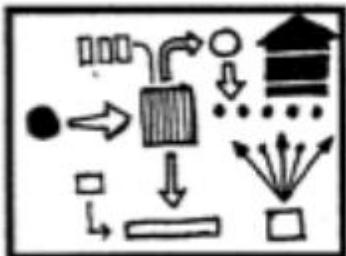
Magazine Style



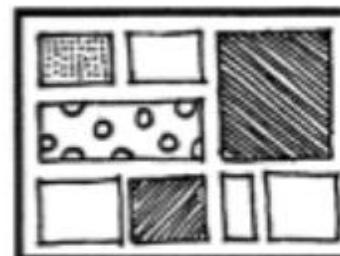
Annotated Chart



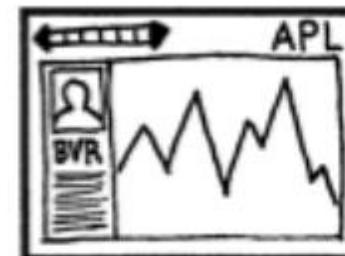
Partitioned Poster



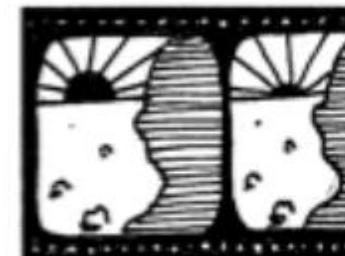
Flow Chart



Comic Strip



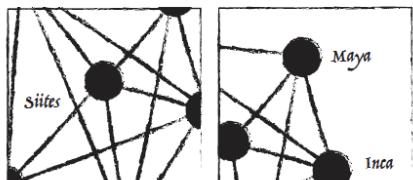
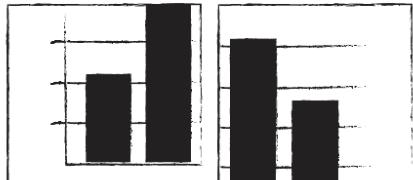
Slide Show



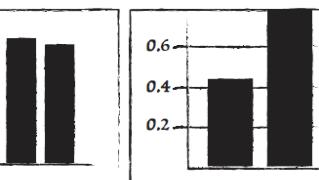
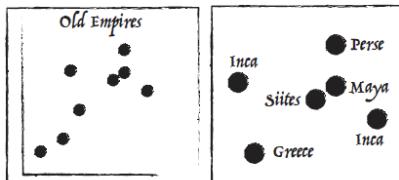
Film/Video/Animation

- Break down the complexity of the story and progressively reveal different facets of the data.

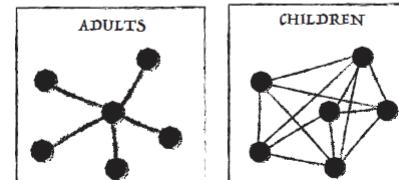
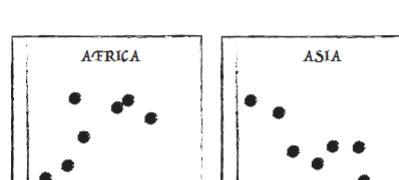
Detail-to-Detail



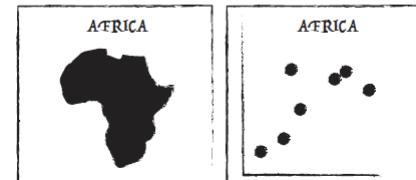
Level-of-Detail



Data-to-Data



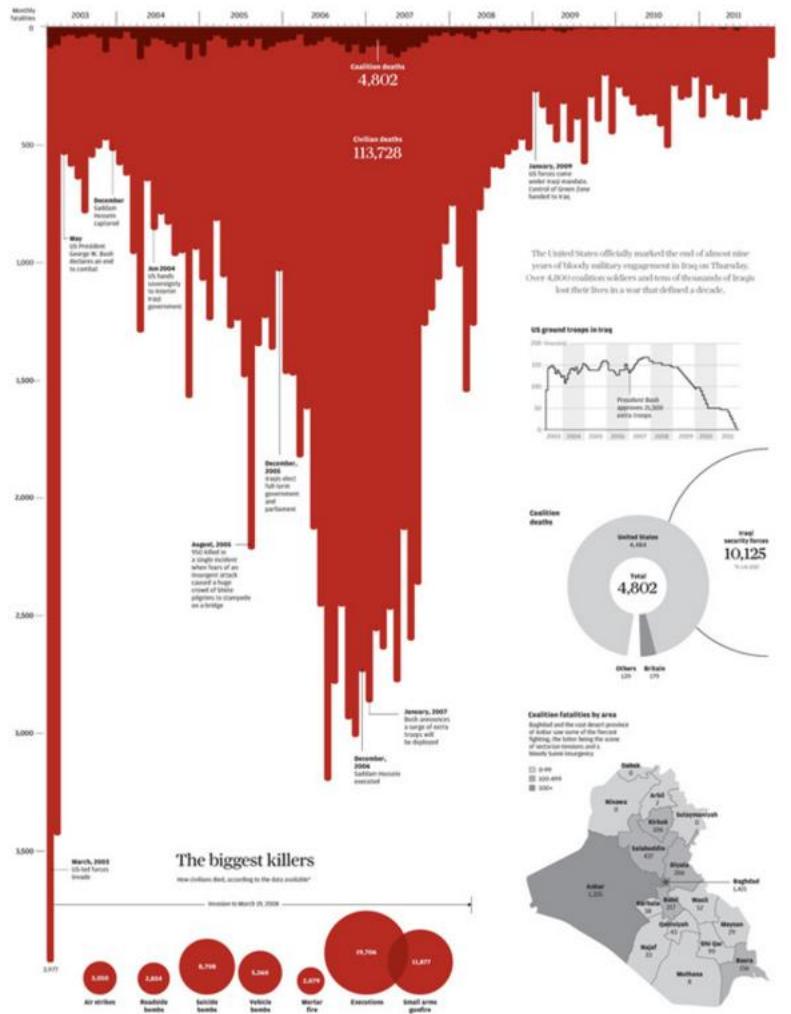
Visualization-to-Context



Bach et al 2017

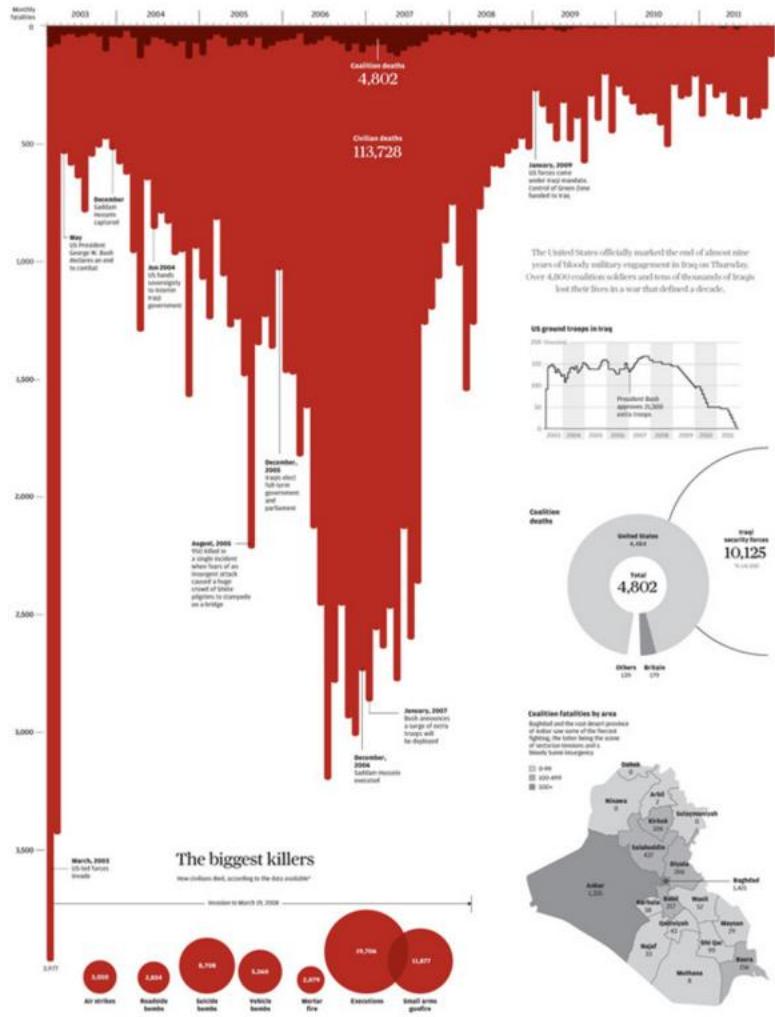
Storytelling, double-edged Sword?

Iraq's bloody toll



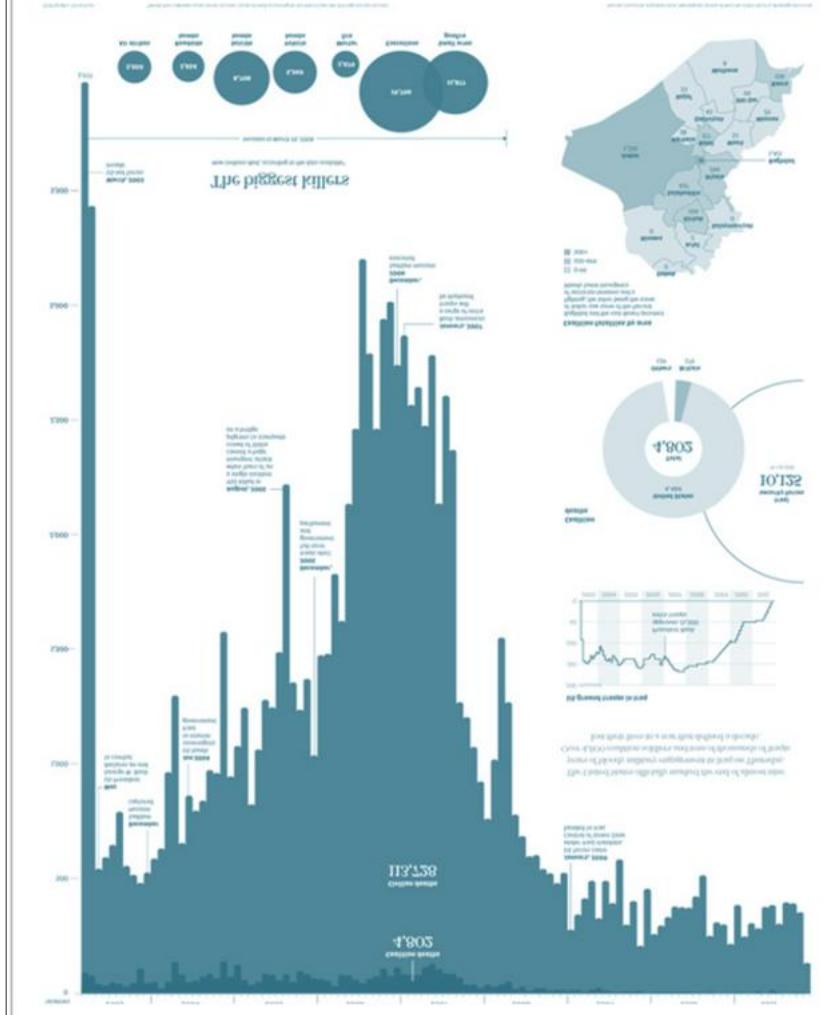
[South China Morning Post
2011]

Iraq's bloody toll



[South China Morning Post
2011]

Iraq: Deaths on the decline



Flipped upside down....

Iraq's bloody toll



Iraq: Deaths on the decline



Same Data, Different Stories

[South China Morning Post
2011]

Flipped upside down....