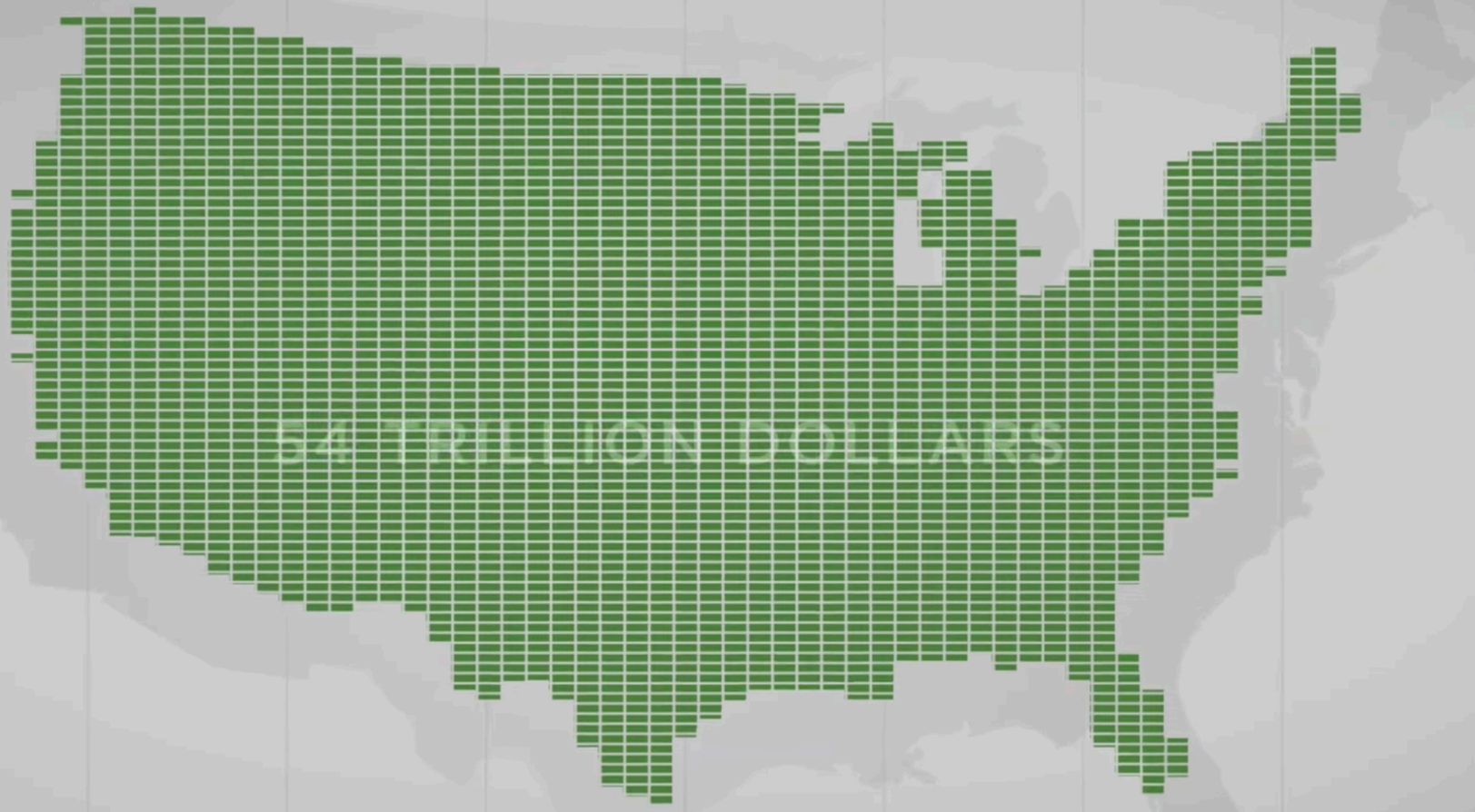
CS-5630 / CS-6630
Visualization for Data Science
Storytelling

Alexander Lex
alex@sci.utah.edu

WEALTH INEQUALITY IN AMERICA

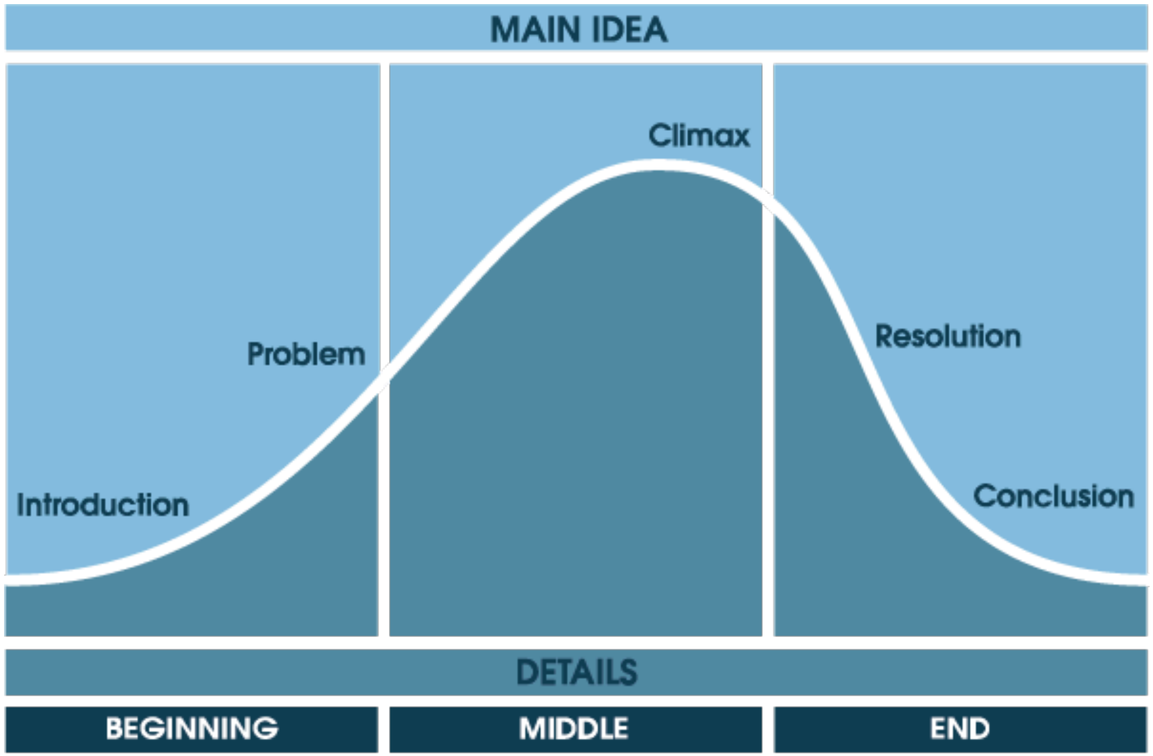


Storytelling

Good stories do more than provide facts and data

- they situate and give context
- they engage
- they educate

Who / What / Where / When
Why / How



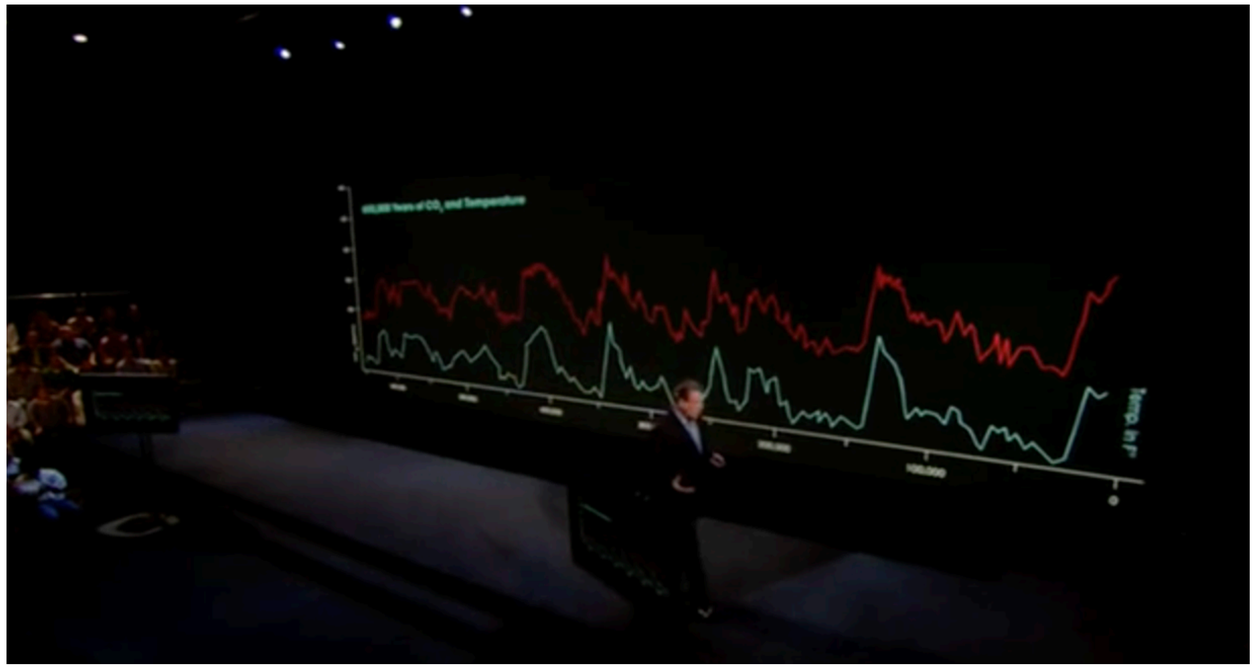
© 2016 Clever Prototypes, LLC

Storytelling with Data / VIS

Underscore your arguments with Data/Facts

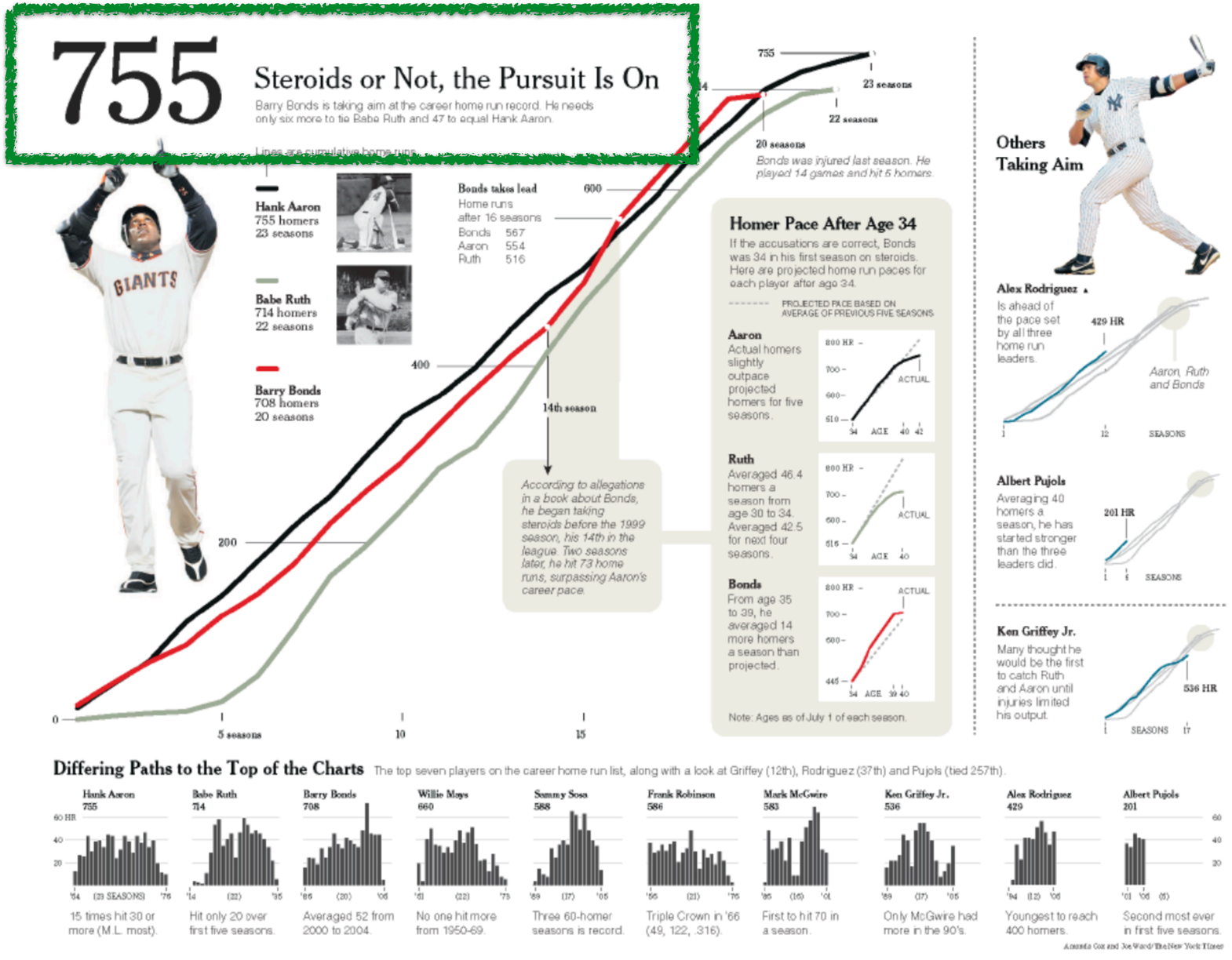
Leverage the power of Visualization

- Show Trends
- Show Correlations
- Show Outliers
- Convey Magnitudes

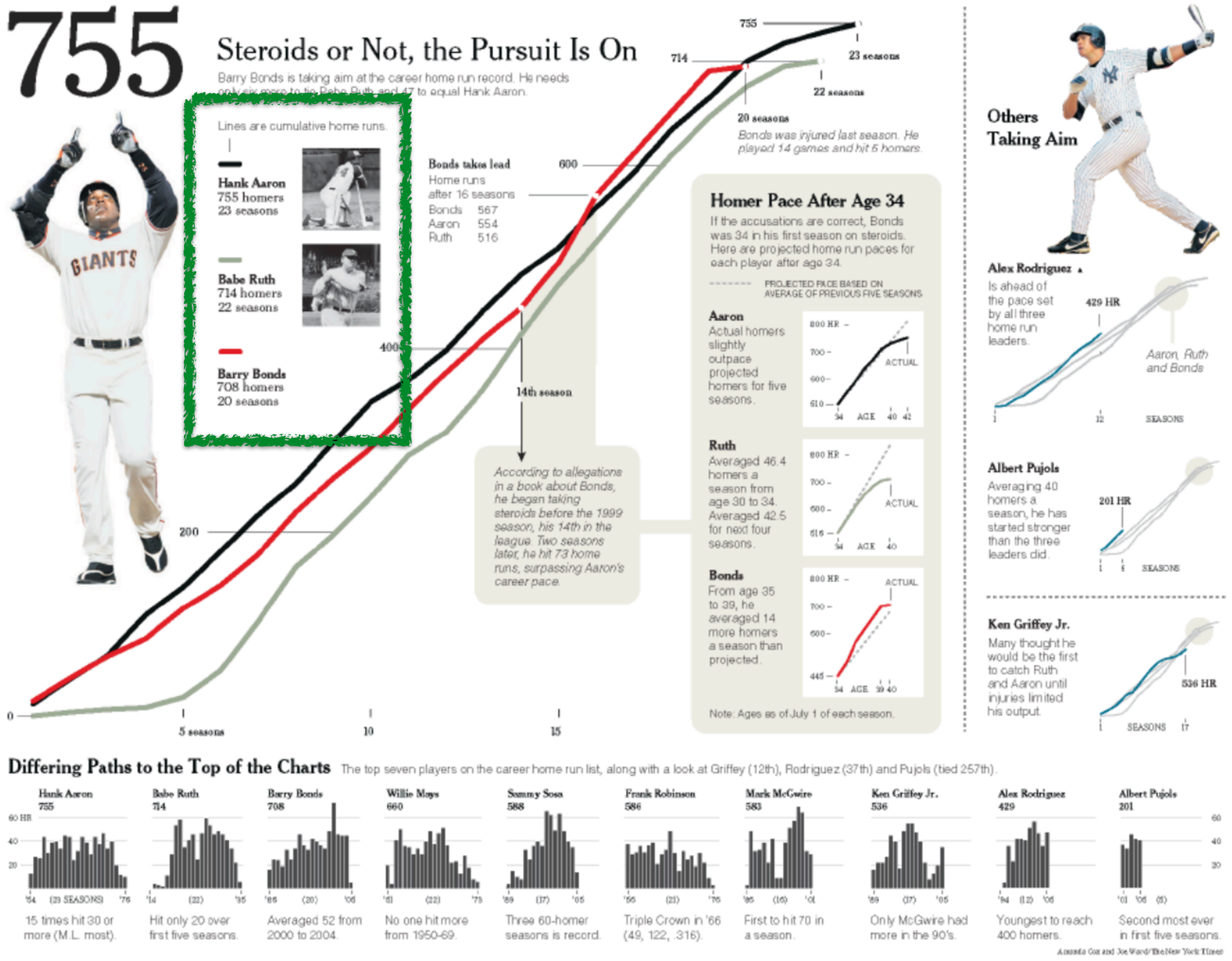


Components of a Data Story

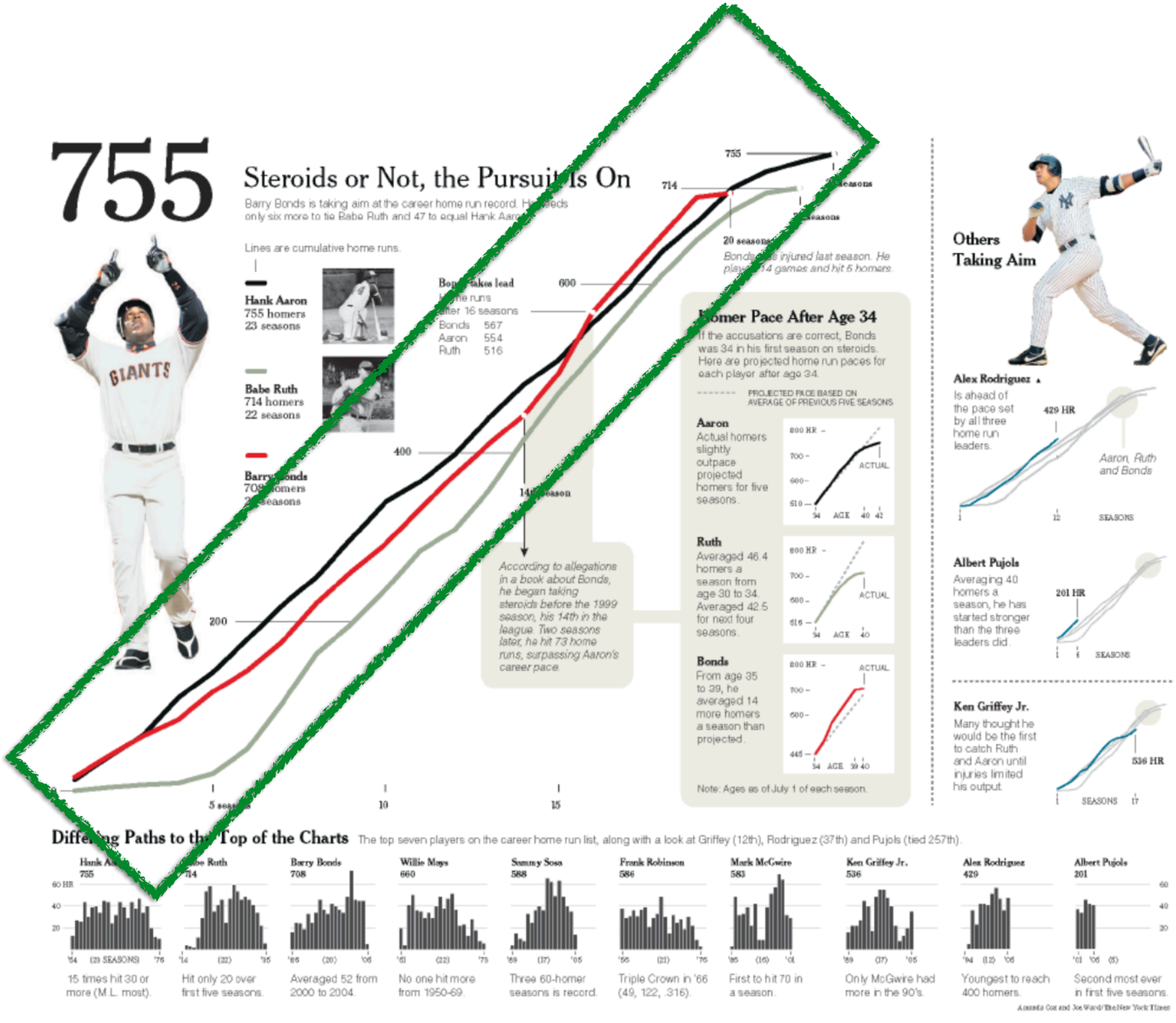
Introduction



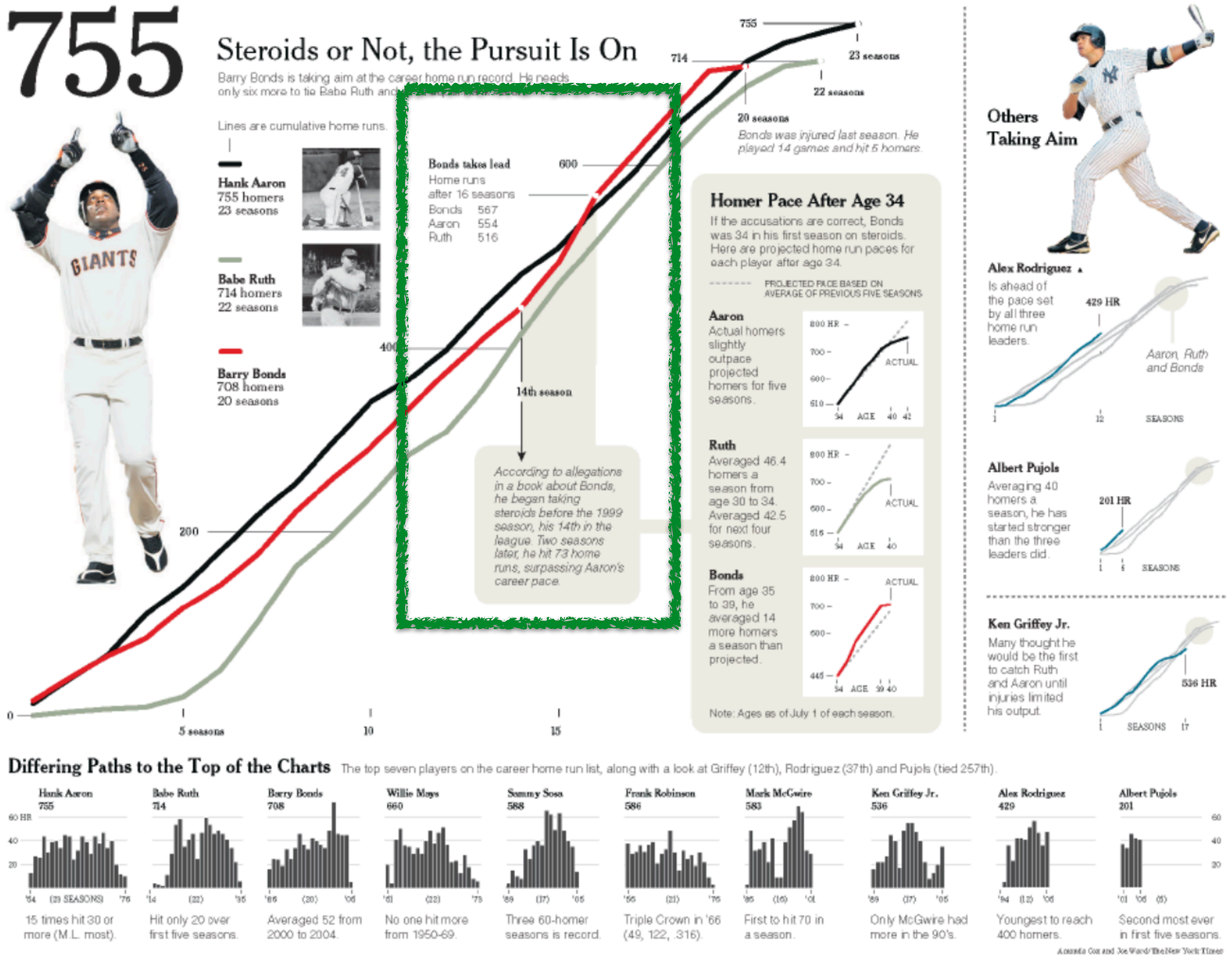
Context / Legend



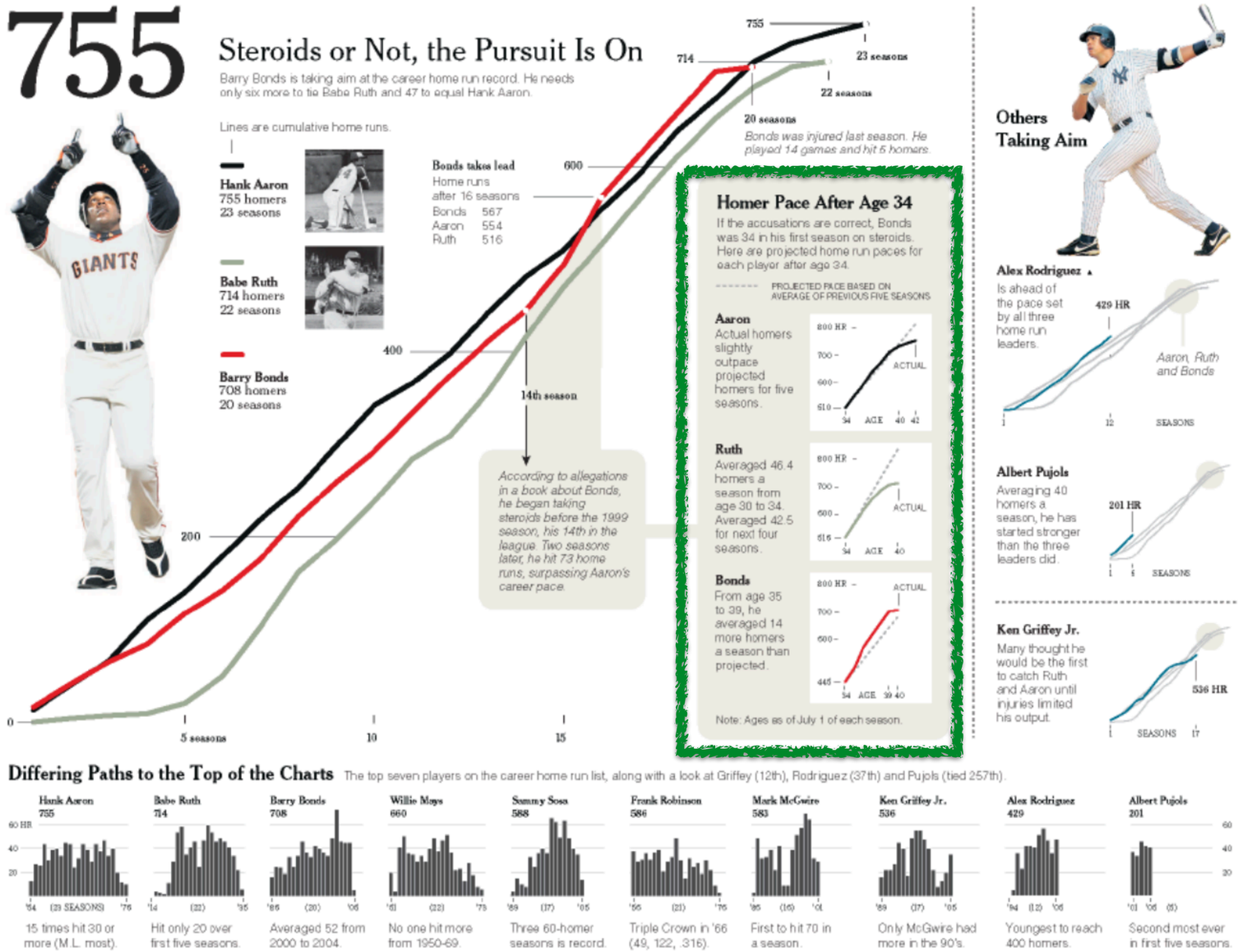
Main Story


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
22 seasons
Bonds takes lead
home runs
after 16 seasons
Bonds 567
Aaron 554
Ruth 516
755
714
600
400
200
20 seasons
Bonds was injured last season. He played 14 games and hit 5 homers.
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 14 more homers a season than projected.
800 HR
700
600
510
445
ACTUAL
AGE
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
Ken Griffey Jr.
Many thought he would be the first to catch Ruth and Aaron until injuries limited his output.
536 HR
Different 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. 536
Alex Rodriguez 429
Albert Pujols 201
60 HR
40
20
(23 SEASONS)
15 times hit 30 or more (M.L. most).
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.
Amanda Cox and Joe Ward/The New York Times

Annotation of key point: a change in trends

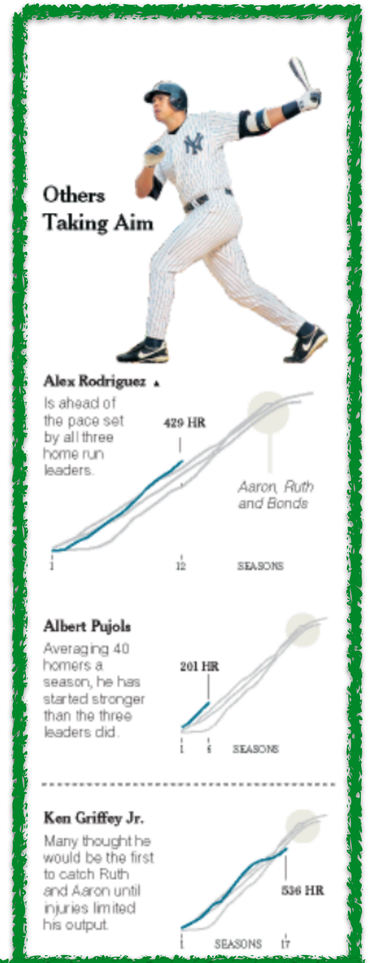


Context, using matching on content (color, trend)

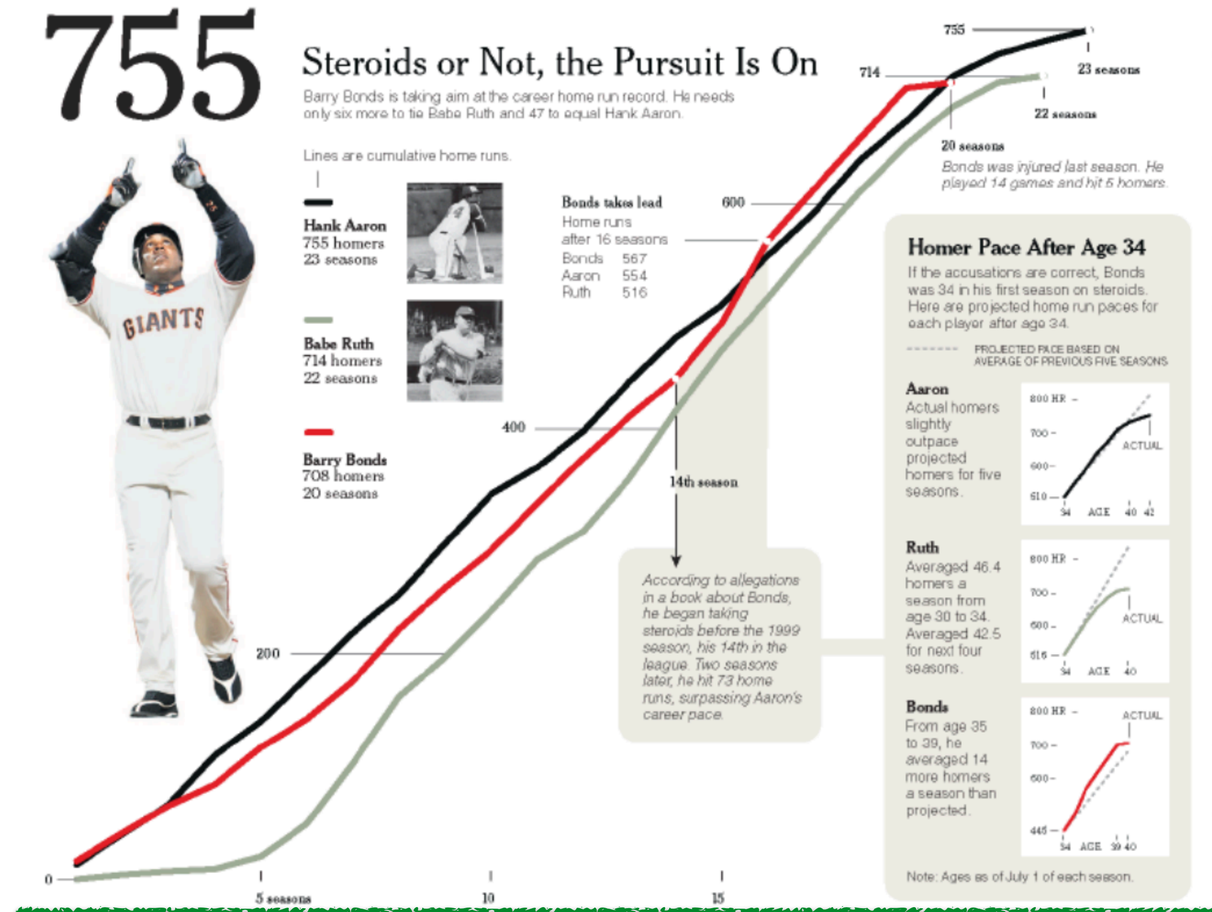


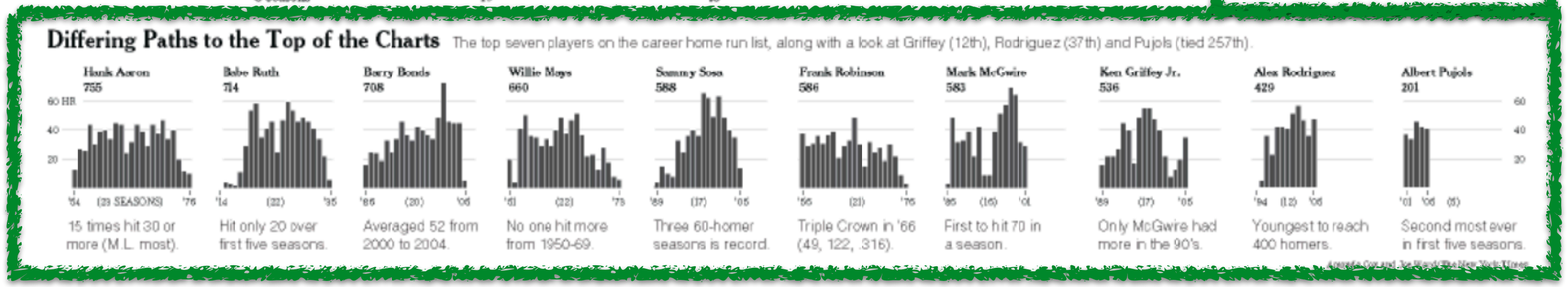
Multi-Messaging – In-Depth Information

Other Players



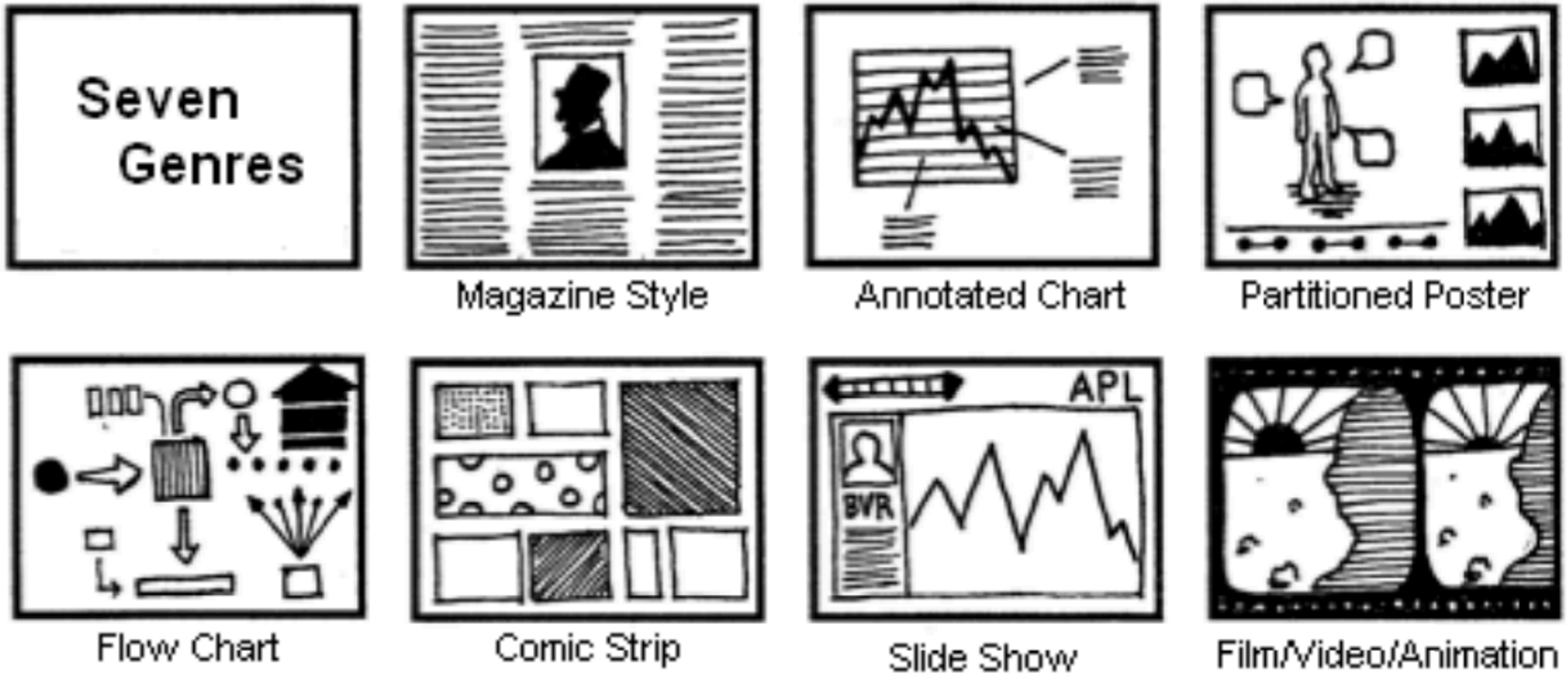
Different Paths
Why a Bar Chart?





Types of Data Stories

Genres



[Segel and Heer, 2010]

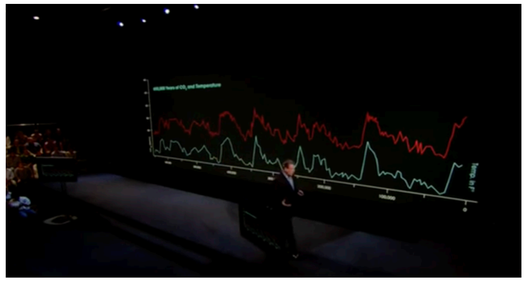
Author or Reader Driven

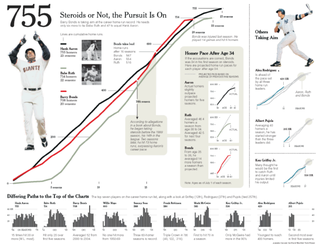
Author Driven

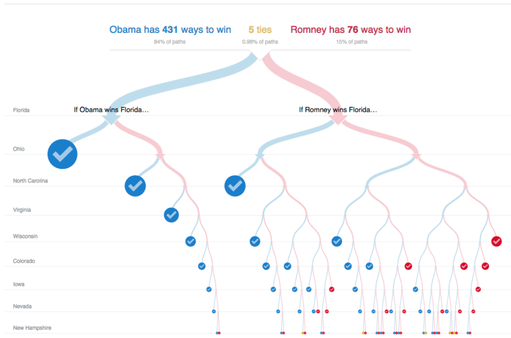
Reader Driven

Linear Ordering
Heavy Messaging
No Interactivity

No Ordering
No Messaging
Free Interactivity



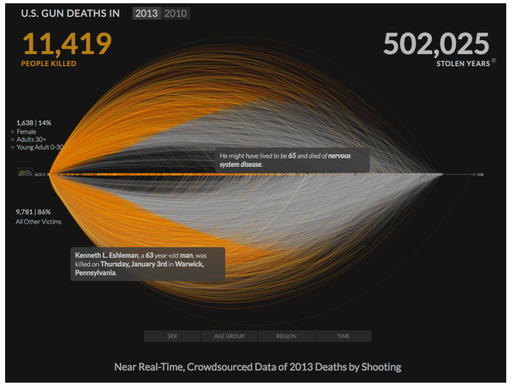




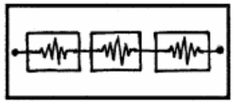
Author or Reader Driven

Martini Glass Structure

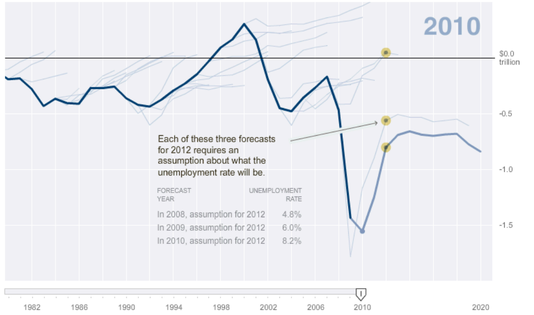
start with author driven, open up for exploration



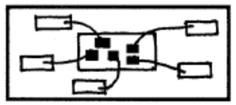
Interactive Slideshow



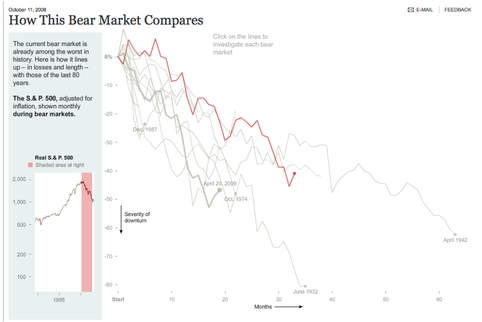
Split into multiple scenes, allow interaction mid-way



Drill-Down Story



Let reader decide which path to follow, all paths are annotated!



Strategies for Storytelling

Layout Principles

Descriptive titles

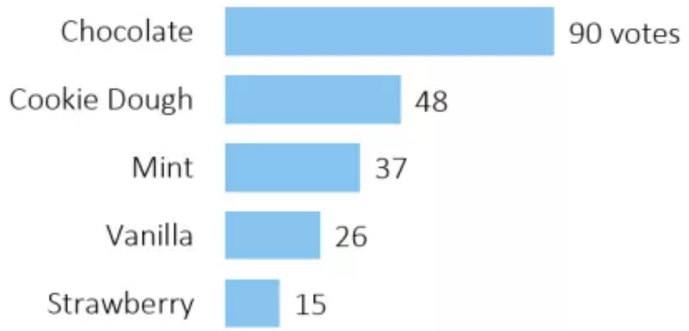
Descriptive subtitles

Annotations

Saturation

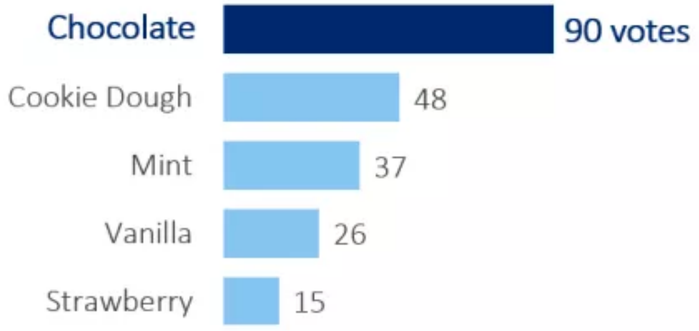
Descriptive title and **saturation** to show how chocolate is the preferred ice cream flavor

Ice cream flavor preferences based on 2014 survey of elementary school students (n=216)



or

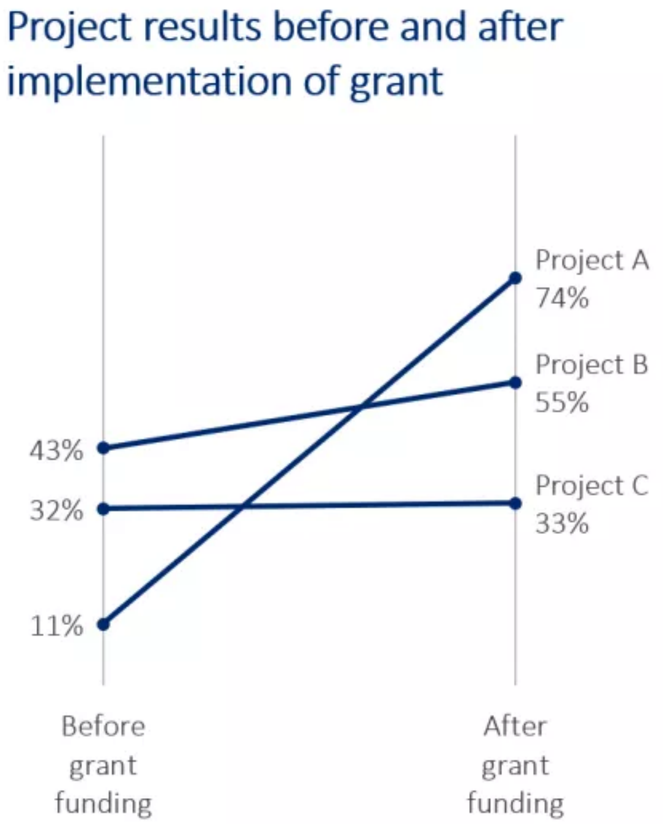
Chocolate was most popular flavor among elementary students surveyed



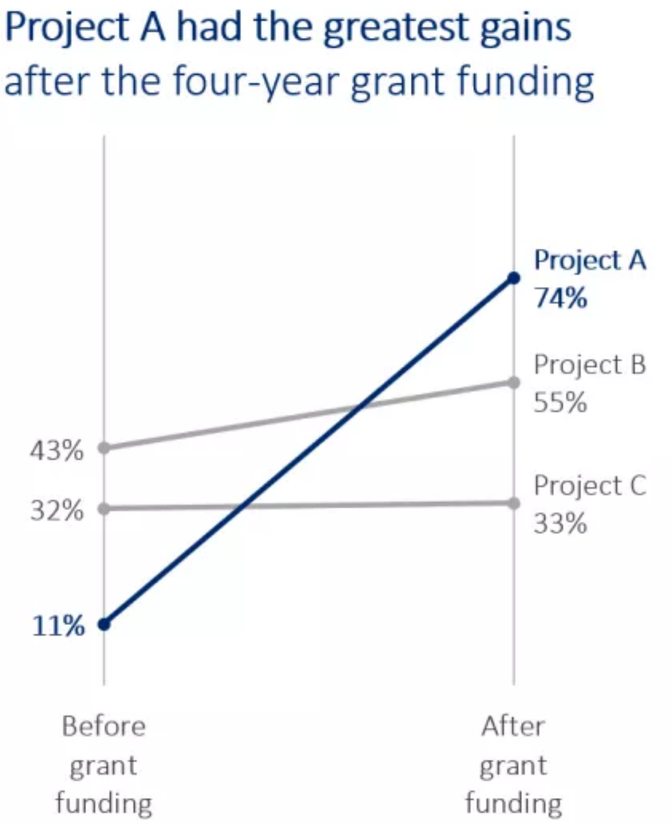
Source: 2014 survey of elementary school students (n=216)

Ann K. Emery

Descriptive title and **saturation** emphasize how Project A is performing particularly well



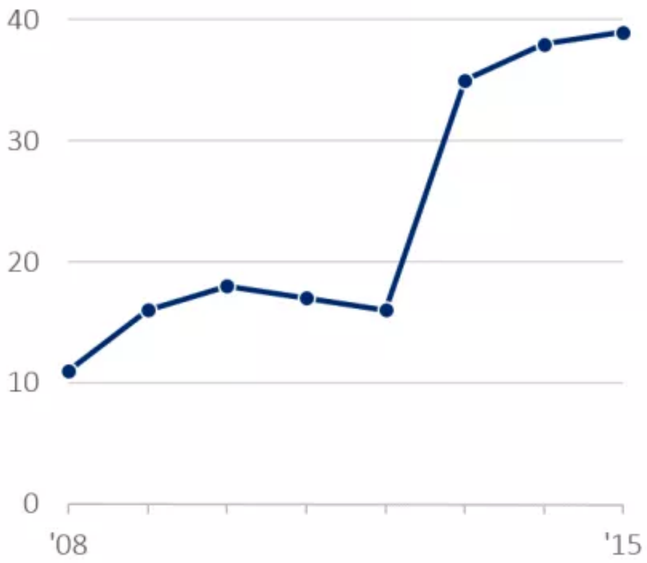
or



Ann K. Emery

Descriptive title, **descriptive subtitle,** and **annotation** for increased understanding

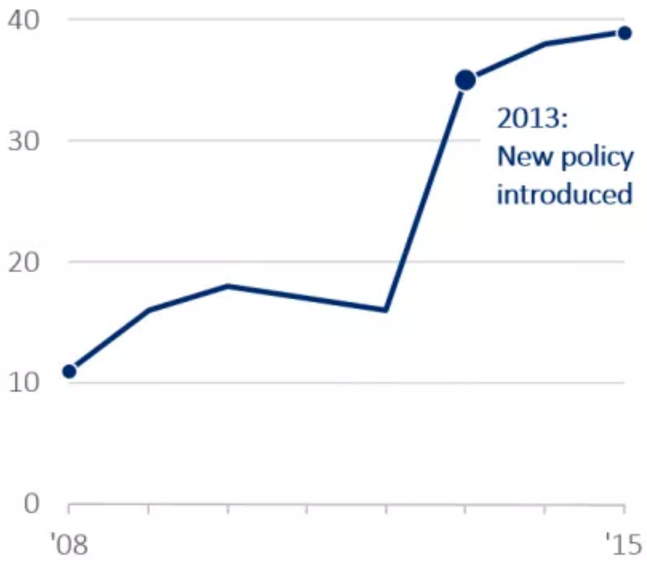
Number of studies funded each year



or

We're funding more studies each year

Beginning in 2013, we set aside new funding to measure the effectiveness of our initiatives – and we evaluated 39 of our programs in 2015 alone.



Ann K. Emery

Interactivity

Navigation

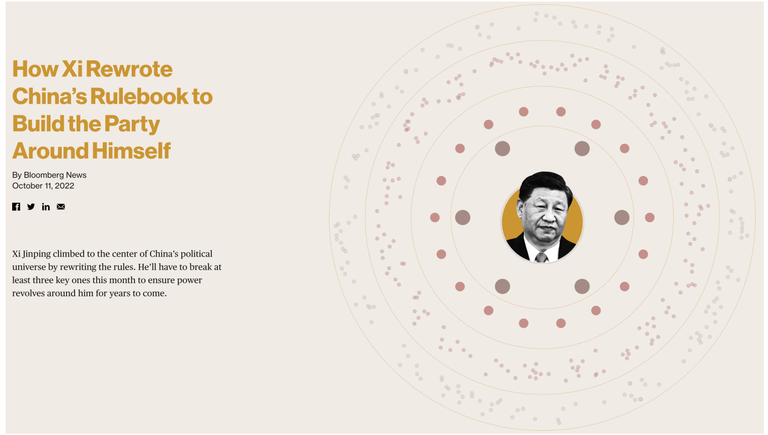
- Steppers, Scrolling, Play/Fwd/Rwd, etc.

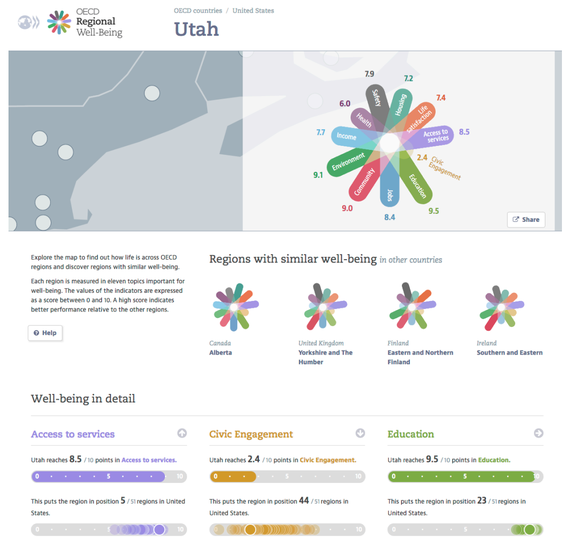
Details on Demand

- Highlights when desired

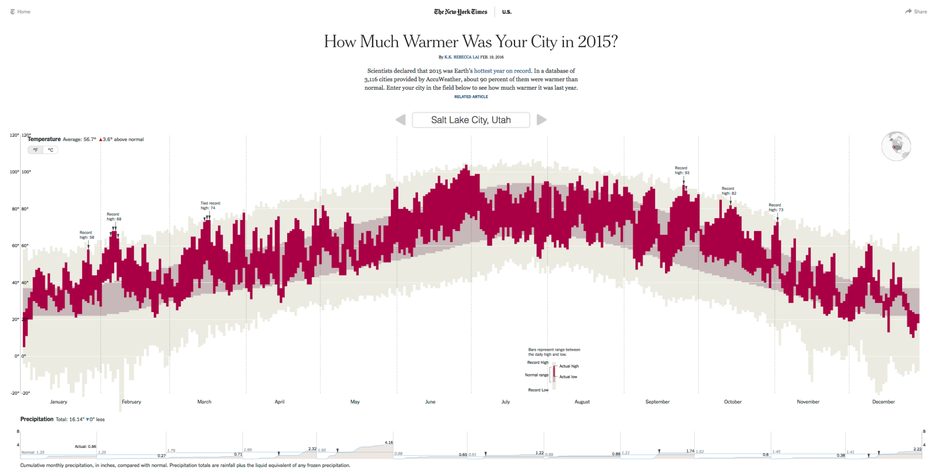
Making it relevant to the reader

- What do you think?
- Who are you?
- Where do you live?





https://www.oecdregionalwellbeing.org/US49.html

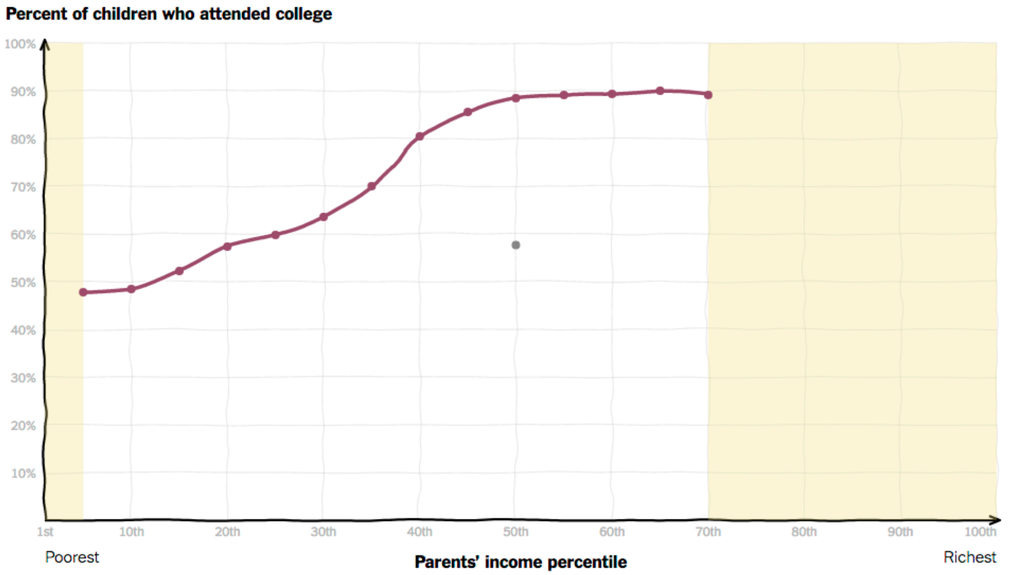


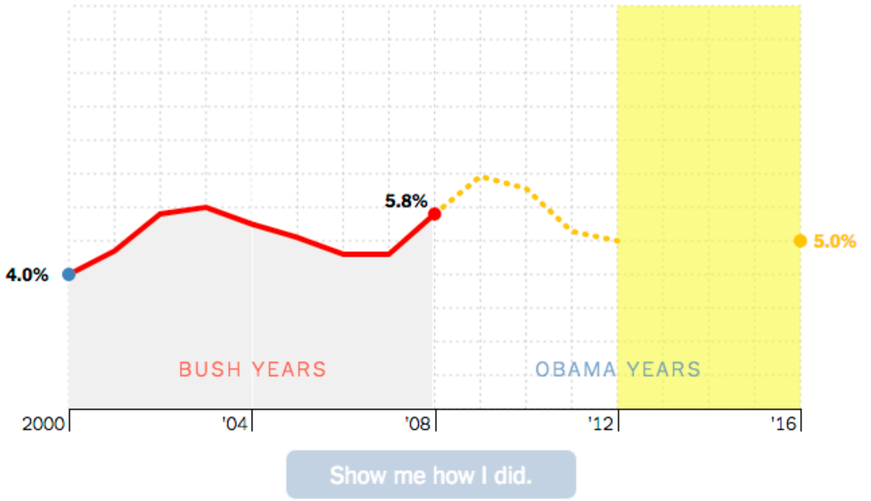
Ask for Opinions / Prior Knowledge

Ask reader to draw a trend

Juxtapose with

- reality
- what other people think



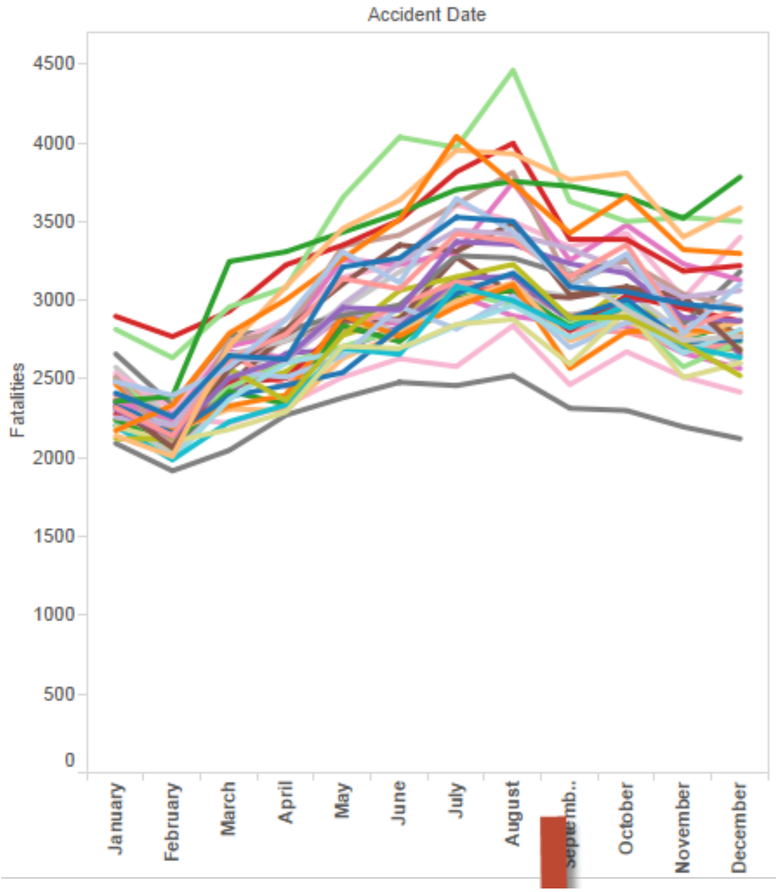


Design Considerations

Fewer people are dying on US roads, but seasonal trends

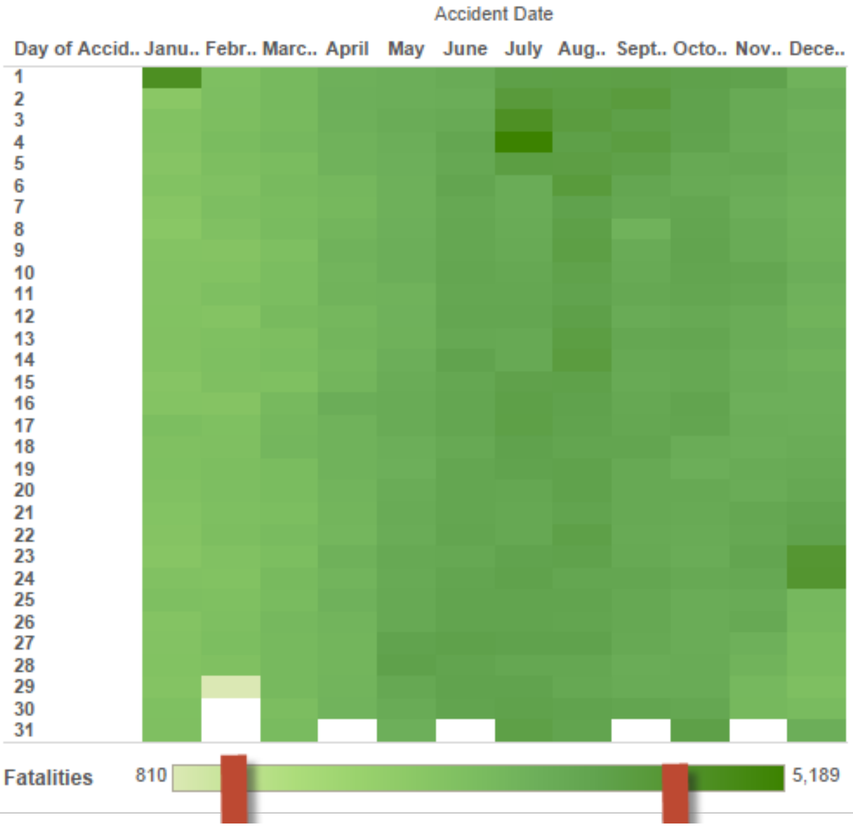
There have been 1,236,670 road fatalities in the US between 1979 and 2011. This dashboard explores the seasonality in the data.

1. Every year sees the same peaks and troughs in fatalities



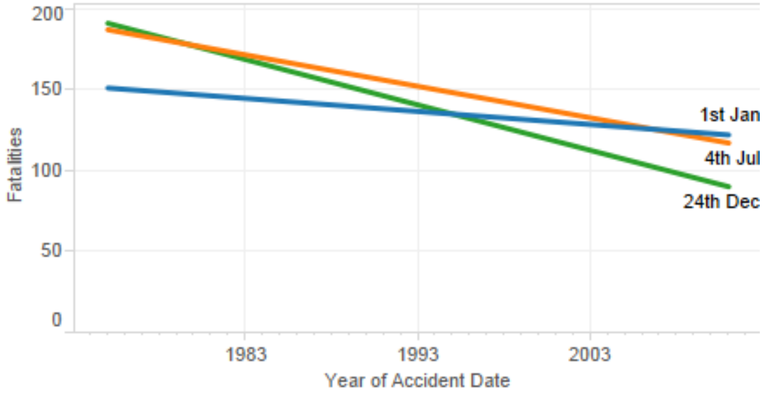
2. Three holidays (New Year, July 4th and Christmas) are particularly lethal

Highlight table of fatalities by day and month for years: All

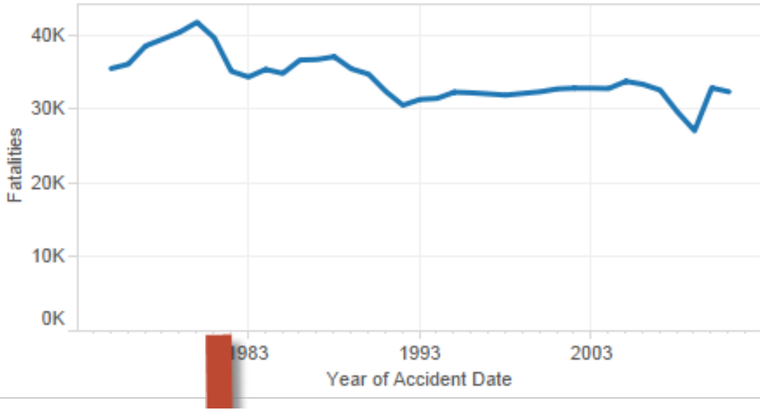


3. Fortunately, those holidays are getting safer...

Slope chart of deaths on the 3 most dangerous days



...which is part of an overall reduction in fatalities



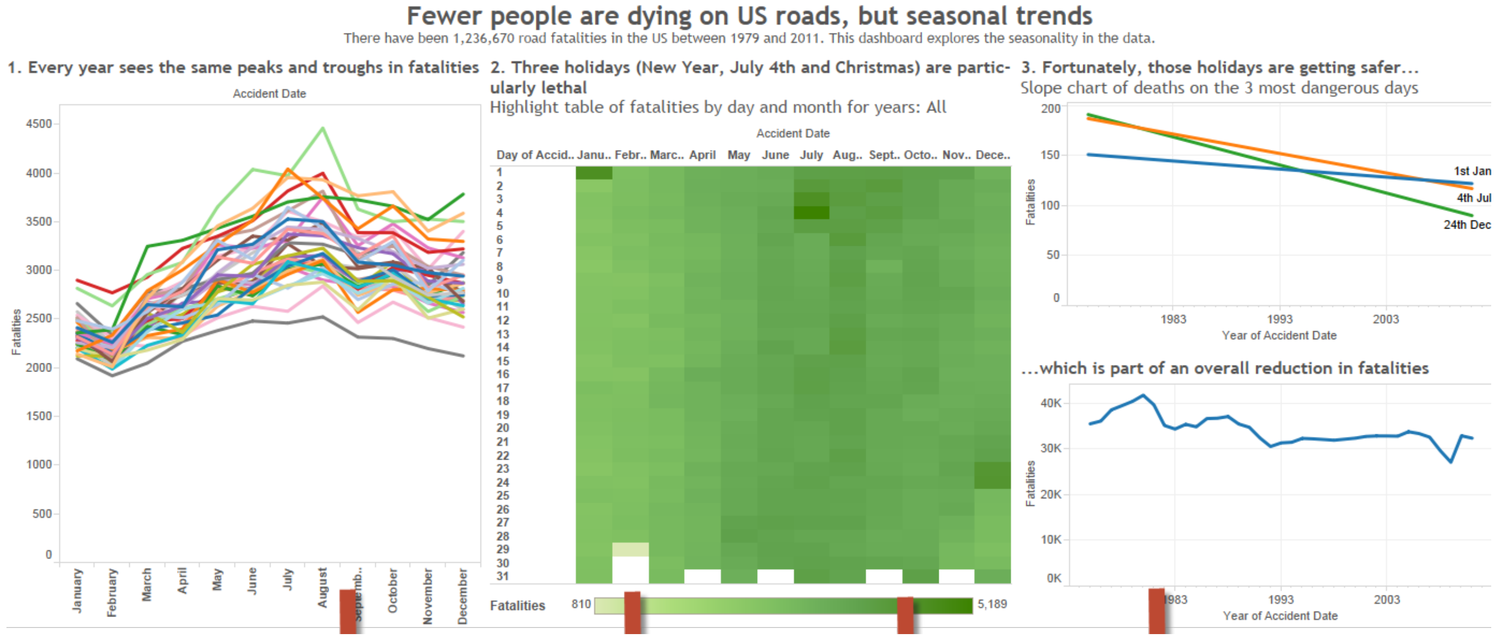
http://gravyanecdote.com/category/tableau-design-month/

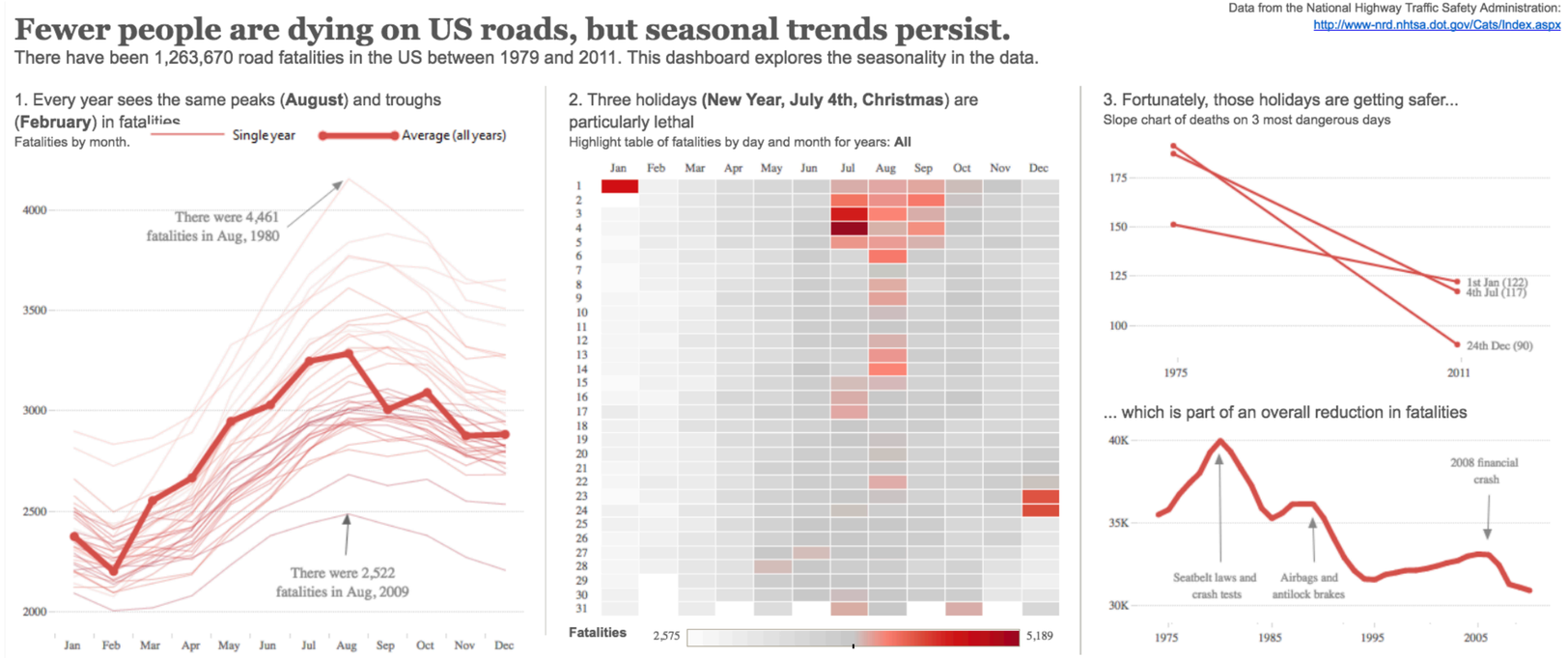
Fewer Colors

Averages for Context

Better Scales

Richer Annotations





http://gravyanecdote.com/category/tableau-design-month/

Engagement

Know Your Audience

People you don’t know are difficult to influence

What do they know?

What motivates them?

What experiences do you share?

What are common goals?

What insights can you give them?

Target Audience

Opinionated

Public Media

Expert Panel

Education

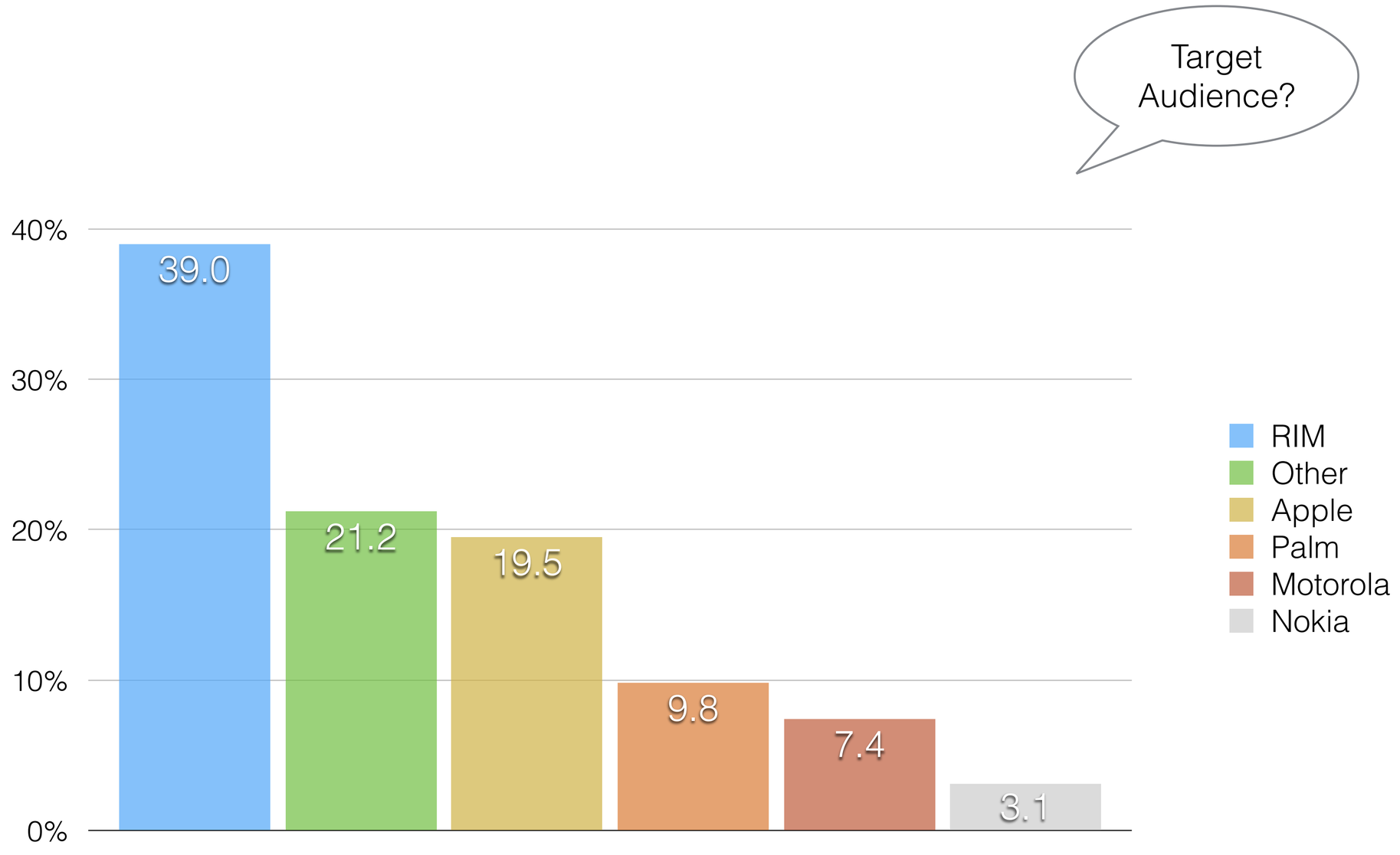
Board Meeting

Group Meeting

Neutral

Low
Information
Density

High
Information
Density


Target Audience?
40%
30%
20%
10%
0%
39.0
21.2
19.5
9.8
7.4
3.1
RIM
Other
Apple
Palm
Motorola
Nokia

Target Audience?

Opportunity Dashboard | Sales Dashboard

Sales Dashboard

Total Sales	Number of Deals	Avg Deal Size	Rev. per Salesperson
$3,190.2M	16,610	$189,545	$20.1M

Date Closed
8/7/2006 – 12/31/2010

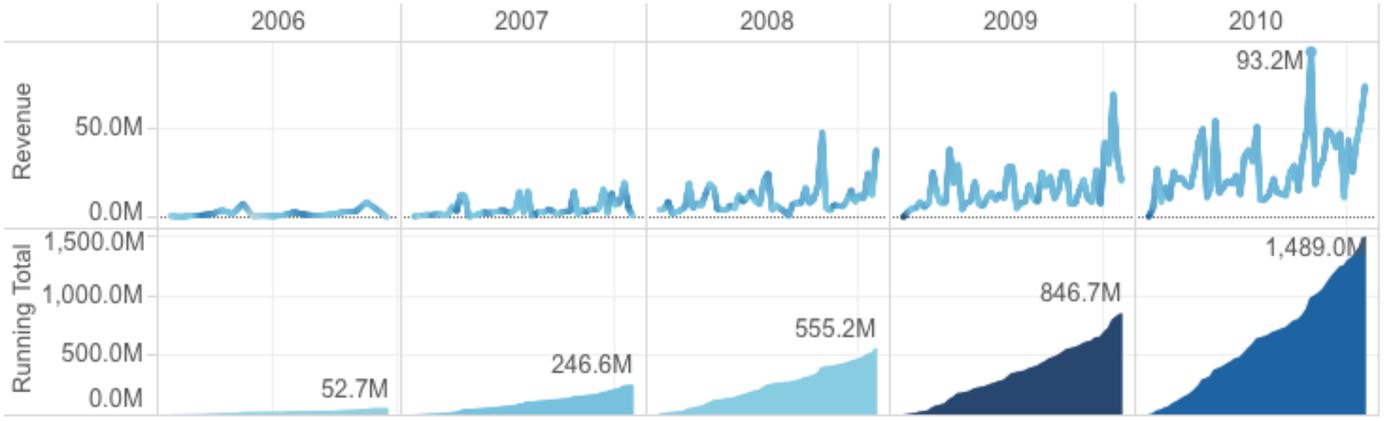
Region
(All)

Country
(All)

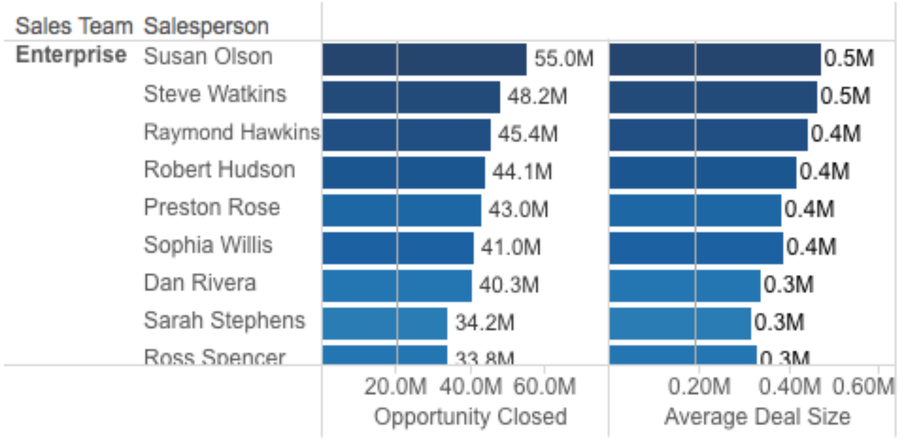
Sales Team
- (All)
- Small and Midmarket
- Enterprise

Avg Deal Size/Salespe...
$130,922 – $336,519

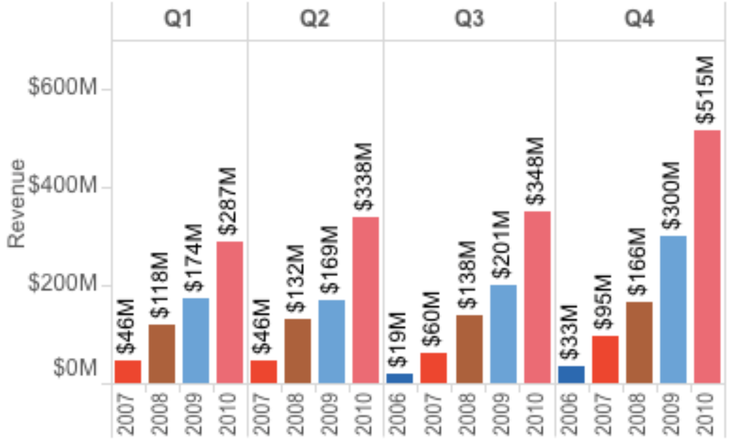
Revenue Over Time



Sales Team Performance



Revenue by Quarter



WHERE THERE'S SMOKE—THERE'S CANCER

Cancer rates are up, but mortality is down. New diagnostics and treatments are responsible for part of this trend. But the greatest single contributing factor is the decline in smoking—rates are at their lowest level in 50 years.

Target Audience?

Men Women

1 Increased incidence

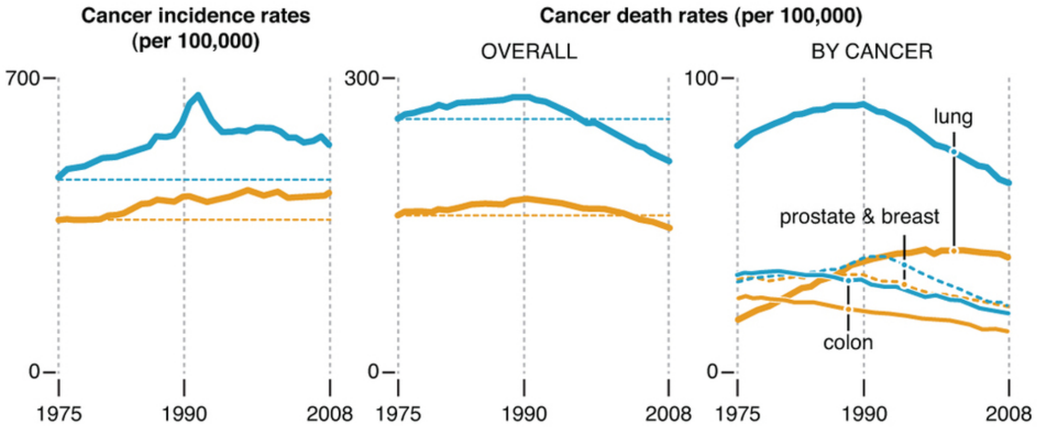
An aging population contributes to rising incidence of cancer.

2 Fewer deaths

Cancer deaths have been dropping since 1991, especially in males.

3 Decline of lung cancer

Drop in lung cancer deaths in males is the primary reason why death rates are down.

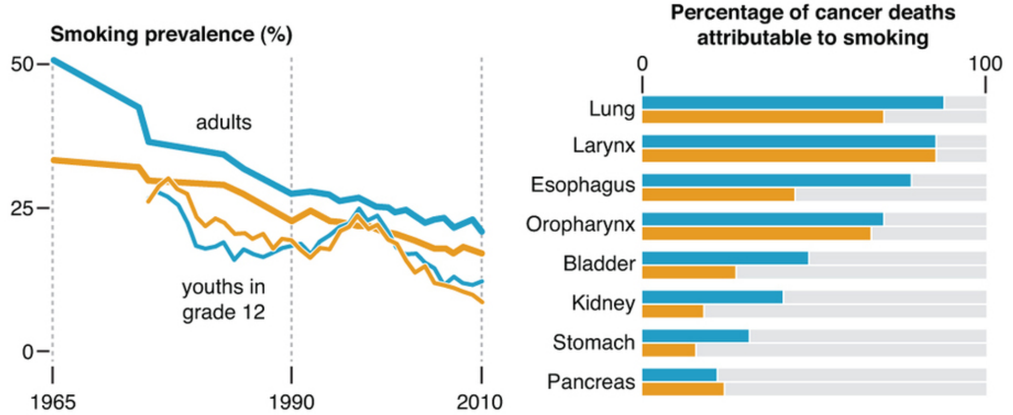


4 Decline in smoking

Since the 1964 first Surgeon General's report, smoking rates have been dropping. By 2010, the rate among males was down to 20%, from 50% at its peak. Among youths, rates have been on an even steeper decline since 1997.

5 Impact of smoking on cancer deaths

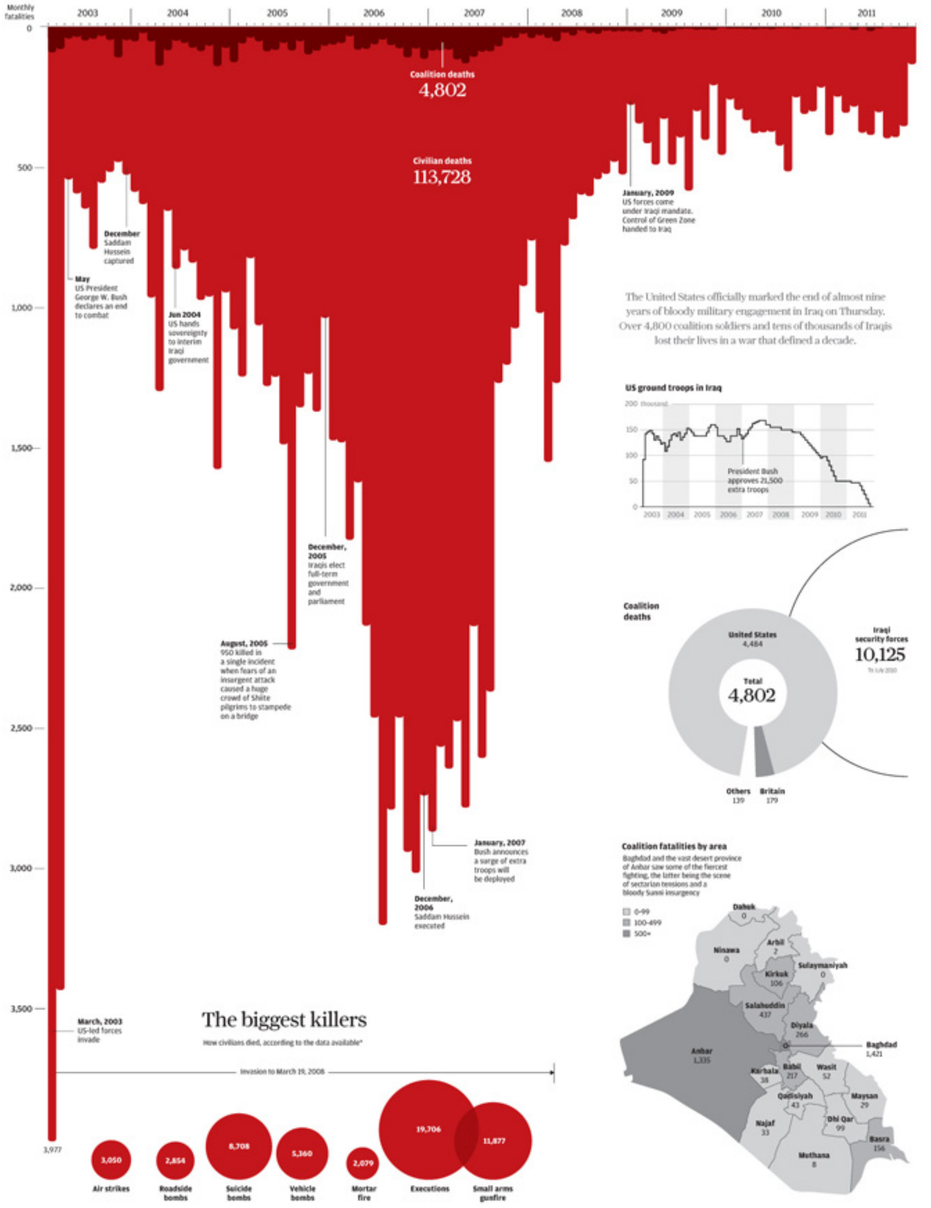
Smoking is a major risk factor for many types of cancer and significant contributor to cancer-related deaths. It remains the single largest preventable cause of disease and premature death in the US.



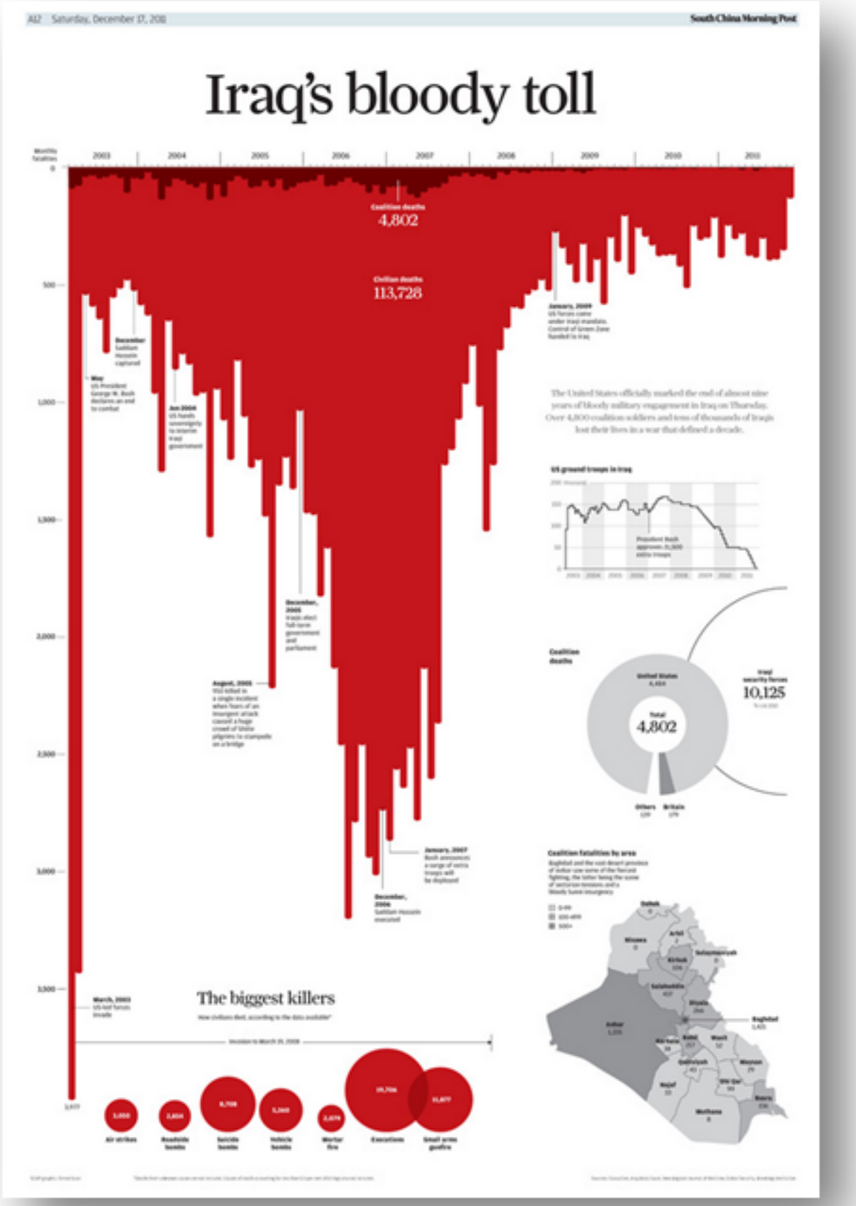
source: American Cancer Society Cancer Statistics 2012; Monitoring the Future (University of Michigan).

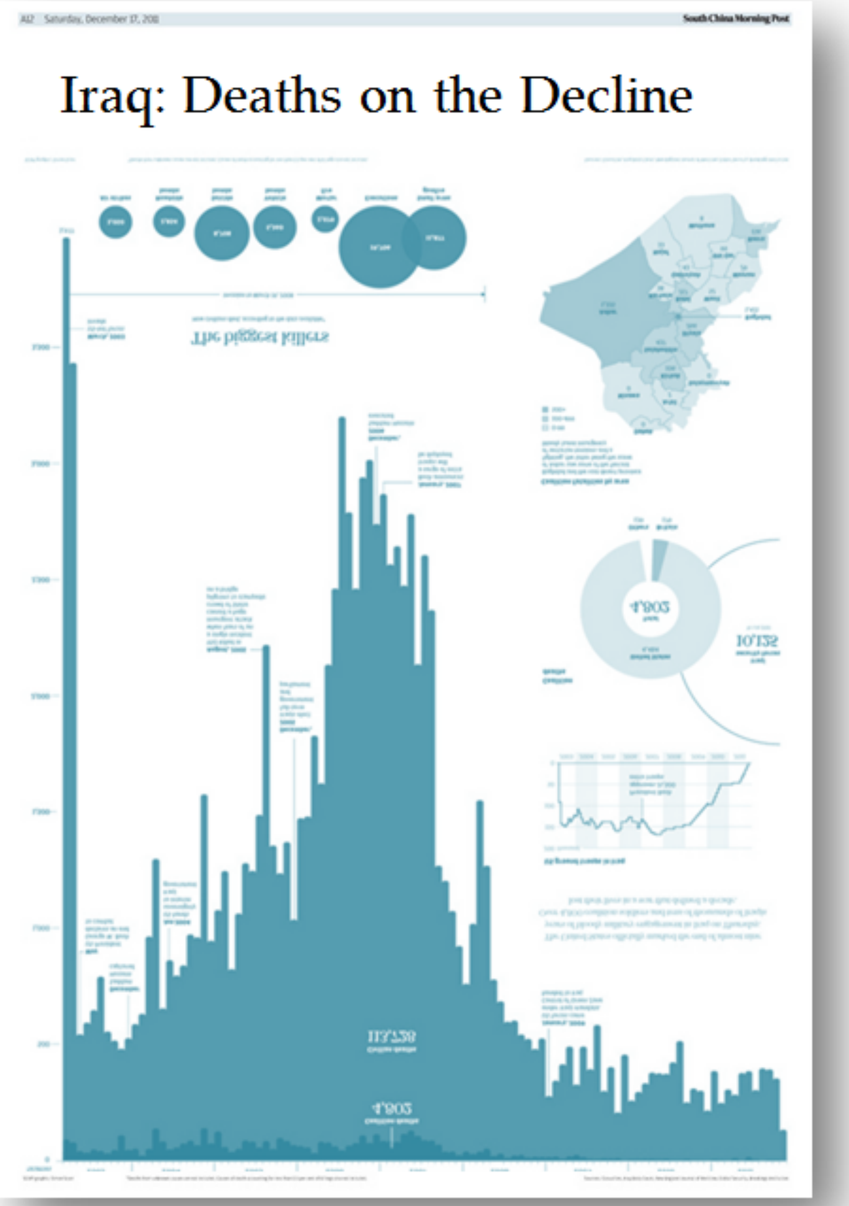
M. Krzywinski & A. Cairo

Iraq's bloody toll



Target Audience?



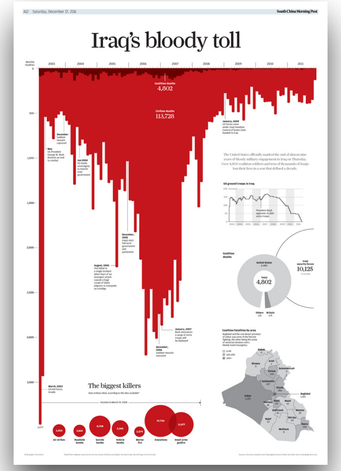


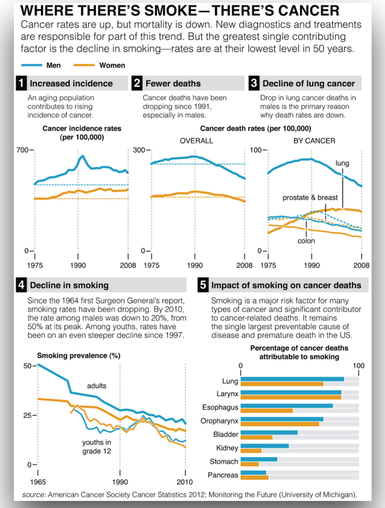
Andy Cotgreave, Tableau

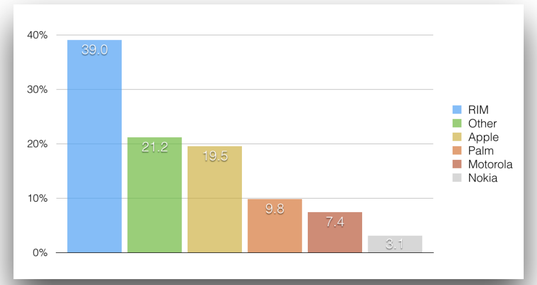
Target Audience

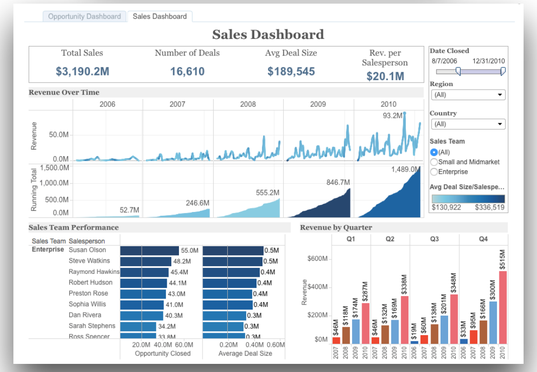
Opinionated

Neutral







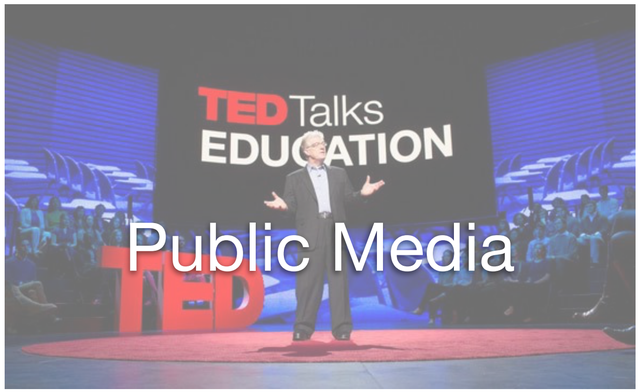


Low Information Density

High Information Density

Target Audience

Opinionated

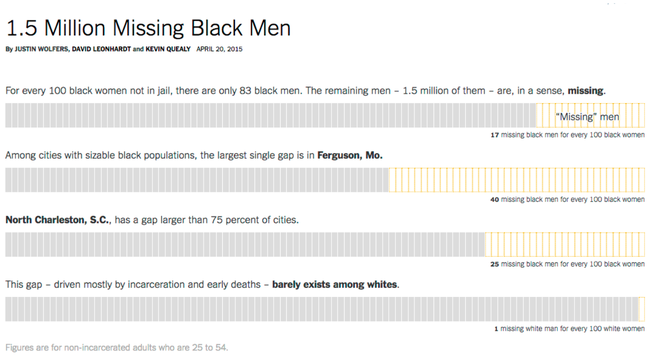


Neutral

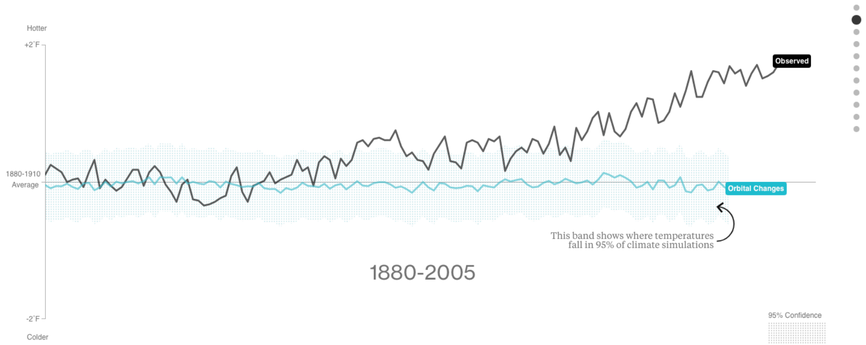
Low
Information
Density

High
Information
Density

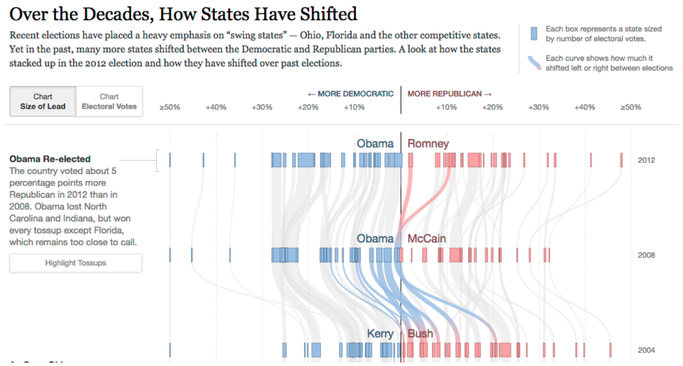
Exercise: Analyze Stories



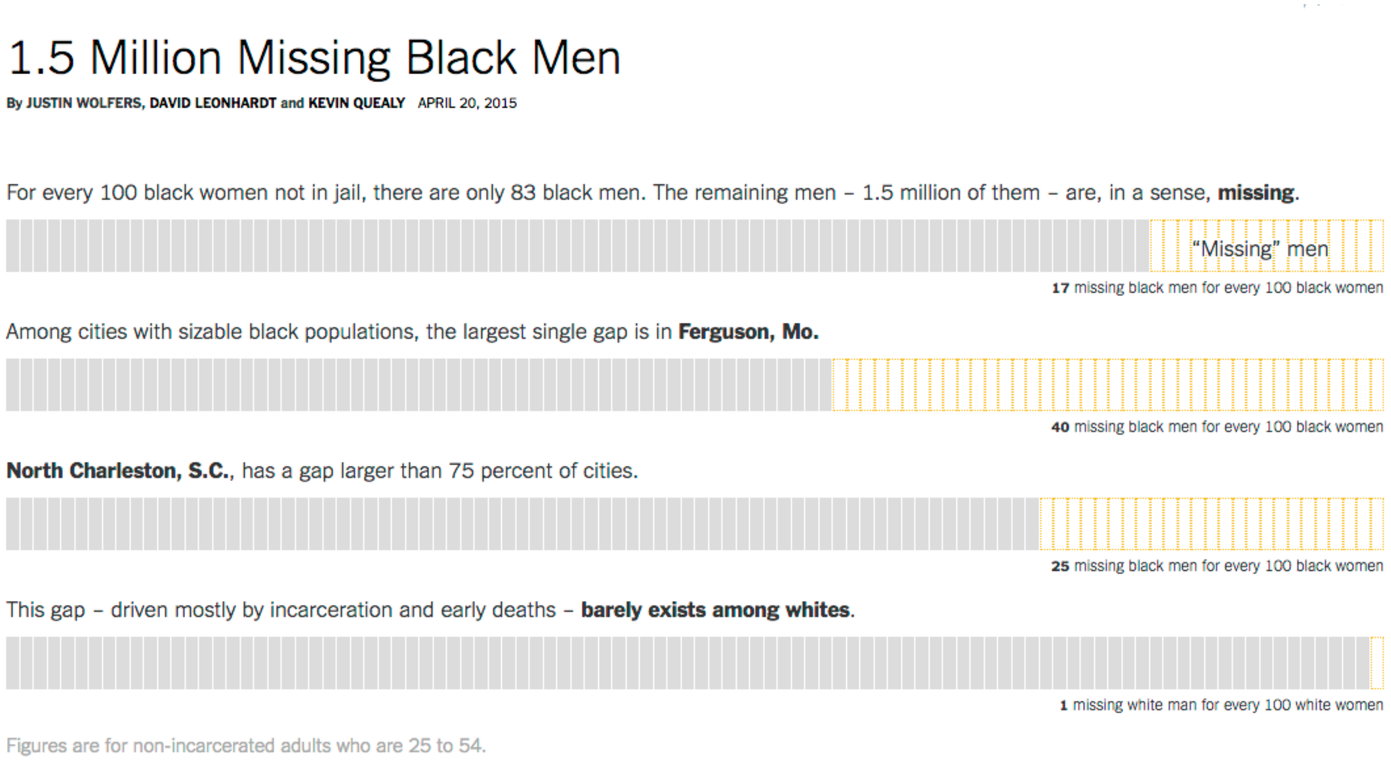
https://www.nytimes.com/interactive/2015/04/20/upshot/missing-black-men.html



https://www.bloomberg.com/graphics/2015-whats-warming-the-world/



http://www.nytimes.com/interactive/2012/10/15/us/politics/swing-history.html



What's the Genre?
(Magazine, Annotated Chart, Partitioned Poster, Flow Chart, Comic Strip, Slide Show, Video/Animation)

Is it Author or Reader Driven?

Can you make out a particular structure?
(Martini Glass, Interactive Slideshow, Drill-Down Story)? Why?

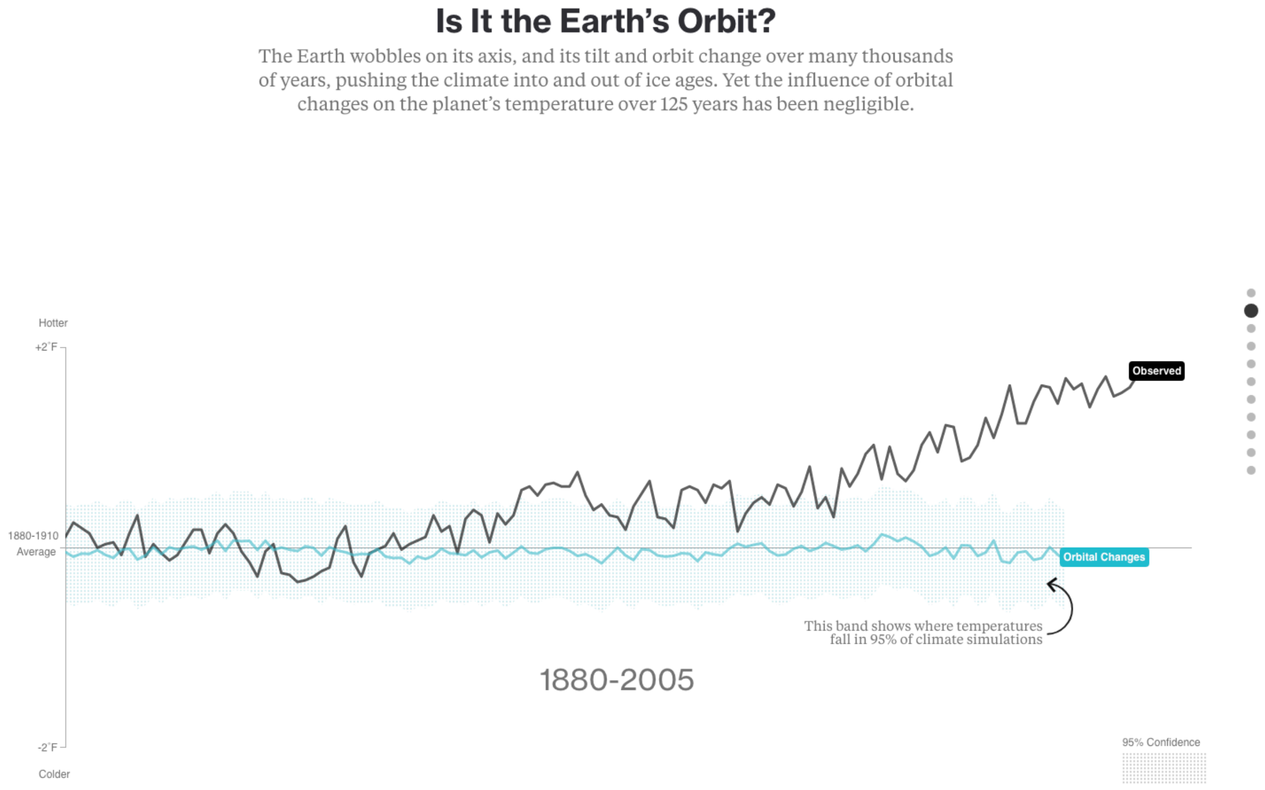
Does it have a good Introduction?
What medium is used for the introduction?

Is the story opinionated or neutral?

Does it use titles, legends, and highlights well?

Does it use interactivity well?

Is the visualization well designed?



What's the Genre?
(Magazine, Annotated Chart, Partitioned Poster, Flow Chart, Comic Strip, Slide Show, Video/Animation)

Is it Author or Reader Driven?

Can you make out a particular structure?
(Martini Glass, Interactive Slideshow, Drill-Down Story)? Why?

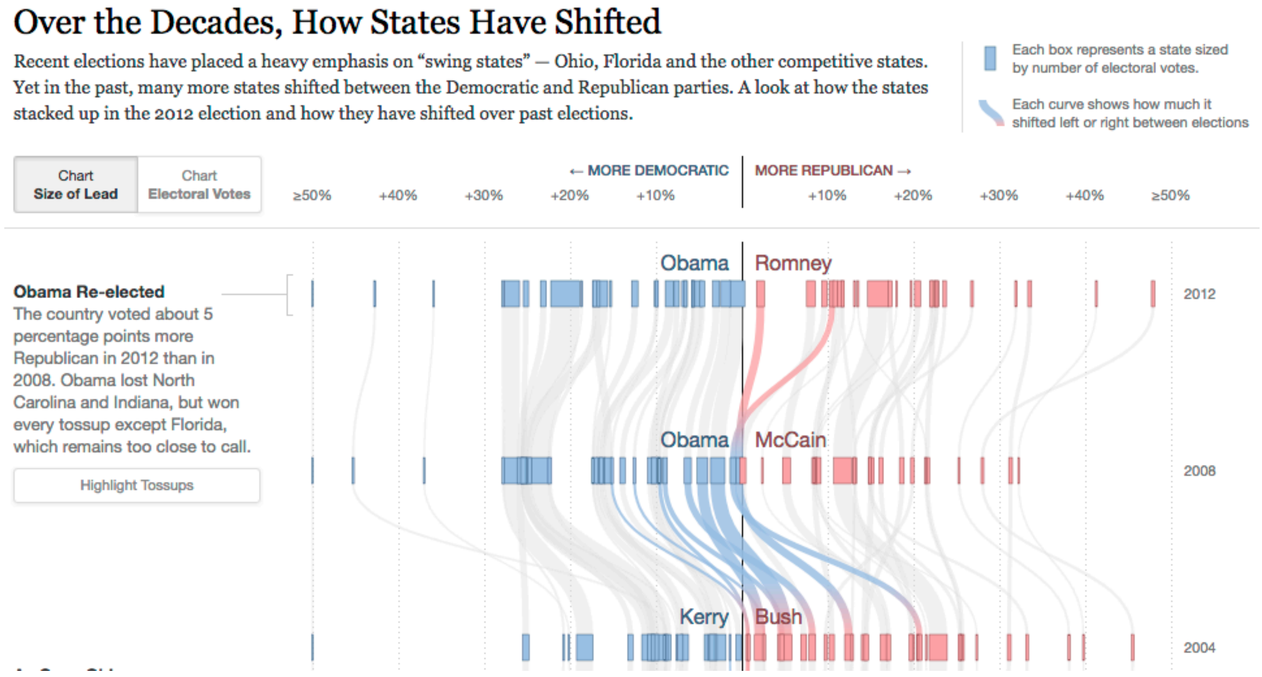
Does it have a good Introduction?
What medium is used for the introduction?

Is the story opinionated or neutral?

Does it use titles, legends, and highlights well?

Does it use interactivity well?

Is the visualization well designed?



What's the Genre?
(Magazine, Annotated Chart, Partitioned Poster, Flow Chart, Comic Strip, Slide Show, Video/Animation)

Is it Author or Reader Driven?

Can you make out a particular structure?
(Martini Glass, Interactive Slideshow, Drill-Down Story)? Why?

Does it have a good Introduction?
What medium is used for the introduction?

Is the story opinionated or neutral?

Does it use titles, legends, and highlights well?

Does it use interactivity well?

Is the visualization well designed?

Please check Slack #announcements for the link to activity