

VISUALIZATION IN PYTHON

Agnes Chang
PSYC GR6130, Fall 2019

AGENDA

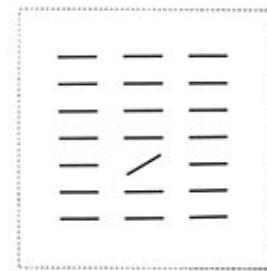
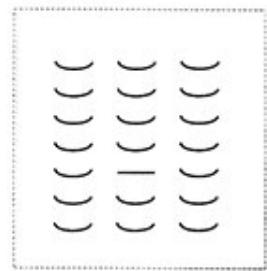
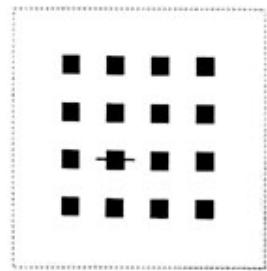
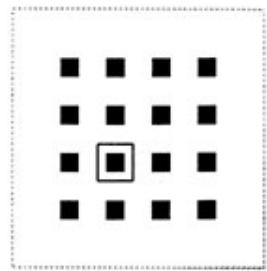
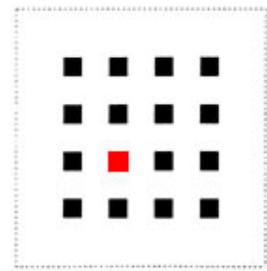
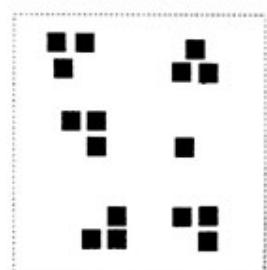
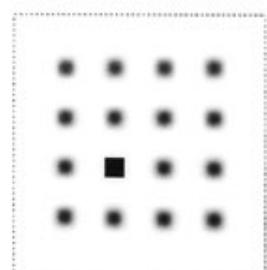
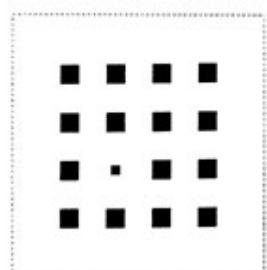
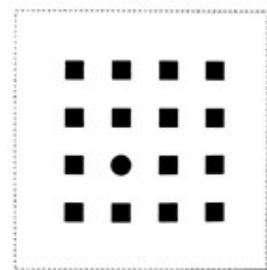
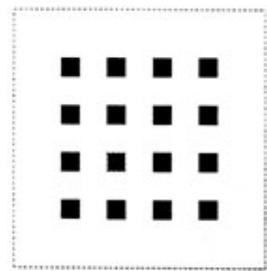
- Perception: pre-attentive attributes, gestalt principles
- Color scales: sequential, diverging, categorical
- Some design rules of thumb, to follow and to break
- Misrepresentation
- Showing uncertainty/probability

PERCEPTION



649772334445286881590457124461225721722
9630183039127152626657049820796240423193
6468364856661829738051526813724187260073

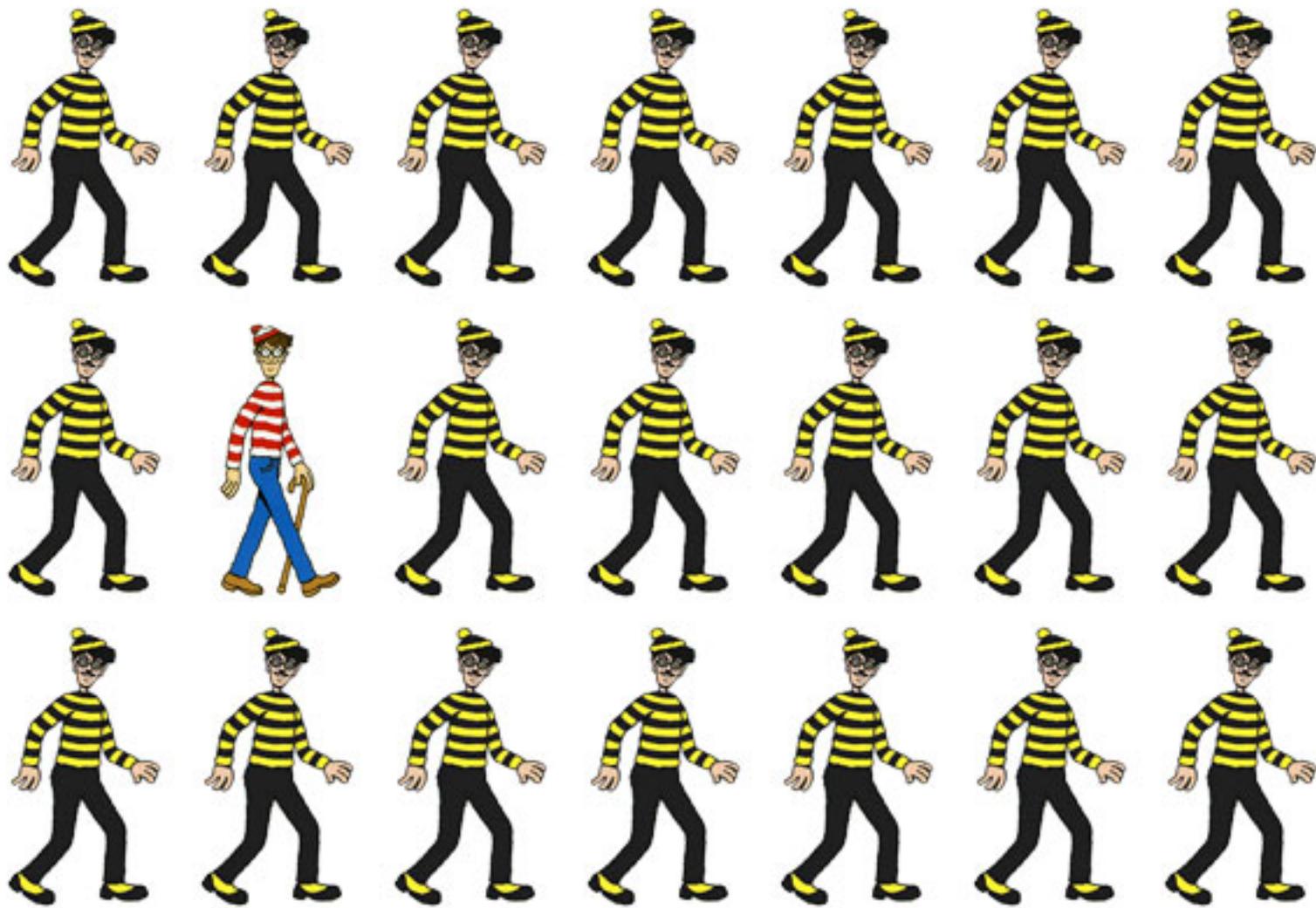
6497723344445286881590457124461225721722
9630183039127152626657049820796240423193
6468364856661829738051526813724187260073

LINE ORIENTATION**LINE LENGTH****LINE WEIGHT****CURVATURE****ADDED MARKS****ENCLOSURE****COLOR/HUE****INTENSITY/VALUE**

Drawn by I. Meirelles

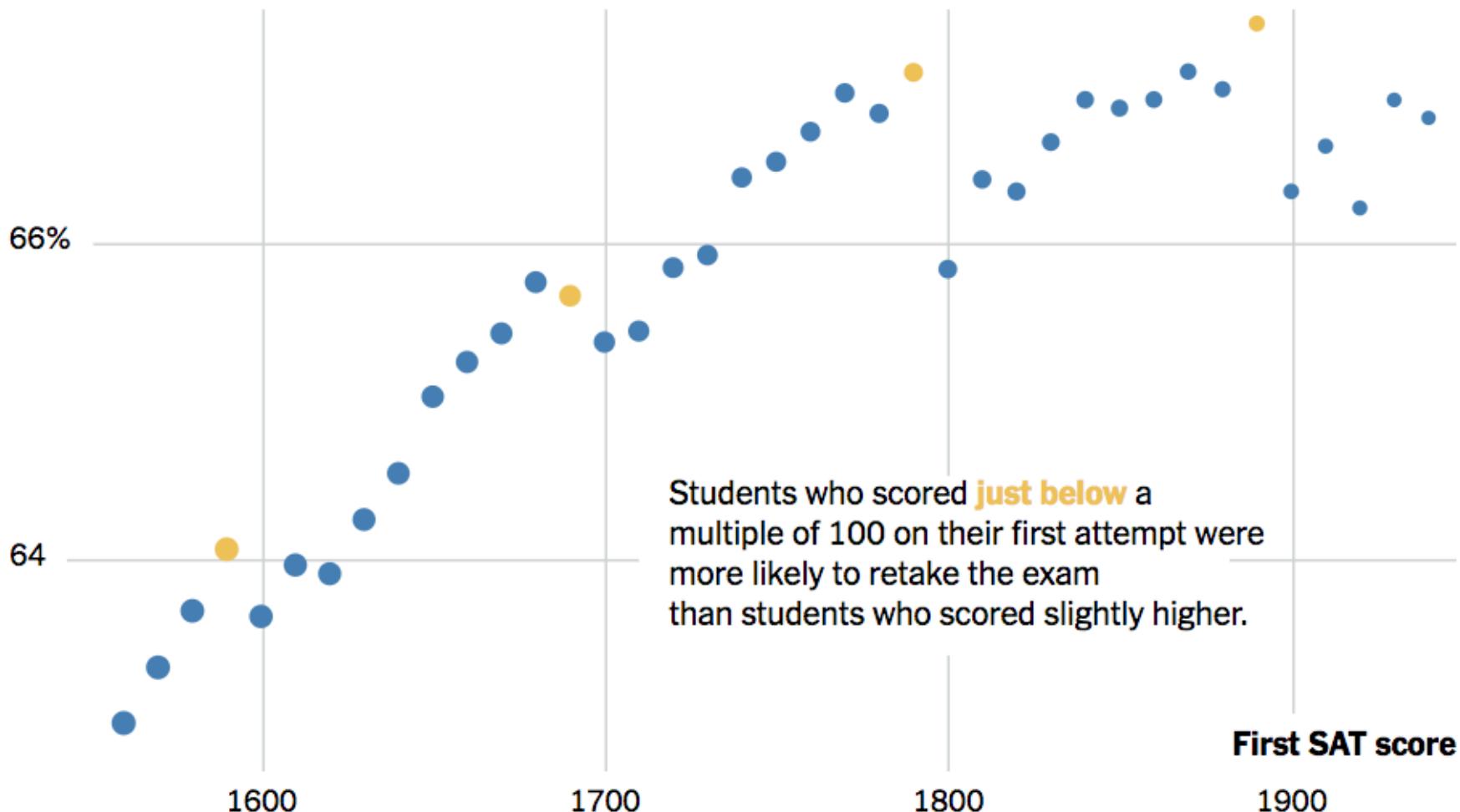
SHAPE**SIZE****SHARPNESS****NUMEROUSITY**





Left-Digit Bias in SAT Retaking

Percent who retook the SAT

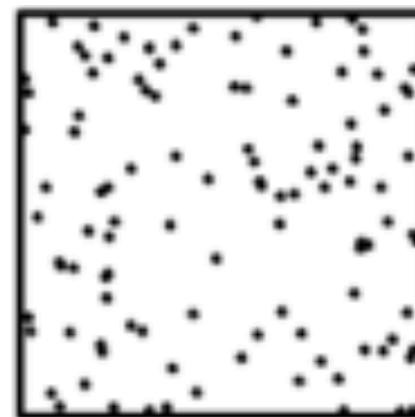


Note: Larger circles indicate more test-takers. Source: Joshua Goodman, Oded Gurantz and Jonathan Smith, "Take Two! SAT Retaking and College Enrollment Gaps." • By The New York Times

CONTEXTUAL AND RELATIVE



10



110



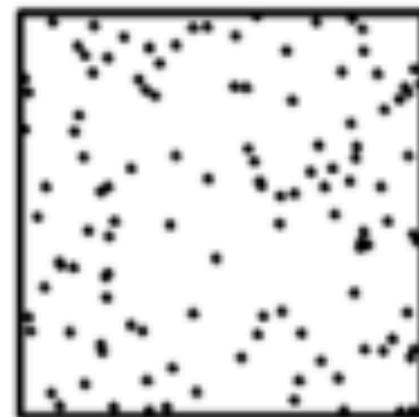
10



110



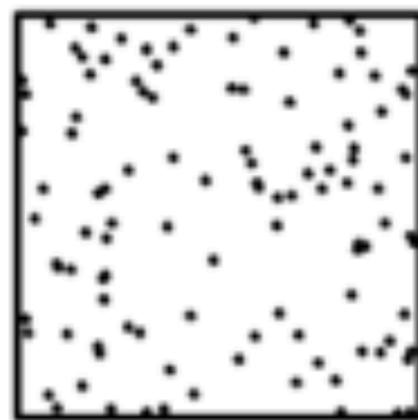
20



120



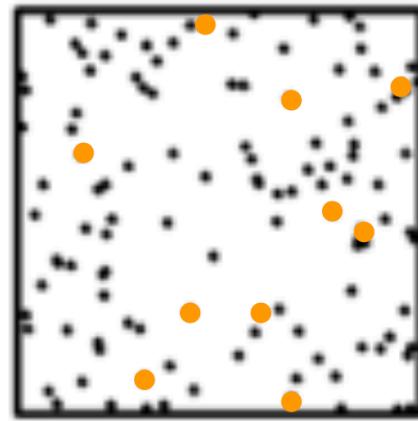
10



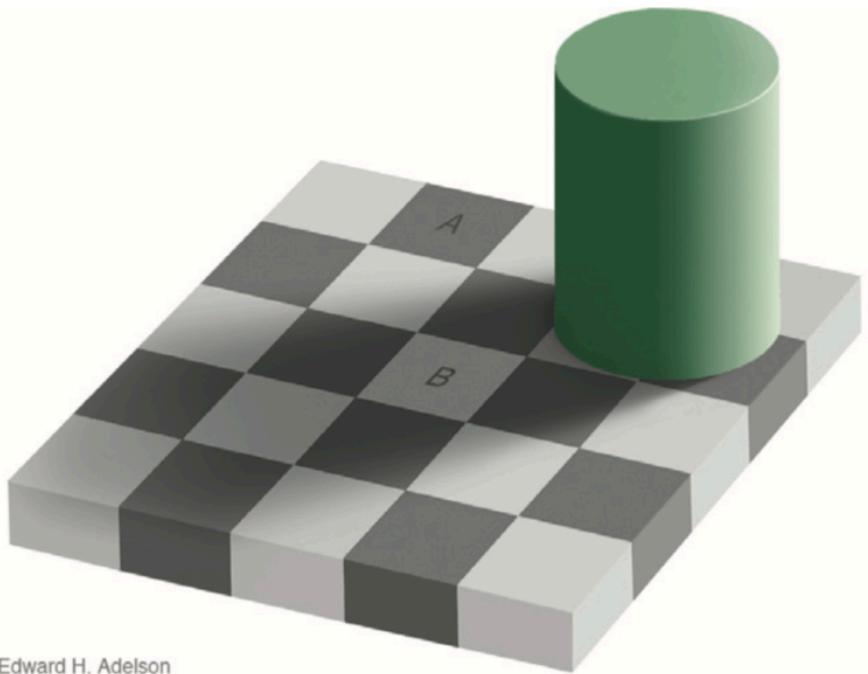
110



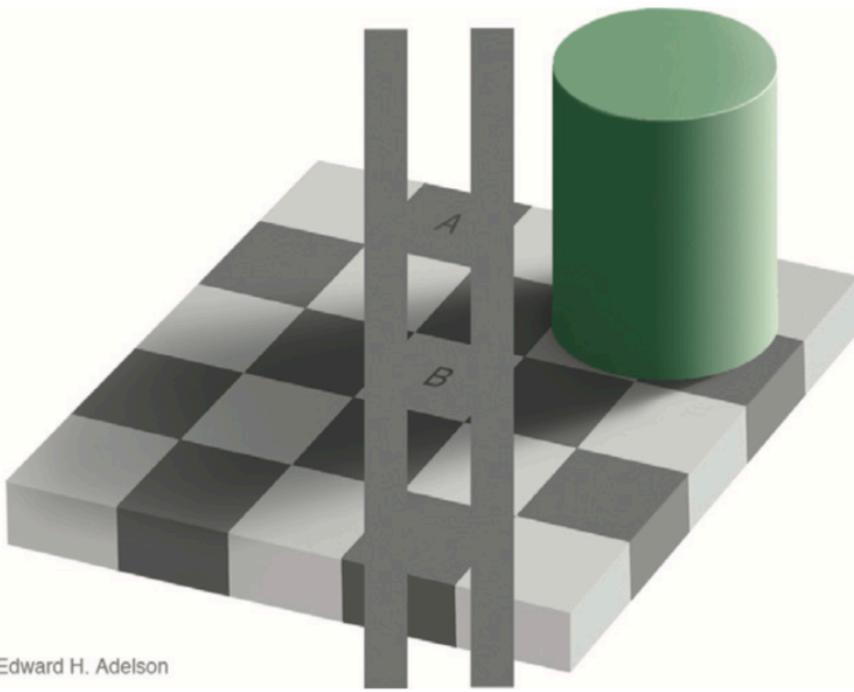
20



120

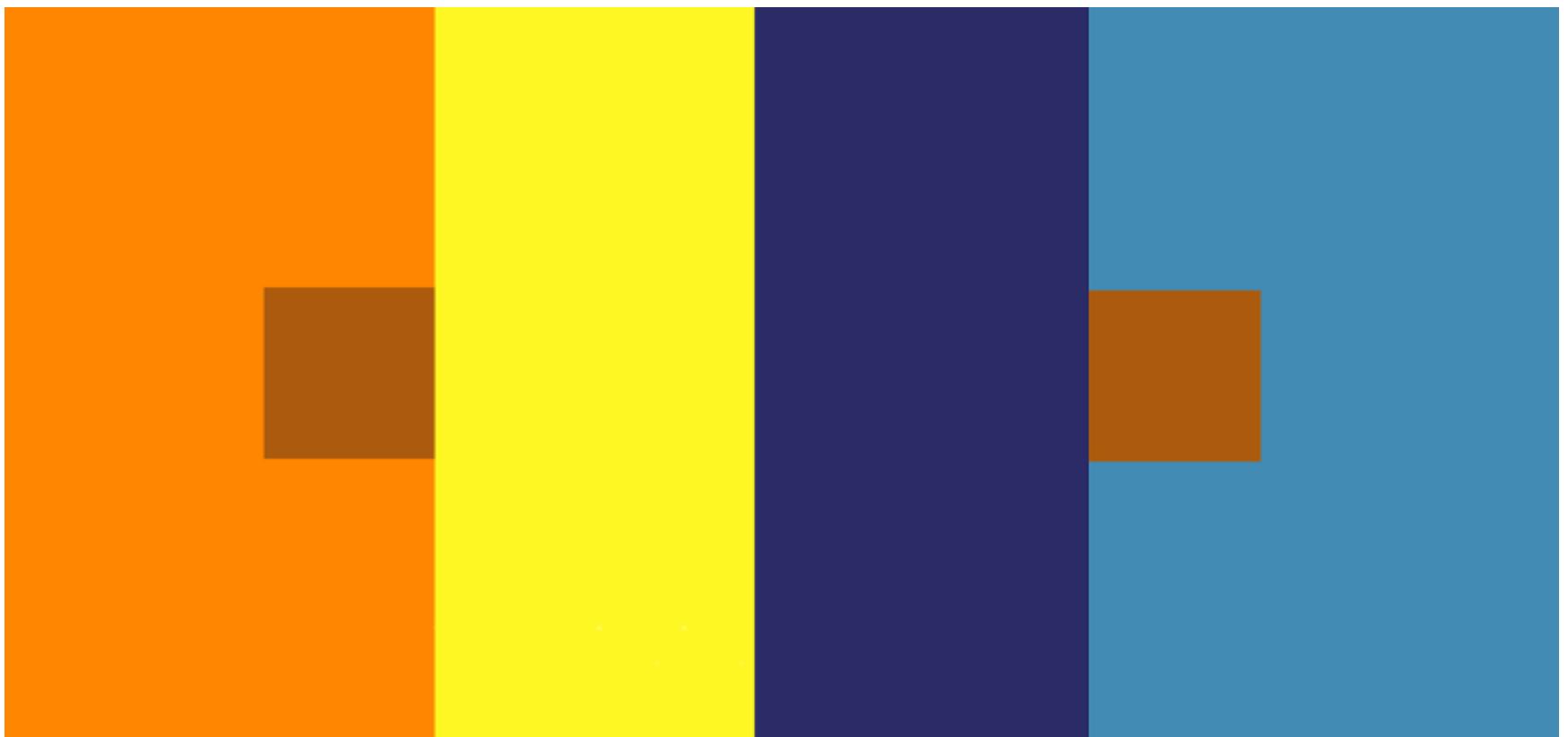


Edward H. Adelson



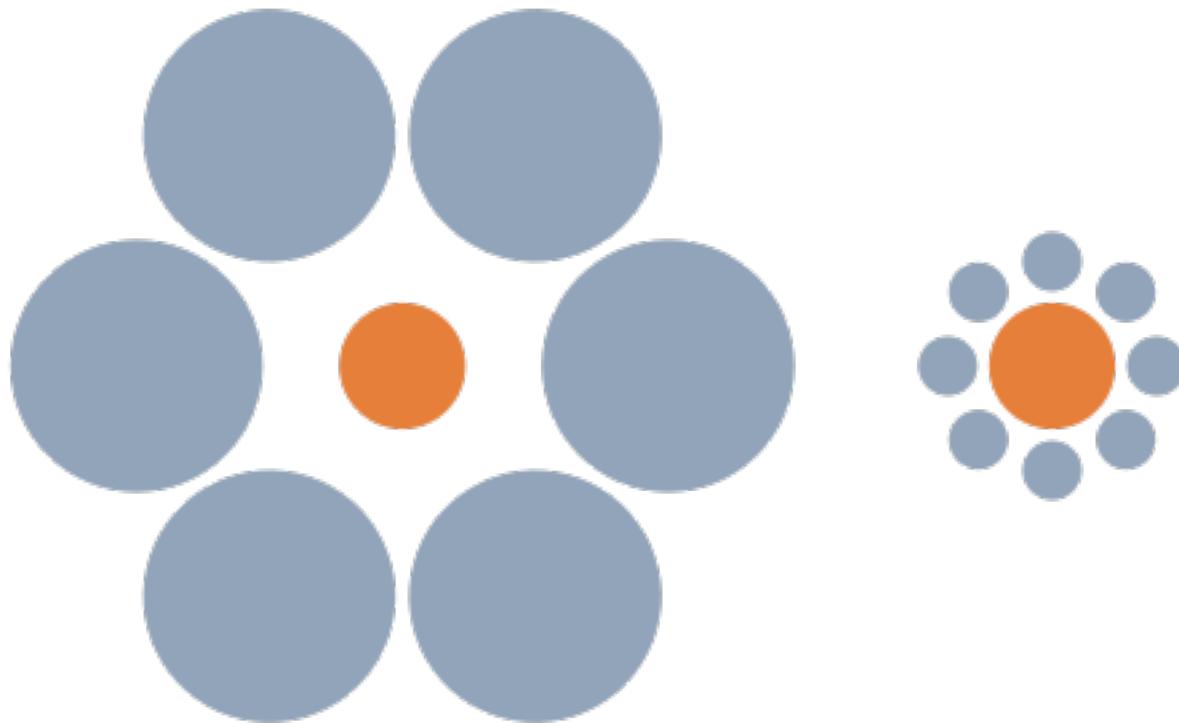
Edward H. Adelson

Via Tamara Munzner



“WAYS TO MAKE 3 COLORS LOOK LIKE 4”



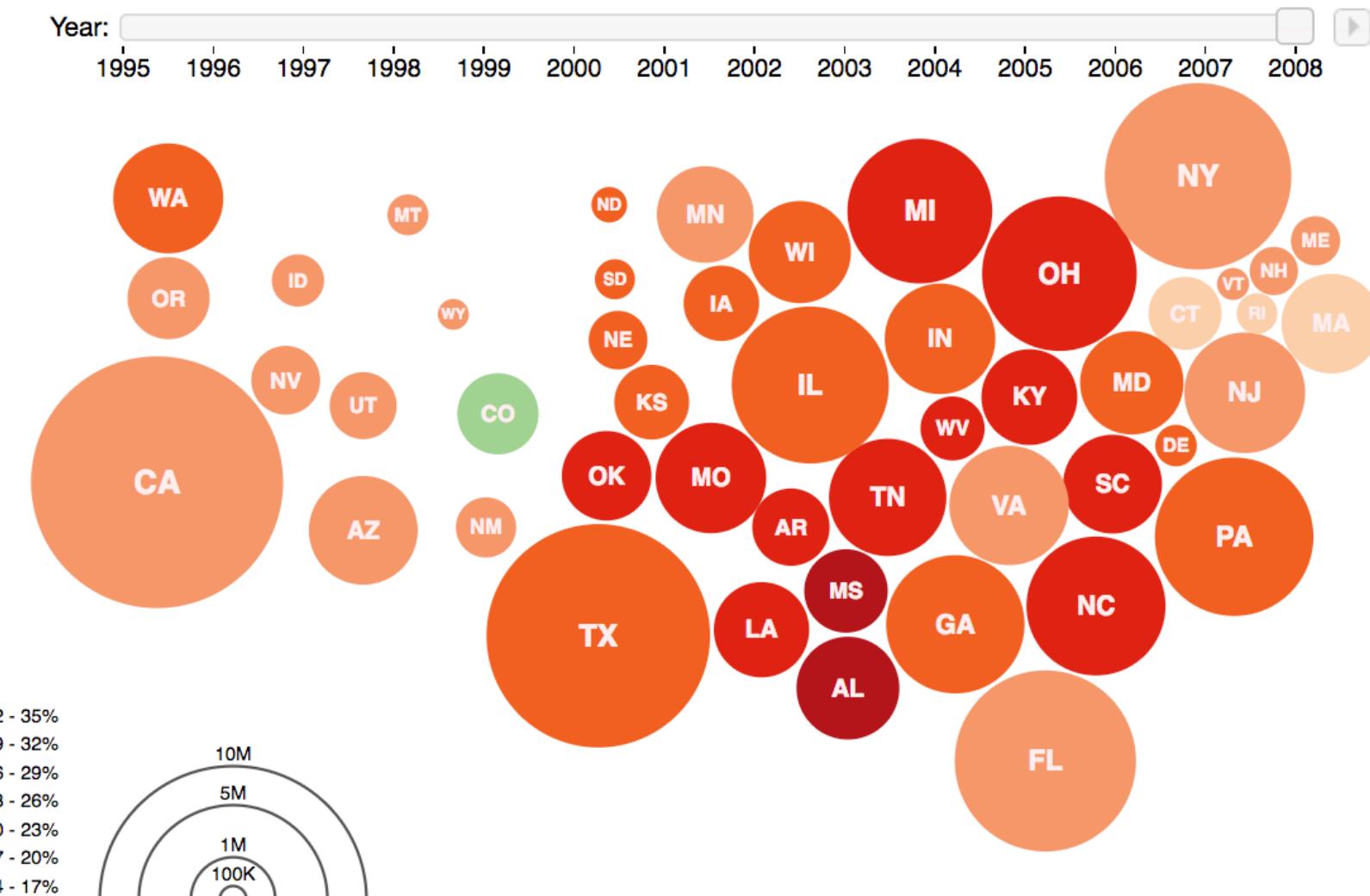


Ebbinghaus / titchener illusion



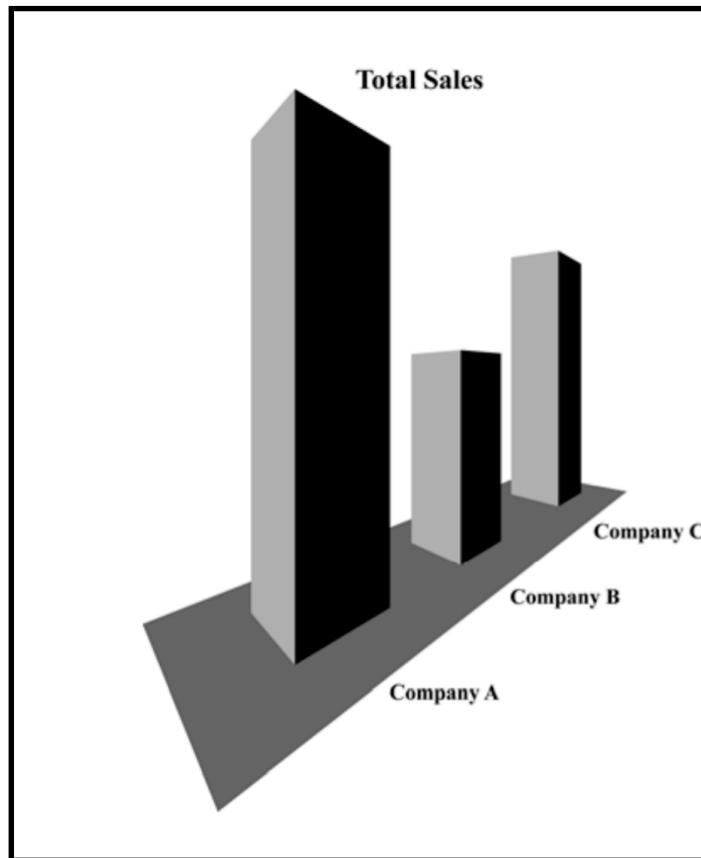
Ebbinghaus / titchener illusion

Dorling Cartogram of Obesity in the United States (1995 to 2008)

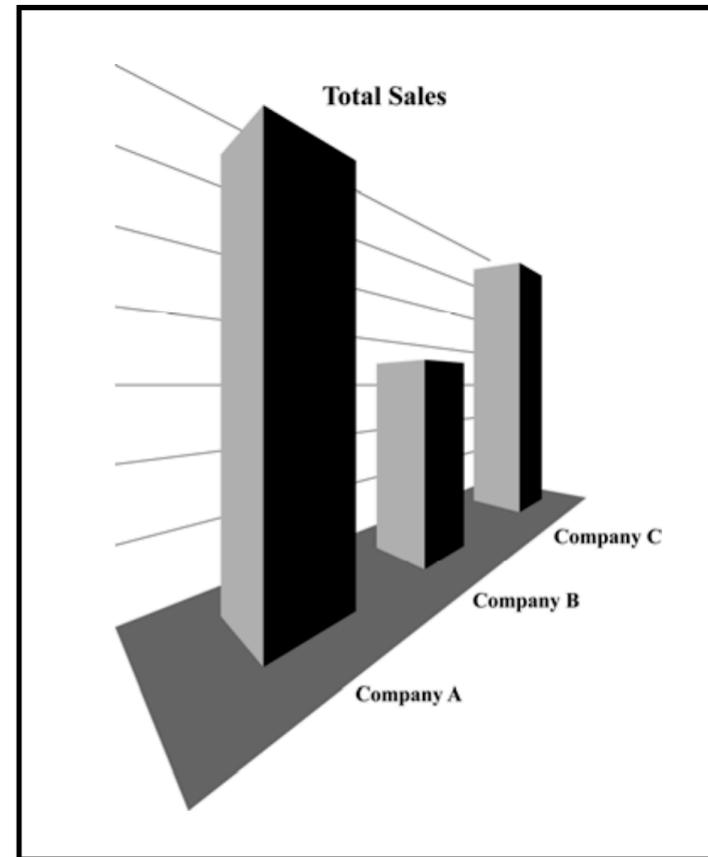
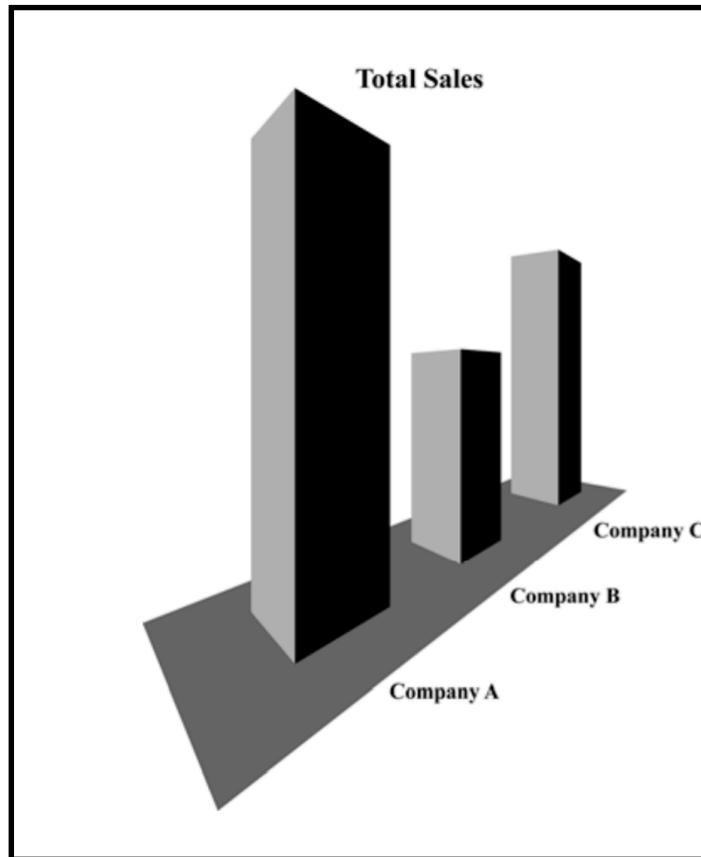


Absolute number of obese people per state and percentage of obese people, represented with circle area and color respectively. California dominates the map due to its large population size, while color indicates that Mississippi and Alabama are the states with the highest obesity rate.

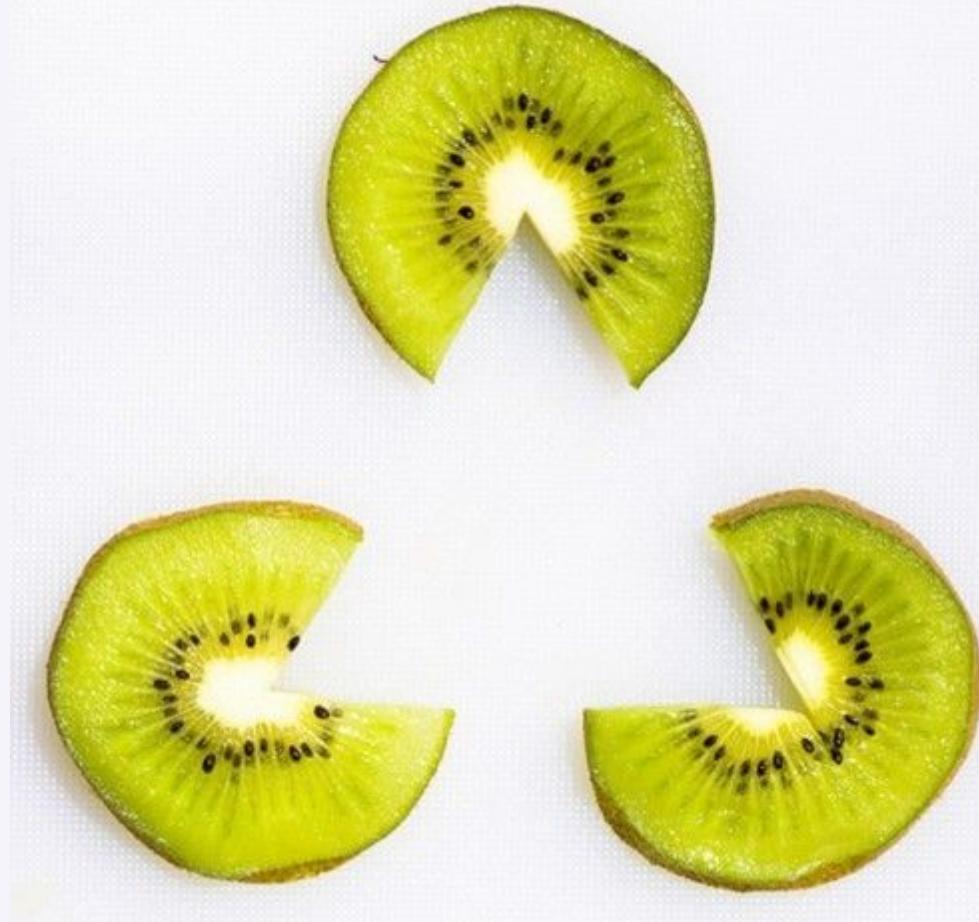
WHICH HAS MORE SALES?



3D PRETTY MUCH ALWAYS BAD IDEA



GESTALT PRINCIPLES



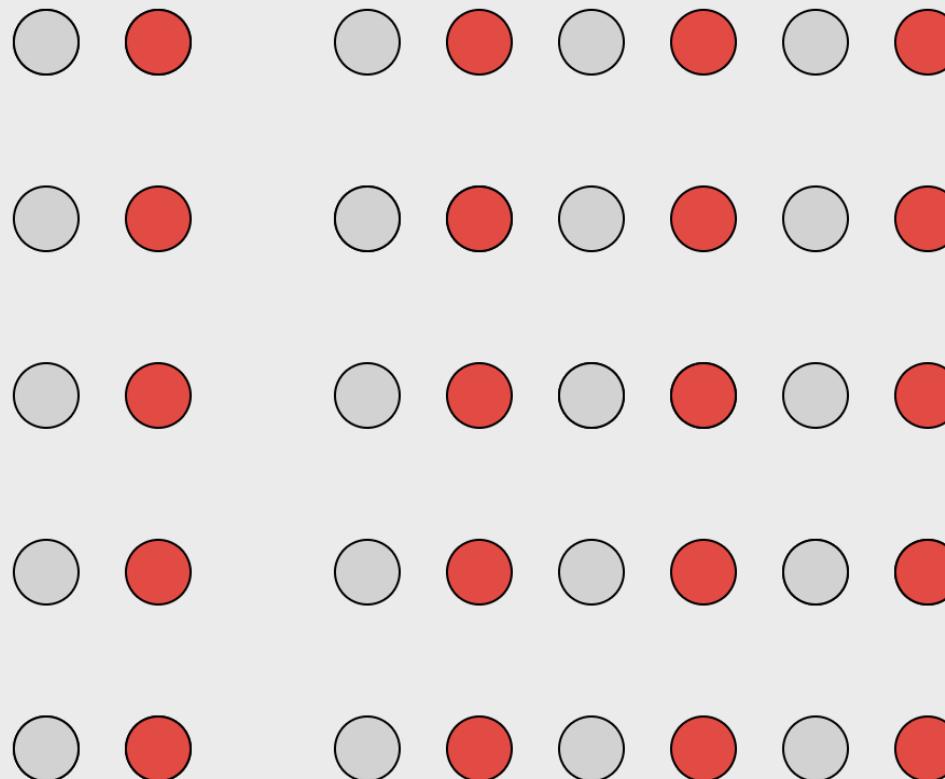
WE SEE PATTERNS AND CONNECTIONS



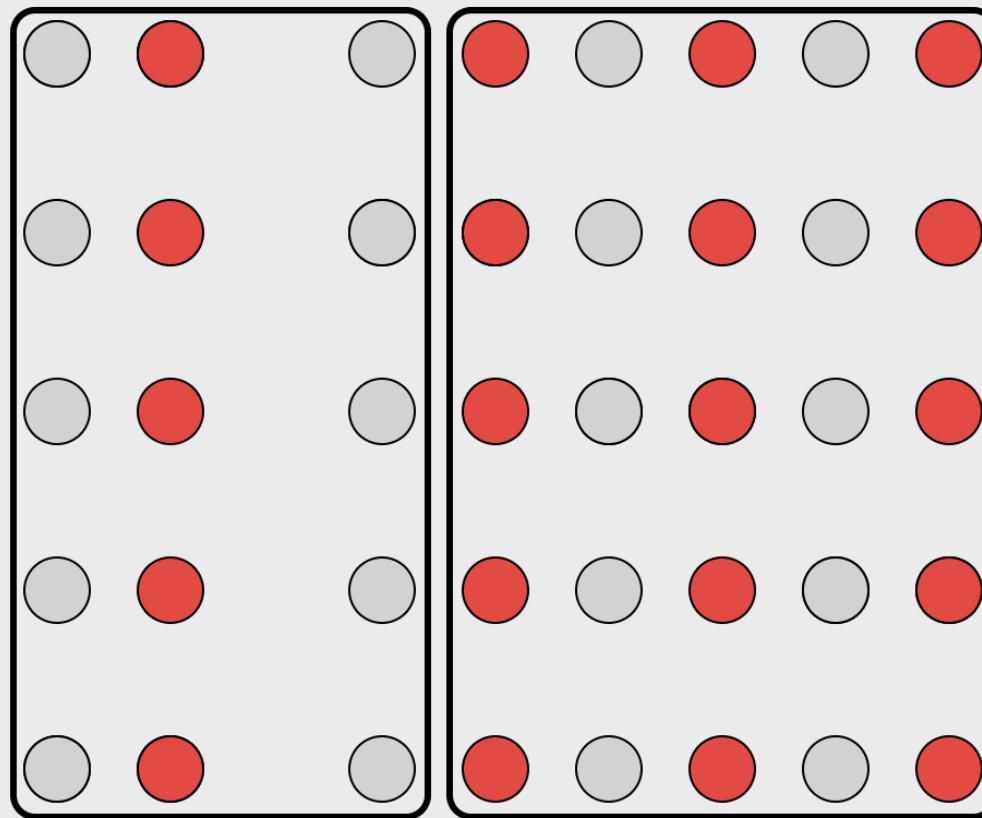
SIMILARITY



PROXIMITY

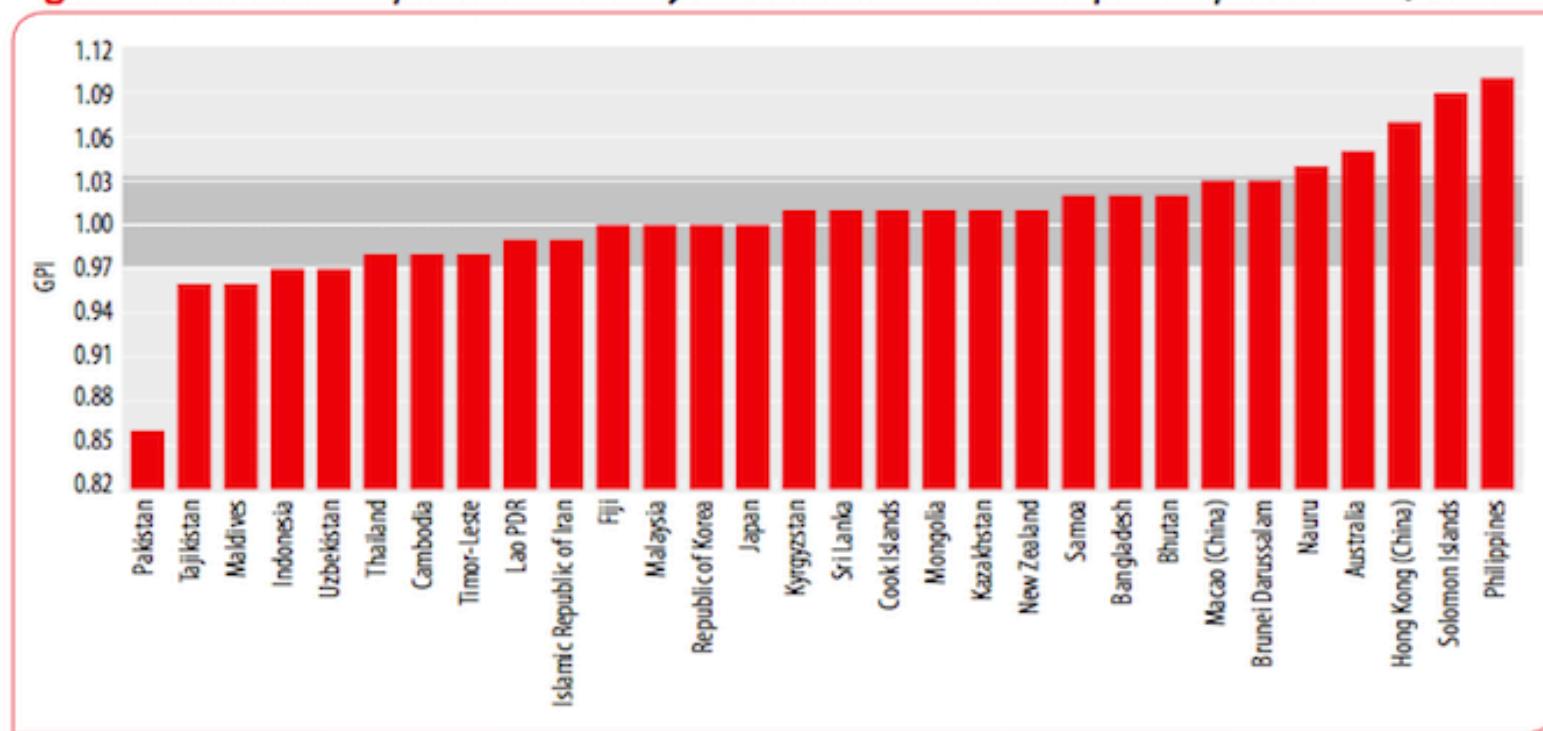


ENCLOSURE



PUTTING PERCEPTION TO PRACTICE...

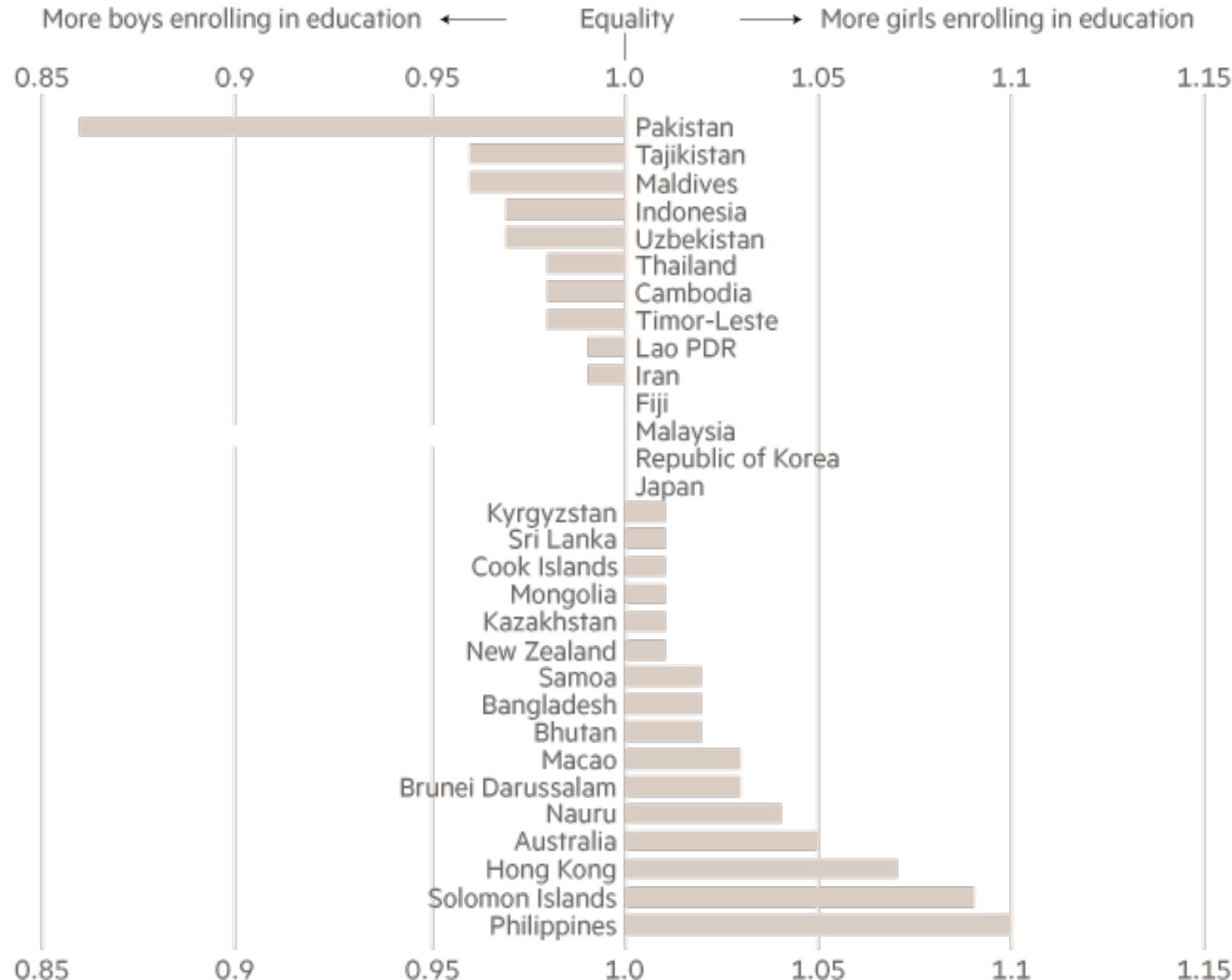
Figure 7: Gender Parity Index of the adjusted net intake rate in primary education, 2009



Source: UIS, 2011, Statistical Table 2.

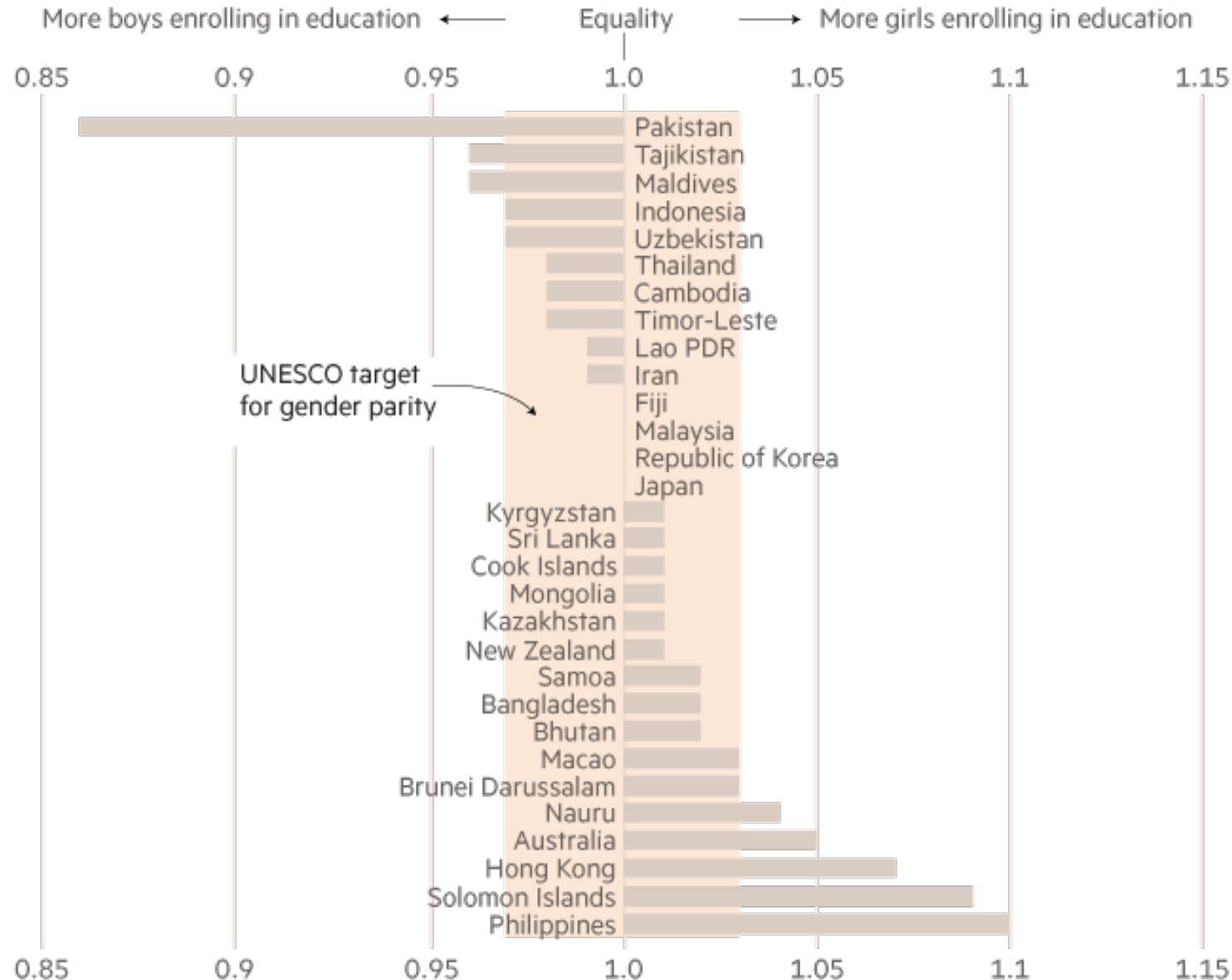
<https://www.ft.com/content/3b59f690-d129-11e7-b781-794ce08b24dc>

Gender parity index*, 2009, Asia-Pacific region



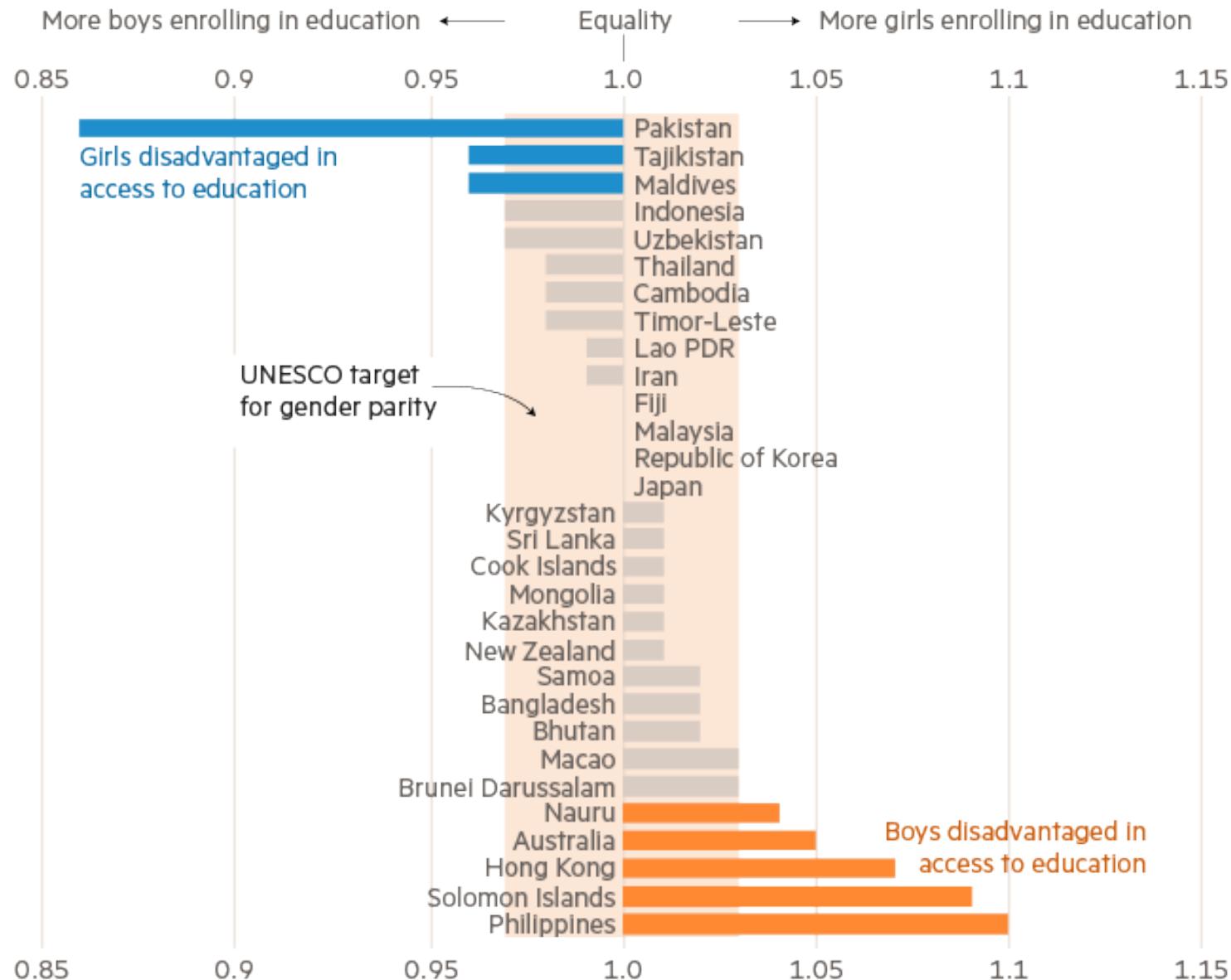
* Of adjusted net intake rate Source: UIS, 2011, Statistical table 2 Graphic by Alan Smith

Gender parity index*, 2009, Asia-Pacific region



* Of adjusted net intake rate Source: UIS, 2011, Statistical table 2 Graphic by Alan Smith

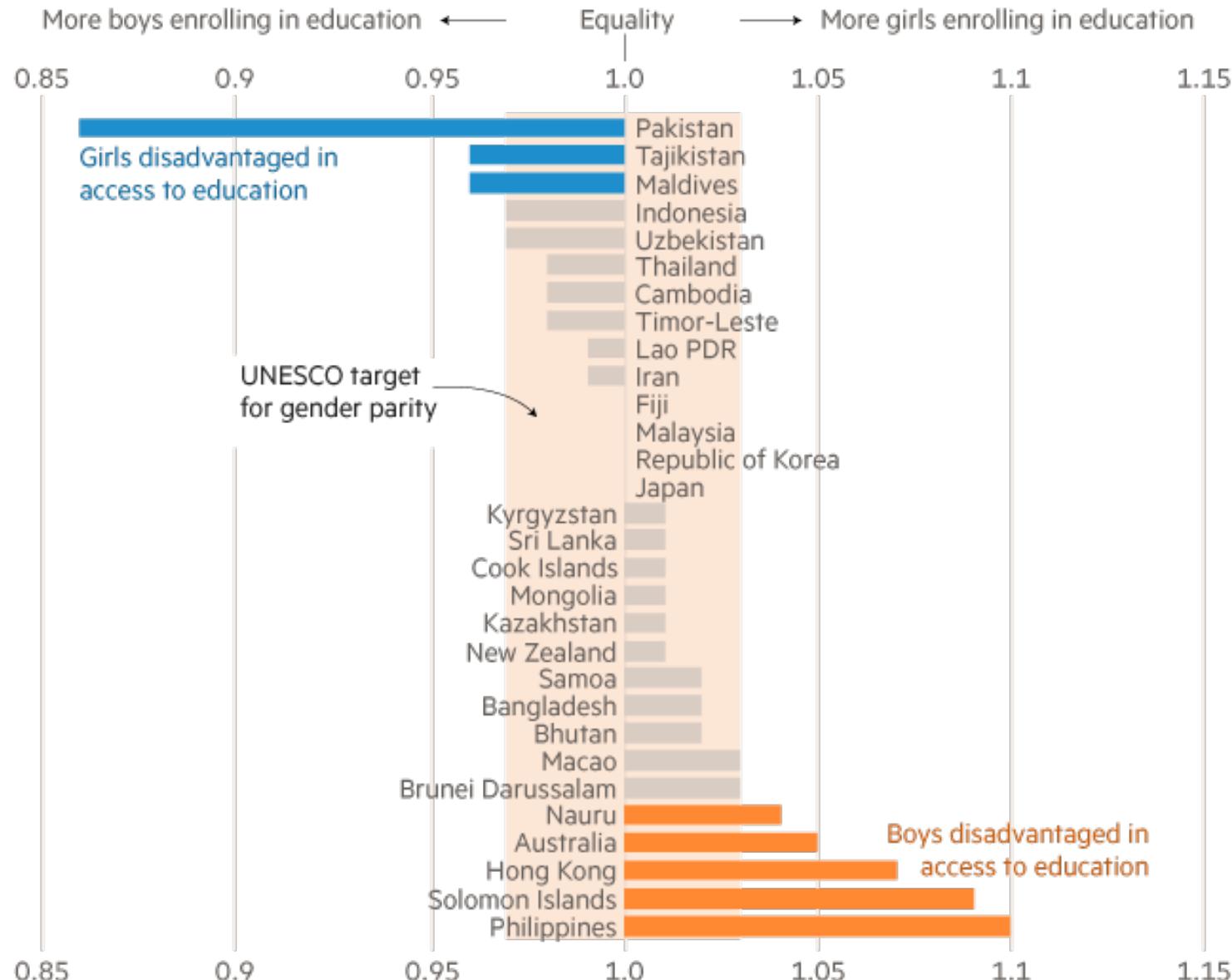
Gender parity Index*, 2009, Asia-Pacific region



* Of adjusted net intake rate Source: UIS, 2011, Statistical table 2 Graphic by Alan Smith

Gender equality in primary education is not yet a reality for all

Gender parity index*, 2009, Asia-Pacific region



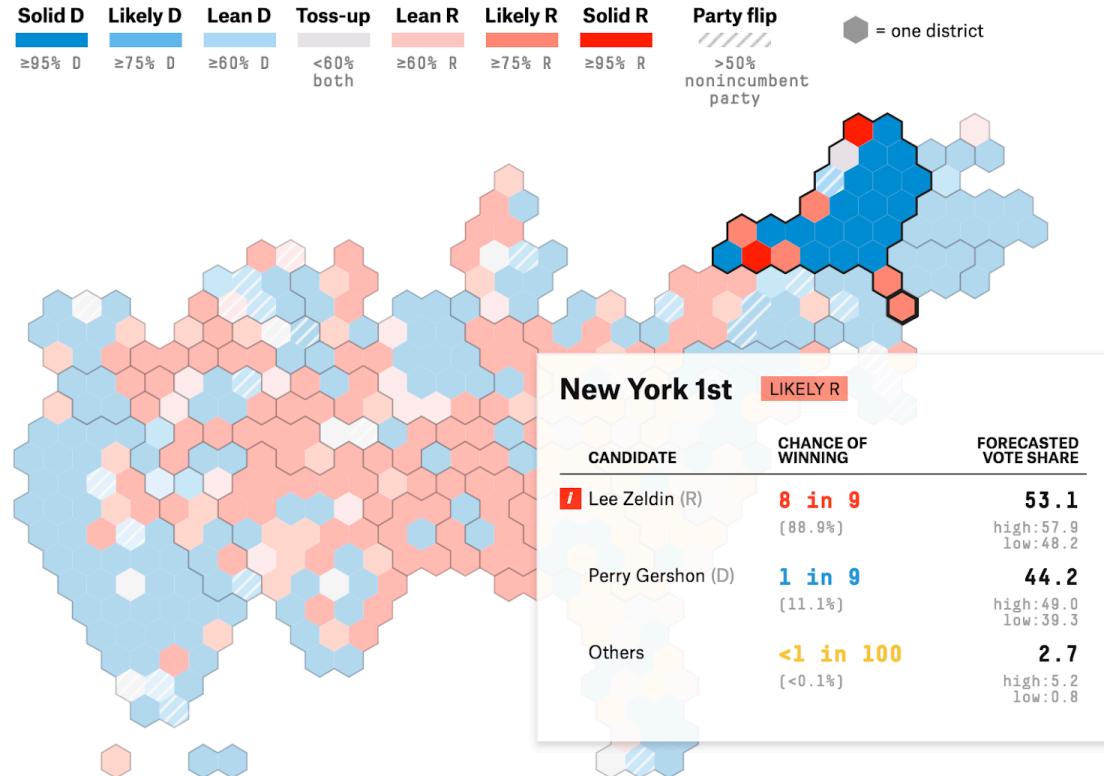
* Of adjusted net intake rate Source: UIS, 2011, Statistical table 2 Graphic by Alan Smith

CARTOGRAMS

Our forecast for every district

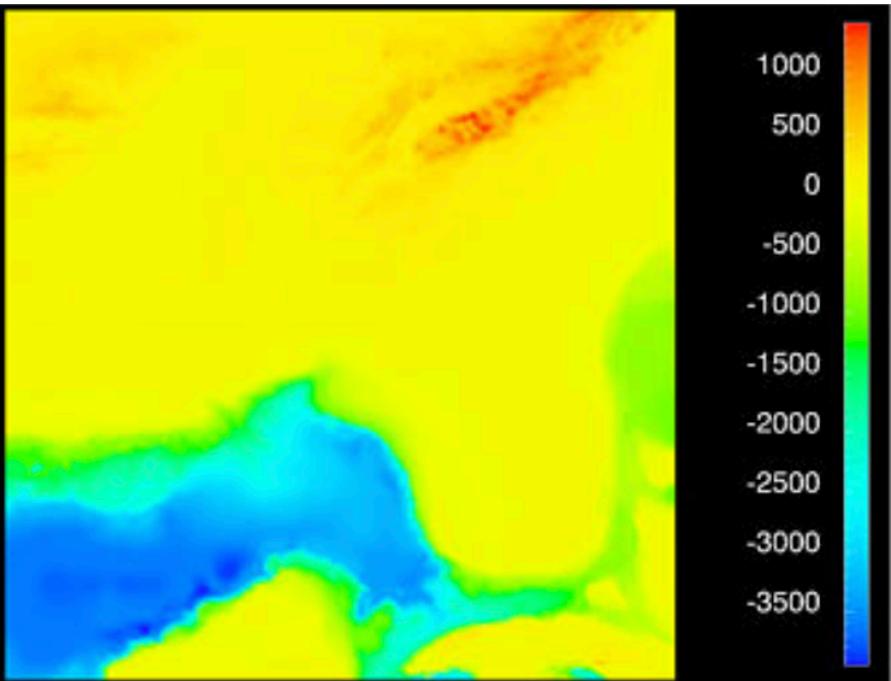
Cartogram Map

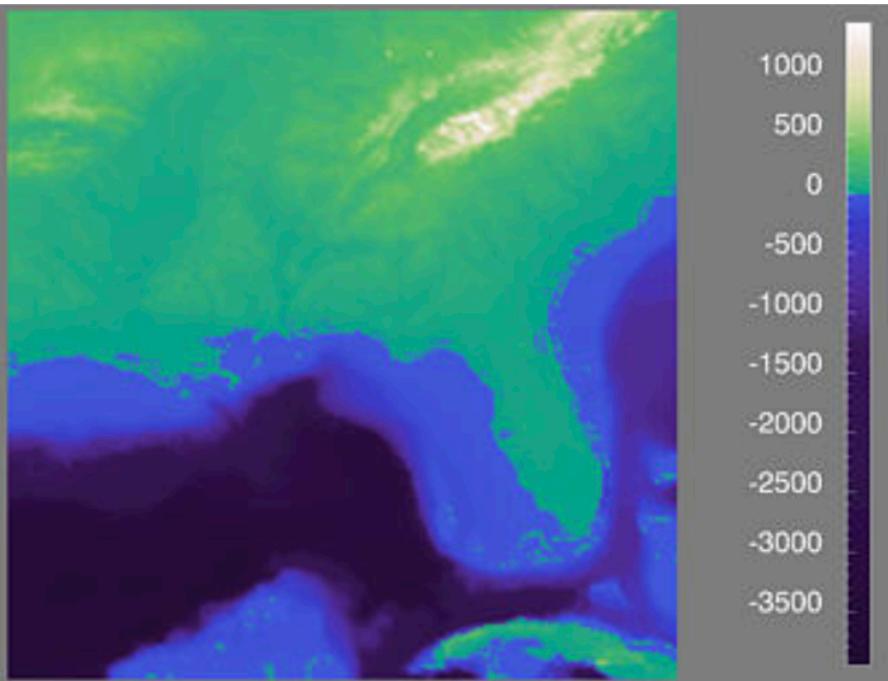
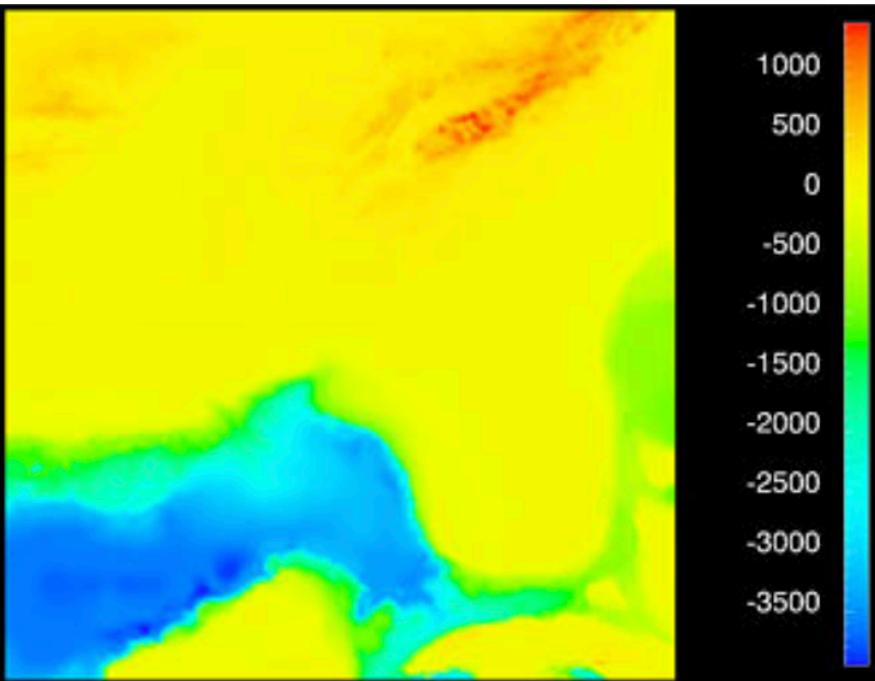
The chance of each candidate winning, with all 435 House districts shown at the same size



COLOR SCALES







Rogowitz and Treinish, 1996.

COLOR SCHEMES



- Sequential
 - ordered
 - e.g. population density: 1.1~6.9, 7.0~79.5, 79.6~299.9



- Diverging
 - ordered, “break” in middle of range
 - e.g. median income: below avg., national avg., above avg.



- Categorical
 - a.k.a. Qualitative
 - equidistant, no order
 - e.g. apples, oranges, pineapples, banana

viridis



magma



plasma



inferno



cividis



Number of data classes: 3

[how to use](#) | [updates](#) | [downloads](#) | [credits](#)

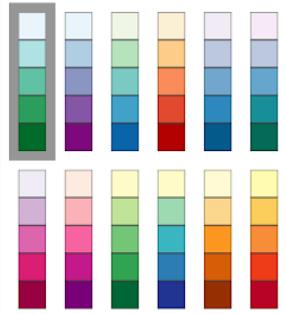
COLORBREWER 2.0
color advice for cartography

Nature of your data:

sequential diverging qualitative

Pick a color scheme:

Multi-hue:



Single hue:

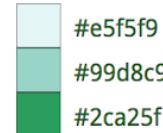


Only show:

3-class BuGn



HEX



Context:

roads

cities

borders



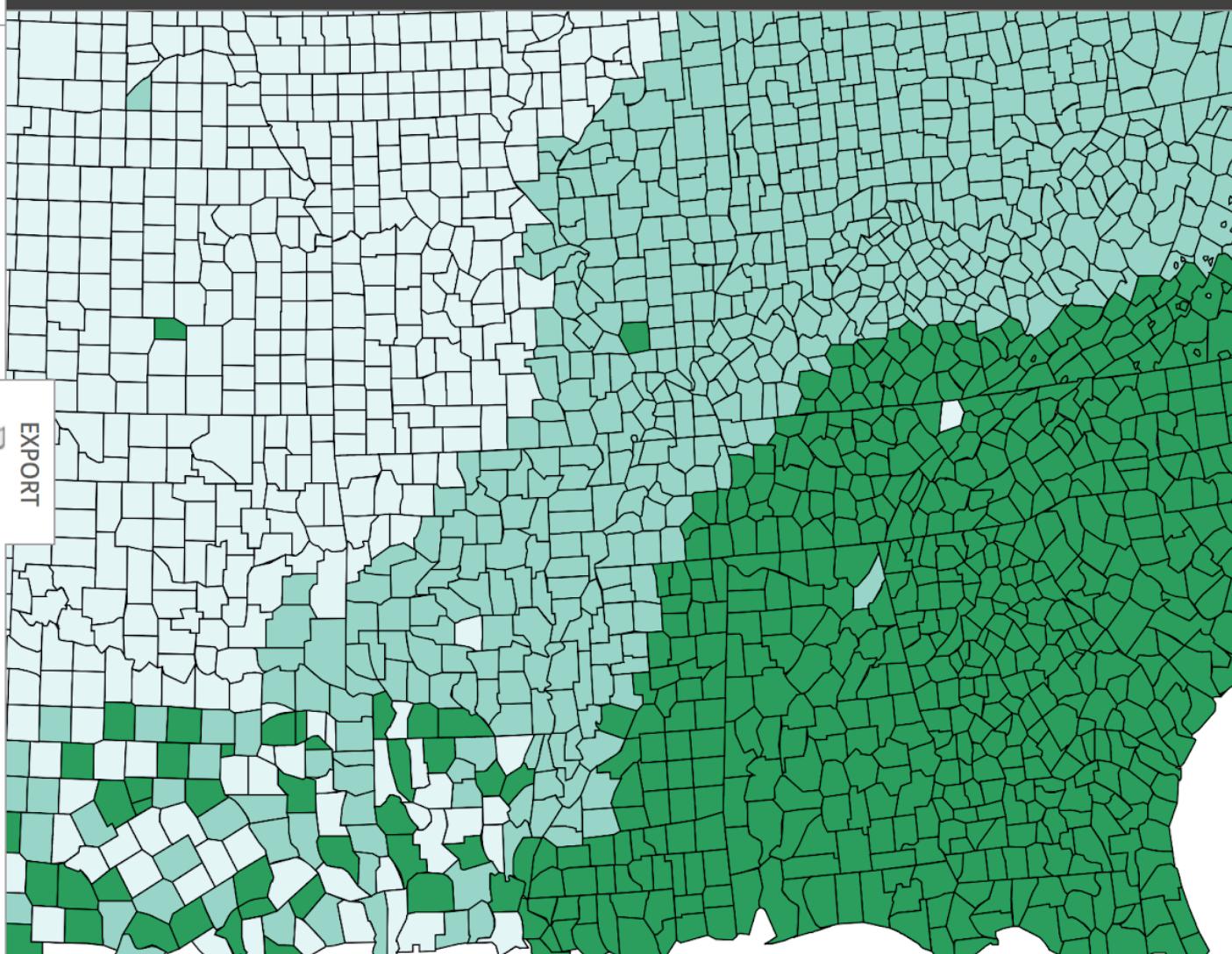
Background:

solid color

terrain



color transparency



STORYTELLING



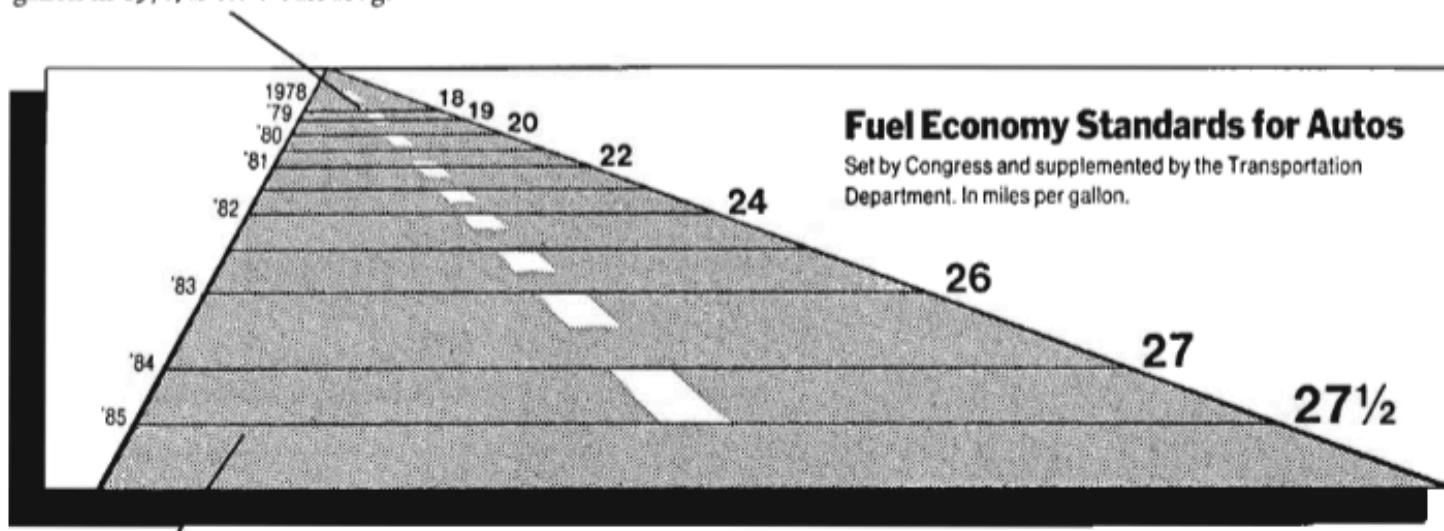
TUFT LIE FACTOR

The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.

Violations of the first principle constitute one form of graphic misrepresentation, measured by the

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

TUFTÉ DATA-INK RATIO

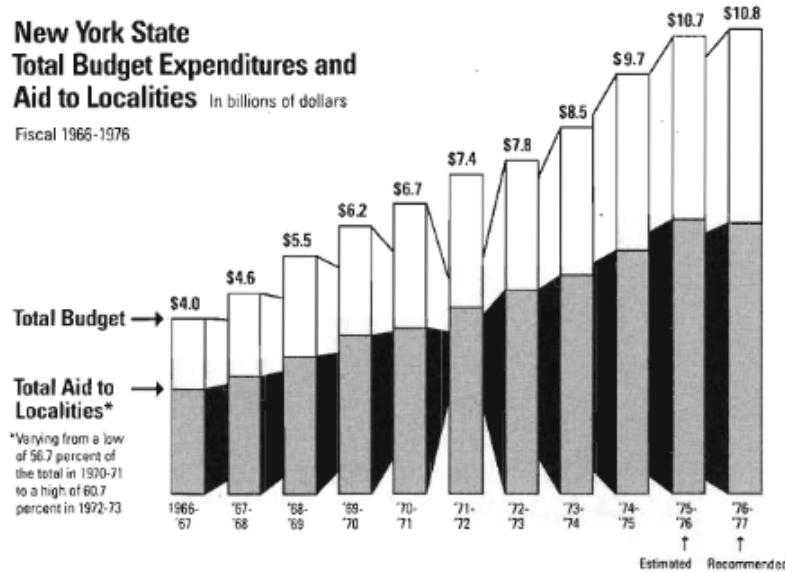
Data-ink ratio = $\frac{\text{data-ink}}{\text{total ink used to print the graphic}}$

= proportion of a graphic's ink devoted to the non-redundant display of data-information

= $1.0 - \text{proportion of a graphic that can be erased without loss of data-information.}$

New York State
Total Budget Expenditures and
Aid to Localities In billions of dollars

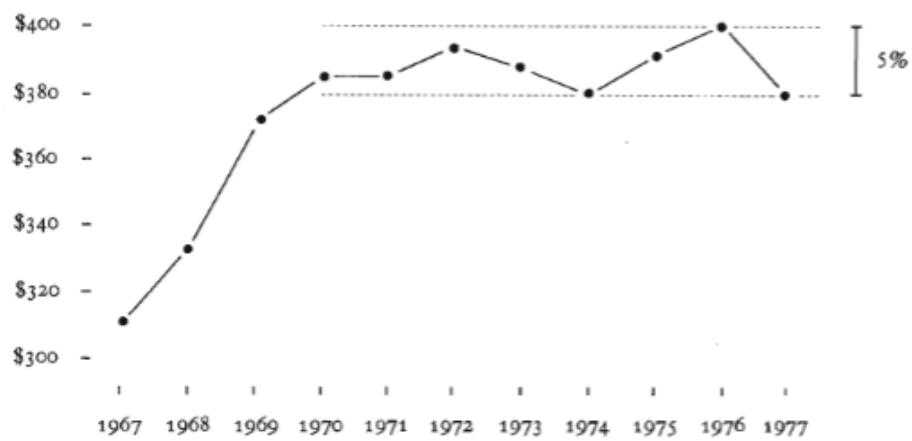
Fiscal 1966-1976



*Varying from a low of 56.7 percent of the total in 1970-71 to a high of 60.7 percent in 1976-77

Per capita

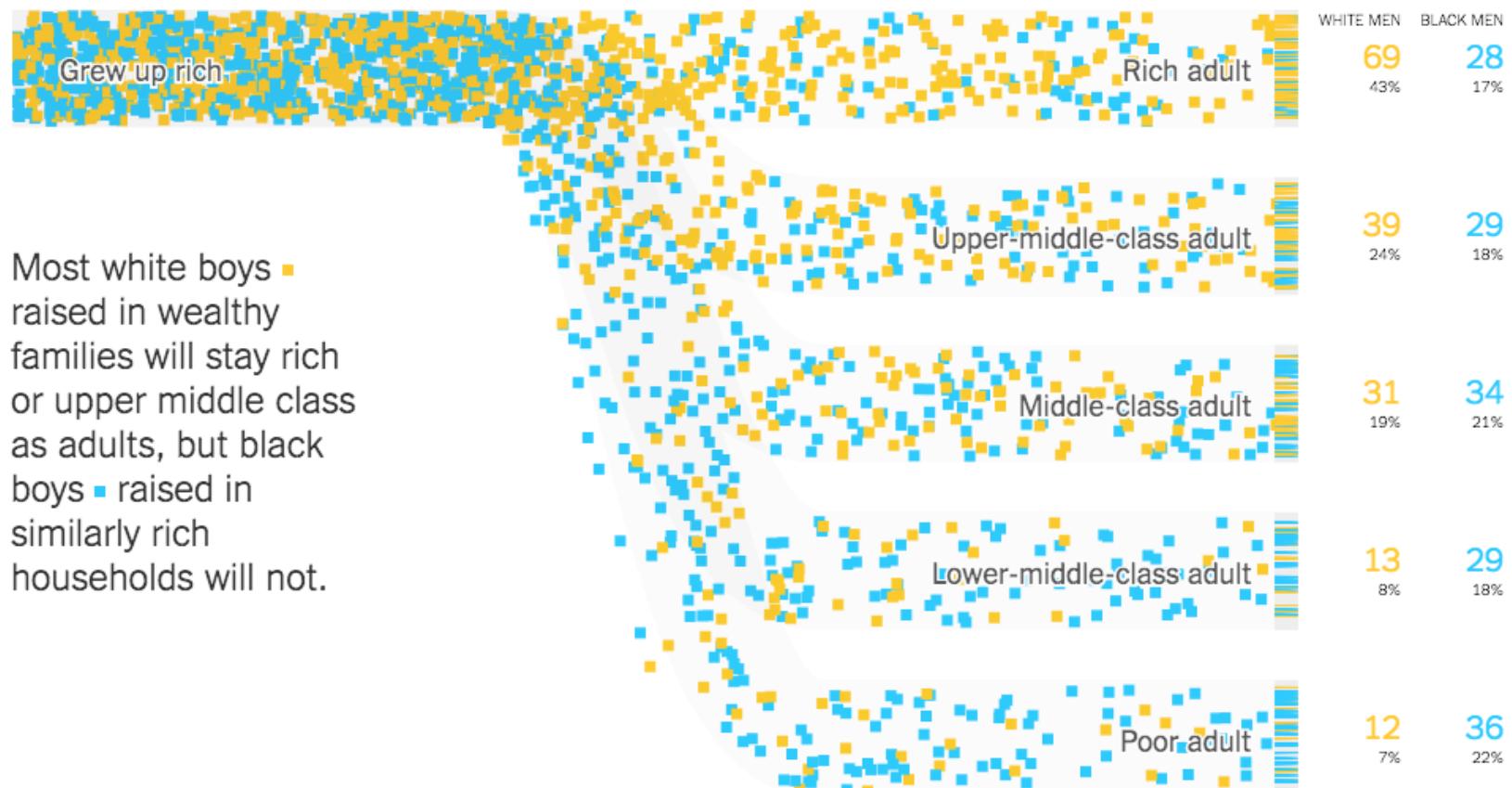
budget expenditures,
in constant dollars



NYT CASE STUDY

Follow the lives of 2,592 boys who grew up in rich families ...

...and see where they end up as adults:



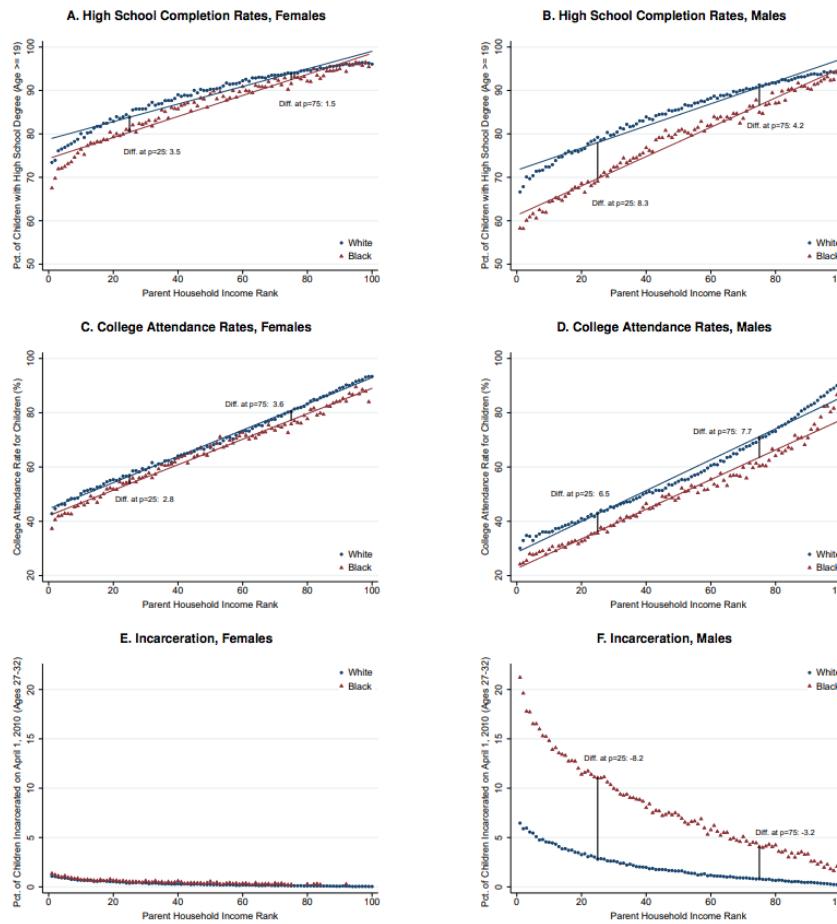
Most white boys ■ raised in wealthy families will stay rich or upper middle class as adults, but black boys ■ raised in similarly rich households will not.

Adult outcomes reflect household incomes in 2014 and 2015.

<https://www.nytimes.com/interactive/2018/03/19/upshot/race-class-white-and-black-men.html>

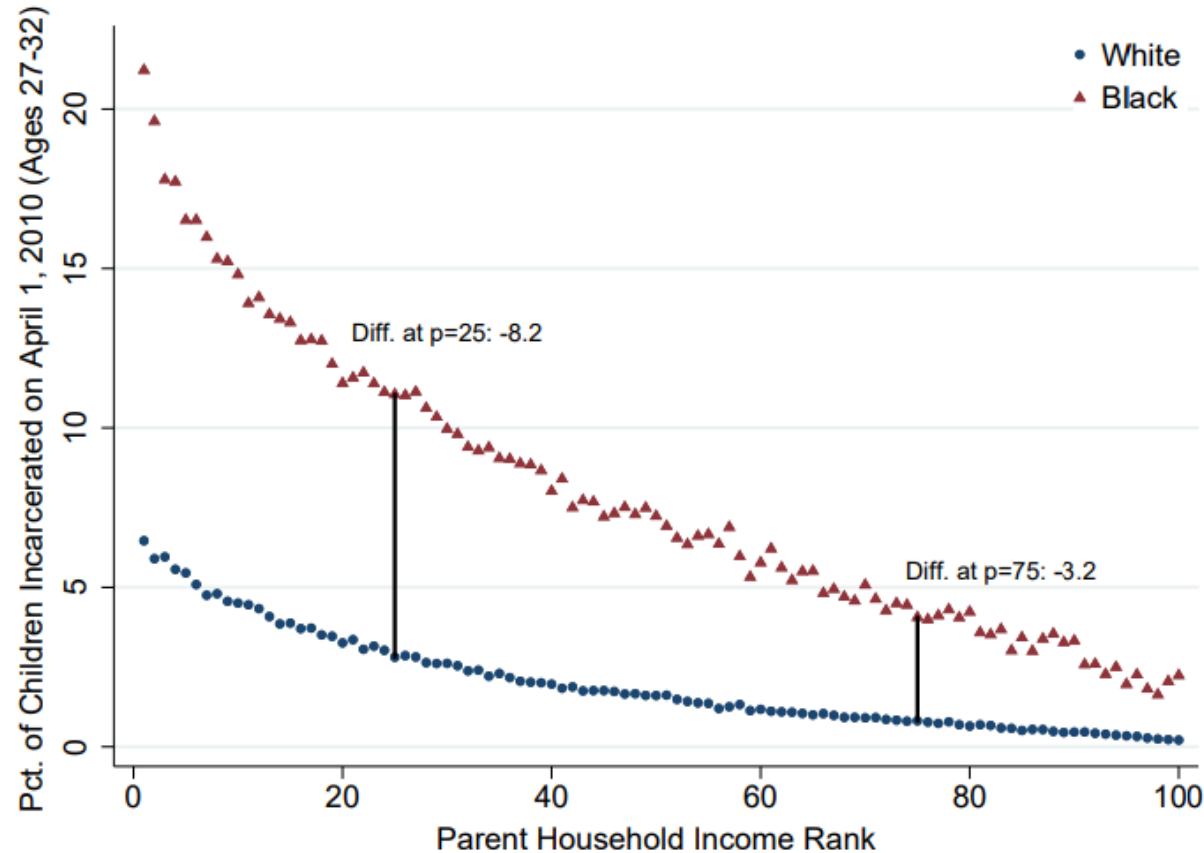
EXPLORATORY CHARTS...

FIGURE VII: Black-White Gaps in Educational Attainment and Incarceration, by Gender



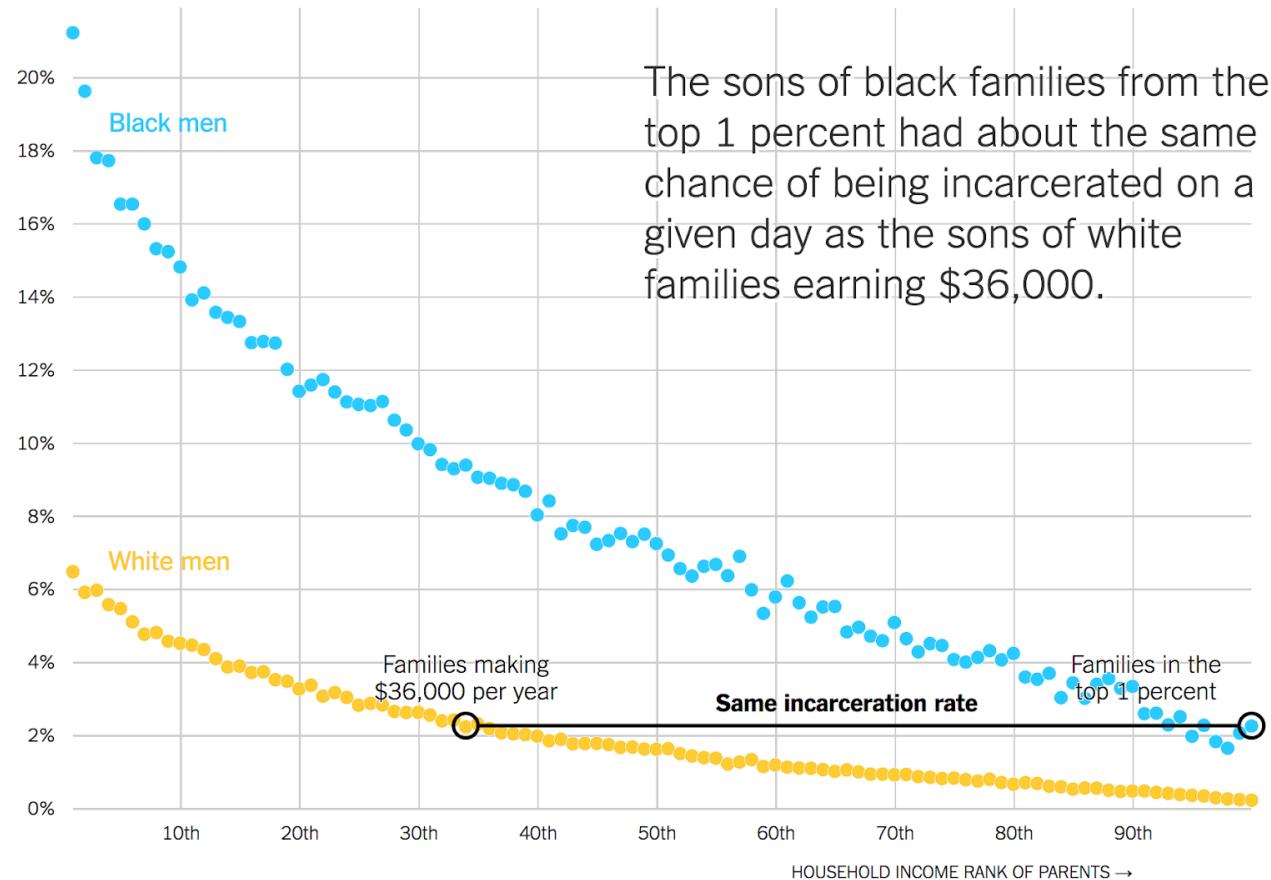
VERSUS...

F. Incarceration, Males



EXPLANATORY

Share of the men incarcerated on April 1, 2010

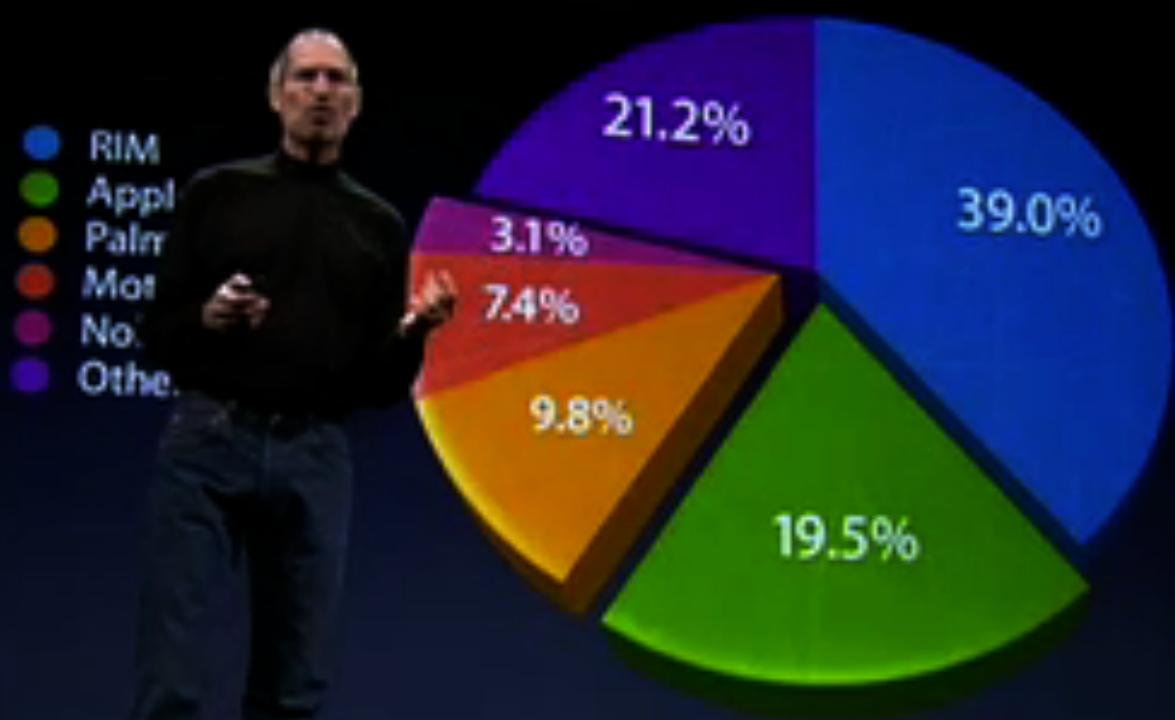


Includes men who were ages 27 to 32 in 2010.

MISREPRESENTATION

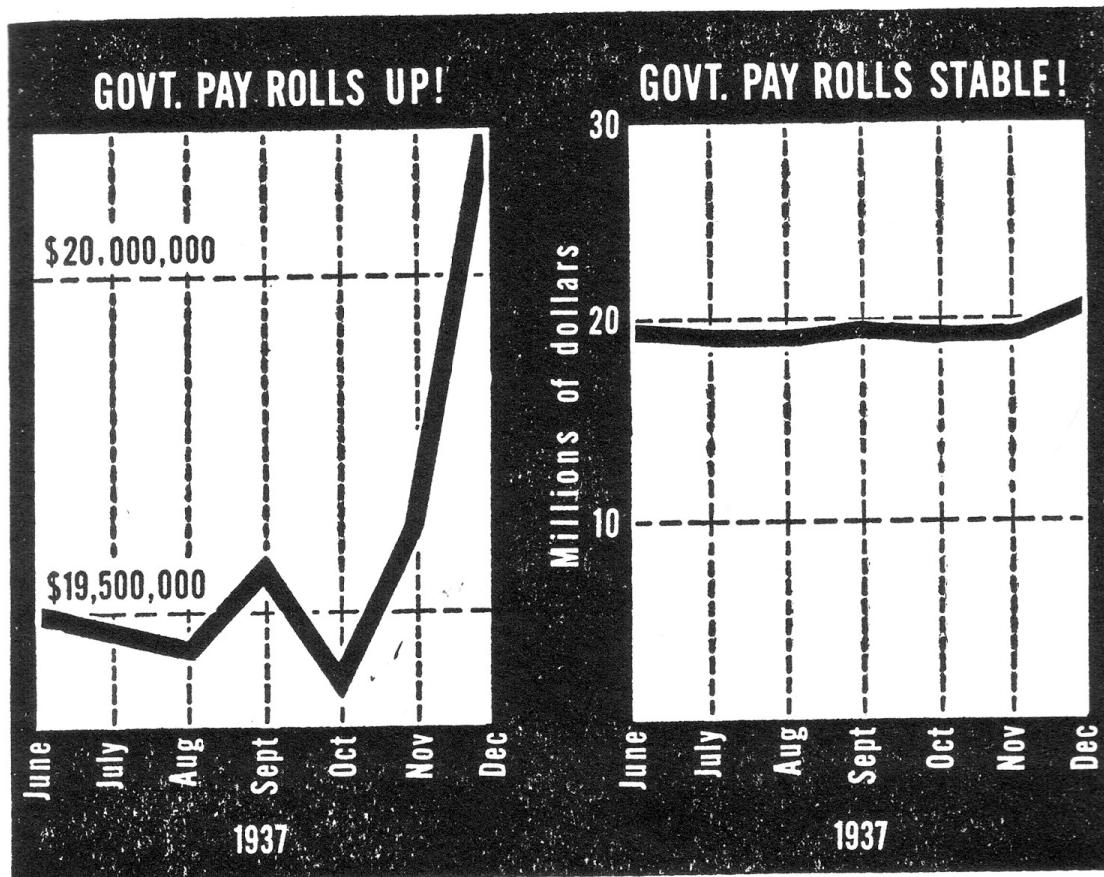


U.S. SmartPhone Marketshare



Source: Gart for 3

COMMON LIES: TRUNCATED AXES



OBAMACARE ENROLLMENT

6,000,000

**AS OF
MARCH 27**

7,066,000

**MARCH 31
GOAL**

SOURCE: HHS

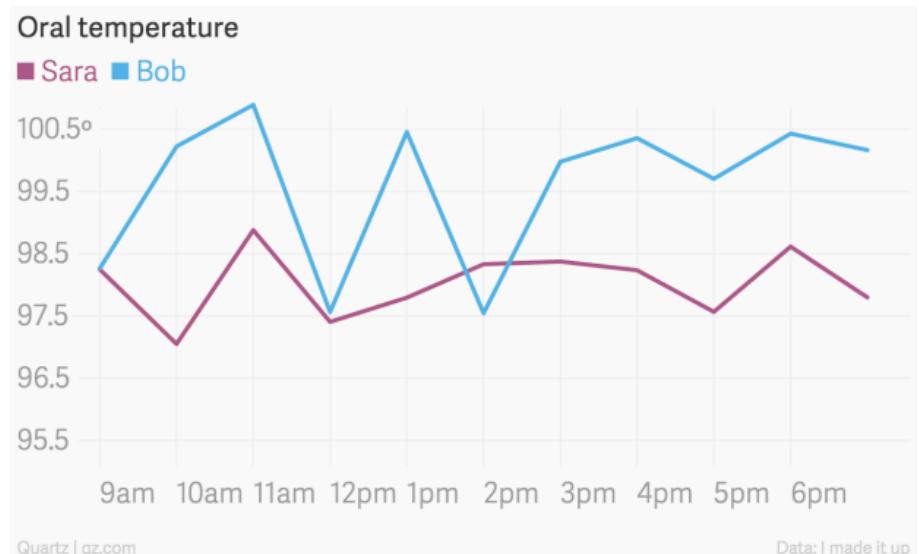
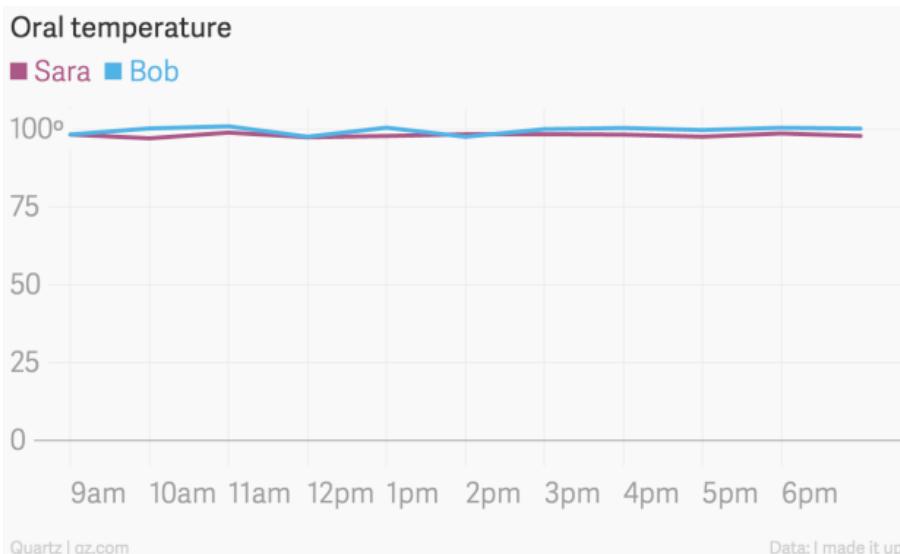


mediamatters.org

E IN SECOND PLACE WITH \$26.5 MIL, WHILE "MUPP DOW FUT 16,325.00

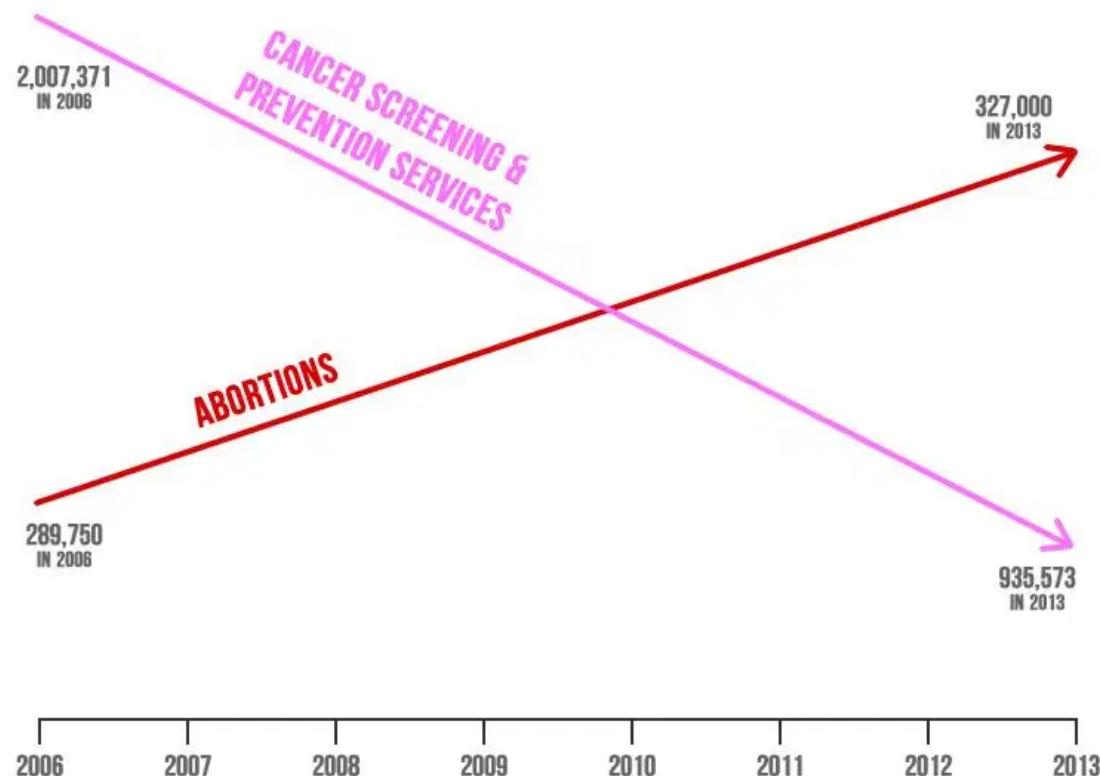
IT'S OK WHEN...

- NOT bar chart
- Small movements are important
- Zero values are ridiculous



<https://qz.com/418083/its-ok-not-to-start-your-y-axis-at-zero/>

PLANNED PARENTHOOD FEDERATION OF AMERICA: ABORTIONS UP – LIFE-SAVING PROCEDURES DOWN



SOURCE: AMERICANS UNITED FOR LIFE

Planned Parenthood services

■ Abortion procedures ■ Cancer screening / preventative services

2.5 million

2.0

1.5

1.0

0.5

0

'06

'07

'08

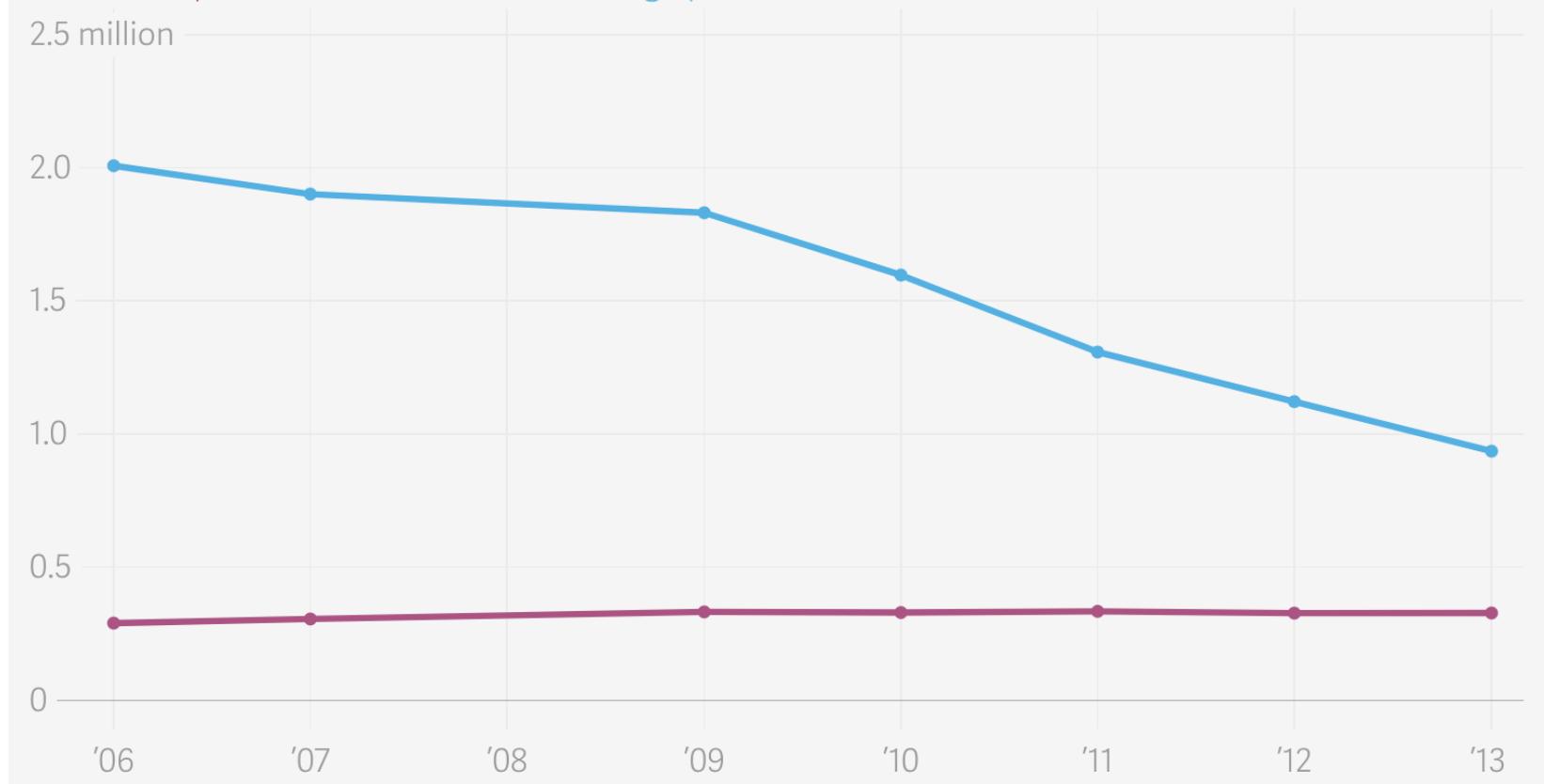
'09

'10

'11

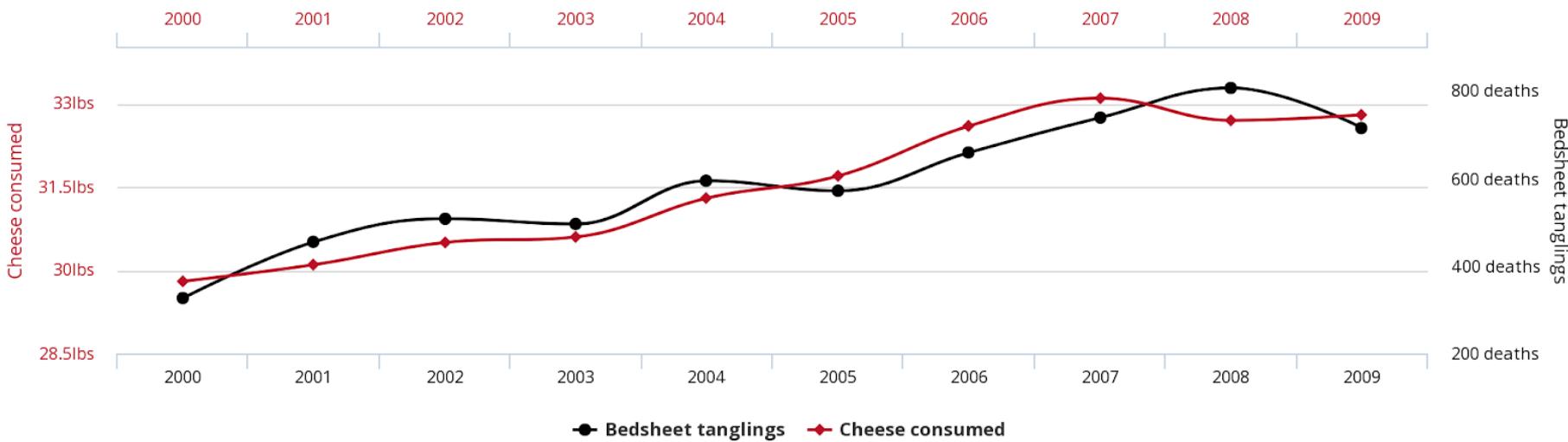
'12

'13



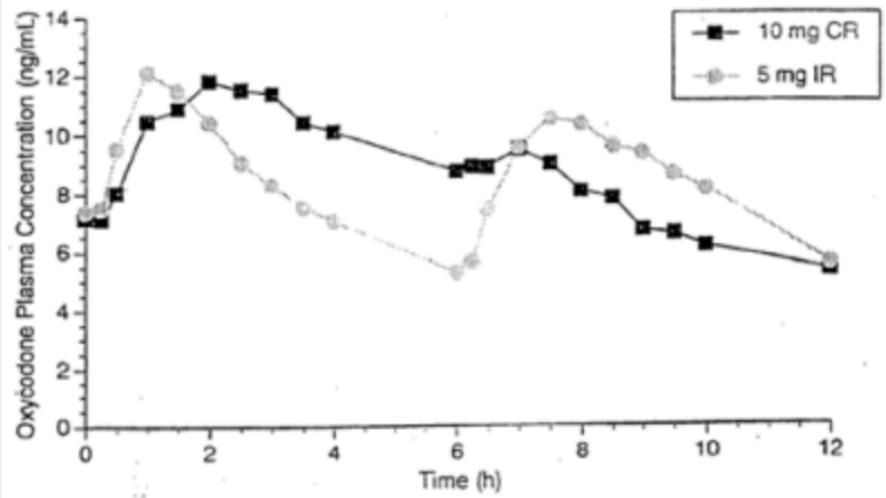
THERE'S A BLOG FOR IT...

Per capita cheese consumption
correlates with
Number of people who died by becoming tangled in their bedsheets



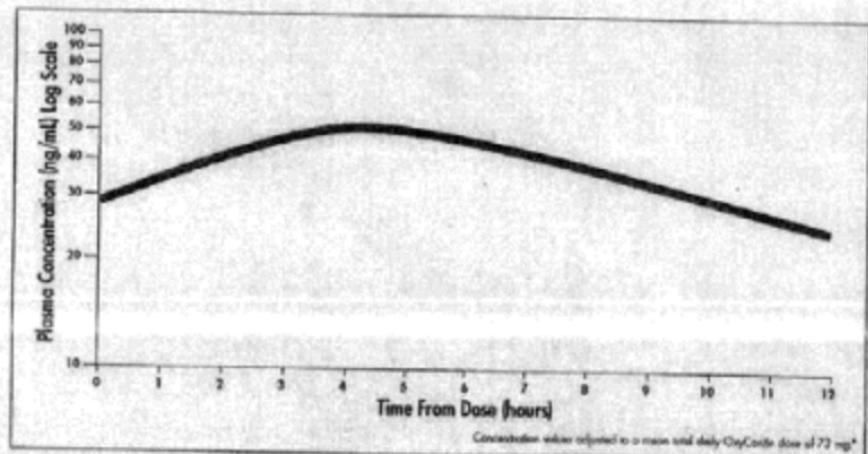
tylervigen.com

BUT IRL...



Purdue's clinical trials results comparing changes in drug concentration levels over time btwn. new drug ('10mg CR') taken every 12 hrs vs. old drug every 6 hrs

Q12h dosing provides smooth and sustained blood levels.



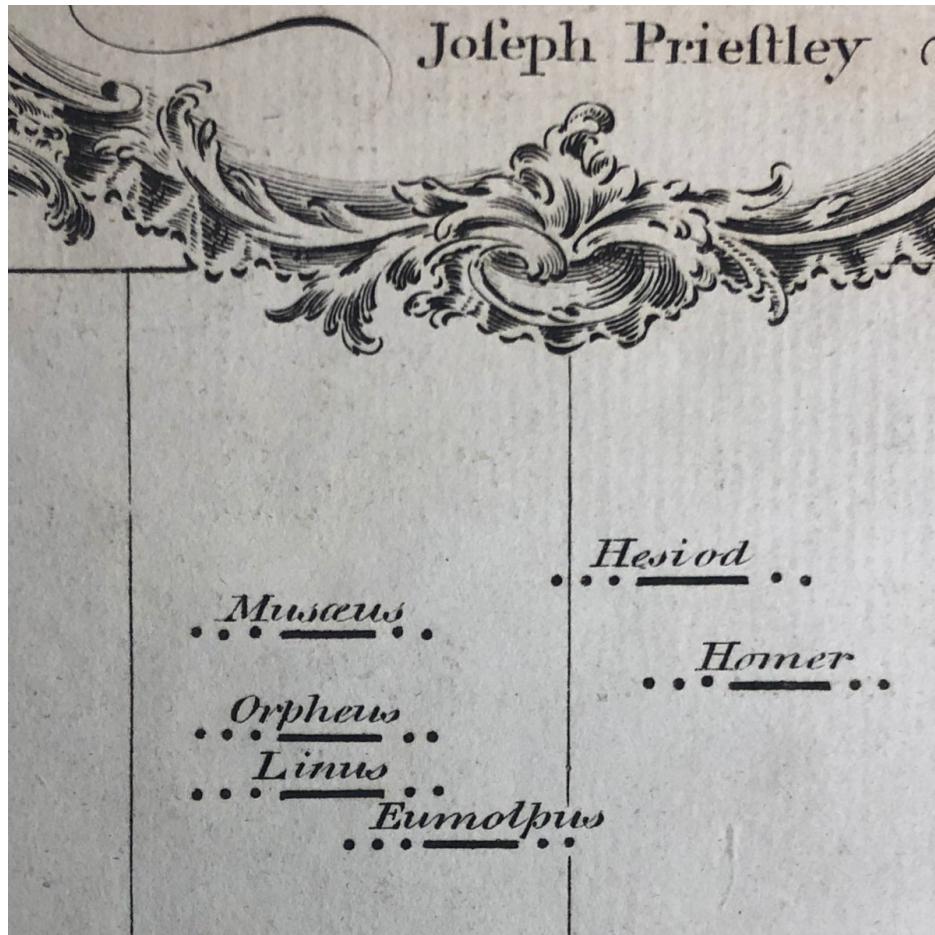
— Fewer "peaks and valleys" than with immediate-release oxycodone

Purdue's marketing materials submitted to FDA

UNCERTAINTY



18TH C. UNCERTAINTY



If it be thought that a person was born or died but a little before or after a certain time, the full line begins or ends at the certain time, and a single dot only is placed at the beginning or end of the line; see the line representing the life of Herodotus in the specimen. If history inform us that he was born or died *about* a certain time, a dot is placed just under the beginning or end of the full line see Philip and Thucydides. If the birth be certain but the death uncertain, the line begins full but ends in dots; as Abul Pharai and Alain in the chart. If, as is often the case, the death be known but the birth unknown, the line begins with dots, and ends full, see Epaminondas. When it is said that a writer flourished at or about a particular time, a short full line is drawn about two thirds before and one third after that particular time, with three dots before and two after it; because, in general, men are said to flourish much nearer the time of their death than the time of their birth. If it be uncertain even in what century a person lived or flourished, there is no full line made at all, but only dots or a broken line where I imagine it is most probable he

SHOW RANGES

Stephen Curry

GOLDEN STATE WARRIORS
POINT GUARD
27 YEARS OLD



WEIGHTED AVERAGE OF
PAST THREE SEASONS

● BAD ○ AVG. ● GOOD

PERCENTILE

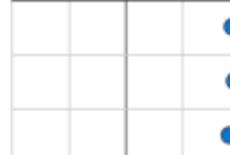
50TH

VITALS	
Height	6'3"
Weight	185
Draft position	7



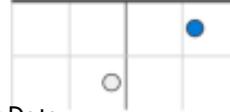
SCORING

True shooting %	62%
Free throw %	90%
Usage %	28%



TENDENCIES

3 pt. frequency	47%
FT frequency	25%



538 NBA projections, via FlowingData

PASSING/BALL HANDLING

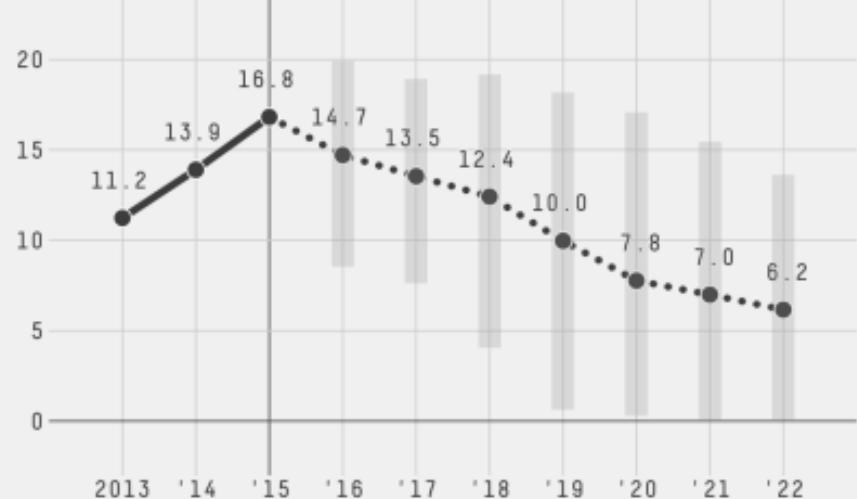
Assist %	38%
----------	-----



WINS ABOVE REPLACEMENT PROJECTION

CATEGORY: MVP CANDIDATE

90TH—
10TH—
CONFIDENCE INTERVAL
PROJECTION

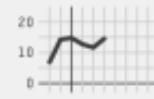


PERFORMANCE OF THE 10 MOST COMPARABLE PLAYERS

1 Chris Paul

YEAR: 2013

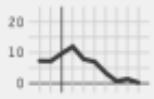
SIMILARITY: 41



6 Terry Porter

YEAR: 1991

SIMILARITY: 31



2 Terrell Brandon

YEAR: 1998

SIMILARITY: 34



7 Clyde Drexler

YEAR: 1990

SIMILARITY: 29

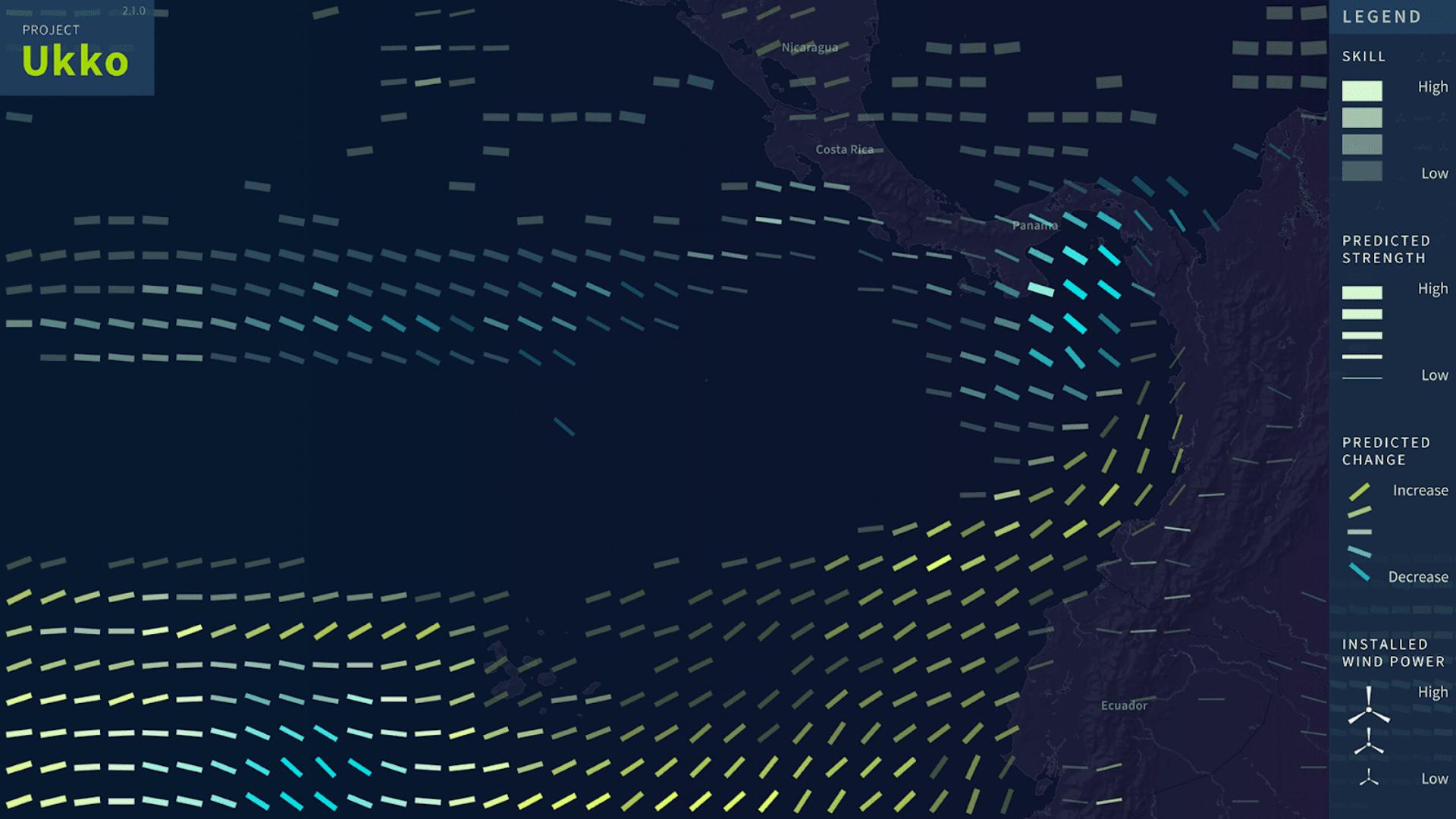


SHOW MULTIPLE OUTCOMES

LAST UPDATED AT 9:19 PM ET

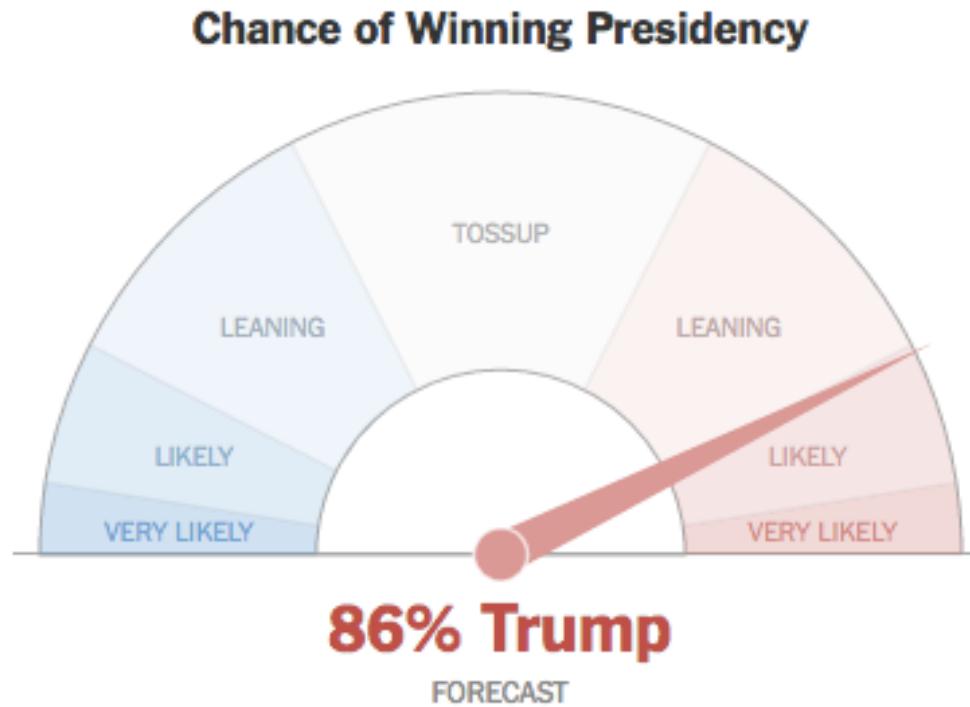
Median estimate:
271 delegates



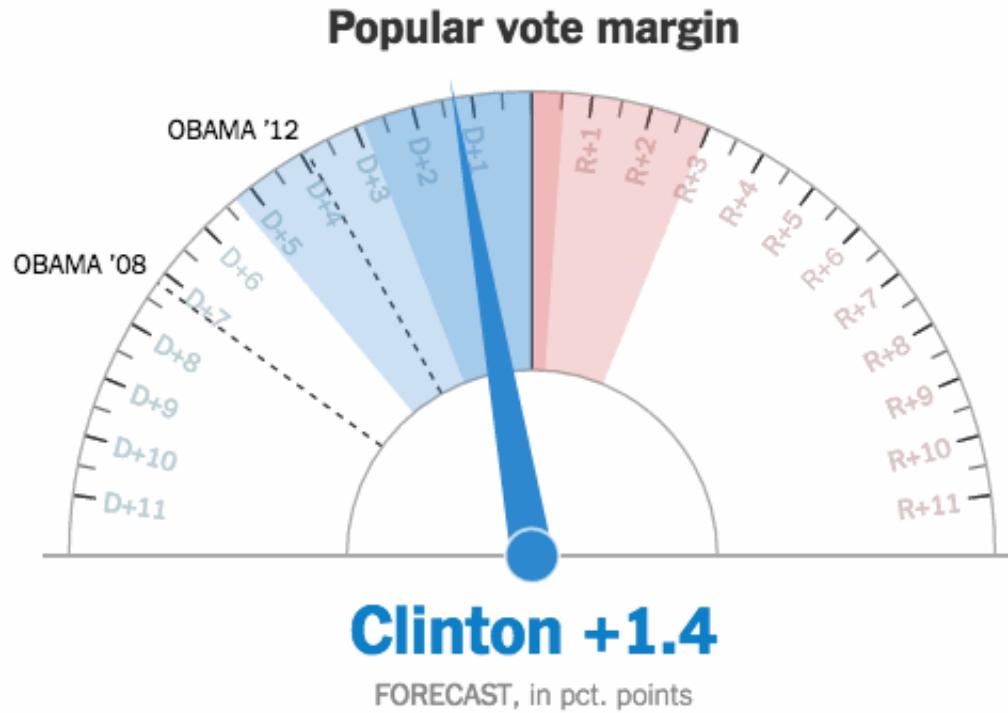


Moritz Stefaner

NYT “ELECTION NEEDLE”



VISUALIZING PROBABILITY



Ultimately, it is impossible to prevent the needle from being “wrong.” It is an inevitability, if the needle is put to work in enough elections. The challenge is to communicate the possibility of an upset, for as long as the possibility remains. Traditionally, we have used probabilities to do so, but we [increasingly doubt](#) whether this helps readers understand the uncertainty.

This was evident before the result of the 2016 election, and as a result we tried something new: a jitter, where the needle quivered to reflect the uncertainty around the forecast. Although many readers disliked it, the jitter reflected an earnest attempt to give tangible meaning to abstract probabilities. Nonetheless, we turned the jitter off for all of our 2017 forecasts.

Tonight, readers will have the option to turn the jitter off. We expect that some readers will opt to do so, but remember this: Switching it off only hides the uncertainty — it doesn’t make it go away.

Ultimately, it’s not over until the race is called.

Nate Cohn is a domestic correspondent for The Upshot. He covers elections, polling and demographics. Before joining The Times in 2013, he worked as a staff writer for The New Republic. [@Nate_Cohn](#)

Josh Katz is a graphics editor for The Upshot, where he covers a range of topics involving politics, policy and culture. He is the author of “Speaking American: How Y’all, Youse, and You Guys Talk,” a visual exploration of American regional dialects. [@jshkatz](#)

Left: NYT [“The Needle's Back...”](#), Right: via Matthew Kay OpenVis 2018

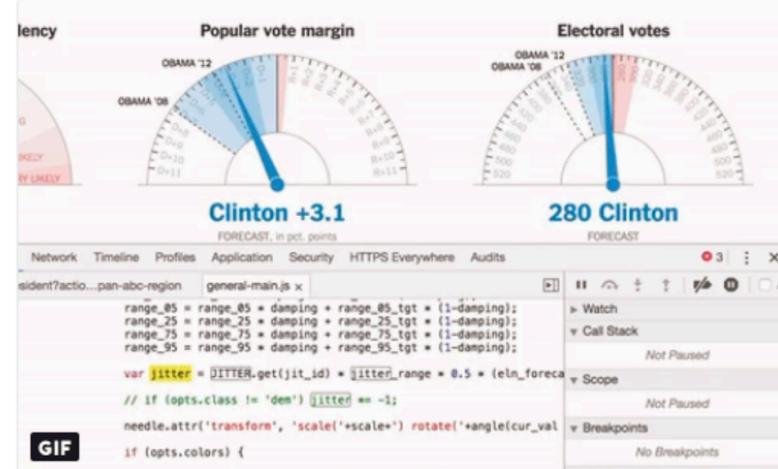


Alp Toker

@atoker

[Follow](#)

Looking for trends in [@nytimes](#)'s presidential forecast needle? Don't look too hard - the bounce is random jitter from your PC, not live data



Richard Porczak

@tsiro

[Follow](#)

straight up: the NYT needle jitter is irresponsible design at best and unethical design at worst and you should stop looking at it

9:58 PM - 8 Nov 2016

509 Retweets 882 Likes



17



509

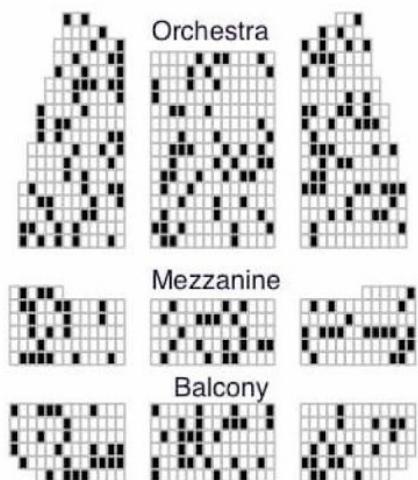


882



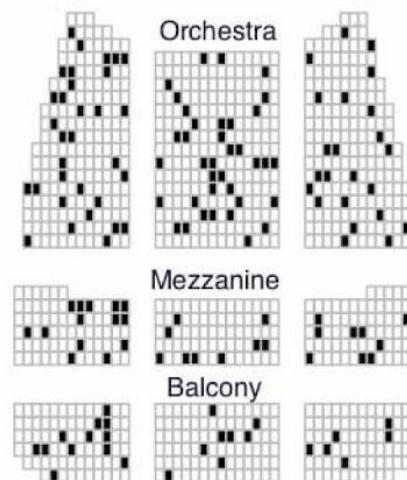
“RISK COMMUNICATION” TECHNIQUES

FiveThirtyEight: Trump's Chances



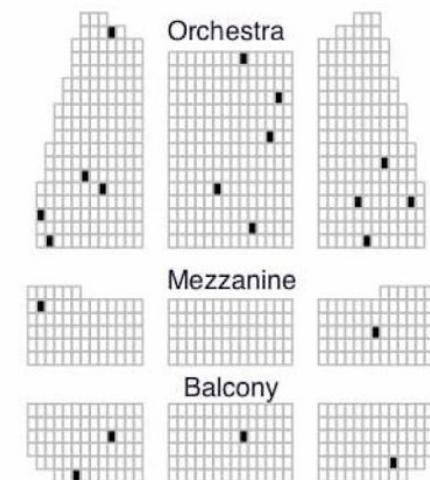
286 cases in 1,000

NYT Upshot: Trump's Chances



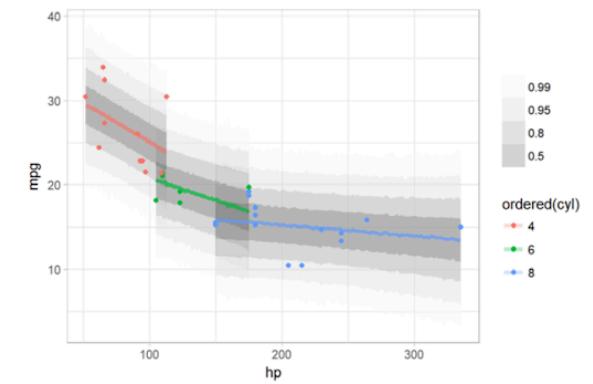
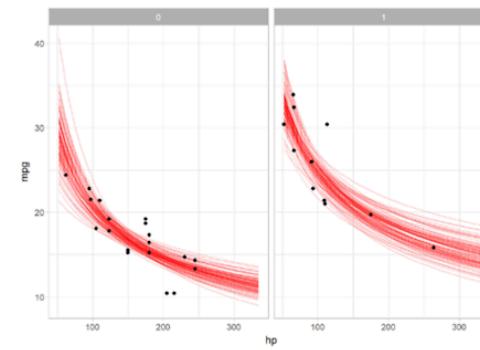
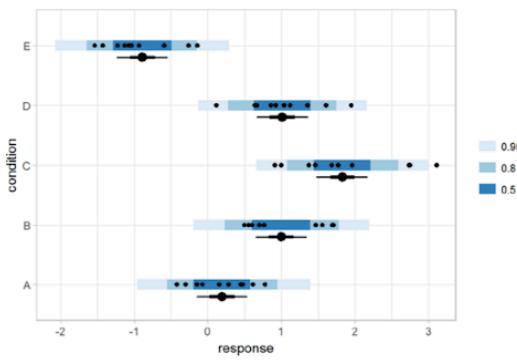
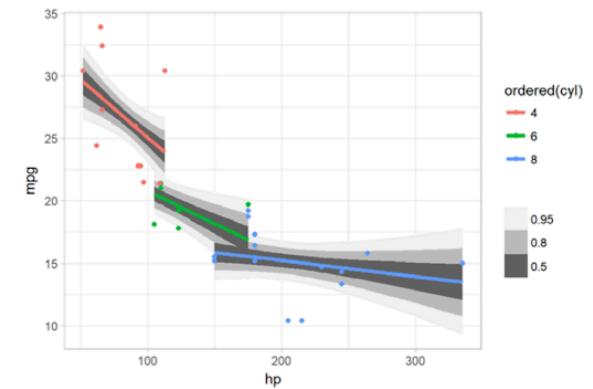
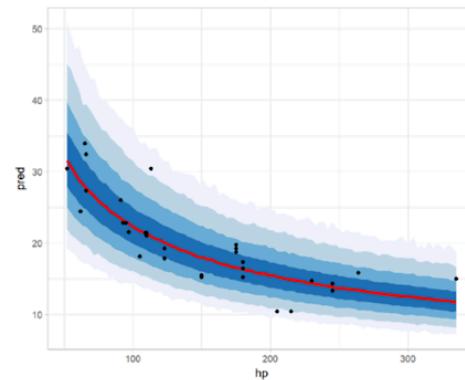
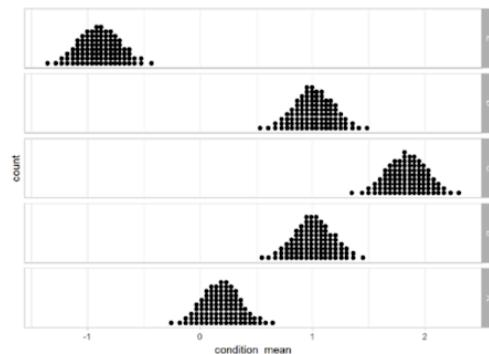
150 cases in 1,000

HuffPo Pollster: Trump's Chances



20 cases in 1,000

UNCERTAINTY IN CHARTS



LAB INSTRUCTIONS

Open assignment: https://classroom.github.com/a/Rmp7D_jC

Click “Clone or Download” to get repository URL

In terminal, type `git clone [URL]` to get a local copy

Start Jupyter notebook, open `4-matplotlib-exercises.ipynb`