

432 Class 26 Slides

thomaselove.github.io/432

2022-04-21

Sources for Today's Material include

- Howard Wainer *Visual Revelations*
- Nathan Yau's work at flowingdata.com
- Andrew Gelman and andrewgelman.com
- Christopher Gandrud and his book *Reproducible Research with R and R Studio*
- Karl Broman *Creating Effective Figures and Tables* at tinyurl.com/graphs2017
- Edward Tufte and edwardtufte.com
- <http://www.datavis.ca/gallery/index.php>
- <https://www.boredpanda.com/world-war-2-aircraft-survivorship-bias-abraham-wald/>
- plus this link to financialgazette.co/zw

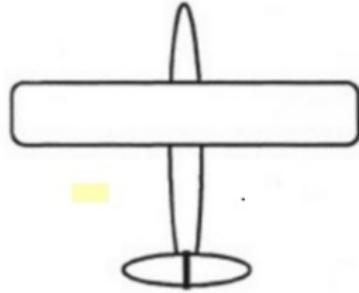
The Abraham Wald Story

During WWII, a group of statisticians had a difficult task to solve



They were asked to evaluate and determine which parts of the aircraft needed to be up-armored in order to minimize the damage of the planes from enemy fire.

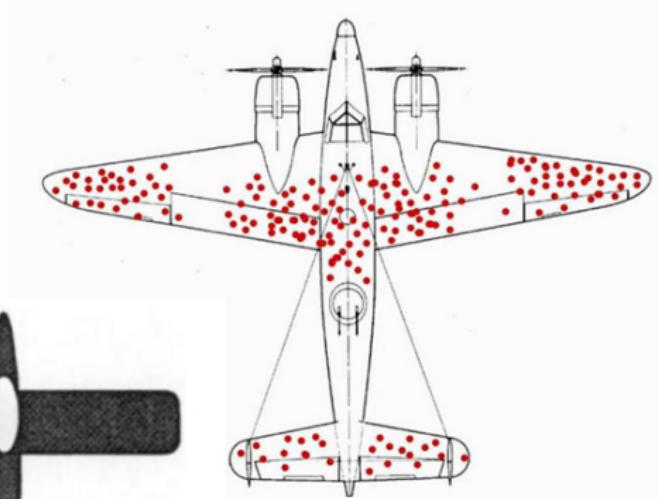
Diagram of all of the places where the planes were damaged the most

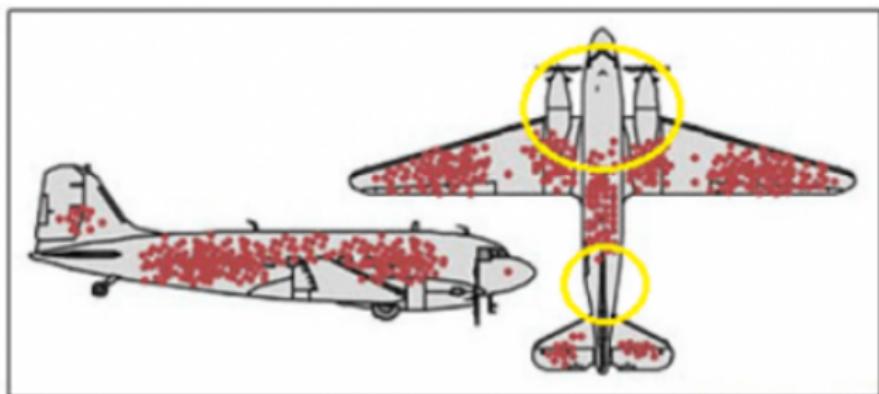


Before



After





Credit: Cameron Moll

Gentlemen, you need to put more armour-plate where the holes aren't because that's where the holes were on the airplanes that didn't return - Abraham Wald 1942.

Presenting Research

- Usually, this is highly abridged
 - Slide shows
 - Abstracts
 - Journal articles
 - Books
 - Websites

Your job is usually to announce the findings and try to convince us that the results are correct.

$P > 0.05$



GAME OVER, TRY AGAIN

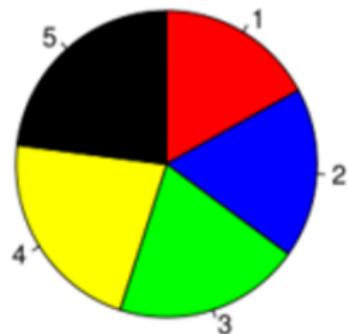
imgflip.com

You Have Ten Minutes?

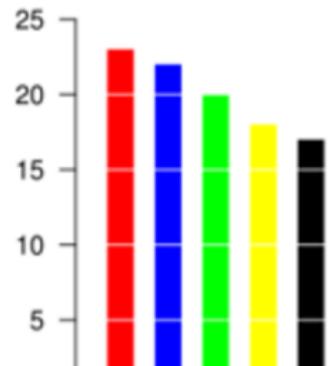
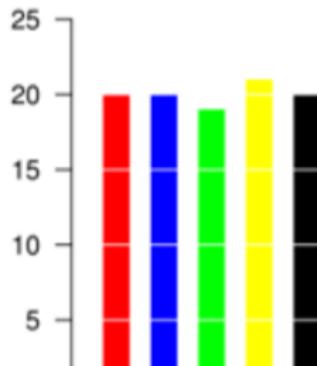
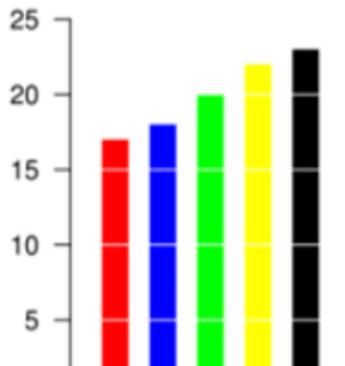
- No time for subtlety.
- Round, a lot.
- Edit, ruthlessly.
 - One pass through software (“default options”) is never enough.
 - Better for people to leave the table hungry than stuffed.
- Have something to say, and say it clearly.
- Some possibilities are never a good choice.

Stay Away from the Pie!

A



Which of these three bar graphs describes the same data as pie graph A?



Stay Away from the Pie!



Not that bars are always better

<https://twitter.com/HWippick/status/1118738492983521286/photo/1>

STATES WITH THE OLDEST NATIONAL PARKS

Year founded

1872

1890

1890

1890

1899

Yellowstone,
Wyoming

Sequoia,
California

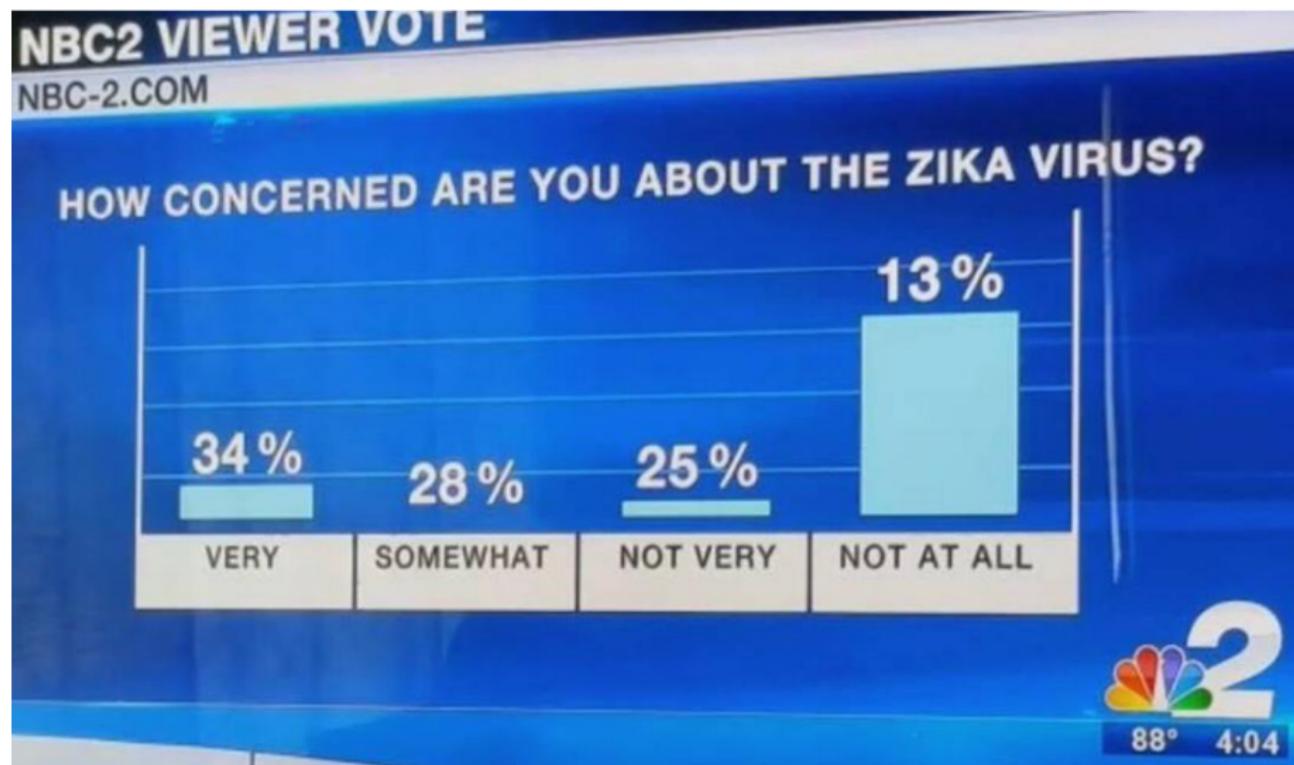
Kings Canyon,
California

Yosemite,
California

Mt. Rainier,
Washington

Not that bars are always better

https://twitter.com/asher_rosinger/status/1119278062804328448/photo/1



What's Wrong With This Picture?



AMERICANS WHO HAVE TRIED MARIJUANA

CBS NEWS POLL

51%
TODAY

43%
LAST YEAR

34%
1997



Source: MOE +/- 4%

HIGH SUPPORT FOR LEGALIZING MARIJUANA
MORE THAN HALF OF AMERICANS SAY THEY'VE TRIED POT

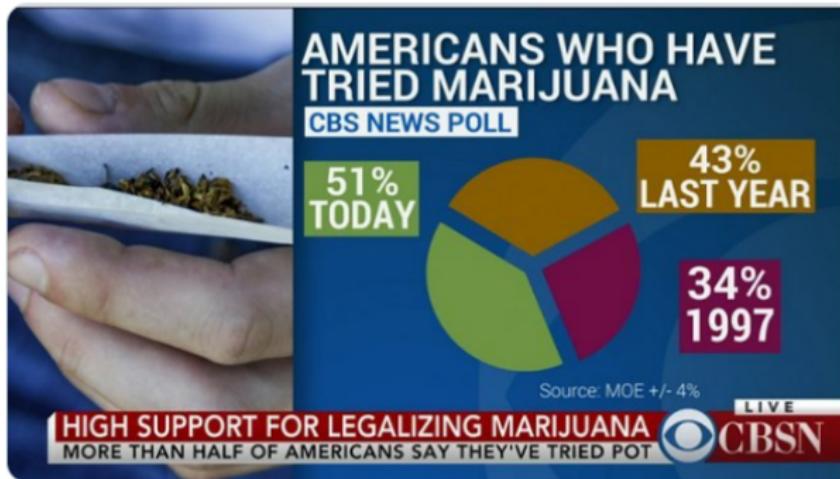


Dorsa Amir

@DorsaAmir

Easily the funniest data viz I've ever seen.

- (1) First of all, it's a pie chart.
- (2) The total is greater than 100%.
- (3) The relevant categories are today, last year, and... the year 1997?
- (4) The margin of error (MOE) is listed as the source.



7:06 PM · Apr 17, 2019 · Twitter Web Client

Clearly Communicating Quantitative Information

- Are the most important elements or relationships visually most prominent?
- Are the elements, symbol shapes and colors consistent with their use in previous graphs?
- Are all of the graphical elements necessary to convey the relationships?
- Are the graphical elements accurately positioned and scaled?

Source: <http://www.datavis.ca/gallery/index.php>

What are you trying to do?

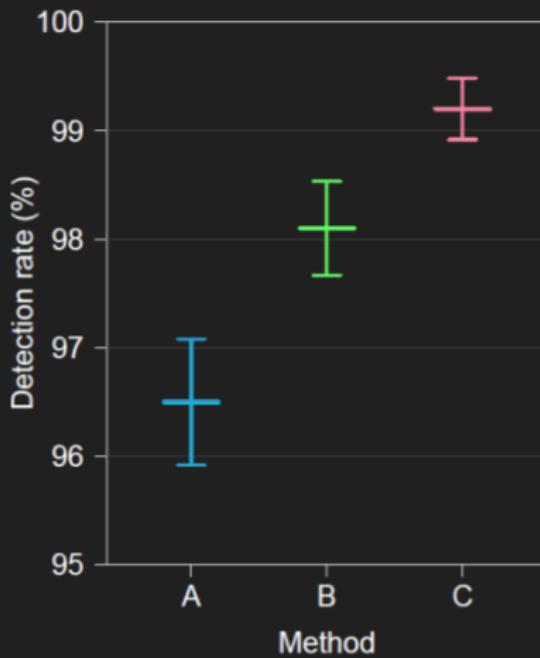
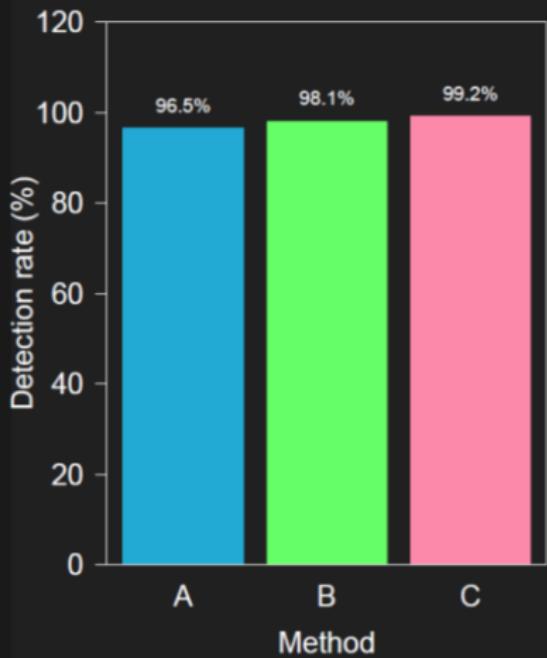
- Is this **information visualization** (grabby, visually striking - dramatize the problem to draw the casual viewer in deeper)
- Or **statistical graphics** (reveal patterns and discrepancies for viewers who are already interested in the problem)

Make tradeoffs carefully - meaningful choices.

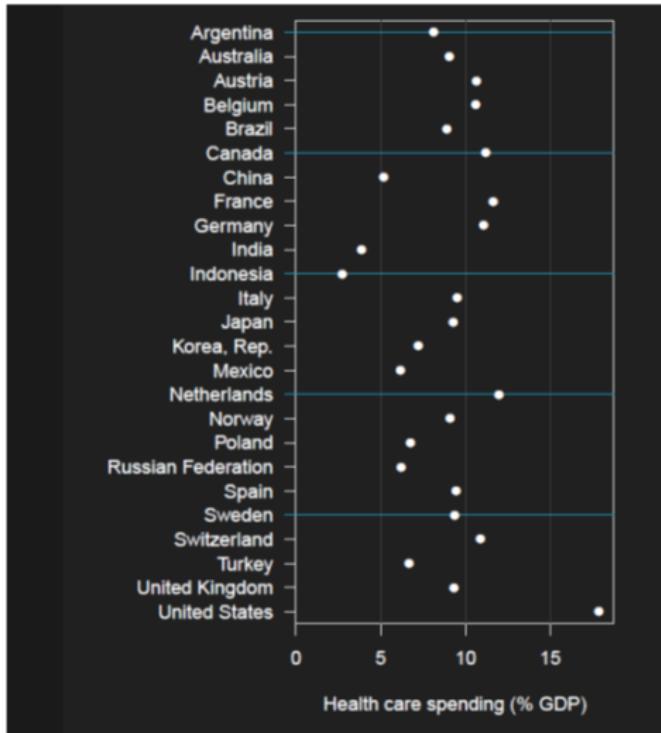
Displaying data well

- Be accurate and clear.
- Let the data speak.
 - Show as much information as possible, taking care not to obscure the message.
- Science not sales.
 - Avoid unnecessary frills (esp. gratuitous 3d).
- In tables, every digit should be meaningful. Don't drop ending 0's.

Must you include 0?

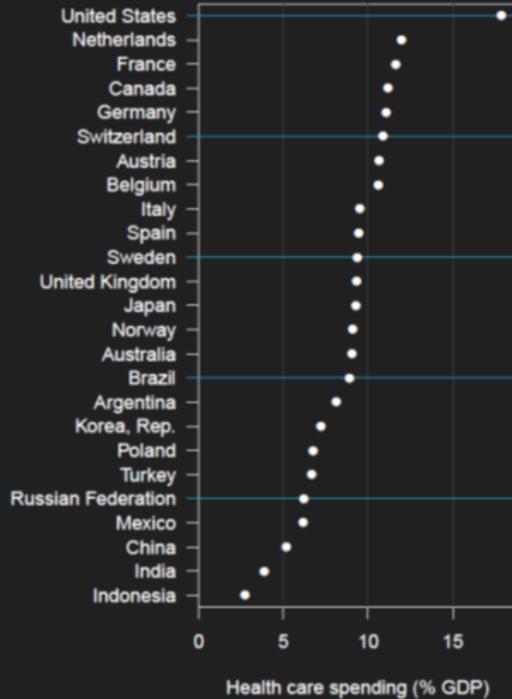
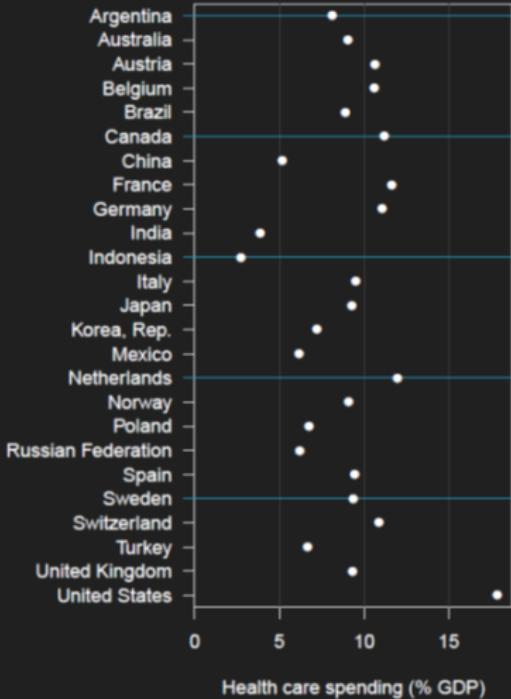


From Karl Broman...



Karl Broman, “Creating Effective Figures and Tables” at tinyurl.com/graphs2017

Don't sort alphabetically



Karl Broman, "Creating Effective Figures and Tables" at tinyurl.com/graphs2017

You Have Ten Minutes?

- No time for subtlety.
- Round, a lot.
- Edit, ruthlessly.
 - One pass through software (“default options”) is never enough.
 - Better for people to leave the table hungry than stuffed.
- Have something to say, and say it clearly.
- Stay away from the pie.

Data Visualization: Napoleon's Russian Campaign

CHAPTER 4 Three Graphic Memorials

"Hear, forget; see, remember." The wisdom of this ancient Confucian saying is apparent. Memorable memorials are visual. Who can ever forget the tragedy chronicled by the austere black granite wall that is the Vietnam Memorial? It is massive in form and content, built from the space taken by the more than 58,000 names inscribed upon it. As the loss of life increases, so too does the height of the wall, and the emotions it evokes. It is a very personal thing. William A. Atwell, Terry Lee Dillard, Ward K. Patton, Jerry Lee Graves, Edward J. Downs, John E. Rice, Jack M. Strong—these names join with thousands of others to form the wall. The interaction of the monument with those who come to it, whether to seek out a particular name or to picnic, often becomes part of the diverse images we take away with us. The tragedy of Vietnam written in the small becomes large and indelible.

The History

It's 1812.

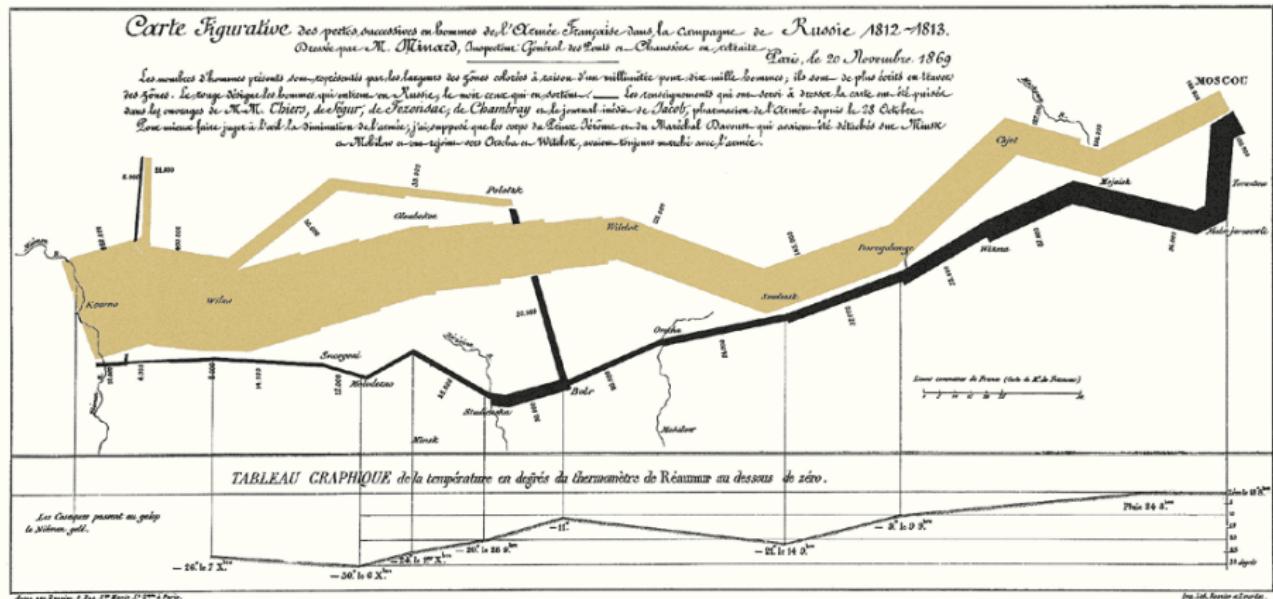
- Napoleon has most of Europe (outside of the United Kingdom) under his control.
- But he cannot break through the defenses of the U.K., so he decides to place an embargo on them.
- The Russian Czar, Alexander, refuses to participate in the embargo.

So Napoleon gathers a massive army of over 400,000 to attack Russia in June 1812.

- Meanwhile, Russia has a plan. As Napoleon's troops advance, the Russian troops burn everything they pass.

Charles Minard's original map

Napoleon's disastrous Russian Campaign of 1812



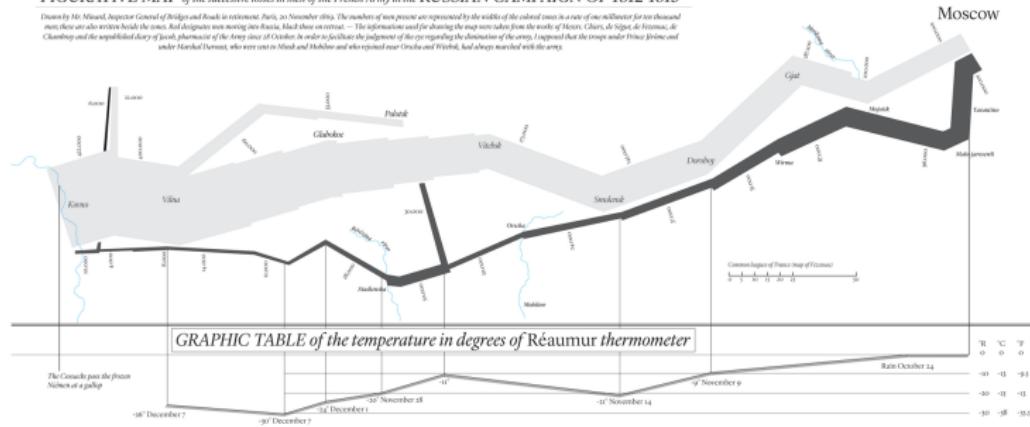
Napoleon's Russian Campaign

Memorializing that portion of the generation of young French men lost in Napoleon's ill-fated Russian campaign was surely part of Charles Joseph Minard's motivation in the construction of his famous 1869 graphic. Minard's plot, shown in [figure 1](#), depicts the movement of the French army from the time it crossed the Polish-Russian border with 422,000 men in June of 1812. The shrinking size of the army is characterized by the progressive narrowing of the broad band stretching across the map. In the original scale, each millimeter of its width represents 10,000 men. When the army reached Moscow in September, only 100,000 remained. The city was deserted, and the army began its retreat, depicted by the darker line below. It is linked to the temperature scale showing quantitatively the depths of the Russian winter. The banks of the Berezina River were littered with the bodies of the 22,000 men who perished as the November temperature dropped to -20° . When the remainder of the army crossed into Poland as the year ended, only 10,000 men remained.

A Modern Redrawing of Minard's Original Map

FIGURATIVE MAP of the successive losses in men of the French Army in the RUSSIAN CAMPAIGN OF 1812-1813

Drawn by Mr. Minard, Inspector General of Bridges and Roads in Government, Paris, on November 1819. The numbers of men present are represented by the widths of the colored routes in a ratio of one millimeter for ten thousand men; these are also written beside the routes. Red designates men marching into Russia, black those en route. — The information used for drawing the map was taken from the works of Hoche, Chabri, de Togot, de Frennes, de Chabord and the unpublished diary of Joseph, pharmacist of the Army since 18 October. In order to facilitate the judgment of the eye regarding the diminution of the army, I repeated that the troops under Prince Jérôme and under Marshal Davout, who were sent to Smolensk and Bobruisk and who rejoined near Orsha and Vitebsk, had always marched with the army.



Source: By Iñigo Lopez - Own work, CC BY-SA 4.0, at [this link](#)

What are we looking at?

- The numbers of Napoleon's troops by location (longitude)
 - Organized by group (at one point they divided into three groups) and direction (advance, then retreat)
- The path that his troops took to Moscow and back again
- The temperature experienced by his troops when winter settled in on the return trip
- Historical context, as shown in the passage of time
- Geography (for example, river crossings)

Wainer: Chapter 4 [c]

The story of the tragedy is clear. We can see the bodies frozen into the snow. Marey told how this graph “brought tears to the eyes of all France.”¹ No wonder; there were few families unaffected.

Minard’s depiction of Napoleon’s Russian campaign has been characterized as perhaps “the best statistical graphic ever drawn.”² Why? It is not the quality of the pen stroke, although it certainly passes muster in that regard. It is the importance and richness of the data. A single page carries six variables that tell the evocative story of where and how thousands of men died. Its poignancy is heightened through the immediate and graphic answer to the question, Compared to what? Ten thousand men returned. A lot or a few? Opposing the returning trickle against the departing torrent answers the question. The difference between them measures the tragedy. But nowhere does the shrinking distance between two lines depict a more touching tragedy than in my next example.

I'll spare you that more tragic story for today.

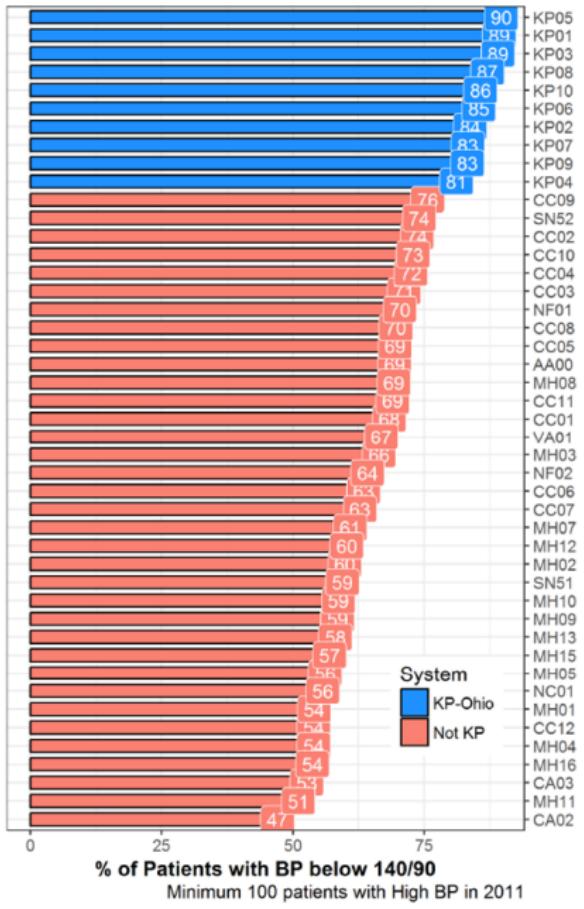
Do as I do? (Better Health Partnership work)

Exhibit A. Better Health Partnership: Patient Characteristics, 2016

	Diabetes		Heart Failure		High Blood Pressure	
	Better Health Partnership	Range by Practice	Better Health Partnership	Range by Practice	Better Health Partnership	Range by Practice
# Health Systems	8		3		9	
# Primary Care Practices	72		35		79	
# Primary Care Providers	748		477		832	
# of Patients	51,305		8,488		183,745	
Insurance (%)						
Medicare	47.6	15 - 77	67.2	49 - 93	52.6	13 - 83
Commercial (and Veterans)	37.2	0 - 62	20.8	3 - 31	36.0	2 - 60
Medicaid	12.7	0 - 70	10.5	0 - 39	9.4	0 - 68
Uninsured	2.5	0 - 36	1.6	0 - 6	2.0	0 - 38
Race / Ethnicity (%)						
White	64.3	2 - 98	63.6	2 - 100	69.4	2 - 99
Black or African-American	28.1	1 - 97	31.8	0 - 98	25.8	0 - 98
Hispanic or Latino	4.7	0 - 68	2.9	0 - 49	2.7	0 - 59
Other Race / Ethnicity	2.8	0 - 9	1.7	0 - 4	2.1	0 - 8
Demographics						
Average Age	60.4	52 - 66	69.1	58 - 79	64.4	51 - 73
% Female	38.7	1 - 75	50.4	32 - 75	40.9	2 - 76
% Low Income*	26.6	0 - 82	31.5	0 - 81	33.2	0 - 83
% Low Education*	25.9	0 - 83	28.3	0 - 75	30.0	0 - 86
% living in Cleveland	33.9	0 - 93	42.5	0 - 87	31.3	0 - 94
% in Cuyahoga County	56.1	0 - 100	74.9	7 - 100	55.7	0 - 99
Population Health						
% with BP below 140/90	75.6	56 - 90	78.2	70 - 92	72.0	36 - 85
% with BMI below 30	33.3	16 - 47	47.8	29 - 57	46.4	9 - 57
% Not Using Tobacco	76.6	42 - 92	87.4	63 - 96	77.1	12 - 93

* Living in neighborhoods with median income < \$33,000; with high school graduation rate < 83%.

Patients with Good BP Control (%), 2011



Social Determinants and % Achieving Good BP Control, 2013

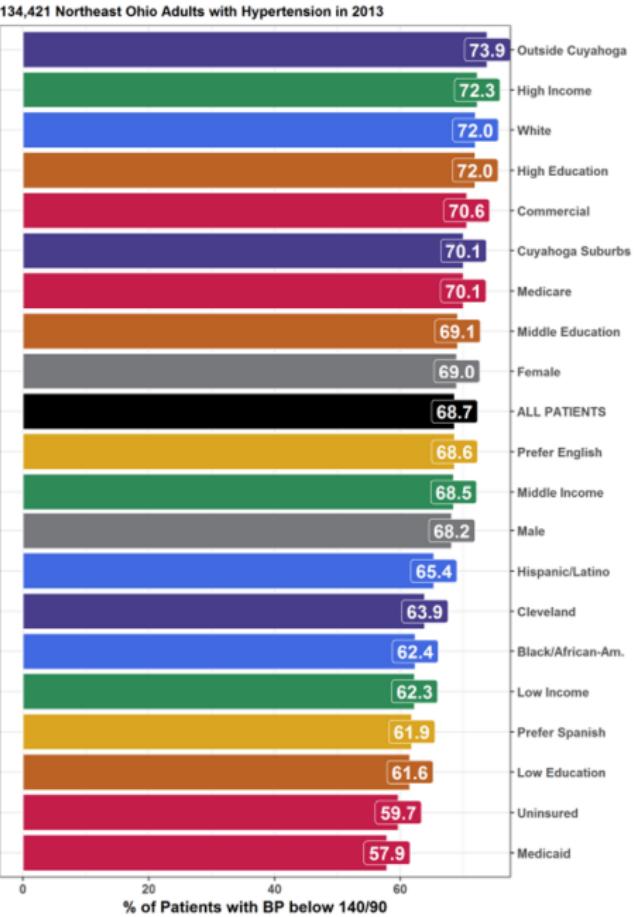
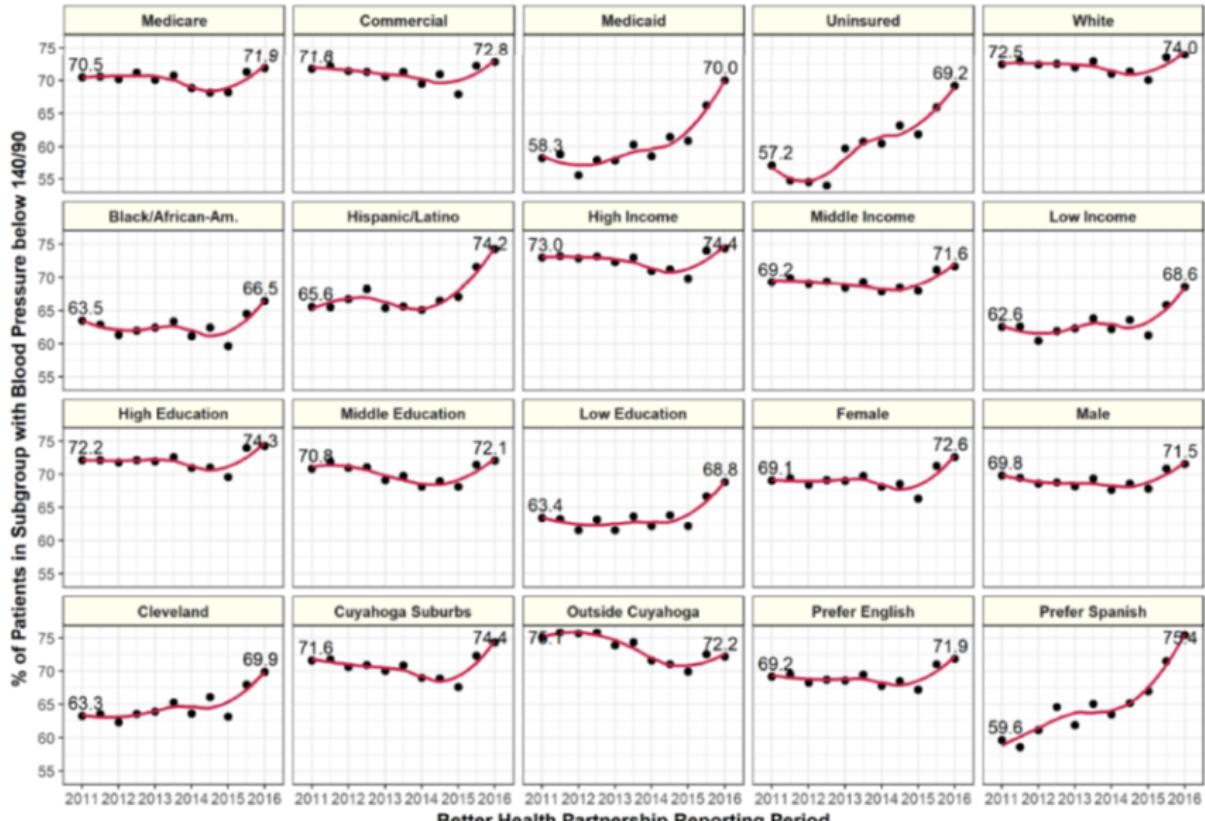


Exhibit E. Trends in Good Blood Pressure Control (< 140/90) by Subgroup, 2011-2016



Shown with fitted loess smooths