



Everyday Stats for Data Wonks

Susan M. McMillan, April 2026



Everyday stats...



DS r/datascience



Search in r/datascience



Home



Popular



Answers BETA



Explore



DS r/datascience • 5y ago
Megatheorist



What kind of math and statistics do you actually use in a daily basis at work?

↑ 315 ↓ · 183



bi_expert • 5y ago

Addition. I kid you not. I have degree in data science, I studied math up to diff eq. I make six figures using nothing more than 3rd grade math.

↑ 20 ↓

16 more replies ▾



Everyday stats...



crackednut • 5y ago

Mostly division. Lots of ratios. This per user, that per user. Cutting edge stuff...

/s



EvanstonNU • 5y ago

Average, weighted average, and Algebra 1.



taguscope • 5y ago

You use algebra? Mr. Fancy pants here.





We all know much less than we'd like to admit
—but the little each of us has learned may be useful to others.

Alberto Cairo, A Conversation Among Friends

<https://openvisualizationacademy.org/courses/introduction-to-information-design-and-data-visualization/understanding-information-design-and-visualization/a-conversation-among-friends/>



Percentages



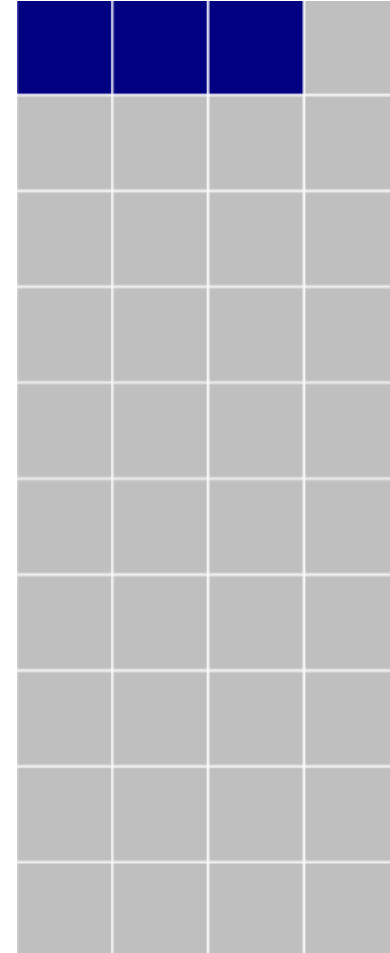
Because your end users may need reminders

Percentages start with fractions

2/10 is 20%



3/40 is 7.5%



- The second set is 50% larger (change from 2 to 3).
- But 3/40 is a smaller proportion (by 12.5 percentage points).
- Both the numerator and denominator change.



Base rates matter

Box C

Sensitivity of PDs to baseline size

Comparison of number of lesions within 1 individual over 2 time points

Comparison Values	Absolute Difference (D)	PD
110 vs 100	10	10%
60 vs 50	10	20%
15 vs 5	10	200%
11 vs 1	10	1000%
10 vs 0	10	Infinity/undefined

- **Percent Change:**

$$(\text{New value} - \text{Old} / \text{Old}) * 100$$

- Small base rates can yield high % change, so it's important to know the starting values...
- While keeping absolute values in sight.



% Change in the wild

HIGHLIGHTS

- The rate of property offenses decreased 9% from 2,019.7 per 100,000 persons in 2023 to 1,835.1 per 100,000 in 2024.



Descriptive Stats & Data Quality





Example employee survey data

File Home Insert Draw Page Layout Formulas Data Review View Automate Help Acrobat										
A1 X ✓ fx ID										
A	B	C	D	E	F	G	H	I	J	
1	ID	Status	Department	Director	Manager	Supervisor	Staff	Question	Response	Response Text
2	1	Complete	Human Resources	0	1	0	0	1. I know what is expected of me at work	4	Strongly Agree
3	2	Complete	Communications Office	0	0	0	0	1. I know what is expected of me at work	4	Strongly Agree
4	3	Complete	Parks and Recreation	0	1	0	0	1. I know what is expected of me at work	0	Not Applicable
5	4	Complete	Human Resources	0	1	0	0	1. I know what is expected of me at work	3	Agree
6	5	Complete	Communications Office	0	0	0	0	1. I know what is expected of me at work	0	Not Applicable
7	6	Complete	Prosecuting Attorney's Office	0	0	0	0	1. I know what is expected of me at work	4	Strongly Agree
8	7	Complete	Prosecuting Attorney's Office	0	0	0	0	1. I know what is expected of me at work	4	Strongly Agree
9	8	Complete	Finance and Performance Management	0	0	0	1	1. I know what is expected of me at work	4	Strongly Agree
10	9	Complete	Finance and Performance Management	0	0	0	0	1. I know what is expected of me at work	2	Disagree
11	10	Complete	Planning and Public Works	0	0	0	0	1. I know what is expected of me at work	4	Strongly Agree
12	11	Complete	Planning and Public Works	0	0	0	0	1. I know what is expected of me at work	4	Strongly Agree

<https://mavenanalytics.io/data-playground/employee-survey-responses>



Frequency tables

- Counts of cases
- Odd or unexpected values
- Check for duplicates

3	Row Labels	Count of Response ID
4	1. I know what is expected of me at work	1472
5	10. Overall I am satisfied with my job	1454
6	10. Overall I am satisfied with my job	22
7	2. At work, I have the opportunity to do what I do best every day	1471
8	3. In the last seven days, I have received recognition or praise for doing good work	1471
9	4. My supervisor, or someone at work, seems to care about me as a person	1471
10	5. The mission or purpose of our organization makes me feel my job is important	1471
11	6. I have a best friend at work	1471
12	7. This last year, I have had opportunities at work to learn & grow	509
13	7. This last year, I have had opportunities at work to learn and grow	962
14	8. My supervisor holds employees accountable for performance	1478
15	9. My department is inclusive and demonstrates support of a diverse workforce	1473
16	(blank)	
17	Grand Total	14725

ID	ID_Item	Status	Department	Director	Manager	Supervisor	Staff	Question	Response	Response.Text
All	All	All	All	All	All	All	All	All	All	All
12	12 1	Complete	Planning and Public Works	0	0	0	0	1. I know what is expected of me at work	3	Agree
12	12 1	Complete	Planning and Public Works	0	0	0	0	1. I know what is expected of me at work	3	Agree



What else to look for

- Are the variables of the expected type?
- Is the range what you anticipated (outliers)?
 - Maximum
 - Minimum
- "Spread"
- Counts, with attention to missing values
- Measures of Central Tendency

```
> summary(maven.analytics.survey)
```

```
      ID
Min.   : 1
1st Qu.: 3681
Median : 7362
Mean   : 7360
3rd Qu.: 11040
Max.   : 14710
```

```
      Status
```

```
Length:14725
```

```
Class :character
```

```
Mode  :character
```

```
      Response
```

```
Min.   :0.000
1st Qu.:3.000
Median :3.000
Mean   :2.984
3rd Qu.:4.000
Max.   :4.000
NA's   :135
```



Cleaned up “Question” frequency table

	Question	Count
1	1. I know what is expected of me at work	1471
2	2. At work, I have the opportunity to do what I do best ever...	1471
3	3. In the last seven days, I have received recognition or prais...	1471
4	4. My supervisor, or someone at work, seems to care about ...	1471
5	5. The mission or purpose of our organization makes me fee...	1471
6	6. I have a best friend at work	1471
7	7. This last year, I have had opportunities at work to learn an...	1471
8	8. My supervisor holds employees accountable for performa...	1471
9	9. My department is inclusive and demonstrates support of ...	1471
10	10. Overall I am satisfied with my job	1471



Measures of Central Tendency (MCT)

- Mean (Average)
 - Add all values, divide by the total count
 - Outliers (unusually high or low values) have an outsized effect
- Median
 - The midpoint of ordered data
- Mode
 - The most common outcome in the data
 - Can have multiple modes
 - This is the only MCT for categorical data
- Evaluate the variation ("spread") around the summary



Descriptive summary of Item 10 data by role

10. Overall I am satisfied with my job

Role	N	Median	Mode
Director	17	3.5	Strongly Agree
Manager	107	3.0	Strongly Agree
Supervisor	166	3.0	Agree
Staff	1181	3.0	Agree





Distributions



For understanding the data

Relying on summary statistics can be dangerous

The same statistics can summarize very different data.

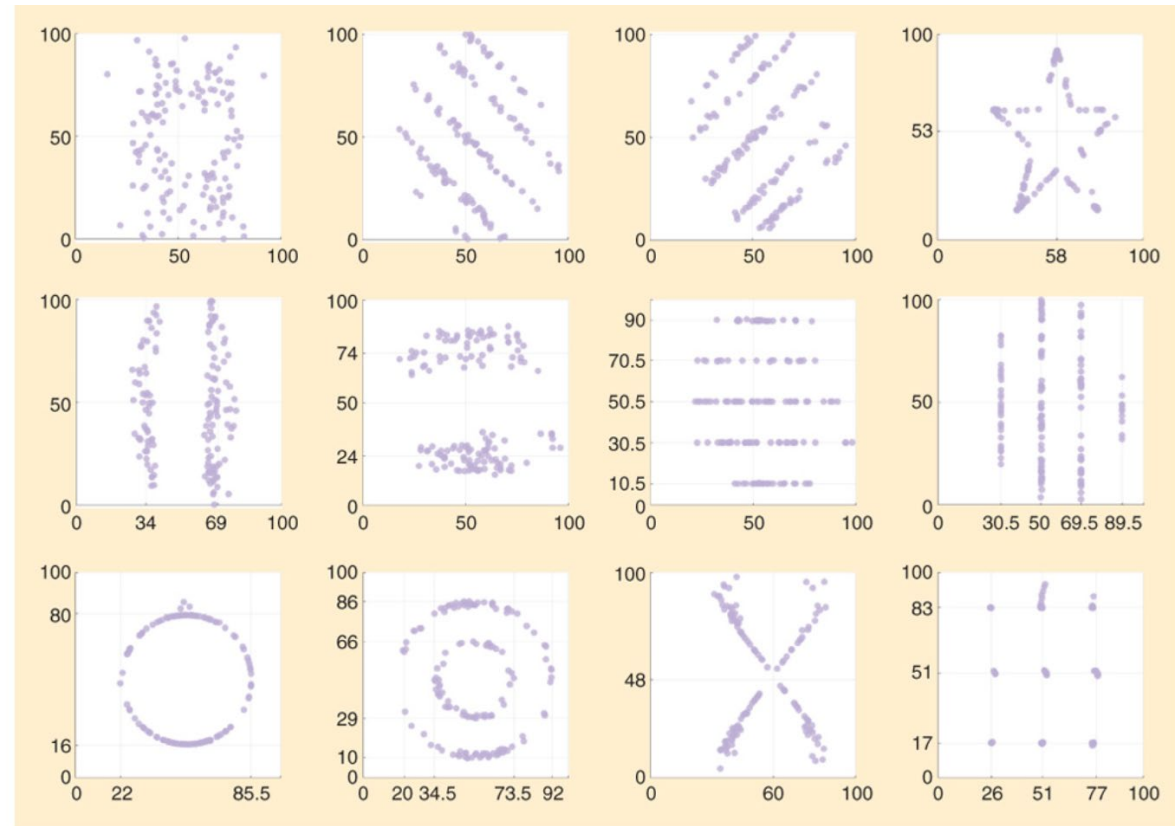
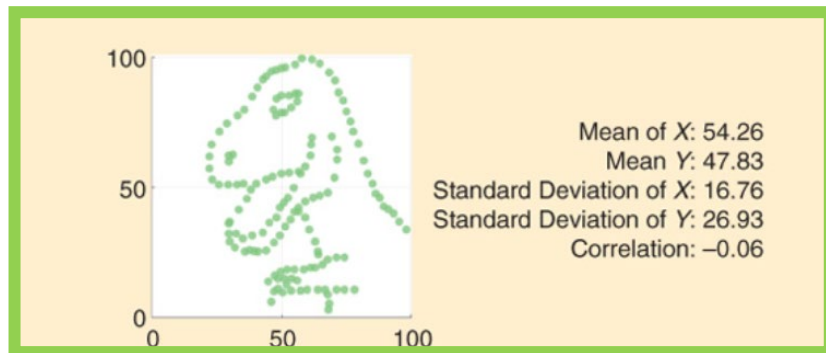


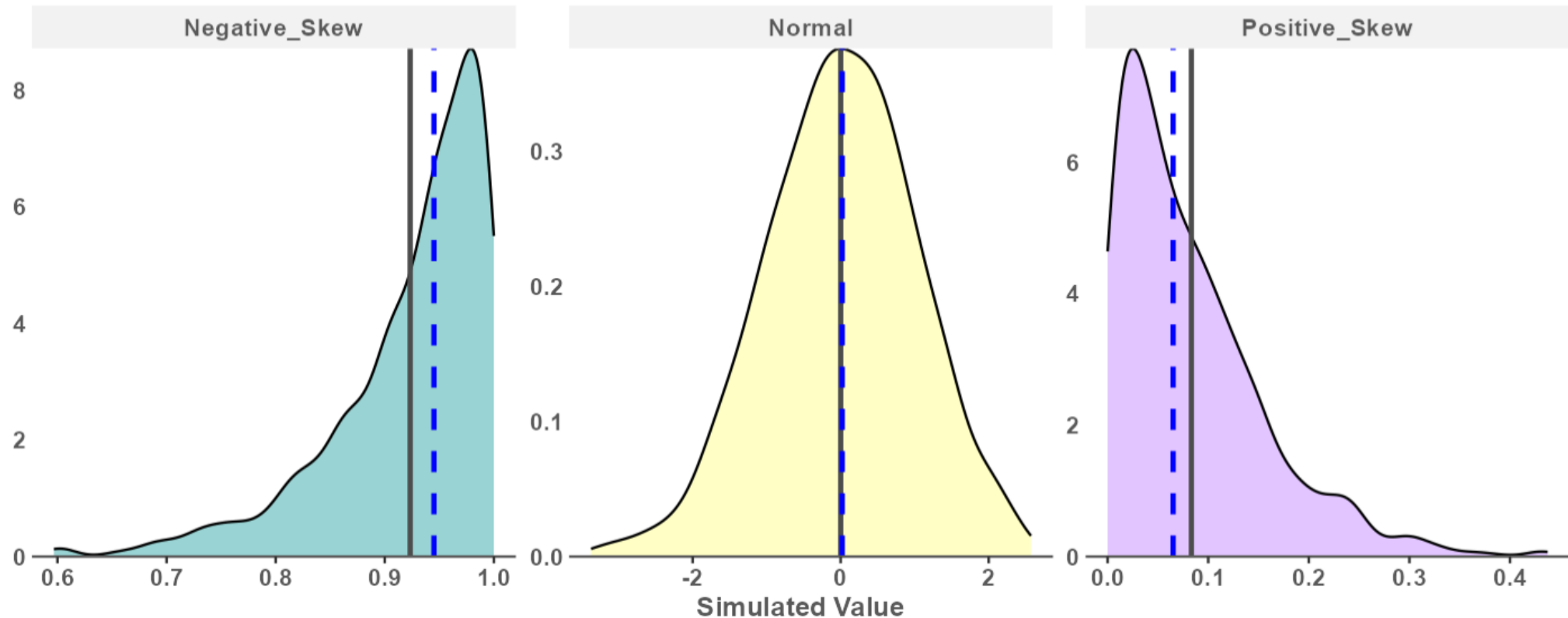
Fig1 Scatterplots showing 13 distinct data sets that share certain summary statistics. The figure was redrawn using the publicly available numerical data from Matejka and Fitzmaurice (2017).

Simulated distributions

The mean and median of skewed distributions diverge.

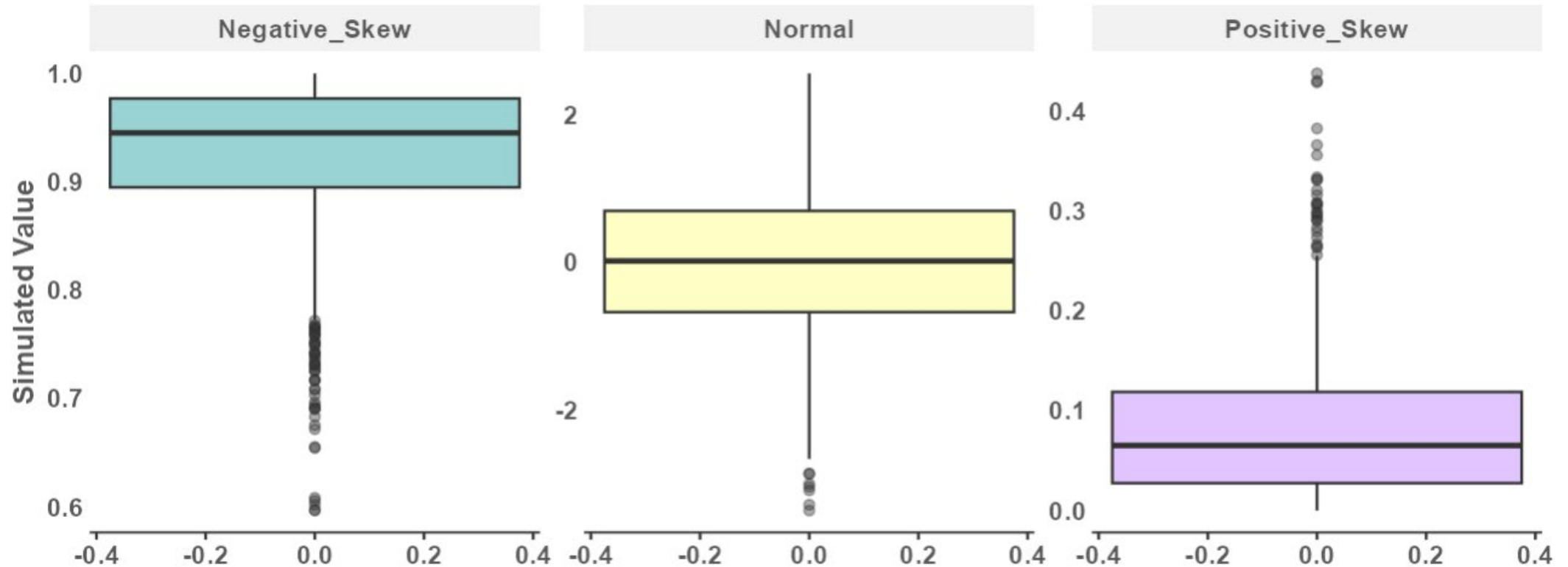
Solid vertical line = Mean

Dashed vertical line = Median



Simulated distribution box plots

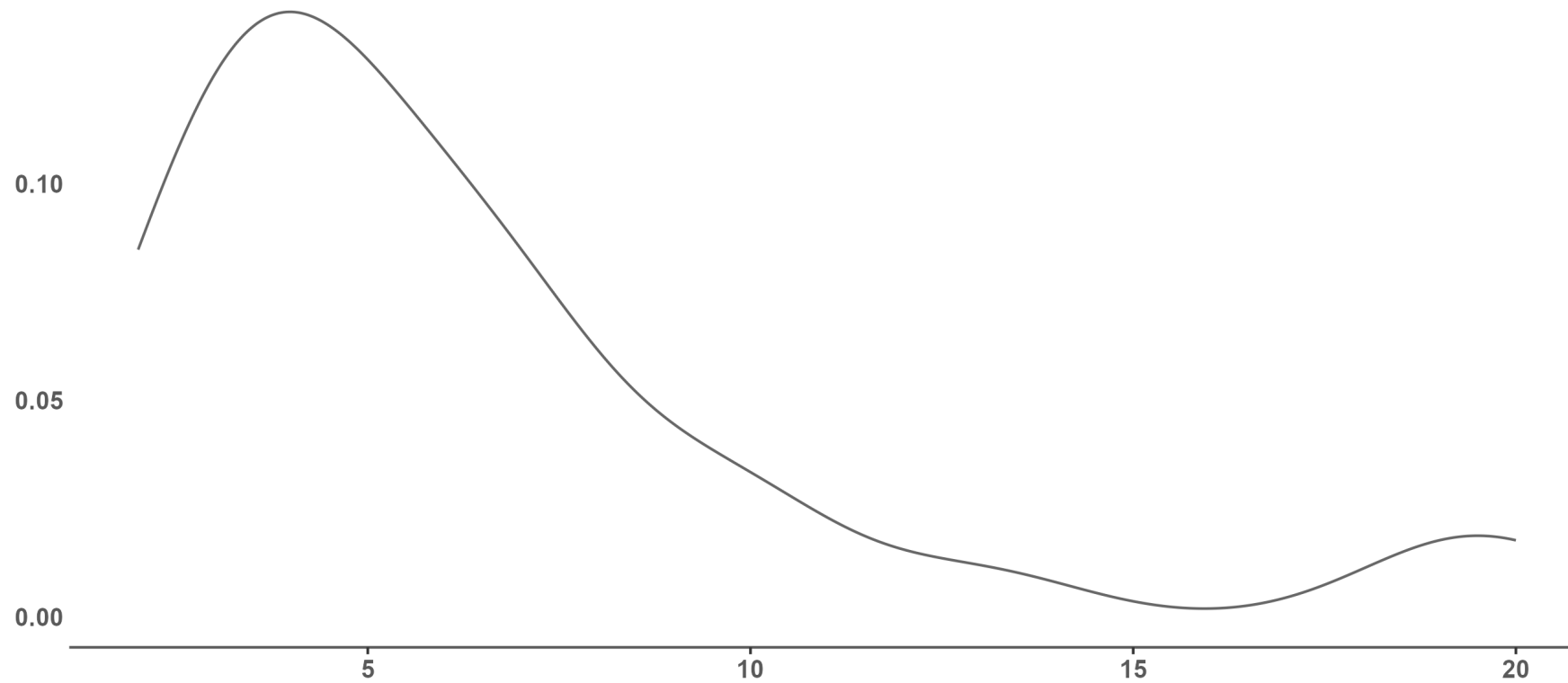
The boxplots also demonstrate skew.



Everyday distribution: AMR deaths in 2020

The distribution of deaths attributable to antimicrobial resistance (AMR) by EU country is...irregular.

Density plot (smoothed histogram): number of AMR deaths per 100,000 people in 29 EU countries.

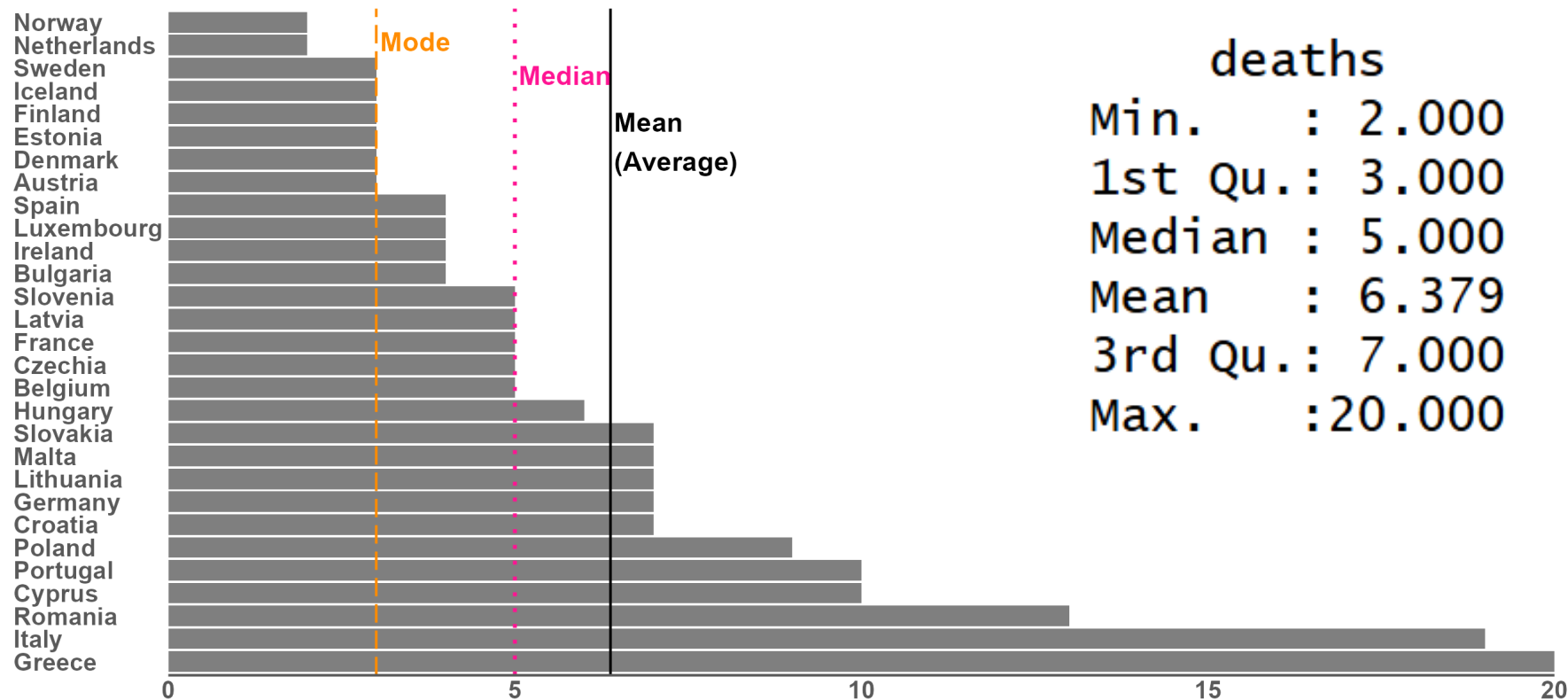


Source: <https://www.consilium.europa.eu/en/infographics/antimicrobial-resistance/#deaths>

AMR deaths with measures of central tendency

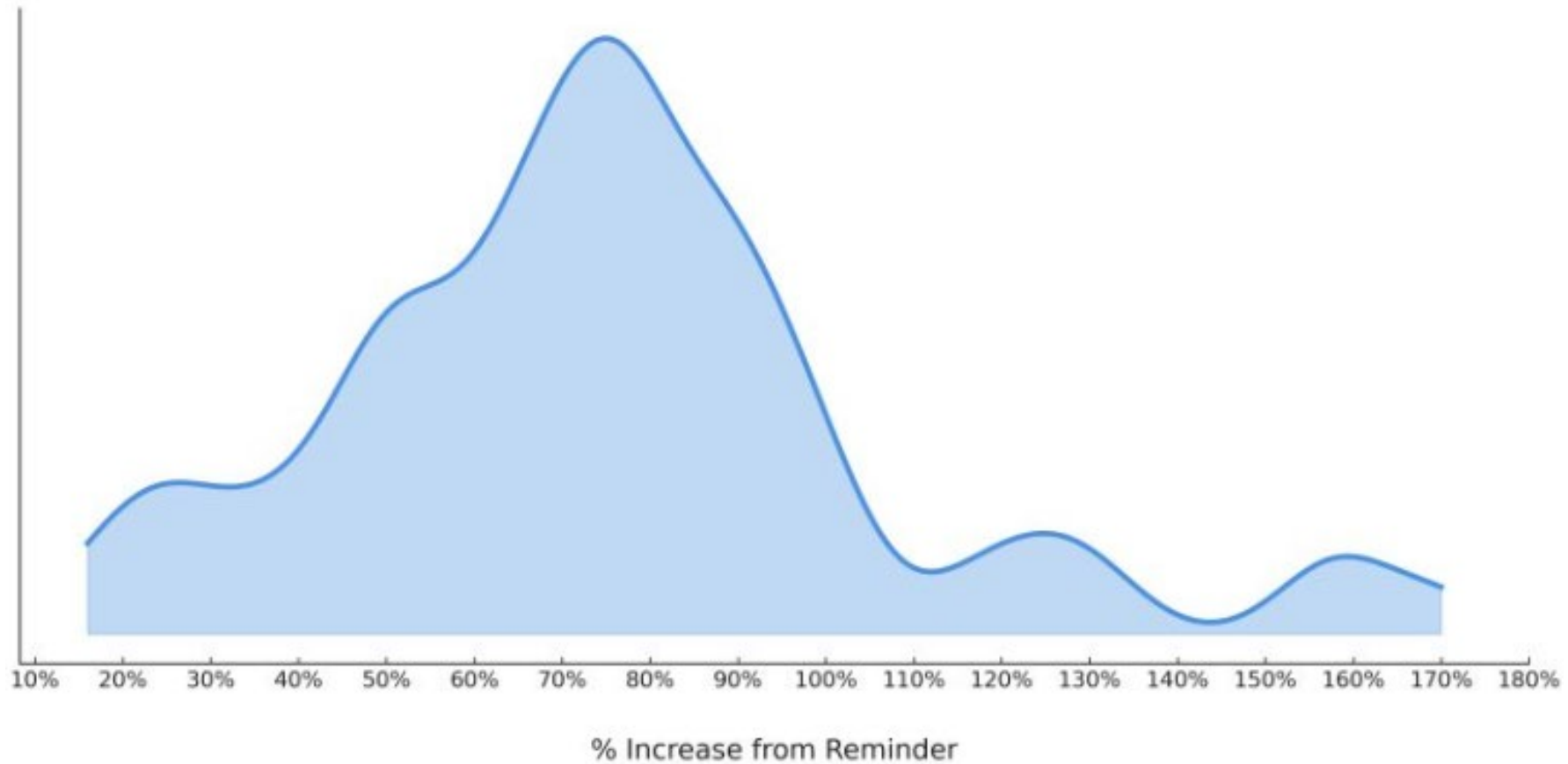
Greece and Italy have the highest antimicrobial resistance (AMR) death rates.

This graph shows the estimated number of deaths attributable to AMR per 100,000 people in 29 countries in 2020.





Everyday distribution: survey response rates



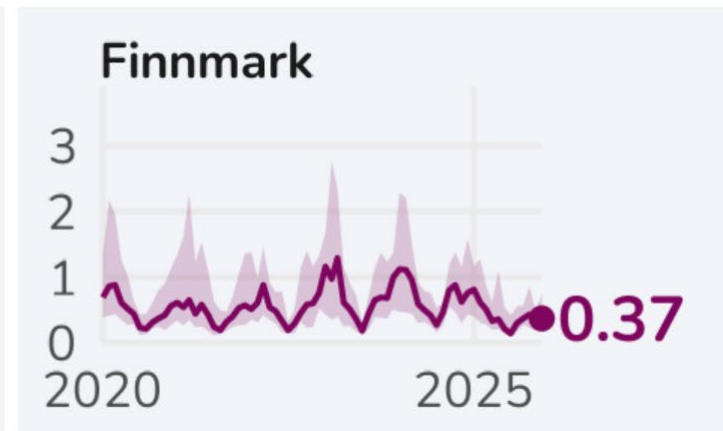
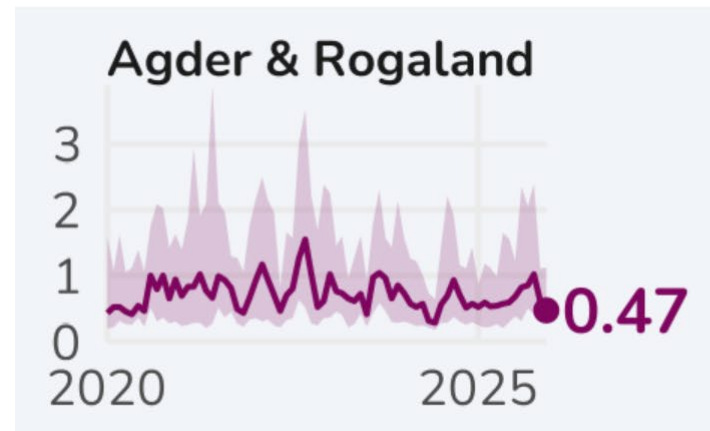
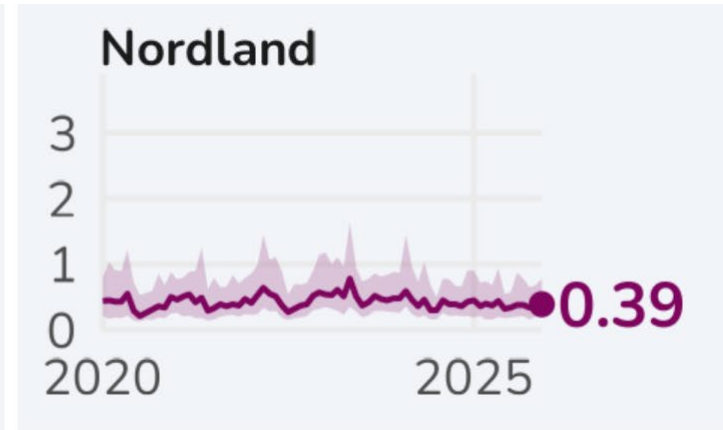
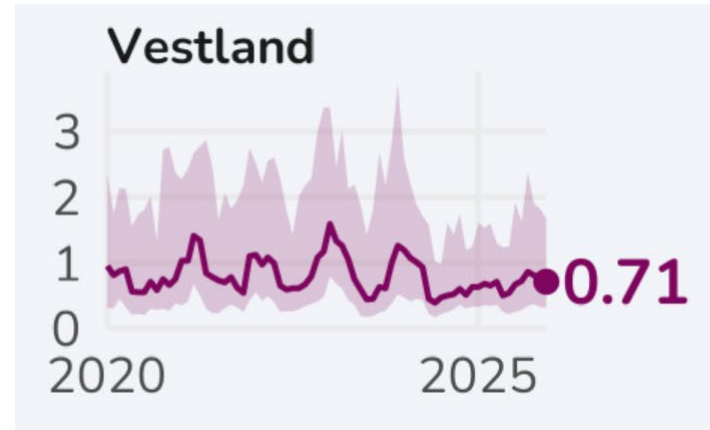
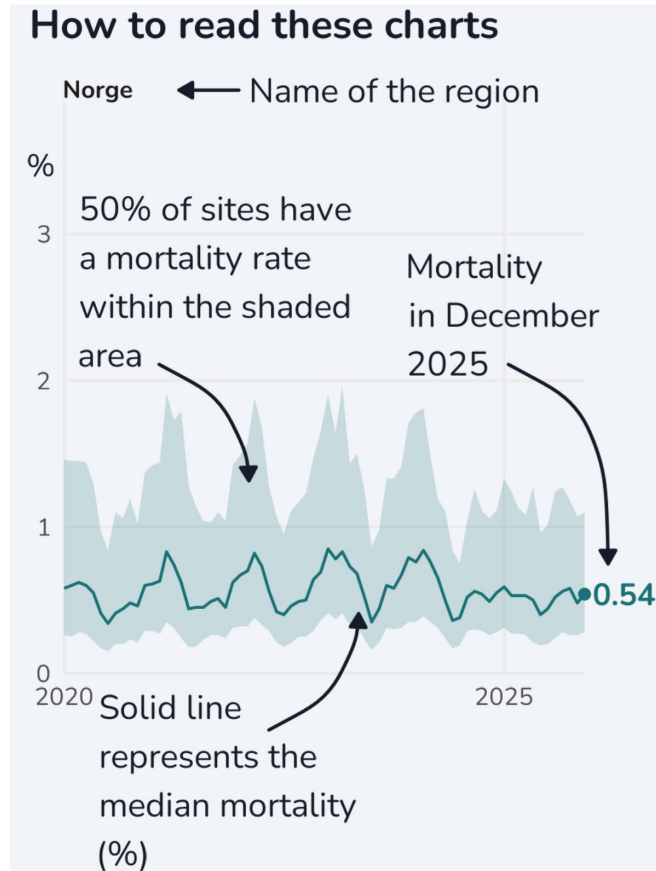
Source: Wes Patterson, LinkedIn post, April 1, 2026. Used with permission.



Trend Data



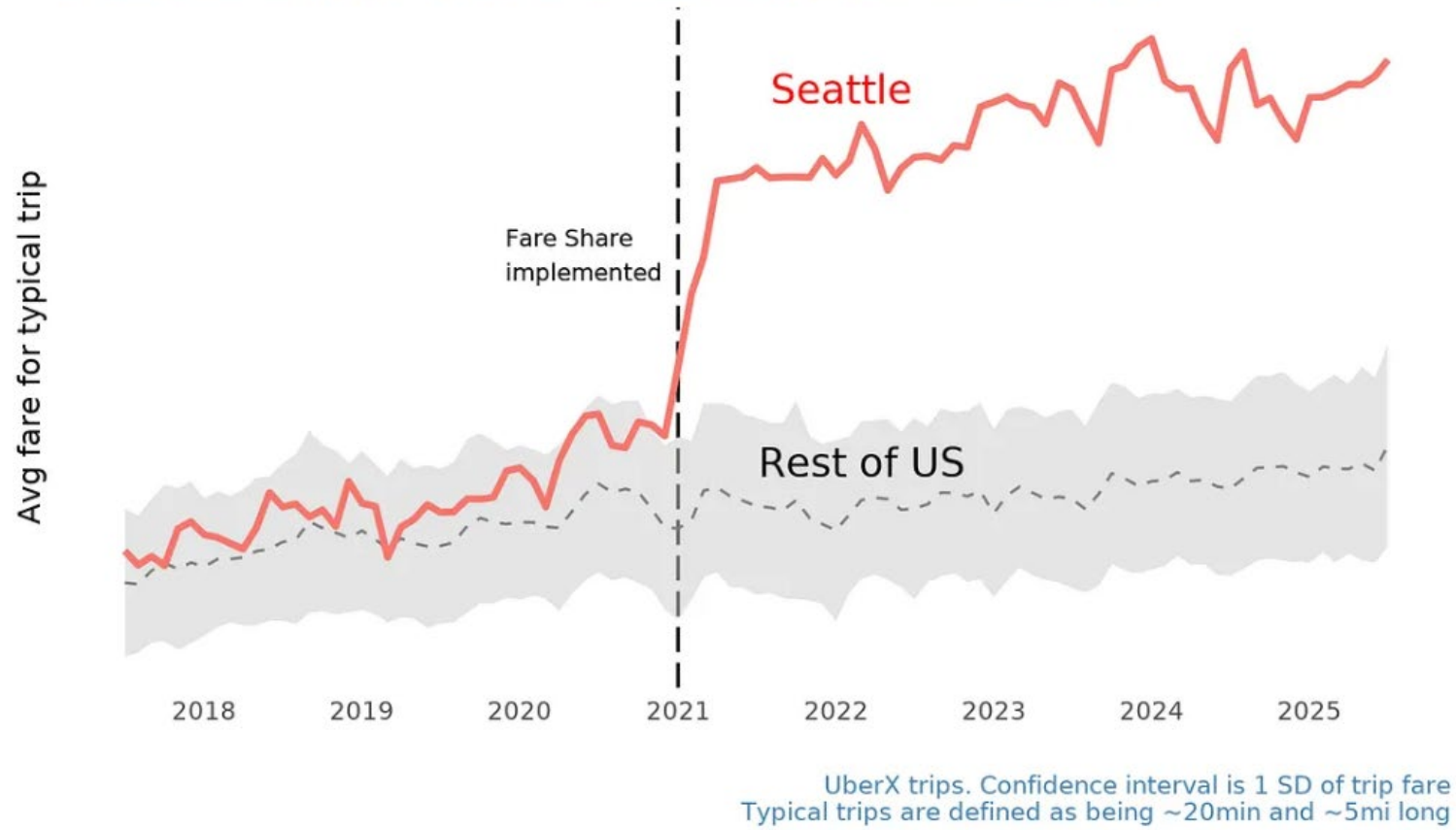
Examining trends: Salmon Mortality Rates





Examining trends: Uber Prices

Seattle trip prices are the highest in the US



<https://medium.com/uber-under-the-hood/seattles-ridesharing-laws-don-t-work-for-anyone-189e3117dc03>



Some Common Complications

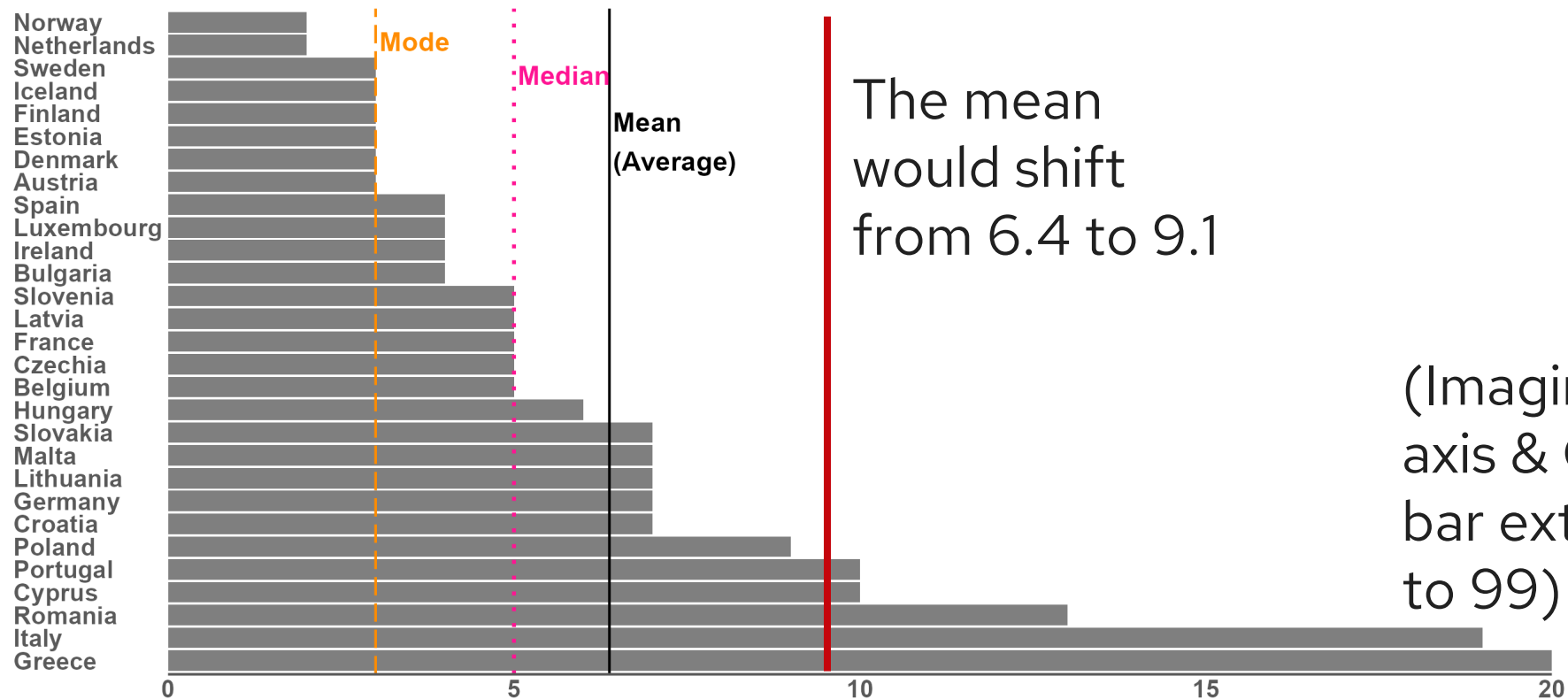


Mathematically simple, but not
uncomplicated

Outliers: If the value for Greece was 99...

Greece and Italy have the highest antimicrobial resistance (AMR) death rates.

This graph shows the estimated number of deaths attributable to AMR per 100,000 people in 29 countries in 2020.

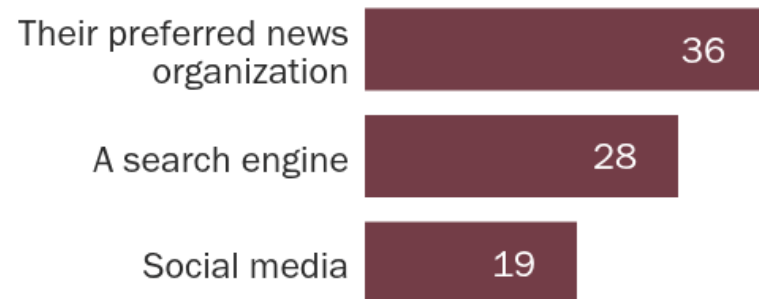




Summaries can hide subgroup differences

Americans turn to news organizations, search engines and social media to learn more about breaking news

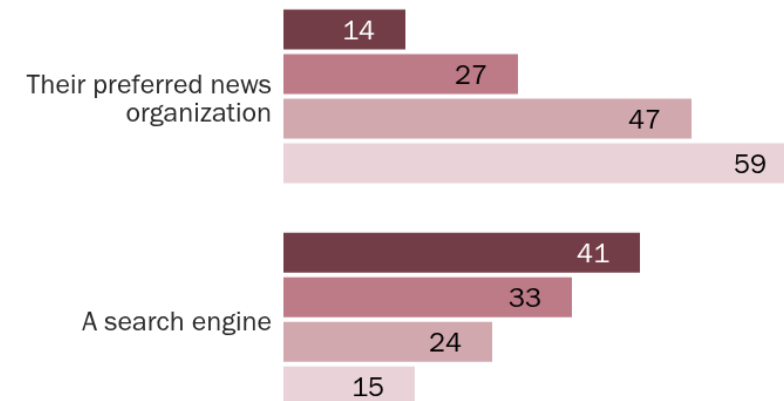
% of U.S. adults who typically go to ___ first to get more information when a breaking news event happens



Where Americans go for breaking news varies by age

% of U.S. adults who typically go to ___ first to get more information when a breaking news event happens

■ Ages 18-29 ■ 30-49 ■ 50-64 ■ 65+

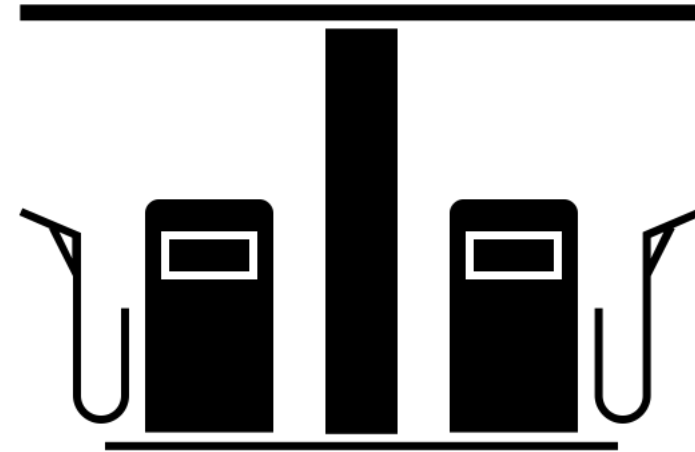


Source: Pew Research Center survey of U.S. adults conducted Dec. 8-14, 2025.

PEW-KNIGHT INITIATIVE

Ecological fallacy(ies)

- My Friend, at the Costco fuel pump:
"I hate it when I get behind a man. I'm 37 seconds faster, on average, getting gas than men."
- Me: 🙄
- Inner Analyst:
 - Sooo many questions...



Created by Hoeda
from Noun Project



Are Two (or more) Things Related?



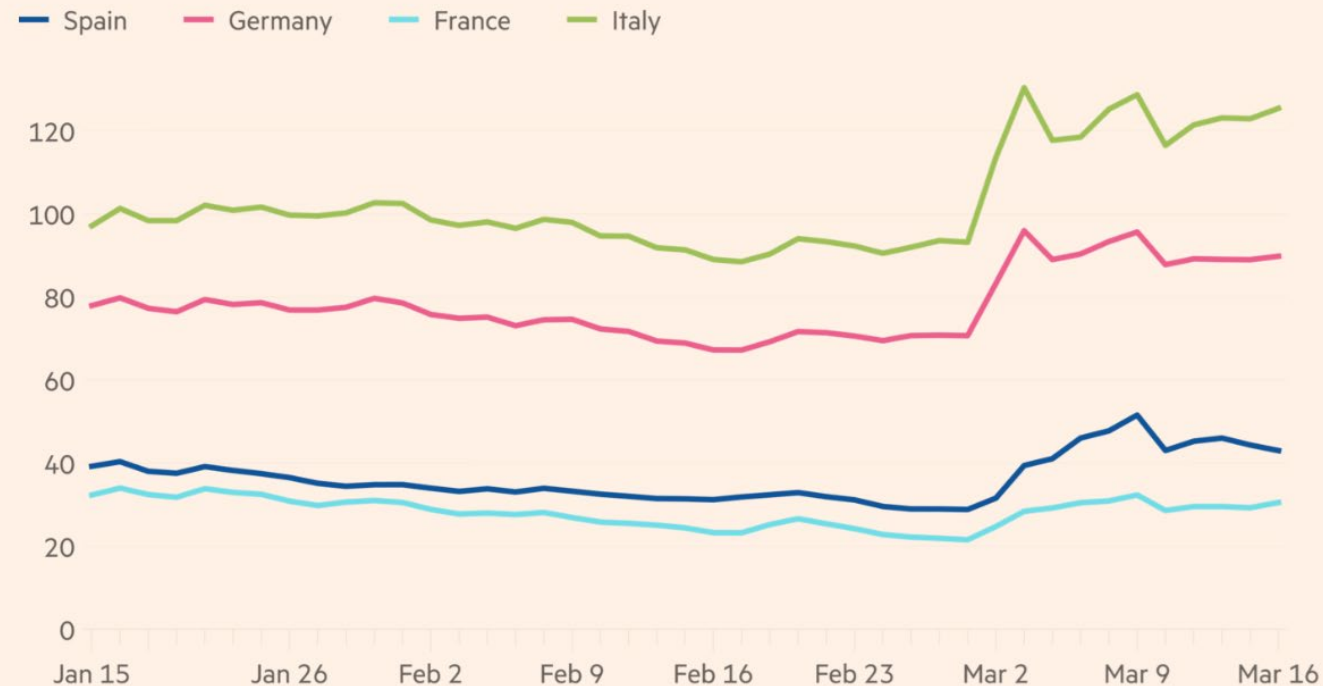
Underlies questions of group differences

If we're lucky, we don't need statistics to answer questions about relationships



A place in the sun

Expected electricity prices for the second quarter of the year (€/MWh)



Source: Bloomberg

<https://www.ft.com/content/19f2ee15-dc86-4964-b23f-d644b18a70a1?syn-25a6b1a6=1>



Scatterplot



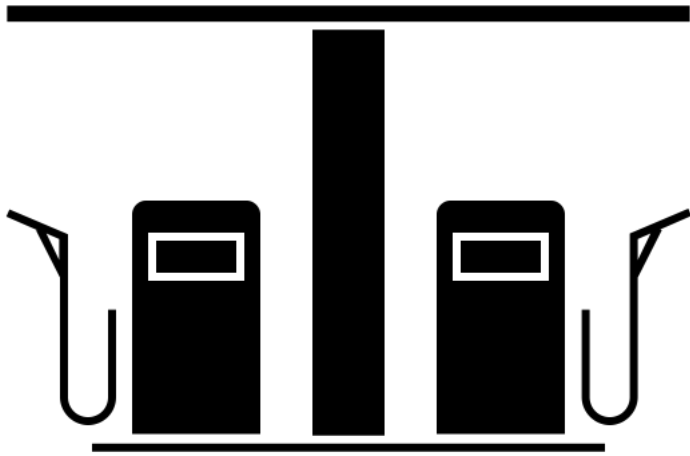
<https://www.the74million.org/article/which-school-districts-do-the-best-job-of-teaching-kids-to-read/>



Most everyday (potential) relationships are less clear

- Must sort out “signal” (a true association) vs. “noise” (an apparent association that is coincidental)
- Frequentist inferential or Bayesian statistics, or simulations
- Requires careful attention to how cases are included, and potential confounding variables

Returning to the fuel pump situation...



Created by Hoeda
from Noun Project

- How many time trials for in/out group members, and how representative are the people?
- What is the variation in times?
- Has she accounted for size of vehicle?
- How to account for the tendency of some people to fill up at $\frac{1}{2}$ full, and others when the "Empty" light is flashing?

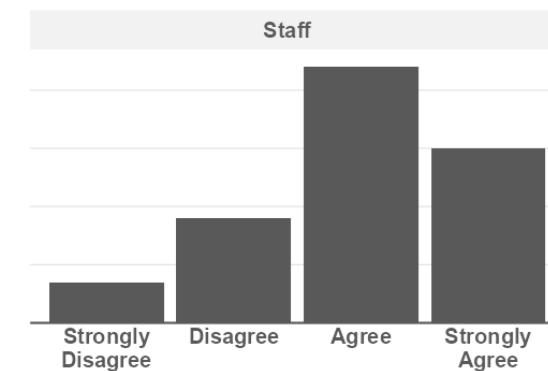
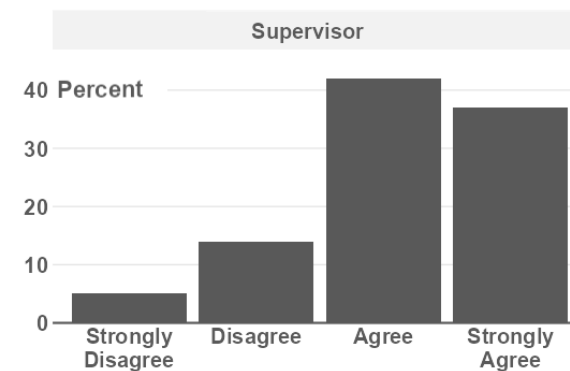
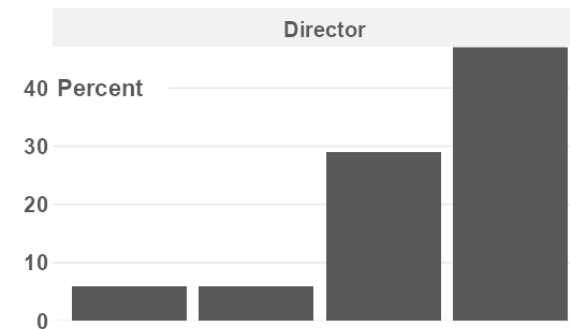


Closer to work related...



Do employees differ in overall job satisfaction by their role?

Role	N	Median	Mode
Director	17	3.5	Strongly Agree
Manager	107	3.0	Strongly Agree
Supervisor	166	3.0	Agree
Staff	1181	3.0	Agree





Simulation example: Null Worlds

Sample size:

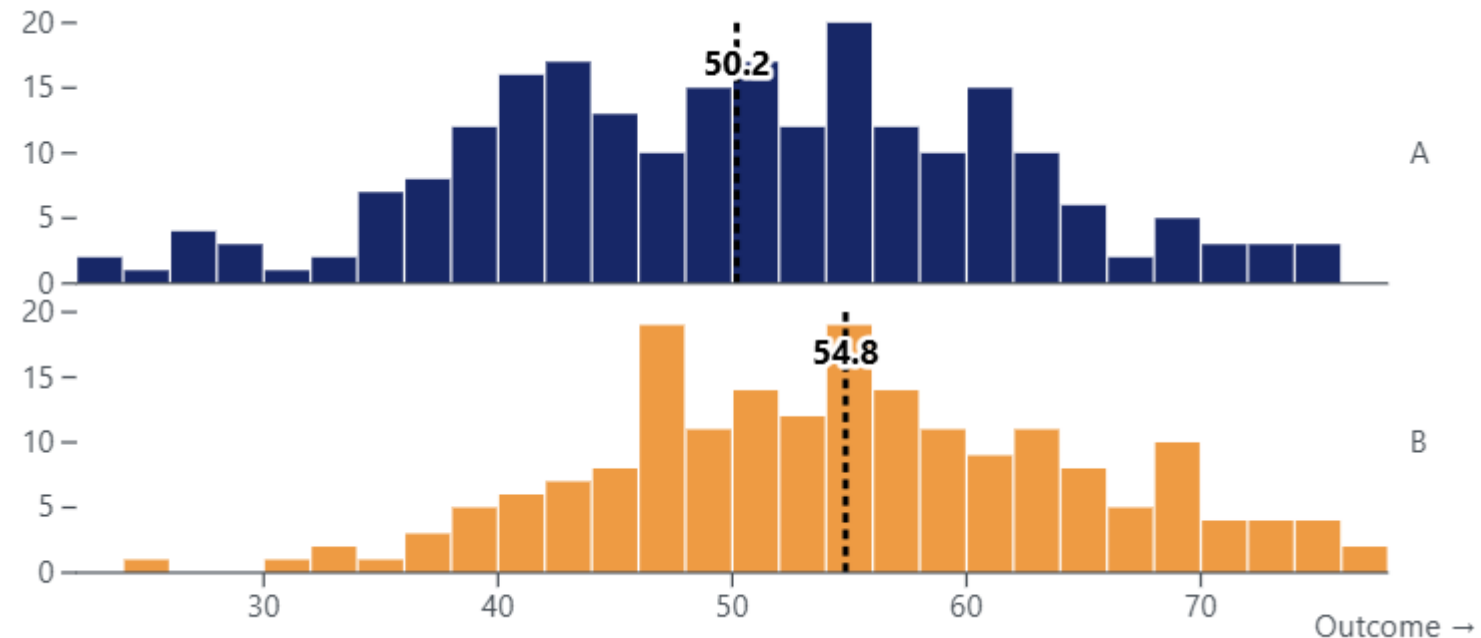
420

Difference:

5

Spread (σ):

10



<https://nullworlds.andrewheiss.com/diff-in-means.html>



Simulate “Null world”

Original data

	Group	Outcome
1	A	61.8
2	B	52
3	A	64.9
4	A	57.4
5	B	74.6
6	A	57.3
7	B	58.7
8	A	48.7

Shuffled data

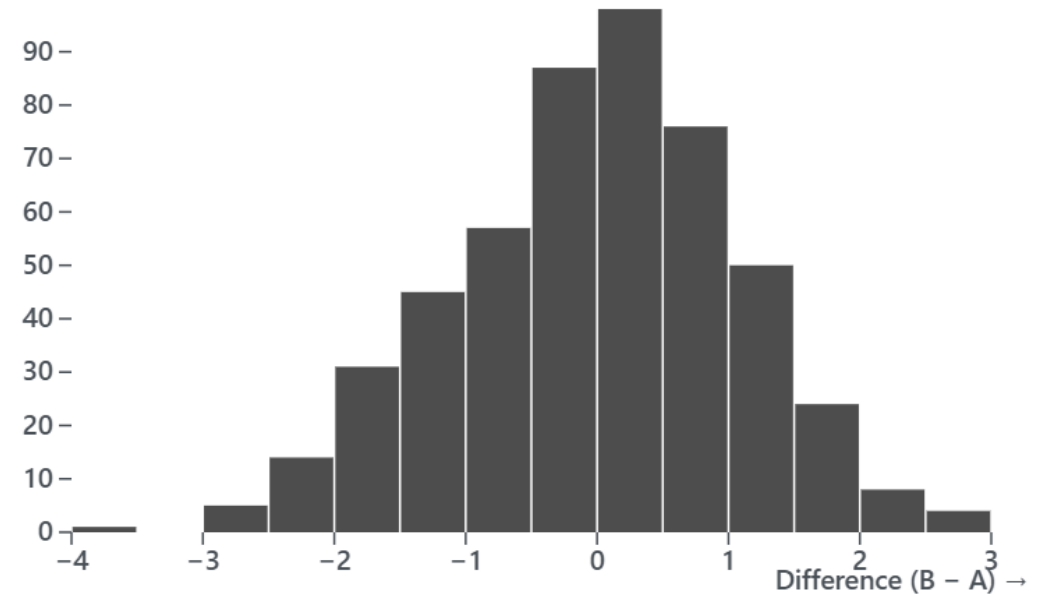
	Group	Outcome
1	A	61.8
2	A	52
3	A	64.9
4	B	57.4
5	B	74.6
6	A	57.3
7	A	58.7
8	B	48.7

Number of
simulations:

500

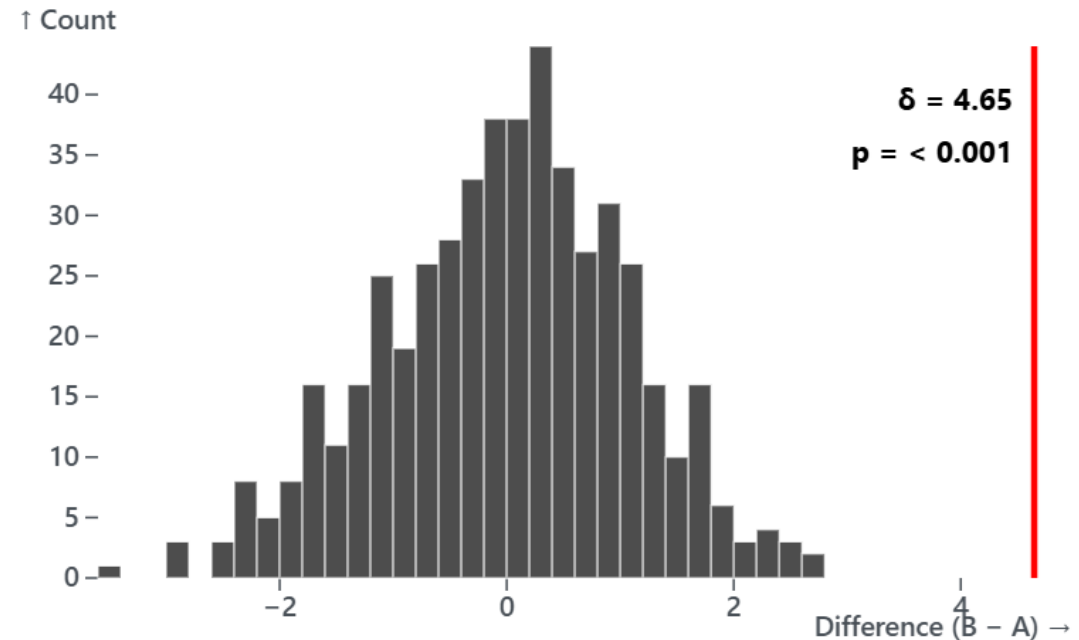


↑ Count



Put the observed value in the Null World

- Compute the probability of seeing a difference at least that large in a world where there is no difference in group average.
- p-value: fraction of times the simulated difference exceeds the observed difference



The p-value is < 0.001