

# Driving with Data



Arnold Barnett  
MIT Sloan

# Four Road Trips

- Murder
- Plane Crashes
- Airline Punctuality
- Political Polls

## How high is the risk of being murdered in American cities?

- “If you live in Detroit, the odds are 2000 to 1 that you will *not* be killed by one of your fellow citizens. Optimists searching for *perspective* in the city's homicide statistics insist that these odds are pretty good.”

---*New York Times*, 1/30/75; *italics* added

- The Times was reacting to the fact that there were **700** homicides in Detroit in 1974, at a time Detroit's population was **1.4 million**.

(700/1.4 million = 1/2000).

But why work out murder risk *per year*, as opposed to be per day, per month, or per decade?

A person born in Detroit who continues to live there is in danger of being murdered *throughout his life*. And if the risk is being sustained over a lifetime, that would seem the *natural time frame* over which to measure it.

## Back of Envelope Calculation:

Assume:

- Randomly chosen newborn would have lifespan of **70** years barring homicide
- Murder risk is **1 in 2000** per year
- Then P(**not** die of homicide)  $\approx \left(\frac{1999}{2000}\right)^{70}$

- $P(\text{not die of homicide}) \approx \left(\frac{1999}{2000}\right)^{70} = \left(1 - \frac{1}{2000}\right)^{70}$
- Do you recall the Taylor series expansion of  $(1 - x)^n$  ?  
(I didn't think you did.)
- Anyway,  $(1 - x)^n \approx 1 - nx$  when  $nx$  is small
- So.  $P(\text{not die of homicide}) \approx 1 - 70 * \left(\frac{1}{2000}\right) = 1 - \frac{70}{2000} = \frac{1930}{2000}$
- $P(\text{die of homicide}) \approx \frac{70}{2000} = 1 \text{ in } 29$
- So much for the “optimists searching for perspective”

“I have been vilified. I have been crucified.

I have even been criticized.”

-----the original Mayor Daley

## Criticisms:

(1) Murder victimization rates vary considerably with **age**, so this simple model is **highly misleading**.

*Actuality, it's not.*

(2) Murder risk varies substantially by gender and race, so this aggregate statistic is **useless**.

*The **overall average risk** is meaningful, and one can perform further calculations about variations around the mean.*

(3) Who knows what will happen over the next 70 years?

*The statistic offers a **snapshot** of what will happen if the status quo prevails. The same objection applies to life expectancy or any number of actuarial projections.*



(4) Why present a statistic that only serves to frighten people?

*Making clear that homicide is a huge problem is important even if it's frightening. Should one **suppress** facts because they are unpleasant?*

(5) “I’m not convinced that the lifetime victimization probabilities are superior to per capita murder rates,... the use of lifetime murder risk does not resonate “

***You ignorant snake.***

# Are things better now?

- Well, consider these data for 50 of the largest American cities:

	Male			Female	
	<u>2019</u>	<u>2020</u>		<u>2019</u>	<u>2020</u>
Black	1 in 22	1 in 16		1 in 161	1 in 122
Non-Black	1 in 153	1 in 115		1 in 529	1 in 462

# Aviation Safety: Dark Clouds Ahead?



How should we measure aviation safety?

The answer to this question is not obvious.

*An e-mail message:*

*“My name is L.S. I would like to know if you, as an expert in aviation safety, fly regularly.”*

“You see, I stopped flying about a year ago and this has affected my life in a significant matter. Just one last question: **what are the odds** of me **(sic)** dying in a plane crash?”

Given that a passenger's greatest fear is of being killed in a plane crash, there is **a natural interest** in statistics about the likelihood of that outcome.

What about the simple ratio of passengers killed to passengers carried?

It's not perfect, but it's better than more "sophisticated" alternatives.

The statistic **passengers killed divided by passengers carried** answers the question:

If we choose **one boarding pass at random** from all those used by the passengers of interest (e.g. Brazilian domestic air travelers over 1990-99), what is the probability that its owner **did not survive** their flight?



The **time-trend** in worldwide passenger death risk is a **joy to behold**:

**Worldwide Death Risk per Boarding for Five Decades from 1968 to 2017**

Decade

Death Risk per Boarding

**1968-77**

**1 in 350,000**

**1978-87**

**1 in 750,000**

**1988-97**

**1 in 1.3 million**

**1998-2007**

**1 in 2.7 million**

**2008-17**

**1 in 8.3 million**

Over 1968-2017, airline safety followed something of a **Moore's Law**, with death risk falling by a factor of two every decade.

But what about the most recent period, from 2018 to 2022?

For 2018-22, the worldwide statistic for death risk per boarding was **1 in 13.4 million**.

Compared to the 2008-17 figure, the improvement (1 in **13.4** million vs. 1 in **7.9** million) was a factor of **1.70**.

*Is the pace of improvement slackening?*

No! That factor concerns a follow-up period of **five** years rather than ten. And an improvement of a factor of 1.7 over five years corresponds to a factor of  $1.7^2 = \mathbf{2.89}$  per decade.

However, much as the center of mass of a circular doughnut is the center of the hole (**where there is no mass**):

- There was **no nation in the world** at which the death risk per boarding was at the worldwide average for 2008-17 of 1 in 8.3 million.
- On the contrary, there was (and has always been) **considerable diversity** in death risk across nations.

Among the **world's safest nations** in passenger aviation, death risk per boarding over 2008-22 has been about **1 in 50 million**.

What are those nations?

US, UK, EU, Canada/Australia/New Zealand,  
Israel, Japan, **China**

At a death risk of **1 in 50 million** per flight, a passenger who took one flight every day could on average travel for **137,000 years** before succumbing to a fatal crash.

When traveling from one of the **safest** aviation nations to that is **less safe**, should one opt for an airline with its home office in the **safer** group?

Well, have you heard of the **Ecological Fallacy**?

# Death Risk per Boarding **Between** Safest-Nation Cities and Cities in Other Countries, 2000-19:

Safest-Nation Carriers: **1 in 2 million**

Other Carriers: **1 in 2 million**



## Rule of Thumb:

When two airlines fly the same route nonstop, **very rarely is there a case related to safety** to prefer one of them over the other.

# What about Covid-19, and the deaths caused by transmissions during flight?

- Considering the period March 2020-May 2023 and Europe and the US, a conservative estimate of the number of **deaths** tied to in-flight infections would be **500**.
- If 500 Covid-19 deaths are added to those arising from air crashes, the mortality risk of aviation in 2018-22 was **only slightly lower** than that in 2008-17. In the US and Europe, the death risk was **higher** in 2018-22.
- However, if the Covid-19 nightmare is behind us, then the mortality risk of air travel continues at its **headlong pace** towards a rate of zero.

# Justice for jetBlue?



jetBlue has been widely criticized for its poor **on-time** record. That record is a large part of the reason that the *Wall Street Journal* ranked jetBlue 9<sup>th</sup> out of 9 among US airlines (1/18/23).

-

For 2022, The US DOT on-time statistics for US airlines were:

<u>Airline</u>	<u>Rate</u>
Hawaiian	83.9 (%)
Delta	80.7
United	79.2
Alaska	78.6
American	75.3
Spirit	73.4
Allegiant	69.2
Frontier	66.5
jetBlue	64.9
Southwest	64.8

But jetBlue contends that its bad reputation for punctuality is **unfair**, because it does most of its flying in the northeast region centered in New York City where everyone does badly.

Is it possible to come up with a **fair** way for rating on-time performance, which might exonerate jetBlue if the facts warrant?

*(I've got a plan for that.)*

Let's see. Suppose we consider the **ten** largest US airlines and the **25** busiest US airports and:

- Measure an airline's punctuality by the **simple average** of its on-time rates over the 25 airports.
- Under this criterion, an airport like JFK would count as **4%** (1/25) of an airline's score, meaning that **all** airlines would be exposed to JFK's troubles to the same extent.

That was easy. **Too easy.**

- The problem with the 25-airport rule is that most of the ten largest airlines do not fly to all 25 airports. For example, Southwest does not fly to Newark, or Frontier to O'Hare.
- If a given airline tends to avoid difficult (lousy) airports, its overall score could be **artificially good**.
- Can we avoid this problem?

*I've got a plan for that.*



Work out **normalized** scores for each airline and the big airports it serves.

- A score of zero would mean that the airline **matches** the average on-time rate for the airport. A score of **+5** would mean it did five points better; a score of **-4** would mean it did four points worse.

## Normalized On-Time Rates for US Airlines Under Equal Weighting of US Airports

<u>Airline</u>	<u>Rate</u>
• Delta	+10.3 (%)
• Hawaiian	+7.2
• United	+5.8
• Alaska	+5.3
• American	+1.3
• Spirit	-0.9
• Allegiant	-4.9
• Frontier	-8.6
• Southwest	-10.0
• jetBlue	-12.7

That was easy. Too easy. *(I've got a plan ..)*

- Much as a difficult airport can **besmirch** the punctuality records of the airlines that fly there, tardy (lousy) airlines can besmirch the punctuality record of an **airport**.
- So why not apply a **correction factor**, to estimate what the airport's on-time record would be if it were served by **average** airlines? Let me explain.

- With that correction factor, now we have a **second** set of normalized scores.
- We're almost there, but....

After several rounds of successive approximations, our “algorithm” converges, and we reach:

<u>Airline</u>	<u>Normalized On-Time Rate</u>
• Delta	+6.9(%)
• United	+6.2
• Hawaiian	+3.8
• American	+3.3
• Alaska	+3.0
• Spirit	-0.6
• Southwest	-5.7
• Frontier	-9.0
• jetBlue	-9.3
• Allegiant	-10.3

- So jetBlue's overall promptness record is not really improved when we act to eliminate the bias it allegedly faces.
- Still, its "New York story" has **surface plausibility**, so it can help in marketing.



# The Polls and the US Presidential Election of 2020....And of 2024



Arnold Barnett and Arnaud Sarfati

In 2020 as in 2016, there was widespread frustration concerning the performance of polls about the US presidential election:

- Former Secretary of State James Baker declared “this time, we were promised that pollsters would get it right. **They didn’t.**”
- *The New York Times* ran an article titled “2016 Dealt a Blow to Polling. **Did 2020 Kill It?**”
- A commentator in Yahoo! Finance thought that “the biggest election takeaway is the **absolutely massive** failure of polling.”



- The prestigious American Association for Public Opinion Research concluded that the presidential polls fared **worse** in 2020 than in 2016.
- This despite the fact that, unlike the polls in 2016, those in 2020 correctly identified the winner of the election.

- But perhaps these negative assessments are overwrought, because:
- They ignore the possibility that, in the hands of a **sophisticated aggregator** like Nate Silver's **538**, the shortcomings of individual polls can be mitigated and accurate forecasts produced.
- If so, the polls might usefully be treated as raw materials in an overall process that is **highly successful**.

How well did 538 perform in the 2020 presidential election race?

In evaluating 538's predictions against actual outcomes , we should note one of its greatest strengths: its extensive use of **probability models**.

In each American state in 2020, 538 estimated:

- The **probability** that Trump would carry the state (a Biden win is the complementary event)
- Trump's **share** of the popular vote (versus Biden and third-party candidates)
- A margin of error for Trump's **vote share**

How well did 538's forecasts perform?

With respect to presidential elections, there are **51** American states rather than 50. The District of Columbia also counts.

(Actually, the congressional districts of Maine and Nebraska also play a role.)

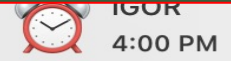
To predict the winner in states like California and Alabama is **belaboring the obvious**. But what about 538's win/loss performance in the **swing states** where the 2020 election would be decided?

- According to 538, there were nine “tipping point” states:

- Arizona      Florida      Georgia      Michigan
- Minnesota      North Carolina      Nevada      Pennsylvania      Wisconsin

Let's say more about them. In particular, what is 538's **probability distribution** for the number of tipping point states that Trump would carry in 2020?

Table 1: *FiveThirtyEight* 's Nine Tipping States in the 2020 Presidential



<u>State</u>	<u>Electoral Votes<sup>a</sup></u>	<u>Five Thirty Eight's Estimated Probability of Trump Victory</u>
Arizona	11	32%
Florida	29	31%
Georgia	16	42%
Michigan	16	5%
Minnesota	10	4%
North Carolina	15	36%
Nevada	6	12%
Pennsylvania	20	16%
Wisconsin	10	6%
<i>Total</i>	133	

- Taking everything into account, we wind up with the following 538-related probability distribution for Trump's number of tipping state victories:

$$S = \begin{cases} 0 \text{ w.p. } .094 \\ 1 \text{ w.p. } .328 \\ 2 \text{ w.p. } .380 \\ 3 \text{ w.p. } .118 \\ 4 \text{ w.p. } .041 \\ 5 \text{ w.p. } .018 \\ 6 \text{ w.p. } .010 \\ 7 \text{ w.p. } .005 \\ 8 \text{ w.p. } .004 \\ 9 \text{ w.p. } .002 \end{cases}$$

w.p. = with probability

$$S = \begin{cases} 0 & \text{w.p. .094} \\ 1 & \text{w.p. .328} \\ 2 & \text{w.p. .380} \\ 3 & \text{w.p. .118} \\ 4 & \text{w.p. .041} \\ 5 & \text{w.p. .018} \\ 6 & \text{w.p. .010} \\ 7 & \text{w.p. .005} \\ 8 & \text{w.p. .004} \\ 9 & \text{w.p. .002} \end{cases}$$

w.p. = with probability

- In the event, Trump carried **two** of the nine tipping states: **Florida and North Carolina**. That outcome is at the mode and median of 538's distribution of  $S$ , and as close as possible to its mean of 1.82.
- In other words, 538 did **extremely well**. It is superficial to say that 538 "erred" in Florida and North Carolina, when it anticipated that Trump would carry two of the tipping point states.



What about 538's estimates of Trump's vote shares by state?

Here 538 fell a bit short. In fact, in **all but four** of the 51 states, 538's point estimate of Trump's vote share fell below Trump's actual achievement. (The exceptions were Alaska, Colorado, Maryland, and Louisiana.).

That doesn't look great. If you continue to assert that a coin is fair when it comes up heads 47 times out of 51, you should go at once to **urgent care**.

538's estimates of Trump's vote share were too low by **an average of two percentage points**.

- In summary, 538 did an **excellent** job in win/loss forecasts, but suffered a **modest** downward bias in estimating Trump's vote shares. Overall, pretty impressive, given the difficulties in polling in 2020, which were exacerbated by the pandemic.

But how badly did the individual statewide polls fare **before** 538 worked its magic?

One simple test of 538's achievements is to compare how well it did compared to *Real Clear Politics*, which takes a **simple average** of various polls close before the election. Real Clear Politics is so **crude** that 538 should do far better if it has a great secret sauce.

- We compare 538 and RCP in 14 swing states (the tipping states according to 538 and/or the toss-up states according to RCP). Here are the results:

Table 2: Trump's Actual 2020 Vote Shares and His Projected Vote Shares in 14 Swing States, For FiveThirtyEight and Real Clear Politics (RCP)

<u>State</u>	<u>Actual Trump Vote Share</u>	<u>Projected Trump Vote Share:</u>	
		<u>Five Thirty Eight<sup>a,b</sup></u>	<u>Real Clear Politics<sup>c</sup></u>
Arizona	49.06(%)	48.1 (-0.96)	47.77 (-1.29)
Georgia	49.24	49.2 (-.04)	48.93 (-0.31)
Florida	51.22	48.4 (-2.82)	49.10 (-2.12)
Iowa	53.08	50.0 (-3.08)	49.37 (-3.71)
Michigan	47.84	45.5 (-2.34)	46.54 (-1.30)
Minnesota	45.31	44.8 (-0.51)	45.53 (0.21)
Nevada	47.67	46.2 (-1.47)	47.23 (-0.44)
North Carolina	49.93	48.8 (-1.13)	48.71 (-1.22)
Ohio	53.27	49.8 (-3.47)	48.63 (-4.63)
Pennsylvania	48.84	47.3 (-1.54)	48.42 (-0.42)
Texas	52.06	50.3 (-1.76)	49.03 (-3.03)
Wisconsin	48.82	45.5 (-3.32)	45.08 (-3.74)
Maine, 2 <sup>nd</sup> CD <sup>d</sup>	52.26	47.8 (-4.46)	46.56 (-5.70)
Nebraska, 2 <sup>nd</sup> CD <sup>d</sup>	45.45	48.4 (2.95)	46.10. (0.65)
Average Forecast Error		-1.71	-1.93
Average Absolute Forecast Error		2.13	2.05

- In the swing states, 538 and RCP performed **about equally!**
- However, RCP is itself an aggregator of polls, much like 538. Is it possible that averaging **masks** huge errors by individual polls **of opposite sign?**
- To investigate that issue, we turn to **mean absolute error** for the individual polls in a given swing state, under which positive and negative errors **cannot cancel** each other out?

- Well, here's what happened with the local polls in various states.

Table 3: Average Absolute Forecast Error and Average Margin of Random Sampling Error for Trump's 2020 Vote Share, Among Key Local Polls in 14 Swing States<sup>a</sup>

State	Average Forecast Error <sup>b</sup>	Average Absolute Forecast Error <sup>c</sup>	Average Margin of Random Sampling Error <sup>d</sup>
Arizona	-1.29(%)	1.65	3.69
Florida	-2.12	2.72	3.54
Georgia	-0.31	1.49	4.00
Iowa	-3.71	3.71	3.82
Michigan	-1.30 (%)	2.10	3.96
Minnesota	0.21	0.54	3.41
Nevada	-0.44	1.50	3.45
N. Carolina	-1.22	1.25	3.67
Ohio	-4.63	4.63	3.33
Pennsylvania	-0.42	1.48	3.84
Texas	-3.03	3.03	3.39
Wisconsin	-3.74	3.74	3.63
Maine CD2	-5.70	5.70	5.59
Nebraska CD2	0.65	1.87	4.30

- The mean absolute error was about **2.5** percentage points, slightly higher than 538 and RCP but **well within** the four percentage point margins of error based on sampling fluctuations alone!
- Put less politely, the **value added** by 538 was **surprisingly close to zero**. The local polls did very well in their own right.

# How Does It All Add Up?

- The assertion that the presidential polls failed grievously in 2020 falls between highly exaggerated and wholly false.
- The polls did well when taken at face value. It was not necessary to perform statistical plastic surgery via 538 to get reliable readings about statewide outcomes.
- Looking ahead to 2024, the 2020 outcomes suggest the rule of thumb:
  - **Believe the polls.**
  - If they suggest a close race between Biden and Trump, believe that the race actually is close.