

Problems in Identifying Causality in Observational Data

Ray Poynter

The Future Place / Nottingham Trent University

Presentation delivered 27 October BigSurv18 Conference, Barcelona, Spain

Observational Data



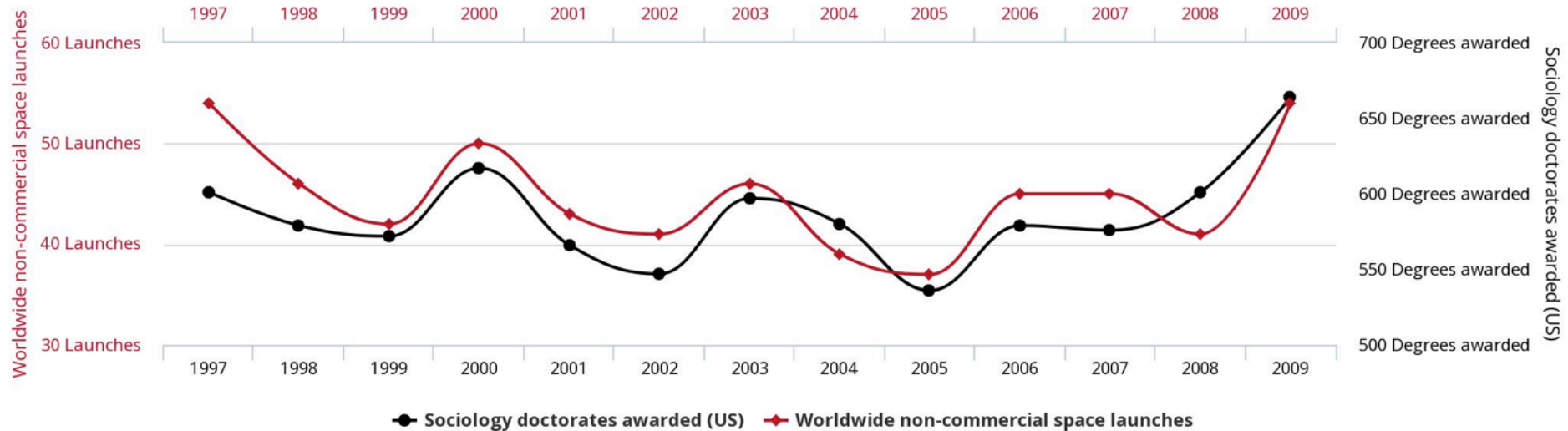
- Most big data (e.g. bank records & social media)
- Census or sample (e.g. all phone calls or a sample of them)
- Objective or subjective (e.g. till receipts or ethnography)
- Structured or unstructured (e.g. phone records or images uploaded to SM)
- Behavioural or motivational (e.g. loyalty cards or motivations from facial coding)
- Naturally occurring or from experiments
- Observational only or with questions (e.g. an advertising test using biometrics as an observational source)

Potential Problems

- Spurious correlation
- When observation tells you the wrong thing
- Coverage error
- Confusing cause and effect
- Ignoring the true driver
- Multicollinearity
- Complex and/or chaotic relationships
- Observer effect
- Survivorship bias
- Feedback loops between cause and effect
- Measurement effects
- Confusing influence and homophily
- Not explaining the why
- Things that have not happened, yet

Spurious Correlations

Worldwide non-commercial space launches
correlates with
Sociology doctorates awarded (US)

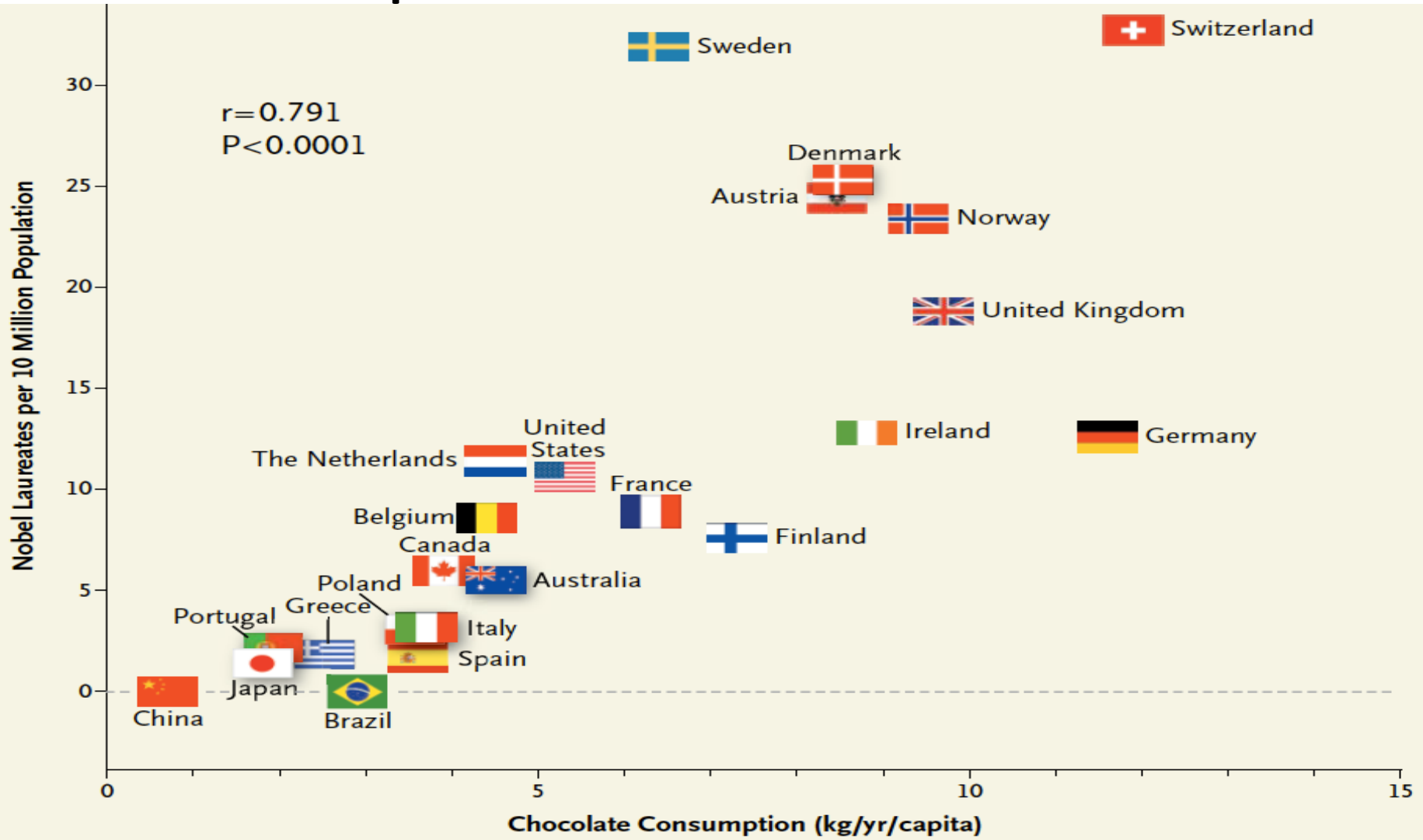


tylervigen.com

As data becomes large “the overwhelming majority of correlations are spurious.”

Calude & Longo, *The Deluge of Spurious Correlations in Big Data*, 2017, Foundations of Science

Spurious Correlations



When observation tells you the wrong thing

The New York Times

SCIENCE

Hormone Studies: What Went Wrong?

By GINA KOLATA APRIL 22, 2003

For nearly nine months, doctors and researchers have been struggling with an intractable problem: how could two large high-quality studies come to diametrically different conclusions about menopause, hormone therapy and heart disease?

The question arose in July, when scientists saw data from a large federal study called the Women's Health Initiative, which was ended early when it became clear that a widely used hormone-replacement drug, Prempro, had risks, including heart attacks, that exceeded its benefits.

When HRT was 1st assessed (Nurses Health Study – large observational study in the USA), seemed to protect the heart.

Doctors were recommended to prescribe it more widely.

Women's Health Initiative (gold standard controlled experiment) – suggested it was slightly bad for the heart.

Why?

Women receiving HRT were systematically healthier and wealthier.

Combinatorial Effects

Region A

- T1, sales = 100
- T2, TV, sales = 110
- T3, TV & Twitter, sales = 130

Region B

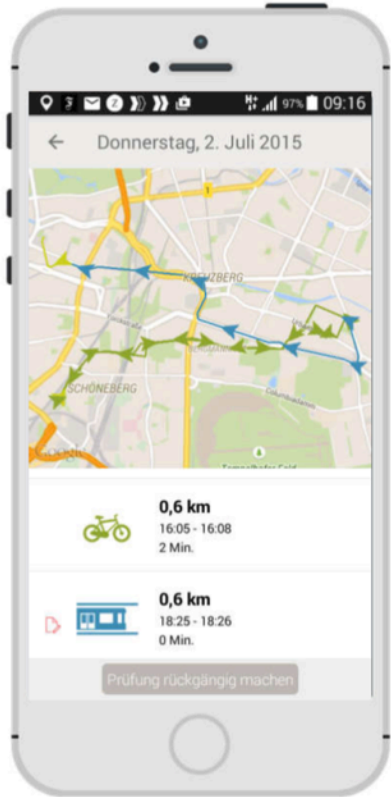
- T1, sales = 100
- T2, Twitter, sales = 110
- T3, TV & Twitter, sales = 130

Region C

- T1, sales = 100
- T2, sales = 105
- T3, sales = 110

The counterfactual = some growth would have happened anyway.

Coverage Error



Dana Gruschwitz & Dr. Robert Schönduwe,
ESRA, 2017, Lisbon, Portugal

Long-standing transport study in Germany.

People have been using PAPI and CAPI to capture journeys – memory based.

Trial with mobiles, to automatically capture information.

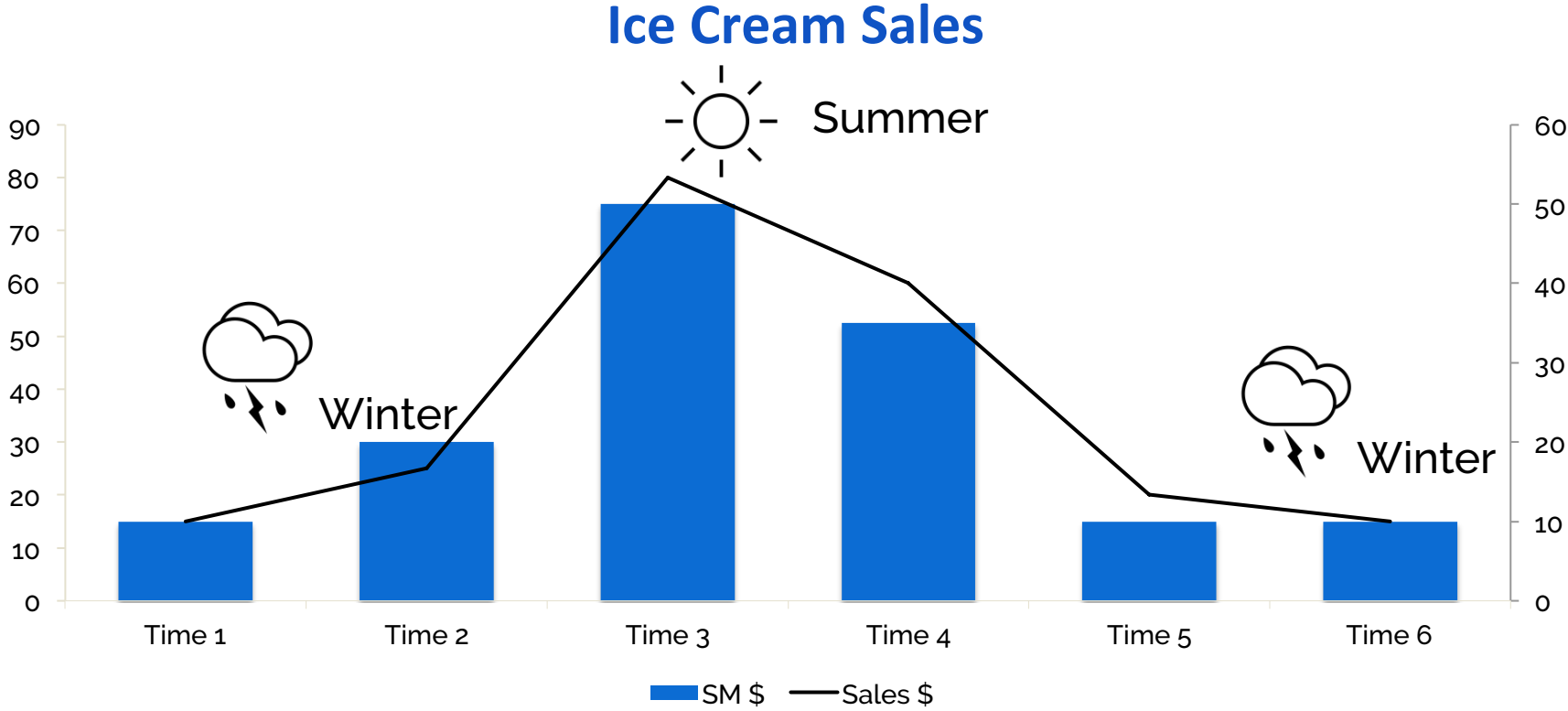
Less ‘heaping’ of the distances and times 😊

But 16% fewer journeys (11% less distance, 18% fewer minutes) were recorded 😞

Why?

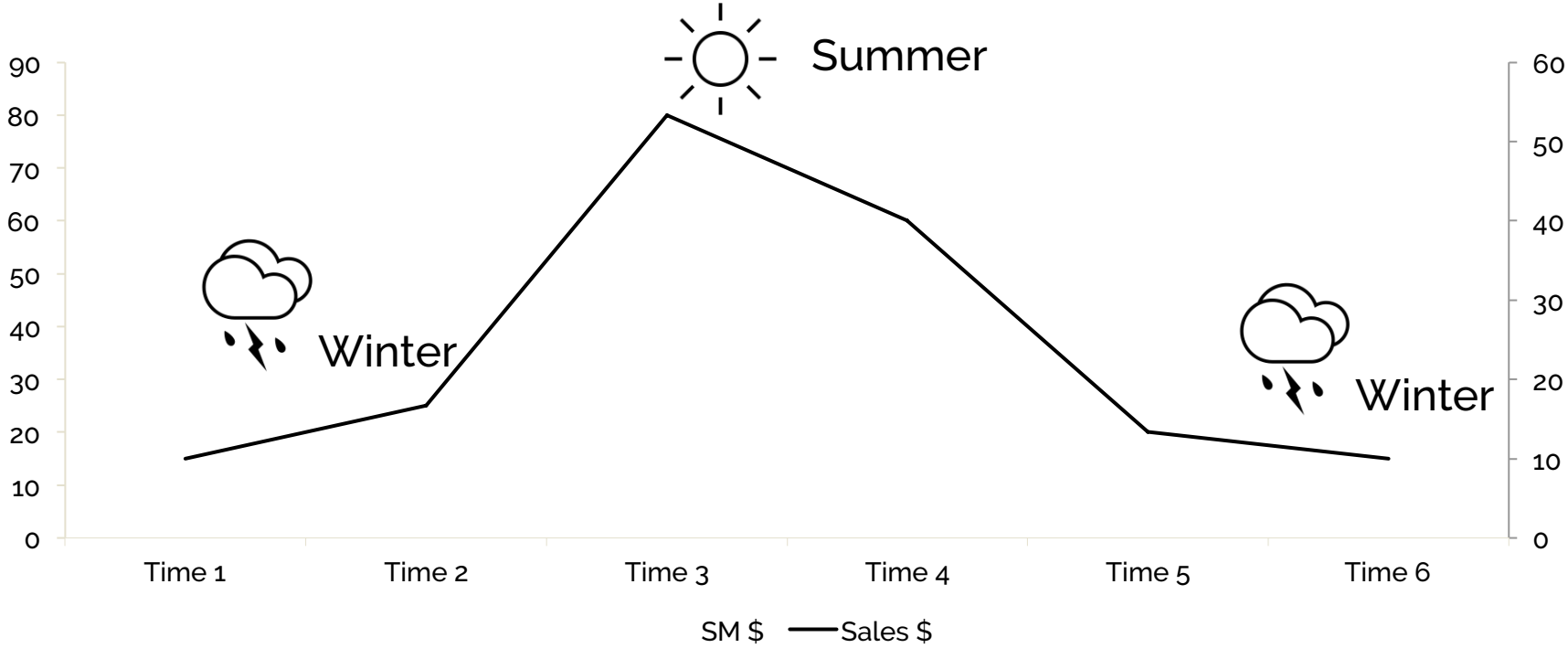
Phone app turned itself off when people’s phone battery reached 20%

Ignoring the True Driver



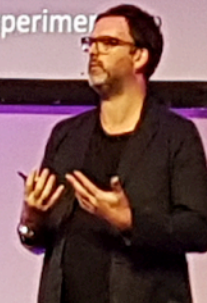
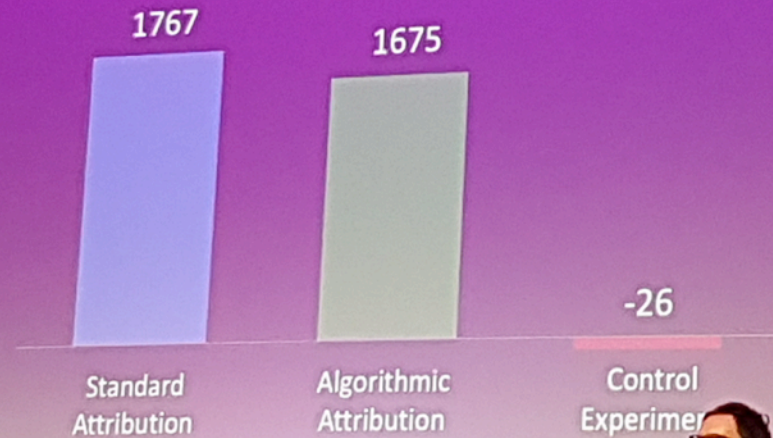
Identifying the Counterfactual

Ice Cream Sales



ATTRIBUTION IS NOT CAUSATION

19 weeks of Retargeting Experiment Data

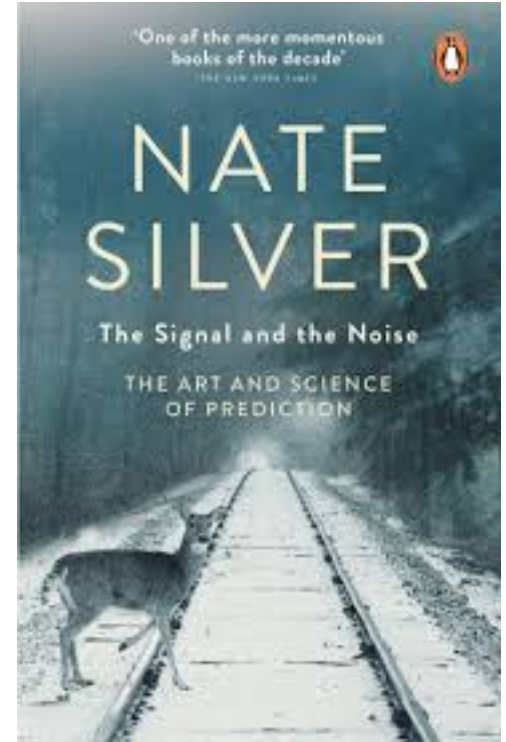


ESOMAR

Brent Smart, CMO IAG (Australian insurance company) – ESOMAR APAC May 2018

Complex and/or Chaotic Relationships

- Baseball, the whole is the sum of the parts
- Weather, complex but models improving 8-day forecast as good as simple models
- Earthquakes, no progress to date, maybe there will be no progress ever



Observer Effects



Watching / measuring behaviour can change behaviour.

UK RAC study of speed cameras, 2013, found 27% reduction in fatal and serious collisions.

Note, nobody was deliberately crashing, it was the underlying behaviour that changed.

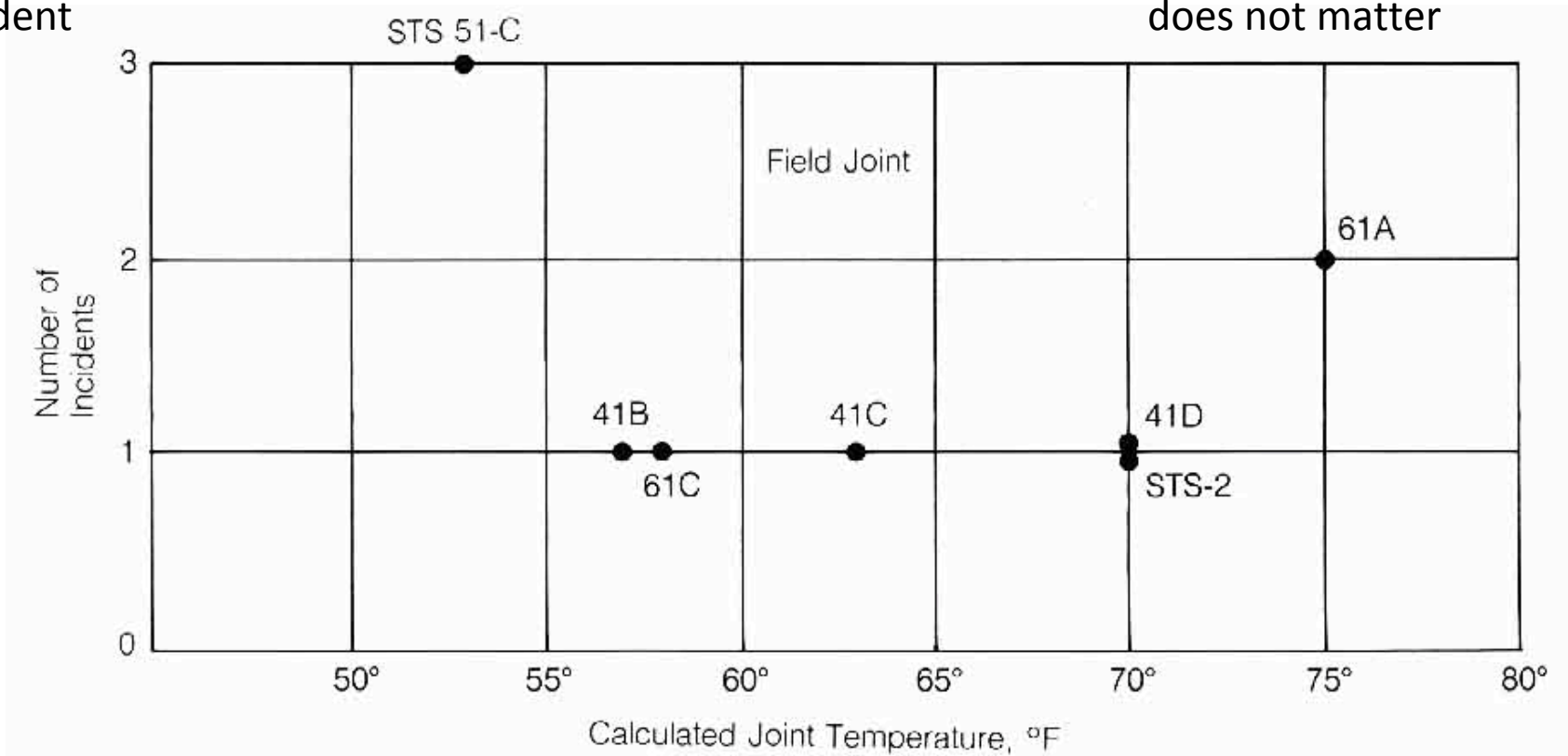
Measurement Effects



Survival Bias

Presidential Commission on
Space Shuttle Challenger
Accident

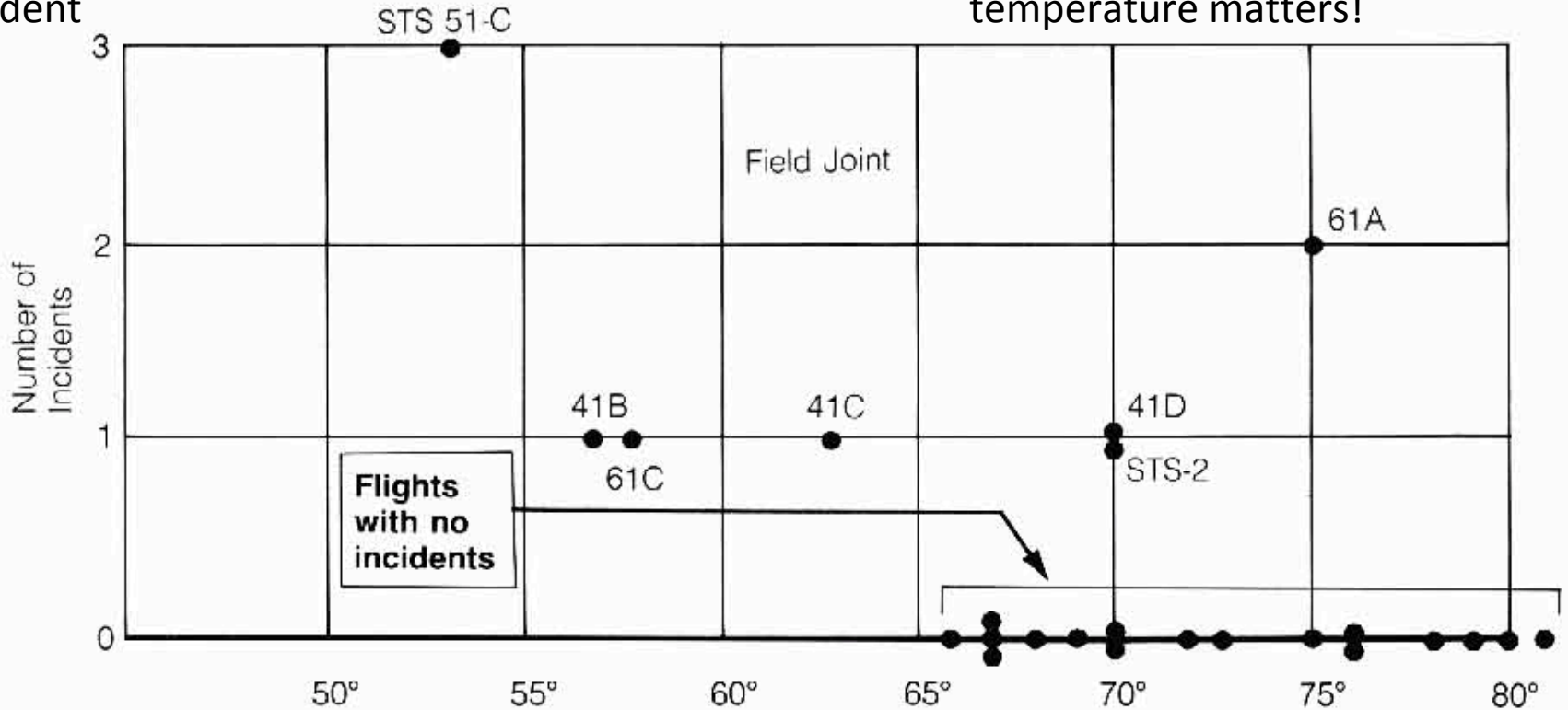
Failures at a range of
temperature – assume temp
does not matter



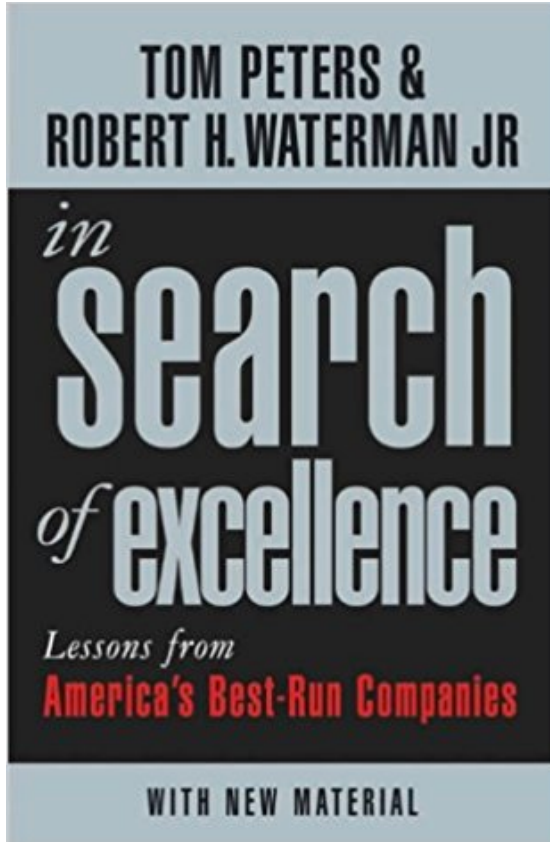
Survival Bias

Presidential Commission on
Space Shuttle Challenger
Accident

Failures rare at higher temperatures,
but common at lower temperatures –
temperature matters!



Survival Bias

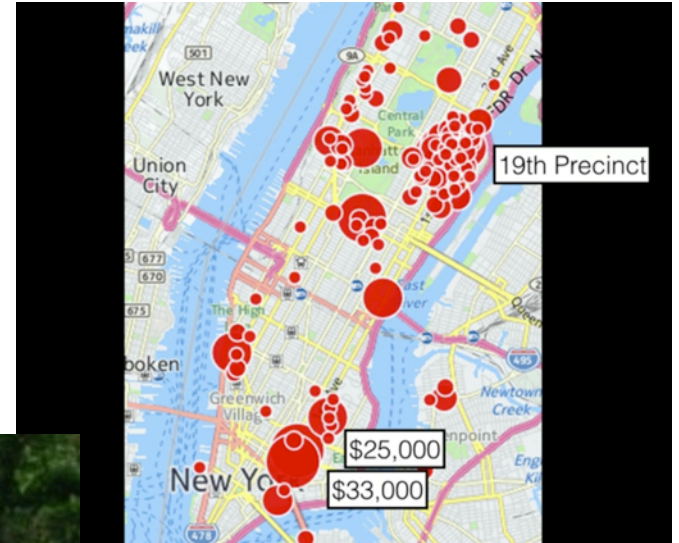


Look at successful companies, see what they have in common, create recommendations.

But, after a few years, many of the companies were failing.

To measure drivers, you must look at failure and success.

Not Explaining the 'Why?'



Things That Have Not Happened (Yet)



Extrapolation. In the Columbia Shuttle disaster the piece of foam that broke free was 620 times bigger than any item tested – but was judged to be harmless.

Type III and IV Errors

Mosteller

- Type I – Rejecting the null hypothesis when it is true
- Type II – Accepting the null hypothesis when it is false
- Type III – Correctly rejecting the null hypothesis for the wrong reason

Marascuilo & Levin

- Type IV – The incorrect interpretation of a correctly rejected hypothesis

Thank You

Follow me

Twitter: @RayPoynter

LinkedIn: www.linkedin.com/in/raypoynter

Connect: ray.poynter@thefutureplace.com