



Data Diaries Don't Lie:

Stats That Spot ePRO Risks

Nechama Katan, Wicked Problem Wizards

Steven Gilbert, Pfizer





Disclosures

- Opinions in this presentation are from Steve Gilbert personally and not from Pfizer.
- Opinions from Nechama Katan represent Wicked Problem Wizards.
- There is no financial relationship between Pfizer and Wicked Problem Wizards.
- All data is simulated, no actual study data was used.



Agenda

- Why Audit Trail Analytics?
- Why pre processing data?
- Common levels of Analysis?
- A worked case study



The Rules of the Road Have Changed

“
L

Procedures for review of trial-specific data, audit trails and other relevant metadata should be in place. It should be a **planned activity**...

(ICH E6(R3), Section 4.2.3)

”

The Mandate: From Ad Hoc Forensics to a Planned, Risk-Based Discipline.

ICH E6(R3) shifts quality management from 'doing everything right' to doing the *right things*, right-sized and right-timed.



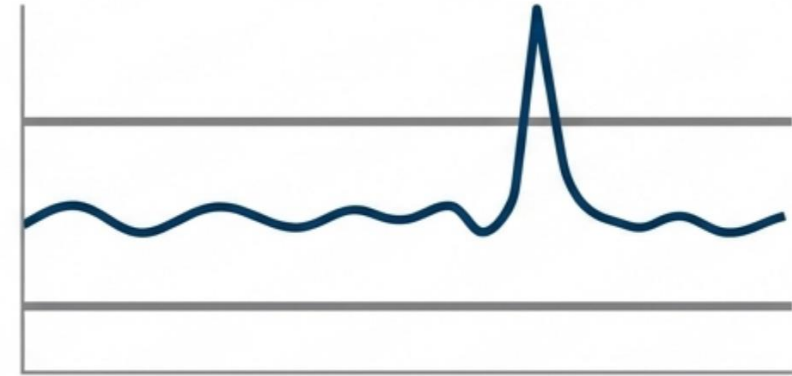
Why Audit Trail and RBQM?

Patient Heart Rate (Outcome Data)



-  Patient Variability
-  Treatment Effect
-  Random Noise

Site Data Changes (Process Data)

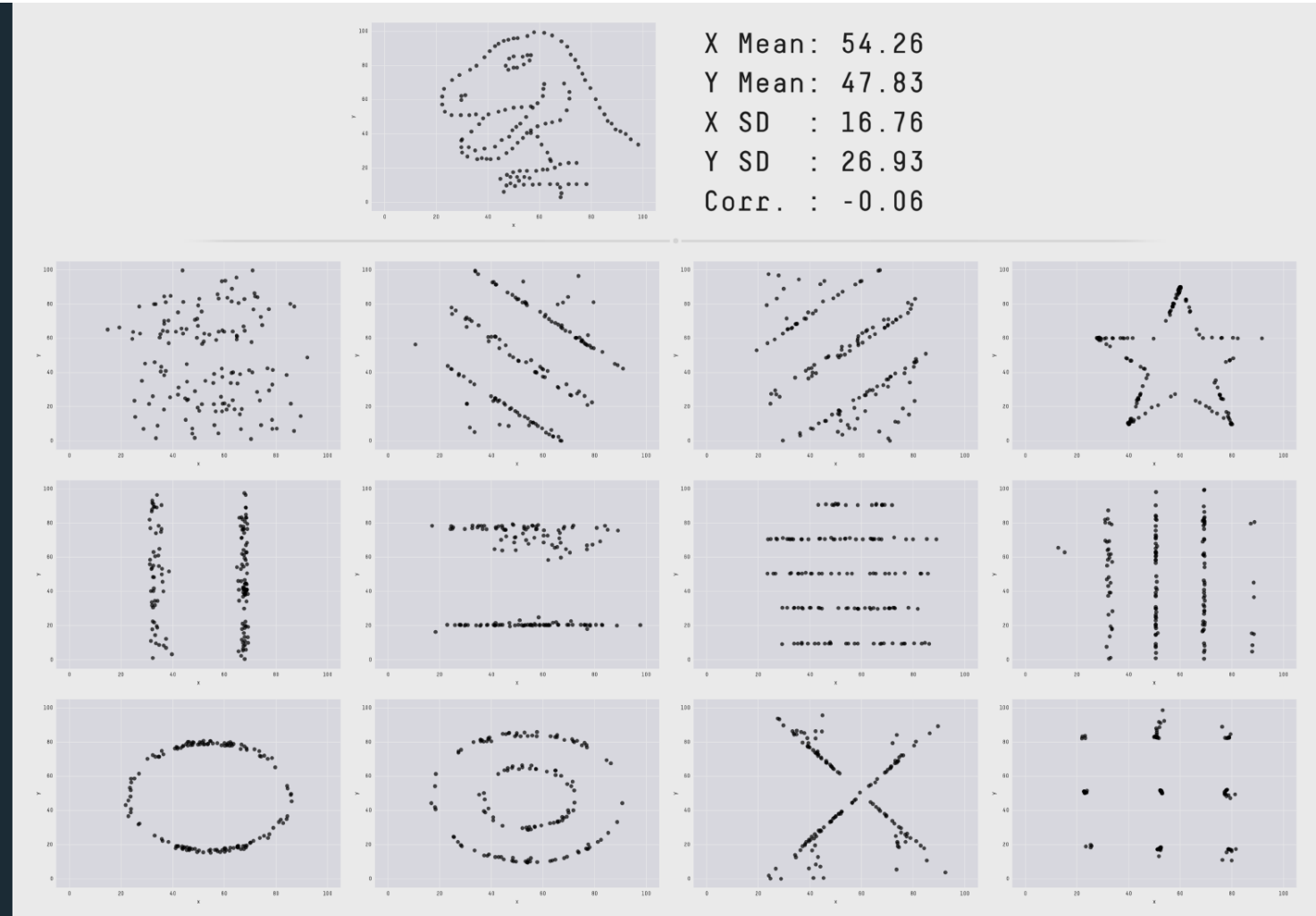


-  Site Behavior

RBQM was always meant to be about Process Control.
Audit Trail data is the best way to measure the *conduct* of a trial.
It needs to be part of your RBQM solution.



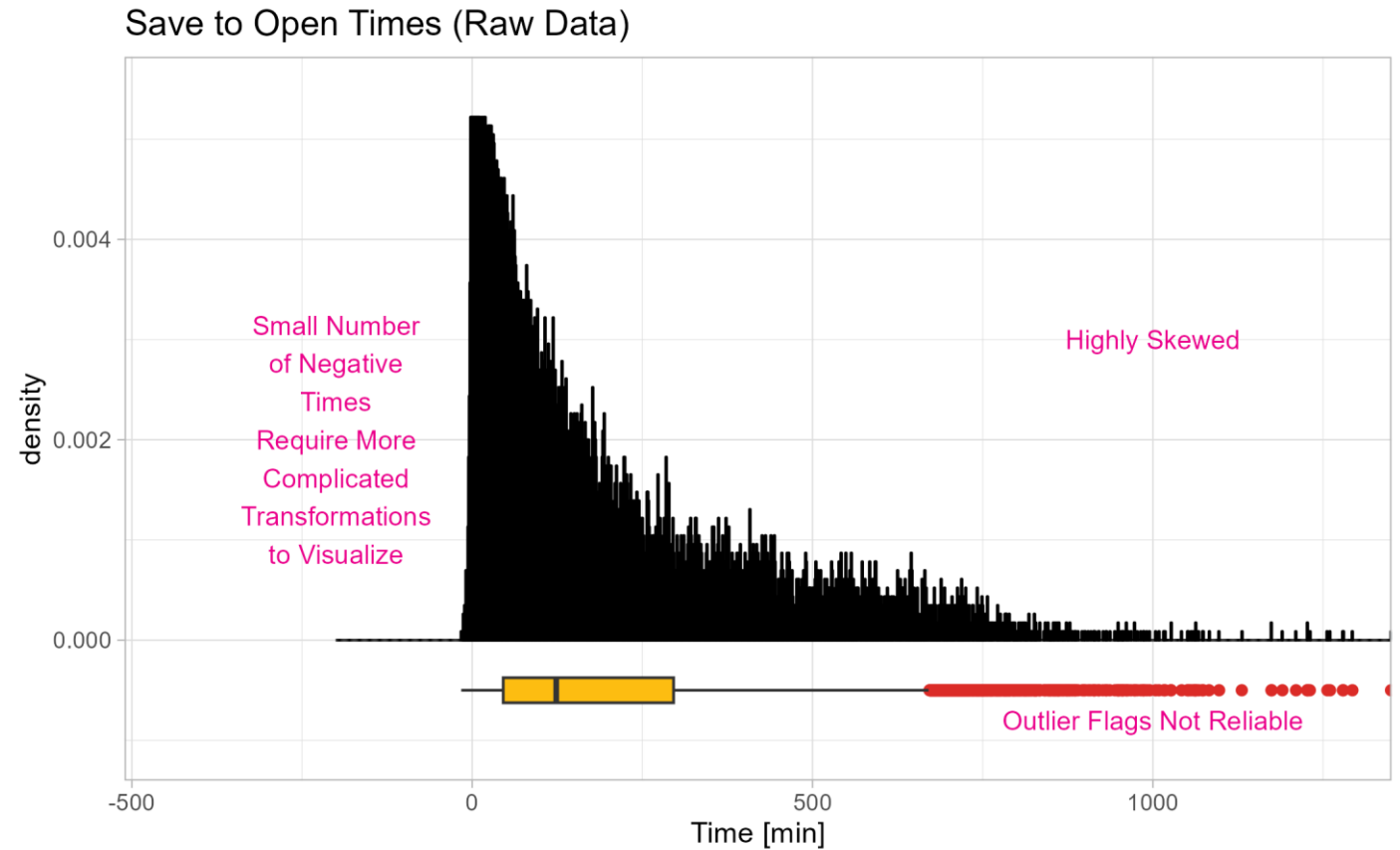
Summary Statistics vs Distribution



The Datasaurus Dozen, *Matejka & Fitzmaurice (2008)*



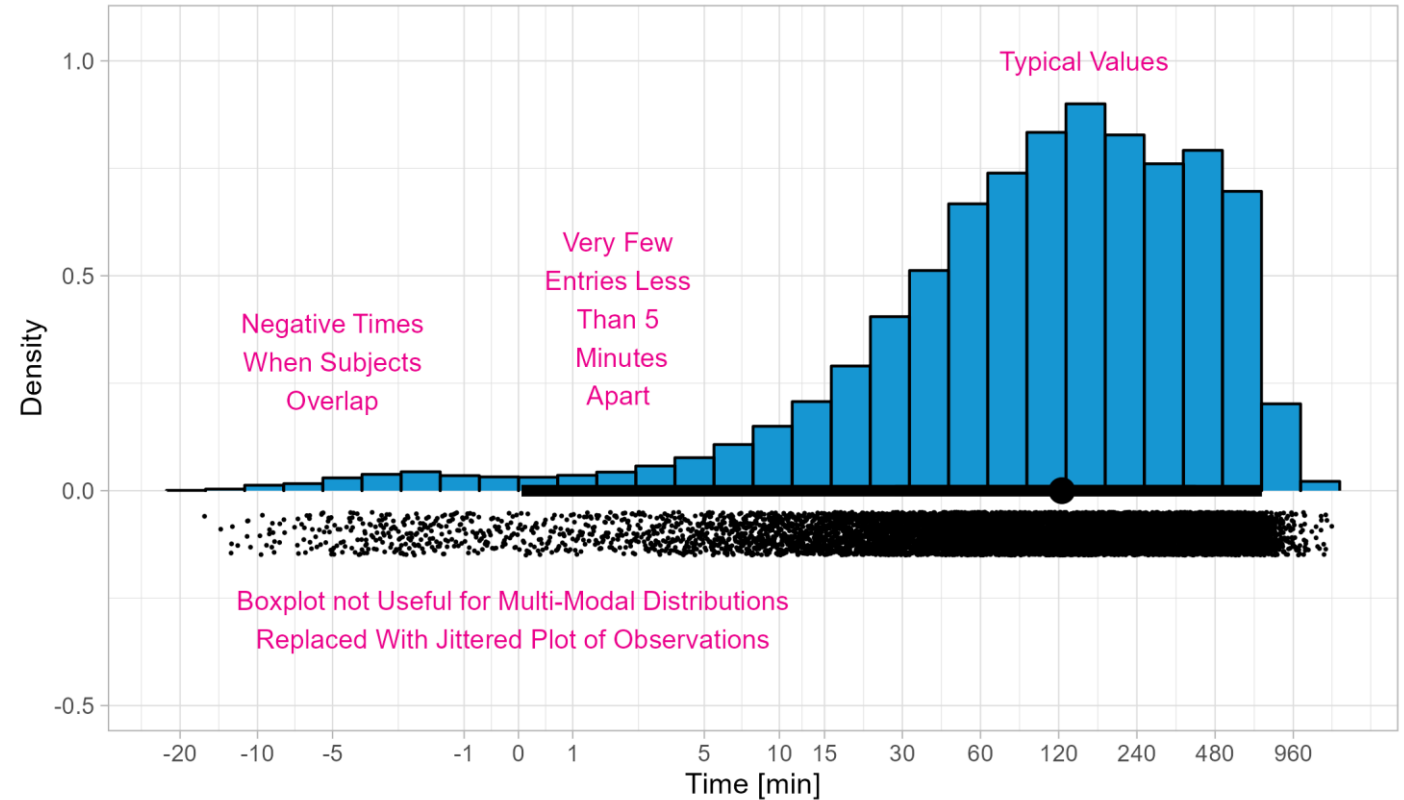
- Save to Open times are highly skewed
- Also contain negative values when subjects overlap.
- **Negative values cannot be transformed with the usual suspects (log, square-root etc)**





- **Pseudo-log transform** used here, works on positive and negative numbers.
- Small negative bump and peak at approximately 2 hours is now visible
- Small times (<5min) very uncommon – jumping off point for deeper dive
- All Data points are visualized in jitter plot below histogram
- Black point and bar denote median and middle 50% of the data

Save to Open Times (Pseudo Log Transformation)





The Analytics Maturity Ladder

Level 1: Descriptive Stats

- How many changes?

Level 2: Time-Trend Analysis

- Is the process stable over time? This is the control chart

Level 3: Advanced Analytics

- Which site is the biggest outlier compared to its peers?

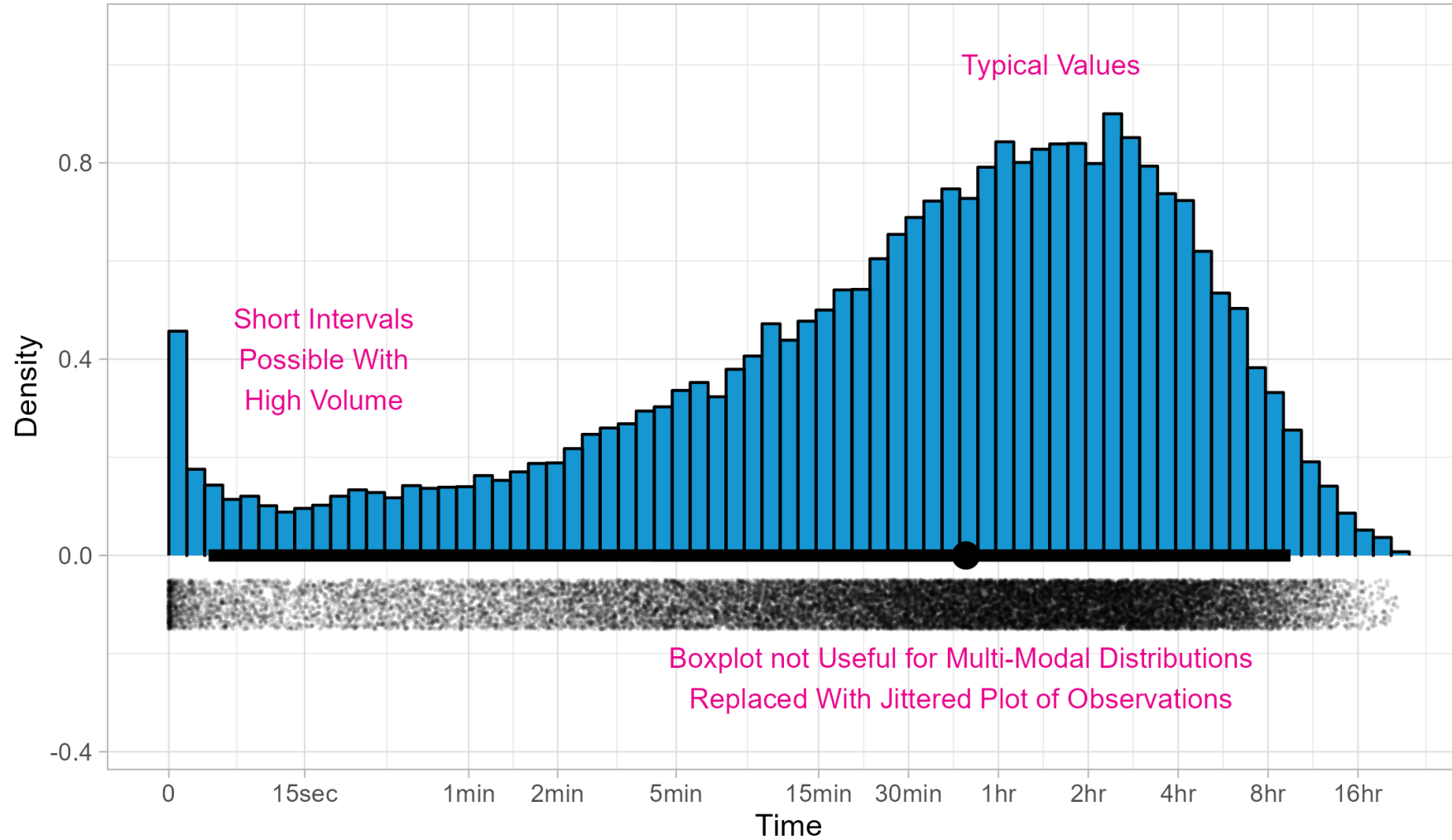
Level 4: Root Cause Analysis

- Why is this happening?



Descriptive Statistics

Open to Open Times (Pseudo Log Transformation)





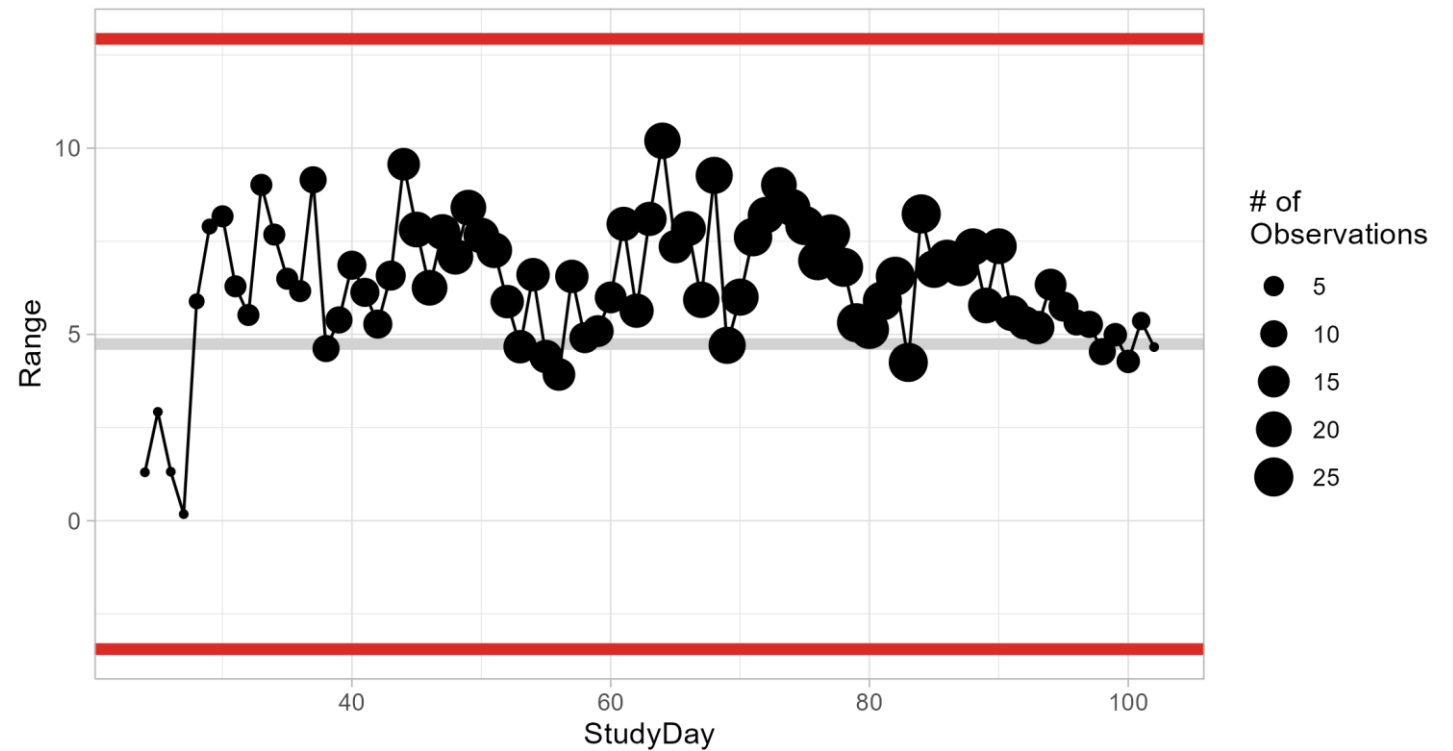
Time Trend Range Data

Open to Open time

- Data log transformed because highly skewed (gamma distribution fits it well)
- R chart instead of standard deviations used because even large site can have very few observations on a single study day

R-Chart by Study Day For a Single Site

Center line and 3 Sigma Lines



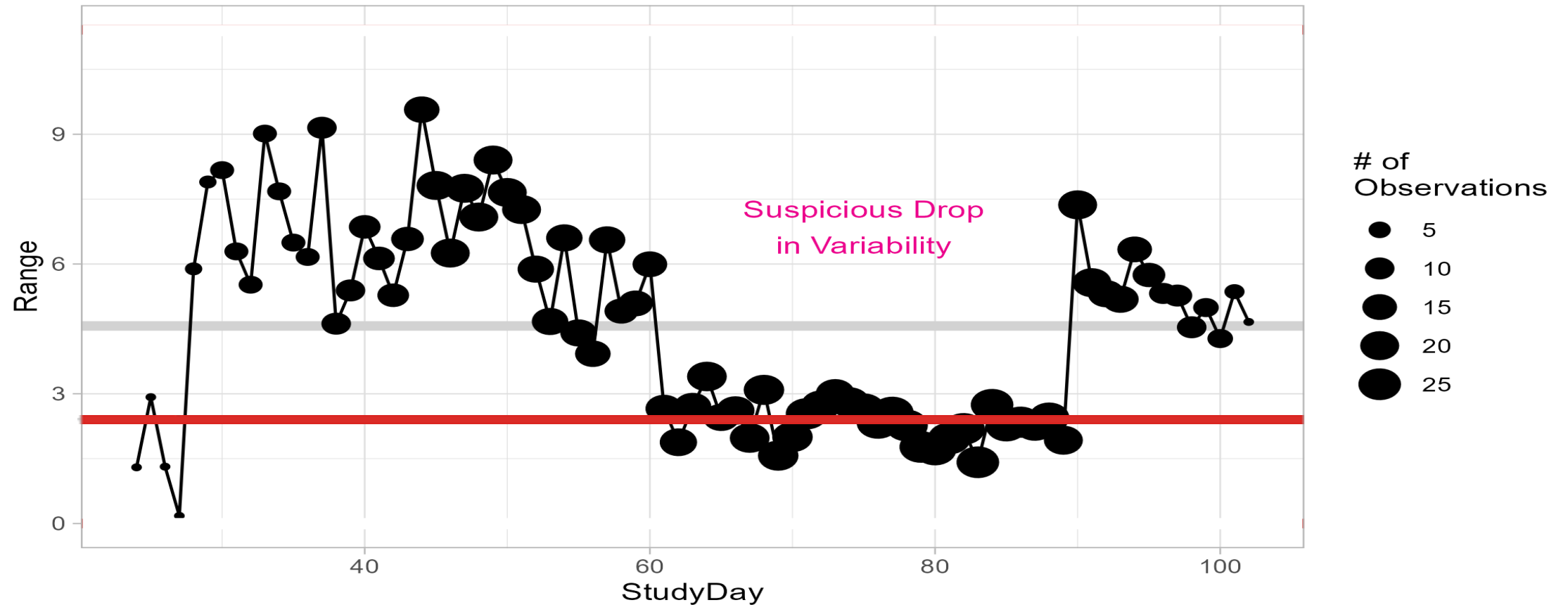
Data Log Transformed



Time Trend Range Data Issue

R-Chart by Study Day For a Single Site

Center line and 3 Sigma Lines



Data Log Transformed



Advanced Analytics

The need:

- > 100s of metrics
- > 100s of sites

Individual review of all charts is not scalable.

There is a need for methods to determine the most unusual site behavior



Root Cause Analysis

Signal to Insight: An Investigative Framework



What factors are associated with the signal?



Signal Detected

The starting point of any investigation—an anomaly, outlier, or pattern in the data that requires further analysis.



Does the signal remain after adjusting for other variables?



Are multiple factors interacting?



What explains why the signal occurs?



**Outcome:
Explanation
and Decision
Support**

Where is this happening?



Which site, user, or visit?



When does it occur most often?



How large is the signal?



**Outcome:
Pattern
Identification**



More ATRA Resources

- <https://www.fda.gov/drugs/news-events-human-drugs/fda-and-mhra-good-clinical-practice-symposium-data-integrity-global-clinical-trials-tackling>
- Playlist <https://www.youtube.com/playlist?list=PLey4Qe-UxcxZvRWJ6dulyw1P5eQWWztz1>
- eClinical Forum / SCDM 2021 paper
<https://eclinicalforum.org/news/public-release-audit-trail-review-key-tool-to-ensure-data-integrity-version-10>
- eClinical Forum 2026 paper
<https://eclinicalforum.org/downloads/public-release-of-strategies-for-implementing-audit-trail-review-industry-position-paper>



Thank you!

Nechama Katan

Nechama@wickedproblemwizards.com

[linkedin.com/in/Nechama](https://www.linkedin.com/in/Nechama)

<https://www.wickedproblemwizards.com/>

Steve Gilbert

Steven.a.gilbert@pfizer.com

[linkedin.com/in/steven-gilbert-110913b](https://www.linkedin.com/in/steven-gilbert-110913b)

