Exponent®
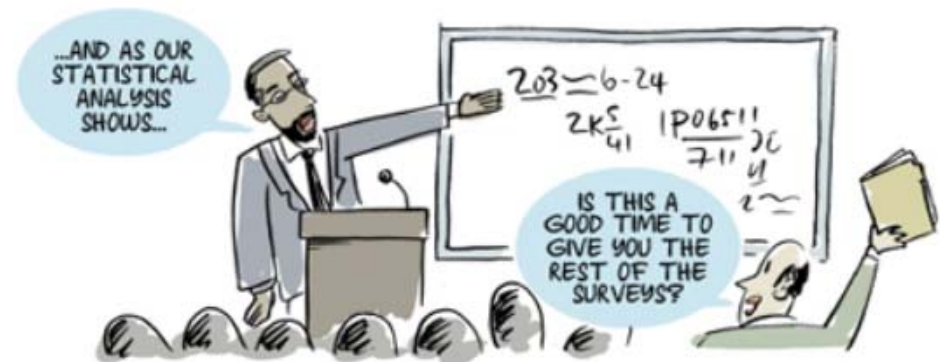
Ensuring Your Data are Fit for Forensic Analysis Use

Linda Cook

Laurie Benton

August 9, 2019

What is Data Fitness?

- Data Fitness ≠ Data Validation

- Data fitness reviews

  - Evaluate data based on intended use

  - Right analyses, quality, quantity, coverage

  - Comparability

- Essential for repurposed data

When the data are not fit … conclusions may be flawed

Ensuring your Data are Fit for Use

- Plan for forensic needs

- Select a qualified laboratory

- Analyze reference materials

- Look at the data

- Determine whether data are fit for intended use

Plan for Forensic Needs

Engage SMEs Early in Data Collection Planning

- Define future data needs

- Assess potential litigation scenarios

- Collect multi-purpose data

- Collect source and ephemeral samples

- Develop user guided data management approaches
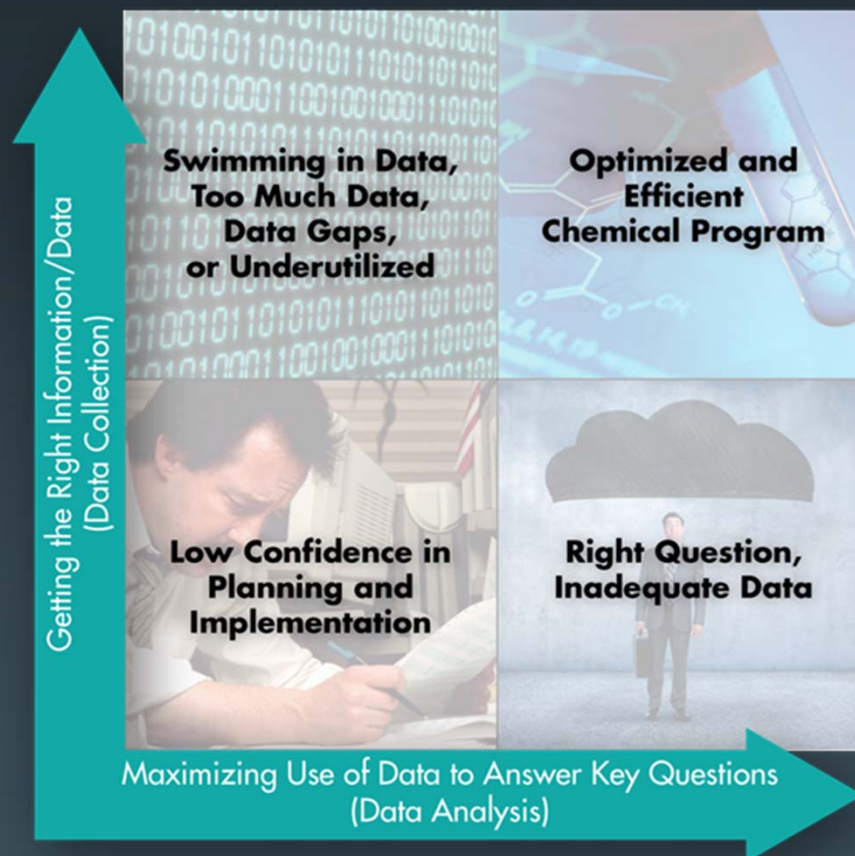
Common Lessons Learned from Forensic Cases

- Limited source/background data

- Wrong or inconsistent methods

- Laboratory variability

- "Messy" data management

What data do we have, what data will we need?

Consider

- Questions that need answering

- Pre-existing site data (usable?)

- Source data

- Background data

- Other types of data needed

Collect Forensic Quality During Investigations to Support Future Objectives
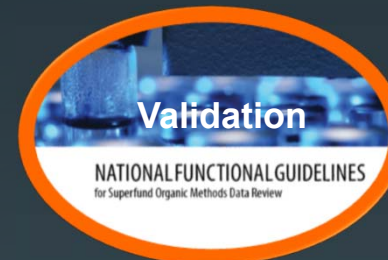
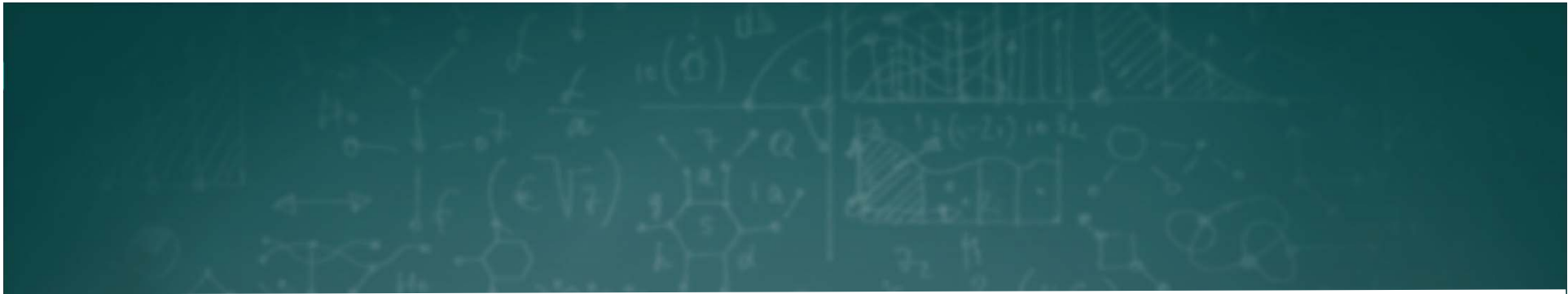Adding forensic quality data to a site investigation may be more expensive

Lack of forensic quality data, when needed, may be more expensive

Where are the real data costs?

Field Work

Shipping

Analysis

Validation

NATIONAL FUNCTIONAL GUIDELINES
for Superfund Organic Methods Data Review

Verification

SHORELINE OILING
ASSESSMENT FIELD BOOK
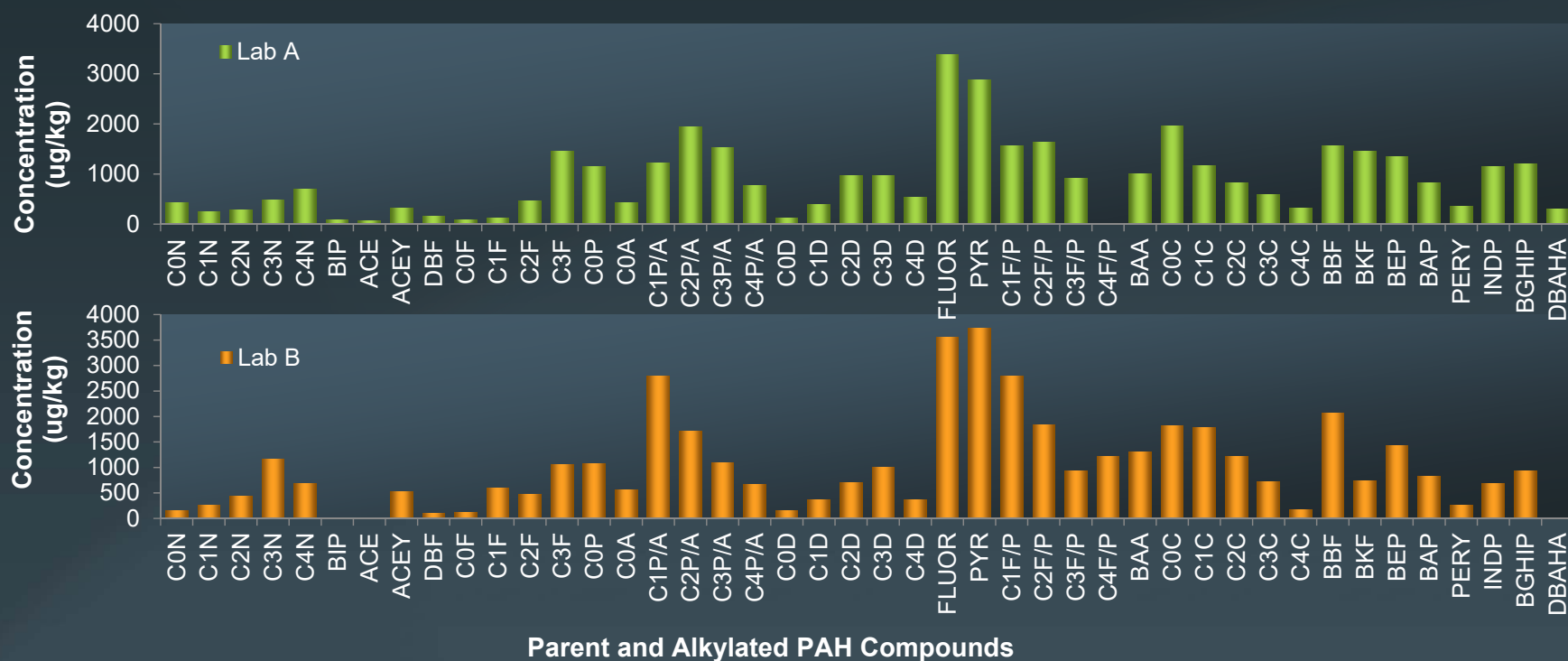
No. X-867

Data Storage

$$$?

Qualified Laboratory

Nothing Beats a Highly Qualified Laboratory

- More laboratories are performing specialty and forensic analyses

- Offering competitive pricing

- But are they introducing data quality or comparability issues?
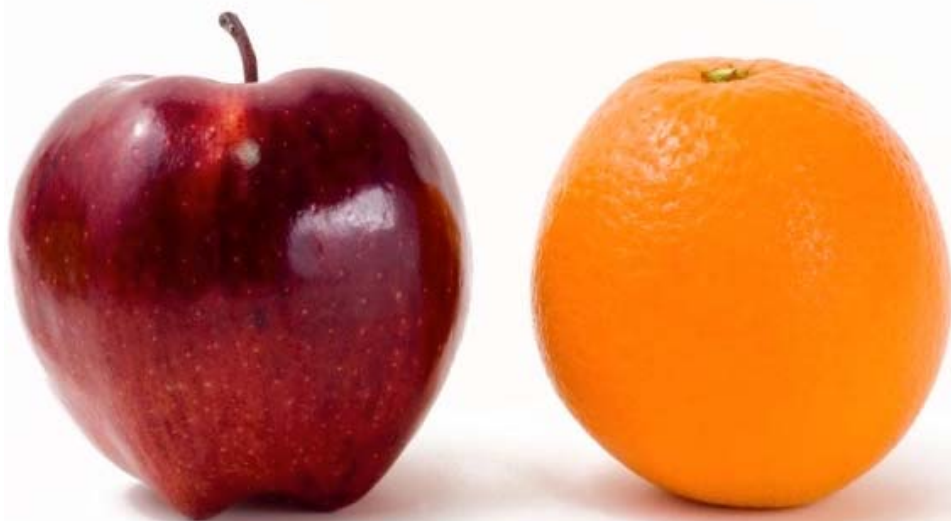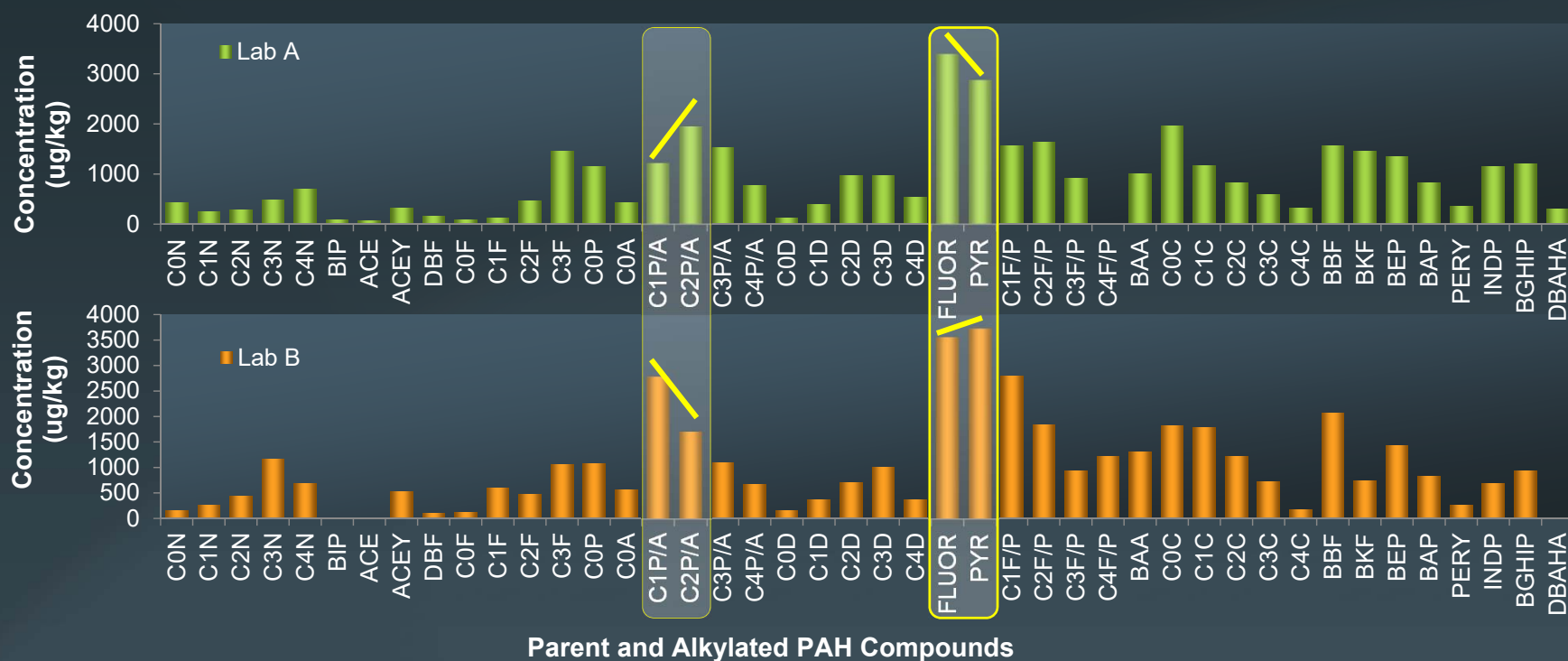
Evidence of Differences in NIST Interlaboratory Studies

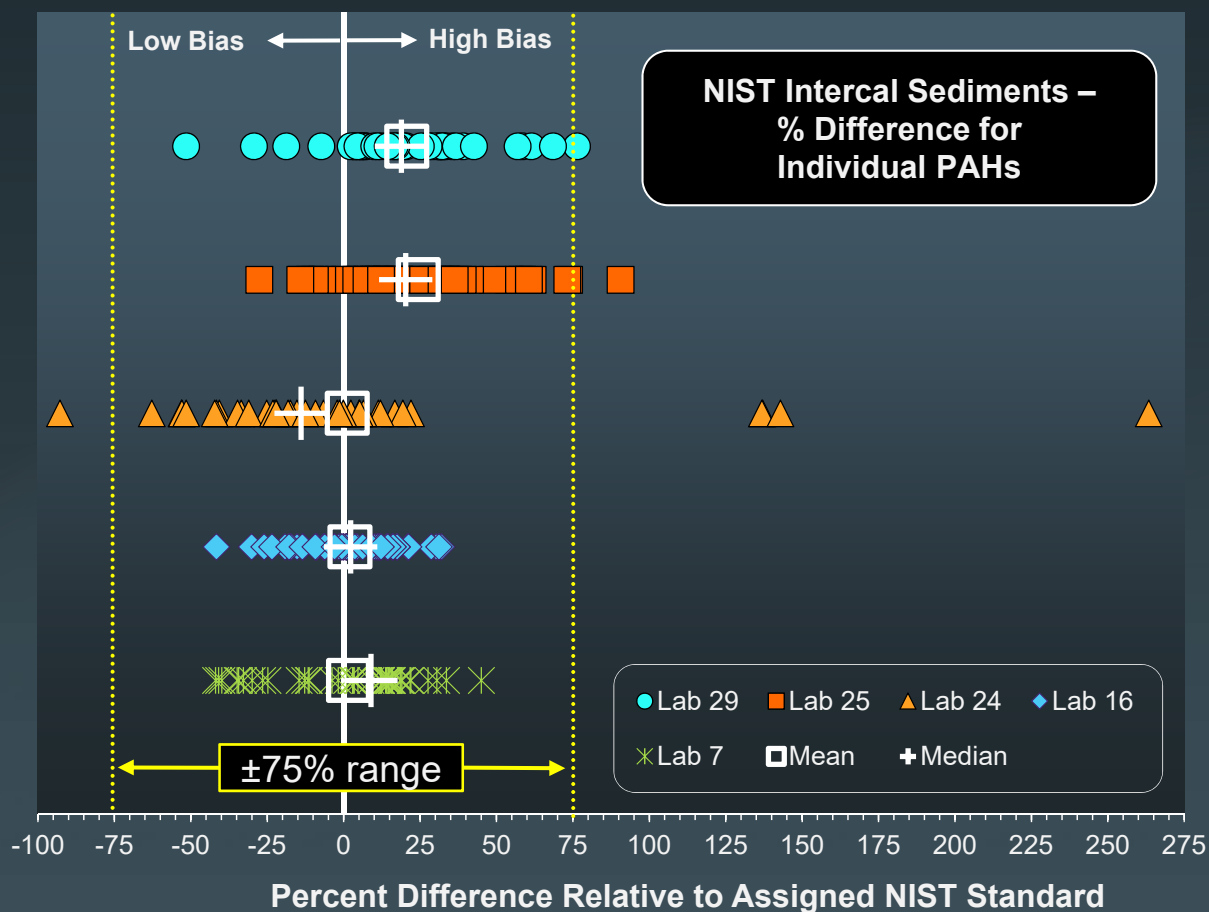Source: NIST Marine Sediment Intercalibration Study QA10SED01

Evidence of Differences in NIST Interlaboratory Studies

Source: NIST Marine Sediment Intercalibration Study QA10SED01

Interlaboratory Variability is Inevitable

How do we minimize impact on your forensic analyses?

NIST Intercal Sediments – % Difference for Individual PAHs

Low Bias ← → High Bias

±75% range

Lab 29   Lab 25   Lab 24   Lab 16
Lab 7   Mean   Median

-100 -75 -50 -25 0 25 50 75 100 125 150 175 200 225 250 275

Percent Difference Relative to Assigned NIST Standard

Strategies to Minimize Impact of Laboratory Variability

- Select qualified laboratory

- Seek advice of SME

- Send "test" sample

- Review NIST intercalibration results

- Stick with same laboratory

Analyze Reference Materials

Standard Reference Material or Control Samples Essential for Assessing Differences in Datasets

- Add to every program

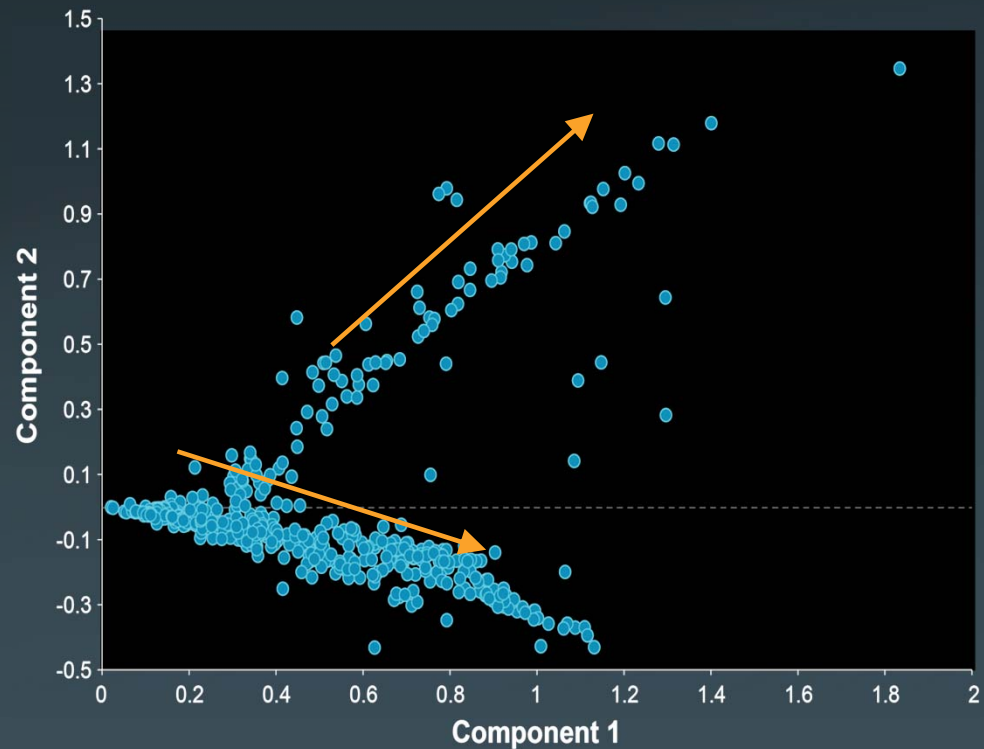- Allows uses to assess difference between labs and over time
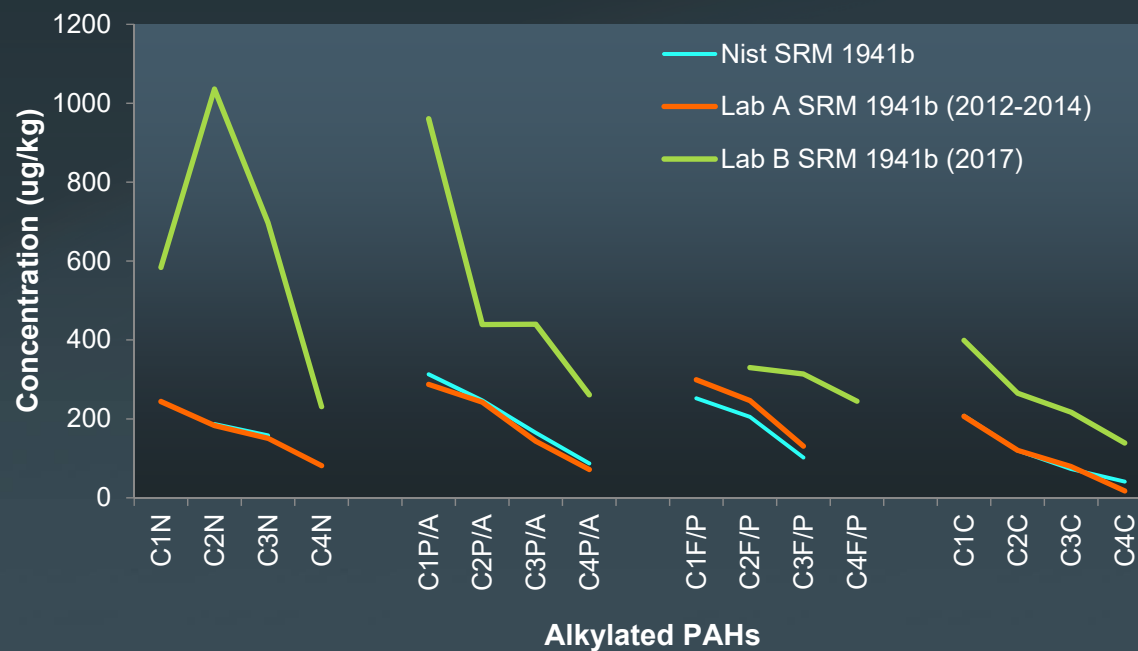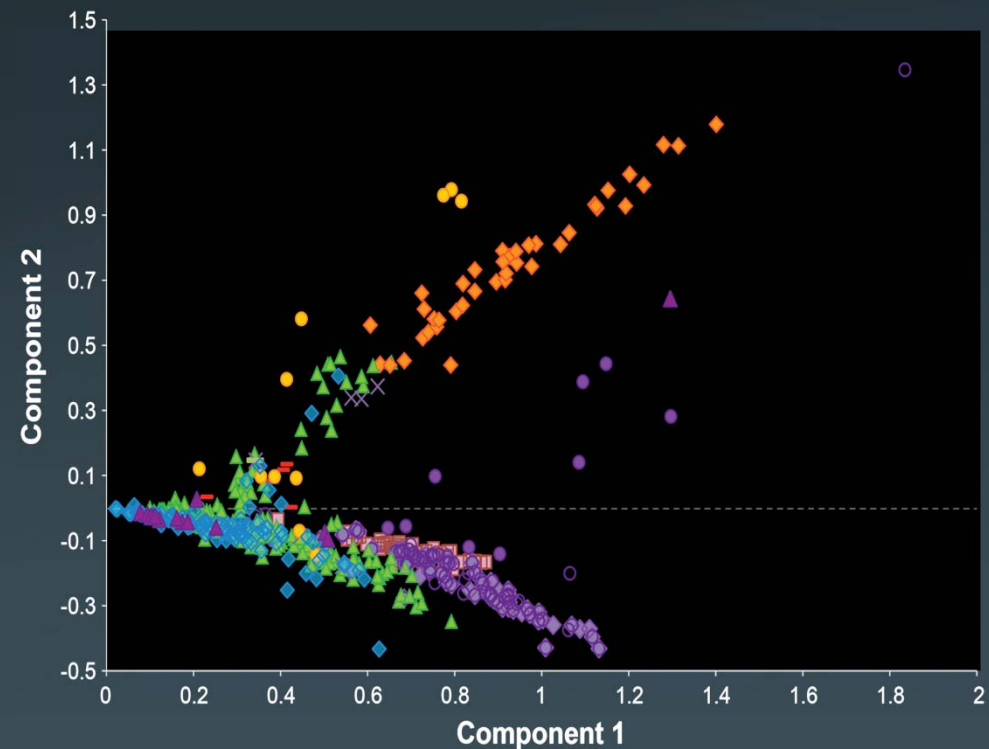

Source: NIST Marine Sediment Intercalibration Study QA10SED01

First Appearances May Be Deceptive

- Common statistical method to evaluate composition

- Two sources?
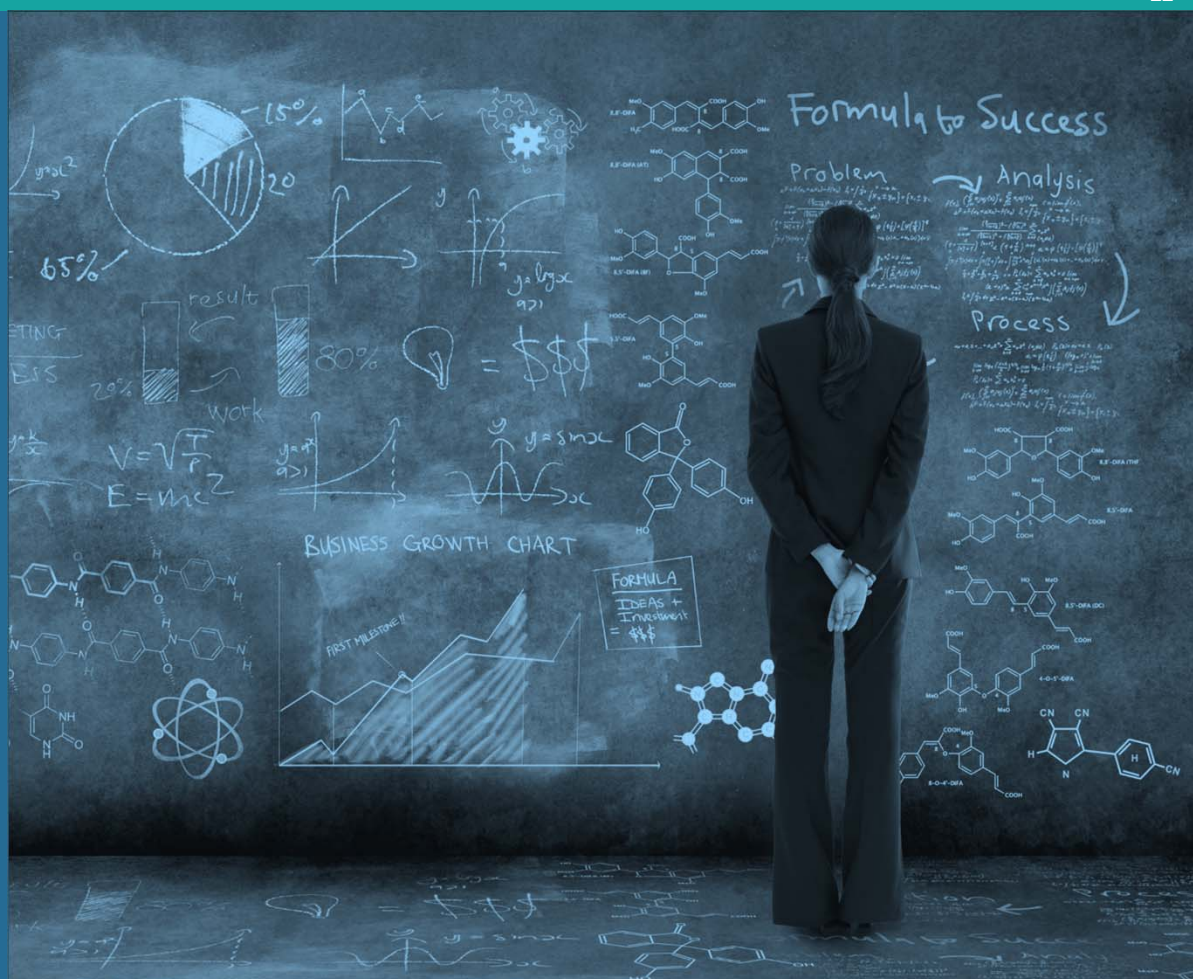
- No SRMs available

Review the Details When Combining

- Multiple sources?

- Study differences?

- Sampling bias?

- Laboratory bias?

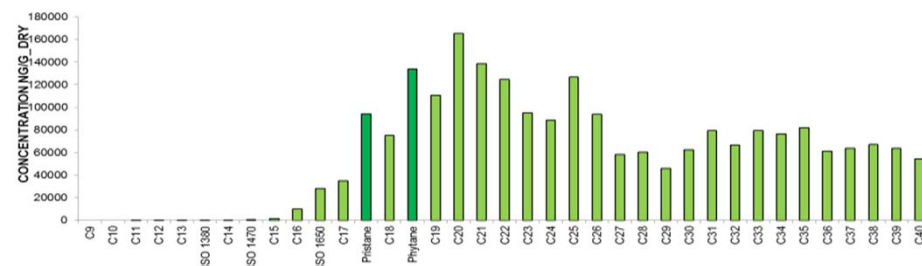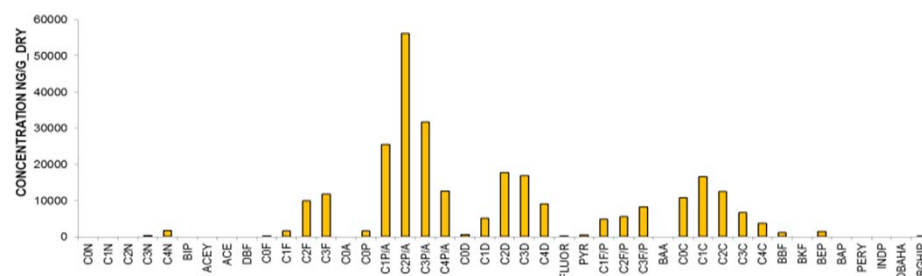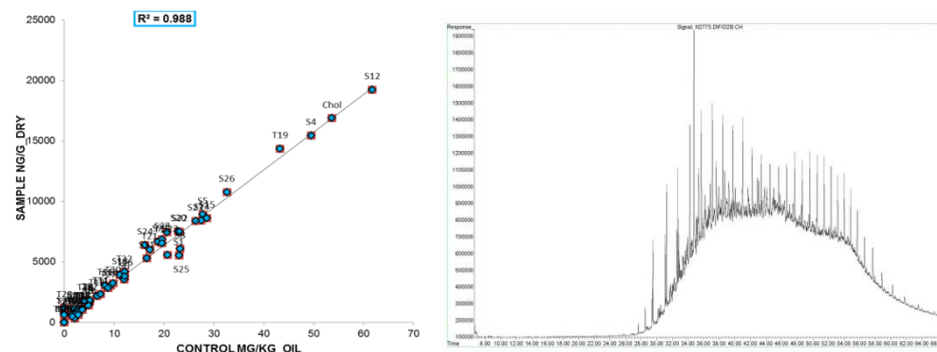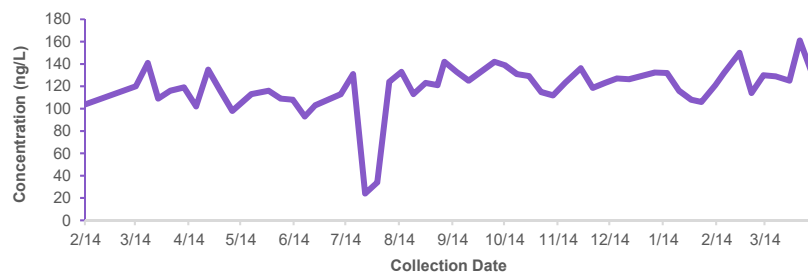- SRMs would have aided in addressing these questions

Data Visualization

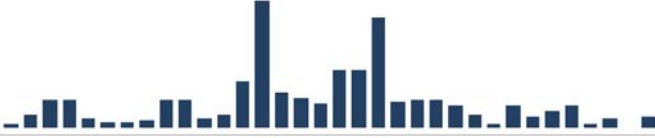# Look at the Data – Do They Make Sense?

Simple Graphics Convey a Wealth of Information

- Concentration histograms
- Scatterplots
- Sample chromatograms
- Trend graphs

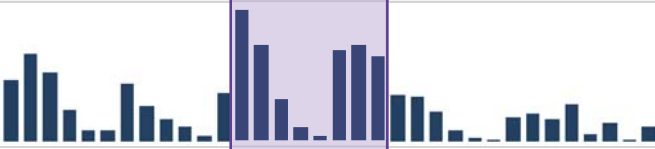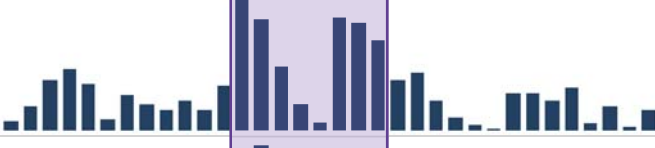# Excel Sparklines—Quick and Simple Data Vis Tool

Sample ID	Total PAH34 (mg/kg)	PAH Histogram
Lab A - Sample 1	260	
Lab A - Sample 2	96	
Lab B - Sample 1	263	
Lab B - Sample 2	663	

PAH Graphs Uncovered Systematic Lab Problem

Sample ID	Total PAH34 (mg/kg)	PAH Histogram
Lab A - Sample 1	260	
Lab A - Sample 2	96	
Lab B - Sample 1	263	
Lab B - Sample 2	663	

# Evidence of Issue Observed in NIST 2011 Interlab Study
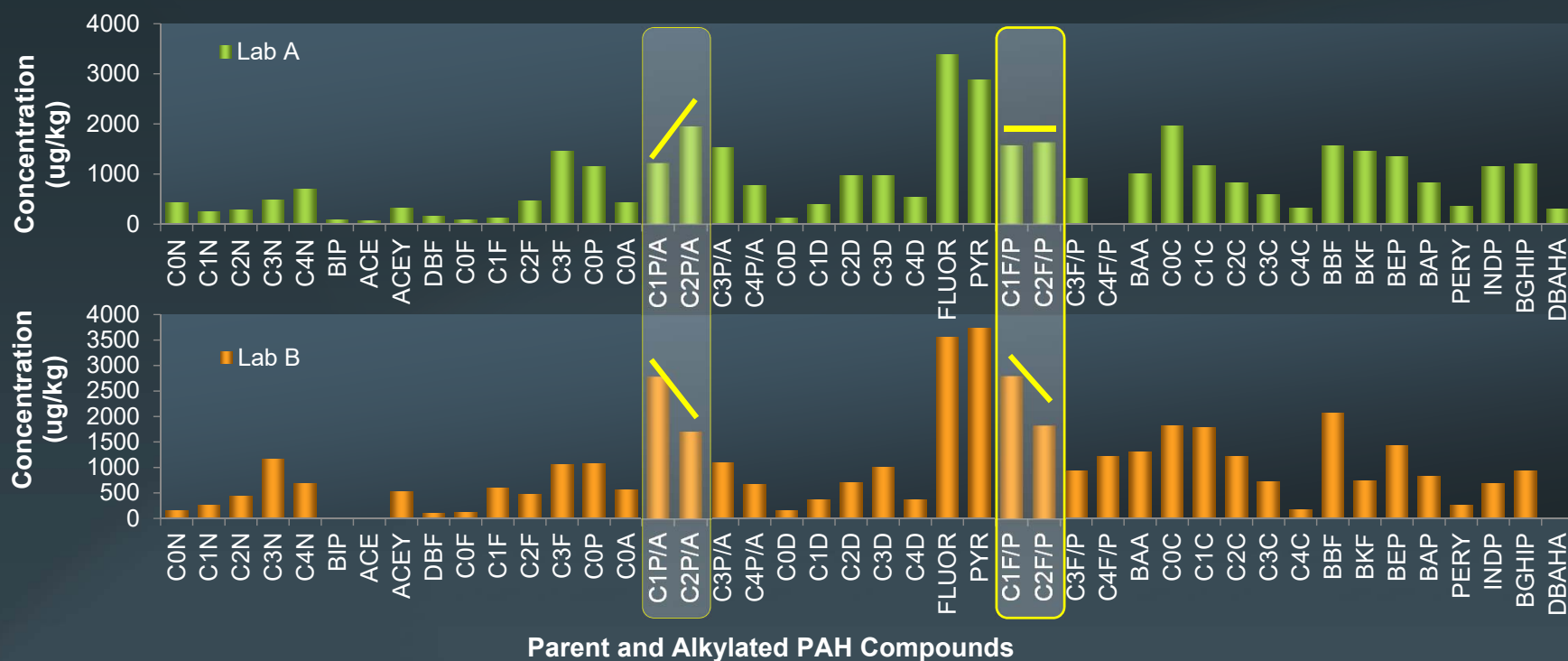
Source: NIST Marine Sediment Intercalibration Study QA10SED01
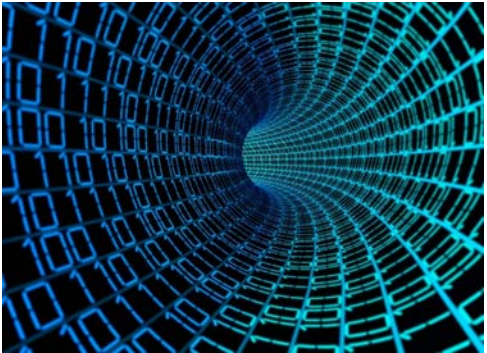
Data Fitness

Are the Quality of My Data Sufficient for My Intended Use?

- How will I use these data?

- Validated Data ≠ Data Fit-for-Use

- Is the accuracy, precision, comparability of these data sufficient for my use?

- Have apparent differences been tested in data subsets?
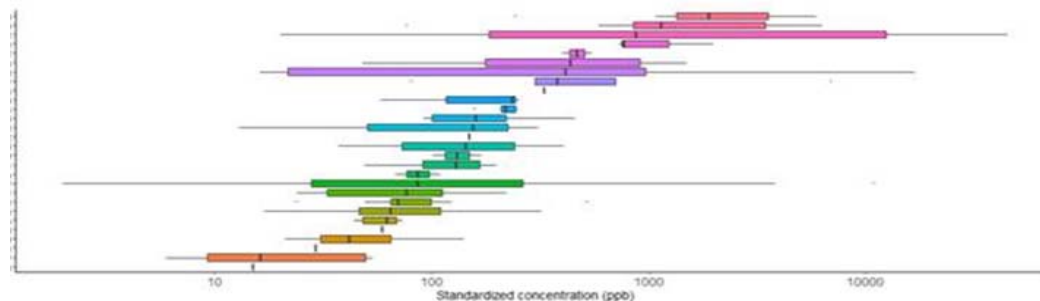
When Data are Not Fit

- Data fitness is not ALL or NOTHING

- Focus on fit bits

- Avoid using unfit bits

- Alter data use plans based on fitness

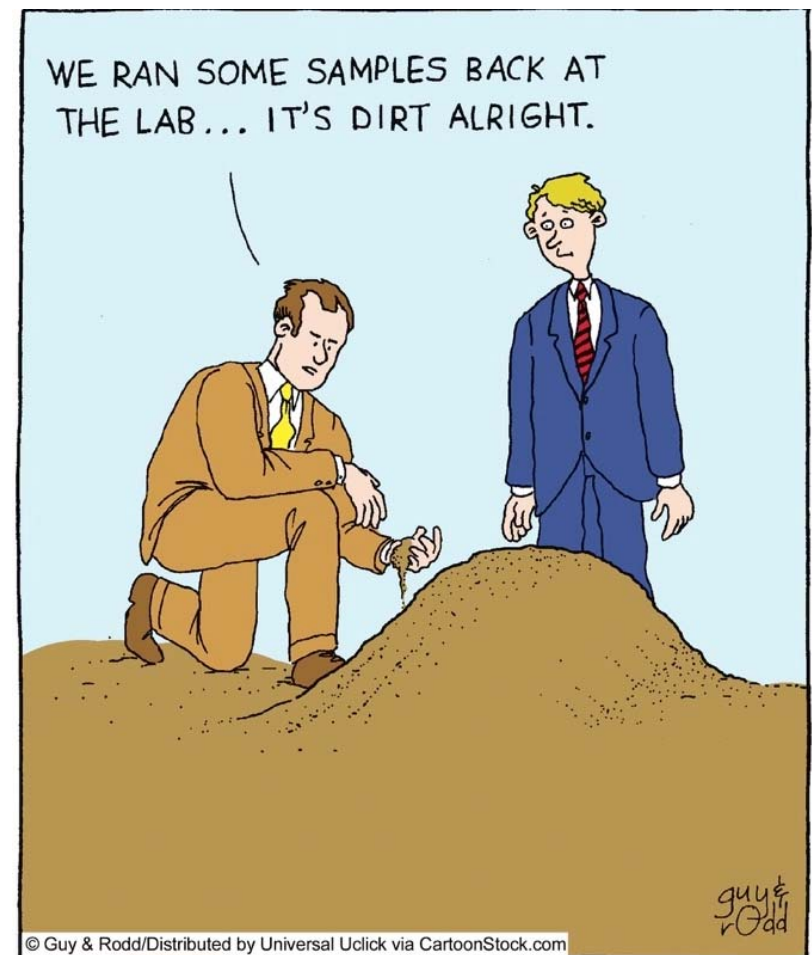Conclusions based on data that are not fit … may be flawed.

Data Optimization

Optimize your Data Collection and Management Efforts

- Integrate

- Normalize

- Organize

- Evaluate for fitness

- Explore

- Visualize

Five Steps for Ensuring Data Fitness

- Plan for forensic needs

- Select qualified laboratory

- Analyze reference materials

- Look at the data

- Assess data fitness for intended use

Thank you!