PANTENE

P&G

Unlocking the Power of Statistical Engineering

Michael Joner

March 28, 2024

DAWN

Michael Joner

R&D DIRECTOR AT PROCTER & GAMBLE
DIGITAL AND DATA DISRUPTION

B.S. Statistics 2002, Brigham Young
M.S. Statistics 2003, Brigham Young
Ph.D. Statistics 2007, Virginia Tech

ISEA Board Member At-Large
ASQ Statistics Division Past Chair

Part-time Instructor in Computer Science at Boston Univ.

What is Digital and Data Disruption?

An R&D organization to more effectively connect data, models, and AI together to enhance the ability of our scientists to fulfill P&G's purpose

P&G's purpose is to provide branded products and services of superior quality and value that improve the lives of the world's consumers, now and for generations to come.

Digital, Data, and AI methods are increasing in importance at P&G. Beyond R&D, we have similar organizations in Manufacturing, Information Technology, Marketing…

P&G Today

Founded	**1837**
Fiscal 2023 Net Sales	**$82.0 Billion**
Countries of Operations	**~70**
Countries and Territories Where Our Products are Sold	**~180**
Number of Employees	**~107,000**

A portfolio of brands in 10 daily-use categories

A Portfolio of Daily-Use Products

Measures of Superiority

PRODUCT
Products so good, consumers recognize the difference. Superior products raise expectations for performance in the category.

PACKAGING
Packaging that attracts consumers, conveys brand equity, helps consumers select the best product for their needs and delights consumers during use.

BRAND COMMUNICATION
Advertising that reaches consumers and communicates the superiority of the brand's product and packaging benefits—attracting consumers to the brand and driving brand and category growth.

RETAIL EXECUTION
In-store: with the right store coverage, product forms, sizes, price points, shelving and merchandising.
Online: with the right content, assortment, ratings, reviews, search and subscription offerings.

CONSUMER & CUSTOMER VALUE
For consumers: all these elements presented in a clear and shoppable way at a compelling price.
For customers: margin, penny profit, trip generation, basket size and category growth.

Innovation cannot be irresistibly superior without also being sustainable… to improve lives now and for generations to come.

https://us.pg.com/environmental-sustainability/

Superiority is a High Standard

PRESENTING COMPLEXITY AND LACK OF STRUCTURE

When we develop superior products, packaging, communication, and execution while also delivering consumer value …

… we frequently encounter **complex, unstructured problems crossing several disciplines and requiring data and data analysis**

Statistical Engineering Examples

Providing superior quality and value

If you cannot make pure goods and **full weight**, go to something else that is honest, even if it is breaking stone.

- James Gamble
  P&G Co-founder
  mid-1800s

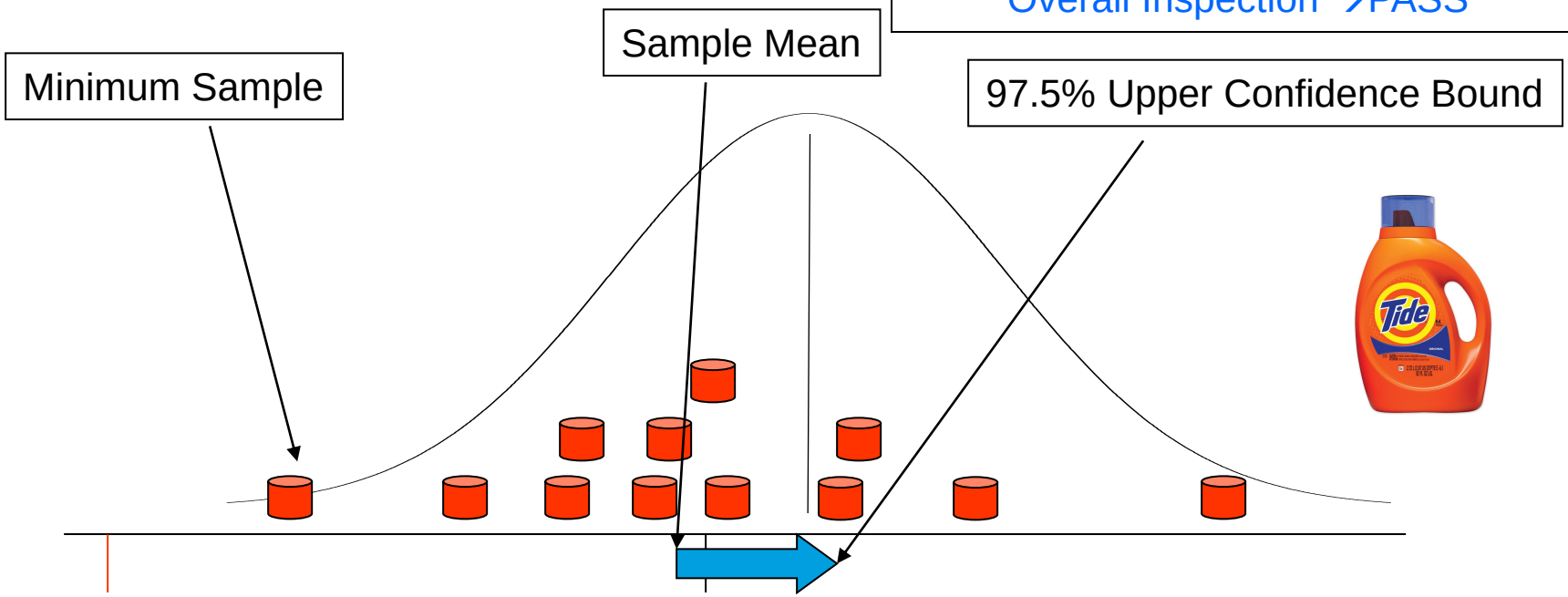
Introduction to Fill Weight Regulations

US Regulations
NIST Handbook 133

Example Result

- Sample n = 12
- Mean = 99.8 oz
- UB = 101.4 oz → PASS Average
- Min = 98.4 oz → PASS Individuals

Overall Inspection →PASS



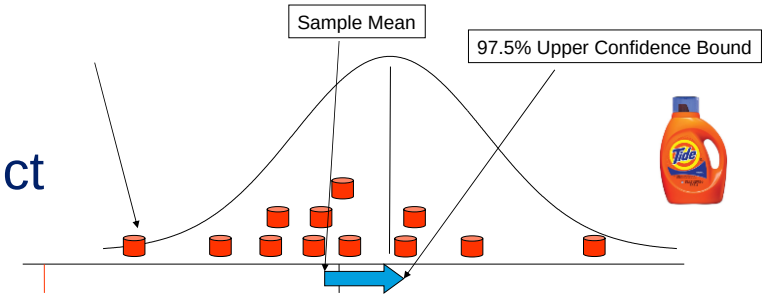
Some Statistical Questions on the Fill Weight Problem

Do we need a census of bottle weights, or is a sample adequate?

- High speed sampling/census may have instrumentation accuracy issues
- Sampling strategy requires understanding how the distribution of product weight/volume changes over time on the manufacturing line

Since regulator will usually obtain bottles from the store:

- How is the distribution of product weight/volume impacted by product distribution and shelf restocking?



Why it matters:

- *Can data collected by the plant be used to build confidence in passing a regulatory inspection?*
- *Can we be assured that we are being honest with our customers, delivering a consistent amount of product and therefore consistent value?*

It’s Statistical Engineering!

Complex?

- Yes

Unstructured?

- Yes, until we established structure

Multiple Disciplines:

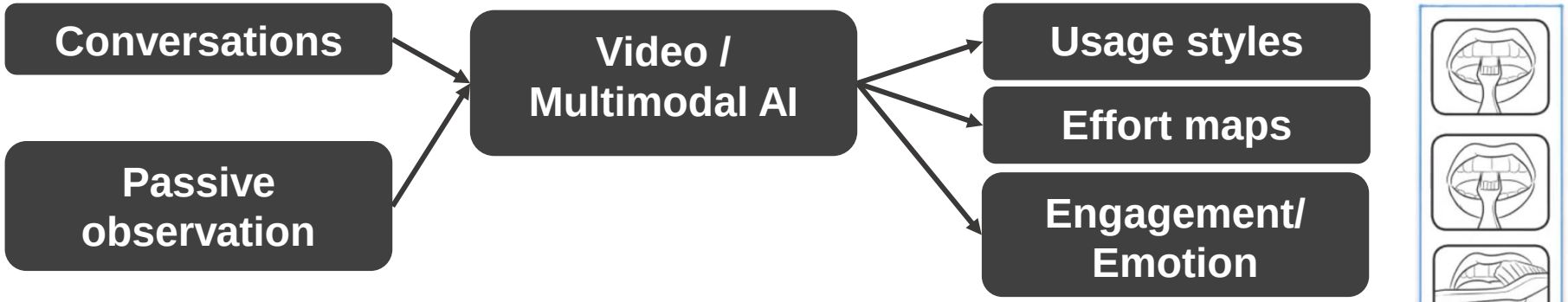
- Process Engineers
- Quality Assurance
- Regulatory
- Statistics
- Programmers

Requiring Data and Data Analysis?

- Yes

Consumer Experience & Multimodal AI

✓ **Understand the consumer:** what they do vs. what they say they do

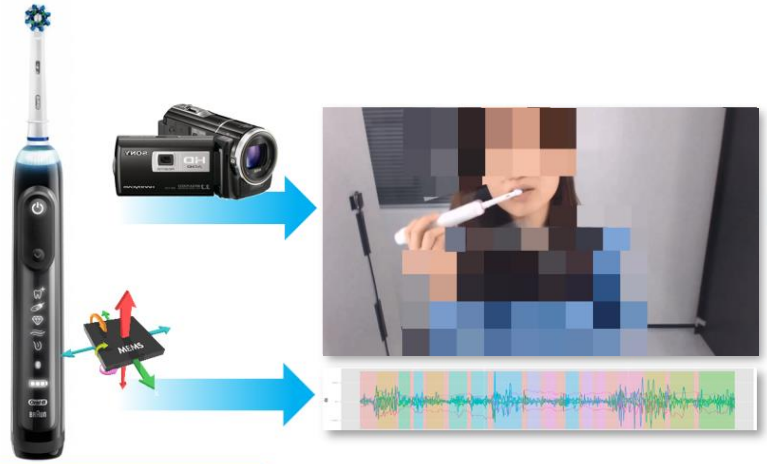


✓ **Outcome**

Dentists recommend:		Reality:
Brush for 2 minutes		<1 min on average
Apply light pressure		2x overpressure
Cover all areas evenly		80% miss at least 1 area, 60% don't cover back molars

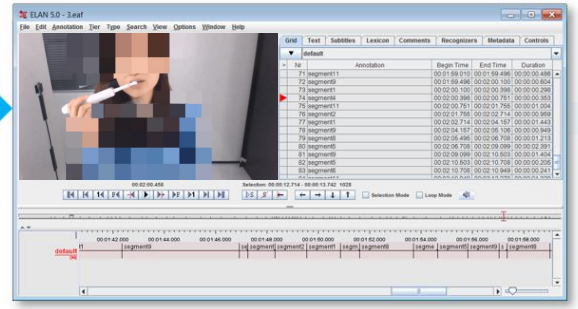
✓ **Create personalised experience**

Oral-B iO – Creating a Smart Toothbrush



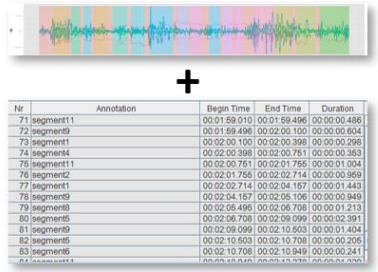
Hardware Development

Data Acquisition

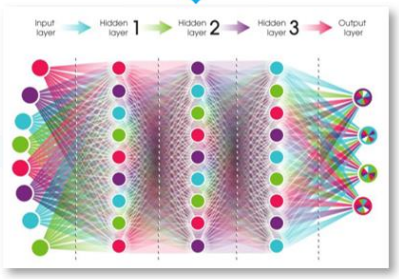


"ELAN (Version 5.0) [Computer software]. (2019). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan"

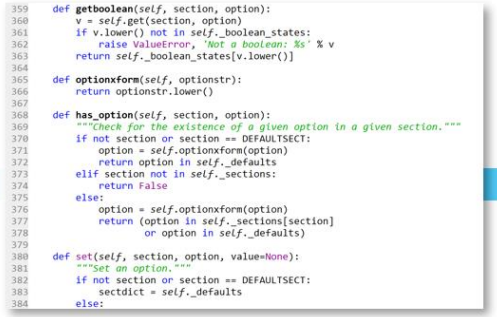
Data Labeling



Data QC and Merging



ML Algorithm Development



```
def getboolean(self, section, option):
    v = self.get(section, option)
    if v.lower() not in self._boolean_states:
        raise ValueError, 'Not a boolean: %s' % v
    return self._boolean_states[v.lower()]

def optionxform(self, optionstr):
    return optionstr.lower()

def has_option(self, section, option):
    """Check for the existence of a given option in a given section."""
    if not section or section == DEFAULTSECT:
        option = self.optionxform(option)
        return option in self._defaults
    elif section not in self._sections:
        return False
    else:
        option = self.optionxform(option)
        return (option in self._sections[section]
                or option in self._defaults)

def set(self, section, option, value=None):
    """Set an option."""
    if not section or section == DEFAULTSECT:
        sectdict = self._defaults
    else:
```

Library Integration [Android & iOS]

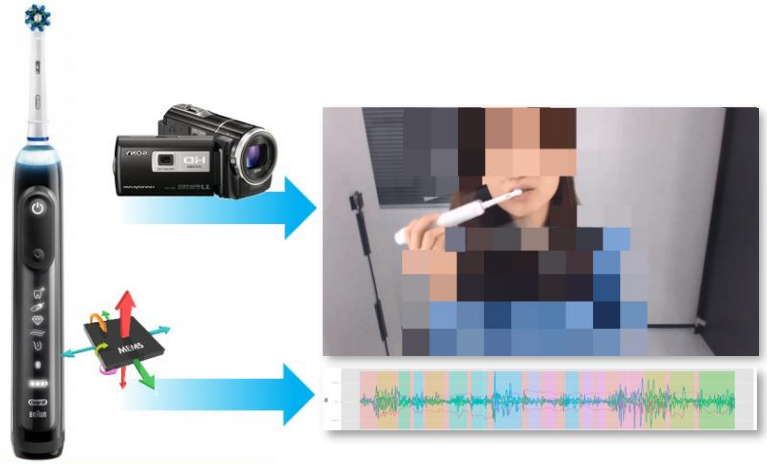
App Development

Consumer App

Bluetooth Data Stream

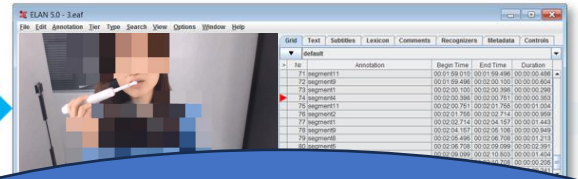
Recurrent NN

Some Stat & ML Questions on the Oral-B iO



Hardware Development

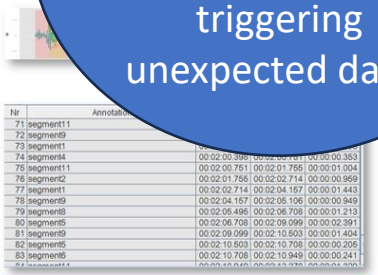
Data Acquisition



How to quantify where toothbrush is in mouth and how much pressure is being applied

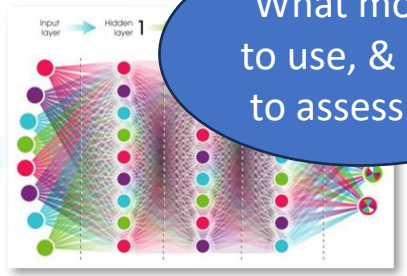
Data Labeling

What to do if a consumer brushes in an unexpected way, triggering unexpected data?



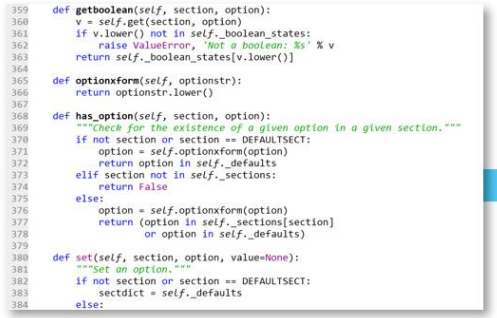
Data QC and Merging

What model to use, & how to assess fit?



ML Algorithm Development

Recurrent NN



Library Integration [Android & iOS]

App Development

Consumer App

Bluetooth Data Stream

Presentation by Matt Barker, Procter & Gamble, August 2023

It’s Statistical Engineering!

Complex?

- Yes

Unstructured?

- Yes, until we established structure

Multiple Disciplines:

- Dentistry
- Hardware / Sensor Development
- Signal Processing
- Statistics
- Machine Learning
- Programmers

Requiring Data and Data Analysis?

- Yes

Putting Statistical Engineering into Practice

Need to Learn Problem Solving *and* Collaboration Skills

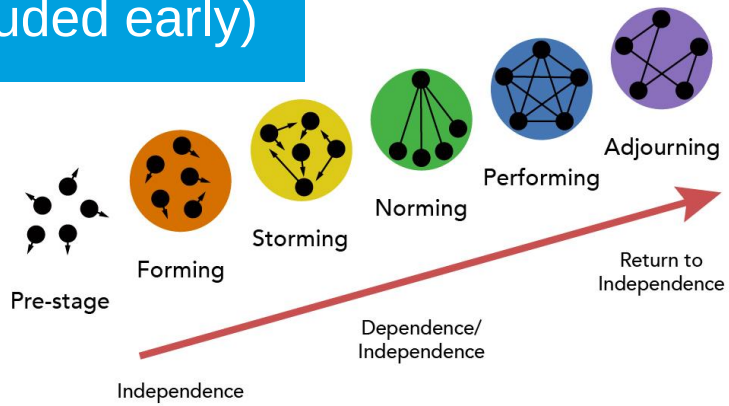
Typical MS/PhD Stats Curriculum:

Stat methods
Stat theory
Linear models
Stat programming
&
Stat consulting
(often focused on executing stats for a well-defined problem, and/or explaining stat concepts to client)

Problem Solving requires:

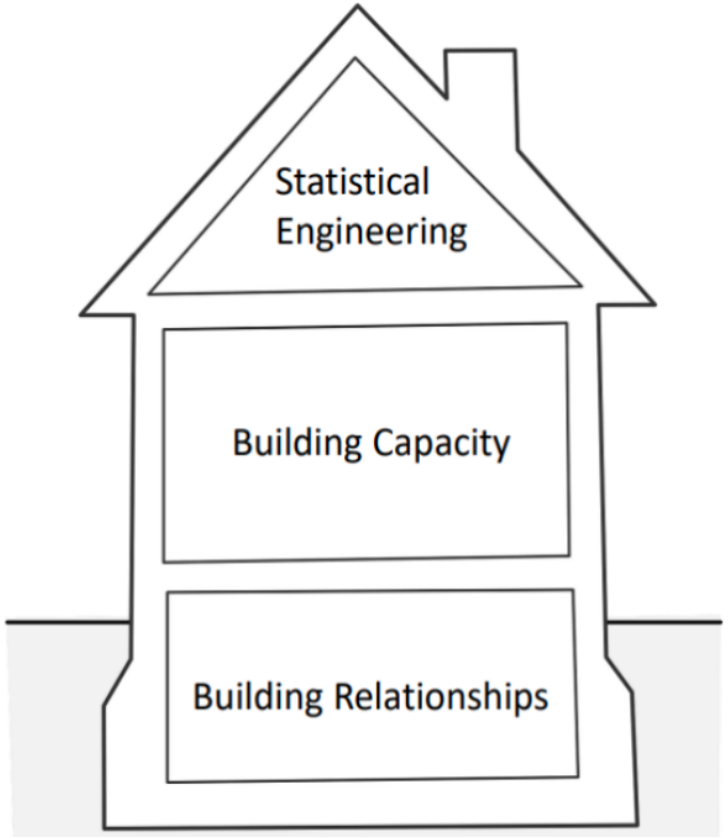
Teams to align on common problem-solving methodology
(If stats to be involved, they must learn it)

Teams to know when to include other disciplines
(Modeling not always included early)



Each Statistical Engineer Grows their Toolbox

NEED (AT A MINIMUM) STATISTICS, PROBLEM DEFINITION, PROJECT MANAGEMENT, COLLABORATION SKILLS



I broaden the idea of the house from Andressa Siroky & Carla Vivacqua (December 2023 ISEA webinar)

To be good statistical engineers, we need to develop across all of these areas

We need a continuous improvement mindset!

Facets of statistical engineering:

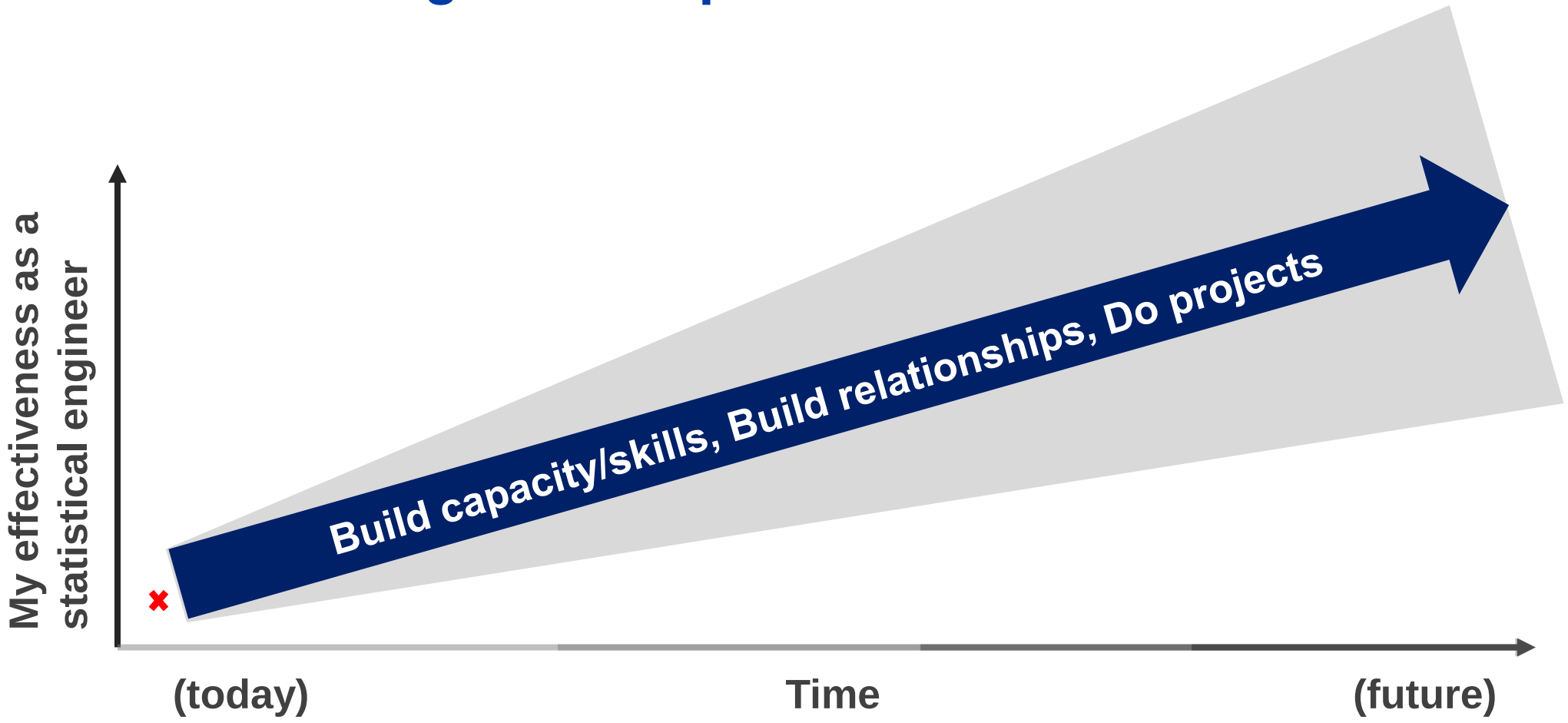
Capacity/skills:

- Statistical toolbox
- Project understanding:
  - Application domains
  - Project management
  - Problem definition and solving

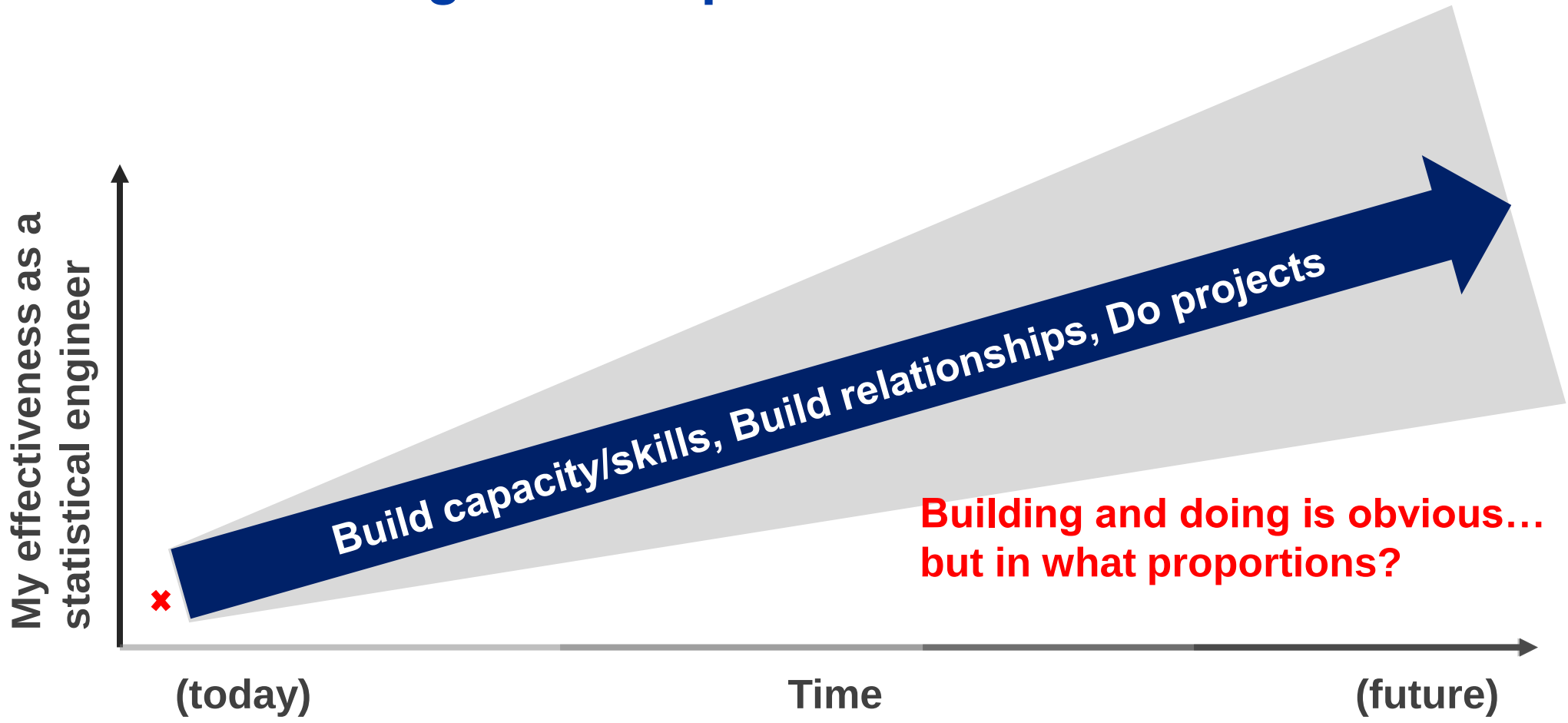
Relationships

- Collaboration
- Teamwork
- Communication

We Grow Through Our Experience



We Grow Through Our Experience



Choosing the Correct Mix to Create Impact & Results?

RESPONSE VARIABLE = AMOUNT OF PROJECT IMPACT

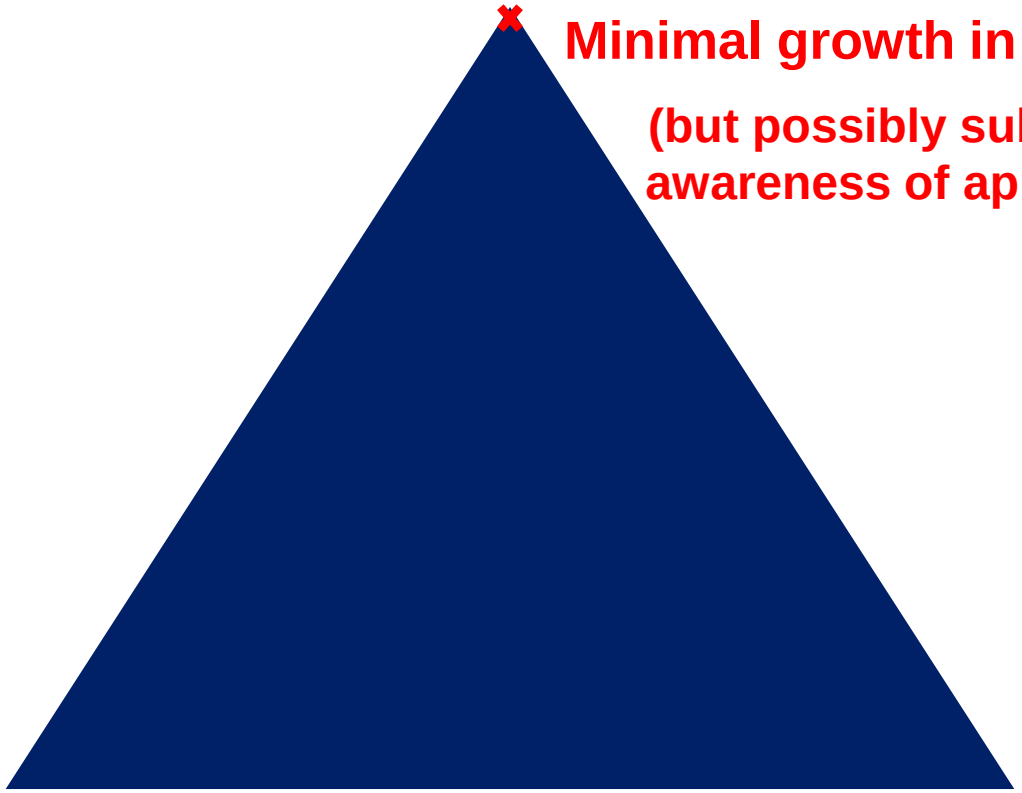
100% time spent growing relationships

100% time spent on statistical toolbox

100% time on problem understanding and execution

What's the Correct Mix to Create Impact? First "experiments"

100% time spent growing relationships



Minimal growth in stat engineering

(but possibly substantial growth in awareness of application domains)

100% time spent on statistical toolbox

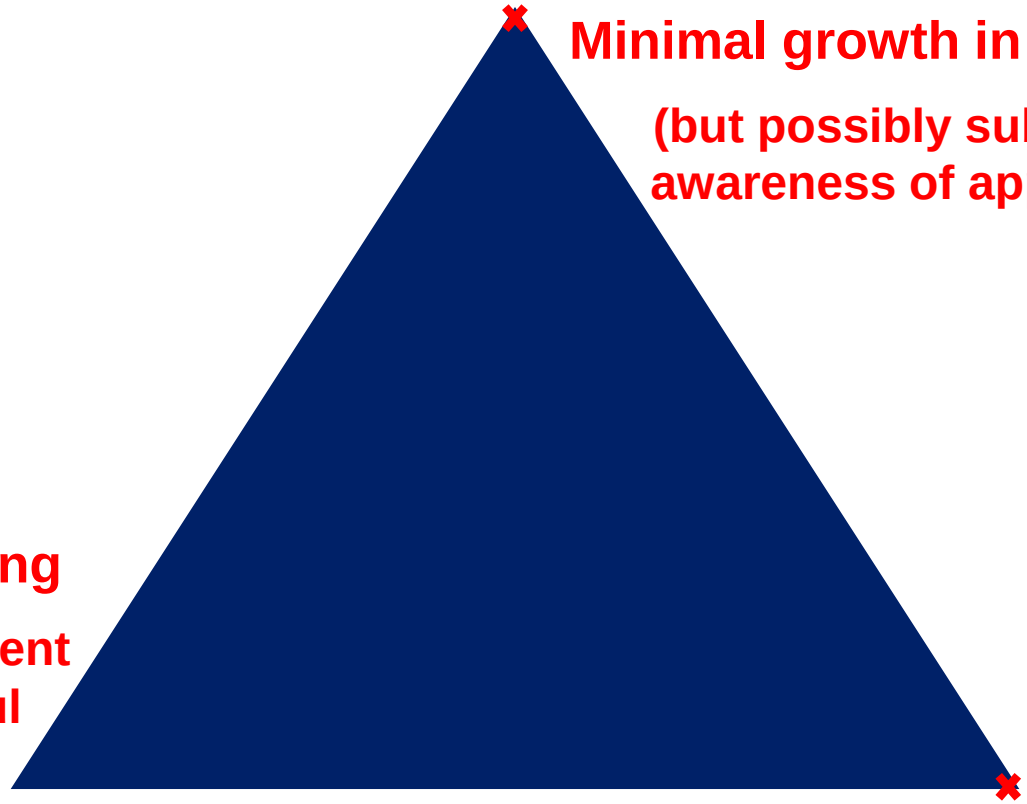
100% time on problem understanding and execution

What's the Correct Mix to Create Impact? First "experiments"

100% time spent growing relationships

Minimal growth in stat engineering

(but possibly substantial growth in awareness of application domains)



Avg. growth in stat engineering

(but potentially [1] too independent and [2] not delivering meaningful innovation/R&D outputs)

100% time spent on statistical toolbox

100% time on problem understanding and execution

What's the Correct Mix to Create Impact? First "experiments"

100% time spent growing relationships

Minimal growth in stat engineering

(but possibly substantial growth in awareness of application domains)

Avg. growth in stat engineering

(but potentially [1] too independent and [2] not delivering meaningful innovation/R&D outputs)

100% time spent on statistical toolbox

Min. growth in stat engineering

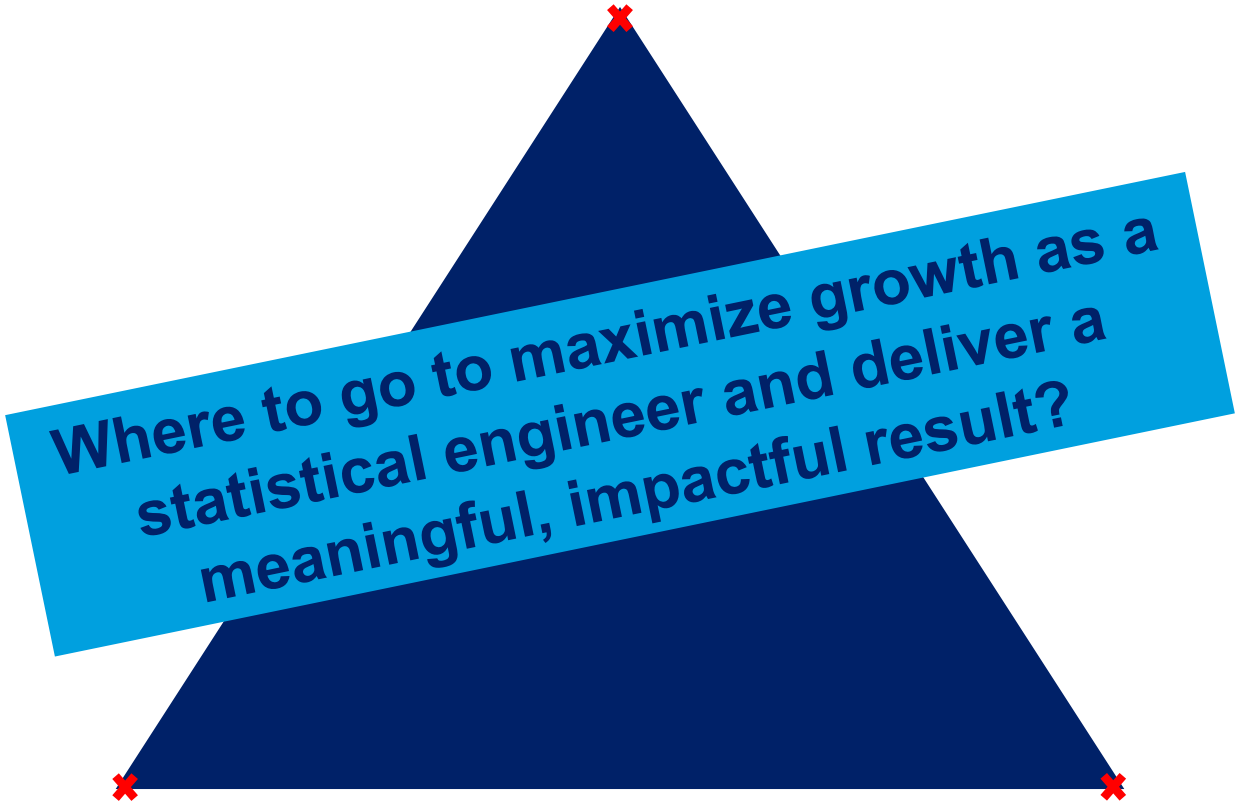
(unless already well trained in stats, project mgmt., problem definition, collaboration, …)

100% time on problem understanding and execution

What's the Correct Mix to Create Impact?

DEPENDS ON WHAT SKILLS YOU ALREADY HAVE AND/OR WHAT YOU NEED TO LEARN

100% time spent growing relationships



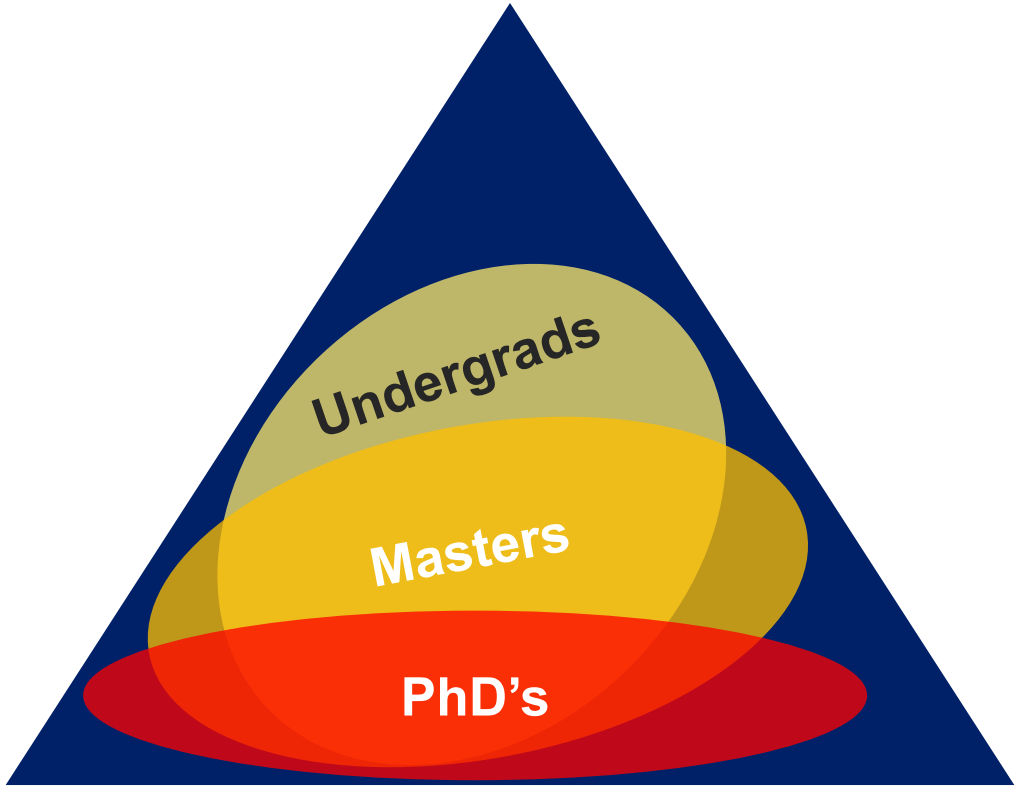
100% time spent on statistical toolbox

100% time on problem understanding and execution

Where Do Newly Graduated Statisticians Fit?

GENERALIZED FROM MY INTERVIEWS WHEN HIRING

100% time spent growing relationships



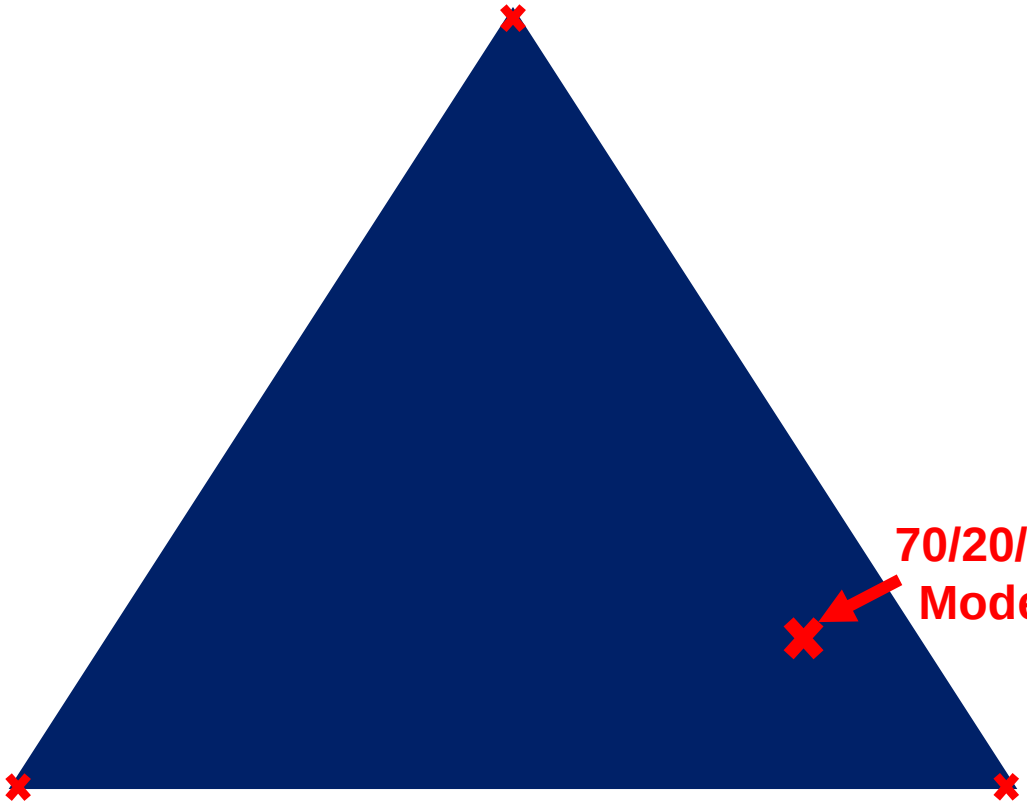
100% time spent on statistical toolbox

100% time on problem understanding and execution

70/20/10 Model

EXPERIENCES, THEN NETWORKING, THEN FORMAL LEARNING

100% time spent growing relationships

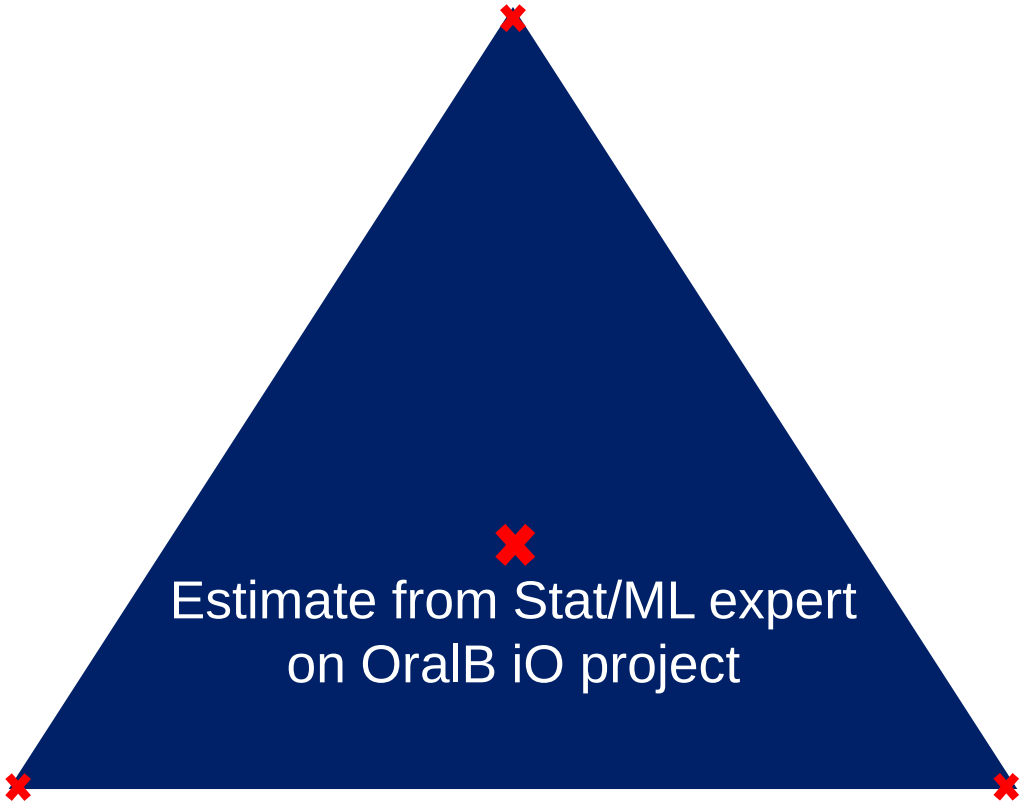


100% time spent on statistical toolbox

100% time on problem understanding and execution

Choosing the Correct Mix to Create Impact (point estimates)

100% time spent growing relationships

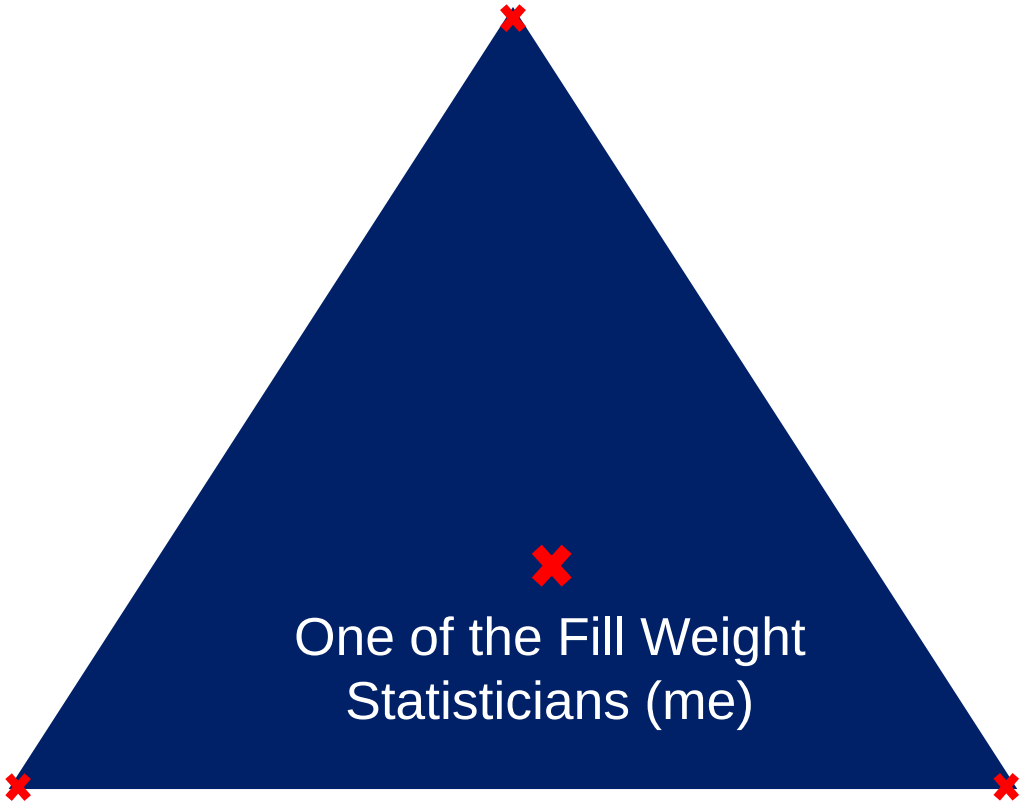


100% time spent on statistical toolbox

100% time on problem understanding and execution

Choosing the Correct Mix to Create Impact (point estimates)

100% time spent growing relationships



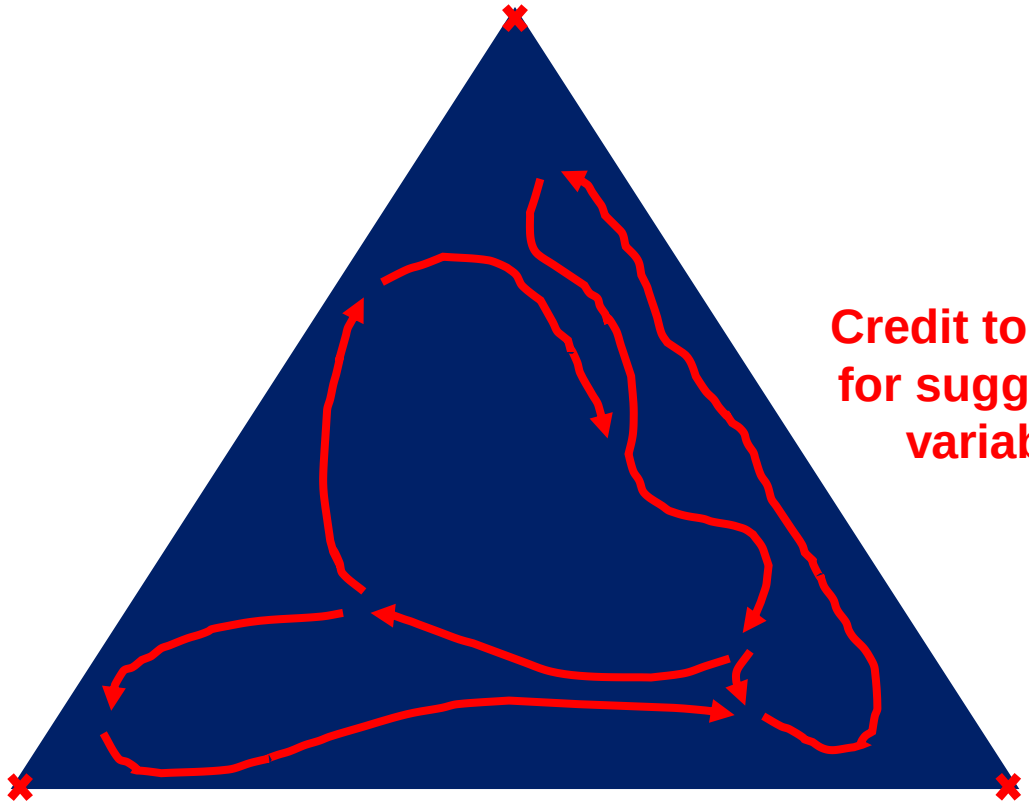
100% time spent on statistical toolbox

100% time on problem understanding and execution

The Correct Mix to Create Impact May Not Be Constant!

MANY POSSIBLE PATHS

100% time spent growing relationships



Credit to William Brenneman for suggesting this idea of a variable or cyclic path

100% time spent on statistical toolbox

100% time on problem understanding and execution

Building Statistical Engineering is a Job for All

Statisticians / Data scientists	• Be aware of the facets of statistical engineering • Define your own sweet spot for growth
Managers / Chairs / Leaders	• Make sure your data people are aware of the facets • Encourage them to find their sweet spot • Give them time for growth
Business/Industry and Government	• Demand the full range of the facets of statistical engineering skills when hiring (go beyond statistical skills)
Academicians	• Continue to grow emphasis on capstone projects, unstructured problem solving, interdisciplinary collaboration, etc.

Facets of statistical engineering:

Capacity/skills:

- Statistical toolbox
- Project understanding:
  - Application domains
  - Project management
  - Problem definition and solving

Relationships

- Collaboration
- Teamwork
- Communication

A special note for those in Business/Industry: Leverage Career Growth Programs

Since most companies offer employee development programs (partial result from simple Google search)…

roivant, Plenty, Trimble, Success Academy, Veeva, stryker, cruise, Heineken, amazon, paloalto networks, intuit, Adobe, Baker Hughes, ICSC Foundation, City National Bank, an RBC company, Chipotle Mexican Grill, Morgan Stanley, yelp, ATLASSIAN, PIXAR, macy's, Cooley, CUSHMAN & WAKEFIELD, airbnb, AmeriCorps, Allstate, randstad, workday, NBCUniversal, Etsy, box, CHUBB, Triage Consulting Group, BCG Boston Consulting Group, Deloitte., mastercard, Urban Company, MARRIOTT, salesforce, McMASTER-CARR, Owens Corning, Goldman Sachs, Dolby, TEXAS INSTRUMENTS, Cigna, M&T Bank, Honeywell, Chick-fil-A, Abbott, vitech, WELLS FARGO

… leverage your employee development program to prioritize statistical engineering skill development

Translate statistical engineering into language that resonates in your company

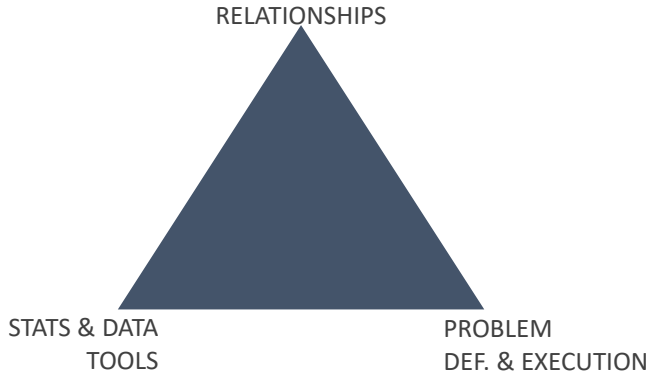
Explain to your leadership (or to your employees) why these skills matter

Declare growth areas

Incentivize and enable the growth

Don’t give up

Conclusion: Adopt a Continual Growth Mindset!



Embrace new statistics, data science, and data-related methods

- What can we compute now that was historically hard?
- How can we better experiment when we don't have much data yet?
- How can we better incorporate a continually growing data set?
- Data Science, Computer Science are not completely foreign to us

Build relationships and be more conversant with cross-disciplinary partners

- Understand their language and teach them to understand yours
- Enjoy learning from them so you understand what the data challenges are

Definition and execution that provide superior benefits

- Translate the unmet need into a problem you can help solve
- Solve the right problem
- High-quality, "irresistible" execution & results so everyone clamors for the solution!

Thank You!

Questions?

joner.md@pg.com