

Metrics and Measurement

17-313: Foundations of Software Engineering

<https://cmu-313.github.io>

Michael Hilton and **Chris Timperley**

Fall 2025

Administrivia

- Homework 1B **due tonight** at 11:59 pm
 - If you just joined the class, **let us know via email or Slack**, and we will extend the 1A deadline till tonight
- If you aren't on the Slack yet, send an email to us!
- We will reveal teams on Monday during recitation
 - if you haven't taken the Team Formation survey yet, see it on Canvas

Smoking Section

- Last full row



Learning Goals

- Explain the importance of measurement and metrics in Software Engineering
- Provide examples of metrics for software qualities and process
- Apply goal-based frameworks for decision making using metrics
- Identify the limitations and dangers of decisions and incentives based on measurements

Measurement in everyday life

- Economics
 - price, inflation rate, stock price, volume
- Medicine
 - heart rate, blood pressure, body temperature, ECG
- Engineering
 - force, torque, heat transfer coefficient, thermal efficiency
- Natural sciences
 - AQI, carbon footprint, soil pH



NS New Scientist

Ants use pedometers to find home

An experiment that involves attaching stilts to ants' legs reveals that the insects somehow keep a record of how many steps they take.

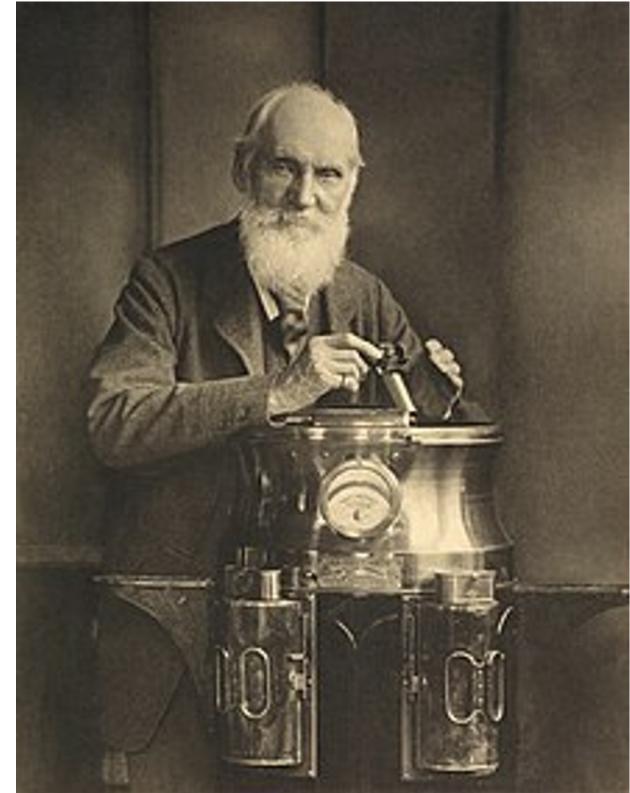
Jun 29, 2006



"To measure is to know; if you can not measure it, you can not improve it"

William Thomson, Lord Kelvin

$$K = \left(\frac{5}{9} (F - 32) \right) + 273.15$$



Software Development...

before Software Engineering



Software Engineering

Principles and practices (technical and non-technical) for
confidently building high-quality software

Software Engineering

Principles and practices (technical and non-technical) for
confidently building **high-quality software**

**What does this mean?
How do we know?**

**We need metrics
and measurement!**

Outline



Measurements and Metrics



How to use Measurements and Metrics



Case Study: Autonomous Vehicle Software



Risks and Challenges



Metrics and Incentives

Outline



Measurements and Metrics



How to use Measurements and Metrics



Case Study: Autonomous Vehicle Software



Risks and Challenges



Metrics and Incentives

What is Measurement?

What is Measurement?

- “Measurement is the empirical, objective assignment of numbers, according to a rule derived from a model or theory, to attributes of objects or events with the intent of describing them”

Kaner, Bond, *Software Engineering Metrics: What Do They Measure and How Do We Know?*

17th INTERNATIONAL SOFTWARE METRICS SYMPOSIUM, METRICS 2014
KANER / BOND - 1

Software Engineering Metrics: What Do They Measure and How Do We Know?

Cem Kaner, Senior Member, IEEE, and Walter P. Bond

Abstract—Contract validity is about the question, how well do we know what we are measuring? This paper is about contract validity in software engineering metrics. Contract validity is a concept that rarely has simpler ways than application by practitioners. Contract validity starts with a thorough analysis of the construct, the attributes, and the measures used to represent the construct. The analysis must consider the context in which measurement of attributes involves a metric that depends only on one or more of the attributes, but not on no software engineering artifacts or tools. In addition, the analysis must consider the context in which the metric is used. The context may be the need to measure an attribute, or the need to measure an attribute in order to measure another attribute. The context may be the need to measure an attribute they are being used to measure. Multidimensional metrics often prove useful as a means of capturing the “whole picture” of a system. We present some thoughts on this topic.

Index Terms—C.2.0 Software Engineering Metrics; Measurement; C.2.1.6 Software Engineering Measurement applied to SQA and Metrics

•

1 INTRODUCTION

We hear too often that for companies, establishing metrics programs, that focus second with failure, or worse, are programs have little to do with confirming to criteria established by management.

One could interpret this as evidence of the immaturity and simplicity of metrics. However, it is also evidence that many metrics programs (Fagan [1] estimates a cost of 4% of the development budget) are not well designed and are not necessarily correct. In other cases, however, metrics programs are well designed and are correct.

John Austin [2] provided an excellent discussion of the problems of measurement distortion and devaluation as goes into his book, *An Introduction to Formal Logic with Special Reference to contemporary Controversies in Quantitative Methodology*. We assert the Software Engineering as a field presents similar difficulties. In this paper, we will provide several examples of metrics that are well designed, but actually measure the attribute in question. This has likely caused confusion in the field. We will also provide several examples of metrics that are not well designed, and actually measure the wrong attribute. These metrics are introduced here to illustrate the importance of understanding what is linked to the attributes they are intended to measure, measure what is intended to measure, and not measure what is not intended to measure.

After justifying our basic assertions, we lay out a model for evaluating the validity and risk of a metric, and apply it to a few metrics common to the field. Not surprisingly given what is asserted, serious problems will show up.

In the final section of this paper, we suggest a different approach to metrics. We believe that the best way to think about measurement of an attribute of interest. The idea of “multidimensional metrics” is introduced. We provide several multidimensional examples that appear to have been used effectively and successfully. We hope that these examples will help you to see the utility emerge from these examples that, we hope, could emerge from your own metrics programs.

2 What are We Measuring?

2.1 Defining Measurement

To provide context for the next two sections, we need a definition of measurement. There are many definitions of measurement, but, we present several current definitions here. We will use the following definition:

“Measurement is the assignment of numbers to objects based on the results of comparison with one or more objects in effect to a normative.” [3, p. 47]

“Measurement is the assignment of numbers to objects based on the results of comparison with one or more objects in effect to a normative.” [3, p. 47]

“Measurement is the assignment of numbers to objects based on the results of comparison with one or more objects in effect to a normative.” [3, p. 47]

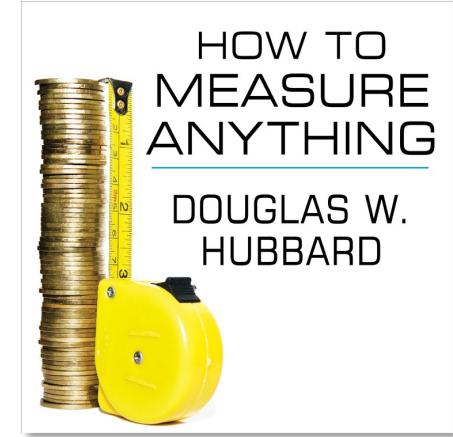
“Measurement is the process by which numbers or symbols are assigned to attributes of entities in the real world in accordance with agreed upon rules.” [4, p. 47]

“Measurement is the process of assigning a number or category to an entity to describe an attribute of the entity.” [5, p. 47]

* Cem Kaner is Professor of Software Engineering at the Florida Institute of Technology, Melbourne, FL, USA. Email: cem@fit.edu
† Walter P. Bond is a Research Scientist at the Florida Institute of Technology, Melbourne, FL, USA. Email: pbond@fit.edu

What is Measurement?

- “A quantitatively expressed **reduction of uncertainty** based on one or more observations.”
Hubbard, *How to Measure Anything* ...



What is Measurement?

- “A **software quality metric** is a function whose inputs are software data and whose output is a single numerical value that can be interpreted as the degree to which the software possesses a given attribute that affects its quality.”

IEEE 1061

Metric

MEASURES

Attribute

DESCRIBES

Entity

What is Measurement?

Method to obtain a number or symbol →

Metric

MEASURES

Quality of interest →

Attribute

DESCRIBES

Object or process →

Entity

Entities represent an Object or Process

Examples

- Software Product
- Modules
- Software Development Process
- People

Metric

MEASURES

Attribute

DESCRIBES

Entity

Attributes represent Qualities of Interest

Examples for Software Quality

- Reliability
- Security
- Scalability
- Extensibility
- Portability
- Availability
- Safety
- Observability
- Accuracy
- Robustness
- Resilience
- Timeliness
- Responsiveness
- Intuitiveness

Metric

MEASURES

Attribute

DESCRIBES

Entity

Attributes represent Qualities of Interest

Examples for Software Dev. Process

- Development efficiency
- Meeting efficiency
- Conformance to process
- Accuracy of predictions
- Fairness in decision making
- Regulatory compliance
- On-time release
- ...

Metric

MEASURES

Attribute

DESCRIBES

Entity

Attributes represent Qualities of Interest

Examples for People

Developers

- Productivity
- Agility
- Well-Being + Job Satisfaction
- Communication and Collaboration
- Creativity
- Morale
- Regulatory Compliance

End Users

- Product Satisfaction
- Ease of Use
- Feature Usage
- Regulatory Compliance

Metric

MEASURES

Attribute

DESCRIBES

Entity

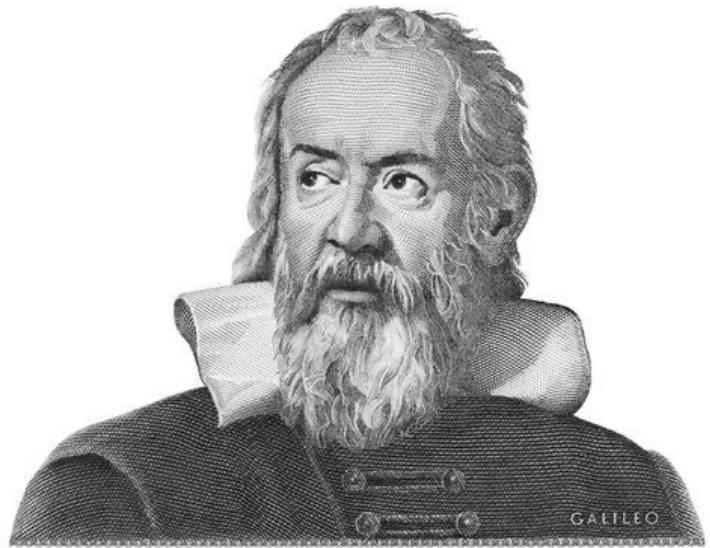
⚠ What about Non-Trivial Qualities?

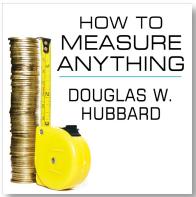
- **Software**
 - code elegance
 - code maintainability
- **Process**
 - development efficiency
 - fairness in decision making
- **Team**
 - productivity
 - collaboration
 - creativity



***“Measure what is
measurable, and make
measurable what is not so.”***

Galileo Galilei



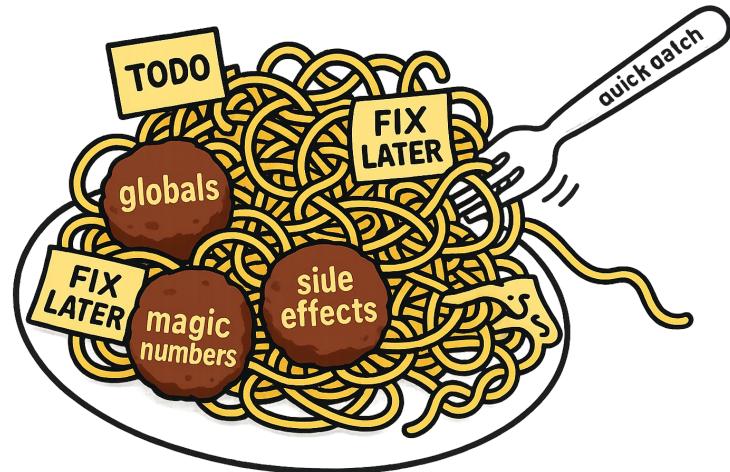


Everything is Measurable

- If we care about X, then, by definition, X **must be detectable**
 - how could we care about things like *quality*, *risk*, *security*, or *public image* if these things were totally undetectable, directly or indirectly?
 - if we have reason to care about some unknown quantity, it is because we think it corresponds to desirable or undesirable results in some way
- If X is detectable, then it **must be detectable in some amount**
 - if you can observe a thing at all, you can observe more of it or less of it
- If we can observe it in some amount, then it **must be measurable**

Douglas Hubbard, How to Measure Anything, 2010

Example: Code Complexity



thanks to ChatGPT for the terrible image

Number of Lines

- Easy to Measure! `wc -l file1 file2...`

LOC	projects
450	Expression Evaluator
2,000	Sudoku
100,000	Apache Maven
500,000	Git
3,000,000	MySQL
15,000,000	gcc
50,000,000	Windows 10
2,000,000,000	Google (MonoRepo)

Normalizing Lines of Code

- Ignore comments and empty lines
- Ignore lines < 2 characters
- Pretty print source code first
- Count statements (logical lines of code)
- See also: cloc

```
for (i = 0; i < 100; i += 1) printf("hello"); /* How many lines of code is this? */
```

```
/* How many lines of code is this? */
```

```
for (
    i = 0;
    i < 100;
    i += 1
){
    printf("hello");
}
```

Normalization by Language

Language	Statement factor (productivity)	Line factor
C	1	1
C++	2.5	1
Fortran	2	0.8
Java	2.5	1.5
Perl	6	6
Smalltalk	6	6.25
Python	6	6.5

Source: "Code Complete: A Practical Handbook of Software Construction", S. McConnell, Microsoft Press (2004)
and <http://www.codinghorror.com/blog/2005/08/are-all-programming-languages-the-same.html> u.a.

Halstead's Metrics (1977)

- Based on number of operators (e.g., `print`, `+`, `-`) and operands (name, “maurice”)
 - n_1 and n_2 : # of distinct operators and operands
 - N_1 and N_2 : # of instances of operators and operands
- Derived metrics include vocabulary (N), length (n), volume (V), difficulty (D), effort (E), time to write (T), number of bugs (B)

$$D = \frac{n_1}{2} + \frac{N_2}{n_2}$$

$$V = N \times \log_2 n$$

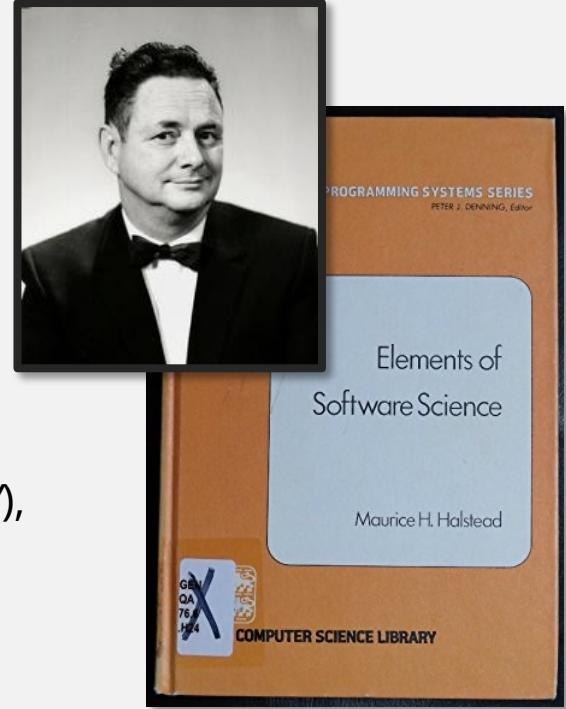
$$E = D \times V$$

$$N = N_1 + N_2$$

$$T = \frac{E}{18}$$

$$B = \frac{V}{3000}$$

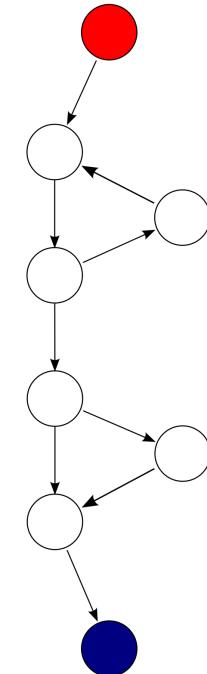
$$n = n_1 + n_2$$



McCabe Cyclomatic Complexity (1976)

- Computed according to **control flow** graph
 - tells you how many tests you need for full branch coverage
- Equal to **number of decision points + 1**
 - if, while, do-while, ?:, catch, switch, case
 - && and || in if condition

"For each module, either limit cyclomatic complexity to [X] or provide a written explanation of why the limit was exceeded."
NIST Structured Testing methodology



Object-Oriented Metrics (1994)

- Number of Methods per Class
 - Depth of Inheritance Tree
 - Number of Child Classes
 - Coupling between Object Classes
 - Calls to Methods in Unrelated Classes
 - ...

476 IEEE TRANSACTIONS ON SOFTWARE ENGINEERING, VOL. 20, NO. 6, JUNE 1994

Takeaways: Code Complexity

- There are **lots of complexity metrics** that measure different things, which may or may not be what we care about
 - there is no single value that fully captures “code complexity”
 - most of these metrics are confounded by code size!
- We need to be **intentional** about what we are measuring
 - how hard is this code to read?
 - how difficult is this code to maintain?
 - does this function need to be refactored?

Outline



Measurements and Metrics



How to use Measurements and Metrics



Case Study: Autonomous Vehicle Software



Risks and Challenges



Metrics and Incentives

"The key result has to be measurable. But at the end you can look, and without any arguments: **Did I do that or did I not do it?** Yes? No? Simple. No judgments in it."

Andy Grove, Inventor of OKRs

What are OKRs?

OBJECTIVES AND KEY RESULTS

OBJECTIVES

An Objective is what you want to accomplish.

A good Objective is significant, concrete, action-oriented and inspirational. Can be set annually or over an even longer-term.

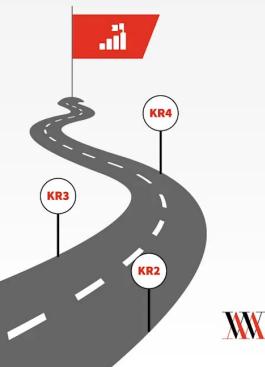
KEY RESULTS

Key Results are how you will accomplish it.

Good Key Results are specific, timebound, aggressive yet realistic measurable and verifiable. Can be set quarterly and evolve as work progresses.

What Matters

OKRs are a management methodology which helps to ensure that your company focuses efforts on the same important issues throughout the organization.



Measure What Matters

3.8721 in

2.6172 in

How Google, Bono, and the Gates Foundation Rock the World with **OKRs**

John Doerr

WITH A FOREWORD BY LARRY PAGE

"OKRs have helped lead us to 10x growth, many times over."

Larry Page, CEO of Google

"Every measurement action must be motivated by a particular goal or need that is clearly defined and easily understandable."

Software Metrics: A Rigorous and Practical Approach
N.Fenton, J.Bieman

CHAPMAN & HALL/CRC INNOVATIONS IN
SOFTWARE ENGINEERING AND SOFTWARE DEVELOPMENT

Software Metrics

A Rigorous and Practical Approach

THIRD EDITION

Norman Fenton
James Bieman

 CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Goal-Based Frameworks

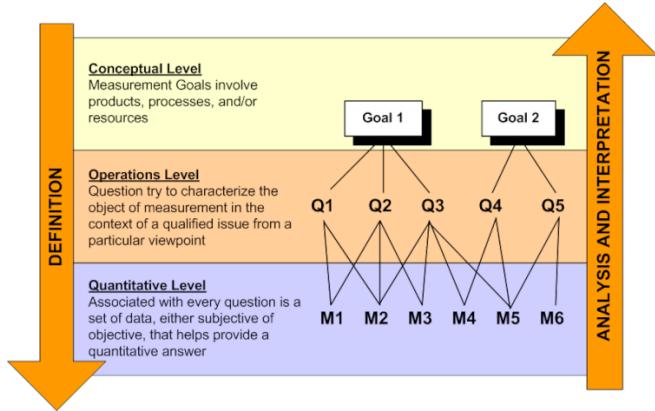
- Objectives and Key Results (OKRs)
- **Goal-Question-Metric (GQM)**
- Assurance Cases
- ...

The GQM Framework

Goal: What do you want to achieve?

Questions: What do you need to answer to know whether your goal is met?

Metrics: What measurements do you need to answer each question?



THE GOAL QUESTION METRIC APPROACH

Victor R. Basili¹ Gianluigi Caldiera¹ H. Dieter Rombach²

(¹) Institute for Advanced Computer Studies
Department of Computer Science
University Of Maryland
College Park, Maryland

(²) FB Informatik
Universität Kaiserslautern
Kaiserslautern, Germany

GQM: Defining Goals

P: **Purpose** (improve, evaluate, monitor, ...)

I: **Issue** (reliability, usability, effectiveness, ...)

O: **Object** (final product, component, process, activity)

V: **Viewpoint** (any stakeholder)

Goal:

Evaluate the effectiveness of the organization's coding standard from the team's perspective

Questions:

How comprehensible are the coding standards?

What is the impact of coding standards on the efficiency and productivity of the team?

Metrics:

Survey results measuring team members' understanding

Number of revisions required to achieve standard compliance

Code size:
of lines of code,
of classes,
of functions

Measurement for Decision Making

- Fund project?
- More testing?
- Fast enough? Secure enough? Safe enough?
- Code quality sufficient?
- Which feature to focus on?
- Should we refactor the code? Rewrite?
- Developer bonus?
- Time and cost estimation? Predictions reliable?

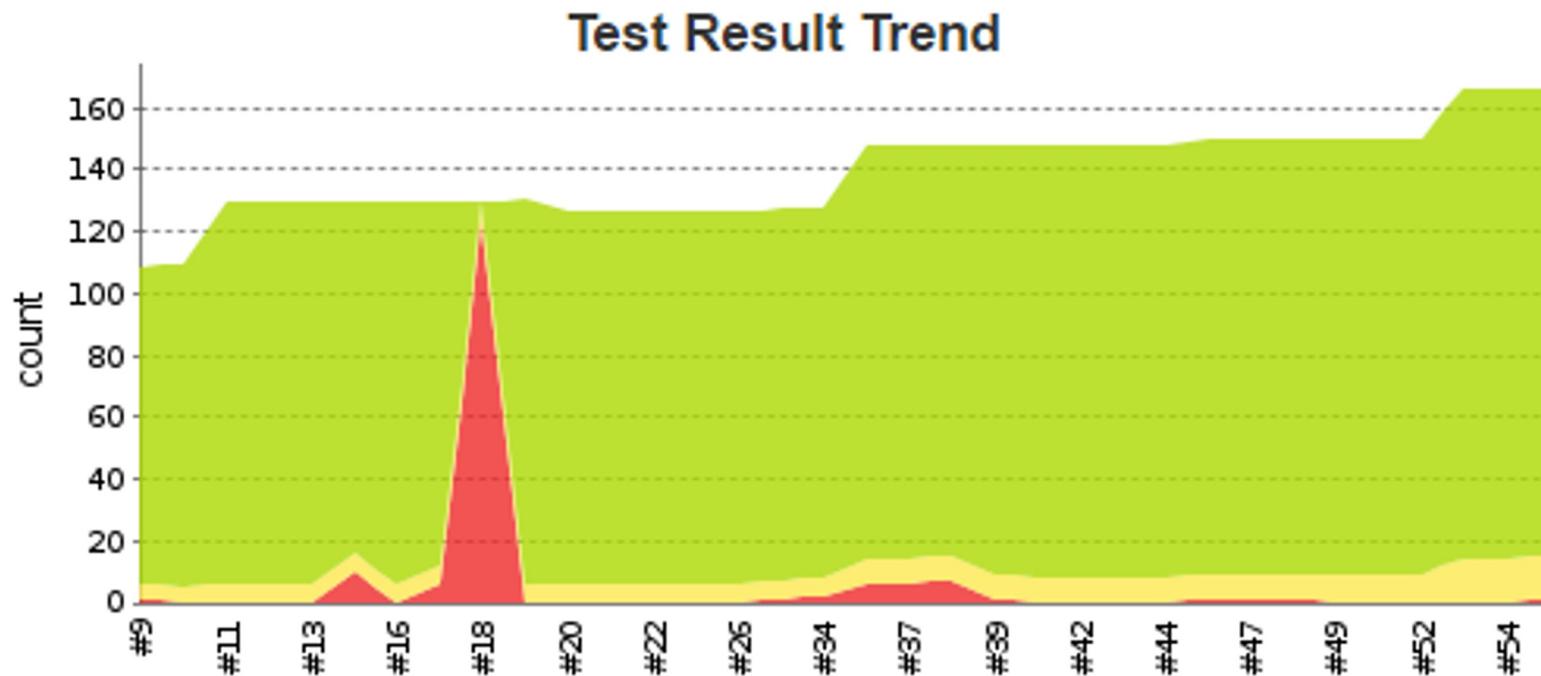
Analysis and Interpretation

Challenge: How do we actually interpret the measurements from our metrics?

Calibrate thresholds based on feedback

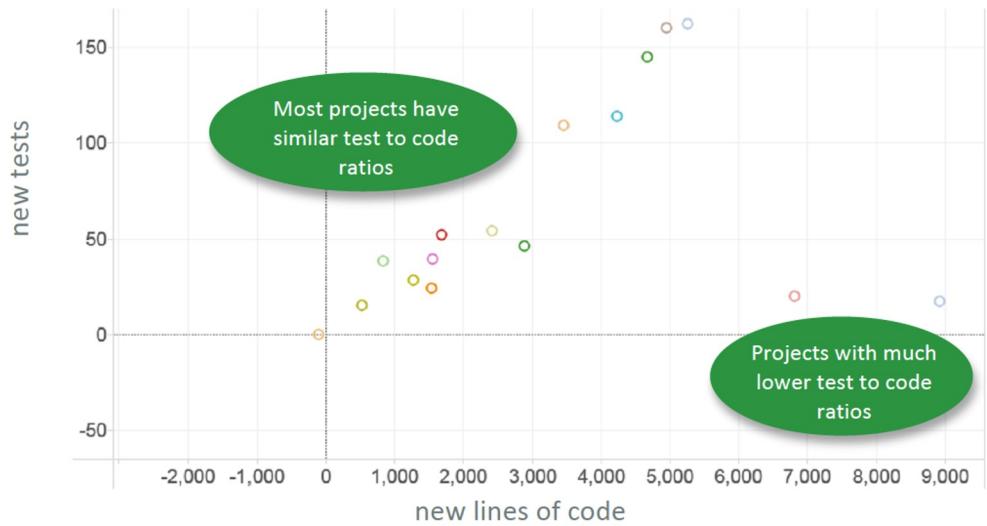
- **Problem:** "I'll know it when I see it"
 - "That function is way too complex!"
- **Solution:** Use existing code to obtain a set of thresholds
 - pick exemplars (clear pass / fail / borderline)
 - present to team, track agreement, and derive thresholds
 - E.g., cyclomatic complexity shouldn't be higher than 10
 - periodically recalibrate over time (e.g., technology changes)

Track metrics over time to assess trends



Benchmark against existing standards

- Monitor **similar** products, projects, modules, teams
- Refer to **external standards** if possible
 - e.g., ISO 26262
- Record typical values for metrics of interest
- Investigate deviations



Outline



Measurements and Metrics



How to use Measurements and Metrics



Case Study: Autonomous Vehicle Software



Risks and Challenges



Metrics and Incentives

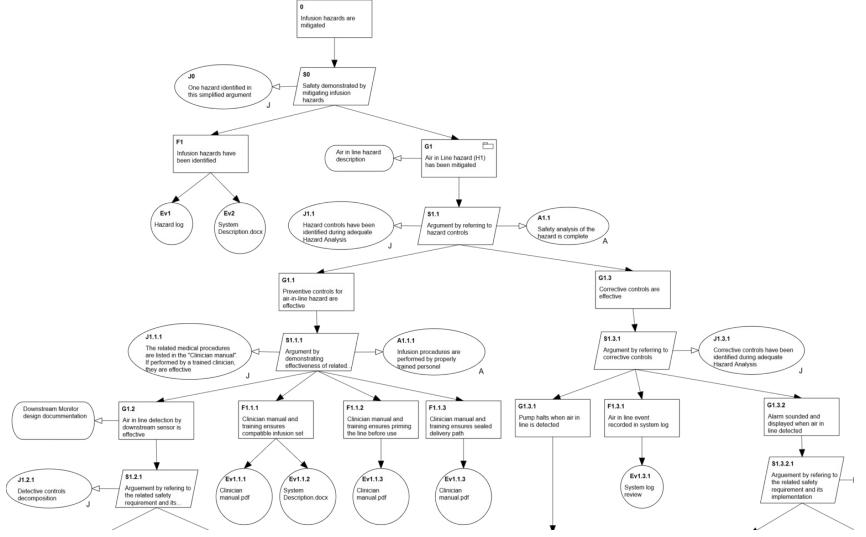
Safety Assurance Cases and AVs

Executing our safety case

At Aurora, we are using a safety case-based approach to inform, guide, and determine that our technology is acceptably safe to operate on public roads.

A safety case is a logical argument, supported by evidence, intended to justify that a system is acceptably safe for a specific application in a specific operating environment. A structured argument includes a specific claim—such as that our self-driving vehicles are acceptably safe to operate on public roads—that is broken down into sub-claims, which are then ultimately supported by evidence. We believe a safety case is the most effective and efficient path to safe driverless operations, and is imperative for any company looking to safely deliver commercial-ready self-driving vehicles at scale.

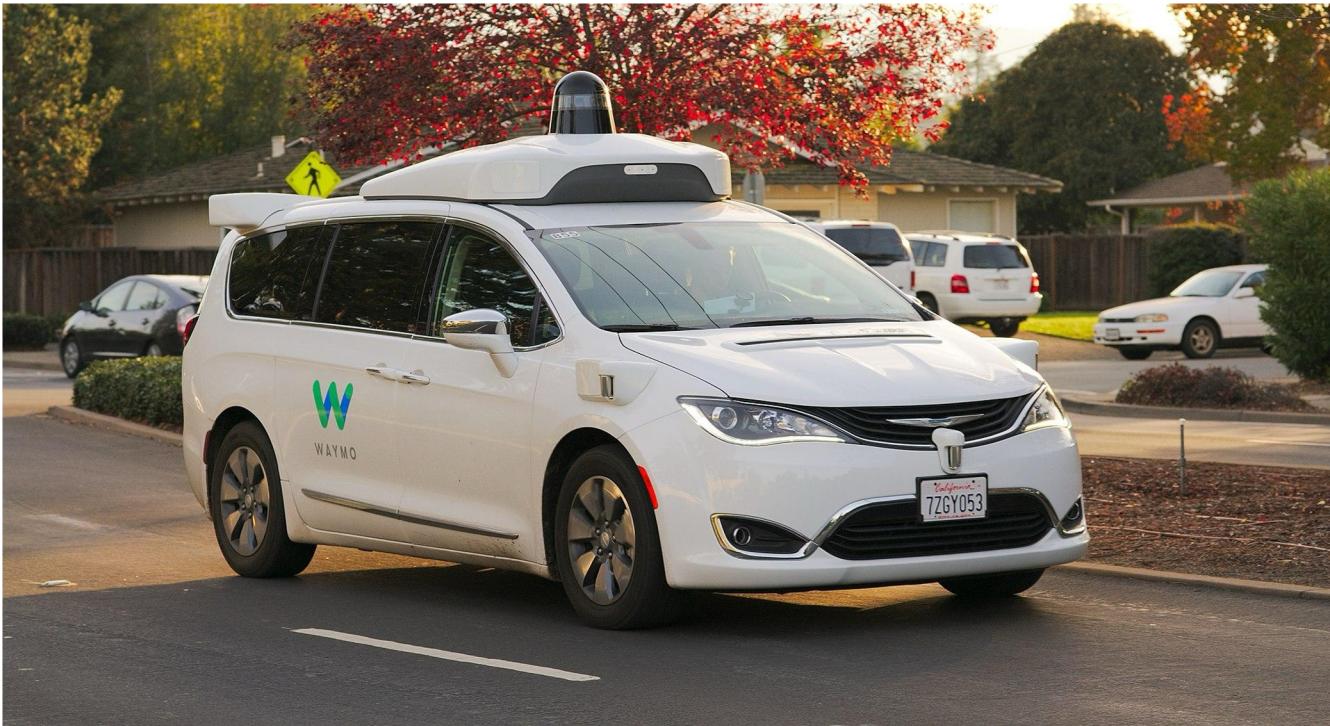
and ready to take over as necessary to ensure operational safety. Therefore our tailored safety case for this use case includes claims focused on vehicle controllability and vehicle operator hiring, training, and operational procedures, among others. However, as we approach the point of removing vehicle operators from the vehicles, these vehicle operator-centric claims will no longer be relevant. At that point, we will have completed a tailored safety case to include other claims from the Safety Case Framework related to demonstrating acceptably safe driverless operations within our ODD.



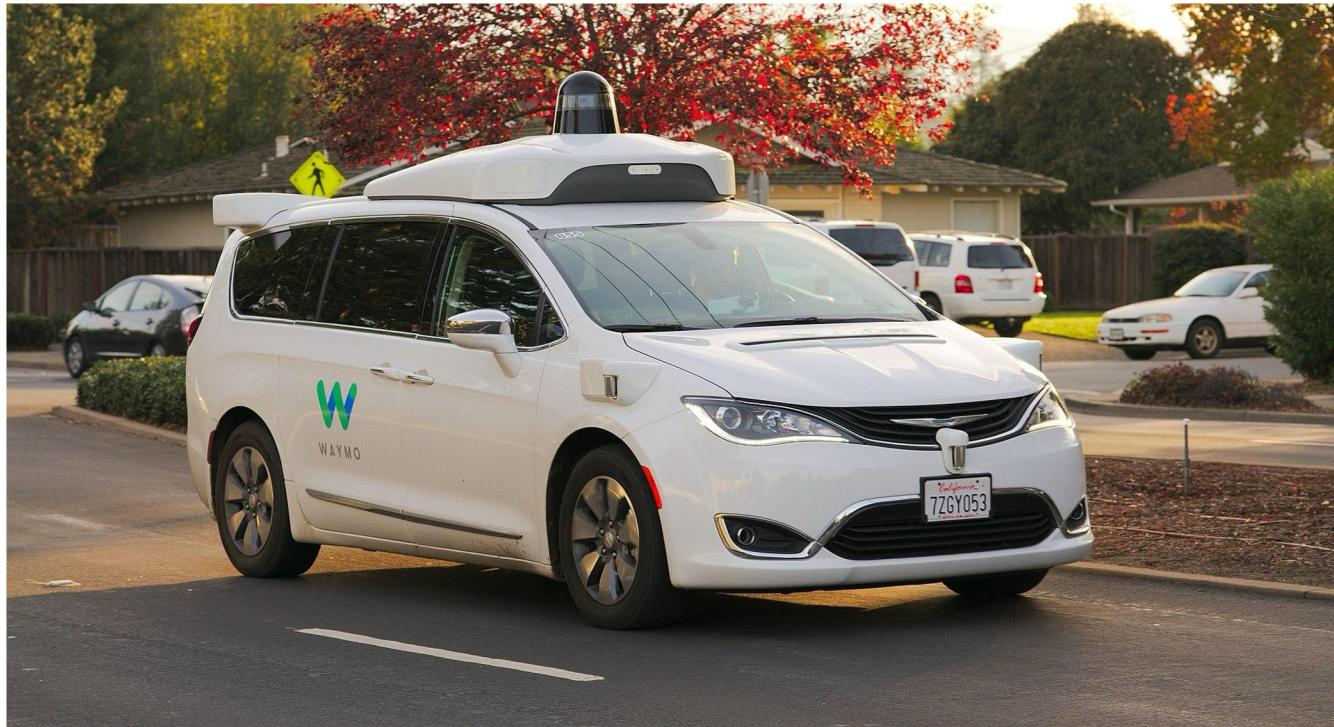
<https://aurora.tech/vssa>



AV Software is _____



By what Metrics can we judge AV software (e.g., safety)?



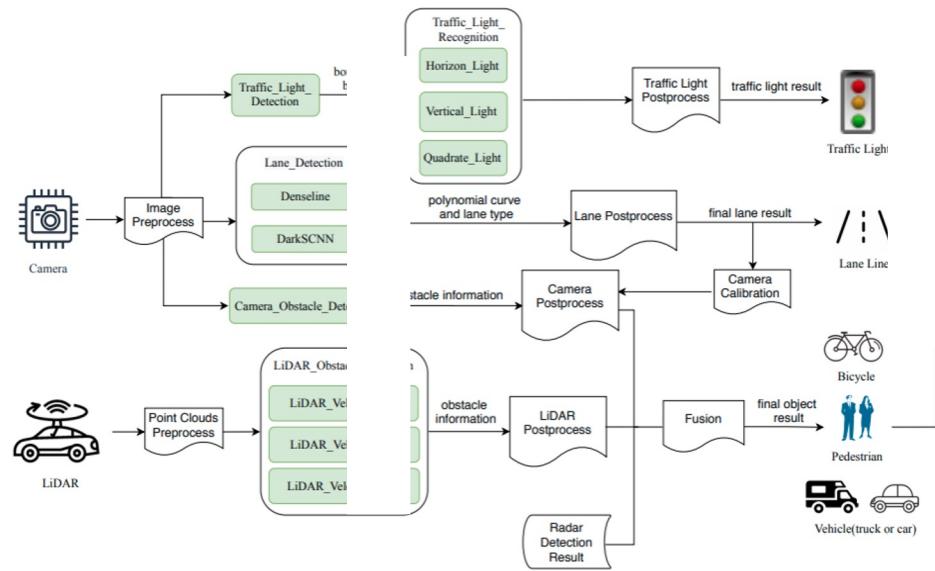
(1) Code Coverage

- Amount of code executed during testing.
- Statement coverage, line coverage, branch coverage, etc.
- E.g., 75% branch coverage ☐ 3/4 if-else outcomes have been executed

```
1698 : const TrajectoryPoint& StGraphData::init_point() const { return init_point_; }
:
2264 : const SpeedLimit& StGraphData::speed_limit() const { return speed_limit_; }
:
212736 : double StGraphData::cruise_speed() const {
212736 :     return cruise_speed_ > 0.0 ? cruise_speed_ : FLAGS_default_cruise_speed;
:
1698 : double StGraphData::path_length() const { return path_data_length_; }
:
1698 : double StGraphData::total_time_by_conf() const { return total_time_by_conf_; }
:
1698 : planning_internal::STGraphDebug* StGraphData::mutable_st_graph_debug() {
1698 :     return st_graph_debug_;
:
566 : bool StGraphData::SetSTDivableBoundary(
:
566 :     const std::vector<std::tuple<double, double, double>& s_boundary,
566 :     const std::vector<std::tuple<double, double, double>& v_obs_info) {
[ + - ]: 566 :     if (s_boundary.size() != v_obs_info.size()) {
[ + - ]:         return false;
:
40752 :     for (size_t i = 0; i < s_boundary.size(); ++i) {
80372 :         auto st_bound_instance = st_drivable_boundary.add_st_boundary();
160744 :         st_bound_instance->set_t(std::get<0>(s_boundary[i]));
120558 :         st_bound_instance->set_s_lower(std::get<1>(s_boundary[i]));
120558 :         st_bound_instance->set_s_upper(std::get<2>(s_boundary[i]));
[ - + ]: 40186 :         if (std::get<1>(v_obs_info[i]) > -kObsSpeedIgnoreThreshold) {
[ - + ]:             0 :             st_bound_instance->set_v_obs_lower(std::get<1>(v_obs_info[i]));
:
40186 :         if (std::get<2>(v_obs_info[i]) < kObsSpeedIgnoreThreshold) {
50254 :             st_bound_instance->set_v_obs_upper(std::get<2>(v_obs_info[i]));
:
:
}
```

(2) Model Accuracy

- Train machine-learning models on labelled data (sensor data + ground truth)
- Compute accuracy on a separate labelled test set
- E.g., 90% accuracy implies that object recognition is right for 90% of the test inputs.



Source: Peng et al. ESEC/FSE'20

(3) Failure Rate

- Frequency of crashes / fatalities
- Per 1,000 rides, per million miles, per month (in the news)

TRANSPO / WAYMO / TECH

Waymo's driverless cars were involved in two crashes and 18 'minor contact events' over 1 million miles



Image: Allen J. Schaben / L

/ The Alphabet-owned company pulls back the curtain on more stats from its public road testing. Of the 20 incidents, only two met the federal government's reporting criteria, and no one was injured.

By Andrew J. Hawkins, transportation editor with 10+ years of experience who covers EVs, public transportation, and aviation. His work has appeared in The New York Daily News and City & State.

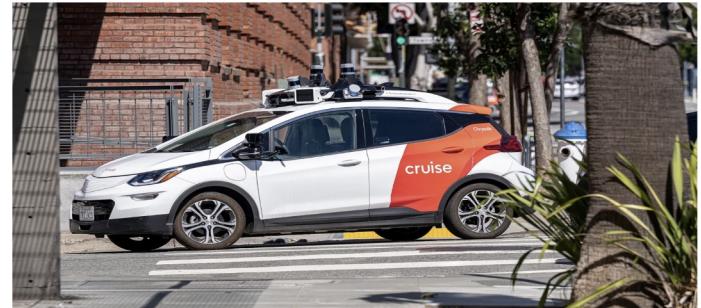
Feb 28, 2023, 8:00 PM GMT-3 | □ 1 Comment / 1 News



'Complete meltdown': Driverless cars in San Francisco stall causing a traffic jam

By Jordan Valinsky, CNN Business

Updated 3:45 PM EDT, Mon August 14, 2023



A Cruise autonomous taxi in San Francisco.

David Paul Morris/Bloomberg/Getty Images



(4) Mileage

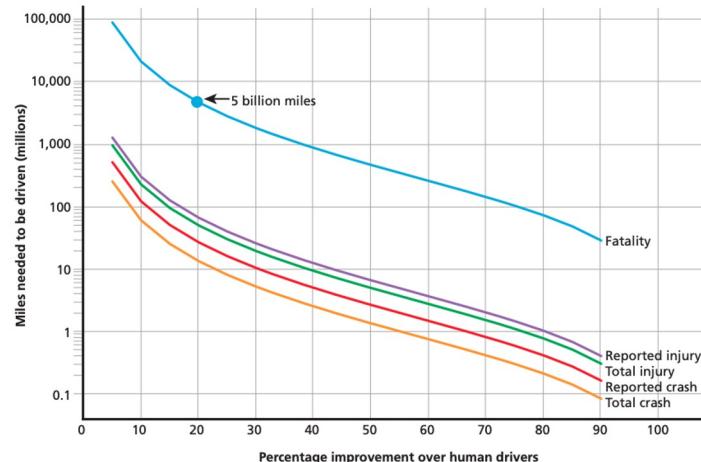


Driving to Safety

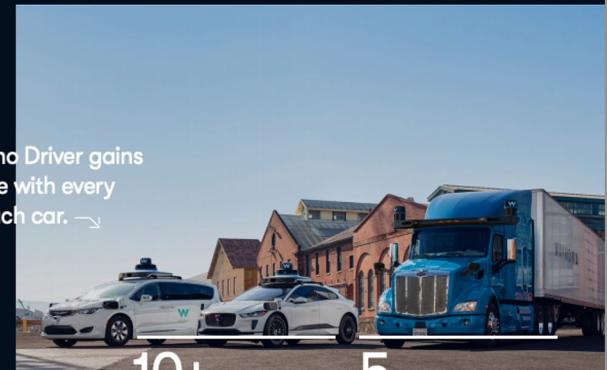
How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?

Nidhi Kalra, Susan M. Paddock

Figure 3. Miles Needed to Demonstrate with 95% Confidence that the Autonomous Vehicle Failure Rate Is Lower than the Human Driver Failure Rate



Building the World's Most Experienced Driver™



Source: waymo.com/safety (September 2021)

Participation Activity

- Apply the **Goal-Question-Metric** framework to explore various aspects of AV software
- Define one goal, two questions, and at least one metric per question
- Write it down on a piece of paper with your Andrew ID(s) on it.
- You can work in **groups of 2 to 3**
- Share with the class!

Example

Goal: Ensure energy efficiency and sustainability from the point of view of the organization and environmental analysts



Q1: What is the energy consumption under different driving conditions?



Metrics: Miles per kWh
{city, highway, mixed}
{winter, summer, fall}



Q2: How efficient is the battery management system?



Metrics: Battery life
number of charge cycles
peak battery temperature

Outline



Measurements and Metrics



How to use Measurements and Metrics



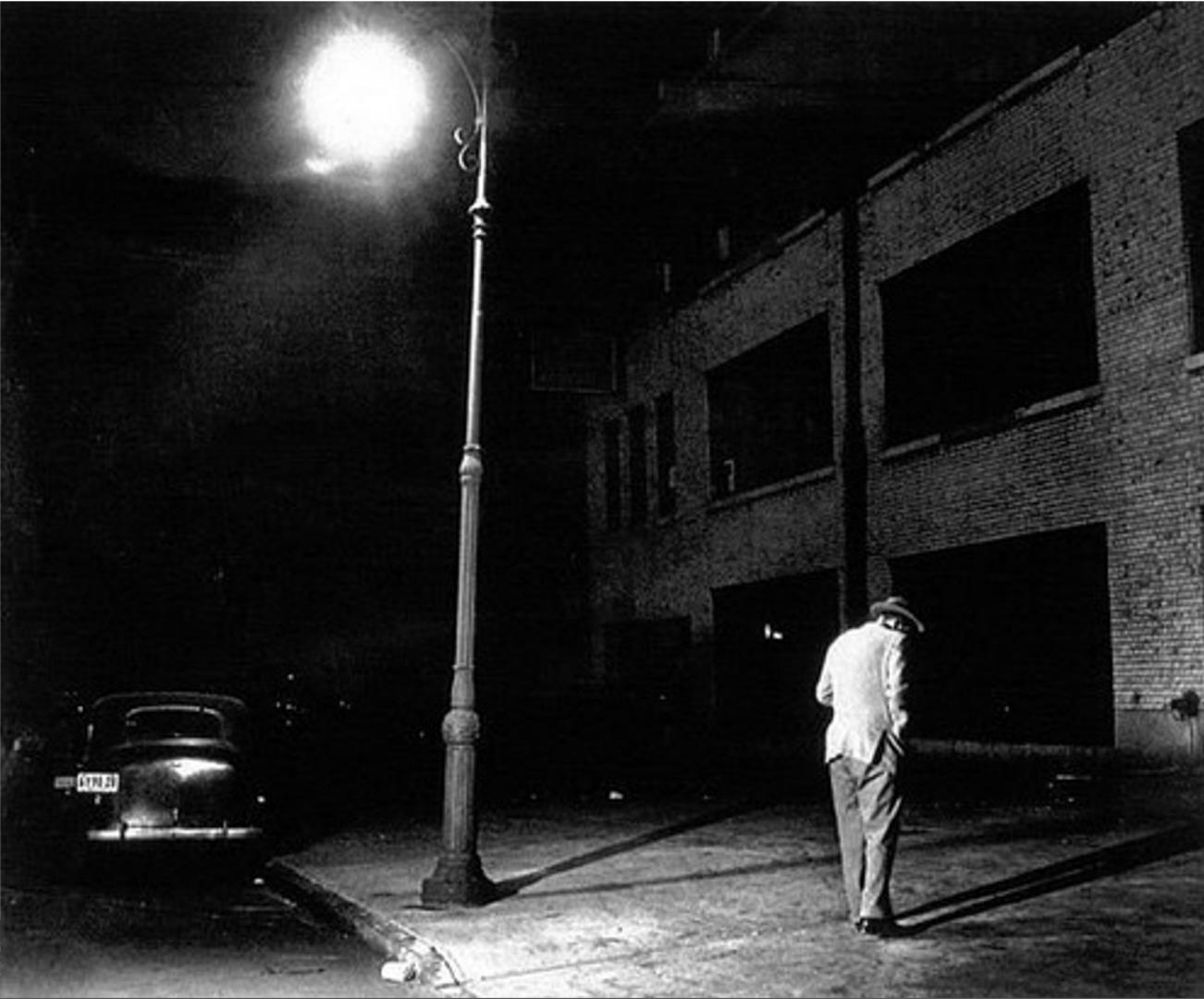
Case Study: Autonomous Vehicle Software



Risks and Challenges



Metrics and Incentives

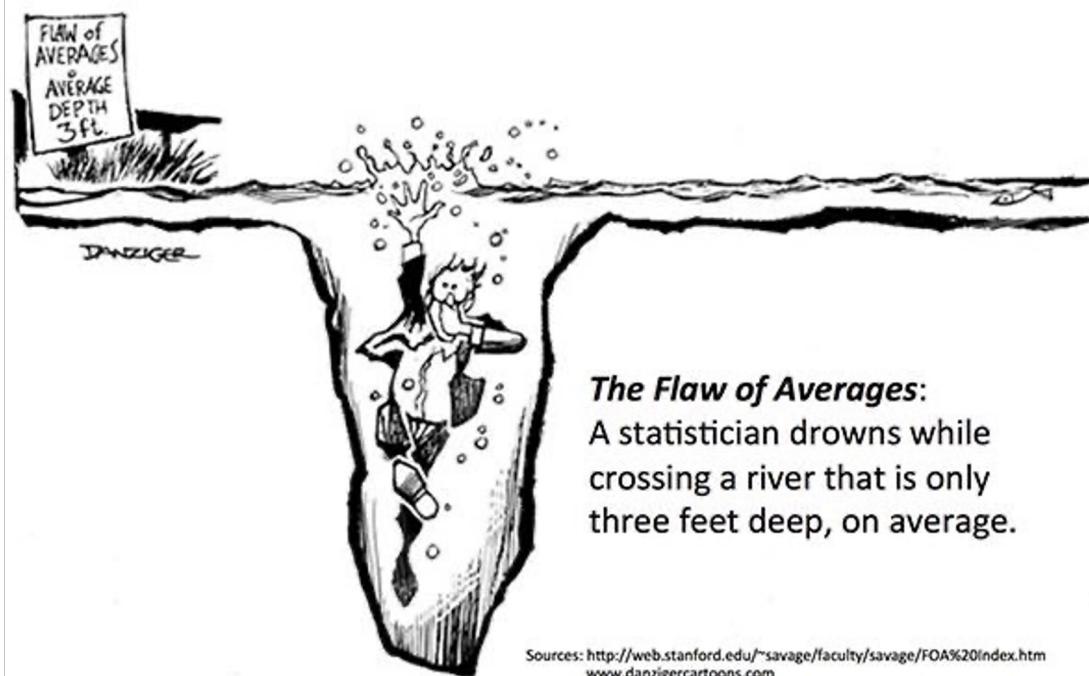


The Streetlight Effect



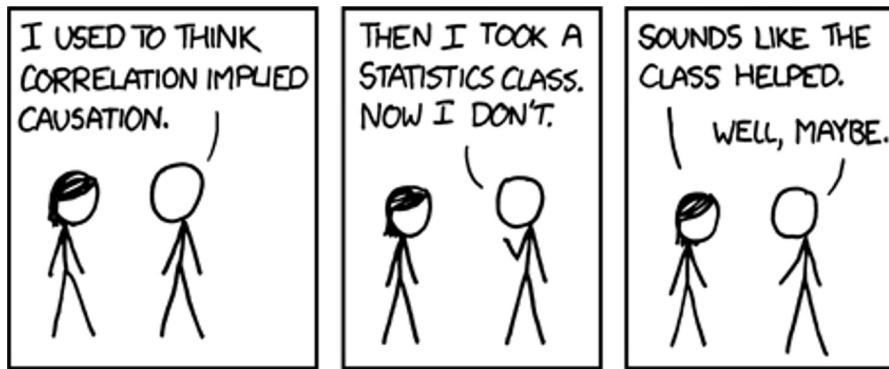
- A known **observational bias**
- People tend to look for something only where it's easiest to do so
 - If you drop your keys at night, you'll tend to look for it under streetlights

Bad Statistics: What could possibly go wrong?



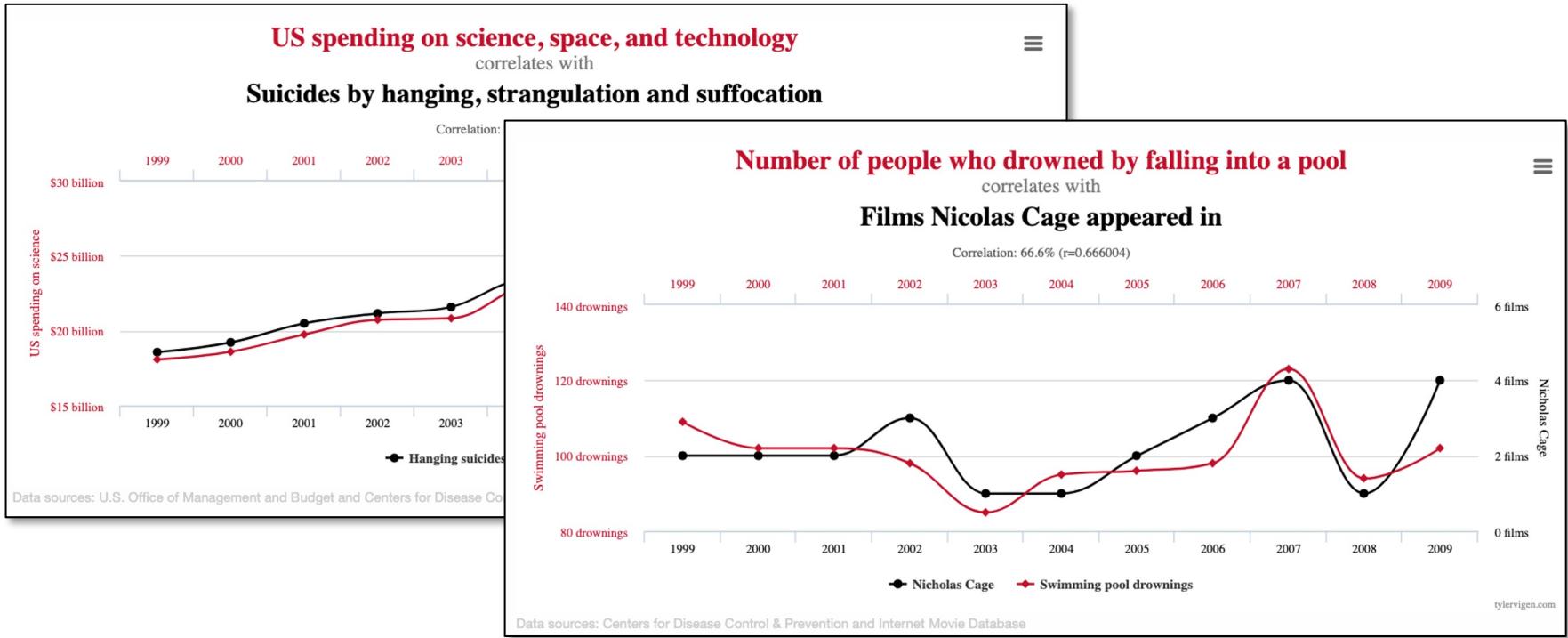
Sources: <http://web.stanford.edu/~savage/faculty/savage/FOA%20index.htm>
www.danzigercartoons.com

Making Inferences

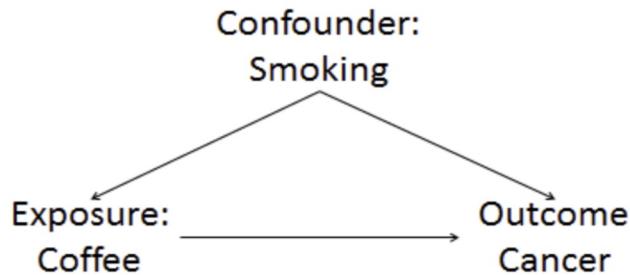
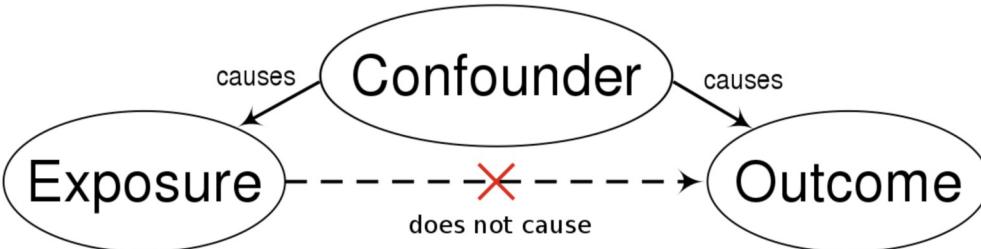


- To infer causation:
 - Provide a theory (from domain knowledge, independent of data)
 - Show correlation
 - Demonstrate ability to predict new cases (replicate/validate)

Spurious Correlations



Confounding Variables



- If you look only at the coffee consumption → cancer relationship, you can get very misleading results
- Smoking is a confounder

Coverage is not strongly correlated with test suite effectiveness

Authors:  [Laura Inozemtseva](#),  [Reid Holmes](#) [Authors Info & Affiliations](#)

ICSE 2014: Proceedings of the 36th International Conference on Software Engineering • May 2014 • Pages 435–445 • <https://doi.org/10.1145/2568225.2568271>

“We found that there is a low to moderate correlation between coverage and effectiveness when the number of test cases in the suite is controlled for.”

61

Most prior studies didn't account for the confounding influence of test suite size

Measurement Reliability

- Extent to which a measurement yields similar results when applied multiple times
- Goal is to reduce uncertainty, increase consistency
- Example: Performance
 - Time, memory usage
 - Cache misses, I/O operations, instruction execution count, etc.
- Law of large numbers
 - Taking multiple measurements to reduce error
 - Trade-off with cost

#1 NATIONAL BESTSELLER

IN

"Unsparing...a clear, concise and extremely interesting look at a crucial period of U.S. decision making. It deserves to be widely read."

—Wall Street Journal

RETROSPECT



THE TRAGEDY AND LESSONS OF VIETNAM

ROBERT S.
MCNAMARA

WITH BRIAN VANDEMARK



ENRICO
NEW YORK
INTERNATIONAL FILM FESTIVAL
AWARD WINNER
CANNES
INTERNATIONAL
FILM FESTIVAL
AWARD WINNER
BERLIN
INTERNATIONAL
FILM FESTIVAL
AWARD WINNER

FROM THE DIRECTOR OF "THE THIN BLUE LINE" AND "GHOSTS OF HAVANA"

"IF THERE IS ONE MOVIE THAT OUGHT TO BE STUDIED BY MILITARY AND CIVILIAN LEADERS AROUND THE WORLD AT THIS TREACHEROUS HISTORICAL MOMENT, IT IS 'THE FOG OF WAR', ERROL MORRIS'S PORTRAIT OF FORMER UNITED STATES DEFENSE SECRETARY, ROBERT S. MCNAMARA." Stephen Holden, The New York Times

"SMART AND COMPLETELY FASCINATING!" Richard Pfeiffer, The Daily Beast

"STUNNING! SUPERBLY MADE!" David Denby, The New Yorker

"SPLENDID! A MASTERPIECE!" Angie Shultz, The Daily Beast

THE FOG OF WAR
Directed by Errol Morris

BEST DOCUMENTARY FEATURE
BEST ORIGINAL SCORE
PHOTOGRAPHY
BEST EDITING
Karen Schmeer
Doug Aibel
David King

2003 Official Selection AFM Film Festival - Fri, Nov 14, 2003 3:00pm



S3D Software and Societal
Systems Department

Carnegie
Mellon
University

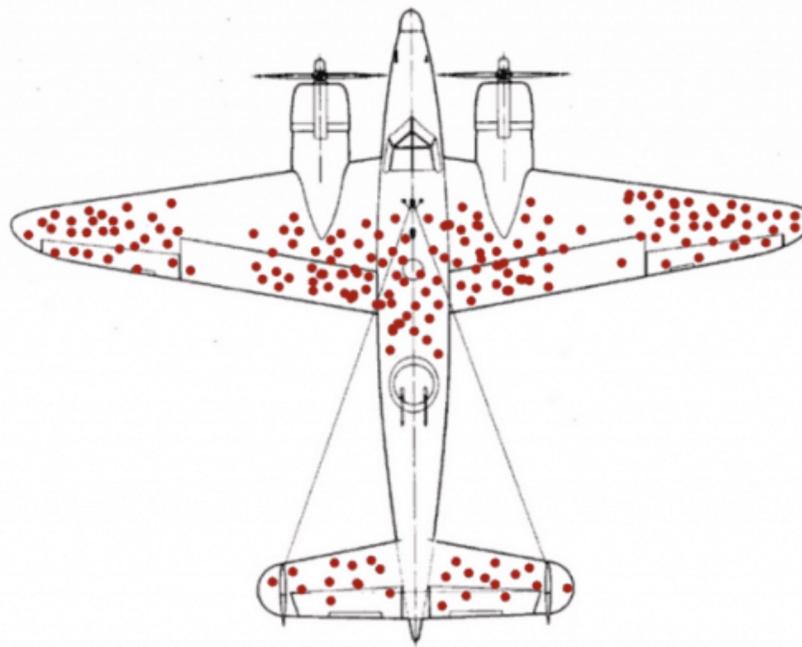
McNamara Fallacy

- Measure whatever can be easily measured
- Disregard that which cannot be measured easily
- Presume that which cannot be measured easily is not important
- Presume that which cannot be measured easily does not exist



<https://chronotopeblog.com/2015/04/04/the-mcnamara-fallacy-and-the-problem-with-numbers-in-education>

Survivorship Bias



Outline



Measurements and Metrics



How to use Measurements and Metrics



Case Study: Autonomous Vehicle Software



Risks and Challenges



Metrics and Incentives

Goodhart's law: "When a measure becomes a target, it ceases to be a good measure."



<http://dilbert.com/strips/comic/1995-11-13>

The Price of Wells Fargo's Fake Account Scandal Grows by \$3 Billion

The bank reached a settlement with federal prosecutors and the Securities and Exchange Commission after abusing customers.

 Share full article



 199



Wells Fargo used fraud to open up fake accounts and force customers into services

Incentivizing Productivity

- What happens when developer bonuses are based on:
 - Lines of code per day?
 - Amount of documentation written?
 - Low number of reported bugs in their code?
 - Low number of open bugs in their code?
 - High number of fixed bugs?
 - Accuracy of time estimates?

What You Need to Know



Metrics are important in Software Engineering



Apply goal-oriented approaches to software metrics



Provide examples of metrics for software qualities and process



Understand limitations and dangers of decisions and incentives based on measurements

Questions to Consider (Projects)

- What properties do we care about and how do we measure them?
- What is being measured? Does it (to what degree) capture the thing you care about? What are its limitations?
- How should it be incorporated into process?
- What are potentially negative side effects or incentives?