Symbiosis of Mobile Analytics & Software Testing

A practical guide to better testing of Mobile Apps

Julian Harty

11 Feb 2016

"Our business branch in 2014 is the 7:01 from Reading to Paddington. Over 167,000 of our customers use our Mobile Banking app between 7am and 8am on their commute to work every day."

RBS

Ross McEwan, CEO, Royal Bank of Scotland

http://www.bbc.co.uk/news/business-your-money-26365616

SOUTH WEST TRAINS

The art of building great user experiences

Maturing to a broader perspective

How to improve your mobile testing beyond the "shift left" recommendation?

How to use mobile analytics to increase user experience and the quality of our work?

How to keep testing laser-focused on user experience and on business success?

How to reduce or remove irrelevant practices?

How to move from traditional "hands-on" testing to the design, analysis, and application of data generated by mobile apps, the users, etc?

How to better prioritize our development and testing efforts?

Mobile Testing

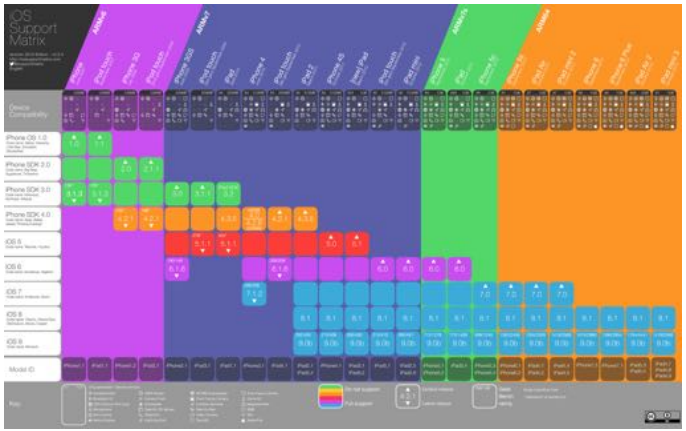
Mobile Analytics

The Confluence

FRAGMENTATION, OR THE GODZILLA IN THE ROOM

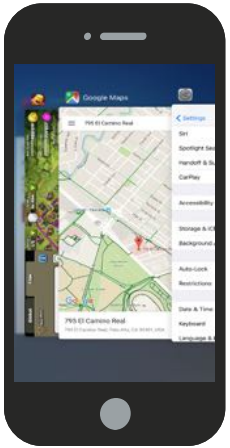
Device Characteristics

Screenshot from Lirum Info



Platform diversities

iOS Support Matrix from OpenSignal



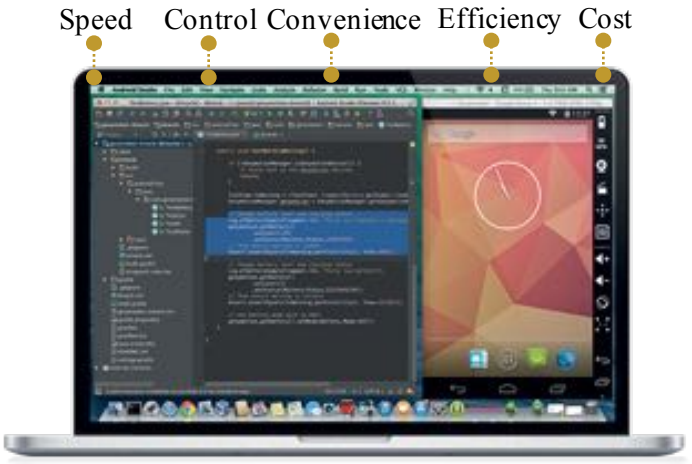
Usage Patterns

Settings, changing contexts

MEASURING QUALITY SO IT CAN BE IMPROVED

Defining metrics

ISO/IEC 25010:2011



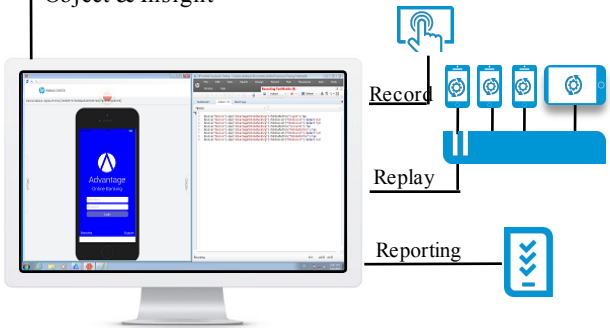
The virtues of emulators

Screenshot from GenyMobile

MOBILE TESTING DISCIPLINES

For Design, Usability, Accessibility, Localization testing

Multi-device, multi OS, multi apps
Object & Insight



https://www.hp.com/go/UFT

Manual to Interactive

The mandate for Automation

1 second page load delay results in 7% drop in conversion rate, 11% drop in page views, 16% drop in customer satisfaction

Performance

Quote from Getelastic.com

Mobile Fragmentation Measure Quality **Disciplines** Boost Testing **Analytics** **Confluence**

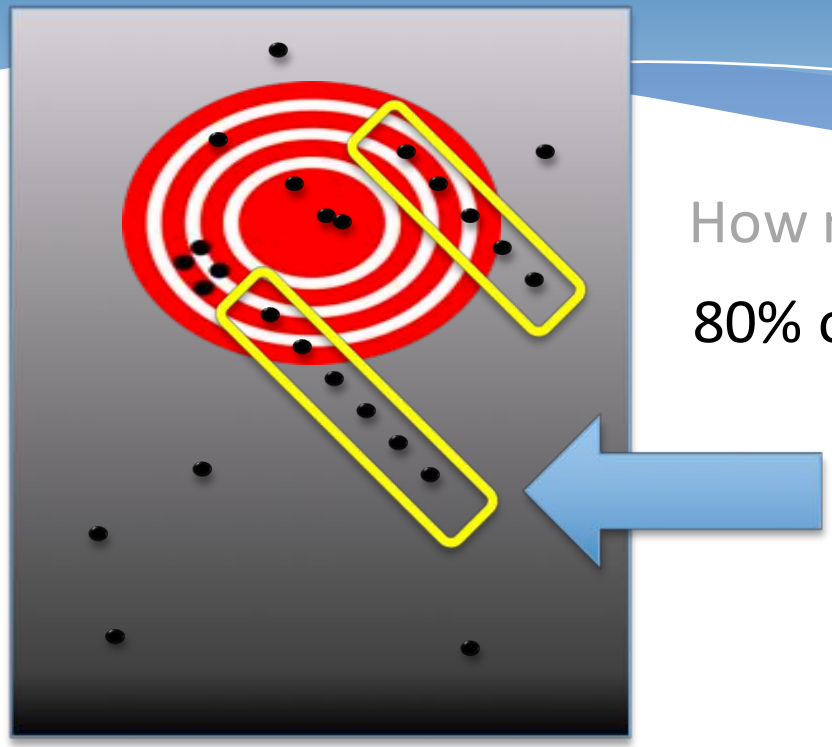
Are you working too hard, on the wrong things?

We struggle to decide what to test, how much testing is enough, etc.

Most of our work is wasted effort

How much of our testing is on target?

Most of our work is wasted effort



How much of our testing is on target?

80% of reported bugs not addressed

Automated Tests ☺

Mobile Testing

Mobile Analytics

The Confluence

Know your users

Custom drink feature removed[1]

=> **1 star feedback** ratings

Parallel Kingdom[2]

Regular users generate **2.5x daily revenues**

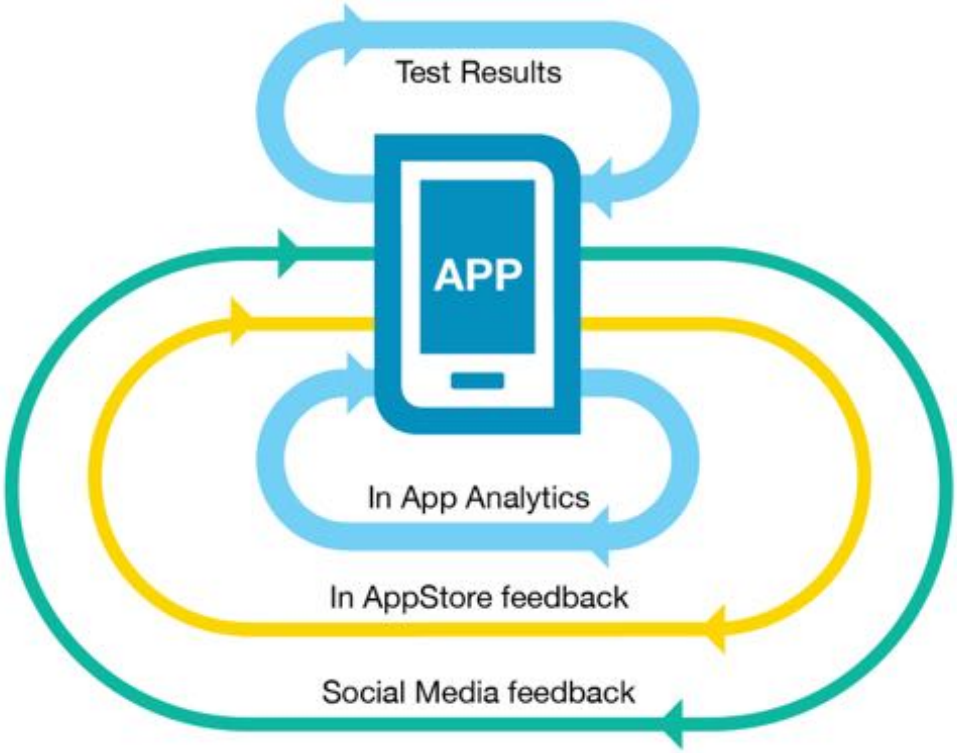
Logos © respective owners

[1] example from App Quality book
[2] example from Tale of Two Apps

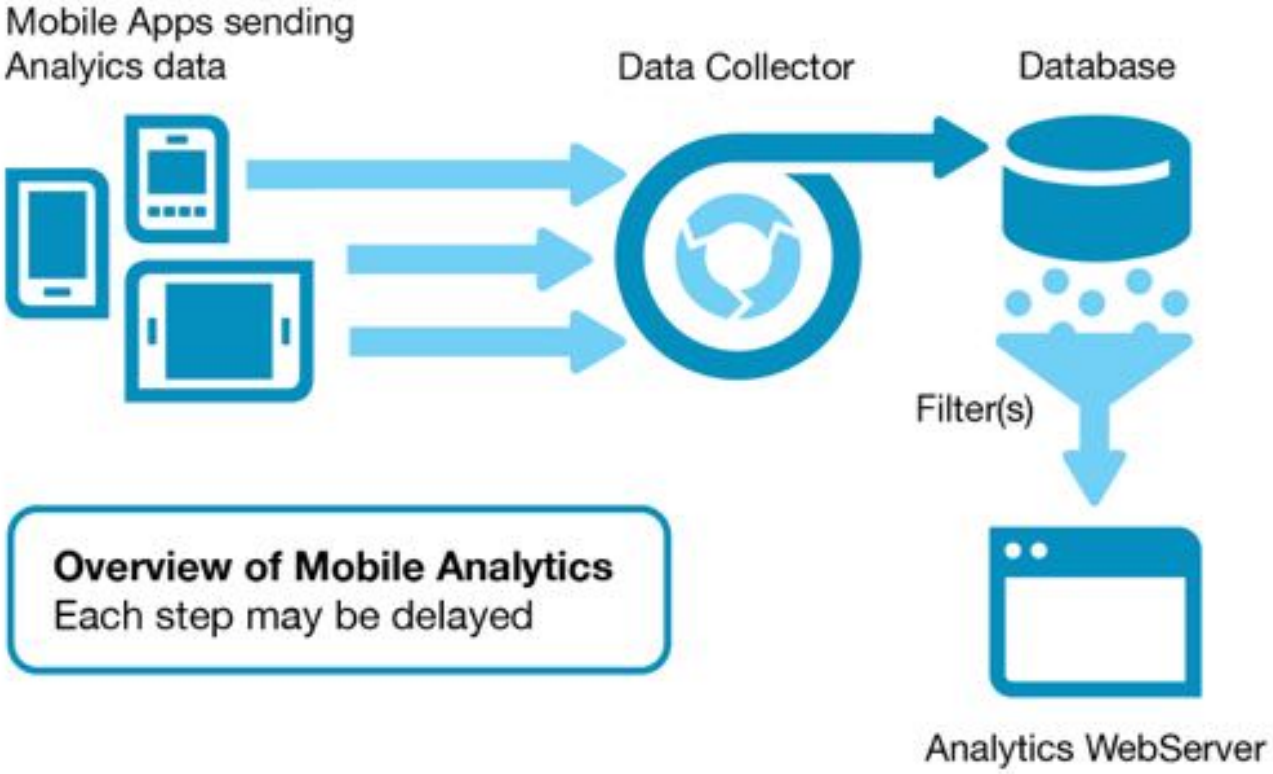
Understand the effects

* **Battery drain varied by 3x** for similar hardware specifications
* Custom code added for **Kindle Fire** to reduce brightness
  * **40% less battery drain**
* Higher network latencies reduced interactivity by 40%
* Users preferred Wi-Fi
  * 69% for Parallel Kingdom, 58% for StudyBlue
* Tablets 2x usage
* Pull-out keyboard also increased usage

Feedback Cycles



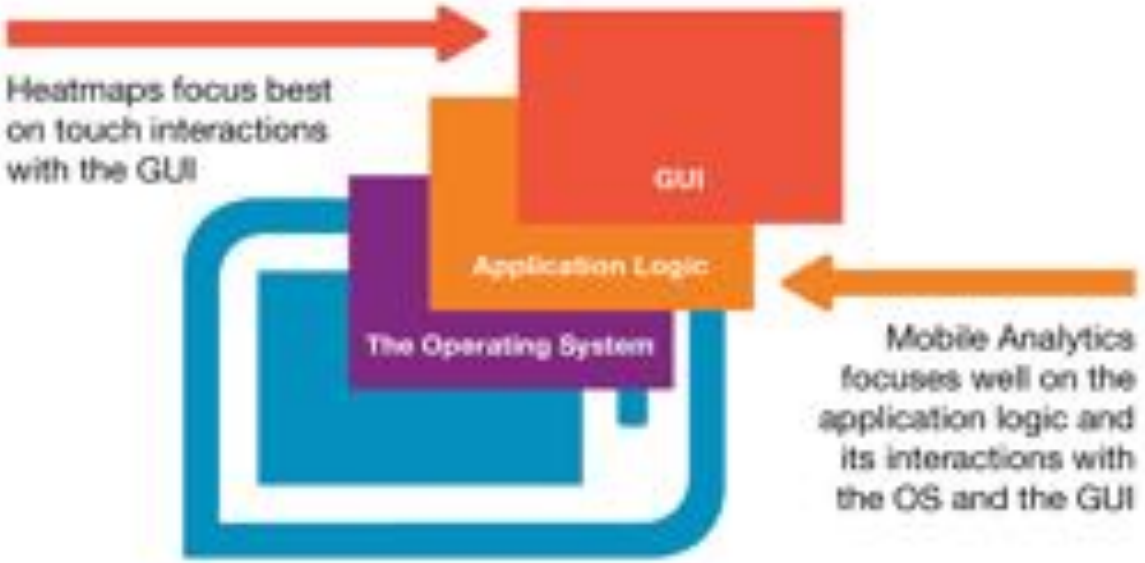
TOPOLOGY


Mobile Apps sending
Analyics data
Data Collector
Database
Filter(s)
Overview of Mobile Analytics
Each step may be delayed
Analytics WebServer

Mobile Analytics

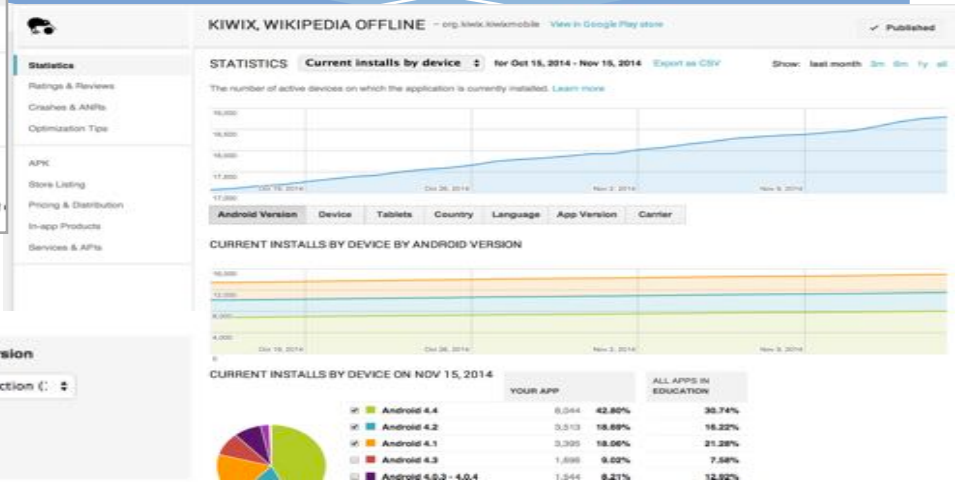
Some examples

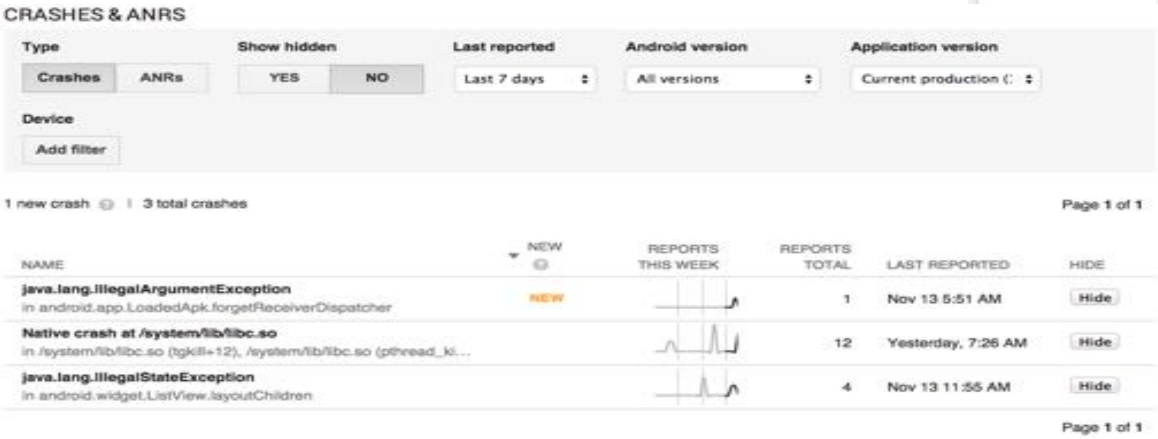
Layers of an App



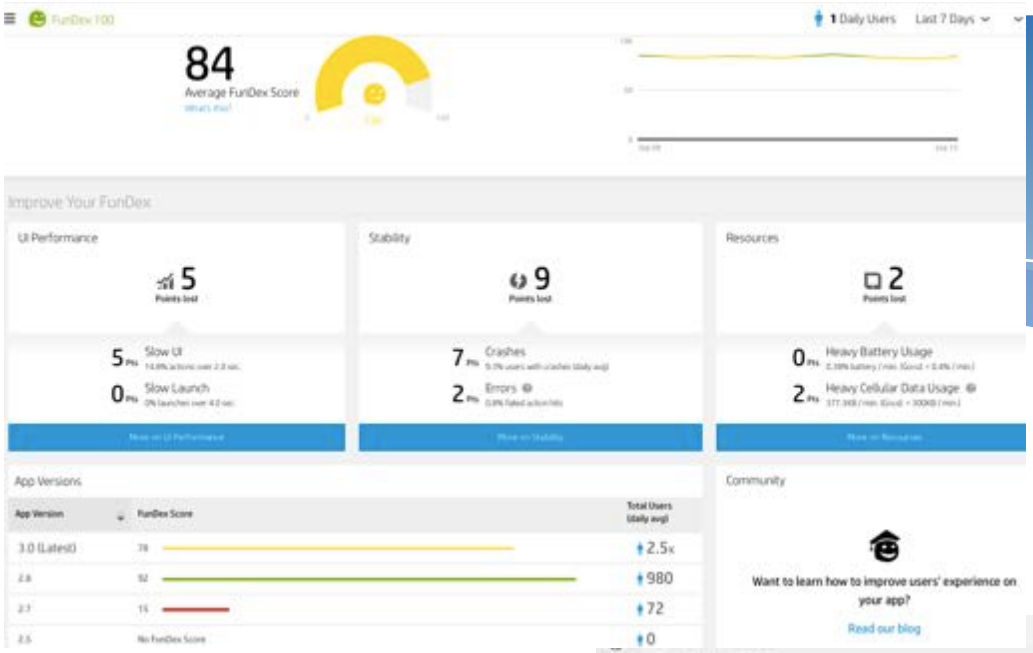
Exploded view of the app running on the platform

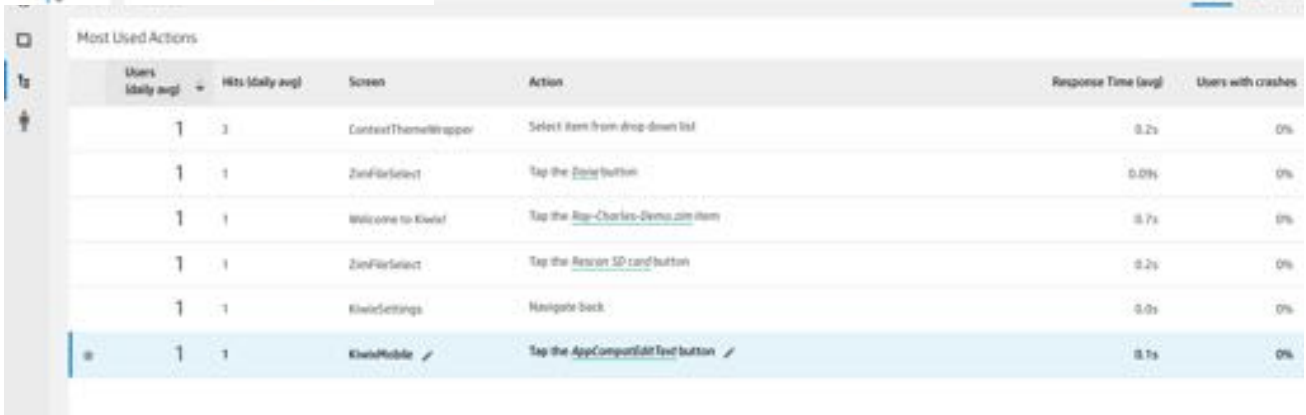
Developer Console (Google Play)





HP AppPulse Mobile





Heatmaps

Slide

Zoom/Dezoom

"8 Heatmap Tests that reveal visitor behavior."

1. The Link Test – Are visitors clicking on dead ends?
2. The Distraction Test – Are irrelevant elements distracting visitors?
3. The Information Test – Is there missing information from your page?
4. The Device Test – Does your page work on different screen sizes?
5. The Depth Test – Is all your content easily reachable?
6. The Engagement Test – What are your site visitors really looking for?
7. The Fold Test – Does your page have what it takes to keep your visitors?
8. The Header Test – Is your page header helping or hurting your site?

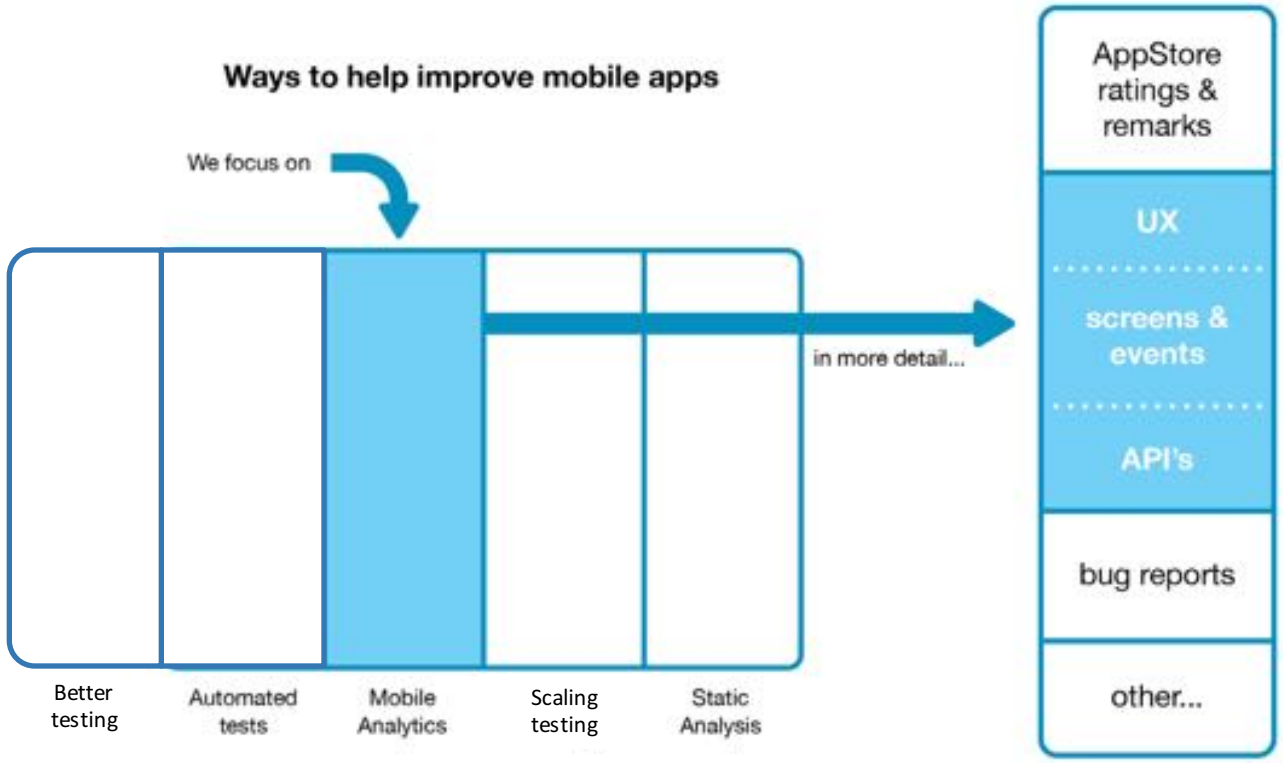
Source: https://www.hotjar.com/heatmaps


Mobile
Testing


Mobile
Analytics


The
Confluence

One Way of Many



Confluence

Tests based on usage

Some benefits of analytics data

* No longer limited to what “we” think “they” need/do/etc. Discover how the app behaves across virtually all the population
* Lower cost of operations
* Real usage can help drive our testing and analysis; brings realism to our testing

Testing based on analytics data

Testing based on information

* Popularity & volumes
* Locales
* User-flows, activities, etc.
* Crashes

Aims include

* Fast reproduction
* 1st-hand learning in controlled env.

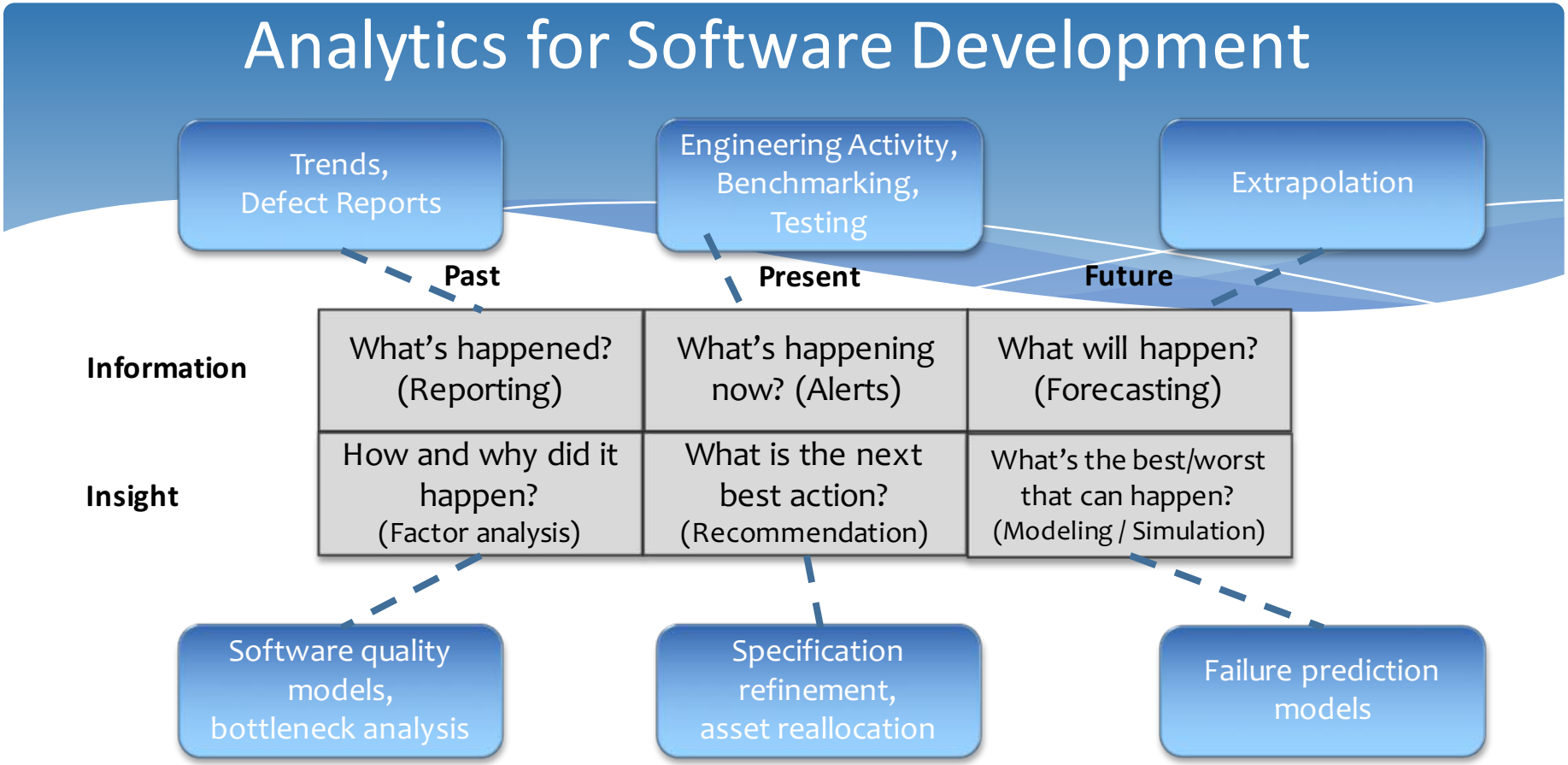
Testing based on insights

* Delta's (rate of change)
* Crossing thresholds
* Anomalies

Aims include:

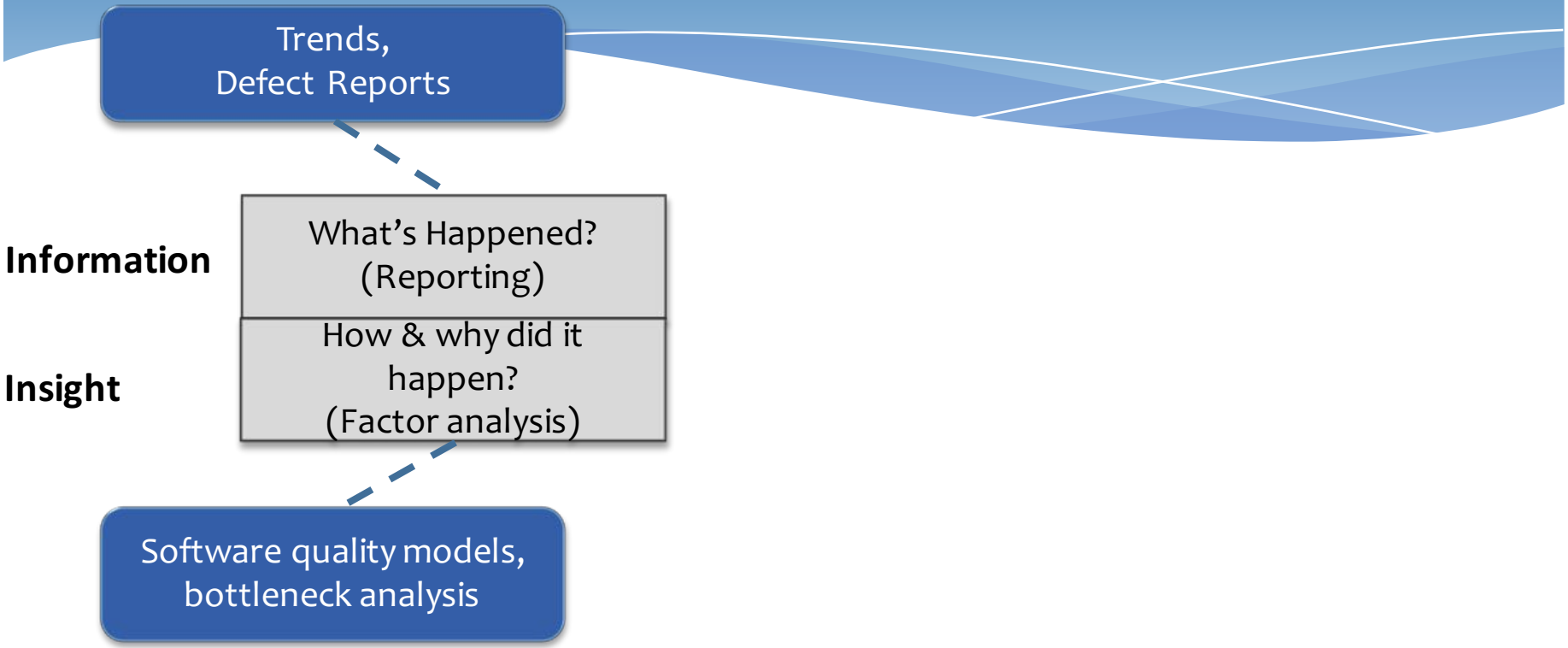
* Maximising insights, agility, and ability to adapt & respond

Analytics for Software Development

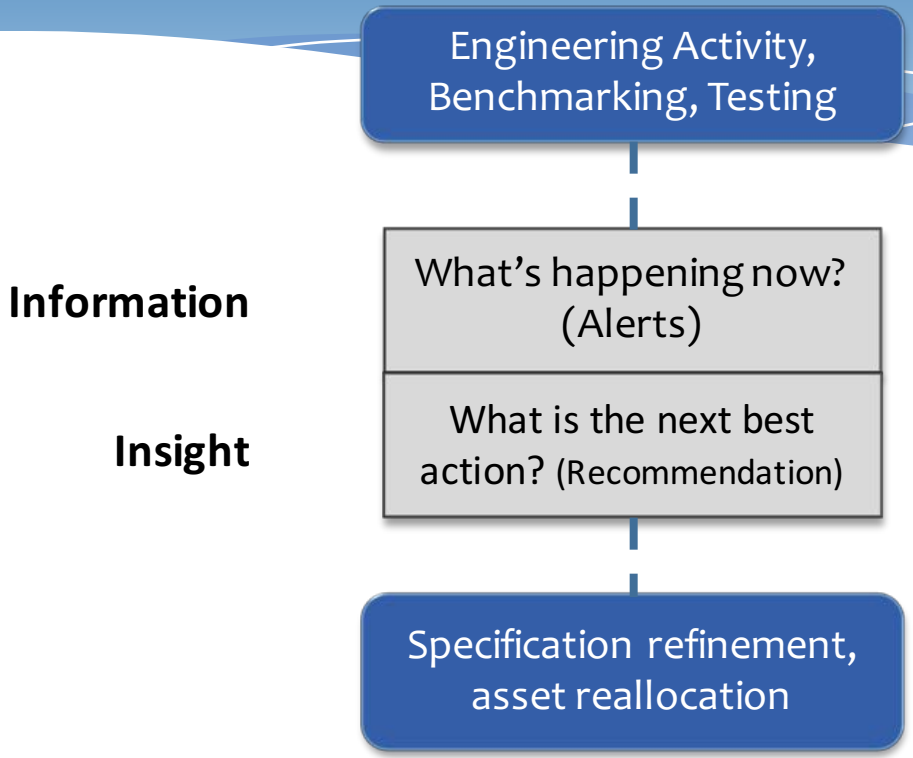


http://research.microsoft.com/pubs/136974/foser-2010-buse.pdf

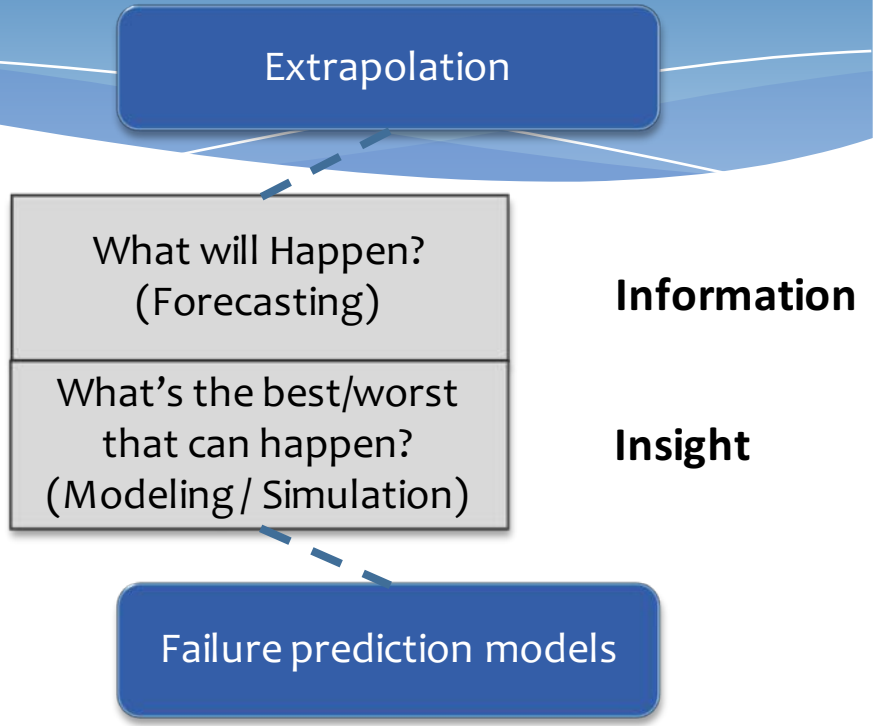
Analytical Questions: Past



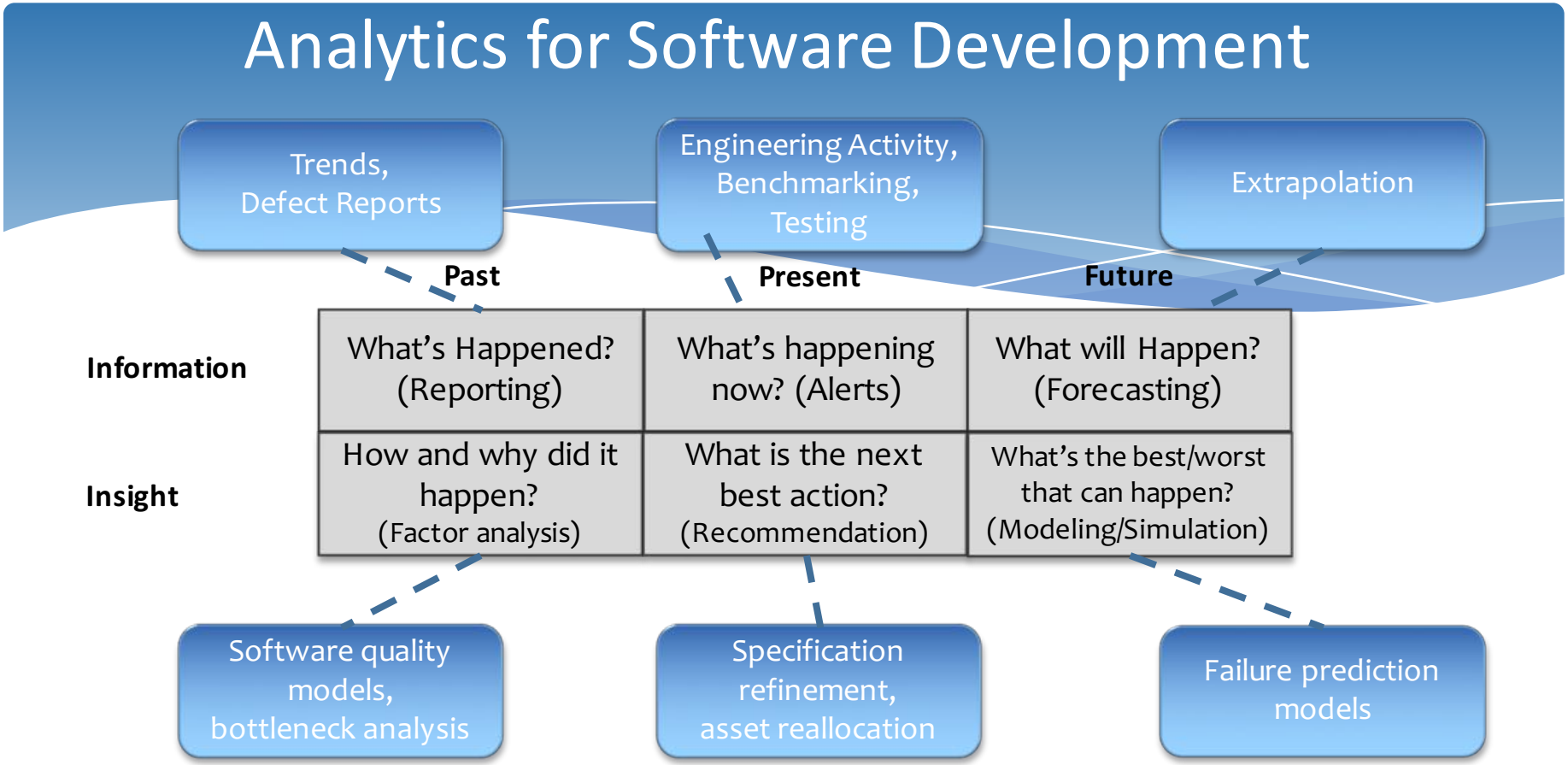
Analytical Questions: Present



Analytical Questions: Future



Analytics for Software Development

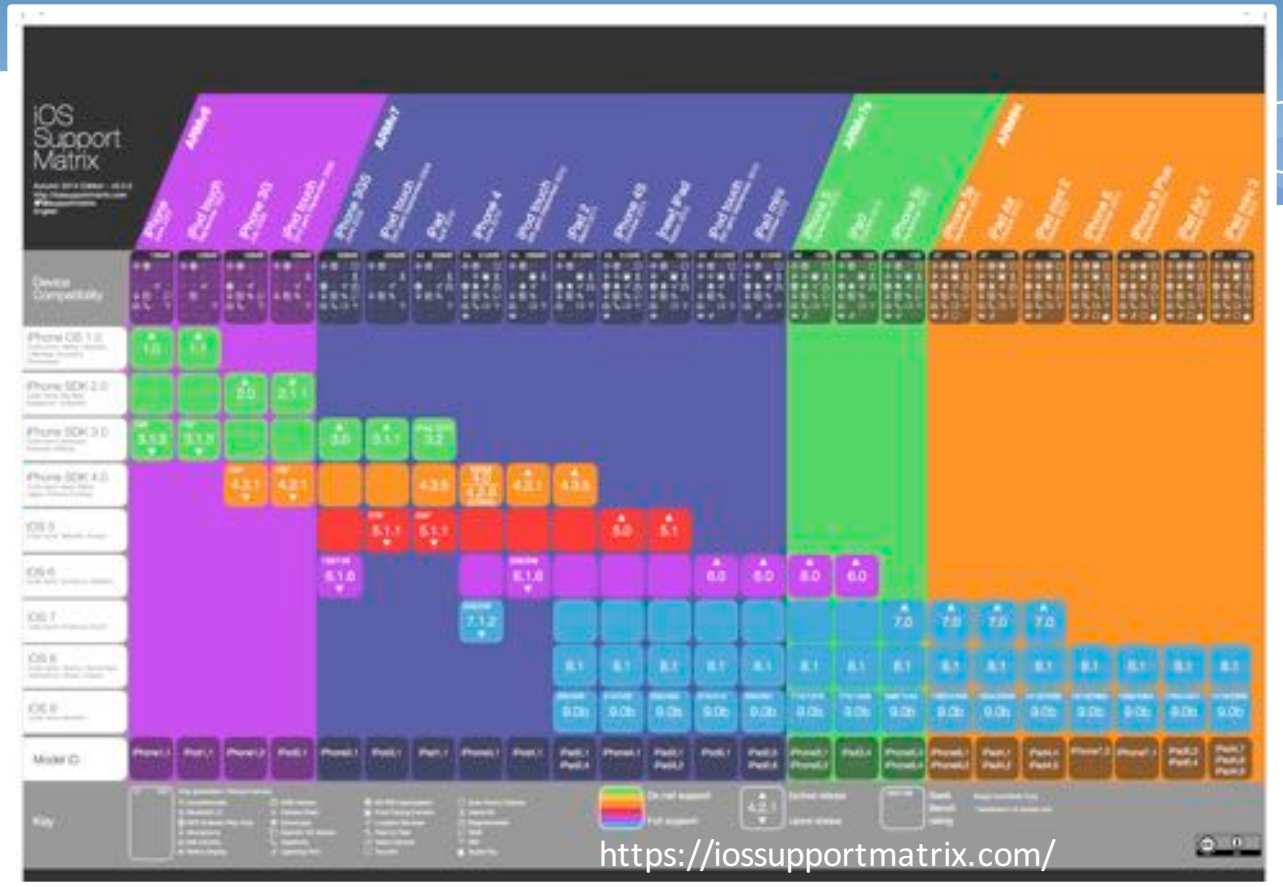


http://research.microsoft.com/pubs/136974/foser-2010-buse.pdf

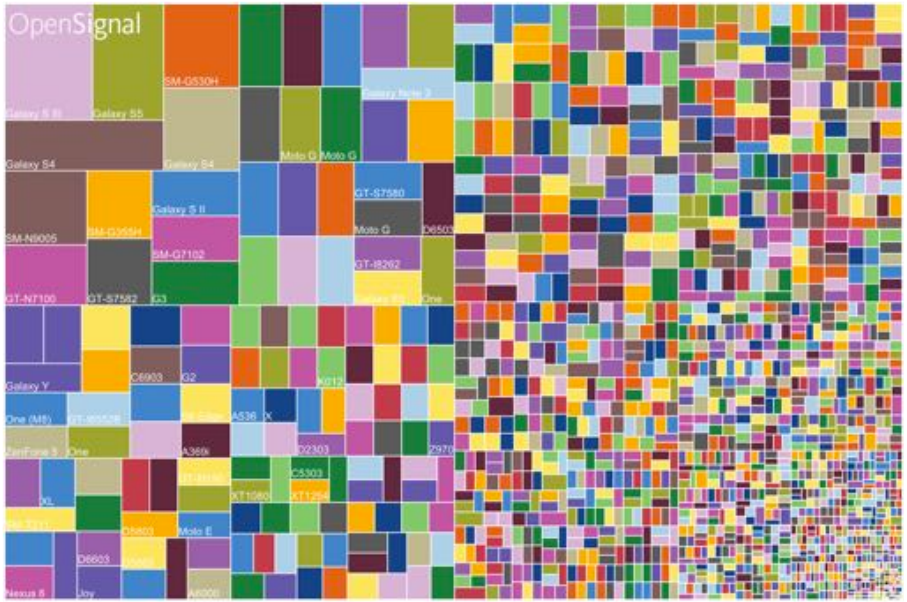
Perennial Question

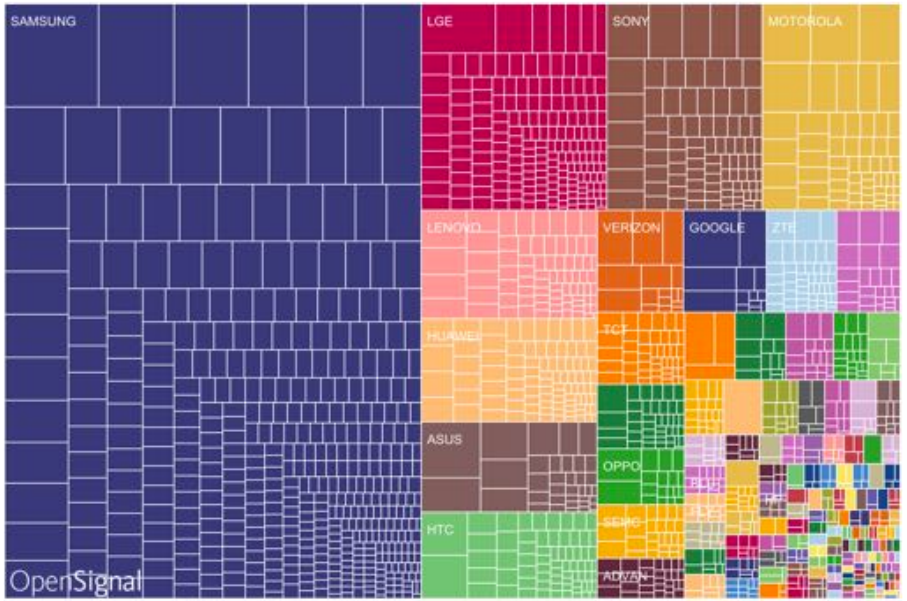
How many devices are enough?

iOS Matrix



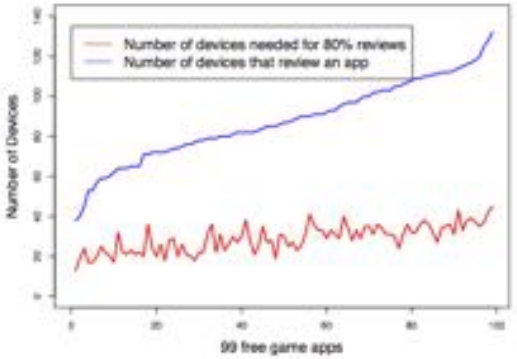
Android: Devices Matrix

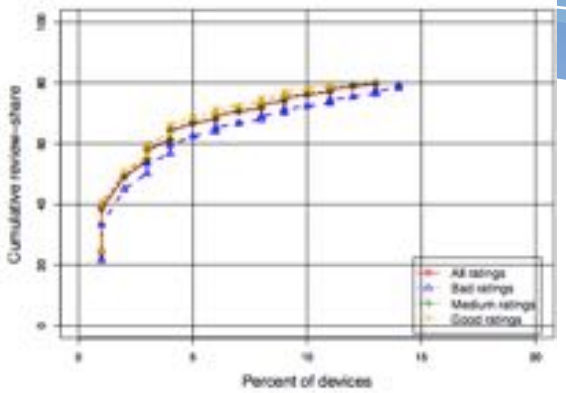




Used with permission from OpenSignal.com

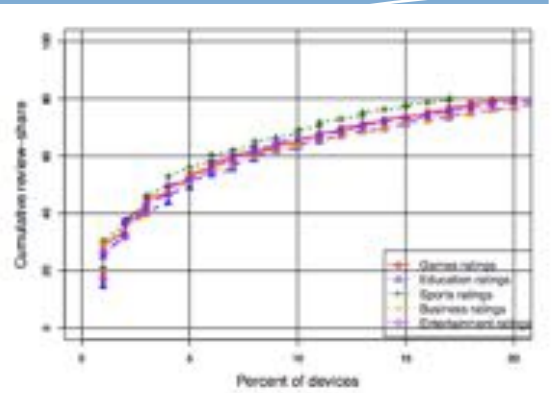
How many devices are enough?





<= Free

Paid=>



Number of devices to cover 80% of reviews

“Prioritizing the devices to test your app on : A case study of android game apps”

Prioritizing The Devices To Test Your App On: A Case Study Of Android Game Apps

Hammad Khalid[1], Meiyappan Nagappan[2], Emad Shihab[1], Ahmed E. Hassan[1]
[1]Software Analysis and Intelligence Lab (SAIL), Queen's University, Kingston, Canada
[2]Department of Software Engineering, Rochester Institute of Technology, Rochester, USA
[3]Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada
[1]hammad@cs.queensu.ca, [2]mei@se.rit.edu, [3]eshihab@cse.concordia.ca,
[1]ahmed@cs.queensu.ca

ABSTRACT

Star ratings that are given by the users of mobile apps directly impact the revenue of its developers. At the same time, for popular platforms like Android, these apps must run on hundreds of devices increasing the chance for device-specific problems. Device-specific problems could impact the rating assigned to an app, given the varying capabilities of devices (e.g., hardware and software). To fix device-specific problems developers must test their apps on a large number of Android devices, which is costly and inefficient.

Therefore, to help developers pick which devices to test their apps on, we propose using the devices that are mentioned in user reviews. We mine the user reviews of 99 free game apps and find that, apps receive user reviews from a large number of devices: between 38 to 132 unique devices. However, most of the reviews (80%) originate from a small subset of devices (on average, 33%). Furthermore, we find that developers of new game apps with no reviews can use the review data of similar game apps to select the devices that they should focus on first. Finally, among the set of devices that generate the most reviews for an app, we find that some

and large companies are developing an enormous amount of applications (called mobile apps), designed to run on Android devices. However, the top-rated or the featured apps in the app markets, are the apps with the most downloads, and hence the most revenue [2, 3]. Also the app market is very competitive, especially for game app developers who have to compete with almost 120,000 game apps already in the Google Play store – more than any other category of apps. To compete in this environment, developers need to get (and maintain) good ratings for their apps [2]. This can be difficult since users are easily annoyed by buggy apps, and that annoyance could lead to bad ratings [4, 5]. Hence, app developers need to test their apps thoroughly on different devices to avoid a poor rating.

To make matters worse, there exists a large number of Android devices, each with its own nuances. In fact, dealing with device specific issues of (the many) Android devices is considered one of the biggest challenges developers face when creating an Android app [6]. A 2013 survey from Appcelerator, which has aggregated results from similar such surveys in the past three years, shows that

Gaining confidence

Fools rush in…

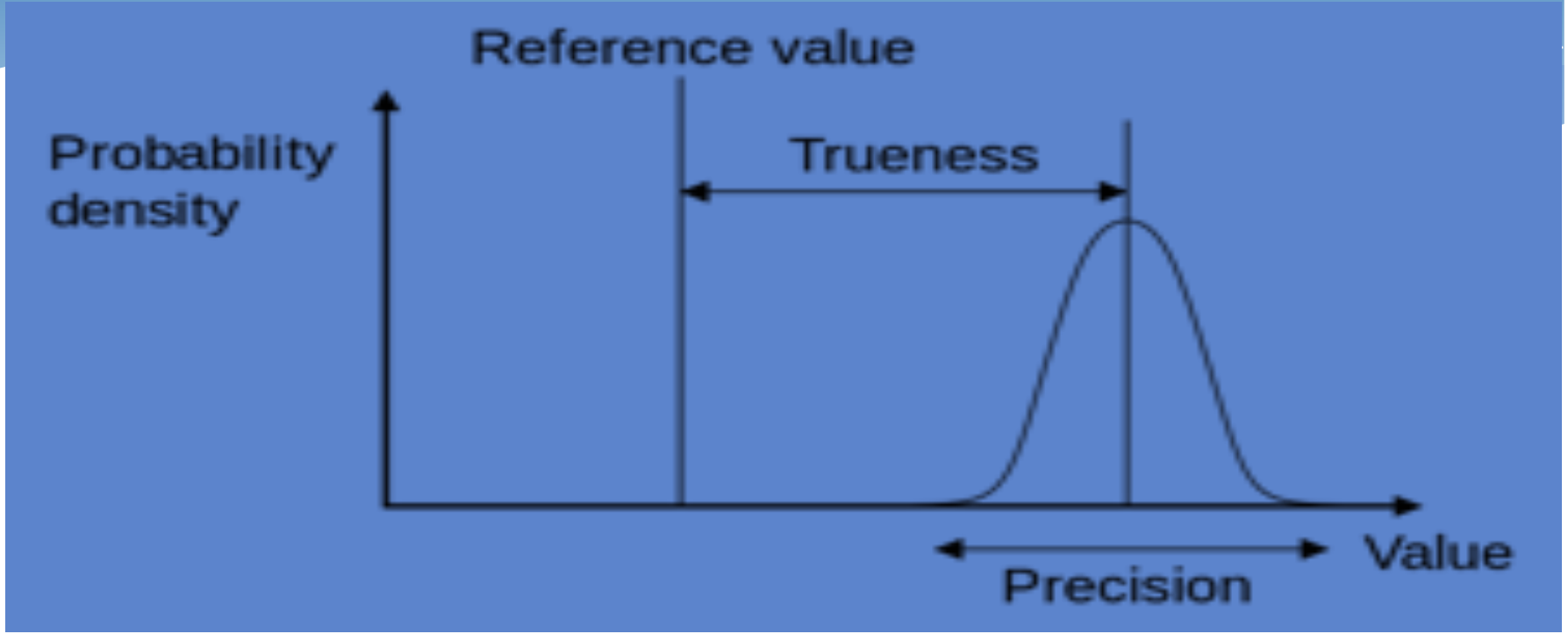
Precision & accuracy

Precision: repeatability

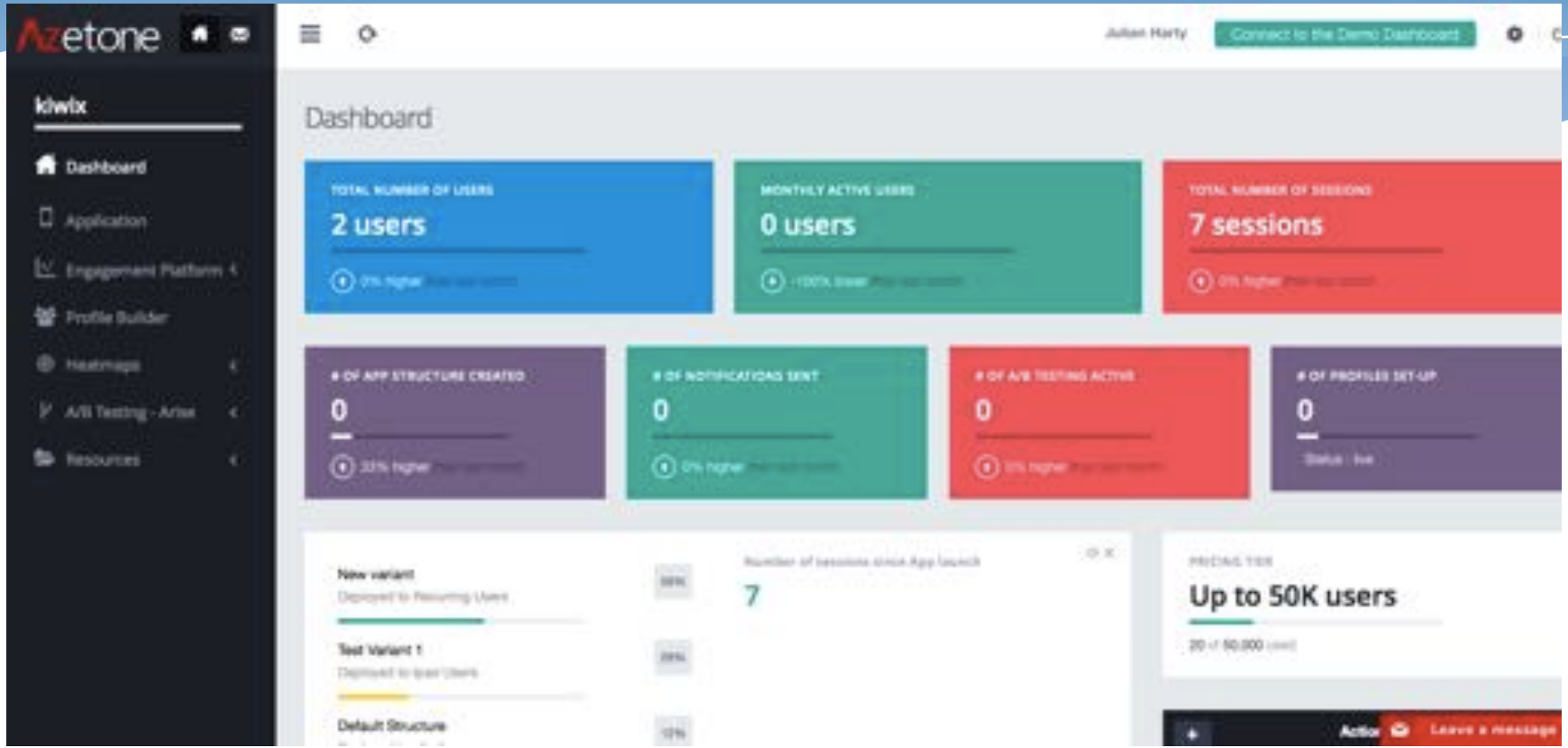
Accuracy: on target

https://en.wikipedia.org/wiki/Accuracy_and_precision

Precision & accuracy



A system at rest?



Divergent answers increase doubt

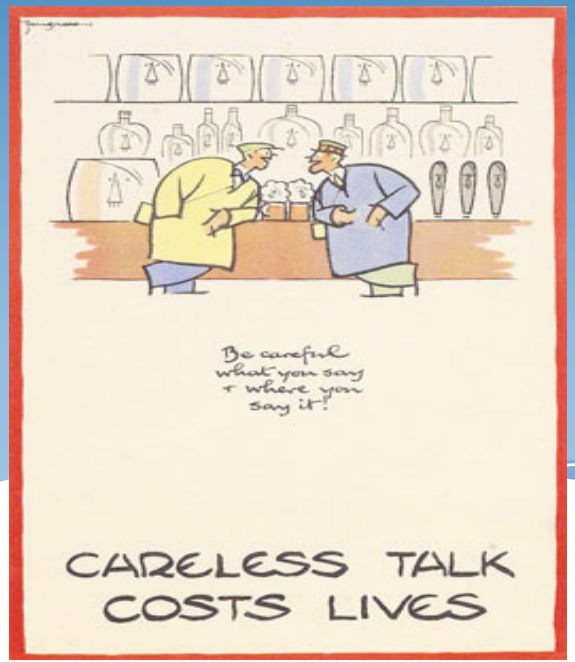
* A tale of ~~two~~ three mobile analytics libraries (and what happens when bonuses are on the line…)

* Where were the testers (part one)?

Of Mobile Analytics

The Dark Side

Do no harm to a neighbour



THIS POPULAR FLASHLIGHT APP HAS BEEN SECRETLY SHARING YOUR LOCATION AND DEVICE ID

MORE THAN 50 MILLION USERS WERE LEFT IN THE DARK ABOUT HOW BRIGHTEST FLASHLIGHT SHARED THEIR LOCATION AND DEVICE INFORMATION WITH THIRD PARTIES.

sourcedna

Blog

iOS Apps Caught Using Private APIs

October 18, 2015

CHINESE TAOMIKE MONETIZATION LIBRARY STEALS SMS MESSAGES

paloalto NETWORKS

Palo Alto Networks WildFire has captured over 18,000 Android apps that contain this library. (October 21, 2015)

Where were the testers (part two)?

SafeDK BETA Marketplace

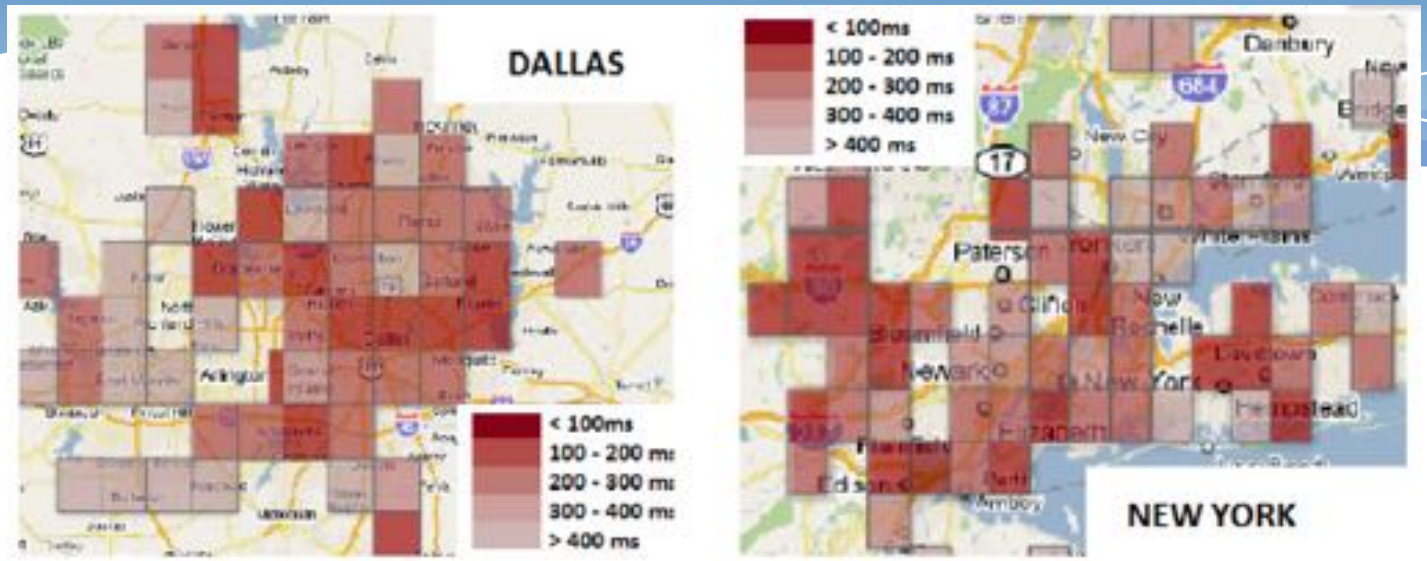
http://www.vam.ac.uk/users/node/1777

A Study of Third-Party Tracking by Mobile Apps in the Wild
ftp://ftp.cs.washington.edu/tr/2012/03/UW-CSE-12-03-01.PDF

https://sourcedna.com/blog/20151018/ios-apps-using-private-apis.html

http://researchcenter.paloaltonetworks.com/2015/10/chinese-taomike-monetization-library-steals-sms-messages/

Reducing precision to protect privacy



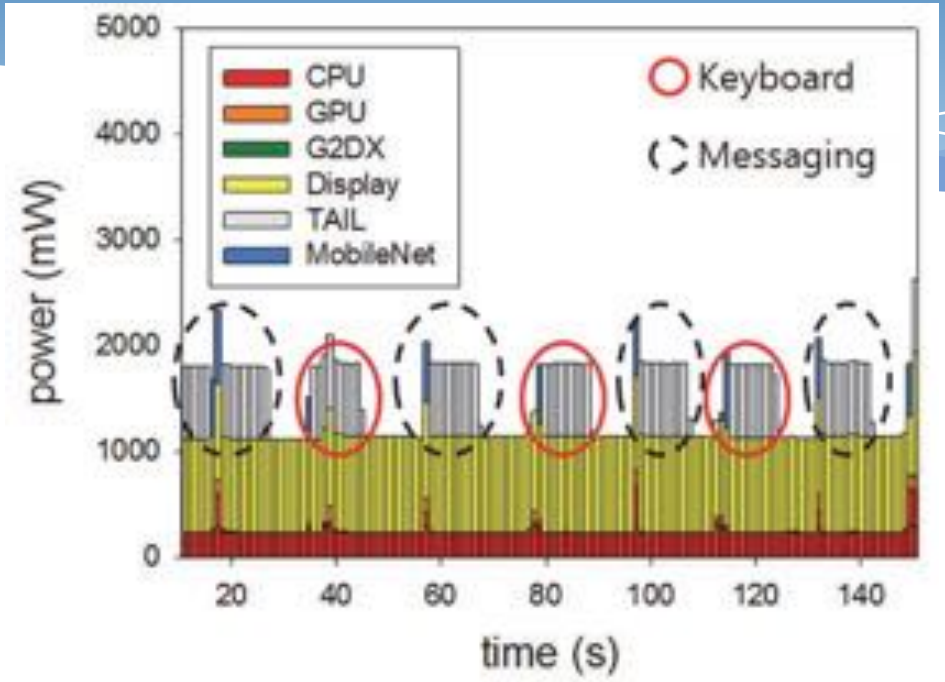
10km x 10km squares

From: Capturing Mobile Experience in the Wild: A Tale of Two Apps

Listening is a means to an end

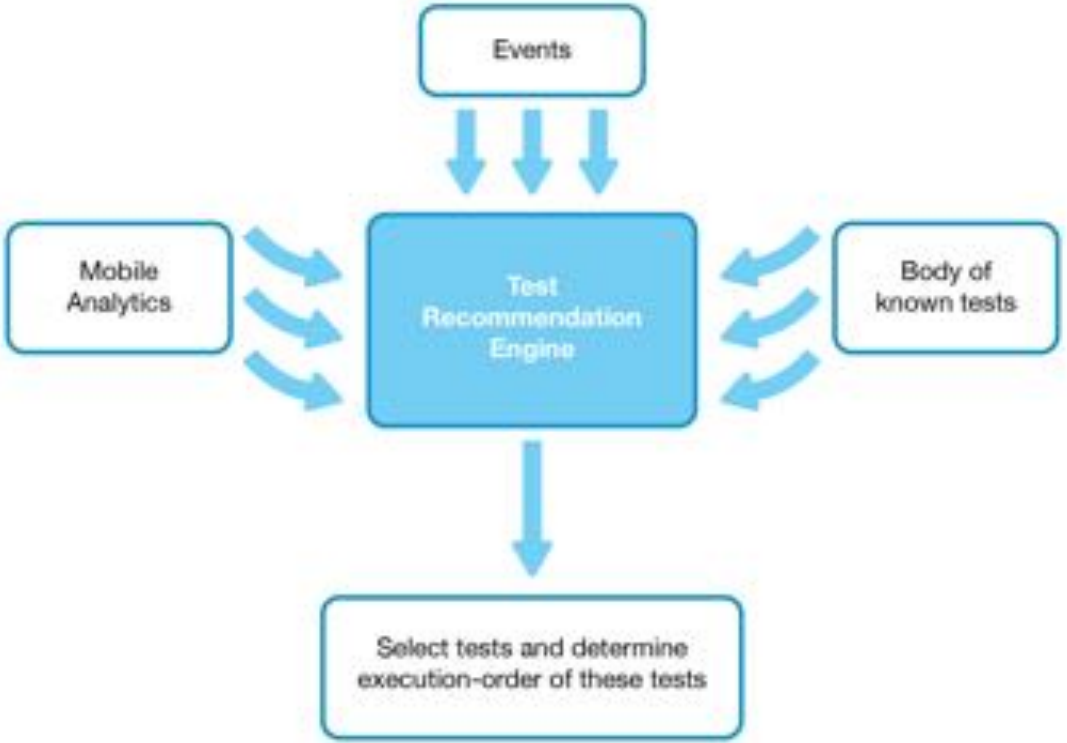
Necessary but not sufficient

Correlation? Causation? Appropriate?



From: User Interaction-based Profiling System for Android Application Tuning

Test Recommendation Engine

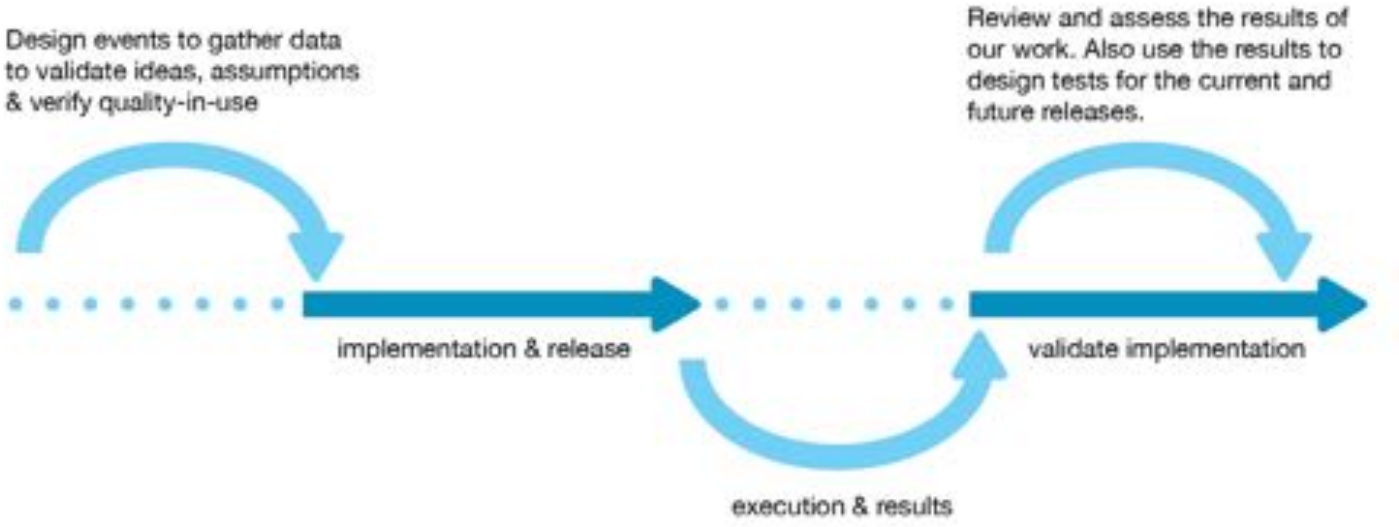


Beware the automation bias

"When presented with an automated solution 40% of pilots reasoned less or none at all"

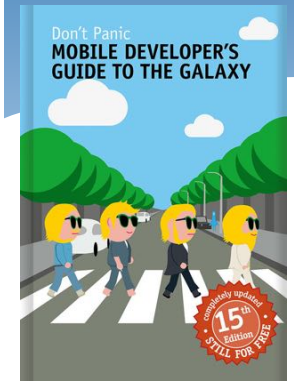
Automation Bias in Intelligent Time Critical Decision Support Systems
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.91.2634&rep=rep1&type=pdf

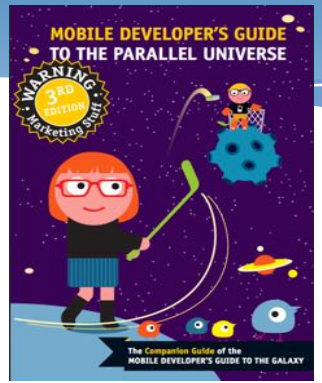
Getting Involved

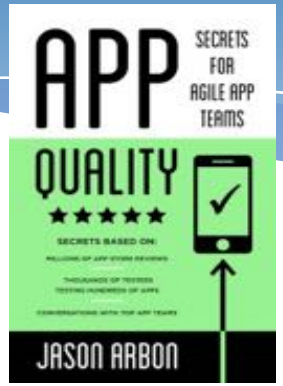


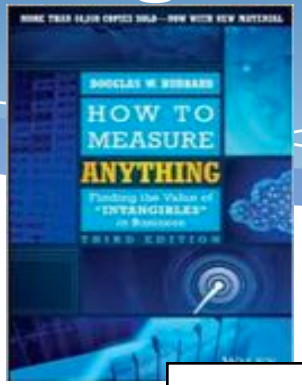
Feedback cycles between testing and mobile analytics for the app

Further reading



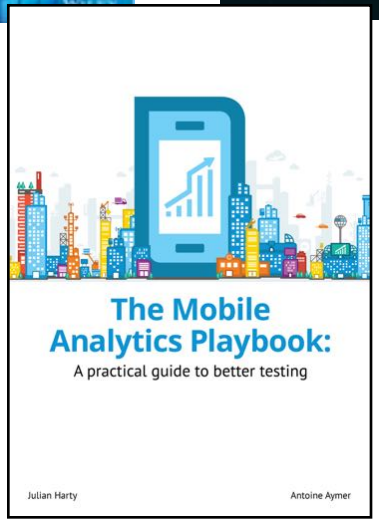






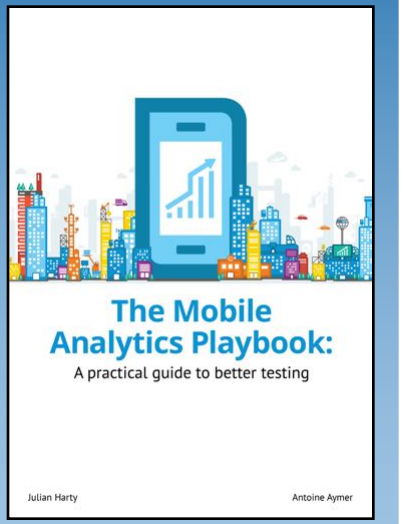
Email me for academic references
These books available at:

[1] http://enough.de
[2] http://wip.org/
[3] http://www.appqualitybook.com/
[4] http://www.howtomeasureanything.com/
[5] https://www.nostarch.com/silence.htm



Later: julianharty@gmail.com

Q&A Now?



https://www.surveymonkey.com/r/853LC3K

Sources of data

For Mobile Devices

Ways to collect data

What	Technique	Tool(s)
Device details	Research, App Store, Mobile Analytics	API calls, published
GUI elements and layout	Static Analysis, Dynamic Analysis	Code Quality utilities Test Automation Tools
Run-time environment	Queries, Mobile Analytics	Shell commands
Run-time behaviours	Instrumentation, Logging, Mobile Analytics	Development Tools, Test Automation Tools
Network traffic	Observation, Logging	Network Analyzer
Test results	Explicit collection	Data collection utilities, Test Automation Tools
App Store Data	Developer account access	Provided by App Store
Ratings & Reviews	Online access, data collection	App Store, other sources
In-app analytics	Integration of Mobile Analytics Library	Development tools

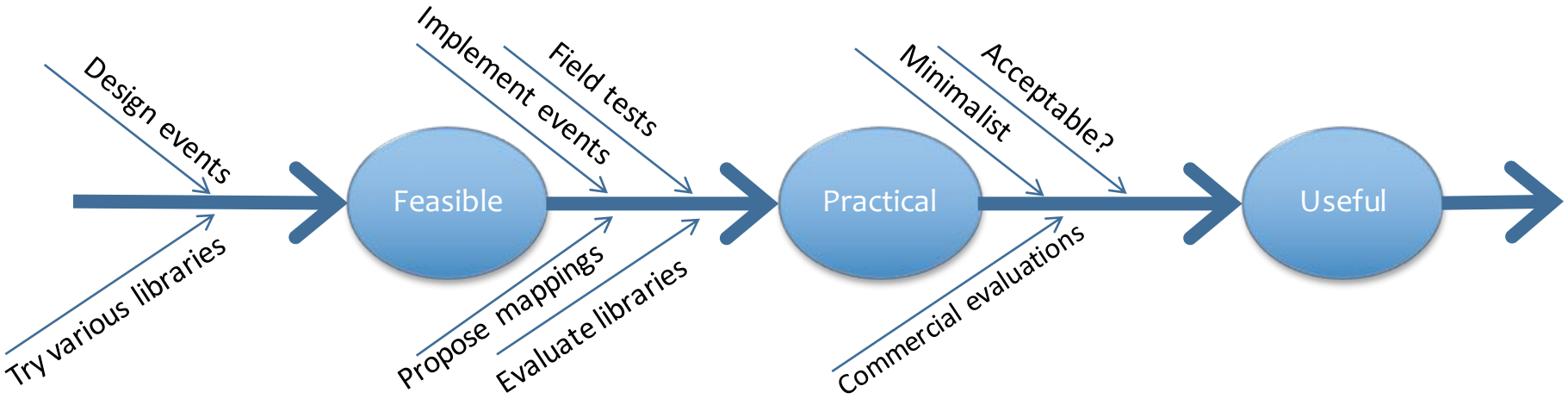
Data sources for mobile apps

What	On device	Off device
Device details	Y	
GUI elements and layout	Y	
Run-time environment	Y	
Run-time behaviours	Y	
Network traffic	Y	Y
Test results	?	Y
App Store Data		Y
Ratings & Reviews		Y
In-app analytics		Y

Sources of the data

What	Static	Runtime	Human
Device details	Y		
GUI elements and layout	Y	Y	
Run-time environment		Y	
Run-time behaviours		Y	
Network traffic		Y	
Test results	Y	Y	Y
App Store data		Y	
Ratings & Reviews			Y
In-app analytics		Y	

Implementation Flow



Questionnaire

Results

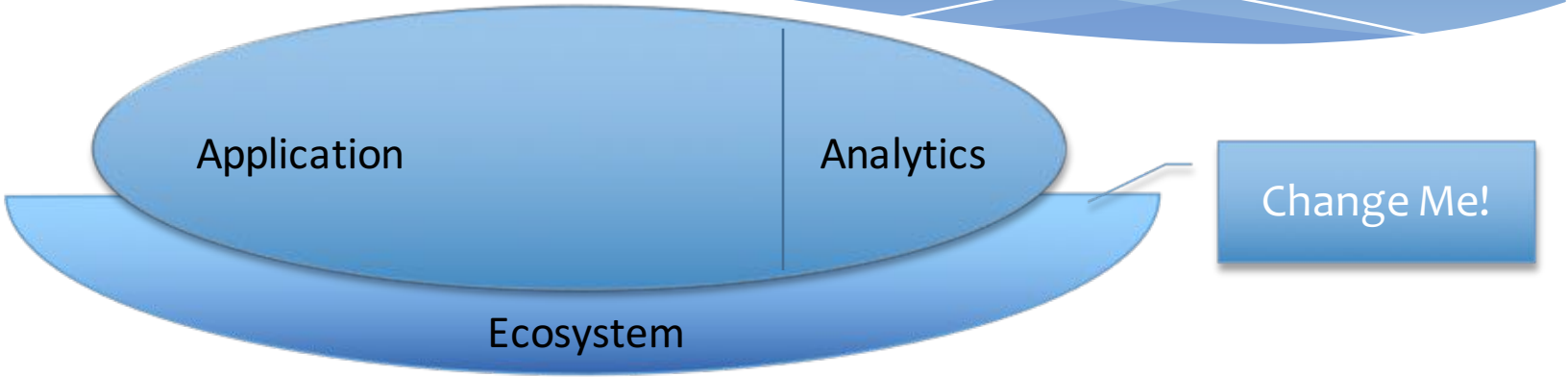
Aims: to glean 2 perspectives

1. As a potential beneficiary
2. As someone being observed

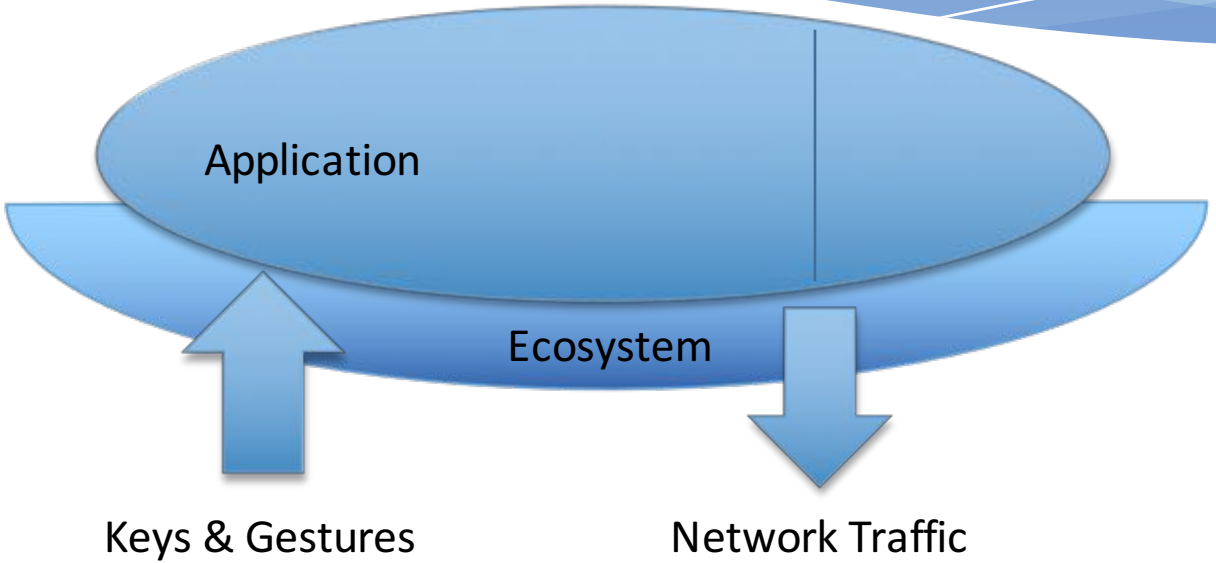
Emerging Patterns

1. Takers 3/10
2. Balanced 7/10
3. Dislike 2/10
4. Lagging what's happening

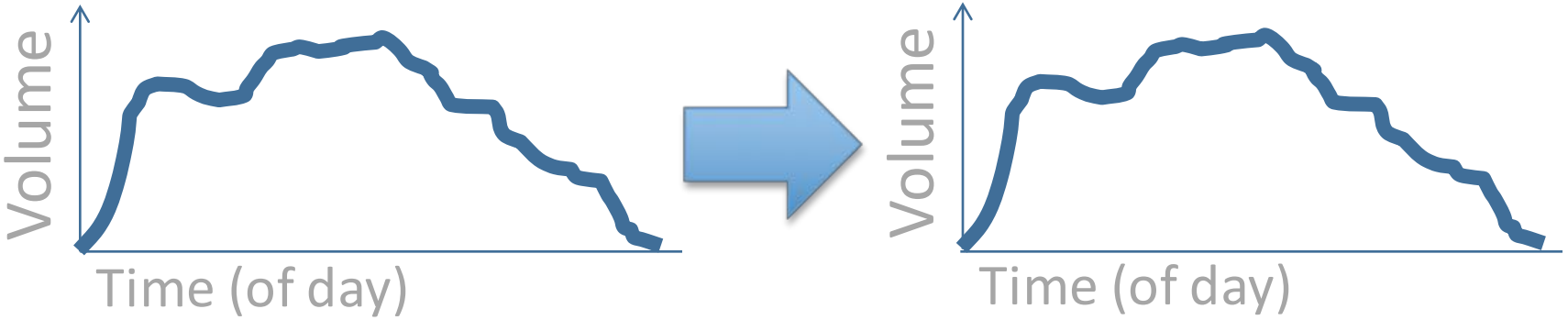
Instrument the Ecosystem



Observe the Behaviours



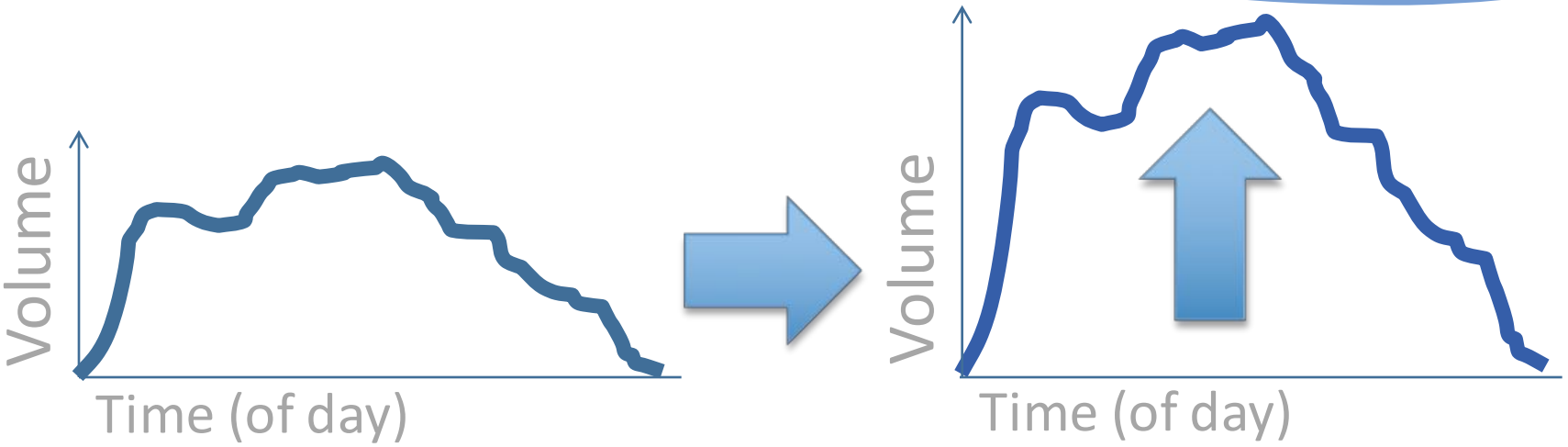
Network profiling



Discover

Use

Network profiling



Transform