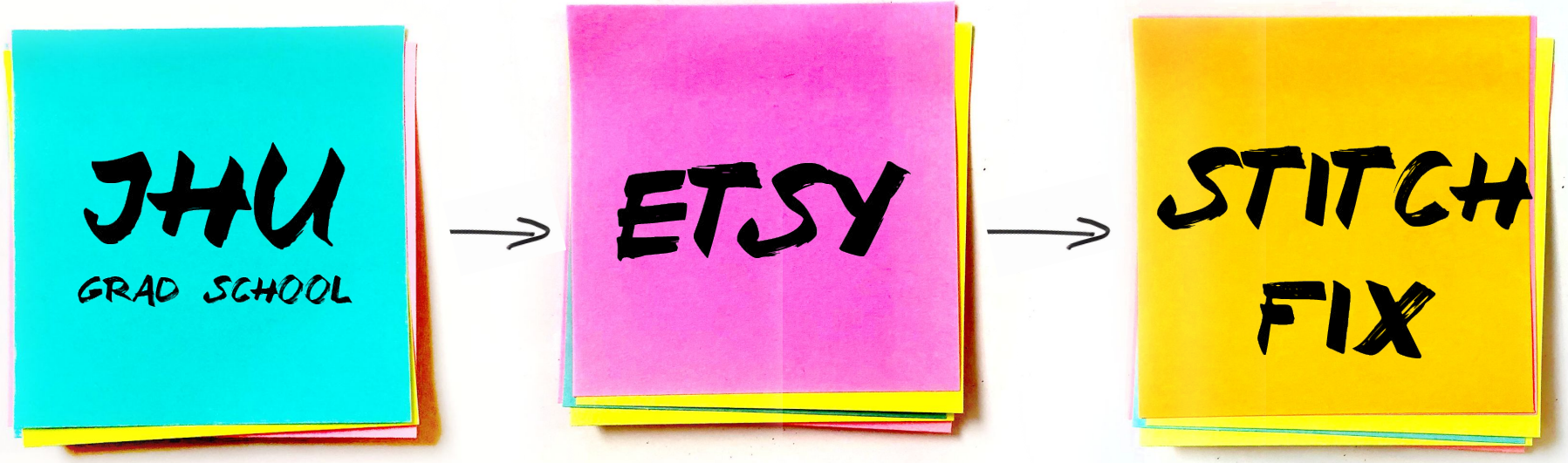
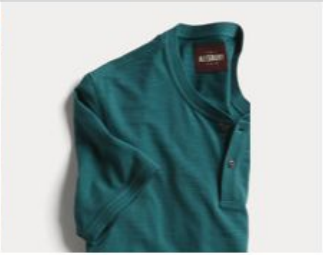
USING DATA EFFECTIVELY: BEYOND ART AND SCIENCE

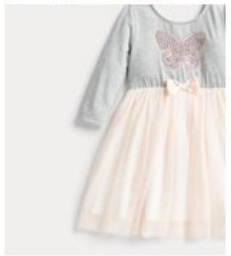
MY PATH












STITCH FIX

YOUR PARTNER IN PERSONAL STYLE

Try the personal styling service for everyone! No matter your age, size or budget we've got styles for you.

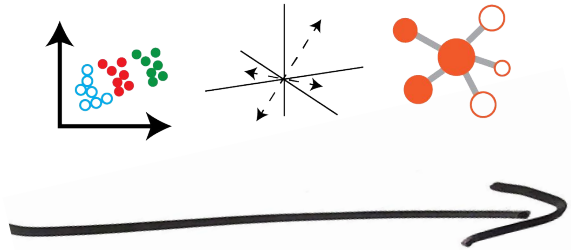
WOMEN → MEN → KIDS →





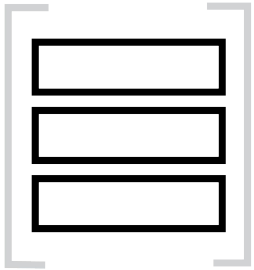
CREATE YOUR STYLE PROFILE
GET 5 HAND PICKED ITEMS
KEEP WHAT YOU LIKE SEND THE REST BACK

INVENTORY



MACHINE LEARNING

ALGORITHMIC RECOMMENDATIONS



“You’ve told me everything **NOT** to do, but how will I know what to do?”

-Anonymous Roger Peng student

DATA SCIENCE IS AN ART

“Data analysis is hard, and part of the problem is that few people can explain how to do it. It’s not that there aren’t any people doing data analysis on a regular basis. It’s that the people who are really good at it have yet to enlighten us about the thought process that goes on in their heads.”

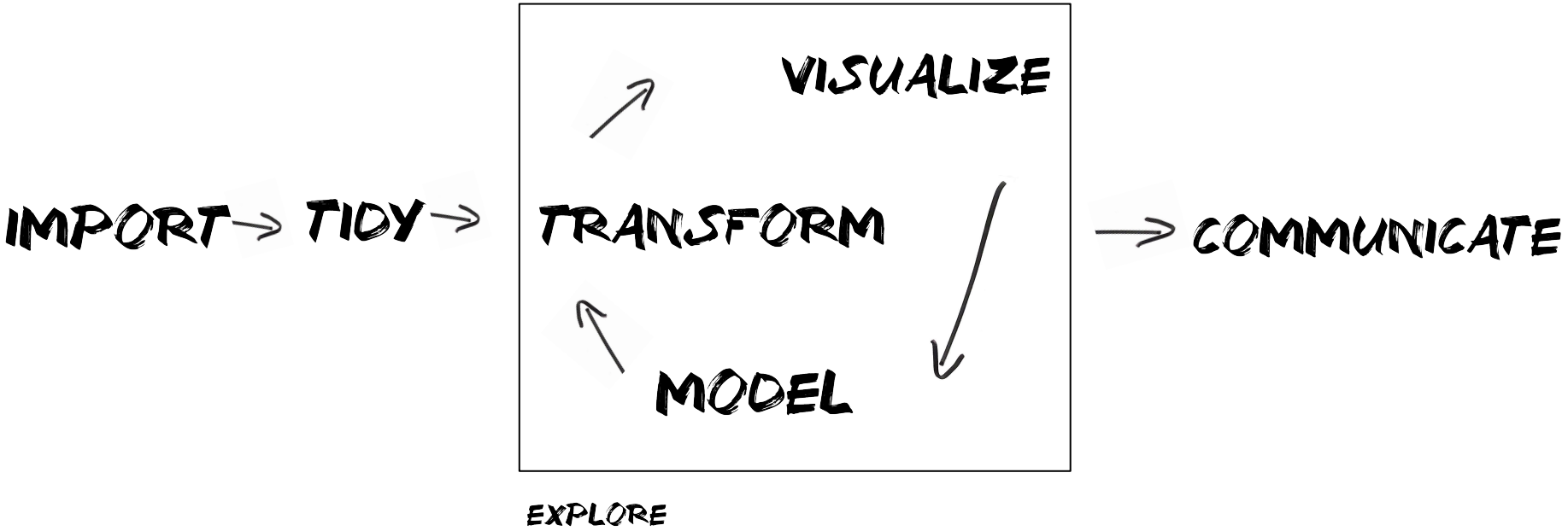
Data Science as an Art

- Intuition
- Qualitative insights
- Exploring a problem through solutions

- “Inspiration exists, but it has to find you working.”
  - Pablo Picasso

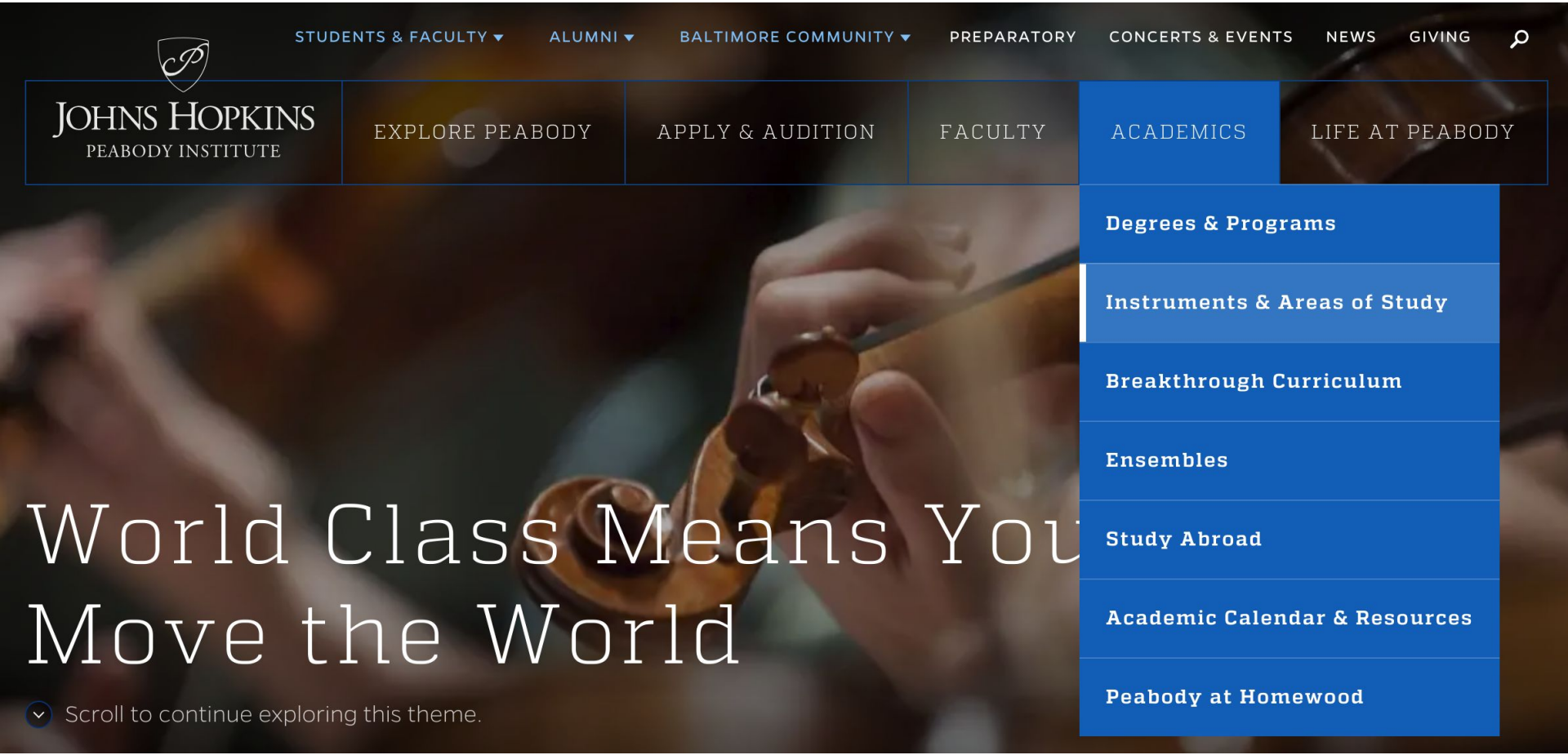
Studio, Tony Wilson (1973) my father. http://www.tonywilsonpainterprintmaker.com/

“The demand for this “right” brain thinking is increasing and in era of increased automation, the need for the “art” of data science will be the increasing cry of business.”



https://r4ds.had.co.nz/explore-intro.html

IF DATA SCIENCE IS AN ART,
WHY DON'T WE TEACH IT
LIKE AN ART?


STUDENTS & FACULTY
ALUMNI
BALTIMORE COMMUNITY
PREPARATORY
CONCERTS & EVENTS
NEWS
GIVING
JOHNS HOPKINS
PEABODY INSTITUTE
EXPLORE PEABODY
APPLY & AUDITION
FACULTY
ACADEMICS
LIFE AT PEABODY
Degrees & Programs
Instruments & Areas of Study
Breakthrough Curriculum
Ensembles
Study Abroad
Academic Calendar & Resources
Peabody at Homewood
World Class Means You
Move the World
Scroll to continue exploring this theme.

Brass

Trumpet, French Horn, Trombone, Tuba, Euphonium

BM MM DMA PC GPD AD

Guitar

BM MM DMA PC GPD AD

Harp

BM MM DMA PC GPD AD

Piano

BM MM DMA PC GPD AD

Organ

BM MM DMA PC GPD AD

Percussion

BM MM DMA GPD AD

Strings

Violin, Viola, Cello, Double Bass

BM MM DMA PC GPD AD

Woodwinds

Flute, Piccolo, Clarinet, Saxophone, Oboe, Bassoon

BM MM DMA PC GPD AD

Music Theory Pedagogy

MM

The Department of Music Theory, including the affiliate areas of Ear-Training and Keyboard Studies, strives to intensify the musicianship of conservatory students and provide the required skills for a profound and concentrated understanding of music.

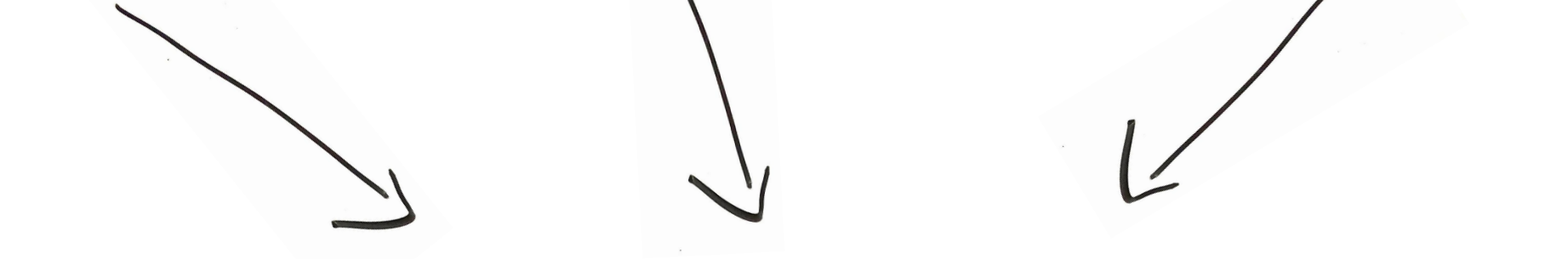
LEARN MORE

Composition

BM MM DMA

Regarded as among the world's elite training grounds for more than a century, the department of composition at Peabody is one of the most sought after programs of its kind in the world. As a division of The Johns Hopkins University, Peabody takes its place alongside the institution's other internationally renowned centers of research and learning, shaping the role of music in the 21st century.

LEARN MORE



music theory :: statistical theory

instrument :: programming language

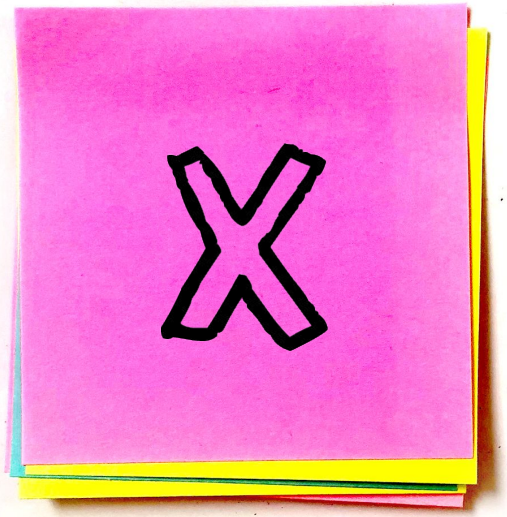
composition :: narrative

music theory :: statistical theory

music theory :: statistical theory

instrument :: programming language

instrument :: programming language



MOSTLY:

WHAT I'VE LEARNED FROM YOU

PROGRAMMING IN R

FOR PURELY STATISTICAL/ACADEMIC PURPOSES

LIMITS CREATIVITY

COOK-BOOKERY

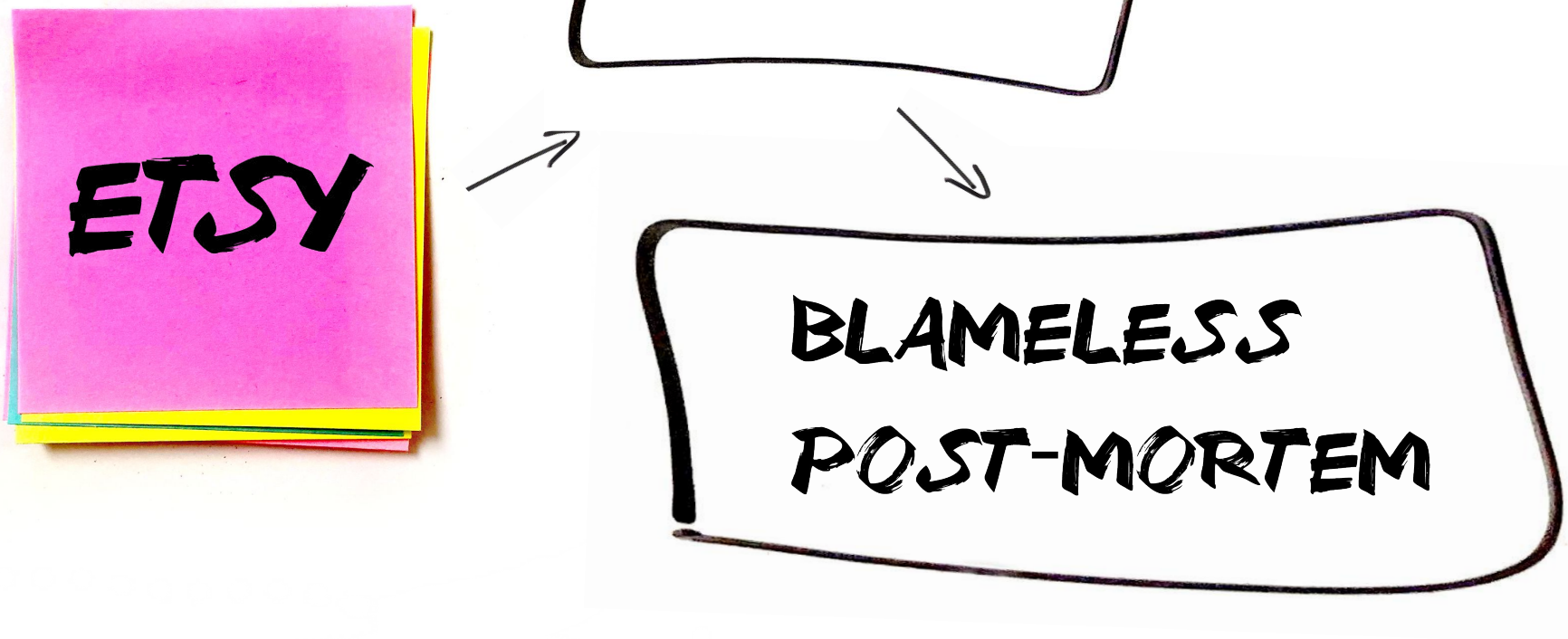
ProjectTemplate

Architecture

ProjectTemplate is based on the idea that you should structure all of your data analysis projects in the same way so that you can exploit conventions instead of writing boilerplate code. Because so much of ProjectTemplate's functionality is based on conventions, it's worth explaining ProjectTemplate's idealized project in some detail.

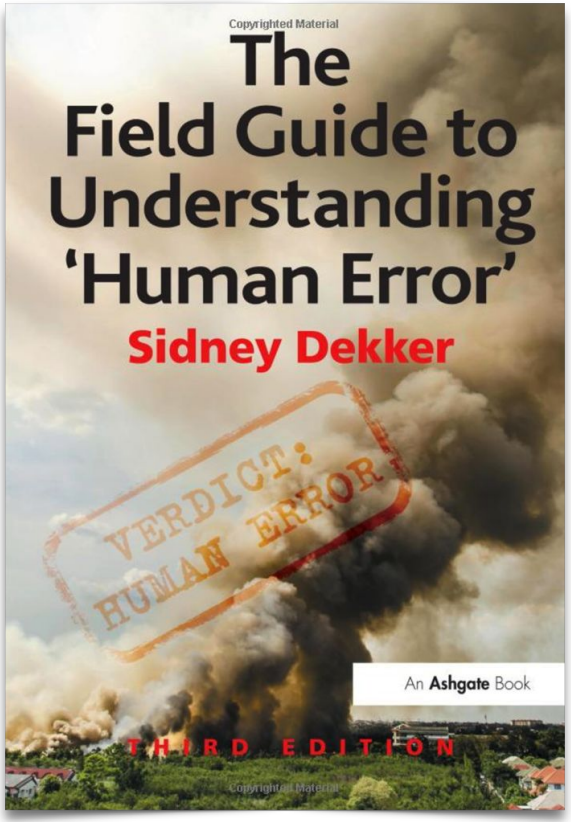
TOOK THE COGNATIVE BURDEN OFF OF FORMATTING

LIKE RUBY ON RAILS, BUT FOR ANALYSIS CODE


ETSY
DEVOPS
BLAMELESS
POST-MORTEM

GOOD ANALYST V. BAD ANALYST

You re-run the analysis and get different results.
Someone else can't repeat the analysis.
You can't re-run the analysis on different data.
An external library you're using is updated, and you can't recreate the results.
You change code but don't re-run downstream code, so the results aren't consistently updated.
You change code but don't re-execute it, so the results are out of date.
You update your analysis and can't compare it to previous results.
You can't point to code changes that resulted in a different analysis results.
A second analyst can't understand your code.
Can you re-use logic in different parts of the analysis?
You change the logic used in analysis, but only in some places where it's implemented.
Your code is not performing as expected, but you don't notice.
Your data becomes corrupted, but you don't notice.
You use a new dataset that renders your statistical methods invalid, but you don't notice.
You make a mistake in your code.
You use inefficient code.
A second analyst wants to contribute code to the analysis, but can't do so.
Two analysts want to combine code but cannot.
You aren't able to track and communicate known next steps in your analysis.
Your collaborators can only make requests in email or meetings, and they aren't incorporated into the project.



"An engineer who thinks they're going to be reprimanded is disincentivized to give the details necessary to get an understanding of the mechanism, pathology and operation of the failure."

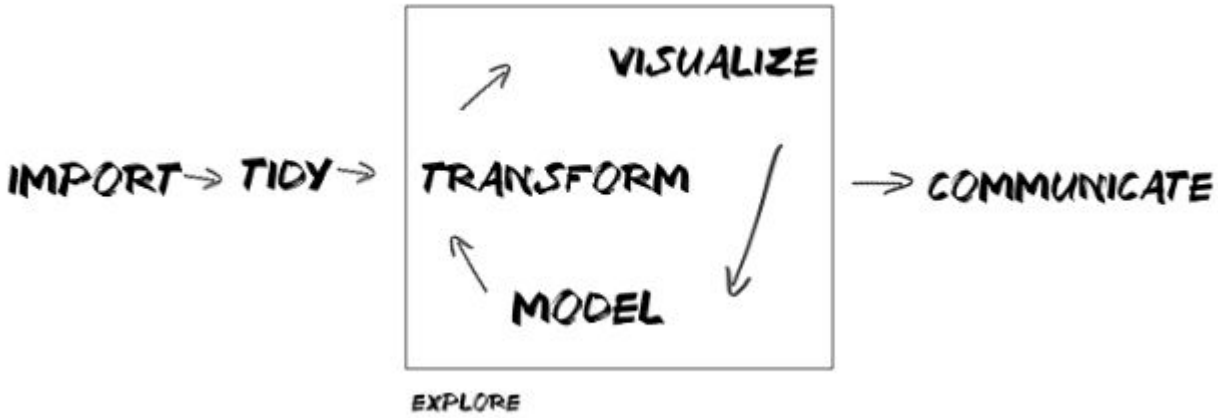
– John Allspaw

BLAMELESS PORTMORTEMS

ALLOWED ME TO TALK ABOUT CONCEPTS IN A WAY THAT DIDN'T MAKE PEOPLE DEFENSIVE

ALLOWED ME TO TALK ABOUT GENERAL PRINCIPLES RATHER THAN SPECIFIC LANGUAGE CHOICES

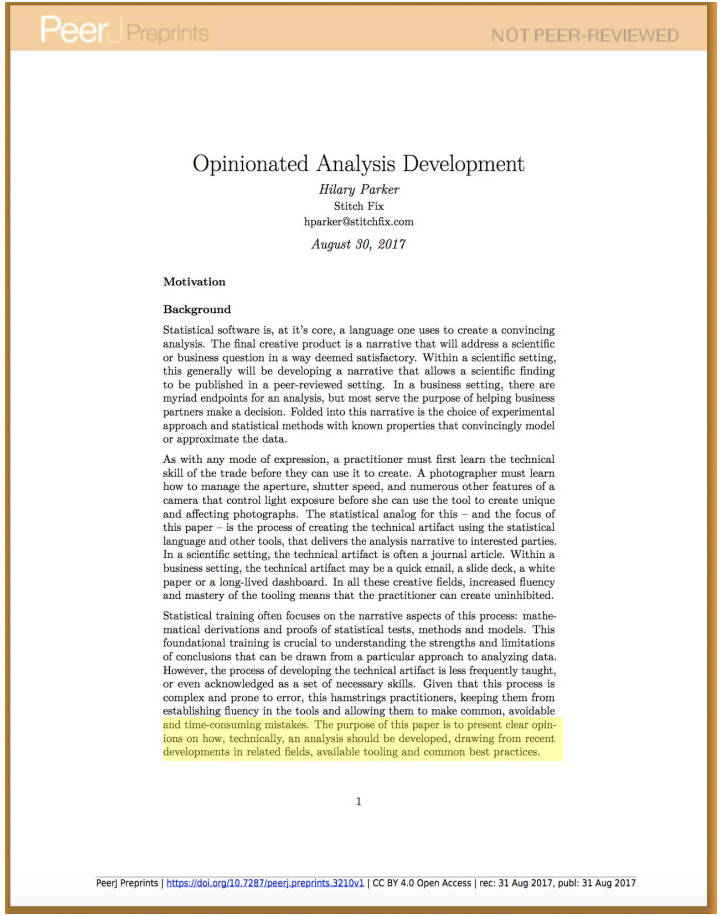
BUILDING A SYSTEM



I HEDGED MY BETS

"the purpose of this paper is to established the opinions for developing the technical artifact, rather than developing the narrative of an analysis."

AND LEFT THE NARRATIVE OUT



PeerJ Preprints NOT PEER-REVIEWED

Opinionated Analysis Development

Hilary Parker
Stitch Fix
hparker@stitchfix.com

August 30, 2017

Motivation

Background

Statistical software is, at it's core, a language one uses to create a convincing analysis. The final creative product is a narrative that will address a scientific or business question in a way deemed satisfactory. Within a scientific setting, this generally will be developing a narrative that allows a scientific finding to be published in a peer-reviewed setting. In a business setting, there are myriad endpoints for an analysis, but most serve the purpose of helping business partners make a decision. Folded into this narrative is the choice of experimental approach and statistical methods with known properties that convincingly model or approximate the data.

As with any mode of expression, a practitioner must first learn the technical skill of the trade before they can use it to create. A photographer must learn how to manage the aperture, shutter speed, and numerous other features of a camera that control light exposure before she can use the tool to create unique and affecting photographs. The statistical analog for this – and the focus of this paper – is the process of creating the technical artifact using the statistical language and other tools, that delivers the analysis narrative to interested parties. In a scientific setting, the technical artifact is often a journal article. Within a business setting, the technical artifact may be a quick email, a slide deck, a white paper or a long-lived dashboard. In all these creative fields, increased fluency and mastery of the tooling means that the practitioner can create uninhibited.

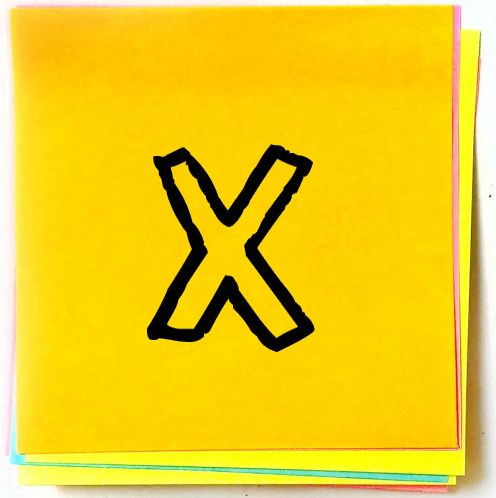
Statistical training often focuses on the narrative aspects of this process: mathematical derivations and proofs of statistical tests, methods and models. This foundational training is crucial to understanding the strengths and limitations of conclusions that can be drawn from a particular approach to analyzing data. However, the process of developing the technical artifact is less frequently taught, or even acknowledged as a set of necessary skills. Given that this process is complex and prone to error, this hamstrings practitioners, keeping them from establishing fluency in the tools and allowing them to make common, avoidable and time-consuming mistakes. The purpose of this paper is to present clear opinions on how, technically, an analysis should be developed, drawing from recent developments in related fields, available tooling and common best practices.

1

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3210v1 | CC BY 4.0 Open Access | rec: 31 Aug 2017, publ: 31 Aug 2017

composition :: narrative

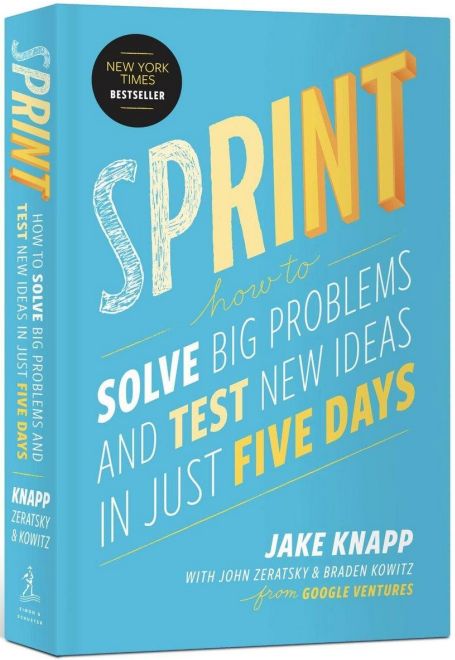
composition :: narrative

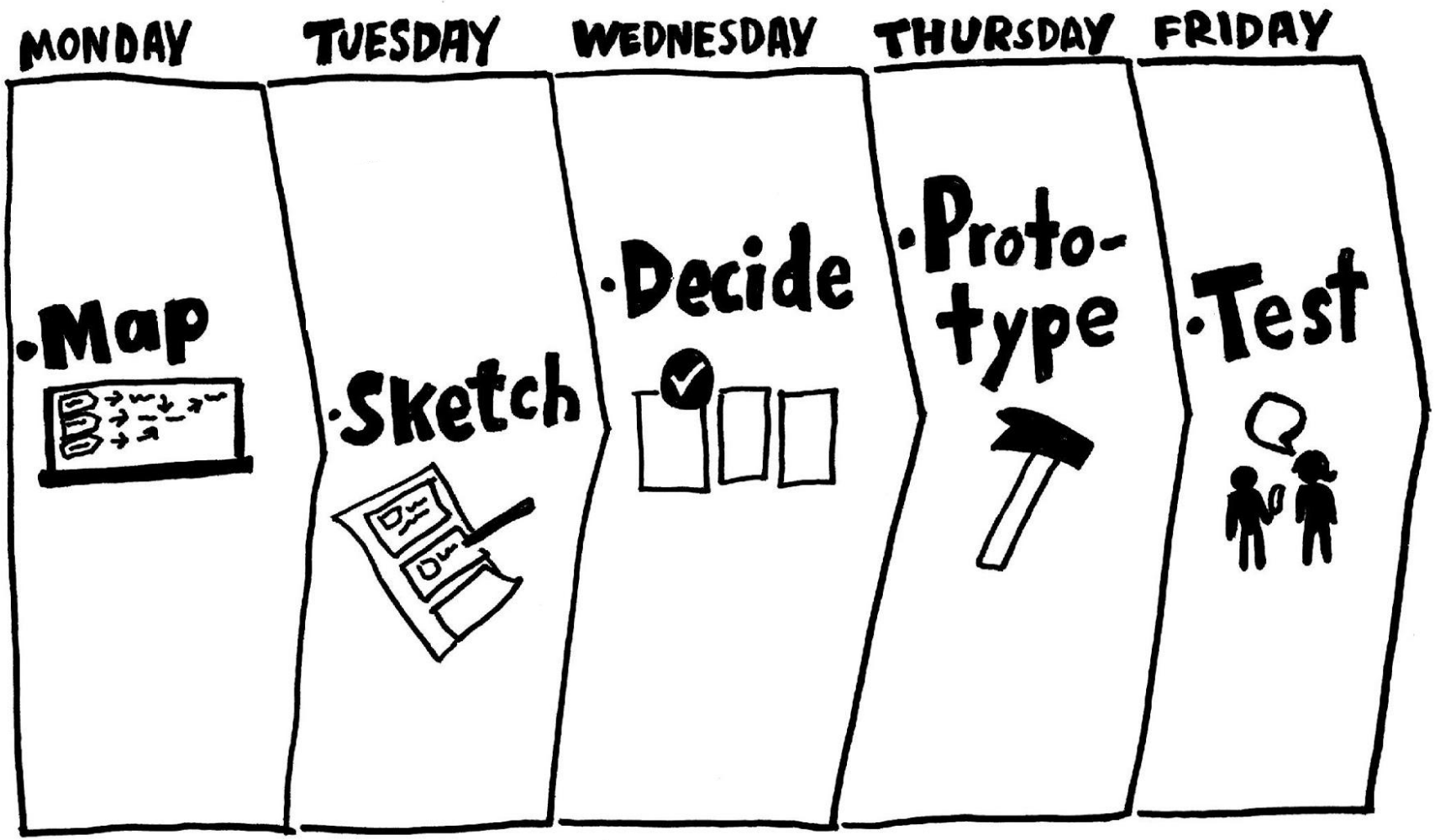


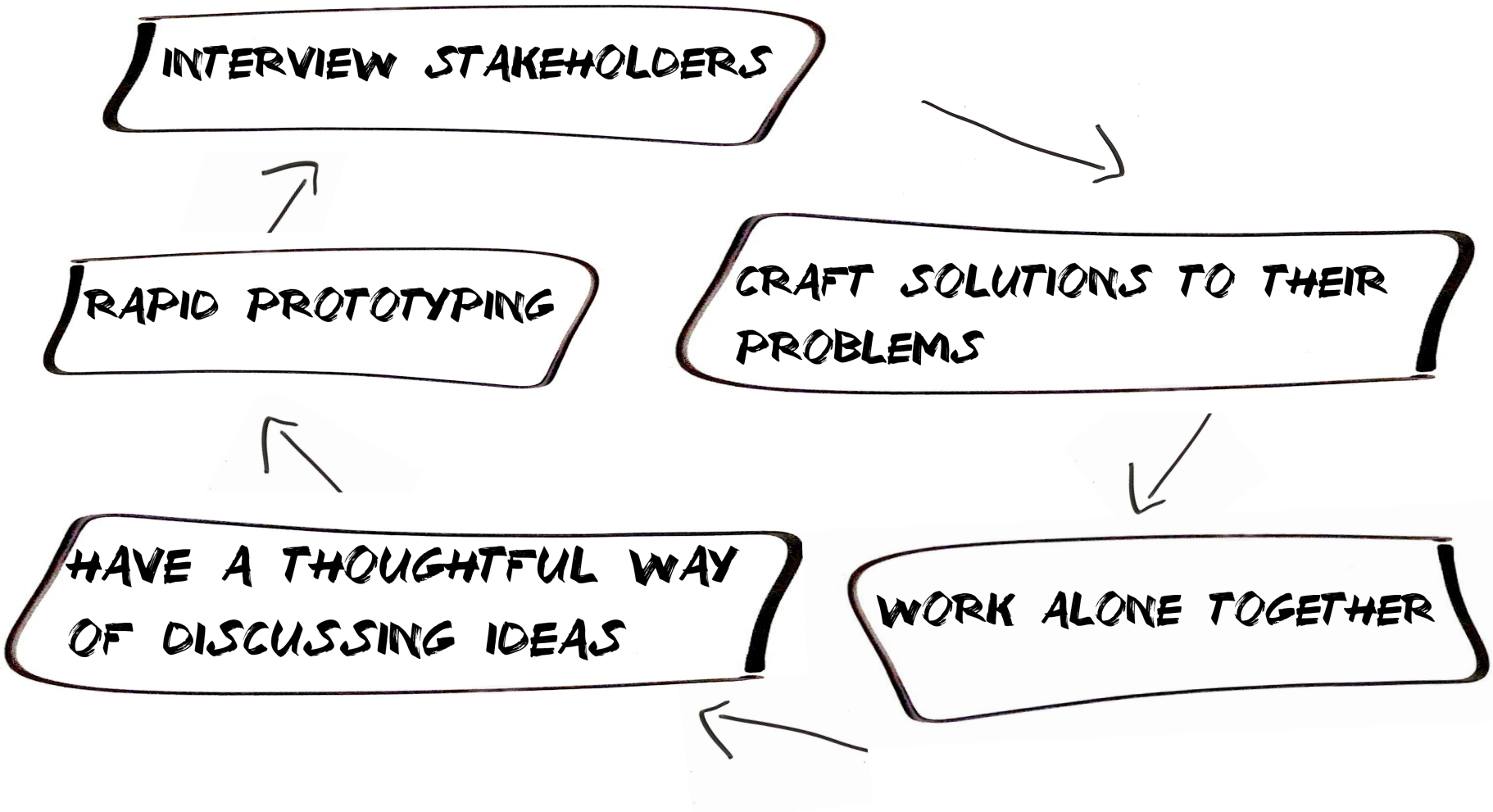
Software is the invisible writing that whispers the stories of possibility to our hardware. You are the storyteller.

Grady Booch

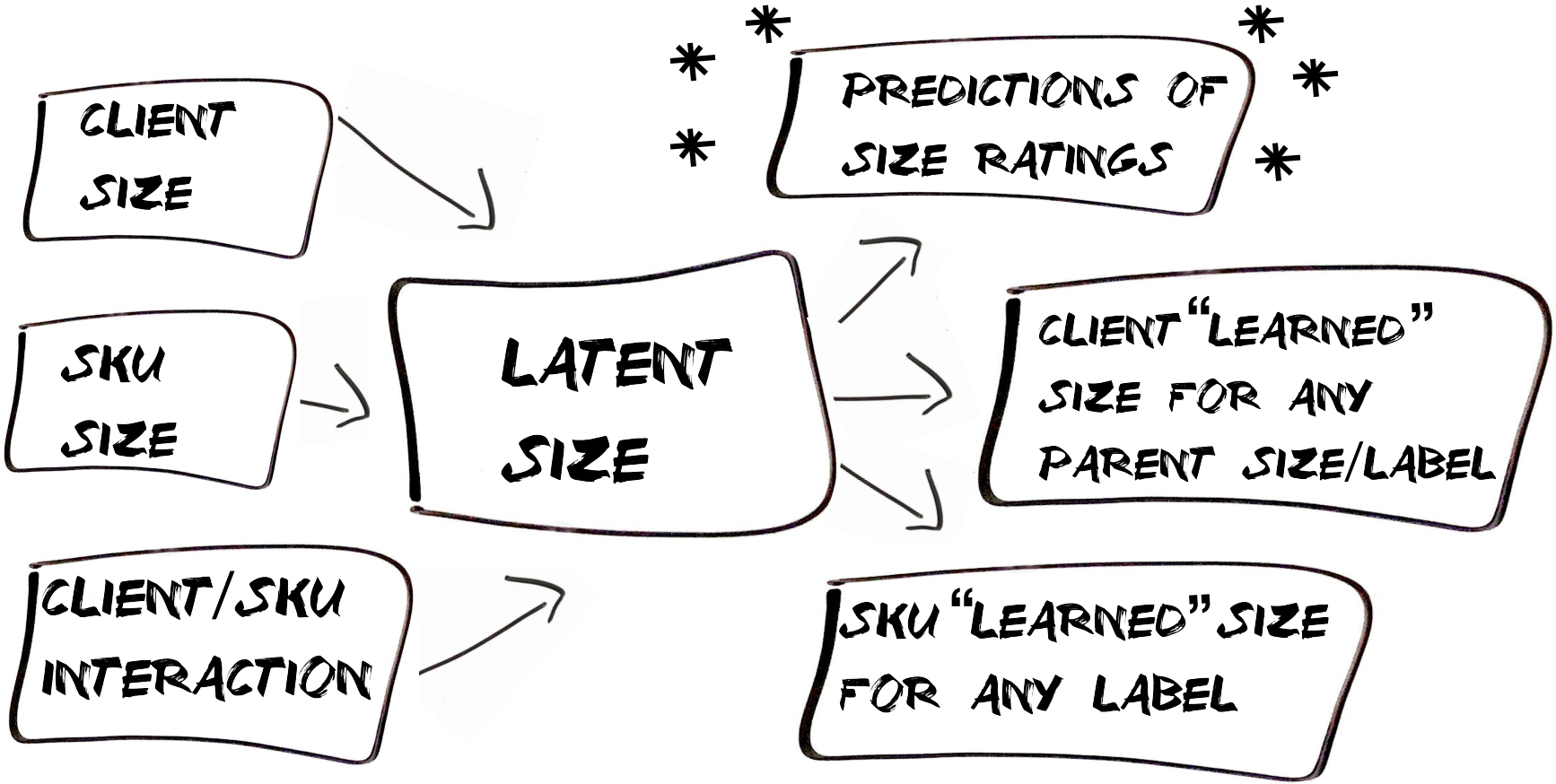
PRODUCT DEVELOPMENT




MONDAY
TUESDAY
WEDNESDAY
THURSDAY
FRIDAY
·Map
·Sketch
·Decide
·Proto-type
·Test


INTERVIEW STAKEHOLDERS
CRAFT SOLUTIONS TO THEIR PROBLEMS
WORK ALONE TOGETHER
HAVE A THOUGHTFUL WAY OF DISCUSSING IDEAS
RAPID PROTOTYPING

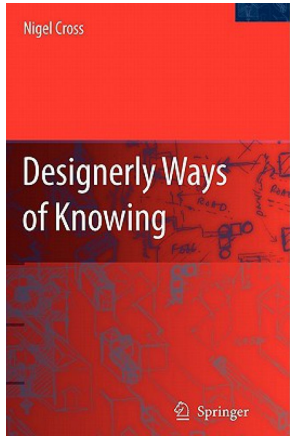

RULES
MERCH INTENT
FIT & AESTHETIC
BRAND LOYALTY
w/in size variation



https://multithreaded.stitchfix.com/blog/2017/12/13/latentsize/

DESIGN IS NOT A "MYTHICAL" OR "MYSTERIOUS TALENT"

IT'S A DISCIPLINE →



“Data analysis is hard, and part of the problem is that few people can explain how to do it. It’s not that there aren’t any people doing data analysis on a regular basis. It’s that the people who are really good at it have yet to enlighten us about the thought process that goes on in their heads.”

THE DESIGN PROCESS

AS OPPOSED TO PROBLEM FOCUSED

“You’ve told me everything **NOT** to do, but how will I know what to do?”

Anonymous Roger Peng student

THE DESIGN PROCESS

YOU NEED TO START BUILDING TO UNDERSTAND THE PROBLEM FULLY

Data Science as an Art

- Intuition
- Qualitative insights
- Exploring a problem through solutions

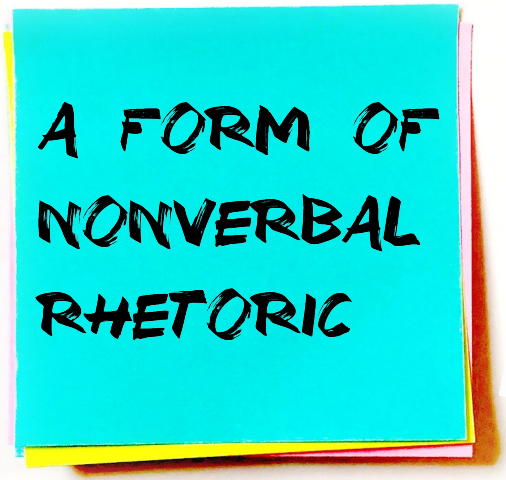
- “Inspiration exists, but it has to find you working.”
  - Pablo Picasso

Studio, Tony Wilson (1973) my father. http://www.tonywilsonpainterprintmaker.com/

THE DESIGN PROCESS

“The demand for this “right” brain thinking is increasing and in era of increased automation, the need for the “art” of data science will be the increasing cry of business.”

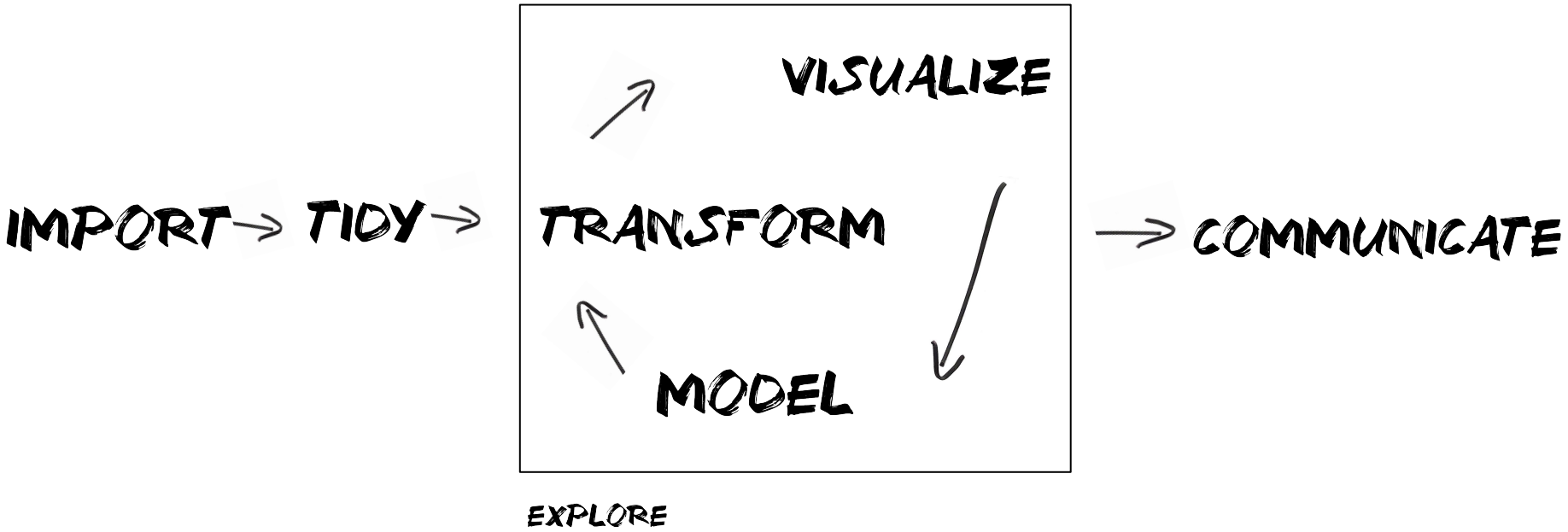
THE DESIGN PROCESS



WITH SKETCHING
AS THE LANGUAGE

One thing that is clear is that sketches enable designers to handle different levels of abstraction simultaneously… Clearly this is something important in the design process. We see that designers think about the overall concept and at the same time think about detailed aspects of the implementation of that concept.

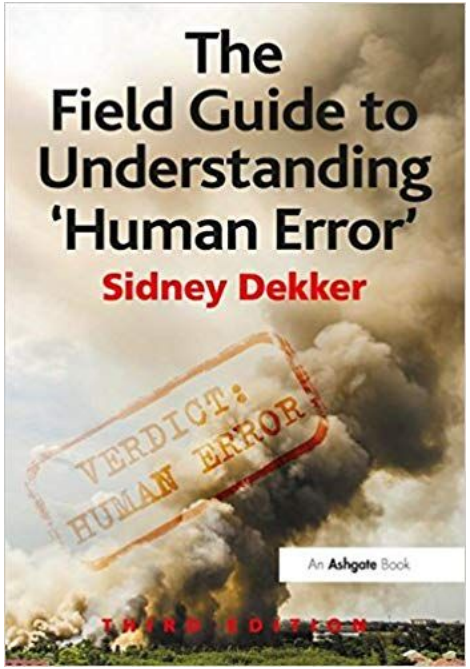
Nigel Cross

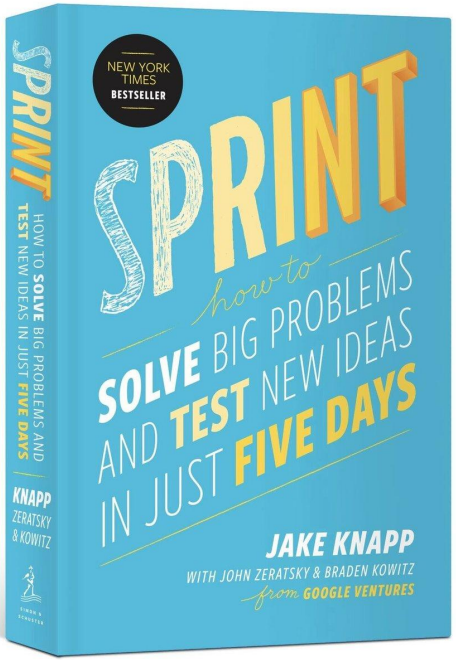

IMPORT
TIDY
VISUALIZE
TRANSFORM
MODEL
COMMUNICATE
EXPLORE

Design ability is, in fact, one of the three fundamental dimensions of human intelligence. Design, science, and art form an ‘and’ not an ‘or’ relationship to create the incredible human cognitive ability.”

Nigel Cross

WHAT DO THEY HAVE IN COMMON





TWO SIDES OF THE SAME COIN

RULES FOR PLAYING NICE

EMPATHY

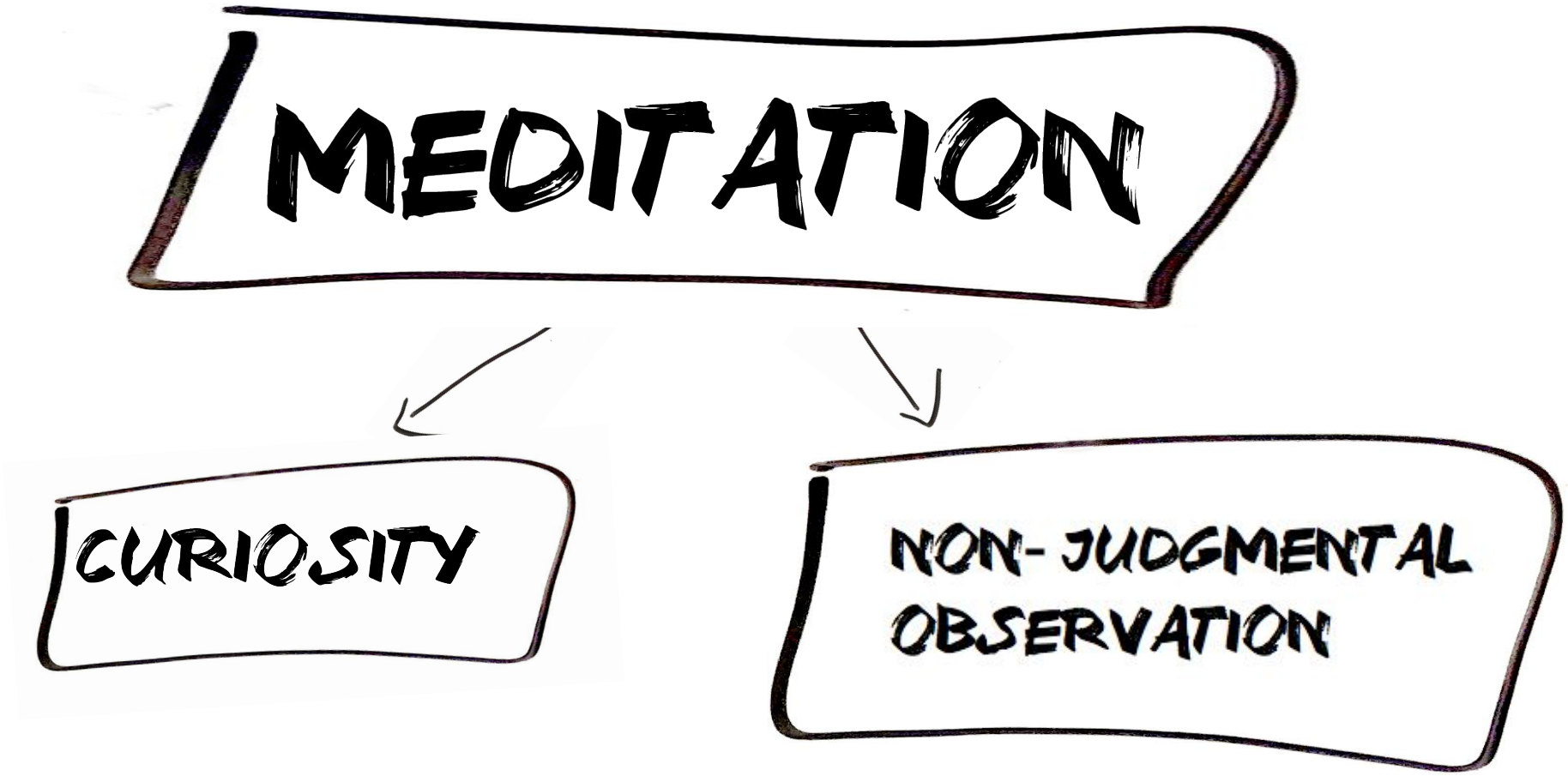
the capacity to understand or feel what another person is experiencing from within their frame of reference, i.e., the capacity to place oneself in another's position.

effective teams

https://rework.withgoogle.com/blog/five-keys-to-a-successful-google-team/

WHAT IF YOU ARE NOT AN EMPATHETIC PERSON?

MY EXPERIENCE WITH MEDITATION


MEDITATION
CURIOSITY
NON-JUDGMENTAL
OBSERVATION

SELF

OTHERS

WITH USERS

WITH CO-CREATORS

WITH SELF

THANK YOU