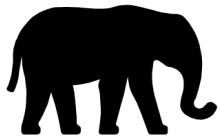
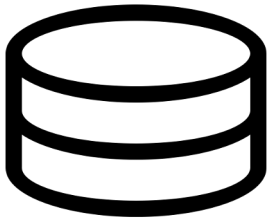
Why experimental design matters and how you can make sure it’s pretty good.



Koonkie

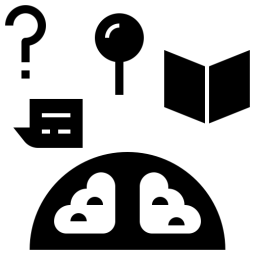
Aria Hahn, PhD
CEO & Co-Founder

… a way to have new thoughts about the world



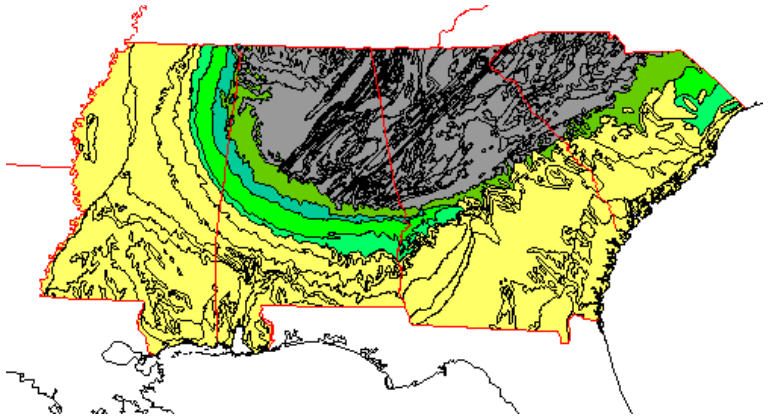
Data

Information

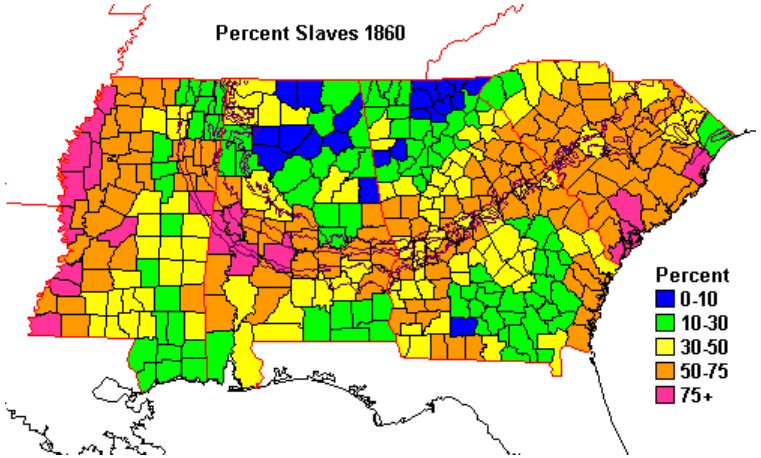


Knowledge

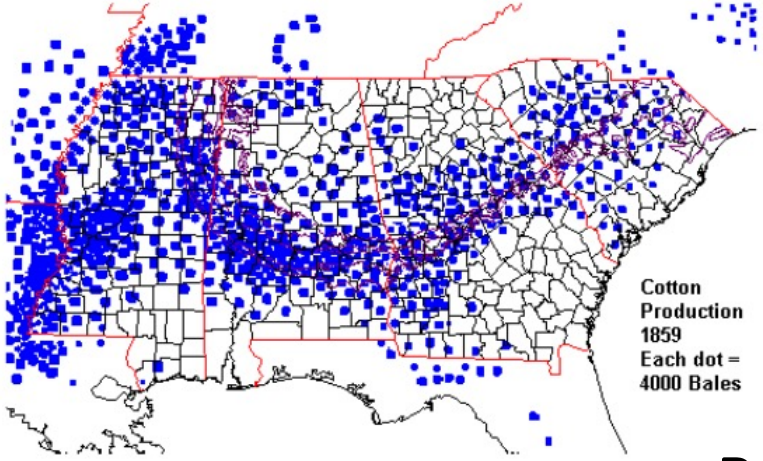
… sometimes it's fun



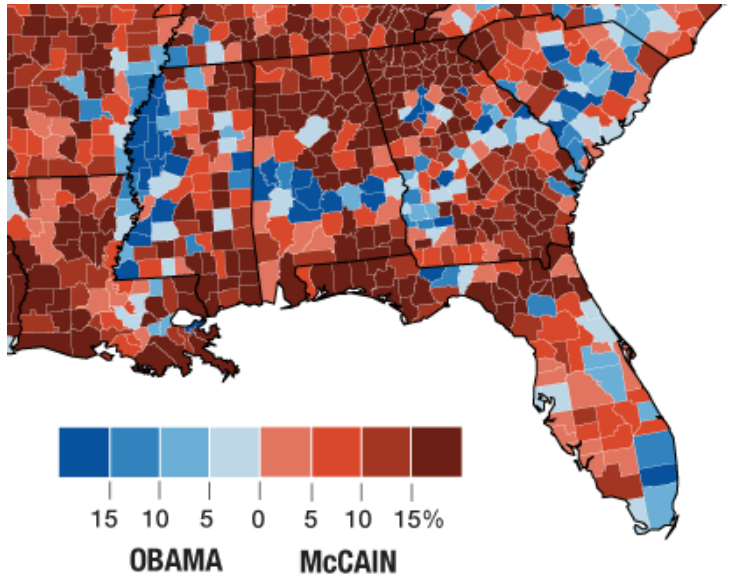
A



C



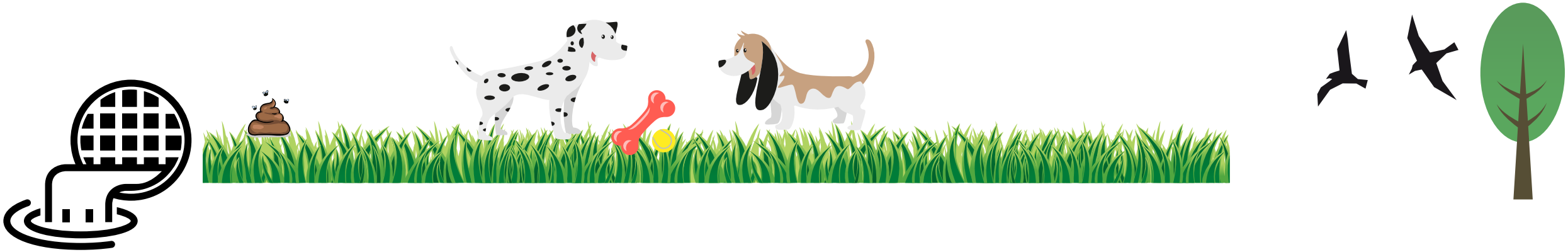
B

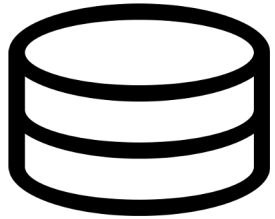


D

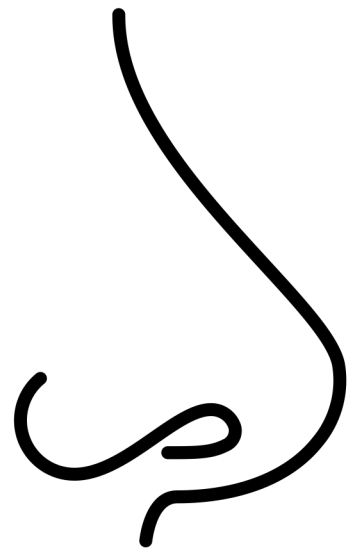
Steven Dutch/2000 Election Results

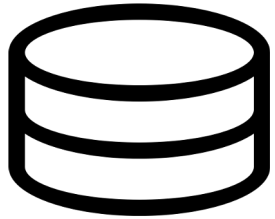
… or necessary



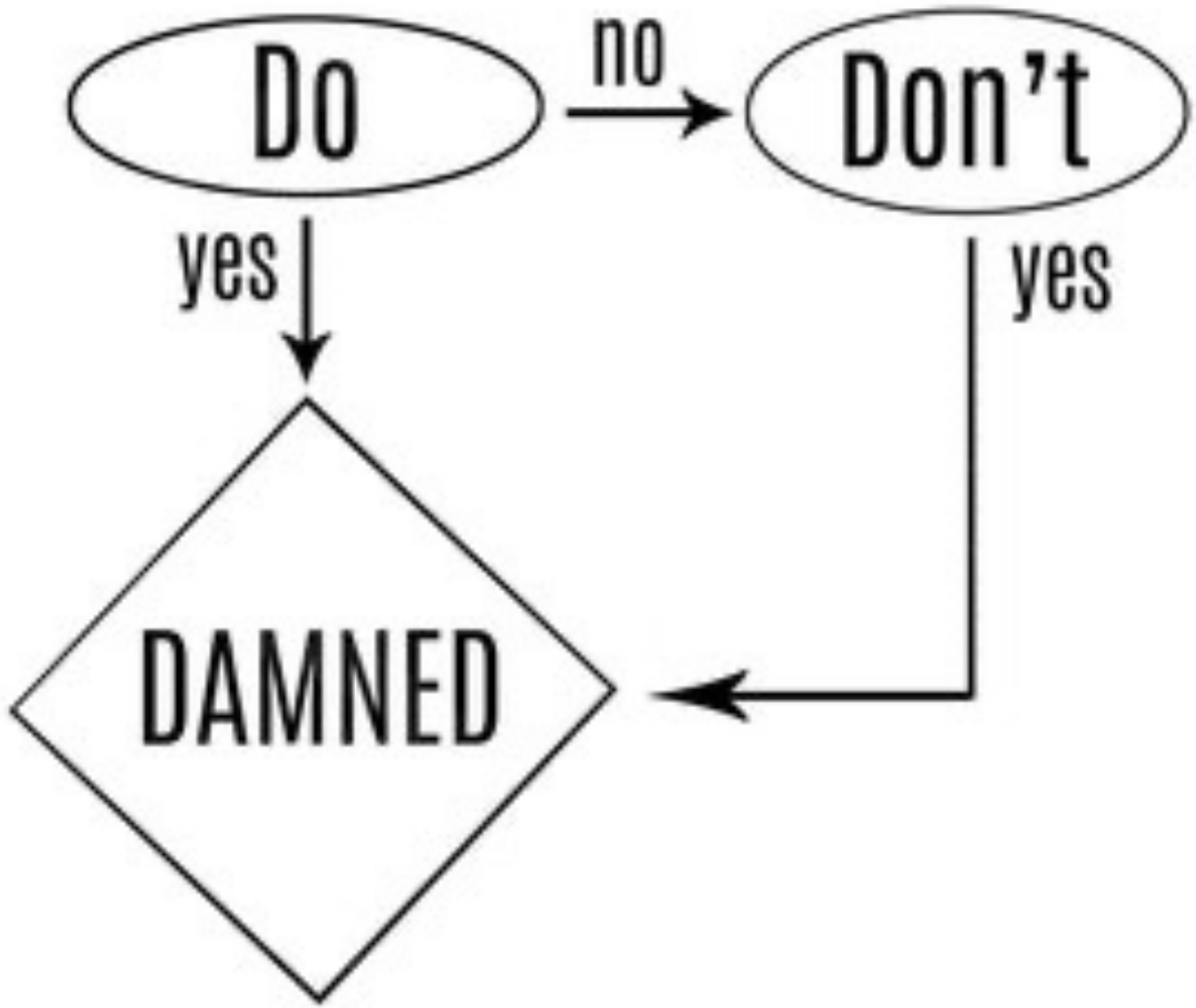


WRNOG





WRNOG


Do
no
Don't
yes
yes
DAMNED

Steps

7 Days

7-Day Average

18,599 steps

7-Day Total

111,046 steps

Weekly Goal Progress

159% of 70,000

58 bpm

resting heart rate

5.5 lbs

to go

Recent Exercise

Running (jog...

Jul 13

10 minutes

97 cals

163 avg bpm

Run

Jul 13

Market Summary > Fitbit Inc

NYSE: FIT

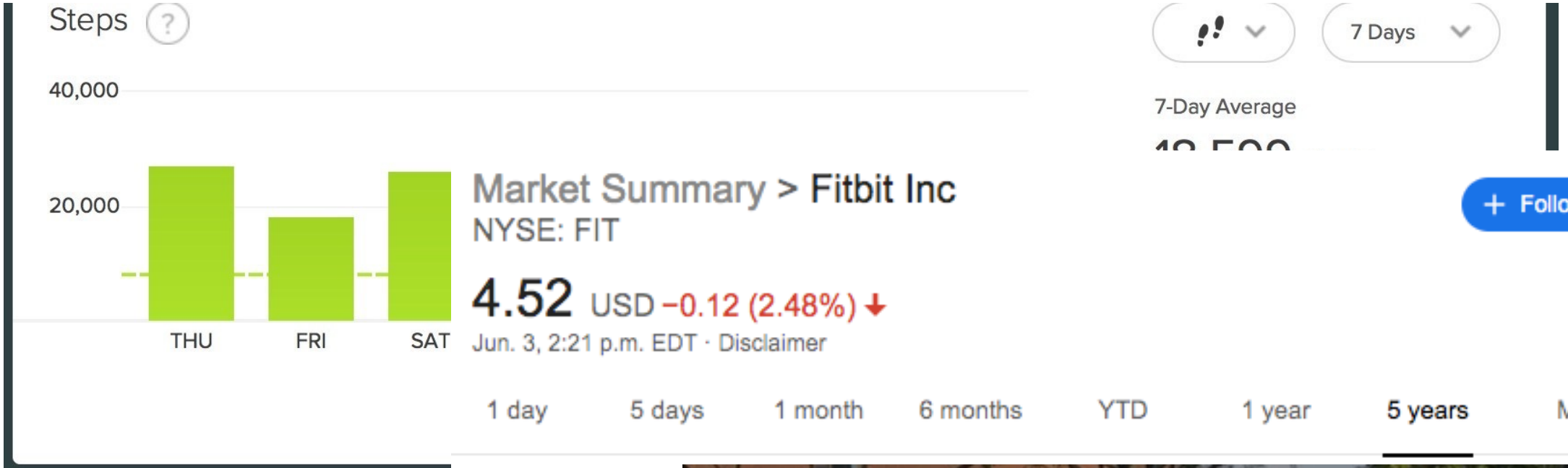
+ Follow

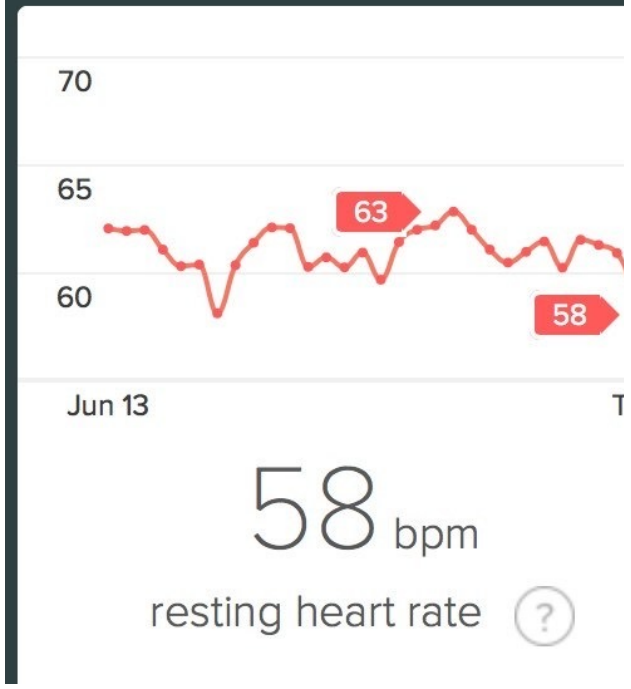
4.52 USD −0.12 (2.48%) ↓

Jun. 3, 2:21 p.m. EDT · Disclaimer

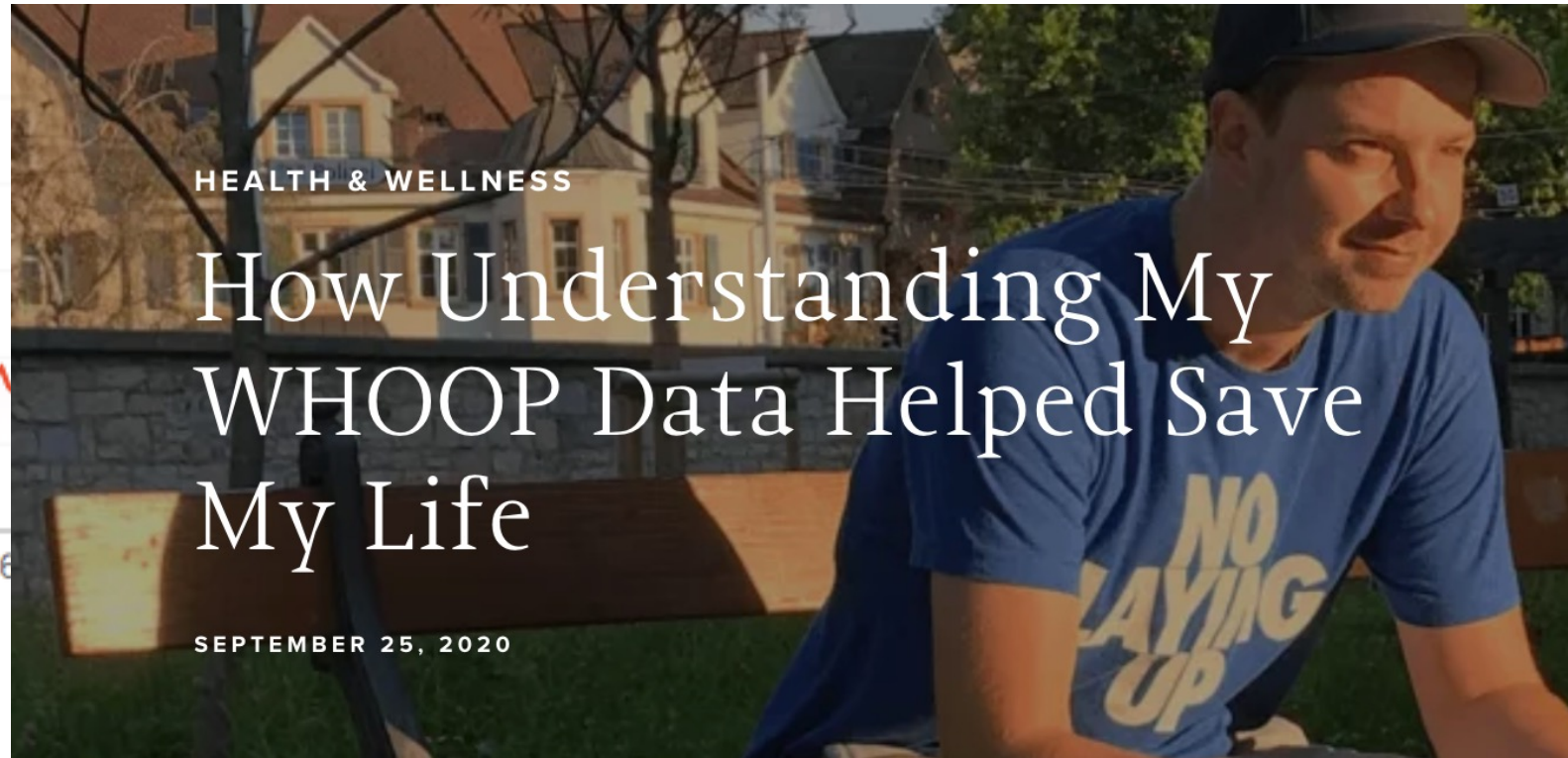
1 day | 5 days | 1 month | 6 months | YTD | 1 year | **5 years** | Max

Open	4.62	Div yield	-
High	4.62	Prev close	4.63
Low	4.46	52-wk high	7.79
Mkt cap	1.15B	52-wk low	4.23
P/E ratio	-		


Steps
40,000
20,000
THU
FRI
SAT
7 Days
7-Day Average
Market Summary > Fitbit Inc
NYSE: FIT
4.52 USD −0.12 (2.48%)
Jun. 3, 2:21 p.m. EDT · Disclaimer
Follow
1 day
5 days
1 month
6 months
YTD
1 year
5 years
Max


70
65
60
63
58
Jun 13
58 bpm
resting heart rate

50
40
30
20
10
0
2016
Open
High


HEALTH & WELLNESS
How Understanding My WHOOP Data Helped Save My Life
SEPTEMBER 25, 2020
NO LAYING UP

10^{30}

Image: Hubble Space Telescope, NASA
10^{22}

1. Provides unbiased estimates of what you care about
2. Enables you to detect a signal
3. Includes the plan for analysis
4. Gives results that are easy to interpret
5. Permits conclusions that have wide validity
6. Is as simple as it can be

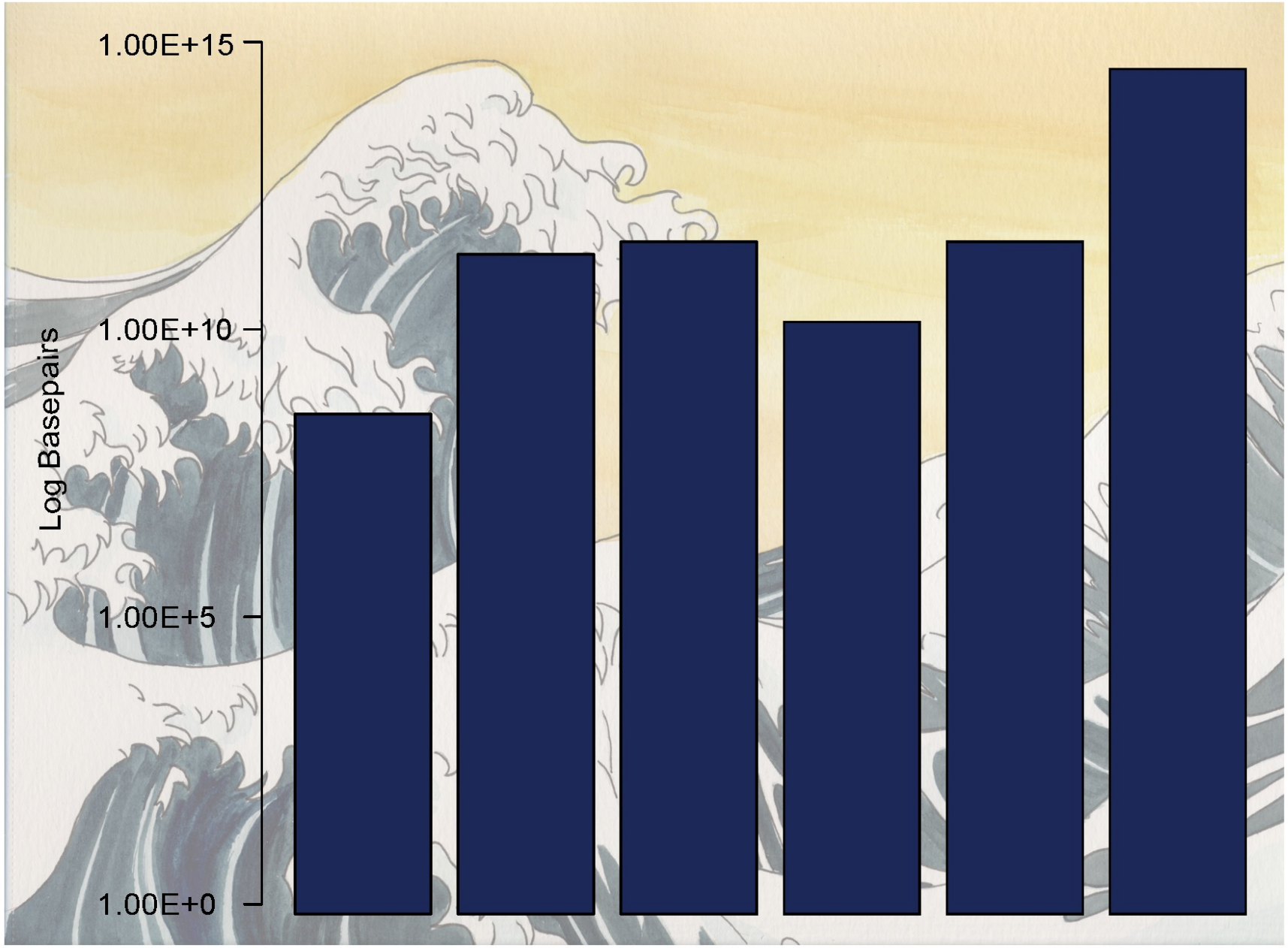
… yeah, but

- Don’t have straight forward questions
- Don’t really know what I am looking for
- Data is (very) imperfect
- No way to remove the noise
- I want to graduate
- Just let me publish/ My supervisor never emails me back
- I have a budget/ it costs how much?
- I don’t know anything about stats/ Can you do it for me?
- I am exploring
- My friend in stats said I couldn’t
- I have to go out in the middle of the ocean/ forest and it’s heavy
- I’d need what kind of computer?
- I don’t know how to do that part
- …..

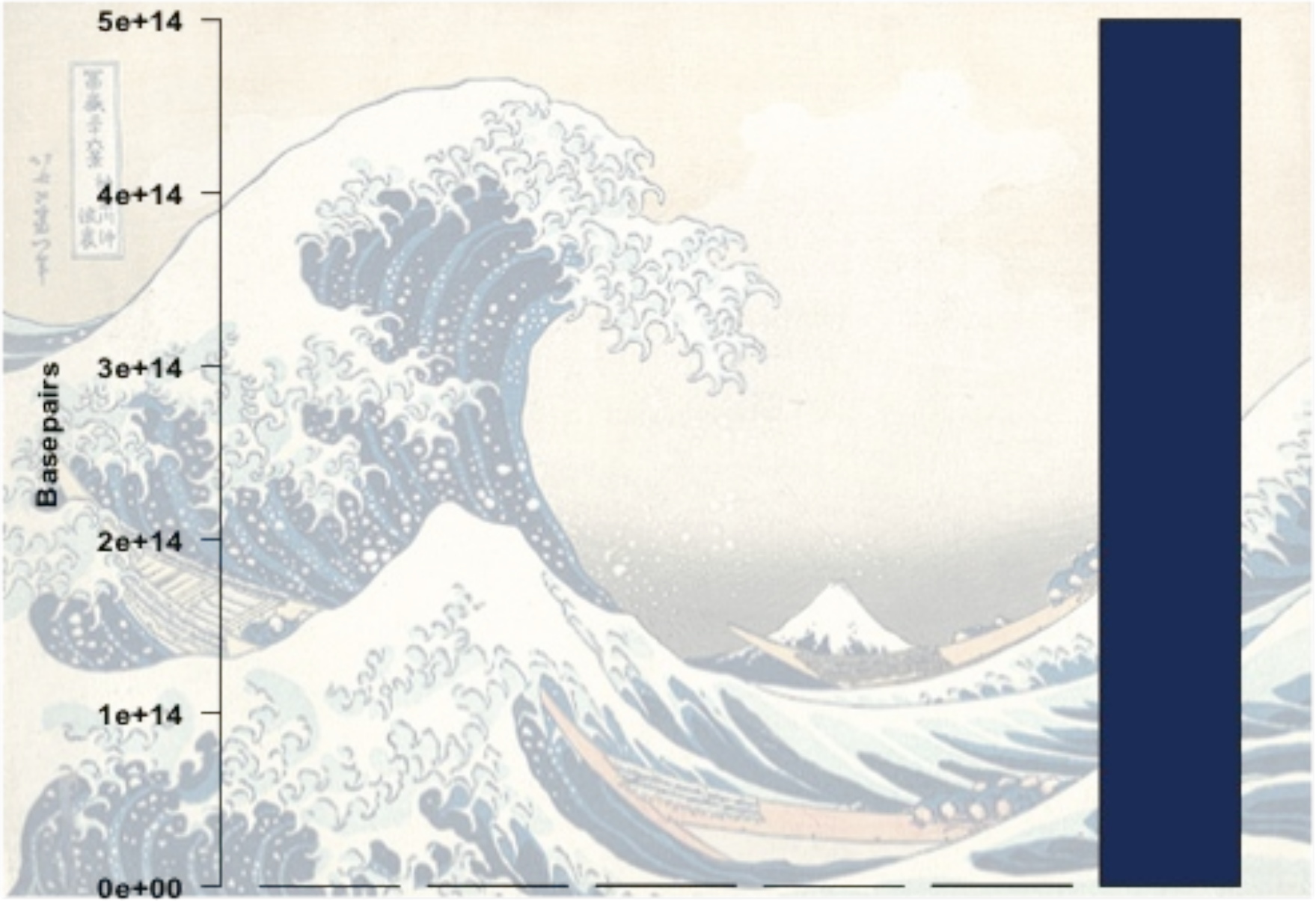
You do know something!

Sample Type	Effective size (bp)
E. coli genome	5,000,000
Human genome	3,000,000,000
Vertebrate transcriptome	5,000,000,000
Human gut	200,000,000
Marine	5,000,000,000
Soil	5,000,000,000,000

How much sequence data would be needed to assemble these metagenomes assuming 100X coverage?



http://ivory.idyll.org/blog/how-much-sequencing-is-needed.html



http://ivory.idyll.org/blog/how-much-sequencing-is-needed.html

1. Provides unbiased estimates of what you care about
    - Spend time understanding what bias you do have in your design (read as much as you can)
2. Enables you to detect a signal
    - Get as much data as you can afford and if you know very little get replicates (BEFORE more treatments)
3. Includes the plan for analysis
    - Test fewer things well rather than more things poorly. Talk to some experts and read ☺ Imagine the figures you would like to create. Save papers that have figures that are informative. Getting help replicating something is much easier than asking open ended questions
4. Gives results that are easy to interpret
    - This requires expertise and experience. Some of this is earned. This is also why we work in teams and collaborate. Be open to learning more and incorporating that (sometimes harsh) feedback. Be ready to iterate.
5. Permits conclusions that have wide validity
    - Select a journal with the right impact. It's not always possible to publish in Science/Nature– but it's almost always possible to generate excellent hypotheses and conceptual models
6. Is as simple as it can be (BIG fan of this)

… yeah, but

- Don’t have straight forward questions – deep breaths, data mining can be learned (practice on public data!)
- Don’t really know what I am looking for - read
- Data is (very) imperfect – yours and everyone elses. Seek to understand not fix this
- No way to remove the noise – if you can get replicates and fewer conditions, stats can help (and its learnable)
- I want to graduate – been there…
- Just let me publish/ My supervisor never emails me back – it’s a grind. Lean on other support
- I have a budget/ it costs how much? – do what you can - Narrow and deeper is generally preferable
- I don’t know anything about stats/ Can you do it for me? – yeah, and most campuses have resources
- I am exploring – write a review or complete a meta study!
- My friend in stats said I couldn’t – understand the rules (then break them)
- I have to go out in the middle of the ocean/ forest and it’s heavy - Do what you can (muscles for science)
- I’d need what kind of computer? – Big Data is a whole skillset…there are times to outsource
- I don’t know how to do that part – read (I know I know…I am broken record)
- …..

- Invest the time before you get going

It’s much easier to graduate if you have a good design or at least understand the pros/cons before you get started.

- Replicates are almost always the best possible thing

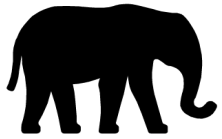
This helps you perform almost every statistical methods and test out there…it also makes for better descriptive statistics.

- Get the work done

The textbook approach is often not possible. That’s what makes this interesting and keeps us all busy.

www.koonkie.com

services@koonkie.com



Koonkie

Aria Hahn, PhD
CEO & Co-Founder