

“Well, how did I get here?”

Talking Heads *Once in a Lifetime*

Gina Merchant

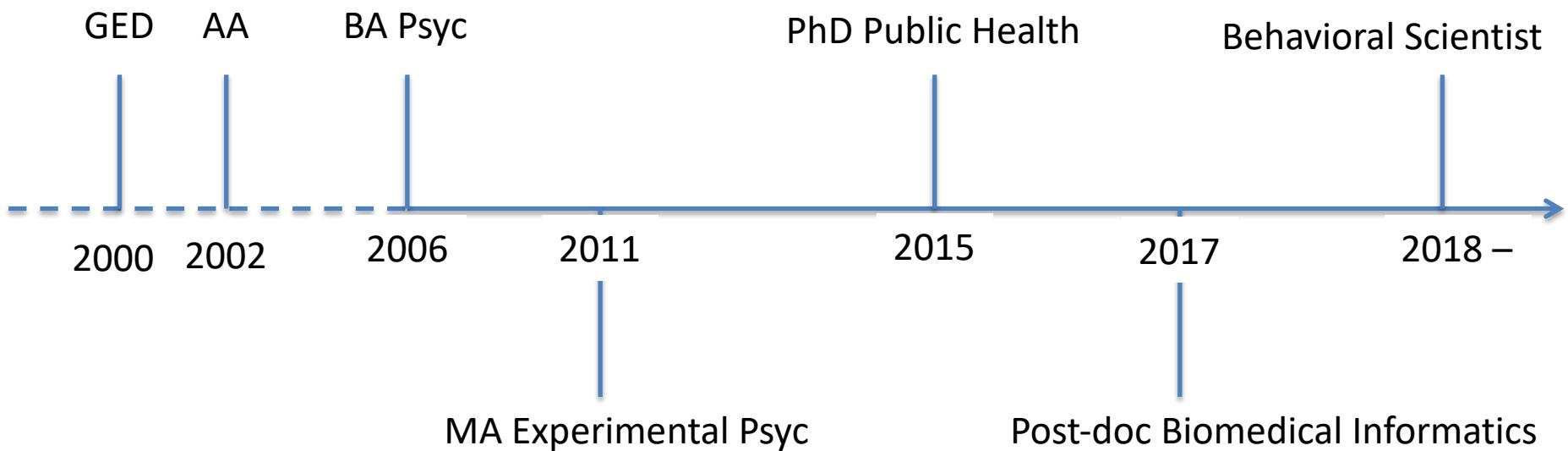
Senior Behavioral Scientist, Medical Affairs, ResMed

1. My background

- How I got into data science
- How data science makes me a better behavioral scientist
- Tools, tips

2. Q & A

Academic path to industry role



1. Know what data you don't have

What is Data Science?

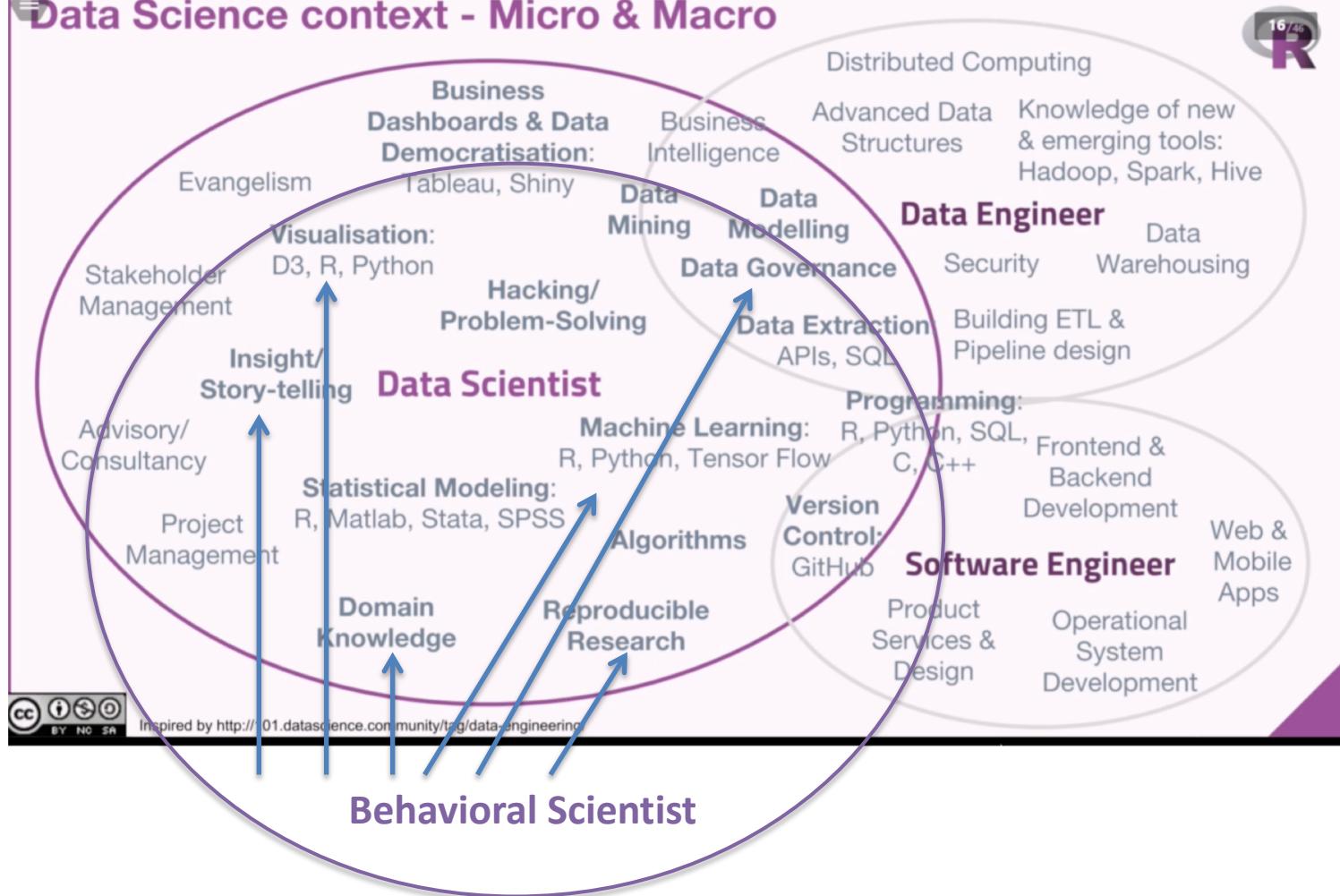


Our suggestion:

“The application of scientific method to Data which enable new discoveries to be made & utilised”

- General consensus = There is no consensus
- DJ Patil & Jeff Hammerbacher purport to first coined the term “Data Scientist” as a title in 2008
- Hybrid & Interdisciplinary
- P.S Don't assume an Organisation recruits for a Data Scientist actually knows what a Data Scientist is.....!

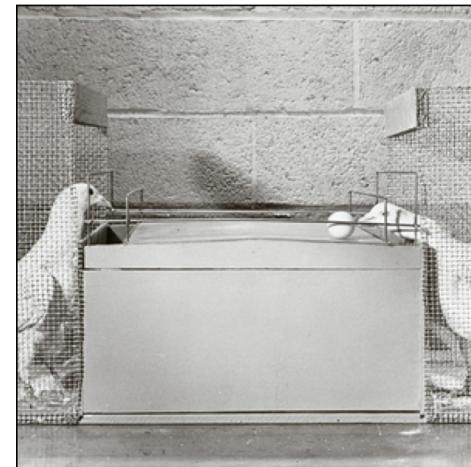
Data Science context - Micro & Macro



2. Know where you contribute

What is behavioral science?

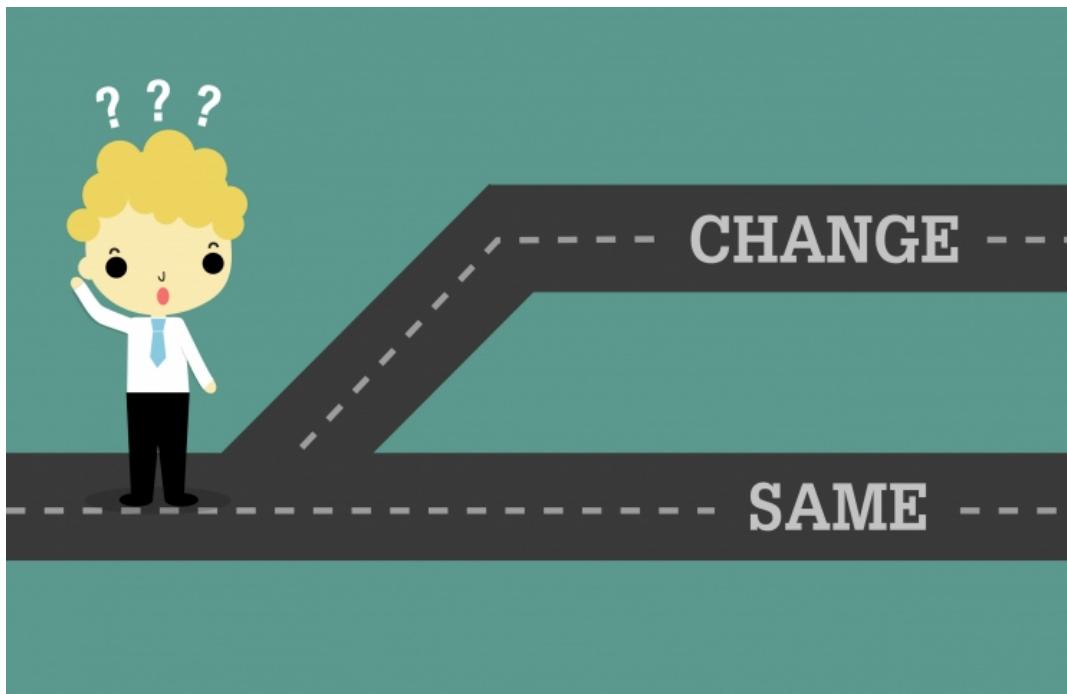
Behavioral science is the study of human behavior across the lifespan with an emphasis on how behavior is heavily influenced by factors beyond the individual



“Behavior” is actually a complex set of actions.

What is behavioral design?

The intentional design of products and services that will influence humans to change their behavior



The early years...
MA Experimental Psychology

What predicts your relationship with exercise?

Running head: HOPE THEORY AND PREDICTING EXERCISE

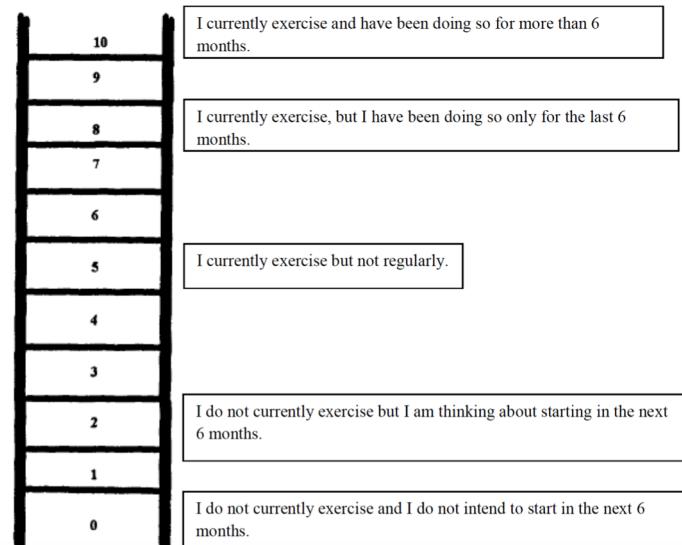
1

Hope and Exercise: The Integration of Hope Theory
with the Transtheoretical Model

Gina Merchant

California State University San Marcos

Figure 1 – Stages of Change Ladder



- Predict relationship between the observable & unobservable
- Compare measurement tools

The early years



Marie Thomas, Ph.D. (1981, Fordham University)

Dr. Thomas was on the faculty at the College of Mount St. Vincent (New York) and had worked as a Personnel Research Psychologist at the Navy Personnel Research and Development Center in San Diego. She joined CSUSM in 1995 and served as Psychology Department Chair from 2000-2003 and 2004-2005. Her areas of expertise are psychological testing, methodology, and statistics. She is currently working on a project with a colleague in the Economics Department to study the subjective well-being of college students. She is also interested in mindfulness, positive psychology, gender issues, and the scholarship of teaching.



P. Wesley Schultz, Ph.D. (1995, Claremont Graduate University)

Dr. Schultz came to CSUSM in 1997 after two years as a Visiting Assistant Professor at St. Lawrence University. His areas of expertise are social psychology and statistics. His current research interests are in the application of social psychology to help understand and solve social problems. Recent projects include studies on environmental programs (energy conservation, water conservation, green marketing, recycling), cross-cultural research on environmental attitudes, and longitudinal research on programs aimed at promoting research careers among underrepresented students. Dr. Schultz is currently serving as Interim Dean of Graduate Studies and Associate Vice President for Research.

The early years...
PhD Program

The early years



SAN DIEGO STATE
UNIVERSITY

Convergent Validity of Self-Reported Sedentary Behavior and Accelerometry: Comparing Multiple Cut Points

Gina Merchant,^{1,2} Humberto Parada,¹ & Simon J. Marshall²

¹San Diego State University, San Diego, CA; ²University of California, San Diego, La Jolla, CA

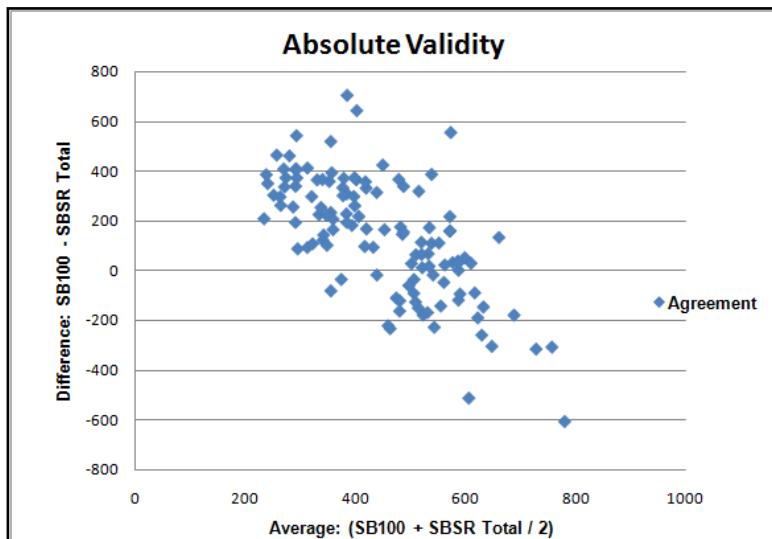


Figure 1. Bland-Altman plot comparing accelerometer (cut point=100) and self-reported SB (Total SBSR). Accelerometer data adjusted for wear time.

3. Know your measurement
4. Your data are meaningless & dangerous without good measurement

Why use accelerometers?



“When you want to know the truth about someone, that someone is probably the last person you should ask”

G. House

Accelerometer data decisions

“Small inconsistencies can have a substantial impact on outcome variables.”

(Colley et al., 2010)

1. Frequency extension

- Normal vs. low-frequency extension

2. “Allowable interruption period”

- 60 or 90 consecutive zeros considered non-wear?

3. Epoch: it depends...

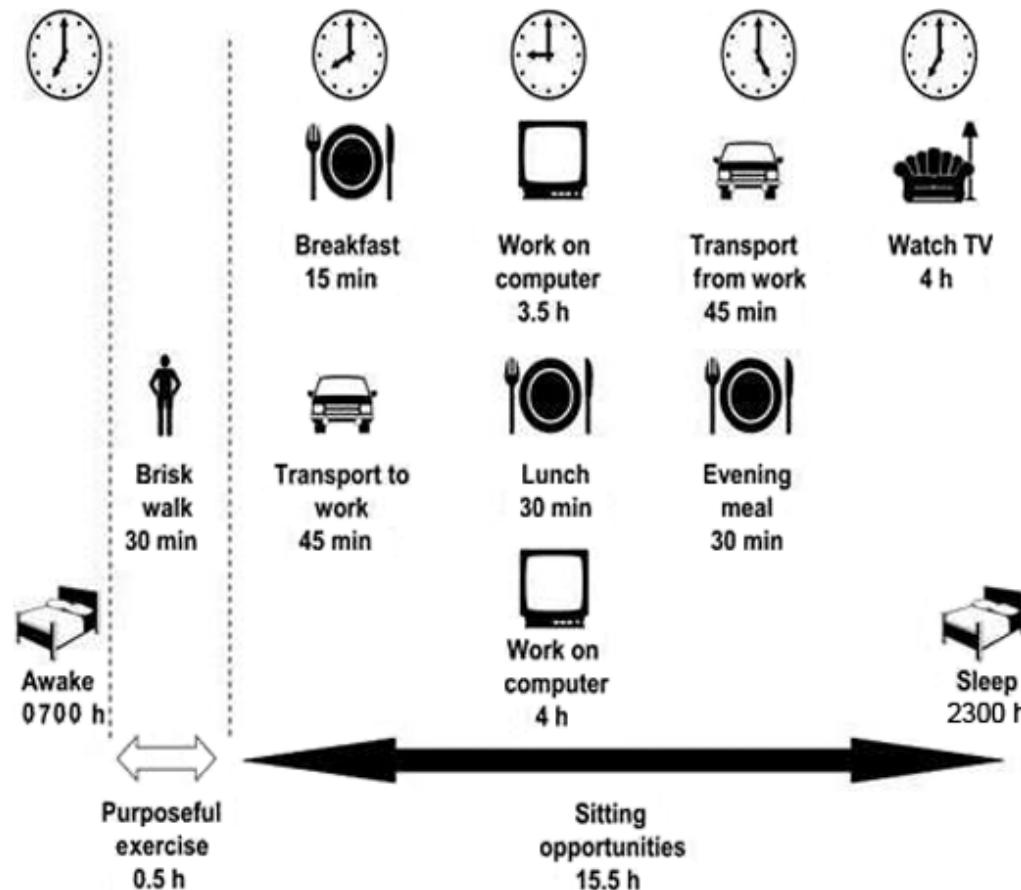
- GT3X+ collects raw acceleration data & epochs can be set during processing

4. Device wear time

- There are within-day & across day considerations!

Wear Time

A waking day = ~ 15.45 hours



(Tremblay et al., 2011)

Epistemology

How do you know what you know?

“Like most people, epistemologists often begin their speculations with the assumption that they have a great deal of knowledge. As they reflect upon what they presumably know, however, they discover that it is much less secure than they realized, and indeed **they come to think that many of what had been their firmest beliefs are dubious or even false.**”

<https://www.britannica.com/topic/epistemology>

The middle years...
PhD Dissertation

Data source: SMART Intervention

1) Facebook



2) Apps



3) Website



4) Texts



5) Email



6) Health coach



To what extent does social influence on health extend to online environments?

- Rationale for investigating:
 - 1.5 billion use Facebook each month
 - 4 billion pieces of content shared each day
 - Dynamic activity (sharing, feedback, exposure)

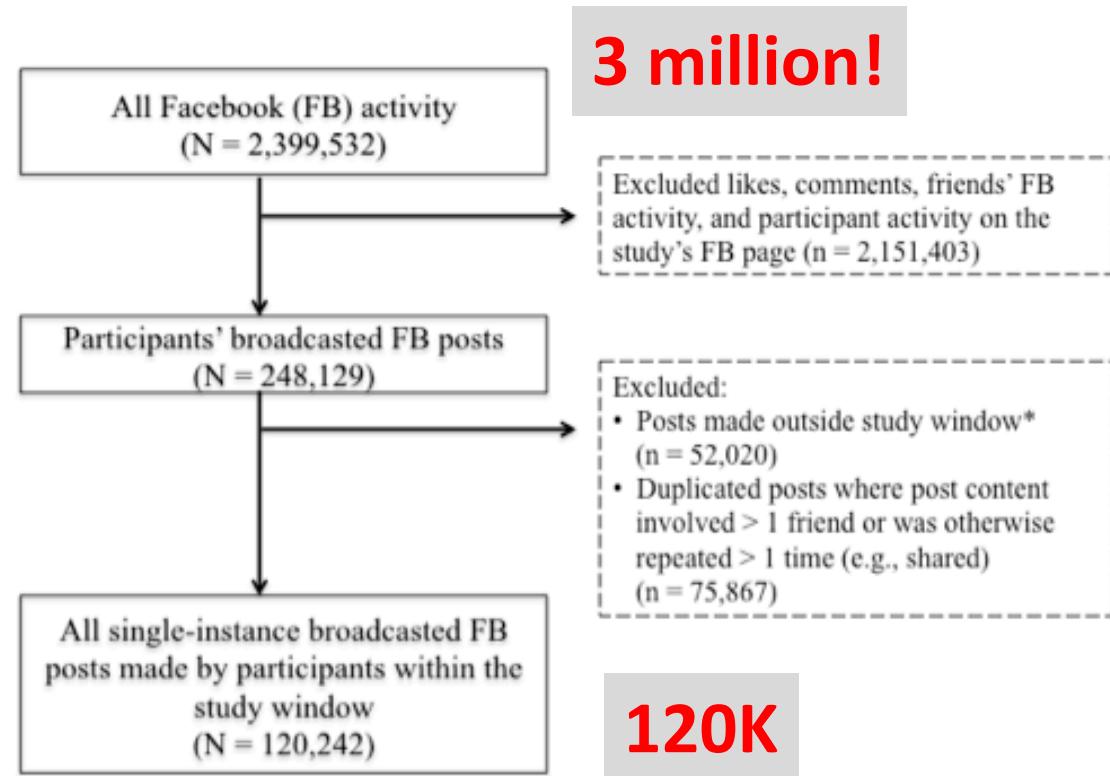
“I mean, we all do it for the likes.”
- *Project SMART study participant*



Most people don't realize this, but you can run without telling Facebook about it.



Figure 1. Processing of Facebook data for HAL dependent variable

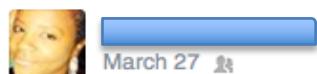


*Study window was defined as posts made six months prior to study entry (i.e., baseline) up to +26 months (i.e., study exit). All post data were anchored to participants' study start date.

Data processing

- Text as data
- Digital engagement
- Social network data

Exercise words (N = 165)					
10k	cheerleading	horsebackriding	raced	soulcycle	workouts
5k	climb	intramural	races	spartan	wrestling
abs	climbing	jærcise	racing	spin	yoga
active	coate	jetski	racquetball	squash	zumba
aerobic	cricket	jetskiing	racquetballs	squats	
anaerobic	crossfit	jog	rafting	stairmaster	
backpacking	cushions	jogging	ran	stain	
badminton	duathalon	kate	recipes	surf	
ballet	elliptical	kayaking	rep	surfboard	
baseball	endurance	kettlebell	rockclimbing	surfboards	
baseballs	ergometer	kettlebells	rollerblading	surfed	
basketball	exercises	kickball	rowing	surfing	
basketballs	exercises	kitesurf	rugby	swan	
biathlon	fencing	kitesurfing	run	swim	
bicep	fitness	lacrosse	runner	swimming	
bicycle	football	lbs	running	taekwando	
bicycled	footballs	hunge	strip	toughmudder	
bicycles	frisbee	hungen	situps	trainer	
bicycling	frisbees	mmader	skateboarding	training	
bike	golf	paddleball	skating	treadmill	
biked	gym	paddleboarding	ski	triathlon	
bikes	gymnasium	pedometer	skiied	ups	
biking	gymnastics	palates	skiing	volleyball	
bodyboarded	gymnastics	plank	skijump	walk	
bodyboarding	gymnastics	planks	skijumping	walked	
bodyboarding	gymnastics	hackysac	skis	walking	
bodyboards	gymnastics	plyos	snowshoe	waterpolo	
bootcamp	handball	pullup	snowshoeing	weights	
bootcamps	healthy	pullups	snowshoes	windsurf	
boxing	hike	pushup	soccer	windsurfer	
cardio	hiking	pushups	softball	windsurfing	
cardiovascular	hockey	race	softballs	workout	



Super excited right now. I came, I saw, I conquered running stocker trail non-stop and ran 1/2 way up valley ridge hill! I'm kinda of a big deal!!! — 😍 feeling excited.

Unlike · Comment · Share

41 10



finishes here first 5k! — with and



Like · Comment · Share

and 27 others like this.



“Big” & messy data

- Saying goodbye to SPSS...
- Hello R!
- Rrrr Club



Suneeta Godbole - Statistician and Data Manager

Suneeta received her MPH with a concentration in Epidemiology and Biostatistics in 2008. She has been working with CWPBS since 2009 and started working for REACH in 2011. On most days you'll find her happily ensconced at her standing desk pouring over data files as she manages and analyzes data for multiple REACH studies and wrangles time variables into alignment. In her free time, she's a keen reader of current events and has started knitting for family and friends. She dreams of a time when people don't have to exercise to be healthy.

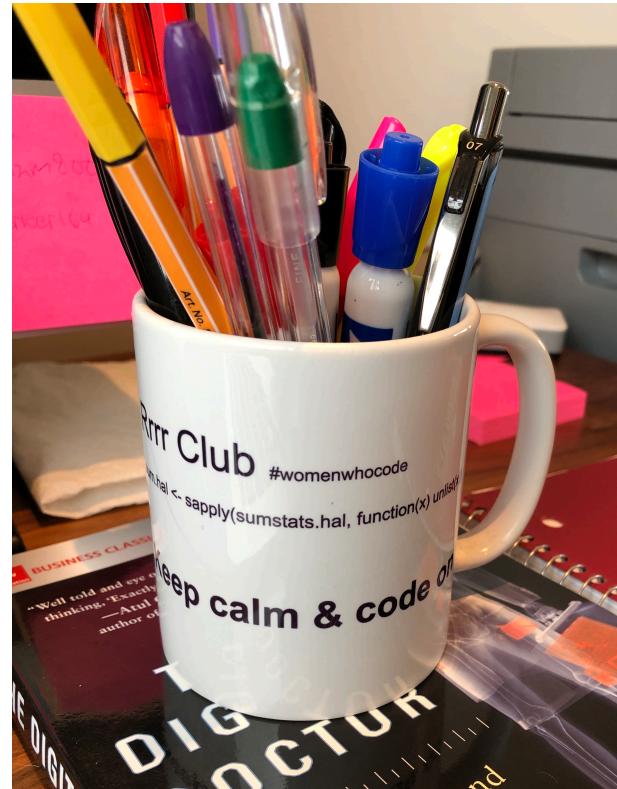
Purpose

- To improve our individual R coding skills by leveraging the group's knowledge.
- To work together to discover solutions that will benefit the group.
- To identify new packages and functions that will increase our coding efficiency.

Rrrrr Club



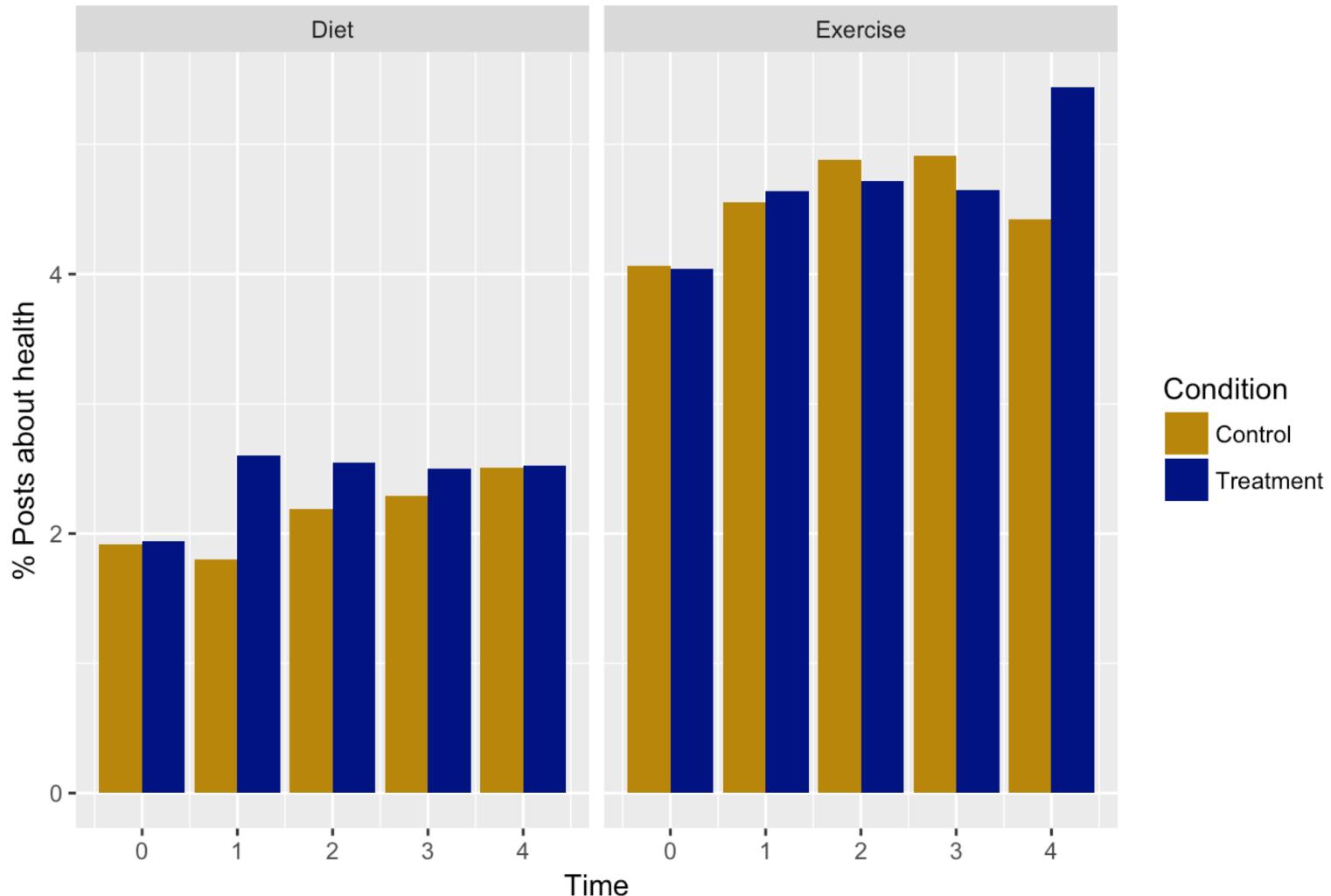
```
sum.hal <- sapply(sumstats.hal,  
function(x) unlist(x))
```



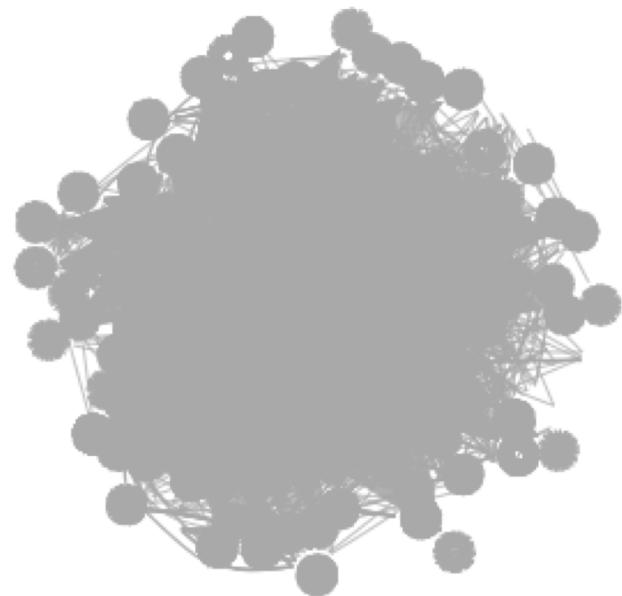
5. Love data unapologetically

Text as data

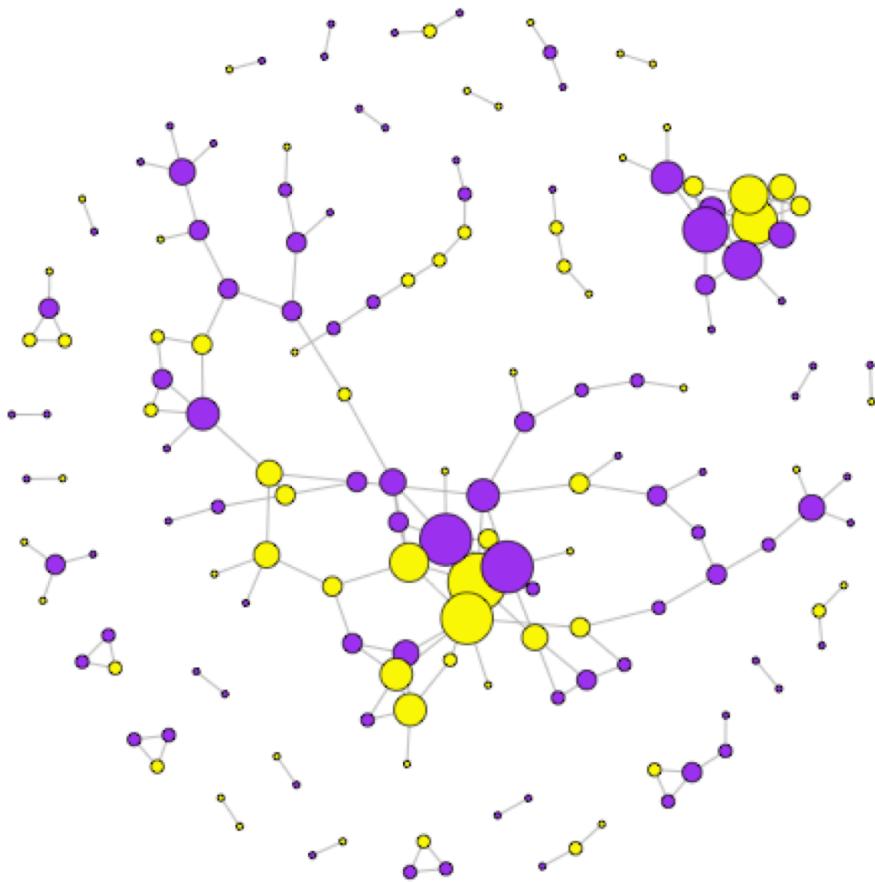
Posts about Health by Condition



Social networks take 1



Social networks take 2



Network Picture by Condition
Yellow = Control | Purple = Intervention

- **Degree**
 - 478 ± 308
- **In-study degree**
 - 1.33 ± 1.83
- **56% have at least 1 study friend**
 - 4 ± 2.7 (among those w/ ≥ 1 friend)
- **Interactions between study friends**
 - 55.31 ± 169
 - Median = 11
 - Range: 1 to 1582
- **Closeness**
 - 73 (35%) strong ties
 - 137 (65%) weak ties
- **Interactions among strong ties**
 - 146 ± 264

Results: ~~Enhanced tx effect~~ (selection effect)

- Having ***at least 1 study friend*** is associated with weight loss at 6 months from baseline
Beta = - 1.28 (p = 0.03)
- Having ***a strong (vs. weak) tie*** is associated with ego weight loss at 6 months
Beta = - 0.76 (p = 0.03)

If you like my run-brag post do I lose weight?

Table 3. Coefficients for the linear mixed model testing for association between receiving social support for talking about HAL on Facebook and change in weight (kg) by sex in the treatment group

	Female treatment group			Male treatment group		
	Beta (CI)	P		Beta (CI)	P	
Intercept	68.14	(53.92, 82.35)	0.00	86.31	(70.82, 101.79)	0.00
T2	0.12	(-1.02, 1.26)	0.84	-1.98	(-3.97, 0.00)	0.05
T3	-0.17	(-1.31, 0.96)	0.76	-2.88	(-4.86, -0.89)	0.01
T4	0.59	(-0.56, 1.74)	0.31	-0.87	(-3.17, 1.43)	0.46
T5	1.35	(0.13, 2.57)	0.03	-0.44	(-2.41, 1.54)	0.66
% HAL SS	0.16	(0.00, 0.32)	0.05	0.04	(-0.15, 0.23)	0.70
T2 * % HALss	-0.20	(-0.37, -0.04)	0.02*	0.01	(-0.20, 0.22)	0.92
T3 * % HALss	-0.13	(-0.28, 0.03)	0.12	0.07	(-0.14, 0.28)	0.51
T4 * % HALss	-0.13	(-0.30, 0.03)	0.11	-0.08	(-0.33, 0.17)	0.55
T5 * % HALss	-0.11	(-0.28, 0.07)	0.22	-0.03	(-0.22, 0.15)	0.72
% HAL	-0.07	(-0.21, 0.08)	0.36	-0.06	(-0.22, 0.11)	0.50
Age	0.36	(-0.28, 1.00)	0.27	0.22	(-0.42, 0.85)	0.51

The reference categories is baseline/T1 (for time)

For every 20% increase in social support on HAL posts, females in the treatment group lost 9 lbs from baseline

Feelings about sharing on FB

Against sharing

*“It’s more like, the fear that it won’t work out. And also, I just don’t like telling people I’m trying to lose weight. **I feel like it affects their perception of me.**”*

Uncertain/ambivalent about sharing

*“I have recently **since I’ve started losing weight to try to like, share what I’ve been doing.** But I’m only doing that because I’ve been successful.”*

Shares regularly

*“I mean, we all do it for the likes... **People will be like, “You go girl!” or something because if they previously saw that you didn’t really do that, but now you do, they’re like, “That’s cool.”**”*

Facebook lurking

“I think I, I look at 90% of stuff and don’t comment on it... Unless I really, you know, am moved to comment. And sometimes I’ll even write a comment on people’s stuff and then delete it. Just like, “Oh my God, I don’t want to be part of that conversation.”

6. Qualitative data is as important as quantitative data

The post-doc

- Spread of online misinformation (anti-vax)
- Dabbled in machine learning
- Decided to leave
 - Read more on that decision [here](#)

7. Don't be afraid to make a big life change

Now...

Getting paid to be a behavioral scientist!

How'd I get my current job?

Twitter.

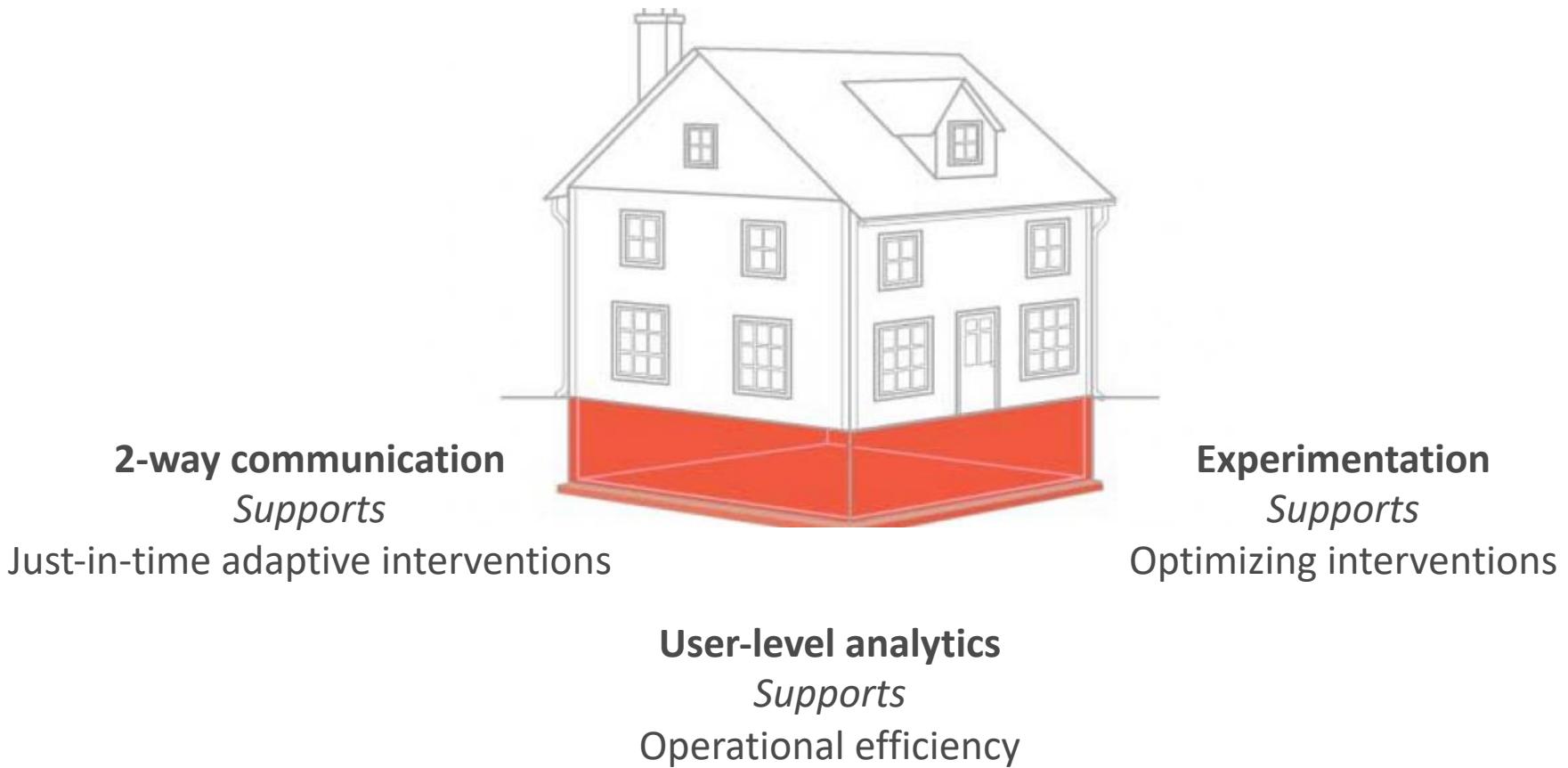


What do I do now?

1. Build behavioral science capacity & competency at our company
2. Design experiments
3. Support the business in applied machine learning
4. Minimal data processing/visualizing/modeling

Designing for behavior change

Personalized patient experience



What tools do I use?

- RMarkdown
- Tidyverse:
 - dplyr
 - ggplot

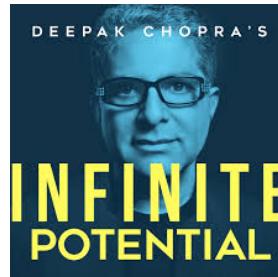


Listen

- Hidden Brain podcast
 - One Head, Two Brains



- Infinite Potential podcast
 - Part Two: Minds and Machines/Christopher Wylie



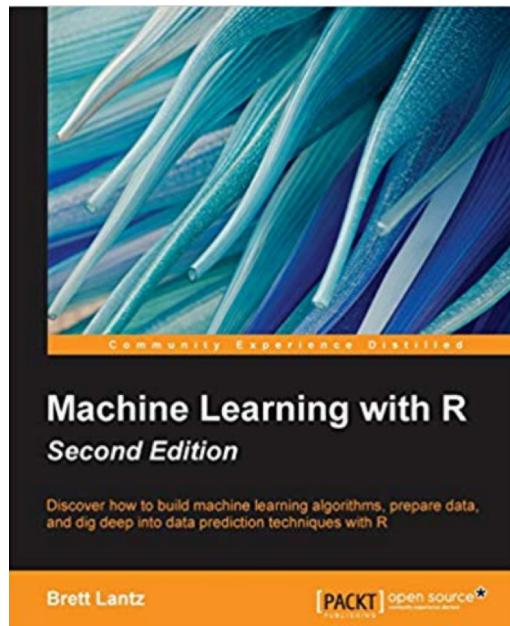
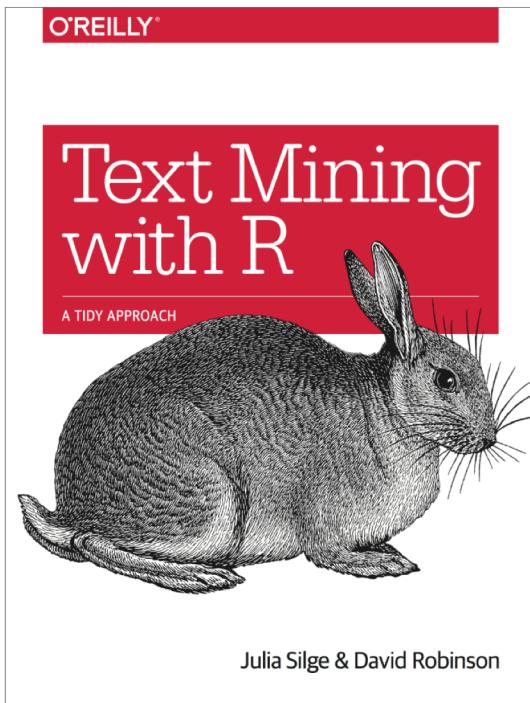
Follow, join

- R Ladies
- R Forwards
- R Bloggers

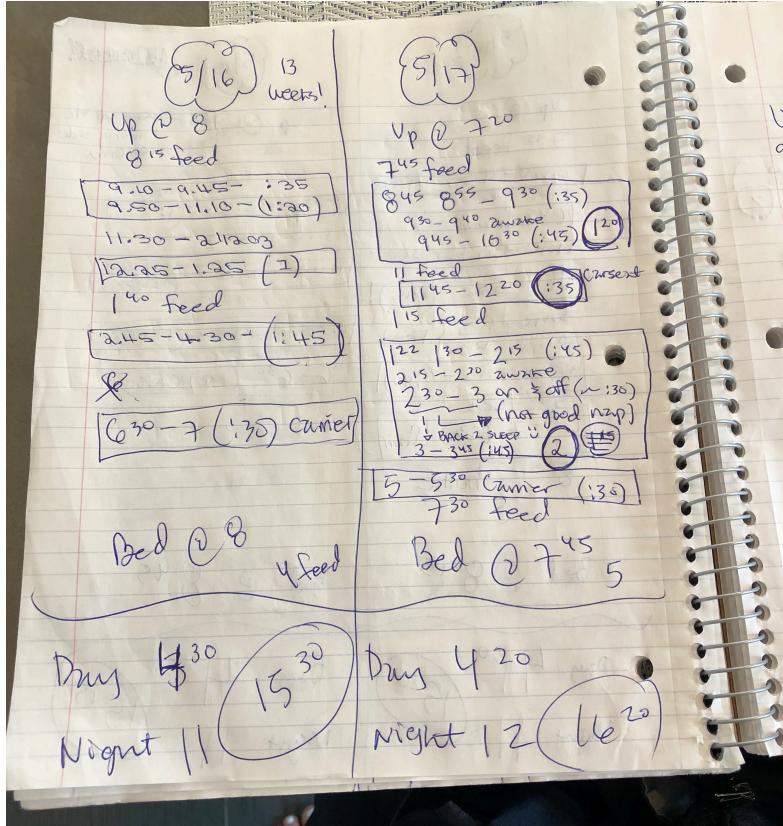
People

- **Mara Averick** @dataandme
- **Julia Silge** @juliasilge
- **David Robinson** @drob
- **Alice Data** @alice_data
- **Sharla Gelfand** @sharlagelfand
- **Hadley Wickham** @hadleywickham
- **Charlotte Wickham** @CVWickham
- **Jenny Bryan** @JennyBryan
- **Daniela Vazquez** @d4tagirl
- **Gabriela de Queiroz** @gdequieroz

Read



9. Read



10. Sleep is important

Thank you!

Gina Merchant

@DrGMerchant

gina.merchant@resmed.com

