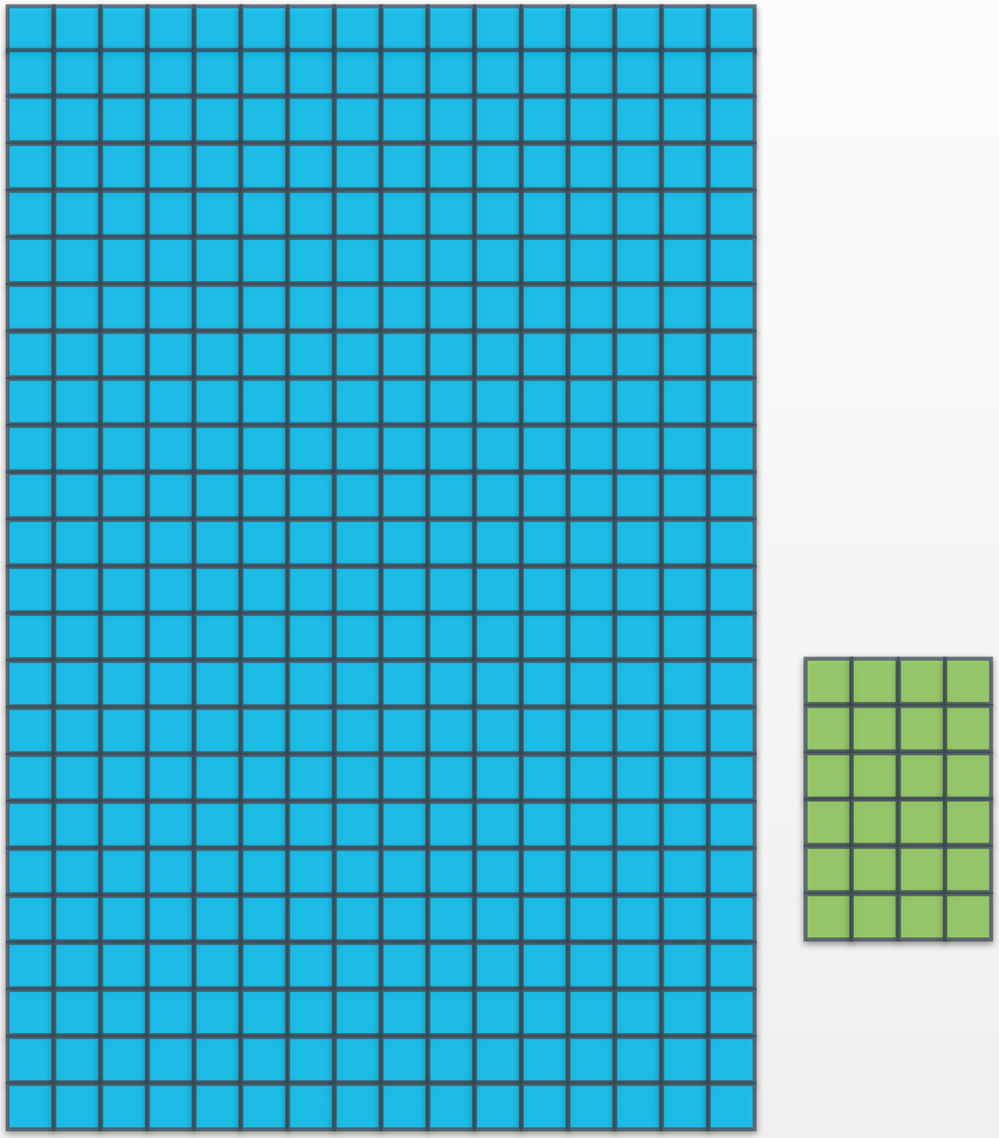
THE SIDEKICK PATTERN:
USING SMALL DATA TO MULTIPLY THE VALUE OF BIG DATA

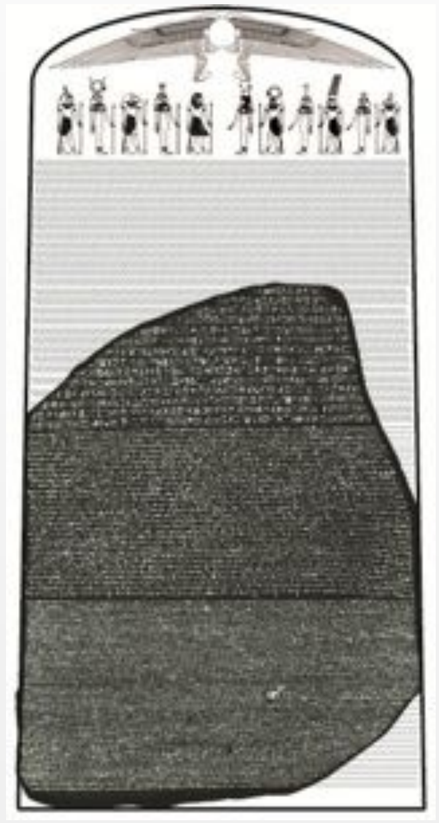
@AbeGong
Data Scientist, Jawbone
Strata - February 2014

Friday, February 14, 14

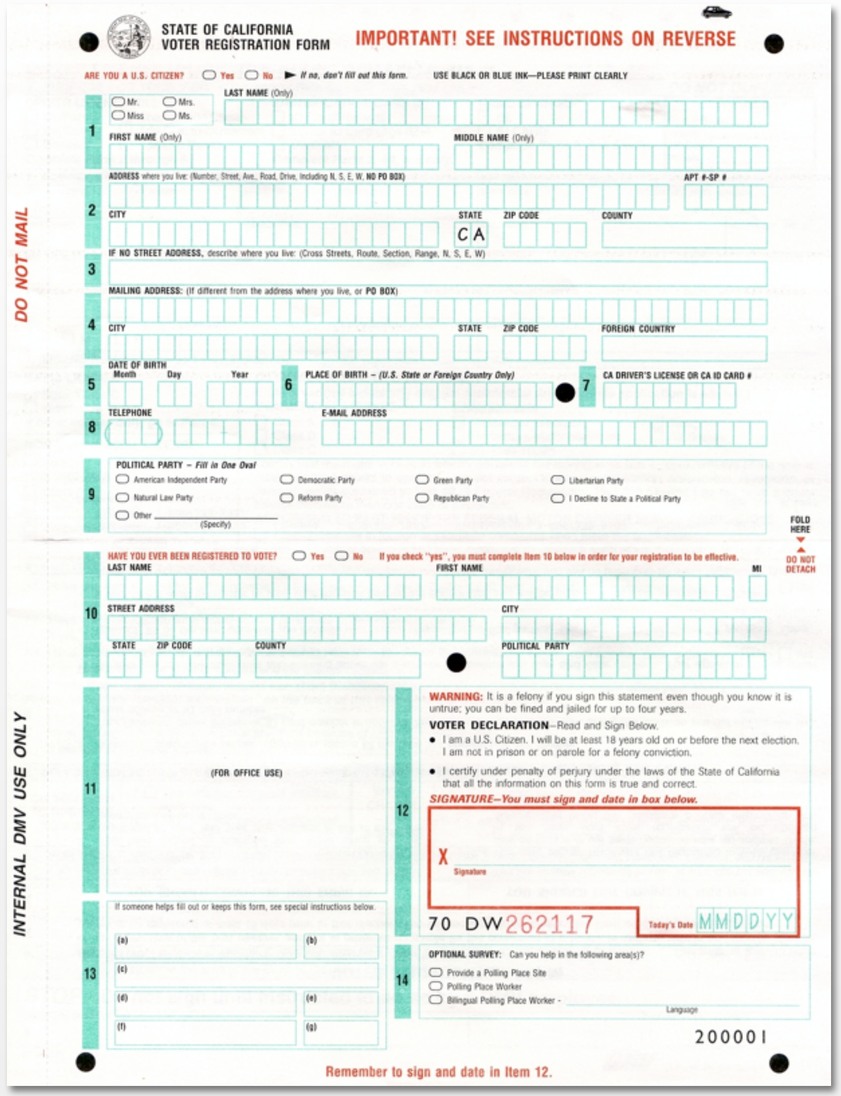
DATA SIDEKICKS



EX: HIEROGLYPH TRANSLATION



EX: CAMPAIGN TARGETING



STATE OF CALIFORNIA
VOTER REGISTRATION FORM

IMPORTANT! SEE INSTRUCTIONS ON REVERSE

ARE YOU A U.S. CITIZEN? ○ Yes ○ No ► If no, don't fill out this form. USE BLACK OR BLUE INK—PLEASE PRINT CLEARLY

1 ○ Mr. ○ Mrs. ○ Miss ○ Ms. LAST NAME (Only) FIRST NAME (Only) MIDDLE NAME (Only)

2 ADDRESS where you live: (Number, Street, Ave., Road, Drive, Including N, S, E, W, NO PO BOX) APT #-SP # CITY STATE CA ZIP CODE COUNTY

3 IF NO STREET ADDRESS, describe where you live: (Cross Streets, Route, Section, Range, N, S, E, W)

4 MAILING ADDRESS: (If different from the address where you live, or PO BOX) CITY STATE ZIP CODE FOREIGN COUNTRY

5 DATE OF BIRTH Month Day Year 6 PLACE OF BIRTH – (U.S. State or Foreign Country Only) 7 CA DRIVER'S LICENSE OR CA ID CARD #

8 TELEPHONE E-MAIL ADDRESS

9 POLITICAL PARTY – Fill in One Oval
○ American Independent Party ○ Democratic Party ○ Green Party ○ Libertarian Party
○ Natural Law Party ○ Reform Party ○ Republican Party ○ I Decline to State a Political Party
○ Other _____ (Specify)

FOLD HERE ▼ DO NOT DETACH

10 HAVE YOU EVER BEEN REGISTERED TO VOTE? ○ Yes ○ No If you check "yes", you must complete Item 10 below in order for your registration to be effective.
LAST NAME FIRST NAME MI
STREET ADDRESS CITY
STATE ZIP CODE COUNTY POLITICAL PARTY

11 (FOR OFFICE USE)

12 WARNING: It is a felony if you sign this statement even though you know it is untrue; you can be fined and jailed for up to four years.
VOTER DECLARATION—Read and Sign Below.
- I am a U.S. Citizen. I will be at least 18 years old on or before the next election. I am not in prison or on parole for a felony conviction.
- I certify under penalty of perjury under the laws of the State of California that all the information on this form is true and correct.

SIGNATURE—You must sign and date in box below.
X Signature
70 DW262117 Today's Date MM D D Y Y

13 If someone helps fill out or keeps this form, see special instructions below.
(a) (b) (c) (d) (e) (f) (g)

14 OPTIONAL SURVEY: Can you help in the following area(s)?
○ Provide a Polling Place Site
○ Polling Place Worker
○ Bilingual Polling Place Worker - _____ Language

200001

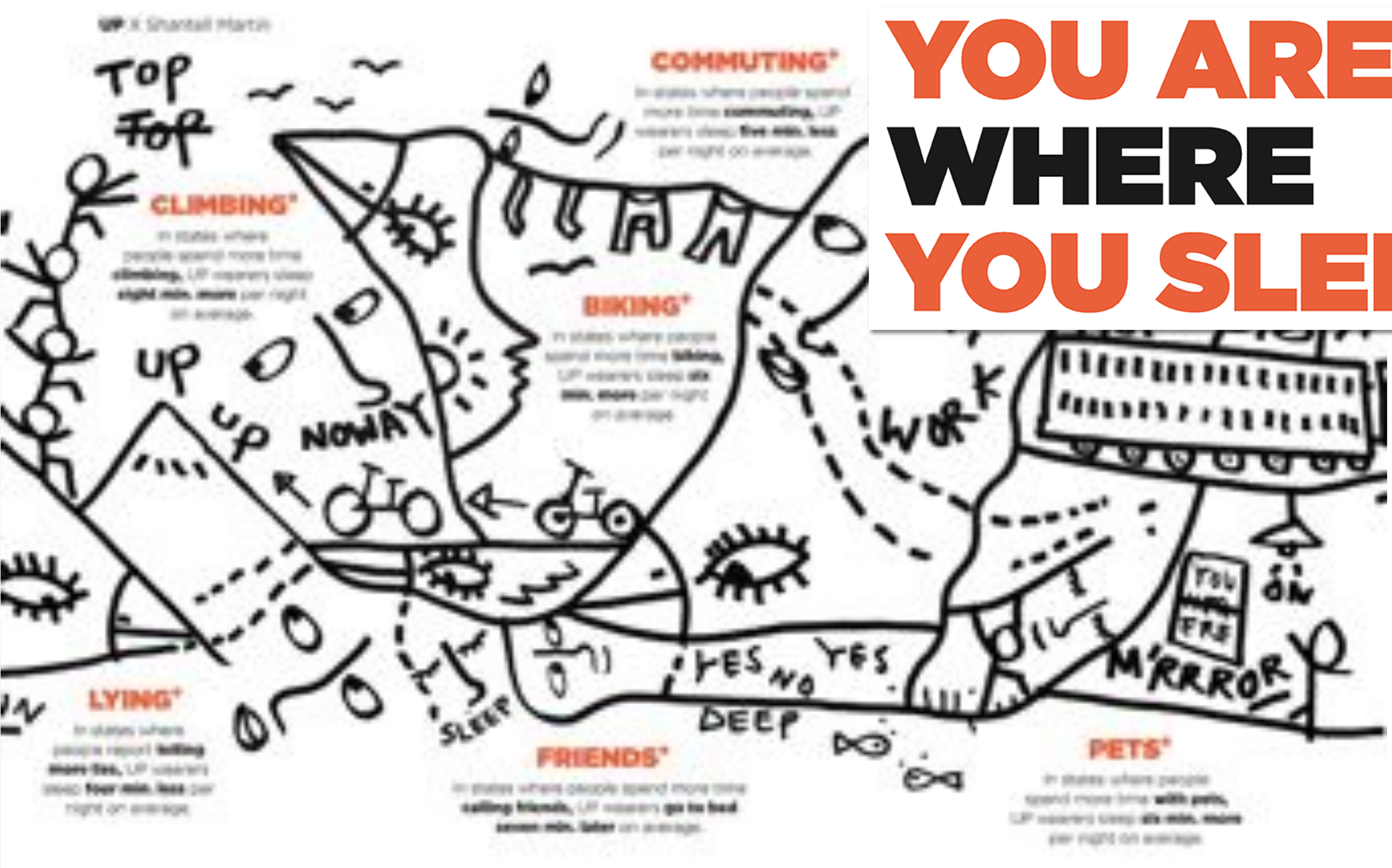
Remember to sign and date in Item 12.

DO NOT MAIL

INTERNAL DMV USE ONLY

EX: SLEEP CONTEXT

YOU ARE WHERE YOU SLEEP.



CLIMBING*

In states where people spend more time **climbing**, UP wearers sleep **eight min. more** per night on average.

COMMUTING*

In states where people spend more time **commuting**, UP wearers sleep **five min. less** per night on average.

BIKING*

In states where people spend more time **biking**, UP wearers sleep **six min. more** per night on average.

LYING*

In states where people report **telling more lies**, UP wearers sleep **four min. less** per night on average.

FRIENDS*

In states where people spend more time **calling friends**, UP wearers **go to bed seven min. later** on average.

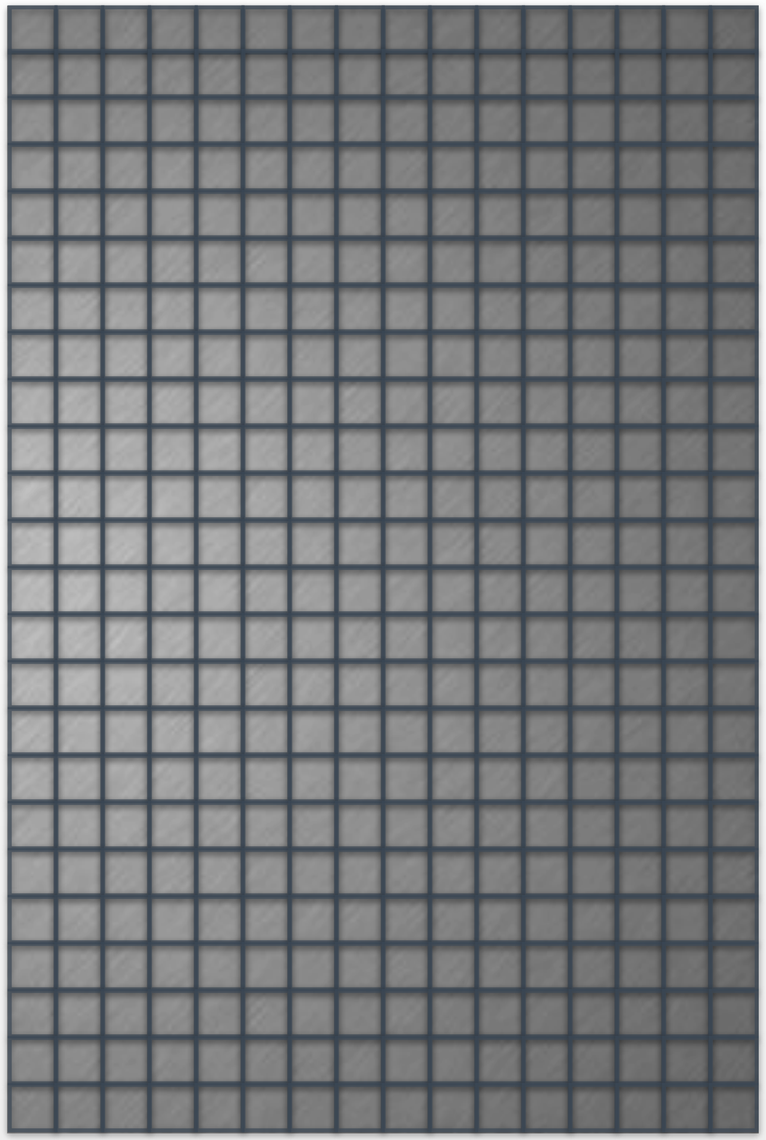
PETS*

In states where people spend more time **with pets**, UP wearers sleep **six min. more** per night on average.

EXAMPLES, PLEASE:
WHICH DATA STREAMS GET BIG?

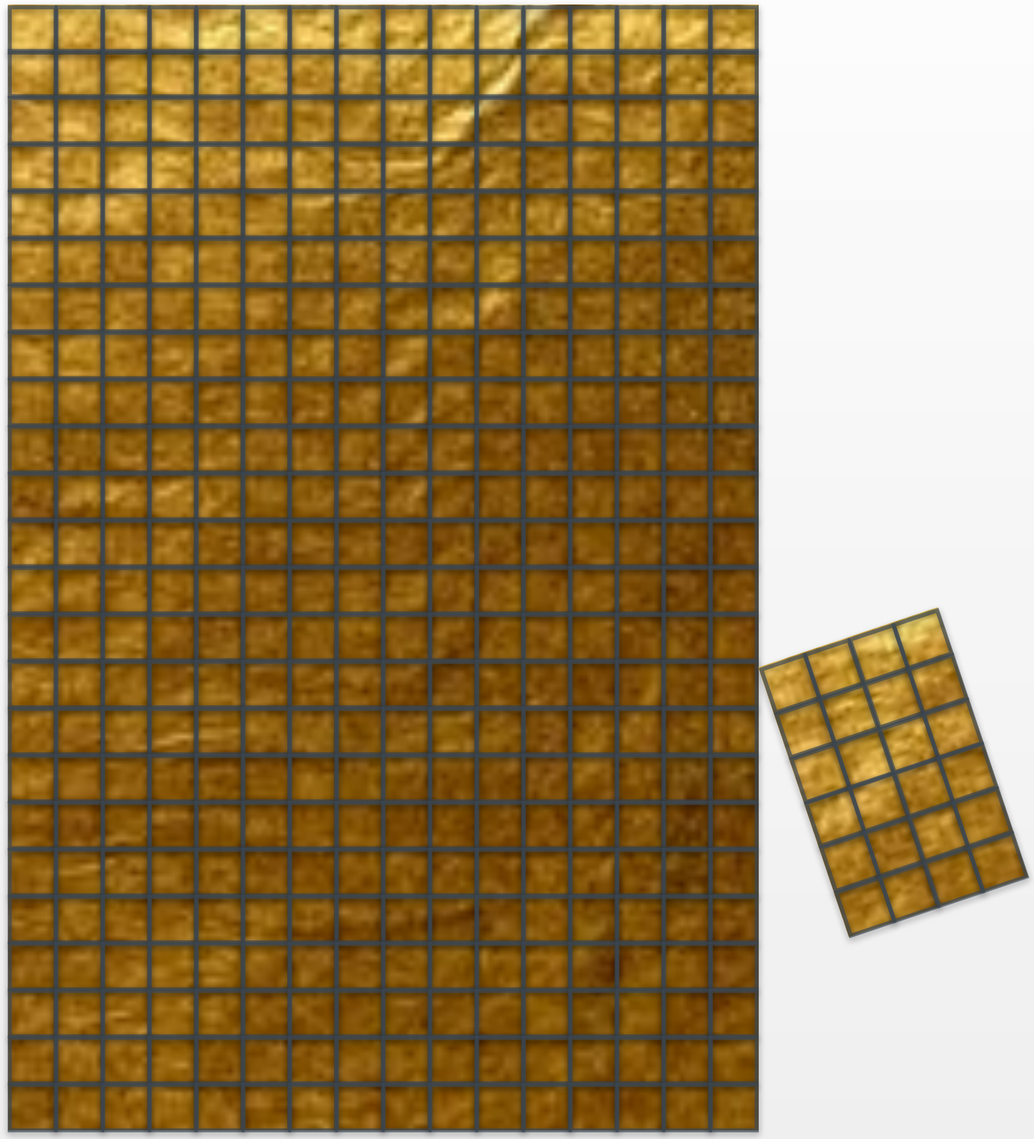
(...AND BESIDES SIZE, WHAT ELSE DO THEY HAVE IN COMMON?)

BIG, RICH, MESSY



CAREFULLY CURATED

TRANSMUTATION!



EX: HUFFPO MODERATION

Michael McAuliff
mike.mcauliff@huffingtonpost.com

Sabrina Siddiqui
sabrina.siddiqui@huffingtonpost.com

Senate Passes Increase Of The Nation's Debt Limit

178 16 1 193

MORE: Congress Debt Limit, Senate Passes Debt Limit, Video, Senate Debt Ceiling, Debt Ceiling, Senate Debt Limit, Debt Limit, Harry Reid , Harry Reid Debt Ceiling, Politics News

WASHINGTON -- The Senate moved quickly Wednesday to boost the nation's debt

Invest in the future of Greece.
Let's tell a new story.
onegreece.org
A campaign by The Hellenic Initiative

MOST POPULAR

This Governor Looks Set To Reject Overdose Prevention For No Good Reason

PHOTOS: Michelle Obama Stuns At State Dinner In A Powder Blue And Black Carolina Herrera

This Guy Learned What Happens When

Headly Lamarr (headly67) 31
SUPER USER · 1,696 Fans · We'll raise my rent

How embarrassing for Rafael Cruz, he thought he actually had some weight in the Senate only to again fail miserably in his juvenile "look at me" foot stomping act.

The Tea Party is over, America watched your clown act long enough and they have spoken.
12 FEB 12:31 PM

REPLY · FAVE · SHARE · MORE

Andy Ballantine (aballantine1) 2
SUPER USER · 230 Fans · Peepeful faction through violent tree-hugging!

It was his act that got it to happen so quickly. He is the new hate subject in the GOP. What he wants will be obstructed from now on.
12 FEB 12:47 PM

REPLY · FAVE · SHARE · MORE

8 PEOPLE IN THE CONVERSATION Read Conversation →

Stacey B. (bernikitty) 11
1,046 Fans · single mom RN, I hate facebook

good, now can we turn to dealing with real problem that exist instead of creating new ones.

this was an expensive lesson for the republicans to learn, but i guess we should be glad the poor simple dears finally learned it.
12 FEB 12:31 PM

REPLY · FAVE · SHARE · MORE

Thomas B. (qualitylumber) 1
194 Fans · I'm against bad policy, not the people writing it.

Rates Rise and Other Fascinating News on the Web

Valentines for People Who Spend Way Too Much Time on the Internet

Meet Nasar, The Horse Who Lives In a House

Time - Politics

THE HUFFPOST/YOUGOV POLL

Do you favor or oppose the federal government more strictly regulating emissions of carbon dioxide and other greenhouse gas emissions?

Strongly favor · Favor · Oppose · Strongly oppose

Join YouGov

Sponsored Links

LifeLock® Services
24/7 Credit Fraud Monitoring Plan. Proactive Credit Fraud Protection.
LifeLock.com

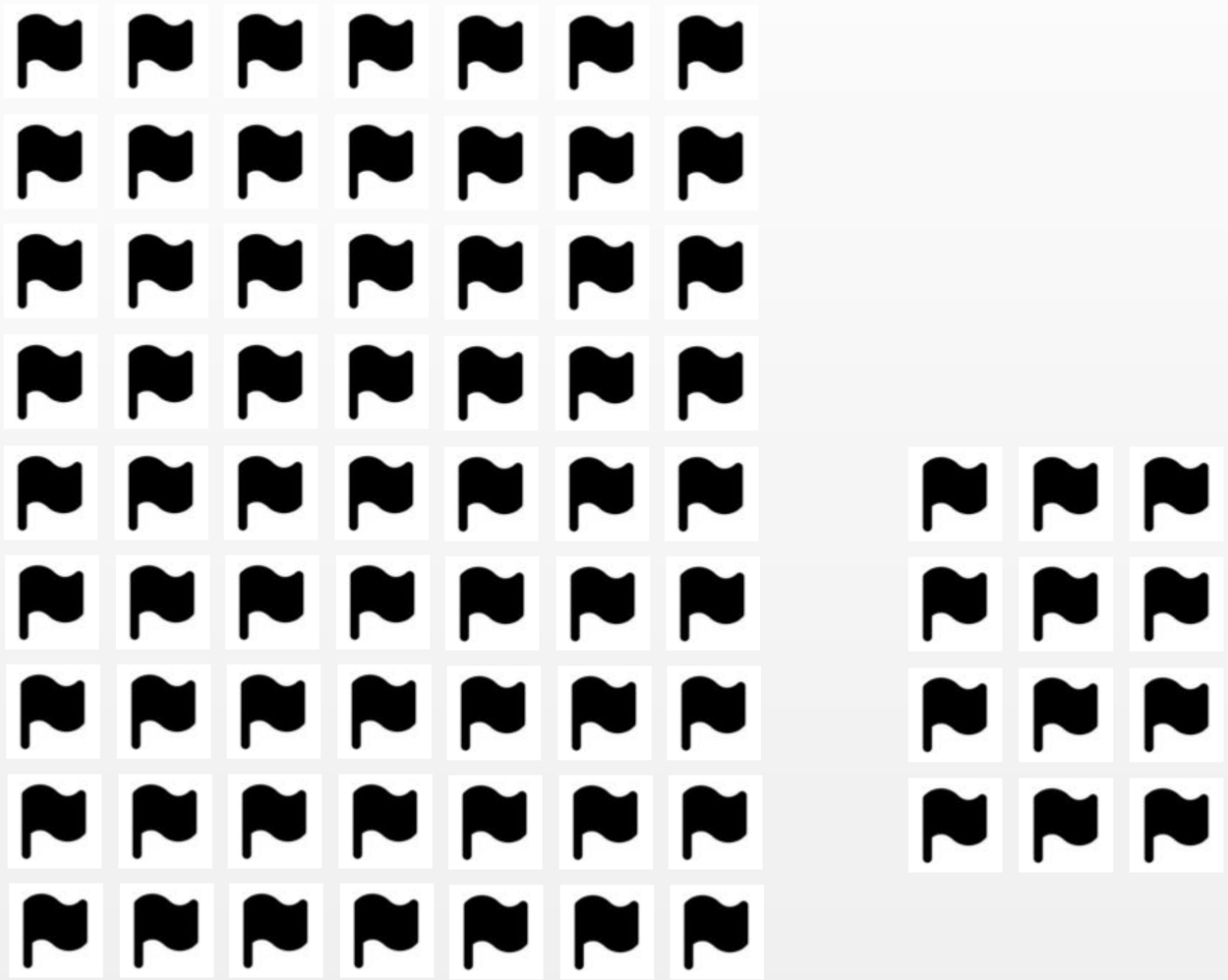
Assistance for Health Coverage
Millions of Californians can now get financial help with their insurance bills.
www.coveredca.com

53yr old Mom Looks 27
Mom releases simple facelift secret that...
http://officialtips.net

Buy a link here

EX: HUFFPO MODERATION

EX: HUFFPO MODERATION

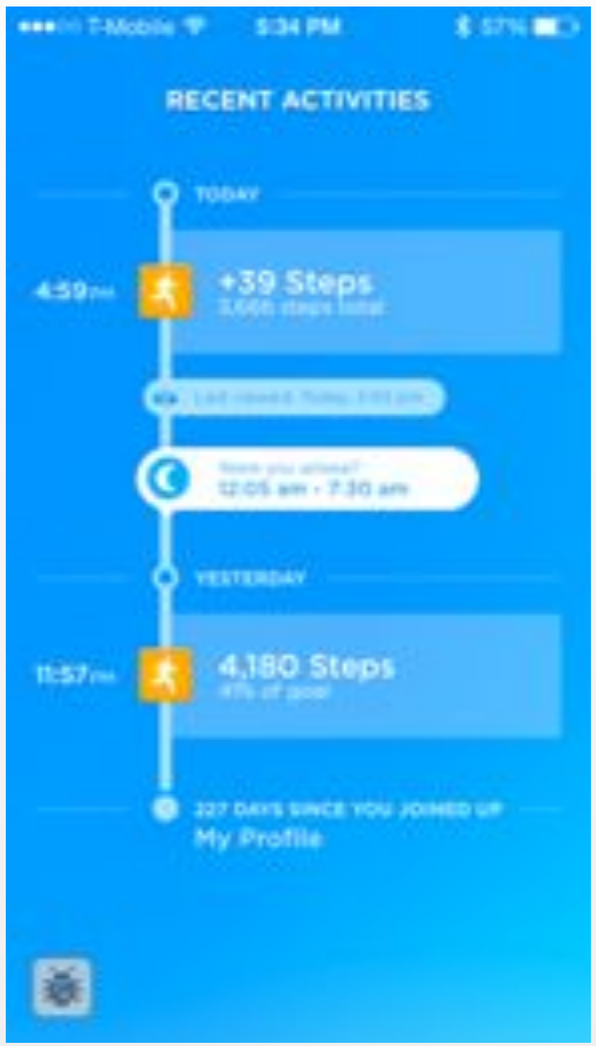


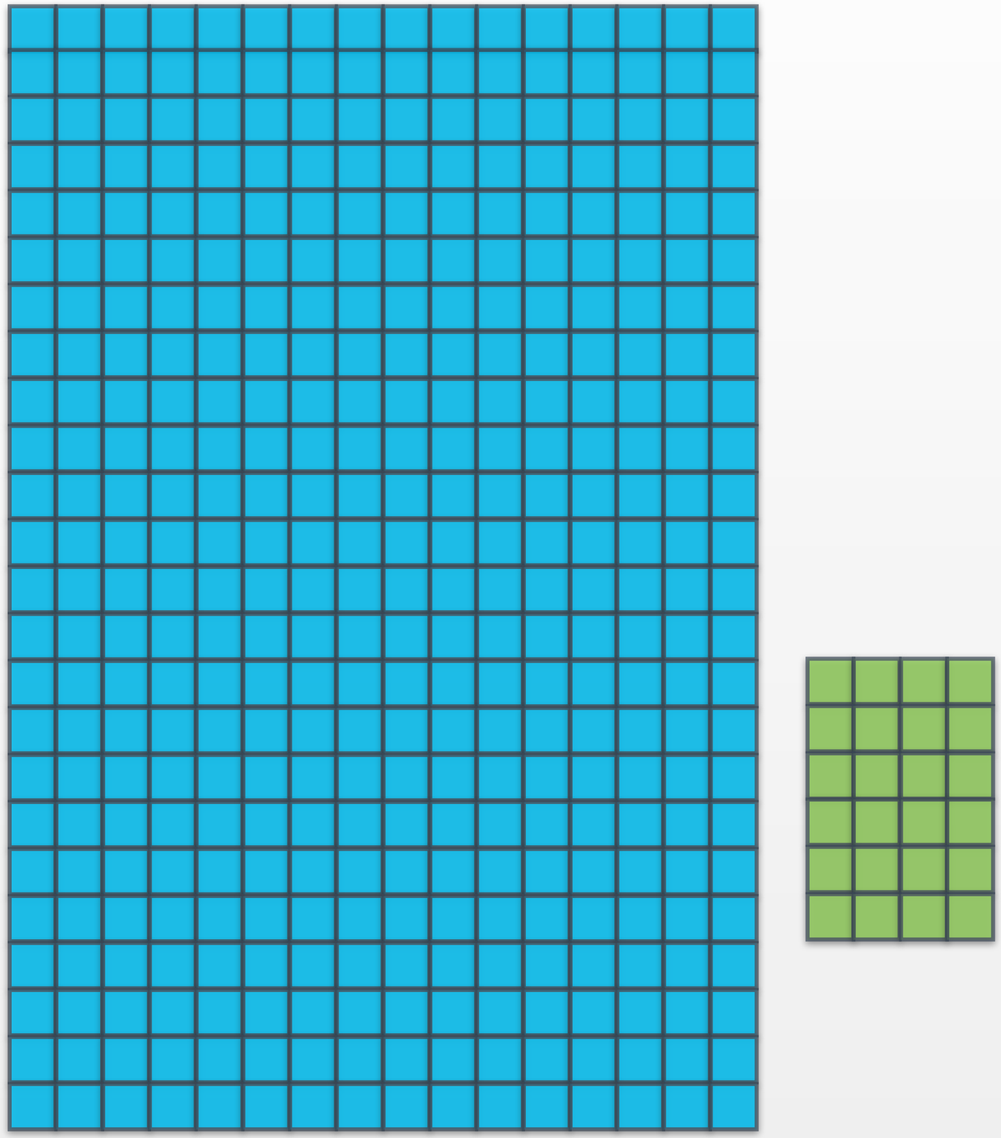
WHEN SHOULD I USE THE SIDEKICK PATTERN?

- To separate munging and cleaning from scaling.
- To bootstrap new data products.
- To leverage variety against volume.

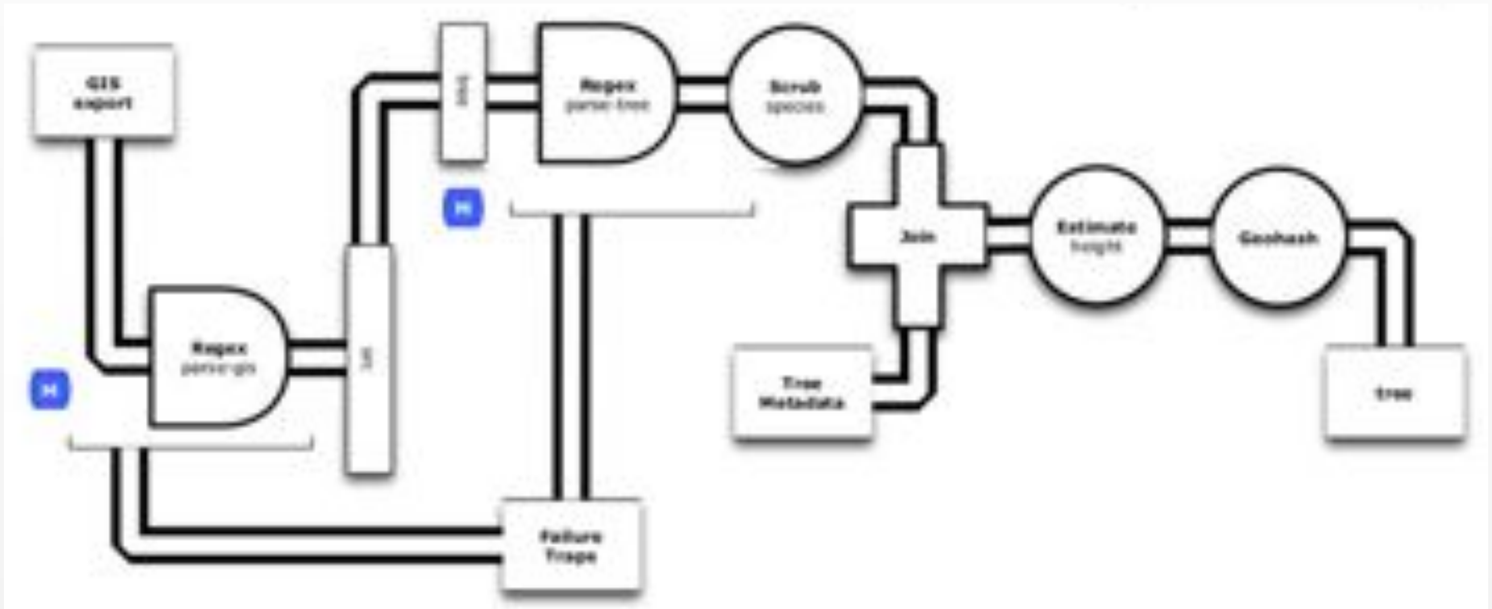
EX: SLEEP RECOVERY

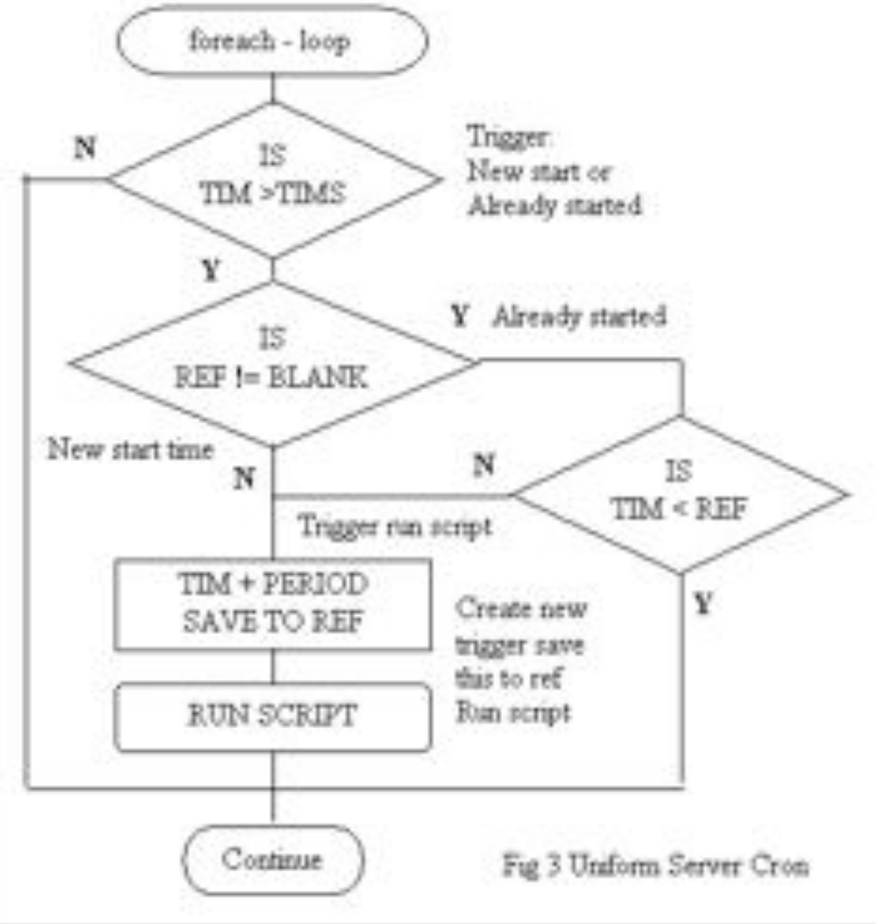
EX: SLEEP RECOVERY





LEVELS OF ABSTRACTION





QUESTIONS? COMMENTS?

@AbeGong
Data Scientist, Jawbone
Strata - February 2014

JAWBONE

Big
Rich
Messy

Small
Focused
Curated

Sensory
User experience
External-facing

Abstract
Business logic
Internal-facing

“Qualitative”
Story-making

“Quantitative”
Science-making

TRANSMUTATION EXAMPLES

Example	Property
Rosetta stone	Synonyms/Comparability
Campaign targeting	Demographic categories
Sleep context	Context
Instrumental variables	Causality
HuffPo moderation	Credibility
Sleep recovery	Clean examples
Economic mobility	Continuity
Crowdflower gold	Credibility

Example	Property
Bridge cases in IRT scaling models	Relative ranking
Sentiment analysis	Categories
Pretty much all supervised learning	Categories/Scales
...	

RECOMMENDED READING

- Pete Skomoroch: http://www.slideshare.net/pskomoroch/strata-endorsements-16939466
- Paco Nathan: http://www.slideshare.net/pacoid/using-cascalog-to-build-an-app-based-on-city-of-palo-alto-open-data
- Jay Kreps: http://engineering.linkedin.com/distributed-systems/log-what-every-software-engineer-should-know-about-real-time-datas-unifying
- Joseph Turian: http://files.meetup.com/1542972/20120202-more-data-same-models-STUDY-SLIDES.pdf
- Me: http://blog.abegong.com/2014/02/wanted-good-examples-of-data-sidekicks.html