

Listen to your Users

Your Data Pipeline

Jeff Bowen

Data Wrangler

huh?
Automattic



<http://www.tor.com/2012/10/09/star-trek-the-next-generation-qa-fistful-of-datasq/>

What my friends think I do



<https://peterskastner.wordpress.com/2011/02/23/cisco-the-lion-king-fights-for-data-center-fabric-leadership/>

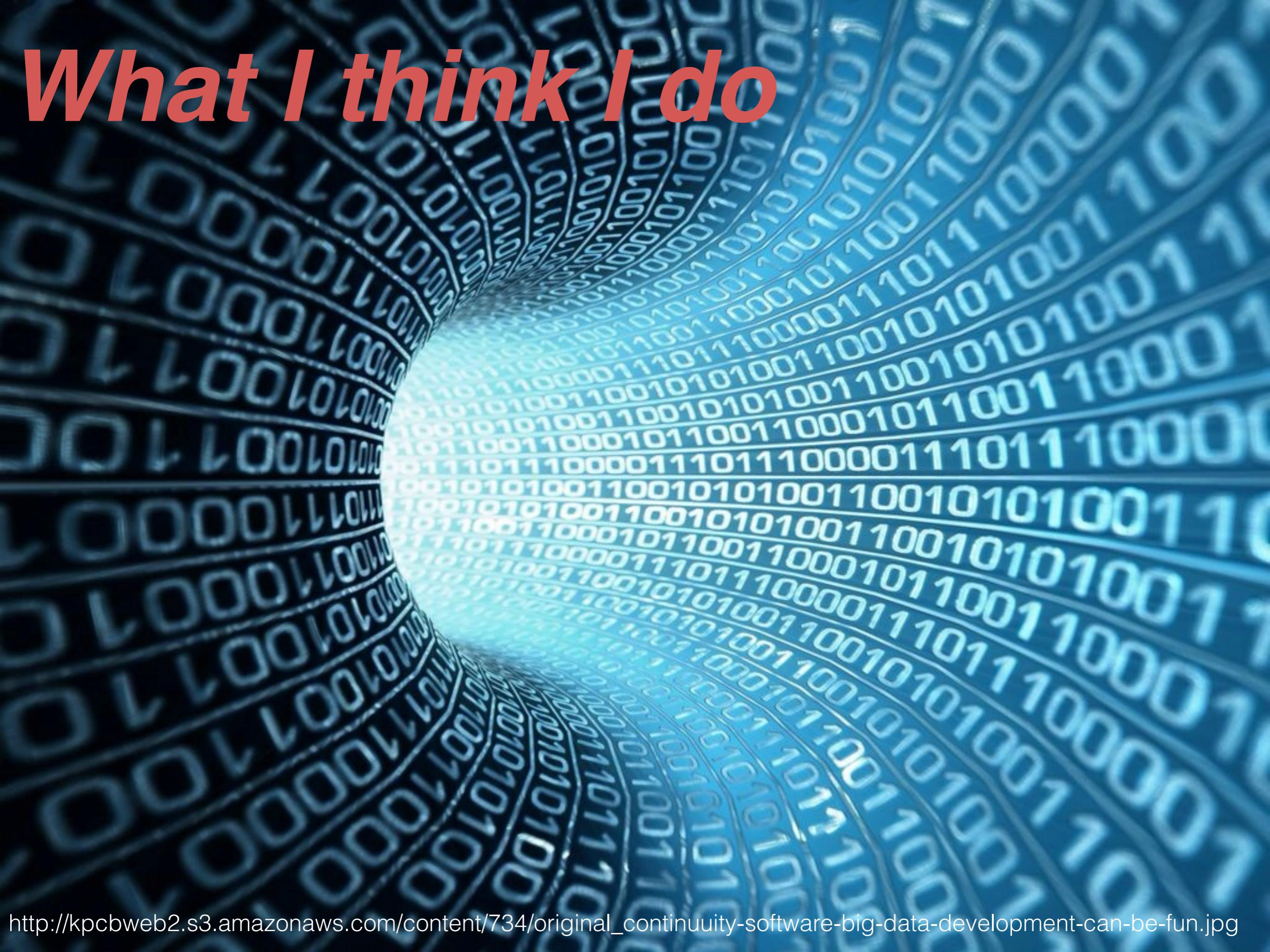


What society thinks I do

What my mom
thinks I do



What I think I do



~~CURSED MAGNETS!~~
Data!



**I AM CONFUNDED AND
PERPLEXED AS TO HOW THEY
DOTH WORK**

What I
actually do

What are "data," anyway?

Facts and statistics collected together for reference or analysis.

http://www.oxforddictionaries.com/us/definition/american_english/data

data are all around us:

- Site profiles for users, customers..
- Content: Documents, posts, pages, comments, images & other attachments (+ associated meta)
- Which products were in your customer's cart when they left your site's store
- Bytes in / out on your bandwidth bill
- The sets of input & output of some function(s) & even the functions themselves

I know what you're
probably thinking...

"OK, smartie pants, how's
about *Data Science*, eh?"

Buzzword?

“'Data Science' could disappear, to be eclipsed by the next buzzword. But it may stick around, just like another incongruous joining of two words, 'computer science,' did.”

<http://www.forbes.com/sites/gilpress/2013/08/19/data-science-whats-the-half-life-of-a-buzzword/>

Data Scientist Merit Badges

- Business Intelligence
- Working with large datasets "Big Data"
- Unstructured / semi-structured data
- "Alternative" DBMS / "nosql"
- Mad IPython Notebook skillz
- Visualizations / charts / graphs / infographics
- Machine Learning / predictive analysis
- Hadoop, Spark, Kafka, Hive, Impala, etc. etc.

Just why..?

“...'data-scientist' is a sexed up term for a statistician. Statistics is a branch of science. Data scientist is slightly redundant in some way and people shouldn't berate the term statistician.”

–Nate Silver

<http://www.statisticsviews.com/details/feature/5133141/Nate-Silver-What-I-need-from-statisticians.html>

“Data science is just science.”

– Jeff Bowen (in my head at some point)

THE SCIENTIFIC METHOD



Question

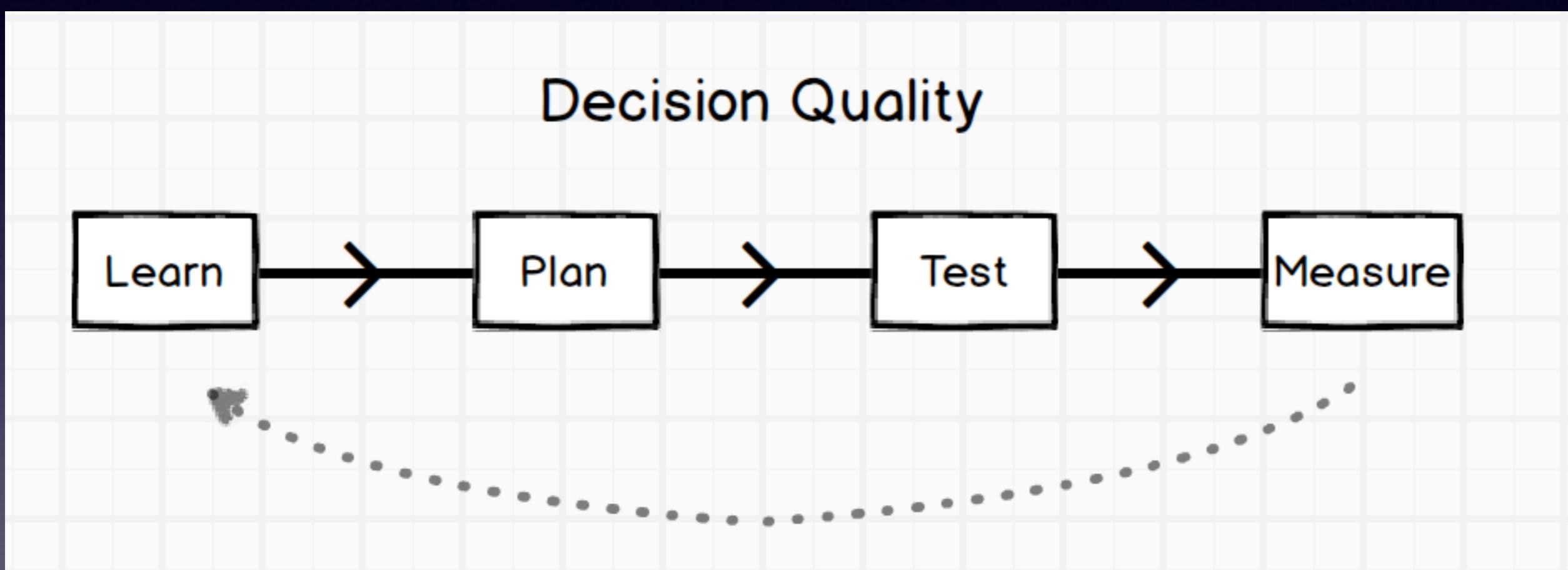
Hypothesis

Procedure

Data

Conclusion

Iterate!



Legality & Ethics

Event Tracking

"Off-the-shelf"

- Jetpack / WordPress.com stats
- Kissmetrics
- Google Analytics
- Mixpanel
- etc.

Pixel logs



<http://design.tutsplus.com/articles/learn-how-to-draw-hand-crafted-pixel-art-in-photoshop--psd-5284>

I mean: Pixel logs

```
1.2.3.4 - - [26/Jun/2015:23:53:45 +0000] "GET /g.gif?  
v=type_of_event&x_fictional_event_name=segment_a&baba=0.15767723880708218  
HTTP/1.0" 200 50 "http://jblz.dev/" "Mozilla/5.0 (Macintosh; Intel Mac OS X  
10_10_4) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/43.0.2357.134 Safari/  
537.36" - 0.002 -
```

<http://design.tutsplus.com/articles/learn-how-to-draw-hand-crafted-pixel-art-in-photoshop--psd-5284>

Front vs. Back-end Tracking

Back-end Tracking

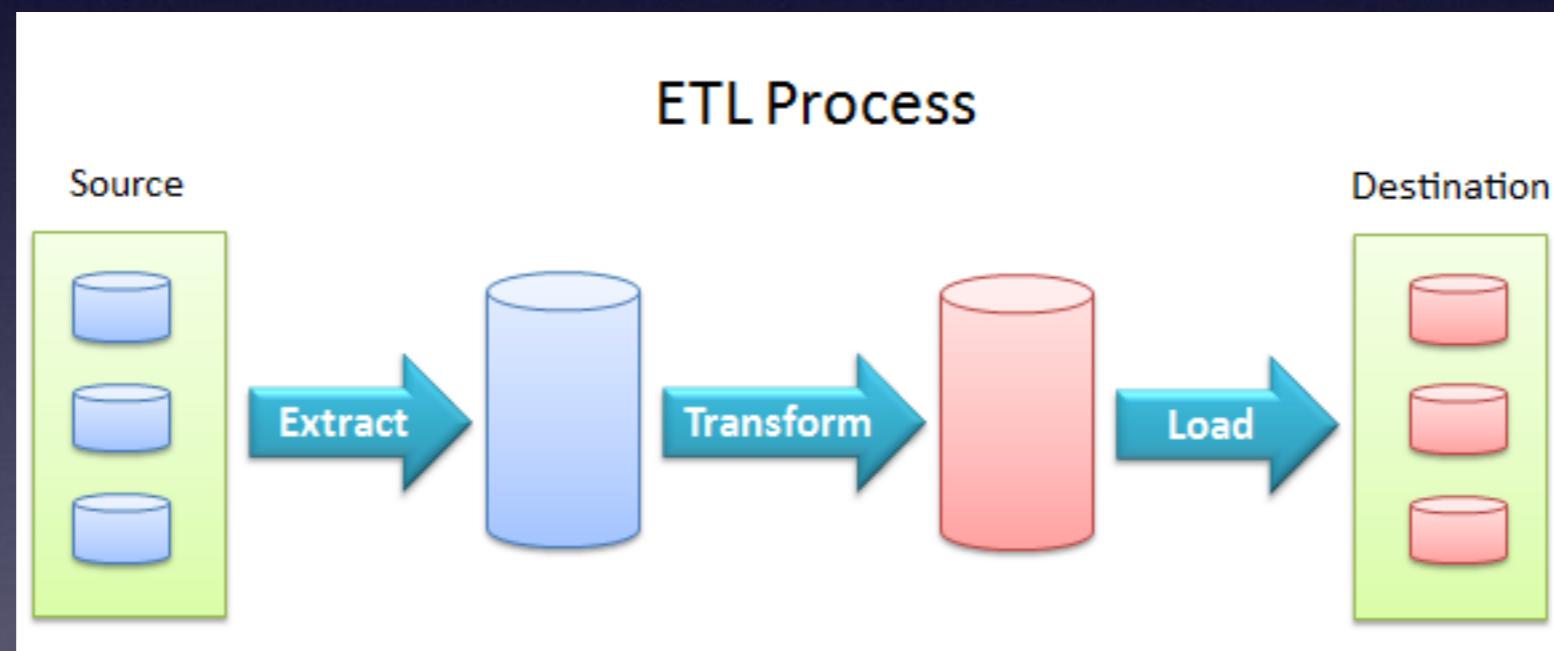
```
function supertracker_track_comment_post( $comment_id, $status )
{
    if ( $status === 'spam' ) {
        return TrackingLibrary::track( 'spam_comment' );
    }
    TrackingLibrary::track( 'comment', $comment_id );
}
add_action( 'comment_post', 'supertracker_track_comment_post' );
add_action( 'edit_comment', 'supertracker_track_comment_post' );
```



quickmeme.com

<http://www.quickmeme.com/meme/3of874>

ETL (Extract, Transform, and Load)



<http://blog.appliedinformaticsinc.com/etl-extract-transform-and-load-process-concept/>

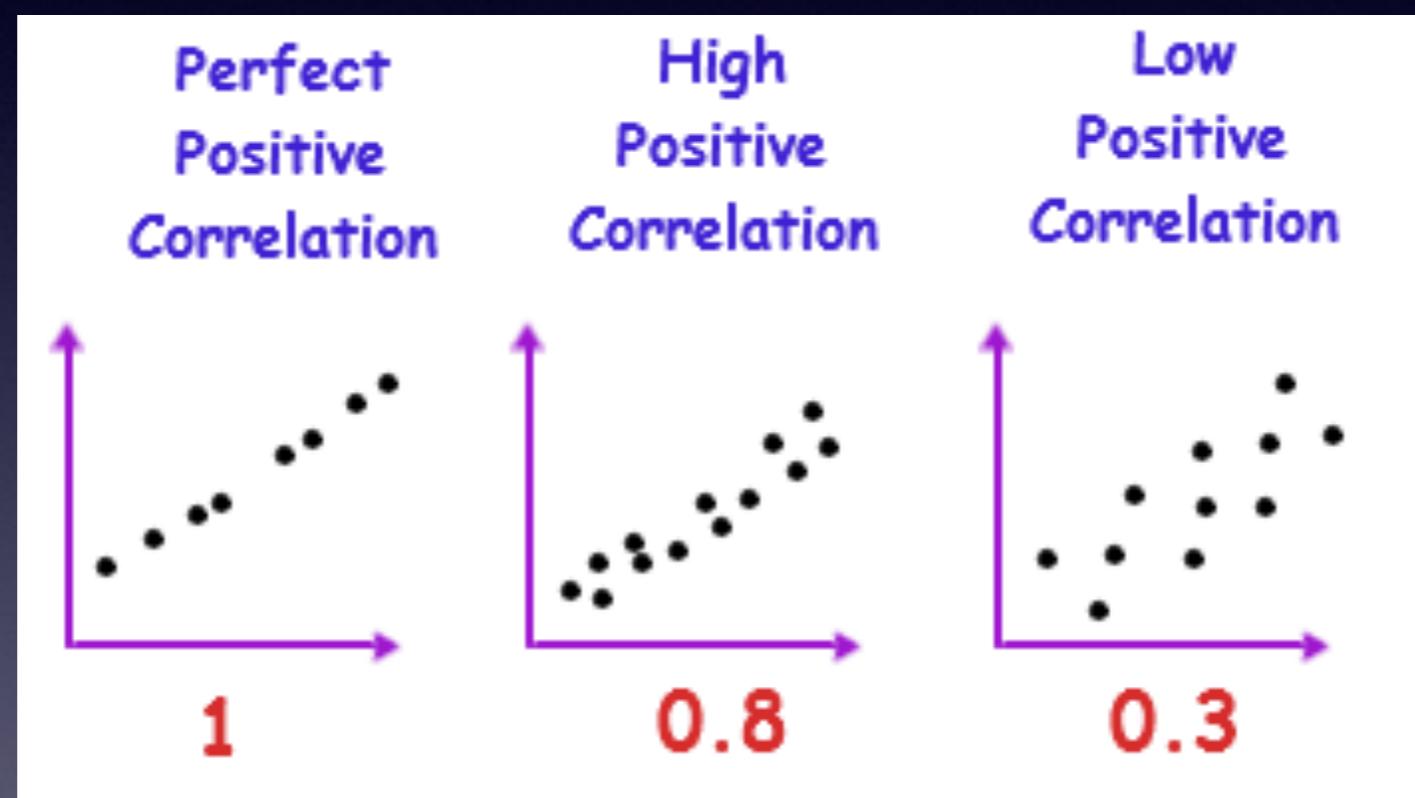
...the elusive "why"

Correlation

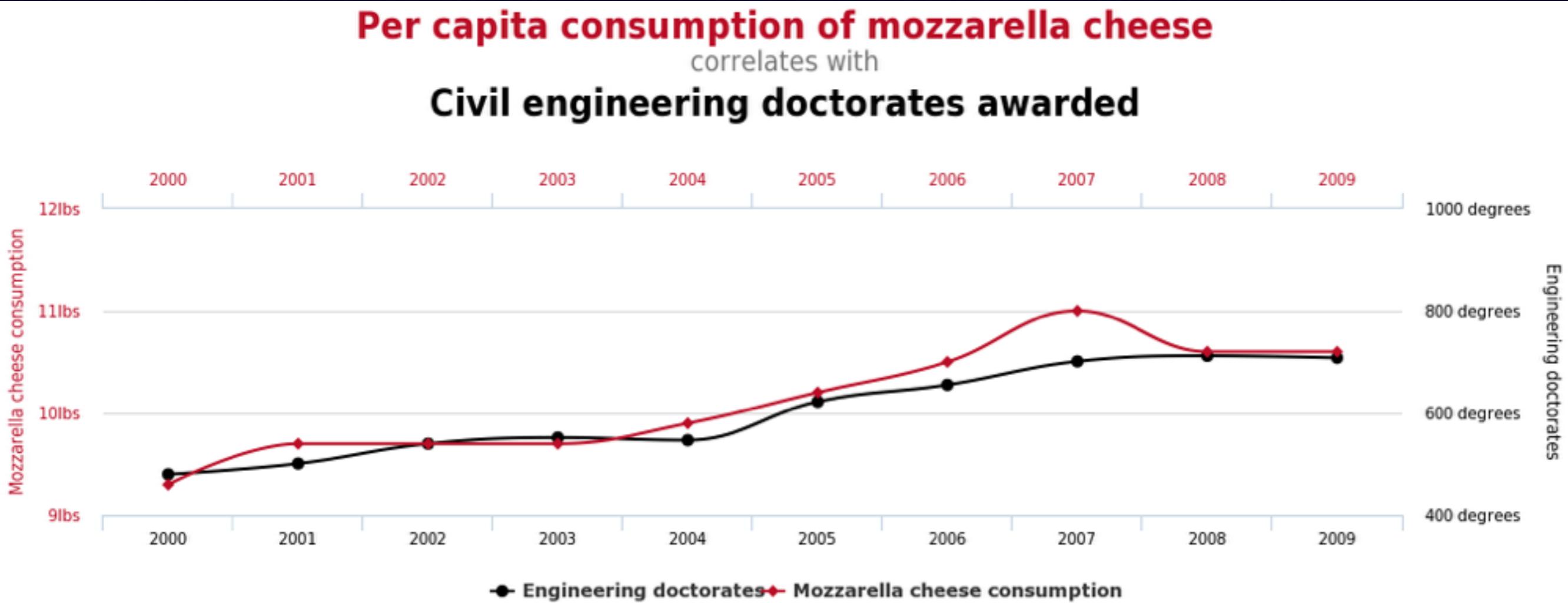
The formula for correlation is

$$r = \frac{N \sum xy - (\sum x)(\sum y)}{\sqrt{[N \sum x^2 - (\sum x)^2][N \sum y^2 - (\sum y)^2]}}$$

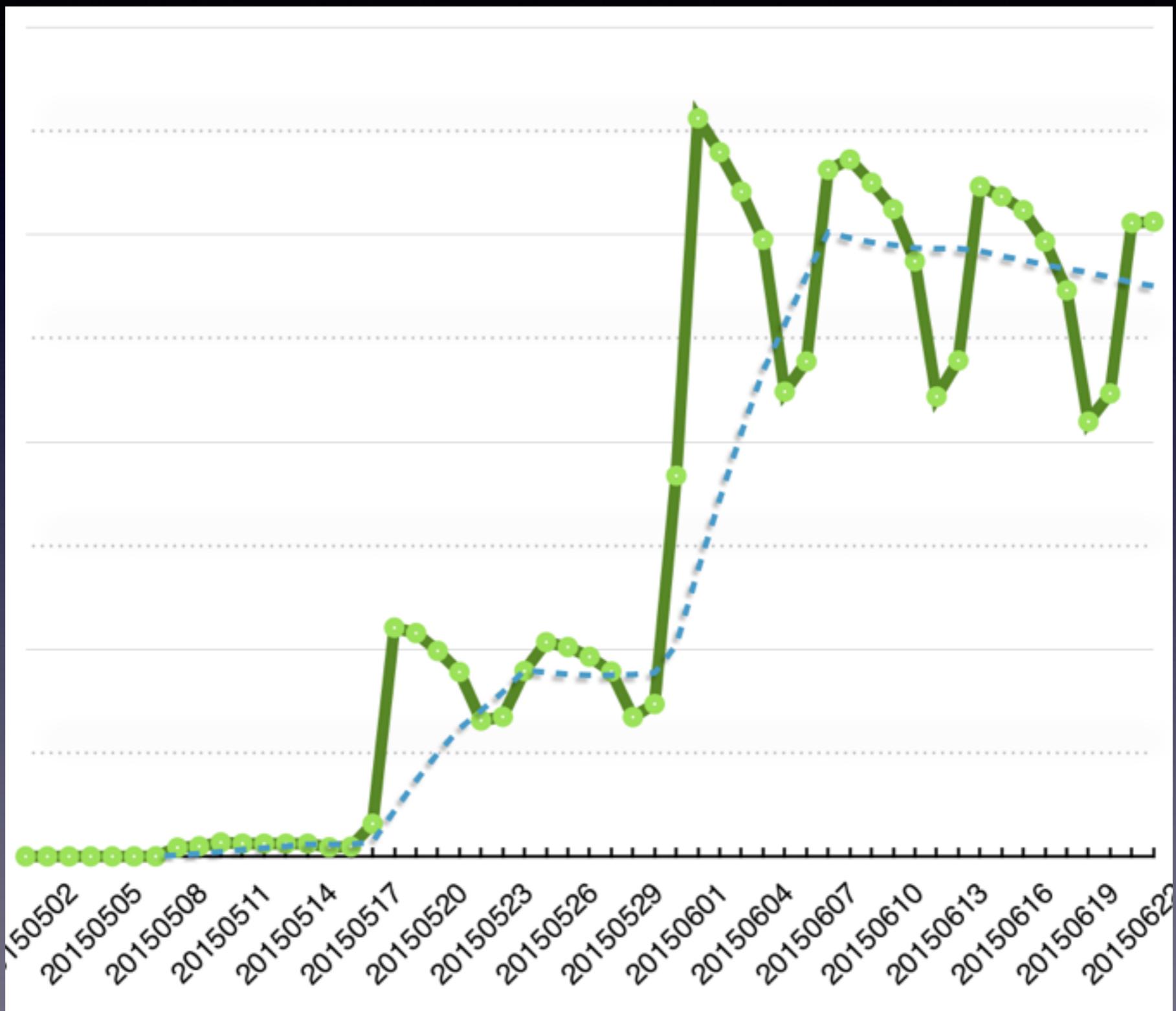
...correlation



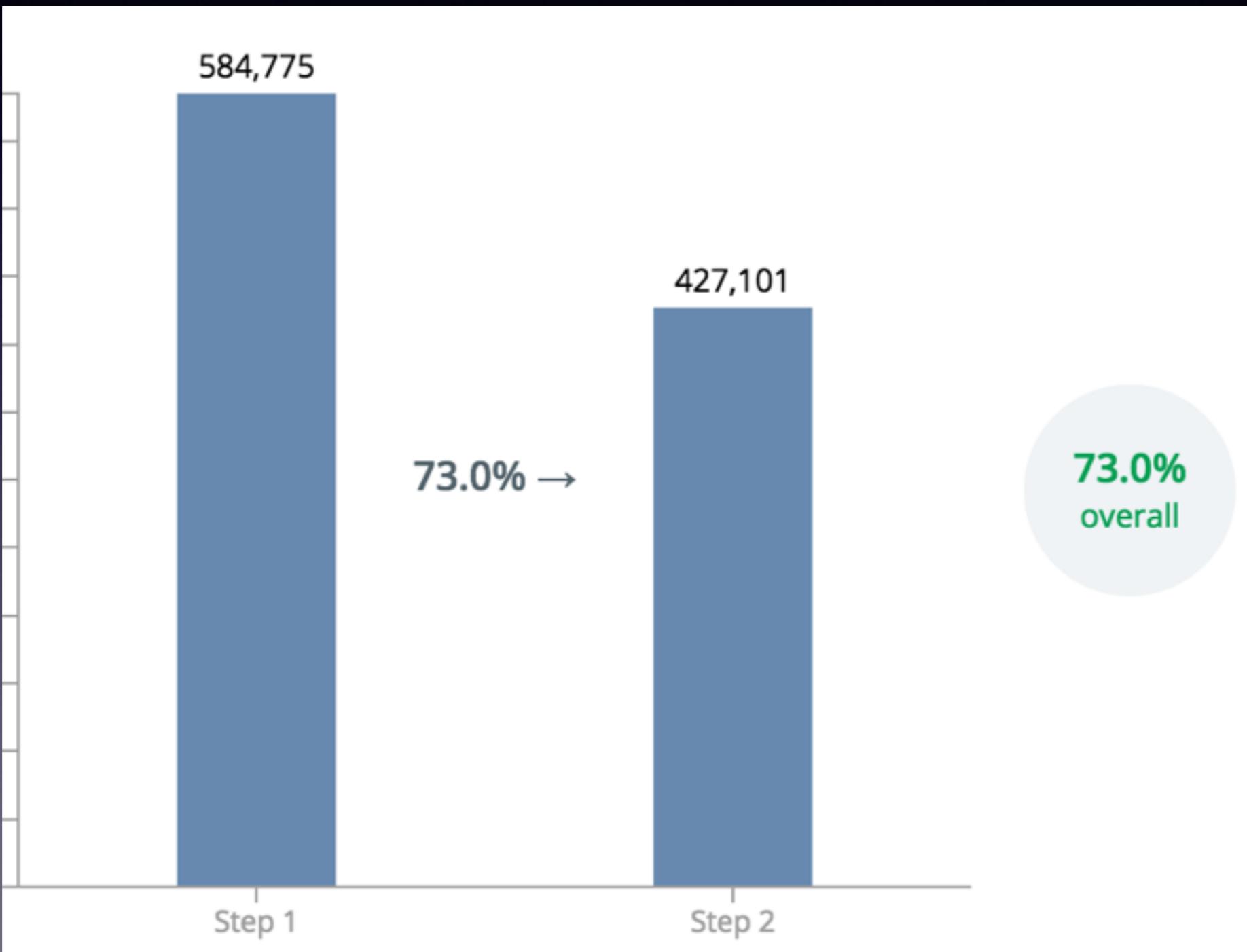
Correlation does not imply causation



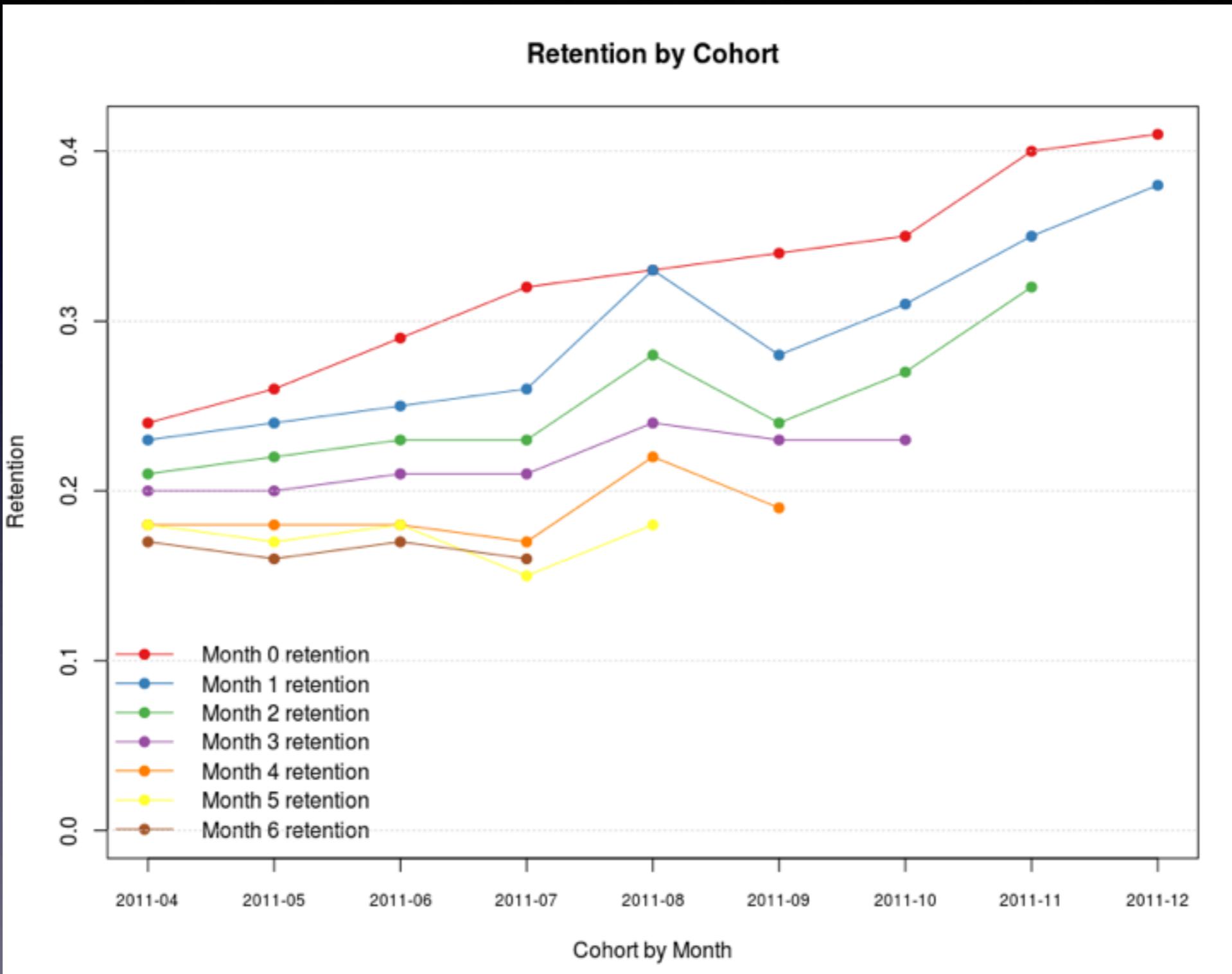
Trends



Funnels



Cohorts

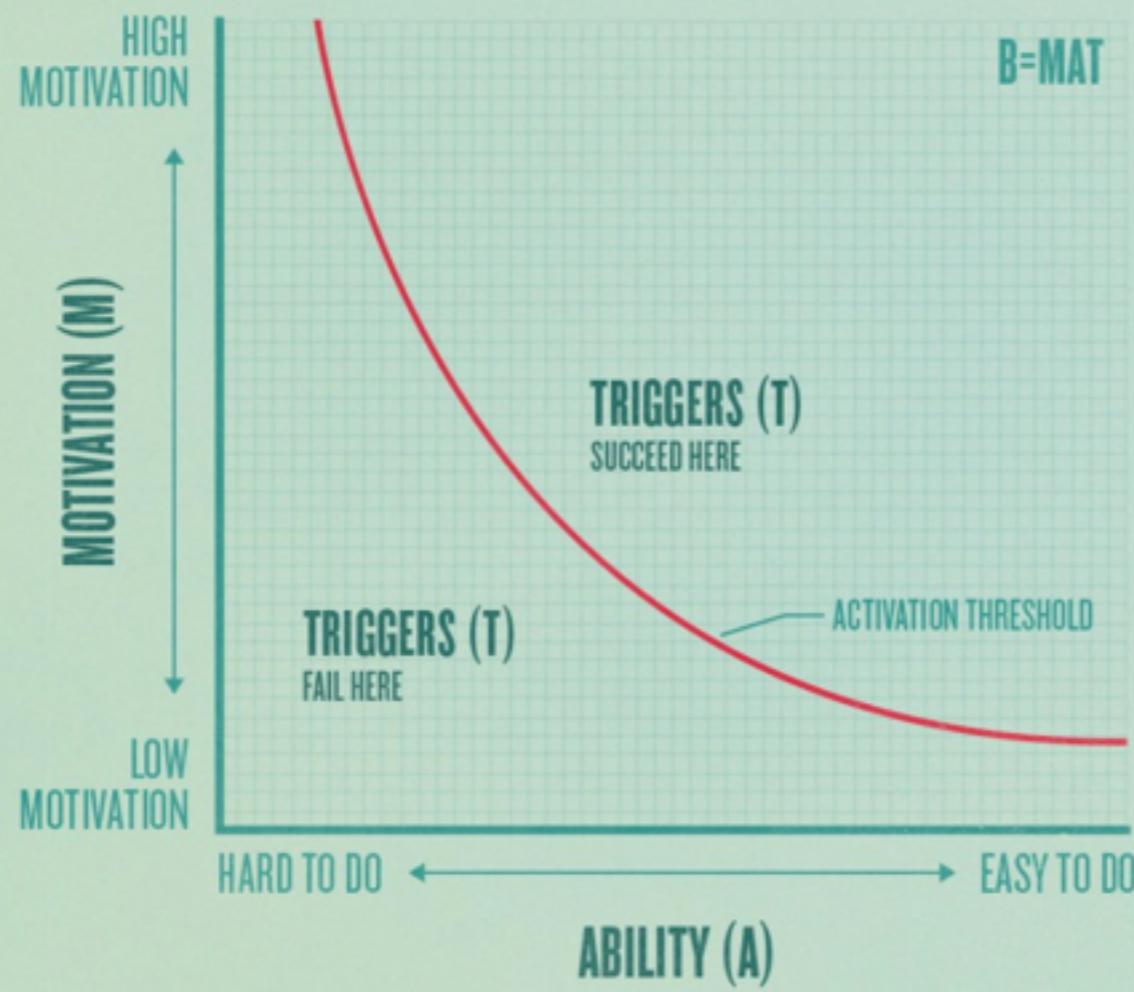


Sampling

WHAT CAUSES BEHAVIOR CHANGE?

THE FOGG BEHAVIOR MODEL (FBM) SHOWS THAT THREE ELEMENTS MUST CONVERGE AT THE SAME MOMENT FOR A BEHAVIOR TO OCCUR: MOTIVATION, ABILITY, AND TRIGGER (MAT). WHEN A BEHAVIOR DOES NOT OCCUR, AT LEAST ONE OF THE THREE ELEMENTS IS MISSING.

THE FOGG BEHAVIOR MODEL



BEHAVIOR CHANGE ELEMENTS



CORE MOTIVATORS



SIMPLICITY FACTORS



“A good data scientist is therefore able to get in the mind of people who use our product and understand their needs.”

<http://nerds.airbnb.com/scaling-data-science/>

Listen to your Users

Your Data Pipeline

Questions?

Jeff Bowen

twitter.com/alapapa

github.com/jblz

profiles.wordpress.org/jblz