

Analytics is the discovery and communication of meaningful patterns in data.

Survey Results

What should we talk about?

Analytics for your app	17
Analytics for marketing	13
Analytics for business	11
Big data	12

First: some context on big data and analytics

Wednesday, February 13, 13

Every company is becoming a **software company.
Every software company is becoming a **data company**.**

“Big Data” and “Analytics” are kind of a thing right now.

WSJ WSJ LIVE MARKETWATCH BARRON'S ALLTHINGS.DI FACTIVA RISK & COMPLIANCE

THE WALL STREET JOURNAL.

U.S. EDITION Sunday, April 29, 2012 As of 9:44 AM EDT

Home World U.S. New York Business Tech Markets Market Data Opinion

TOP STORIES IN Technology 1 of 12 Apple, Samsung Back in Court Kodak Gets Bid of More Than \$500 Million for Patent

April 29, 2012, 9:44 a.m. ET

Big Data's Big Problem: Little Talent

Article Video Stock Quotes Comments (46)

Email Print Save [f](#) [t](#) [g](#) [in](#) A A

It seems that the markets are as much in love with "Big Data"—the ability to acquire, process and sort vast quantities of data in real time—as the technology industry.



Hilary Mason, chief scientist for the URL shortening service Bitly, outlines the key skills that data scientists must have.

The first Big Data initial public offering hit the market last week to roaring approval. [Splunk Inc.](#), [SPLK +0.31%](#) which helps businesses organize and make sense of all the information they gather, soared 109% on its first day of trading. Big Data, big price.

And this week, in cities in the U.S. and the U.K., Big Data Week events are being held to proselytize the unbelievers.

Big Data refers to the idea that an enterprise can mine all the data it collects right

Harvard Business Review

THE MAGAZINE BLOGS AUDIO & VIDEO BOOKS WEBINARS COURSES

Guest | limited access Register today and save 20%* off your first order! [Details](#)

THE MAGAZINE October 2012

ARTICLE PREVIEW To read the full article: [Sign in](#) or [Register](#) for free. HBR Subscribers [activate your free archive access »](#)

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

Comments (6) [Email](#) [Print](#) [Save](#) [f](#) [t](#) [g](#) [in](#)



RELATED

Executive Summary

ALSO AVAILABLE

Buy PDF

Artwork: Tamar Cohen, *Andrew J. Buboltz*, 2011, silk screen on a page from a high school yearbook, 8.5" x 12"

When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives

Wednesday, February 13, 13

<http://online.wsj.com/article/SB10001424052702304723304577365700368073674.html>
<http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>



POSSIBLE @POSSIBLE

We have a Sr. Analytics & Optimization Manager opportunity in #Seattle office [ow.ly/fSQcP](#) #analytics #jobs

[Expand](#)

1h



iMedia Jobs @imediajobs

#jobs #imedia Business Analyst - Web Analytics - TechnoSphere, Inc
- Atlanta, GA: Deep understa... [bit.ly/120EoDF](#) #web #analytics

[Expand](#)

4h

job security.



nyttjobb.tw @nyttjobb_tw

Web Analyst for LBI! [wp.me/p2jx2J-9e](#) via @NuCreatives
#webanalyst #jobs #LBI #Nucreatives #analytics [dlvr.it/2bMzc3](#)

[View media](#)

7h



Terry @TerryJobs

5 Dec

Use Google #Analytics to Improve Your #SocialMedia Strategy ...
#Recruiter #Jobs #SocialRecruiting [lnkd.in/SI-C6h](#)

[Expand](#)



Anametrix @anametrix

5 Dec

Join the #Digital #Analytics Revolution! @anametrix is hiring! Lots of positions available. --> [bit.ly/VjwZ1O](#) #jobs

[Expand](#)



Zillion Jobs @ZillionJobs

5 Dec

Business Analytics Manager Job (Olathe, KS) [zillionjobs.com/Job-6665643.htm](#)... #Analytics #Manager #job #jobs #Olathe

[Expand](#)



RosettaCareers @RosettaCareers

4 Dec

We are looking for an Analytics Director in Cleveland w/8+ yrs #measure exp, apply here [lnkd.in/uzBE-M](#) or DM me #jobs
#analytics

[Expand](#)



Social Media Club @SMCSTL

3 Dec

Know a good fit for this role? > Senior Manager, Digital Bear Marketing #STL #Jobs #SMCSTL
[careers.buildabear.com/Careers.aspx](#) #Analytics #SocialMedia

[Expand](#)



Michael Young @michaelyoungMBN

2 Dec

"@mbnrecruitment: #Jobs Senior Analytics & Optimization Manager - £50000 - £60000 pa [bit.ly/VgoB3L](#)" #bigdata #sql #SAS #Analytics
@ from Glasgow City Glasgow City

Wednesday, February 13, 13

Cool data stories

Wednesday, February 13, 13

There are entire professions and really amazing work happening in different areas of analytics.

I have some stories to illustrate work in each of these realms.

Data modeling - Brahe & Kepler

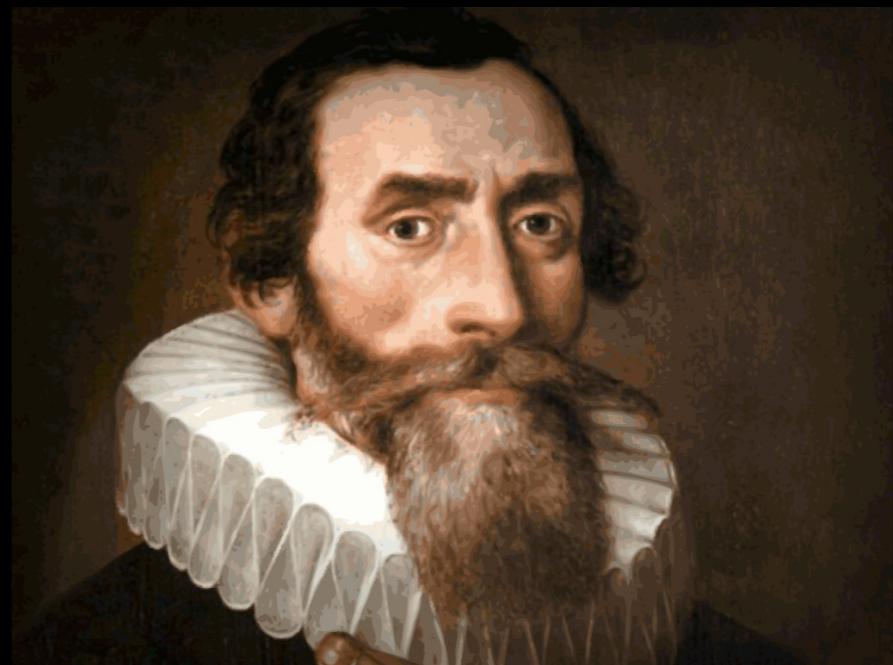
Data mining/analysis - Linked In

Data viz - Infosthetics

Communication - Broad Street Pump



Tycho Brahe



Johannes Kepler

Wednesday, February 13, 13

Story 1: Data Modeling - what data should we record and how should we record it?

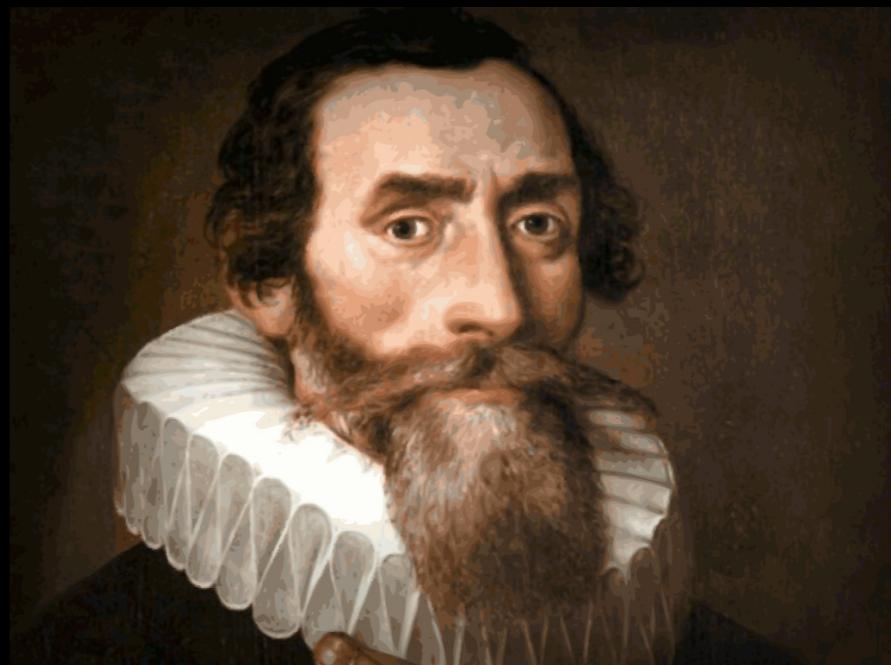
Tycho Brahe – collected an astounding amount of astronomical data. Every night, he would write down the location of every star and every planet in the sky. After 30 years of doing this... he died.

Johannes Kepler – took Brahe's data and single-handedly discovered the laws of planetary motion.

The point: recording stuff is important!



Tycho Brahe



Johannes Kepler



aka science walrus

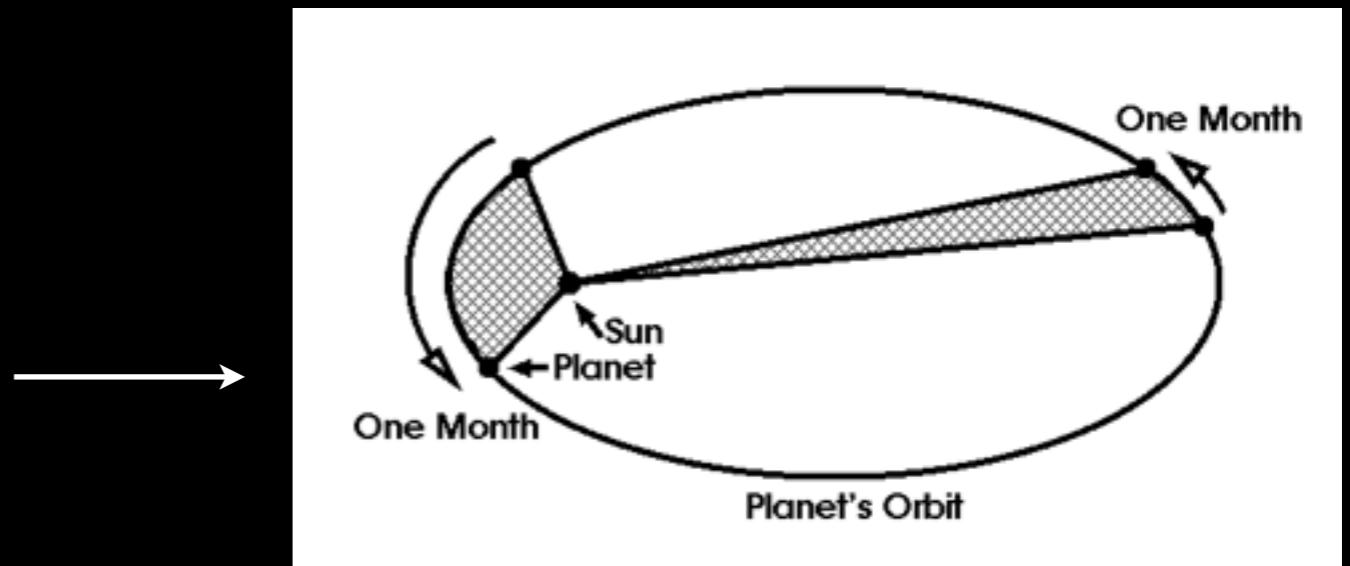
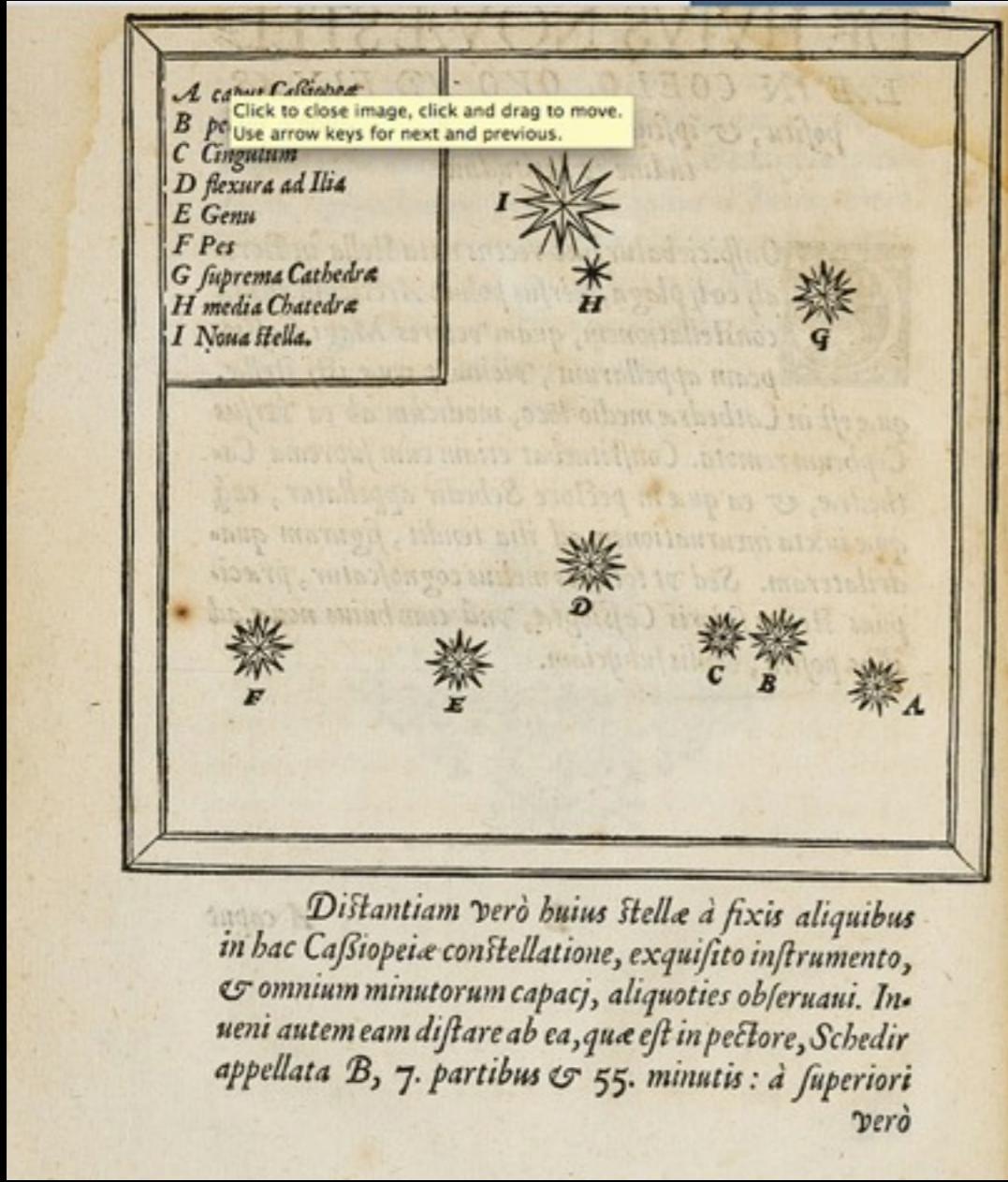
Wednesday, February 13, 13

Story 1: Data Modeling - what data should we record and how should we record it?

Tycho Brahe – collected an astounding amount of astronomical data. Every night, he would write down the location of every star and every planet in the sky. After 30 years of doing this... he died.

Johannes Kepler – took Brahe's data and single-handedly discovered the laws of planetary motion.

The point: recording stuff is important!



Tycho Brahe

Johannes Kepler

Wednesday, February 13, 13

Second point: sometimes you don't know how the data you are collecting will be used.

Other cool data collection: quantified self movement, 23andme, sensor data on vehicles,

The screenshot shows the LinkedIn homepage. At the top, it displays 'Account Type: Basic | Upgrade' and a notification bar with '23' and '1'. The main navigation menu includes Home, Profile, Contacts, Groups, Jobs, Inbox, Companies, News, and More. A search bar with 'Search...' and 'Advanced' options is also present. Below the menu, a banner reads 'Want \$ for a Trust Deed - Get cash for real property secured loans, performing or non-performing.' On the left, there's a 'Share an update' box and a 'LinkedIn Today' news recommendation featuring a fire icon and a thumbnail for '£50m new investment in London's Tech City' by David C. On the right, a section titled 'PEOPLE YOU MAY KNOW' is highlighted with a red circle. It lists three profiles with 'Connect' buttons: Valerie Serrin, Izzy Johnston, and Peggy Hills. Below this is an 'ADS BY LINKEDIN MEMBERS' section with a thumbnail for 'Build Huge Email List?'.

Wednesday, February 13, 13

Story 2: Data Analysis - what can we learn from data? what tools and algorithms can we use to unlock new meaning?

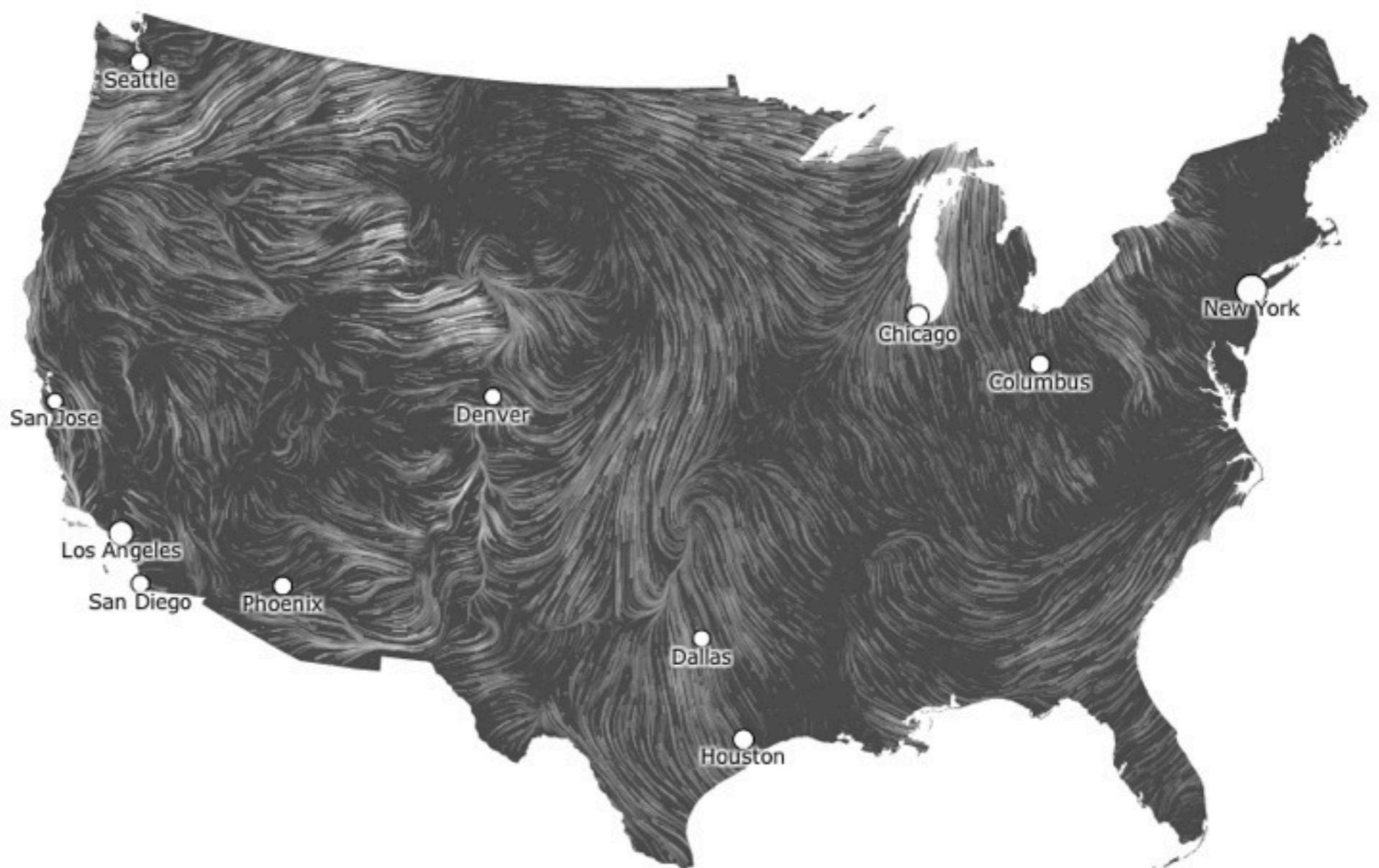
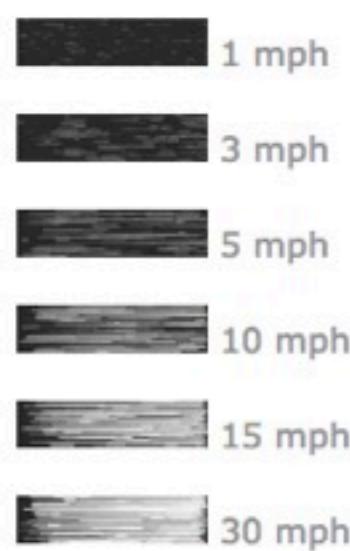
- Goldman, a PhD in physics from Stanford, was intrigued by the linking activity he saw happening on LinkedIn. He began exploring people's connections, forming theories, testing hunches, and finding patterns that allowed him to predict whose networks a given profile would land in.
- He could imagine that new features capitalizing on the heuristics he was developing might provide value to users. But LinkedIn's engineering team, caught up in the challenges of scaling up the site, seemed uninterested. Some colleagues were openly dismissive of Goldman's ideas. Why would users need LinkedIn to figure out their networks for them? The site already had an address book importer that could pull in all a member's connections.
- LinkedIn's cofounder and CEO at the time had faith in the power of analytics because of his experiences at PayPal, and he had granted Goldman a high degree of autonomy. For one thing, he had given Goldman a way to circumvent the traditional product release cycle by publishing small modules in the form of ads on the site's most popular pages.
- Goldman started to test what would happen if you presented users with names of people they hadn't yet connected with but seemed likely to know—for example, people who had shared their tenures at schools and workplaces. He did this by ginning up a custom ad that displayed the three best new matches for each user based on the background entered in his or her LinkedIn profile. Within days it was obvious that something remarkable was taking place.
- The click-through rate on those ads was the highest ever seen. Goldman continued to refine how the suggestions were generated, incorporating networking ideas such as "triangle closing"—the notion that if you know Larry and Sue, there's a good chance that Larry and Sue know each other. Goldman and his team also got the action required to respond to a suggestion down to one click.
- It didn't take long for LinkedIn's top managers to recognize a good idea and make it a standard feature. That's when things really took off. "People You May Know" ads achieved a click-through rate 30% higher than the rate obtained by other prompts to visit more pages on the site. They generated millions of new page views. Thanks to this one feature, LinkedIn's growth trajectory shifted significantly upward.

Dec. 6, 2012

5:59 pm EST

(time of forecast download)

top speed: **30.2 mph**
average: **6.2 mph**



<http://hint.fm/wind/>

<http://infosthetics.com/>

Wednesday, February 13, 13

<click link for animation>

Story 3: Data Visualization - how can we use visuals to discover trends?

Sometimes visuals can tell us things we can't see from raw data.

There are some amazing projects out there (see infosthetics)

<http://hint.fm/wind/>

<http://infosthetics.com/>

Wednesday, February 13, 13

Story 4: Data Communication - How do we share what we have learned with others? What does the data tell us (and what does it leave out)? One critical skill required of a data analyst is how to communicate what the results mean and what we should do next. The following is the story of how data sampling and visualization were used to communicate something very important.

Cholera hit London in 1854. Waste management systems were really bad and consisted of vats in people's basements. In 3 days, 127 people near Broad Street Died.

The mortality rate was 12.8 percent in some parts of the city.

By the end of the outbreak, 616 people had died.

http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak

Le Petit Journal

ADMINISTRATION
61, RUE LAFAYETTE, 61
Les manuscrits ne sont pas rendus
On s'abonne sans frais
Renseignez les bureaux de poste.

5 CENT. SUPPLÉMENT ILLUSTRÉ 5 CENT.
23^e Année — 44 — Numéro 1.150
DIMANCHE 1^{er} DÉCEMBRE 1912

ABONNEMENTS
PARIS
SEINE ET SEINE-ET-OISE... 2 fr. 3 fr. 80
DÉPARTEMENTS... 2 fr. 4 fr. 80
ÉTRANGER... 2.50 5 fr. *



LE CHOLÉRA

Wednesday, February 13, 13

Story 4: Data Communication - How do we share what we have learned with others? What does the data tell us (and what does it leave out)? One critical skill required of a data analyst is how to communicate what the results mean and what we should do next. The following is the story of how data sampling and visualization were used to communicate something very important.

Cholera hit London in 1854. Waste management systems were really bad and consisted of vats in people's basements. In 3 days, 127 people near Broad Street Died.

The mortality rate was 12.8 percent in some parts of the city.

By the end of the outbreak, 616 people had died.

http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak



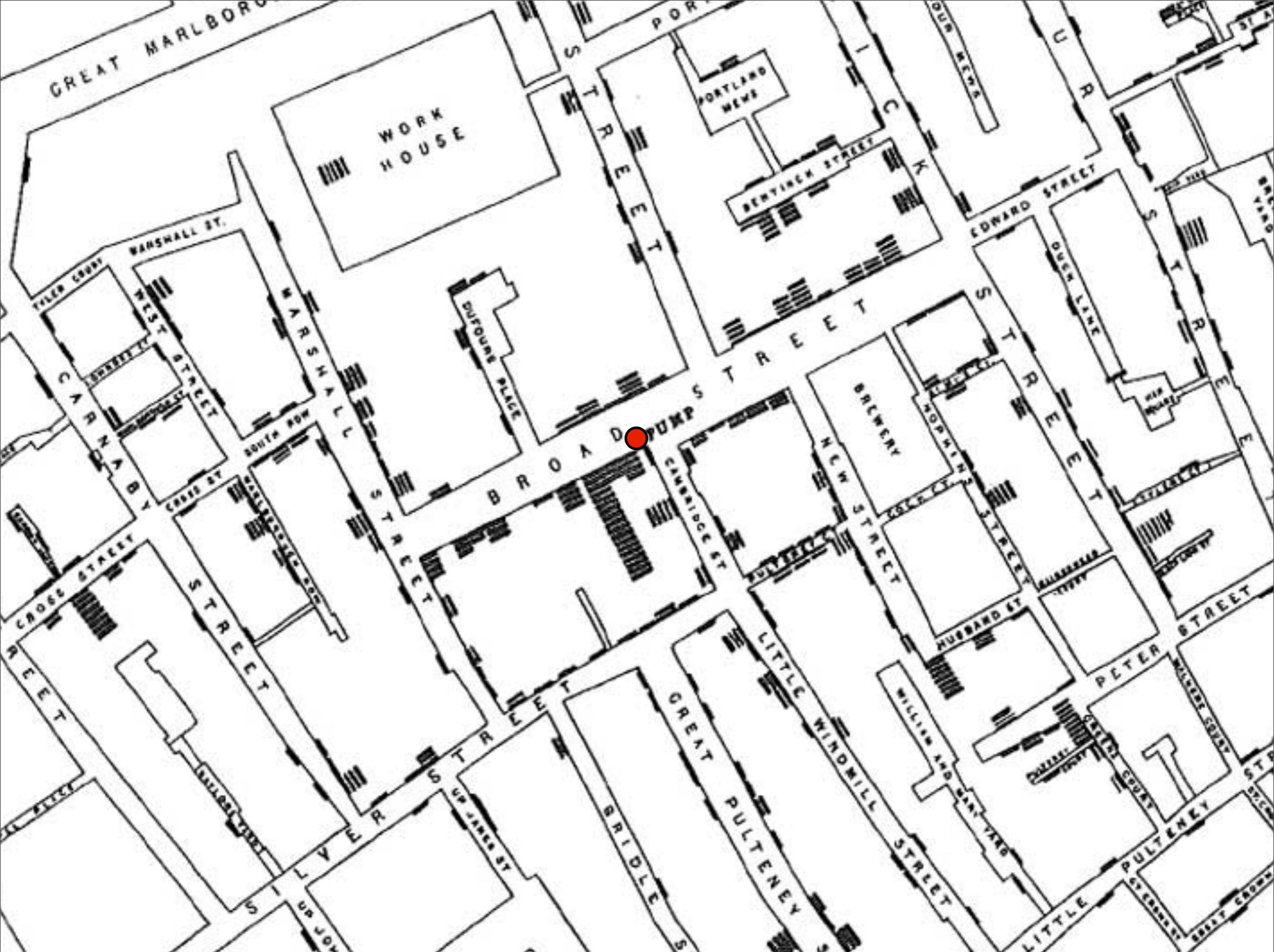
Figured out Cholera
spreads through water.

No one believed him :(

John Snow

Wednesday, February 13, 13

John Snow hypothesized that Cholera was being spread through the water supply, but no one believed him.



Wednesday, February 13, 13

Until he created this awesome image.
He counted deaths in each building and used this data visualization to prove his theory.
Little black bars are deaths.
The broad street pump was shut down and deaths stopped.

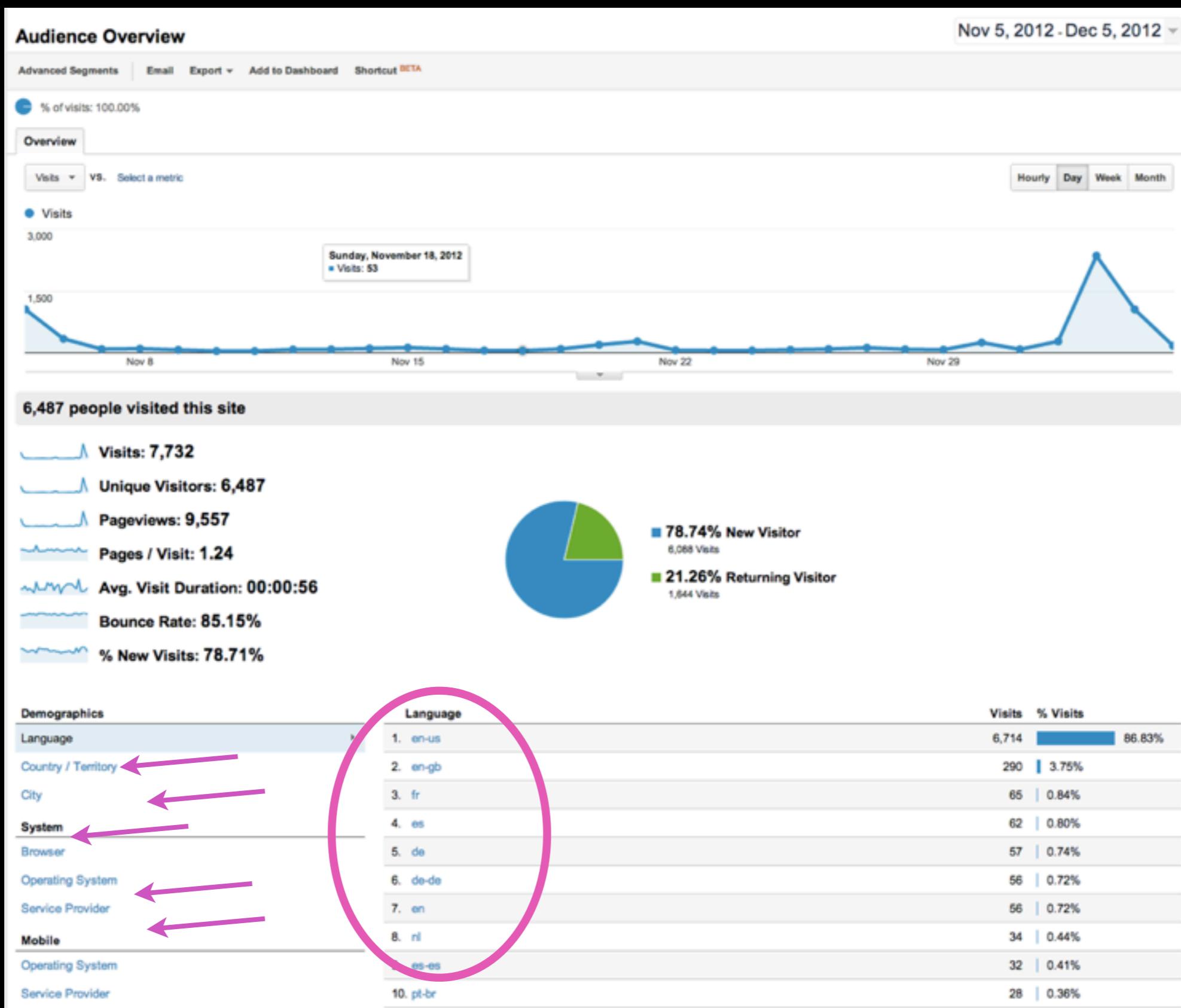
Next: analytics in your app.

Wednesday, February 13, 13

Use analytics to make your app better

1. Figure out who is using your app
2. Figure out what features are most used
3. Figure out where people are quitting

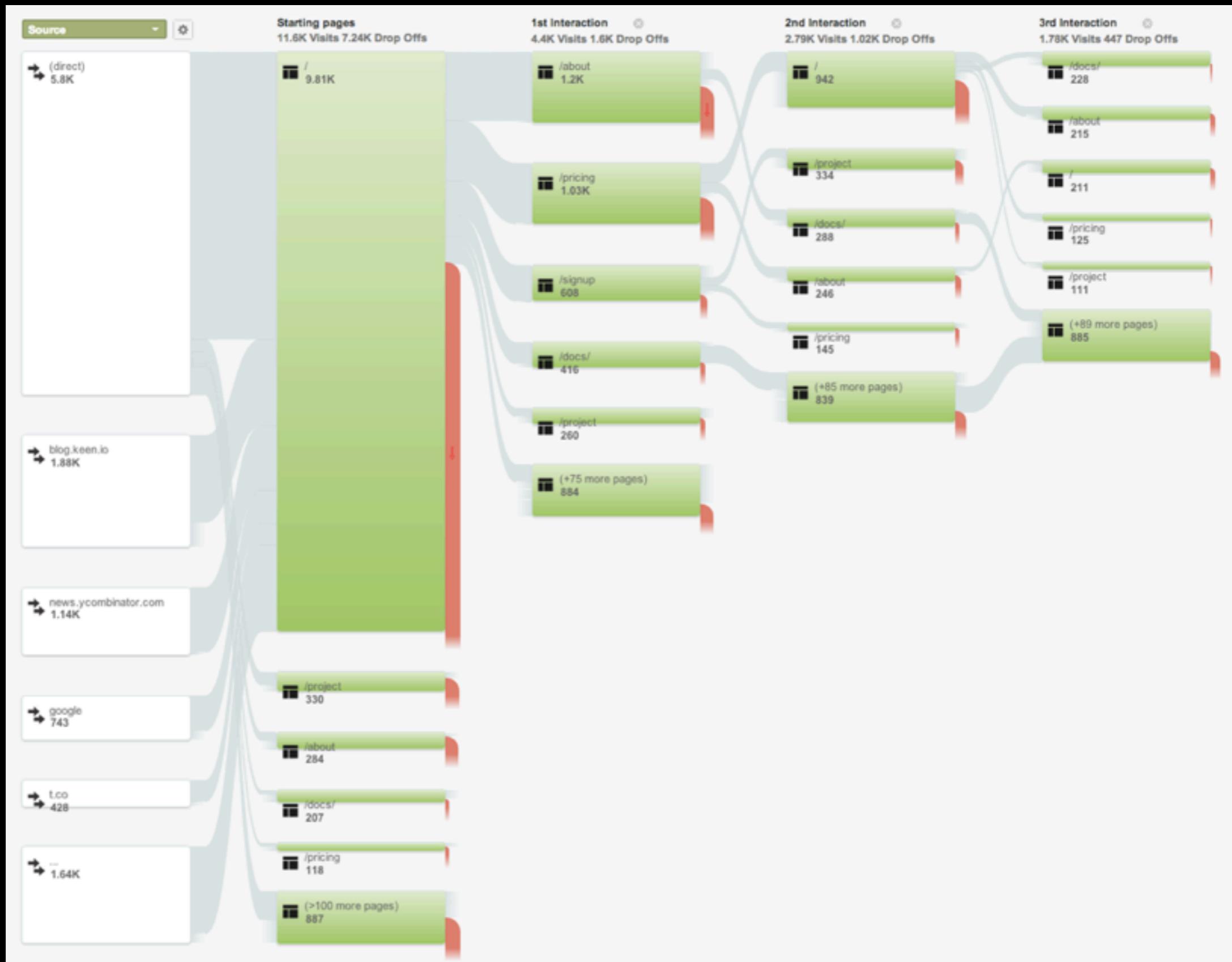
Google analytics is freeeeeeee!



Wednesday, February 13, 13

<live demo>

Google analytics “flows” show user click paths

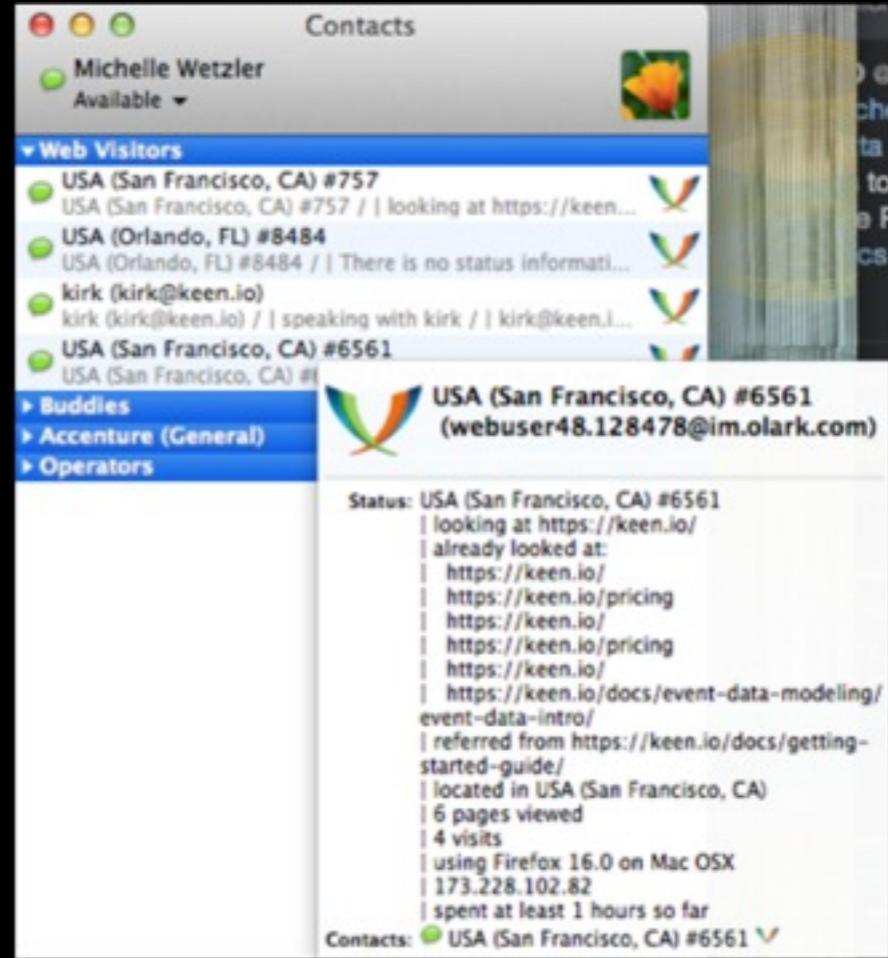


Wednesday, February 13, 13

Who is using your app?

USER	COL2
thomas@github.com	
micah@uber.com	
shelly@google.com	
sam.livingston@urban.uk	
keith@urbanairship.com	
kelly@myspace.com	
janglingbones@gmail.com	

use a DB to
store login data



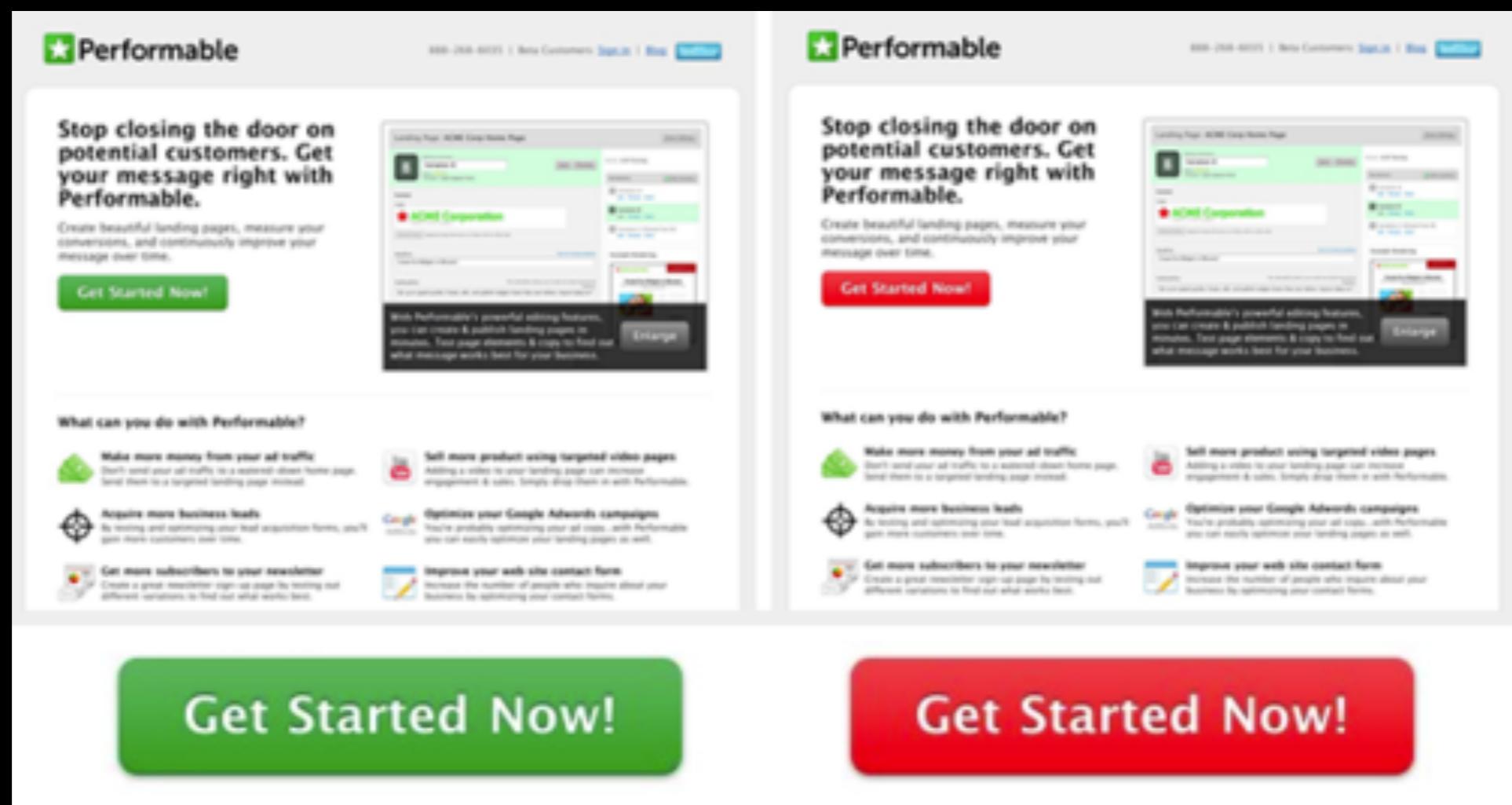
Olark shows live visits

Wednesday, February 13, 13

Record logins to your app. Use tools like KissMetrics, Mixpanel, Keen IO.

- who is using your app the most?
- who hasn't logged in for a long time?
- what companies do your users work at?
- track last login for each customer

A/B Testing aka Split Testing

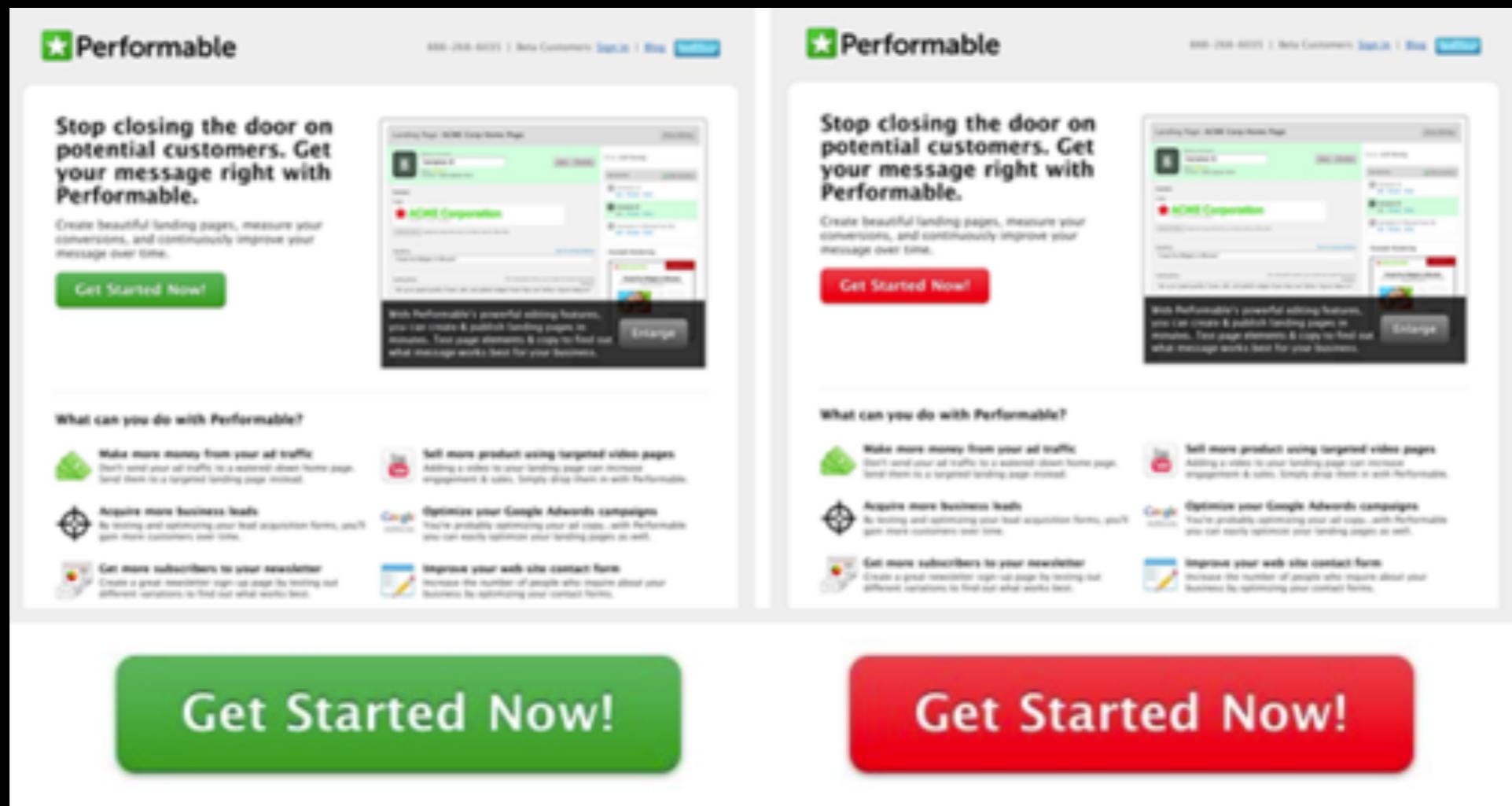


Wednesday, February 13, 13

http://en.wikipedia.org/wiki/A/B_testing

- can be applied to even the smallest things (e.g. button color)
- used extensively in advertising & marketing
- used to get “the last 20%” in app optimization
- tells you which option is better, but won’t tell you if both of them suck!
- some tools now automate this and will automatically serve the most popular version of site

A/B Testing aka Split Testing



21% more people clicked on the red button than on the green button!

Wednesday, February 13, 13

http://en.wikipedia.org/wiki/A/B_testing

- can be applied to even the smallest things (e.g. button color)
- used extensively in advertising & marketing
- used to get “the last 20%” in app optimization
- tells you which option is better, but won’t tell you if both of them suck!
- some tools now automate this and will automatically serve the most popular version of site

Example of split testing data

Signup Events

user.id	user.name.first	user.name.last	form.version	form.fields
223655	zach	morris	A	[first name, middle name, last name, organization, gender, age, email, password]
223656	kelly	kapowski	B	[email, password]
223657	screech	powers	A	[first name, middle name, last name, organization, gender, age, email, password]
223658	lisa	turtle	A	[first name, middle name, last name, organization, gender, age, email, password]
223659	ac	slater	B	[email, password]
223660	jessie	spano	B	[email, password]
223661	mr.	belding	B	[email, password]
223662	mrs.	culpepper	B	[email, password]
223663	stacey	carosi	A	[first name, middle name, last name, organization, gender, age, email, password]
223664	allison	fox	B	[email, password]
223665	tori	scott	B	[email, password]
223666	mr.	dewey	B	[email, password]
223667	ollie	creeky	B	[email, password]
223668	violet	bickerstaff	B	[email, password]
223669	rhonda	robistelli	B	[email, password]

Which version of the form was more effective?

Name an example of something in your app that you would like to A/B test.

What is Event Data?



Actions & State + Time

Wednesday, February 13, 13

now we can record more stuff than ever!

we can store information about *every user interaction*.

new analytics tools like KissMetrics, MixPanel, and Keen IO allow you to record event data about your application.

traditional db stores state in tables: users, inventory, accounting data

event data bases store Actions, State, and Time!

Ref: <https://speakerdeck.com/benbjohnson/behavioral-databases>

```
purchase = {  
    timestamp: 2012-06-06T19:10:39.205000,  
    item: sophisticated orange turtleneck with deer on it,  
    cost: 469.5,  
    payment_method: Bank Simple VISA,  
    customer: {  
        name: Francis Woodbury,  
        age: 28,  
        personality: sullen,  
    },  
    store: {  
        name: Yupster Things,  
        city: San Francisco,  
        state: California,  
        address: 467 West Portal Ave,  
    }  
};
```

Wednesday, February 13, 13

provide rich contextual information with every event.
now you can count the number of purchases by people under the age of 21. In a single query.
you can calculate the average purchase amount for businesses within 10miles of SF.
you can graph the number purchases per house for a given store or state.

Action

Time

purchase = {
 timestamp: 2012-06-06T19:10:39.205000,
 item: sophisticated orange turtleneck with deer on it,
 cost: 469.5,
 payment_method: Bank Simple VISA,
 customer: {
 name: Francis Woodbury,
 age: 28,
 personality: sullen,
 },
 store: {
 name: Yupster Things,
 city: San Francisco,
 state: California,
 address: 467 West Portal Ave,
 }
};

State
(nouns)

Wednesday, February 13, 13

provide rich contextual information with every event.
now you can count the number of purchases by people under the age of 21. In a single query.
you can calculate the average purchase amount for businesses within 10miles of SF.
you can graph the number purchases per house for a given store or state.

Rules of thumb for event data

- 5-10 nouns are important in your app
- 5-10 verbs are important in your app
- Nouns and their properties are your data model.
- Verbs are the actions you should record.
- Try to record maximum information about nouns with every event.

Examples

- Nouns: User, Company, Organization, Team, Platform, Device, App, Level, Garden, Favorites, Interests, Inventory, Cart, Location, Item, Product, Account, Form, Picture, Story
- Verbs: Login, Upgrade, Submit Content, Scroll, Send, Share, Search, Check-In, Vote, Purchase, Level Up, Favorite, Vote, Crash, Rate, Start, Modify, Check, View, Capture

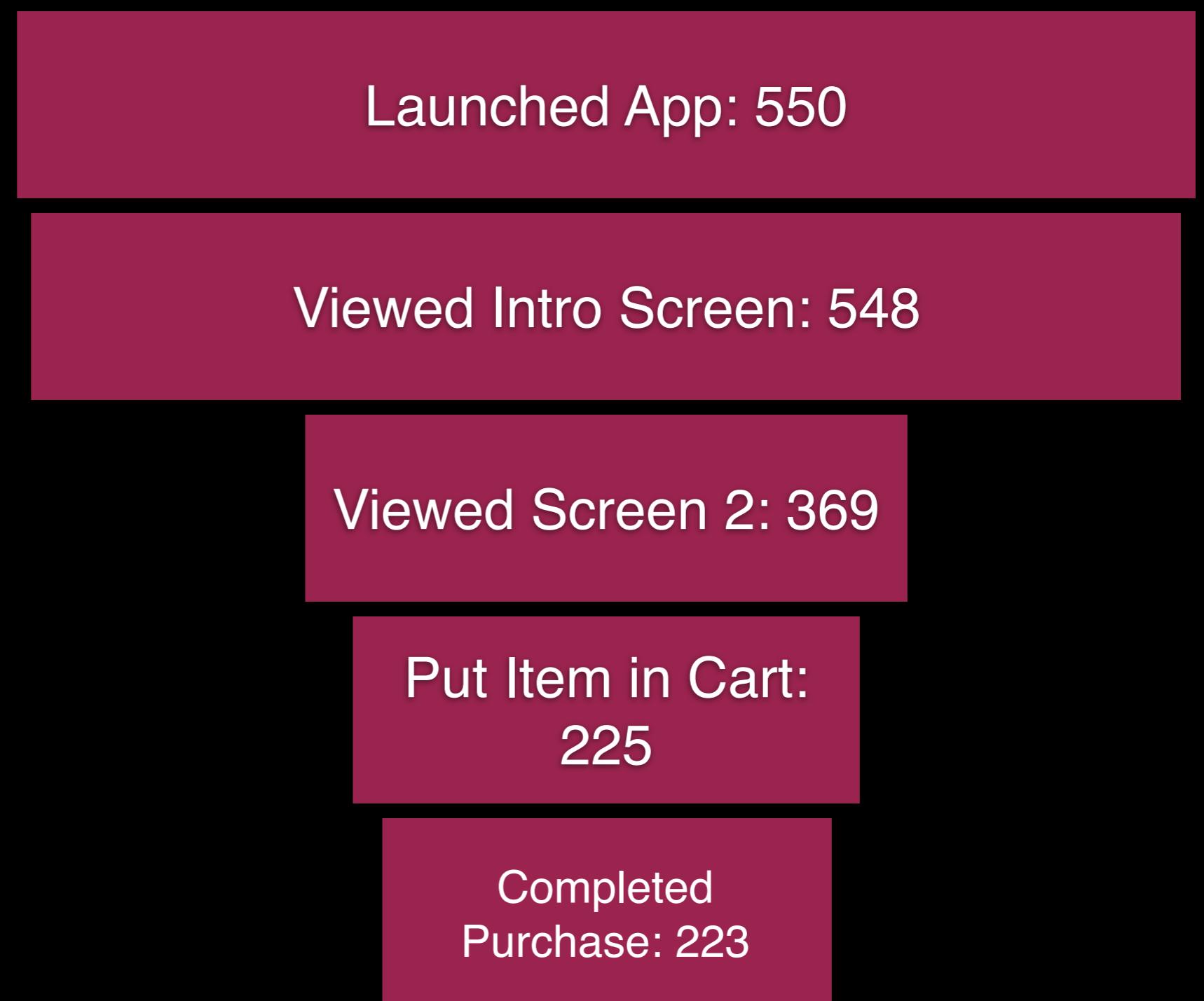
Exercise I

Pair with the person next to you.

1. Pick 1 event (action) in your app and create a data model for it. What nouns and properties can you record for this event?
2. Name something in your app you would like to A/B test.

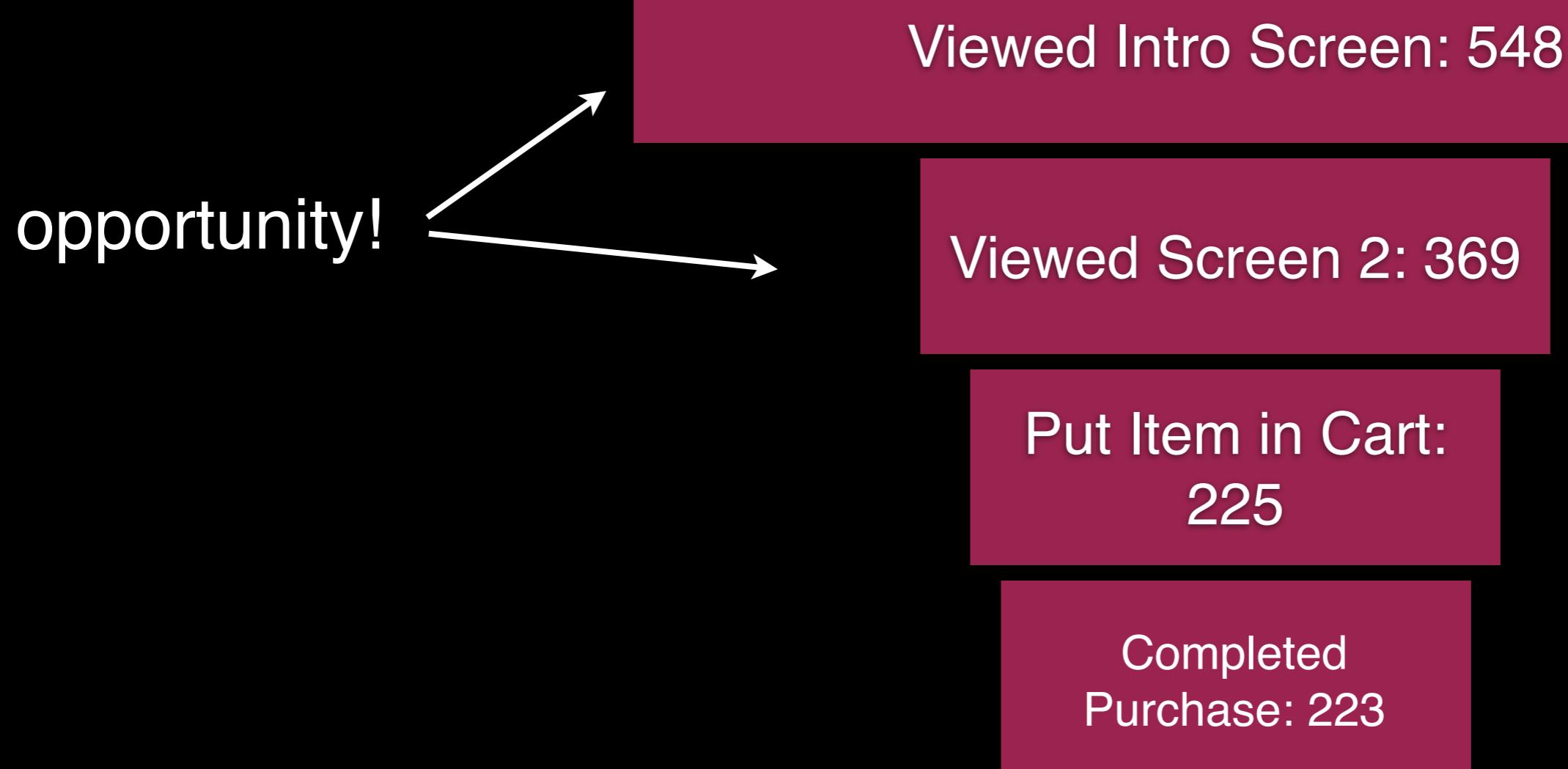
Prepare to share your results with the class.

Funnels

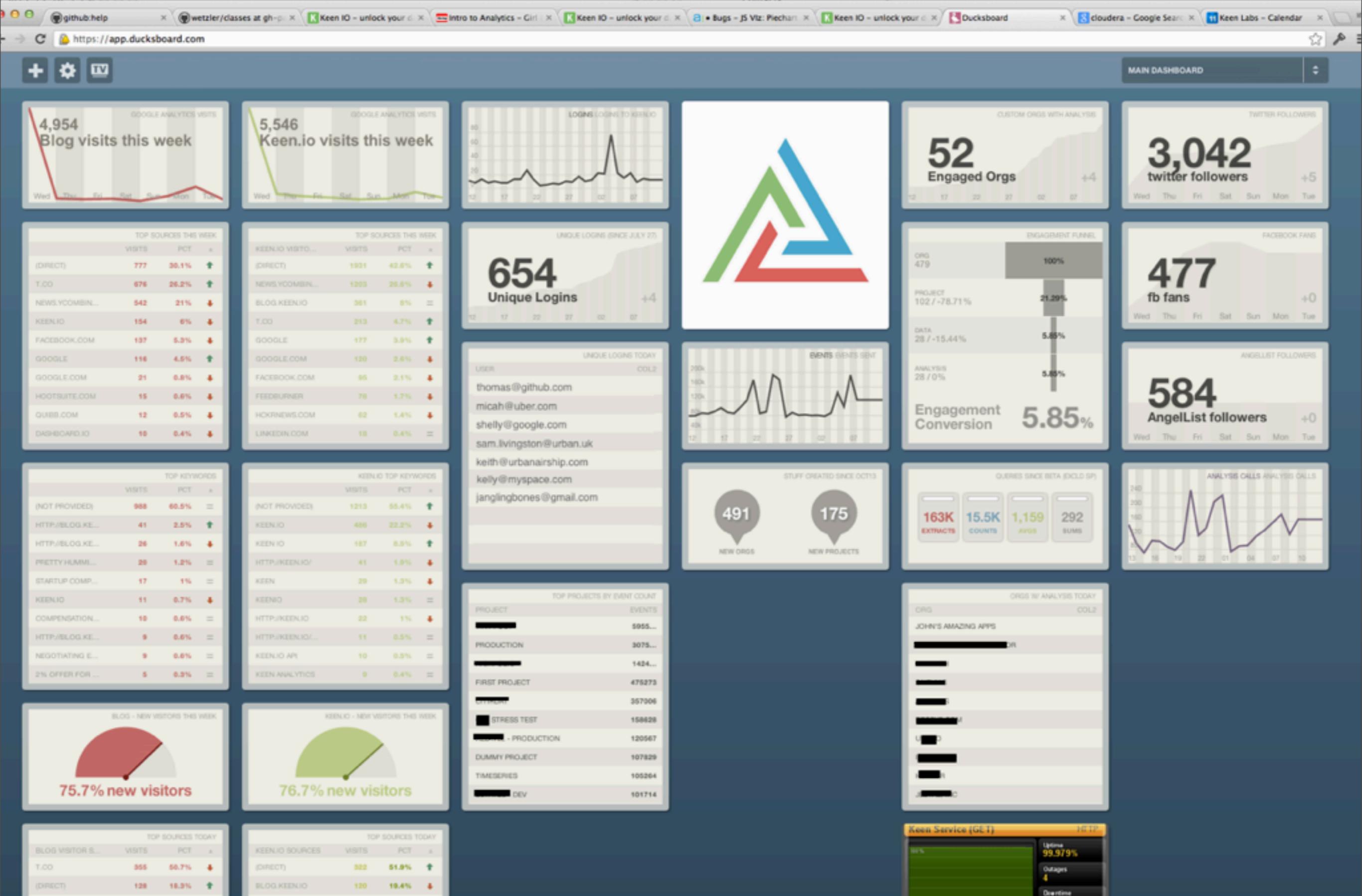


Wednesday, February 13, 13

Funnels



Wednesday, February 13, 13



Wednesday, February 13, 13

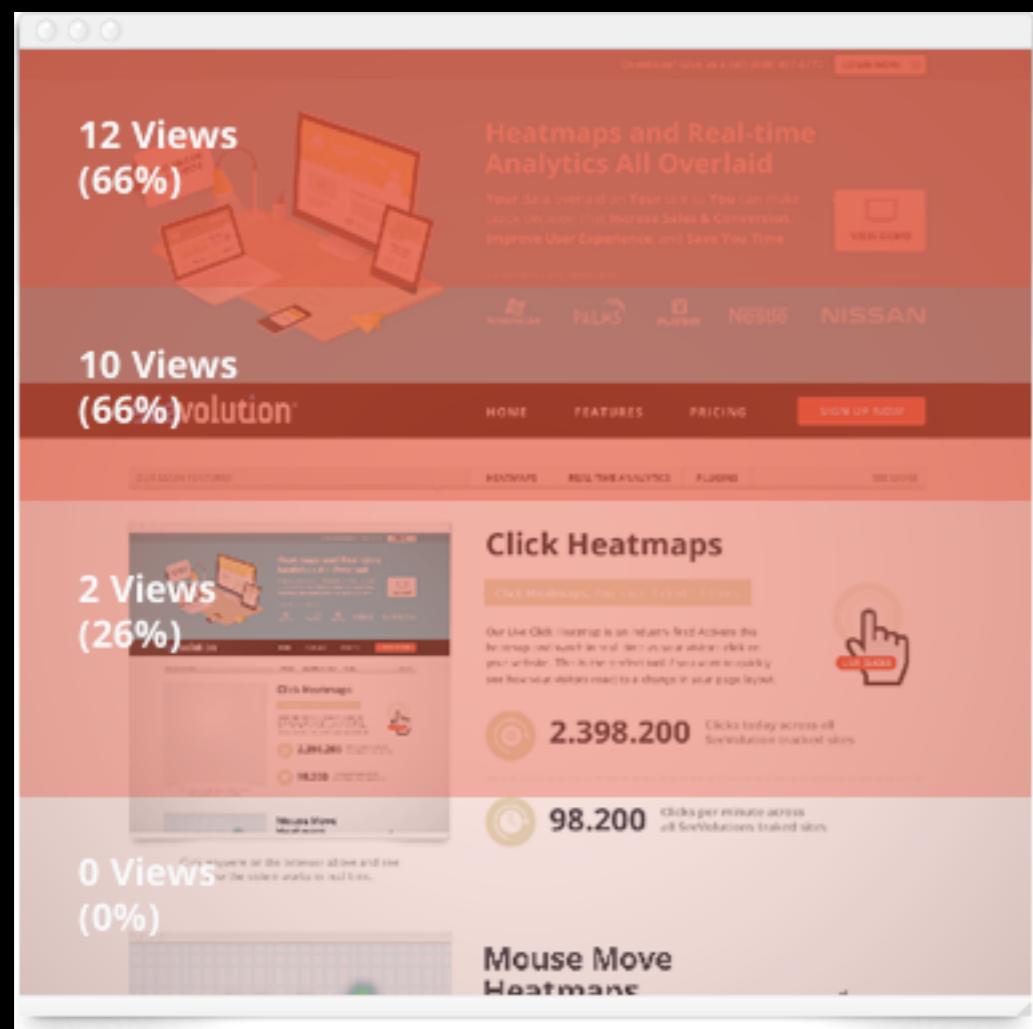
Use a dashboard to share analytics and motivate your team

New Cool Stuff

What are people highlighting? <http://markerly.com/>

Where are people looking? <http://www.crazyegg.com/>

Where are people mousing? <https://www.seevolution.com/>



Wednesday, February 13, 13

Cohort Analysis

A cohort is a group of people who share a common characteristic over a certain period of time.

	Engagement over time											
	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12
(Joined in) January	100%	20%	19%	13%	13%	10%	12%	11%	7%	7%	7%	?
February	100%	21%	16%	13%	11%	9%	9%	7%	7%	7%	7%	?
March	100%	24%	20%	17%	15%	13%	11%	10%	10%	10%	?	
April	100%	31%	27%	24%	19%	15%	12%	12%	?			
May	100%	31%	27%	25%	21%	18%	16%	?				
June	100%	39%	28%	24%	20%	19%	?					
July	100%	40%	33%	27%	23%	?						
August	100%	47%	41%	32%	?							
September	100%	52%	43%	?								
October	100%	53%	?									
November	100%	?										
December	?											

In this case engagement is improving nicely.
Of the January cohort, only 20% were engaged in month 2.
Of the October cohort, 53% were engaged in month 2.

<http://redeye.firstround.com/2008/01/after-the-teche.html>

Wednesday, February 13, 13

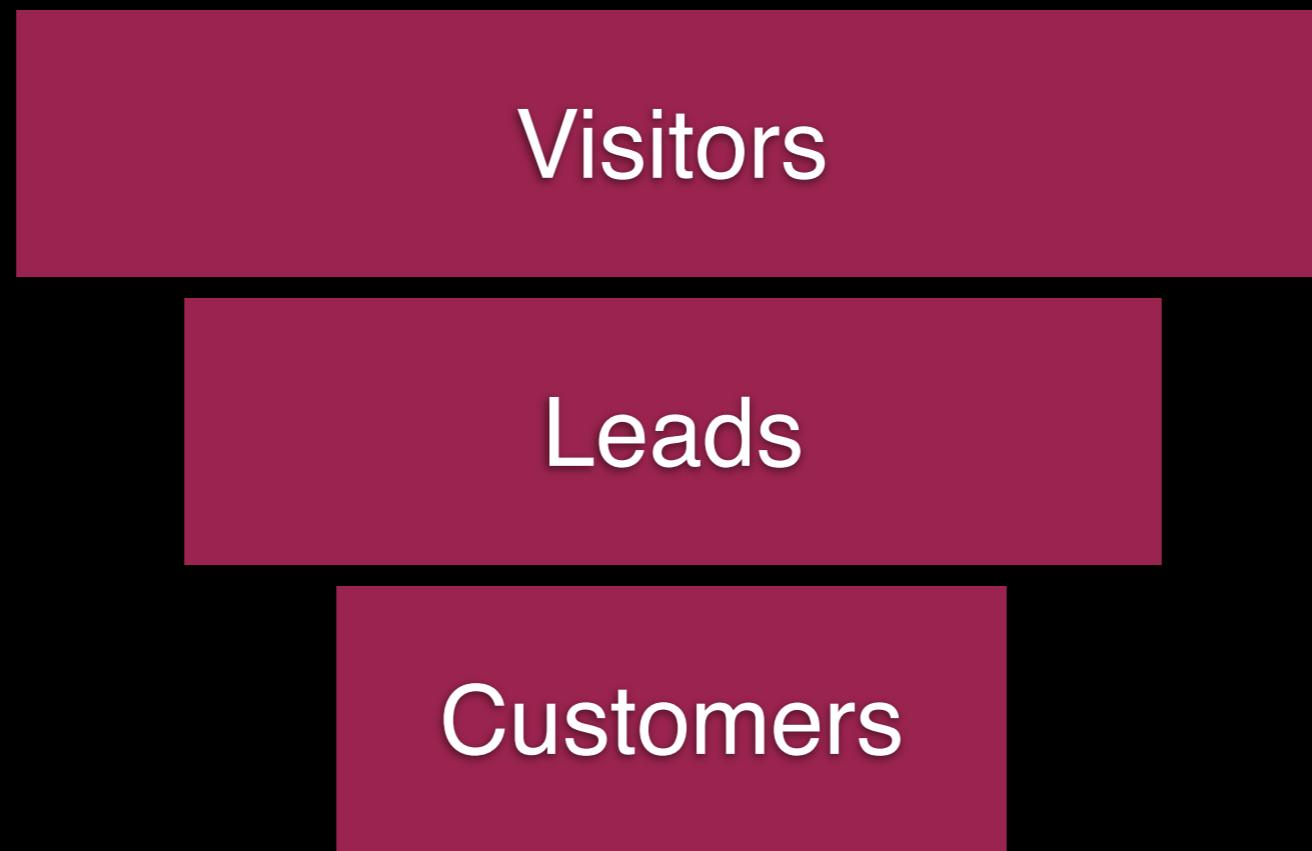
<http://www.cohortanalysis.com/>

<http://jonathonbalogh.com/2012/03/24/introduction-to-cohort-analysis-for-startups/>

<http://52weeksofux.com/post/646711369/cohort-analysis-measuring-engagement-over-time>

Next: analytics for marketing & sales

Conversion Funnel



Wednesday, February 13, 13

You can use analytics to improve conversion throughout the funnel.
So far we've talked about how to make your app better for people who have made it to your app.
You can also use analytics to increase the traffic to your app or site.

where are people coming from? use google analytics “sources” feature

Source / Medium	Visits
1. news.ycombinator.com / referral	5,985
2. (direct) / (none)	3,410
3. google / organic	959
4. t.co / referral	854
5. facebook.com / referral	828
6. keen.io / referral	814
7. google.com / referral	424
8. hckrnews.com / referral	170
9. Newsletter / Email	166
10. news.ycombinator.org / referral	132

Wednesday, February 13, 13

“dark social” = email and IM sharing

t.co = twitter

<https://pi.pardot.com>

Pardot Marketing Prospects Reports Admin search Help michelle@keen.io

Dashboard

All Prospects

Date Range: Last 7 days

Prospects Created	15
Conversions	0
All Prospects	1

Active Prospects for Review

Name	Company
No new prospects	

[More Prospects »](#)

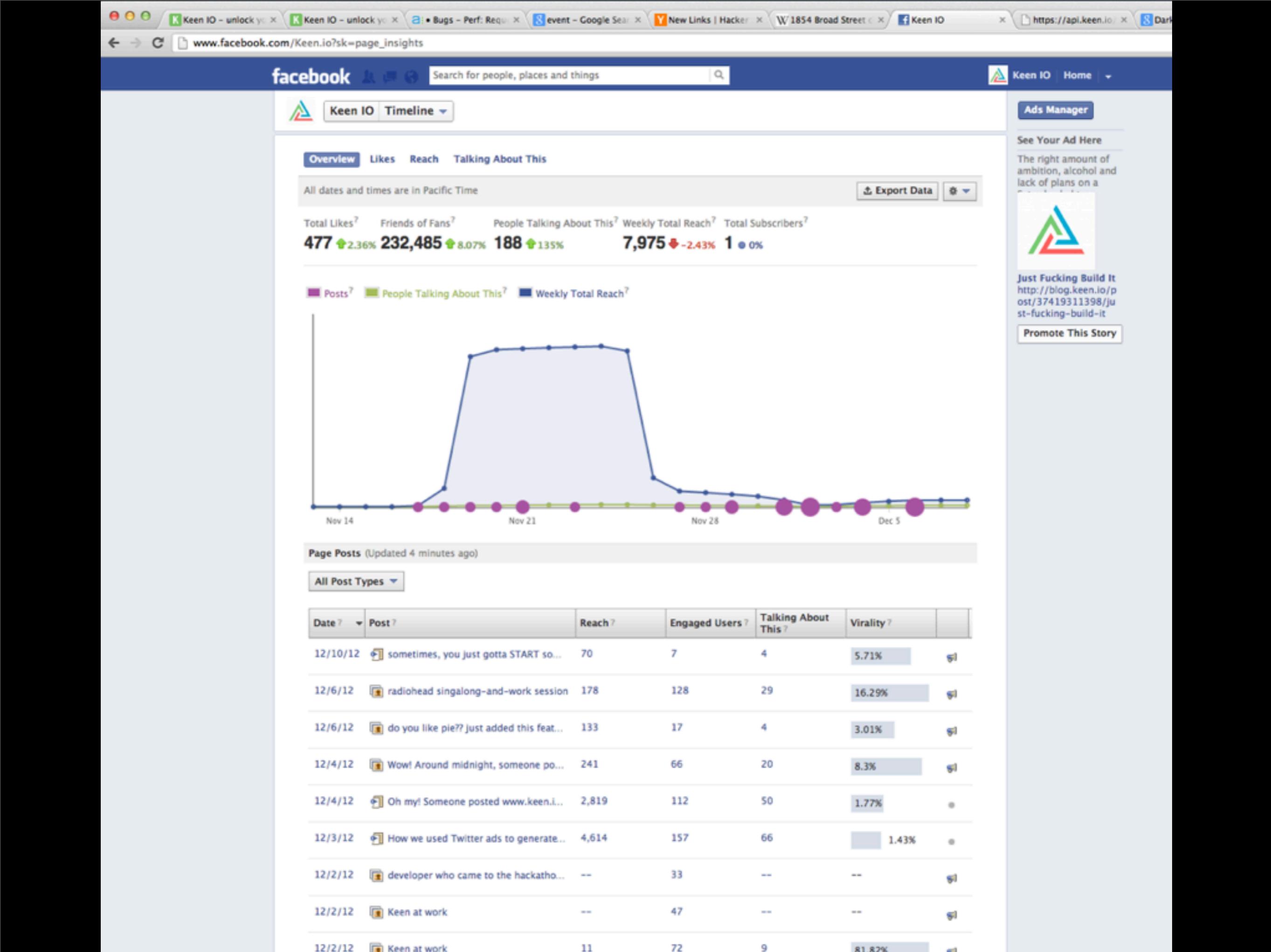
Identified Companies

Company	Location
Three	Chesterfield, Derbyshire, United Kingdom, threemb.co.uk
DELOITTE TOUCHE TOHMATSU	Melbourne, Victoria, Australia, reach.com
Red Hat	Raleigh, NC, United States, redhat.com
Turner Broadcasting Sy...	Santa Monica, CA, United States
USF DSLAM South	Lahore, Punjab, Pakistan

[More Visitors »](#)

Wednesday, February 13, 13

pardot tracks visitors to your site and organizes your leads so you can do drip campaigns



Wednesday, February 13, 13

social media sites have AMAZING dashboards for measuring the effectiveness of your efforts.
See also: Google Ads, Twitter

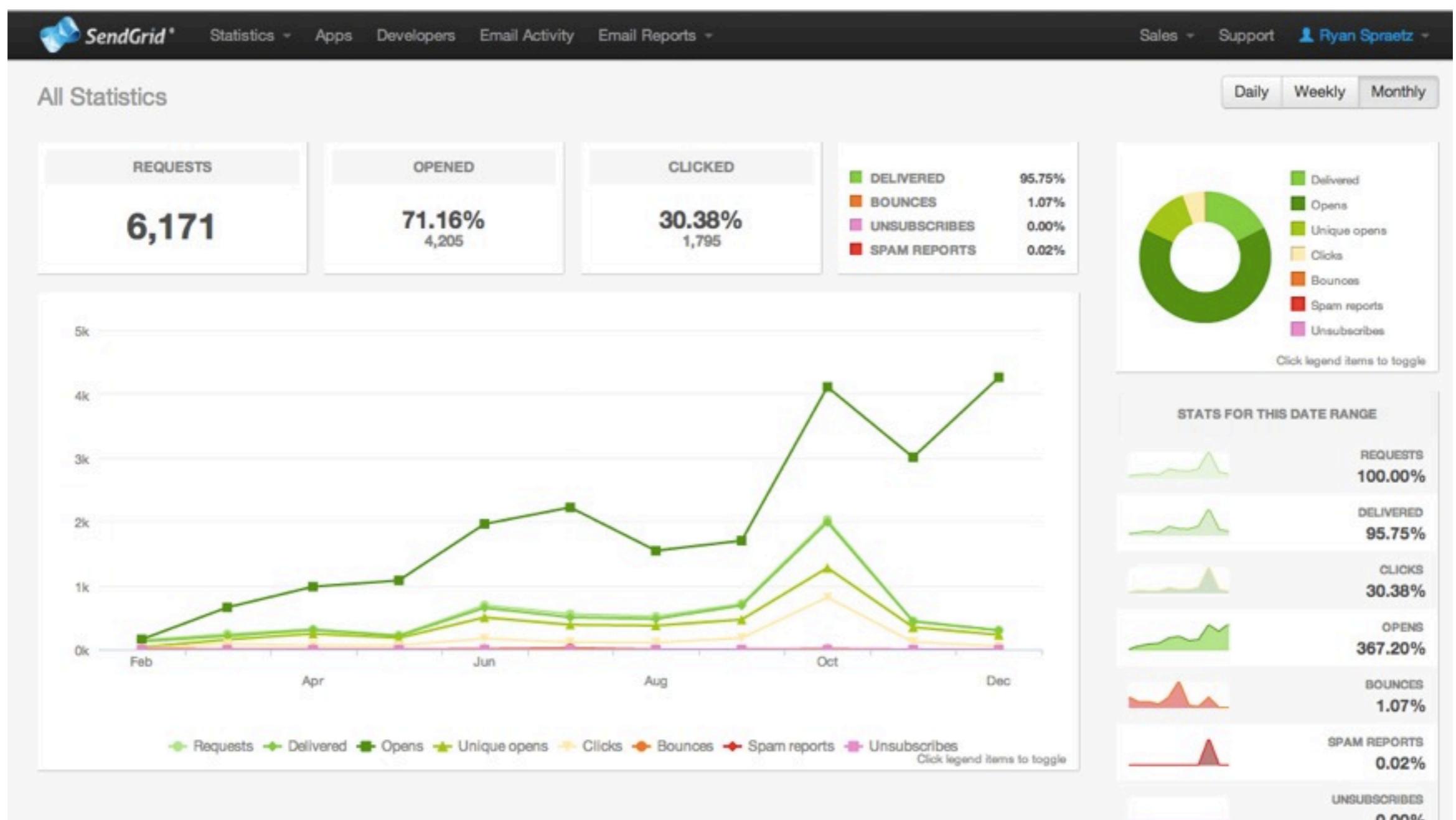


Send emails via web or APIs

Receive emails via API

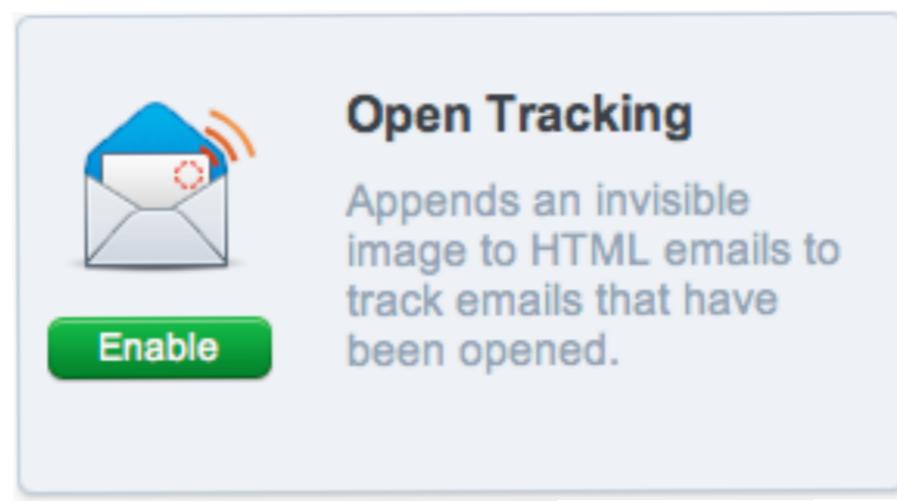
Track email analytics via web or APIs

SendGrid Email Analytics

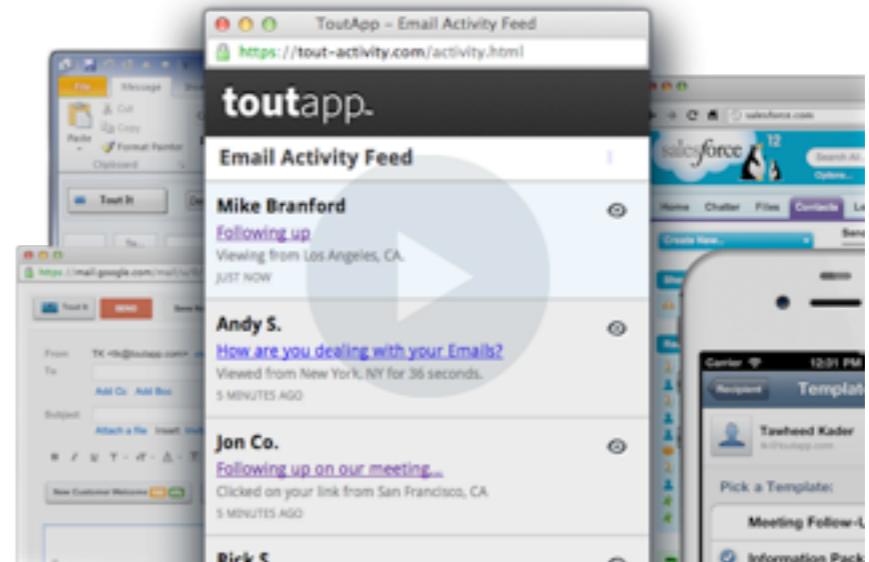


Wednesday, February 13, 13

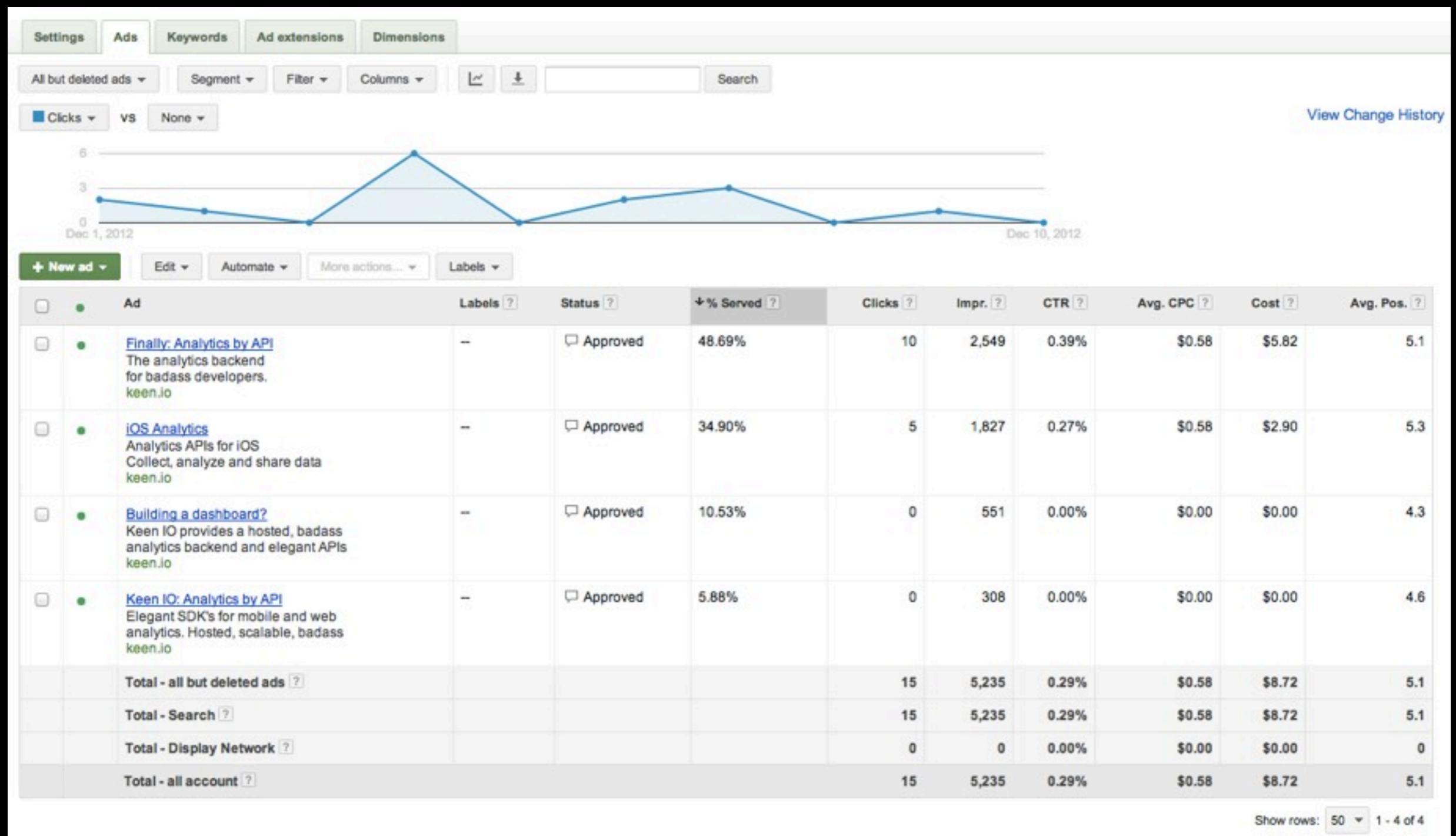
Email Tracking



...used by ToutApp:

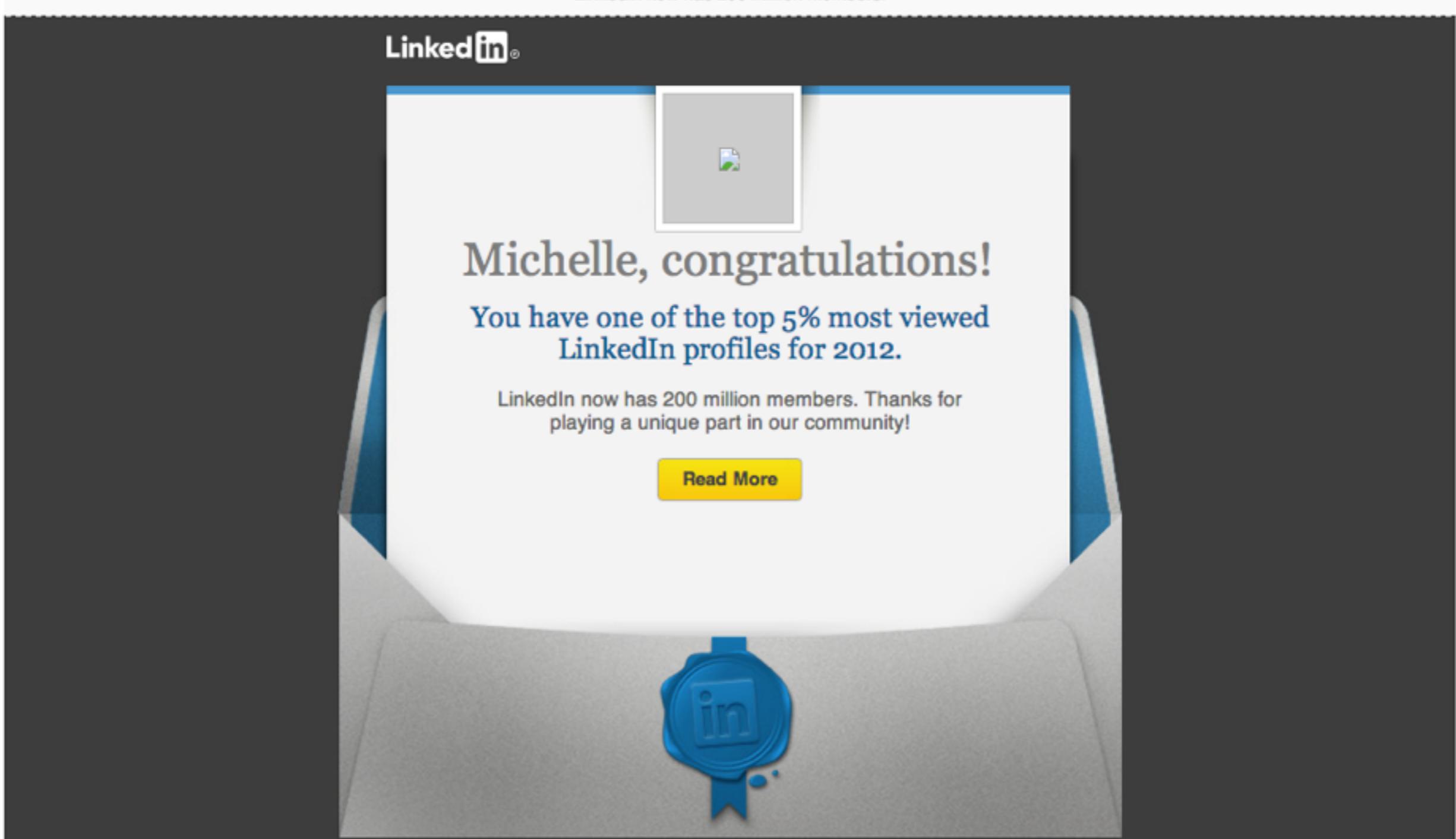


Google Adwords



Wednesday, February 13, 13

Google adwords is cool because it automatically chooses ads and keywords for you.



This is an occasional email to help you get the most of LinkedIn. [Unsubscribe](#)
This email was intended for Michelle Wetzler (Director at Keen IO). [Learn why we include this](#).
If you need assistance or have questions, please contact [LinkedIn Customer Service](#).
© 2013, LinkedIn Corporation. 2029 Stierlin Ct., Mountain View, CA 94043, USA.

Wednesday, February 13, 13

Exercise Time!

Wednesday, February 13, 13

Conversion Rate - Percentage of people who made it from one part of the funnel to another part of the funnel

MAU - Unique monthly active users. Common in gaming and consumer apps (Facebook has 1B MAU).

Segmentation - Dividing things into groups for analysis (e.g. Average purchase amount segmented by age group)

CPC - Cost per click. This is how ads are priced on twitter, facebook, and google.