

BIG DATA, INDUSTRY AND DISEASE SURVEILLANCE

Alessio Signorini

alessio-signorini@uiowa.edu

@a_signorini

How to Make Millions with Big Data

Alessio Signorini
@a_signorini



My Data

is Bigger

than Yours



Alessio Signorini

@a_signorini

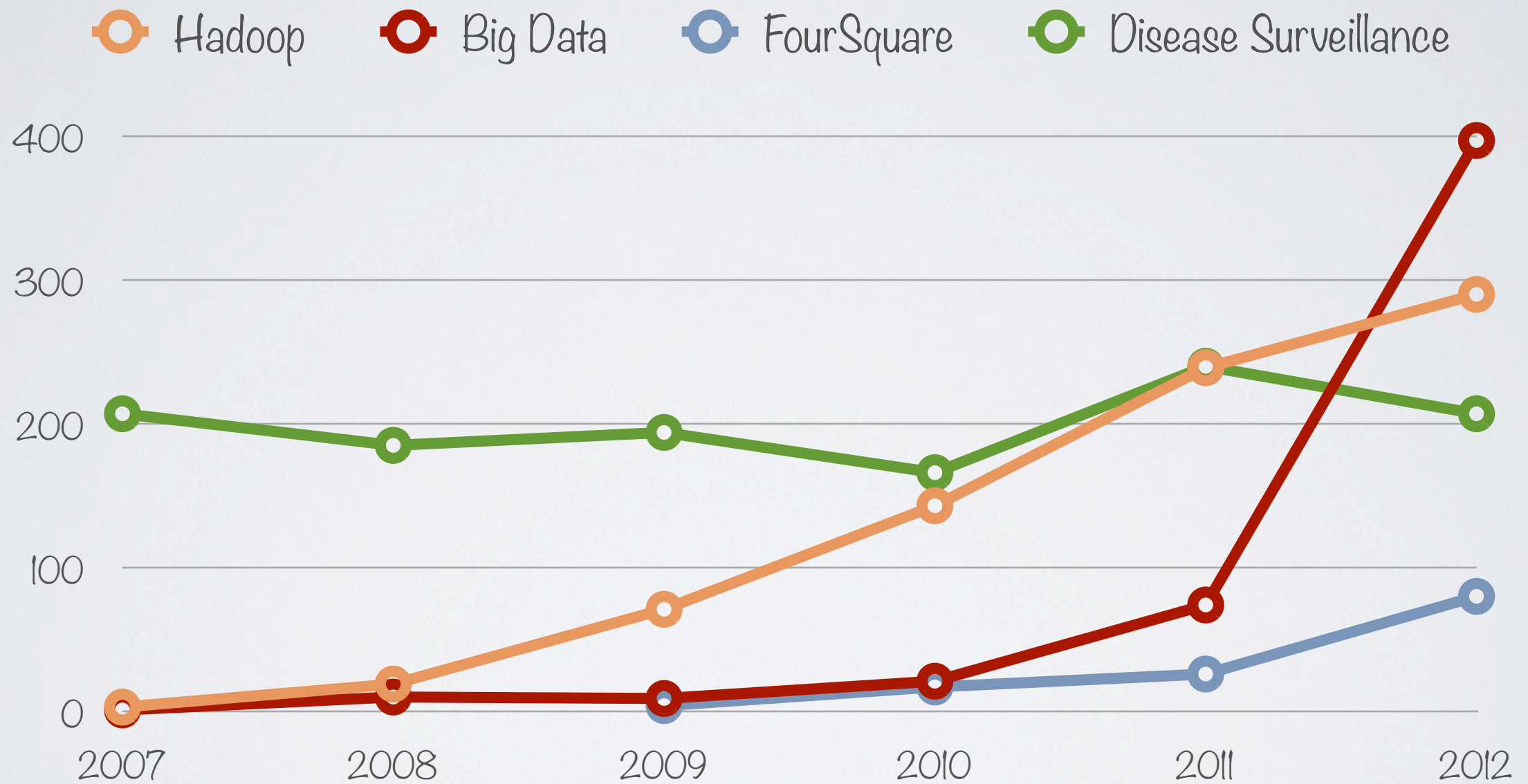
BIG DATA, INDUSTRY AND DISEASE SURVEILLANCE

Alessio Signorini


alessio-signorini@uiowa.edu


@a_signorini

Number of Articles with Keyword in Title



Google Scholar + python



allintitle: "big data" 

Scholar About 555 results (0.08 sec)


Articles

Legal documents

Any time
Since 2013
Since 2012
Since 2009
Custom range...

Sort by relevance
Sort by date

☐ include patents
☐ include citations

 Create alert

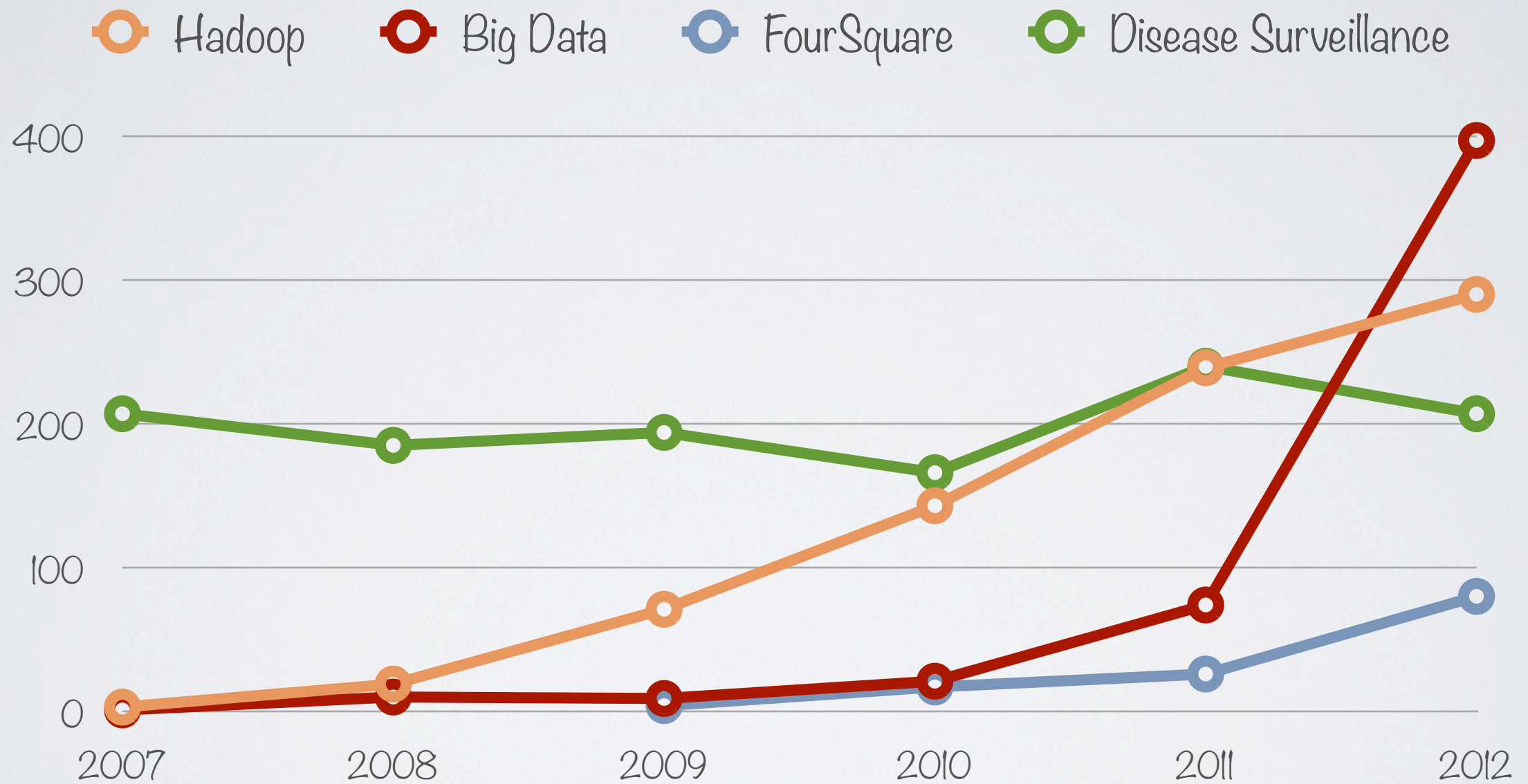
Big data: The future of biocuration
D Howe, M Costanzo, P Fey, T Gojobori, L Hannick... - Nature, 2008 - nature.com
We propose three urgent actions to advance this key field. First, authors, journals and curators should immediately begin to work together to facilitate the exchange of data between journal publications and databases. Second, in the next five years, curators, researchers and ...
Cited by 145 Related articles All 17 versions Cite

Big data: How do your data grow?
C Lynch - Nature, 2008 - nature.com
Data can be 'big' in different ways. National and international projects such as the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva in Switzerland, or the Large Synoptic Survey Telescope planned for northern Chile, are frequently cited for ...
Cited by 63 Related articles All 7 versions Cite

[PDF] **Big data: wikiomics**
M Waldrop - Nature, 2008 - wiki.myexperiment.org
Alexander Pico remembers just when the idea hit him. In January 2007, he and his boss, Bruce Conklin, were discussing how to push their software tool for visualizing intracellular signalling pathways to the next level of interactivity—when Pico blurted out, "What we ...
Cited by 60 Related articles All 5 versions Cite More▼

MAD skills: new analysis practices for big data
J Cohen, B Dolan, M Dunlap, JM Hellerstein... - Proceedings of the ..., 2009 - dl.acm.org
Abstract As massive data acquisition and storage becomes increasingly affordable, a wide

Number of Articles with Keyword in Title



17,000+

“Big Data” companies

The logo for SAILTHRU, featuring the word "SAILTHRU" in a bold, blue, sans-serif font.

\$28M



\$23M



\$2M



\$4M

\$80 Million

raised in last 30 days



Let data speak

\$70K



\$2M



\$5M



\$3M



\$100K

The logo for AYASDI, featuring the word "AYASDI" in a bold, black, sans-serif font.

\$10M

CrunchBase + Perl 5

CrunchBase

[Advanced Search](#)[Register](#) | [Login](#) or  [Login](#)

Home

What is CrunchBase?

CrunchBase is the free database of technology companies, people, and investors that anyone can edit.

Here, you can learn and edit everything about companies like Facebook, YouTube, Twitter, MySpace and Tagged, products like Droid and Google Wave, and people like Mark Zuckerberg and Steve Jobs.

CrunchBase Stats

Companies	113,886
People	148,269
Financial Organizations	8,872
Service Providers	5,322
Funding Rounds	33,207
Acquisitions	7,615

CrunchBase News

[Crunching CrunchBase](#)

Hot on CrunchBase

Trending Now

Popular Pages

In The News

real-time data

Facebook



Xoom Corporation



GitHub



Virool



Marin Software



Everest



Airbnb



Contribute

[Add Company](#)
[Add Financial Organization](#)
[Add Person](#)

Latest Milestones Added

 **Battery Park City Authority**
— Company added to CrunchBase (2/15/13)
Posted 2/15/13 at 12:25pm

 **Mattone Group Ltd.**
— Company added to CrunchBase (2/15/13)
Posted 2/15/13 at 12:25pm

 **Tradebound International**
— Company added to CrunchBase (2/15/13)
Posted 2/15/13 at 11:54am

 **VAULT29**
— Company added to CrunchBase (2/15/13)
Posted 2/15/13 at 11:54am

 **Fastvibe/ Telus**

The logo for SAILTHRU, featuring the word "SAILTHRU" in a bold, blue, sans-serif font.

\$28M



\$23M



\$2M



\$4M



Let data speak

\$70K

\$80 Million

raised in last 30 days



\$2M



\$5M



\$3M



\$100K

The logo for AYASDI, featuring the word "AYASDI" in a bold, black, sans-serif font, set against a white rectangular background.

\$10M





THE UNIVERSITY
OF IOWA

PhD Candidate



Web Search Engine



Social Search Engine



Digital Advertising



Health Reward Platform



Data is Everywhere



I am only at Slide #10
and lots of you
are checking their phone



Bill Gates ✓
@BillGates

Follow

@a_signorini talk at #IMED2013 is awesome! If you aren't at the conference today, you are really missing out...
bit.ly/WpSyyh

Reply Retweet Favorite More

416
RETWEETS

269
FAVORITES



11:10 AM - 18 Feb 2013

Twitter



Twitter



Facebook/Email

"Promiscuous" Mode

The screenshot displays the Colasoft Capsa 7.1 software interface, titled "Analysis Project 1 - Full Analysis - Colasoft Capsa 7.1". The interface is divided into several sections:

- Top Bar:** Contains tabs for Analysis, System, Tools, and View. Below these are icons for Adapter, Filter, Start, and Stop, along with a "Capture" button.
- Left Panel:** Includes a "Network Profile" section with options like Network Group, Name Table, and Alarm Settings. It also has a "Packet Storage" section with Log Settings and Diagnosis Settings.
- Right Panel:** Features two gauges for "Utilization (0%)" and "pps (1)", a "Traffic History Chart(...)", and a "Packet Buffer (16 MB)" indicator.
- Bottom Section:** Contains a "Diagnosis" tab with sub-tabs for Protocol, Physical Endpoint, IP Endpoint, Physical Conversation, IP Conversation, TCP Conversation, UDP Conversation, Matrix, Packet, Log, and Report. The "TCP Conversation" tab is currently selected, showing a table of network traffic.

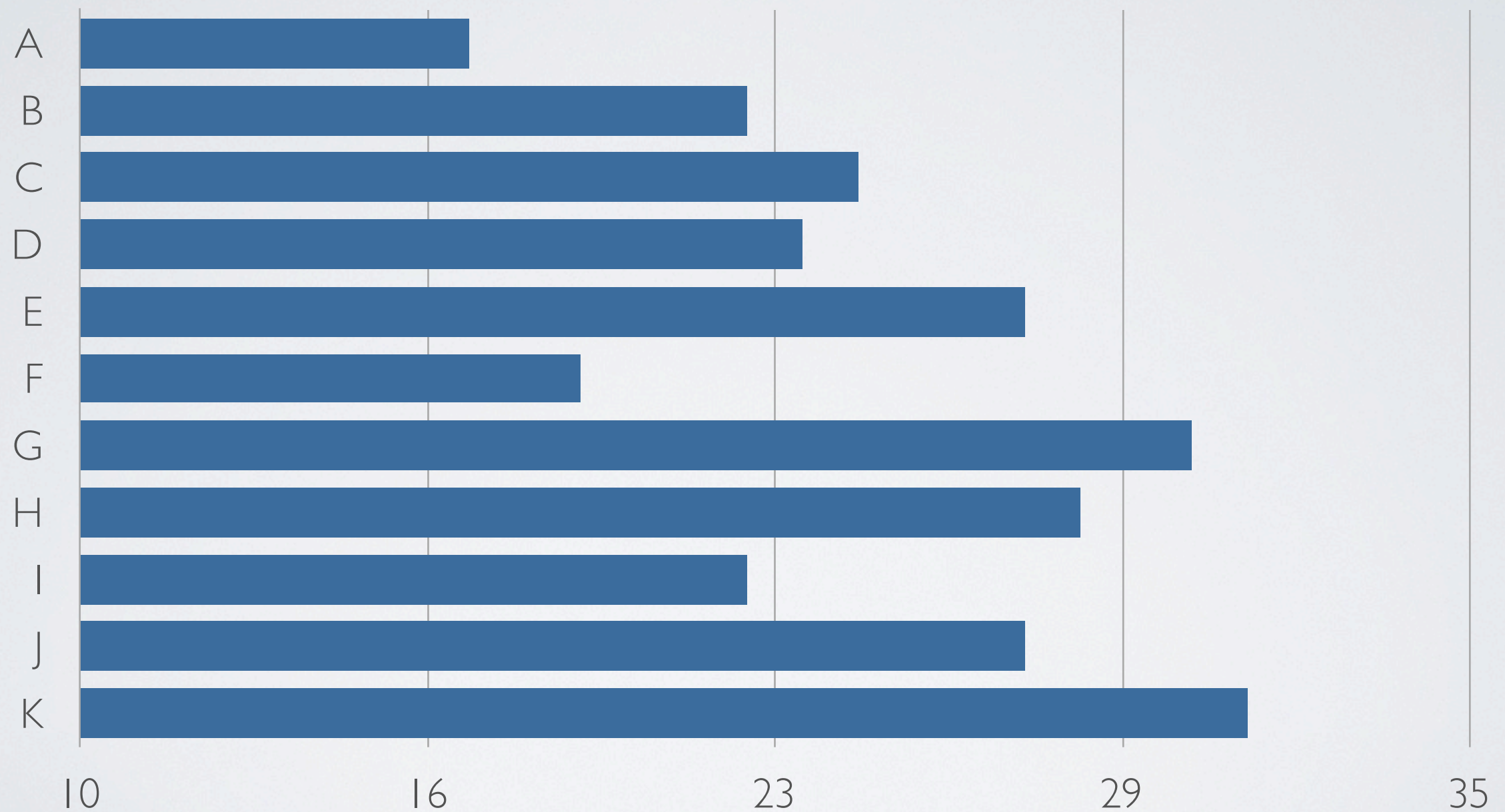
The "TCP Conversation" table displays the following data:

Endpoint 1 ->	<- Endpoint 2	Duration	Bytes	Bytes ->	<- Bytes	Packets	Packets ->	<- Packet
192.168.6.40:3655	FANC-PC:139	00:00:00	5.612 KB	2.729 KB	2.883 KB	33	17	1
FANC-PC:51738	74.53.97.226:995	00:00:04	3.565 KB	1.348 KB	2.218 KB	28	14	1
FANC-PC:51740	192.168.0.227:80	00:00:00	1.532 KB	1.064 KB	479 B	9	5	
192.168.6.39:2709	FANC-PC:139	00:00:00	3.354 KB	1.647 KB	1.707 KB	21	11	1
192.168.6.39:2710	FANC-PC:139	00:00:00	3.178 KB	1.647 KB	1.530 KB	21	11	1
FANC-PC:49268	192.168.0.211:445	03:00:10	62.355 MB	18.288 MB	44.068 MB	185.533	88.656	96.87

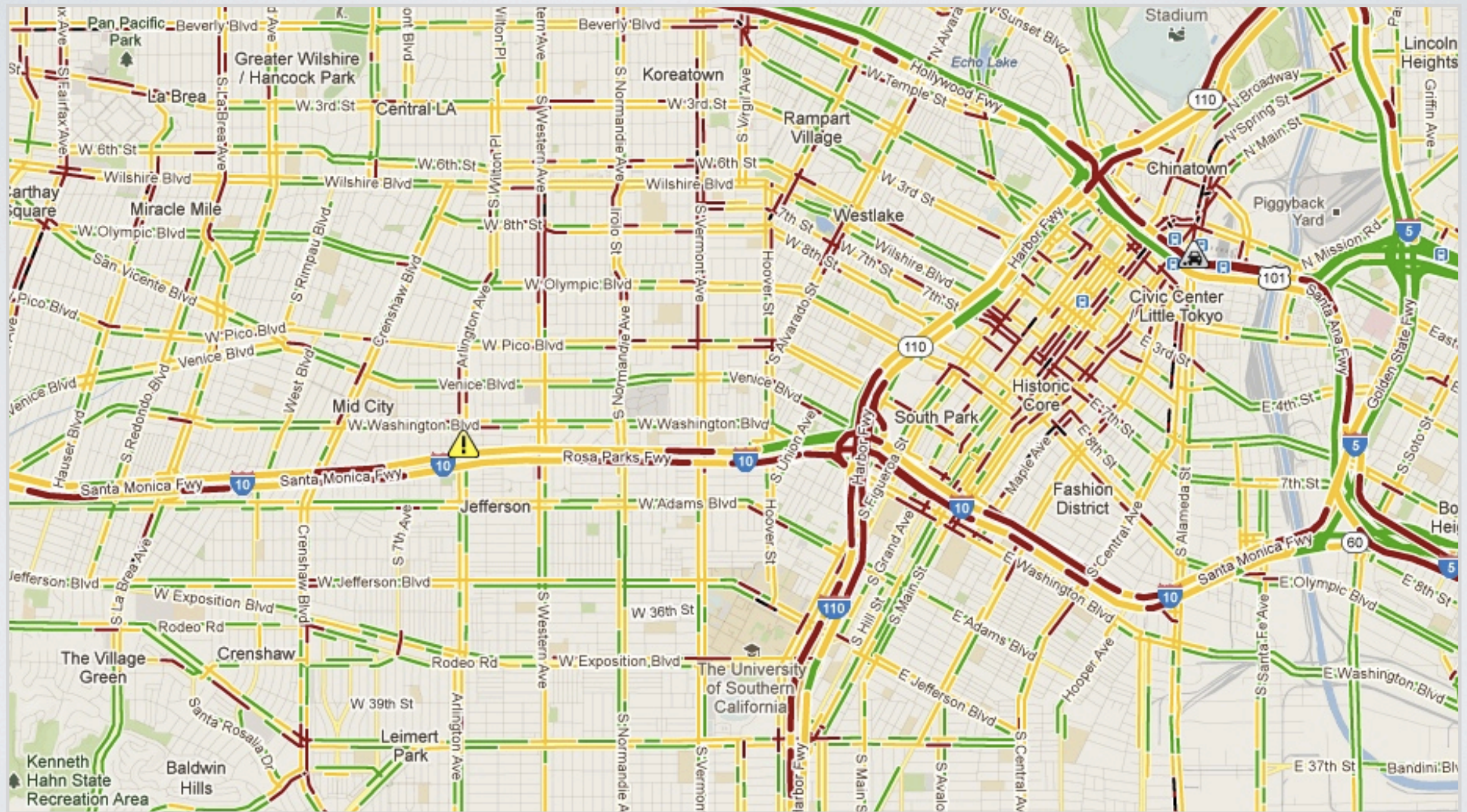
Below the table, the "Time Sequence" tab is selected, showing a detailed view of the network traffic between 192.168.6.7:49177 and 65.55.50.101:443. The sequence includes various TCP flags and window sizes, such as PSH, ACK, and Load Length.

The bottom status bar indicates the current capture status: "Capture - Full Analysis", "Local Area Connection", "Inactive", "Duration: 00:51:49", "5,459,500", "0", "Ready", and "Alarm Explorer".

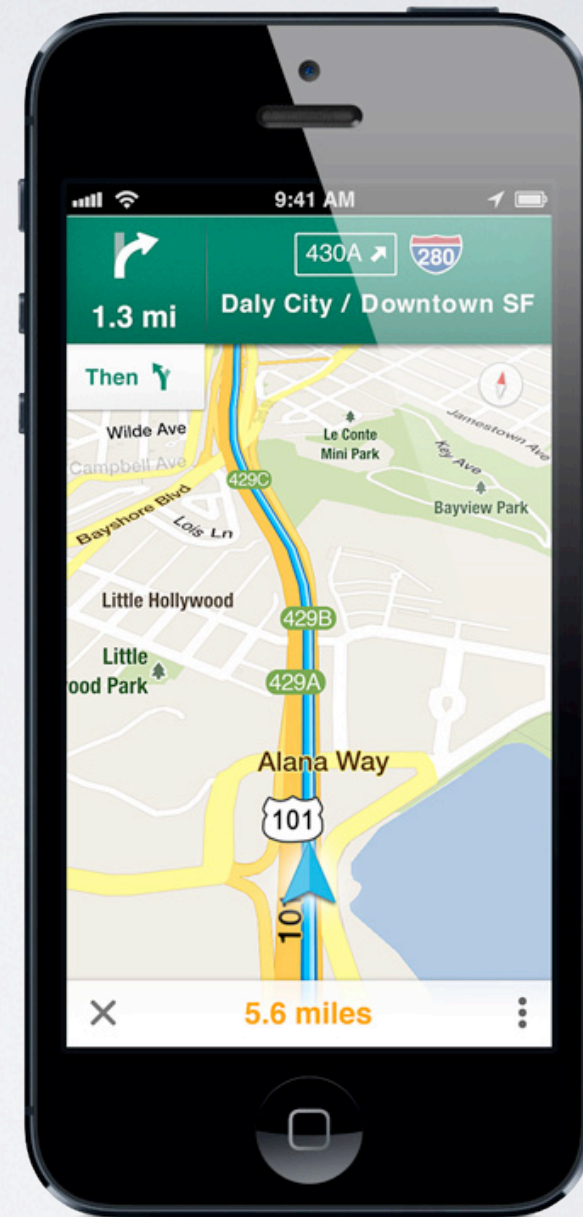
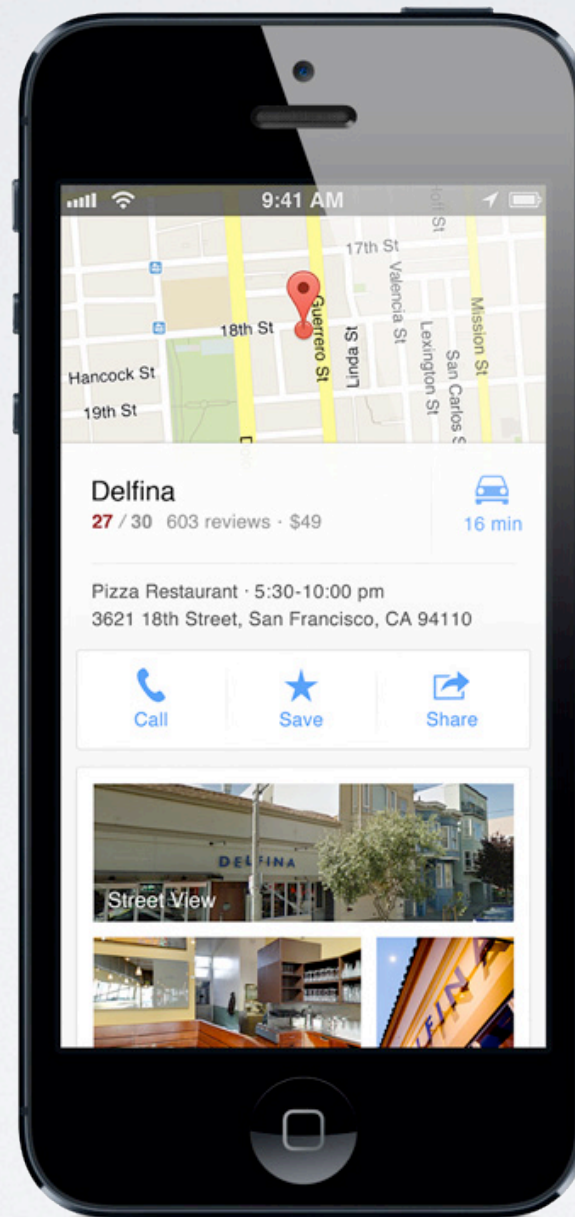
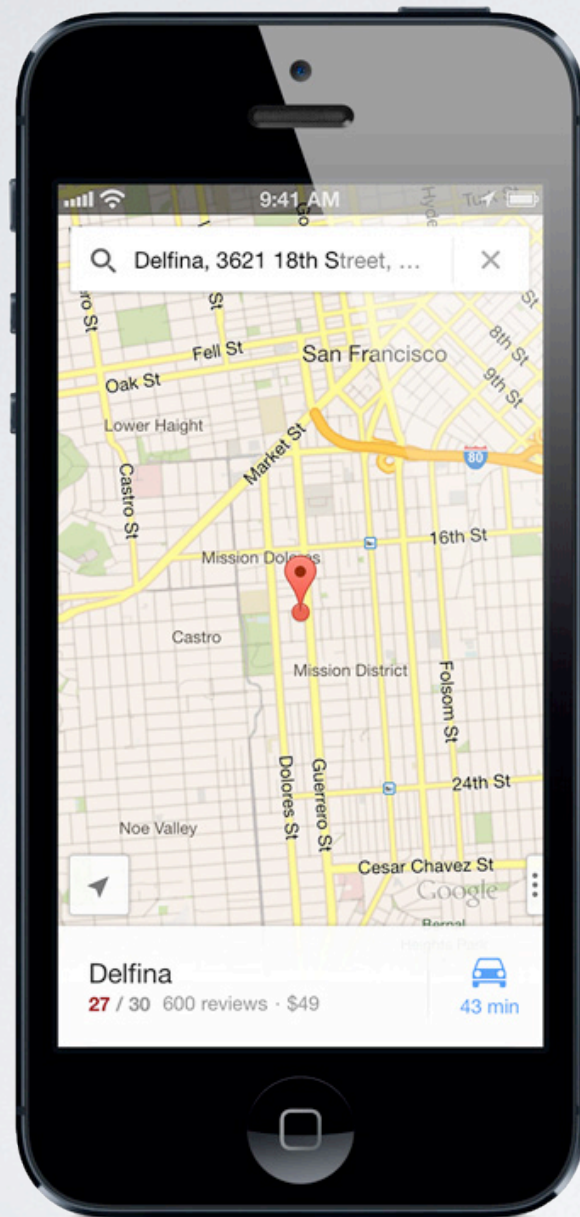
Number of Users on Wifi during Session 12



Google maps - Los Angeles Traffic

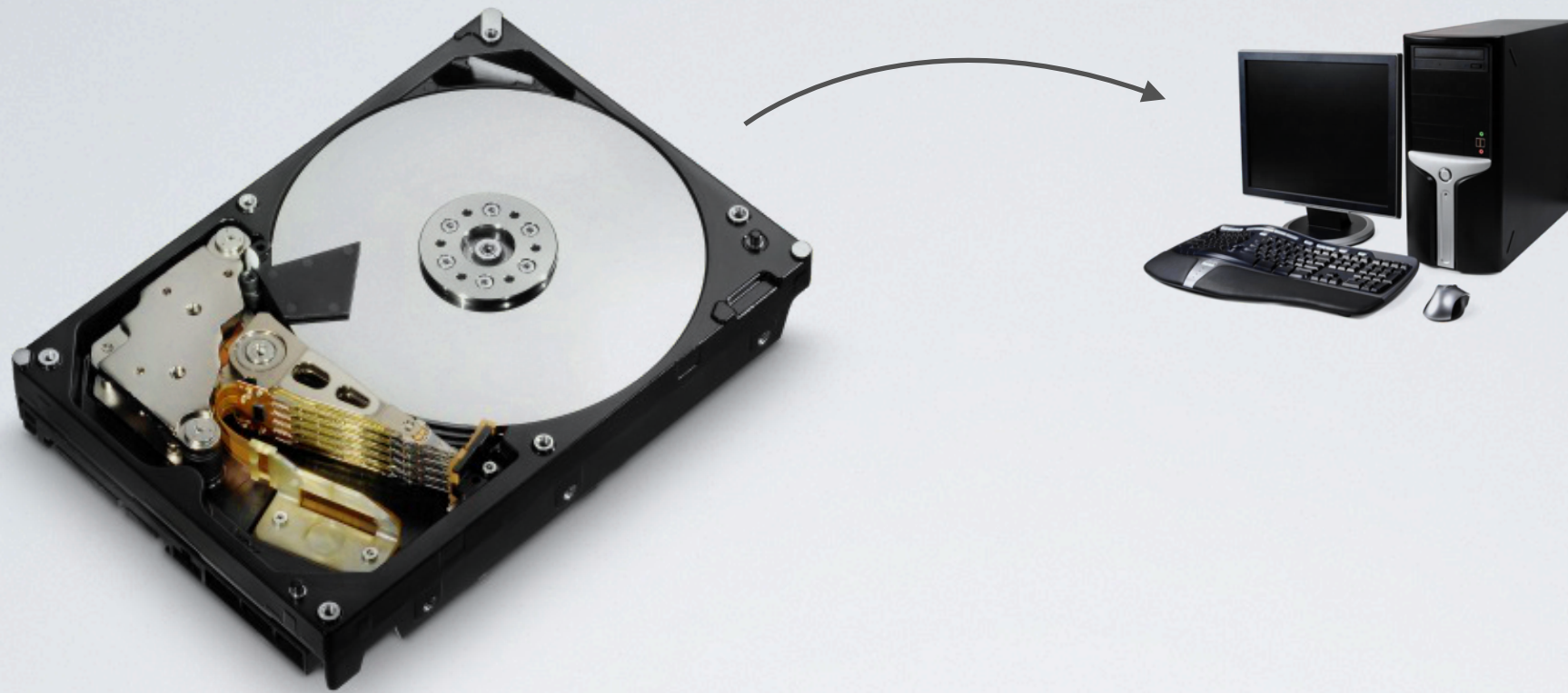


Google Mobile Apps sends back GPS data

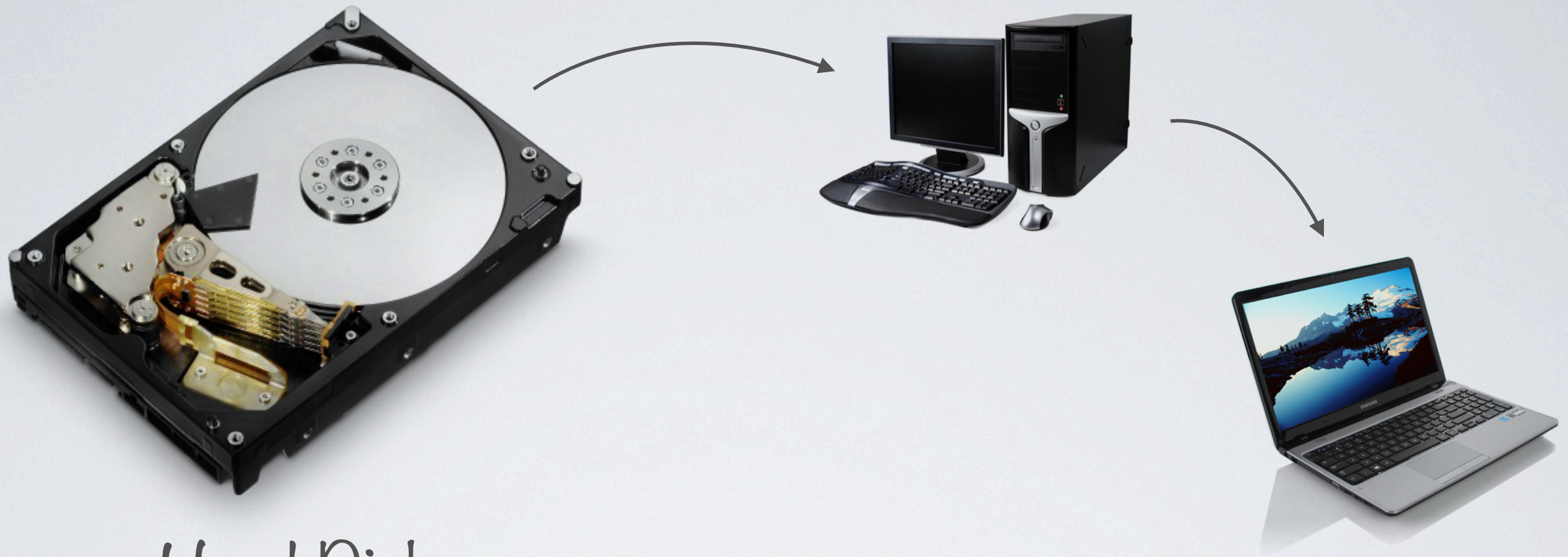




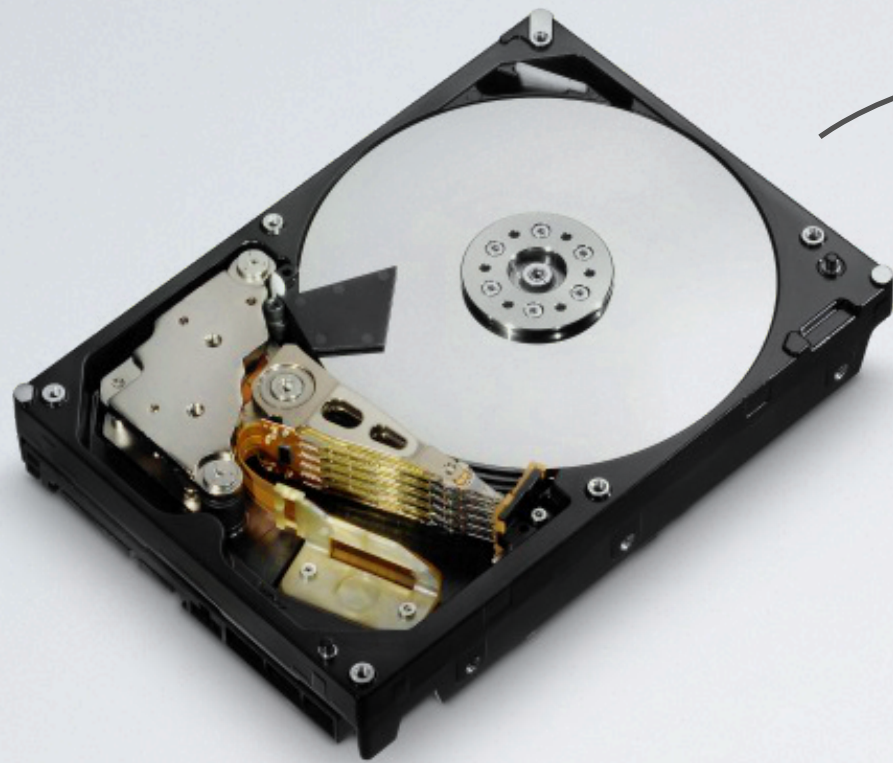
Hard Disk



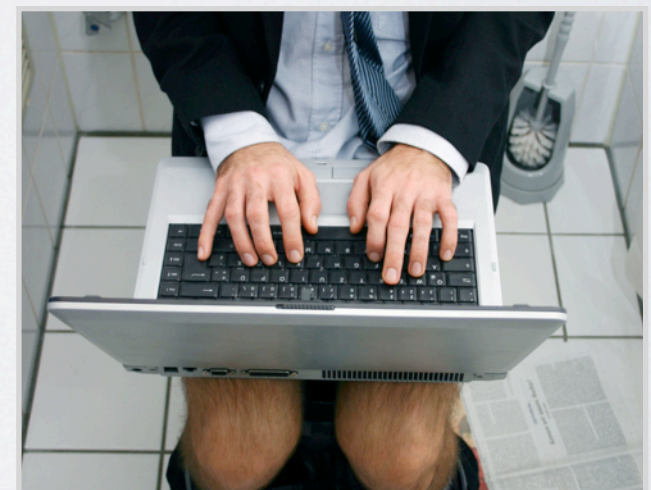
Hard Disk



Hard Disk



Hard Disk

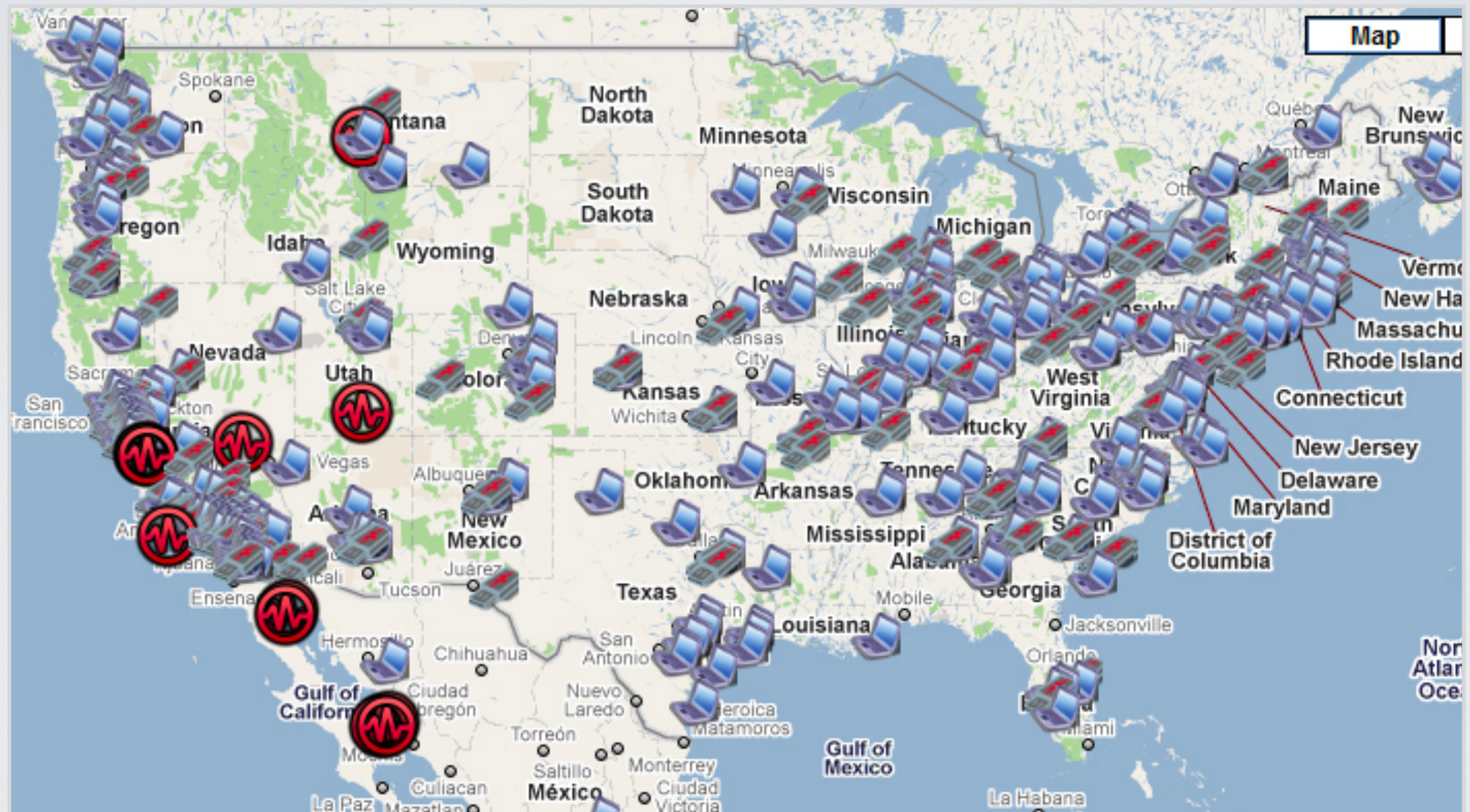




Hard Disk



Quake-Catcher Network - Stanford University





Data is Valuable

Big Data before anybody else...



AT&T

VISA





73% stock trades are made by algorithms

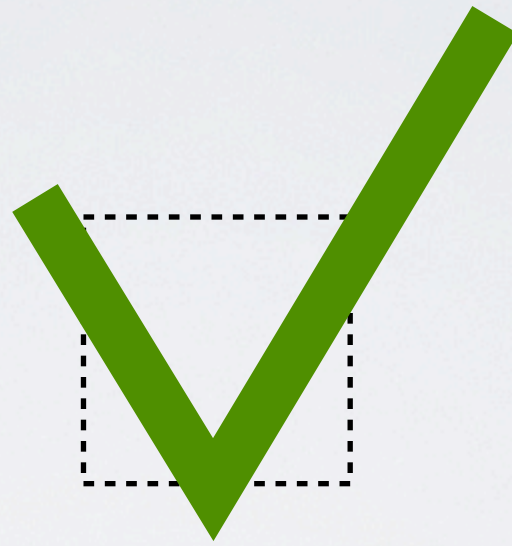


Oxford Health

\$200/year discount if you connect FourSquare

PROGRESSIVE®

discounts if you install device to track driving habits



I have Read and Agree to the Term of Service

Men Keep Cellphones in their Pockets





Terms of Service

- 1) free but we can make you pay
- 2) if phone explodes not our fault
- 3) we may take your temperature



\$5/year cost of Gmail Account

500 Million of Active Users



Free in-flight WiFi



Fast Browser



Free Wifi - Chelsea NYC



Google Analytics

Web Traffic Reports

Google - 94% of Google Revenue

AdSense

Leaderboard (728 x 90)

Affordable Hotels Special rates until the end of the month. No booking fees, book your room now! www.example.com

Dublin Hotels Browse hundreds of hotels in Dublin, sort by price, location and user reviews. www.example.com

Hotels in Ireland Compare prices of 1000s of hotels all over ireland. Offers start from €100 for 2 nights! www.example.com

Hotels Great weekend offers for hotels in Dublin. Hundreds of hotels to choose from. www.example.com

< > AdChoices ▶

Banner (468 x 60)

<u>Affordable Hotels</u> Special rates this season. No booking fees, book your room now! www.example.com	<u>Dublin Hotels</u> Browse 100+ hotels in Dublin, sort by location and user reviews www.example.com
------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------

AdChoices ▶

Button (125x125)

Affordable Hotels
Special rates until the end of the month. No booking fees.
www.pets2u.com.au

< > AdChoices ▶

Half Banner (234x60)

Dublin Hotels
Browse 100+ hotels in Dublin, sort by location and user reviews
www.example.com

AdChoices ▶



Free in-flight WiFi



Fast Browser



Google Analytics

Web Traffic Reports

Google™



Free Wifi

facebook[®]

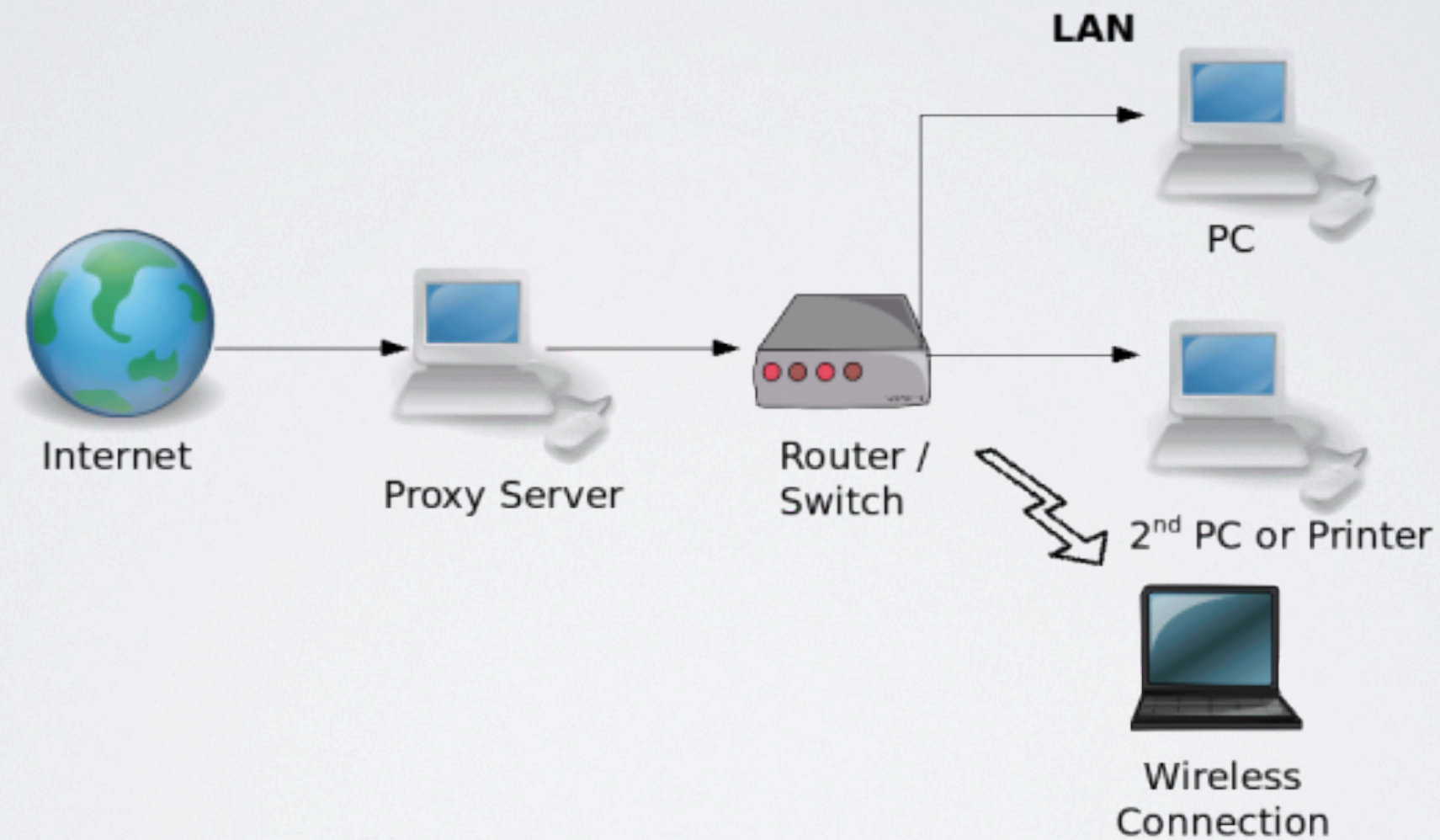
1 Billion Active Users

800 Million Posts per Day

80% users browse the web
while still logged on Facebook



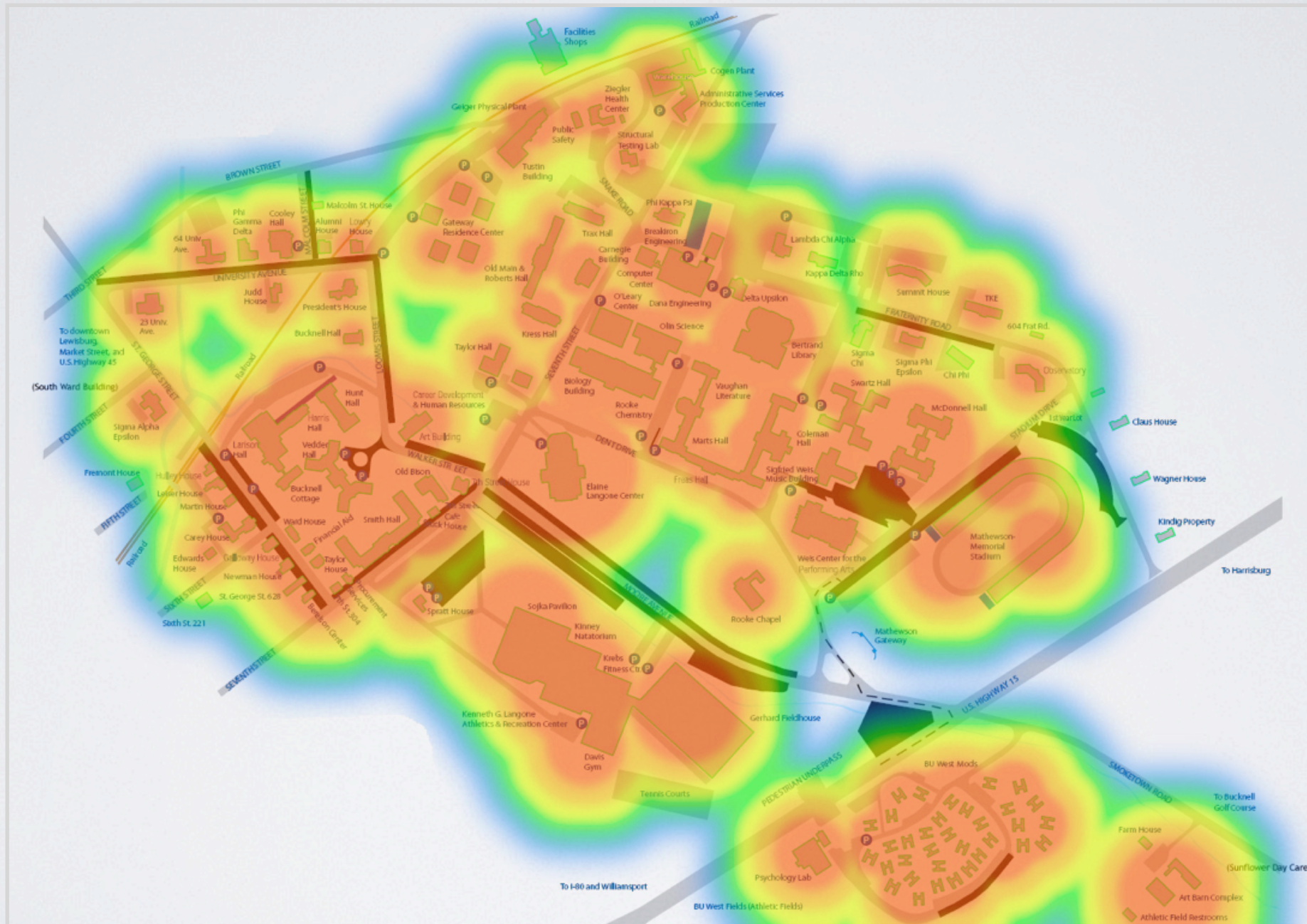
Get Proxy Logs from your IT Department



Guess what this guy was concerned about?

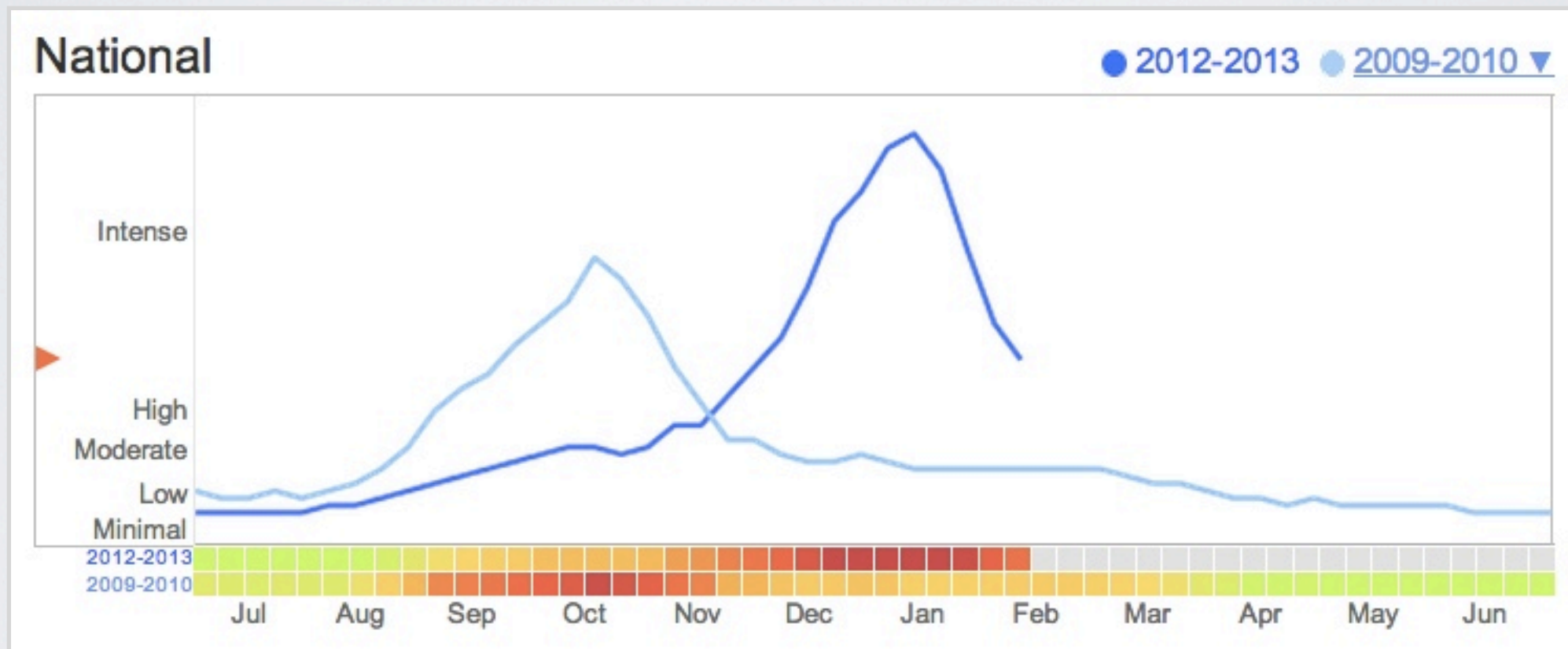
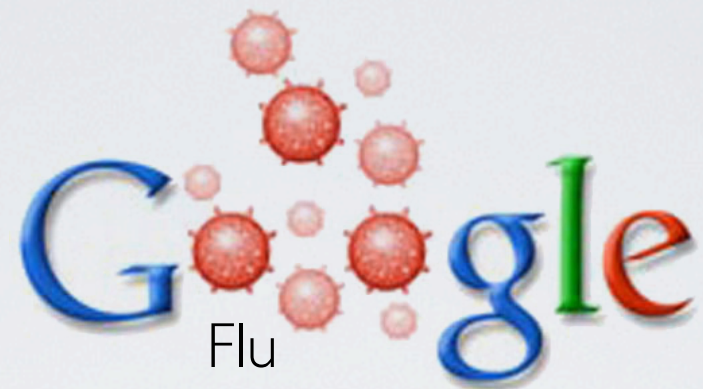
```
192.168.3.174 [21/Sep/2011:14:35:10]
  "GET http://www.espn.com/football HTTP/1.1" 200 677
192.168.3.174 [21/Sep/2011:15:15:50]
  "GET http://www.facebook.com/ HTTP/1.1" 200 14287
192.168.3.174 [21/Sep/2011:15:28:41]
  "GET http://www.facebook.com/login HTTP/1.1" 200 4567
192.168.3.174 [21/Sep/2011:15:43:00]
  "GET http://www.google.com/search?q=genital+rash HTTP/1.1" 200 1677
192.168.3.174 [21/Sep/2011:15:44:33]
  "GET http://www.cdc.gov/std/herpes/stdfact-herpes.htm HTTP/1.1" 200 18632
192.168.3.174 [21/Sep/2011:15:48:21]
  "GET http://www.cdc.gov/std/herpes/screening.htm HTTP/1.1" 200 14632
192.168.3.174 [21/Sep/2011:15:49:27]
  "GET http://www.webmd.com/std/pictures-and-facts HTTP/1.1" 200 17397
192.168.3.174 [21/Sep/2011:15:53:54]
  "GET http://www.google.com/search?q=dermatologist+52240 HTTP/1.1" 200 1693
```


Use Wireless Access Logs to Track Movements





User Generated Data



twitter



lexarod Lexa Mendoza

Off to work going home now long day in work so happy I'm going home now so tired I'm still **sick** with **flu** ☹

48 minutes ago



jordanbby7717 Jordan Daniel

I seriously can't take working at **my job** anymore. So **sick of** dealing with the people I work with and customers [#getmeoutofhere](#)

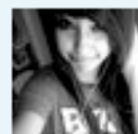
34 minutes ago



viagraviagras cheap viagra

order **Viagra** buy dublin, buy cheap online Sildenafil Citrate generic ...: We offer **Viagra** online for lowest pric... bit.ly/etelnY

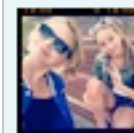
35 minutes ago



SelenaSeesStars Selena Martinez

Eating a **hamburger** that my mommy made me. ^^

1 hour ago



rubygreen1 ruby green

i don't think i have ever been this **sick** had the **flu** last night, woke up this morning with the biggest and most painful cramps in my tommy.

2 hours ago

twitter



lexarod Lexa Mendoza

Off to work going home now long day in work so happy I'm going home now so tired I'm still **sick** with **flu** ☹

48 minutes ago



jordanbby7717 Jordan Daniel

I seriously can't take working at **my job** anymore. So **sick of** dealing with the people I work with and customers [#getmeoutofhere](#)

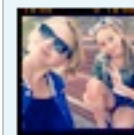
34 minutes ago



viagraviagras cheap viagra

order **Viagra** buy dublin, buy cheap online Sildenafil Citrate generic ...: We offer **Viagra** online for lowest pric... bit.ly/etelnY

35 minutes ago



rubygreen1 ruby green

i don't think i have ever been this **sick** had the **flu** last night, woke up this morning with the biggest and most painful cramps in my tommy.

2 hours ago



SelenaSeesStars Selena Martinez

Eating a **hamburger** that my mommy made me. ^.^

1 hour ago

twitter



lexarod Lexa Mendoza

Off to work going home now long day in work so happy I'm going home now so tired I'm still **sick** with **flu** ☹

48 minutes ago



jordanbby7717 Jordan Daniel

I seriously can't take working at **my job** anymore. So **sick of** dealing with the people I work with and customers [#getmeoutofhere](#)

34 minutes ago



viagraviagras cheap viagra

order **Viagra** buy dublin, buy cheap online Sildenafil Citrate generic ...: We offer **Viagra** online for lowest pric... bit.ly/etelnY

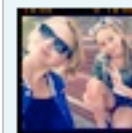
35 minutes ago



SelenaSeesStars Selena Martinez

Eating a **hamburger** that my mommy made me. ^.^

1 hour ago



rubygreen1 ruby green

i don't think i have ever been this **sick** had the **flu** last night, woke up this morning with the biggest and most painful cramps in my tommy.

2 hours ago

twitter



lexarod Lexa Mendoza

Off to work going home now long day in work so happy I'm going home now so tired I'm still **sick** with **flu** ☹

48 minutes ago



jordanbby7717 Jordan Daniel

I seriously can't take working at **my job** anymore. So **sick of** dealing with the people I work with and customers [#getmeoutofhere](#)

34 minutes ago



viagraviagras cheap viagra

order **Viagra** buy dublin, buy cheap online Sildenafil Citrate generic ...: We offer **Viagra** online for lowest pric... bit.ly/etelnY

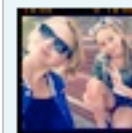
35 minutes ago



SelenaSeesStars Selena Martinez

Eating a **hamburger** that my mommy made me. ^.^

1 hour ago



rubygreen1 ruby green

i don't think i have ever been this **sick** had the **flu** last night, woke up this morning with the biggest and most painful cramps in my tommy.

2 hours ago

twitter



lexarod Lexa Mendoza

Off to work going home now long day in work so happy I'm going home now so tired I'm still **sick with flu** ☹

48 minutes ago



jordanbby7717 Jordan Daniel

I seriously can't take working at **my job** anymore. So **sick of** dealing with the people I work with and customers [#getmeoutofhere](#)

34 minutes ago



viagraviagras cheap viagra

order **Viagra** buy dublin, buy cheap online Sildenafil Citrate generic ...: We offer **Viagra** online for lowest pric... bit.ly/etelnY

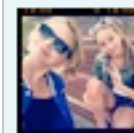
35 minutes ago



SelenaSeesStars Selena Martinez

Eating a **hamburger** that my mommy made me. ^.^

1 hour ago

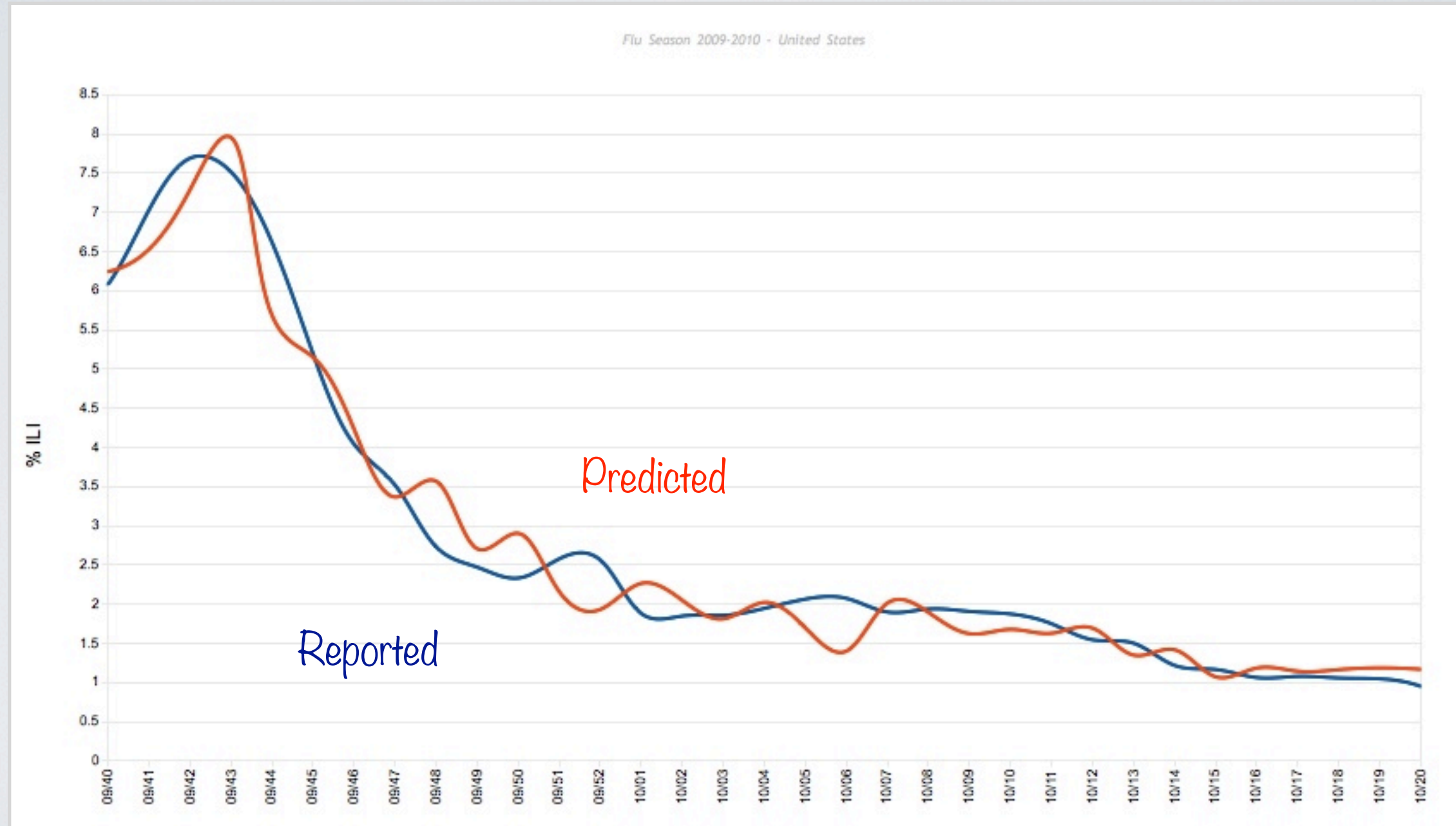


rubygreen1 ruby green

i don't think i have ever been this **sick** had the **flu** last night, woke up this morning with the biggest and most painful cramps in my tommy.

2 hours ago

Predicting ILI% using **twitter**



foursquare



Doctor



Gym



Restaurant

foursquare



Doctor



Gym



Restaurant



Drunken



Mayor



Movie



Pizza

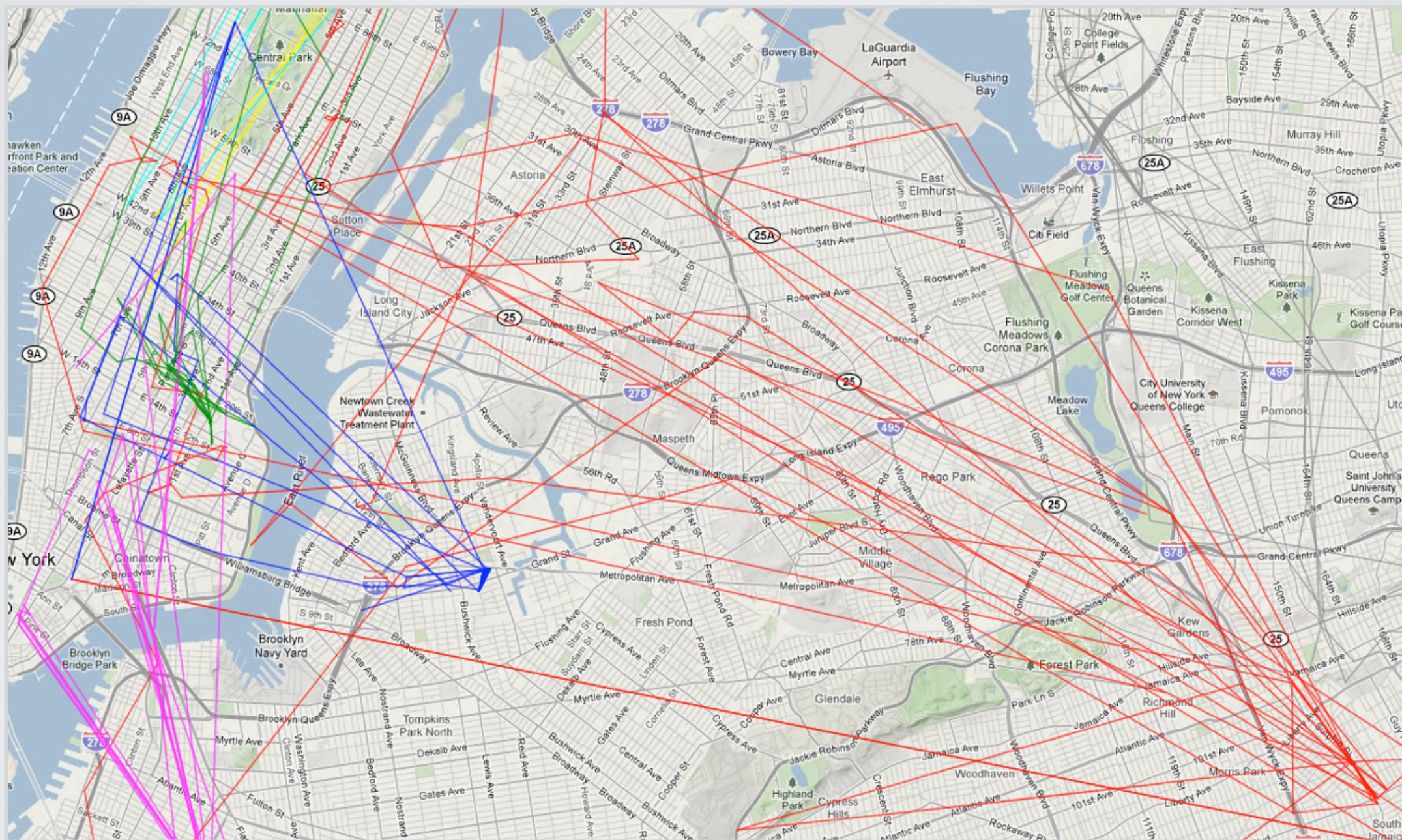


Traveler

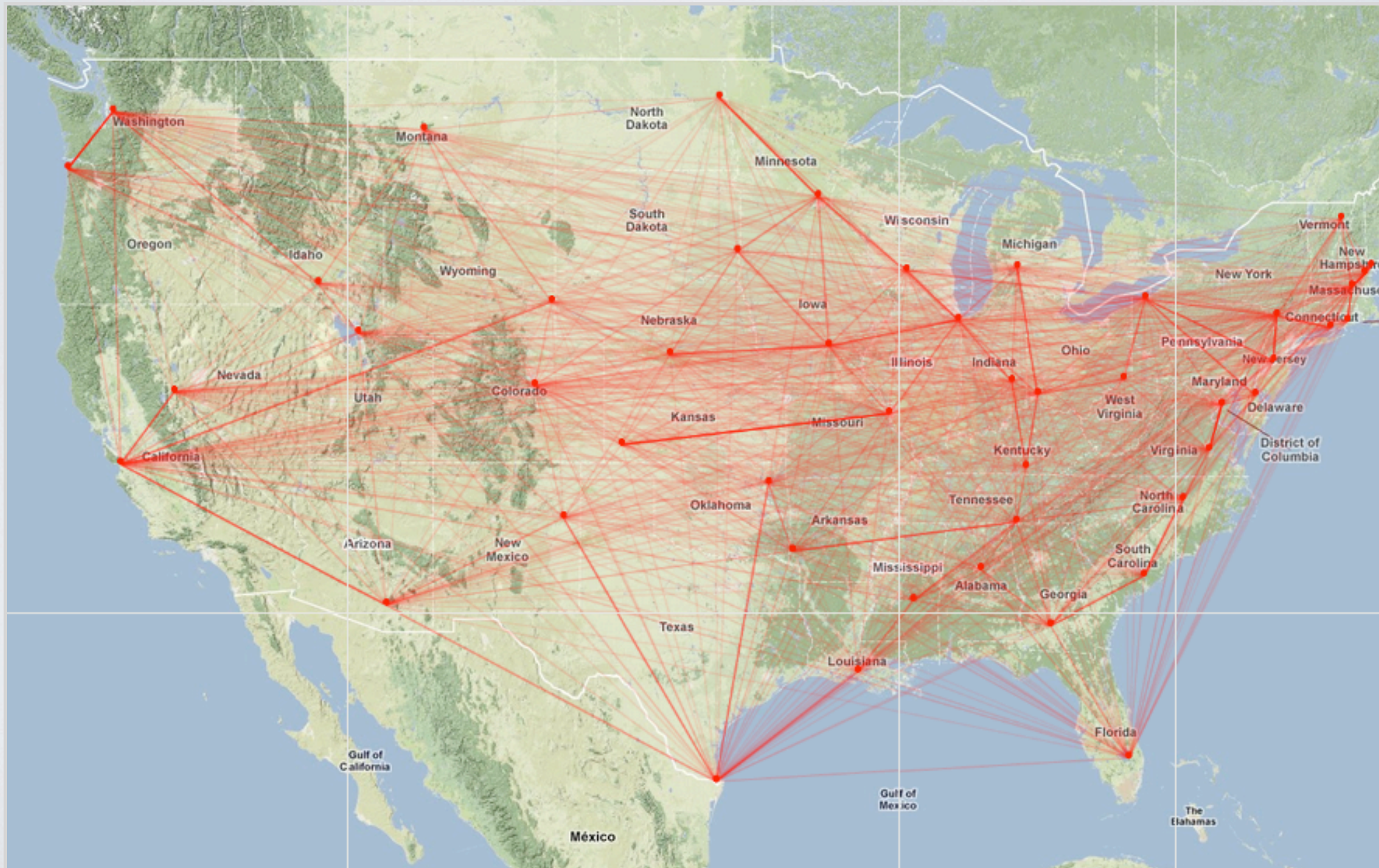


Gym-rat

foursquare



foursquare



twitter



Matt @apple4ever

33 mins

My weight: 228.1 lb. 45.1 lb to go. Started at 234.6... withings.com

Expand



Brad Wood @randombrad

37 mins

My weight: 223.2 lb. 22.6 lb to go. [#gettingInShape withings.com](http://withings.com)

Expand



Shawn Hutchinson @Hutchbot1984

36 mins

I ran 6.30 km with @MapMyFitness. Check out my route in Carlisle ! [#run](#) [#running](#) [#HWD](#) [#Sweat](#)

Expand



bigbadbootyman @marckrajewski

22 mins

I ran with MapMyRun! Distance: 17.74km, time: 02:04:12, pace: 7:34min/km, speed: 7.93km/h. mapmyrun.com/view_route?r=6...

Expand



Chelsea Grothe. @ChelseaGrothe

16 mins

I just ran 8.4 km

Expand



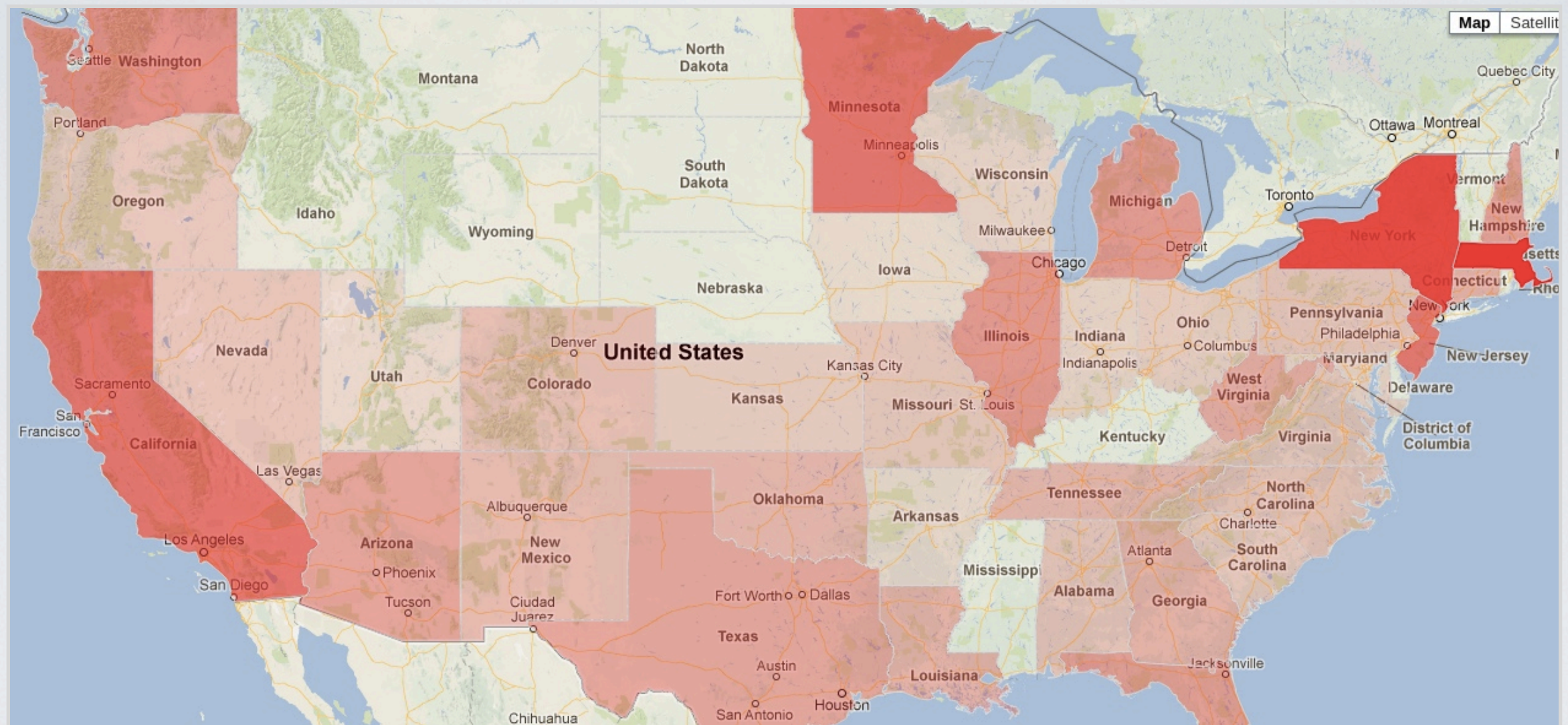
John McWaldi @McWaldi

7 mins

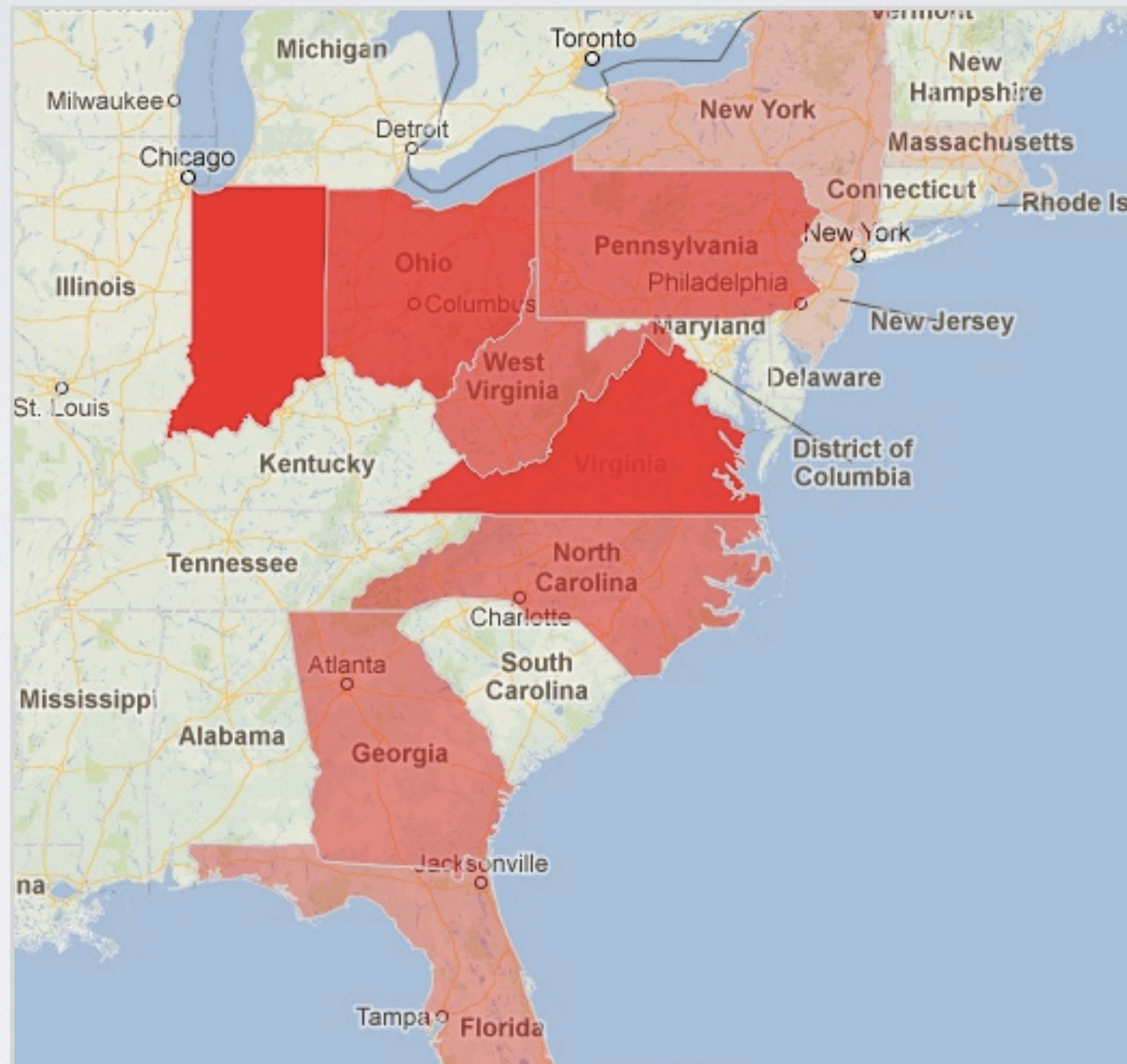
Mein Gewicht: 66.2 kg. first try :) withings.com

Expand

Average Length of Runs - **twitter**

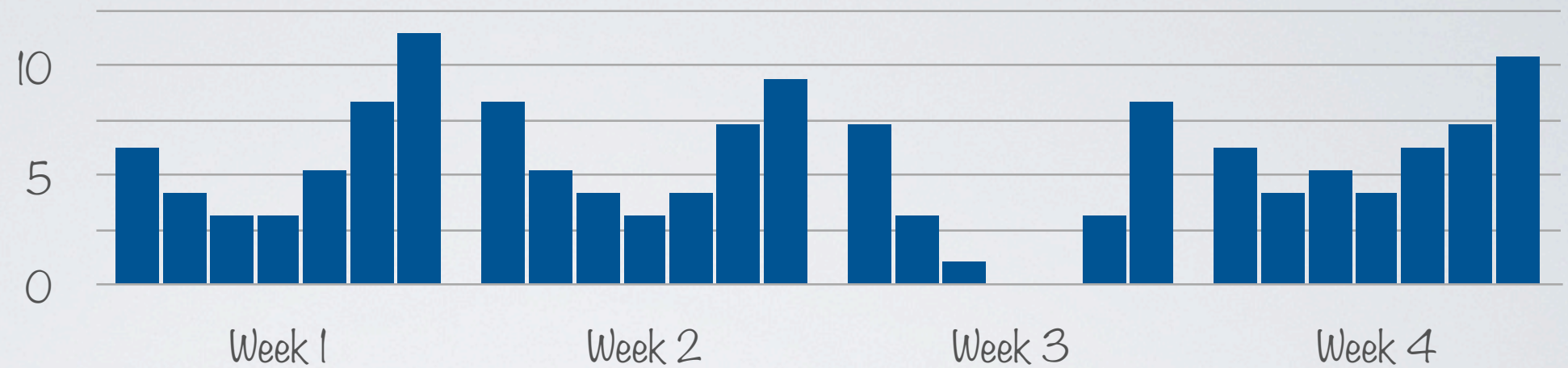


Average Weight - **twitter**

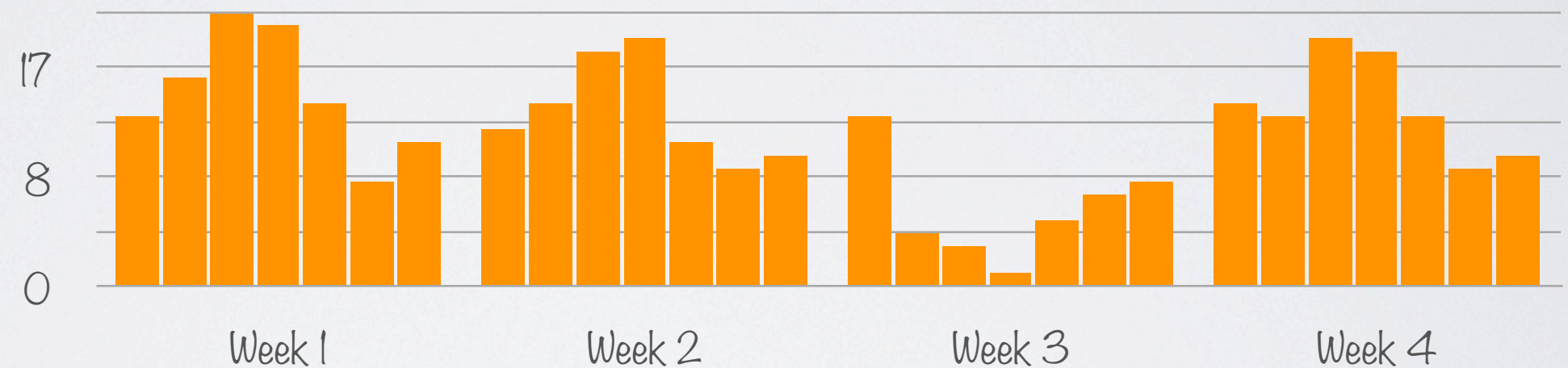


Social Media Monitoring

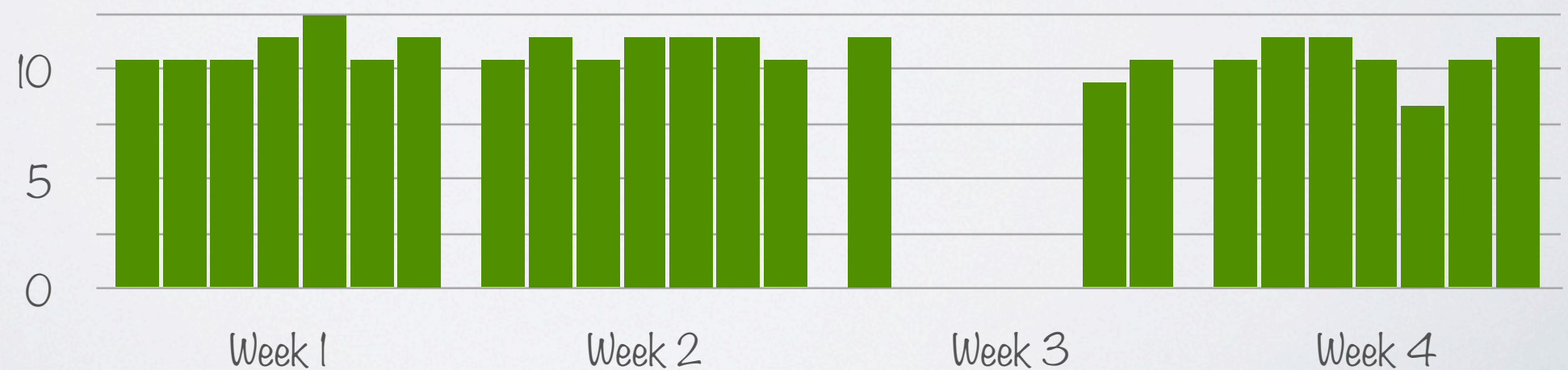
foursquare
checkins



twitter
tweets

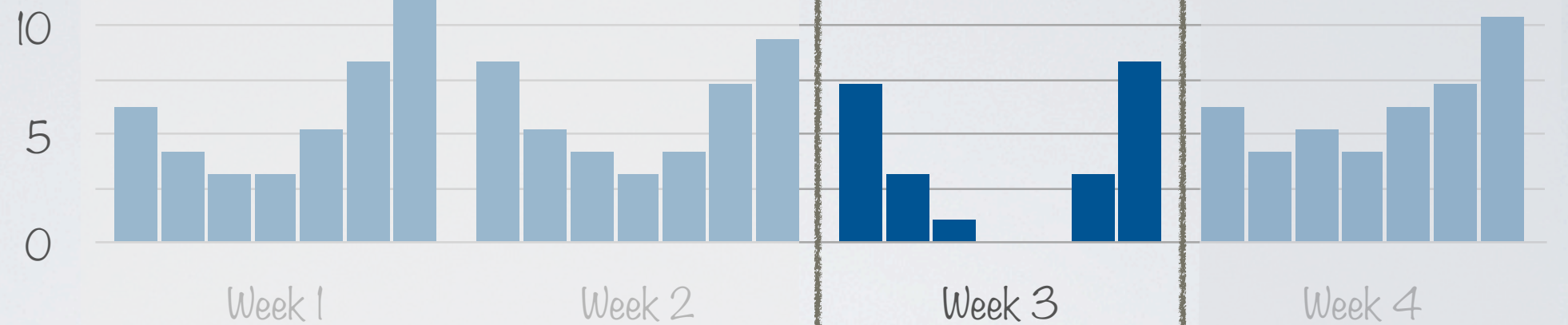


RunKeeper
km

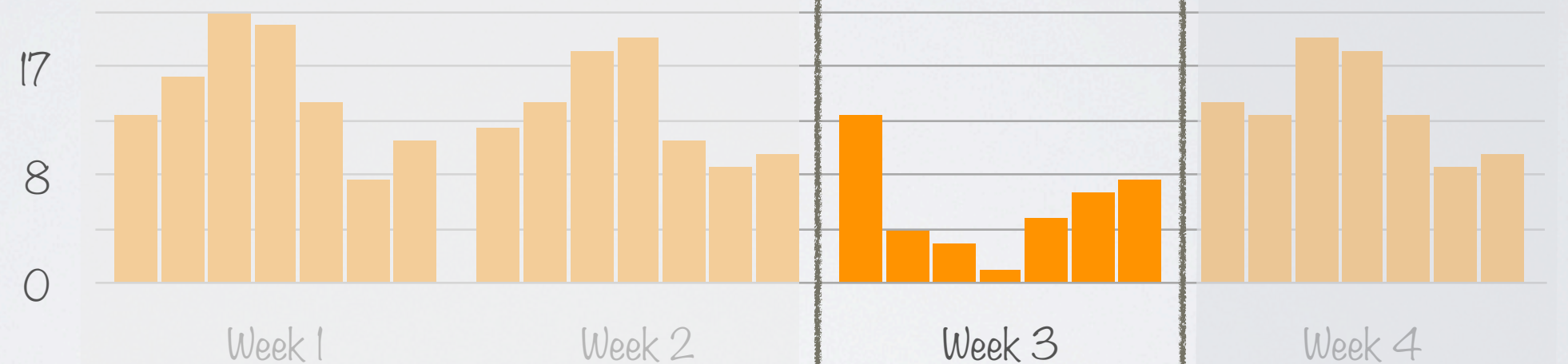


Social Media Monitoring

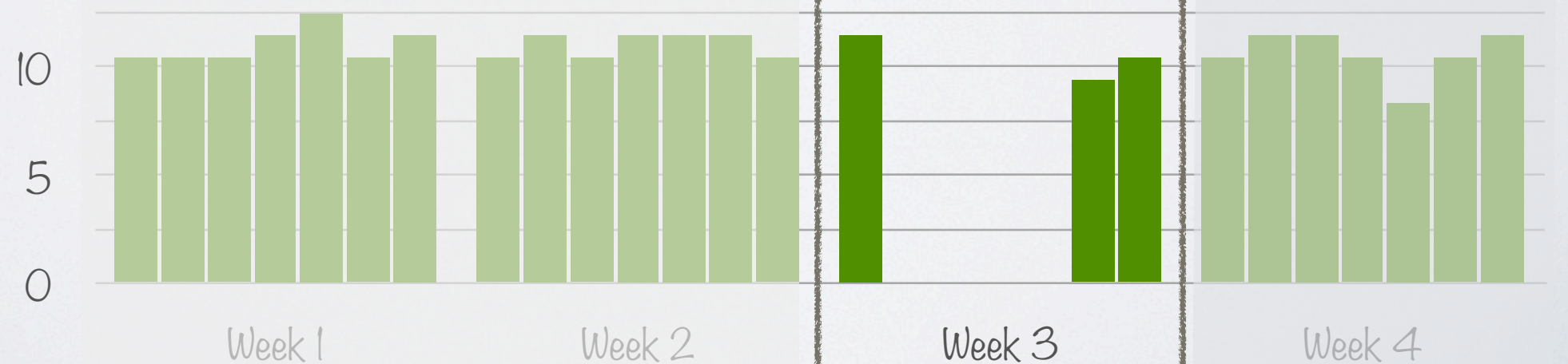
foursquare
checkins

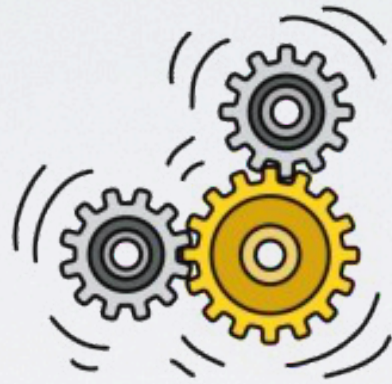


twitter
tweets



RunKeeper
km





New Devices



FitBit



BodyMedia



Withings



Nike Fuel



iHealth



Ki



Zeo

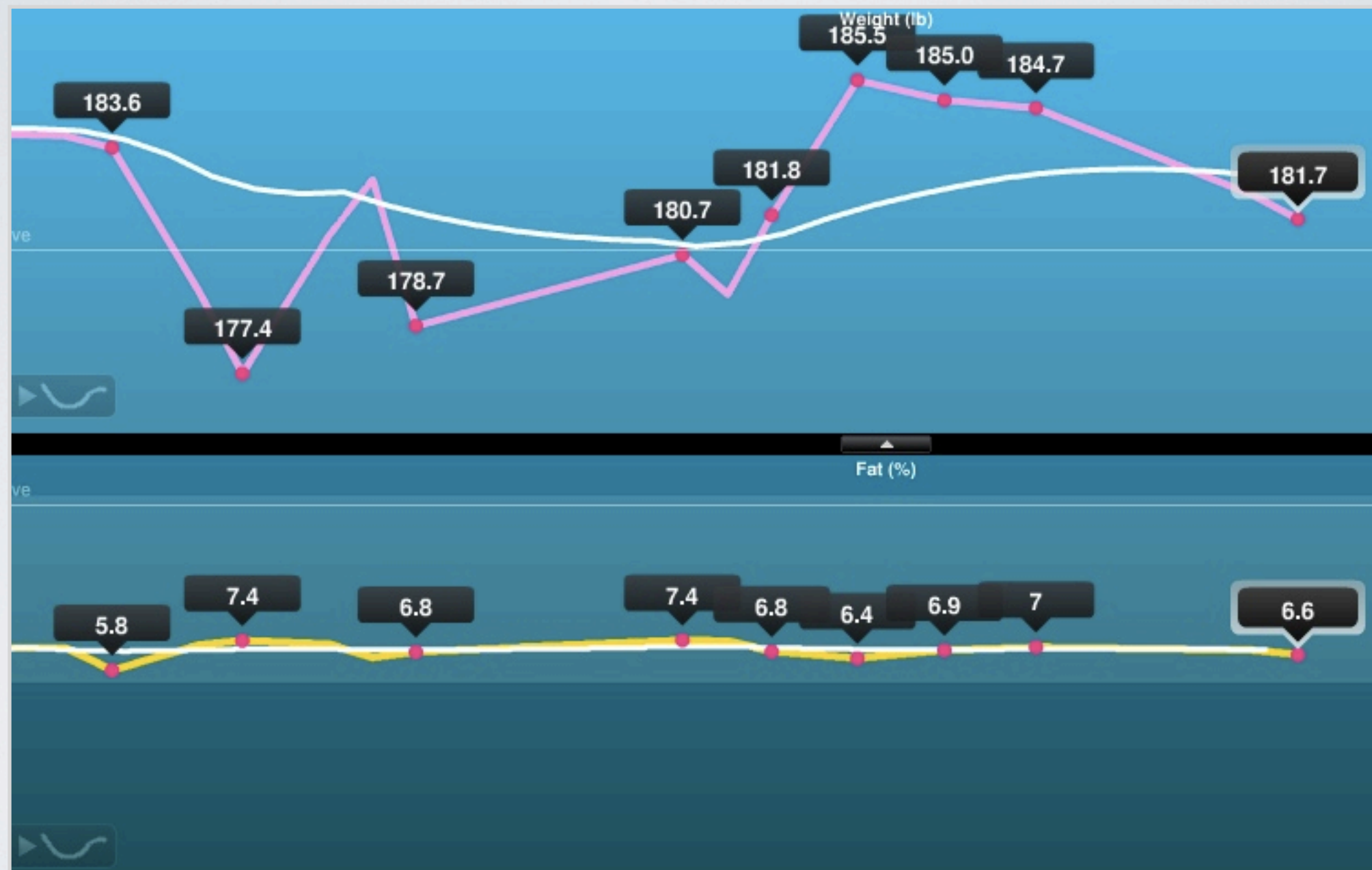


Jawbone Up



Withings Scale

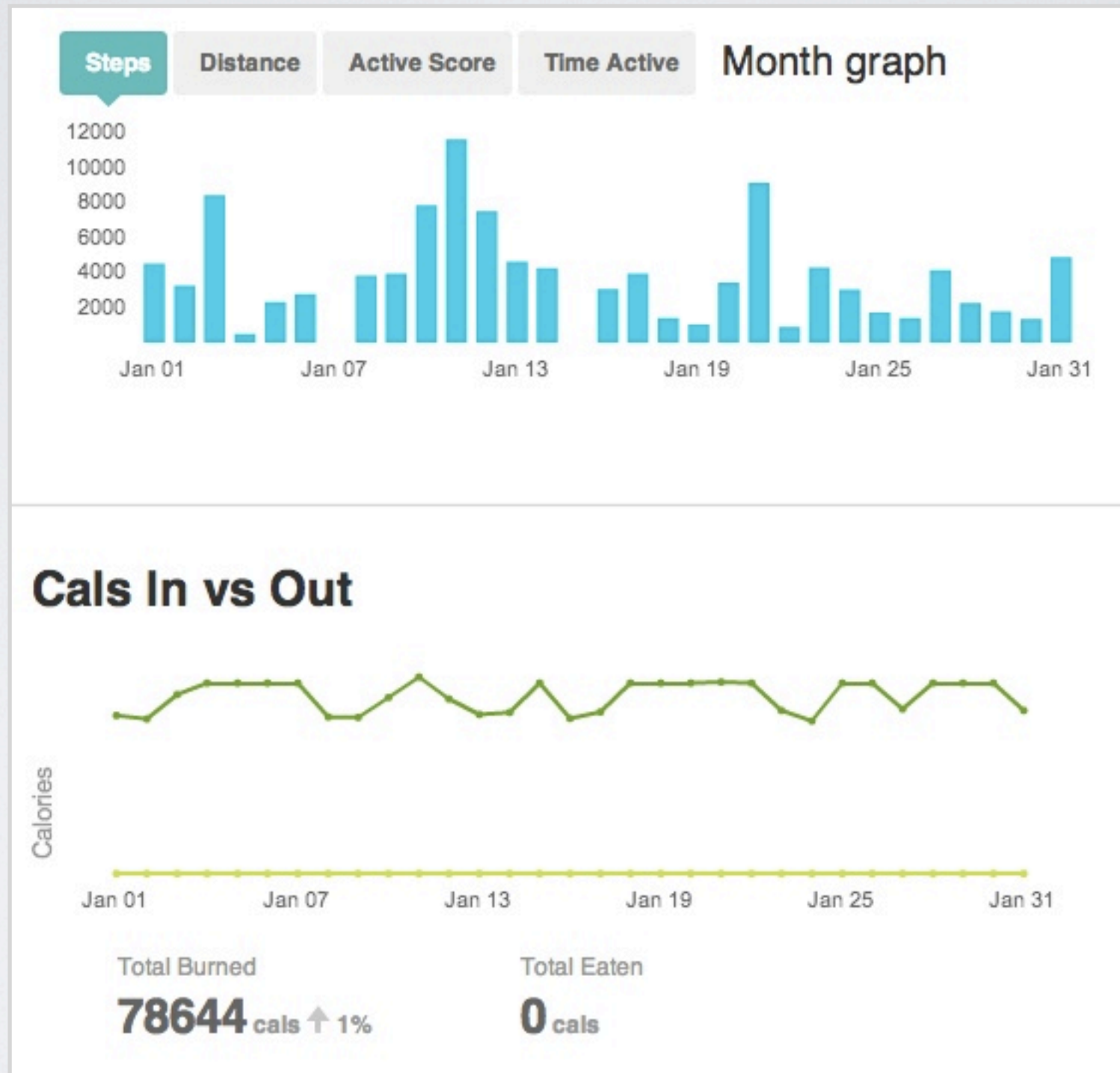
Withings - Weight/Fat Statistics





FitBit Zip

FitBit - Daily Steps





iHealth Pressure Monitor



Zeo Sleep Trainer



iPhone



Android



Instant Heart Monitor



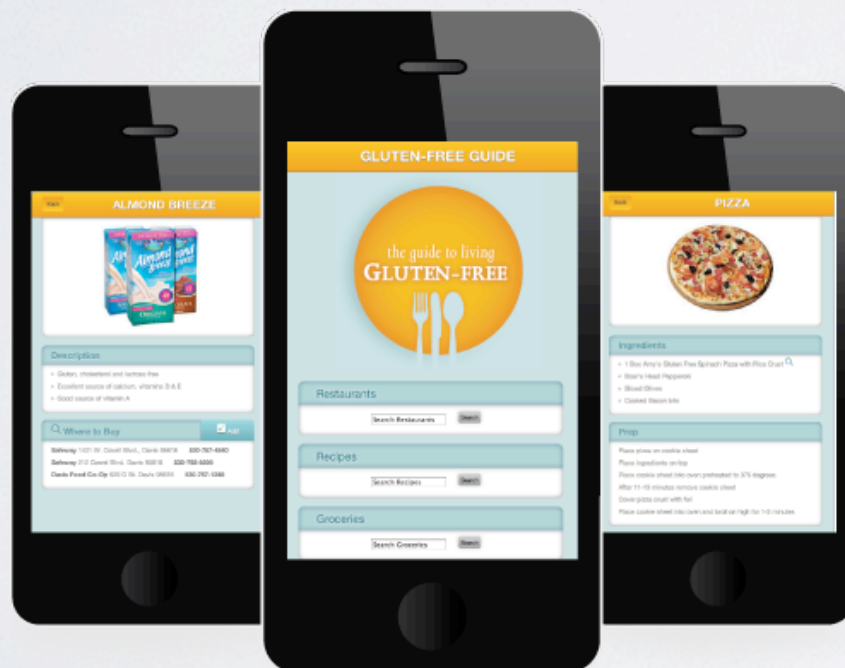
Cardio



Calories Tracking



Running Partner

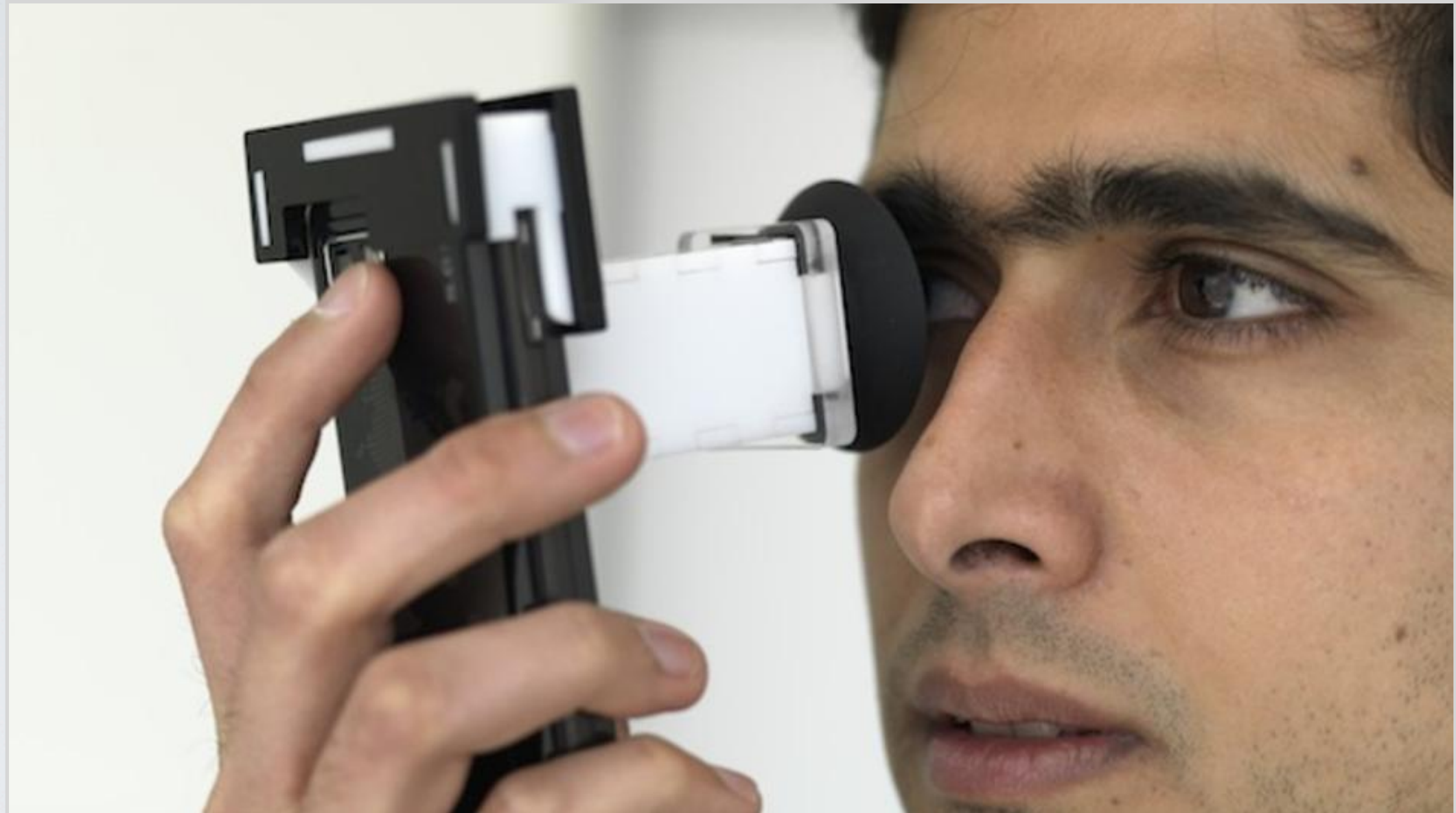


Dietary Restriction



Addiction Management

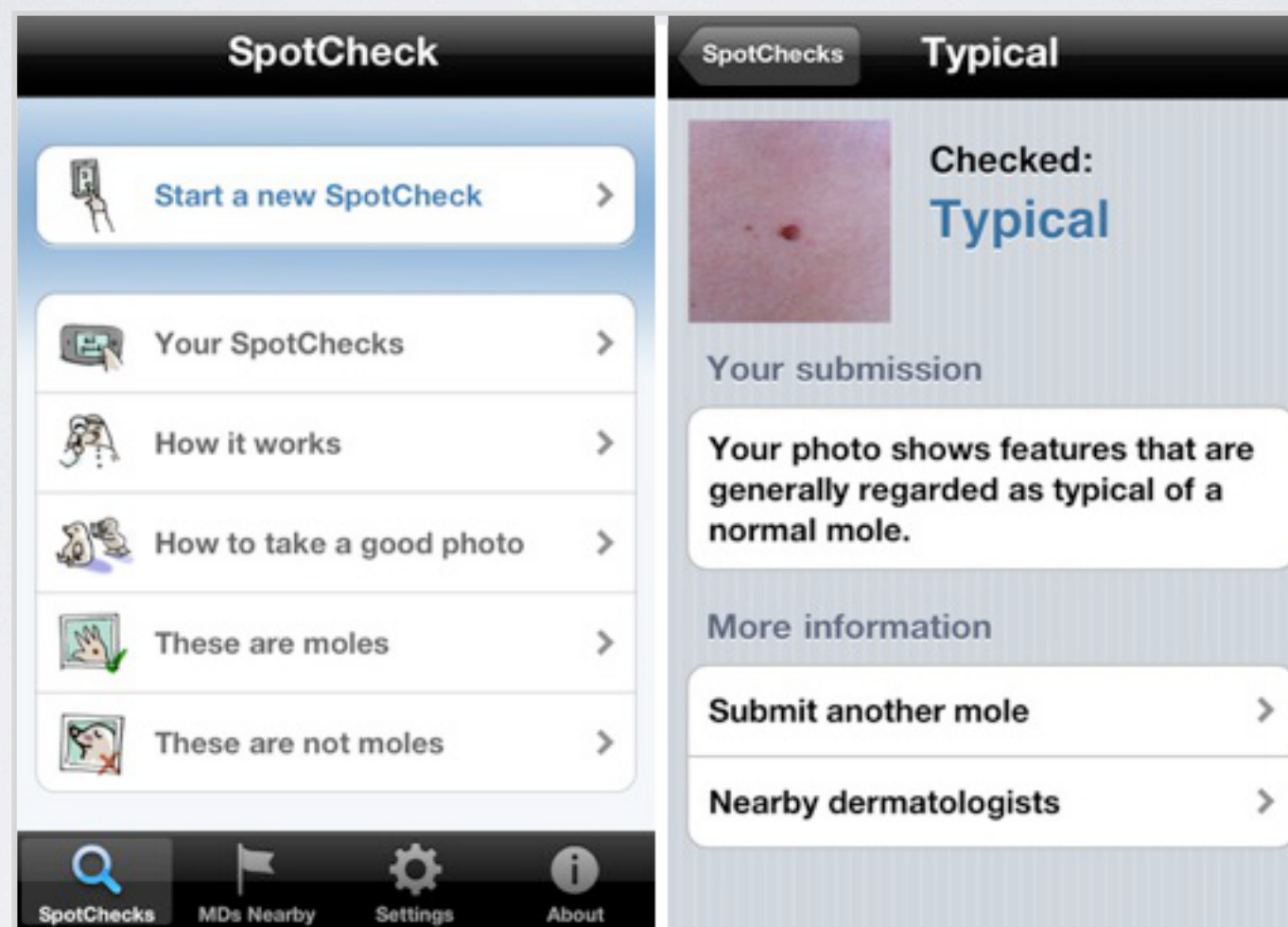
Eye Exam

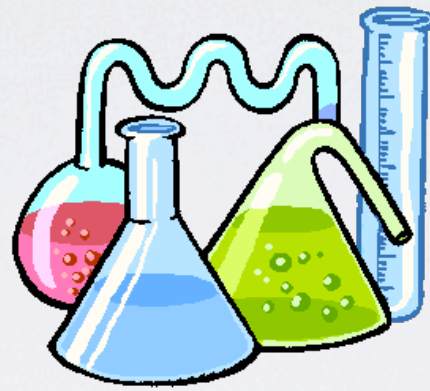


Glucodock



Spotcheck





New Diagnostic Paths

"Dystextia": Stroke diagnosed with help of garbled text message



ISTOCKPHOTO

Monitoring Computer Interactions to Detect Early Cognitive Impairment in Elders

Holly Jimison, *Member, IEEE*, Nicole Jessey, James McKanna, and Tracy Zitzelberger, Jeffrey Kaye

Abstract—Maintaining cognitive performance is a key factor influencing elders' ability to live independently with a high quality of life. We have been developing unobtrusive measures to monitor cognitive performance and potentially predict decline using information from routine computer interactions in the home. Early detection of cognitive decline offers the potential for intervention at a point when it is likely to be more successful. This paper describes recommendations for the conduct of studies monitoring cognitive function based on routine computer interactions in elders' home environments.

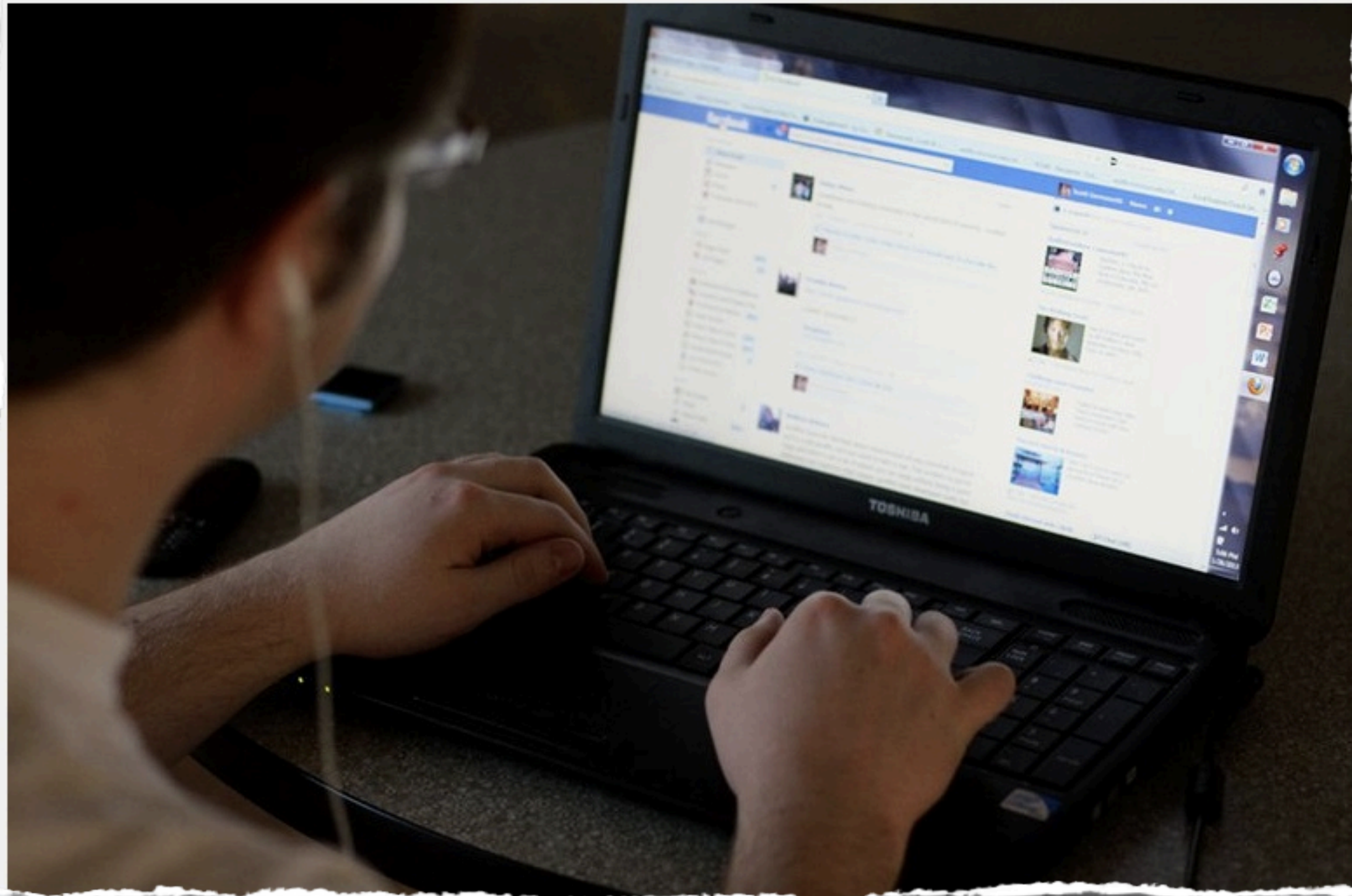
I. INTRODUCTION

ELDERS are the fastest growing demographic in many countries, with a concomitant escalation of health care resources being spent on conditions associated with aging. One of the key functional losses at risk with aging is a decline in cognitive abilities. Estimates vary depending on assessment protocols and populations studied, but up to 50% of all individuals over age 85 are found to have measurable decline in cognitive function [1]. These individuals are at high risk for dementia, requiring increasing levels of assisted living. Even mild cognitive declines lead to degraded quality of life. Thus a major goal of seniors and their families is to optimize their quality of life and remain

useful in predicting an elder's degree of cognitive health. These metrics include relative typing speed and accuracy, relative mouse movement efficiency, and relative performance on computer games that have been designed for cognitive monitoring. Our continuous monitoring of cognitive indicators allows us to perform trend detection of individual performance, so that we are less susceptible to unwanted variability due to educational, language, and cultural backgrounds, as compared to traditional cognitive assessment tests. In this pilot project, we deployed our monitoring software in the homes of 15 elders in the Portland, Oregon metropolitan area for a period of 6 months and tested our ability to monitor computer activities and keep elders motivated and engaged in using the computer. This test was in preparation for a larger prospective long-term study that will determine our ability to use these types of metrics to predict future cognitive decline. Based on our pilot test of the monitoring software in the homes of elders, we developed a set of recommendations for effectively managing a large number of elderly participants in a long-term study, and eventually, as a routine service available to the public at large.

II. BACKGROUND AND MOTIVATION

Facebook activity may reflect mental health issues according to MU psychology study



 **RunKeeper**


OpenTable™

foursquare

 **cardiio**

 **fitbit**

twitter

NETFLIX

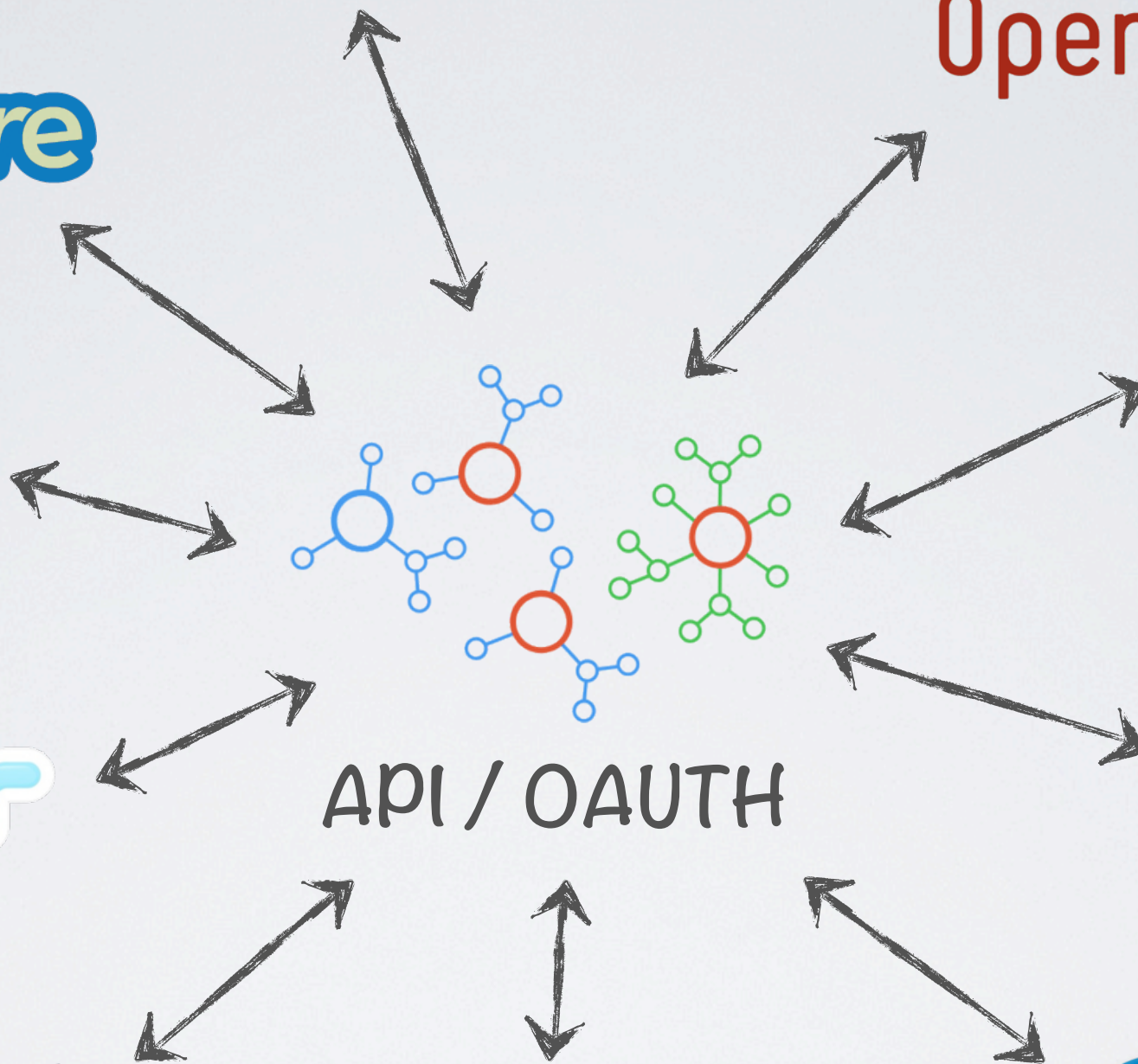
API / OAUTH

facebook®

Gmail™
by Google™



Instagram



Alessio Signorini

alessio-signorini@uiowa.edu
@a_signorini

Computational Epidemiology Group
University of Iowa