



Alessio Signorini

Technical Co-founder

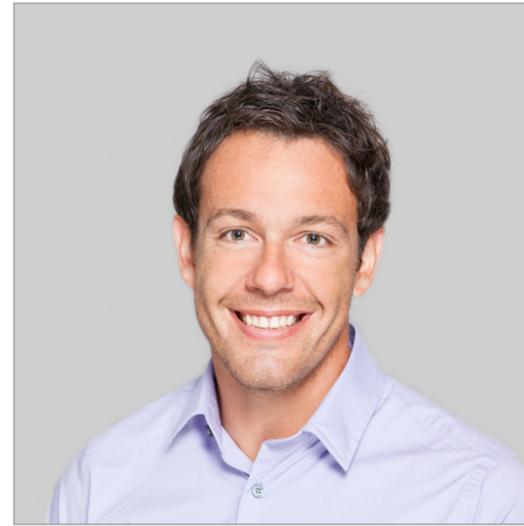


www.evidation.com

1036 Santa Barbara Street
Santa Barbara, CA 93103



Christine Lemke
President



Luca Foschini
Chief Data Scientist



Mikki Nasch
VP of Business Dev.



Alessio Signorini
Chief Architect

Started in a house off Milpas St. 6 years ago

2 offices, 85+ employees, \$30M+ raised

Work with Hospitals, Clinics, Universities, Insurance,
Pharma, Governmental Organization



Christine Lemke
President



Mikki Nasch
VP of Business Development

Sense Networks

CHANNELIQ

Microsoft

brightspark

BACKWEB
TECHNOLOGIES

FairIsaac



Luca Foschini

PhD, Theoretical Computer Science





Approximation Algorithms for Problems on Networks and Streams of Data

Luca Foschini

In this dissertation we investigate approximation algorithms for problems defined on networks and streams of data. The unifying theme of all problems discussed is that finding an exact solution to them is impractical. Such impracticality may stem from a provable time complexity characterization in a general computation model, or it may be dictated by constraints on the resources available to the algorithm.

The first problem concerns finding shortest paths in a network with time-variant transit times. We show that under minimal assumptions on the transit time functions, the complexity of the structure of the shortest paths between two network nodes over a time interval may overwhelm a decision maker; thus, it is crucial we devise strategies that

30+ Publications



Alessio Signorini
PhD, Artificial Intelligence



IMRSV
DATA LABS

OneRiot
Search for the pulse.



ROCK
HEALTH





The Indexable Web is More than 11.5 Billion Pages

A. Gulli

Universita' di Pisa, Dipartimento di Informatica
gulli@di.unipi.it

A. Signorini

University of Iowa, Computer Science
alessio-signorini@uiowa.edu

ABSTRACT

In this short paper we estimate the size of the public indexable web at 11.5 billion pages. We also estimate the overlap and the index size of Google, MSN, Ask/Teoma and Yahoo!.

Categories and Subject Descriptors

H.3.3 [Information Storage And Retrieval]: Information Search and Retrieval

General Terms

Design, Experimentation, Measurement

Keywords

Search Engines, Index Sizes, Size of the Web

1. INTRODUCTION

What is the current size of the Web? At the time of this writing, Google claims to index more than 8 billion pages, MSN claims about 5 billion pages, Yahoo! at least 4 billion and Ask/Teoma more than 2 billion. Two sources for tracking the growth of the Web are [8, 7], although they are not

2. ESTIMATING SIZES AND OVERLAPS

We review [2] and point out where our approach differs. The idea is quite simple: suppose we have two search engines A and B with size $s(A)$ and $s(B)$, respectively, and intersection $A \& B$. Let $Pr(A)$ represent the probability that an element belongs to the set A , and let $Pr(A \& B|A)$ represent the conditional probability that an element belongs to both sets given that it belongs to A . Then, $Pr(A \& B|A) \approx s(A \& B)/s(A)$ and similarly, $Pr(A \& B|B) \approx s(A \& B)/s(B)$, and therefore the relative size is $s(A)/s(B)$ that is approximately $Pr(A \& B|B)/Pr(A \& B|A)$. The methodology estimates $\frac{s(A)}{s(B)}$ by the ratio between the fraction of URLs sampled from B found in A and the fraction of URLs sampled from A found in B . It also estimates the overlap (fraction of search engine A index, indexed by search engine B) as fraction of URLs sampled from A found in B .

To implement this idea, one needs a procedure for picking pages uniformly at random from the index of a particular engine - i.e. a *sampling procedure* -, and a procedure for determining whether a particular page is indexed by a



The Indexable Web is More than 11.5 Billion Pages

Using Twitter to Estimate H1N1 Influenza Activity

Alessio Signorini[†], Philip M. Polgreen[‡], Alberto Maria Segre[†]

Departments of Computer Science[†] and Internal Medicine[‡], The University of Iowa

Objective

This paper describes a system that uses Twitter to estimate influenza-like illness levels by geographic region.

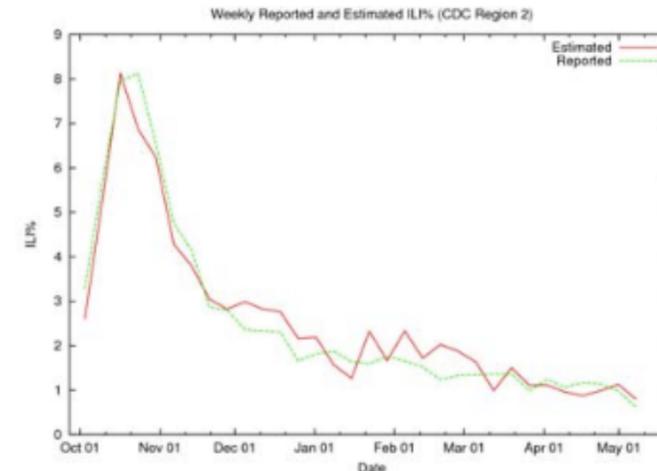
Background

Twitter is a free social networking and micro-blogging service that enables its millions of users to send and read each other's "tweets," or short messages limited to 140 characters. The service has more than 190 million registered users and processes about 55 million tweets per day [1]. Despite a high level of chatter, the Twitter stream does contain useful information, and, because tweets are often sent from handheld platforms on location, they convey more immediacy than other social networking systems.

Methods

We collected and stored all public tweets beginning October 1, 2009 that matched a set of pre-specified search terms (*e.g.*, flu, swine, influenza, tamiflu, oseltamivir, h1n1, etc.). After culling, *e.g.*, non-U.S. and non-English tweets, and applying appropriate stemming algorithms, the remaining tweets were used to produce a dictionary of English word equivalents.

max=1.25%) and a standard deviation of 0.26%. Similar results were obtained when estimating ILI at a national level.



Conclusions

Our results demonstrate that Twitter traffic can be used to provide real-time estimates of disease activity. Our ability to quickly detect trends which are then confirmed by observations from traditional surveillance approaches make this new form of surveillance a promising area of research at the interface between computer science, epidemiology,



The Indexable Web is More than 11.5 Billion Pages

Using Twitter to Estimate H1N1 Influenza Activity

OPEN ACCESS Freely available online

PLoS one

The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic

Alessio Signorini¹, Alberto Maria Segre¹, Philip M. Polgreen^{2,3*}

1 Department of Computer Science, University of Iowa, Iowa City, Iowa, United States of America, **2** Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, Iowa, United States of America, **3** Department of Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City, Iowa, United States of America

Abstract

Twitter is a free social networking and micro-blogging service that enables its millions of users to send and read each other's "tweets," or short, 140-character messages. The service has more than 190 million registered users and processes about 55 million tweets per day. Useful information about news and geopolitical events lies embedded in the Twitter stream, which embodies, in the aggregate, Twitter users' perspectives and reactions to current events. By virtue of sheer volume, content embedded in the Twitter stream may be useful for tracking or even forecasting behavior if it can be extracted in an efficient manner. In this study, we examine the use of information embedded in the Twitter stream to (1) track rapidly-evolving public sentiment with respect to H1N1 or swine flu, and (2) track and measure actual disease activity. We also show that Twitter can be used as a measure of public interest or concern about health-related events. Our results show that estimates of influenza-like illness derived from Twitter chatter accurately track reported disease levels.

Citation: Signorini A, Segre AM, Polgreen PM (2011) The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. PLoS ONE 6(5): e19467. doi:10.1371/journal.pone.0019467

Editor: Alison P. Galvani, Yale University, United States of America

Received: September 20, 2010; **Accepted:** April 7, 2011; **Published:** May 4, 2011

Copyright: © 2011 Signorini et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Support for this research was provided by a National Institutes of Health Career Award NIAID K01 AI75089 (PMP), by the National Institutes of Health

The Indexable Web is More than 11.5 Billion Pages

Using Twitter to Estimate H1N1 Influenza Activity

OPEN ACCESS Freely available online

PLoS one

Inferring travel from social media

Alessio Signorini*, Philip Polgreen and Alberto Segre

Computer Science, University of Iowa, Iowa City, IA, USA

the period from September 11, 2010 through January 28, 2011. The Twitter API provides a random sample of tweets; non-geocoded tweets or tweets originating from outside the United States were discarded. In addition, users with fewer than 6 records, or those who check in too frequently (more than once in 5 seconds) or travel too quickly (faster than 1800 km/hr) were removed to exclude automated bots or other location spam.

*Alessio Signorini
E-mail: alessio-signorini@uiowa.edu

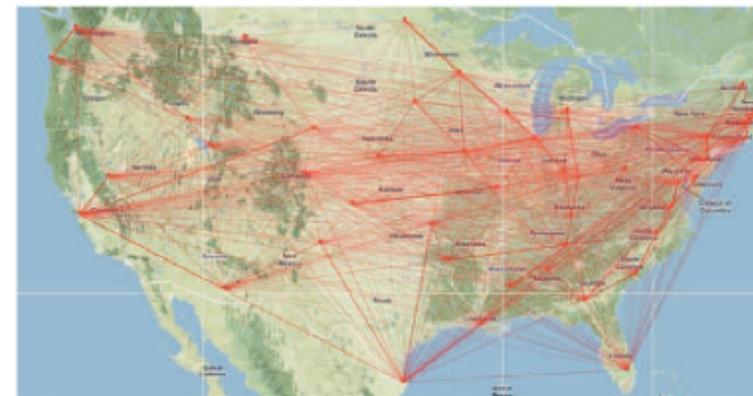
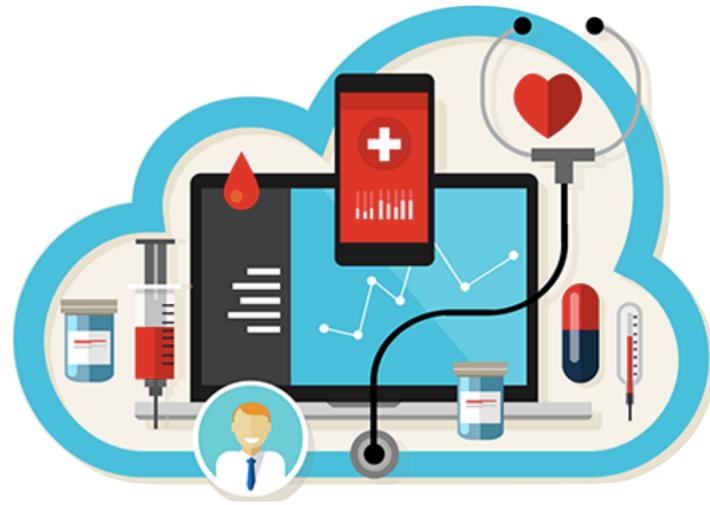


Fig. 1. User Transitions between U.S. States.



Fig. 2. Density of Check-ins over Manhattan, New York City, NY.



Data Experts with a dream

Achievement **HOME** ABOUT LOG IN

WE SERIOUSLY PAY YOU TO TRACK YOUR STEPS.

We pay you (via PayPal or direct deposits) for doing healthy things! Earn for tracking steps, sleep, meals, tweets & more! It's free to sign up for Achievement and start earning.

SIGN UP

"My sister told me about it, and it seemed too good to be true at first. But when she said that she had already received \$20, I signed up. It was so easy and a monetary motivation really

Achievement

2M+ Members | iOS/Android/Web | Research

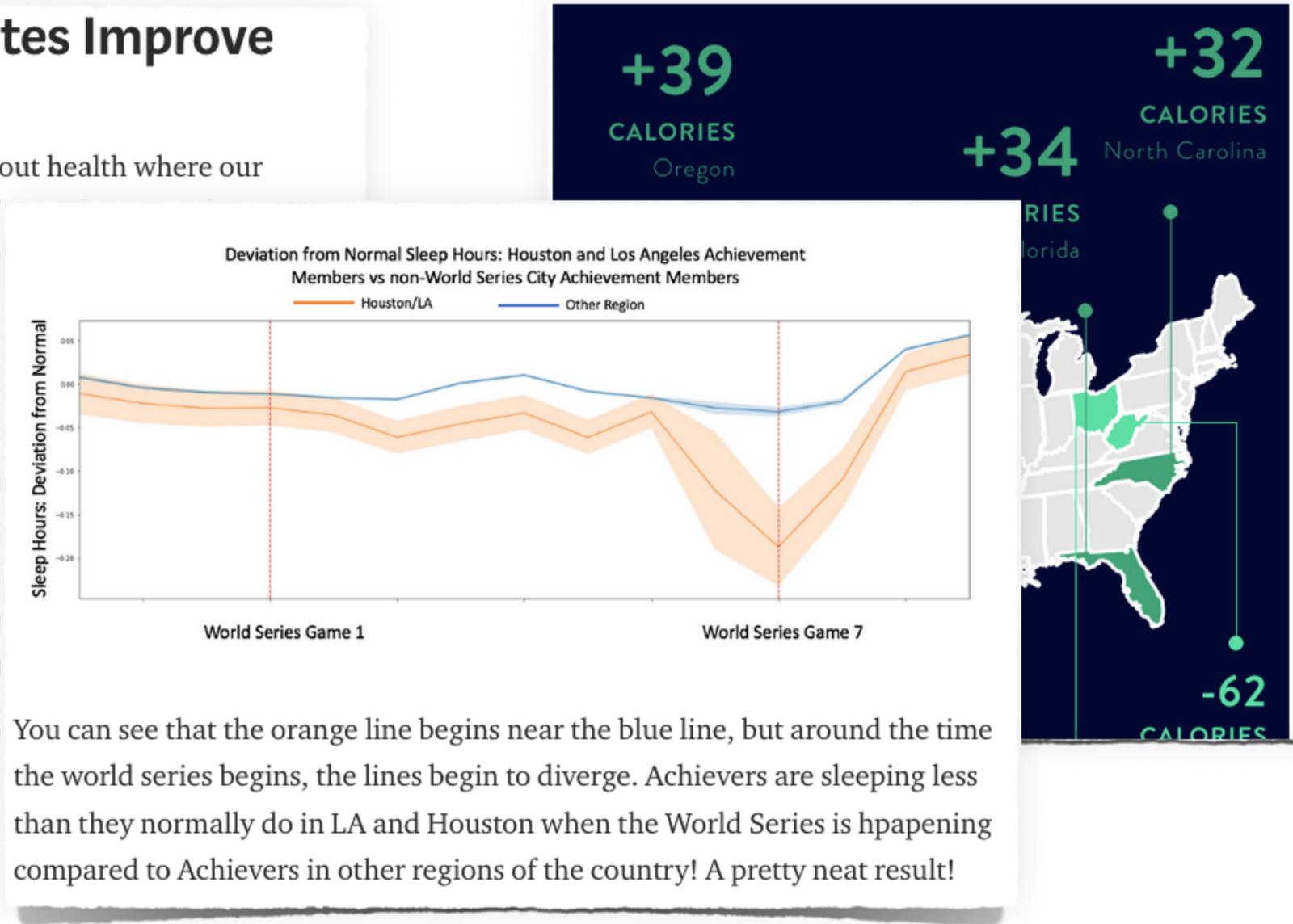
Achievement Research

Can a Mobile App for Diabetes Improve A1C Levels?

At Achievement, we often run research studies about health where our community can choose to participate. We're also full of insights to better understand your personal health. When exciting things, we think it seems natural to share them.

Welcome to our first research digest! Stay tuned for more.

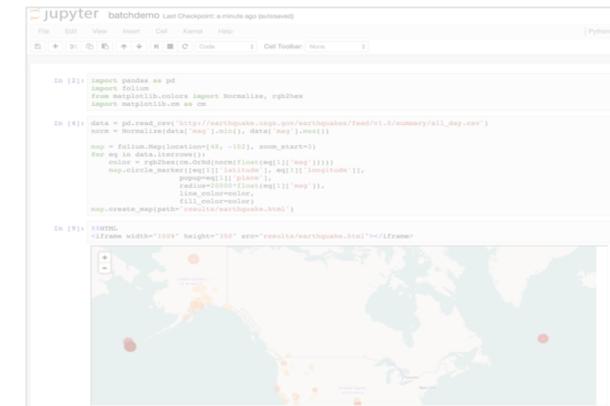
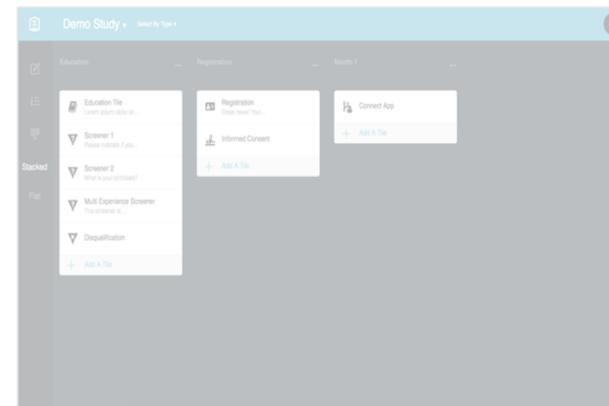
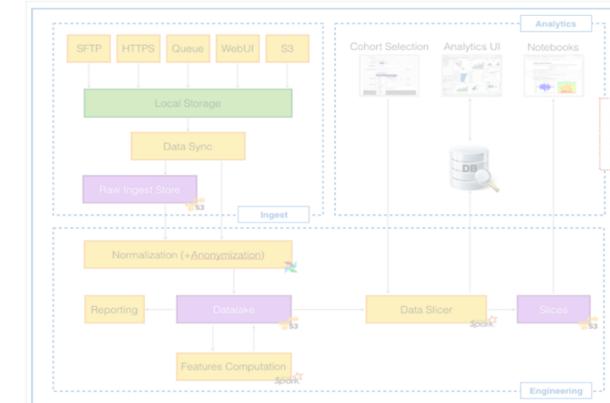
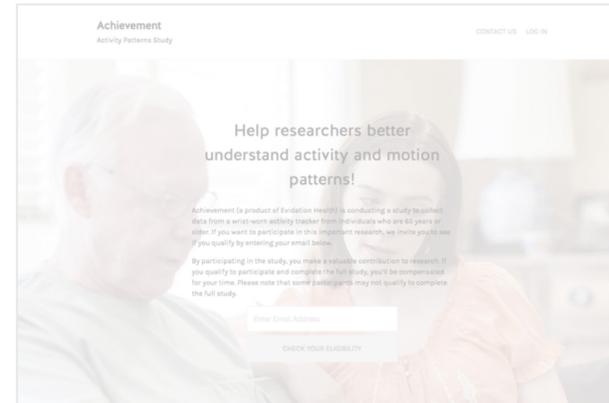
We've heard that type 2 diabetes is an area of health research. With type 2 diabetes accounting for 95% of diabetes cases, it's important to figure out there about how to best manage A1C levels. That's why we thought it would be great to partner with a company that has a management app and coaching program, called Otonomi, to see in a scientific way how it might benefit people with type 2 diabetes.



You can see that the orange line begins near the blue line, but around the time the world series begins, the lines begin to diverge. Achievers are sleeping less than they normally do in LA and Houston when the World Series is happening compared to Achievers in other regions of the country! A pretty neat result!

calories trick-or-treating? The answer is yes! We saw that overall average steps/day increase across all states in the US on Halloween.

While everyone seems to be walking around more on Halloween, moms take the real trophy for Halloween steps. Moms walk a whopping 29% more than usual on Halloween, while dads take only 18% more steps than they do on average on the holiday.



Data Connectivity

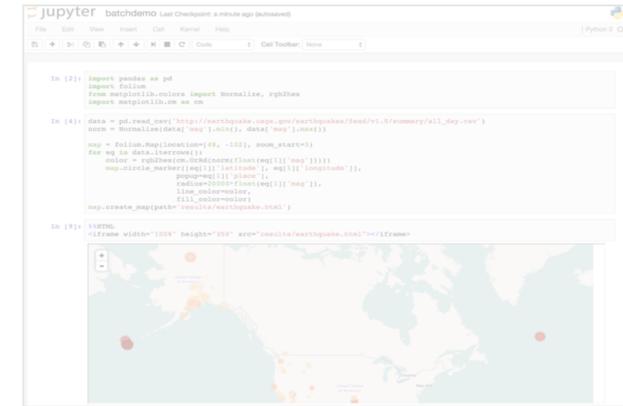
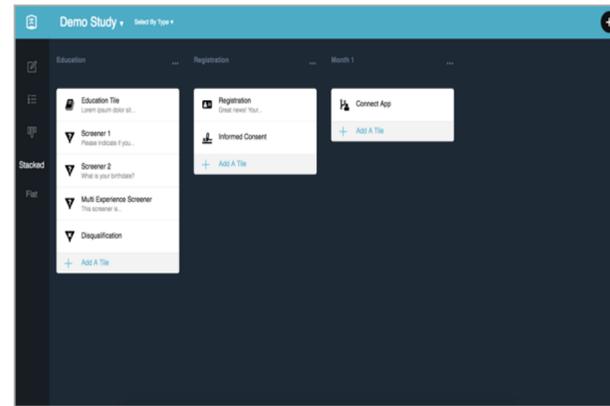
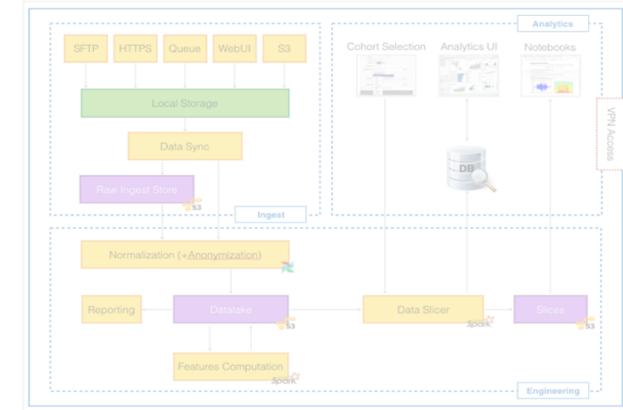
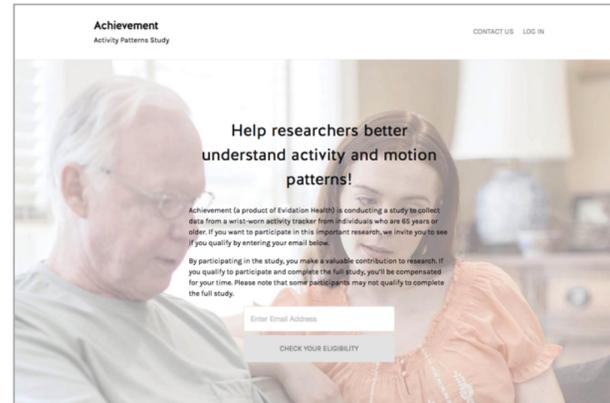
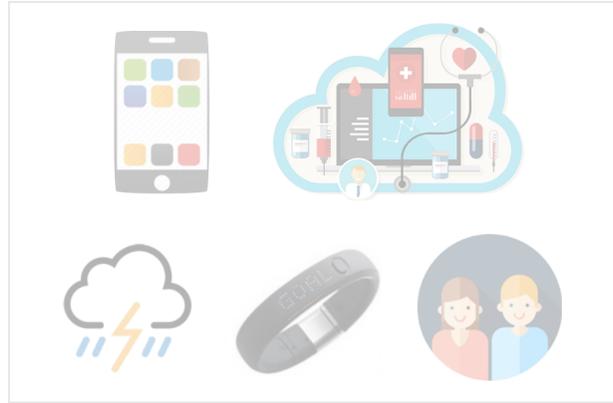
100+ Devices/Apps
Highly Scalable
Weather, Census, EMR

Study Management

Design, Management
Multi-Site, Physical/Virtual
Protocol + IRB Approval

Analysis Platform

HIPAA Compliant
Exploration & Research
Python, R, MATLAB



Data Connectivity

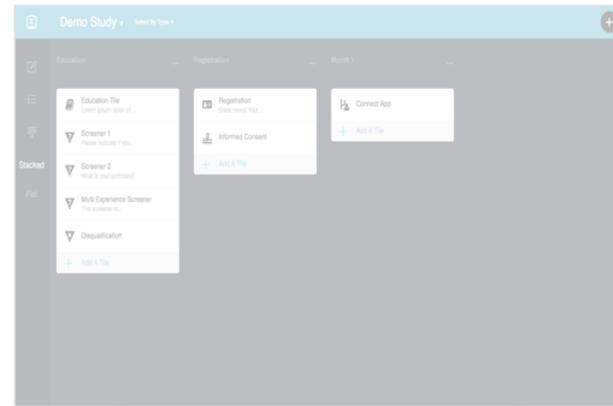
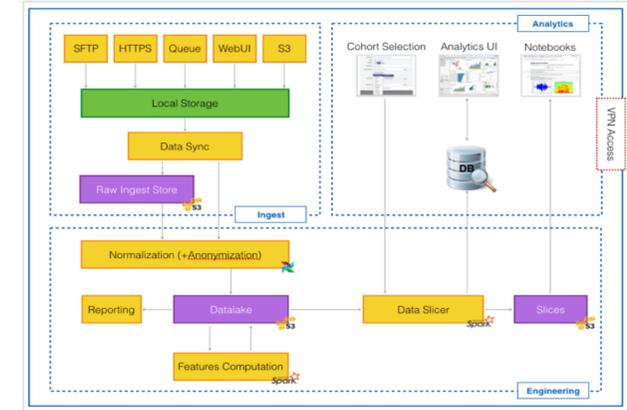
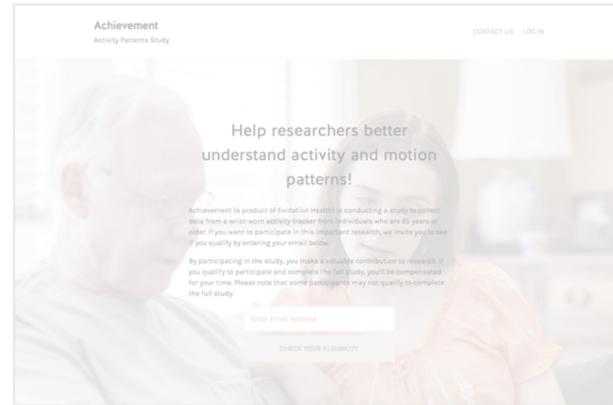
100+ Devices/Apps
Highly Scalable
Weather, Census, EMR

Study Management

Design, Management
Multi-Site, Physical/Virtual
Protocol + IRB Approval

Analysis Platform

HIPAA Compliant
Exploration & Research
Python, R, MATLAB



```

jupyter batchdemo Last Checkpoint: a minute ago (autosaved)
File Edit View Insert Cell Kernel Help Python 2.0
In [2]: import pandas as pd
import folium
from matplotlib.colors import Normalize, rgb2hex
import matplotlib.cm as cm

In [4]: data = pd.read_csv('https://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/all_day.csv')
norm = Normalize(data['mag'], data['mag'].max())

map = folium.Map(location=(40, -102), zoom_start=3)
for eq in data.iterrows():
    color = rgb2hex(cm.OldHsvNorm(float(eq['mag'])))
    map.circle_marker([eq['lat'], eq['lon']], eq['mag'],
                    radius=5000, line_color=norm(float(eq['mag'])),
                    fill_color=norm(float(eq['mag'])),
                    fill_opacity=0.5)
map.create_map(path='results/earthquake.html')

In [19]: %HTML
<iframe width="100%" height="350" src="results/earthquake.html" />
    
```

Data Connectivity

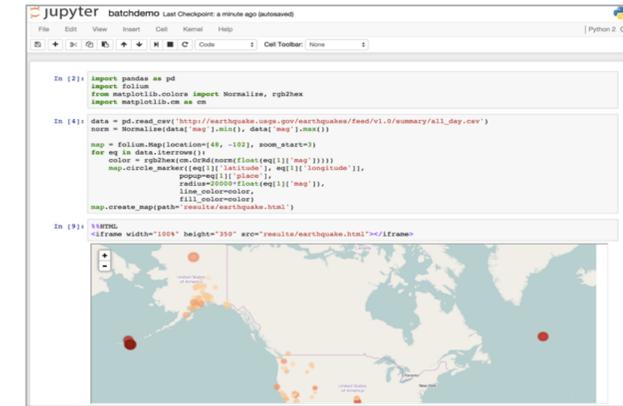
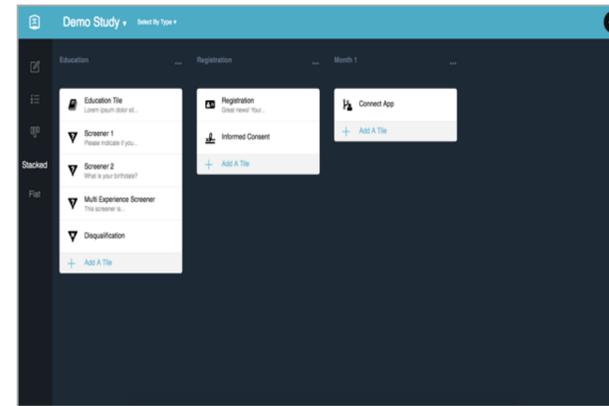
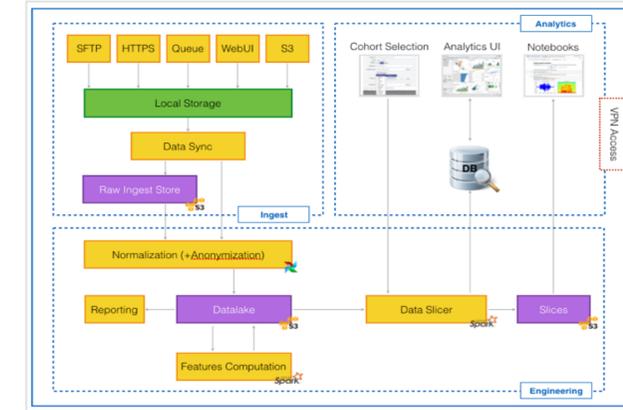
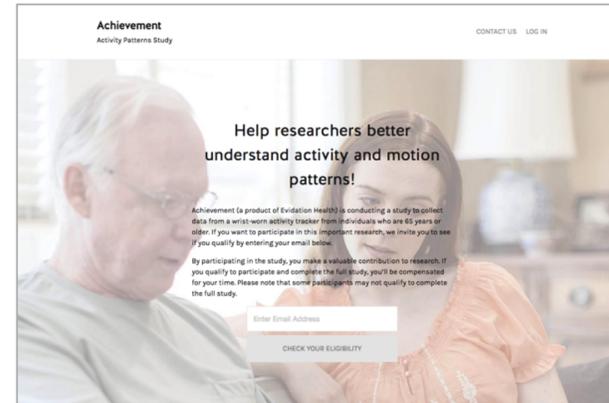
100+ Devices/Apps
Highly Scalable
Weather, Census, EMR

Study Management

Design, Management
Multi-Site, Physical/Virtual
Protocol + IRB Approval

Analysis Platform

HIPAA Compliant
Exploration & Research
Python, R, MATLAB



Data Connectivity

100+ Devices/Apps
Highly Scalable
Weather, Census, EMR

Study Management

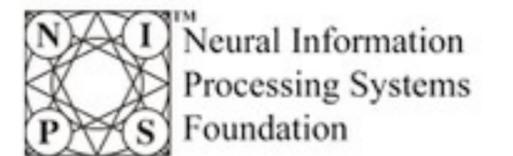
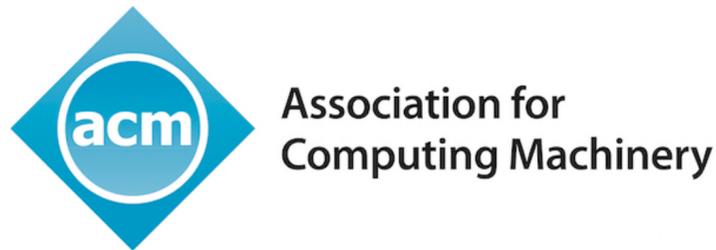
Design, Management
Multi-Site, Physical/Virtual
Protocol + IRB Approval

Analysis Platform

HIPAA Compliant
Exploration & Research
Python, R, MATLAB

Published 35+ Papers

in Medical Journals and Conferences



evidation.com/research

Study? Research? AI?

Let's talk!



alessio@evidaction.com
www.alessiosignorini.com