Can GIS help provide better health ?

Roberto Lattuada - CEO
myHealthbox
Milan
Italy

GISTAM 2016

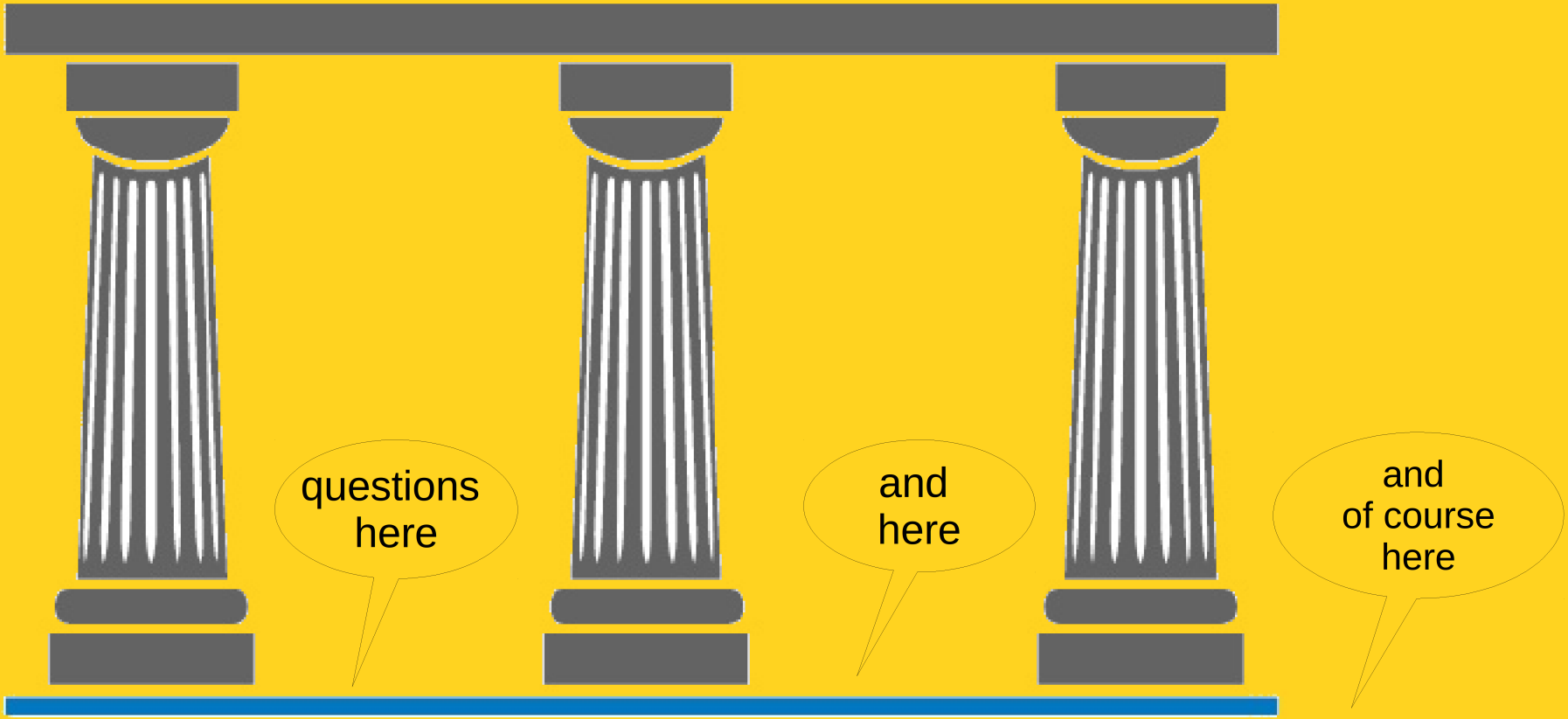
Abstract

The need to provide better, personal health information and care at lower costs is opening the way to a number of digital solutions and applications designed and developed specifically for the healthcare. Under the mHealth (Health 2.0) umbrella we find web and mobile applications and services that can help manage chronic conditions or improve ones lifestyle; at the same time doctors, hospitals and health-services are using digital devices and solutions to improve quality of care, improve efficiency and reduce costs.

But what happens when we apply GIS techniques to health data ?

MedsTrends: an analytical tool to map and analyze access to health-related information.

Summary

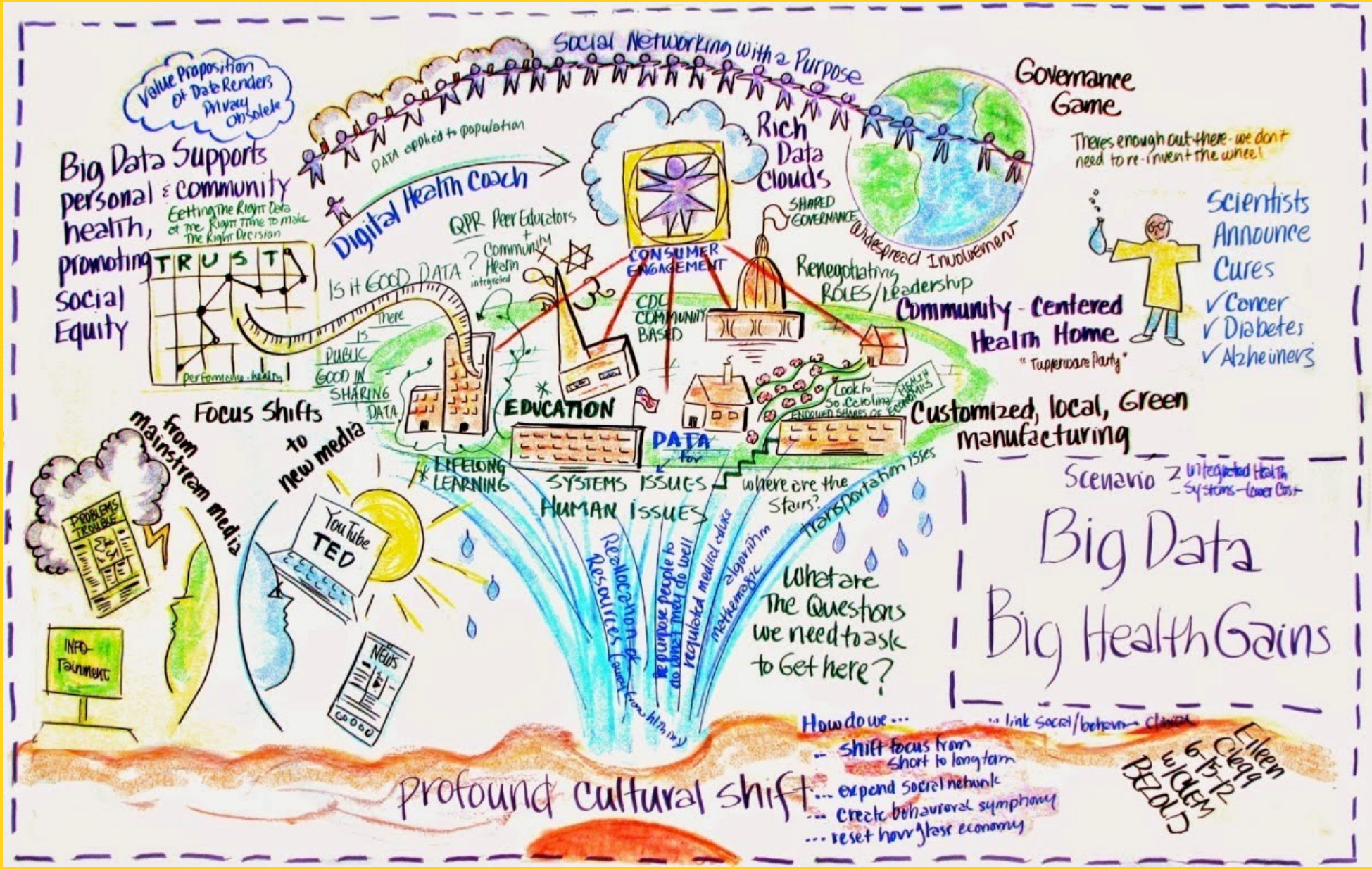

questions here
and here
and of course here
Who we are
What others are doing
What we are doing

We are a Health 2.0 company

Health 2.0 - Ideal scenario

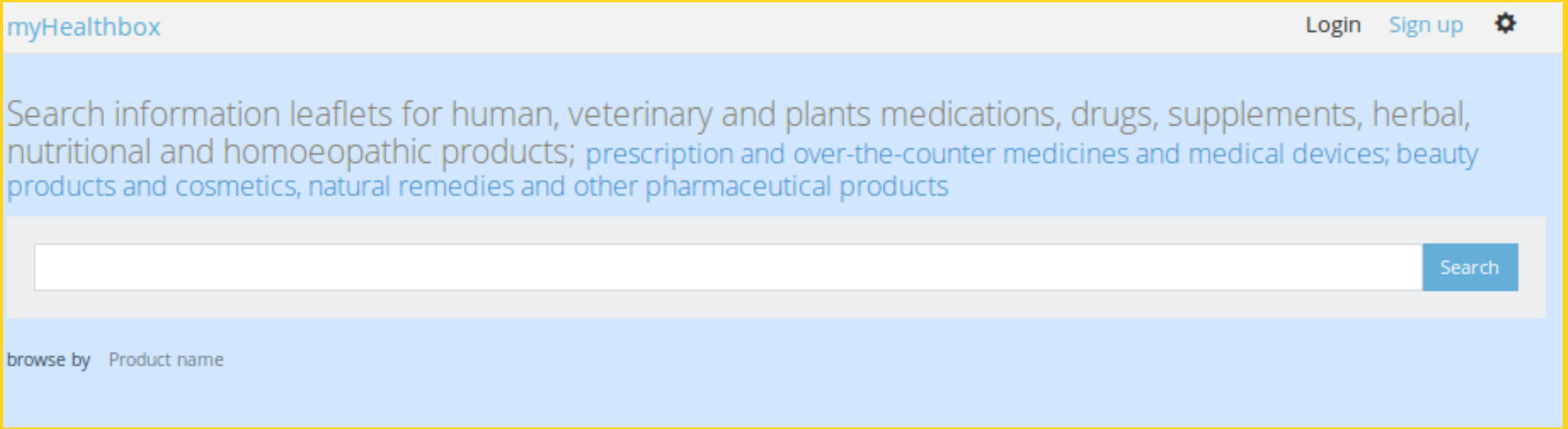
Health 2.0 - Realistic scenario

Health 2.0 - Reality



What is myHealthbox

Essentially a search engine, like Google but specialized in healthcare products, plus a huge database of information about those product.



myHealthbox – extended info

myHealthbox is a leading provider of health information services to consumers and healthcare professionals, our platform includes.

- a **specialized search engine** to find and retrieve information on how to use medicines and healthcare products
- A **website** where companies can **add**, **update** and **publish information** and documents about their products
- a set of W**eb-API**s providing access to our high-quality content to third parties (health web sites, web and mobile applications, company web sites)

For any company in the healthcare space joining the myHealthbox platform means being able to provide to the general public and to healthcare professionals high quality, guaranteed information that can be easily updated and is optimized for access via any web or mobile device.

What information do we provide?

Official information on the use of medicines and any healthcare product (*human, animal and plants medications; herbal, homeopathic and natural remedies; prescription and over-the-counter; beauty products and cosmetics; supplements and medical devices*)

Targeted at the general public and healthcare professionals

Mainly PIL (like the paper leaflets you find in most medicines packaging) and SPC documents in digital format (pdf or eLeaflet)

Cardioaspirin

compresse gastroresistenti 100mg

acido acetilsalicilico

Aa Aa Aa

FOGLIO ILLUSTRATIVO | © PRODUTTORE

BAYER

Science For A Better Life

Legga attentamente questo foglio prima di prendere questo medicinale perché contiene importanti informazioni per lei.

- Conservi questo foglio. Potrebbe aver bisogno di leggerlo di nuovo.
- Se ha qualsiasi dubbio, si rivolga al medico o al farmacista.
- Questo medicinale è stato prescritto soltanto per lei. Non lo dia ad altre persone, anche se i sintomi della malattia sono uguali ai suoi, perché potrebbe essere pericoloso.
- Se si manifesta un qualsiasi effetto indesiderato, compresi quelli non elencati in questo foglio, si rivolga al medico o al farmacista. Vedere paragrafo 4.

Contenuto di questo foglio:

1. Che cos'è Cardioaspirin e a cosa serve

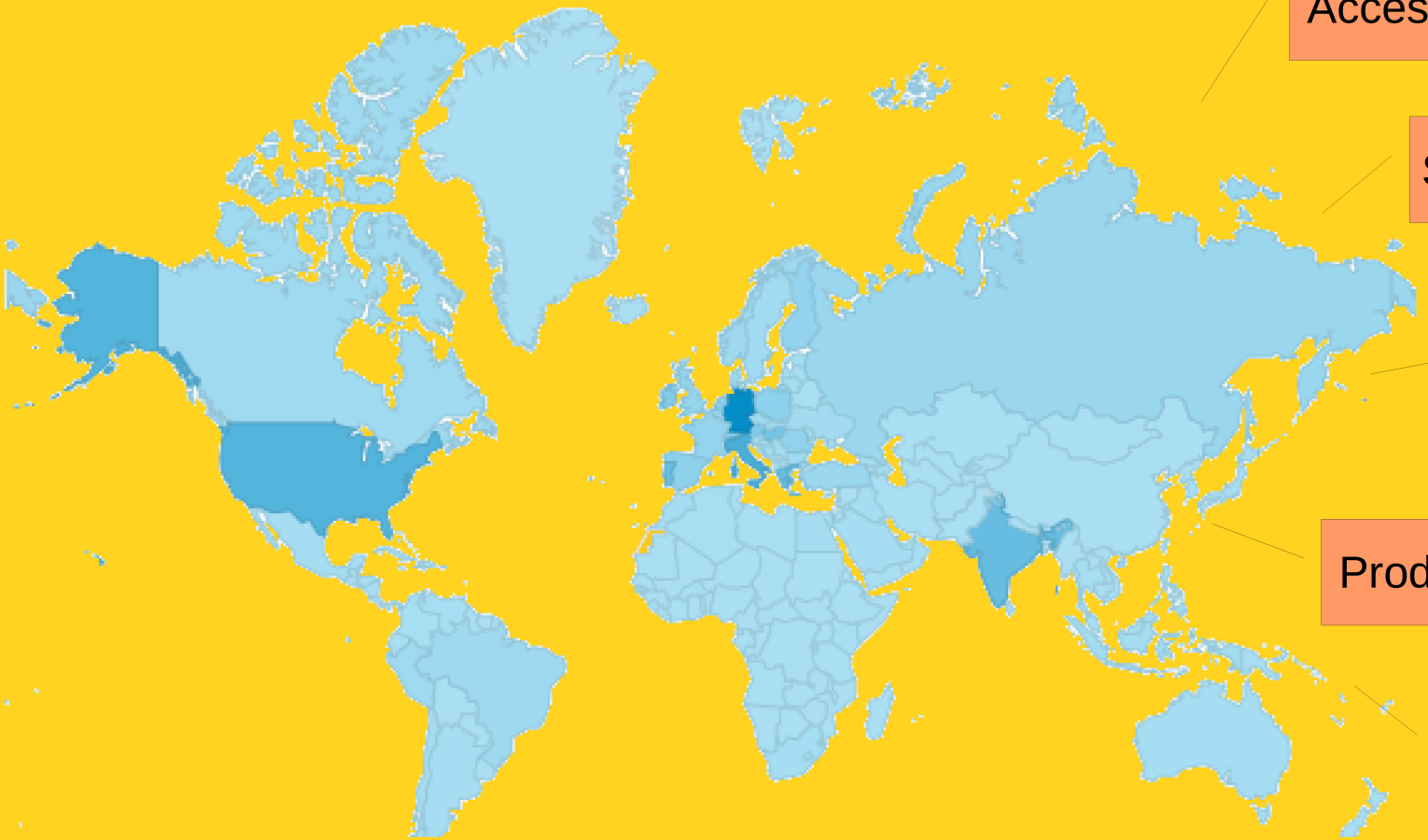
Cardioaspirin è un antitrombotico, cioè un medicinale che previene la formazione di coaguli di sangue (trombi) nei vasi sanguigni.
Cardioaspirin si usa nella:

- Prevenzione degli eventi atero-trombotici maggiori:
  - - dopo infarto del miocardio (attacco di cuore);
  - - dopo ictus cerebrale (improvvisa interruzione del flusso di sangue al cervello) o attacchi ischemici transitori (TIA) (temporanea interruzione o riduzione del flusso di sangue al cervello);
  - - in pazienti con angina pectoris instabile (attacchi che si manifestano con dolore e senso di oppressione al torace in corrispondenza dello sterno, a riposo);
  - - in pazienti con angina pectoris stabile cronica (o angina da sforzo, che si manifesta con forte dolore toracico e senso di oppressione nella regione dietro lo sterno, generalmente in seguito ad uno sforzo).
- Prevenzione della riocclusione dei by-pass aorto-coronarici (la tecnica chirurgica che consente di ripristinare il corretto afflusso di sangue al cuore) e nell'angioplastica coronarica percutanea transluminale (PTCA) (la tecnica d'intervento non chirurgica che consente di ripristinare il corretto afflusso di sangue al cuore).
- Prevenzione degli eventi cardiovascolari nei pazienti con malattia ateromasica conclamata (malattia che porta all'irrigidimento ed al deposito di grassi all'interno delle arterie), nei pazienti sottoposti alla dialisi del sangue e nella prevenzione della trombosi durante circolazione extracorporea.
- Prevenzione degli eventi cardiovascolari in pazienti ad elevato rischio.

The type of information that helps people use products in a safer way (dosage, excipients for allergies, therapeutic indications, side effects etc.)

"better information for better health"

Reach & numbers



Accesses from 200 countries

Search in 35 languages

UI in 35 languages

Products from 56 countries

1.5 million documents

2 million products

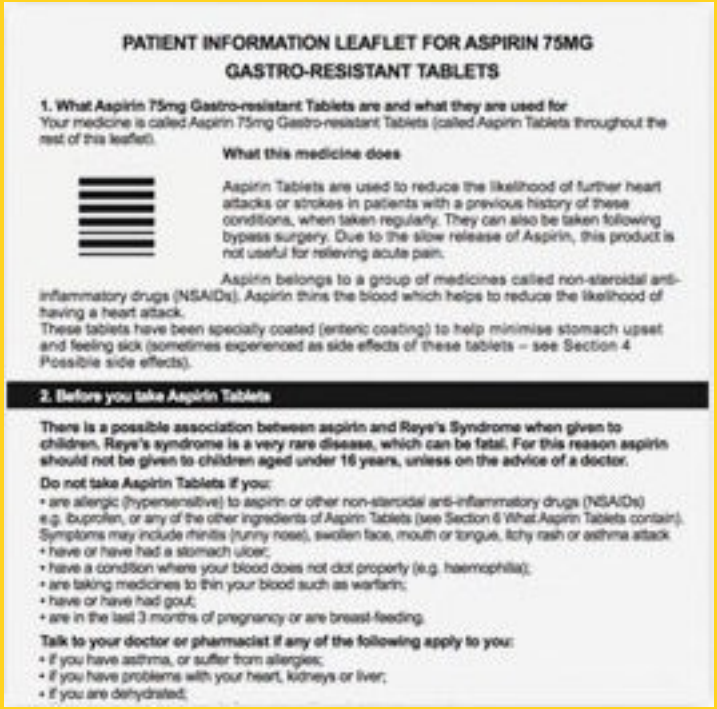
Over 1 million product leaflets read every month

What people do on our sites

Search information leaflets for human, veterinary and plants medications, drugs, supplements, herbal, nutritional and homoeopathic products; prescription and over-the-counter medicines and medical devices; beauty products and cosmetics, natural remedies and other pharmaceutical products

Fever flu symptoms Search

browse by Product name



PATIENT INFORMATION LEAFLET FOR ASPIRIN 75MG GASTRO-RESISTANT TABLETS

1. What Aspirin 75mg Gastro-resistant Tablets are and what they are used for
Your medicine is called Aspirin 75mg Gastro-resistant Tablets (called Aspirin Tablets throughout the rest of this leaflet).

What this medicine does

Aspirin Tablets are used to reduce the likelihood of further heart attacks or strokes in patients with a previous history of these conditions, when taken regularly. They can also be taken following bypass surgery. Due to the slow release of Aspirin, this product is not useful for relieving acute pain.

Aspirin belongs to a group of medicines called non-steroidal anti-inflammatory drugs (NSAIDs). Aspirin thins the blood which helps to reduce the likelihood of having a heart attack.
These tablets have been specially coated (enteric coating) to help minimise stomach upset and feeling sick (sometimes experienced as side effects of these tablets – see Section 4 Possible side effects).

2. Before you take Aspirin Tablets

There is a possible association between aspirin and Reye's Syndrome when given to children. Reye's syndrome is a very rare disease, which can be fatal. For this reason aspirin should not be given to children aged under 16 years, unless on the advice of a doctor.

Do not take Aspirin Tablets if you:
- are allergic (hypersensitive) to aspirin or other non-steroidal anti-inflammatory drugs (NSAIDs) e.g. ibuprofen, or any of the other ingredients of Aspirin Tablets (see Section 6 What Aspirin Tablets contain). Symptoms may include rhinitis (runny nose), swollen face, mouth or tongue, itchy rash or asthma attack
- have or have had a stomach ulcer;
- have a condition where your blood does not clot properly (e.g. haemophilia);
- are taking medicines to thin your blood such as warfarin;
- have or have had gout;
- are in the last 3 months of pregnancy or are breast-feeding.

Talk to your doctor or pharmacist if any of the following apply to you:
- if you have asthma, or suffer from allergies;
- if you have problems with your heart, kidneys or liver;
- if you are dehydrated;

Show only results in English

Only products available in United Kingdom

1) search

2) refine

3) read

Logging access data

Disclaimer
For the sake of privacy rules and sensitive data handling: registration on our platform is optional, you do not have to provide any personal data if you do not want to.

We do log everything (and I really mean everything) that is done on our platform:

- what users search
- what they read
- how long they read a specific document for
- when
- where (down to zip code precision in some regions)

All this log data is collected and made available to our analytics tools,

The Data Problem

We have the answer

...

what is your question ?

The Data problem

We have the 3 Vs (Big data ?)

Volume: The quantity of generated and stored data. The bigger the better, in our case millions of interactions
Variety: The type and nature of the data (several types of interactions to weight, several languages
Velocity: the speed at which the data is generated generated in real time 24/7, live data

The Data Problem is not a storage problem but, as typical with most big data problems, we need to find the answers (**value indicators**) that are hidden under millions and millions of records

extract value from data

This should lead to more confident decision making and thus greater operational efficiency, cost reduction and reduced risk and ...

End of part 1 - Questions ?

Before we move to part 2, are there any questions ?

Some projects worth mentioning

- Google Flu Trends and Google Dengue Trends
- healthmap.org
- Columbia School of Public Health

and many more of course …

Google Flu Trends - Intro

Ended as a modeling and prediction tool in 2015

Still going as a data collection source

Data is available to third parties

“When a small team of software engineers first started working on Flu Trends in 2008, we wanted to ***explore how real-world phenomena could be modeled using patterns in search queries****. Since its launch, Google Flu Trends has provided useful* ***insights*** *and served as one of the early examples for “****now-casting****” **based on search trends**, which is increasingly used in health, economics, and other fields.*

Over time, we’ve used search signals to create prediction models, updating and improving those models over time as we compared our prediction to real-world cases of flu.”

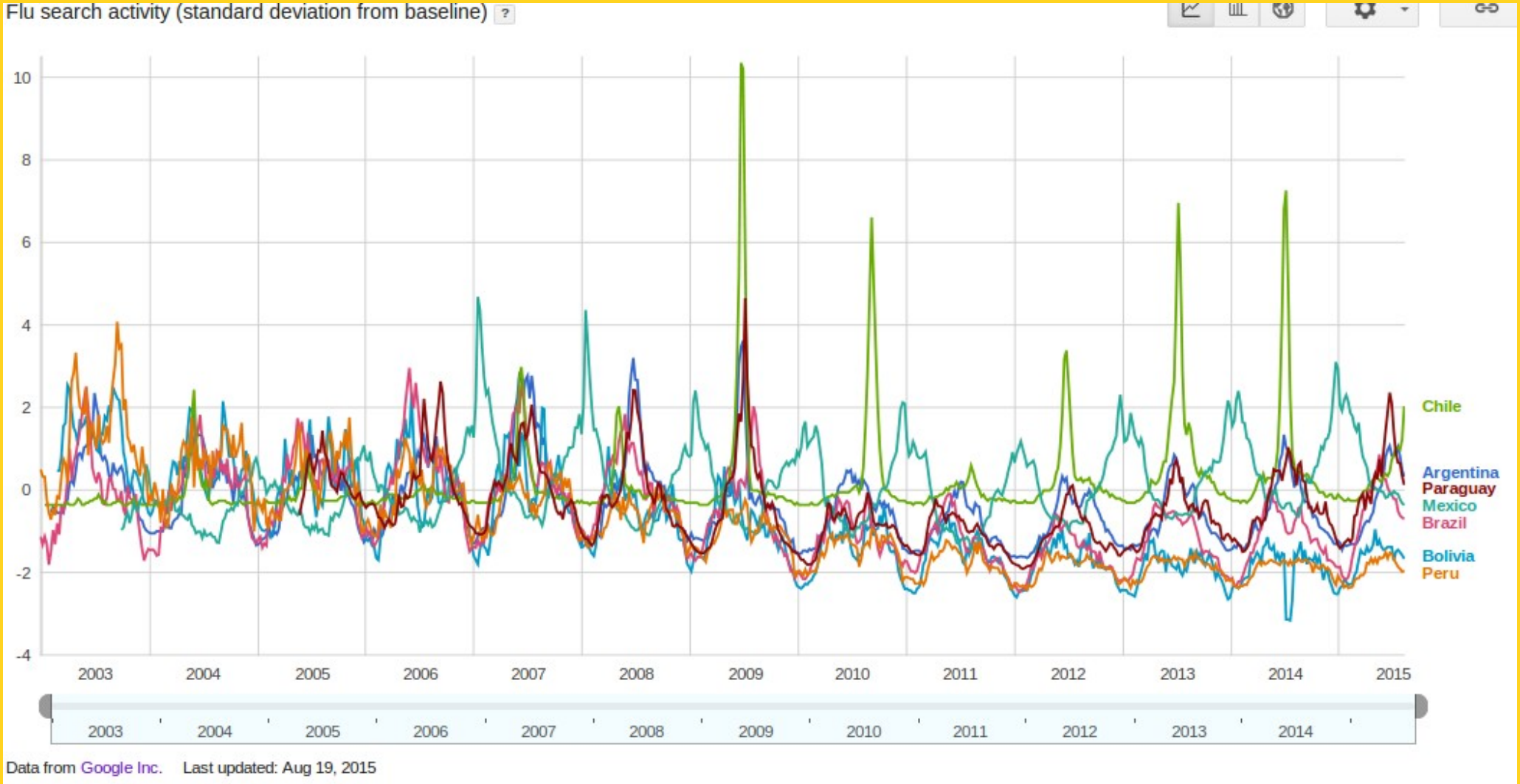
Google Flu Trends - status

“Instead of maintaining our own website going forward, we’re now going to empower institutions who specialize in infectious disease research to use the data to build their own models. Starting this season, we’ll provide Flu and Dengue signal data directly to partners including Columbia University’s Mailman School of Public Health (to update their dashboard), Boston Children’s Hospital/Harvard, and Centers for Disease Control and Prevention (CDC) Influenza Division. We will also continue to make historical Flu and Dengue estimate data available for anyone to see and analyze.”

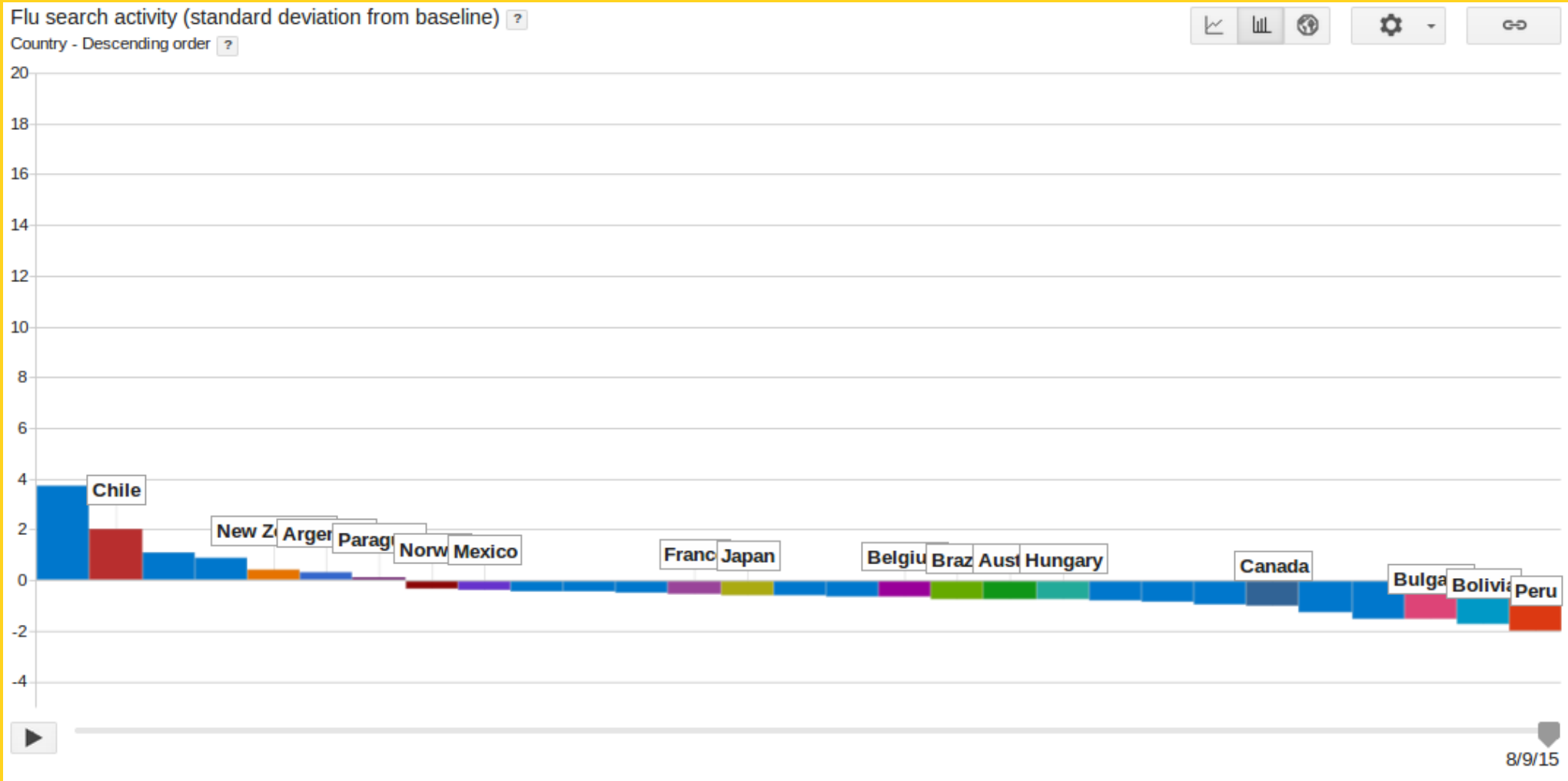
Google Flu Trends - method

Flu search activity **based on aggregated Google Search query data**, standardized to make data more comparable across regions. The '**baseline**' level for each region (shown as 0) is its **average flu search activity**, measured over many seasons. **Activity** levels for each region **represent how much flu search activity differs** from that region’s 'baseline' level.

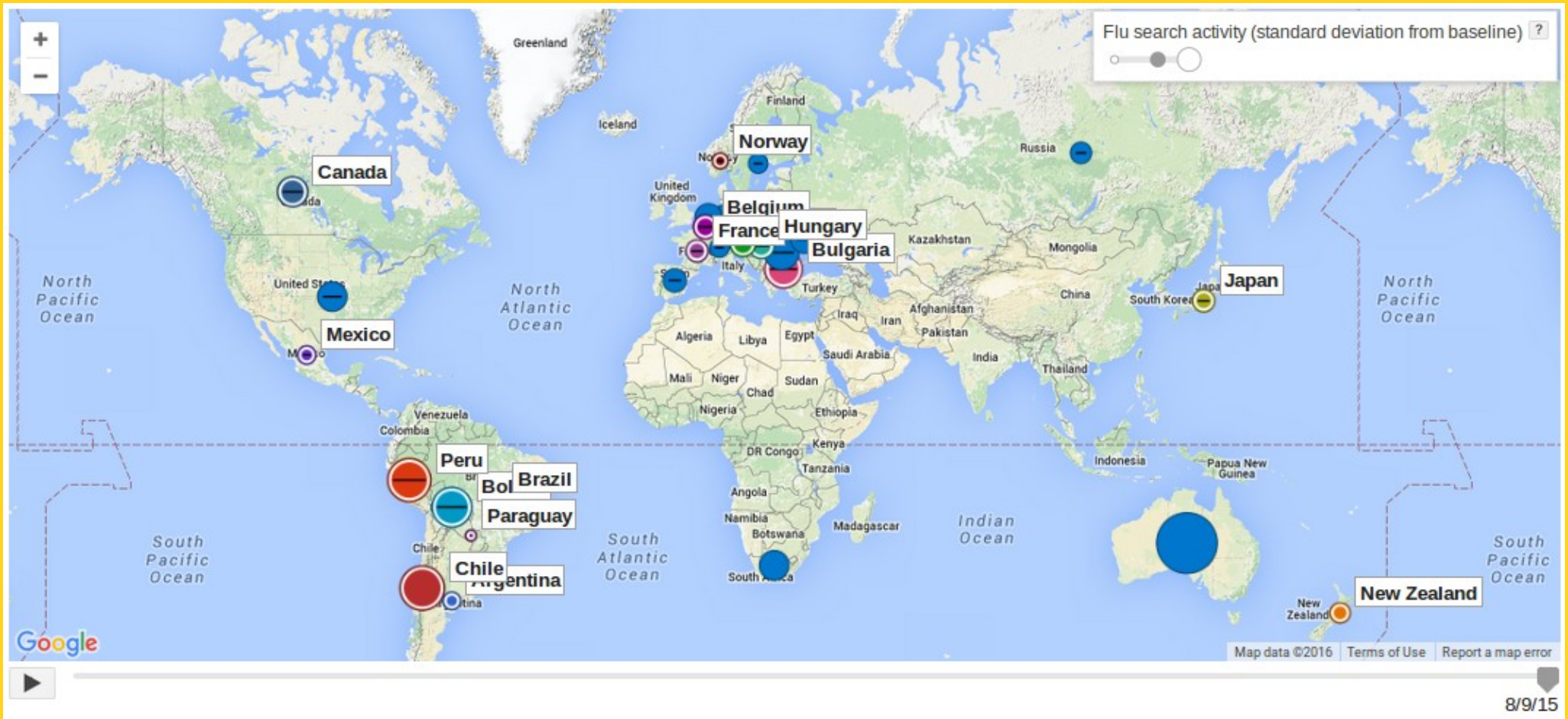
Google Flu Trends - Results



Google Flu Trends - Results

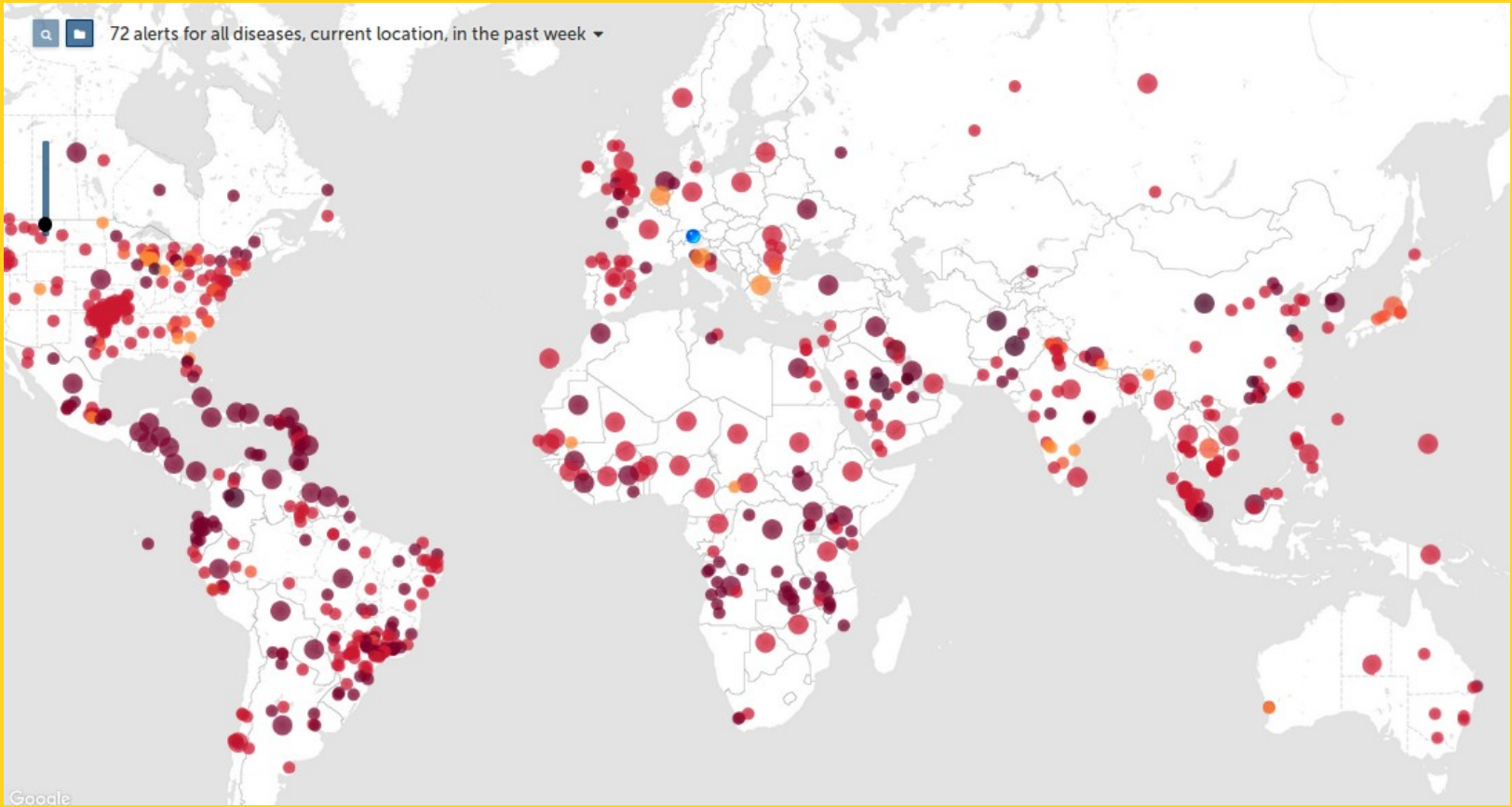


Google Flu Trends - Results



healthmap.org - Intro

*HealthMap, a team of researchers, epidemiologists and software developers at **Boston Children's Hospital** founded in 2006, is an established global leader in utilizing on-line informal sources for disease outbreak monitoring and real-time surveillance of emerging public health threats.*



healthmap.org - Projects

Projects

- Crowd Clinical
- HealthMap Flu Trends
- DengueMap
- MedWatcher
- PREDICT
- Thermia
- Flu Near You
- HealthMap Vaccine Finder
- ProMED mail

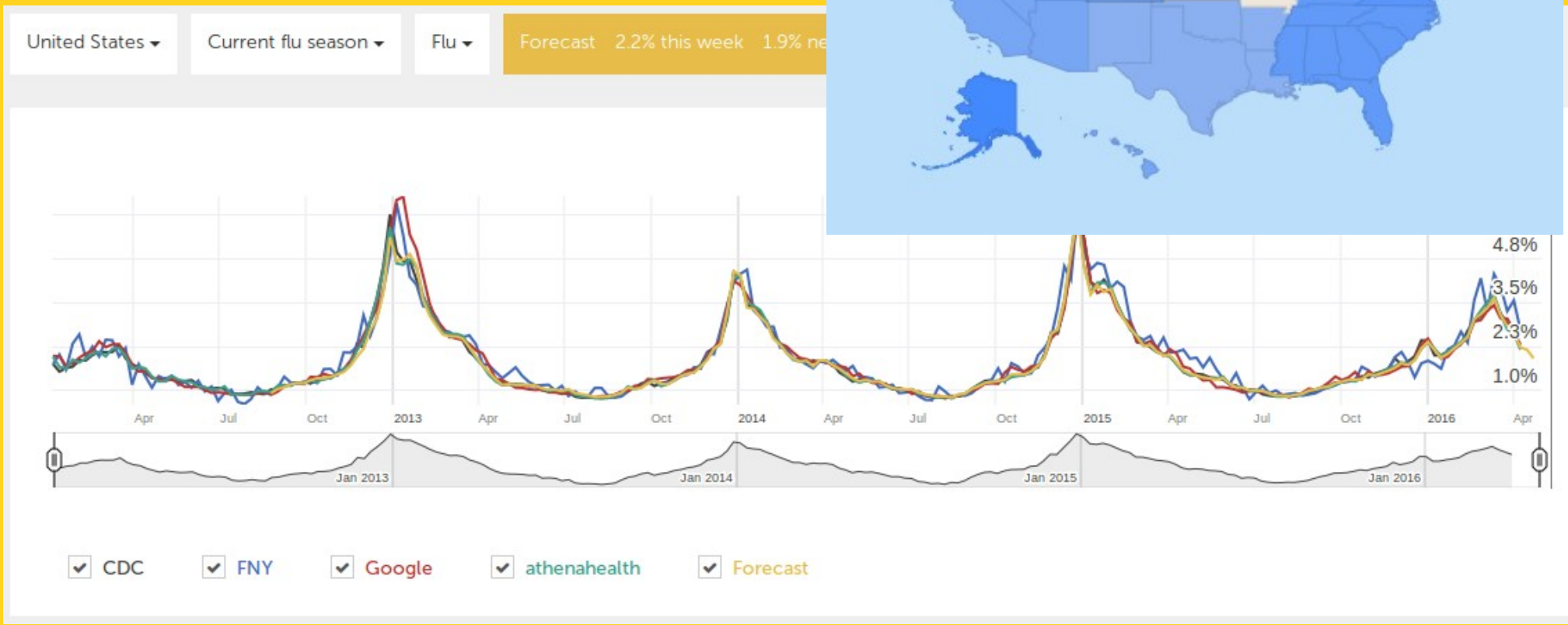
healthmap.org - Method

Data from **several sources**:

- The **US Centers for Disease Control** (CDC) continuously records the percentage of patients seen in clinics who exhibit influenza-like illnesses (ILI) according to physicians' reports.
- Search data provided by Google
- Flu Near You (FNY) is a crowd-sourced participatory disease surveillance system. It gathers information from users about their health in real-time.
- Athenahealth (ATH) is a company specialized in medical practices management and electronic health records management.

Produce a **forecast** by combining data from CDC, Google, athenahealth, and FNY, flu activity is estimated for last week, current week, next week, and the week after next.

healthmap.org – Results
flu activity

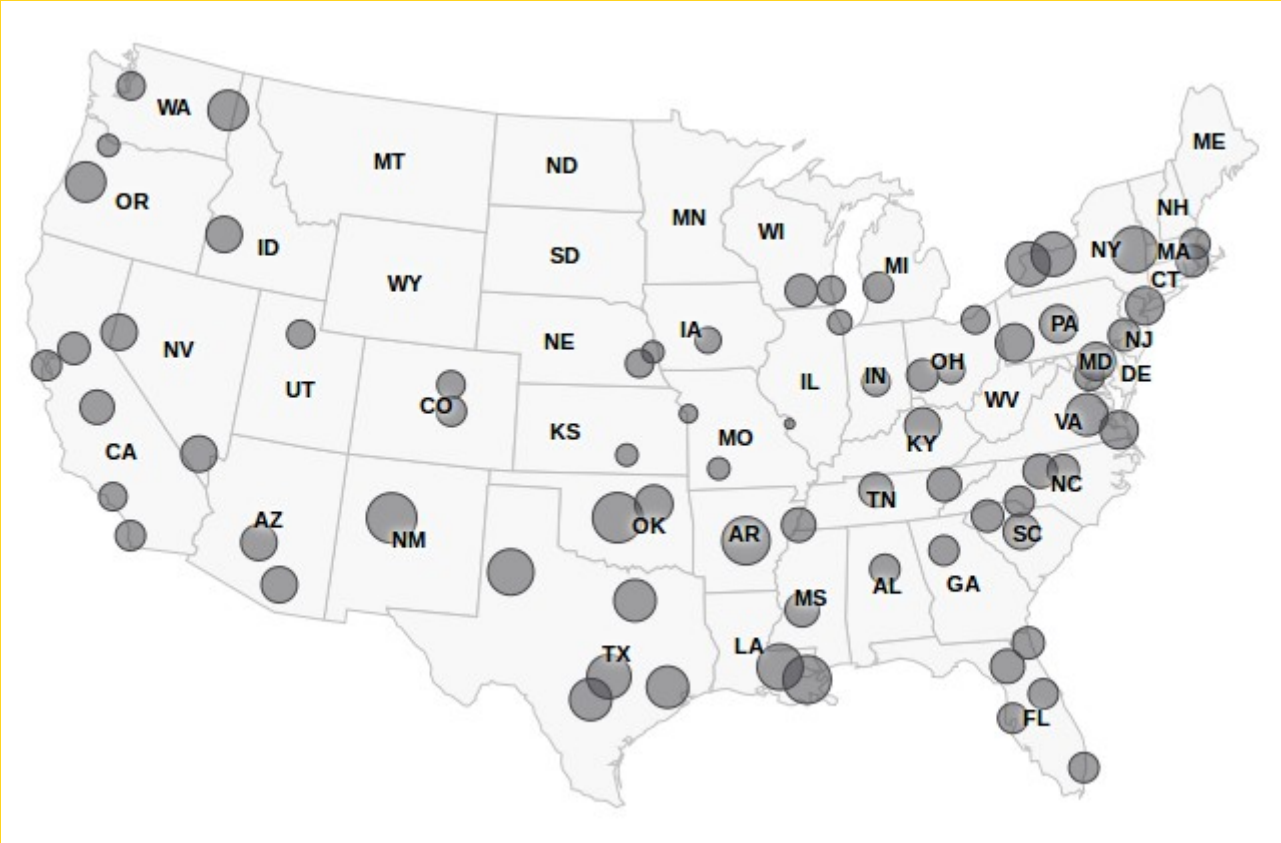


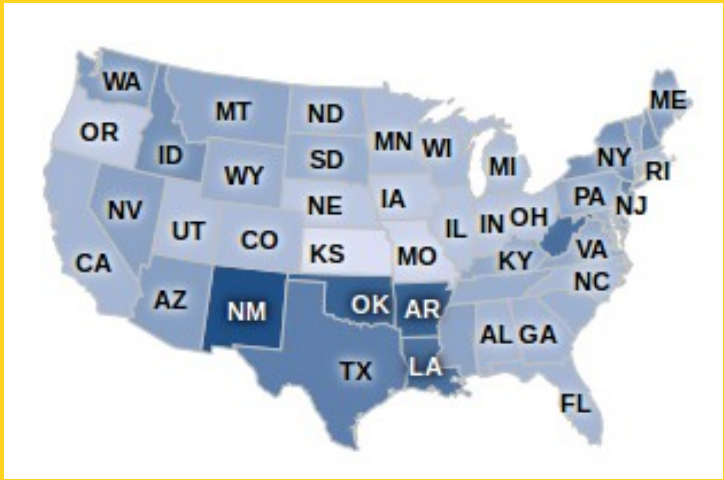
Columbia School of Public Health - Intro

Similar to healthmap but different model (added lab tests to confirm cases of flu)

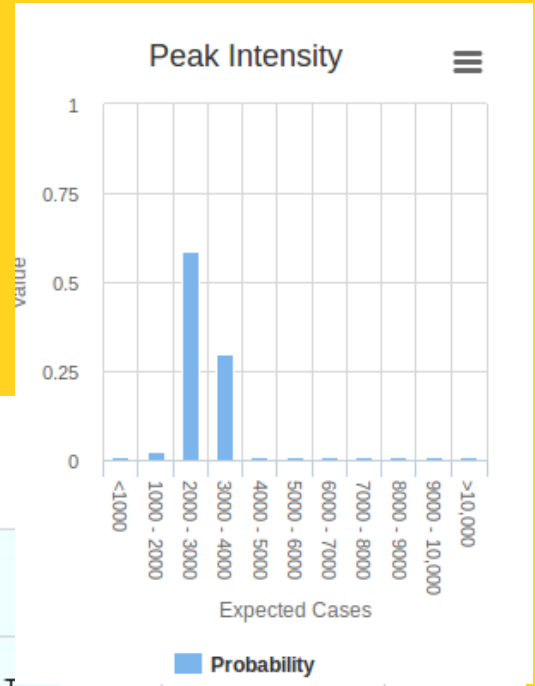
“We construct local estimates of influenza incidence at the state and municipal levels using Centers for Disease Control and Prevention (CDC) ILINet weekly regional reports of influenza like illness (ILI), WHO/NREVSS collaborating laboratory influenza positive tests and Google search activity.”

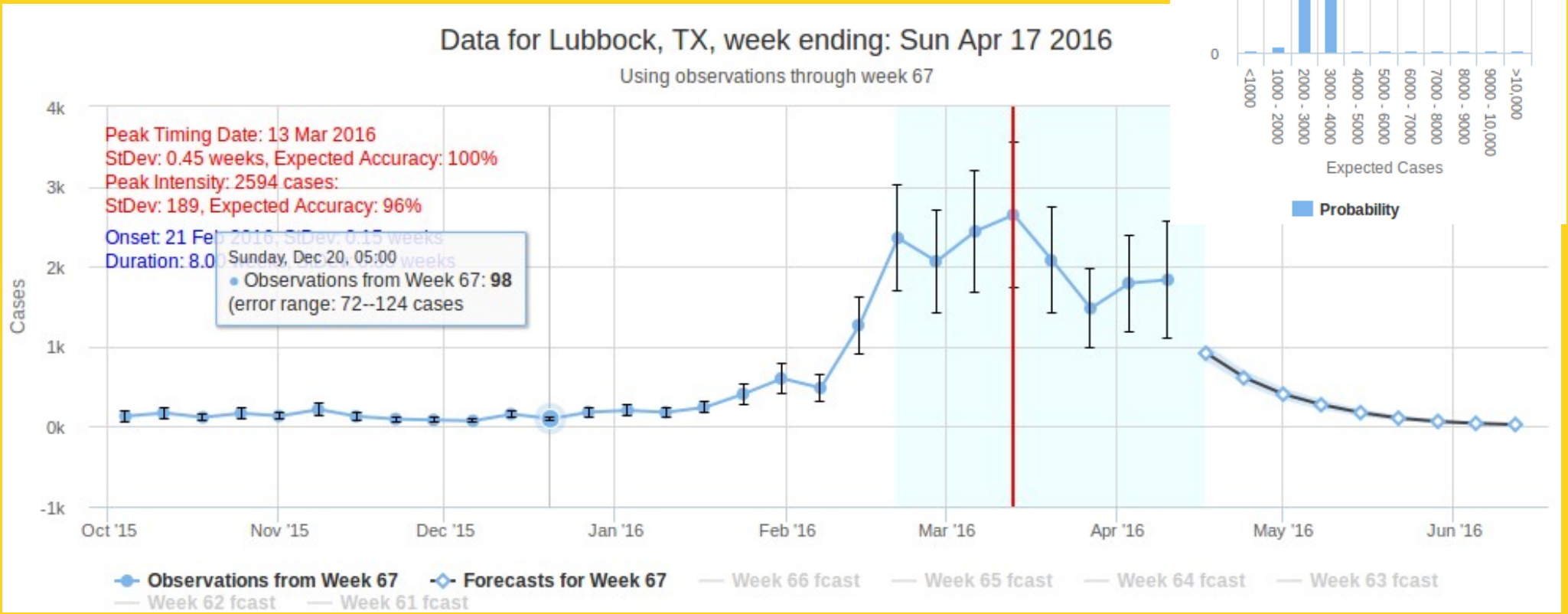
Columbia School of Public Health - Results





Columbia School of Public Health - Forecasts





Limitations of these approaches

Search data proven to be high unstable and unreliable if used by itself

"... Google Flu Trends massively inaccurately predicted the number of flu cases at the end of 2012, as first reported by Nature Magazine..."

"... A comparison with traditional surveillance data showed that Google Flu Trends, which estimates prevalence from flu-related Internet searches, ***had drastically overestimated peak flu levels****. The glitch is no more than a temporary setback for a promising strategy, experts say, and Google is sure to refine its algorithms. But as flu-tracking techniques based on mining of web data and on social media proliferate, the episode is a reminder that they will* ***complement, but not substitute for, traditional epidemiological surveillance networks****...."*

" ...Google Flu Trends had missed last season by quite a lot, over 100 percent ..."

Limitations of these approaches

- Disease centric, large epidemics (i.e. flu, Ebola)
- Spatial granularity (state/region level)
- Easily put off by media campaigns (like in the Google case)
- **Signal-to-noise ratio** is usually very low
- Demographic bias (most active web users are young adults and so are not representative of the general public)

Without context or other strong complementary signals it is very difficult to tell if a tweet message indicates that the tweeter is actually ill rather than simply pointing to news articles about flu.

Why does it matter ?

- So we can time vaccine production and distribution
- Resource planning
- Emergency plans
- Health policies and investments
- More confident decision making
- Greater operational efficiency
- Cost reduction
- Reduced risk

Overall … Big money is involved

End of part 2 - Questions ?

Before we move to part 3, are there any questions ?

What we are doing

MedsTrends - intro

An analytical tool to monitor and predict health patterns related to healthcare products use based on **health-related queries** and **information consumption**.

Status

Currently in alpha mode (development, debugging, testing...)

MedsTrends - signals

Data points:

- queries (like the Google Flu data) - low weight
- information access (documents) – high weight

We believe our signal-to-noise ratio is high as it is not based just on search queries but on the actual reading of information related to a product.

Weights are added depending on the duration of the interaction and the sequence of events.

Typical search engines record what you search and where you go next but NOT what you do at your destination.

MedsTrends – ATC code analysis

Data points:

- time
- location
- product information access

Clustering:

- by ATC code

The Anatomical Therapeutic Chemical (**ATC**) Classification System is used for the classification of active ingredients of drugs according to the organ or system on which they act and their therapeutic, pharmacological and chemical properties.

For example: acetylsalicylic acid (aspirin), has A01AD05 as a drug for local oral treatment, B01AC06 as a platelet inhibitor (heart disease) and N02BA01 as an analgesic and antipyretic.

Great method to overcome language and name differences.

MedsTrends – example 1

Compare the use of aspirin (acetylsalicylic acid) as an analgesic and antipyretic versus platelet inhibitor (*myocardial infarction, atrial fibrillation, following coronary bypass, angioplasty and stenting. It is also used as prophylaxis to prevent myocardial infarction and stroke*).

ATC code: (A01AD05 + N02BA01) vs B01AC06

Therapeutic main groups:
A01 STOMATOLOGICAL PREPARATIONS
N02 ANALGESICS
B01 ANTITHROMBOTIC AGENTS

Time: Jan-Apr 2015
Location: Italy - Hungary

MedsTrends – example 1

Use of aspirin (acetylsalicylic acid) as analgesic and antipyretic

ATC code: A01AD05 + N02BA01

Time: Jan-Apr 2015

Location: Italy vs Hungary

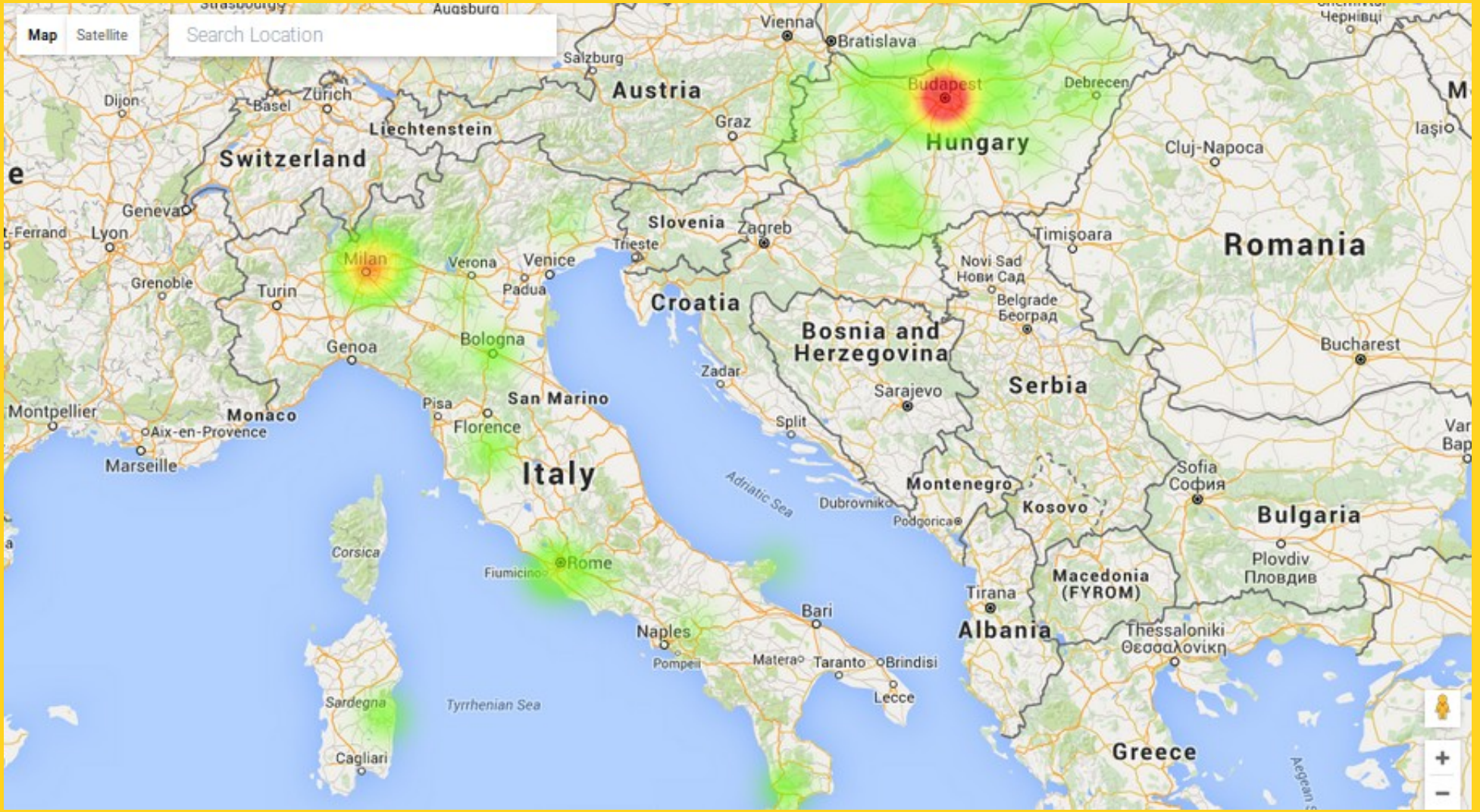
MedsTrends – example 1

Use of aspirin (acetylsalicylic acid) as platelet inhibitor

ATC code: B01AC06

Time: Jan-Apr 2015

Location: Italy vs Hungary



MedsTrends – example 2

Use of anti-acne products

ATC code: D10

Identify areas of interest for anti-acne products

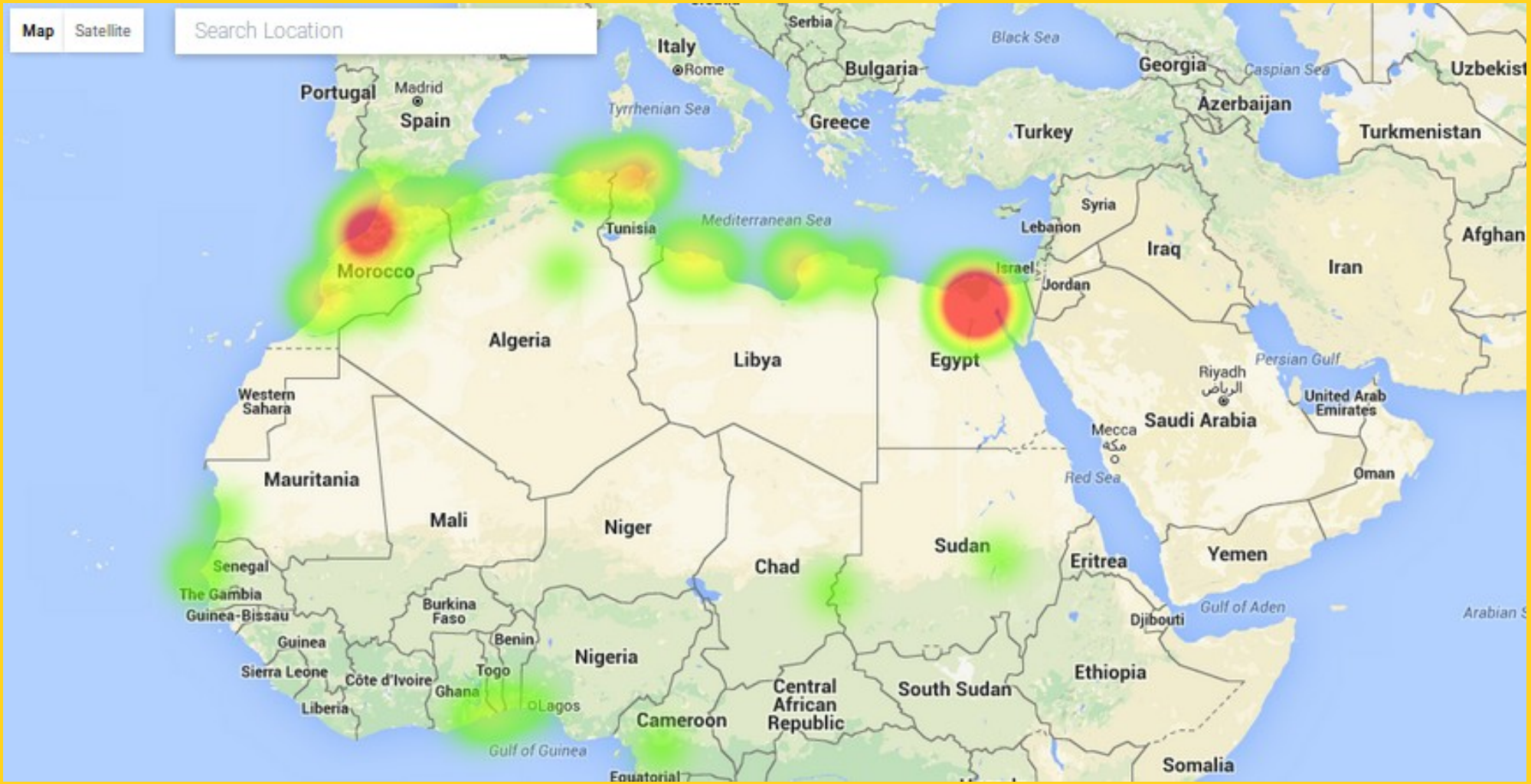
MedsTrends – example 2

Use of anti-acne products

ATC code: D10 ANTI-ACNE

Time: Jan 2014 to Dec 2015

Location: Africa



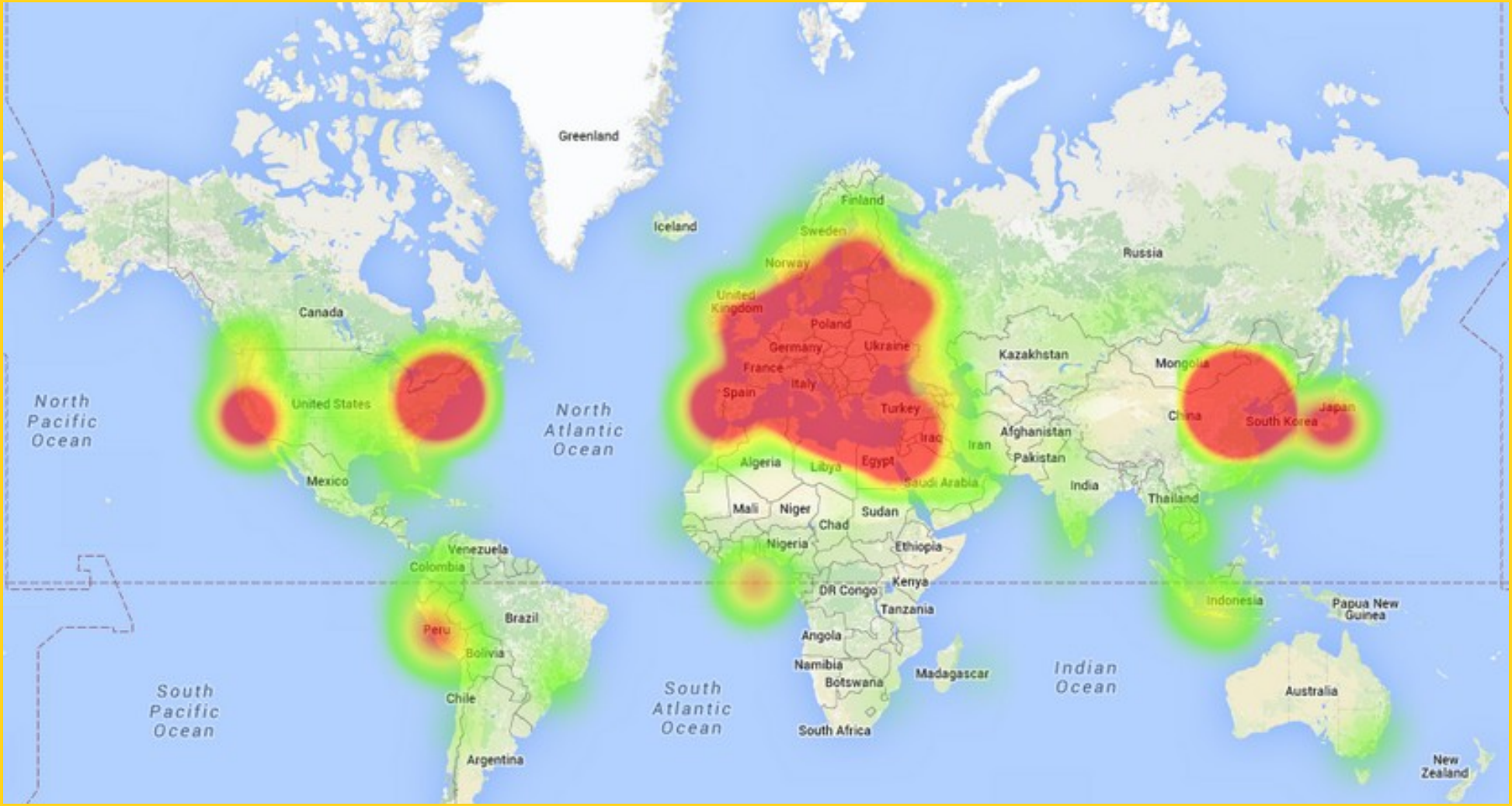
MedsTrends – example 2

Use of anti-acne products

ATC code: D10 ANTI-ACNE

Time: Jan 2015 to Dec 2015

Location: World



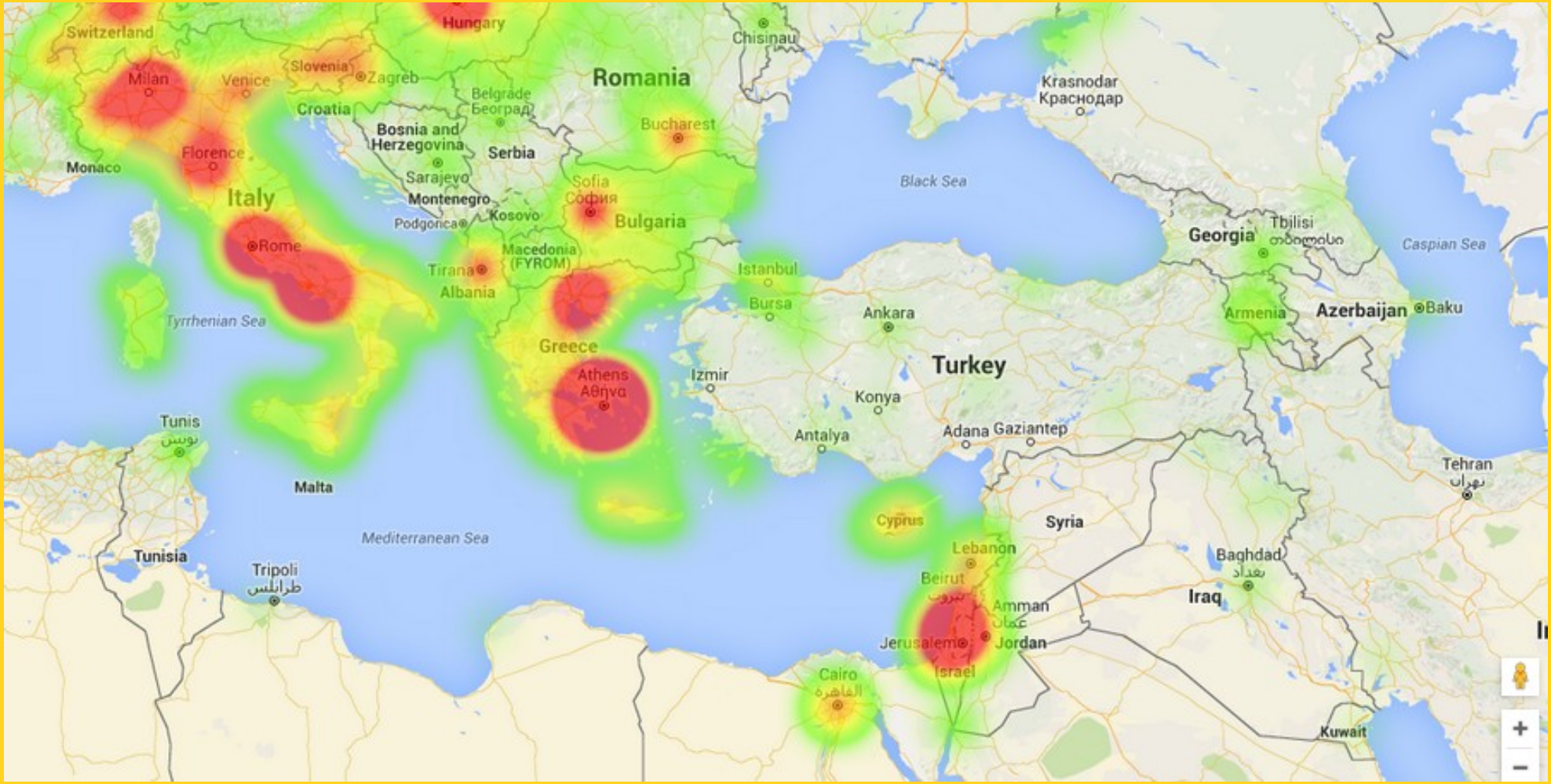
MedsTrends – example 2

Use of anti-acne products

ATC code: D10 ANTI-ACNE

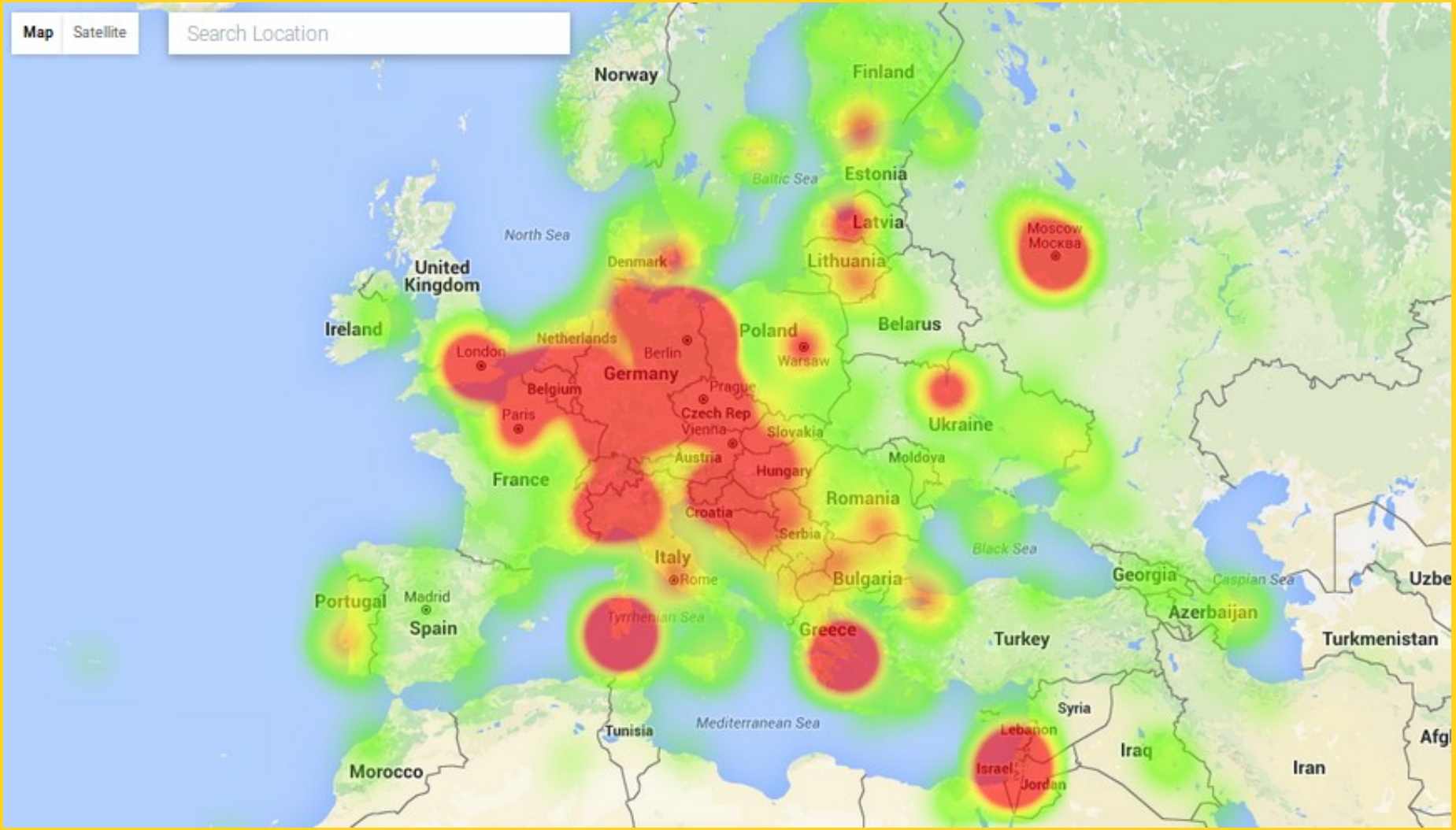
Time: Jan 2015 to Dec 2015

Location: Middle East



MedsTrends – example 3

Use of products for diabetes
ATC code: A10 (DRUGS USED IN DIABETES)
Identify areas of interest for diabetes-related products



MedsTrends
Other questions we can try to answer

Most “popular” product in each country, region by week/month/year

Most “popular” active ingredient in each country, region by week/month/year

Time lapse of seasonal healthcare products

Time lapse of seasonal healthcare illnesses

etc...

Next steps

Limitations
- Strong population density bias (more people=more searches?)
- Geographical coverage

Integrate population density data to work with (events/user) instead of total events.

Spot anomalies in data

Make our big data bigger with:

- other open datasets (Google trends data, Hospital data, Disease Centers data)
- actual product sales and usage

Find more interesting questions to answer (collaboration with epidemiology centers)

Partnerships with healthcare solutions providers

Conclusions and take away points

- 175 million **daily** searches on Google are health related (every day !)
- Huge interest from companies (producers, insurance etc...) and governments in health data
- With health data it is difficult to find good answers but a lot more difficult to find the good questions
- Education in healthcare is a big problem area: more and more people self diagnose through the Internet more and more products freely available (no doctor consultation required)

It is an incredibly interesting, innovative and growing area to work on !

Questions ?

Last chance …

Contacts

Roberto Lattuada
myHealthbox

Roberto.Lattuada@myhealthbox.eu

eleaflet.eu
myhealthbox.eu

GISTAM 2016