

Silicon Valley Big Data Science & Personal Analytics



Personal Big Data

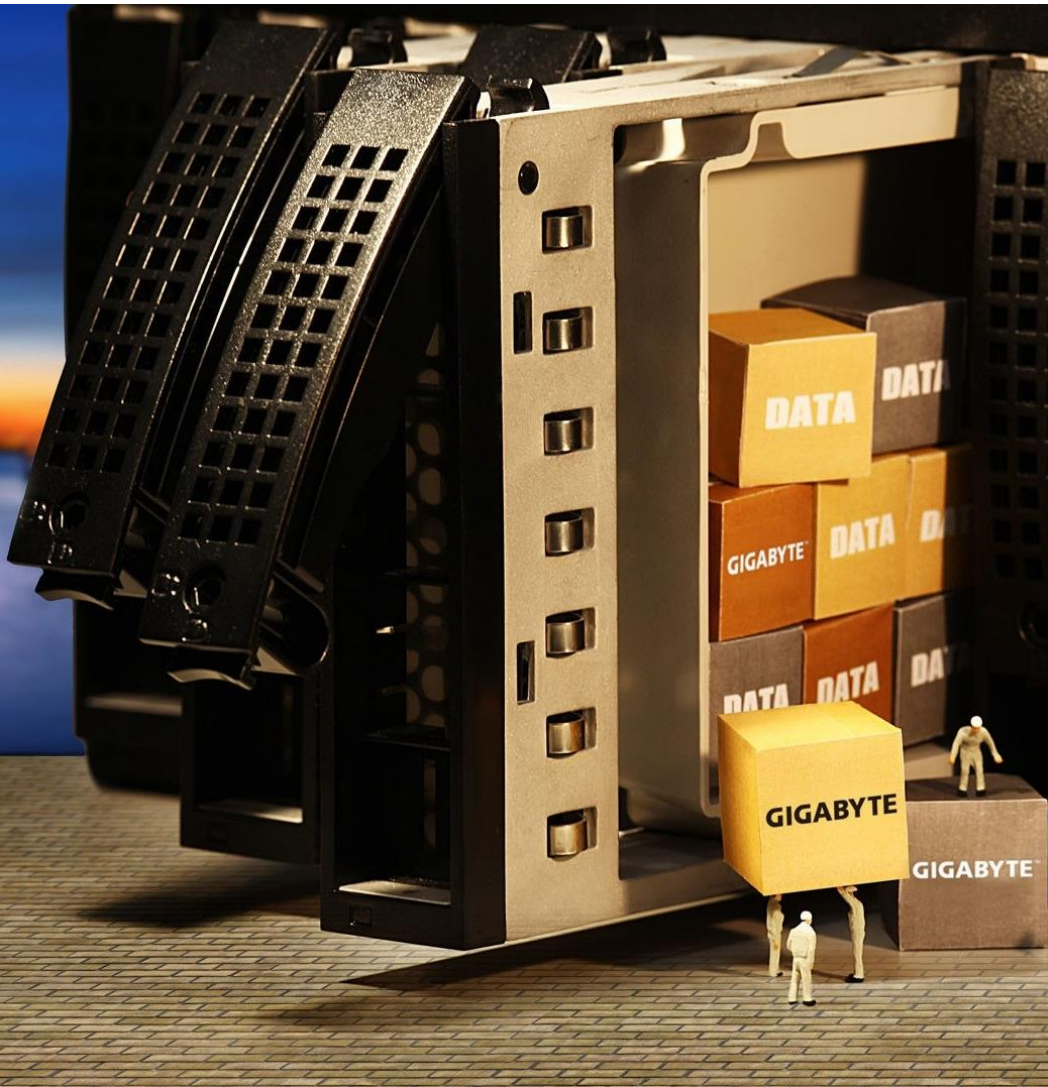
29/5/14 Sunnyvale

Meetup scope

- ▶ Mobile Users manage an always growing mass of documents and data of heterogeneous types. We want to create an active group to analyze the specific aspects of Personal Big Data science.
- ▶ The objective is to discuss how a person can have a generalizable extraction of knowledge and recommendation from personal data to improve the everyday life quality.



Present and Future



- ▶ Big data and personal information are converging
 - ▶ They'll predict your needs, store your memories, and improve your life—if you let them.
 - ▶ There is so much more data out there that you can afford to tailor it to the individual.
 - ▶ Personal analytics is what you can figure out applying analytics to the person, to understand the operation of the person. (—Stephen Wolfram)
-

Gartner's definition

- ▶ Analytics exploits information to identify patterns, create possible change scenarios, make predictions about the future, and prescribe actions based on predicted results.
- ▶ **Analytics is the key to successfully driving changes in people, processes, and business systems.**
- ▶ Thanks to simulation we can examine alternative outcomes and scenarios not only before but also **during and after implementation and execution.**



Growing Amount of Personal Data

- ▶ **Personal data** is linkable to an individual **directly or indirectly**.
- ▶ Humanity created 2.8 zettabytes of digital data in 2012.
- ▶ The average U.S. citizen creates **5,000 megabytes of data per day**
- ▶ Legal issues: California “**Right to Know**” bill would give people access to copies of all personal data. It’s opposed by the technology industry.



Big Data for Mobile Apps: Satisfy a Need

- ▶ The whole promise of big data is that you'll never perceive that you're getting spam again. You'll never receive an offer that isn't for you because there will be so much intelligence in the system.
- ▶ Games, fitness app, documents, photos, ... will be in one place and thanks to this intelligence they will not bother you when you don't need them.
- ▶ Time is priceless, people are getting used to live in a fast moving world, so they need intelligence to cut the old waste of time and embrace what they really need.



Big data is growing

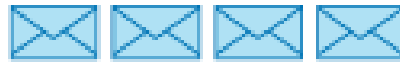
- ▶ The amount of digital data being produced globally is growing exponentially.
- ▶ Most can be considered “**personal**” or linked to individuals.
- ▶ Mining such data is leading to important new types of **highly personalized products and services**.



 **65 billion**

Location-tagged payments
made in the U.S. annually

154 billion



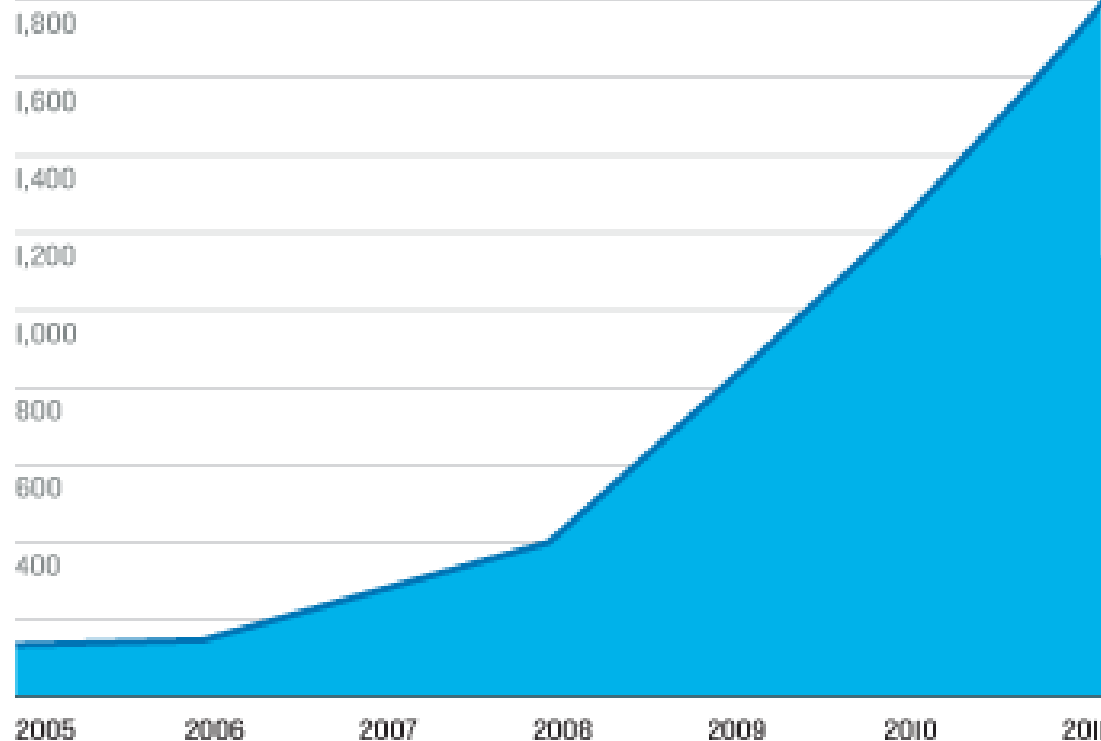
E-mails sent per day

 **87%**

U.S. adults whose location is
known via their mobile phone

Digital Information Created Each Year, Globally

2,000 BILLION GIGABYTES



2,000%

Expected increase in
global data by 2020

**III
Megabytes**

Video and photos stored
by Facebook, per user

75%

Percentage of all digital
data created by consumers

Predictive Apps

- ▶ It works via **data mining** in e-mail, contact lists, and calendars, combined with public databases, such as flight schedules, weather, traffic.
- ▶ Privacy tradeoff: **predictive apps work better the more information** you give them.
- ▶ **Expect combination** with Google Glass and other new interfaces.



Anticipatory Systems Guess Your Next Move

Smartphone apps that mine personal data in order to anticipate a person's needs

NAME	Cue	Google Now	Osito	Tempo AI	Dark Sky
RAISED	\$4.7 million	N/a	\$1.1 million	Incubated at SRI International	\$39,376
FOUNDED BY	Y Combinator graduates Daniel Gross and Robby Walker	An internal Google team	Bill Ferrell, a former Google developer	Raj Singh, Corey Hulen, and Thierry Donneau-Golencer	Jack Turner and Adam Grossman
PREDICTIONS	Summarizes a person's day based on information scavenged from calendar, e-mail, and documents	Directions, traffic, and weather based on a person's location and calendar	Handles transactions like checking in for a flight or calling a cab after you land at the airport	Directions to appointments. Also sends messages if you're running late	Provides minute-by-minute weather forecasts for user's exact location

Sources: Cue, Google Now, Osito, Tempo AI, Dark Sky



Mobile Apps and Personal Analytics

- ▶ **Mobile apps are not ready to manage big data, that's why we need services to analyze and summarize data to build comprehensive diagrams and views.**
- ▶ End users love to track their personal data in the field of health, fitness and every day life.
- ▶ **Moreover, end users love statistics and analytics to have a quick way to understand what's happening in their personal contexts.**



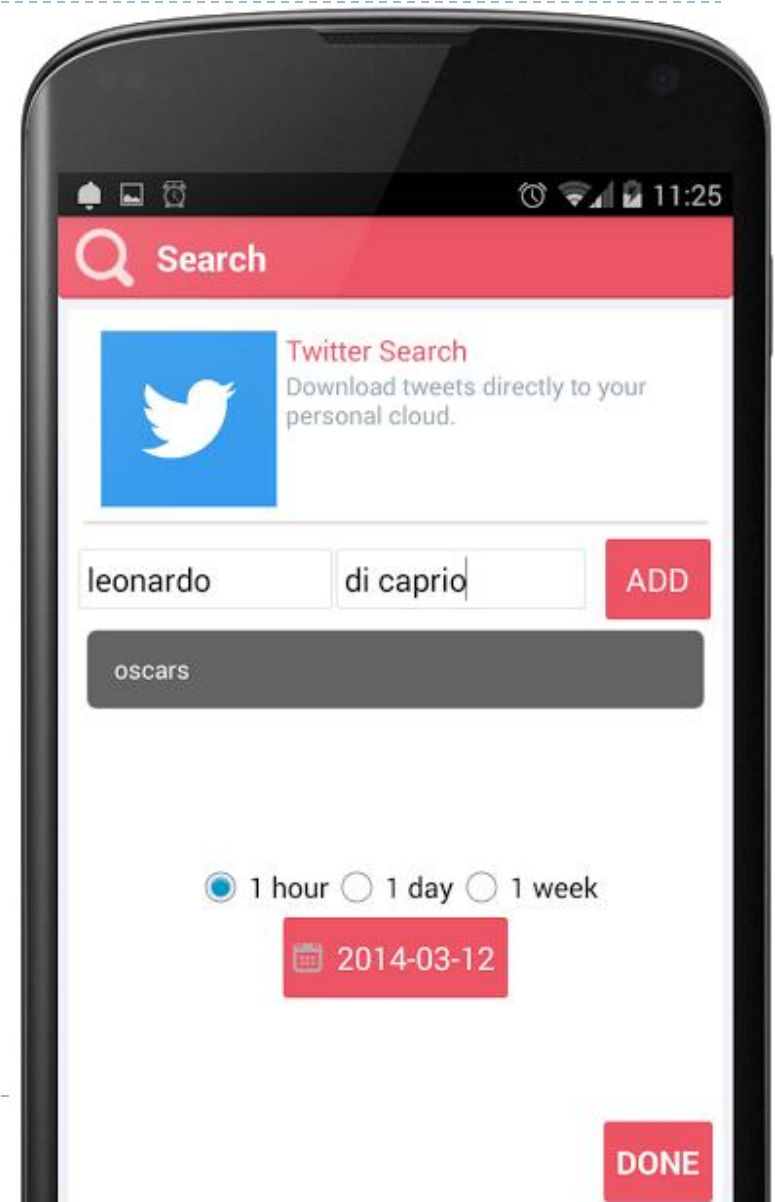
Mobile Apps versus Old Fashioned PCs

- ▶ “The big difference is that desktops are stationary and laptops do not in general have GPS sensors, so there is a location awareness in the data that can be captured and mined from a mobile device,” says Andrew Purtell, principal architect at Intel.
- ▶ “Increasingly, people live within the virtual world presented by their mobile devices as they move through the physical world,” Purtell says.
- ▶ The virtual world becomes a personal world that creates the need of an **intelligent companion**, something that sorts out what is good or not from complex and big sets of data.



From Twitter to analytics report

- ▶ A real example is to search for something on twitter using a mobile app.
- ▶ In this case the result will contain thousands and thousands of tweets, that are impossible to understand and read in 2 minutes!
- ▶ The only way to understand what's going on, is to generate a report.



Analytics report to create a sense from big data

- ▶ Tweets and Rss are statistically processed generating cross relationship between major frequency word inside texts using the technique of distance. Words and txt are clustered giving the possibility to understand where the main “information” is contained applying in full the paradigm of big data and related analytics.
- ▶ The report generated from a twitter search consists of a summary and some diagrams about words correlations and tweets clustering.



Correlation Matrix

- ▶ A matrix of correlations can be displayed schematically in a variety of forms: as numbers, shaded squares, bars, ellipses, or as circular 'pac-man' symbols. These schemes all attempt to show **both the sign and magnitude of the correlation value, using a color mapping of two hues in varying lightness** (Cleveland, 1993), where the intensity of color increases uniformly as the correlation value moves away from 0.





- ▶ In the example color (blue for positive values, red for negative values) is used to encode the sign of the correlation between words to understand in the messages what words are contained together.

Correlation Matrix

- ▶ In the shaded row, each cell is shaded blue or red depending on the sign of the correlation, and with the intensity of color scaled 0-100% in proportion to the magnitude of the correlation.



Creation of Correlograms in R

- ▶ **Correlograms** help us visualize the data in correlation matrices.

```
> library(corrgram);  
> library(ellipse);  
> R = cor(x[, 2:30]);  
> R<-round(R, 3);  
> jpeg(filename = jpeg3, width = 1920, height = 1080, units = "px",  
pointsize = 12, quality = 100,bg = "white");  
> corrgram(R,main = "Bigger Frequency words Correlation");
```



Cluster Analysis

- ▶ *“The notion of a "cluster" cannot be precisely defined, which is one of the reasons why there are so many clustering algorithms”. [Vladimir Estivill-Castro]*
- ▶ However there is a common denominator: **it’s a group of data objects.**



Cluster Analysis

- ▶ Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).
- ▶ It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.



Cluster Analysis in R

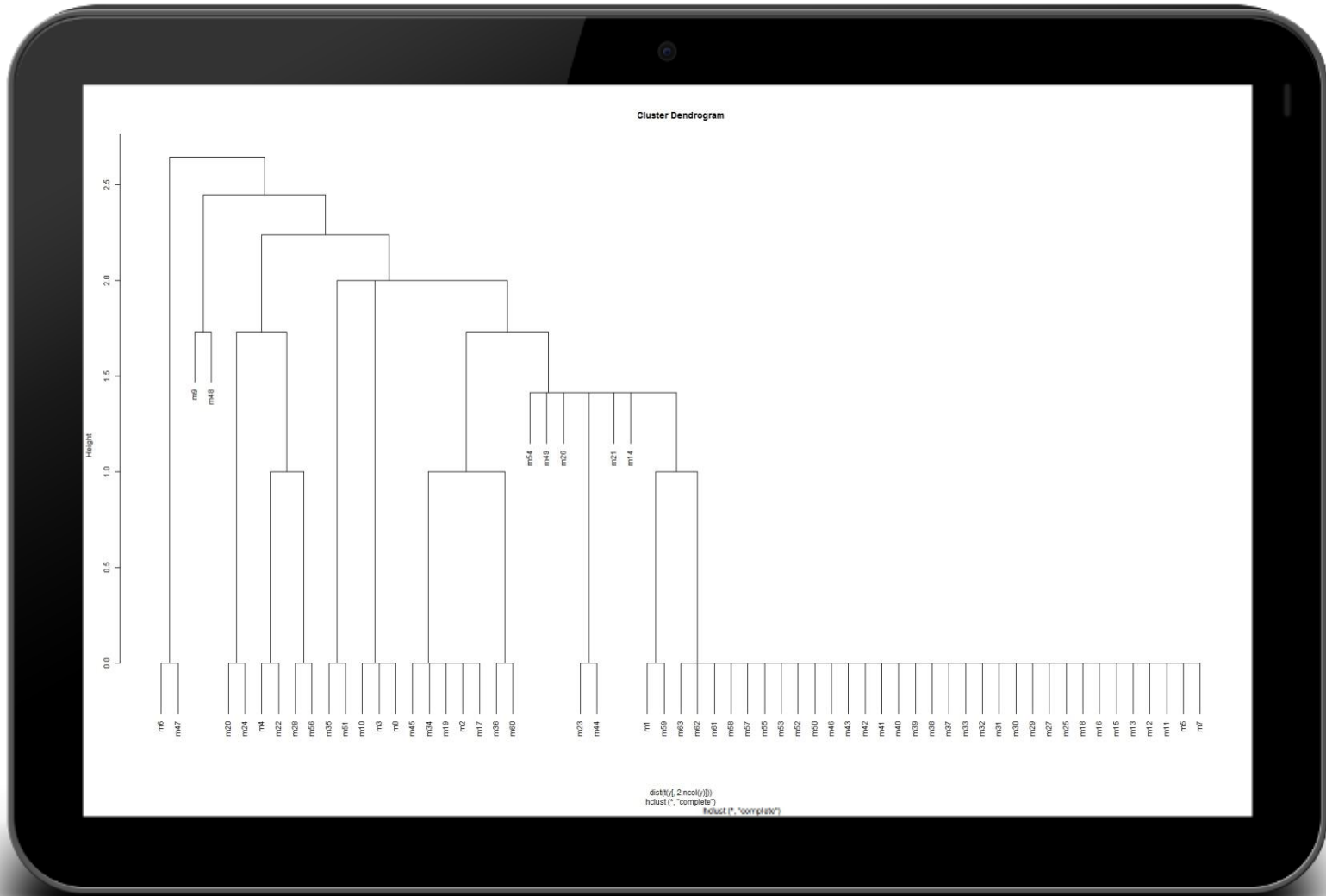
- ▶ With a distance matrix, we can use various techniques of cluster analysis for relationship discovery.
- ▶ We can run the distance matrix with `hclust`, and plot a dendrogram that displays a hierarchical relationship among the vehicles.

Example:

```
> d <- dist(as.matrix(mtcars)) # find distance matrix
> hc <- hclust(d) # apply hierarchical clustering
> plot(hc) # plot the dendrogram
```

- ▶ The next image will clearly describes how messages are clustered depending on words.
-

Cluster Analysis



Data Analytics & Decision Process

Past

Observe

Descriptive Analytics

Questions:

What happened?
Why did it happen?

Solutions:

- Data mining & forensics
- Real-time analytics & mining
- Market segmentation
- Reporting & dashboards

Near Future

Predict

Predictive Analytics

Questions:

What is likely to happen?

Solutions:

- Simulation ←
- Statistics & linear regression
- Predictive data-mining
- Forecasting & trend reporting

Future

Influence

Prescriptive Analytics

Questions:

What should I do about it?
How do I influence the future?
What are the consequences?

Solutions:

- Simulation ←
- Optimisation



Simulation-Based Analytics

- ▶ A rapid modeling tool that produces data on each alternative, allowing you to get a glimpse into the future of your project.
- ▶ Imagine software that answers the most important question: **What will happen?**
- ▶ These scans can be made with rapid personal modeling simulation tools.



Predictive power

- ▶ Analytics employs statistical methods to both draw conclusions from historical data, as well as the predictive power of simulation models to accurately predict future outcomes.
- ▶ The predictive power of rapid modeling software is that essential capability that can help people companies see the future and make informed decisions to gain a competitive edge.



Simulations for people

- ▶ With rapid modeling simulation tools, people can use simulation modeling as an analytics tool for predicting what can or will happen as a result of decisions across the enterprise.
- ▶ The role of simulation is expanding to exploit information and predict the impact of change at all levels of the business.



An example: running



Six Defining Equations

1. $D = \int_0^T v(t) dt$

4. $f(t) \leq F$

2. $\frac{dv}{dt} + \frac{v}{\tau} = f(t)$

5. $\frac{dE}{dt} = \sigma - f(t)v(t)$

3. $v(0) = 0$

6. $E(t) \geq 0$

We search Partnership

- **If you have knowledge on big data, analytics and recommendation systems, please contact us! We can work together, run tests, organize meetups and find new and better solutions.**

- **If you are a Company that works in the field of health monitoring, mobile fitness or smartbands production, contact us! We can work together to find synergies on cloud services, analytics and recommendation system.**

Thank you for the attention!

Stefano Rago
250 Oakmead Parkway
Sunnyvale, 94085, Ca
415 644 5197
stefano.rago@megatris.com

