

Big Data – מגמות, אתגרים ויישומים אפשריים



כנס מידע 2013,
תל-אביב, 8.5.13

תמי נויטל, האוניברסיטה הפתוחה
tamine@openu.ac.il

נושאים

Big Data - על קצה המזלג

נתונים פתוחים

פרויקטים לתועלת הציבור

מידענים ו-data science

What Happens in an Internet Minute?



And Future Growth is Staggering



ולא לה יש להוסיף...



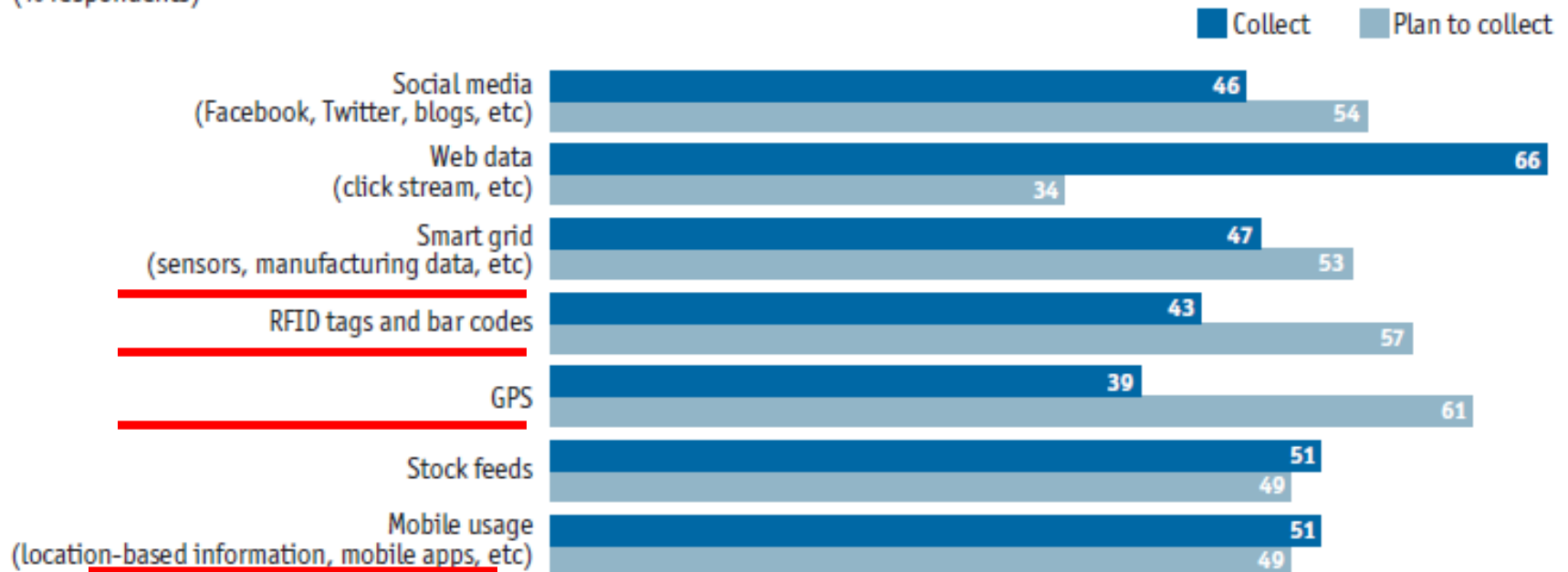


Q

Data collection

Which of the following sources of data does your organisation collect or plan to collect?

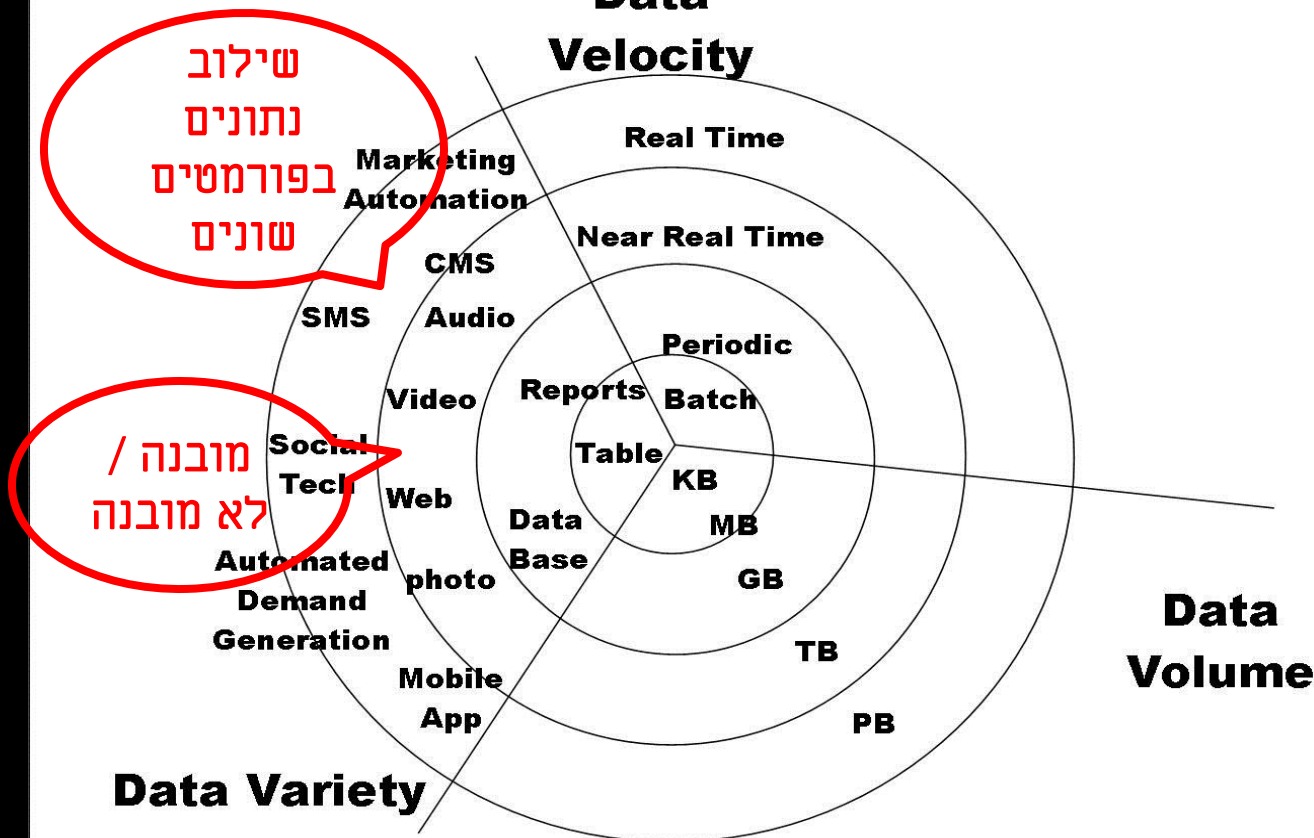
(% respondents)



Source: Economist Intelligence Unit survey, March 2012.

752 senior executives from a broad range of sectors and countries interviews were conducted with 17 executives, consultants and specialists who are regarded as data pioneers

Data Velocity



BIG

Is relative

Veracity

Value?

<http://jeffhurlblog.com/wp-content/uploads/2012/07/3VsBigData.jpg>

When volume, velocity and variety of data exceed an organization's storage or compute capacity for accurate and timely decision making...

מה עושים בארגונים?



לא רק שינויים טכנולוגיים:



KLM Like

Wall Info I LOVE KLM KLM Surprise Video Vote Photos >>

We like you Do you like us?

100,000 friends!

Wanna bet?

Want to join the 100,000 countdown bet? Predict the exact date and time (Central European Time) we hit 100,000 fans and win a nice surprise! The winner will be the one with the closest guess. You can participate until Friday by midnight (24:00 hours CET).

Add a comment...

Raza M Qureshi 15:10
My favourite airline is KLM, & I use it on my all tours, Europe, Canada & even Asia. I wish KLM must start its route to Pakistan.

Message - Report

Arnoud Raeven 14:54

http://youtu.be/HD_W3EMuC1U?t=32s



■ מידע רפואי, אישי ובאוכלוסיות גדולות. תרומות
נתוני בריאות אישיים לקידום המחקר. איתור מגיפות
בהתהוות

■ מידע על משתמשים והעדפותיהם – שיווק, פיתוח
מוצרים, שיפור מערכות קיימות, התאמת שירותים,
פרסונליזציה מרבית ומערכות המלצה

■ נתוני תנועה – כלי רכב, הולכי רגל, אמצעי תחבורה
המונניים... (חב' ביטוח, אנשי לוגיסטיקה, מהנדסי
תחבורה)

■ Digital Humanities ומד"ח – בלשנות, ספרות,
היסטוריה... מפעלי דיגיטציה מספקים מערכי נתונים

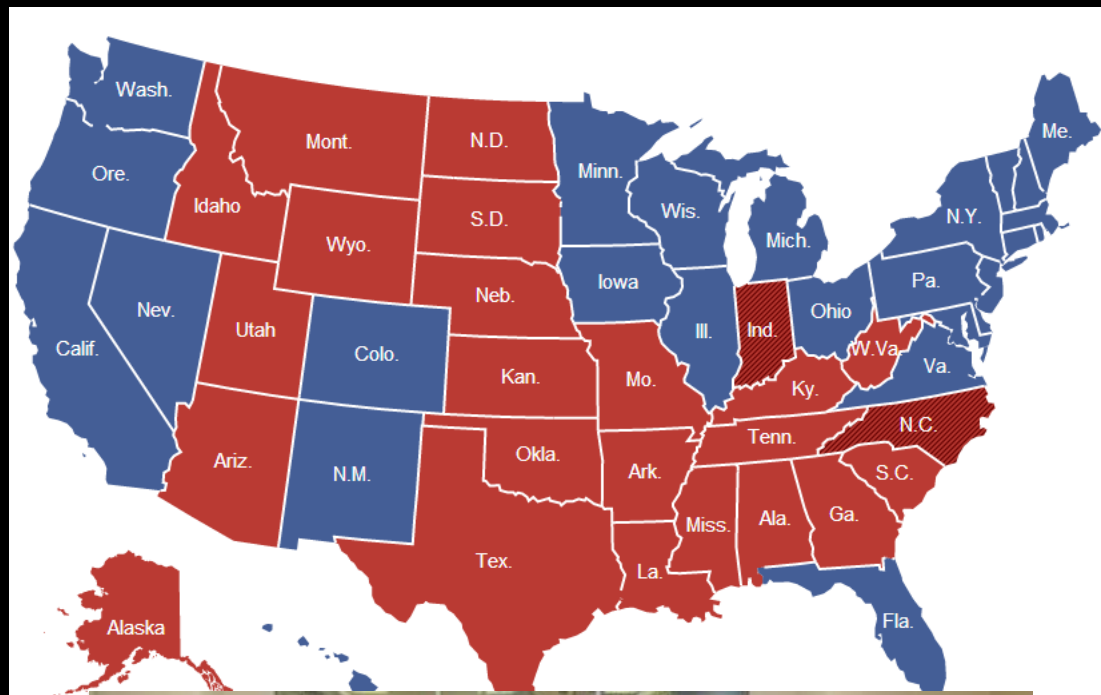
Researchers at IBM have redrawn the bus routes of Ivory Coast's largest city using mobile phone data.

The research was completed as part of the Data for Development competition run by Orange which released 2.5 billion call records from five million mobile phone users in Ivory Coast.

The anonymised data is the largest of its kind ever released.

Such data could be used by urban planners for new infrastructure projects said IBM.

Currently the project is just a research exercise although IBM is hoping to implement it in a number of cities.



WHY IT MATTERS

The Obama 2012 campaign used data analytics and the experimental method to assemble a winning coalition vote by vote. In doing so, it overturned the long dominance of TV advertising in U.S. politics and created something new in the world: a national campaign run like a local ward election, where the interests of individual voters were known and addressed.



Purdue University achieves remarkable results with big data

Every student generates a unique data trail. This data trail can be analyzed in real-time to deliver an optimal learning environment for the student as well to gain a better understanding in the individual behaviour of the students.

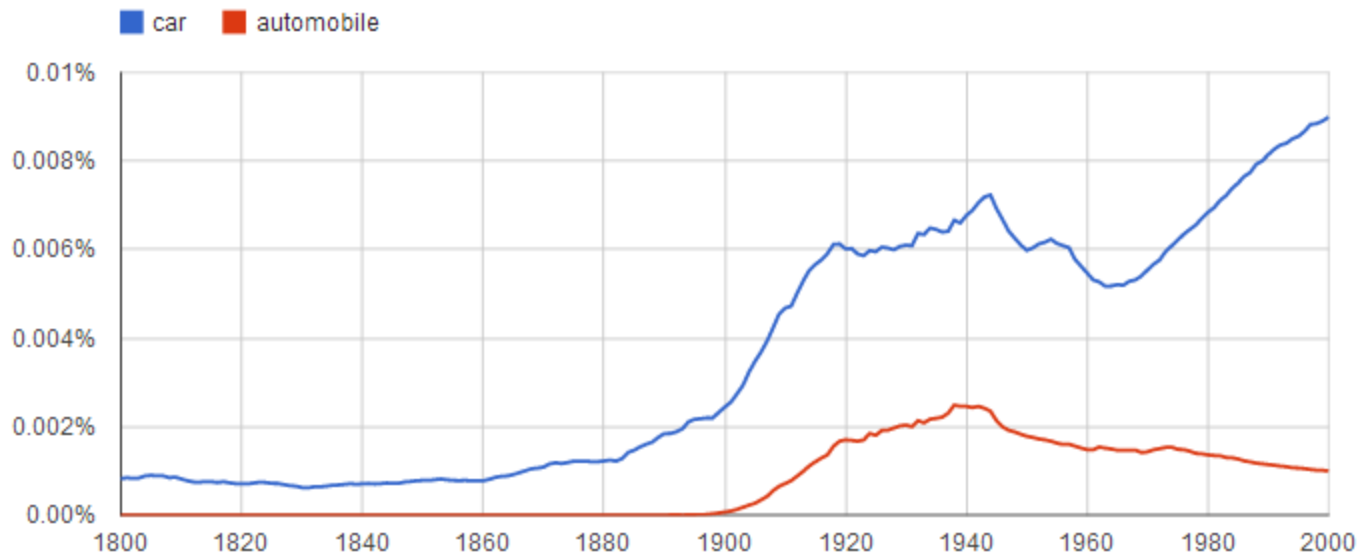
The screenshot displays the 'Course Signals' web application. The top navigation bar includes 'Home', 'Course Progress', 'Progress History', and 'Settings'. Below this, a 'Term(s)' dropdown is set to '201130-Summer 2011'. A table lists course sections with columns for 'Section', 'Parent Section', 'Status', 'Last Signal Updated', and 'By'. Two sections are visible: 'Summer-2011-ENGL-421' and 'Summer-2011-PHYS-17200'. Below the table, a 'Section Details' tab for '12-Aug-2011' is active. The main content area shows 'Summer-2011-ENGL-421' with a 'Filter B' option. On the left, a 'Risk Group Summary' table shows risk levels (High, Medium, Low, Unknown) and their corresponding student counts and grade cutoffs. On the right, a table lists students and their performance across different sections, with columns for 'Student', 'Cross-Listed Child Section', and three columns for signal status (8/8/2011, 5/8/2011, 8/12/2011). The bottom of the page features a copyright notice: '© 2000 - 2011 SunGard All rights reserved. SUNGARD HIGHER EDUCATION'.

Google books Ngram Viewer

Graph these [case-sensitive](#) comma-separated phrases:

between and from the corpus with smoothing of .

[Search lots of books](#)



Search in Google Books:

[1800 - 1884](#)

[1885 - 1973](#)

[1974 - 1984](#)

[1985 - 1993](#)

[1994 - 2000](#)

[car](#)

[1800 - 1919](#)

[1920 - 1940](#)

[1941 - 1946](#)

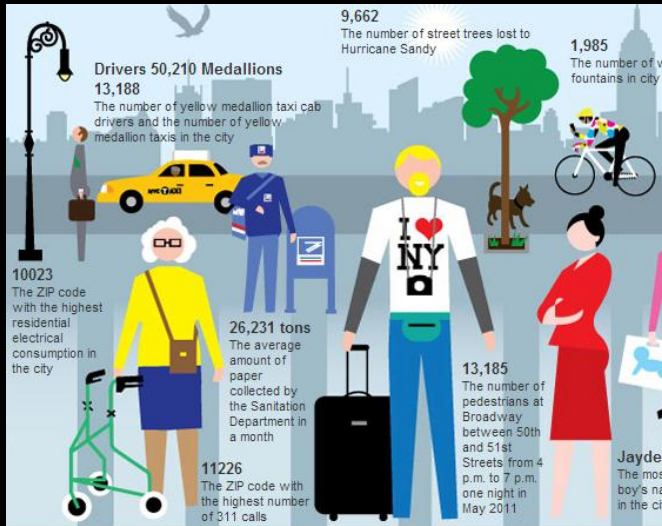
[1947 - 1985](#)

[1986 - 2000](#)

[automobile](#)

Run your own experiment! Raw data is available for download [here](#).

<http://books.google.com/ngrams>



The Mayor's Geek Squad

By ALAN FEUER
Published: March 23, 2013



HOME ABOUT DATA METRICS OPEN GOVERNMENT



Search public data



Tami Neuthal 0

+ Share

Public Data

Datasets

Metrics

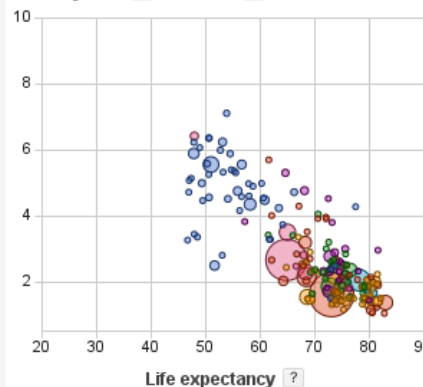
Any data provider (103)

Eurostat (10)
Destatis (6)
U.S. Bureau of Labor Statistics (5)
U.S. Census Bureau (5)
U.S. Bureau of Economic Analysis (5)

My Datasets

Fertility rate

Countries



Living longer with fewer children

This chart correlates life expectancy and number of children per woman for each country in the world. The bubbles are sized by population and colored by region. Over time, most countries have moved towards the bottom right corner of the chart, corresponding to long lives and low fertility. Note the progression of the bubble for China- in the late 60's and 70's life expectancy rose quickly, then the implementation of the one-child policy caused a drop in the number of children per woman.

Explore the data

Dataset: World Development Indicators

נתונים כמוצר – שיתוף ומכירת נתונים

“שווקים משניים” - חברה שאוצרת נתונים יכולה למכור או לפתוח אותם לאחר “ניקוי” הולם.
למשל:

<http://www.bigdata-startups.com/public-data/>

מערכי נתונים גולמיים לצרכי מחקר או פרויקטים חופשיים –
רשויות, ממשלות, סוכנויות ותאגידים

BIG DATA WEEK

22-28 APRIL

25+ Cities
175+ Events
20,000+ Participants

VIEW EVENTS IN YOUR CITY

"... a global community and festival of data ..."

AMSTERDAM | ATLANTA | BARCELONA | BERLIN | BEIRUT | BRIGHTON | BIRMINGHAM | BUDAPEST | BUCHAREST | CHARLOTTE | CHICAGO | COLOGNE |
COLORADO | DRESDEN | GLASGOW | HOUSTON | KUALA LUMPUR | LEEDS | LONDON | MADRID | MOSCOW | NEW YORK | PERTH | SAN FRANCISCO |
SAN JOSE | TORONTO | UNIVERSITY PARK | UTRECHT | WASHINGTON D.C.

A GLOBAL FESTIVAL OF DATA



#bdw13

25+ CITIES

Get Involved!

Amsterdam	Chicago	Moscow
Atlanta	Cologne	Perth
Barcelona	Colorado	Shanghai
	Houston	Toronto
Birmingham	Kuala Lumpur	Utrecht
London	Madrid	Washington DC

"Big Data Week is one of the most unique global platforms of interconnected community events focusing on the social, political, technological and commercial impacts of data."

למה מידענים?

□ תחום מתחדש, כלי עזר משתנים, הרחבה של
"עולם המידע"

□ מקורות מידע חדשים ושימושיים כנראה – ללא
מאגרים מסודרים

□ צורך ב"תרגום" ותיווך מידע

□ למצוא "שאלות נכונות" כדי להגיע מדאטה
להחלטות

□ אוצרות שוטפת, אוצרות "מורשת"

למתעניינים:

□ קורסים אקדמיים – מעט מיועדים לחסרי ידע בתכנות. בשנה הבאה אולי בחיפה.

□ קורסים פתוחים והדרכות ברשת -

Data Science ,SNA – Coursera

□ הרצאות מוקלטות מקורסים וכנסים

□ ניירות עמדה, דו"חות, פורטלים של SAS ,IBM ,

ועוד



□ טוויטר - #big data, #bdw13, #smartdata