

Data Analytics

CS301

Introduction to Data Analytics

Week 1

Fall 2018

Oliver BONHAM-CARTER

Data Analytics CMPSC*301 / CMPSC*390

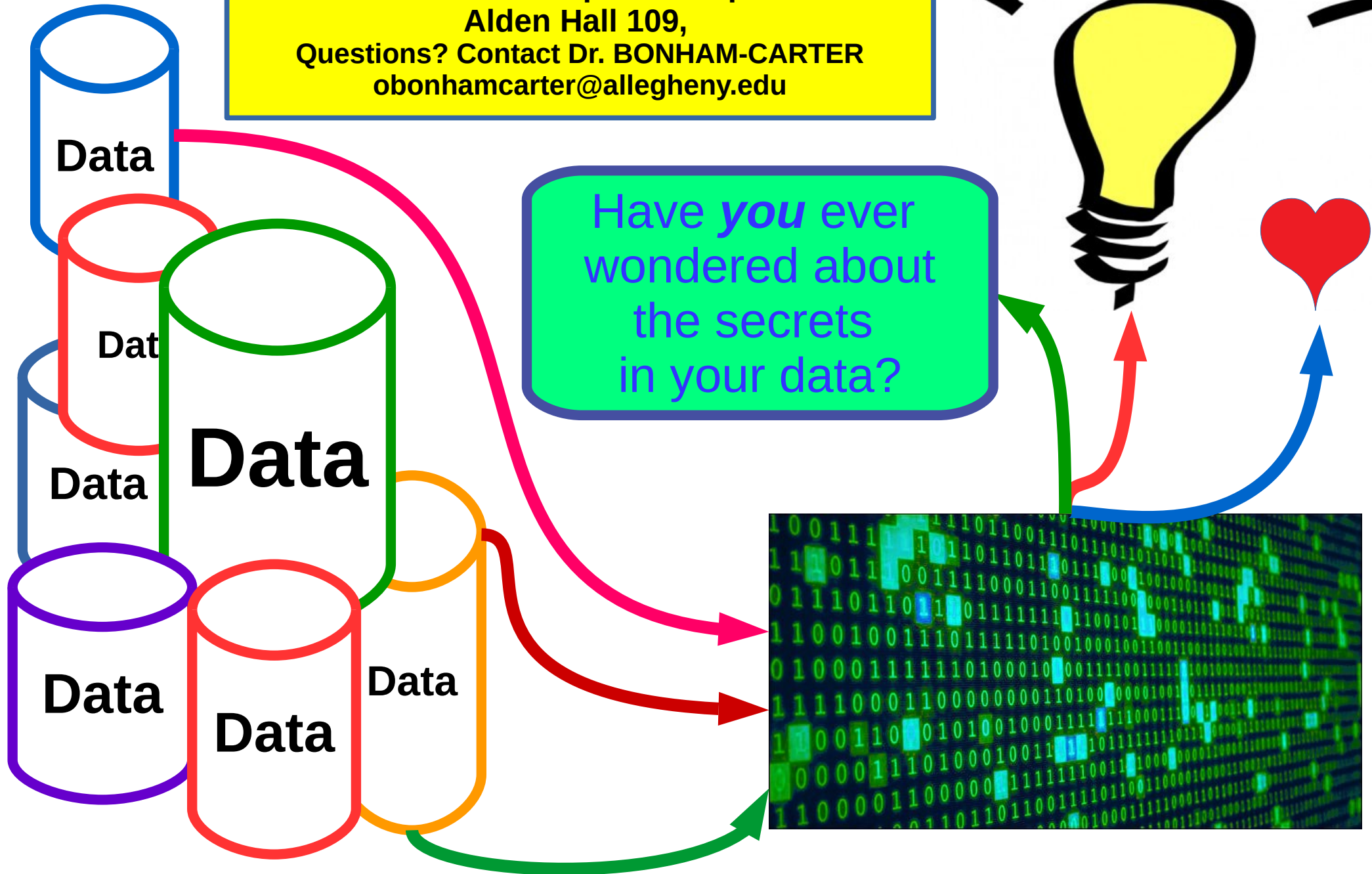
Lect: Tues and Thurs: 9:30am – 10:45am

Lab: Fri: 2:30pm – 4:20pm

Alden Hall 109,

Questions? Contact Dr. BONHAM-CARTER

obonhamcarter@allegheny.edu



- In this class, you will learn how to use machines to harness the power of data.



Meaningful Information

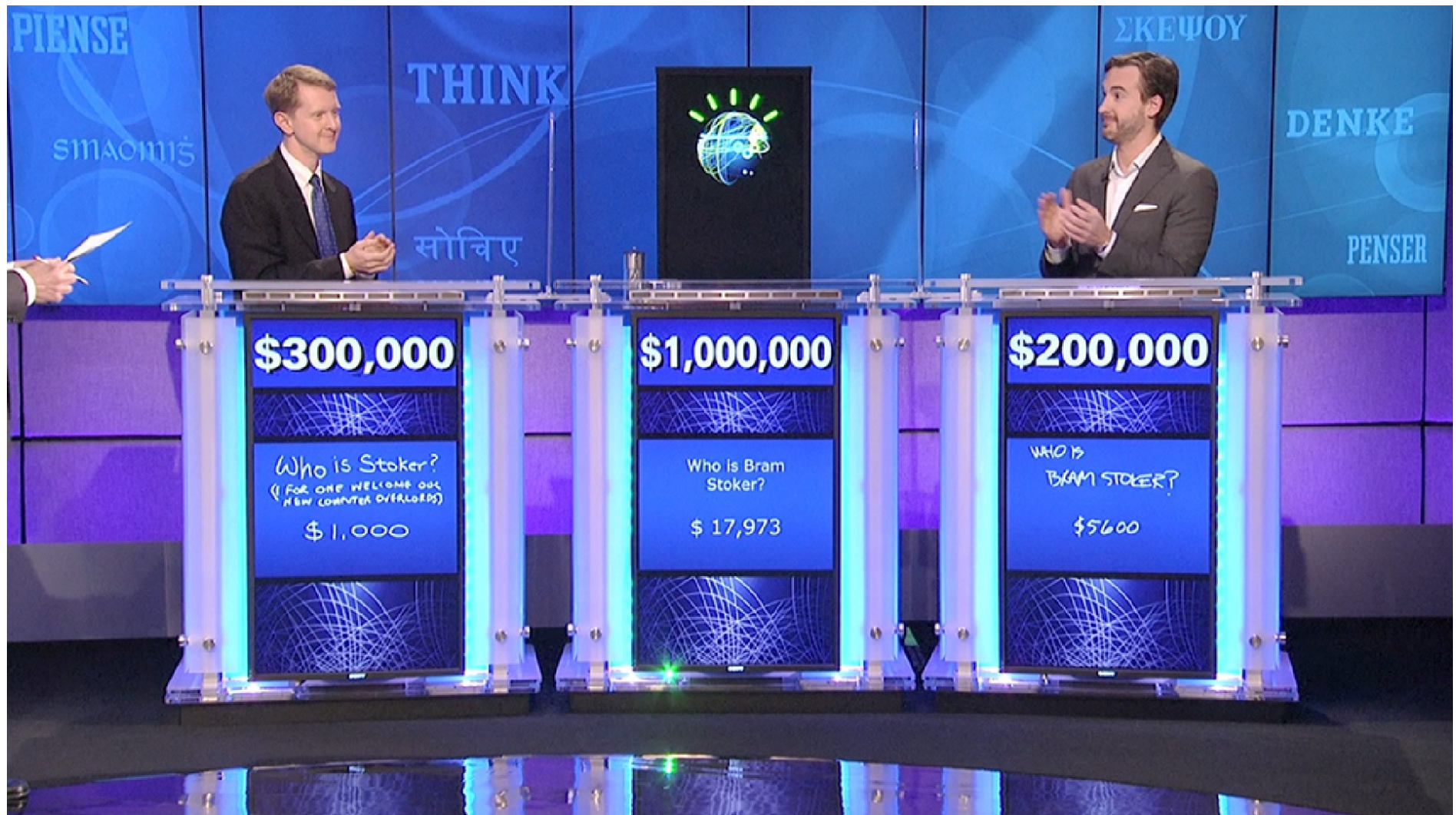


Analytics in Action

- The Jeopardy! Challenge of February 2011
- IBM's Watson beat the show's greatest champions: Ken Jennings and Brad Rutter.



Machines, Data and Information



WATSON for PRESIDENT



Is Watson magic??

<http://watson2016.com/>
(The Electronic Frontier Foundation)



Surrounded by DATA!

- We live in the “Information age”
- Actually, we live in the “Data age” since there is more data available than information
- Data != Information



Surrounded by DATA!

- It is cheap (and free or even lucrative) for businesses to collect data concerning:
 - in e-commerce,
 - customer behaviors,
 - purchase interests,
 - health and medical data.



Meet the Fitbit Family



EVERYDAY FITNESS

ACTIVE FITNESS

PERFORMANCE FITNESS

SMART SCALE

zip

one

flex

alta

chargeHR

blaze

surge

aria

We Voluntarily Give Away Our Health Data

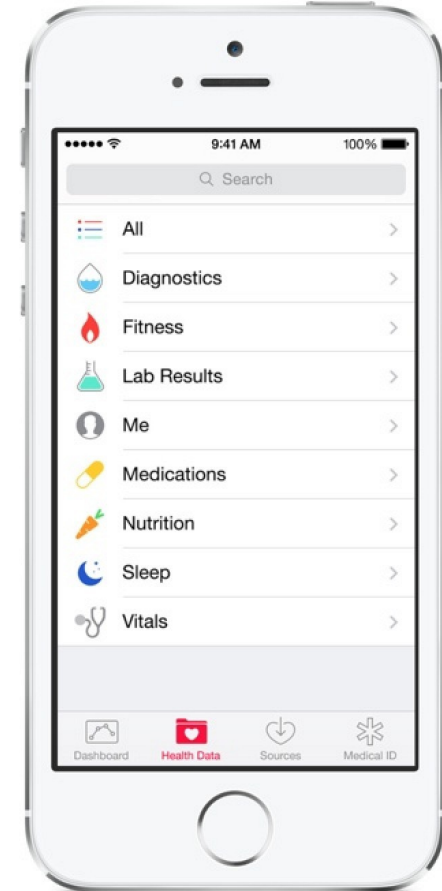




Our Phones Create Data



- Smart phones constantly monitor us and keep data.
- **How does the iPhone decide whether we are actually getting enough sleep?**
- **Who keeps the data?**



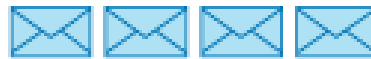


And So, Data is Increasing

 **65 billion**

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

154 billion



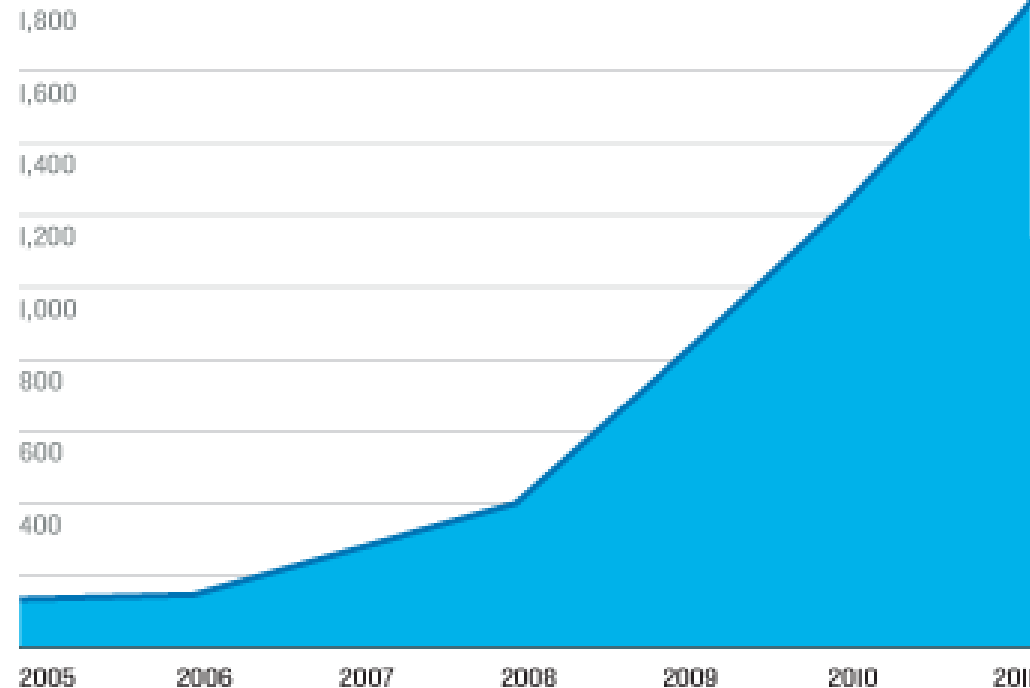
E-mails sent per day

 **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**

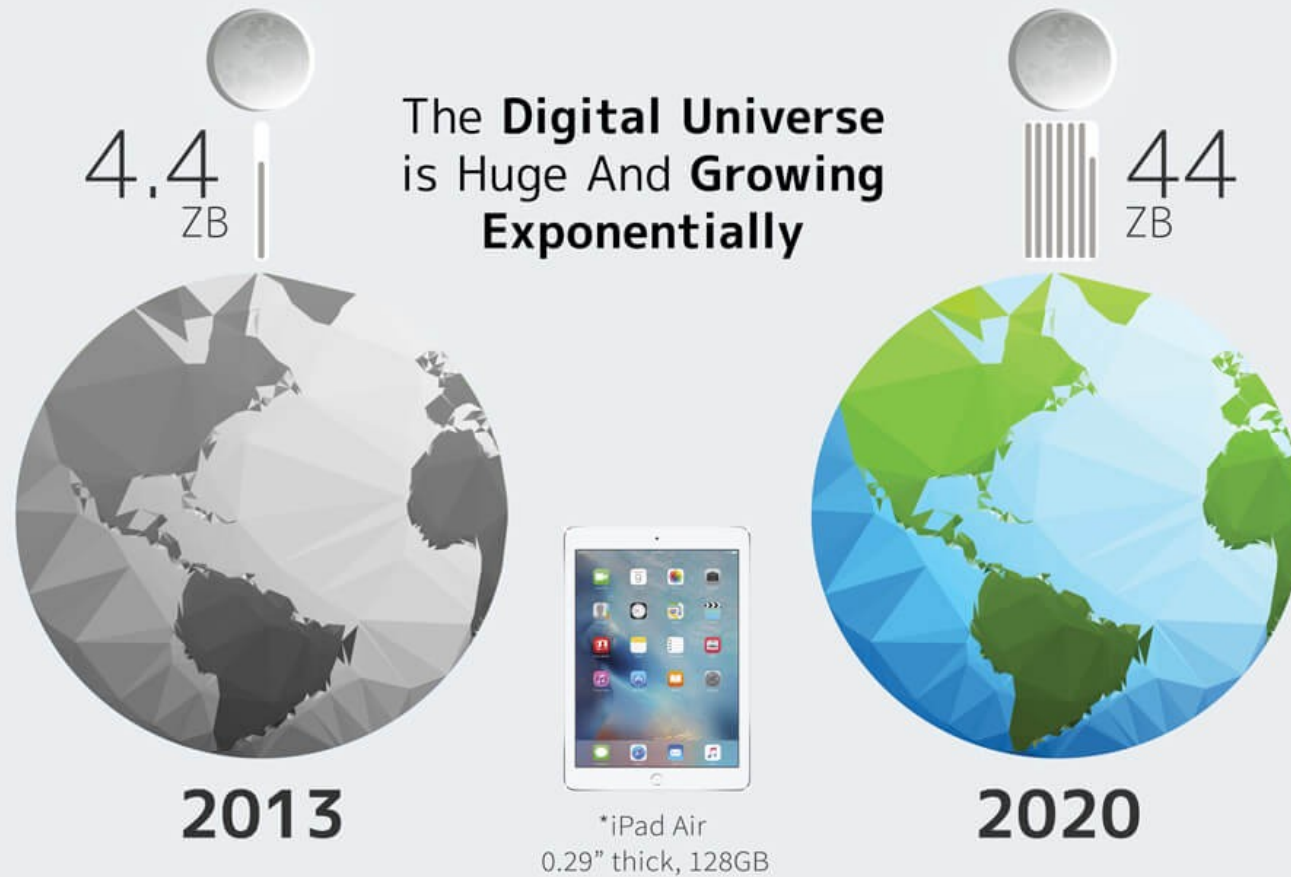
Video and photos stored
by Facebook, per user

75%

Percentage of all digital
data created by consumers



Data, Data, Data, Data!



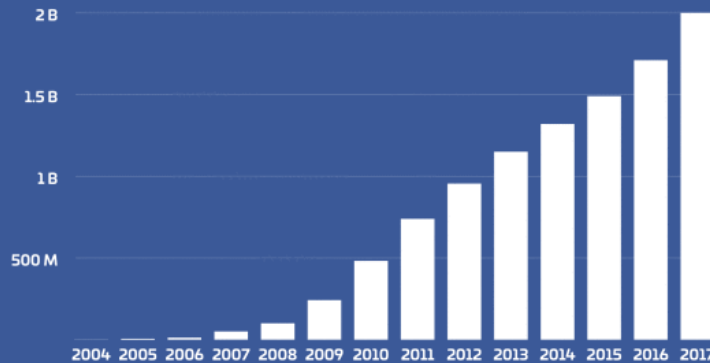
If the Digital Universe were represented by the memory in a stack of tablets, in **2013** it would have stretched two-thirds the way to the moon*.

By **2020**, there would be 6.6 stacks from the Earth to the Moon*.

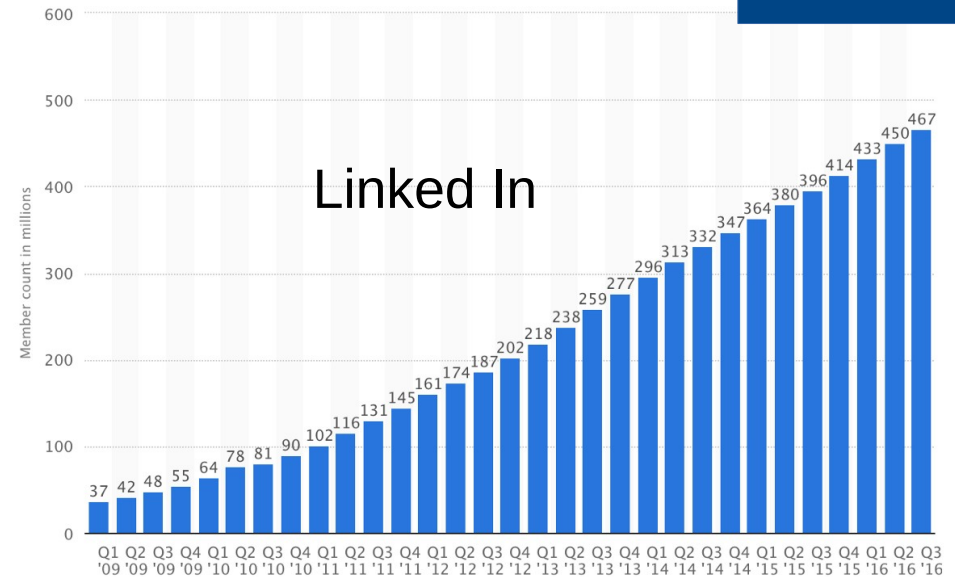


Sources of Data

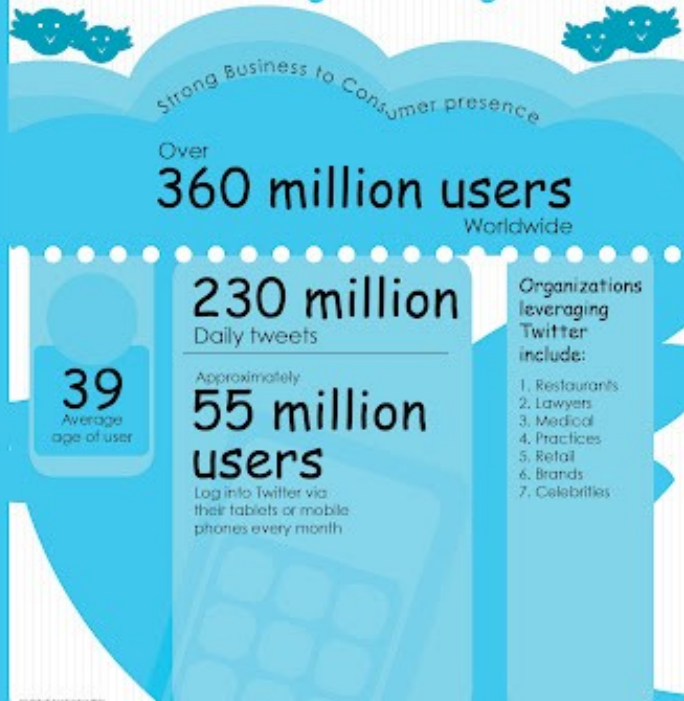
FACEBOOK MONTHLY ACTIVE USERS
JUNE 2017



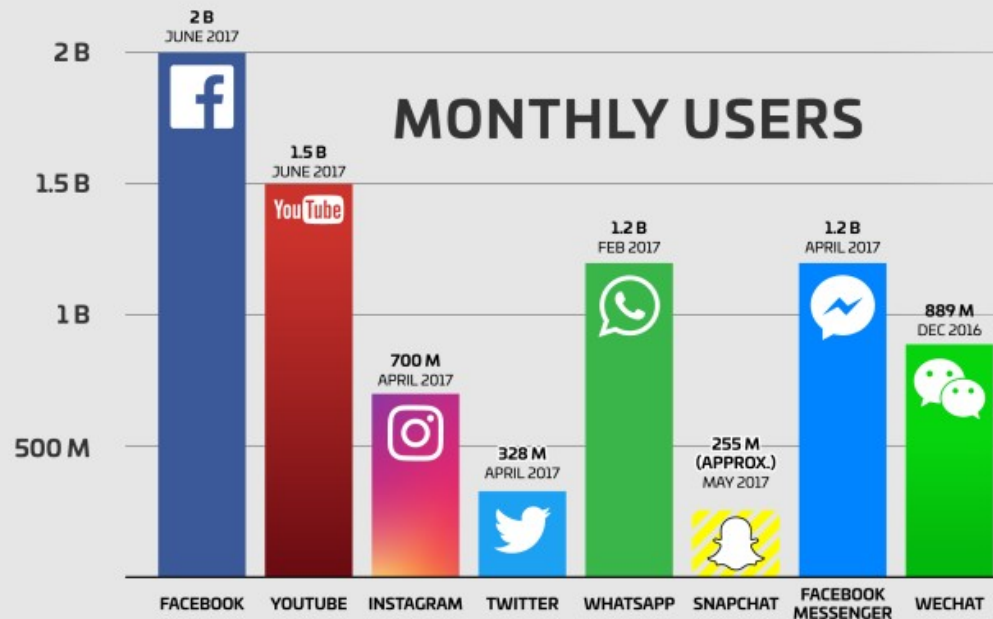
Linked In



twitter fast facts



MONTHLY USERS

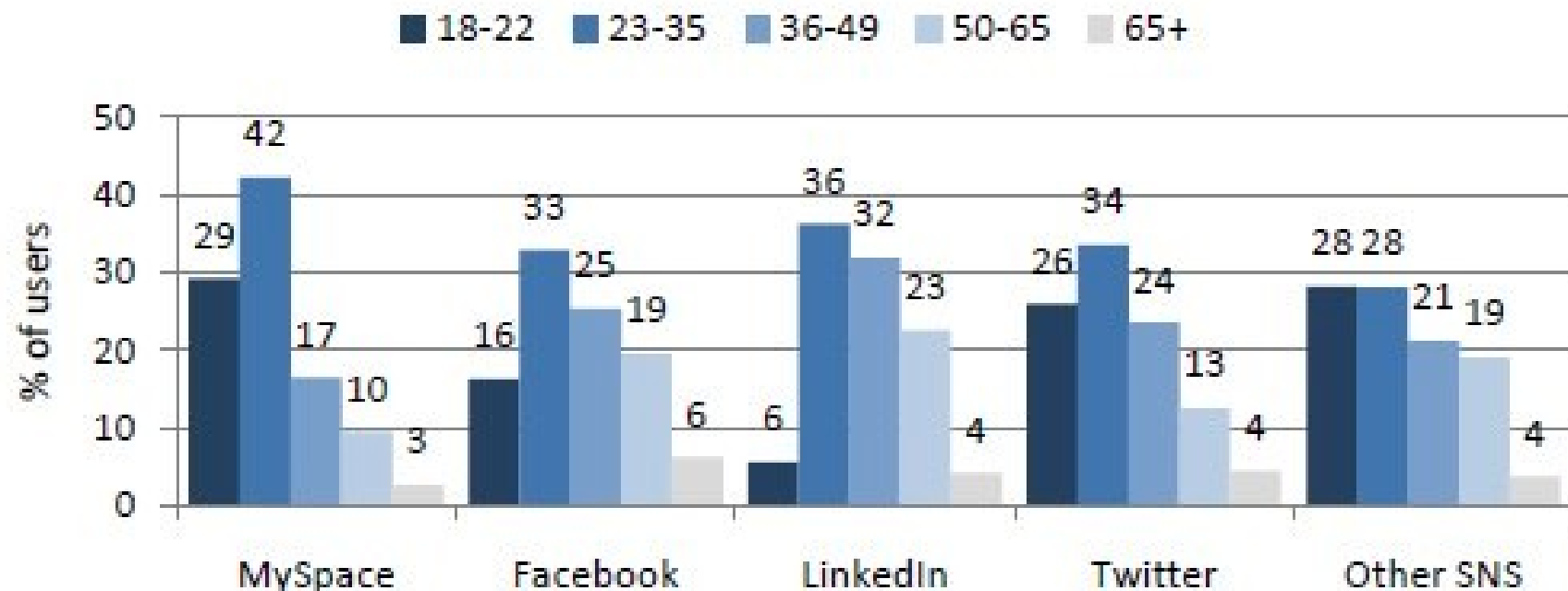




Data of User Ages

Age distribution by social networking site platform

% of social networking site users on each site who are in each age group. For instance, 29% of MySpace users are 18-22 years old.



Source: Pew Research Center's Internet & American Life Social Network Site survey conducted on landline and cell phone between October 20-November 28, 2010. N for full sample is 2,255 and margin of error is +/- 2.3 percentage points. N for social network site and Twitter users is 975 and margin of error is +/- 3.5 percentage points.

By the way: These last slides discussed some trends...

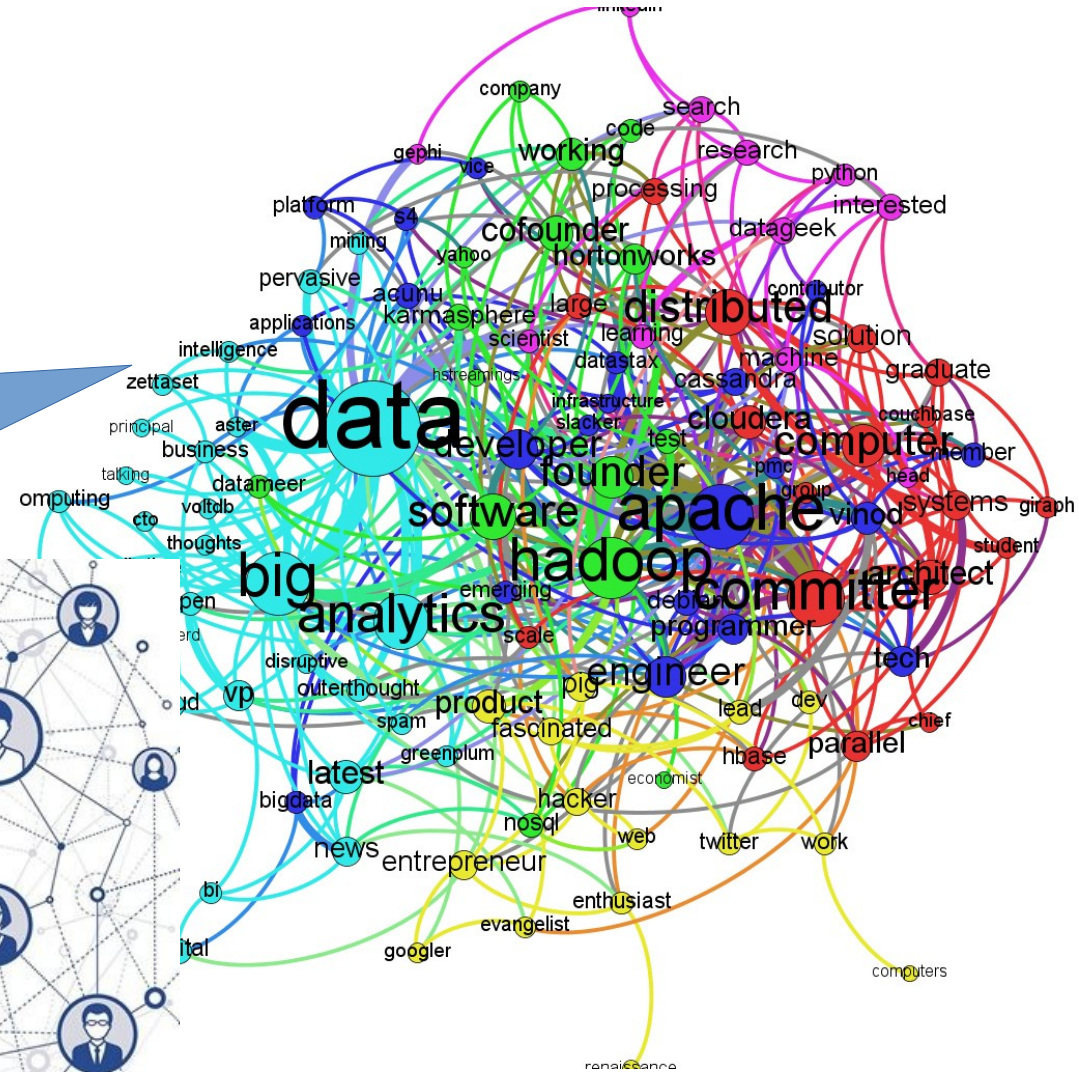
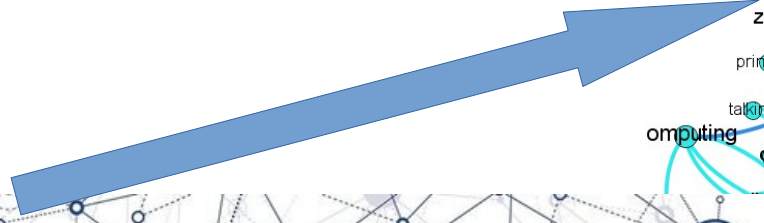
- Graphics have informed us:
 - Which apps are popular
 - Number of people in age groups for social networking sites
 - How much data is created each year, in relation to other years
 - Twitter “fast-facts”
 - Monthly users of services
 - Increases in Linked-In membership
- **How did we learn this information to make these previous visualizations?**

**Seriously, where did this
information come from???**



From Raw DATA!!

- Algorithms processed seemingly unconnected data to filter out unimportant material.





How Do We Know?

- The previous graphs came to us via raw Big Data from sites like Google, Facebook, Twitter and others.
- **Raw Data:** Seemingly meaningless clutter-like gibberish in which patterns are masked.

*Big data is high-**volume**, high-**velocity** and high-**variety** information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.*

-- Gartner





So, It Looks Like We Need **Data** to *Live Intelligently*

- Some basic questions:
 - *Can we make reliable decisions without **data**?*
 - *Is the quality of our society diminished by bad or missing **data**?*
 - *How can we improve commerce, trade without knowledge from **data**?*
 - *How can we make better health decisions without knowledge from **data**?*
- You could give surveys to gather ideas from people but few are likely to respond...

***When was the last time
YOU took a survey?***

Thus, Much Interest in Data Analytics

- The present and future are information-driven
- Some of the decisions made after studying trends in a population
 - Commerce: what have customers already bought?
 - Media: What themes of films, music make money?
 - Industry: What products should we make to build, satisfy a market? Which market?
 - Health and Medicine: Reasons for sickness? Bad types of foods? Exposures to toxins?





Your Career Could Be Here!

- *“Big Data & Analytics Is The Most Wanted Expertise By 75% Of IoT (Internet of Things) Providers”*
 - <https://www.forbes.com/sites/louiscolumbus/2017/08/21/big-data-analytics-is-the-most-wanted-expertise-by-75-of-iot-providers/#52082a4e5188>
- *“75% of IoT providers are prioritizing big data and analytics expertise in their hiring decisions.”*
 - <http://www.forbes.com/sites/louiscolumbus/2017/08/21/big-data-analytics-is-the-most-wanted-expertise-by-75-of-iot-providers/>
- *“68% of vendors developing IoT solutions are struggling to find and recruit employees with development expertise.”*
 - <http://www.forbes.com/sites/louiscolumbus/2017/08/21/big-data-analytics-is-the-most-wanted-expertise-by-75-of-iot-providers/>



Forbes

- ***“75% of firms are prioritizing big data and analytics expertise in their hiring decisions, stating that having these skills is critical for any candidate to be considered an IoT (Internet of Things) expert.”***



Hard to Hire Skilled People

