

DS2000 – Programming with Data

01. Computers and Python



DS 2000: Programming with Data

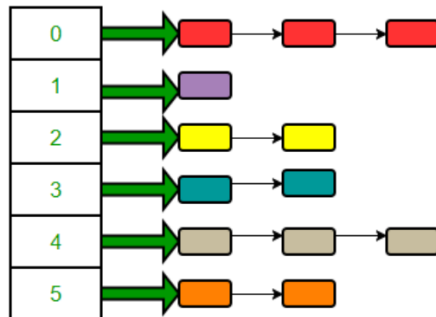
Tue and Fri with Profs. Felix Muzny and John Rachlin

- *DS 2000 is the first course in Northeastern's Programming with Data sequence.*
- *DS 2000 is for DS majors, minors, and non-CS / non-DS majors. No programming experience is assumed!*
- *DS 2000 is the one course to take if you are only going to take one programming course, or you want a broad introduction to Data Science.*

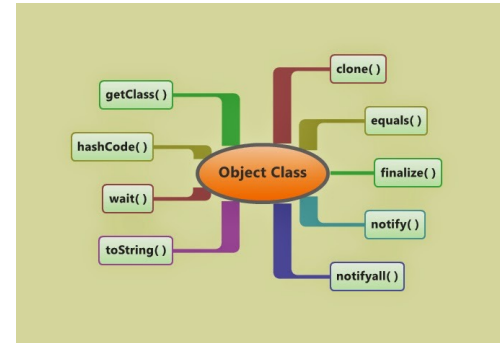
Some of the topics explored in DS2000:



Reading and Processing Data



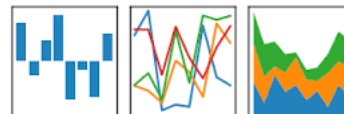
Data Structures and Algorithmic Thinking



Introduction to Object-Oriented Design



Creating Insightful Visualizations



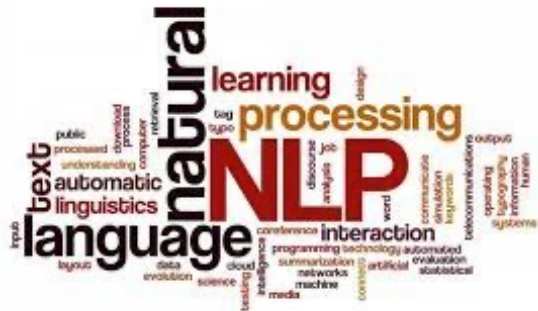
Northeastern University
Khoury College of
Computer Sciences

DS 2500: Intermediate Programming with Data

Tue and Fri 9:50-11:30 / 1:35p – 3:15p ET with Prof. Matt Higger

- *DS 2500* is the second course in Northeastern's *Programming with Data* sequence.
- *DS 2500* is for DS majors, minors, and non-CS / non-DS majors who have taken *DS 2000* or have equivalent programming experience.
- *DS 2500* is the course to take if you want to explore a broad range of data science topics using python.

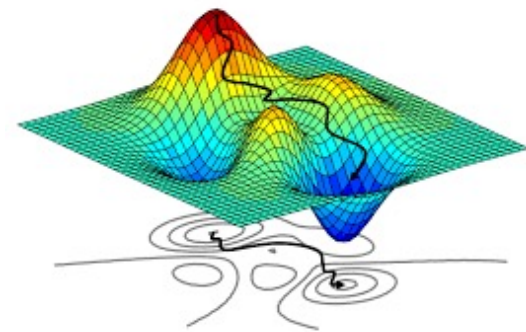
Some of the topics usually explored in DS2500:



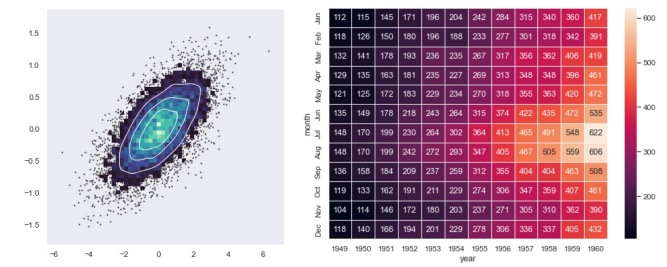
Natural Language Processing (NLP)



Graphs and Networks: Data structures and Analysis



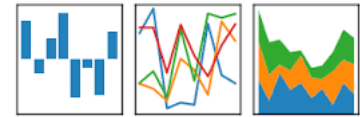
Optimization and Machine Learning



Visualization Techniques



pandas
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



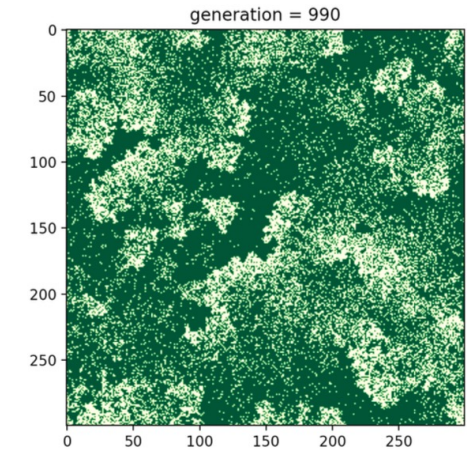
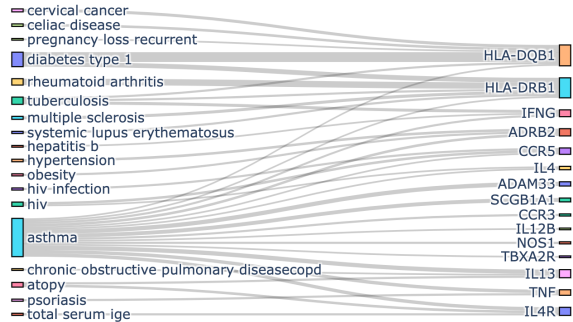
Northeastern University
Khoury College of
Computer Sciences

DS 3500: Advanced Programming with Data

Tue and Fri 1:35p – 3:15p ET / Online with Prof. John Rachlin

- *DS 3500* is the third and final course in Northeastern's *Programming with Data* sequence.
- *DS 3500* is for DS majors, minors, and non-CS / non-DS majors who have taken *DS 2500* or have equivalent programming experience.
- *DS 3500* is the course to take if you want to become a professional software developer or data scientist.

Some of the topics we will explore this Fall 2022:



Building Interactive Visualizations & Dashboards

Working with relational and non-relational databases

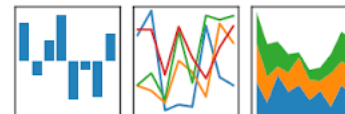
Object-oriented and functional paradigms.

Animation, Simulation, and Modeling



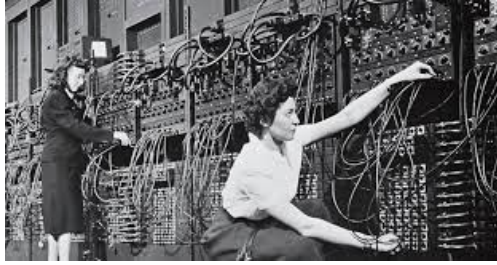
pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Northeastern University
Khoury College of
Computer Sciences

The history of computing



Early Electronic Computing
(1950-1980)

The PC Revolution
(1980-1995)

The Internet Revolution
(1995 - 2010)

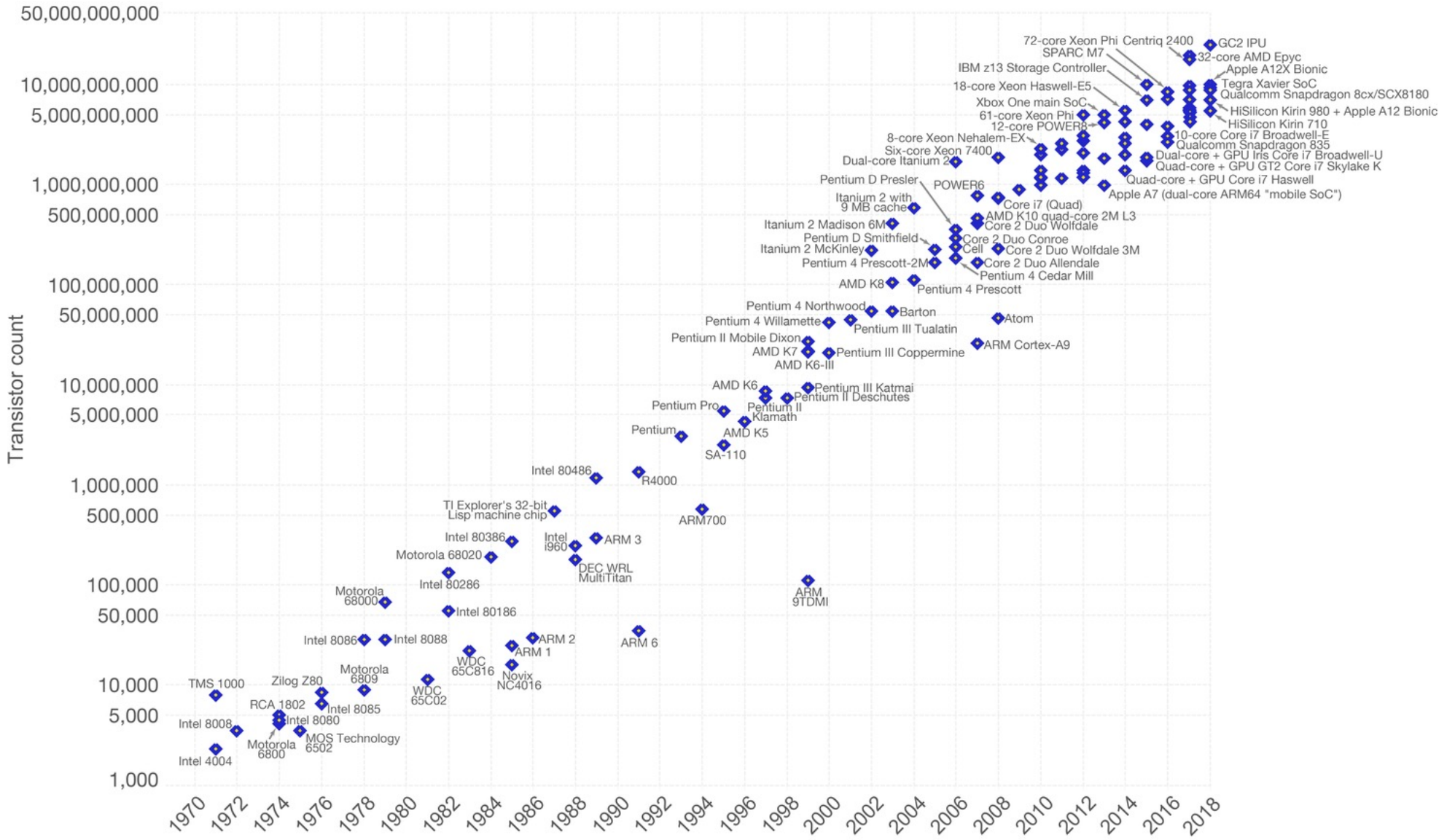
The Smartphone Revolution
(2007-present)

The AI / Cloud / Data Science
Revolution (2010 -)

Moore's Law – The number of transistors on integrated circuit chips (1971-2018)



Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are linked to Moore's law.



Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
 The data visualization is available at [OurWorldinData.org](https://www.ourworldindata.org). There you find more visualizations and research on this topic.

Licensed under CC-BY-SA by the author Max Roser.

Internet Stats – Noon on Jul 31st, 2020

internet live stats

live

1 second

watch

trends & more

Get our Counters!



4,633,646,754

Internet Users in the world



1,789,114,300

Total number of Websites



133,468,656,166

Emails sent [today](#)



3,703,048,613

Google searches [today](#)



3,550,922

Blog posts written [today](#)

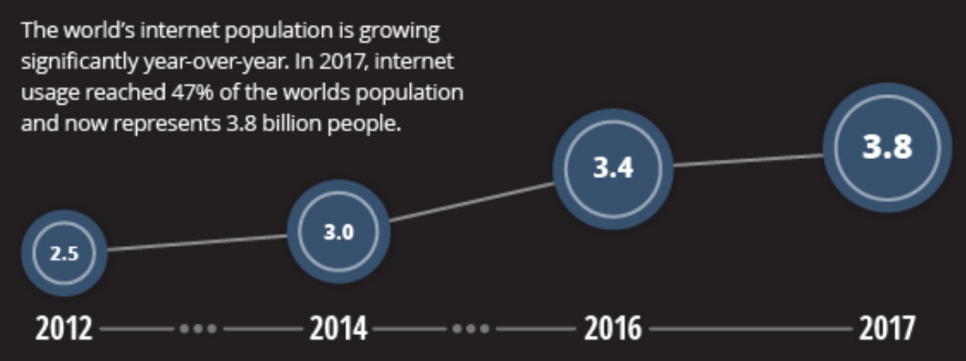


396,932,949

Tweets sent [today](#)

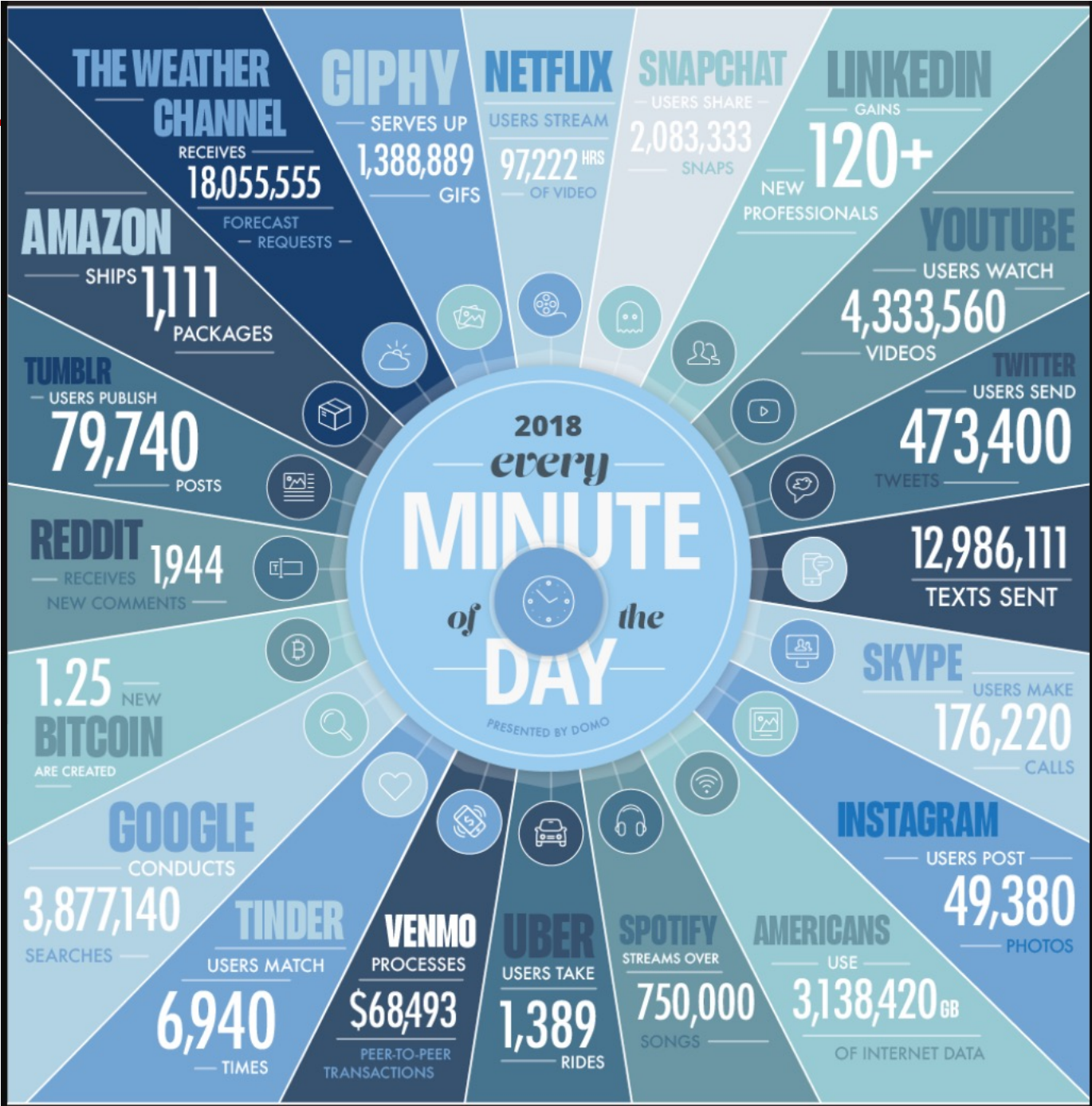


More data every minute



Source:

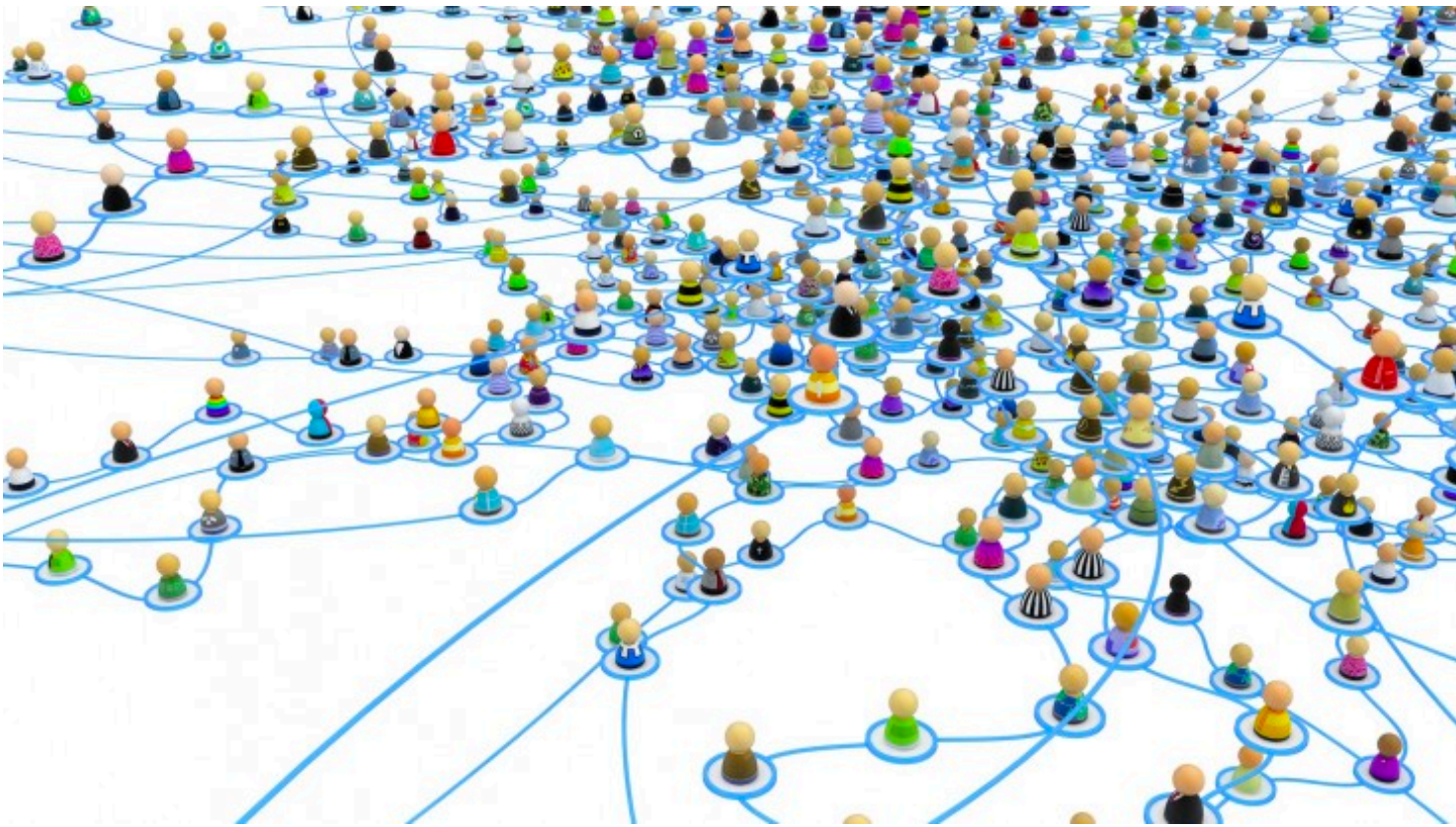
<https://www.domo.com/learn/data-never-sleeps-6>



You are worth \$82 to Facebook...

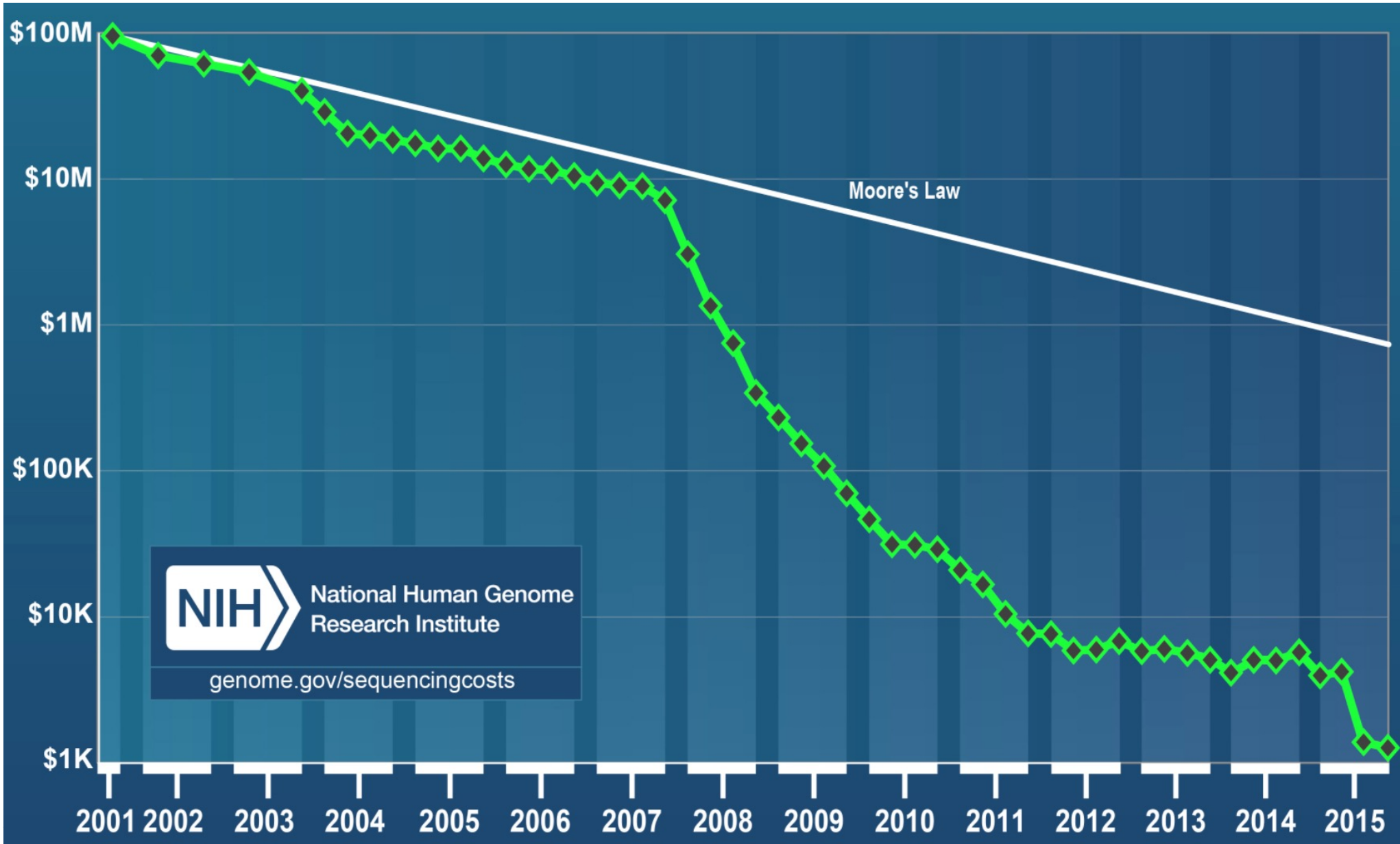
facebook

Connecting the world takes every one of us.



...how much is Facebook worth to you?

Cost to sequence a single human genome.

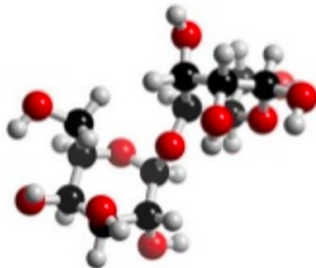


Data Science and Bioinformatics

Sequences

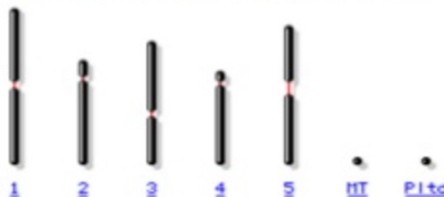


Structure

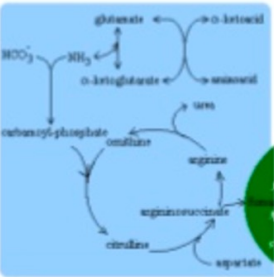


Genome

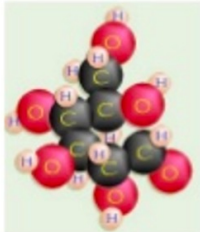
Complete Genome of Arabidopsis thaliana



Pathways



Lipids, Carbohydrates



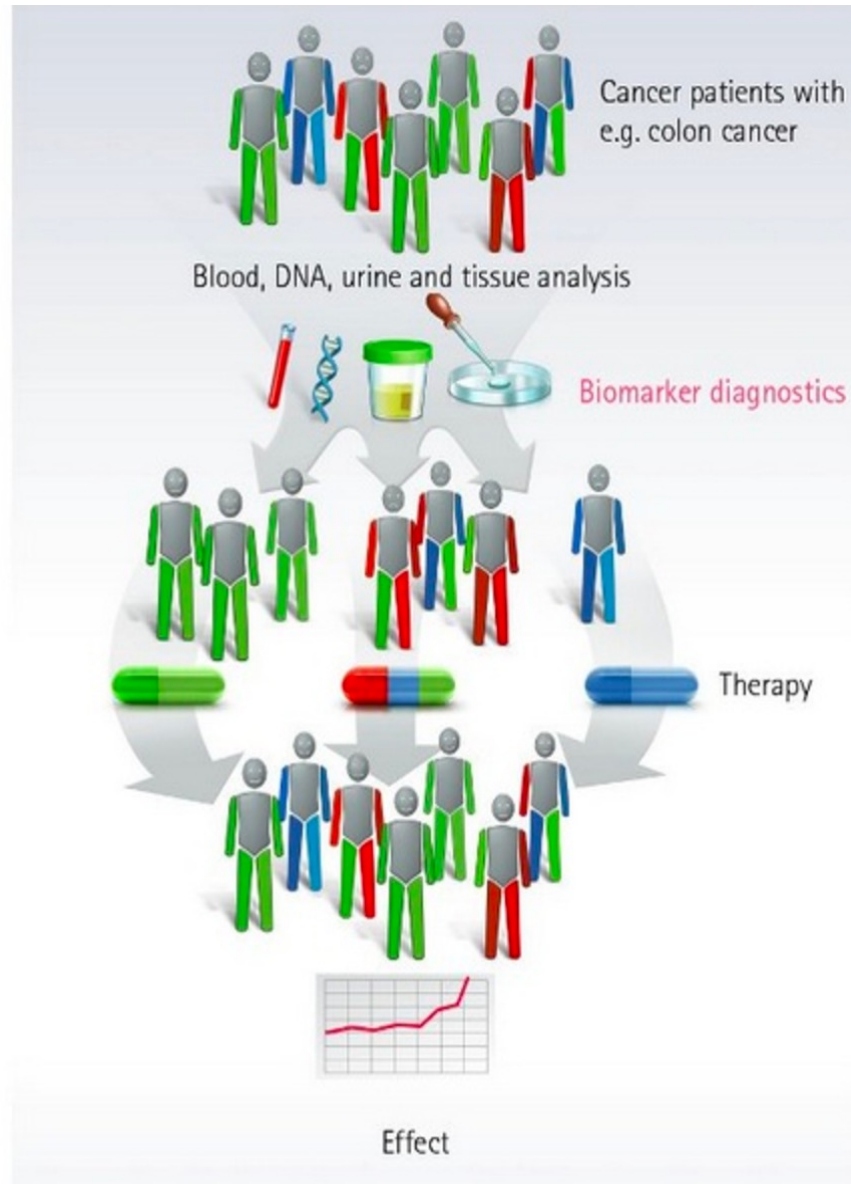
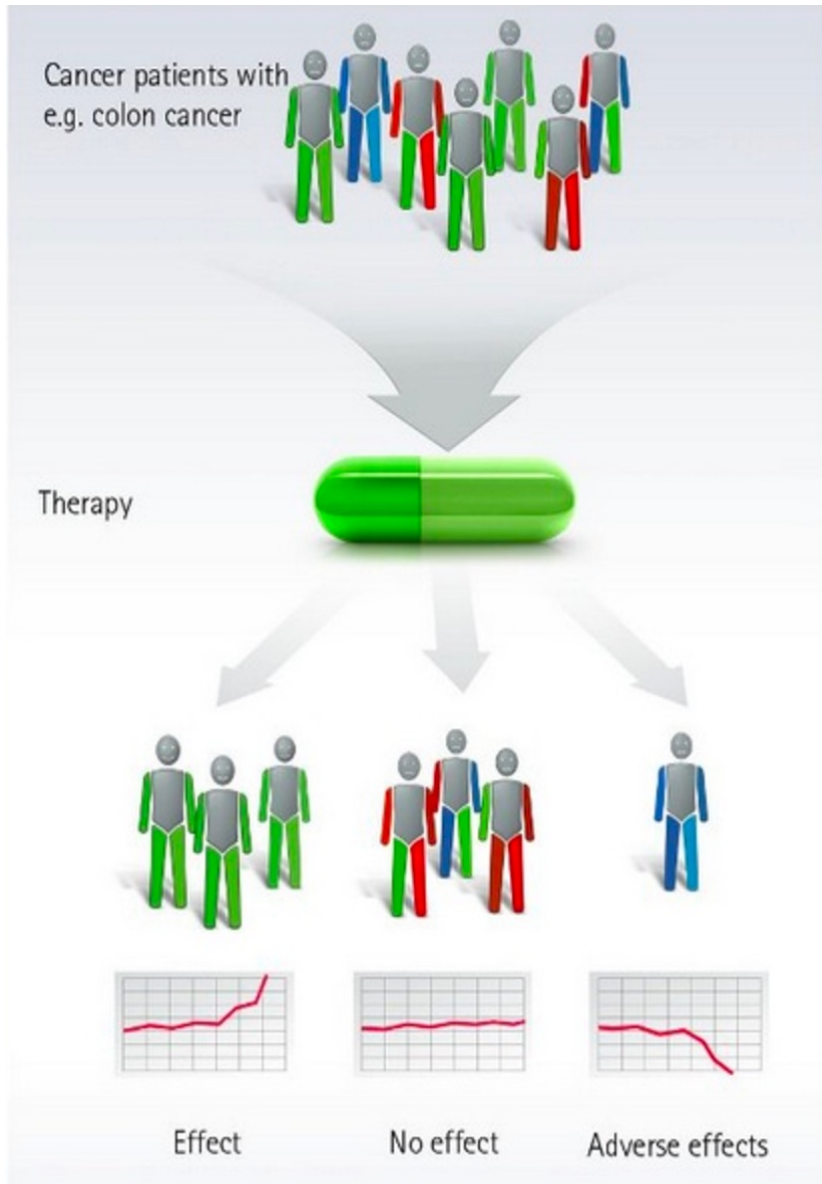
Literature



Primary Database: Direct experimental results
Secondary Database: Derived databases from transformation & analysis

Source: <https://www.slideshare.net/shwetakagliwal/biological-databases-11267007>

Data Science and personalized medicine



The promise of personalized medicine:

- Improved efficacy
- Reduced adverse side-effects
- Reduced “trial-and-error” delays in the treatment of time-critical diseases such as cancer.

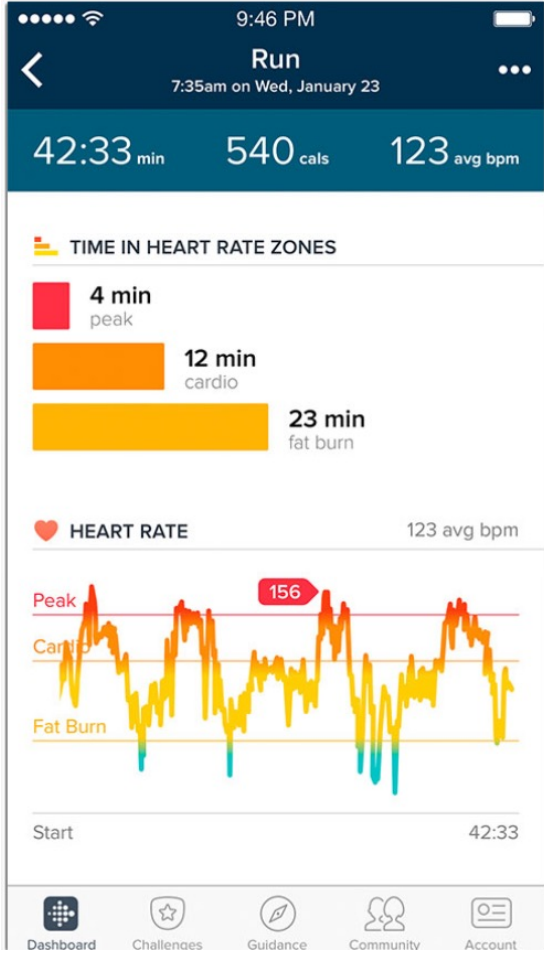
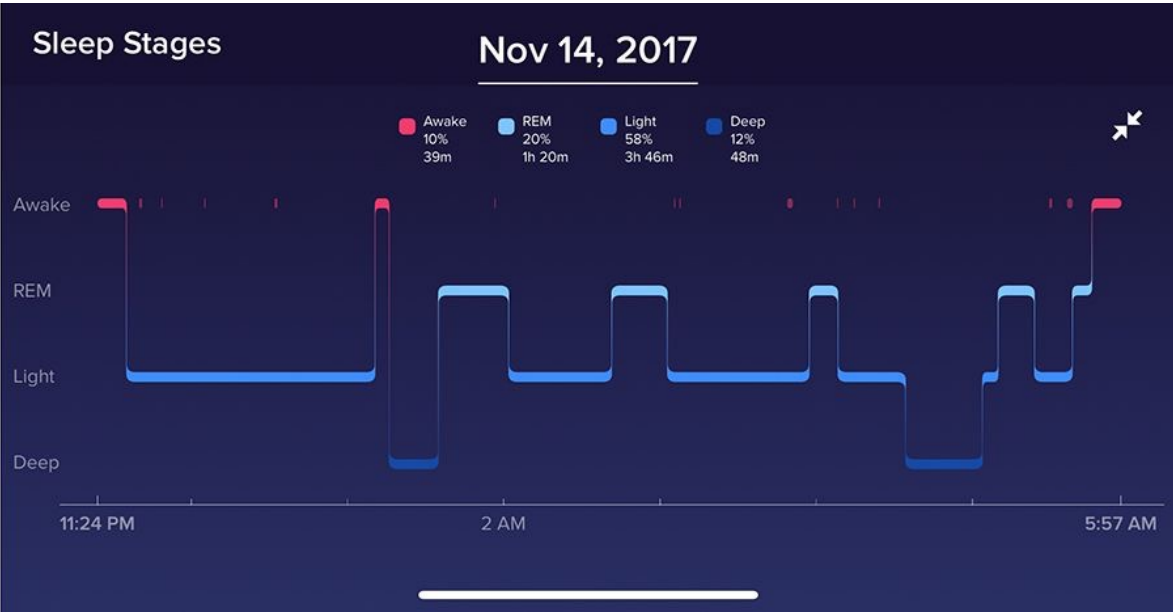
Data Science and Medicine: Electronic Medical Records



- Patient Demographics
- Doctors
- Insurance
- Medical history
- Allergies
- Procedures / Measurements (BP, Temp, O₂)
- Order Lab Tests → Lab Results
- Dr. Notes

Next time you get a physical, notice how much time your doctor spends at a computer terminal!

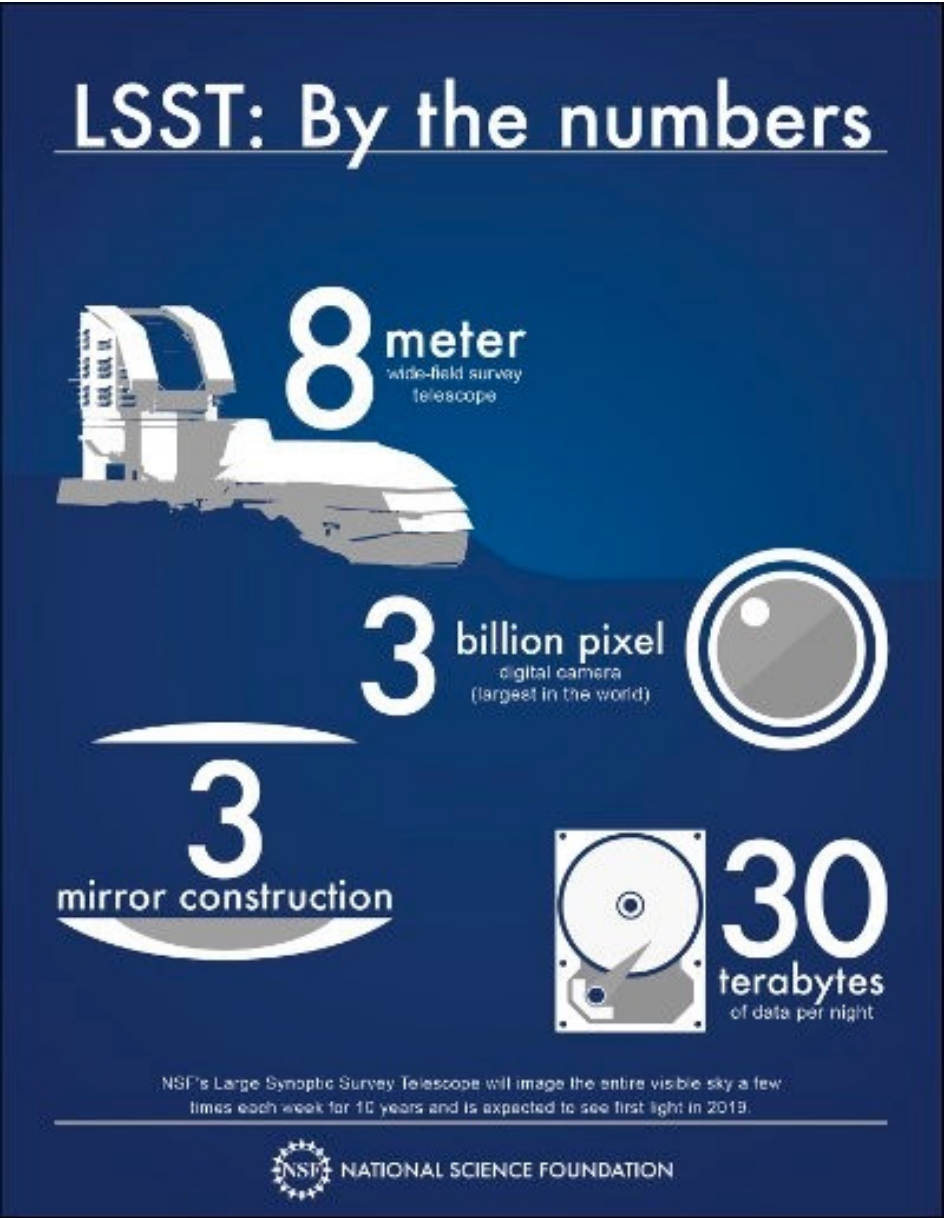
Data Science and Wearable Sensors for better health



<https://finance.yahoo.com/news/exclusive-fitbits-6-billion-nights-sleep-data-reveals-us-110058417.html>

Astronomical Data Science: Understanding our place in the Universe

LSST: By the numbers



The infographic features four main statistics arranged in a 2x2 grid. Top-left: '8 meter wide-field survey telescope' with a 3D cutaway of the telescope. Top-right: '3 billion pixel digital camera (largest in the world)' with a circular camera lens icon. Bottom-left: '3 mirror construction' with a curved mirror icon. Bottom-right: '30 terabytes of data per night' with a hard drive icon. At the bottom, there is a line of text and the NSF logo.

8 meter wide-field survey telescope

3 billion pixel digital camera (largest in the world)

3 mirror construction

30 terabytes of data per night

NSF's Large Synoptic Survey Telescope will image the entire visible sky a few times each week for 10 years and is expected to see first light in 2019.

NSF NATIONAL SCIENCE FOUNDATION

The LSST will be an 8-meter wide-field survey telescope that will image the entire visible sky a few times each week for 10 years, providing an unprecedented amount of information while transforming the emerging discipline of data-enabled science. It is expected to see first light in 2019 and begin full operation in 2022.

