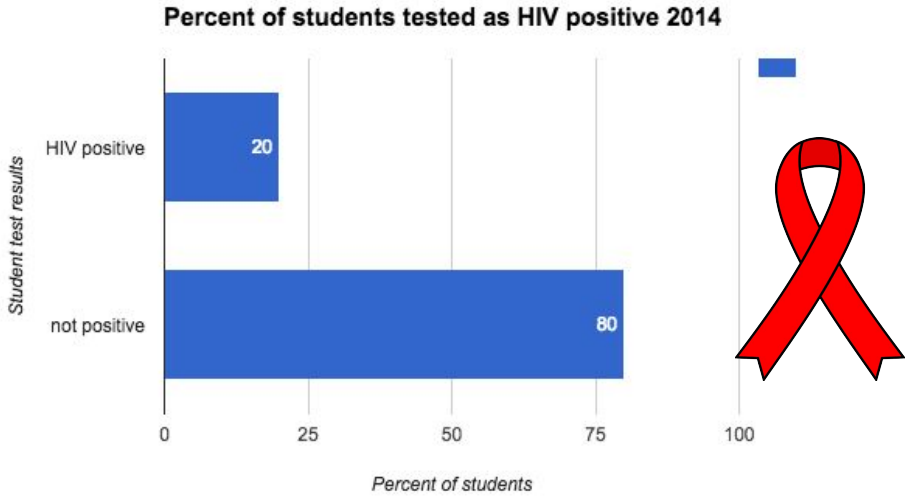
1/8/2019

Finding data

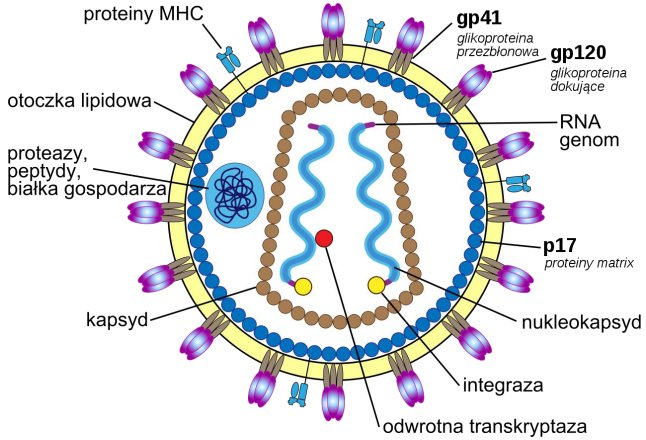


Journal: Given this data:

Blood drive at the local high school reveals that 20% of the students were HIV positive.

- What is your immediate reaction?
- What questions do you have?

Possible questions:



Where was the school?

Who collected the data?

When was the data from?

Was it only kids 16 and older who can give blood?

Is it true?

Article from snopes.com (Snopes Article).

What generates data online?

- Personal
- Business
- Government
- Internet of Things

What kinds of data do these generate online?

Calendar

7	8 4-1-A Data Acquisition & Analysis HW: Google CR	9 10 4-1-B More Data Acquisition & Analysis HW: Google CR	11 4-2 Models & Simulations HW: Google CR	
14 4-3 Using Data Simulations HW: Google CR	15 4-4-A File Input & Output w/Python HW: Google CR	16 17 4-4-B More File Input & Output w/Python HW: Finish In-Class Work	18 4-5-A Data Collection, Analysis & Simulation HW: Finish Classwork	
21 NO SCHOOL MLK Jr. Day	22 4-5-B More Data Collection, Etc. HW: Finish Classwork	23 24 4-6-A Hypothesis Testing with Simulations Using NetLogo NO HW	25 4-6-B More Hypothesis Testing HW: Finish Classwork (Due 1/29)	
28 4-6-C Hypothesis Testing Conclusion HW: Finish Classwork	29 4-7 Review For Test	30 31 4-8 **Test** Work on Explore Task	Finish Explore Task	Computer is used today

Python Everywhere Textbook
https://bit.ly/2gSWkSS

Python 3.6 Quick Reference
https://bit.ly/1QbX27o

Video: Using files in Python
https://bit.ly/2TrEAMb

AP CSP

Unit 4: Data Acquisition

Numbers have an important story to tell. They rely on you to give them a voice."
-Stephen Few

When we are looking at data

Explore

Question - I wonder

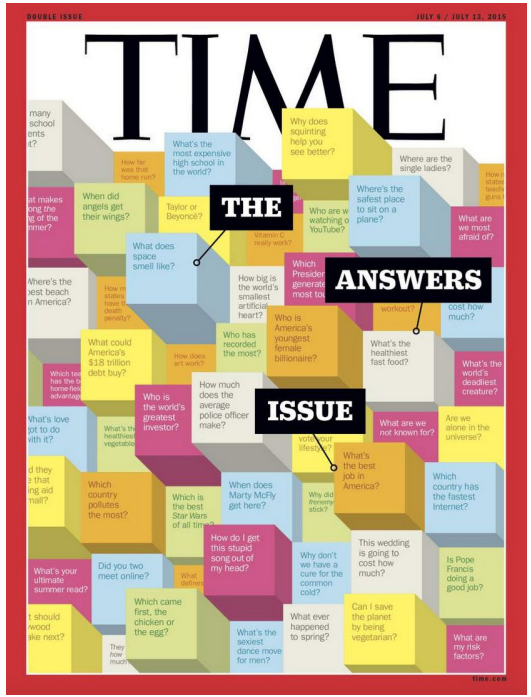
Hypothesis - I think

Test

How do we get information from data?

From Time magazine July 6-13, 2015:

"every 2 days, humanity creates a quantity of data equivalent to the entire amount created from the dawn of time up until 2003."



Some organizations are trying to organize and make data available:

What is Wolfram|Alpha?

Wolfram|Alpha computes answers and provides knowledge.

Video Introduction:

Wolfram|Alpha in a nutshell

Wolfram|Alpha is an engine for **computing answers** and **providing knowledge**.

It works by using its vast store of expert-level knowledge and algorithms to automatically answer questions, do analysis, and generate reports.

Wolfram|Alpha's long-term goal is to make all systematic knowledge computable and broadly accessible.

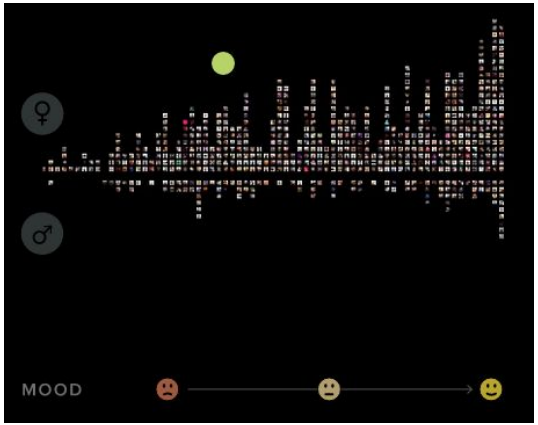
Analyzing data

You have to gather the data and analyze it to create meaning.

http://selfiecity.net/ what do we know from analyzing selfies?
(see results from 5 different cities)

How did they figure it out?

- Bangkok
- Berlin
- Moscow
- New York
- Sao Paulo



Which city smiles the most?

Could your photo be in there?

Did everybody give their permission to have their photo used?

"Is it just me, or do Sao Paulo women actually tilt their heads more? Do New Yorkers or Berliners look older?"

In order to answer these questions, and supplement our rudimentary automatic face analysis with human judgment, we had thousands of photos inspected by people, who estimated age and gender of the people on the photos.

Collaborate

Share/Refine/Retest

Reflect

Data is pervasive ... and growing

Data is everywhere and constantly being generated. Beyond a point, the more data you have, the harder it is to extract meaning from it.

Explore a 365-gigapixel image of the Mont Blanc. If the picture is square, how big is each side? Will it fit on a HD TV screen?
http://www.in2white.com/

Quick review:
Define these and put them in order:

MB, bit, TB, ZB, byte, GB, pixel, GB, KB, PB

http://highscalability.com/blog/2012/9/11/how-big-is-a-petabyte-exabyte-zettabyte-or-a-yottabyte.html

Everything leaves a trail, even soda machines, pacemakers, webcams, home alarms, everything that's connected. Can anybody make use of all that data or is it too big to look at?

Homework - google classroom

For one of the TED talk videos listed:

1. Write a short 2-3 sentence original summary of the video.
2. What is the most striking information in the video?
3. What kind of data is being discussed?
4. How is the data being generated, collected and analyzed?
5. What additional insight and knowledge could be gained from this data beyond what is described in the video?
6. What are some considerations and trade-offs that come to mind about this data? (is it complete? does it invade any privacy? it is accurate? is it up to date? can we be sure that the conclusions are true? does it change rapidly?)