

1/10/2019

Finding and analyzing data

Journal: Describe at least 2 ways that we create meaning out of data.

Present findings from homework



David McCandless

The beauty of data visualization

David McCandless turns complex data sets (like worldwide military spending, media buzz, Facebook status updates) into beautiful, simple diagrams that tease out unseen patterns and connections. Good design, he suggests, is the best way to navigate information glut — and it may just change the way we see the world.



Anne Milgram

Why smart statistics are the key to fighting crime

When she became the attorney general of New Jersey in 2007, Anne Milgram quickly discovered a few startling facts: not only did her team not really know who they were putting in jail, but they had no way of understanding if their decisions were actually making the public safer. And so began her ongoing, inspirational quest to bring data analytics and statistical analysis to the US criminal justice system.

 Watch later · 155 comments >



Aaron Koblin

Visualizing ourselves ... with crowd-sourced data

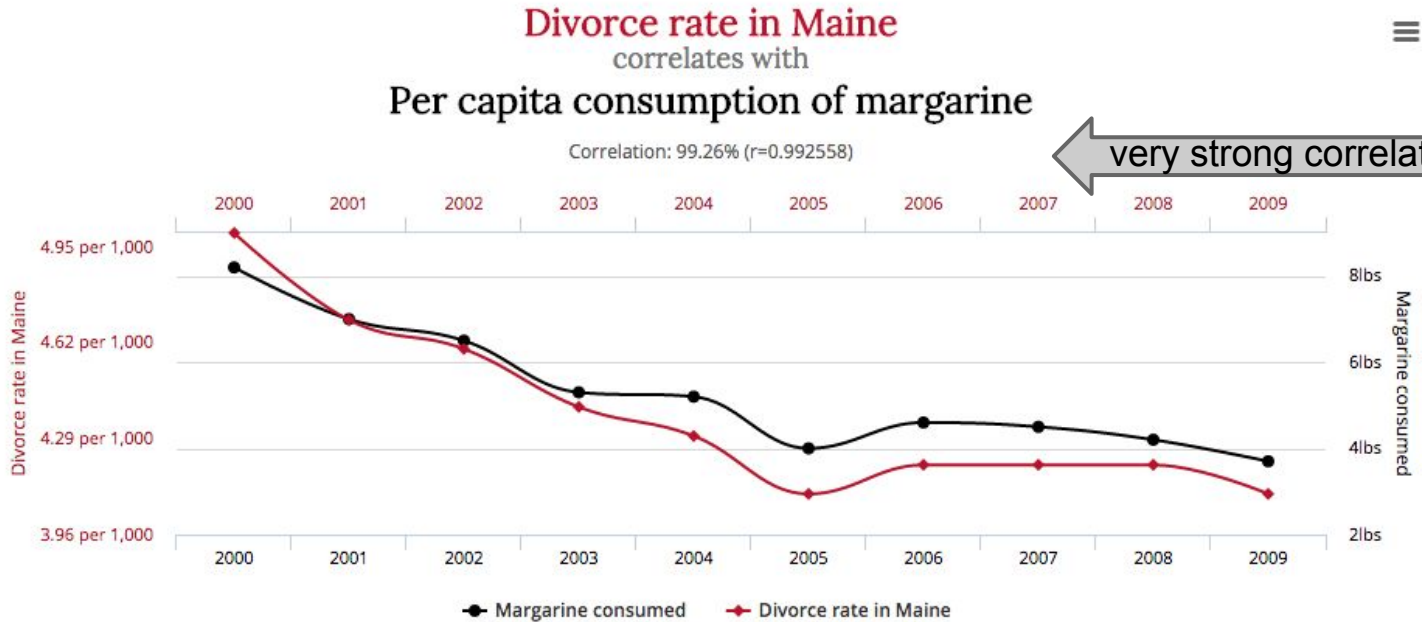
Artist Aaron Koblin takes vast amounts of data — and at times vast numbers of people — and weaves them into stunning visualizations. From elegant lines tracing airline flights to landscapes of cell phone data, from a Johnny Cash video assembled from crowd-sourced drawings to the "Wilderness Downtown" video that customizes for the user, his works brilliantly explore how modern technology can make us more human.

TED.com

For one of the TED talk videos:

1. Write a short 2-3 sentence summary about 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?)

Does this data have meaning?



Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

Does eating margarine cause divorce in the state of Maine?

Does less divorce in Maine lead to fewer people buying margarine?

What facts can we state from this data? A statistical analysis states there is > 99% correlation.

<http://www.tylervigen.com/spurious-correlations>

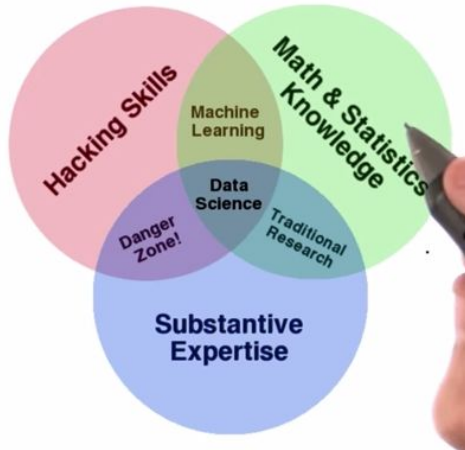
Data Science: a rapidly growing new field

Video from Udacity: What is a data scientist (0:54) and what does a data scientist do (0:39)

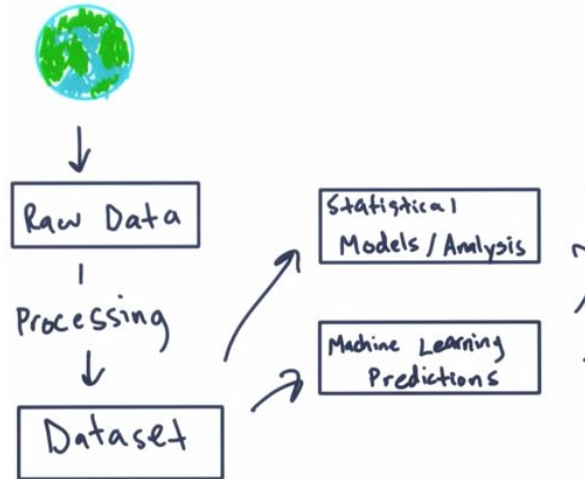
<https://www.youtube.com/watch?v=9PIqjaXJo7M>

<https://www.youtube.com/watch?v=vowXaEDh1uk>

What is a data scientist?



What does a data scientist do?



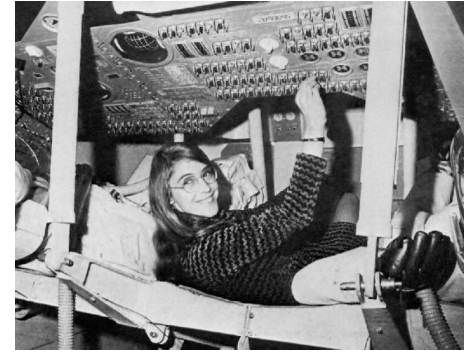
What considerations and tradeoffs arise in the computational manipulation of data?

3 false assumptions about big data

1. It's complete and accurate
2. It tells the whole story
3. Bigger is better

Things to decide about computing data

1. How do you account for missing data?
How do you certify your sources?
2. How do you decide which data to include and which to exclude?
3. How much data is enough? (time is money!)
4. Are your processing algorithms accurate?



Hamilton during her time as lead Apollo flight software designer

Do I have the right data? Am I taking everything into consideration? Are the calculations accurate? Am I taking the limits of the computer into consideration?

Margaret Hamilton standing next to listings of the actual Apollo Guidance Computer (AGC) source code



What are some other ways data is used?

Businesses that succeed because they know their customers.

NETFLIX
amazon.com[®]
and you're done.[™]

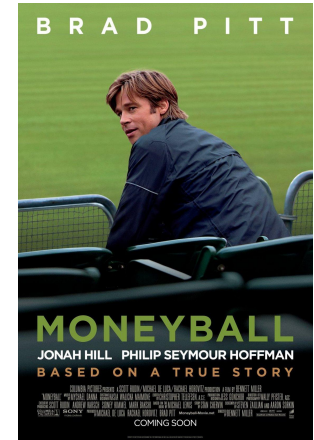
<http://www.smartdatacollective.com/bernardmarr/312146/big-data-how-netflix-uses-it-drive-business-success>

<https://commons.wikimedia.org/wiki/File:NewNetflixLogo.png>

<http://www.fastcompany.com/3024655/pitch-perfect-and-how-analytics-are-transforming-movie-marketing>

https://commons.wikimedia.org/wiki/File:Amazon_logo.jpg

Teams that make decisions using data



[https://en.wikipedia.org/wiki/Moneyball_\(film\)](https://en.wikipedia.org/wiki/Moneyball_(film))

<https://www.flickr.com/photos/wolfgangkuhnle/6837360817>

<http://www.forbes.com/sites/aliciajessop/2013/05/28/why-the-kings-are-staying-in-sacramento-meet-vivek-ranadive/>

http://commons.wikimedia.org/wiki/File:Phoenix_suns_kings.jpg

Data analysis requires an algorithm

What is a possible algorithm for making a decision about choosing what movies Netflix might suggest for a customer?

Describe at least 2 calculations needed

- a. To find which genre of movie has the highest like rating for this customer
add up all the star ratings for every movie rated and find the genre with the highest average star rating
- b. To find how much does the customer like to vary the types of movies they watch from one week to another or on a specific day of the week or time of day?
Add up star ratings by the day of the week and then by the time of day, calculate if there are significant differences between days or times. Compare to similar viewers.

Describe some of the data you'd need to collect.

1. Date and time a movie is watched (streaming), star rating, genre of movie, length of movie, do they like to rewatch the same movies? Data from similar viewers.

Data analysis requires an algorithm

What is a possible algorithm for making a decision about choosing:

1. what movie to produce
or
2. what sports player to
hire

Choose one of the options and write an outline of an algorithm.

1. Describe at least 2 calculations needed
2. Describe some of the data you'd need to collect.

Share and discuss.

Do you own Data Search and Analysis

On Google Classroom

One Document

Work as a team to complete the entire data collection/analysis

Homework: 4-1-B (2)



Watch the following video: <https://www.youtube.com/watch?v=2LPboySOSvo> (5:15)

Answer the following questions about the video:

- What data had to be gathered before the simulation was created?
- What data can be gathered from the simulation?
- What is an advantage to having so many data points?
- What is a disadvantage?
- Brainstorm at least 5 other systems (body or otherwise) that a simulation might be helpful to create.