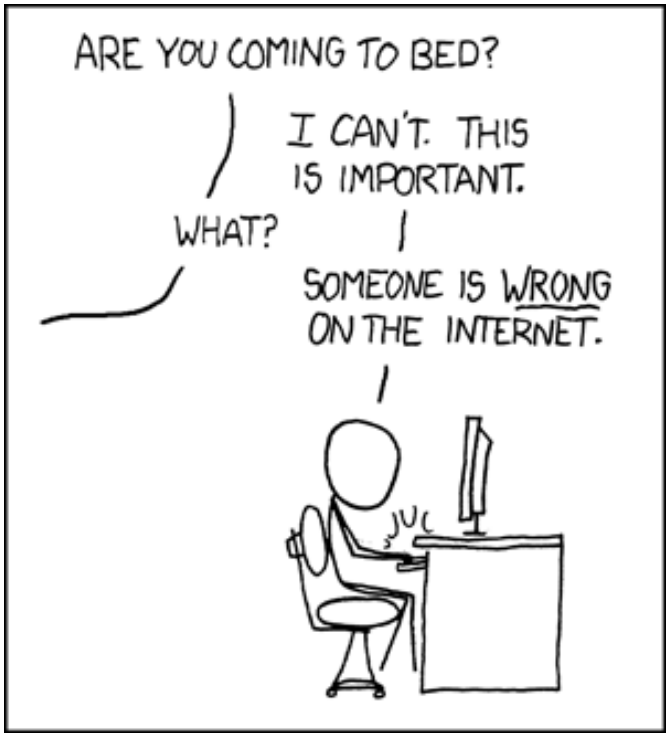
Evaluating Information

Michael Paul
Feb 10, 2016

University of Colorado
CMCI 1020




THE WØRD
Truthiness

How do we evaluate information?

Two aspects:

- Look at the quality of the **content**
- Look at the **metadata** for more context and background
  - Metadata is “data about data” – additional information about data such as author and date

Metadata

Some questions for evaluating metadata:

- Who wrote the information?
- Who edited/published the information?
- When was the information written?
- How was the information created/changed?

Metadata

- **Author:**
  - Writes the majority of the information; writes first draft
- **Editor:**
  - Makes changes to the information
  - Decides what information gets published
- **Publisher:**
  - The institution that releases the information

Scavenger hunt!

- Who authored the article?
- Who edited the article?
- Who published the article?
- When was the article published?
- When was the article written?
- Was the article revised? When?

Scavenger hunt!

Group A: New York Times

- http://tinyurl.com/cmci1020a

Group B: Wikipedia

- http://tinyurl.com/cmci1020b

Group C: BuzzFeed

- http://tinyurl.com/cmci1020c

Metadata

Other things to consider…

Authorship

Evaluating authorship in **social media:**

- Is the person posting reputable?
  - One clue: look at the social network of the user
    - Do they have a high number of friends/followers?
    - Are the friends/followers themselves reputable?
  - Another clue: how recently was the account created?
    - Spam/troll accounts are often newly created

Authorship

Evaluating authorship in **social media:**

- What is the location of the person posting?
  - Is it relevant to the information?

Example: if you're reading Twitter to learn about floods in CO, tweets from users who live in CO may be more reliable

Publishing

How was the information published?

- Is it published by an individual or an institution?
- If published by an institution, who is the institution?
  - **.com** – commercial company
  - **.edu** – educational institution
  - **.org** – non-profit organization
  - **.gov** – government agency

Publishing

How was the information published?

Academic websites are usually reliable, but difference between institution pages and personal pages

cmci.colorado.edu – the official College website
cmci.colorado.edu/~mpaul – personal website

Publishing

How was the information published?

.org websites are often public interest groups and political think tanks, and may present biased info.

- You should investigate the credentials and agendas of organizations just as you would investigate authors

Currency

When was the information written?

Make sure the date is when you think!

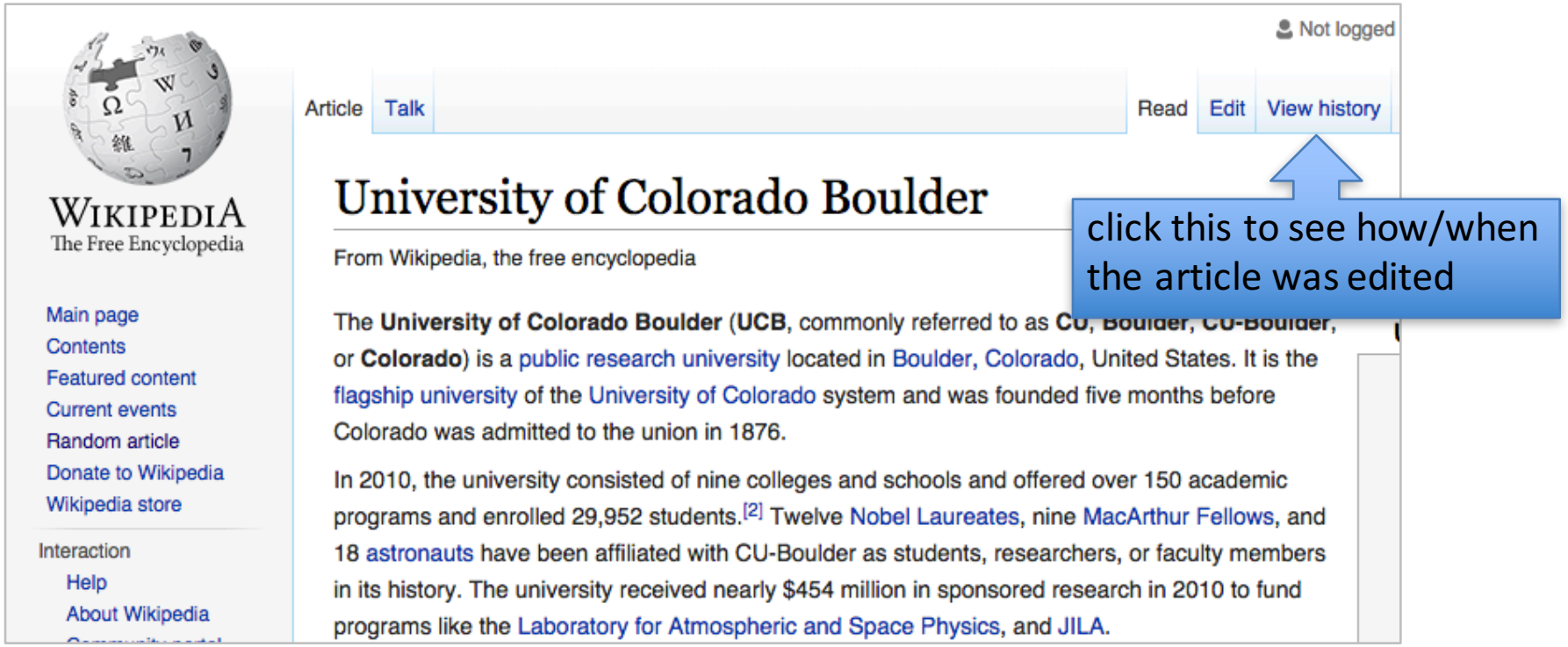
- Some people link to old information as if it's new

And be aware that the "truth" changes over time

  - based on new research, new discoveries
- Old information might be wrong, even if it was correct when it was published

Currency

How has information **changed over time**?



University of Colorado Boulder: Revision history

View logs for this page

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Atom
Upload file
Special pages
Page information
Wikidata item

Languages

Browse history

From year (and earlier): 2015 From month (and earlier): all Tag filter: Show

For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary.
External tools: Revision history statistics · Revision history search · Edits by user · Number of watchers · Page view statistics

(cur) = difference from current version, (prev) = difference from preceding version,
m = minor edit, → = section edit, ← = automatic edit summary
(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

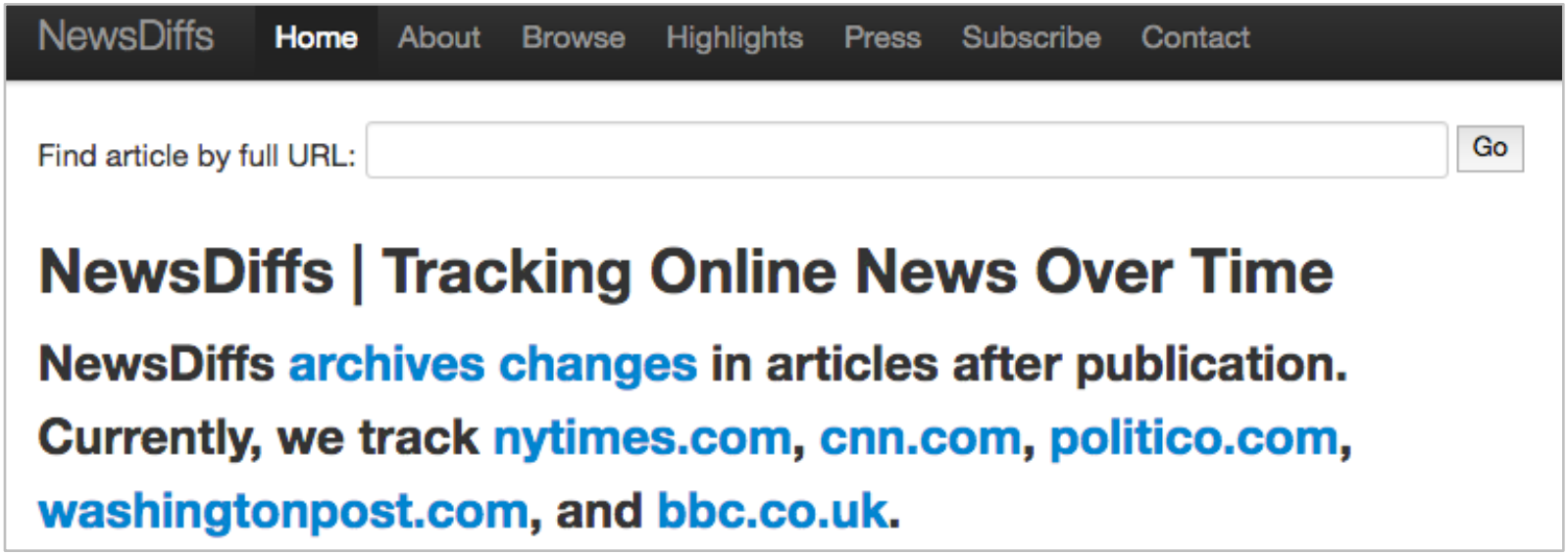
- (cur | prev) 18:51, 18 December 2015 Ccawblake (talk | contribs) . . (61,067 bytes) (+436) . . *(links)* (undo) *(Tag: Visual edit)*
- (cur | prev) 15:11, 10 December 2015 SSTflyer (talk | contribs) . . (60,631 bytes) (-62) . . *(Cat-a-lot: Removing from Category:Association of Public and Land-Grant Universities)* (undo)
- (cur | prev) 01:25, 9 December 2015 Contributor321 (talk | contribs) . . (60,693 bytes) (+9) . . *(→Rankings: correction)* (undo)
- (cur | prev) 04:54, 7 December 2015 98.245.125.79 (talk) . . (60,684 bytes) (+221) . . *(→Rankings)* (undo)
- (cur | prev) 07:38, 29 November 2015 Xenophrenic (talk | contribs) . . (60,463 bytes) (-372) . . *(verdict of 'wrongly fired' was not vacated; the $1 award was vacated due to quasi-judicial immunity)* (undo)
- (cur | prev) 04:39, 29 November 2015 Luftmensch~enwiki (talk | contribs) . . (60,835 bytes) (+39) . . *(→Faculty: Updated information about court case involving W Churchill)* (undo)
- (cur | prev) 04:34, 29 N nber 2015 Oshwah (talk | contribs) **m** . . (60,796 bytes) **(+3,710)** . . *(Reverted edits by Luftn t (HG) (3.1.18))* (undo)

click a date to see how the article looked at that time

Currency

How has information **changed over time**?

Useful tool: **newsdiffs.org**



Currency

How has information **changed over time**?

Useful tool: **archive.org**

Currency

How has information **changed over time**?

Useful tool: **archive.org**

Limitations:

- Doesn't have complete revision history
  - Only periodic snapshots
- Doesn't store the entire Internet
  - But, it's a good option when websites don't provide their own revision history (most don't – Wikipedia is exception)

Currency

How has information **changed over time**?

Conclusion: the web is not static!

- Not enough to know **where** information came from, but also **when**

Provenance

How was the information created?

In the context of data, **provenance** may include:

- revision history (as you just saw)
- a record of authors/owners over time
- additional notes about the design/creation of the data

Provenance is not a common part of metadata, but more digital systems are creating a place for it

How do we evaluate information?

Conclusion:
Metadata can help a lot with evaluation!

Next:
What about the actual content of the data?

Content

Some questions for evaluating content:

- Who was the information written for?
- Is the information biased?
- Is the information accurate?
- Is the information complete?

Audience

Who was the information written for?

- Written for experts or the general public?
- Does the information use technical language?
- Is it written at the right level for your needs?
- Is it intended to argue a viewpoint?
  - watch out for **bias**

Accuracy

Is the information accurate?
How can you tell?

- Is the content presented clearly?
  - Poor presentation and typos are red flags
- Are sources cited?
  - If so, those should be examined as well
- Does the content match other sources?
  - Very important to look at multiple sources for the same information, see how they compare!

Accuracy: Fact Checking

Fact checkers verify if statements are accurate

Fact checking is a standard part of journalism

- Articles are verified before being published
  - This isn't always perfect

Fact checking can also happen after events such as political debates

- A number of organizations exist:

FACTCHECK.ORG

POLITIFACT
WINNER OF THE PULITZER PRIZE

The Washington Post
Fact Checker

POLITIFACT
WINNER OF THE PULITZER PRIZE

"Limberbutt McCubbins (a five-year-old cat) is a candidate" in the 2016 presidential election.

— *Limberbutt McCubbins on Thursday, July 9th, 2015 in comments on his website.*

Can a cat run for president?

By Linda Qiu on Tuesday, July 14th, 2015 at 4:37 p.m.

Our ruling

McCubbins, a five-year-old cat, said, "Limberbutt McCubbins is a candidate" in the 2016 presidential election. Yes, this is the first time we're fact-checking a claim made by a cat. (Though we have fact-checked a terrier who supported Mitt Romney in 2008.)

Limberbutt's campaign manager has filed official paperwork, but the FEC doesn't deem him formally a candidate, because he hasn't spent or received $5,000. (This is also the case for some human candidates.) Experts told us it's very unlikely that he'll appear on any ballots as a candidate, and it's even more unlikely that his candidacy will stand in a court of law.

We rate Limberbutt's claim Half True.

Sources:

Limberbutt2016.Com, Limberbutt 2016, accessed July 13, 2015

Email interview with Isaac Weiss, July 13-14, 2015

Email interview with Michael Gilbert, professor of election law at the University of Virginia, July 14, 2015

Email interview with Sarah Duggin, professor of Constitutional law at Catholic University, July 13, 215

Interview with Richard Winger, publisher and editor of Ballot Access News, July 14, 2015

Email interview with Katherine Sibley, professor of American studies at St. Joseph University, July 13, 215

Email interview with Christian Hilland, spokesperson for the Federal Election Committee, July 13, 2015

Federal Election Committee, 2016 Presidential Form 2 Filers, July 14, 2015

Federal Election Committee, Quick Answers to Candidate Questions, accessed July 13, 2015

Federal Election Committee, Statement of Candidacy, May 6, 2015

Accuracy: Fact Checking

Good resource for non-political fact checking: **Snopes**

- Specializes in rumors, urban legends

CLAIM: A fan was beaten up by moviegoers after spoiling the new Star Wars movie.

FALSE

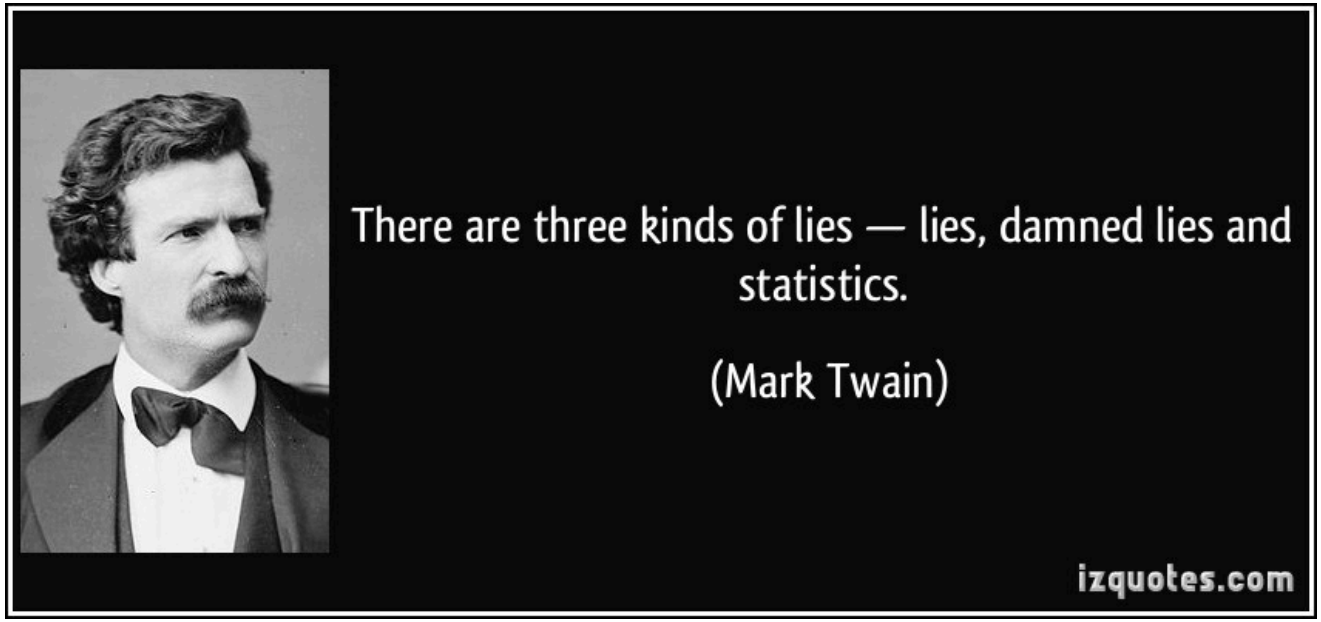
Accuracy: Statistics

What about evaluating **data and statistics**?

Accuracy: Statistics

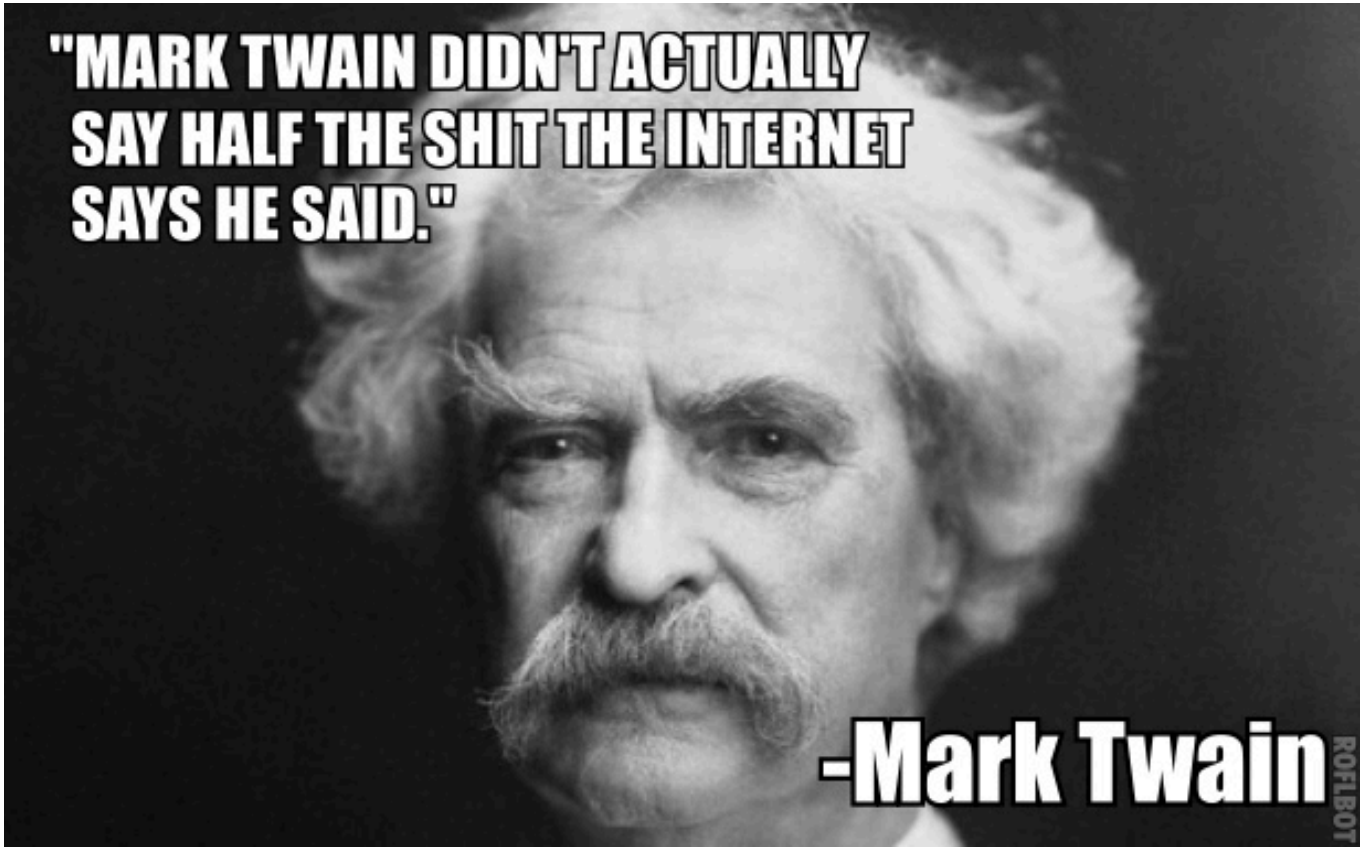
What about evaluating **data and statistics**?

Often, numbers can be technically accurate BUT the way they are **presented/interpreted** is inaccurate



Accuracy: Statistics

(disclaimer)

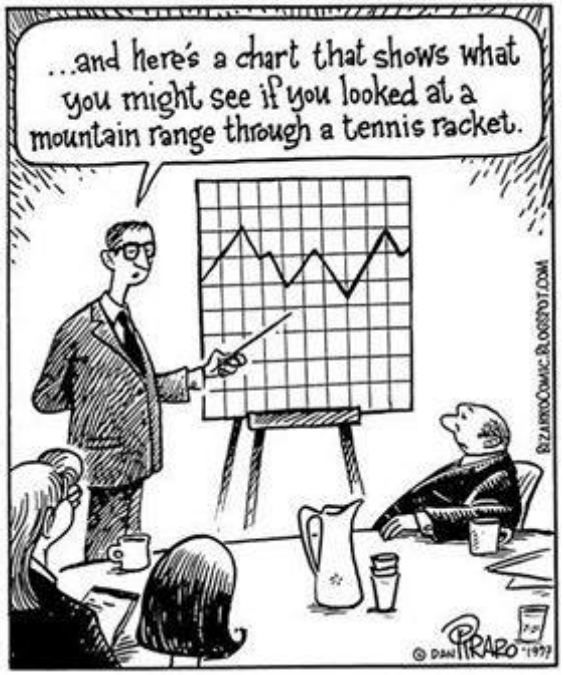


Accuracy: Statistics

Why might statistics be misleading?

People often don't evaluate statistics critically:

- Information looks more accurate if it is supported by numbers
- Statistics are perceived as "objective" metrics, with no room for human bias
- Statistics can be hard to understand, even by experts

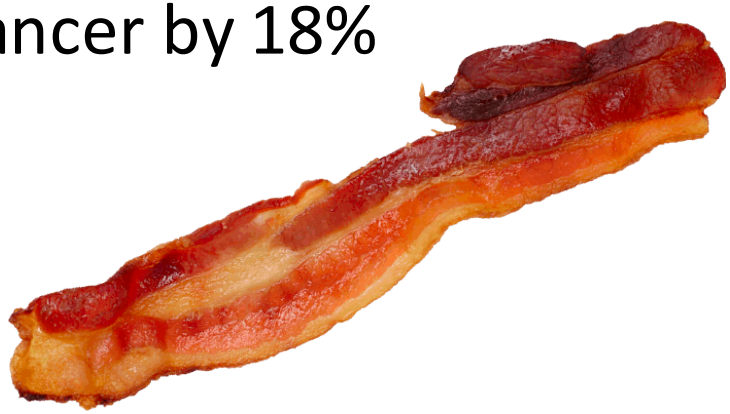


Accuracy: Statistics

Know the difference between **relative** and **absolute** percentages

Bacon (and other processed meats) were recently classified as carcinogens

- Daily consumption of processed meat will increased chance of getting colorectal cancer by 18%
  - So what does that mean?

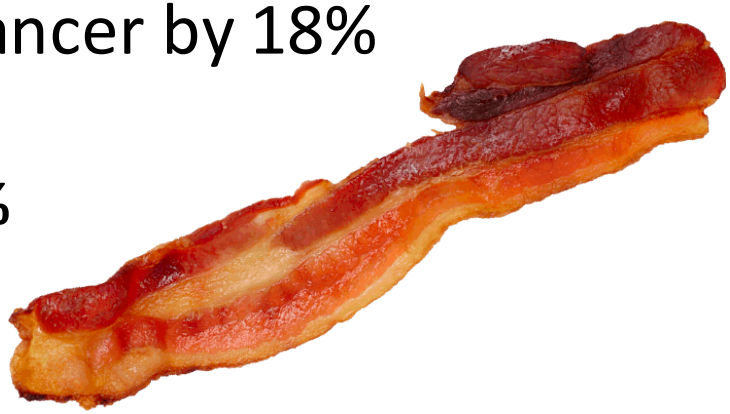


Accuracy: Statistics

Know the difference between **relative** and **absolute** percentages

Bacon (and other processed meats) were recently classified as carcinogens

- Daily consumption of processed meat will increased chance of getting colorectal cancer by 18%
  - Chance of colorectal cancer in the US general population: **4.5%**
  - Chance after daily consumption of processed meat: **5.3%**

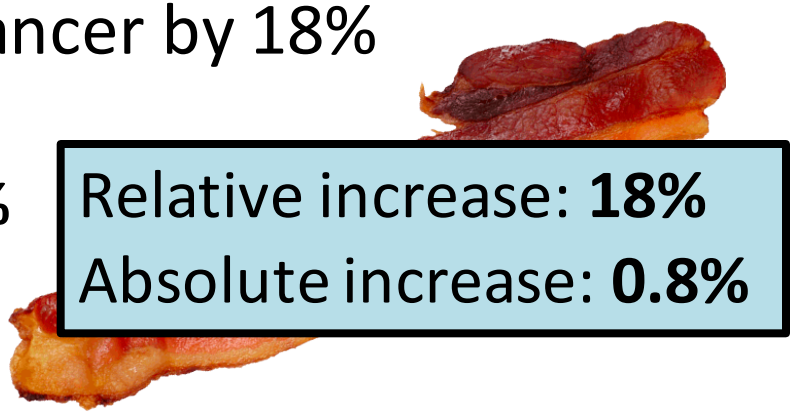


Accuracy: Statistics

Know the difference between **relative** and **absolute** percentages

Bacon (and other processed meats) were recently classified as carcinogens

- Daily consumption of processed meat will increased chance of getting colorectal cancer by 18%
  - Chance of colorectal cancer in the US general population: **4.5%**
  - Chance after daily consumption of processed meat: **5.3%**



Relative increase: **18%**
Absolute increase: **0.8%**

Accuracy: Statistics

Know the difference between **relative** and **absolute** percentages

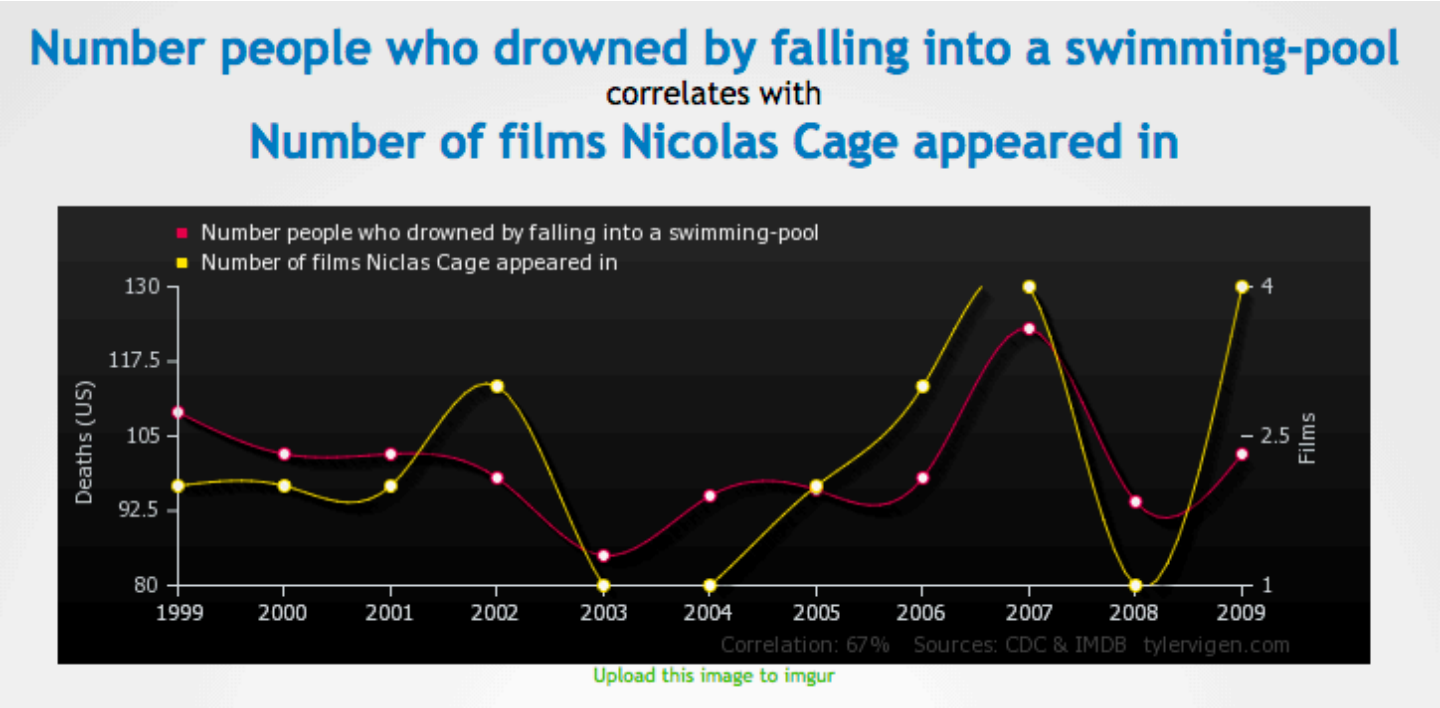
Recent study:

- Taking antidepressants during pregnancy increases risk of autism by 72%
  - Risk of autism in the general population: **0.71%**
  - Chance after taking antidepressants in 2nd or 3rd trimester: **1.22%**

Relative increase: **72%**
Absolute increase: **0.5%**

Accuracy: Statistics

Some things happen by **coincidence**!



Proving that one event causes another requires additional experimentation beyond simple statistics

Accuracy: Statistics

What about evaluating **data and statistics**?

Conclusion: Statistics can be misleading or misunderstood, and should be evaluated critically

Note: statistics are not always lies! (sorry Mark Twain)

- Statistics can be very informative and they often ARE more objective than other types of information
- But proceed carefully

Coverage

Is the information complete?

- Are you seeing all of the relevant information?
- Do you have the “big picture”?

This can be hard:

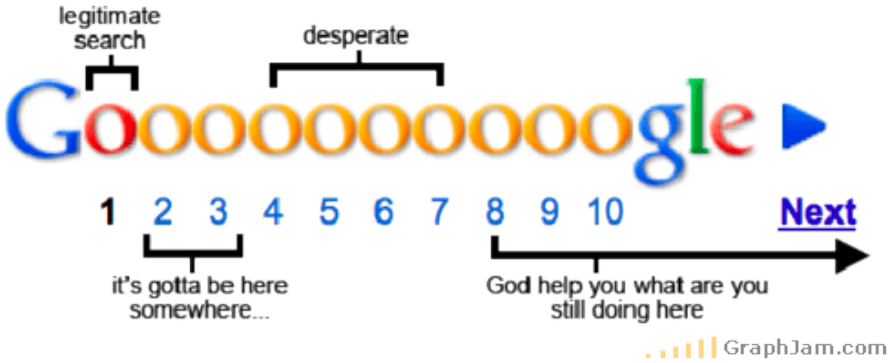
There is more information out there than you can read!

- But it helps to know about common pitfalls

Coverage: Filtering

There is too much information out there to read, so we rely on search engines to find what we need

- But how many search results do you really read?



What shows up on the first page matters!

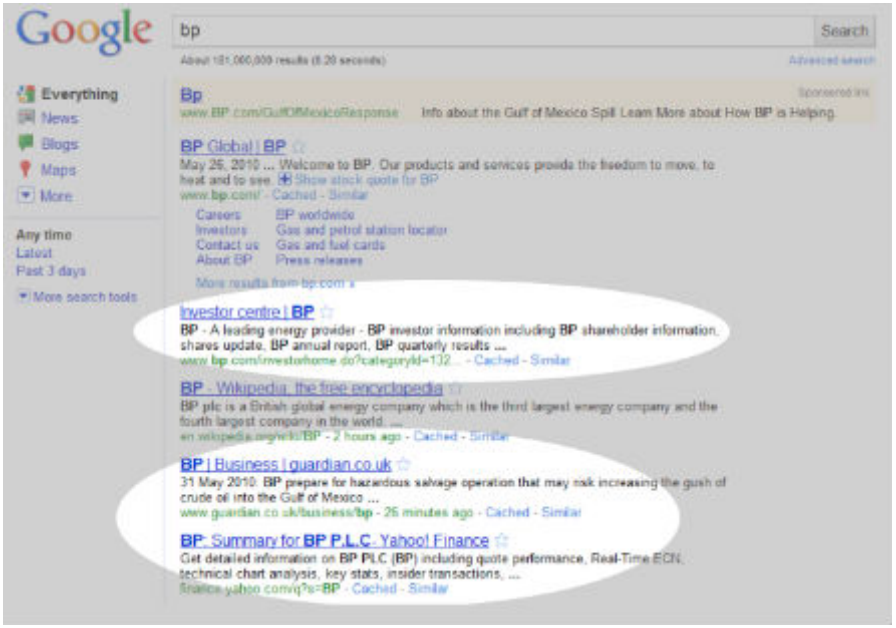
- But not everyone even sees the same results

Coverage: Filtering

Different people get different Google results when searching for “BP”

Oil spill news:

Financial/stock info



Source: http://dontbubble.us/

Coverage: Filtering

Why do people get different results?

A common reason: **personalization**

- Google's system guesses what you want to see, based on:
  - Your location
  - Your demographics
  - Your interests (what have you searched in the past?)

Coverage: Filtering

Why do people get different results?

A common reason: **personalization**

- Google's system guesses what you want to see, based on:
  - Your location
  - Your demographics
  - Your interests (what have you searched in the past?)

This is also true of Facebook's news feed:
it algorithmically selects what updates to show you

Coverage: Filtering

Why do people get different results?

Another reason: **A/B testing**

- Google gives different users different results as a way of experimenting with their system
  - Google will give users in the “A” group one version of the results, and users in the “B” group get a different version
  - If users in the “A” group click on more results than “B”, then next time Google will use the “A” results

Coverage: Filtering

Why do people get different results?

Another reason: **A/B testing**

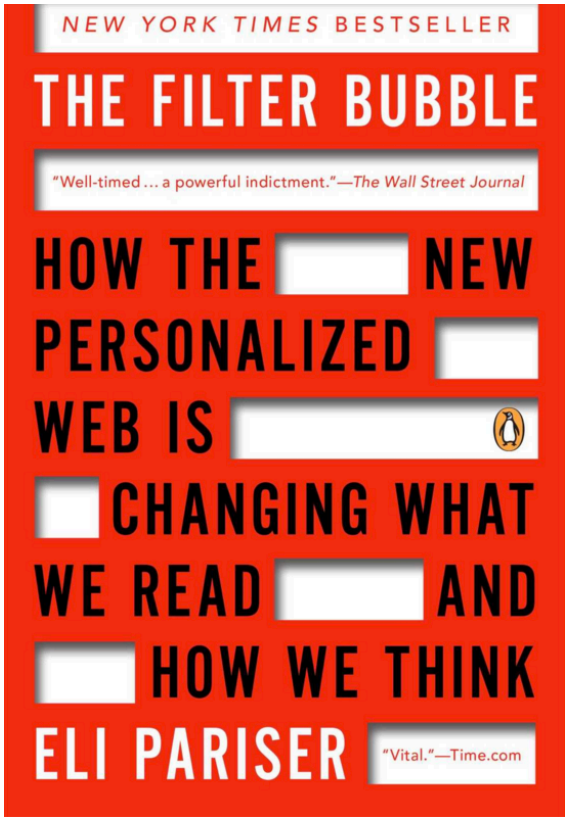
- Google gives different users different results as a way of experimenting with their system
  - Google will give users in the "A" group one version of the results, and users in the "B" group get a different version
  - If users in the "A" group click on more results than "B", then next time Google will use the "A" results

Companies like Google and Facebook are constantly experimenting with their systems

Coverage: Filtering

What is the consequence of this kind of filtering?

Some people worry about a **“filter bubble”:** people will only access information that conforms to their existing views

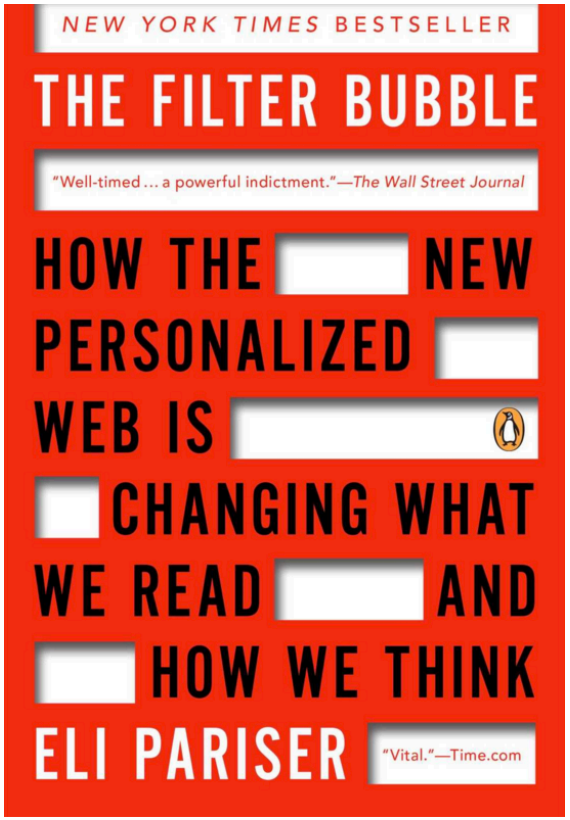


Coverage: Filtering

What is the consequence of this kind of filtering?

Some people worry about a **"filter bubble":** people will only access information that conforms to their existing views

- Newer systems address this by combining personalization with common information
- And newer research has shown that people don't self-select as much as some have speculated



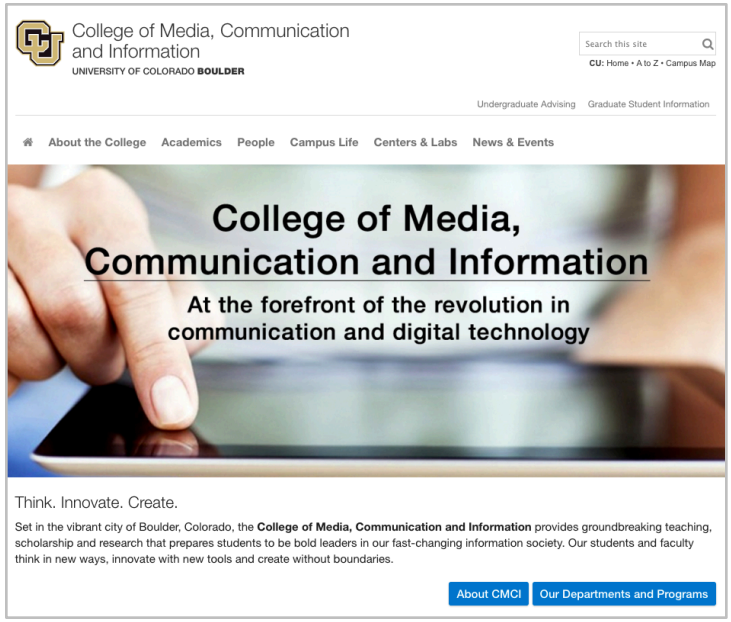
Coverage: Device Dependence

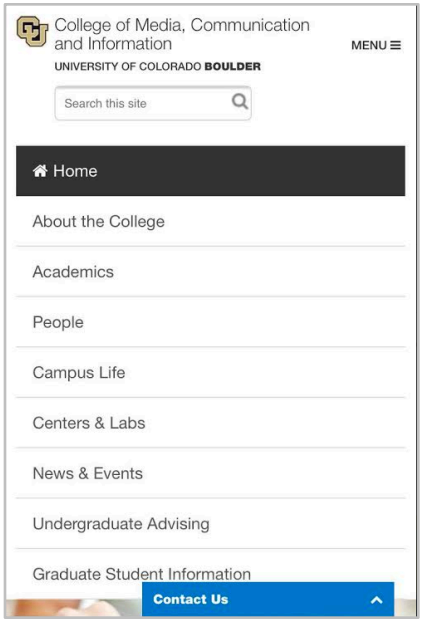
Another challenge:
The information you view can even change depending on the device you use!

Coverage: Device Dependence

Mobile websites usually look different than desktop websites

- Sometimes mobile websites even exclude content that's available in the full version

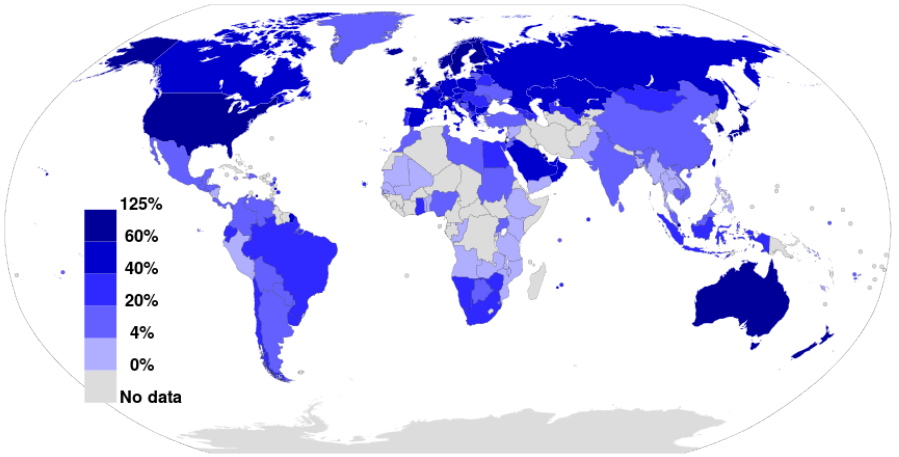




Coverage: Device Dependence

Mobile websites usually look different than desktop websites

- Sometimes mobile websites even exclude content that's available in the full version
- This matters because some people only access the web from their phone
  - especially in some countries



Coverage: Device Dependence

Emoji appear differently on different phones

Apple

Android

Source: https://medium.com/matter/lost-in-emoji-translation-apple-vs-android-648fdd57ca25

Coverage

Is the information complete?

Conclusion: the information you see may be different from the information I see

- sometimes in ways that you don’t expect!

Summary

- Evaluate information holistically
  - To understand the content, you also need to **understand the context:** who created/published the information, and when?
- Read **numbers as critically** as you read words
- Notice where your information comes from
  - Has an **algorithm filtered the information** for you? How might this affect your perspective?