LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
FAKULTÄT FÜR MATHEMATIK, INFORMATIK UND STATISTIK
INSTITUT FÜR INFORMATIK
LEHRSTUHL FÜR DATENBANKSYSTEME UND DATA MINING

Lecture Notes to

Big Data Management and Analytics

Winter Term 2018/2019

The Flip Side of the Coin

© Matthias Schubert, Matthias Renz, Felix Borutta, Evgeniy Faerman, Christian Frey, Klaus Arthur Schmid, Daniyal Kazempour, Julian Busch

DBS

A general problem of the information age

- social and economic systems are often based on a certain level of information being available to certain roles
- getting information is connected to a certain effort which is assumed to grow linearly for additional cases
- machines are considered to behave similar to humans
- If these preconditions get invalidated by technical advancements, existing social and economic systems can be
    - exploited
    - or even threatened in their existence

Target Advertising and Recommender Systems

- Collect data on personal behavior:
  - Click streams on the internet (Where do you go online?)
  - Smartphone trajectories (Where do you go in real life?)
  - Search history on Search Engines (What are you interested in?
  - Social Networks (Who do you know?)
- Big paradigm break in advertisement (measure the success of your marketing efforts)
- Tasks:
  - Select the best suitable add for a given customer
  - Offer a customer interesting products
- Techniques: Collaborative Filtering, Association Rules, etc.

The value of personal data

Personal data allows to:

- know costumer interests
- schedule the next time a product is needed
- analyze the interest of family and friends

$\Rightarrow$ useful for selling you the goods you need

Costumer profiles are based on :

- movement information (handy OS, telco provider, car,..)
- text messages (email, sms, what's app,..)
- census data (age, sex, address, ..)
- social relations (friends, colleagues, family...)
- surf profiles...

What can be learned from profiles?

- Where do you live? Where do you work? (trajectories)
- Do you have one night stands ? (trajectories)
- What maladies do you have? (browser history)
- What political preferences do you have? (twitter, facebook, browser history)
- Did you commit speeding? (trajectories)
- Sexual preferences? (trajectories, facebook,..)
- Which medication did you buy recently? (sales data,..)
- What does your credit history look like? (agencies)

What can you do with all this information?

- personal information is used for identification
- identity theft and account kidnapping
  - personal questions
  - passwords with personal reference
- personalize spam and fishing
  - messages from family and friend are considered trust worthy
- check for insurances or employment
  - known health issues or life style
- check for unlawful, criminal or politically unwanted actions
  - is somebody close to the opposition
  - did speeding or public drinking occur

Beyond Knowing towards Manipulating

- knowing personal info is useful, but
  - if a customer does want to buy something
  - if people do not want to behave like I want them to it is useless. Is It?

- Do people always have a strong opinion about everything?

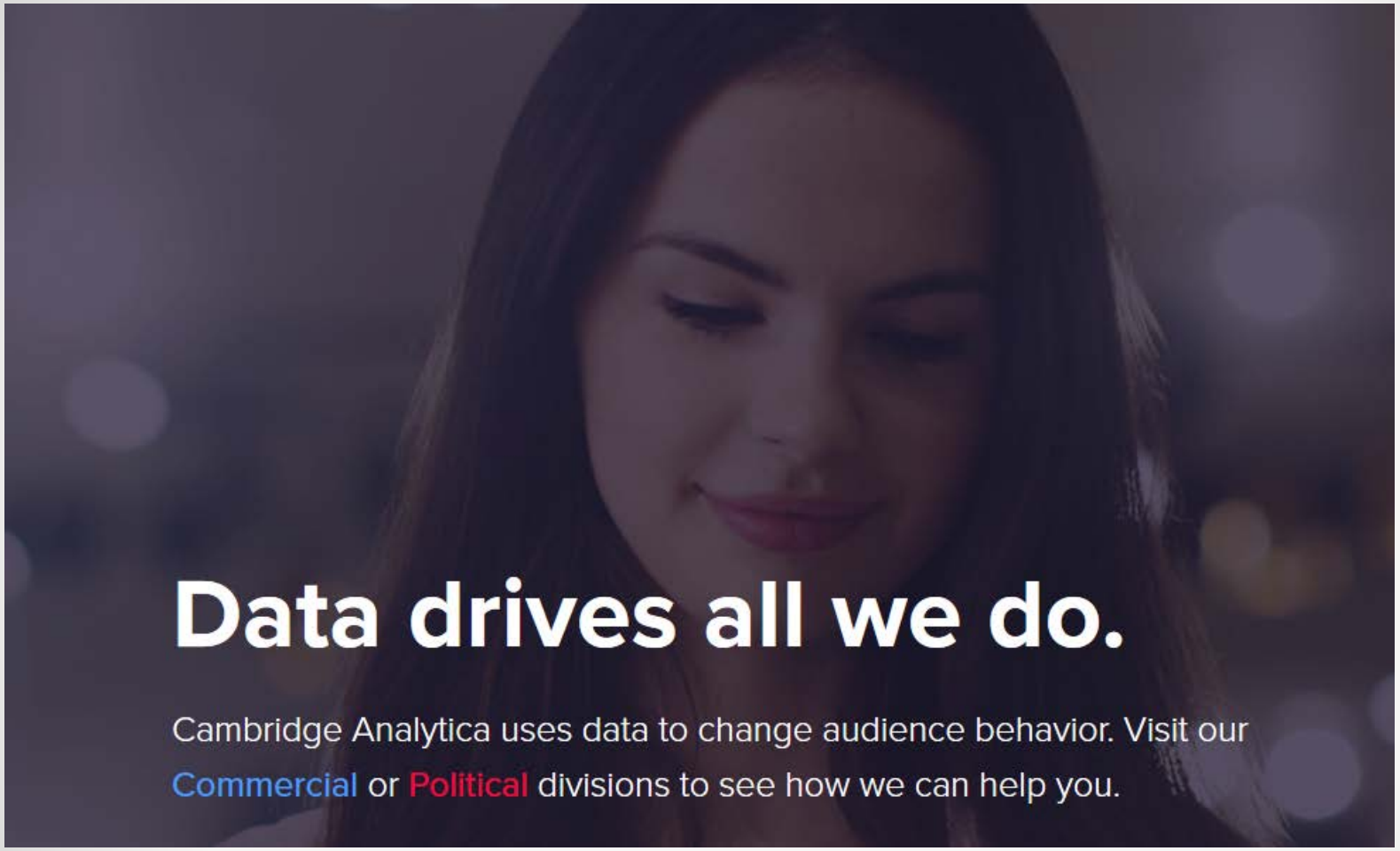
- nobody can be an expert for everything

  => rely on the expertise of others
- Who do you trust?
  - people believe want makes sense to them
  - people like to believe positive news
  - people prefer simple answers

Fake news or not?

https://ig.ft.com/sites/quiz/fake-news-or-not/

cambridgeanalytica.org



method is based on the research of psychologist Michal Kosinski about psychologic profiling based on social media

Steps in Micro Targeting

Micro targeting: *Address thousands of social media users by using data analytics to generate individual psychological profiles and then confront them with selected content to trigger a certain reaction.*

naïve example:

- collect facebook data
- profile users w.r.t. the big five traits (c.f. next slide)
- find people with a high level of **neuroticism**
- feed them with (disconnected) cases of criminal refugees
- show them fake statistics of strongly increased criminality caused by refugees
- => increased likelihood to support right-wing populists

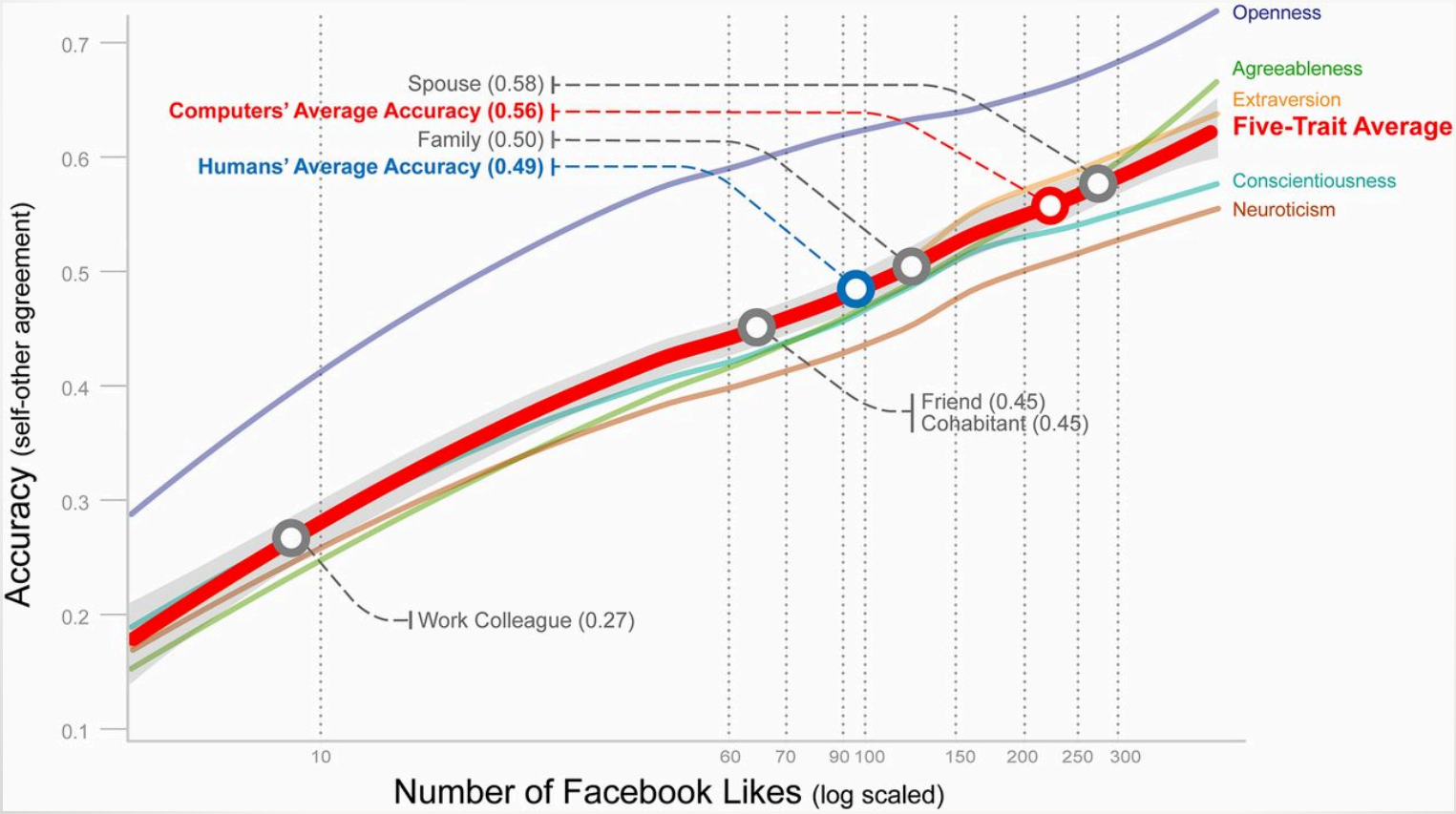
Psychologic Profiles

example profile: the big 5 traits

- **Openness to experience**: Appreciation for art, emotion, adventure, unusual ideas, curiosity, and variety of experience.
- **Conscientiousness**: A tendency to be organized and dependable, show self-discipline, act dutifully, aim for achievement, and prefer planned rather than spontaneous behavior.
- **Extraversion**: Energy, positive emotions, surgency, assertiveness, sociability and the tendency to seek stimulation in the company of others, and talkativeness.
- **Agreeableness**: A tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others.
- **Neuroticism**: The tendency to experience unpleasant emotions easily, such as anger, anxiety, depression, and vulnerability.

Computer made Profiles

Wu Youyou, Michal Kosinski, and David Stillwell:
Computer-based personality judgments are more accurate than those made by humans, Proc Natl Acad Sci U S A. 2015 Jan 27; 112(4): 1036–1040. Published online 2015 Jan 12. doi: 10.1073/pnas.1418680112



How to influence people given the profiles?

Hirsh JB, Kang SK, Bodenhausen GV: **Personalized persuasion: tailoring persuasive appeals to recipients' personality traits.** Psychol Sci. 2012 Jun;23(6):578-81. doi: 10.1177/0956797611436349. Epub 2012 Apr 30.

key outcomes:

- personal traits show large value to tailor persuasive messages
- in the described experiment the tailored advertisement showed higher success
- does not work every time, but often enough

Example: Spam Email, I recently received

To the

German Presserat
Fritschestraße 27-28

10585 Berlin, Germany

Dear Sir or Madam,

Anja Stahmann is planning an excessive reduction of staff positions in nursing homes, which is expected to result in a zero supply and euthanasia. Because a single nurse is unable to provide fifty patients or to supervise, to save and so on.

When Ms Stahmann enters the canary's rights, others will follow.

The new and presumably unconstitutional law, in which the personnel key planned by Mrs Stahmann is to be committed, is to be adopted in September. Already at the beginning of August the lecturer of the university of applied sciences for Diakonie in Bielefeld Christopher Kesting had posted the petition below, which turns against the planned new law. Because of the holiday season, the petition was largely ignored for weeks, which is why I would like to inform you about the topic.

....

FAKE NEWS

What was the connection to Data Science?

- social and economic systems are build on the assumption that manipulating a single person is coupled with a large effort
- social networks and personal profiling allow for scalable personal profiling
- scientists start to develop strategies to influence behavior on large scale via electronic media

How bad is it ?

- assume every 20th person can be influenced at an election into:
  - voting a populist party
  - not participating

$\Rightarrow$ 5% of the votes might be different

$\Rightarrow$ might have played a role in Brexit and the 2016 US election

Ok, who cares ?

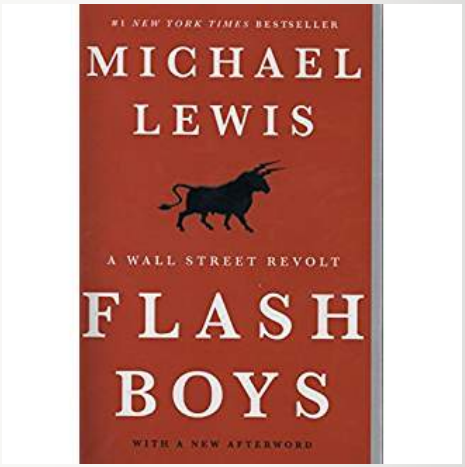
- climate change triggered an emotional public discussion led by none-experts:
  destroying our children's future vs.
  an egg-head tale to boss around people and destroy honest jobs

- investment and participation into future technologies depend on the trust of investors and consumers

- manipulating the people based on micro targeting is a method which can be employed regardless of facts and intentions

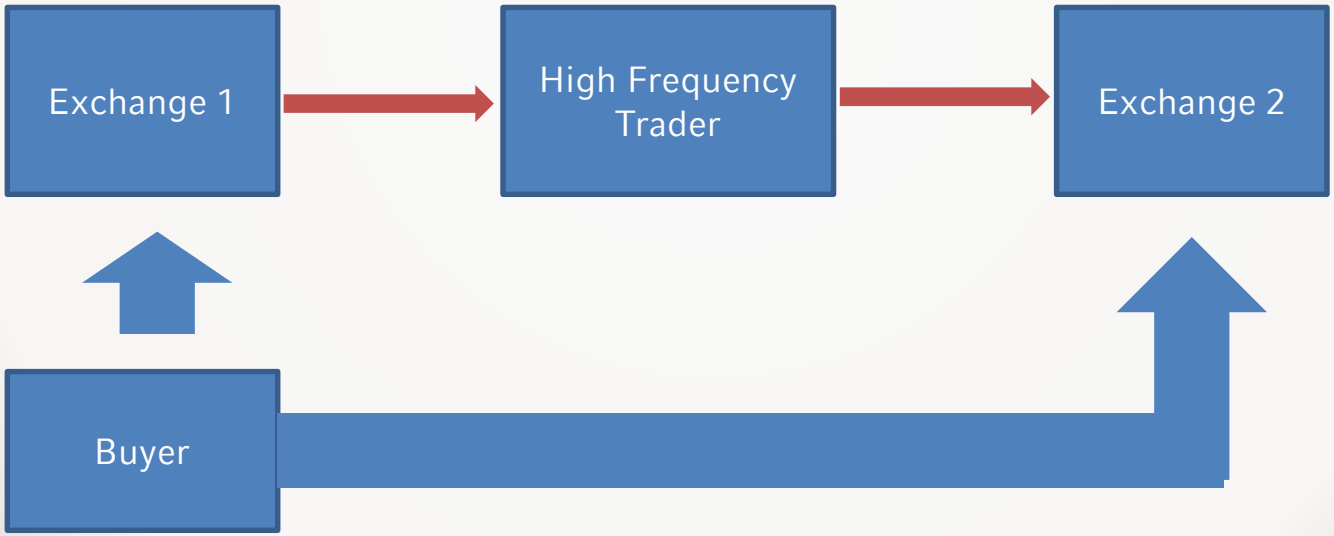
High Frequency Trading

- The stock market is built on the assumption that all offers are visible to all traders simultaneously. Sales and buy offers meet at the middle price on average.
- in 2007: Wallstreet switched to electronic trading systems
  => matching buy and sales offers is done by algorithms
- electronic trading is done on several stock exchanges throughout the larger New York area and New Jersey
- for human traders all offers appear simultaneously on all stock markets due to high speed .
  (there is a latency of ca. 20 ms)
- even though the trade should just get more volatile, prices began to run away (buy orders would often hit the maximum price instead the middle price)



High Frequency Trading

- What happened?
- Pre-running: High Frequency Trader sees offer at exchange 1 buys available stock at exchange 2 and sells the stocks to the buyer at a larger price. (just one example)
- What has this to with Data Analytics?
  - the HFT has to predict the larger order from monitoring Exchange1



High Frequency Trading

What is bad about this?

- draws capital from corporate investment
- investment of the stock market gets more expensive
- milks money from pension fonts
- no macro economic advantage to the financial system
- destabilizing electronic markets by techniques like spoofing or layering which try to provoke price chances on the market

example: Flash Crash of May 6, 2010

- United States trillion-dollar stock market crash
- started at 2:32 p.m. EDT and lasted for approximately 36 minutes
- S&P 500, Dow Jones Industrial Average and Nasdaq Composite, collapsed and rebounded very rapidly
- multiple theories involving HFT trading algorithms
- conviction of one person which used an automated program to generate large sell orders, pushing down prices, which were then cancelled to buy at the lower market prices

High Frequency Trading

Solutions:

- require an exchange protocol for making offers apparent simultaneously to all users (regulations required)

- in 2013 IEX opened: Exchange that adds additional delay to negate advantages of HFT companies and only allows 3 order types to prevent various arbitrage effects.

- system designers and system users should not be the same

Why is interesting to us?

- Decision makers, regulators and legislations did not understand what they were actually doing and where there decision would lead

- Expertise on CS should have been involved in the decision process not just in the development step

- Technically there would have been solutions which would have been easy to install before starting with electronic trading

Message: CS people have a responsibility to society to involve themselves into these kinds of political decisions.

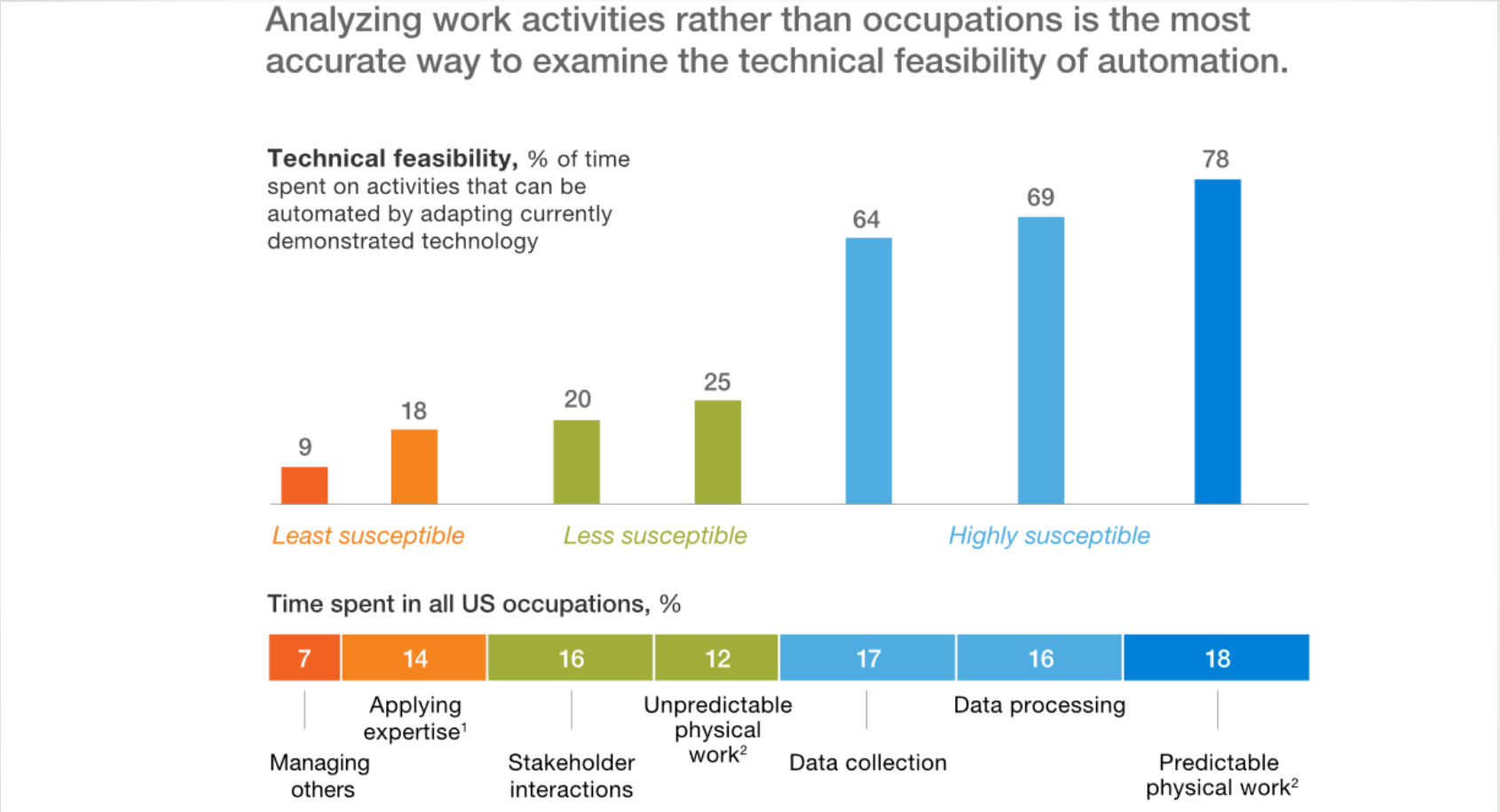
Robotics, AI and the Work Force

Where machines could replace humans
—and where they can't (yet)

Michael Chui, James Manyika, and Mehdi Miremadi, McKinsey quarterly
http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet

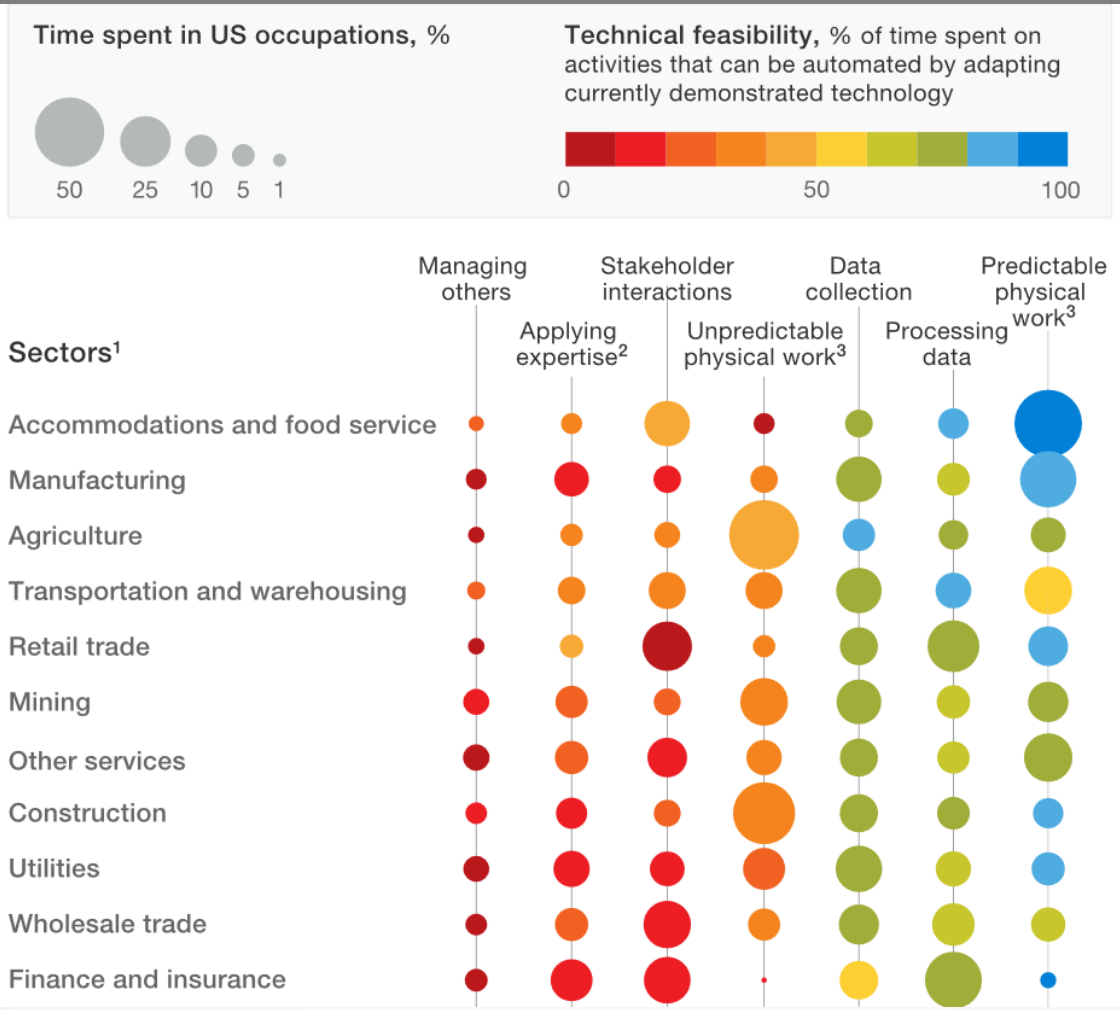
- splits any job into activities
- examine activity whether automation is techn. feasible
- also considers: costs, regulations and social acceptance
- potential for predictable physical work, data collection and processing is the largest (64% -78 %)
- potential for unpredictable physical work still 25 %

Technical Feasibility of Activities



http://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Where%20machines%20could%20replace%20humans%20and%20where%20they%20cant/SVGZ-Sector-Automation-ex1.ashx

How does this aggregate over job sectors?



http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/where-machines-could-replace-humans-and-where-they-cant-yet

What is the connection to Computer Science

- currently we are fine off because all these robots and computers have to be programmed
  => Does this remain a valid statement for next 50 y?
- for now education more people in CS will be mandatory to make people fit for the information age
- the impact on the society may cause rejection to novel development cooling down the speed of the automatization
- current social systems have to be reevaluated
- societies might split due the gap between people profiting from digitization and those of the loosing side

Conclusions

- advance in CS is often been made to “make the wolrd a better place”
- during the life cycle of multiple useful services fraud and misuse limit the advantages
- a lot of traditional imaginations on social, political and economical systems have to be reevaluated due digitization and the information age
- CS people have the responsibility to involve themselves more into political and regulatory processes because other people lack the expertise