Big Data Science

Course: 18-788

Patrick McSharry
patrick@mcsharry.net
www.mcsharry.net
Twitter: @patrickmcsharry

Spring 2023

ICT Center of Excellence
Carnegie Mellon University

Big Data Science

WEEK 3A

Course outline

Week	Lecture A	Lecture B
1	Weather & agriculture	Climate change
2	Climate scenarios	Catastrophe models
3	Social trends	Finance
4	Sentiment analysis	Health
5	Telemedicine	Mobile data
6	Data4Dev	Socioeconomic status

Today's Lecture

No.	Activity	Description	Time
1	Challenge	Surveys versus big data	10
2	Discussion	Google trends	10
3	Case study	Flu forecasting	10
4	Analysis	Unemployment	20
5	Demo	CPI portal	20
6	Q&A	Questions and feedback	10

Survey

- According to the Oxford English Dictionary, a survey is:
- 1. Look closely at or examine (someone or something)
- 2. Examine and record the area and features of (an area of land) so as to construct a map, plan, or description
- 3. Investigate the opinions or experience of (a group of people) by asking them questions

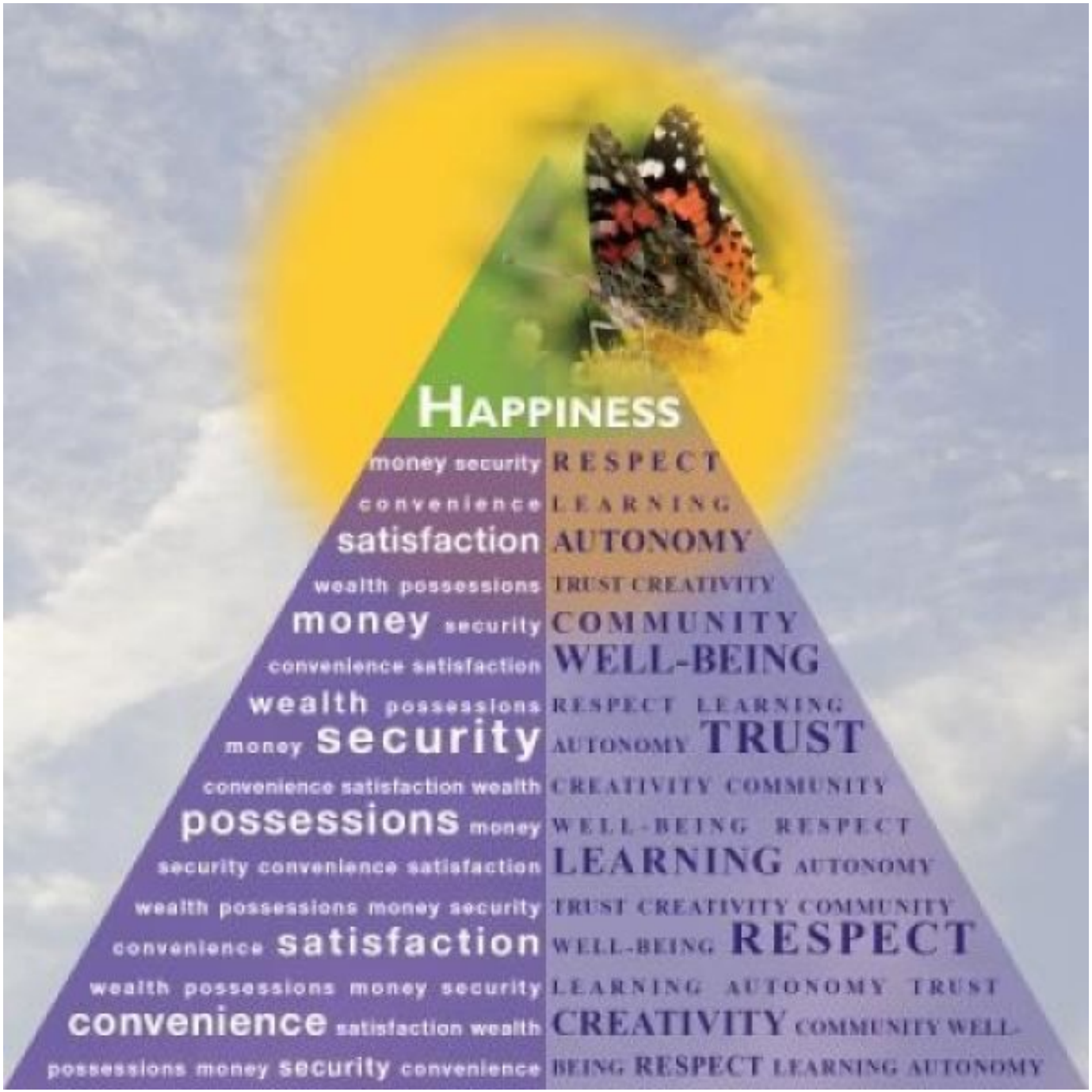
Problems with surveys

- Is there an incentive to take the survey and might this bias the result?
- Respondents may be required to pick particular answers to specific questions
- Respondents may answer what they think the surveyor wants to hear
- Respondents limited to a certain group and not representative of the whole population
- Sample size issues

Happiness Poll

- What would you measure in order to quantify happiness?

- **Slido.com #836 812**


HAPPINESS
money security
convenience
satisfaction
wealth possessions
money security
convenience satisfaction
wealth possessions
money security
convenience satisfaction wealth
possessions money
security convenience satisfaction
wealth possessions money security
convenience satisfaction
wealth possessions money security
convenience satisfaction wealth
possessions money security convenience
RESPECT
LEARNING
AUTONOMY
TRUST CREATIVITY
COMMUNITY
WELL-BEING
RESPECT LEARNING
AUTONOMY TRUST
CREATIVITY COMMUNITY
WELL-BEING RESPECT
LEARNING AUTONOMY
TRUST CREATIVITY COMMUNITY
WELL-BEING RESPECT
LEARNING AUTONOMY TRUST
CREATIVITY COMMUNITY WELL-
BEING RESPECT LEARNING AUTONOMY

Gross National Happiness (GNH)

- Gross National Happiness (GNH) is a phrase coined in 1972 by Bhutan's fourth Dragon King, Jigme Singye Wangchuck.
- It represents a commitment to building an economy that would serve Bhutan's culture based on Buddhist spiritual values instead of western material development gauged by gross domestic product (GDP).
- Bhutan recently ranked eighth out of 178 countries in Subjective Well-Being - the only country in the top 20 "happiest" countries that has a very low GDP.

UN and GNH

- In July 2011, the United Nations passed Resolution 65/309, that was adopted unanimously by the General Assembly in July 2011, placing "happiness" on the global development agenda.
- The four pillars of the GNH philosophy are:
  - Sustainable development
  - Preservation and promotion of cultural values
  - Conservation of the natural environment, and
  - Establishment of good governance.

Happiness indices

Year	Country	Description
1972	Bhutan	Bhutan's King introduced the Gross National Happiness (GNH)
2007	Thailand	Green and Happiness Index (GHI)
2009	US	Gallup Well-Being Index
2010	UNDP	Oxford University launch multidimensional poverty index (MPI)
2011	OECD	OECD launched "Better Life Index" (BLI)
2011	Canada	The Canadian Index of Wellbeing (CIW)
2011	South Korea	Happiness Index
2014	UK	Launches well-being and happiness statistics
2016	Thailand	Launch of GNH imitative
2016	Dubai	Ministry of Happiness motivated by GNH

The first World Happiness Report was released on April 1, 2012 as a foundational text for the UN *High Level Meeting: Well-being and Happiness: Defining a New Economic Paradigm.*

Measuring happiness

- Daniel Kahneman: happiness can be measured using the day reconstruction method, which consists in recollecting memories of the previous working day by writing a short diary.
- Adam Kramer: a behavioral model of "Gross National Happiness" based on the use of positive and negative words in social network status updates, resulting in a quantitative GNH metric.
- Adrian G. White (2007): "A Global Projection of Subjective Well-being: A Challenge to Positive Psychology?”
- Participants in the various studies were asked questions related to happiness and satisfaction with life. The meta-analysis is based on the findings of over 100 different studies around the world, which questioned 80,000 people worldwide.

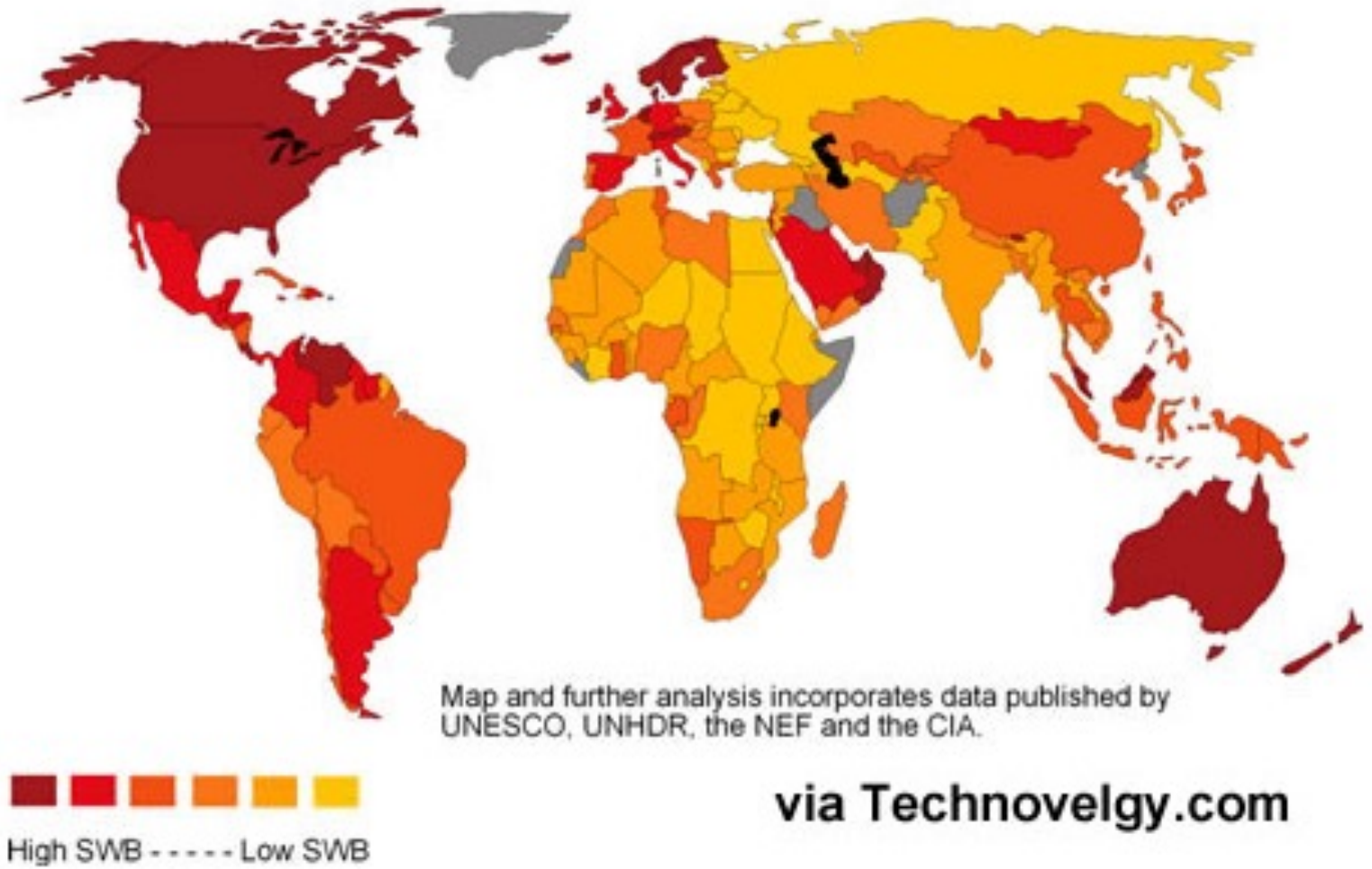
Happiest Nations in the World

Rank	Country	Rank	Country
1	Denmark	11	Ireland
2	Switzerland	12	Luxembourg
3	Austria	13	Costa Rica
4	Iceland	14	Malta
5	The Bahamas	15	The Netherlands
6	Finland	16	Antigua and Barbuda
7	Sweden	17	Malaysia
8	Bhutan	18	New Zealand
9	Brunei	19	Norway
10	Canada	20	The Seychelles

23 – USA, 35 – Germany, 41 – UK, 62 – France, 82 – China, 90 – Japan, 125 – India, 167 – Russia
176 – Democratic Republic of the Congo, 177 – Zimbabwe, 178 – Burundi

Source: Adrian White, World Map of Happiness

World Happiness Map

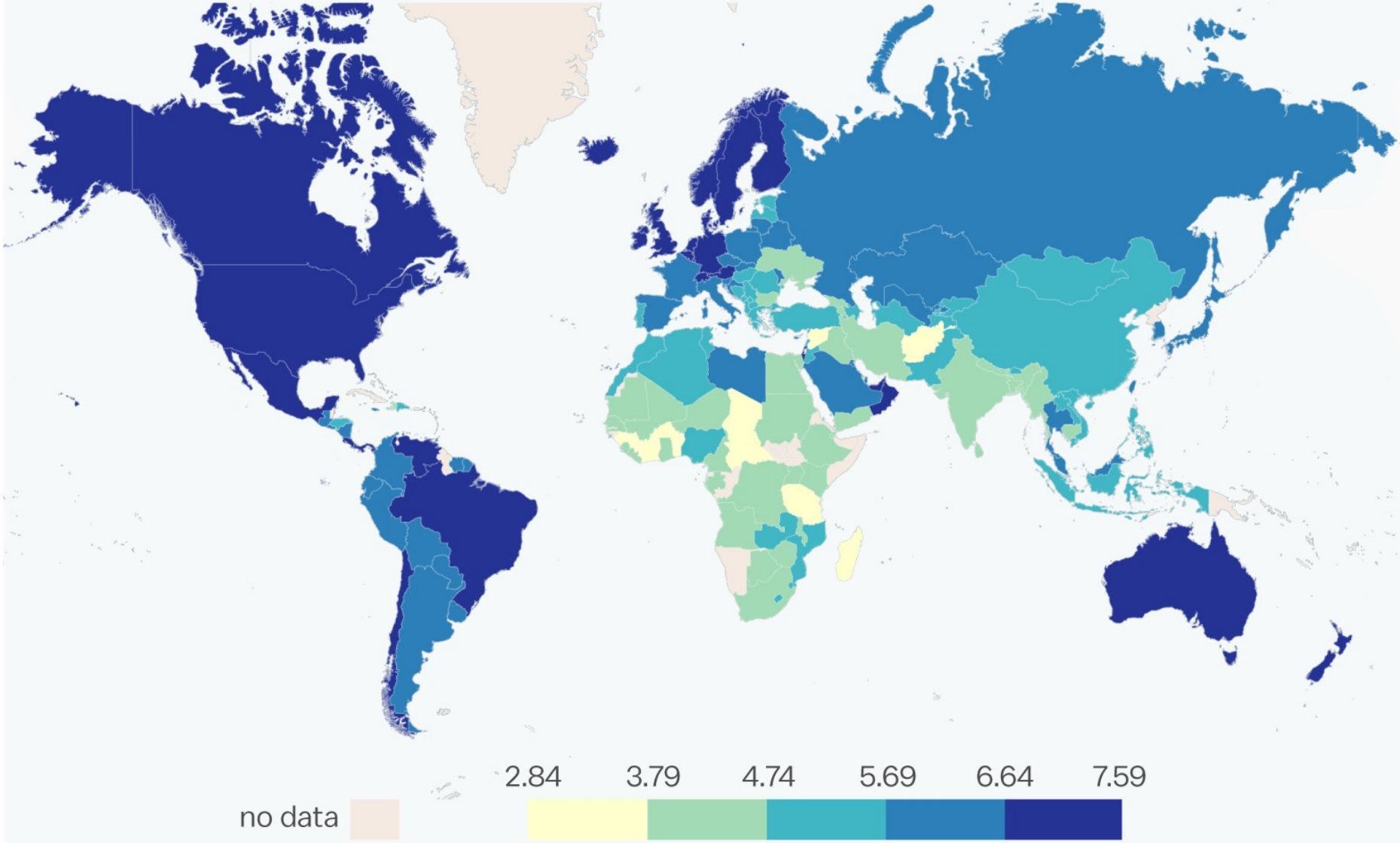


Happiness is ...being Healthy, Wealthy and Wise

- A nation's level of happiness was most closely associated with health levels (correlation of .62), followed by wealth (.52), and then provision of education (.51).
- Many of the largest countries in terms of population do quite badly.
- A recent BBC survey found that 81% of the UK population think the Government should focus on making us happier rather than wealthier.

Happiness level by country

Self-reported, with 1 as the "worst possible life for you," and 10 the best

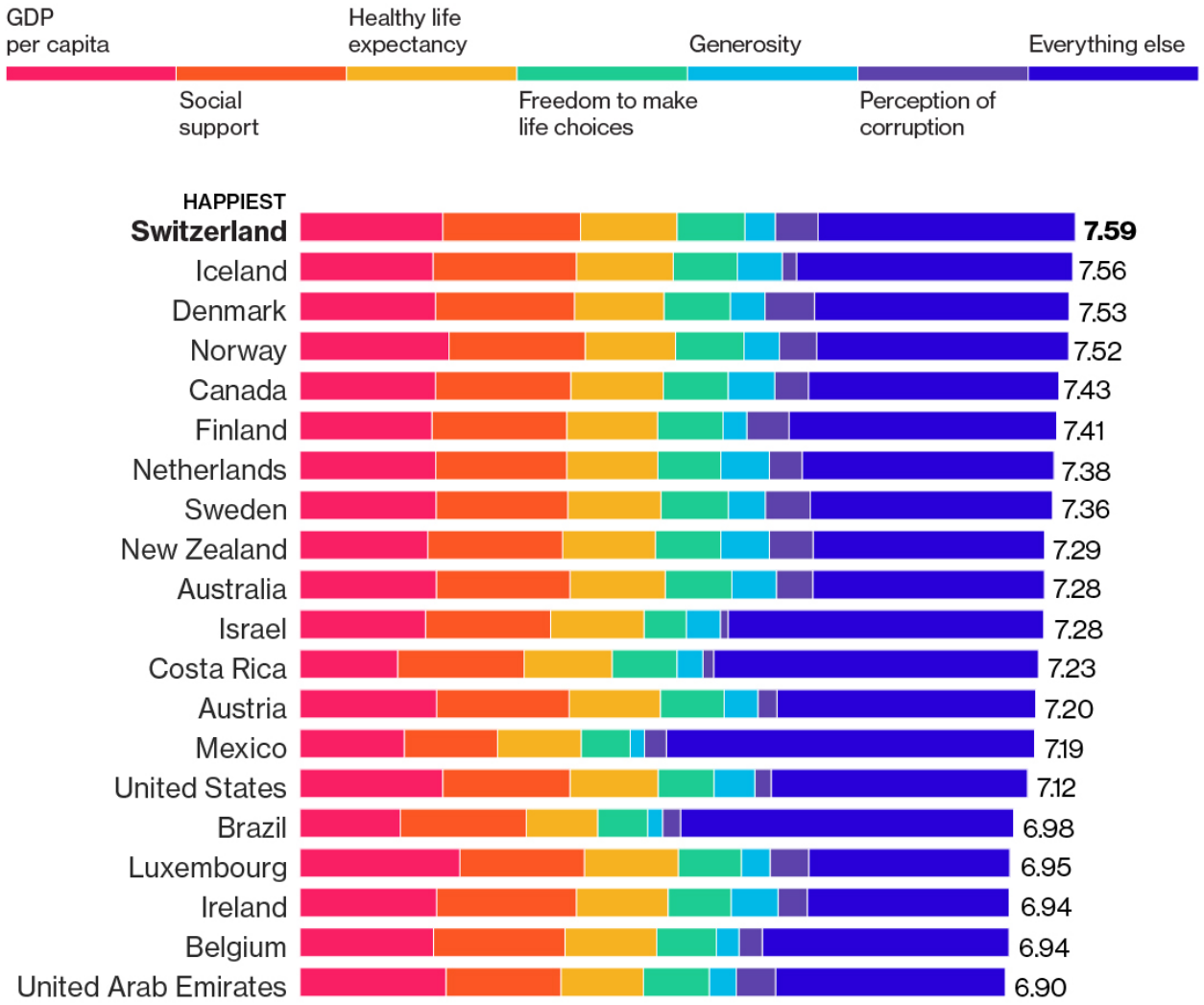


SOURCE: Gallup World Poll 2012-2014 via World Happiness Report 2015. Made with CartoDB.

Vox

20 Happiest Countries in the World

Residents of each country were asked to score their quality of life on a scale of 0 to 10.



Note: Researchers estimated how much each factor contributed to people's happiness.
Source: World Happiness Report 2015

GNH problems

- Research demonstrates that markers of social and individual well-being are remarkably transcultural.
- People generally report greater subjective life satisfaction if they have strong and frequent social ties, live in healthy ecosystems, experience good governance, etc.
- Reliance on national measures of GNH would render international comparisons of relative well-being more problematic, since there is not and is not likely ever to be a common scale as "portable" as GDP.
- Alternative indicators of emotion as an analog to economic progress are being used in the UK, Europe and Canada.

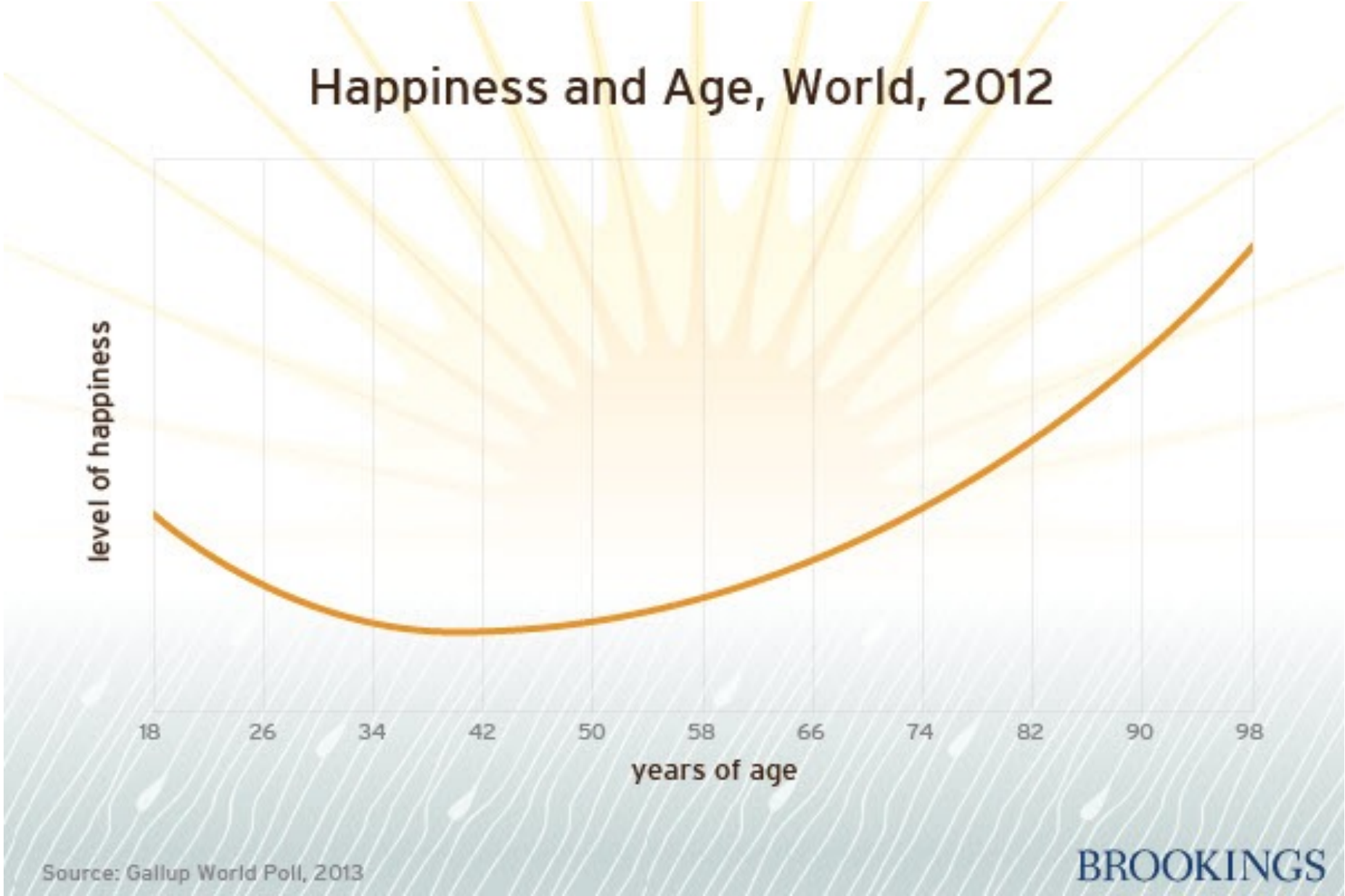
Gallup Daily Mood: Eight Happiest Days in 2009

(In order by % happiness and enjoyment without stress and worry)

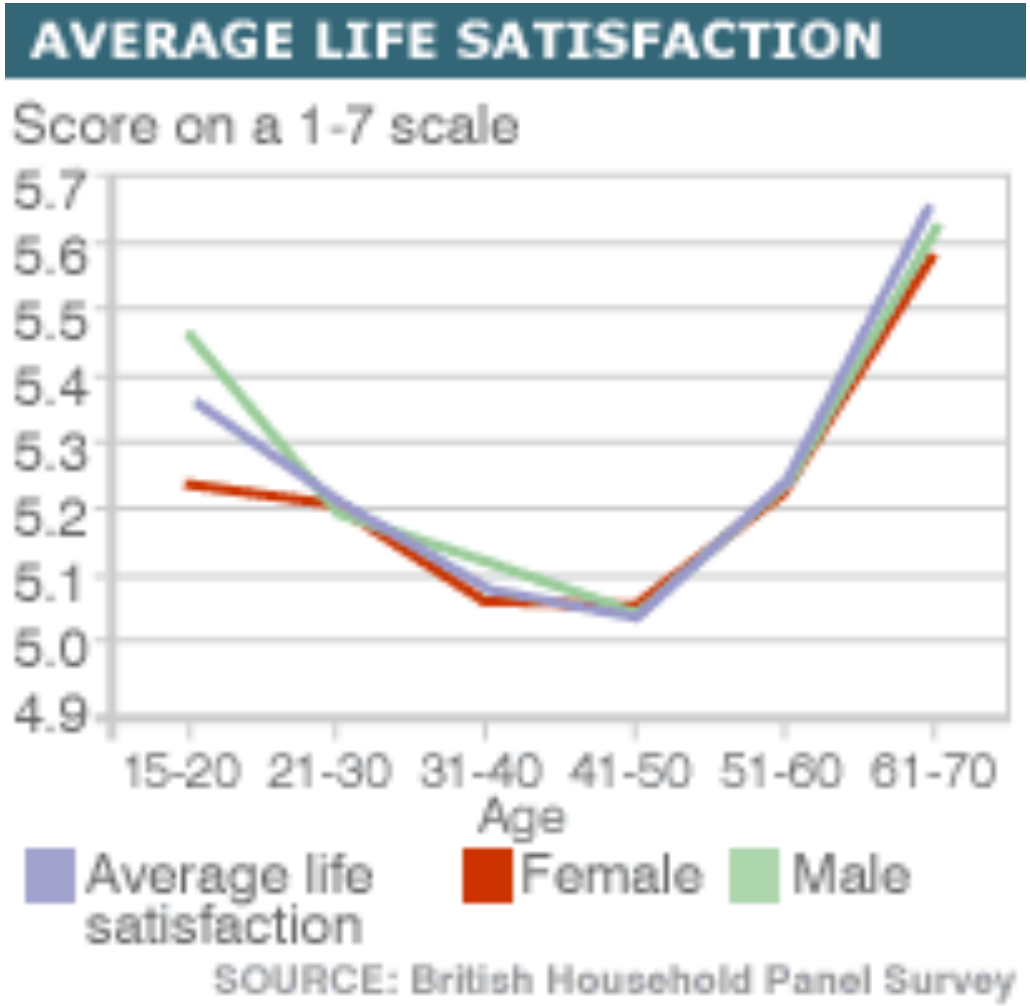
Date	% Happiness and enjoyment without stress and worry	% Stress and worry without happiness and enjoyment	Day of the year
Apr 12, 2009	64	6	Easter
Jun 14, 2009	63	8	Flag Day
May 10, 2009	62	7	Mother's Day
Nov 26, 2009	62	5	Thanksgiving
Jul 4, 2009	61	7	Independence Day
Jan 1, 2009	61	7	New Year's Day
Jun 21, 2009	61	9	Father's Day
Oct 11, 2009	61	9	Sunday (day before Columbus Day)

Gallup-Healthways Well-Being Index

GALLUP

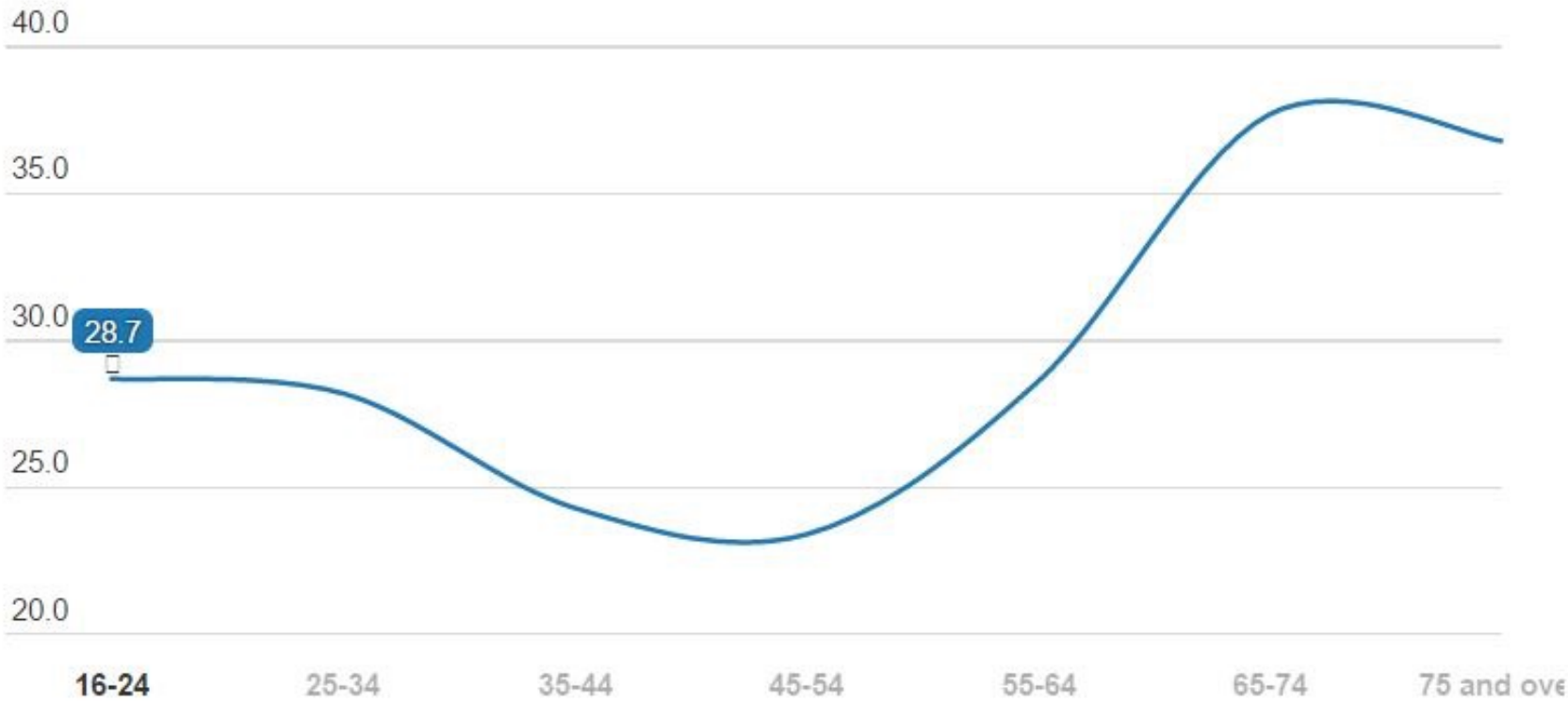

Happiness and Age, World, 2012
level of happiness
18
26
34
42
50
58
66
74
82
90
98
years of age
Source: Gallup World Poll, 2013
BROOKINGS

U-shaped happiness



The happiness U-shape

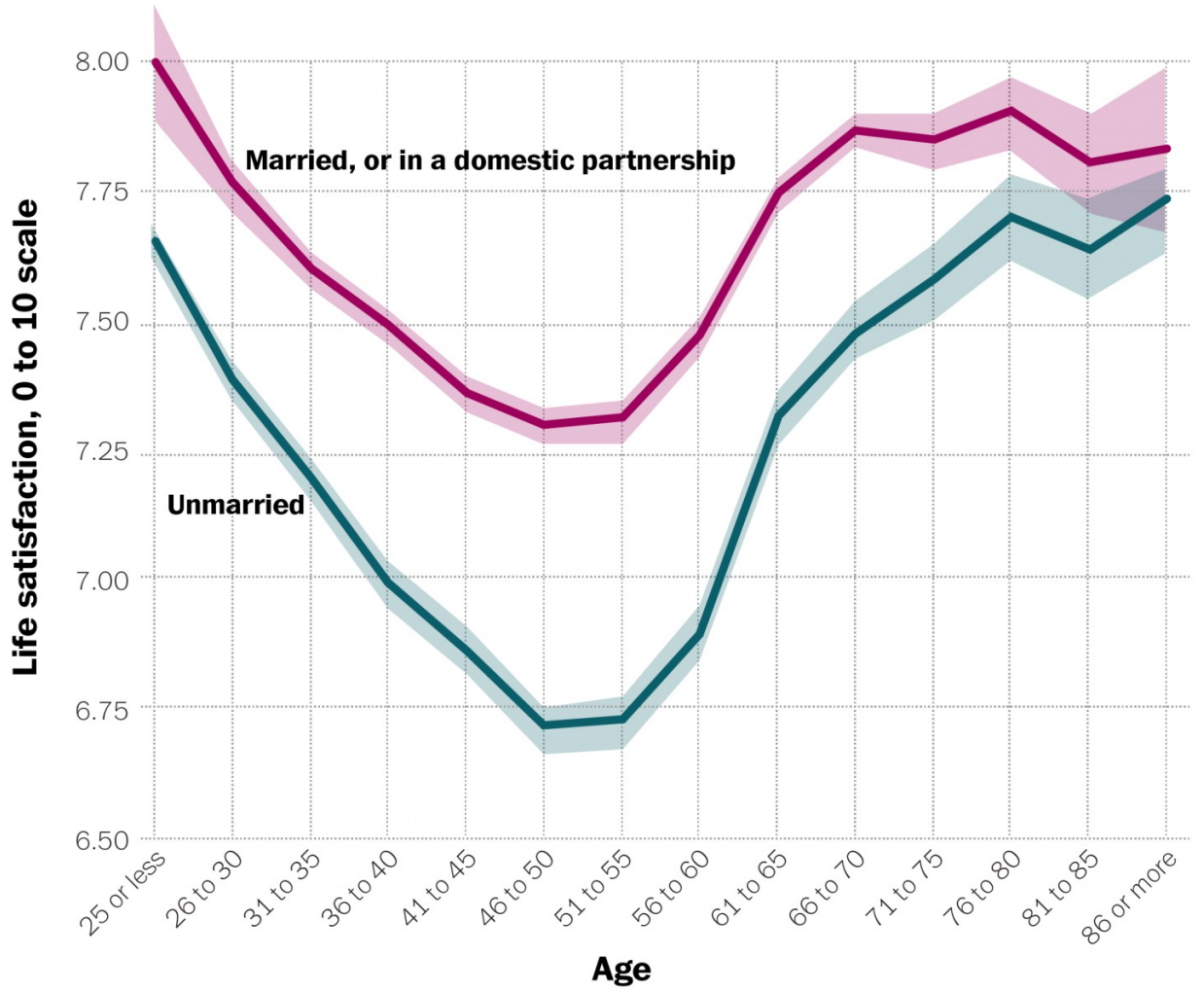
Overall life satisfaction in the UK: self-reports from four surveys between April 2014 and March 2015.



Source: UK Office of National Statistics Get the data

Married people are happier – especially in middle age

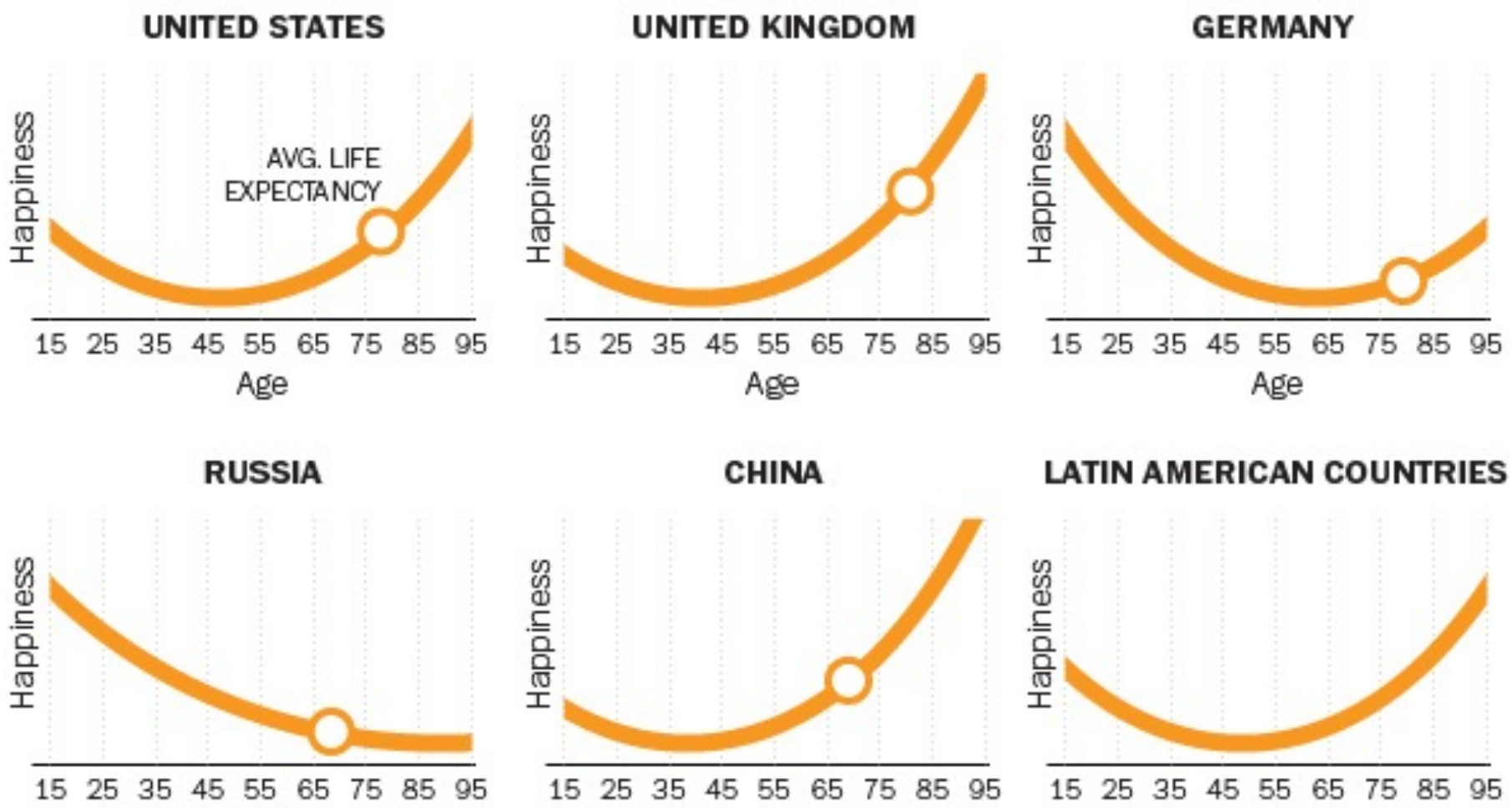
Self-reported life satisfaction among married and unmarried people in the U.K., by age. Shaded areas represent confidence intervals.



WASHINGTONPOST.COM/**WONKBLOG**

Source: Grover and Helliwell

Happiness by country



Surveys versus big data

- Sources of big data such as social media provide an alternative means of accessing information
- Google trends tells us about what people are searching for
- Twitter tells us about what people are thinking and doing
- Loyalty schemes provide financial transactions to tell us about how people spend money

Google trends

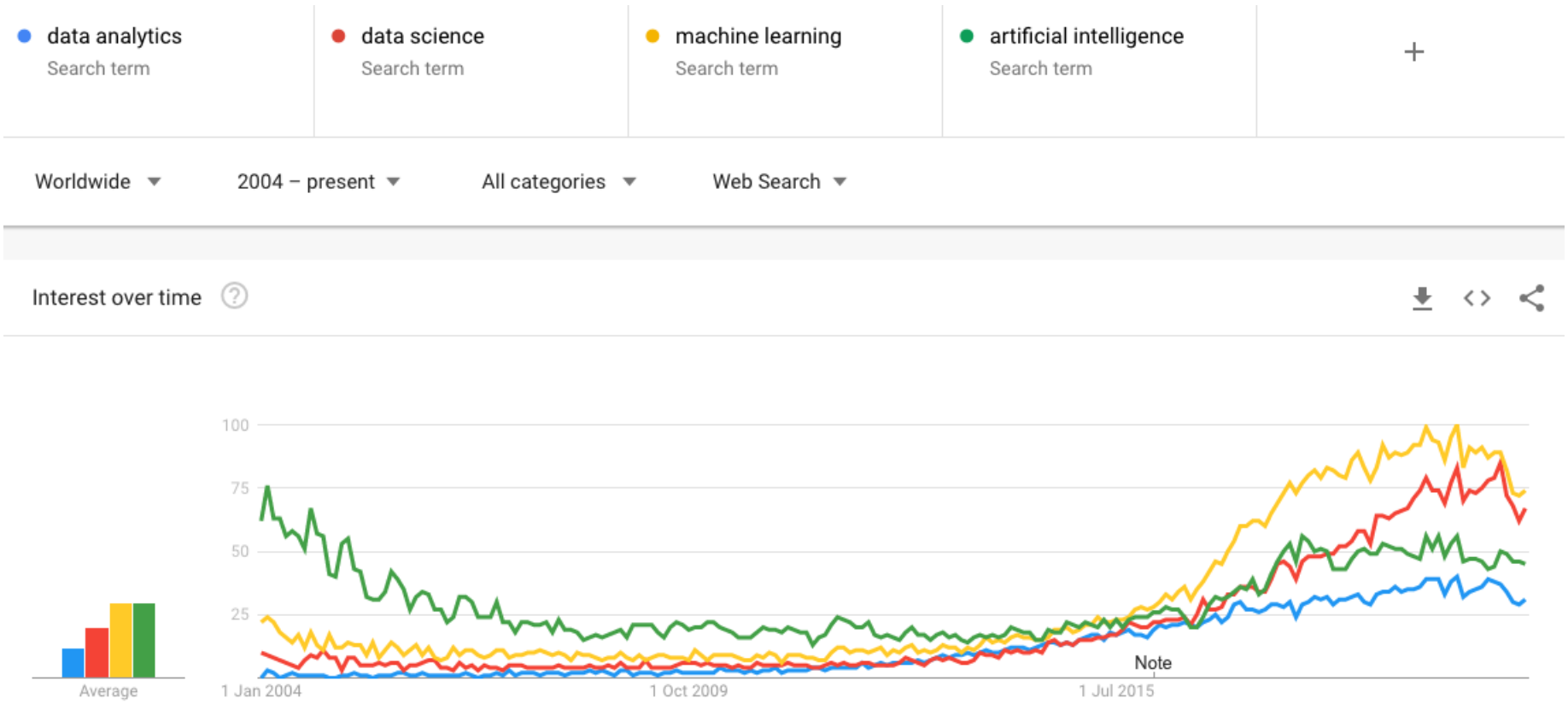
- Google Trends is a public web facility of Google Inc., based on Google Search, that shows how often a particular search-term is entered relative to the total search-volume across various regions of the world, and in various languages.
- The data is reported weekly since 2004.
- Varian & Choi (2009) “Predicting the Present with Google Trends” showed that retail, auto, home and travel sales are now-castable with search data.
- Hot Trends is updated hourly and displays the top 20 hot, i.e., fastest rising, searches (search-terms) of the past hour in various countries.

Source: https://www.**google**.com/**trends**/

Specialization Poll

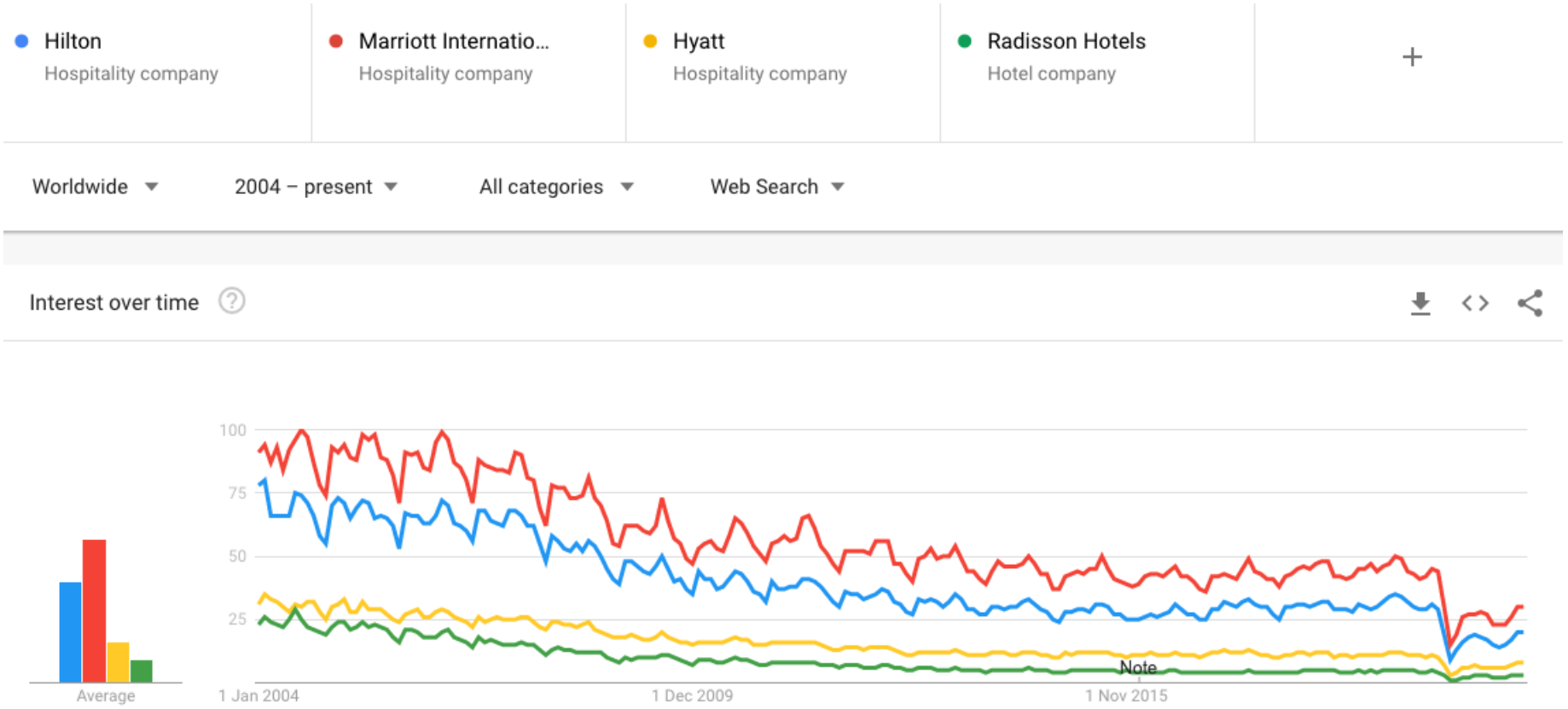
- Which of the following do you think is most popular?
  a) Data analytics
  b) Data science
  c) Machine learning
  d) Artificial intelligence

Measuring Popularity: Google Trends

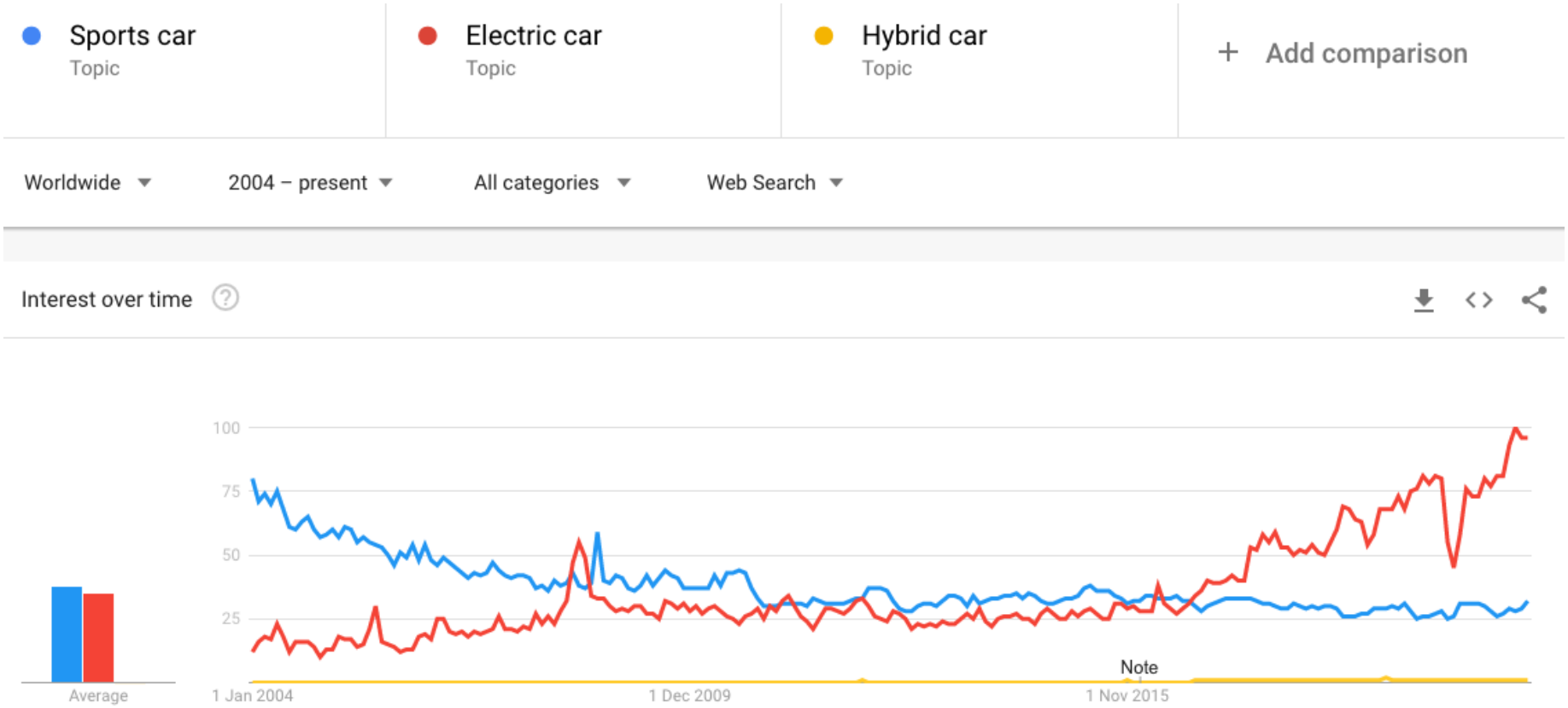


Source: Google Trends, Worldwide, 2014 – 2021

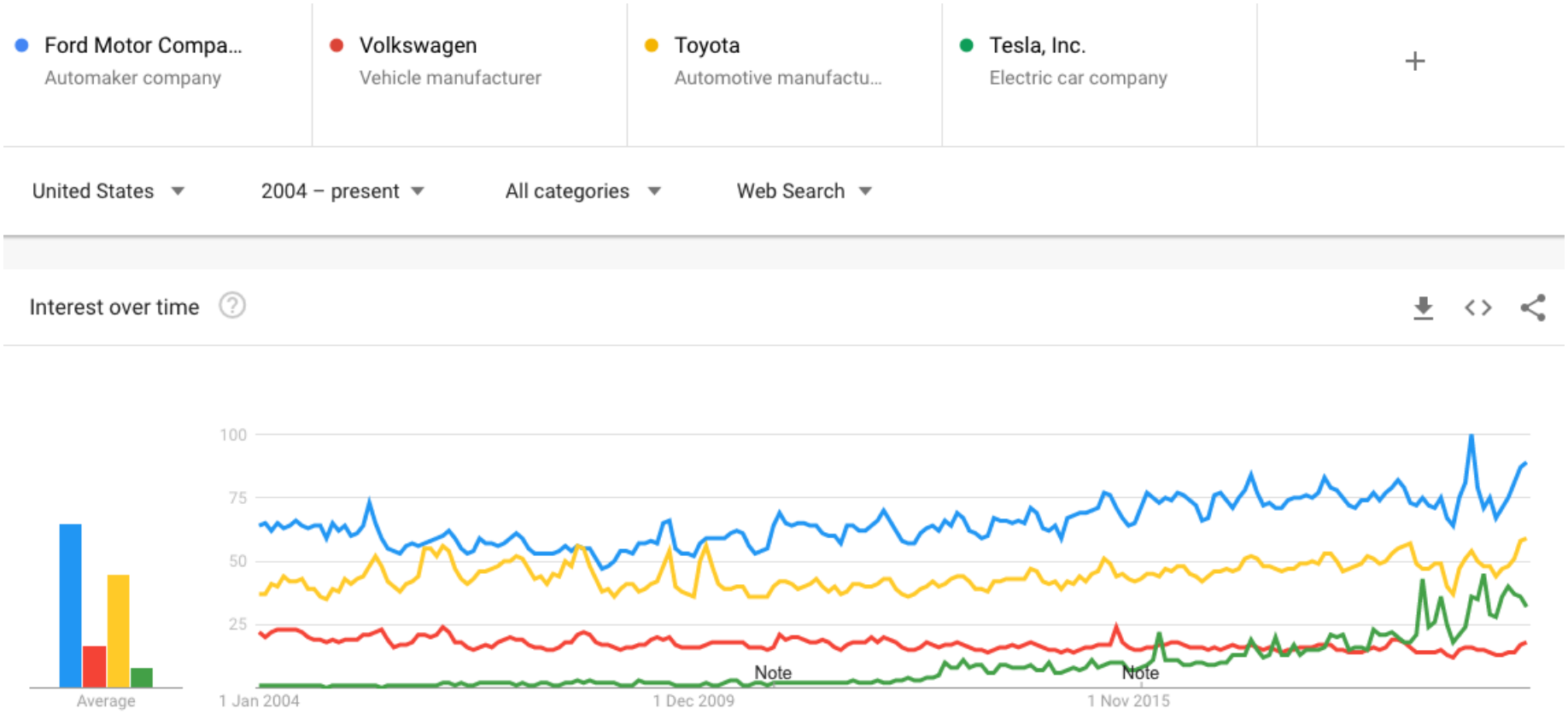
Hotel Trends



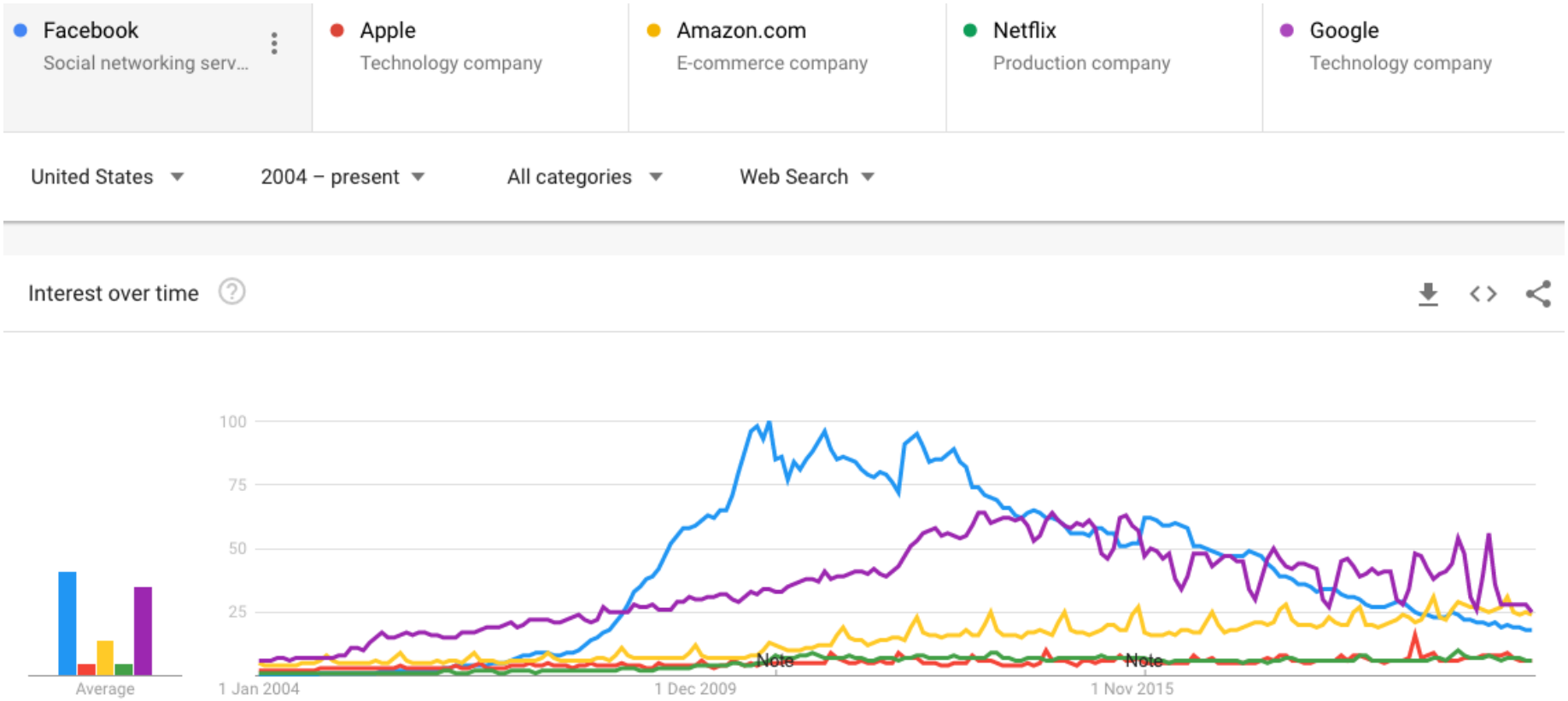
Car Trends


Sports car
Topic
Electric car
Topic
Hybrid car
Topic
Add comparison
Worldwide
2004 – present
All categories
Web Search
Interest over time
100
75
50
25
Note
Average
1 Jan 2004
1 Dec 2009
1 Nov 2015

Car Company trends



Tech Company Trends



Poll

- What research could you undertake using Google trends that might offer a positive social impact?

- **Slido.com #836 812**

Google trends for social good

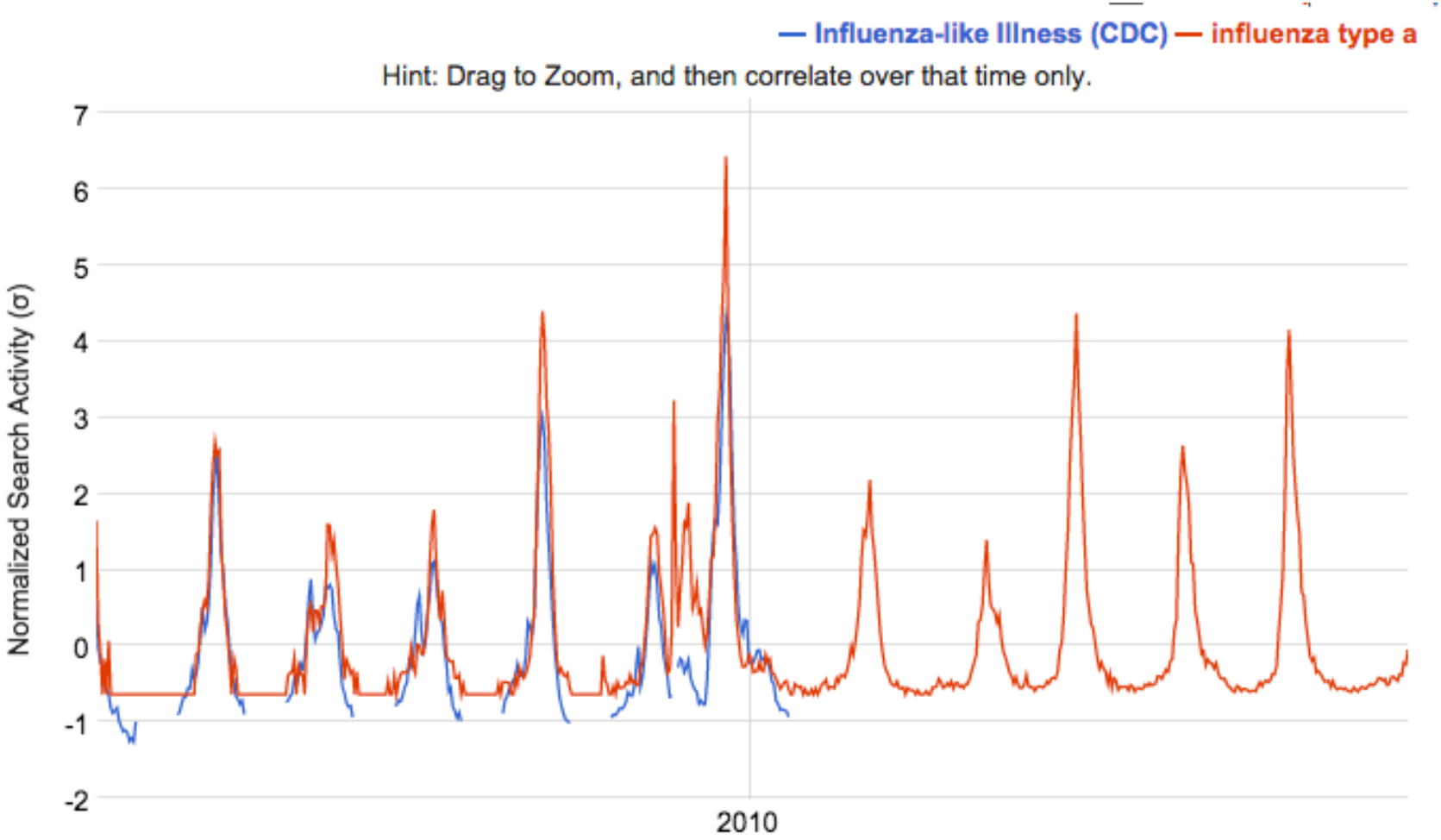
- Every day, millions of people use Google to dig up information that drives their daily lives, from how long their commute will be to how to treat their child's illness.
- This search data reveals a lot about the searchers: their wants, their needs, their concerns—extraordinarily valuable information.
- If these searches accurately reflect what is happening in people's lives, analysts could use this information to track diseases, unemployment, predict sales of new products, or even anticipate the results of elections.

Influenza-like illness

- Using data from the CDC about influenza, Google found correlations:

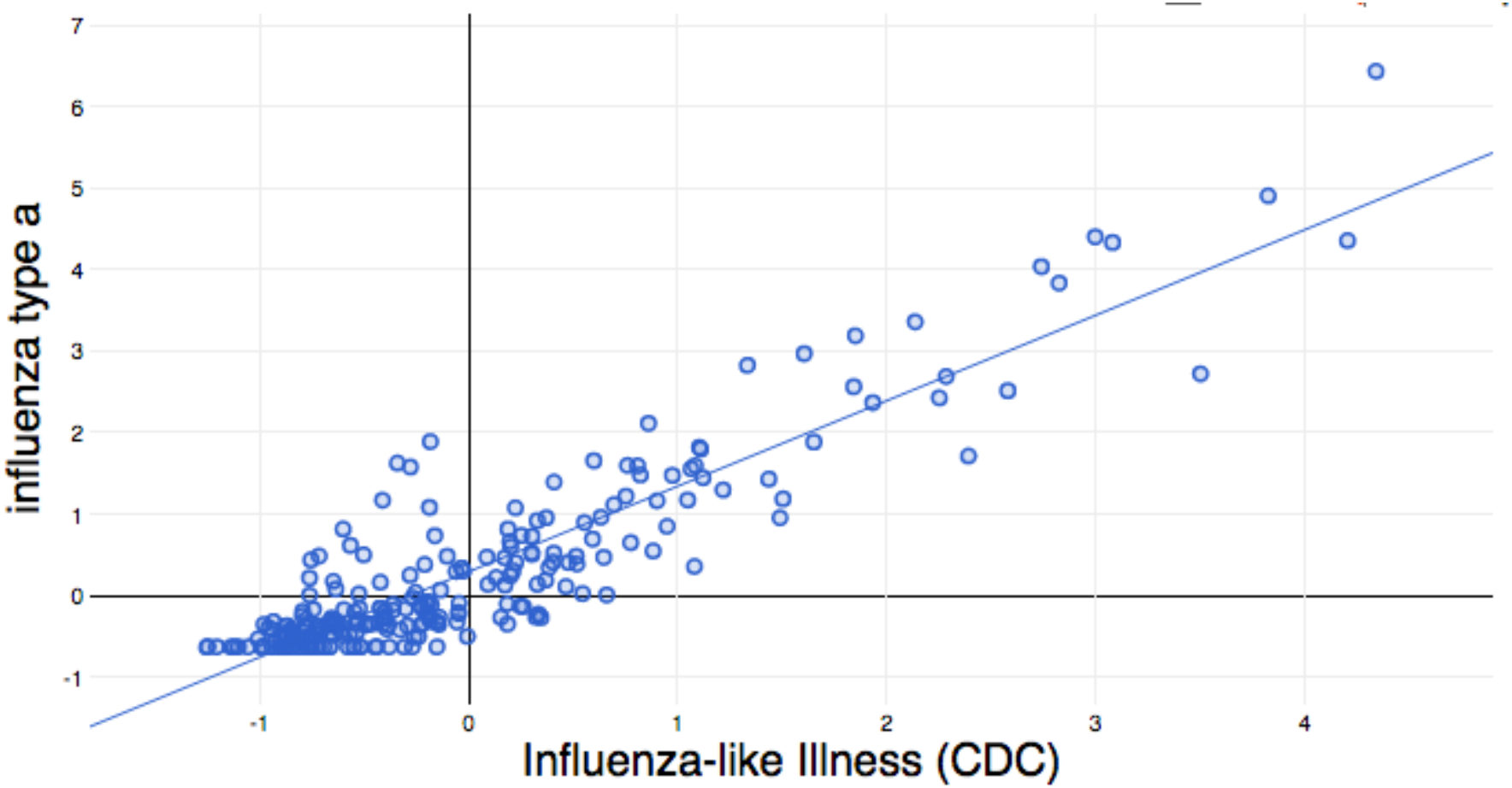
Search term	Correlation
Influenza type a	0.9069
Symptoms of flu	0.9038
Flu duration	0.9033
Flu contagious	0.8919
Flu fever	0.8851
Treat the flu	0.8831
How to treat the flu	0.8830
Signs of the flu	0.8815
How long is the flu	0.8775
Symptoms of the flu	0.8741

Flu time series



User uploaded activity for influenza-like illness (CDC) and US searches for **influenza type a** has a correlation of 0.9069.

Flu scatter plot



Google Flu Trends

- Google Flu Trends (GFT) tracks flu-related search terms as an early predictor of flu outbreaks.
- In 2008, researchers from Google published in Nature that:
- In the U.S., Google's data has a 92 percent correlation with actual U.S. government data from the Centers for Disease Control and Prevention (CDC).
- Google claimed accurate estimates of flu prevalence two weeks earlier than the CDC's data—turning the digital refuse of people's searches into potentially life-saving insights.
- And then, GFT failed spectacularly—missing at the peak of the 2013 flu season by 140 percent.

Source: https://www.google.org/flutrends/about/

Lessons from failure of GFT

- Lazer & Kennedy (2014) published in Science a report that documented and deconstructed the failure of Google to predict flu prevalence.
- The point was not to bury big data—their own research demonstrates the value of big data in modeling disease spread, real time identification of emergencies, and identifying macro economic changes ahead of traditional methods.
- But while Google's efforts in projecting the flu were well meaning, they were remarkably opaque in terms of method and data—making it dangerous to rely on Google Flu Trends for any decision-making.

GFT was overfitting

- Google's algorithm was quite vulnerable to overfitting to seasonal terms unrelated to the flu, like "high school basketball."
- With millions of search terms being fit to the CDC's data, there were bound to be searches that were strongly correlated by pure chance, and these terms were unlikely to be driven by actual flu cases or predictive of future trends.

GFT suffered from non-stationarity

- Google also did not take into account changes in search behavior over time.
- After the introduction of GFT, Google introduced its suggested search feature as well as a number of new health-based add-ons to help people more effectively find the information they need.
- While this is great for those using Google, it also makes some search terms more prevalent, throwing off GFT's tracking.

Unemployment

- The “Initial Jobless Claims” is a weekly report issued by US Department of Labor: http://www.dol.gov/ui/data.pdf.
- It tracks the number of people filed for the unemployment benefits and it is considered a leading indicator of the labor market.
- The claims are reported at the national and state level and the data data is available from first week of 1987.
- Advance initial claims weekly reports at the national level data are released 5 days after the week ends.
- Google trends is 7 days ahead of government release.

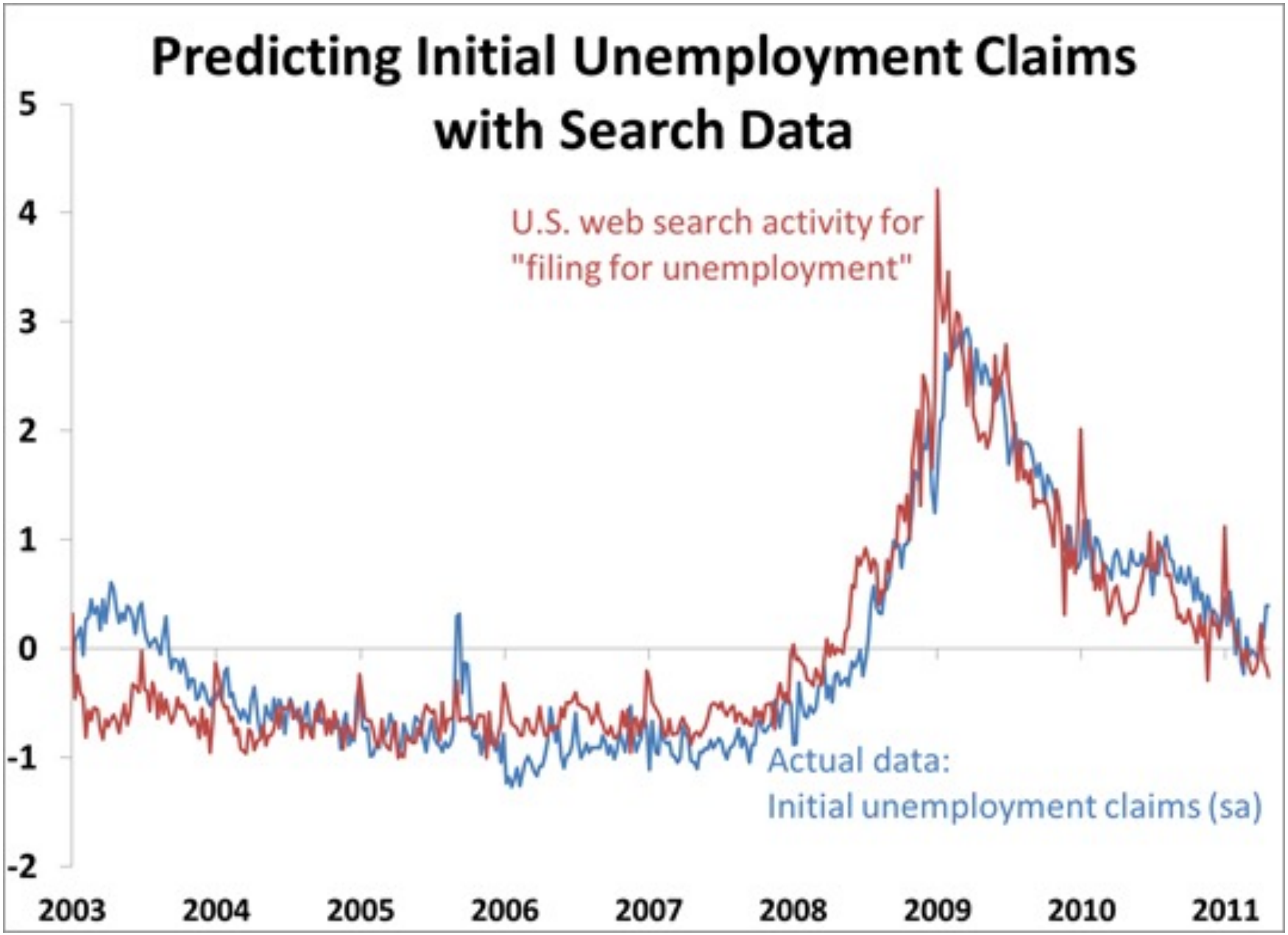
Model comparison

- Denote y_t, $Jobs_t$ and $Welfare_t$ as the time series of Initial Claims, 'Jobs' and 'Welfare & Unemployment' categories in Google Trends respectively.
- Baseline: $\ln(y_t) = a + b\ln(y_{t-1}) + \varepsilon_t$
- Alternative: $\ln(y_t) = \alpha + \beta\ln(y_{t-1}) + \gamma Jobs_t + \delta Welfare_t + \varepsilon_t$

	a	b	σ	MAE	α	β	γ	δ	σ	MAE
Estimate	0.1269	0.9902	0.0443	3.24%	1.6498	0.8727		0.0014	0.0429	2.73%
SE	0.1618	0.0126			0.3754	0.029		0.0003		

- With the Google Trends series, the model fit is improved significantly and the out-of-sample mean-absolute-error estimated with the rolling window for the past 24 weeks is decreased by 15.74%

Source: Choi & Varian (2009).



Searches for “filing for unemployment” & weekly initial unemployment claims (Correlation of 0.91)

Source: www.freakonomics.com

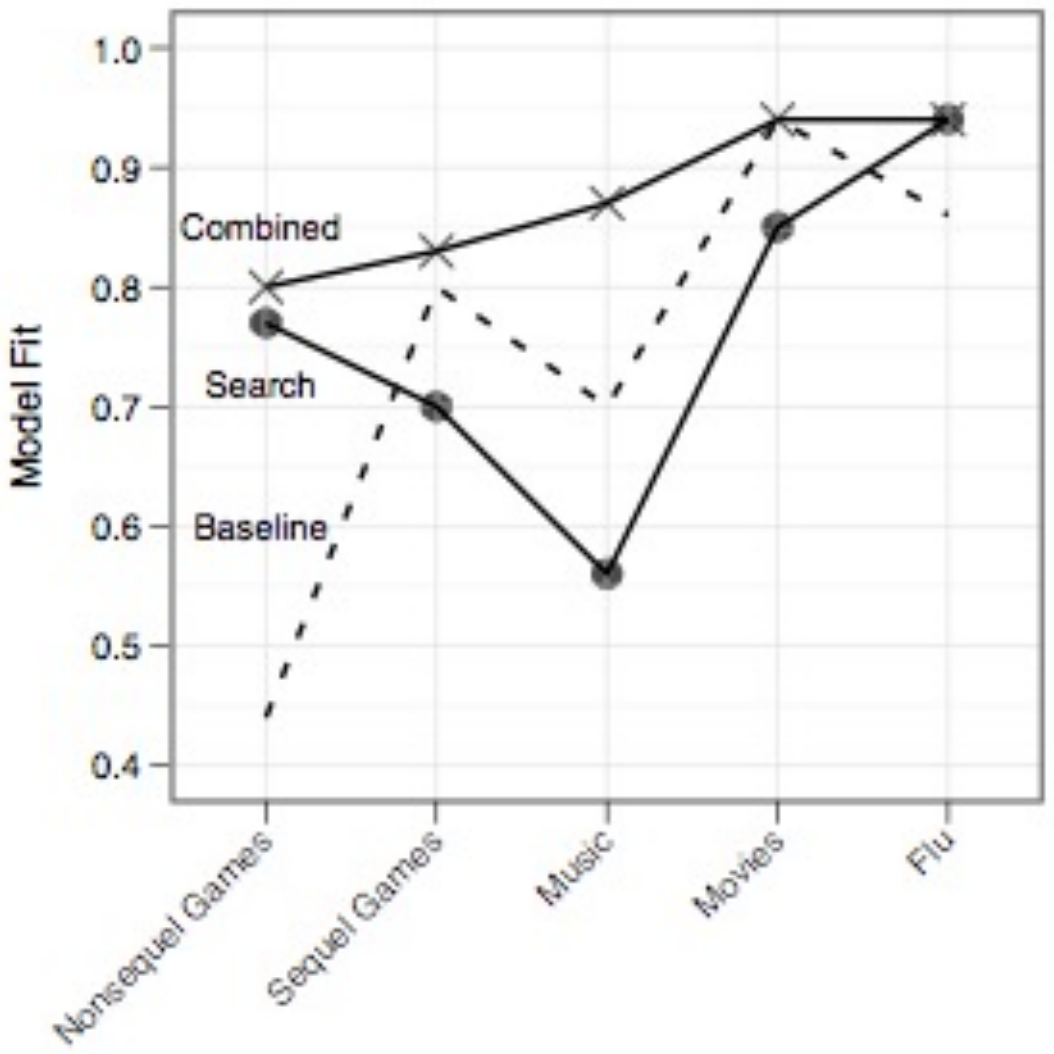
Quiz

- Do you think web search data can improve predictability of consumer behaviour?

  a) Yes
  b) No

- **Slido.com #836 812**

Correlation for different predictions



Baseline model shows performance without web search variable.

Goel et al. (2010). Predicting consumer behavior with web search. PNAS www.pnas.org/cgi/doi/10.1073/pnas.1005962107

Consumer price index

- The Consumer Price Index (CPI) is a measure that examines the weighted average of prices of a basket of consumer goods and services, such as transportation, food and medical care.
- The CPI is calculated by taking price changes for each item in the predetermined basket of goods and averaging them; the goods are weighted according to their importance.
- Changes in CPI are used to assess price changes associated with the cost of living.
- CPI is also known as headline inflation.

Why is CPI important?

- CPI is one of the most frequently used statistics for identifying periods of inflation or deflation.
- Large rises in CPI during a short period of time typically denote periods of inflation.
- Large drops in CPI during a short period of time usually mark periods of deflation.
- Inflation erodes the value of a national currency meaning that the cost of basic goods like milk will increase substantially over time.

Hedges against inflation

- For investors, the key to making money in an inflationary environment is to hold investments that increase in value at a rate in excess of the rate of inflation.
- A number of investments are historically viewed as hedges against inflation.
- These include real estate, gold, oil, stocks and inflation-indexed bonds.

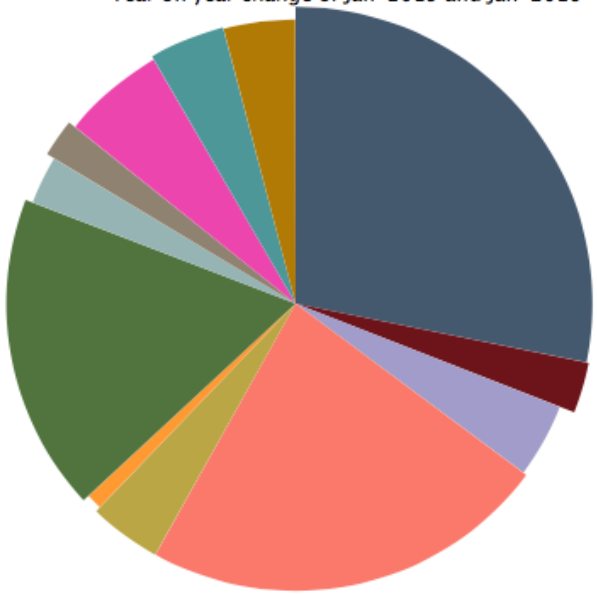
Consumer Price Index - Rwanda

.csv About

Rwanda's CPI rose 4.5% year on year change of January-2016

General CPI (4.5)%

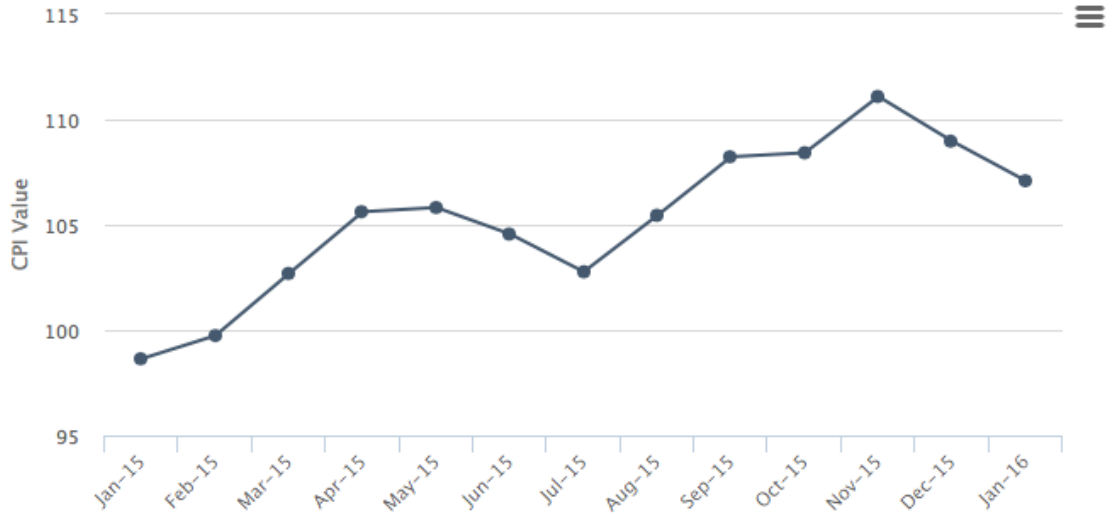
Year on year change of Jan-2015 and Jan-2016

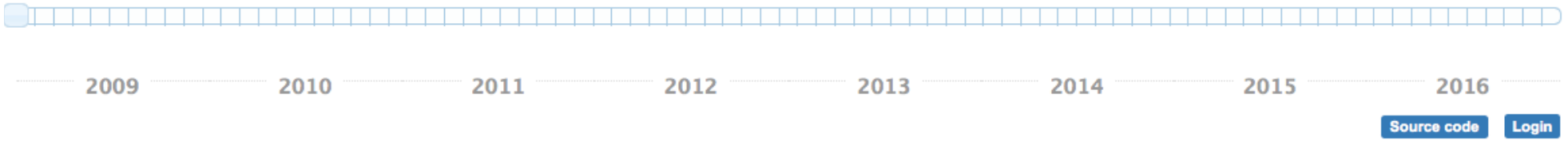


Food and non-alcoholic beverages

Food and non-alcoholic beverages contains 2 subgroups and was 28.19% of the Jan-2015 CPI Basket.

Prices fell by 1.73% in Jan-2016, and were 8.57% higher than they were in Jan-2015





Source code Login

Source: http://41.74.167.198/CPI/

Big Data Science

WEEK 3B

Course outline

Week	Lecture A	Lecture B
1	Weather & agriculture	Climate change
2	Climate scenarios	Catastrophe models
3	Social trends	Finance
4	Sentiment analysis	Health
5	Telemedicine	Mobile data
6	Data4Dev	Socioeconomic status

Today’s Lecture

No.	Activity	Description	Time
1	Challenge	Forecasting the stock market	10
2	Discussion	Web searches	10
3	Case study	Google trends and the market	10
4	Analysis	Anomalies	20
5	Demo	Environmental impact	20
6	Q&A	Questions and feedback	10

Efficient Market Hypothesis

- The efficient market hypothesis (EMH) states that asset prices fully reflect all available information.
- This implies that it is impossible to "beat the market" because stock market efficiency causes existing share prices to always incorporate and reflect all relevant information.
- Motivation for “no free lunch”.

Prices and a random walk

- The EMH implies that changes in the stock price reflect release of new information, changes in the market generally, or random movements around the value that reflects the existing information set.
- In 1973, Burton Malkiel published “A Random Walk Down Wall Street”, claiming that stock prices could therefore not be accurately predicted by looking at price history.
- As a result, Malkiel argued, stock prices are best described by a statistical process called a "random walk" meaning each day's deviations from the central value are random and unpredictable.
- This implies that paying managers to predict the market actually degrades the net portfolio return.
- Many portfolios managed by professional stock predictors do not outperform the market average return after accounting for the managers' fees.

Poll

- Do you think it is possible to forecast the stockmarket?

  a) Yes
  b) No

- **Slido.com #984 848**

Forecasting the stockmarket

- Pesaran & Timmerman (1994, 1995) found statistical and economic significance for the predictability of excess returns to common stocks.
- Granger (1992) identified benefits from taking a longer horizon, from using disaggregated data, from carefully removing outliers or exceptional events and especially from considering non-linear models.

Predictors

- Cenesizoglu and Timmermann (2008) divide the predictor variables into four broad categories:
  1) Valuation ratios capturing some measure of fundamental value to market value.
  2) Bond yield measures capturing the level or slope of the term structure or measures of default risk.
  3) Estimates of equity risk.
  4) Corporate finance variables.

Stockmarket crashes

- Joe Kennedy’s (President John F. Kennedy's father) got out of the market just before the 1929 crash and actually made a fortune shorting it.
- He explained that the trigger came when the shoeshine boy was talking to him about stock tips.
- When everyone, including those who really shouldn’t be in it, is in the market then that’s when it’s going to crash.
- Sir James Goldsmith (Anglo-French billionaire financier) explained his selling out of the market just before the 1987 crash.
- He simply didn’t understand the market, couldn’t work out what was going on and therefore decided to get out.

Google and Search Volume Index (SVI)

- Wu and Brynjolfsson (2009): SVI predicts home purchases
- Choi and Varian (2009): SVI predicts car purchases
- Choi and Varian (2009): SVI predicts unemployment benefit claims
- Goel et al. (2010): SVI predicts video game sales
- Da et al. (2011) demonstrate that SVI for company names and tickers provides a direct measure of *active* investor attention

Google Trends data

- *Volume limits:* You can only retrieve data for about 250 search terms per Google account or IP-address.
- *Relative scale:* It is scaled between 0 and 100, representing the minimum and maximum over the query period. Additionally, there is some some truncation of lower search volumes (0 does not mean no search volume).
- *Periodicity:* By default, it is weekly data. You can only retrieve *daily* SVI for 3-month blocks, so to construct a longer time series, you have to stitch together multiple time-series.

Investor attention

- Many variables can be measured to understand how much attention is being given by investors to individual stocks.
- Examples of indirect proxies include:
  a) turnover
  b) extreme returns
  c) news, and
  d) advertising expense.

SVI and stock prices

- Da et al. (2011) propose a new and direct measure of investor attention using search frequency in Google (Search Volume Index (SVI)).
- In a sample of Russell 3000 stocks from 2004 to 2008, they find that SVI (1) is correlated with but different from existing proxies of investor attention; (2) captures investor attention in a more timely fashion and (3) likely measures the attention of retail investors.
- An increase in SVI predicts higher stock prices in the next 2 weeks and an eventual price reversal within the year.
- It also contributes to the large first-day return and long-run underperformance of IPO stocks.

Da, Engelberg & Gao (2011). “In search of attention”. Journal of Finance LXVI No 5: 1461-1499.

Noisy tickers

- Da et al (2011) are cautious about using tickers with a generic meaning such as “GPS,” “DNA,” “BABY,” “A,” “B,” and “ALL.”
- They manually go through all the Russell stock tickers in the sample and flag such “noisy” tickers.
- These tickers are usually associated with abnormally high SVIs that may have nothing to do with attention paid to the stocks with these ticker symbols.
- While they report the results using all tickers to avoid subjectivity in sample construction, they confirm that the results are robust to the exclusion of the “noisy” tickers identified (about 7% of all Russell 3000 stocks).

Google trends and the market

- Preis et al. (2013) introduced a method to identify online precursors for stock market returns (DJIA), using trading strategies based on search volume data provided by Google Trends.
- Their analysis of Google search volume for 98 terms of varying financial relevance, suggests that increases in search volume for financially relevant search terms tend to precede large losses in financial markets.
- Out of these terms, only three were significant at the 5% level and all negative predictors.

Tobias Preis, Helen Susannah Moat and H. Eugene Stanley (2013). "Quantifying Trading Behavior ir Financial Markets Using Google Trends". Scientific Reports 3: 1684.

Poll

- Might Wikipedia be useful for forecasting the stockmarket?

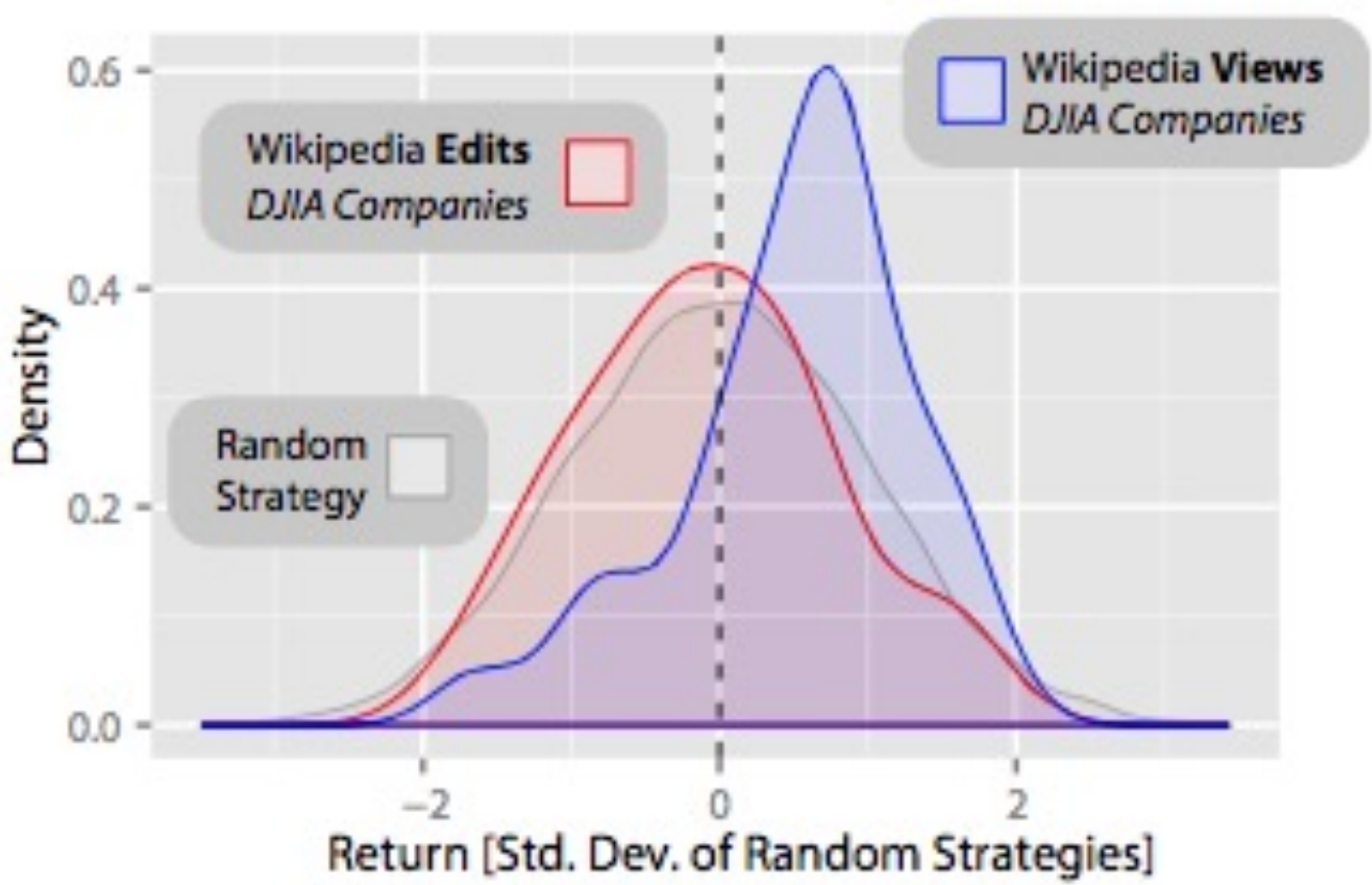
  a) Yes
  b) No

- **Slido.com #984 848**

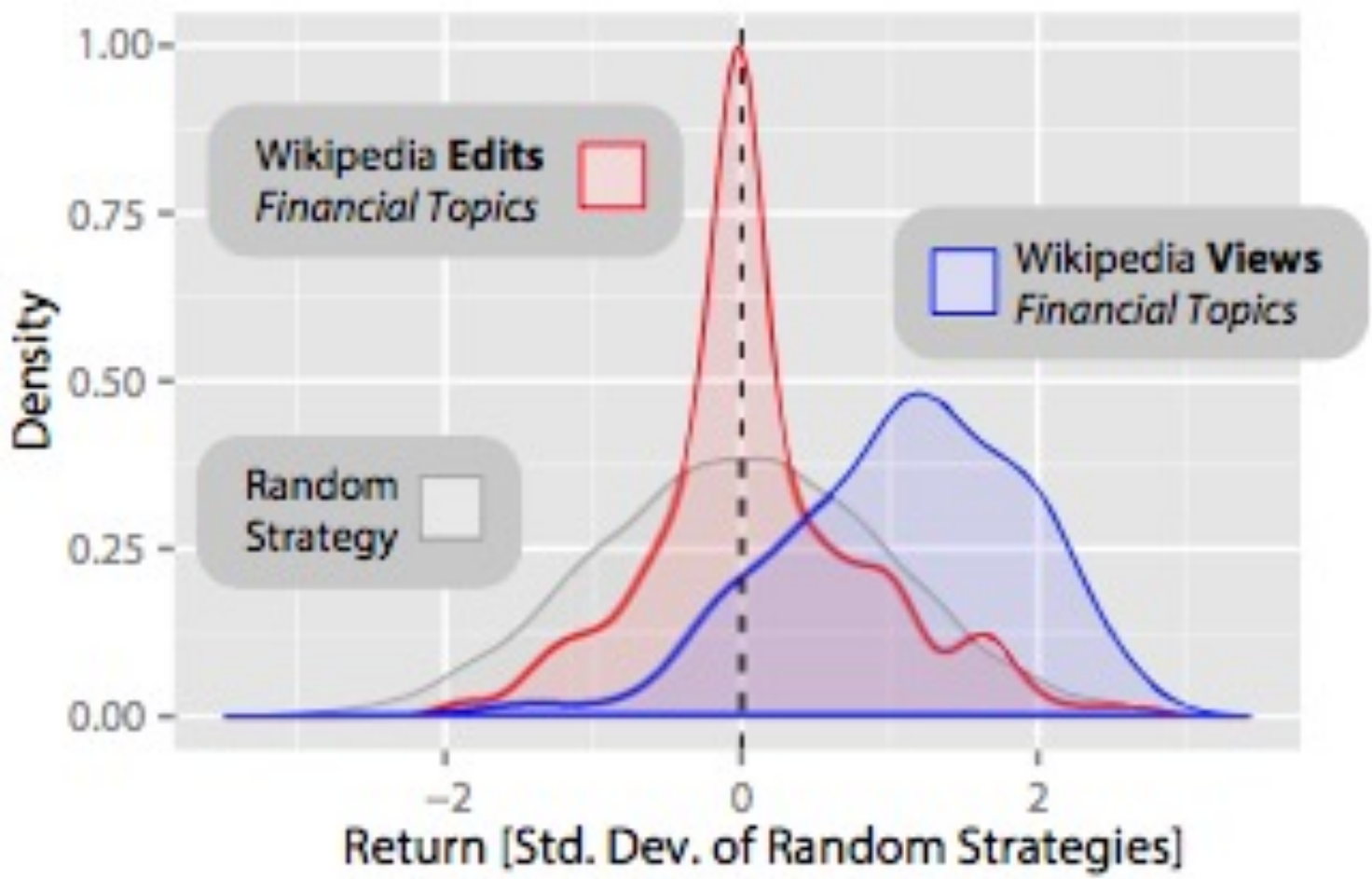
Wikipedia and the market

- Data generated through Internet usage contain traces of attempts to gather information before trading decisions were taken.
- Moat et al. (2013) present evidence that data on changes in how often financially related Wikipedia pages were viewed may have contained early signs of stock market moves.
- Their results suggest that online data may allow us to gain new insight into early information gathering stages of decision making.
- Weekly trading of DJIA (Dec-2007 to Apr-2012) using volume change versus previous 3-week average.

Wikipedia trading strategy I



Wikipedia trading strategy II

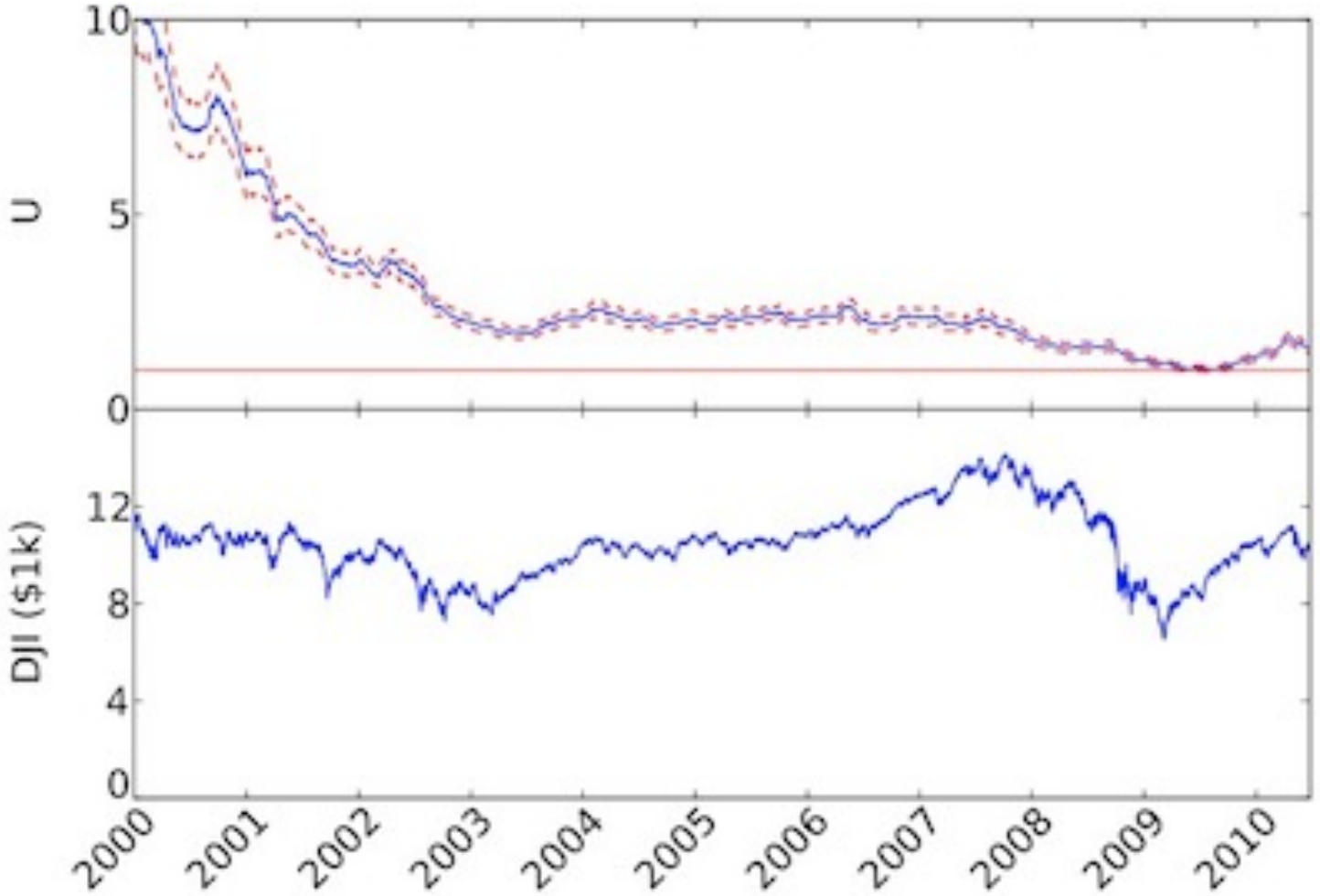


Crashes as phase transitions

- Bar-Yam (2013) used a phase transition model to motivate a measure of co-movement, or the likelihood of stocks to move in the same direction.
- When a market is healthy, co-movement is low.
- But in the months and years before a crash, co-movement seems to grow.
- If people act in tandem, by copying each other, a small nudge can send everyone in the same direction.
- This tells us that the system appears primed for collapse.

"Predicting economic market crises using measures of collective." By Dion Harmon, Marcus A. M. de Aguiar, David D. Chinellato, Dan Braha, Irving R. Epstein, Yaneer Bar-Yam. arXiv, Feb. 13, 2011.

Co-movement predicts crashes



A metric of stock co-movement during the 2000s (top). As it gets closer to zero, individual stocks are more likely to move up or down in the same direction. The Russell 3000 Index (bottom).

Financial anomalies

- Certain tradable anomalies seem to persist in the stock market.
- The discovery and testing of these anomalies relies on quantitative analysis of large quantities of financial data.
- Each anomaly is generally based on an intuitive reason (psychological, emotional etc) and supported by empirical evidence.

Size effect

- Smaller firms, measured by smaller capitalization, tend to outperform larger companies.
- Smaller firms have greater potential for growth and rapid than larger firms.
- If growth follows an S-curve, then one might prefer small firms that are about to accelerate up the curve than those that are slowing down.

Value effect

- Unusually cheap stocks should attract buyers' attention and revert to the mean.
- Stocks with below-average price-to-book ratios tend to outperform the market.
- The price-to-book ratio (P/B Ratio) is a ratio used to compare a stock's market value to its book value.
- It is calculated by dividing the current closing price of the stock by the latest quarter's book value per share.
- Book value is the total assets minus intangible assets (patents, goodwill) and liabilities.

Momentum effect

- Prior stock returns have been shown to have explanatory power in the cross section of common stock returns.
- Stocks with prices on an upward (downward) trajectory over a prior period of 3 to 12 months have a higher than expected probability of continuing on that upward (downward) trajectory over the subsequent 3 to 12 months.

Reversal effect

- Stocks that have performed well (or poorly) over certain periods of time tend to reverse to the mean in the following period.
- The past’s top performers become the future's underperformers, and vice versa.
- Reversal effects arise from the over-reaction of investors to information and news.
- Pairs trading (Coca Cola and Pepsi) are an example of a mean-reversion strategy.

January effect

- Stocks that underperformed in the fourth quarter of the prior year tend to outperform the markets in January.
- The reason for the January effect is that investors will often look to sell underperforming stocks late in the year so that they can use their losses to offset capital gains taxes.
- Excess selling pressure before January and excess buying pressure after 01-Jan causes this effect.

Day of the week effect

- Stocks tend to move more on Fridays than Mondays, and that there is a bias toward positive market performance on Fridays.
- Perhaps an end-of-week optimism permeates the market as traders and investors look forward to the weekend.
- Alternatively, perhaps the weekend gives investors a chance to catch up on their reading, stew and fret about the market, and develop pessimism going into Monday.

Capital asset pricing model

- The capital asset pricing model (CAPM) is a model that describes the relationship between risk and expected return:

$$r = r_f + \beta(r_m - r_f) + \alpha$$

- where r is the expected return,
- r_f is the risk free interest rate
- r_m is the expected market return
- β is the sensitivity to the market

Fama-French 3 factor model

- The Fama–French three-factor model describes stock returns using three factors:

$$r = r_f + \beta_m(r_m - r_f) + \beta_s SMB + \beta_v HML + \alpha$$

- where β_m, β_s, β_v relate to market risk, company size and company value (Price-to-Book Ratio).
- SMB is Small [market capitalization] Minus Big" HML is "High [book-to-market ratio] Minus Low" These variables measure the historic excess returns of small caps over big caps and of value stocks over growth stocks.

Other factor models

- The Carhart (1997) four-factor model contains an additional momentum factor (MOM), which is long prior-month winners and short prior-month losers.
- The inclusion of MOM has been recommended by Cliff Asness, co-founder of AQR capital.
- The profitability factor (RMW) is the difference between the returns of firms with robust (high) and weak (low) operating profitability
- The investment factor (CMA) is the difference between the returns of firms that invest conservatively and firms that invest aggressively.

Poll

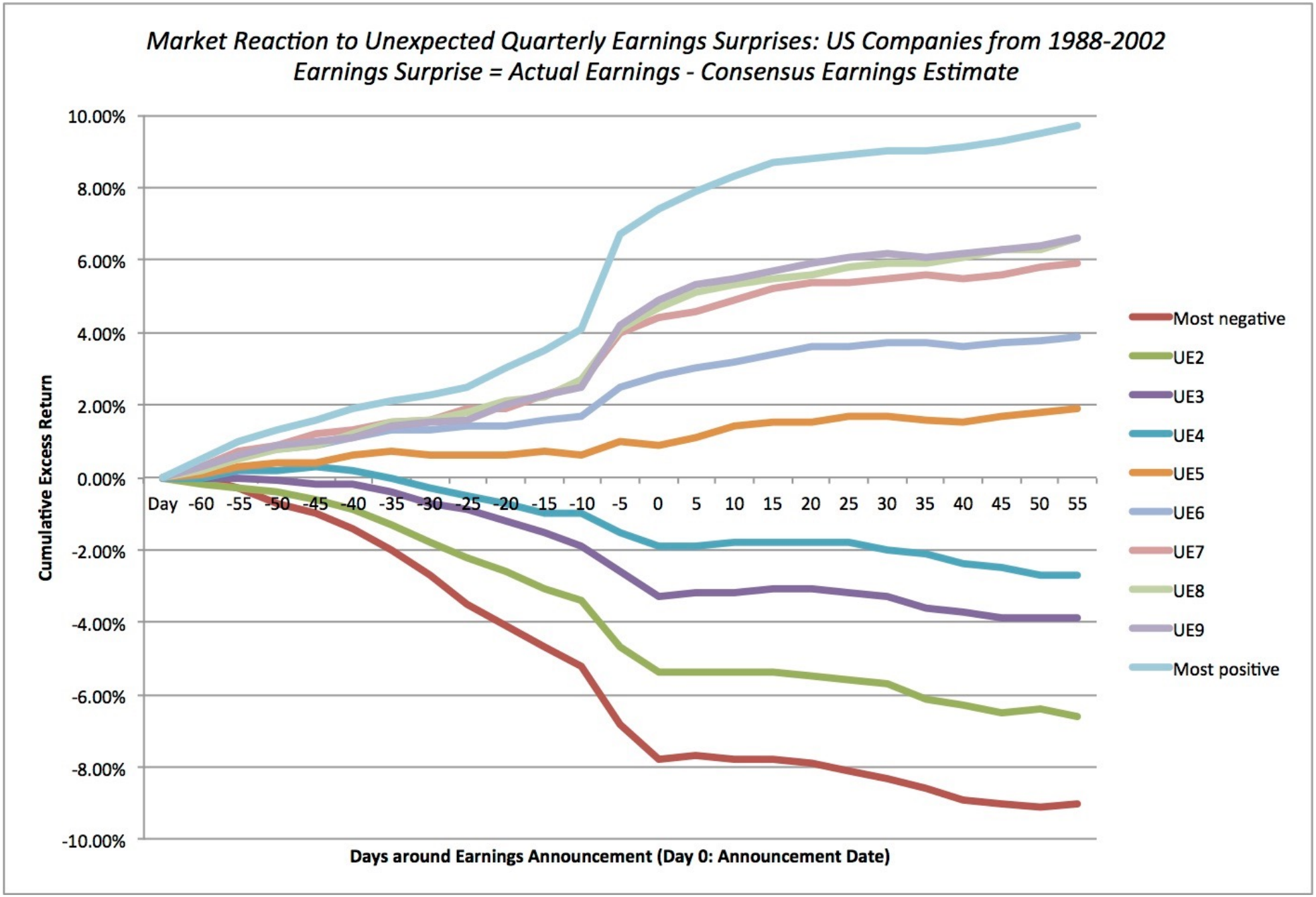
- The Efficient Market Hypothesis claims that information is factored into the market immediately. Do we have any evidence of earnings news propagating slowly enough to trade upon?

  a) Yes
  b) No

- **Slido.com #984 848**

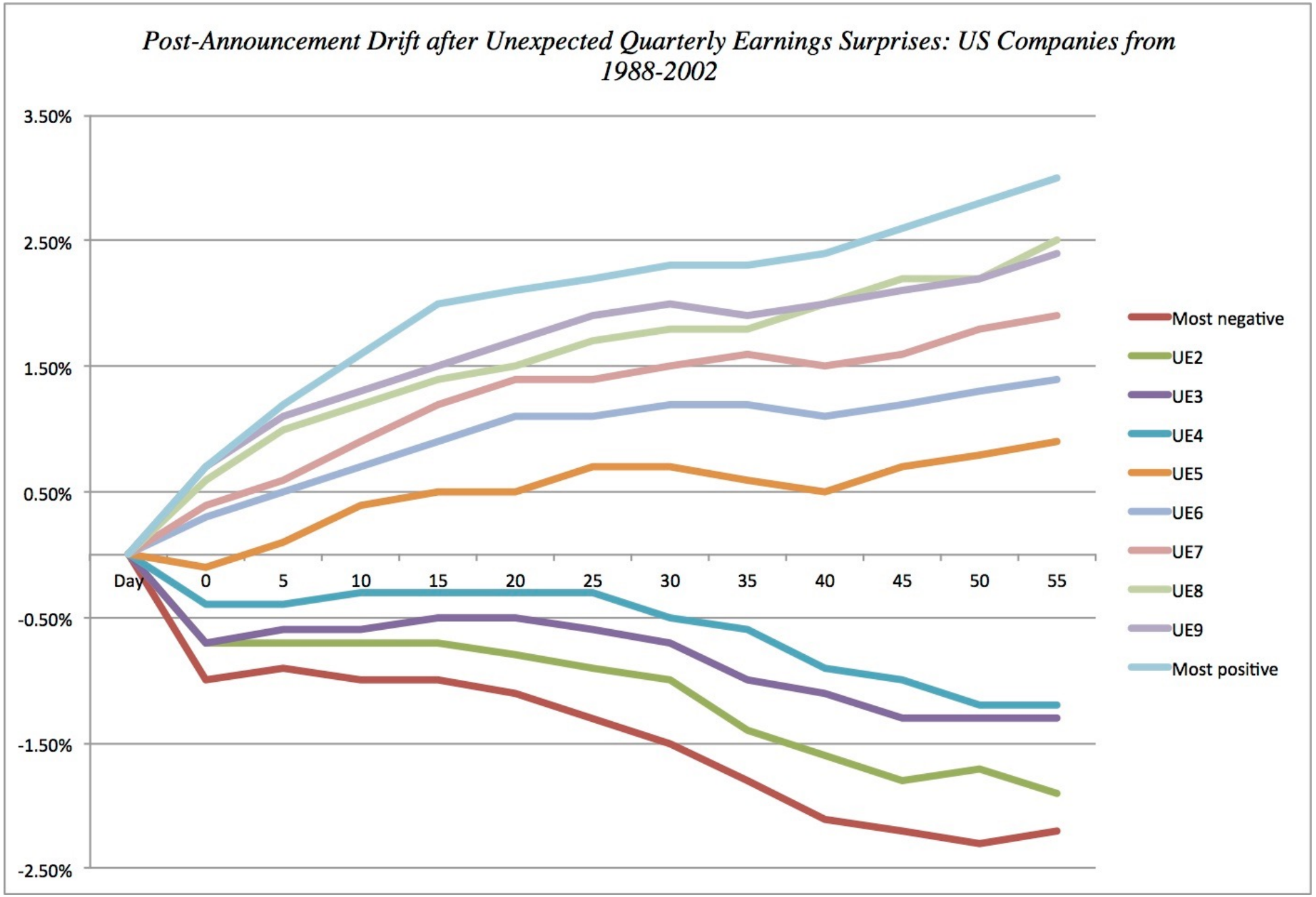
Post–earnings-announcement drift

- Post–earnings-announcement drift (PEAD) is the tendency for a stock's cumulative abnormal returns to drift in the direction of an earnings surprise for several weeks following an earnings announcement.
- One explanation for the effect is investor under-reaction to earnings announcements.
- The level of surprise is measured using the standardized unexpected earnings (SUE).
- SUE is the standardized difference between reported earnings and expected earnings, where expected earnings is modelled based on the assumption that earnings follows a seasonal random walk with a trend.


Market Reaction to Unexpected Quarterly Earnings Surprises: US Companies from 1988-2002
Earnings Surprise = Actual Earnings - Consensus Earnings Estimate
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
-2.00%
-4.00%
-6.00%
-8.00%
-10.00%
Cumulative Excess Return
Day -60 -55 -50 -45 -40 -35 -30 -25 -20 -15 -10 -5 0 5 10 15 20 25 30 35 40 45 50 55
Days around Earnings Announcement (Day 0: Announcement Date)
Most negative
UE2
UE3
UE4
UE5
UE6
UE7
UE8
UE9
Most positive

Earnings announcements

- 1. Pre-announcement drift: There is a mild drift in stock prices before earnings reports that is consistent with the eventual surprise: prices move up before positive surprises and down before negative surprises.
- 2. Announcement effect: The announcement still contains news. On the announcement, the price effect reflects the magnitude and the direction of the surprise, with stock prices going up about 3%, on average, in reaction to the most positive surprises.
- 3. Post-announcement drift: The most surprising finding is that stock prices continue to drift after the announcement in response to the surprise.


Post-Announcement Drift after Unexpected Quarterly Earnings Surprises: US Companies from 1988-2002
3.50%
2.50%
1.50%
0.50%
-0.50%
-1.50%
-2.50%
Day
0
5
10
15
20
25
30
35
40
45
50
55
Most negative
UE2
UE3
UE4
UE5
UE6
UE7
UE8
UE9
Most positive

Environmental, social and corporate governance (ESG)

- Sustainable, responsible and impact investing (SRI) is an investment discipline that considers environmental, social and corporate governance (ESG) criteria to generate long-term competitive financial returns and positive societal impact.
- A 2012 study by Deutsche Bank Group Climate Change Advisors found that incorporating ESG data in investment analysis is “correlated with superior risk-adjusted returns at a securities level.”

Environmental Profit and Loss

- An Environmental Profit and Loss Account (EP&L) places a financial value on environmental impacts along the entire value chain of a business to help companies combine sustainability metrics with traditional business management.
- PUMA developed EP&L and found the total impact of its direct and supply chain operations was valued at EUR 145 million (one third each to water use and GHGs).
- PUMA believes that businesses should account for and, ultimately, pay for the cost to nature of doing business.

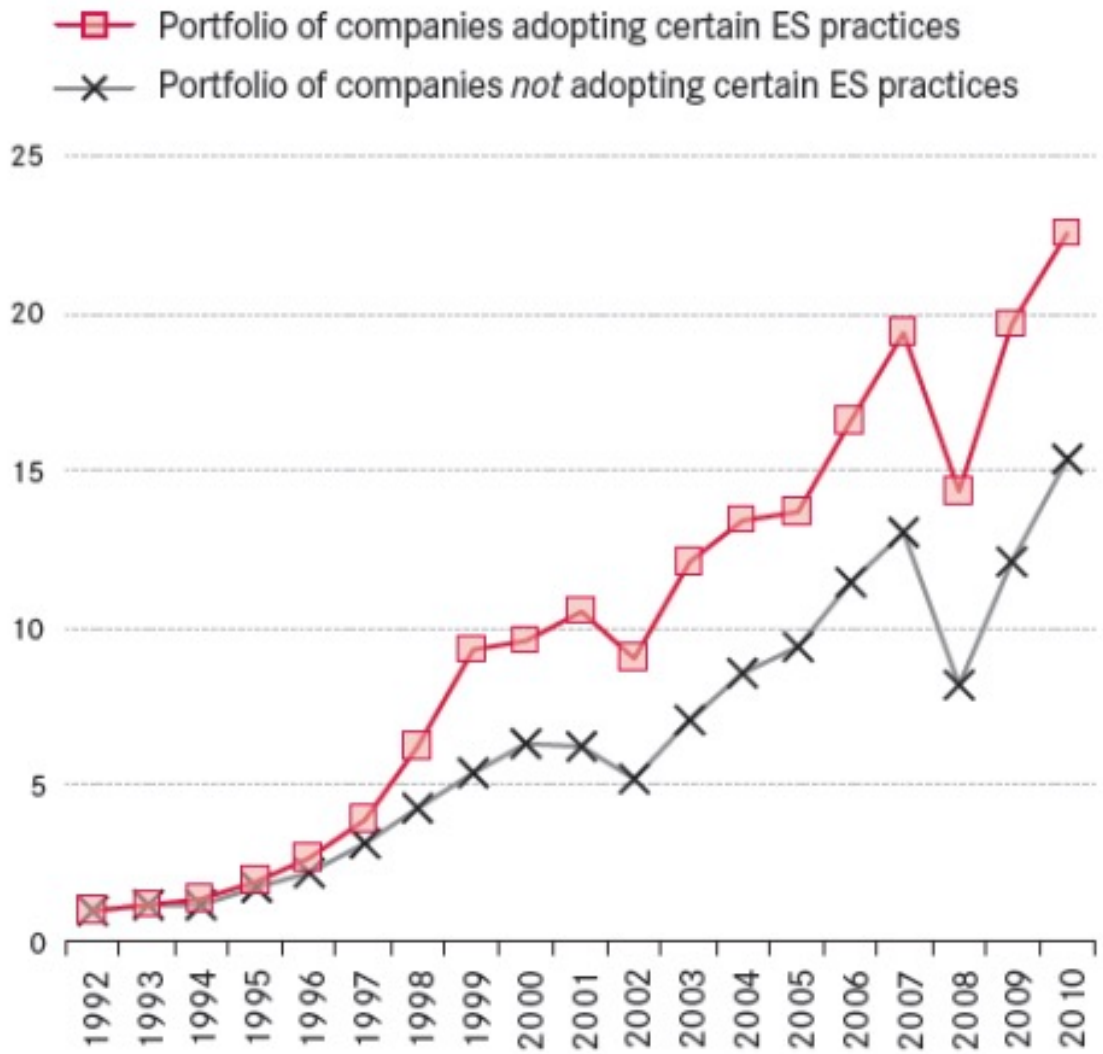
Environmental impact

- New indices focus on selecting firms that have little environmental impact.
- The FTSE4Good Index is a series of ethical investment stock market indices launched in 2001 by the FTSE Group.
- The Dow Jones Sustainability Indices (DJSI) are a family of indices evaluating the sustainability performance of the largest 2,500 companies.
- MSCI ESG indexes are designed to represent the most prevalent environmental, social and governance (ESG) investment strategies.

Pension funds

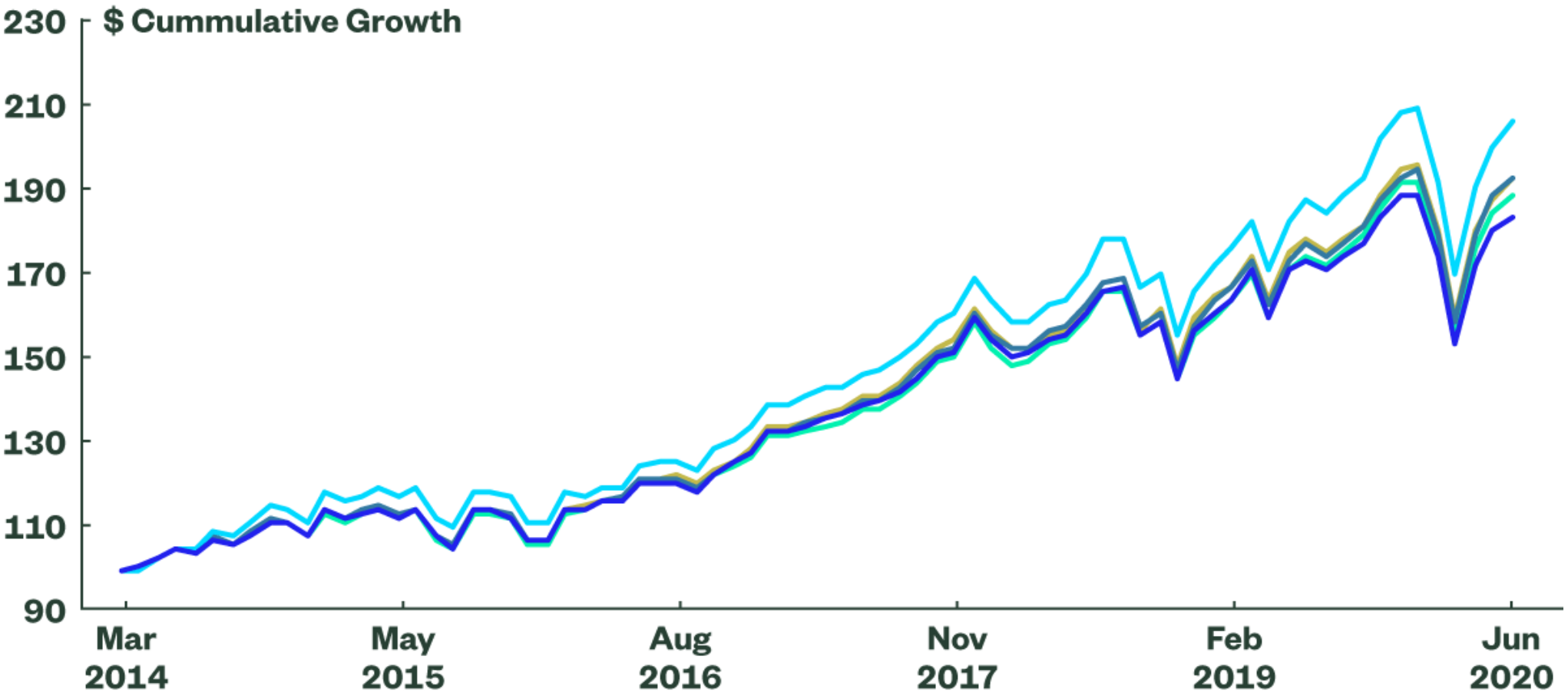
- Pension funds are allocating more capital to firms with good ESG characteristics.
- Fossil fuel divestment campaigns are helping to drive some of this behavior change.
- Pension funds take long-term decisions with investment horizons of a few decades.
- Furthermore, there are now approaches to measure the substantial risks due to owning firms with carbon intensive operations.

Evolution of $1 Invested in the Stock Market in Equal-Weighted Portfolios of Companies Adopting/Not Adopting Certain Environmental and Social Practices



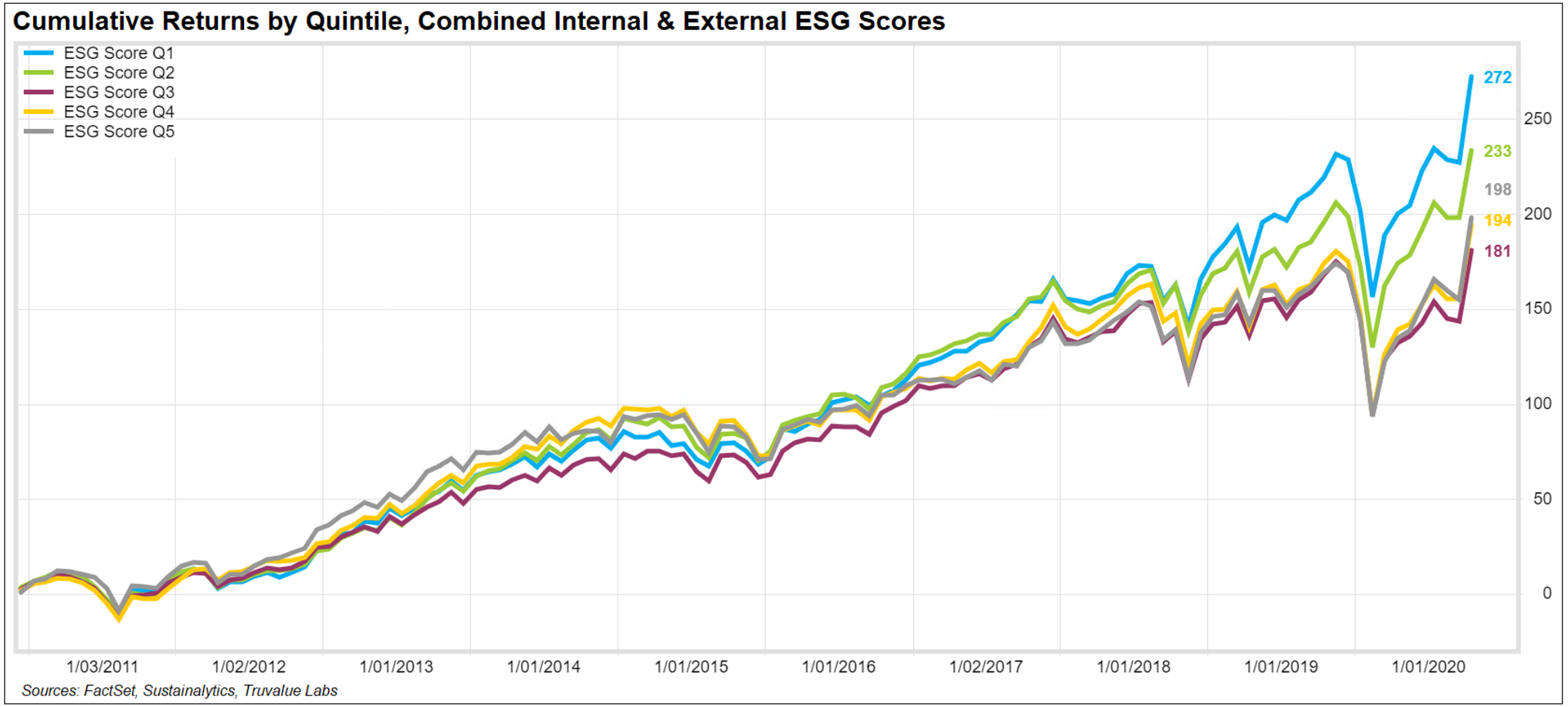
Source: Robert G. Eccles, Ioannis Ioannou, and George Serafeim, "The Impact of Corporate Sustainability on Organizational Processes and Performance," *Management Science* 60, no. 11, November 2014.

Figure 2: Cumulative Performance of ESG Index Strategies (%)
(March 31, 2014 to June 30, 2020)



Source: MSCI, FTSE/Russell/Bloomberg; S&P; State Street Global Advisors; all calculations by State Street Global Advisors.

ESG Quintiles



ESG Quiz

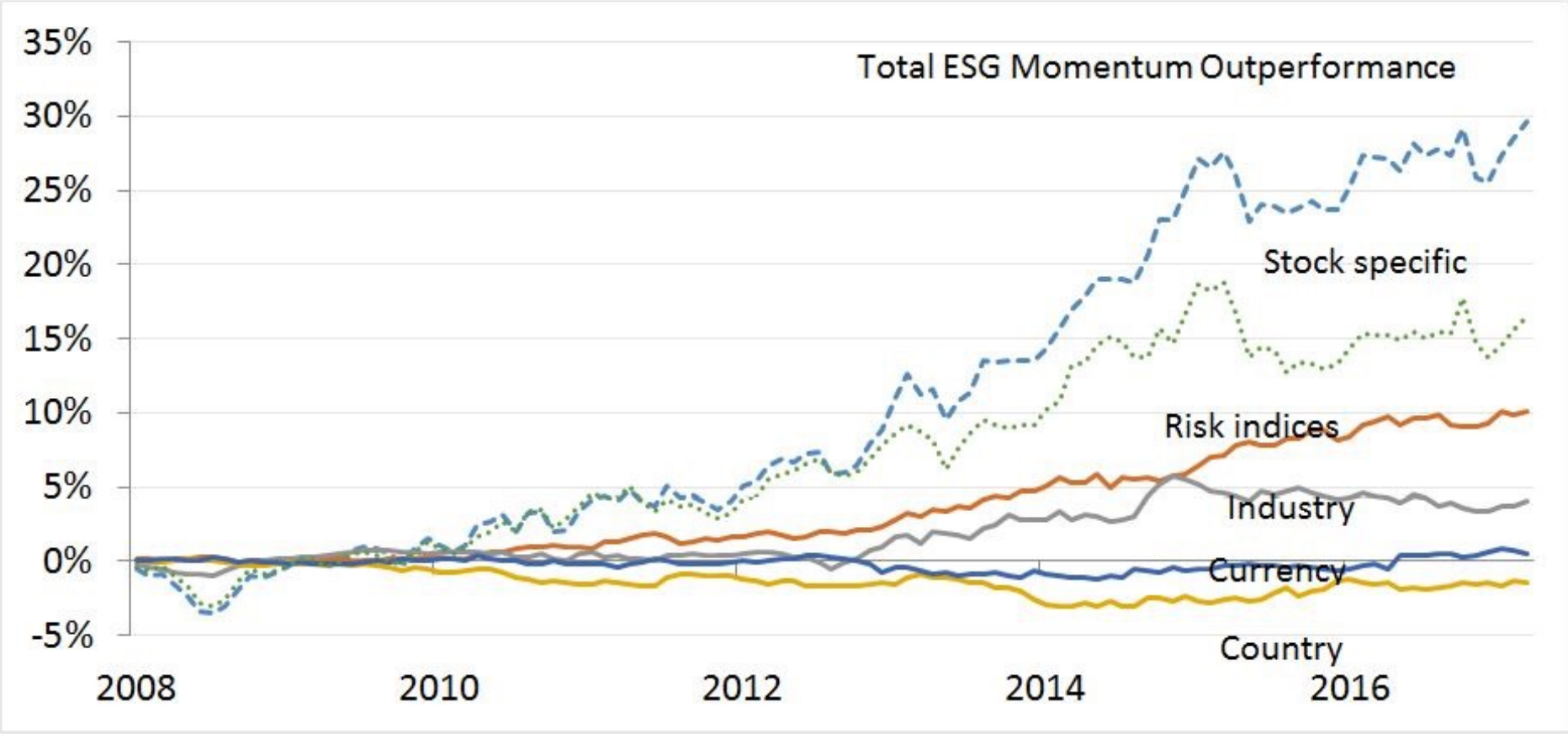
- ESG information can be used to select stocks that will perform better in the future because of the following factors

  a) Country
  b) Currency
  c) Industry
  d) Risk
  e) Stock specific information

- **Slido.com #984 848**

ESG performance

ESG Momentum Strategy vs. MSCI World cumulative relative performance decomposed into factor returns, Feb 2008- April 2017.



Source: MSCI.

Stranded assets

- Stranded assets are "assets that have suffered from unanticipated or premature write-downs, devaluations or conversion to liabilities".
- Environment-related risk factors that could result in stranded assets are:
  - environmental challenges (e.g. climate change, natural capital degradation)
  - changing resource landscapes (e.g. shale-gas abundance, phosphate scarcity)
  - new government regulations (e.g. carbon pricing, air pollution regulation, carbon bubble)
  - falling clean-technology costs (e.g. solar PV, onshore wind, electric vehicles)
  - evolving social norms (e.g. fossil fuel divestment campaign) and consumer behaviour (e.g. certification schemes)
  - litigation (e.g. carbon liability) and changing statutory interpretations (e.g. fiduciary duty, disclosure requirements)

Carbon reporting

- Emphasis has been primarily on scope1 reductions and overall environmental impact.
- Resilience refers to the capacity to adapt, respond and transform.
- Existing indices aim to measure companies ability to adapt and respond (reducing damage of existing operations).
- We focus on their contribution to reduced emissions in society though innovation via new products and delivered services.

Value Creation

- Companies survive by generating value such that profits exceed costs.
- Cost of environmental impact (GHG emissions) remains a risk that could be realised in the near future.
- There is a fine balance between innovation and environmental impact (new versus old firms).
- Innovation is required in order to create value without causing environmental impact.
- By predicting the optimal level of innovation, we can assess the ability of companies to seize opportunities and reduce environmental impacts.

Data for measuring impact and innovation

- Trucost measures the environmental impact (normalised by revenue to make companies comparable).
- Innovation and potential to seize opportunities is quantified using R&D intensity (R&D expenditure divided by revenue).
- Innovation is negatively correlated with environmental impact (correlation of -0.42).
- The transformative journey that companies need to make in order to remain resilient is mapped out.
- We can predict innovation using level of impact and type of industry (correlation of 0.71).

Innovation versus environmental impact