Common pitfalls in quantitative research

Ivan Smirnov
Chair for Computational Social Sciences and Humanities, RWTH Aachen
Laboratory for Computational Social Science, HSE University

Improves your memory by 10%

Meanwhile **Chinese exports** soared by 32% in April compared with the same month last year. The economic recovery in America, combined with factory closures in other parts of the world, spurred demands for goods made in China. Chinese imports also jumped, rising 43% compared with April 2020, the fastest growth since January 2011.

A federal grand jury indicted four former Minneapolis police officers on charges of violating the civil rights of **George Floyd**, a black man whose killing set off a summer of massive racial-justice protests last year. Derek Chauvin—who was recently convicted of second- and third-degree murder and manslaughter after kneeling on Mr Floyd's neck for more than nine minutes—was among those indicted.

Copper prices hit a record high. The commodity, seen as a bellwether for the health of the global economy, rose by 1.4% to $10,361 a tonne, surpassing the previous peak set in 2011, during a commodities boom. Demand comes from China and the green transition in rich countries: the metal is needed for electric vehicles, wind turbines and solar panels.

Fact of the day: Over 50 monetary authorities, representing the bulk of global GDP, are exploring digital currencies. Read the full story.

Corrections: We reported that IBM's new chip was projected to achieve 45% higher performance while using 75% less energy. That should have been 45% percent higher performance or 75% less energy, but not both at the same time. In another item, we referred to the Welsh Assembly; since last year it has been the Welsh Parliament. Apologies.

Big news.

11 %
thinner

15 %
smaller

16 %
lighter

Improves your memory by 10%

How it was measured?

00:05

48154724439

Do Men Think About Sex Every Seven Seconds?

An old bit of accepted psychology holds that men, on average, think about sex every seven seconds.

By Snopes Staff

Published 18 April 2002

Sex on the brain?: An examination of frequency of sexual cognitions as a function of gender, erotophilia, and social desirability

Fisher, T. D., Moore, Z. T., & Pittenger, M. J. 2012 //Journal of Sex Research

0002

How it was measured?

Experience sampling method would probably be better

3.4 million real-world learning management system logins reveal the majority of students experience social jet lag correlated with decreased performance

3.4 million real-world learning management system logins reveal the majority of students experience social jet lag correlated with decreased performance

Logins to LMS

3.4 million real-world learning management system logins reveal the majority of students experience social jet lag correlated with decreased performance

Logins to LMS

Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures

3.4 million real-world learning management system logins reveal the majority of students experience social jet lag correlated with decreased performance

Logins to LMS

Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures

Number of certain words

3.4 million real-world learning management system logins reveal the majority of students experience social jet lag correlated with decreased performance

Logins to LMS

Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures

Number of certain words

Online social integration is associated with reduced mortality risk

3.4 million real-world learning management system logins reveal the majority of students experience social jet lag correlated with decreased performance

Logins to LMS

Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures

Number of certain words

Online social integration is associated with reduced mortality risk

Number of friends

Historical analysis of national subjective wellbeing using millions of digitized books

Hills TT, Proto E, Sgroi D, Seresinhe CI 2019 *// Nature human behaviour*

Idea: sentiment analysis

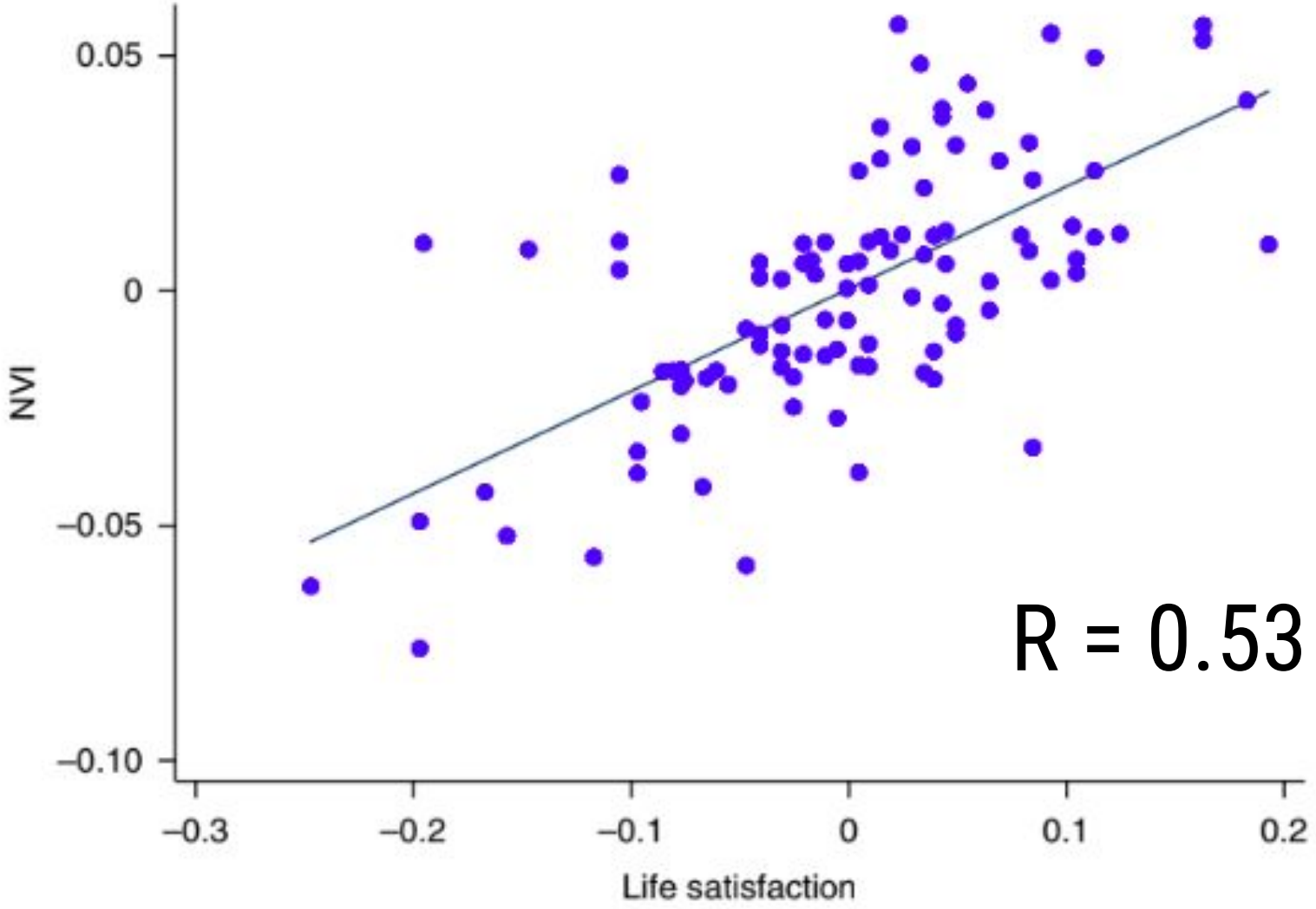
Google books

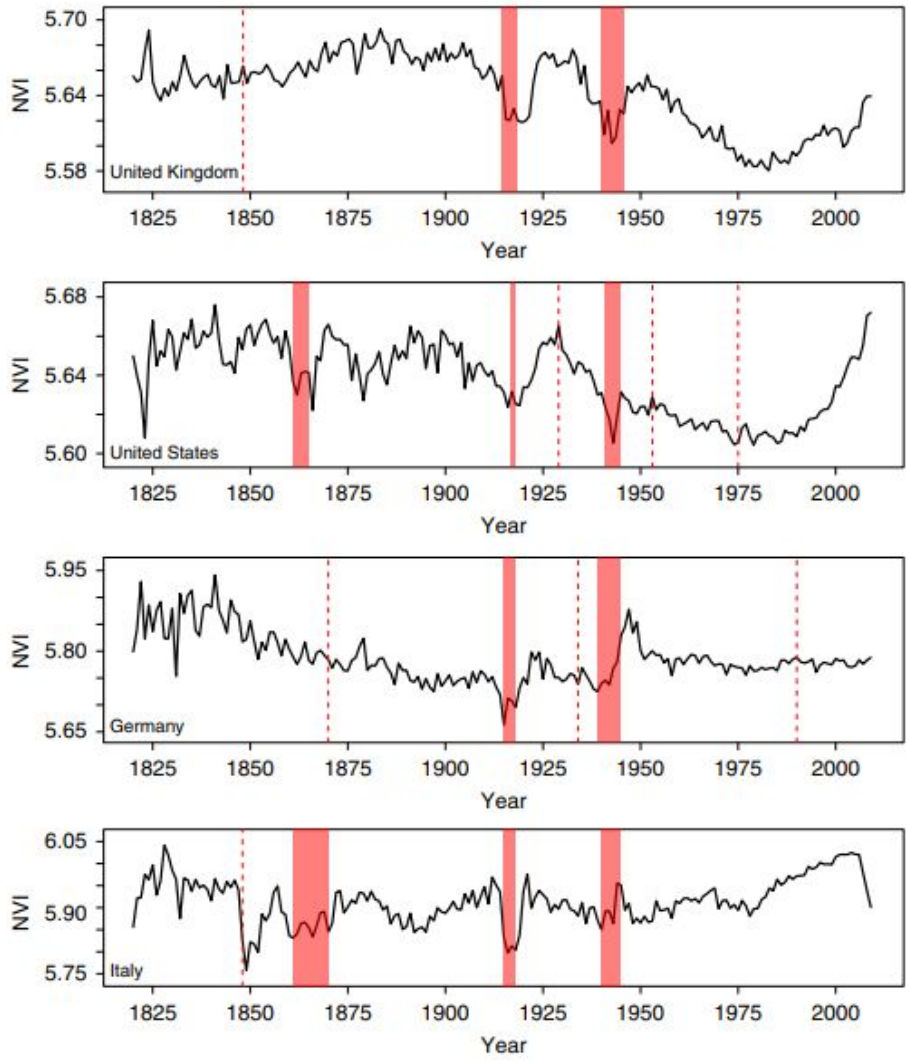
Dictionary method

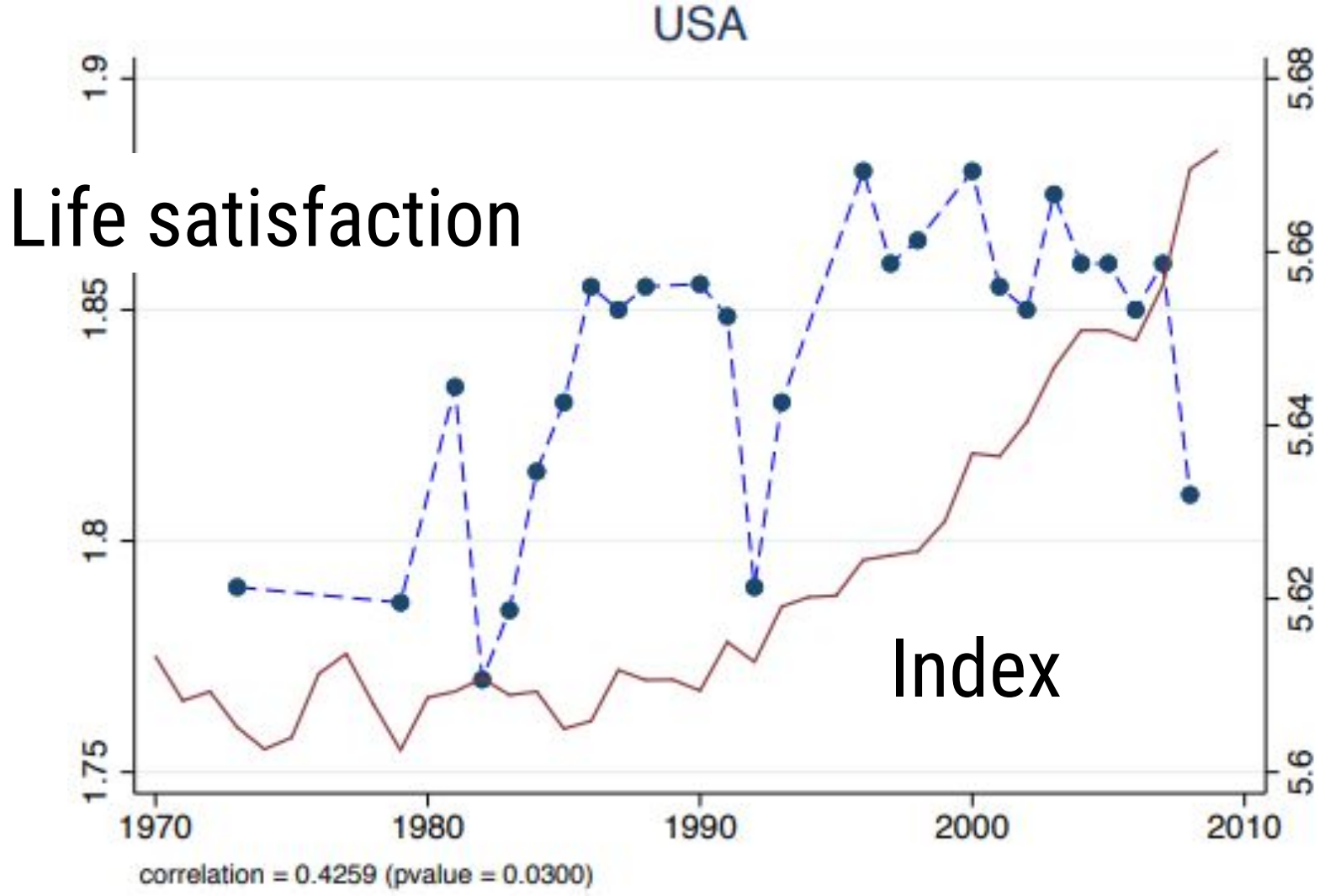
Validation

Eurobarometer survey (3 countries, 1973-2009)

How satisfied with your live?


0.05
0
−0.05
−0.10
NVI
−0.3
−0.2
−0.1
0
0.1
0.2
Life satisfaction
R = 0.53


5.70
5.64
5.58
NVI
United Kingdom
1825
1850
1875
1900
1925
1950
1975
2000
Year
5.68
5.64
5.60
NVI
United States
1825
1850
1875
1900
1925
1950
1975
2000
Year
5.95
5.80
5.65
NVI
Germany
1825
1850
1875
1900
1925
1950
1975
2000
Year
6.05
5.90
5.75
NVI
Italy
1825
1850
1875
1900
1925
1950
1975
2000
Year


USA
Life satisfaction
Index
1.9
1.85
1.8
1.75
5.68
5.66
5.64
5.62
5.6
1970
1980
1990
2000
2010
correlation = 0.4259 (pvalue = 0.0300)

Controls

Table 2 | Historical determinants of the NVI from 1820 to 2009

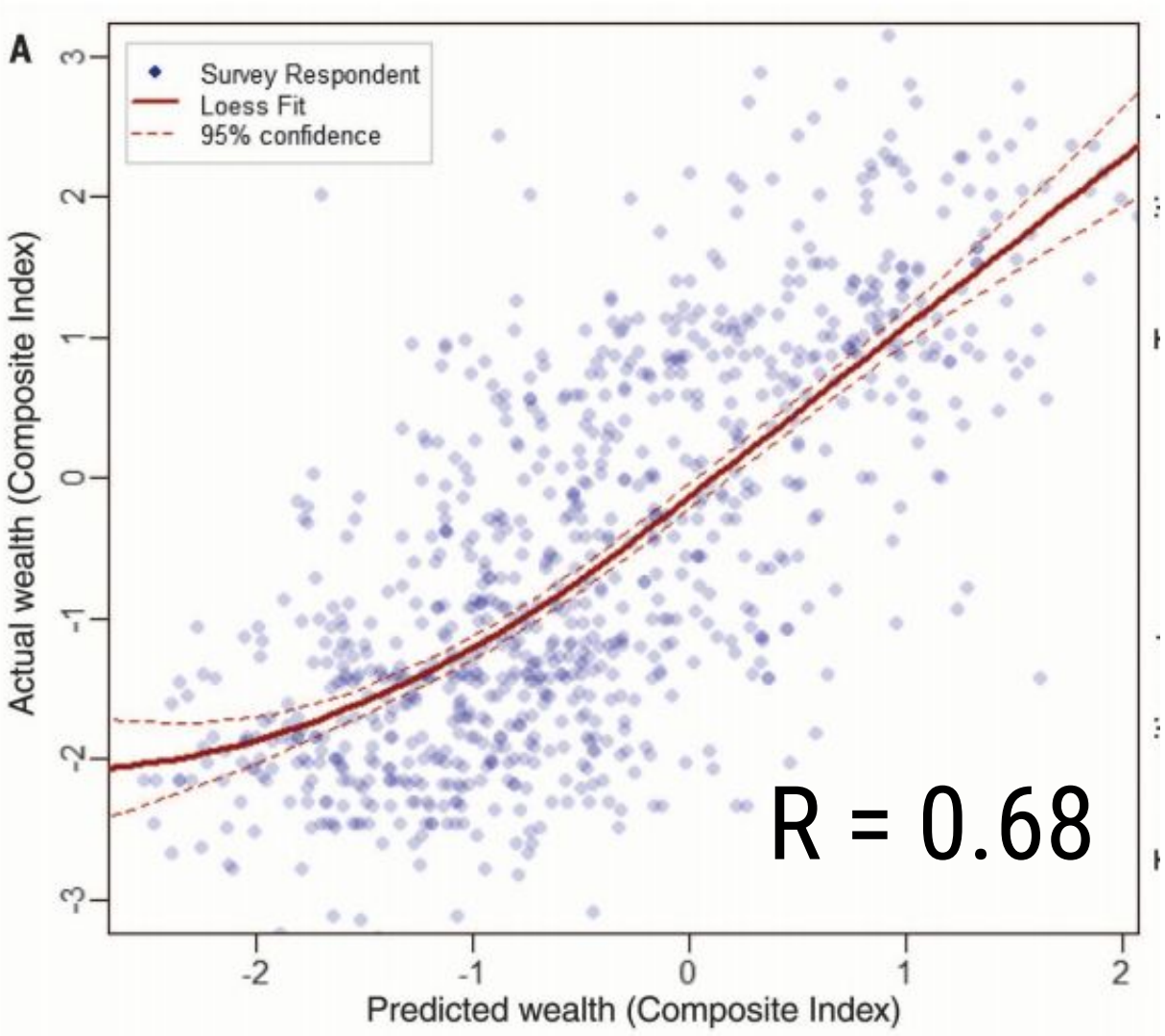
	Year fixed effects	Year fixed effects	Year fixed effects	Country-specific trends
log[GDP($t-5$)] (β (s.e.))	0.0826*** (0.0090)		0.0698*** (0.0106)	0.0550** (0.0130)
Life expectancy ($t-1$) (β (s.e.))		0.0048** (0.0013)	0.0030 (0.0014)	0.0016 (0.0013)
Internal conflict ($t-1$) (β (s.e.))				−0.0184** (0.0040)
Words covered (t)	Yes	Yes	Yes	Yes
Democracy (t)	Yes	Yes	Yes	Yes
Educational inequality (t)	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	No
Country-specific trends	No	No	No	Yes
r^2	0.752	0.705	0.774	0.571
n	412	412	412	412

One extra year of life expectancy is worth as much as 4.3% annual growth in GDP per capita

One fewer year of internal conflict is worth as much as 30% annual growth in GDP per capita

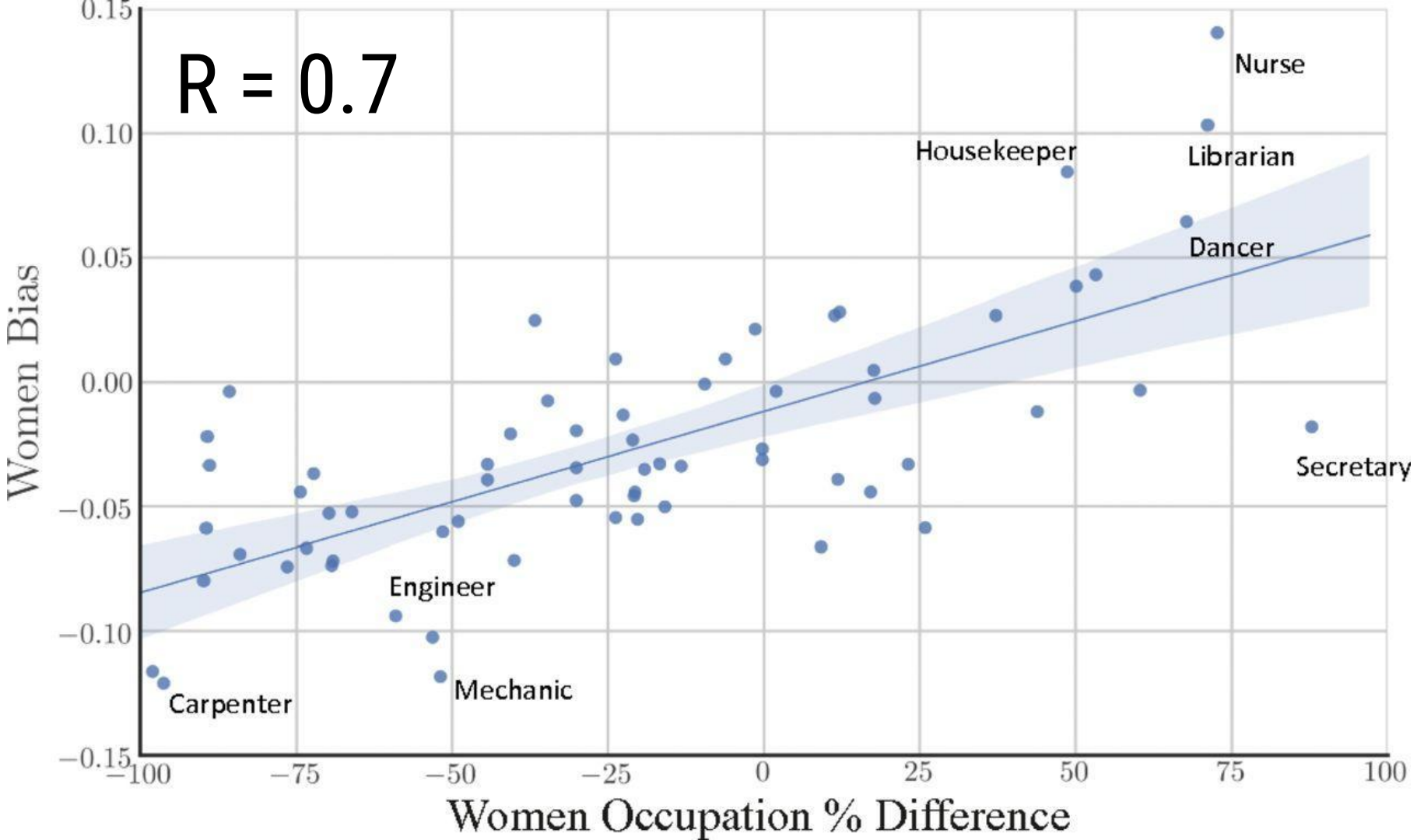
Predicting poverty and wealth from mobile phone metadata

Blumenstock J, Cadamuro G, On R 2015 // *Science*


A
Survey Respondent
Loess Fit
95% confidence
Actual wealth (Composite Index)
Predicted wealth (Composite Index)
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
R = 0.68

Word embeddings quantify 100 years of gender and ethnic stereotypes

Garg N, Schiebinger L, Jurafsky D, Zou J 2018 // *PNAS*


R = 0.7
Nurse
Librarian
Housekeeper
Dancer
Secretary
Engineer
Mechanic
Carpenter
Women Bias
Women Occupation % Difference
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−100
−75
−50
−25
0
25
50
75
100

For some reason it is okay to have R = 0.5-0.7 for validating that X ~ Y

e.g. sentiment ~ well-being

This is probably wrong

Conceptually: being a man = being a prisoner

This is probably wrong

Conceptually: being a man = being a prisoner

Practically: correlation is not transitive

The effects of errors of measurement on correlation coefficients

Thouless RH. 1939 // *British Journal of Psychology*

Corruption of the Pearson correlation coefficient by measurement error and its estimation, bias, and correction under different error models

Saccenti E, Hendriks MH, Smilde AK 2020 // *Scientific reports*

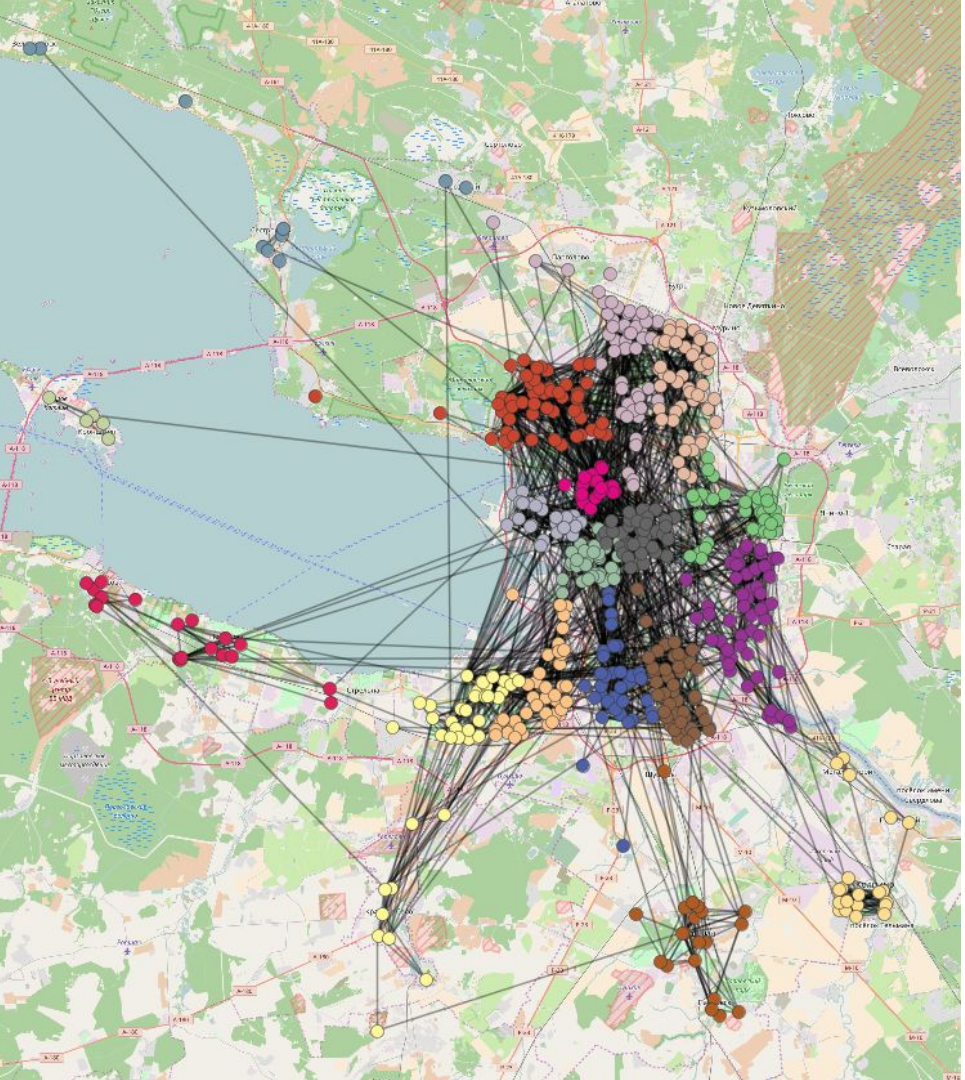
The proof and measurement of association between two things

Spearman C 1904 // *The American Journal of Psychology*

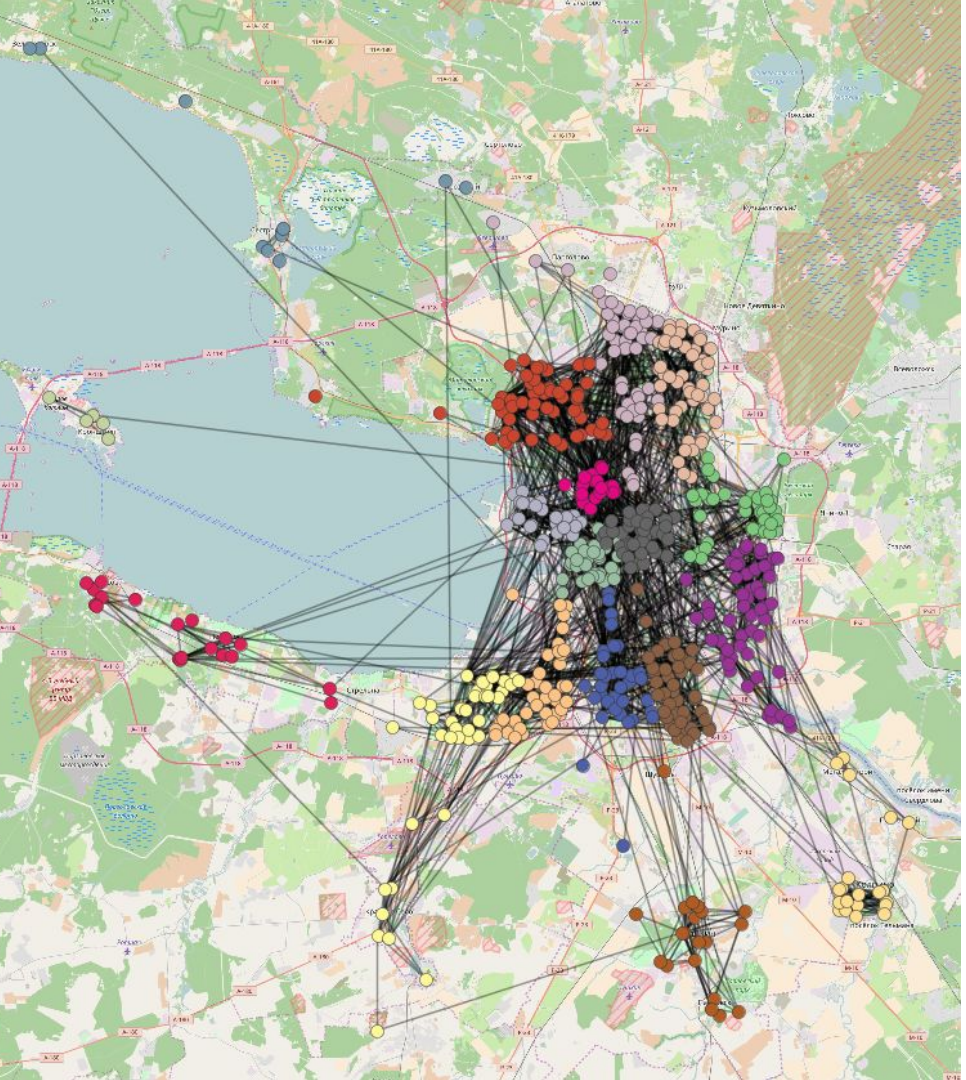
Known but ignored

if error structure is simple it is not problematic

complex error structure is not convincing



Real-world case



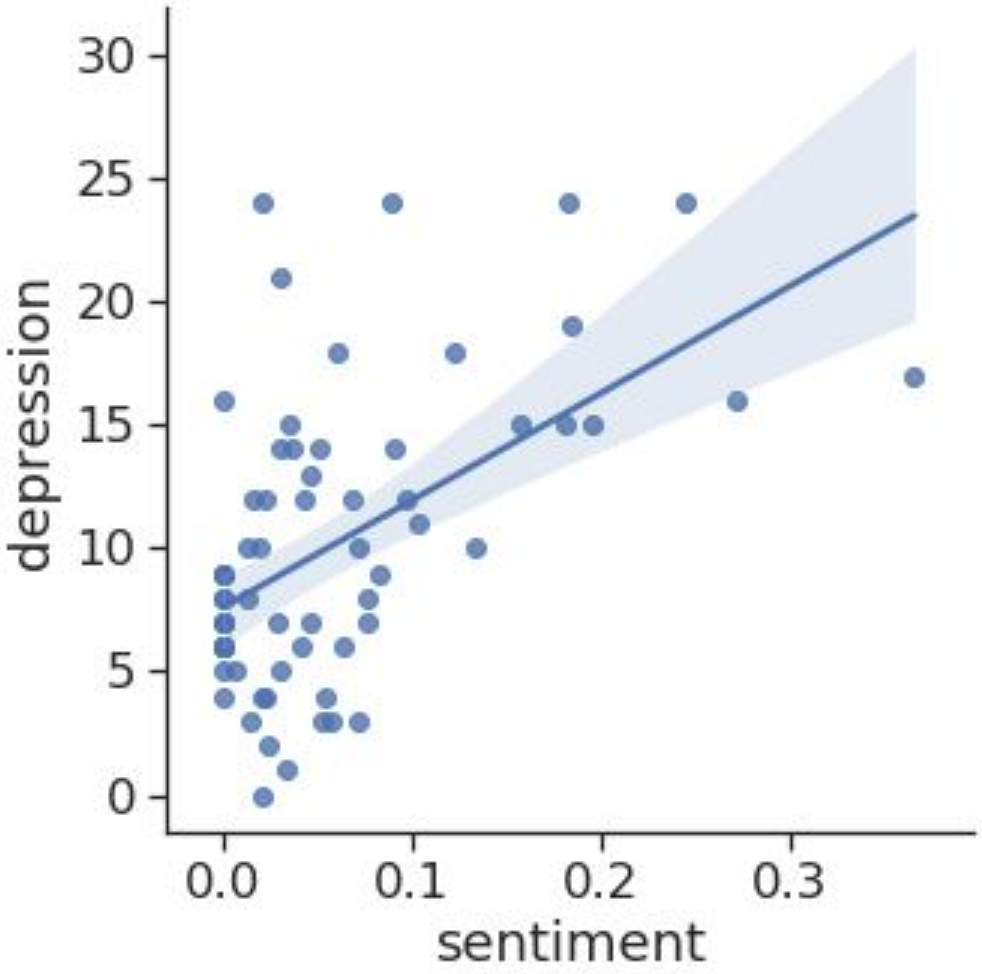
Real-world case

Depression level at schools

Depression ~ academic performance

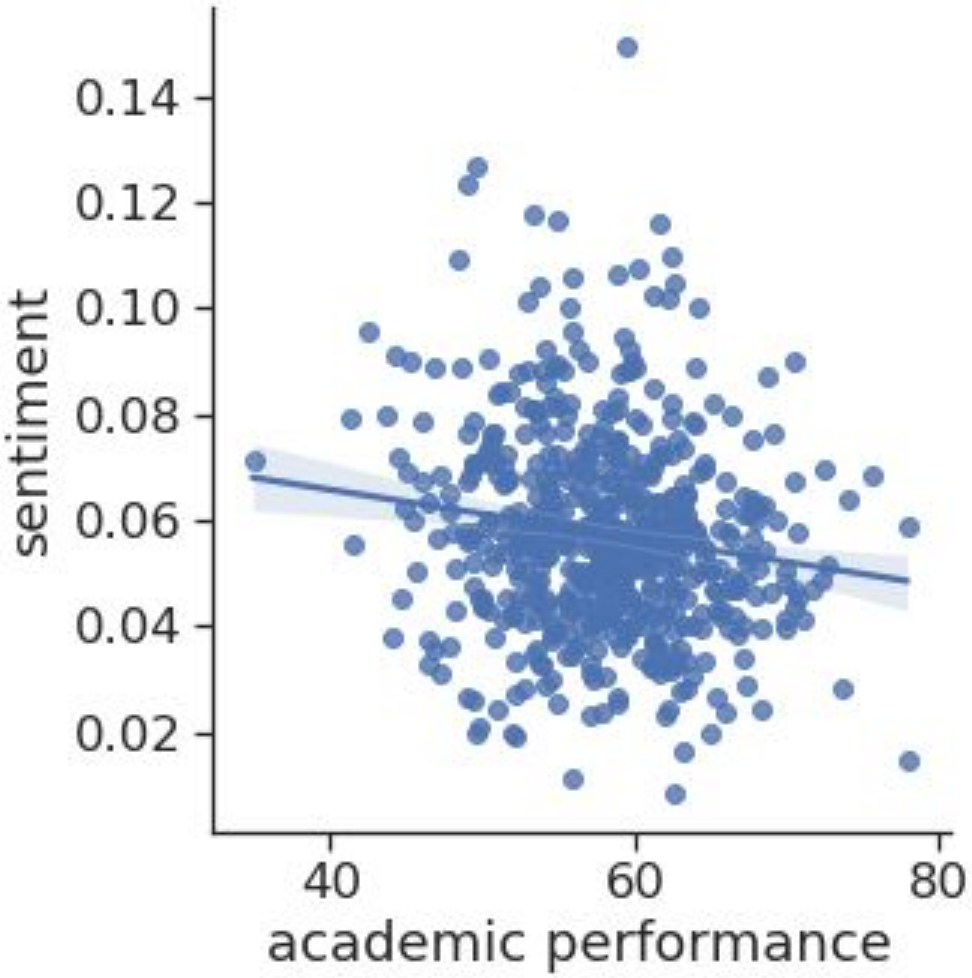
Let’s use sentiment of post on VK

Validation: PHQ-8



R = 0.54

Let’s apply the model

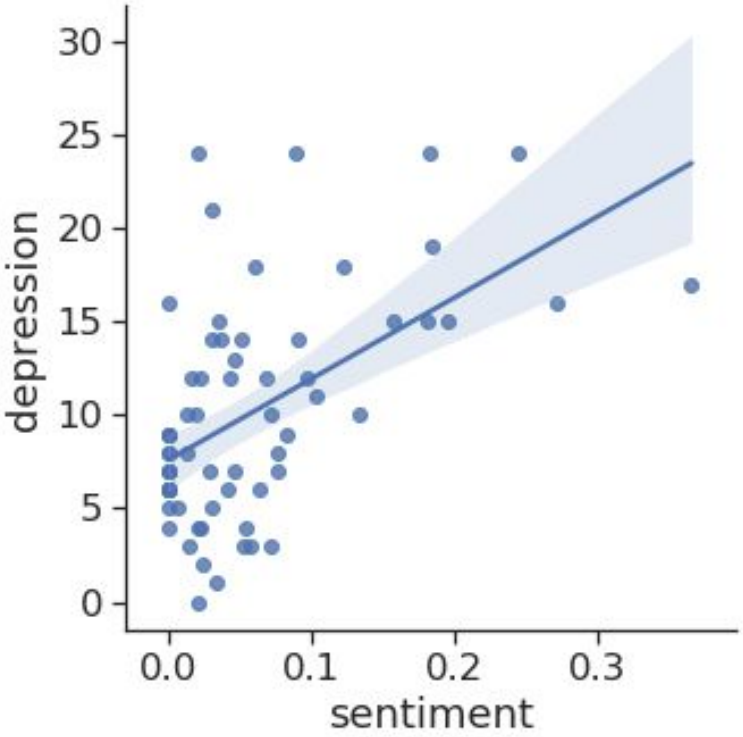

0.14
0.12
0.10
0.08
0.06
0.04
0.02
sentiment
40
60
80
academic performance

Students in higher performing schools are less depressed

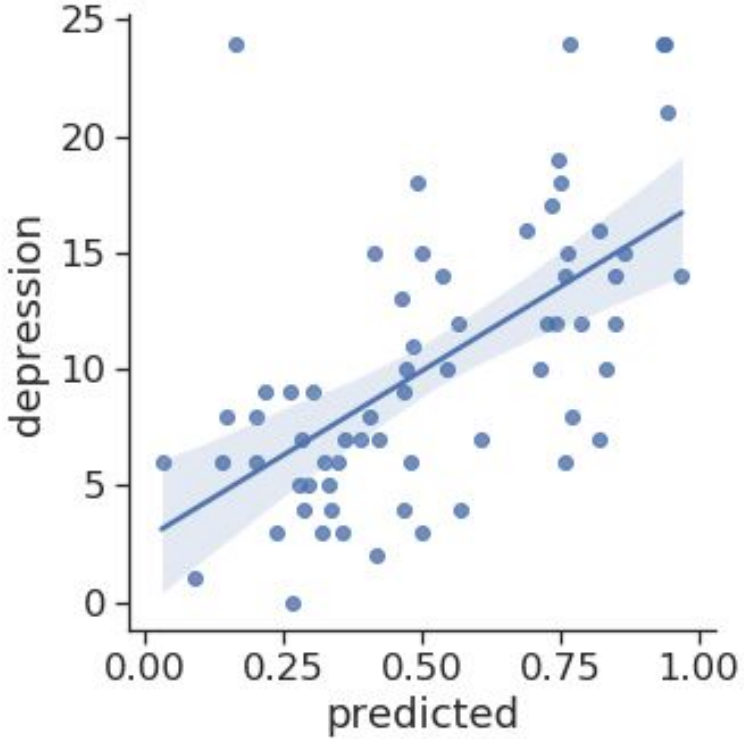
R = -0.15, P = 0.0005

Let’s build a model to predict depression

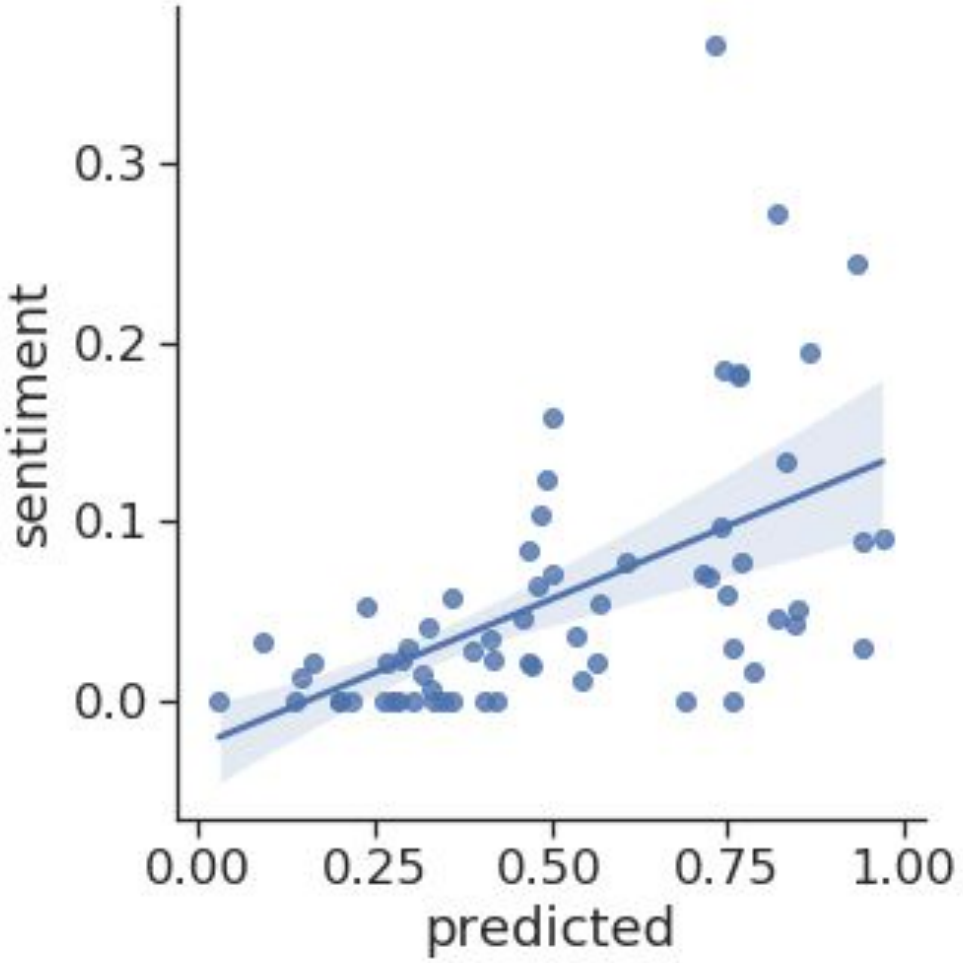
Instead of sentiment


depression
sentiment
0
5
10
15
20
25
30
0.0
0.1
0.2
0.3

R = 0.54

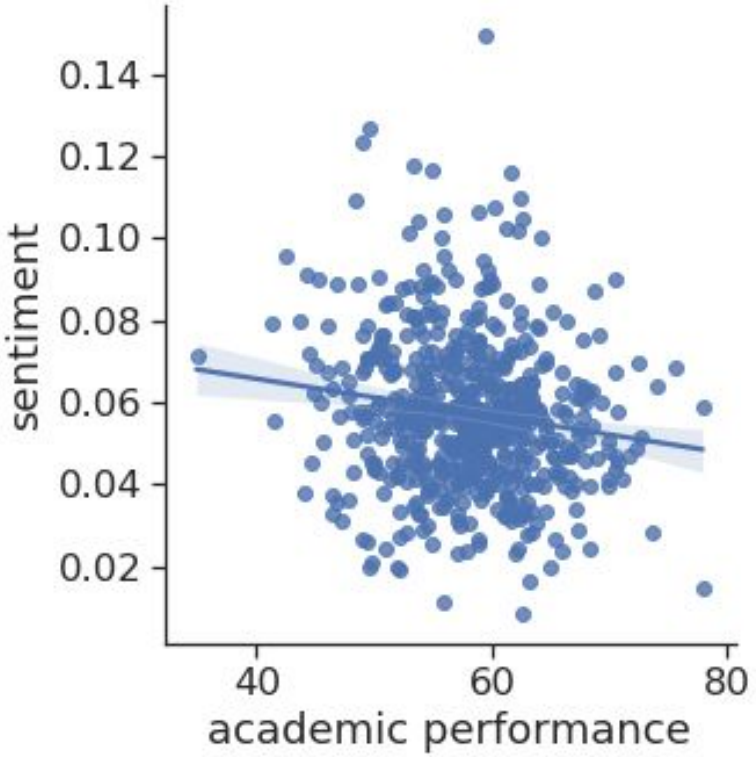

depression
predicted
0
5
10
15
20
25
0.00
0.25
0.50
0.75
1.00

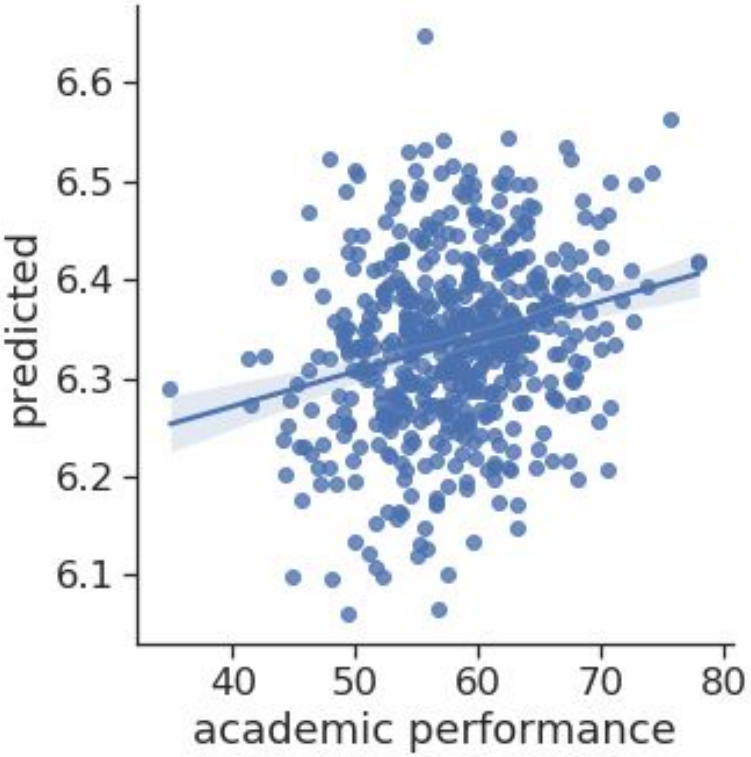
R = 0.60



R = 0.56

Let’s apply the model again to confirm our findings regarding academic performance





$R = -0.15, P = 0.0005$

$R = 0.25, P < 10^{-8}$

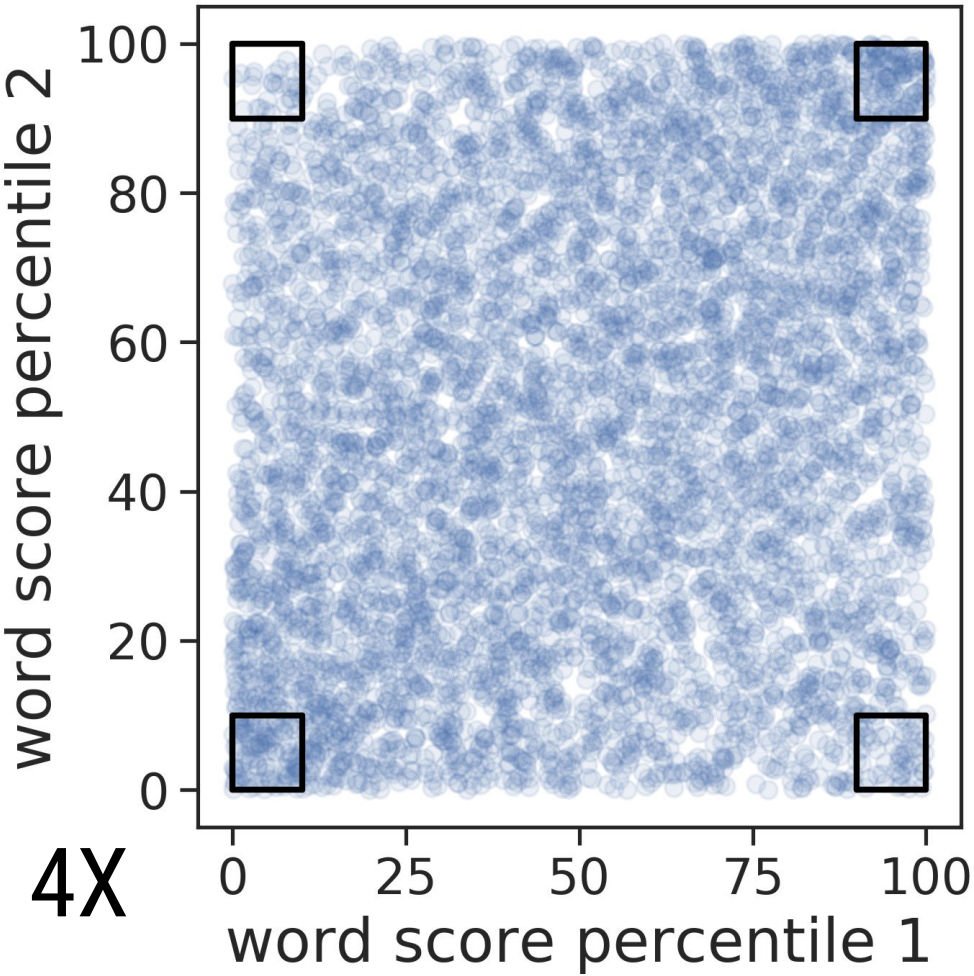
Proxies are substantially different

Proxies are substantially different

Let’s explore by computing word scores

Not-depressive negative

X



4X

depressive negative

4X

2X

Not-depressive positive

depressive positive

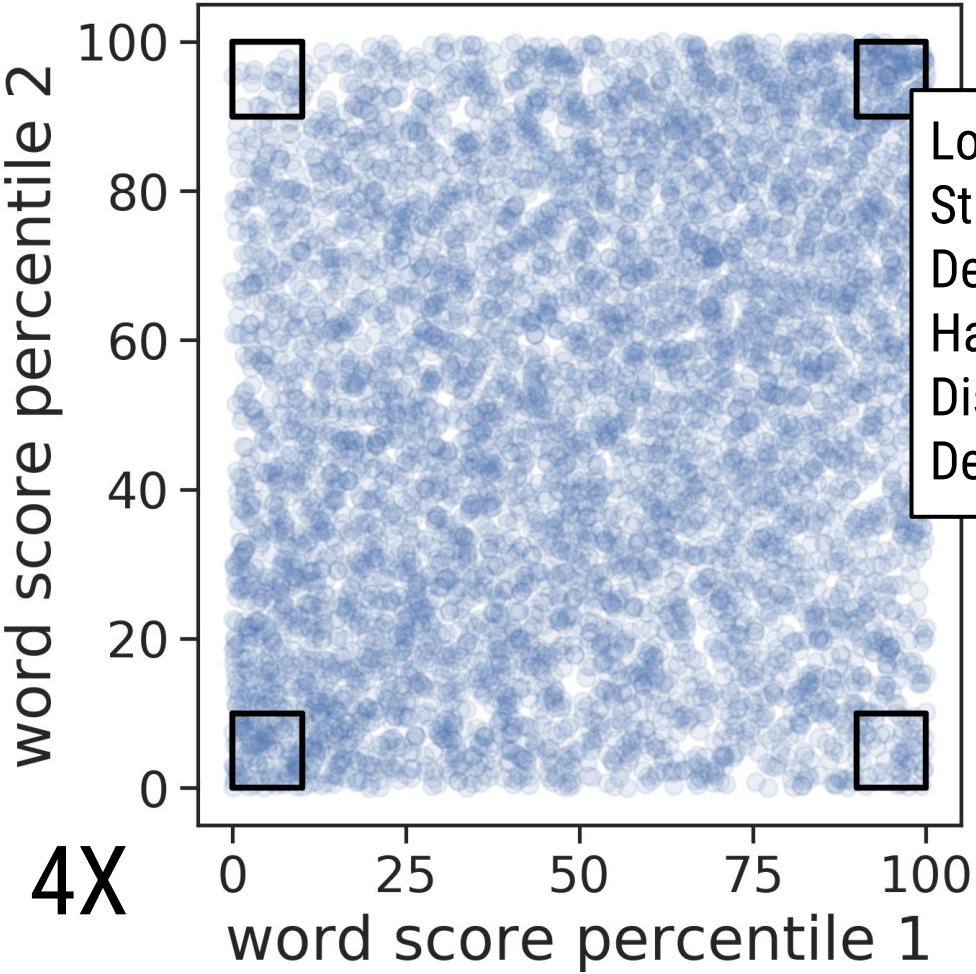
Not-depressive negative

X

depressive negative

4X

Loneliness
Stupid
Depression
Hate
Disappointment
Death



4X

2X

Not-depressive positive

depressive positive

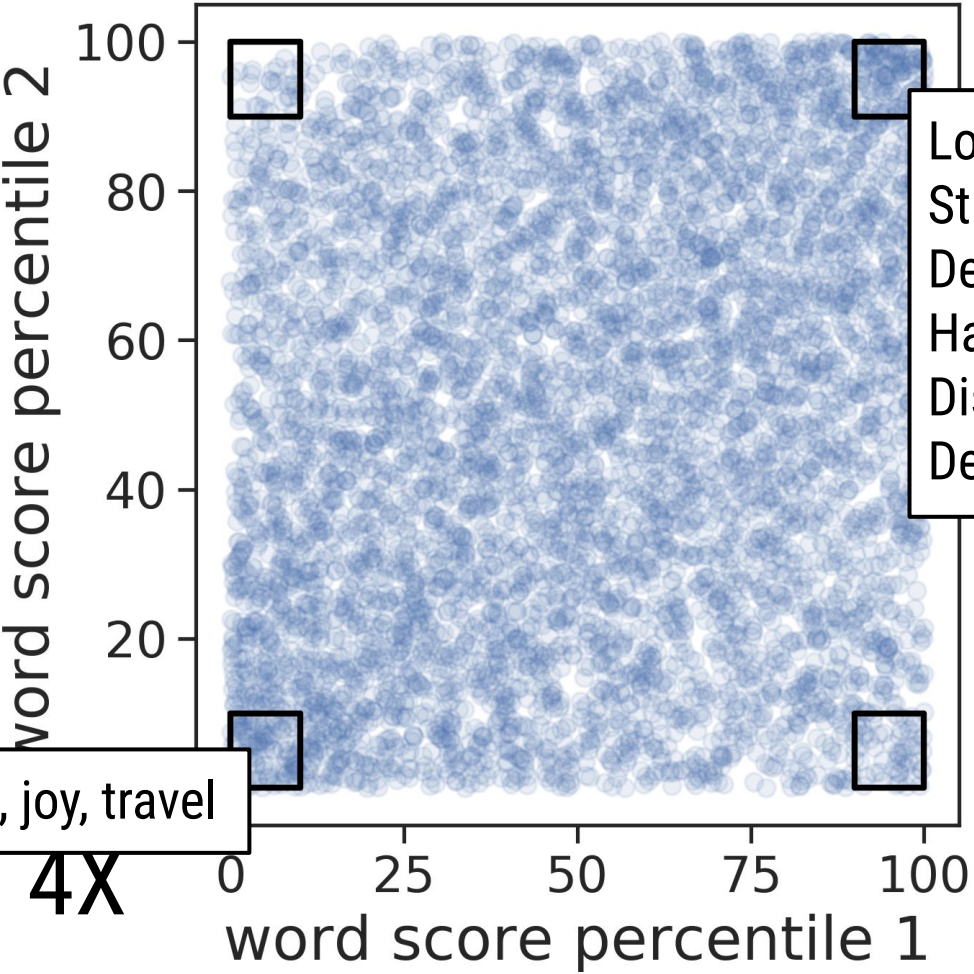
Not-depressive negative

X

depressive negative

4X

Loneliness
Stupid
Depression
Hate
Disappointment
Death



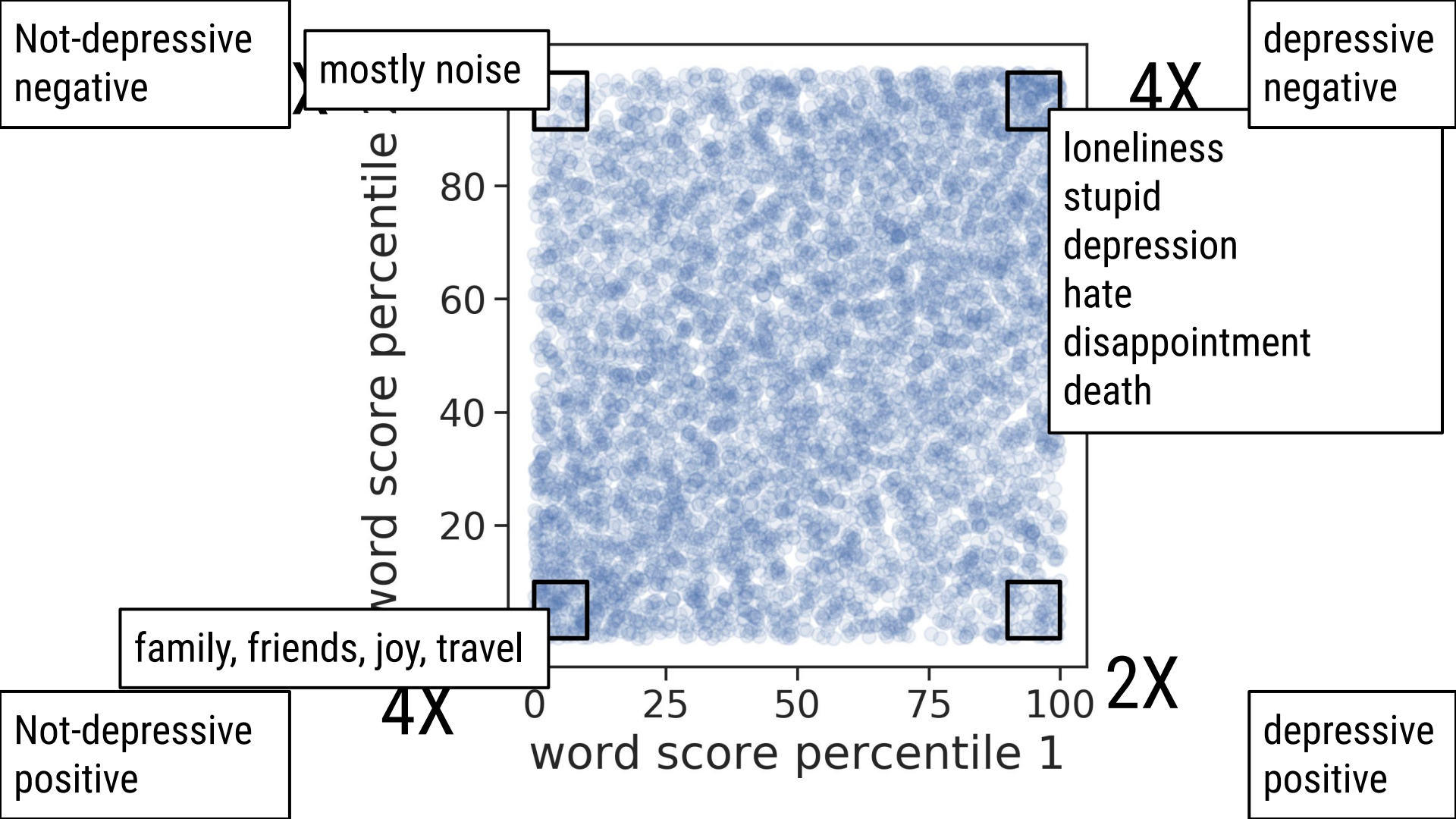
family, friends, joy, travel

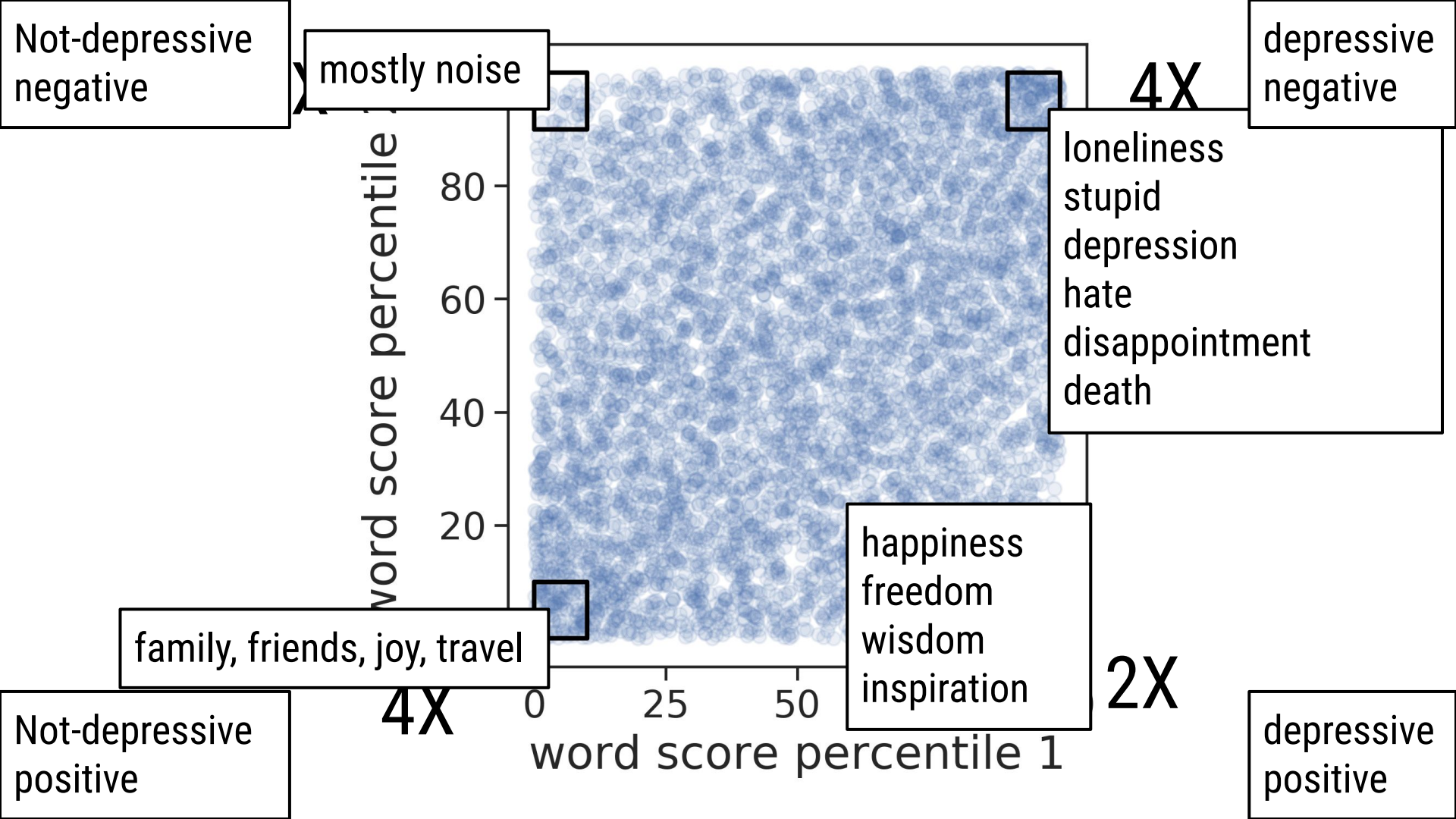
4X

2X

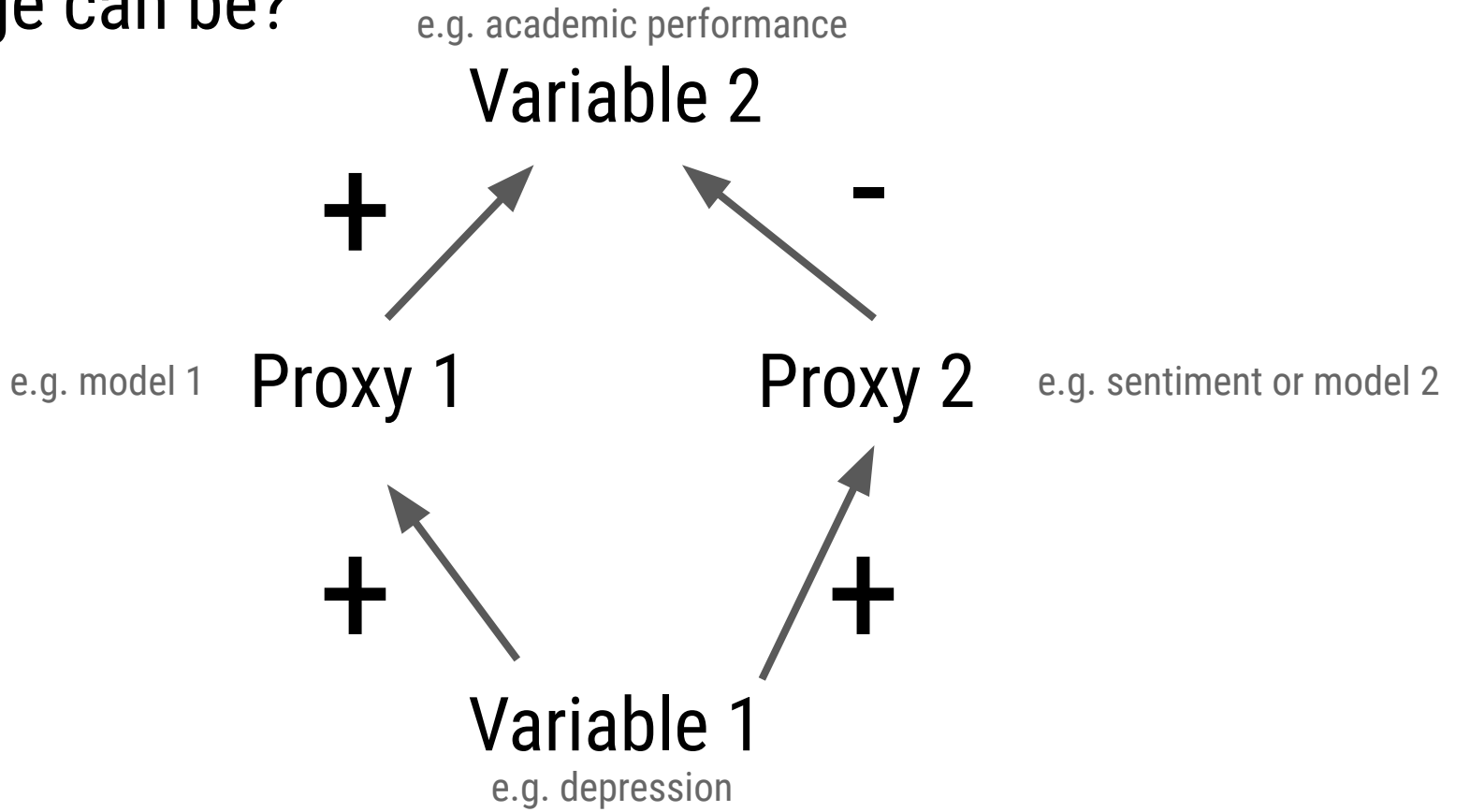
Not-depressive positive

depressive positive

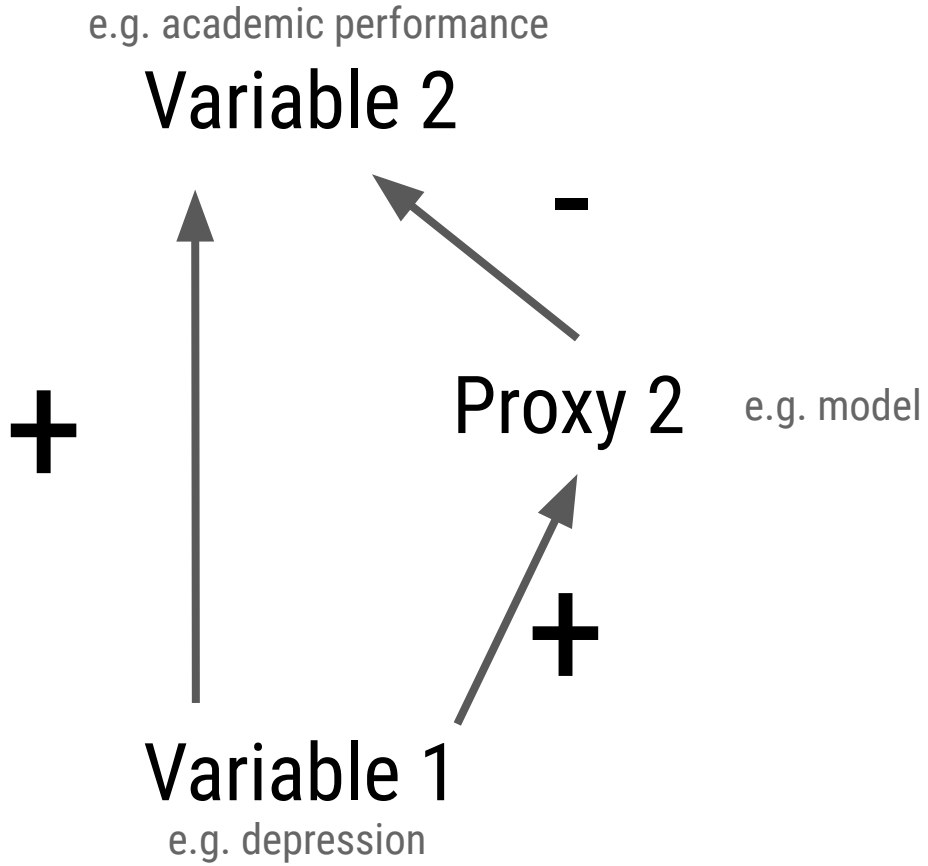

Not-depressive negative
mostly noise
depressive negative
4X
loneliness
stupid
depression
hate
disappointment
death
80
60
40
20
word score percentile
family, friends, joy, travel
4X
2X
0
25
50
75
100
word score percentile 1
Not-depressive positive
depressive positive


Not-depressive negative
mostly noise
depressive negative
4X
loneliness
stupid
depression
hate
disappointment
death
80
60
40
20
happiness
freedom
wisdom
inspiration
family, friends, joy, travel
4X
2X
0
25
50
word score percentile 1
Not-depressive positive
depressive positive

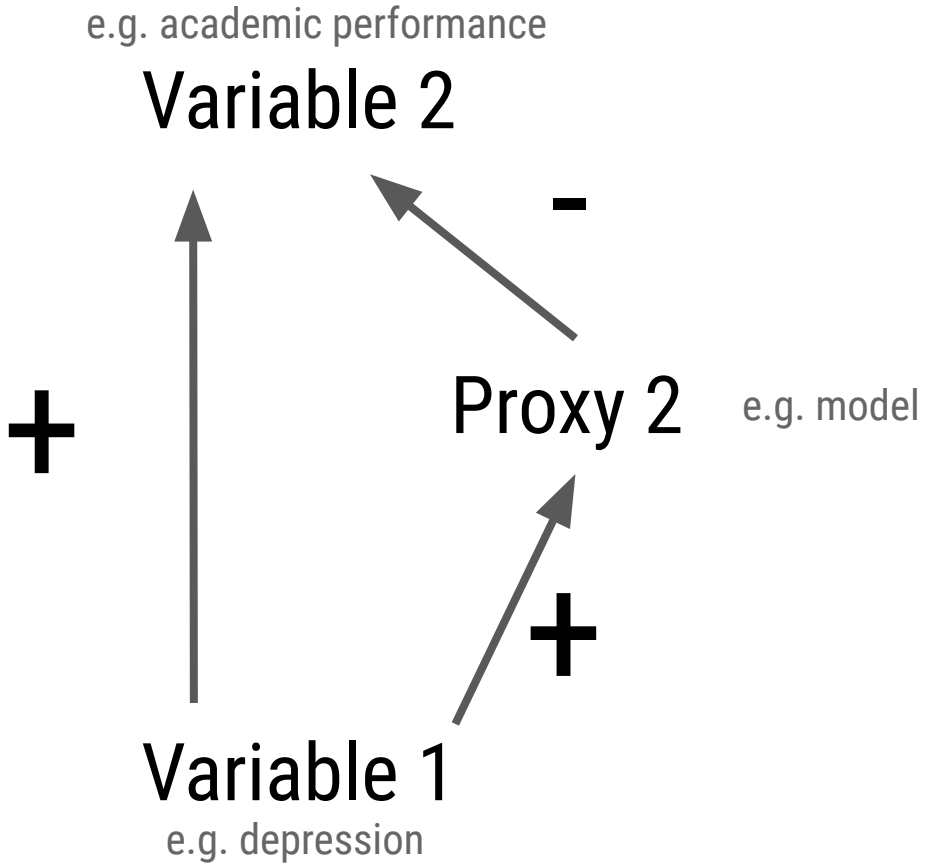
How large can be?



How large can be?



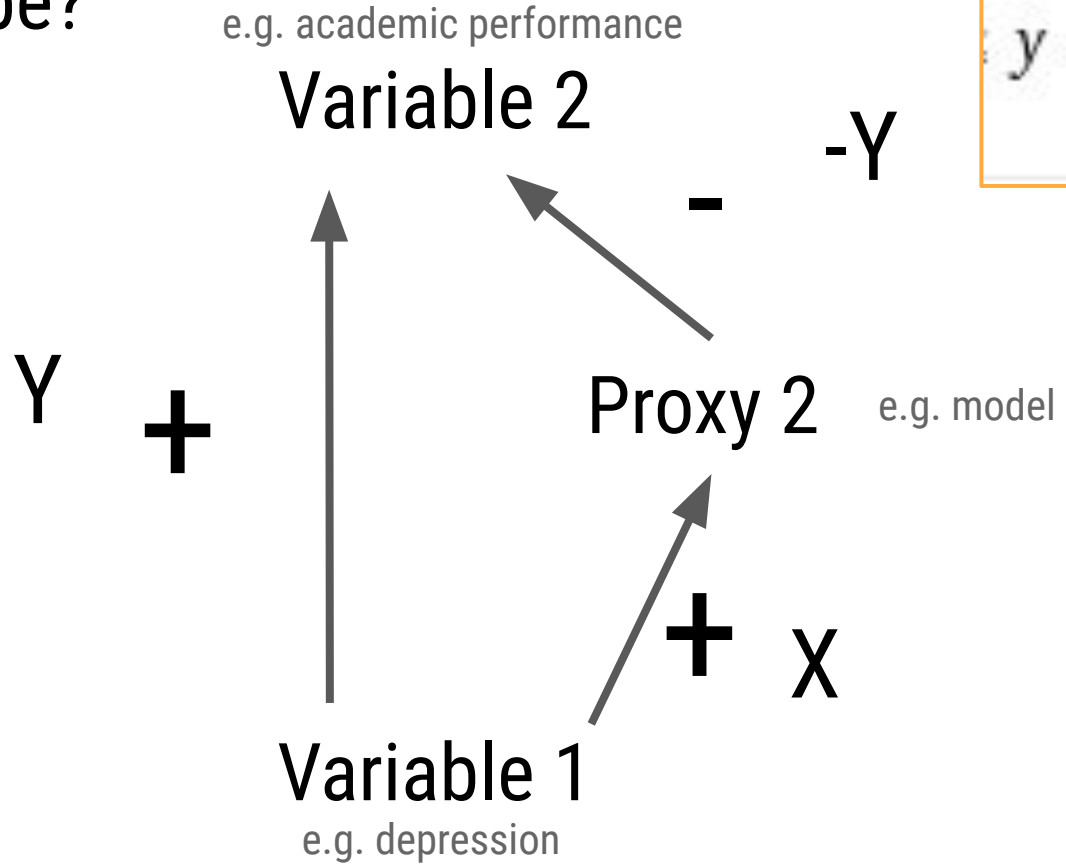
How large can be?



$$1 + 2\rho\sigma\tau - \left(\rho^2 + \sigma^2 + \tau^2\right) \geq 0$$

How large can be?

$$y \leq \frac{\sqrt{1-x}}{\sqrt{2}}$$



$$1 + 2\rho\sigma\tau - \left(\rho^2 + \sigma^2 + \tau^2\right) \geq 0$$

If proxy is 0.5 one can have -0.50; 0.50 effects

If proxy is 0.5 one can have -0.50; 0.50 effects

If proxy is 0.8 one can have -0.31; 0.31 effects

If proxy is 0.5 one can have -0.50; 0.50 effects

If proxy is 0.8 one can have -0.31; 0.31 effects

If proxy is 0.9 one can have -0.22; 0.22 effects

Ivan Smirnov

ivan@ismirnov.eu

#computational social science #Fair AI
#education #inequality #social networks
#emotional well-being #gender #ML

Bonus: AI

Criminal machine learning

Calling Bullshit. Data Reasoning in a Digital World

https://www.callingbullshit.org/case_studies/case_study_criminal_machine_learning.html

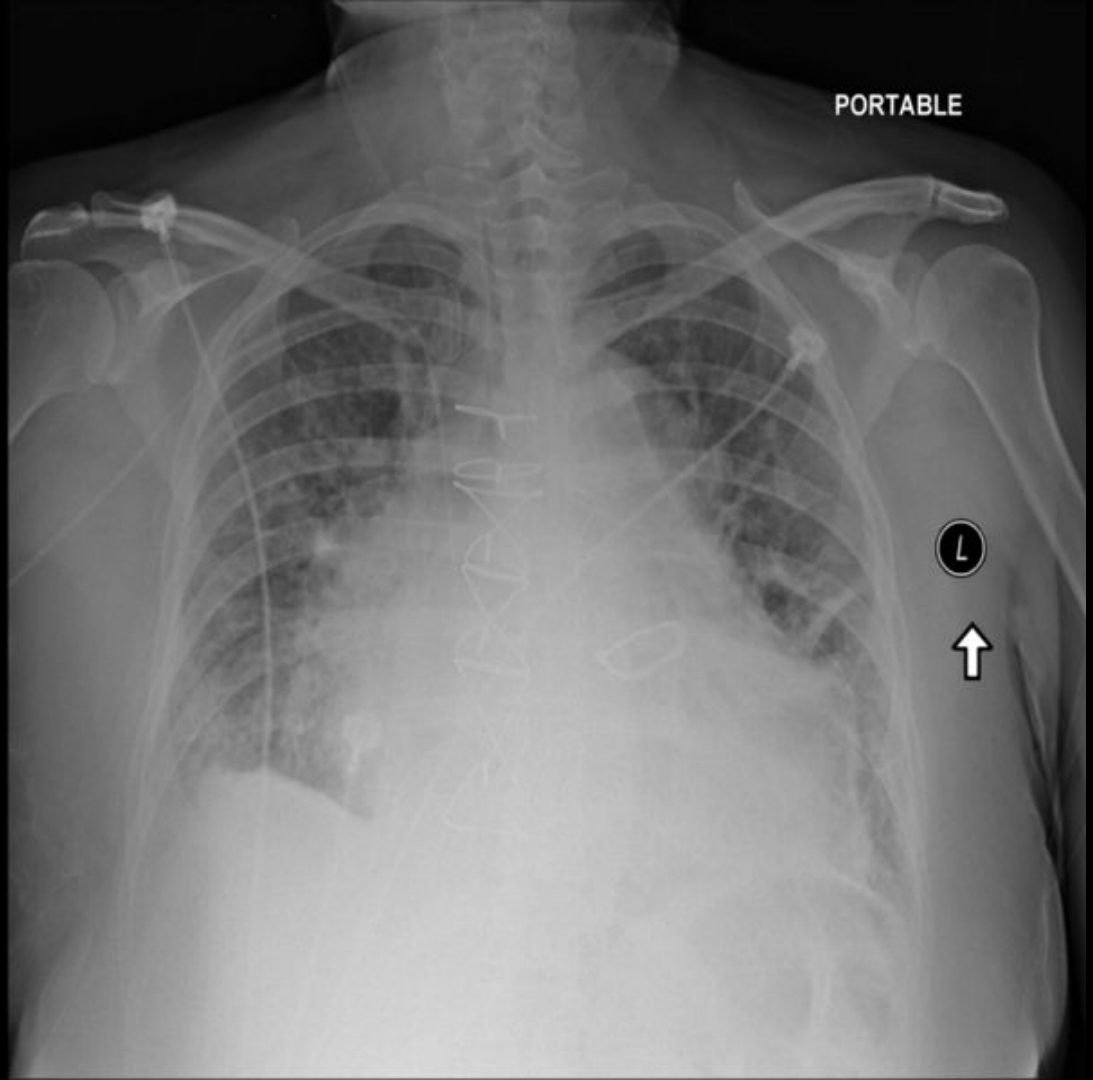
(a) Three samples in criminal ID photo set S_c.

(b) Three samples in non-criminal ID photo set S_n

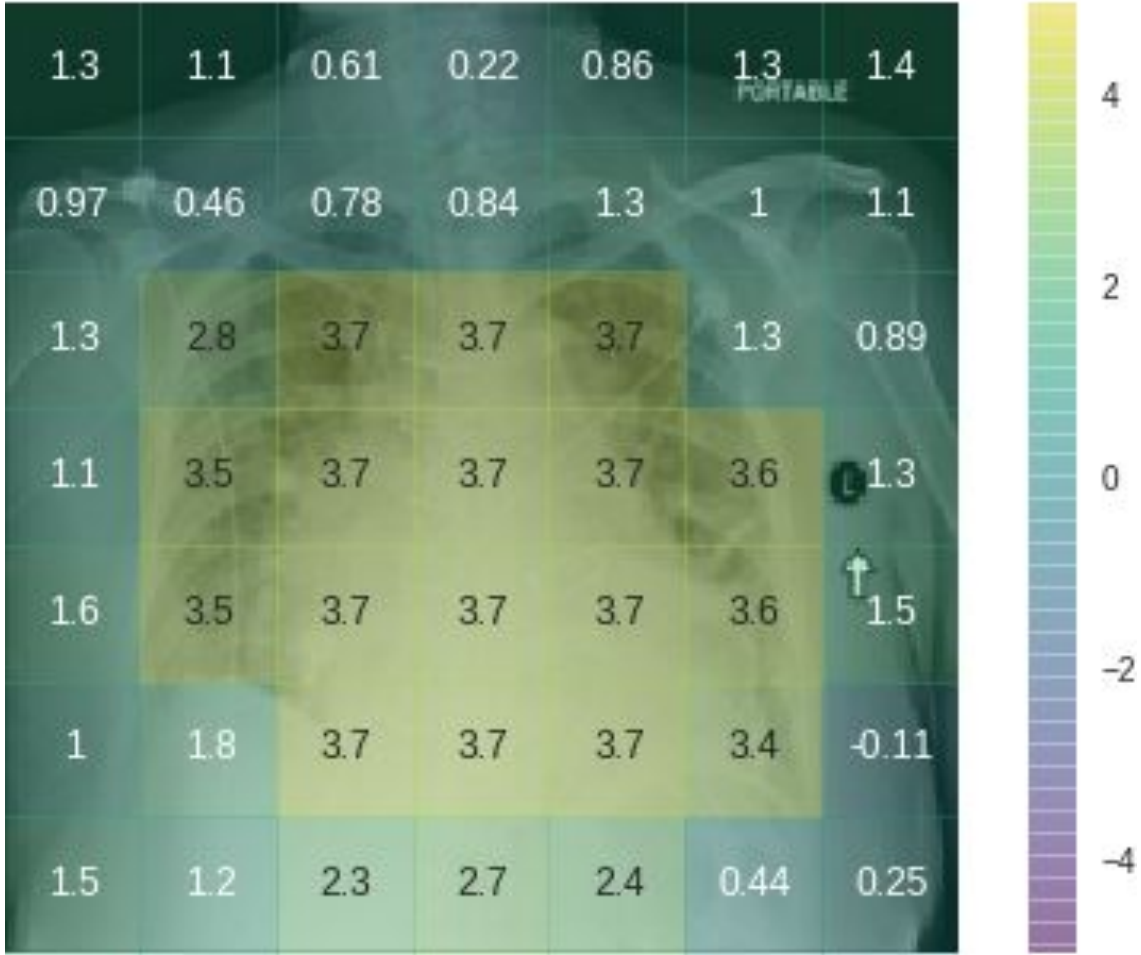
What are radiological deep learning models actually learning?

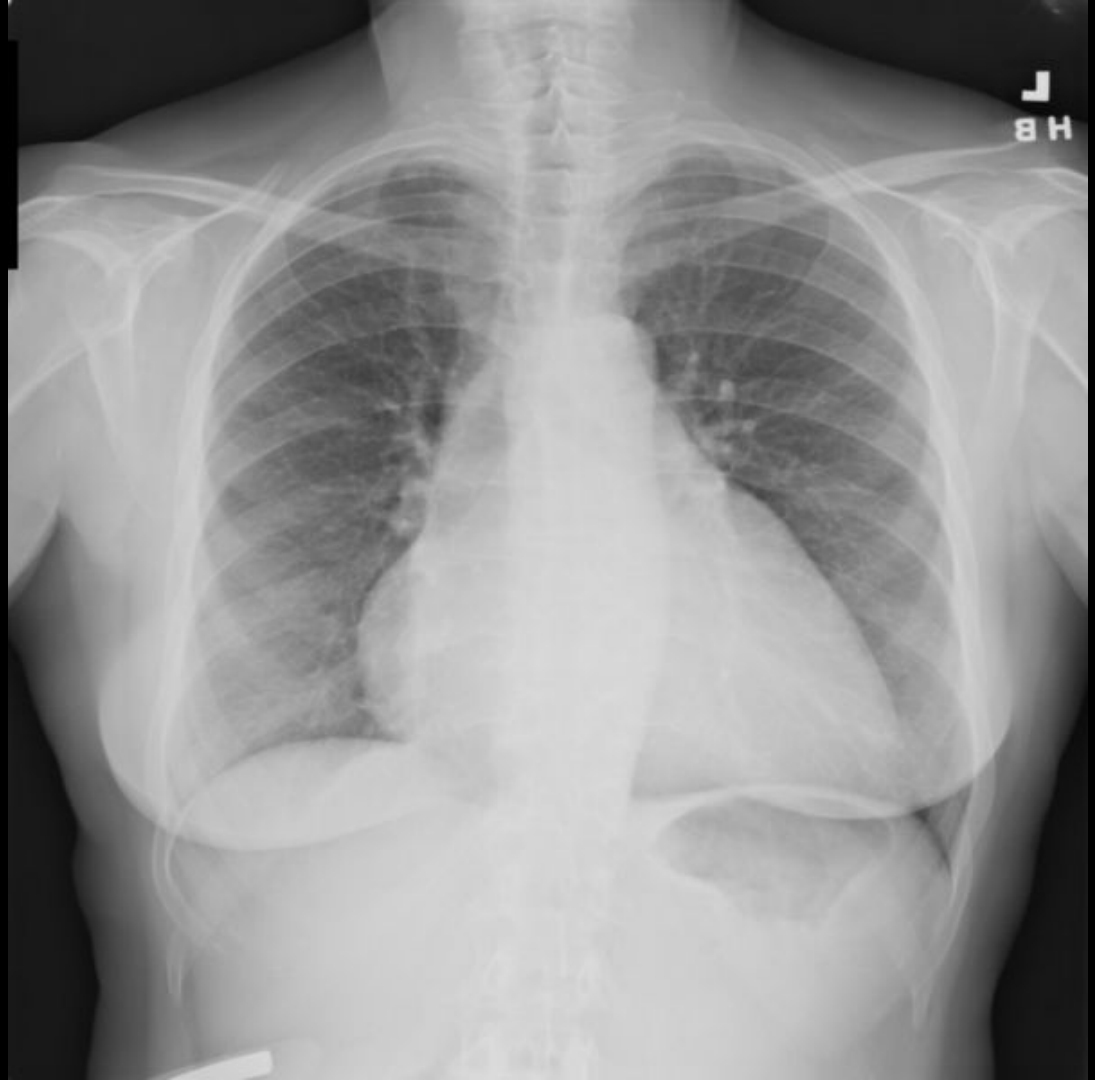
John Zech

https://jrzech.medium.com/what-are-radiological-deep-learning-models-actually-learning-f97a546c5b98

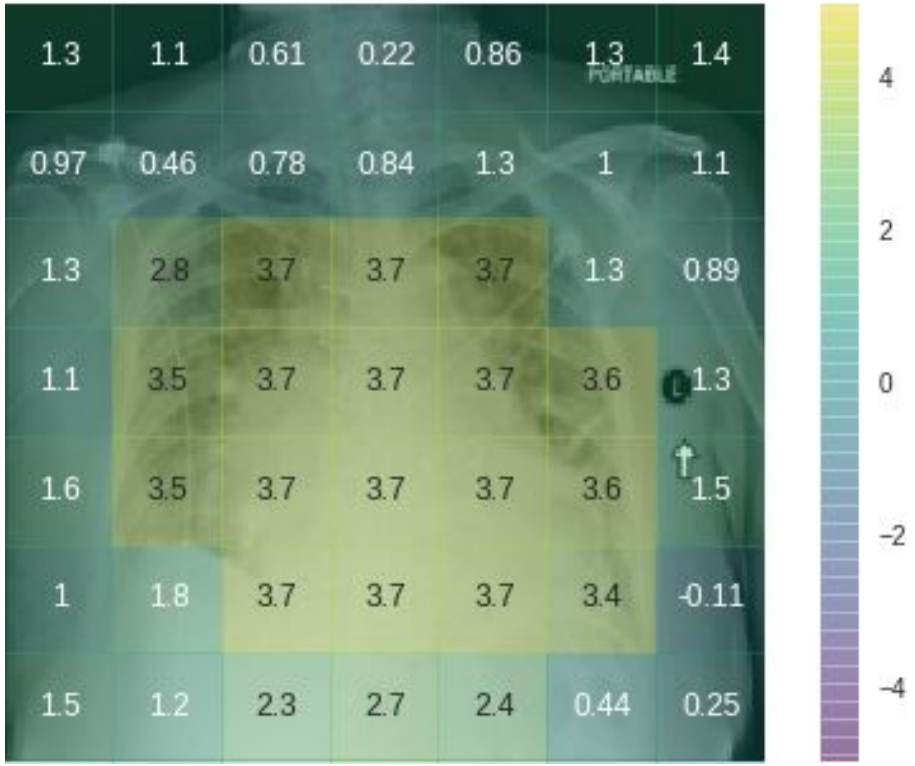

PORTABLE
L

P(Cardiomegaly)=0.752

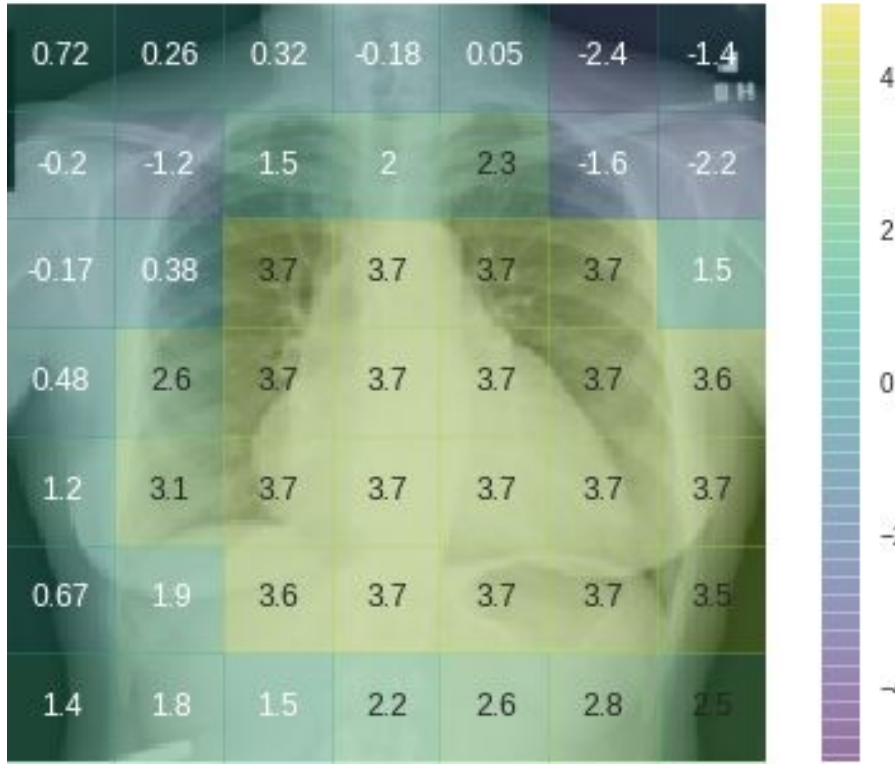



L
H B

P(Cardiomegaly)=0.752



P(Cardiomegaly)=0.937

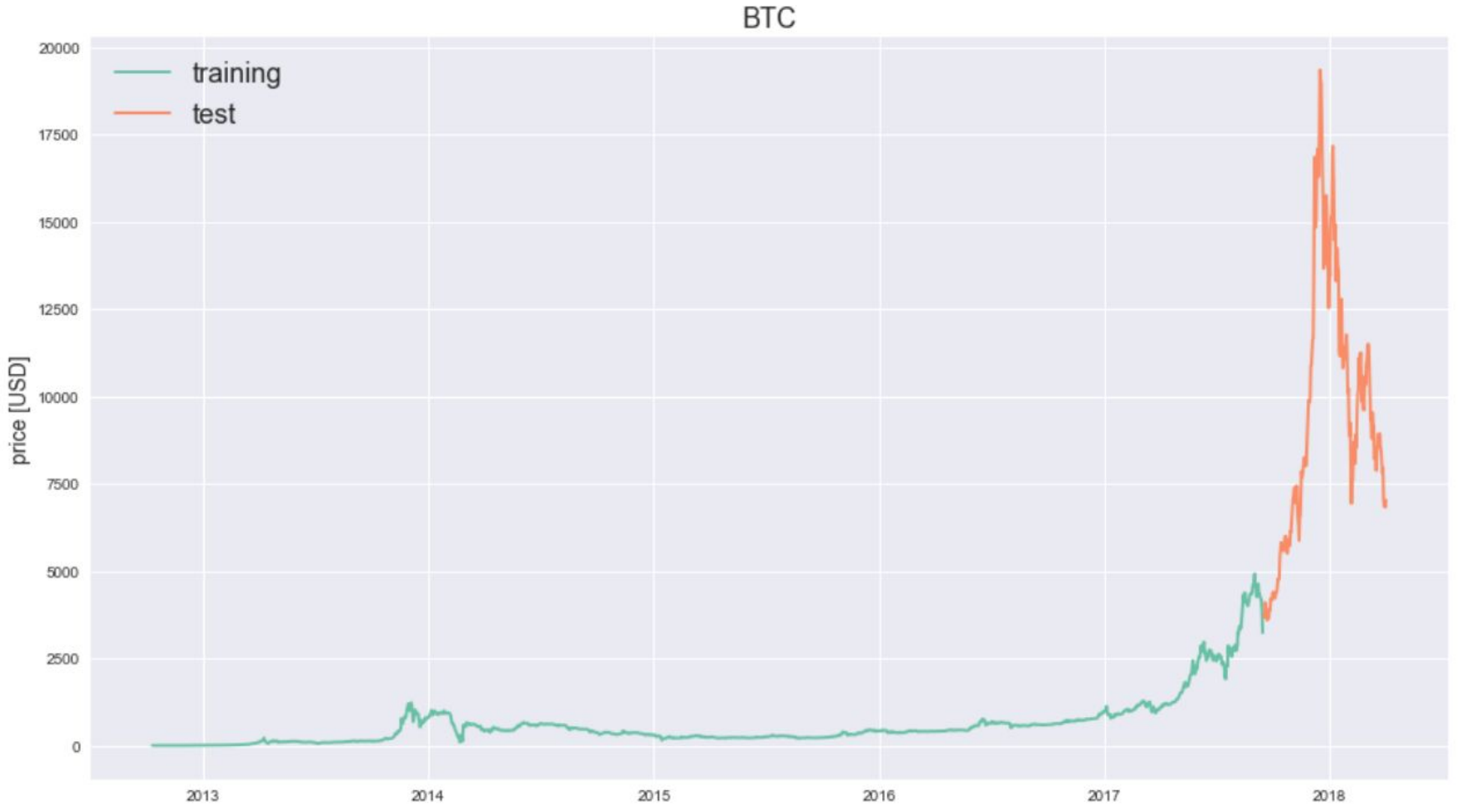


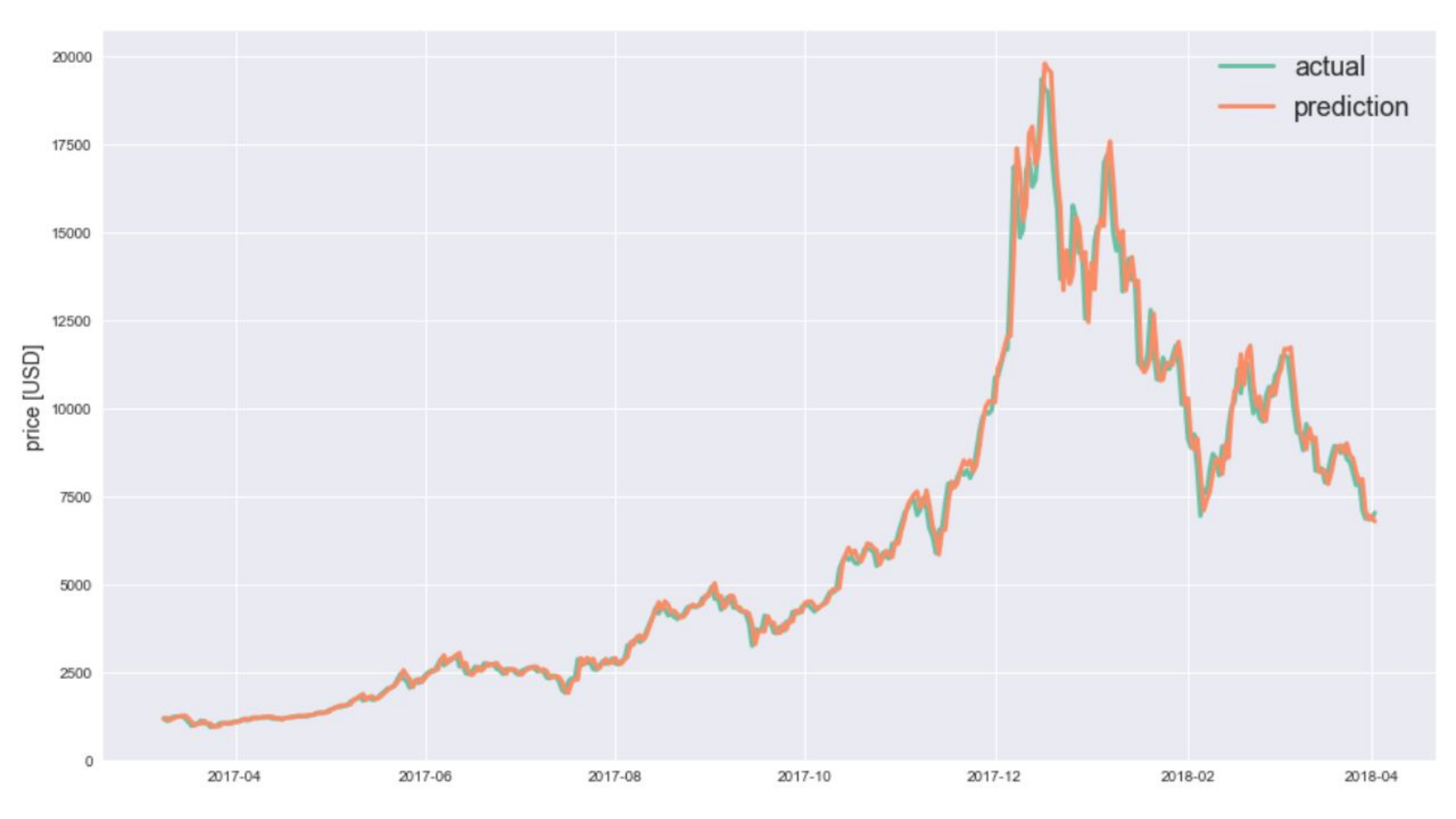
Don’t be fooled — Deceptive Cryptocurrency Price Predictions Using Deep Learning

Rafael Schultze-Kraft

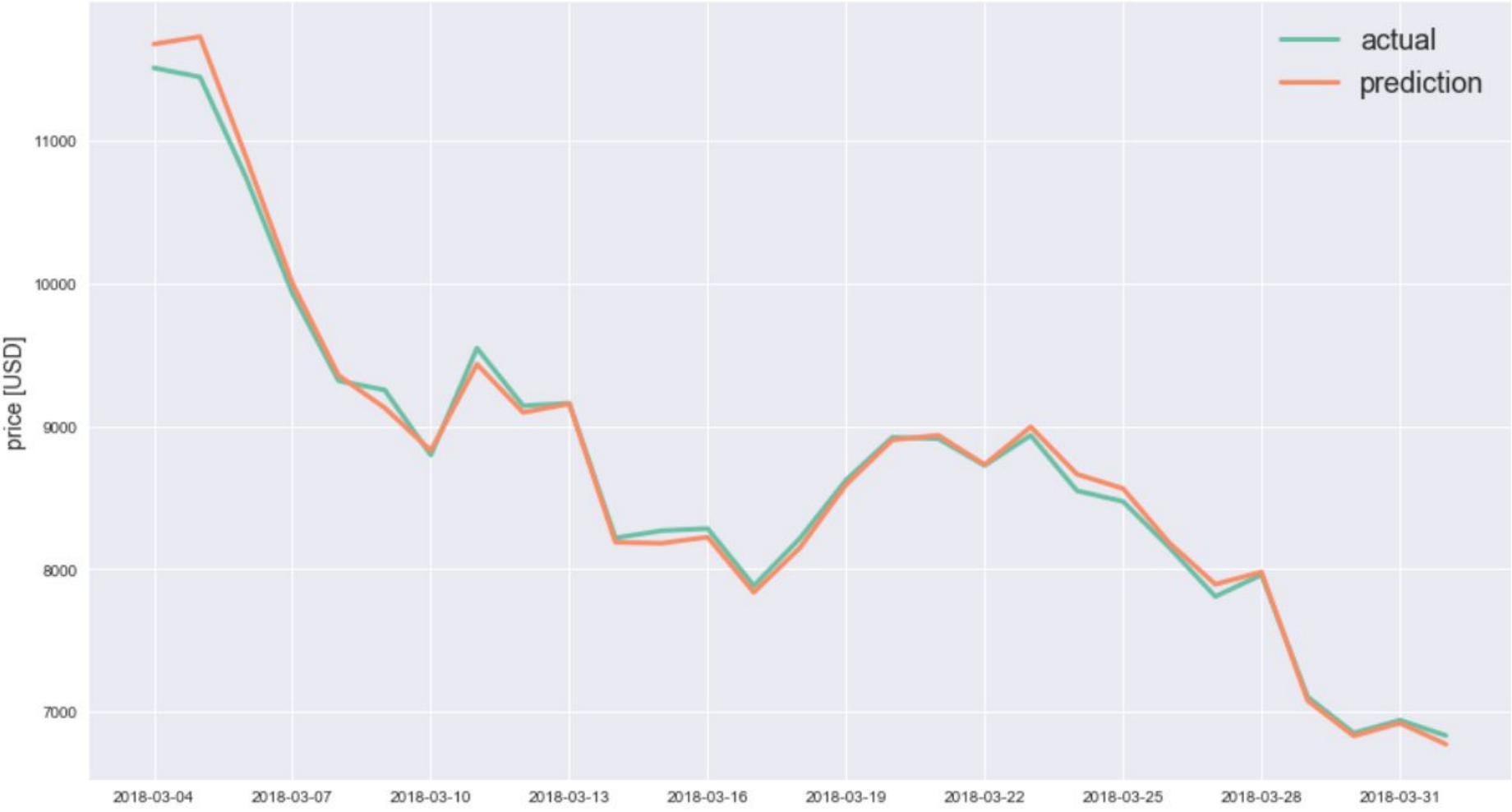
https://medium.com/hackernoon/dont-be-fooled-deceptive-cryptocurrency-price-predictions-using-deep-learning-bf27e4837151

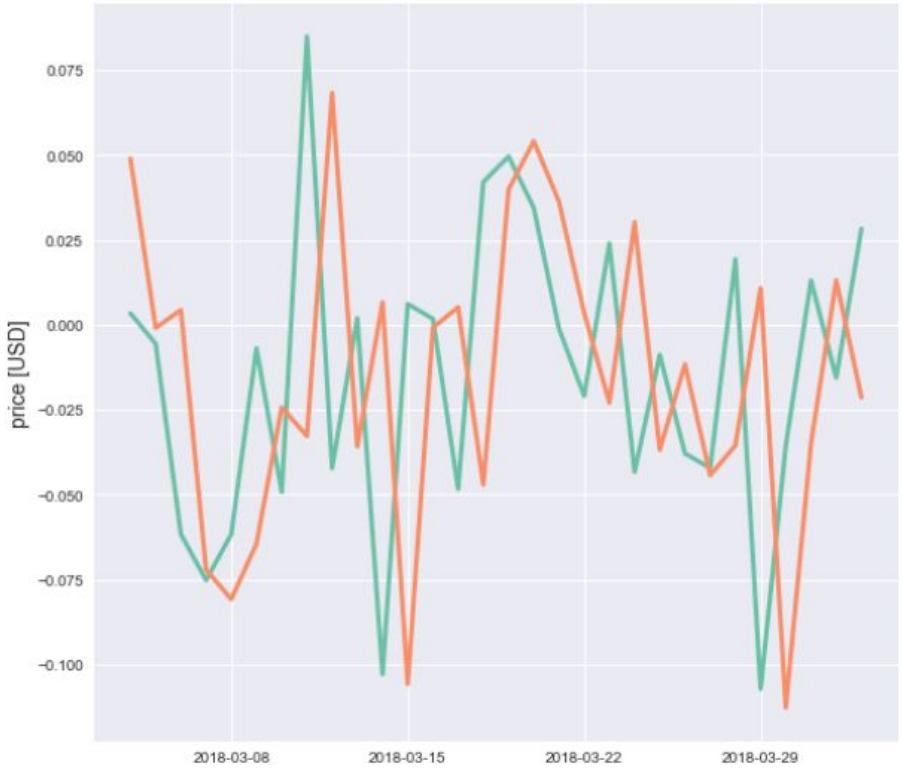
Predicting Bitcoin Price

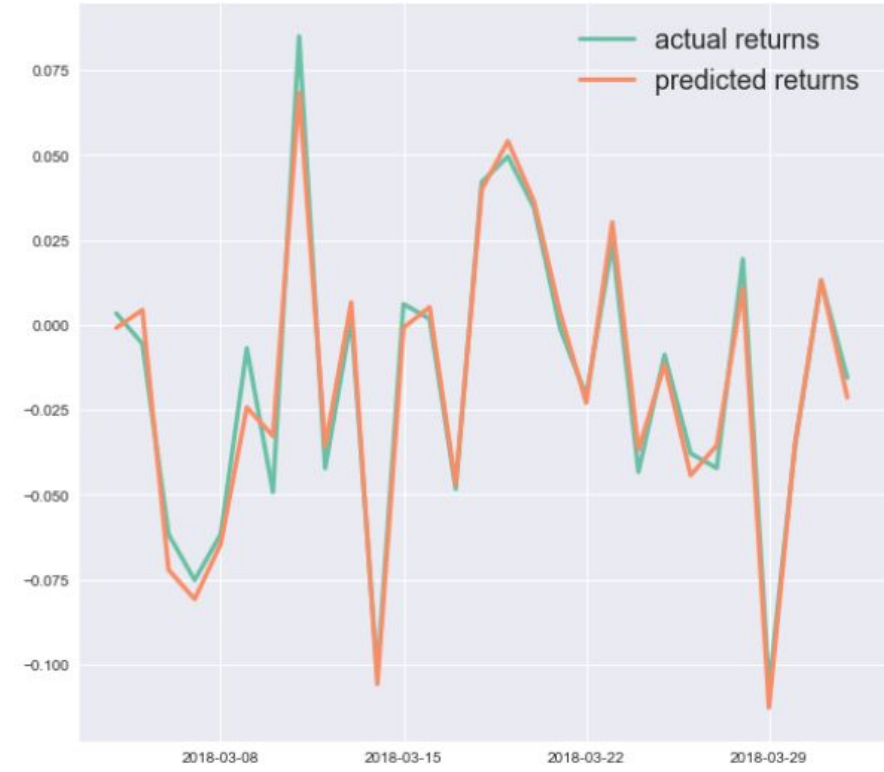

BTC
training
test
price [USD]
20000
17500
15000
12500
10000
7500
5000
2500
0
2013
2014
2015
2016
2017
2018

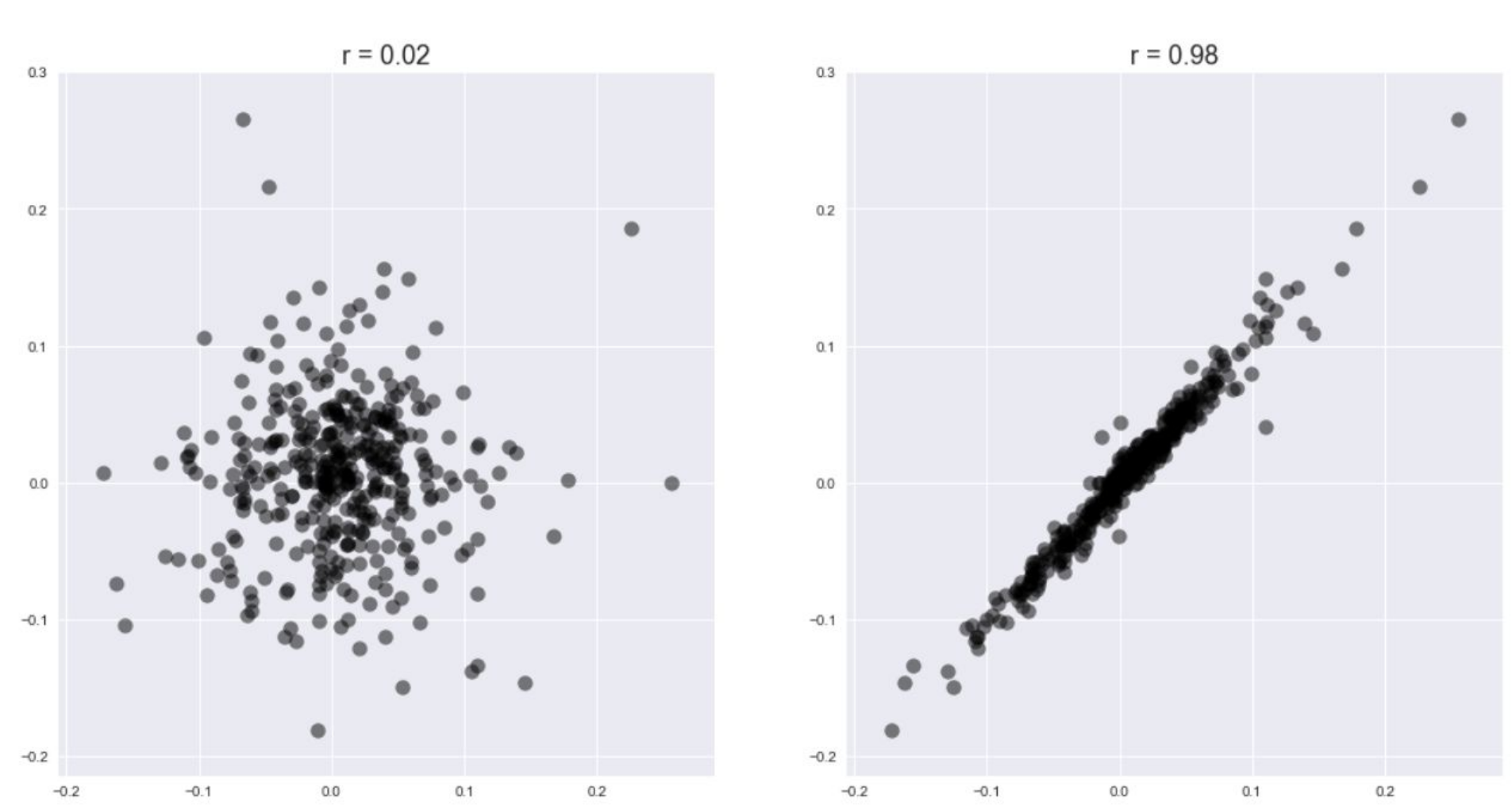

actual
prediction
price [USD]
20000
17500
15000
12500
10000
7500
5000
2500
0
2017-04
2017-06
2017-08
2017-10
2017-12
2018-02
2018-04

actual
prediction
price [USD]
11000
10000
9000
8000
7000
2018-03-08
2018-03-15
2018-03-22
2018-03-29


actual
prediction
price [USD]
11000
10000
9000
8000
7000
2018-03-04
2018-03-07
2018-03-10
2018-03-13
2018-03-16
2018-03-19
2018-03-22
2018-03-25
2018-03-28
2018-03-31


price [USD]
0.075
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
2018-03-08
2018-03-15
2018-03-22
2018-03-29


actual returns
predicted returns
0.075
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
2018-03-08
2018-03-15
2018-03-22
2018-03-29


r = 0.02
r = 0.98
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.2
−0.1
0.0
0.1
0.2

Exploring limits to prediction in complex social systems

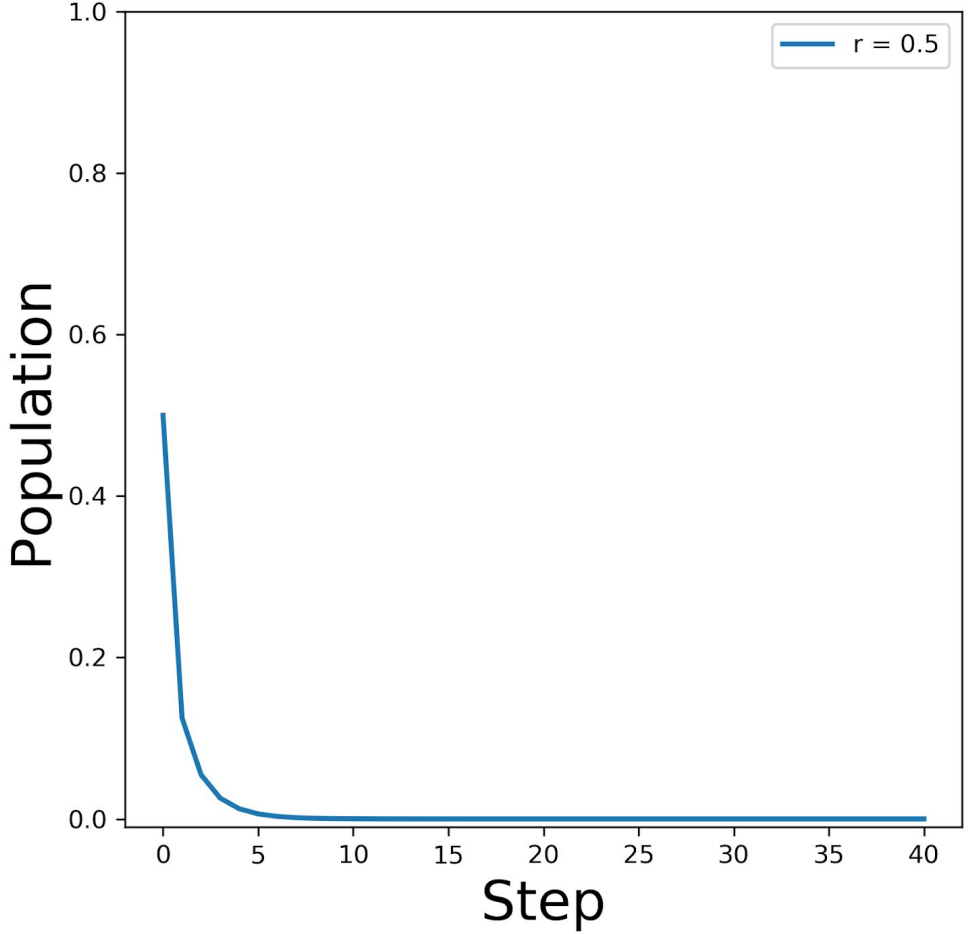
Martin T, Hofman JM, Sharma A, Anderson A, Watts DJ 2016

Proceedings of the International Conference on World Wide Web

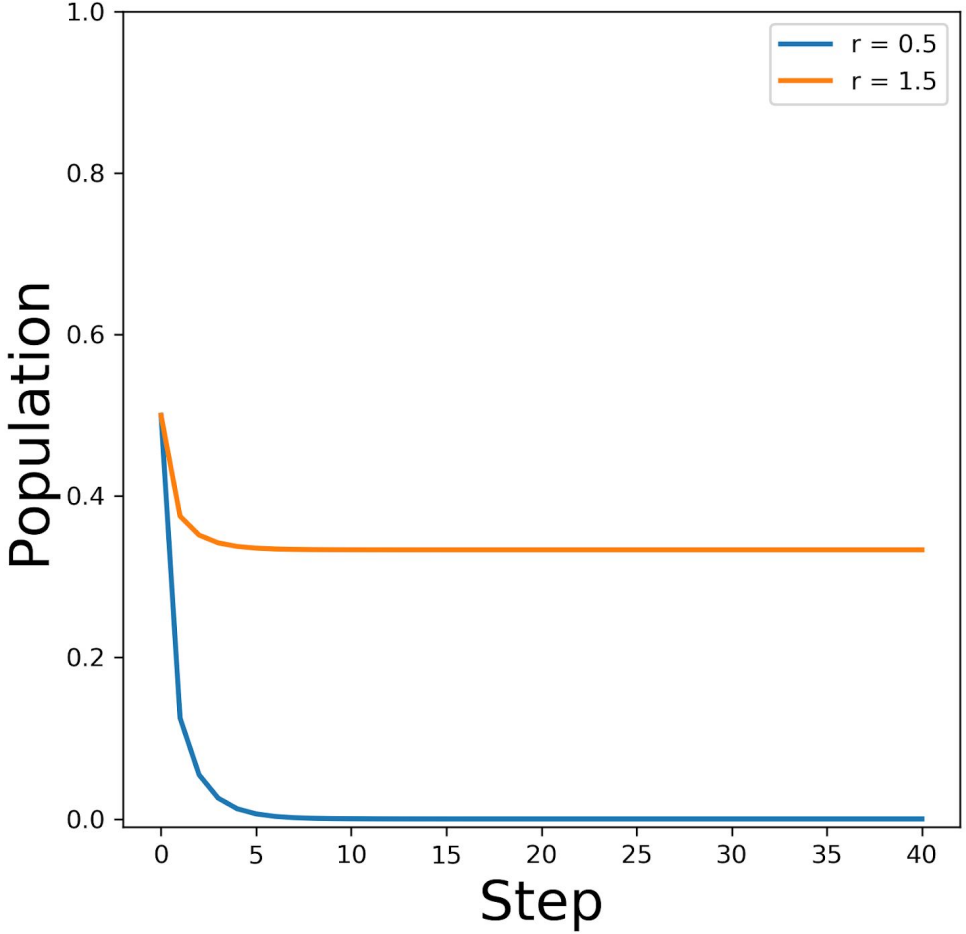
Logistic map

$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$

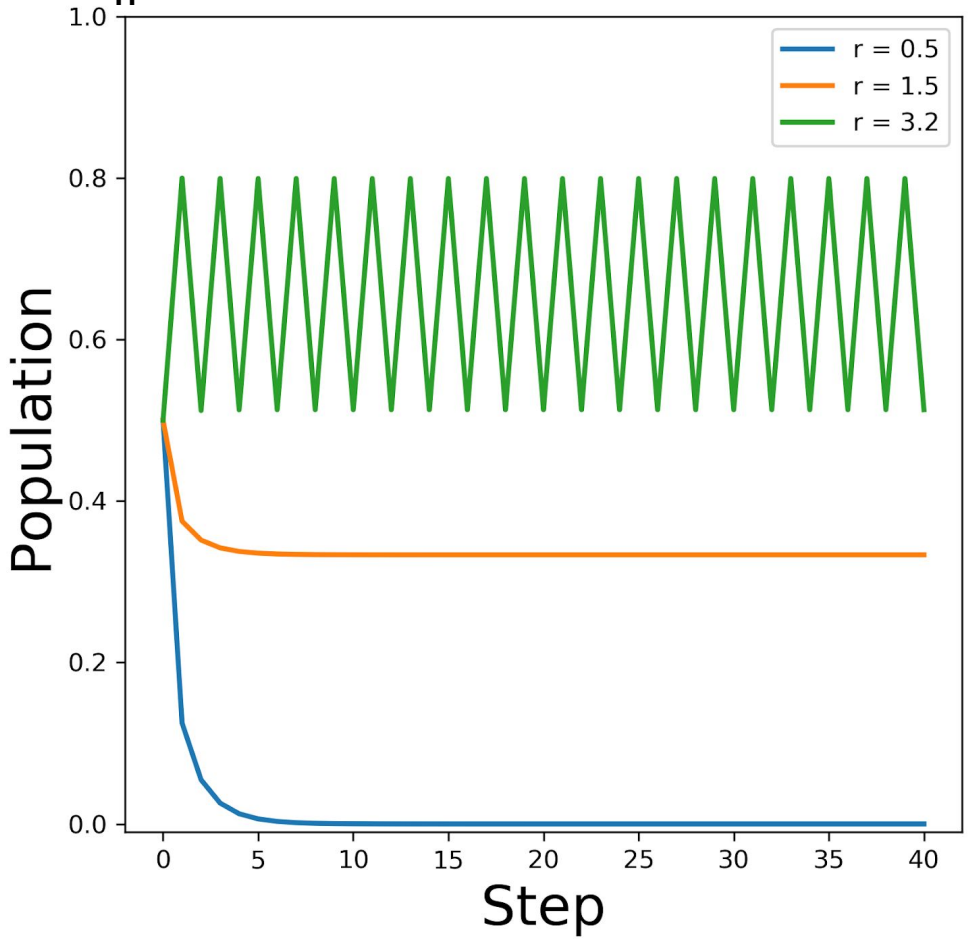
$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



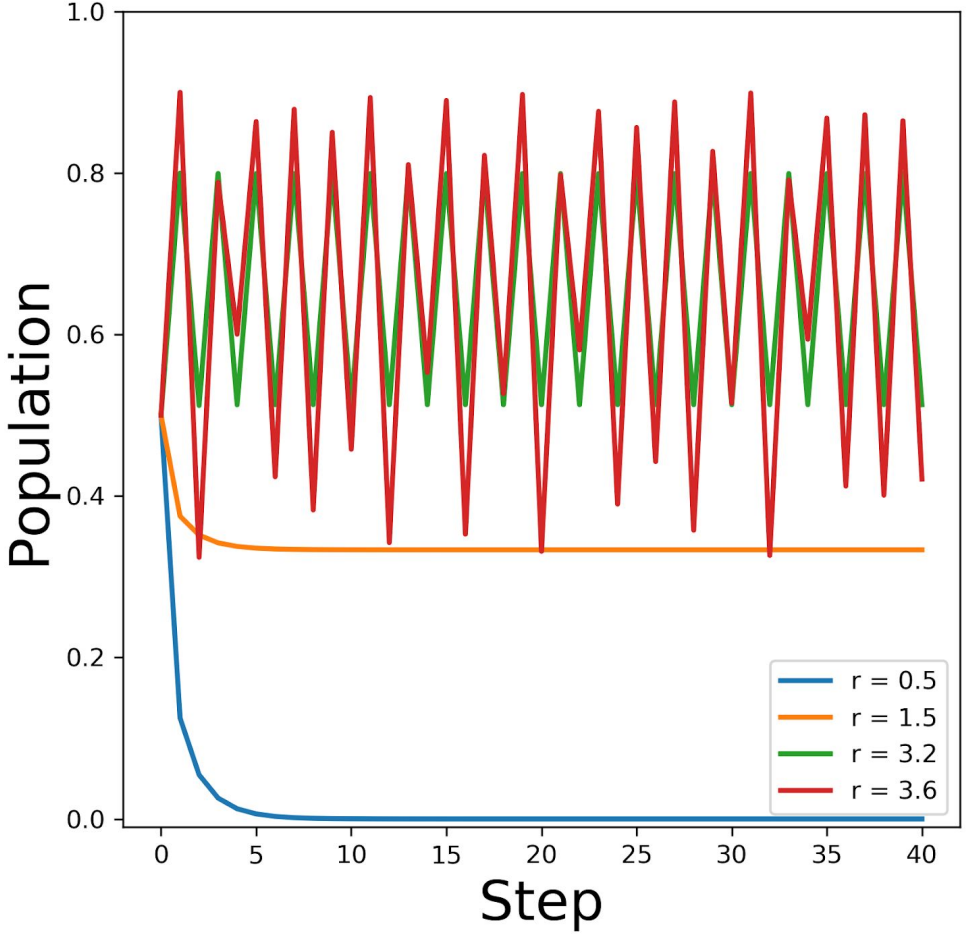
$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



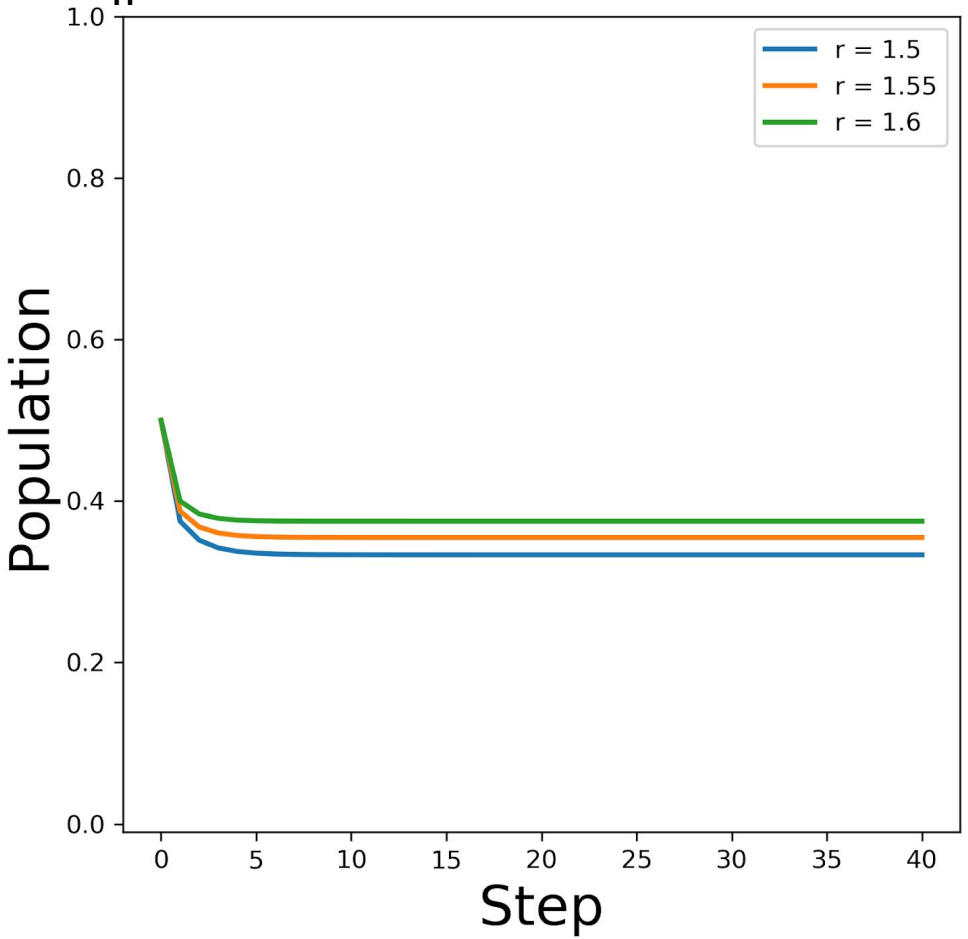
$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



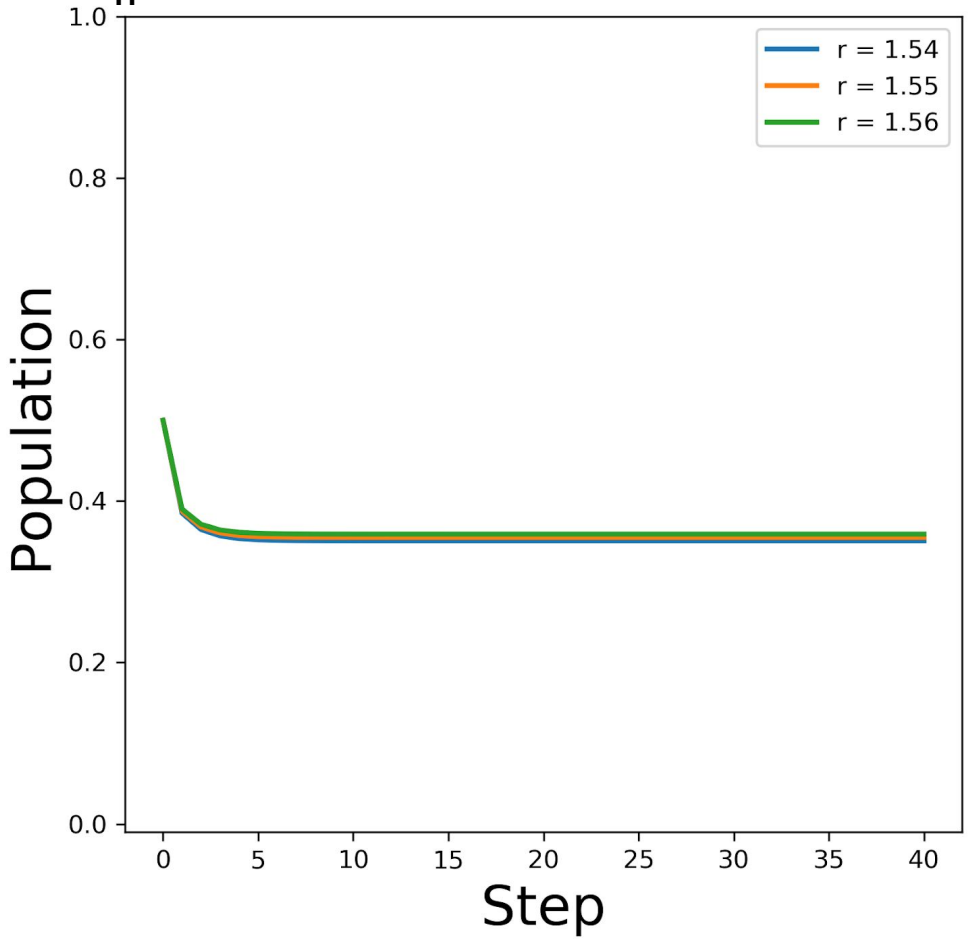
$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



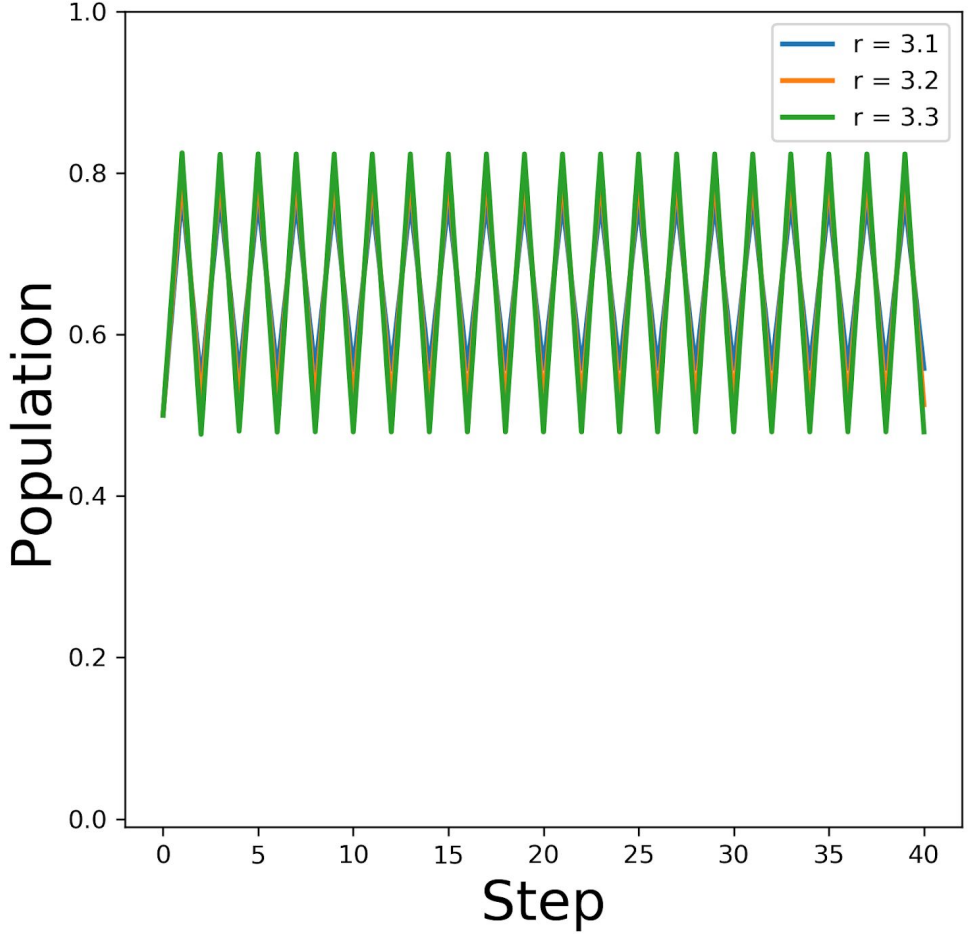
$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



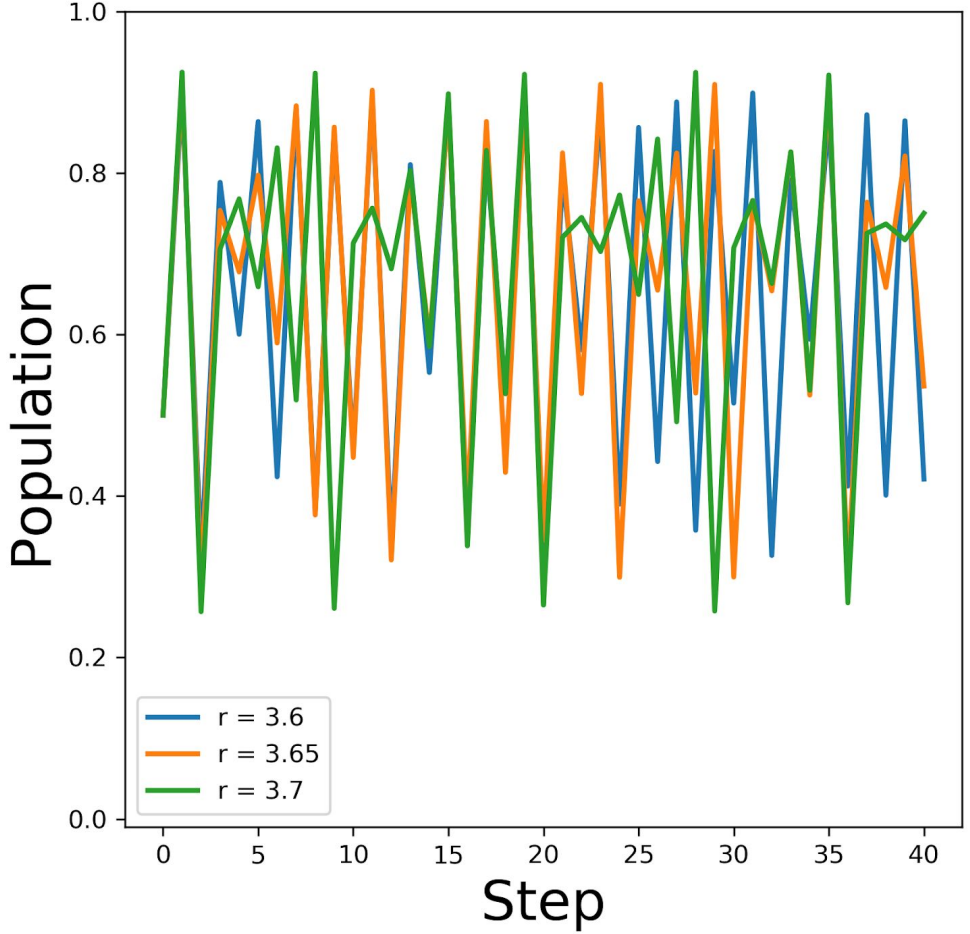
$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



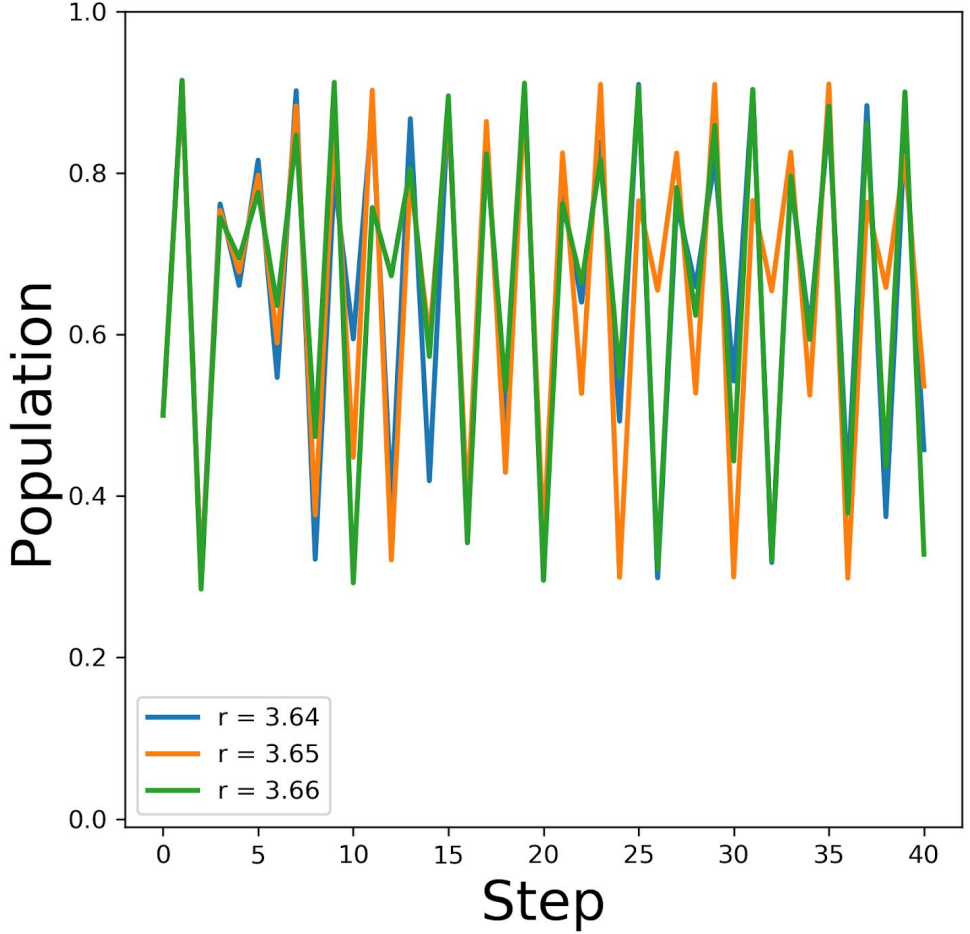
$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



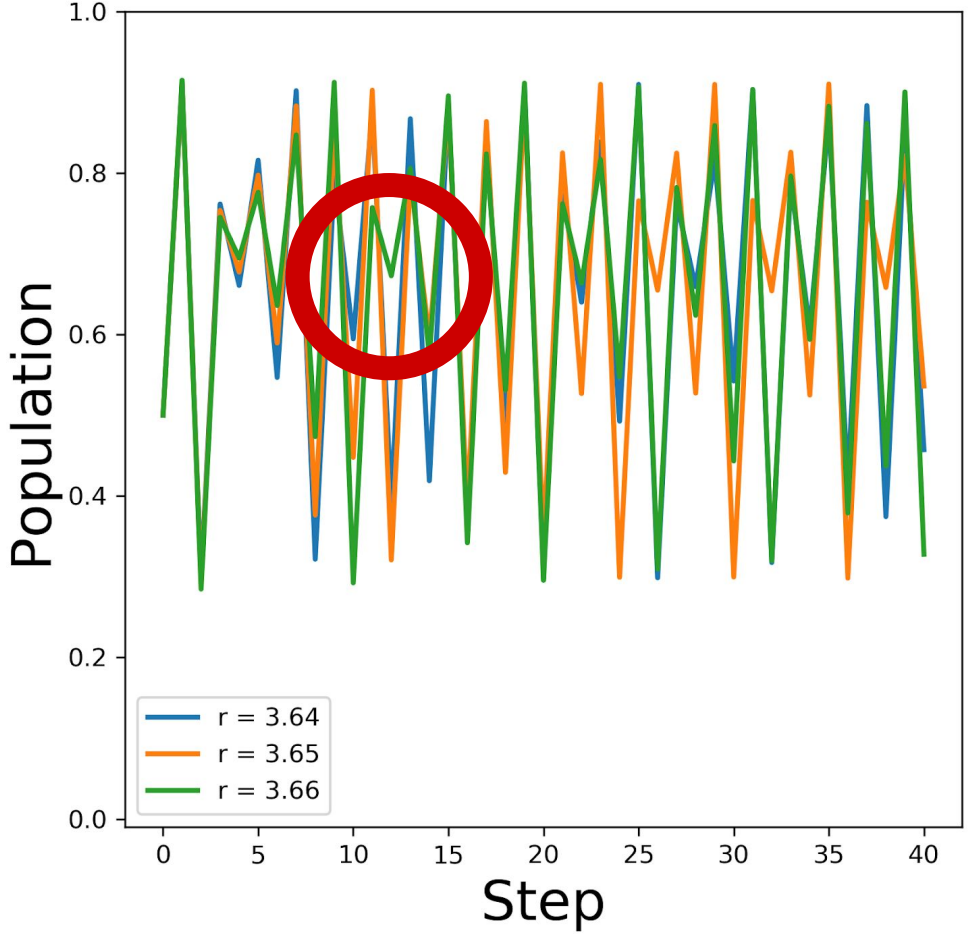
$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



$$X_{n+1} = r \cdot X_n \cdot (1 - X_n)$$



Neil deGrasse Tyson

@neiltyson

Follow

1916: Einstein predicts Gravity Waves. 1917: He lays the foundation for Lasers. 2016: Gravity Waves discovered using Lasers.

6:48 PM - 13 Feb 2016

?? Retweets ?? Likes

Neil deGrasse Tyson

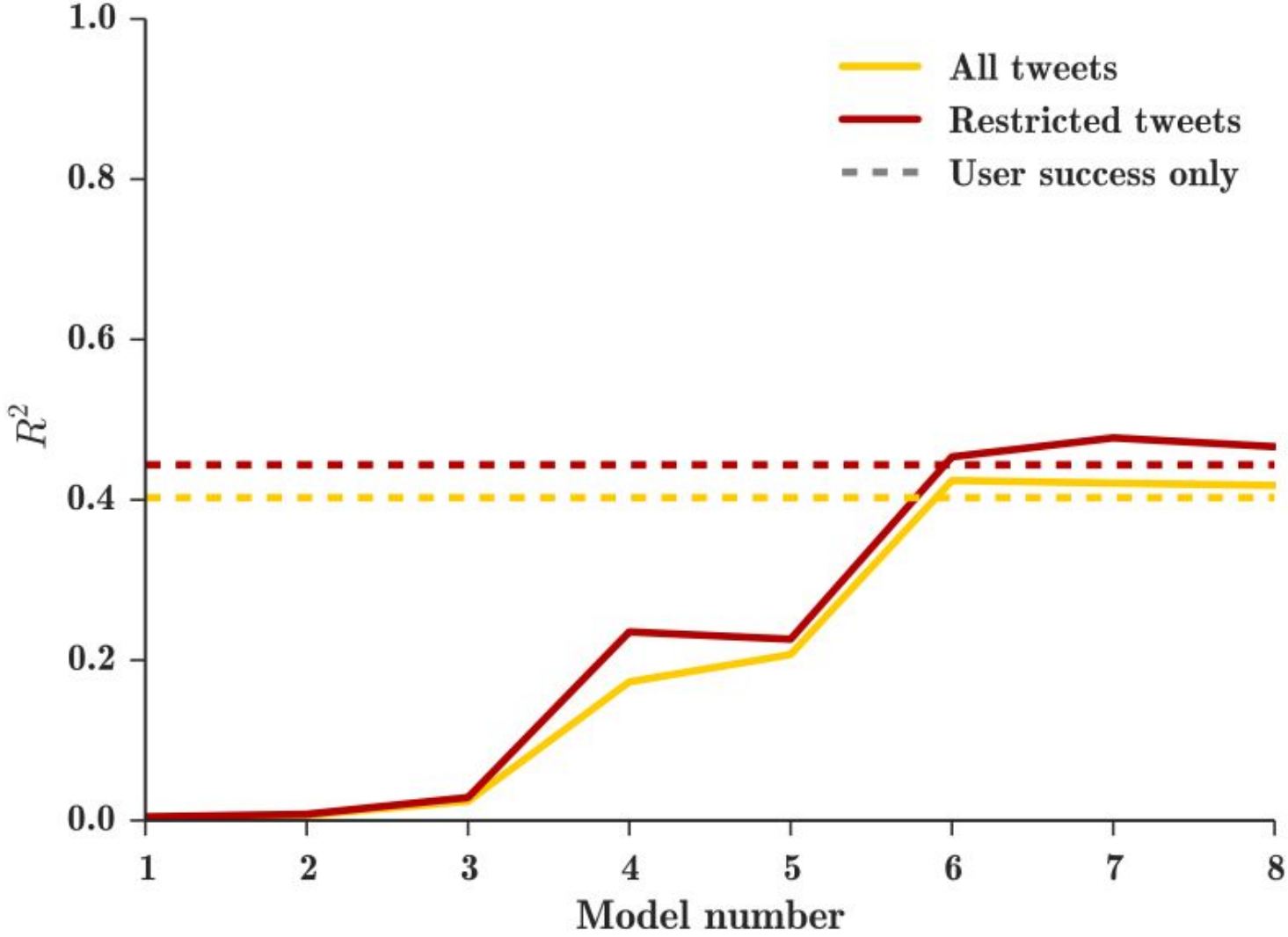
@neiltyson

Follow

1916: Einstein predicts Gravity Waves. 1917: He lays the foundation for Lasers. 2016: Gravity Waves discovered using Lasers.

6:48 PM - 13 Feb 2016

19,764 Retweets **33,949** Likes


All tweets
Restricted tweets
User success only
1.0
0.8
0.6
0.4
0.2
0.0
R^2
1
2
3
4
5
6
7
8
Model number

Measuring the predictability of life outcomes with a scientific mass collaboration

Matthew J. Salganik, ProfileIan Lundberg, Alexander T. Kindel, Caitlin E. Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M. Altschul, Jennie E. Brand, Nicole Bohme Carnegie, Ryan James Compton, Debanjan Datta, Thomas Davidson, Anna Filippova, Connor Gilroy, Brian J. Goode, Eaman Jahani, Ridhi Kashyap, Antje Kirchner, Stephen McKay, Allison C. Morgan, Alex Pentland, Kivan Polimis, Louis Raes, Daniel E. Rigobon, Claudia V. Roberts, Diana M. Stanescu, Yoshihiko Suhara, Adaner Usmani, Erik H. Wang, Muna Adem, Abdulla Alhajri, Bedoor AlShebli, Redwane Amin, Ryan B. Amos, Lisa P. Argyle, Livia Baer-Bositis, Moritz Büchi, Bo-Ryehn Chung, William Eggert, Gregory Faletto, Zhilin Fan, Jeremy Freese, Tejomay Gadgil, Josh Gagné, Yue Gao, Andrew Halpern-Manners, Sonia P. Hashim, Sonia Hausen, Guanhua He, Kimberly Higuera, Bernie Hogan, Ilana M. Horwitz, Lisa M. Hummel, Naman Jain, Kun Jin, David Jurgens, Patrick Kaminski, Areg Karapetyan, E. H. Kim, Ben Leizman, Naijia Liu, Malte Möser, Andrew E. Mack, Mayank Mahajan, Noah Mandell, Helge Marahrens, Diana Mercado-Garcia, Viola Mocz, Katariina Mueller-Gastell, Ahmed Musse, Qiankun Niu, William Nowak, Hamidreza Omidvar, Andrew Or, Karen Ouyang, Katy M. Pinto, Ethan Porter, Kristin E. Porter, Crystal Qian, Tamkinat Rauf, Anahit Sargsyan, Thomas Schaffner, Landon Schnabel, Bryan Schonfeld, Ben Sender, Jonathan D. Tang, Emma Tsurkov, Austin van Loon, Onur Varol, Xiafei Wang, Zhi Wang, Julia Wang, Flora Wang, Samantha Weissman, Kirstie Whitaker, Maria K. Wolters, Wei Lee Woon, James Wu, Catherine Wu, Kengran Yang, Jingwen Yin, Bingyu Zhao, Chenyun Zhu, Jeanne Brooks-Gunn, Barbara E. Engelhardt, Moritz Hardt, Dean Knox, Karen Levy, Arvind Narayanan, Brandon M. Stewart, Duncan J. Watts, and Sara McLanahan

It is hard to predict life trajectories

For GPA, $R^2 \sim 0.2$

The problem is
not “predictability”

The problem is
not “predictability”
not ML

The problem is
not “predictability”
not ML
But rather
GPA is bad measure

The problem is
not “predictability”
not ML
But rather
GPA is bad measure

Properly measured academic performance
is highly predictable

Properly measured academic achievements are highly predictable

For GCSE Scores heritability is 58%

Krapohl E et al. 2014 The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence // *PNAS*

Properly measured academic achievements are highly predictable

For GCSE Scores heritability is 58%

Krapohl E et al. 2014 The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence // *PNAS*

Simple model trained on posts from social media explains 30% of PISA scores

Smirnov I 2020 Estimating educational outcomes from students' short texts on social media // *EPJ Data Science*

Properly measured academic achievements are highly predictable

For GCSE Scores heritability is 58%

Krapohl E et al. 2014 The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence // *PNAS*

Simple model trained on posts from social media explains 30% of PISA scores

Smirnov I 2020 Estimating educational outcomes from students' short texts on social media // *EPJ Data Science*

Just one variable (socio-economic status) could explain more than 20% of PISA scores

OECD, PISA 2018 Results (Volume II) Where All Students Can Succeed