Using Consumer's Digital Exhaust and other Non-Traditional Data

MERVYN KOPINSKY

November 19, 2019

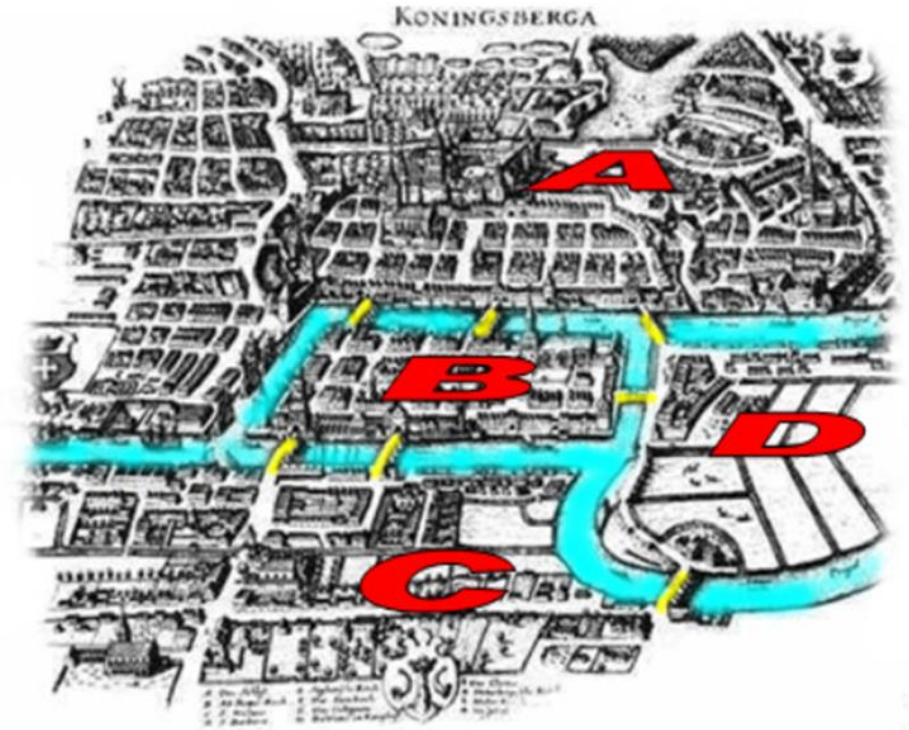
SOCIETY OF ACTUARIES

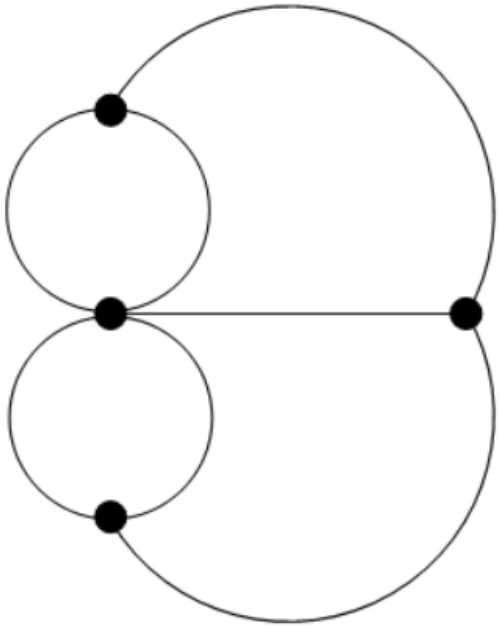
Agenda

- Introduction: some perspectives on big data
- Nontraditional data sources
- AI, Insurtech and looking forward
- Actuary of the future
- Some recent research

Some Background . . . Graph Theory

- Leonard Euler's brilliant solution to the 7 bridges of Konigsberg problem in 1735
- Must be no more than 2 vertices with an odd number of lines





Application of Graph Theory

EXAMPLES OF NETWORKS AND THEIR COMPONENTS

NETWORK	VERTICES	VERTEX ATTRIBUTES	EDGES	EDGE ATTRIBUTES
Airlines Network	**Airports**	Footfall, Terminals, Staff, City population, International/Domestic, Freight, Hangar capacity	**Airplanes / Routes**	Frequency, # Passengers, Plane Type, Fuel Usage, Distance covered, Empty seats
Banking Network	**Account Holders**	Name, demographics, KYC Document, Products, Account status, balance and other details	**Transactions**	Type, Amount, Authentication (pass/OTP), Time, Location, Device
Social Network	**Users**	Name, demographics, # connections, likes, circles belong to, subscriptions	**Interactions**	Medium (like/comment/direct message), time, duration, type of content, topic
Physician Network	**Doctors**	Demographics, speciality, experience, affiliation (type and size), Weekly patient intake	**Patients**	Demographics, Diagnosis history, visit frequency, purpose, referred to, insurance
Supply Chain Network	**Warehouses**	Location, size, capacity, storage type, connectivity, manual/automated	**Trucks**	Load capacity, # wheels, year of make, geographical permit, miles travelled. Maintenance cost, driver experience

https://www.analyticsvidhya.com/blog/2018/04/introduction-to-graph-theory-network-analysis-python-codes/

Phone Call Fraud Detection

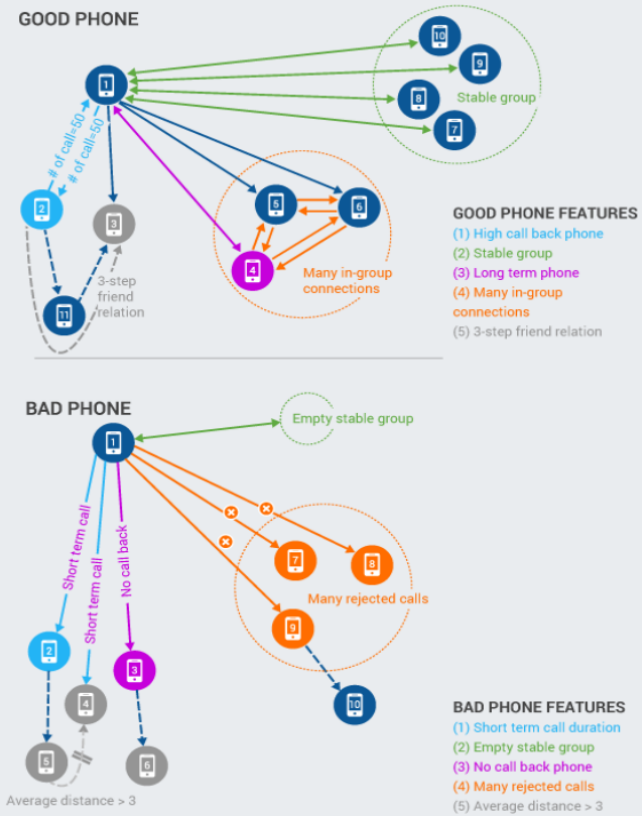
TigerGraph

PRODUCT SOLUTIONS RESOURCES COMPANY DEMO DOWNLOAD DEVELOPERS SEARCH

Improving Fraud Detection with AI and Machine Learning

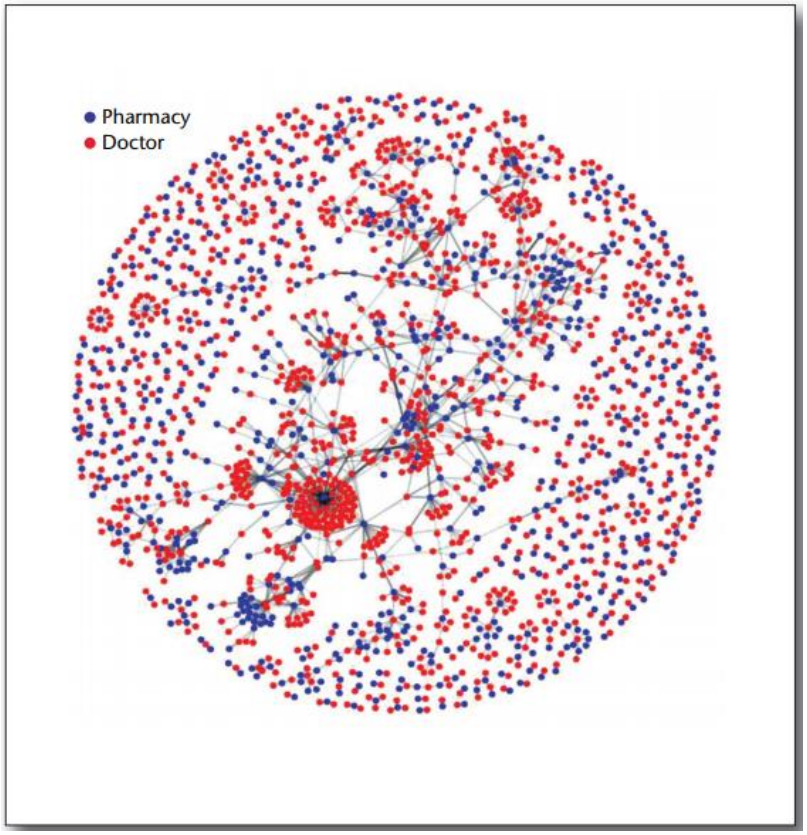
Confirmed fraud events are rare, with less than 1% of total call volume for Telecom or claims data for healthcare and government benefits or payment transactions for financial services. This means that the machine learning models do not have sufficient training data with confirmed fraud activity to learn and improve accuracy of fraud detection. TigerGraph with its native parallel architecture is purpose-built to address this challenge. Consider the example for phone-based fraud detection at China Mobile, where TigerGraph creates over 118 features for each phone in real-time by analyzing relationships among subscribers over time, identifying a good phone owned by a regular customer and bad phone suspected to be a fraudster.

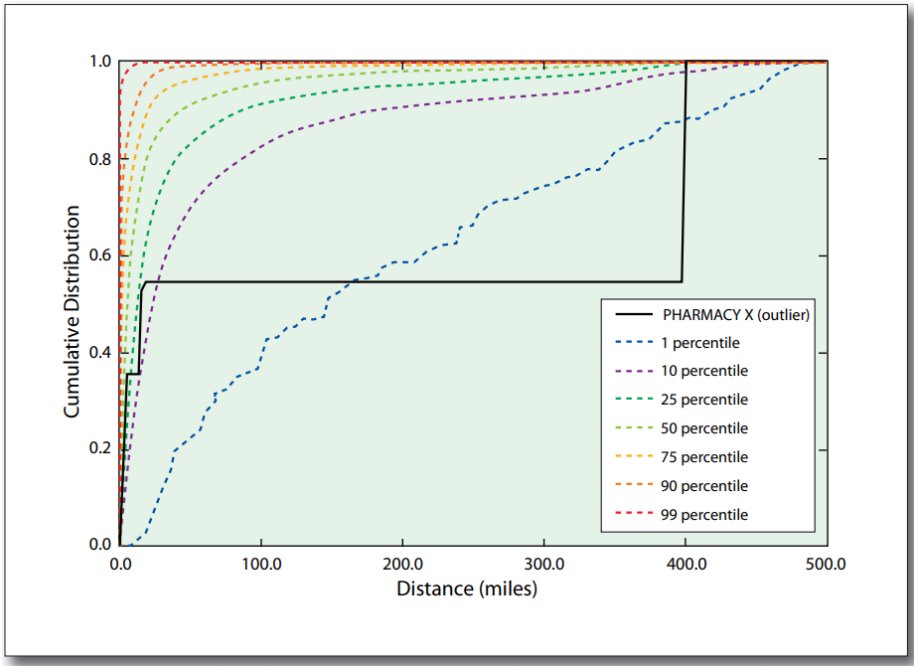
READ THE EXECUTIVE BRIEF



https://www.tigergraph.com/solutions/fraud-detection/

Graph Theory: Fraud detection for narcotics prescriptions



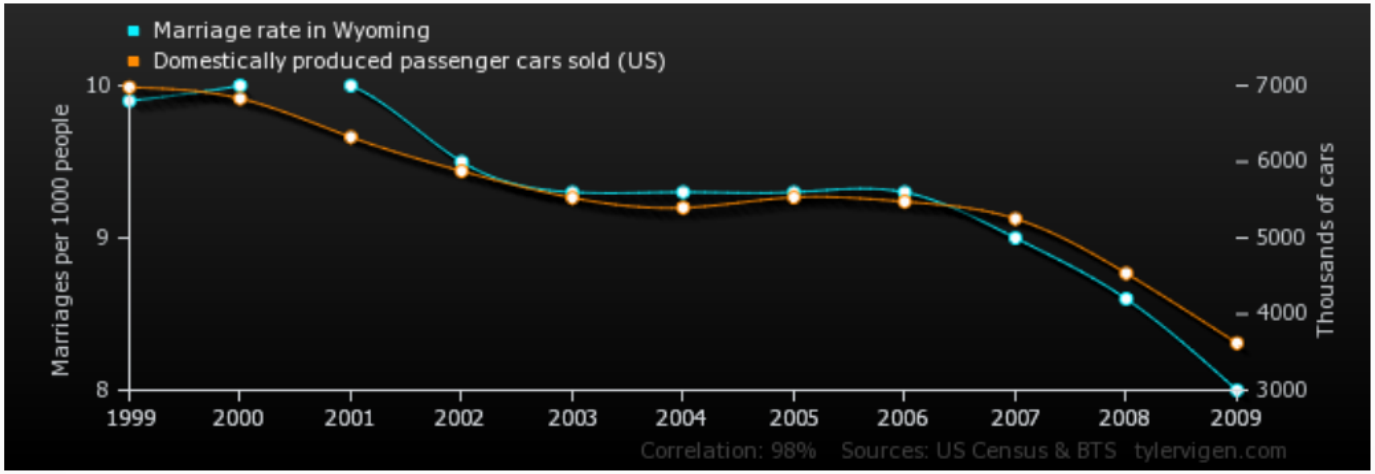


https://www.aaai.org/ojs/index.php/aimagazine/article/view/2630/2554

Graph theory demo: https://www.youtube.com/watch?v=uA525G3beFo&list=PLq4l3NnrSRp4IGO-CgwjRa1JqxN-vWATx&index=13

Find All Types of Relationships . . .

Marriage rate in Wyoming
correlates with
Domestically produced passenger cars sold (US)



Upload this chart to imgur

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Marriage rate in Wyoming *Marriages per 1000 people (US Census)*	9.9	10	10	9.5	9.3	9.3	9.3	9.3	9	8.6	8
Domestically produced passenger cars sold (US) *Thousands of cars (BTS)*	6,979	6,831	6,325	5,878	5,527	5,396	5,533	5,476	5,253	4,535	3,619

Correlation: 0.976233

https://www.tylervigen.com/spurious-correlations

Nontraditional Data Sources

- Different types of data in use:
  - Demographic
  - Financial
  - Government
  - Climate
  - Medical
  - Motor vehicle records
  - Public records
  - Telematics
- AI and Looking Forward
  - Epigenetics
  - Digital behavioral data
  - Internet of Things

Demographic Data

- Examples of Demographic data to append to company data
  - Home ownership
  - Occupation
  - Gender
  - Race
  - Estimated Income
  - Phone numbers
  - Credit card information
  - Hobbies
  - E-mail append services
  - Languages spoken
  - Purchase behavior
  - Lifestyle interests
  - Investment information
  - Family information
  - Life events
- Information cannot be used to discriminate
- Information could be useful *to show* nondiscrimination
  - SOA sponsoring current research projects to get industry and regulator input regarding validation of algorithmic models

Financial Data

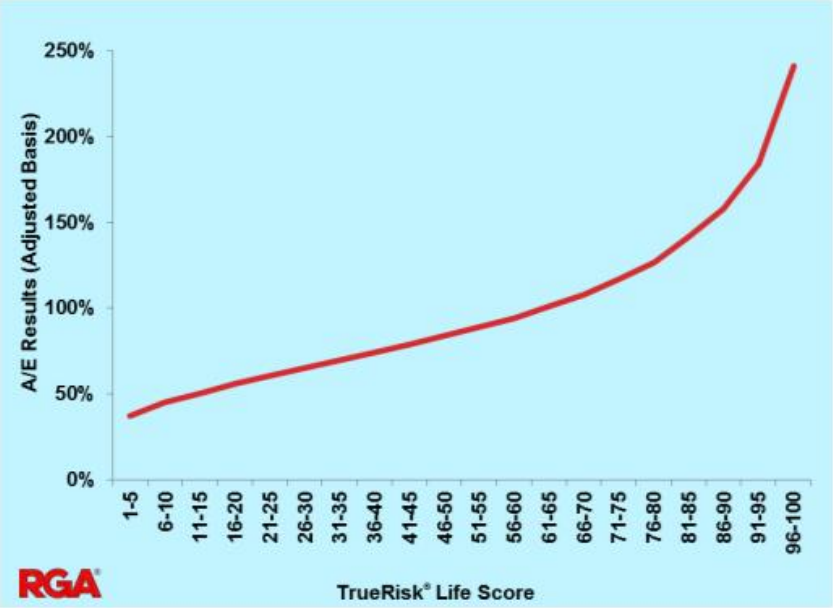
- Credit scores have been shown to predict mortality risk
- Transunion TrueRisk Life Score
  - Credit-based behavioral index
  - 25 Selected credit attributes selected from 800 attributes
  - TrueRisk represented as a number between 1 (low risk) and 100 (high risk).

https://www.soa.org/globalassets/assets/files/e-business/pd/events/2017/underwriting-seminar/underwriting-seminar-presentation-risk-assessment-2.pdf

TrueRisk Correlated with A/E's . . . .

Traditional Mortality Study (prepared by RGA)

Overall Mortality



Details

- Mortality study performed on holdout sample of 18 million lives using a 1998 TransUnion archive and studying the lives during 1999-2010
- Score buckets are set to be uniform across the population
- Study shows 5 times segmentation (96-100 compared to 1-5)
- SSMDF used as source of deaths; used population mortality tables

. . .After Allowing for Preferred Status

Mortality by Underwriting Class (prepared by RGA)

- Segmentation exists within risk classes
- Mortality for highest TrueRisk® Life scores (71-100) are over double that of lowest scores (1-10)
- Non-smokers are shown, but results are similar for smokers

Term, UL & VUL; Face Amounts ≥ $100,000; Issue Ages < 70

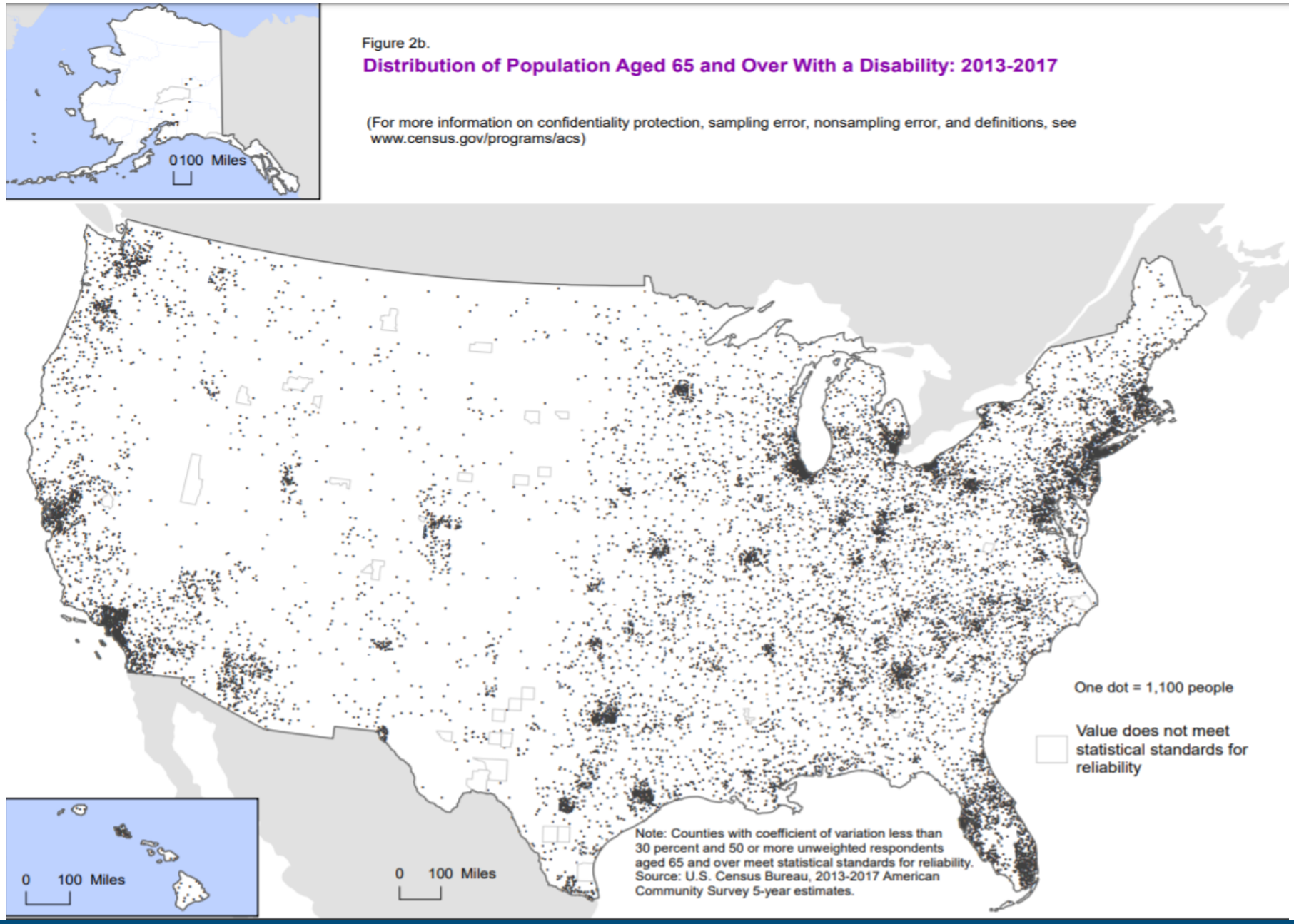
Government Data

- Government statistics
  - Health
  - Education
  - Worker-safety
  - Housing information
  - Better visibility into geographic + demographic effects
  - https://www.census.gov/en.html
  - Image on next page: https://www.census.gov/content/dam/Census/library/visualizations/time-series/demo/older-population/Figure%202%20Population%20Aged%2065%20and%20Over%20With%20a%20Disability.pdf

Figure 2b.

Distribution of Population Aged 65 and Over With a Disability: 2013-2017

(For more information on confidentiality protection, sampling error, nonsampling error, and definitions, see www.census.gov/programs/acs)



Note: Counties with coefficient of variation less than 30 percent and 50 or more unweighted respondents aged 65 and over meet statistical standards for reliability.
Source: U.S. Census Bureau, 2013-2017 American Community Survey 5-year estimates.

Climate Data

- Insurance typically covers extreme events – earthquakes or hurricanes
- Parametric insurance – does not indemnify losses but pays after a triggering event
  - Index insurance – payout when a measurable index triggers a claim payment.
  - Payment could be triggered on too much or too little rain
  - Historical rainfall data needed to asses risk
  - Not all farmers are near a rain gauge – may need satellite data
  - Possible to combine rain gauge information with satellite imagery to create usable datasets

https://iri.columbia.edu/news/enacts-transforms-insurance-projects-in-africa/

Property Level Data

- Hazardhub is one example
  - New services offering detailed data based on big data sources
  - More information available to P&C underwriters
  - Delivers overs 475 hazard risks and 100 property data elements that can be used by insurers

HazardHub in action
637 Hibiscus Glen, Escondido, CA, 92025 – the basics

Nearest Fire Station	1.1 Miles
Fire Station Name	Escondido FD Station 4
Fire Station Staffing	Full Time
Fire Response Area	Escondido Fire Protection Zone
Number of Fire Stations in 5 miles	9
Nearest Hydrant	Less than 250 Feet
Number of Hydrants in 1000 Ft	18
AAIS Fire Protection Class	P2
Distance to Coas	13.48 Miles
Geocoding Accuracy	Rooftop

https://www.willistowerswatson.com/en-US/Insights/2019/05/quarterly-insurtech-briefing-q1-2019

Medical Data

- Electronic drug records quickly available
  - Drug, fill dates, dosage, pharmacy, physician
- Medical data
  - Claims data, procedure codes, diagnosis codes
- Different EHR platforms have slowed adoption
  - SOA currently working with a researcher on an EHR white paper
- Proposed VM-51 changes
  - Blood pressures, LDL, HDL, family history, cause of death, underwriting details, whether Rx information used, whether credit data used etc.
  - Phase-in depending on anticipated data retrieval effort
  - Example (newer exposure version in process):
    - https://www.naic.org/meetings1808/cmte_a_latf_2018_summer_nm_accelerated_underwriting_data_request.xlsx

Telematics

- Pay as you drive; usage-based insurance; pay per mile
- Metromile, Esurance, Travelers, Safeco
- Dongles not cheap – substantial reduction in loss ratio needed to pay for it
- Aggregated telematics data used for pricing
- A UK Company reported 30% drop in number of claims
- Future connectivity
  - To internet – 90% of new cars by 2020 (?) *
  - Vehicle to vehicle
  - Vehicle to everything

* https://medium.com/iotforall/7-connected-car-trends-fueling-the-future-946b05325531

Traditional vs Telematics Data

Traditional Insurance	Telematics
Age	Miles driven
Gender	Hard braking
Past driving record	Hard cornering
Education level	Rapid acceleration
Address location	Locations – urban, city, traffic density
Marital status	Traffic patterns
	Time of day when car drive
	Week or weekend use

Obs.	Date	UTC	Degrees Latitude	Degrees Longitude	Seconds	Miles	Gallons	Miles per hour	g-force
1	4/8/2012	14:45:30	-27.117	-109.367	0:00:00	0.000	0.000	0	0.000
2	4/8/2012	14:46:30	-27.118	-109.364	0:01:00	0.000	0.050	0	0.000
3	4/8/2012	14:47:30	-27.117	-109.371	0:02:00	0.080	0.055	20	0.030
4	4/8/2012	14:48:30	-27.150	-109.379	0:03:00	0.552	0.060	50	0.060
5	4/8/2012	14:49:30	-27.109	-109.388	0:04:00	1.499	0.078	65	-0.100
6	4/8/2012	14:50:30	-27.098	-109.398	0:05:00	2.538	0.091	59	0.010

https://towardsdatascience.com/telematics-in-auto-insurance-a886a03b5a88

AI, Insurtech and Looking Forward

SOCIETY OF ACTUARIES

AI – In the blink of an eye . . .

Some Chess History	Date	Elapsed
Modern chess rules	1475	
Bobby Fisher ELO Rating – 2,785	1972	497 Yrs
Magnus Carlson ELO Rating – 2,882	2017	542 Yrs
First computer chess Engine	1957	
Deep Blue beats Garry Kasparov	1997	40 Yrs
Stockfish ELO Rating – 3390	2017	60 Yrs
Google's Alphazero – AI, no human heuristics. ELO Rating 3,400+	2017	4 Hours

All human thought

Human thought with computer help

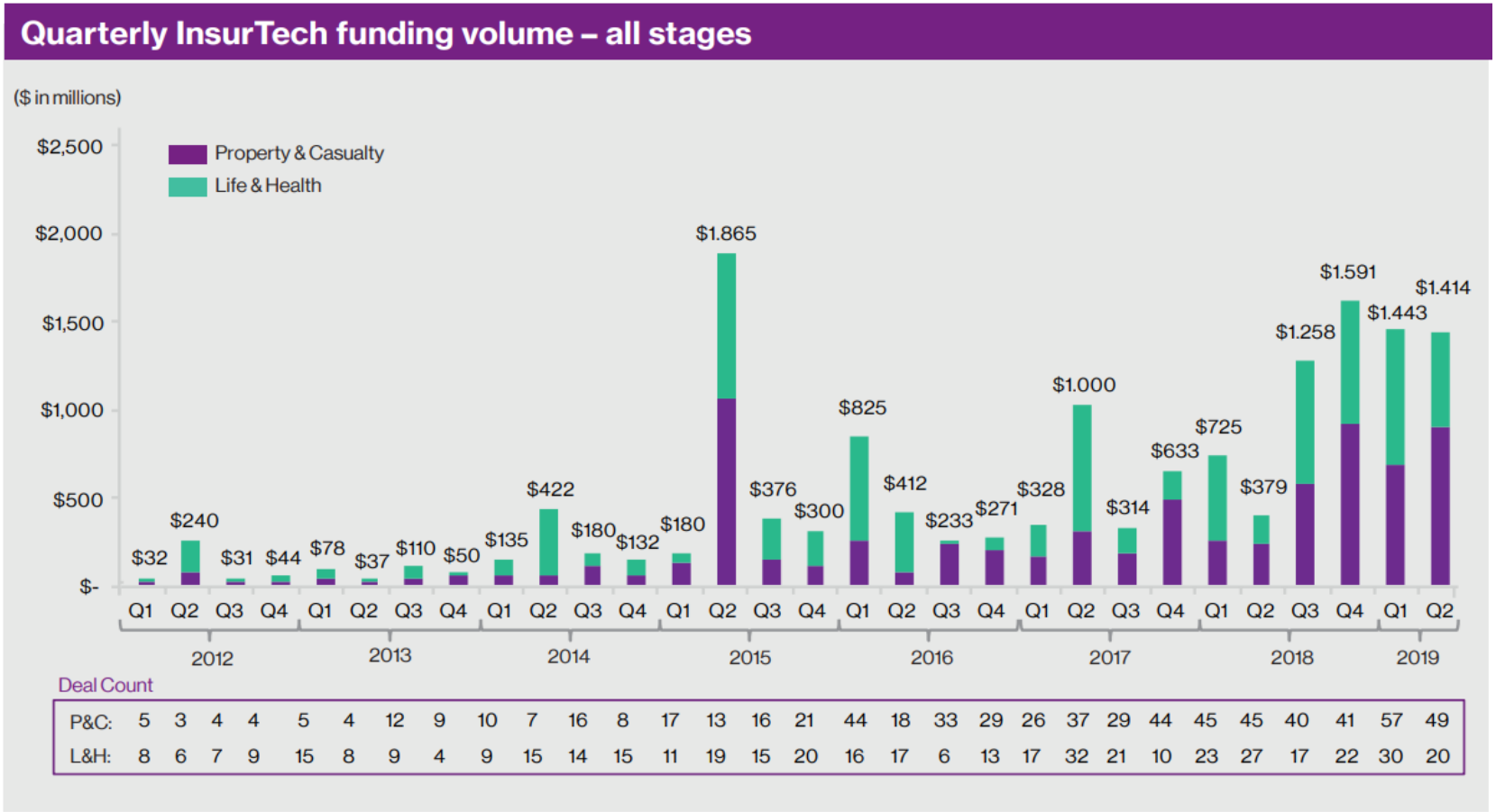
AI "thought" alone

Quotes about AlphaZero:

- " . . . . . likened AlphaZero's play to that of a superior alien species."
- " . . . . ."insane attacking chess" with profound positional understanding."

Example of AI in action: https://www.youtube.com/watch?v=bo5plUo86BU

Insurtech Funding



https://www.willistowerswatson.com/en-US/Insights/2019/07/Quarterly-InsurTech-Briefing-Q2-2019

Looking Forward – Life Insurance

- Epigenetics
- Wearables and apps
- Marketing
  - Social information utilization
- Risk management and wellness programs
  - John Hancock Vitality program
- Likely impact on numerous practice areas:
  - Product development
  - Pricing
  - Reserving

Looking Forward – Health & Health Insurance

- "Detect and Repair" to "Predict and Prevent"
- Fitbits / activity tracker use reduces health premiums
- Enormous AI advances in health care
  - Radiology
  - Phone apps
  - Remote surgery
- Value-based payment models
  - CMS already there for some time
  - Aetna / Medtronic insulin pumps
  - Medtronic – antibacterial sleeve

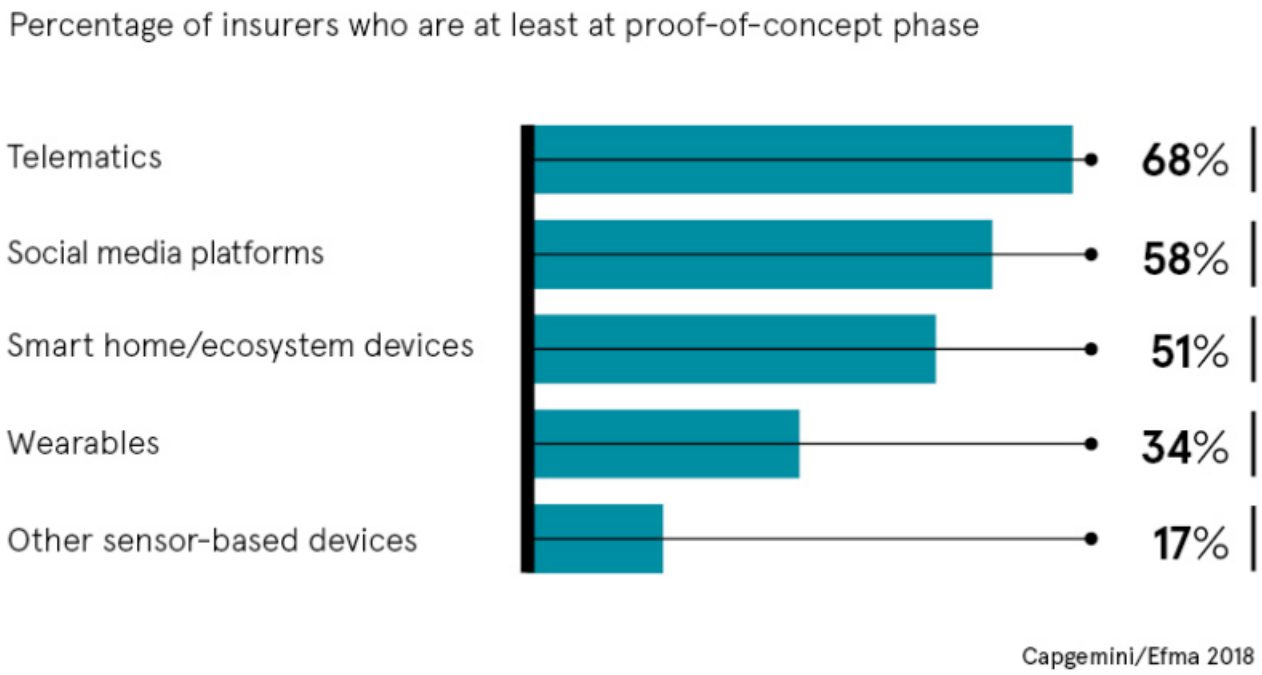
Looking Forward – Pay As You Live

- Turn coverage on and off via a phone app
- Autonomous vehicles
- Peer-to-Peer insurance
  - Small groups get together
  - Fraud reduction
  - Lemonade / Teambrella / Friendsurance
- Internet of Things
  - Install smart devices to warn and prevent

https://www.ft.com/content/bb9f1ce8-f84b-11e6-bd4e-68d53499ed71

Looking Forward – Real-Time Data

- How insurers are capturing real-time customer data



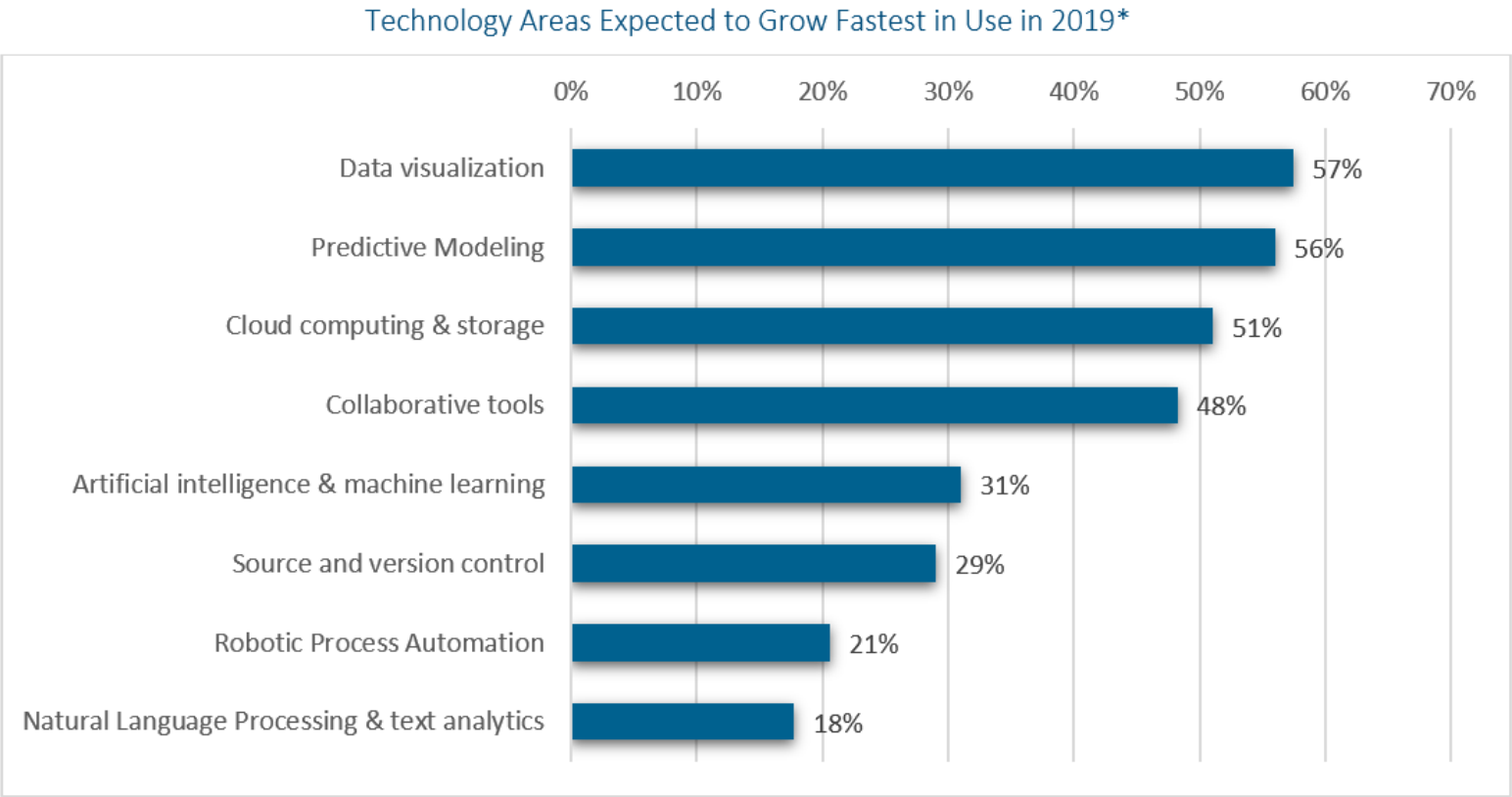
https://www.raconteur.net/risk-management/personalisation-insurance

Actuary of the Future

- New skills
  - Predictive Analytics
  - Visualization software – e.g. Tableau
  - Programming languages – e.g. R
- Who will talk who's language? *
  - SOA - 32,000 members.
  - Kaggle – 2+ million members
- Future interactions with data will change completely
  - Cloud based computing
  - Collaboration
- Automated machine learning

* https://www.kdnuggets.com/2018/09/how-many-data-scientists-are-there.html

Top Actuarial Technologies

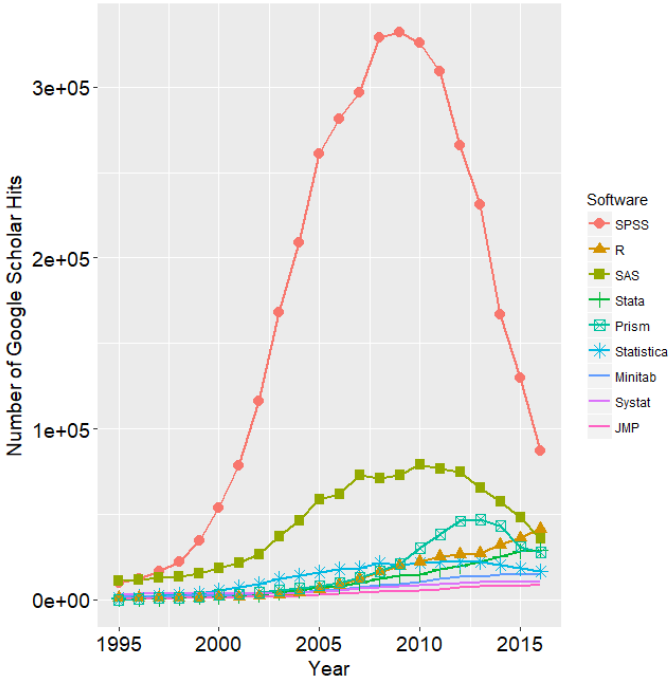


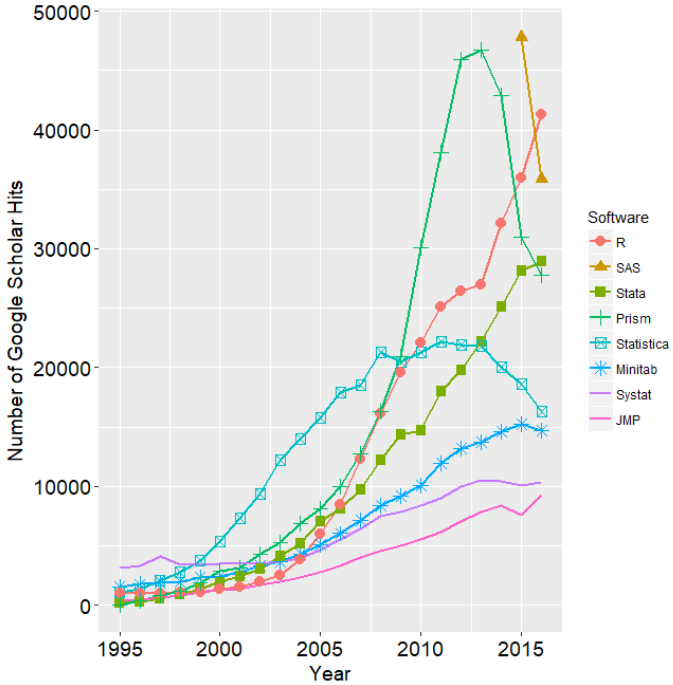
*Percentages represent the percent of actuaries surveyed who believe their usage will increase between March 2019 and March 2020.

https://www.soa.org/resources/research-reports/2019/actuarial-innovation-technology/

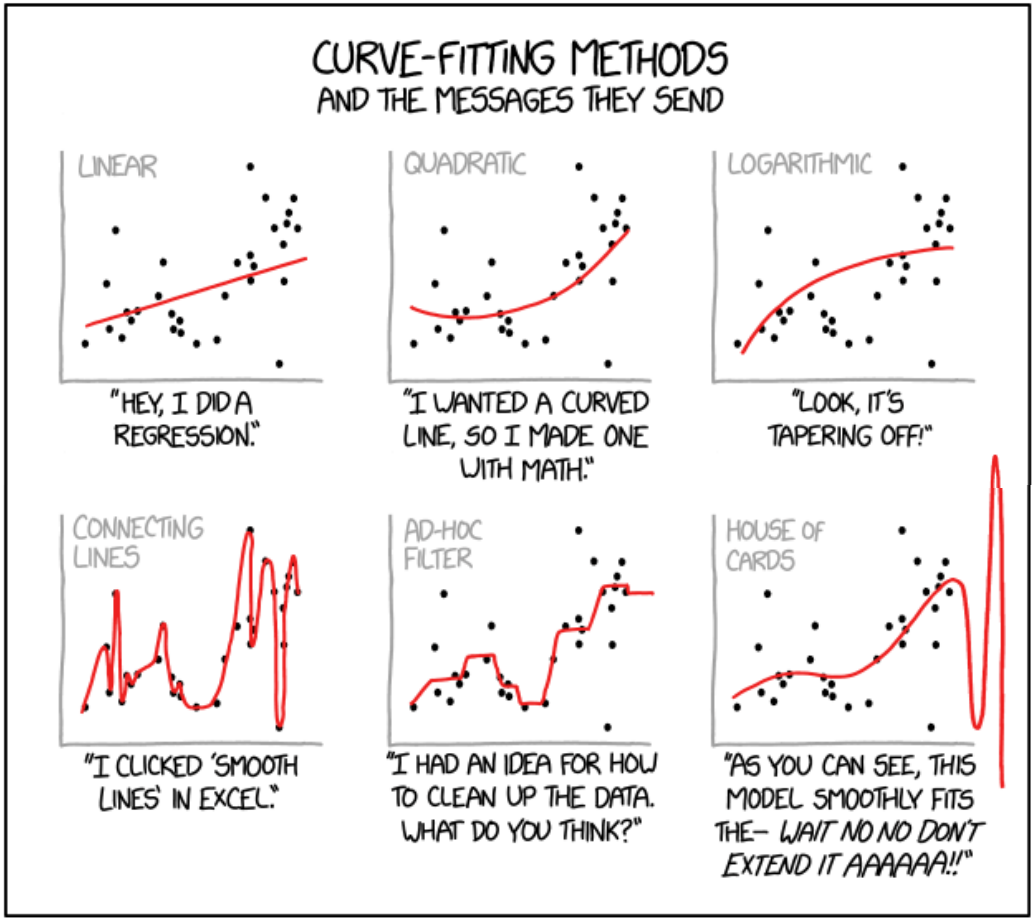
Actuary of the Future - R

- Open source language with ability to quickly add powerful statistical and data analysis packages
- Free!
- Rapid growth (Graphs from Muenchen, Robert (19 June 2017). "The Popularity of Data Science Software".)





Modeling is getting easier . .



Subset of graphs from https://xkcd.com/2048/

Some Recent Research

SOCIETY OF ACTUARIES

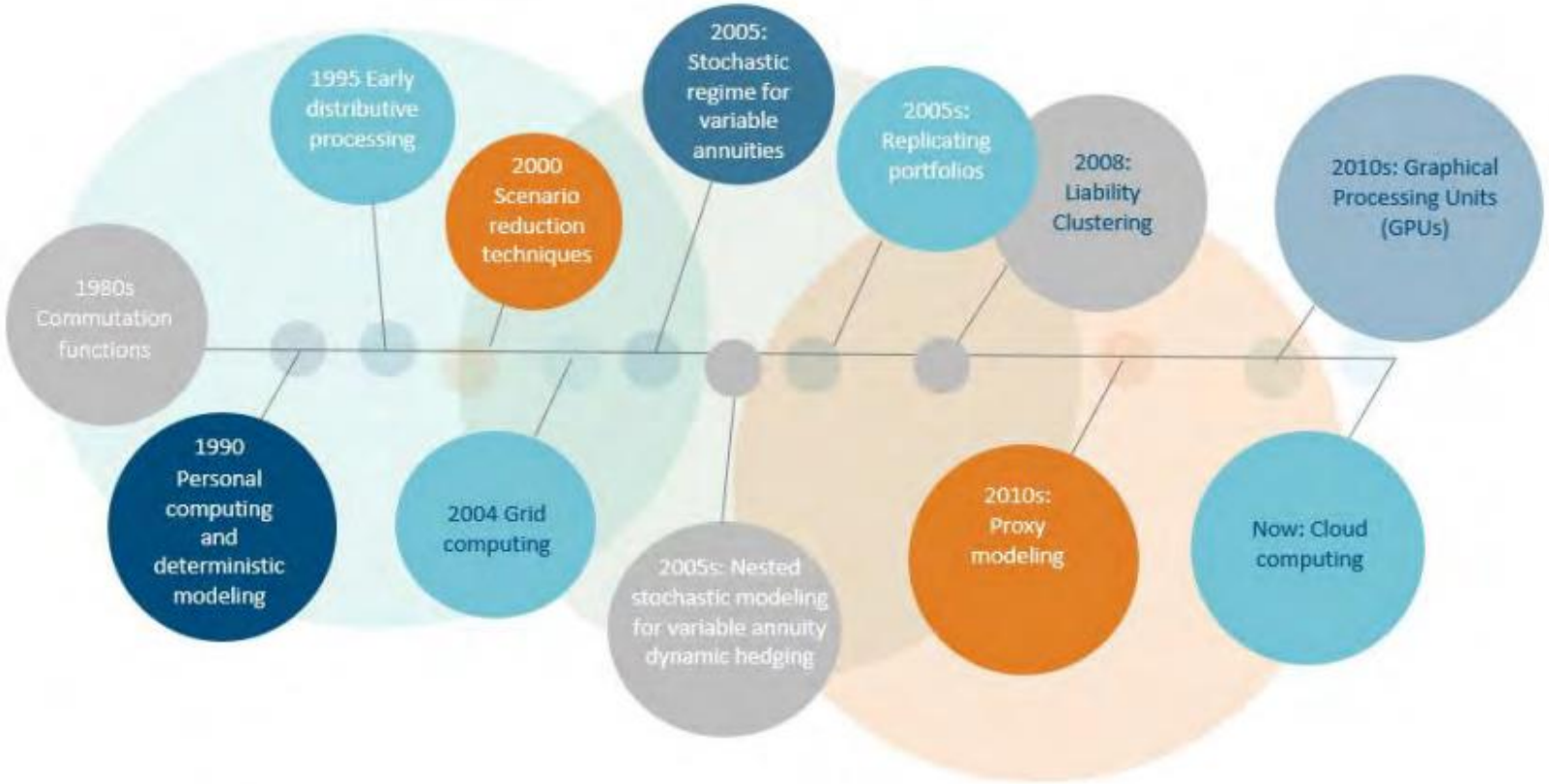
Ethical Use of AI for Actuaries

- Social context
- AI Risks
  - Data – imperfect or misunderstood
  - Bias in AI development process
  - Process automation can have unintended consequences
  - Amateur development
- Even with similar biases, humans will make different decisions
- Organizational level considerations

https://www.soa.org/globalassets/assets/files/resources/research-report/2019/ethics-ai.pdf

Cloud Computing

USE OF TECHNOLOGY IN THE INSURANCE SECTOR, HISTORICAL MILESTONES



https://www.soa.org/globalassets/assets/files/resources/research-report/2019/cloud-computing.pdf

Technology in Microinsurance

Applications to Address Microinsurance Challenges	
Remote sensing (incl satellites)	Administrative Systems
Big data, machine learning, data mining	Smartphone apps
Pricing tools	Internet of Things
Mobile networks	Blockchain
Telemedicine	Mobile Money
Digital marketing channels	Airtime
AI and chatbots	

https://www.soa.org/globalassets/assets/files/resources/research-report/2019/2019-technology-microinsurance.pdf

SOCIETY OF ACTUARIES