



Analytics and Applications

Lecture 1: Introduction to Data Science and Business Analytics

First, let me introduce myself...

Prof. Dr.
Wolfgang
Ketter



- **Chaired Professorship at University of Cologne and Rotterdam School of Management**
- **Director of the Institute of Energy Economics (EWI) at the University of Cologne**
- **Director of the Erasmus Centre for Future Energy Business and Learning Agents Research Group at Rotterdam School of Management**
- **Visiting Professorship at UC Berkeley**
- **PhD in Computer Science and Information Systems from the University of Minnesota**

Nastaran Naseri, the TA for this course

Nastaran
Naseri

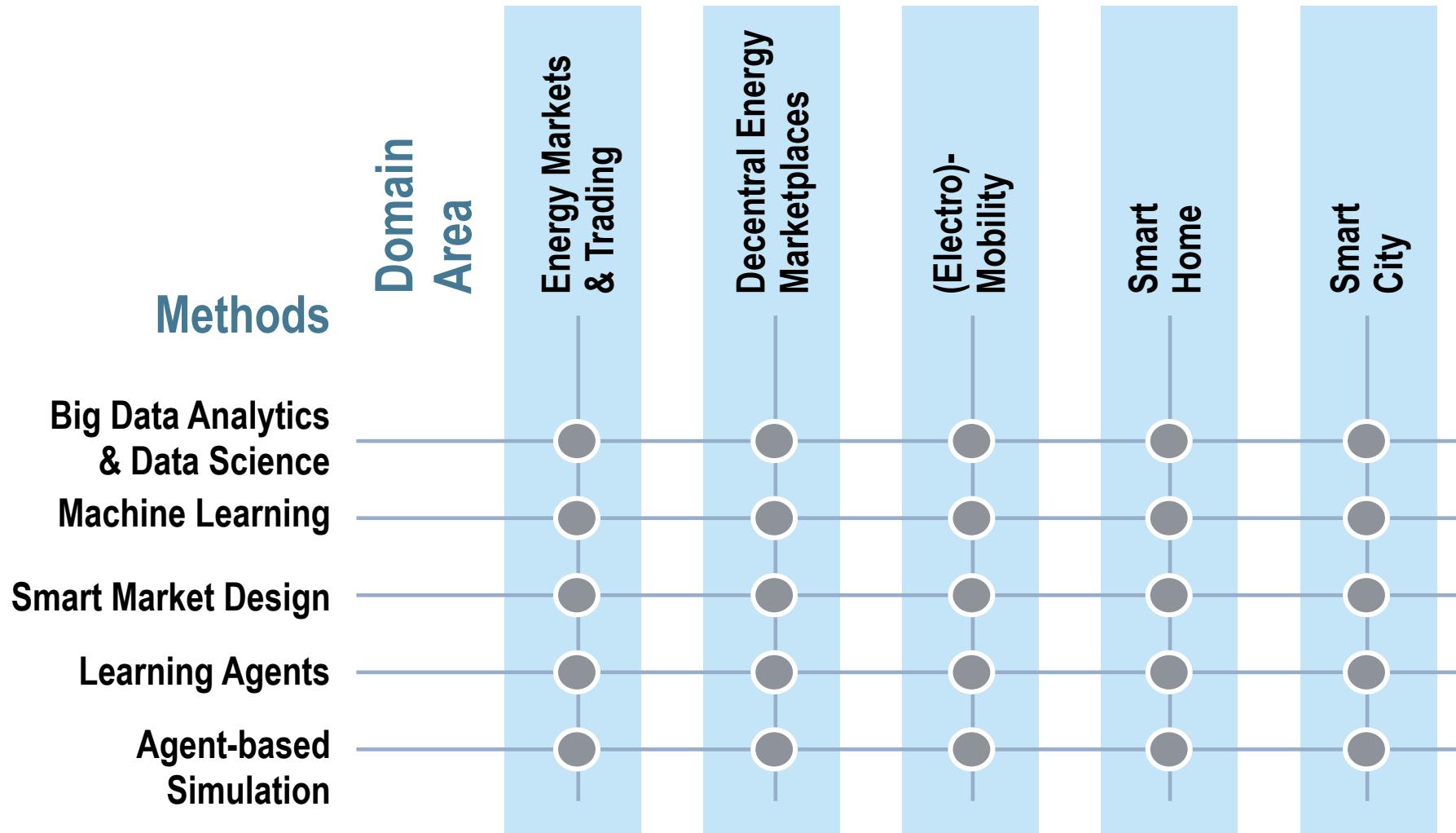


- **PhD Researcher** at the Chair for Information Systems for Sustainable Society
- B.Sc. And M.Sc. In **Industrial Engineering** from Ferdowsi University, Iran
- Two years experience in discrete event simulation and game design for educational purpose
- Research interest in analytical solution and market design for implementation of flexibility and storage in electricity markets
- Methodological focus on machine learning and mathematical programming

We interface and **collaborate closely** with **international research groups** from top institutions around the world



Our research is highly interdisciplinary – We apply **big data**, **data science** and **machine learning** to societal challenges in **energy** and **mobility**



In these domains we cover a broad spectrum of research topics – An overview

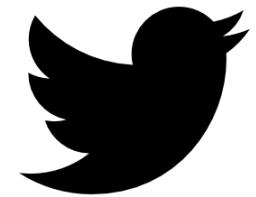
- 1 **Energy Markets & Trading** – Large-scale agent-based simulation of energy markets to derive policy and strategy recommendations; **Price** and economic regime **forecasting** in energy markets; **Market design**
- 2 **Decentralized Energy Markets** – Design of decentral marketplaces for peer-to-peer energy trading in industrial and residential context; Design of **energy cooperative models**
- 3 **(Electro)-Mobility** – Defining **solutions** for the **integration challenge** of electric vehicles into the **power system** (virtual power plants, smart charging); **Forecasting models** for load and storage availability; **New business models**
- 4 **Smart Home** – Identification of load balancing and shifting potential; Development of **intelligent learning agents** for **smart home energy management**
- 5 **Smart City** – Data-driven **public transport electrification strategies** (e.g. Rotterdam electric bus strategy); Charging **infrastructure placement** based on real-world mobility patterns

Stay up-to date with our research via our [homepage](#) or Twitter and feel free to contact us via e-mail



Visit our website at is3.uni-koeln.de to learn more about our **research**, our **teaching**, **job openings** and new **projects**

Follow us on **Twitter** at [@IS3_UniCologne](https://twitter.com/IS3_UniCologne) to stay up to date with recent publication and presentations of our group as well as new developments in **#DataScience**, **#BigData** and **#MachineLearning** for **#Sustainability**



Now let's get
to know you!

- Join at www.kahoot.it
or with the Kahoot! App
- Game pin: **6456940**

Kahoot!

Now let's get to know you!



What is your background?

1. Information Systems
2. Mathematics or Statistics
3. Management
4. Software engineering
5. Others

Now let's get to know you!



Which programming languages are you familiar with?

1. Python
2. Java
3. Julia
4. R
5. Others



Now let's get to know you!



How familiar are you with Data Science and Machine Learning?

1. Advanced
2. Intermediate
3. Basic
4. No fundamental knowledge



CALL FOR APPLICATIONS FOR THE ROLE OF STUDENT ASSISTANT

The Department for Information Systems for Sustainable Society ([IS3](#)) is inviting applications for the role of student assistant (8 h/week) starting this semester.

We are looking for advanced **Bachelor's or early stage Master's students** with information systems, computer science, economics or business management backgrounds who are interested to join our team.

Role : Development Student Assistant

- As a development student assistant, you will support our researchers in data acquisition, data base management, machine learning and analytics tasks and work hands-on with large-scale research simulation platforms, such as [Power TAC](#).
- You should have a strong technical background (e.g., Java, Python), analytical skills and readiness to acquire new technical skills.

Please send any inquiries and your applications (CV summarized in one .pdf file or link to your LinkedIn Profile + overview of grades) to is3-teaching@wiso.uni-koeln.de



Agenda

Course Content & Goals

Course Guidelines & Admin

Introduction to Data Science

Course Goal



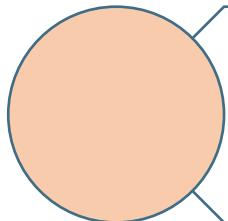
This course will help you to possess valuable practical analytical skills that will equip you with a competitive skill in almost any contemporary workplace.

Course Objectives

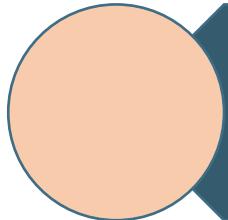
The course will provide participants with the following skills and knowledge:

- Become familiar with the potential and dangers of data mining in todays data rich-environment
- Gain experience using key data mining methods of classification, prediction, data reduction and exploration
- Know how to decide when to use which technique
- Be able to implement major data mining techniques using software
- Become a smart and critical user and consumer of data mining techniques
- Gain the intellectual capital required to provide responsible business analytics services
- Be conscious of humanistic and societal implications of deploying data mining in organizations

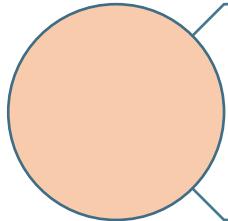
Agenda



Course Content & Goals



Course Guidelines & Admin



Introduction to Data Science

Not on ILIAS? – Make sure you join!

- ILIAS Course Name: **Analytics and Application** (can be accessed [here](#))
- ILIAS Password: **AA_2022!**

We will now run through some of the key admin points of the course syllabus – Make sure you download a copy from ILIAS and read it thoroughly

#	Title	Topics	Recommended reading
#1 Oct 12	Kick-off	Introducti & announ	
** #2 Oct 19	Data Mining Process	Core idea diction	
#3 Oct 26	Supervised Learning: Regression (1/2)	Linear reg	
** #4 Nov 02	Supervised Learning: Regression (2/2)	Performar	
#5 Nov 09	Supervised Learning: Classification	Naive Bay Logistic re	
** #6 Nov 16	Supervised Learning: Classification and Decision Trees	Classificat evaluatio	
#7 Nov 23	Supervised Learning: Combining methods	Ensembles	
** #8 Nov 30	Supervised Learning: Artificial Neural Network	Intro to A	
#9 Dec 07	Unsupervised Learning: Cluster Analysis	Hard Clu Hierarchic	
** #10 Dec 14	Unsupervised Learning: Cluster Analysis	Soft clust with PCA	
#11 Dec 21	Unsupervised Learning: Association Rules	Introducti filtering	
** #12 Jan 11	Times Series Analysis	Handling Smoothing	
#13 Jan 18	Social Network Analysis	Primer Gi & Eigenve	
** #14 Jan 25	Text Mining & NLP	Introducti Text Mini	
#15 Feb 01	Wrap-up	Course sy exam pre	
Feb 01	Deadline: Group Projects	Submit y (12.00h)	
Feb 07	Exam 1	15:00 to inspection	
Mar 10	Exam 2	15:00 to inspection	

Table 1. Tentative Schedule (** =

Analytics and Application [AA] 2021/22

Master of Science WI / IS
Information Systems II
Faculty of Management, Economics, and Social Sciences
Department of Information Systems for Sustainable Society
University of Cologne
Version - September 20, 2021

Instructor Prof. Dr. Wolfgang Ketter
TA Nastaran Nasori
Website [www.is3.uni-koeln.de](http://is3.uni-koeln.de) and ILIAS Class Location H113 & Zoom

Welcome to **Analytics and Application [AA]**. This course and the accompanying reading materials aim to provide you with knowledge and skills required for data analytics using information systems that drive business success.

Business analytics is the use of data-driven decision making. Companies, governments, and other organizations now collect and have access to large amounts of data about suppliers, clients, employees, citizens, transactions, etc. Data Mining and predictive analytics provide a powerful toolkit for detecting actionable patterns in data and generating predictions. These methods are used in many industries: Mobile companies use their customer database to predict customer churn or to personalize SMS messages for improving customer service; Financial institutions use past loan data to predict defaulting chance for loan applicants; Charities use data from a campaign in one location to target the right people in another location; Politicians use databases of supporters to segment and best target each audience; Movie rental and e-Commerce websites provide recommendations based on users' online behavior; Renewable energy providers use weather forecasts to predict their electricity generation to better trade in different types of markets.

Most business analytics applications are geared towards benefiting the company, often at the expense of the individual, community, and society. The focus of this course is on human and socially-responsible business analytics, especially with the focus on sustainability.

In this course we will work with real business problems and real data. We will examine types of questions that data mining can answer and will develop a variety of data-driven tools to answer these questions. The emphasis is on understanding the concepts and logic behind a wide set of data mining techniques and their relation to specific business analytics situations. The course is not about mastering the theoretical underpinnings of the techniques. The gained knowledge will be applied on a business analytics team project that encapsulates the learnings, while is not expected to master the theoretical underpinnings of the techniques.

You will learn about the process of data analytics. You will learn to identify problems, to define the structure of an information system, to evaluate competing solutions for a business problem, and you will start to "speak business and analytics". You will gain an insight into how important these activities are in creating information systems that are truly aligned with business needs. Throughout the course, you will also learn about selected topics in sustainability, especially in terms of renewable electricity.

Some of the most important thoughts that we will present throughout the lecture are summarized in this reader and the references therein. This material is designed to help you find additional references, and to help you recapture what was taught in the lectures. It is not designed, however, to serve as an exclusive source of exam preparation material. In our lectures and workshops, we will present facts, examples, and teach you skills that are not on the following pages, and we reserve the right to ask you about them during the final exam of your assignments. Please make sure you take advantage of all modes of learning that we offer in this course!

Cologne, September 20, 2021

Prof. Dr. Wolfgang Ketter, Nastaran Nasori
Department of Information Systems for Sustainable Society

Key points covered

- Lecture Schedule
- Lecture Content & Goals
- Recommended Reading
- Assessment Guidelines & Grading
- Other Course Guidelines (setup, communication, attendance, submitting work, etc.)

Weekly Lectures and Bi-weekly Workshops

Wed
Lectures
Tutorials

Lectures: Wed, 14:00 - 15:30

Tutorials: Wed, 15:45 - 17:00

Check Syllabus for
detailed schedule!

Note: In this semester, we offer mostly **biweekly in person workshops**, and for other sessions we provide you **recorded workshop**. Please check the syllabus for the schedule.

Syllabus: Course Schedule

#	Title	Topics	Recommended reading
#1 Oct 12	Kick-off	Introduction to course contents & objectives; admin & announcements	Shmueli et al., Ch. 1
** #2 Oct 19	Data Mining Process	Core ideas in Data Mining & Explanation vs. Prediction	Shmueli et al. Ch. 2
#3 Oct 26	Supervised Learning: Regression (1/2)	Linear regression & Polynomial regression	Shmueli et al., Ch. 6
** #4 Nov 02	Supervised Learning: Regression (2/2)	Performance evaluation & Regularization	Shmueli et al., Ch. 6
#5 Nov 09	Supervised Learning: Classification	Naive Bayes classifier & Maximum Likelihood & Logistic regression	Shmueli et al., Ch. 8 & 10
** #6 Nov 16	Supervised Learning: Classification and Decision Trees	Classification tree & Regression tree & Performance evaluation	Shmueli et al., Ch. 9
#7 Nov 23	Supervised Learning: Combining methods	Ensembles: Bagging, Boosting & Uplift modeling	Shmueli et al., Ch. 13
** #8 Nov 30	Supervised Learning: Artificial Neural Network	Intro to Artificial Neural Networks & applications	Shmueli et al., Ch. 11
#9 Dec 07	Unsupervised Learning: Cluster Analysis	Hard Clustering methods: K-Mean, K-means ++, Hierarchical clustering	Shmueli et al., Ch. 15
** #10 Dec 14	Unsupervised Learning: Cluster Analysis	Soft clustering methods, Dimensionality reduction with PCA	Shmueli et al., Ch. 15
#11 Dec 21	Unsupervised Learning: Association Rules	Introduction to association rules & Collaborative filtering	Shmueli et al., Ch. 15
** #12 Jan 11	Times Series Analysis	Handling time series & Forecasting methods & Smoothing methods	Shmueli et al. Ch. 16-19
#13 Jan 18	Social Network Analysis	Primer Graph Theory & Dijkstra for SP problems & Eigenvector centrality & Link Prediction	Shmueli et al. Ch. 19
** #14 Jan 25	Text Mining & NLP	Introduction to Natural Language processing & Text Mining	Shmueli et al. Ch. 20
#15 Feb 01	Wrap-up	Course synthesis and announcements regarding exam preparations	none
Feb 01	Deadline: Group Projects	Submit your group projects via ILIAS by noon (12.00h)	none
Feb 07	Exam 1	15:00 to 16:00 , sign up via KLIPS by Jan. 27, inspection date on March 3 (upon request only)	none
Mar 10	Exam 2	15:00 to 16:00 , sign up via KLIPS by March 5, inspection date on April 14 (upon request only)	none

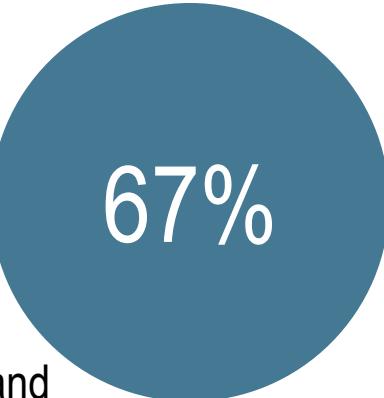
Table 1. Tentative Schedule (** workshops will be held in offline mode)



Assessment Guidelines

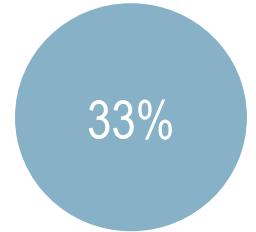
Written Exam

- **Scope:** 60-minutes, closed book
- **Content:** All materials provided in the Lectures, Tutorials, Required readings, and Assignments.
- **Format:** Written and multiple choices questions
- **Date:**
 - **Exam 1:** February 7
 - **Exam 2:** March 10

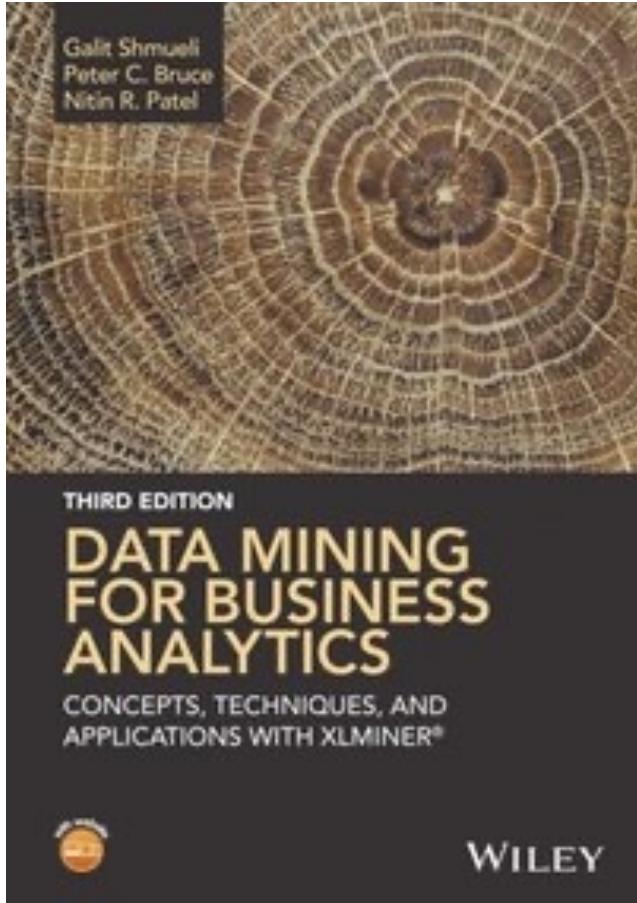


Team assignment

- **Set-up:** Teams of 5 students
- **Tasks:**
 - **Project report:** details the team project, from the business problem through the data mining problem and solution, to recommendations.
 - **Deliverables:** Project reports as well as the Python code.
 - **Deadline:** February 1

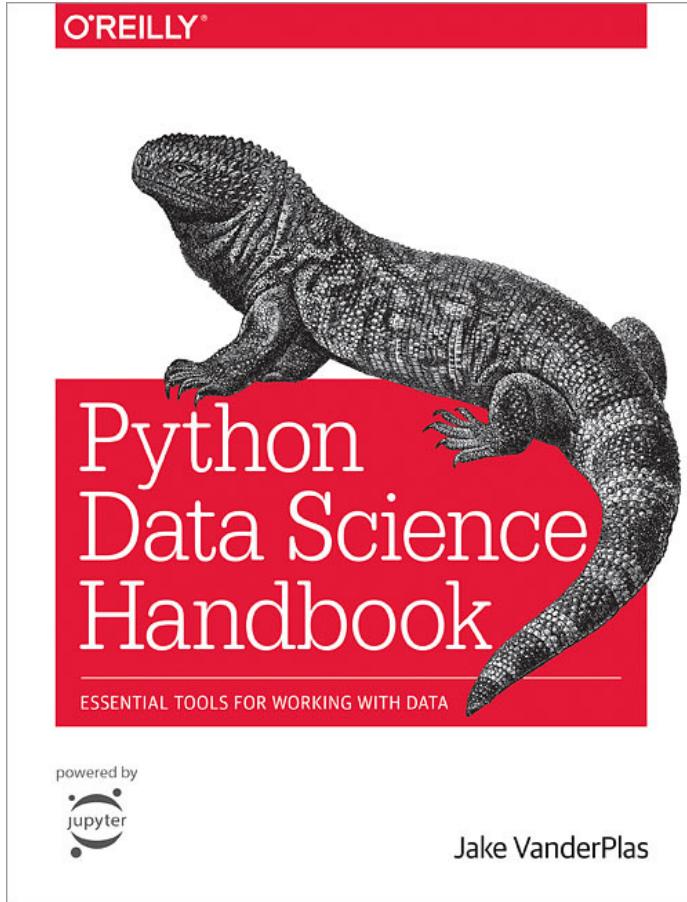


Core reading



- Data Mining for Business Analytics – Concepts, Techniques, and Applications with XLMiner
- Galit Shmueli; Peter C. Bruce; Nitin R. Patel
- Publisher: John Wiley & Sons

Recommended reading for learning Python



- Python Data Science Handbook – Essential Tools for Working with Data
- By Jake VanderPlas
- Publisher: O'Reilly Media

Contact



For general questions and enquiries on **research**, **teaching**, **job openings** and new **projects** refer to our website at www.is3.uni-koeln.de



For specific enquiries regarding this course contact us by sending an email to the **IS3 teaching** address at is3-teaching@wiso.uni-koeln.de

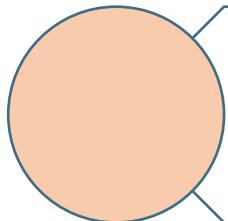
To help us process your request efficiently, use the following subject line format:

[AA] <Your request subject>

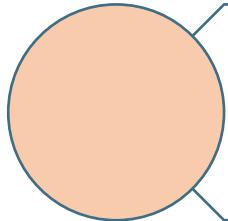


The full **syllabus** is available on ILIAS –
Make sure to **download** a copy and **read**
it thoroughly!

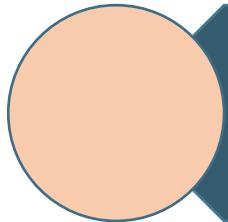
Agenda



Course Content & Goals



Course Guidelines & Admin



Introduction to Data Science

Some Examples of Data Science in Every Day Life

The screenshot shows a Google search results page. The search bar at the top contains the query "how do i know if i am intelligent". Below the search bar, there are tabs for "Web", "News", "Images", "Videos", "More", and "Search tools". A black oval highlights the text "About 12,80,00,000 results (0.46 seconds)". The search results list several links:

- 20 Scientifically-Proven Signs You're Smarter Than Average ...**
www.businessinsider.com/20-scientifically-proven-signs-youre-smarter-t... ▾
Dec 10, 2010 - Many scientific studies have been conducted to determine the cause of high IQs, and a lot of it boils down to genetics. According to researchers ...
- How do I know if I'm smart? - Quora**
https://www.quora.com/How-do-I-know-if-Im-smart
If you ever stop and think, "Hold on a minute, am I actually smart? ... "Among highly intelligent people, there are two kinds of minds, the sharp and the soft.
- How to know if I am intelligent - Quora**
https://www.quora.com/How-do-I-know-if-I-am-intelligent
First, read up on what it means to be intelligent, the good and the bad. ... How do I know if I am actually smart and my intelligence is not a delusion resulting from ...
- Are you Intelligent? - GoToQuiz.com**
www.gotoquiz.com/are_you_intelligent_3 ▾
Or are you the type of person who doesn't even know what the alphabet is? Or maybe ... If you have 5 apples and take away 2, how many do you have? 5. 7. 3. 2.
- Am I Dumb Test - Intelligence Test - How Smart Are You?**
www.am-i-dumb.com/ ▾
The Am I Dumb Test is a free intelligence test that will reveal your true ... Find out how much you're worth on the open human market · See if you're a ... Take the intelligence test and we'll tell you how you compare to the rest of the world.

<https://www.analyticsvidhya.com/blog/2015/09/applications-data-science/>

Search is not an issue anymore
→ getting results that are individually relevant is

Some Examples of Data Science in Every Day Life

The image is a composite of two screenshots illustrating data science applications.

Left Side (Netflix): A screenshot of the Netflix homepage. It features a "Congratulations!" message: "Movies we think You will ❤️". Below it, there's a grid of movie thumbnails with "Add" buttons. At the bottom, there's a "People You May Know" section with profiles for Harry Tamacas Perla and Gina Spano, each with a red arrow pointing to their profile pictures. The URL <https://www.analyticsvidhya.com/blog/2015/09/applications-data-science/> is visible at the bottom of this section.

Right Side (YouTube): A screenshot of the YouTube "Up Next" sidebar. It shows a list of recommended videos with their titles, uploaders, and view counts. The first video is "YouTube Rewind 2014: Behind the Scenes" by YouTube Spotlight with 3,212,337 views. The second is "#REWIND2013" by YouTube Spotlight with 106,428,985 views. The third is "Rewind YouTube Style 2012" by YouTube Spotlight with 154,722,053 views. The fourth is "Girls' Generation "I Got a Boy" Wins Video of the Year - Live at the YTMAs" by YouTube Spotlight with 350,421 views. The fifth is "Mind the Gap: The Making of unReal Episode 1" by Teton Gravity Research with a "Recommended for you" label. The sixth is "BEST NEWS BLOOPERS 2014 Part 2" by NewsBeFunny with 1,010,904 views.

~27 likely to create
quite accurate
personal profile

Some Examples of Data Science in Near Future!



<https://www.greentechmedia.com/articles/read/fully-autonomous-vehicles-decade-away-experts#gs.iwDj=Dc>

What is Data Science?

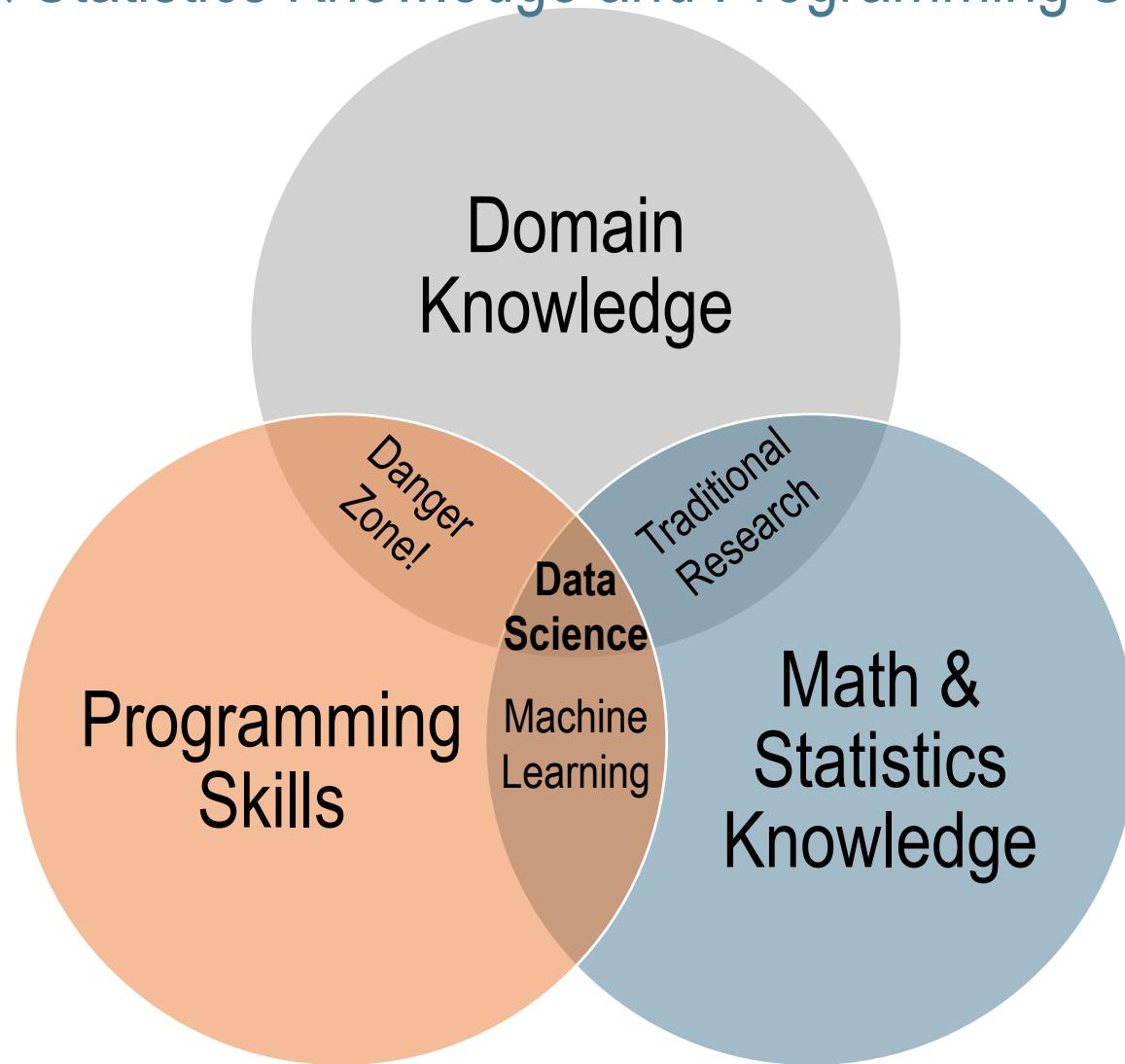
- Possible definition:

An interdisciplinary field of **computer science** and **statistics** that uses scientific methods, processes, algorithms and systems to **extract knowledge or insights** from **data** in various forms, either structured or unstructured

→ transform (raw) data into information

Data Science = statistics + data processing + machine learning + scientific inquiry +
visualization + business analytics + big data + ...

In essence a good Data Scientists needs to combine three core skills – Domain Knowledge, Math & Statistics Knowledge and Programming Skills



Source: Drew Conway

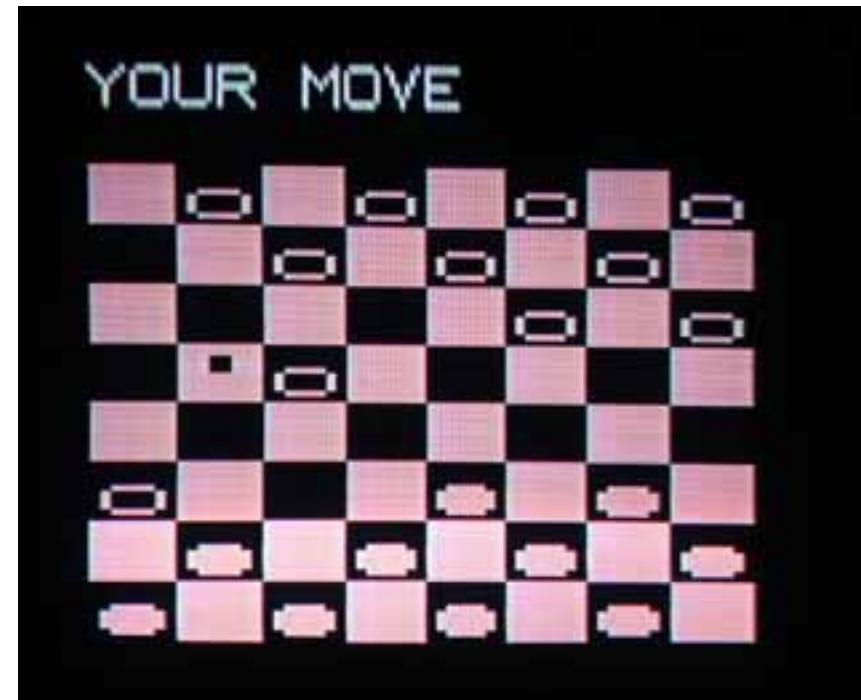
Information Systems for Sustainable Society (is3) | WiSo Faculty | Univ.-Prof. Dr. Wolfgang Ketter | 13.10.22

What is Machine Learning?

Even among machine learning practitioners, there isn't a well accepted definition of what is and what isn't machine learning. Here is some definitions of Machine learning:

- Arthur Samuel(1959): Field of study that gives computers the ability to learn without being explicitly programmed.

Example: Checkers playing program that sees tens of thousands of examples of board positions and learns over time what the good positions are.



<http://incompleteideas.net/book/first/ebook/node109.html>

What is Machine Learning?

- Tom Mitchell(1998) Well-posed Learning Problem: A computer program is said to learn from **experience E** with respect to some **task T** and some **performance measure P**, if its performance on T, as measured by P, improves with **experience E**.
- **Example:** For the checkers play, the experience E would be the experience of having the program play tens of thousands of games itself. The task T would be the task of playing checkers, and the performance measure P will be the probability that wins the next game of checkers against some new opponent.

Poll: ML definition

Suppose your email program watches which emails you do or do not mark as spam, and based on that, learns how to filter spam better. What is task T in this setting?

- Classifying emails as spam or not spam.
- Watching you label emails as spam or not spam.
- The number (or fraction) of emails correctly classified as spam/not spam.
- None of the above—this is not a machine learning problem.

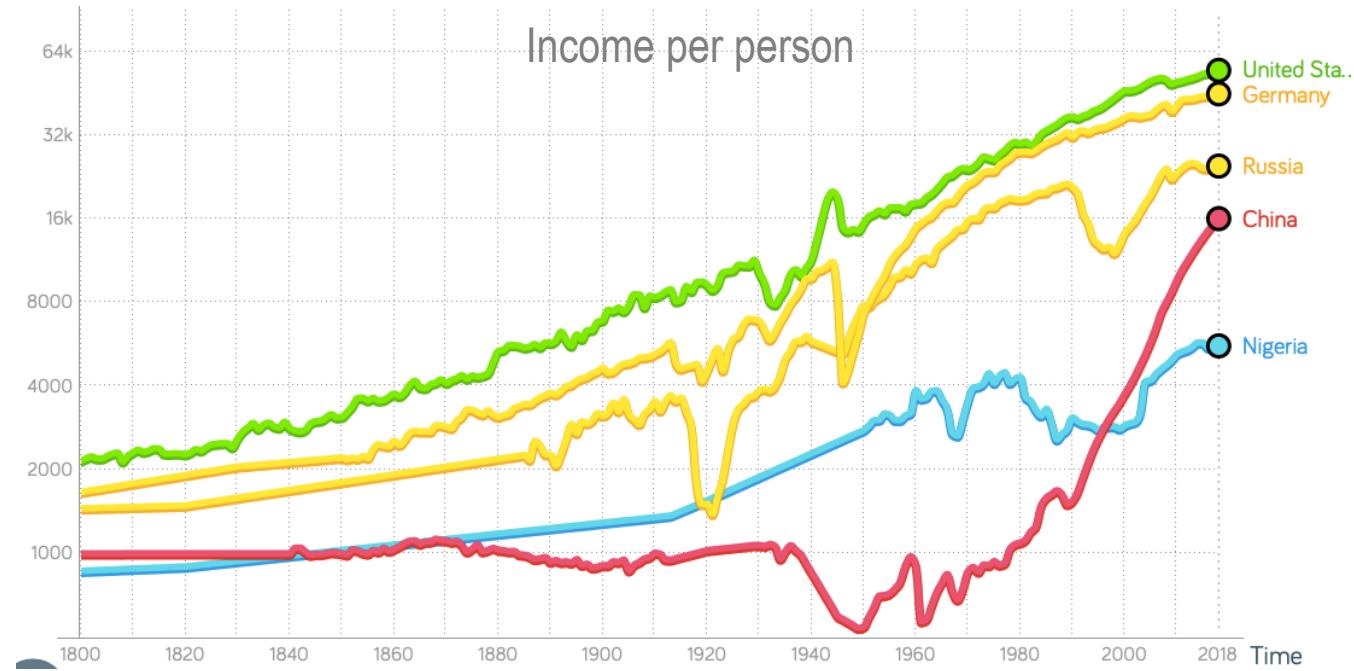
Answer

Suppose your email program watches which emails you do or do not mark as spam, and based on that, learns how to filter spam better. What is task T in this setting?

- Classifying emails as spam or not spam. This is *Task T*
- Watching you label emails as spam or not spam. This is *Experience E*
- The number (or fraction) of emails correctly classified as spam/not spam. This is *Performance P*
- None of the above—this is not a machine learning problem.

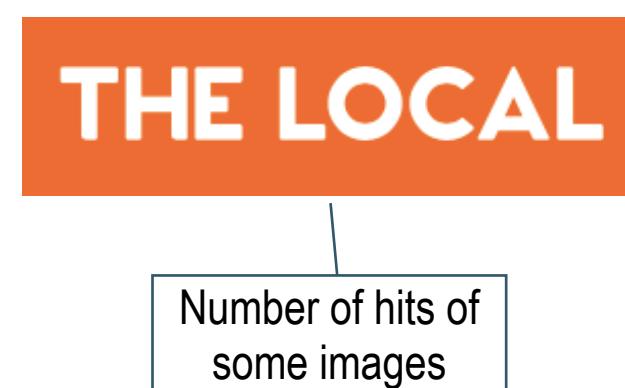
Business Intelligence

- *Business Intelligence (BI)* refers to data visualization and reporting for understanding “what happened and what is happening”.
- BI has evolved over years from static reports into more user-friendly and effective tools, such as creating interactive dashboards that allow the user not only to access the data but also interact with it. See www.gapminder.org

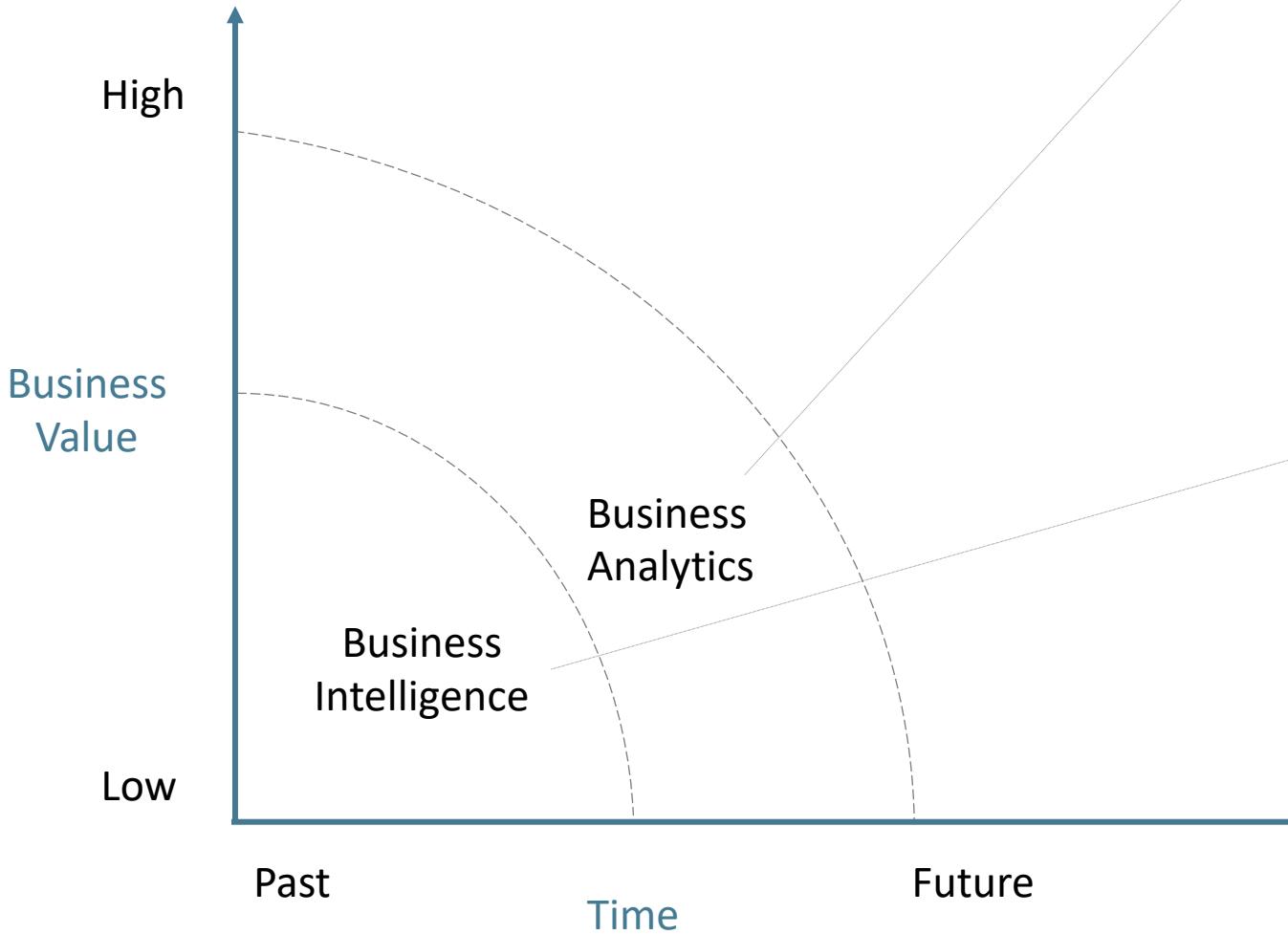


What is Business Analytics?

- *Business Analytics (BA)* is the practice and the art of bringing quantitative data to bear on decision making. However, the term can be defined differently for different organizations



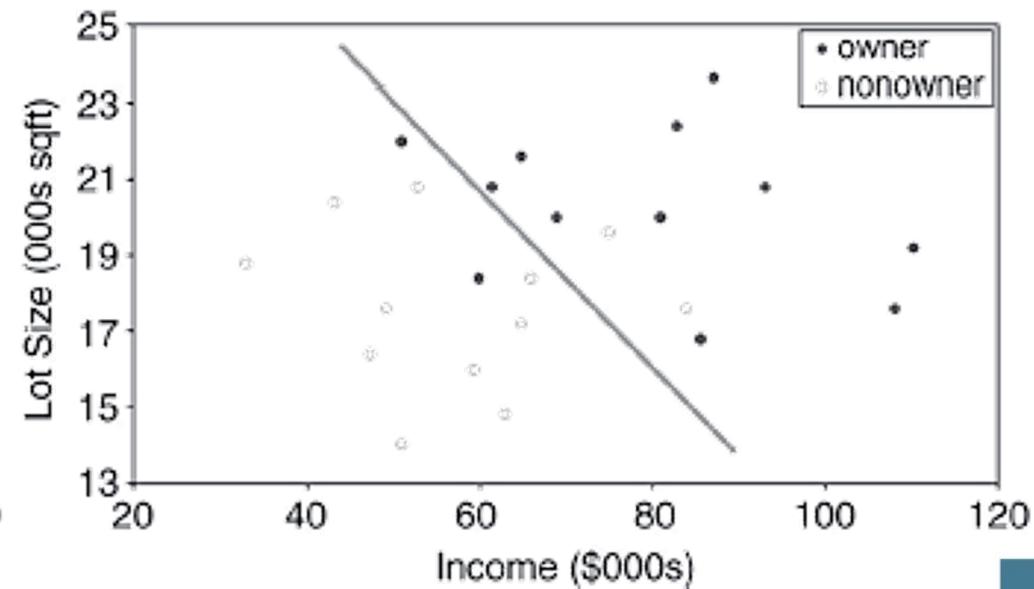
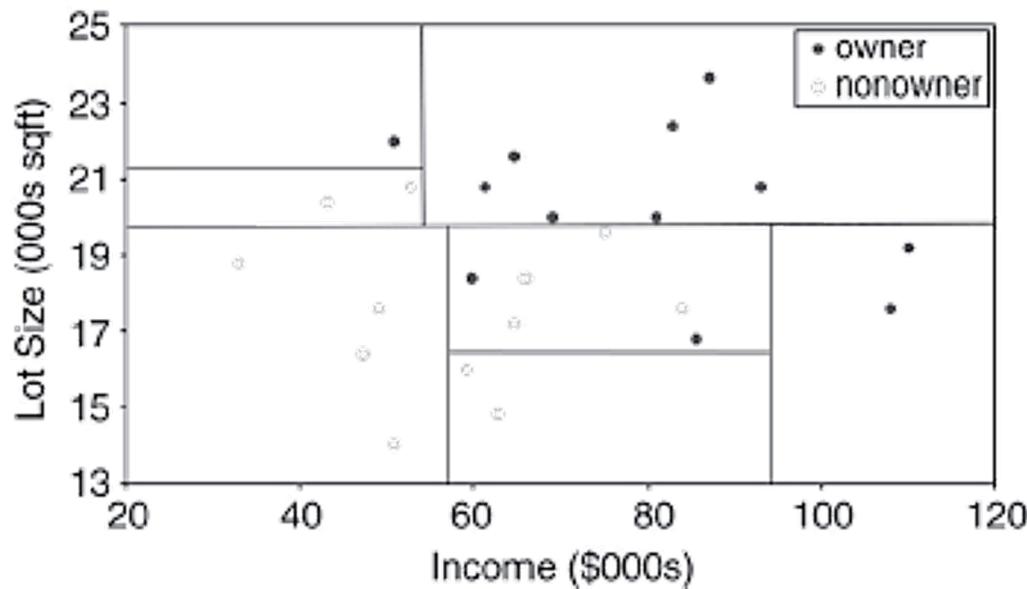
Business Analytics and Business Intelligence



Business Analytics	
Typical technologies	<ul style="list-style-type: none">Optimization, Predictive modeling, Forecasting, Statistical analysis
Data types	<ul style="list-style-type: none">Structured/ unstructured data, internal/ external data sets (massive data sets)
Common Questions	
	<ul style="list-style-type: none">What if ...?What if these trends continue?What is the optimal scenario for our business?What will happen next?
Business Intelligence	
Typical technologies	<ul style="list-style-type: none">Standard and ad hoc reporting, dashboards, alerts, queries
Data types	<ul style="list-style-type: none">Structured data, traditional sources (manageable data sets)
Common Questions	
	<ul style="list-style-type: none">What happened so far?How many did we sell?Where's the problem?

Why there are so many Methods?

- Each method has advantage and disadvantages. The usefulness of a method can depend on factors such as size of the dataset, the type of patterns that exist in the data, whether the data meet some **underlying assumptions** of the method and how noisy the data are, and the particular goal of the analysis.



2 cases of impactful data science – Health and Sustainability Analytics

1

Mapping the spread of the Corona Virus for actionable policy recommendations

- At the onset of the Corona Virus health crisis, computational virologists and data scientists provided **actionable policy recommendations for Western countries** from data – Possibly saving countless lives
- This **world-wide effort** has resulted in cutting-edge **data-driven decision support tools** to **inform policy decisions** in real time

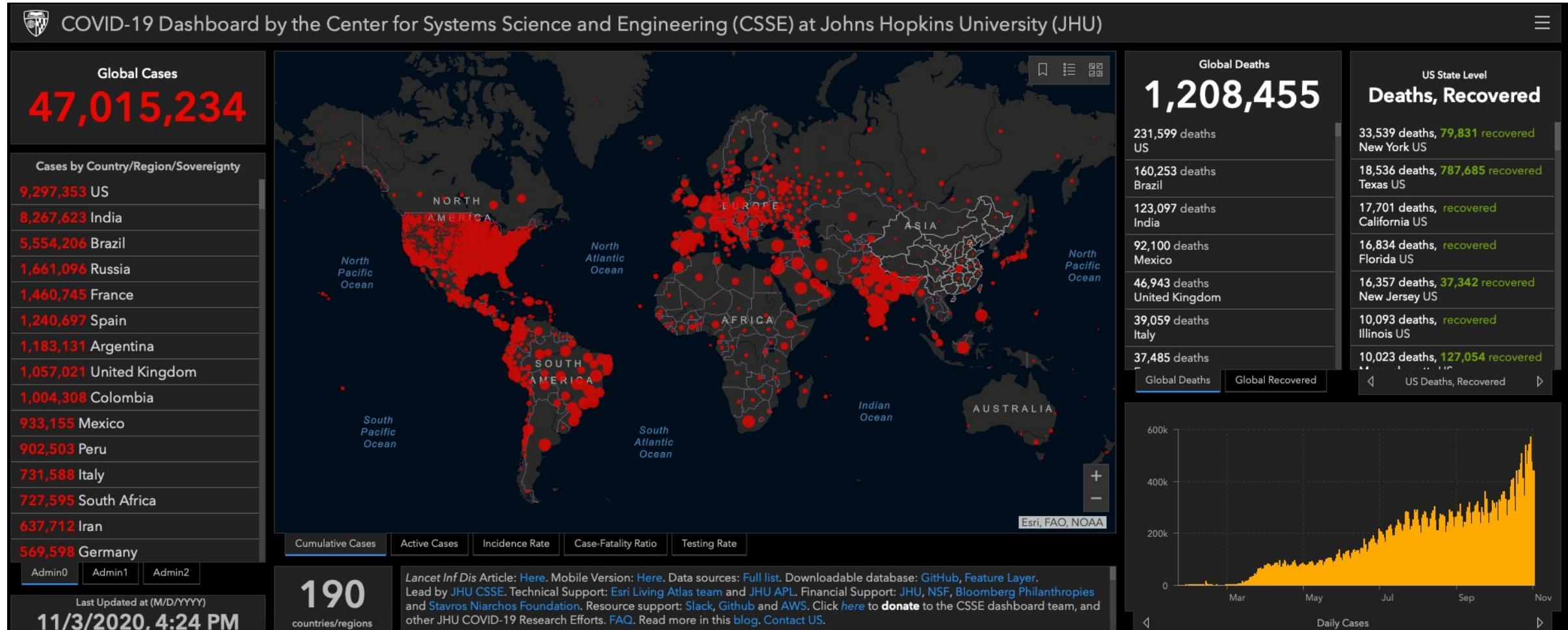
2

Virtual Power Plants of EVs to ensure grid stability

- Development of a **data-driven agent-based framework** to manage **EV vehicle fleet charging** in a **grid-optimal and profitable manner**
- **Battery resources** are committed in **real-time** to either the mobility or electricity services markets
- As a result the **electricity grid is stabilized** and **integration of renewable energy carriers** into the grid is facilitated



As a first step, data scientists provided close to real-time transparency on infection rates – All data available open-source for download

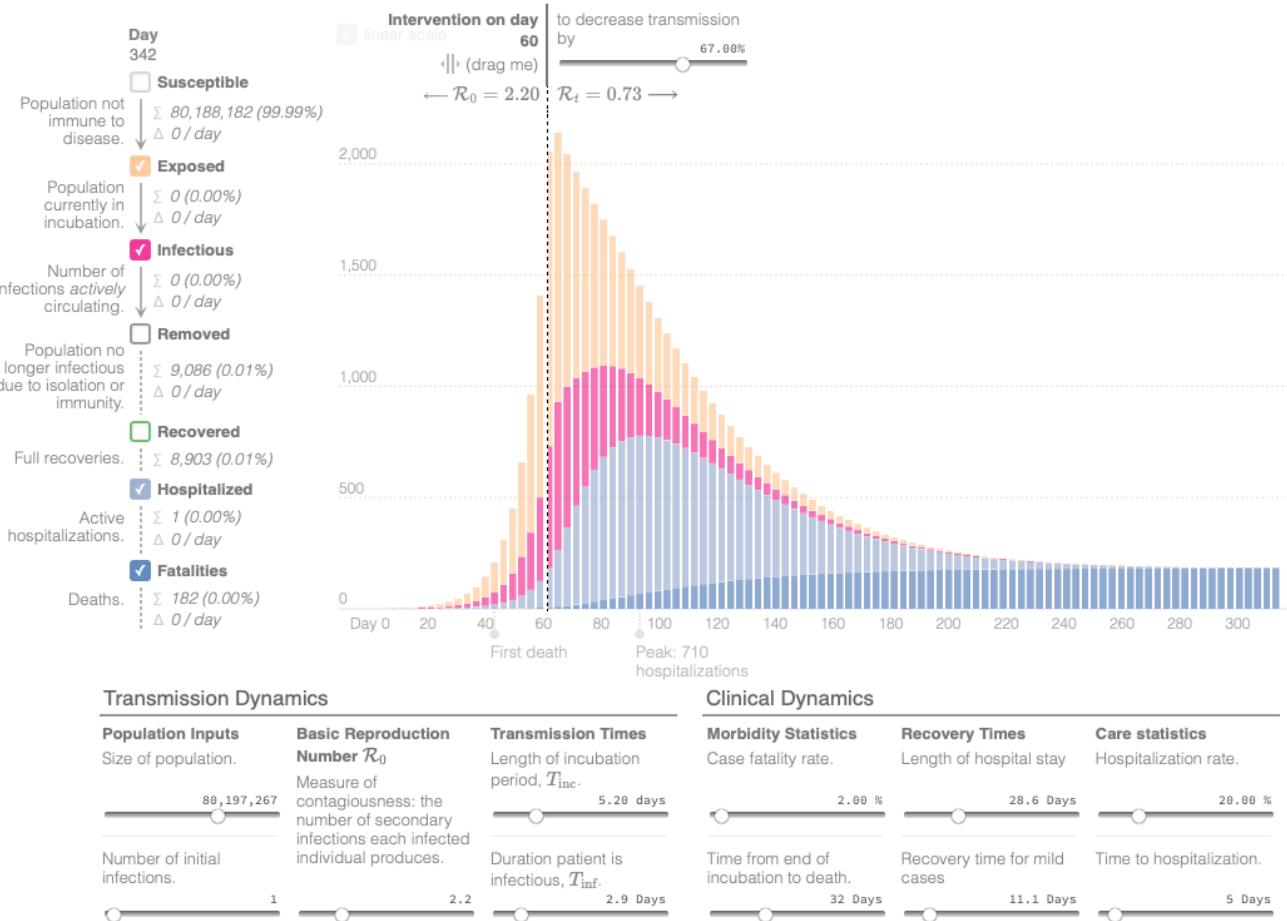


Source: Johns Hopkins University (<https://coronavirus.jhu.edu/map.html>)

Information Systems for Sustainable Society (is3) | WiSo Faculty | Univ.-Prof. Dr. Wolfgang Ketter | 13.10.22

They were among the first to warn of the exponential nature of the disease – Data Scientists at Open AI soon published the **Epidemic Calculator**

Epidemic Calculator



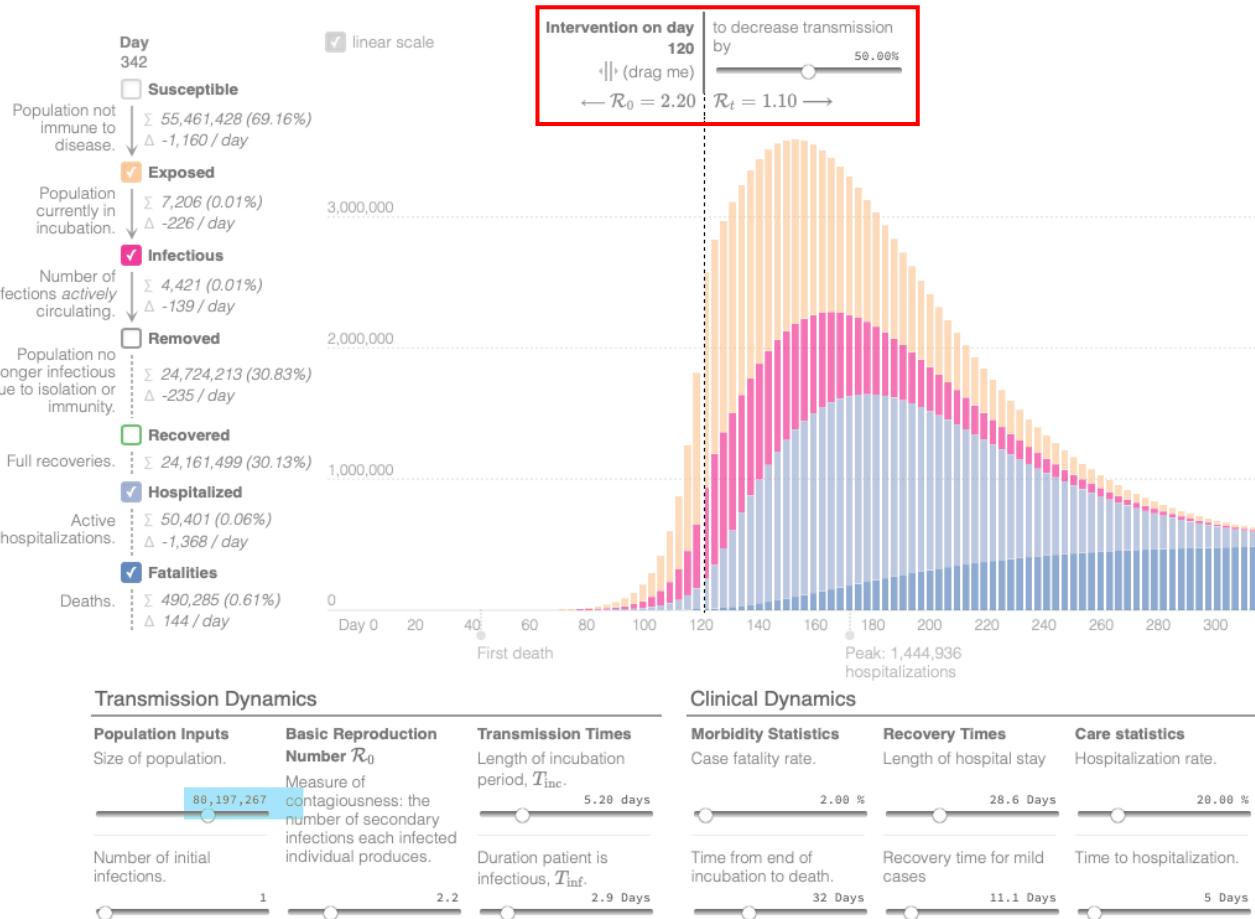
Source: Gabriel Goh (<http://gabgoh.github.io/COVID/index.html>)

Information Systems for Sustainable Society (is3) | WiSo Faculty | Univ.-Prof. Dr. Wolfgang Ketter | 13.10.22

- Researchers at Open AI soon published an Epidemic Calculator Tool, which allows to simulate various types of intervention in terms of
 - The reduction in R_0 , the reproduction number
 - The time of intervention
- We simulate an intervention at 60 days after case 1 that reduces transmission by 67% in a country of 80m inhabitants such as Germany

Note how case numbers rise exponentially as the intervention is further delayed and becomes less effective – Example Germany

Epidemic Calculator



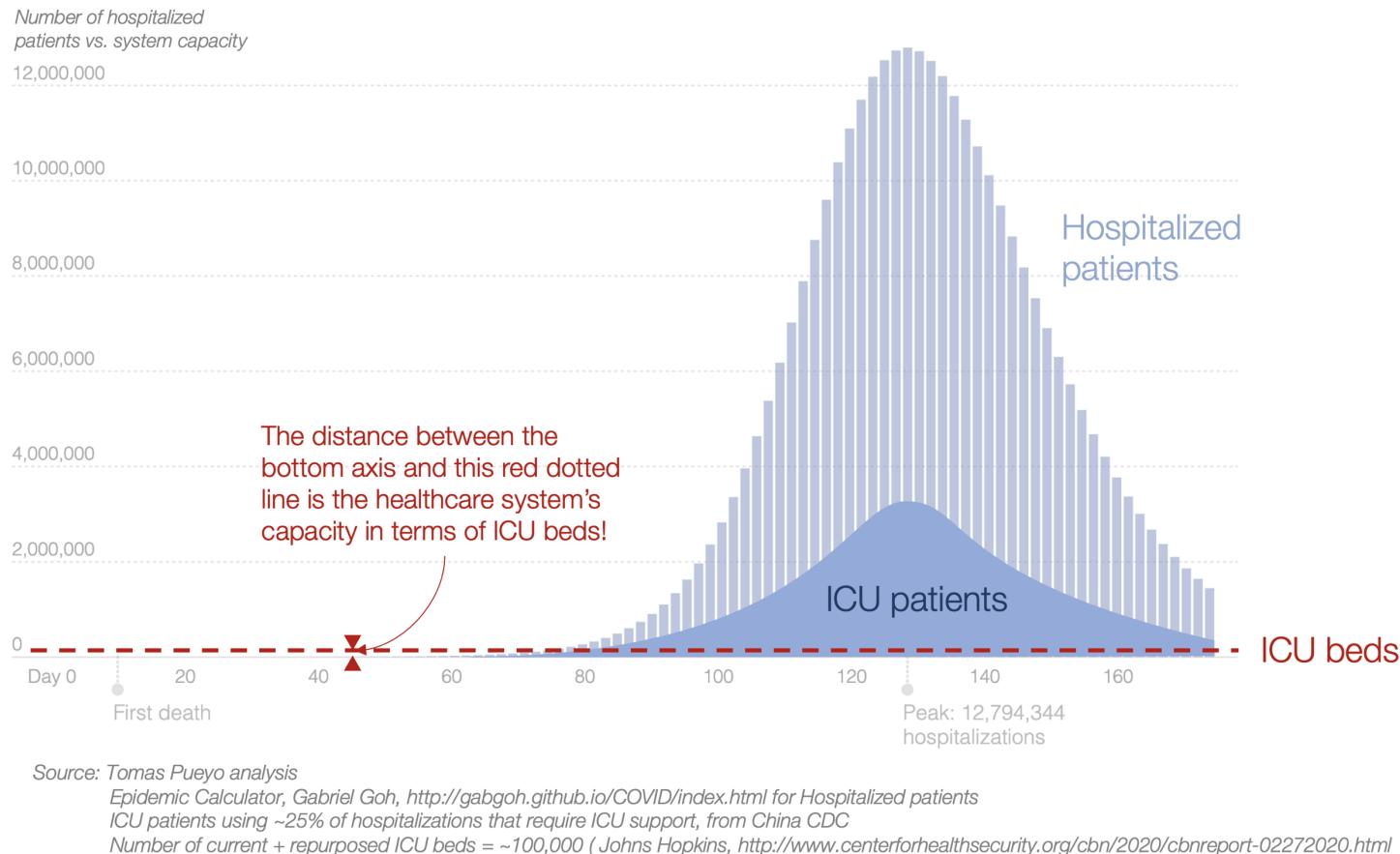
Source: Gabriel Goh (<http://gabgoh.github.io/COVID/index.html>)

Information Systems for Sustainable Society (is3) | WiSo Faculty | Univ.-Prof. Dr. Wolfgang Ketter | 13.10.22

- Let's look at that same country of 80m (e.g. Germany) inhabitants and **see what happens if we act later and less stringently**
- A **delay in intervention by 2 months** and a **reduction of effectiveness in transmission reduction form 67% to 50%** results in a **peak number of infections of >2.0m** (up from) with **peak hospitalization of circa 1.8m** – Way above capacity

This is a capacity constrained problem – Avoid overburdening of intensive care units (ICU) to minimize fatality rates!

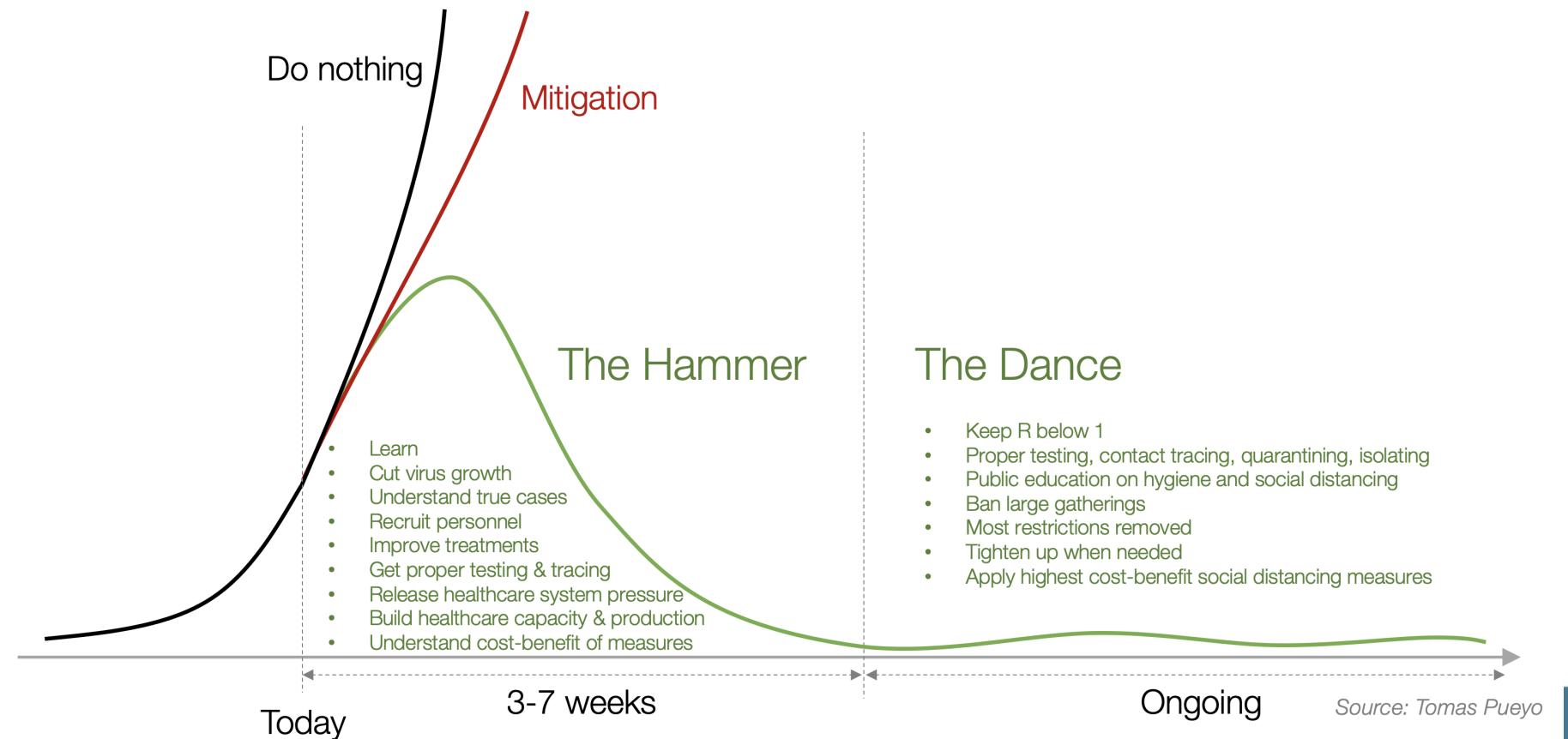
Chart 4: Hospitalized Coronavirus Patients vs. System Capacity (US)



- Countries that are prepared will see a fatality rate of ~0.5% (South Korea) to 0.9% (rest of China).
- Countries that are overwhelmed will have a fatality rate between ~3%-5%

It was also a data scientist who coined the phrase “Flattening the curve” – His predictions and recommendations have largely remained true

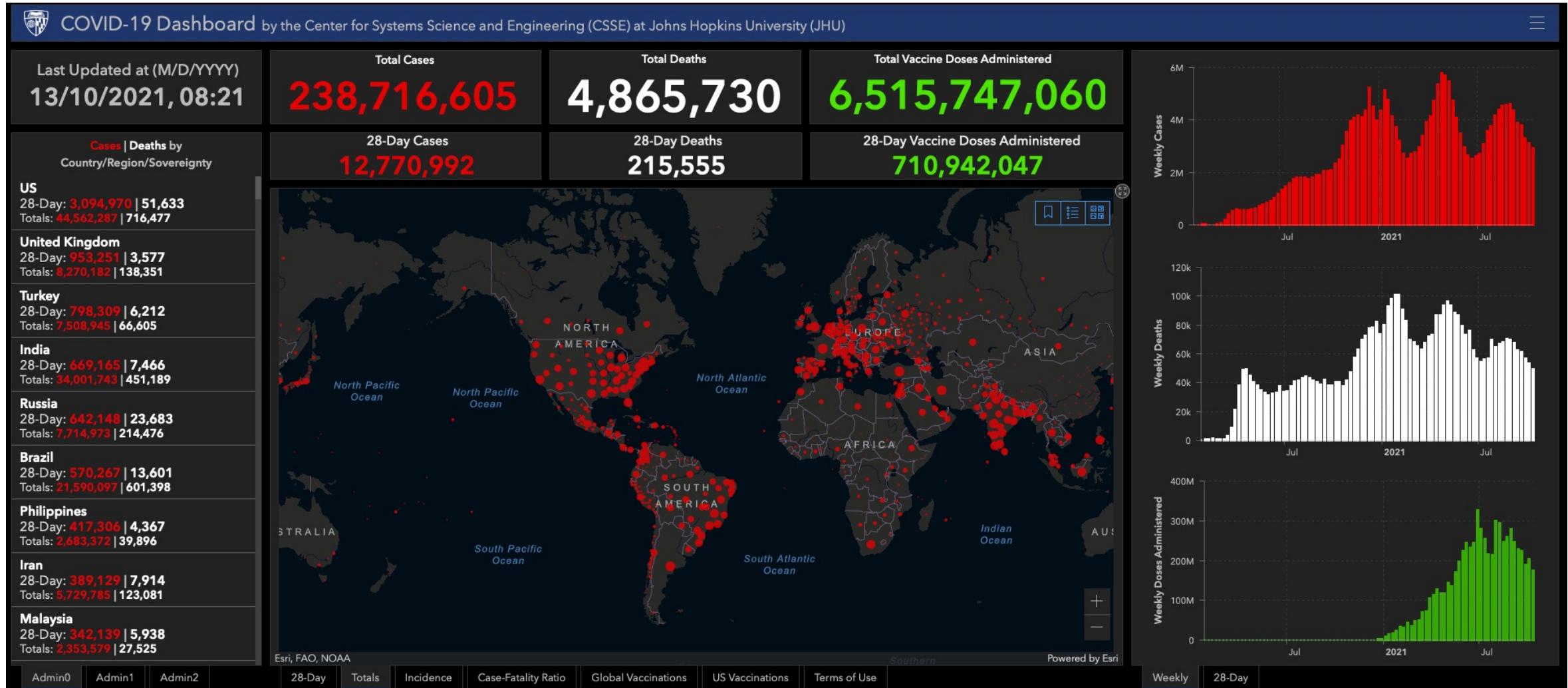
Chart 13: Suppression vs. Mitigation vs. Do Nothing – early on



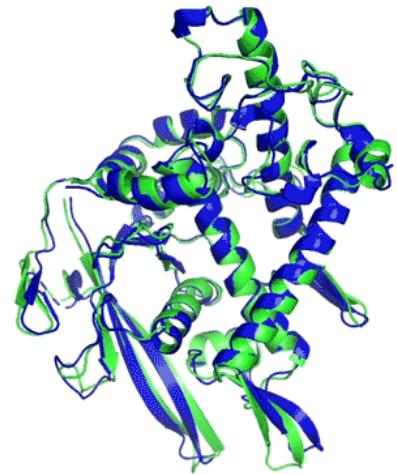
Source: Tomas Pueyo (<https://medium.com/@tomaspueyo/coronavirus-the-hammer-and-the-dance-be9337092b56>)

Information Systems for Sustainable Society (is3) | WiSo Faculty | Univ.-Prof. Dr. Wolfgang Ketter | 13.10.22

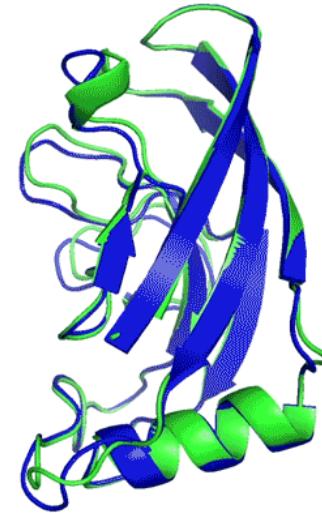
One year on the numbers look quite different but dashboards like Johns Hopkins' COVID-19 Map continue to provide decision support for the world



AlphaFold, an AI system that is recognised as a solution to "protein folding", a grand scientific challenge for more than 50 years.



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



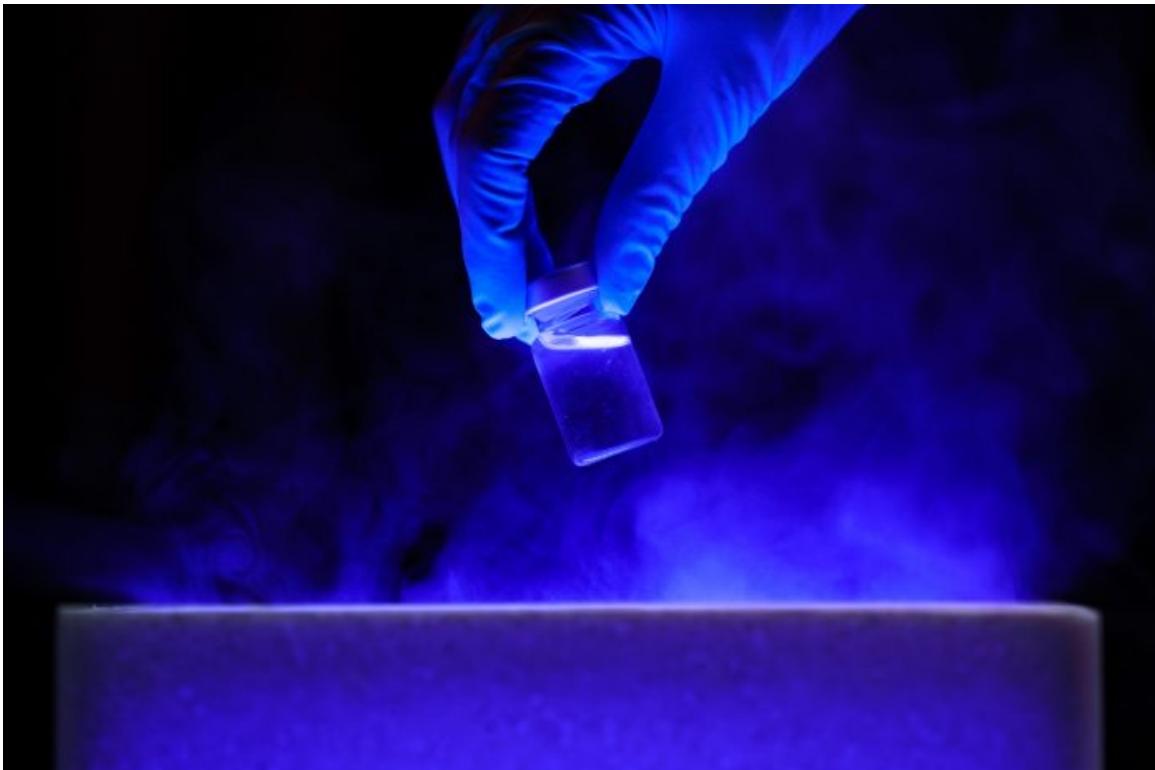
T1049 / 6y4f
93.3 GDT
(adhesin tip)

- Experimental result
- Computational prediction

Source: <https://deepmind.com/research/case-studies/alphafold>

- AlphaFold is a groundbreaking AI tool can accurately predict 3D models of protein structures.
- AlphaFold could predict one of the coronavirus proteins with a very high accuracy (98%).
- This tool can predict the structure of millions of unknown proteins and help tackling disease and more quickly find new medicines.

Data scientist took a further step: Scientists invent AI that creates COVID vaccine candidates within seconds



- Essentially, a machine-learning model can accomplish vaccine design cycles that once took months or years in a matter of seconds and minutes.
- When applied to SARS-CoV-2, the computer model quickly eliminated 95% of the compounds that could've possibly treated the pathogen and pinpointed the best options.

Source: <https://www.openaccessgovernment.org/covid-vaccine-candidates/103053/>

2 cases of impactful data science – Health and Sustainability Analytics

1

Mapping the spread of the Corona Virus for actionable policy recommendations

- At the onset of the Corona Virus health crisis, computational virologists and data scientists provided **actionable policy recommendations for Western countries** from data – Possibly saving countless lives
- This **world-wide effort** has resulted in cutting-edge **data-driven decision support tools** to **inform policy decisions** in real time

2

Virtual Power Plants of EVs to ensure grid stability

- Development of a **data-driven agent-based framework** to manage **EV vehicle fleet charging** in a **grid-optimal and profitable manner**
- **Battery resources** are committed in **real-time** to either the mobility or electricity services markets
- As a result the **electricity grid is stabilized** and **integration of renewable energy carriers** into the grid is facilitated



2 – EV VPPs for grid balancing



What does it cost?

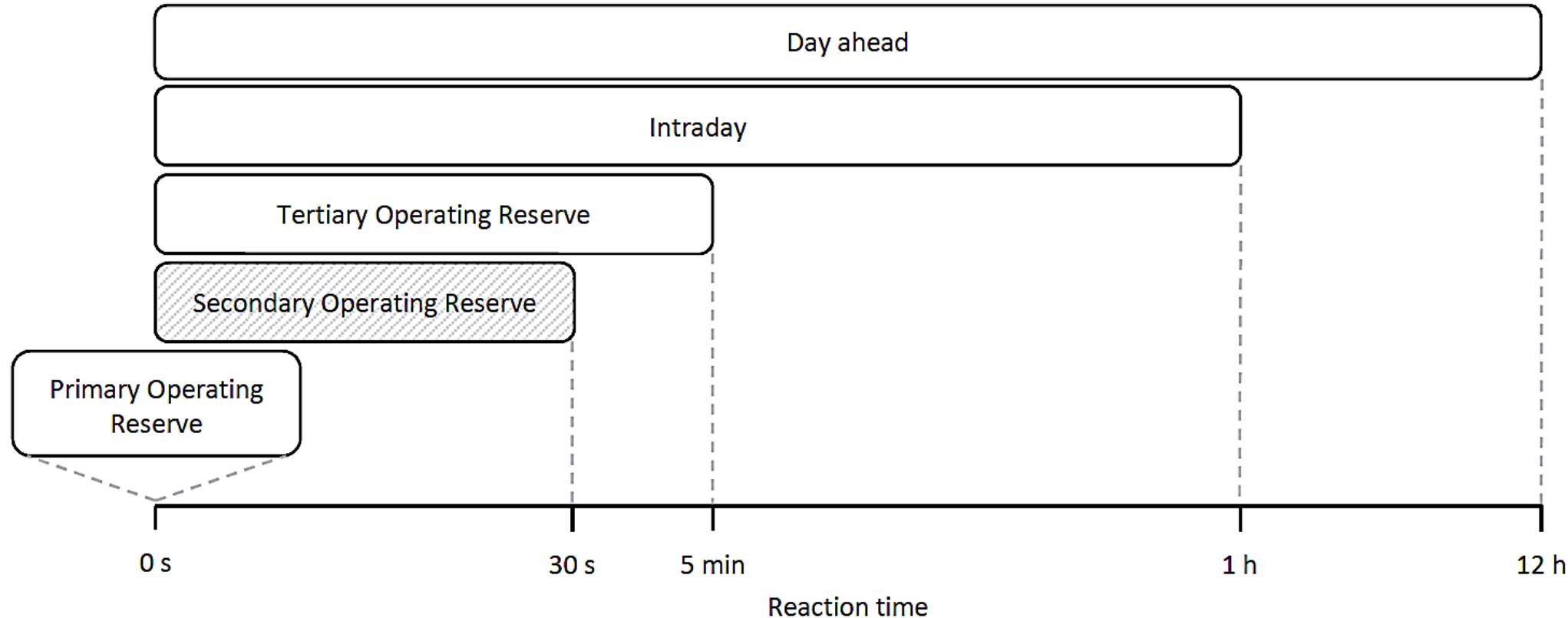
€ 0.29
per minute



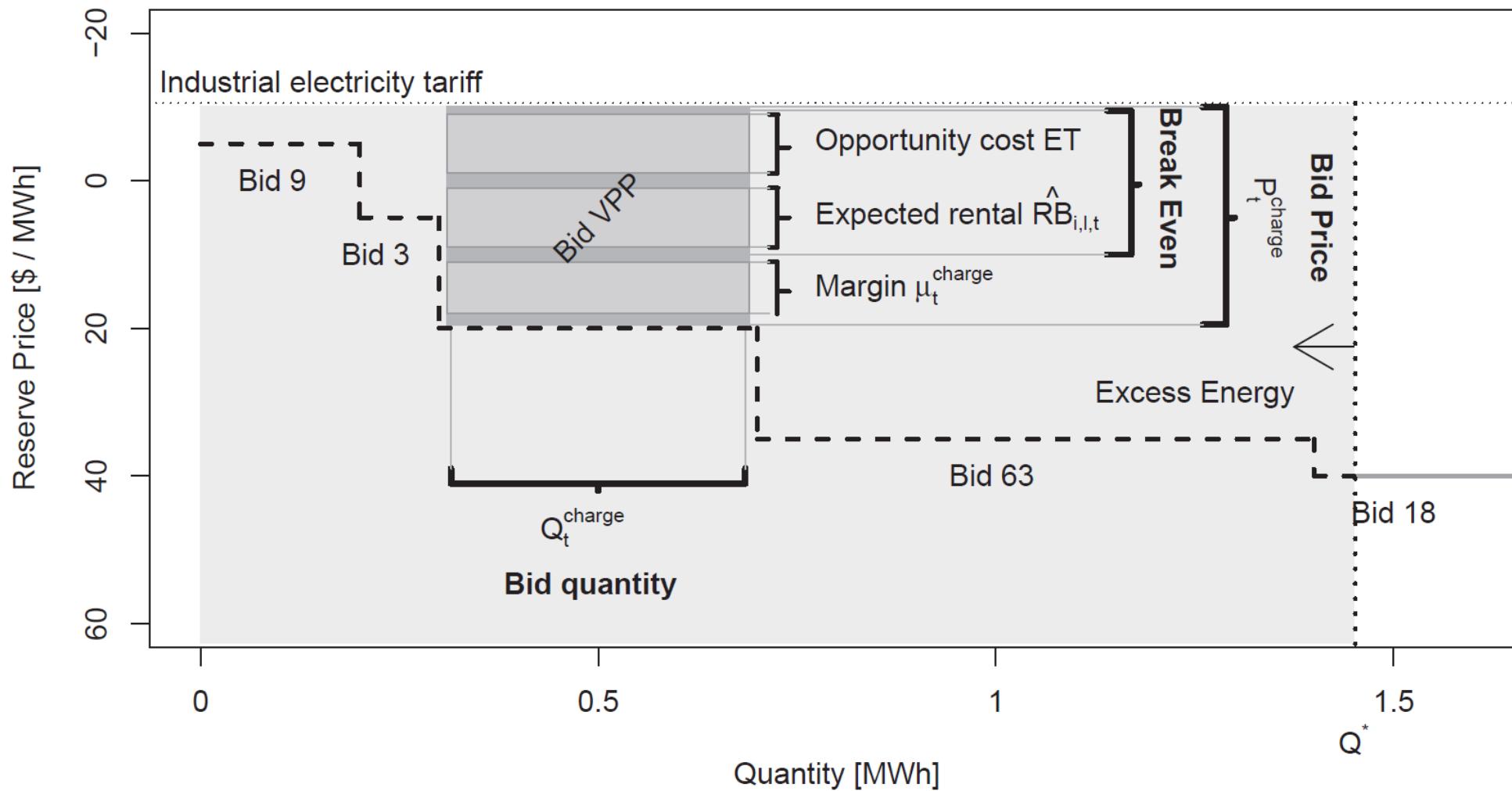
€ 14.90
per hour

€ 59
per day

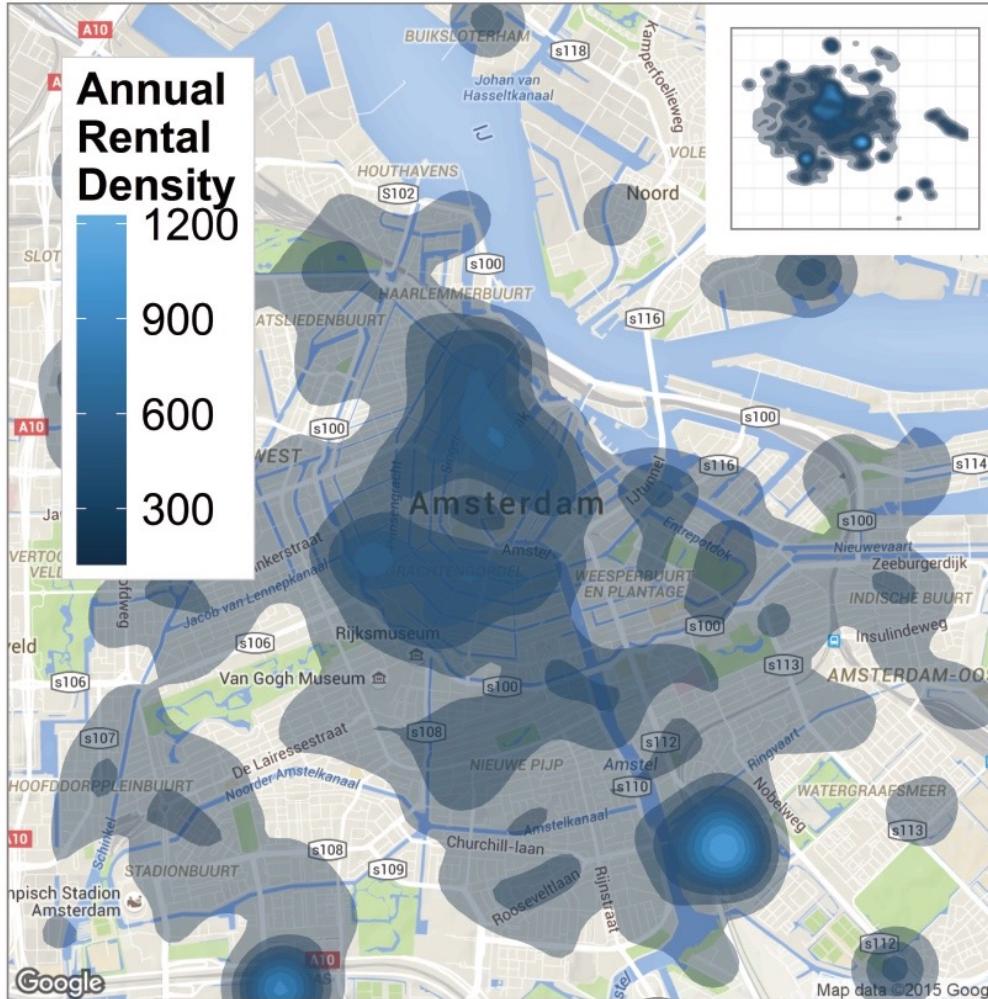
The flexibility that the car batteries offer can be marketed for



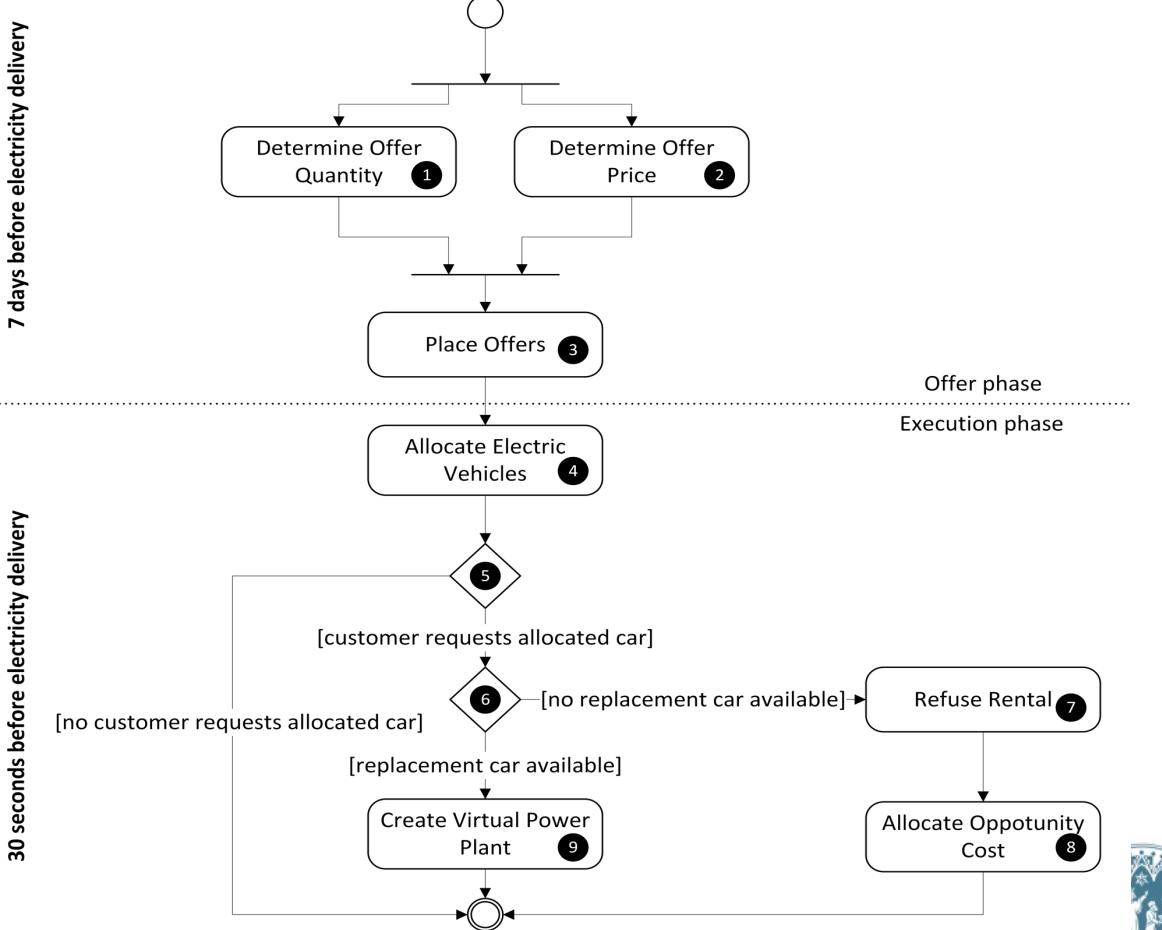
The VPP operator issues bids for upwards and downwards regulation – Example Charging



We predict rental demand and use this as input to allocate fleet vehicles to the virtual power plant



Activity Diagram Electric Vehicle Allocation Process



The VPP business model provides an additional revenue stream to the fleet operator – Overview of financial results

	Stuttgart	Amsterdam	San Diego
Battery technology 2015 (0.1 \$/kWh)	4.4% Gross profit increase	3.4% Gross profit increase	3.7% Gross profit increase
Battery technology 2020 (0.05 \$/kWh)	5.0% Gross profit increase	3.9% Gross profit increase	4.1% Gross profit increase
Battery technology 2025 (0.025 \$/kWh)	6.1% Gross profit increase	4.4% Gross profit increase	4.8% Gross profit increase