



Digital Technologies and Value Creation

Dr. Philippe Blaettchen
Bayes Business School (formerly Cass)

Why analytics now?

Major opportunities for businesses:

- Personalized services to customers/enhanced customer experience
- Automation of tasks that were previously costly (Forbes, 2018: AI will generate \$2.9 trillion in business value and recover 6.2 billion hours of worker productivity by 2021)
- Increased efficiency of processes and tasks
→ Has led to a lot of hype around the topic



"Artificial intelligence is the future, not only for Russia, but for all humankind. It comes with colossal opportunities, but also threats that are difficult to predict. Whoever becomes the leader in this sphere will become the ruler of the world," Russian President Vladimir Putin said.

Why analytics now?

Ongoing research into Algorithms

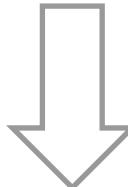
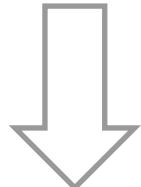
Page Rank for Google Search, Deep Learning Algorithms (ANNs, CNNs), Reinforcement Learning, ADMM

Increasing access to Data

1991: Internet
1997: Google
2000s: Home PC
2004: Facebook
2005: YouTube
2007: Iphone
Etc.

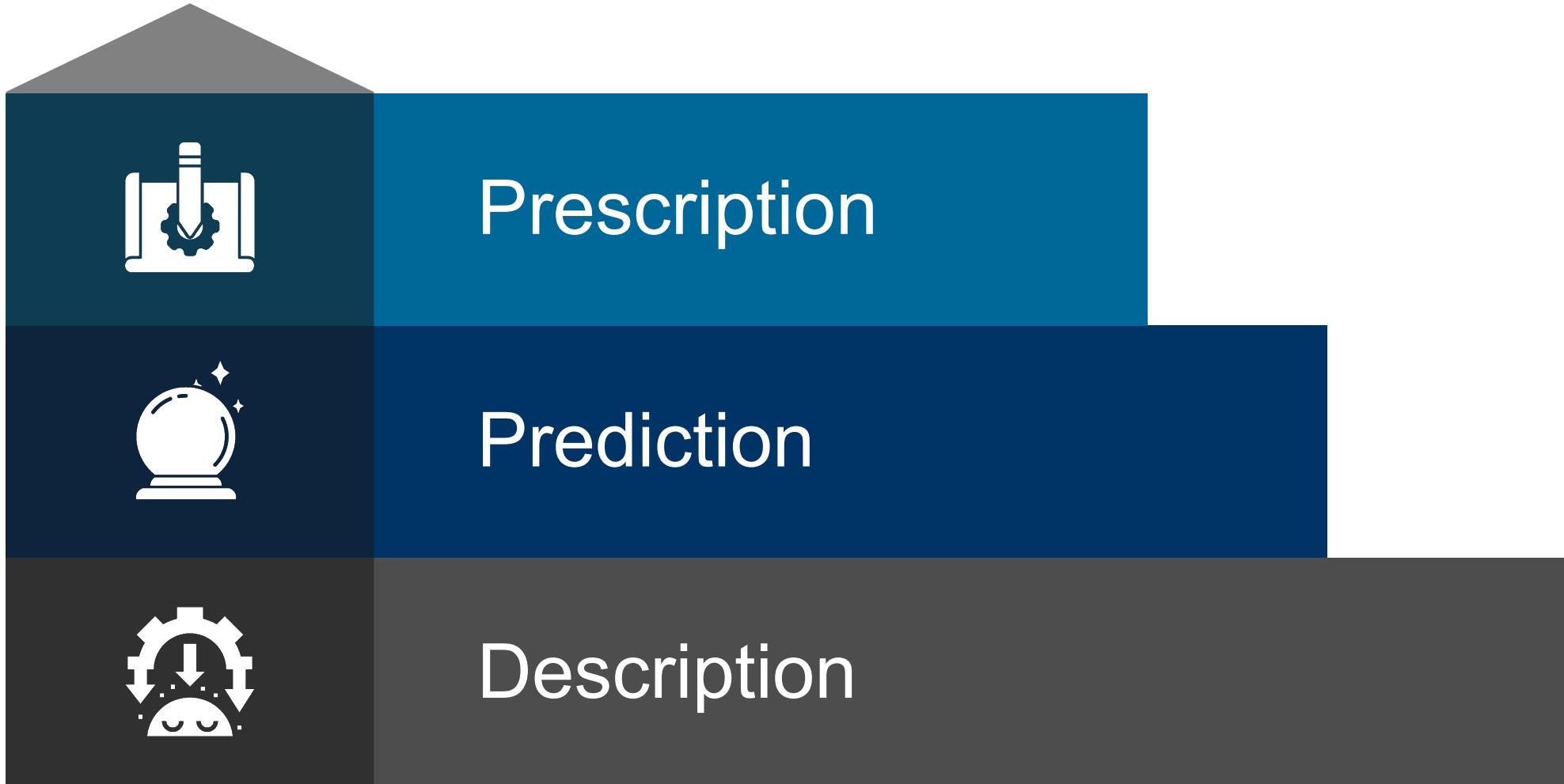
Better computing capabilities

More processors per sq. inch (Moore's law), cloud-based services (storage and computation)



“Perfect storm” for Analytics

Descriptive, predictive, or prescriptive – where is the value?



Overview – subject to change

Overarching theme	Week	
Introduction	1	Introduction to analytics applications and coding basics
Gathering data	2	Scraping web data
Gathering data / descriptive analytics	3	Data pre-processing and descriptive analytics
Gathering data / descriptive analytics	4	Descriptives in marketing analytics, and using social media APIs
Descriptive analytics	5	Descriptives in people analytics
NO LECTURE	6	NO LECTURE
Predictive analytics	7	Retaining employees and customers
Predictive analytics	8	Valuing a (social media) customer base
Predictive analytics	9	Segmenting customers and positioning products
Prescriptive analytics	10	Optimizing products and organizations
Prescriptive analytics	11	A/B-testing in practice

Learning objectives of the module

Goals: Provide you with the knowledge to

- feel comfortable in any environment where analytics and coding is discussed
- interact with (“speak the same language”) as specialists and set you on the path to become one
- be aware of the tools and technologies that allow to extract value from analytics
- identify business opportunities which could use analytics, especially in organizational design and marketing
- know when you’re being taken for a ride

How will we do this?

- Hands-on approach to analytics tools and technologies
- Experiencing the whole analytics pipeline
- Coding and practice on various datasets

Hybrid class norms

Please arrive on time – or even a couple of mins ahead of schedule

Please make sure to wear a mask at all times to protect yourselves and others

Please only use your computers for the task at hand: no social media (except the one we analyze), no browsing

Camera on please on Zoom. Otherwise I don't get to know you (and I'll randomly invite people to contribute to the discussion)! Use chat for raising clarification questions. I'll also pause every 5-10 minutes to answer questions that I missed

Come prepared to class: lecture learnt, homework done

You will need multiple browser windows:

- Jupyter notebook (as needed to run code)
- PollEverywhere for in class polls (will give instructions)

Office hours

- For everyone's safety (and the sanity of my co-workers), only through Zoom
- Every Tuesday, 11 am – 1 pm, Link on Moodle
- Strictly FCFS (if there is a lot going on, write your name in the chat when you arrive so we can keep track)
- Feel free to stay around and listen to other students' questions. Also, helping each other out is strongly encouraged
- I expect you to have carefully revised the materials beforehand

Learning objectives of today

Goals: Familiarize yourselves with the course and Python (and let me familiarize myself with the class and students' background)

- Resolve any outstanding questions about what this class is about
- Be able to run programs in Jupyter Notebooks (or at least Google Colab)

How will we do this?

- Quick walk-through of the main application areas we will talk about
- A deeper discussion on the assigned reading, as a teaser to the People Analytics part
- A chance for everyone to practice their programming and for me to see where possible problems are

Let's start by me familiarizing myself with the class



Go to pollev.com/dtvc2021

People Analytics and Org 2.0

The universal problem of organizing



Source: Puranam & Clement (2020)

Organization Design & Development Issues



Individual Level

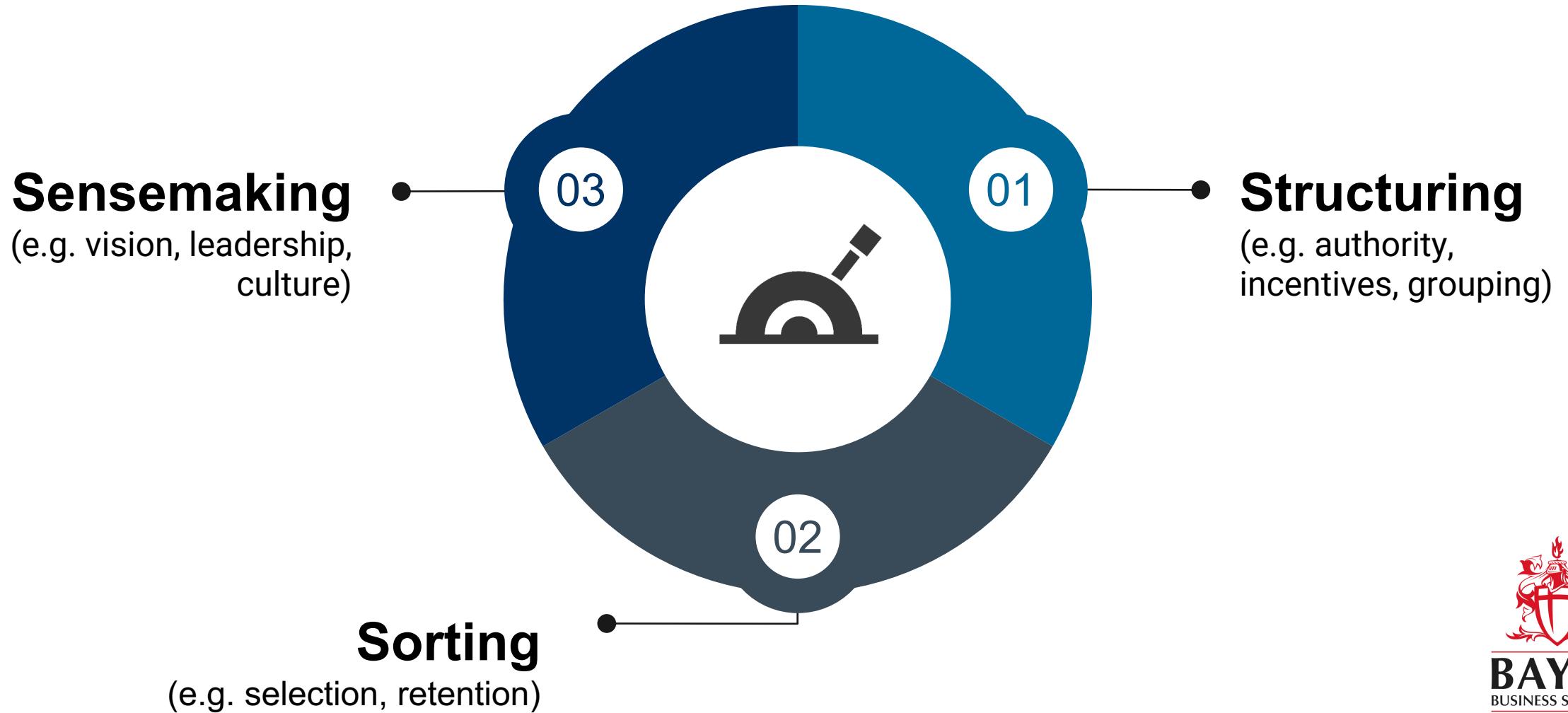
- ✓ Executive assessment
- ✓ Executive Search
- ✓ Executive Pay
- ✓ Employer Branding
- ✓ Onboarding
- ✓ Talent Acquisition
- ✓ Employee engagement
- ✓ Talent Management
- ✓ Leadership Development
- ✓ Compensation Design
- ✓ Communication Management



Group Level

- ✓ Organization Design
- ✓ Measures and Incentive systems
- ✓ Design Effectiveness
- ✓ Organization Diagnostics
- ✓ Performance Improvement
- ✓ Change management
- ✓ Strategy Execution
- ✓ Culture Management
- ✓ Post-Merger and Acquisition

Three levers of organization design



Example: Project Oxygen at Google

Managers
at Google:
a tough job

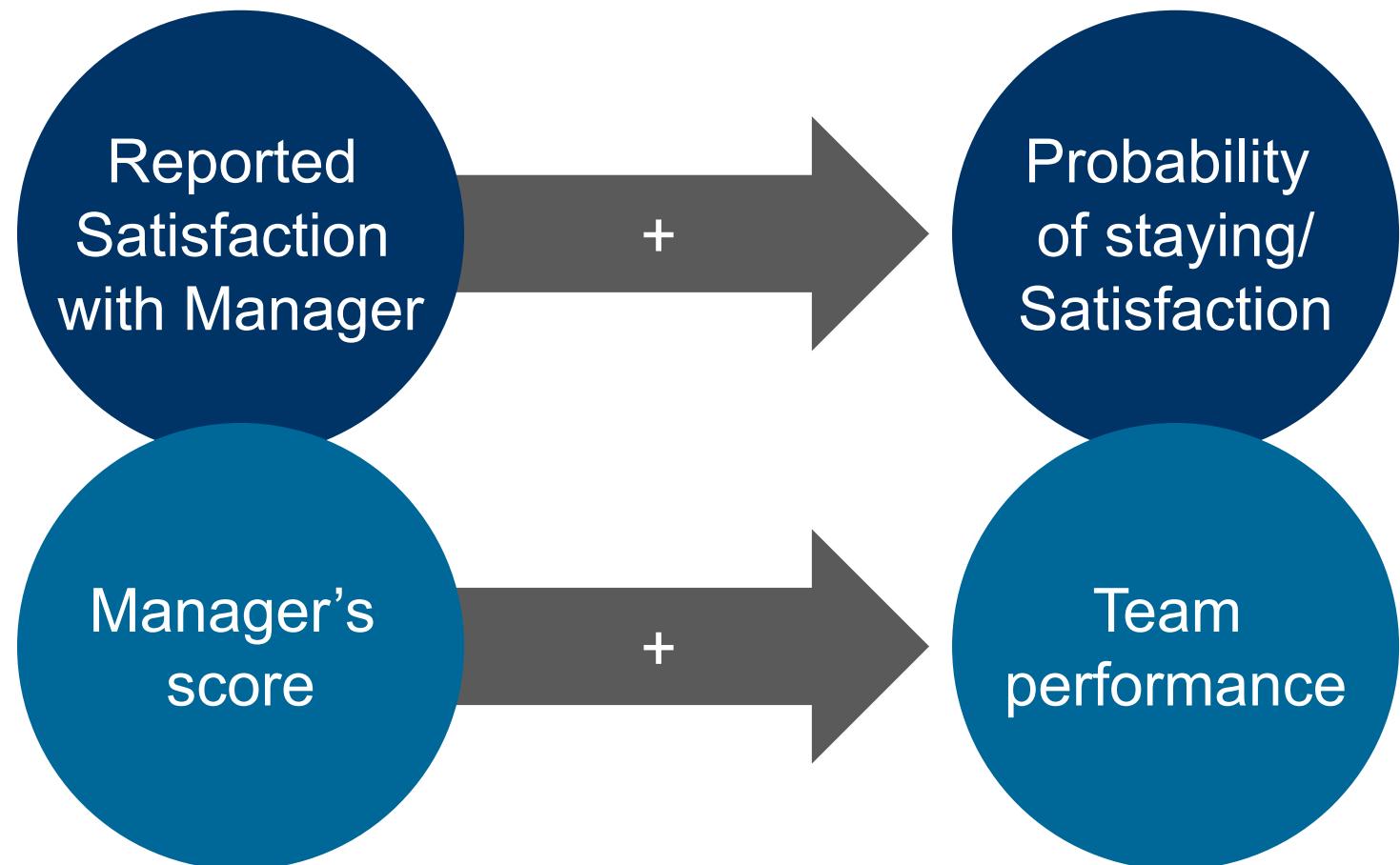


Need for autonomy of employees

Must assist career development without promotions

Managers need to “earn their living”

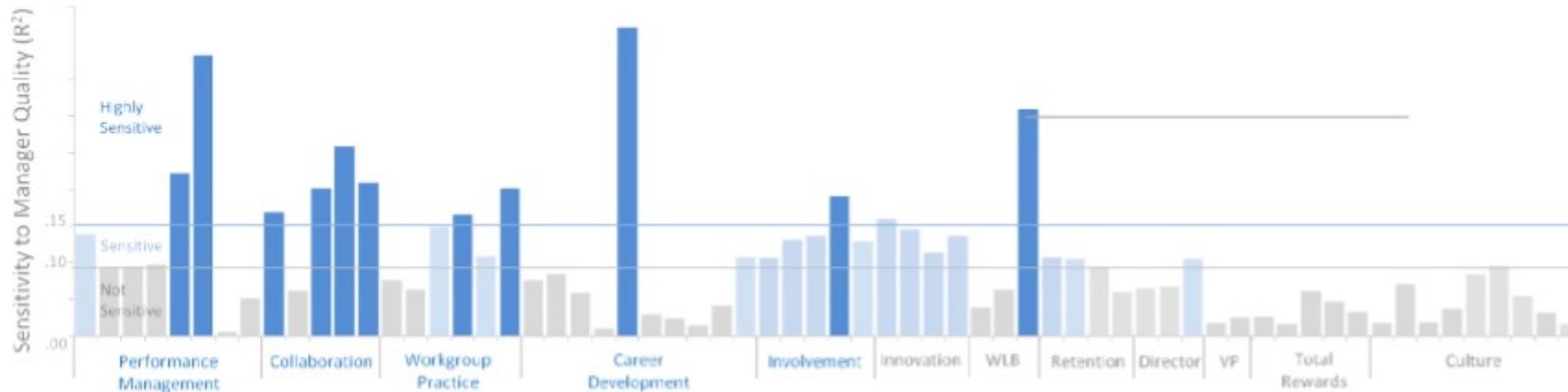
Patterns in the data



Managers have a significant impact on how Googlers perceive life at Google.



Each bar represents a Googlegeist question, grouped by theme. The height of each bar corresponds to how much managers influence how a Googler will answer that question.



Highly Influenced by Managers

Performance management & career development: recognition and transparency around performance evaluation & promotions, regular feedback & support

Collaboration: value diverse perspectives, quick resolution, resource allocation

Work-Life Balance: manager supports balance

Influenced by Managers

Career Development: career goals can be met

Innovation: encouraged to innovate, take risks, improve processes, prioritize innovation

Not Influenced by Managers

Total Rewards: paid fairly, understand compensation

Culture: Google is fun, there is a climate of trust

Why wasn't this enough?

Follow up – 2009

Double blind qualitative interviews
With a script (vs. Open ended)

“Great manager” award nominations

Inductively derived 8 attributes

Follow up – 2009: Oxygen attributes of a good manager

1



Coaches

2



Empowers

3



Is concerned with
team's wellbeing

4



Is results
oriented

5



Is a good
communicator

6



Helps with career
development

7



Has a
vision/strategy

8



Has technical
skills



Follow up – 2010

Roll out list and include in “Upward Feedback Survey” (UFS)

Linked to great managers nominations

Remedial (targeted) training

Do not link to salary; collected off cycle from performance review

Follow up – 2010: Impact



83 to 88%
average
increase in
manager score



Some
Managers
surprised by
disconnect
between UFS
scores and own
evaluation from
their boss

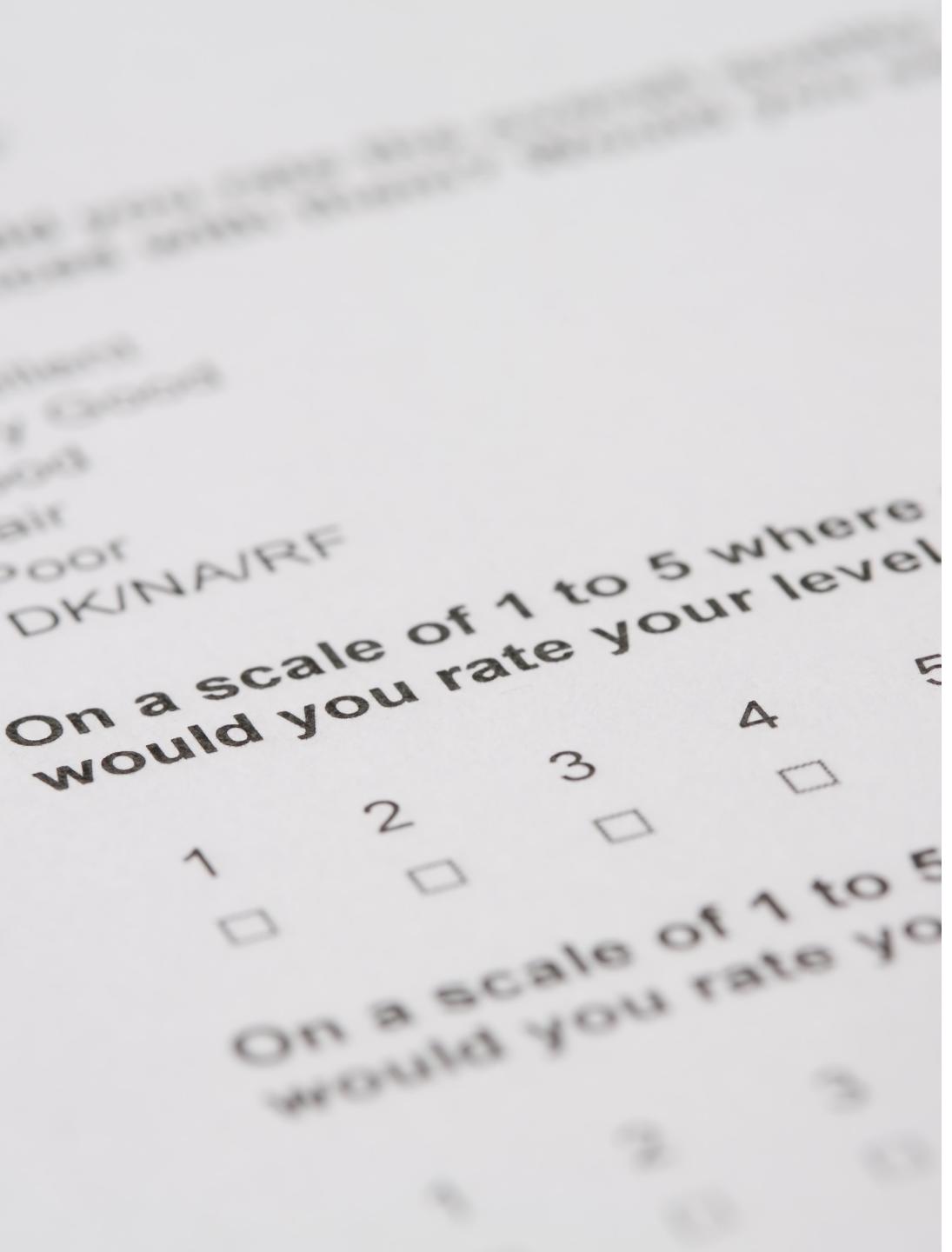


No evidence of
retributive UFS



ARE YOU CONVINCED THAT GOOGLE HAS
IDENTIFIED THE CHARACTERISTICS OF
EFFECTIVE MANAGERS?

Go to pollev.com/dtvc2021



Survey data: The challenge of subjectivity

How to improve the use of subjective data

Pre-test

Measure same construct in multiple ways

CRONBACH ALPHA (correlation between measures)

Use multiple evaluators for same construct if possible

INTER-RATER RELIABILITY (correlation between raters)

How to improve the use of subjective data

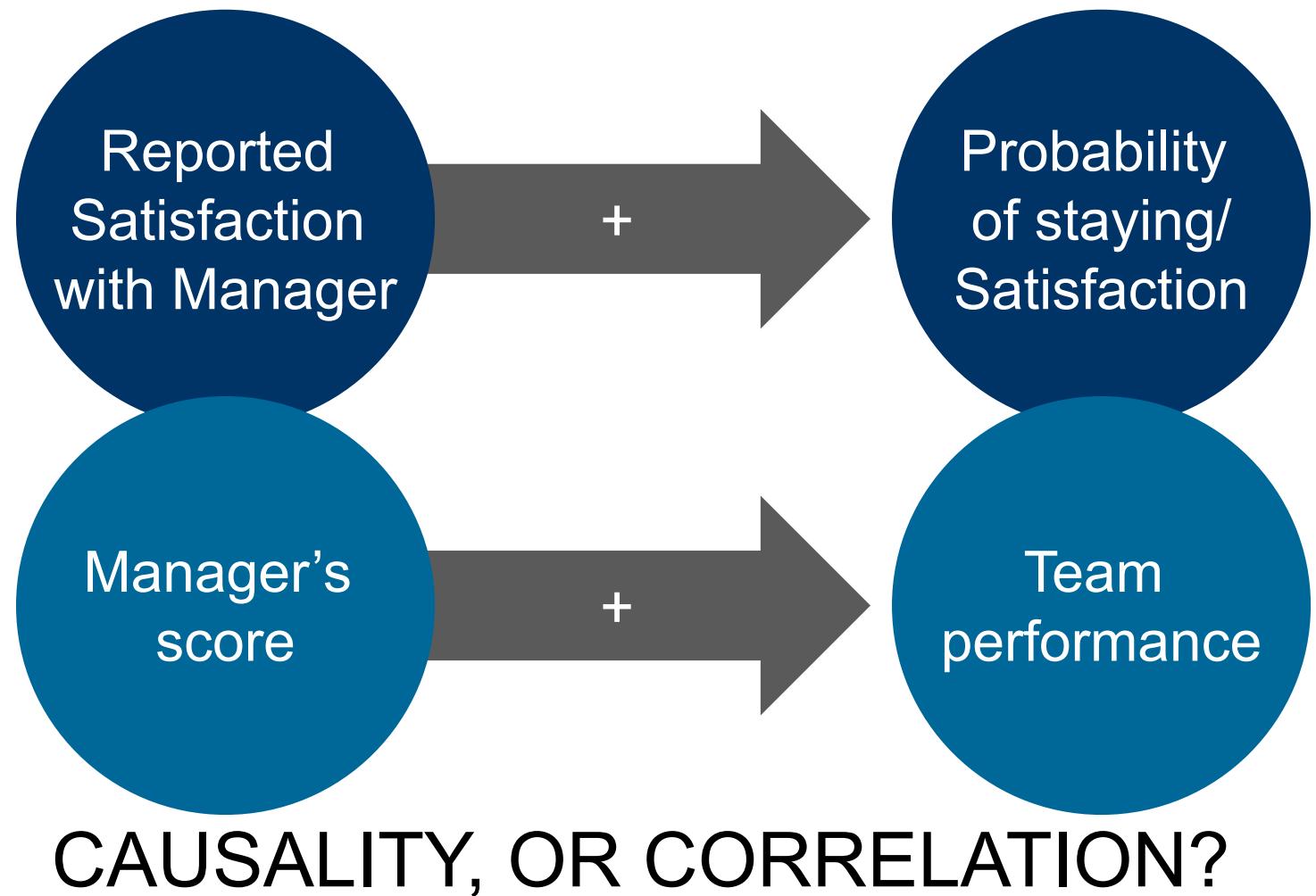
Triangulate with other (objective) measures

Treat associations between two subjective measures reported by the same individual with caution

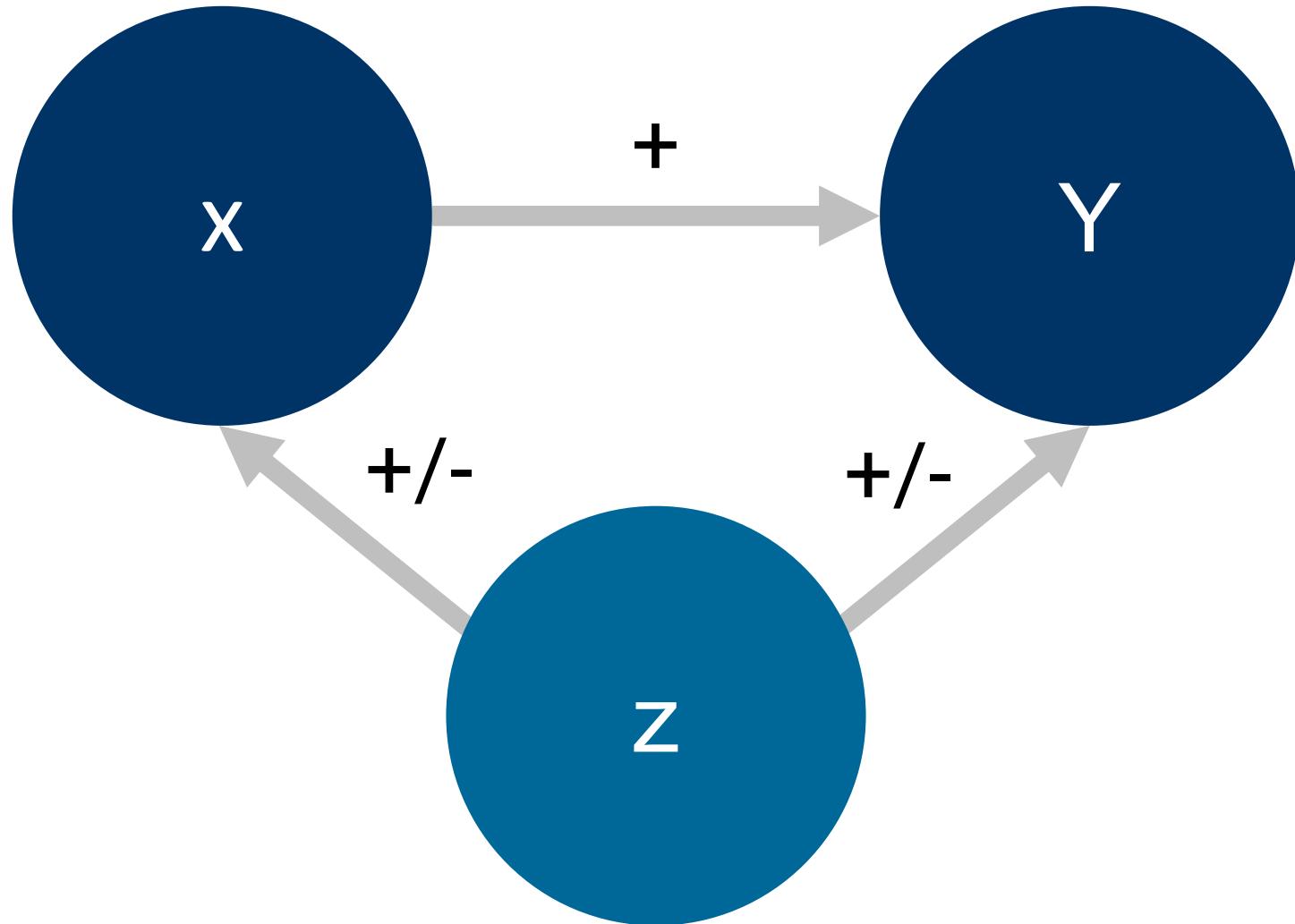
Beware of “**COMMON METHOD BIAS**”

You may be capturing an attribute of respondent, not reality

Patterns in the data



Patterns in the data: 2008



What's the hidden variable?



Ice cream sales
and drowning



Stork nesting and human
baby deliveries



Number of fire engines at
fire site and \$ of damage



Graduate degree
and salary

What could cause spurious relationships in Project Oxygen (between Oxygen 8 and Satisfaction/Retention)?

HIDDEN VARIABLE	EXAMPLE?
A feature of team	
A feature of manager	
A fit between the two	

What happens then?

WE IMPLEMENT PROJECT
OXYGEN ANYWAY, AND TEAM
PERFORMANCE

INTERPRETATION?

Worsens

Stays the same

Improves



What data from within an organization can tell us, that external consultants cannot

- » Unique to organization
- » Surprise also lies in what we do not find, not what we find
- » Debunking current “wisdom”
- » Legitimizes and creates buy-in



Convergence around data (instead of unproductive diversity of opinion)

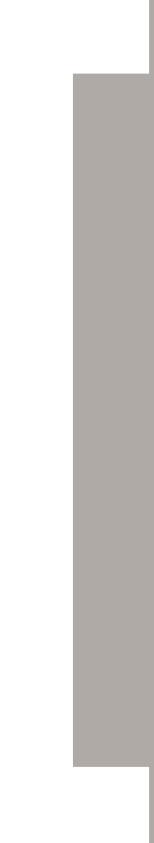


Using data to understand what is currently going on in the organization-perception

- » How to use descriptive statistics to gain insights
- » How to formulate and test simple hypotheses to diagnose what might be going on



Teams as the basic organizational building block



Analytics in Marketing / MarTech

Marketing analytics?

“Building models of seller and buyer preferences to make predictions about future marketplace behaviors, and to use these predictions to optimize a firm’s offering”

A fundamental measurement issue

“Building models of seller and **buyer preferences** to make predictions about future marketplace behaviors, and to use these predictions to optimize a firm’s offering”

→ How do we measure preferences?

The old world

- Customer surveys (live, phone, mail, online, etc.)
- Choice experiments



The new world

- Revealed preferences observed in (big) data
- Low-cost field experiments on websites
- Social media information

Technology (esp. software) solutions that focus on

- **Marketing analytics**
- Customer experience
- Marketing attribution
- Content and content management
- Email marketing
- Marketing management

Social Media Analytics

What is Social Media Analytics?

“The art and science of extracting valuable hidden insights from vast amounts of **semistructured and unstructured** social media data to enable informed and insightful decision making”

How would you predict the success of a new movie with data?

Features	Adjusted R^2	p-Value
Average weekly tweet-rate	0.80	3.65e-09
Average daily tweet-rate	0.93	5.27e-09
Average daily tweet-rate + theaters count	0.97	9.14e-12

A supporting function

Org 2.0: emails, work chat, online employee reviews, ...

Marketing: influencers, networks, reactions, ...

Coding basics

Before getting to coding, one more question about you.

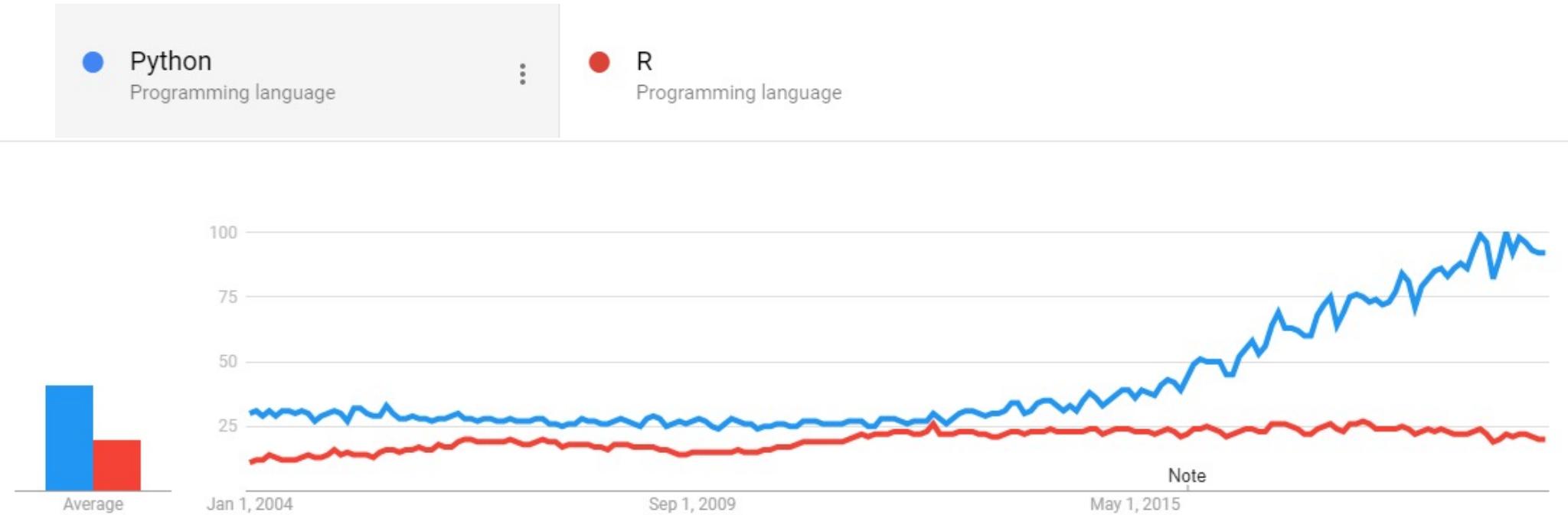


Go to pollev.com/dtvc2021

The choice of a language

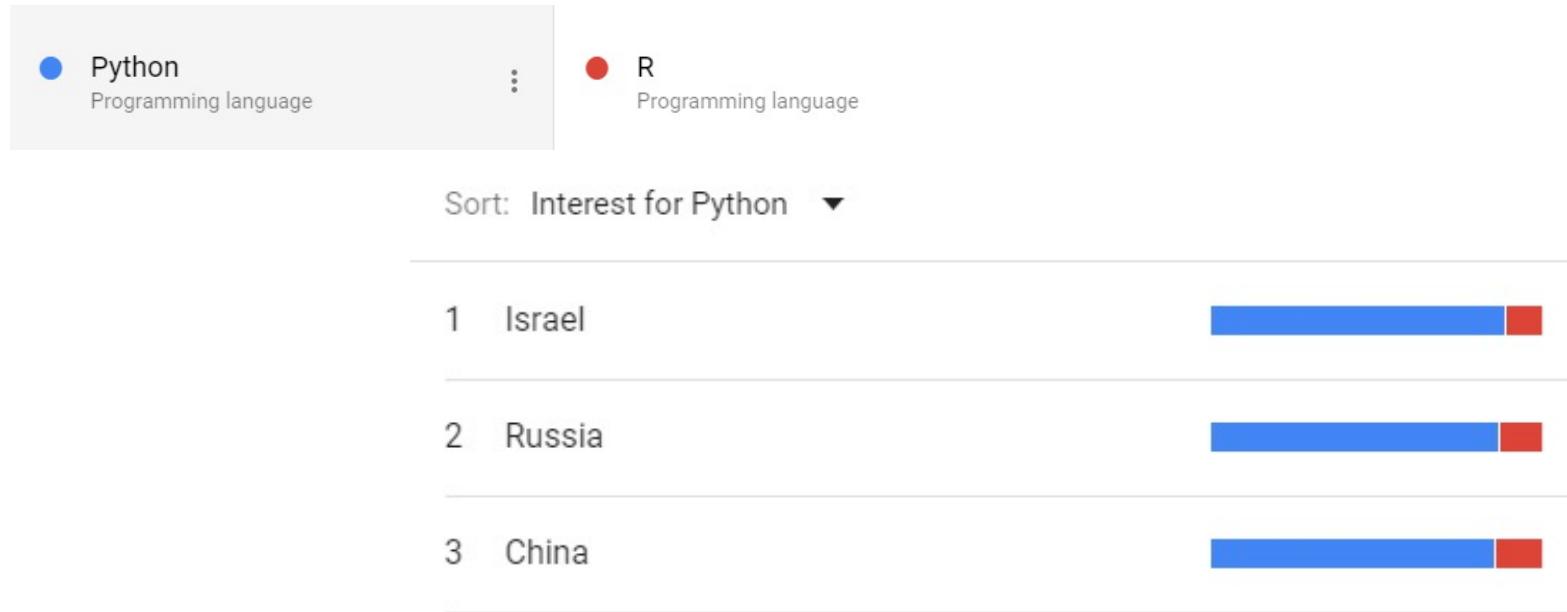
You have worked with two languages already: Python and R.

In this course, I have decided to go with Python.



Source: Google trends

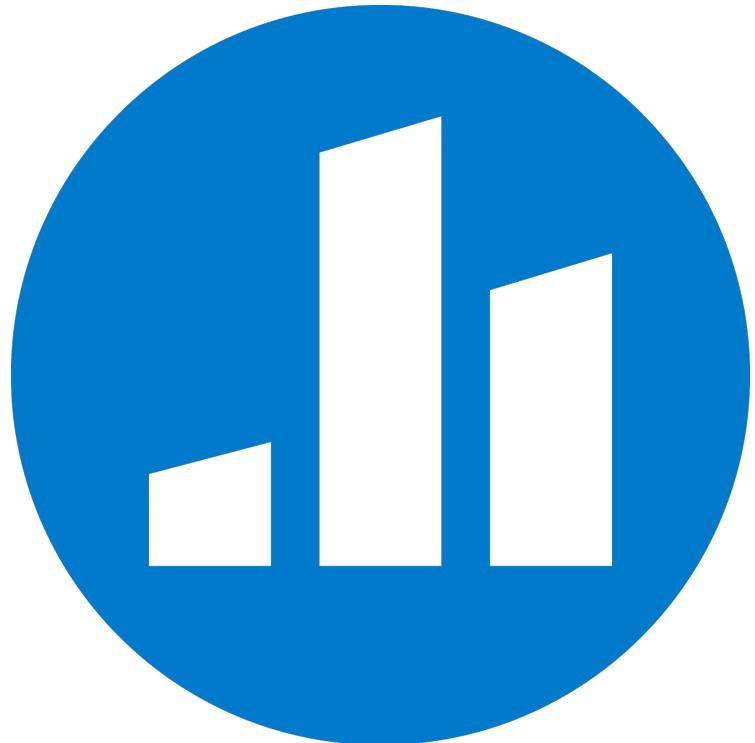
The choice of a language



- Not as important as people think
- Learning one coding language is hard
- Switching from one language to another is much easier
- **Key thing: learn how to think as a coder!**

Source: Google trends

What are your experiences coding in Python?



Go to pollev.com/dtvc2021

Logistics of coding

- **Make sure to download the software on time**
 - For today, you should have Jupyter Notebooks working, optimally with Python 3.7 distributed through Conda. Pip should also be up and running to manage the packages we need
 - When we get to crawling data with Scrapy and Splash, we will need to exceptionally use the command line tool
- **For ease of access:** download the data and notebooks of the current lecture or tutorial from Moodle before class. In general, for each class, I will provide two IPython notebooks:
 - the “**Exercise**” notebook:
 - contains the code you need to run/complete
 - typically posted before class
 - the “**Lecture**” notebook:
 - contains solutions to exercises that I ask of you
 - typically posted a couple of days after lecture (to give time for homeworks)
- Some videos will come with only a “**video**” notebook, but some will also have a “**video-exercise**” notebook

Dataframes

A **dataframe** in Python looks like:

	Feature 1	Feature 2	Feature 3
Observation 1	221	Small	Blue
Observation 2	157	Large	Green
Observation 3	236	Medium	Red
Observation 4	50	Extra-Small	Green
Observation 5	122	Large	Red

Numerical values Categorical values

The tools we will mostly be using for programming



colab

Let's get started on some programming!



Analytics pipeline

The very first step: A problem to solve

“[Analytics] also entails applying data patterns towards effective decision-making” - Wikipedia

Now we need some data

From within the organization

- Sensors in factories and offices
- Bookkeeping and accounting
- Slack, Emails
- Website traffic
- Customer data
- ...

From outside the organization

- Competitive intelligence: reports, statements
- Social media: Google, Twitter, Instagram, Facebook, Glassdoor, LinkedIn, ...
- Consumer studies
- ...

We don't just need to have the data, we also need to understand it

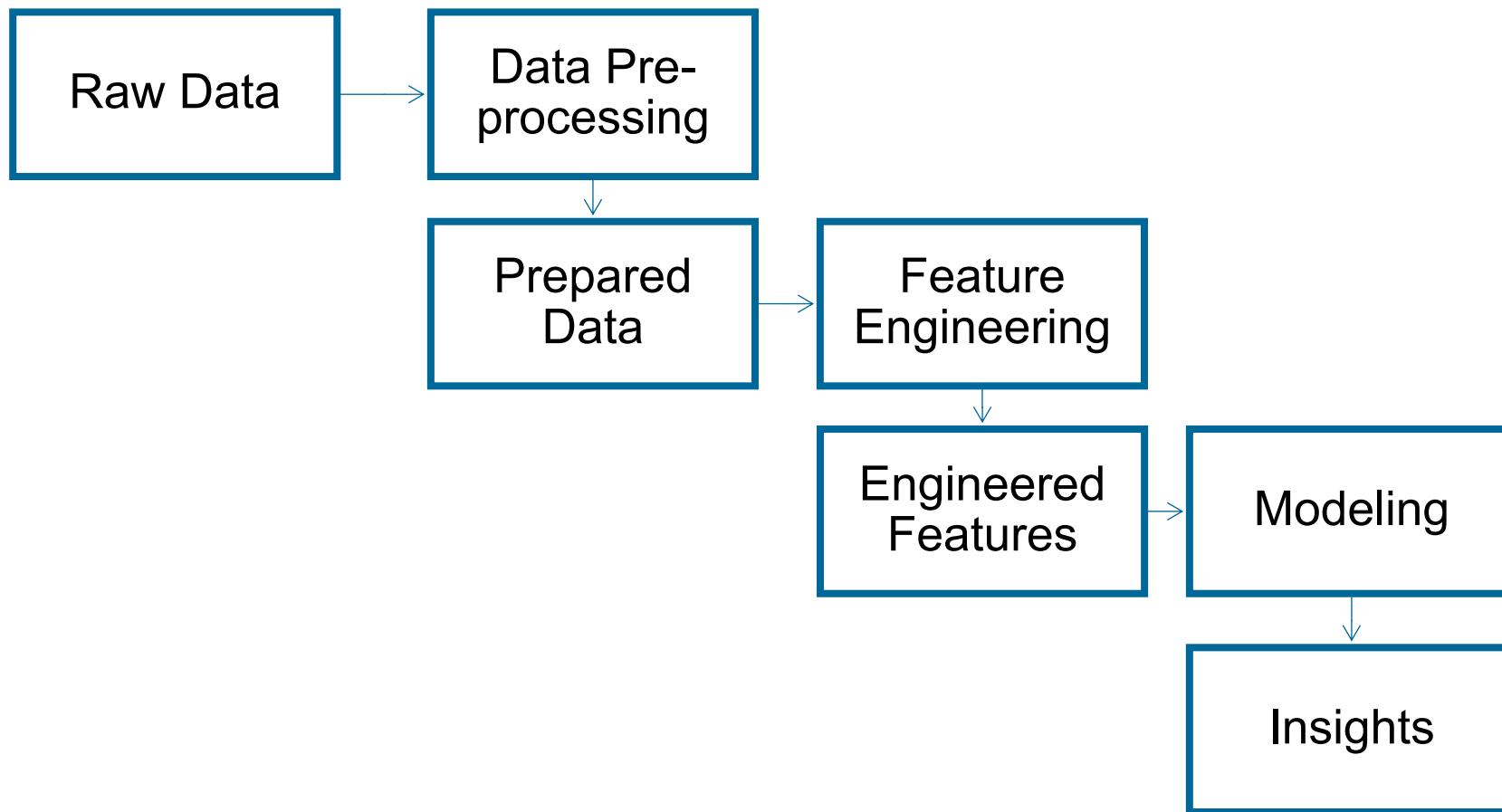
**Understand how the dataset was put together.
Otherwise: garbage in = garbage out**

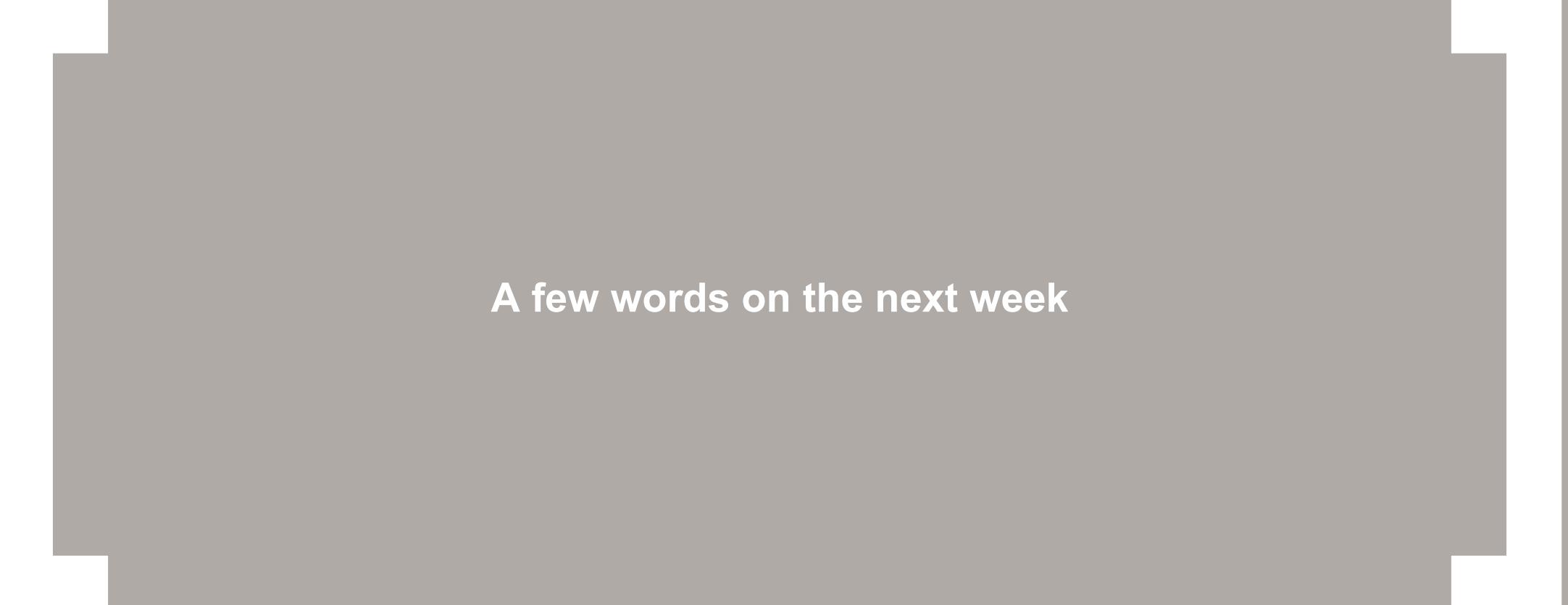
What could you want to know about how the dataset was put together?

- How are features measured?
- How were features mapped to observations?
- Are all observations included?

These are **crucial** questions to give context to your analysis

The analytics pipeline





A few words on the next week

A few words on the next week

- Send me your group selections
- Note the homework on Moodle – the quiz is graded!
- Make sure to go through the Exercise book, including the questions at the end
- It is really important to train your coding abilities early – the next lecture is probably going to be the biggest step
- Come prepared to the next class – videos watched, notebooks downloaded
- Also, I will give instructions on downloading Splash, which is one of the tools we will use for web scraping!

Until next time – have a great week!



The homework for the first week will be a brief
quizz

