Smart Mobility Transition?

Prozesse und Konzepte eines digitalen Wandels der Mobilität

Dr. Robert Schönduwe

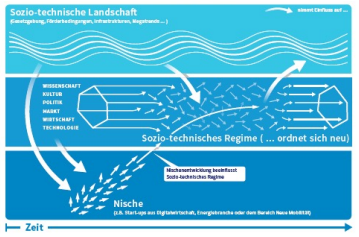
18.06.2021, Berlin

Previously on Smart Mobility Transition? And the plan for today

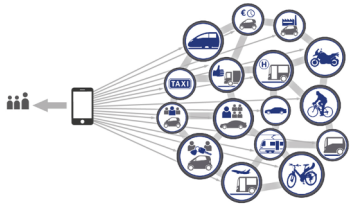
Definition Nachhaltige Mobilität

Komplexität von Ursache-Wirkungs-Beziehungen

Theorie des Wandels: Wenn die Aufgabe so komplex ist, wie kommt dann überhaupt das Neue in die Welt?



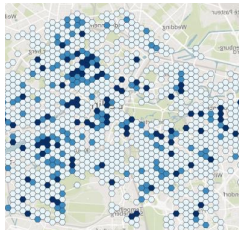
Praxis des Wandels anhand eines Beispiels: Wie kann der Wandel praktisch gelingen? Eine konzeptionelle Annäherung über das Bsp. MaaS.



Praxis des Wandels: Wie wird ein Produkt entwickelt?

Praxis des Wandels: Welche Rolle spielen Daten bei der Produktentwicklung?

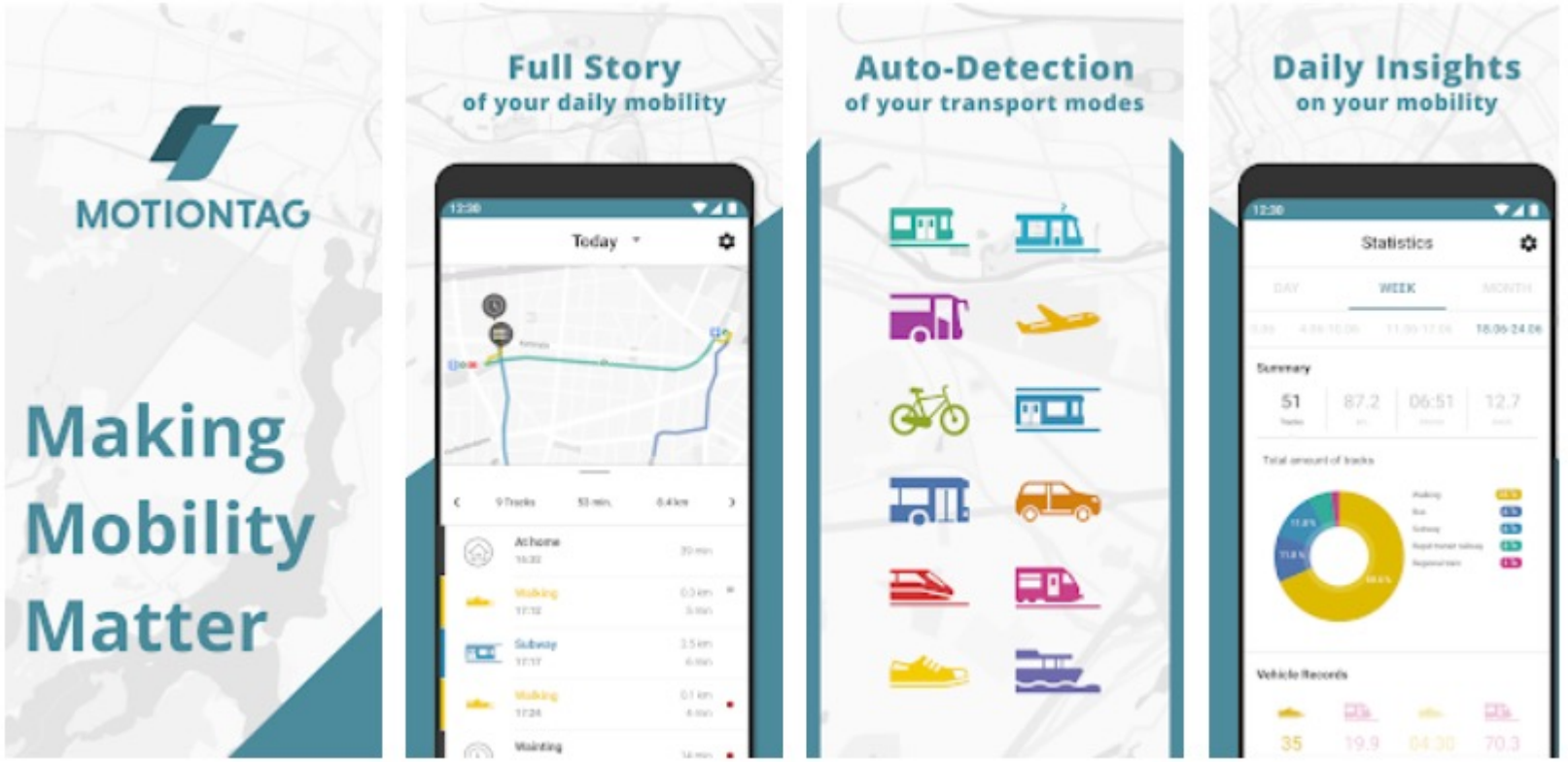
Praxis des Wandels: Woher kommen Daten für die Entwicklung von Mobilitätsprodukten?



Praxis des Wandels: Wie arbeitet man mit Mobilitätsdaten?

Homework 1: Data collection for next lecture

Smartphone-based GPS tracking data



Download the MOTIONTAG-App from the store.
Use it for some days.

We will use the data in the next lecture on 18.06.2021

Download on the App Store

GET IT ON Google Play

Homework 2
Install R and RStudio

Install the language and the interpreter R.

https://cran.rstudio.com/

The second step is to install the IDE

https://rstudio.com/products/rstudio/download/#download

[9]

Nutzerzentrierte und datenbasierte Produktentwicklung

If I’d asked my customers what they wanted, they’d have said a faster horse.

Henry T. Ford

Customers don’t know what they want. It’s very hard to envision the solution you want without actually seeing it.

Marty Cagan

- Silicon Valley Product Group -

Customers don’t know what’s possible. Most have no idea about the enabling technologies involved

The critical failing of user interviews is that you're asking people to either remember past use or speculate on future use of a system

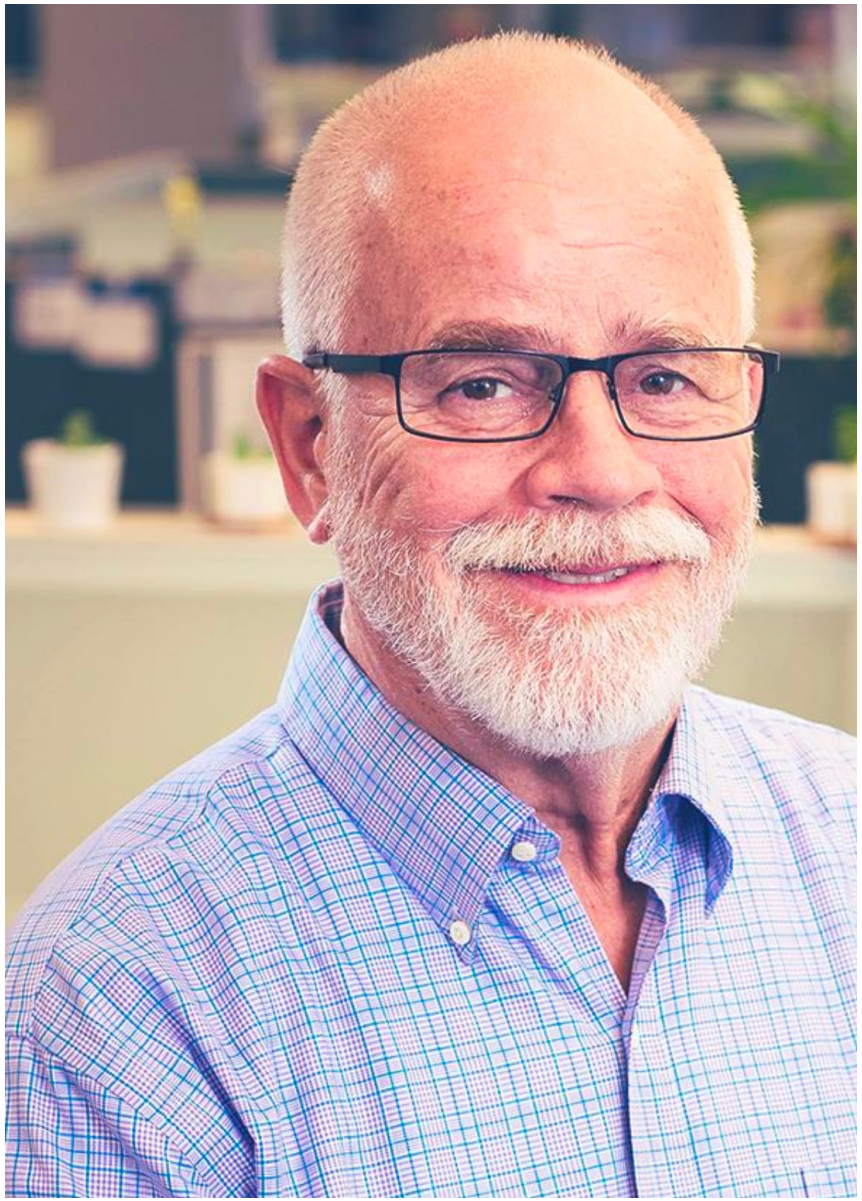
Jakob Nielsen

- UI Expert-

[The assumption that a] reasonably well-defined set of requirements exists, if only we take the time to understand them, is wrong

Dean Leffingwell

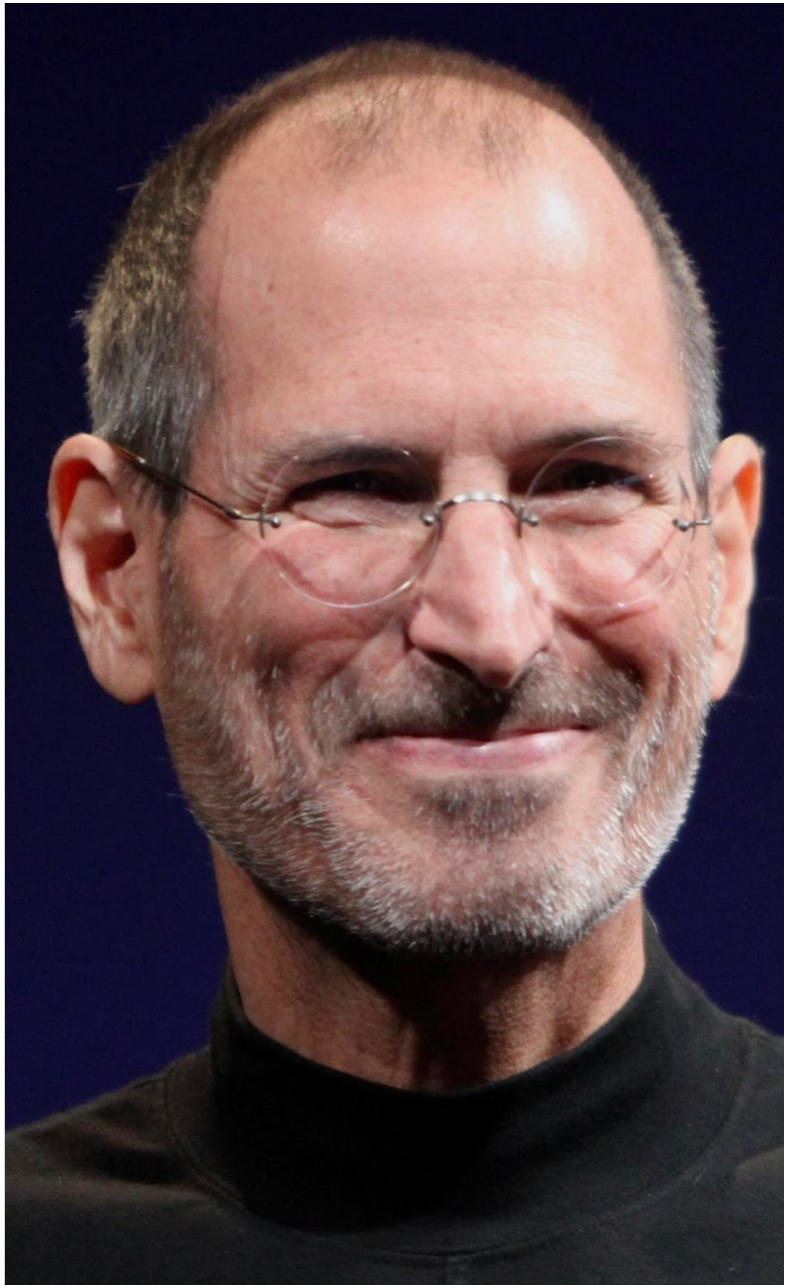
- Scaled Agile Expert-



You can't just ask customers what they want and then try to give that to them. By the time you get it built, they'll want something new.

Steve Jobs

- Founder of Apple-



Underlying insights

- Customers don’t know what they want
- You need to show “it” to them
- And then measure their behavior
- Or, if you must, talk to them

The Requirements Engineering Crisis

Main reasons why RE is „under attack“

- RE is expensive
  - Mostly manual work
  - Strongly based on experience
- RE needs time (vs. the need for speed)
- RE is detached from the product
  - Specifications get outdated
  - No feedback from field operations
  - RE neglected after product is shipped
- Plus: An entirely complete and unambiguous requirements specification will never exist in practice

The Software Crisis
In the 1960s

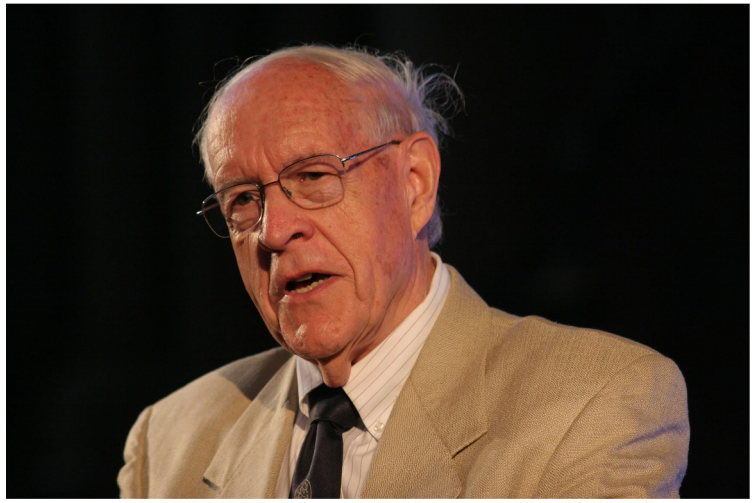
- Projects running over-budget
- Projects running over-time
- Software was very inefficient
- Software was of low quality
- Software often did not meet requirements
- Projects were unmanageable and code difficult to maintain
- Software was never delivered

The Software Crisis
In the 1960s

The major cause of the software crisis is that the **machines have become** several orders of magnitude **more powerful**! To put it quite bluntly: as long as there were no machines, programming was no problem at all; when we had a few weak computers, programming became a mild problem, and now we have gigantic computers, programming has become an equally gigantic problem.

— Edsger Dijkstra

The Software Crisis

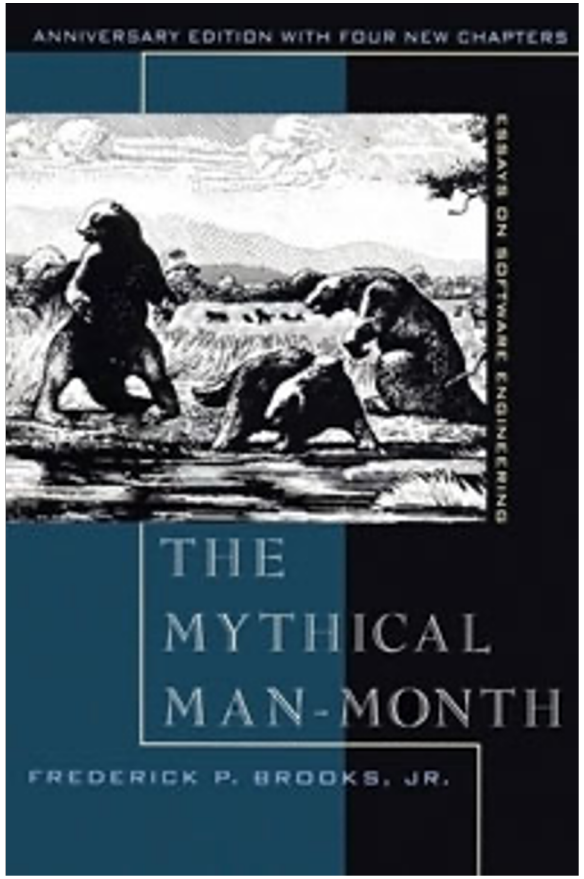


Fred Brooks

Managed the development of IBM’s System/360 and OS/360 (1966).

Brooks Law

Adding manpower to a late software project makes it later.



The vision of data-driven requirement engineering

Traditional RE

Stakeholder-centric

Decisions made by experts

Decisions based on intuition or rationale schemes

Yes/No Decisions

Upfront RE

Data-Driven RE

User-centric

Group/Crowd-based decisions

Decisions based on observations

Incremental control of features

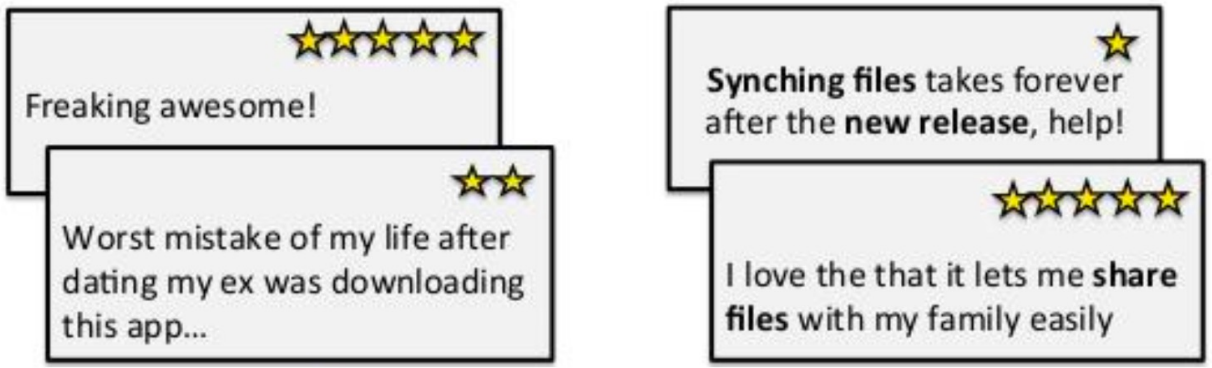
Continuous RE (real-time)

The core is data:

Collection, processing, analysis, presentation

Data sources for RE
Explicit feedback

- Example: app stores

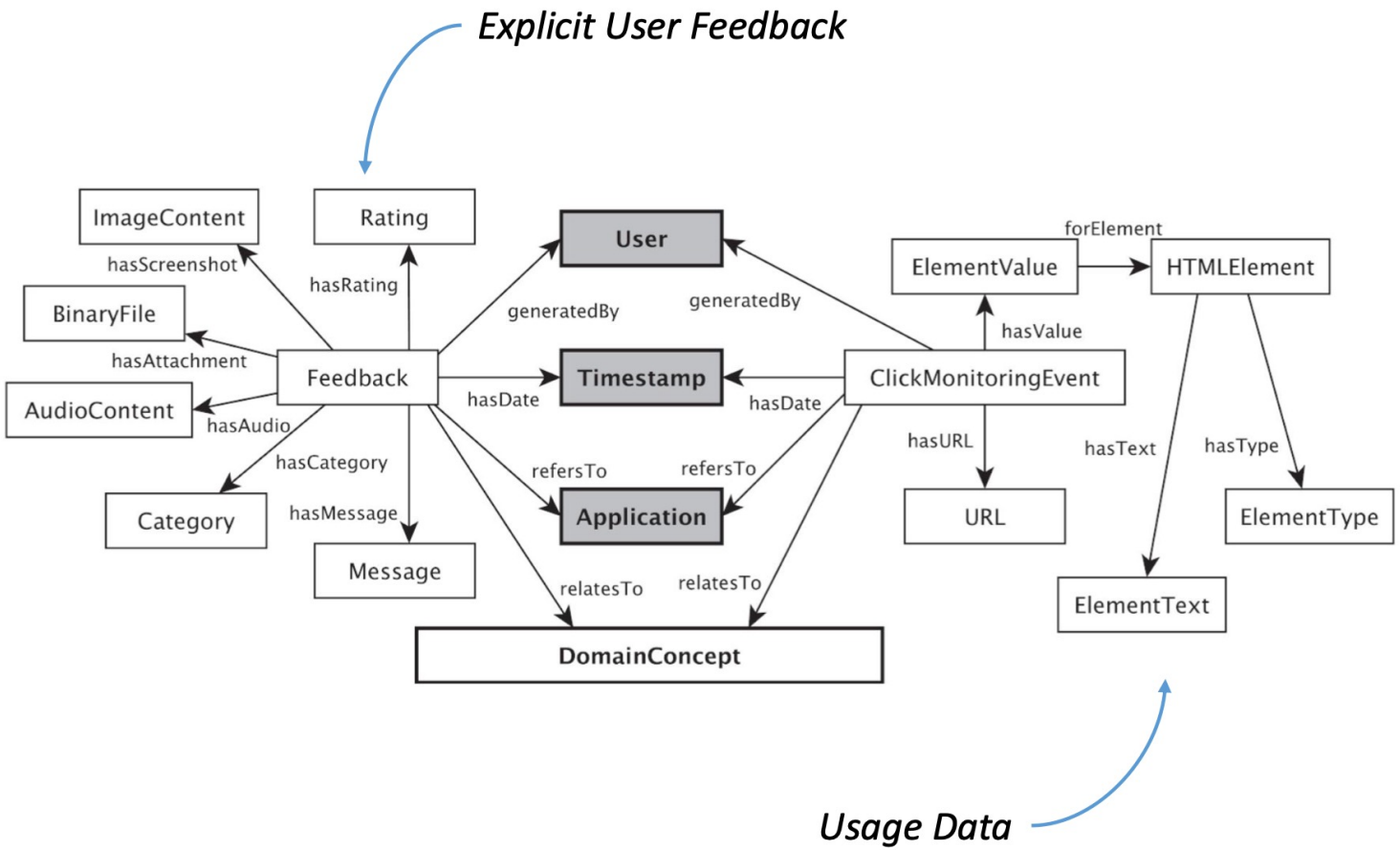


- About 1/3 of all reviews in app stores contain information related to requirements [2]
- There are tools for analyzing large sets of user feedback automatically [1,2]
  - Identifying sentiment (positive, negative)
  - Identifying types (bugs, feature requests, ideas)

[1] E. Guzman, W. Maalej: "How do Users Like my Feature?", *RE'14*
[2] D. Pagano, W. Maalej: "User feedback in the appstore: An empirical study", *RE'13*

Data sources for RE

Implicit feedback from usage data



More sources: Log data, test execution data, sensor input, capturing emotions, …

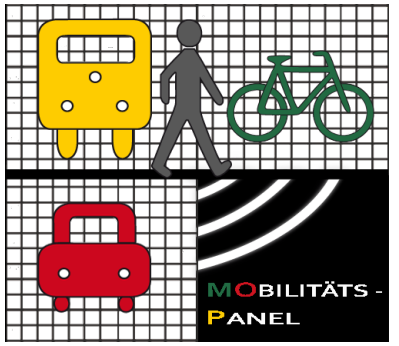
13

M. Oriol et al.: "FAME: Supporting Continuous Requirements Elicitation by Combining User Feedback and Monitoring", *RE'18*

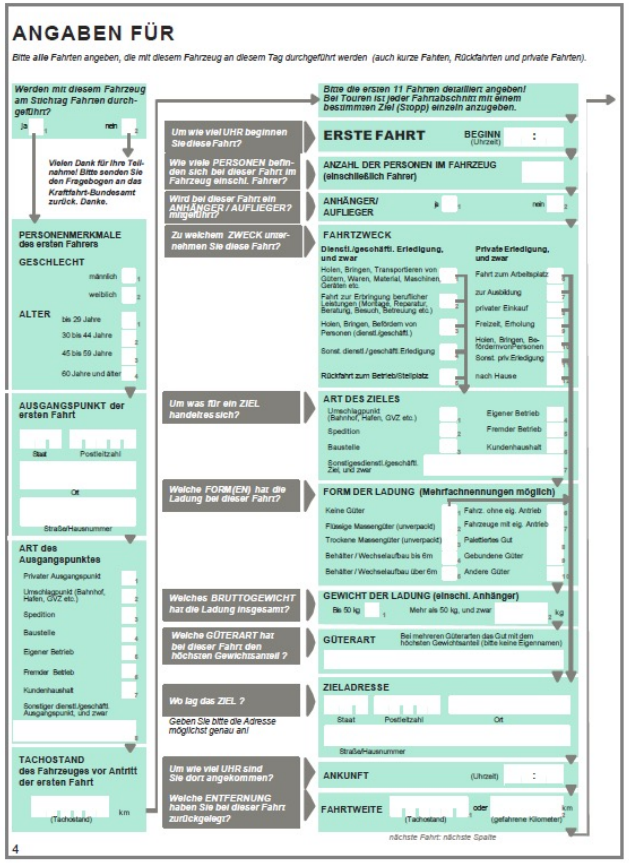
[10]

Mobilitätsdaten: von standardisierten Befragungen zu passiver und automatisierter Mobilitätsdatenerhebung

Data sources:
Traditional transport and market surveys

	MiD **M**obilität **i**n **D**eutschland	SrV **Mobilität in Städten**	MOP **Deutsches Mobilitätspanel**
			
Interval	Every six to ten years	Every five years	Every year
Participants	ca. 300.000 people	ca. 120.000 people	ca. 3.500 people
Mobility	1 day per participant	1 day per participant	1 week per participant
Field time	1 year	1 year	2 months
Survey area	Germany	ca. 75 German cities	Germany
Time to market	2 years	2 years	1 year

Survey methods:
CAWI, CATI, PAPI



Paper and pencil survey - PAPI

Computer assisted telephone interview - CATI

Computer assisted web interview - CAWI

Traffic flow
Manual traffic counting

Technology
- Manual counting

Use case
- Point-based survey of traffic situation
- Preparation of planning process
- Surveys in connection with the elimination of accident hotspots
- Surveys in connection with the installation of bus lanes
- Surveys in connection with the construction and planning of traffic signals

KPI collected
- Traffic density by number of cars per street and direction

Pro
- Very flexible, no need to install infrastructure

Cons
- Doesn‘t scale
- Selective measurement
- Only reflects traffic situation at a specific location

Traffic flow
Manual traffic counting

Technology
- Manual counting

Use case
- Point-based survey of traffic situation
- Preparation of planning process
- Surveys in connection with the elimination of accident hotspots
- Surveys in connection with the installation of bus lanes
- Surveys in connection with the construction and planning of traffic signals

KPI collected
- Traffic density by number of cars per street and direction

Pro
- Very flexible, no need to install infrastructure

Cons
- Doesn‘t scale
- Selective measurement
- Only reflects traffic situation at a specific location

Traffic flow

Stationary, road-based, automatic traffic counter

Technology
- Inductive loop detector
- Infrared sensor
- Radar sensor

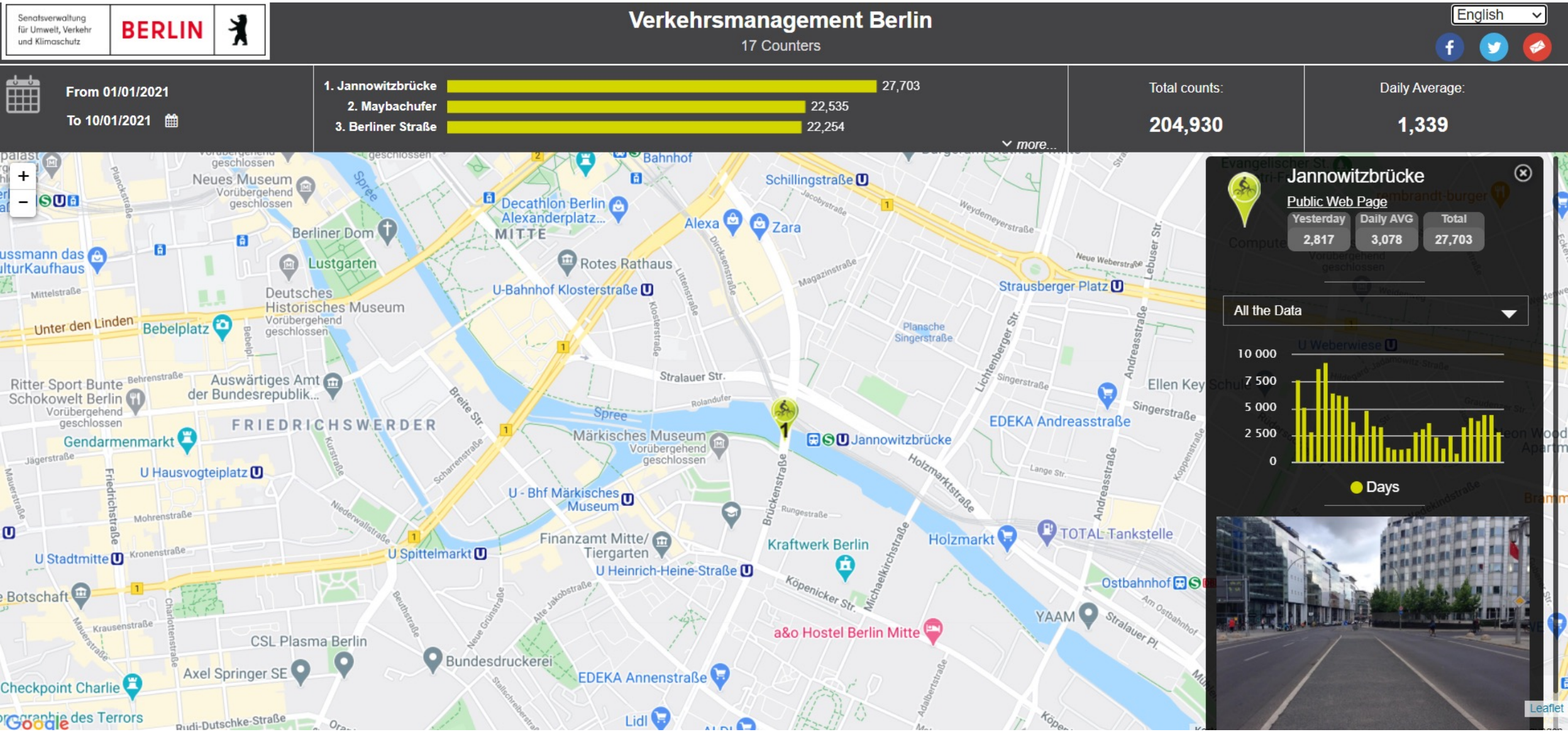
Use case
- Used as strategic detectors for traffic management
- Used for traffic signal control

KPI collected with stationary counter
- Traffic volume / intensity
- Speed
- Congestion
- Occupation time
- Traffic Density
- Vehicle class

Pro
- Proven, robust and calibrated technology used since more than 60 years

Cons
- Point-based measurement:
  - transport model needed to get information about the whole traffic network, secondary road network not covered
  - OD-information is missing
  - Route information is missing

Passive data collection: e.g. Bicycle traffic counter



Source: https://www.eco-public.com/ParcPublic/?id=4728

Traffic flow

Camera-based automatic traffic counter

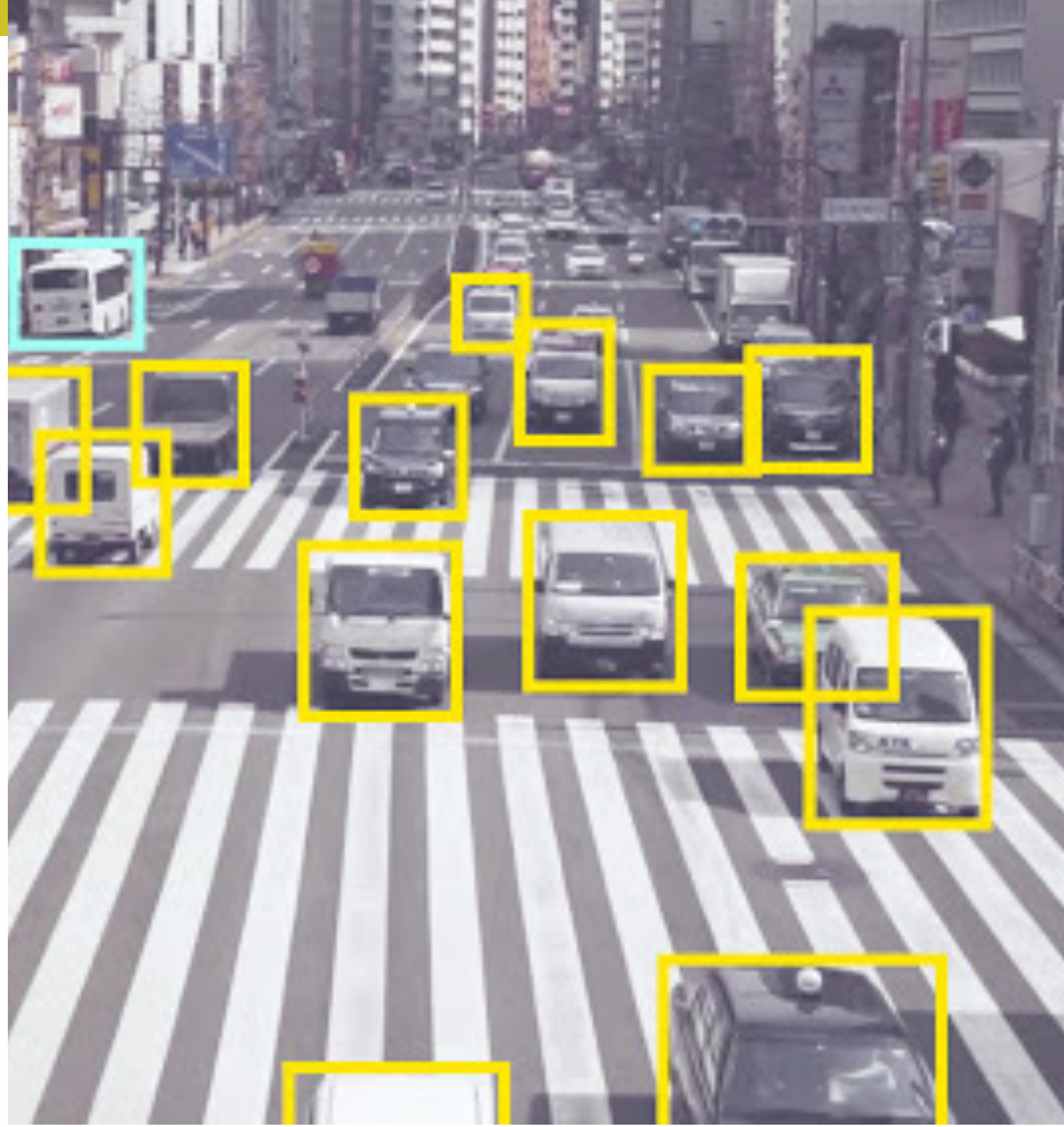
Technology
- Camera + Machine learning

Use case
- Same as manual counting

KPI collected with stationary counter
- Traffic volume / intensity
- Speed
- Congestion
- Traffic Density
- Vehicle type
- Number plate

Pro
- Scalable architecture

Cons
- Reliability and precision depends on calibration



Open source project: build your own camera-based traffic counter
https://github.com/opendatacam/opendatacam
https://www.move-lab.com/project/opendatacam/

Automatic counter systems

Overview

	Walking	Cycling	MIV	PT	Lane accurate	Realtime	Accuracy	Simple installation
swiss ANPR		■■	■■■	■■	■■■	■■■	■■■	■■
swiss LASER	■■■	■■■	■■■		■■■	■■■	■■■	■■■
swiss 3D-SENSOR	■■■	■■■	■■■		■■■	■■■	■■■	■■
swiss KAMERA	■■■	■■■	■■■	■■■	■■■		■■■	■■■
swiss DRONE	■■■	■■	■■■	■■	■■■		■■■	■■
swiss RADAR	■	■■	■■■		■■	■■■	■■	■■■
swiss IoT-RADAR	■■	■■	■■■		■■	■■■	■■	■■■
swiss IoT-INFRARED	■■■	■■				■■■	■■	■■■
swiss IoT-Parking			■■■		■■■	■■■	■■■	■■
swiss Broadcasting	■■■	■■■	■■■	■■■	■■■	■■■	■■■	■■■
BlueScan	■■	■■	■■	■■■		■■■	■■■	■■
MetroCount	■■	■■■	■■■		■■■	■■■	■■■	
Sensys IoT			■■■		■■■	■■■	■■■	■■■

Source: https://www.swisstraffic.ch/content/news/multimodale-zaehler-4_0/

Operation data of mobility service provider

Technology
- Transactional data which get‘s created during operational processes of transport operators and delivered over an API, e.g. by booking of vehicles or E-Ticket-transactions

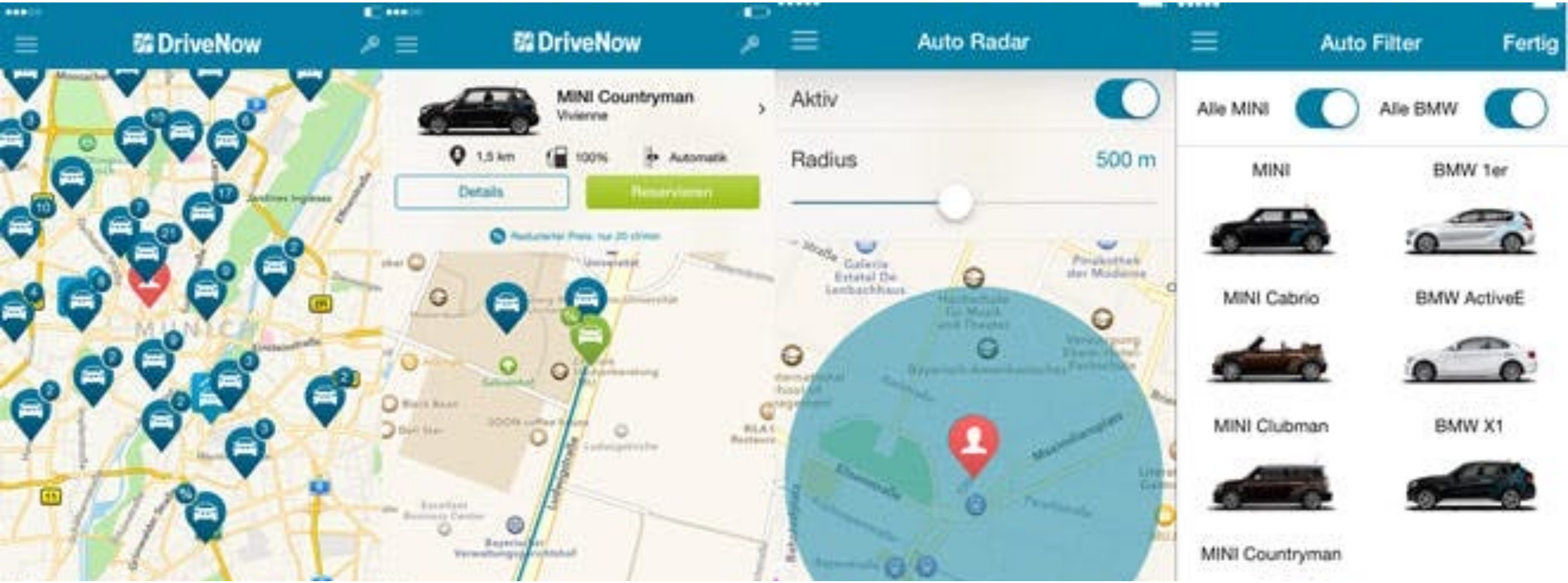
Use case
- Dynamic pricing
- Fleet utilization

KPI collected
- Depending on Use case

Pro
- Scalable architecture

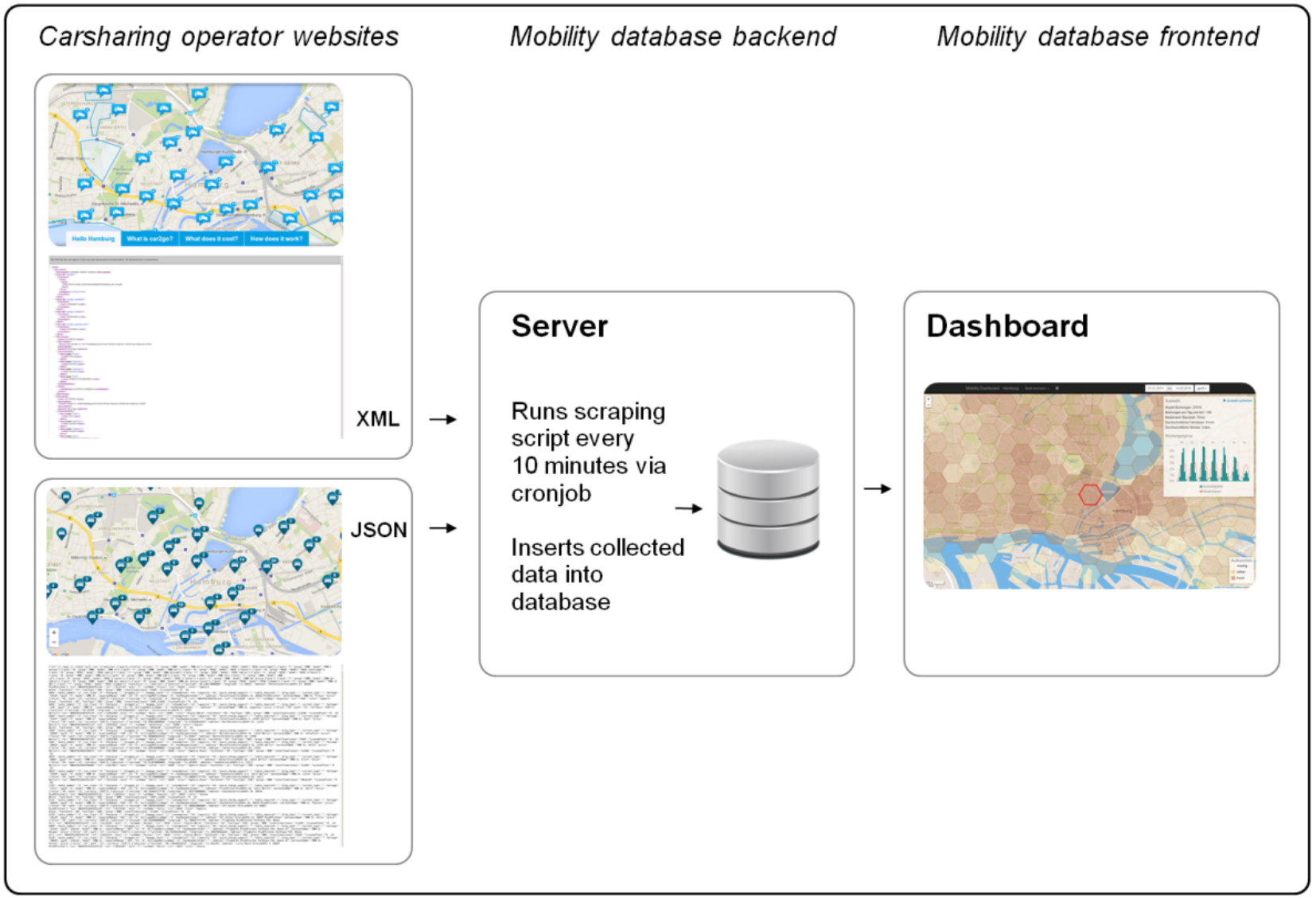
Cons
- Usually only company-internal availability

Using operation data while not being the operator
Webscraping



Using operation data while not being the operator

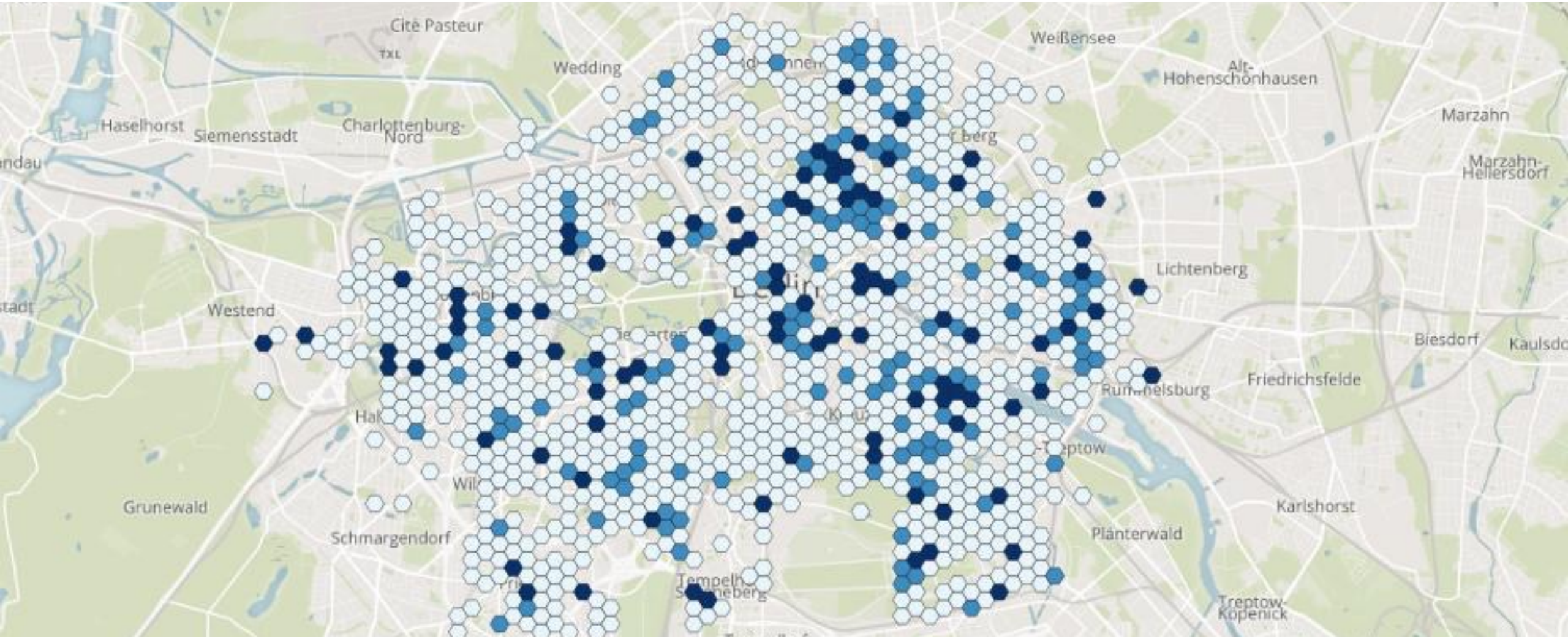
Webscraping



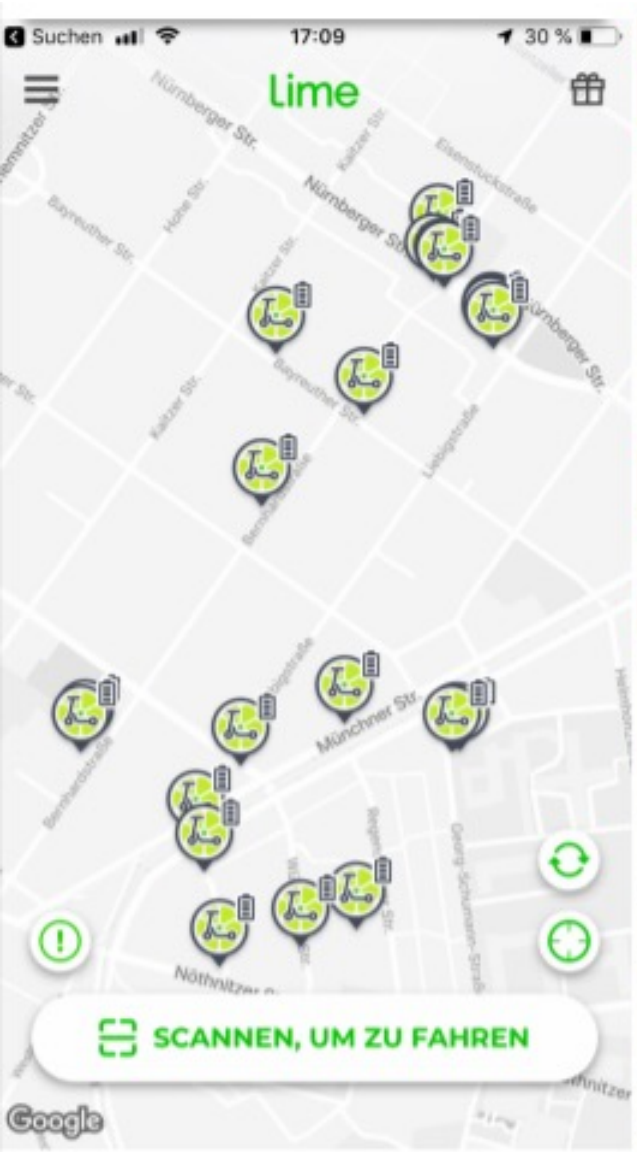
Kortum, Katherine, Robert Schönduwe, Benjamin Stolte & Benno Bock (2016): Free-Floating Carsharing: City-Specific Growth Rates and Success Factors. In: Transportation Research Procedia 19, S. 328-340.

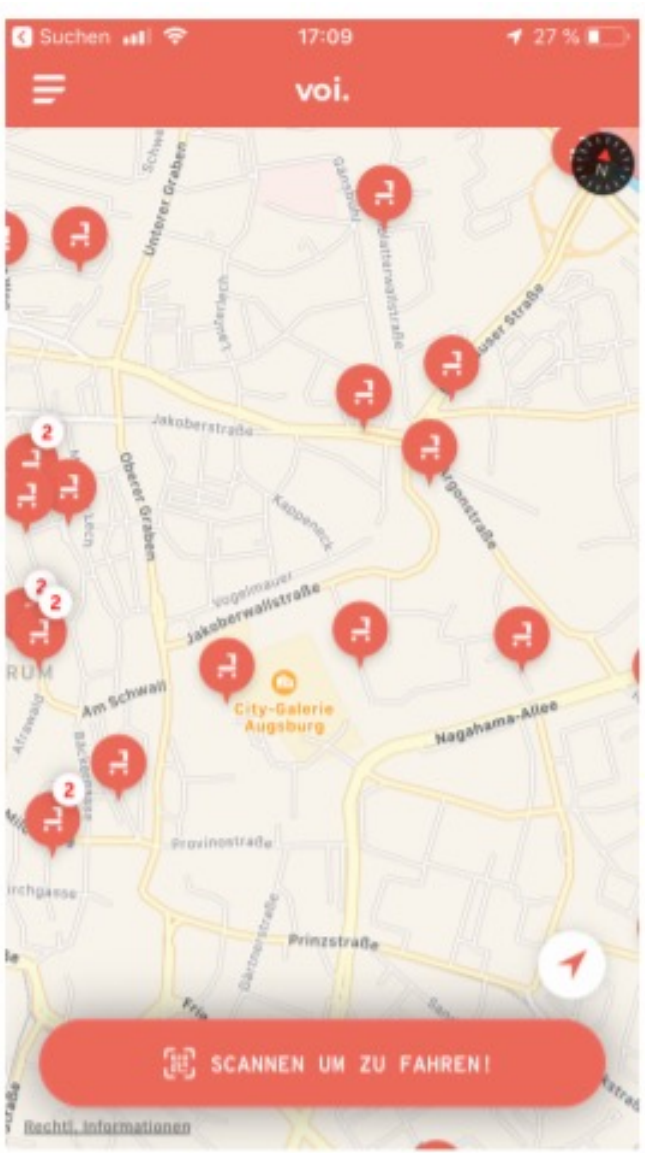
Analysis of carsharing using hotspots

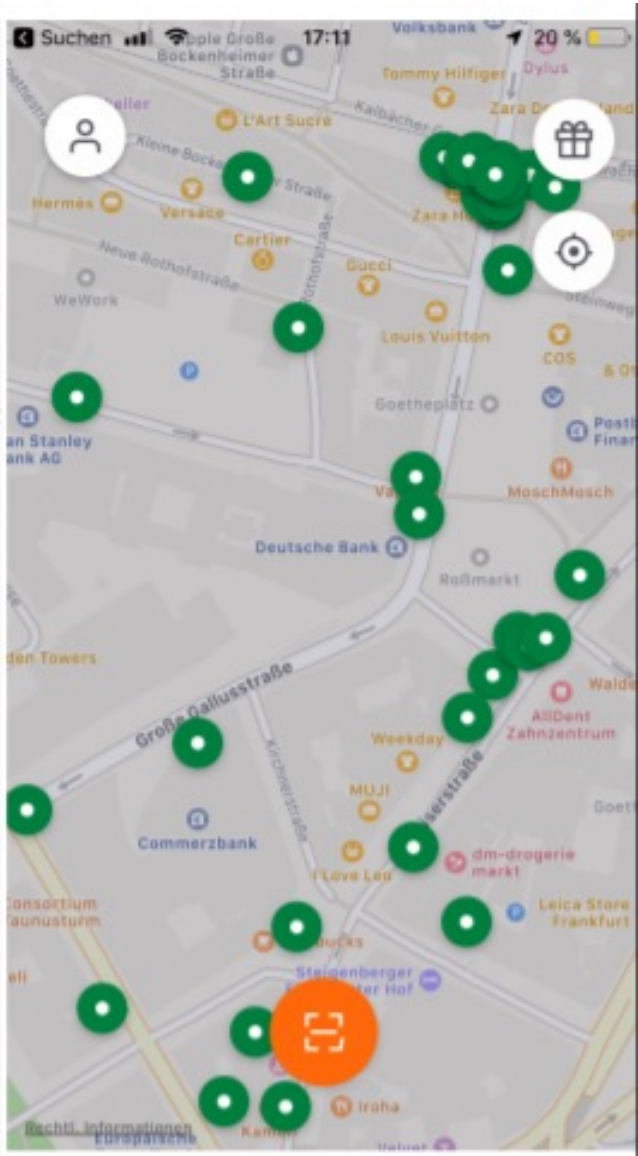
Webscraping



Regulations for the publication of mobility operator data
Mobility data specification (MDS)



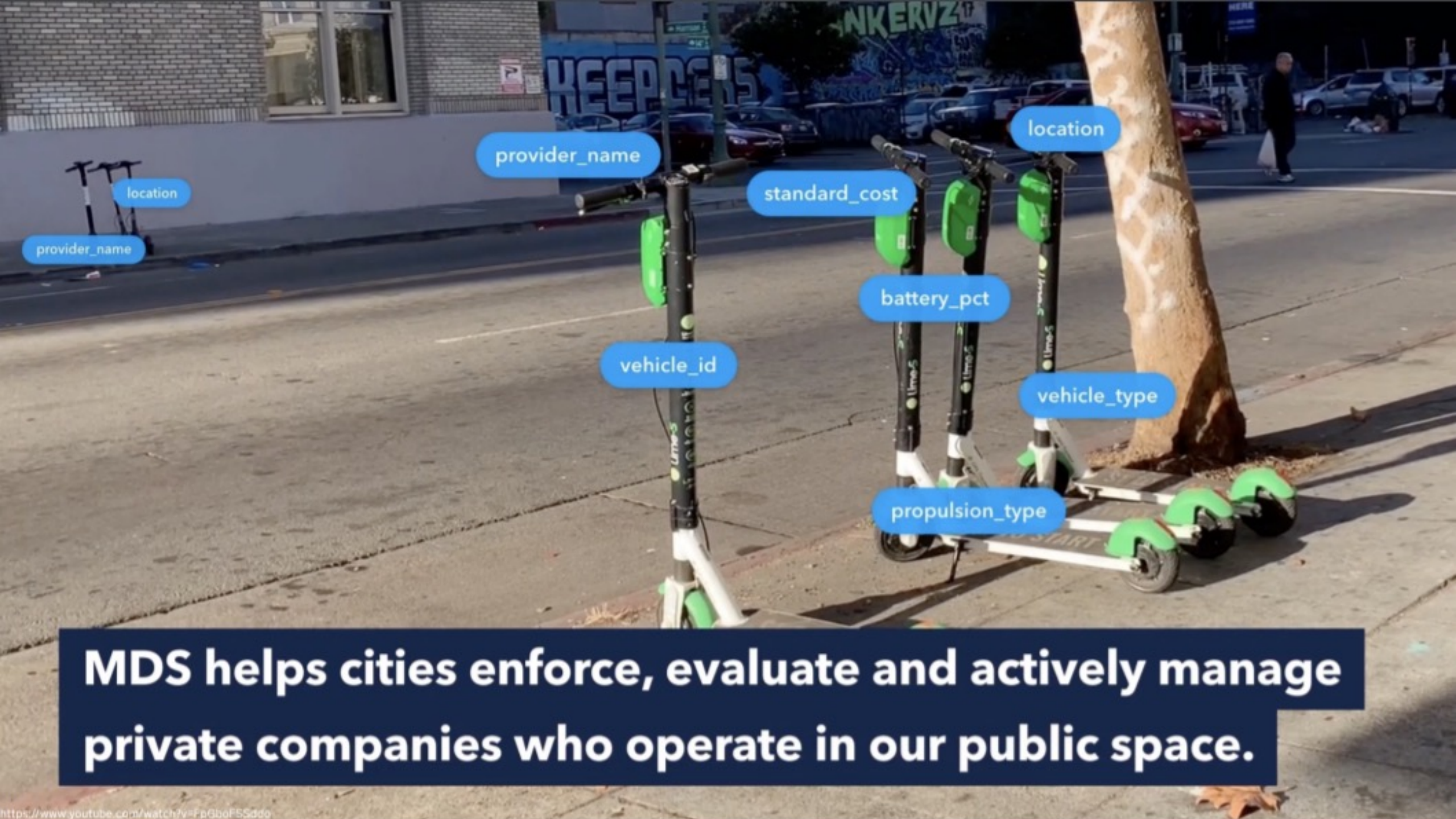


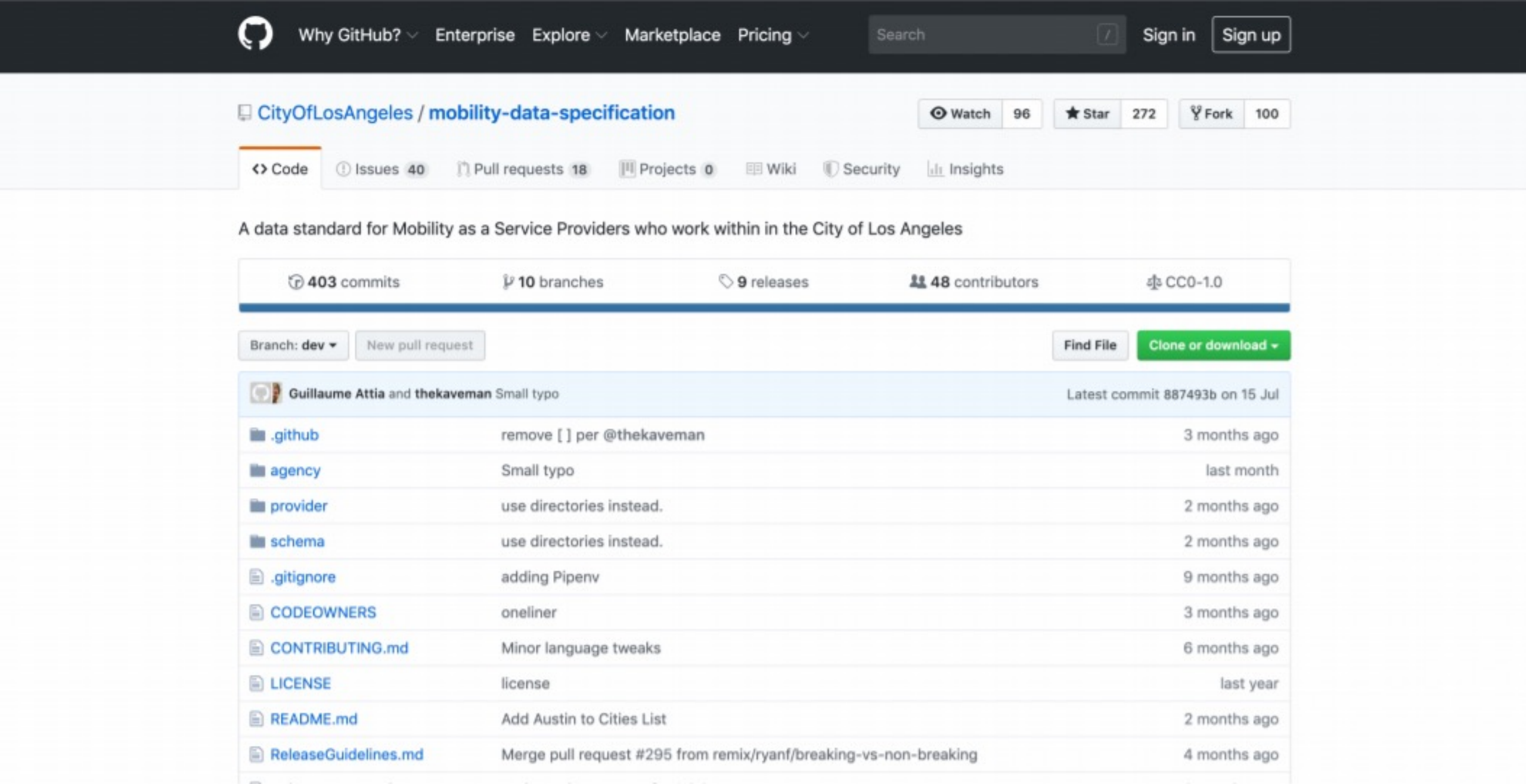


Regulations for the publication of mobility operator data

Mobility data specification (MDS)


```
}, {
    "id": "f4e29ae3-275b-4a75-868f-35c55bdc67ae",
    "short": "ps82",
    "name": "VOI",
    "zone": 145,
    "type": "como",
    "registration_plate": "",
    "added": "2019-06-12T13:20:19.239036Z",
    "serial": null,
    "model_specification": null,
    "status": "ready",
    "bounty": 0,
    "location": [52.50433349609375, 13.33386516571045],
    "battery": 97,
    "locked": true,
    "updated": "2019-09-02T06:04:23Z",
    "mileage": 0
}, {
```



location
provider_name
provider_name
standard_cost
location
battery_pct
vehicle_id
vehicle_type
propulsion_type
MDS helps cities enforce, evaluate and actively manage private companies who operate in our public space.



github.com/CityOfLosAngeles/mobility-data-specification

Operation data of non-mobility companies

Technology
- Transactional data which get‘s created during operational processes of non-transport operators and delivered over an API

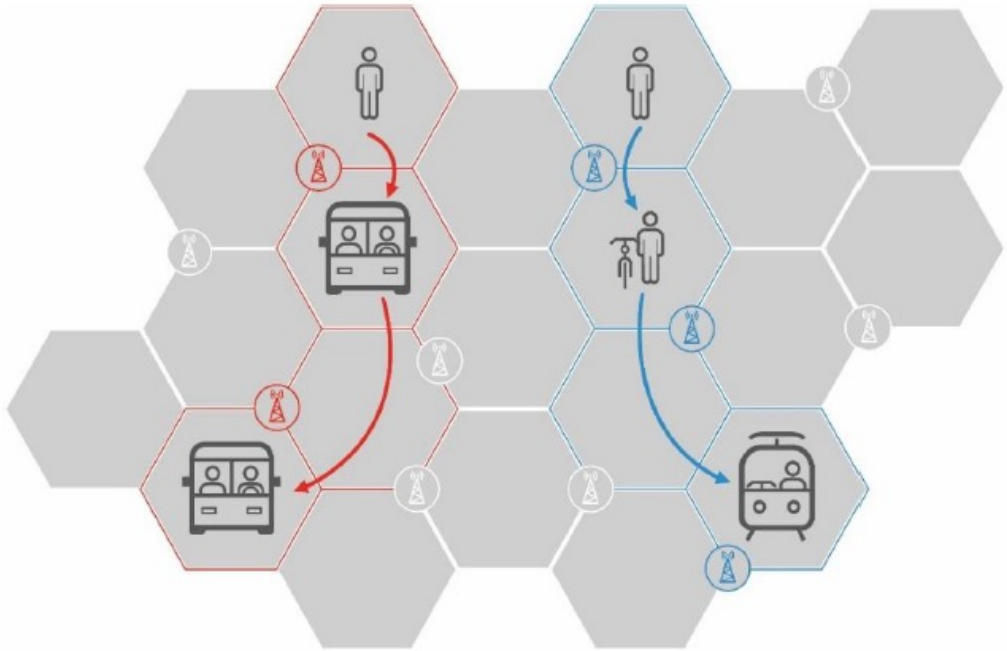
Use case
- OD Matrices
- Transport modelling

KPI collected
- Traffic volume / intensity
- Traffic Density
- Mode of transport
- OD Matrices

Pro
- (Very) big samples
- High availability in near real time

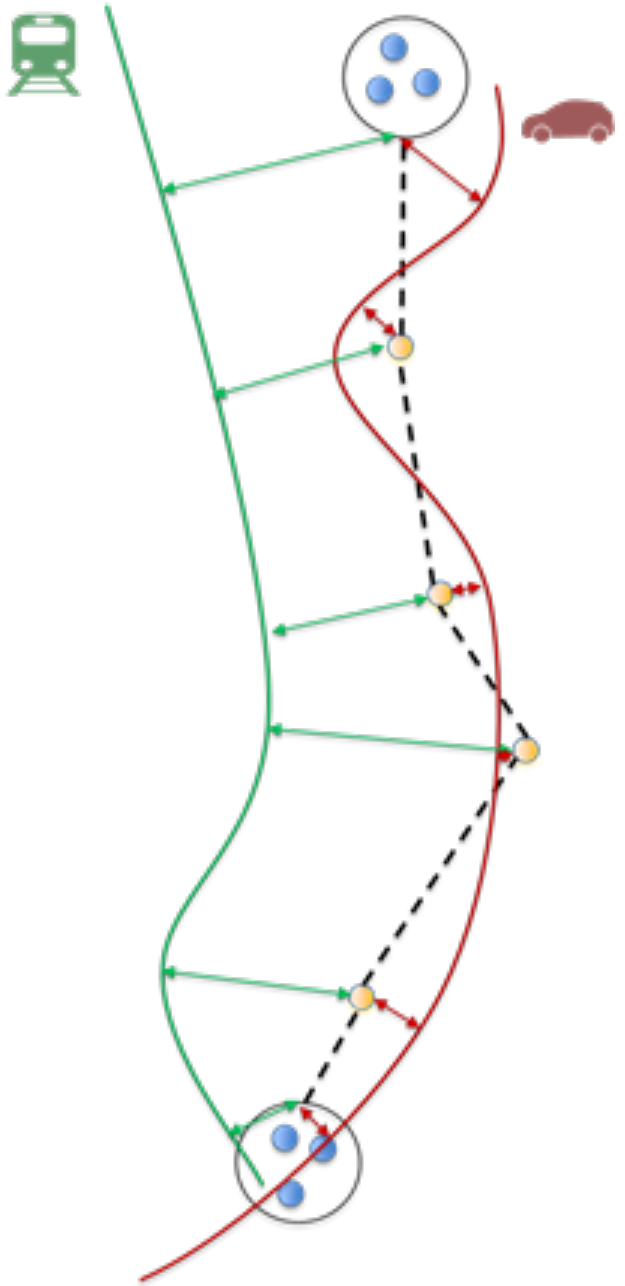
Cons
- High costs
- Only limited information about mode of transport

Use of mobile network data in the transport sector



Cell Events

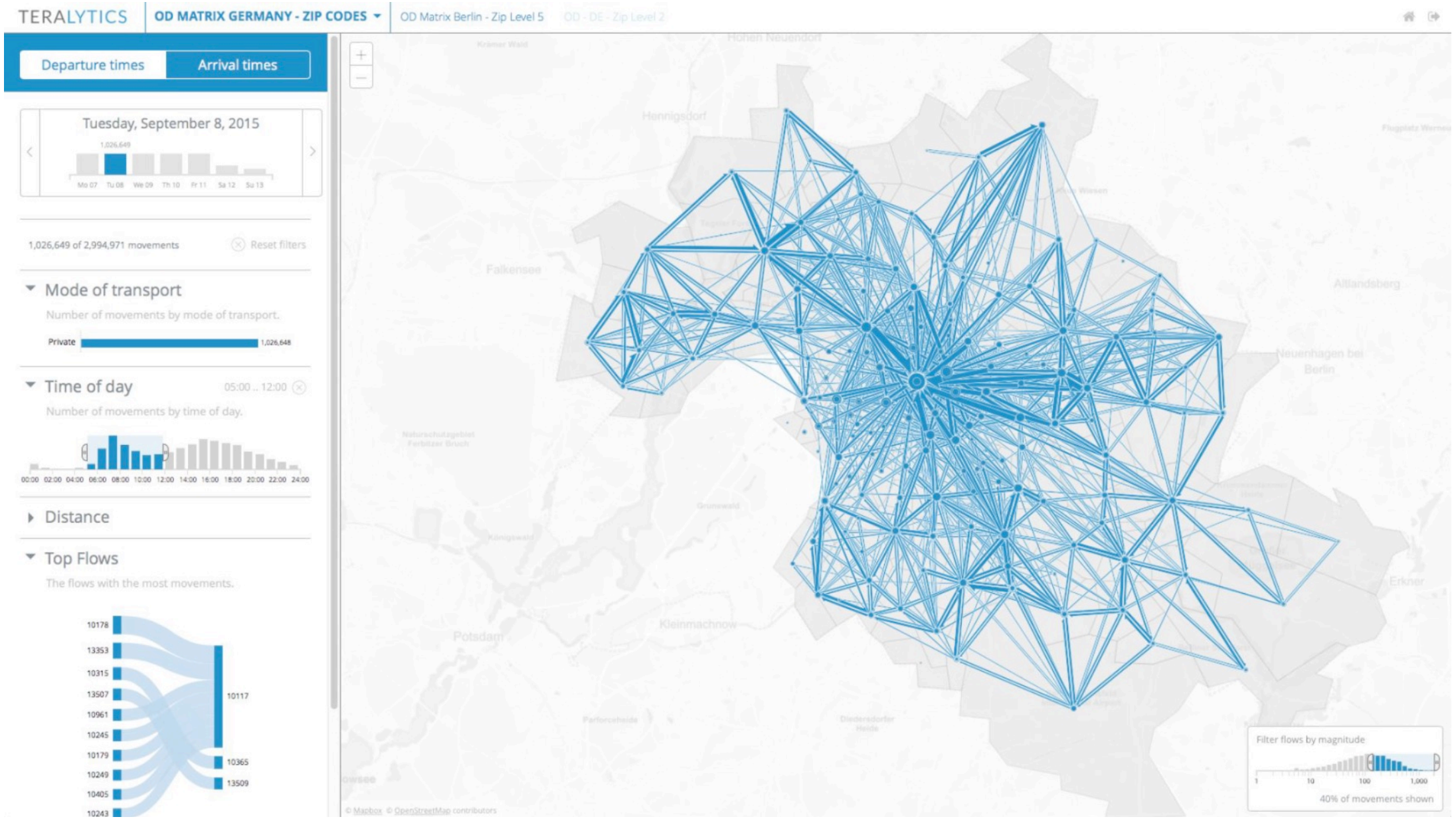
- Calls
- Short messages
- Data transfer
- Phone turned on and switched off
- Cell change



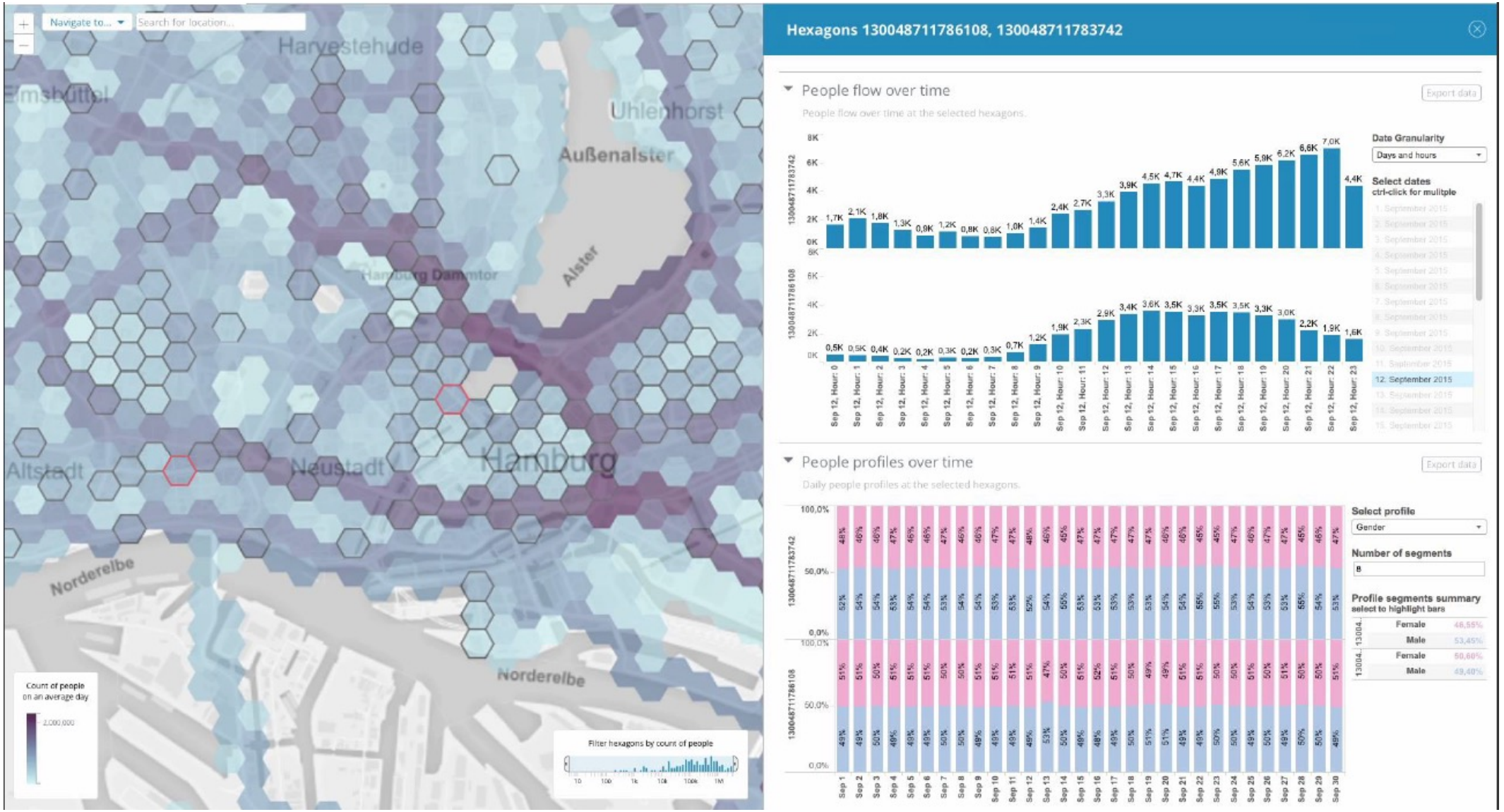
Real-world example of mobile network cell distribution

Berlin-Mitte

Mobile network data



Mobile network data



Floating Car Data: Real-time traffic flow information

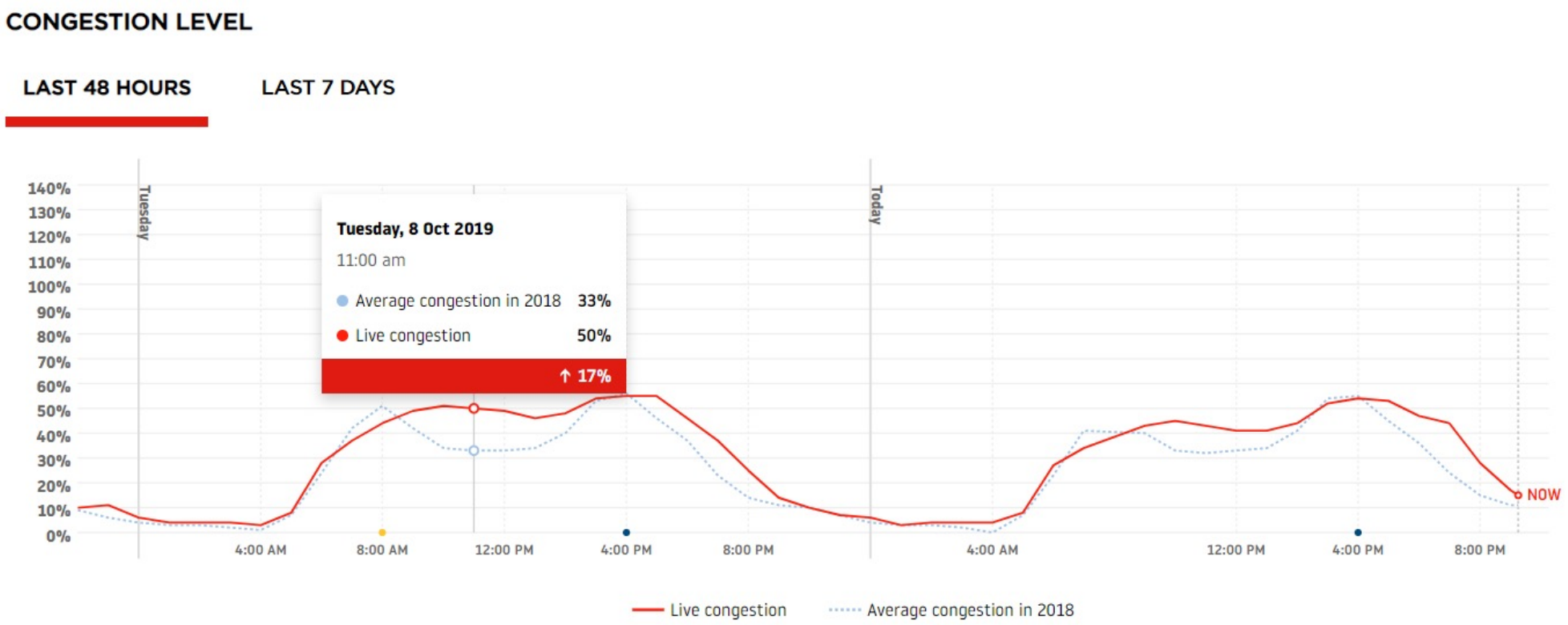
https://www.tomtom.com/en_gb/traffic-index/berlin-traffic

Berlin

12 °C

LOCAL TIME

9:29 PM



Floating Car Data: Real-time traffic flow information

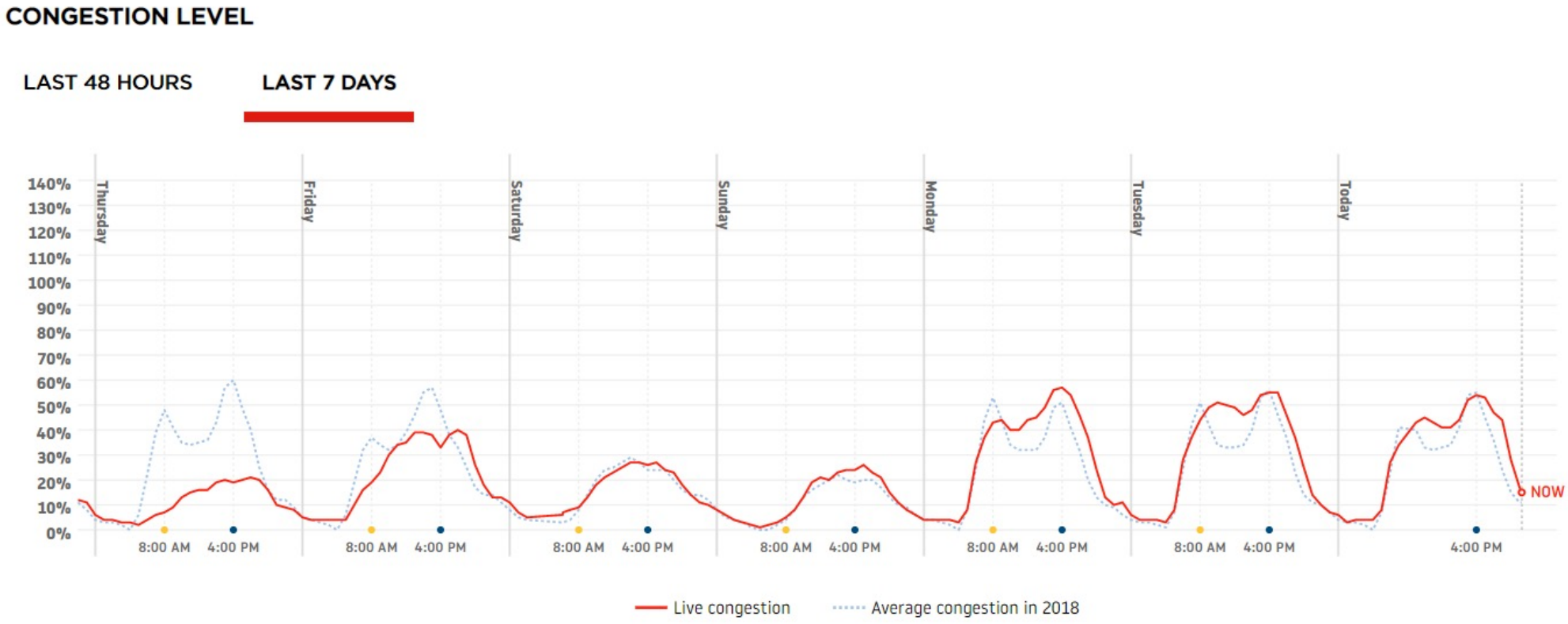
https://www.tomtom.com/en_gb/traffic-index/berlin-traffic

Berlin

12 °C

LOCAL TIME

9:24 PM



"The pandemic as a catalyst for digitization".
Roadshow of the mobility data providers

Mobile network data

TERALYTICS

Smartphonebased GPS-data

umlaut

Google

Transaction data and logfiles

moovit

Second Measure

Automatic counter

hystreet.com

Floating car data

INRIX

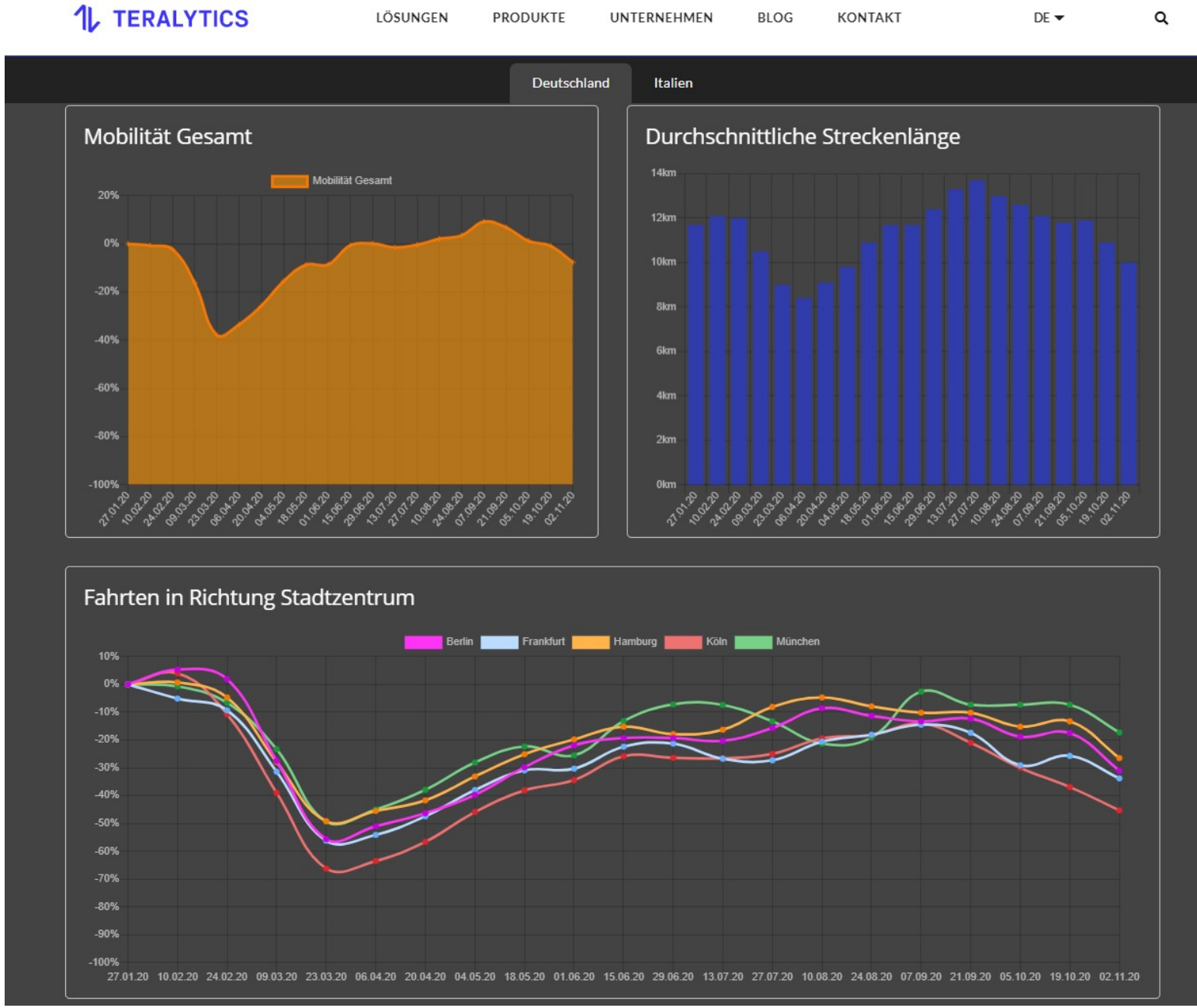
TOMTOM

Mobile network data

TERALYTICS

www.teralytics.net/de/ressourcen-zu-covid-19

Looking at mobility as a whole and between regions, how is it developing?



Transaction data and logfiles

covid19.apple.com/mobility

How is the mobility behavior of certain population groups developing?

Maps

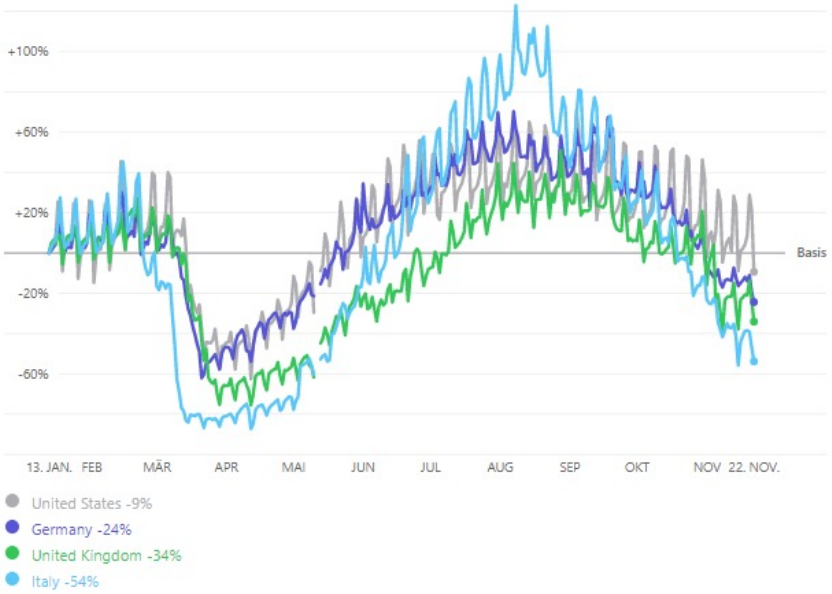
Berichte zu Mobilitätstrends

Informiere dich über die durch Covid-19 veränderten Mobilitätstrends. Die Berichte werden täglich veröffentlicht und spiegeln Anfragen nach Wegbeschreibungen in Apple Karten wider. Datenschutz zählt bei uns zu den Kernwerten. Daher bringt die App „Karten" deine Daten nicht mit deiner Apple-ID in Verbindung und Apple speichert nicht die Orte, an denen du warst.

Mobilitätstrends

Änderung in den Anfragen zur Routenführung seit 13. Januar 2020

Suche (z. B. Italy, California oder New York City)

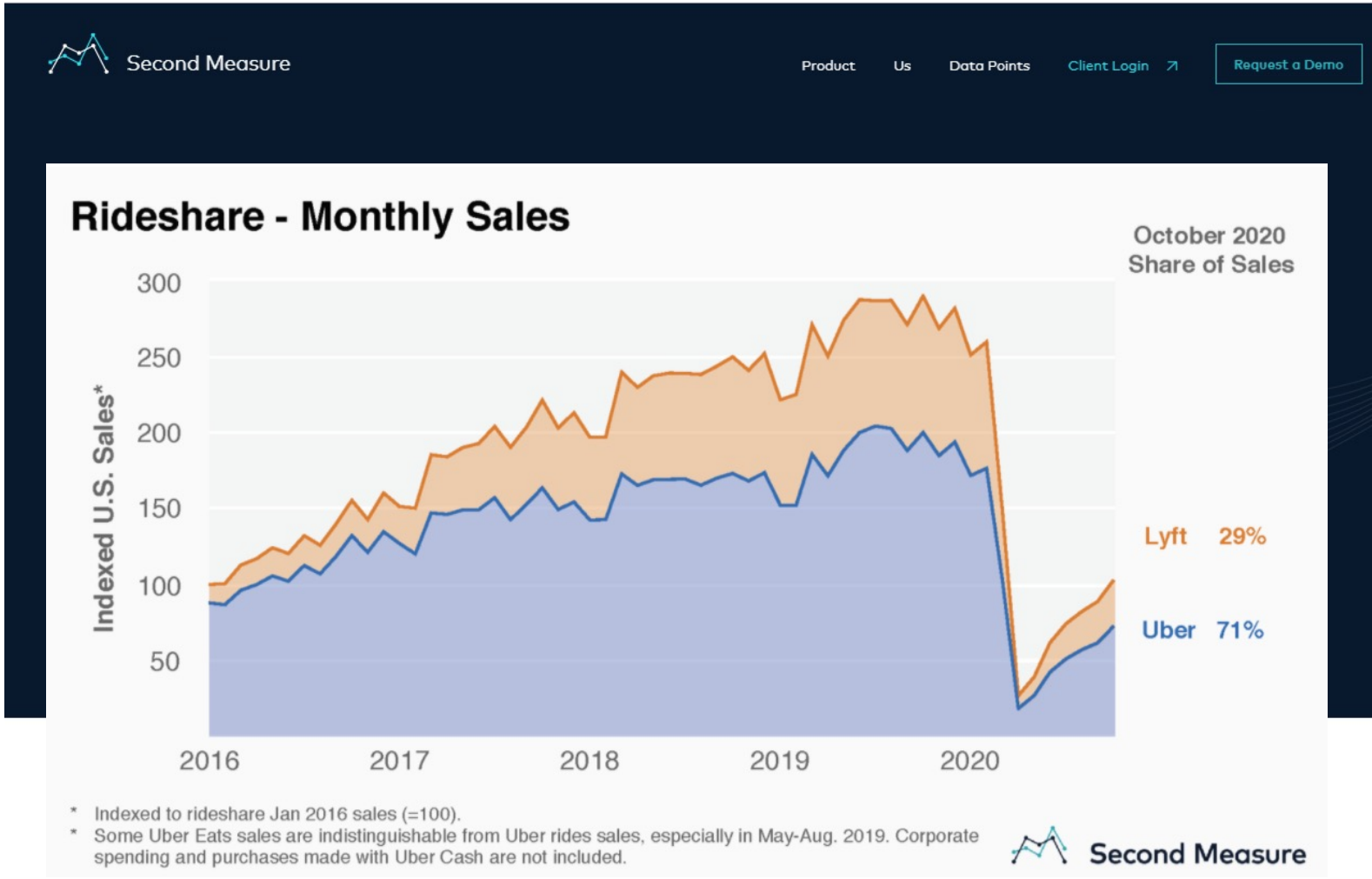


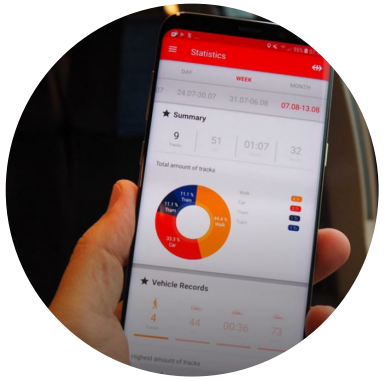
Transaction data and logfiles

Second Measure

secondmeasure.com/datapoints/rideshare-industry-overview/

How are new business models of mobility service providers developing?





Smartphone based GPS data

Google

www.google.com/covid19/mobility/

What changes are taking place in people's everyday lives?

Google Mobilitätsberichte zur Coronakrise

Bayern

Einzelhandel und Freizeit

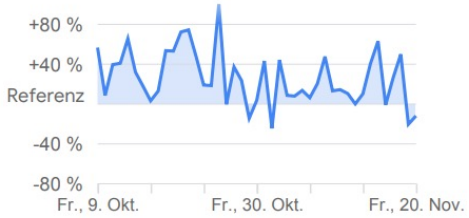
-34 % im Vergleich zur Referenz

Läden für den tägl. Bedarf

-1 % im Vergleich zur Referenz

Parks

-12 % im Vergleich zur Referenz



Bahnhöfe und Haltestellen

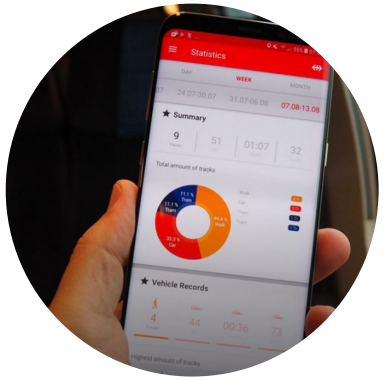
-35 % im Vergleich zur Referenz

Arbeitsstätten

-19 % im Vergleich zur Referenz

Wohnorte

+11 % im Vergleich zur Referenz



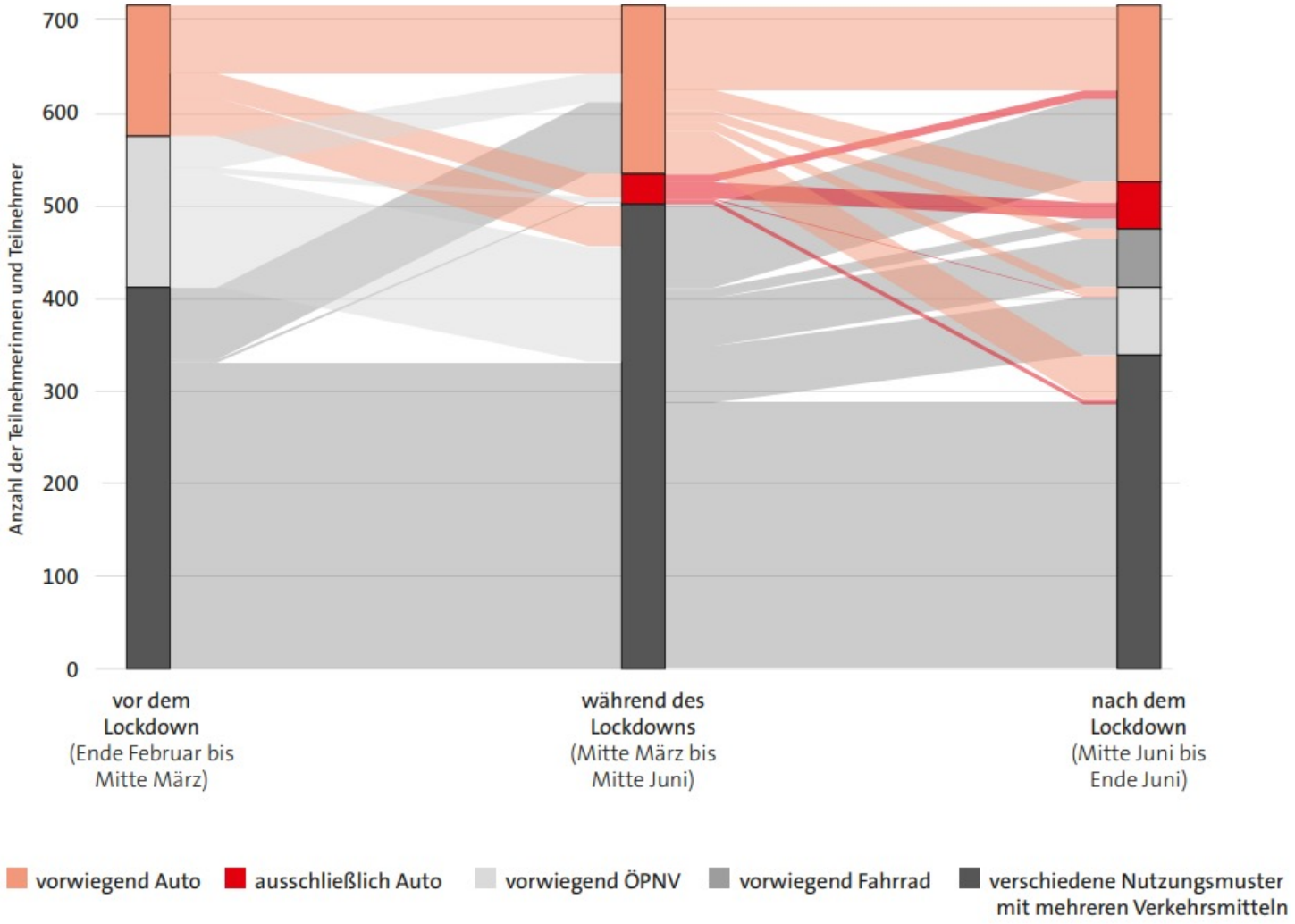
Smartphonebasierte GPS-Daten

ivtmobis.ethz.ch/mobis/covid19/reports/

www.infas.de/neuigkeit/die-mobilitaet-und-corona-schneller-als-gedacht-zurueck-zur-normalitaet/

What are the long-term implications of the pandemic for transportation behavior?

Verkehrsmittelnutzung im Tracking – Wanderungen der Nutzertypen vor, während und nach dem Lockdwon



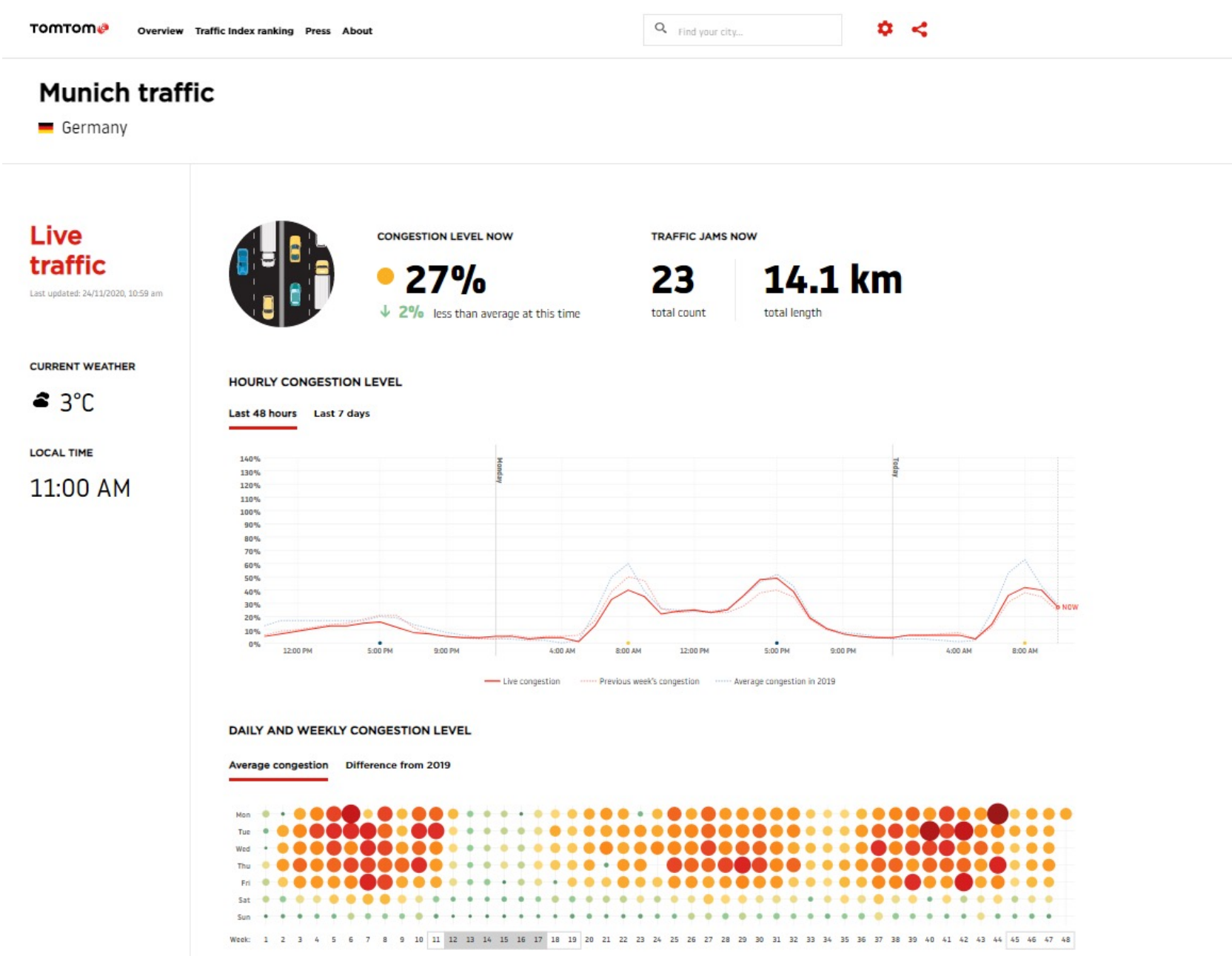
Basis ca. 700 Teilnehmende ohne Unterbrechung des Trackings, Tracking-Ergebnisse Eigenstudie MOTIONTAG

Floating car data

TOMTOM

www.google.com/covid19/mobility

What impact will the pandemic have on car traffic?

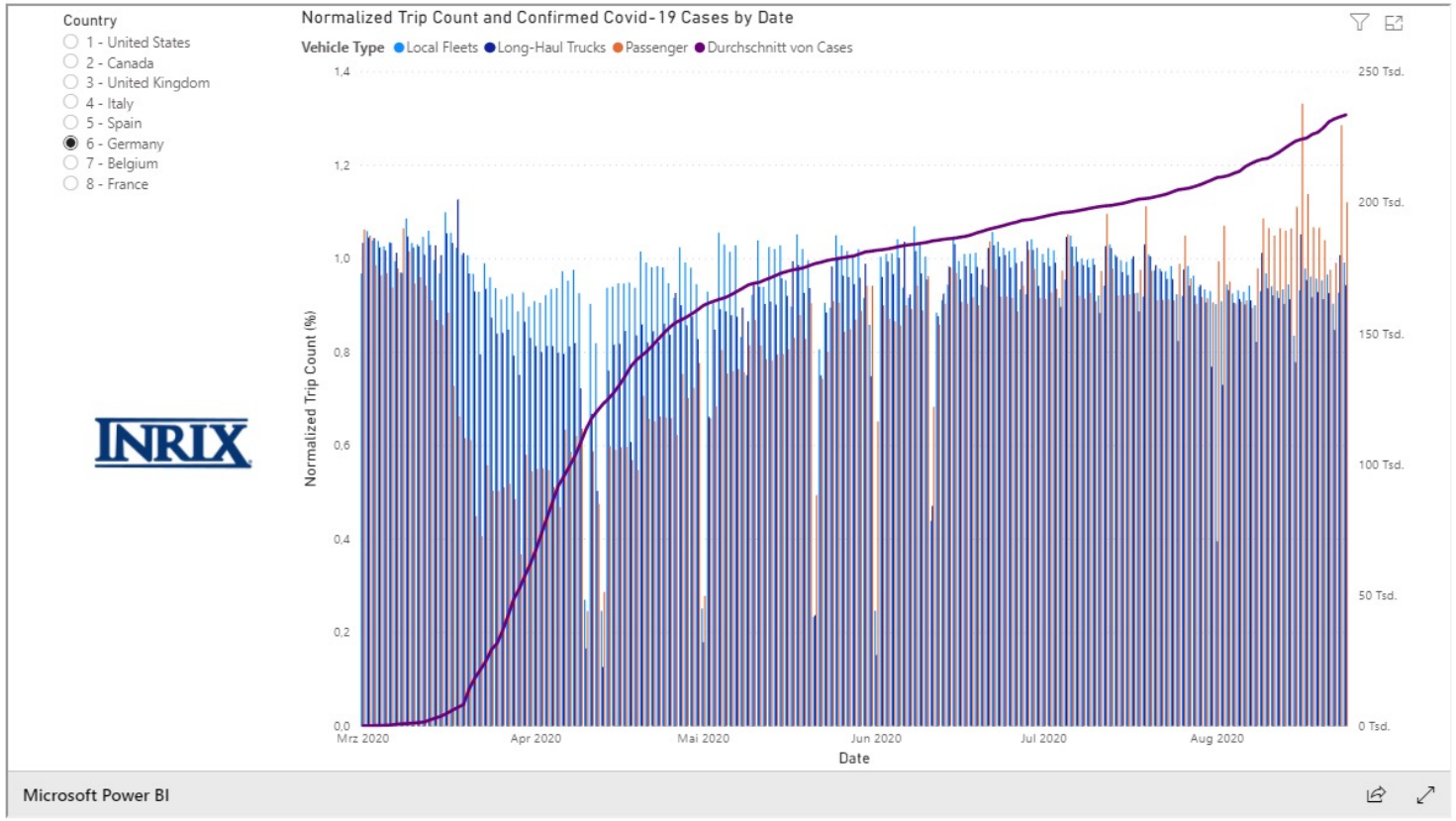


Floating car data

INRIX

inrix.com/covid-19-transportation-trends/

What impact will the pandemic have on car traffic?

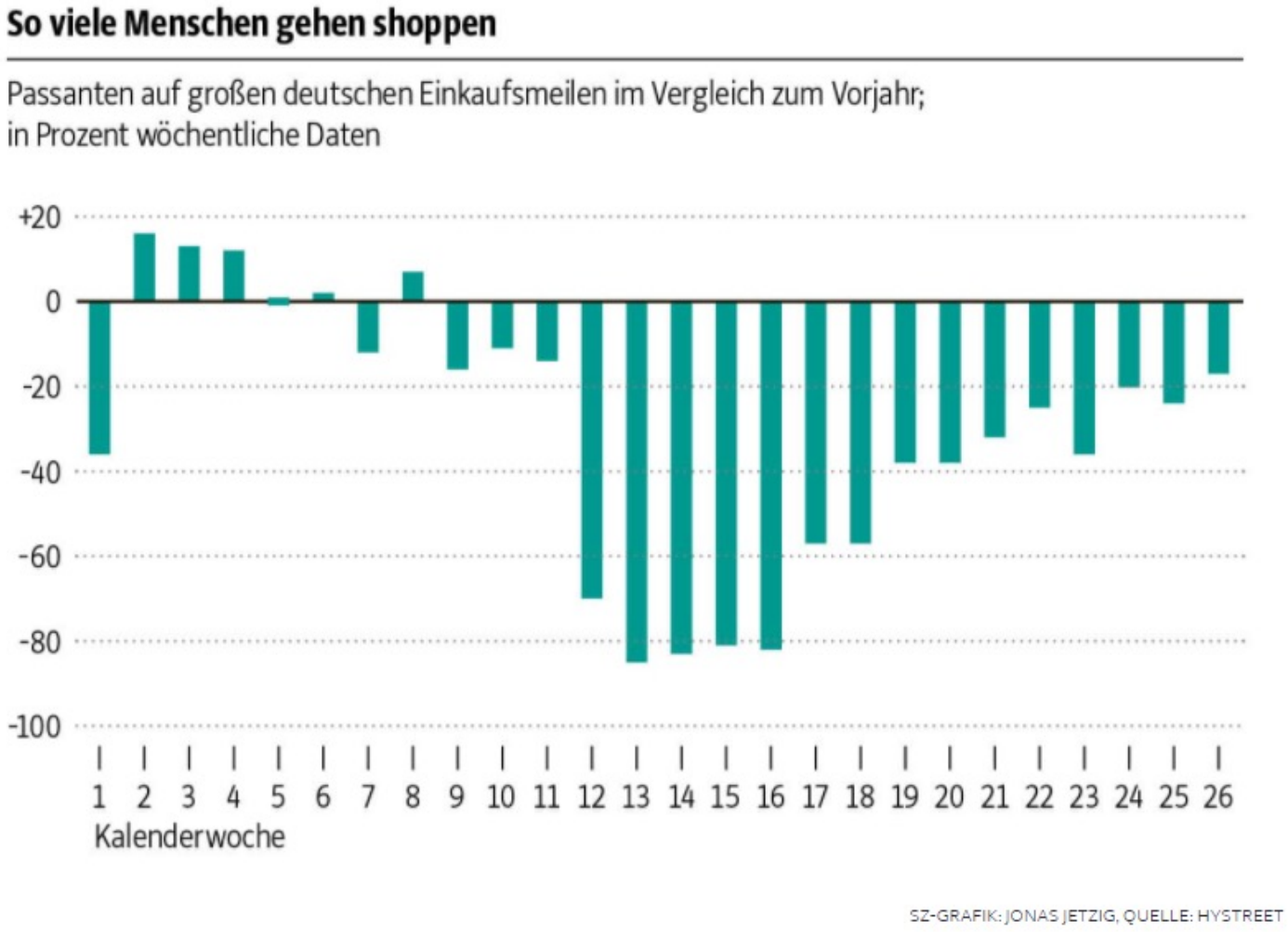


Automatische Zählstellen

hystreet.com

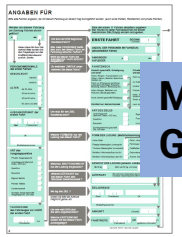
www.sueddeutsche.de/wirtschaft/aktienmarkt-alternative-daten-wirtschaft-1.4958199

What impact will the pandemic have on shopping behavior and city centers?



Advantages of new data sources

- Timeliness
- Availability
- Volume
- Level of detail



MiD – Mobility in Germany

(several) millions

Transaction-data

Floating Phone / Car Data

Apps

Freqency of data collection

(Several) Years

SrV

German Mobility Panel

Movements

Seconds

Automatic counter

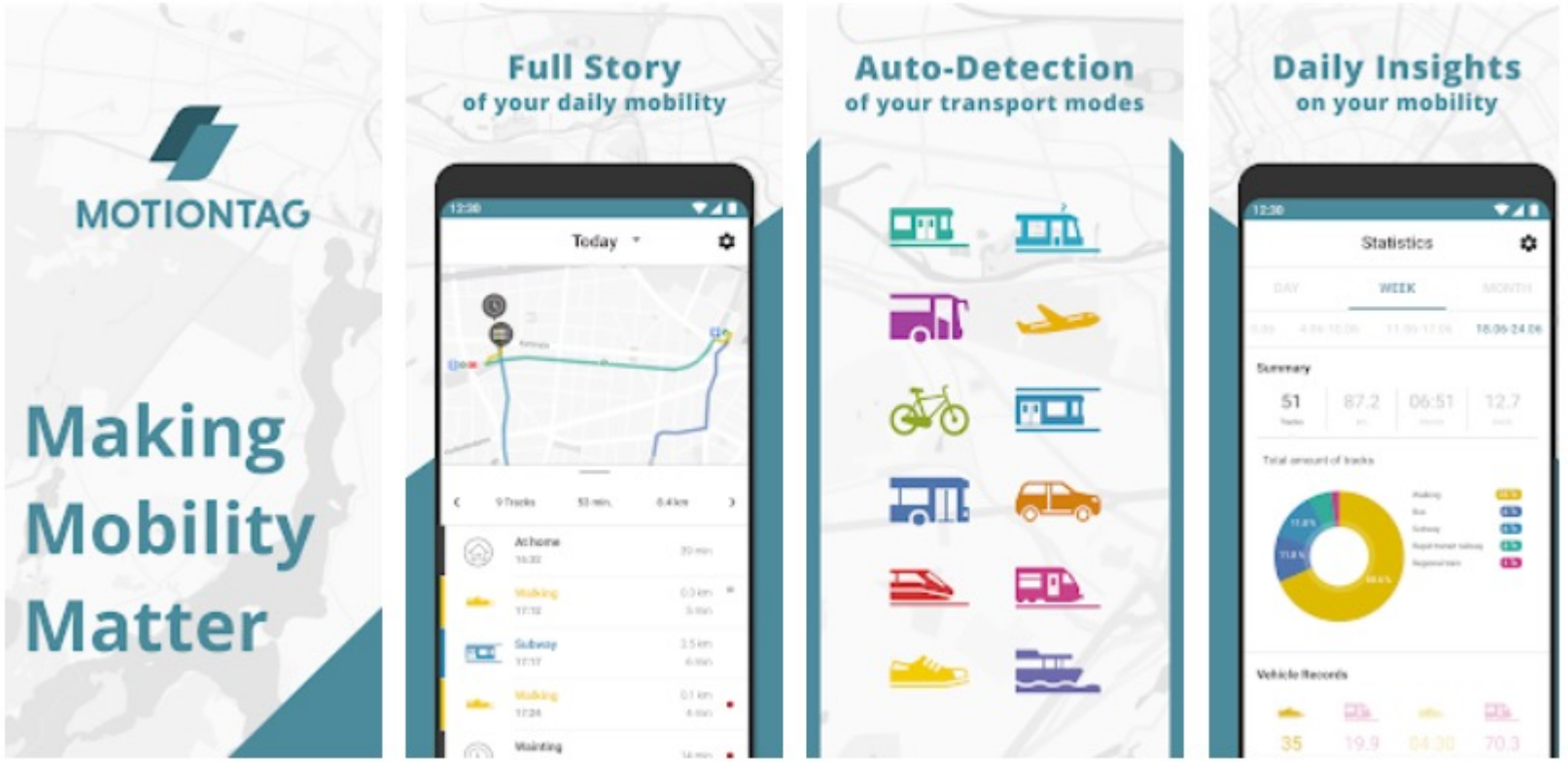
(several) Thousands

[5]

Introduction to the data sources used in this course

Homework: Data collection for next lecture

Smartphone-based GPS tracking data



Download the MOTIONTAG-App from the store.
Use it for some days.

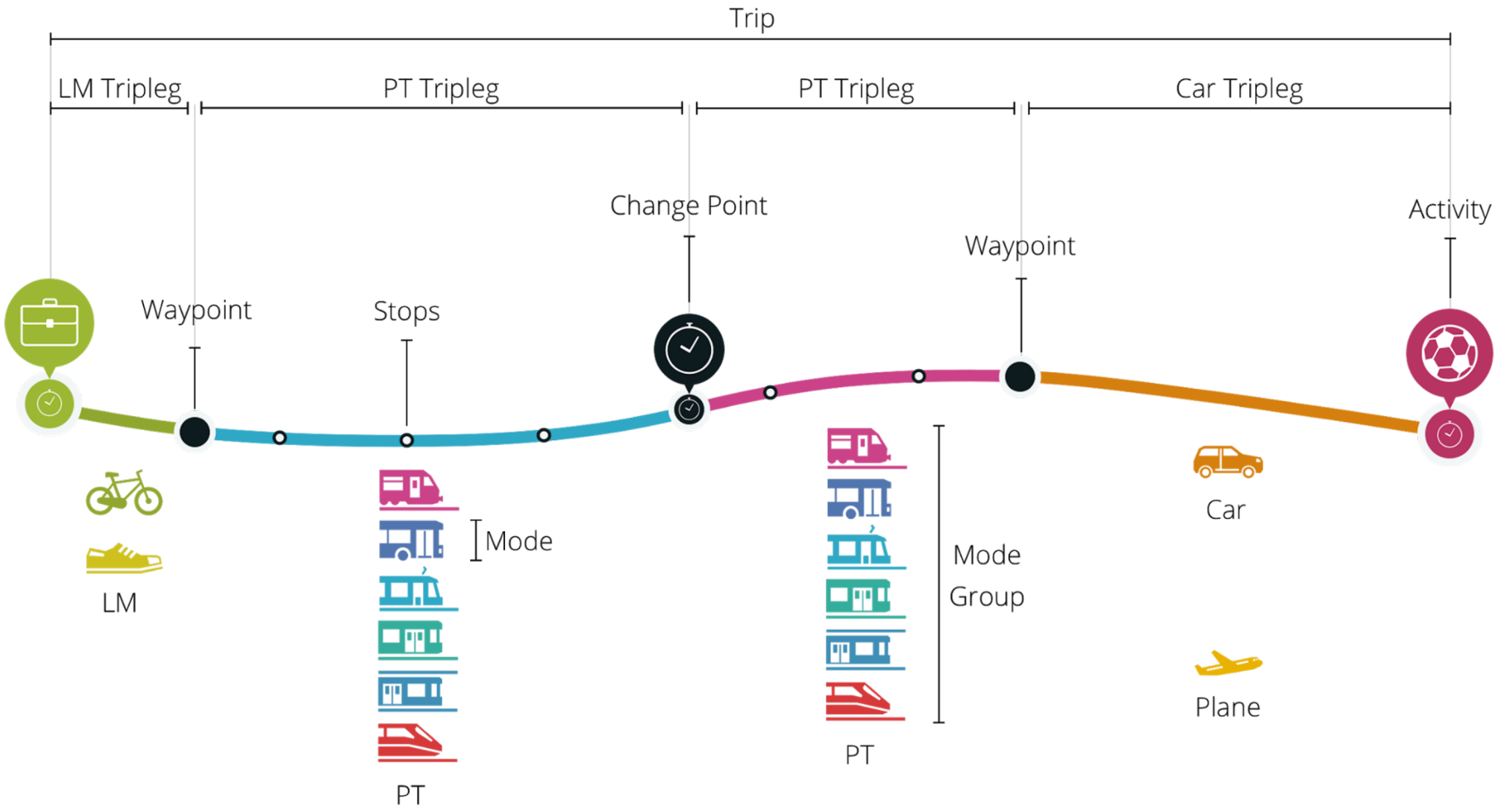
We will use the data in the next lecture on 02.02.2021

Download on the App Store

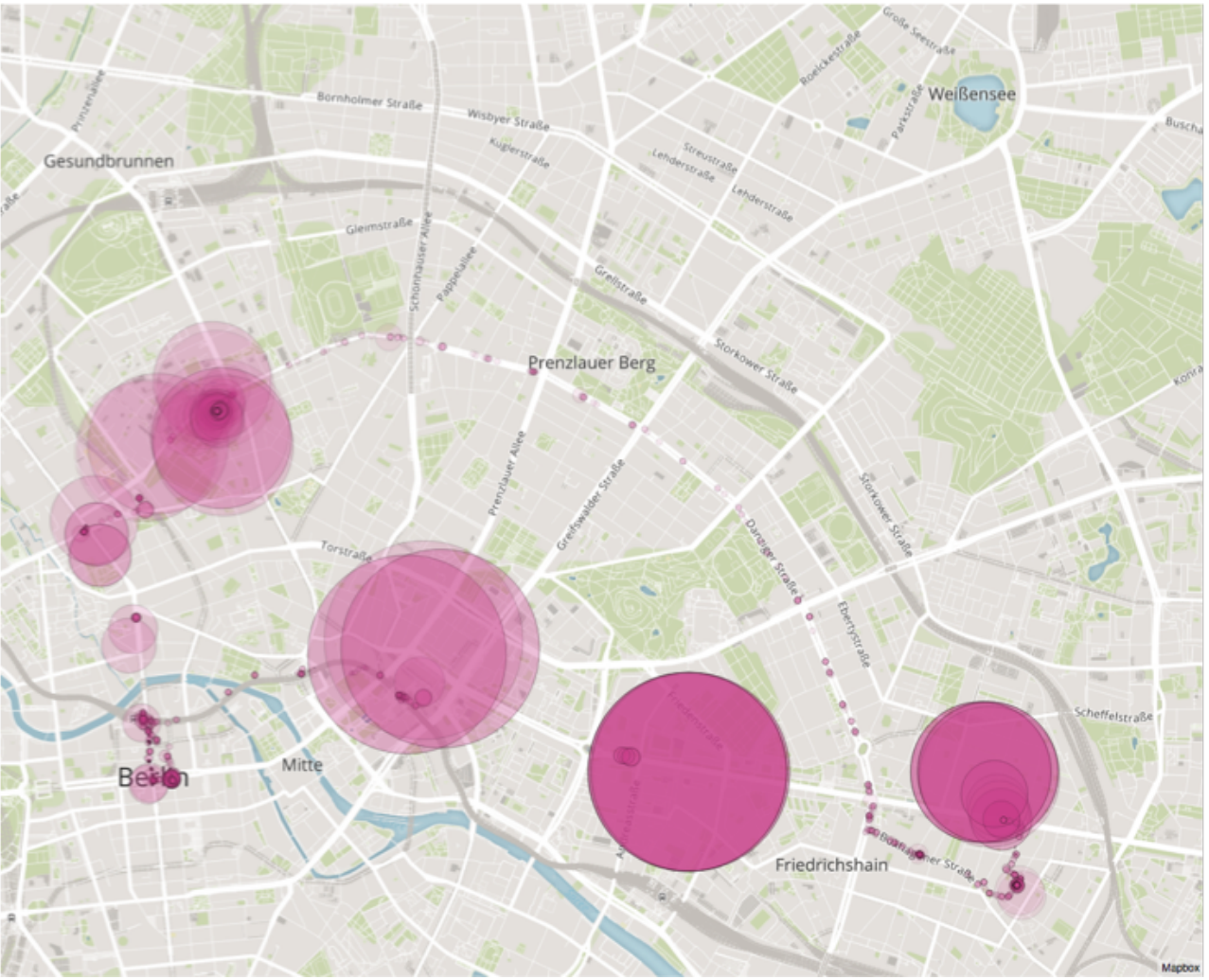
GET IT ON Google Play

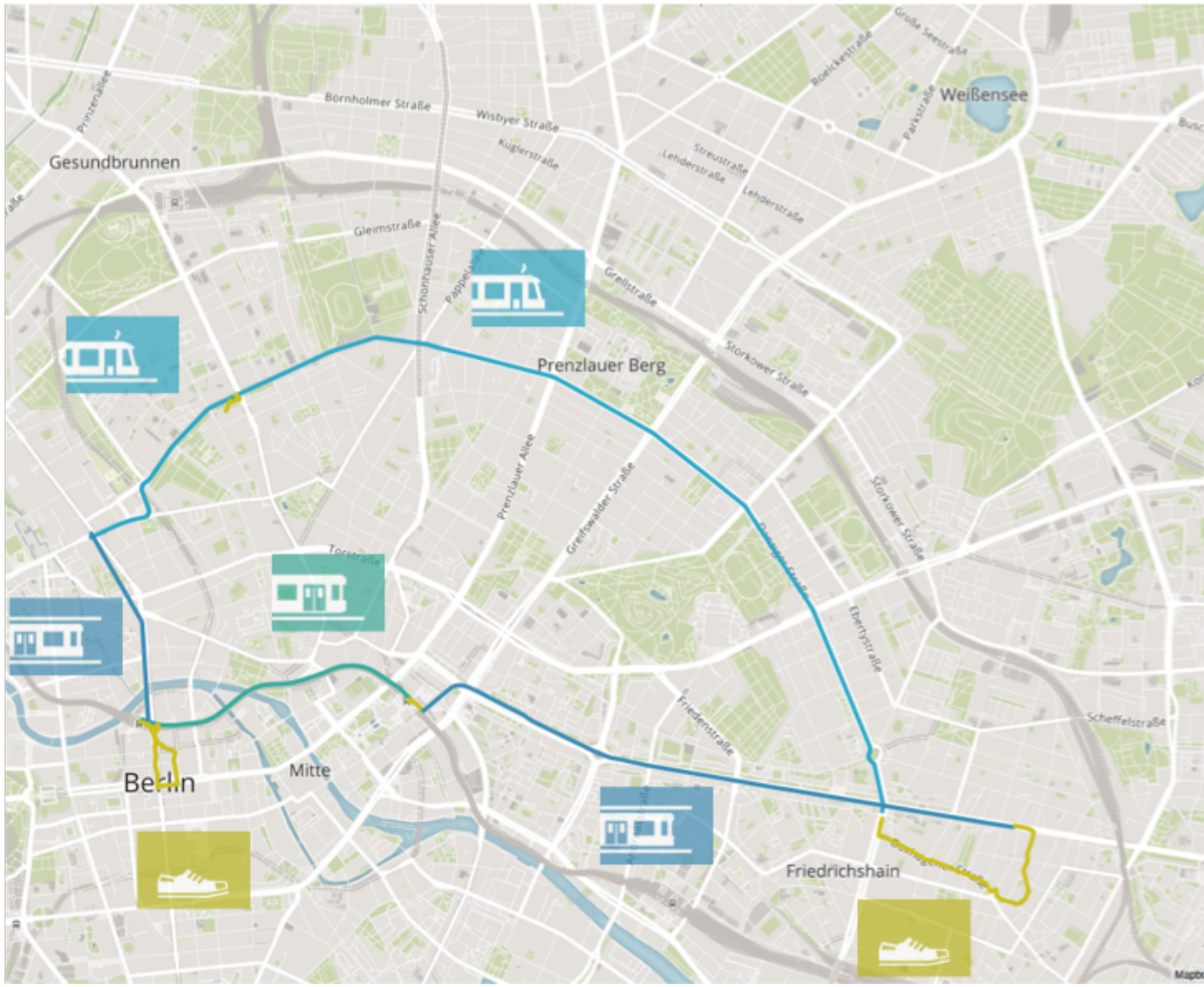
From Smartphone sensor data to actionable insights

Automatic detection of tracks and locations

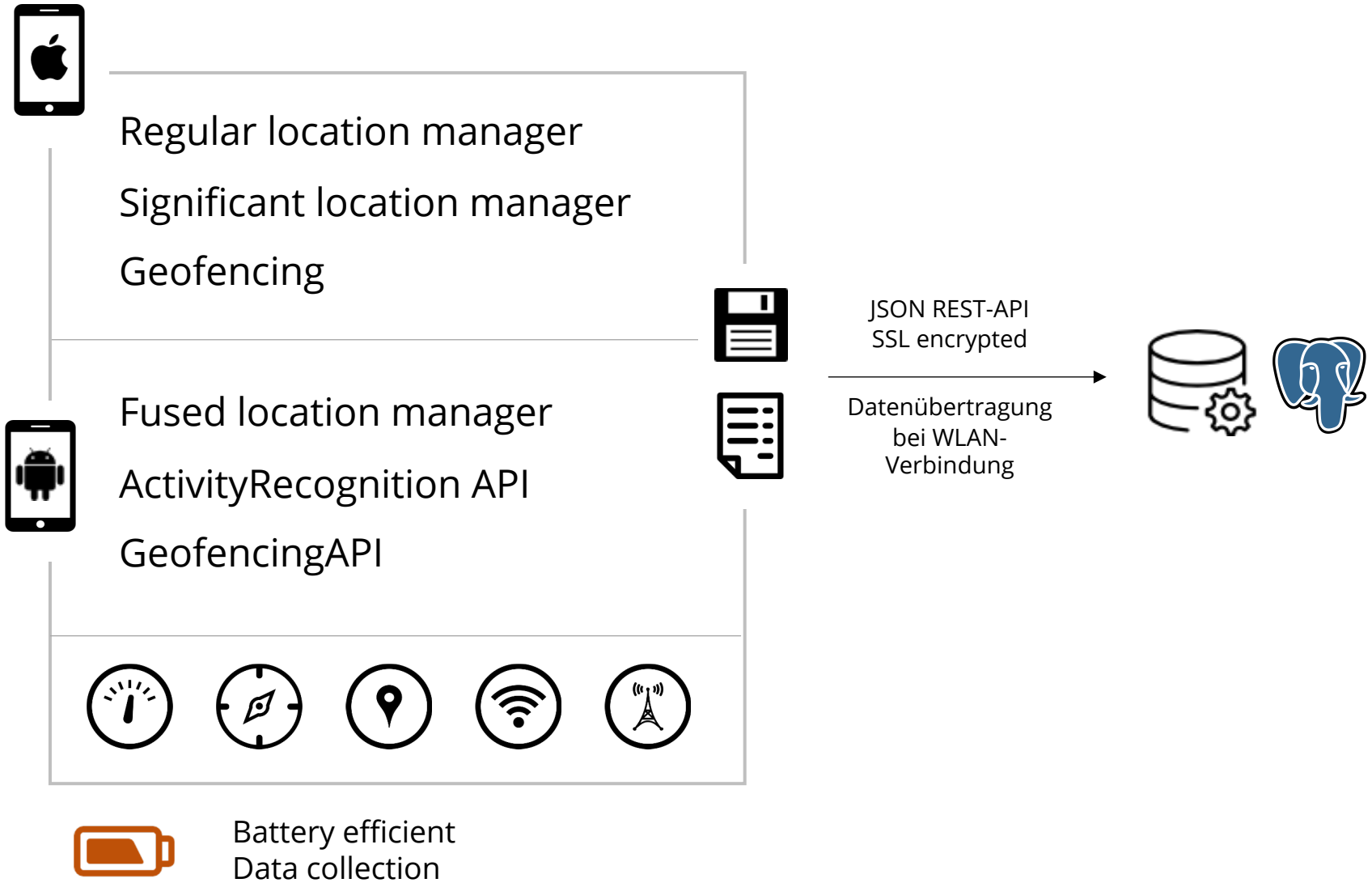


From sensor data to labeled tracks

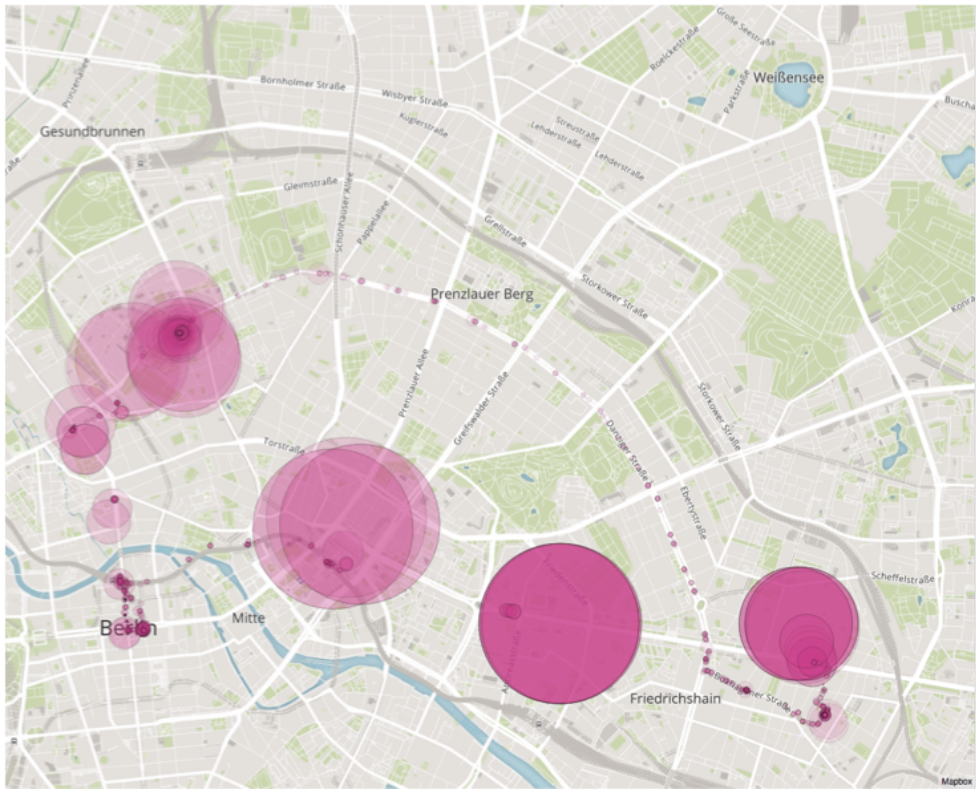




Process of data collection and processing:
1. Efficient data collection



Process of data collection and processing:
2. Location filtering und Interpolation

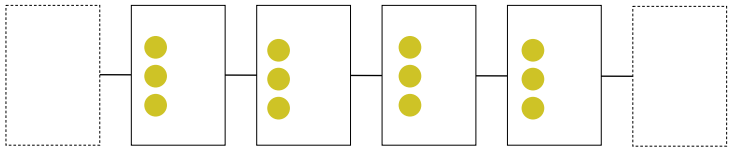


Filtern von Ausreißern

- Double filter:
  Identische Waypoints entfernen
- Accuracy filter:
  Waypoints auf Basis der Genauigkeit auswählen
- Downsampling filter:
  Anzahl der Waypoints verringern
- Local outlier filter:
  benachbarte Waypoints analysieren und auswählen

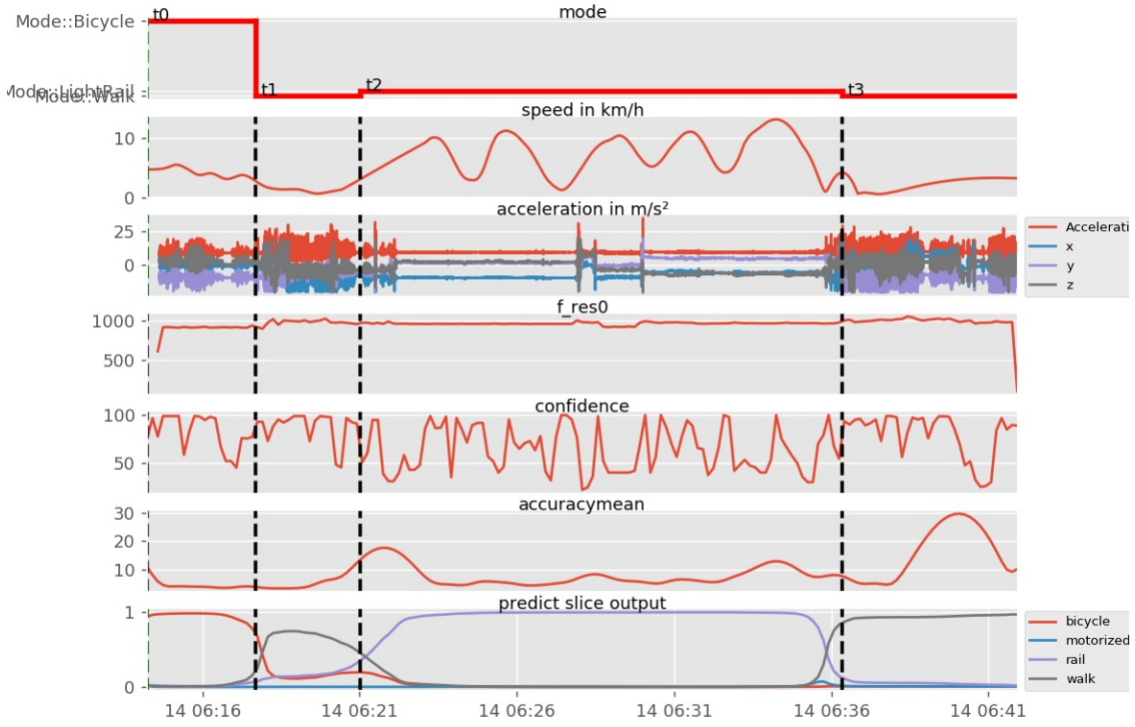
Interpolation

Process of data collection and processing:
3. Machine learning: Time based RNN



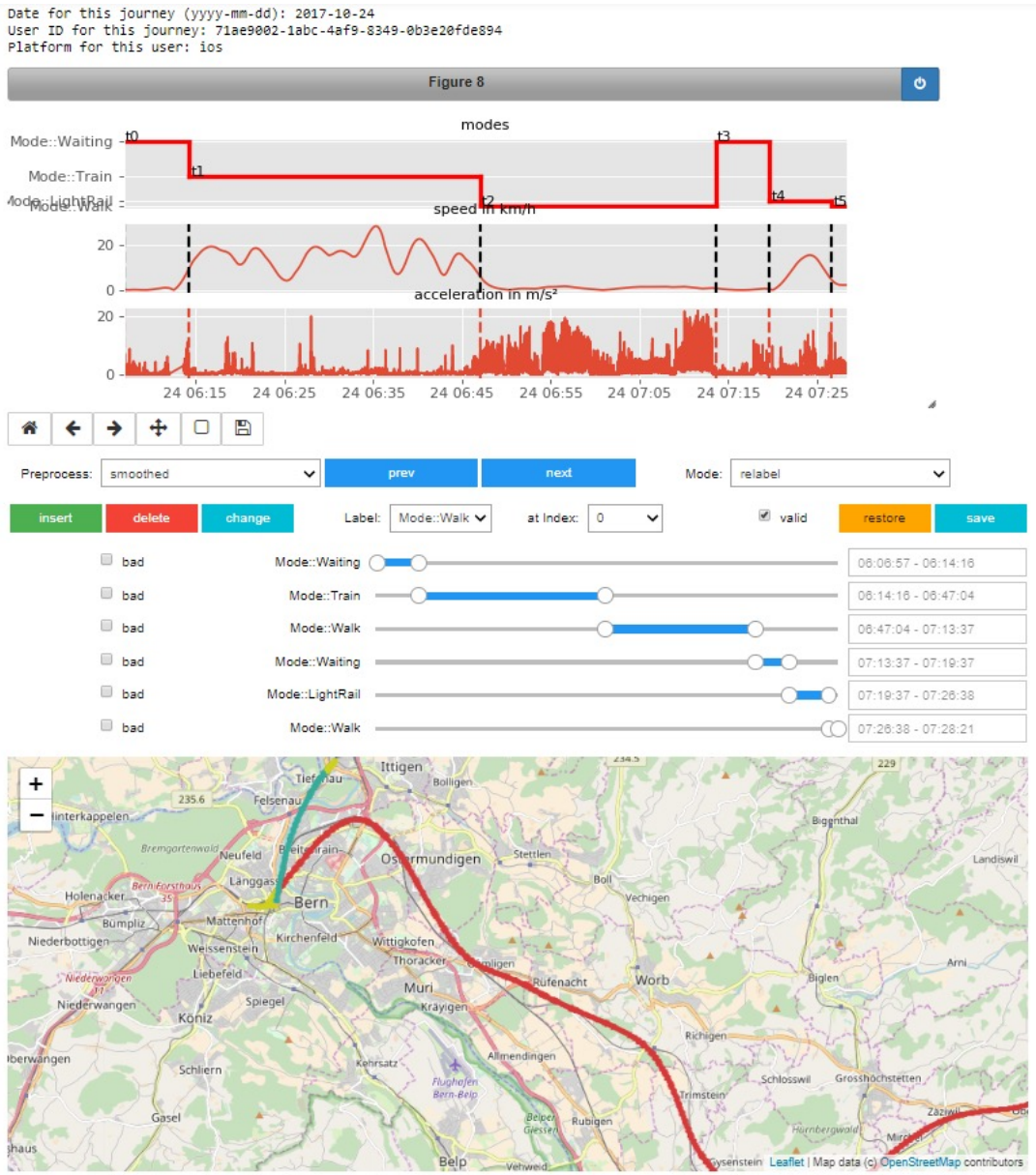
Data segments in a continuous time interval with approx. 80 features for each time segment

Statistical features
Time bases features
Frequency features



Process of data collection:
3. Machine learning: relabeling

High quality training data is generated using a self-developed relabeling tool.



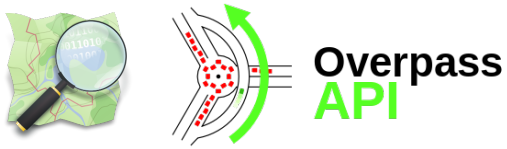
Process of data collection:
4. Fine classification, integration of external data sources and map matching

Timebased clustering
- Identification of stops
- Cluster size dependent on detected modes

Grouping and merging
- Linking segments with the same mode of transport

Fine classification
- Based on features like:
  - Average speed
  - Density of points
  - Length

Map matching
- OSM-Data for determination of stops and routes via Overpass API
- Mapmatching via JGraphT

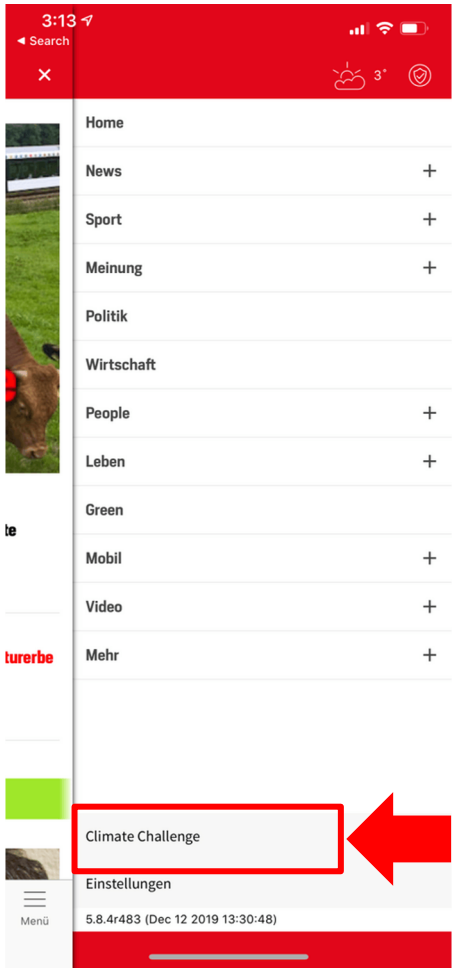


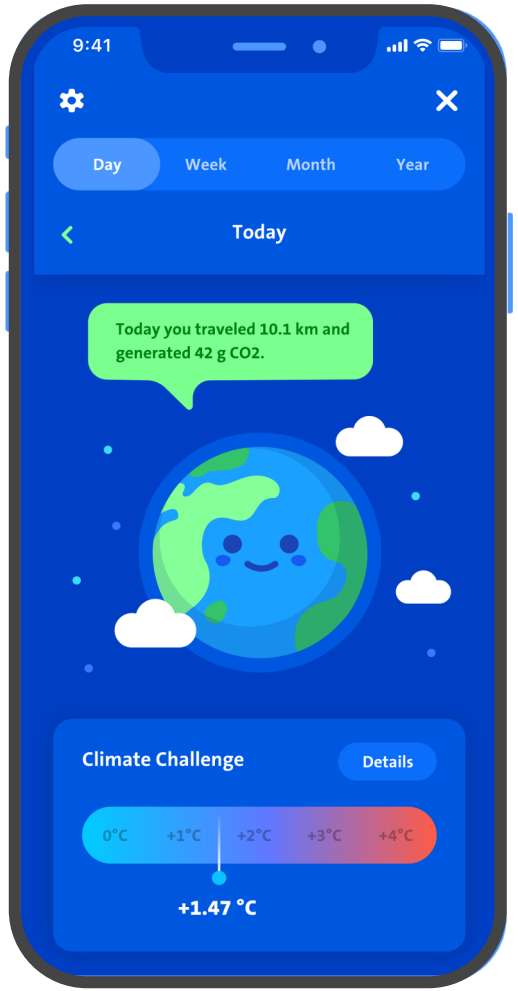
Stay classification
- Stops get classified based on routines

How to engage the user to share movement data

Example: Swiss Climate Challenge

www.swissclimatechallenge.ch





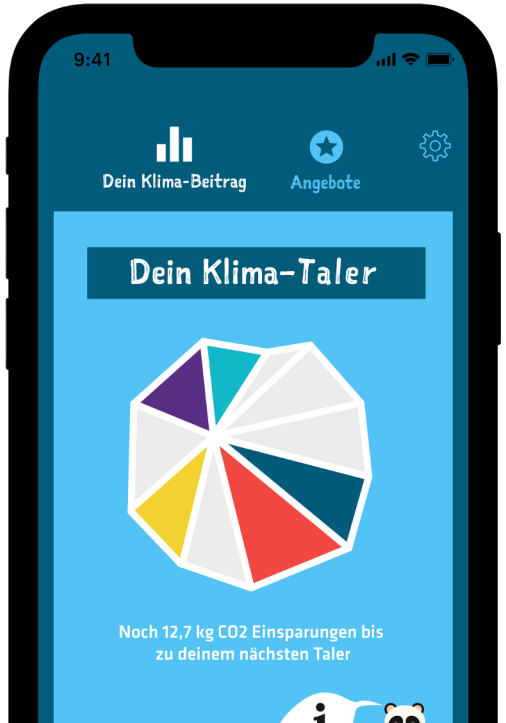
How to engage the user to share movement data

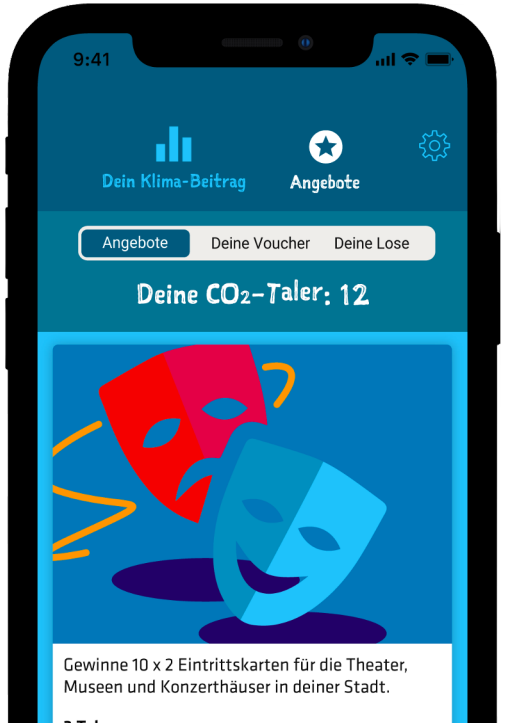
Example: Muenster bewegt - Reward program for sustainable behavior

Users download the App Münster bewegt.

The app measures all distances travelled by modes of transport, calculates the resulting CO2 emissions and rewards journeys made by public transport, bicycle and on foot with Climate Coins.

Climate Coins can be traded in for attractive local offers, lotteries and vouchers.

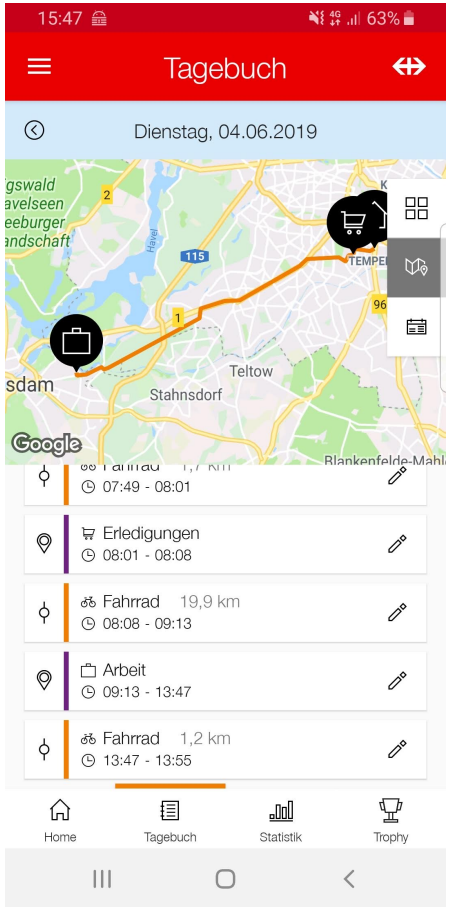




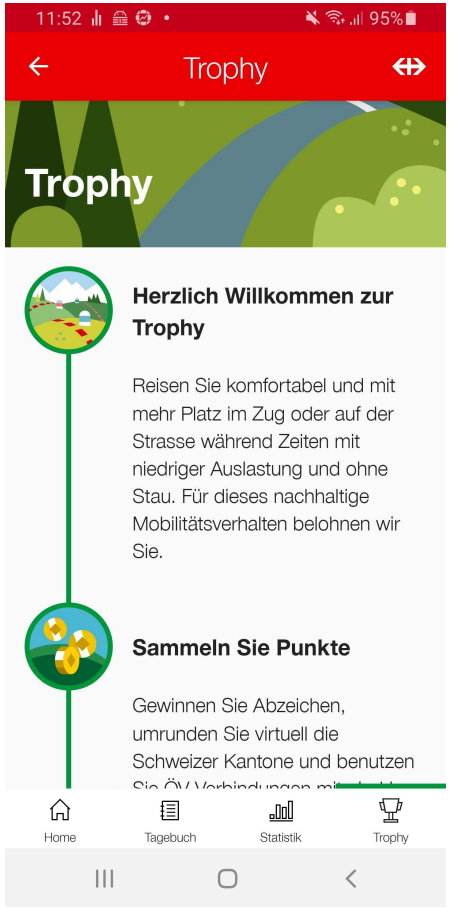
How to engage the user to share movement data

SBB MyWay: Data-based promotion of behavioral change

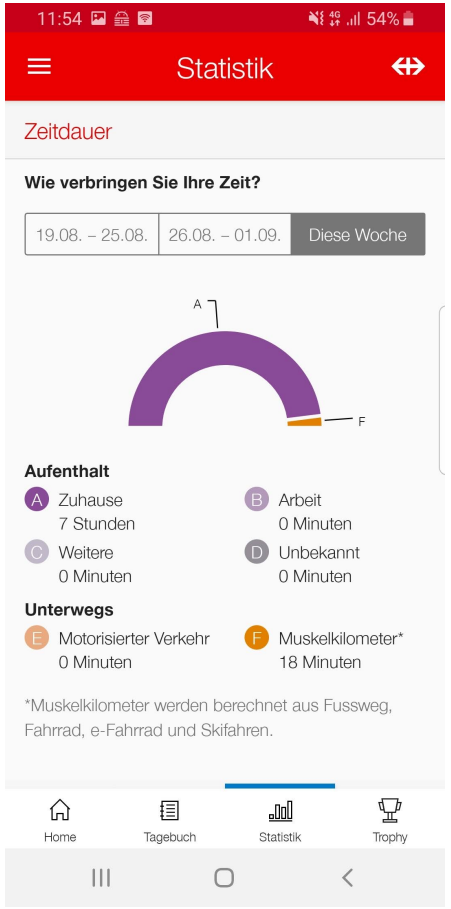
SBB CFF FFS



Mobility diary



Challenges to optimise individual mobility

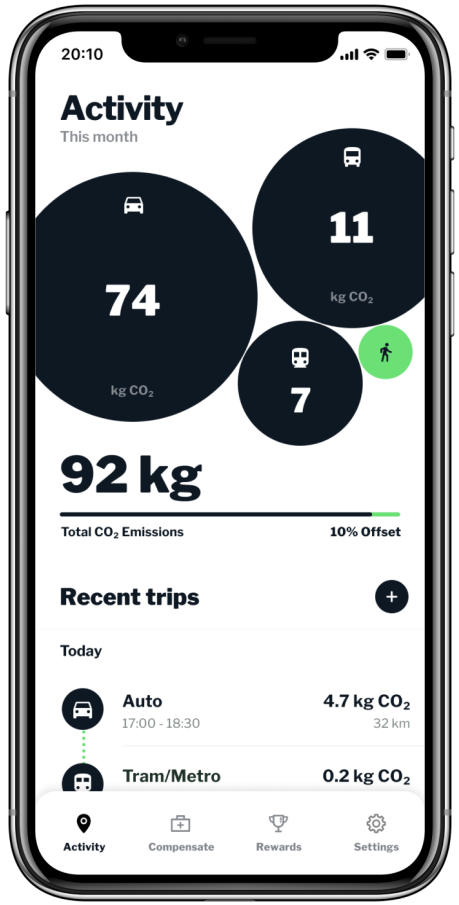


Motivate by individual statistics

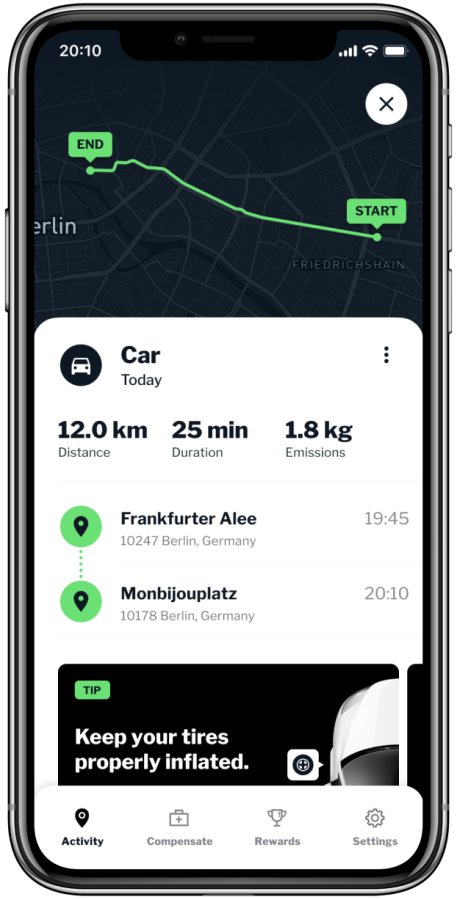
MyWay records which means of transport users use to get around. The app thus shows how and when users can best travel, avoid rush hours and make their mobility more efficient and ecological overall.

How to engage the user to share movement data

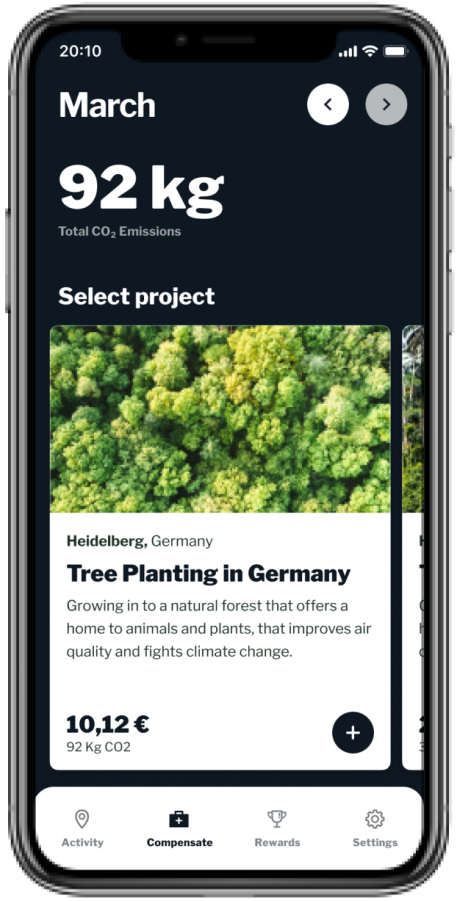
Movefriendly by AUDI: Promote sustainable mobility



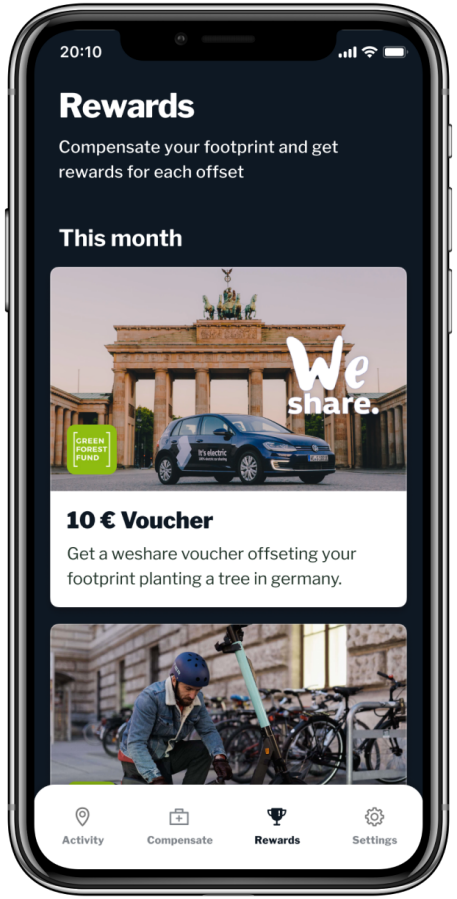
Mobility diary



Information about possibilities to save CO2



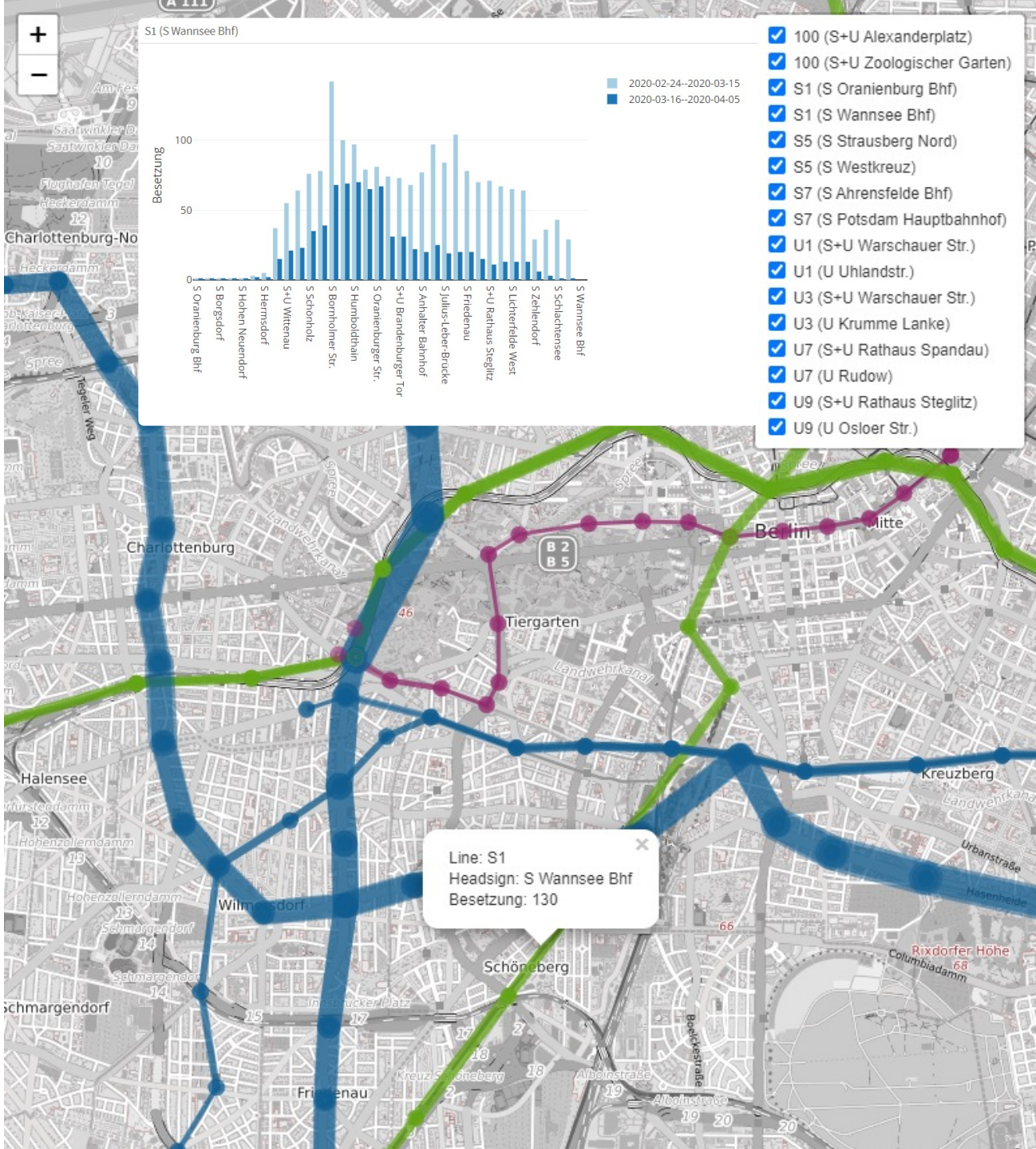
CO2-Offsetting in the App

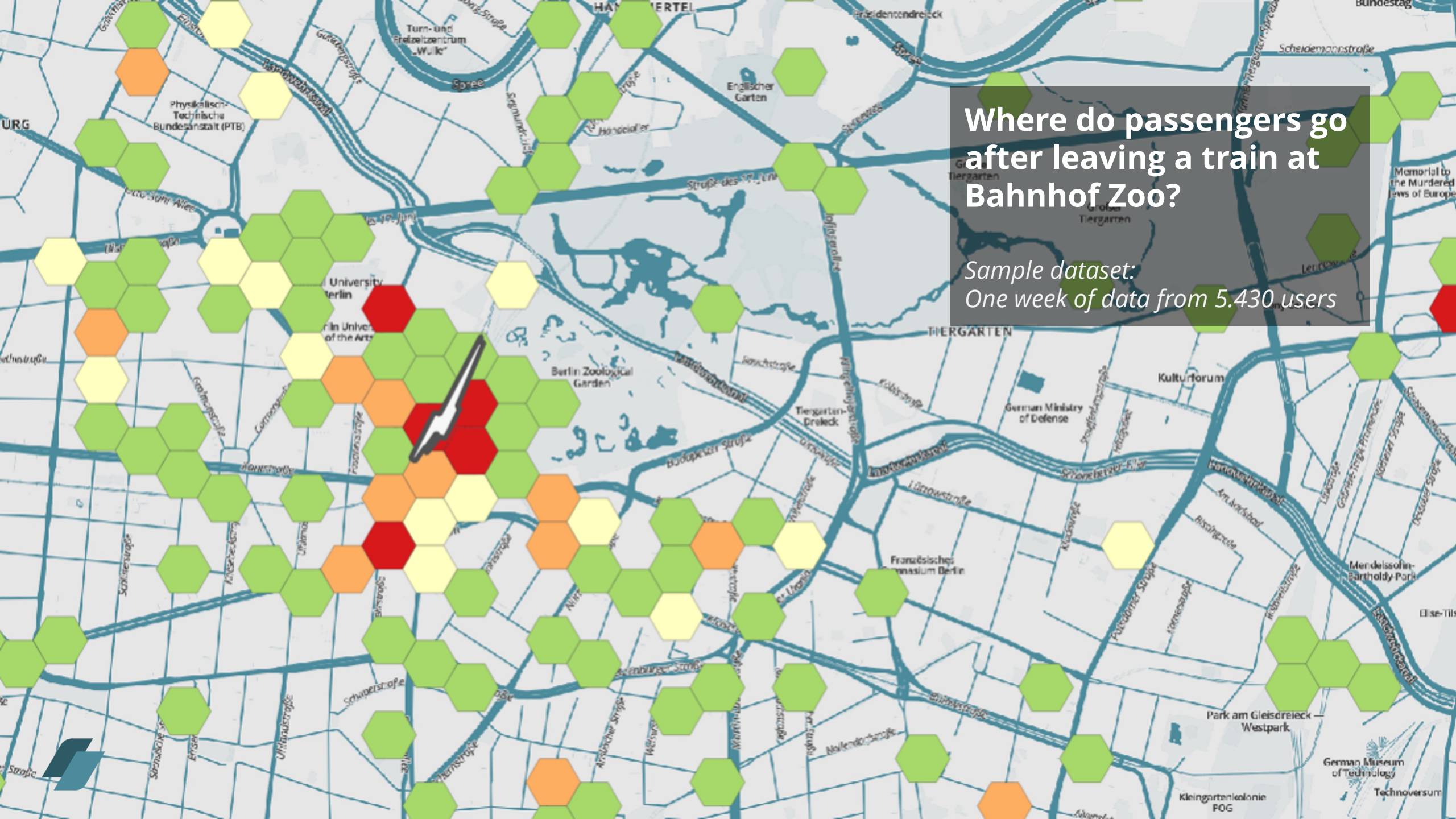


Support of offsetting through rewards

Station-based statistics

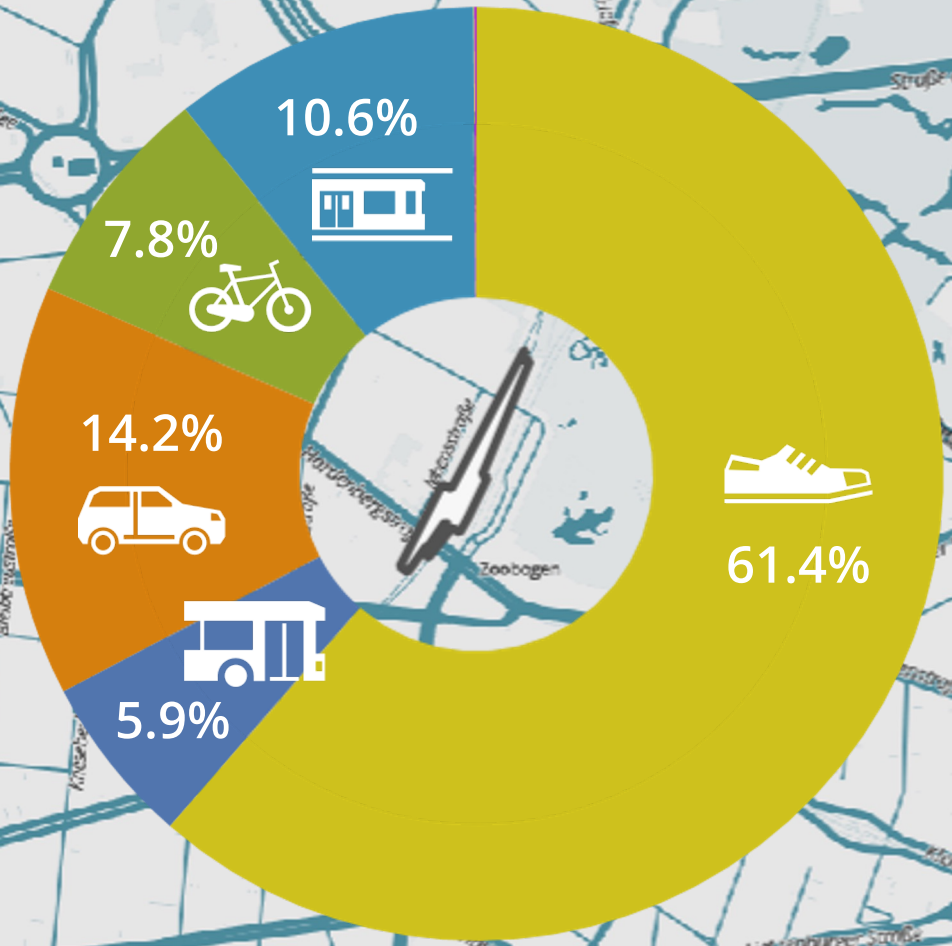
Spatial and line-based statistics



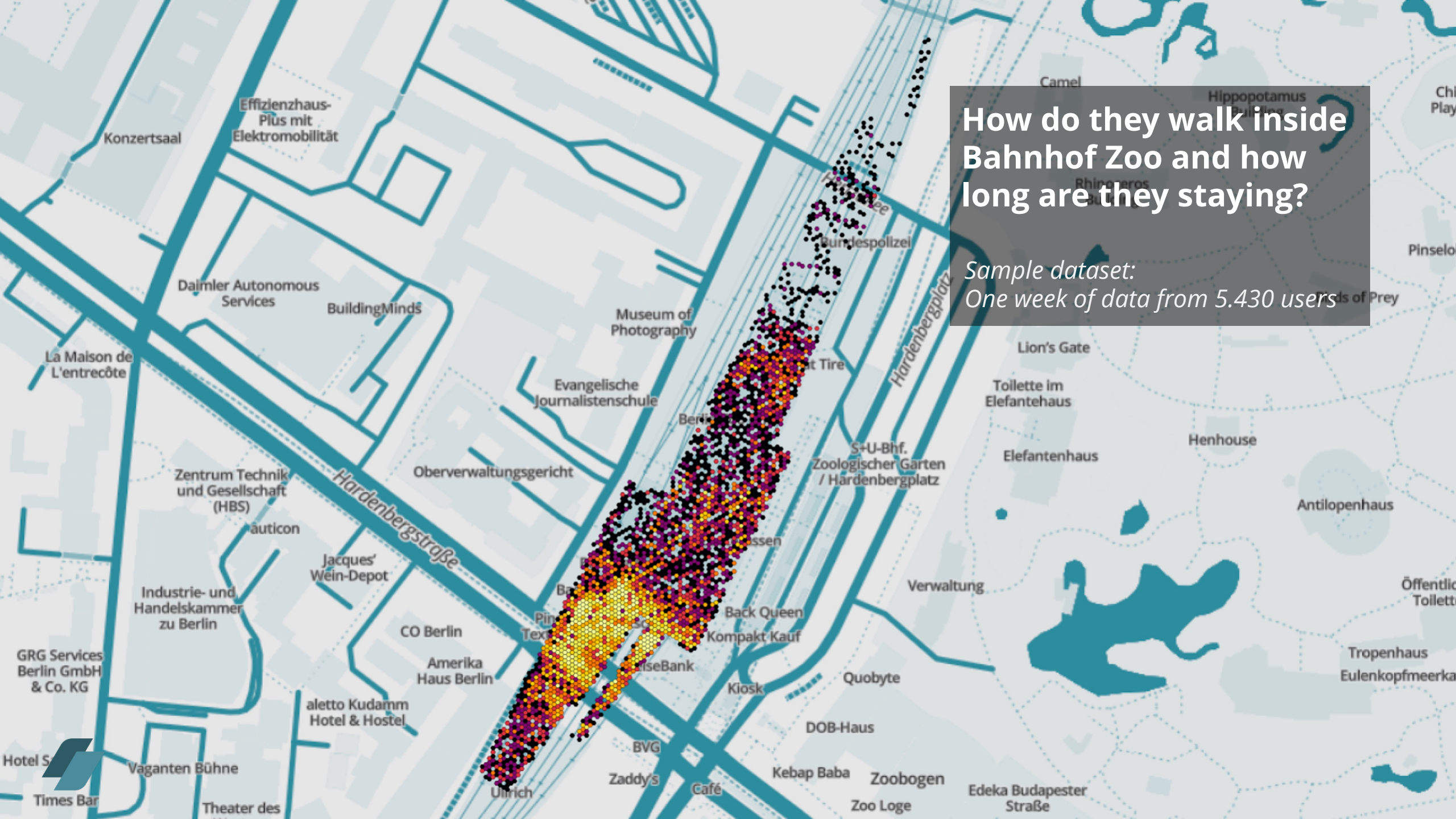

Where do passengers go after leaving a train at Bahnhof Zoo?
Sample dataset:
One week of data from 5.430 users
Spree
TIERGARTEN
Berlin Zoological Garden
Kulturforum
German Ministry of Defense
Tiergarten-Dreieck
Englischer Garten
Physikalisch-Technische Bundesanstalt (PTB)
Turn- und Freizeitzentrum „Wuhle"
Park am Gleisdreieck — Westpark
Mendelssohn-Bartholdy-Park
German Museum of Technology
Technoversum
Kleingartenkolonie POG

What mode do they use after leaving a train at Bahnhof Zoo?

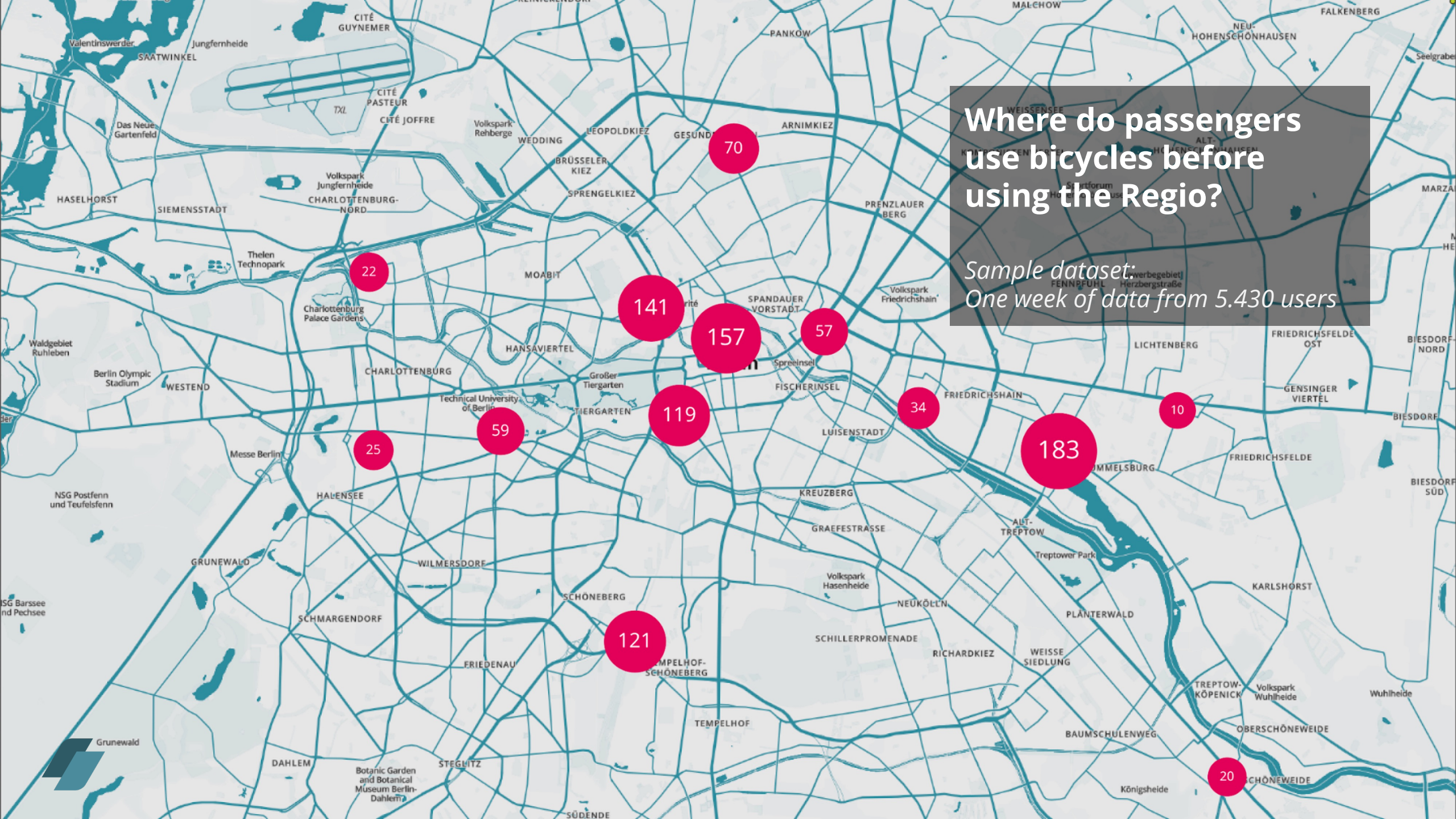
Sample dataset:
One week of data from 5.430 users




Where do they work, shop, live and eat?
Sample dataset:
One week of data from 5.430 users
Work
Shop
Home
Gastronomy
Technical University of Berlin
Berlin University of the Arts
Berlin Zoological Garden
TIERGARTEN
Großer Tiergarten
Englischer Garten
Kulturforum
German Ministry of Defense
Tiergarten-Dreieck
Deutscher Bundestag
Sony Center
Spree
Straße des 17. Juni
Landwehrkanal
Park am Gleisdreieck
Westpark
Mendelssohn-Bartholdy-Park
German Museum of Technology
Physikalisch-Technische Bundesanstalt (PTB)
Turn- und Freizeitzentrum „Wuhle"


How do they walk inside Bahnhof Zoo and how long are they staying?
Sample dataset:
One week of data from 5.430 users
Hardenbergstraße
Hardenbergplatz
Konzertsaal
Effizienzhaus-Plus mit Elektromobilität
Daimler Autonomous Services
BuildingMinds
La Maison de L'entrecôte
Museum of Photography
Evangelische Journalistenschule
Oberverwaltungsgericht
Zentrum Technik und Gesellschaft (HBS)
Jacques' Wein-Depot
Industrie- und Handelskammer zu Berlin
CO Berlin
Amerika Haus Berlin
GRG Services Berlin GmbH & Co. KG
aletto Kudamm Hotel & Hostel
Vaganten Bühne
Times Bar
Theater des
Bundespolizei
S+U-Bhf. Zoologischer Garten / Hardenbergplatz
Lion's Gate
Toilette im Elefantenhaus
Elefantenhaus
Henhouse
Antilopenhaus
Verwaltung
Back Queen
Kompakt Kauf
Kiosk
Quobyte
DOB-Haus
BVG
Zaddy's
Café
Kebap Baba
Zoobogen
Zoo Loge
Edeka Budapester Straße
Tropenhaus
Camel
Hippopotamus
Rhinoceros
Pinselo
Ch
Play

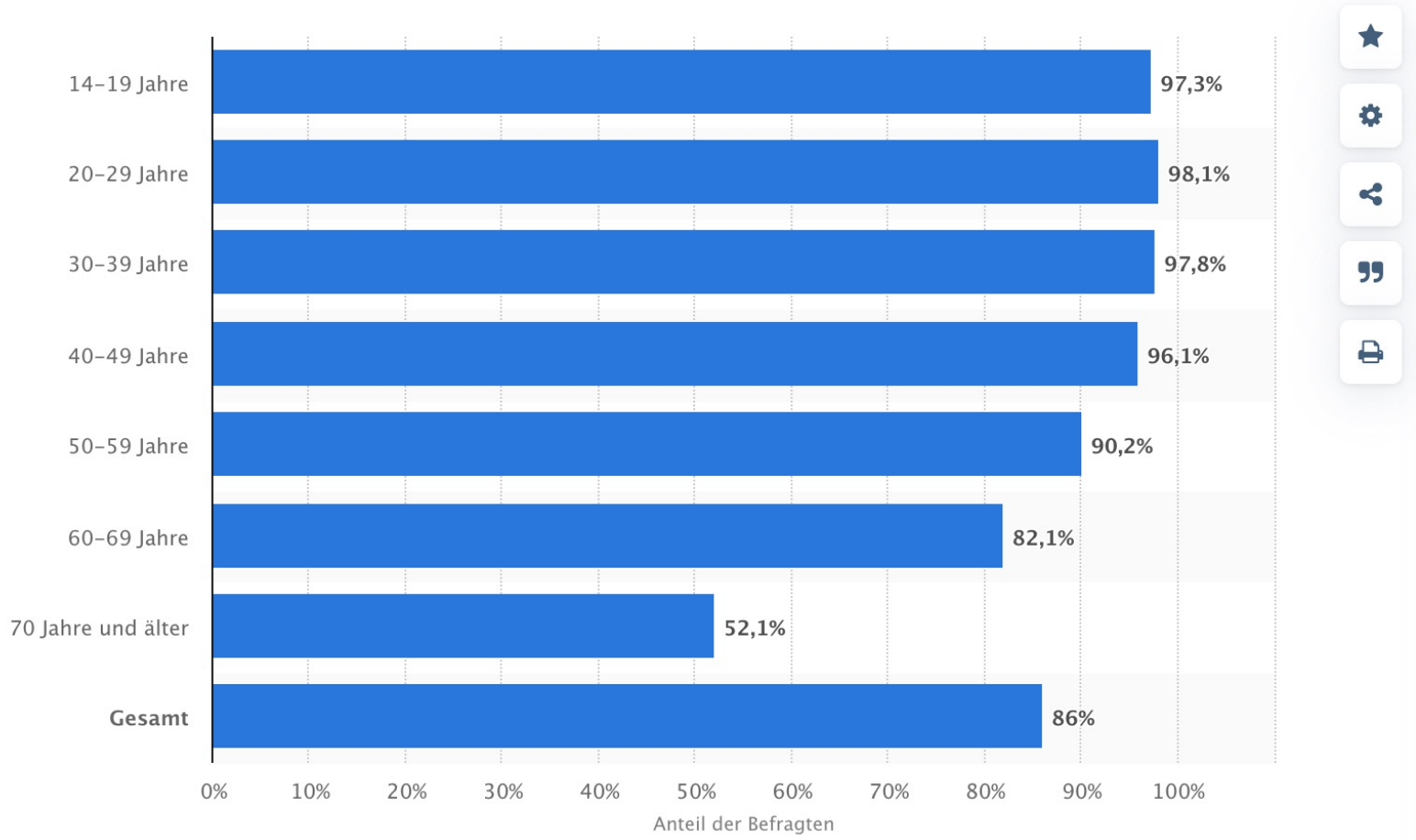

Where do passengers park their bike before getting on a train?
Sample dataset:
One week of data from 5.430 users
Daimler Autonomous Services
BuildingMinds
Museum of Photography
Helmut Newton Foundation
Evangelische Journalistenschule
Berlin Zoo Bhf.
Fat Tire
Toilette im Elefantenhaus
Henhouse
Elefantenhaus
Industrie- und Handelskammer
Zentrum Technik und Gesellschaft (HBS)
Oberverwaltungsgericht
Hardenbergstraße
Jebensstraße
OnNoYa Ishin
auticon
Prediger Lichtberater
Jacques' Wein-Depot
DB Reisezentrum
Zooterassen
Backwerk
Ditsch
Presse Buch
DB Information
ReiseBank
CO Berlin
Amerika Haus Berlin
The Flint
aletto Kudamm Hotel & Hostel
Verwaltung
Quobyte
Risa Chicken
Hekticket
DOB-Haus
Kebap Baba
Zoobogen
Edeka Budapester Straße
Zoo Loge
Jim Block Restaurant
ZooBa
Super
Bikini Berlin
Spirit Coffee
North Face
BVG
Zaddy's
Delphi LUX
Ullrich
Café
Berlin
Starbucks
Juwelier
Theater des Westens


Where do passengers use bicycles before using the Regio?
Sample dataset:
One week of data from 5.430 users
70
22
141
157
57
119
59
34
10
25
183
121
CITÉ GUYNEMER
Valentinswerder
Jungfernheide
SAATWINKEL
CITÉ PASTEUR
CITÉ JOFFRE
TXL
Das Neue Gartenfeld
Volkspark Rehberge
WEDDING
LEOPOLDKIEZ
BRÜSSELER KIEZ
SPRENGELKIEZ
ARNIMKIEZ
PANKOW
MALCHOW
NEU-HOHENSCHÖNHAUSEN
FALKENBERG
Seelgraben
WEISSENSEE
HASELHORST
SIEMENSSTADT
Volkspark Jungfernheide
CHARLOTTENBURG-NORD
PRENZLAUER BERG
Thelen Technopark
MOABIT
Charlottenburg Palace Gardens
SPANDAUER VORSTADT
Volkspark Friedrichshain
FENNPFUHL
Herzbergstraße
MARZAHN
Waldgebiet Ruhleben
HANSAVIERTEL
LICHTENBERG
FRIEDRICHSFELDE OST
BIESDORF-NORD
Berlin Olympic Stadium
WESTEND
CHARLOTTENBURG
Großer Tiergarten
Spreeinsel
FISCHERINSEL
FRIEDRICHSHAIN
GENSINGER VIERTEL
BIESDORF
Technical University of Berlin
TIERGARTEN
LUISENSTADT
Messe Berlin
FRIEDRICHSFELDE
NSG Postfenn und Teufelsfenn
HALENSEE
KREUZBERG
GRAEFESTRASSE
ALT-TREPTOW
Treptower Park
GRUNEWALD
WILMERSDORF
Volkspark Hasenheide
KARLSHORST
SCHÖNEBERG
NEUKÖLLN
PLÄNTERWALD
SCHMARGENDORF
SCHILLERPROMENADE
RICHARDKIEZ
WEISSE SIEDLUNG
FRIEDENAU
TEMPELHOF-SCHÖNEBERG
TREPTOW-KÖPENICK
Volkspark Wuhlheide
Wuhlheide
TEMPELHOF
BAUMSCHULENWEG
OBERSCHÖNEWEIDE
Grunewald
DAHLEM
STEGLITZ
Botanic Garden and Botanical Museum Berlin-Dahlem
Königsheide
SÜDENDE

[11]

Big data: Verheißungen und Tücken

Anteil der Smartphone-Nutzer in Deutschland nach Altersgruppe im Jahr 2020



Details: Deutschland; IFAK; GfK Media and Communication Research; forsa marplan; GIM; 19.082 Befragte; ab 14 Jahre; deutschsprachige Bevölkerung; Besitzer eines internetfähigen Smartphones oder Handys

Promises and pitfals of big data

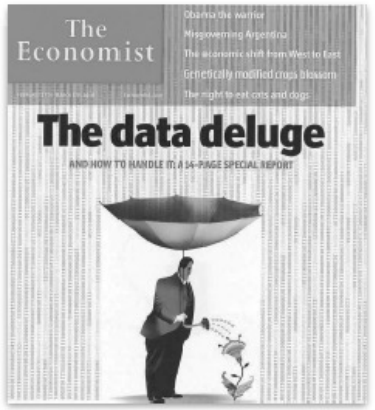
The (second) rise of big data

2008

2011

2012

2018



2010

2011

Promises and pitfalls of big data

What is Big Data?

Big data definition in traditional computer science:

Data that will not fit in main memory.

1 byte

Kilobyte
~1000 (10^3) bytes

Megabyte
~1000000 (10^6) bytes

Gigabyte
~1000000000 (10^9) bytes

25 gigabytes: data analysed per hour by Ford's Ford Fusion Energi plug-in hybrid

60 gigabytes: data gathered per hour by Google's self-driving car

140 gigabytes: data gathered per day by Nokia HERE mapping car

Terabyte
~1000000000000 (10^{12}) bytes

30 terabytes: data produced by Boeing 777 on a transatlantic trip

Petabyte
~1000000000000000 (10^{15}) bytes

Several petabytes: traffic data stored by INRIX to produce traffic analysis for e.g. Google traffic

Exabyte
~1000000000000000000 (10^{18}) bytes

Zettabyte
~1000000000000000000000 (10^{21}) bytes

1 zettabyte: Total amount of visual information conveyed from the eyes to the brains of all humans per day in 2013.

4.4 zettabytes: Estimated size of the digital universe in 2013

44 zettabytes: Projected size of the digital universe in 2020

Yottabyte
~1000000000000000000000000 (10^{24}) bytes

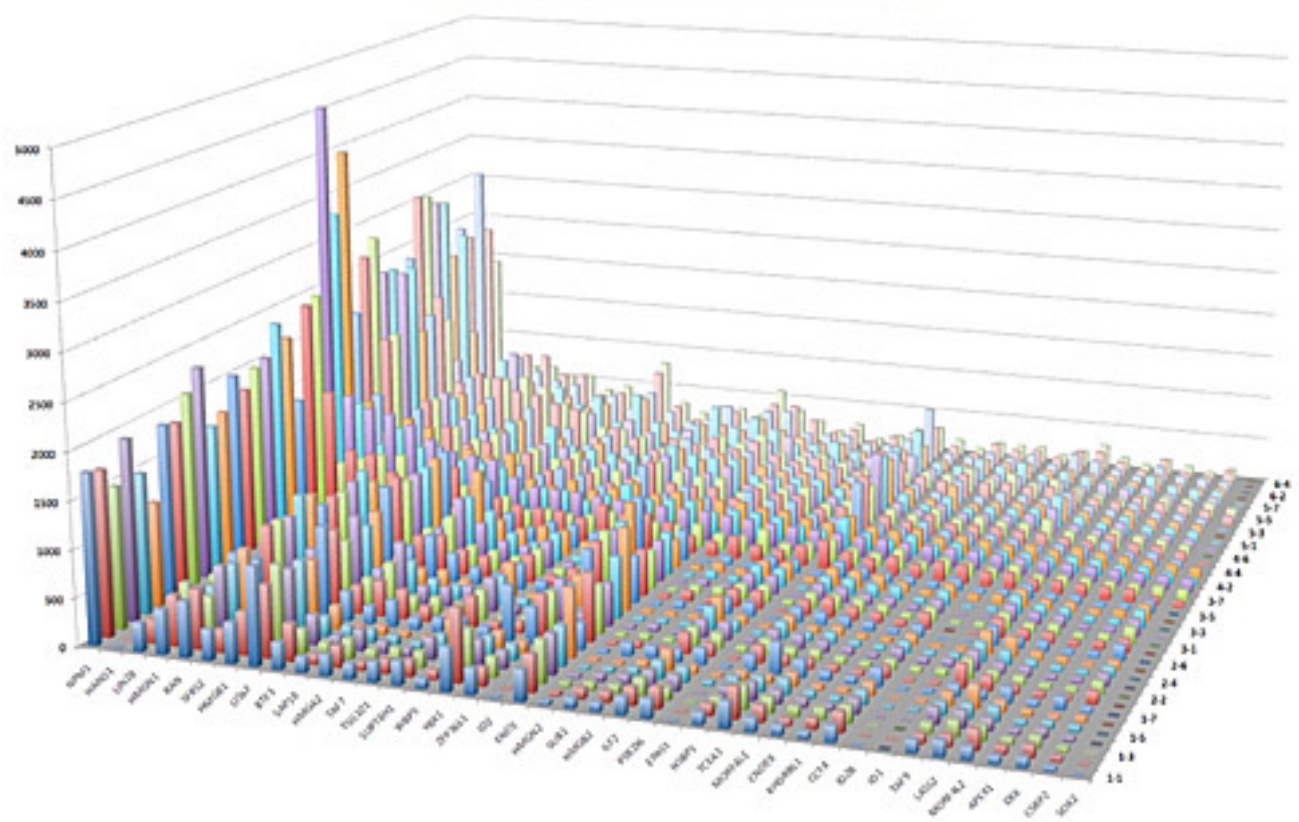
Source: Nokia HERE, Forbes, Idealab, GE, ITF calculations.

Promises and pitfalls of big data

What is Big Data? View of statistics

Big data definition in statistics:

Complex data with a large number of observations and/or features

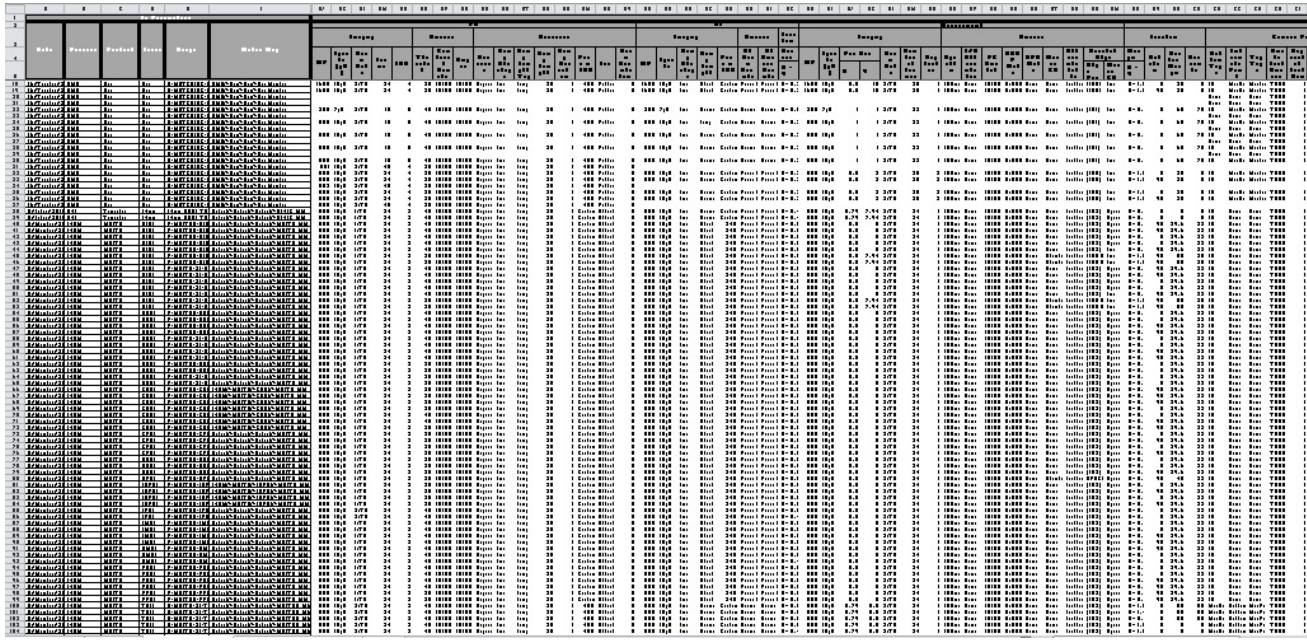


Promises and pitfals of big data

What is Big Data? View of other fields

Big data definition in other fields:

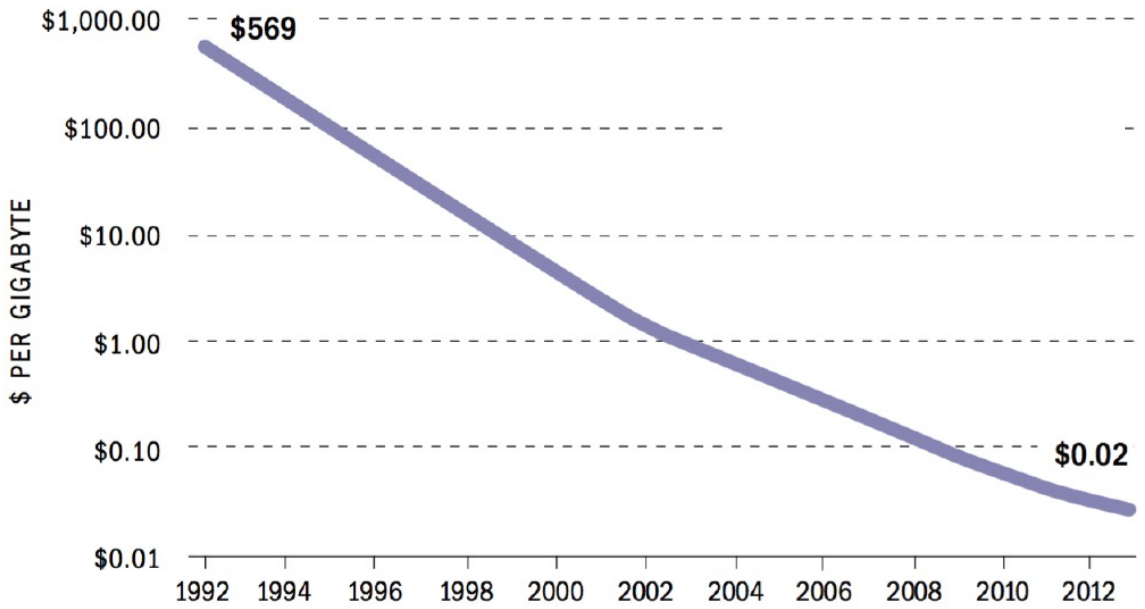
Non traditional sample sizes.
Everything that doesn‘t fit in an excel sheet.



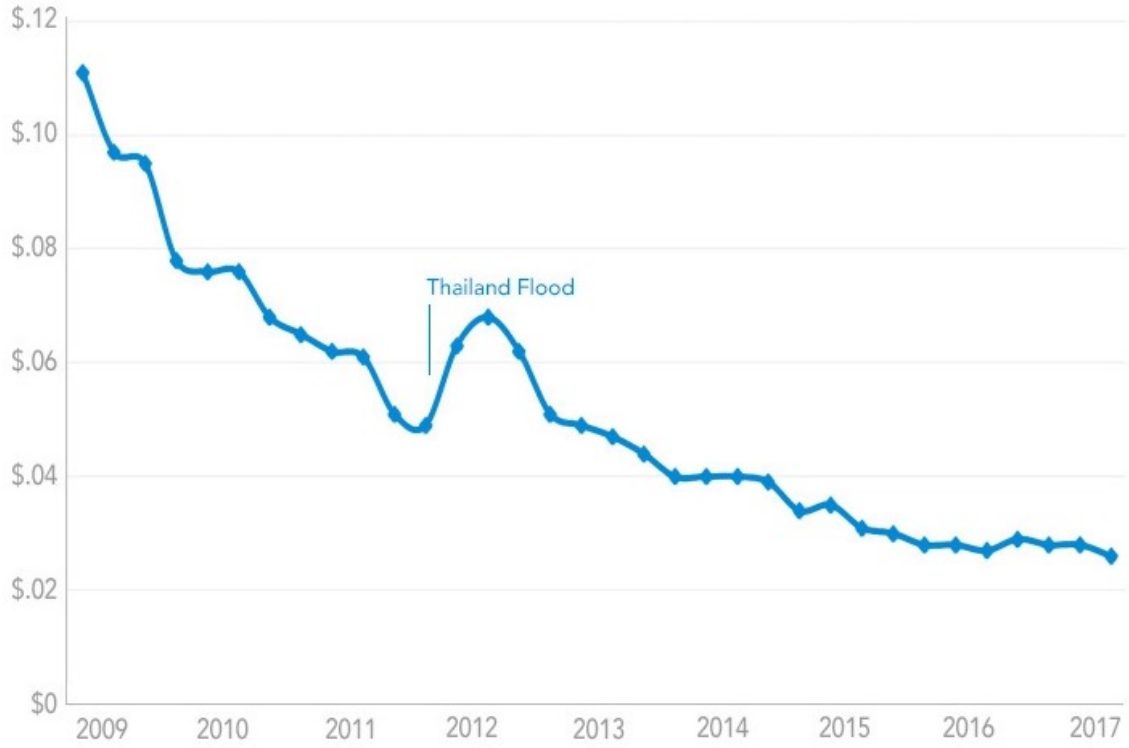
Big data drivers: how did we get there and where does it end?

Sinking cost

Global data storage costs (in US $ per gigabyte of storage)



Source: Deloitte, May 2014, as reported by Internet Trends 2014 by Mary Meeker for Kleiner Perkins Caufield & Byers

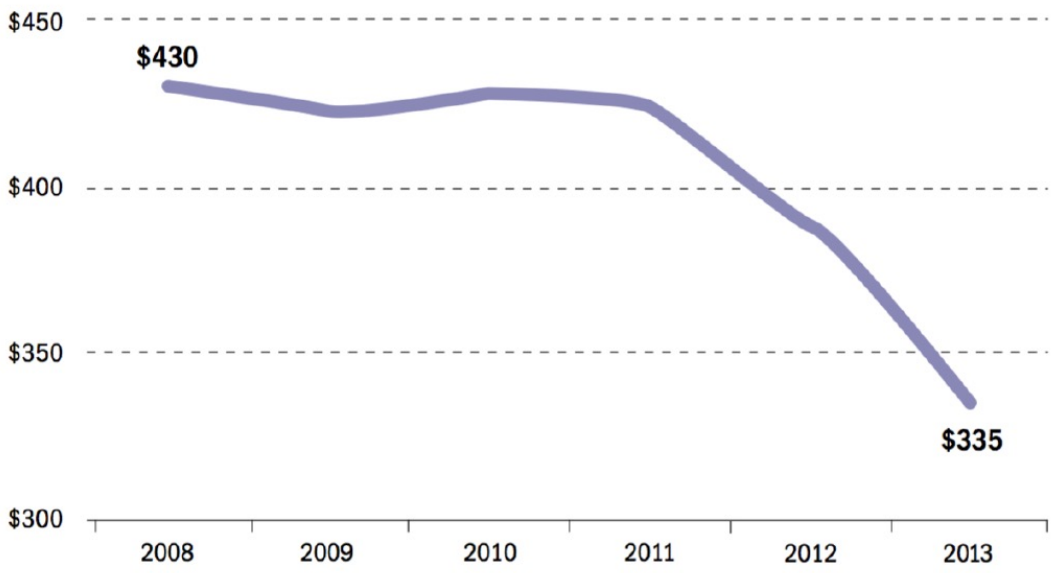


https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/

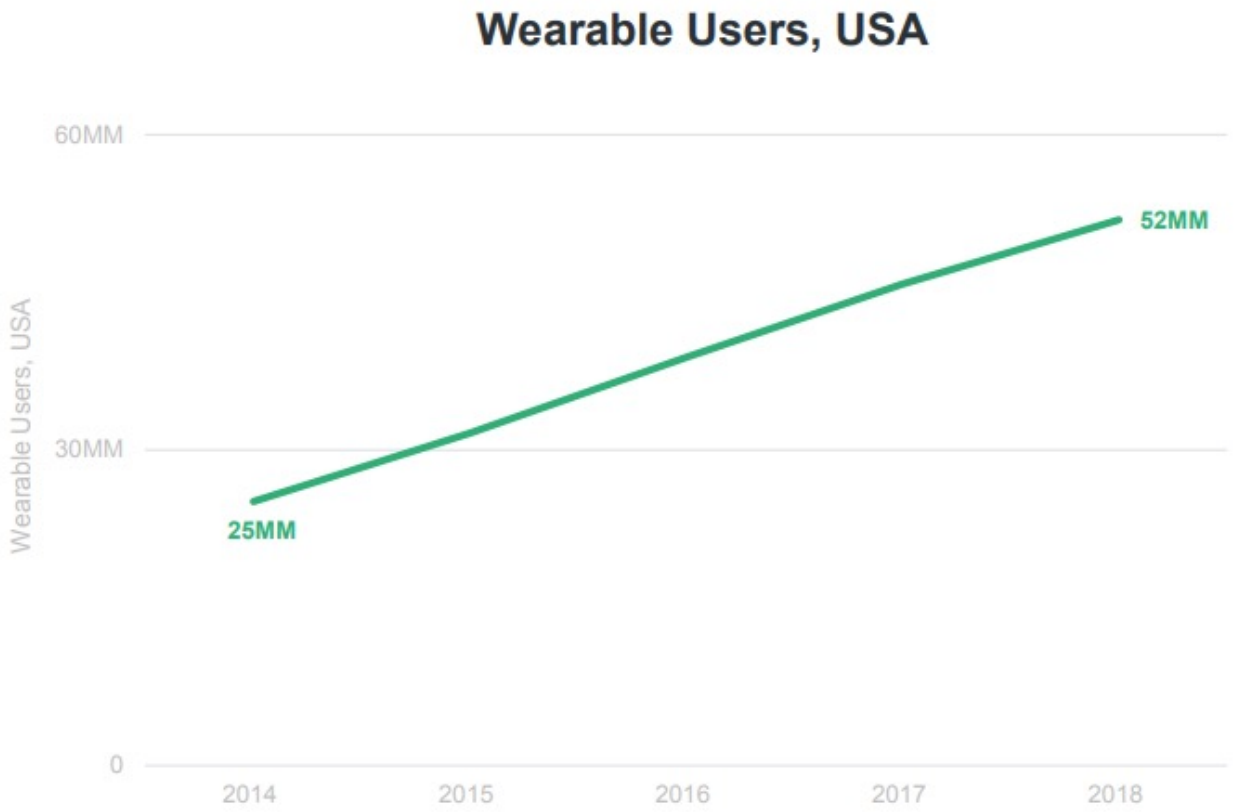
Big data drivers: how did we get there and where does it end?

Sinking cost

Global smartphone costs
(Average price per unit, in US $)



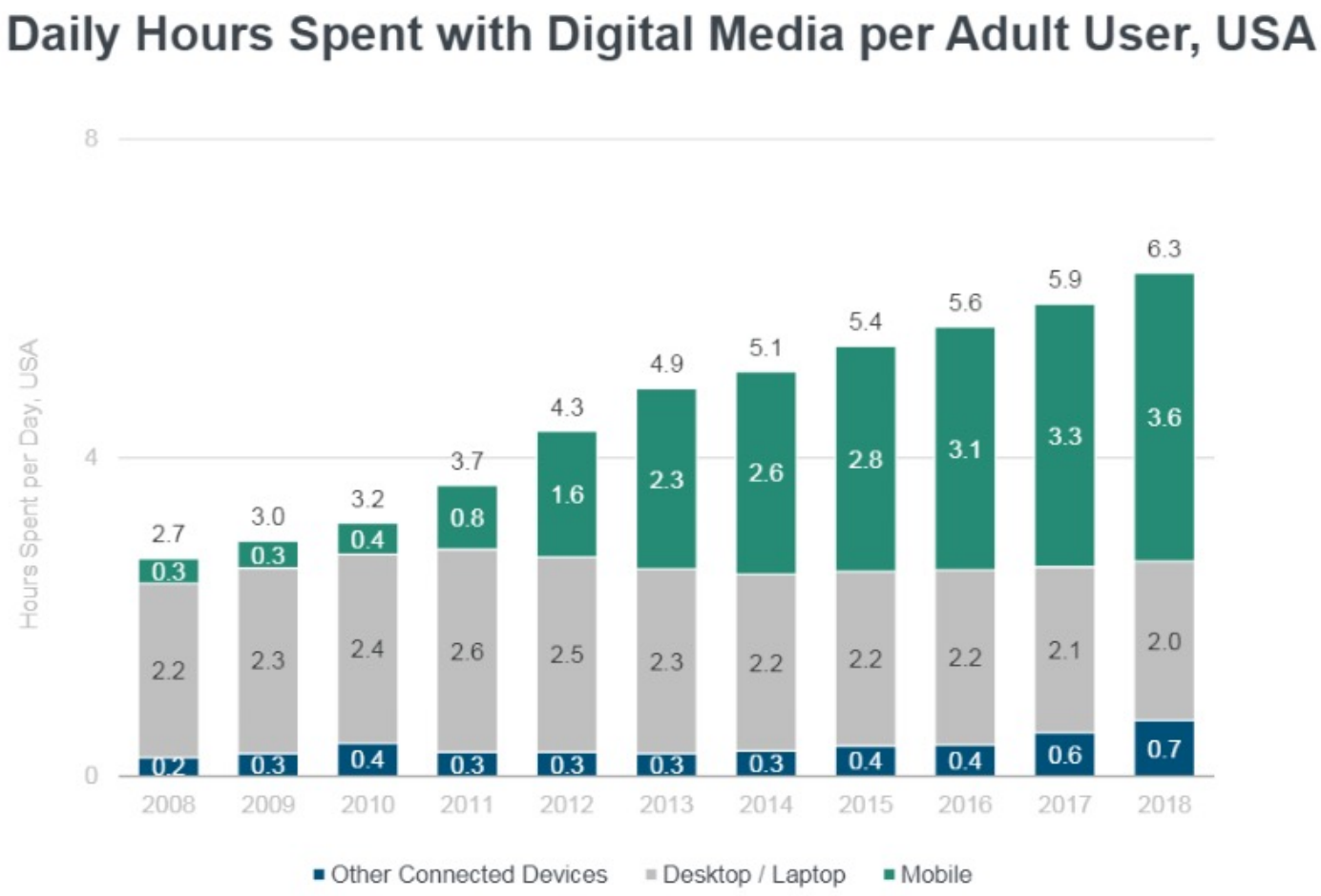
Source: Deloitte, May 2014, as reported by Internet Trends 2014 by Mary Meeker for Kleiner Perkins Caufield & Byers



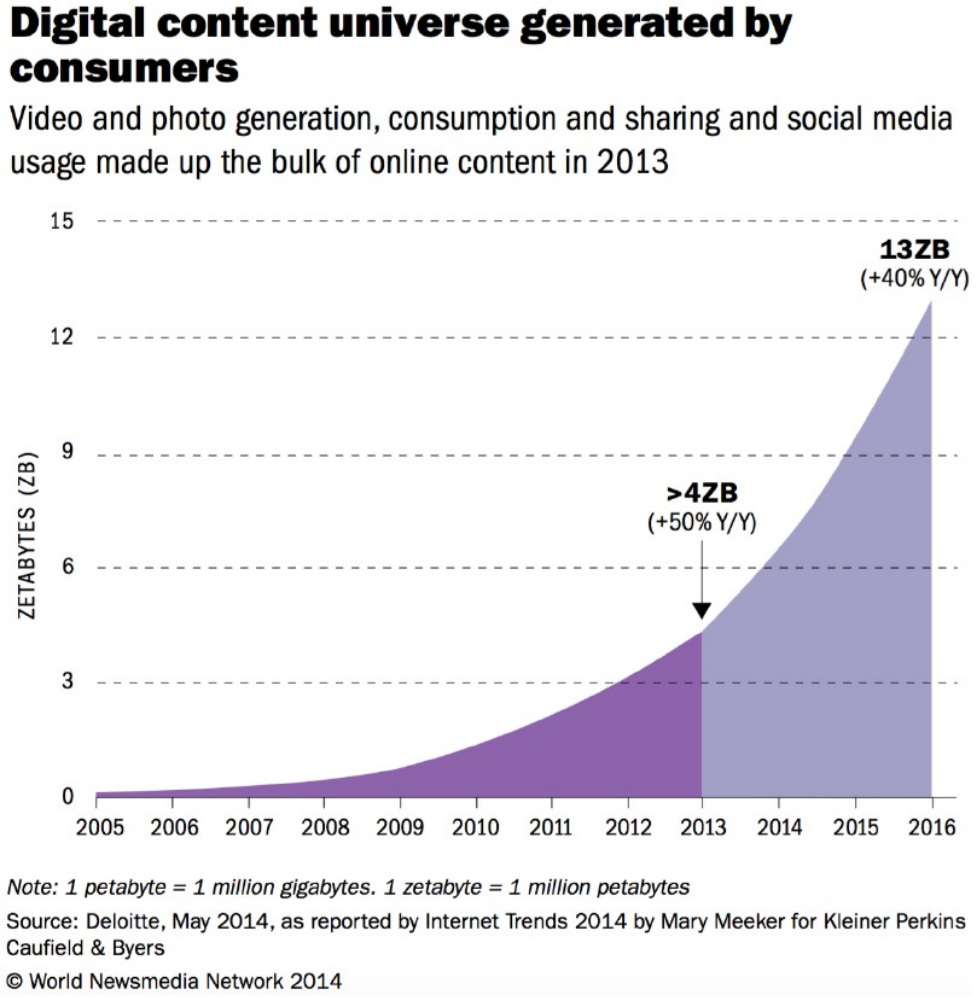
https://www.bondcap.com/report/itr19/#view/52

Big data drivers: how did we get there and where does it end?

Sinking cost + new use cases = more data

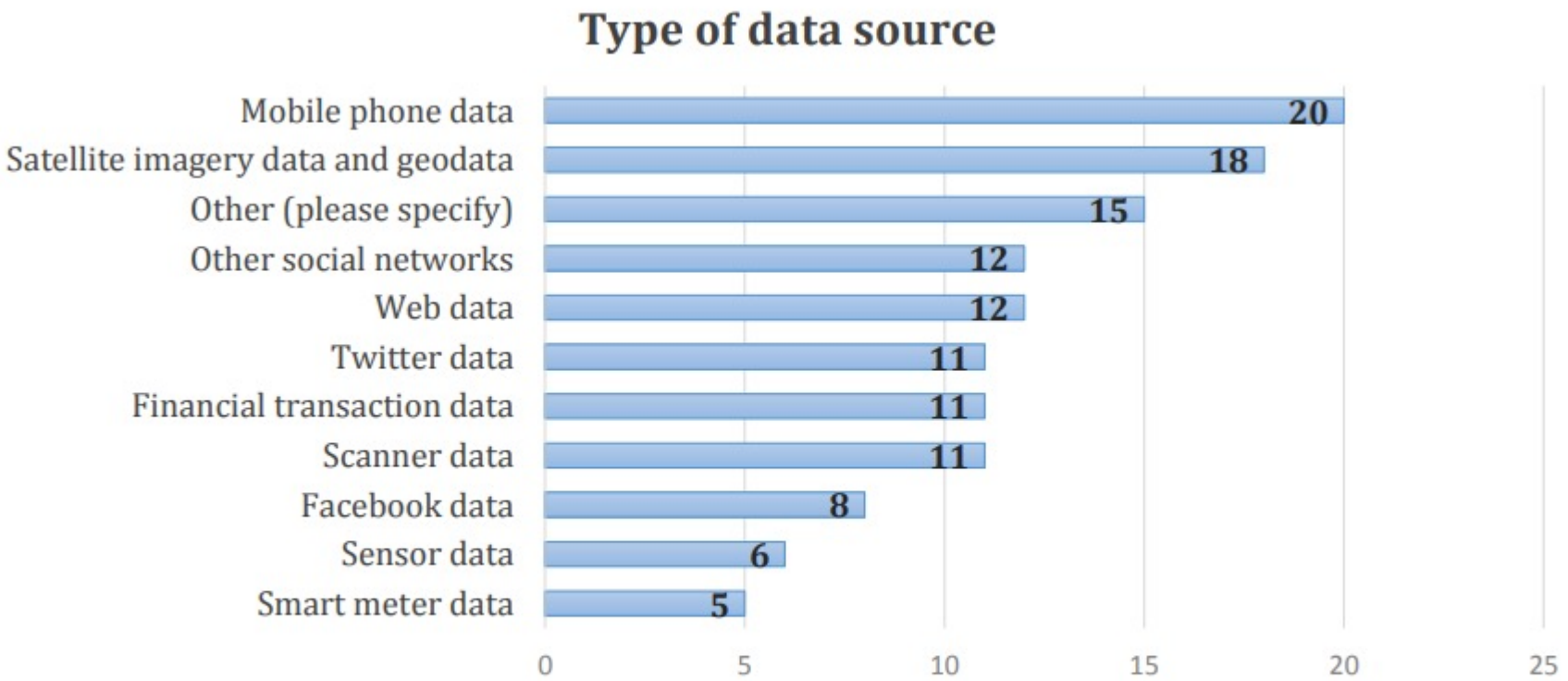


https://www.bondcap.com/report/itr19/#view/41



Promises and pitfalls of big data

SDG related big data projects (Worldbank, 2016)

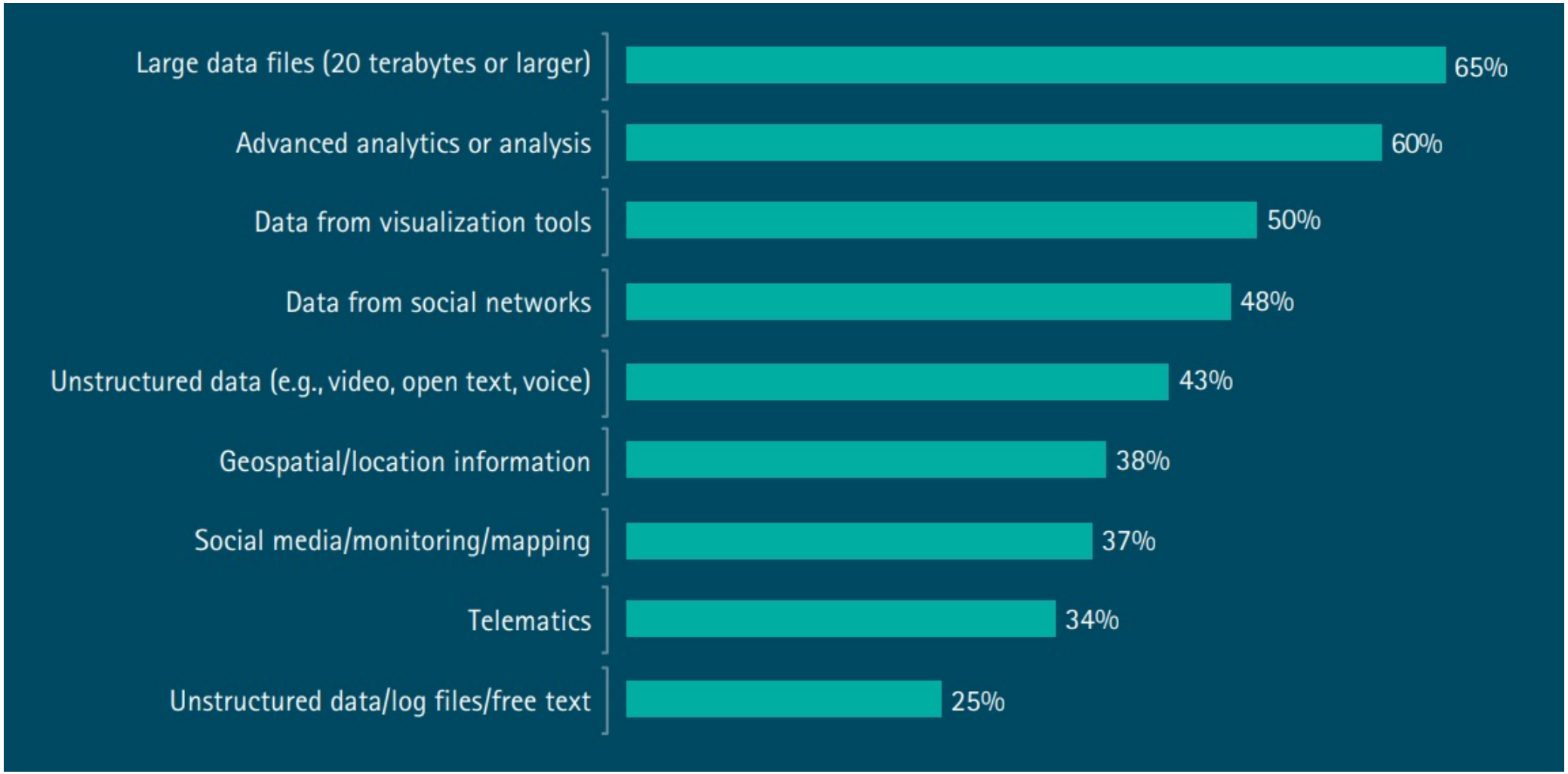


- Mobile phones (20), satellite imagery (18) and social media (11+8) are the most prominent sources

https://unstats.un.org/unsd/trade/events/2015/abudhabi/presentations/day3/02/2b%20A-Using%20Big%20Data%20for%20the%20Sustainable%20Development%20Goals%2010222015.pdf

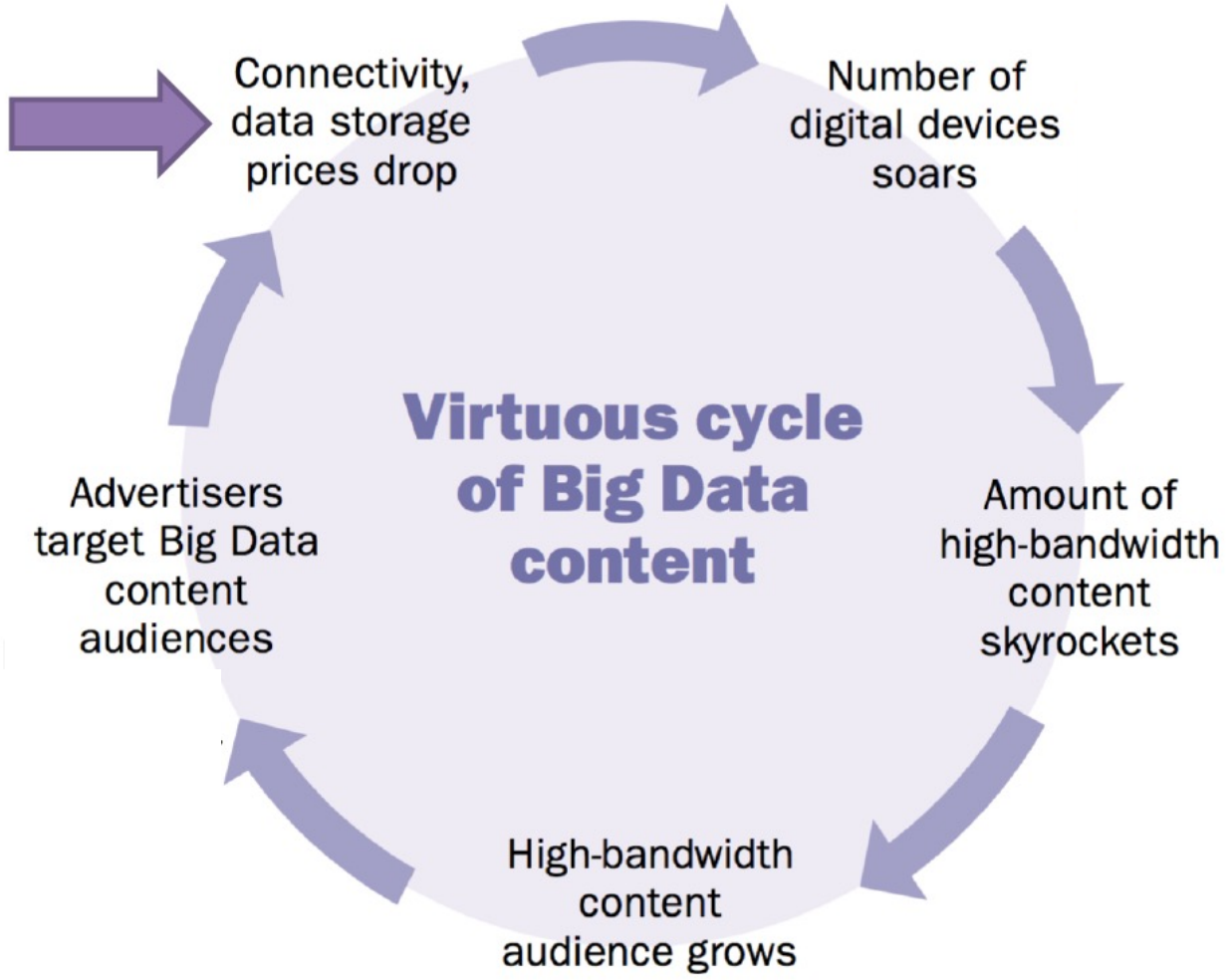
Promises and pitfals of big data

Sources of big data



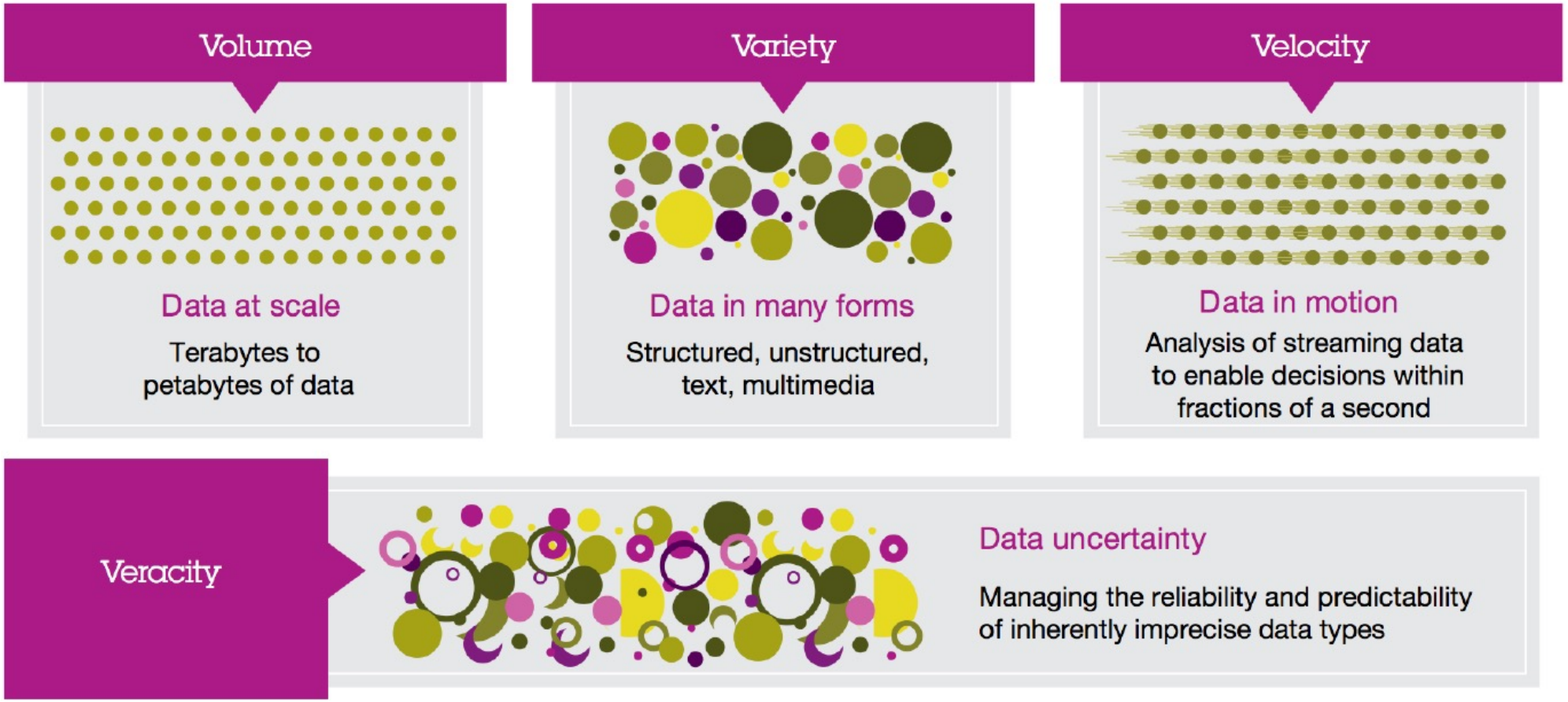
https://www.accenture.com/us-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_14/Accenture-Big-Data-POV.pdf

Promises and pitfals of big data
Virtuous cycle of big data content



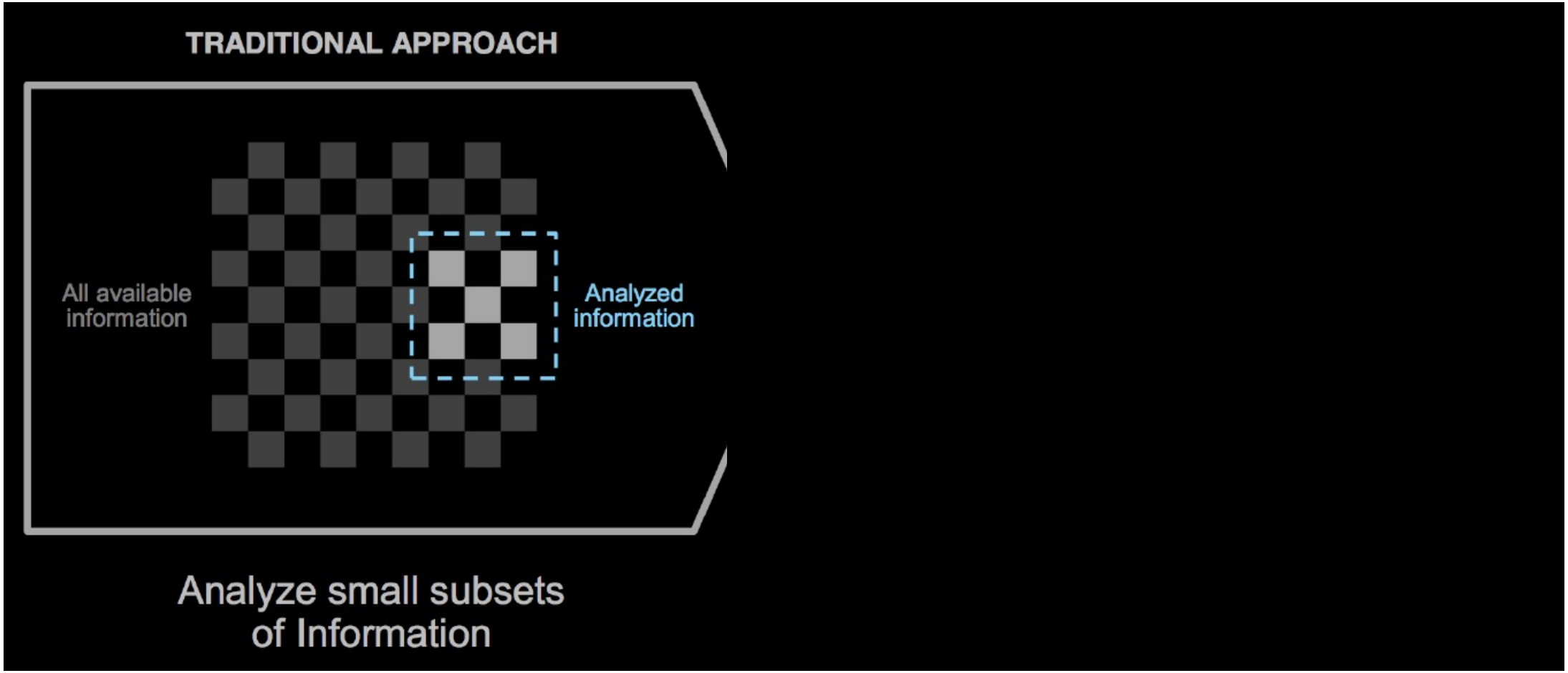
Promises and pitfalls of big data

Big data: a type of analytics



Promises and pitfalls of big data

Promise: leverage more of the data being captured



Andrews, 2014

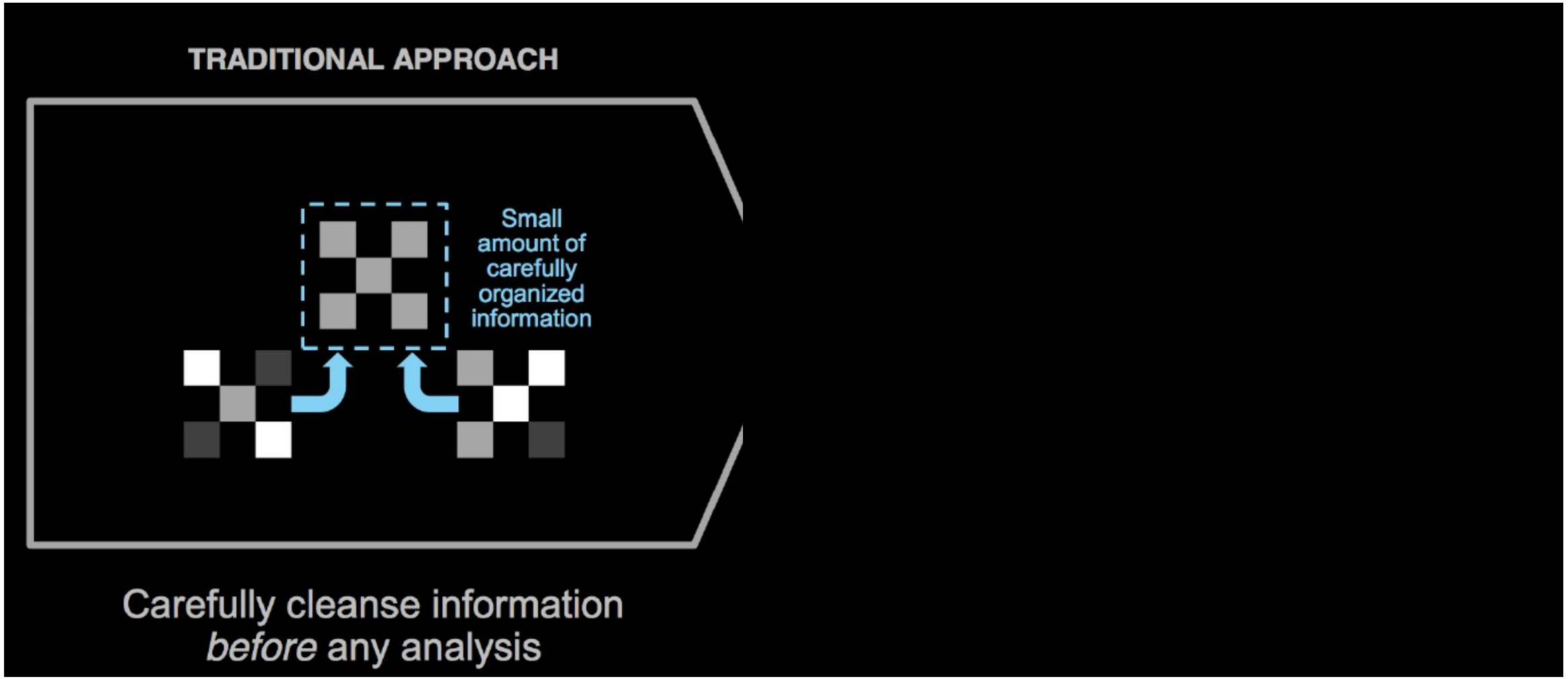
Promises and pitfalls of big data

Promise: leverage more of the data being captured

Andrews, 2014

Promises and pitfalls of big data

Promise: reduce effort required to leverage data

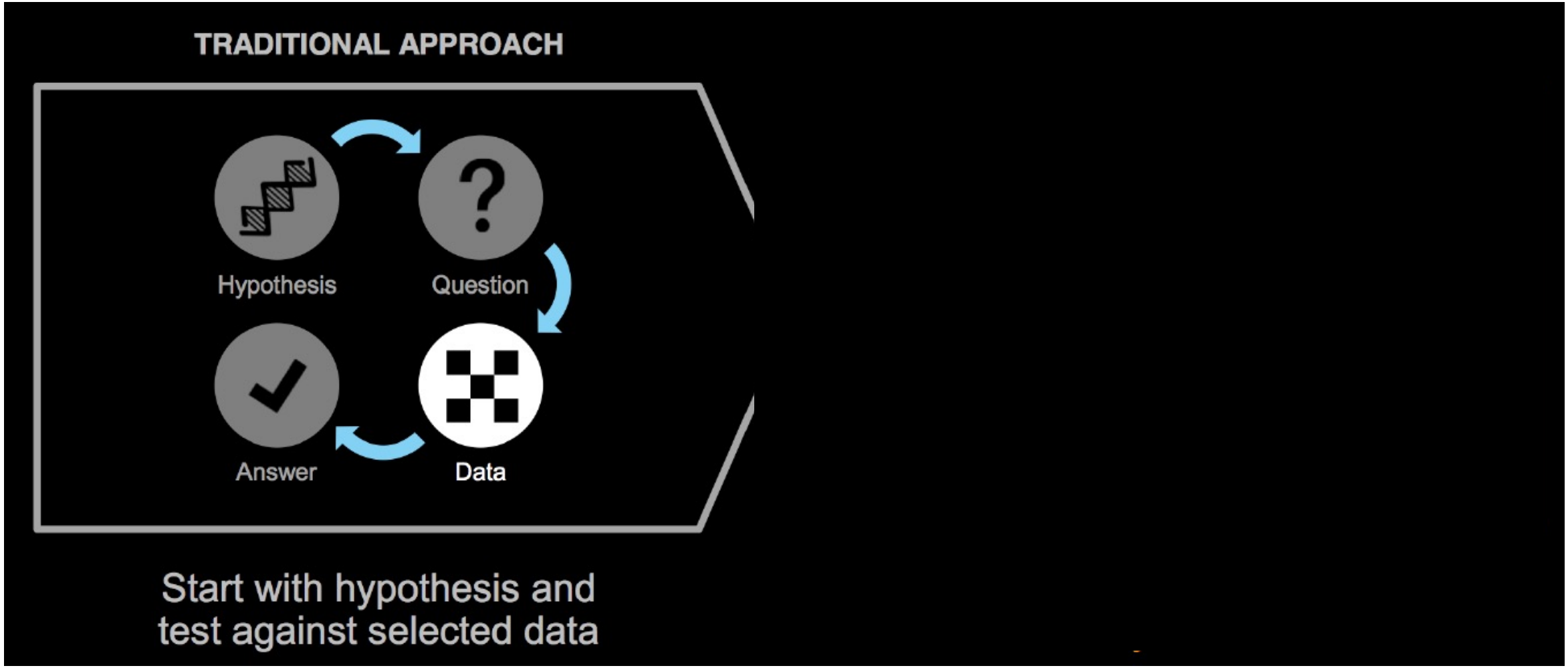


Andrews, 2014

Promises and pitfals of big data
Reduce effort required to leverage data

Promises and pitfals of big data

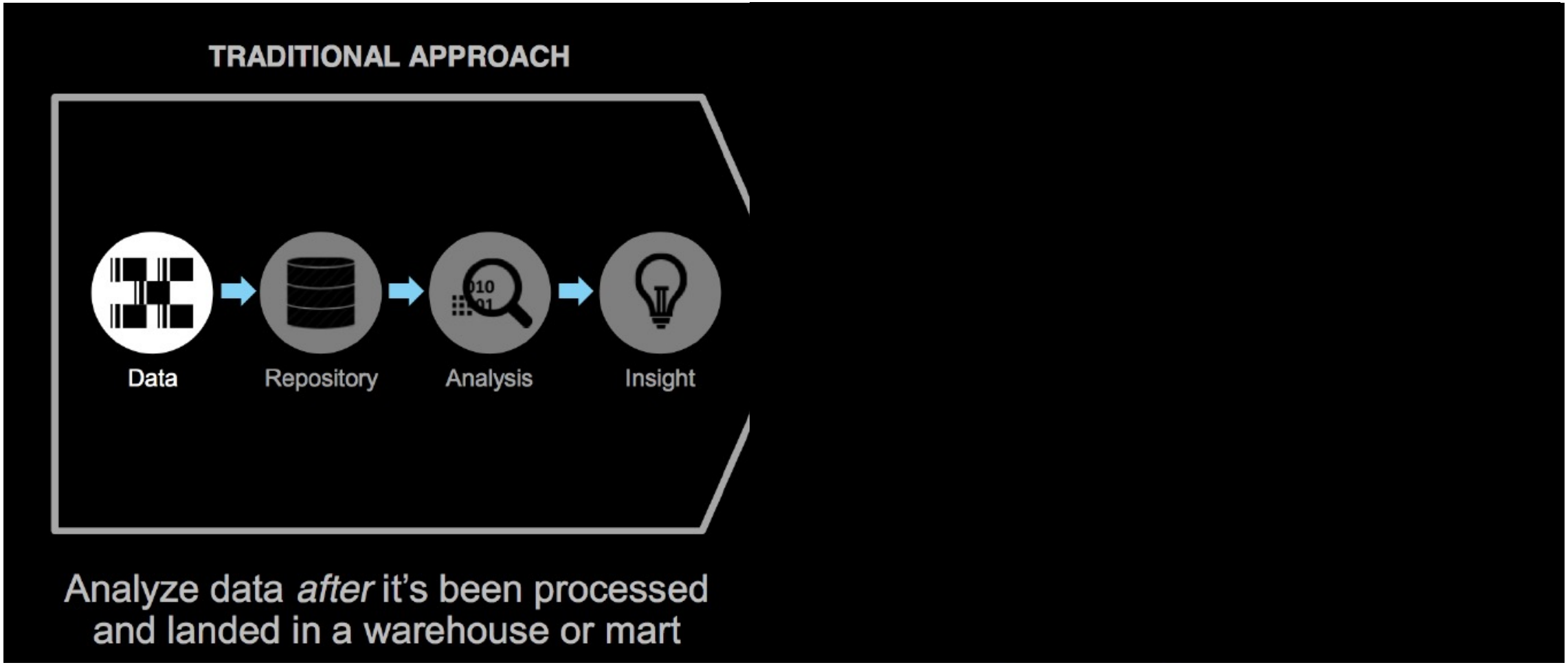
Promise: data-driven exploration looking for correlation



Andrews, 2014

Promises and pitfals of big data

Promise: leverage data as it is captured

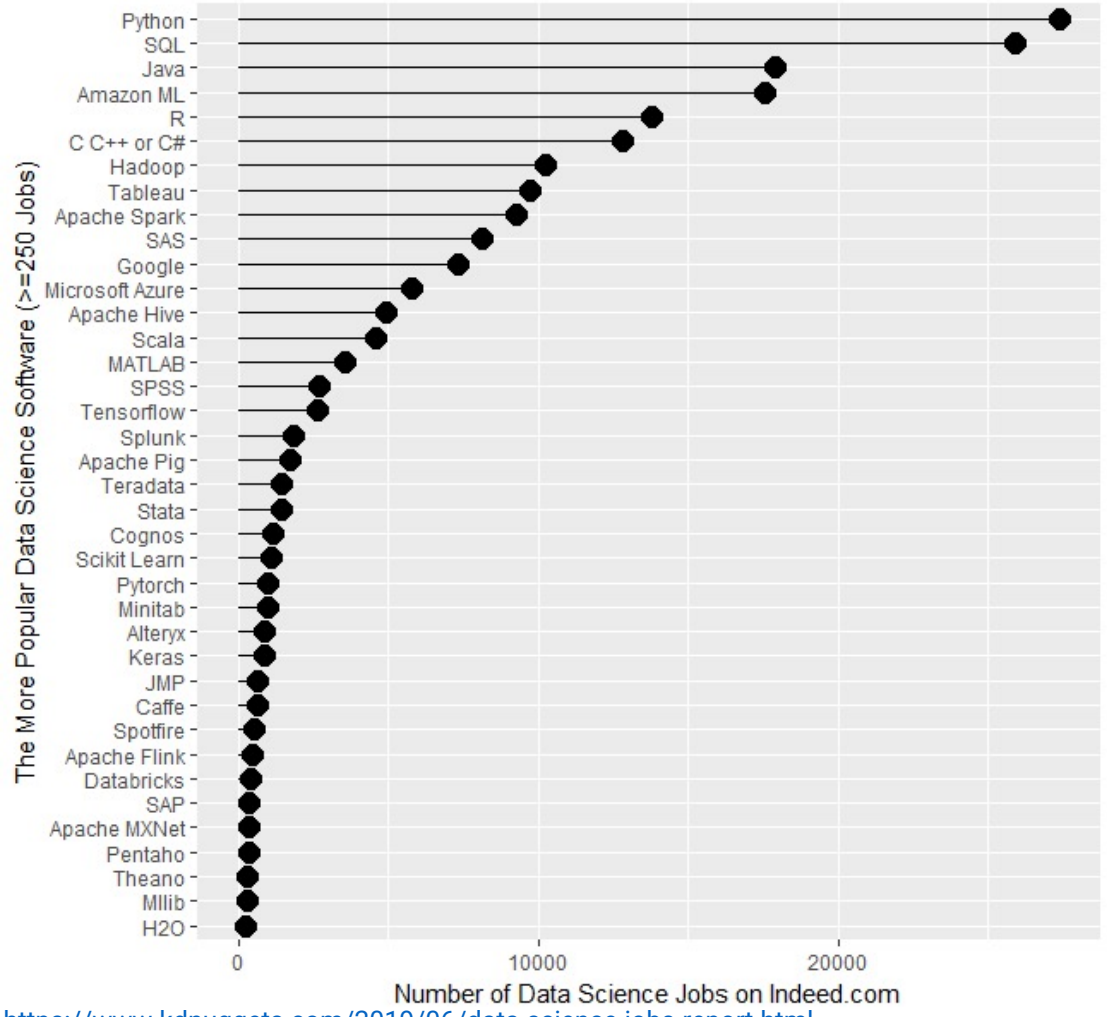


Andrews, 2014

Promises and pitfals of big data

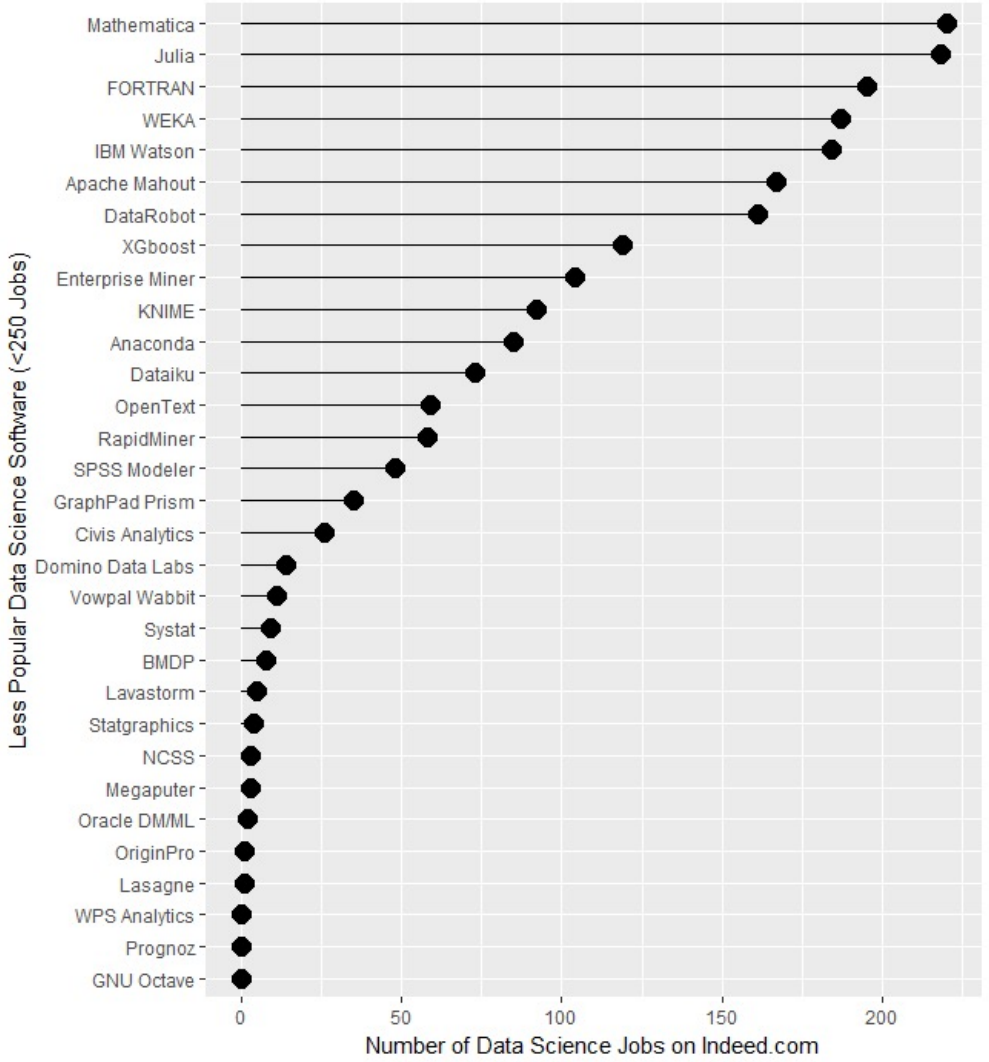
Diverse skills needed

The upper end



https://www.kdnuggets.com/2019/06/data-science-jobs-report.html

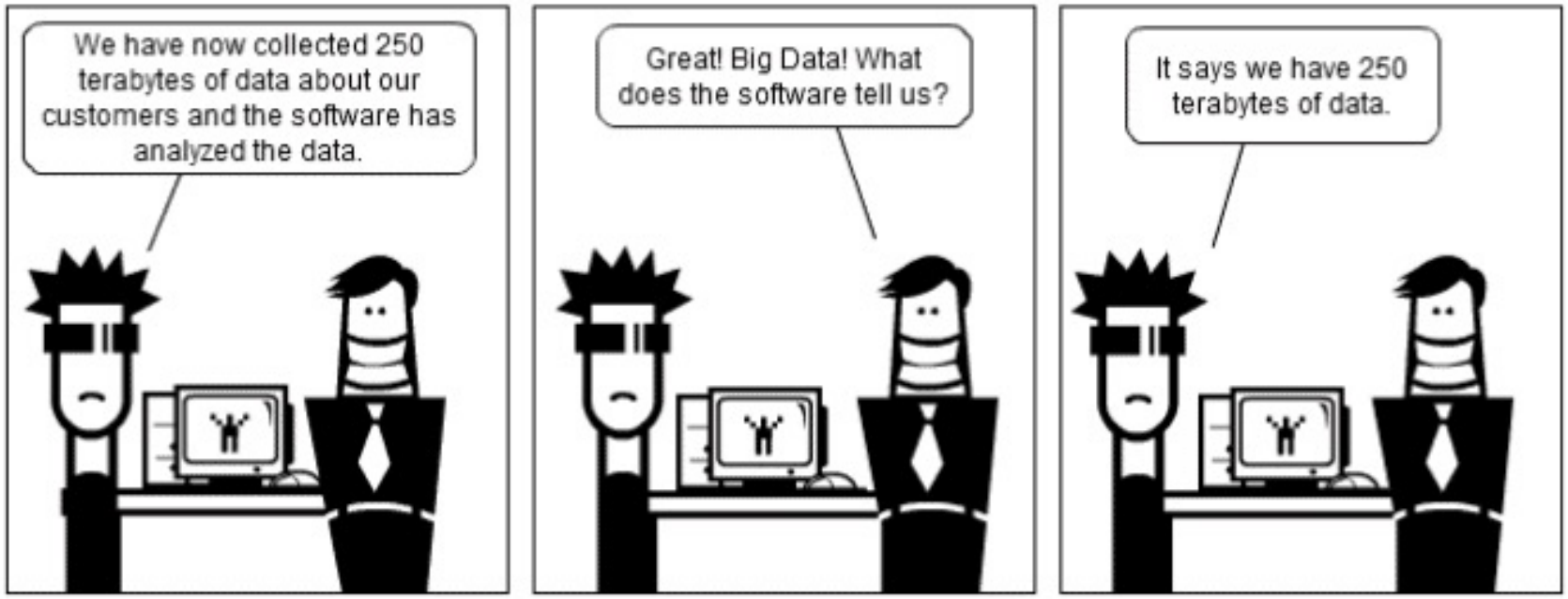
The long tail



Promises and pitfals of big data
Drinking from the data firehose

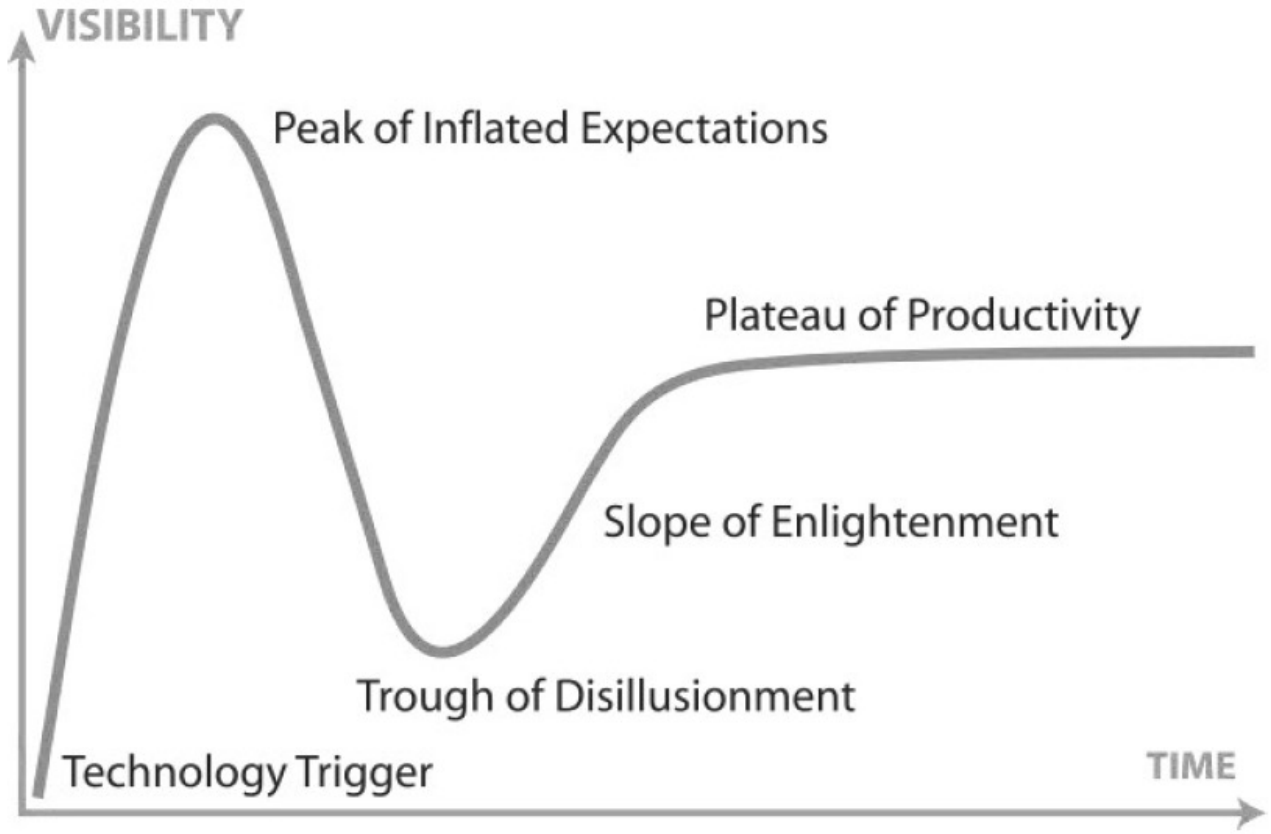
https://www.bondcap.com/report/itr19/#view/157

Promises and pitfals of big data
The big data challenge



https://www.socmedsean.com/comic-the-critical-element-of-a-successful-big-data-strategy/

Promises and pitfals of big data
Big data: a buzz word?



Gartner Hype Cycle

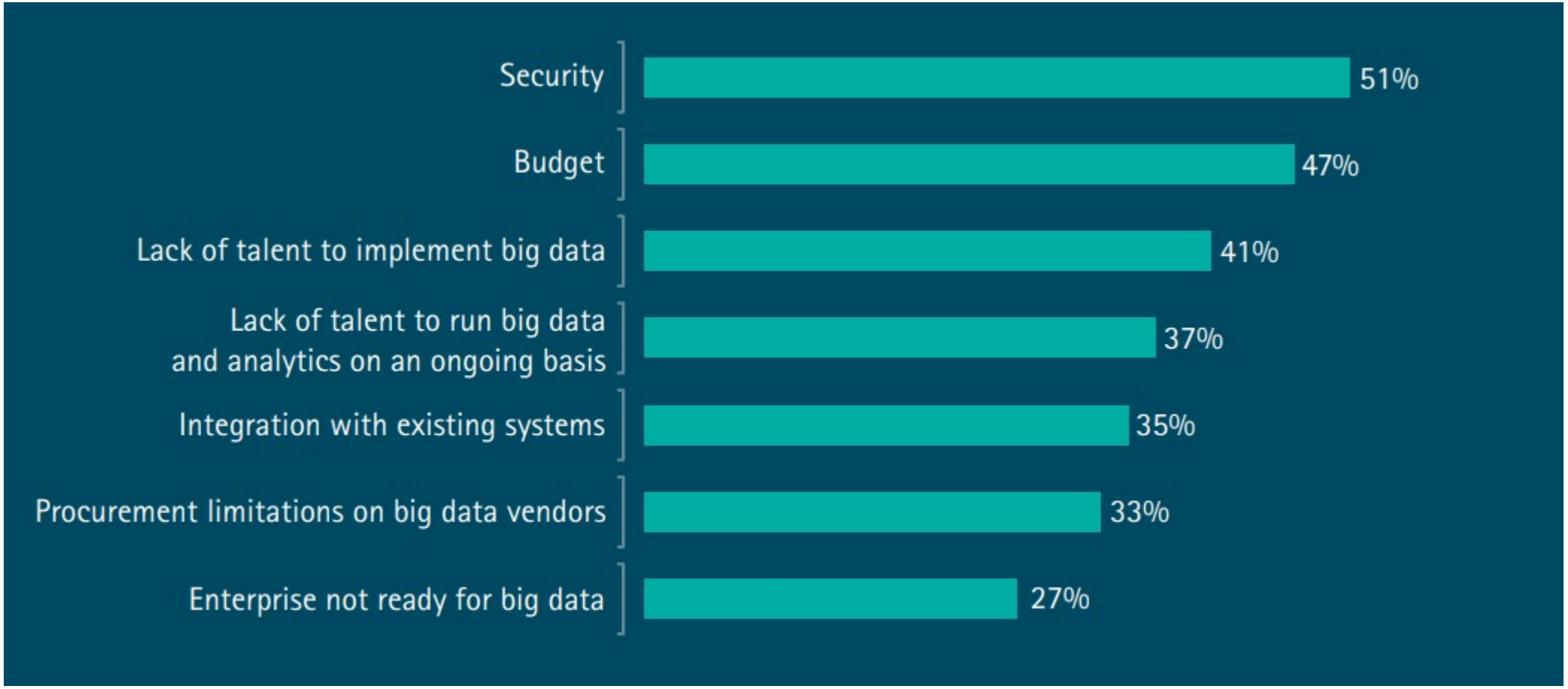
Data is now fundamental to how people work & the most successful companies have intelligently integrated it into everyone's daily workflow…

Data is the new application.

Frank Bien – CEO & President, Looker, 6/19

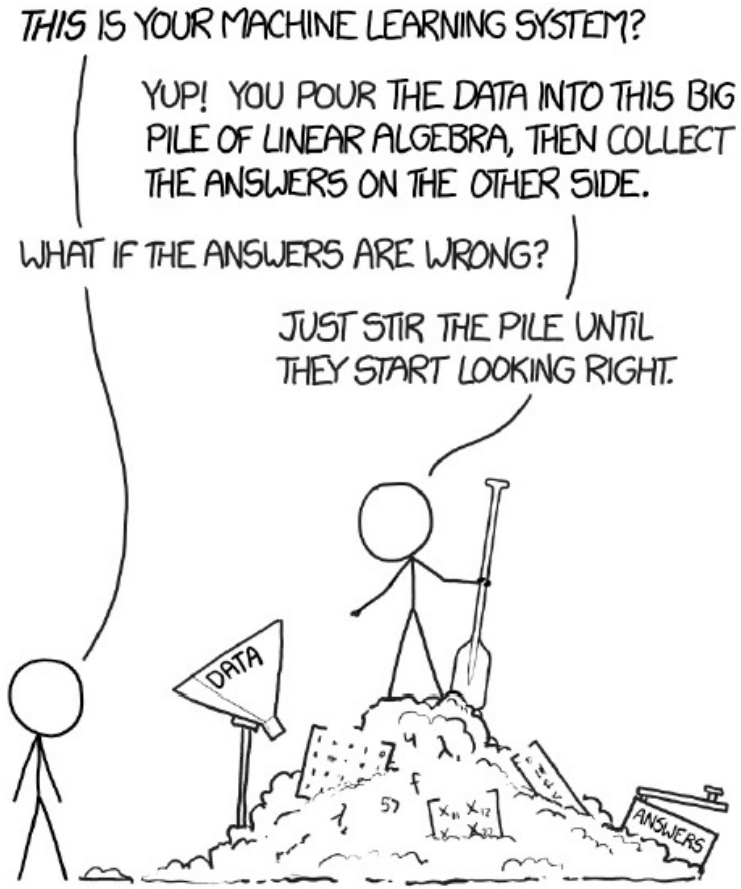
Promises and pitfals of big data

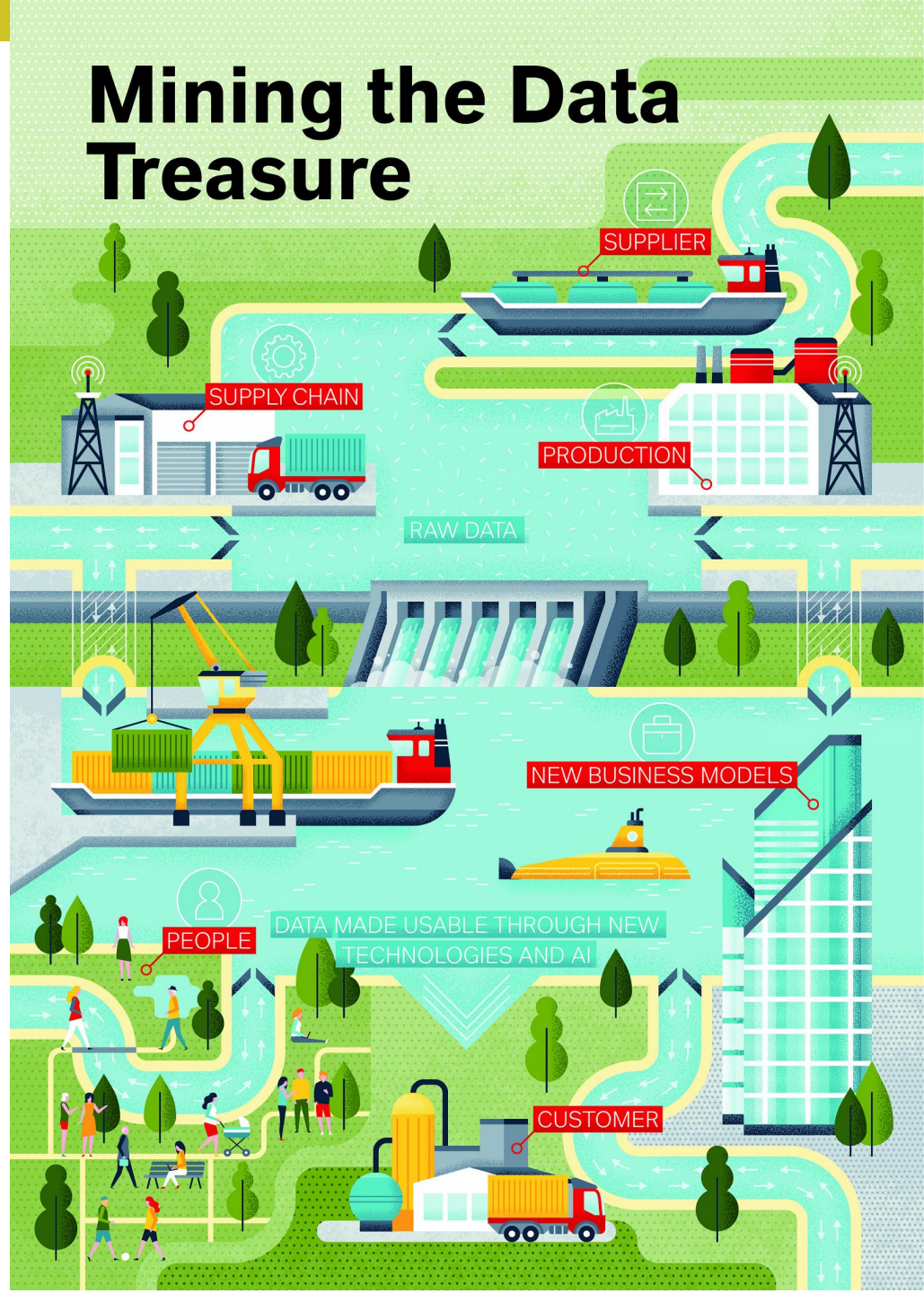
Main challenges with big data projects?



https://www.accenture.com/us-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_14/Accenture-Big-Data-POV.pdf

Promises and pitfals of big data
Other buzzword „data lake“





https://webmagazine.lanxess.com/the-treasure-of-the-data-lake/

Why privacy for mobility data is a concern?

- Mobility is a *sensitive* type of information
  Depending on the location visited, one could infer religious preferences, daily habits, health problems.
- Mobility data is *abundant* and readily available
  Location based services, social media access etc.

Short Answer:

Big Data ≈ Data Mining ≈ Predictive Analytics ≈ Data Science

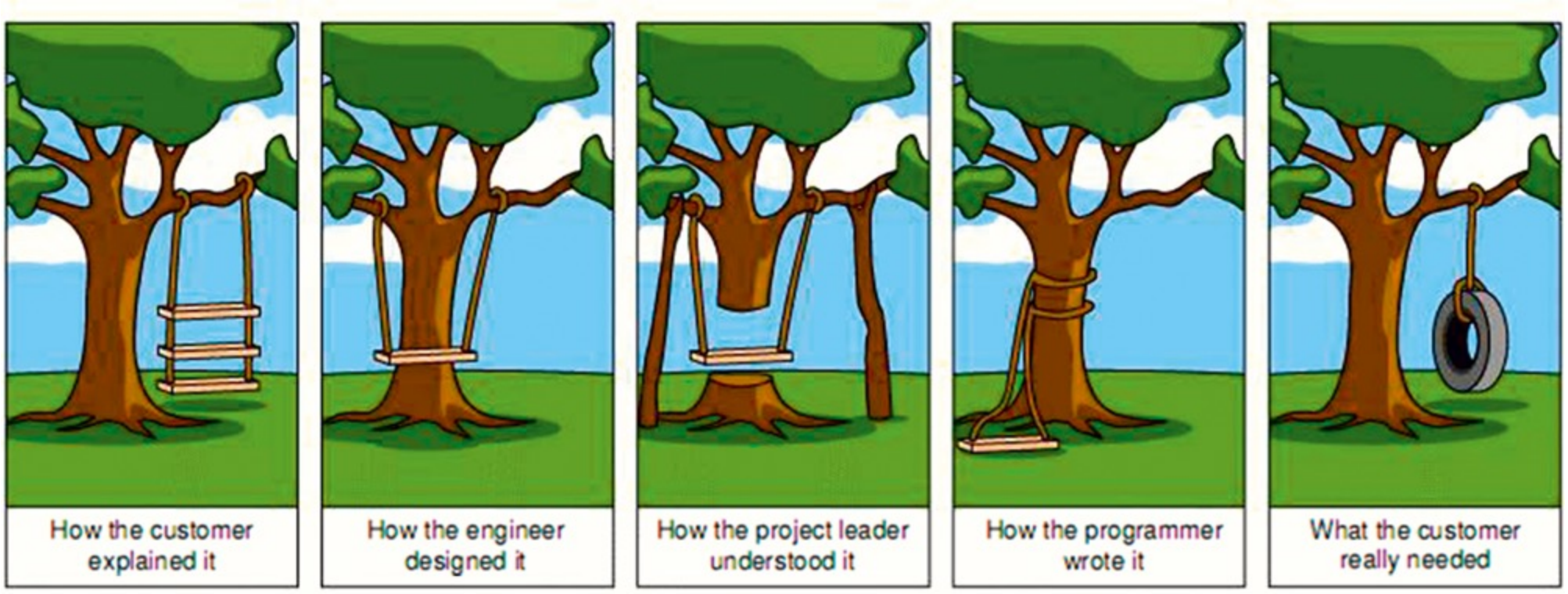
(Leskovec et al., 2014)

[12]

Tools für die Arbeit mit Mobilitätsdaten

Roles in a data project

How to translate stakeholders interests between different domains



Roles in a data project

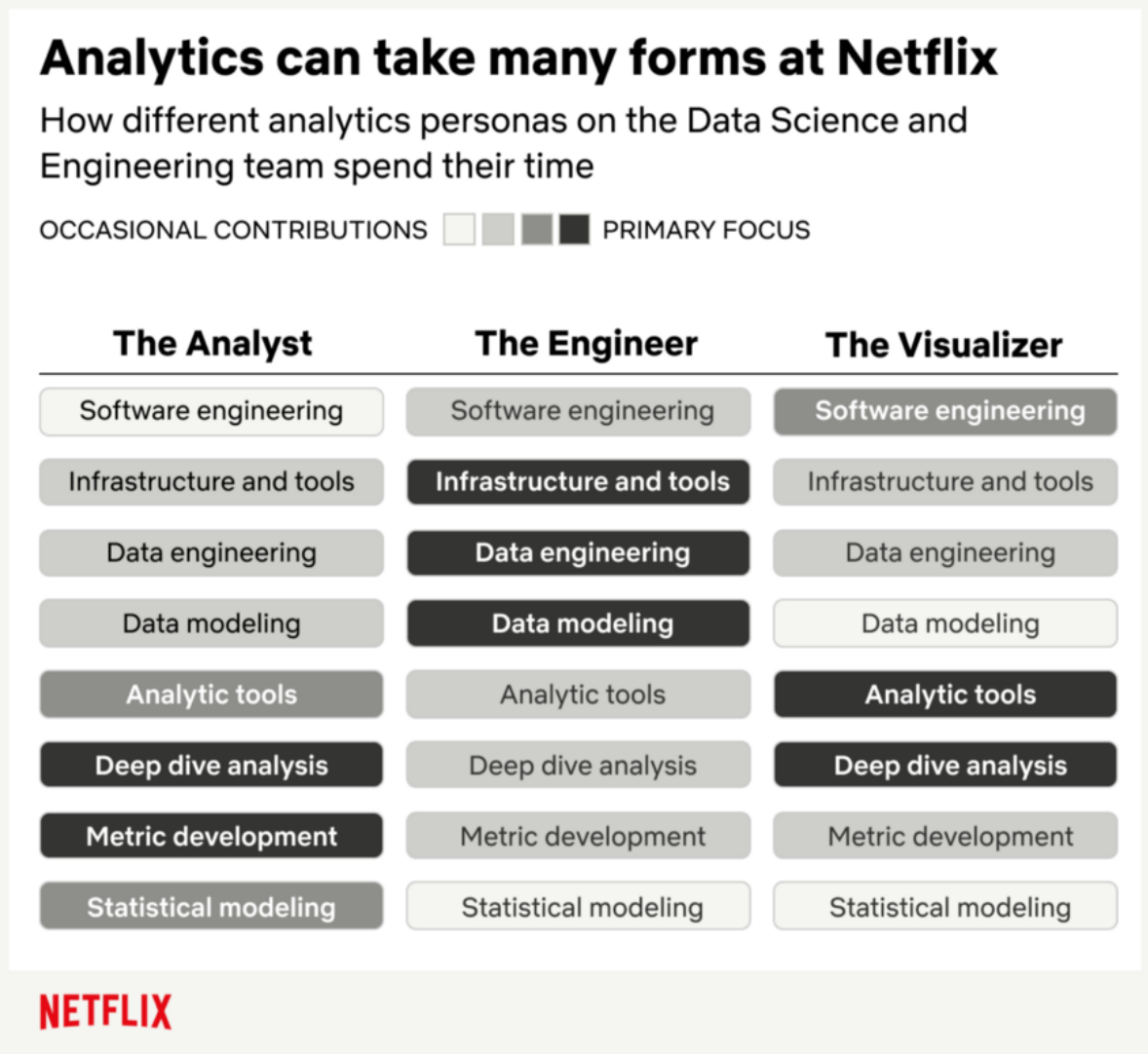
Who are the stakeholder and how do you find a common language?

Role	Responsibilities
Project sponsor	Represents the business interests; champions the project
Client	Represents end users' interests; domain expert
Data scientist	Sets and executes analytic strategy; communicates with sponsor and client
Data architect	Manages data and data storage; sometimes manages data collection
Operations	Manages infrastructure; deploys final project results

Source: Zumel, Mount, 2019, Practical Data Science with R.

Roles in a data project

Who are the stakeholder and how do you find a common language?



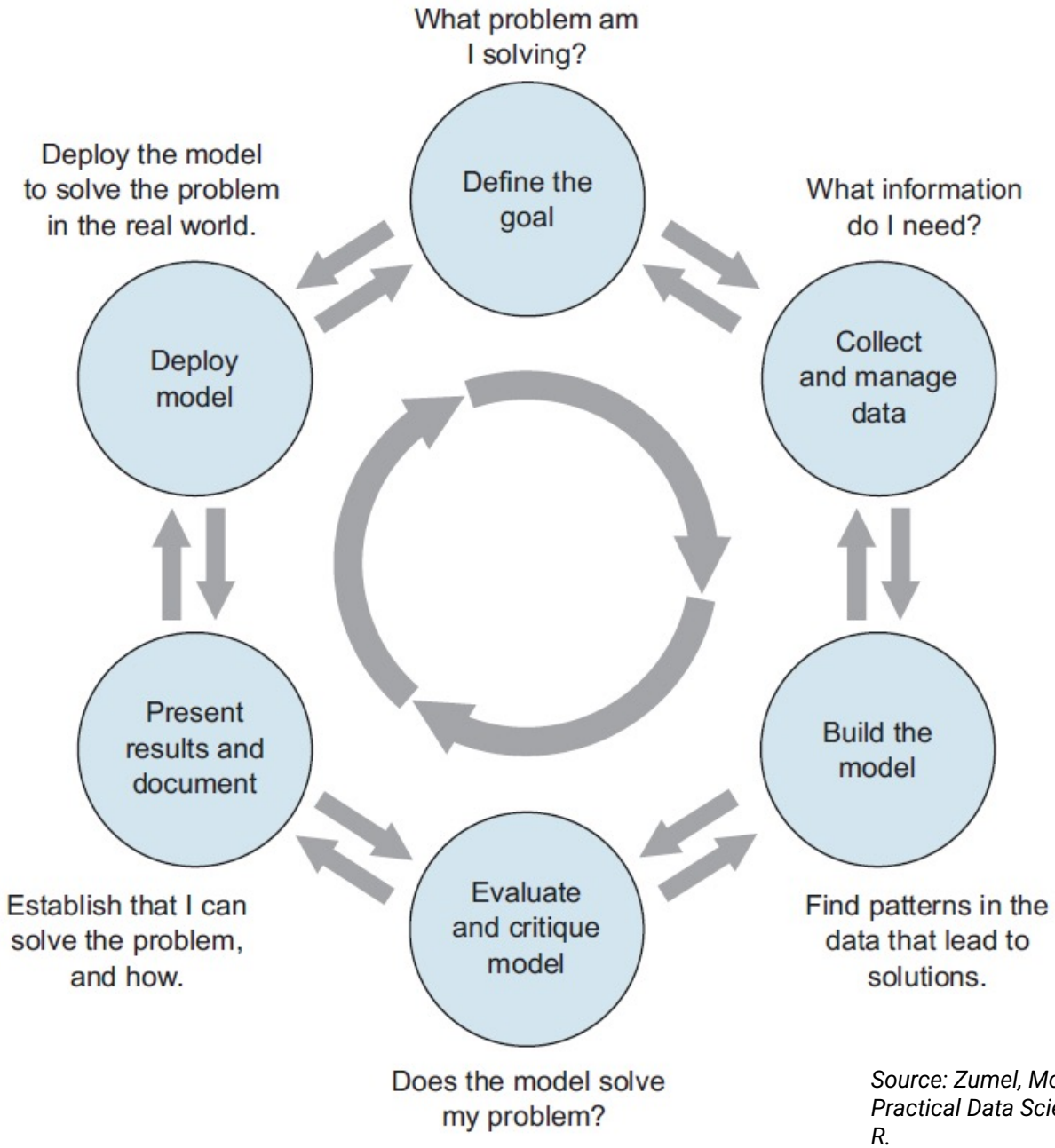
Source: https://netflixtechblog.com/analytics-at-netflix-who-we-are-and-what-we-do-7d9c08fe6965

Stages of a data project

The lifecycle of a data science project

Defining the goal

- Why do the sponsors want the project in the first place? What do they lack, and what do they need?
- What are they doing to solve the problem now, and why isn't that good enough?
- What resources will you need: what kind of data and how much staff? Will you have domain experts to collaborate with, and what are the computational resources?
- How do the project sponsors plan to deploy your results? What are the constraints that have to be met for successful deployment?



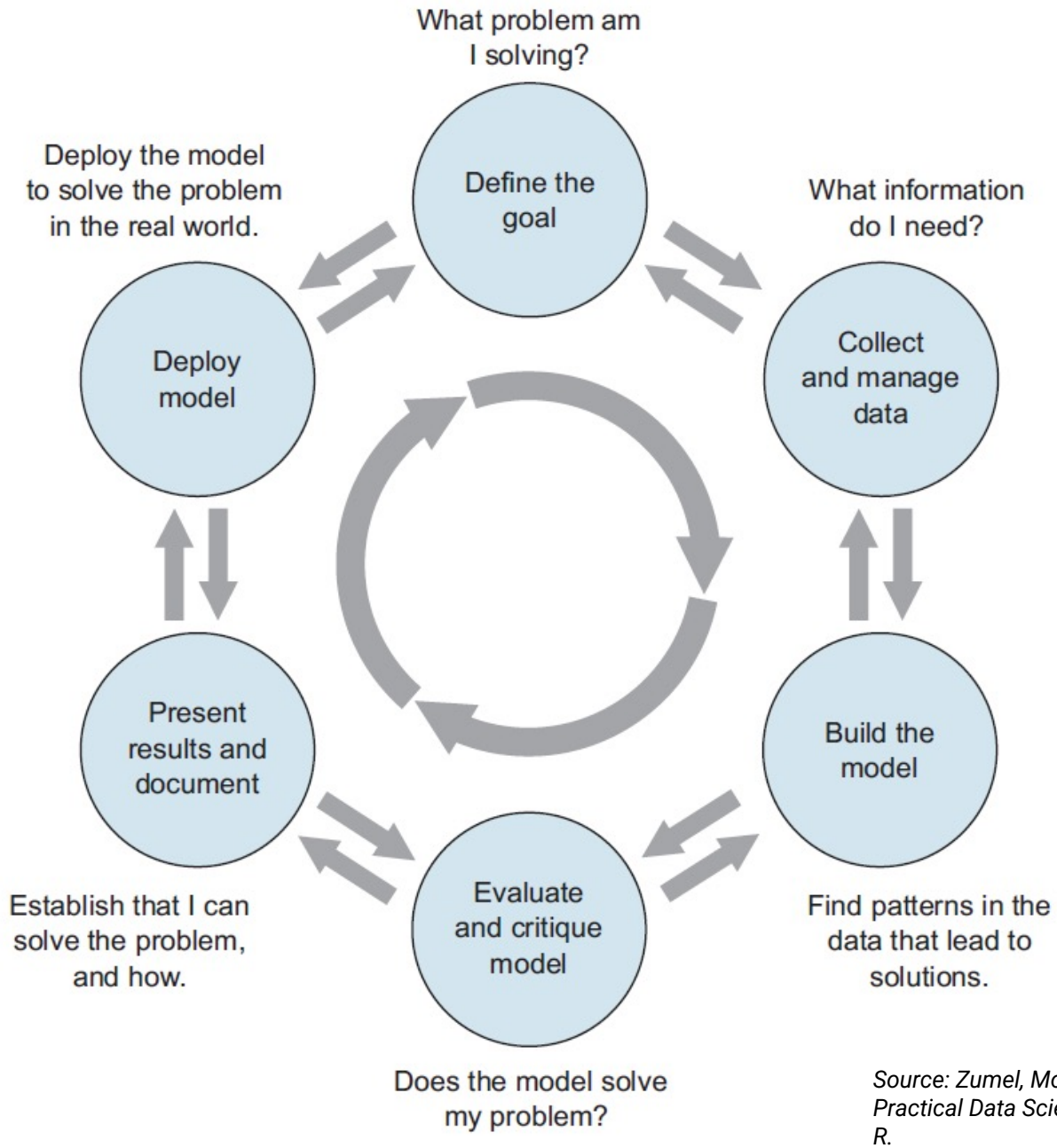
Source: Zumel, Mount, 2019, Practical Data Science with R.

Stages of a data project

The lifecycle of a data science project

Collect and manage data

- What data is available to me?
- Will it help me solve the problem?
- Is it enough?
- Is the data quality good enough?

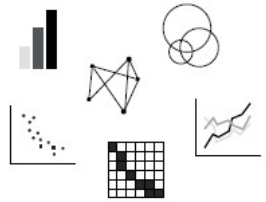


Source: Zumel, Mount, 2019, Practical Data Science with R.

Mobility analytics:
Basic components

(Random) sample

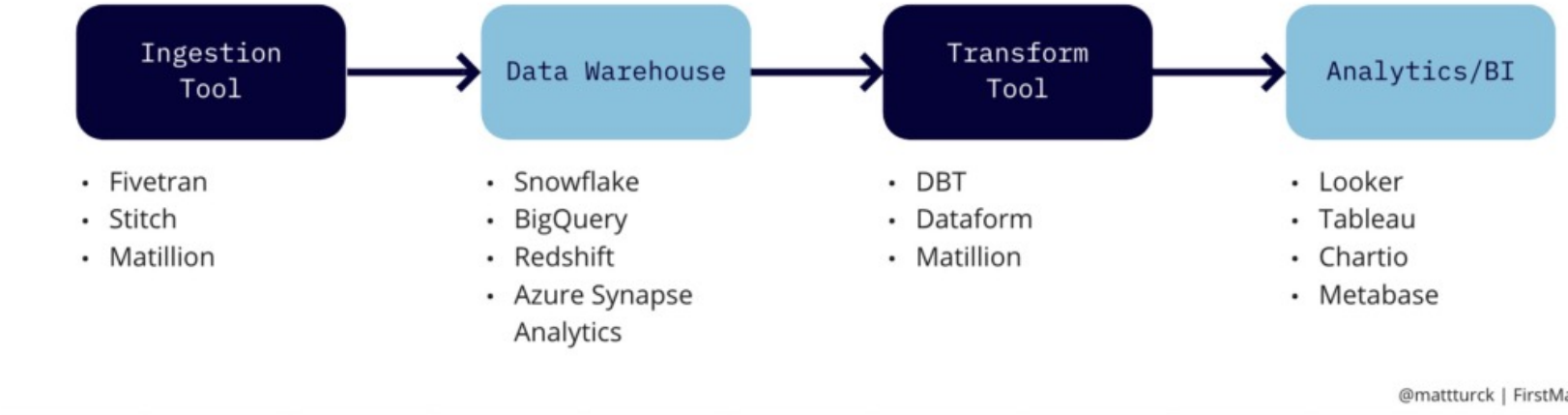
Different Data sources



MATSim
Multi-Agent Transport Simulation

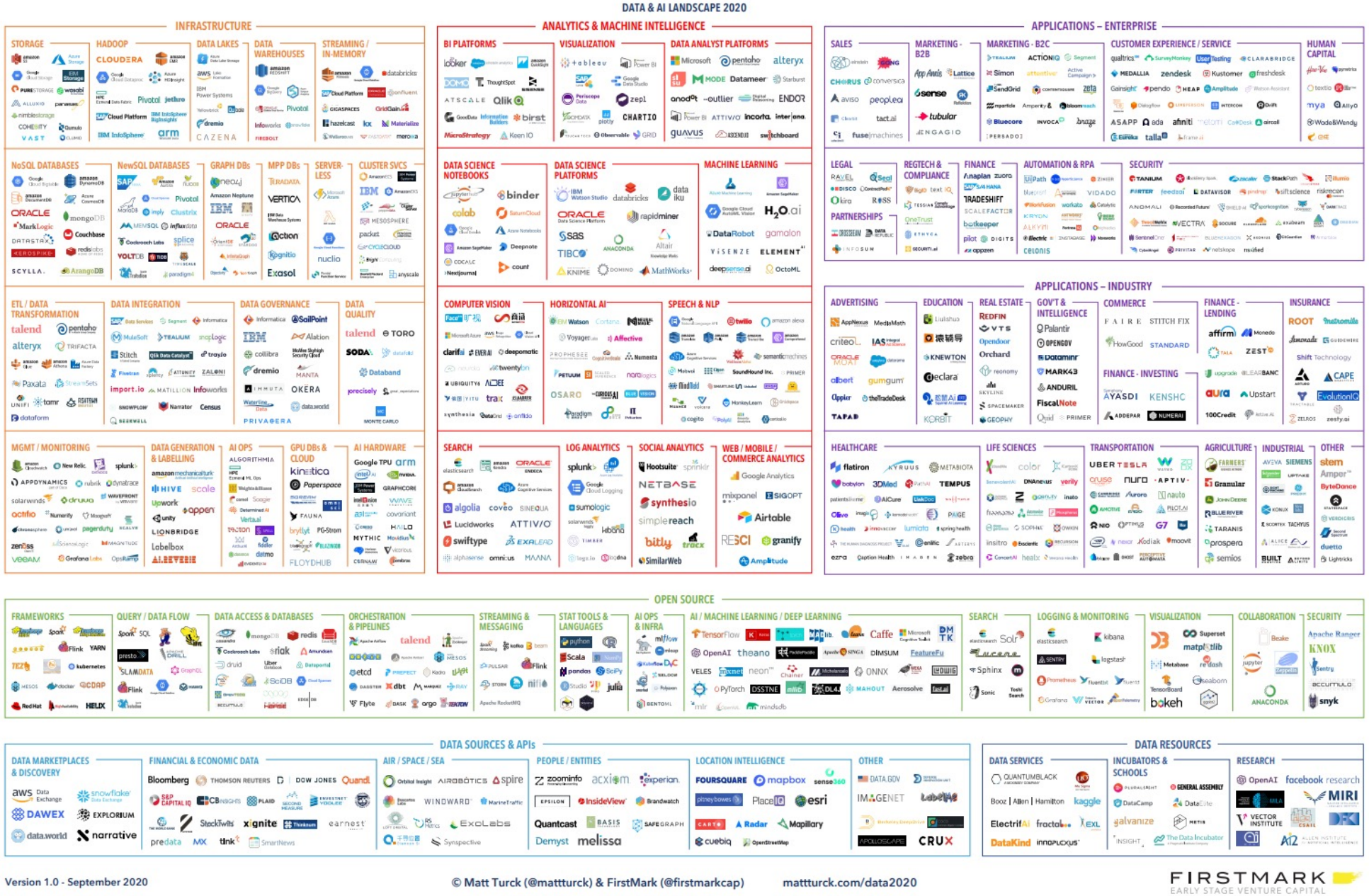
Data analysis, Visualisation, modelling

Mobility analytics:
Unbundling the analytics process



Tools

The big unbundling

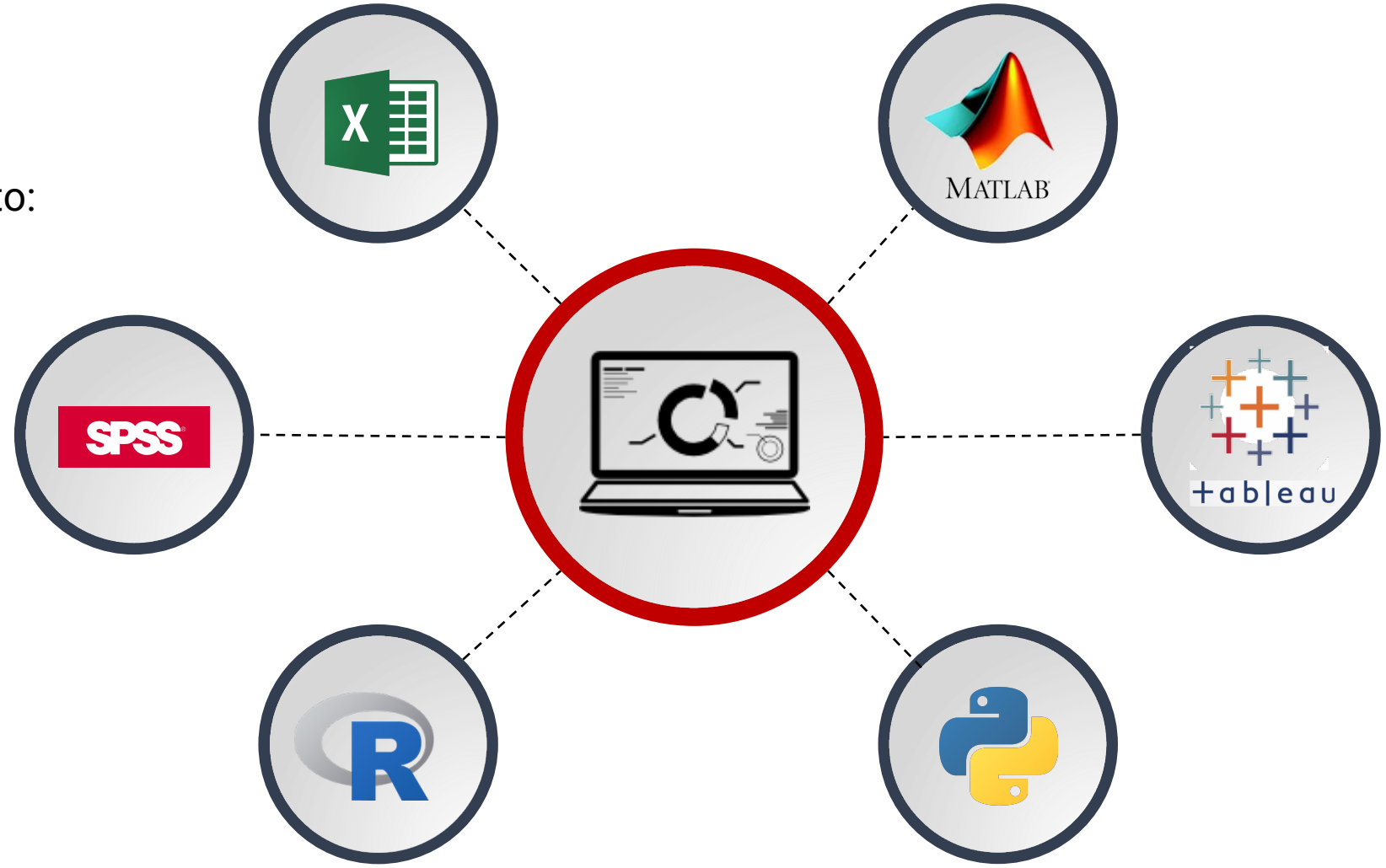


Why R?
Data Analytics and visualization software

Software that enables to:

- Prepare
- Filter
- Select
- Analyze
- Visualize

data.



Why R?
Data Analytics and visualization software

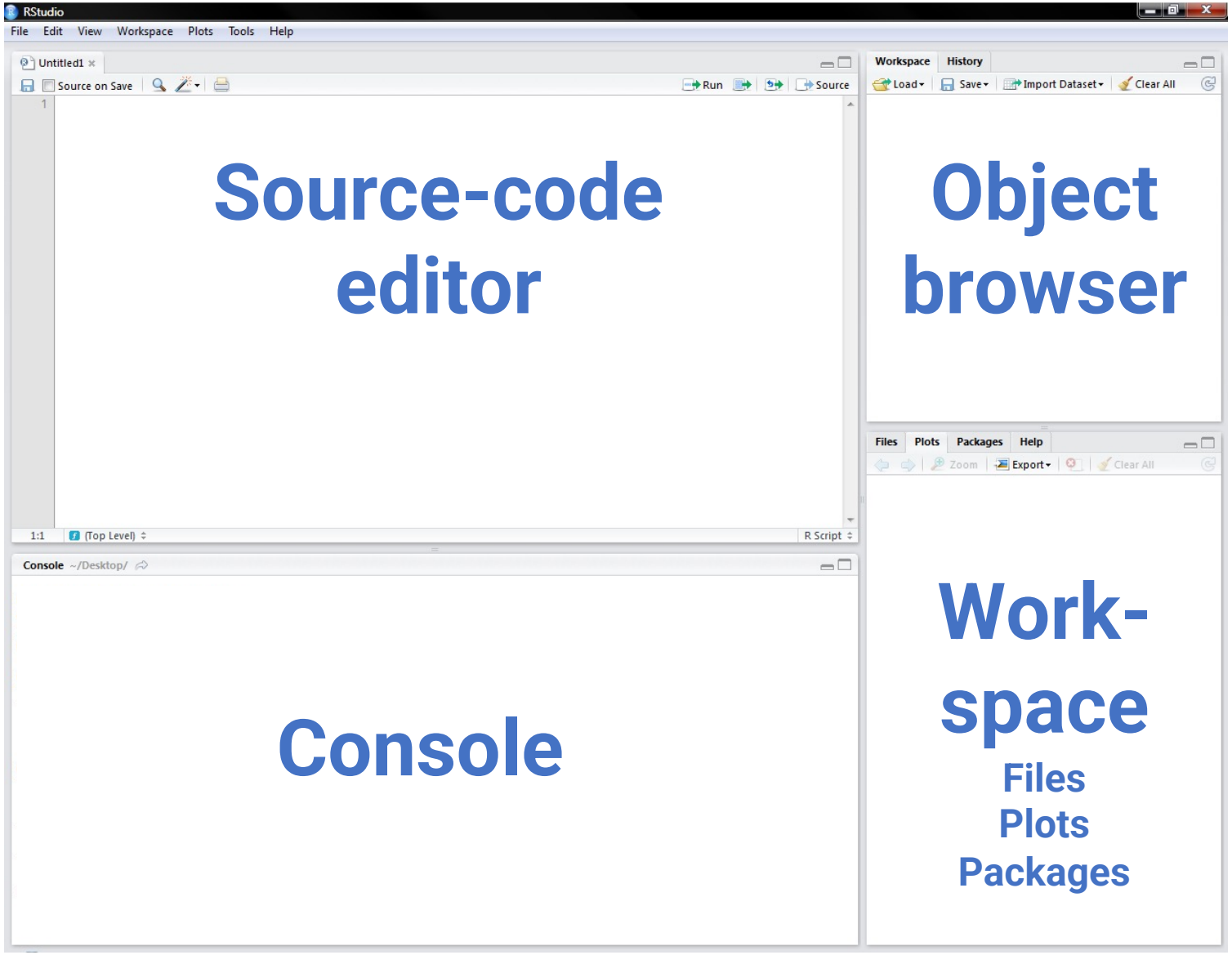
- Can be used as a **statistical language** and a **programming language**
- Very powerful and diverse data analysis and **visualization possibilities**
- **Simple and easy to learn**
- Very powerful but free and **open source**
- **user contributed add-on packages** to supplement the base distribution: worldwide repository system: Comprehensive R Archive Network (CRAN)

Why Rstudio?

- Most famous IDE (integrated development environment)
- working environment to analyse data with R

Everyone installed R and RStudio? Let‘s start!

The R Studio workspace



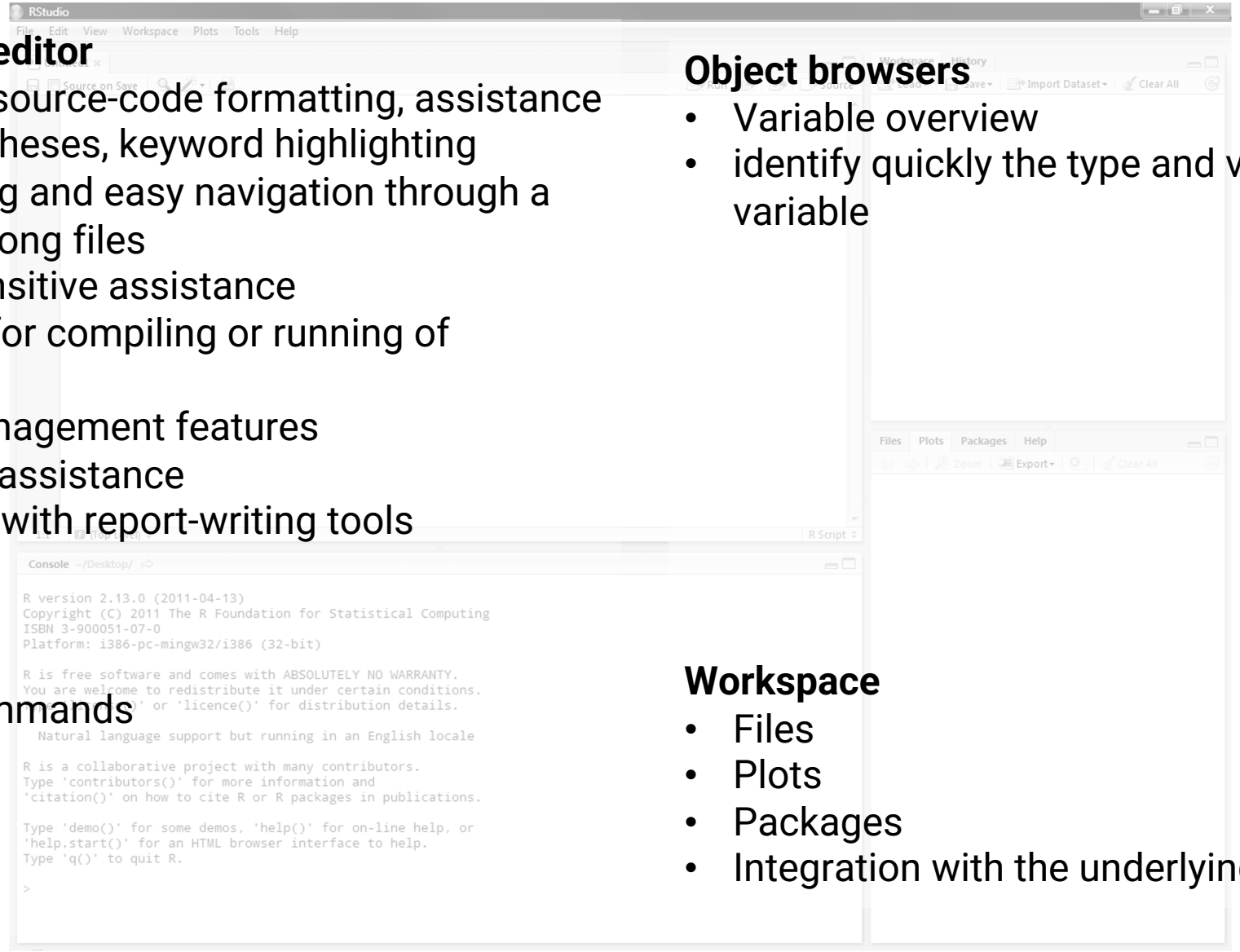
The R Studio workspace

Source-code editor

- automatic source-code formatting, assistance with parentheses, keyword highlighting
- code folding and easy navigation through a file and among files
- context-sensitive assistance
- interfaces for compiling or running of software
- project-management features
- debugging assistance
- integration with report-writing tools

Console

- issuing commands



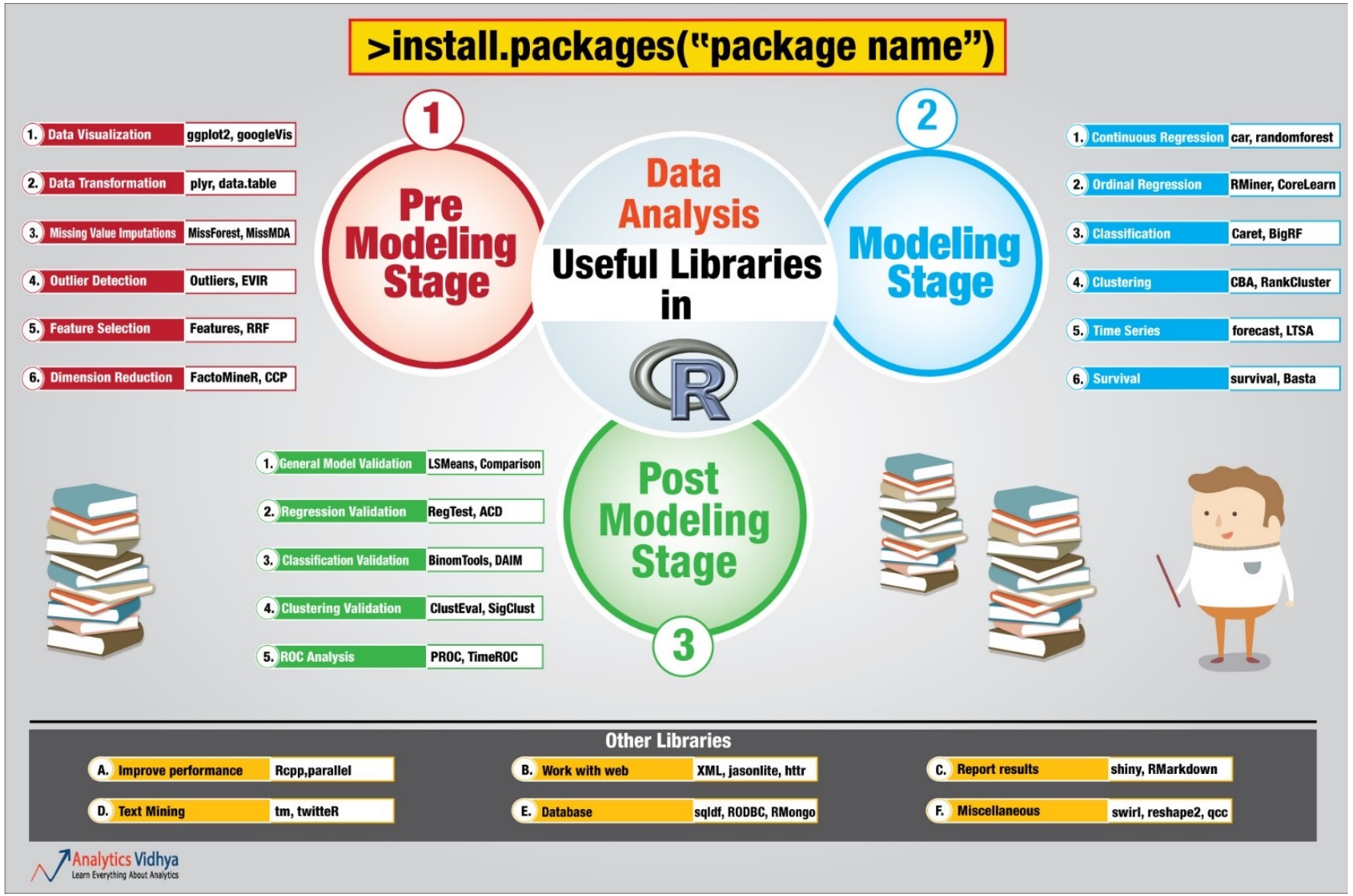
Object browsers

- Variable overview
- identify quickly the type and values for each such variable

Workspace

- Files
- Plots
- Packages
- Integration with the underlying documentation

Packages



Package
reusable R function(s) with standard and self-explanatory documentation on how to use it.

[13]

Auswertung der erhobenen Mobilitätsdaten

Daten

https://hessenbox-a10.rz.uni-frankfurt.de/dl/fi88asBdwcm9oRT5tWy3rFy9

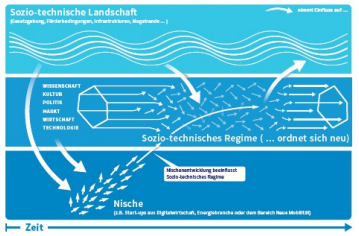
https://hessenbox-a10.rz.uni-frankfurt.de/dl/fizfxmvtQ87MXYmT4EeeiB7/tracks.geojson

Wrap up

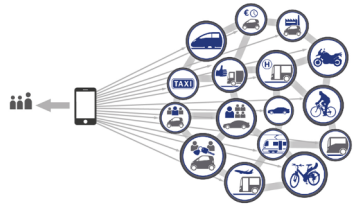
Definition Nachhaltige Mobilität

Komplexität von Ursache-Wirkungs-Beziehungen

Theorie des Wandels: Wenn die Aufgabe so komplex ist, wie kommt dann überhaupt das Neue in die Welt?



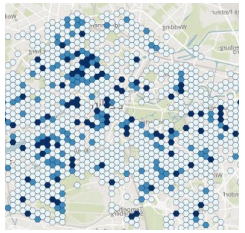
Praxis des Wandels anhand eines Beispiels: Wie kann der Wandel praktisch gelingen? Eine konzeptionelle Annäherung über das Bsp. MaaS.



Praxis des Wandels: Wie wird ein Produkt entwickelt?

Praxis des Wandels: Welche Rolle spielen Daten bei der Produktentwicklung?

Praxis des Wandels: Woher kommen Daten für die Entwicklung von Mobilitätsprodukten?



Praxis des Wandels: Wie arbeitet man mit Mobilitätsdaten?