bandicoot

> an open-source Python toolbox to analyze mobile phone metadata

Luc Rocher (luc@rocher.lc / @cynddl)

in collaboration with Yves-Alexandre de Montjoye
and Alex 'Sandy' Pentland

UCL Université catholique de Louvain

mit media lab

Quantifying the Impact of Human Mobility on Malaria

Amy Wesolowski,[1,2] Nathan Eagle,[3,4] Andrew J. Tatem,[5,6,7] David L. Smith,[6,8]
Abdisalan M. Noor,[9,10] Robert W. Snow,[9,10] Caroline O. Buckee[4,11]*

NATURE | NEWS FEATURE

Computational social science: Making the links

From e-mails to social networks, the digital traces left by life in the modern world are transforming social science.

Jim Giles

Impact of human mobility on the emergence of dengue epidemics in Pakistan

Amy Wesolowski[a,b], Taimur Qureshi[c], Maciej F. Boni[d,e], Pål Roe Sundsøy[c], Michael A. Johansson[b,f],
Syed Basit Rasheed[g], Kenth Engø-Monsen[c], and Caroline O. Buckee[a,b,1]

SOCIAL SCIENCE

Computational Social Science

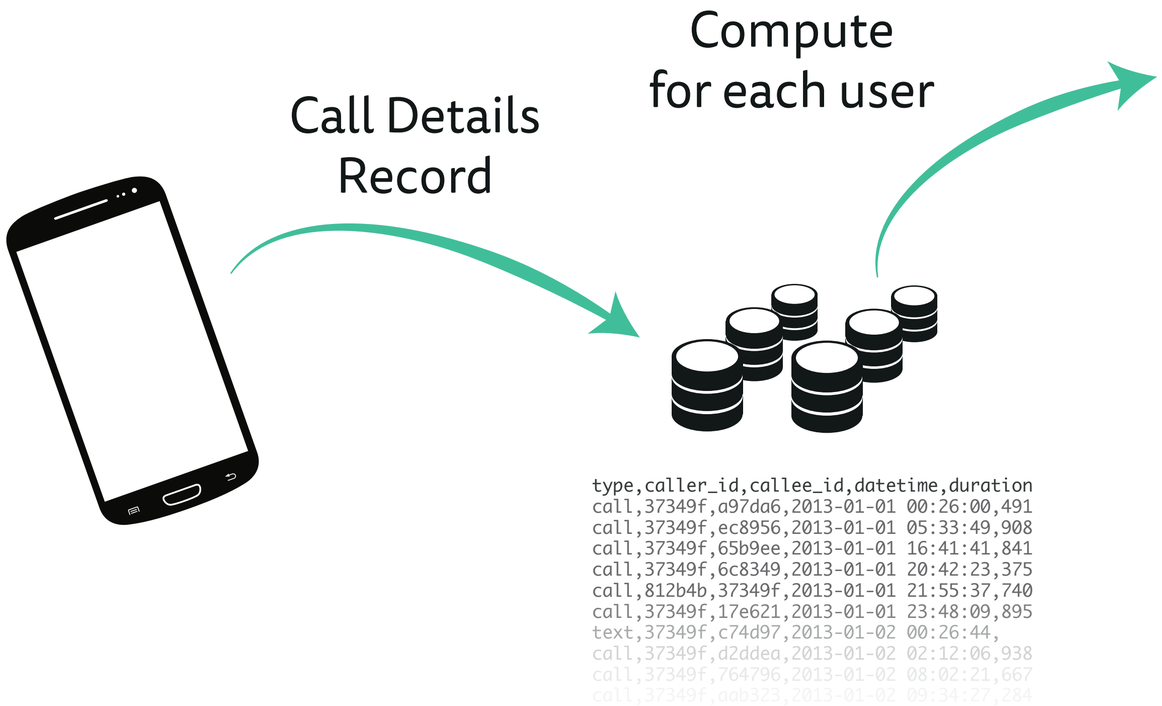
David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5]
Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3]
Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne[2,11]

A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

Two issues for researchers

1. How to access such sensitive data sets?
2. How to replicate results?

bandicoot toolbox

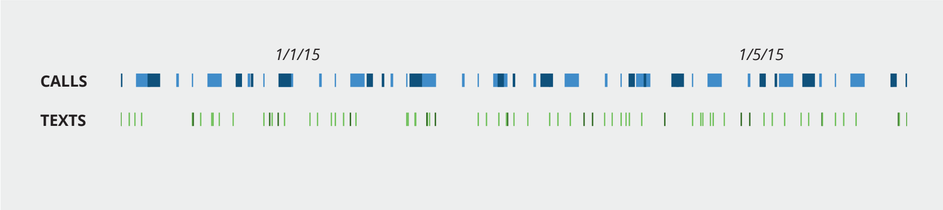


Robust behavioral indicators

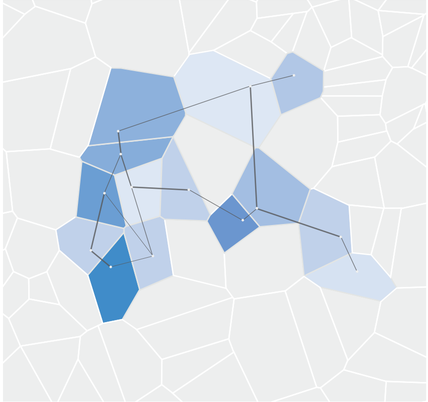
Entropy of contacts
Radius of gyration
Response delay
Percentage at home
Percentage during night
Percentage of text
Number of places visited
Number of contacts

1400+ bandicoot indicators

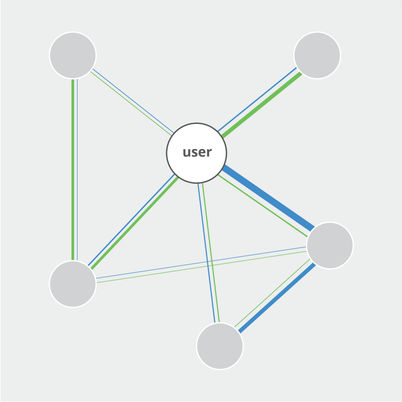
Call and text interactions



Mobility

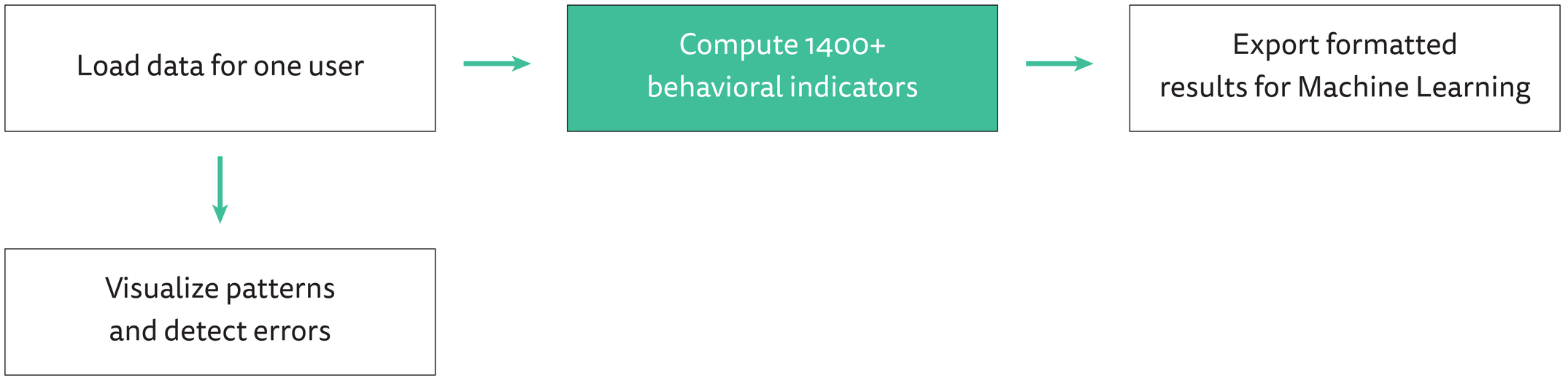


Network



& Mobile phone recharges (top-up)

bandicoot workflow



bandicoot workflow > 4 lines of code

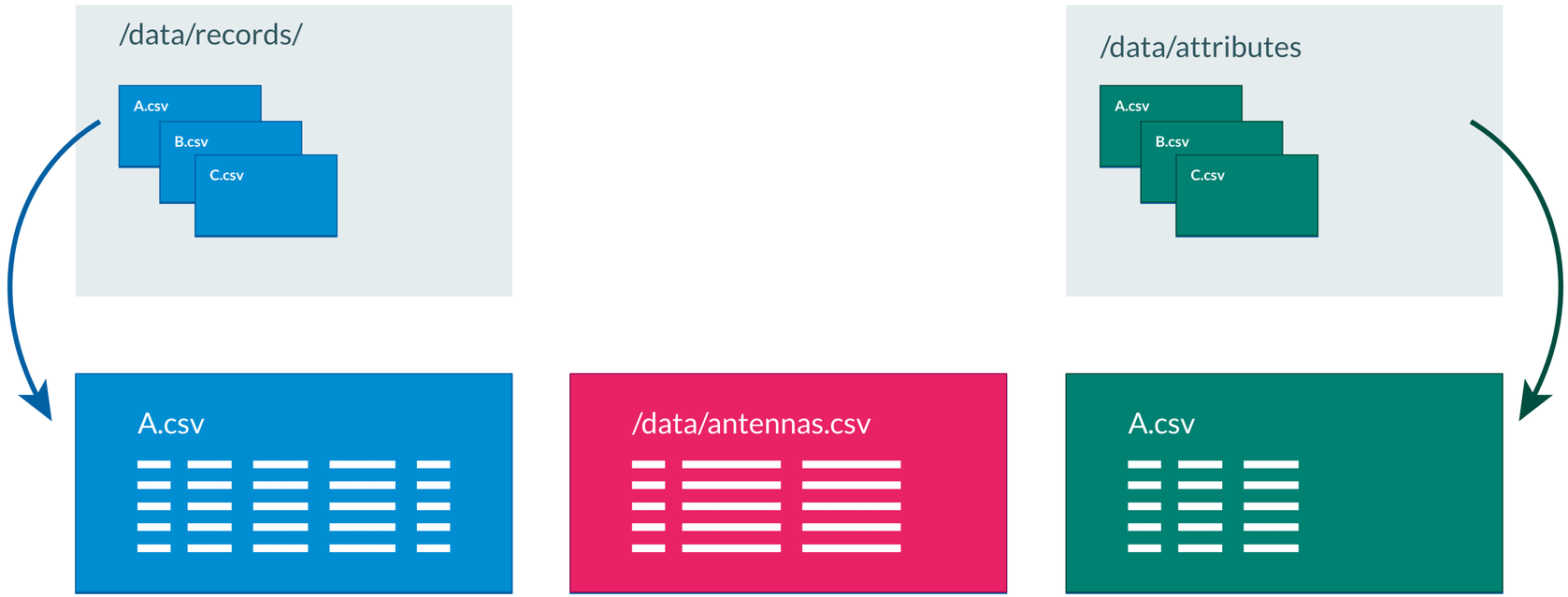
```
import bandicoot as bc

my_user = bc.read_csv('my_user', '/data/records')
features = bc.utils.all(my_user)
bc.to_json(features, 'bandicoot_indicators.json')
```

bandicoot workflow > 4 lines of code

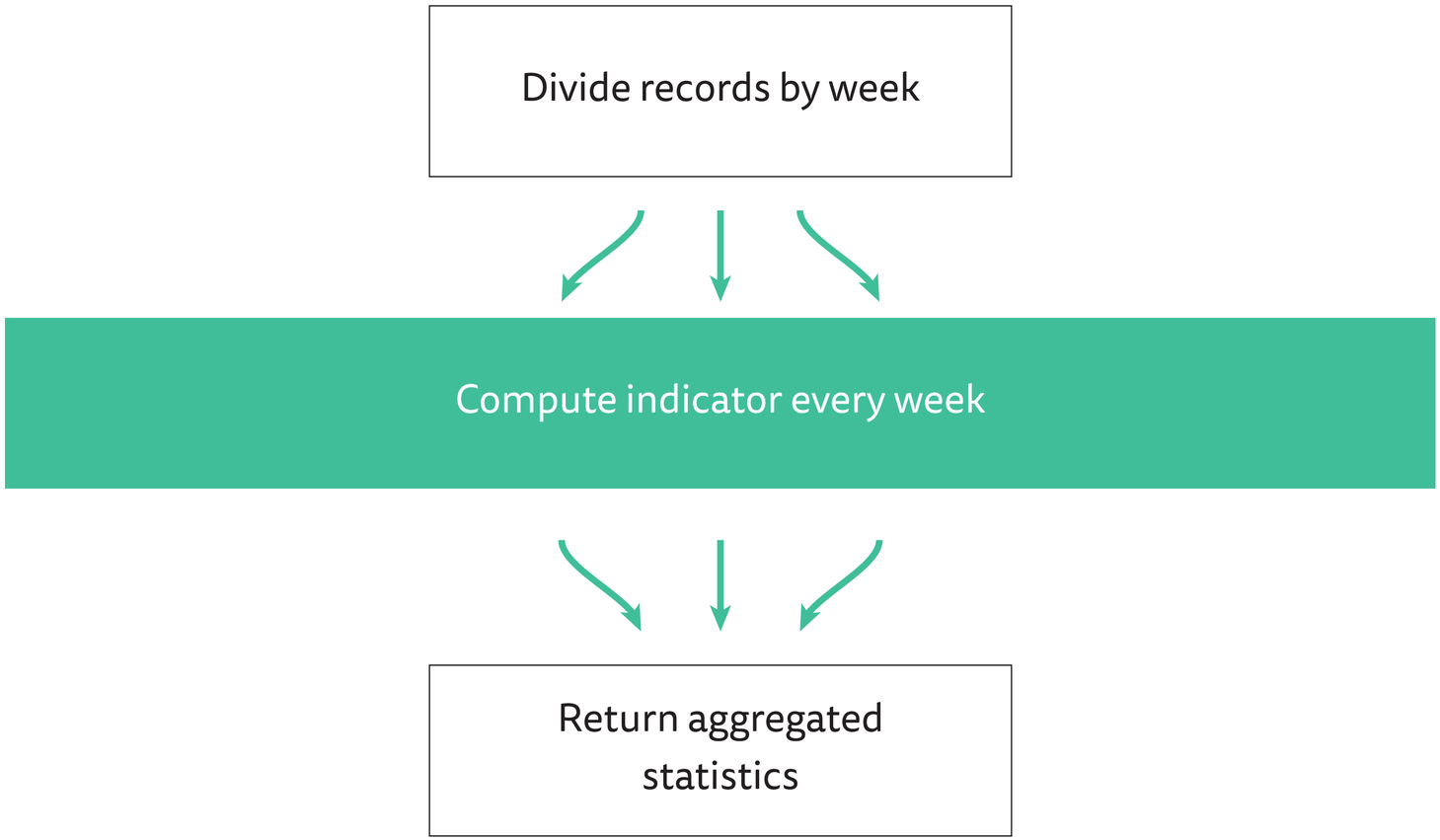
```
{
  "my_user": {
    "reporting__start_time": "2014-08-20 20:30:37",
    "reporting__end_time": "2014-08-21 21:18:31",
    "reporting__bins_without_data": 0,

    [...]

    "response_delay_text__weekday__allday__callandtext__max__mean": 300.0,
    "response_delay_text__weekday__allday__callandtext__median__std": 0.0,
    "response_delay_text__weekday__allday__callandtext__median__mean": 180.0,

    [...]

    "number_of_recharges__allweek__night__std": 0.9428090415820634,
    "number_of_recharges__allweek__night__mean": 1.6666666666666667,
    "average_balance_recharges": 5.619047619047619
  }
}
```

bandicoot workflow > loading data

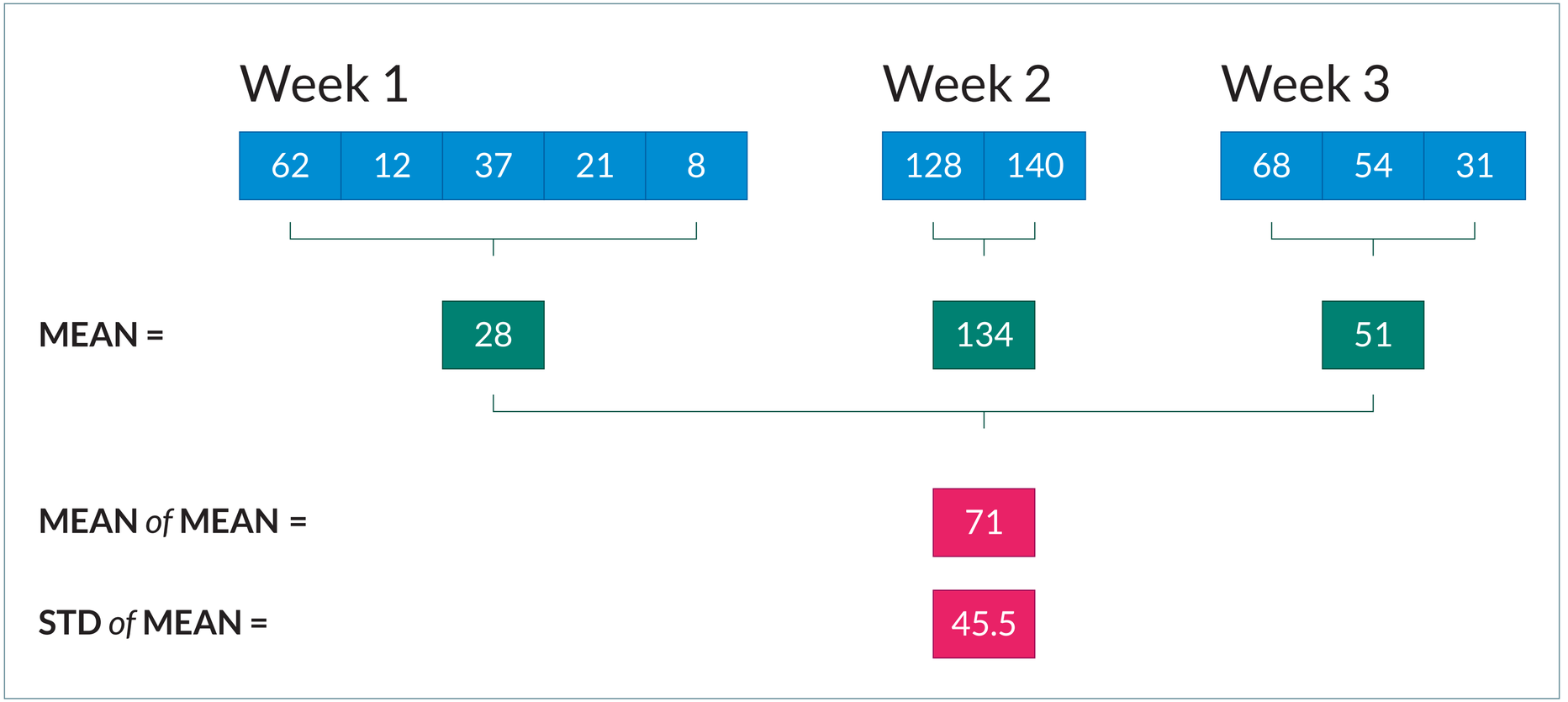
```
>>> bc.read_csv('A', '/data/records/', '/data/antennas.csv', '/data/attributes/')
```



bandicoot workflow > computations



bandicoot workflow > computations



Aggregating the indicator “call_duration”

dive into bandicoot > notebook demo

Notebook added and removed here! :-)

Replicability

Standardized open-source toolbox to analyze mobile phone metadata

Same metrics, same code for all
Tested and robust source code

Privacy

Less sensitive information than raw data
Limits re-identification and inference risks

Real-life use cases

Use cases > predicting behaviors


```
[…]

"number_of_interaction_out__weekend__night__call": 48,
"number_of_interaction_out__weekend__night__text": 54,
"number_of_antennas__allweek__allday": 7,
"number_of_antennas__allweek__day": 7,
"number_of_antennas__allweek__night": 7,
"number_of_antennas__weekday__allday": 7,
"number_of_antennas__weekday__day": 7,
"number_of_antennas__weekday__night": 7,
"number_of_antennas__weekend__allday": 7,
"number_of_antennas__weekend__day": 7,
"number_of_antennas__weekend__night": 7,
"entropy_of_antennas__allweek__allday": 1.9419867718126902,
"entropy_of_antennas__allweek__day": 1.902707353033661,
"entropy_of_antennas__allweek__night": 1.9414986247438955,
"entropy_of_antennas__weekday__allday": 1.942101476455784,
"entropy_of_antennas__weekday__day": 1.8934579035596322,
"entropy_of_antennas__weekday__night": 1.9426970127693166,
"entropy_of_antennas__weekend__allday": 1.9376901490208147,
"entropy_of_antennas__weekend__day": 1.822553031185579,
"entropy_of_antennas__weekend__night": 1.9258841213419076,
"percent_at_home__allweek__allday": 0.1563467492260062,
"percent_at_home__allweek__day": 0.11428571428571428,
"percent_at_home__allweek__night": 0.1645101663585952,
"percent_at_home__weekday__allday": 0.15711252653927812,
"percent_at_home__weekday__day": 0.13157894736842105,
"percent_at_home__weekday__night": 0.1620253164556962,
"percent_at_home__weekend__allday": 0.15428571428571428,
"percent_at_home__weekend__day": 0.06896551724137931,
"percent_at_home__weekend__night": 0.17123287671232876,
"radius_of_gyration__allweek__allday": 1.1261249033689014,
"radius_of_gyration__allweek__day": 1.1103898231623104,
"radius_of_gyration__allweek__night": 1.1286971424946113,
"radius_of_gyration__weekday__allday": 1.1240418740290543,
"radius_of_gyration__weekday__day": 1.1605772518571211,
"radius_of_gyration__weekday__night": 1.1166371509744721,
"radius_of_gyration__weekend__allday": 1.131521050714622,
"radius_of_gyration__weekend__day": 0.9502406200901717,
"radius_of_gyration__weekend__night": 1.1591646504866928,
"frequent_antennas__allweek__allday": 6,
"frequent_antennas__allweek__day": 5,

[…]
```


Gender
Age
Socio-economic status
Employment status
Personality
…

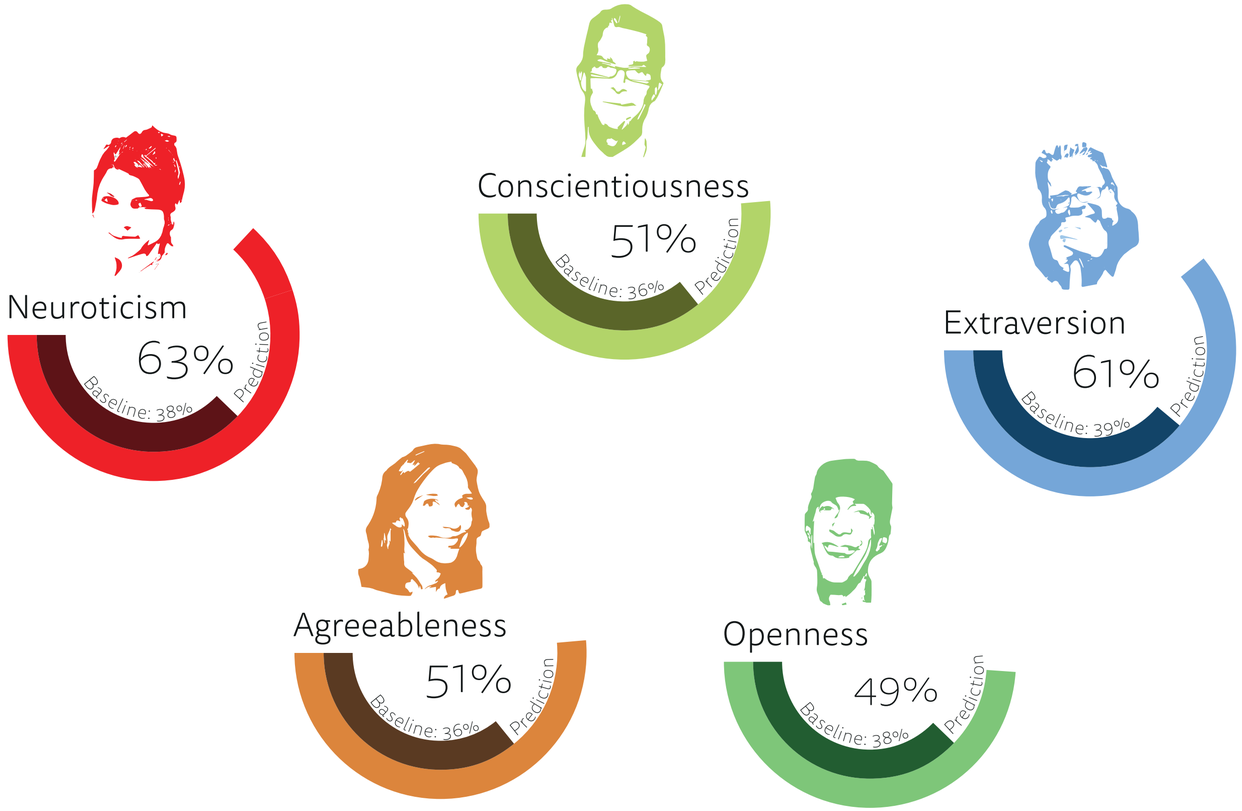
?

Use cases > predicting personality

I see myself as someone who...

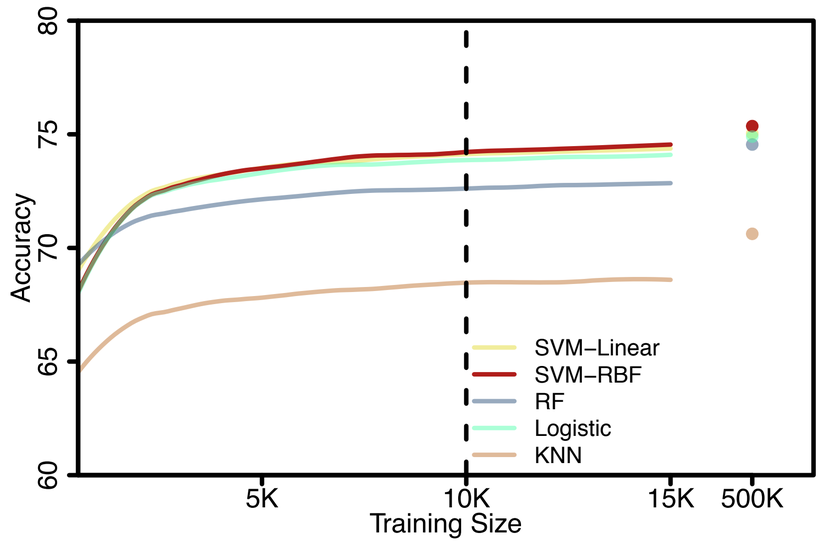
Is talkative (01) _____
Tends to find fault with others (02) _____
Does a thorough job (03) _____
Is depressed, blue (04) _____
Is original, comes up with new ideas (05) _____
Is full of energy (06) _____
Is helpful and unselfish with others (07) _____
Is a reliable worker (08) _____
Is relaxed, handles stress well (09) _____
Is curious about many different things (10) _____
Has an assertive personality (11) _____
Can be cold and aloof (12) _____
Tends to be disorganized (13) _____
Can be tense (14) _____
Is ingenious, a deep thinker (15) _____
Is sometimes shy, inhibited (16) _____
Is considerate and kind to almost everyone (17) _____
Perseveres until the task is finished (18) _____
Worries a lot (19) _____
Has an active imagination (20) _____
Is outgoing, sociable (21) _____
Is sometimes rude to others (22) _____
Does things efficiently (23) _____
Is emotionally stable, not easily upset (24) _____
Values artistic, aesthetic experiences (25) _____
Tends to be quiet (26) _____
Likes to cooperate with others (27) _____
Makes plans and follows through with them (28) _____
Gets nervous easily (29) _____
Likes to reflect, play with ideas (30) _____

Use cases > predicting personality



de Montjoye, Y. A., Quoidbach, J., Robic, F., & Pentland, A. S. (2013). Predicting personality using novel mobile phone-based metrics. In Social computing, behavioral-cultural modeling and prediction (pp. 48-55). Springer Berlin Heidelberg.

Use cases > predicting gender



Jahani, E., Sundsøy, P. R., Bjelland, J., Iqbal, A., Pentland, A., & de Montjoye, Y. A. Predicting Gender from Mobile Phone Metadata. 2015

bandicoot.mit.edu

5900 lines of code and 10 contributors

1442 indicators

Python 2 & 3

In collaboration with Yves-Alexandre de Montjoye, Alex ‘Sandy’ Pentland,
Florent Robic, Eaman Jahani, Brian Sweatt & many others

Use cases > Orange D4D Challenge

150 teams of researchers in 2014 working on transportation, healthcare, agriculture, energy...

One year of data from Senegal

1. **aggregated mobility** data sets
2. **bandicoot indicators** for 300,000 customers