


Peer production of community science with personal data

Bastian Greshake Tzovaras
@gedankenstuecke
2019-08-18

Universal Declaration of Human Rights



*Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to **share in scientific advancement and its benefits.***

Mertonian Norms

- **communism:** all scientists should have *common ownership* of scientific goods (intellectual property), to promote *collective collaboration*; secrecy is the opposite of this norm.
- **universalism:** scientific validity is independent of the sociopolitical status/ personal attributes of its participants

Citizen/Community Science



- an extension of
 - the Mertonian Norms?
 - the Declaration of Human Rights?

Bonney et al. Can citizen science enhance public understanding of science? (2016)

Greshake Tzovaras et al. The Personal Data is Political (2019)

Citizen/Community Science



Are citizen science participants actually taking part in scientific advancement?

prevalent CCS types






- Crowdfunding
- Human Computation
- Crowdsourcing

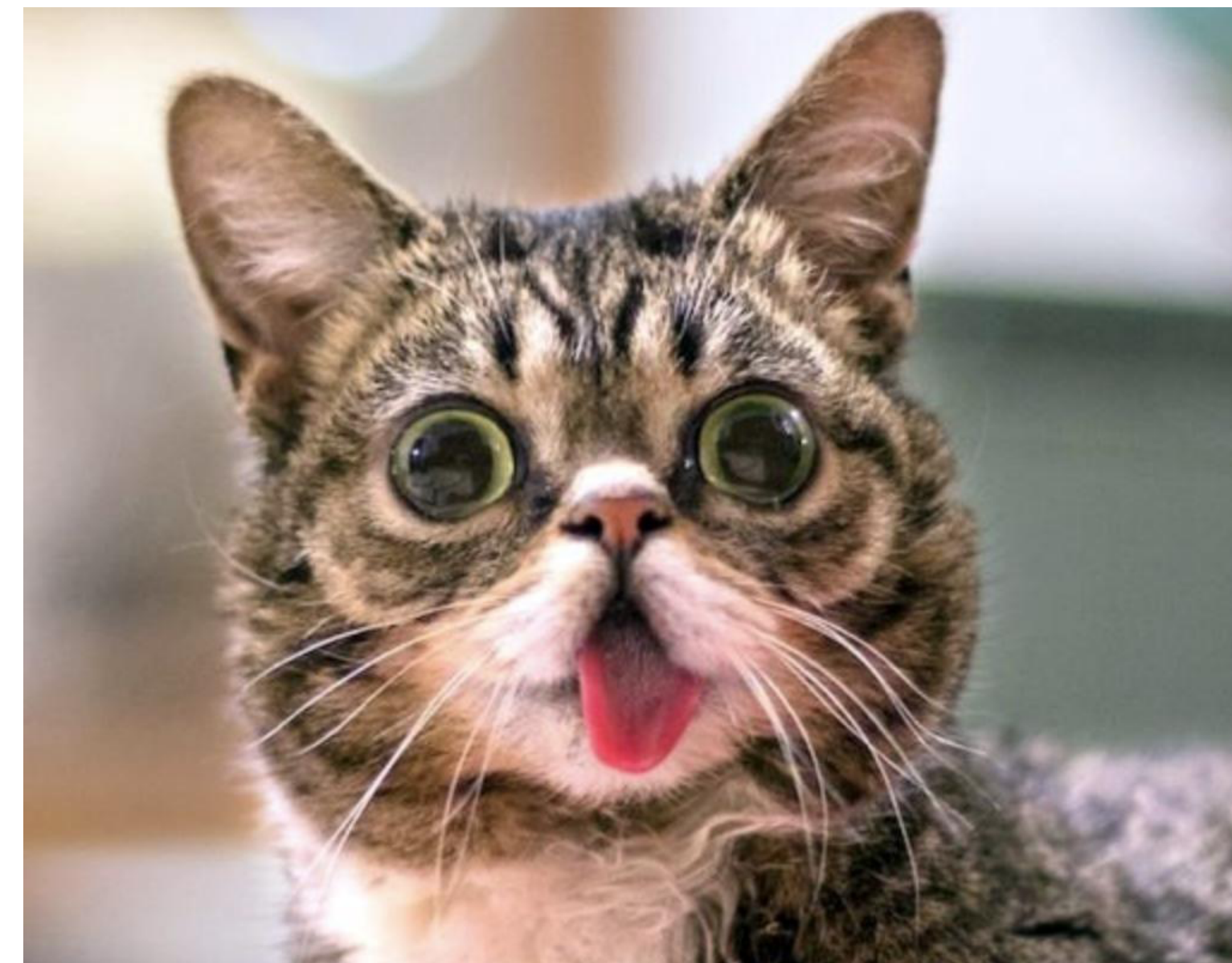
Bonney et al. Can citizen science enhance public understanding of science? (2016)

Hecker et al. The European citizen science landscape – a snapshot (2019)

prevalent CCS types

Crowdfunded whole-genome sequencing of the celebrity cat Lil BUB identifies causal mutations for her osteopetrosis and polydactyly

Mike Bridavsky, Heiner Kuhl, Arthur Woodruff, Uwe Kornak, Bernd Timmermann, Norbert Mages, 99 Lives Consortium,  Darío G Lupiáñez,  Orsolya Symmons,  Daniel M Ibrahim
doi: <https://doi.org/10.1101/556761>



prevalent CCS types

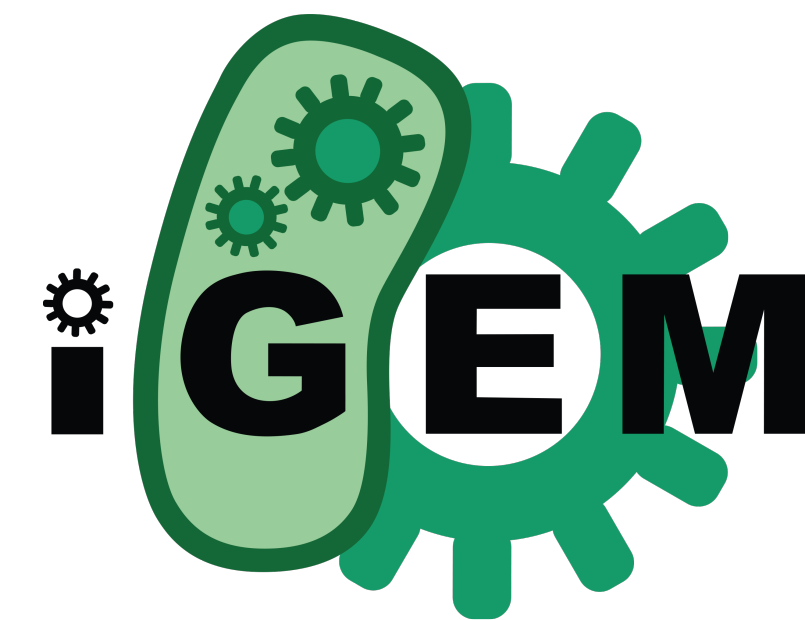


ZOO~~N~~IVERSE

prevalent CCS types

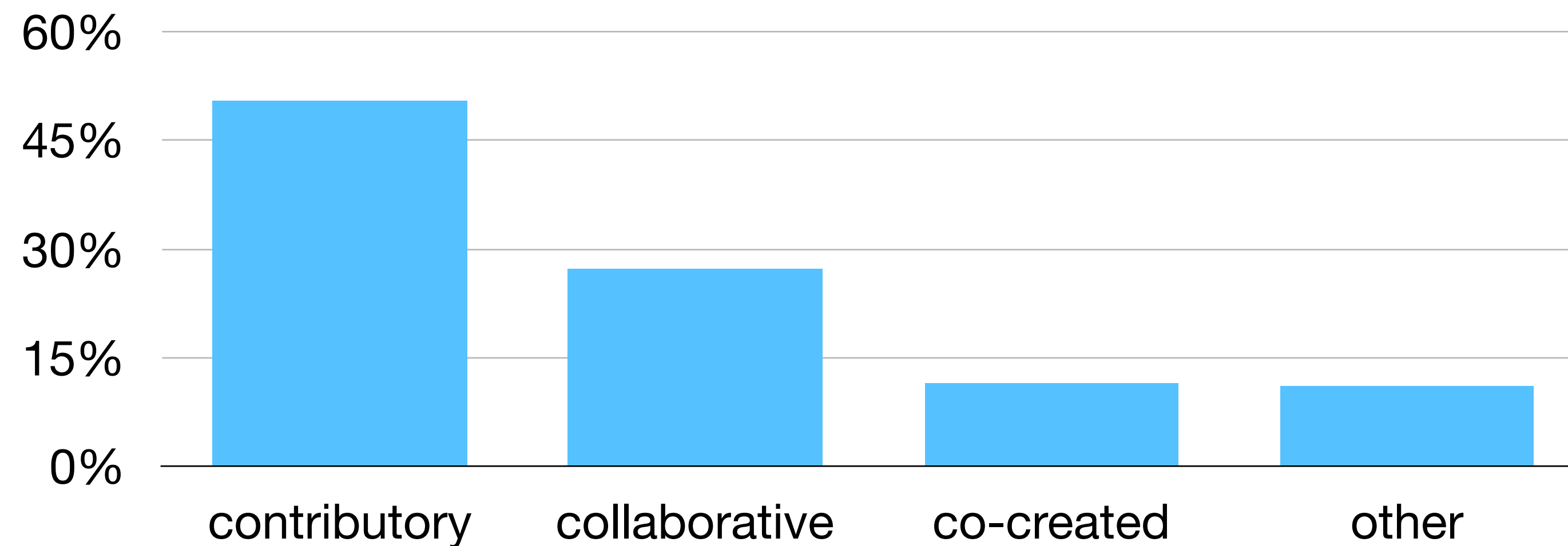
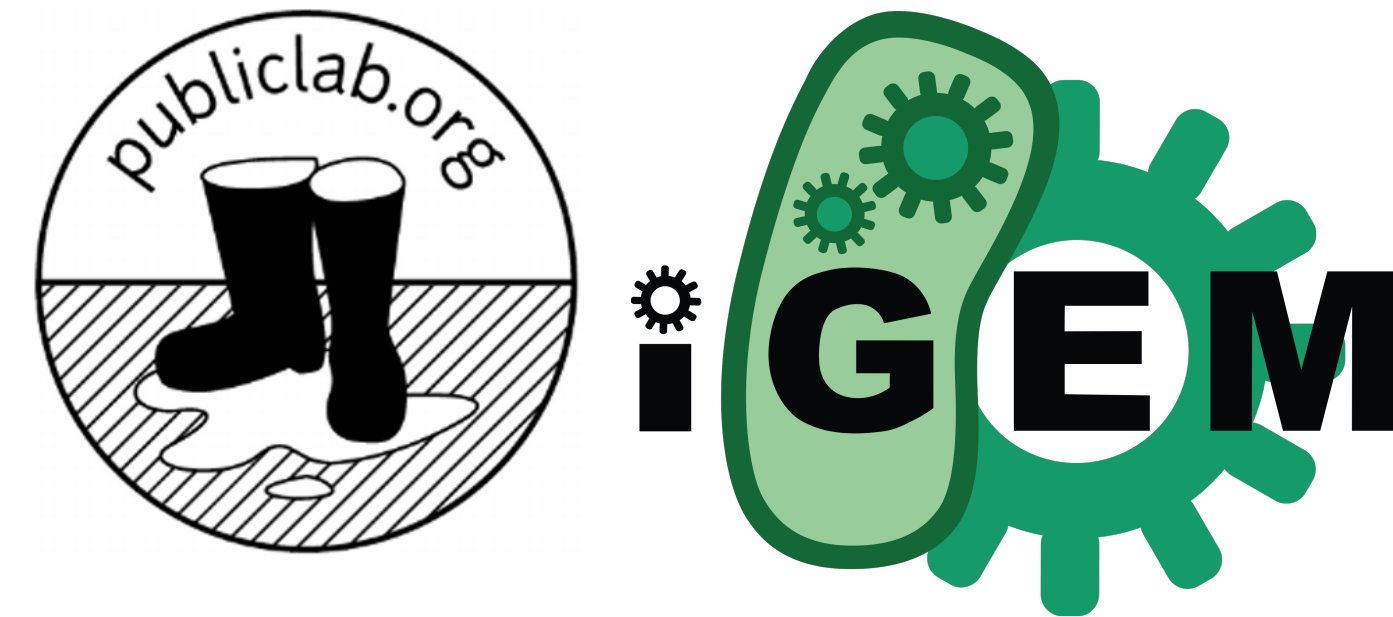


co-created community science



Bonney et al. Can citizen science enhance public understanding of science? (2016)
Hecker et al. The European citizen science landscape – a snapshot (2019)

co-created community science



Bonney et al. Can citizen science enhance public understanding of science? (2016)

Hecker et al. The European citizen science landscape – a snapshot (2019)

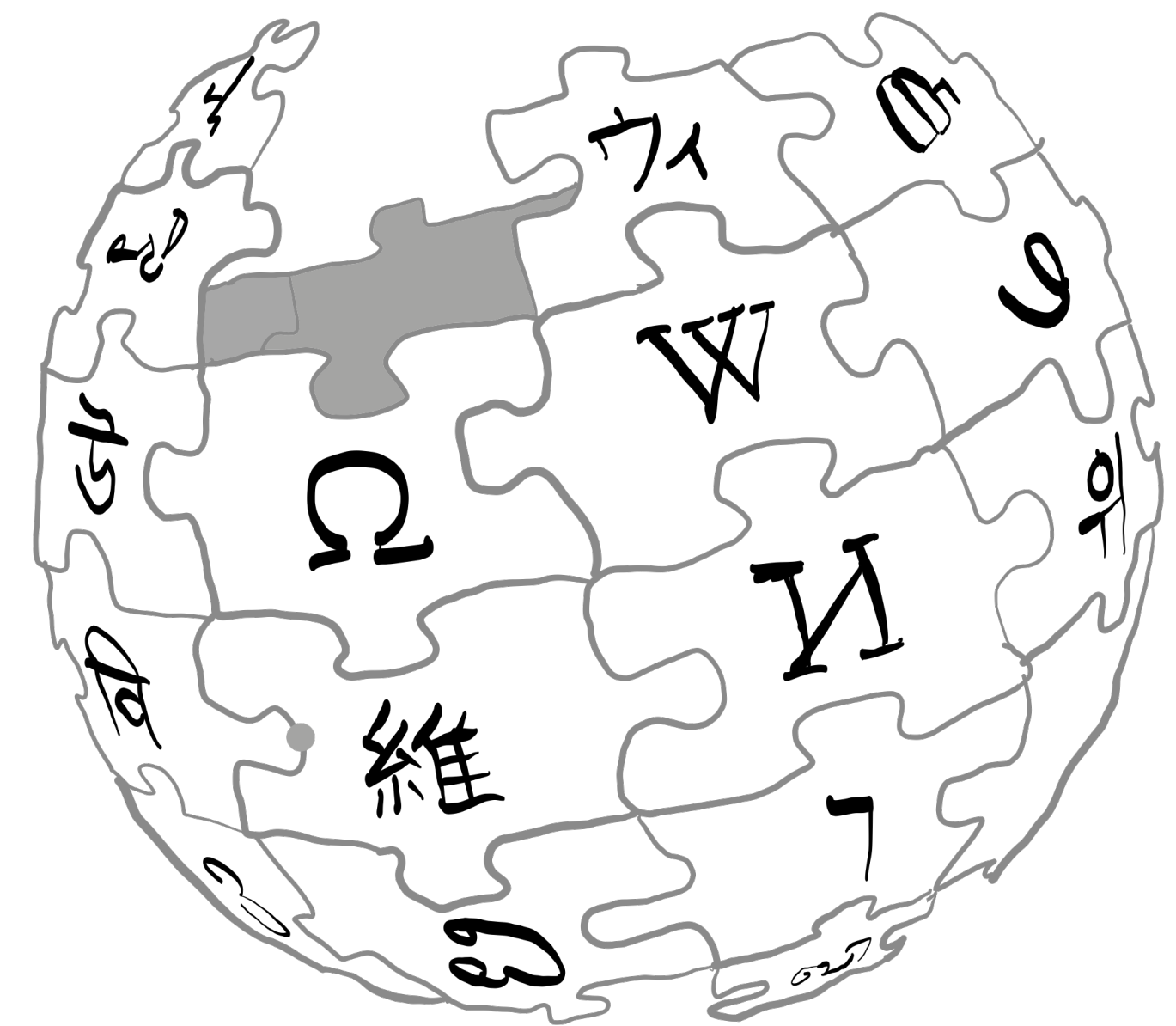
**true co-creation & peer-production is rare inside
research, but outside research...**

co-creation & peer-production



open source

**non-hierarchical
negotiated coordination
for-benefit**



Wikipedia

co-creation & peer-production *& community science*

- *divide researcher / volunteers*
- *top-down structure*
- *one-sided data flow*

co-creation & peer-production & community science

- divide researcher / volunteers
- top-down structure
- one-sided data flow



Clickworkers
Classic

Mertonian Norms



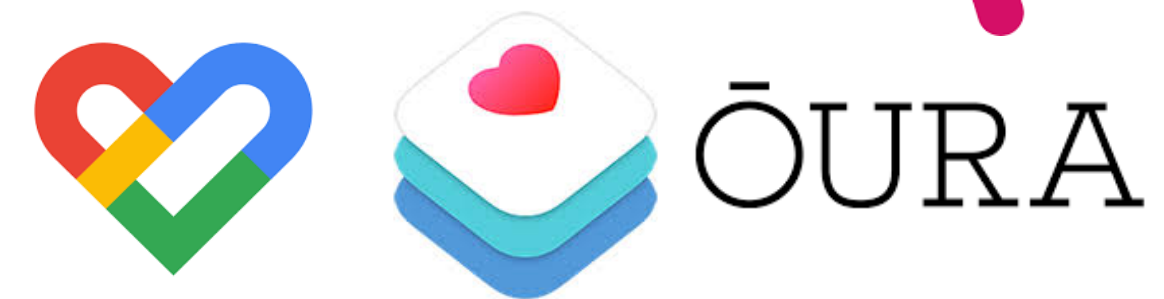
- **communism: all scientists** should have *common ownership* of scientific goods (intellectual property), to promote *collective collaboration*; secrecy is the opposite of this norm.
- **universalism:** scientific validity is independent of the sociopolitical status/ personal attributes of its participants

Can everyone be a scientist?

Can everyone be a scientist?

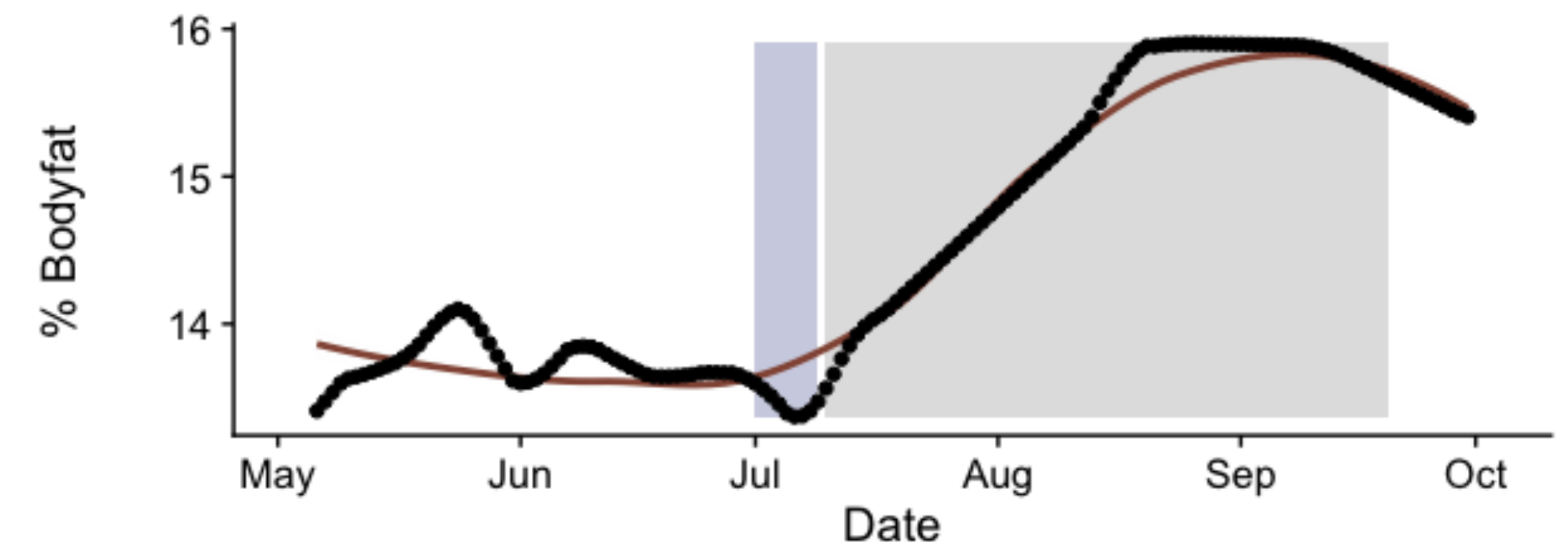
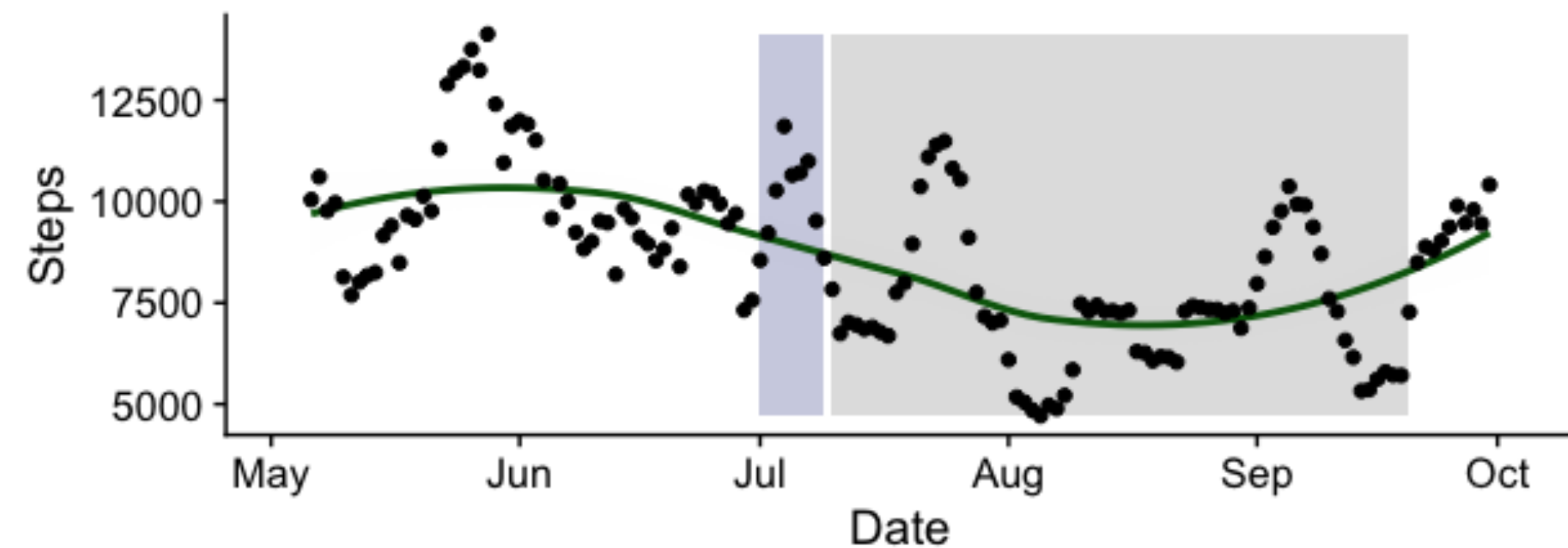
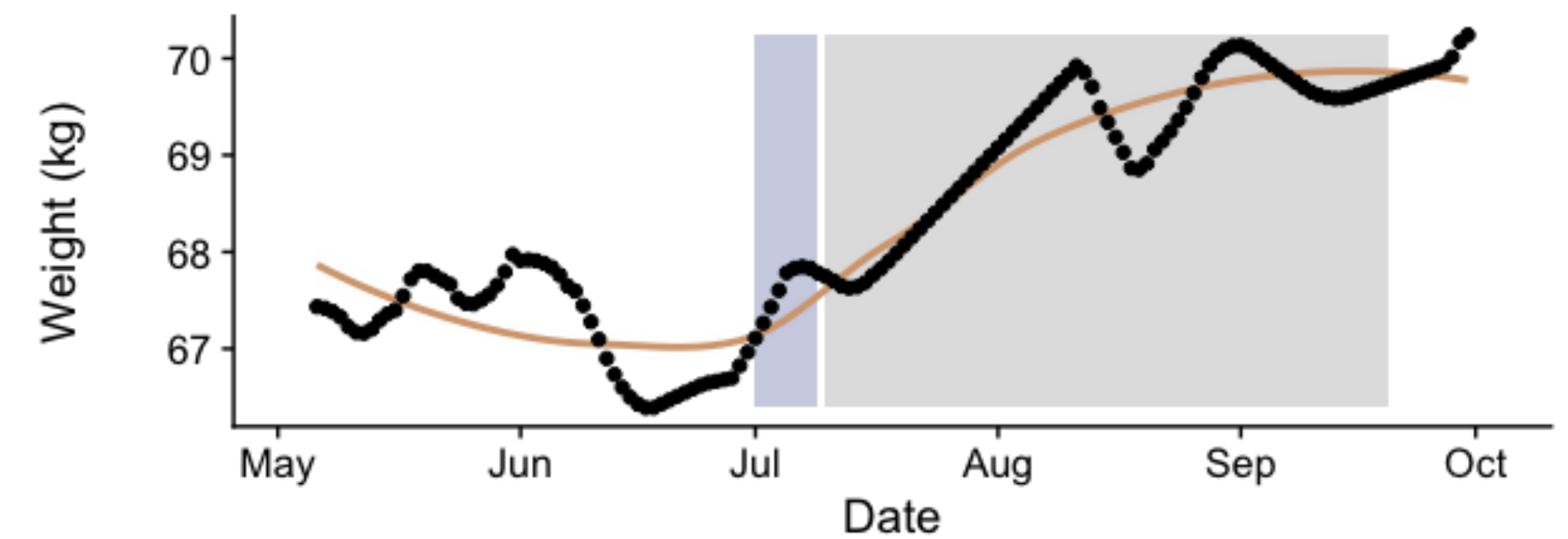
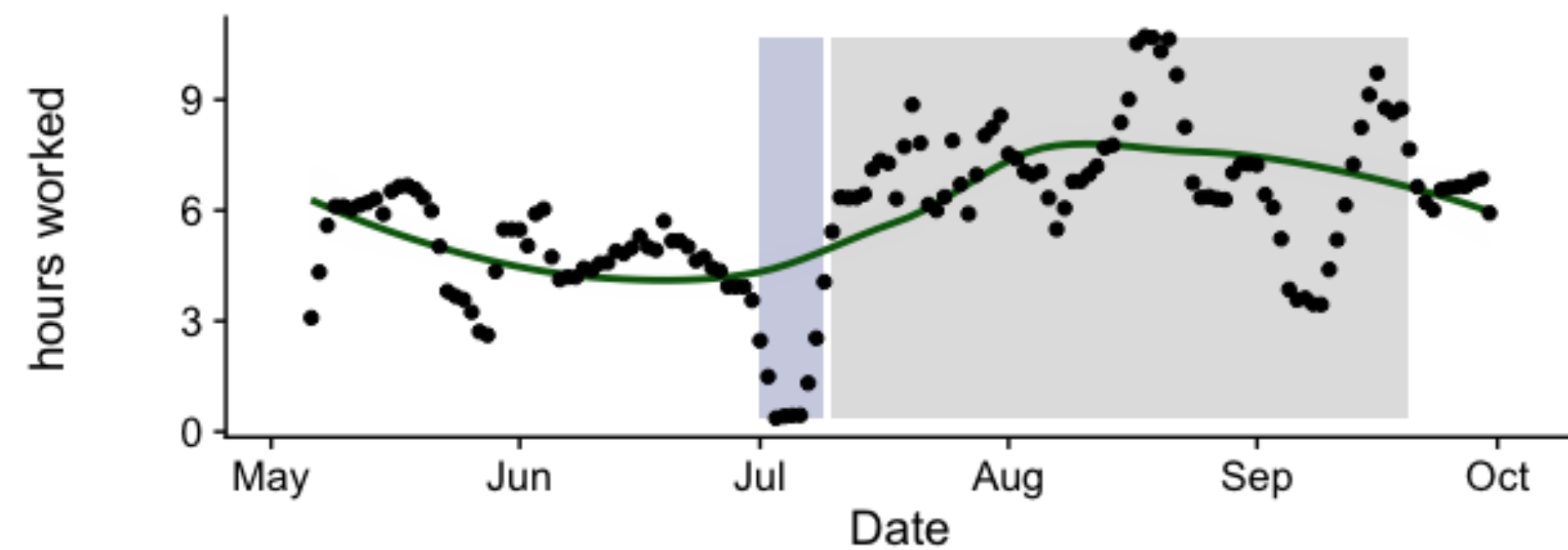
QUANTIFIED SELF

SELF KNOWLEDGE THROUGH NUMBERS



Can everyone be a scientist?

before, during (gray box) and after PhD thesis writing
(points are rolling average over 7 days)



Going from *Quantified Self* to *Quantified Us*?



OPEN HUMANS

Greshake Tzovaras et al. Open Humans: A platform for participant-centered research and personal data exploration (2018)

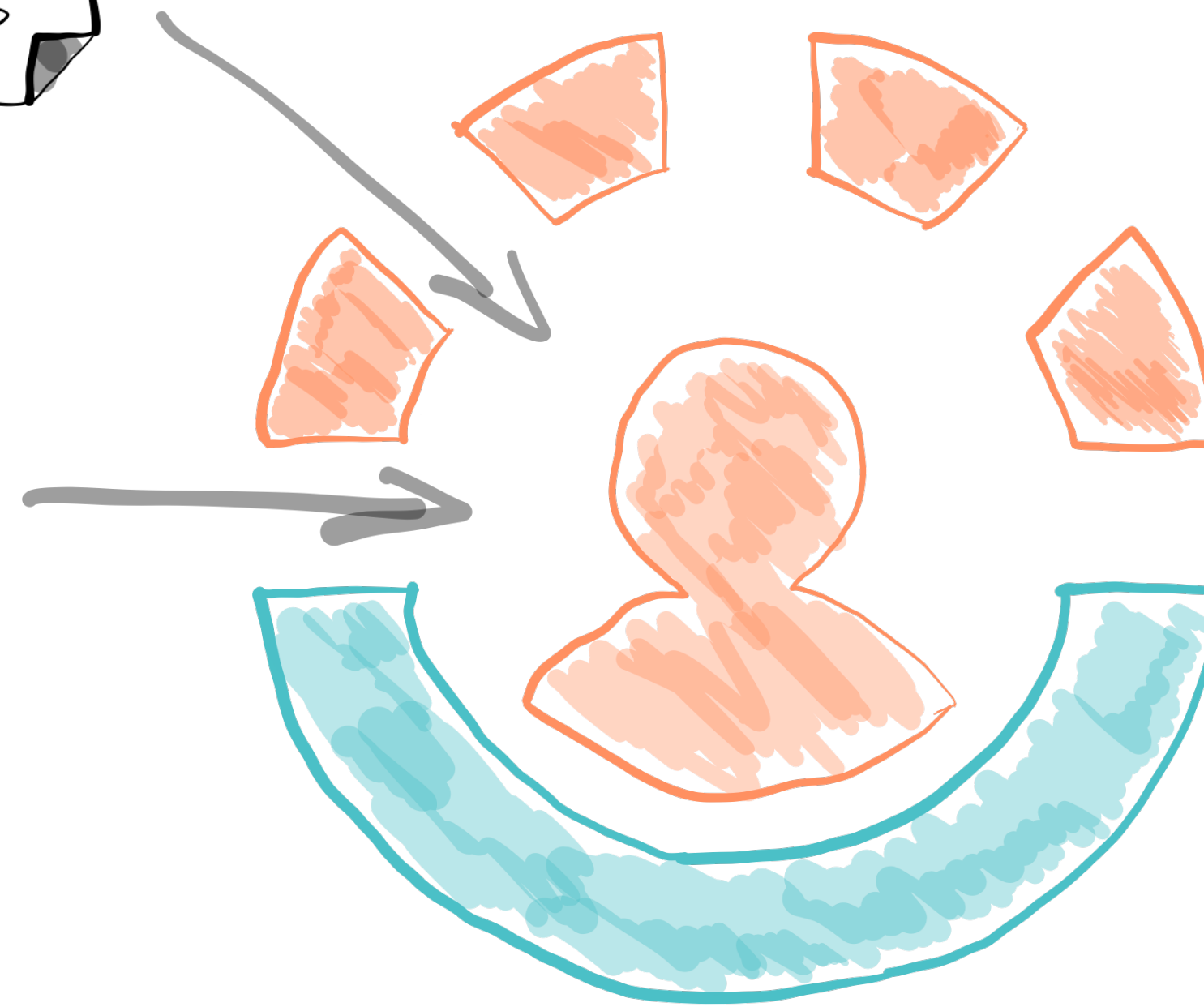
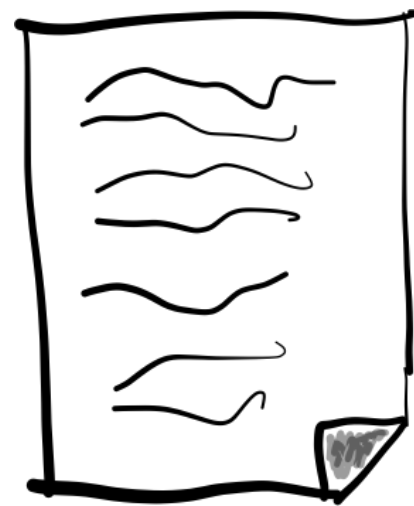


OPEN HUMANS

Greshake Tzovaras et al. Open Humans: A platform for participant-centered research and personal data exploration (2018)



individual imports
personal data,



OPEN HUMANS

Greshake Tzovaras et al. Open Humans: A platform for participant-centered research and personal data exploration (2018)

Data Sources



23andMe Upload
Open Humans
Open Humans Foundation
Connected by **1152** members

23andMe is a direct-to-consumer genetic testing company that tests about one million genetic locations.

[Learn more](#)



Twitter Archive Analyzer
Open Humans
Bastian Greshake Tzovaras
Connected by **519** members

The TwArxiv is a Twitter Archive Analyzer. Upload your Twitter archive and get new insights.

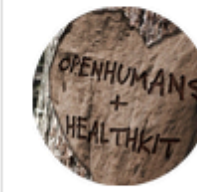
[Learn more](#)



AncestryDNA Upload
Open Humans
Open Humans Foundation
Connected by **411** members

Ancestry.com's AncestryDNA is a direct-to-consumer genetic testing product that tests about 700,000 genetic locations.

[Learn more](#)



Open Humans Healthkit Integration
James Turner
self
Connected by **183** members

Install this third-party app on your iPhone or iPad to upload HealthKit data to Open Humans.

[Learn more](#)



Data Selfies
Open Humans
Open Humans Foundation
Connected by **124** members
Upload Data Selfies

[Learn more](#)



Nokia Health (Withings) Connection
Open Humans
Open Humans Foundation
Connected by **101** members

Add your Nokia Health (Withings) data to Open Humans

[Learn more](#)



Gencove
Joseph K. Pickrell
Gencove, Inc.
Connected by **295** members

Your genome app - get your ancestry, microbiome, and more! Contribute your data to OpenHumans.

[Learn more](#)



Fitbit Connection
Open Humans
Open Humans Foundation
Connected by **398** members

Connect your Fitbit account to add data from your Fitbit activity trackers and other Fitbit devices.

[Learn more](#)



openSNP
Bastian Greshake Tzovaras
openSNP
Connected by **290** members

openSNP allows you to put your genetic and phenotypic data into the public domain. Connect your openSNP account to Open Humans.

[Learn more](#)



uBiome Upload
Open Humans
Open Humans Foundation
Connected by **69** members

uBiome is a company based in San Francisco that performs microbiome sequencing

[Learn more](#)



FamilyTreeDNA integration
Bastian Greshake Tzovaras
Connected by **74** members
Upload your FamilyTreeDNA data to Open Humans

[Learn more](#)



Imputer
Kevin Arvai
Connected by **55** members

Imputer will perform imputation on user's genotype data to estimate genotypes at missing loci.

[Learn more](#)



Runkeeper connection
Open Humans
Open Humans Foundation
Connected by **210** members

RunKeeper is a free smartphone app for GPS fitness-tracking. You can use it to record GPS timepoint data for runs, walks, bicycling, and oth

[Learn more](#)



Nightscout Data Transfer
James Wedding, P.E.
The Nightscout Foundation
Connected by **280** members

A tool to easily enable upload of data from individual Nightscout databases to the Open Humans platform

[Learn more](#)



Genome/Exome Upload
Open Humans
Open Humans Foundation
Connected by **209** members

Do you have genome or exome data? You can upload genome, exome, and genotyping data in VCF format.



Overland connection
Open Humans
Open Humans Foundation
Connected by **15** members

Import GPS data from Overland

[Learn more](#)



Spotify integration
Open Humans
Open Humans Foundation
Connected by **47** members
Collect the songs you listen to on Spotify

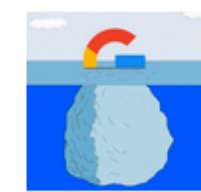
[Learn more](#)



RescueTime connection
Bastian Greshake Tzovaras
Connected by **31** members

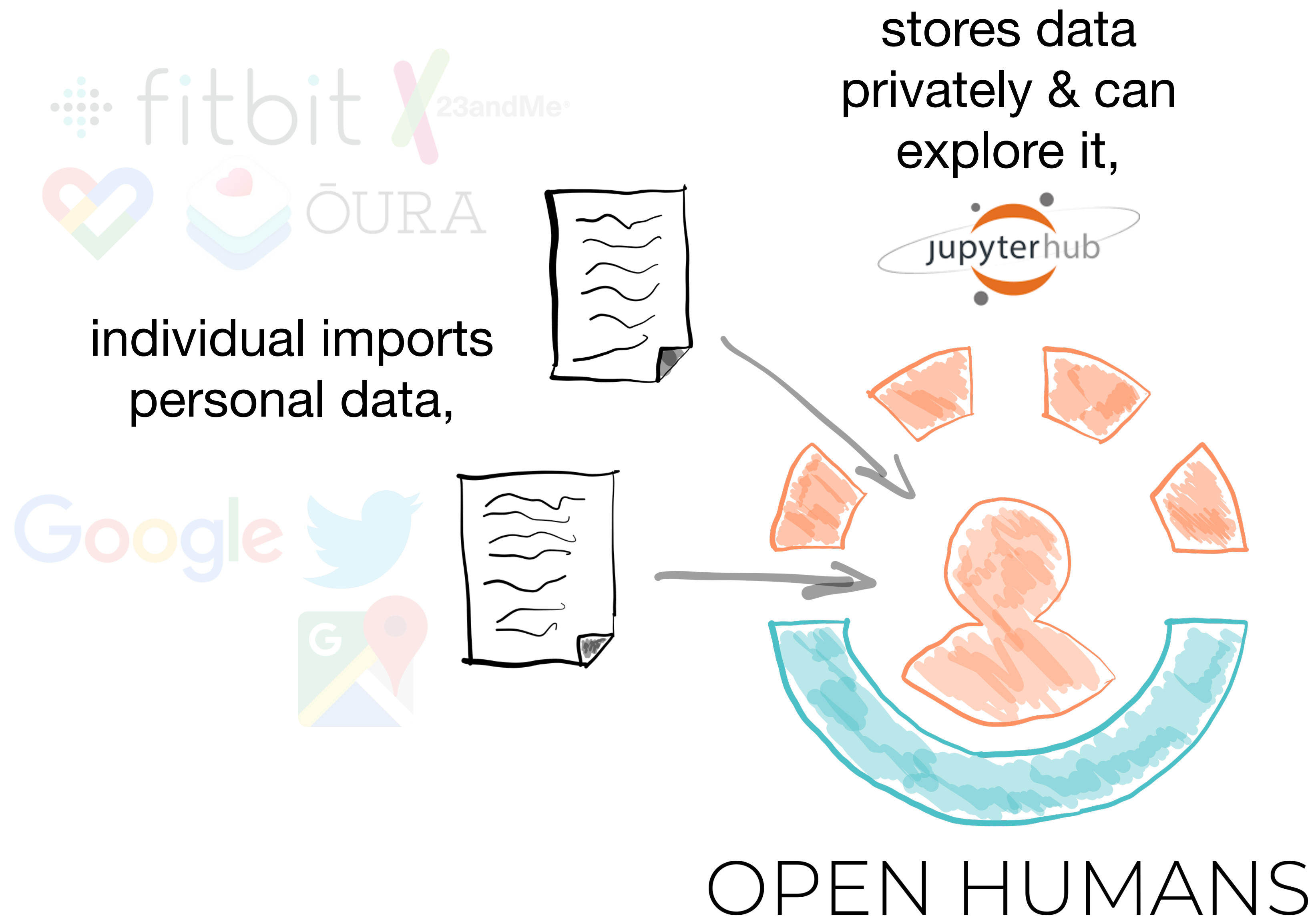
RescueTime is a web service tracks your productivity and computer usage over time.

[Learn more](#)



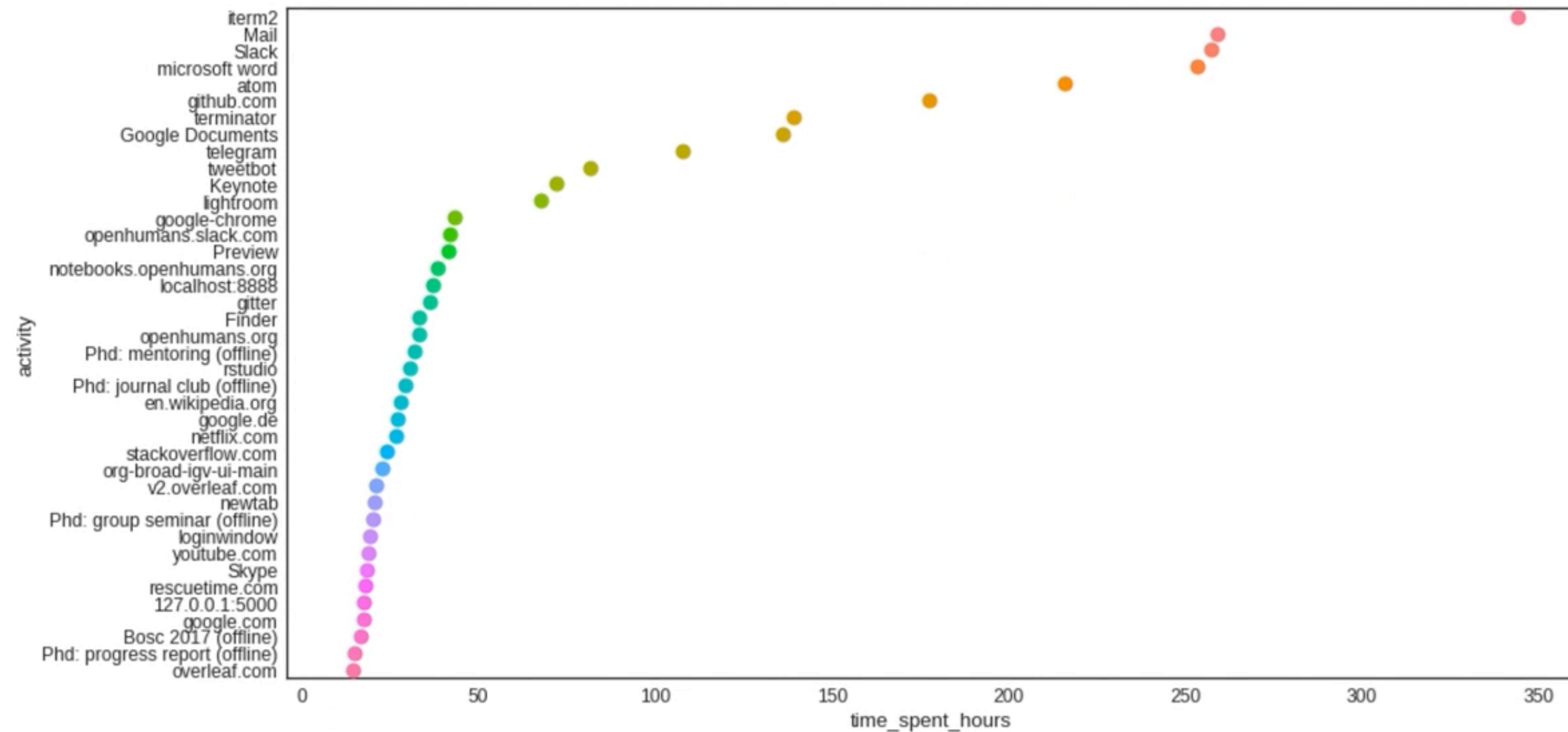
Google search history analyzer
Athina Tzovara
Connected by **12** members

Google search history analyzer



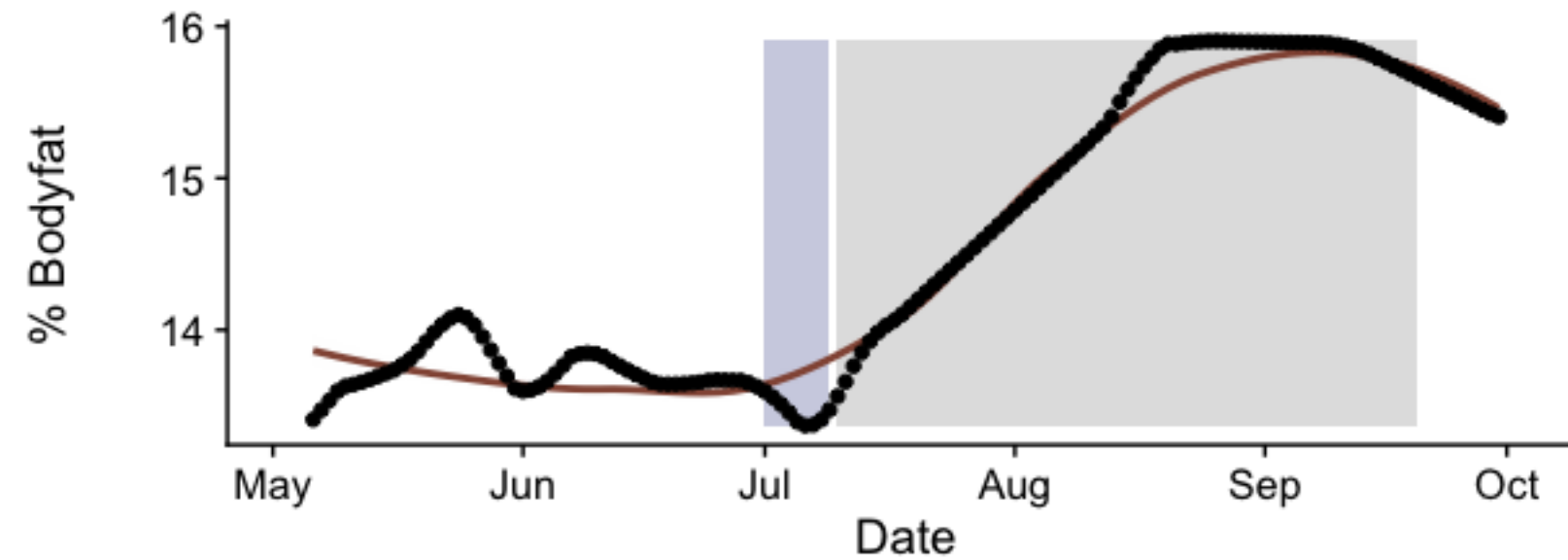
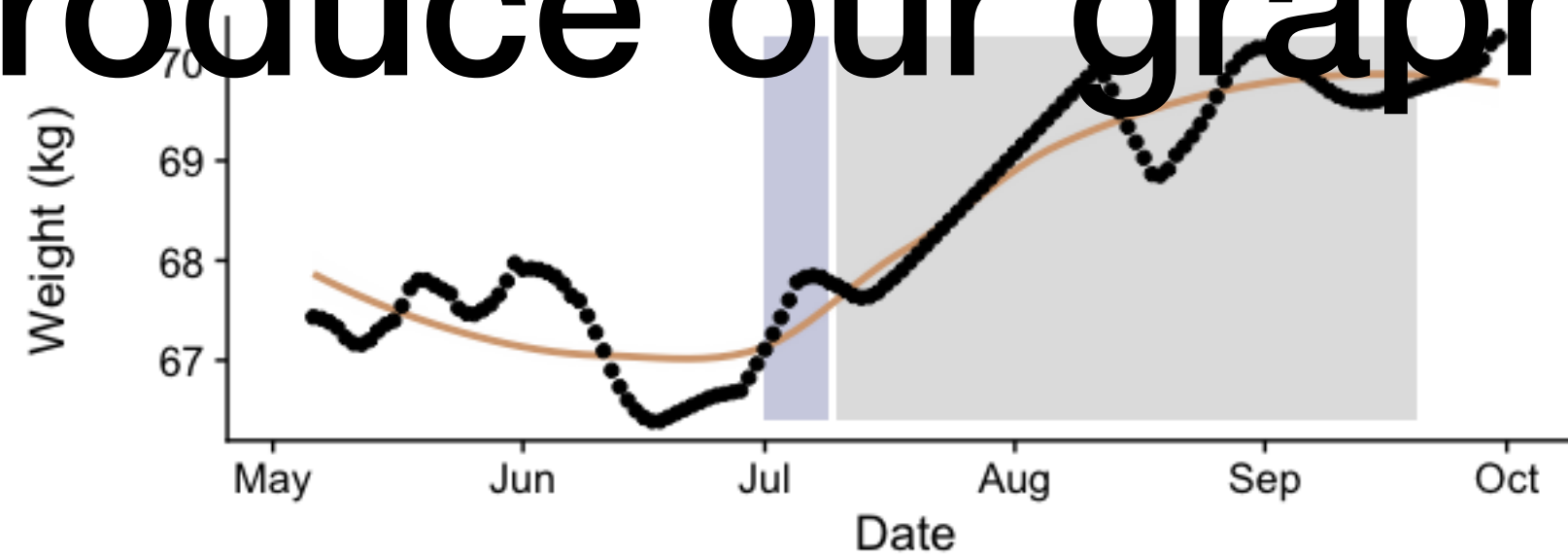
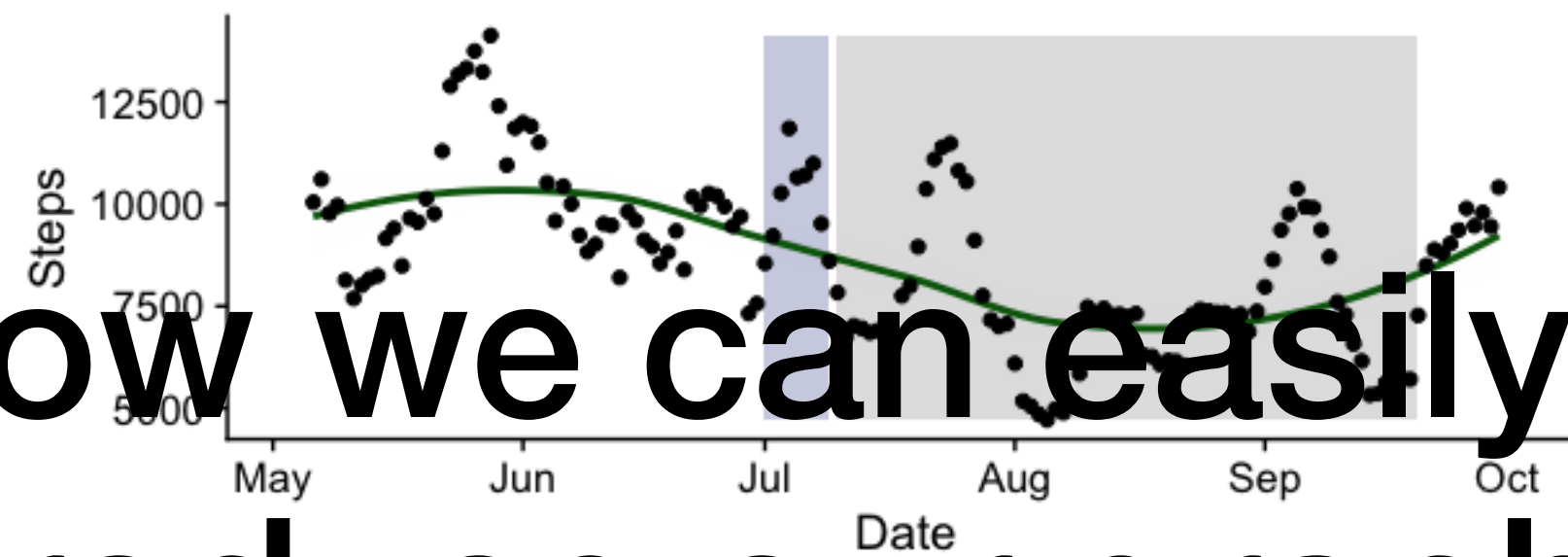
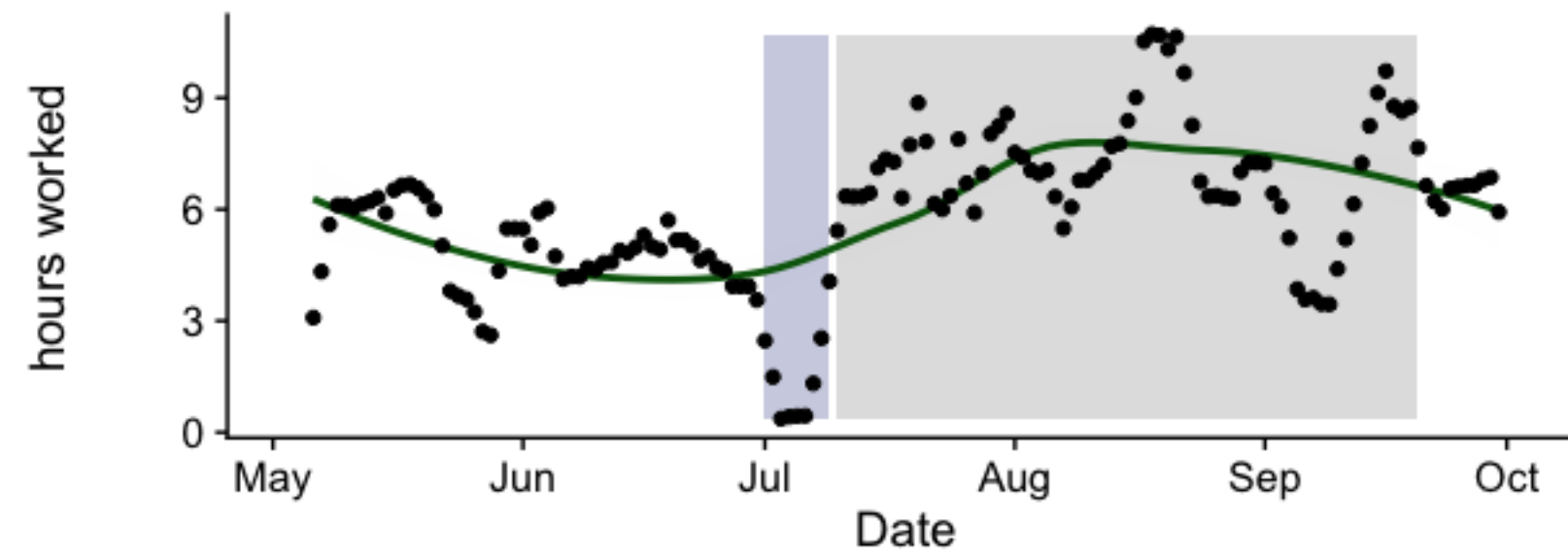
Greshake Tzovaras et al. Open Humans: A platform for participant-centered research and personal data exploration (2018)


```
plt = sns.stripplot(y='activity',  
                  x='time_spent_hours',  
                  data=summed_activity_time.sort_values(by=['time_spent_hours'], ascending=False)[:40],  
                  size=10)
```

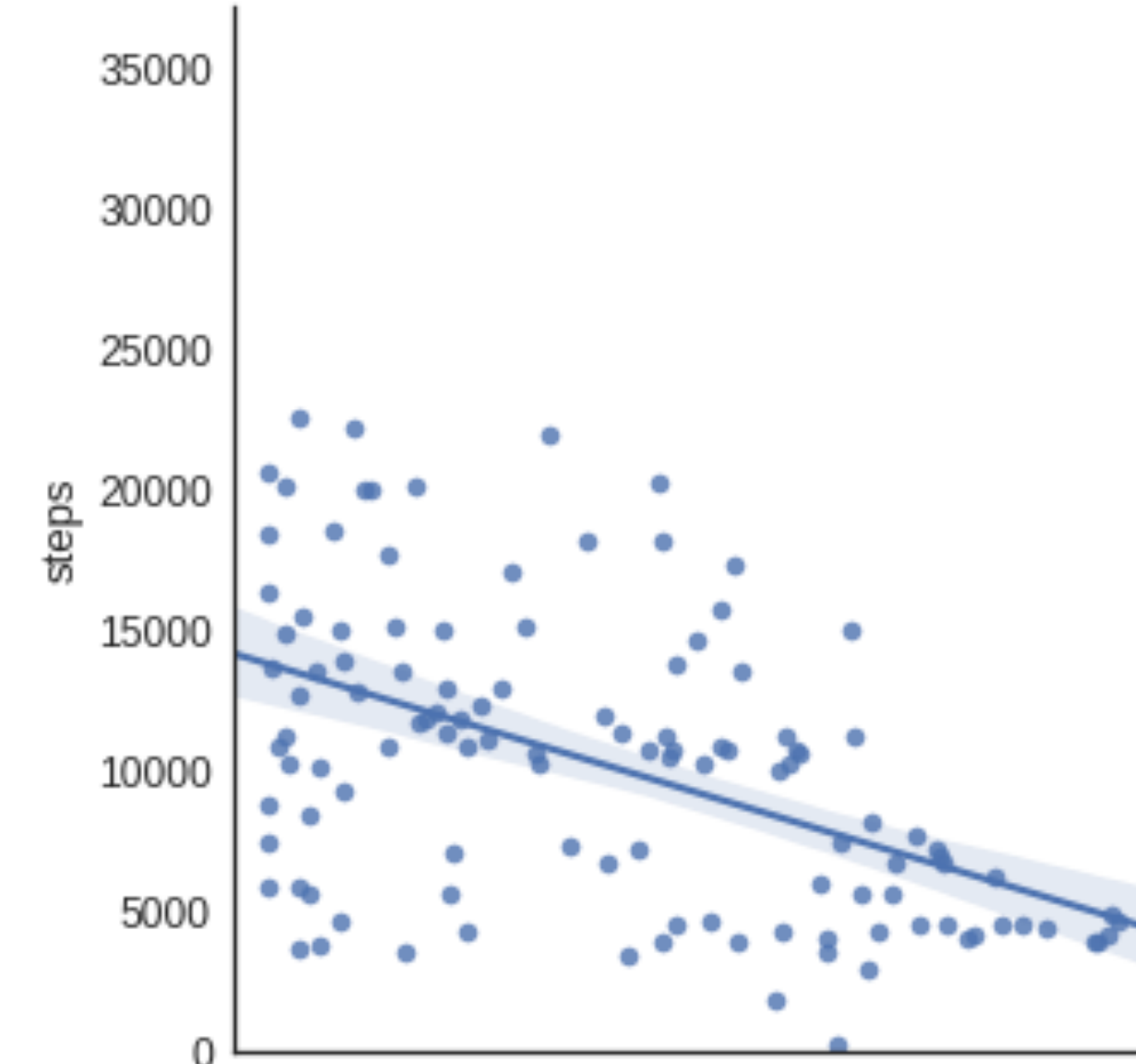


`item2` and `terminator` are my two command line tools of choice, while `MS Word` is clearly in second place due to the inclusion of my PhD writing time. `atom` as my current favorite text editor for programming and the high ranking of `github` is also expected. Similarly, lots of time spent on eMail & Slack aren't surprising either, given that much of my work is done remotely with collaborators across the globe.

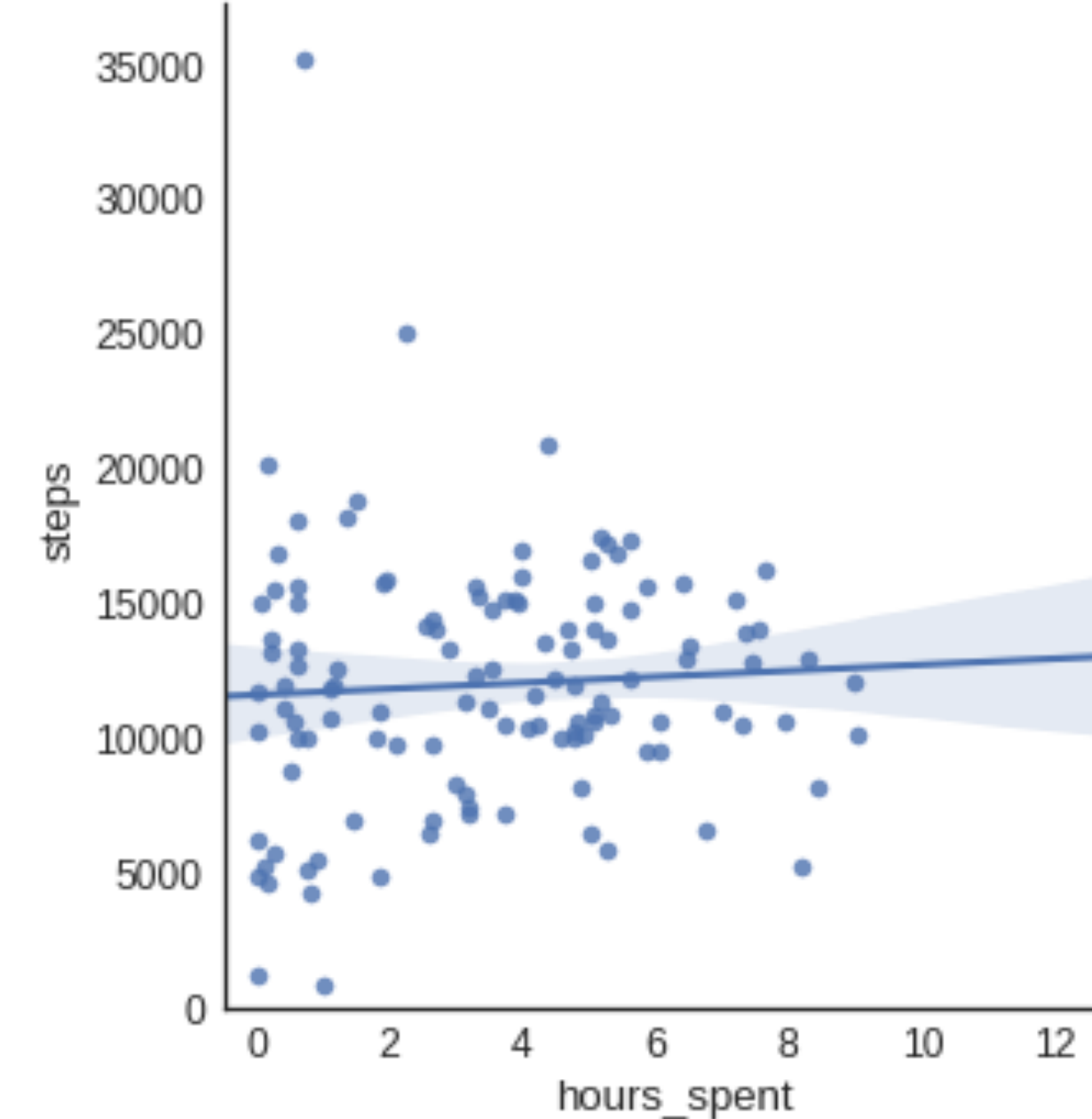
before, during (gray box) and after PhD thesis writing
(points are rolling average over 7 days)



year = 2017



year = 2018



Now we can easily
reproduce our graphs

rescuetime-vs-step-counts-2019-05-01-22:14:17 (autosaved)

LogoutControl Panel

FileEditViewInsertCellKernelWidgetsHelp

New Notebook

Open...

Make a Copy...

Rename...

Save and Checkpoint

Revert to Checkpoint

Print Preview

Download as

Share Notebook

Trust Notebook

Close and Halt

Run

Code

13.042686893113581)

year = 2017

Upload to Open Humans

```
plot(data=joined_data,
      y='steps',
      x='hours_spent',
      row='year')
t.ylim(-0.5, None)
t.xlim(-0.5, None)
```

https://notebooks.openhumans.org/user/gedankenstuecke/notebooks/rescuetime-vs-step-counts-2019-05-01-22:14:17.ipynb#

day

Highlight AllMatch CaseWhole Words4 of 19 matches

Juno's Personal Data Exploratory

https://exploratory.openhumans.org/notebook/3/

120%

Log In

Juno's Exploratory

Explore

Search Notebook

Submit

About

Details for **rescuetime-vs-step-counts.ipynb**

Open

Published by [gedankenstuecke](#)

Description

How do your productivity and screen time influence your physical activity? Find out by looking at your **RescueTime** and **Fitbit** or **Apple Healthkit** data.
Important: You only need one of the activity trackers to use this notebook.

101

1

Tags & Data Sources

activity tracking

productivity tracking

work life balance

RescueTime connection

Fitbit Connection

Open Humans Healthkit Integration

Open in Personal Data Notebooks

Comments

Please [log in to comment](#).

Notebook

Last updated 10 months, 4 weeks ago

Using **RescueTime** to analyze your productivity and computer usage

RescueTime is a service that monitors how you're using your computer. Split into 5 minute intervals you can't only see what you did with your computer, but it also classifies these activities automatically in productive/unproductive time.
Let's use this data to see

- How our productivity changes over time
- How we're spending most of our computer time

and if we're also counting steps with **Fitbit** and/or **Apple HealthKit** to see

- How our computer usage influences our activity

To get started let's find out which data sources we have on *Open Humans* and identify & load the data from **RescueTime**
In [1]:

```
import json
import os
```

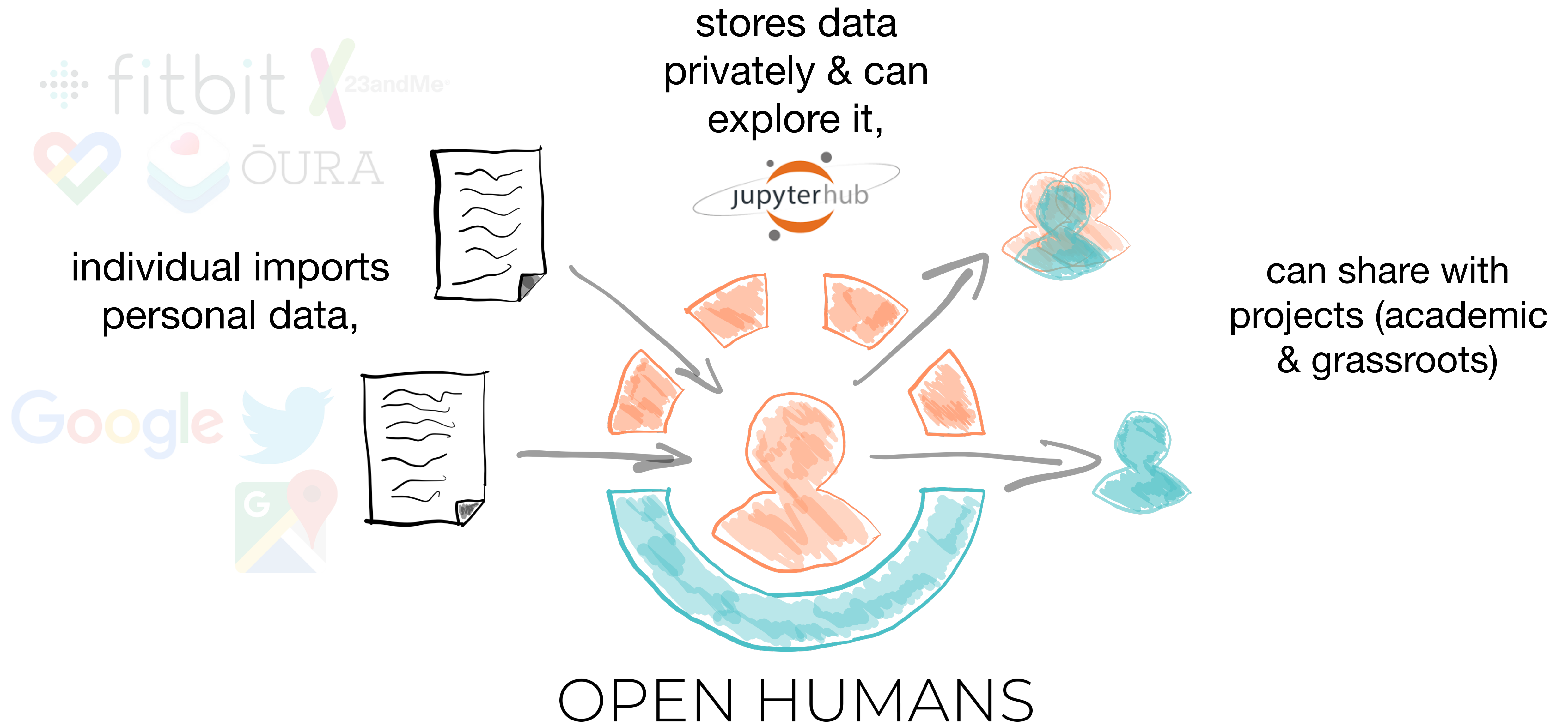
day

Highlight All

Match Case

Whole Words

4 of 19 matches



Greshake Tzovaras et al. Open Humans: A platform for participant-centered research and personal data exploration (2018)

Projects using Data



Genevieve Genome Report

Mad Ball

Connected by **807** members

Genevieve matches your genome against public variant data, and invites you to contribute to shared notes.

[Learn more](#)



Twitter Archive Analyzer

Bastian Greshake Tzovaras

Connected by **519** members

The TwArxiv is a Twitter Archive Analyzer. Upload your Twitter archive and get new insights.

[Learn more](#)



Keeping Pace

Dr. Rumi Chunara
New York University

Joined by **400** members

Keeping Pace seeks to study data about how we move around, to understand how seasons and local environment influence our movement patterns.

[Learn more](#)



Juno's Personal Data Exploratory

Bastian Greshake Tzovaras
Open Humans Foundation

Connected by **96** members

Showcase and share your Personal Data Notebooks with others

[Learn more](#)



UbiQomix Microbiome Exploration

Orit Shaer, Mad Price Ball,
Oded Nov
Wellesley College, Open Humans,
NYU

Joined by **123** members

This research survey study aims to gain insight into user needs as they engage with their personal and social microbiome data.

[Learn more](#)



Gut Instinct

Vineet Pandey
UCSD Design Lab

Connected by **93** members

Learn about current science of the gut microbiome with expert-curated learning materials.

[Learn more](#)



Circles

Cliff Tabin and Abigail Wark
Harvard Medical School

Joined by **316** members

Circles is a research study that aims to discover the genetic basis for a mysterious and remarkable human trait: the areola.

[Learn more](#)



Personal Data Notebooks

Bastian Greshake Tzovaras
Open Humans Foundation

Connected by **466** members

Play with code for analyzing personal data! Your data stays private on a personal virtual machine.

[Learn more](#)



GenomiX Genome Exploration

Orit Shaer & Oded Nov
Wellesley College & NYU

Joined by **356** members

Is your 23andMe data public? Register interest in a study of how folks interact with their genome data using GenomiX, our visualization tool

[Learn more](#)



Imputer

Kevin Arvai

Connected by **55** members

Imputer will perform imputation on user's genotype data to estimate genotypes at missing loci.

[Learn more](#)



Galileo

Vineet Pandey
UCSD Design Lab

Connected by **54** members

Learn about current science of the gut microbiome with expert-curated learning materials.

[Learn more](#)



Nightscout Data Commons

The Nightscout Data
Commons Committee
The Nightscout Foundation

Joined by **66** members

The Nightscout Data Commons collects data from Nightscout users and facilitates research in partnership with the DIY diabetes community.

[Learn more](#)



Data sharing and ethical oversight

Prof. Dr. Effy Vayena
ETH Zurich

Joined by **210** members

Through a quick survey we aim at understanding what the Open Humans community thinks about ethical oversight.

[Learn more](#)



openSNP

Bastian Greshake Tzovaras
openSNP

Connected by **290** members

openSNP allows you to put your genetic and phenotypic data into the public domain. Connect your openSNP account to Open Humans.

[Learn more](#)



OpenAPS Data Commons

Dana Lewis
OpenAPS

Joined by **128** members

The OpenAPS Data Commons collects data from DIY closed loopers and facilitates research in partnership with the DIY closed loop community.

[Learn more](#)



lineage

Andrew Riha
Scedastic Software, LLC

Connected by **35** members

lineage provides tools for genetic genealogy - find shared DNA and genes between individuals, merge raw data files, and remap SNPs

[Learn more](#)

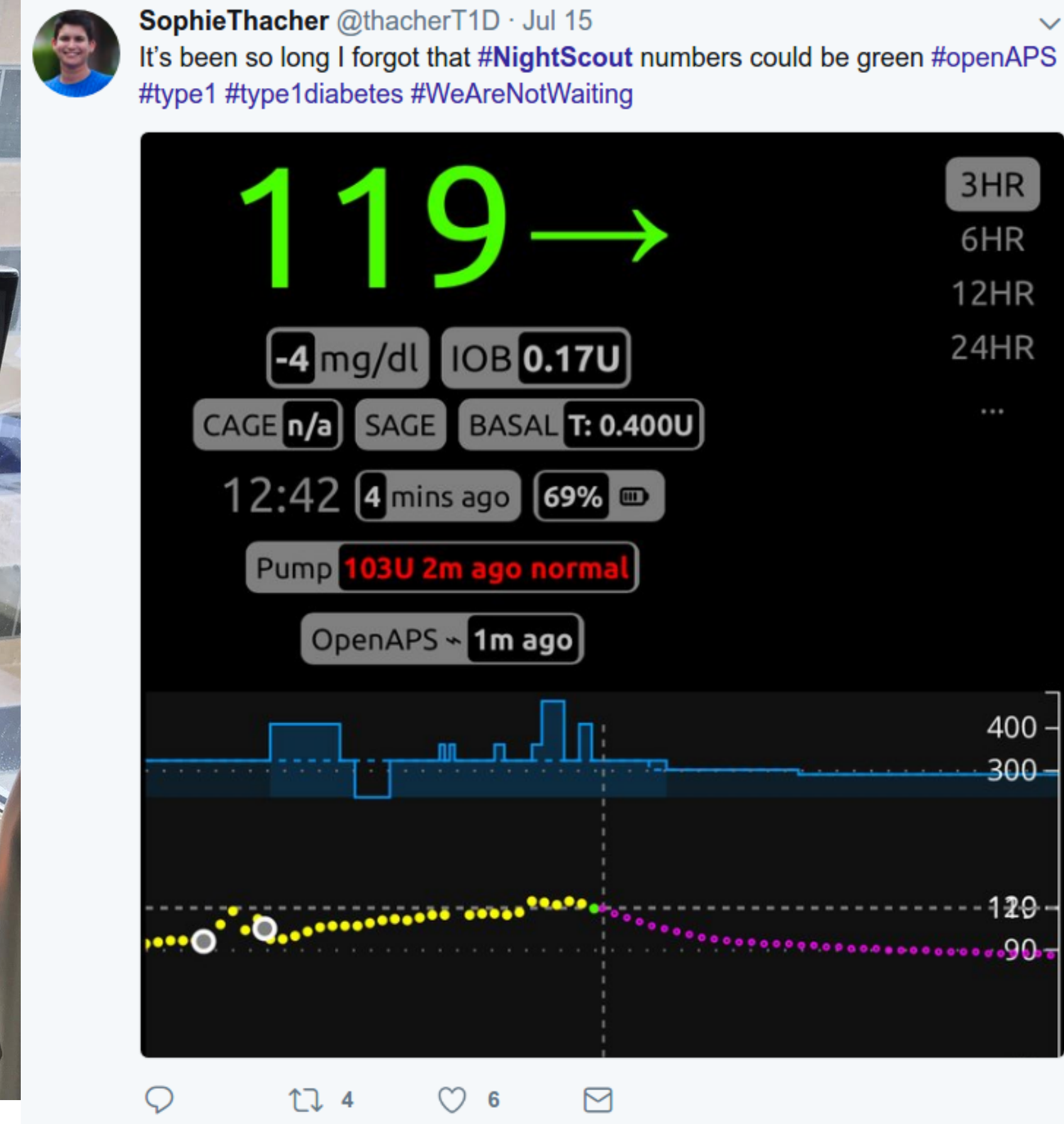


Resilience Project Study

Jason Bobe
Icahn School of Medicine at Mount
Sinai

Connected by **6** members

The Resilience Project aims to discover why some people are more able to overcome challenges to their health and escape disease.



This is **Dana Lewis** and her Open Artificial Pancreas System.

CGM data = continuous glucose monitor

#OpenAPS



NIGHTSCOUT
#WeAreNotWaiting



Welcome to QCycle!

QCycle is a participant-led research (PLR) project focusing on physiological change across the ovulatory cycle. We're a mix of researchers, computer scientists, engineers, and self-trackers looking to learn about ourselves, predict our LH surges, investigate links between our cycles and mood, HRV, productivity and more.

This is our private website for project resources. You can come here to find:

[Background information](#) on our project.

[Instructions](#) on how to use your devices and test kits.

[Important links.](#)

[Iterative Data Visualizations.](#)

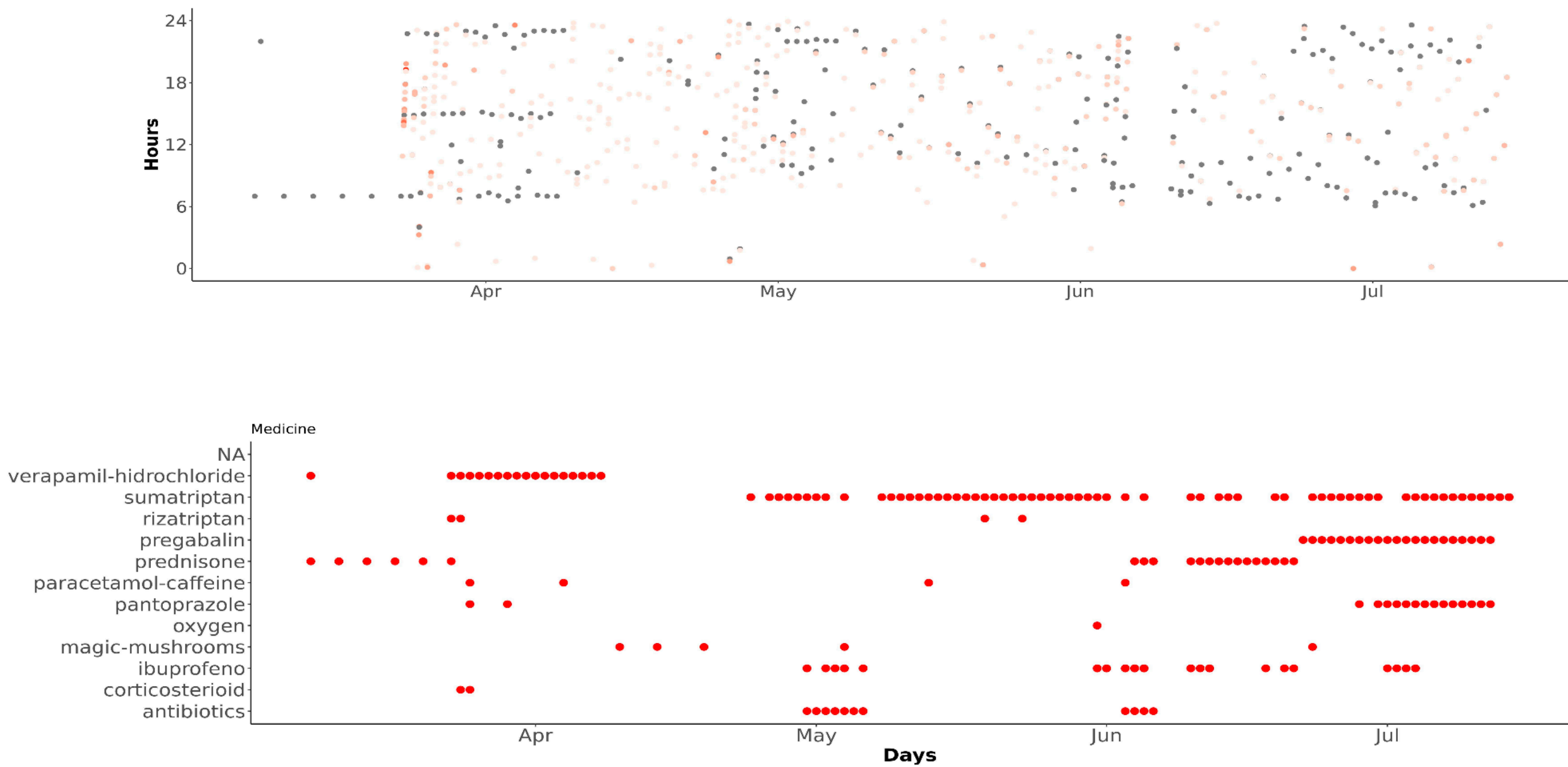


- collect their data with own app
- store & share data through *Open Humans*
- create & share reports

Resume of Treatments

It's interesting to know what treatments you took, its also good to see when you took these treatments and how your attacks responded to it. So we display here a timeline of attacks together with a timeline of your treatment intake.

This month you used **12** different treatments.



tl;dr

- most of citizen science these days heavily limits how participants are engaged
- The *Quantified Self* community does their own research, but a lack of shared tool chains makes it hard to scale
- Bringing together private data stores, standardized tools & ways to share tools/data enables co-creation & peer-production of scientific knowledge

thanks to:



Mad Ball

Mairi Dulaney

Mike Escalante

Our Outreachy interns: Rosy
Gupta, Manaswini Das, Jasmine
Tamak, Tarannum Khan
& over 7,500 community
members!

QUANTIFIED SELF

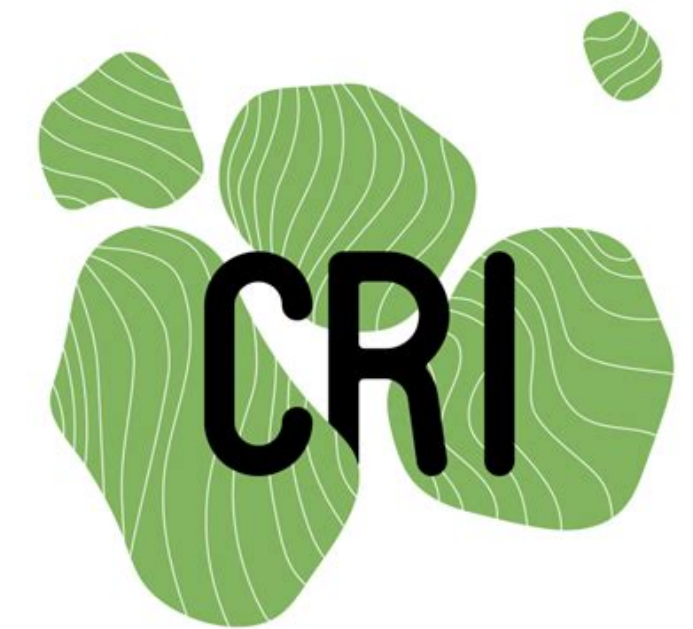
SELF KNOWLEDGE THROUGH NUMBERS

Gary Wolf

Azure Grant

I'm hiring people:

- PhD students
- Post-Docs
- Engineers



Talk to me if you're interested to
work with me on building a
community science peer-
production system at the *Center
for Research & Interdisciplinarity*
in Paris!

bastian@openhumans.org

@gedankenstuecke

+49 176 213 044 66