Quantified Self: Analyzing the Big Data of our Daily Life

Andreas Schreiber <Andreas.Schreiber@dlr.de>

PyData Berlin 2014

DLR

Introduction

Scientist,
Head of department

Deutsches Zentrum für Luft- und Raumfahrt
German Aerospace Center

Co-Founder, CEO

medando

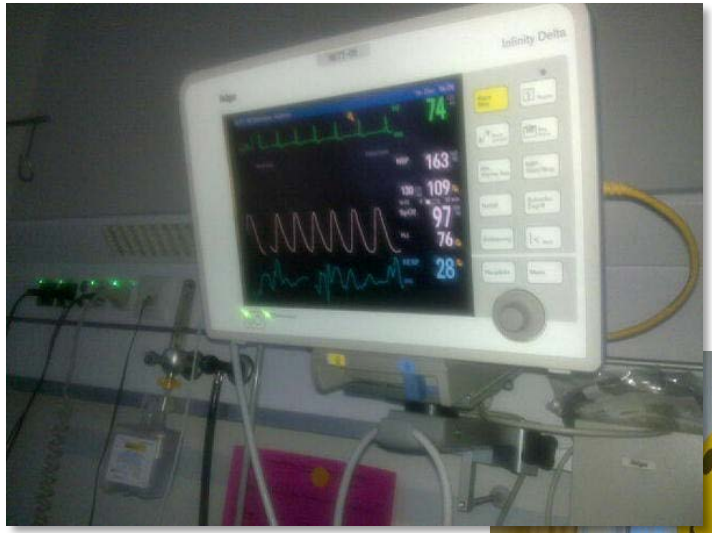
Co-Founder

Quantified Self Meetup Cologne

Python User since 1992

Related Background

Personal – Stroke 2009
https://twitter.com/onyame/status/6664357458



DLR – Telemedicine, AAL, Space medicine

Quantified Self

self knowledge through numbers

What is The Quantified Self?

Self-knowledge through numbers

- Analyze trends and set goals to improve yourself

Recording of daily activities

- Fitness, sleep, location, …
- Monitoring and display of information from various devices, services, and applications

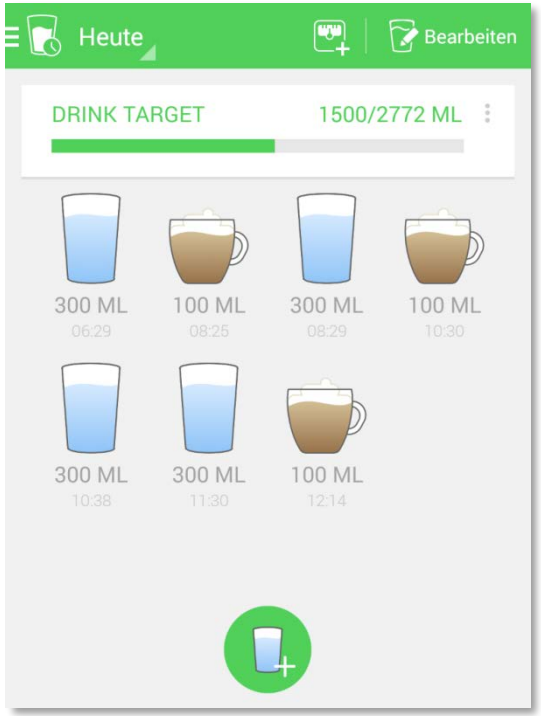
Other Terms

- Self Tracking
- Life Hacking
- Life Logging
- Self Optimization
- …

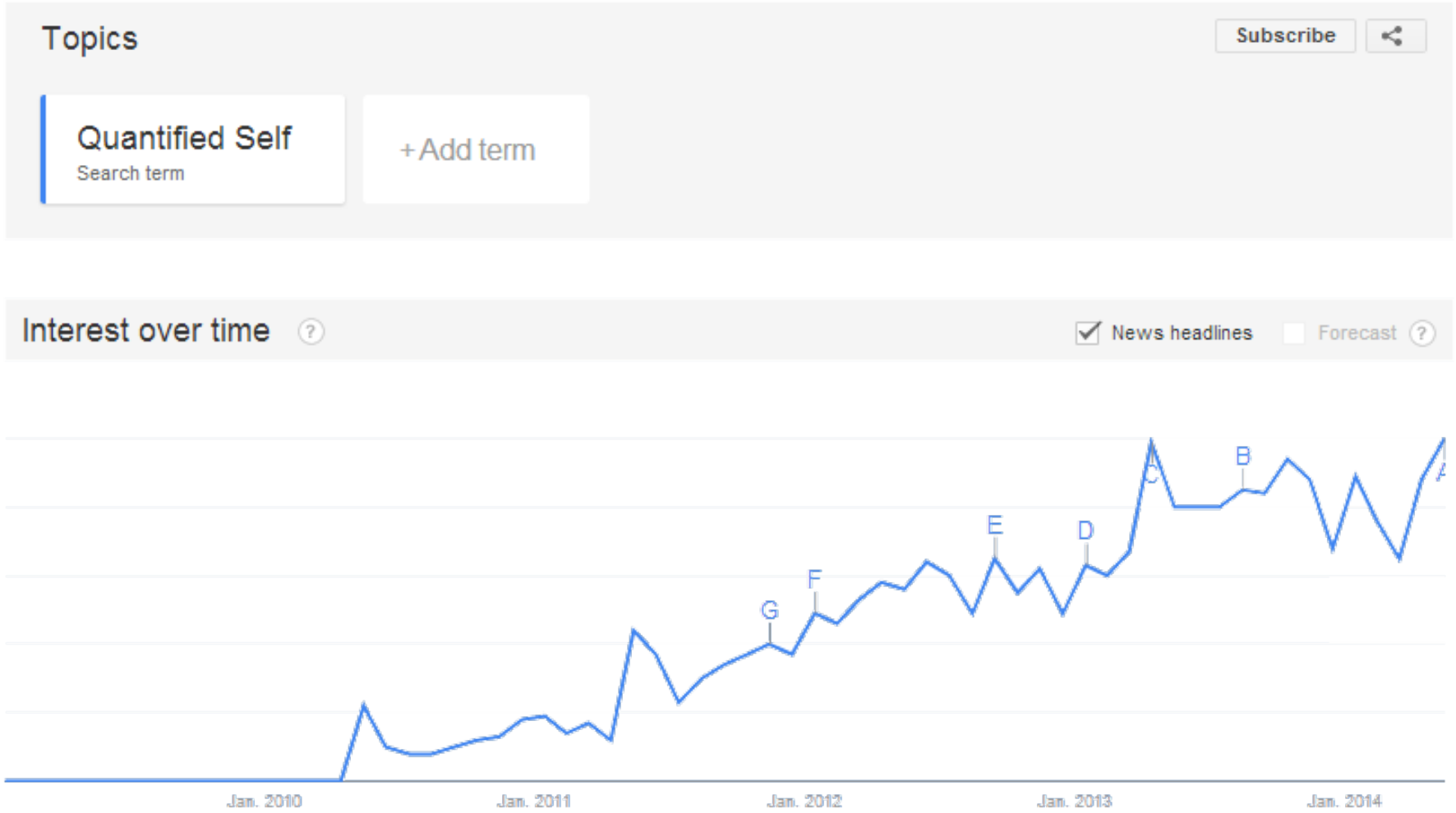
Self Optimization Example
Drinking Water

Optimization task:

- How much water is enough to **not** collapse during the heat period

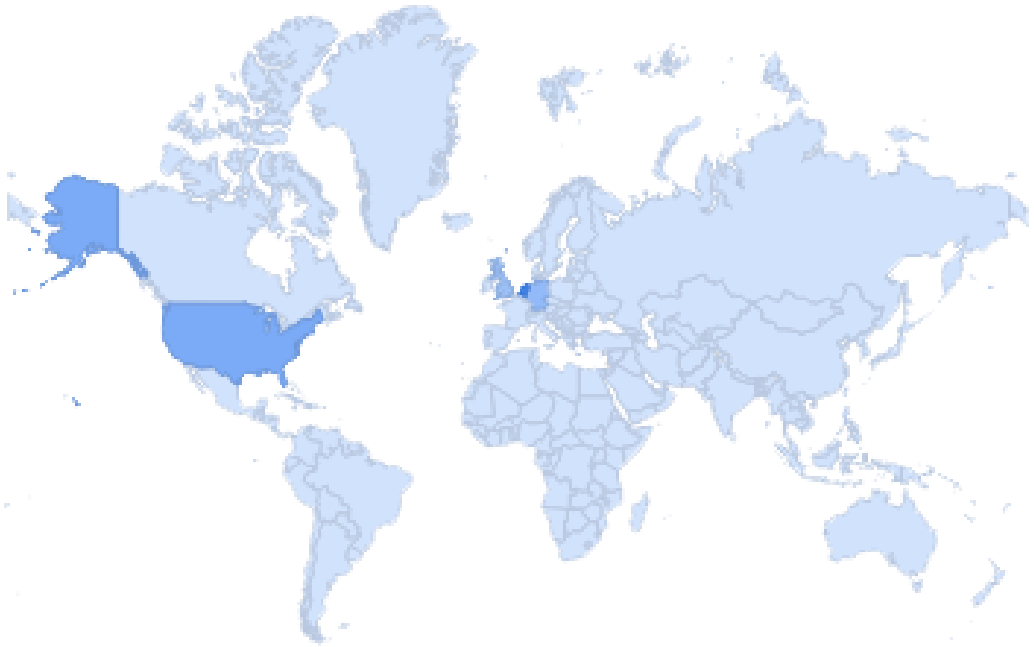


Google Trends: “Quantified Self”



Regional Interest

Regional interest



Region | City

Netherlands	100	
United States	59	
Germany	43	
United Kingdom	36	

Quantified Self Meetups

Quantified Self Meetup Groups

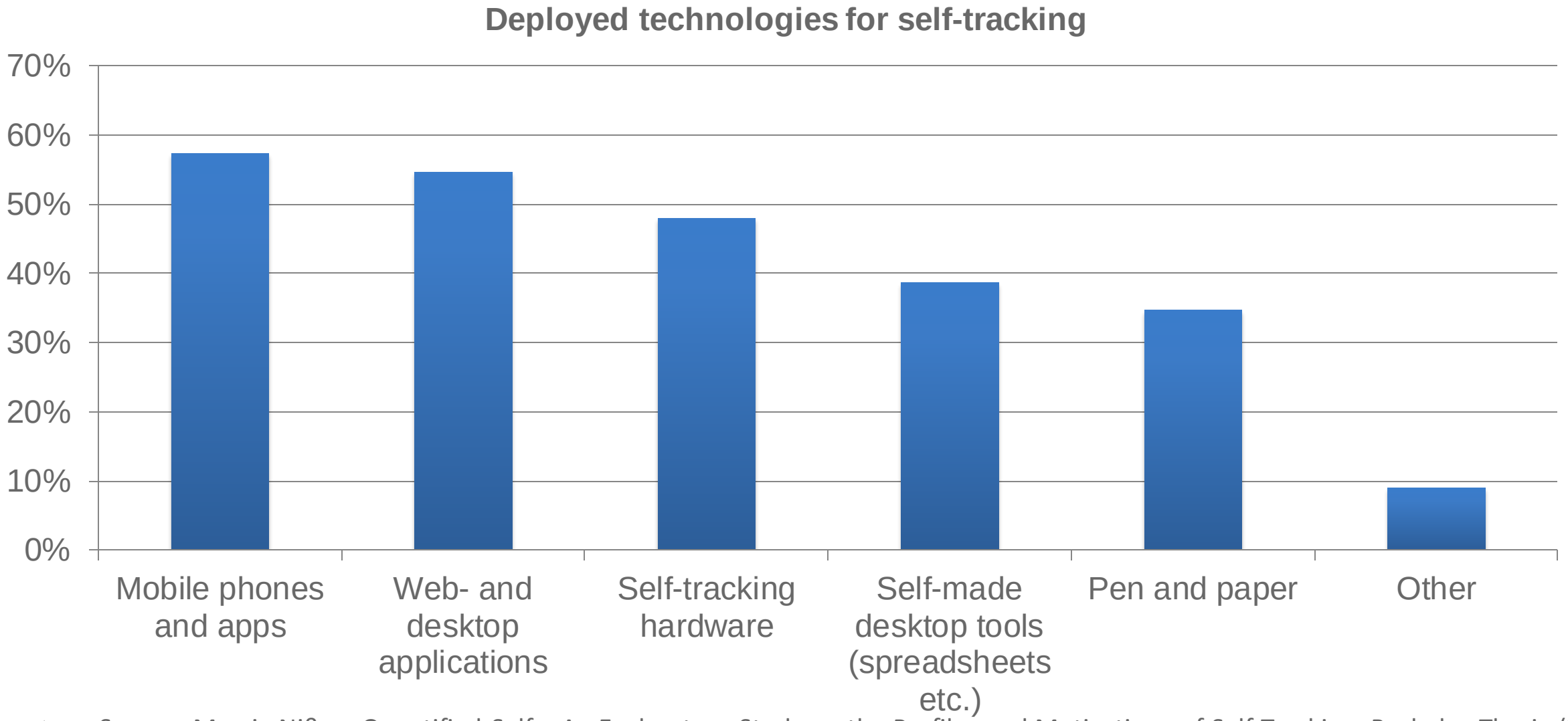
Groups	Members	Interested	Cities	Countries
169	31,502	8,550	121	38

http://quantified-self.meetup.com

Objects of Tracking

Addictions
ToDos
Temperature
Chronic diseases
Finance
Sex
Body
Delays (Train, etc.)
Blood test results
General medication
Rain
General daily records about health state
Physical activities
Mood
Clouds
Insulin intake
Location
Blood sugar
Symptoms
Ozone conzentration
No-chronic diseases
Nutrition
Atmospheric pressure
Relationship Quality

Technologies for Self-Tracking

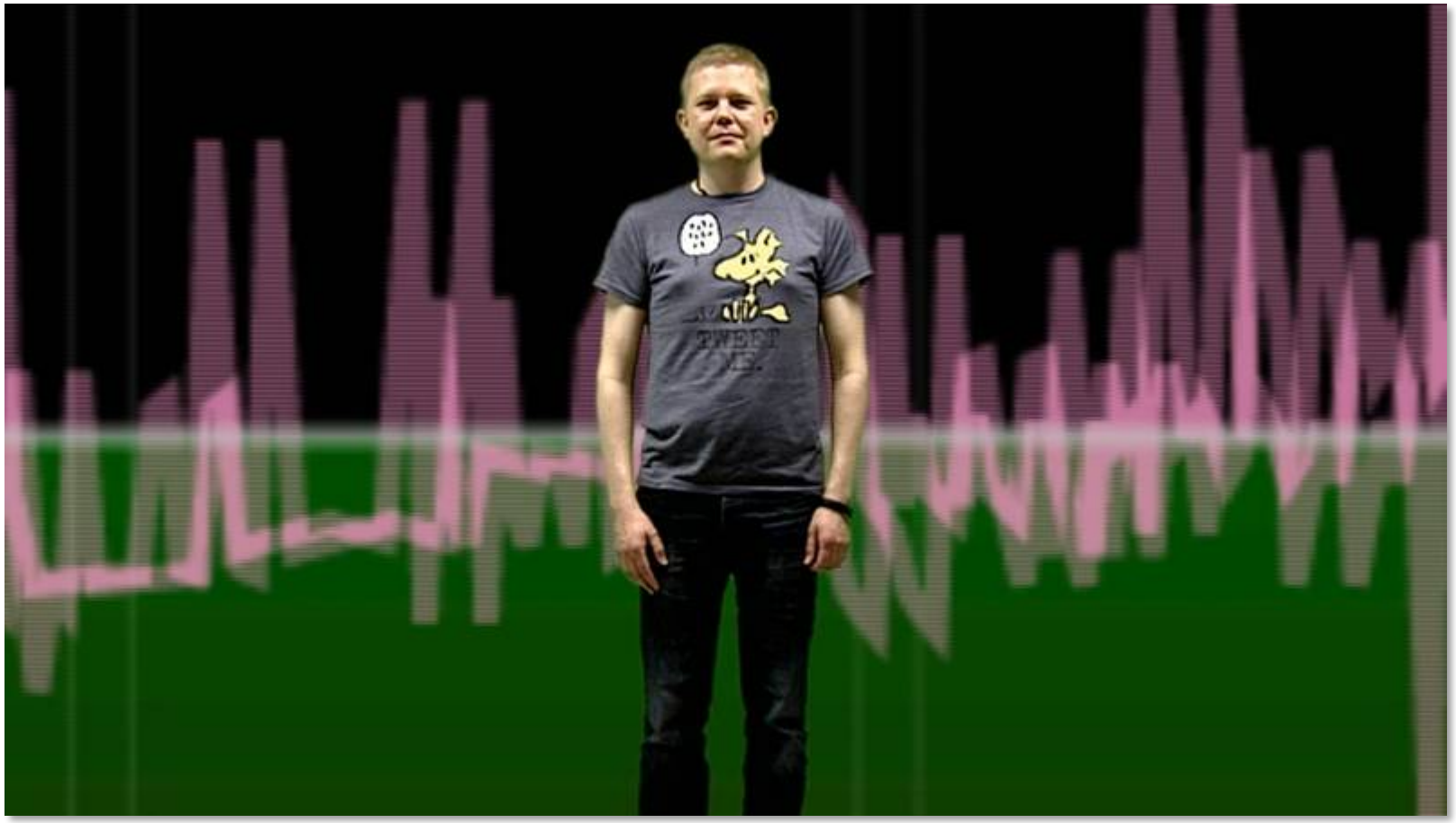


Source: Marcia Nißen, Quantified Self – An Exploratory Study on the Profiles and Motivations of Self-Tracking, Bachelor Thesis (2013)

Rise of the Wearables

My Self Tracking

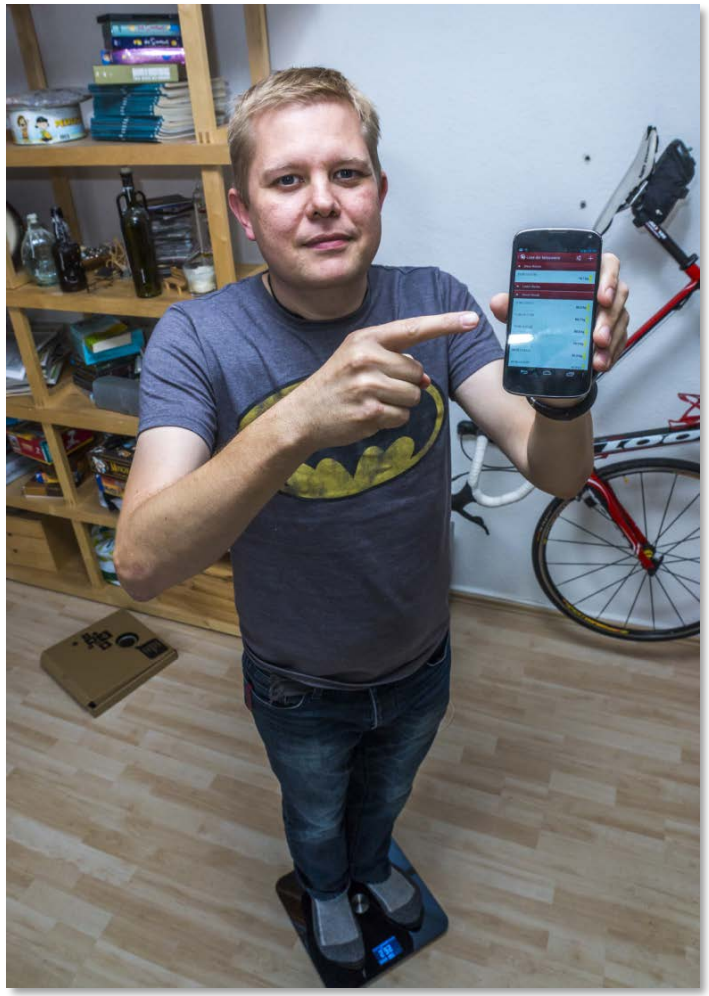
- With sensors
- With smartphone apps



Source: SAT.1/Weckup, http://bit.ly/10CEfUX

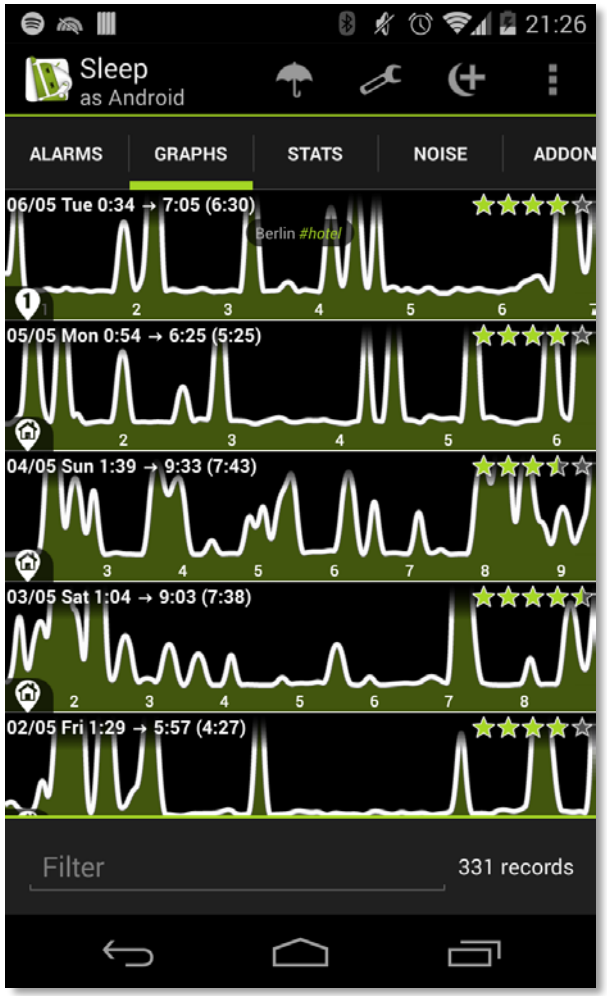
Steps (Fitbit)

Weight (Withings)

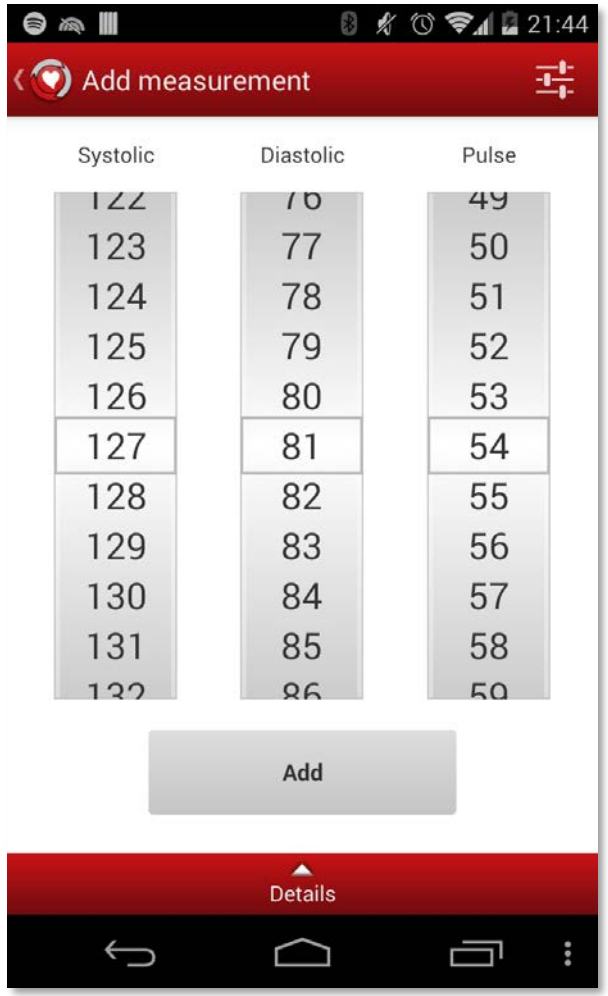


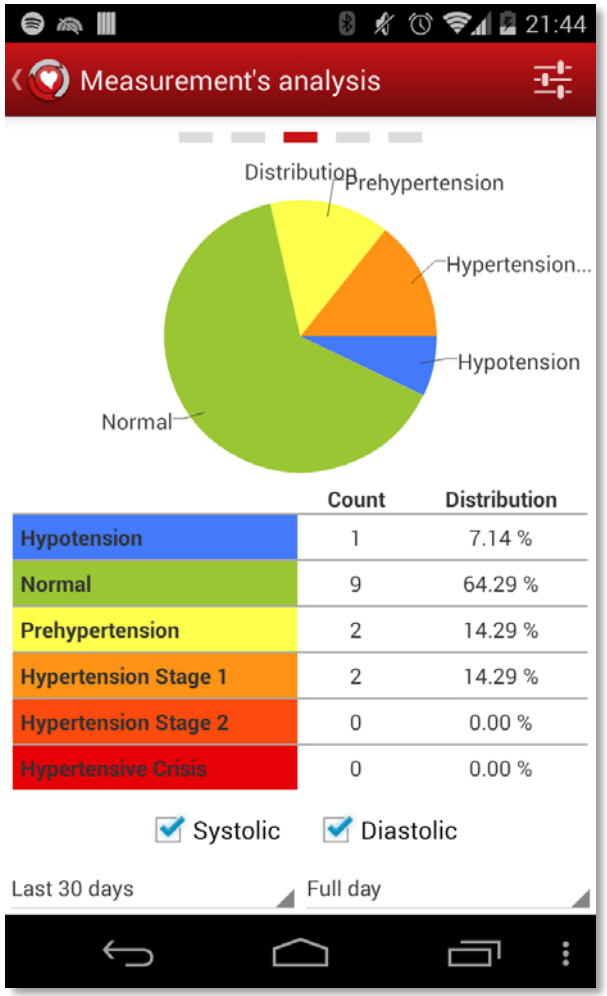
Stress (W/Me)

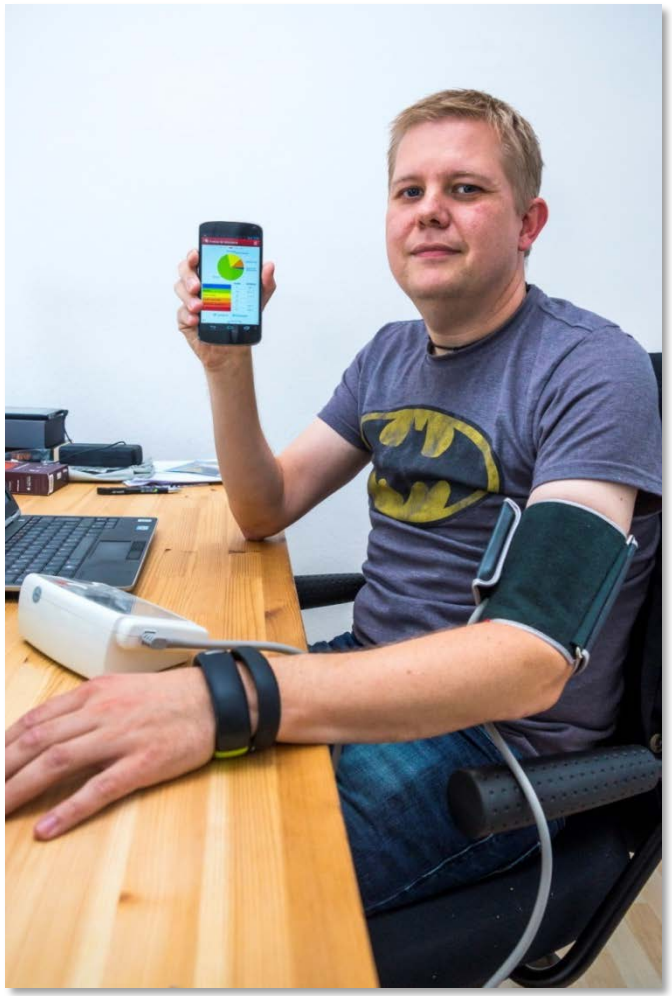
Sleep (Sleep as Android)



Blood Pressure (BloodPressureCompanion)



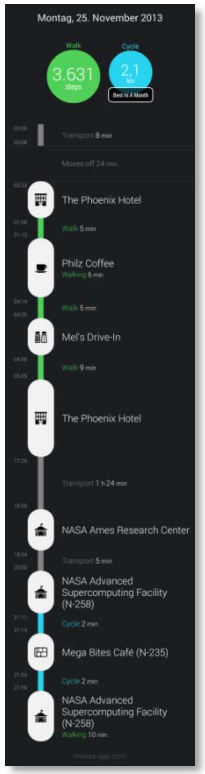


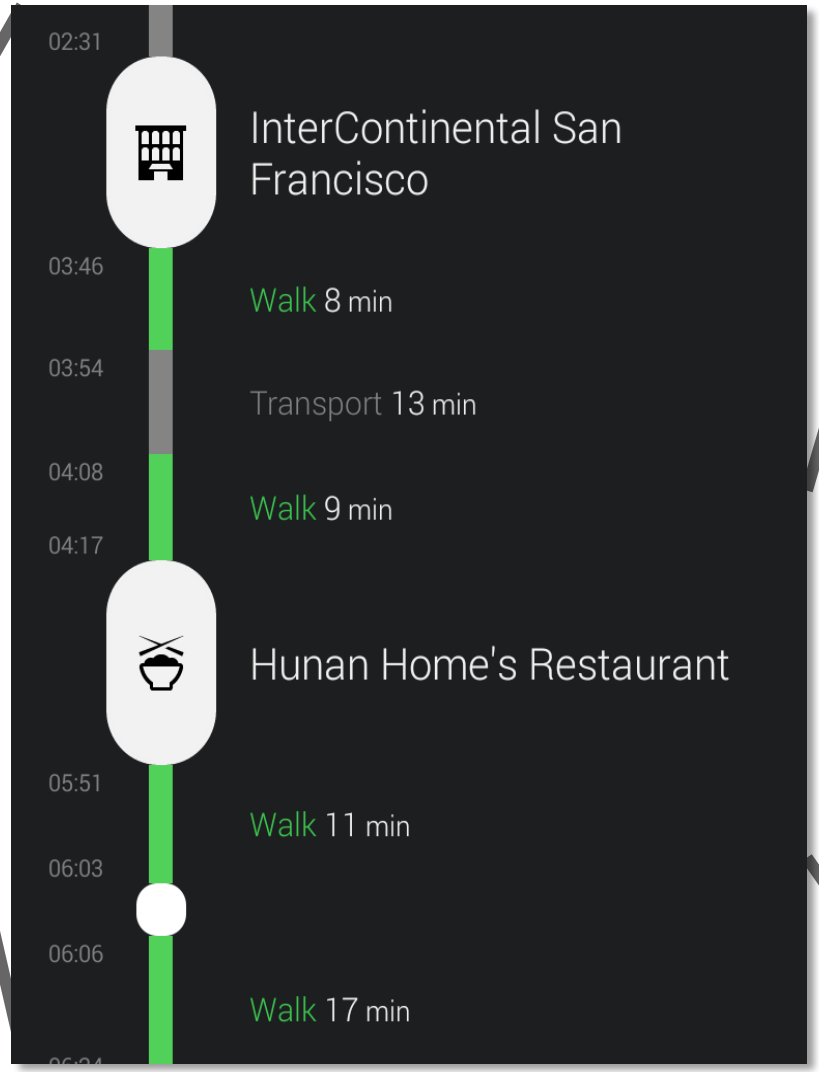


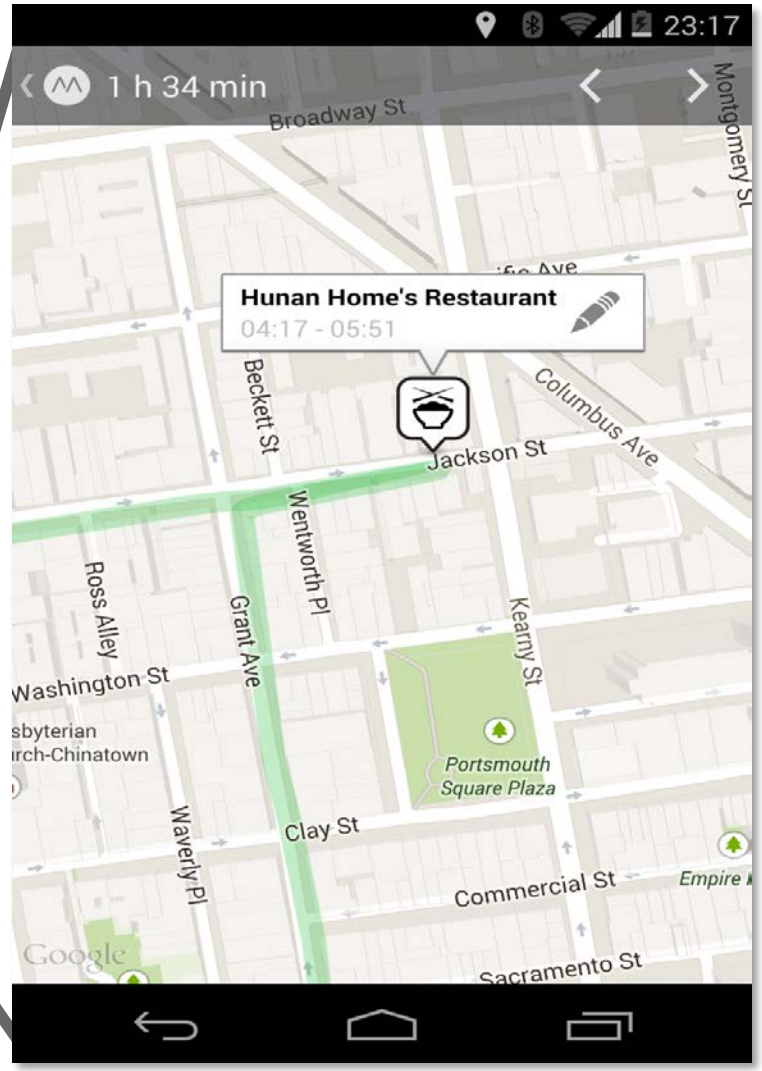
Activity & Location (Moves)

Source: WDR/Servicezeit, http://bit.ly/DigitaleSelbstvermessung

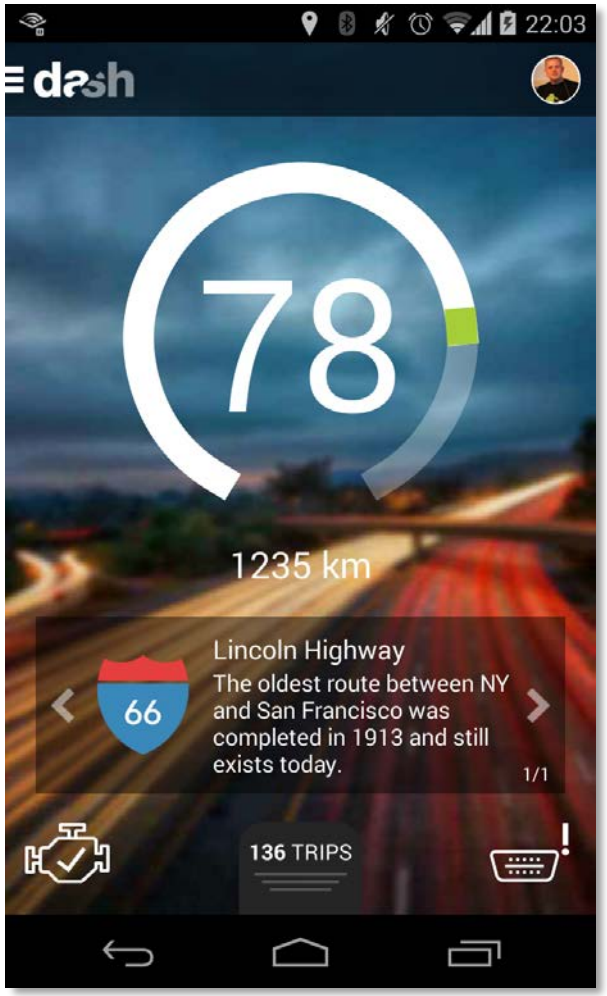
Activity & Location (Moves)

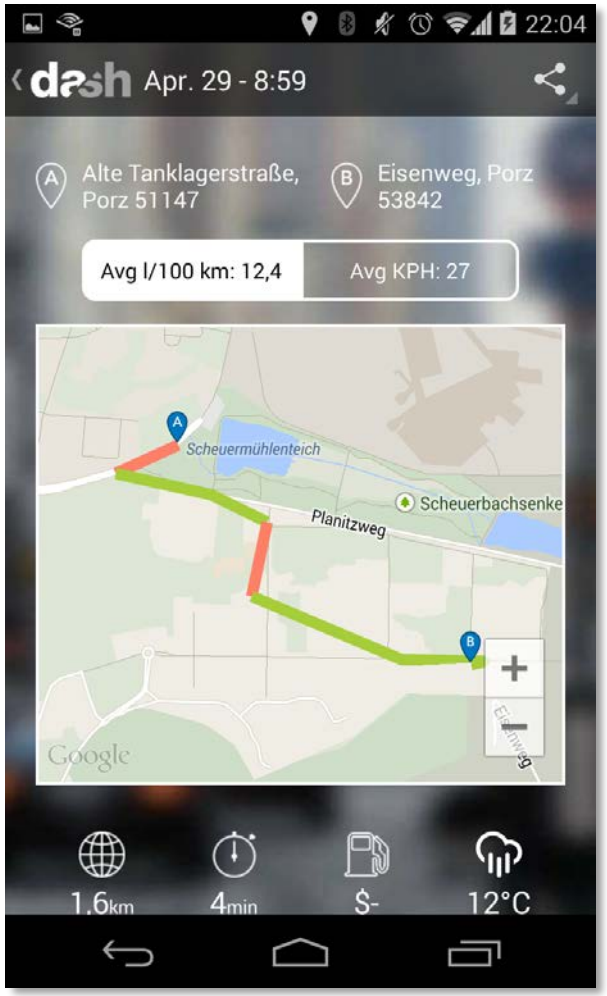






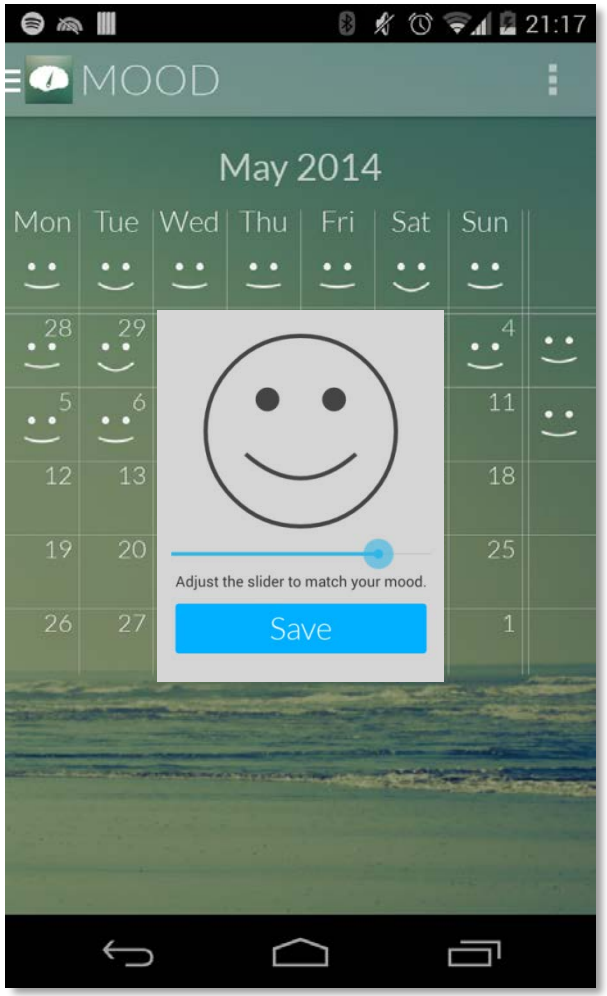
Car (Dash)

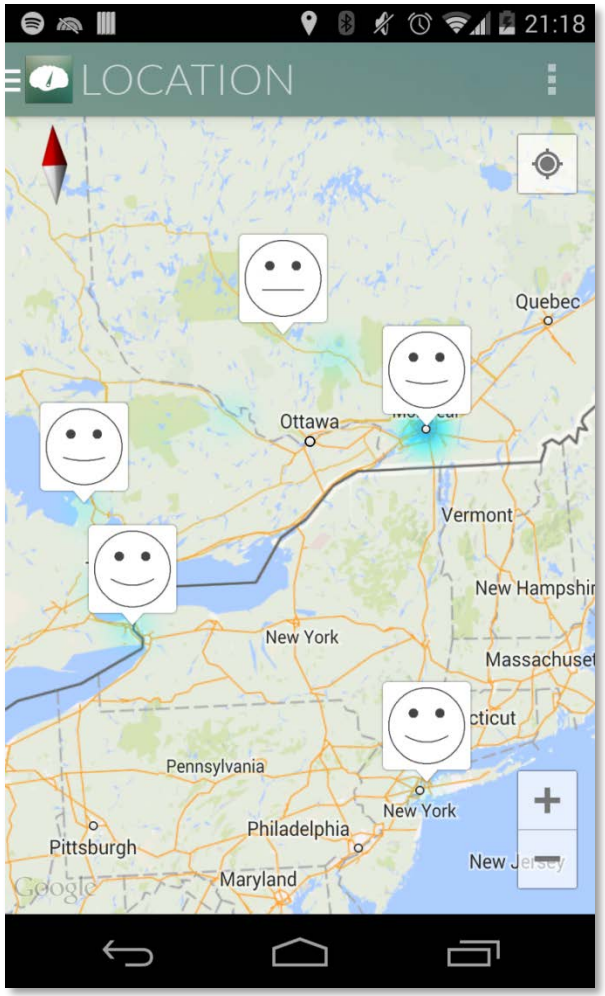

22:03
dash
78
1235 km
Lincoln Highway
The oldest route between NY and San Francisco was completed in 1913 and still exists today.
66
1/1
136 TRIPS


22:04
dash Apr. 29 - 8:59
Alte Tanklagerstraße, Porz 51147
Eisenweg, Porz 53842
Avg l/100 km: 12,4
Avg KPH: 27
Scheuermühlenteich
Scheuerbachsenke
Planitzweg
Google
1.6km
4min
$-
12°C

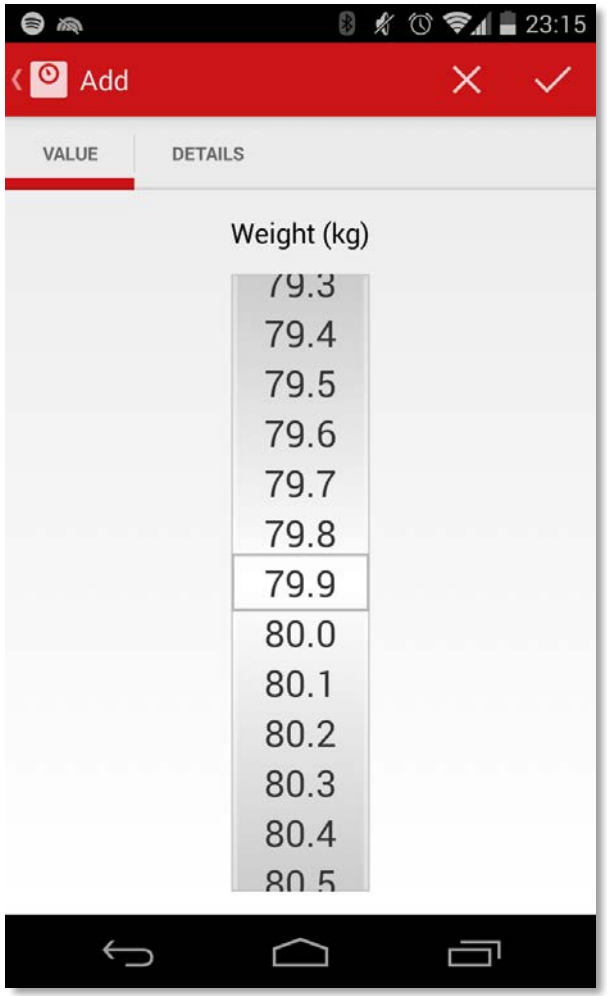

ELM 327 MINI
Interface
Supports all OBDII protocols

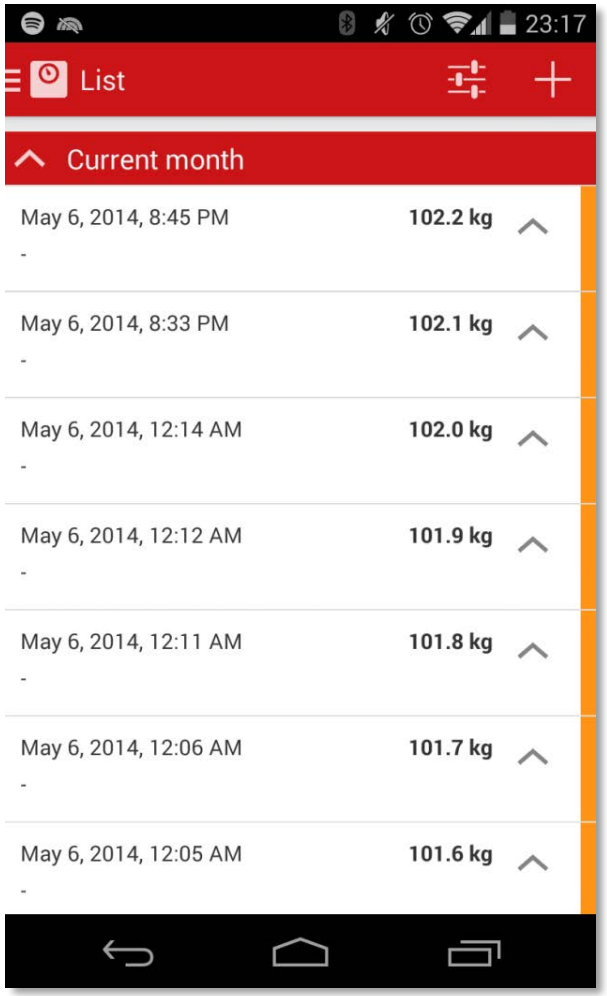
Mobile Phone Usage & Well-being (*Menthal*)

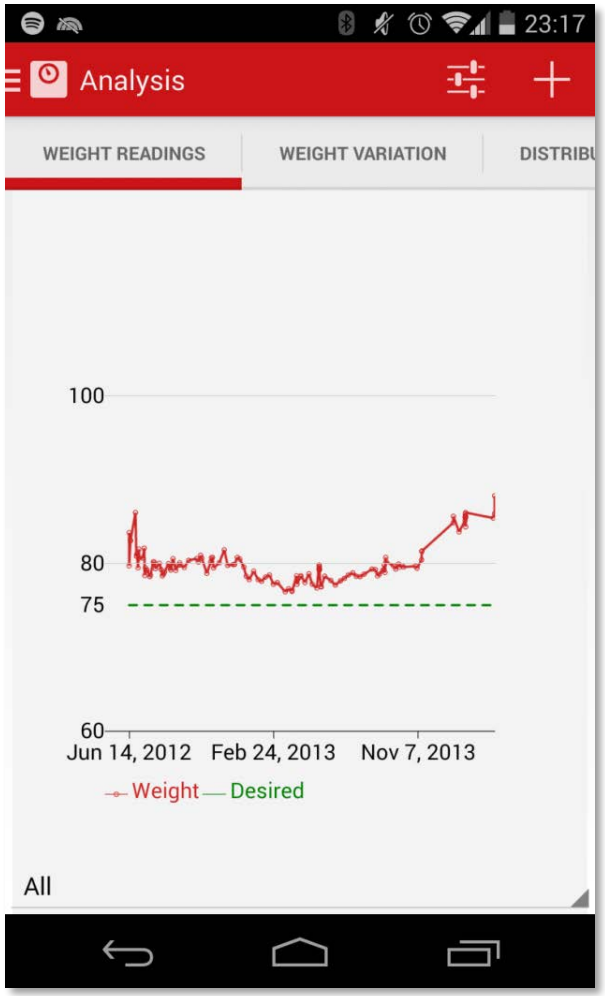




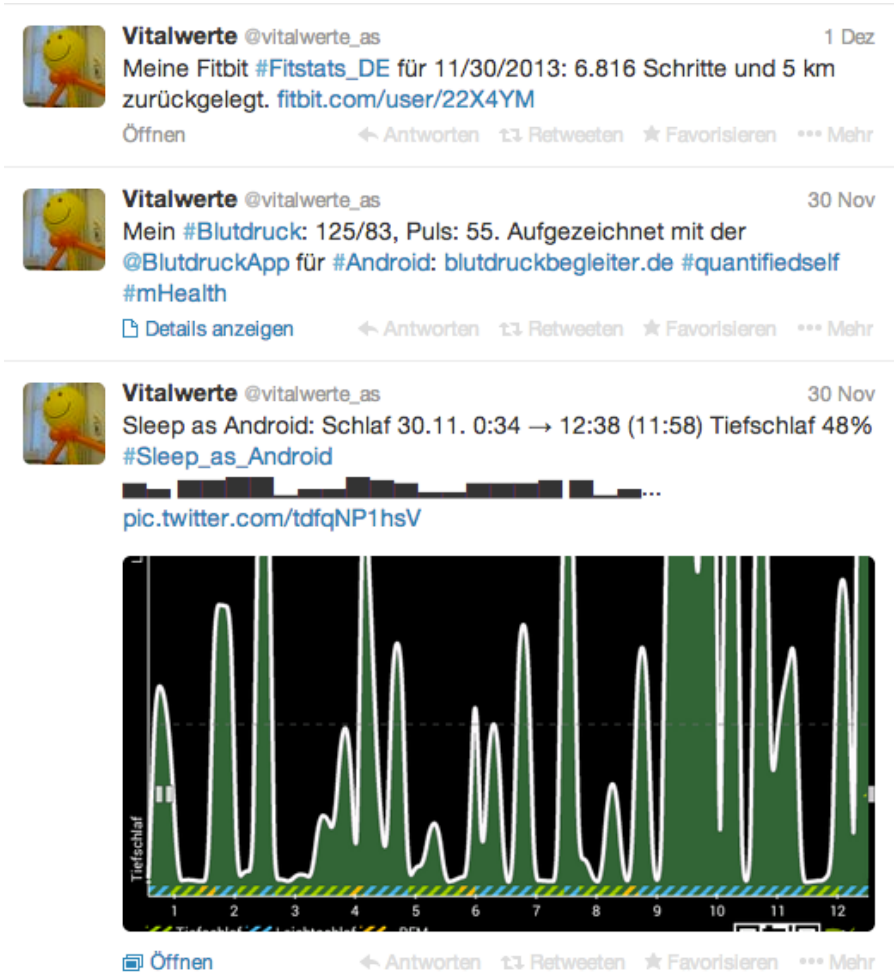
Weight (*WeightCompanion*)







Sharing


Vitalwerte @vitalwerte_as 1 Dez
Meine Fitbit #Fitstats_DE für 11/30/2013: 6.816 Schritte und 5 km zurückgelegt. fitbit.com/user/22X4YM
Öffnen Antworten Retweeten Favorisieren Mehr
Vitalwerte @vitalwerte_as 30 Nov
Mein #Blutdruck: 125/83, Puls: 55. Aufgezeichnet mit der @BlutdruckApp für #Android: blutdruckbegleiter.de #quantifiedself #mHealth
Details anzeigen Antworten Retweeten Favorisieren Mehr
Vitalwerte @vitalwerte_as 30 Nov
Sleep as Android: Schlaf 30.11. 0:34 → 12:38 (11:58) Tiefschlaf 48% #Sleep_as_Android
pic.twitter.com/tdfqNP1hsV
Tiefschlaf
1 2 3 4 5 6 7 8 9 10 11 12
Öffnen Antworten Retweeten Favorisieren Mehr

Human Centered

Location

Environment

Weather

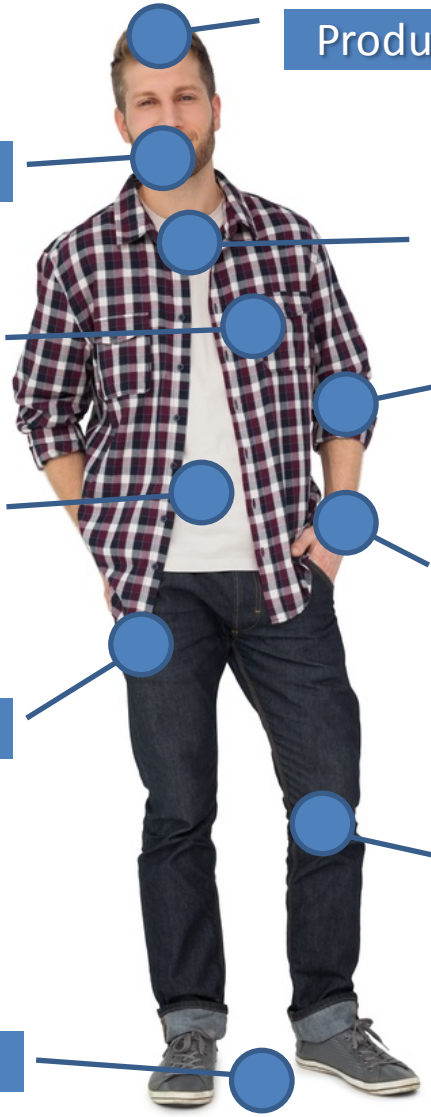
(Social) Interactions

Dog

Car

IoT (Smart Home)

Urban Data (Smart City)



Productivity

Food Intake

Photos

Heart Rate

Lab Values

Mood

Pulse

Activity

Posture

Weight

Cloud

Smartphone

Doctor

Family

Health Insurance

Photo: © WavebreakmediaMicro - Fotolia.com

The Data

Data Sources

Heterogeneous data sources

- Data from Wearables and other devices
- Data from smartphone apps
- Environmental data
- Social data
- IoT and urban data (smart home, smart car, smart city, …)

Heterogeneous storage

- Local files and databases (smartphone, device, desktop app)
- Cloud

Accessing Local and Distributed Data Sources

Export of files

- Most apps and web services allow file export (CSV, Excel, JSON, …)

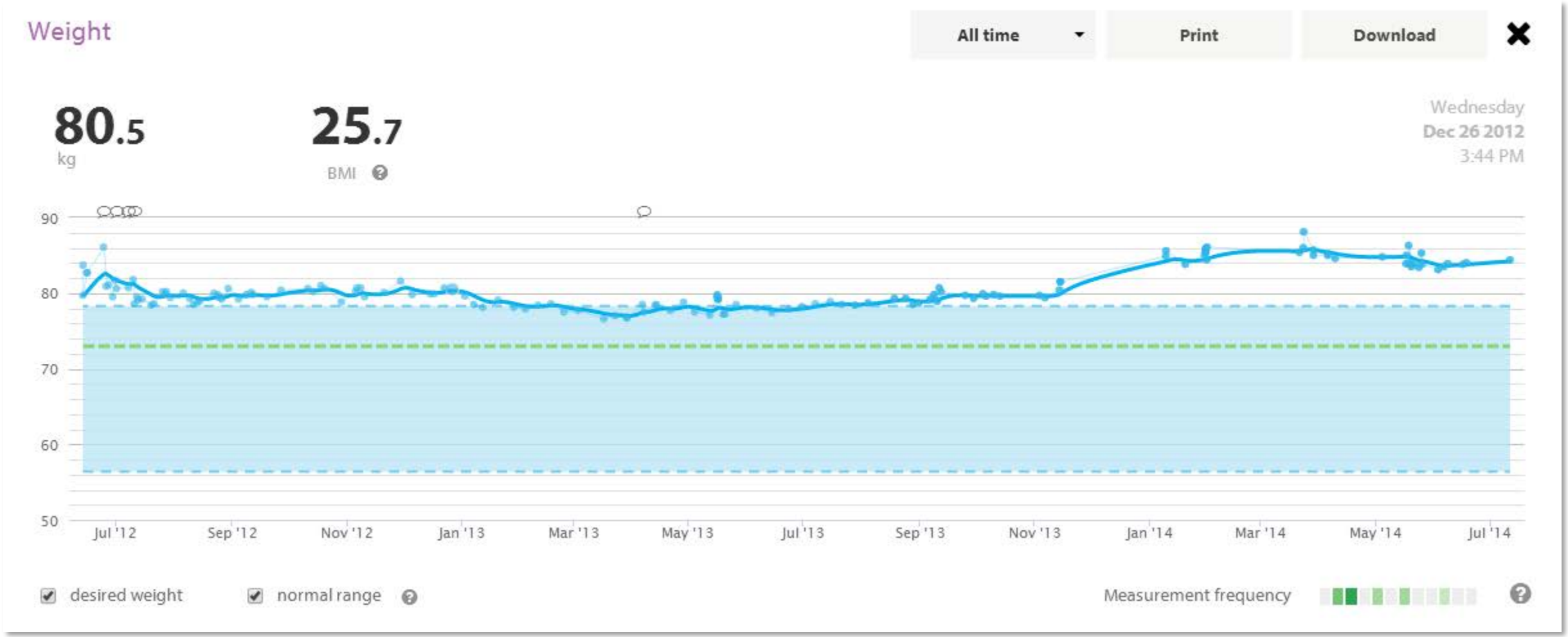
APIs

- Some vendors store date in their cloud only
- Access via vendor APIs

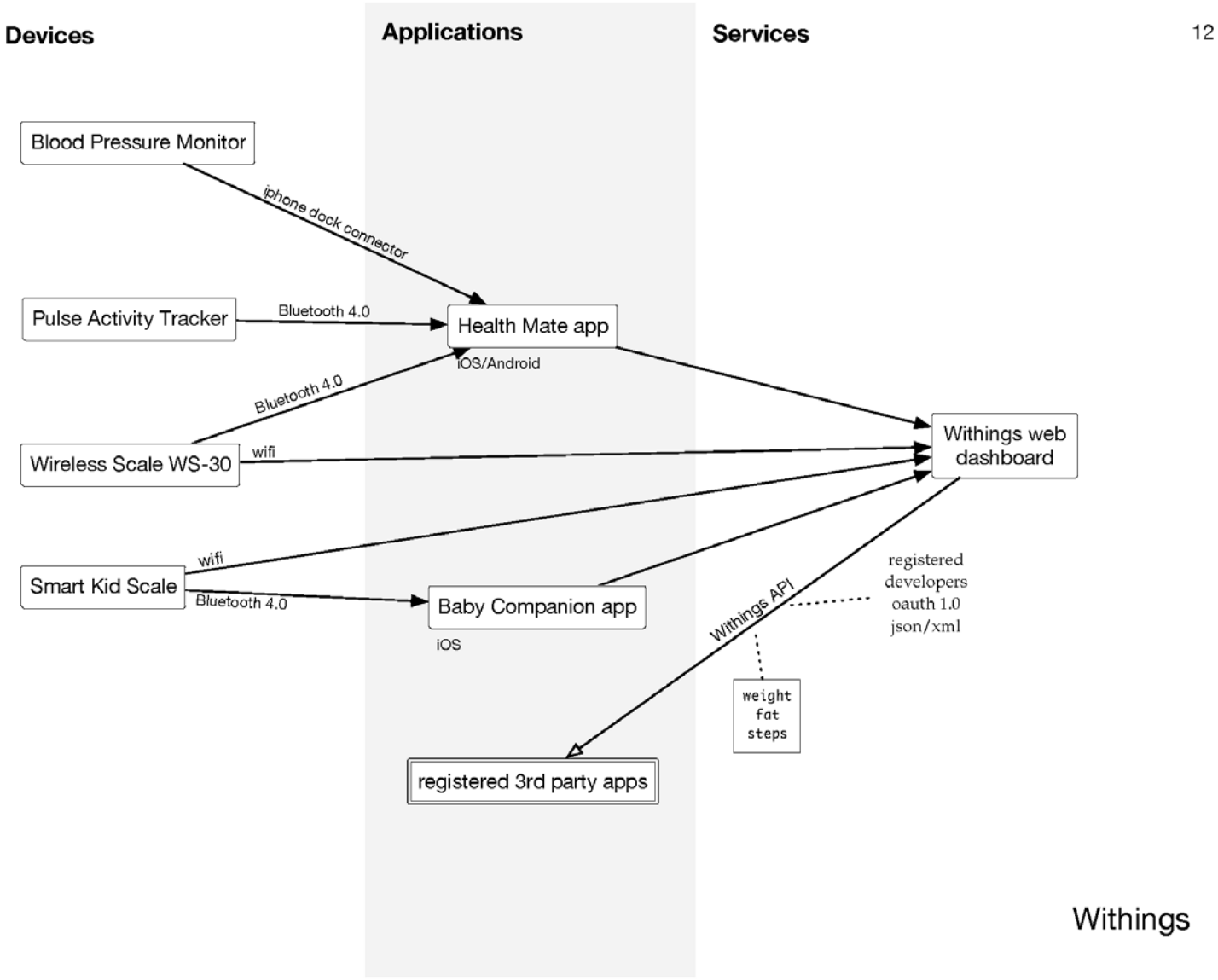
Unfortunately…

- Some (good) apps don’t have any export functionality or API
- APIs are very dissimilar for different vendors

Web Frontend
Withings

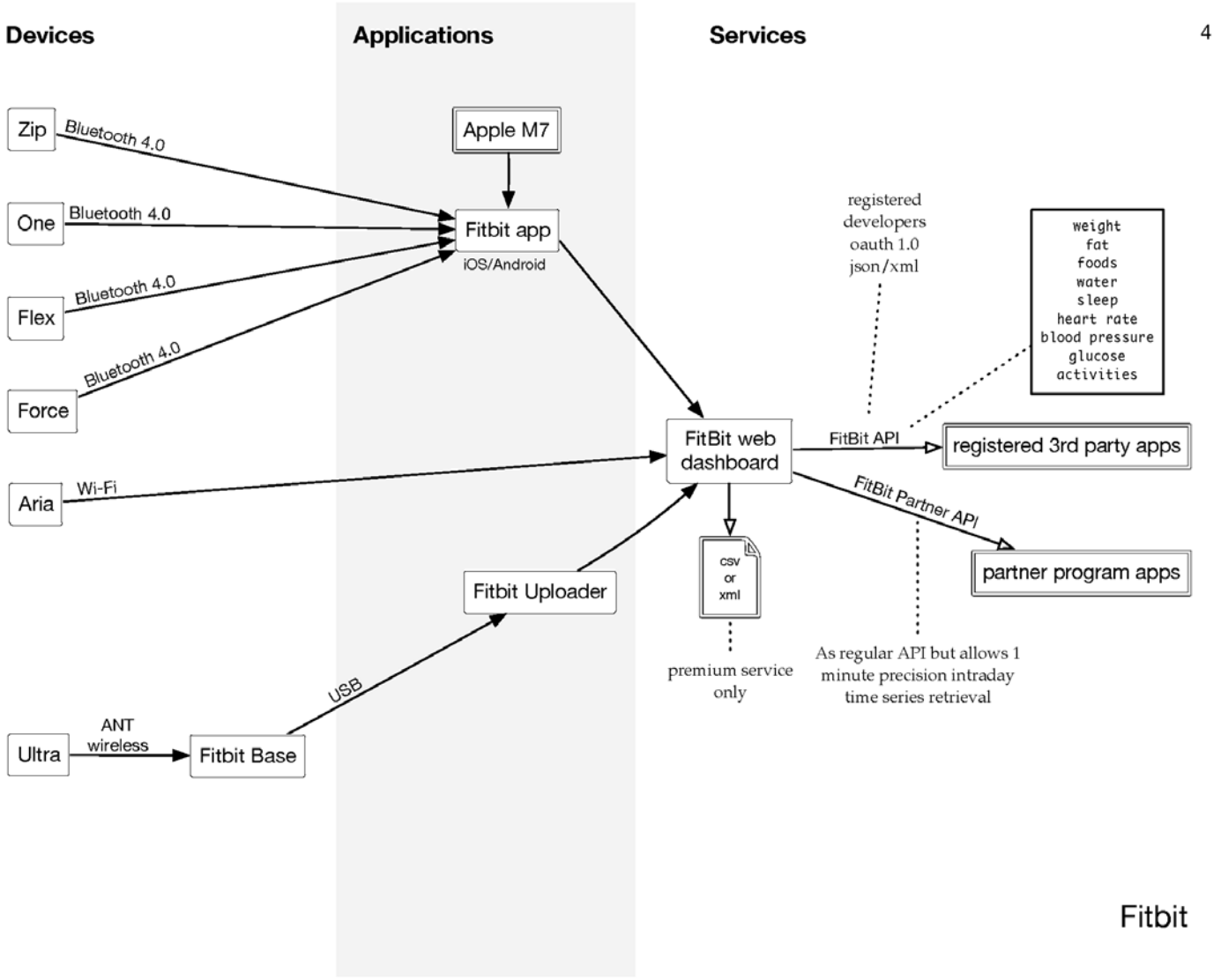


API Diagram
Withings



Source: https://forum.quantifiedself.com/thread-breakout-mapping-data-access

API Diagram
Fitbit



Source: https://forum.quantifiedself.com/thread-breakout-mapping-data-access

Example
Fitbit

IP[y]: Notebook Fitbit Steps Last Checkpoint: Jul 25 23:48 (autosaved)

File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None

Fitbit Steps

In [1]:
```
from fitbit import Fitbit
import pandas as pd
from pandas import DataFrame, Series
import simplejson as json
```

In [2]:
```
user_id = '22X4YM'
access_token_secret = '631063ab83b6b9'
access_token = 'ba7 4b530'
consumer_key = 'df7 9a160'
consumer_secret = '5c 1240d516252'
```

In [3]:
```
fitbit_client = Fitbit(consumer_key, consumer_secret, user_key=access_token, user_secret=access_token_secret)
```

In [14]:
```
user_profile = fitbit_client.user_profile_get(user_id=user_id)
print user_profile
```

```
{u'user': {u'weight': 77, u'memberSince': u'2012-05-26', u'locale': u'de_DE', u'strideLengthWalking': 73.5, u'height': 177, u's
trideLengthRunning': 92.10000000000001, u'glucoseUnit': u'METRIC', u'timezone': u'Europe/Berlin', u'avatar150': u'https://d6y8z
fzc2qfsl.cloudfront.net/8A3C66FA-61D2-5D41-97D9-A1140AFD7B0C_profile_150_square.png', u'dateOfBirth': u'1970-06-08', u'foodsLoc
ale': u'de_DE', u'distanceUnit': u'METRIC', u'heightUnit': u'METRIC', u'offsetFromUTCMillis': 7200000, u'fullName': u'Andreas S
chreiber', u'startDayOfWeek': u'MONDAY', u'displayName': u'Andreas', u'gender': u'MALE', u'weightUnit': u'METRIC', u'avatar': u
'https://d6y8zfzc2qfsl.cloudfront.net/8A3C66FA-61D2-5D41-97D9-A1140AFD7B0C_profile_100_square.png', u'waterUnit': u'METRIC', u'
country': u'DE', u'encodedId': u'22X4YM'}}
```

Example
Fitbit

IP[y]: Notebook **Fitbit Steps** Last Checkpoint: Jul 25 23:48 (autosaved)

File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None

```
'https://d6y8zfzc2qfs1.cloudfront.net/8A3C66FA-61D2-5D41-97D9-A1140AFD7B0C_profile_100_square.png', u'waterUnit': u'METRIC', u'
country': u'DE', u'encodedId': u'22X4YM'}}
```

Get values from Fitbit

See https://wiki.fitbit.com/display/API/API-Get-Time-Series

In [6]:
```
ts = fitbit_client.time_series('activities/steps', period='6m')
```

In [7]:
```
steps_dict = {}
for step in ts['activities-steps']:
    steps_dict[step['dateTime']] = int(step['value'])
```

In [13]:
```
steps = Series(steps_dict, name='Steps')
steps.head()
```

Out[13]:
```
2014-01-21    15686
2014-01-22     9674
2014-01-23    13849
2014-01-24    14562
2014-01-25    15015
Name: Steps, dtype: int64
```

In [9]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

In [10]:
```
plt.figure(); steps.plot(label='Steps'); plt.legend(loc='best')
```

Out[10]:
```
<matplotlib.legend.Legend at 0x1095c9810>
```

Example
Fitbit

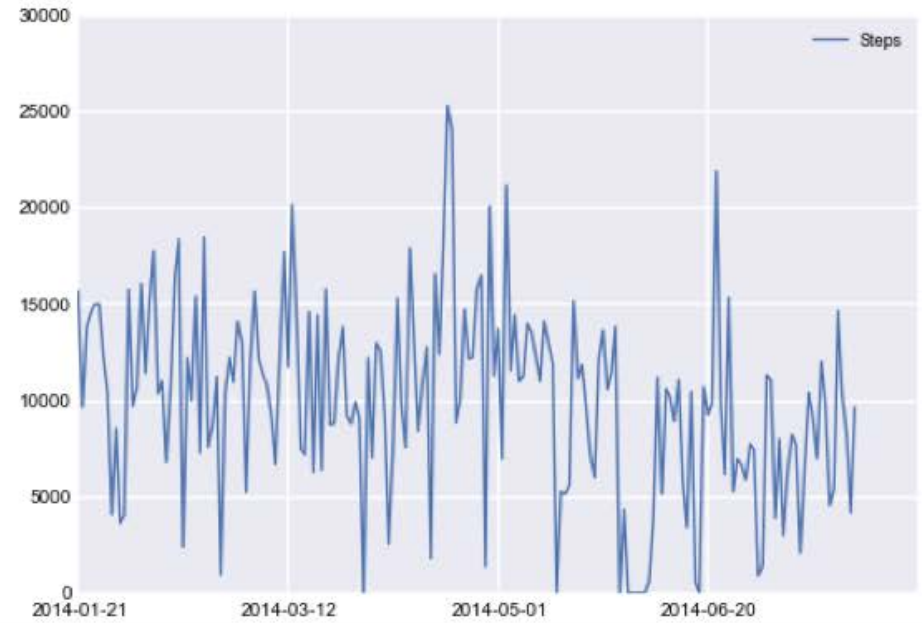
IP[y]: Notebook Fitbit Steps Last Checkpoint: Jul 25 23:48 (autosaved)

File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None

```
In [9]: import matplotlib.pyplot as plt
        import seaborn as sns

In [10]: plt.figure(); steps.plot(label='Steps'); plt.legend(loc='best')
Out[10]: <matplotlib.legend.Legend at 0x1095c9810>
```



Homogenizing the Data

Different data formats

- Almost every app and every Wearable has its own format!
- No standardization
- No access to raw data
- Best practice: Import into pandas DataFrame, then work with it

Example

TapLog Data

```
latitude,longitude,altitude,accuracy,gpstime,street,city,state,country,zip,samples
,_id,timestamp,DayOfYear,DayOfMonth,DayOfWeek,TimeOfDay,catOne,catTwo,catThree,num
ber,rating,note
"50.92","6.95982","0","31.544","07/21/2014 08:18","Bonner Straße 7","Köln","no
data","DE","50677","2","1517","21.07.2014 08:18","202","21","Montag",
"8.30833333333334","Kaffee",,,,,
"52.5231","13.4133","0","30","07/22/2014 08:18","Alexanderplatz 7","Berlin","no
data","DE","10178","1","1518","22.07.2014 08:19","203","22","Dienstag",
"8.3175","Kaffee",,,,,
"52.5206","13.4158","0","23","07/22/2014 11:48","Alexanderstraße 11","Berlin","no
data","DE","10178","2","1519","22.07.2014 11:48","203","22","Dienstag",
"11.8161111111111","Kaffee",,,,,
"52.5225","13.4095","0","29.69","07/22/2014 13:23","Karl-Liebknecht-Straße
15","Berlin","no data","DE","10178","1","1520","22.07.2014 13:23","203","22",
"Dienstag","13.385","Kaffee",,,,,
```

Example
Fitbit Data

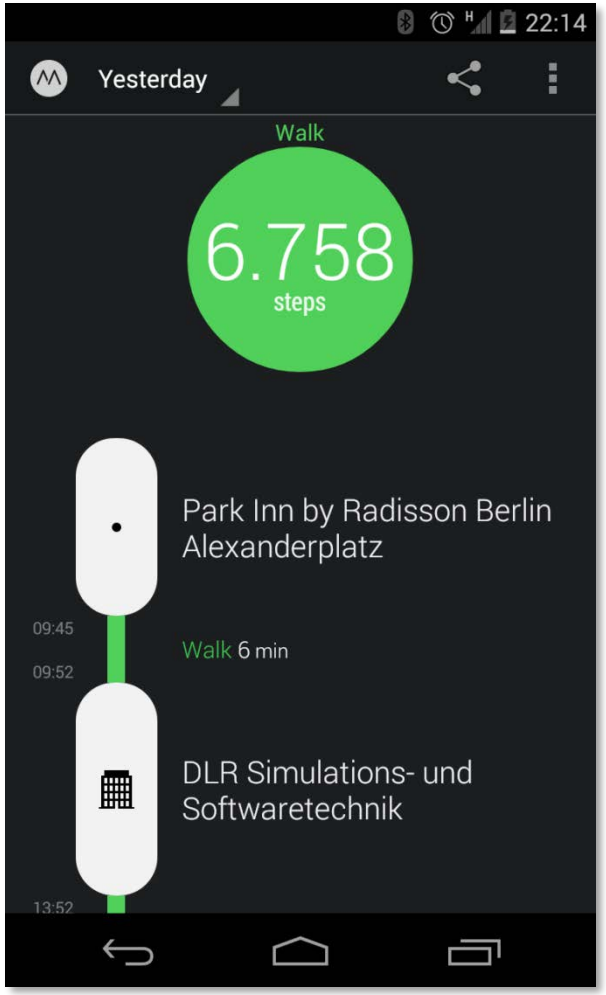
```
Aktivitäten
Datum,Verbrannte Kalorien,Schritte,Strecke,Stockwerke,Minuten im Sitzen,Minuten mit leichter Aktivität,Minuten mit relativ hoher Aktivität,Minuten mit sehr hoher Aktivität,Aktivitätskalorien
"01-04-2013","2.439","0","0","0","1.440","0","0","0","0"
"02-04-2013","2.083","3.871","2,85","4","1.273","109","48","10","604"
"03-04-2013","2.324","8.068","5,93","8","1.224","106","87","23","902"
"04-04-2013","2.805","17.190","12,63","23","1.135","113","128","64","1.485"
"05-04-2013","2.264","6.811","5,01","3","1.237","111","73","19","826"
"06-04-2013","2.507","11.261","8,28","18","1.208","93","99","40","1.118"
"07-04-2013","2.988","19.962","14,67","31","1.076","117","187","60","1.737"
"08-04-2013","3.020","19.186","14,1","19","1.089","108","172","71","1.754"
```

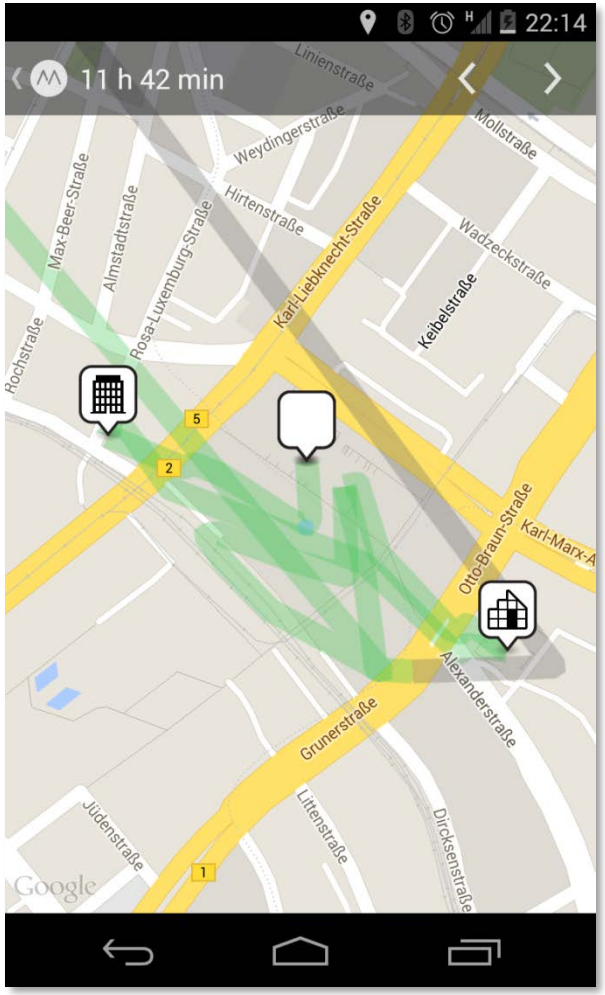
Example
Moves Data

Exports many formats

- CSV, geojson, georss, gpx, ical, json, kml
- daily, weekly, monthly, yearly, full

- **Different number of files for each format**
  - For CSV: activities, places, storyline, summary





Moves Data

places.csv

```
Date,Name,Start,End,Duration,Latitude,Longitude,Category,Link
24.07.14,Park Inn by Radisson Berlin Alexanderplatz,2014-07-
24T00:00:00+02:00,2014-07-24T09:45:33+02:00,35133,52.5227623262832 5,
13.412772417068481,,
24.07.14,DLR Simulations- und Softwaretechnik,2014-07-24T09:52:06+02:00,2014-07-
24T13:52:57+02:00,14451,52.523047921834 11,13.409121930599213,,
24.07.14,bcc Berliner Congress Center,2014-07-24T14:08:14+02:00,2014-07-
24T17:42:27+02:00,12853,52.520647229439 5,13.416452407836914,,
24.07.14,Park Inn by Radisson Berlin Alexanderplatz,2014-07-
24T17:53:11+02:00,2014-07-24T18:07:40+02:00,869,52.5227623262832 5,
13.412772417068481,,
24.07.14,Factory,2014-07-24T18:28:25+02:00,2014-07-24T22:08:39+02:00,13214,
52.537250304678 5,13.395079791885648,,
24.07.14,St. Oberholz,2014-07-24T22:19:29+02:00,2014-07-24T22:39:47+02:00,1218,
52.529624770283 07,13.401576783271219,,
```

Moves Data

activities.csv

```
Date,Activity,Group,Start,End,Duration,Distance,Steps,Calories
24.07.14,walking,walking,2014-07-24T09:17:35+02:00,2014-07-24T09:20:33+02:00,178,0.134,159,0
24.07.14,walking,walking,2014-07-24T09:45:33+02:00,2014-07-24T09:52:06+02:00,393,0.383,641,0
24.07.14,walking,walking,2014-07-24T10:00:24+02:00,2014-07-24T10:04:03+02:00,219,0.307,409,0
24.07.14,walking,walking,2014-07-24T13:13:42+02:00,2014-07-24T13:14:42+02:00,60,0.045,91,0
24.07.14,walking,walking,2014-07-24T13:52:57+02:00,2014-07-24T14:08:14+02:00,917,0.820,1404,0
24.07.14,walking,walking,2014-07-24T14:22:52+02:00,2014-07-24T14:26:57+02:00,245,0.221,295,0
24.07.14,walking,walking,2014-07-24T15:28:11+02:00,2014-07-24T15:30:16+02:00,125,0.126,168,0
24.07.14,walking,walking,2014-07-24T16:36:34+02:00,2014-07-24T16:37:04+02:00,30,0.015,30,0
24.07.14,walking,walking,2014-07-24T17:42:27+02:00,2014-07-24T17:53:11+02:00,644,0.346,649,0
24.07.14,walking,walking,2014-07-24T18:05:01+02:00,2014-07-24T18:05:31+02:00,30,0.015,27,0
24.07.14,walking,walking,2014-07-24T18:07:40+02:00,2014-07-24T18:14:47+02:00,427,0.316,501,0
24.07.14,transport,transport,2014-07-24T18:14:47+02:00,2014-07-24T18:21:49+02:00,422,2.593,0,0
24.07.14,walking,walking,2014-07-24T18:21:49+02:00,2014-07-24T18:28:25+02:00,396,0.434,491,0
24.07.14,walking,walking,2014-07-24T18:33:21+02:00,2014-07-24T18:33:51+02:00,30,0.015,20,0
24.07.14,walking,walking,2014-07-24T21:03:01+02:00,2014-07-24T21:04:31+02:00,90,0.060,121,0
24.07.14,walking,walking,2014-07-24T22:05:26+02:00,2014-07-24T22:06:26+02:00,60,0.045,90,0
```

DLR

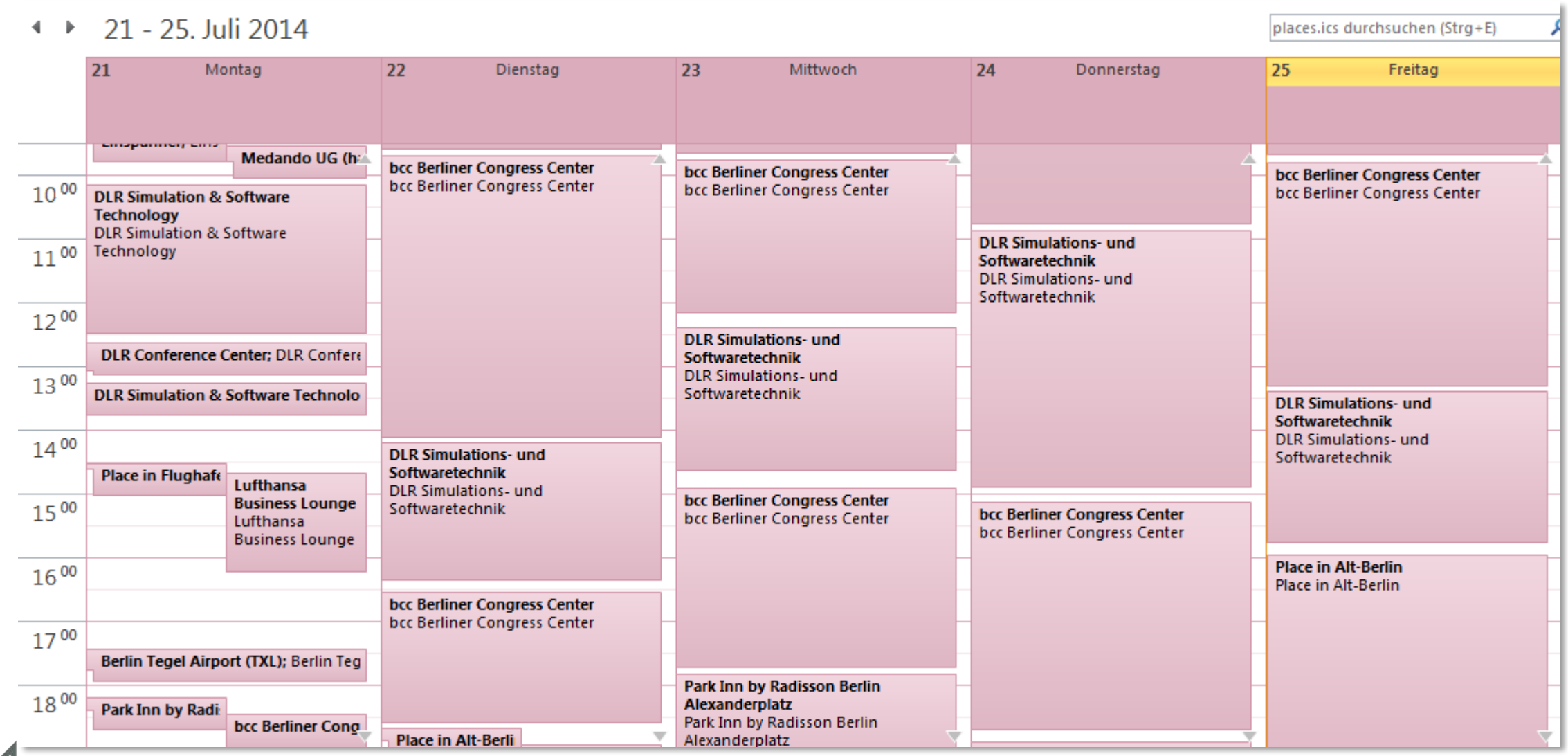
Moves Data

storyline.csv

```
Date,Type,Name,Start,End,Duration
24.07.14,place,Park Inn by Radisson Berlin Alexanderplatz,2014-07-24T00:00:00+02:00,2014-07-24T09:45:33+02:00,35133
24.07.14,move,walking,2014-07-24T09:45:33+02:00,2014-07-24T09:52:06+02:00,393
24.07.14,place,DLR Simulations- und Softwaretechnik,2014-07-24T09:52:06+02:00,2014-07-24T13:52:57+02:00,14451
24.07.14,move,walking,2014-07-24T13:52:57+02:00,2014-07-24T14:08:14+02:00,917
24.07.14,place,bcc Berliner Congress Center,2014-07-24T14:08:14+02:00,2014-07-24T17:42:27+02:00,12853
24.07.14,move,walking,2014-07-24T17:42:27+02:00,2014-07-24T17:53:11+02:00,644
24.07.14,place,Park Inn by Radisson Berlin Alexanderplatz,2014-07-24T17:53:11+02:00,2014-07-24T18:07:40+02:00,869
24.07.14,move,walking,2014-07-24T18:07:40+02:00,2014-07-24T18:14:47+02:00,427
24.07.14,move,transport,2014-07-24T18:14:47+02:00,2014-07-24T18:21:49+02:00,422
24.07.14,move,walking,2014-07-24T18:21:49+02:00,2014-07-24T18:28:25+02:00,396
24.07.14,place,Factory,2014-07-24T18:28:25+02:00,2014-07-24T22:08:39+02:00,13214
24.07.14,move,walking,2014-07-24T22:08:39+02:00,2014-07-24T22:19:28+02:00,649
24.07.14,place,St. Oberholz,2014-07-24T22:19:29+02:00,2014-07-24T22:39:47+02:00,1218
24.07.14,move,walking,2014-07-24T22:39:47+02:00,2014-07-24T22:45:40+02:00,353
```

DLR

Moves Data: Places in Outlook (ical)



Example
Date Formats

Luckily, **pandas** can already handle many of these!

- Example:

```
2014-07-24T09:17:35+02:00

"2014-03-23 21:35 Uhr"

"01-04-2013"

21.04.2012,23:16

"07/21/2014 08:18"

"Europe/Amsterdam","25. 07. 2014 0:11"
```

```
pandas.read_csv("filename.csv“,
    parse_dates=True
    dayfirst=True
    parse_dates=[[0, 1]]
)
```

Analyzing and Visualizing the Data

Data analytics and visualization is essential to get insights

- What does the data mean?
- How does data correlate to some other data?
- What can I learn for my self?

Currently, many web sites for analytics and visualization arise

zenobase.com

addapp.io

pryv.com

traqs.me

fluxstream.org

Exploring the Data with Python, IPython, pandas, …

IP[y]: Notebook Fitbit Steps Last Checkpoint: Jul 25 23:48 (autosaved)

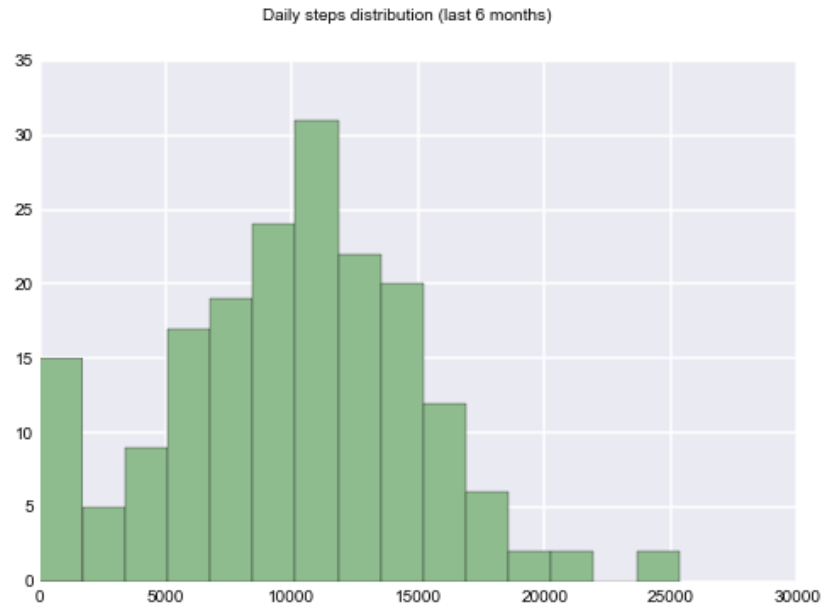
File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None

0
2014-01-21 2014-03-12 2014-05-01 2014-06-20

```
In [11]: steps.hist(bins=15, color='darkseagreen')
         plt.suptitle('Daily steps distribution (last 6 months)')
```

```
Out[11]: <matplotlib.text.Text at 0x1095c9190>
```



In [ ]:

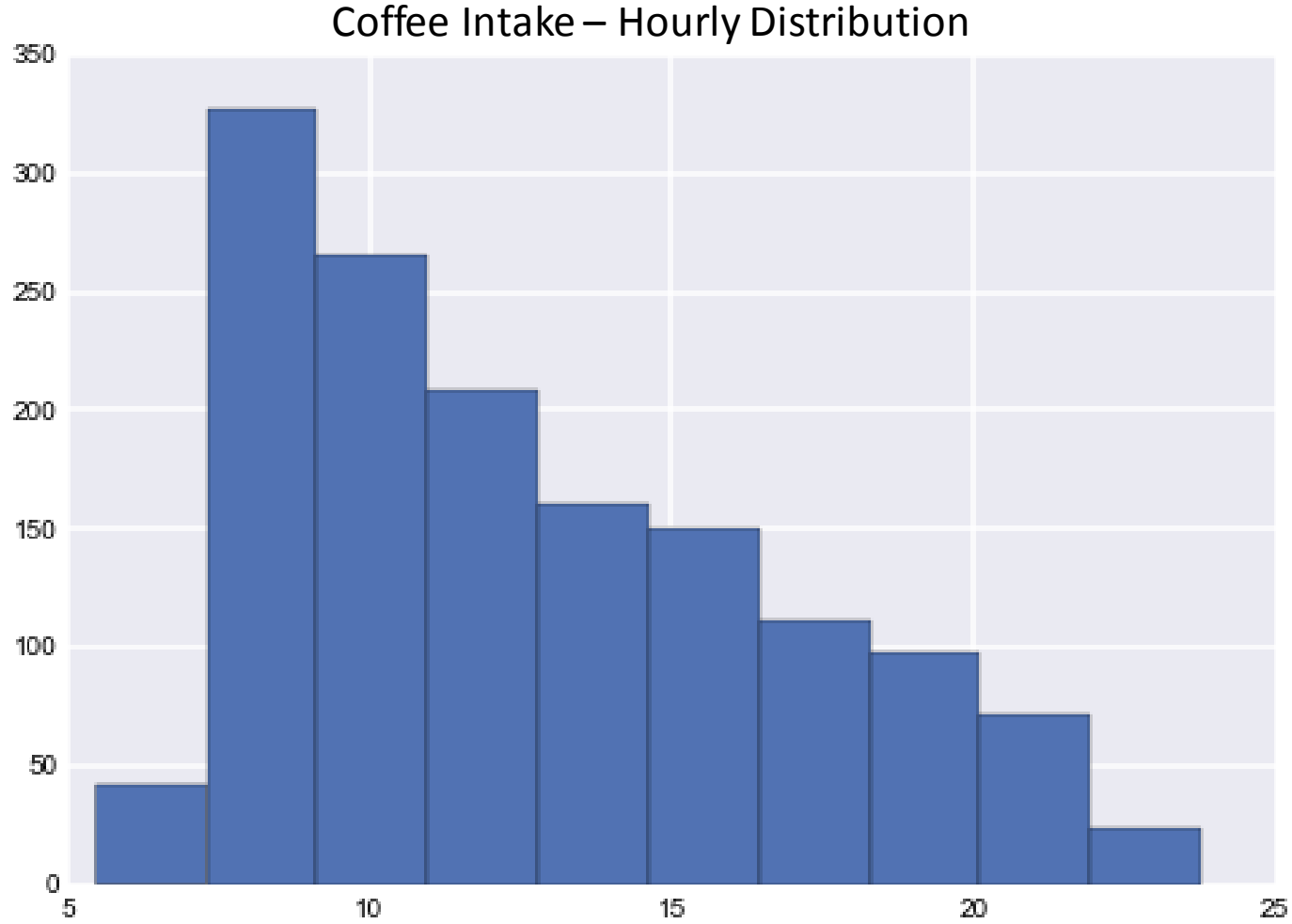
Sleep Efficiency vs Duration

Finding Answers within the Data

… even if I didn't know, that there was a questions ;)

New questions arise while tracking

New possibilities with all the data



Visualization and Data Analytics on Mobile Devices

Many people use only apps

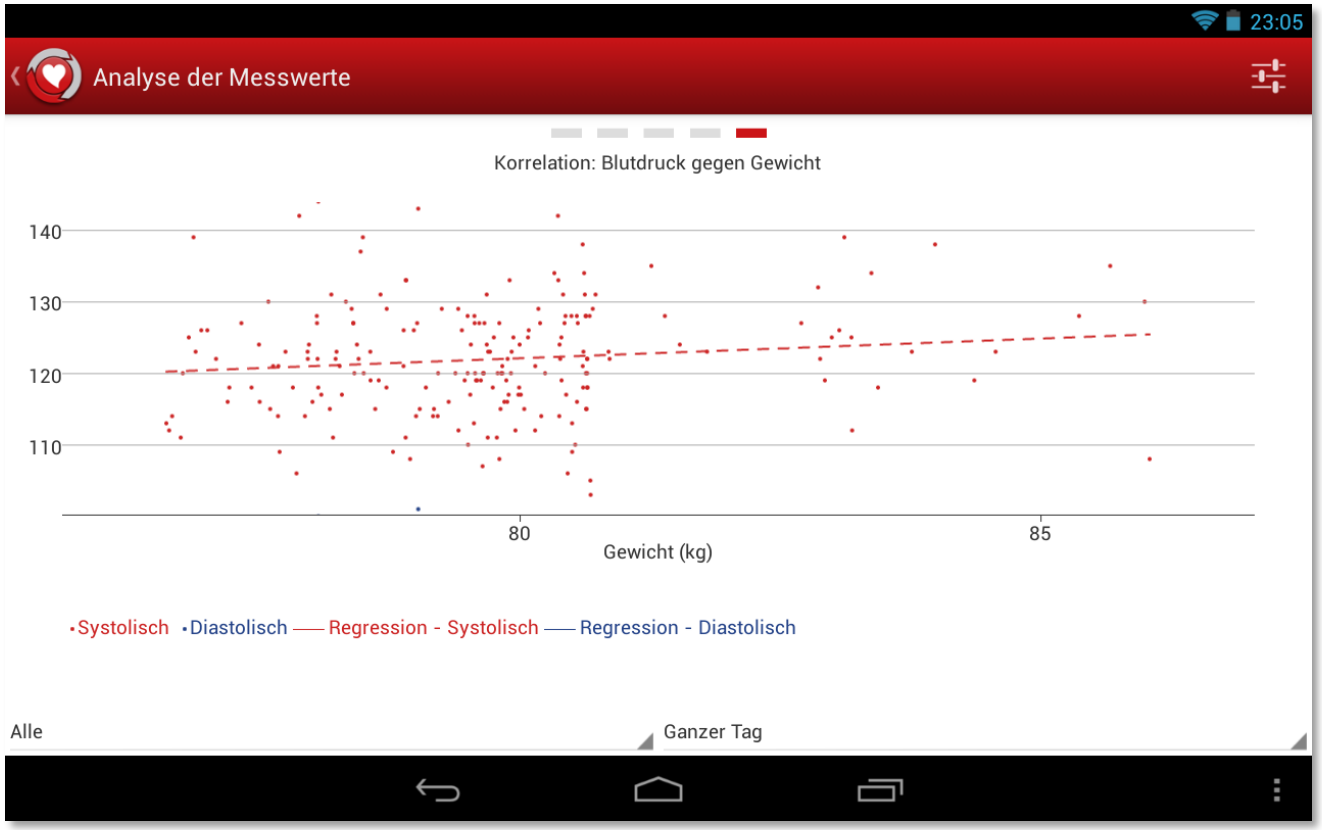
Quantified Self apps

- Should have visualization
- Should have some data analysis

As app developers, we would love to do that with Python

Requirements

- Good visualization components for mobile platforms (Android, iOS)
- Tools like NumPy, pandas, scikit-learn etc. available on mobile platforms



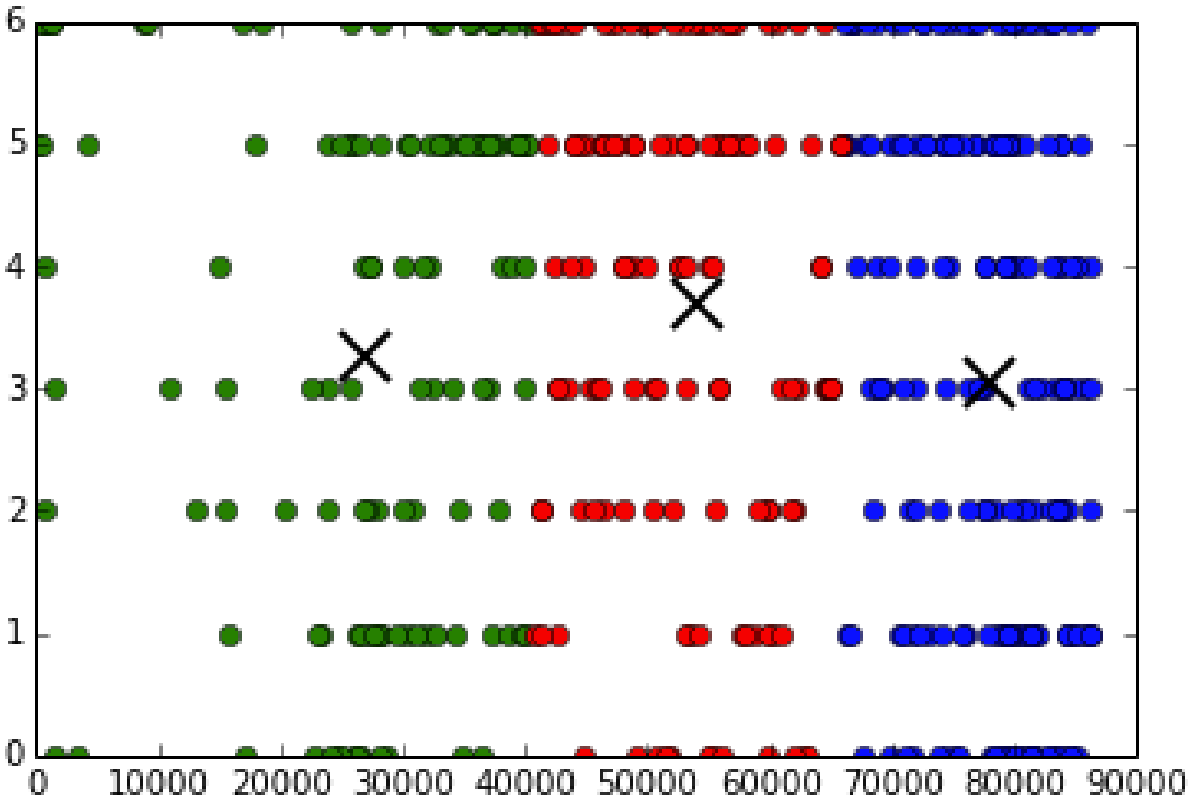
Machine Learning
Typical Use Cases (for Mobile Apps)

Predicting Behavior Patterns

- Detecting medication non-adherence
- Remind users for measuring blood pressure
- Advice users to get some sleep

Detecting stress and depression

- Stress detection based on heart rate variability (in combination with suitable Wearables)
- Depression detection based on communication behavior



Fit4Duty *(under development)*
Measuring cognitive performance

Pilots and others

- Sleep Tracking
- Psychomotor Vigilance Task (PVT)

Machine Learning for

- Predicting Fitness for Duty
- Predicting fatigue (during flights)

Hardware and smartphone app

Current Research

Provenance of QS data

- Trust and Traceability of data
- Privacy audits

Standardized APIs

Code Offloading

- Moving code for computations to the cloud

Thank You!

Questions?

Andreas.Schreiber@dlr.de

www.dlr.de/sc | @onyame