



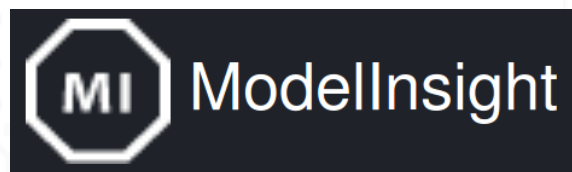
# Citizen Science with Python

## EuroPython 2018

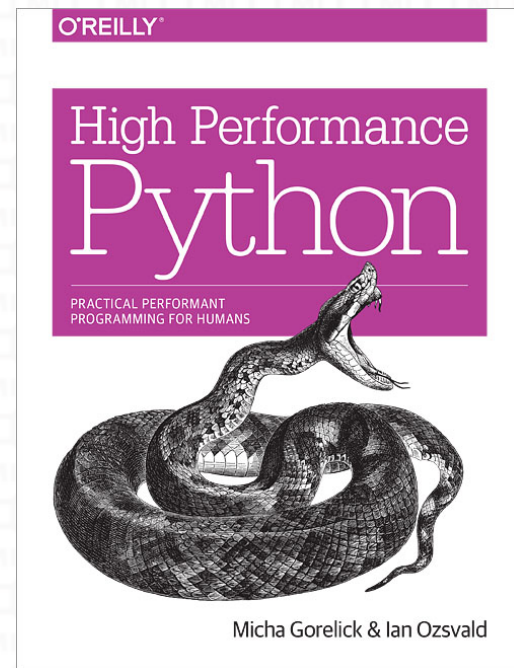
Ian Ozsvald @IanOzsvald ModellInsight.io

# Introductions

- I'm an engineering data scientist
- Coaching, training & consulting in AI + Data Science for 15+ years



Blog->[lanOzsvald.com](http://lanOzsvald.com)



# Goals today

---

- Short stories on Citizen Science
- Crowd-led demo with Jupyter Lab
  - Please check your wifi connection
- Ideas on how to start with your own data explorations
- References & links in the Appendix



# Macedonian air quality

---



The “smelly fog” in Skopje  
Gorjan Jovanovski (when 21)  
Government open data showed “4\* more pollution than Beijing  
and 20\* EU limits”  
[www.theair.app](http://www.theair.app)

Photo: <https://www.facebook.com/ilijoski>



# Political debate and change

- Initially a single JSON dump
- 1 mil. people in 1 mo.
- Visualisations in Parliament
- Challenged by Minister for Ecology
- How did Gorjan get the data?
- App goes from a single dump to frequent updates
- Drove government policy



# Macedonian air quality

---

- Located highly-polluting incinerator
- Got it “fixed”
- ESA Copernicus satellite collaboration

## Macedonia

**Paul Brown**, *environment correspondent*

Mon 21 May 2001 01.27 BST

## UK makes toxic gift to the Balkans

### Waste incinerator for Macedonia breaches EU regulations

The UK government has provided **Macedonia** with an incinerator to burn hospital waste that would be illegal under British law because of its toxic emissions.

<https://www.theguardian.com/environment/2001/may/21/globalwarming.europeanunion>

Ian.Ozsvald@ModelInsight.io @IanOzsvald[.com]

EuroPython 2018



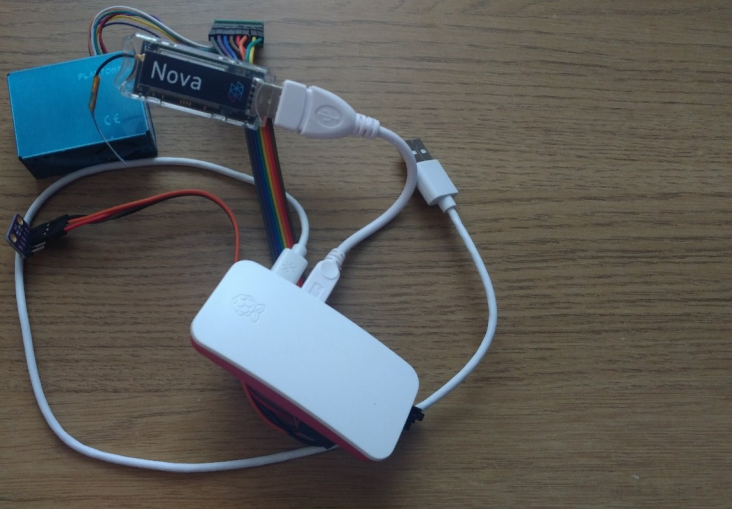
# Lessons

---

- Graph “unseen” data
  - Most people don’t know what JSON is
- Tell a story
- For change – recruit others to the project
- See Appendix for data sources

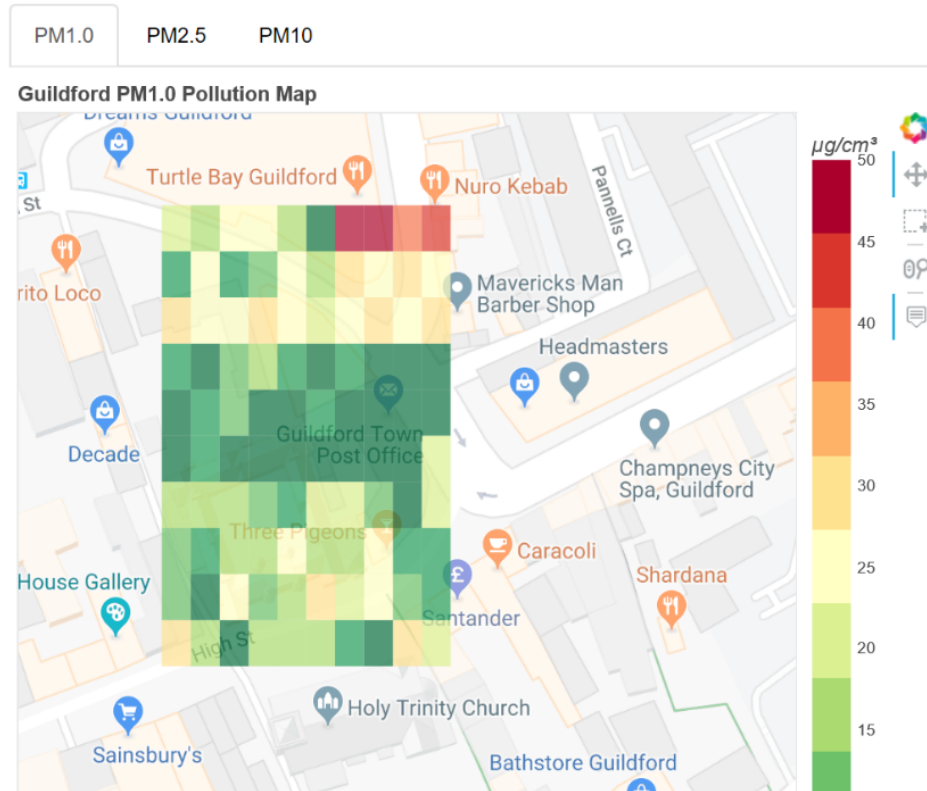


# “Monitoring Personal Air Quality”



“Air Quality and Python” this afternoon at EuroPython, Douglas Finch

PyLondinium 2018 talk by Robin & Oliver



[Github.com/OxygenLithium](https://github.com/OxygenLithium)

# Guess the weight - v1

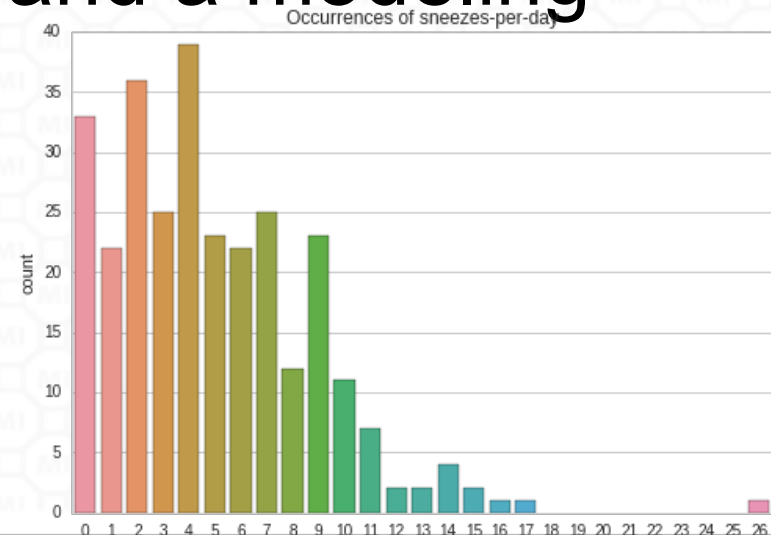
---

- Visit this URL <http://bitly.com/keynoteada1>
- No sign-in, please share the page with colleagues on your device
- Guess the weight of my dog *in kg*
- No other information (yet)
- We'll explore the results later

# Diganosing my wife's sneezing

---

- Emily sneezes *a lot*
- Can we gather data to diagnose correlated (and possibly causal) factors?
- We had to build an app and a modeling process

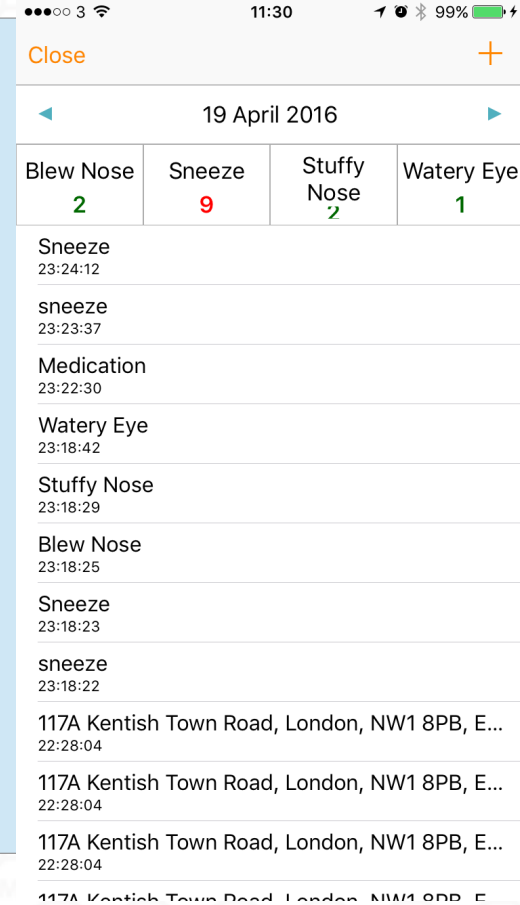
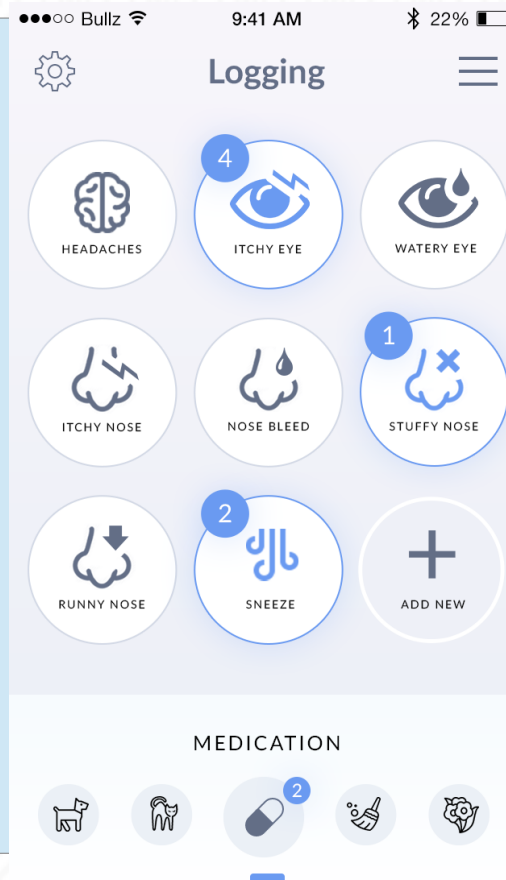


<http://bitly.com/keynoteada1>



# Diganosing my wife's sneezing

- iOS
- Event logs
- GPS trace
- Editable history
- Open Src
- >1yr old

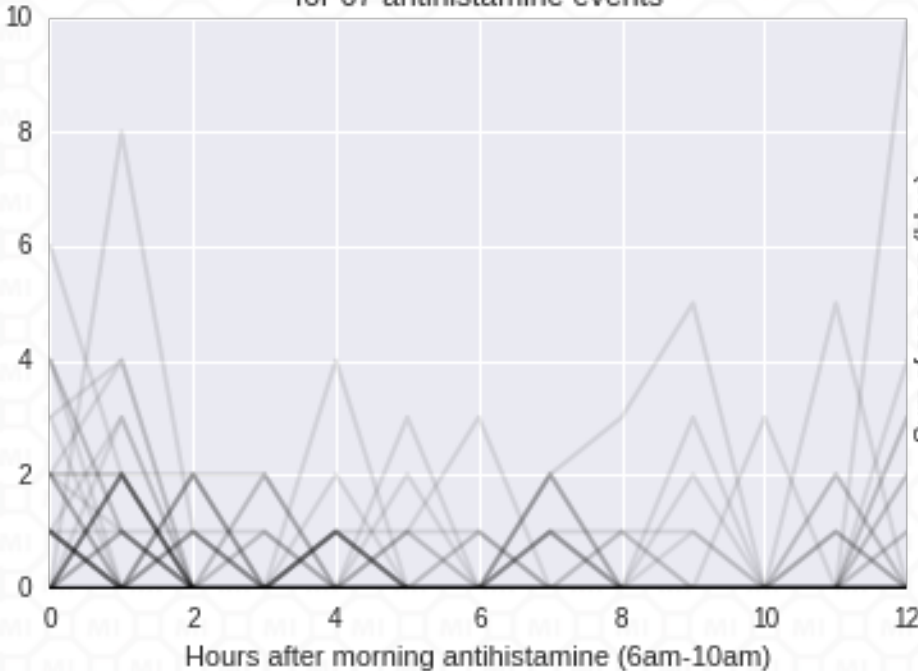


<https://github.com/radicalrobot/allergy-tracker>

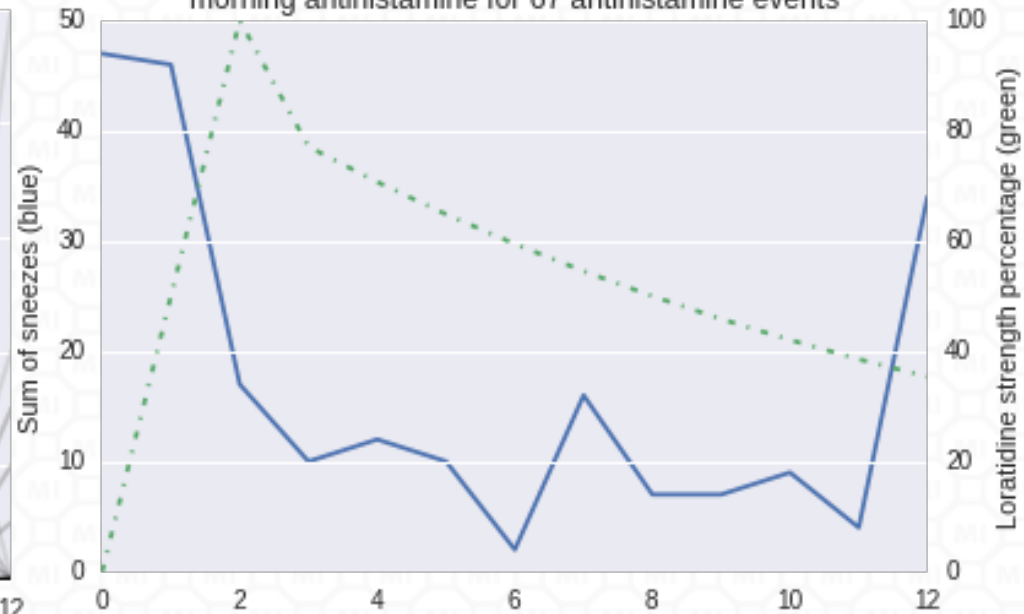
# Diagnosing my wife's sneezing

- “Single patient” antihistamine effect

Sneezes per hour relative to taking a morning antihistamine for 67 antihistamine events



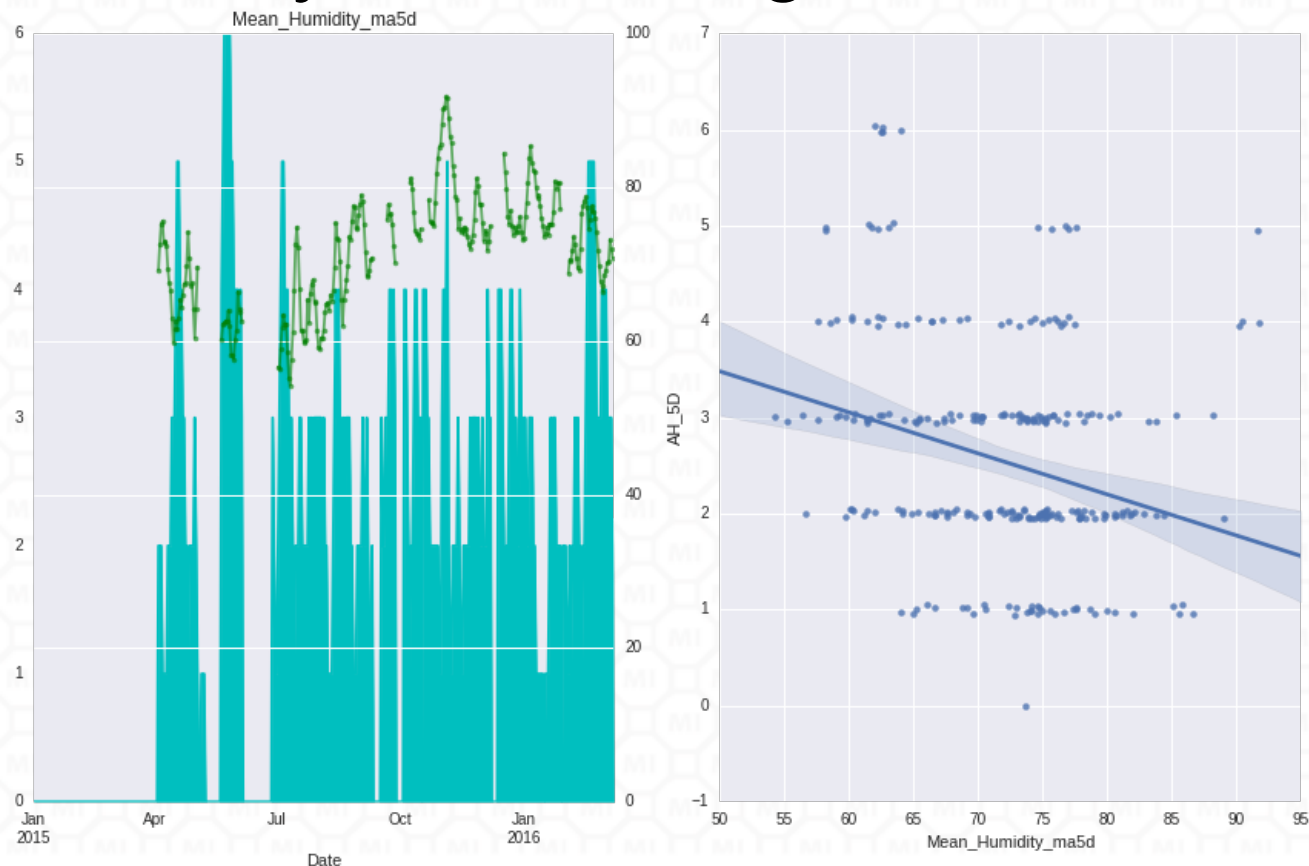
Total sneezes per hour relative to taking a morning antihistamine for 67 antihistamine events



<http://bitly.com/keynoteada1>

# Humidity was a predictive factor

- Humidity vs sneezing





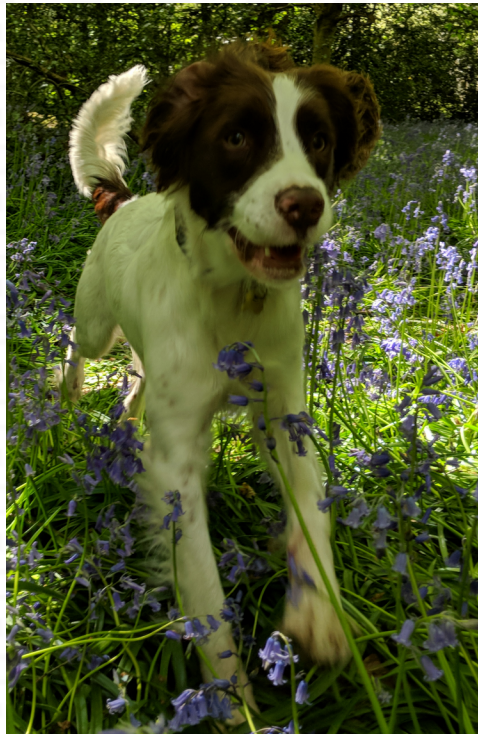
# Lessons

---

- Escalated to Kings College professor - “Great result! Clearly this is non-allergic, chronic persistent rhinitis”
- Suggested new treatment (Nasalcrom) sadly didn’t do anything interesting
- Graphing was enough to get a diagnosis, the machine learning was overkill
- See Bonzanini’s “Lies, Damned Lies” talk this afternoon

# Guess the weight (kg) - v2

- Visit this URL  
<http://bitly.com/keynoteada2>





# Updating outdated medical results

## Fighting Friedman's Curve: towards data driven childbirth assessment

**Anna Sztyber**<sup>1</sup>, Monika Sieradzan<sup>2,3</sup>, Beata Sztyber<sup>2</sup>

<sup>1</sup> WUT, [sztyber.anna@gmail.com](mailto:sztyber.anna@gmail.com)

<sup>2</sup> Medical University of Warsaw

<sup>3</sup> St Sophia Hospital in Warsaw

20 October 2017

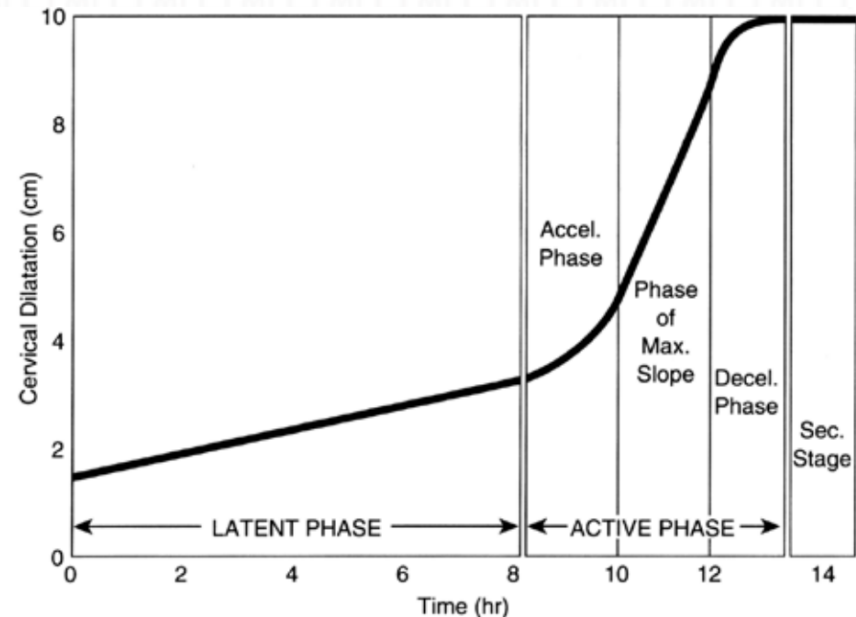
PyData Warsaw 2017





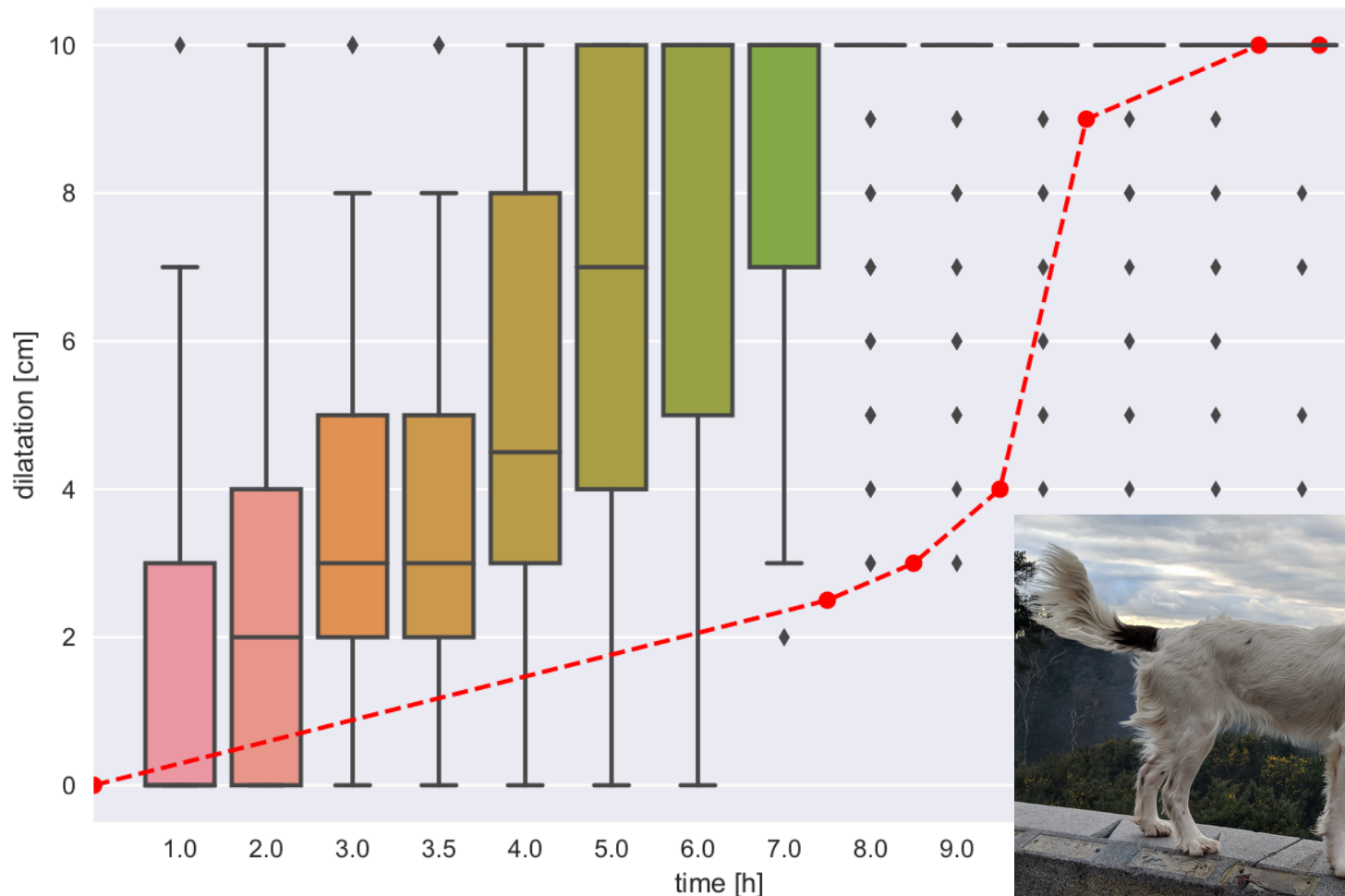
# Updating outdated medical results

- Friedman 1955
- Stages of labour by cervix dilation
- Different drugs, ages, technologies
- Significant medical decisions based on the result

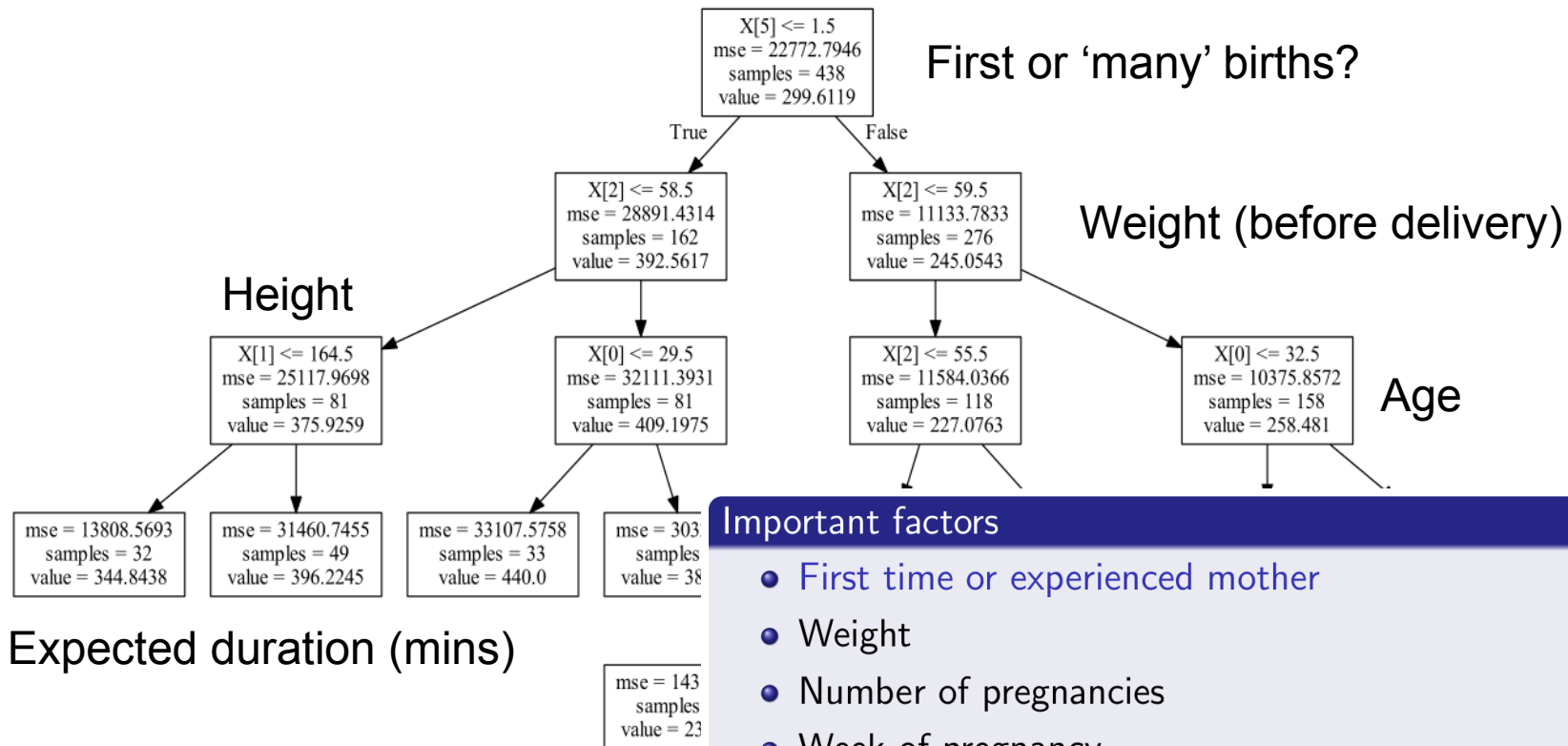


Friedman E: Labor: Clinical Evaluation and Management, 2nd ed. New York, Appleton-Century-Crofts, 1978

# Updating outdated medical results



# Actionable result



## Important factors

- First time or experienced mother
- Weight
- Number of pregnancies
- Week of pregnancy
- Height
- Age

# Lessons

---

- Check for out-dated assumptions
- Gather data to demonstrate what's missing
- Draw graphs to gain trust
- Produce interpretable advice

<http://bitly.com/keynoteada2>





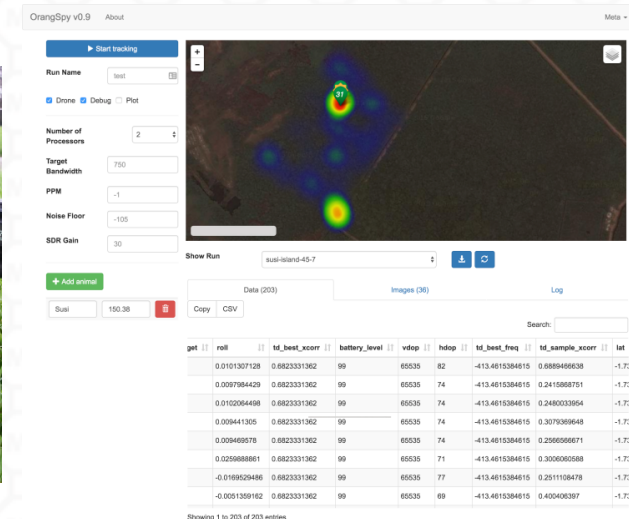
# Where are the Orangutangs?

- Dirk Gorissen – track 6 Orangutangs in 2000 km<sup>2</sup> of Bornean jungle
- Radio pings, drones and signal processing



# Where are the Orangutangs?

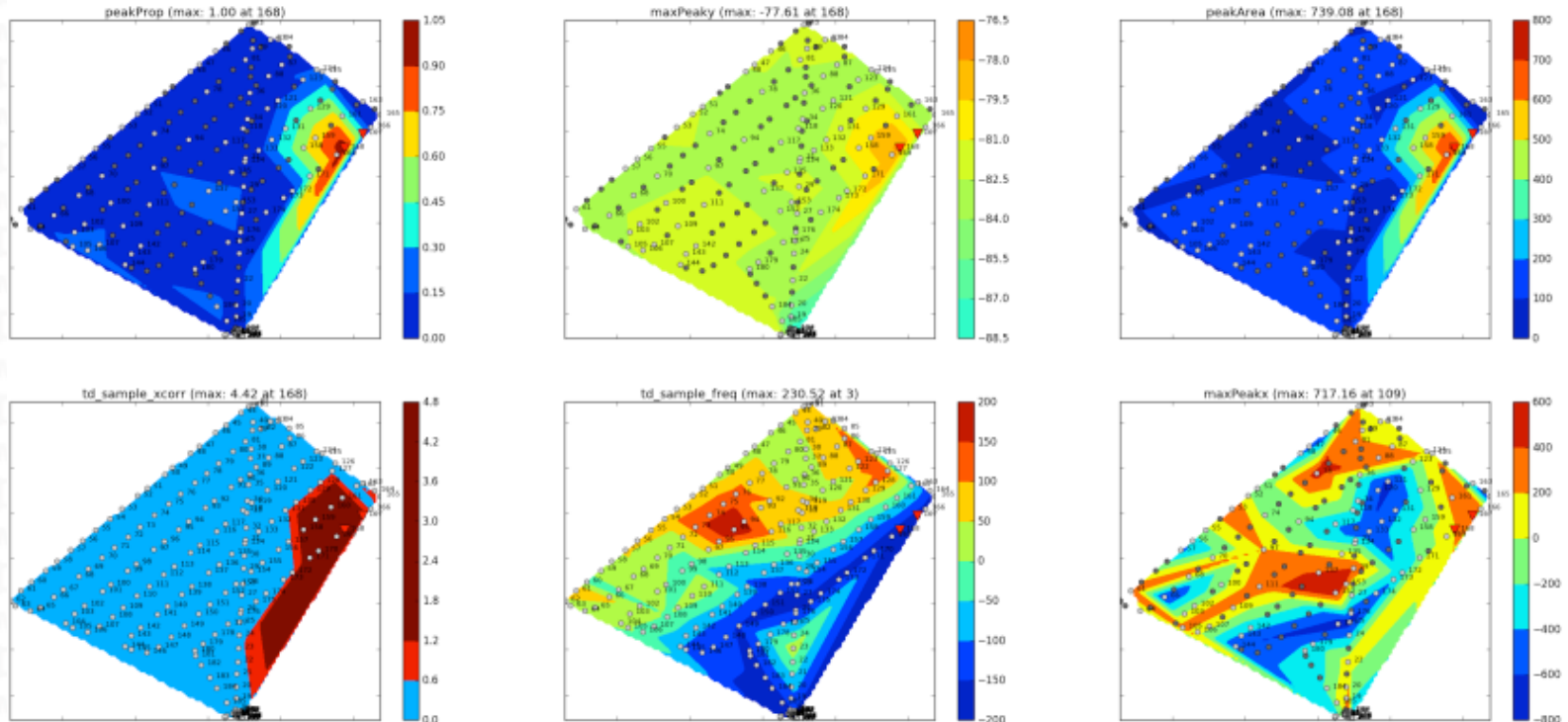
- Fixed search pattern, possibly detect many radio beacons
- Software defined radio, robust kit, post-return data processing





# Where are the Orangutangs?

- Tracking Susi (test runs) on her way to the test site



# Videos

---

- Demo with 2<sup>nd</sup> tracking camera
- Test flight in jungle
- Post-test-flight...not so good

<http://bitly.com/keynoteada2>





# Lessons?

---

- Hardware is hard
- Freeing up human time is valuable
- Expect to iterate a lot (so tackle something you can achieve in stages)

# Did we guess Ada's weight?

---

- Jupyter Lab live demo...
- Do we recreate Francis Galton's "vox populi" result?

# Closing...

---

- Collect, visualise and share your data
- Try datasets in Appendix
- Learnt something? Please send me a postcard!





# Closing...

---

- Please thank the volunteers & speakers!
- Write-up + more: <http://ianozsvald.com/>





# Appendix

---

- Gorjan Jovanovski TheAir.app <https://youtu.be/GQOmyKwhd4I>
- “The Data I Breathe”  
<https://pydata.org/amsterdam2018/schedule/presentation/16/>
- Anna Sztyber "Fighting Friedman's curve"  
[https://www.youtube.com/watch?list=PLGVZCDnMOq0oe0eD-edj\\_2CuBIZ938bWT&v=6qe2gtndJS4](https://www.youtube.com/watch?list=PLGVZCDnMOq0oe0eD-edj_2CuBIZ938bWT&v=6qe2gtndJS4)
- Ian “Solving sneezes”  
<http://ianozsvald.com/2016/05/07/statistically-solving-sneezes-and-sniffles-a-work-in-progress-report-at-pydata-london-2016/>
- Dirk Gorissen "Python vs Orangutang"  
[https://www.youtube.com/watch?v=vBHq3\\_C6uMM](https://www.youtube.com/watch?v=vBHq3_C6uMM)
- <http://robohub.org/wheres-susi-airborne-orangutan-tracking-with-python-and-react-js/>

# Appendix

---

- UK Government open data:  
<https://data.gov.uk/>
- Awesome public data sets:  
<https://github.com/awesomedata/awesome-public-datasets>
- 50 machine learning data sets:  
<https://blog.cambridgespark.com/50-free-machine-learning-datasets-part-one-government-data-portals-e39524ba601b>