



DATA SCIENCE

FOR

SOCIAL GOOD



THE BEST MINDS OF MY GENERATION  
ARE THINKING ABOUT HOW TO MAKE  
PEOPLE CLICK ADS. THAT SUCKS.

— Jeff Hammerbacher (Co-founder Facebook)

SOCIAL ORGANIZATIONS NEED TO  
ALLOCATE SCARCE RESOURCES FOR  
MAXIMUM IMPACT

- predicting where services will be needed
- prioritizing resources based on expected impact
  - forecasting trends and changes
- identifying supporters likely to donate

# BASICALLY REGRESSION PROBLEMS

Social organizations often don't have

DATA SCIENTISTS



WHAT IF DATA SCIENTISTS WOULD

DONATE  
THEIR SKILLS?



This is already happening...\*

\* ...but not so much in Germany



A GUIDED TOUR THROUGH THE  
DATA SCIENCE FOR SOCIAL GOOD  
MOVEMENT

# DSSSG FELLOWSHIP

- supported by the Eric & Wendy Schmidt Foundation
  - 6 month fellowship
- currently in 3rd year (started 2013)
  - Partners: NGOs, Governments

<http://dsssg.io/>

# EXAMPLES (PROJECT HEADLINE COPY&PASTE)

- World Bank Group – Prediction & Identification of Collusion in International Development Projects
- Chicago Public Schools – Student Enrollment Prediction for Budget Allocation
- Pecan Street , WikiEnergy – Building Open Source Tools to Analyze Smart Meter Data

# MORE EXAMPLES

**visit** <http://dssg.io/projects/>

# DSSG ATLANTA

- like DSSG Chicago
- mainly funded by Oracle and Georgia Tech
  - started 2014 (one year after Chicago)

<http://dssg-atl.io/>

# DSSG WASHINGTON

- **University of Washington just announced their DSSG Summer Program**

**[http://escience.washington.edu/what-we-do/  
data-science-for-social-good](http://escience.washington.edu/what-we-do/data-science-for-social-good)**



# BAYES IMPACT (YC-BACKED)

Bayes Impact is a nonprofit that deploys data scientists to solve big social problems with civic and nonprofit organizations

- founded 2014
- started as (full-time-)fellowship
- now hiring long-term employees

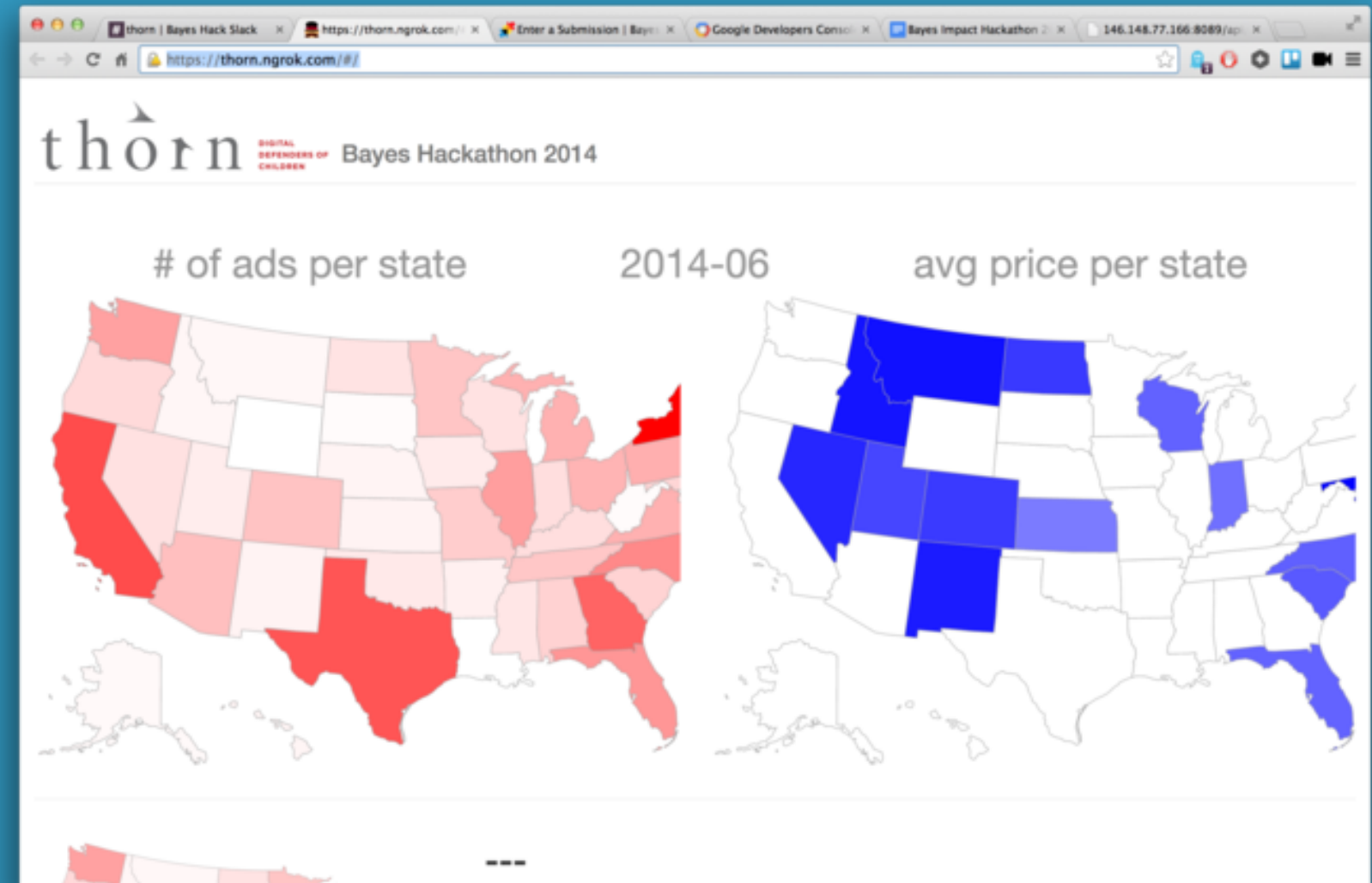
# EXAMPLES (PROJECT HEADLINE COPY&PASTE)

- Increasing Graduation Rate And Optimizing Class Offerings For UC Riverside
- Improving Outcomes For Emotionally And Behaviorally Challenged Children With Youth Villages
- Stratification Of Parkinson's Disease Patients
- Optimizing Ambulance Response Times In Sf

# BAYES HACK NOV 15TH

One weekend,  
impact the world

[http://  
bayeshack.challengepost.com  
/submissions](http://bayeshack.challengepost.com/submissions)



# DRIVENDATA

- started this year (2014)
- currently 4 open competitions

<http://www.drivendata.org/>

# MORE OPPORTUNITIES TO GET INVOLVED

<http://blog.datalook.io/definitive-guide-data-science-good/>

# DATAKIND

We're tackling the world's biggest problems through data science.

— <http://www.datakind.org>

DataKind connects charities with data scientists by organizing two-day data dives where those data scientists help solve the charities' data problems.





## Jake Porway at TEDxMontreal

# EXAMPLES

# ANALYZING POVERTY THROUGH SATELLITE IMAGERY

**DataKind helped GiveDirectly**  
– an NGO making  
unconditional cash transfers  
to poor households via mobile  
phones in Kenya and Uganda<sup>1</sup>  
– to identify especially needy  
villages through satellite  
image analysis<sup>2</sup>.



<sup>1</sup> [http://www.ted.com/talks/joysunshouldyoudonate\\_differently](http://www.ted.com/talks/joysunshouldyoudonate_differently)

<sup>2</sup> <http://www.datakind.org/projects/using-the-simple-to-be-radical/>

VIEW THE PRESENTATION  
OR READ THE PAPER



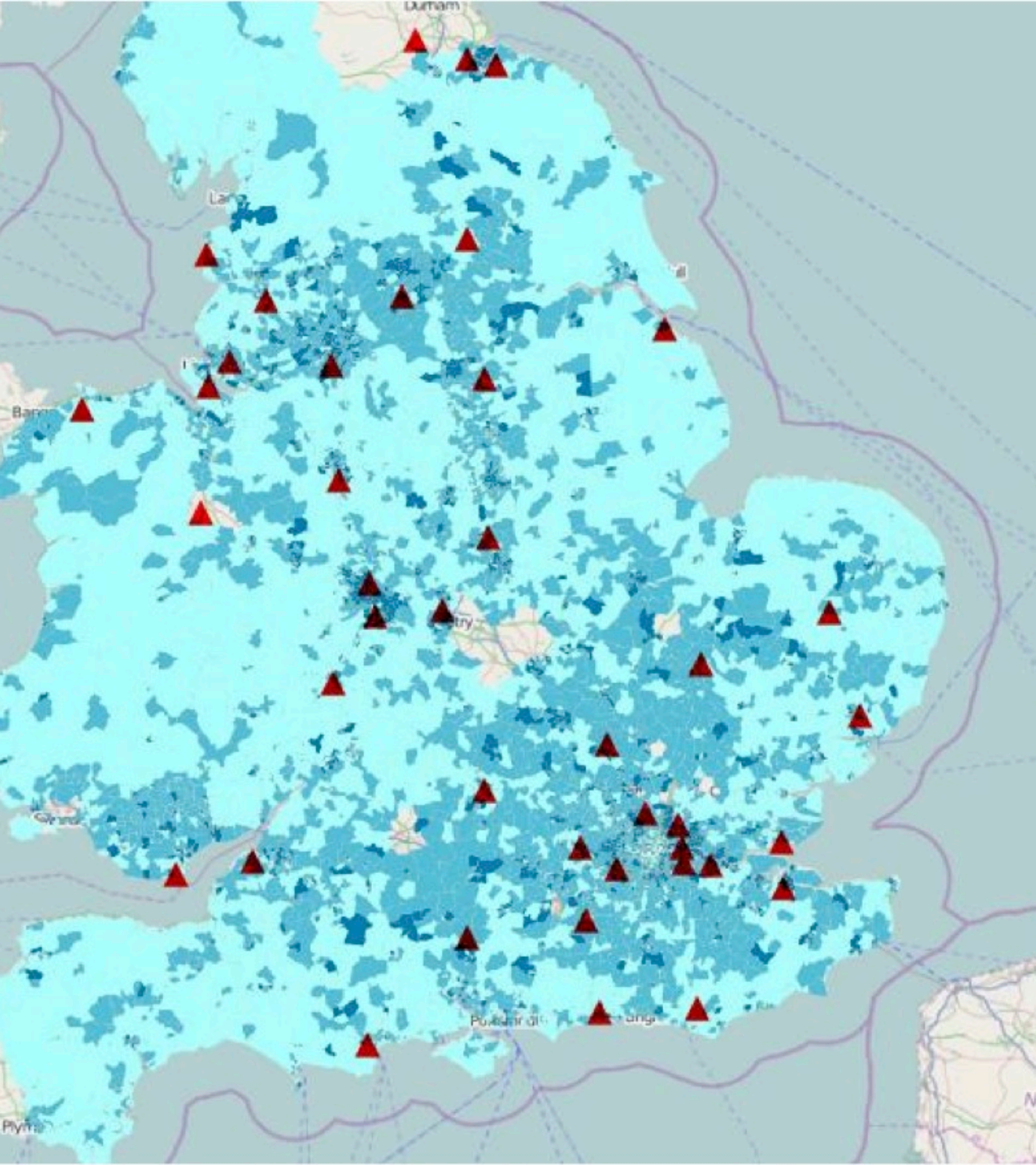
# EARLY WARNING SYSTEM FOR HUMAN RIGHTS ABUSES

To help prioritize the many calls for help reaching Amnesty International's Urgent Action Network DataKind volunteers have created a predictive model that analyzes messages for potential escalation.<sup>34</sup>

<sup>3</sup> <http://www.datakind.org/projects/using-predictive-analytics-to-prevent-human-rights-abuses/>

<sup>4</sup> [http://www.washingtonpost.com/business/on-it/amnesty-international-considers-using-big-data-to-predict-human-rights-violations/2013/11/22/3f4f1a1e-5388-11e3-a7f0-b790929232e1\\_story.html](http://www.washingtonpost.com/business/on-it/amnesty-international-considers-using-big-data-to-predict-human-rights-violations/2013/11/22/3f4f1a1e-5388-11e3-a7f0-b790929232e1_story.html)





# PREDICTING DEMAND FOR CHILDREN'S HOSPICES

Combining data from Shooting Star Chase, public data about the hospice and healthcare sector and demographic data DataKind volunteers calculated predicted demand against hospice capacity to reveal areas of possible shortage.<sup>5</sup>

<sup>5</sup> <http://www.datakind.org/finding-30000-missing-children/>



**Most of DataKinds projects have been tackled by volunteers on 2-day data dives.**

WHAT'S A

DATAADDITIVE?

# BEFORE: PREPARATION OF THE DATA

(by voluntary data ambassadors in collaboration with the challenge partner – starting ~2 month before the data dive)

- anonymization/pseudonymization
  - cleaning/fixing
- ensuring proper (machine readable) data formats

BEFORE:

PREPARATION OF THE  
QUESTIONS

ANY DATA SCIENTIST WORTH THEIR  
SALARY WILL TELL YOU THAT YOU  
SHOULD START WITH A QUESTION, NOT  
THE DATA.

— Jake Porway in <https://hbr.org/2013/03/you-cant-just-hack-your-way-to/>

# THE DATA DIVE

- Challenge partners pitch their problems
- Volunteers create analyses, models and visualizations (led by data ambassadors) in two intense days of hacking
- solutions are being presented at the end



# AFTERMATH: IMPLEMENTATION

**Social organizations still don't have the expertise: data ambassadors must help implement the solutions**

# AFTERMATH: IMPACT ANALYSIS

Did it help at all?

**There is currently no organization in Germany comparable to DataKind.**

There is currently no organization in Germany comparable to DataKind.



© RLY?



Open Knowledge  
Foundation

{ODE} for  
Germany

WE WANT TO START THE  
DATA SCIENCE FOR SOCIAL GOOD  
MOVEMENT IN GERMANY

# WHO WE ARE?

- Daniel Kirsch
- Marit Brademann
  - Jana Kludas
- Richard Lawrence
  - Georg Walther



# SUPPORTERS

- Klaas Bollhöfer,  
**Chief Data Scientist**  
@ The Unbelievable Machine Company
- Adam Drake,  
**Chief Data Officer @ Skyscanner**
- Dr. Alexander Weiß  
**Head of Data Analytics @ Trademob**

WHAT WE NEED

# WE NEED DATA AMBASSADORS (HOW DATAKIND CALLS THEM)

- **to prepare data before data dives**
  - **lead teams at data dives**
- **help with the implementation afterwards**

# WE NEED PROBLEMS

The international of the **Data Science for Social Good**-movement shows that data scientists are eager to donate their skills.

Social organizations need to understand how we can help them. Are you in contact with NGOs?  
Spread the word!

# CONTACT

<http://dssg-berlin.org/>  
[@dssgber](mailto:@dssgber)

**Daniel Kirsch**  
[mail@danielkirs.ch](mailto:mail@danielkirs.ch)

THANK YOU!

QUESTIONS?

# LINKS:

- <http://datakind.org>
  - <http://dssg.io>
  - <http://dssg-atl.io>
- <http://bayesimpact.org>
  - <http://codefor.de>
  - <http://datalook.io>

# ATTRIBUTION

- Foto of Jeff Hammbacher **by** Fred Brenenson  
licensed under **CC BY 2.0**