



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

- identifying supporters likely to donate
- predicting where services will be needed
 - predicting impact of campaigns
 - forecasting trends and changes

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 in Germany (afaik)

A GUIDED TOUR THROUGH THE DATA SCIENCE FOR SOCIAL GOOD MOVEMENT

DSSSG FELLOWSHIP

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

<http://dssg.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

SELECTED DSSG EXAMPLE

Partner: Nurse-Family Partnership

**DSSG analyzed who was „benefiting the most“
from NFP's Program.**

<http://dssg.uchicago.edu/2014/08/27/nfp-undefinable-unmeasurable.html>

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/>

BAYES IMPACT (YC-BACKED)

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

- started this year
- 12-month Fellowship

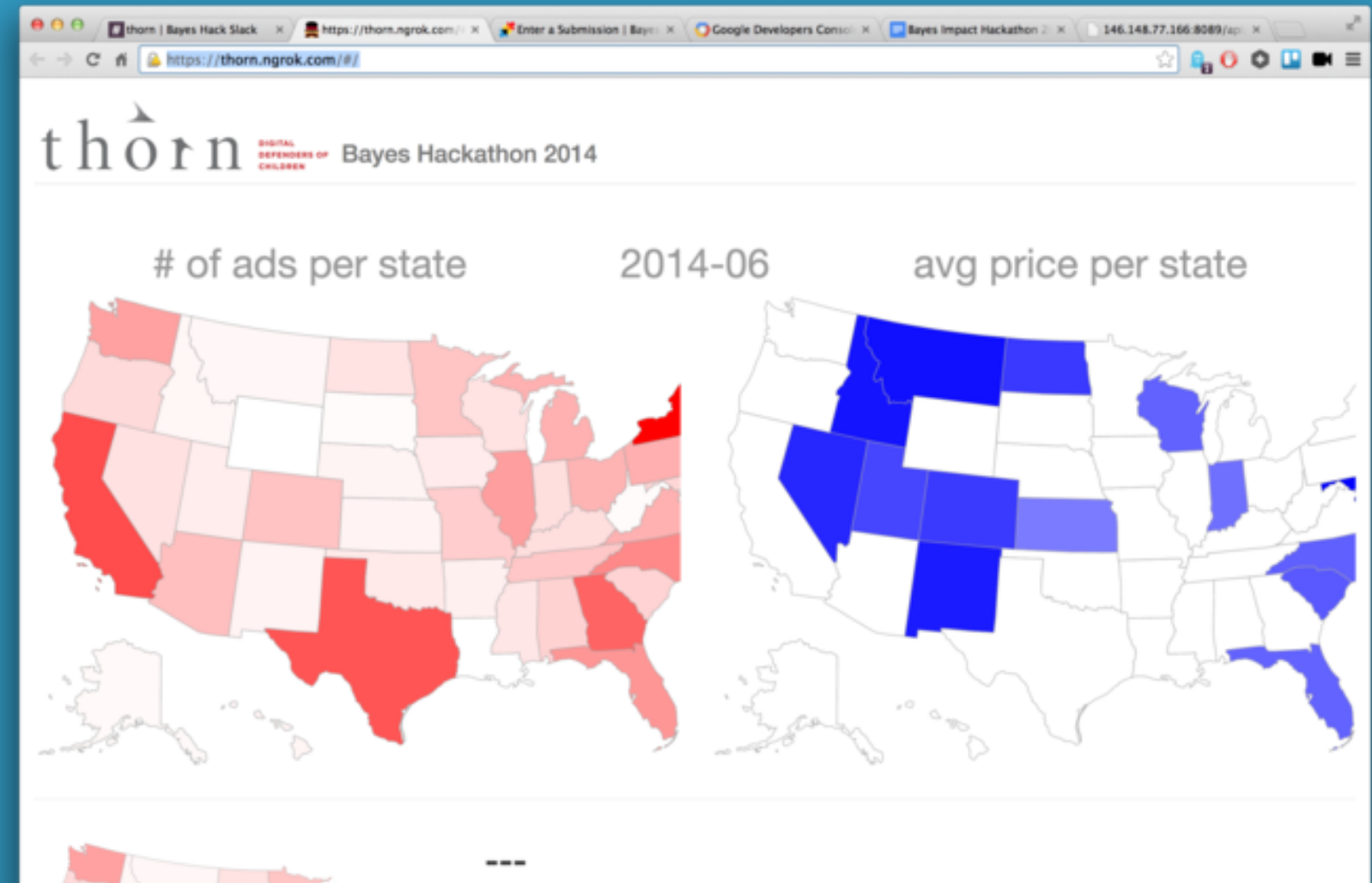
EXAMPLES (PROJECT HEADLINE COPY&PASTE)

- 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)



THORN FELLOWSHIP

drive technology innovation to fight child sexual
exploitation

— <http://www.wearethorn.org/thorn-innovation-lab>

<http://www.wearethorn.org/thorn-innovation-lab/>

DRIVENDATA

- started this year (2014)
- currently 3 competitions

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

KDD 2014

had a

Workshop on Data Science for Social Good



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 Uganda¹
– to identify especially needy
villages through satellite
image analysis².



¹ http://www.ted.com/talks/joysunshouldyoudonate_differently

² <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.³⁴

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

⁴ 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.⁵

⁵ <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

DATA DIVE?

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



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

WHO WE ARE?

- Daniel Kirsch
- Marit Brademann
- Richard Lawrence
- Tobias Pfaff (of **dataforgood.co**)
 - You?

SUPPORTERS

- Klaas Bollhöfer,
Chief Data Scientist,
The Unbelievable Machine Company
- Adam Drake,
Chief Data Scientist,
Zanox

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

Daniel Kirsch
mail@danielkirs.ch
@kirel

No website yet... No name...
just contact me!



THANK YOU!
QUESTIONS?

LINKS:

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

ATtribution

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