Delta Analytics Community Celebration Lightning Talks

Agenda - Welcome!

3:00 PM - Networking + Food & Drink

3:20 PM - Lightning Talks

4:20 PM - Poster Session, Networking + Food & Drink

Lightning Talks

- Delta Analytics - Who we Are
- Social Justice Collaborative
- Democratizing Machine Learning for Good
- Rainforest Connection
- BUILD

Welcome to our 2017 community celebration!

DATA
INSTITUTE

UNIVERSITY OF SAN FRANCISCO

Delta Analytics

Our Impact

Sara Hooker - Executive Director at Delta Analytics, Brain Resident at Google

Data can change the world.

Delta fills the skill gap by enabling nonprofits to accelerate their impact and building long term technical capacity.

Our Impact

29 projects with nonprofits and social impact organizations

70+ Fellows volunteering part-time over 3 years

Over **15,000** hours donated

$0.00 charged for services

19 US and 10 International projects (Tanzania, UK, Kenya, and more)

Some of our nonprofit grant recipients

Community Engagement

- United Way — United Way of the Bay Area
- Tenants Together
- fsd — foundation for sustainable development
- SJC social justice collaborative

Education

- Roots of Success
- eneza education — spreading education everywhere
- KUOW.ORG
- City Year

Economic Development

- Village Enterprise
- TechnoServe — Business Solutions to Poverty
- BUILD — In Business to Learn
- sama

Environmental

- LivelyHoods
- BioLite
- Karibu
- BGCI — Plants for the Planet
- Rainforest Connection

Almost all of our fellows work full time.

facebook

STITCH FIX

AON

udemy

NETFLIX

twitter

Google

Brattle
THE POWER OF ECONOMICS

CORNERSTONE RESEARCH

3Q DIGITAL

... and more.

Delta Data Fellows

Partner with a non-profit grant recipient for 6 months.

&

Delta Teaching Fellows

Democratize access to machine learning education by building technical capacity around the world.

In 2018… we will be partnering with the Data Institute to involve graduate students passionate about leveraging data for good as fellows!

DATA INSTITUTE

UNIVERSITY OF SAN FRANCISCO

2017... full of data, fun and impact.

We started the Delta Teaching Fellows program!

In 2018, we taught our pilot course in Nairobi, Kenya.

We intend to continue teaching all over the world to empower communities to use data for good.

We worked with 7 incredible grant recipients!

RAINFOREST CONNECTION

CITY YEAR

BUILD
www.build.org

SJC social justice collaborative

BGCI
Plants for the Planet

TECHNOSERVE
BUSINESS SOLUTIONS TO POVERTY

sama SOURCE

40 incredible data and teaching fellows gave their time and talent to Delta.

Site visits to understand the work of our nonprofit partners.

Delta Analytics kickoff in January

Fellows worked in small teams for 6 months.

Hackathons once a month.

Team monthly social budget.

We hosted incredible speakers at our monthly hackathons.

Speakers from leading nonprofits using data to make an impact: Watsi, Khan Academy, Kiva, Mozilla.

We gave back to the wider community at conferences and events.

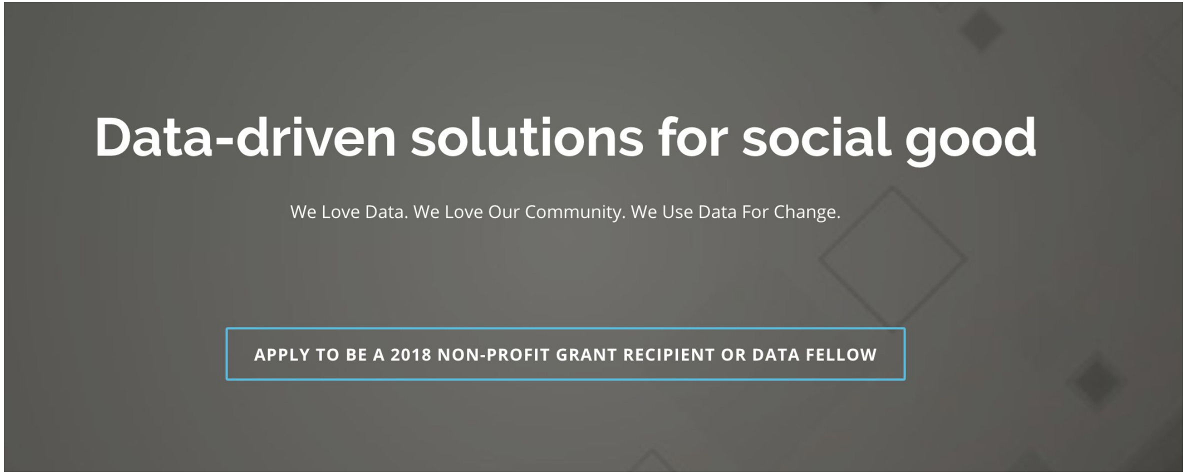
We spoke at Northwestern Global Engagement Summit, USF lightning talks, MLConf, Open Data Science Conference.

Looking forward to 2018!

Teaching fellows

Data fellows

Nonprofit grant recipients



Applications are already live! You can apply on our website at deltanalytics.org.

Deep dive into some of our projects!

- Social Justice Collaborative
- Democratizing Machine Learning for Good
- Rainforest Connection
- BUILD

Social Justice Collaborative

Applying our skill set to the social sector

Roberto Sanchis Ojeda - Project Lead at Delta Analytics,
Senior Data Scientist at Netflix

Amanda Su - Data Science Fellow at Delta Analytics,
Senior Analyst at Analysis Group

Delta Team

Roberto Sanchis Ojeda
Senior Data Scientist
Netflix

Amanda Su
Senior Analyst
Analysis Group

Zanele Munyikwa
Research Fellow
Stanford University

Bharath Srikanth
Data Scientist
Uber

Our Partner: Social Justice Collaborative

SJC social justice collaborative

Gautam Jagannath & Emily Abraham
Founders and Directing Attorneys

SJC's Data Question

SJC's Mission: to **protect and advance the rights of immigrants** and their families through legal representation in immigration and criminal court

"Given the resources that we have, how can we offer **cheaper services to more people in a consistent way**?"

Adapting to our partner's needs

In our day jobs:

- **Large and rich data**
  - Every show showed to each Netflix user when they log in
  - Microseconds of trading data

- **Proprietary software**
  - Statistical packages
  - Internal data processing and analysis tools

- **Technical capability**
  - Dedicated teams of data scientists, economists, etc.

Big data-biased expectations vs. small nonprofit reality

- **A new pricing model for SJC**
  - *Expectation:* Sophisticated machine learning model
  - *Reality:* No training data set available → human-driven model

- **Revenue forecast**
  - *Expectation:* Time series analysis with short- and medium-term forecasting
  - *Reality:* Static model based on most clients

- **Model visualization**
  - *Expectation:* R/Shiny dashboard with multiple input fields
  - *Reality:* Excel spreadsheet with hidden information

3 reasons why: Wildly applicable to smaller companies

3 reasons why: Working with people with a good heart

3 reasons why: Helping those in need

And, of course, we finished the project, see our poster!

	Case Type	Discount	Amount
Base Service	Asylum		1,000.00
Extra Services	**Include it?**		
Derivatives	Yes		100.00
Removal defensive	No		0.00
Custody	Yes		300.00
Mandatory Bar	No		0.00
Relationship Status	Married Parent	10%	
Criminal Record	Yes	0%	
Time in US	5 years or less	5%	
Family in US	No	5%	
Indigenous Language	Yes	5%	
Experienced Abuse	No	0%	
	Calculated Discount	25%	
	Actual Discount	**25%**	(350.00)
	TOTAL		**$1,050.00**
	MONTHLY PLAN		$100
	DEPOSIT		$350.00
	RESIDUE FOR PLAN		$700.00
	MONTHS OF PAYMENT		7.0

Conclusion

To any data professional in the audience: If you are thinking about applying to Delta's fellow program, do it! We have given you plenty of reasons to do so.

To all the other fellows and to Delta: Thanks for letting us contribute and for all your efforts, it made it worthwhile.

To SJC: We wish you the best in your quest for justice to those that need it the most, we hope that our small contribution gets you one step closer to your goals.

Rainforest Connection

Deep Learning for Good

Steven Troxler - Project Lead at Delta Analytics,
Data Platform Engineer at Stitch Fix

Sean McPherson - Data Science Fellow at Delta Analytics,
Data Analyst at Northrop Grumman

Delta Team

Sara Hooker
Machine Learning
Researcher at Google

Sean McPherson
Data Analyst
Northrop Grumman

Steven Troxler
Data Platform Engineer
Stitch Fix

Cassandra Jacobs
Data Scientist
Stitch Fix

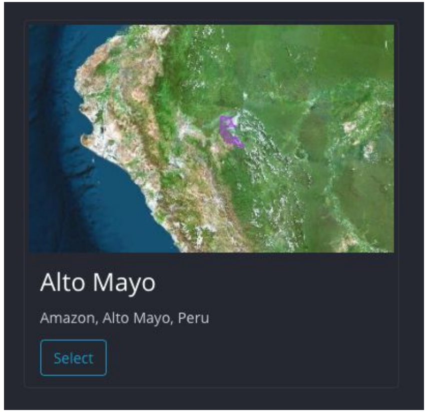
Rainforest Team

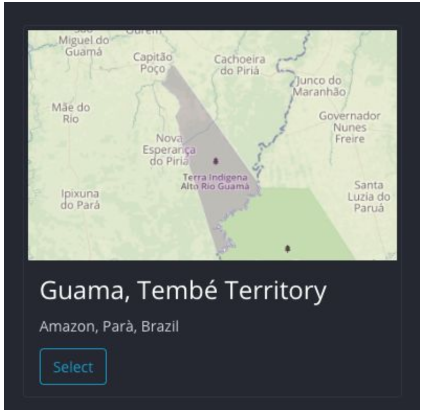
Close partnership with Rainforest Eng team
Stefan Zapf and Christopher Kaushaar

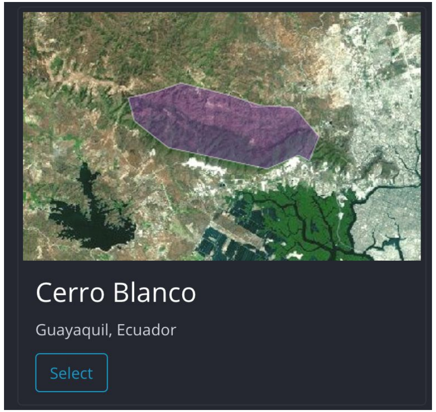
Rainforest Connection (RFCx)

- collects recycled cellphones, modifies them into guardians with microphones and solar panels
- deploys guardians to rainforests around the world
- the cellphones stream audio to the RFCx cloud
- machine learning models detect events such as chainsaws

Streaming audio data from rainforests in Ecuador, Peru and Brazil



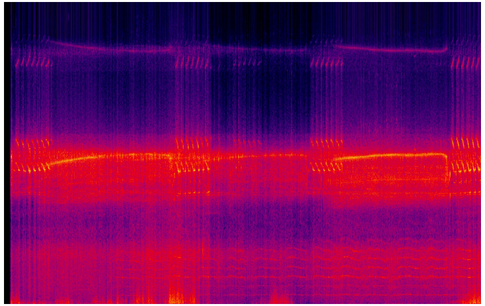




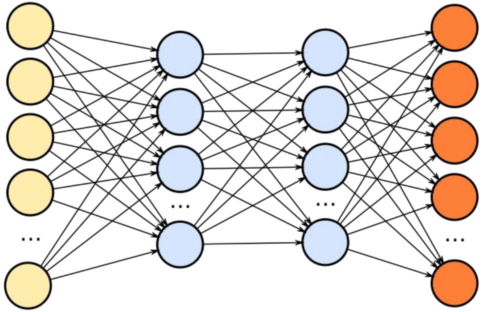
Additional deployments are made available as conservationists and research partners request them.

approach: turn audio detection into an image classification problem. apply a deep learning architecture.

→



→



Audio streamed from conservation sites in Ecuador, Peru and Brazil

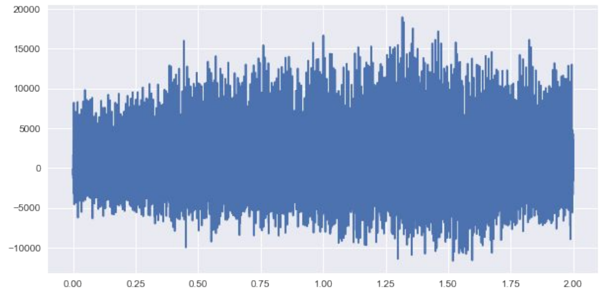
Convert audio to features, typically a spectrogram (visual way to represent the signal strength of a sound)

Fit machine learning models, typically feed-forward neural networks, to the features

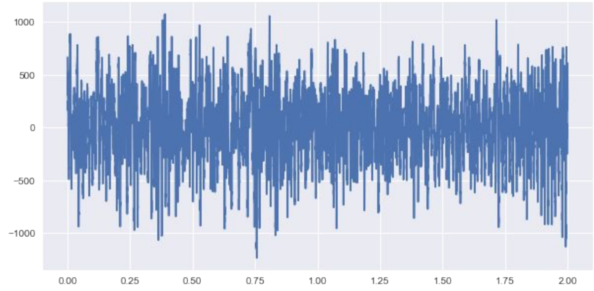
Example Spectrograms

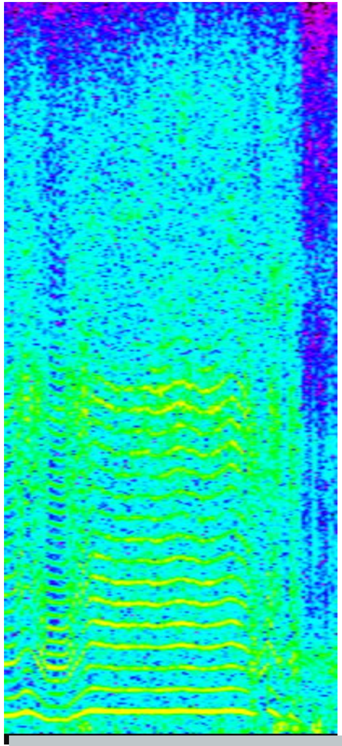
frequency vs time

Chainsaw Audio

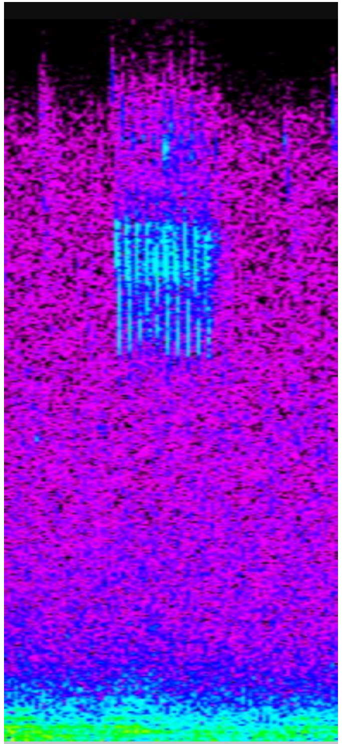


Birdsong Audio





Chainsaw Spectrogram



Birdsong Spectrogram

Rainforest Connection Goals:

- Research how to use multiple guardians
  - Separation of chainsaw sound from noise
  - Determine the direction and distance of chainsaw sound

- Improve model training capabilities
  - Automate data downloads
  - Make data processing more efficient
  - Enable development of new features
  - Allow research into data augmentation

Building a model training prototype

Challenges with model training

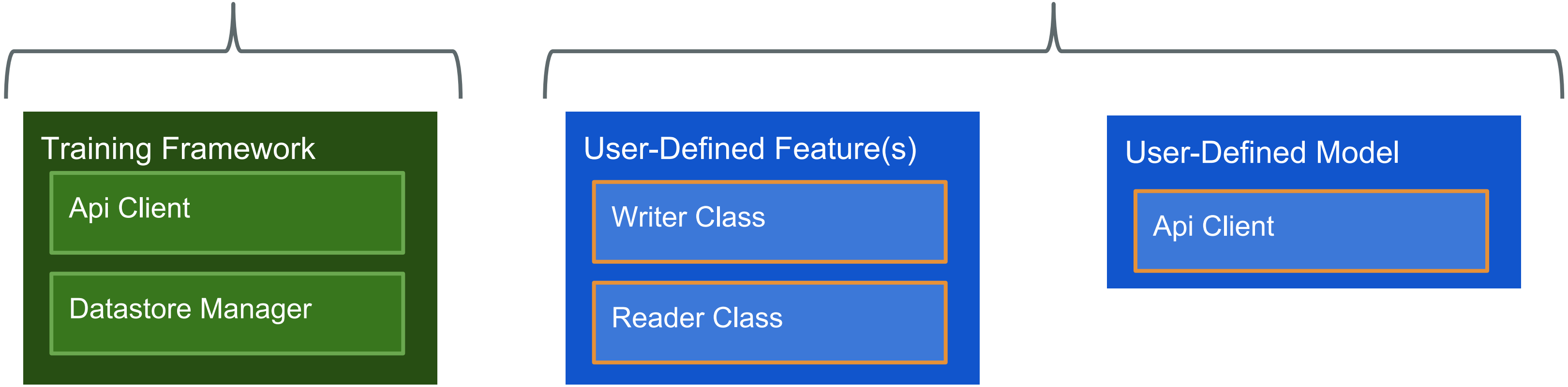
- Manual data downloads, and slow spectrogram generation using `sox`
- Model code tightly coupled to features and training, which made it hard to experiment with changes other than neural network structure
  - Data augmentation
  - New Features
  - New ways of splitting training and test data
- Reliant on manual changes on a laptop - not repeatable, not easy to move to the cloud

Solution: Separation of concerns (Modelling with legos!)

Solution: Separation of Concerns

Provided by RFCx

Provided by user code (uploaded or pulled from github)



- Finds user code via conf file
- Control access to datastore
- Define training actions
  - pull data
  - create features
  - train model

- Features with any structure
- Readers have fine-grained control
  - trade speed for RAM
  - randomize for data augmentation
- Control selection of training, validation data

Command line client for common actions

```
> drfc -- --help
Type:           FireSpec
String form:    <drfc.cli.FireSpec object at 0x10235bac8>
File:           ~/kode/c4g/rainforest-connection-2017/drfc/drfc/cli.py
Docstring:      drfc command-line tool for audio processing and modelling

Usage:          /Users/steventroxler/.virtualenvs/drfc-dev/bin/drfc
                /Users/steventroxler/.virtualenvs/drfc-dev/bin/drfc add-feature
                /Users/steventroxler/.virtualenvs/drfc-dev/bin/drfc api
                /Users/steventroxler/.virtualenvs/drfc-dev/bin/drfc train-model
```

Command-line tool with actions to

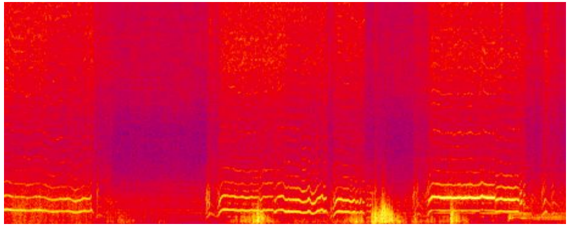
- pull data from RFCx Api
- create features (scipy for fast spectrograms)
- train models

Standardized data layout and support for custom features

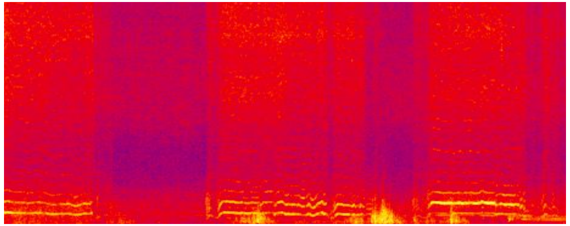
```python
class PitchShiftWavWriter(BaseFeatureWriter):

    def __init__(self, conf, archive):
        super(PitchShiftWavWriter, self).__init__(conf, archive)
        self.bins_per_octave = self.conf['bins_per_octave']
        self.shift_amounts = self.conf['shift_amounts']
        self.sampling_rate = '_not_known_till_we_read_wavs_'

    @staticmethod
    def feature_name(conf):
        return 'pitch_shift_wav'

    def write(self, dao):
        wav_reader = dao.feature_reader('wav')
        self.sampling_rate = wav_reader.sampling_rate
        self.archive.attrs['sampling_rate'] = self.sampling_rate
        for audio_guid in dao.audio_guids:
            self._write_one(audio_guid, wav_reader)
```

```
.
├── audio
│   ├── 35bcbfa0-b255-4ffe-bb62-8e590a6141ca.json
│   ├── 35bcbfa0-b255-4ffe-bb62-8e590a6141ca.wav
│   ├── conf.yaml
│   ├── d9ec8a41-6a4b-4db4-bd78-c7c7d97759ec.json
│   ├── d9ec8a41-6a4b-4db4-bd78-c7c7d97759ec.wav
│   └── download.yaml
├── features
│   ├── spectrogram
│   │   ├── archive.h5
│   │   └── conf.yaml
│   └── wav
│       ├── archive.h5
│       └── conf.yaml
├── labels.csv
└── model
    ├── info.yaml
    ├── model.json
    ├── validation.csv
    └── weights.h5
```

original spectrogram



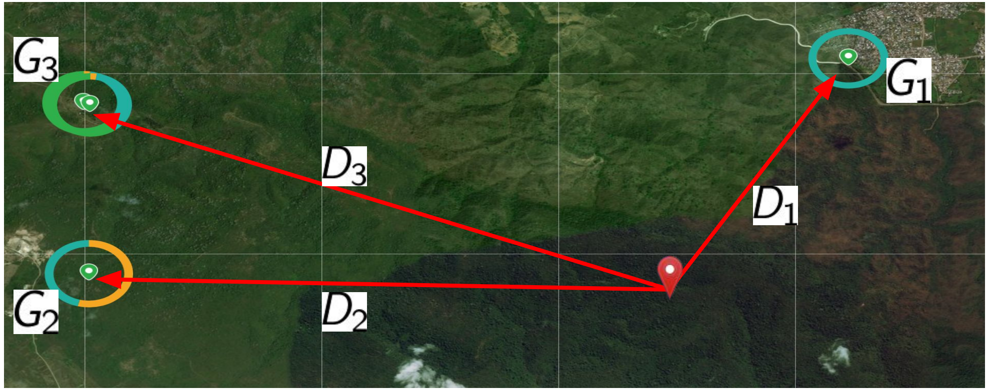
pitch shifted down

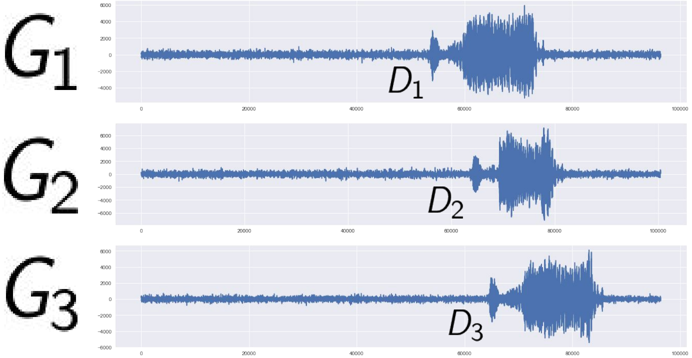


Research topics with multiple Guardians

What are the benefits of combining multiple Guardians?

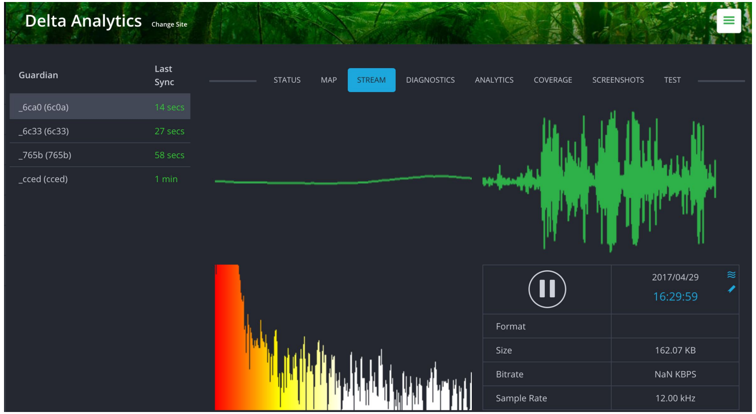
- **Propagation of sound:** Each Guardian G records the same chainsaw sound attenuated and delayed based on the distance D between Guardian G and source, plus uncorrelated noise





1. Improve detection by separating chainsaw from noise
2. Estimate the direction of chainsaw sound

Delta Analytics Test Site



We were given four Guardians and our own RFCx site

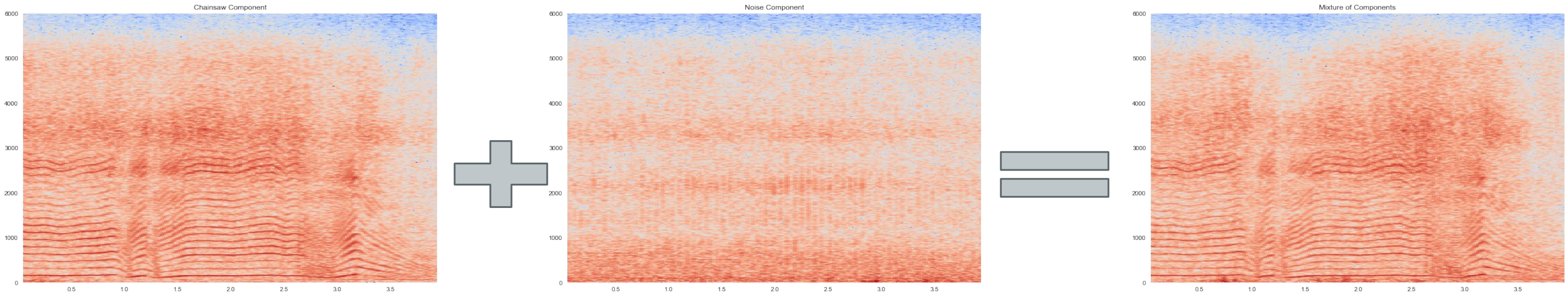
Goal of Delta site is to generate training data to support our research areas

Guardians are not attached

Portable, could try multiple configurations

Separate Chainsaw from Noise

- Each Guardian records chainsaw plus uncorrelated noise
- Goal: reduce noise to improve detection accuracy
- Simple assumption:
  - Two components present in positive examples
  - Chainsaw-like component
  - Background noise component

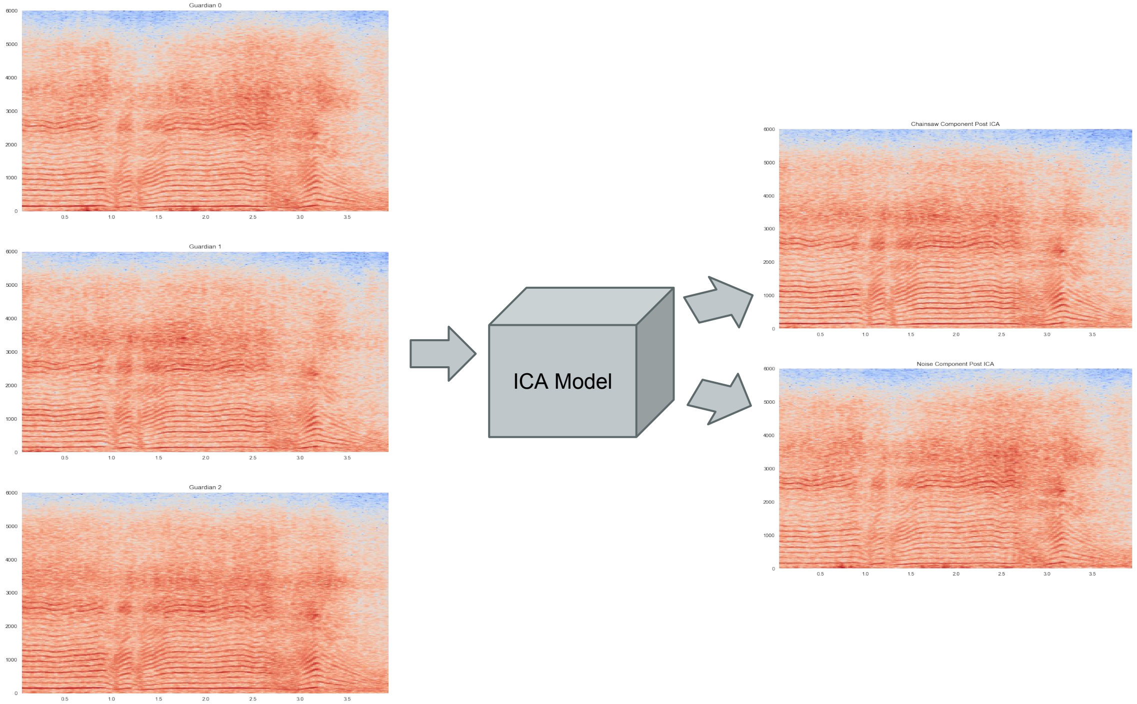


Chainsaw Component

Noise Component

ICA allows us to reduce noise component

- Take audio from multiple Guardians, and separate into chainsaw and noise components
- Current impact on classification accuracy unknown



Due to ‘non-Gaussian’ assumption made by ICA it may not be the best decomposition

However, with the user-defined feature interface testing different decompositions will be easier

Estimate direction and distance of sound

Providing direction and distance of sound reduces the area that rainforests conservationists have to search when they receive an alert.

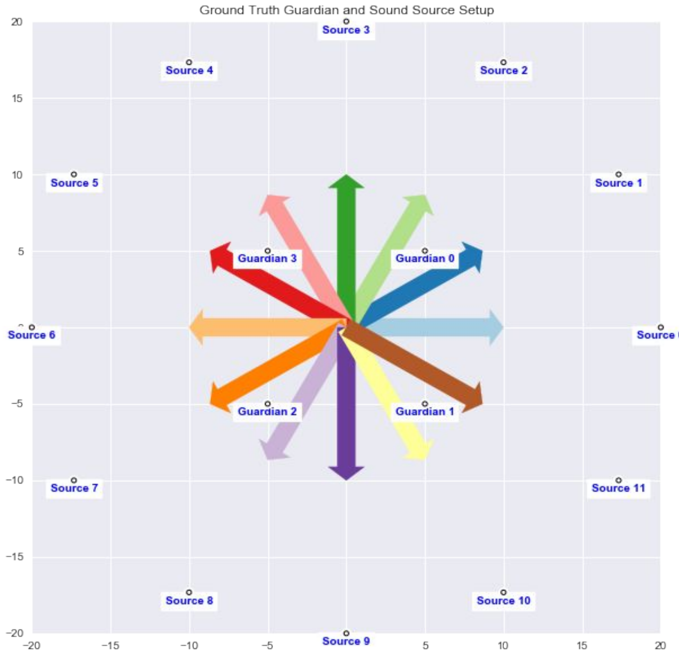
Approach: Source localization based on time delay of arrival

We locate the sound using the time delay between Guardians

Standard techniques use triangulation to determine source location.

However, we assume exact Guardian position is unknown

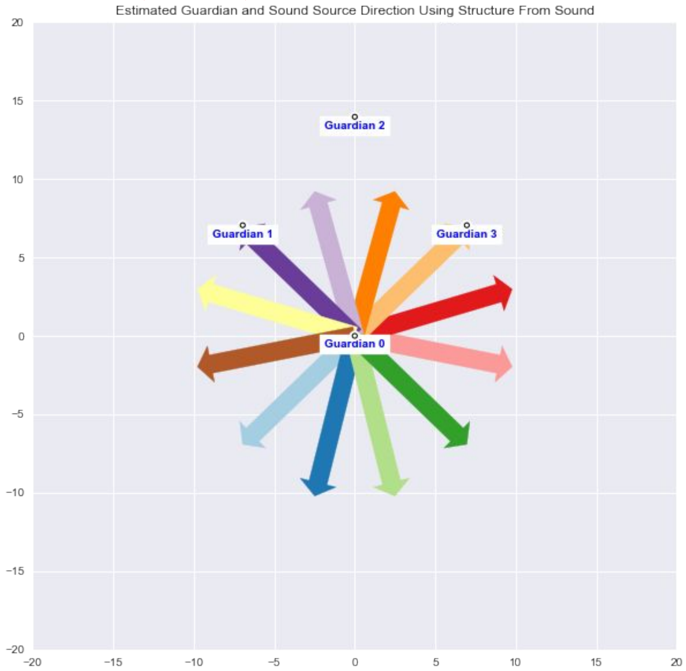
Affine Structure from Sound Algorithm (SfS)*



Toy Example:

Ground Truth

SfS Estimate

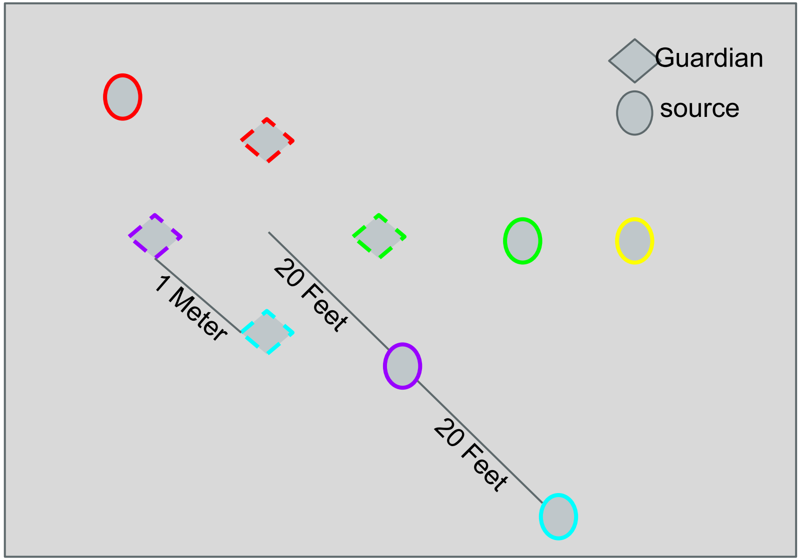


Use the time delay between Guardians relative to a reference Guardian

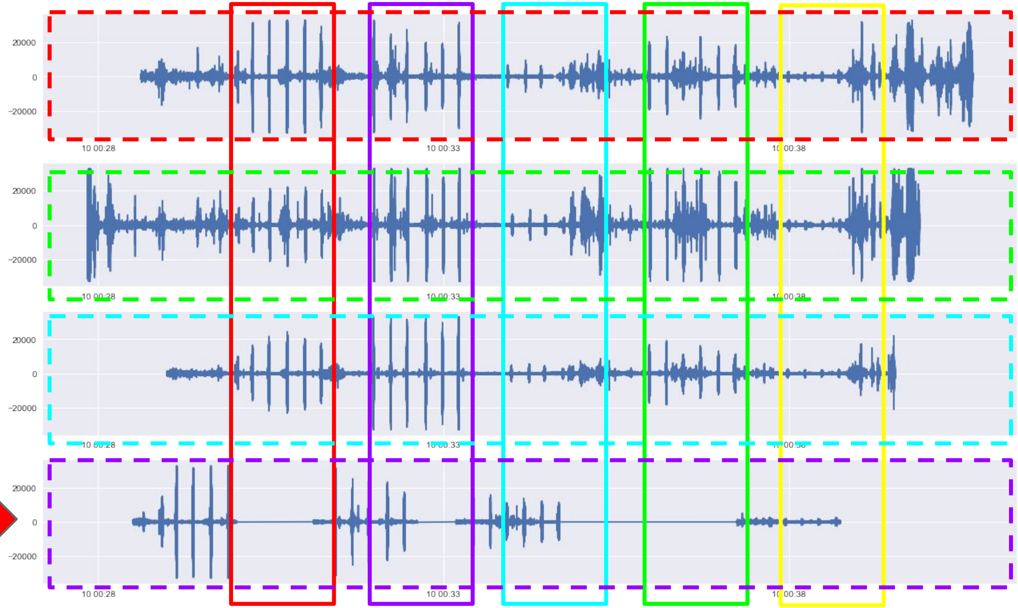
Simultaneously solve relative position of Guardians, and location of multiple sound sources

* Sebastian Thrun - Affine Structure from Sound
robots.stanford.edu/papers/thrun-structure-from-sound05.pdf

Initial direction estimation experiments



Conducted multiple experiments with 'ground truth', e.g., known Guardian and source locations

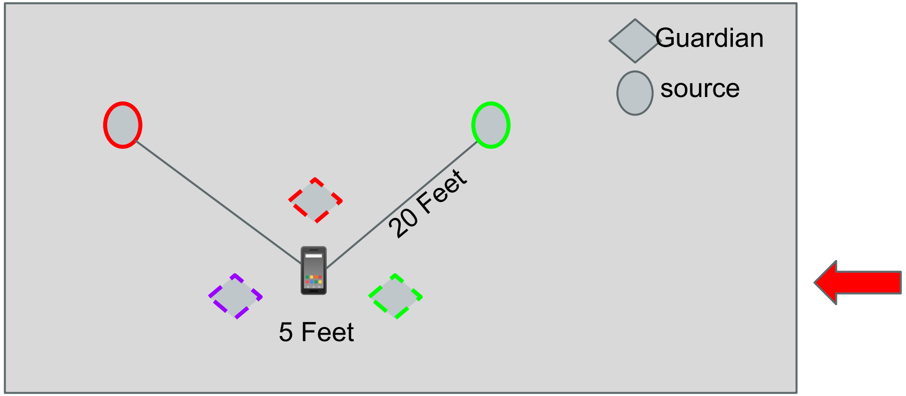


Guardians can lose time sync or completely drop data which results in incorrect estimation of Guardian location and source location

A temporary solution

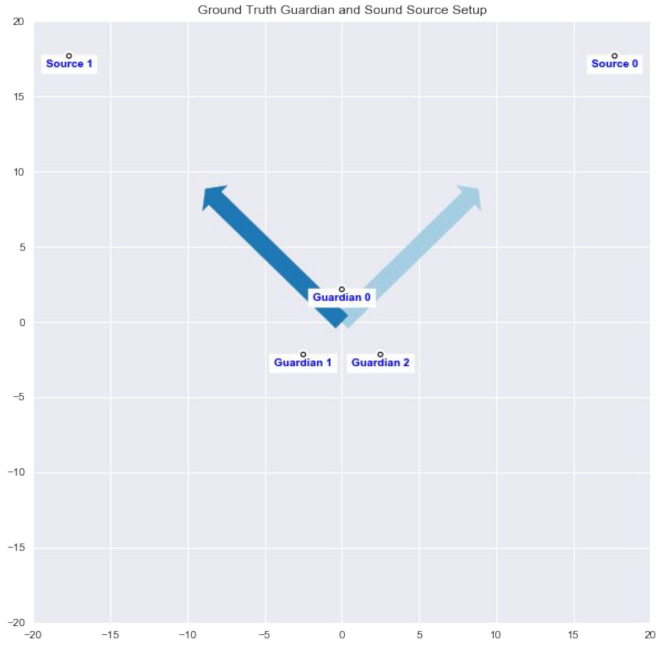
- Guardian timing received from cellular network
  - Issue when limited connectivity
- RFCx plan to use NTP in the future

- Added a ‘sync tone’ played from exact center of Guardians
  - Used to align audio from Guardians before running SfS



Cell phone to play sync tone

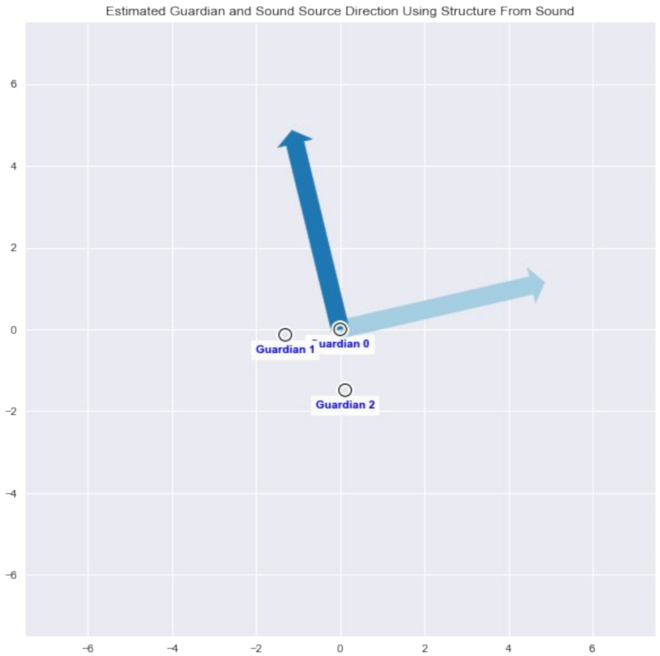
Success!



Actual Example:

Ground Truth

SfS Estimate



Plenty of work before SfS can be used in practice, however, we have demonstrated the algorithm as a viable solution to direction estimation

Some work to do finalizing the algorithm for distance estimation

Machine Learning for Good

Building technical capacity all over the world

Sara Hooker
Executive Director at Delta
Brain Resident at Google

Delta Teaching Fellows


Sara Hooker


Emily Rourke


Kevin Pan


Jack Pfeiffer


Patrick Sun


Hannah Song


Amanda Su


Rosina Norton


Melissa Fabros

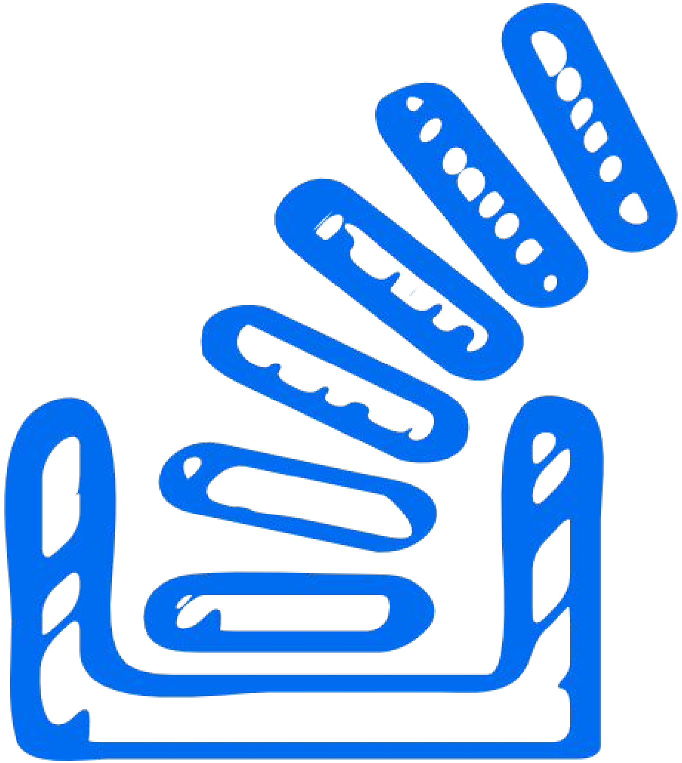
Our goal is simple: **empower anyone, anywhere to leverage data for good in their communities.**

We aim to build technical capacity around the world.

We started this journey for two key reasons.

1. Delta has been working for the last 3 years to help non-profits around the world with their data.

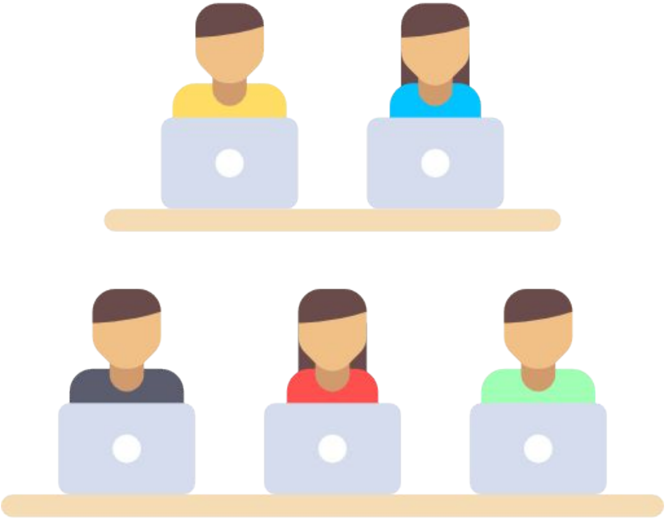
Our non-profit grant recipient program is small and selective. ***We only work with 7-10 nonprofit grant recipients every year.***



There is an incredible amount of unmet demand from nonprofits we say no to.

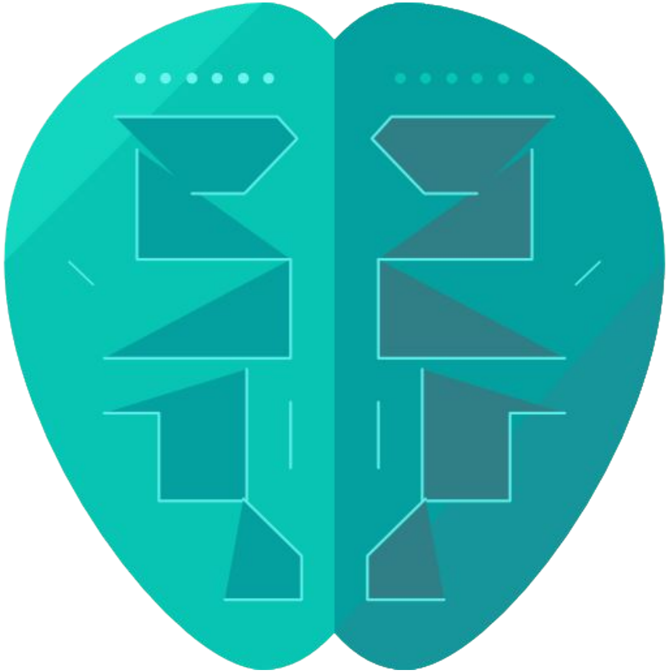
We don’t have the capacity to respond to all the requests we receive.

To address this gap, we aim to build technical capacity around the world to meet demand.



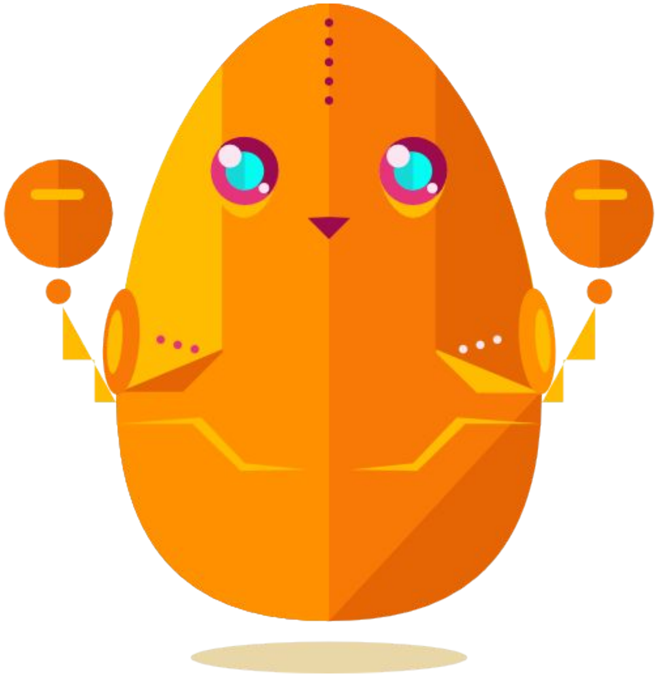
We believe anyone can use data for good. Our team of Delta Teaching Fellows are working to **empower communities around the world to leverage their own data for good.**

2. Machine learning is an incredibly powerful and increasingly complex field.



As the amount of data available in the world increases, machine learning techniques will be increasingly used to extract insight. **However, the community equipped to make sense of these algorithms tends to be concentrated in a few places.**

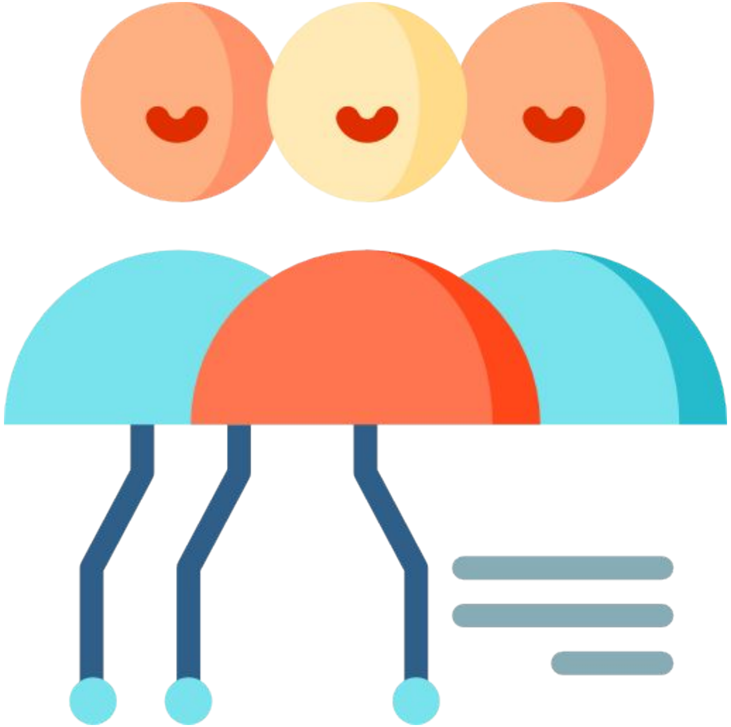
Like all powerful tools, machine learning requires responsibility & accountability.



It is easy to mislead with complex algorithms and we often don't understand the implications of our work.

We believe the face of data science has to be **more representative** of the world it serves.

We want nonprofits and communities to be able to own their data and be part of the conversation about data.



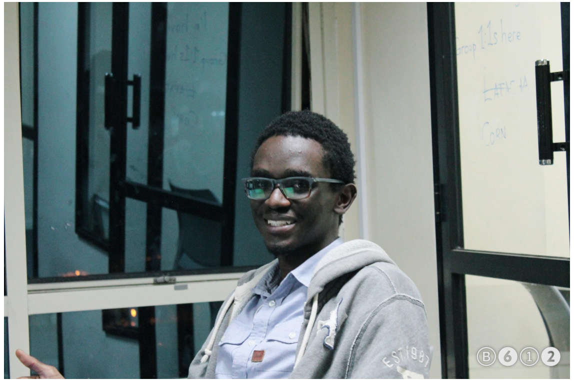
- Democratize access to machine learning tools.
- Regularly share tutorials, code and insights
- Teach in person classes to communities around the world

In 2017, we started a pilot program in Nairobi, Kenya to teach machine learning for good.

Over 3 months, we:

1. Developed a curriculum,
2. Travelled to Nairobi, Kenya to teach a group of 30 students
3. Learned a lot about the challenges and joys of making machine learning accessible.

We partnered with Moringa school, an engineering school in Nairobi.



moringa SCHOOL

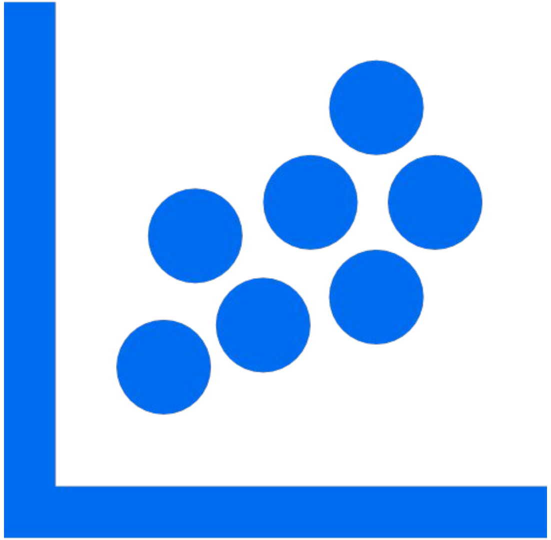
Our students met us more than halfway, many of them travelling for over an hour to attend class

We taught a night class from 6:00 to 9:00 pm on Mondays and Wednesdays, and hosted office hours every Saturday.

Almost 100% of our students were working full time in addition to taking our class. Most hoped to use the class as a stepping stone to getting a job as a data scientist.

What did we teach?

Focus on key machine learning concepts and combine with hands on python coding labs.



We started at ground zero.

We explained the ***underpinnings***, ***assumptions***, and ***applications*** of a number of algorithms, including:

- Linear regression
- Regression trees
- Random Forest
- Clustering algorithms
- NLP algorithms

Example of a module outline for a class.

Module Checklist

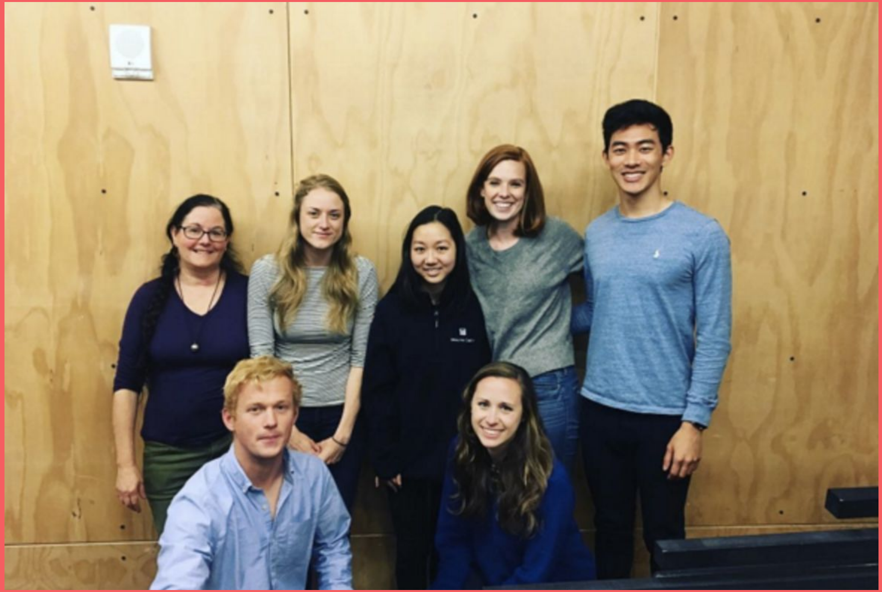
- ❏ Linear regression
    - ❏ Relationship between two variables (x and y)
        - ❏ Formalizing f(x)
        - ❏ Correlation between two variables
        - ❏ Assumptions
    - ❏ Feature engineering and selection
    - ❏ Learning process: Loss function and Mean Squared Error
    - ❏ Univariate regression, Multivariate regression
    - ❏ Measures of performance (R2, Adjusted R2, MSE)
    - ❏ Overfitting, Underfitting

What did we learn?

1. Three months = extremely compressed timeline

We would spend the time in the car on the way to class finalizing the code and slides! Luckily(?), Nairobi's traffic jams made our commute to class take up to 90 minutes.

Hannah and Sara were on the ground in Nairobi, Kenya for our pilot program. The rest of the team taught remotely through Slack.

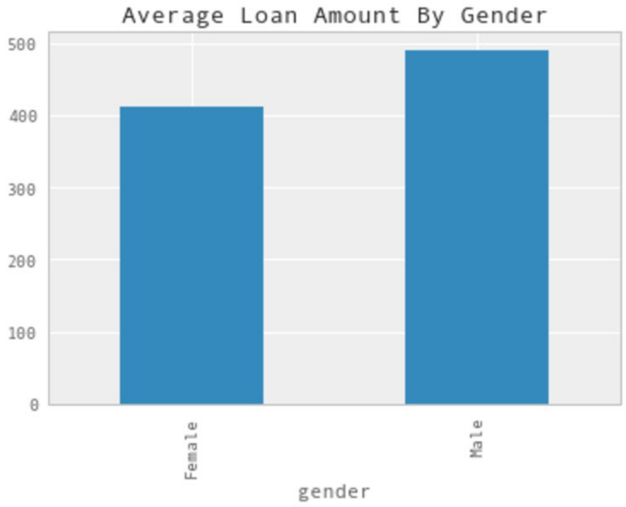


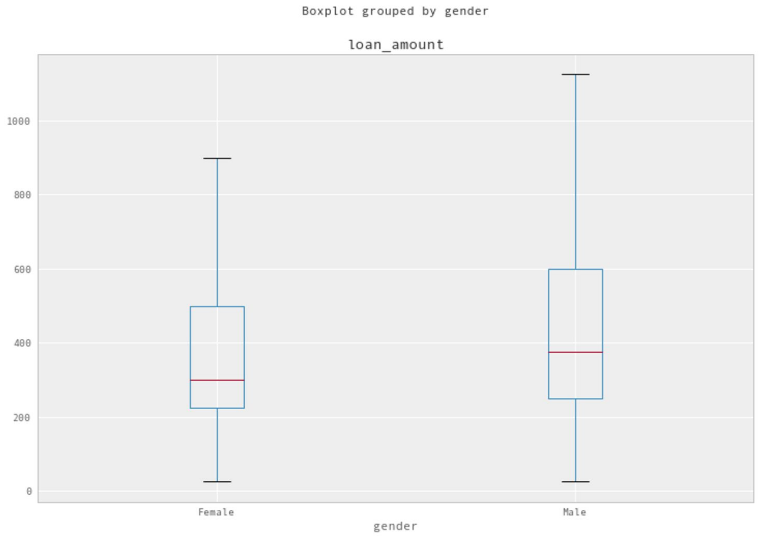
2. We made cultural missteps, and also encountered bias in surprising ways.

We used McDonald's as an example in our first lesson. **McDonald's is not available in Kenya!**

We relied on Google Image cartoons to demonstrate some concepts. A student pointed out that almost **all of our images were of white males.**

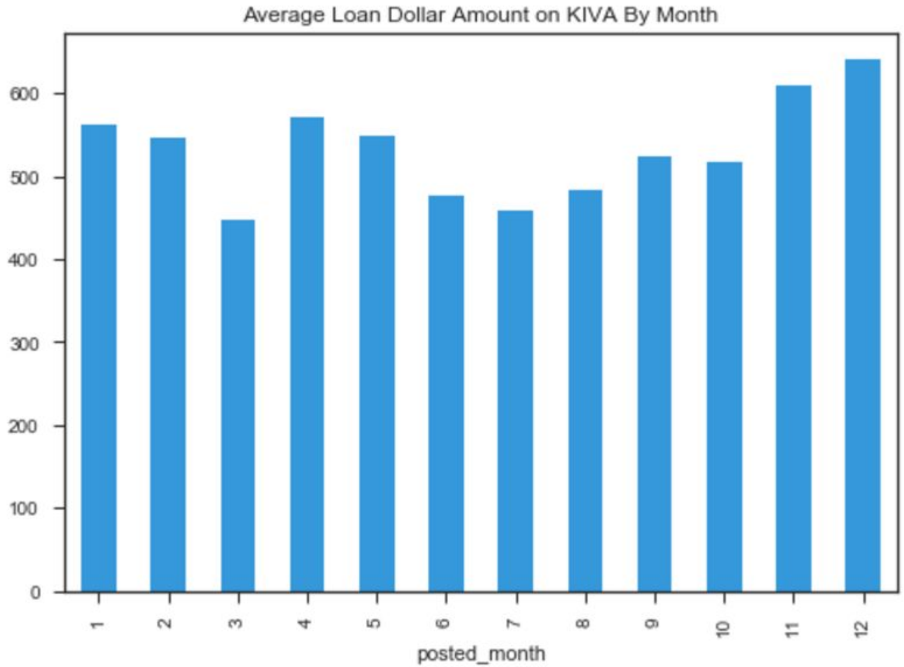
3. Using local data empowers the student to become the teacher.





We used Kiva loan data from Kenya for all of our coding labs. **Students provided invaluable cultural context and insight in interpreting the raw data.**

4. Innovative insights and feature engineering emerges because of domain expertise.

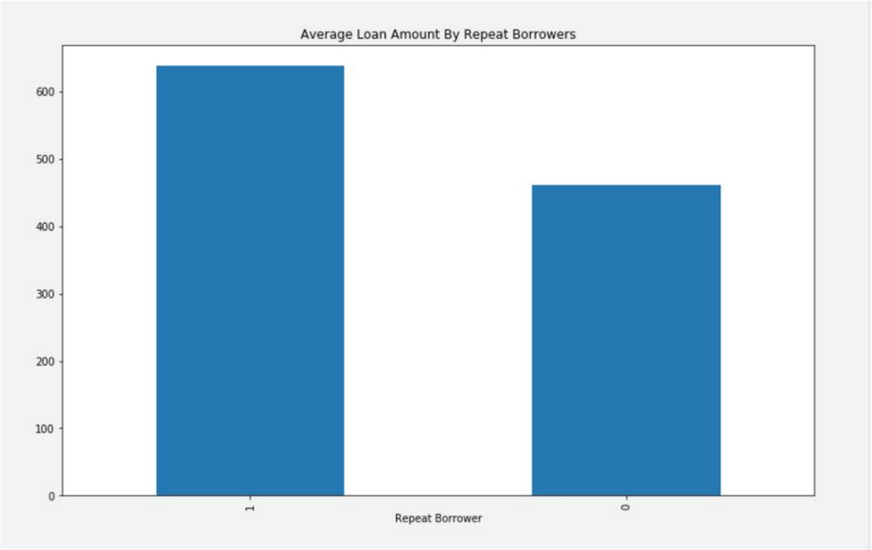


Average loan amount goes up at the end of the year in November-February.

The amount requested spikes again in April.

Why?

Example of one of our student's exploratory charts.



Repeat borrower

First time borrower

Anthony's chart shows that repeat borrowers request on average ~$120.00 more than first time borrowers.

Why would this be the case?

5. Clearly communicate expectations and your students will meet you there.

We did not make compromises on the difficulty of the course but asked our students to take what they wanted from it. Our students pushed themselves every step of the way.

You set your learning path.

We will be over-providing resources & content since students have different goals in this class:

- Become a data scientist
- Learn to code
- Understand core concepts and vocabulary

6. Teaching students with different perspectives is an incredible thing.

Each student's **drive**, **persistence**, **curiosity**, and **hard work** kept us on our toes

Hannah was given bananas as a thank you from one of our students.

What’s next?

- Over the next few months we will make our content free and available online.
- In 2018, we will be teaching our curriculum in collaboration with 3 partners in the Bay Area & abroad.

Interested in getting involved?

Teaching fellows

Beta testers

International/Local partners

Applications are already live! You can apply on our website at deltanalytics.org.

BUILD

What factors influence high school students' success in college and beyond?

Delta Team

From left to right:

Rosina Norton

Physicist and clean energy analyst

Jessie Huang

CX at Udemy

Hannah Song

Researcher at Columbia Law

Kevin Pan

Analyst at Brattle Group

BUILD

Who is BUILD?

BUILD enables high school students to start their own small businesses to learn critical skills for success in college and beyond.

Dataset:

- Student demographics
- Student college enrollment data
- Student program participation data
- Collected between 2011-2016
- From multiple schools in the Peninsula, Boston, Oakland, and DC areas

Who is BUILD?

BUILD Sales Bazaar; Oakland, March 2017

BUILD asked Delta to help with program evaluation

What factors, in particular BUILD program elements, determine student success?

BUILD's metrics of success for their students:

- College acceptance and enrollment
- College persistence
- Transferring from 2-year to 4-year college

We had two equally important goals:

1. Producing results based on the dataset as of the current date
2. Equipping BUILD staff for future analyses

Regression Analysis Setup

What factors, in particular BUILD program elements, determine student success?

Success metric = B_1*Demographic factors + B_2*Individual student merit + B_3*BUILD program elements + e

Success metrics = {College acceptance, College dropout, Persisted through first two years of college, Successfully transferred}

Regression Analysis Results

	Dependent variable:			
	Accept	Drop	Persist	Transfer
	(1)	(2)	(3)	(4)
role_in_student_teamOfficer	0.006	-0.069	-0.351^{***}	-0.236^{***}
	(0.049)	(0.044)	(0.055)	(0.086)
role_in_student_teamVP	0.047	-0.042	-0.332^{***}	-0.279^{***}
	(0.053)	(0.047)	(0.059)	(0.093)
pct_attended	-0.212	0.034	-0.282^{**}	-0.006
	(0.130)	(0.109)	(0.143)	(0.213)
most_recent_gpa	0.402^{***}	-0.050^{*}	0.118^{***}	-0.349^{***}
	(0.029)	(0.025)	(0.032)	(0.050)
not_in_another_college_prep_programYes	0.076^{**}	0.077^{**}	0.214^{***}	-0.048
	(0.038)	(0.035)	(0.042)	(0.069)
genderMale	0.020	-0.045	0.057	-0.045
	(0.038)	(0.034)	(0.042)	(0.067)
qualifies_for_free_reduced_lunchYes	0.075^{*}	0.0003	0.019	0.021
	(0.045)	(0.041)	(0.050)	(0.079)
has_older_sibling_in_build	0.018	-0.001	-0.036	-0.024
	(0.046)	(0.041)	(0.050)	(0.079)
Constant	-0.453^{***}	0.202^{*}	0.165	1.802^{***}
	(0.127)	(0.117)	(0.140)	(0.229)
Observations	310	187	310	187
R^2	0.420	0.079	0.238	0.280
Adjusted R^2	0.404	0.037	0.218	0.247
F Statistic	27.232^{***} (df = 8; 301)	1.902^{*} (df = 8; 178)	11.750^{***} (df = 8; 301)	8.631^{***} (df = 8; 178)
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Regression Analysis Results

	Dependent variable:			
	Accept	Drop	Persist	Transfer
	(1)	(2)	(3)	(4)
role_in_student_teamOfficer	0.006	-0.069	-0.351***	-0.236***
	(0.049)	(0.044)	(0.055)	(0.086)
role_in_student_teamVP	0.047	-0.042	-0.332***	-0.279***
	(0.053)	(0.047)	(0.059)	(0.093)
pct_attended	-0.212	0.034	-0.282**	-0.006
	(0.130)	(0.109)	(0.143)	(0.213)
most_recent_gpa	0.402***	-0.050*	0.118***	-0.349***
	(0.029)	(0.025)	(0.032)	(0.050)
not_in_another_college_prep_programYes	0.076**	0.077**	0.214***	-0.048
	(0.038)	(0.035)	(0.042)	(0.069)
genderMale	0.020	-0.045	0.057	-0.045
	(0.038)	(0.034)	(0.042)	(0.067)
qualifies_for_free_reduced_lunchYes	0.075*	0.0003	0.019	0.021
	(0.045)	(0.041)	(0.050)	(0.079)
has_older_sibling_in_build	0.018	-0.001	-0.036	-0.024
	(0.046)	(0.041)	(0.050)	(0.079)
Constant	-0.453***	0.202*	0.165	1.802***
	(0.127)	(0.117)	(0.140)	(0.229)
Observations	310	187	310	187
R^2	0.420	0.079	0.238	0.280
Adjusted R^2	0.404	0.037	0.218	0.247
F Statistic	27.232*** (df = 8; 301)	1.902* (df = 8; 178)	11.750*** (df = 8; 301)	8.631*** (df = 8; 178)
Note:				*p<0.1; **p<0.05; ***p<0.01

Note:
Selection bias?
Missing data?

Transfer Analysis Setup

Transfers defined as students who enrolled first in 2-year colleges, then successfully transferred to 4-year colleges.

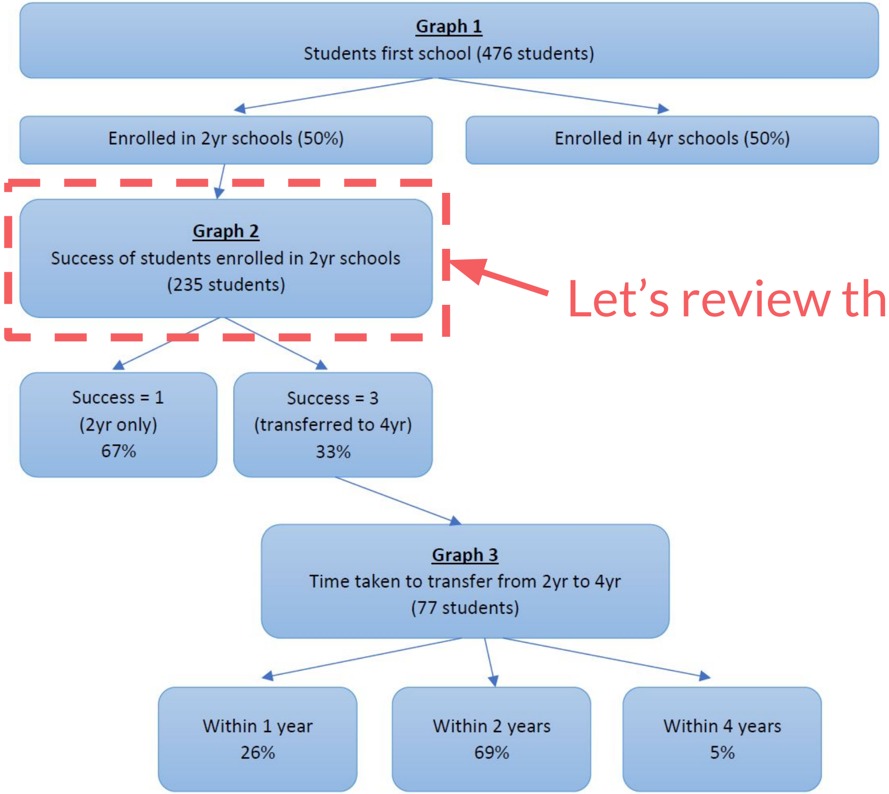
What factors led to successful transfers? What should BUILD do to encourage successful transfers?

Transfer Analysis Setup

Defining "***transfer success***" by ranking outcomes for students who enrolled in 2-year colleges:

Success Metric	**Transfer Score**
Enrolled in 2-year college and transferred to a 4-year college within 4 years	3
Enrolled in 2-year college and transferred to other establishment within 4 years	2
Enrolled in 2-year college only	1
Never enrolled in 2-year college	0

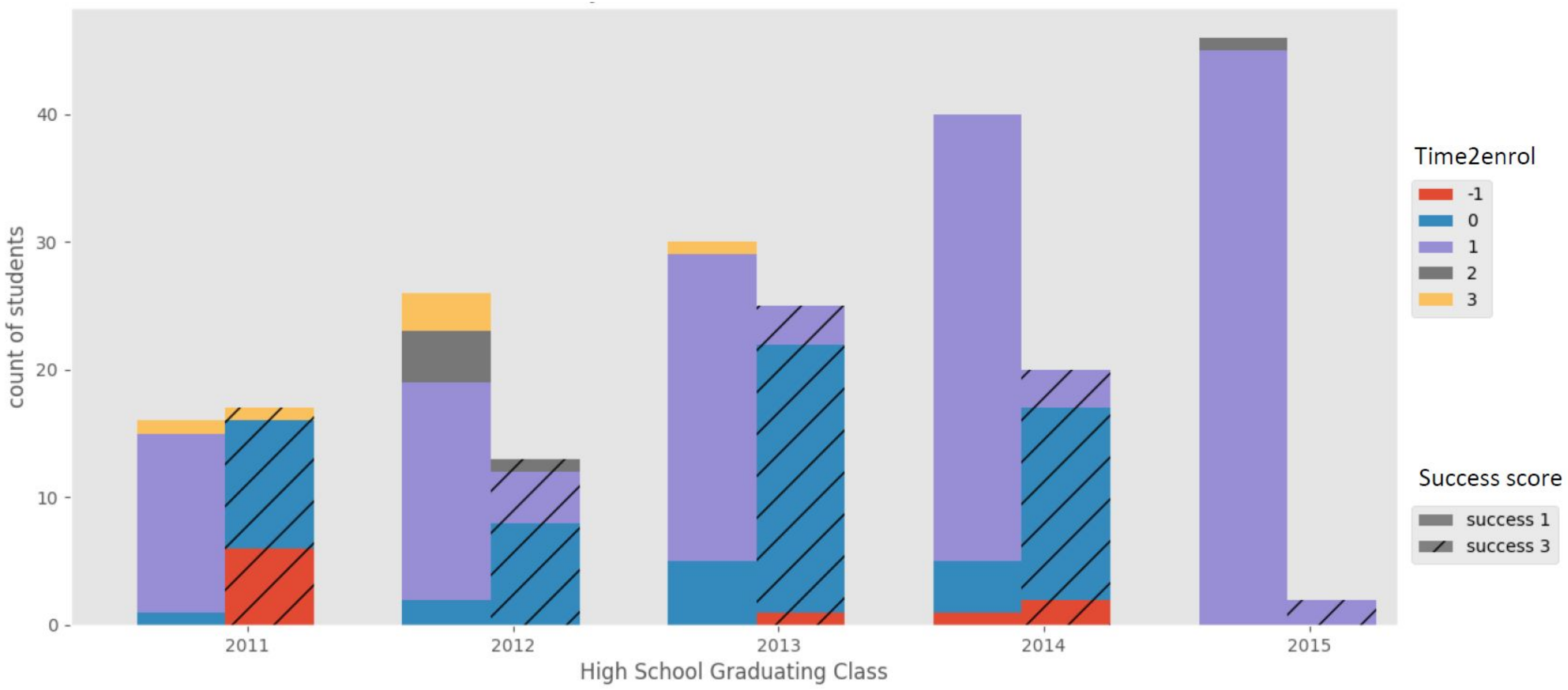
Transfer Analysis Setup



Let's review this population in detail

Transfer Analysis Results (n = 235)

- Progressively fewer transfers (more direct enrollees to 4-year)
- Students with -1 or 0 time are more likely to transfer



Transfer Analysis Results

Takeaways:

- We recommend that BUILD continue collecting this data to allow for large-scale analysis
- Results appear to suggest BUILD is having success improving number of students enrolling directly in 4-year schools
- Results appear to suggest that high school students taking community college classes is relatively determinative in transfer success

Equipping BUILD for their own analyses

How can BUILD take our work further?

Documentation, documentation, documentation!

Equipping BUILD for their own analyses

Delta turned over the following materials to BUILD:

1. **Final report of results and recommendations**
2. **Code package**

Equipping BUILD for their own analyses

- **Retaining all code from exploratory analyses**
  - *What sanity checks did we conduct on the data?*
  - *What questions yielded fruitful research questions?*
  - *Which questions led to dead ends?*
- **Setting up code for future regression analyses**
  - *What R packages were useful?*
  - *What visualizations did we make that BUILD can recreate?*
- **Commenting code extensively**

From here, BUILD will take the lead on their own analyses!

For more information on BUILD and the work they do, check out:

[*https://build.org/*](https://build.org/)

Thanks!

Ways you can get involved

Data Professionals

Apply to be a Fellow!

http://tiny.cc/df_2018

Non-Profits

Apply for our 2018 Service Grant!

http://tiny.cc/dg_2018

Donate!

http://bit.ly/2tXEtPA

Thanks to our Partners!

breather

moringa
SCHOOL

DATA
INSTITUTE
UNIVERSITY OF SAN FRANCISCO

Some special thanks…

Enjoy the poster presentations!

Appendix (additional slides)

Our course required no prior knowledge of coding.

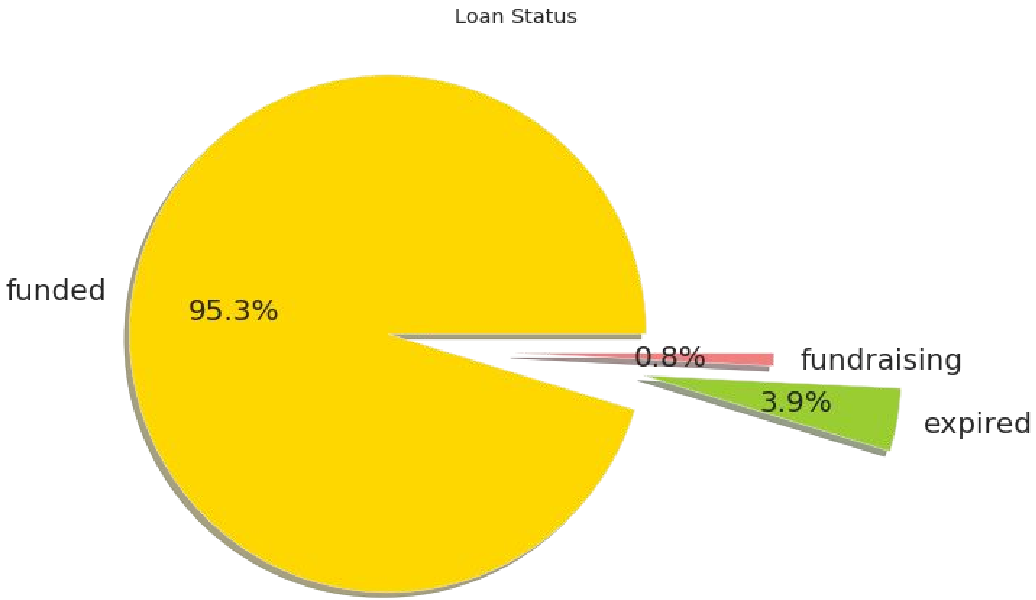
We relied on intuitive examples, coding labs to expose students to applied machine learning and discussion based classes to expose gaps in understanding.

This course was only possible because of the support of 9 Delta fellows working around the clock.

Delta Teaching Fellows are data professionals who want to make machine learning accessible.

Our students used python code labs to understand how ML could be applied to real problems.

Using Kiva microfinancing data from Kenya, we walked the students through the steps (and missteps) of a real research problem.



An excellent visualization on loan status by Jim, one of our students!

We taught students code by exposure to a real data science project.

To our leadership team!

To our project leads!






To our fellows!