

# Center for Data Science & Public Policy



THE UNIVERSITY OF  
**CHICAGO**

## Preparing for Collaborative Data Driven Projects

December 9, 2016



**Lauren  
Haynes**



**Hareem  
Naveed**



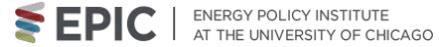
**Erika  
Salomon**

# Agenda

- 5 Minutes Introductions
- 20 Minutes Intro to Project Scoping (DSaPP)
- 20 Minutes Intro to Data Maturity (DSaPP)
- 20 minutes Intro to Data Governance (McClean) - How to get legal agreements, stakeholder buy in
- 20 Minutes Q&A



**NOBLE**



Pecan Street Inc.



# 171 Fellows



“We are used to using data to justify funding decisions. Now we can use data to improve what we do.”

amazon

High Roller Bicycles Prestige Bike (20-inch Wheel)

Key Features:

- 20-inch wheels
- 21-speed drivetrain
- Front and rear disc brakes
- Multi-position seat
- Multi-position handlebars
- Front and rear fenders
- Front and rear mudguards
- Front and rear lights
- Front and rear reflectors
- Front and rear fenders
- Front and rear mudguards
- Front and rear lights
- Front and rear reflectors

Frequently Bought Together

Customers Who Bought This Item Also Bought



NETFLIX

JULIE ANDREWS ANNE HATHAWAY

Walt Disney PICTURES PRESENTS

THE PRINCESS DIARIES 2

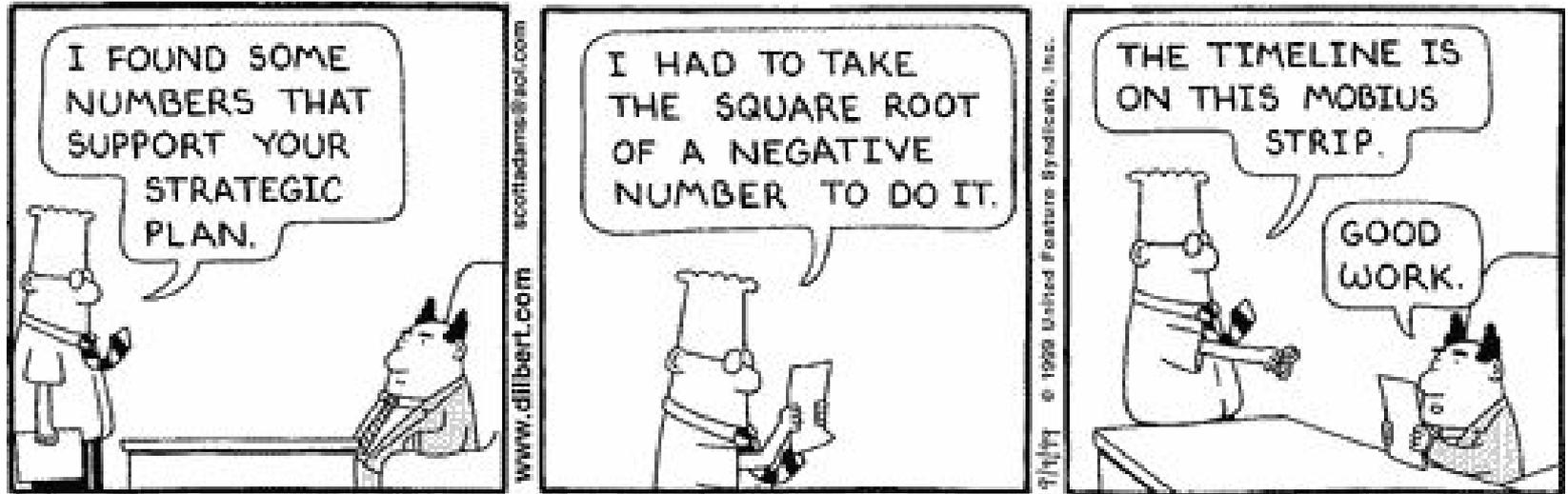
ROYAL ENGAGEMENT

ORIGINAL SOUNDTRACK

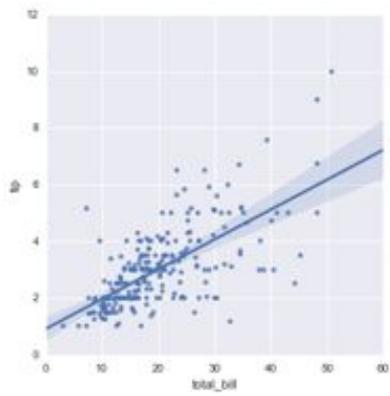
New Hit Song from KELLY CLARKSON "Breakaway"



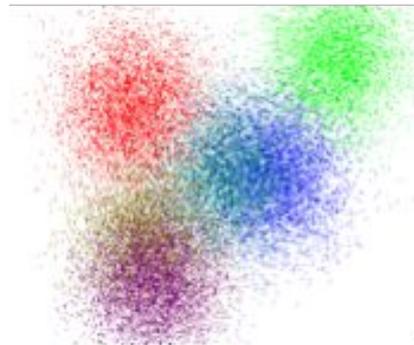
# This May Sound Daunting ...



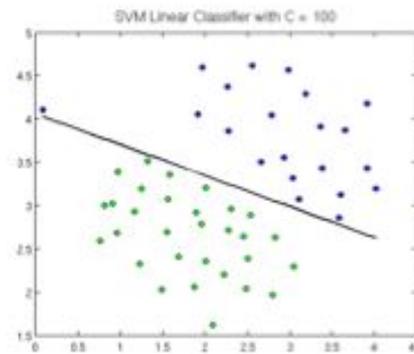
# But Most Common Machine Learning Tasks...



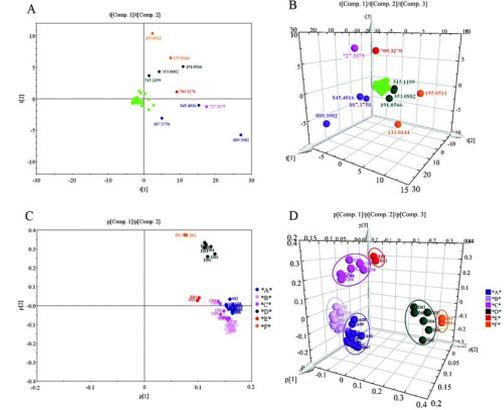
**Regression**  
Using trends to  
predict outcomes



**Clustering**  
Finding existing  
groups or  
categories



**Classification**  
Labeling and  
sorting into  
groups



**Dimension Reduction**  
Finding important  
predictors

# ... You Actually Learned In Kindergarten

## Regression

Using trends to predict outcomes



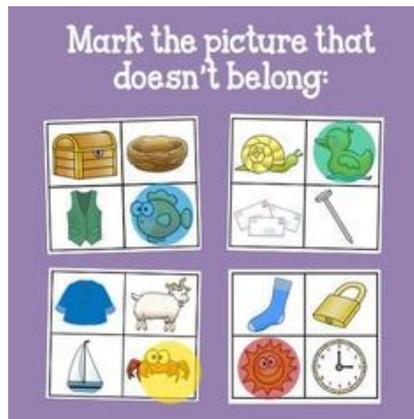
## Clustering

Finding existing groups or categories



## Classification

Labeling and sorting into groups



## Dimension Reduction

Finding important predictors



# Most Projects Fall in a Few Categories

- **Early** warning & intervention
- **Efficient** resource allocation & targeted action
- **Effective** advocacy & fundraising
- **Data-driven** policy recommendation & evaluation

STAR  
TREK

# Data are People



BRENT SPINER  
LIEUTENANT COMMANDER DATA

*“Predictive analytics is emerging as a game-changer. Instead of looking backward to analyze “what happened?” predictive analytics help executives answer “What’s next?” and “What should we do about it?”*

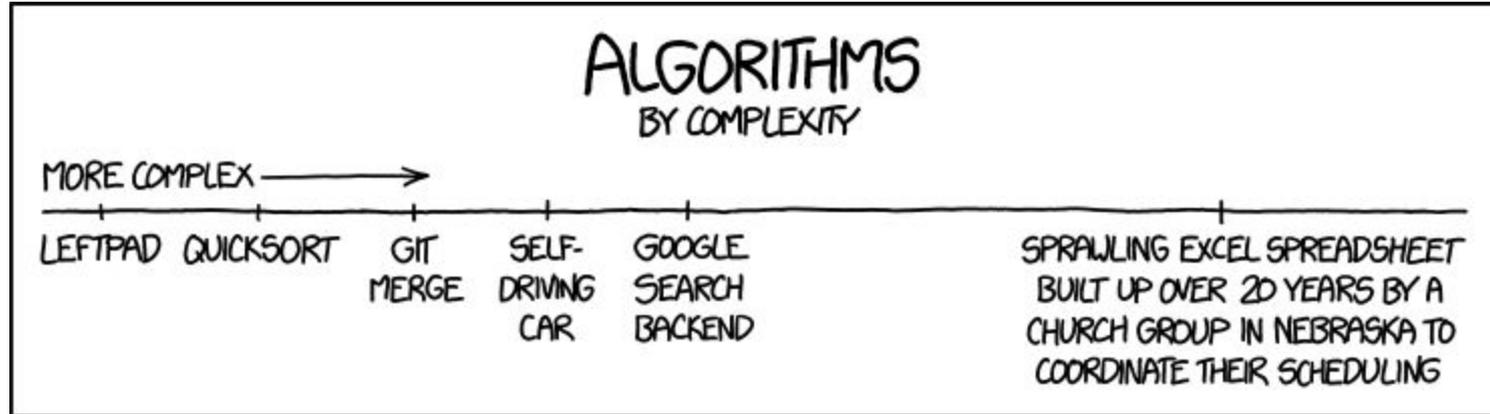
***Forbes Magazine***

*Why Predictive Analytics Is A Game-Changer*

# Project Scoping Workshop

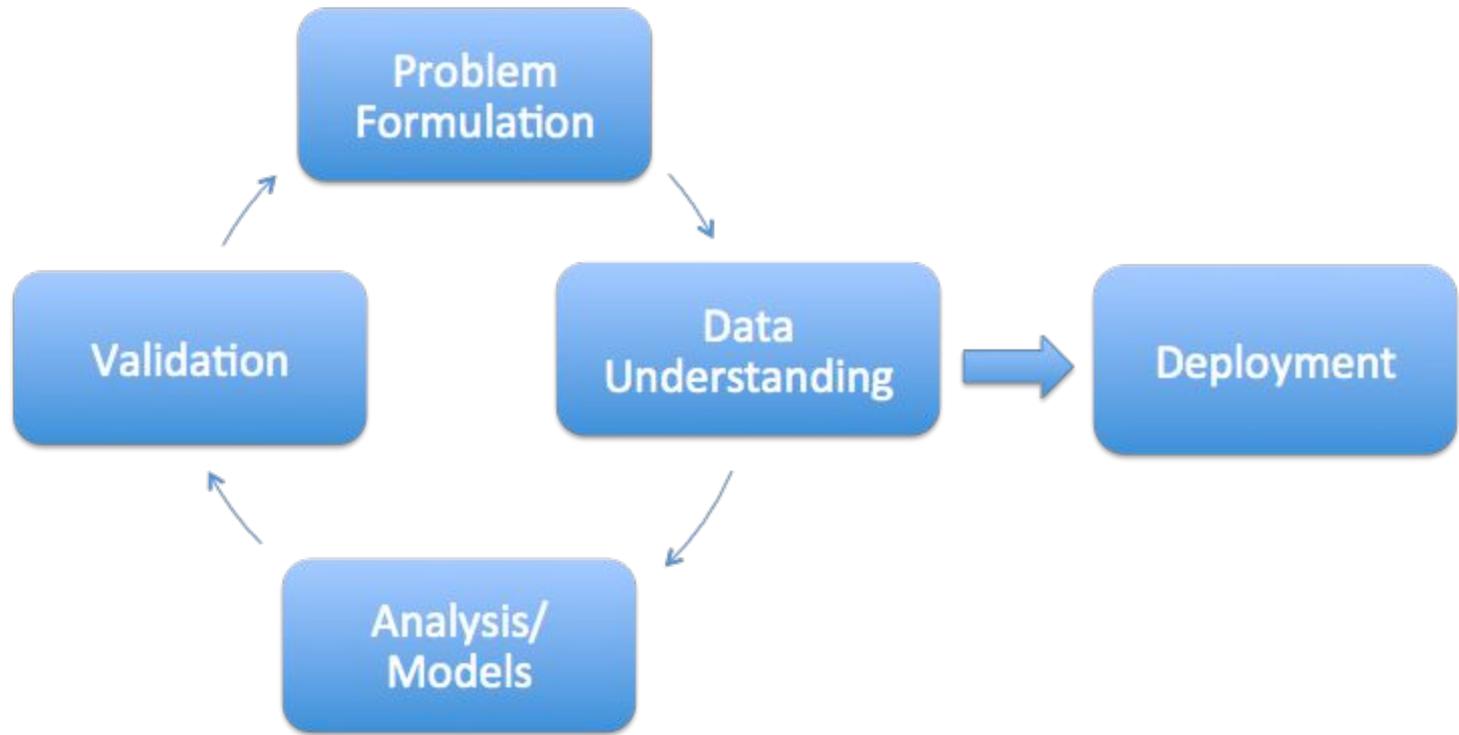
“We are used to using data to justify funding decisions. Now we can use data to improve what we do.”

# Scoping a Good Project is Easier Said than Done



# What Makes a Good Project

- A solvable problem.
- A challenging problem.
- An important problem with social impact.
- A motivated, capable, and committed partner.
  - a. Domain/business resources
  - b. Data understanding resources
  - c. Commitment to implementation
- Appropriate, relevant, available data.



# Relevance and Sufficiency



Irrelevant and Insufficient



GOAL

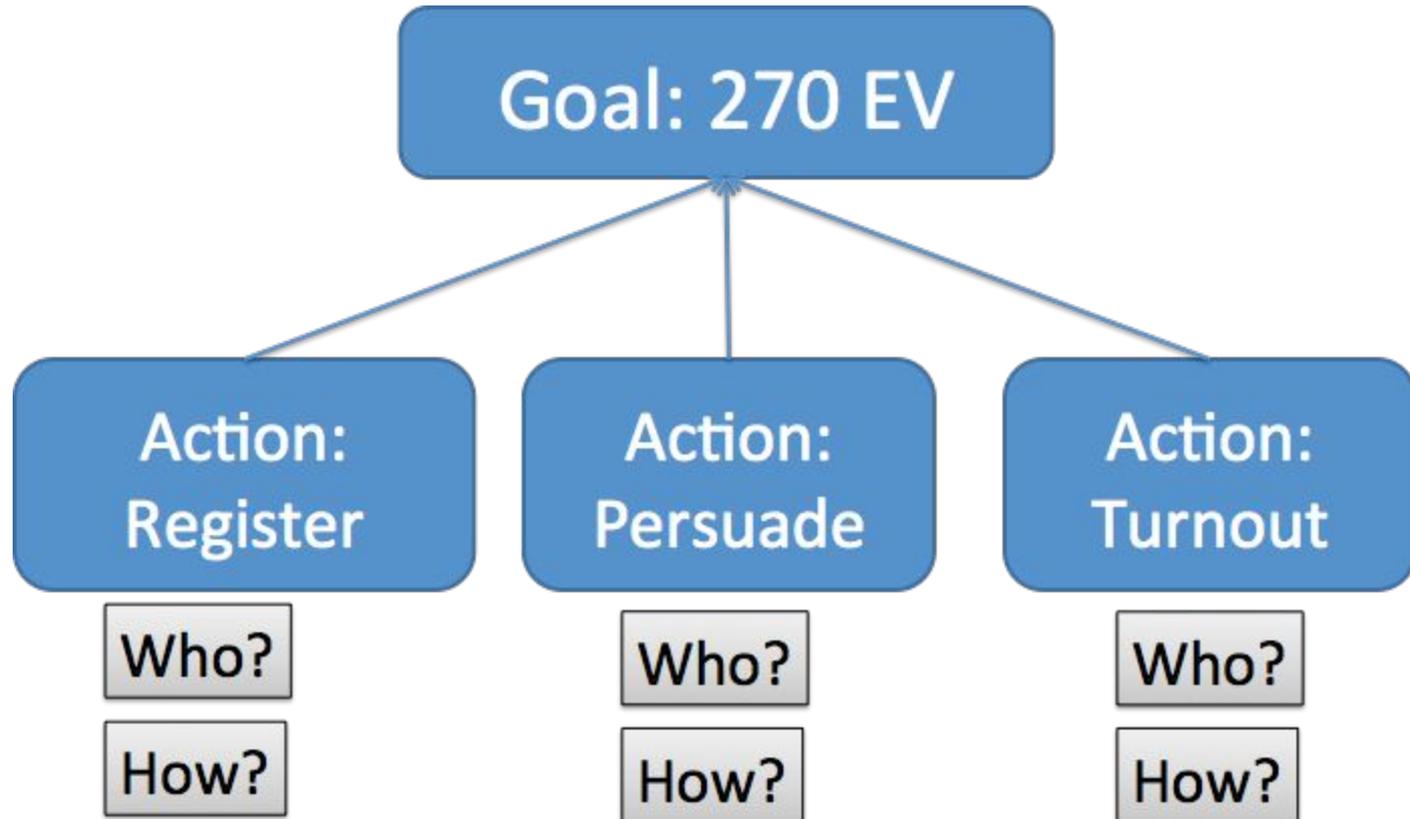


Relevant but Insufficient



Relevant and Sufficient





# Data Analytics: Problem Formulation

- Identify Goals
- Identify Actions you can take to achieve those goals
  - Break down actions into fine-grained questions/subactions
- Identify Data Sources Resources you need and have
- Identify Analysis/Modeling that needs to be done

# Workshop

- Scope an analytics-driven project for a problem your organization is facing
  - Identify goal(s)
  - Actions (persuasion for example)
  - Data sources (data you have, data you need to collect, relationships to get that data)
  - Models
- Who? (to target for each action)
- What? (to say to them)

# Workshop: Identify Actions that can be taken to achieve the goal

- What programs do I have access to? What would they do differently if they had more information/knew where their interventions were most likely to be effective
- Be sure to name out which individuals are taking the action

# Workshop: Data Sources

- Data sources (data you have, data you need to collect, relationships to get that data)

# Workshop : Changing Actions

- Who? (to target for each action)
- What? (to say to them)
- How? (to use different communication channels)

# Partner Data Maturity Framework

# Data Maturity Framework

**Center for Data Science & Public Policy**



**DataKind**

## Problem Definition

Definition

Data Governance

Implementation and Maintenance

## Data Readiness

Accessibility

Storage

Integration

Relevance and Sufficiency

Quality

Collection Frequency

Granularity

History

Privacy

Documentation

## Organizational Readiness

Staff Buy In

Data Collector Buy In

Leadership Buy In

People Resources

Intervener Buy In

Funder Buy In

# How to Arrive at Data Nirvana

- **Commit to data-driven decisionmaking.**
  - Data as a first class citizen
  - Ready for tough love from data
  - Willing to take data risks
  
- **Understand what data supports the mission and how.**
  - Has the data it needs or can obtain it
  - Technical and organizational capacity
  - Outcome linked to action



# Data Readiness

# Relevance and Sufficiency



Irrelevant and Insufficient



GOAL



Relevant but Insufficient



Relevant and Sufficient



# Data Capture

population with data



population of interest

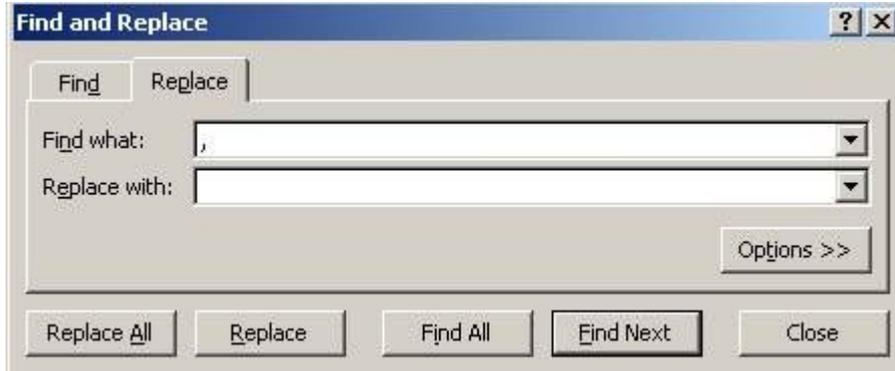
# Data Storage / Format



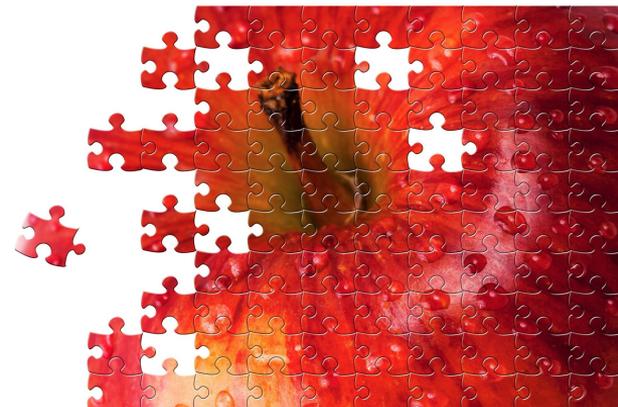
High cost of use

Low cost of use

# Data Quality



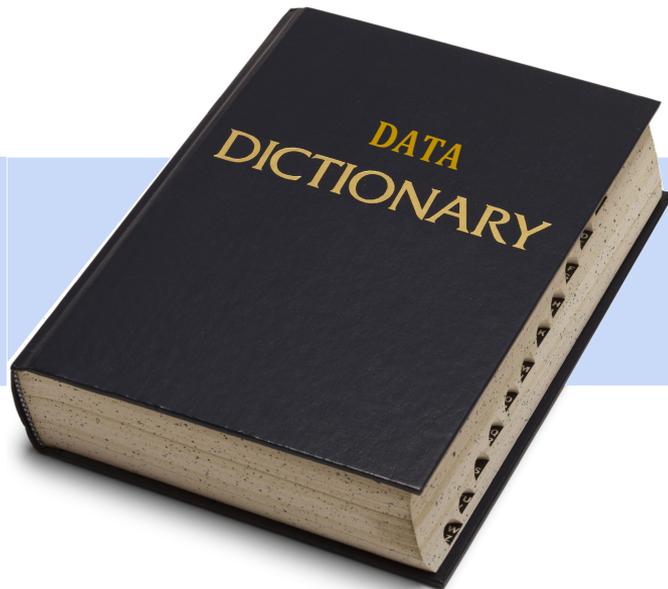
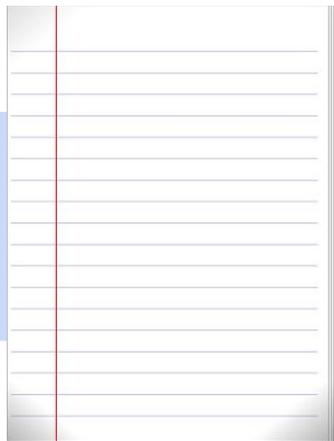
# Integration



# Accessibility



# Documentation



```
<metadata>data  
<metadata>data  
</metadata>  
</metadata>
```

# Organizational Readiness

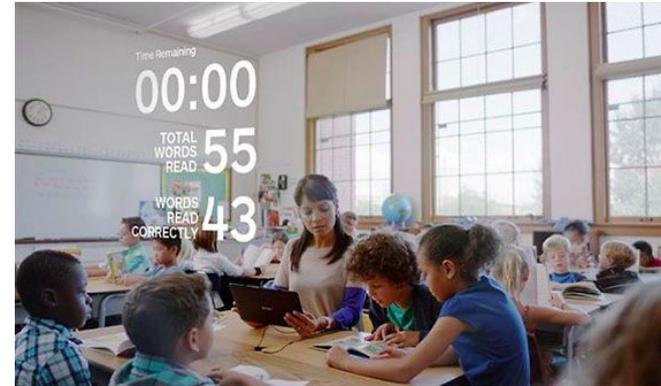
# Leadership Buy In



# Individual Buy In



# Stakeholder Buy In



# Thank You!

<http://dssg.uchicago.edu/>

<http://dsapp.uchicago.edu/>

DSSG 2017 Project Partner  
Applications Due 1/31/17!

# Case Studies



# WHO IS DROPPING OUT?



EVERY YEAR, 1.3 MILLION STUDENTS DROP OUT OF SCHOOL.

**7,000** EVERY DAY  
OF SCHOOL **OR** **1** EVERY 26  
SECONDS





Impaired Attention

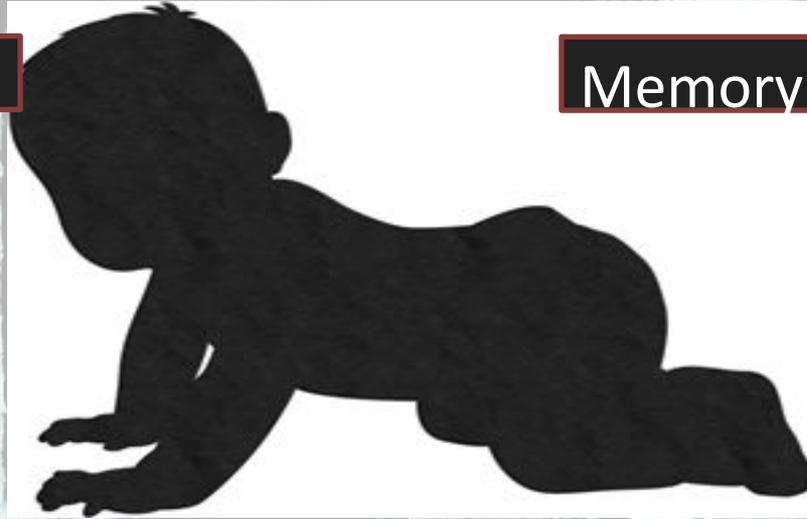
Lack of Motor Skills

Hearing Loss

Learning Disability

Lower IQ

Memory Problems





Constituents



**AUSTRALIAN  
CONSERVATION  
FOUNDATION**



Elected Representatives





# Input

Colonia name

ZUMPANGO, ZUMPANGO, MEXICO, 28728

Age

29

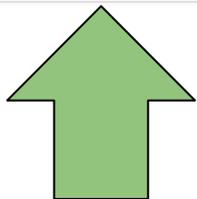
Risk index

3.2

Daily wage

120.5

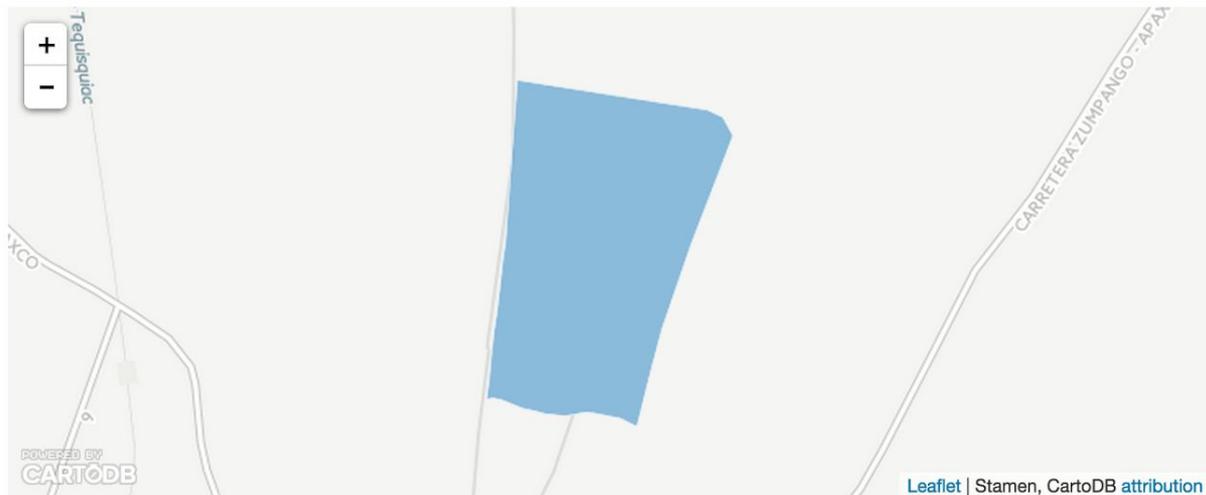
Predict



Personal features and location

# Prediction

Zumpango, Zumpango, Mexico



Probability of abandonment

28.2%

Top factors contributing to home abandonment

#	Factor
0	Long distance to schools
1	Lack of healthcare services
2	High crime rate

