The Two Cultures of People Science

Michelangelo D'Agostino
mdagostino@civisanalytics.com
@MichelangeloDA

CIVIS™ ANALYTICS

“For thirty years I have had to be in touch with scientists not only out of curiosity, but as part of a working existence…I believe the intellectual life of the whole of western society is increasingly being split into two polar groups…at one pole, we have the literary intellectuals…at the other scientists…Between the two a gulf of mutual incomprehension—sometimes (particularly among the young) hostility and dislike, but most of all lack of understanding.”

C. P. Snow, *The Two Cultures*, 1959

"For several years now I have had to be in touch with social scientists not only out of curiosity, but as part of a working existence…I believe the intellectual life of those who want to understand and predict human behavior is increasingly being split into two polar groups…at one pole, we have the social scientists… at the other data scientists…Between the two a gulf of mutual incomprehension—sometimes (particularly among the young) hostility and dislike, but most of all lack of understanding."

Me, 2015

Data Scientists

- tree ensembles, neural networks, SVM's, and other machine learning algorithms
  - regularization and variable selection

Social Scientists

- linear regressions and hand-crafted, theory-generated models
  - careful model specification

Data Scientists

- tree ensembles, neural networks, SVM's, and other machine learning algorithms
  - regularization and variable selection
- emphasis on prediction and out-of-sample validation

Social Scientists

- linear regressions and hand-crafted, theory-generated models
  - careful model specification
- inference: understanding and interpretation of in-sample coefficients

Data Scientists

- tree ensembles, neural networks, SVM's, and other machine learning algorithms
  - regularization and variable selection
- emphasis on prediction and out-of-sample validation
- "data exhaust"

Social Scientists

- linear regressions and hand-crafted, theory-generated models
  - careful model specification
- inference: understanding and interpretation of in-sample coefficients
- careful survey and experiment design

Data Scientists

- tree ensembles, neural networks, SVM's, and other machine learning algorithms
  - regularization and variable selection
- emphasis on prediction and out-of-sample validation
- "data exhaust"
- A/B testing

Social Scientists

- linear regressions and hand-crafted, theory-generated models
  - careful model specification
- inference: understanding and interpretation of in-sample coefficients
- careful survey and experiment design
- causal reasoning: natural experiments, regression discontinuity analysis, instrumental variables…

Data Scientists

- tree ensembles, neural networks, SVM's, and other machine learning algorithms
  - regularization and variable selection
- emphasis on prediction and out-of-sample validation
- "data exhaust"
- A/B testing
- optimization and massively parallel computing

Social Scientists

- linear regressions and hand-crafted, theory-generated models
  - careful model specification
- inference: understanding and interpretation of in-sample coefficients
- careful survey and experiment design
- causal reasoning: natural experiments, regression discontinuity analysis, instrumental variables…

Data Scientists

- tree ensembles, neural networks, SVM's, and other machine learning algorithms
  - regularization and variable selection
- emphasis on prediction and out-of-sample validation
- "data exhaust"
- A/B testing
- optimization and massively parallel computing
- software engineering and version control

Social Scientists

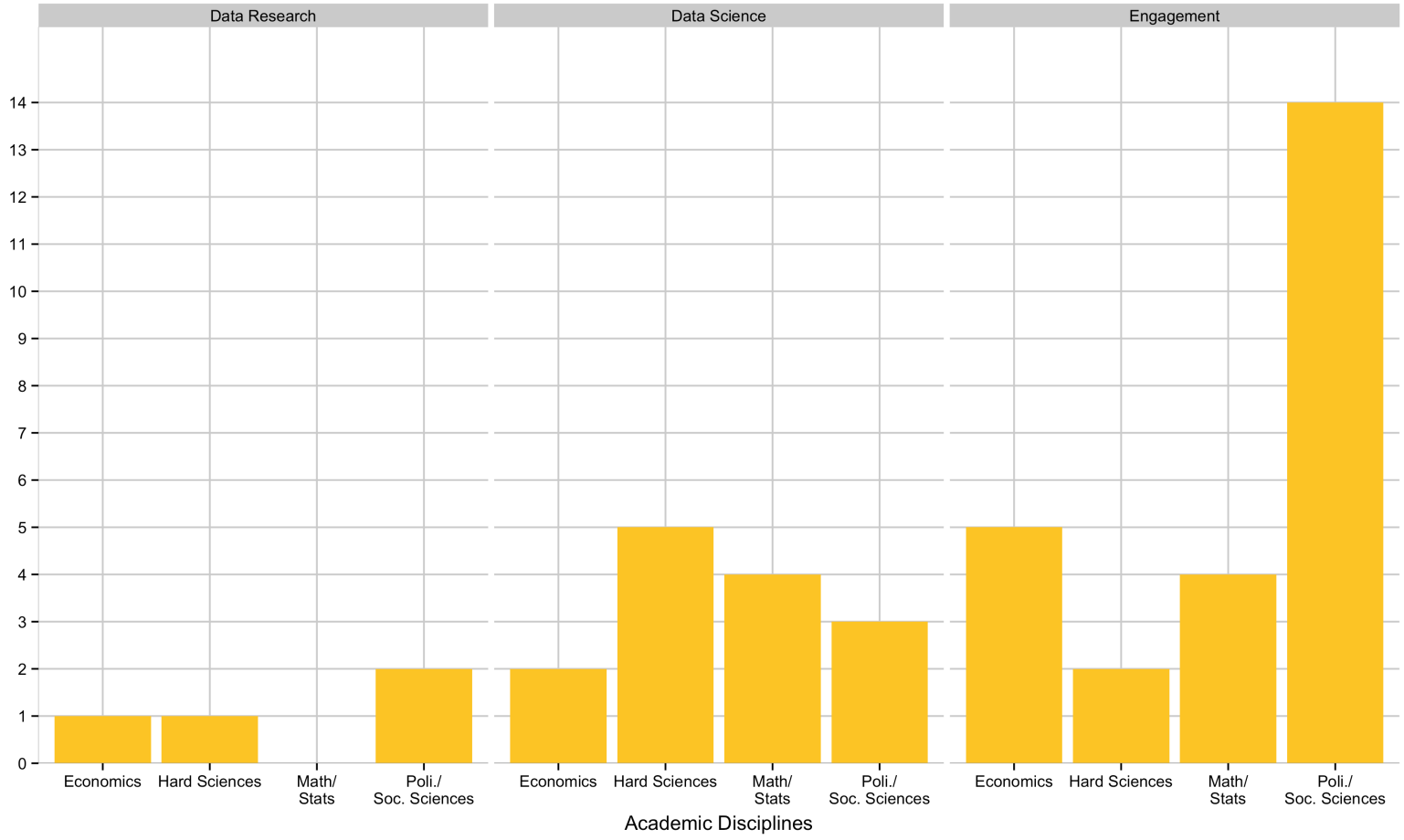
- linear regressions and hand-crafted, theory-generated models
  - careful model specification
- inference: understanding and interpretation of in-sample coefficients
- careful survey and experiment design
- causal reasoning: natural experiments, regression discontinuity analysis, instrumental variables…

We ran the first individualized presidential campaign.

We built a scientific understanding of each voter.

Our data science targeted voters through paid media, direct mail, social media, communications and fundraising.

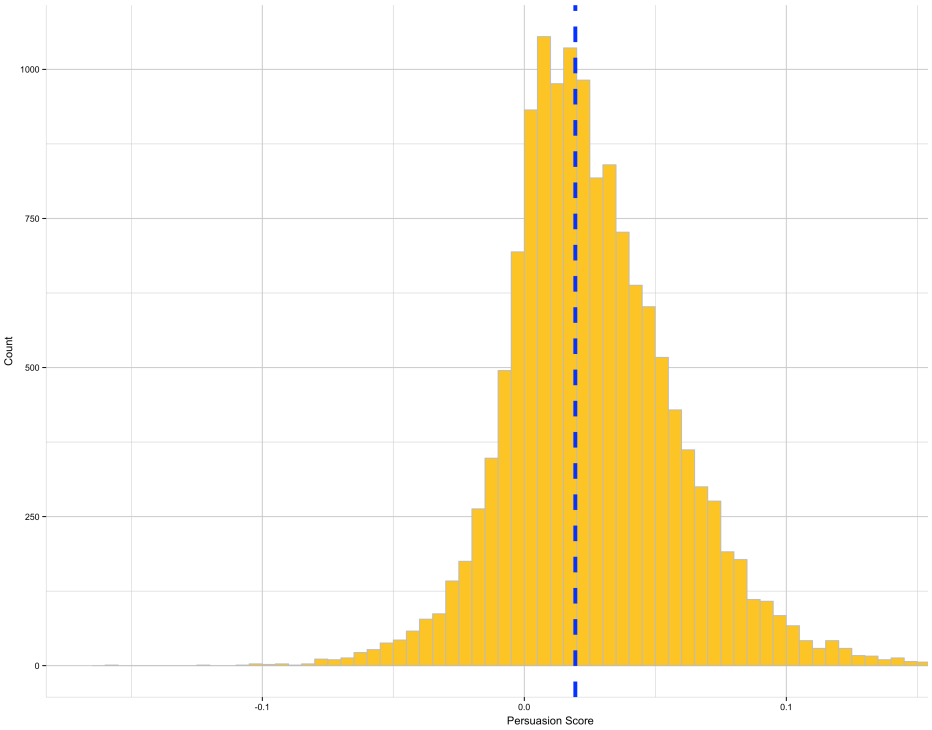
Our data science directed decision makers' strategies and tactics.


Data Research
Data Science
Engagement
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
Economics
Hard Sciences
Math/ Stats
Poli./ Soc. Sciences
Academic Disciplines

Opportunities for Collaboration

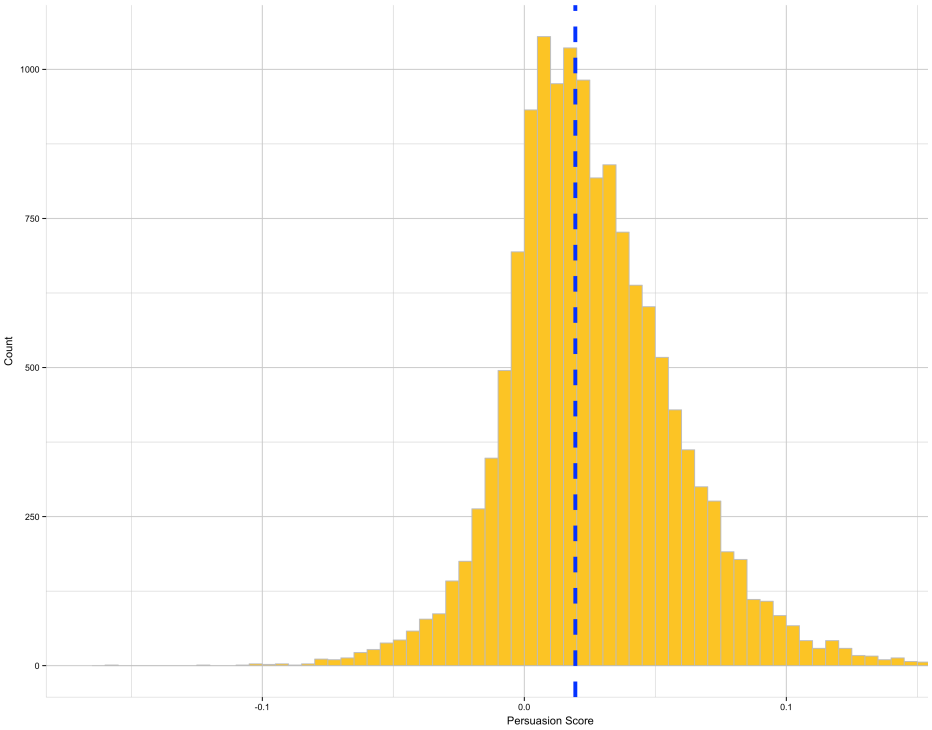
- Data science can help traditional social science.
  - modeling techniques, large scale computation, optimization methods
- Social science can help general data teams.
  - proper survey and experimental design
  - understanding biases
  - causal reasoning and methods like instrumental variables, regression discontinuity…

Finding Persuadable Voters



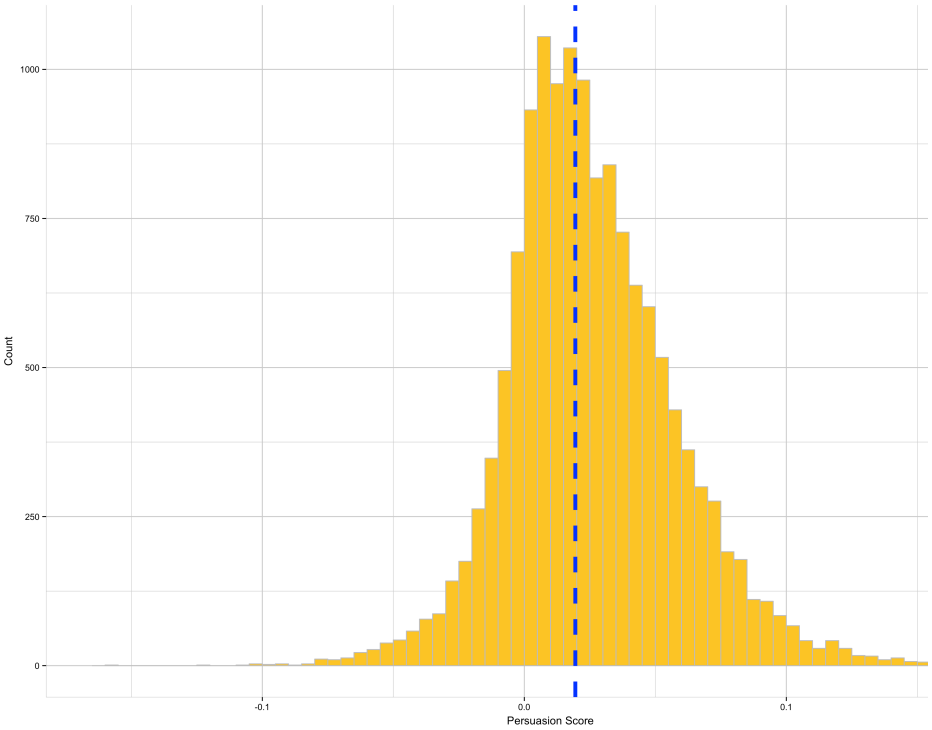
Problem: We want to be able to determine how much of an effect a TV commercial, a phone conversation, or an in-person canvassing conversation will have on an *individual voter's* perception, given all of that voter's known characteristics.

Finding Persuadable Voters



Solution: We intelligently designed a randomized controlled experiment and then used techniques from machine learning and large scale computation to assign a persuasion score to each individual in the country.

Finding Persuadable Voters



Our Approach: Robust python framework utilizing machine learning techniques like tree ensembles paired with massive computational infrastructure on AWS

Team: Political scientists, statisticians, and a data scientist

Poll Aggregation and Forecasting

Problem: What's the best way to combine information from (biased) public polling, other external sources, and internal polling to estimate true latent public sentiment and to forecast election results?

Solution: A Bayesian, multi-level dynamic linear model, along with a python framework for managing and distributing the MCMC computations and updating the models daily

Team: A statistician, a data scientist, and numerous social scientists

Survey Raking

Problem: Even well-designed surveys suffer from issues of bias: the people that respond are not usually representative of the target population that you want to study (all adults, registered voters, likely voters…).

Solution: Assign each respondent a weight to get back to the desired target universe.

Survey Raking

Traditional Approach: Often done by hand in a spreadsheet or by considering only simple one-way marginal distributions on characteristics like age or race

Our Approach: Python package utilizing modern optimization techniques to find the best set of weights that matches the samples on an arbitrary number of traits and combinations of traits

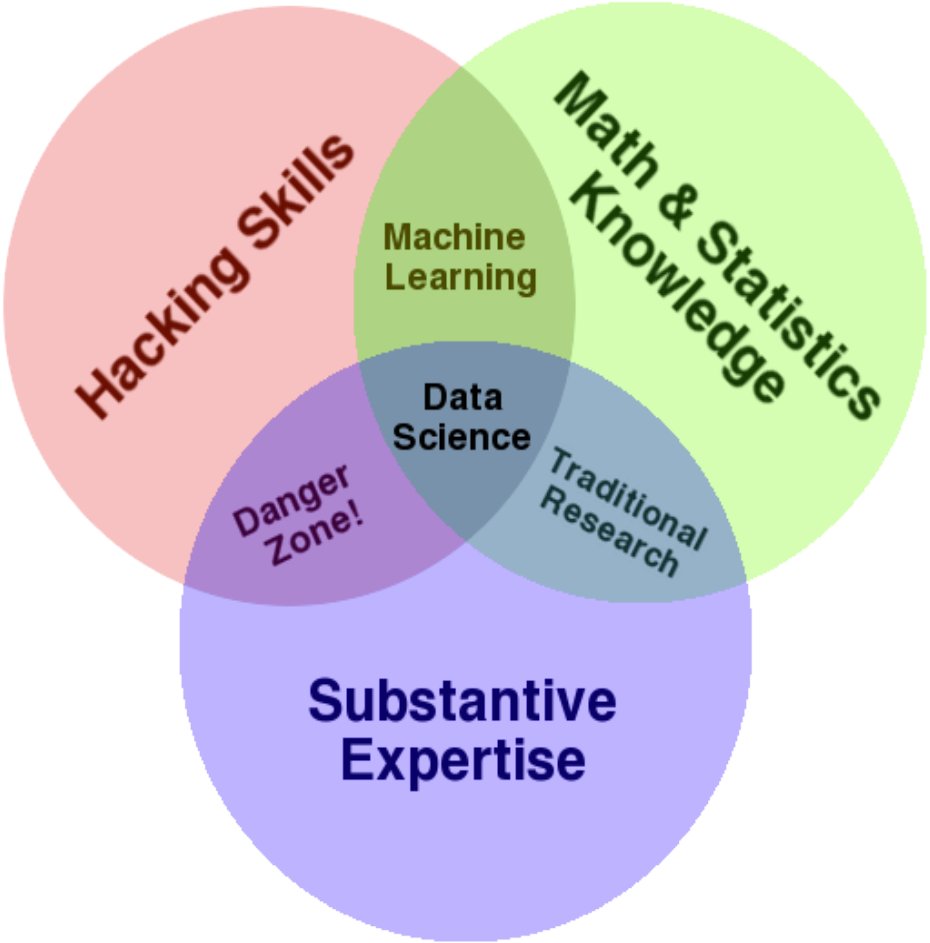
Team: Survey scientists, a statistician, and a data scientist

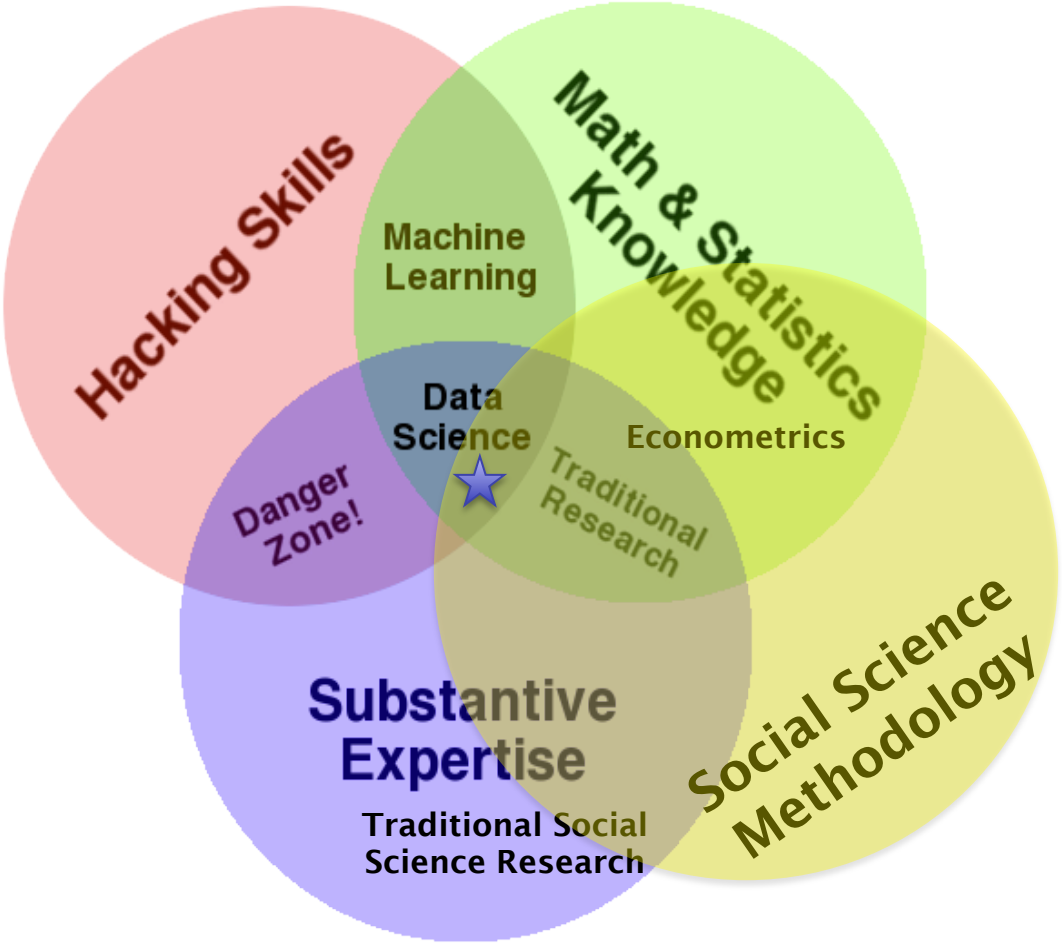
Transfer Learning for Small Polls

Problem: Sometimes, we have very small surveys with which to model public opinion. However, we also have a large corpus of public opinion polls on related issues and candidates.

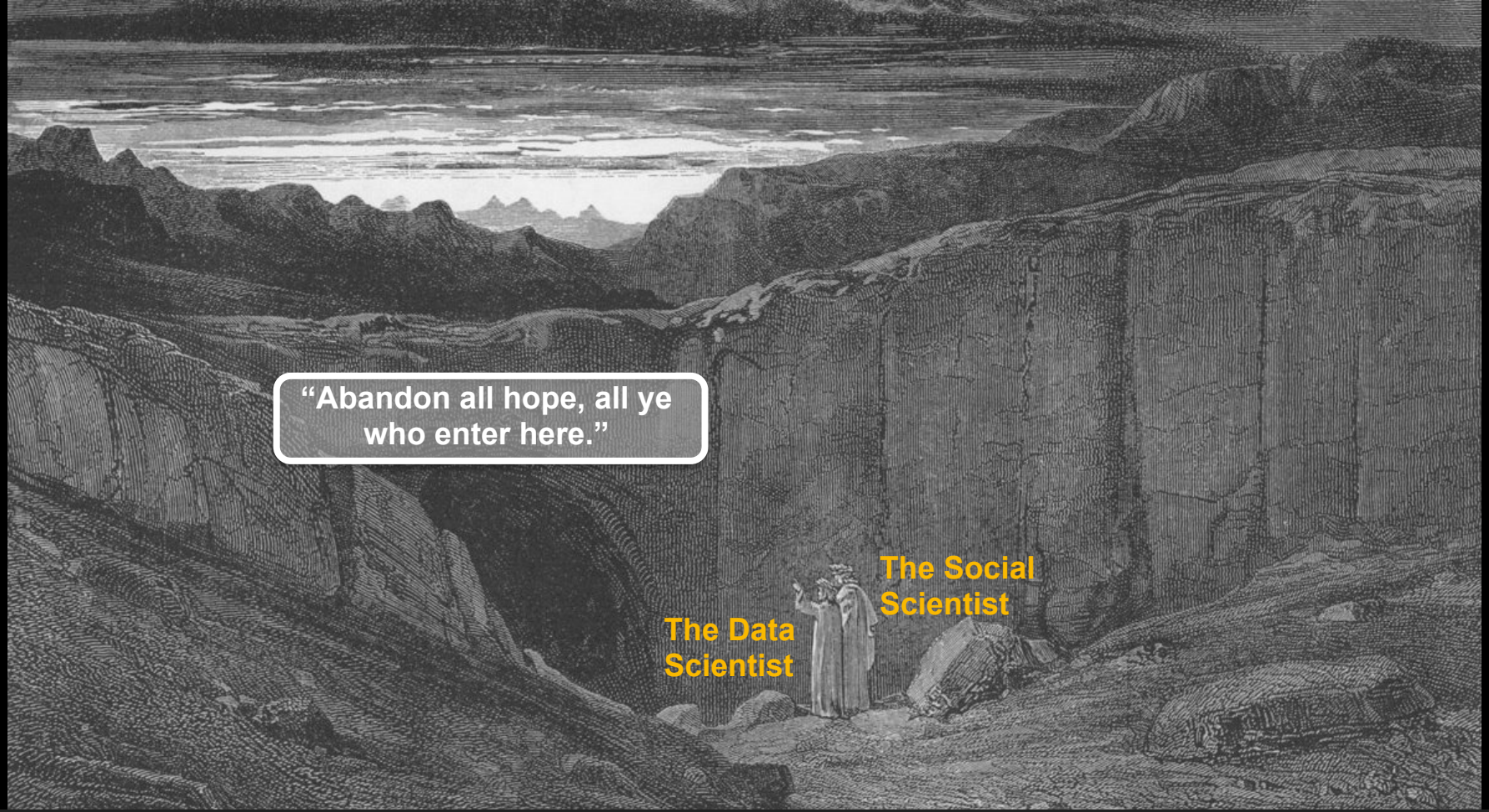
Our Approach: Transfer Learning: an intelligent way to determine related polls and to construct Bayesian priors for modeling the new, small survey

Team: A statistician, a machine learning theoretician, a data scientist, and numerous political scientists


Hacking Skills
Math & Statistics Knowledge
Machine Learning
Data Science
Danger Zone!
Traditional Research
Substantive Expertise

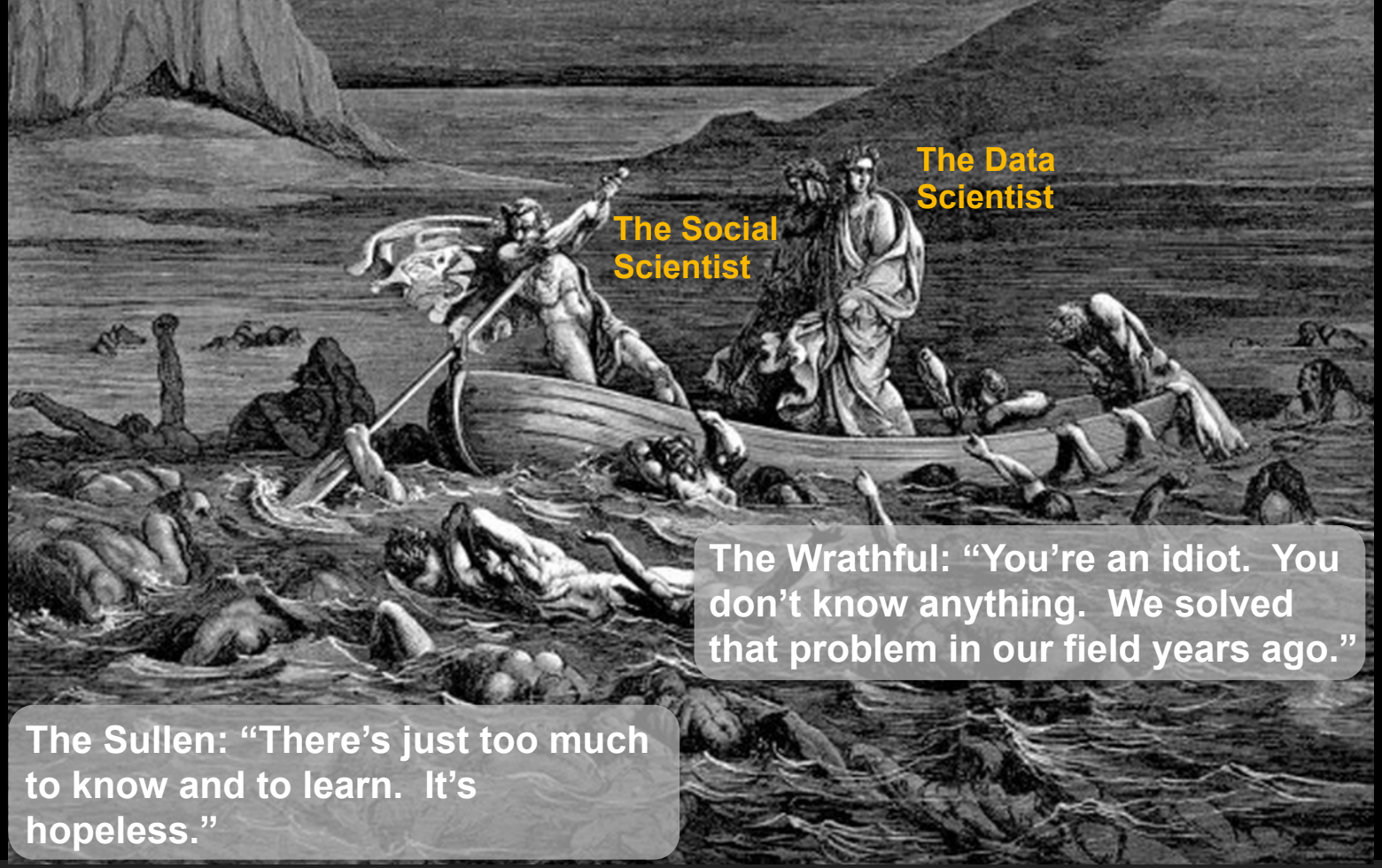

Hacking Skills
Math & Statistics Knowledge
Machine Learning
Data Science
Econometrics
Traditional Research
Danger Zone!
Substantive Expertise
Traditional Social Science Research
Social Science Methodology

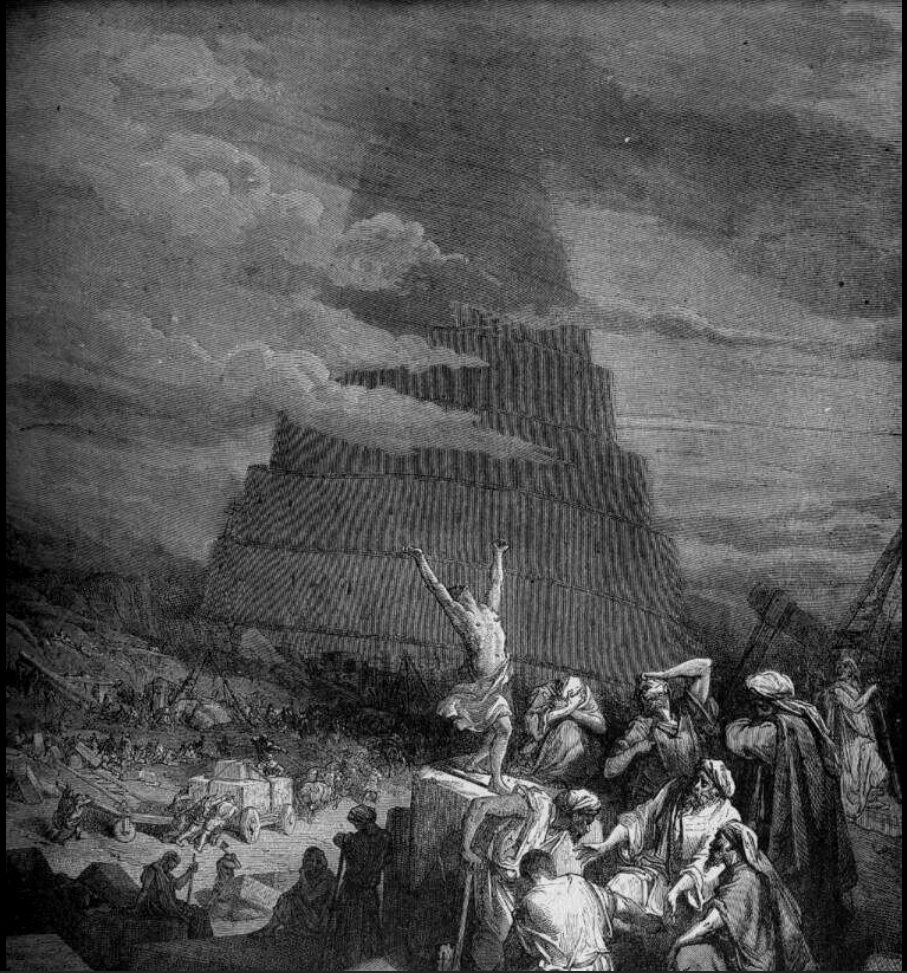

The Data Scientist
The Social Scientist

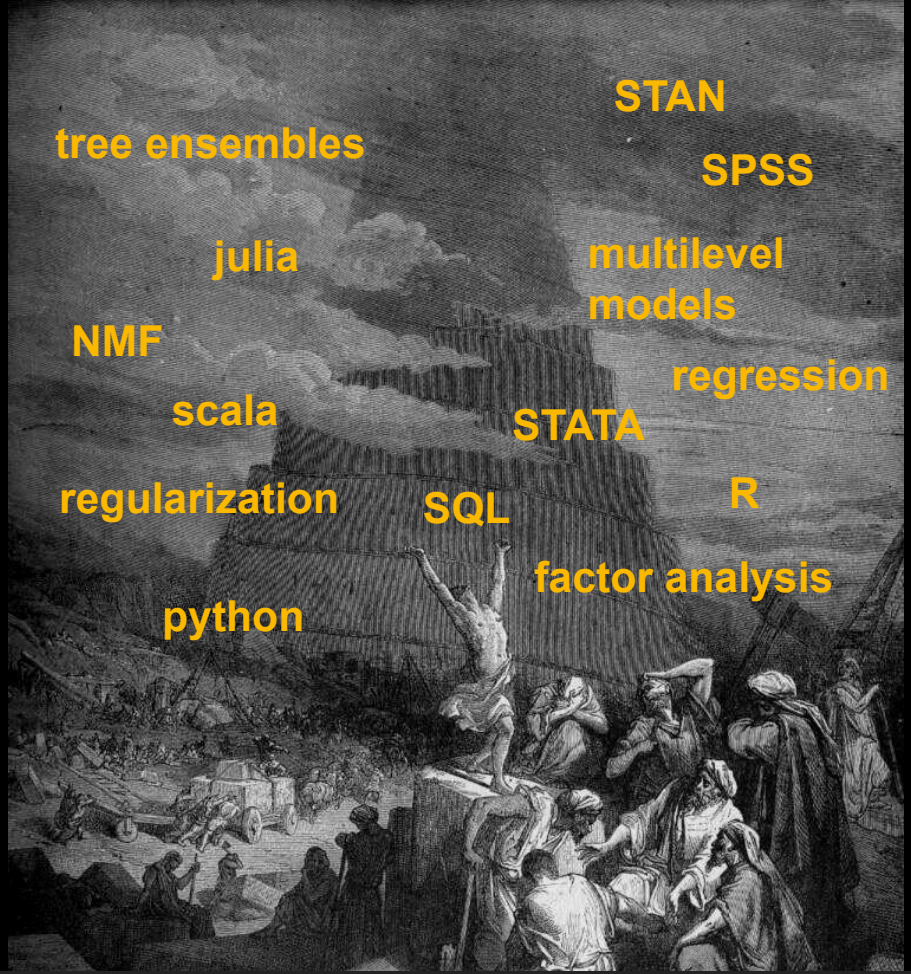

"Abandon all hope, all ye who enter here."
The Data Scientist
The Social Scientist

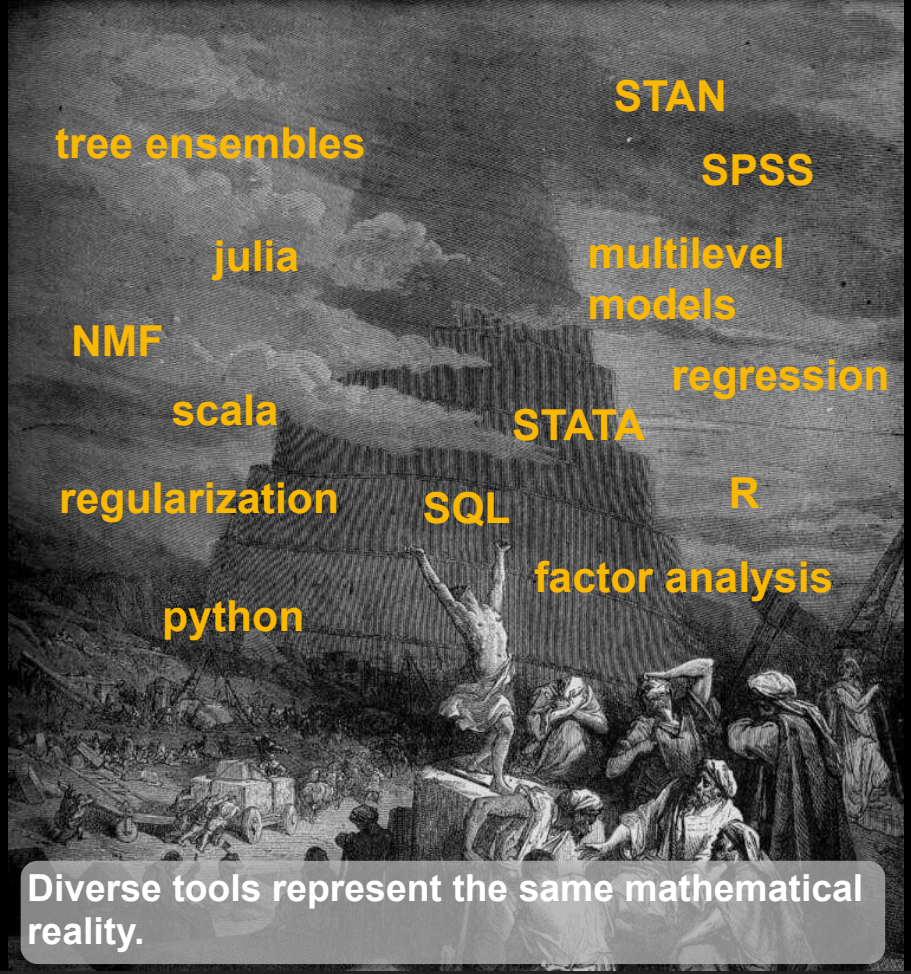

"Abandon all arrogance,
all ye who enter here."
The Data Scientist
The Social Scientist


The Data Scientist
The Social Scientist


The Data Scientist
The Social Scientist
The Wrathful: “You’re an idiot. You don’t know anything. We solved that problem in our field years ago.”
The Sullen: “There’s just too much to know and to learn. It’s hopeless.”




STAN
tree ensembles
SPSS
julia
multilevel models
NMF
regression
scala
STATA
regularization
SQL
R
factor analysis
python



Diverse tools represent the same mathematical reality.

Final Tips

- Leave your arrogance at the door.
- Focus on ideas, not diversity of tools.
- Aggressively, but politely, ask questions and try to reframe techniques and ideas in your own “language”.
- If there’s a person that can bridge the two sides, with deep expertise in both areas, that’s ideal.
- Most importantly, always keep talking…


Blissful
Collaboration
The Data
Scientist
The Social
Scientist