Bayesian Machine Learning

Quantifying uncertainty and robustness at scale

Tamara Broderick
ITT Career Development Assistant Professor
tbroderick@mit.edu


Raj Agrawal


Trevor Campbell


Lorenzo Masoero


Will Stephenson

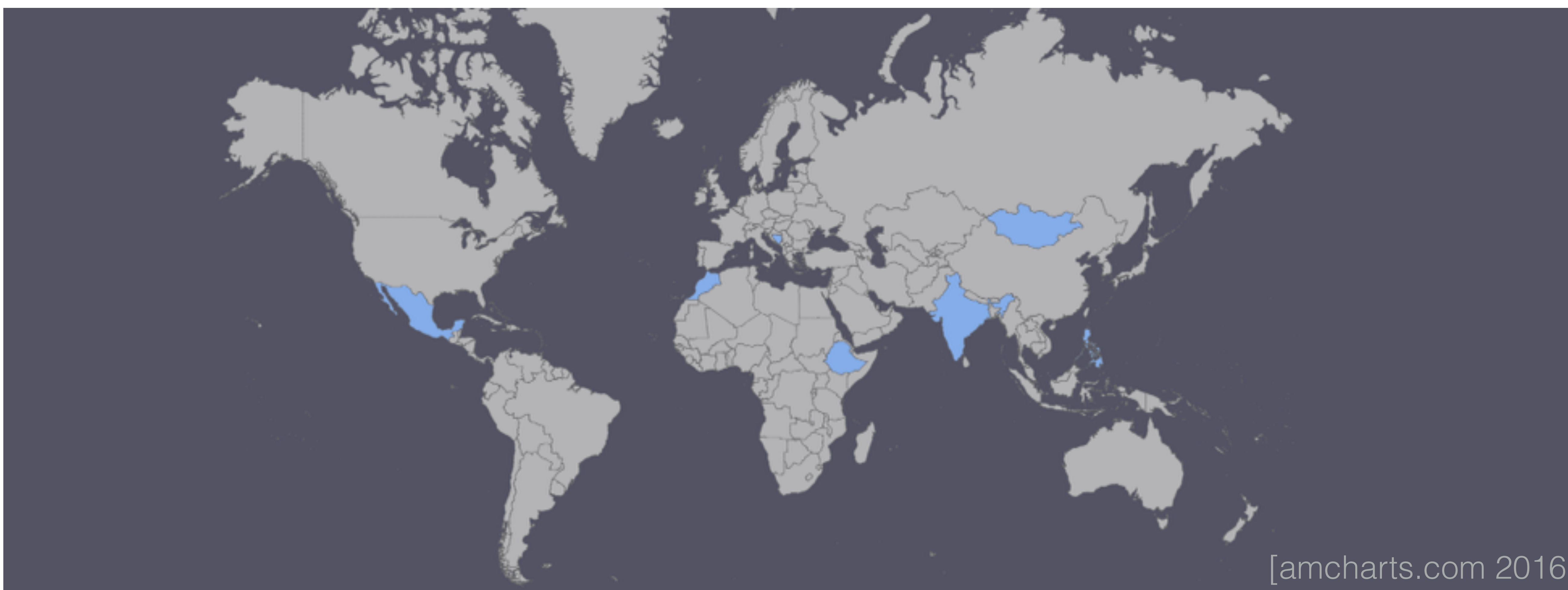
Microcredit Experiment

Microcredit Experiment

- Simplified from Meager (2016)

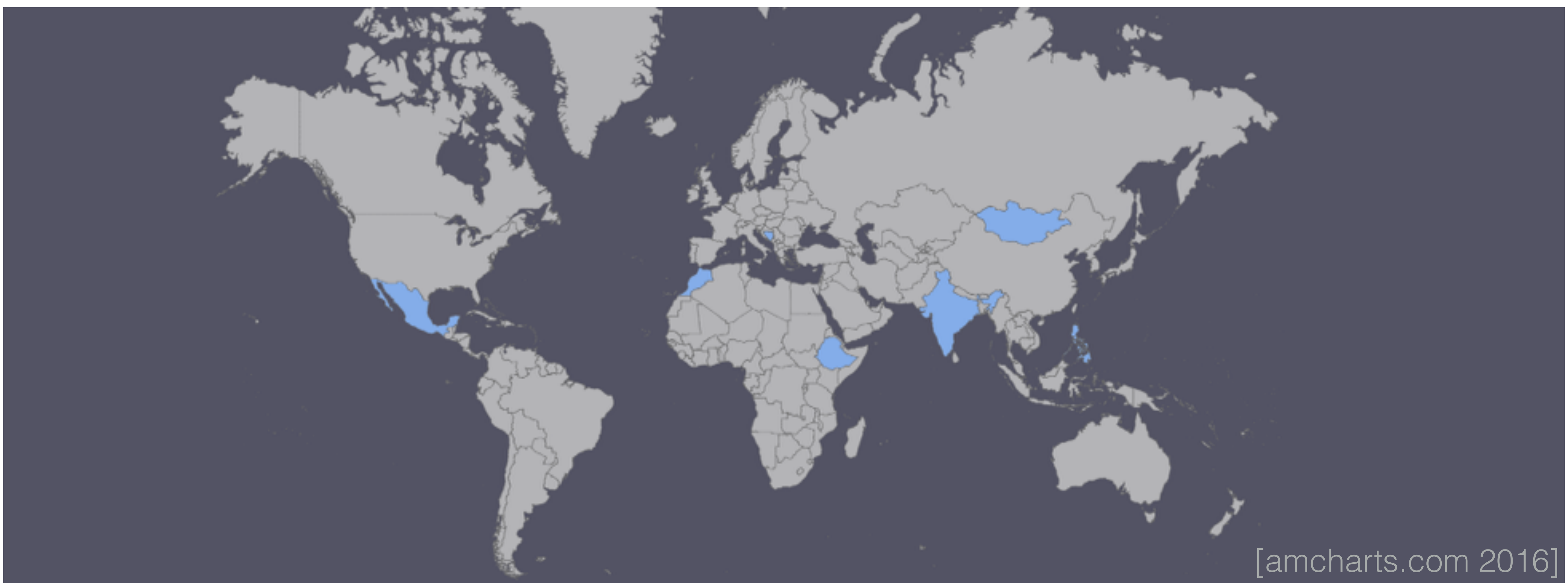
Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)



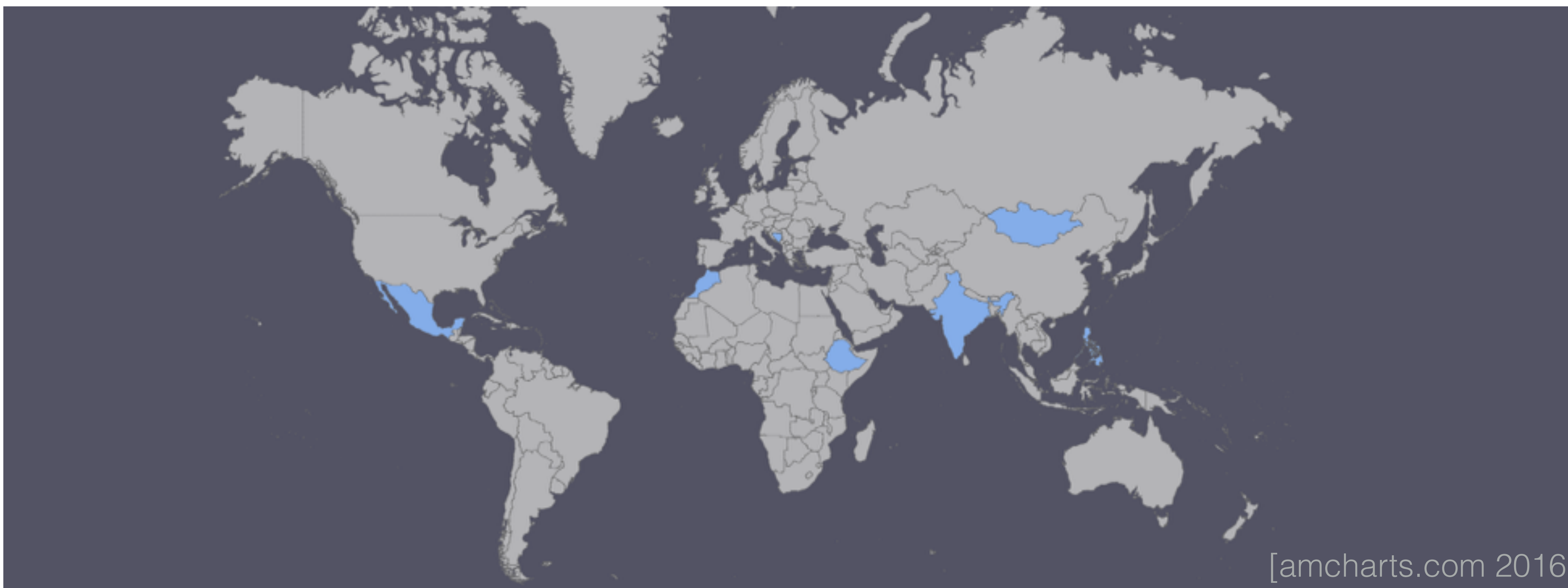
Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site



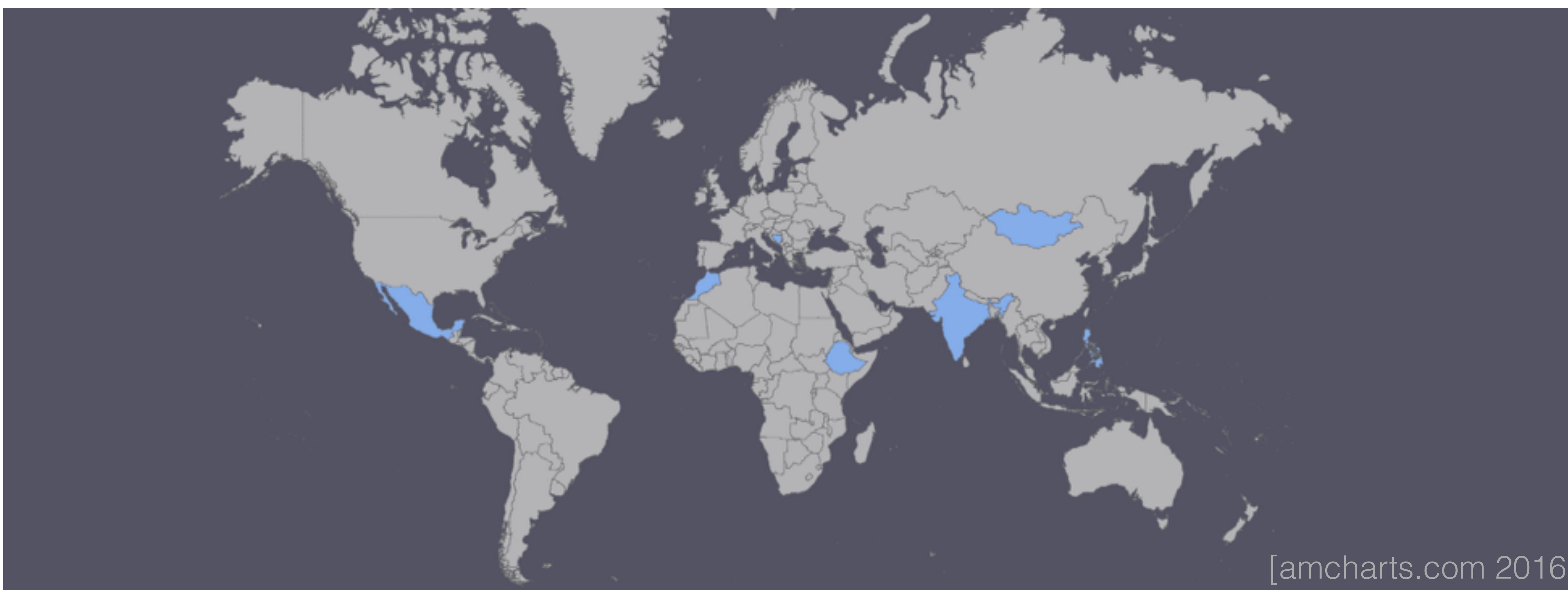
Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit?



Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ



Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:
  - Interpretability

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:
    - Interpretability; incorporate expert info

[amcharts.com 2016]

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:
  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)

[amcharts.com 2016]

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:
  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Share information across experiments

[amcharts.com 2016]

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:
  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Share information across experiments

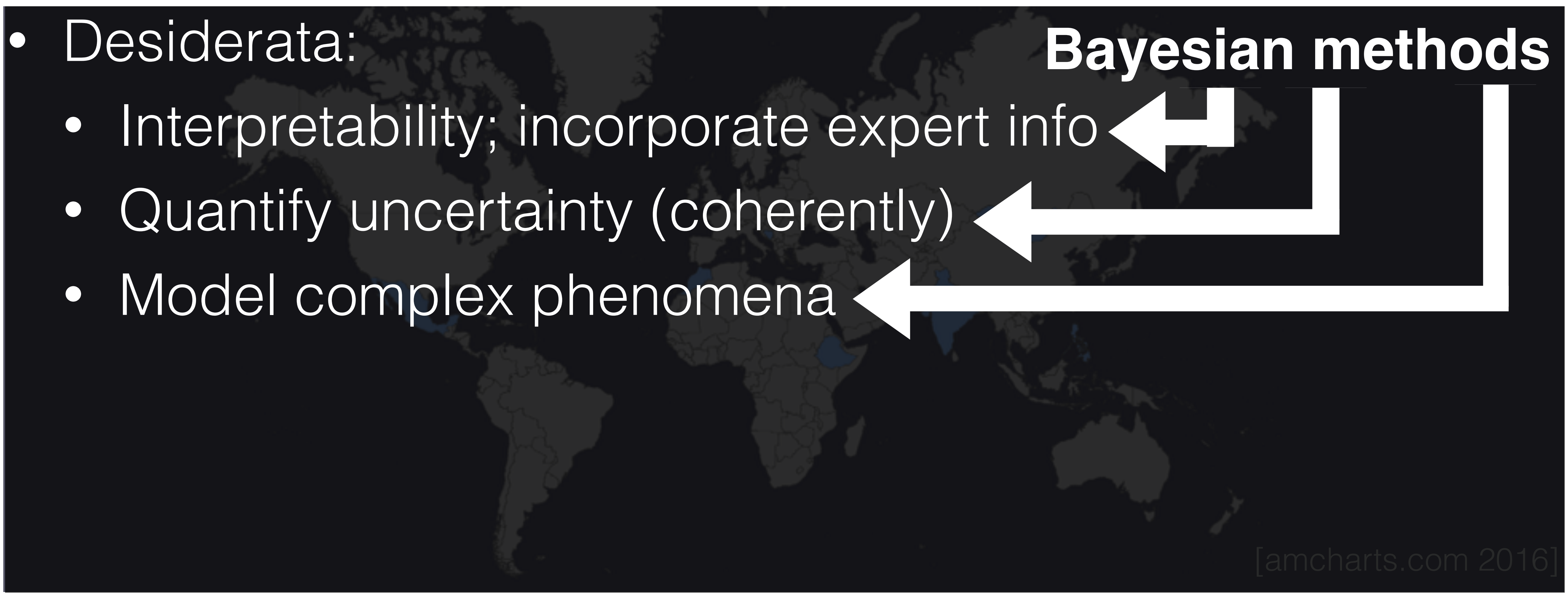
Bayesian methods

[amcharts.com 2016]

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:
  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Model complex phenomena

Bayesian methods



[amcharts.com 2016]

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:
  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Model complex phenomena
  - Get fast results (K, M, B+)

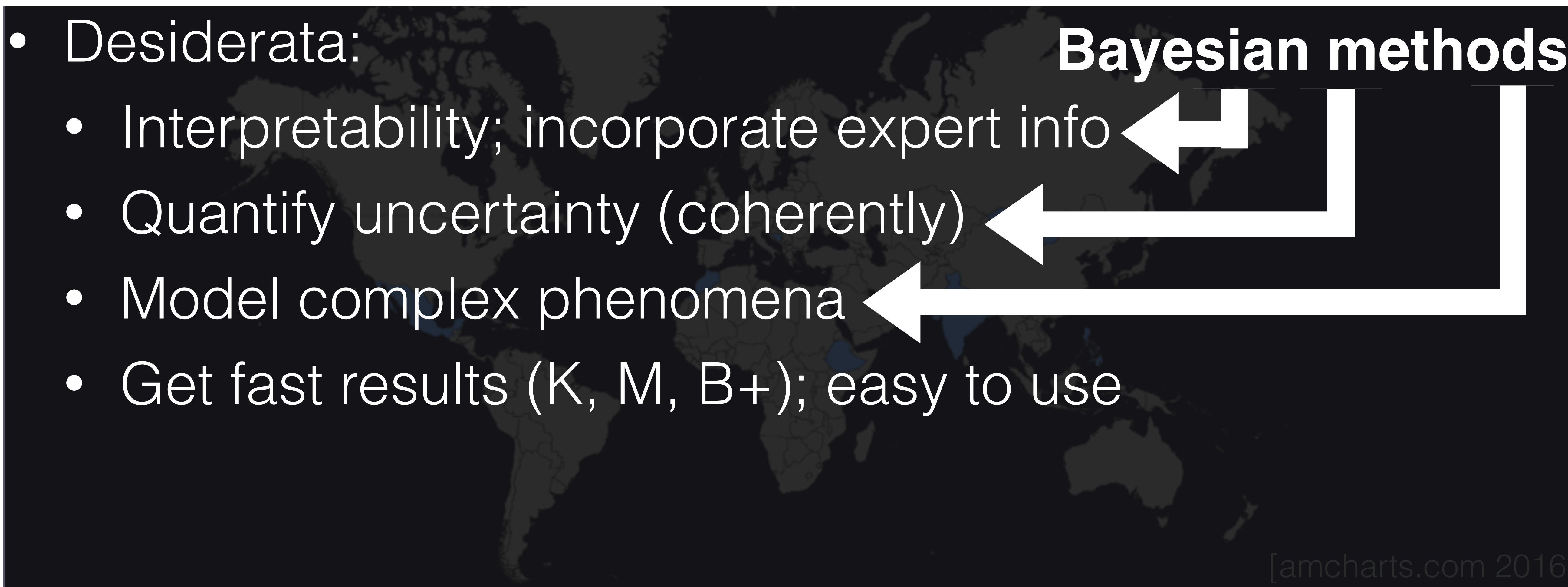
Bayesian methods

[amcharts.com 2016]

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:
  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Model complex phenomena
  - Get fast results (K, M, B+); easy to use

Bayesian methods



[amcharts.com 2016]

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:

  Bayesian methods

  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Model complex phenomena
  - Get fast results (K, M, B+); easy to use
  - Finite-data, finite-run-time theoretical guarantees

[amcharts.com 2016]

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:
  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Model complex phenomena
  - Get fast results (K, M, B+); easy to use
  - Finite-data, finite-run-time theoretical guarantees
  - Robustness quantification

Bayesian methods

[amcharts.com 2016]

Microcredit Experiment

- Simplified from Meager (2016)
- 7 sites with microcredit trials (in Mexico, Mongolia, Bosnia, India, Morocco, Philippines, Ethiopia)
- ~900 to ~17K businesses at each site
- Q: how much does microcredit increase business profit? τ
- Desiderata:

  Bayesian methods

  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Model complex phenomena
  - Get fast results (K, M, B+); easy to use
  - Finite-data, finite-run-time theoretical guarantees
  - Robustness quantification

[amcharts.com 2016]

Bayesian machine learning

- Desiderata:

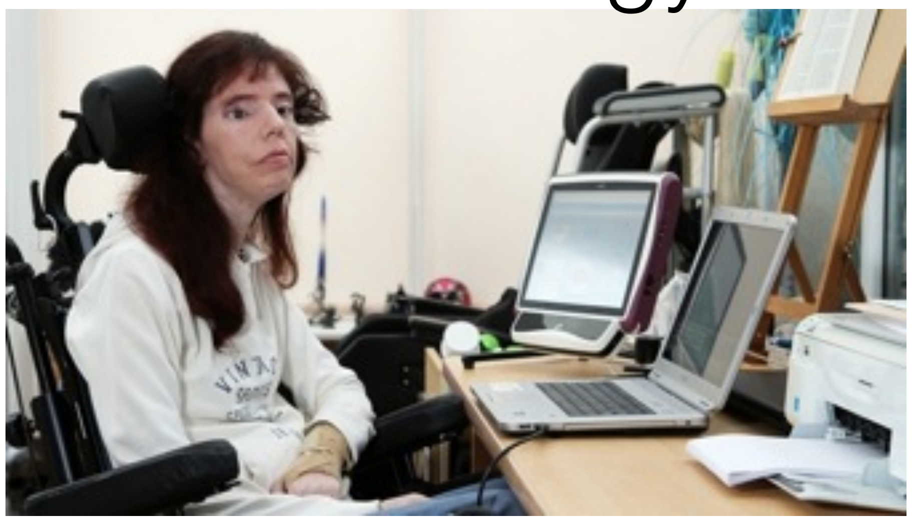
Bayesian methods

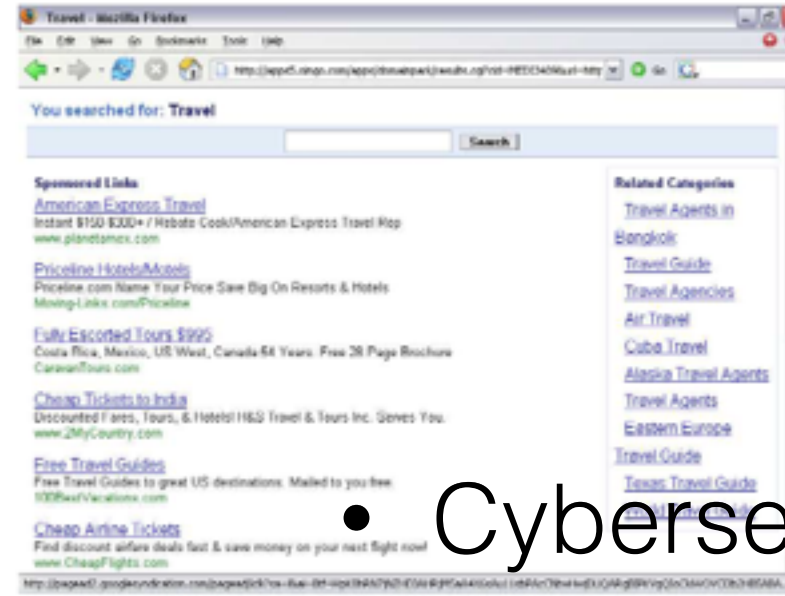
  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Model complex phenomena
  - Get fast results (K, M, B+); easy to use
  - Finite-data, finite-run-time theoretical guarantees
  - Robustness quantification

[amcharts.com 2016]

Bayesian machine learning

- Assistive technology





- Cybersecurity

- Desiderata:
  - Interpretability; incorporate expert info
  - Quantify uncertainty (coherently)
  - Model complex phenomena
  - Get fast results (K, M, B+); easy to use
  - Finite-data, finite-run-time theoretical guarantees
  - Robustness quantification

Bayesian methods

[amcharts.com 2016]

[image sources: Webb, Caverlee, Pu 2006 , www.itv.com/news/central/story/2014-08-05/locked-in-syndrome-woman-earns-degree-by-blinking]

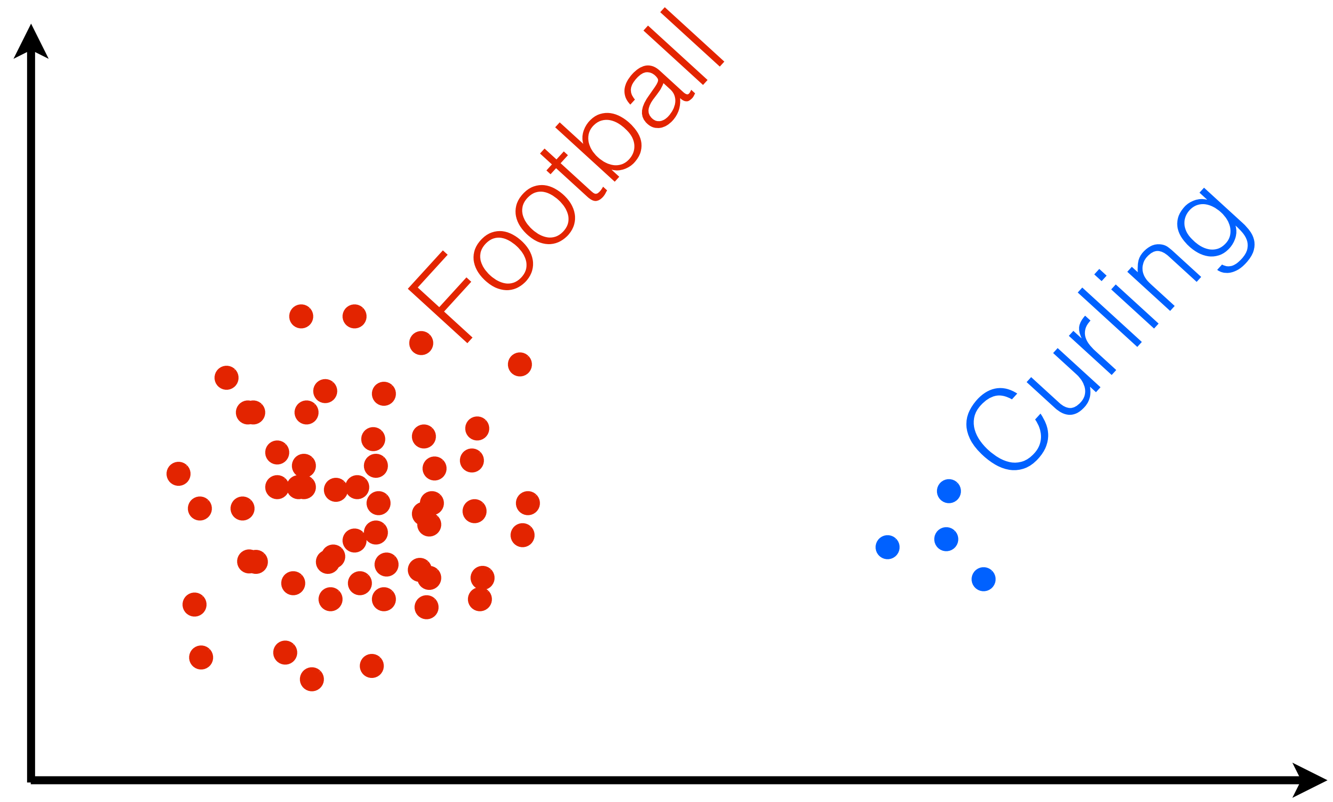
Data summarization

Data summarization

- Observe: redundancies can exist even if data isn't "tall"

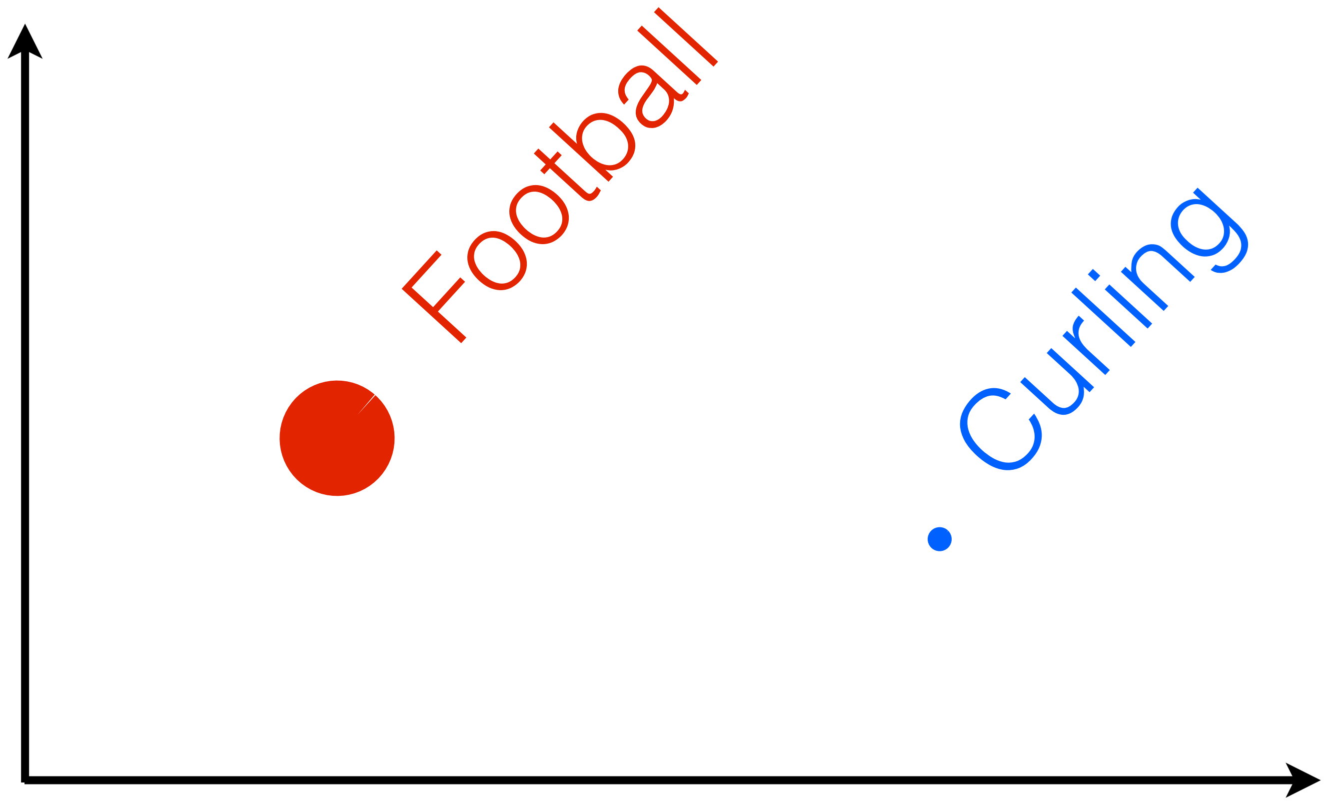
Data summarization

- Observe: redundancies can exist even if data isn't "tall"



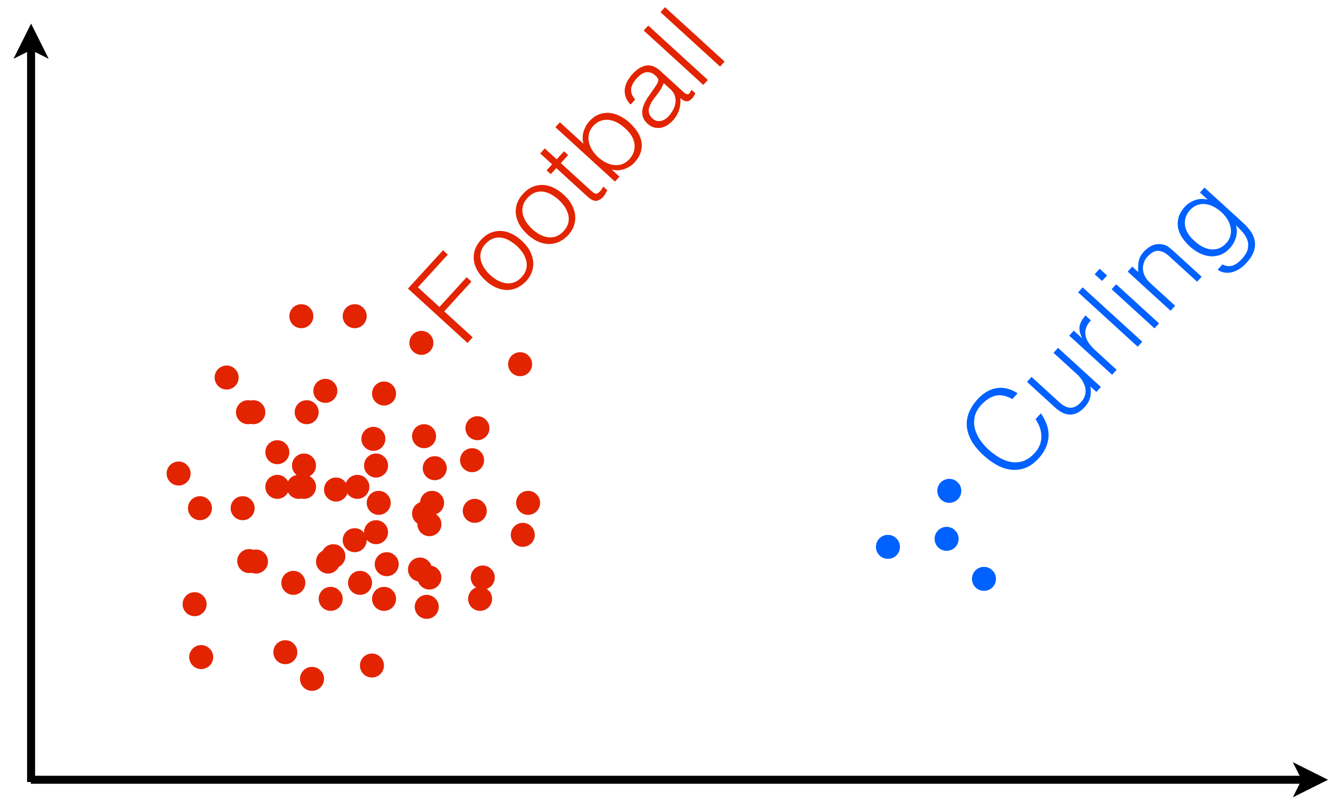
Data summarization

- Observe: redundancies can exist even if data isn't "tall"



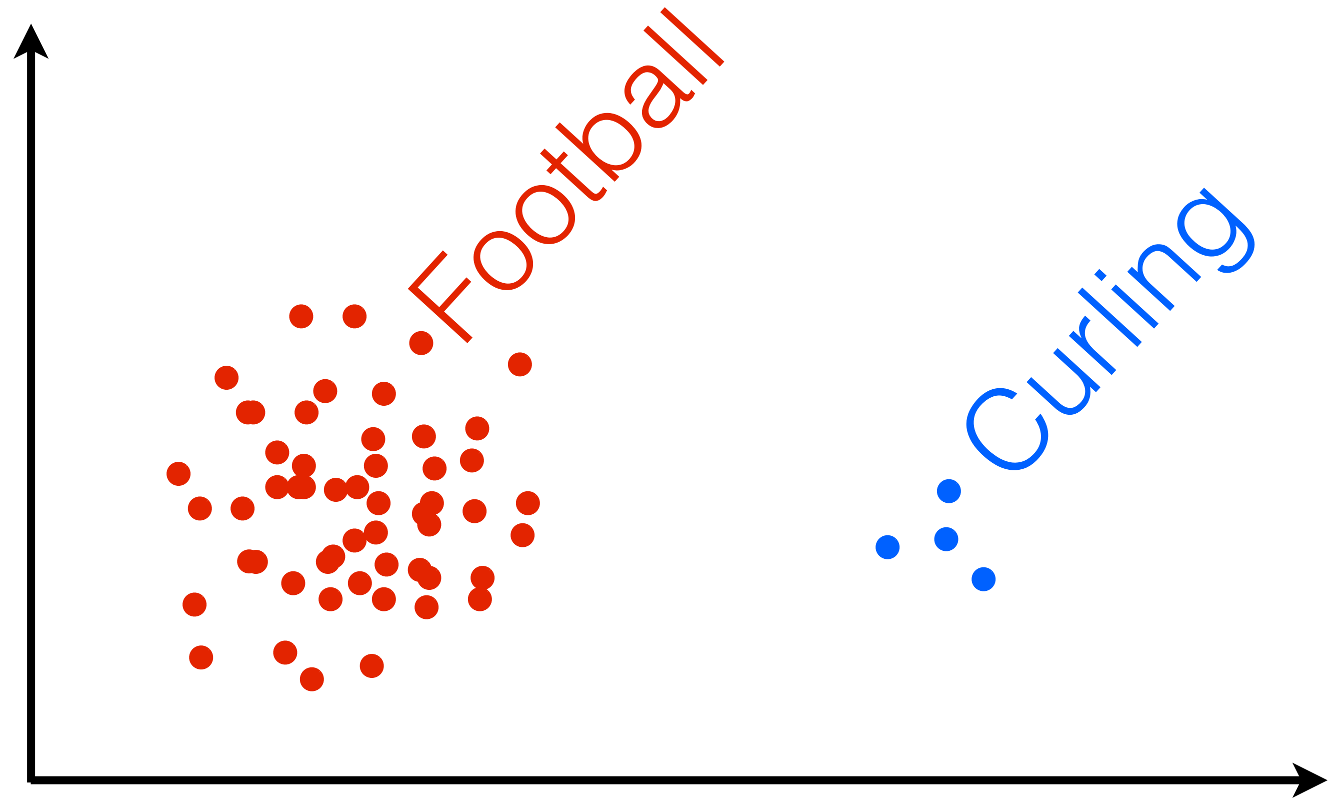
Data summarization

- Observe: redundancies can exist even if data isn't "tall"



Data summarization

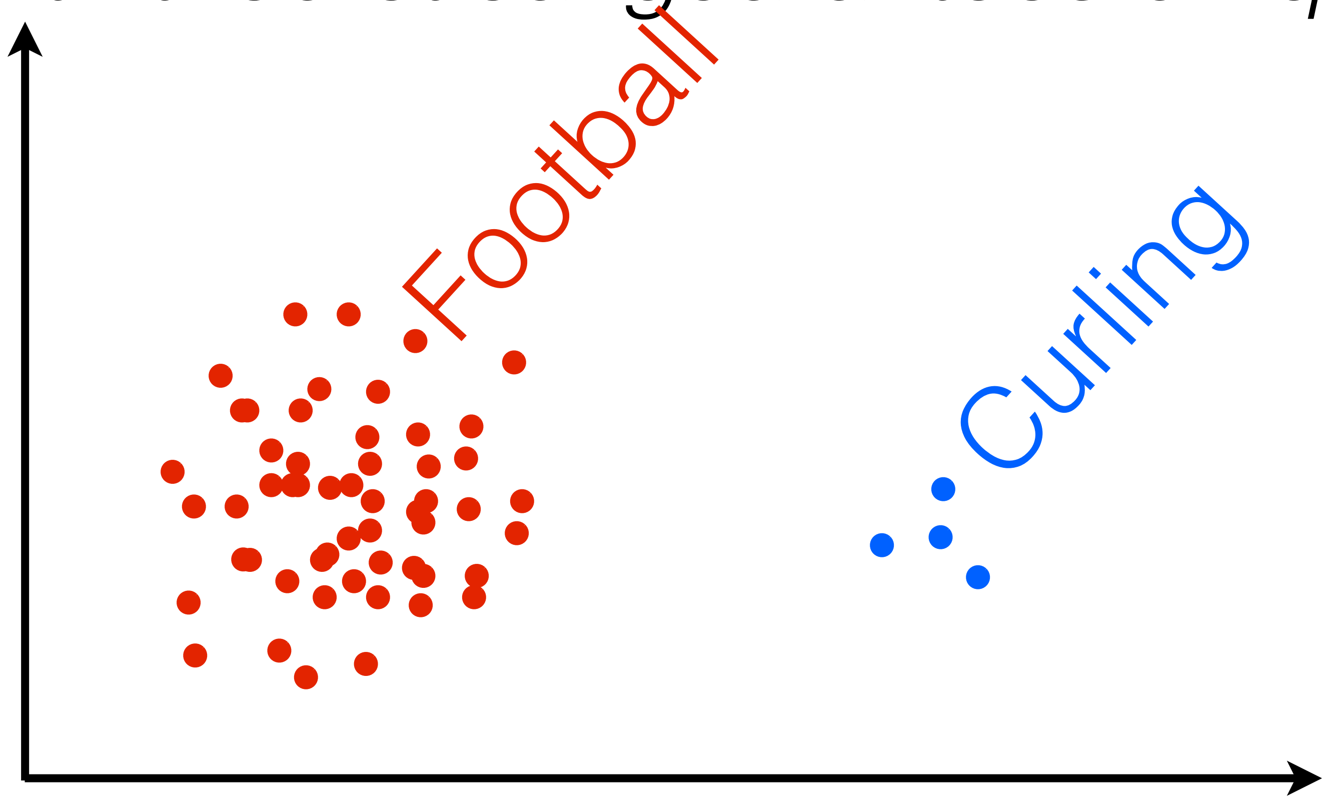
- Observe: redundancies can exist even if data isn't "tall"
- Coresets: pre-process data to get a smaller, weighted data set



[Agarwal et al 2005; Feldman & Langberg 2011]

Data summarization

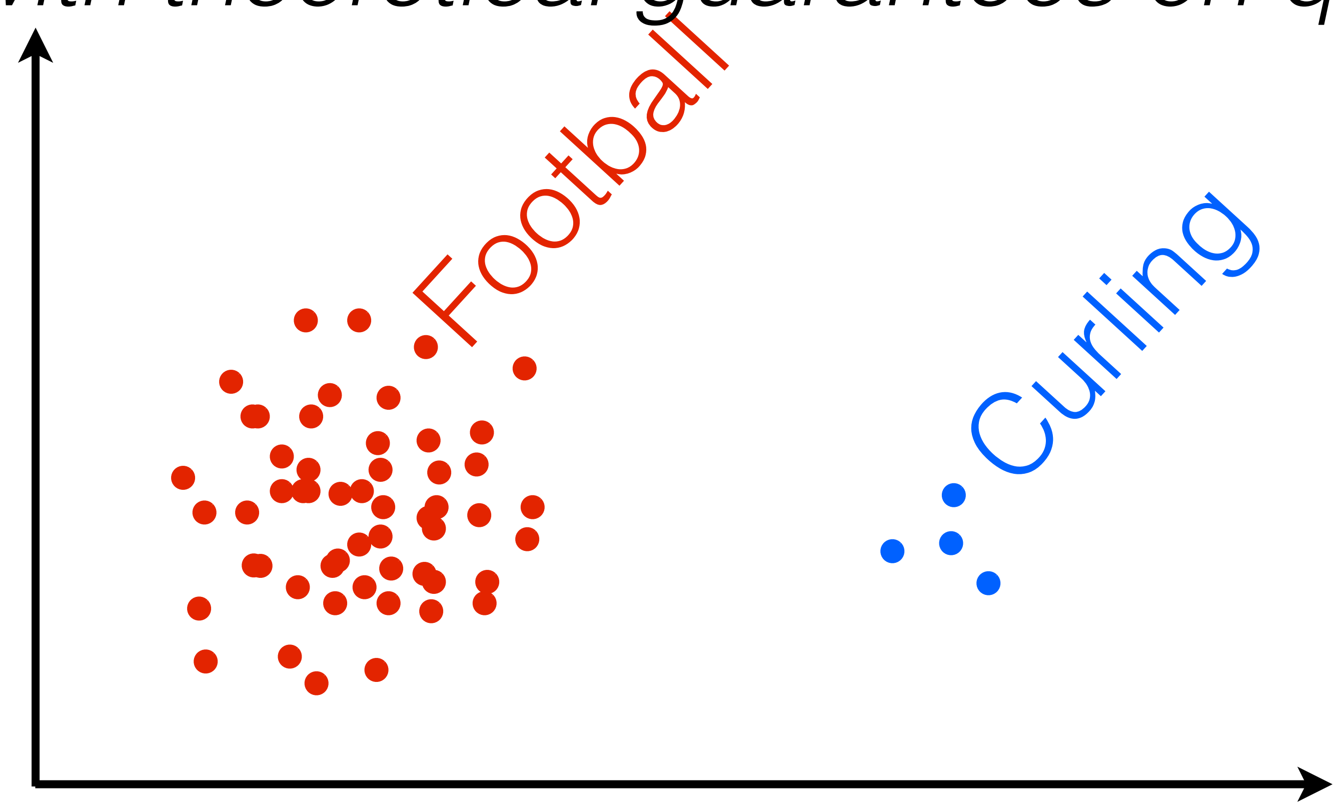
- Observe: redundancies can exist even if data isn't "tall"
- Coresets: pre-process data to get a smaller, weighted data set *with theoretical guarantees on quality*



[Agarwal et al 2005; Feldman & Langberg 2011]

Data summarization

- Observe: redundancies can exist even if data isn't "tall"
- Coresets: pre-process data to get a smaller, weighted data set *with theoretical guarantees on quality*

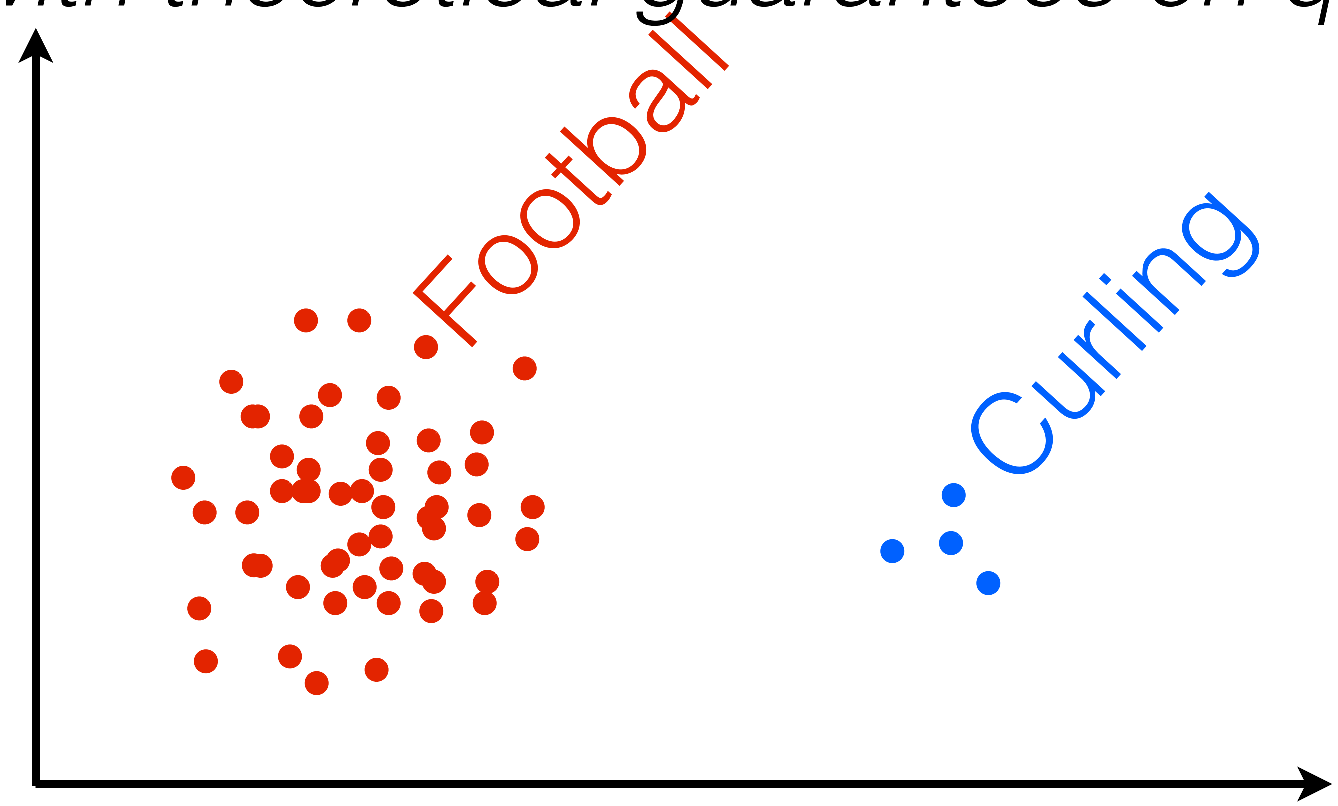


- Previous heuristics: data squashing, inducing points

[Agarwal et al 2005; Feldman & Langberg 2011; DuMouchel et al 1999; Madigan et al 1999]

Data summarization

- Observe: redundancies can exist even if data isn't "tall"
- Coresets: pre-process data to get a smaller, weighted data set *with theoretical guarantees on quality*



- Previous heuristics: data squashing, inducing points
- Cf. subsampling

[Agarwal et al 2005; Feldman & Langberg 2011; DuMouchel et al 1999; Madigan et al 1999]

Data summarization

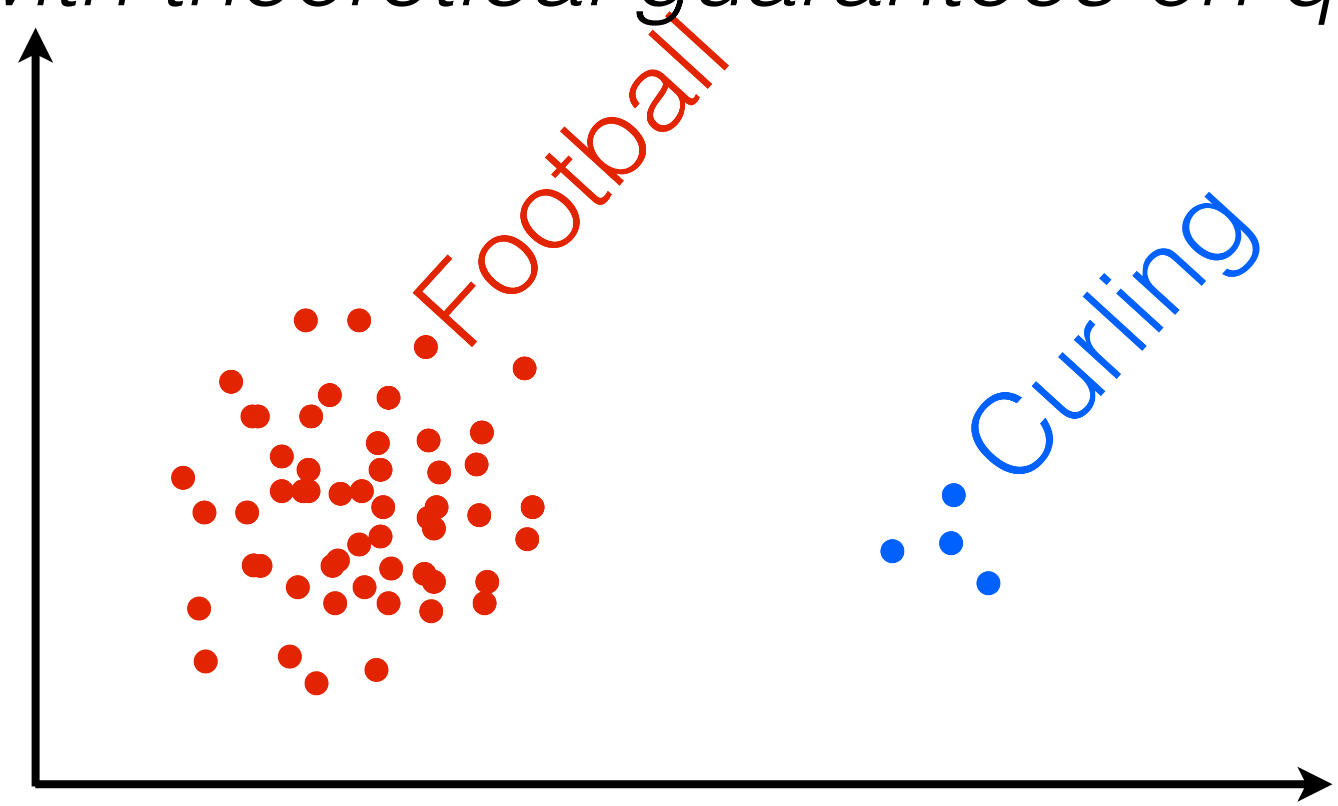
- Observe: redundancies can exist even if data isn't "tall"
- Coresets: pre-process data to get a smaller, weighted data set *with theoretical guarantees on quality*

- Previous heuristics: data squashing, inducing points
- Cf. subsampling

[Agarwal et al 2005; Feldman & Langberg 2011; DuMouchel et al 1999; Madigan et al 1999]

Data summarization

- Observe: redundancies can exist even if data isn't "tall"
- Coresets: pre-process data to get a smaller, weighted data set *with theoretical guarantees on quality*

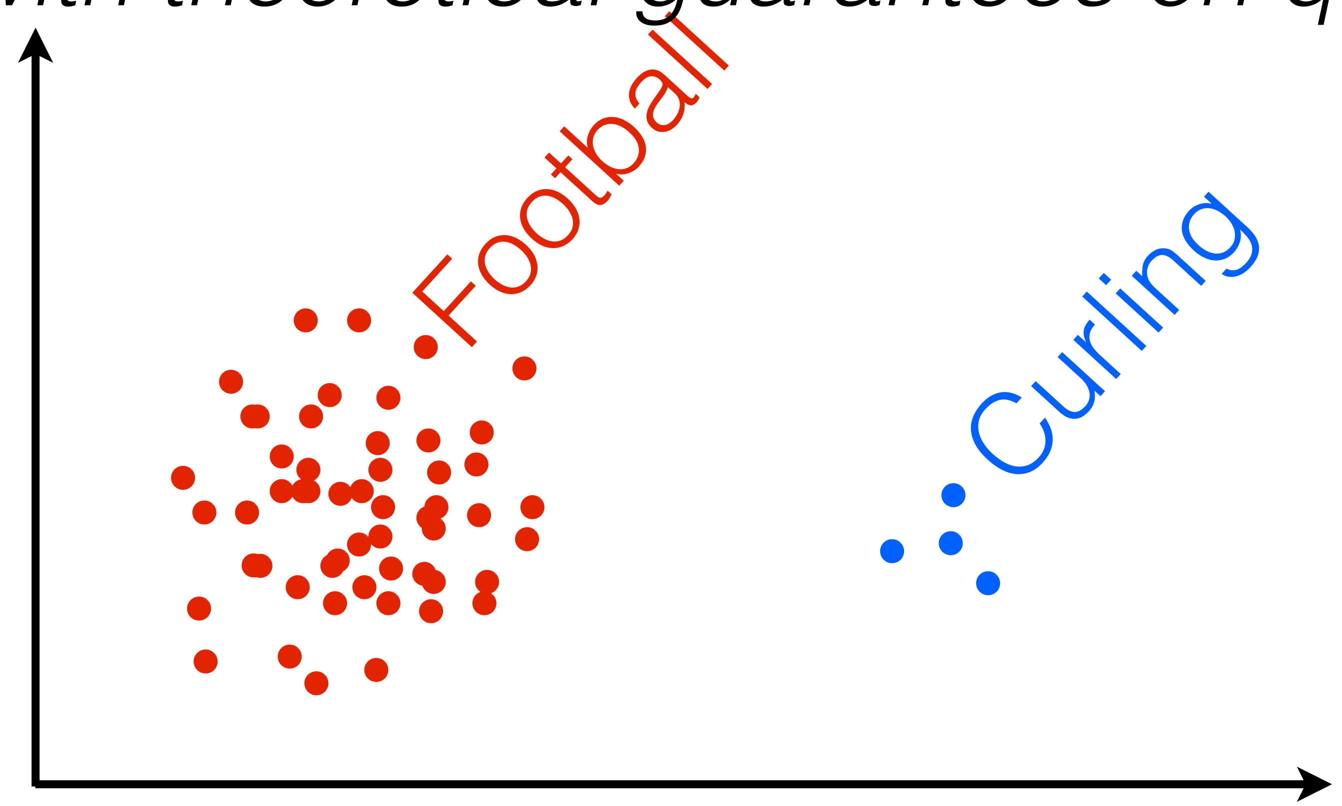


- Previous heuristics: data squashing, inducing points
- Cf. subsampling

[Agarwal et al 2005; Feldman & Langberg 2011; DuMouchel et al 1999; Madigan et al 1999]

Data summarization

- Observe: redundancies can exist even if data isn't "tall"
- Coresets: pre-process data to get a smaller, weighted data set *with theoretical guarantees on quality*

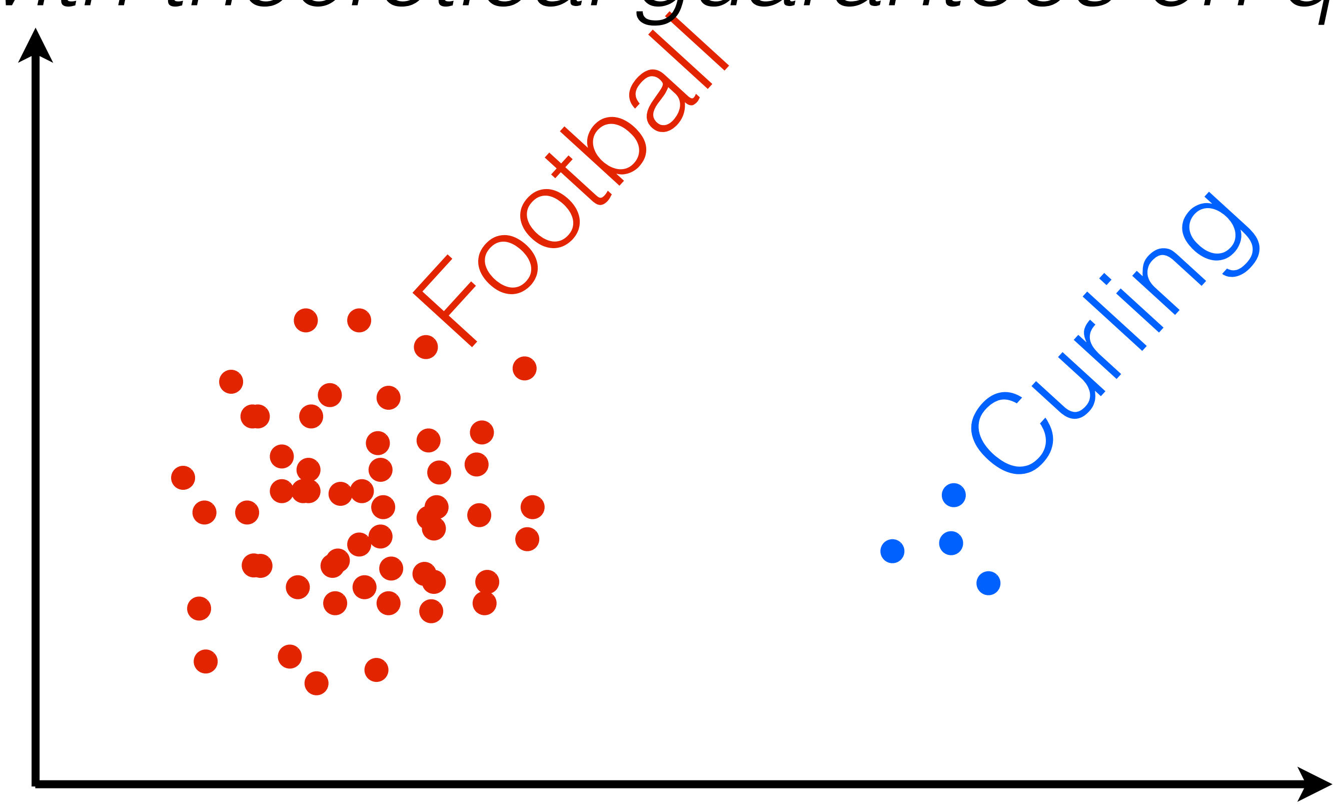


- Previous heuristics: data squashing, inducing points
- Cf. subsampling
- Coresets for Bayes

[Agarwal et al 2005; Feldman & Langberg 2011; DuMouchel et al 1999; Madigan et al 1999; Huggins, Campbell, Broderick 2016; Campbell, Broderick 2017; Campbell, Broderick 2018]

Data summarization

- Observe: redundancies can exist even if data isn't "tall"
- Coresets: pre-process data to get a smaller, weighted data set *with theoretical guarantees on quality*



- Previous heuristics: data squashing, inducing points
- Cf. subsampling
- Coresets for Bayes
- Next up: distributed computing, beyond Bayes

[Agarwal et al 2005; Feldman & Langberg 2011; DuMouchel et al 1999; Madigan et al 1999; Huggins, Campbell, Broderick 2016; Campbell, Broderick 2017; Campbell, Broderick 2018]

Gaussian processes

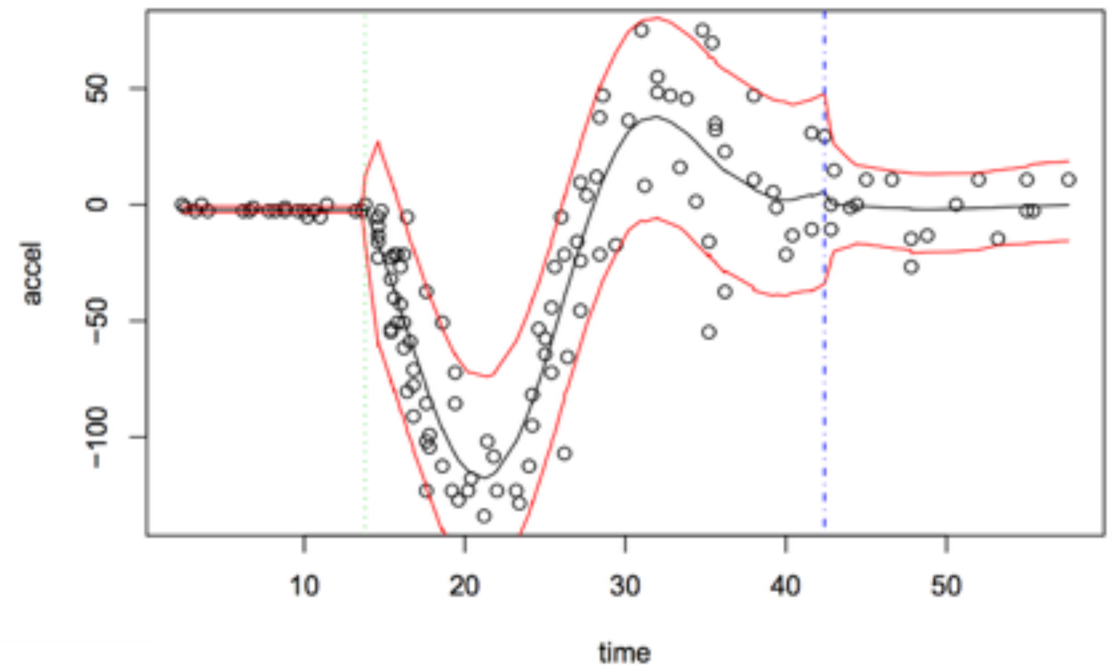
- Fuel consumption

[Julian Hertzog 2016]

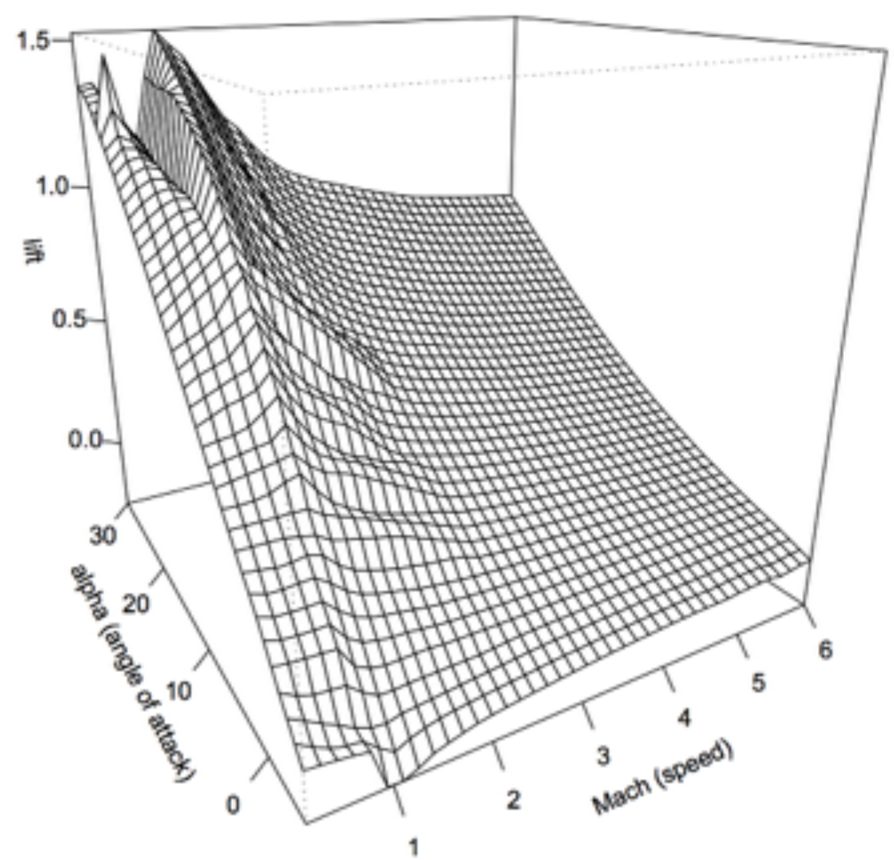
Gaussian processes

- Fuel consumption

- Motorcycle acceleration



- NASA rocket boosters

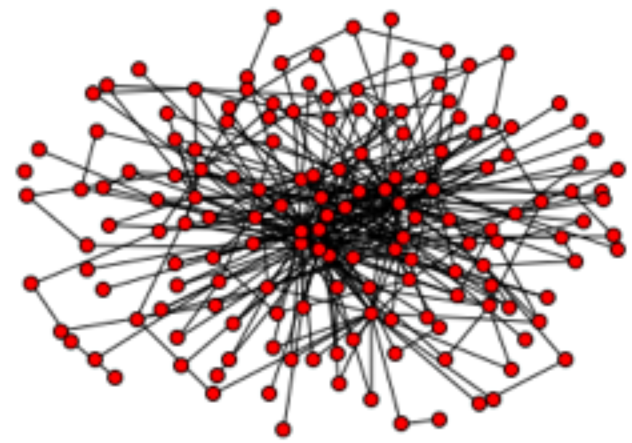


[Julian Hertzog 2016; Gramacy, Lee 2009]

Nonparametric Bayes

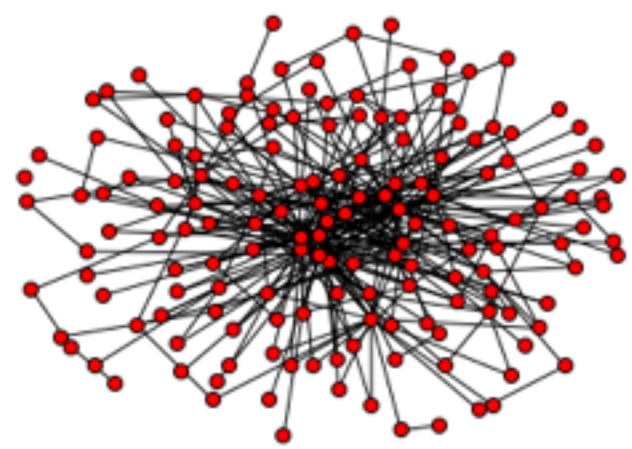
- Can adapt automatically to learn *more from data* as we accrue more data

WIKIPEDIA
The Free Encyclopedia

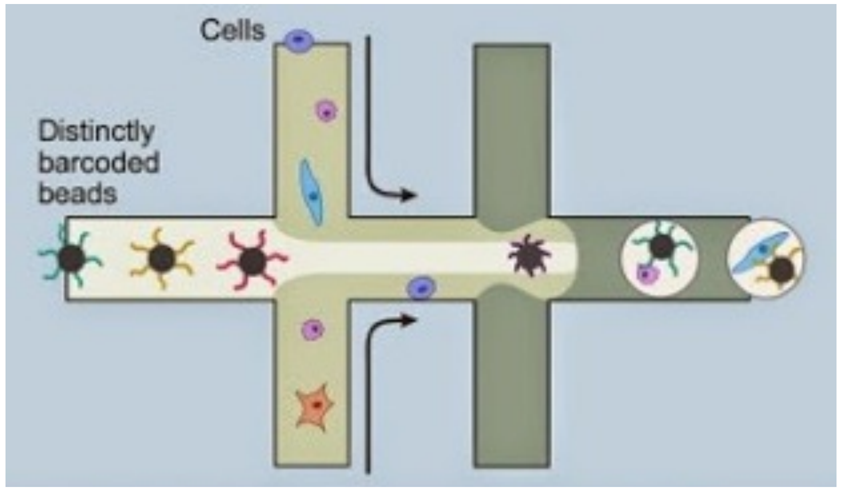


Nonparametric Bayes

- Can adapt automatically to learn *more from data* as we accrue more data




- Beyond clustering; Complex structures in data
    - Latent trees / Single-cell RNA sequencing



    - Latent probabilistic graphical models / gene regulatory networks

[Shiffman, Stephenson, Schiebinger, Campbell, Huggins, Regev Broderick 2017; Agrawal, Broderick, Uhler 2018]

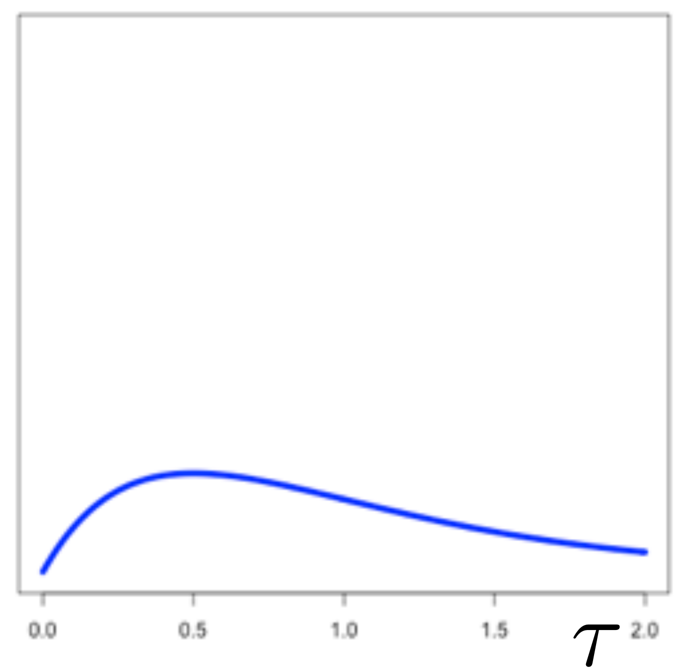
Evaluation / Quantifying robustness

Evaluation / Quantifying robustness

- How sensitive is our output to our assumptions?

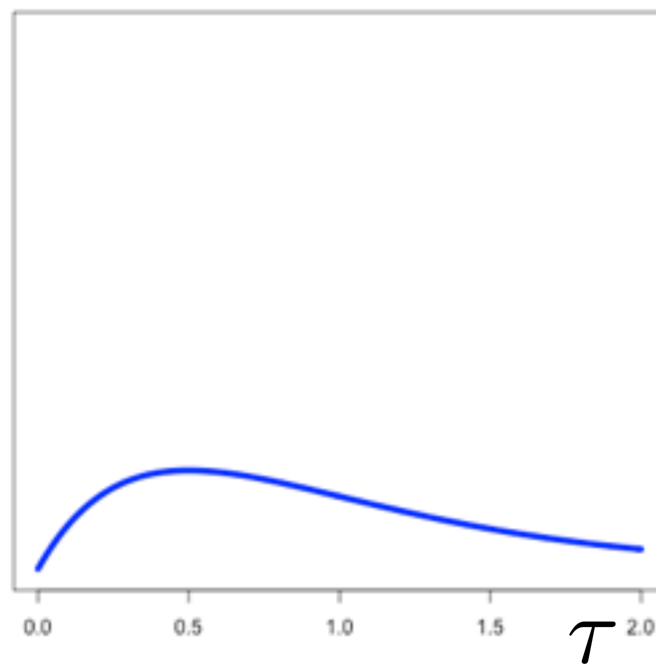
Evaluation / Quantifying robustness

- How sensitive is our output to our assumptions?



Evaluation / Quantifying robustness

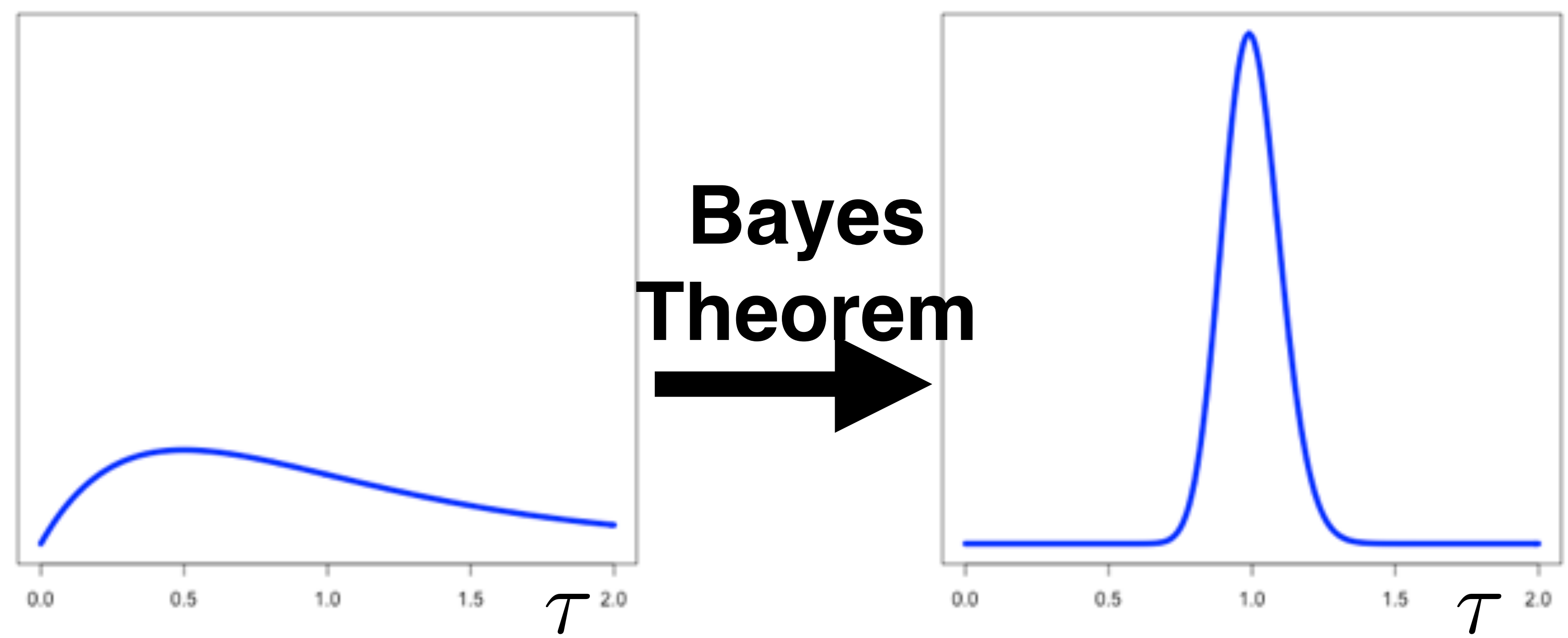
- How sensitive is our output to our assumptions?



Bayes Theorem

Evaluation / Quantifying robustness

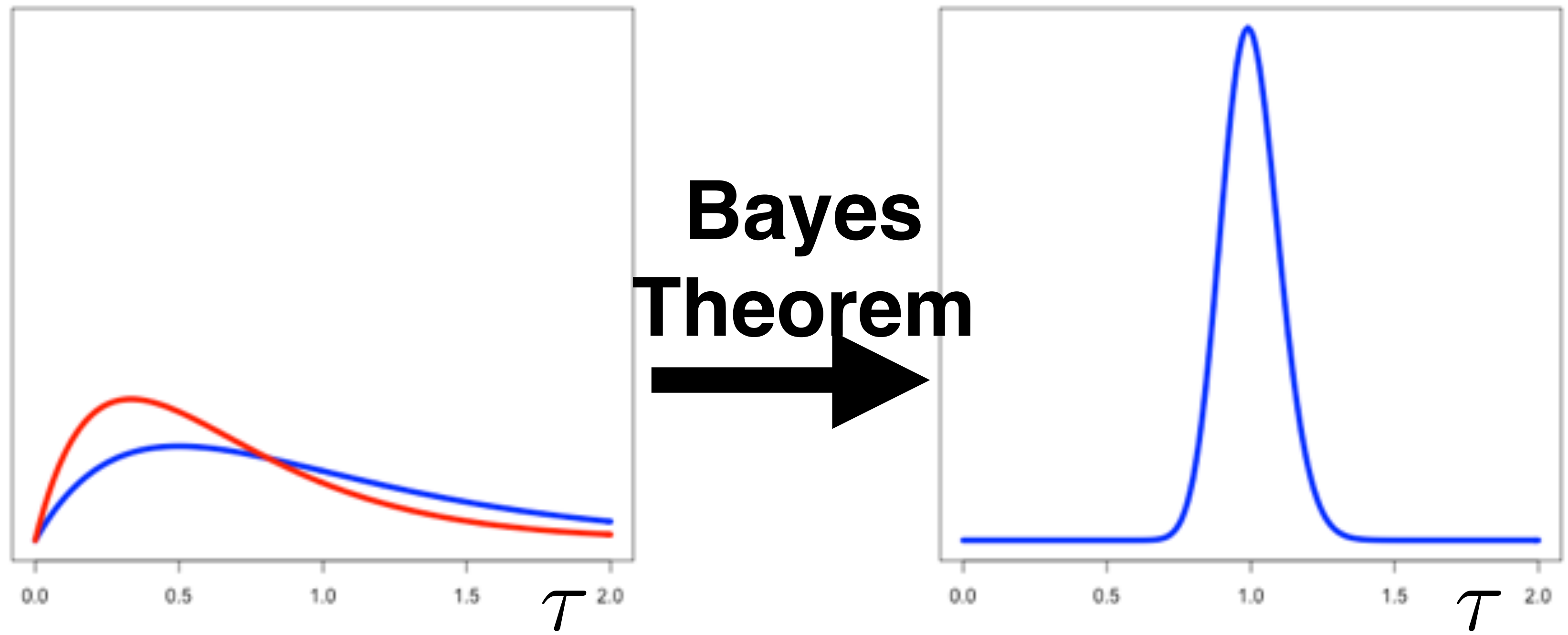
- How sensitive is our output to our assumptions?



Evaluation / Quantifying robustness

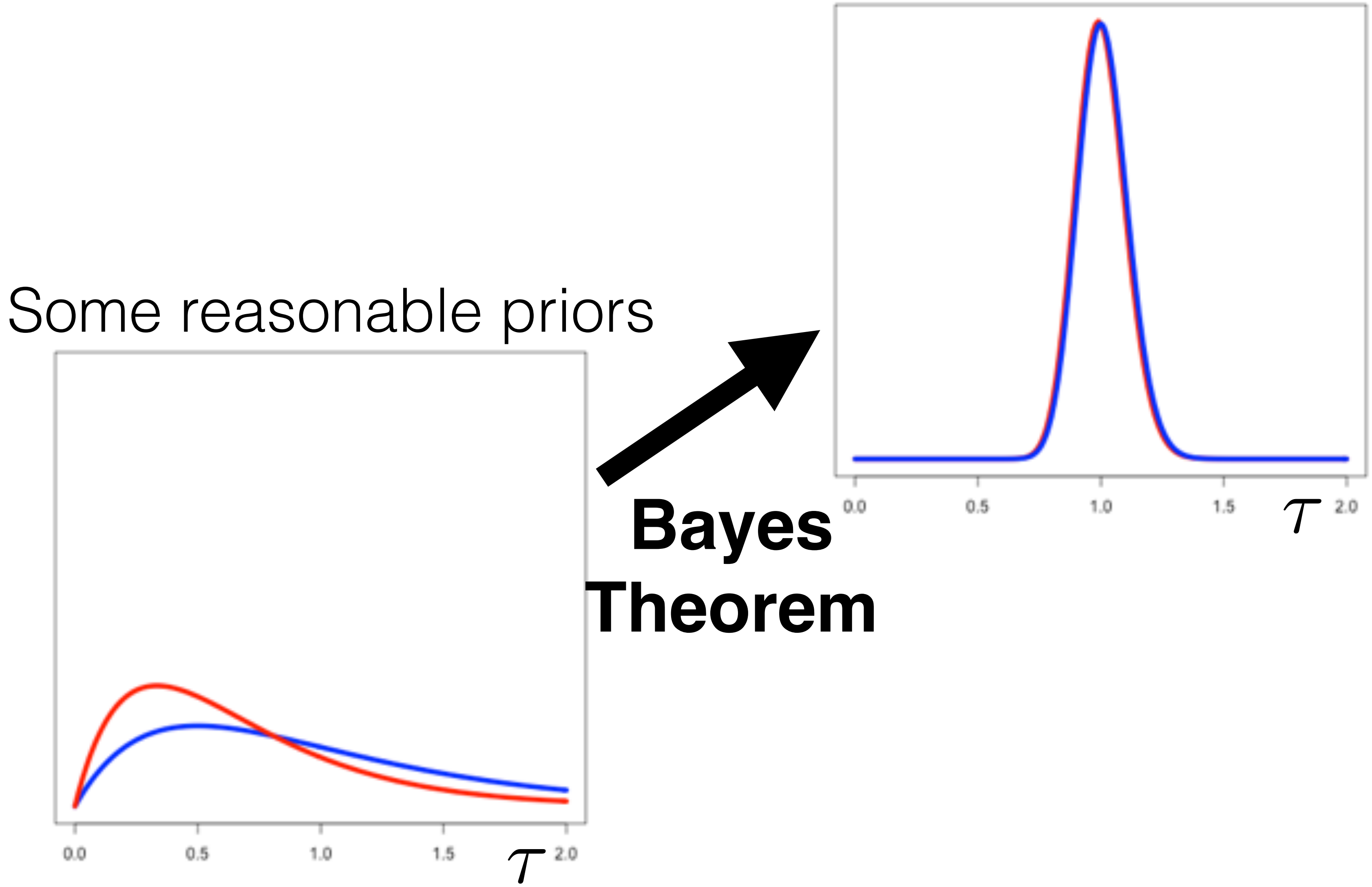
- How sensitive is our output to our assumptions?

Some reasonable priors



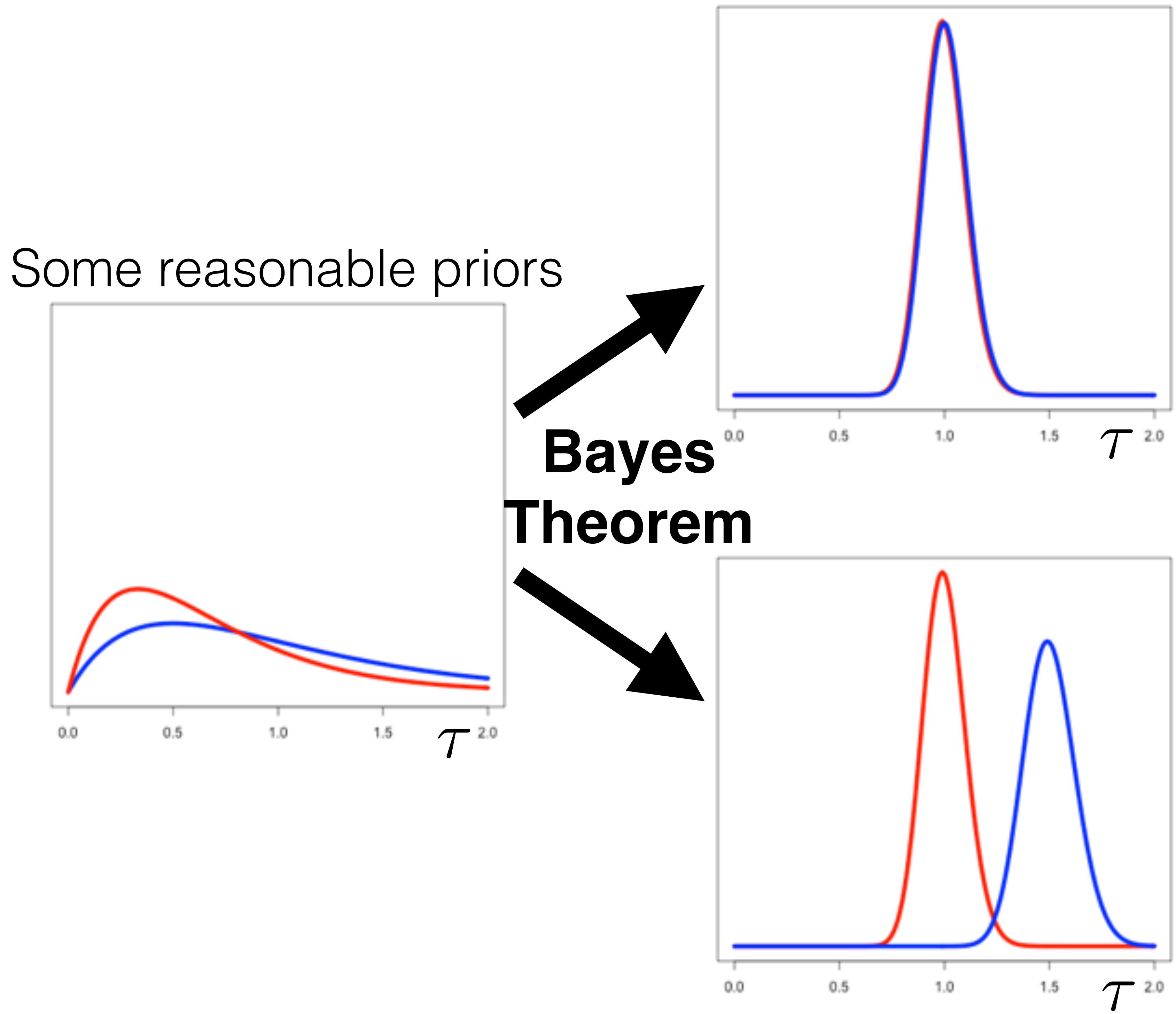
Evaluation / Quantifying robustness

- How sensitive is our output to our assumptions?



Evaluation / Quantifying robustness

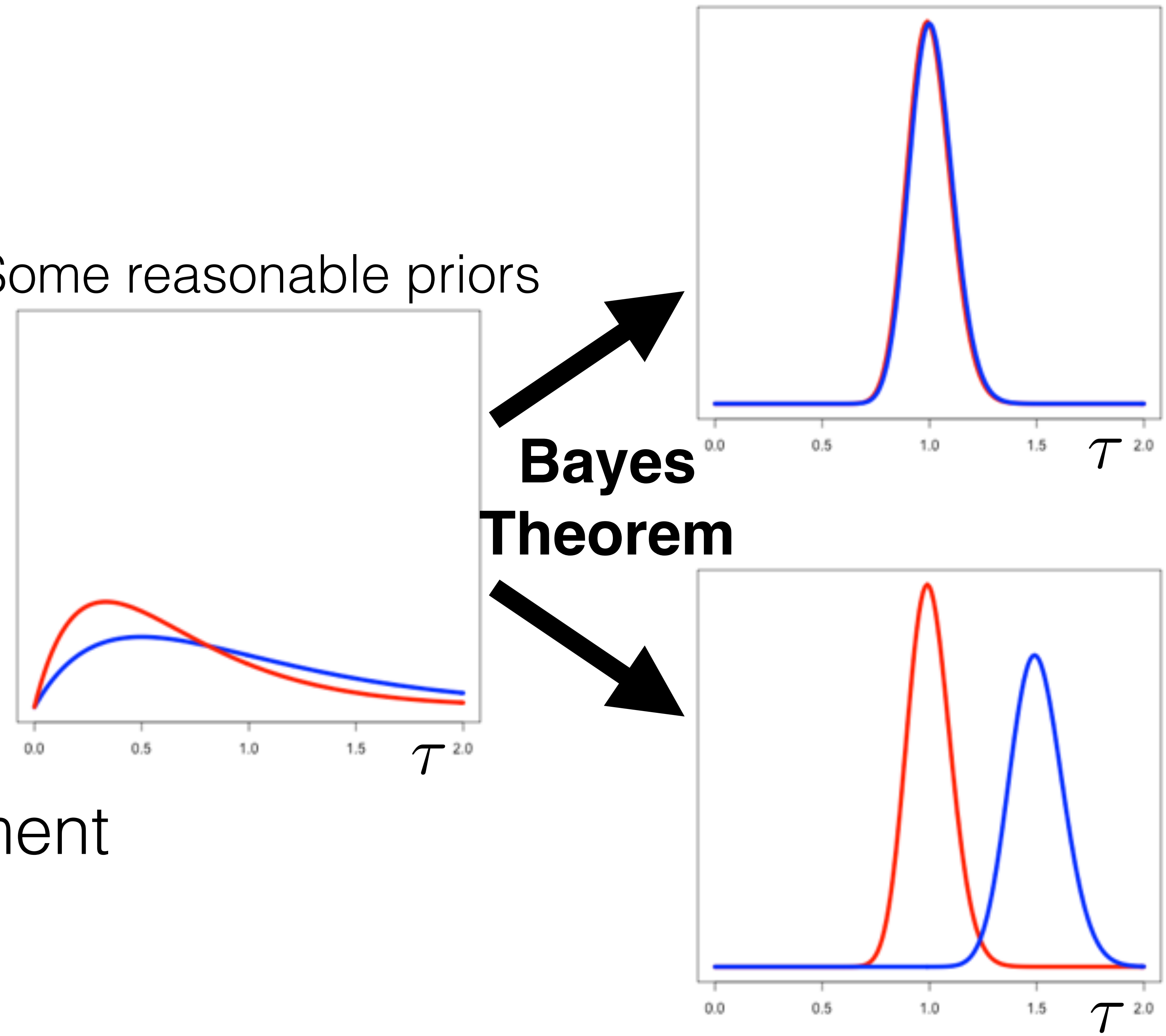
- How sensitive is our output to our assumptions?



Evaluation / Quantifying robustness

- How sensitive is our output to our assumptions?

Some reasonable priors



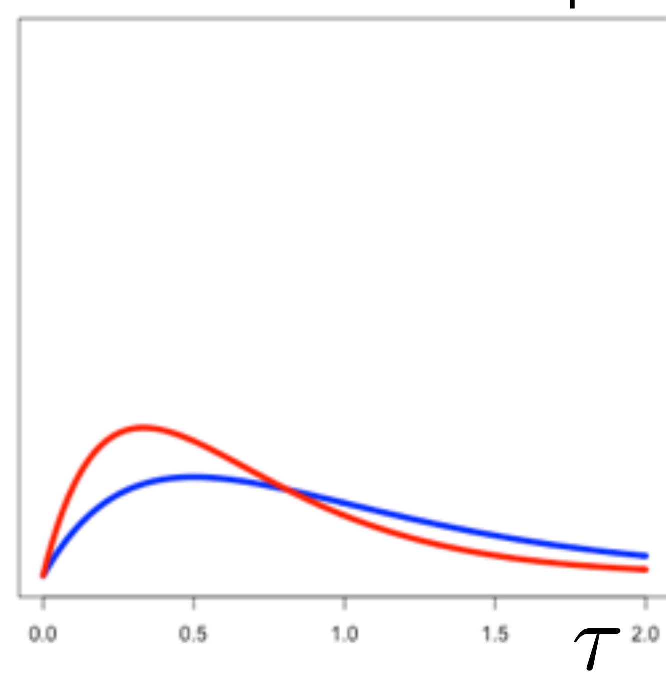
- A perturbation argument

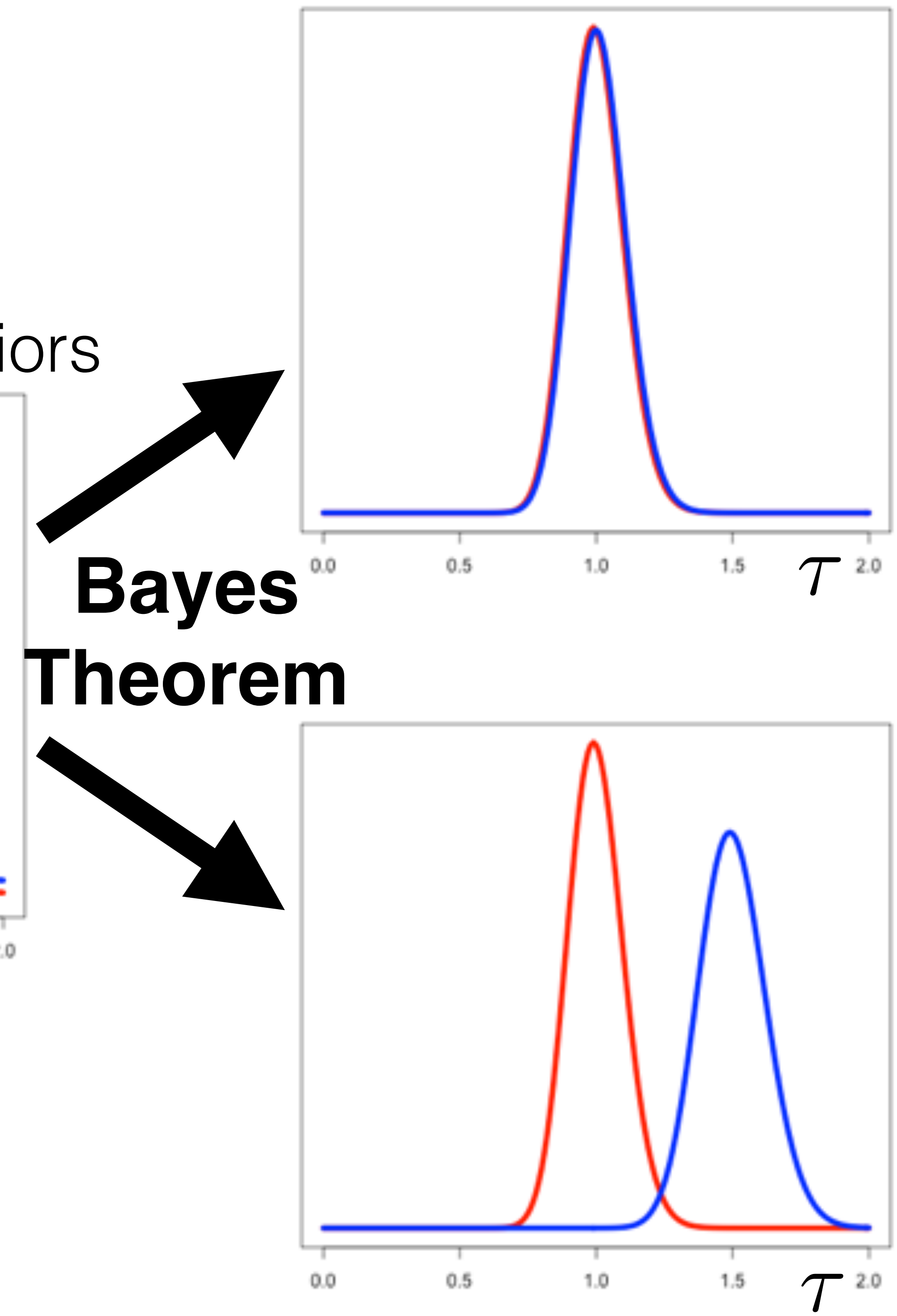
[Giordano, Broderick, Jordan 2015; Giordano, Broderick, Jordan 2017]

Evaluation / Quantifying robustness

- How sensitive is our output to our assumptions?

Some reasonable priors





- A perturbation argument

- Fast cross validation and bootstrap

[Giordano, Broderick, Jordan 2015; Giordano, Broderick, Jordan 2017]

Bayesian Machine Learning

Quantifying uncertainty and robustness at scale

Tamara Broderick
ITT Career Development Assistant Professor
tbroderick@mit.edu


Raj Agrawal


Trevor Campbell


Lorenzo Masoero


Will Stephenson

(Partial) References

R Agrawal, T Broderick, and C Uhler. Minimal I-MAP MCMC for scalable structure discovery in causal DAG models, 2018. ArXiv:1803.05554.

T Campbell and T Broderick. Bayesian coreset construction via Greedy Iterative Geodesic Ascent, 2018. ArXiv:1802.01737.

T Campbell and T Broderick. Automated scalable Bayesian inference via Hilbert coresets, 2017. ArXiv:1710.05053.

R Giordano, T Broderick, and MI Jordan. Linear response methods for accurate covariance estimates from mean field variational Bayes. *NIPS* 2015.

R Giordano, T Broderick, and MI Jordan. Covariances, robustness, and variational Bayes. ArXiv:1709.02536.

JH Huggins, T Campbell, and T Broderick. Coresets for scalable Bayesian logistic regression. *NIPS 2016*.

JH Huggins, RP Adams, and T Broderick. PASS-GLM: Polynomial approximate sufficient statistics for scalable Bayesian GLM inference. *NIPS 2017*.

M Shiffman, W Stephenson, G Schiebinger, T Campbell, J Huggins, A Regev, T Broderick. Probabilistic reconstruction of cellular differentiation trees from single-cell RNA-seq data. *NIPS Workshop on Machine Learning in Computational Biology* 2017.