Introduction to BART Bootcamp

Purushottam (Prakash) Laud
Medical College of Wisconsin, Division of Biostatistics

with rest of the day presented by

Rodney Sparapani, Brent Logan &
Medical College of Wisconsin
Division of Biostatistics

Rob McCulloch
Arizona State Universiy
School of Mathematical &
Statistical Sciences

Beer, Brats & Biostats BART Bootcamp, MCW, OktoberFest 2017

Objectives and Outline

Objectives

- Warm-up act for the rest of the day!
- Get acquainted with conceptual differences between traditional and modern regression methods
- Understand implications of parameter-free regression

Outline

- Quick look at traditional regression
- Regression with a single tree
- Ensembles of trees: Bagging, Boosting, Random Forests and BART
- Example(s)

Today's Agenda

You already know it!

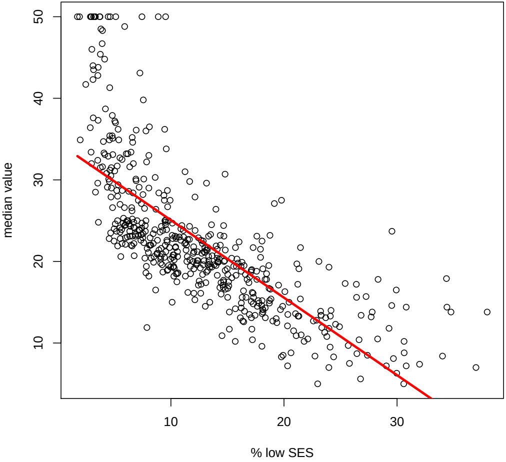
- Dr. McCulloch: BART, what it is, how it works, how to use it, its advantages
- Dr. Sparapani: Models for binary data, survival data
- Dr. Logan: Individualized treatment rules

Boston Housing Data

Median value of houses and neighborhood characteristics

- Dependent (outcome) variable: median value (medv)
- Predictor variable: percent of population with low SES (lstat)

Scatter diagram and linear fit

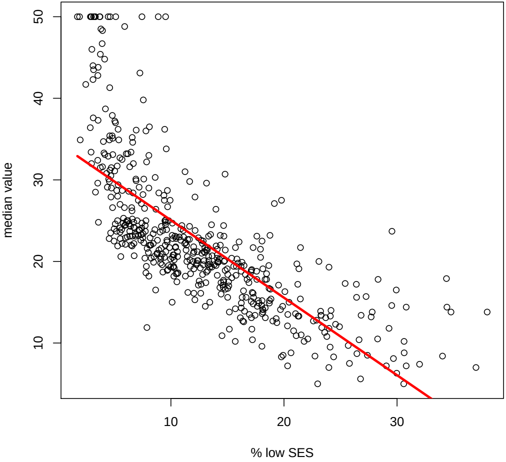


Boston Housing Data

Median value of houses and neighborhood characteristics

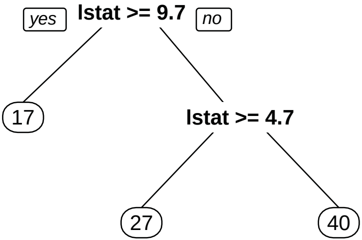
- Dependent (outcome) variable: median value (medv)
- Predictor variable: percent of population with low SES (lstat)

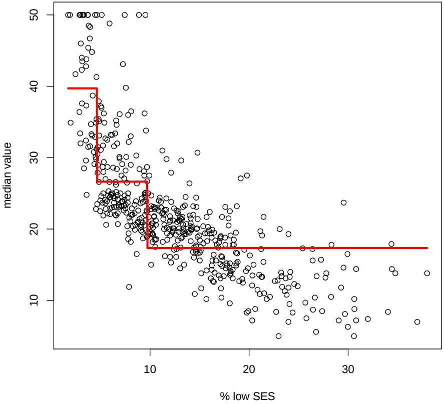
Scatter diagram and linear fit



Carl Friedrich Gauss
Pierre-Simon Laplace
About 220 years ago

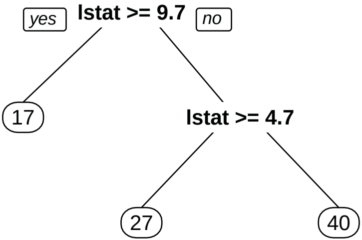
Single small tree





Breaking up predictor axis into pieces
Fit constant values of (mean) outcome on these pieces

Single small tree



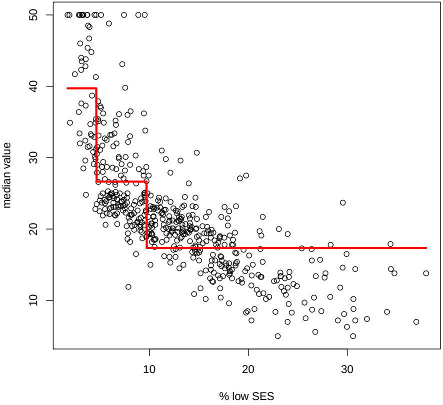
yes **lstat >= 9.7** no

17

lstat >= 4.7

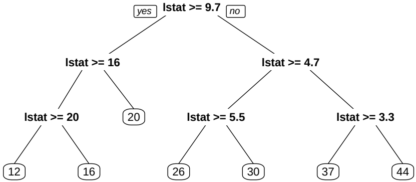
27

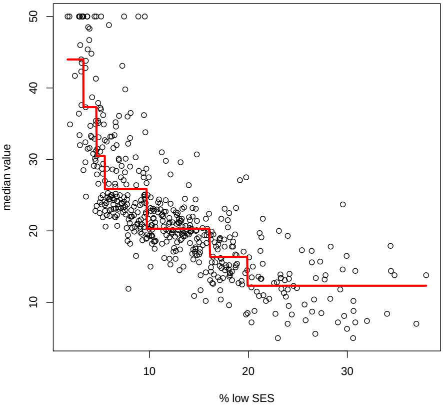
40



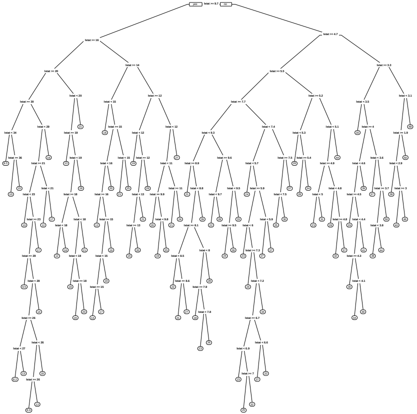
Breaking up predictor axis into pieces
Fit constant values of (mean) outcome on these pieces
Leo Breiman   About 1980

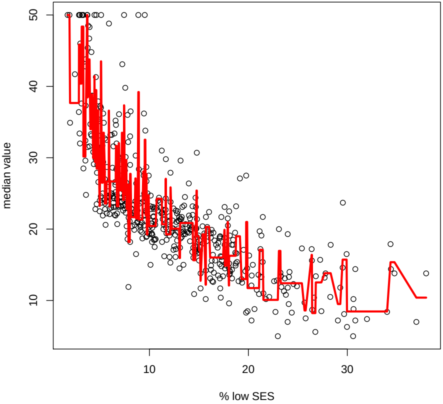
Single medium tree



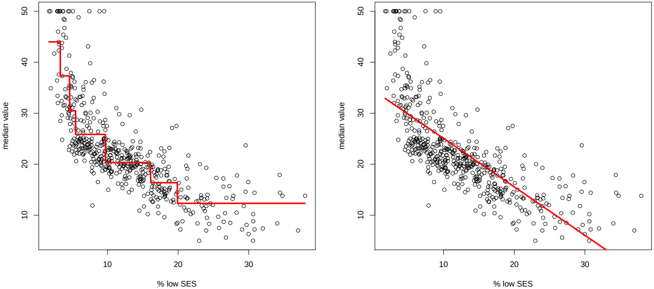


Single large tree



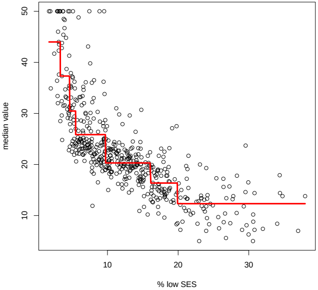


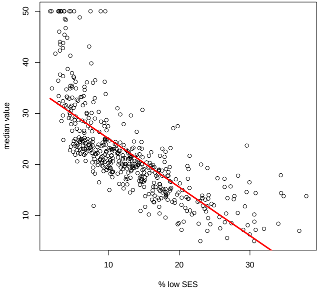
Back to 7-node tree: interpretation?



- Linear regression does not fit well but we can describe the fit: Median value drops by 6 units for every 5% increase in persons of low SES.
- If we had transformed to log(median value), we could say something like: for every 5% increase in persons of low SES median value drops by x percent.

Back to 7-node tree: interpretation?

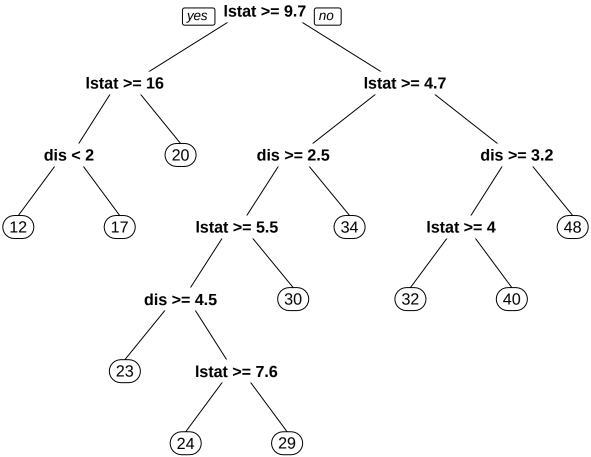




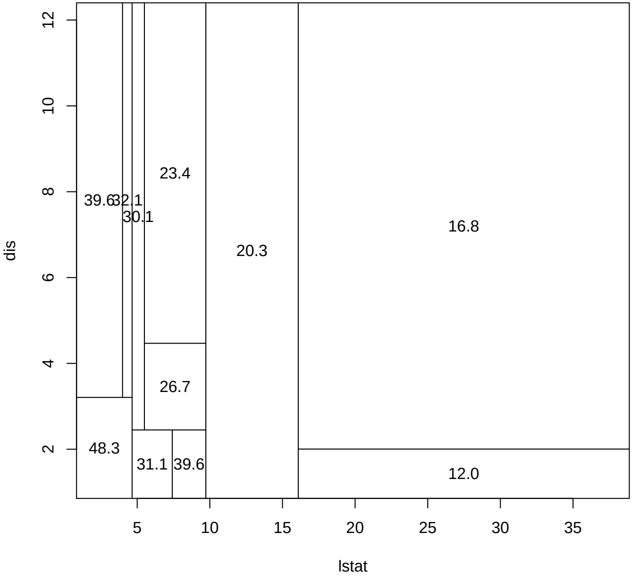
- Linear regression does not fit well but we can describe the fit: Median value drops by 6 units for every 5% increase in persons of low SES.
- If we had transformed to log(median value), we could say something like: for every 5% increase in persons of low SES median value drops by x percent.
- For tree regression we do not need to hunt for the right transformation, if even such exists. But simple interpretation is a challenge. One can describe tree fit by line segments.
- For example, additional increase in percent low SES beyond 20% does not affect median value. Fairly constant slope between about 7 and 20 percent.

More than one predictor

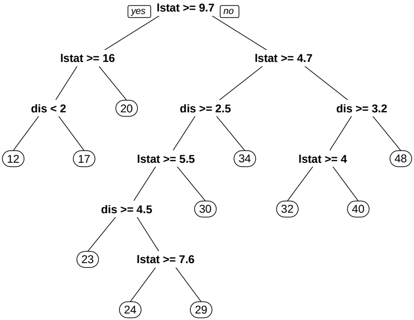
Boston data: weighted mean distance from 5 major employment centers

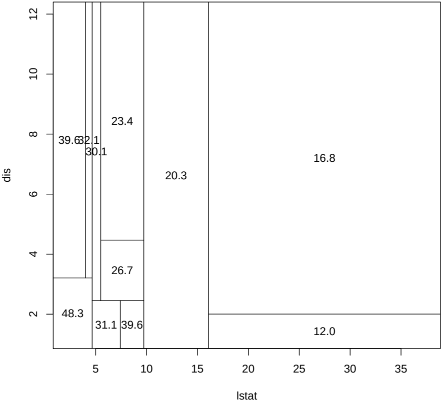


Bivariate plane cut into pieces

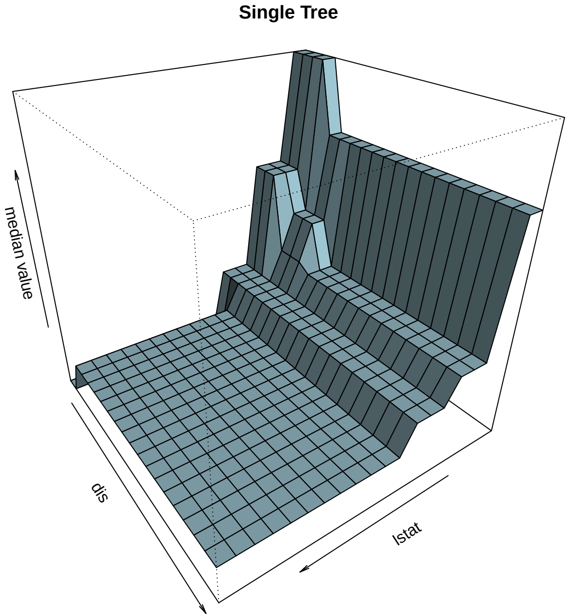


Tree and partition





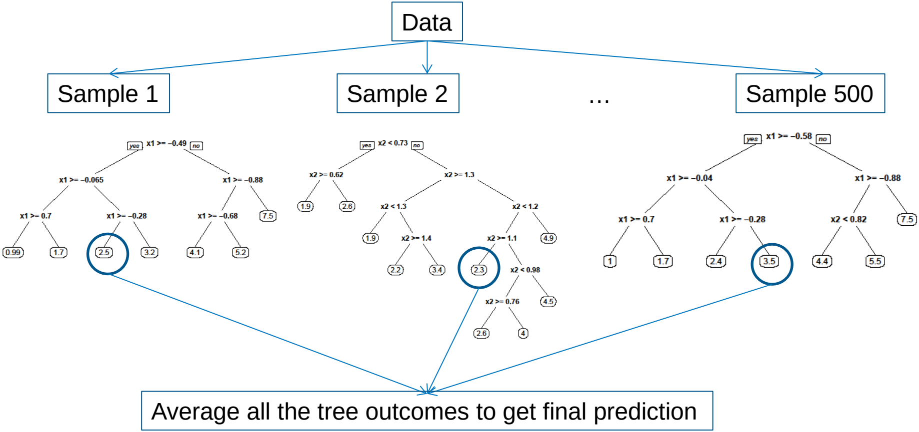
3-d fit



From single tree to ensemble of trees

- Single trees lead to jagged fits, spiky functions
- Use multiple trees, then average or add after scaling
- How to choose these multiple trees?
- Four main methods:
  - Bagging (**B**ootsrap **ag**gregate)
    Leo Breiman About 1995
  - Boosting
    Schapire 1990
  - Random Forests (Bagging plus random predictor selection)
    Leo Breiman About 2000
  - BART (Bayesian additive regression trees)
    Chipman, George, McCulloch 2010

Bagging and Random Forests



Data

Sample 1

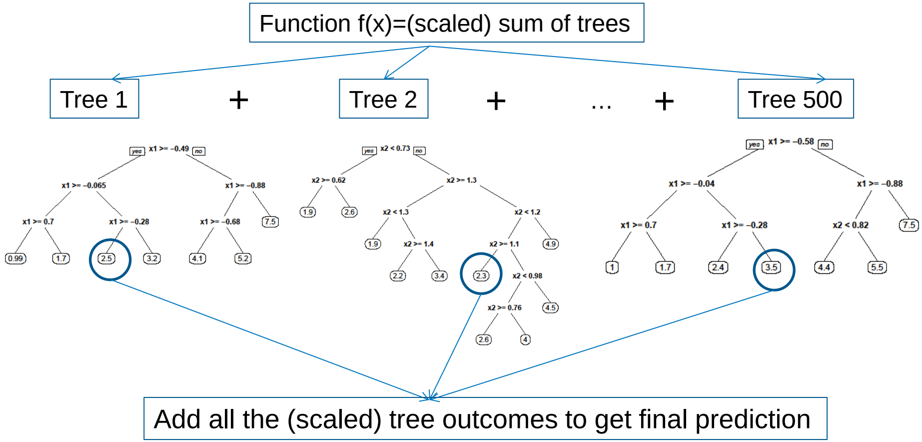
Sample 2

…

Sample 500

Average all the tree outcomes to get final prediction

Boosting and BART

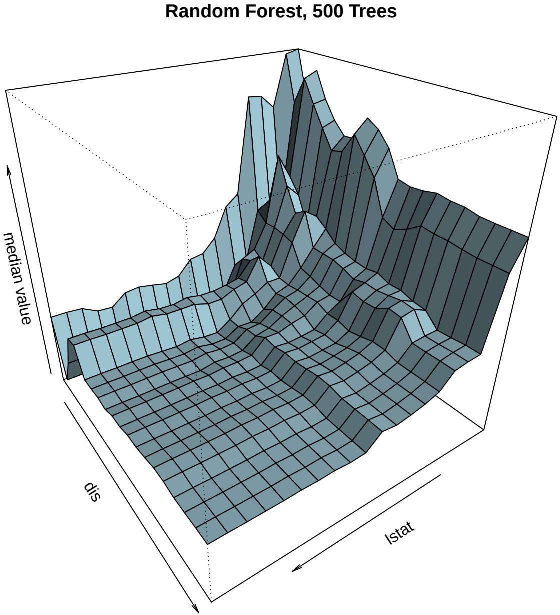


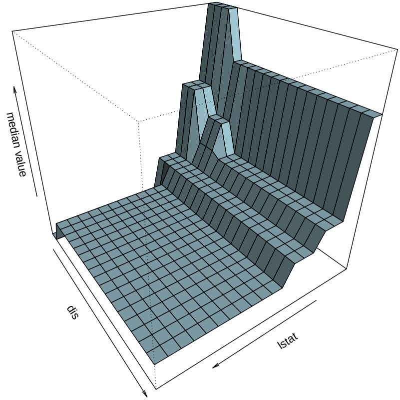
Function f(x)=(scaled) sum of trees

Tree 1 + Tree 2 + … + Tree 500

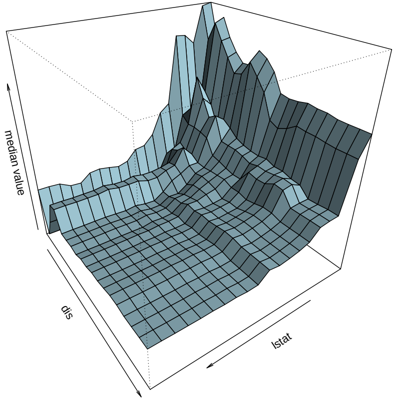
Add all the (scaled) tree outcomes to get final prediction

Random Forest fit to Boston housing data





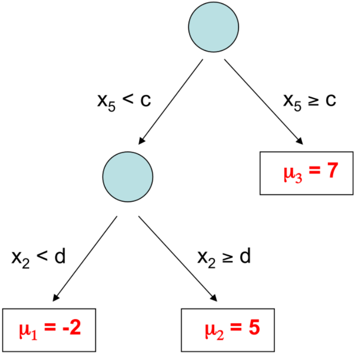
Single Tree



Random Forest, 500 Trees

Regression Tree Model

T denotes the tree structure and branch decision rules
$M = \{\mu_1, \mu_2, \ldots, \mu_b\}$ denotes the terminal nodes
$g(x_i; T, M)$ is a regression tree function

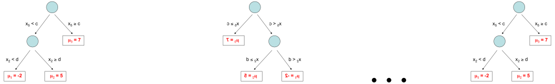


$$y_i = g(x_i; T, M) + \epsilon_i$$

BART: Bayesian Additive Regression Trees

Chipman, George & McCulloch (CGM10);
2010 *AnnApplStatist*

$$y_i = g(x_i; T_1, M_1) + g(x_i; T_2, M_2) + \cdots + g(x_i; T_m, M_m) + \epsilon_i$$



where m is somewhat large (50 or 200 or 1000).

$$\begin{aligned} f(x_i) &= \sum_j g(x_i; T_j, M_j) \\ f &\sim \text{BART} \end{aligned}$$

BART Prior

Prior 1: Distribution for tree depth
Prior 2: Distribution for covariate selection
Prior 3: Distribution for split point
Prior 4: Distribution for value at terminal node
Prior 5: Distribution for ϵ

Details in CGM10; “many small trees” principle

Default prior settings					
Number of terminal nodes	1	2	3	4	5+
Prior probability	0.05	0.55	0.28	0.09	0.03

LVH (Left Vetricular Hypertrophy) Example

- Diagnosing LVH using patient chararacteristics and ECG data
- Gold standard based on LVM index - ratio of observed LVM from MRI and body surface area
- Multi-Ethnic Study of Atheroclerosis (MESA), 2000-2013, of 4953 subjects, 7855 MRI's with ECG
- Two randomly split groups

Visit	Group 1	Group 2	Total
Baseline	2501	2452	4953
10 years	1430	1472	2902
Total	3931	3924	7855

Predicting LVH with ECG with MESA criteria

Jain et.al., American Heart Journal, 2010

LVH by ECG	LVH by MRI				
MESA	Yes	No	Total	Rates	
Yes	29	870	899	True Pos.	3.2%
No	18	6938	6956	True Neg.	99.7%
Total	47	7808	7855		

	MESA		
	Sens.	Spec.	
	61.7%	88.9%	
Predictive	Pos.	Neg.	
Value	43.0%	94.4%	

LVH Prevalence $\approx$ 12%

Schirmer, Lunde and Rasmussen. European Heart Journal 1999

Predicting LVM Index with ECG data using BART

- 193 covariates: $R^2 = 0.533$ (0.443) in-sample (out-of-sample)
- Demographics: age and gender
- Body size: height, weight and body mass index (BMI)
- ECG global: PR interval, P axis, QRS interval and axis
- ECG lead measurements: 46 variables X 4 leads

- 8 covariates: $R^2 = 0.464$ (0.408) in-sample (out-of-sample)
- Demographics: gender
- Body size: BMI
- ECG global: QRS interval
- ECG wave amplitude: T (V2 and I), STJ (V2), S (V2), R (V5)

Predicting LVH with ECG data using BART

LVH by ECG	LVH by MRI				
BART	Yes	No	Total	Rates	
Yes	31	197	228	True Pos.	13.6%
No	16	7611	7627	True Neg.	99.8%
Total	47	7808	7855		

	MESA		BART	
	Sens.	Spec.	Sens.	Spec.
	61.7%	88.9%	66.0%	97.5%
Predictive	Pos.	Neg.	Pos.	Neg.
Value	43.0%	94.4%	78.3%	95.5%

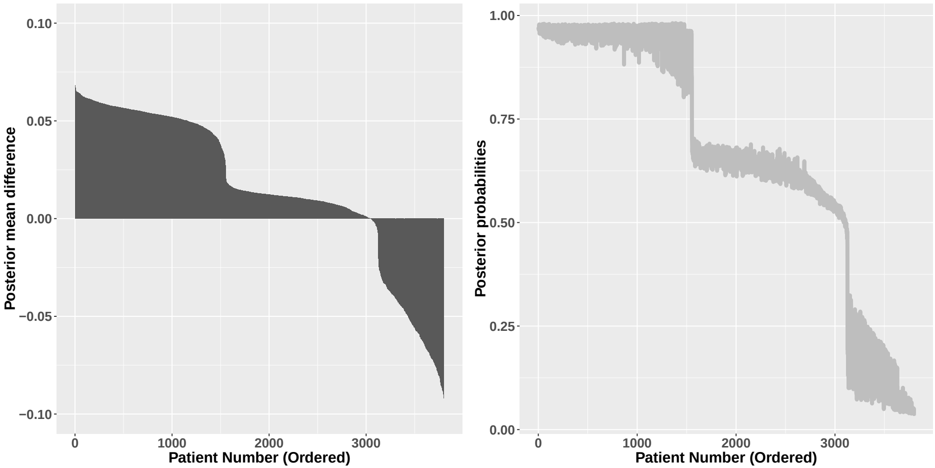
Brief preview of individualized treatment selection

- Two treatments: Flu/Bu and Flu/Mel
- Outcome: 1-year survival
- Patient, donor and disease factors: Age, Race/ethnicity, KPS, CMV status, disease, remission status, disease subtypes, chemosensitivity, interval from dx to tx, donor type, HLA matching, comorbidity score, prior autotx, gender matching, year of tx
- 3802 patients receiving Flu/Bu or Flu/Mel RIC regimens between 2011-2013

What can BART produce?

Dr. Logan will elaborate on the example and ITRs and related inference. Here is a quick preview of things to look for:

ITR Example: FluMel minus FluBu



Enjoy the rest of the day with BART!

laud@mcw.edu