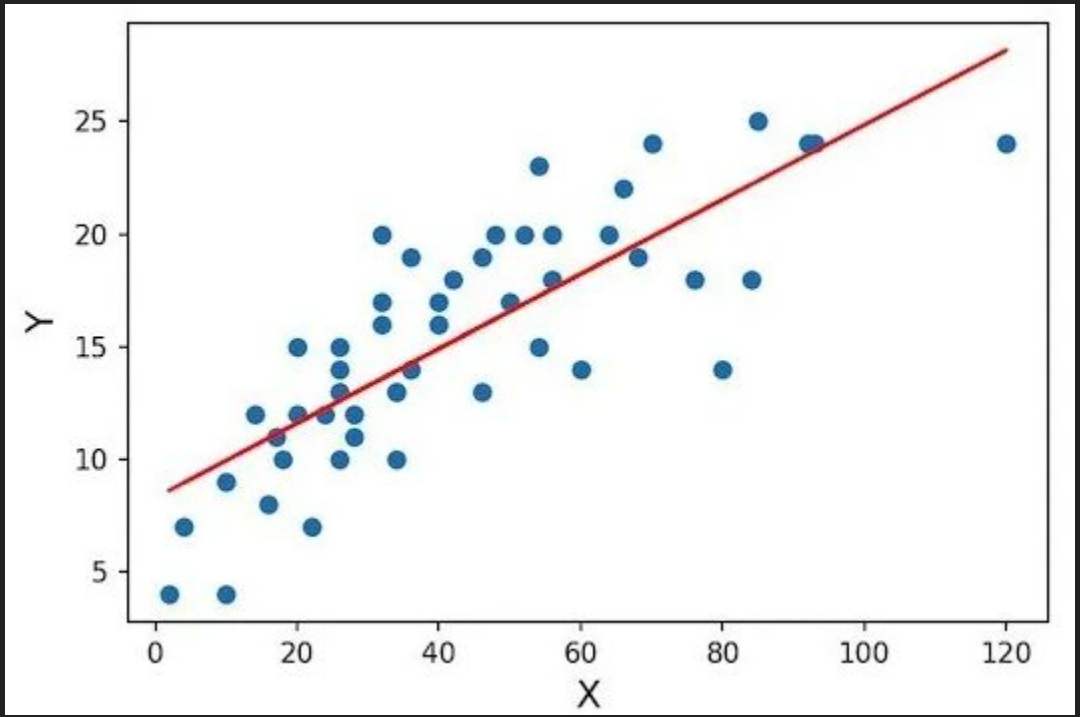
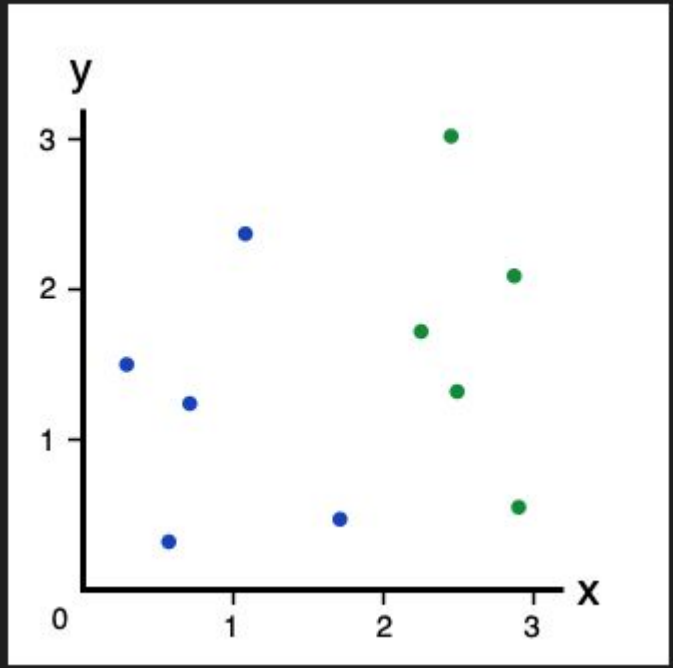
An Introduction to Random Forest Models

Created for the Methods Workshop
Spring 2022
Ojashwi Pathak

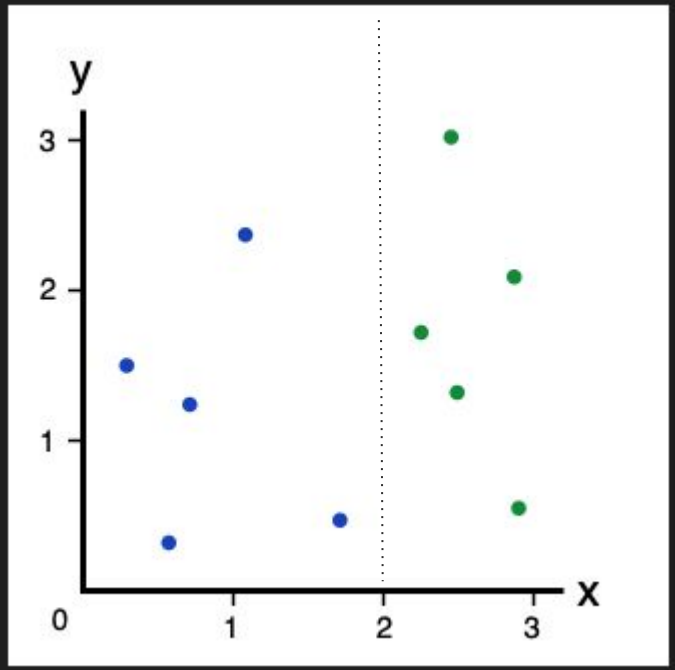
What is the intuition?



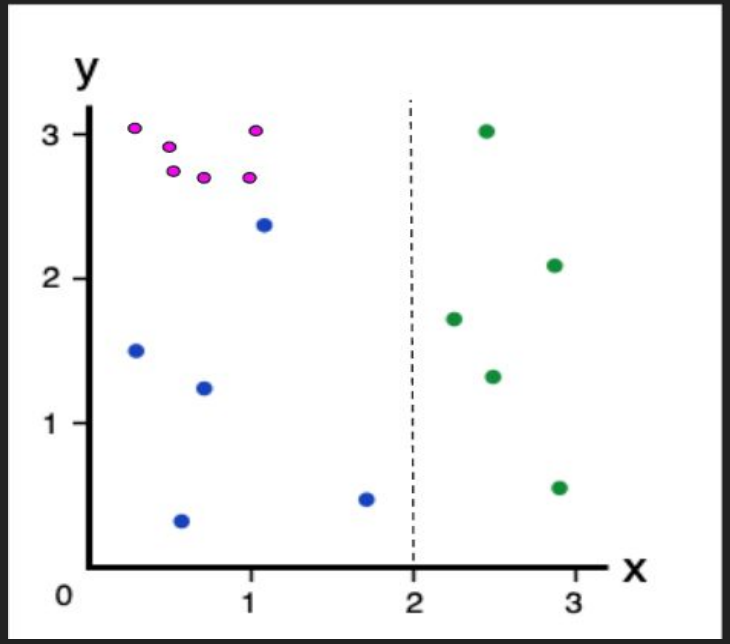
What is the intuition?

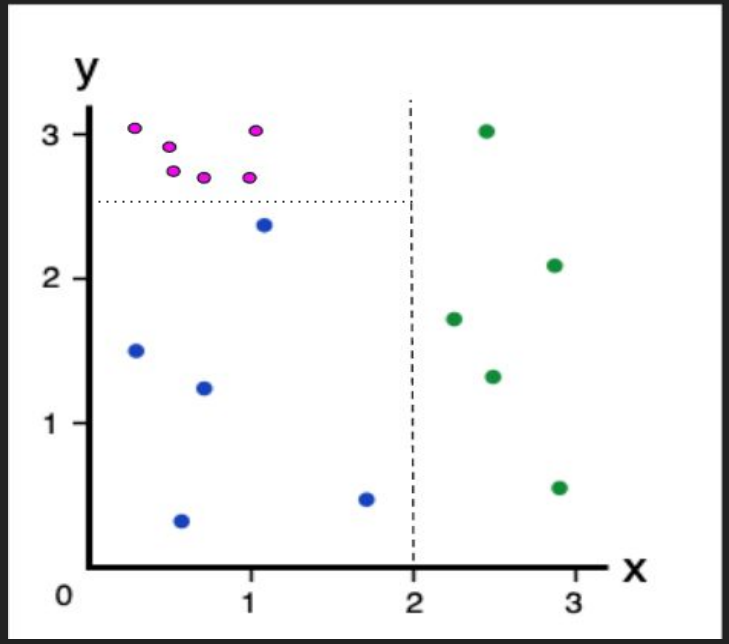


What is the intuition?

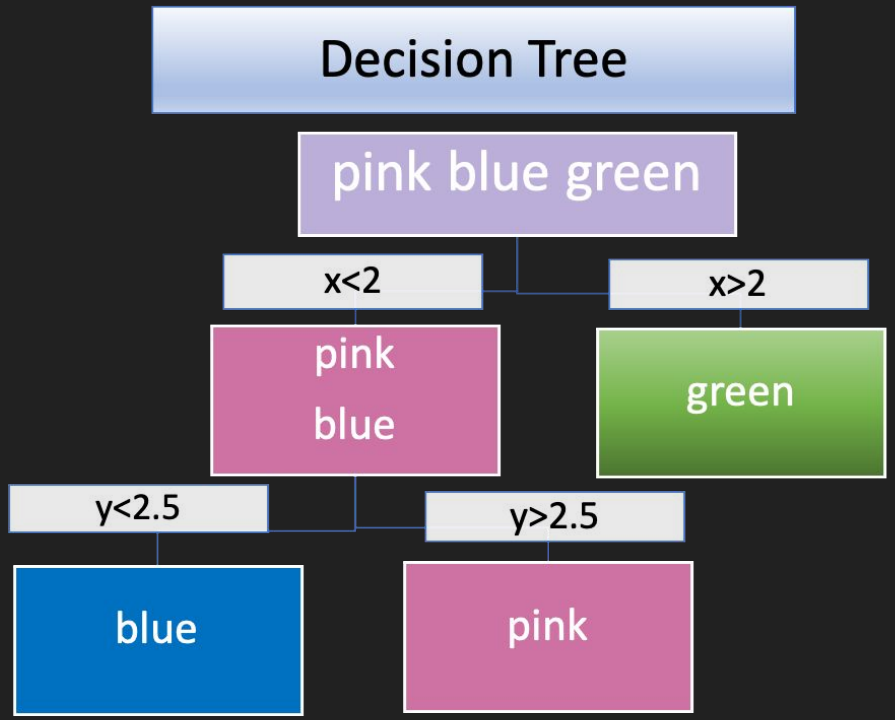


What happens if there are more than two colors now?

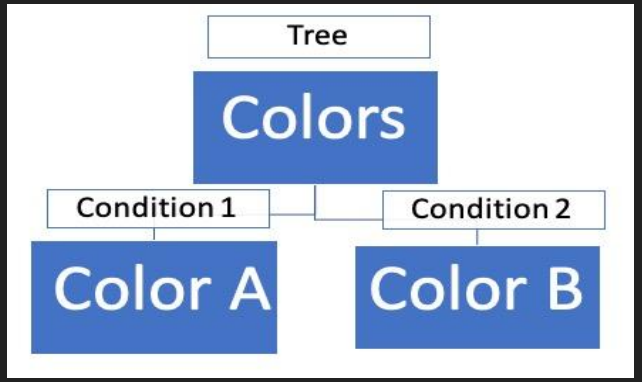


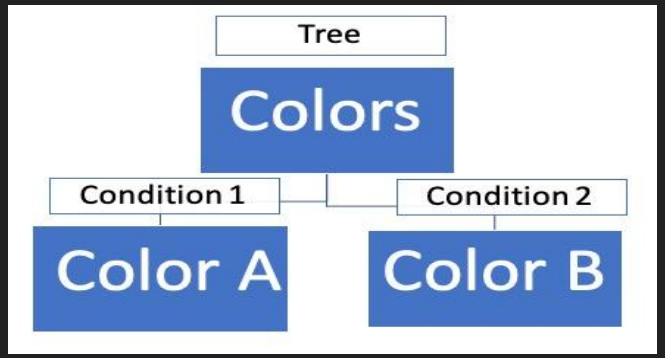


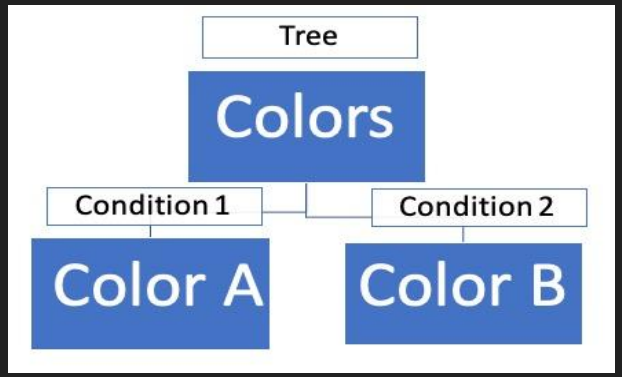
Let’s see that in a decision tree



What about a random forest?



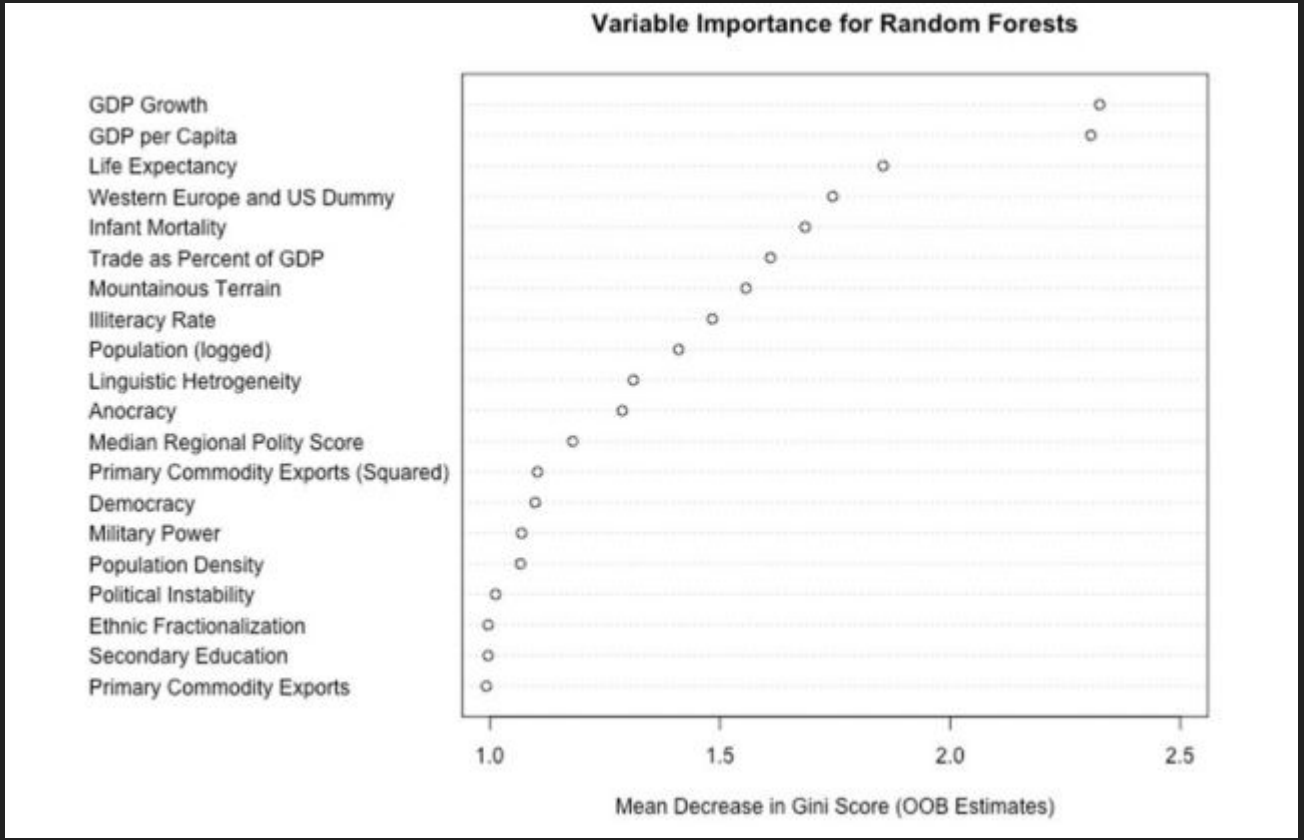




So What? How can I use it in Political Science?

Example: Muchlinski et. al 2017 (Political Analysis)

- *Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data*
  - Does our theory make correct predictions?
  - Causality v. Predictive Power
  - DV: civil war onset; IV: 88 predictors (Civil War Data)
  - Compare RF against Collier and Hoeffler, Fearon and Laitin, and Hegre and Sambanis
  - How can we tell which predictor is the best?
    - Gini Score (0-1)
    - Variables with greater mean decrease



Variable Importance Plot from Muchlinsky et. al

Steps to conduct Random Forest analysis

- Randomly pick k data points from a training dataset
- Build tree with those data points
- Choose the number of trees (N tree) you want to build and repeat the steps above
- For a new data point,
  - Make each of your Ntree predict Y
  - Assign new data point the average across all of the Y values

Let’s look at a simple example in R

Interested in learning more?

An Introduction to Statistical Learning: With Applications in R

Machine Learning @ JPSM

Machine Learning @GWU

Coursera and other online learning