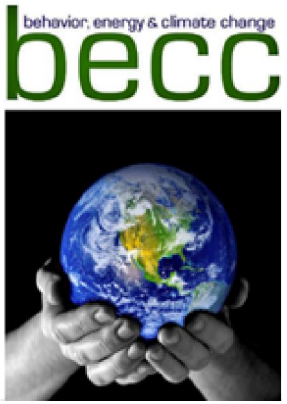
Using Open Data to Predict Energy Usage

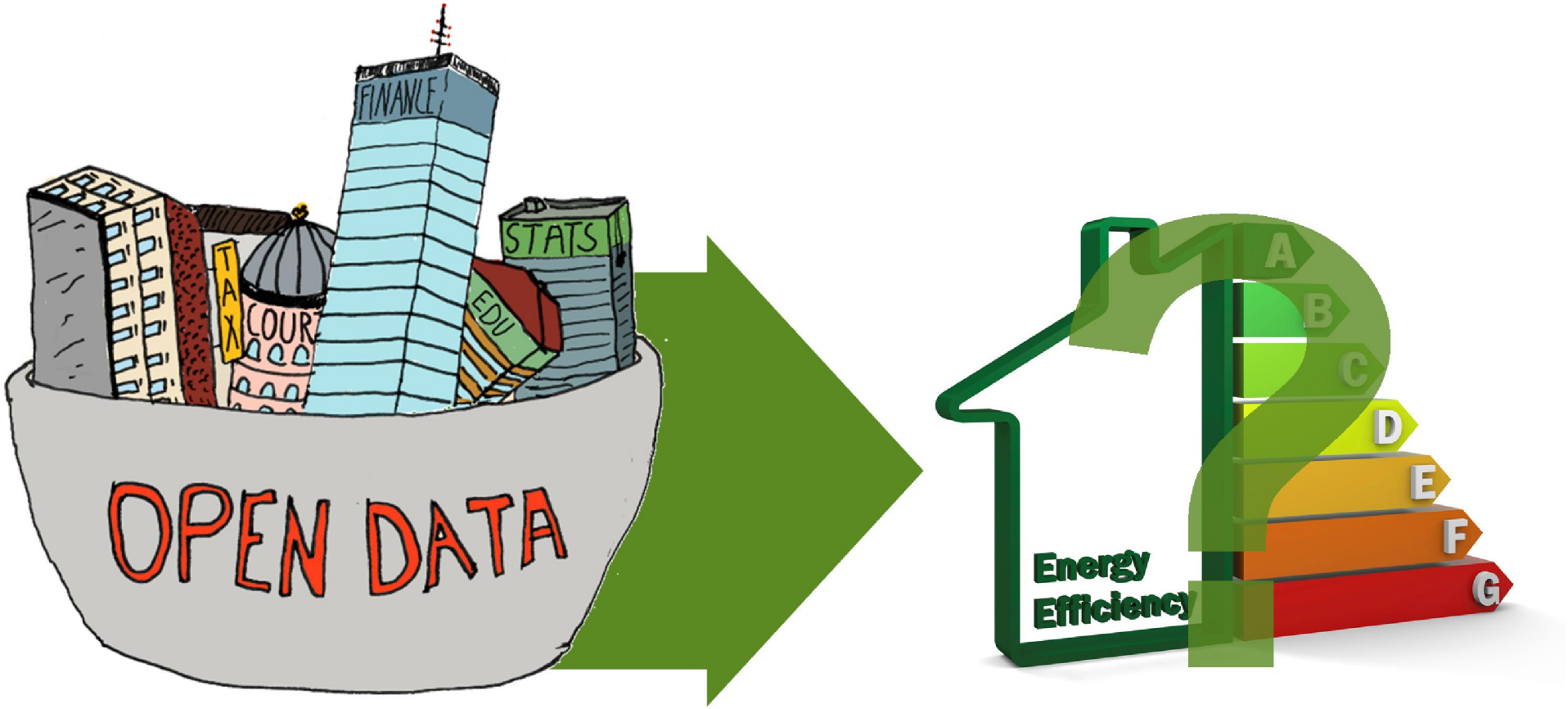
What tax lot data can tell us about energy usage intensity in New York City

Theo Love

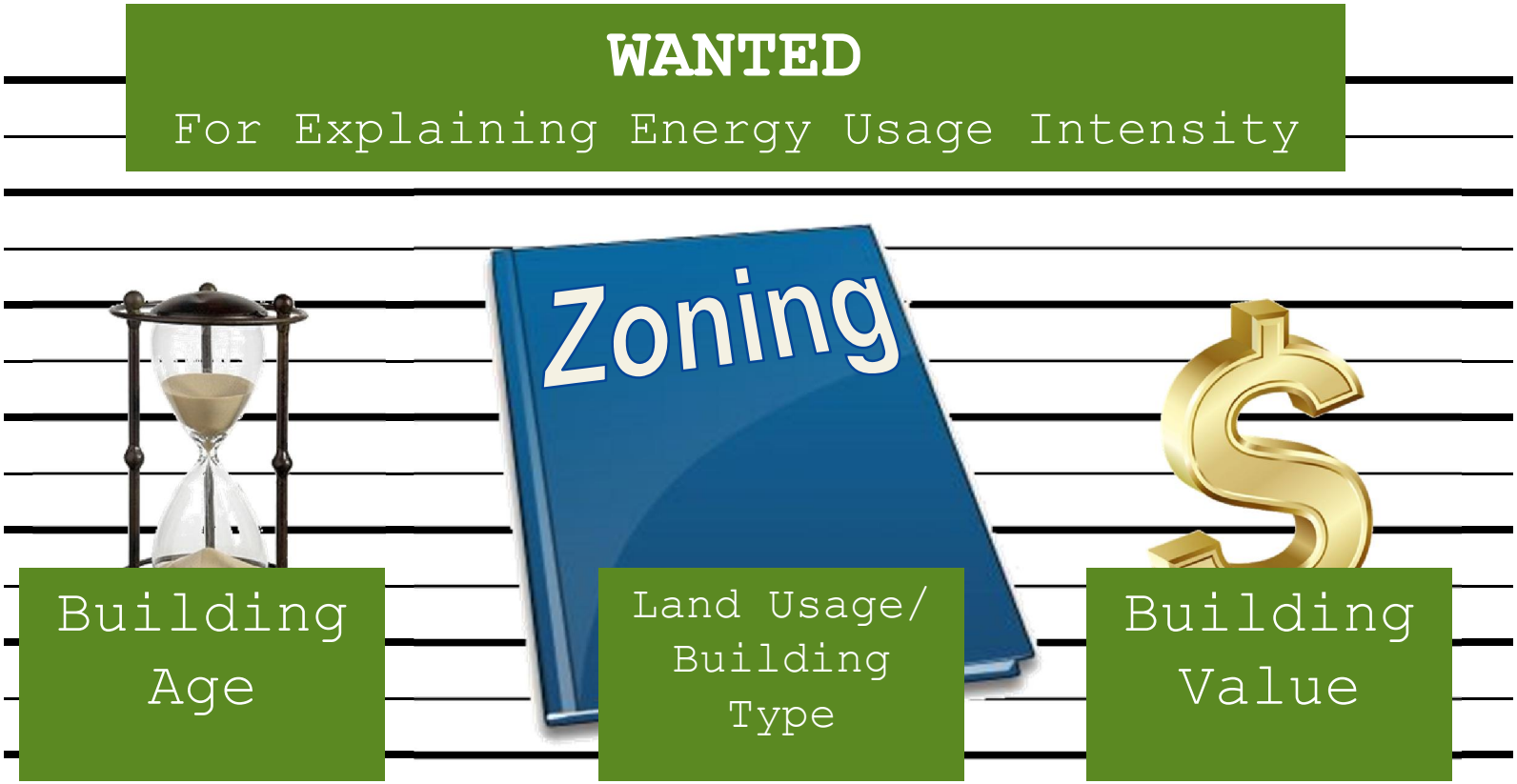
Green Energy Economics Group



What can all this data tell us about how we use energy?



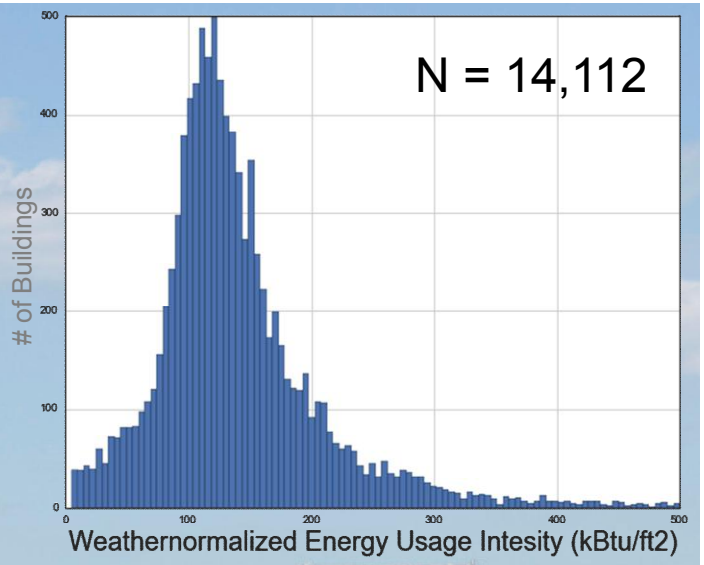
It's the Usual Suspects



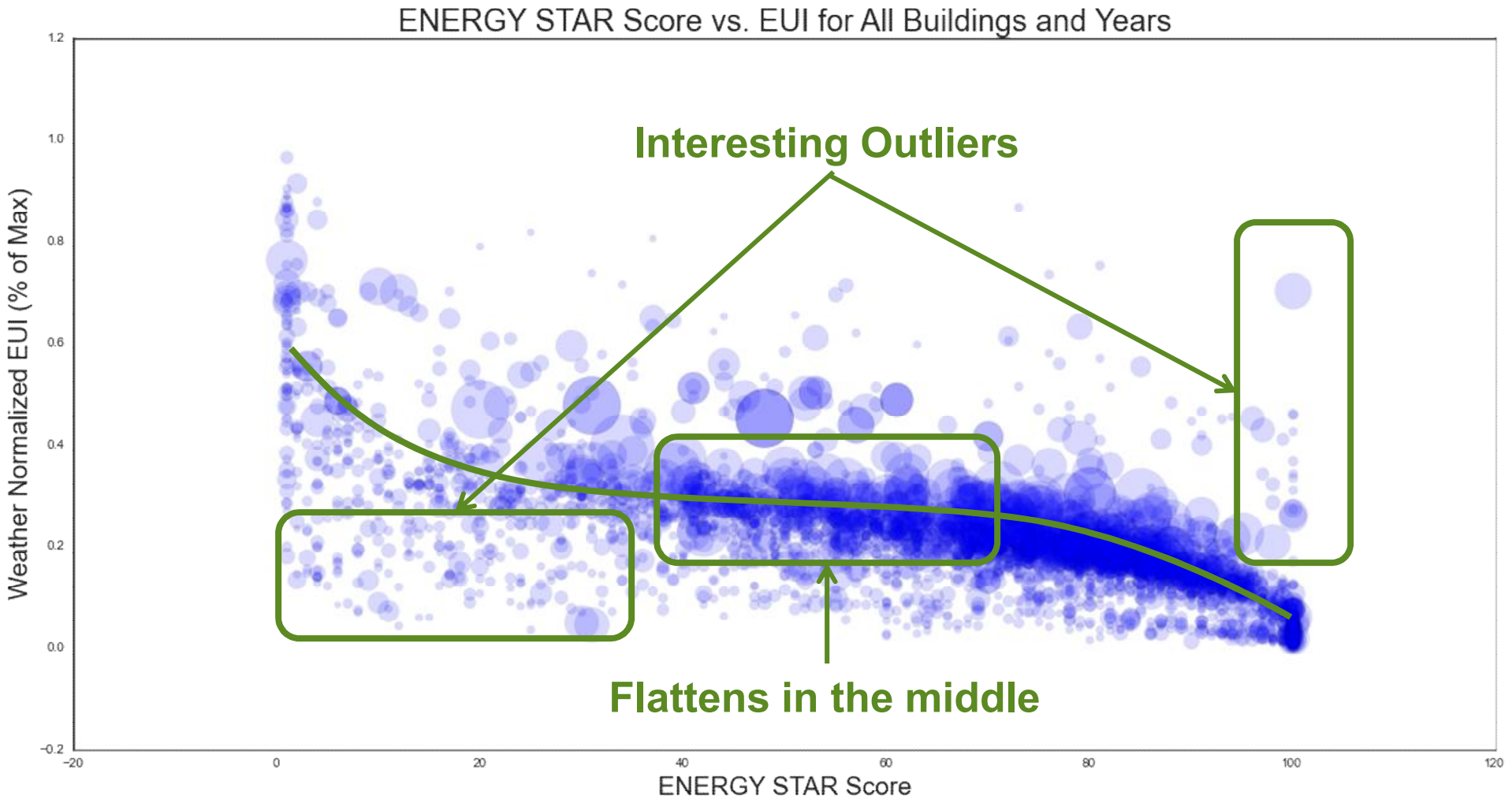
But that's only part of the story…

GreenEnergy EconomicsGroup

Local Law 84

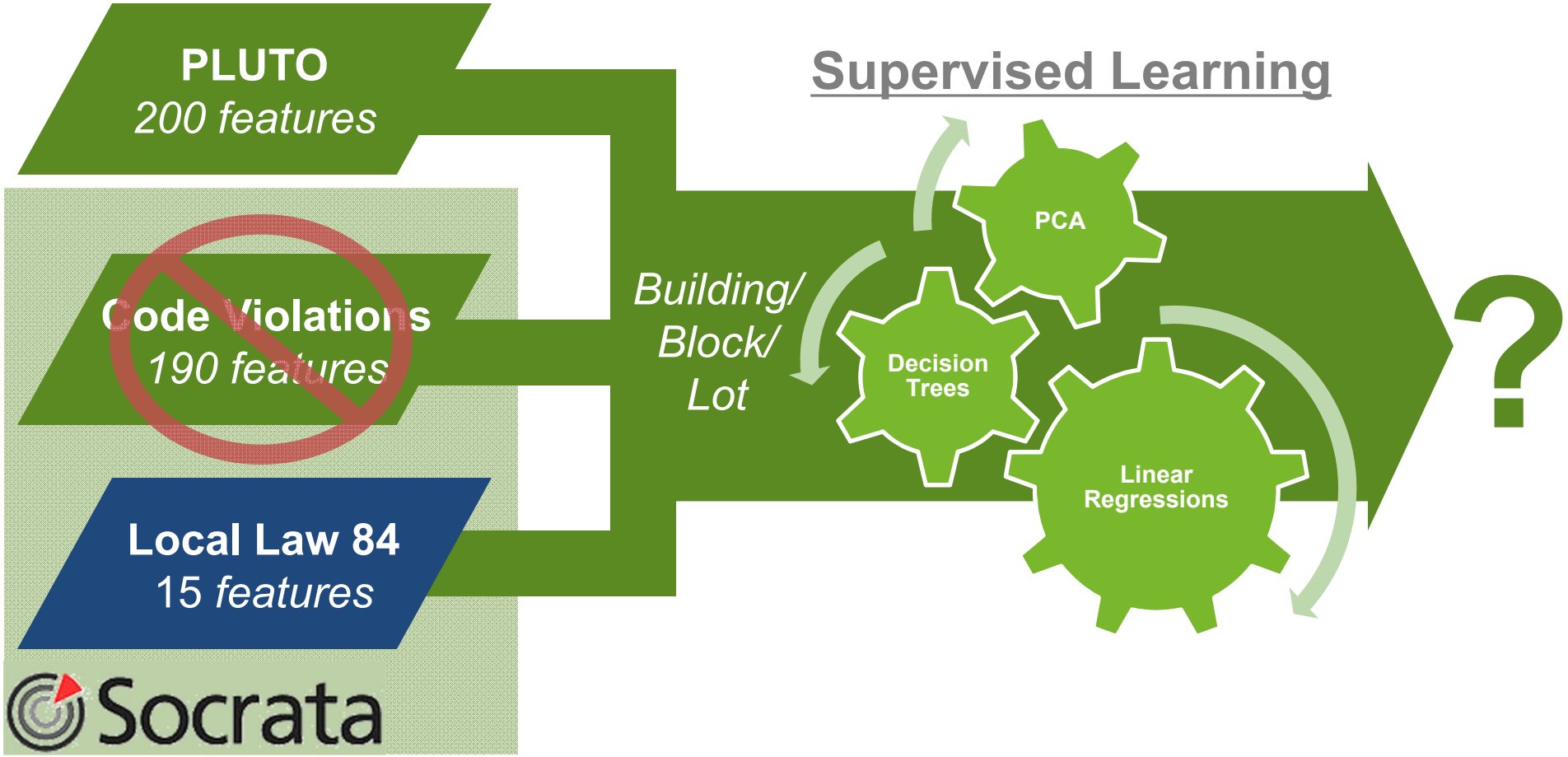

N = 14,112
# of Buildings
0
100
200
300
400
500
Weathernormalized Energy Usage Intesity (kBtu/ft2)

Existing Benchmarks

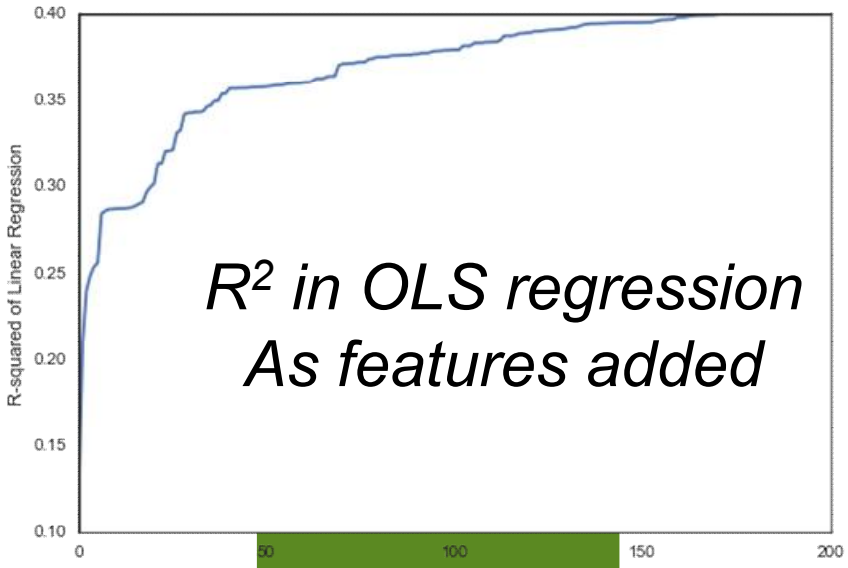


GreenEnergy EconomicsGroup

Going Further with New Data



What Was Found?



Cross-validation

Gini Importance

Tier 1 Features

- Is it a hospital?

Tier 2 Features

- Is it an office?
- Year built

Tier 3 Features

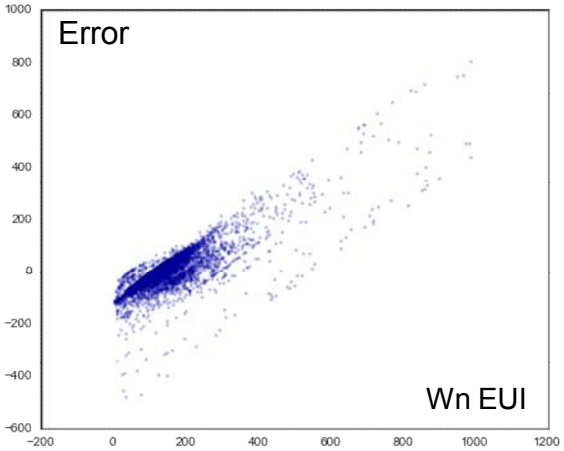
- Value of building/land
- Year of last renovation
- Building size/usage type

Best Model: Random Forest Regression

R^2 from 23% to 29% with standard deviation ~3%

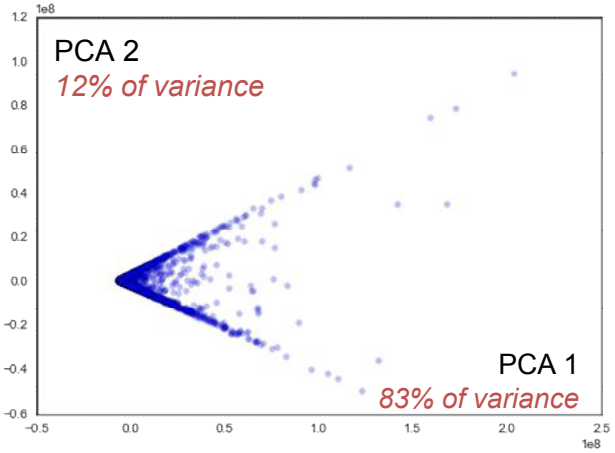
Hints in the Data

Residuals



- Clear linear relationship
- Further classification striations

Primary Component Analysis



- Collapses to two main features
- First component explains vast majority of variance

GreenEnergy EconomicsGroup

What Does it Mean?

- Around **30%** of variance explained by features in PLUTO
- **Sub-sector** analysis **crucial**
- Type **>** Age **>** Value **>** Size

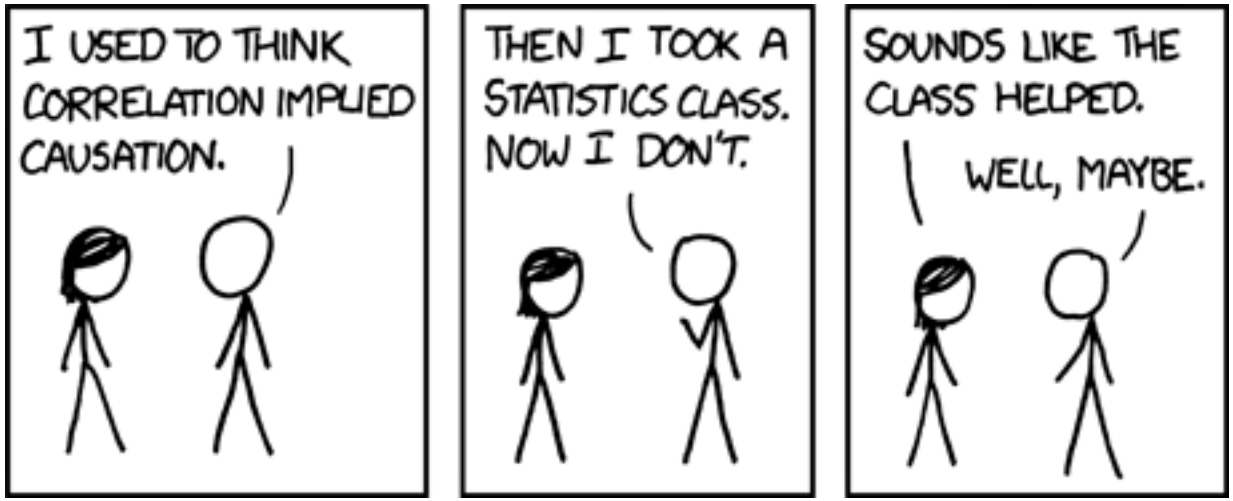
What Now?

- **White paper** with additional details
- **Predicting** out of sample (new years)
- Bring in **more data** to improve model
- Applying and examining in **new areas**

"*Large repositories of public data can help improve our understanding of energy usage.*"

» Digging for **efficiency opportunities**

- Significant features could show hidden usage drivers
- Outliers may mean opportunity



Theo Love

» tlove@greenenergyeconomics.com

GreenEnergy EconomicsGroup