What are the different groups of CO_2 emission countries?

worldbank → transpose → [year | col]

worldbank → raw → [col | year]

feature selection → Boruta (Ryan)

"important" columns → missing value analysis → Kyle

imputation → 50%

EDA

ML + Clustering
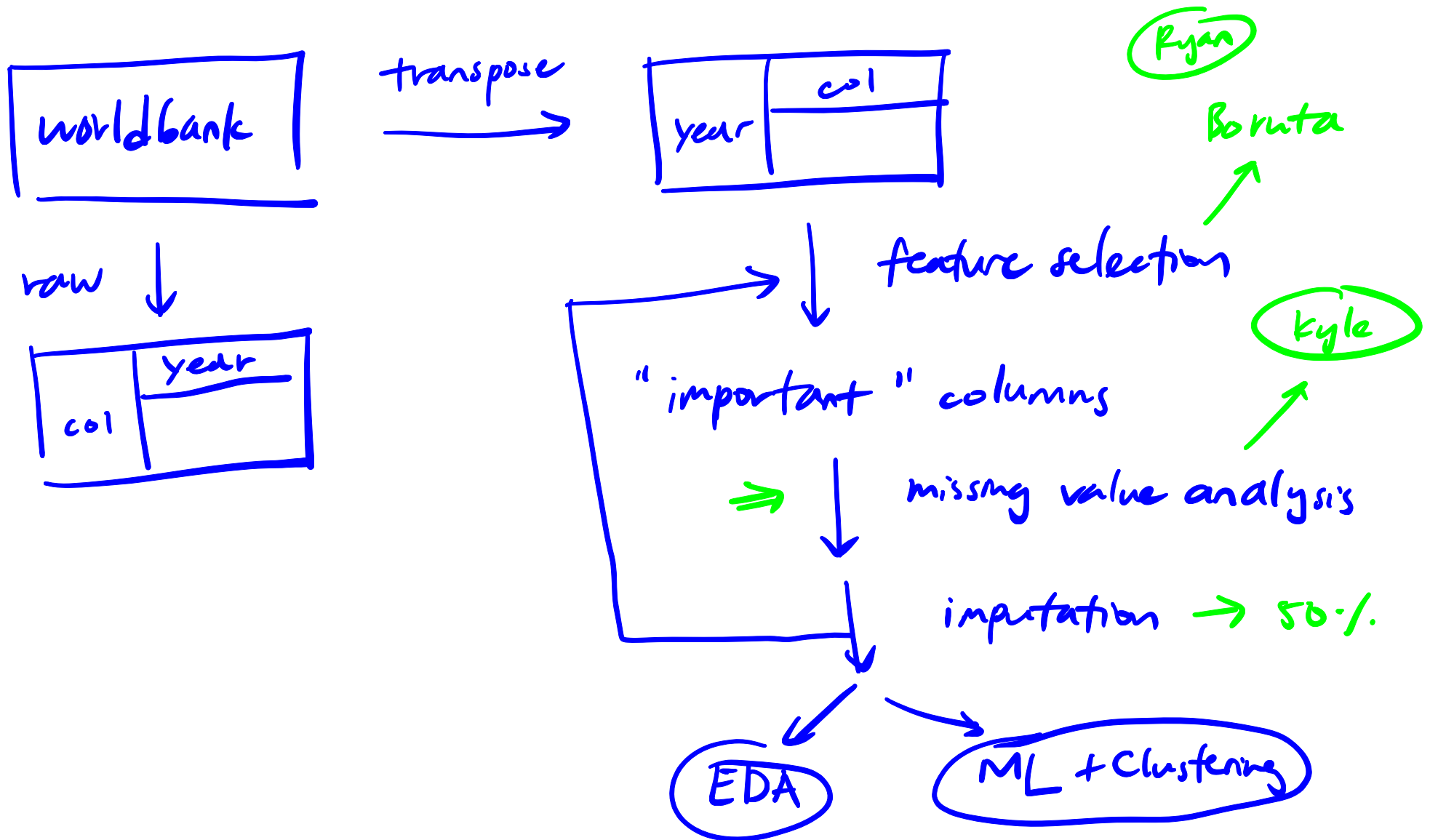
New Dataflow

worldbank

col	year

→ transpose →

year	col

↓ remove noisy country

↓ drop col with >50% missing

fill NA with median

[df]

feature selection
(BORUTA)

[columns to keep]

"new" Multivariate imputation "clean"

[df] → [df]

EDA

ML + Clustering