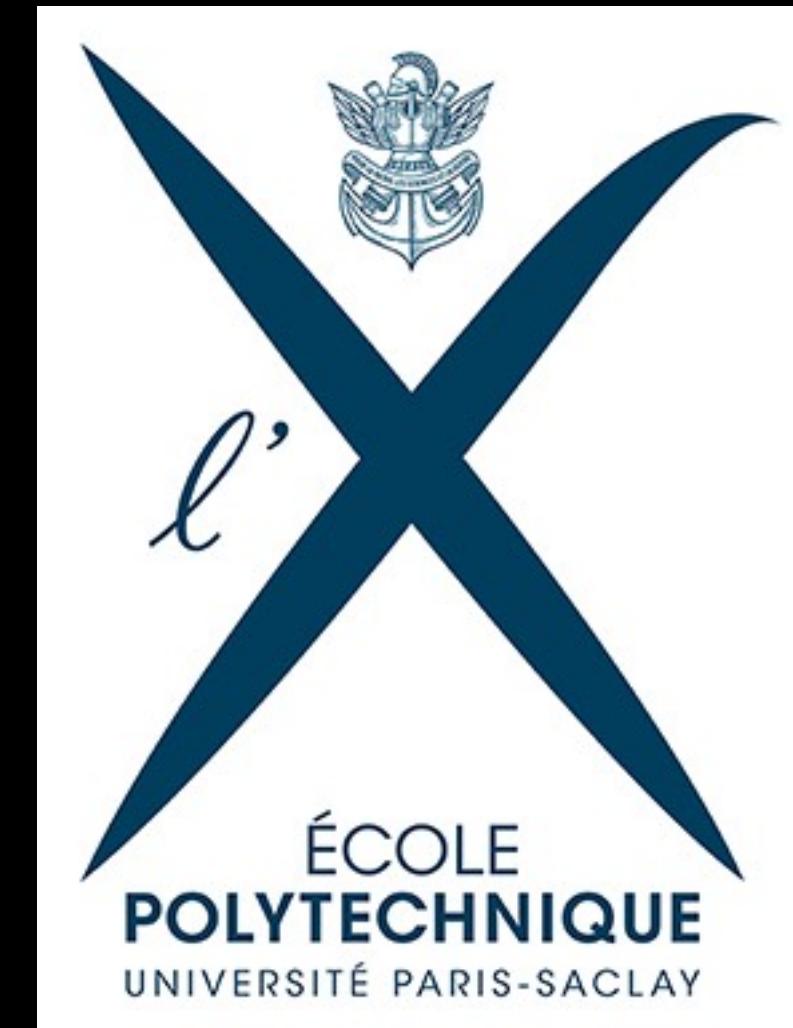
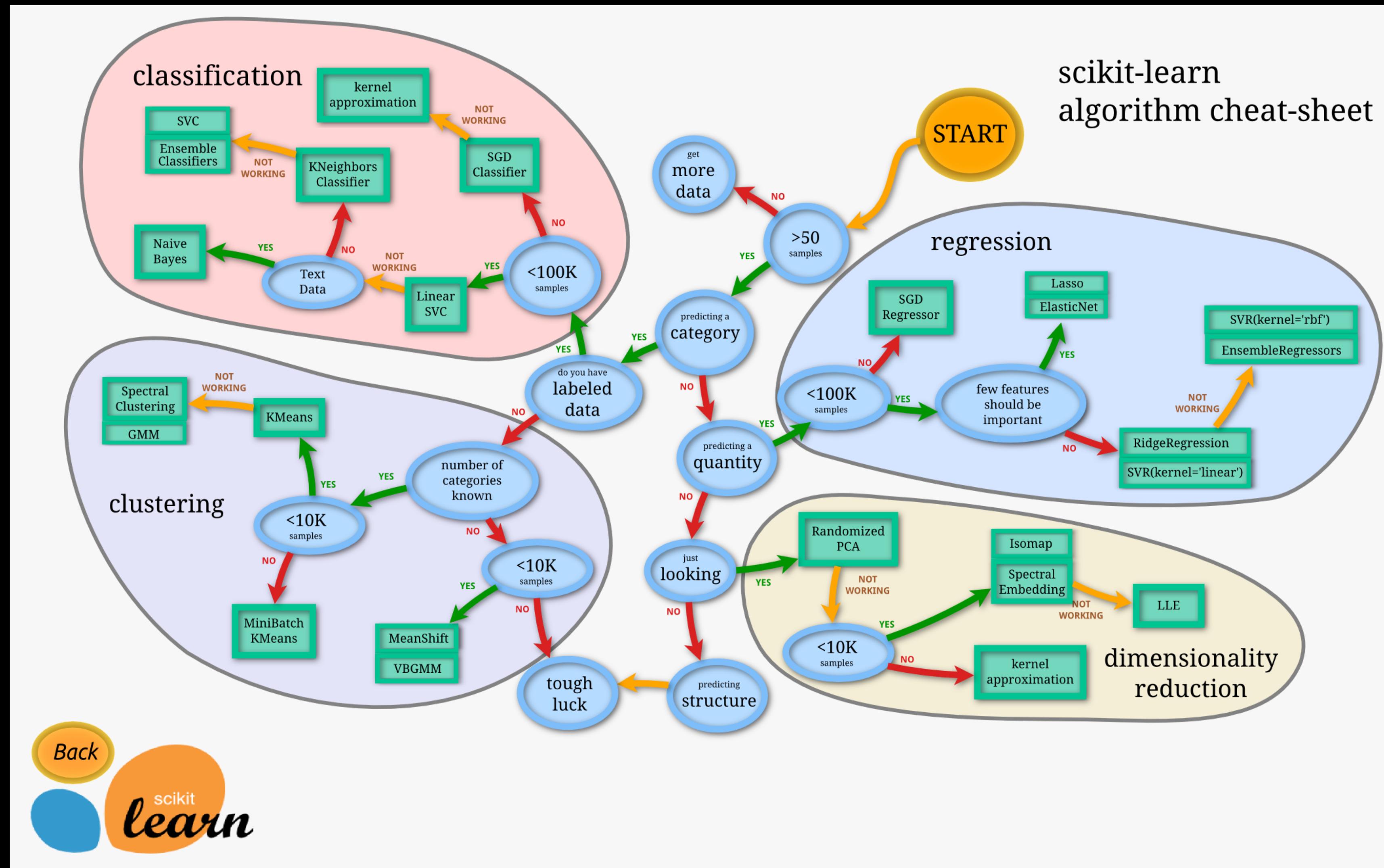


Supervised Learning Classification

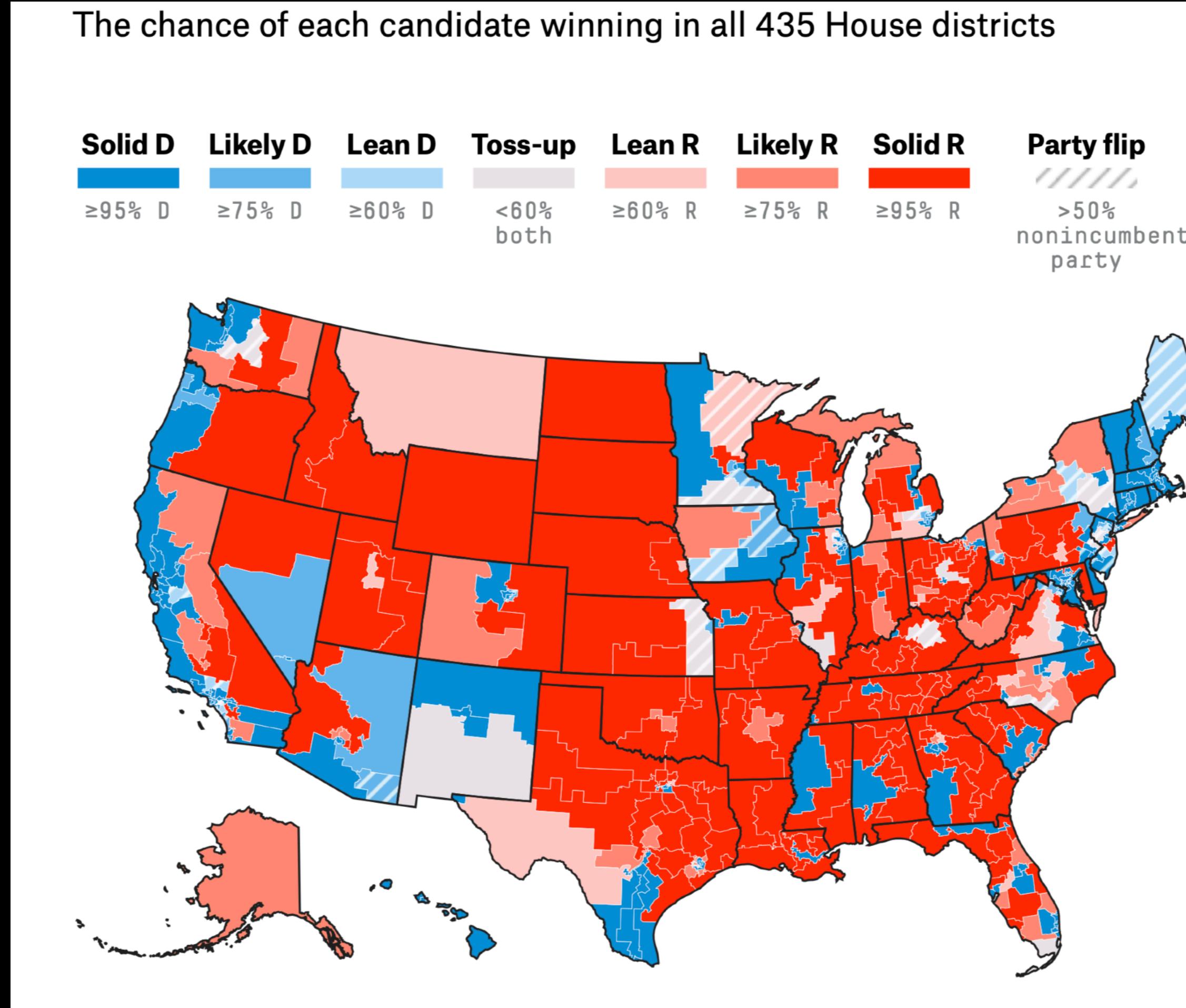


X-Ed
Demian Wassermann, Claire Boyer
Inria, France

Where are we in the Data Science Landscape?

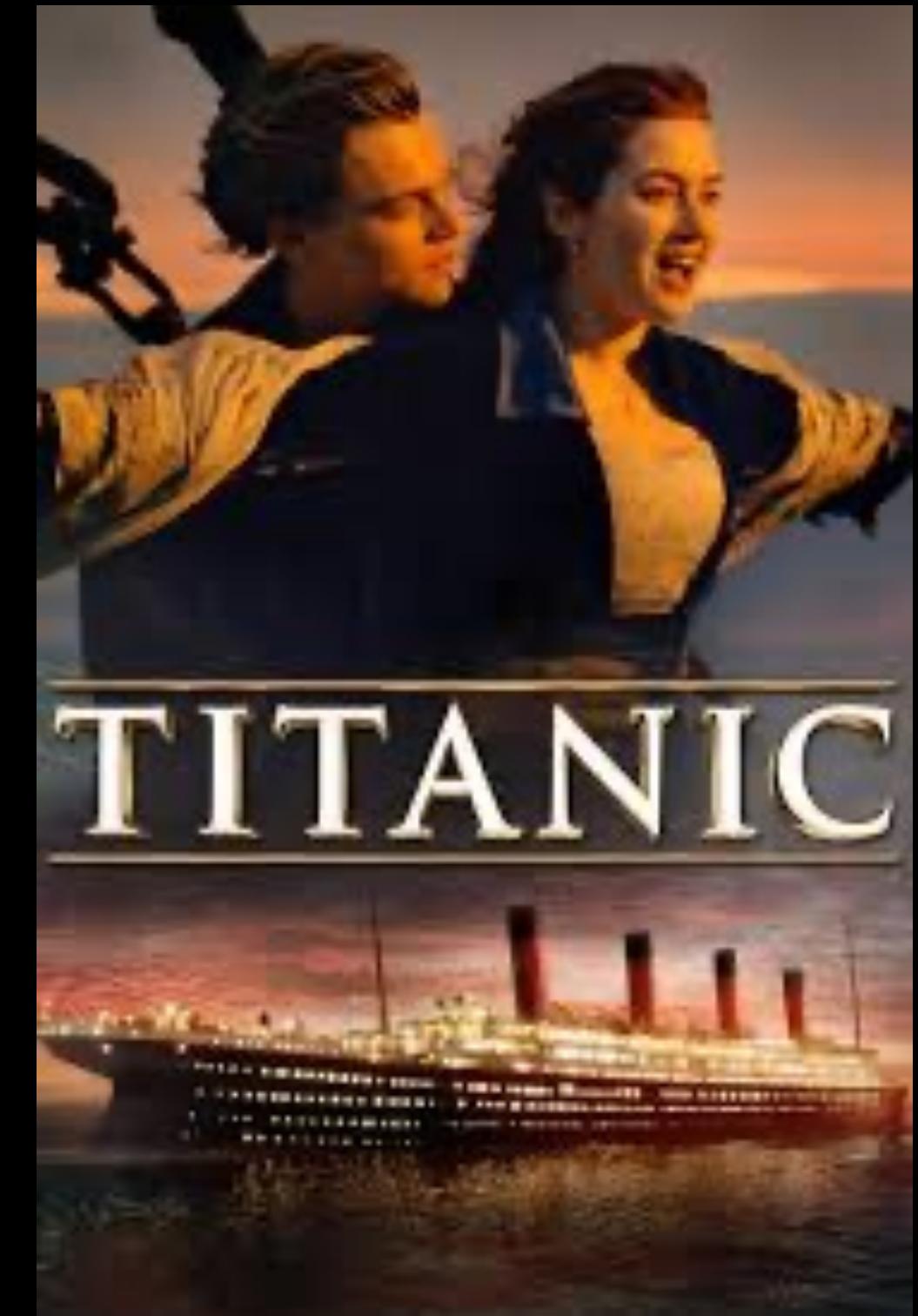
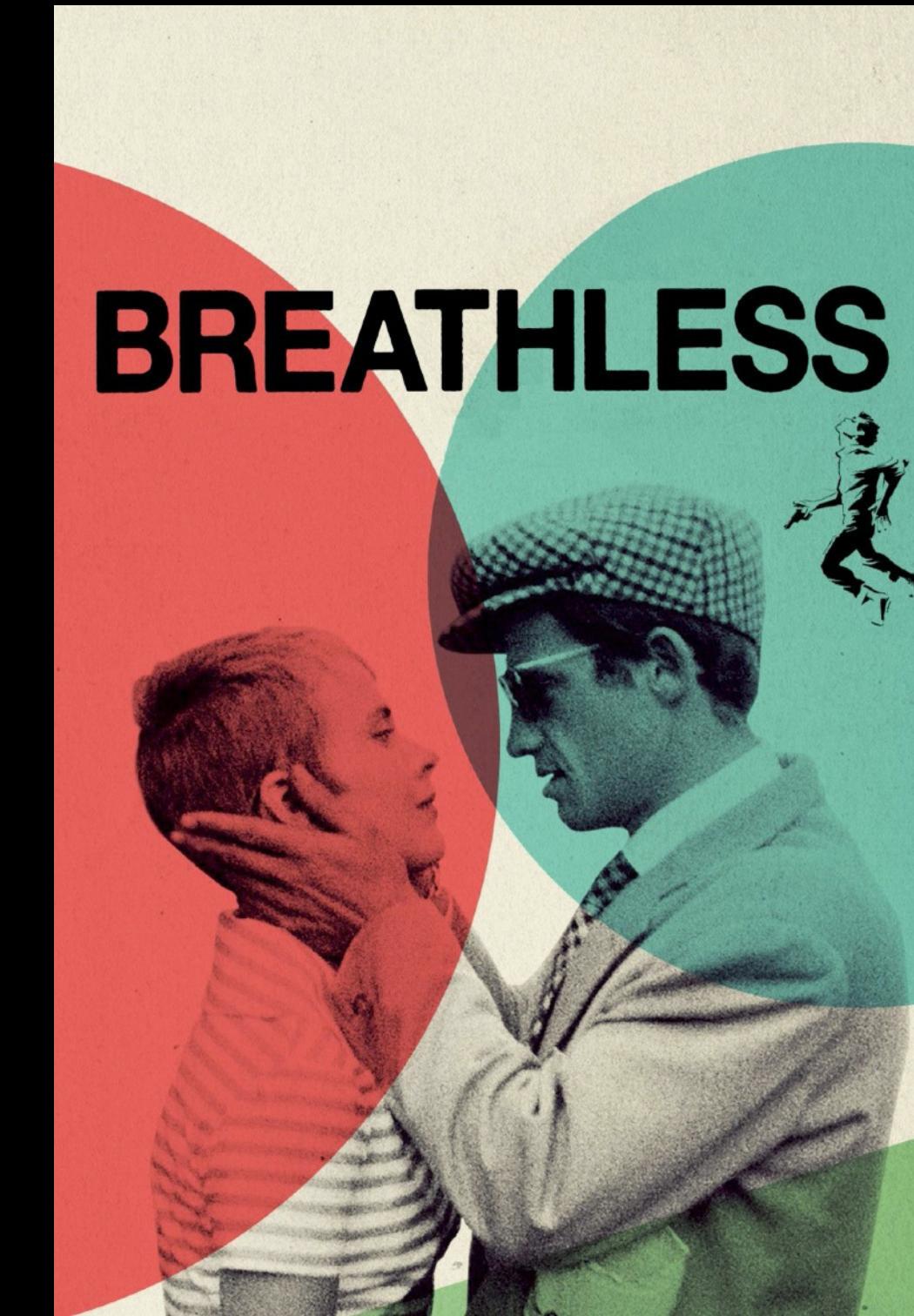
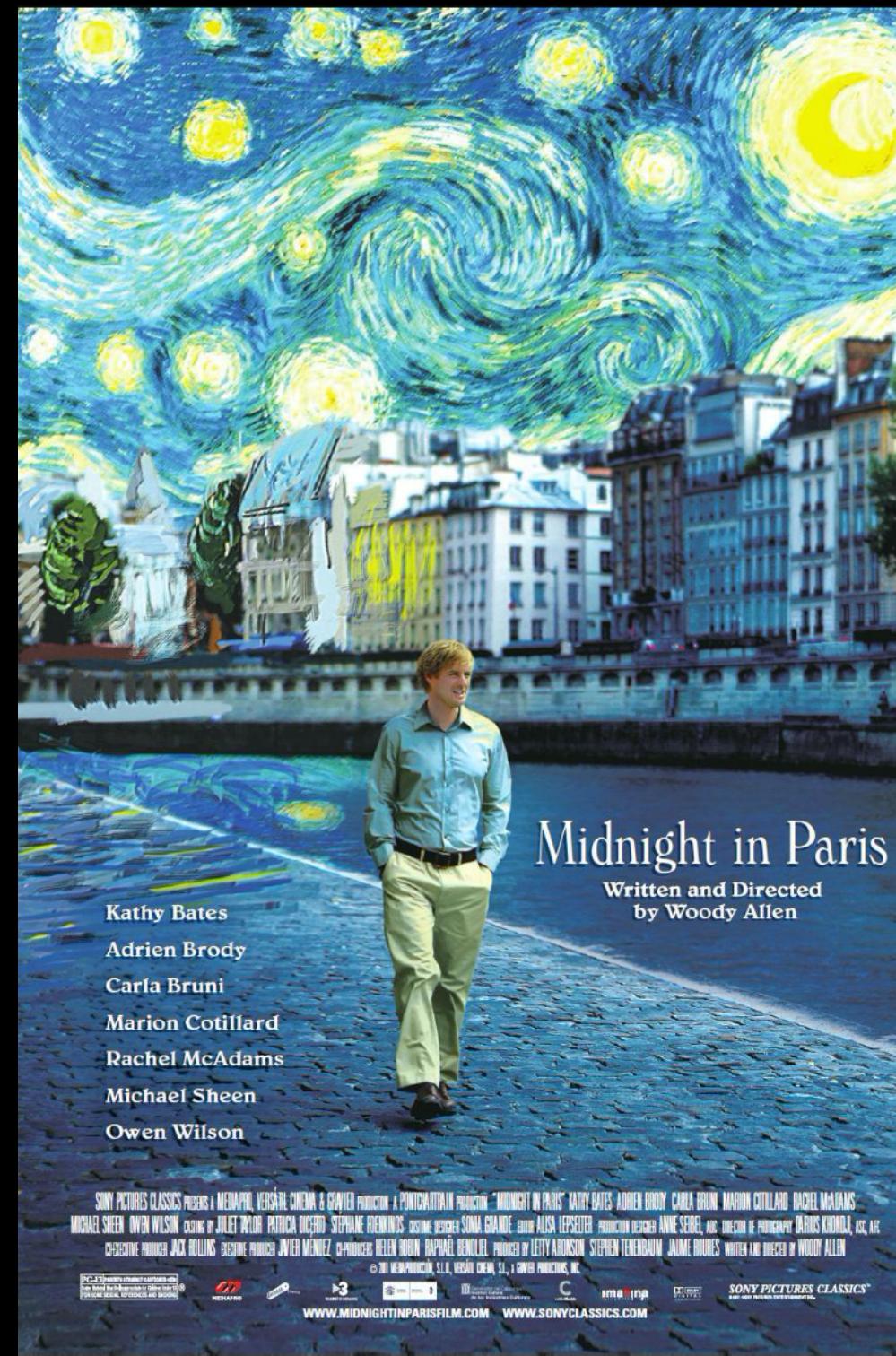


Motivating Example



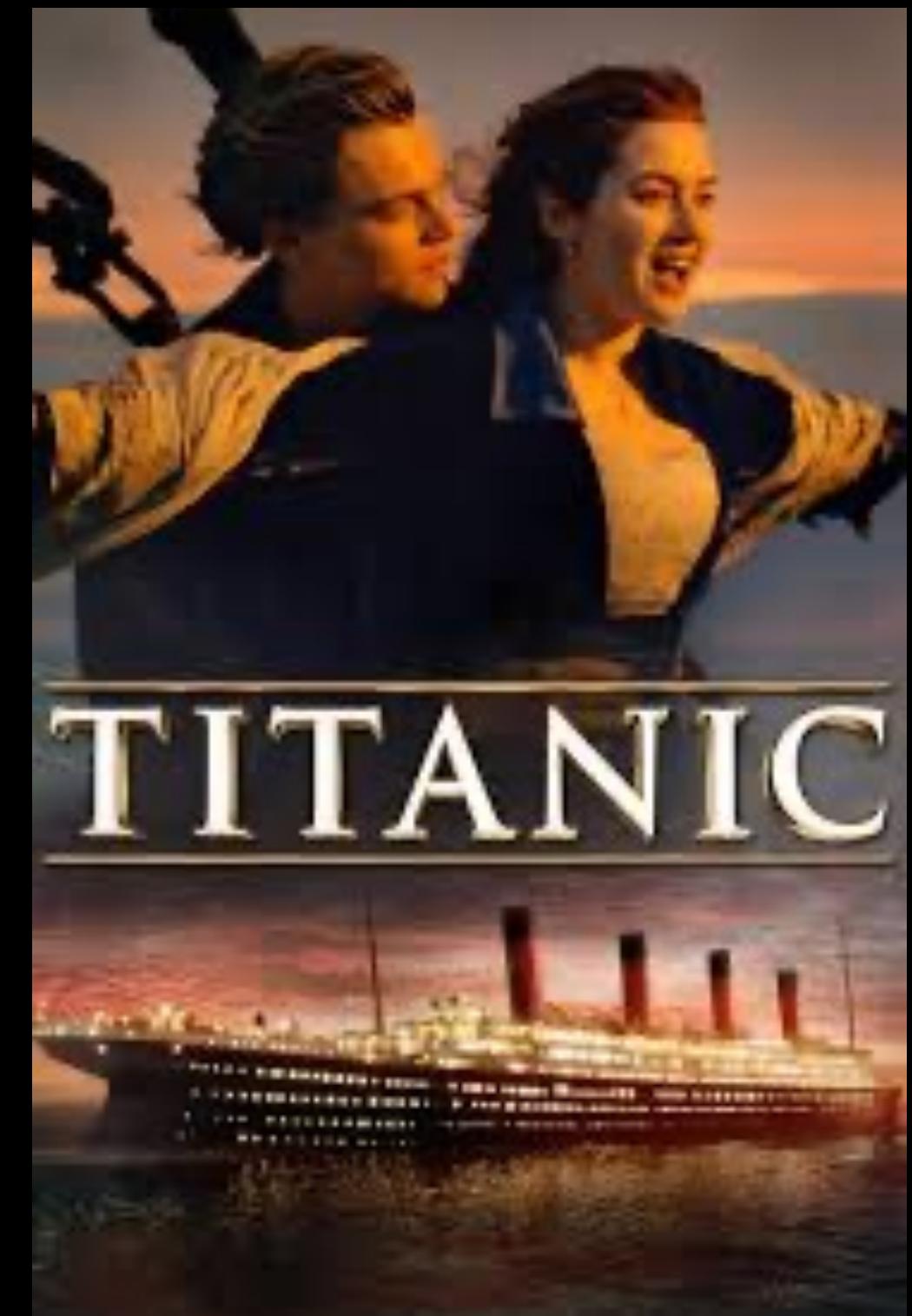
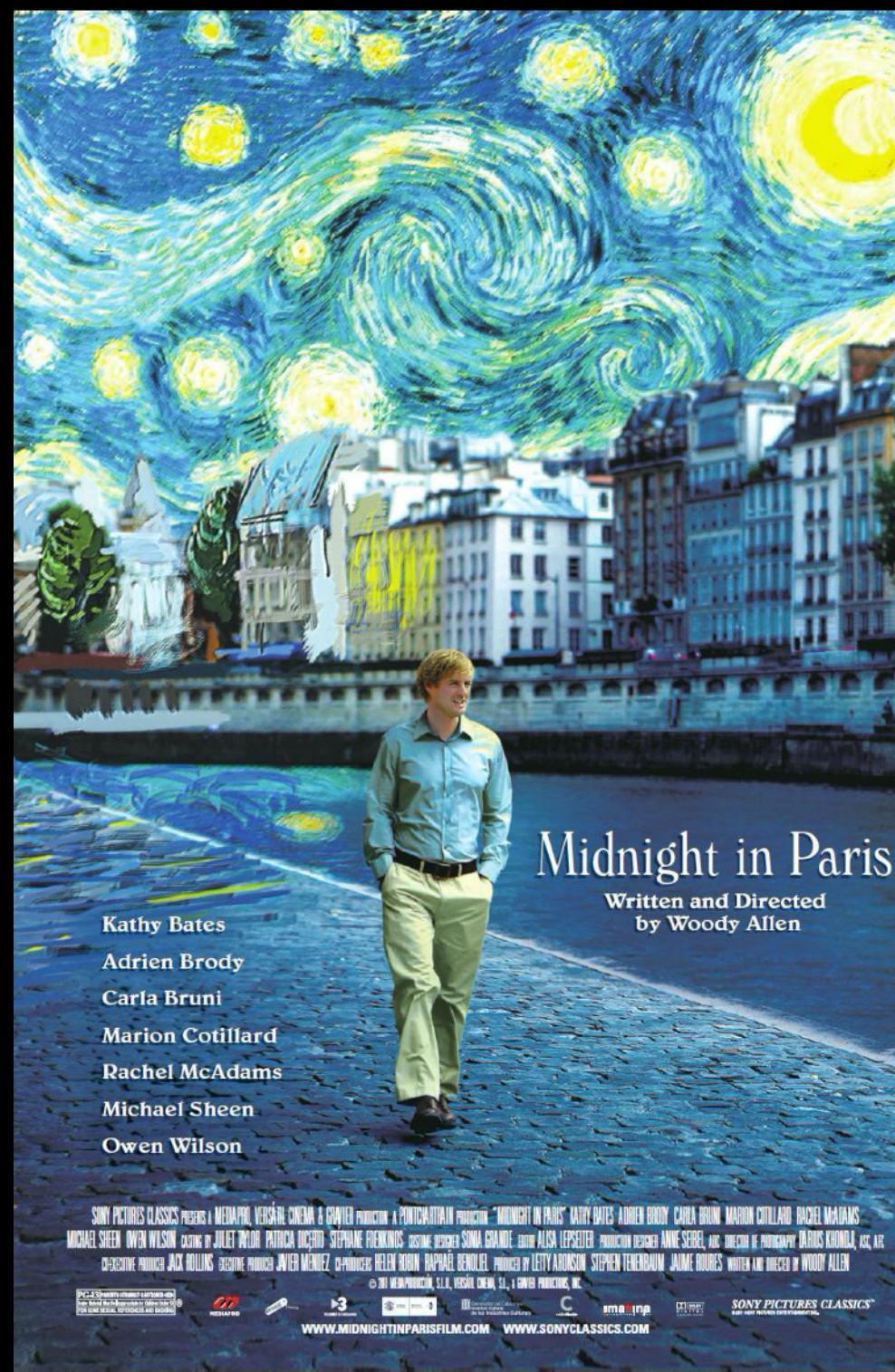
Regression Implies a Model of Reality

“A movie is better when there are more explosions”



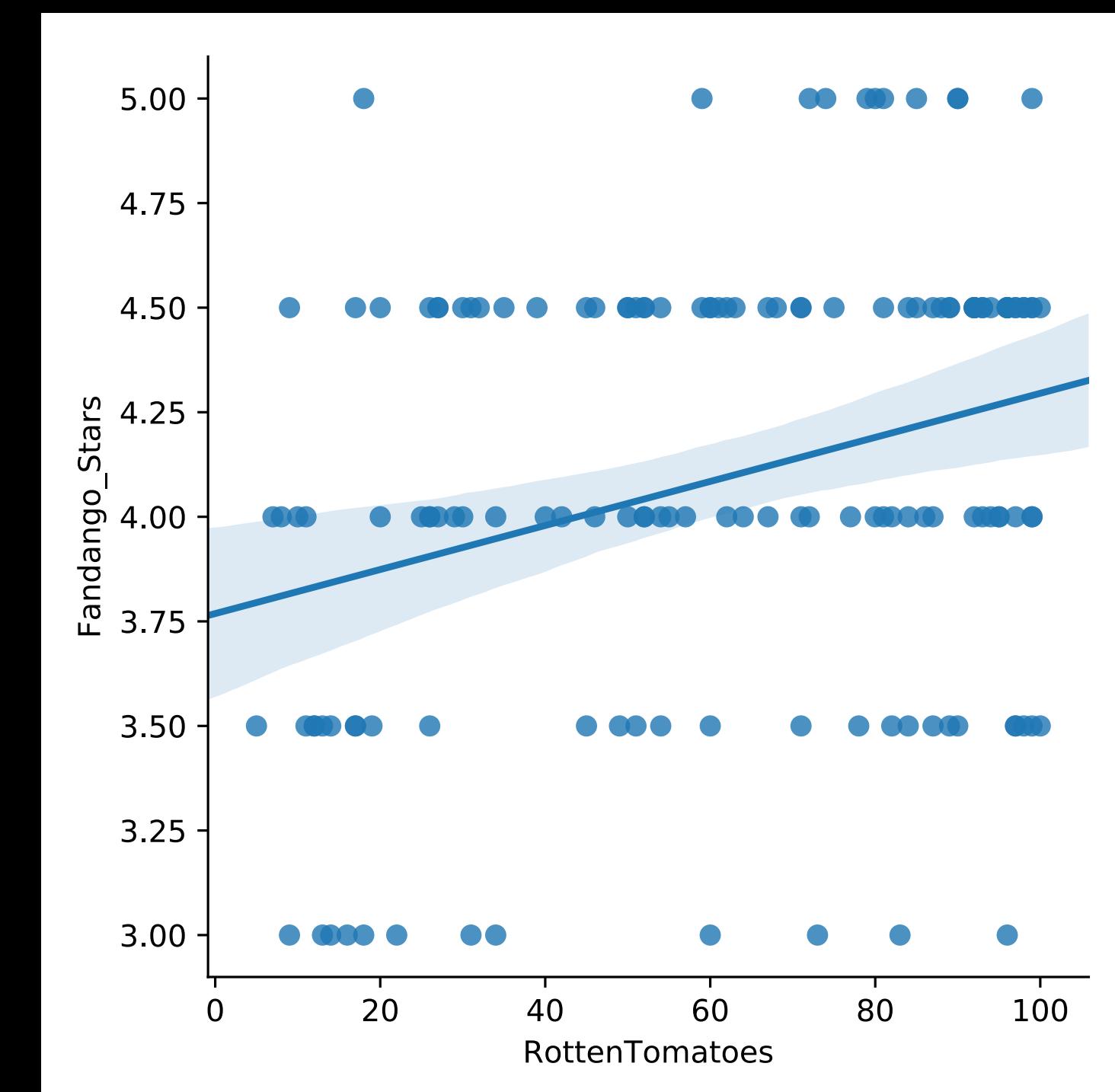
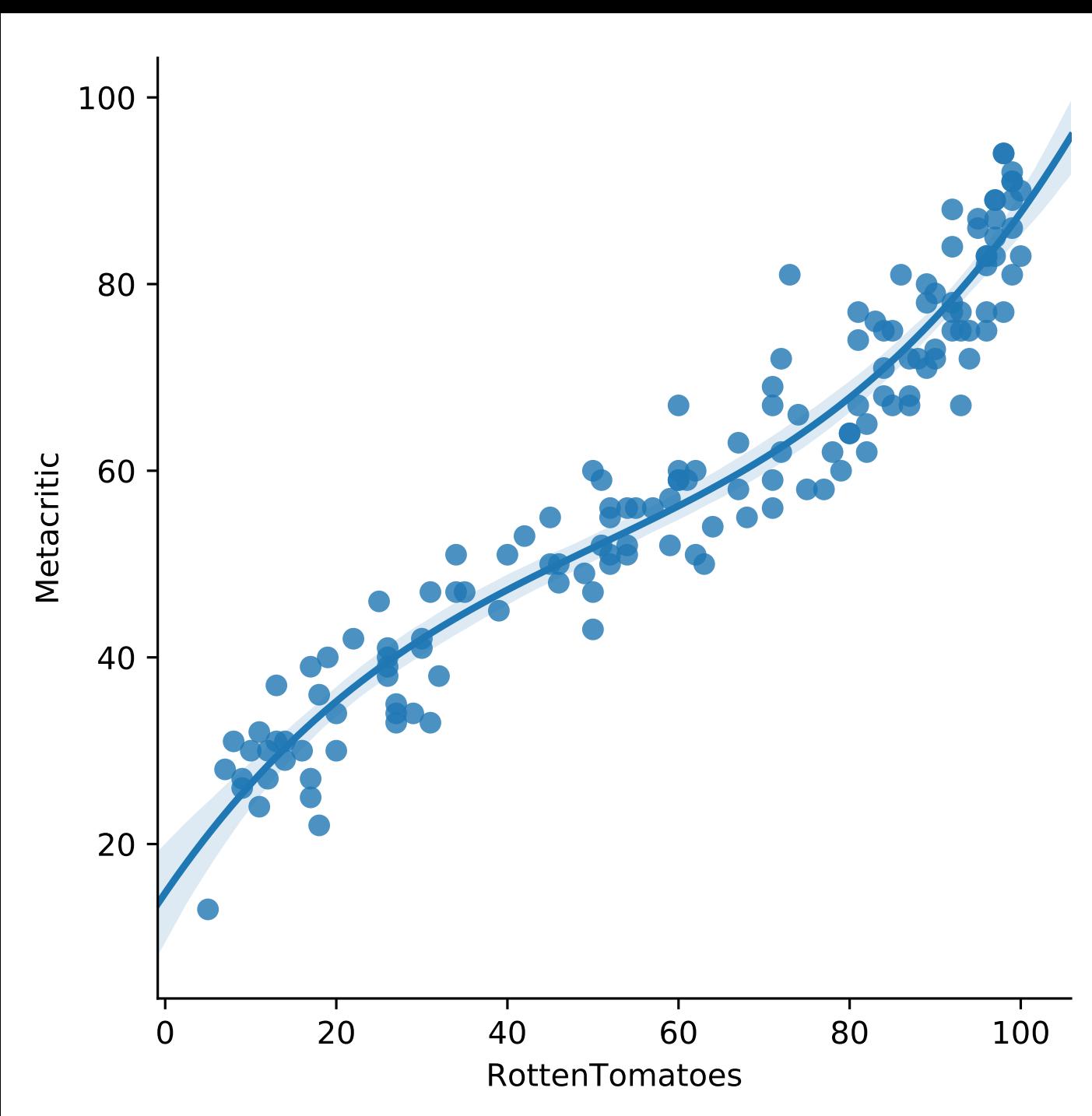
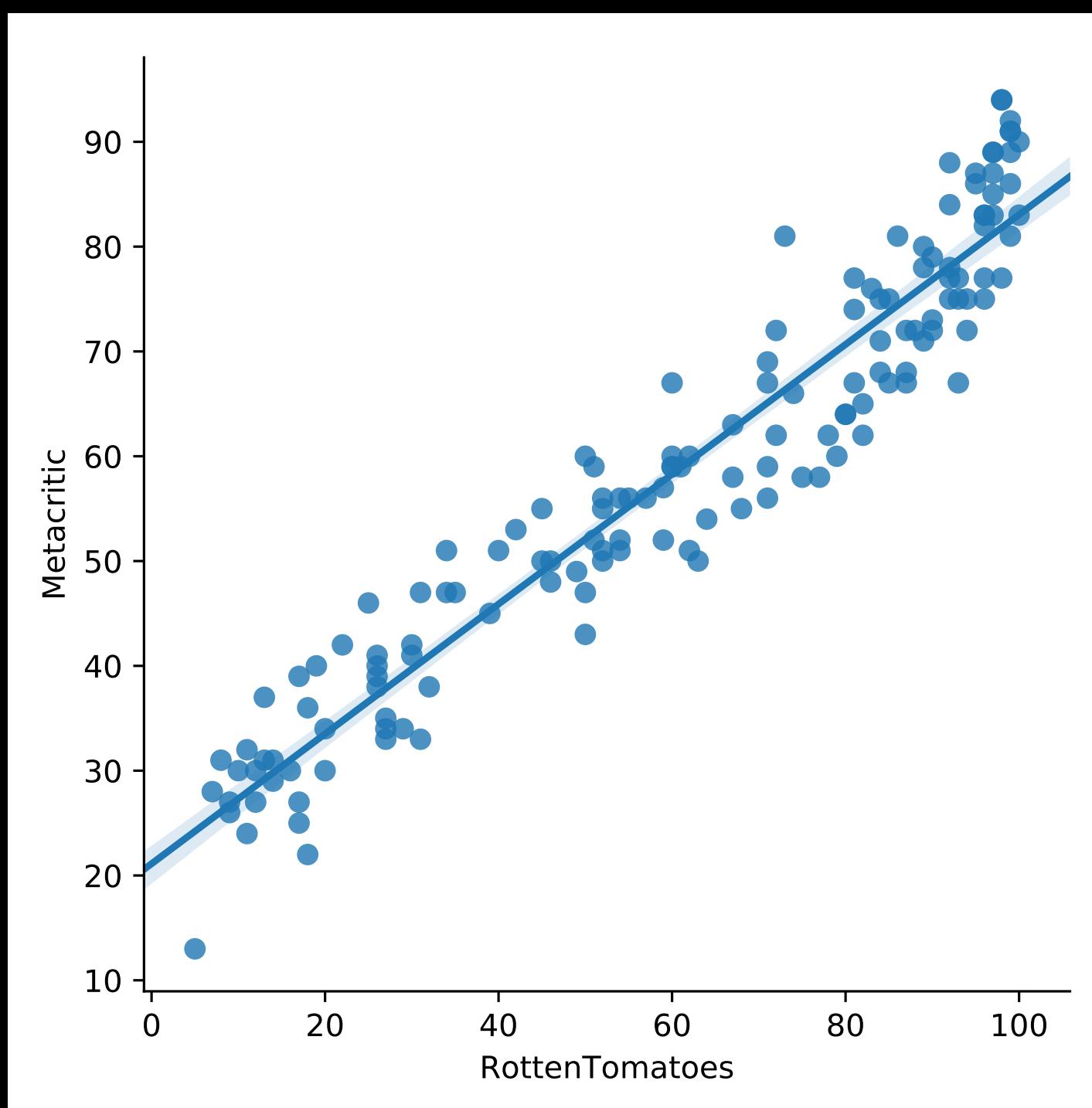
Regression Implies a Model of Reality

“A movie is **good** when there are more explosions”



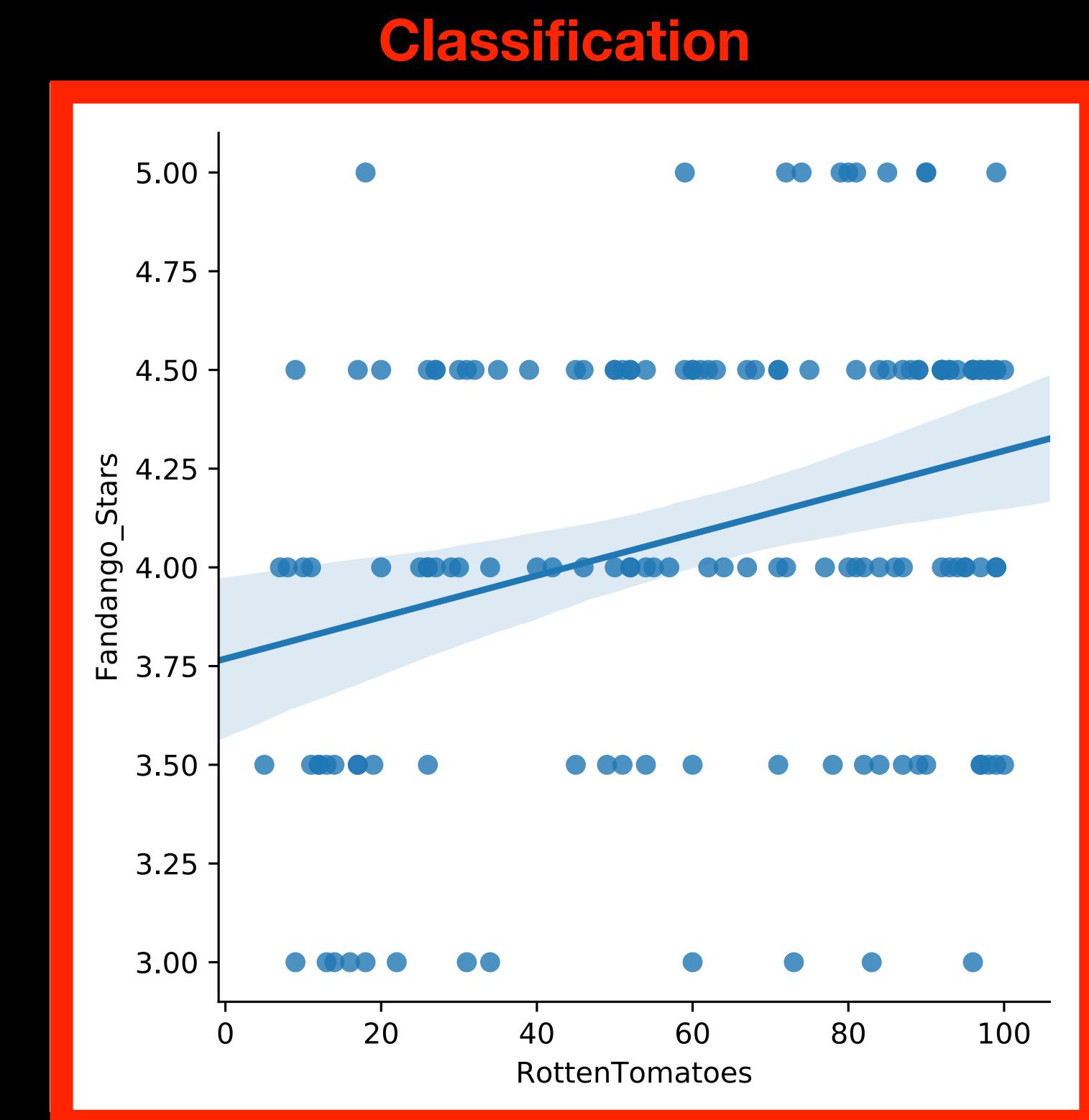
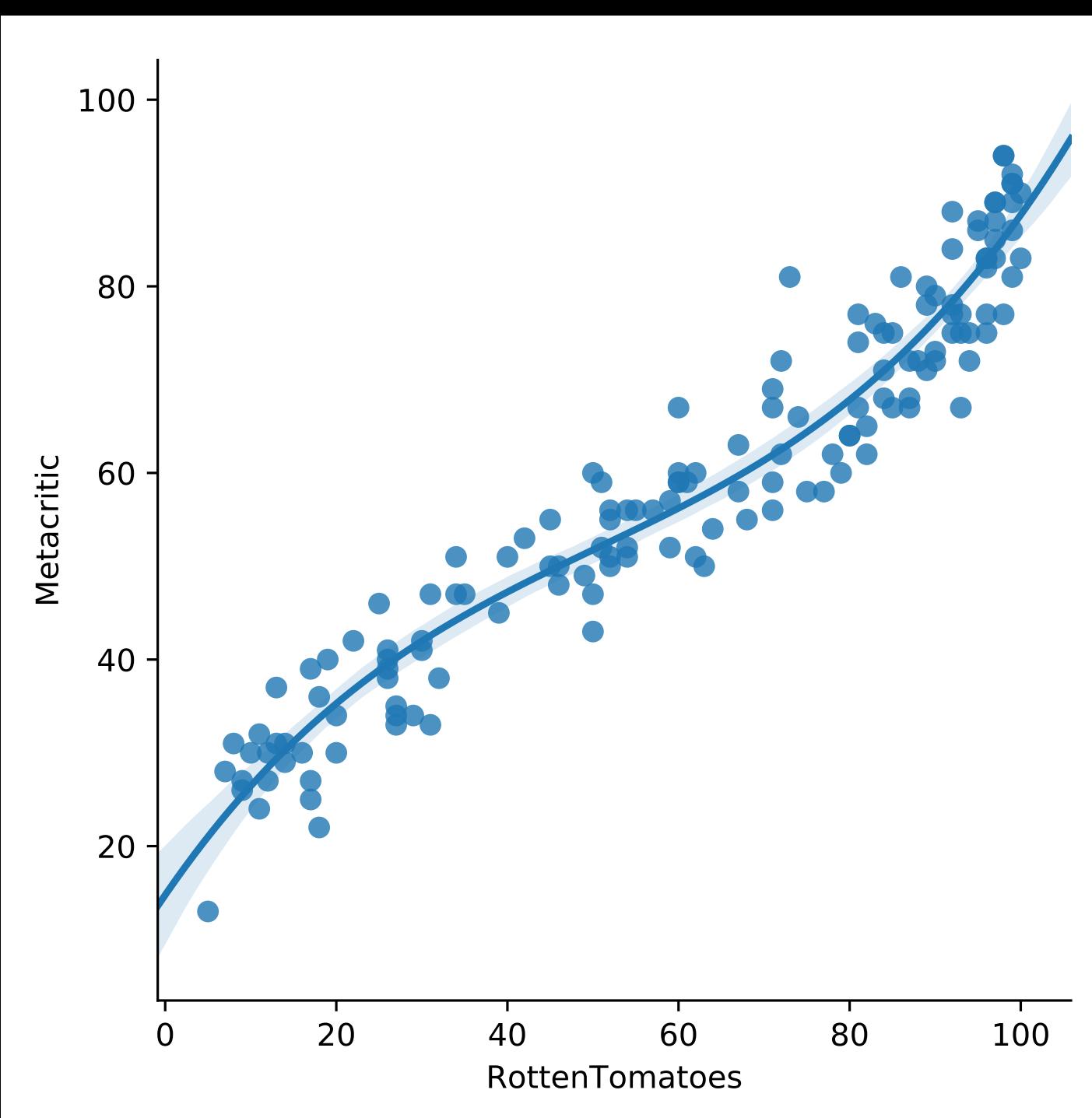
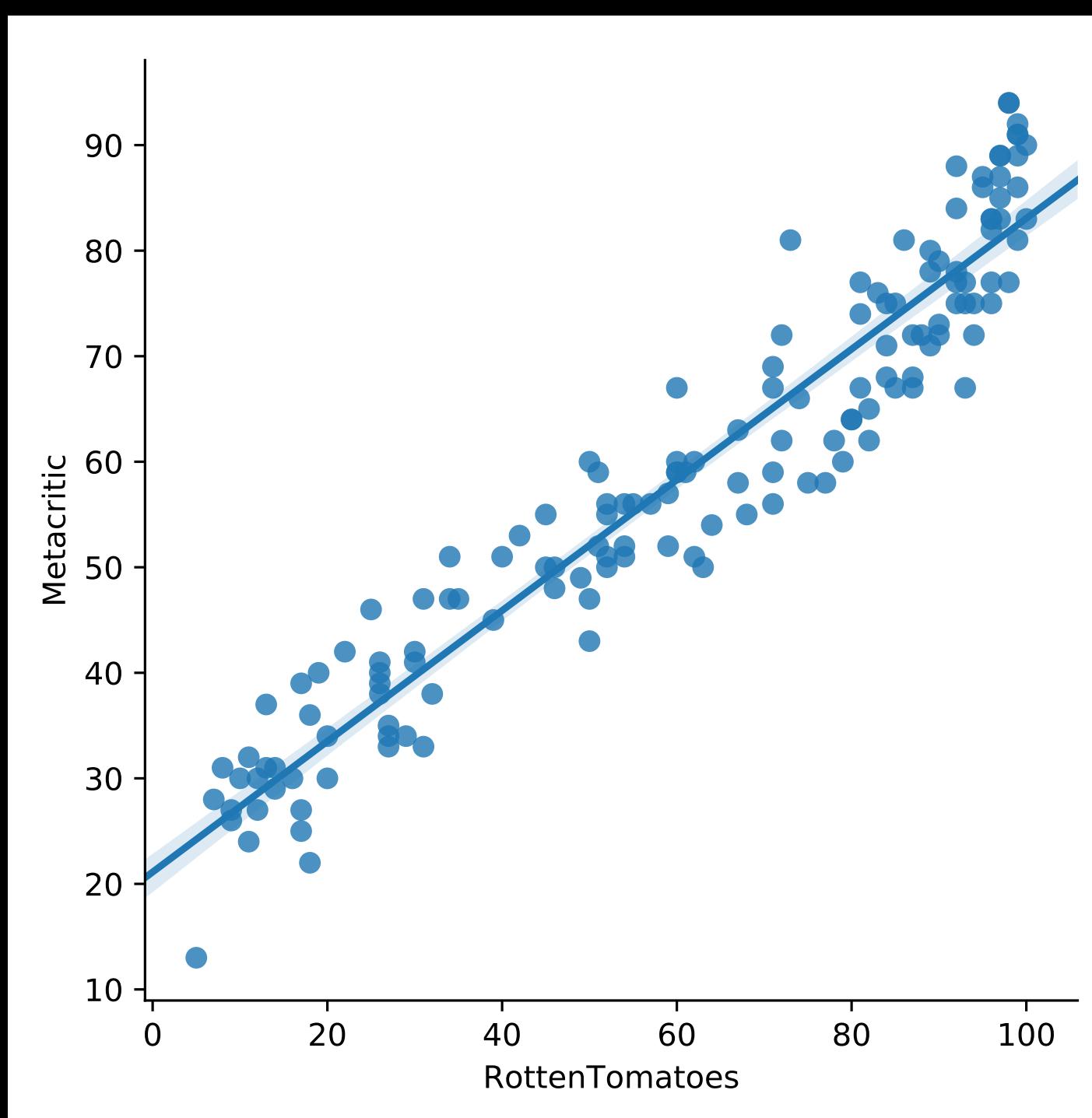
Model Selection - Regression Model

Are movie recommendation systems consistent?



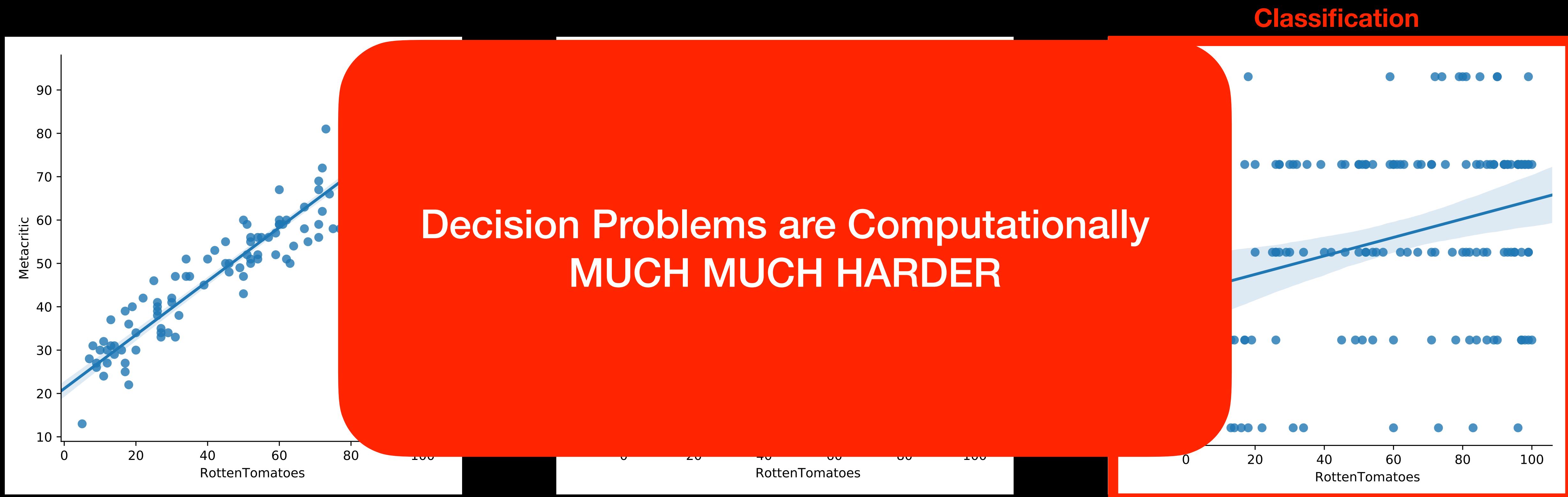
Model Selection - Regression Model

Are movie recommendation systems consistent?

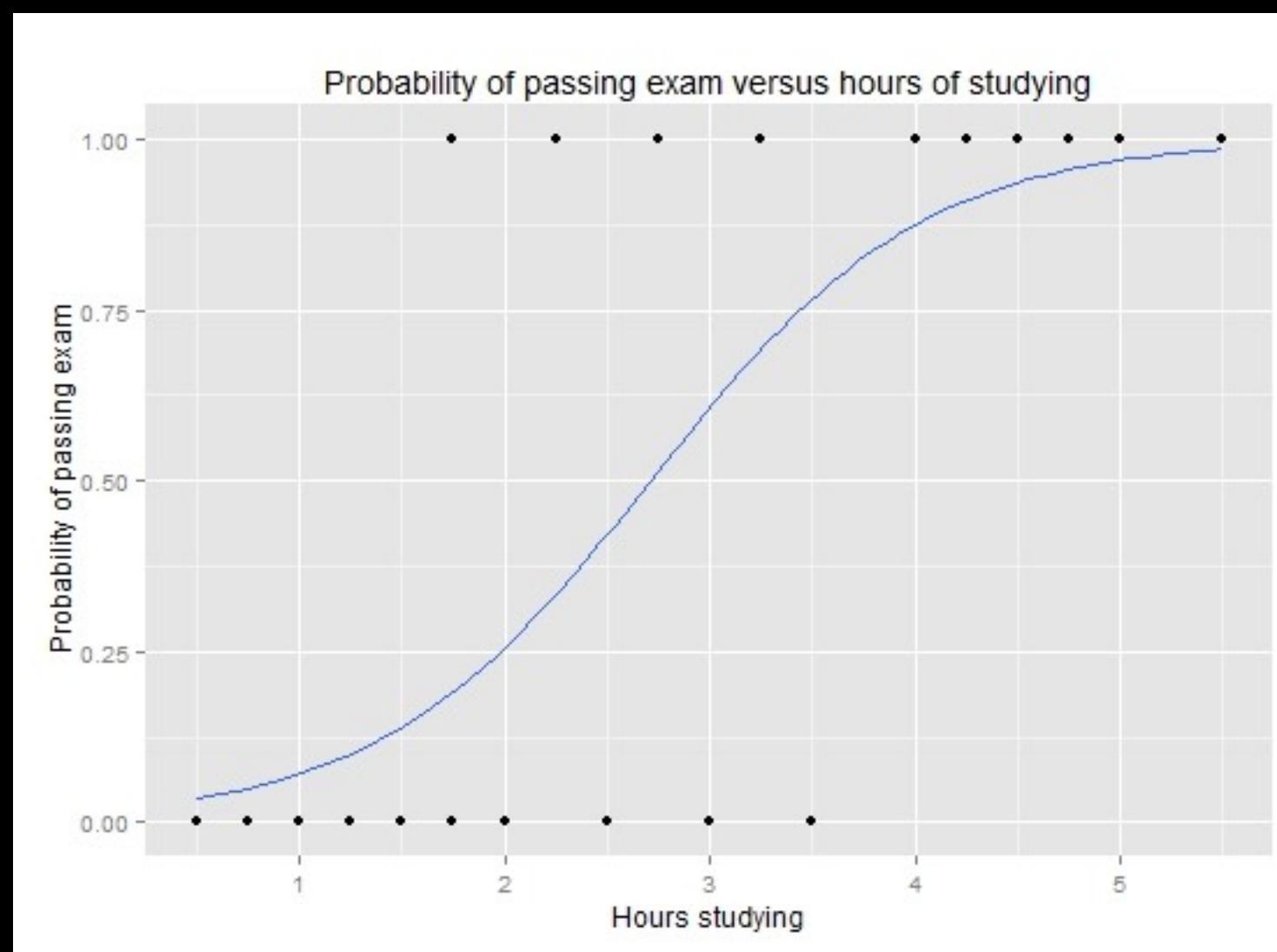


Model Selection - Regression Model

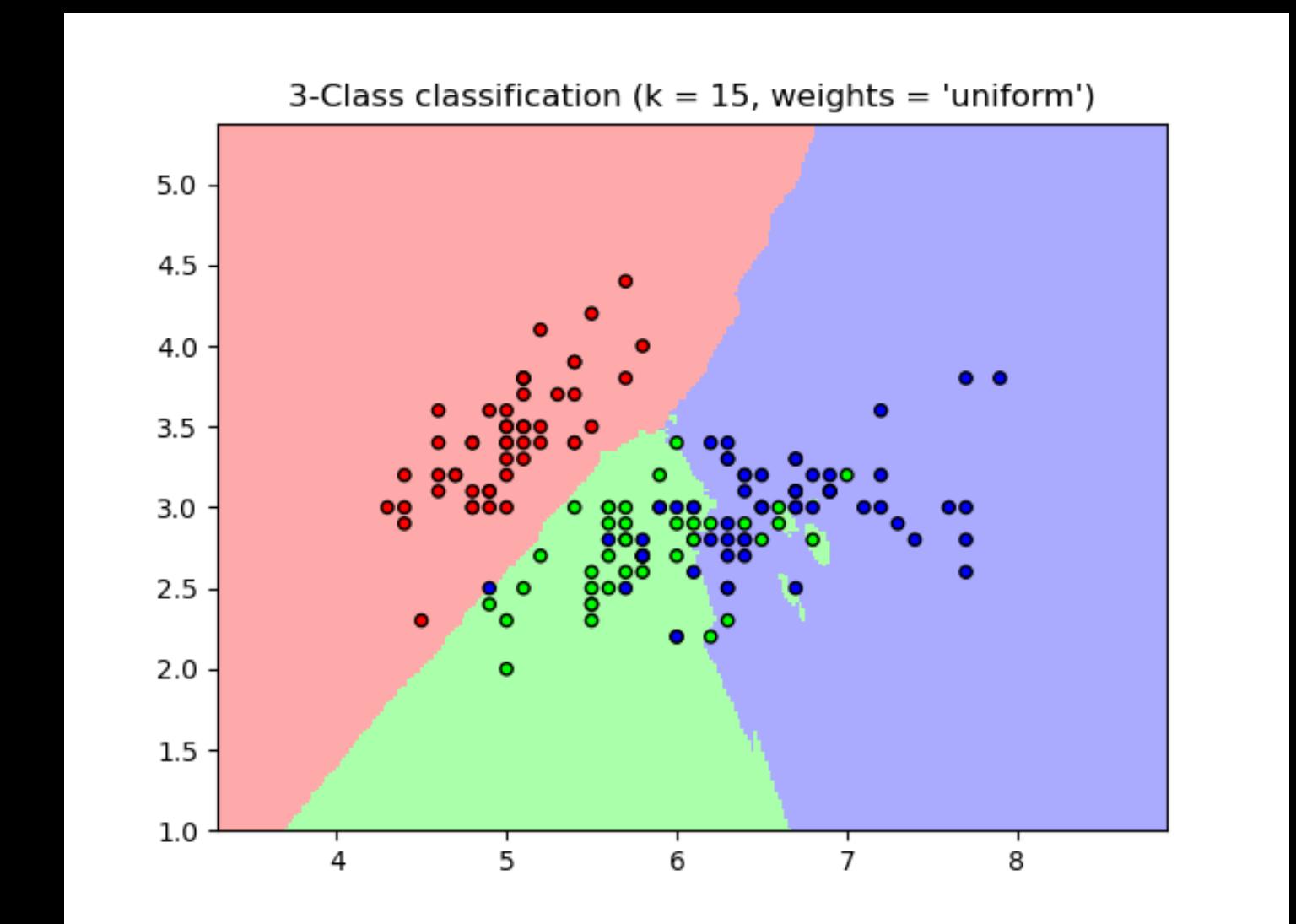
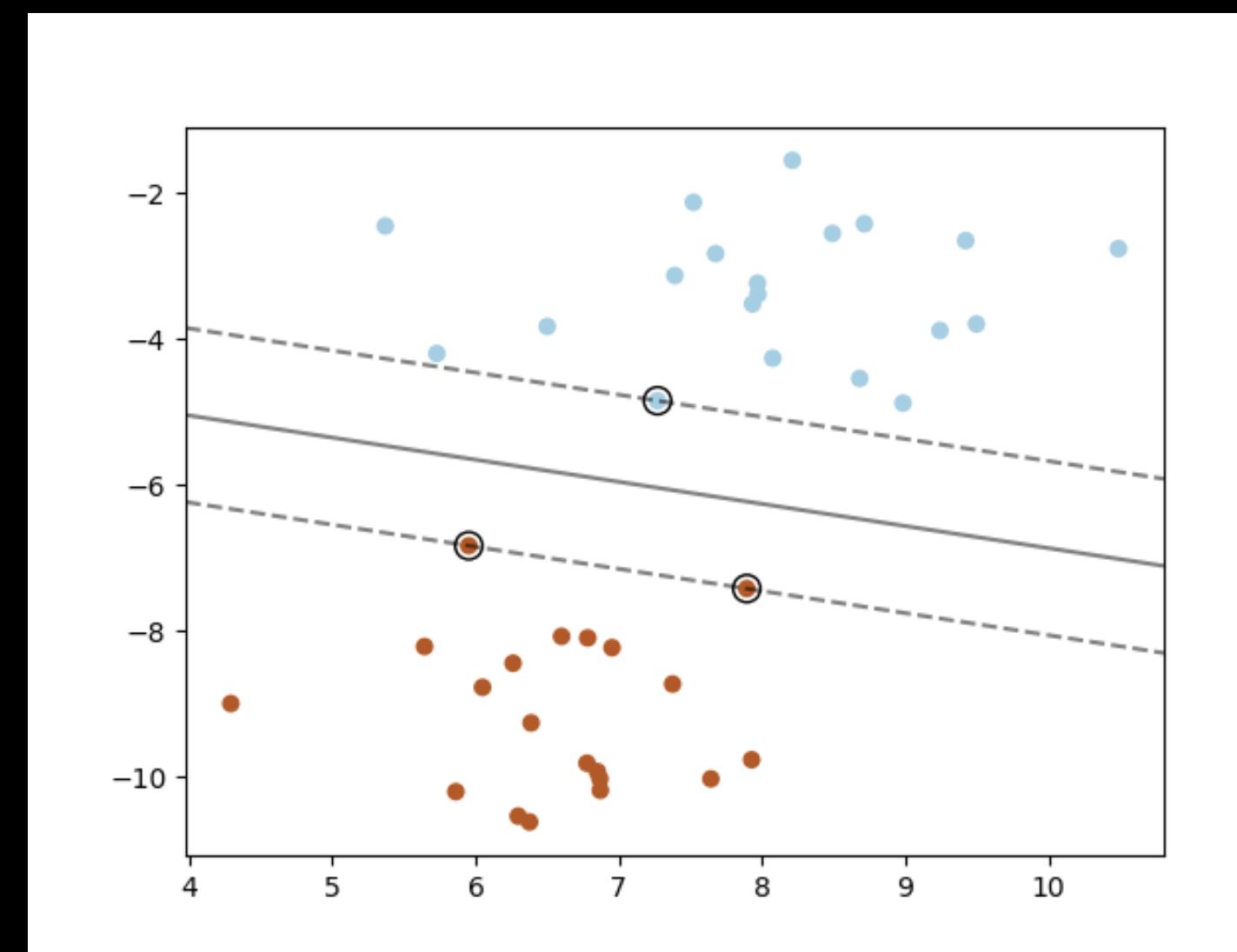
Are movie recommendation systems consistent?



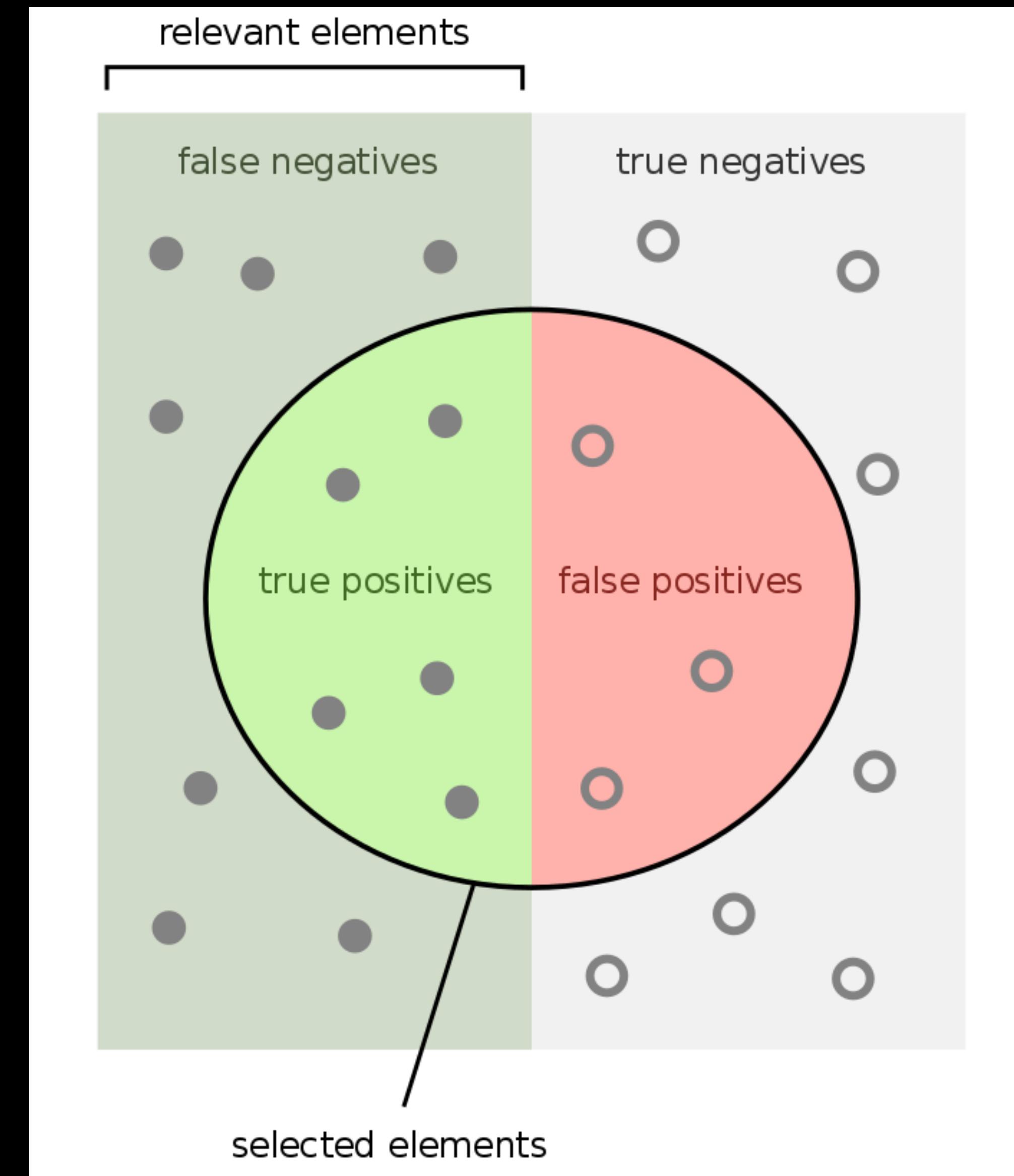
(Some) Classification Models



Logistic Regression



Model Selection and Validation in Classification



Model Validation in Classification

Precision and Recall

Precision

$$P = \frac{T_p}{T_p + F_p}$$

T_p **True Positive:** “I predicted survival and s/he survived”

F_p **False Positive:** “I predicted survival and s/he did not survived”

F_n **False Negative:** “I predicted death and s/he did survived”

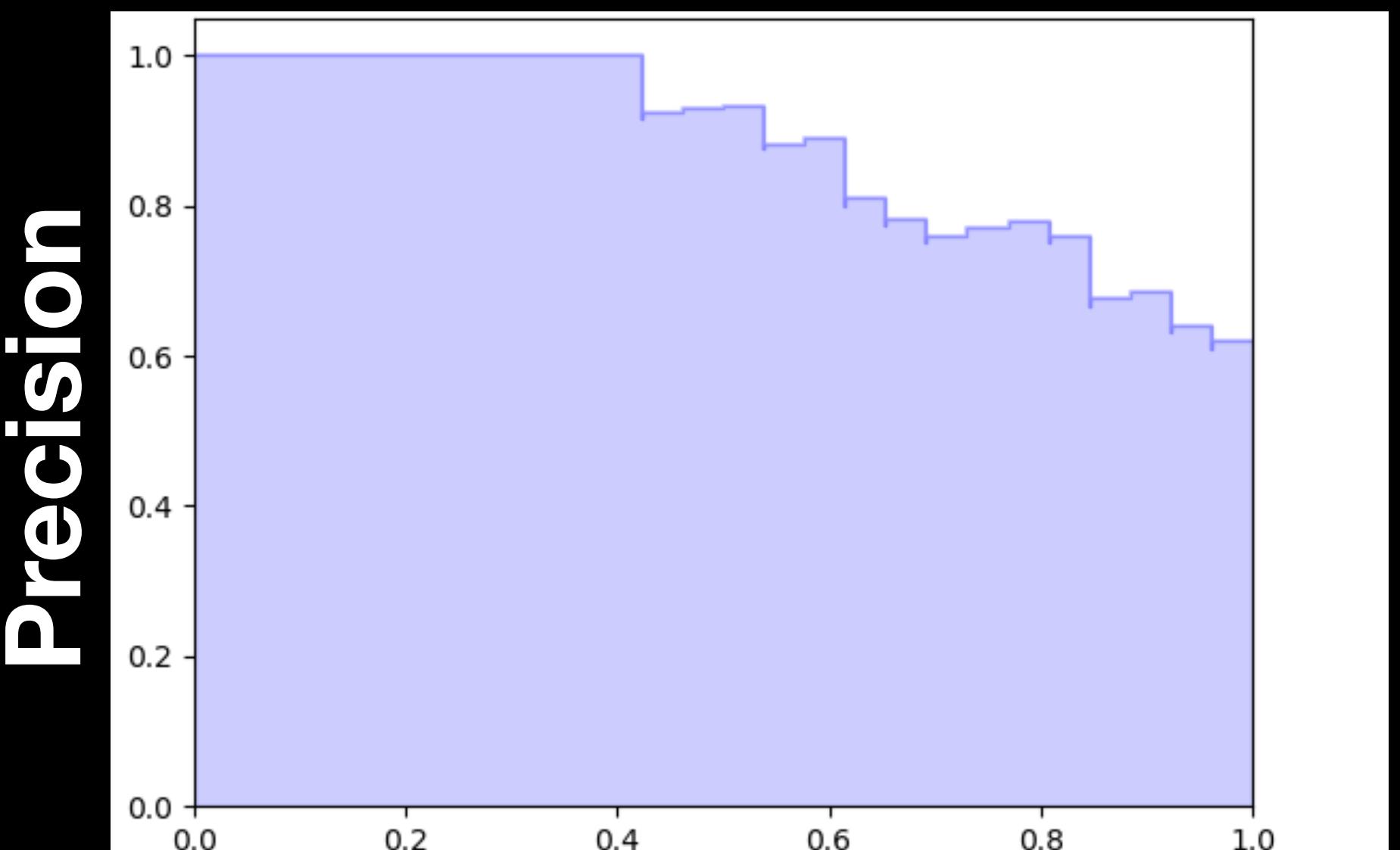
Recall

$$R = \frac{T_p}{T_p + F_n}$$

Score

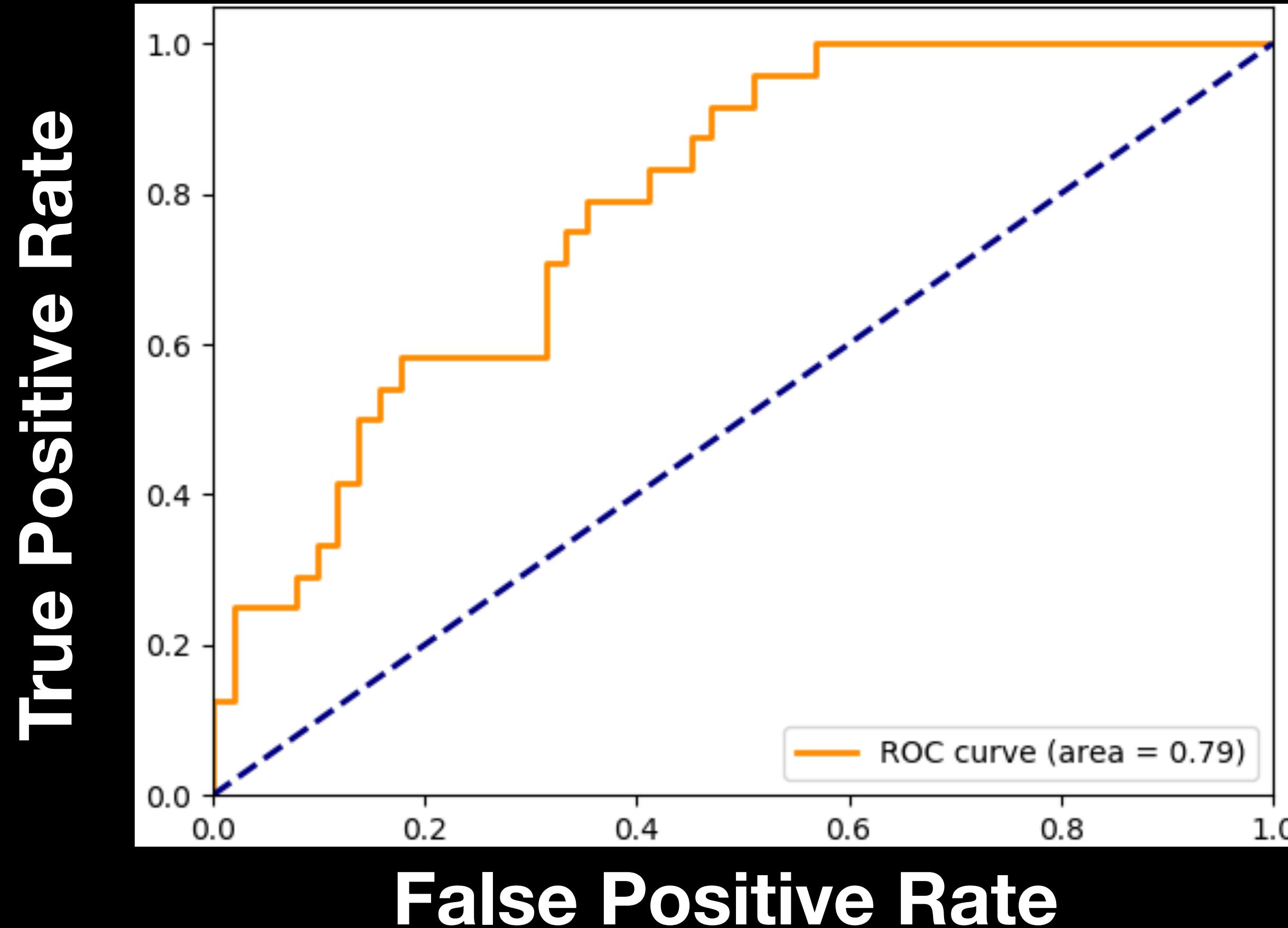
$$F_1 = 2 \frac{P \times R}{P + R}$$

Recall



Model Validation in Classification

Receiver Operator Characteristic

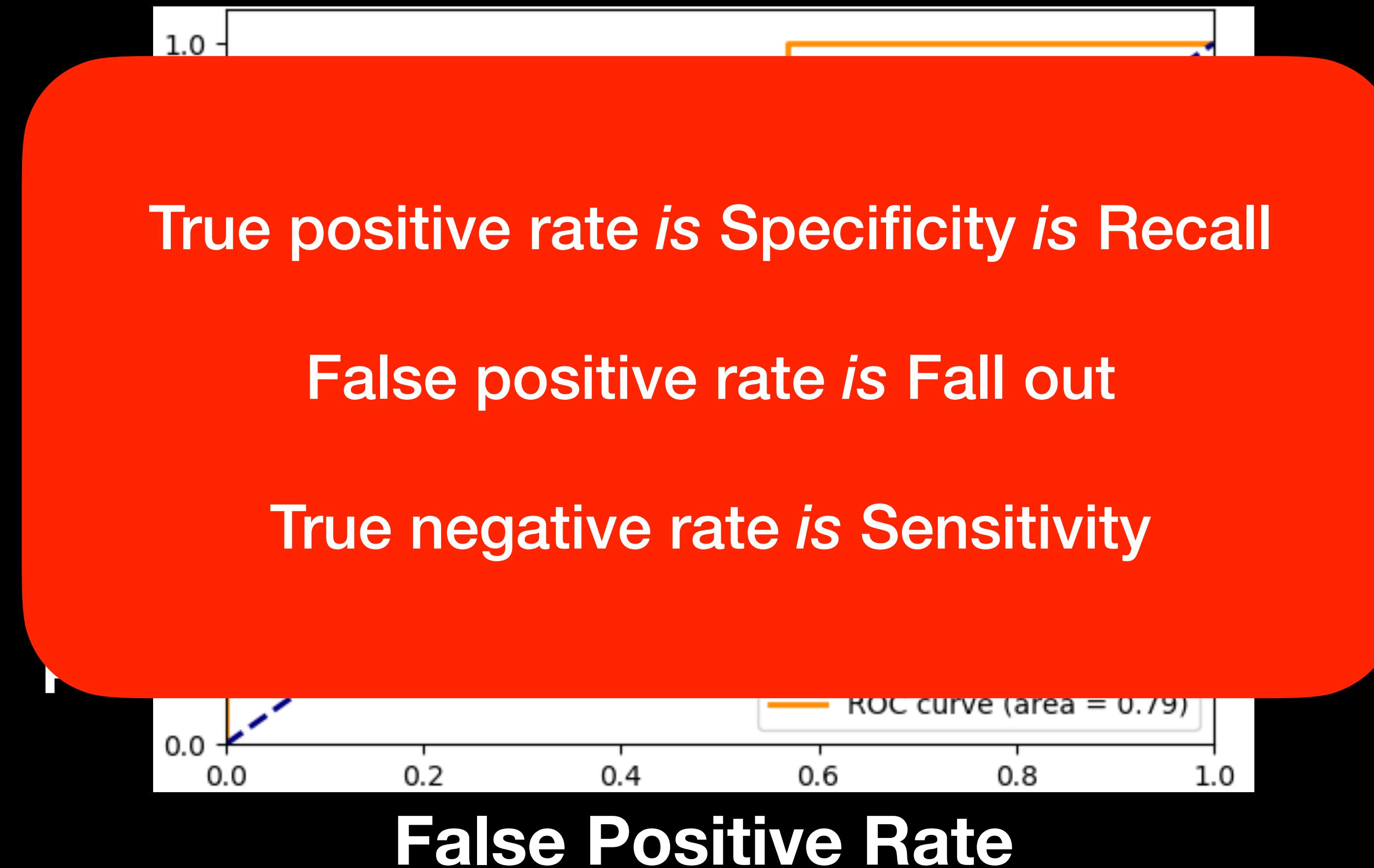


T_p **True Positive:** “I predicted survival and s/he survived”

F_p **False Positive:** “I predicted survival and s/he did not survived”

Model Validation in Classification

Receiver Operator Characteristic

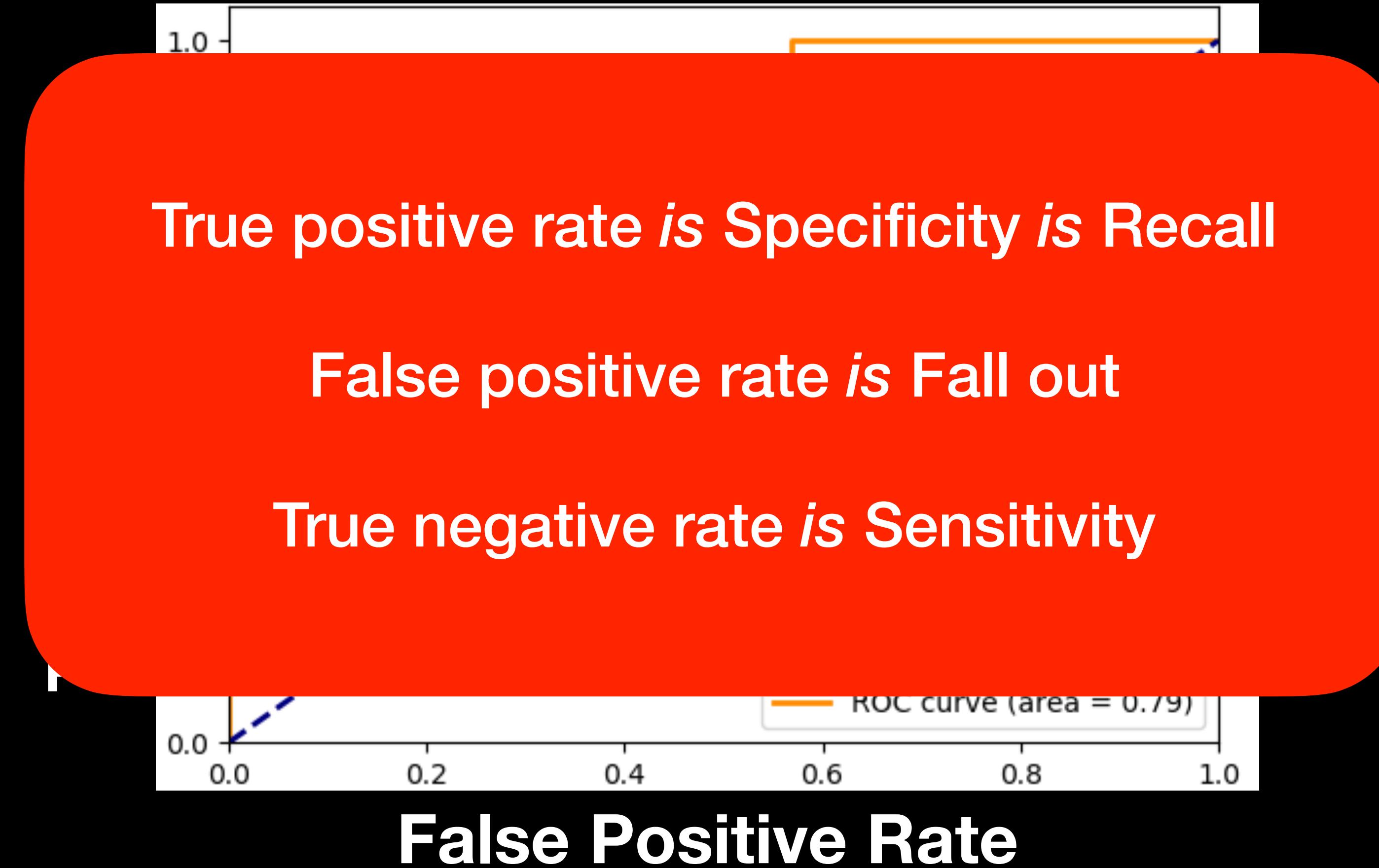
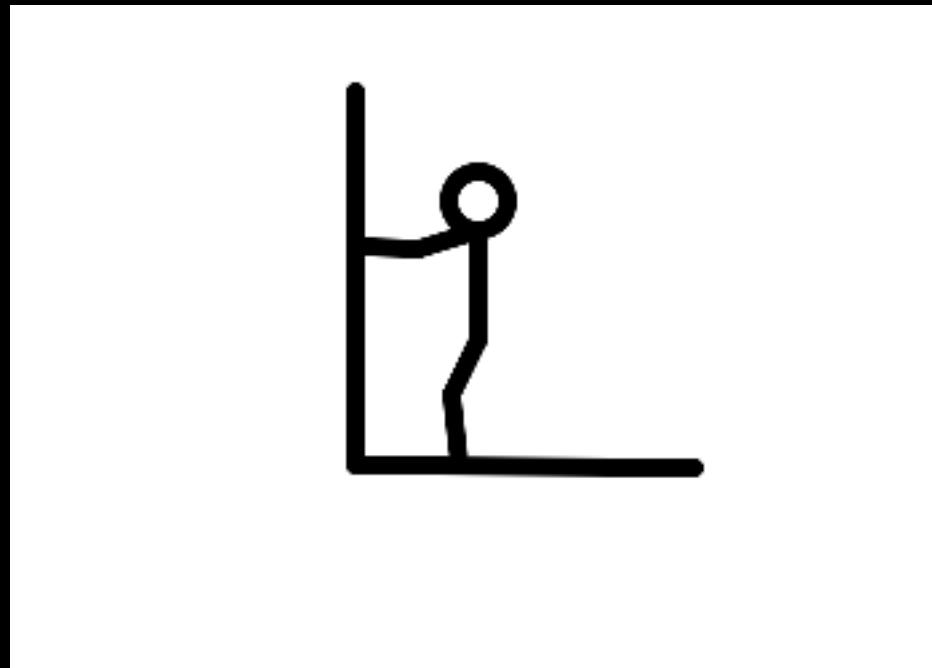


T_p **True Positive:** “I predicted survival and s/he survived”

F_p **False Positive:** “I predicted survival and s/he did not survived”

Model Validation in Classification

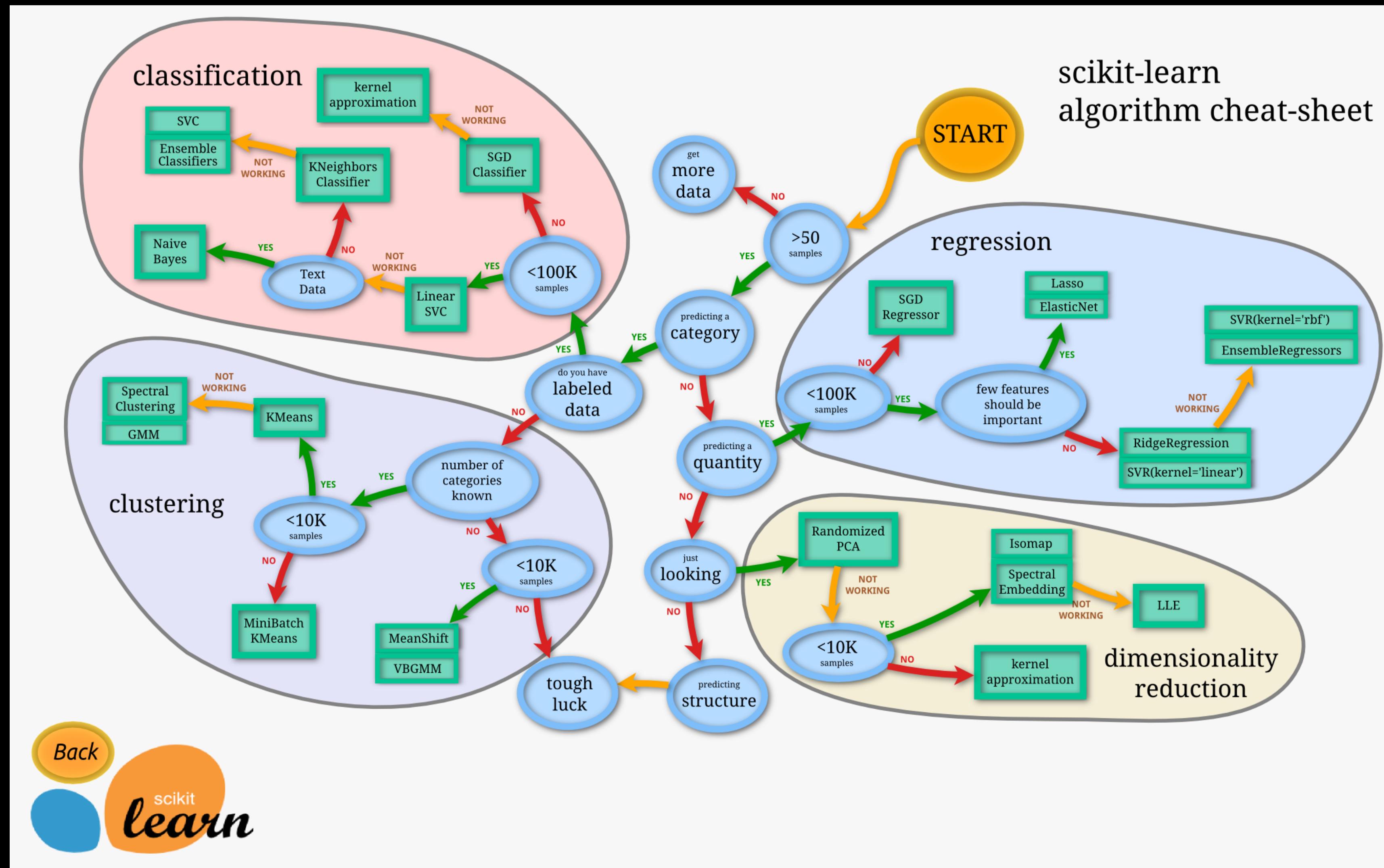
Receiver Operator Characteristic



T_p **True Positive:** “I predicted survival and s/he survived”

F_p **False Positive:** “I predicted survival and s/he did not survived”

Where are we in the Data Science Landscape?



Thanks!