



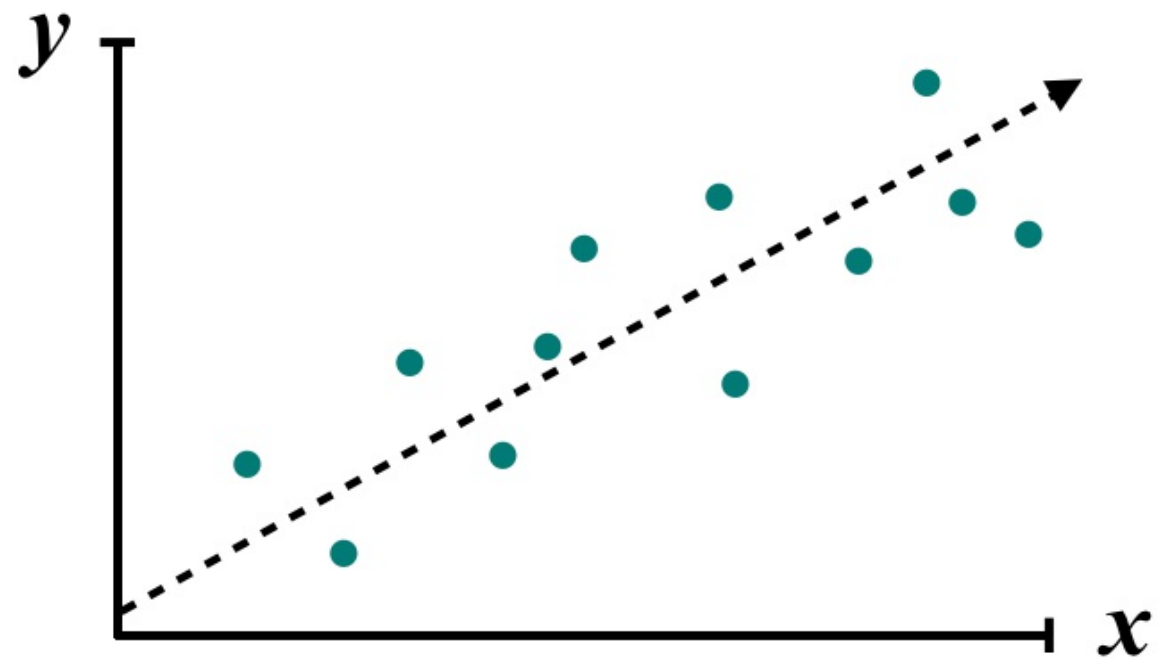
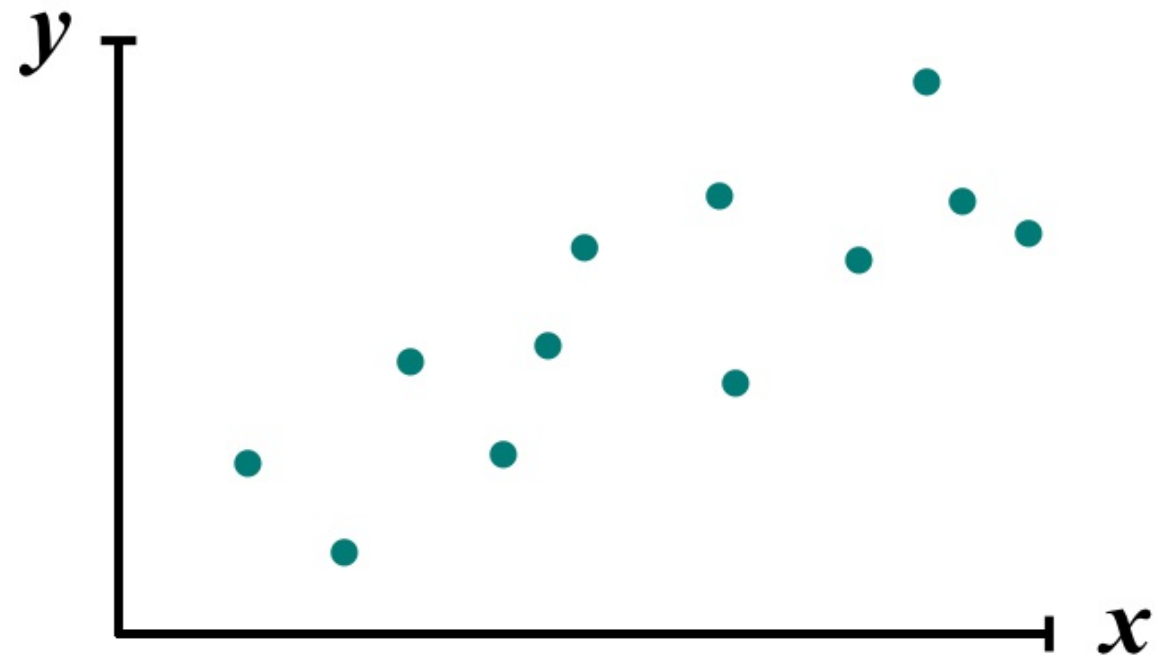
## SUPERVISED LEARNING IN R: CLASSIFICATION

# Making binary predictions with regression

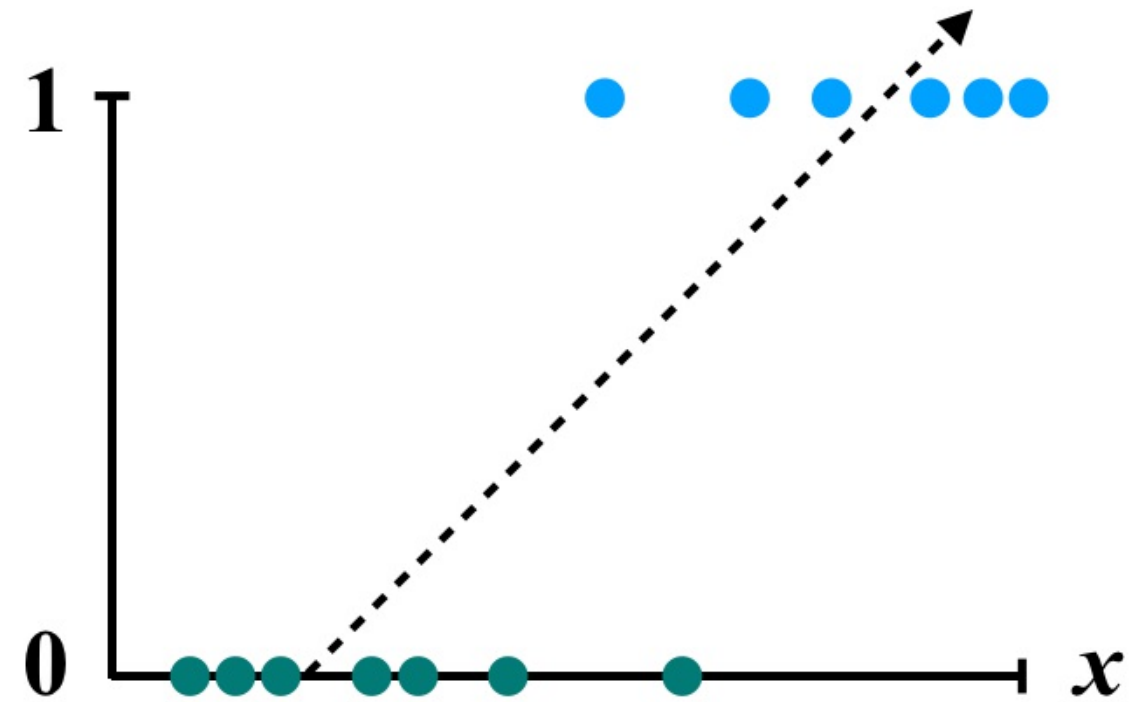
Brett Lantz  
Instructor



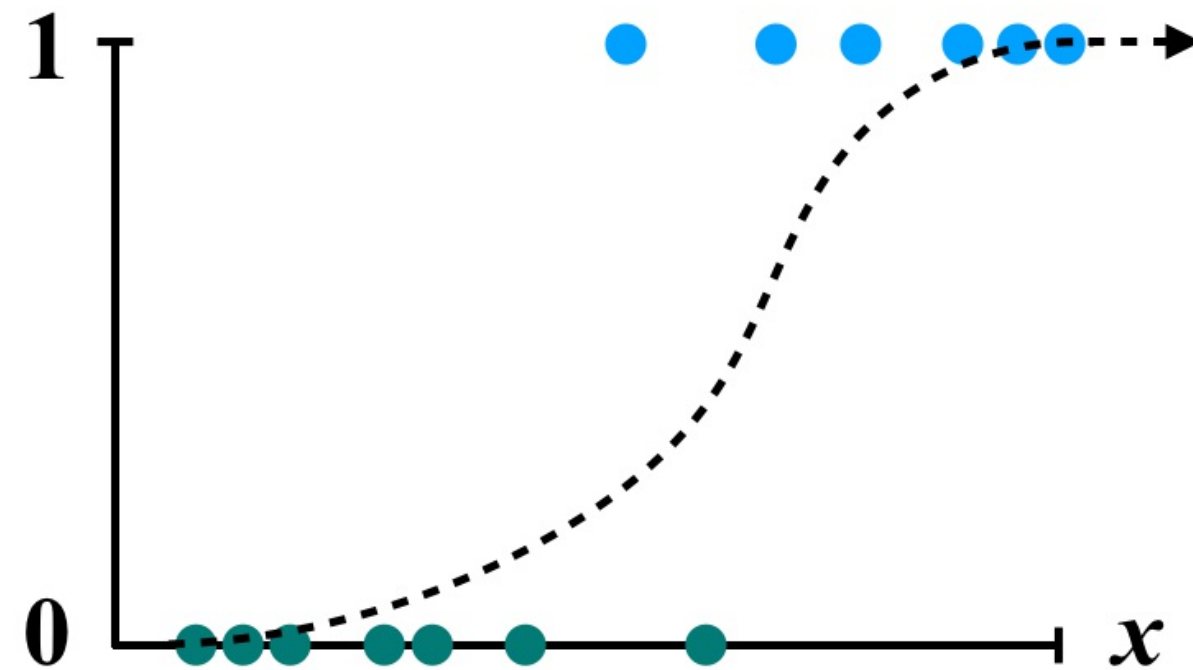
# Introducing linear regression



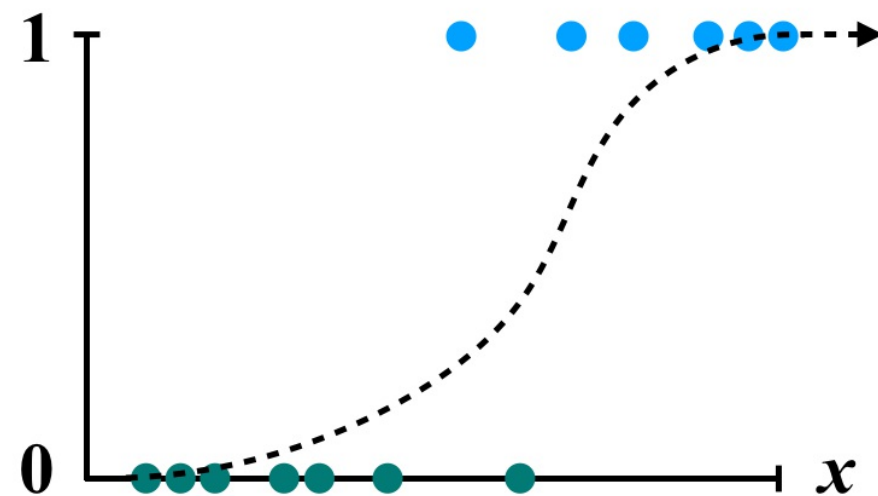
# Regression for binary classification



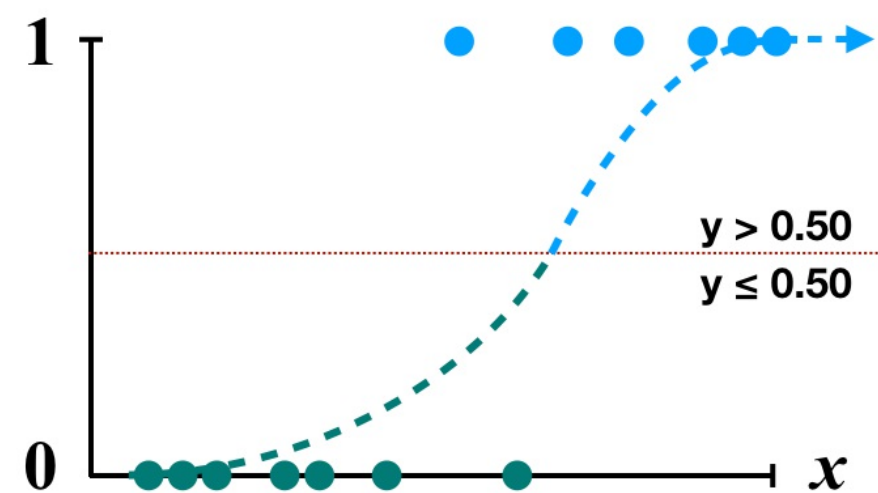
# Introducing logistic regression



# Making predictions with logistic regression



```
m <- glm(y ~ x1 + x2 + x3,
          data = my_dataset,
          family = "binomial")
```



```
prob <- predict(m, test_dataset,
                type = "response")
```

```
pred <- ifelse(prob > 0.50, 1, 0)
```



## SUPERVISED LEARNING IN R: CLASSIFICATION

**Let's practice!**

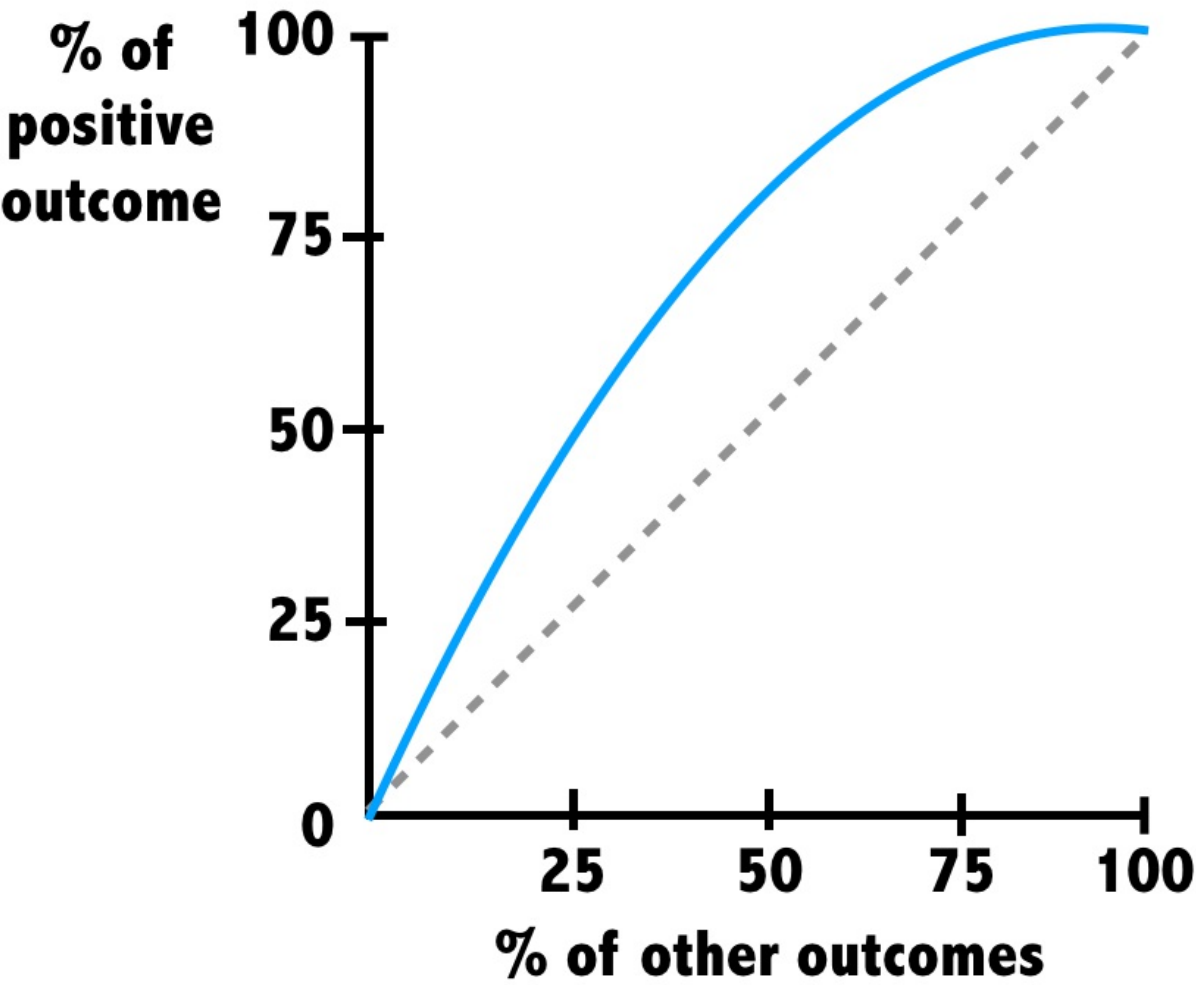


SUPERVISED LEARNING IN R: CLASSIFICATION

# Model performance tradeoffs

Brett Lantz  
Instructor

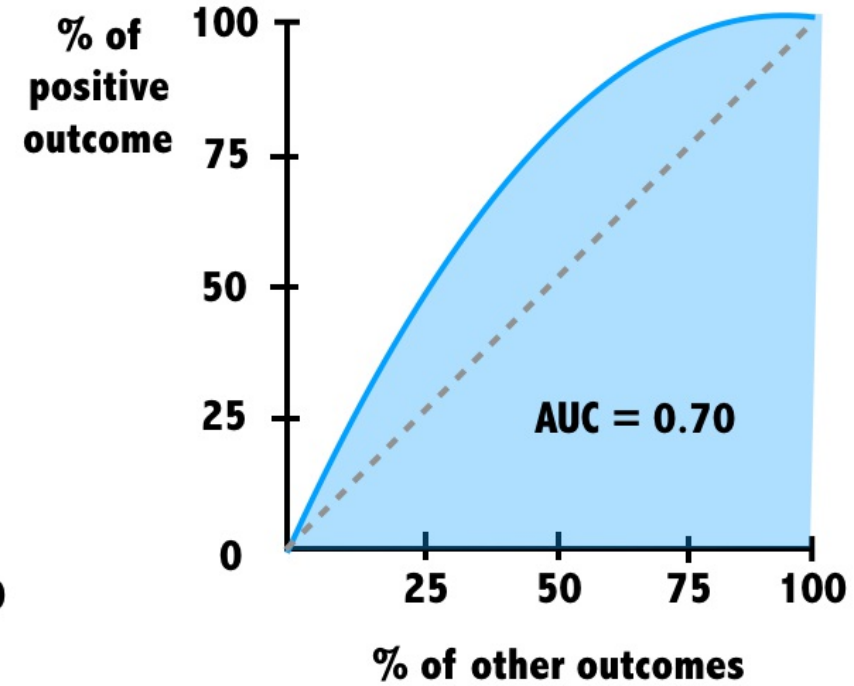
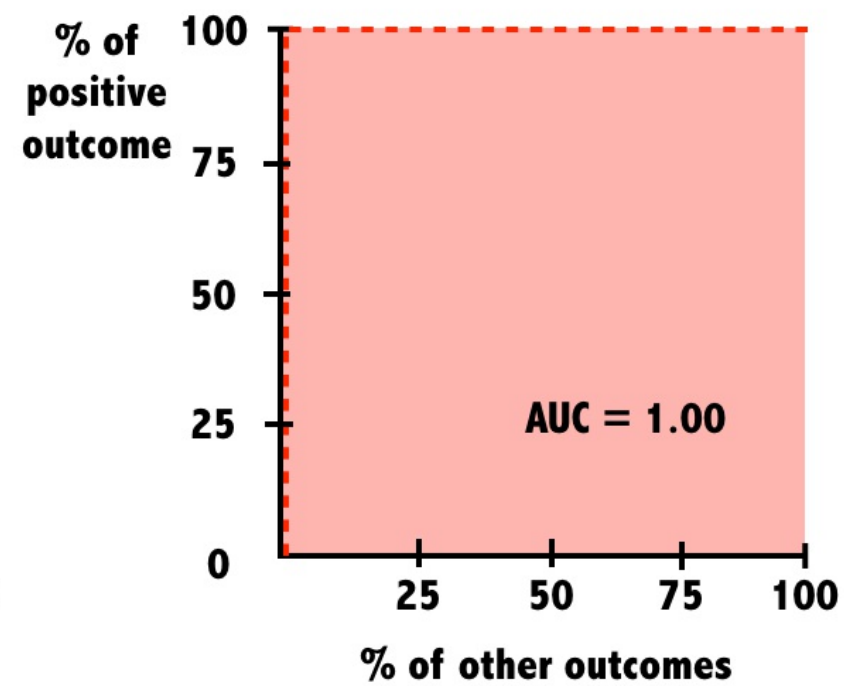
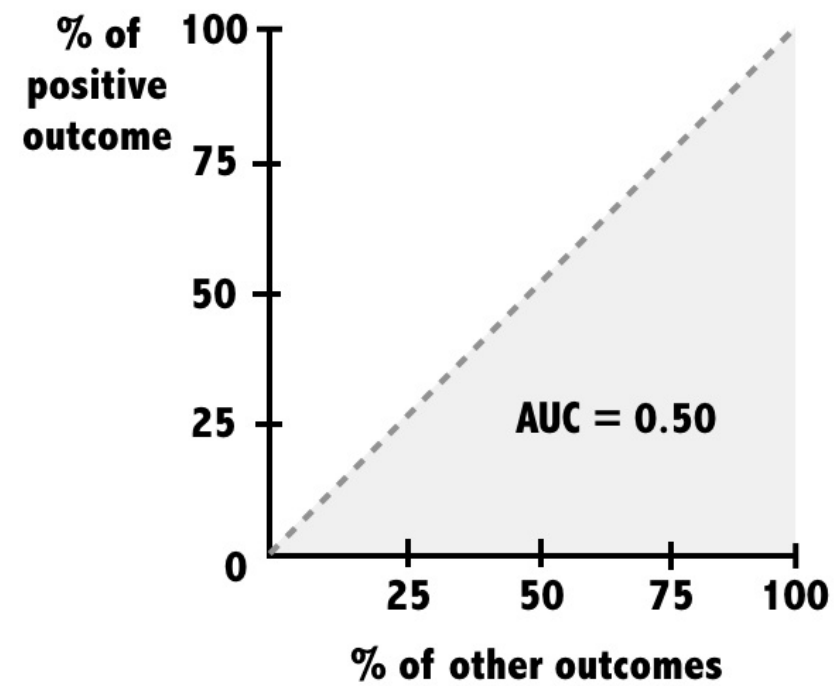
# Understanding ROC curves





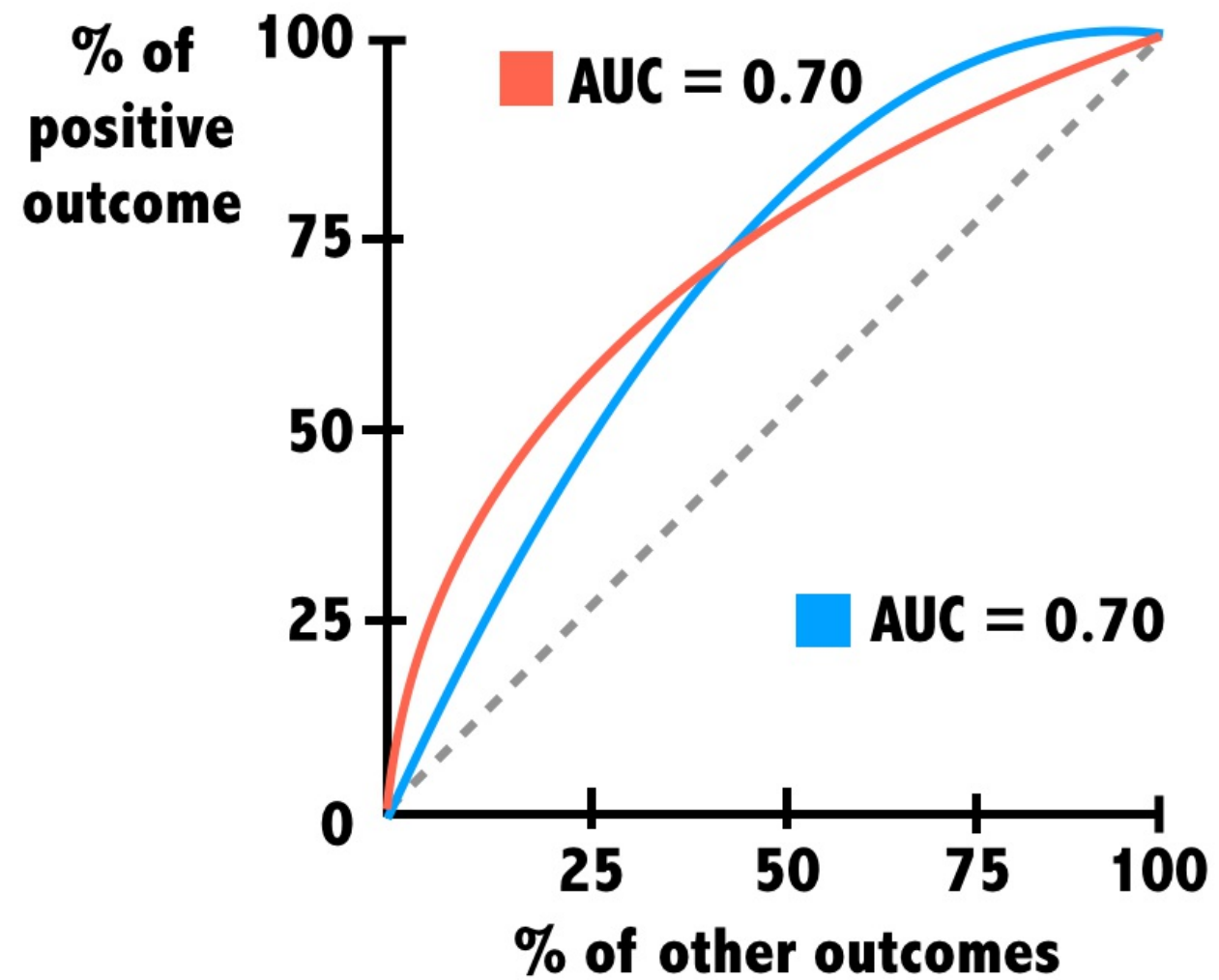


# Area under the ROC curve





# Using AUC and ROC appropriately





## SUPERVISED LEARNING IN R: CLASSIFICATION

**Let's practice!**



## SUPERVISED LEARNING IN R: CLASSIFICATION

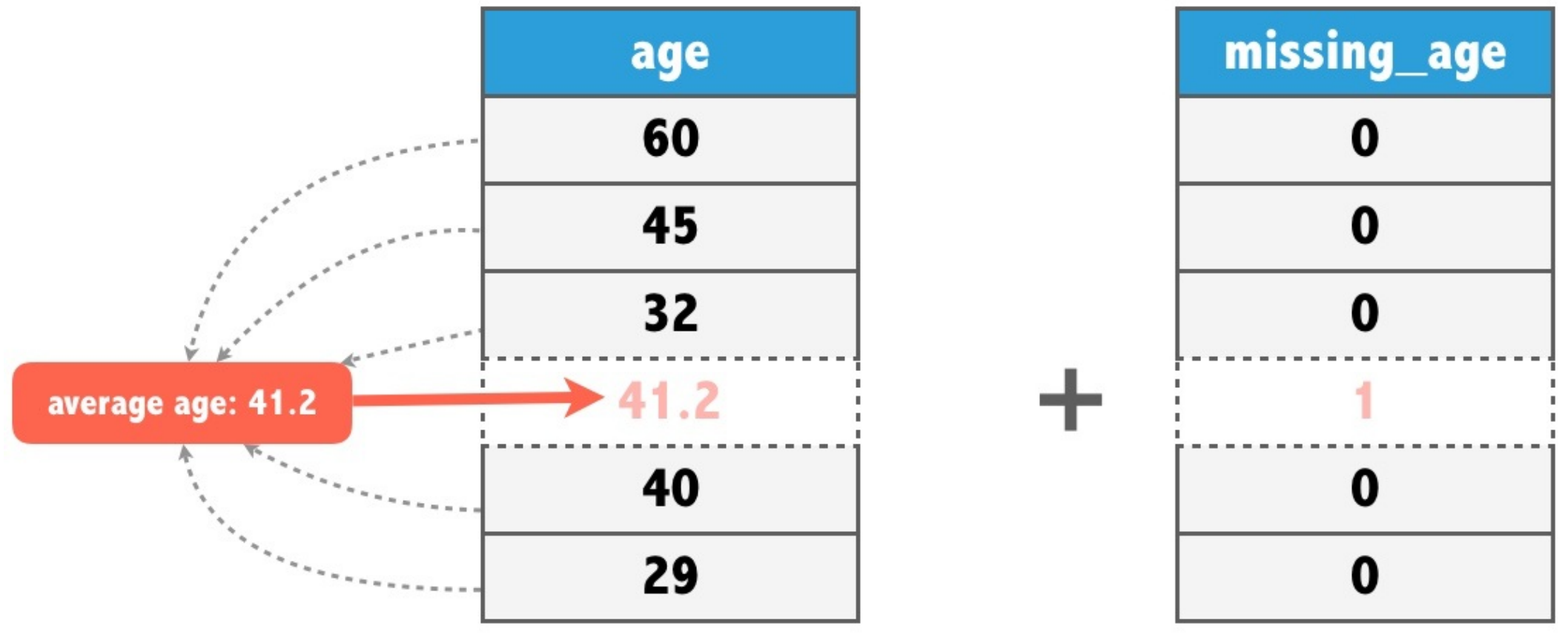
# **Dummy variables, missing data, and interactions**

Brett Lantz  
Instructor



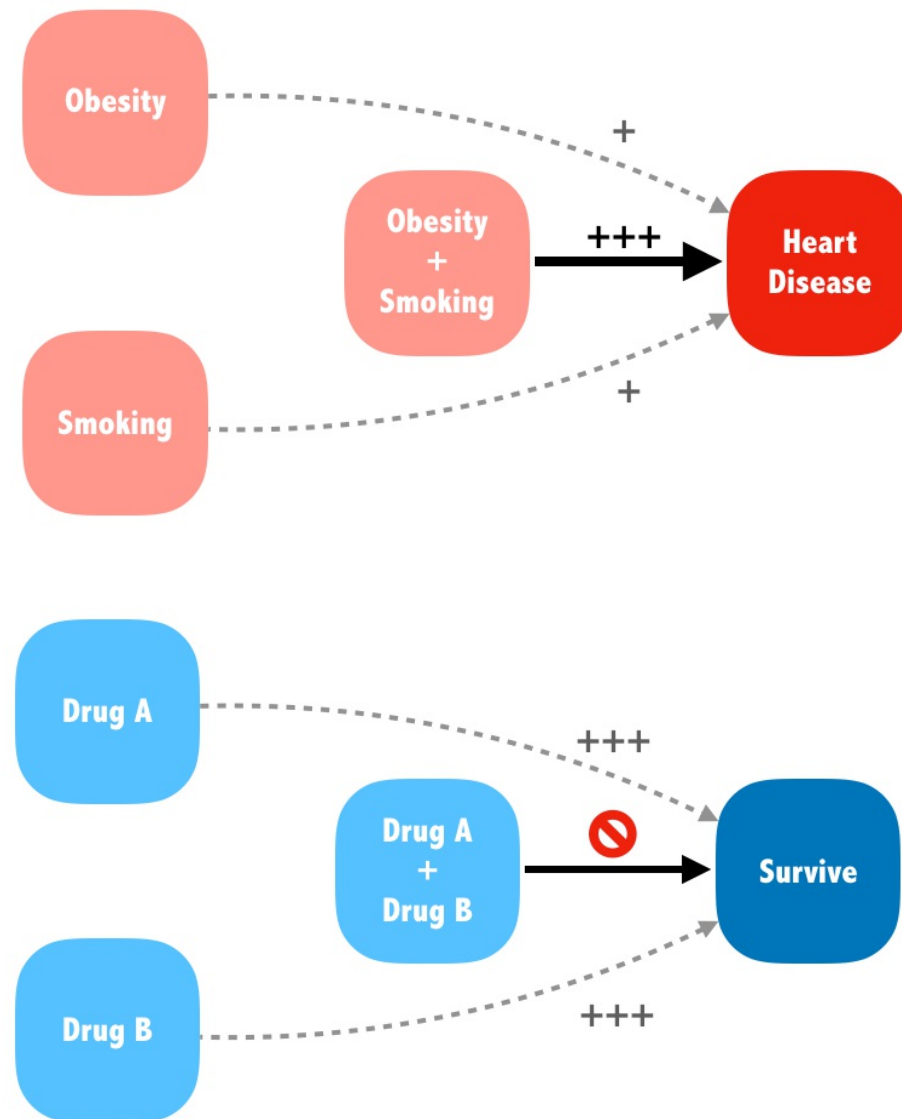


# Imputing missing data





# Interaction effects



```
# interaction of obesity and smoking
glm(disease ~ obesity * smoking,
    data = health,
    family = "binomial")
```



## SUPERVISED LEARNING IN R: CLASSIFICATION

**Let's practice!**





## SUPERVISED LEARNING IN R: CLASSIFICATION

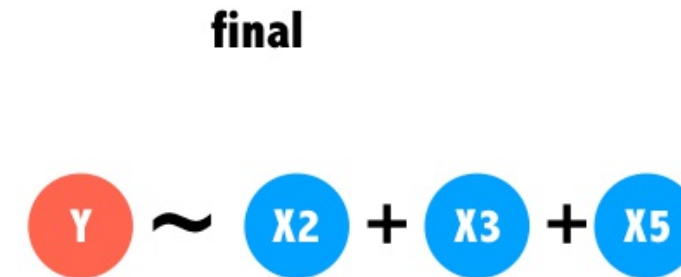
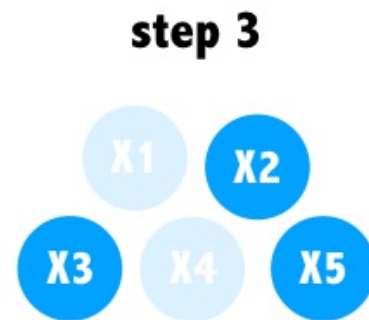
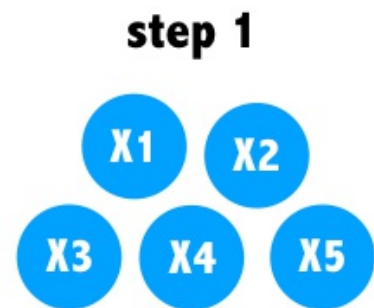
# Automatic feature selection

Brett Lantz  
Instructor

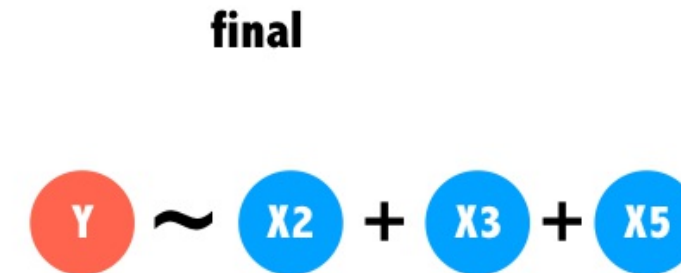
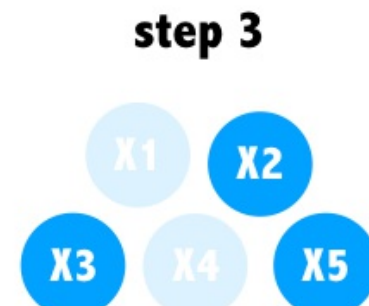


# Stepwise regression

**backward  
stepwise**

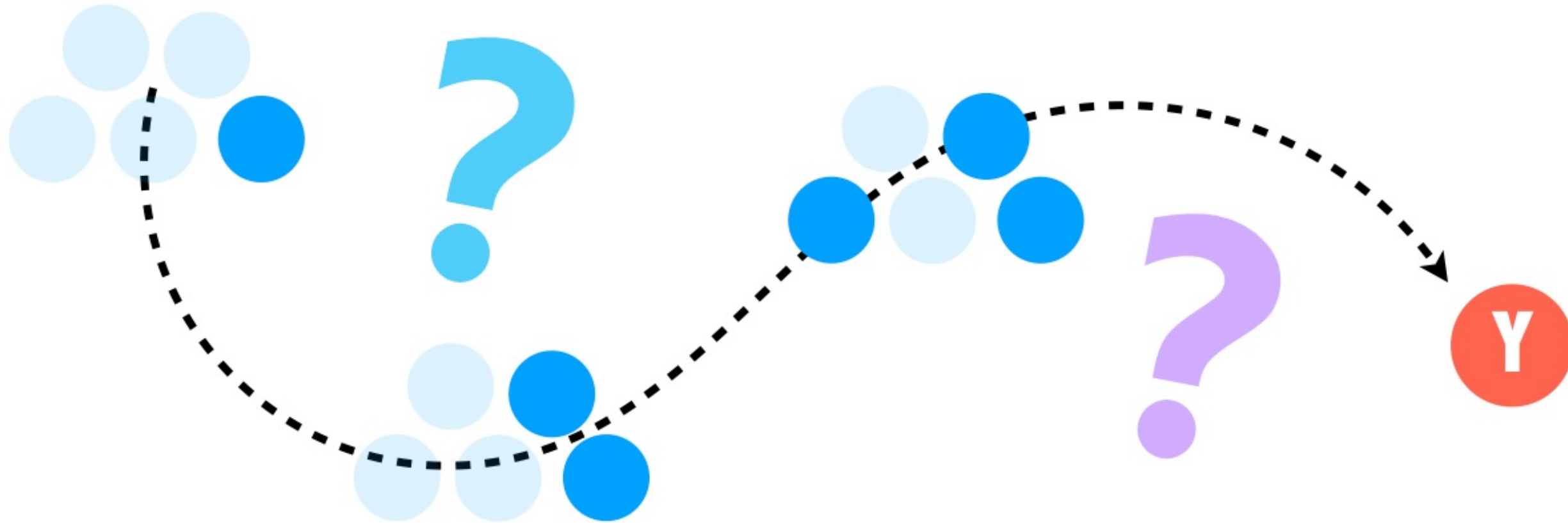


**forward  
stepwise**





# Stepwise regression caveats





## SUPERVISED LEARNING IN R: CLASSIFICATION

**Let's practice!**