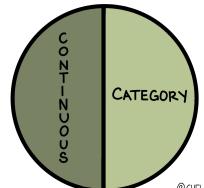
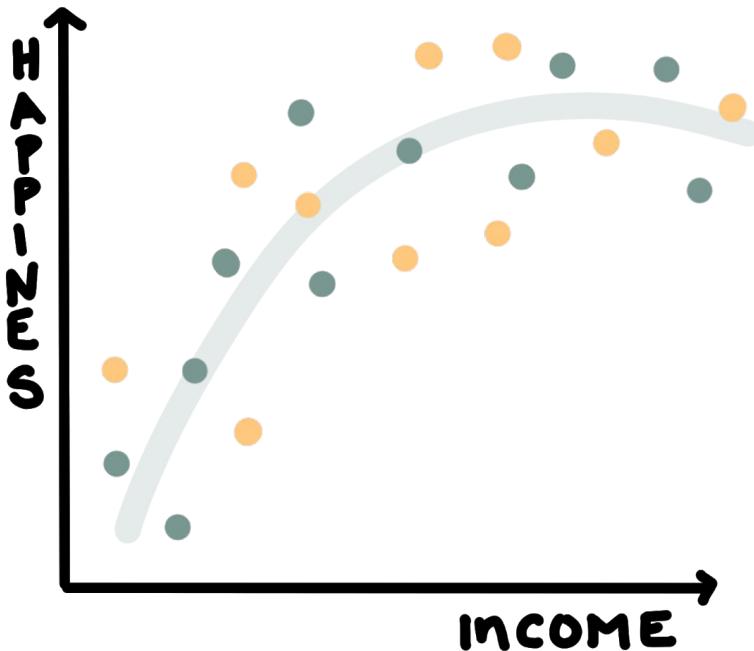


PREDICT



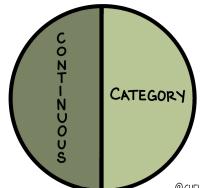
Bias and Variance



- data we have now
- future data

@CHELSEAPARLETT

PREDICT



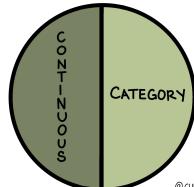
Bias and Variance



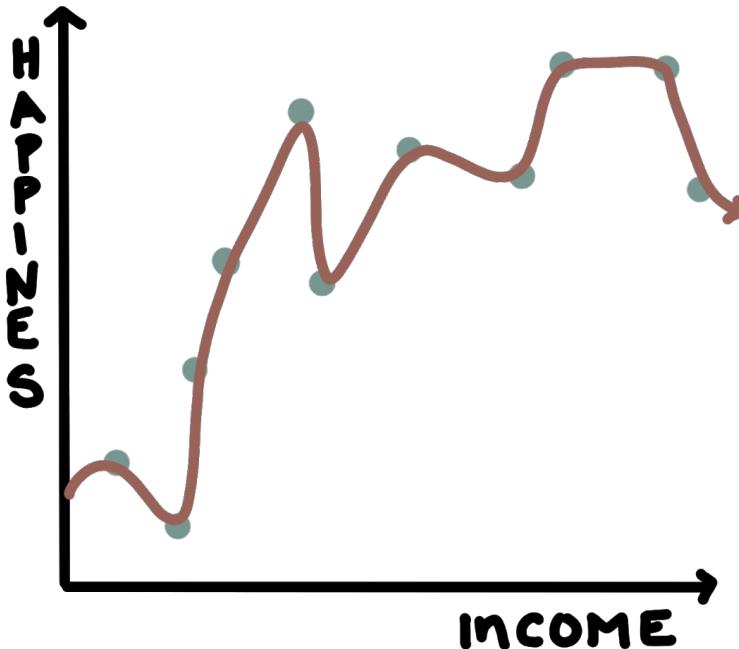
- data we have now
- future data

@CHELSEAPARLETT

PREDICT



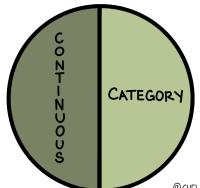
Bias and Variance



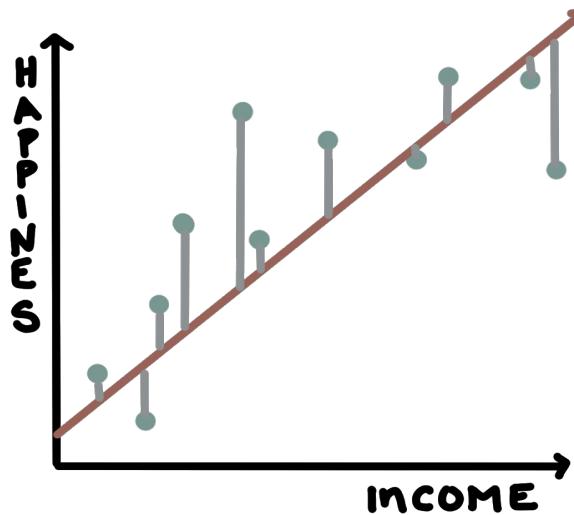
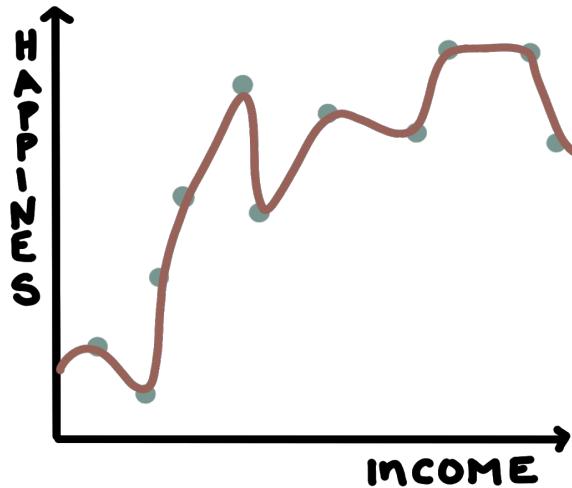
- data we have now
- future data

@CHELSEAPARLETT

PREDICT

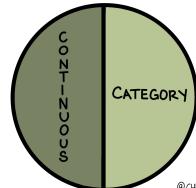


Bias and Variance

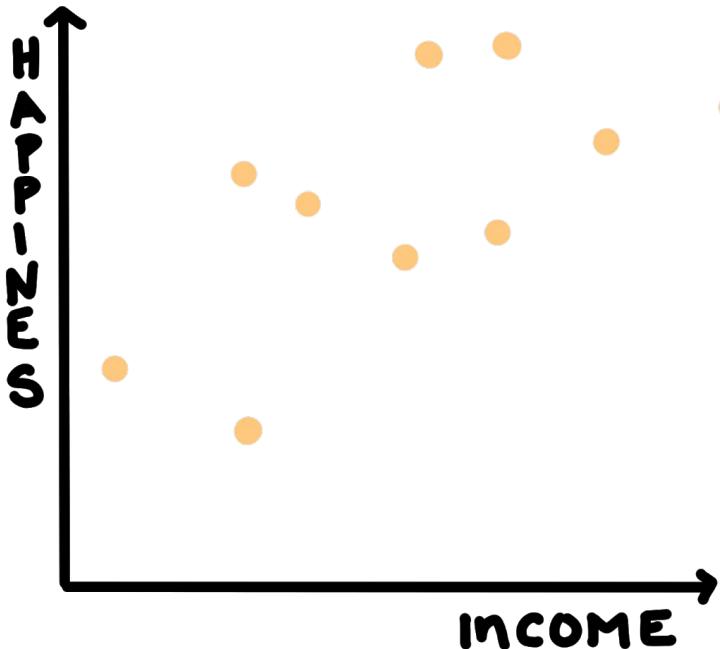


- data we have now
- future data

PREDICT



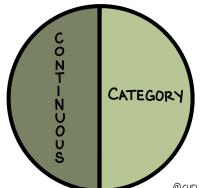
Bias and Variance



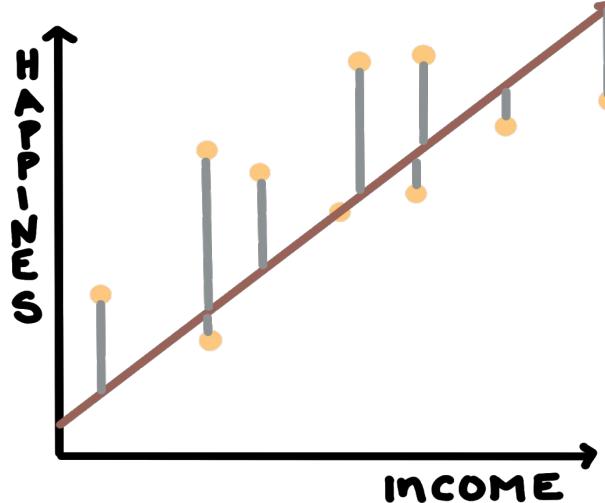
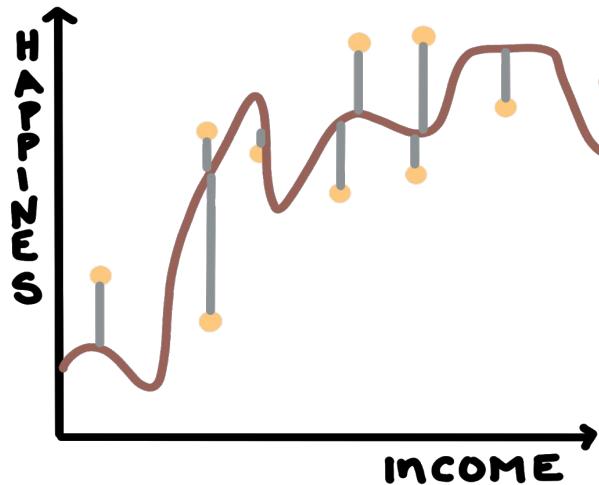
- data we have now
- future data

@CHELSEAPARLETT

PREDICT

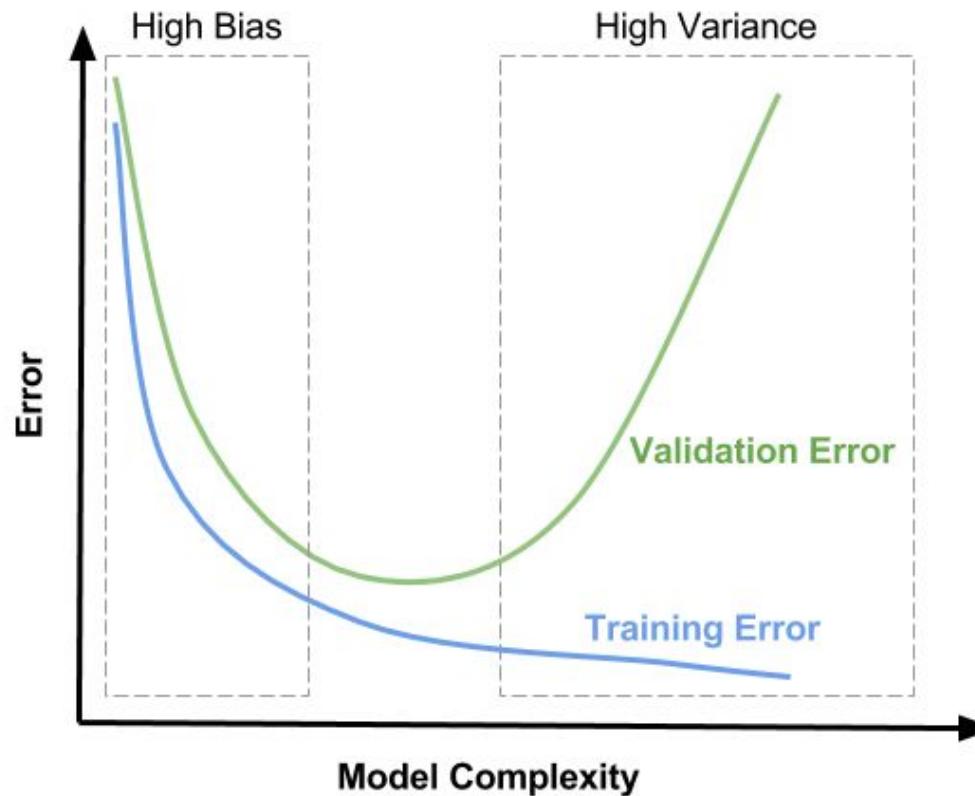


Bias and Variance

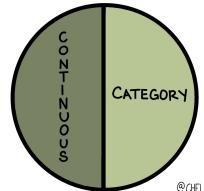


- data we have now
- future data

Bias and Variance



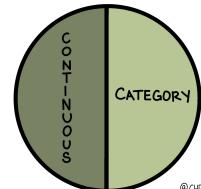
PREDICT



Cross Validation

1. **Split data** into train and test, build model on train, test on test
2. Use performance metrics on test data (or average over multiple) to **evaluate model performance**

PREDICT



Cross Validation

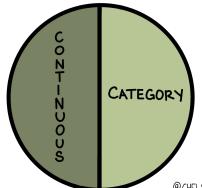
DATA

■ train
■ test



@CHESEAPARLETT

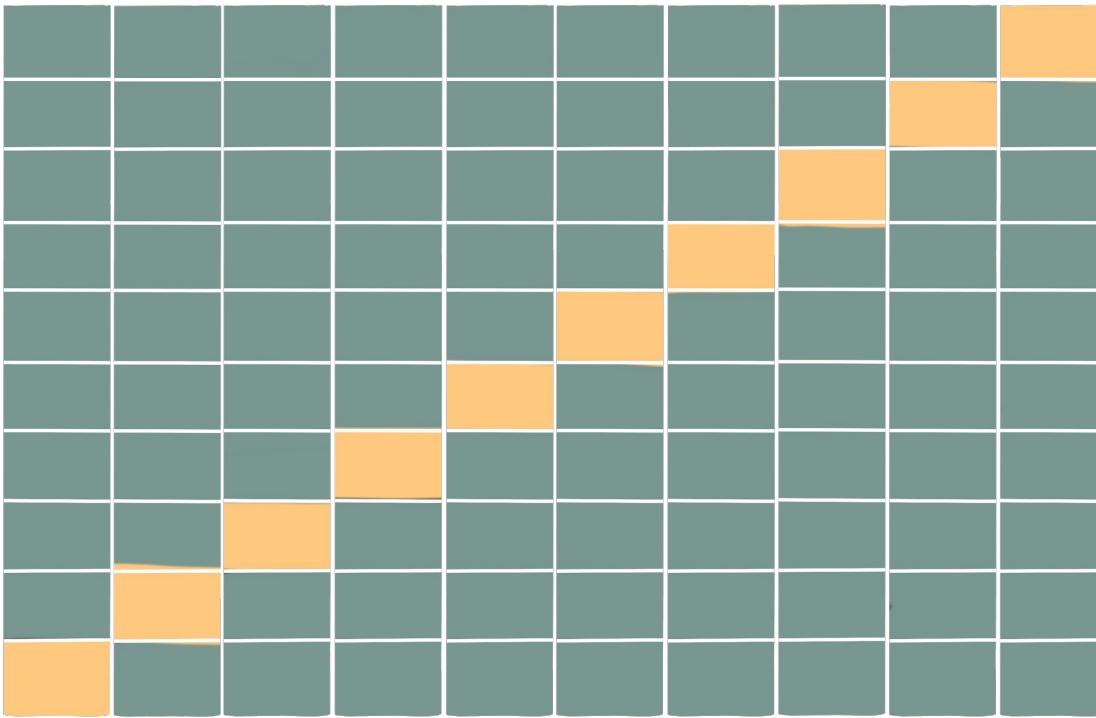
PREDICT



Cross Validation

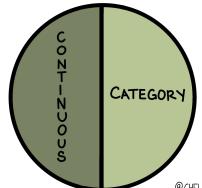
DATA

■ train
■ test



@CHESEAPARLETT

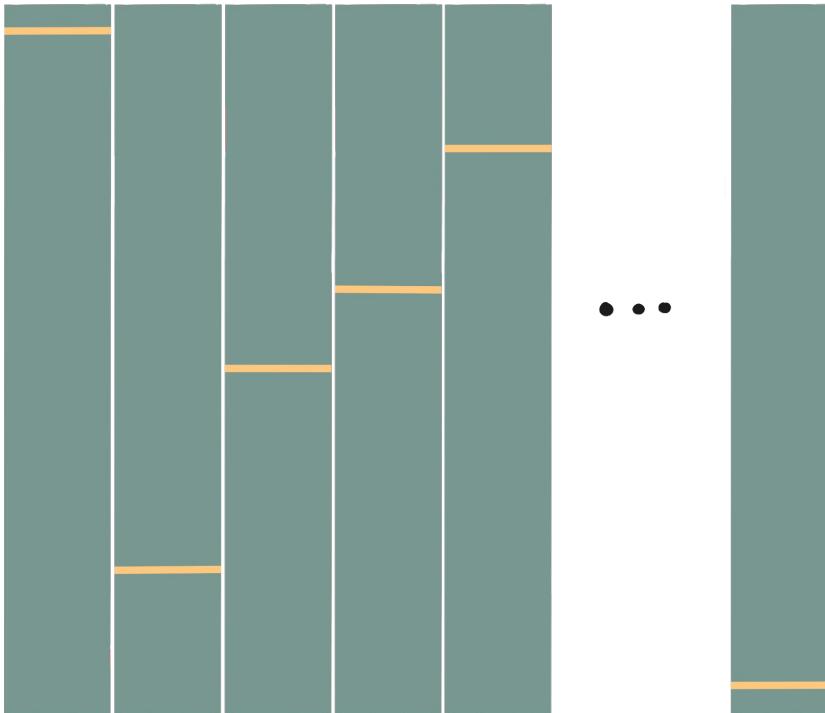
PREDICT



Cross Validation

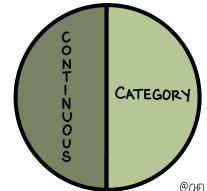
DATA

train
test



@CHESEAPARLETT

PREDICT



Cross Validation

How to decide (Things to Think About)

- Size of your Dataset (rows AND columns)
- Computational Expense