

# Statistics or Machine Learning: What's the Difference?

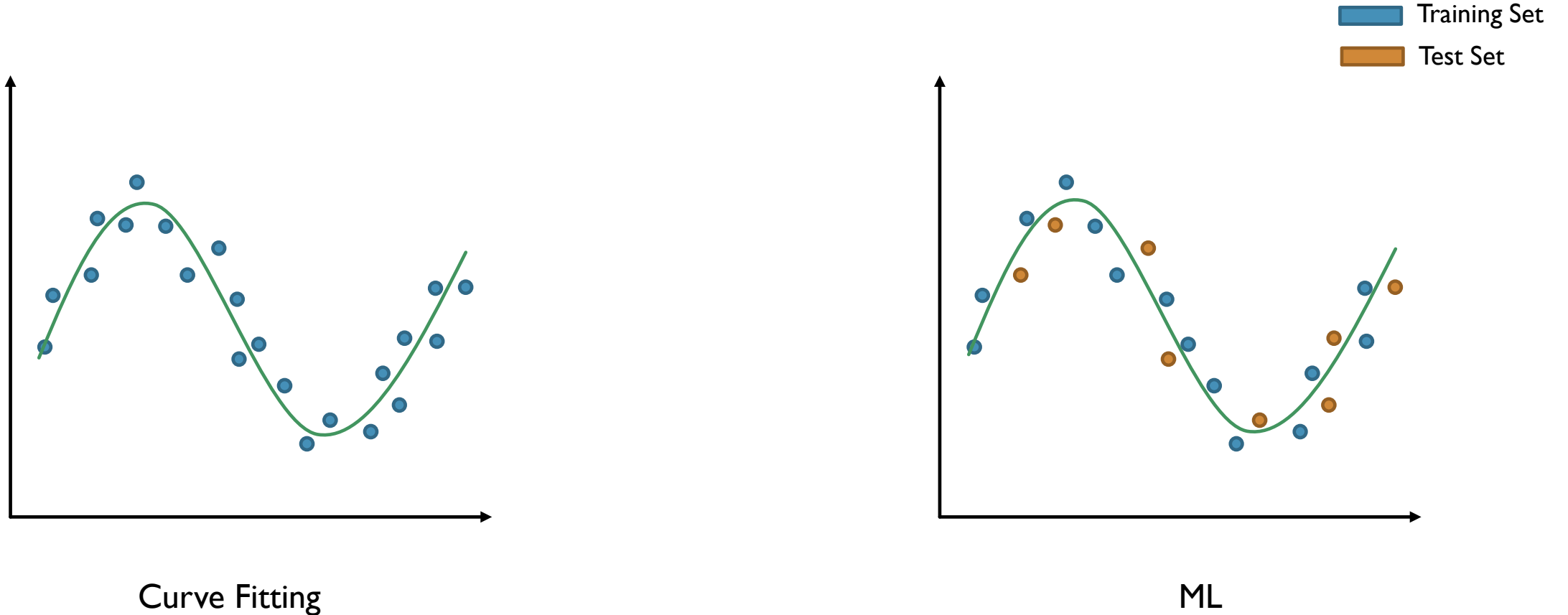
presented by:

Pascal N. Tyrrell

Trisal (Yanlin) Li

Ernest (Khashayar) Namdar

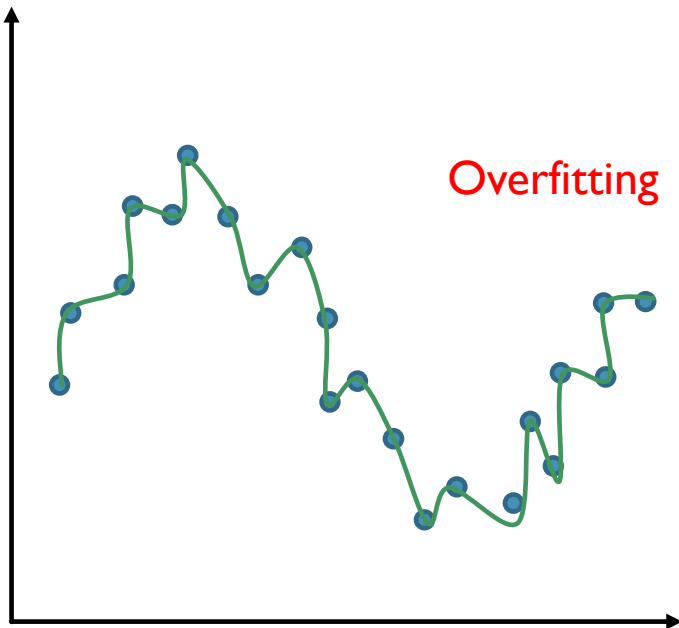
# The Essence of Machine Learning (ML)



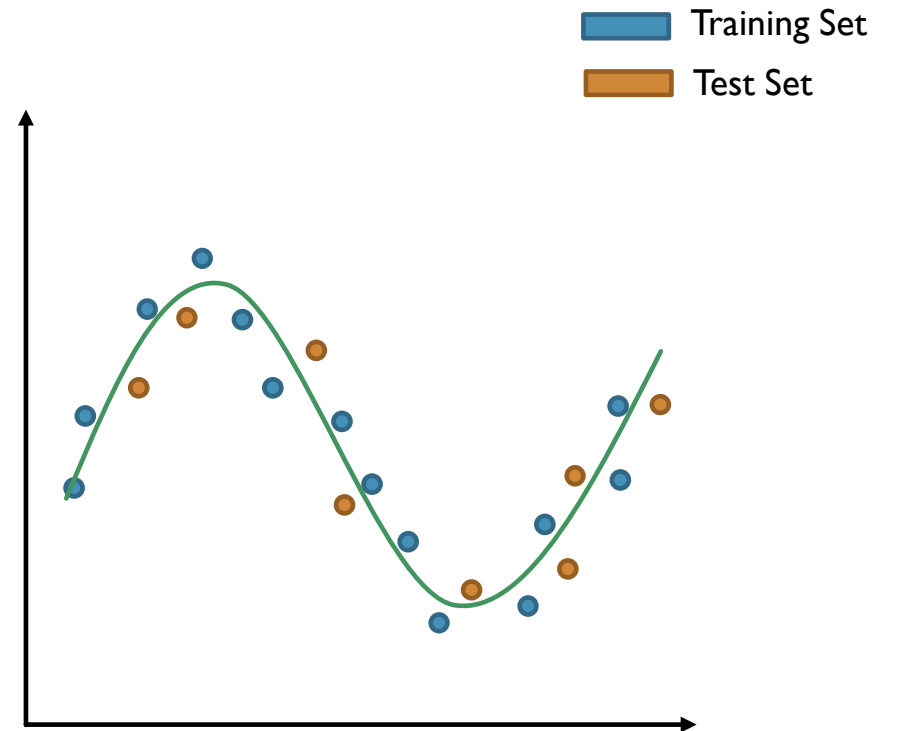
**Making sure the solution is Generalizable**

# The Essence of ML

Keywords: Unseen Data, Generalizability



Curve Fitting



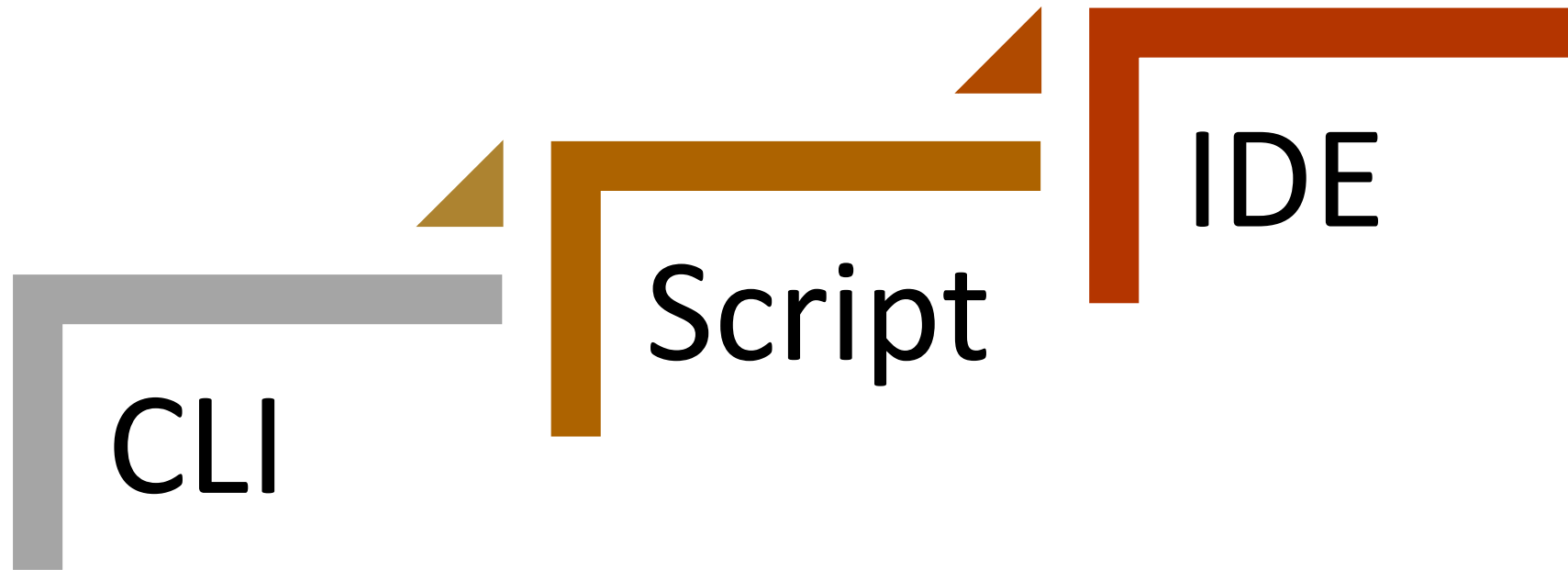
ML

**Making sure the solution is Generalizable**

# Plan for the Day



Warming up for the hands-on section



**How python (&R) can be utilized**

# We will use Google Colab, but why?

Because:

A) Package management can become a nightmare when you use open-source tools, and cloud-based options handle it (as much as possible)

B) Jupyter Notebook is great for classroom

# The Last Point: Dataset

Tabular

155 Examples (observations)

19 Features (variables)

1 Binary Label (outcome variables)

Group	Variables [Variable Number in the Dataset]
Demographics	AGE[2], SEX[3]
Treatment	STEROID[4], ANTIVIRALS[5]
Symptom	FATIGUE[6], MALAISE[7], ANOREXIA[8]
Physical Examination	LIVER BIG[9], LIVER FIRM[10], SPLEEN PALPABLE[11], SPIDERS[12], ASCITES[13], VARICES[14]
Laboratory Test	BILIRUBIN[15], ALK PHOSPHATE[16], SGOT[17], ALBUMIN[18], PROTINE[19], HISTOLOGY[20]

<https://archive.ics.uci.edu/ml/datasets/Hepatitis>

All we have from the past is stats,  
and  
all we can say about the future is probability

Enjoy your moment!

