



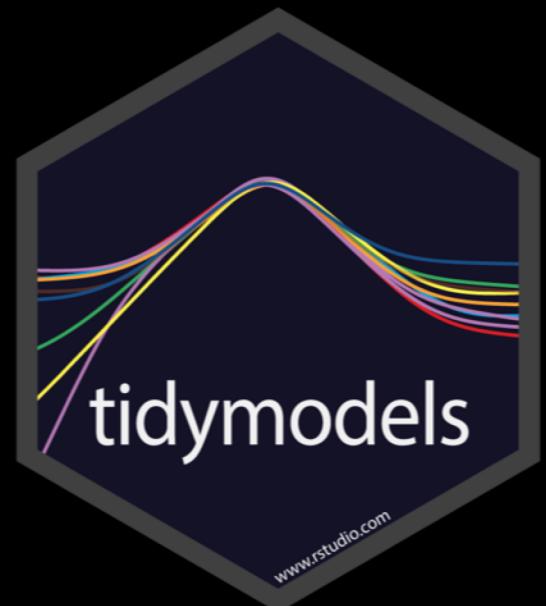
Modelling tidyverse way

Jelena Ilic

Senior Data Scientist

Twitter: @MangoTheCat

Email: info@mango-solutions.com



Agenda

Introduction

Importance of being tidy

tidymodels demo

questions



Introduction



About Mango

- Data Science Services
- Cross Sector
- ~80 people
- London & HQ in Chippenham
- R, Python, Spark



Mango Structure



Data Science
Training
Consulting
App Dev



Data Engineering
Analytic Infrastructure
DevOps for Data
Modernisation



Data Products
App Dev
Validation
Testing



Active in R community

Organise

LondonR

ManchesterR

BaselR

Data engineering London

Bristol Data Scientists

Sponsor

- BarcelonaR
- BirminghamR
- OxfordR

EARL

The Enterprise Applications of the R Language Conference

<https://earlconf.com/>



About Me

- Particle Physicist (> 10 years)
- Insightful (C++)
- Mango (R, Python)
- App Development
- Data Analysis



Importance of being tidy



A photograph of a complex railway junction. The scene is filled with numerous sets of railway tracks of various types, including main lines, sidings, and switchers, all curving and crossing each other in a dense web. Overhead, a complex network of catenary lines and insulators spans the area, supported by tall poles. In the background, there are some industrial buildings and a few people walking on a paved path. The lighting suggests it's either early morning or late afternoon.

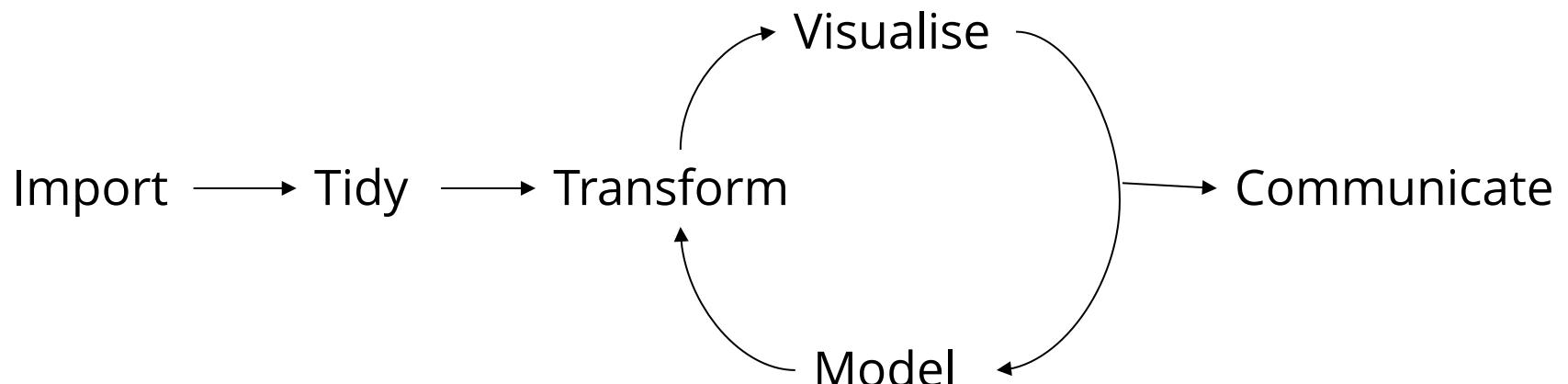
So many
paths...



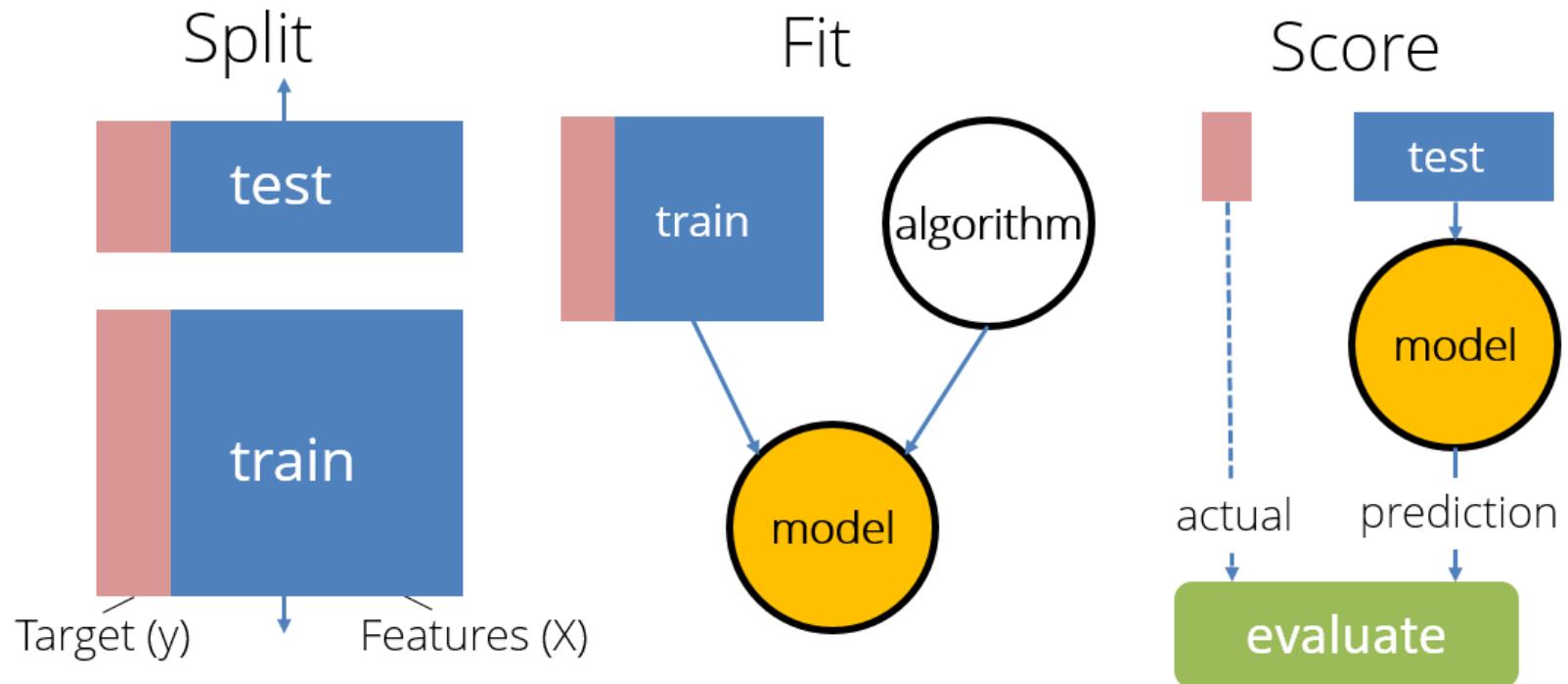
Time to tidy up

Tidy Data Analysis

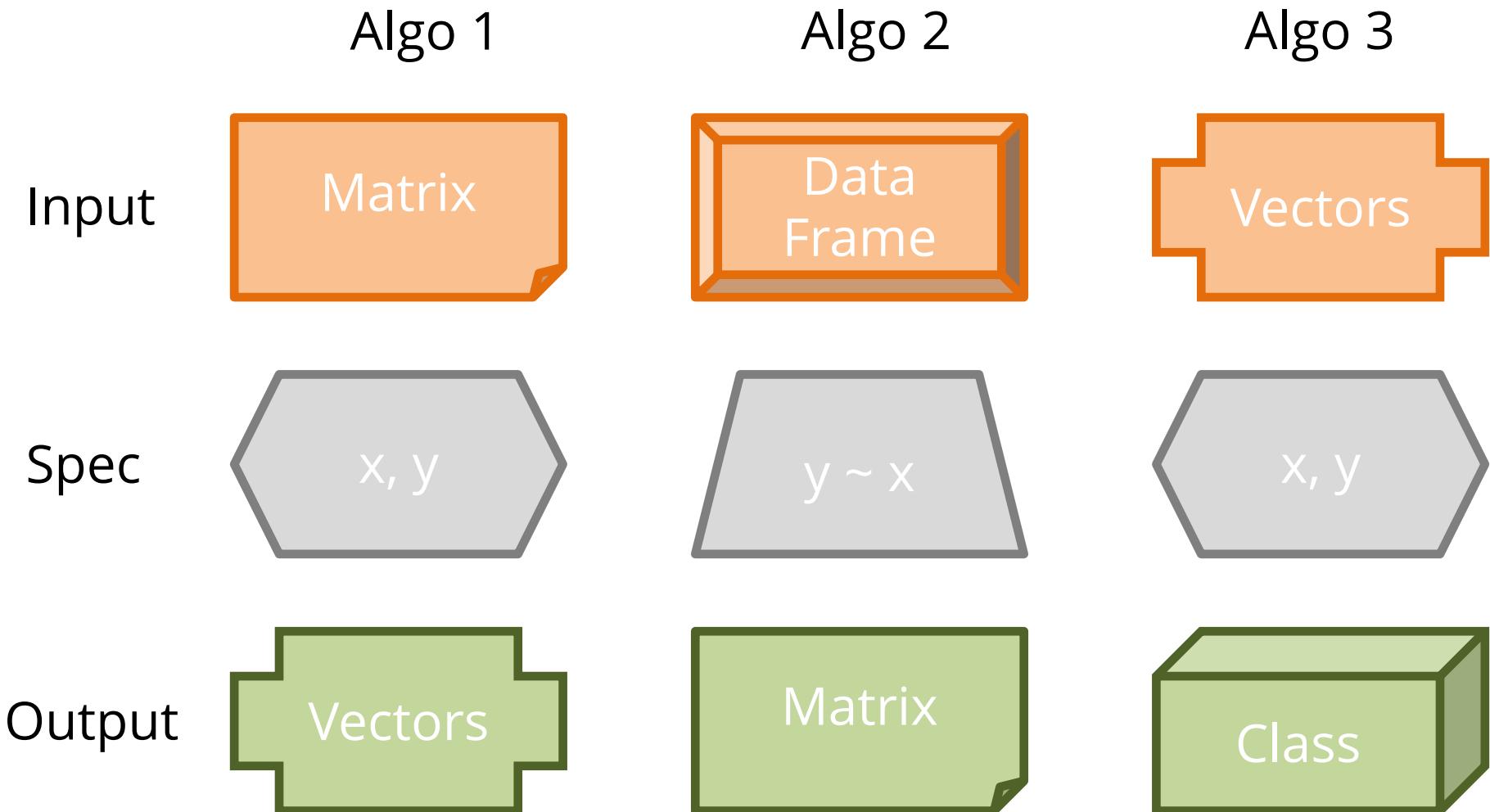
- tidyverse: A bundle of packages covering the data analysis workflow



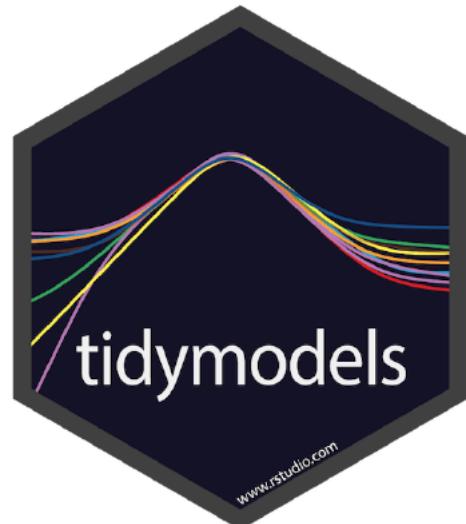
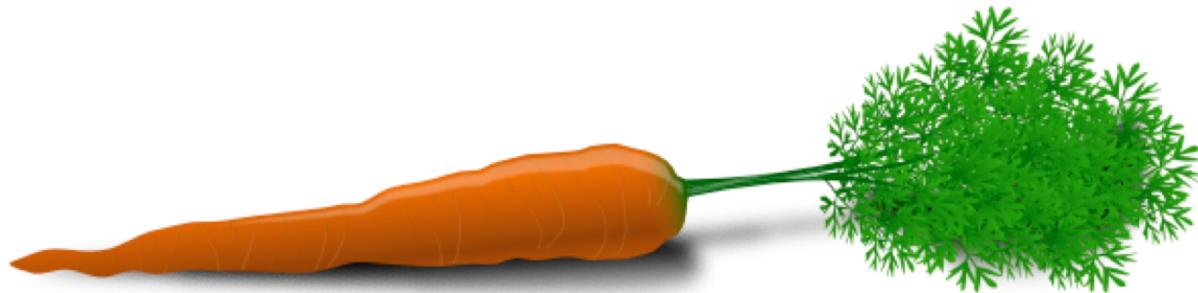
Tidy Modelling



Modelling algorithms in R behave in many curious ways



Modelling frameworks



Max Kuhn

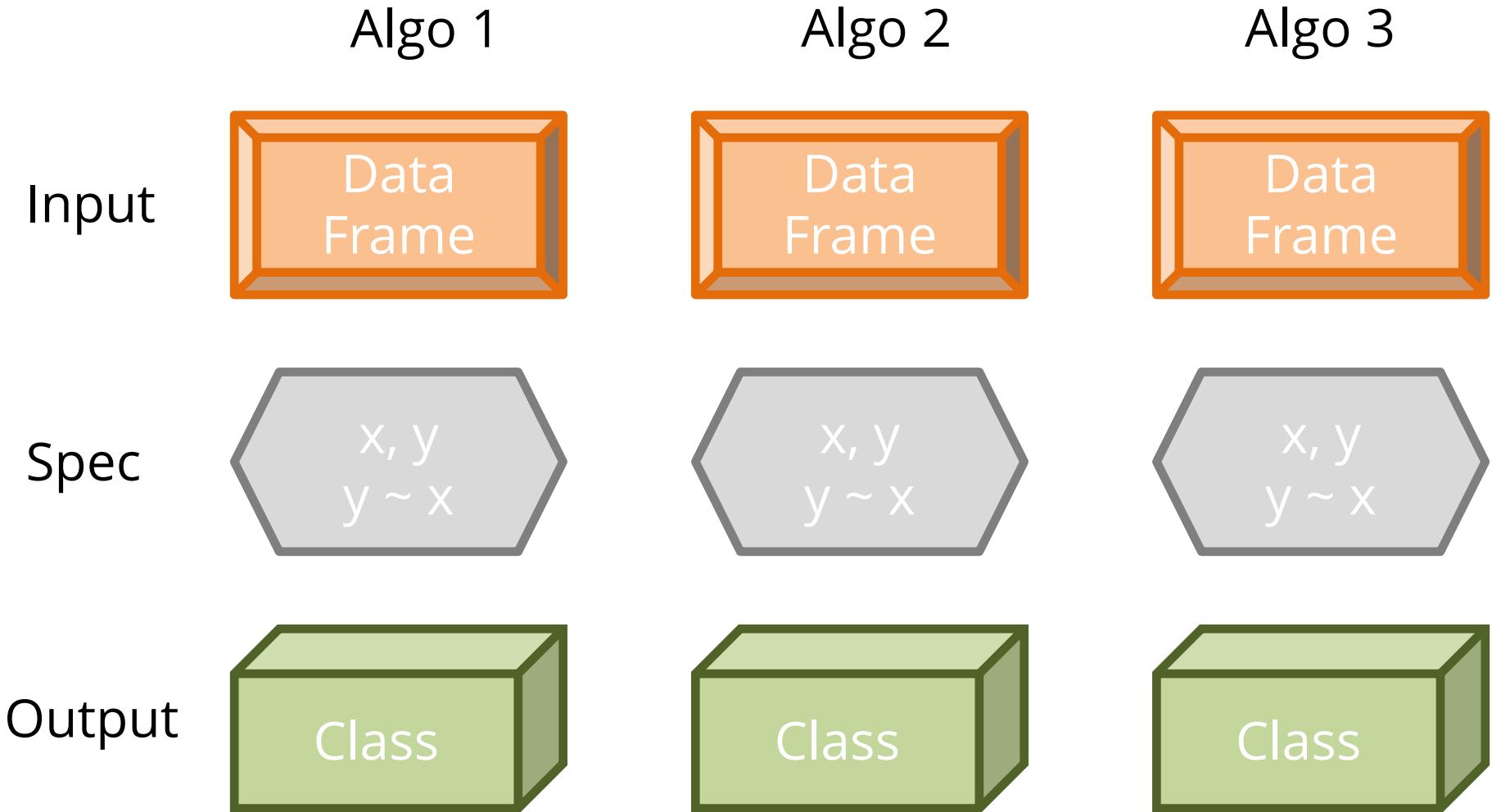


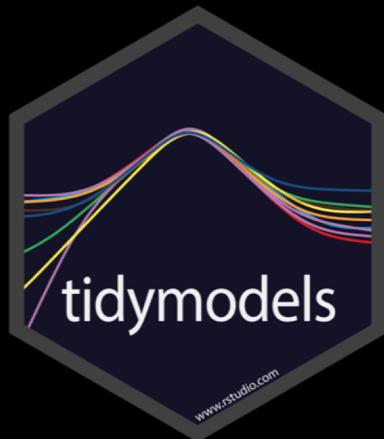
tidymodels

- Create a unified interface to models tidy way
- Organise by model type
- Generalise how the models are fitted
- Always return the same object type



So tidy...





Demo

