

Unsupervised Learning

Clustering



X-Ed

Demian Wassermann
Inria, France

Disgression

What are Processing Pipelines?

Artisanal Development

- Case-by-case development
- Extra effort to re-apply to a different dataset
- Process depiction less clear
- Modularity / model exploration is harder

Pipeline-based Development

- General-case development
- Simpler reuse
- Steps organised sequentially
- Simpler production of a modular design
- Interchangeable, modifiable workflow

Disgression

What are Processing Pipelines?

Artisanal Development

- Case-by-case development
- Extra effort to re-apply to a different dataset
- Process depiction less clear
- Modularity / model exploration is harder

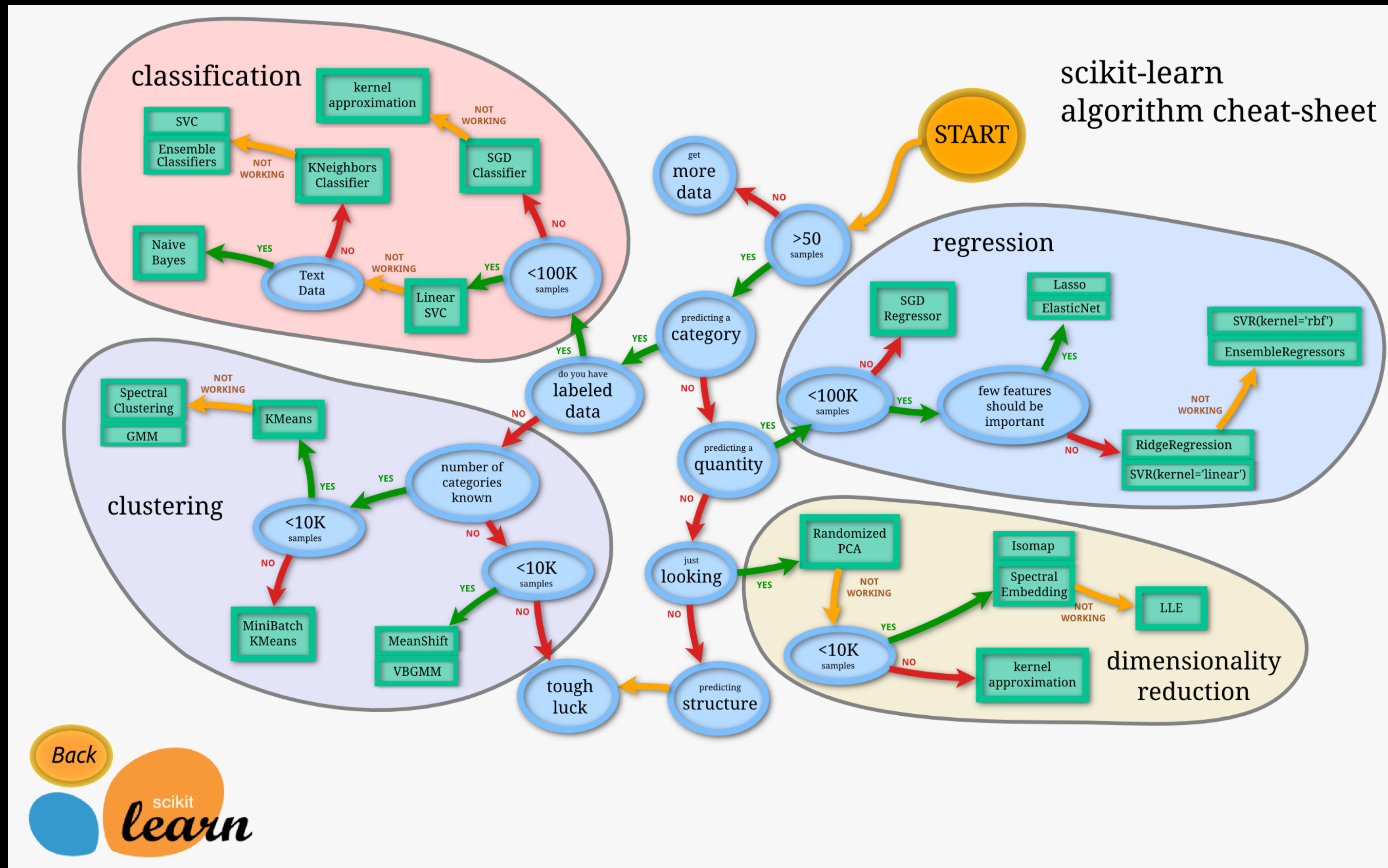


Pipeline-based Development

- General-case development
- Simpler reuse
- Steps organised sequentially
- Simpler production of a modular design
- Interchangeable, modifiable workflow



Where are we in the Data Science Landscape?



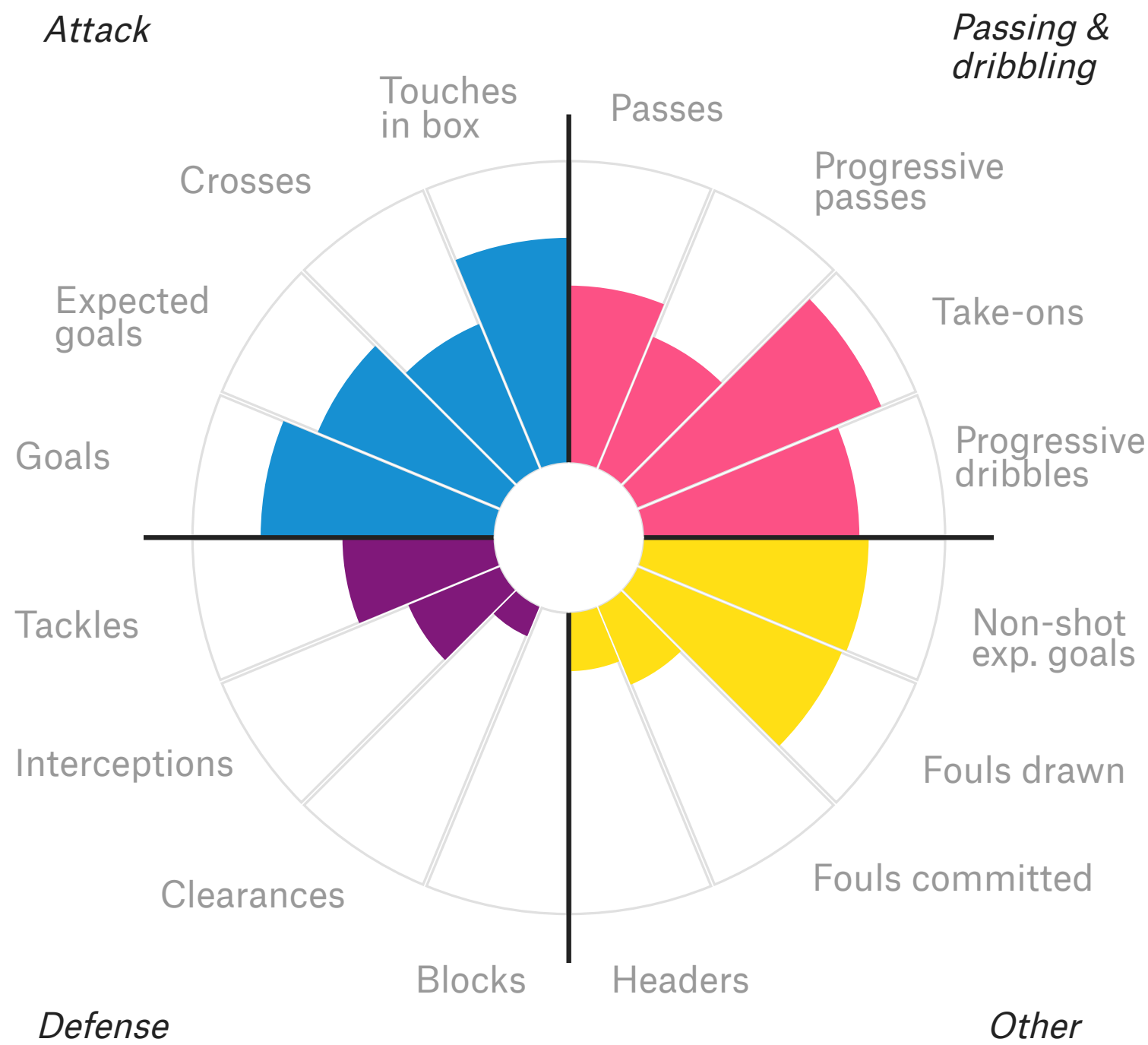
Motivating Example



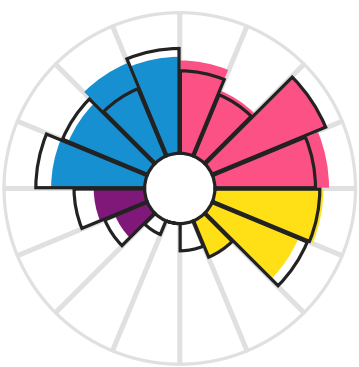
Kylian Mbappé

TEAM: FRANCE 2018

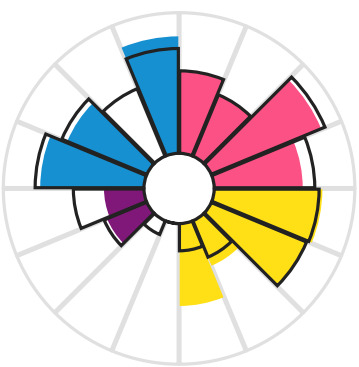
PLAY STYLE: DYNAMIC PLAYMAKER



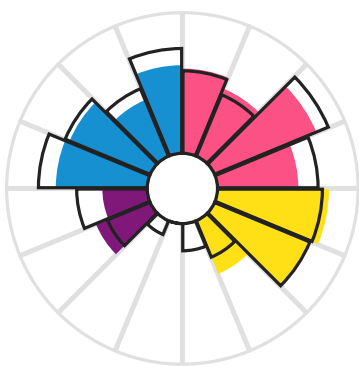
Most similar World Cup performances



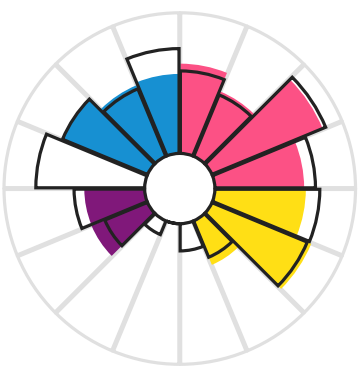
1. Brian Laudrup
DENMARK 1998



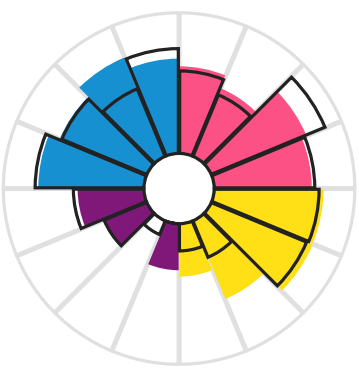
2. Ladislav Petras
CZECHOSLOVAKIA 1970



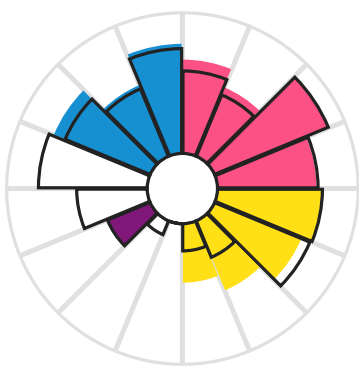
3. Aleksandr Zavarov
USSR 1990



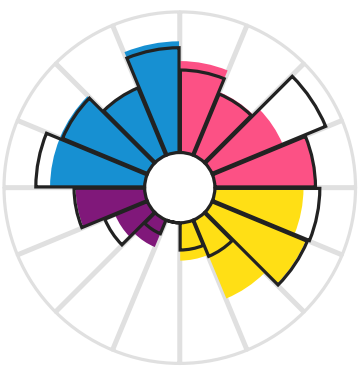
4. Jorge González
EL SALVADOR 1982



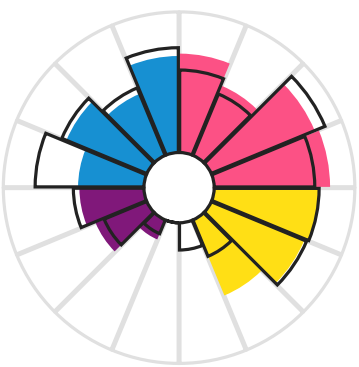
5. Neymar
BRAZIL 2014



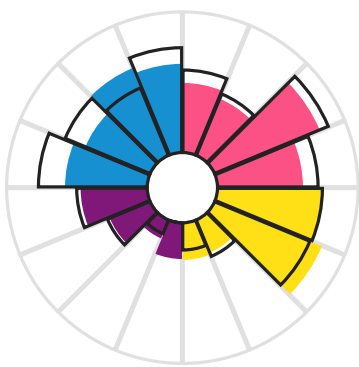
6. Safet Susic
YUGOSLAVIA 1982



7. Landon Donovan
USA 2002



8. Franck Ribéry
FRANCE 2006



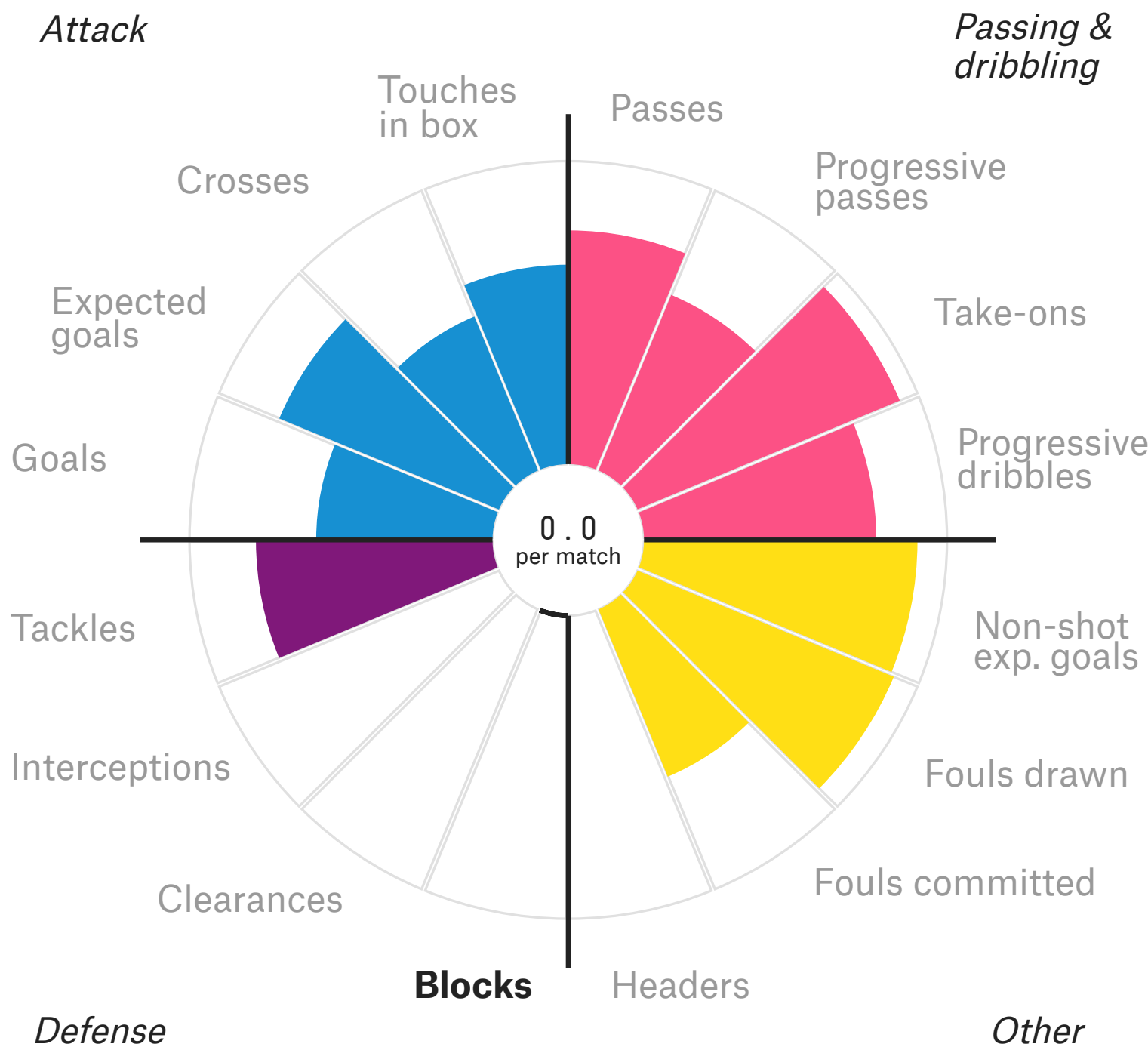
9. Daniel Bertoni
ARGENTINA 1978

Motivating Example

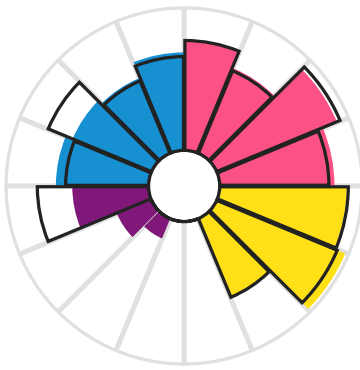


Lionel Messi

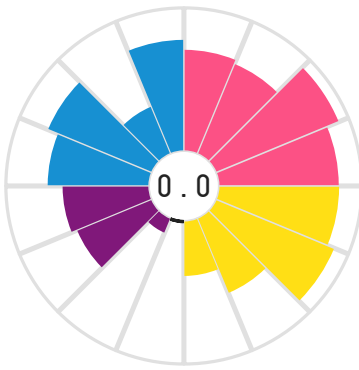
TEAM: ARGENTINA 2018
PLAY STYLE: DYNAMIC PLAYMAKER



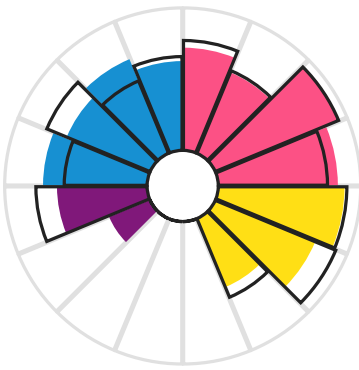
Most similar World Cup performances



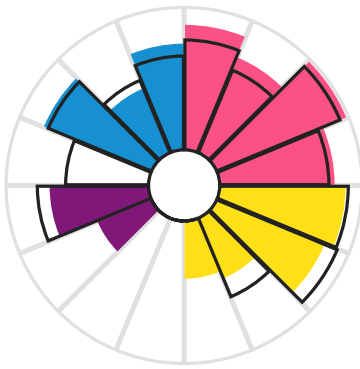
1. Ariel Ortega
ARGENTINA 1998



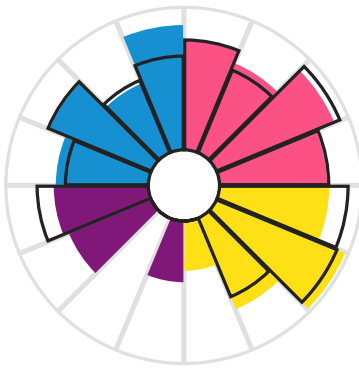
2. Eden Hazard
BELGIUM 2018



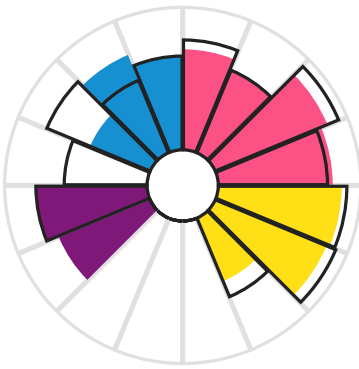
3. Lionel Messi
ARGENTINA 2014



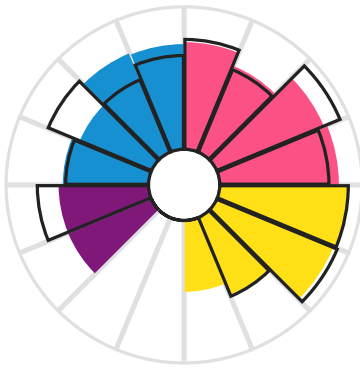
4. Lionel Messi
ARGENTINA 2010



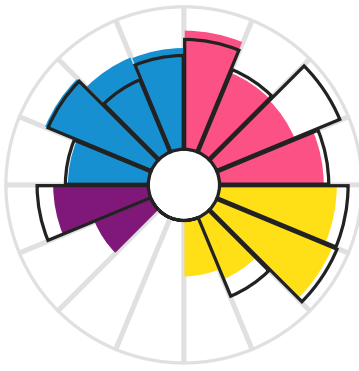
5. Diego Maradona
ARGENTINA 1982



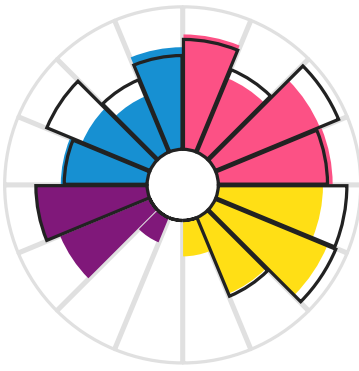
6. Franck Ribéry
FRANCE 2010



7. Fredrik Ljungberg
SWEDEN 2006



8. Cuauhtémoc Blanco
MEXICO 2002

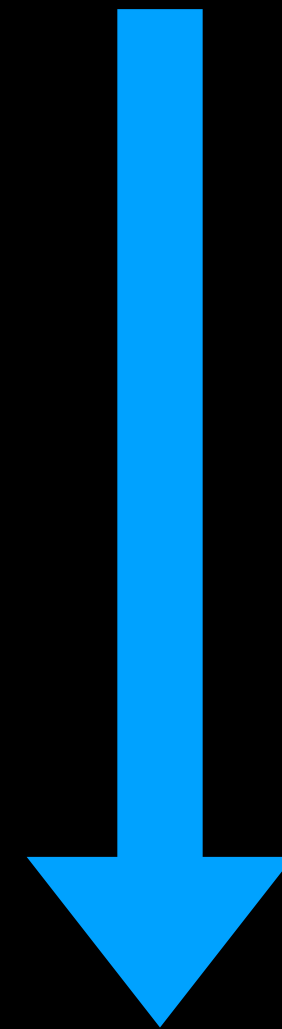


9. Andrés Iniesta
SPAIN 2010

Regression Implies a Model of Reality

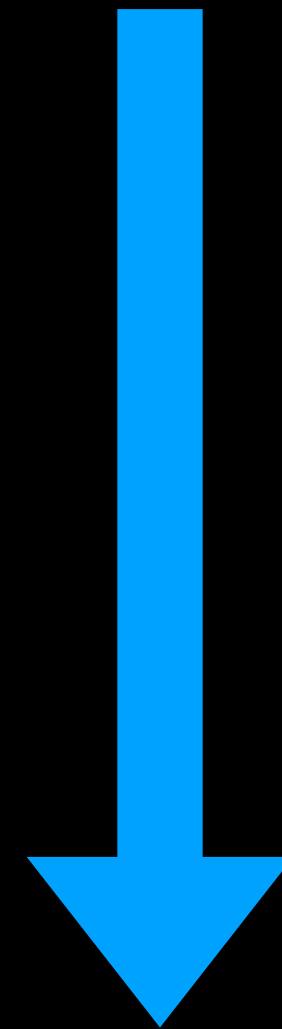
Supervised Implies a Model of Reality

Supervised Implies a Model of Reality



Unsupervised Implies a Exploring
Automatically-Inferred Reality Models

Supervised Implies a Model of Reality



~~Unsupervised Implies a Exploring
Automatically-Inferred Reality Models~~

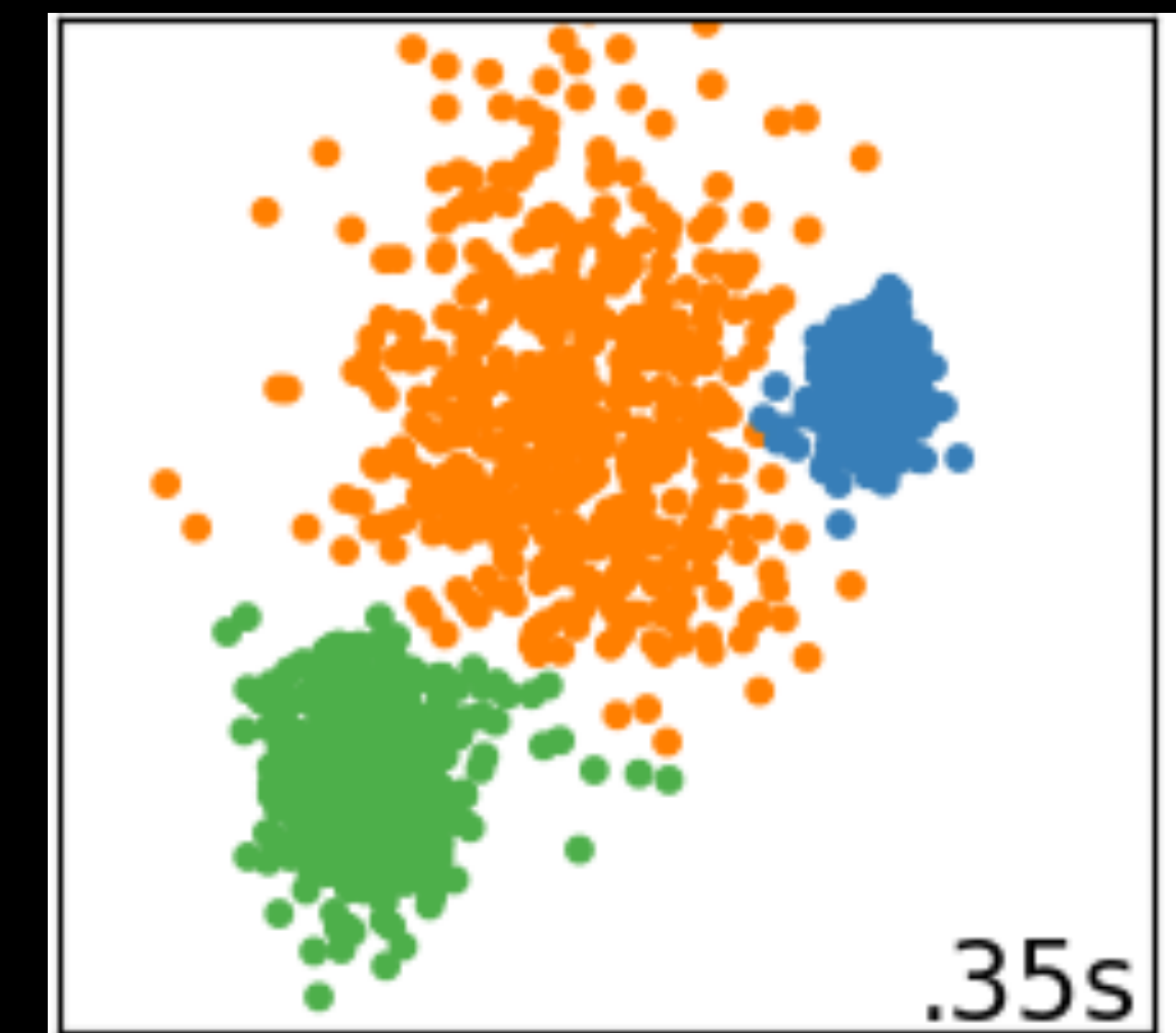
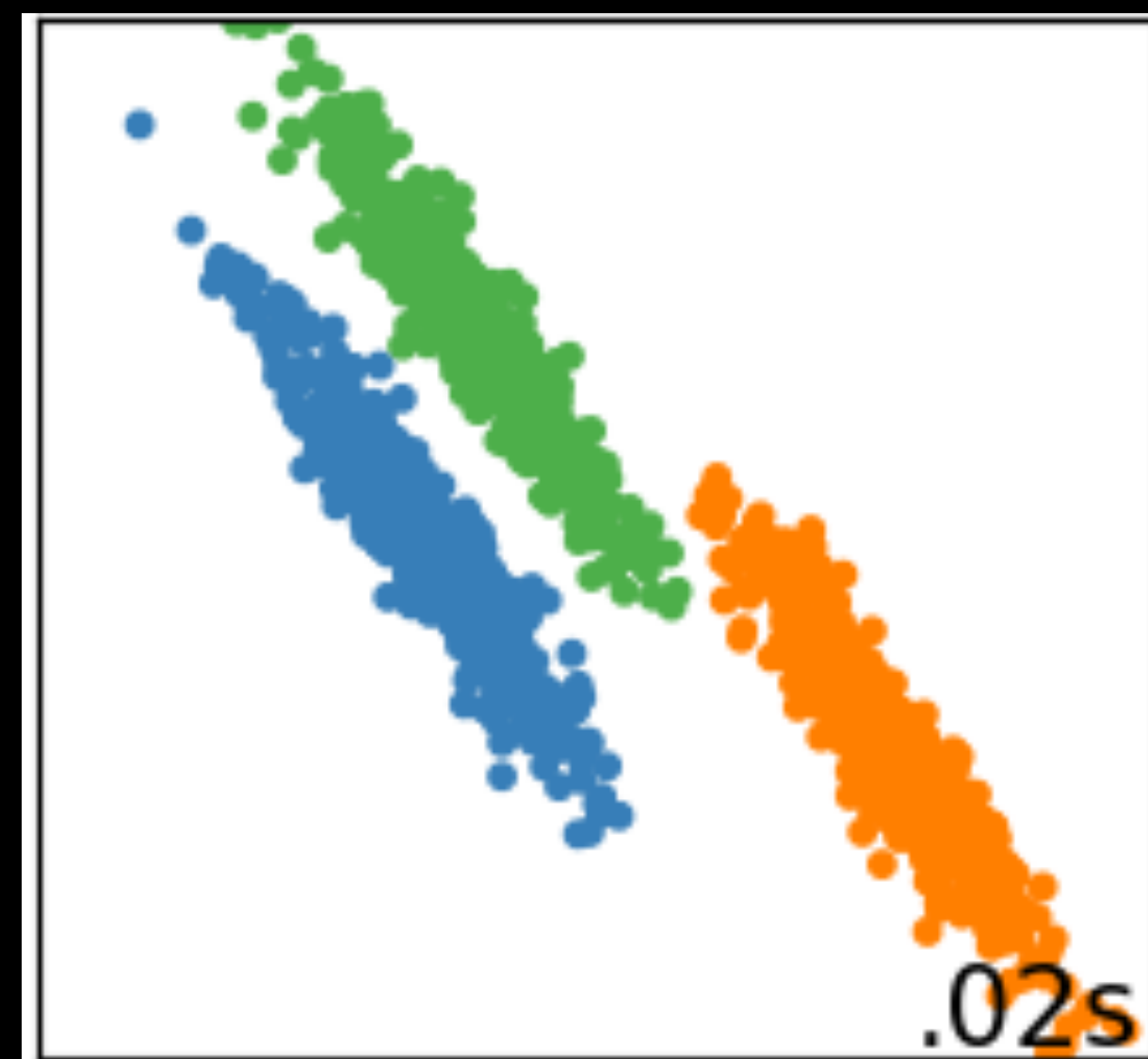
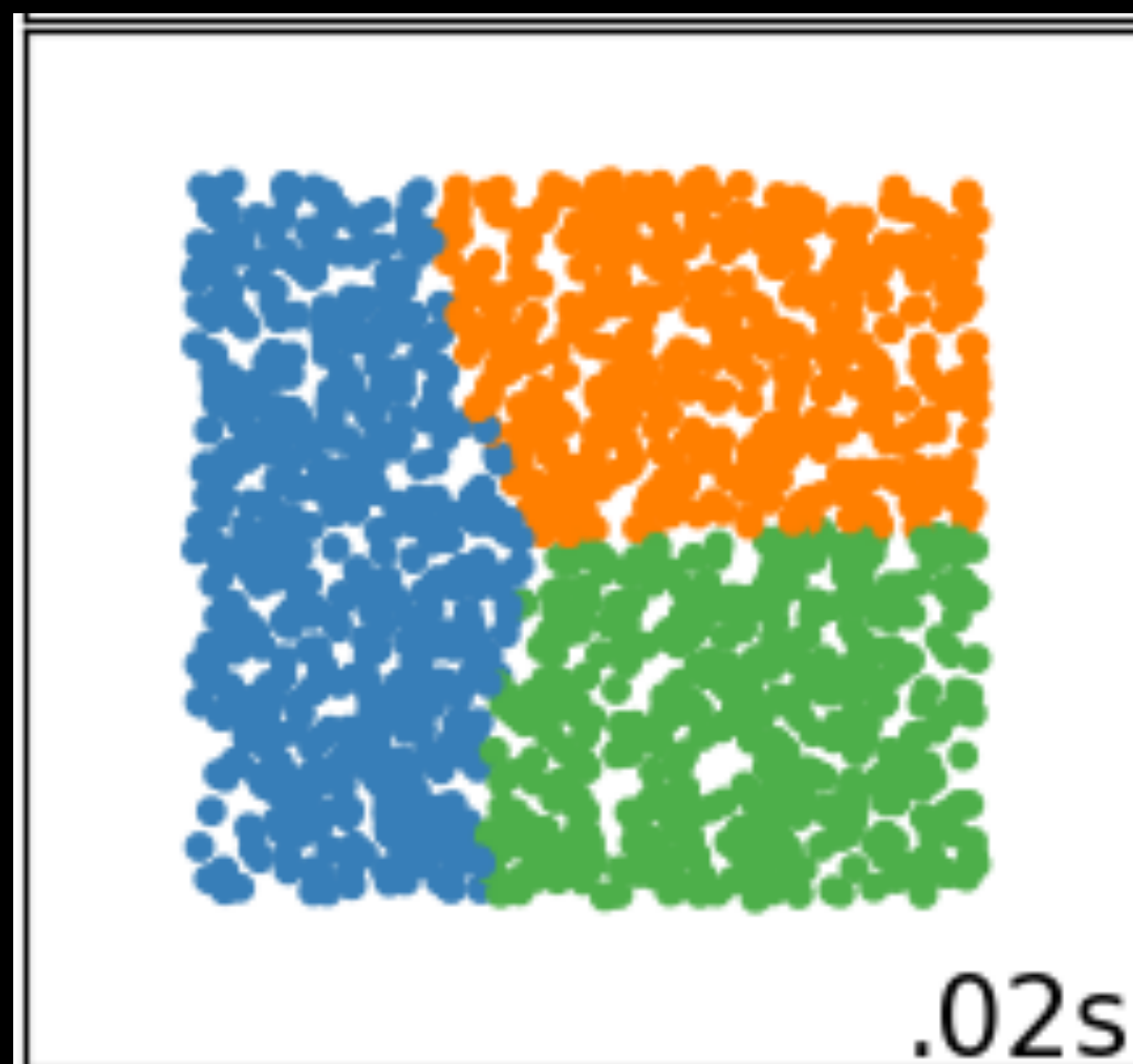
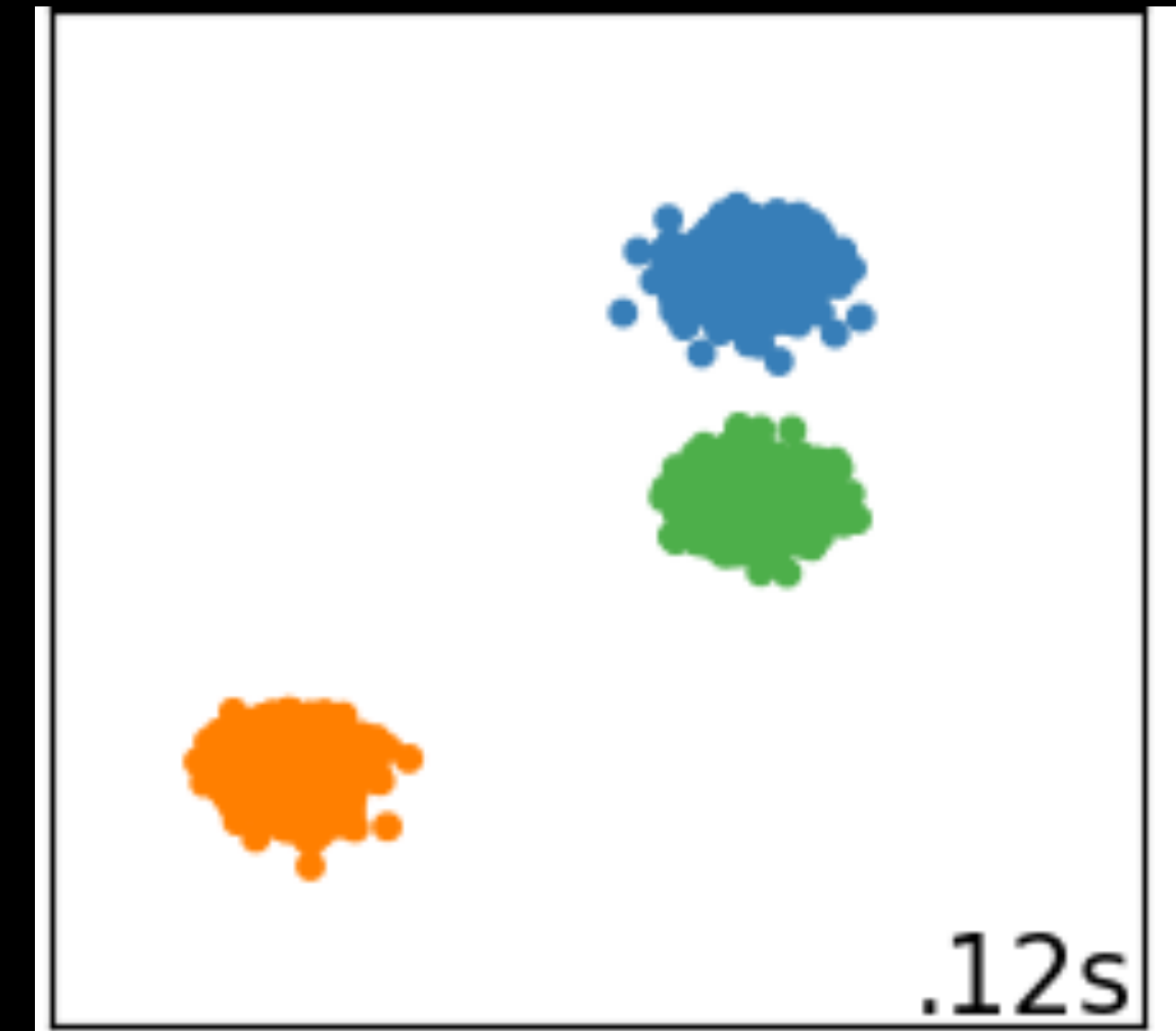
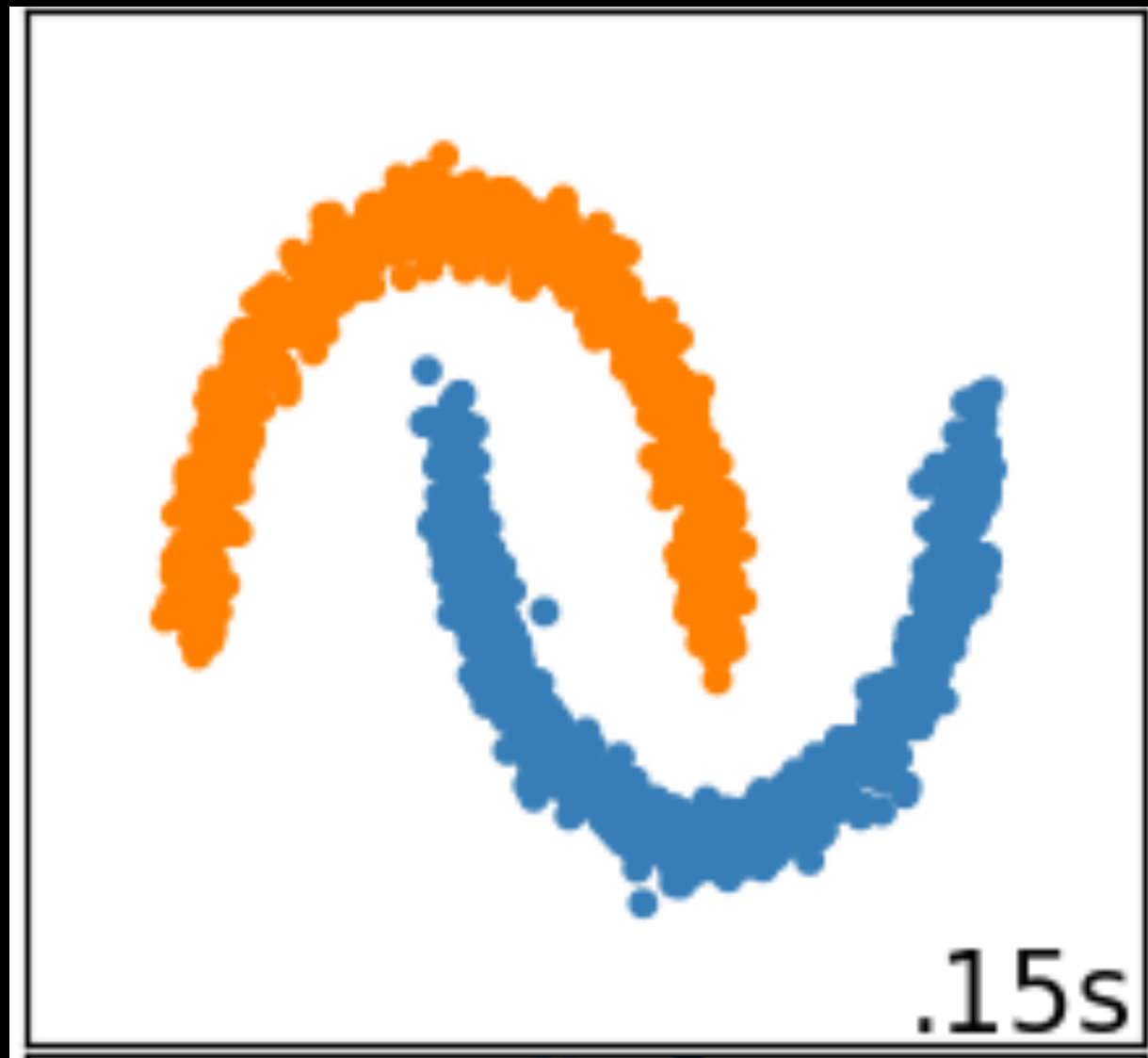
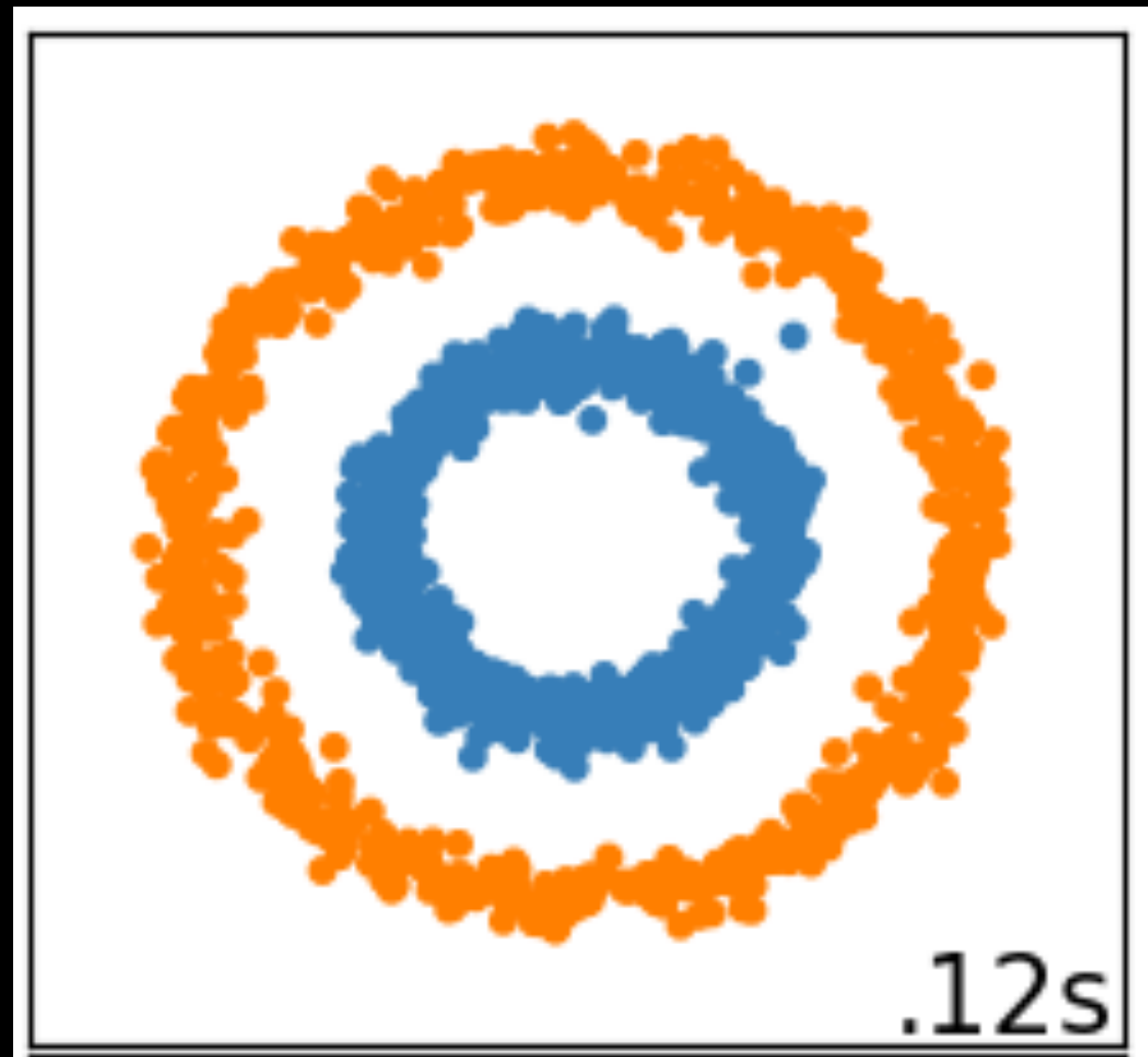
Supervised Implies a Model of Reality

In Unsupervised Methods the model of reality is implicit in the method choice.

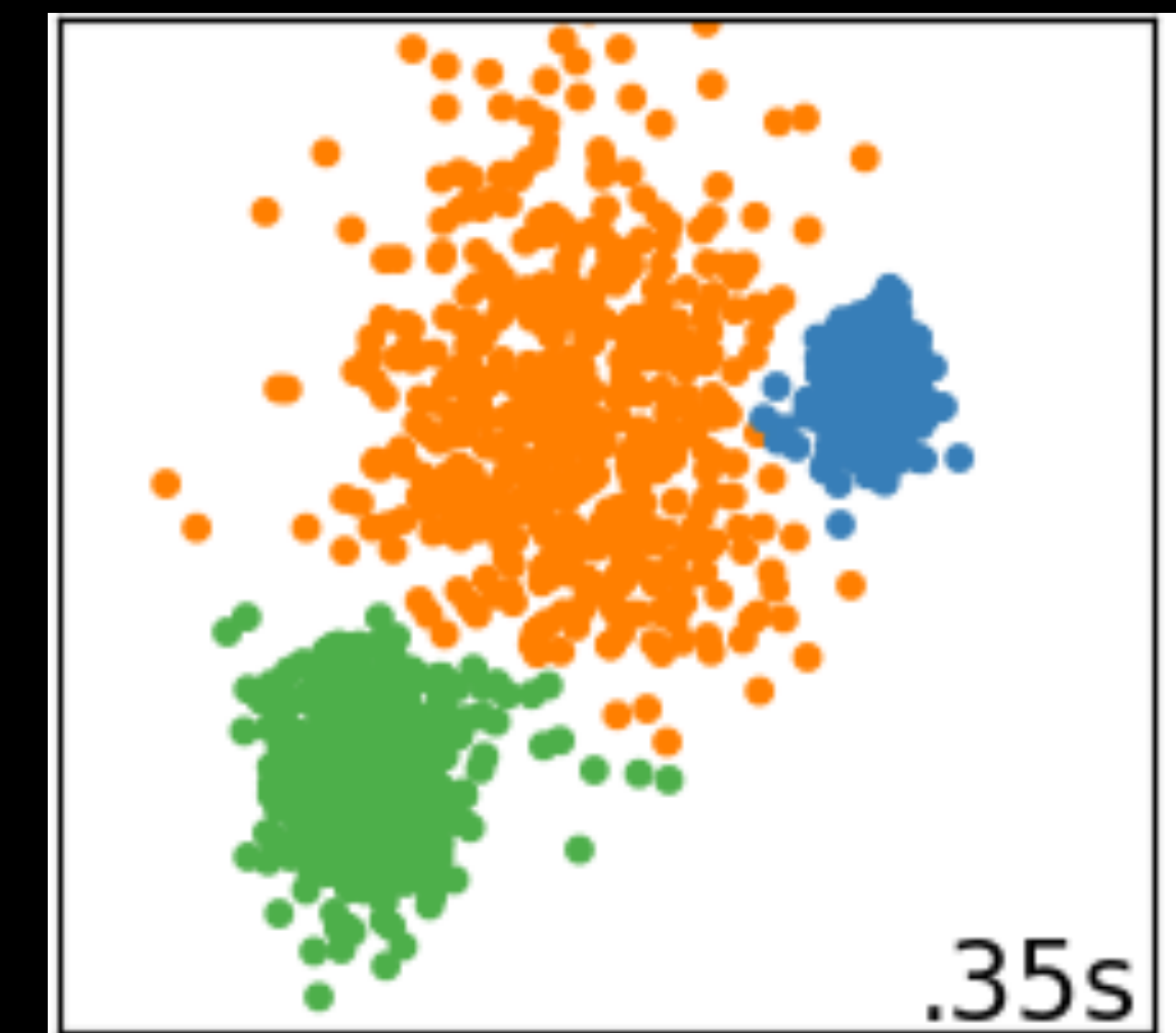
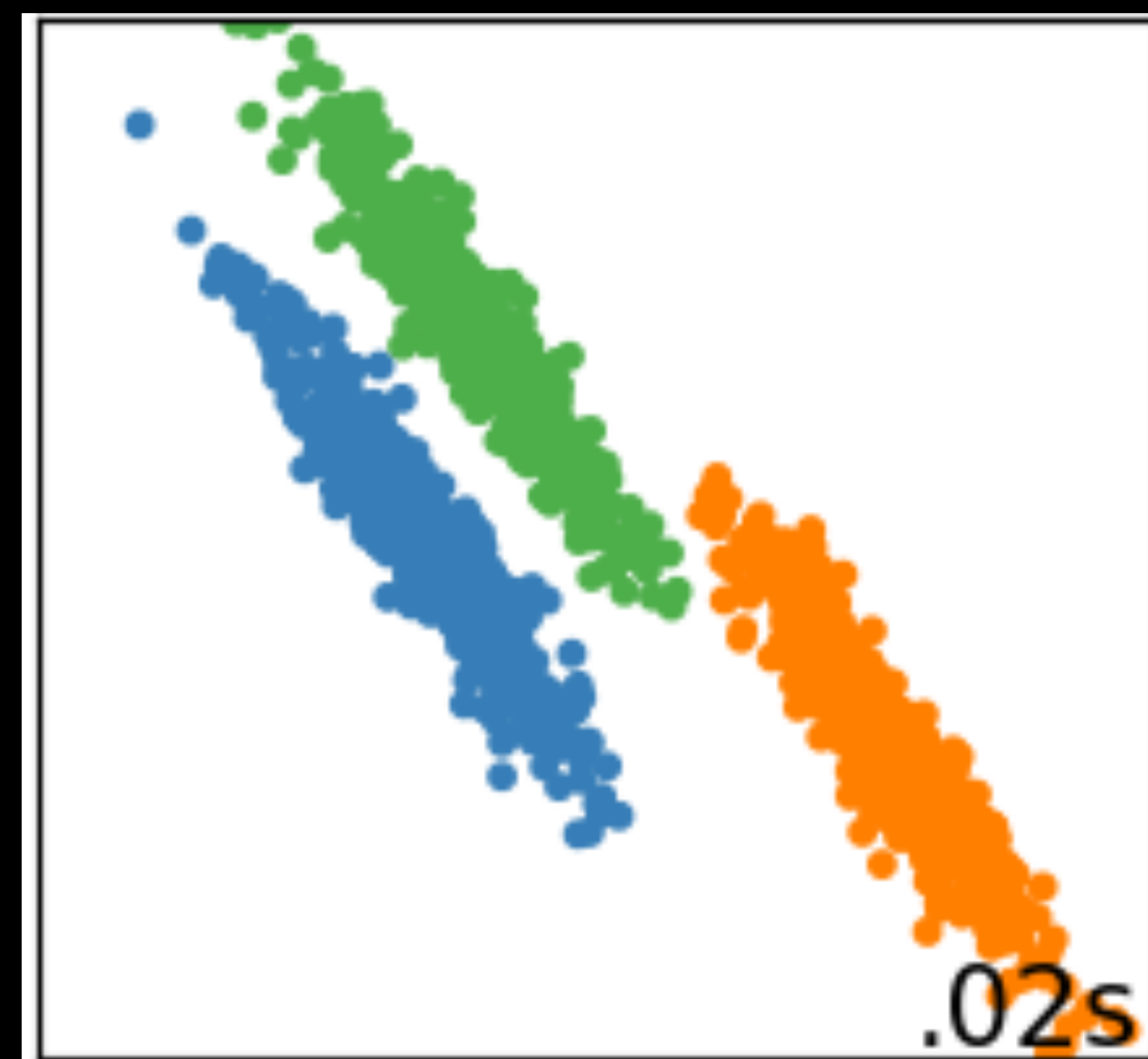
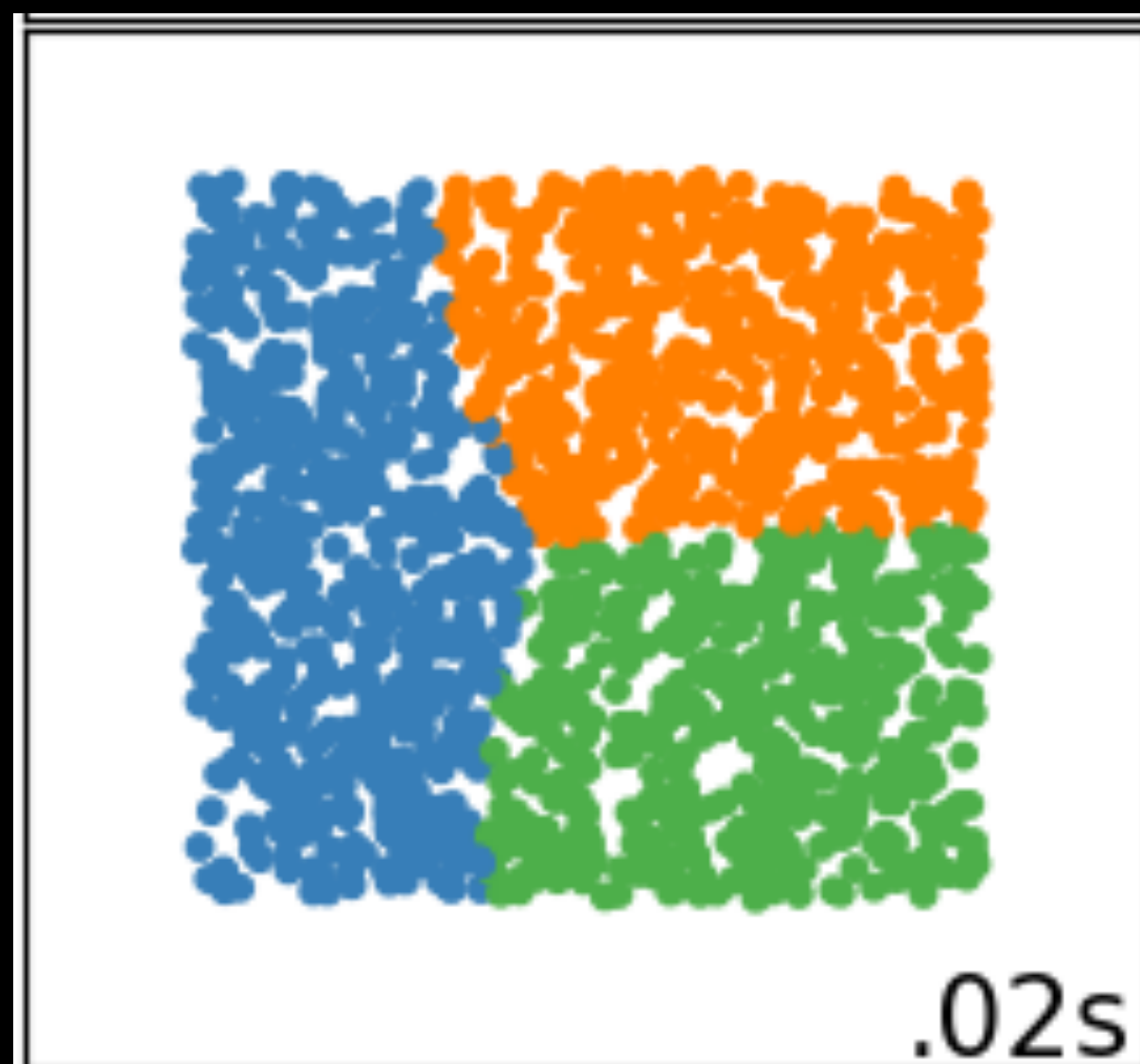
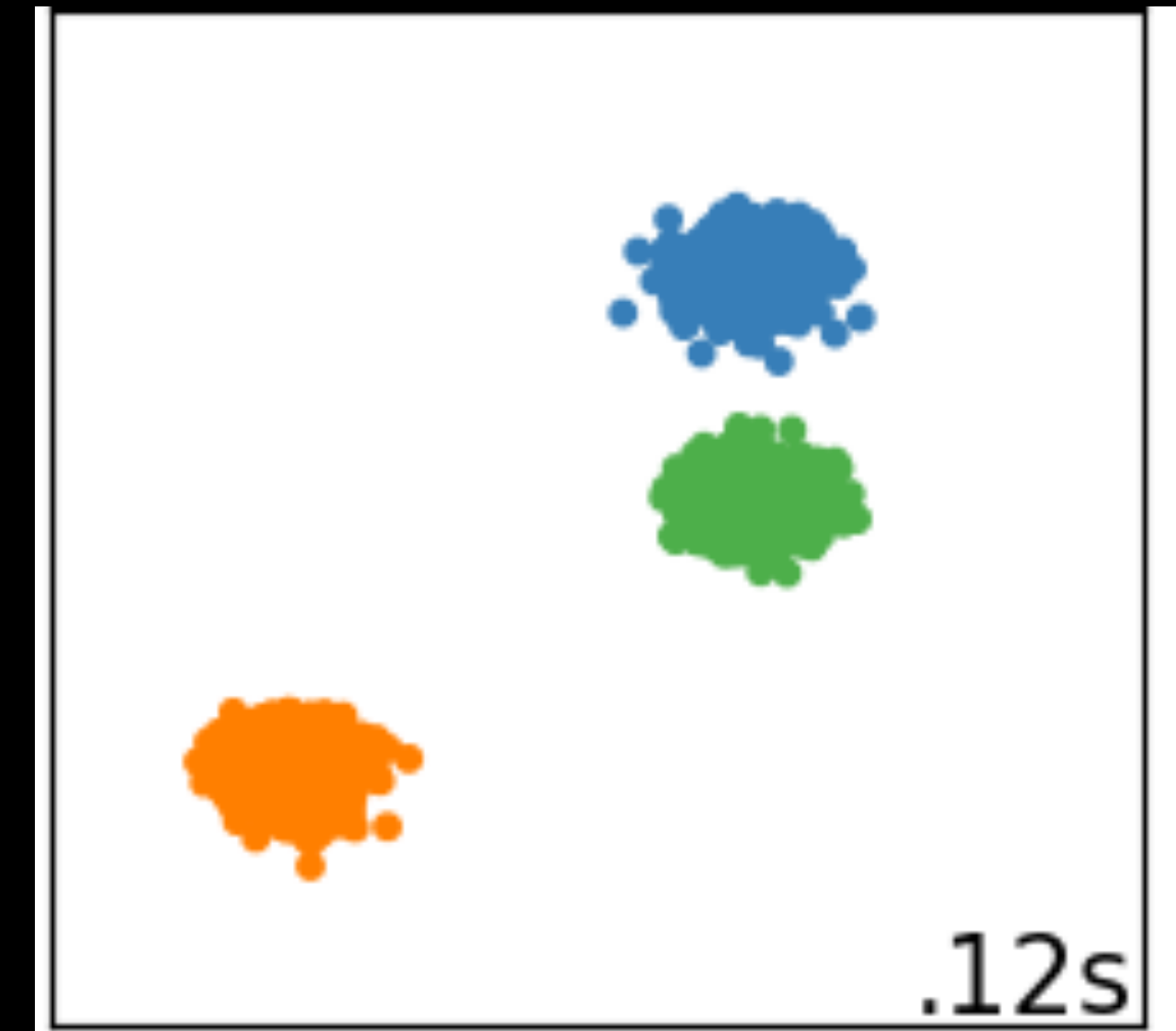
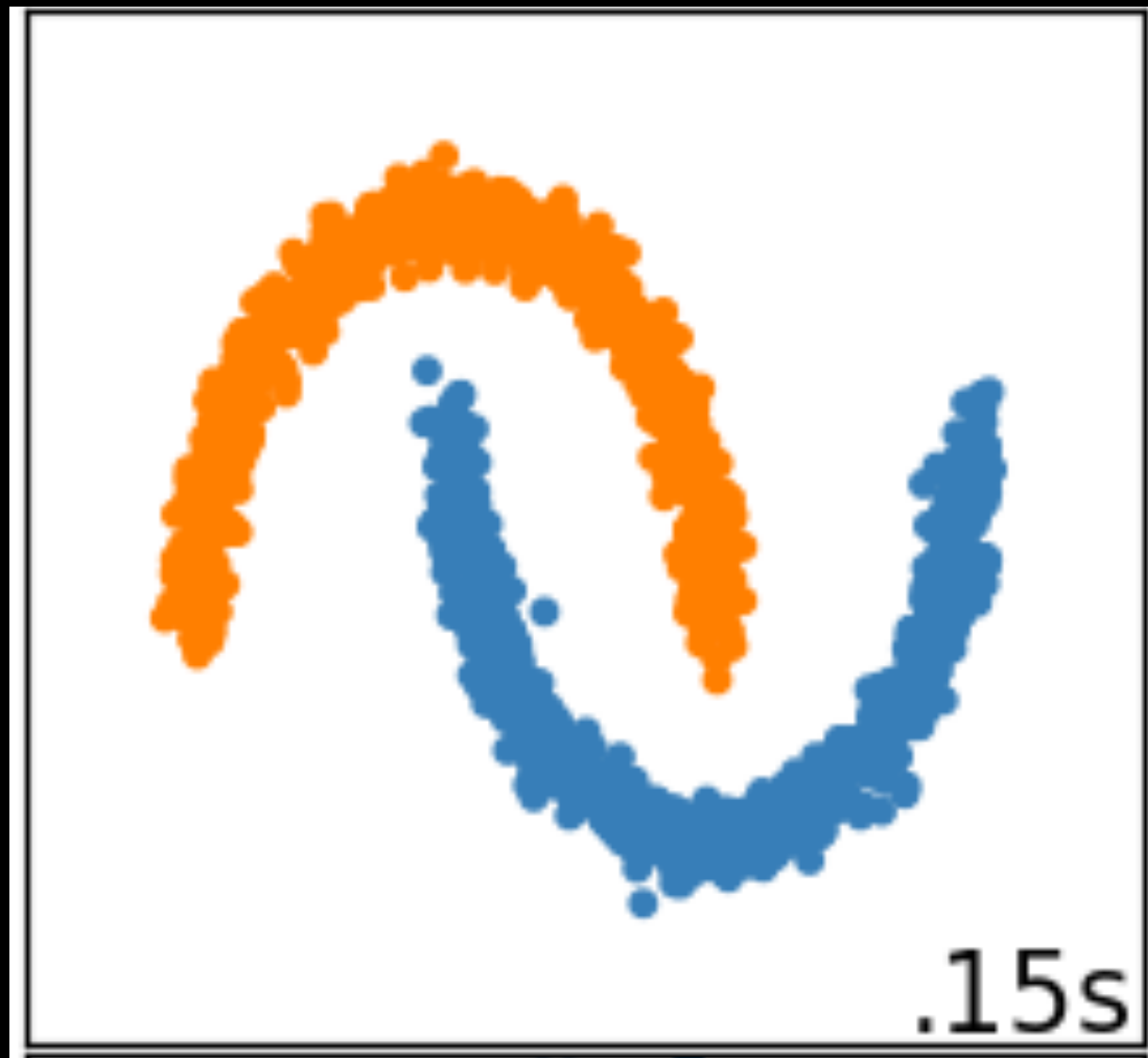
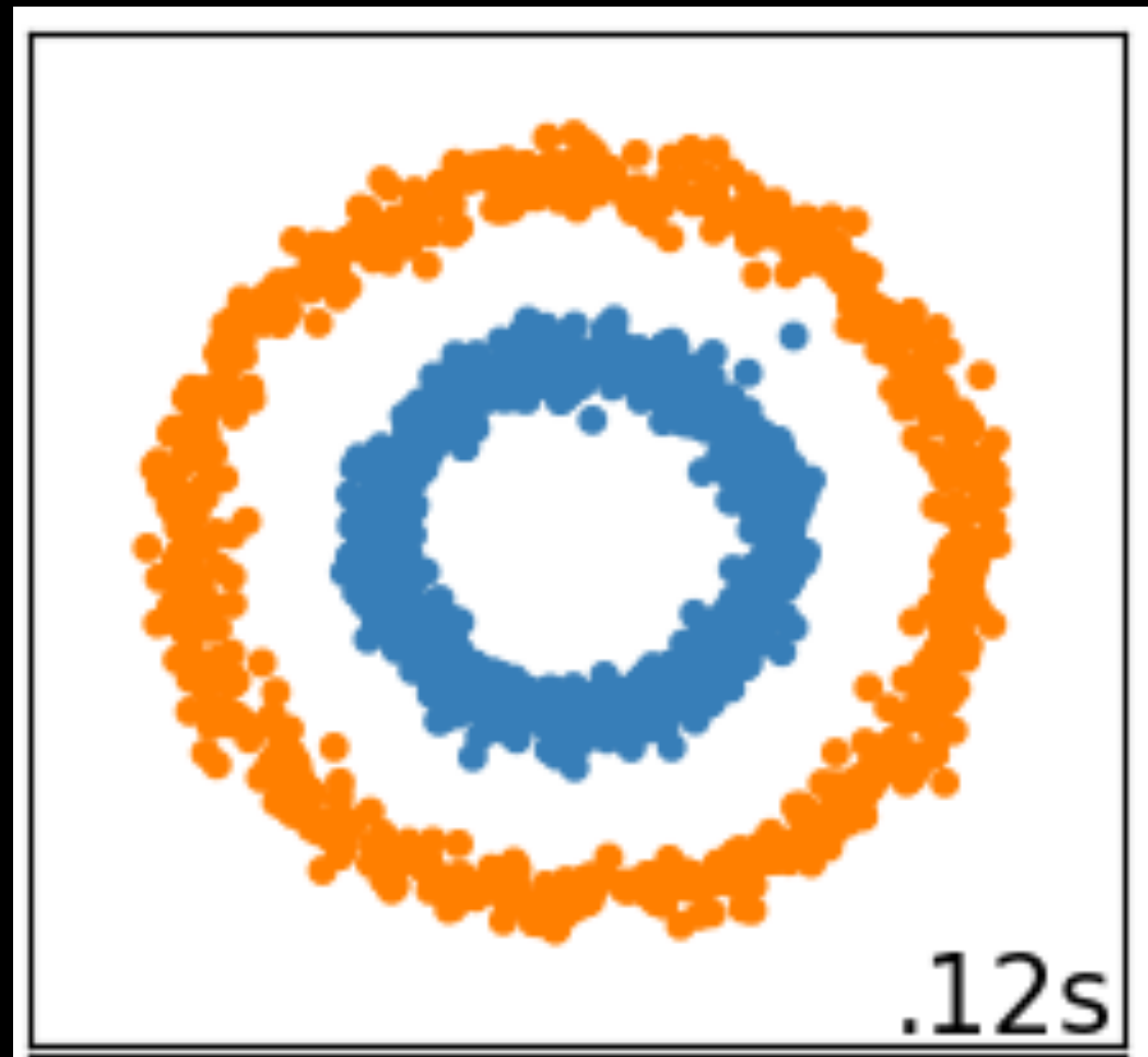
Each method mathematically represents an inherent reality viewpoint.

~~Unsupervised Implies a Exploring
Automatically-Inferred Reality Models~~

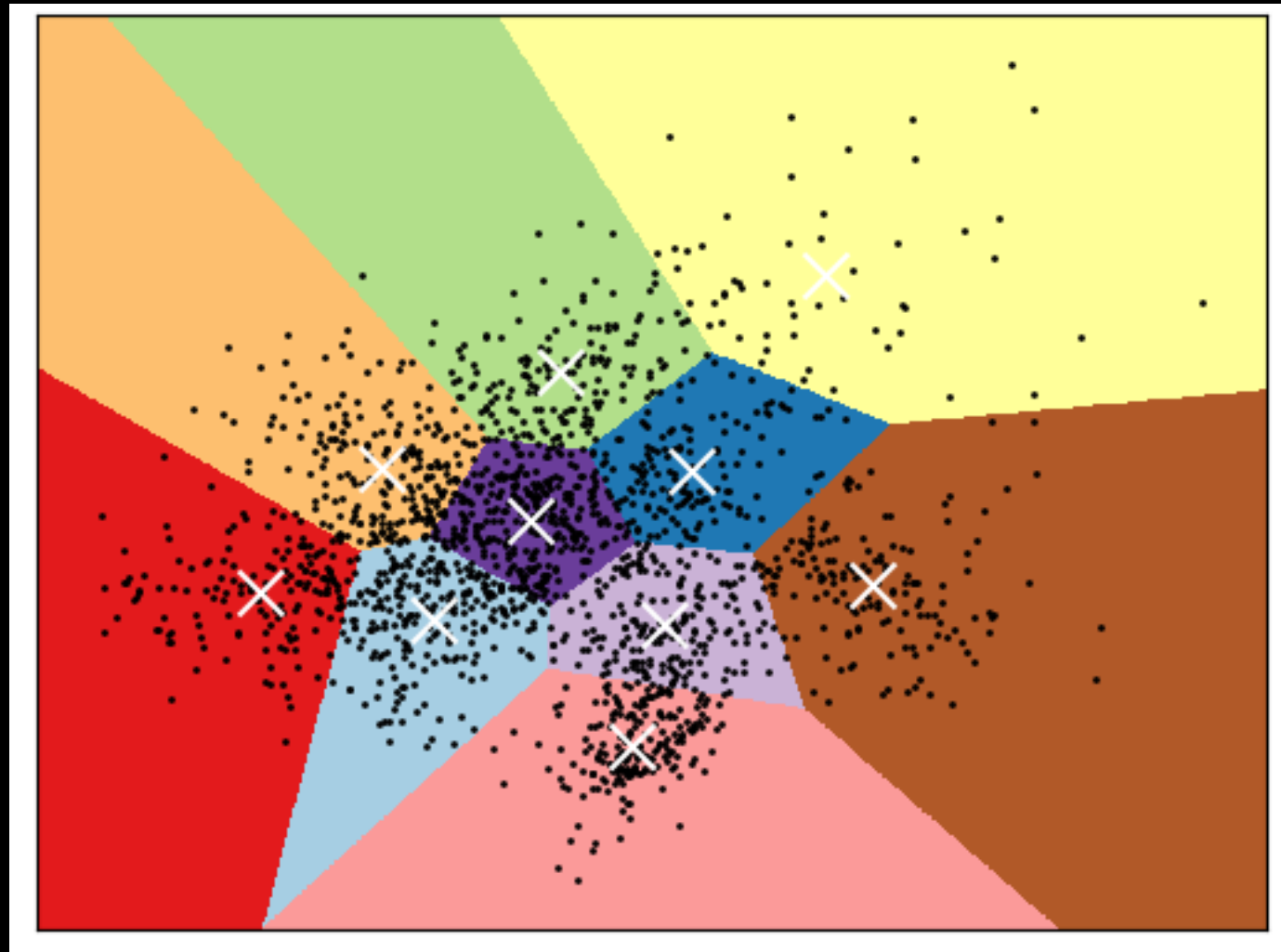
Regression Implies a Model of Reality



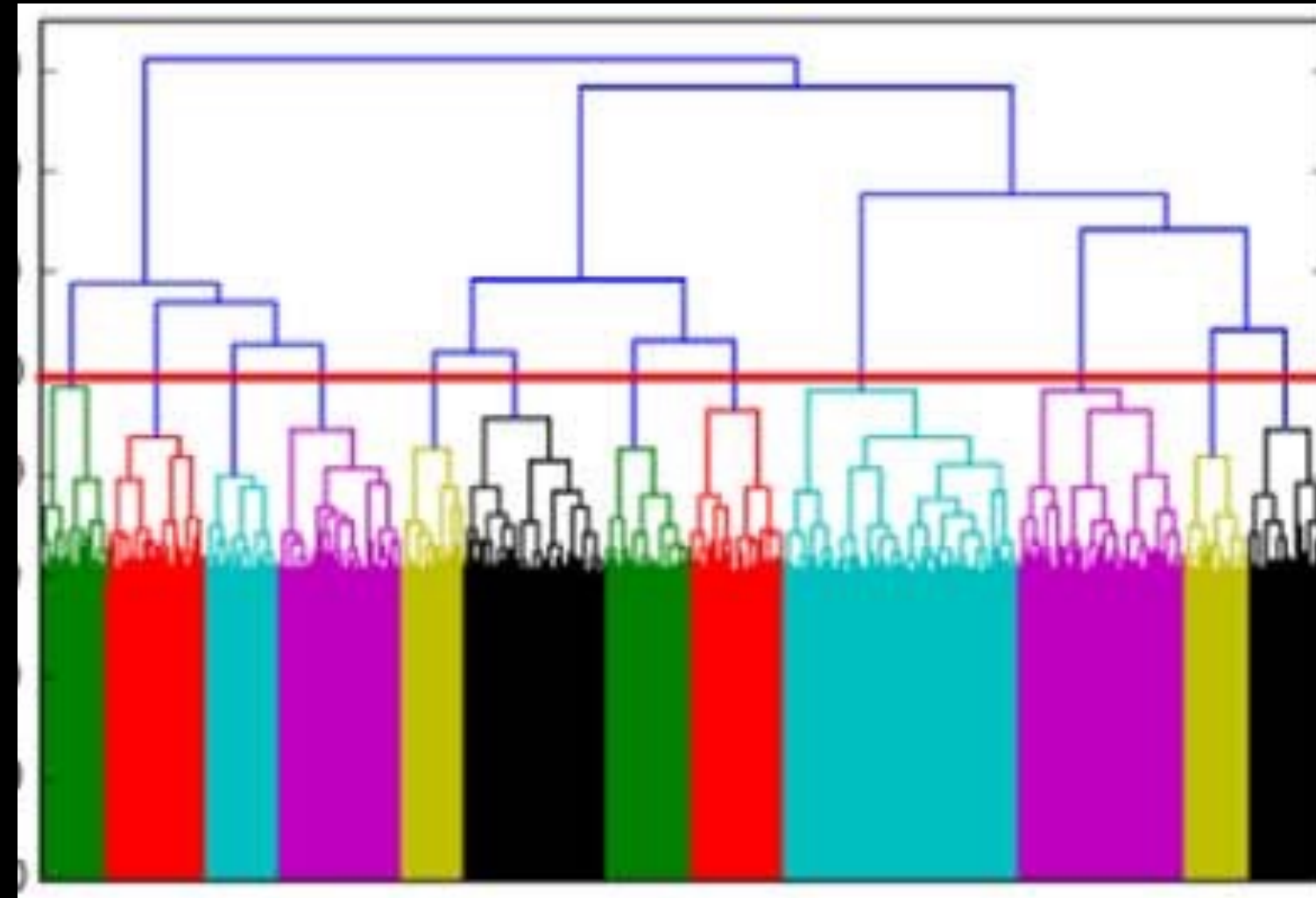
Examples of Inherent reality models



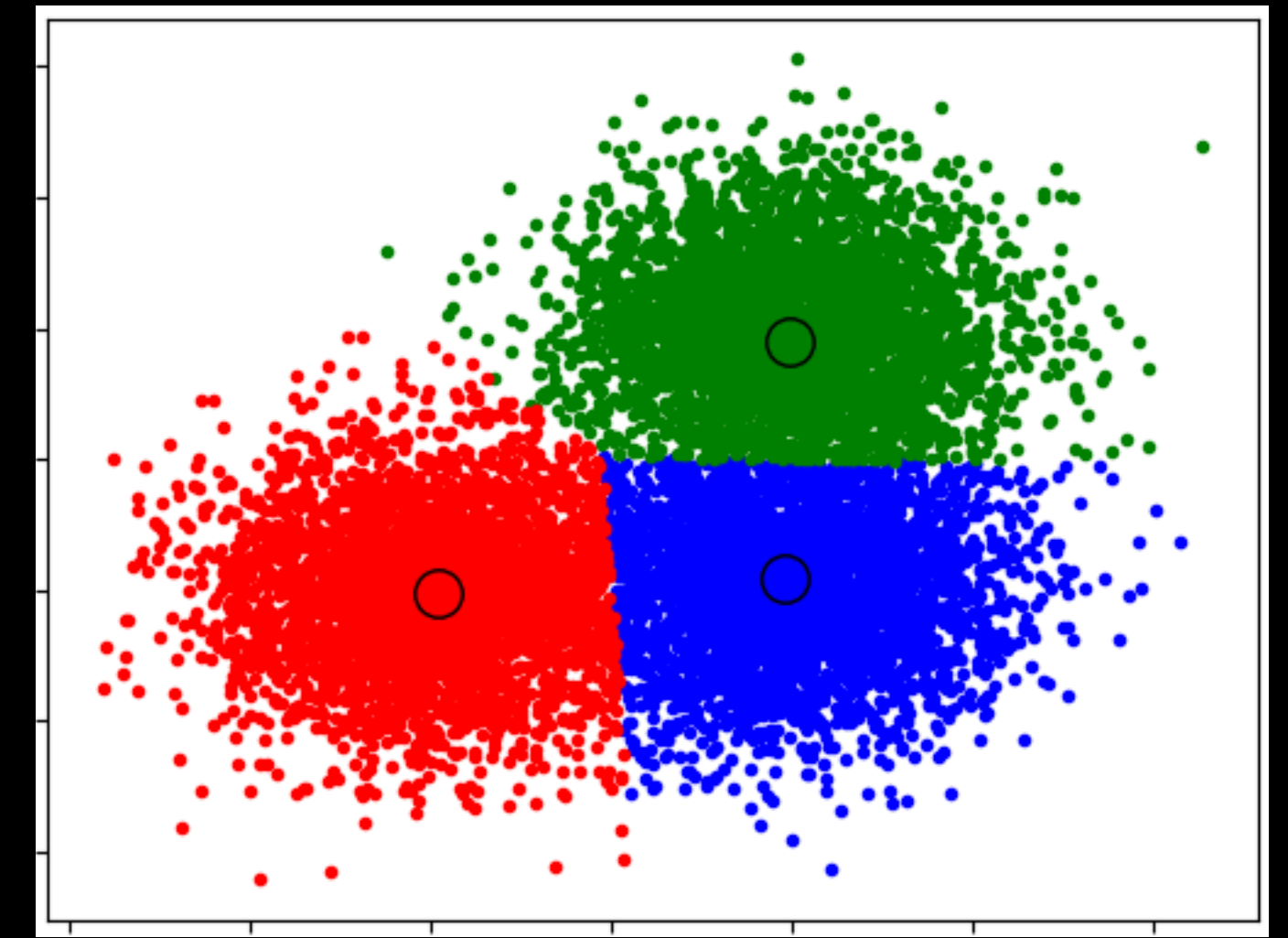
(Some) Clustering Models



K-means

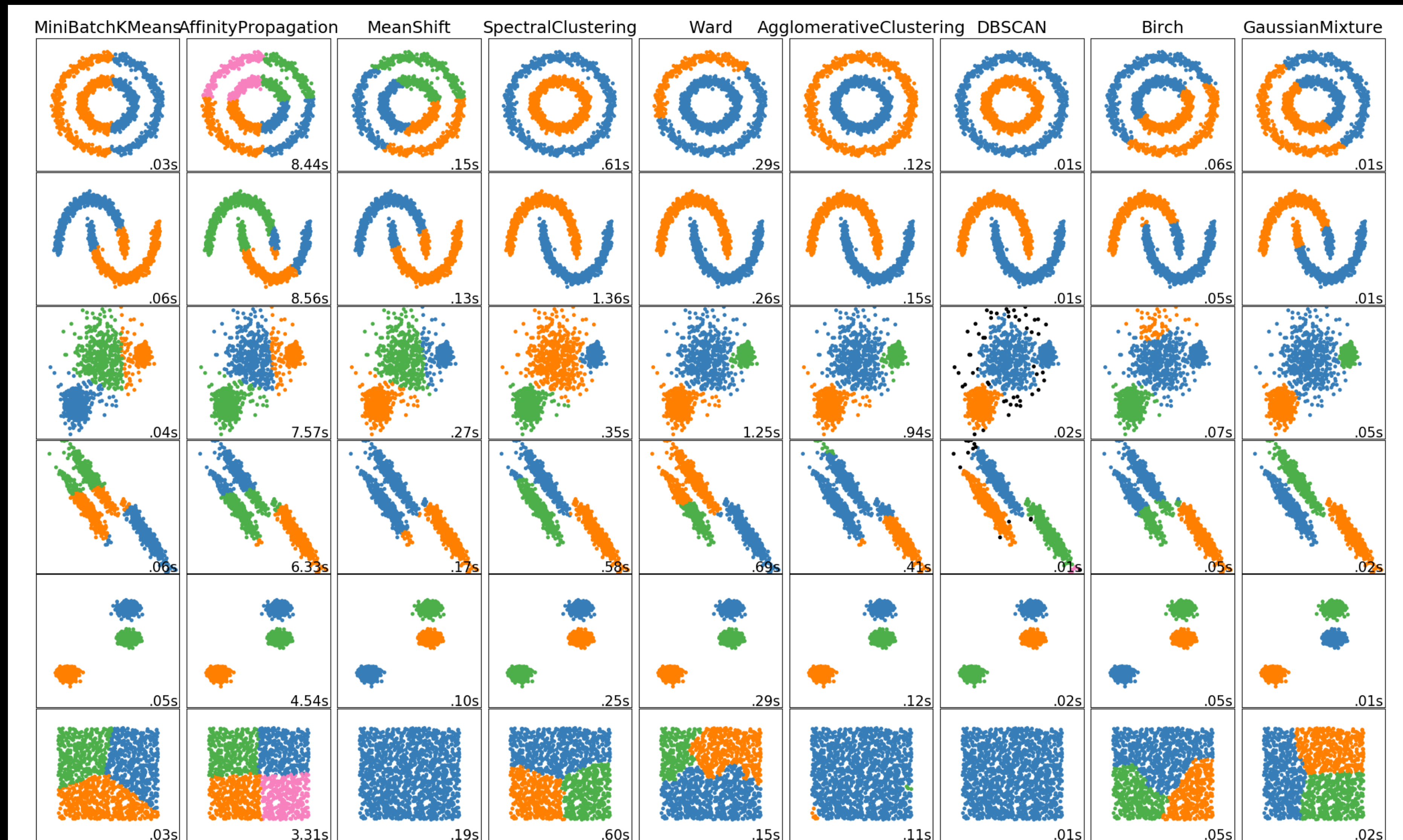


Hierarchical Clustering



Mean Shift

Model Selection and Validation in Clustering

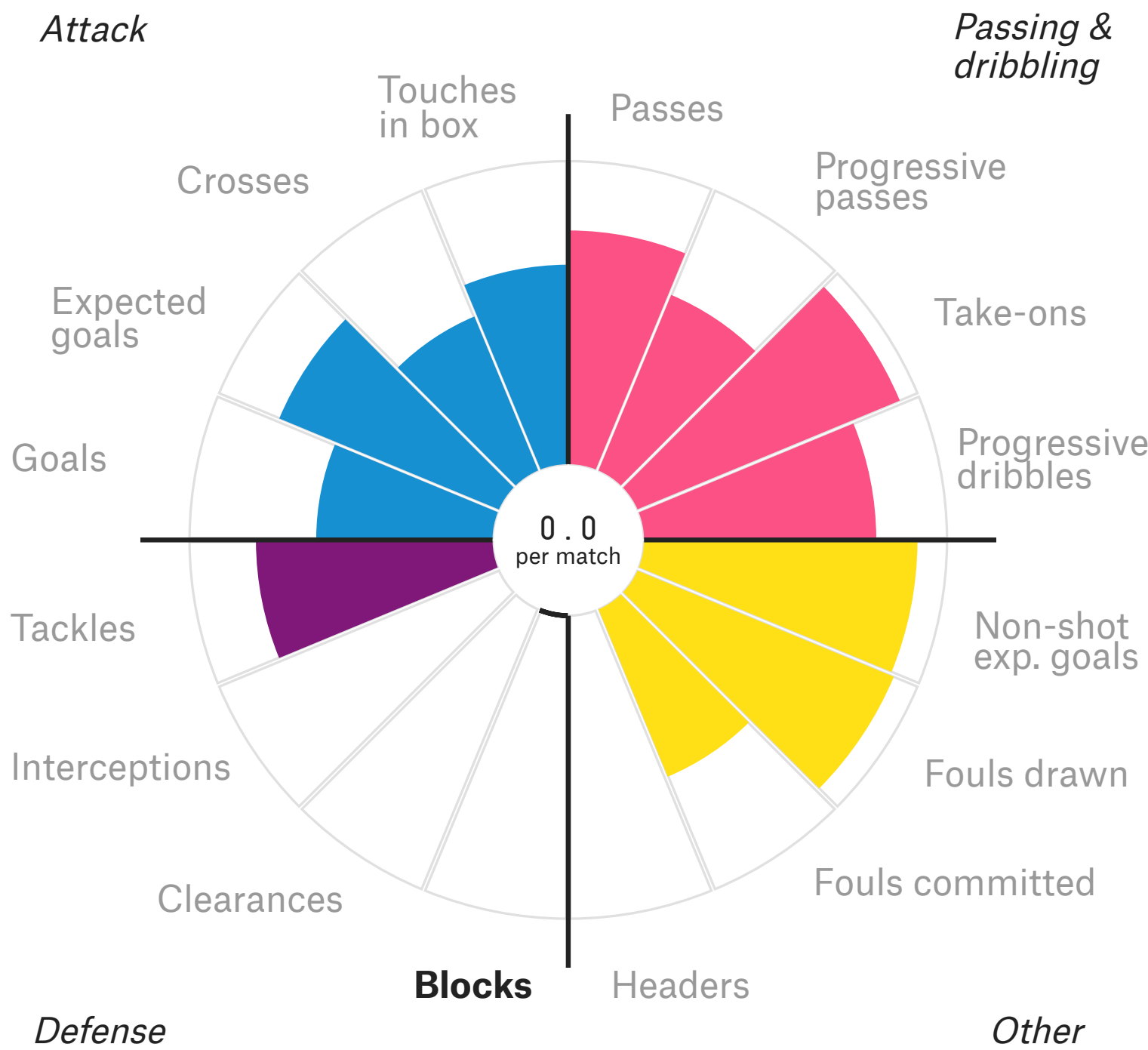


Motivating Example

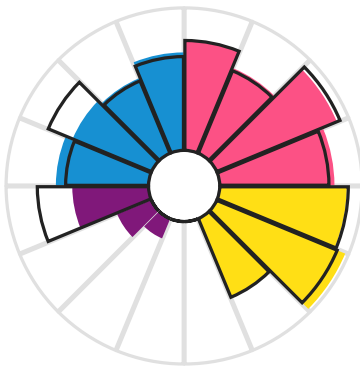


Lionel Messi

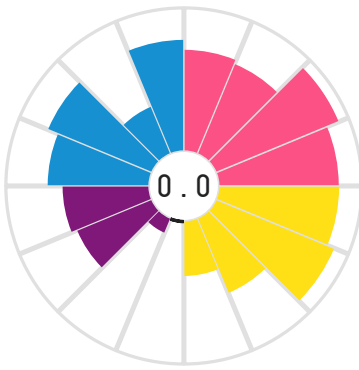
TEAM: ARGENTINA 2018
PLAY STYLE: DYNAMIC PLAYMAKER



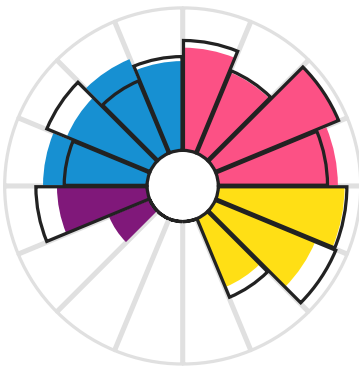
Most similar World Cup performances



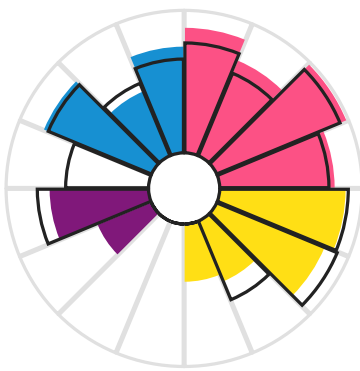
1. Ariel Ortega
ARGENTINA 1998



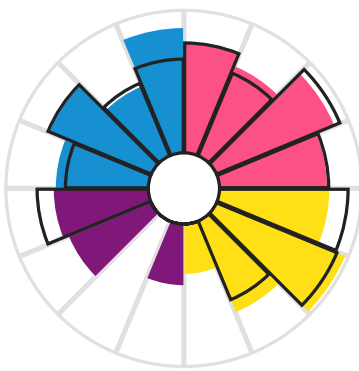
2. Eden Hazard
BELGIUM 2018



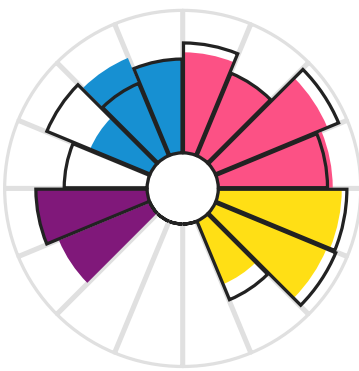
3. Lionel Messi
ARGENTINA 2014



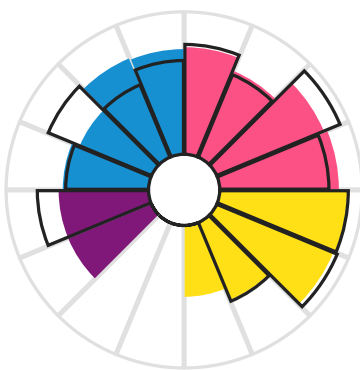
4. Lionel Messi
ARGENTINA 2010



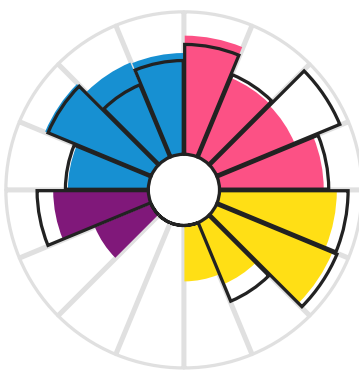
5. Diego Maradona
ARGENTINA 1982



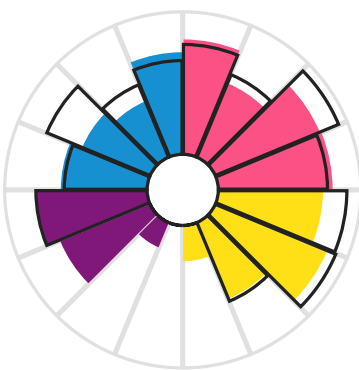
6. Franck Ribéry
FRANCE 2010



7. Fredrik Ljungberg
SWEDEN 2006

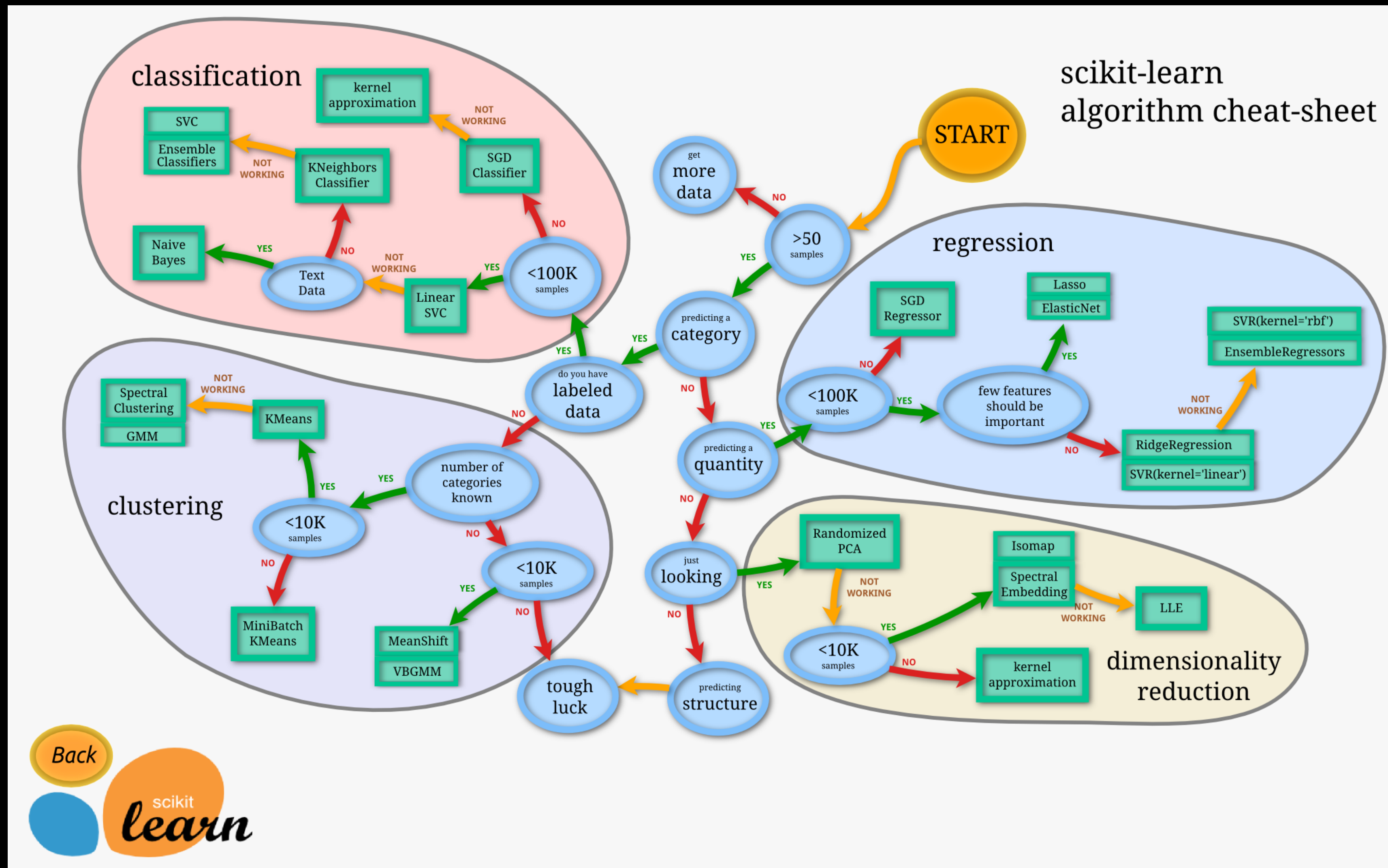


8. Cuauhtémoc Blanco
MEXICO 2002



9. Andrés Iniesta
SPAIN 2010

Where are we in the Data Science Landscape?



Thanks!