Tecnológico de Monterrey
Escuela de Ciencias Sociales
y Gobierno

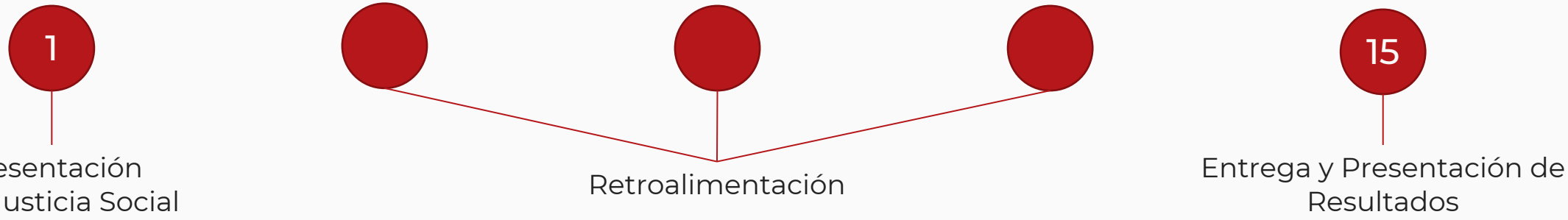
Ciencia de Datos para la Toma de Decisiones I

Mtro. René Rosado González
Director de Programa LTP

Nuestro Reto:

Desarrollar una metodología basada en datos para la generación de un índice que permita entender los procesos de desigualdad que ocurren en las comunidades de origen o retorno de las personas privadas de la libertad.

Nos reuniremos con el socio formador los días miércoles



Design Sprint

day 1

understand

- who are the users
- what are their needs
- what is the context
- competitor review
- formulate strategy

2

diverge

- envision
- develop lots of solutions
- ideate

3

decide

- choose the best idea
- storyboard the idea

4

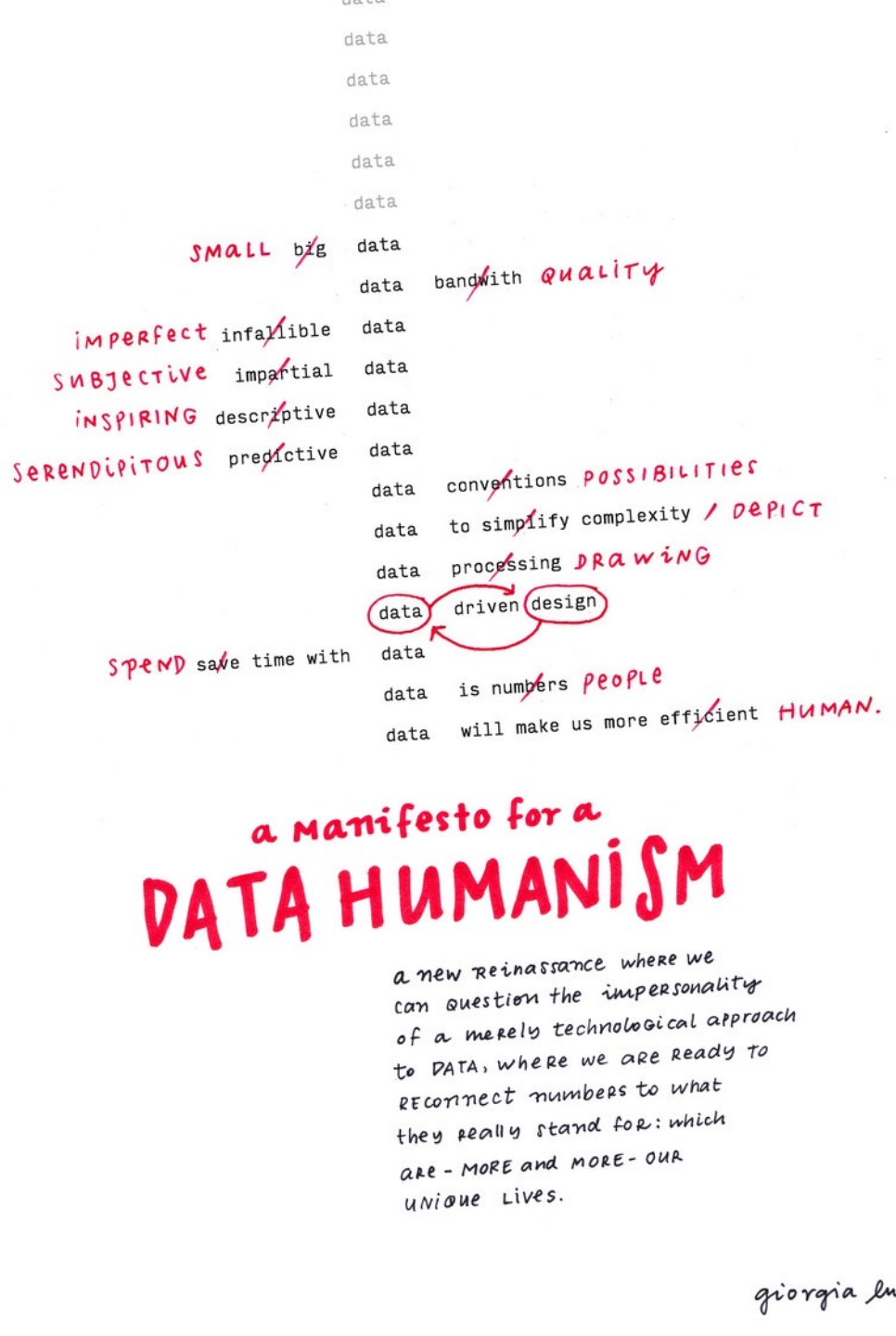
prototype

- build something quick and dirty to show to users
- focus on usability not making it beautiful

5

validate

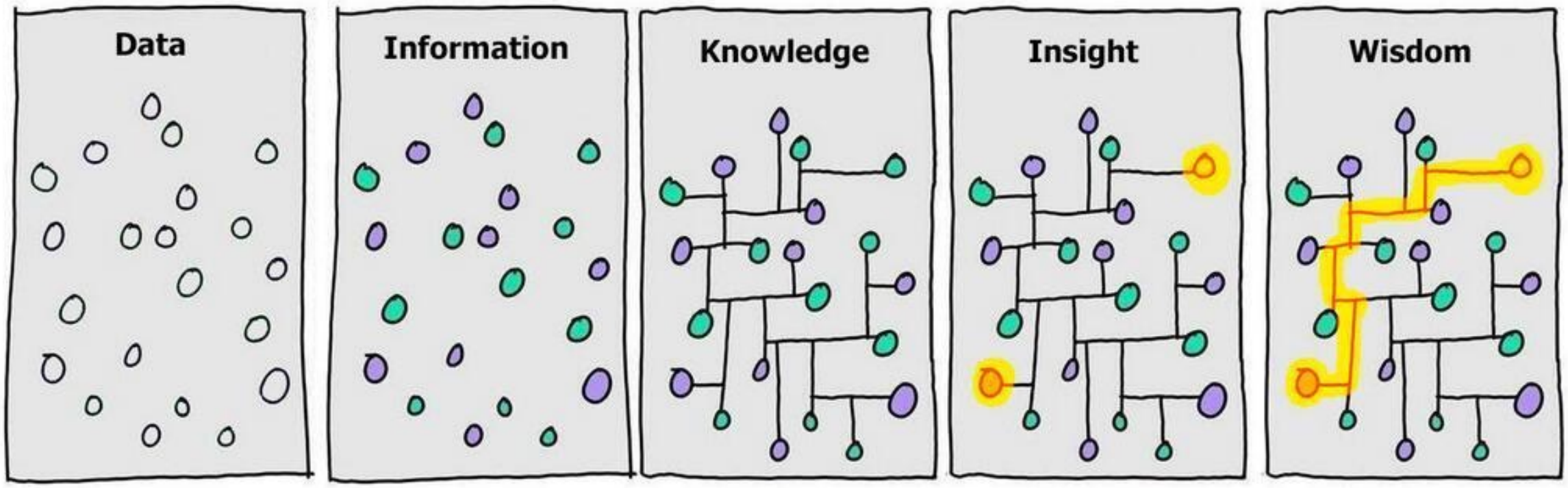
- show the prototype to real users outside the organisation
- learn what doesn't work



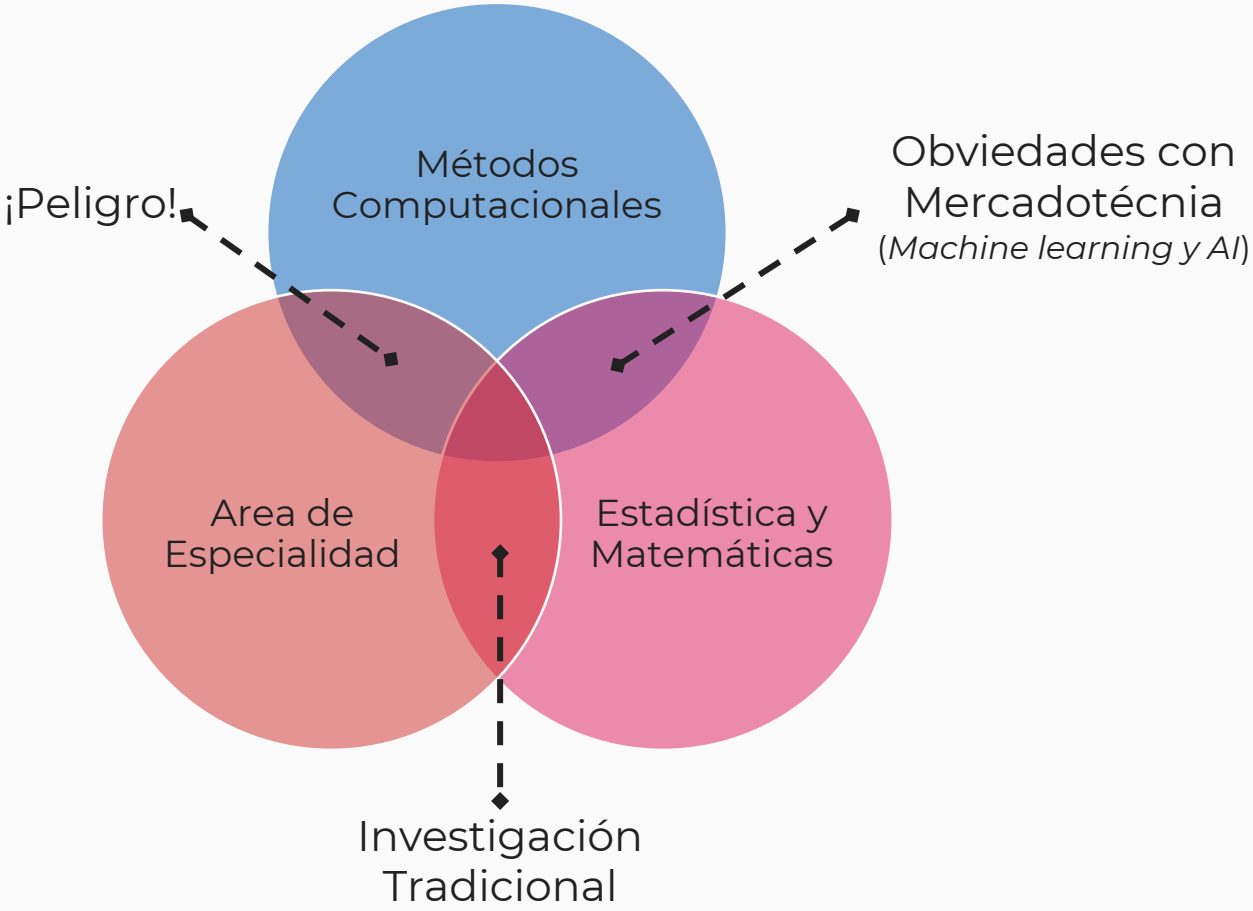
¿Qué es un dato?

Una expresión de deseo por almacenar una vivencia

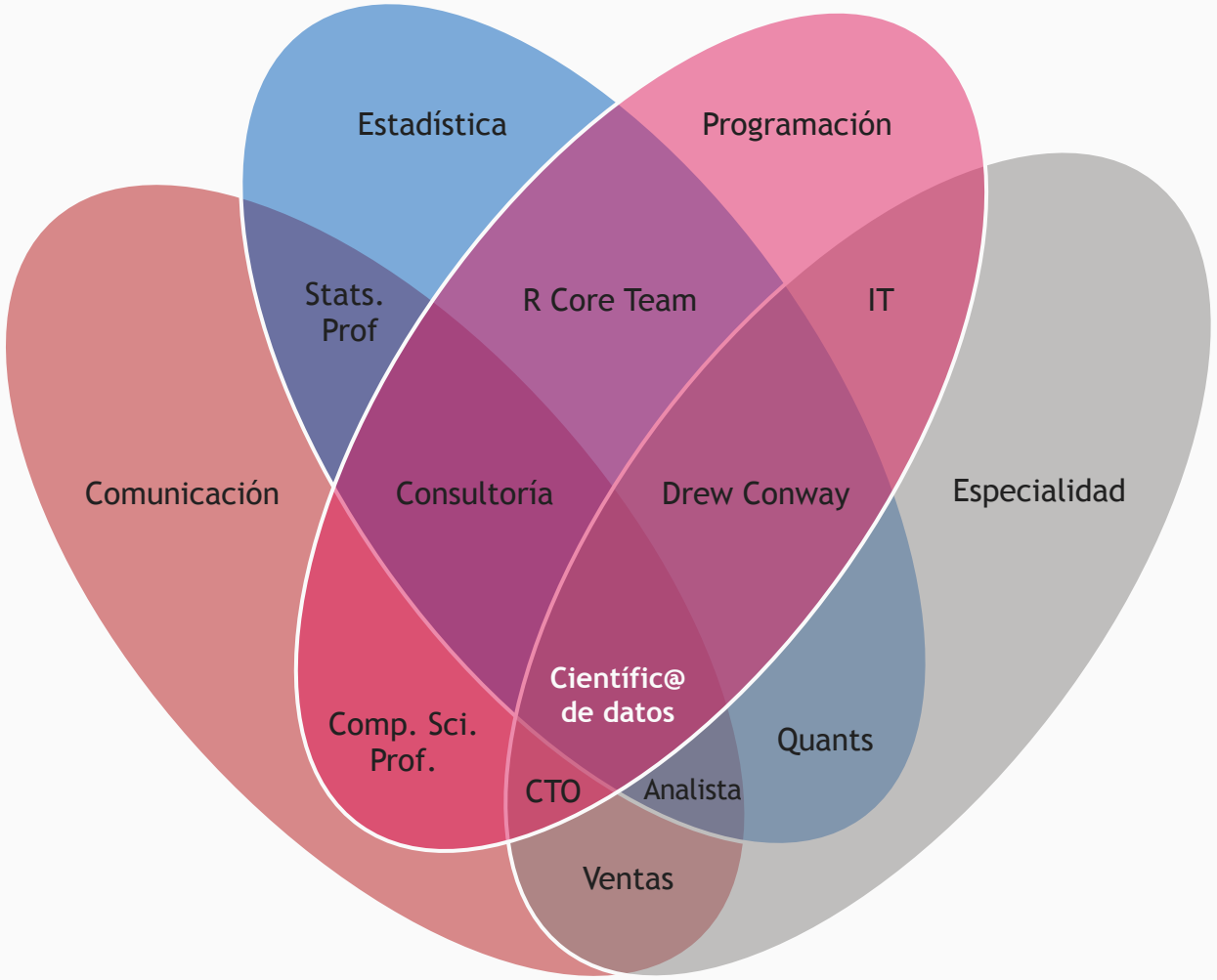
¿Qué hace la ciencia de datos?



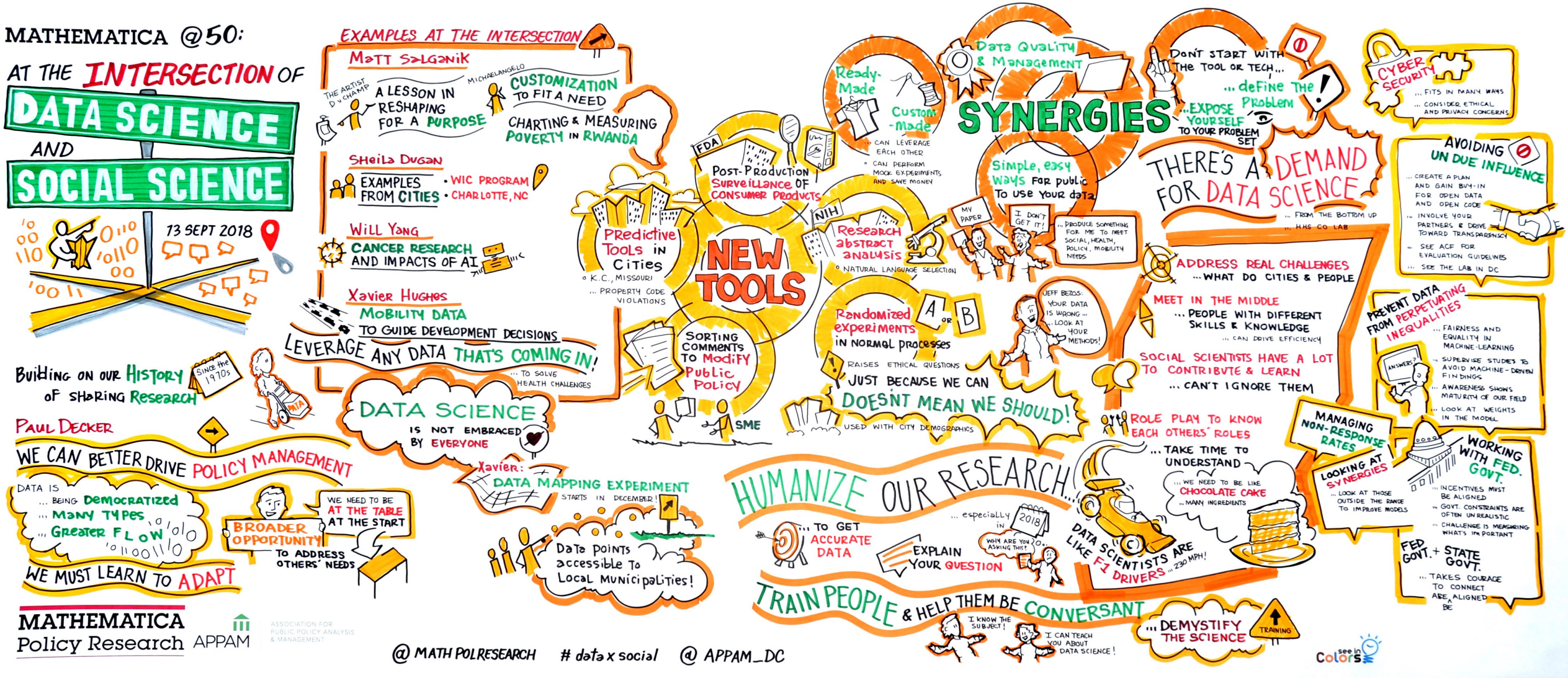
El diagrama de Drew Conway



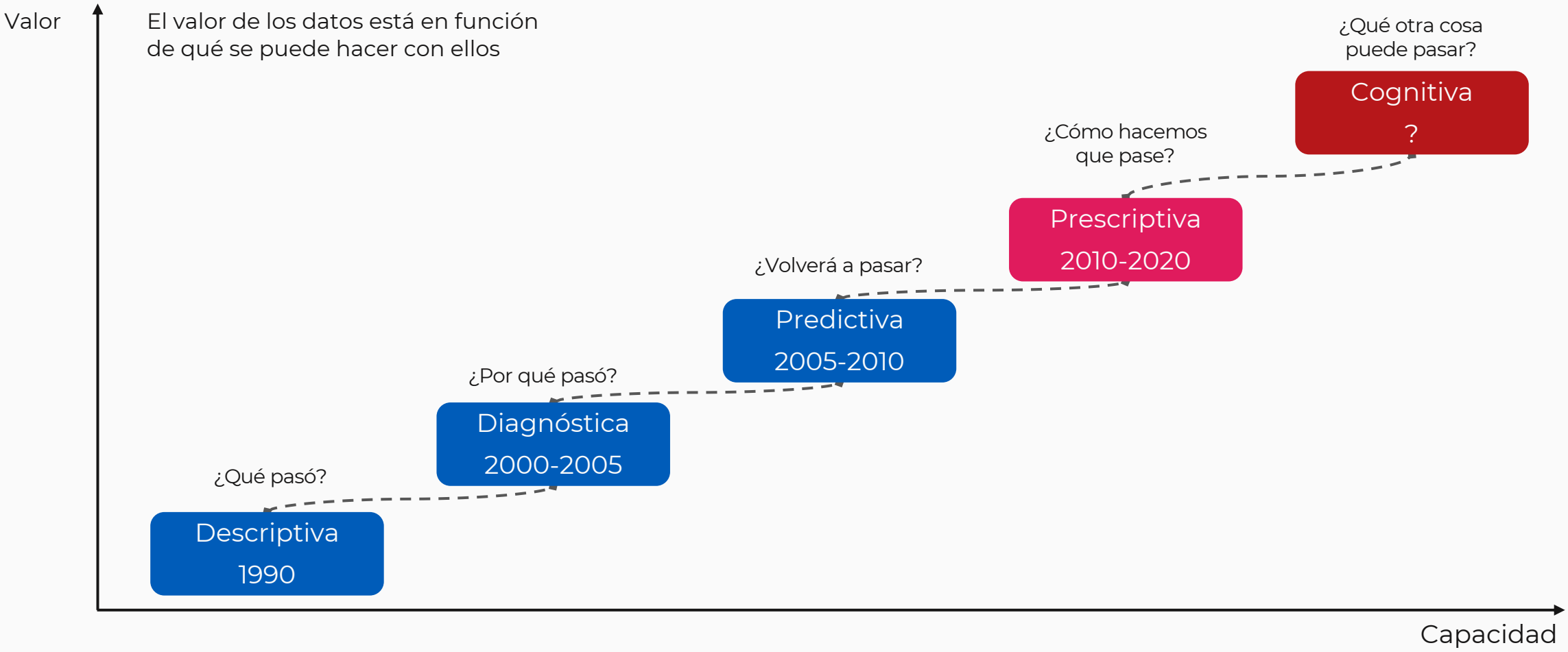
“There is no I in Team”



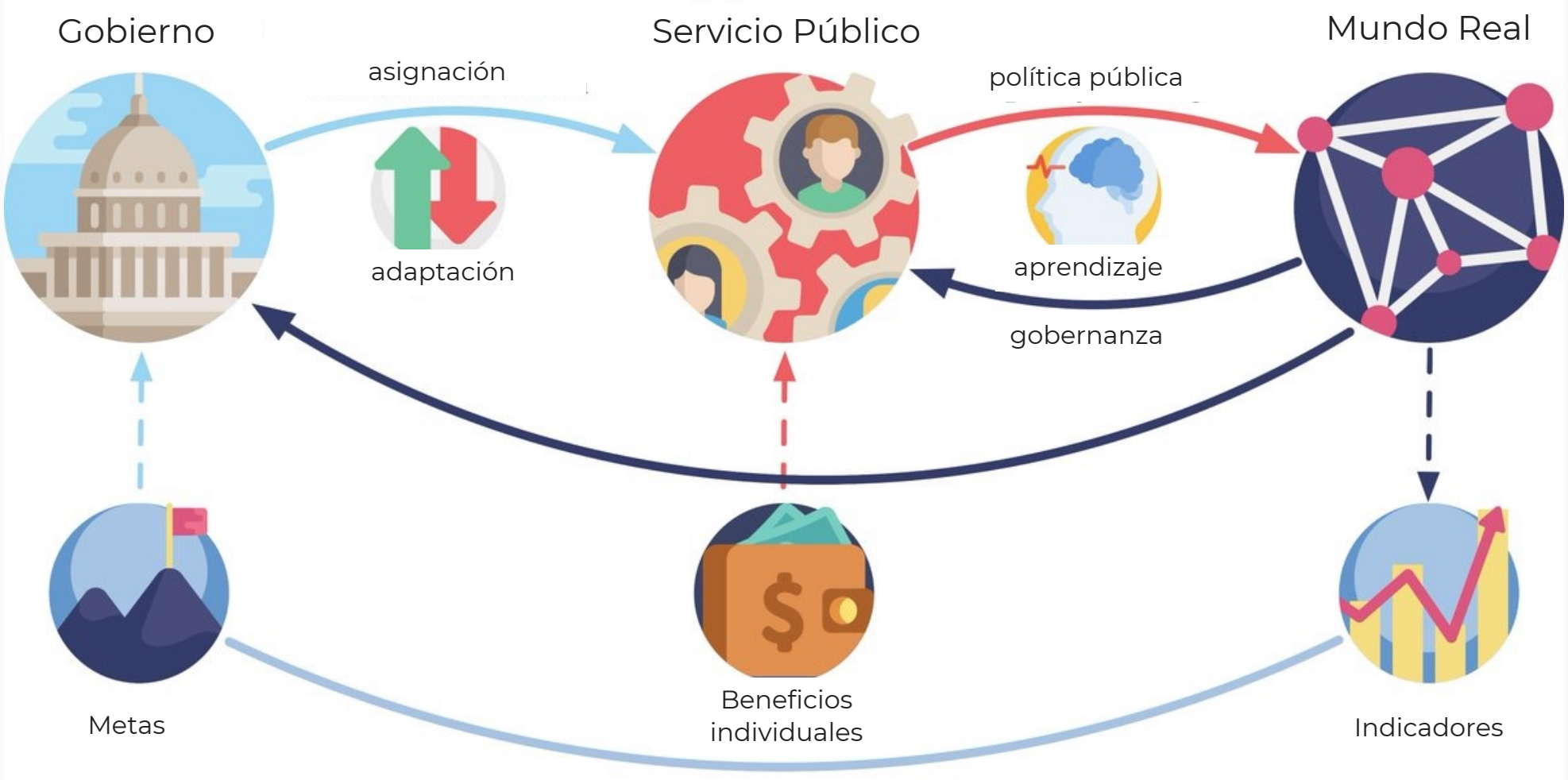
Cientific@ de Datos Ideal


MATHEMATICA @50:
AT THE INTERSECTION OF
DATA SCIENCE
AND
SOCIAL SCIENCE
13 SEPT 2018
EXAMPLES AT THE INTERSECTION
Matt Salganik
THE ARTIST DUCHAMP
A LESSON IN RESHAPING FOR A PURPOSE
MICHAELANGELO
CUSTOMIZATION TO FIT A NEED
CHARTING & MEASURING POVERTY IN RWANDA
Sheila Dugan
EXAMPLES FROM CITIES
• WIC PROGRAM
• CHARLOTTE, NC
Will Yang
CANCER RESEARCH AND IMPACTS OF AI
Xavier Hughes
MOBILITY DATA
TO GUIDE DEVELOPMENT DECISIONS
LEVERAGE ANY DATA THAT'S COMING IN!
... TO SOLVE HEALTH CHALLENGES
Building on our History of sharing Research
Since the 1970s
DATA SCIENCE
IS NOT EMBRACED BY EVERYONE
Paul Decker
WE CAN BETTER DRIVE POLICY MANAGEMENT
DATA IS
... BEING Democratized
... Many Types
... Greater Flow
WE MUST LEARN TO ADAPT
BROADER OPPORTUNITY
TO ADDRESS OTHERS' NEEDS
WE NEED TO BE AT THE TABLE AT THE START
Xavier:
DATA MAPPING EXPERIMENT
STARTS IN DECEMBER!
Data Points accessible To Local Municipalities!
FDA
Post-Production Surveillance OF Consumer Products
Predictive Tools in Cities
• K.C., MISSOURI
... PROPERTY CODE VIOLATIONS
NEW TOOLS
SORTING COMMENTS TO Modify Public Policy
SME
NIH
Research abstract analysis
• NATURAL LANGUAGE SELECTION
Randomized experiments IN NORMAL processes
A or B
RAISES ETHICAL QUESTIONS
JUST BECAUSE WE CAN DOESN'T MEAN WE SHOULD!
USED WITH CITY DEMOGRAPHICS
Ready-Made
Custom-Made
... CAN LEVERAGE EACH OTHER
• CAN PERFORM MOCK EXPERIMENTS AND SAVE MONEY
Data Quality & Management
SYNERGIES
Simple, easy ways FOR public To use Your data
MY PAPER
I DON'T GET IT!
... PRODUCE SOMETHING FOR ME TO MEET SOCIAL, HEALTH, POLICY, MOBILITY NEEDS
Don't start with The tool or tech...
... define The Problem
...Expose Yourself To Your Problem Set
THERE'S A DEMAND FOR DATA SCIENCE
... FROM THE BOTTOM UP
... HHS CO LAB
CYBER SECURITY
... FITS IN MANY WAYS
... CONSIDER ETHICAL AND PRIVACY CONCERNS
AVOIDING UN DUE INFLUENCE
... CREATE A PLAN AND GAIN BUY-IN FOR OPEN DATA AND OPEN CODE
... INVOLVE YOUR PARTNERS & DRIVE TOWARD TRANSPARENCY
... SEE ACF FOR EVALUATION GUIDELINES
... SEE THE LAB IN DC
ADDRESS REAL CHALLENGES
...WHAT DO CITIES & PEOPLE
MEET IN THE MIDDLE
... PEOPLE WITH DIFFERENT SKILLS & KNOWLEDGE
... CAN DRIVE EFFICIENCY
SOCIAL SCIENTISTS HAVE A LOT TO CONTRIBUTE & LEARN
... CAN'T IGNORE THEM
JEFF BEZOS: YOUR DATA IS WRONG ... LOOK AT YOUR METHODS!
PREVENT DATA FROM PERPETUATING INEQUALITIES
... FAIRNESS AND EQUALITY IN MACHINE-LEARNING
... SUPERVISE STUDIES TO AVOID MACHINE-DRIVEN FINDINGS
... AWARENESS SHOWS MATURITY OF OUR FIELD
... LOOK AT WEIGHTS IN THE MODEL
ANSWERS?
MANAGING NON-RESPONSE RATES
LOOKING AT SYNERGIES
... LOOK AT THOSE OUTSIDE THE RANGE TO IMPROVE MODELS
WORKING WITH FED. GOVT.
... INCENTIVES MUST BE ALIGNED
... GOVT. CONSTRAINTS ARE OFTEN UNREALISTIC
... CHALLENGE IS MEASURING WHAT'S IMPORTANT
FED GOVT. + STATE GOVT.
... TAKES COURAGE TO CONNECT ARE ALIGNED BE
HUMANIZE OUR RESEARCH...
... TO GET ACCURATE DATA
EXPLAIN YOUR QUESTION
... especially in 2018
WHY ARE YOU ASKING THIS?
ROLE PLAY TO KNOW EACH OTHERS' ROLES
... TAKE TIME TO UNDERSTAND
... WE NEED TO BE LIKE CHOCOLATE CAKE
... MANY INGREDIENTS
DATA SCIENTISTS ARE LIKE F1 DRIVERS ... 230 MPH!
TRAIN PEOPLE & HELP THEM BE CONVERSANT
I KNOW THE SUBJECT!
I CAN TEACH YOU ABOUT DATA SCIENCE!
... DEMYSTIFY THE SCIENCE
TRAINING
MATHEMATICA
Policy Research
APPAM
ASSOCIATION FOR PUBLIC POLICY ANALYSIS & MANAGEMENT
@ MATH POLRESEARCH
# data x social
@ APPAM_DC
see in Colors

El valor de los datos en la políca pública



Diseño e implementación de política pública



Tecnológico de Monterrey
Escuela de Ciencias Sociales y Gobierno

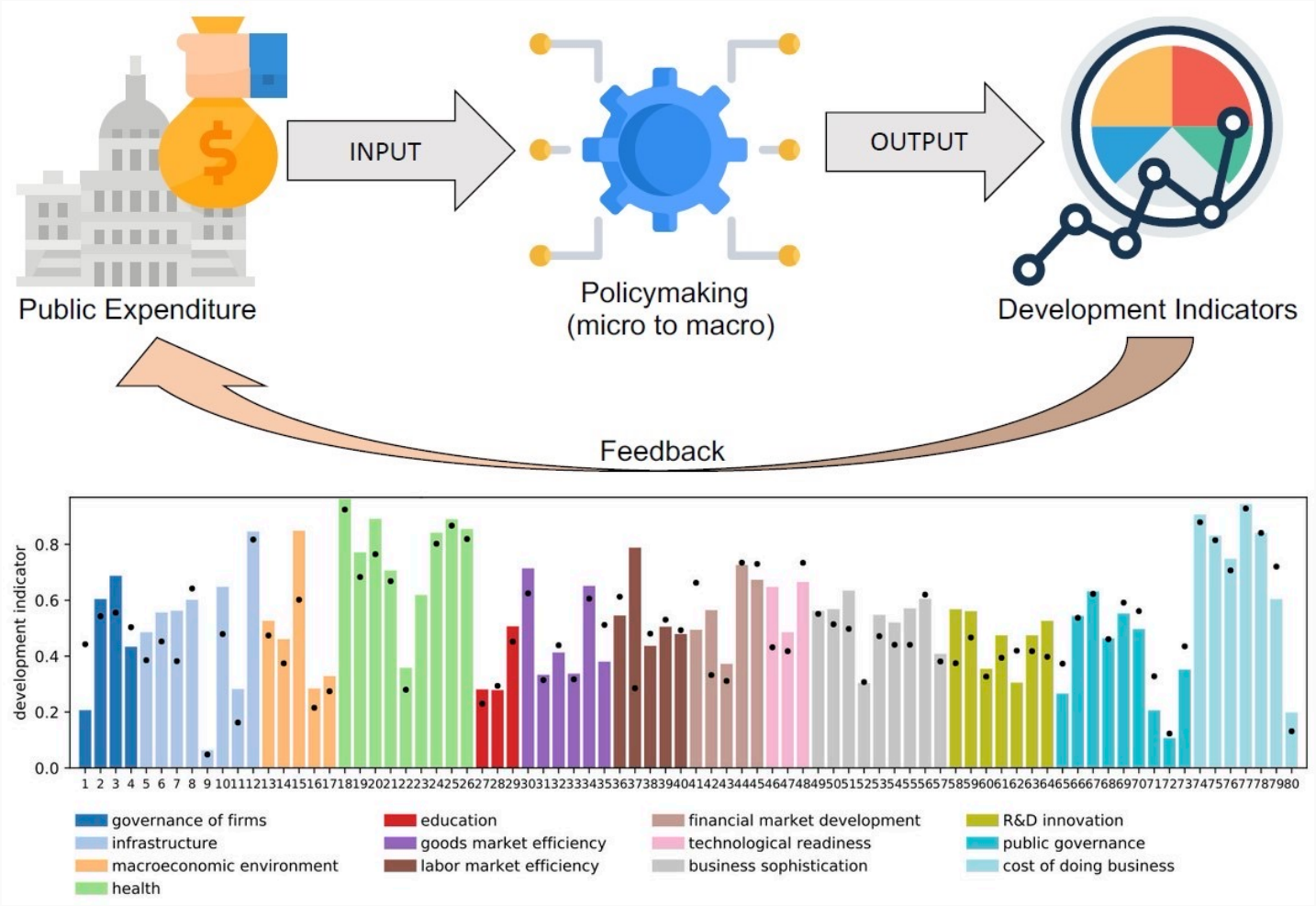
Policy Priority Inference

for sustainable development

PhD. Omar A. Guerrero

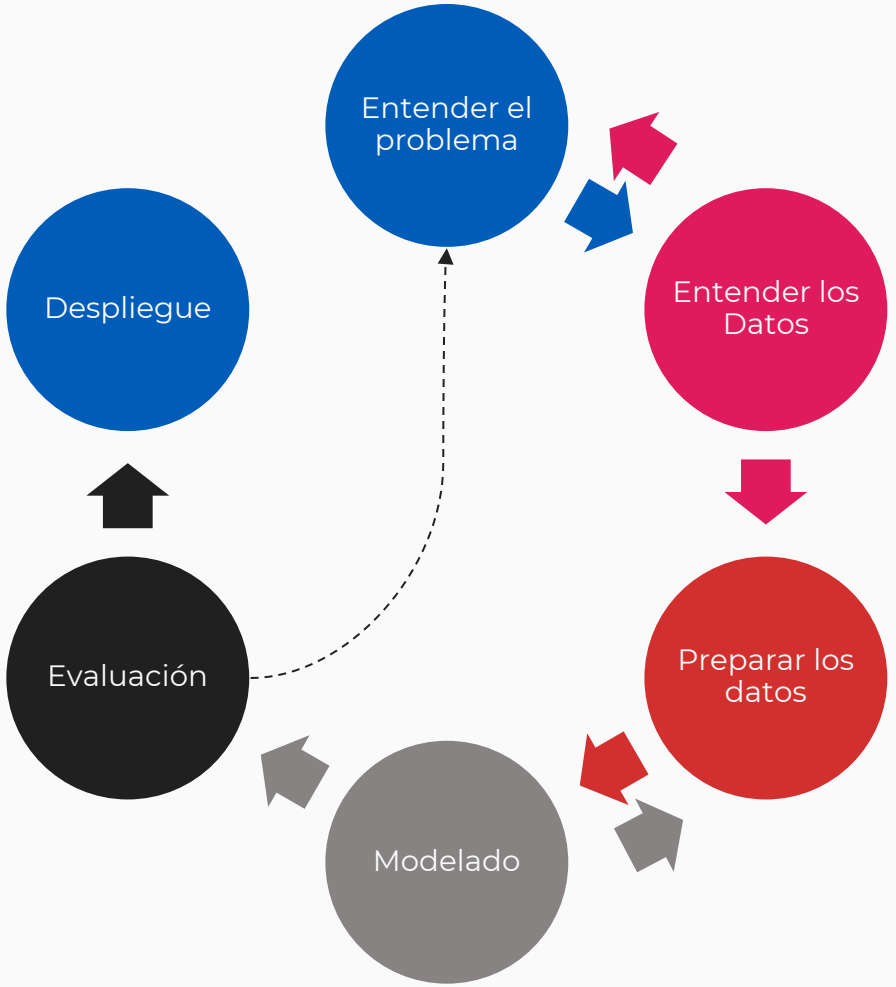
Head of Computational Social Science Research

The Alan Turing Institute



CRISP - DM

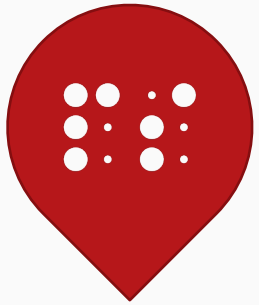
Cross-industry Standard Process for Data Mining



Objetivos de un Proyecto de Datos

Automatizar

Descurbir Patrones



Explicar

Optimizar

Predecir

Aprendizaje de Máquina

Machine Learning

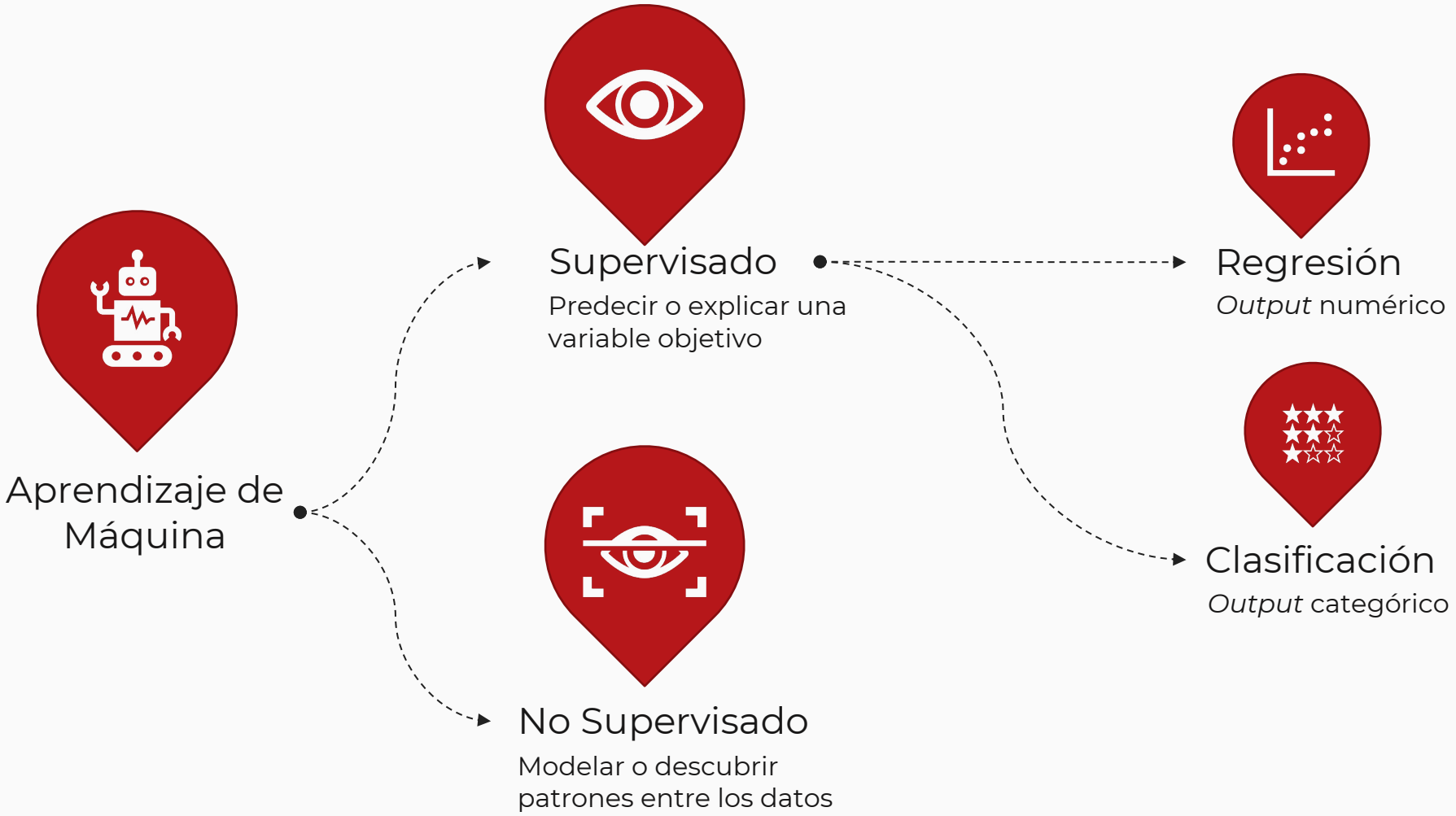
Métodos Computacionales

Datos

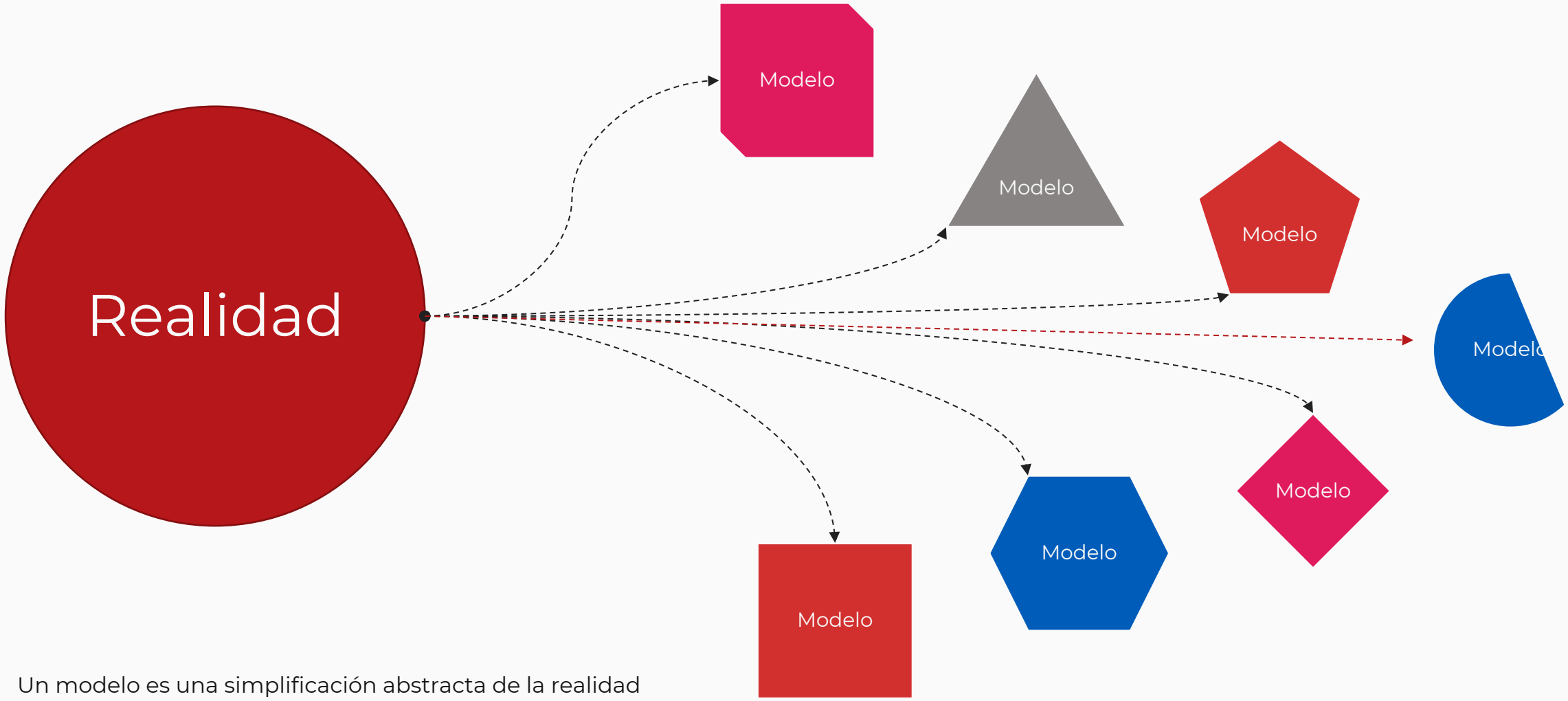
Herramientas para Toma de Tecisiones

Aprendizaje de Máquina

Machine Learning

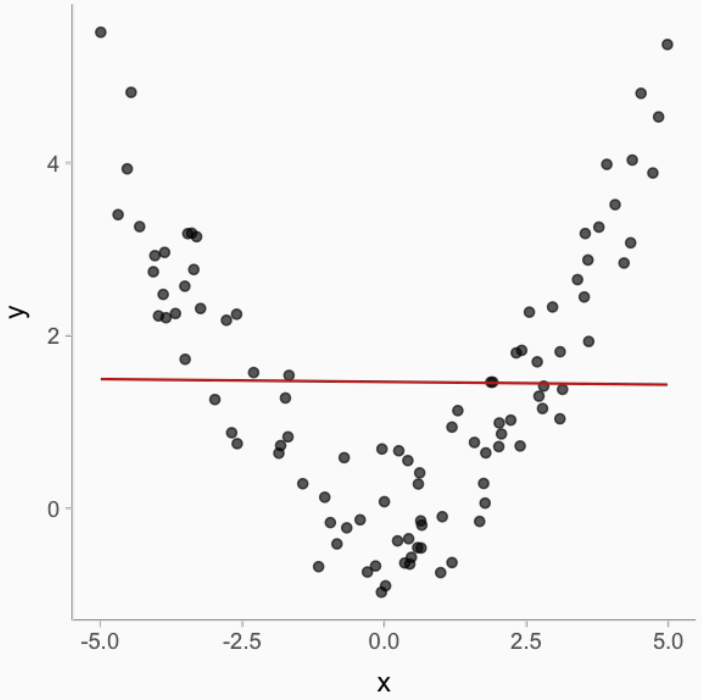


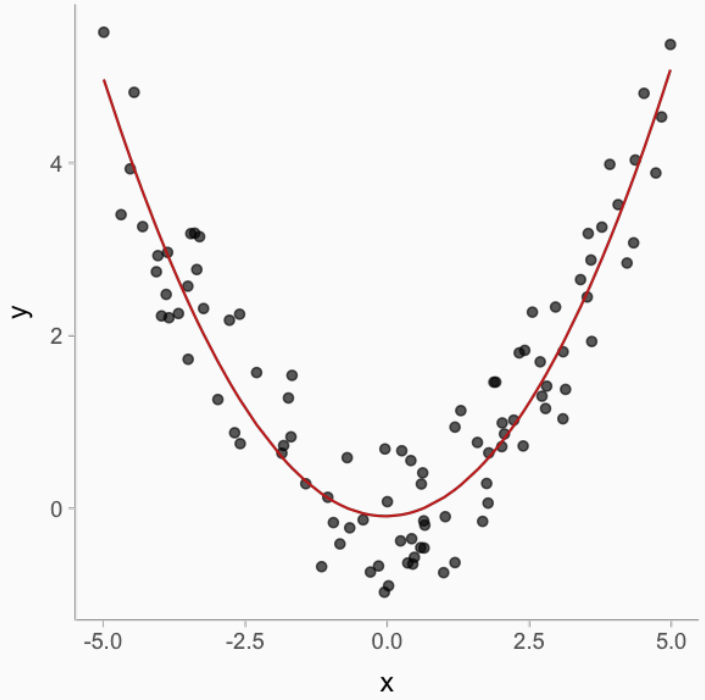
¿Qué es un modelo?

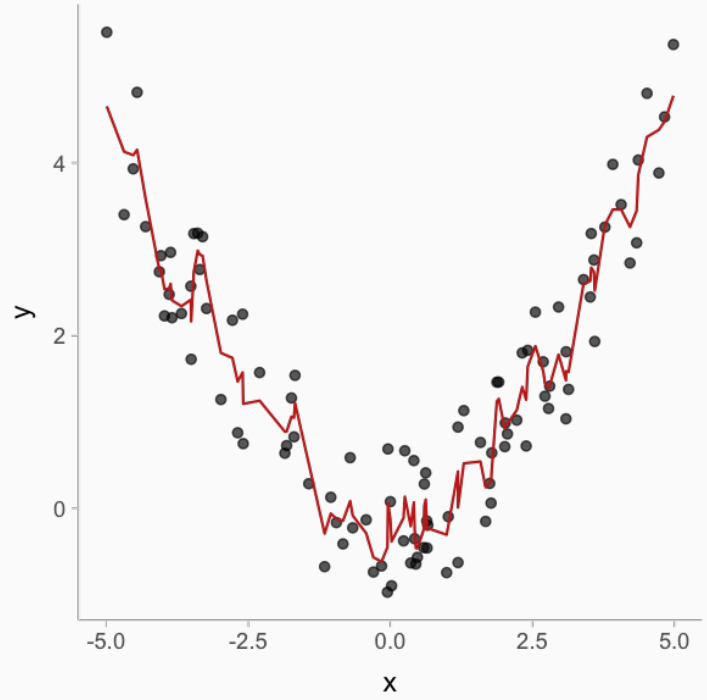


Un modelo es una simplificación abstracta de la realidad

La elección de un modelo ideal

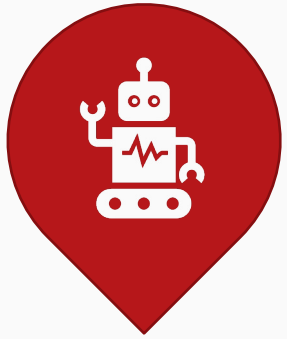






Funciones de Pérdida

Loss functions



Aprendizaje

$$\min\left(f\left(y-\hat{y}\right)\right)$$

Evaluación

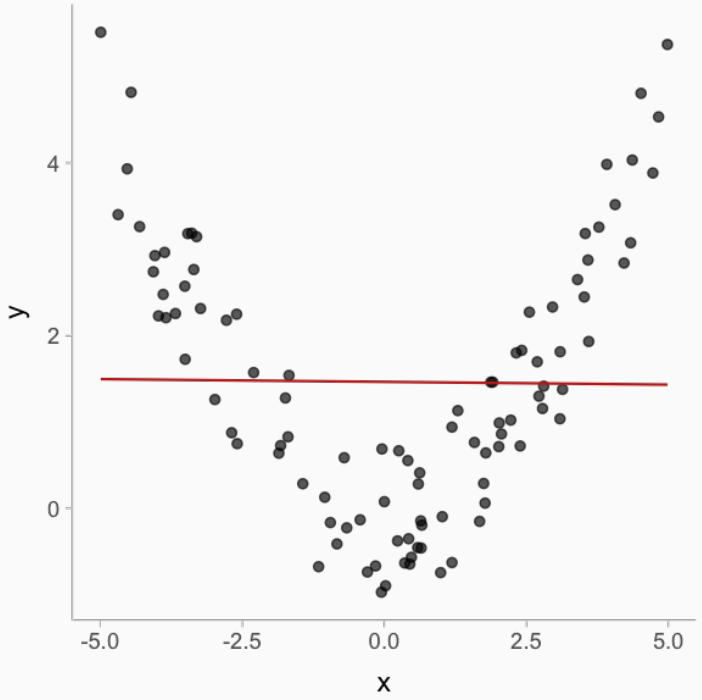
Ejemplo: Error Cuadrático Medio

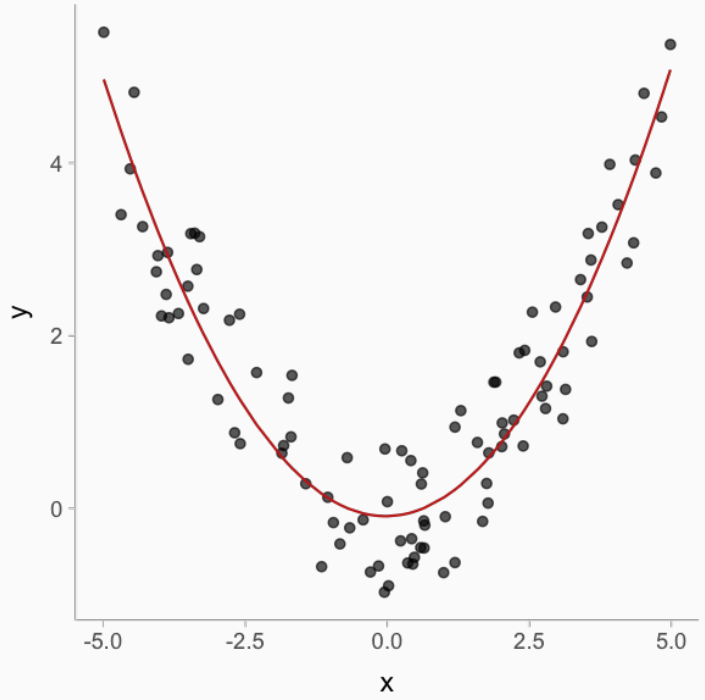
$$MSE(\hat{y}) = E[(\hat{y} - y)^2]$$

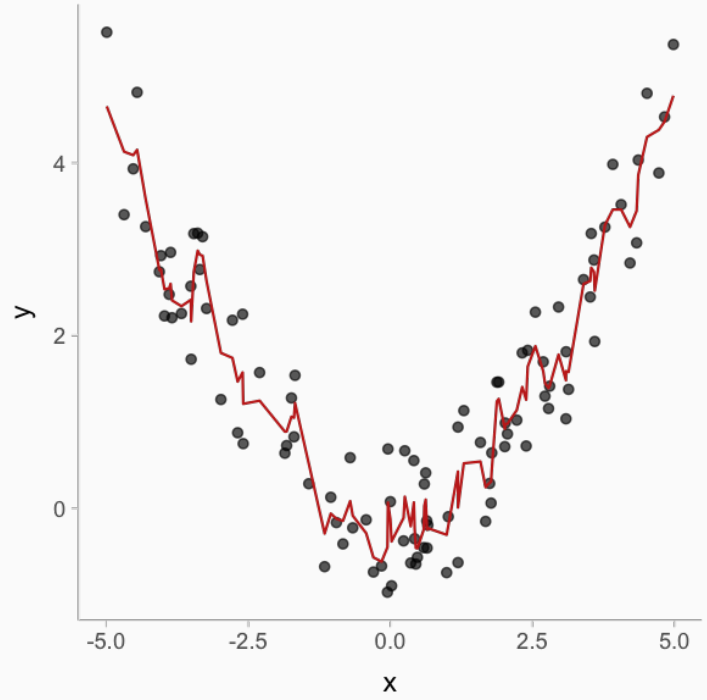
$$MSE(\hat{y}) = \underbrace{E[(\hat{y} - E(\hat{y}))^2]}_{\text{Varianza}} + \underbrace{E[(E(\hat{y}) - y)^2]}_{\text{Sesgo}}$$

Varianza

Qué tanto se acopla el modelo al recibir nueva información



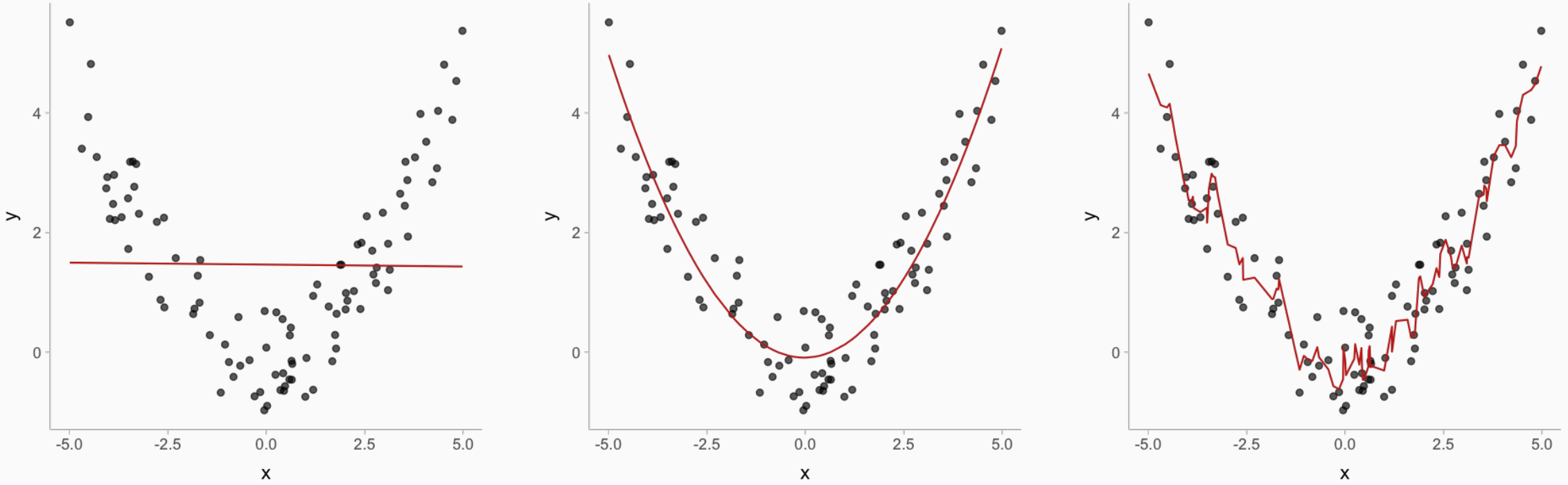




Varianza Baja ◄----------► Varianza Alta

Sesgo

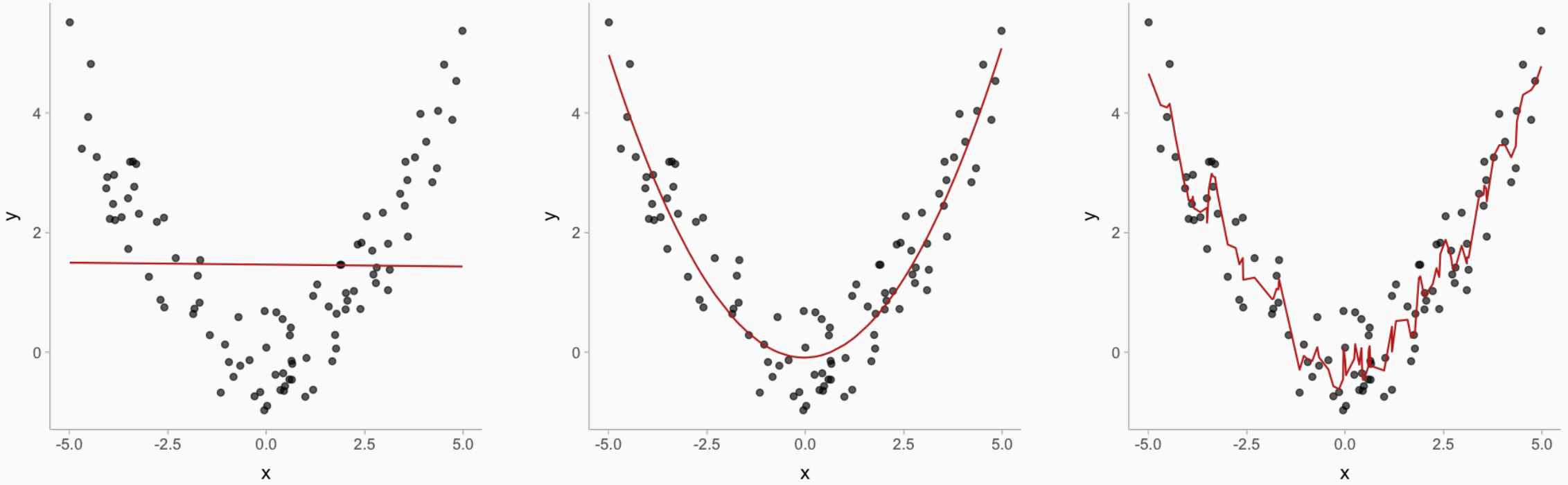
Qué tan lejos se encuentra el valor estimado del valor real



Sesgo Alto ⟷ Sesgo Bajo

La elección de un modelo ideal

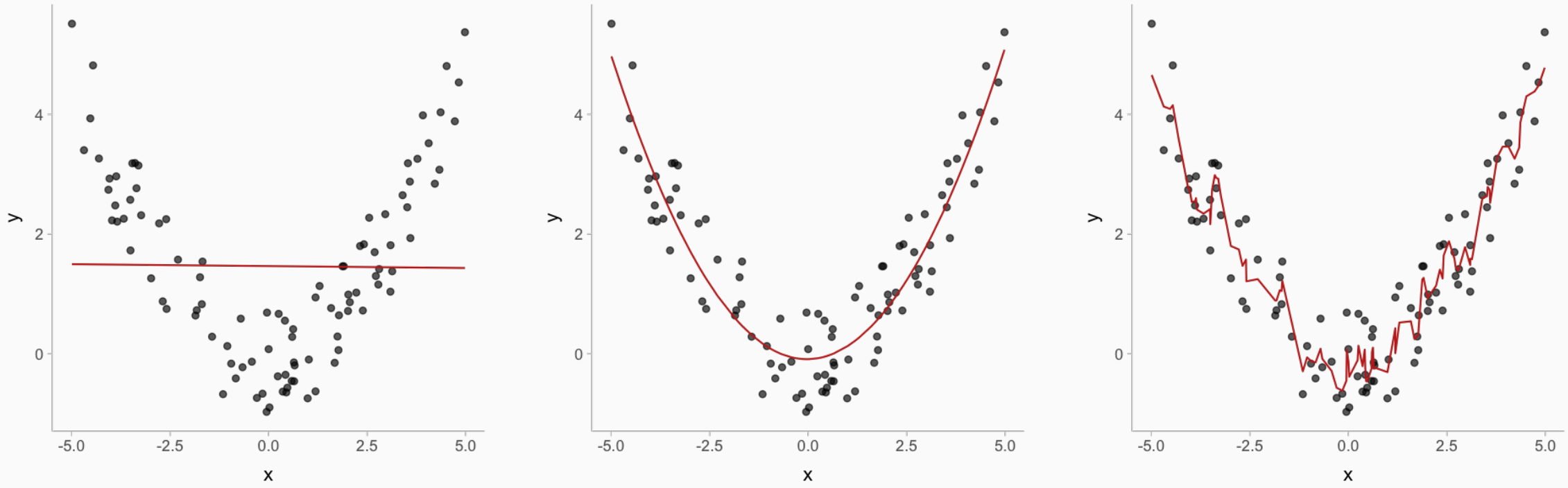
Buscamos un modelo que aprenda de los datos con los que se le alimentaron
y que sea capaz de generalizar las predicciones ante nuevos datos



Subajuste ◄----------► Sobreajuste

	Poca Varianza	Mucha Varianza
Sesgo Bajo		
Sesgo Alto		

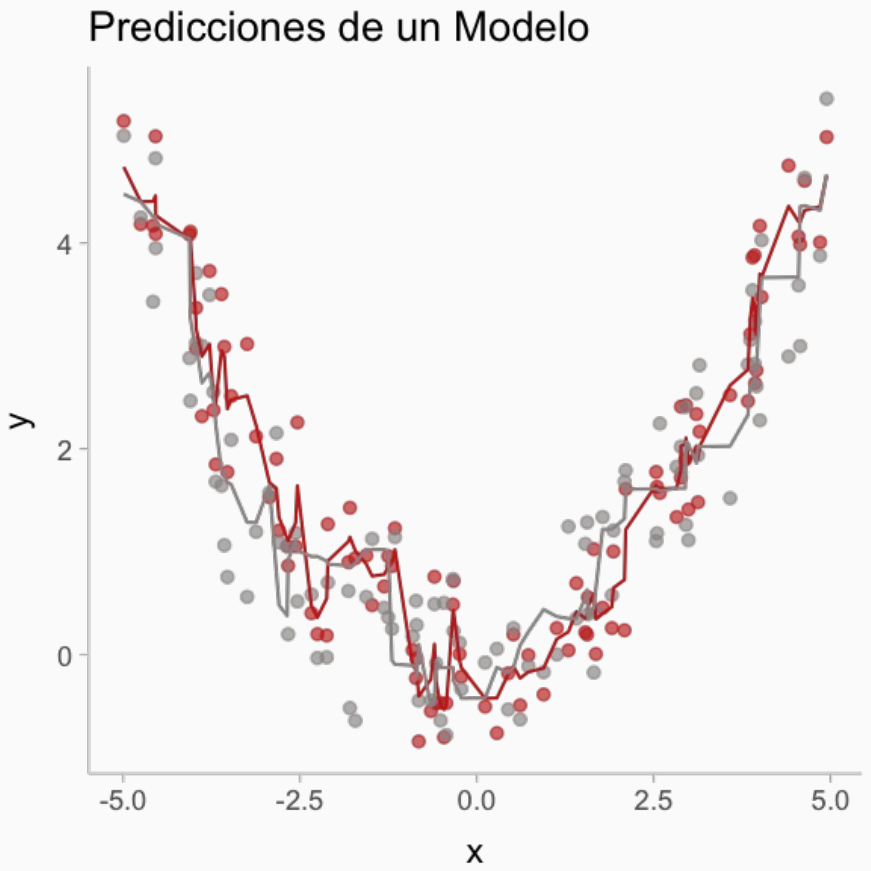
La elección de un modelo ideal

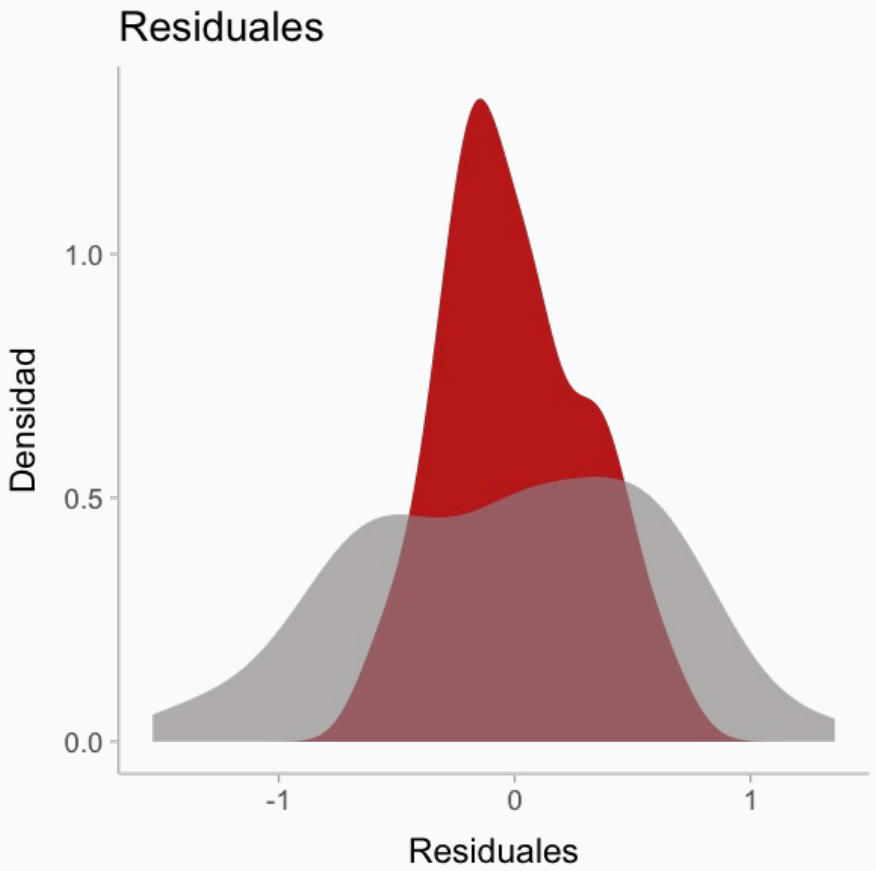


Uno Modelo Sin Validar

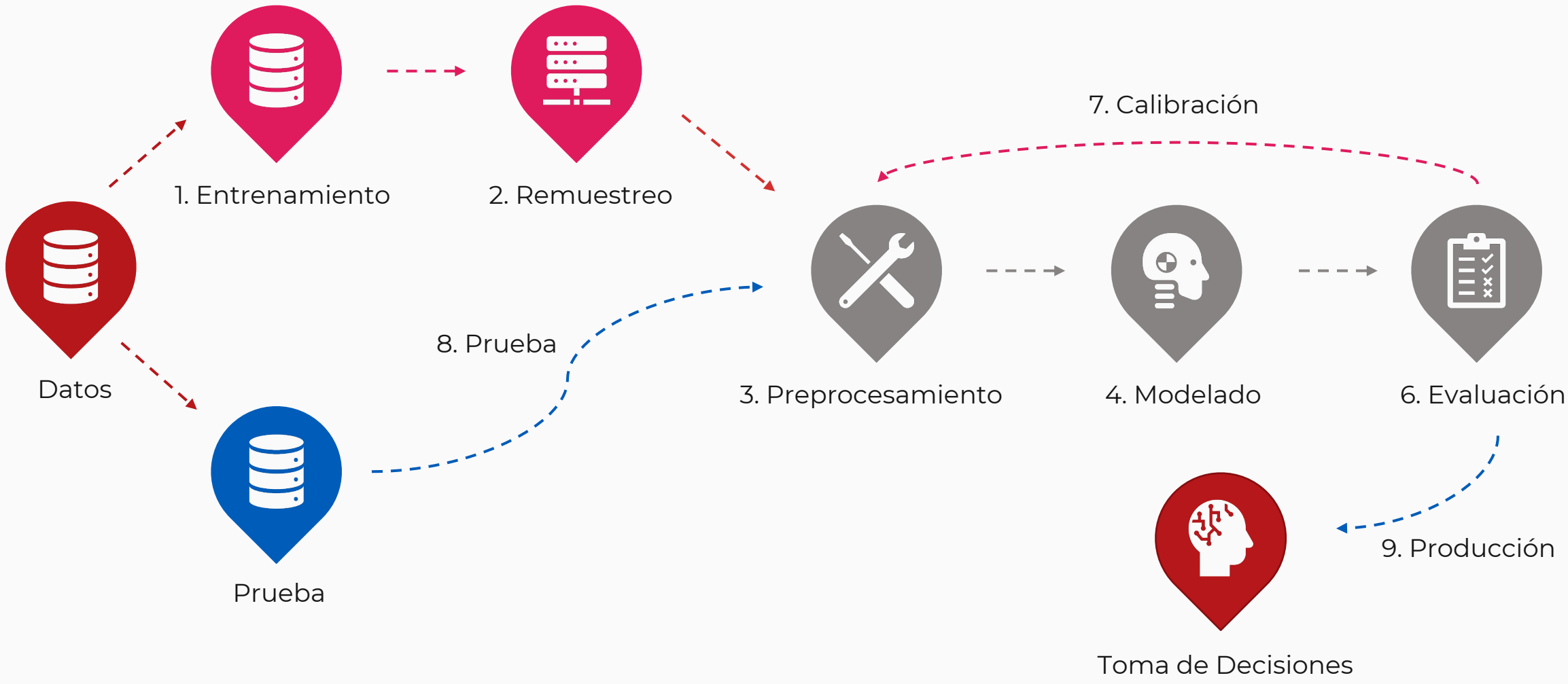
Datos Conocidos

Datos Nuevos



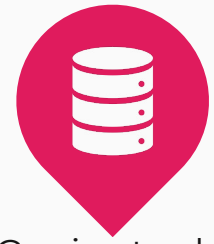


Encontrar el mejor modelo



Conjuntos de Entrenamiento y Prueba

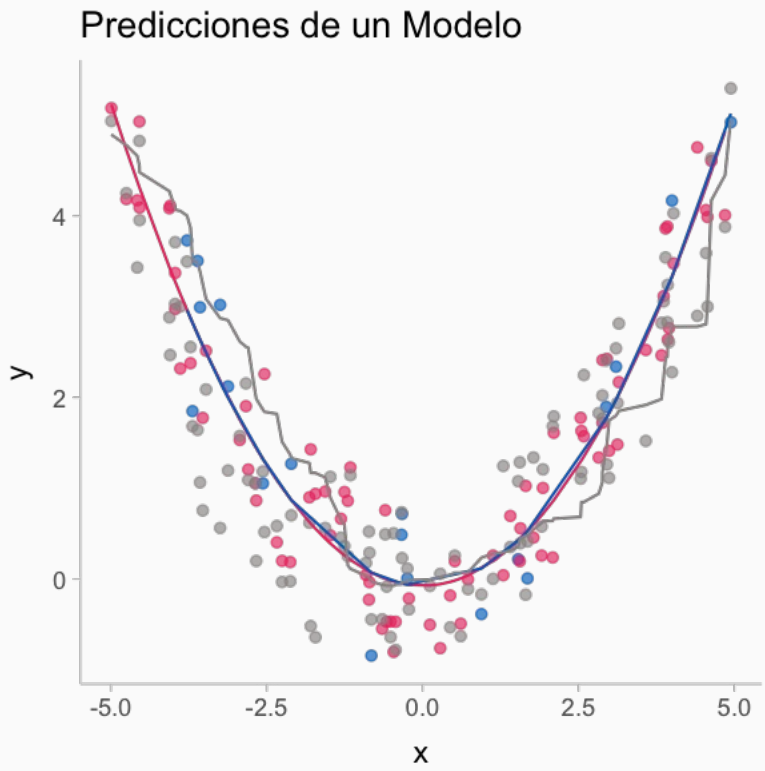
Train & Test Splits

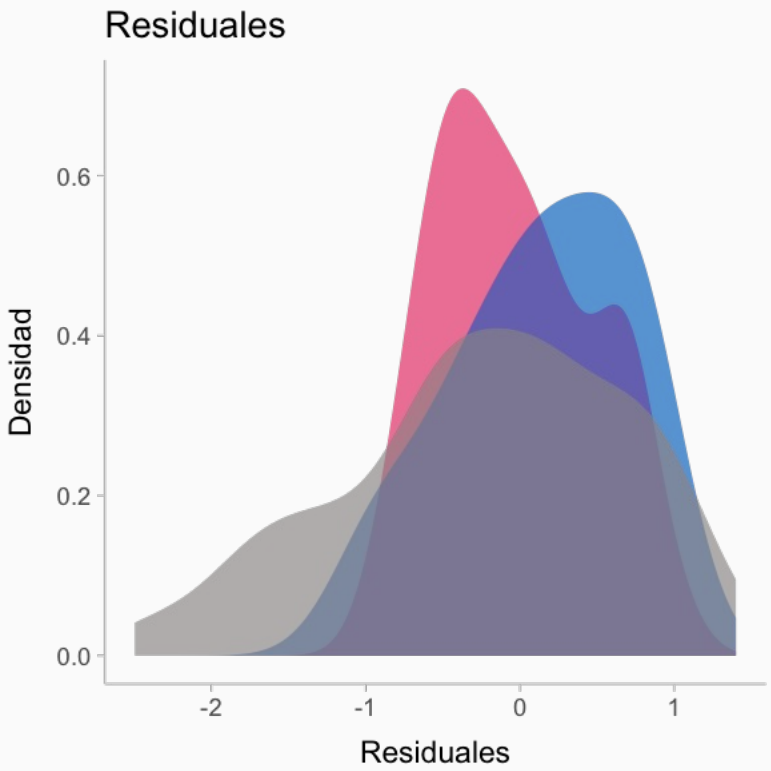


Conjunto de Entrenamiento

Conjunto de Prueba

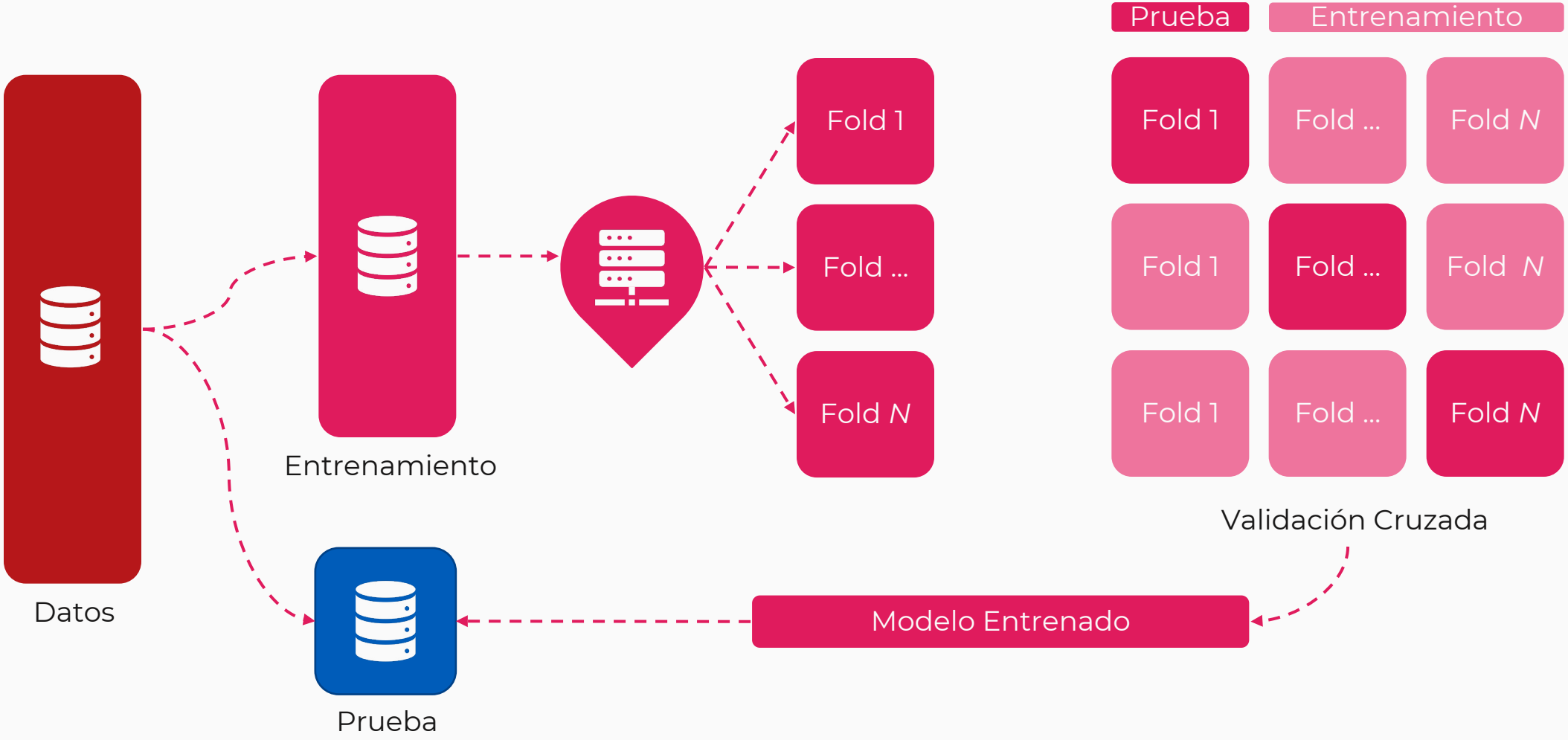
Datos Nuevos



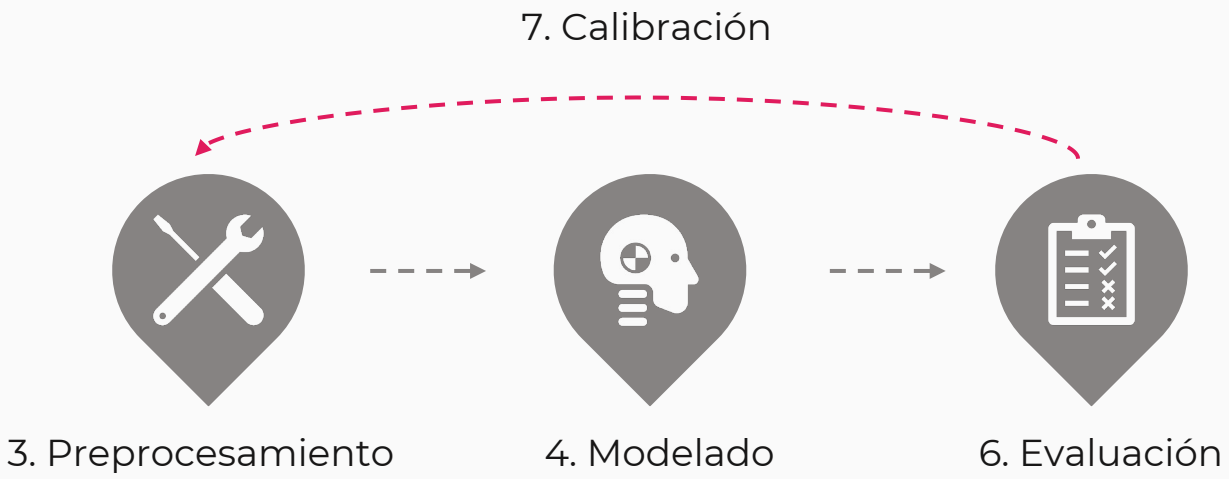


Validación Cruzada

Cross Validation



Este no es un curso de inferencia causal



Límites del Aprendizaje de Máquina

Limits of Machine Learning

No prevé soluciones a largo plazo

Soluciones específicas para casos específicos

Optimizaciones en un sistema dado