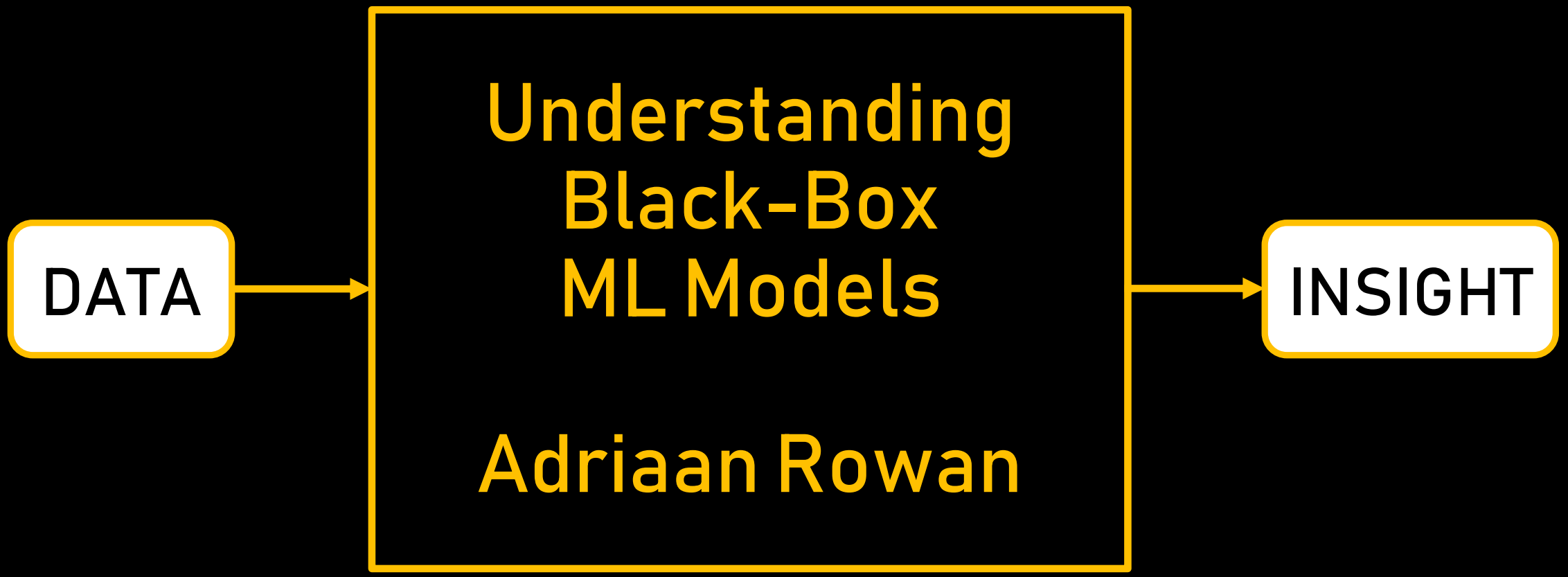
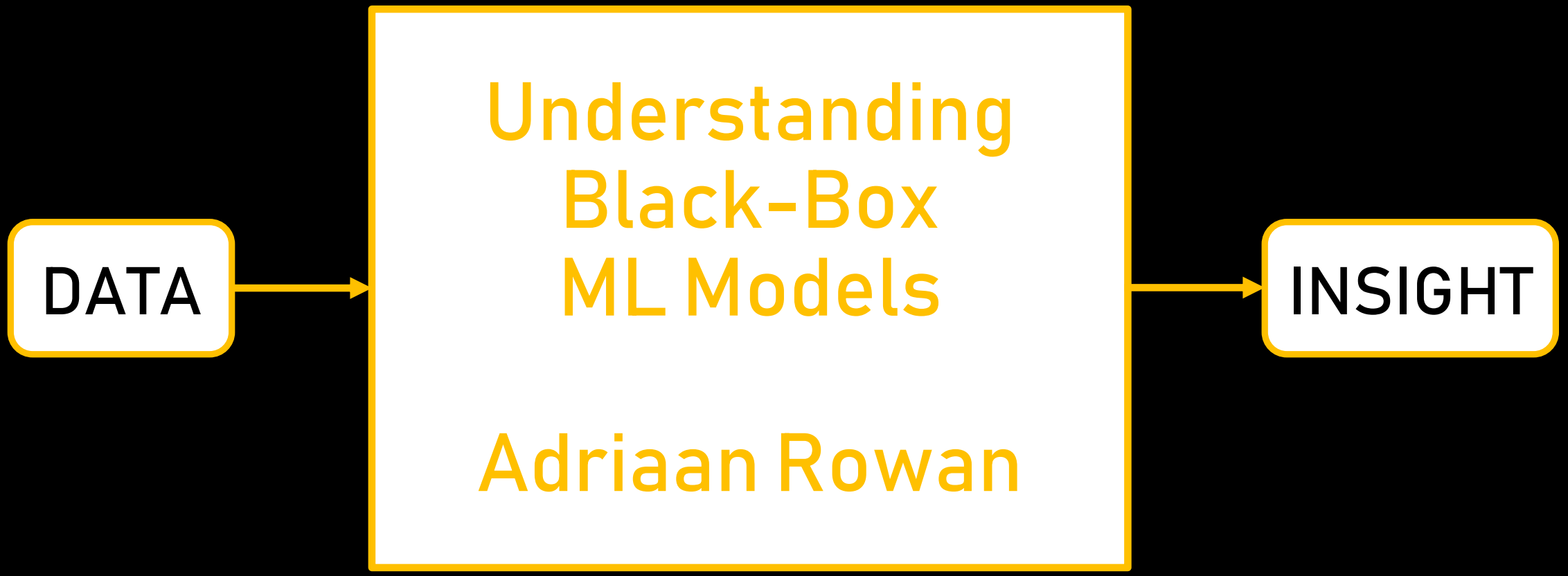
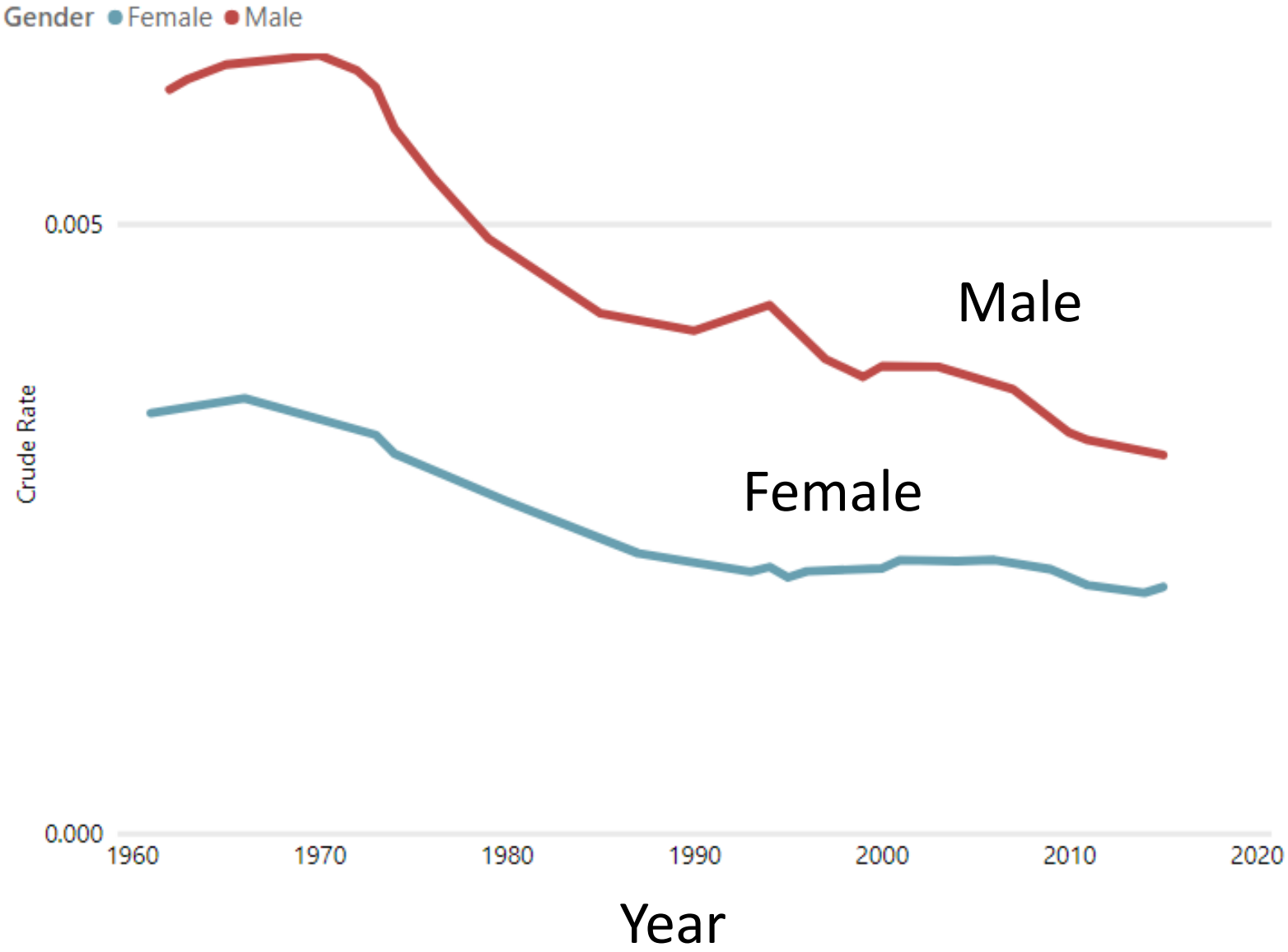
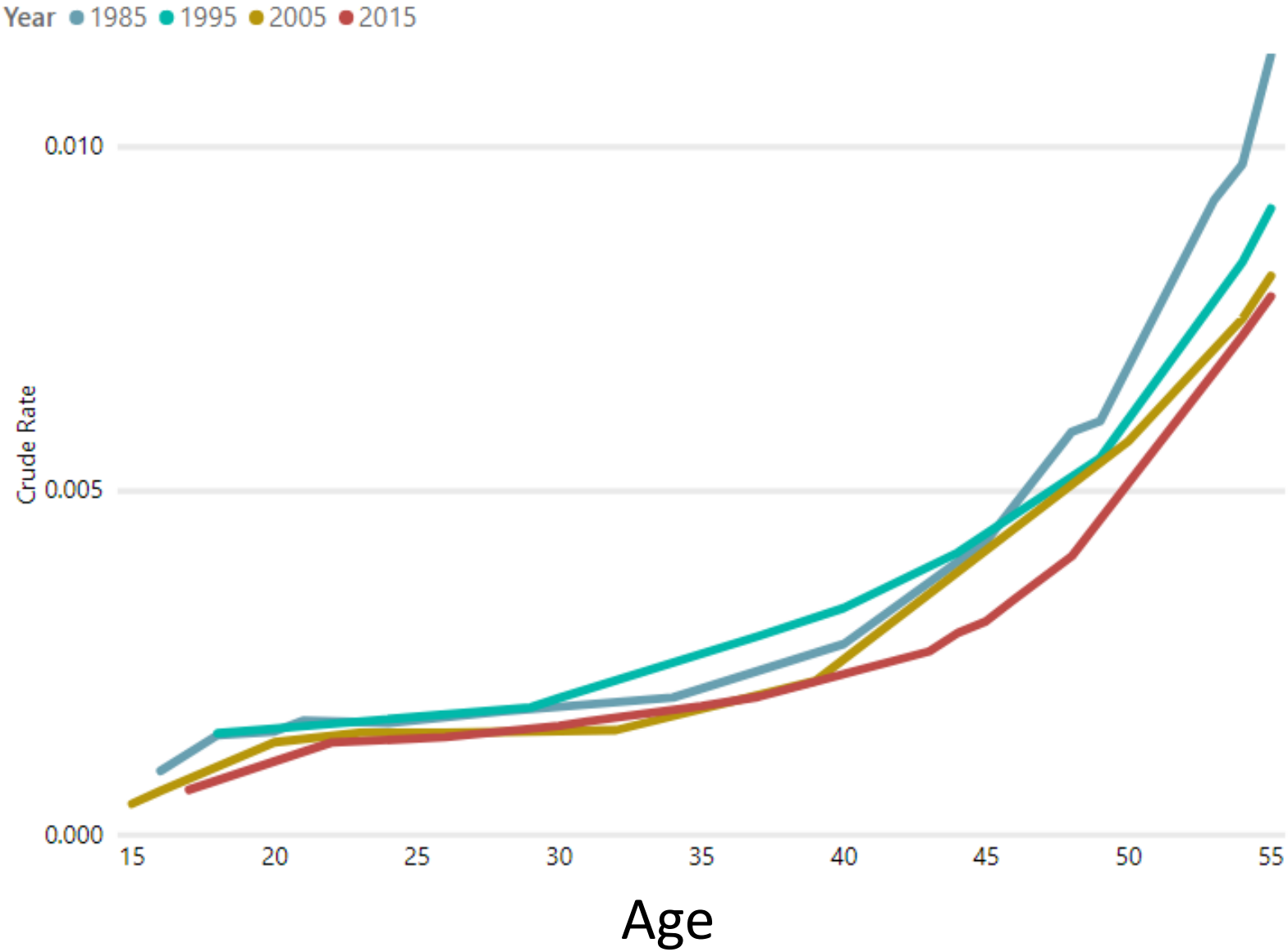

DATA
Understanding
Black-Box
ML Models
Adriaan Rowan
INSIGHT


DATA
Understanding
Black-Box
ML Models
Adriaan Rowan
INSIGHT

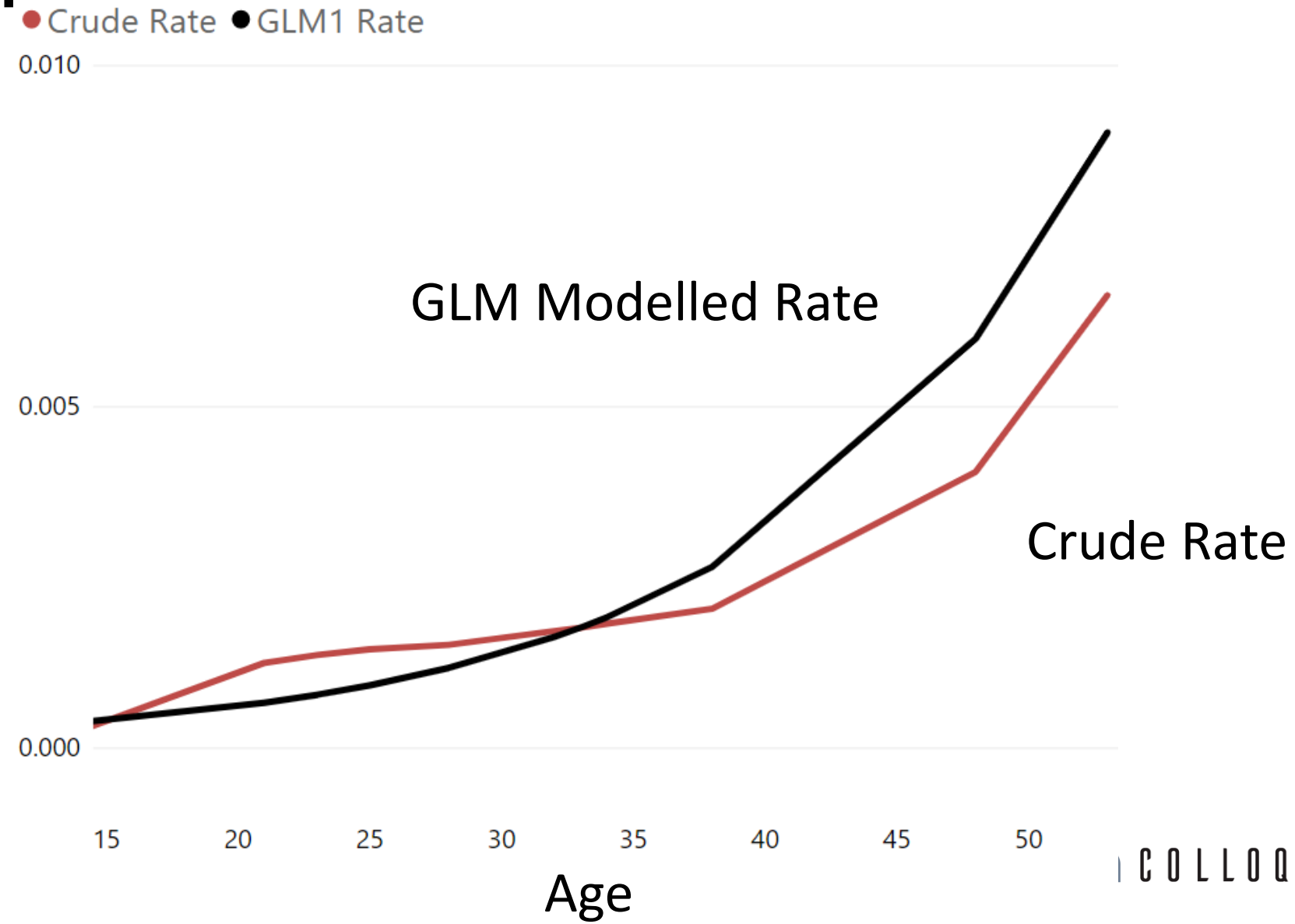
# Graph of mortality by Year and Gender



# Graph of mortality by Age and Year



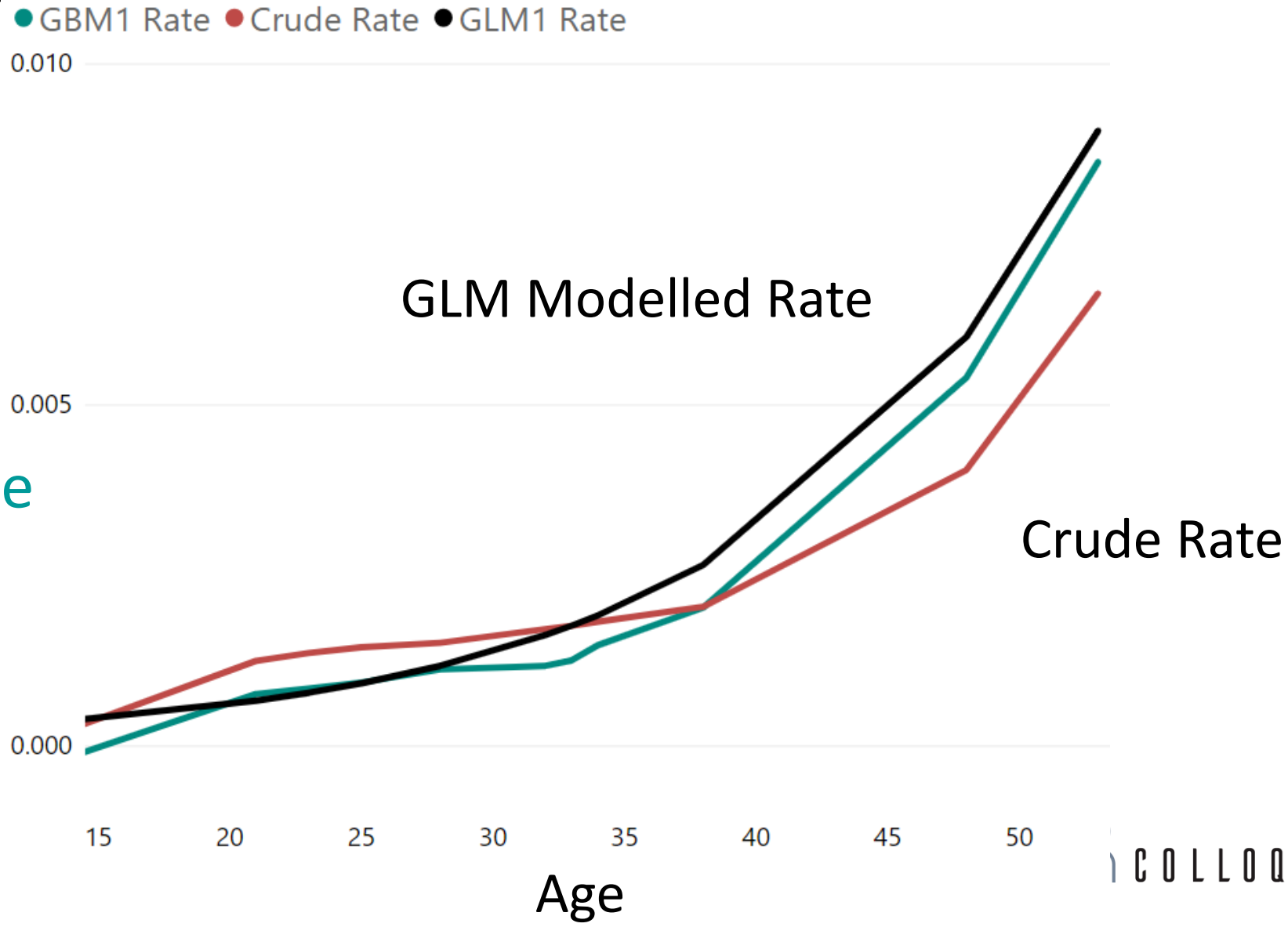
# Graph of GLM model vs Crude Rate



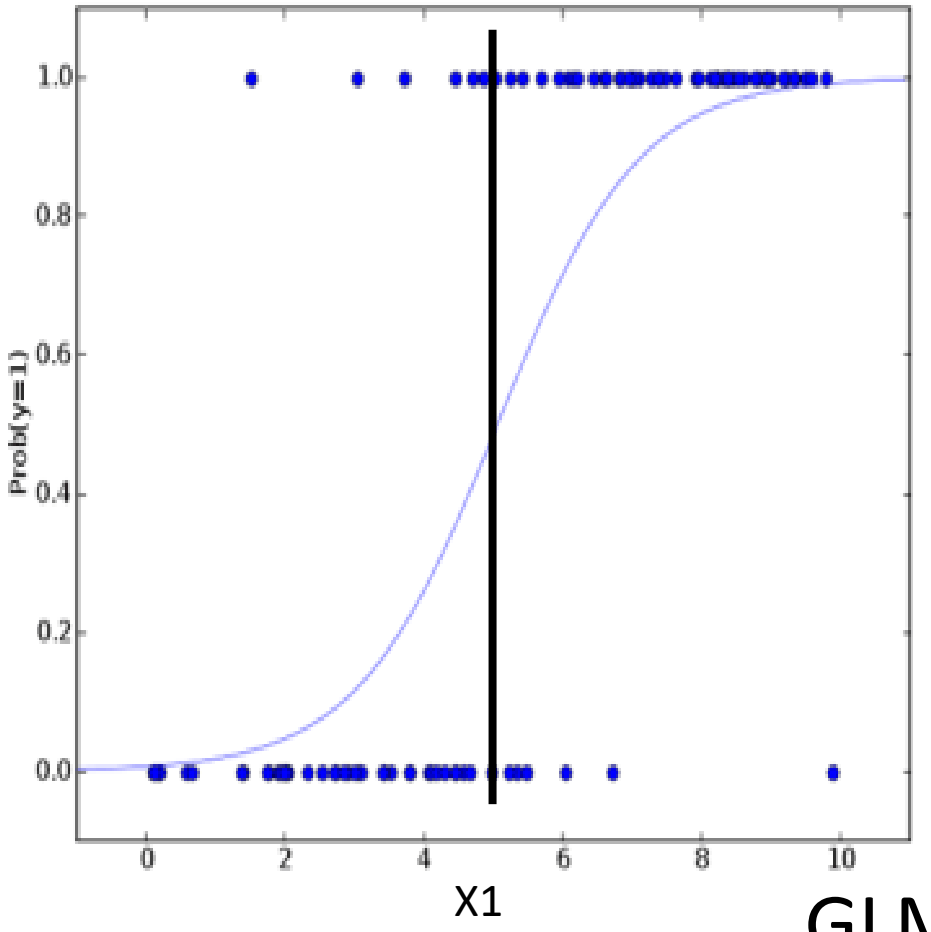
# Graph of GBM model vs Crude Rate

+

Gradient Boosting Modelled Rate



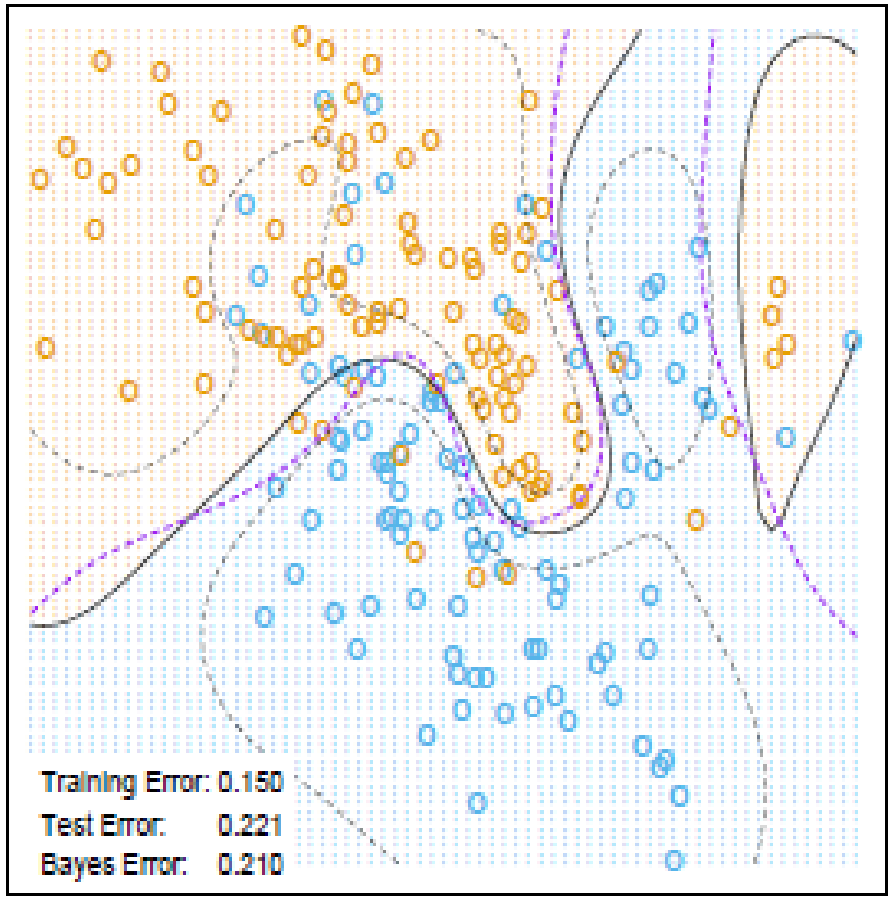
# Visualising Log Linear Regression



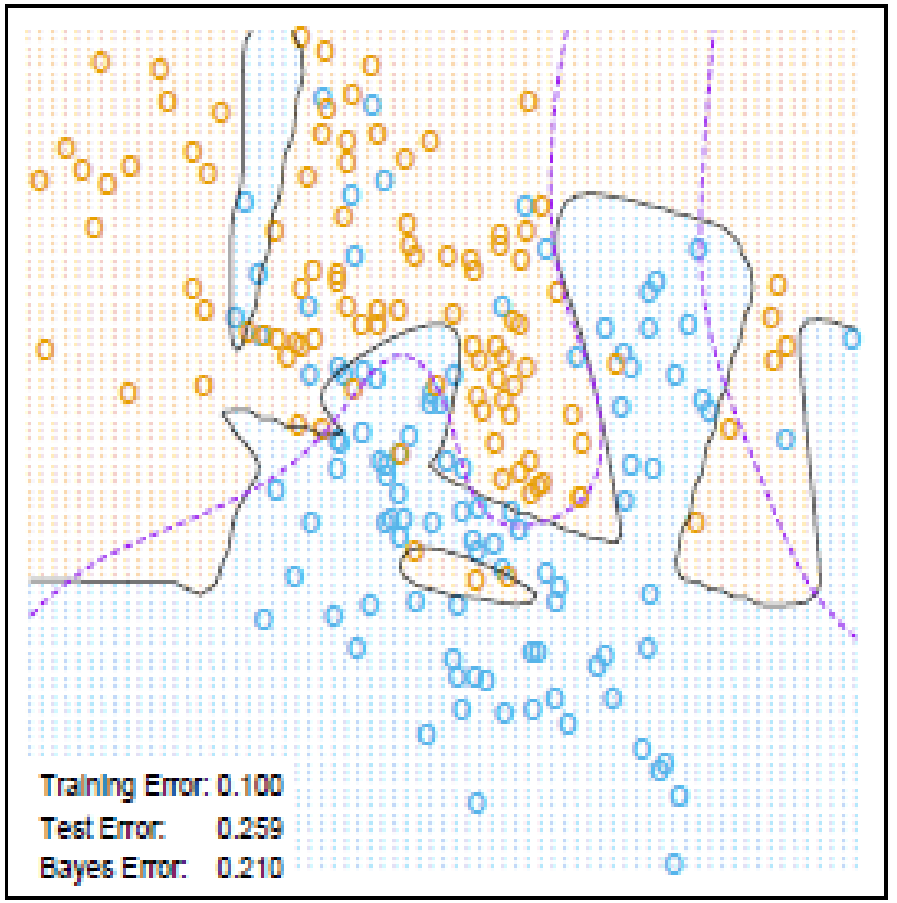


GLM Model

# Comparing Machine Learning Models

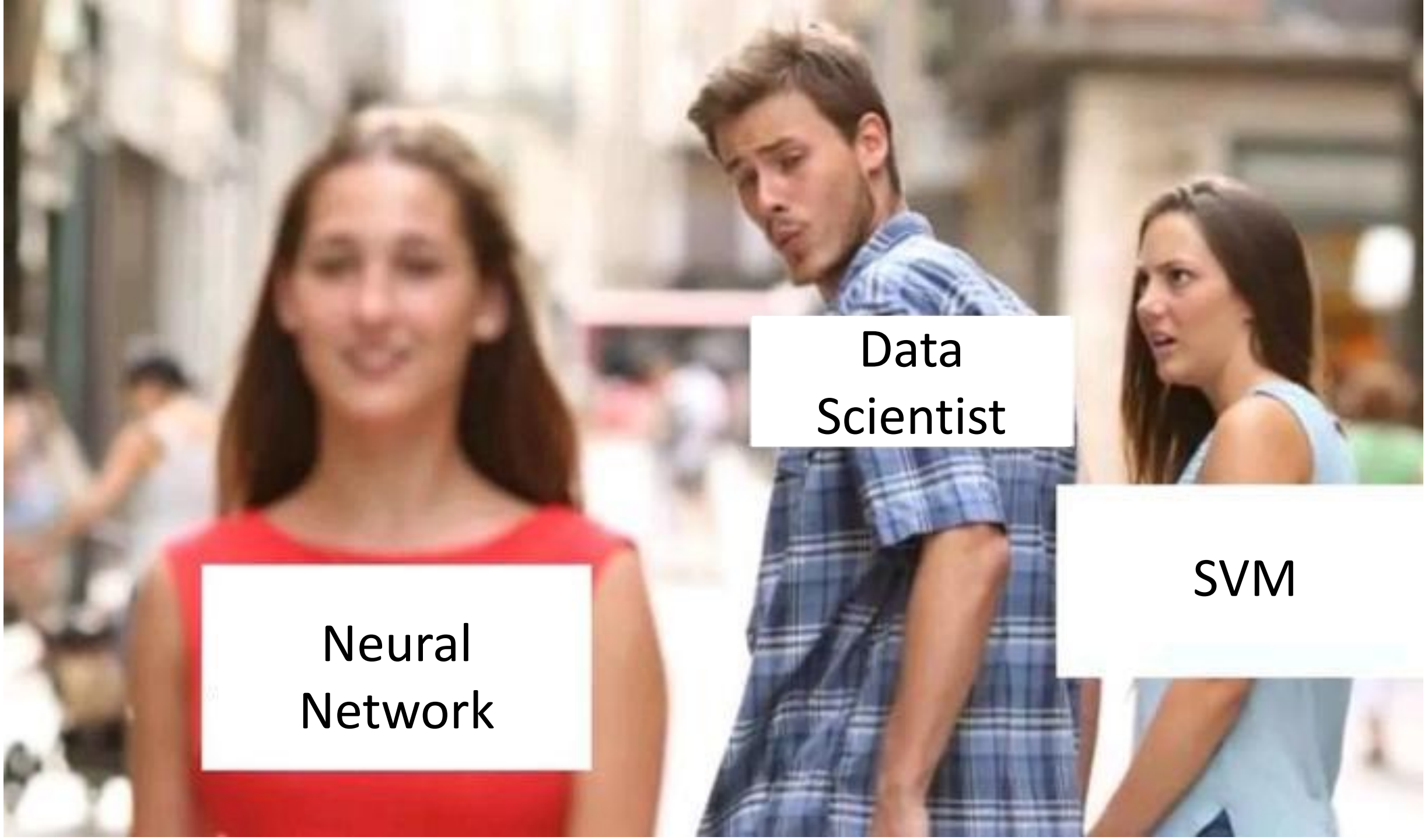



SVM Model



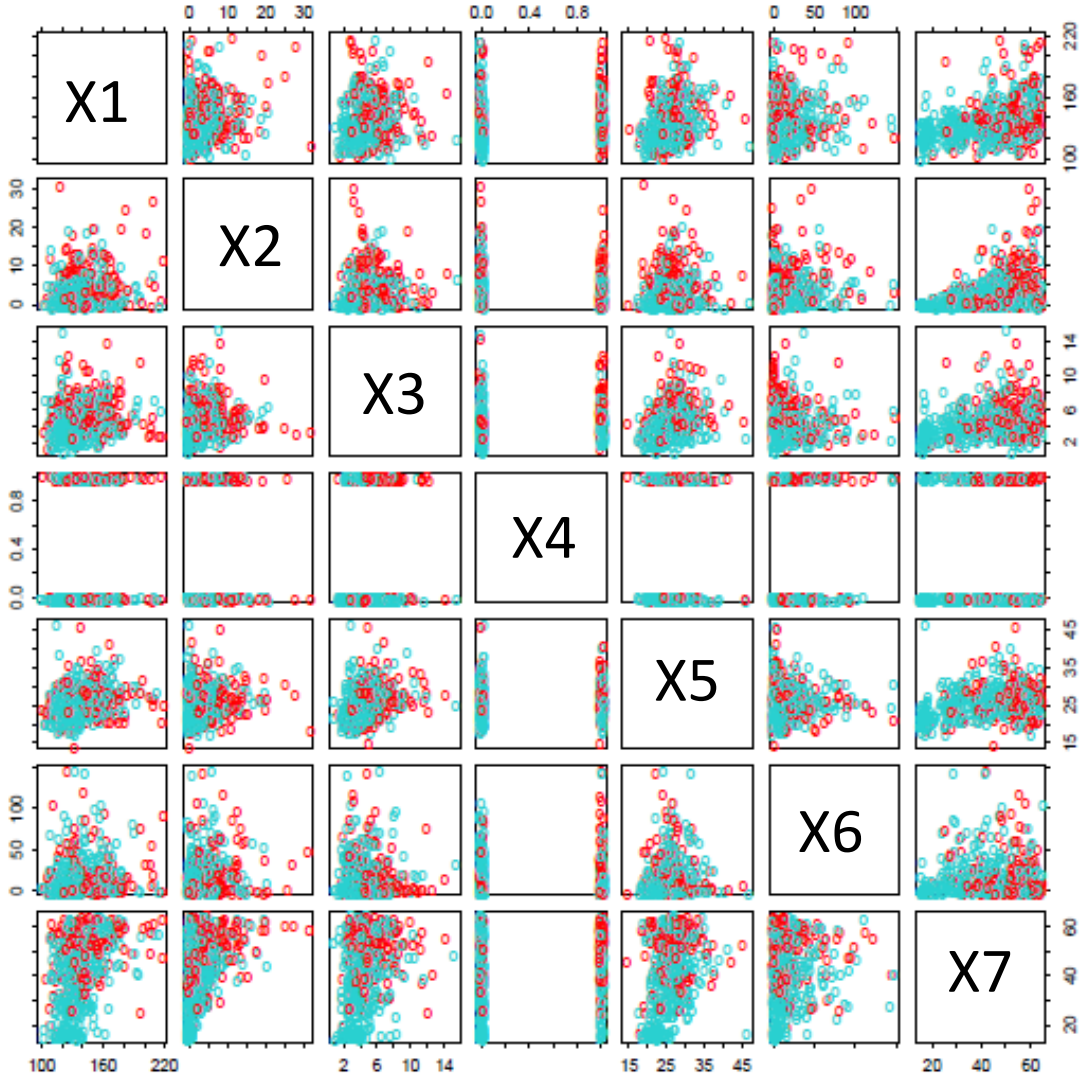

Neural Network Model

Hastie, Tibshirani, Friedman


Data
Scientist
Neural
Network
SVM


STILL WAITING
FOR MY NEURAL NETWORK TO TRAIN
memegenerator.net

# Trellis of 2x2 results




Hastie,
Tibshirani,
Friedman

# A solution is the use of a Biplot

## Optimal 2-dim representation of p-variable data

- Bi refers to showing both Data Points & Variables

## Data points via Multi-Dimensional Scaling

- Distance between points = measure of similarity

## Variables via Regression

- Direction of increase = Location of name
- Correlation = direction of increase
- Scale of variables = Label on the axes

# Multi-Dimensional Scaling

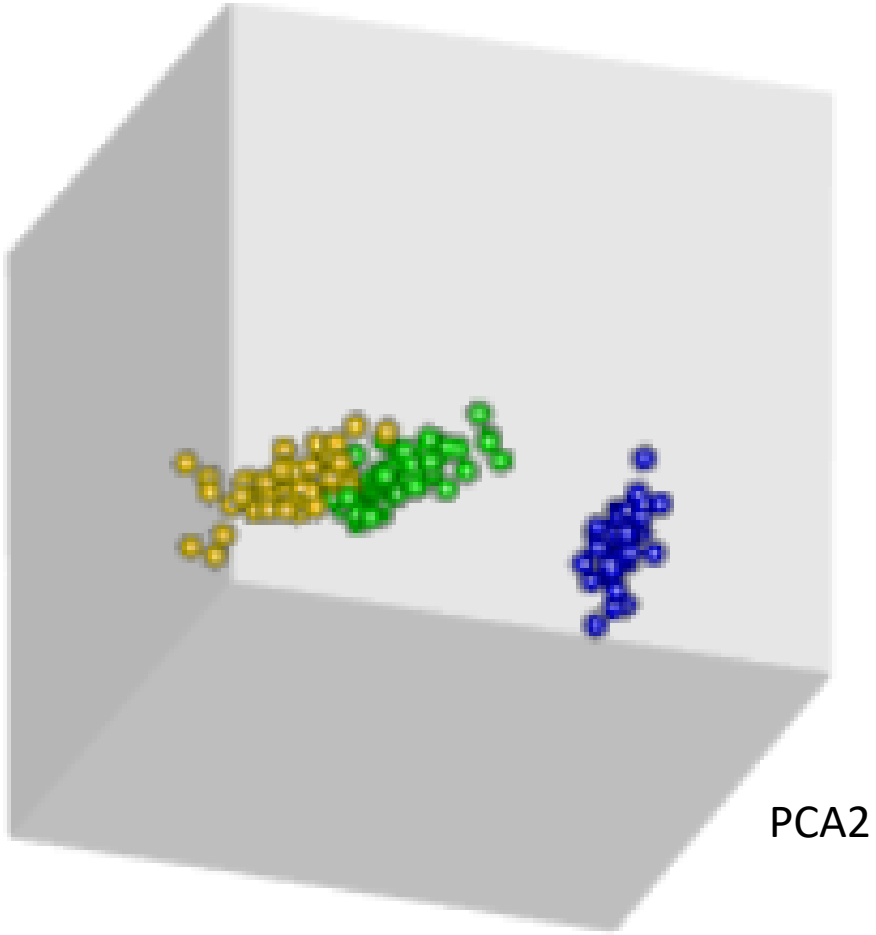
Setosa

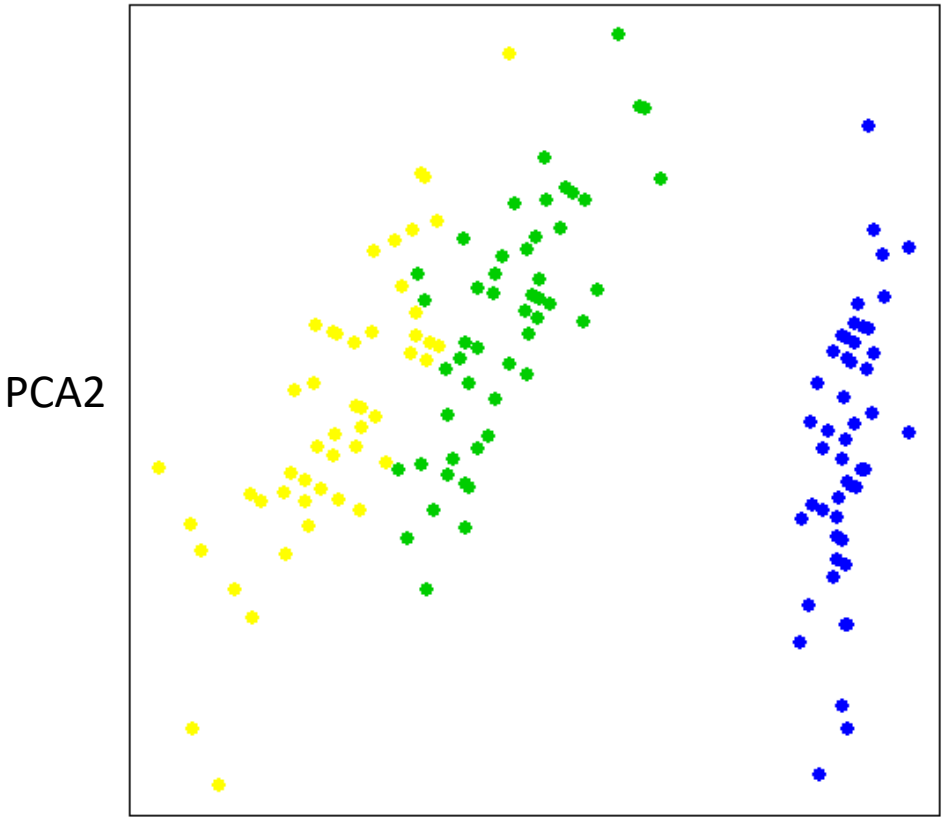
Versicolor

Virginica

| Species | Sepal Length | Sepal Width | Petal Length | Petal Width |
|---|---|---|---|---|
| Setosa | 5.1 | 3.5 | 1.4 | 0.2 |
| Setosa | 4.9 | 3 | 1.4 | 0.2 |
| setosa | 5 | 3.3 | 1.4 | 0.2 |
| | | | | |
| Versicolor | 6.4 | 3.2 | 4.5 | 1.5 |
| Versicolor | 5.7 | 2.8 | 4.1 | 1.3 |
| | | | | |
| Virginica | 7.2 | 3.6 | 6.1 | 2.5 |
| Virginica | 6.8 | 3 | 5.5 | 2.1 |

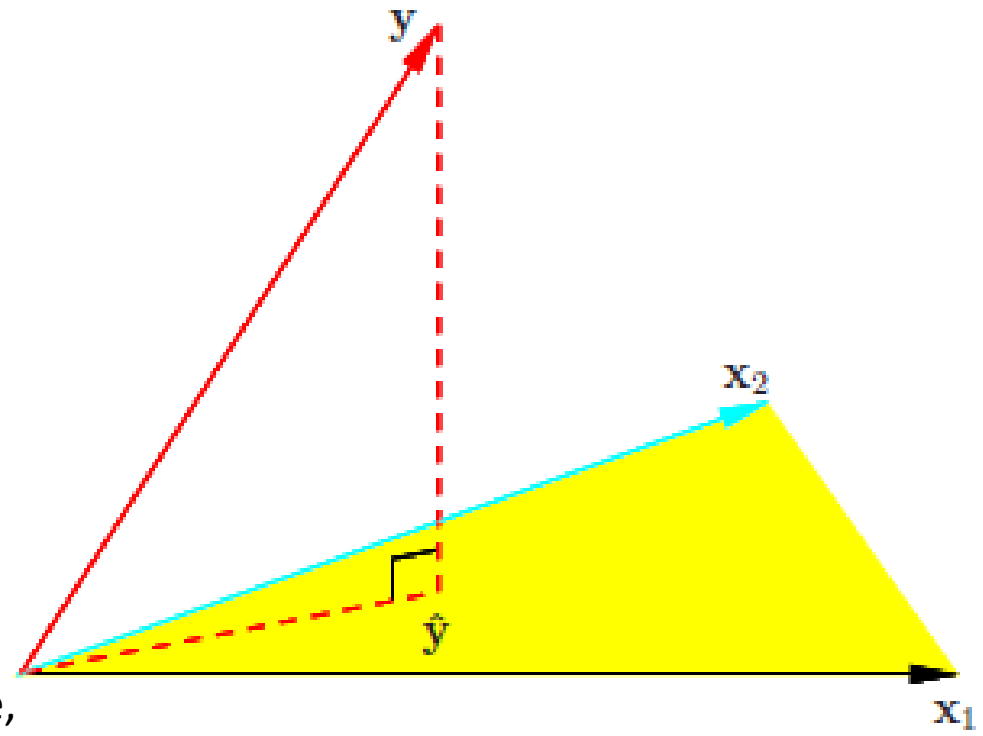
# Multi-Dimensional Scaling



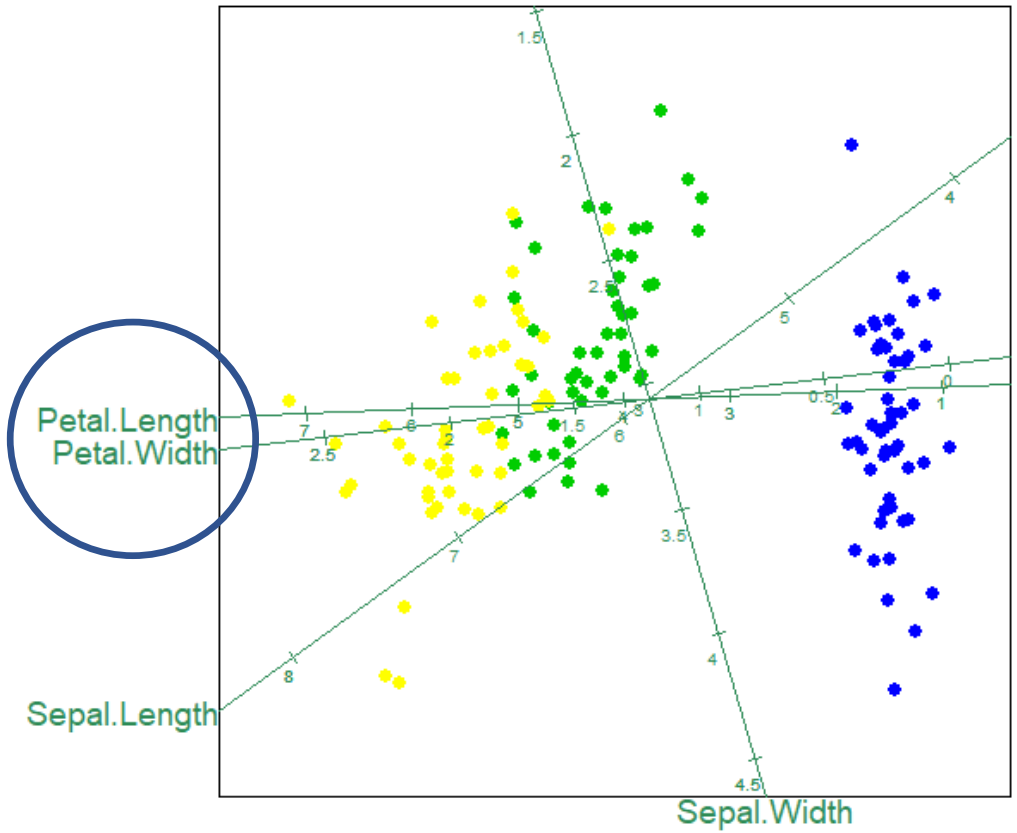
# Regression to calculate Axes

Y = XB: B (p x **1**)



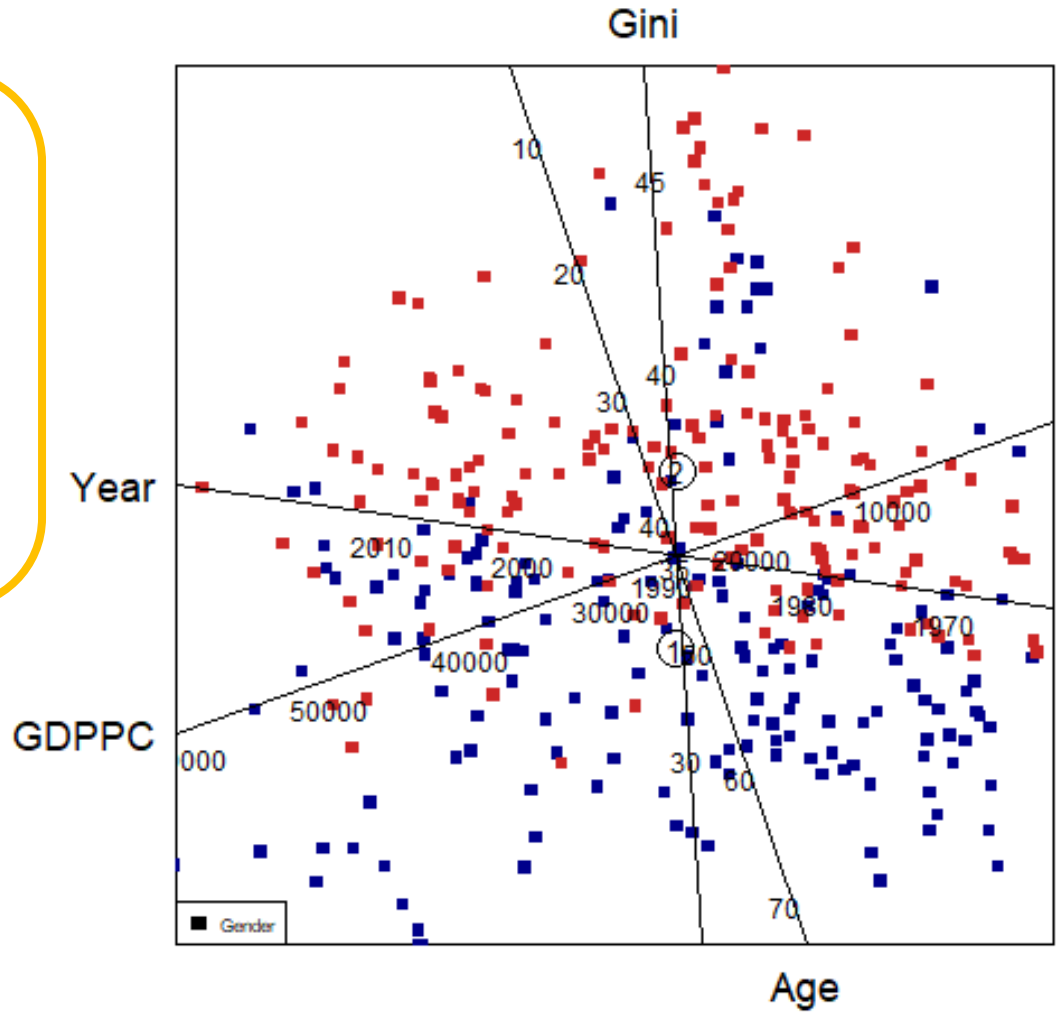

Hastie,
Tibshirani,
Friedman

X = ZR: R (p x **2**)



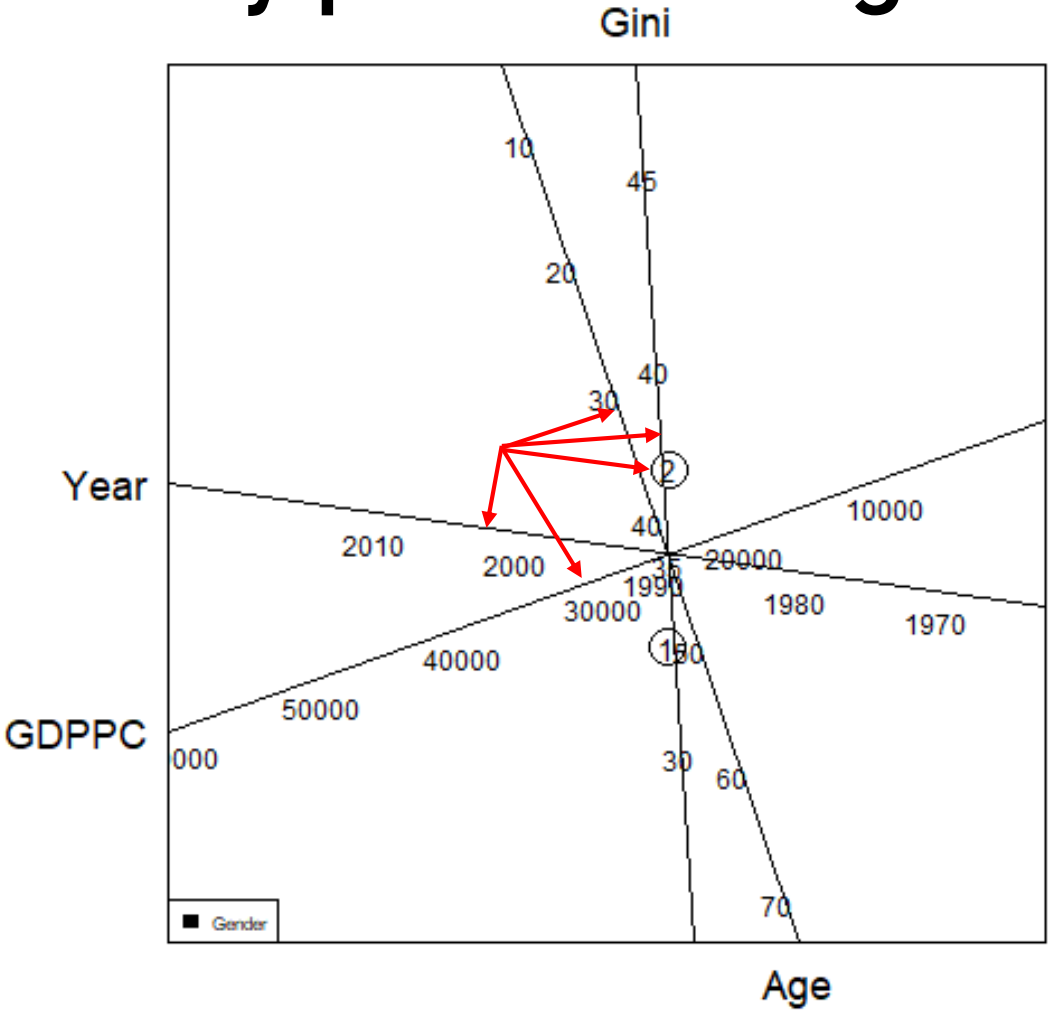
# Creates a 5-variable Biplot

5 Variables:
- Age
- Gini Coefficient
- Year
- GDP Per Capita
- Gender




Gender indicated by (2) Males And (1) Females
Middle point of Data points

# Convert any point to original 5 variables



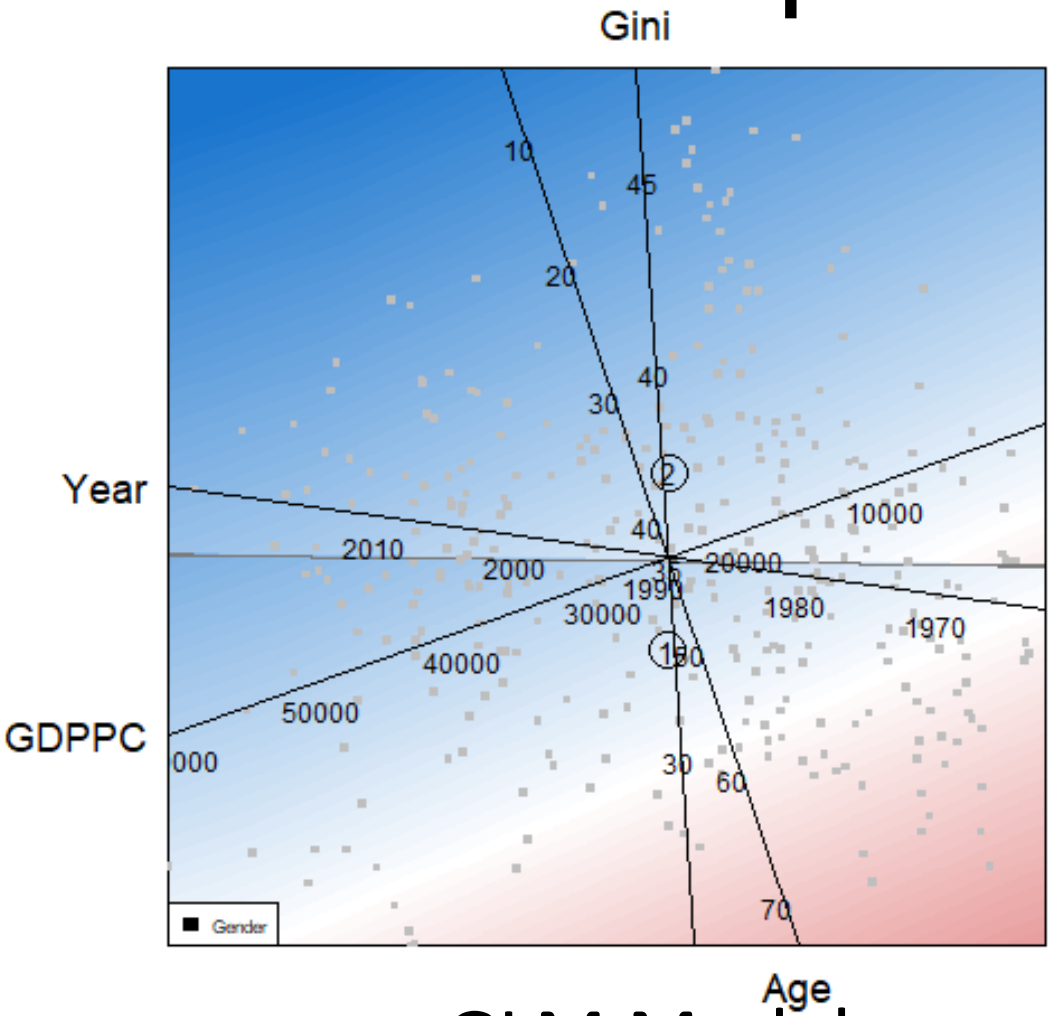

Z = (-1.5; 0.7)

X: Age = 30;
Gini = 38;
Gender = (2) Male;
GDPPC = 30000;
Year = 2003

[redacted]

m = 0.002

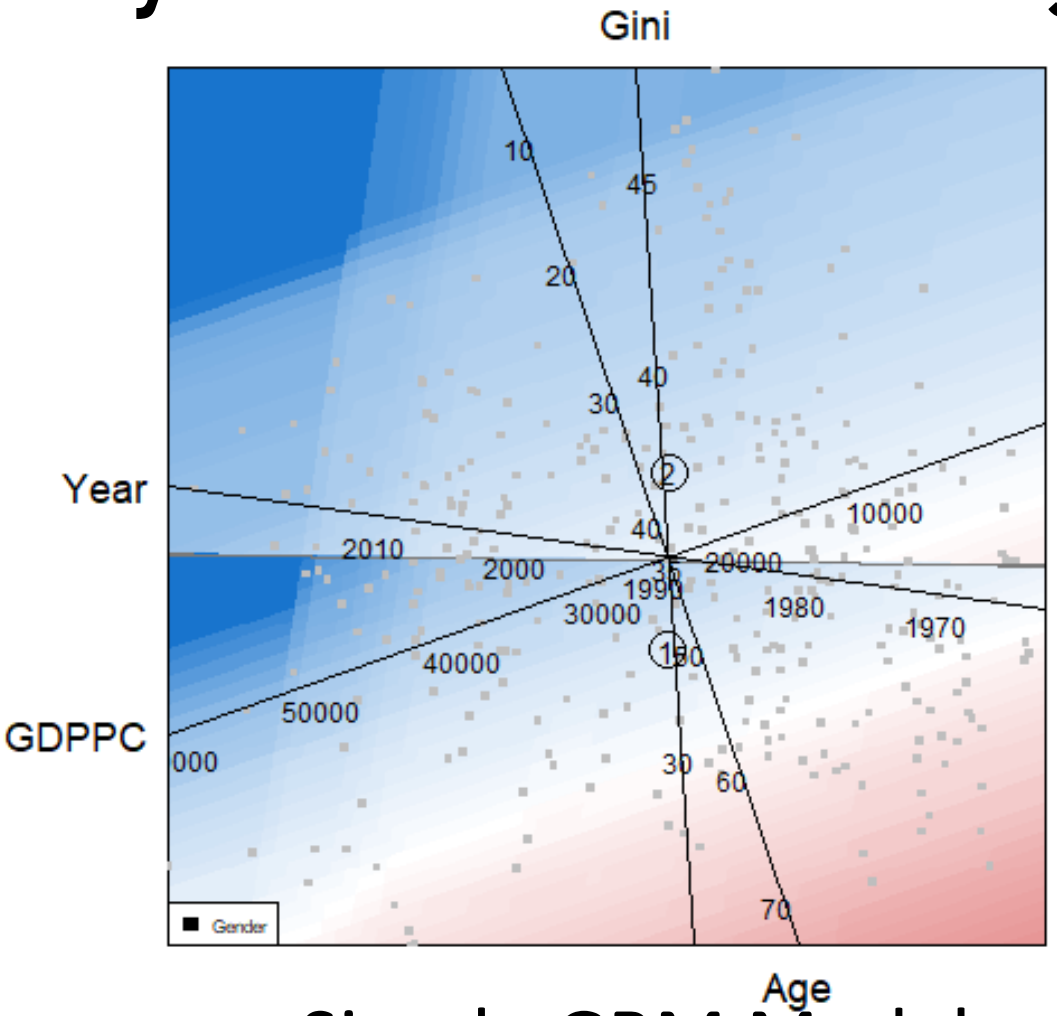
# … and use values as input in GLM model




GLM Model

Overall Error rate:
1.3%

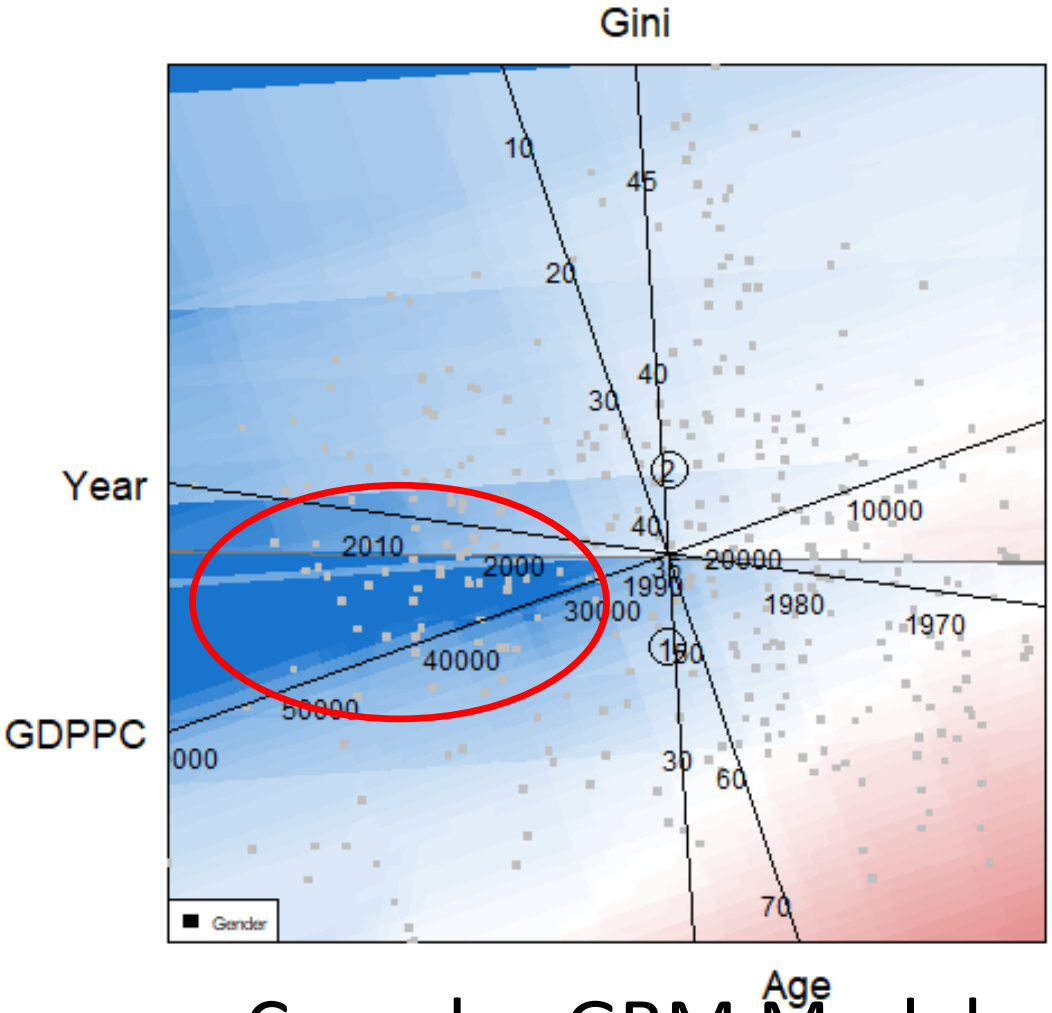
# …or in any Machine Learning model




Overall Error rate:
1.1%

Simple GBM Model

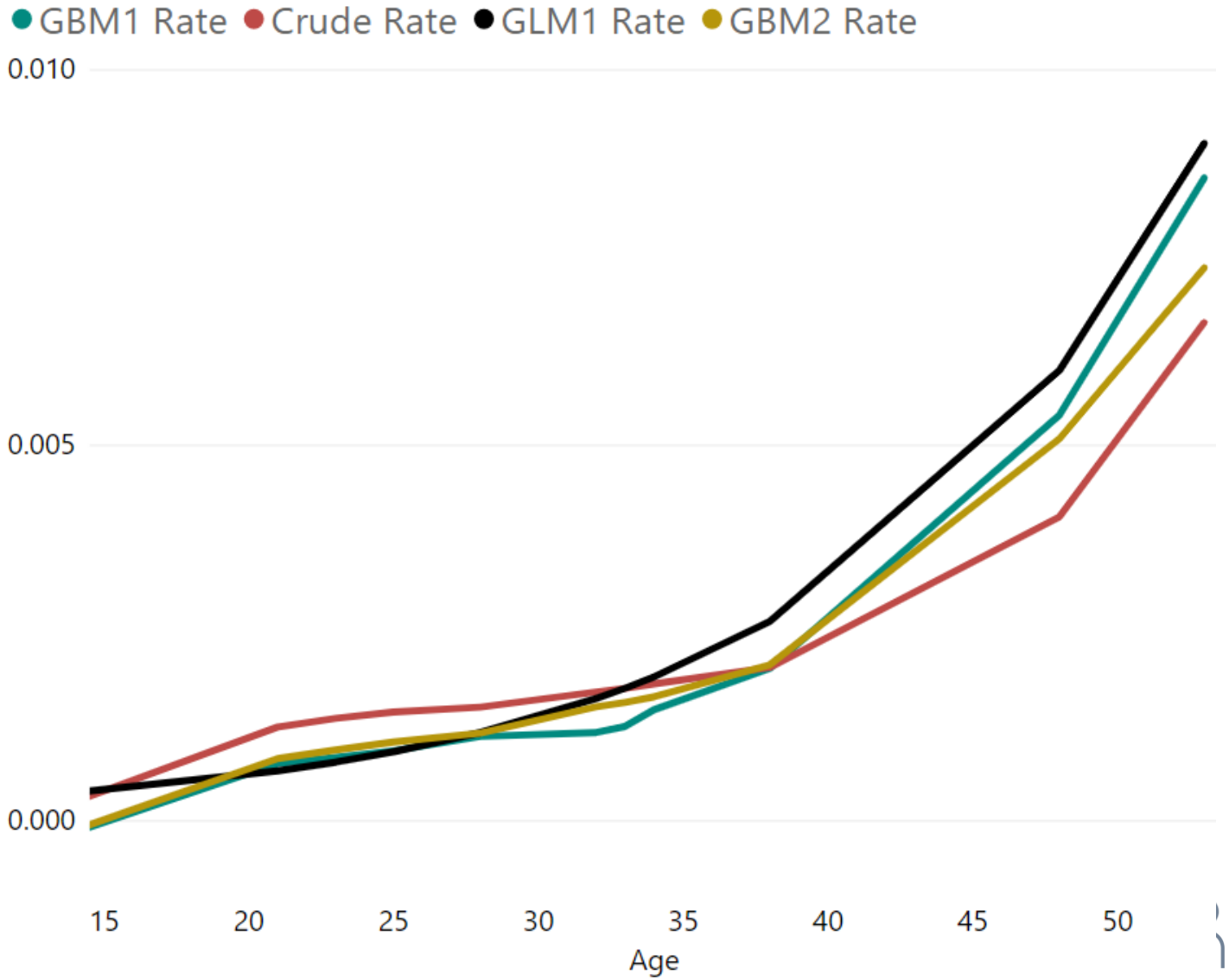
# Well calibrated GBM model predictions




Overall Error rate:
0.7%

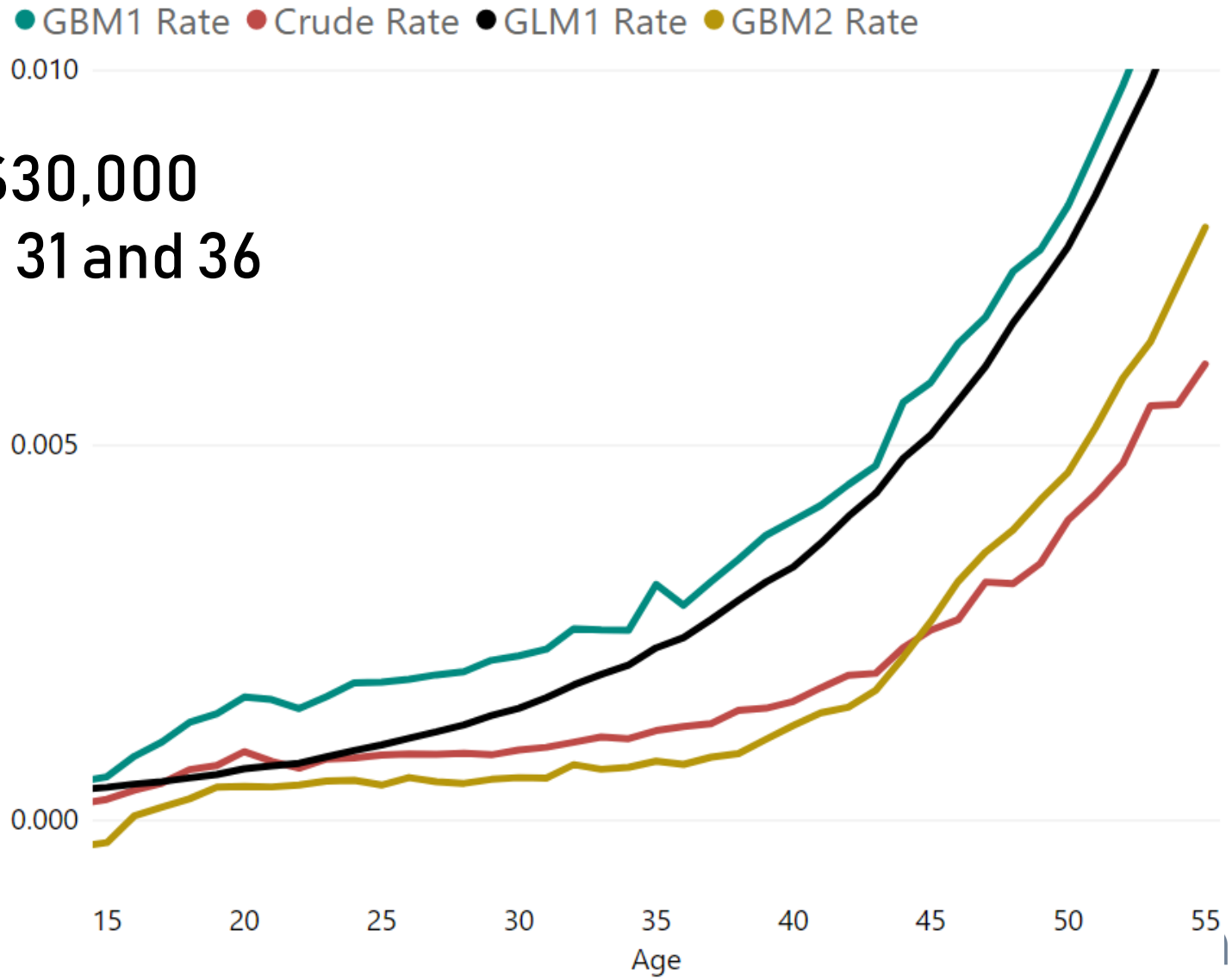
Complex GBM Model

# …Better follows overall mortality



# …And groups with unique mortality

GDP-PC > $30,000
GINI between 31 and 36




No such thing as a free black lunch box

# Understanding Black Box models

- No such thing as a free lunch – Both for ML models and Biplots
- Need to change how model output is viewed and assessed
- Judgement and domain knowledge gives Actuaries an edge
- BUT Need for simplistic and smooth outcome should not be at expense of accuracy and false certainty

If Alpha-Infinity could identify, understand [and predict the outcome] of any problem, it would seem to us like an oracle. We would sit at its feet and listen intently.

We would not understand why the oracle was always right, but we would check its calculations…and confirm its revelations.

Science, that signal human endeavor, would reduce our role to that of spectators, gaping in wonder and confusion.

https://www.nytimes.com/2018/12/26/science/chess-artificial-intelligence.html


QUESTION
Adriaan@
i3actuaries
.co.za
ANSWER