

# Visualising ML Models

---

Amit Kapoor  
@amitkaps

# Story

---

“We don’t see things as they are, we see them as we are.”

— *Anais Nin*

# The Blind Men & the Elephant

---

“And so these men of Indostan  
Disputed loud and long,  
Each in his own opinion  
Exceeding stiff and strong,  
Though each was partly in the right,  
And all were in the wrong.”

— *John Godfrey Saxe*

# The Elephant: Data

---

**“Data is just a clue to the end  
truth”**

— *Josh Smith*

# The Men: Building Models

---

"All models are wrong, but some  
are useful"

— *George Box*

# **Layers of Abstraction**

---

**Data Abstraction**

**Visual Abstraction**

**Model Abstraction**

# Machine Learning (ML) Speak

---

Data Transformation

Visual Exploration

Model Building

# ML Pipeline

---

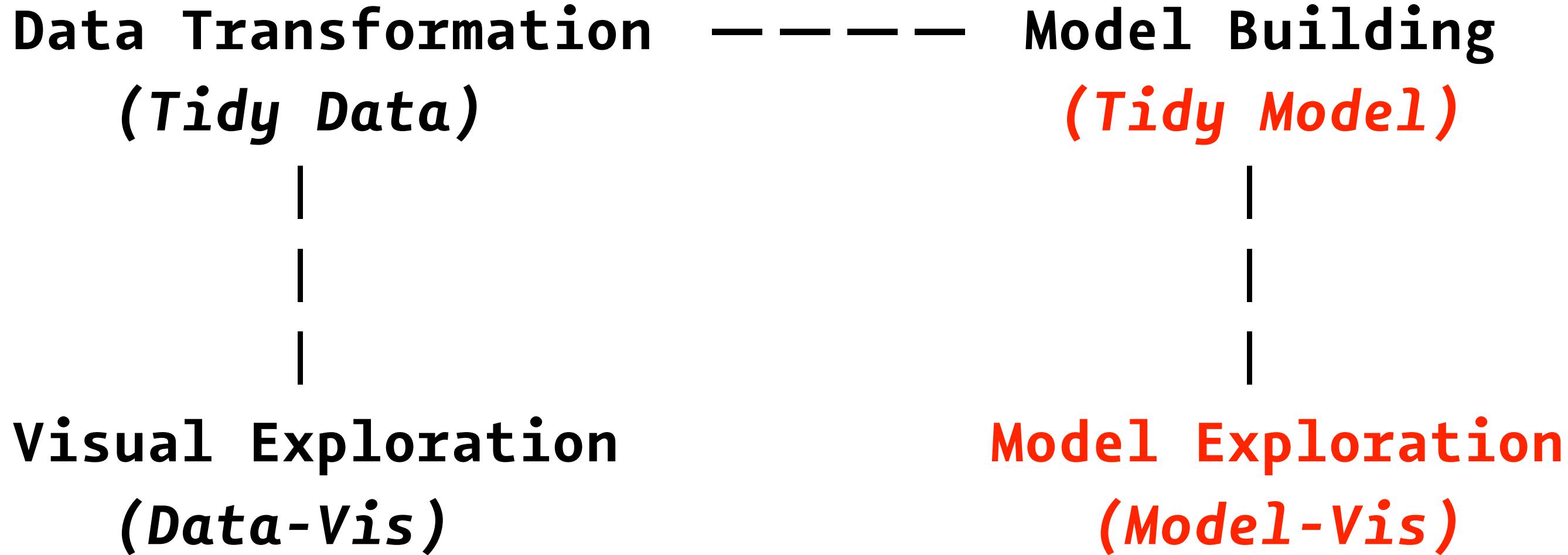
Data Transformation ————— Model Building  
*(Tidy Data)*



Visual Exploration  
*(Data-Vis)*

# ML Pipeline++

---



# Model-Viz Approach

---

- [0] Visualise the **data space**
- [1] Visualise the **predictions in the data space**
- [2] Visualise the **errors in model fitting**
- [3] Visualise with **different model parameters**
- [4] Visualise with **different input datasets**
- [5] Visualise the **entire model space**
- [6] Visualise the **entire feature space**
- [7] Visualise the **many models together**

# Model-Viz Examples

---

Regression ( $n < 50$ ,  $p = 4$ )

Classification: 2 class ( $n \sim 5K$ ,  
 $p = 785$ )

# Regression: Small

---

Cars dataset - price vs kmp1

Scraped from comparison website

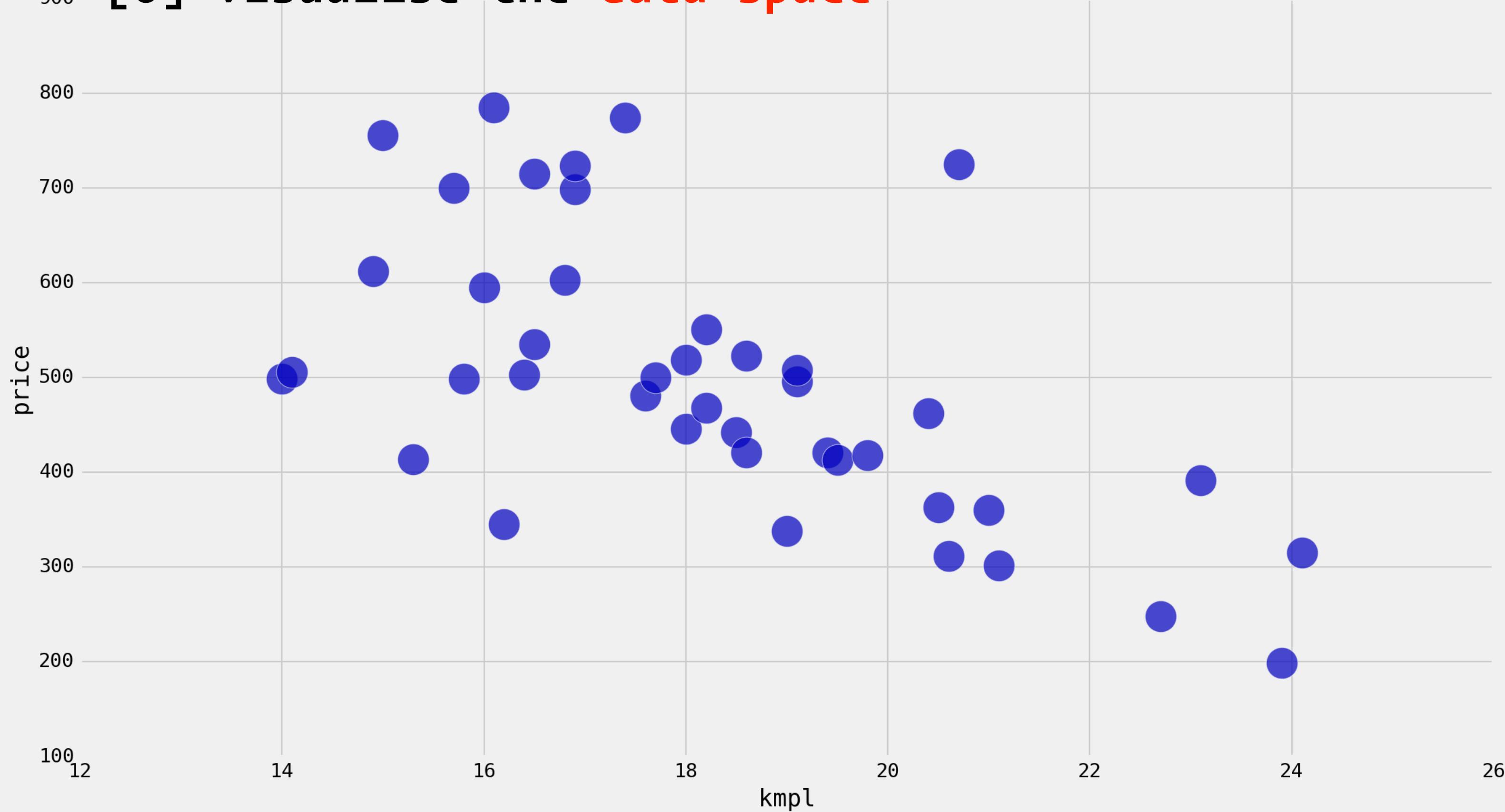
Refined & tidied up

Base version for petrol cars

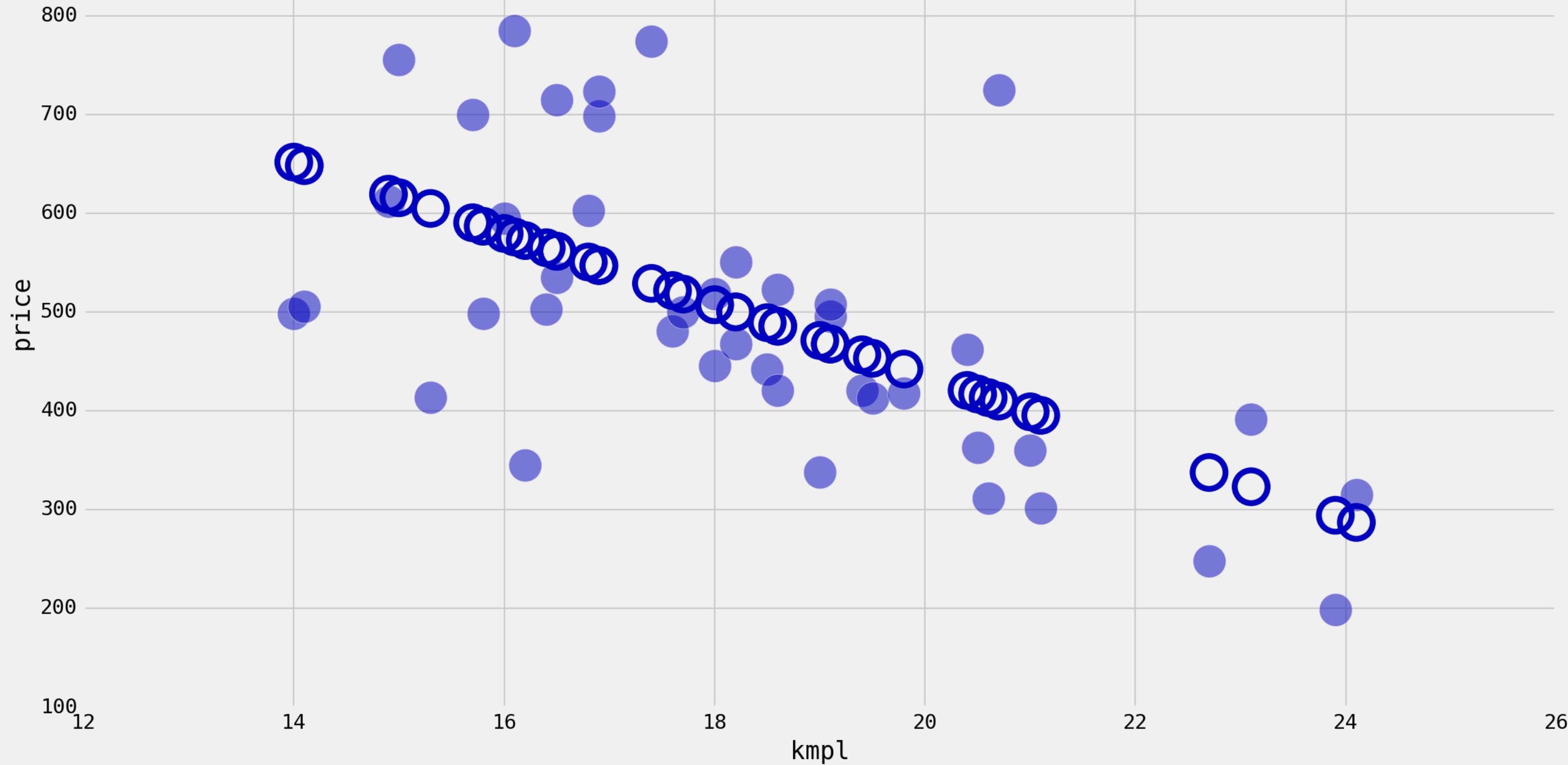
Price < ₹ 1,000K, n = 42

brand	model	price	kmpl	type	bhp
Tata	Nano	199	23.9	Hatchback	38
Suzuki	Alto800	248	22.7	Hatchback	47
Hyundai	EON	302	21.1	Hatchback	55
Nissan	Datsun	312	20.6	Hatchback	67
...	...	...	...	...	...
Suzuki	Ciaz	725	20.7	Sedan	91
Skoda	Rapid	756	15.0	Sedan	104
Hyundai	Verna	774	17.4	Sedan	106
VW	Vento	785	16.1	Sedan	104

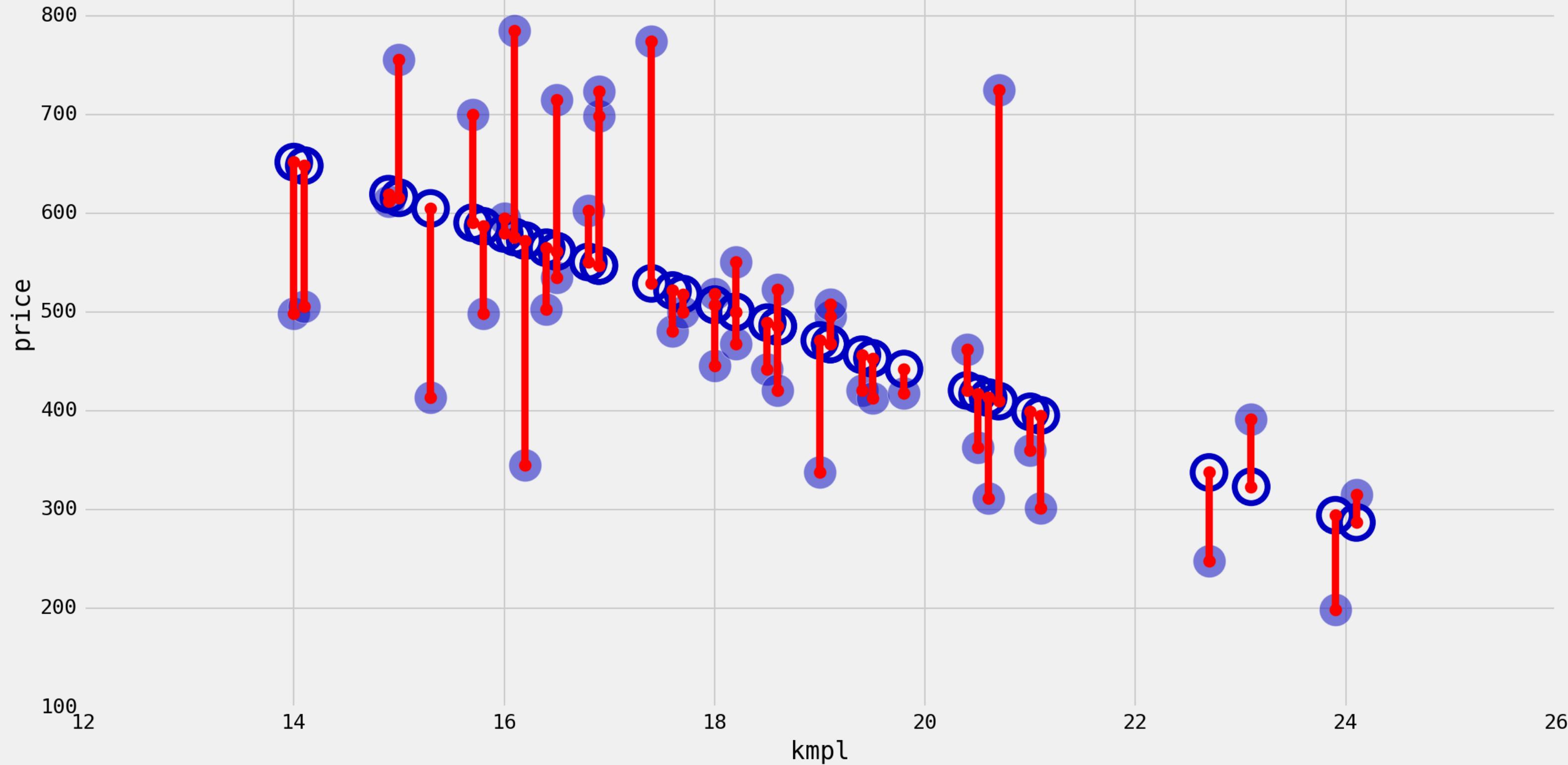
# [o] Visualise the data space



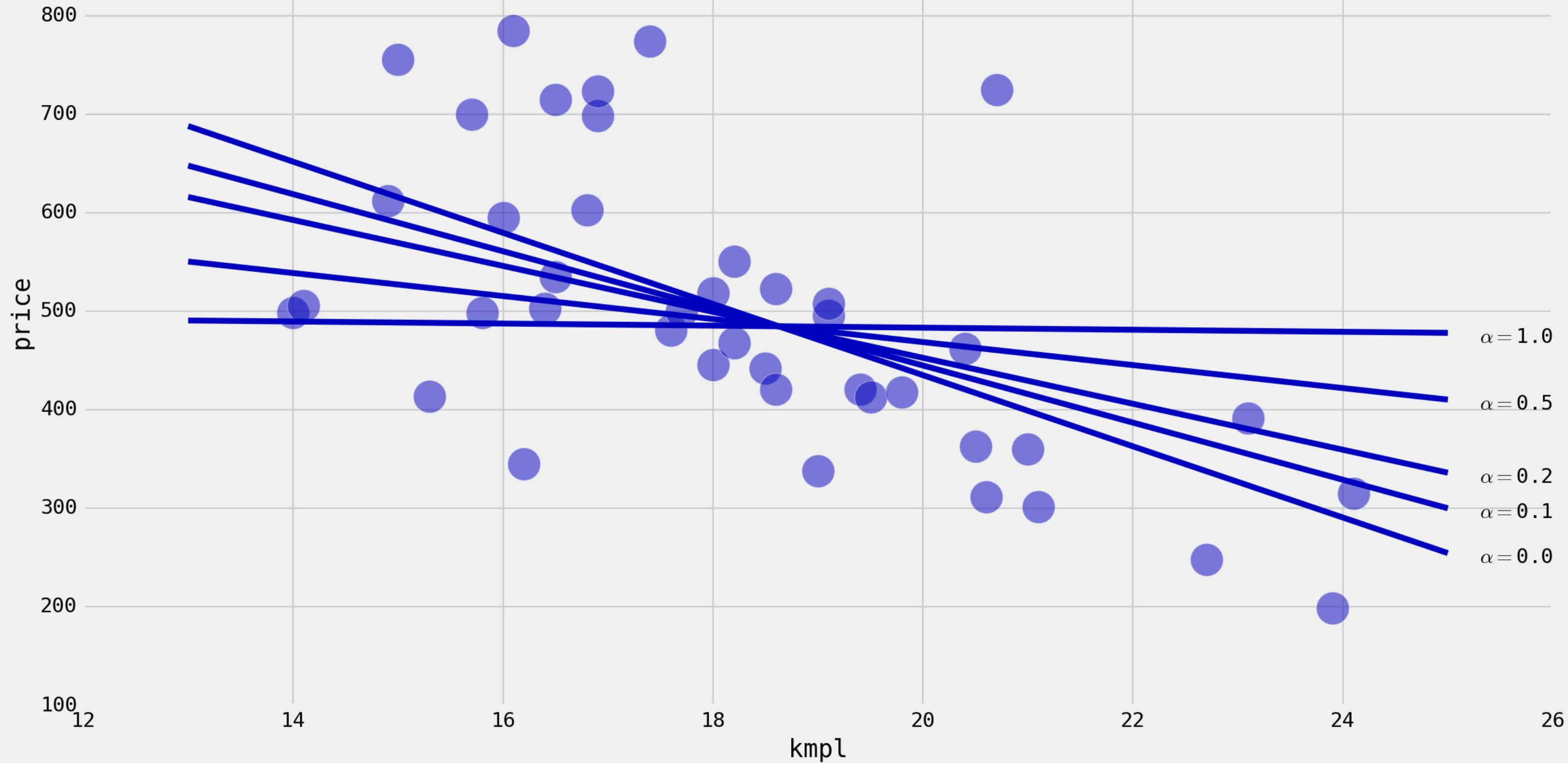
# [1] Visualise the predictions in the data space



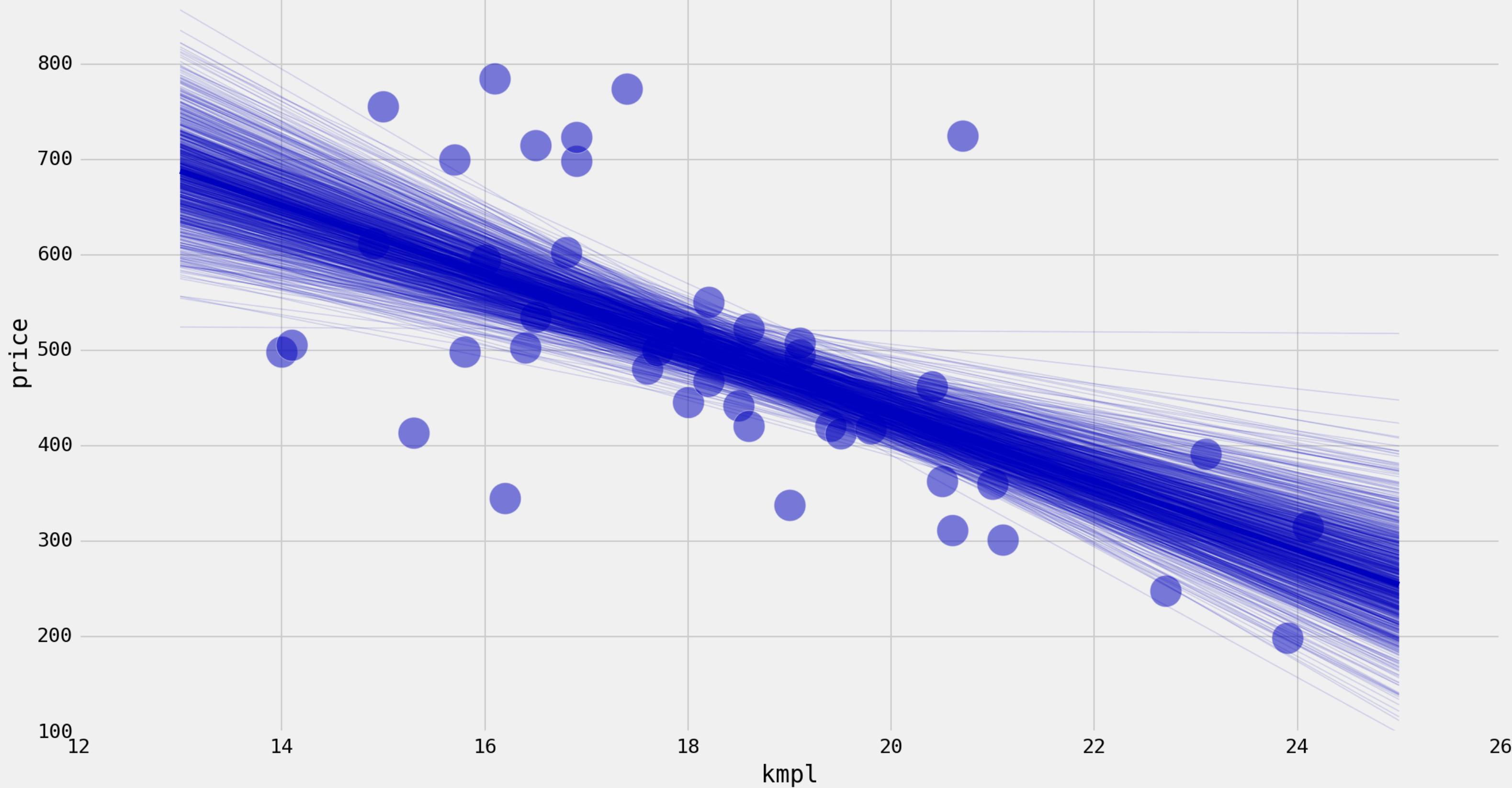
## [2] Visualise the errors in model fitting



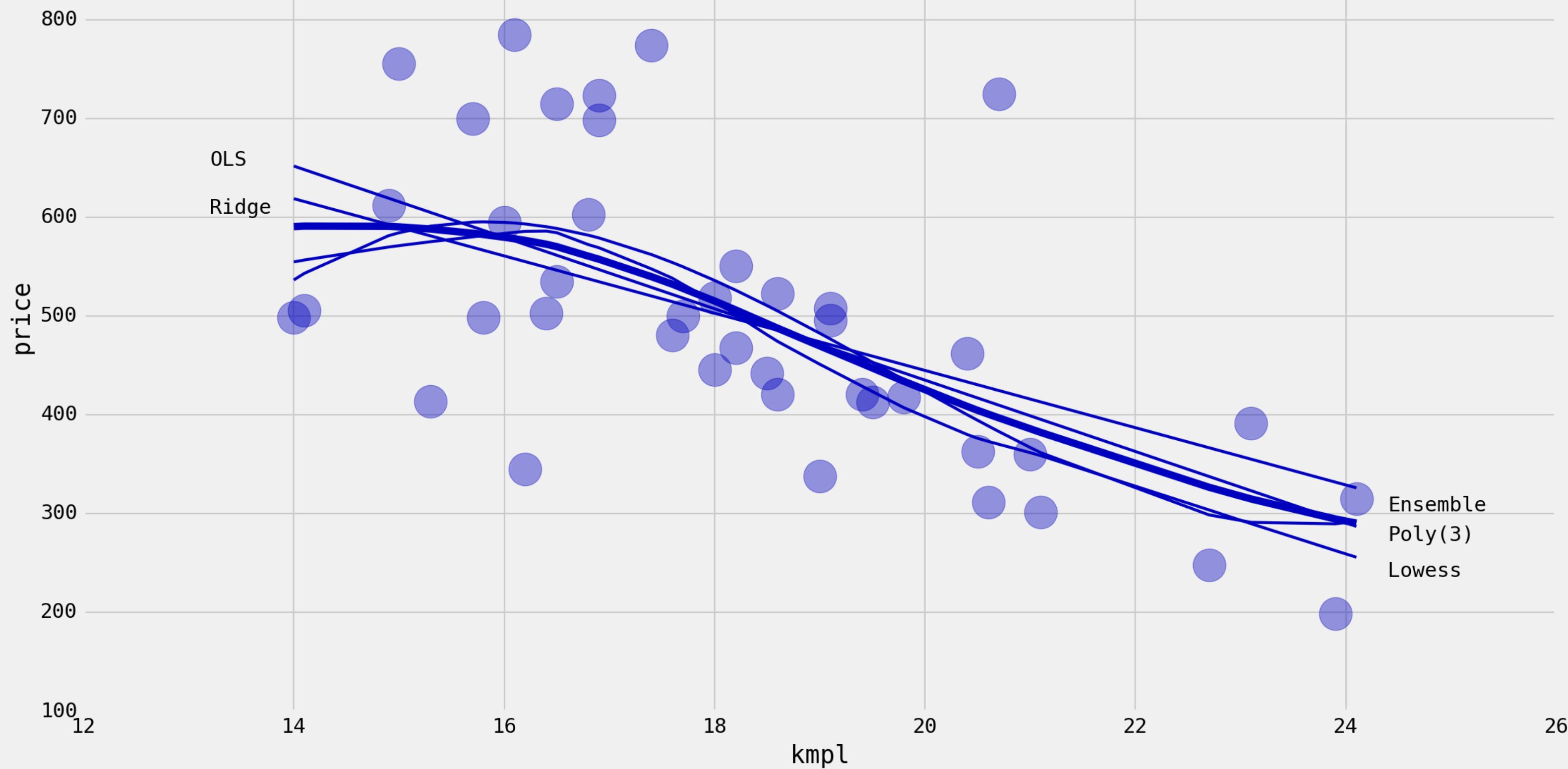
### [3] Visualise with different model parameters



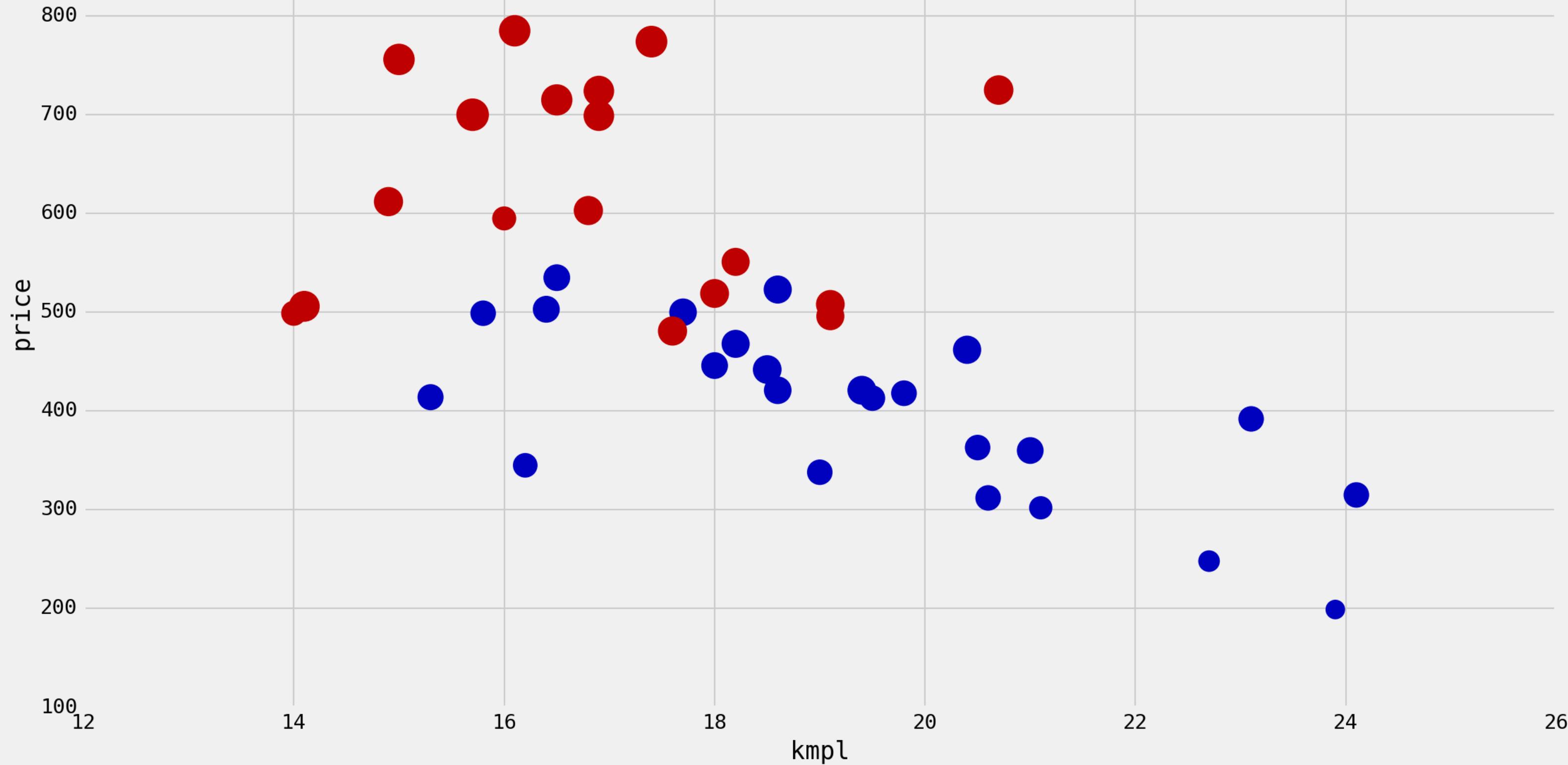
## [4] Visualise with different input datasets



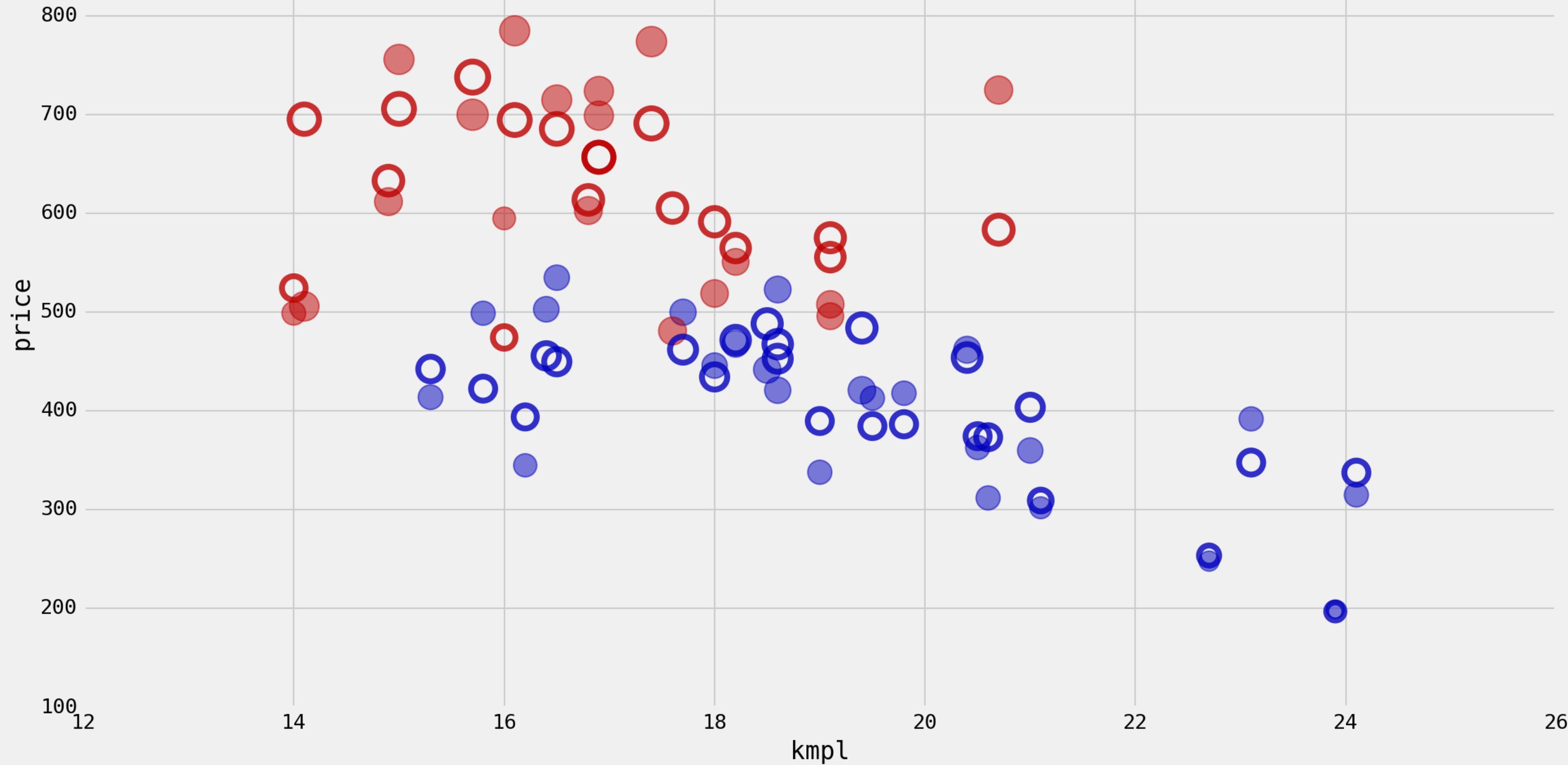
## [5] Visualise the entire model space



## [6] Visualise the entire feature space



## [7] Visualise the many models together



# Model-Viz Approach

---

- [0] Visualise the **data space**
- [1] Visualise the **predictions in the data space**
- [2] Visualise the **errors in model fitting**
- [3] Visualise with **different model parameters**
- [4] Visualise with **different input datasets**
- [5] Visualise the **entire model space**
- [6] Visualise the **entire feature space**
- [7] Visualise the **many models together**

# Model-Viz & ML Approach

---

- [0] **DATA VIS**: the data space
- [1] **PREDICTION**: the predictions in the data space
- [2] **VALIDATION**: the errors in model fitting
- [3] **TUNING**: with different model parameters
- [4] **BOOTSTRAP**: with different input datasets
- [5] **ENSEMBLE**: the entire model space
- [6] **FEATURES**: the entire feature space
- [7] **N-MODELS**: the many models together

# Model-Vis Methods

---

Limited **standard** methods

articulated

Adapt to **data** and **domain** type

Scope for **innovation** and

**development**

# Model-Vis Key Concept

---

Use visualisation to aid the transition of **implicit knowledge** in the data and your head to **explicit knowledge** in the model.

# Classification: 2 Class

---

MNIST - digit recognition

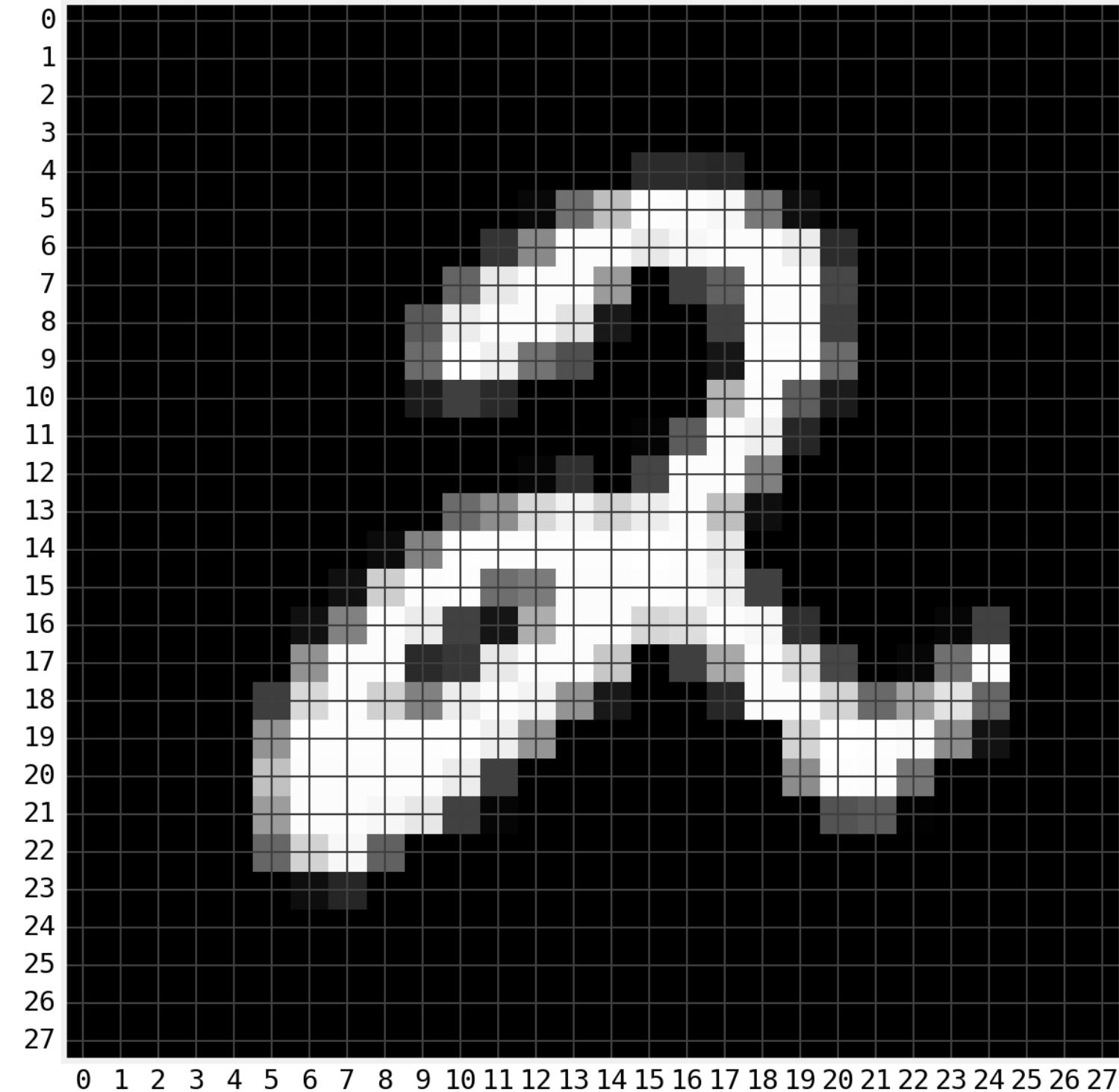
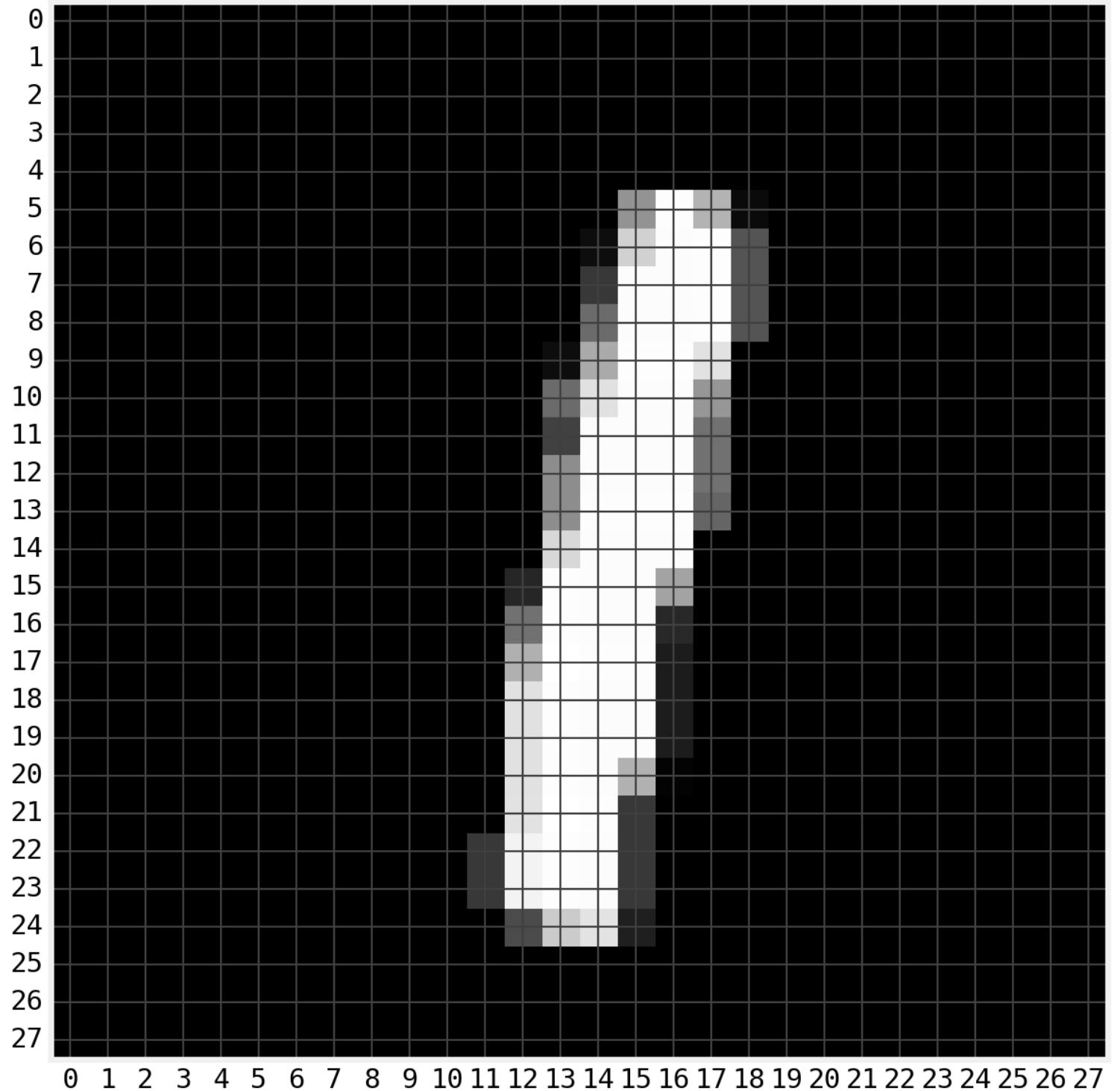
Reduced to 2-class - 1 and 2

784 dimensions - 28 x 28 gray

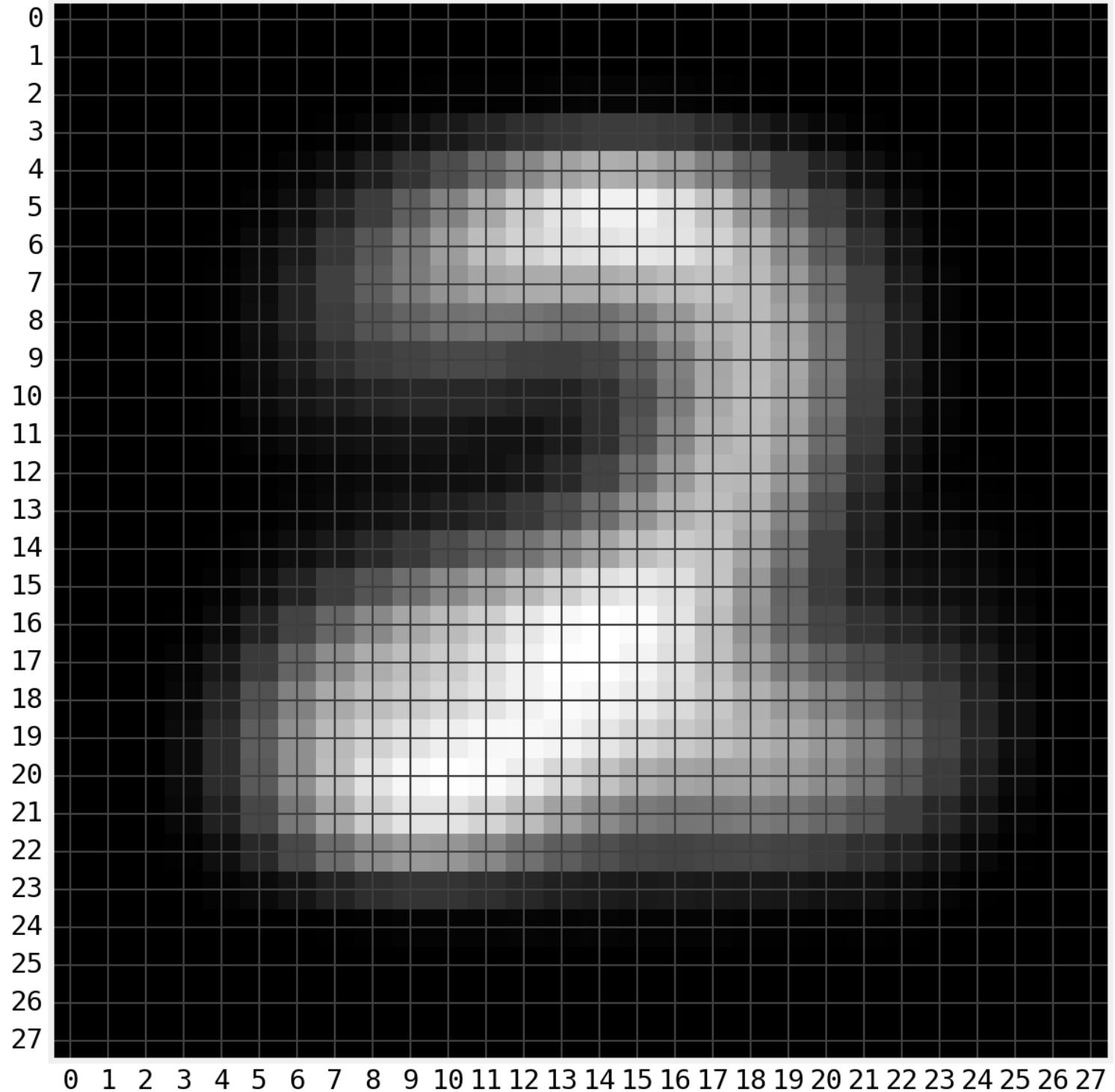
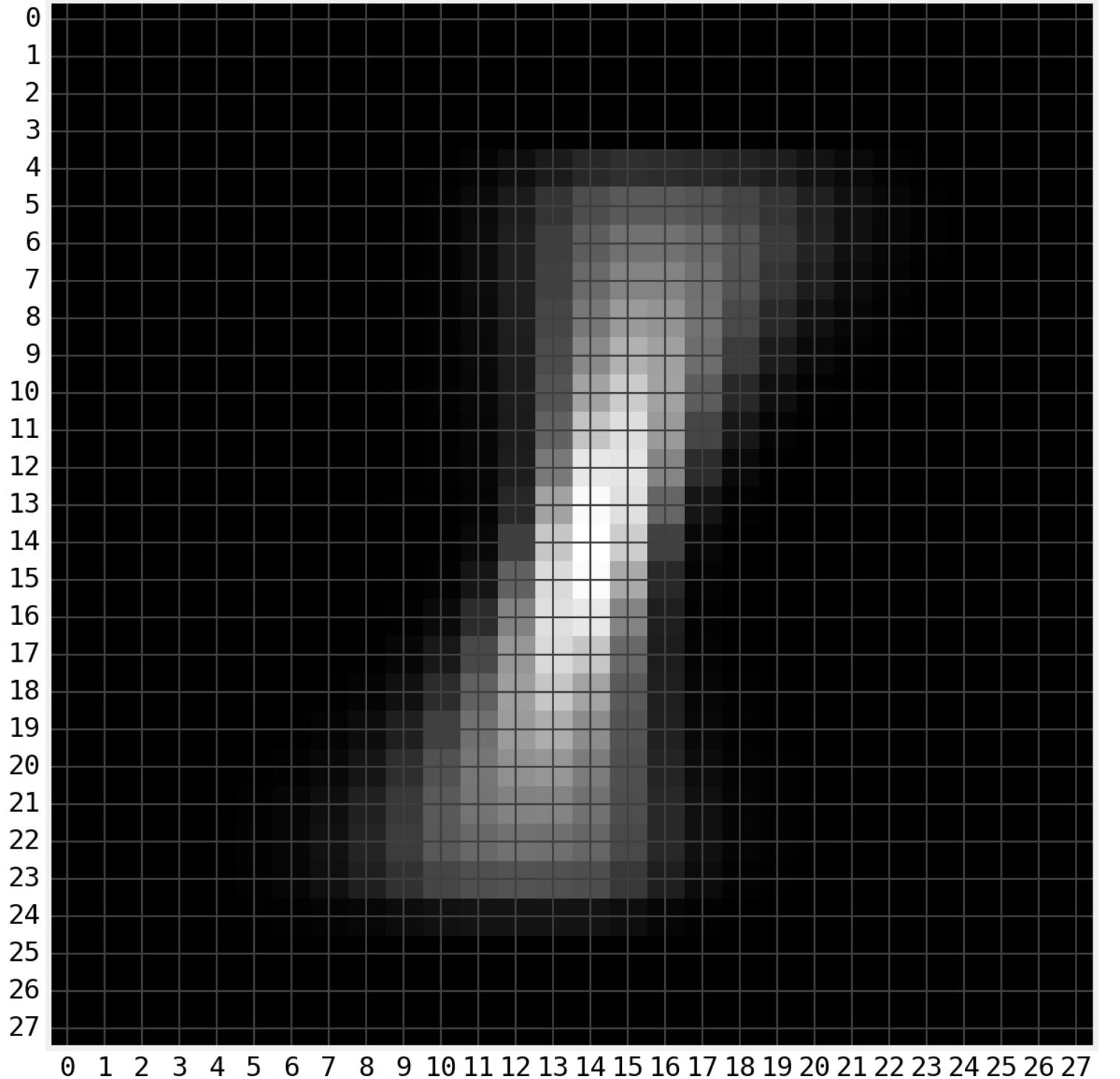
pixel map

n > 5000

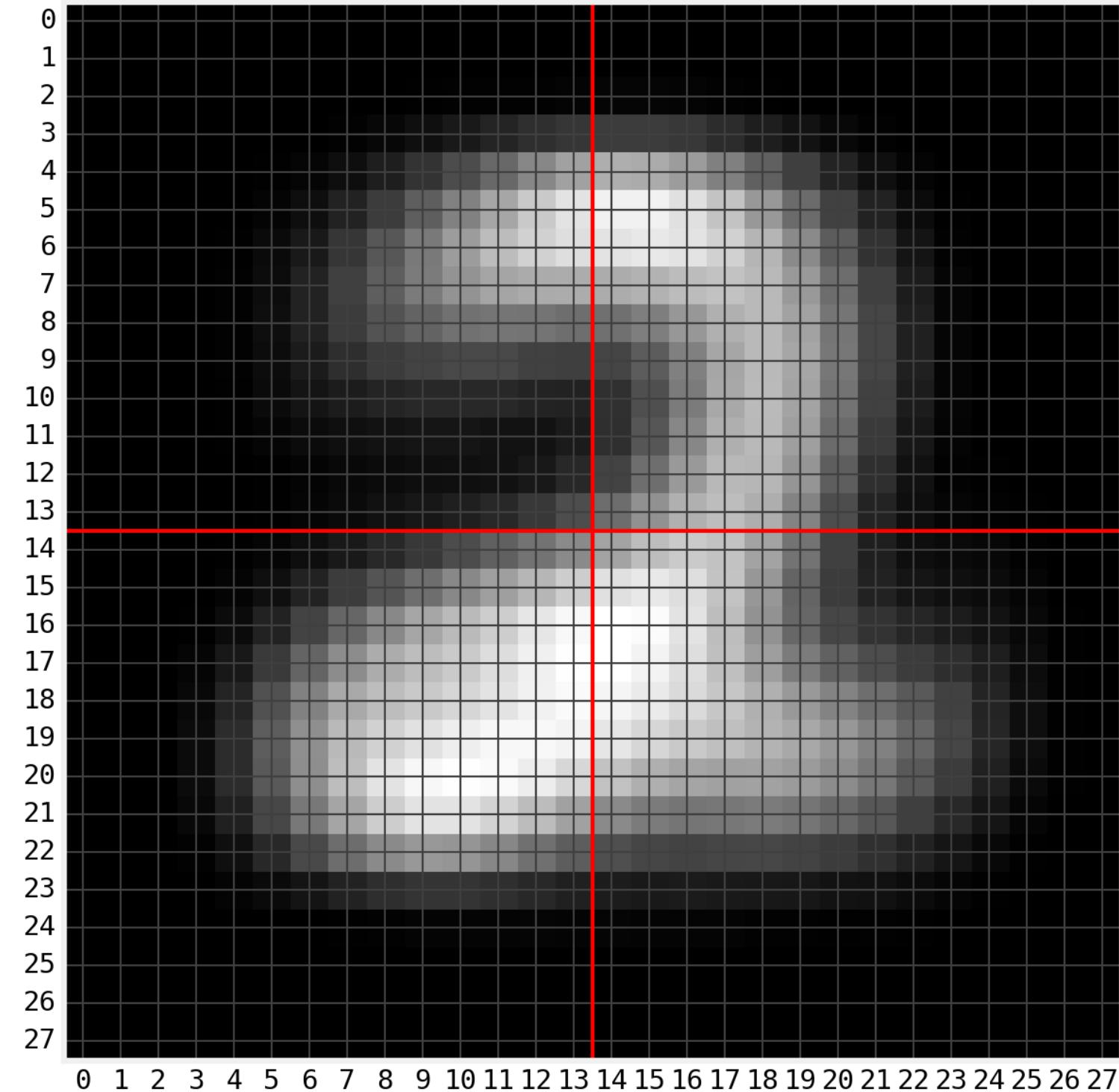
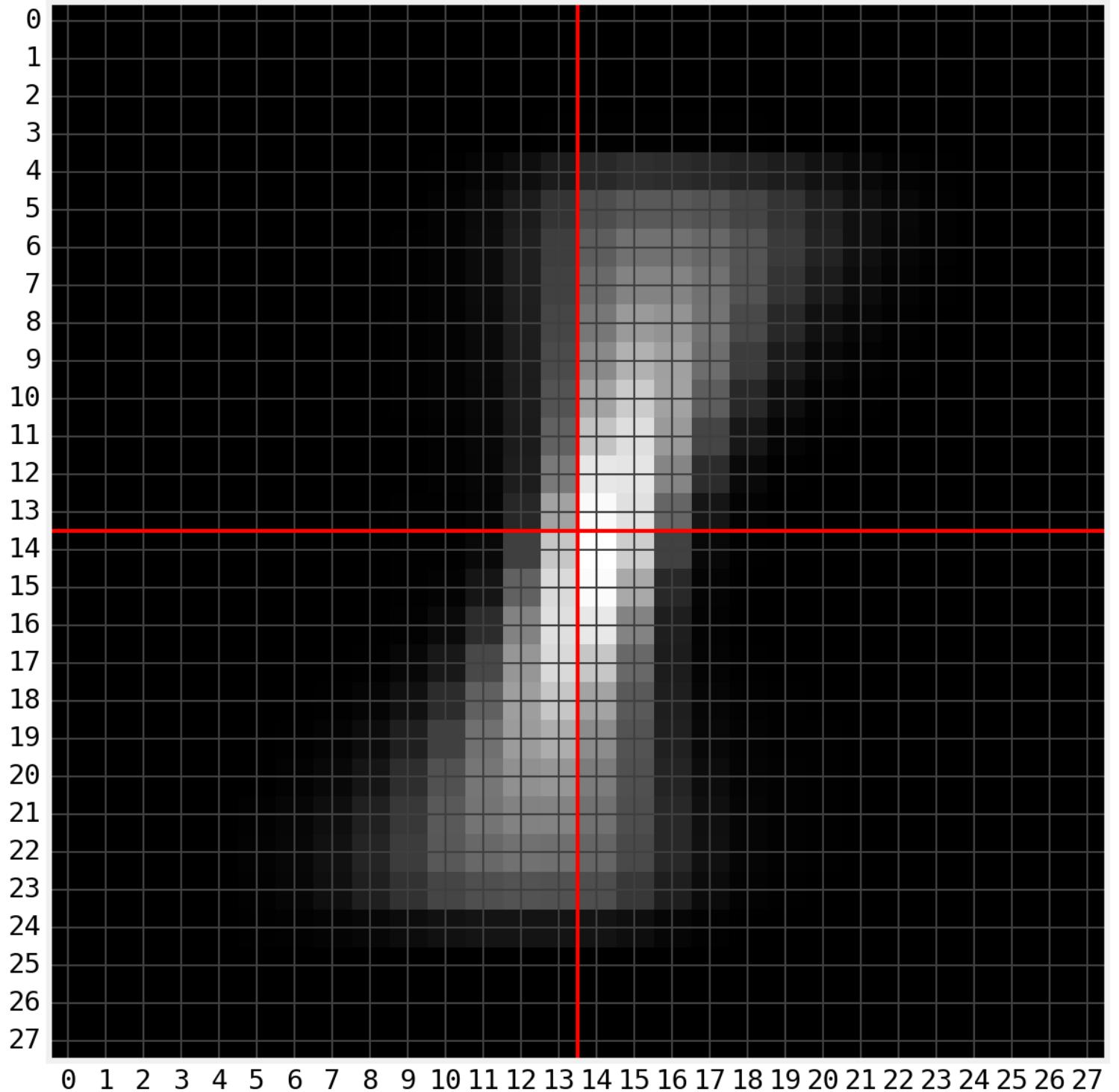
# MNIST dataset: Examples of number **1** and **2**



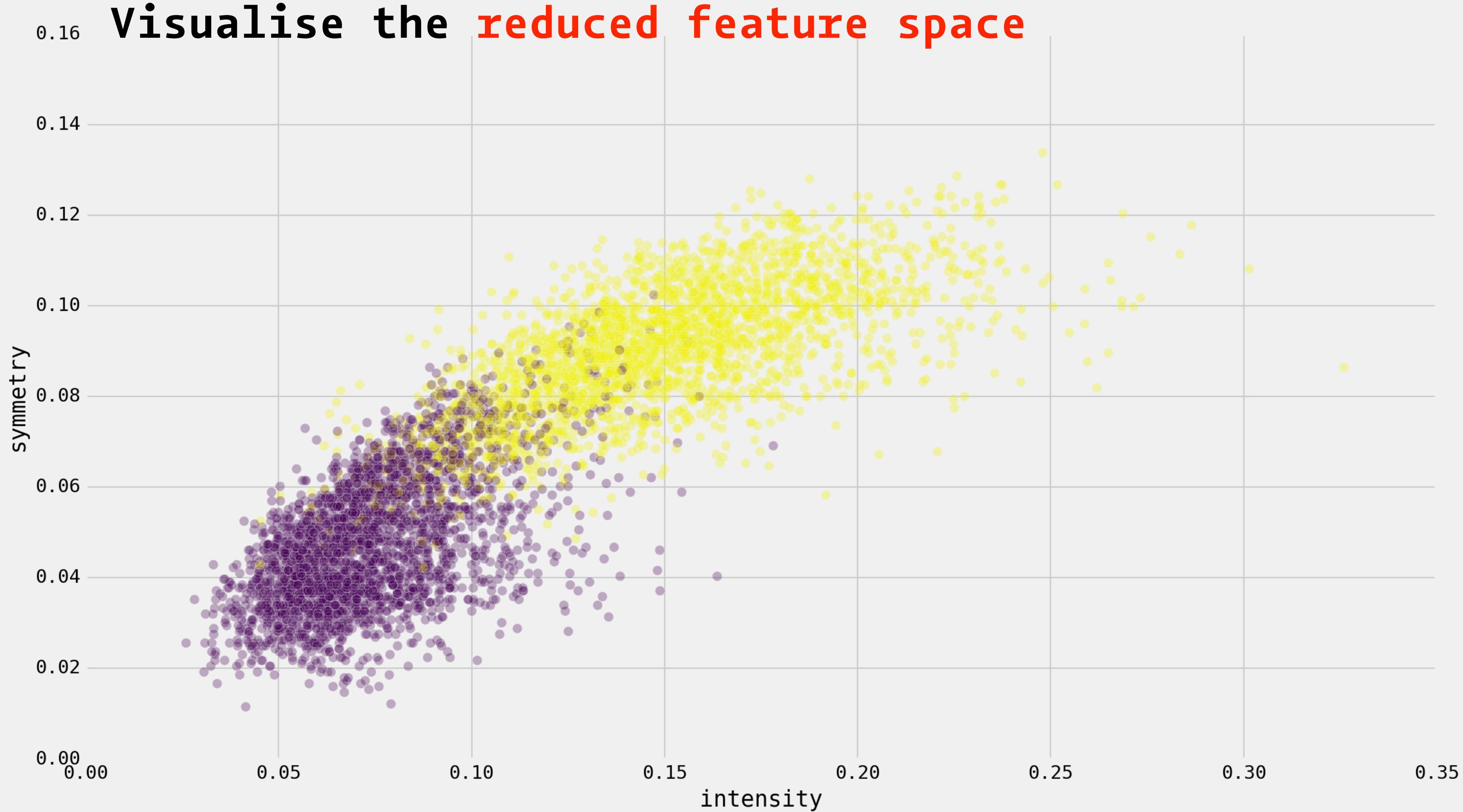
# Visualise the data space



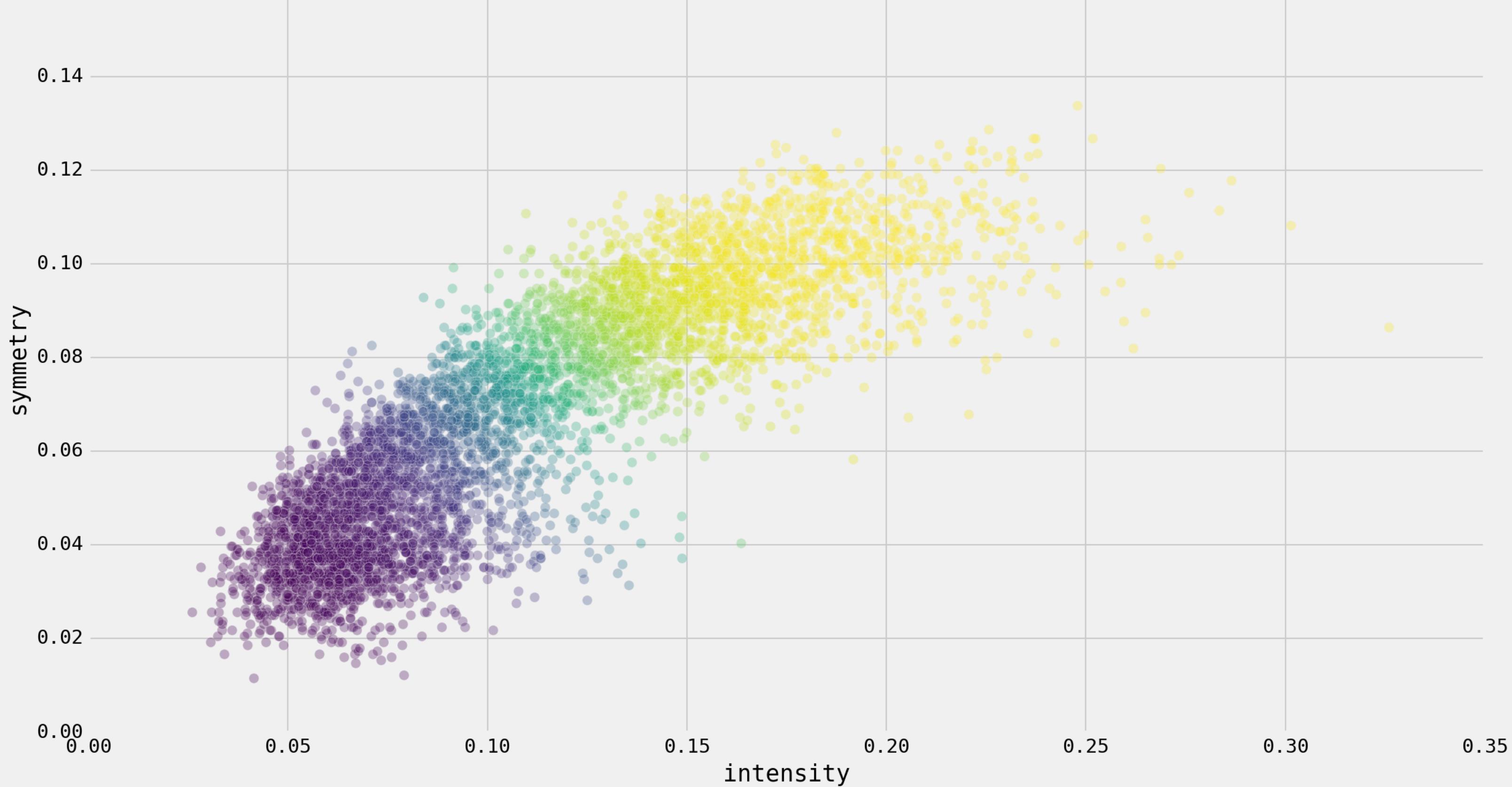
# Identify the features - Symmetry & Intensity



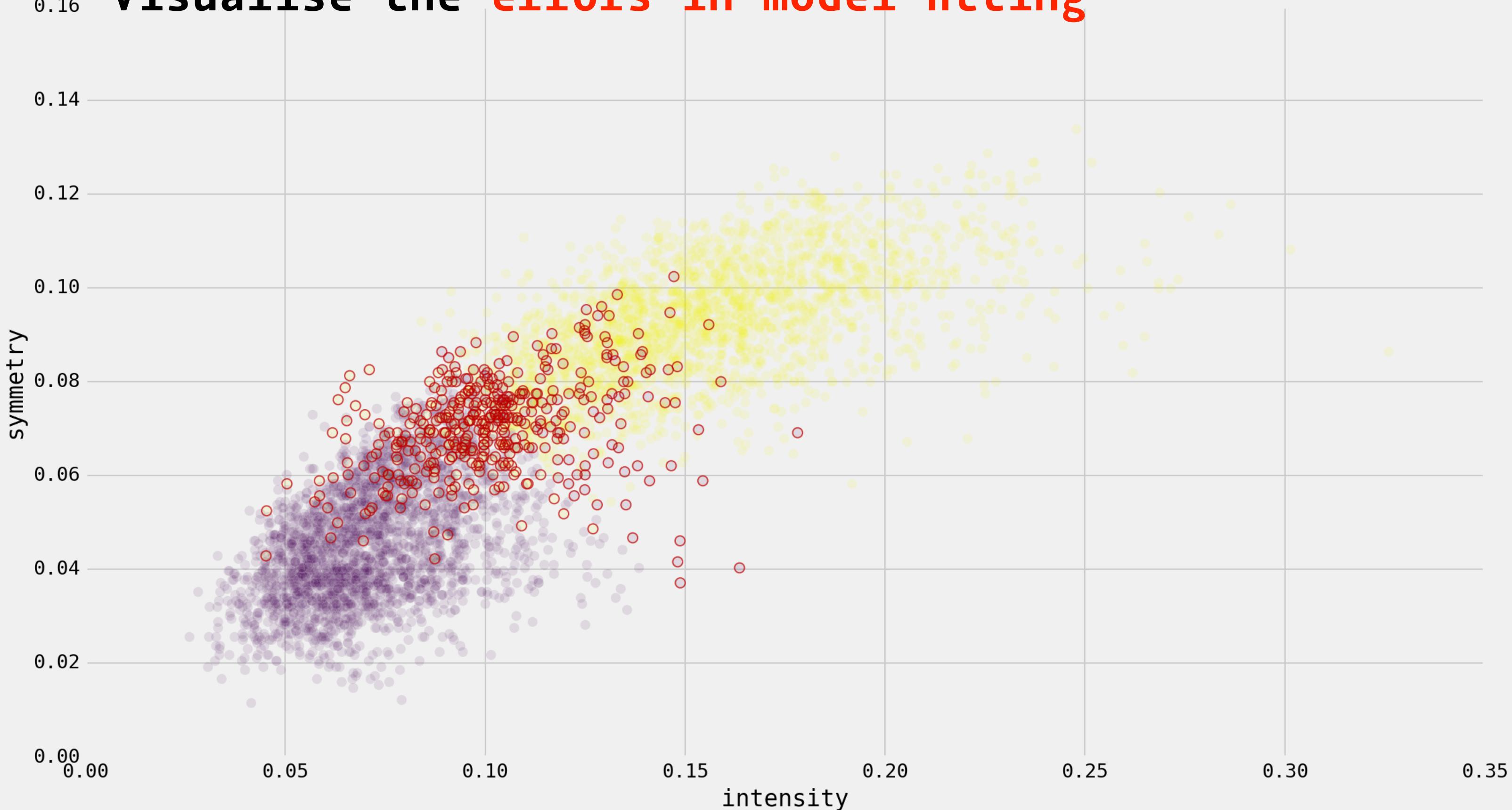
# Visualise the reduced feature space



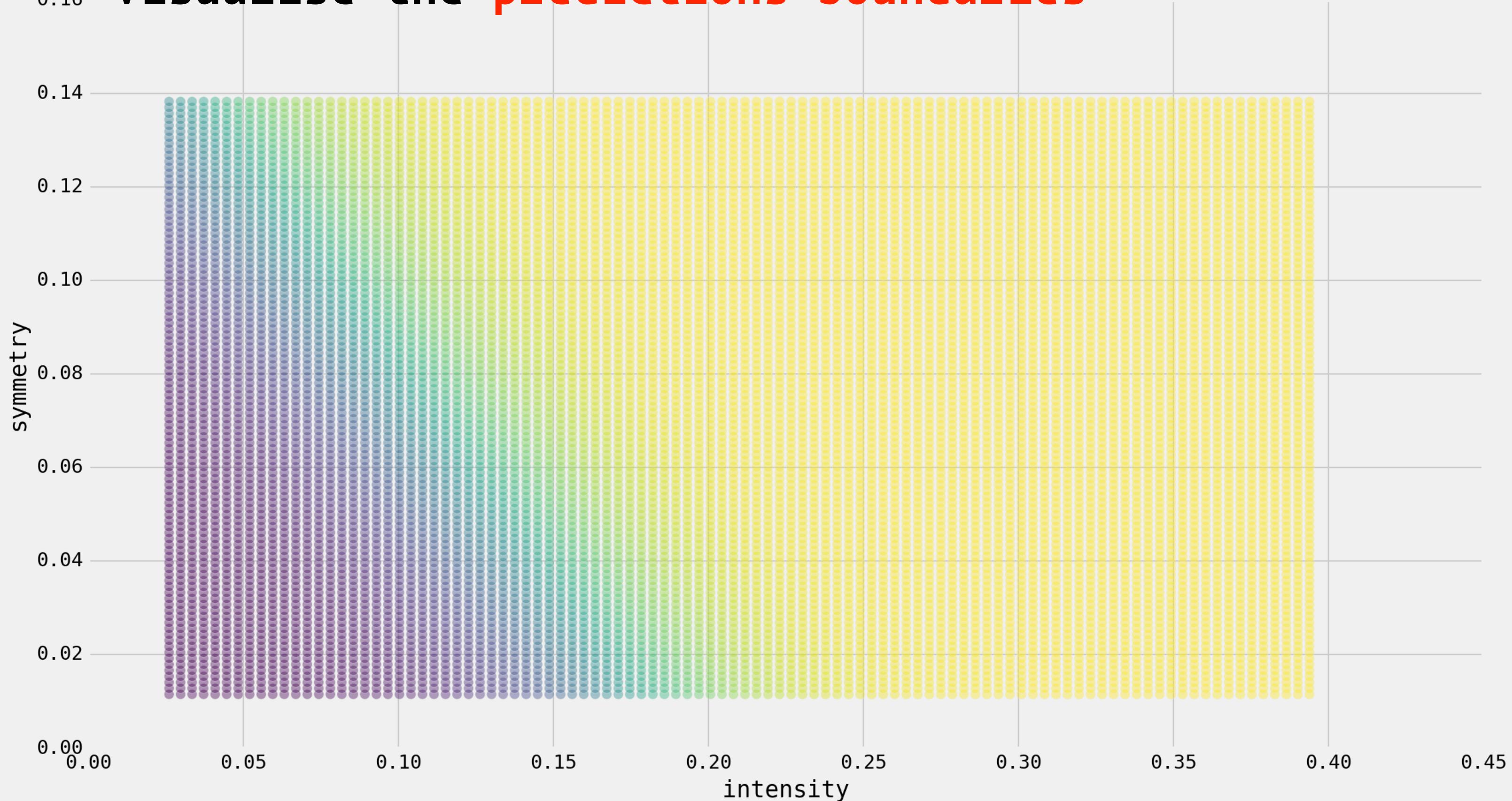
# Visualise the predictions in the data space



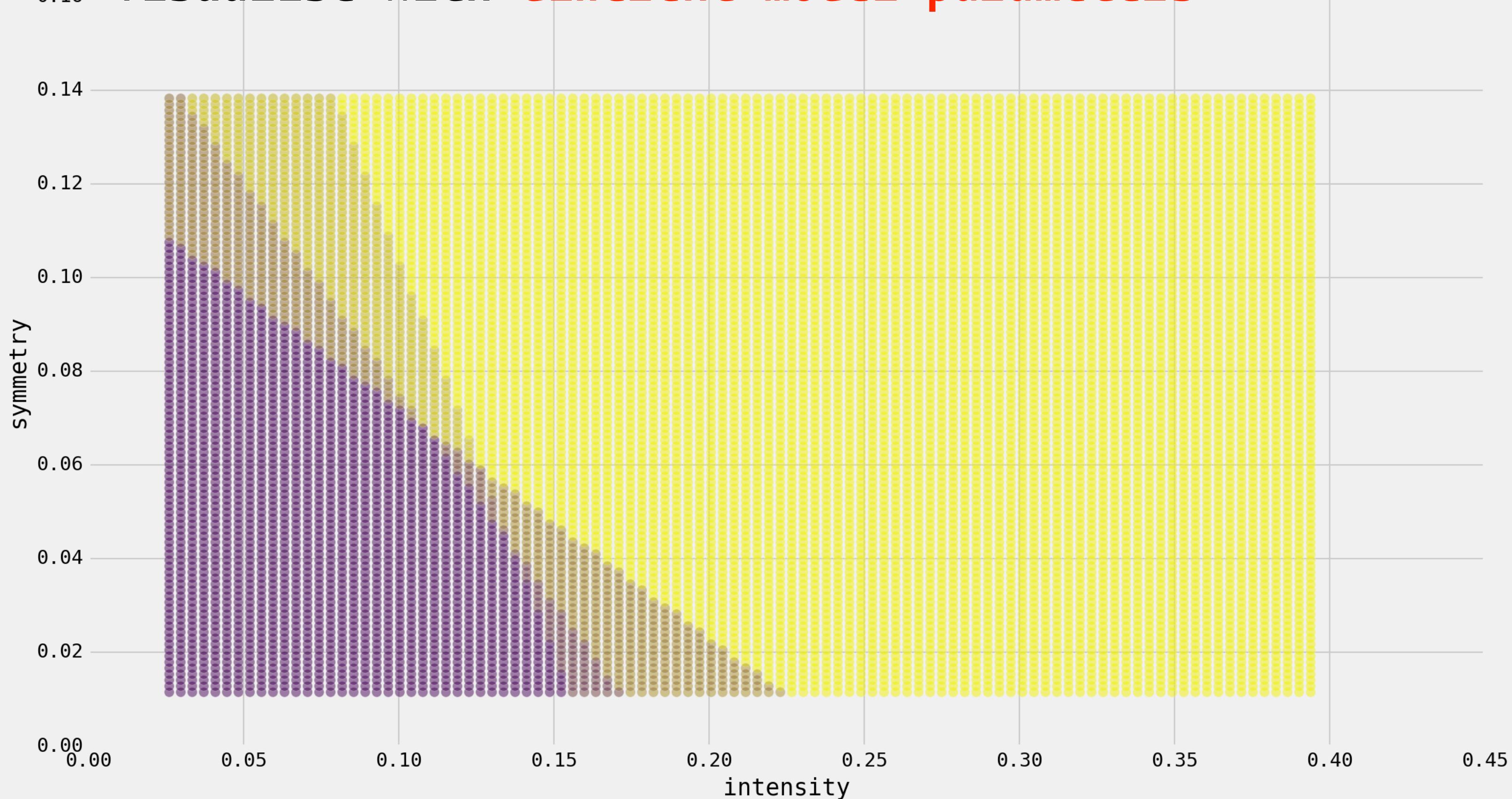
# Visualise the errors in model fitting



# Visualise the predictions boundaries



# Visualise with different model parameters



# n/p/N Model-Vis challenge

---

**n** -- Large and big data

**p** -- High dimensional data

**N** -- Multiple models

# How to scale for large p?

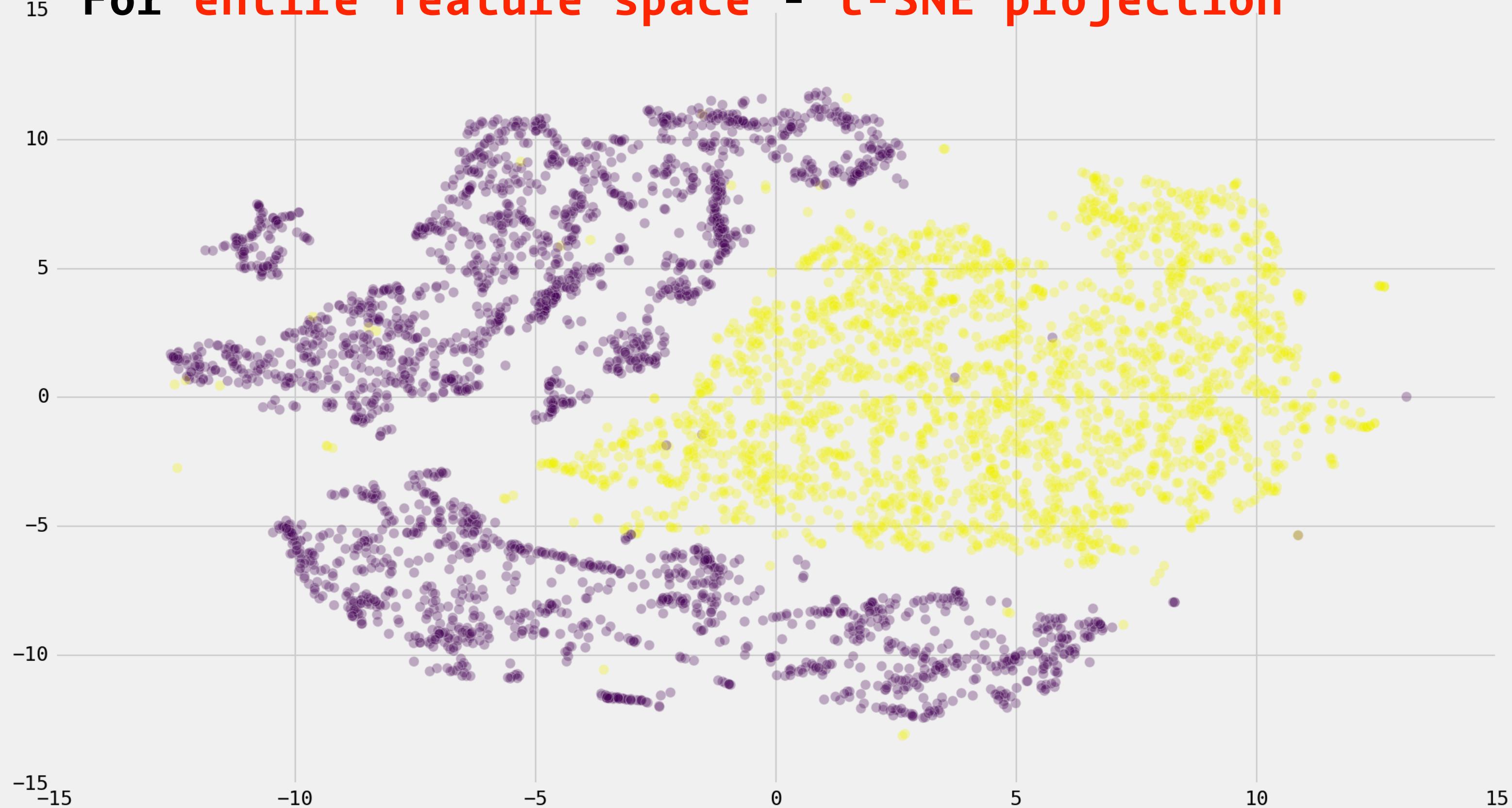
---

**Curse of dimensionality**

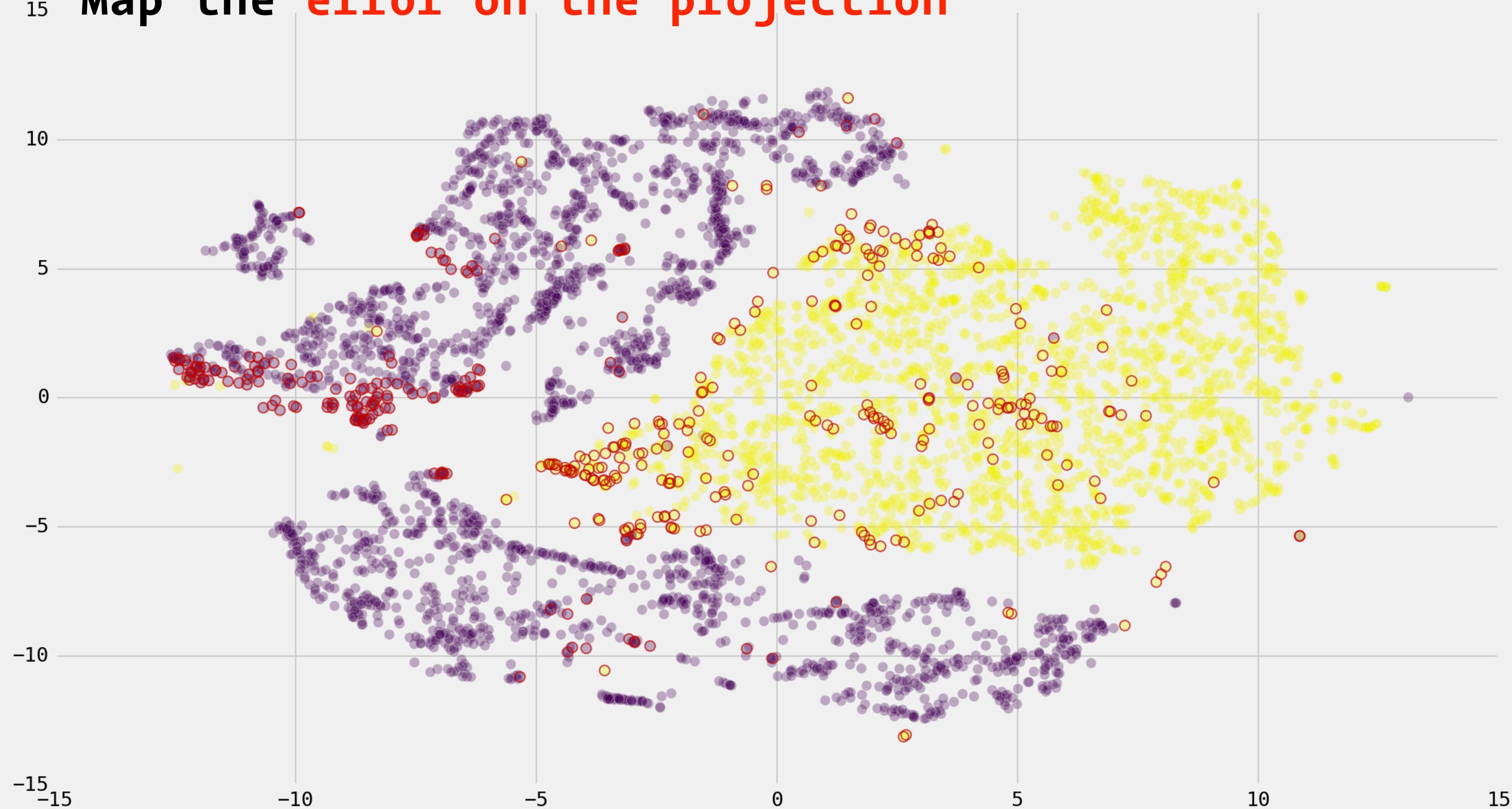
Mesh approach **computationally expensive**

Need to use **projections**

# For entire feature space - t-SNE projection



# Map the error on the projection



# n/p/N Model-Vis approach

---

**n** -- use **Binning or Sampling**

**p** -- use **Projections**

**N** -- use **Summaries**

# In Practice - Model Explosion

---

Entire Model Space

- + Add Tuning Models
- + Add Bootstrap Models
- + Add Ensemble Models
- + Add Cross-Validation Models

# Challenge with Model-Vis

---

Keep track of prediction &  
errors

Keep track of model output  
parameters

# Tidy Model

---

Augment predictions & errors to  
dataset

Create output parameters data  
frame

Visualise like Tidy Data

# Tooling still nascent

---

R: **tidyverse** (esp. **broom** and  
**purr**)

Python: **pybroom**, experimental  
packages

# Model-Viz

---

Similar challenges to Data-Viz

More an Art, than a Science

Essential in ML Model Pipeline

Both to Explain or to Predict

Scope for easier tooling

# Model-Vis

---

Model Visualisation Mini-Site (Coming Soon! )  
<http://modelvis.amitkaps.com>

Code

<https://github.com/amitkaps/modelvis>

# Visualising ML Models

---

Amit Kapoor  
@amitkaps

[amitkaps.com](http://amitkaps.com)