Model Thinking

Scott E Page

Categorical Models

Amazon

“Lump to Live”

Broccoli

Grasshopper

Banana

Candy Bar

Orange

Asparagus

Pear

Strawberry

	Calories
Pear	100
Cake	250
Apple	90
Banana	110
Pie	350

Mean = 180

Pear	100-180	= 80
Cake	250-180	= 70
Apple	90-180	= 90
Banana	110-180	= 70
Pie	350-180	= 170
		480

Pear	$(100-180)^2$
Cake	$(250-180)^2$
Apple	$(90-180)^2$
Banana	$(110-180)^2$
Pie	$(350-180)^2$

$(100-180)^2 = 6400$

$(250-180)^2 = 4900$

Pear	$(100-180)^2 = 6400$
Cake	$(250-180)^2 = 4900$
Apple	$(90-180)^2 = 8100$
Banana	$(110-180)^2 = 4900$
Pie	$(350-180)^2 = 28900$

Total Variation = 53,200

90
100
110

250
350

mean = 100

$(90-100)^2 = 100$

$(100-100)^2 = 0$

$(110-100)^2 = 100$

mean 300

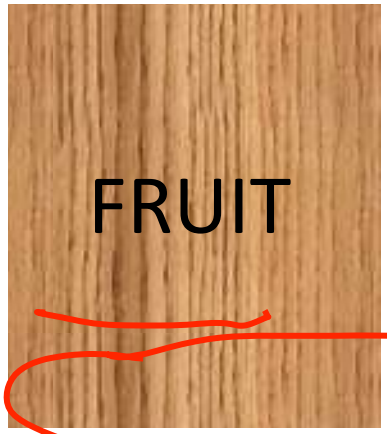
$(250-300)^2 = 2500$

$(350-300)^2 = 2500$

5000

mean = 180

V = 53,200



Mean = 100

Variation = 200

Mean = 300

Variation = 5000

Total Variation = 53,200

Fruit Variation = 200

Dessert Variation = 5000

5200

How much did I explain?

$$\frac{53,200 - 5,200}{53,200} = \frac{48,000}{53,200}$$

R-Squared

% Variation Explained

1 – 5200/53,200

90.2%

48,000 / 53,200

R-Squared

R-squared near 1
model explains a lot

R-squared near 0
model explains little

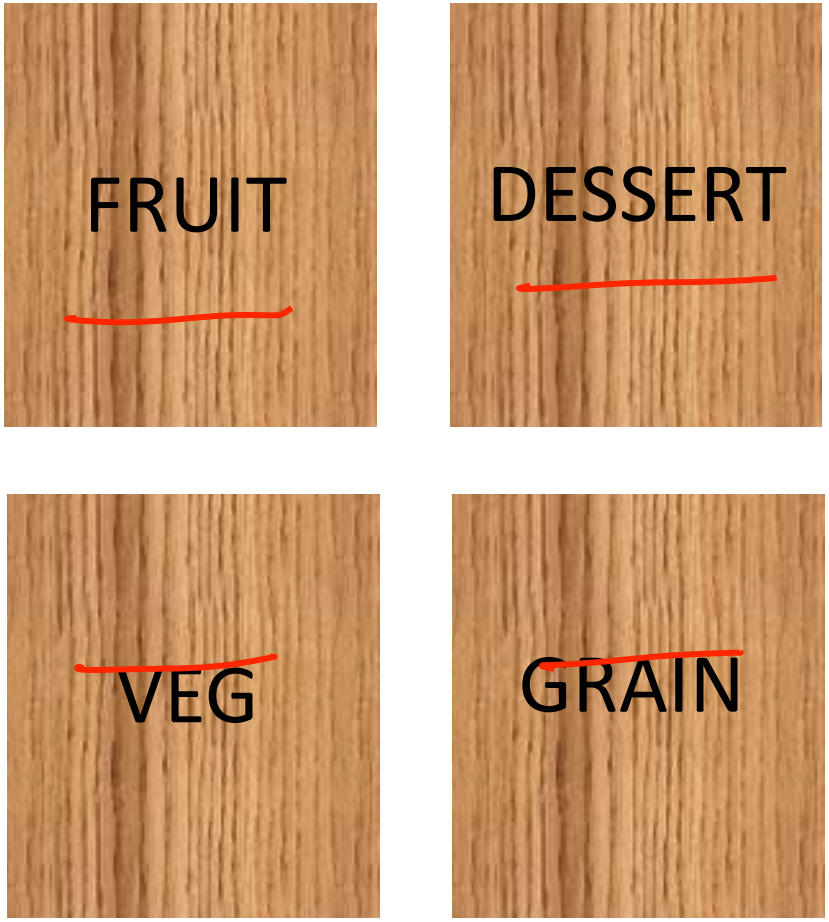
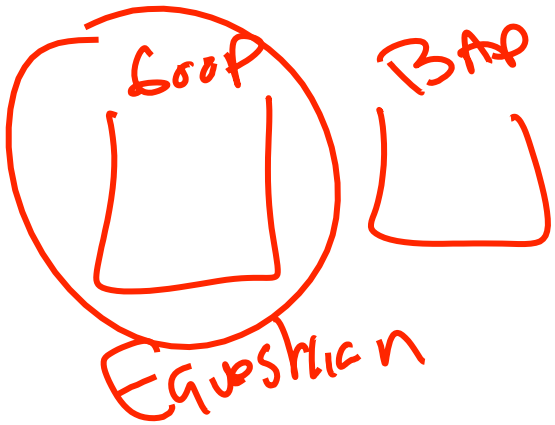

FRUIT
DESSERT
VEG
GRAIN

Photo Simon Howden

Correlation

is not

Causation



Model Thinking

Scott E Page