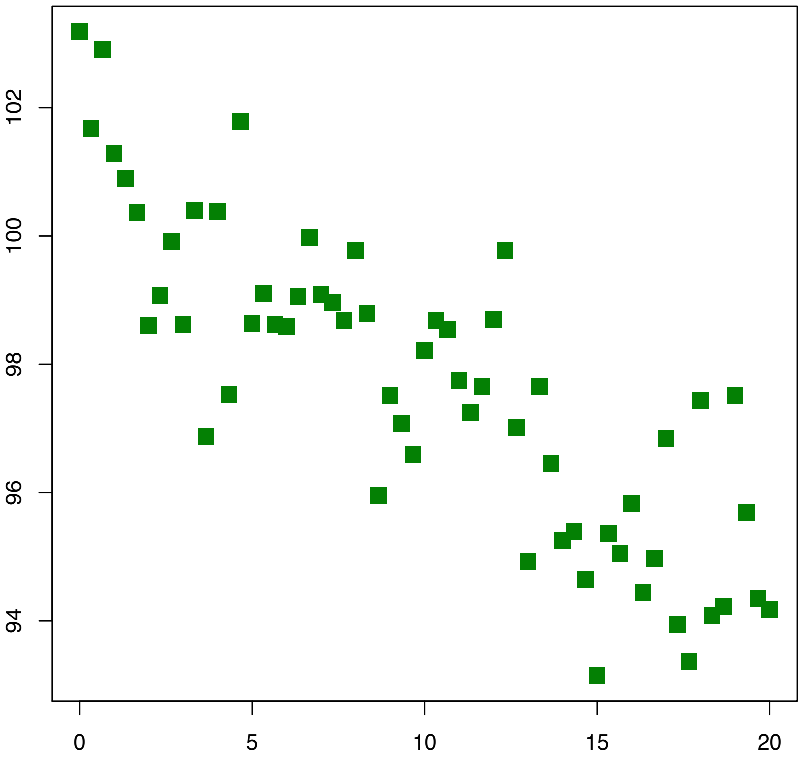
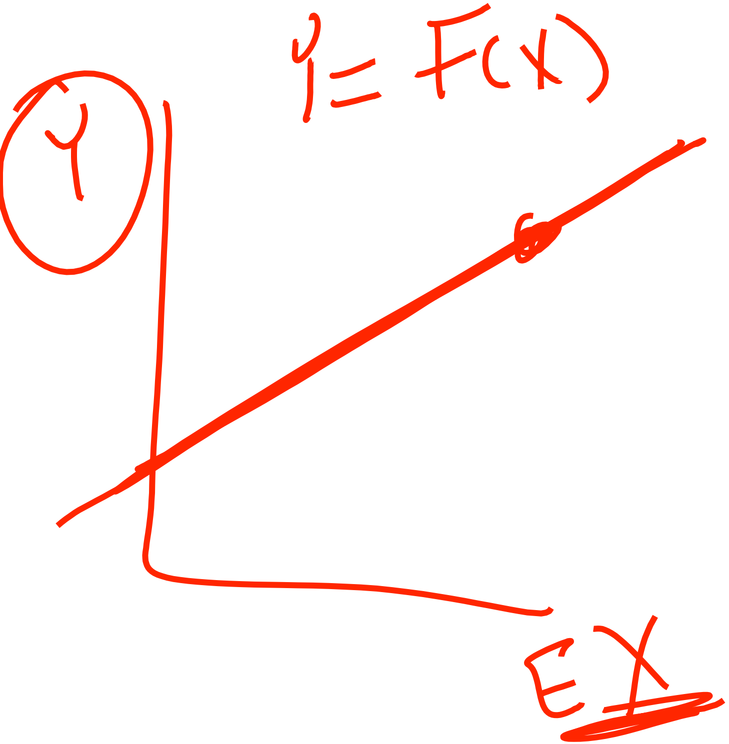


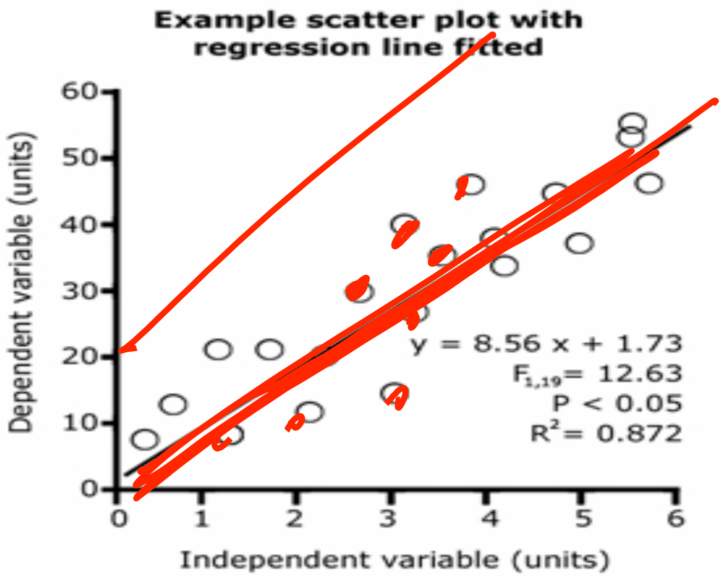
www.wikimediacommons

Categorical Models



Linear Models

Fitting Lines to Data



http://web.anglia.ac.uk

Non Linear Models

$y = a_1 x_1 + a_2 x_2$

Big Coefficient

Model Thinking

Scott E Page

Model Thinking

Scott E Page

Categorical Models

Amazon

INFO

DELIVERY

“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

FRUIT

DESSERT

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

FRUIT

DESSERT

Mean = 100

Mean = 300

Variation = 200

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%

$\frac{48,000}{53,200}$

R-Squared

R-squared near 1

model explains a lot

R-squared near 0

model explains little

FRUIT

DESSERT

VEG

GRAIN

Equestrian

Photo Simon Howden

Correlation

is not

Causation

GOOD

BAD

Equestrian

Model Thinking

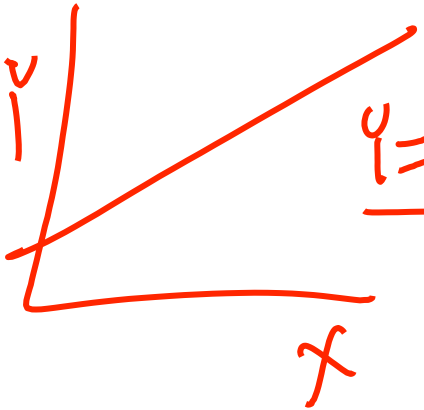
Scott E Page

Model Thinking

Scott E Page

X

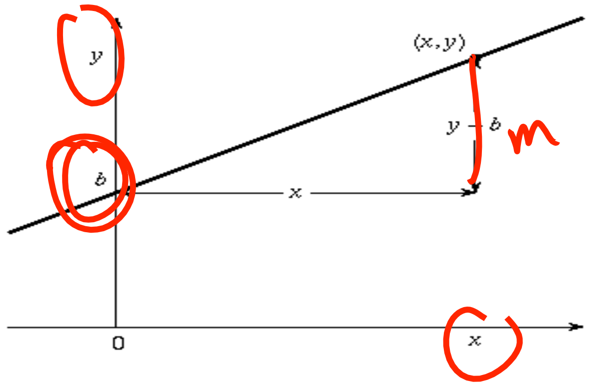
$Y = F(X)$



$y = mx + b$

Linear Models

$y = m\underline{x} + b$

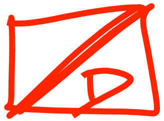


Linear Model vs Line

X = Independent Variable

Y = Dependent Variable

Y depends on X



X = Length of Diagonal

Y = Cost of TV

Linear Model:

Cost = 15*Length + $100

$Y = 5X$

Sign: does Y increase or decrease in X?

Magnitude: how much does Y increase for each one unit increase in X?

x

$$\text{Cost} = 15*\text{Length} + 100$$

Predict

Understand Data

Cost = 15*Length + 100

30 inch TV?

$C = 15(30) + 100$

$450 + 100$

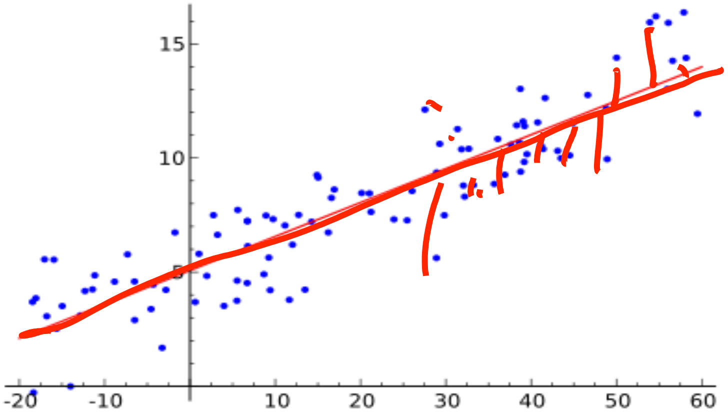
$\$550$

Cost = 15*Length + 100

100 inch TV?

$C = 15(100) + 100$

$1600



www.wikimedia.org

Robyn Dawes 1979: "The Robust Beauty of Improper Linear Models in Decision Making"

43 bank loan officers predict which 30 of 60 firms would go bankrupt. They see financial statements.

Bankers: 75 % accurate

Linear Model: ratio of assets to liabilities 80%

Mehl (1954) 20
studies of clinicians

Sawyer (1966) 45
studies of predictions
in the social world.

Experts NEVER did
significantly better

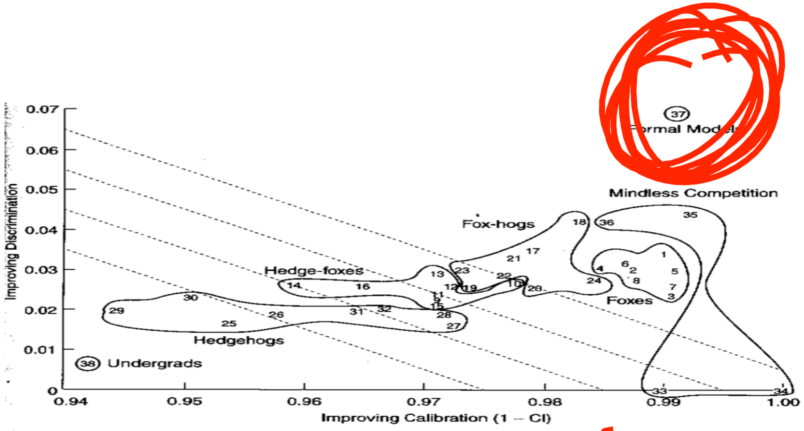
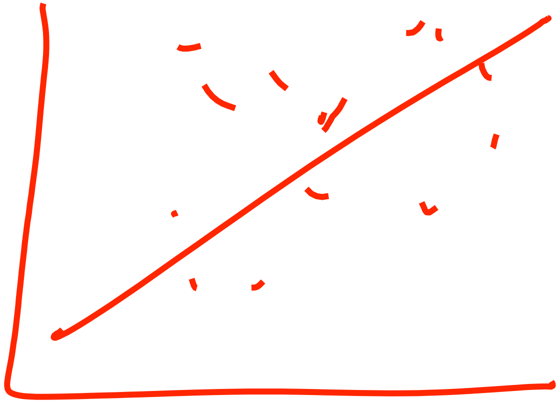


Figure 3.2. How thoroughly foxes and fox-hog hybrids (first and second



Model Thinking

Scott E Page

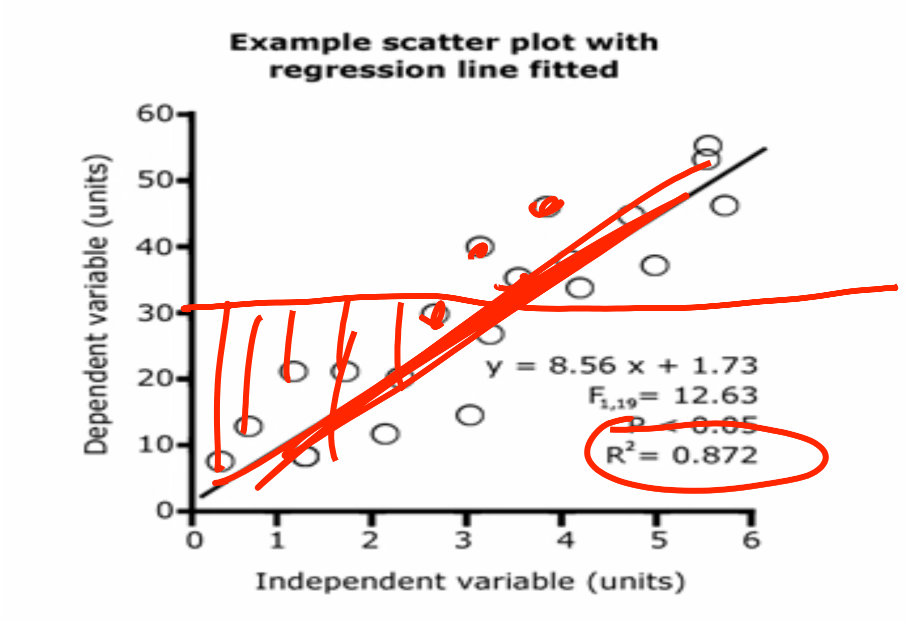
Model Thinking

Scott E Page

Fitting Lines to Data

R-squared:

% Variation Explained



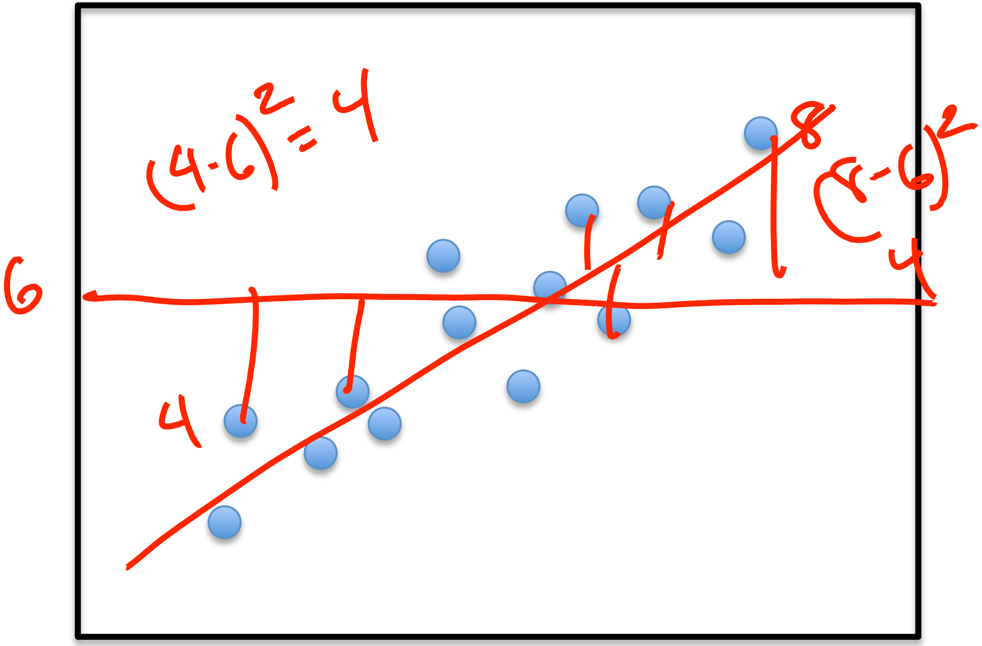
http://web.anglia.ac.uk

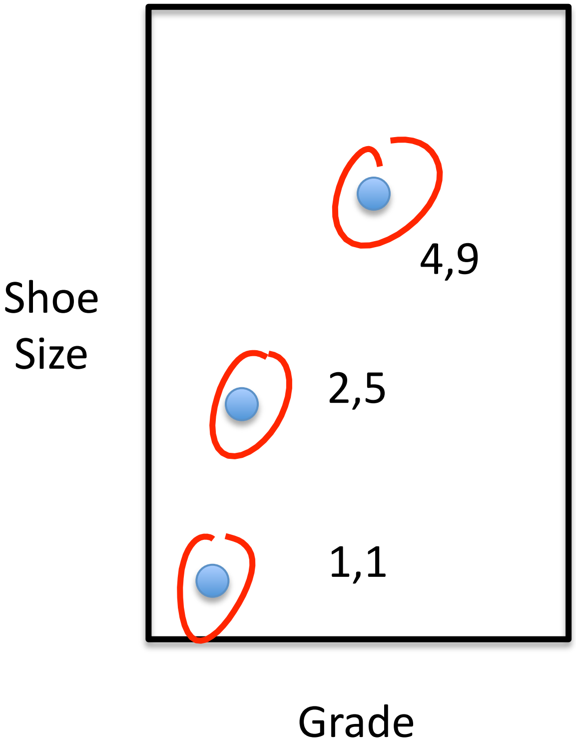
R-Squared

% Variation Explained

1 – 5200/53,200

90.2%


$(4-6)^2 = 4$
$(8-6)^2$
4
6
4
8
4


Shoe
Size
4,9
2,5
1,1
Grade

Variation?

G S

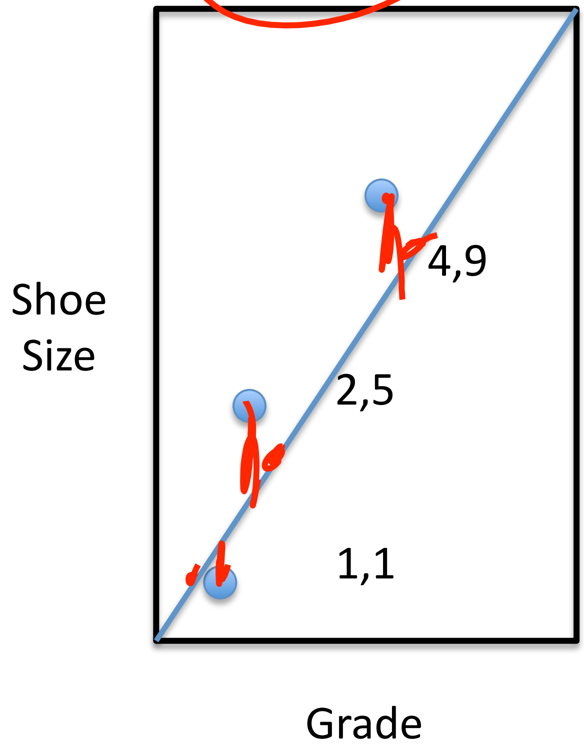
(1,1) $(1-5)^2 = 16$

(2,5) $(5-5)^2 = 0$

(4,9) $(9-5)^2 = 16$

$15/3 = 5$ 32

Y=2X



Y=2X

32

(X,Y,2X)

(1,1,2) $(2-1)^2 = 1$

(2,5,4) $(4-5)^2 = 1$

(4,9,8) $(8-9)^2 = 1$

3

$1 - \frac{3}{32} = 90 + \%$

(1,1,2)

(2,5,4)

(4,9,8)

R-Squared

$Y = mX + b$

$x=1$ $(m+b-1)^2 = (m^2+2mb+b^2-2m-2b+1)$

$x=2$ $(2m+b-5)^2$

$(4m+b-9)^2$

$21m^2+14mb+3b^2$ -94m-30b +81

Total Variation

b= -1

M = 8/3

m

X,Y, MODEL

(1,1, 5/3)

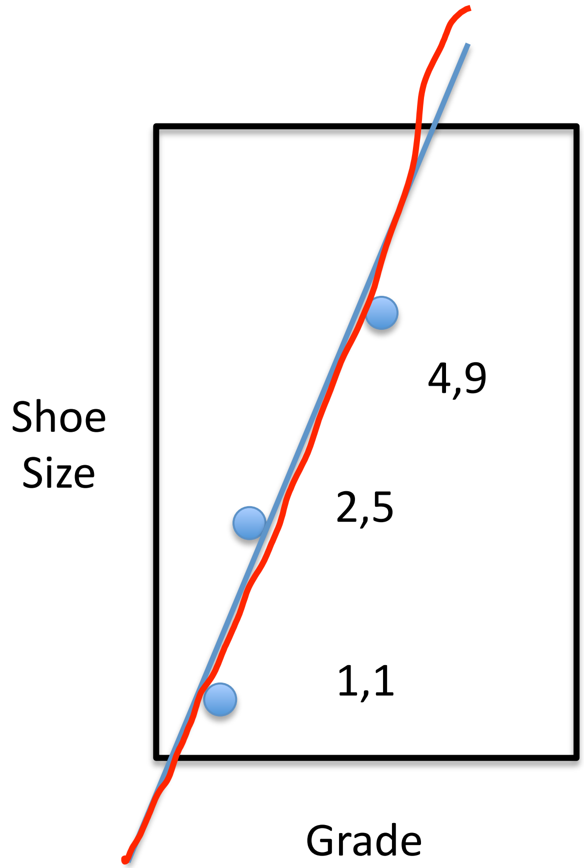
(2,5, 13/3)

(4,9, 29/3)

R-Squared

$1 - \frac{4/3}{32}$

95%


Shoe
Size
4,9
2,5
1,1
Grade

Model Thinking

Scott E Page

Model Thinking

Scott E Page

R-Squared:	0.72
Standard Error:	24.21
Observations:	50

	Coeff	SE	P-value
Intercept	25	2	0.000
X1	20	1	0.000
X2	10	4	0.014

$Y = mx + b$

$Y = \underline{m_1 x_1} + \underline{m_2 x_2} + b$

Multiple Variables

Y = Test Score

T = Teacher Quality

Z = Class Size

$$Y = cT + dZ + b$$

$c > 0$ $d < 0$

$$Y = \overbrace{a}^{\text{Coef}}X_1 + \overbrace{b}X_2 + \underbrace{c}_{\text{Int}}$$

R-Squared: 0.72
Standard Error: 24.21
Observations: 50

	Coeff	SE	P-value
Intercept	25	2	0.000
X1	20	1	0.000
X2	10	4	0.014

$Y = 20X_1 + 10X_2 + 25$

Sign: does Y increase or decrease in X?

Magnitude: how much does Y increase for each one unit increase in X?

VARIATION R^2

SIGN

magnitude

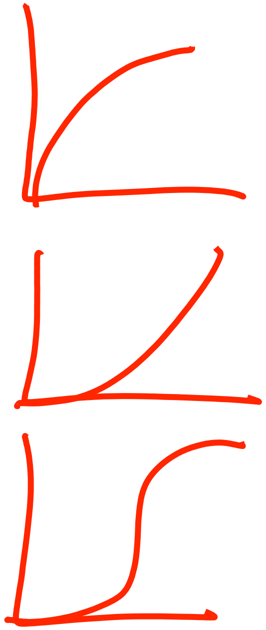
P-value

Model Thinking

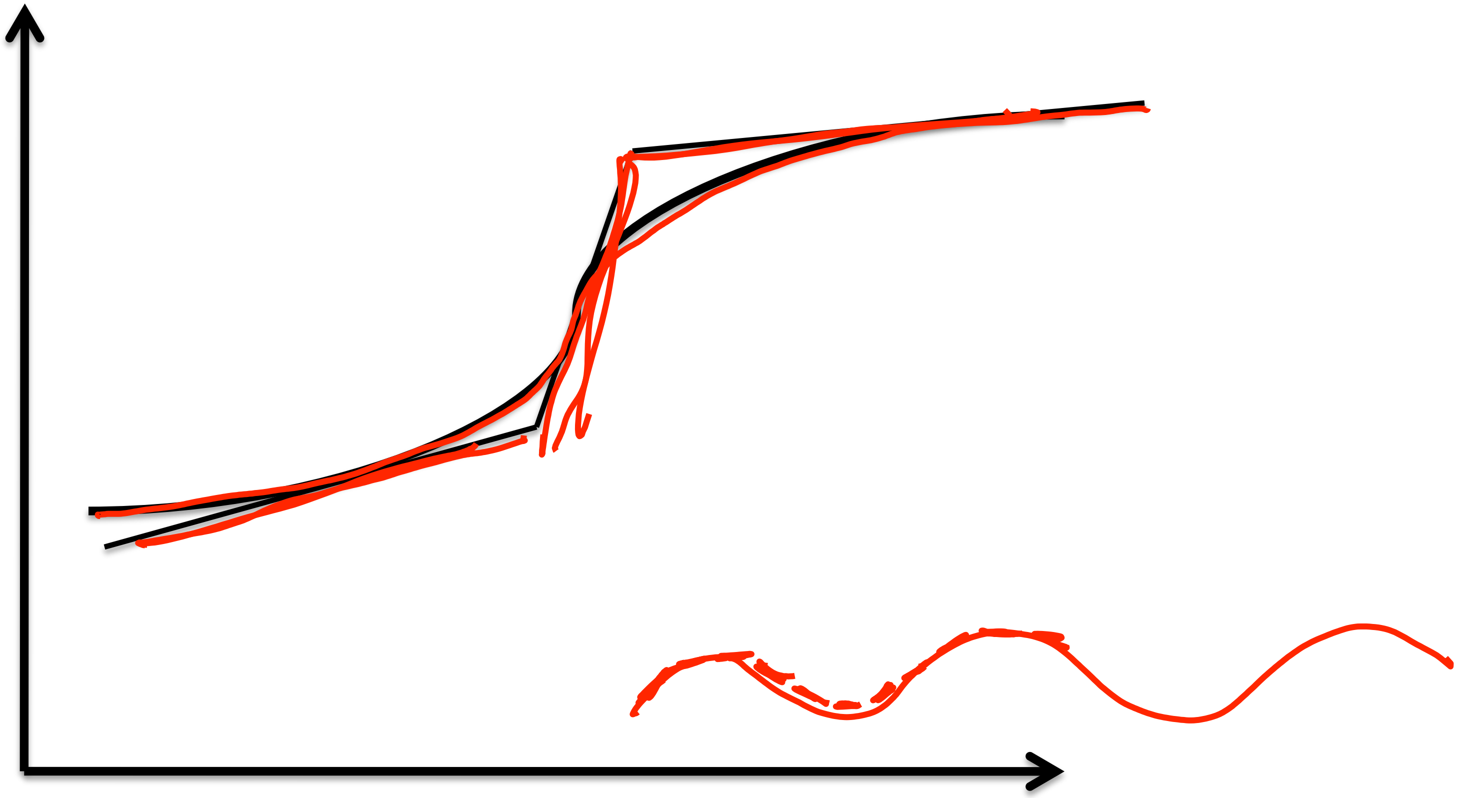
Scott E Page

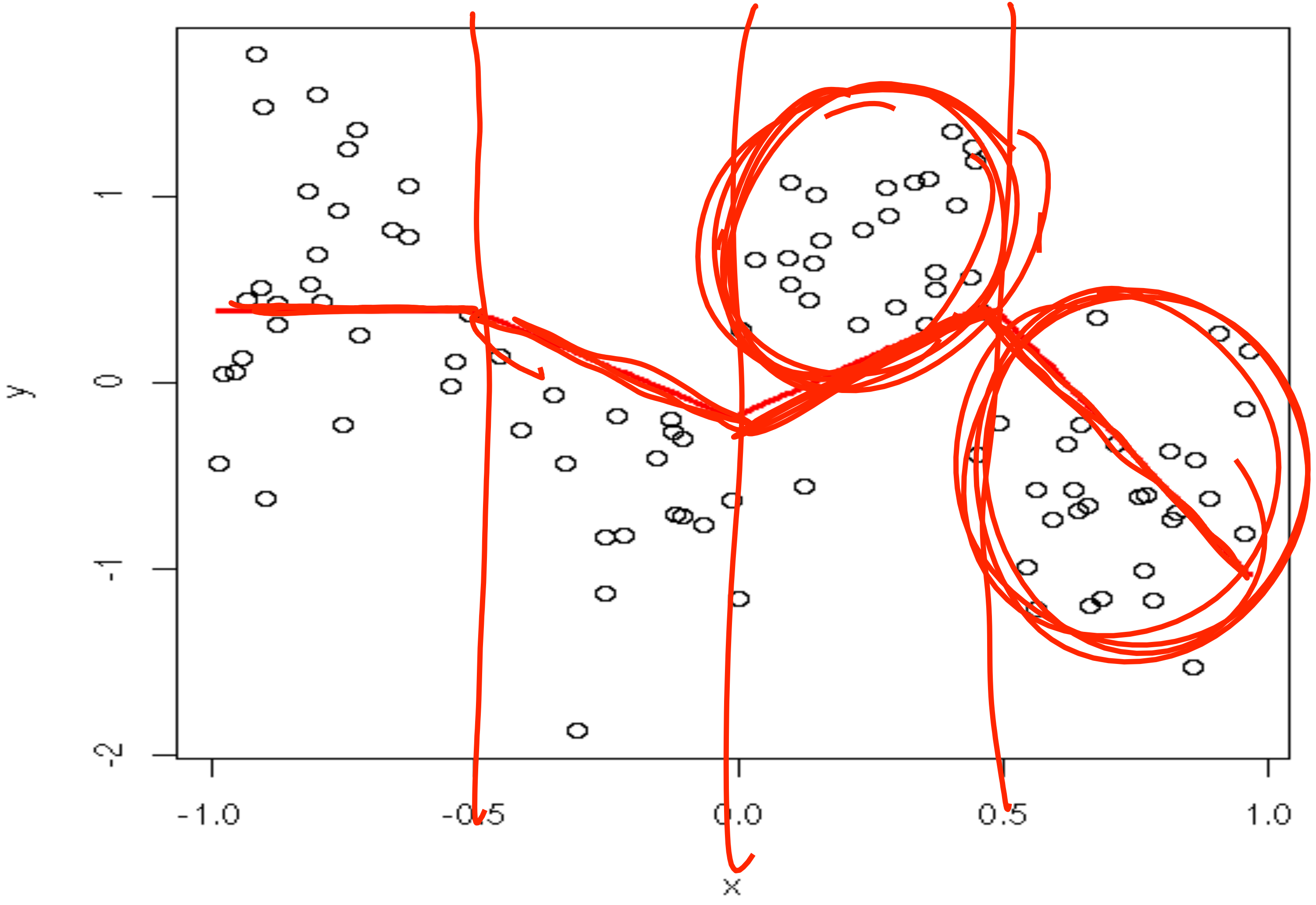
Model Thinking

Scott E Page

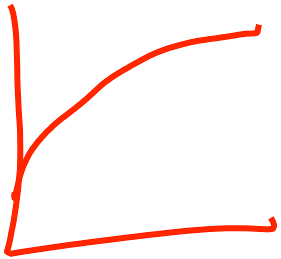


Non Linear





y
1
0
-1
-2
-1.0
-0.5
0.0
0.5
1.0
x

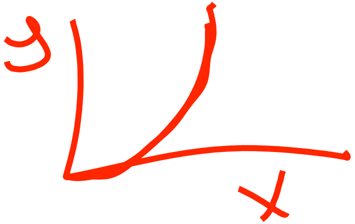


$y = m\sqrt{x} + b$

$y = mz + b$

"$z = \sqrt{x}$"

$z = x^2$ $\quad$ $y = mz + b$



Non Linear Terms

Model Thinking

Scott E Page

Model Thinking

Scott E Page

$$Y = a_1 x_1 + a_2 x_2 + b$$

The Big Coefficient

Evidence Based_____

Medicine

Philanthropy

Education

Management

Construct Model

Gather Data

Identify important variables

Change those variables

Big Data

1 Gather Data

Find Pattern

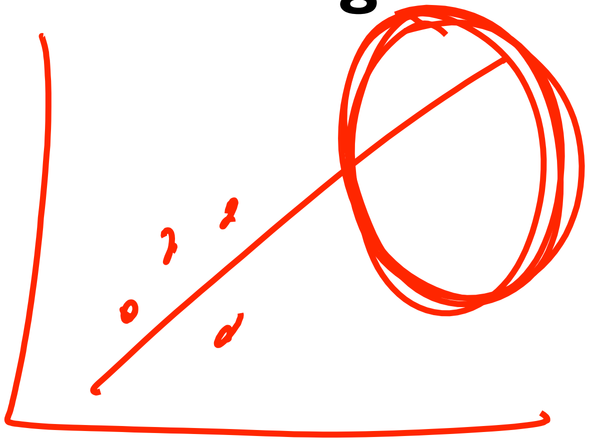
Identify important variables

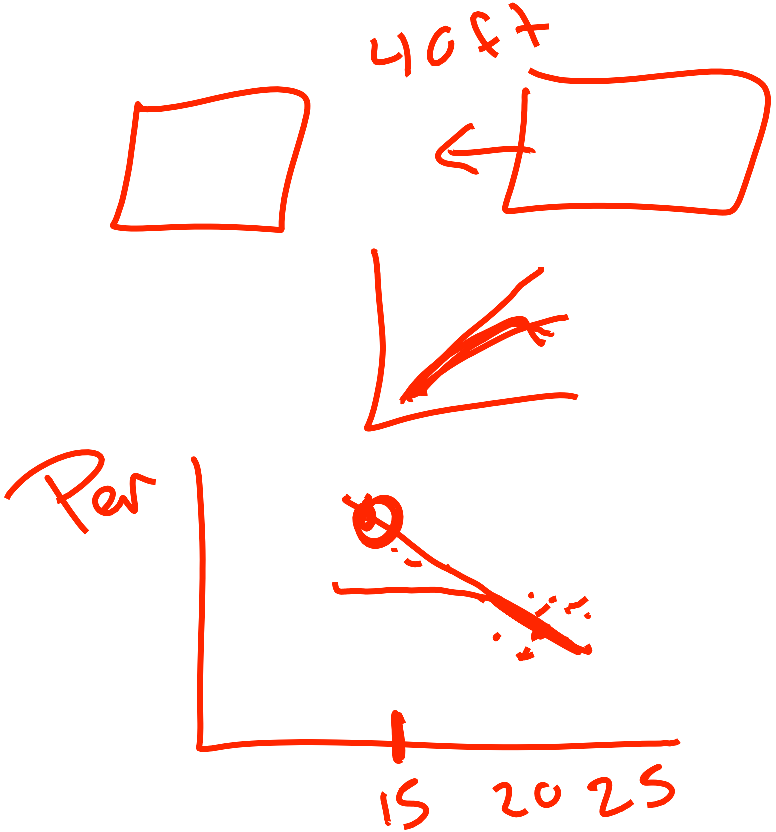
Change those variables

Big data does not obviate the uses of models

Correlation
is not
Causation

Linear models tell sign and magnitude of changes in independent variables **within data range**





Feedbacks

Multiple Peaks

The New Reality

Big Coefficient

Tax Cigarettes

New Reality

Universal Health Care

Big Coefficient

Increase HOV Lanes

New Reality

Rail System

Big Coefficient

Oat Bran Pretzels

New Reality

Fitness Regimine

American Jobs Act

1. Tax Cuts to Help America's Small Businesses Hire and Grow
   - Cutting the payroll tax
   - Payroll tax holiday for new workers and higher wages
   - 100% expensing
2. Rebuilding and Modernizing America
   - Subsidies to hire veterans
   - Save 280,000 teachers jobs
   - Infrastructure and infrastructure bank
   - Modernize Schools and buildings
   - High speed wireless
3. Pathways to Work
   - Rewrite unemployment insurance
   - $4000 tax credit for new employees
4. Tax Relief
   - Cuts in payroll taxes
   - Allowing more refinancing

Interstate Highway System

41,000 miles of roads

$25 billion

CPI: $207 Billion today

Mile: $10 million a mile

$410 Billion

Model Thinking

Scott E Page