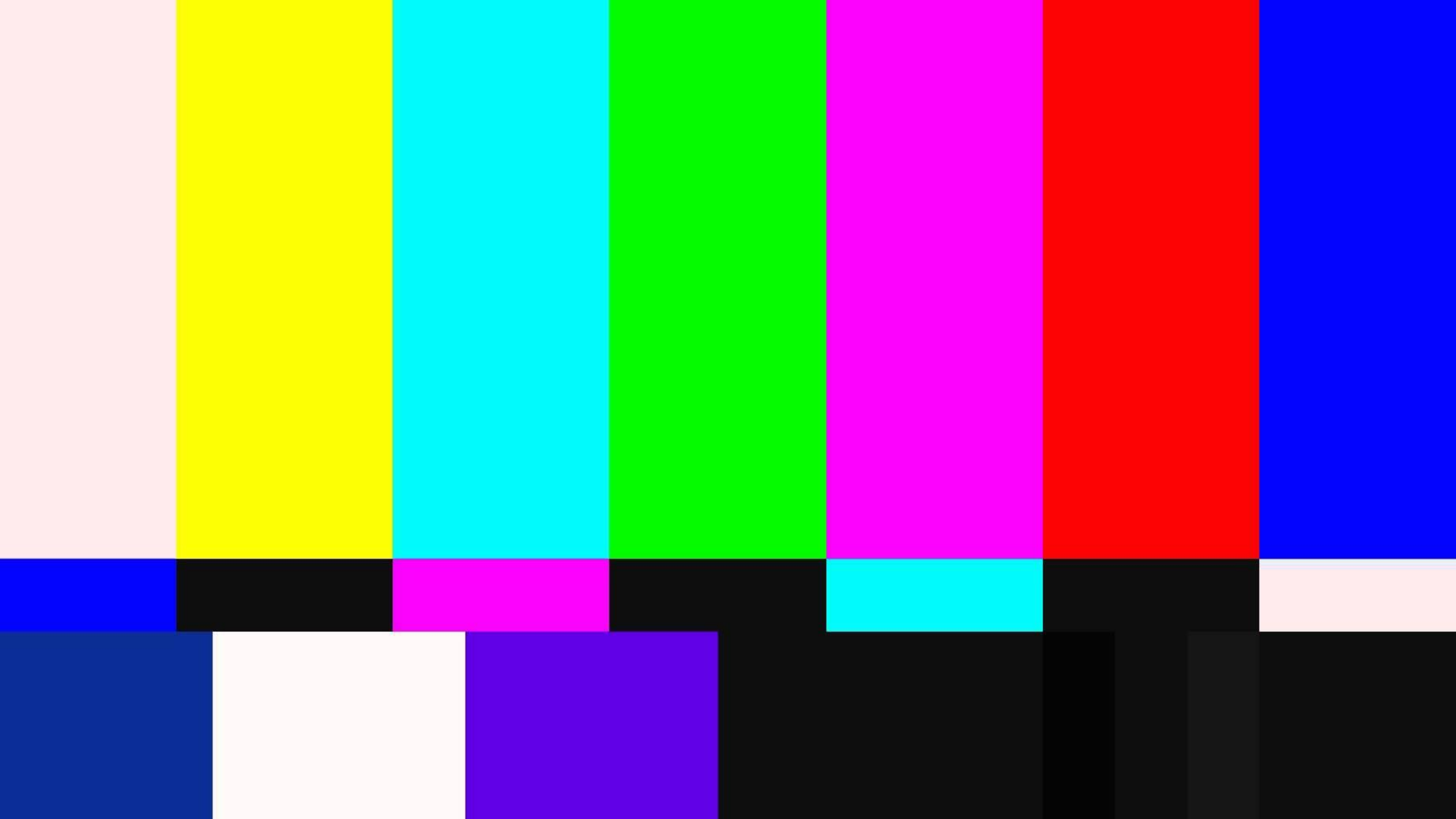


KEEP
CALM
AND
EAT A
COOKIE

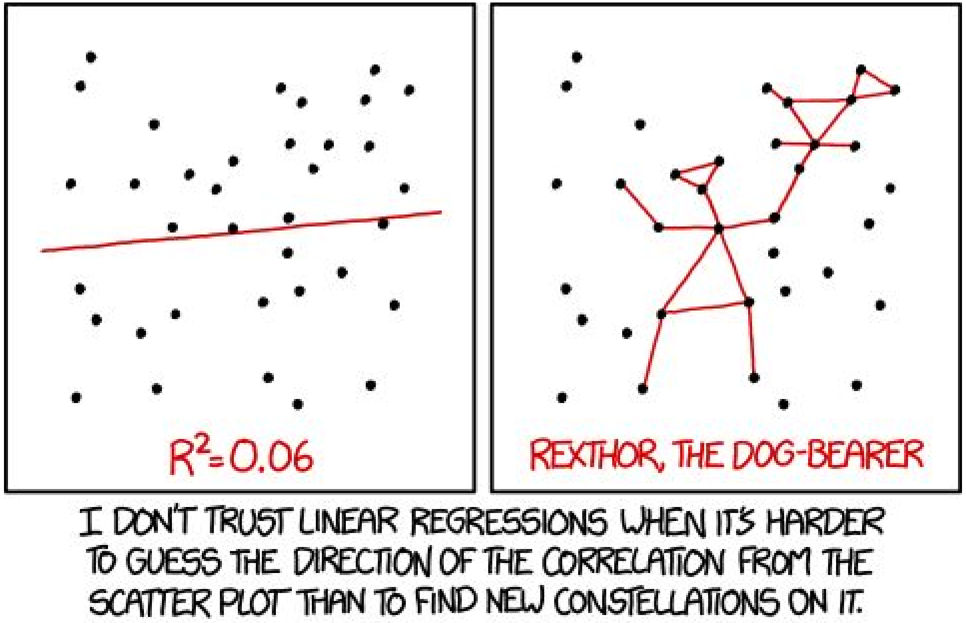
Linear Regression

or: How I Learned to Stop Worrying and Love Data Science

[A work of possibly some fiction]

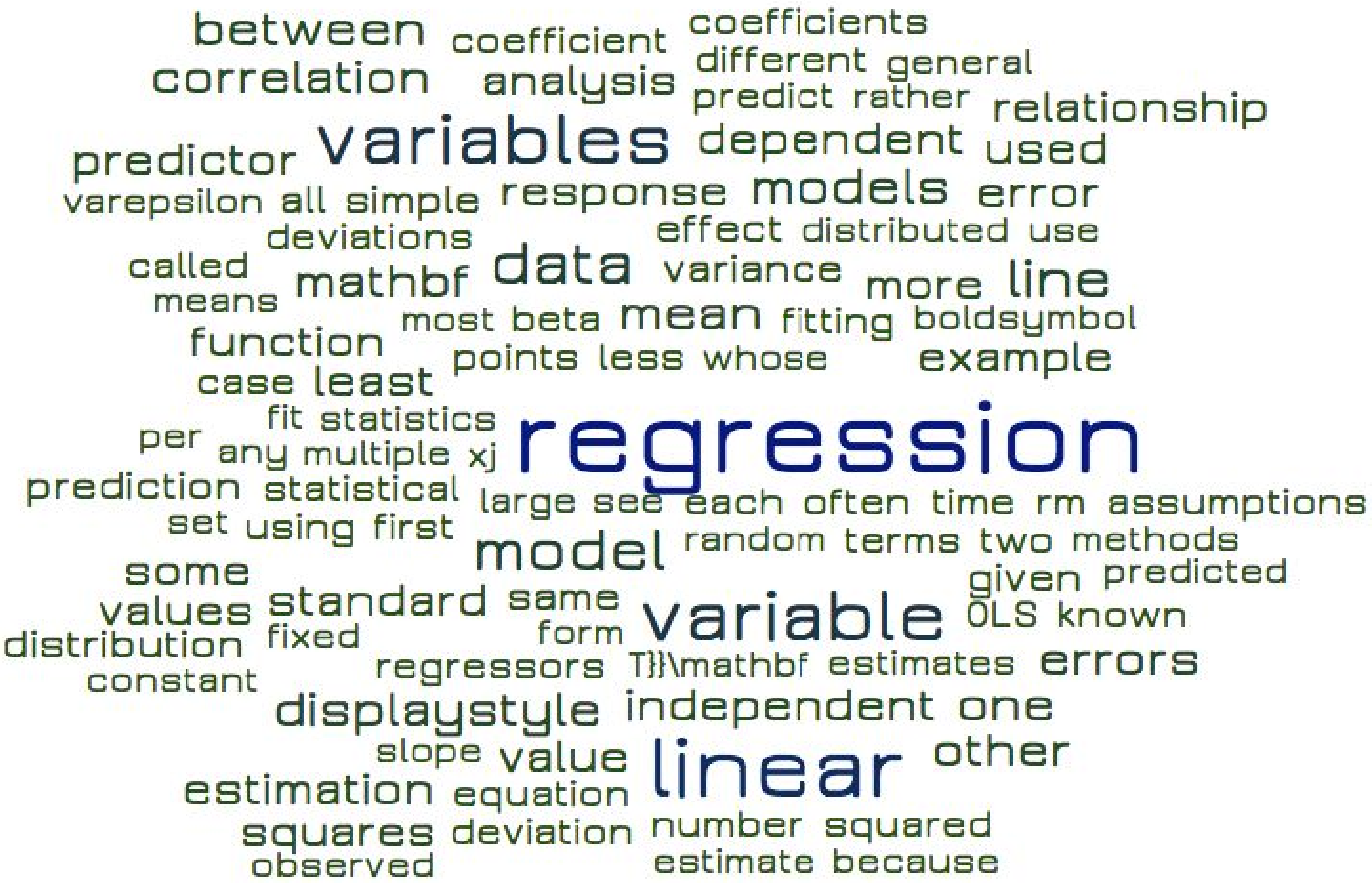
November 23, 2016

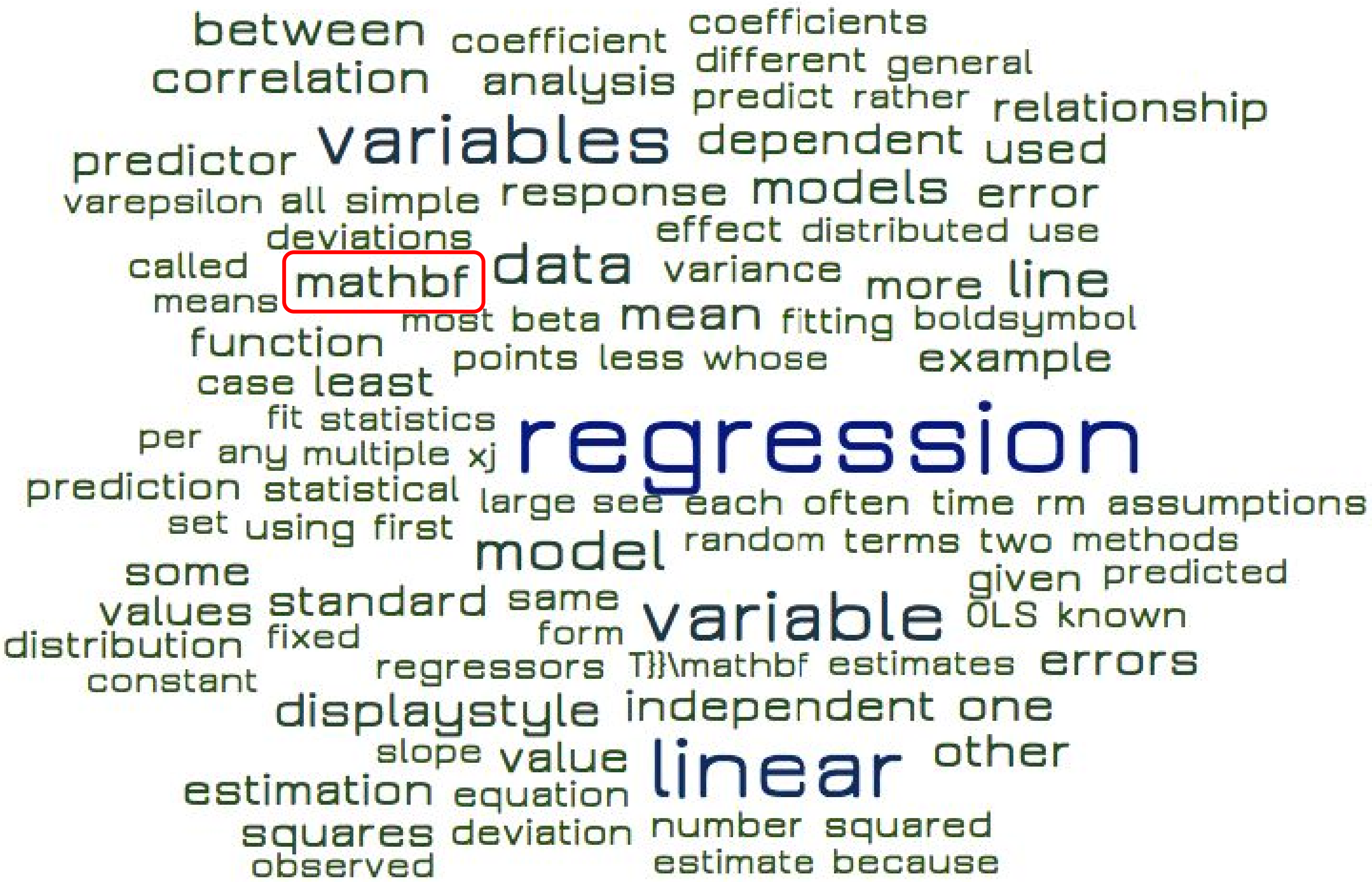
What is linear regression?



What is linear regression?

1) https://en.wikipedia.org/wiki/Linear_regression
2) http://www.statisticssolutions.com/what-is-linear-regression/
3) http://onlinestatbook.com/2/regression/intro.html
4) http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm
5) http://people.duke.edu/~rnau/regintro.htm


between coefficient coefficients
different general
correlation analysis predict rather relationship
variables dependent used
predictor
varepsilon all simple response models error
deviations effect distributed use
called data variance
mathbf more line
means
most beta mean fitting boldsymbol
function
points less whose example
case least
fit statistics
per regression
any multiple xj
prediction statistical
large see each often time rm assumptions
set using first
model random terms two methods
some given predicted
standard same
values variable OLS known
fixed form
distribution
regressors T}}\mathbf estimates errors
constant
displaystyle independent one
slope value linear other
estimation equation
squares deviation number squared
observed estimate because



between coefficient coefficients
different general
correlation analysis
predict rather
relationship
predictor variables dependent used
varepsilon all simple response models error
deviations effect distributed use
called data variance
mathbf more line
means
most beta mean fitting boldsymbol
function
points less whose example
case least
fit statistics
per regression
any multiple xj
prediction statistical
large see each often time rm assumptions
set using first
model random terms two methods
some given predicted
standard same
values variable OLS known
distribution fixed form
regressors T}}\mathbf estimates errors
constant
displaystyle independent one
slope value linear other
estimation equation
squares deviation number squared
observed estimate because

Linear Regression

~~Linear~~ Regression

The term "regression" was coined by Francis Galton in the nineteenth century to describe a biological phenomenon: the heights of descendants of tall ancestors tend to regress down towards a normal average (a phenomenon also known as “regression towards the mean”).

In statistical modeling, regression analysis is a process for estimating the correlation relationships among variables where a **target** (aka ‘dependent’) variable is determined by one or more **predictor** (aka ‘independent’) variables.

This relationship is encapsulated in the “beta” for your predictor, aka:

The coefficient

The amount your predictor value is multiplied by to get your target value.

More specifically, regression analysis explains how the average value of the target variable changes when any individual predictor variable is varied as the other predictor variables are held fixed.

LOSS or COST function

LOSS function

In mathematical optimization, statistics, decision theory and machine learning, a loss function or cost function is a function that maps an event or values of one or more variables onto a real number intuitively representing some "cost" associated with the event.

More specifically, regression analysis explains how the average value of the target variable changes when any individual predictor variable is varied as the other predictor variables are held fixed.

Regression attempts to minimize the loss function.

1. Estimate the relationships between predictor and target variables

1. Estimate the relationships between predictor and target variables
2. Incurs loss as a result of this estimation

1. Estimate the relationships between predictor and target variables
2. Incurs loss as a result of this estimation
3. Attempts to minimize loss

~~Linear~~ Regression

Linear Regression

What is linear regression?

1635–45; < Latin ***līneāris* –** of, belonging to lines

What is linear regression?

1635-45; < Latin ***līneāris* -** of, belonging to lines

- Linear regression attempts to **model** the **relationship** between two variables by fitting a linear equation to observed data.
- Is a **statistical** method that allows us to summarize relationships between two **continuous** (quantitative) variables.
- Is the most widely used of all statistical techniques, which studies linear, **additive** relationships between variables.
- A **mathematical** technique for finding the **straight line** that **best-fits** the values. This line can be used for estimating the **future** values of the function by extending it while maintaining its **slope**.

What is linear regression?

1635–45; < Latin ***līneāris* –** of, belonging to lines

- The **statistical** and **mathematical** process by which we **model** the **slope** of the **best-fit**, **additive relationship**, represented by a plotted **straight line**, between two sets of **continuous** variables where one set determines or predicts the other, in order to estimate **future** values.

What is linear regression?

1635–45; < Latin ***līneāris* – of, belonging to lines**

- The **statistical** and **mathematical** process by which we **model** the **slope** of the **best-fit**, **additive relationship**, represented by a plotted **straight line**, between two sets of **continuous** variables where one set determines or predicts the other, in order to estimate **future** values.

What is linear regression?

1635-45; < Latin ***līneāris* – of**, belonging to lines

- **Simple** linear regression:

  1 predictor variable

What is linear regression?

1635–45; < Latin ***līneāris* –** of, belonging to lines

- **Simple** linear regression:

    1 predictor variable

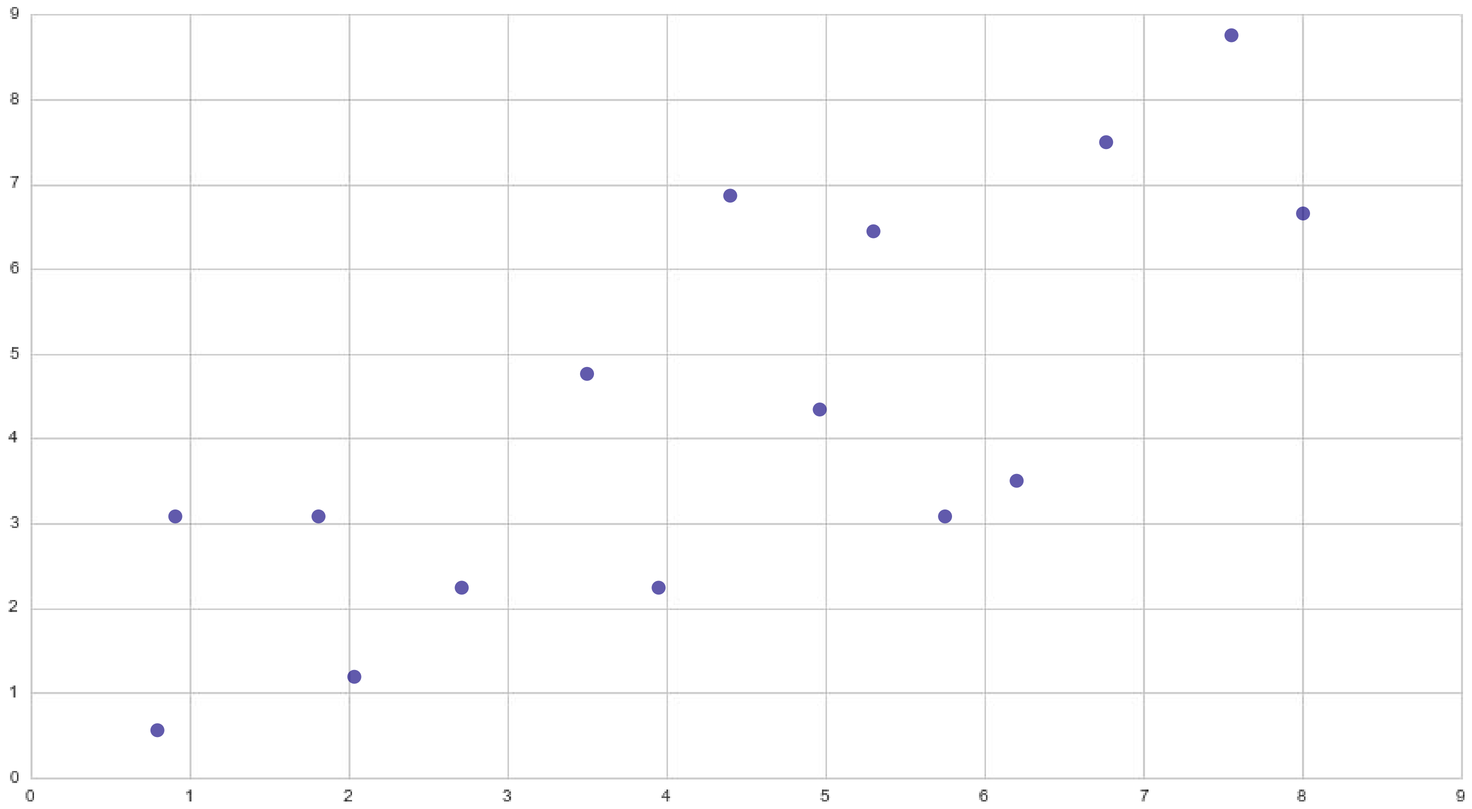
- **Multiple** linear regression:

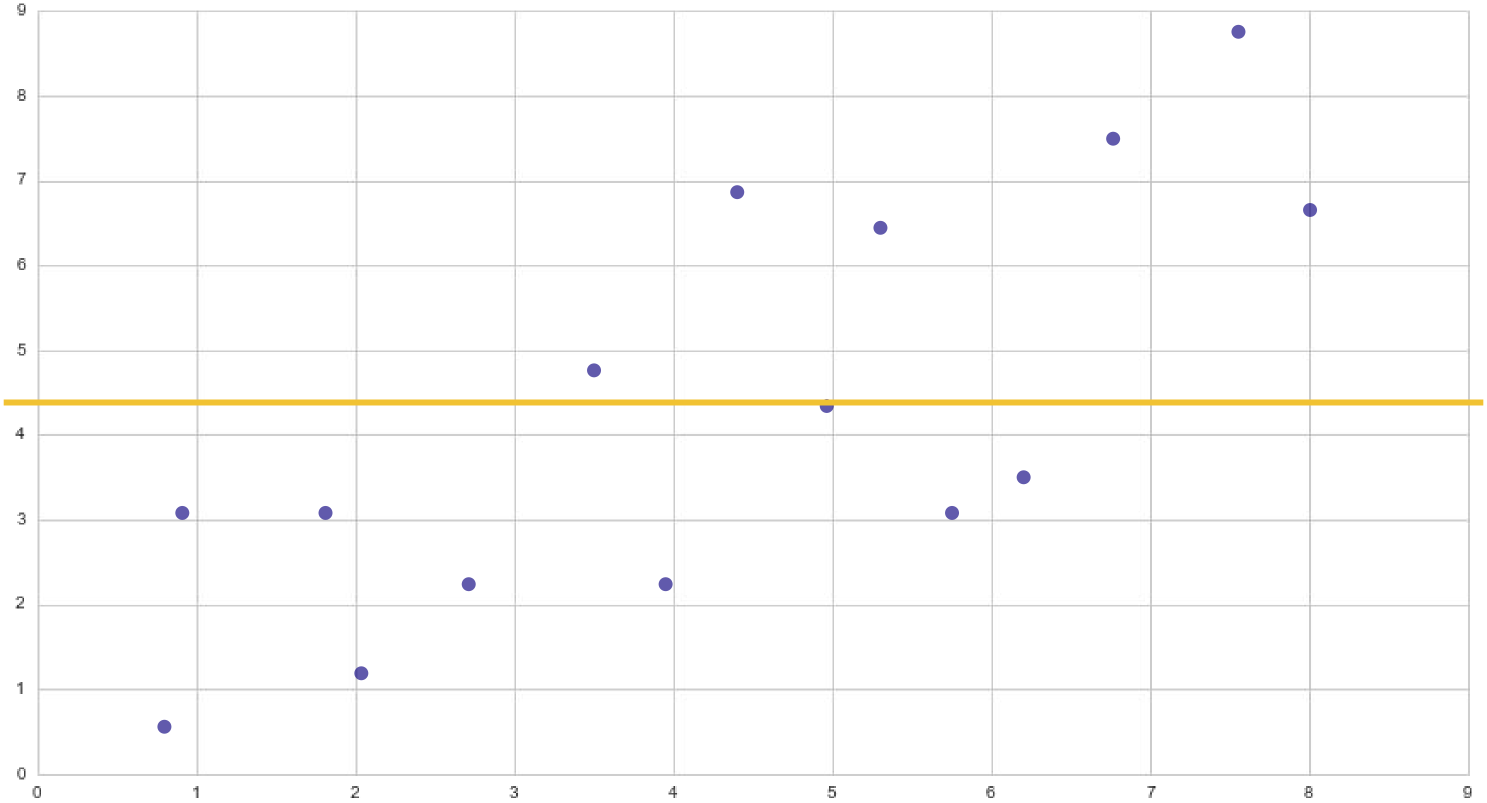
    >1 predictor variable

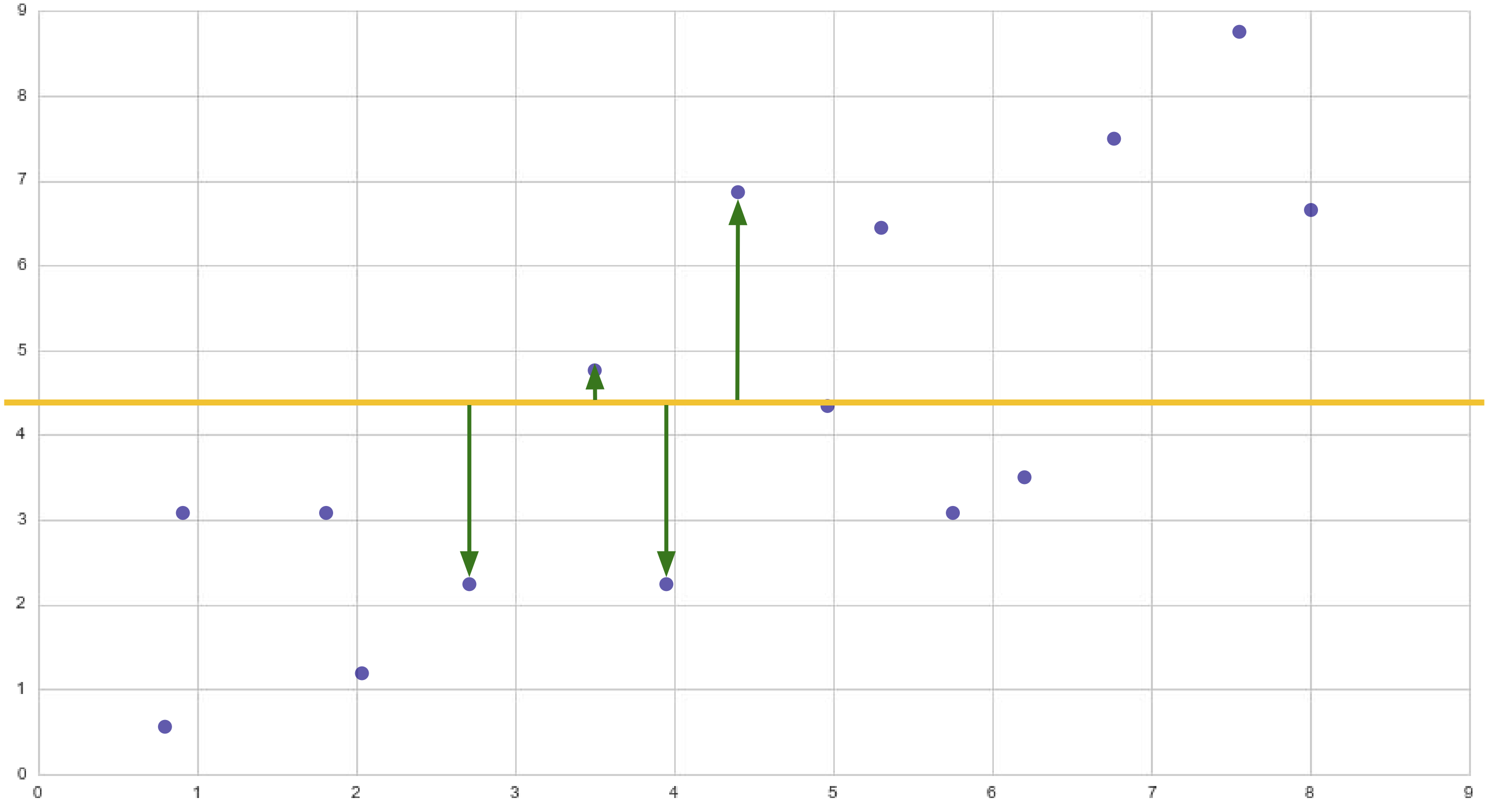
Ordinary Least Squares

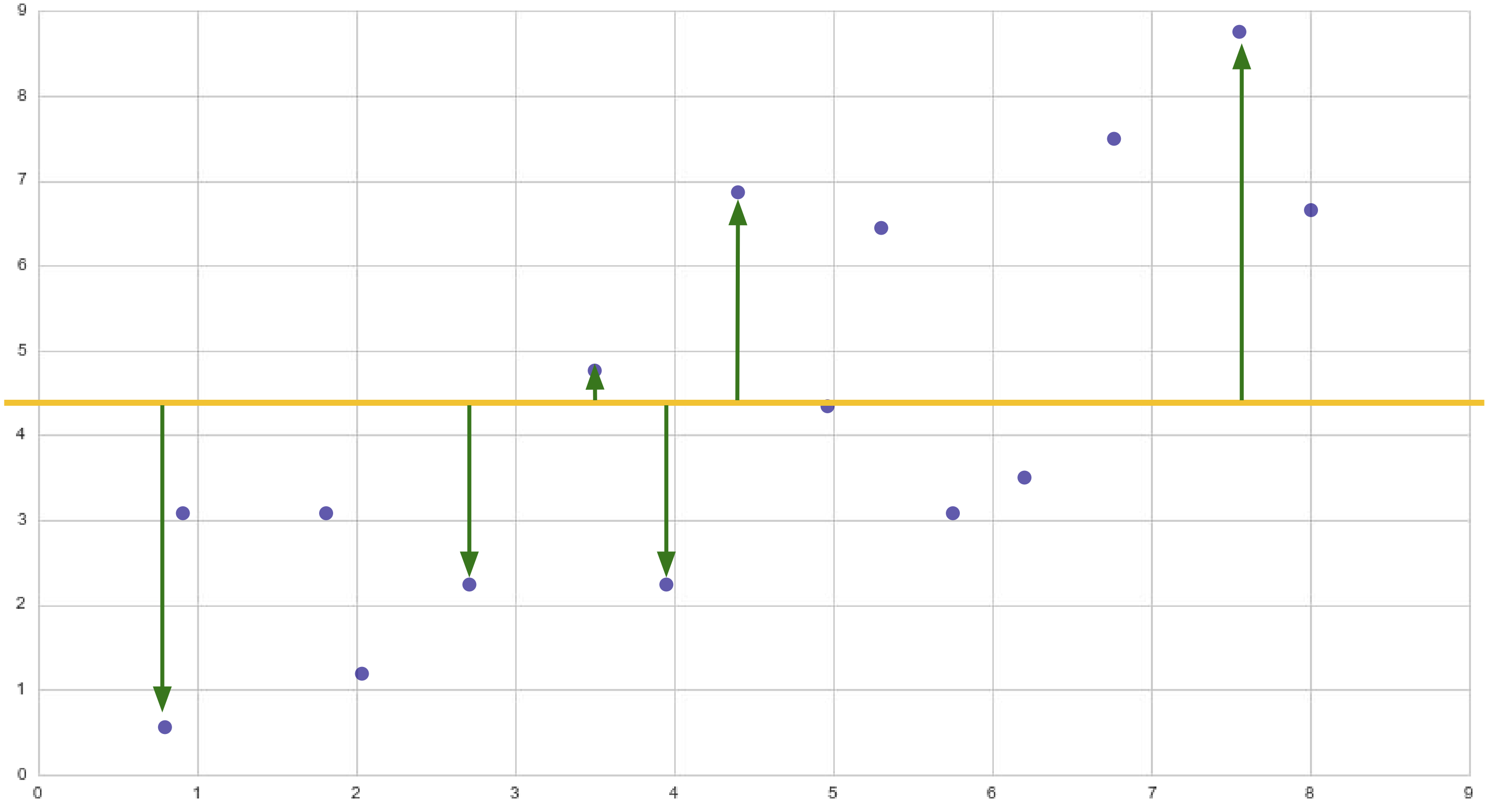
The earliest form of regression was the *method of least squares*, which was published by Adrien Legendre in 1805 as he applied the method to the problem of determining, from astronomical observations, the orbits of bodies about the Sun.

The specific form used for linear regression is “**Ordinary** Least Squares,” which comes from the original French term, "*méthode des moindres carrés*".

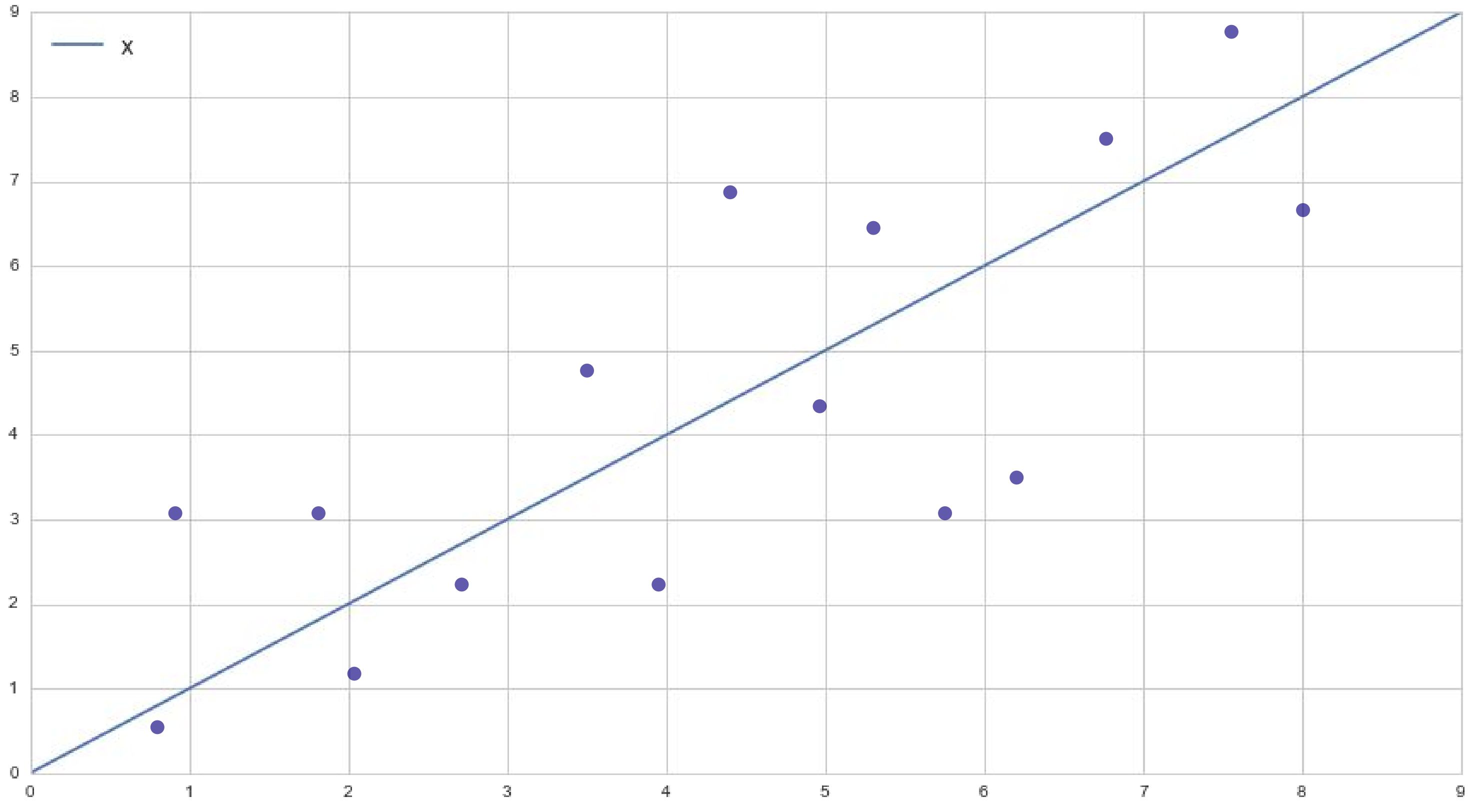

9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5
6
7
8
9

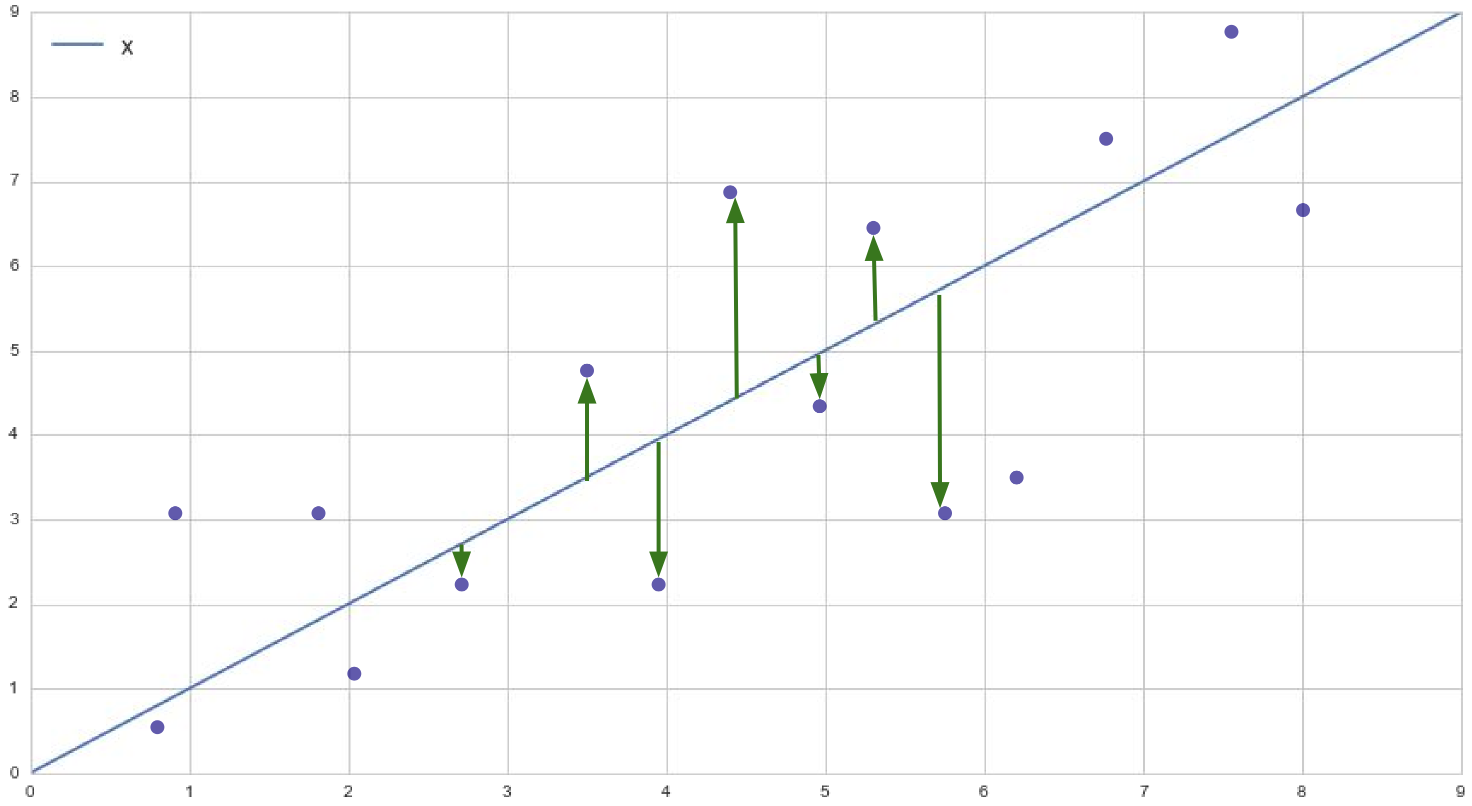

0
1
2
3
4
5
6
7
8
9
0
1
2
3
4
5
6
7
8
9


0
1
2
3
4
5
6
7
8
9
0
1
2
3
4
5
6
7
8
9

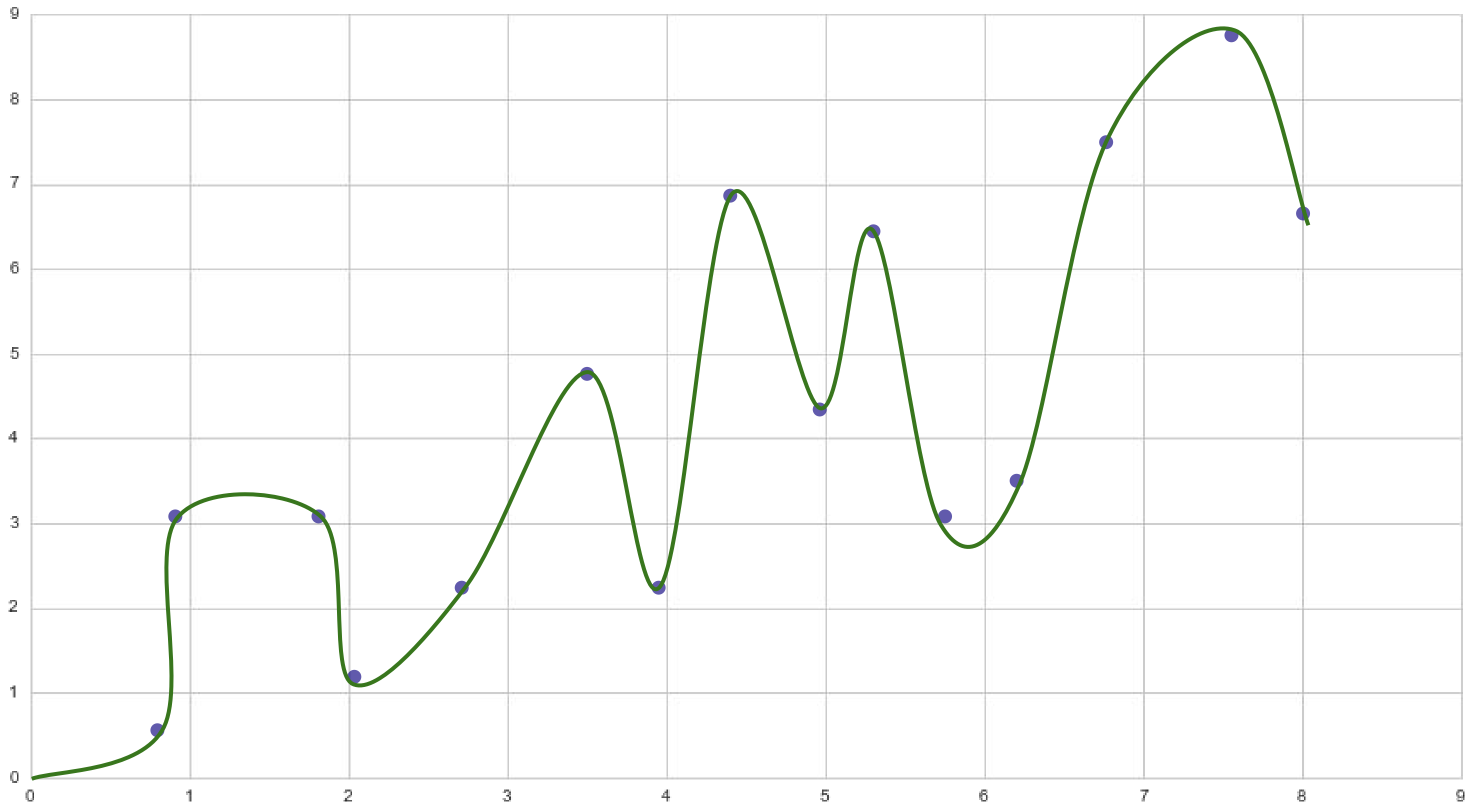


0
1
2
3
4
5
6
7
8
9

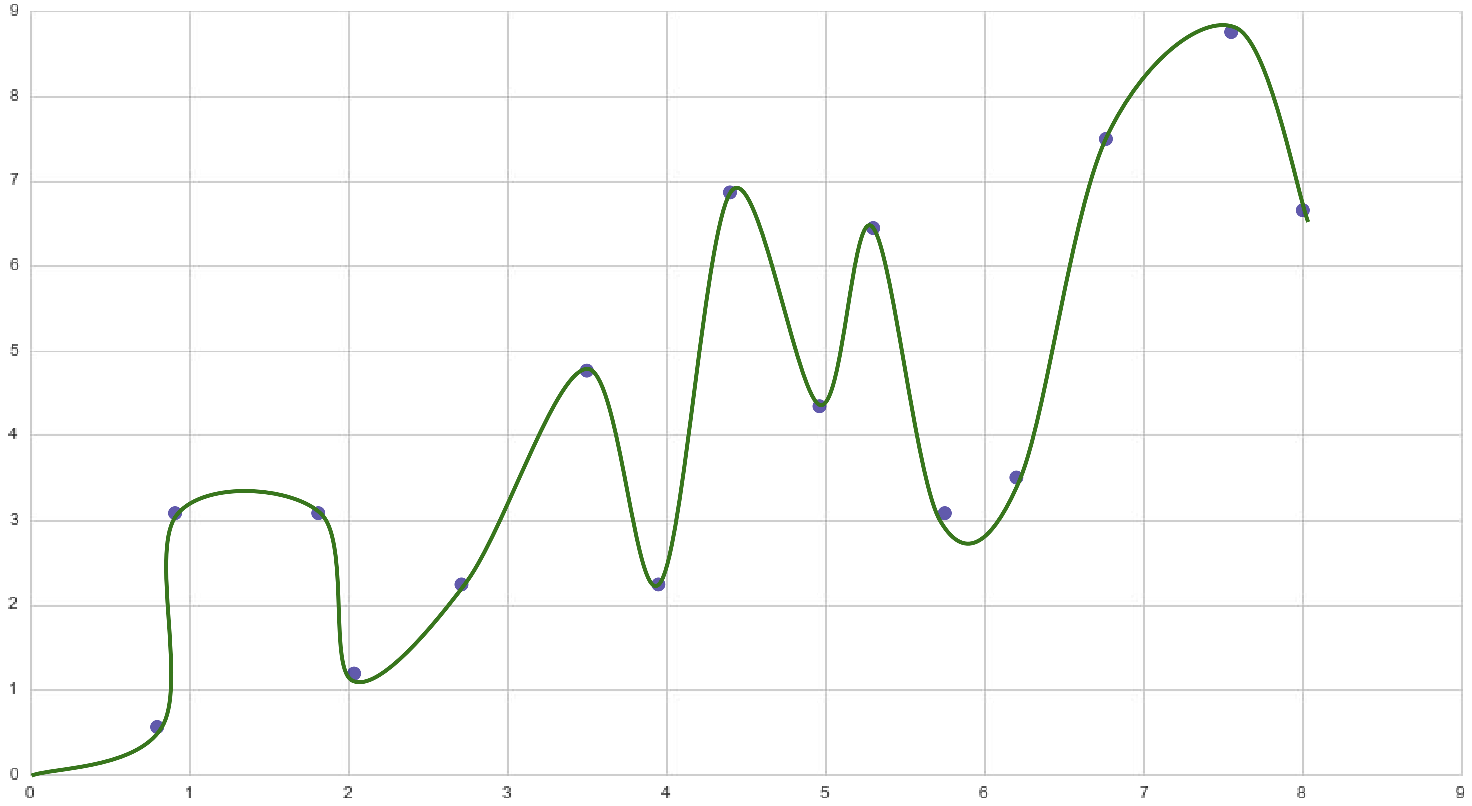

X
0
1
2
3
4
5
6
7
8
9



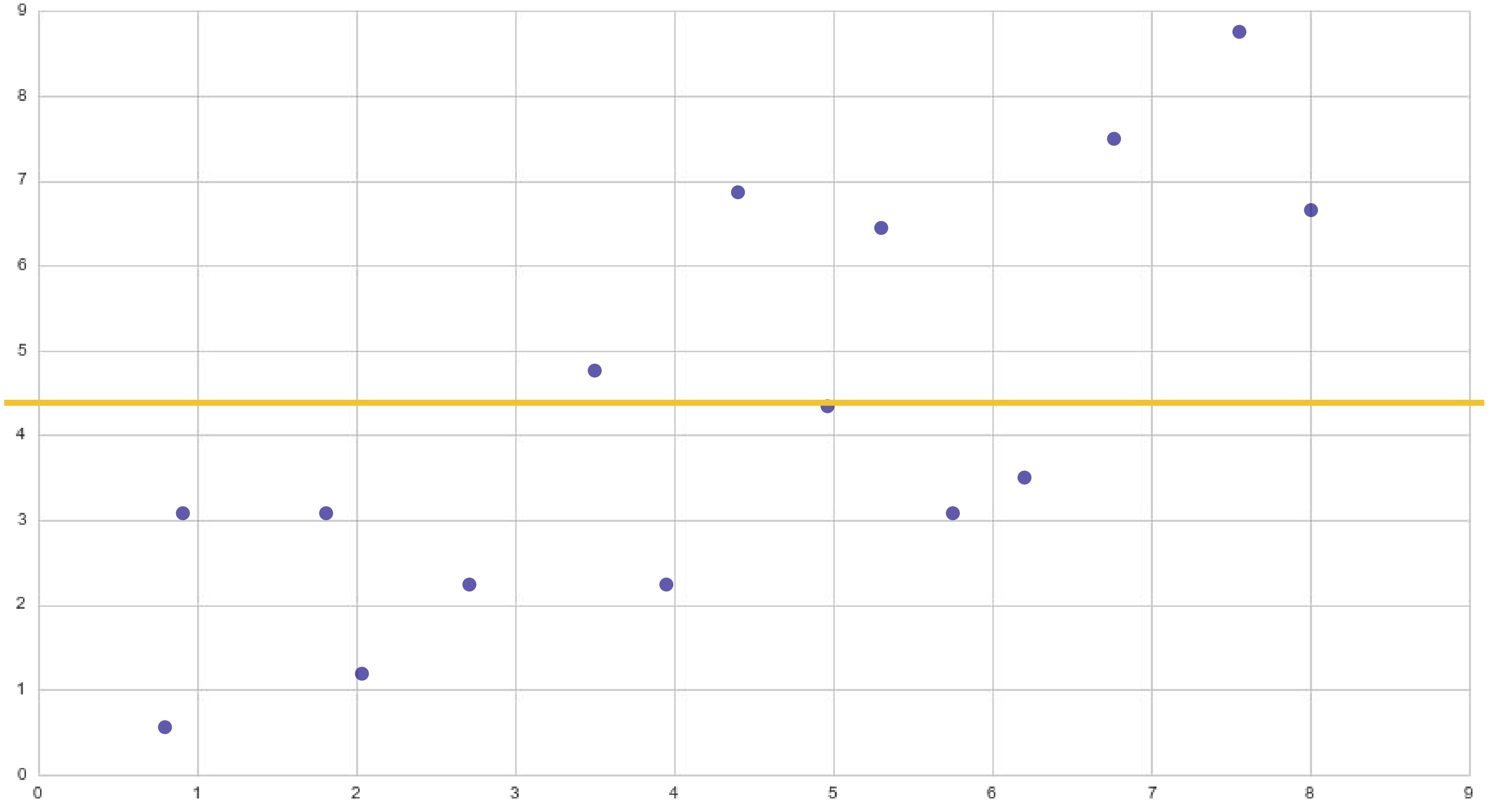
X
0
1
2
3
4
5
6
7
8
9


0
1
2
3
4
5
6
7
8
9
0
1
2
3
4
5
6
7
8
9

Overfitting



Underfitting



Bias and Variance

Bias

- The bias (or bias function) of an estimator is the difference between the estimator's expected value and the true value of the parameter being estimated.

Bias

- The bias (or bias function) of an estimator is the difference between the estimator's expected value and the true value of the parameter being estimated.
- Bias is caused by the simplifying assumptions made by a model to make the target function easier to learn

Bias

- The bias (or bias function) of an estimator is the difference between the estimator's expected value and the true value of the parameter being estimated.
- Bias is caused by the simplifying assumptions made by a model to make the target function easier to learn
- Quicker and easier to model, at the potential expense of accuracy

Variance

- Variance is error from sensitivity to small fluctuations in the training set.

Variance

- Variance is error from sensitivity to small fluctuations in the training set.
- Can seductively resemble more “accuracy,” as it may seem to better fit your current data, but does so at the expense of being less accurate on new data

Variance

- Variance is error from sensitivity to small fluctuations in the training set.
- Can seductively resemble more “accuracy,” as it may seem to better fit your current data, but does so at the expense of being less accurate on new data
- More flexibility with form, less assumptions made about the data

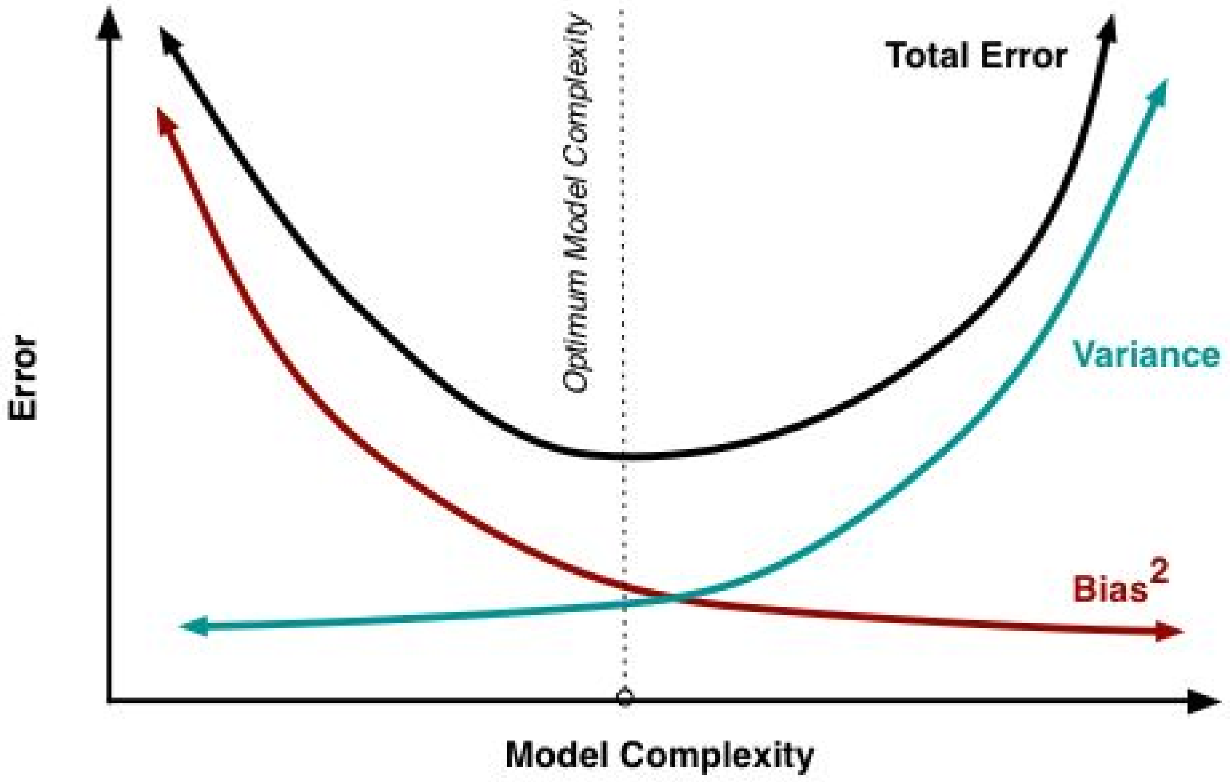
Bias vs Variance

- Tend to be inverses of each other - as one increases, the other decreases.

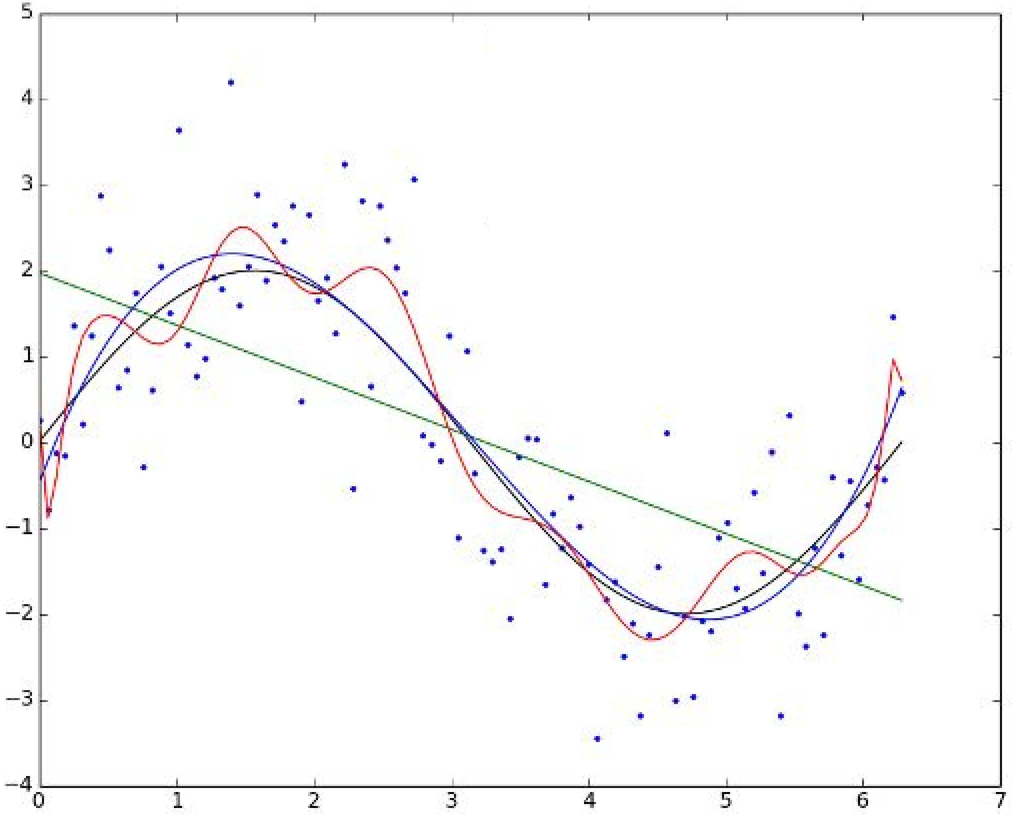
Bias vs Variance

- Tend to be inverses of each other - as one increases, the other decreases.
- The trick is balance where they are both as minimized as possible.

Bias vs Variance



Bias vs Variance



Math

$$\beta_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\beta_1 = \frac{\boxed{\sum_{i=1}^{n}}(y_i - \bar{y})(x_i - \bar{x})}{\boxed{\sum_{i=1}^{n}}(x_i - \bar{x})^2}$$

$$\beta_1 = \frac{\text{Sum} \quad (y_i - \bar{y})(x_i - \bar{x})}{\text{Sum} \quad (x_i - \bar{x})^2}$$

$$\beta_1 = \frac{\text{Sum} \quad \boxed{(y_i - \bar{y})(x_i - \bar{x})}}{\text{Sum} \quad (x_i - \bar{x})^2}$$

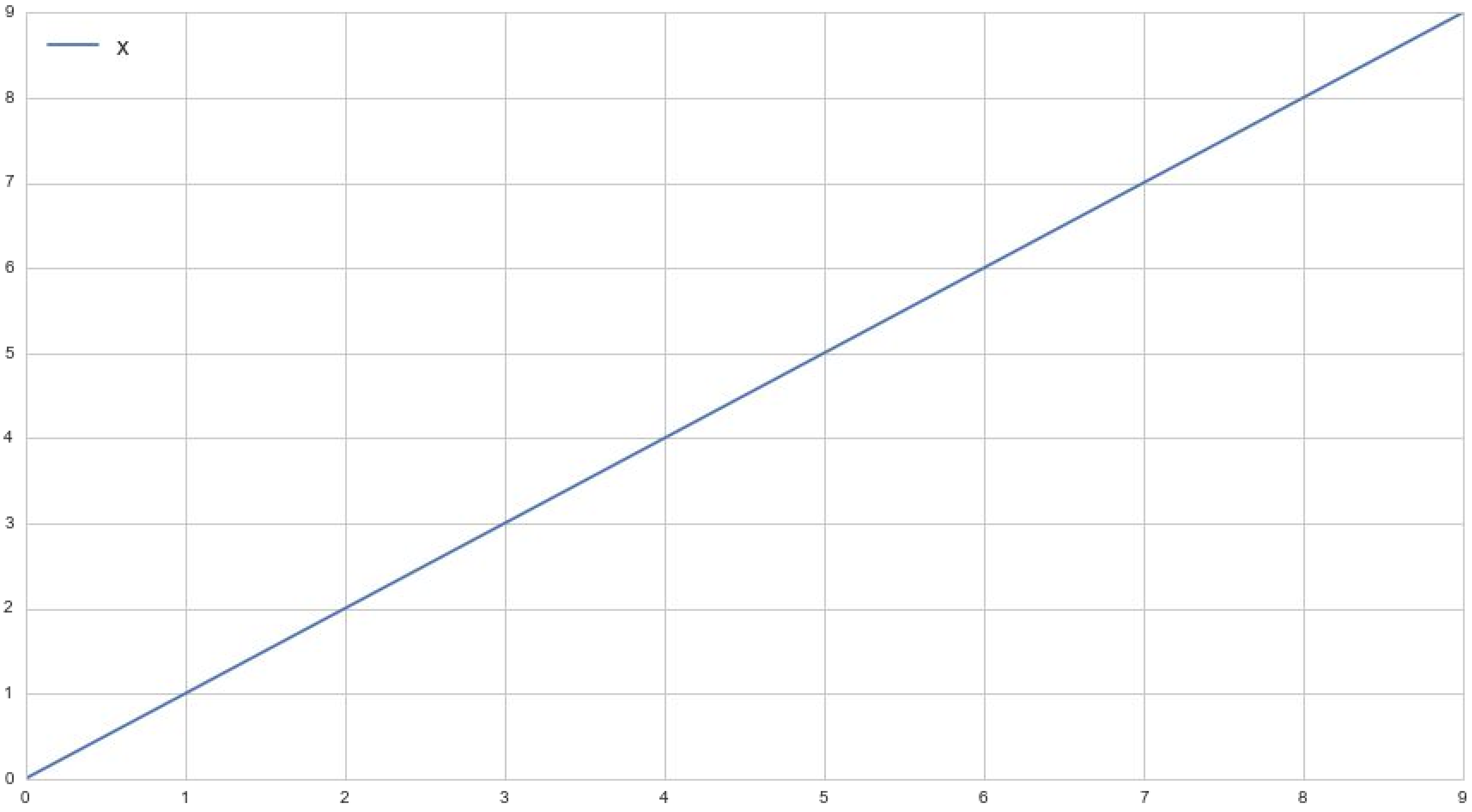
$$\beta_1 = \frac{\text{Sum } (y - \text{mean of } y)(x - \text{mean of } x)}{\text{Sum } (x_i - \bar{x})^2}$$

$$\beta_1 = \frac{\text{Sum} \quad (y - \text{mean of } y)(x - \text{mean of } x)}{\text{Sum} \quad (x - \text{mean of } x)^2}$$

Simple Linear Regression

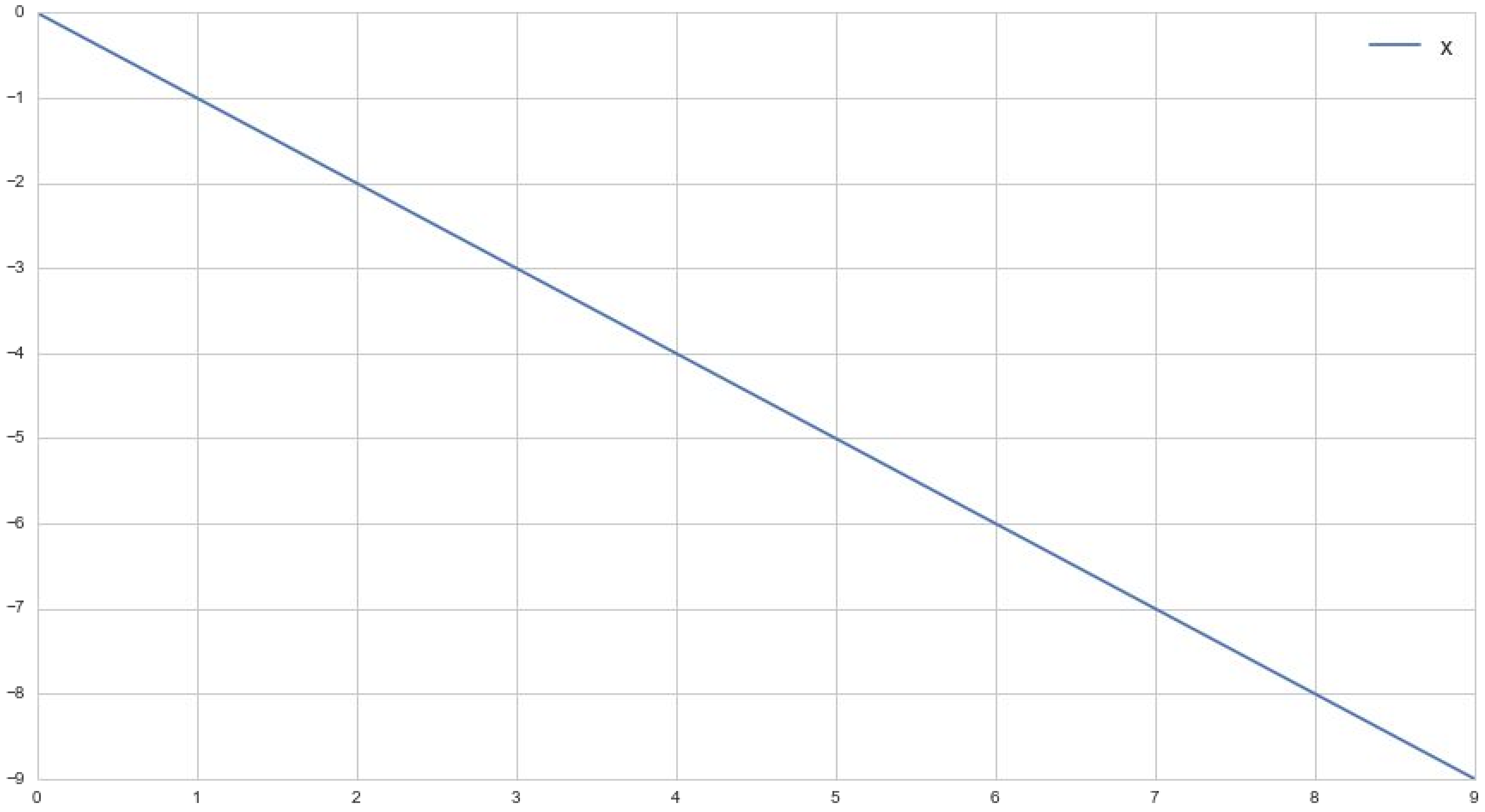
$$y = x$$

	x	y
0	0	0
1	1	1
2	2	2
3	3	3
4	4	4


x
0
1
2
3
4
5
6
7
8
9

$$y = -x$$

	x	y
0	0	0
1	-1	1
2	-2	2
3	-3	3
4	-4	4


0
−1
−2
−3
−4
−5
−6
−7
−8
−9
0
1
2
3
4
5
6
7
8
9
x

Deterministic (or functional) relationships

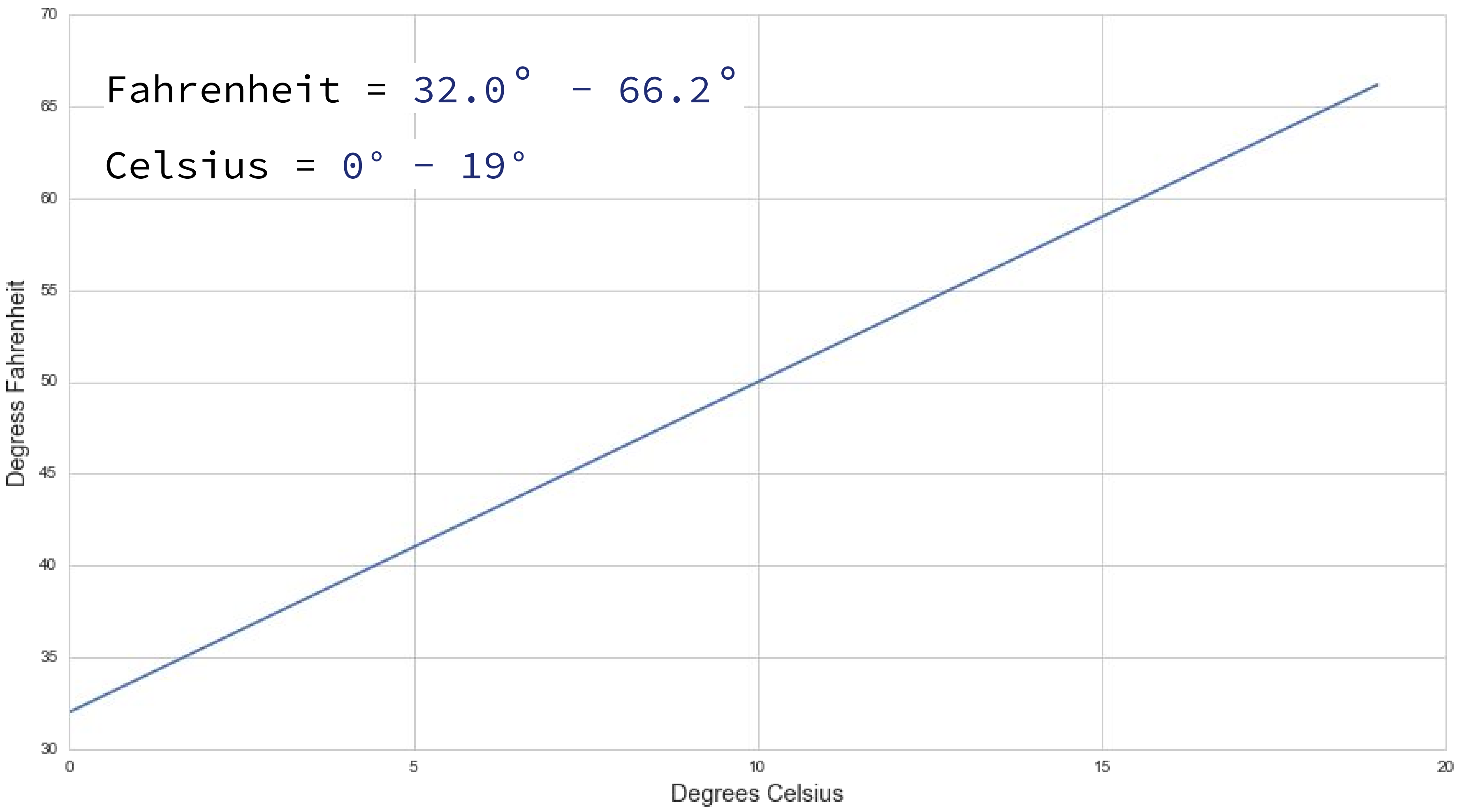
Deterministic (or functional) relationships

are fixed and predict their data *exactly*.

$$F = (9/5)C + 32$$

Converting Celsius to Fahrenheit

$$°F = (9/5)°C + 32$$



Fahrenheit = 32.0° - 66.2°
Celsius = 0° - 19°
70
65
60
55
50
45
40
35
30
Degress Fahrenheit
0
5
10
15
20
Degrees Celsius

Converting Celsius to Fahrenheit

$$°F = (9/5)°C + 32$$

Converting Celsius to Fahrenheit

$$°F = (9/5)°C + 32$$

$$y = mx + b$$

$$y = mx + b$$

y = target variable

$$y = mx + b$$

y = target variable

x = predictor variable

$$y = mx + b$$

y = target variable

x = predictor variable

m = coefficient

$$y = mx + b$$

y = target variable

x = predictor variable

m = coefficient

b = y intercept

$$y = mx + b + \in$$

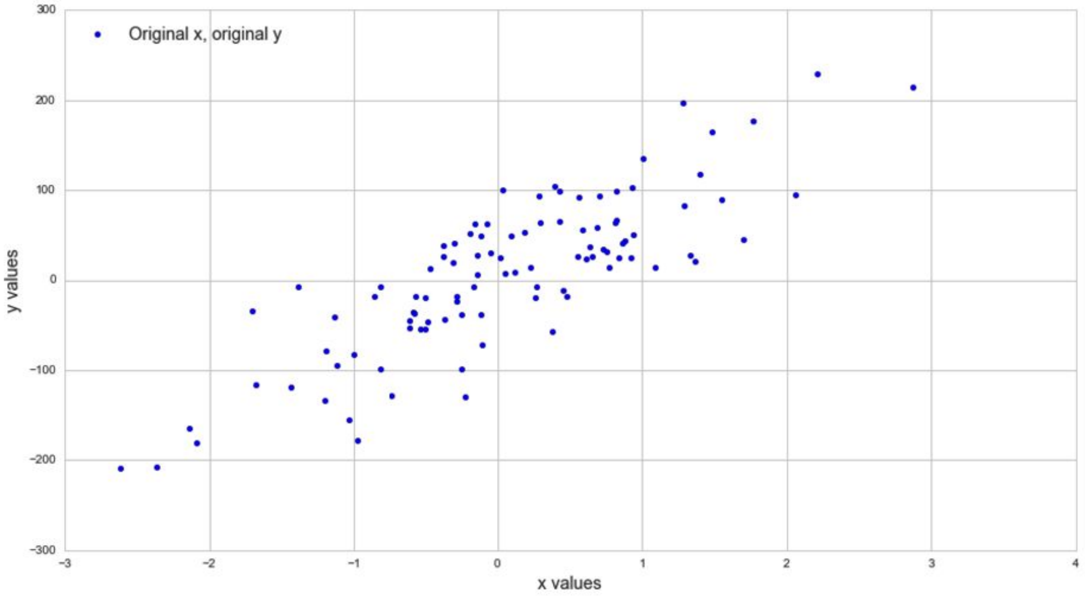
	x	y
0	-2.613661	-208.496884
1	-2.367838	-207.317293
2	-2.139704	-180.131111
3	-2.088335	-177.826539
4	-1.704292	-164.514399
5	-1.678502	-154.957639
6	-1.435008	-133.865052
7	-1.389613	-129.010627
8	-1.198706	-128.643283
9	-1.191934	-119.141774

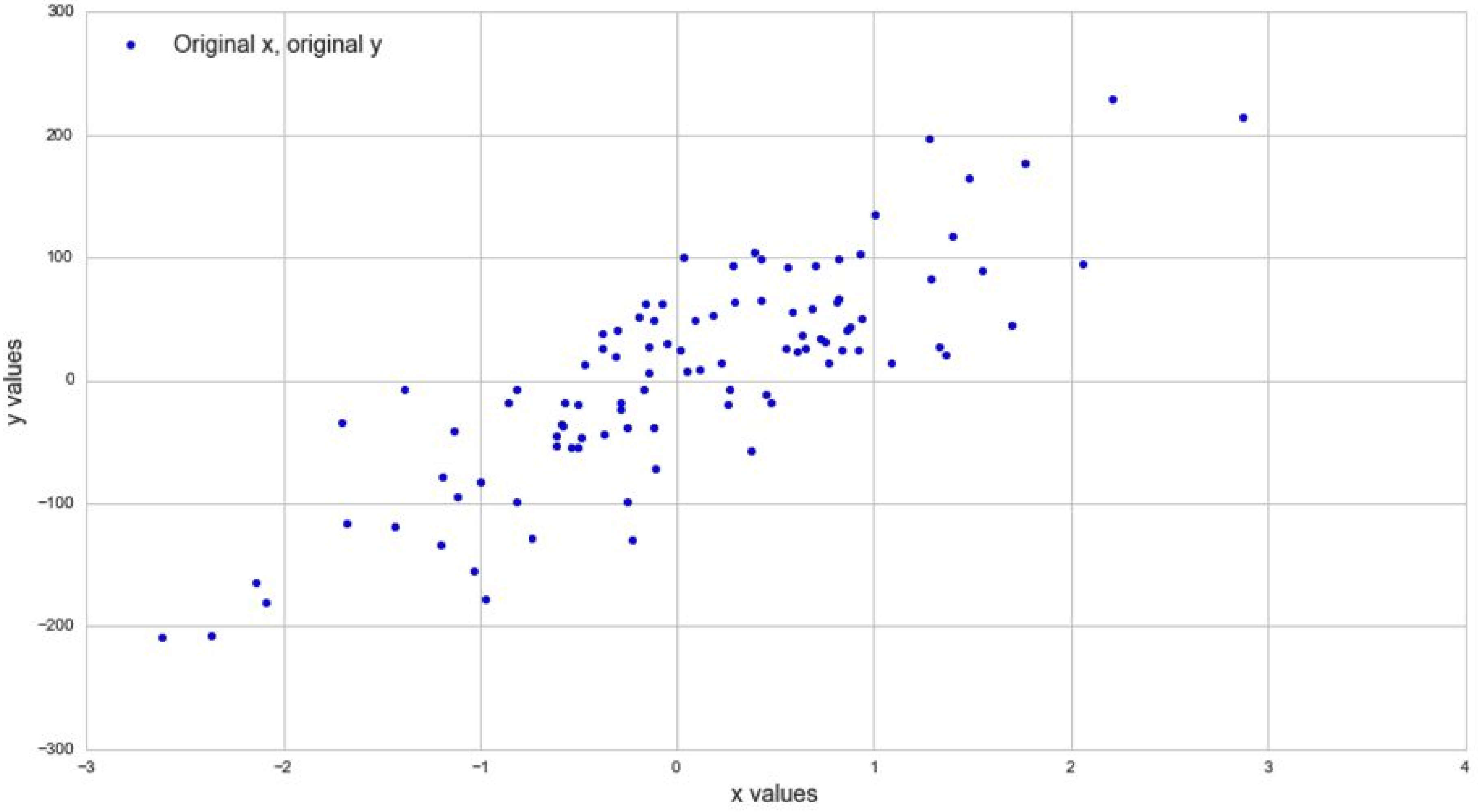
“Observed data”

(functionally generated with added random noise)

	x	y
0	-2.613661	-208.496884
1	-2.367838	-207.317293
2	-2.139704	-180.131111
3	-2.088335	-177.826539
4	-1.704292	-164.514399
5	-1.678502	-154.957639
6	-1.435008	-133.865052
7	-1.389613	-129.010627
8	-1.198706	-128.643283
9	-1.191934	-119.141774

“Observed data”
(functionally generated with added random noise)




Original x, original y
300
200
100
0
−100
−200
−300
−3
−2
−1
0
1
2
3
4
x values
y values

	x	y
0	-2.613661	-208.496884
1	-2.367838	-207.317293
2	-2.139704	-180.131111
3	-2.088335	-177.826539
4	-1.704292	-164.514399
5	-1.678502	-154.957639
6	-1.435008	-133.865052
7	-1.389613	-129.010627
8	-1.198706	-128.643283
9	-1.191934	-119.141774

$$y = mx + b$$

	x	y
0	-2.613661	-208.496884
1	-2.367838	-207.317293
2	-2.139704	-180.131111
3	-2.088335	-177.826539
4	-1.704292	-164.514399
5	-1.678502	-154.957639
6	-1.435008	-133.865052
7	-1.389613	-129.010627
8	-1.198706	-128.643283
9	-1.191934	-119.141774

$$-208.496884 = mx + b$$

$$y = mx + b$$

	x	y
0	$\boxed{-2.613661}$	-208.496884
1	-2.367838	-207.317293
2	-2.139704	-180.131111
3	-2.088335	-177.826539
4	-1.704292	-164.514399
5	-1.678502	-154.957639
6	-1.435008	-133.865052
7	-1.389613	-129.010627
8	-1.198706	-128.643283
9	-1.191934	-119.141774

$$-208.496884 = m(-2.613661) + b$$

$$y = mx + b$$

```python
from sklearn.linear_model import LinearRegression

lr = LinearRegression()

lr.fit(x,y)
```

```
from sklearn.linear_model import LinearRegression

lr = LinearRegression()

lr.fit(x,y)

print lr.coef_
print lr.intercept_
```

```
[72.72244833]
3.07326447646
```

$$\hat{y} = 72.73(x) + 3.072$$

$$y = mx + b$$

	x	y	predicted y
0	-2.613661	-208.496884	-186.998546
1	-2.367838	-207.317293	-169.121683
2	-2.139704	-180.131111	-152.531234
3	-2.088335	-177.826539	-148.795591
4	-1.704292	-164.514399	-120.867000
5	-1.678502	-154.957639	-118.991513
6	-1.435008	-133.865052	-101.284041
7	-1.389613	-129.010627	-97.982828
8	-1.198706	-128.643283	-84.099591
9	-1.191934	-119.141774	-83.607106

$$\hat{y} = 72.73(x) + 3.072$$

$$y = mx + b$$

	x	y	predicted y
0	-2.613661	-208.496884	-186.998546
1	-2.367838	-207.317293	-169.121683
2	-2.139704	-180.131111	-152.531234
3	-2.088335	-177.826539	-148.795591
4	-1.704292	-164.514399	-120.867000
5	-1.678502	-154.957639	-118.991513
6	-1.435008	-133.865052	-101.284041
7	-1.389613	-129.010627	-97.982828
8	-1.198706	-128.643283	-84.099591
9	-1.191934	-119.141774	-83.607106

$$\hat{y} = 72.73(x) + 3.072$$

$$y = mx + b$$

	x	y	predicted y
0	-2.613661	-208.496884	-186.998546
1	-2.367838	-207.317293	-169.121683
2	-2.139704	-180.131111	-152.531234
3	-2.088335	-177.826539	-148.795591
4	-1.704292	-164.514399	-120.867000
5	-1.678502	-154.957639	-118.991513
6	-1.435008	-133.865052	-101.284041
7	-1.389613	-129.010627	-97.982828
8	-1.198706	-128.643283	-84.099591
9	-1.191934	-119.141774	-83.607106

$$\hat{y} = 72.73(x) + 3.072$$

$$y = mx + b$$

	x	y	predicted y
0	-2.613661	-208.496884	-186.998546
1	-2.367838	-207.317293	-169.121683
2	-2.139704	$\boxed{-180.131111}$	$\boxed{-152.531234}$
3	-2.088335	-177.826539	-148.795591
4	-1.704292	-164.514399	-120.867000
5	-1.678502	-154.957639	-118.991513
6	-1.435008	-133.865052	-101.284041
7	-1.389613	-129.010627	-97.982828
8	-1.198706	-128.643283	-84.099591
9	-1.191934	-119.141774	-83.607106

$$\hat{y} = 72.73(x) + 3.072$$

$$y = mx + b$$

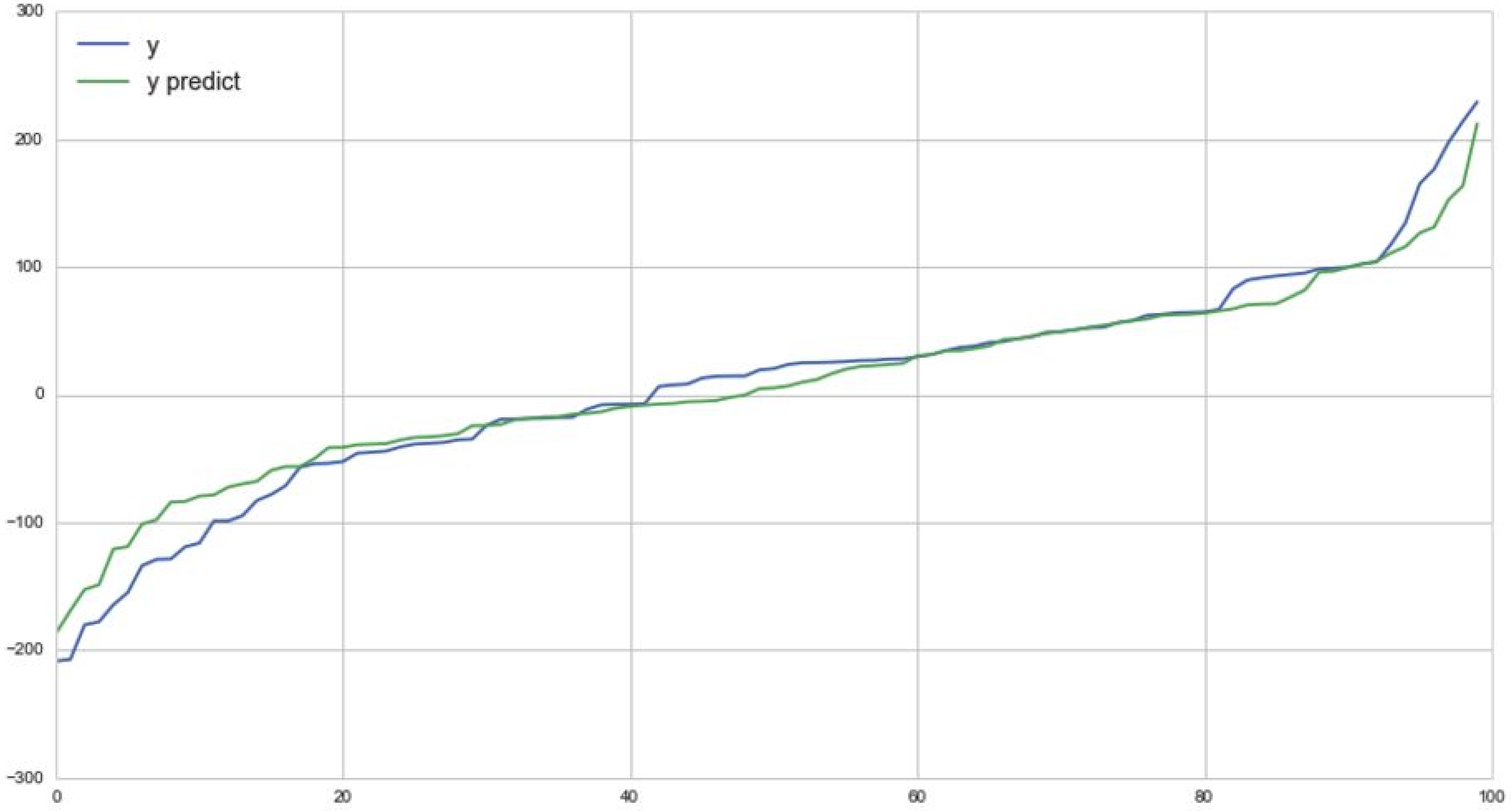
	x	y	predicted y
0	-2.613661	-208.496884	-186.998546
1	-2.367838	-207.317293	-169.121683
2	-2.139704	-180.131111	-152.531234
3	-2.088335	-177.826539	-148.795591
4	-1.704292	-164.514399	-120.867000
5	-1.678502	-154.957639	-118.991513
6	-1.435008	-133.865052	-101.284041
7	-1.389613	-129.010627	-97.982828
8	-1.198706	-128.643283	-84.099591
9	-1.191934	-119.141774	-83.607106

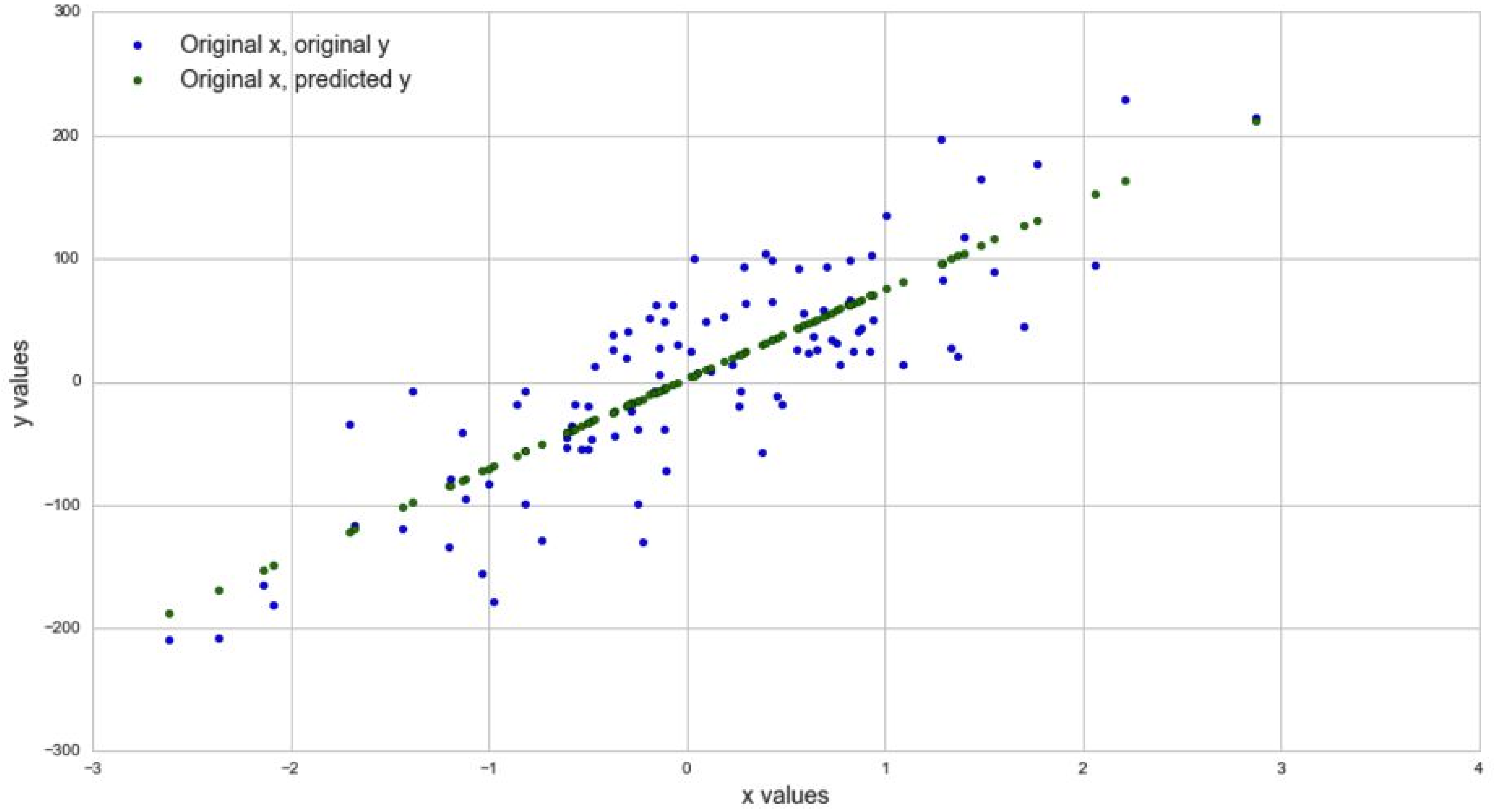
$$\hat{y} = 72.73(x) + 3.072$$

$$y = mx + b$$

$$y = mx + b + \in$$

$$y = mx + b + \boxed{\in}$$


y
y predict
300
200
100
0
−100
−200
−300
0
20
40
60
80
100



Original x, original y
Original x, predicted y
300
200
100
0
−100
−200
−300
y values
−3
−2
−1
0
1
2
3
4
x values

Measuring Accuracy

Accuracy

1. Score

```python
from sklearn.linear_model import LinearRegression

lr = LinearRegression()

lr.fit(x,y)

print lr.coef_
print lr.intercept_
```

```
[72.72244833]
3.07326447646
```

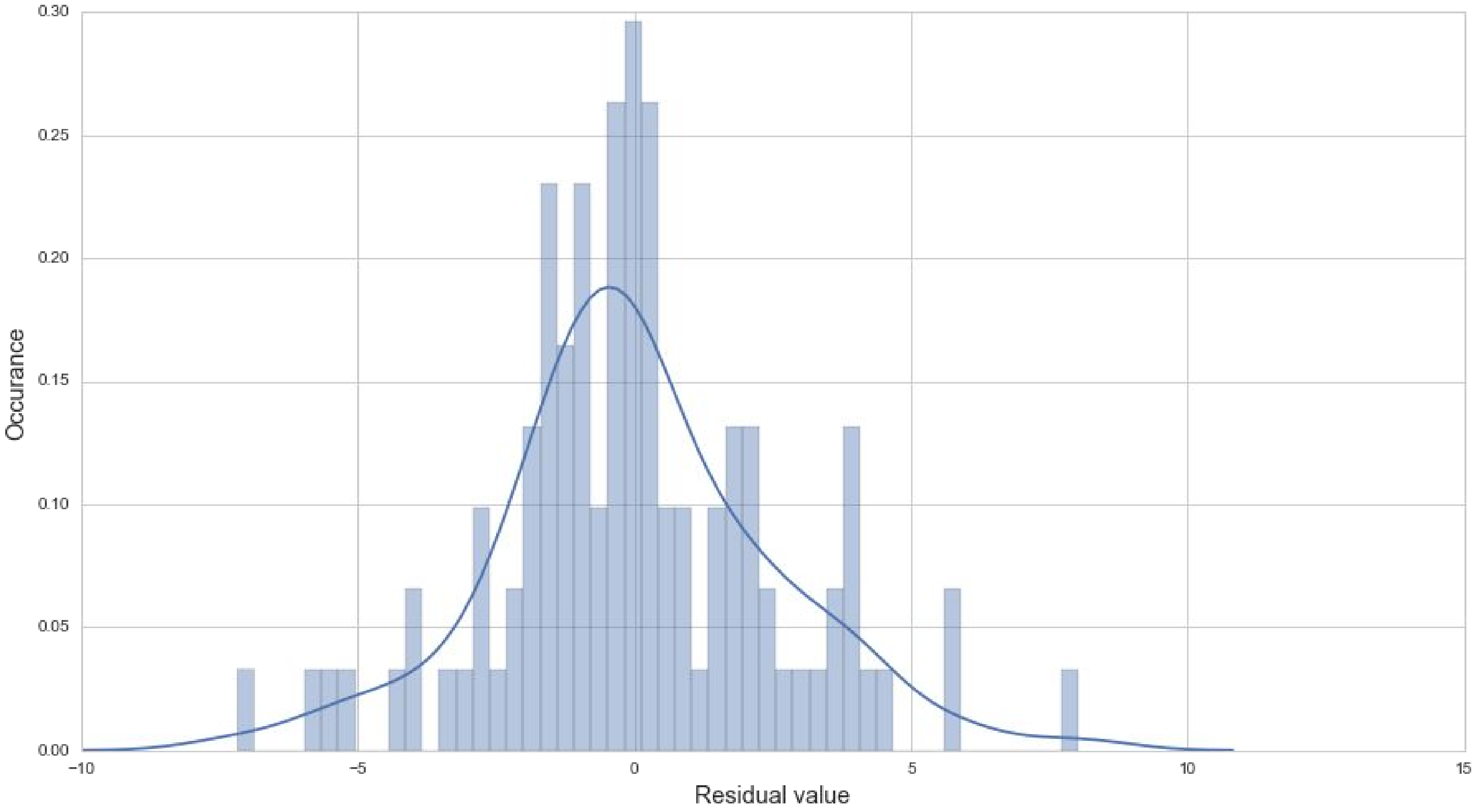
```python
lr.score(x,y)
```

```
0.89193846362
```

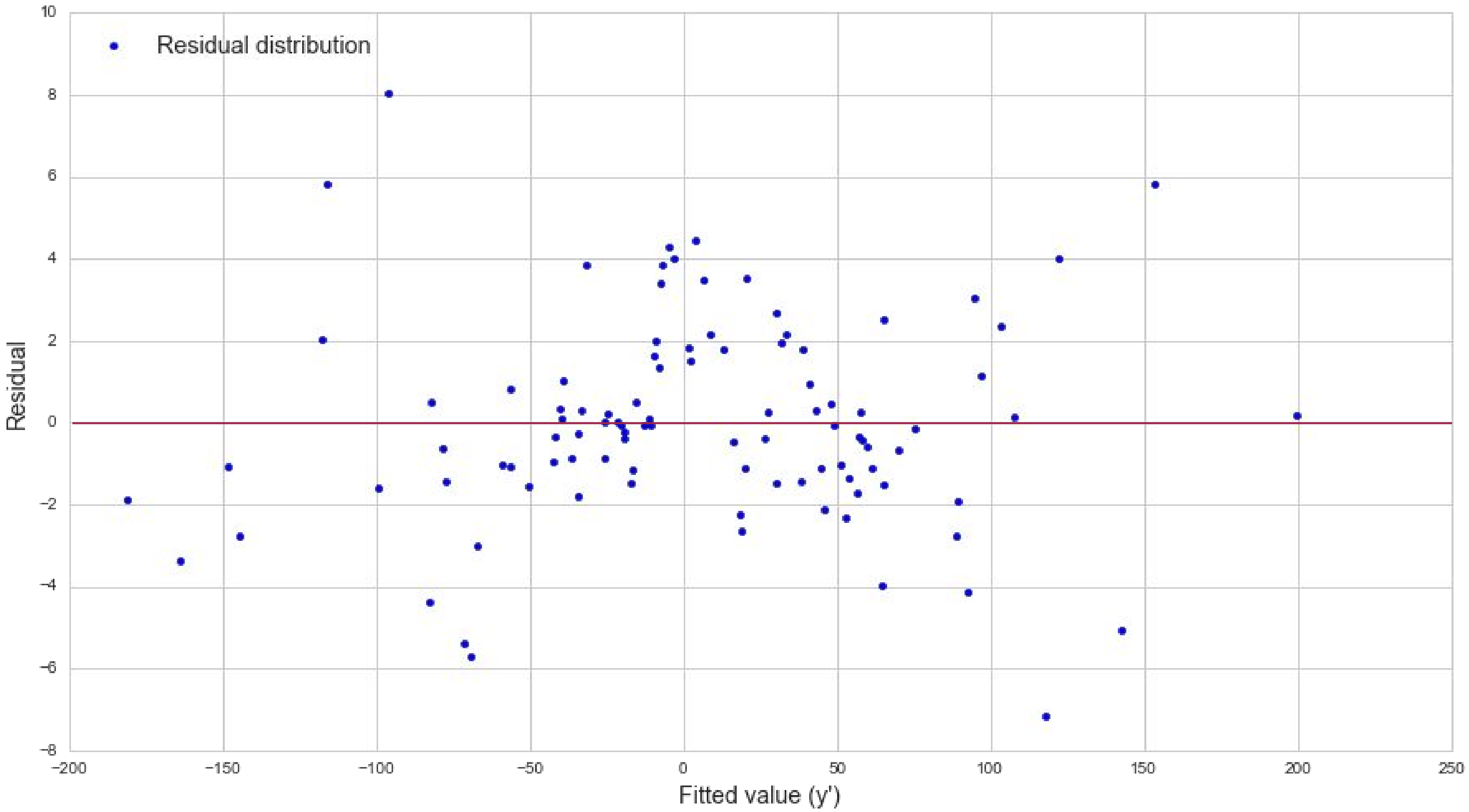
Accuracy

1. Score
2. Plot your residuals

Accuracy

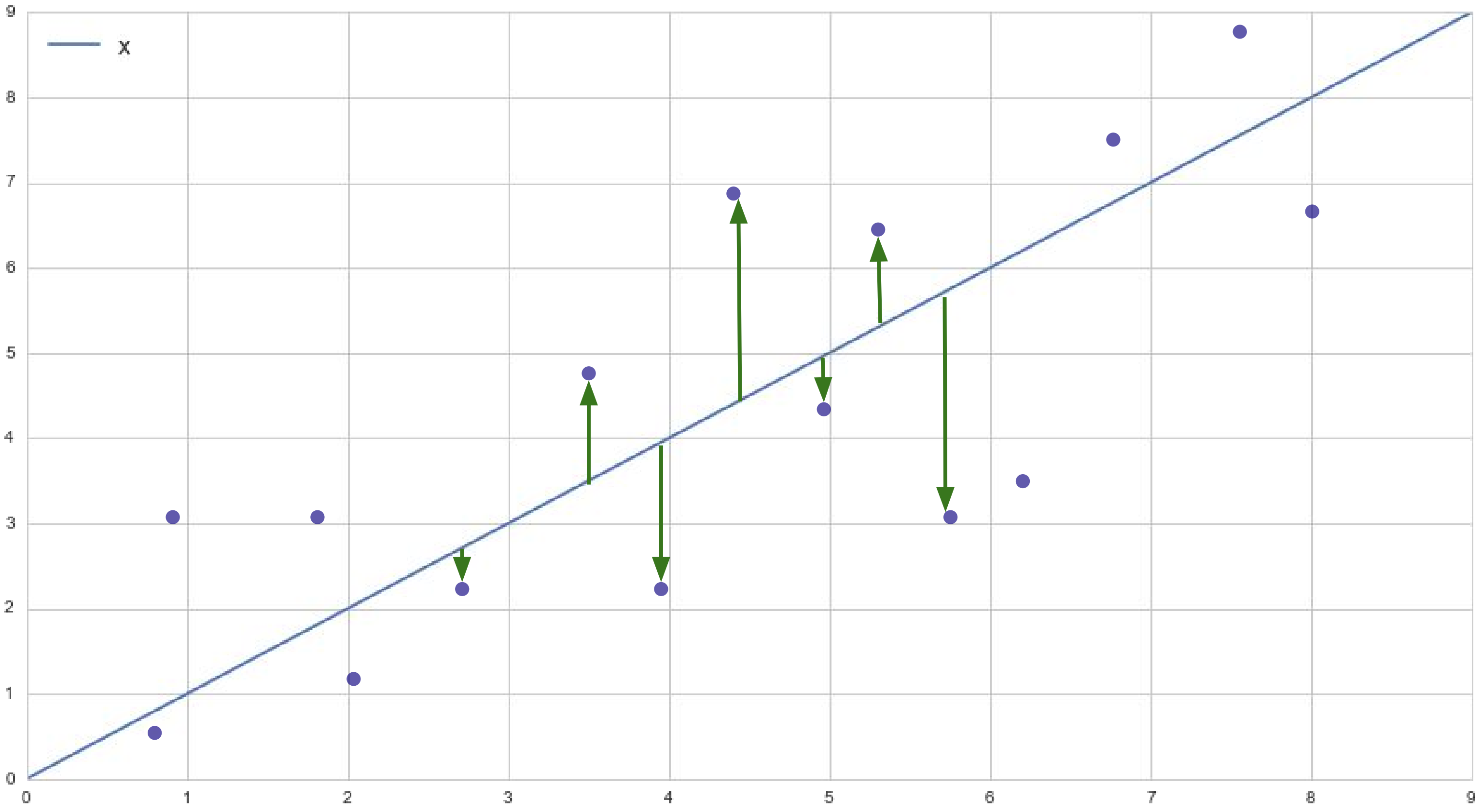

0.30
0.25
0.20
0.15
0.10
0.05
0.00
Occurance
−10
−5
0
5
10
15
Residual value

Accuracy

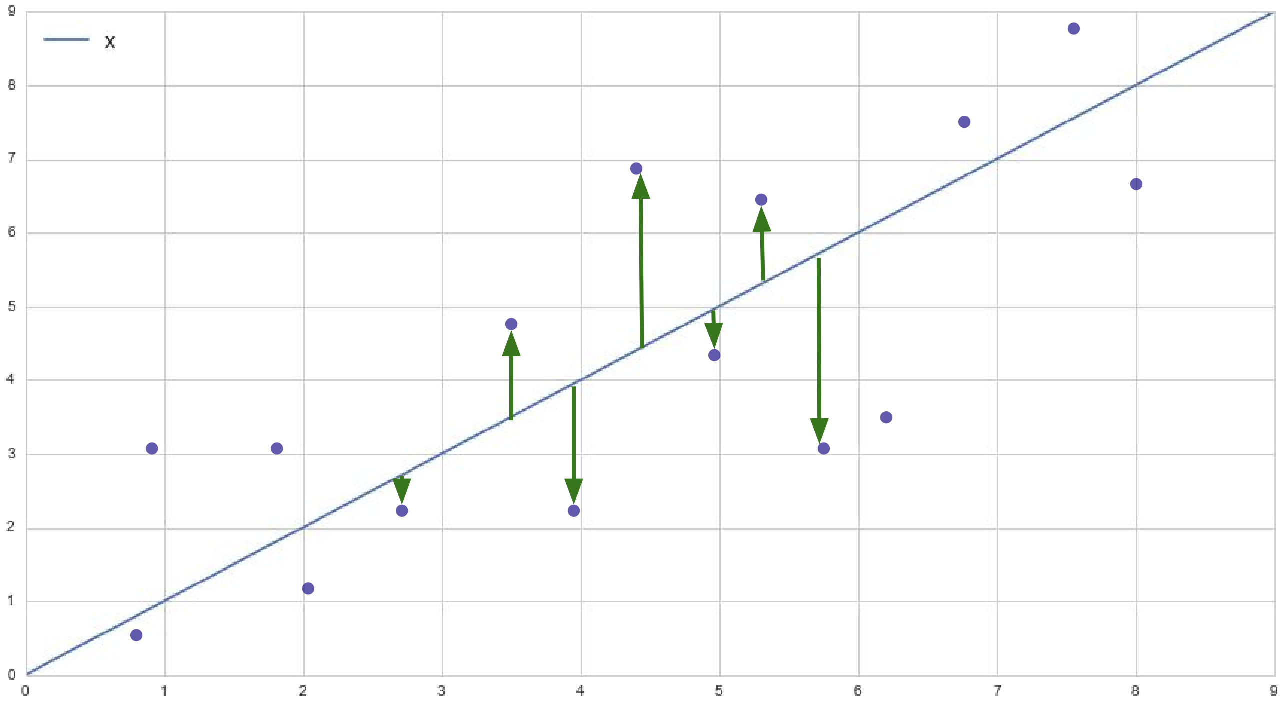

Residual distribution
Residual
Fitted value (y')
10
8
6
4
2
0
−2
−4
−6
−8
−200
−150
−100
−50
0
50
100
150
200
250

Accuracy

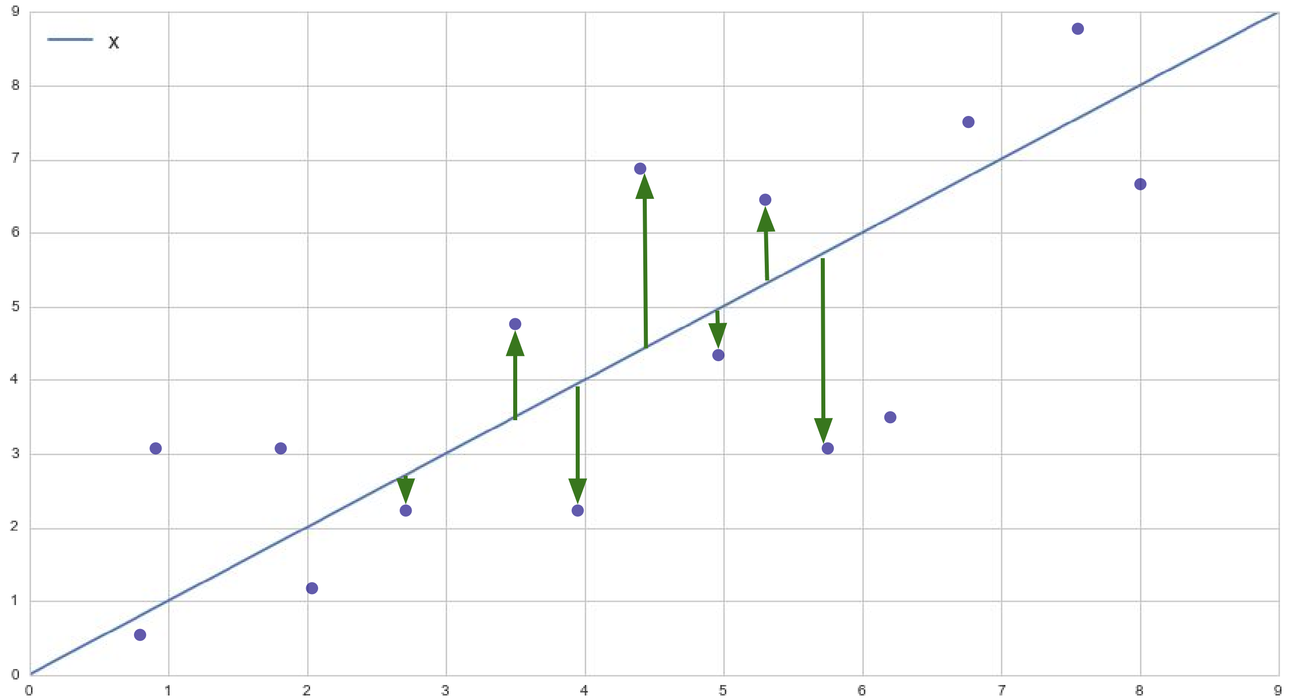
1. Score
2. Plot your residuals
3. R^2 and Adjusted R^2

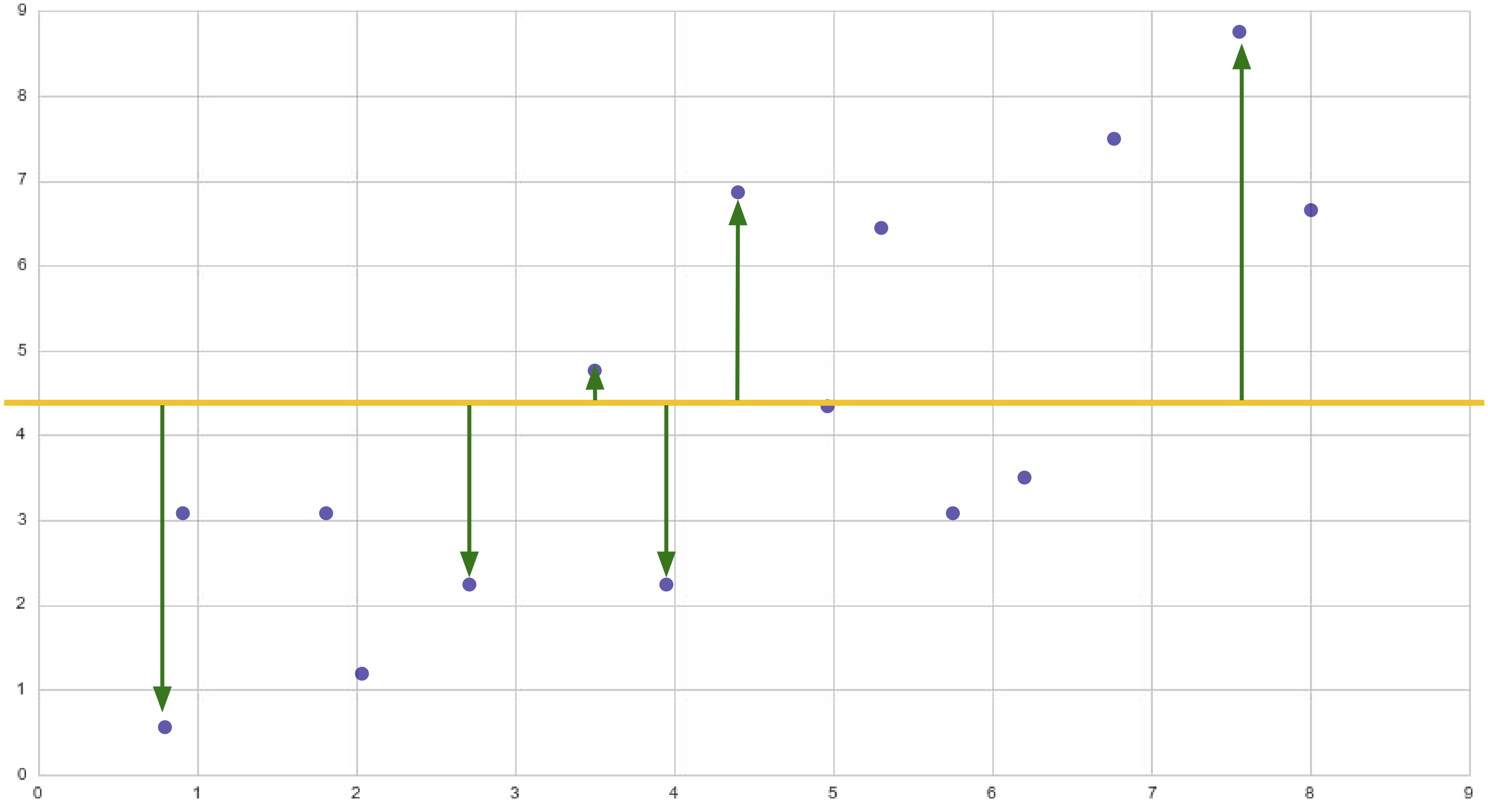

x
0
1
2
3
4
5
6
7
8
9

$$SUM(y - \hat{y})^2$$

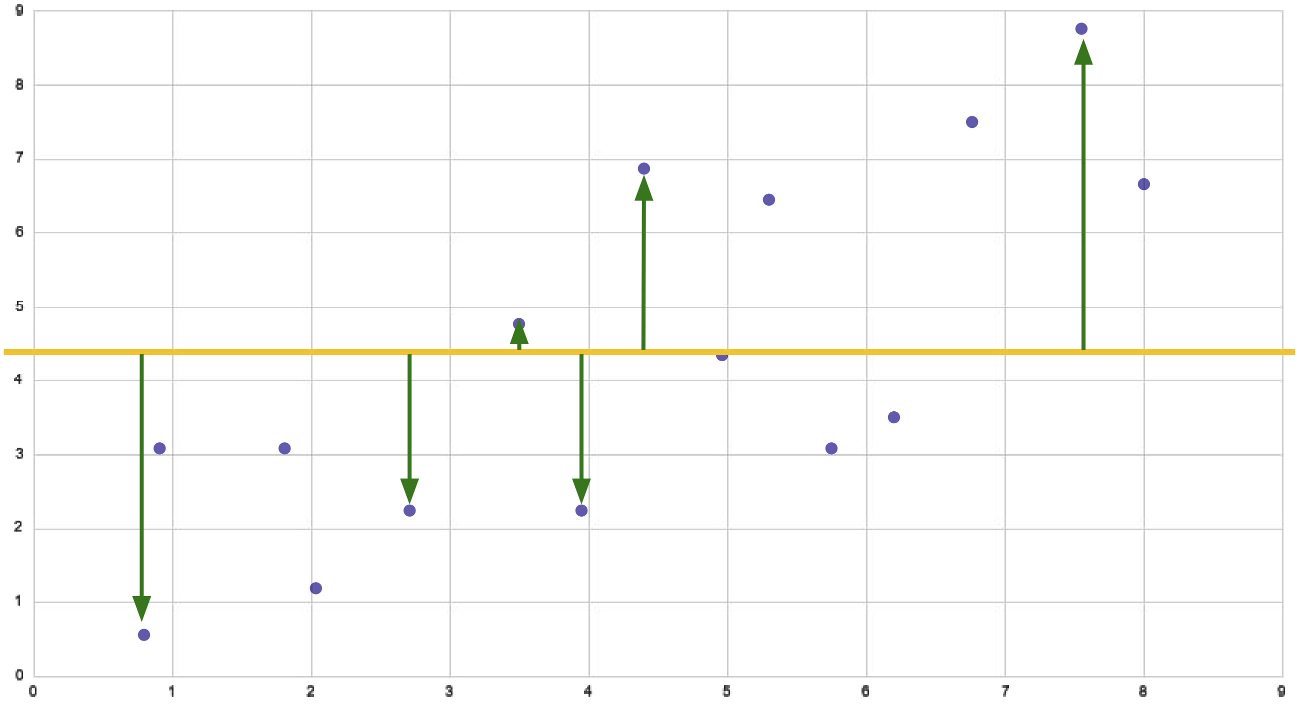


Residual Sum Squares = SUM$(y - \hat{y})^2$

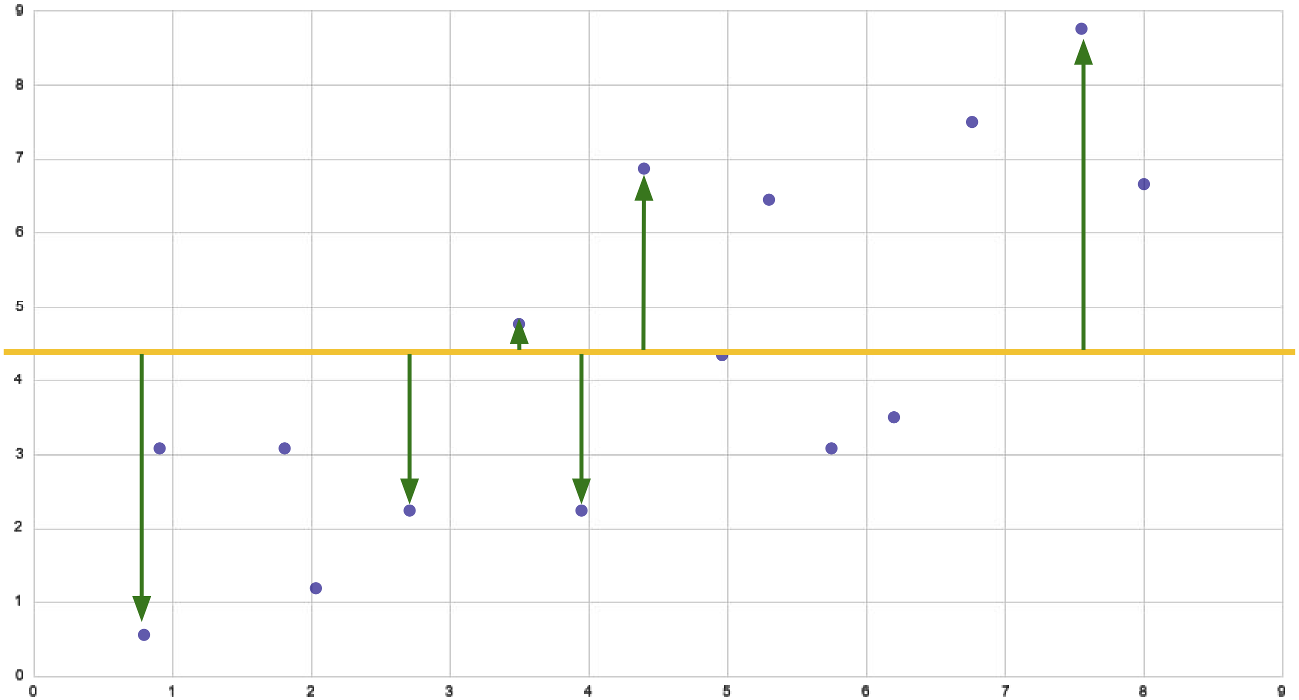



0
1
2
3
4
5
6
7
8
9
0
1
2
3
4
5
6
7
8
9

$$SUM(y - \bar{y})^2$$



Total Sum Squares = $SUM(y - \bar{y})^2$



$$R^2 = 1 - \frac{\text{RSS } [\text{SUM}(y - \hat{y})^2]}{\text{TSS } [\text{SUM}(y - \bar{y})^2]}$$

$$R^2 = 1 - \frac{\text{RSS } [\text{SUM}(y - \hat{y})^2]}{\text{TSS } [\text{SUM}(y - \bar{y})^2]}$$

- works for simple linear regression

$$R^2 = 1 - \frac{\text{RSS } [\text{SUM}(y - \hat{y})^2]}{\text{TSS } [\text{SUM}(y - \bar{y})^2]}$$

- fine for simple linear regression
- has problems with multiple linear regression

$$y = m_1 x_1 + b + \in$$

$$y = m_1 x_1 + m_2 x_2 + b + \in$$

$$y = m_1 \boxed{x_1} + m_2 \boxed{x_2} + b + \in$$

Features

$$y = \boxed{m_1} x_1 + \boxed{m_2} x_2 + b + \in$$

Coefficients

Regression

1. Estimate the relationships between predictor and target variables

2. Incurs loss as a result of this estimation

3. Attempts to minimize loss

Regression

1. Estimate the relationships between predictor and target variables

2. Incurs loss as a result of this estimation

3. Attempts to minimize loss

$$y = \boxed{m_1} x_1 + \boxed{m_2} x_2 + b + \in$$

Coefficients

$$y = m_1 x_1 + \boxed{0} x_2 + b + \in$$

Coefficients

$$\text{Adj } R^2 = 1 - (1 - R^2)\frac{n - 1}{n - p - 1}$$

$$\text{Adj } R^2 = 1 - (1 - R^2)\frac{n - 1}{n - p - 1}$$

n = sample size

p = number of predictors

$$\text{Adj } R^2 = 1 - (1 - R^2)\frac{\boxed{n - 1}}{n - p - 1}$$

n = sample size

p = number of predictors

$$\text{Adj } R^2 = 1 - (1 - R^2)\frac{n - 1}{\boxed{n - p - 1}}$$

n = sample size

p = number of predictors

$$\text{Adj } R^2 = 1 - (1 - R^2)\boxed{\frac{n - 1}{n - p - 1}}$$

n = sample size

p = number of predictors

$$\text{Adj } R^2 = 1 - \boxed{(1 - R^2)\frac{n - 1}{n - p - 1}}$$

n = sample size

p = number of predictors

$$\text{Adj } R^2 = 1 - (1 - \boxed{R^2})\frac{n - 1}{n - p - 1}$$

n = sample size

p = number of predictors

$$\text{Adj } R^2 = 1 - \boxed{(1 - R^2)} \frac{n - 1}{n - p - 1}$$

n = sample size

p = number of predictors

Assumptions

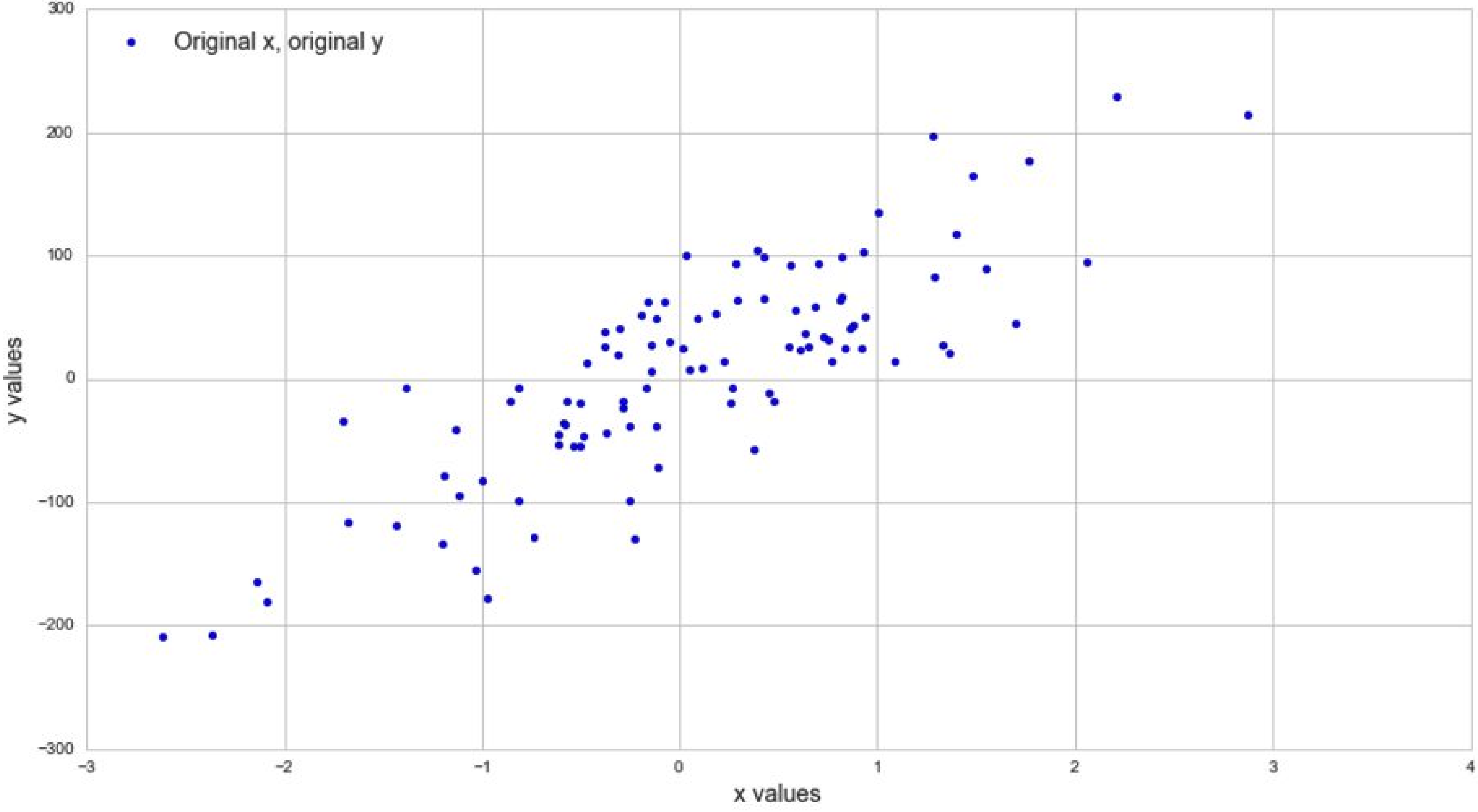
or: Is That A Linear Regression in Your Pocket or Are You Just Happy to See Me?

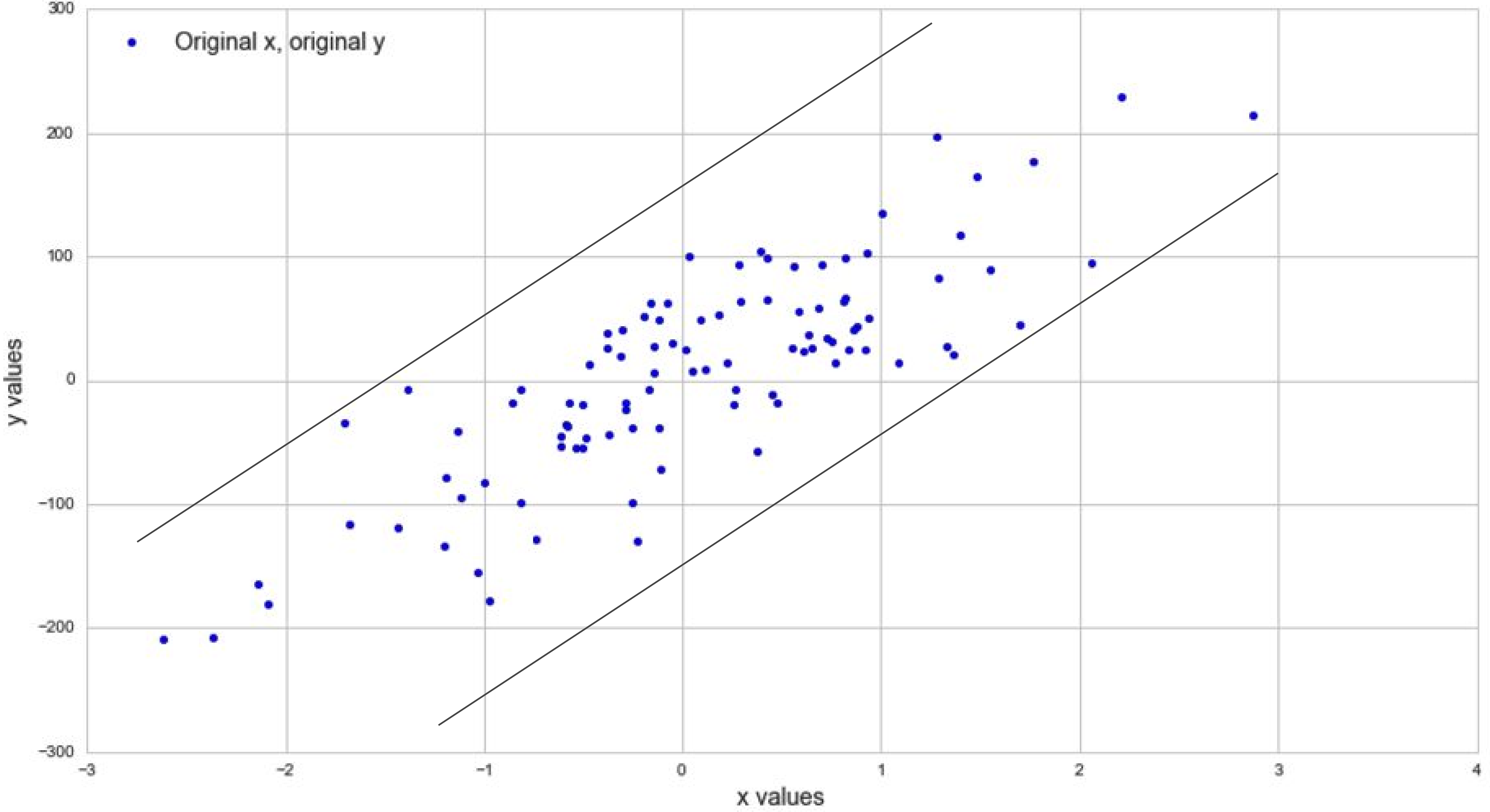
- That your data are linearly correlated.

- That your data are linearly correlated.
- That the weights of the predictor values can be summed meaningfully.

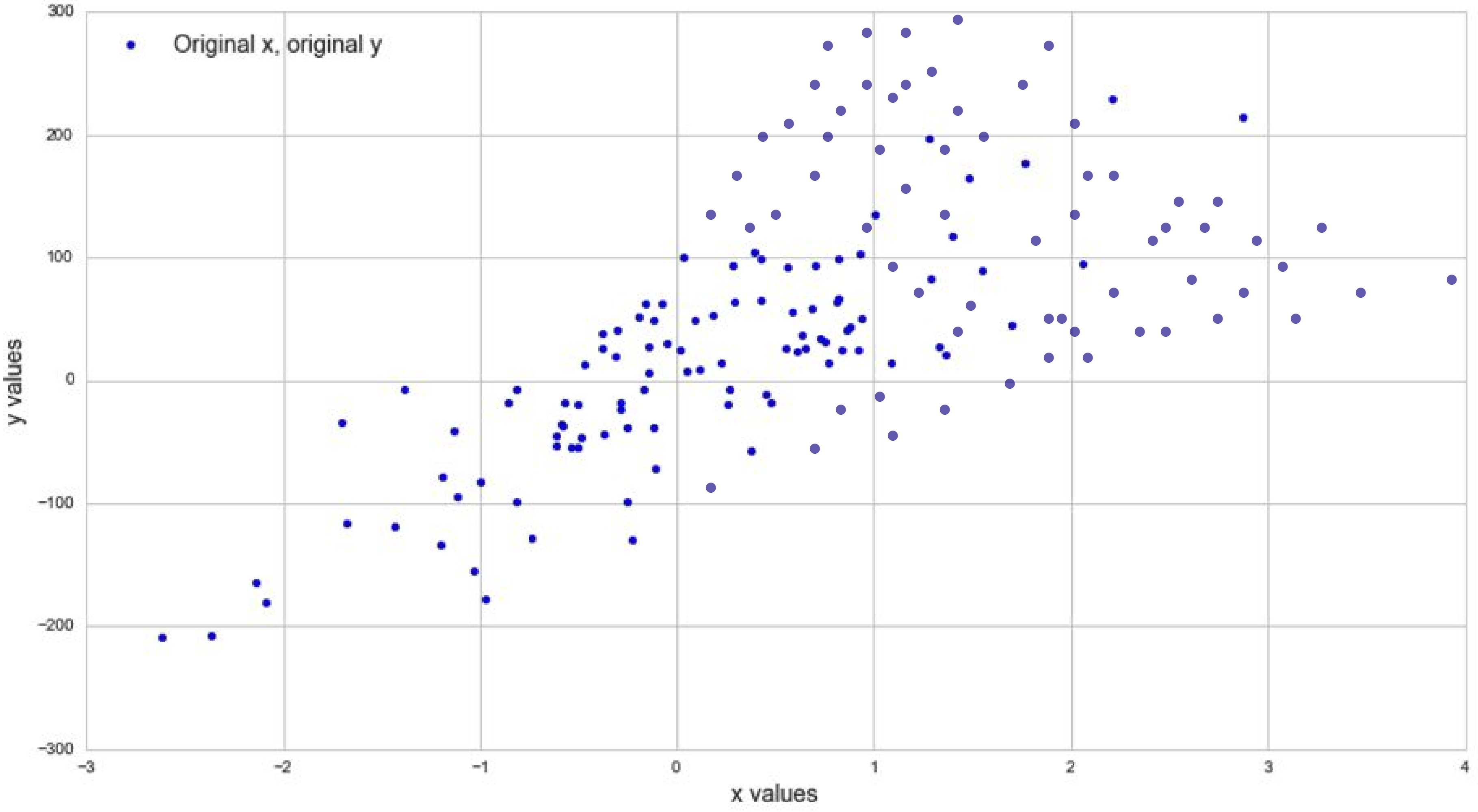
- That your data are linearly correlated.
- That the weights of the predictor values can be summed meaningfully.
- No linear dependence, aka multicollinearity.

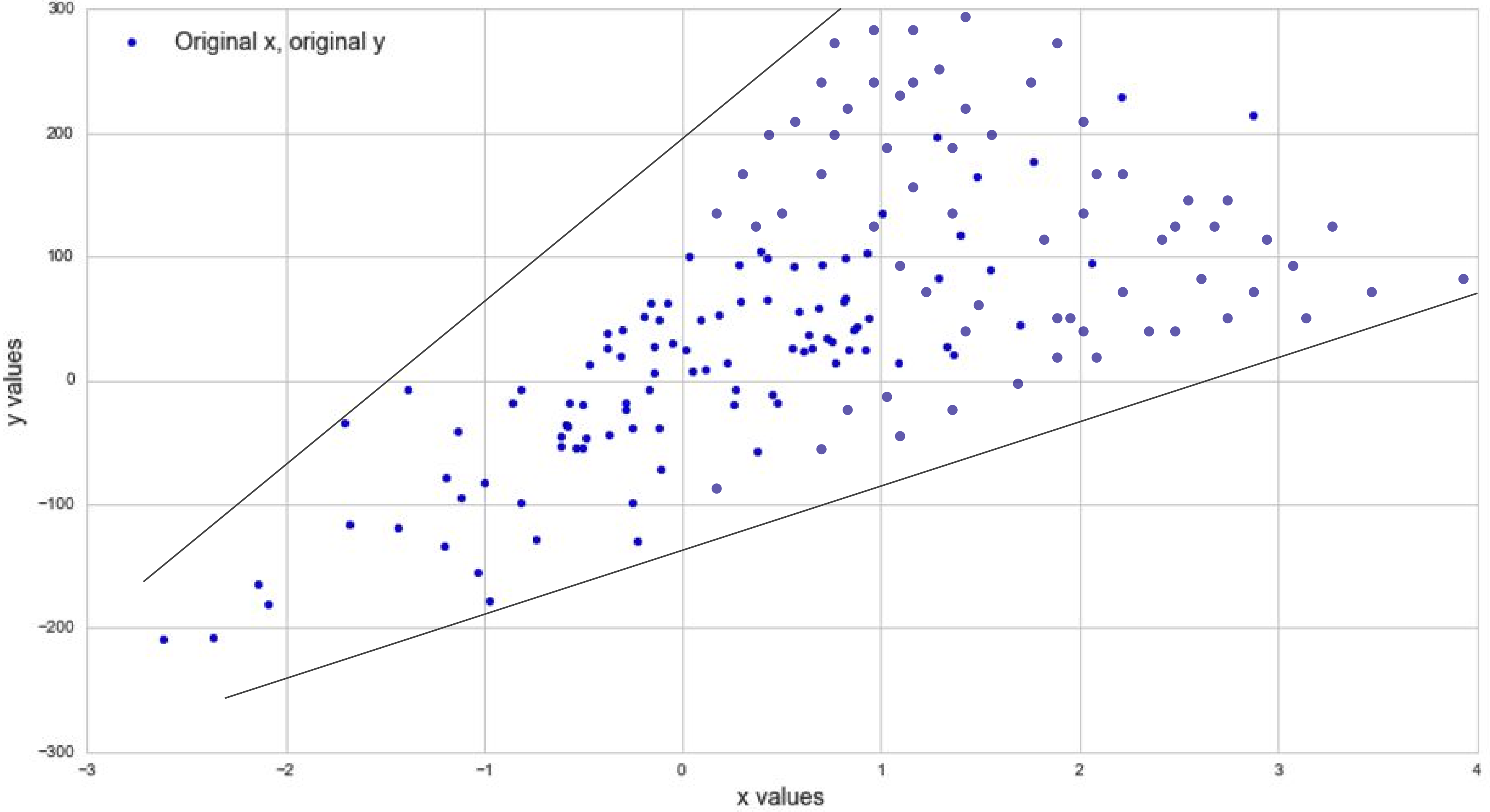
- That your data are linearly correlated.
- That the weights of the predictor values can be summed meaningfully.
- No linear dependence, aka multicollinearity.
- Homoscedasticity


Original x, original y
300
200
100
0
−100
−200
−300
−3
−2
−1
0
1
2
3
4
x values
y values



Original x, original y
300
200
100
0
−100
−200
−300
y values
−3
−2
−1
0
1
2
3
4
x values


Original x, original y
y values
x values
300
200
100
0
−100
−200
−300
−3
−2
−1
0
1
2
3
4



Original x, original y
y values
x values
300
200
100
0
−100
−200
−300
−3
−2
−1
0
1
2
3
4

Final Thoughts

“Correlation doesn't imply causation...

Anscombe’s Quartet

Anscombe's quartet comprises four datasets that have nearly identical simple statistical properties, yet appear very different when graphed.

They were constructed in 1973 by the statistician Francis Anscombe to demonstrate both the importance of graphing data before analyzing it and the effect of outliers on statistical properties.

He described the graphs as being intended to attack the impression among statisticians that "numerical calculations are exact, but graphs are rough."

Anscombe’s Quartet

All the summary statistics are close to identical:

- The average *x* value is 9

Anscombe’s Quartet

All the summary statistics are close to identical:

- The average x value is 9
- The average y value is 7.50

Anscombe's Quartet

All the summary statistics are close to identical:

- The average x value is 9
- The average y value is 7.50
- The variance for x is 11 and the variance for y is 4.12

Anscombe’s Quartet

All the summary statistics are close to identical:

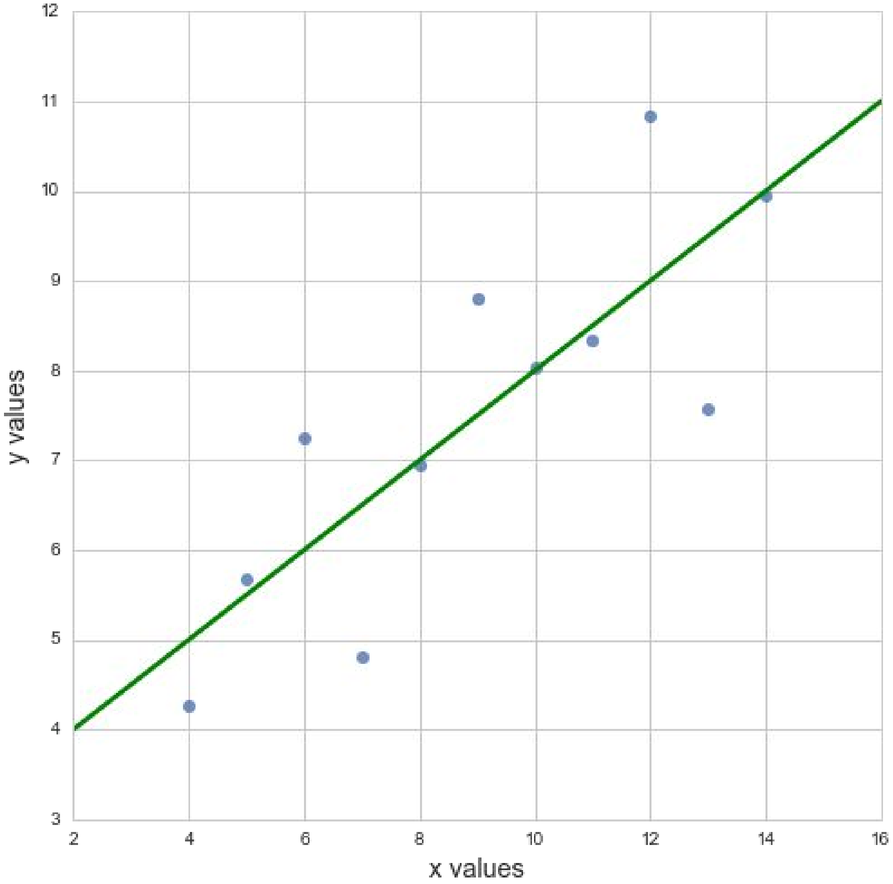
- The average x value is 9
- The average y value is 7.50
- The variance for x is 11 and the variance for y is 4.12
- The correlation between x and y is 0.816

Anscombe’s Quartet

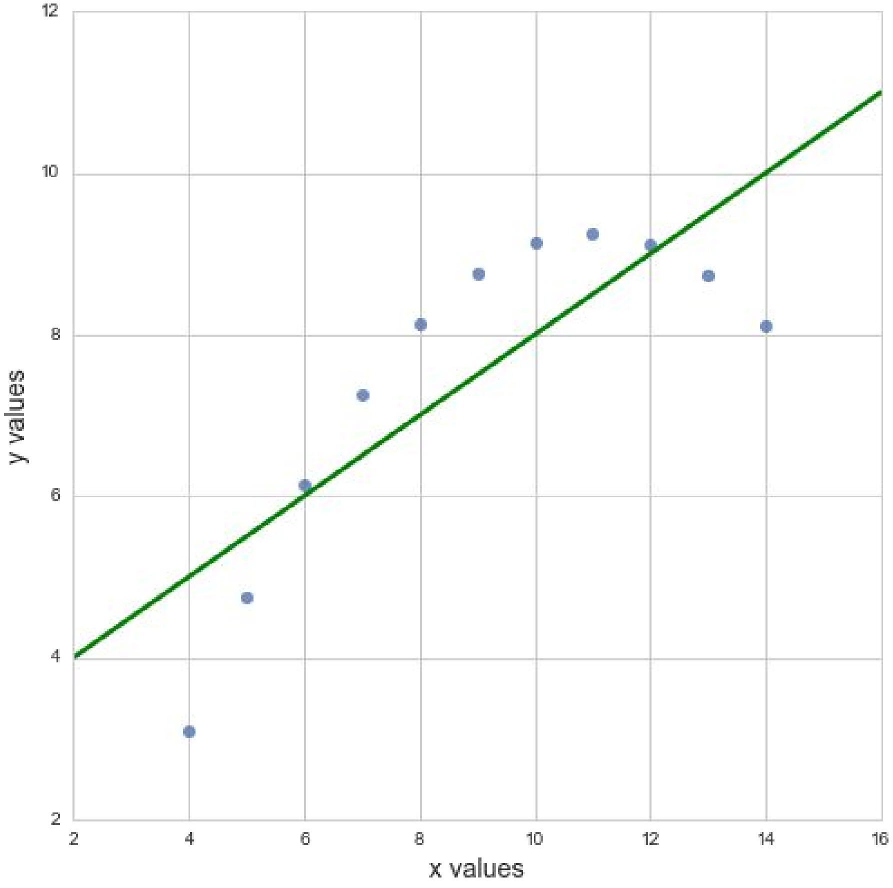
All the summary statistics are close to identical:

- The average x value is 9
- The average y value is 7.50
- The variance for x is 11 and the variance for y is 4.12
- The correlation between x and y is 0.816
- A linear regression (line of best fit) follows the equation $y = 0.5x + 3$

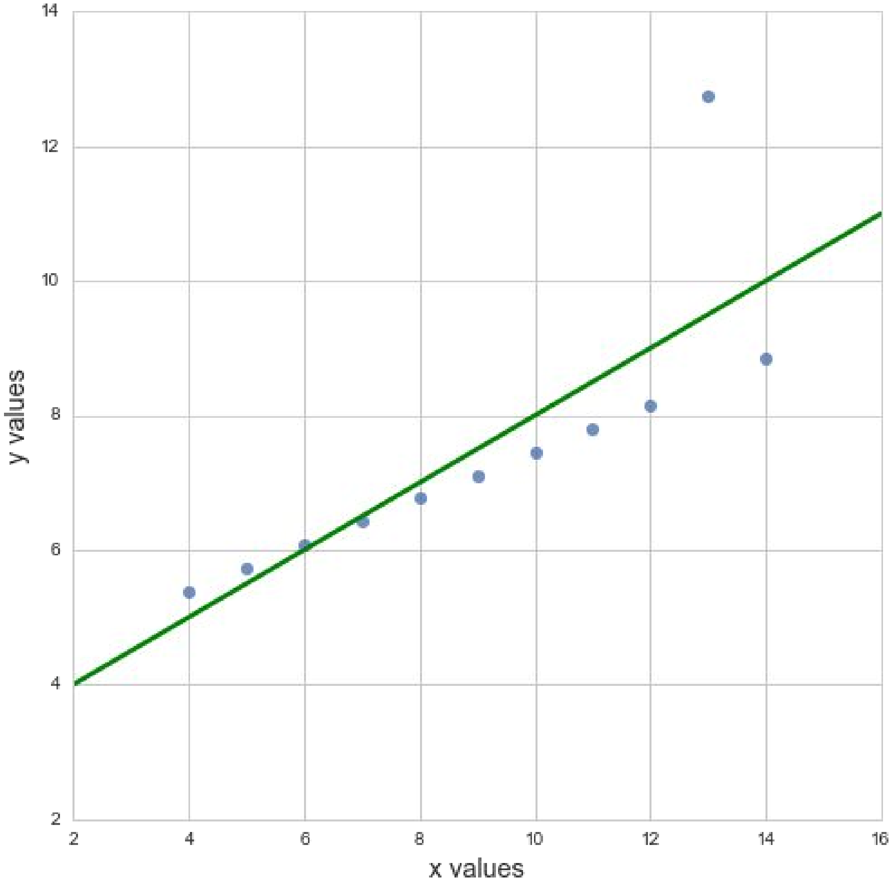
Anscombe's Quartet



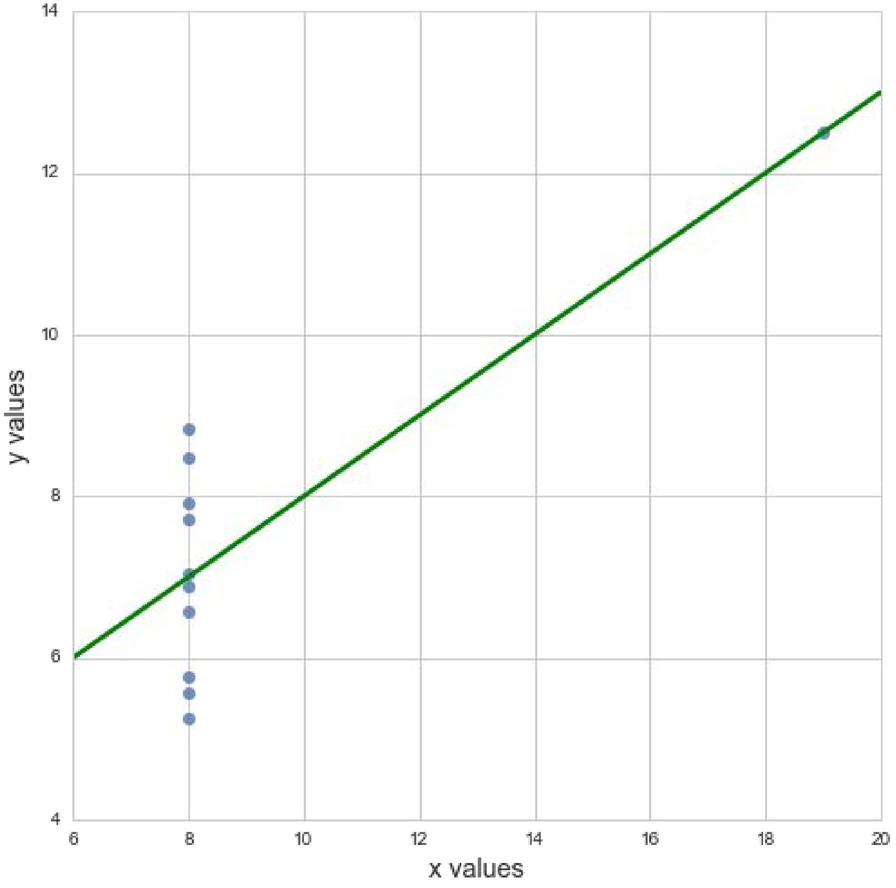
Anscombe’s Quartet



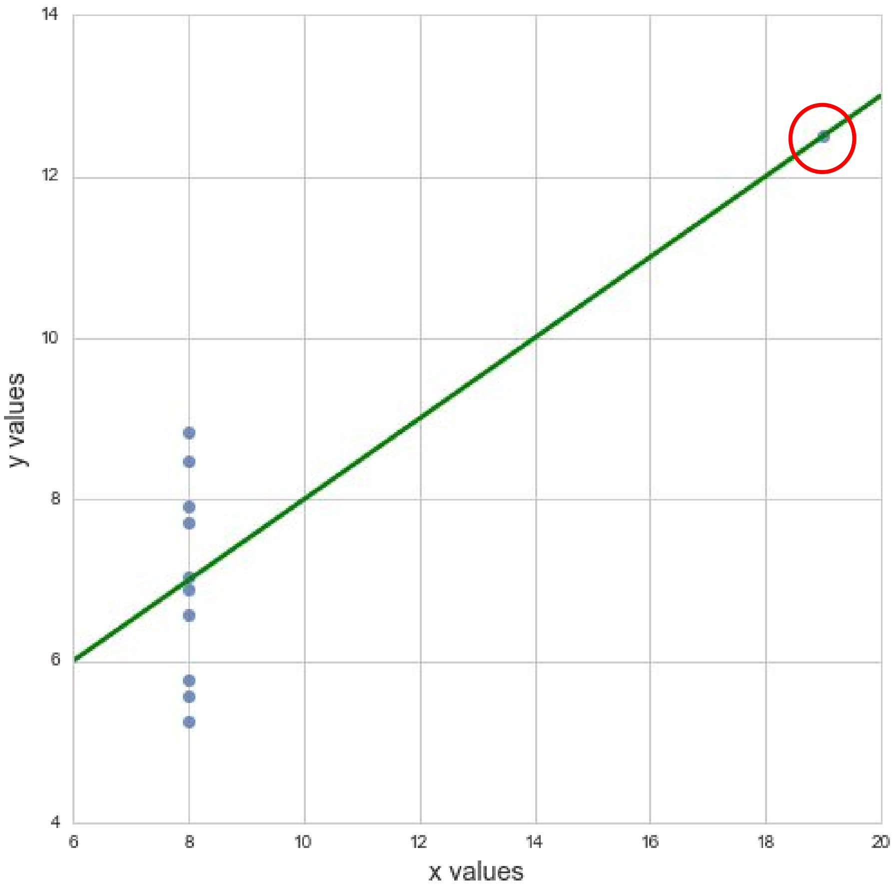
Anscombe's Quartet



Anscombe’s Quartet

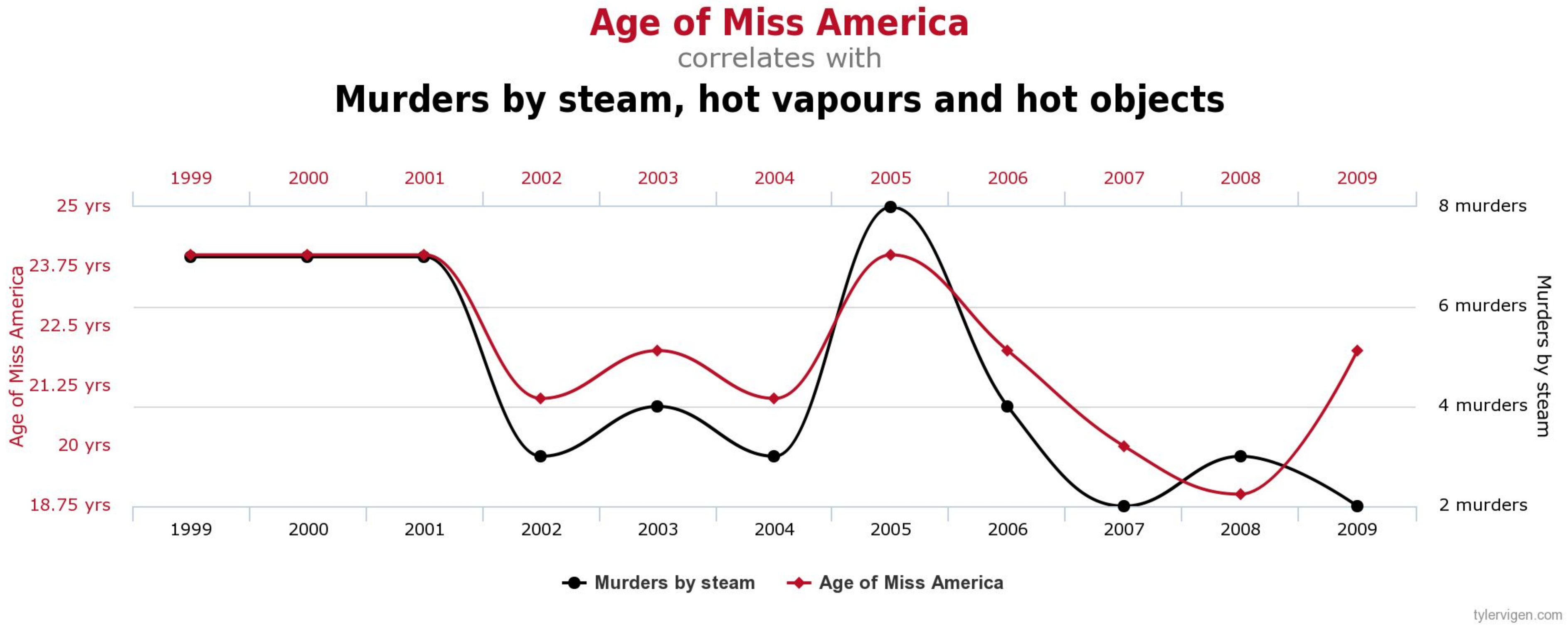


Anscombe’s Quartet



"There are three kinds of lies: lies, damned lies, and *statistics*."

- British Prime Minister Benjamin Disraeli


Age of Miss America
correlates with
Murders by steam, hot vapours and hot objects
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
25 yrs
23.75 yrs
22.5 yrs
21.25 yrs
20 yrs
18.75 yrs
Age of Miss America
8 murders
6 murders
4 murders
2 murders
Murders by steam
Murders by steam
Age of Miss America
tylervigen.com

“Correlation doesn't imply causation...

“Correlation doesn't imply causation…

...but it does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there'."

– Randall Munroe, XKCD


coda
goodbye
bye-bye
Auf-Wiedersehen
thankyou
envoi
adieu
seeya
cheers
godspeed
toodle-oo
Au-revoir
regards
leave-taking
Arrivederci
conclusion
Sayonara
farewell
so long
culmination
cheerio
ciao
adios