

PSY9511: Seminar 4

The basics of regression and classification

Esten H. Leonardsen

29.05.2024



UNIVERSITETET
I OSLO

Recap

What is statistical learning?



What is statistical learning?

Inferential view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y .



Recap

What is statistical learning?

Inferential view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y .

Predictive view: Finding a function $\hat{f}(X)$ that, when given a new set of inputs X allows us to predict an output y .



Recap

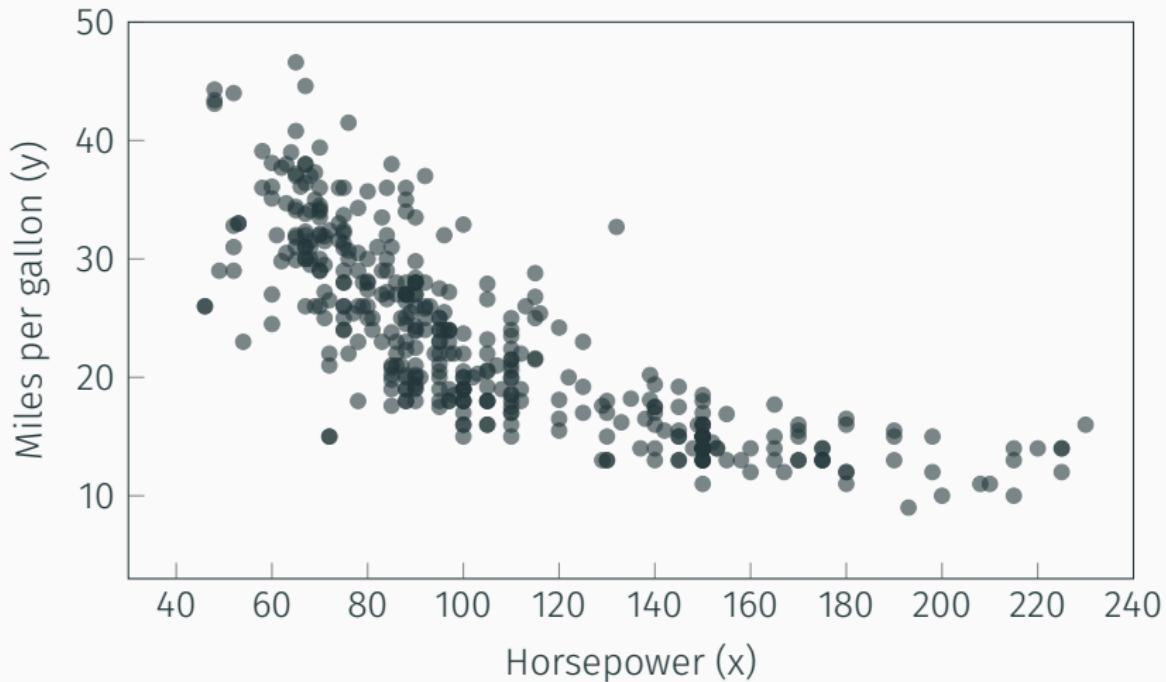
What is statistical learning?

Inferential view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y .

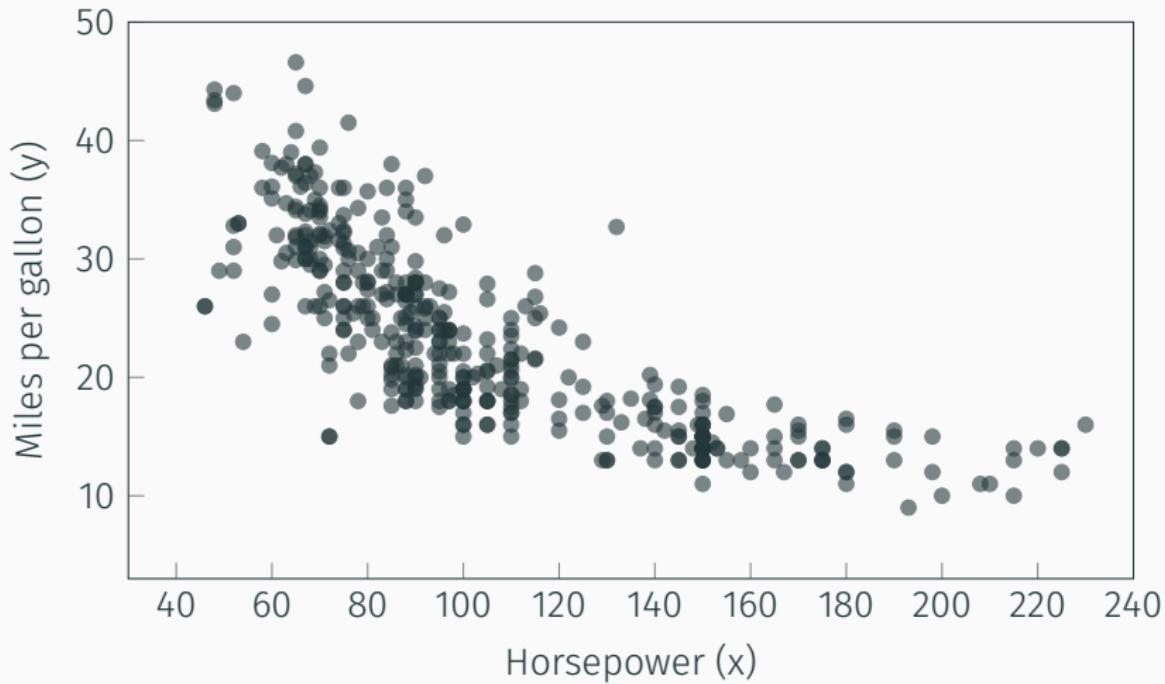
Predictive view: Finding a function $\hat{f}(X)$ that, when given a new set of inputs X allows us to predict an output y .



Recap



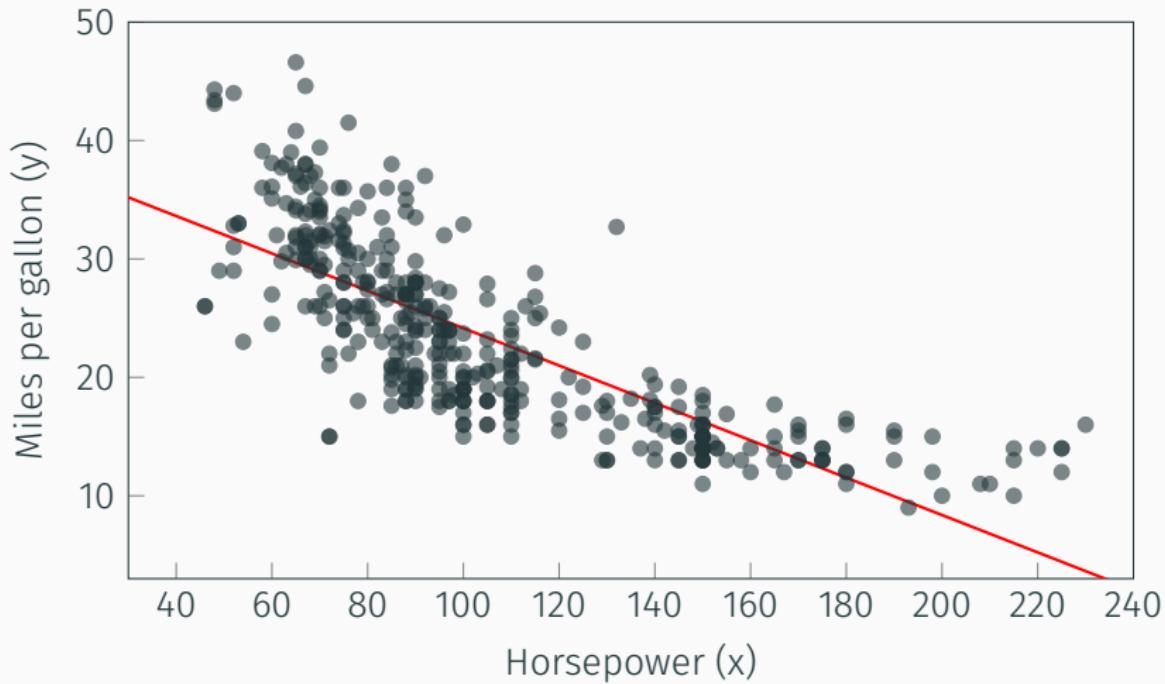
Recap



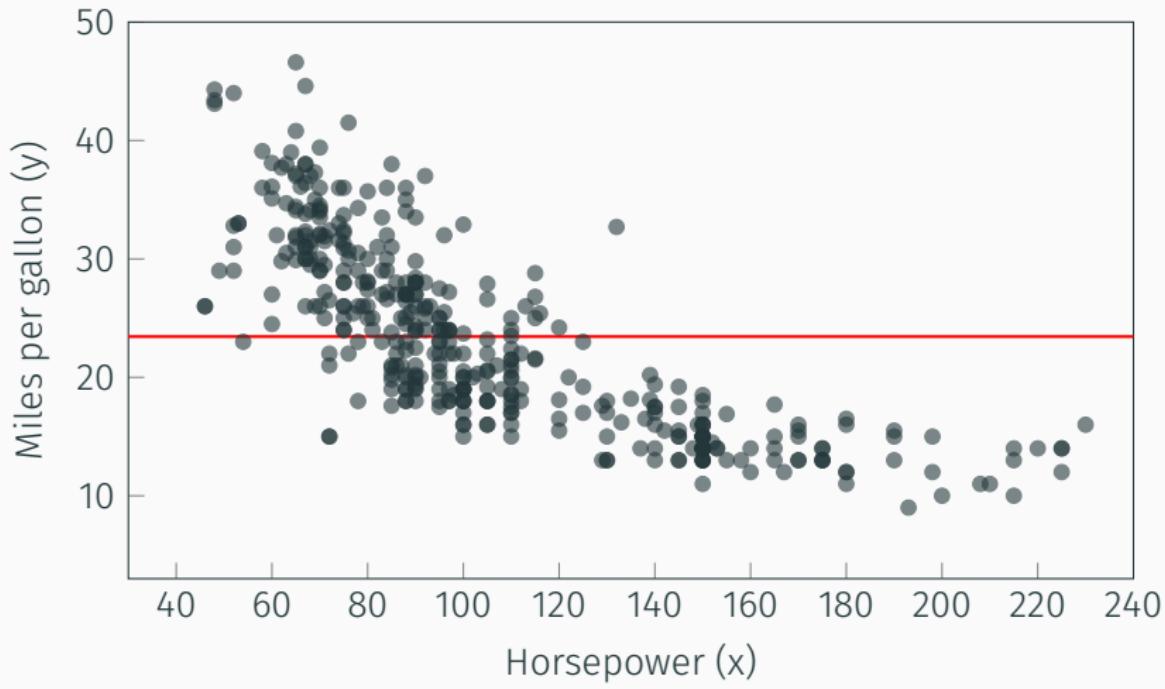
$$\hat{y} = \hat{f}(x)$$



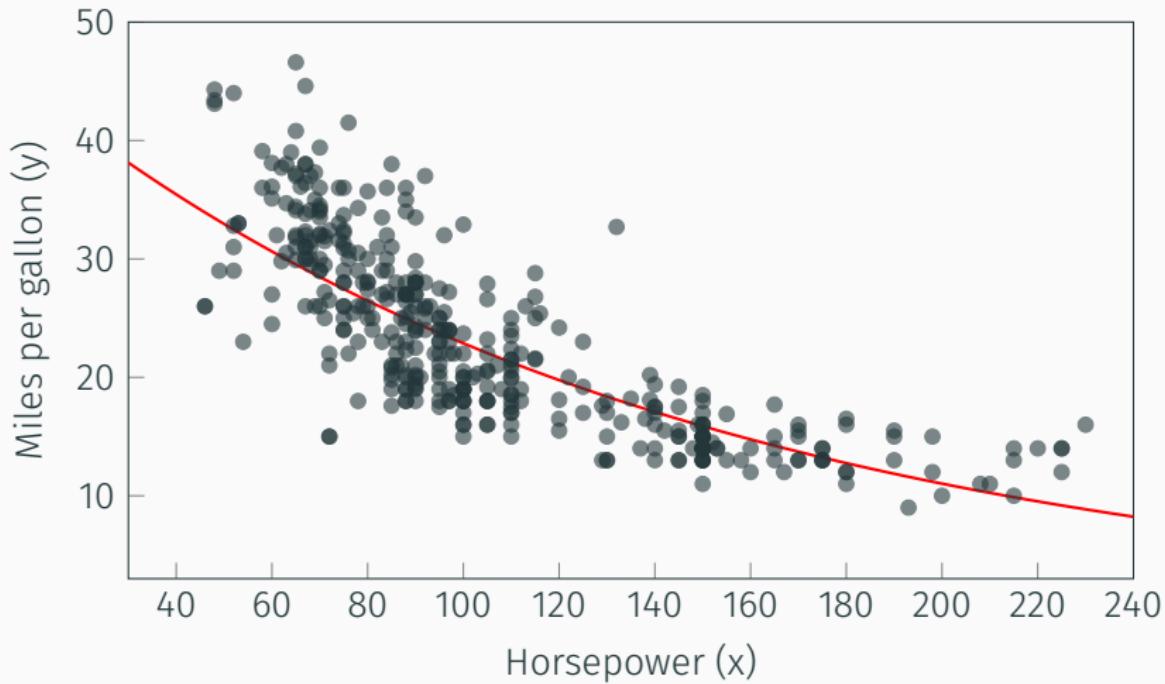
Recap



Recap



Recap



Outline

Plan for the day:

- Different types of outputs y : Regression vs classification
- Linear regression: Restricting the scope of $\hat{f}(X)$
 - Finding $\hat{f}(X)$: Training machine learning models
 - Live coding
- k Nearest Neighbours
- Logistic regression: Extending linear regression to classification
 - Live coding
- Generative models

Plan for future lectures:

- How do we evaluate how good our models are? (Lecture 3)
- Complex solutions to regression and classification problems (Lecture 4 and onwards)



Regression vs. classification

Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet



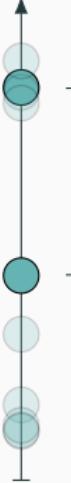
Regression vs. classification



Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet



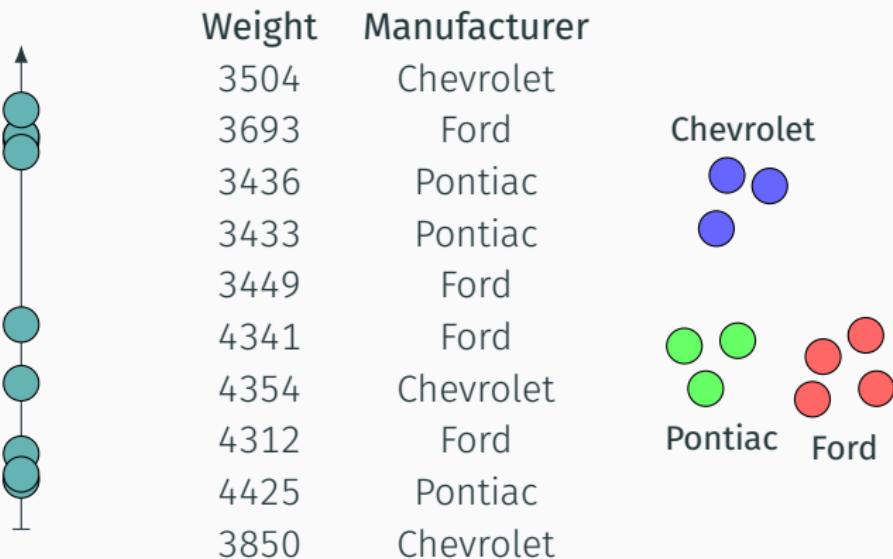
Regression vs. classification



Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet



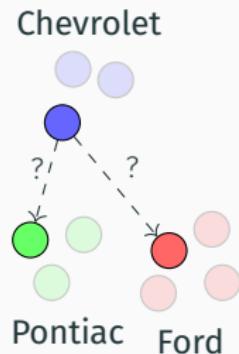
Regression vs. classification



Regression vs. classification



Weight	Manufacturer
3504	Chevrolet
3693	Ford
3436	Pontiac
3433	Pontiac
3449	Ford
4341	Ford
4354	Chevrolet
4312	Ford
4425	Pontiac
3850	Chevrolet



Regression vs. classification

Mean squared error (MSE):

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Accuracy

$$\mathbb{1}_{eq}(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}$$



Regression vs. classification



Regression vs. classification

The quick brown fox jumps over the lazy



Regression vs. classification

The quick brown fox jumps over the lazy _____



Regression vs. classification

"Students taking
a machine learning
class"

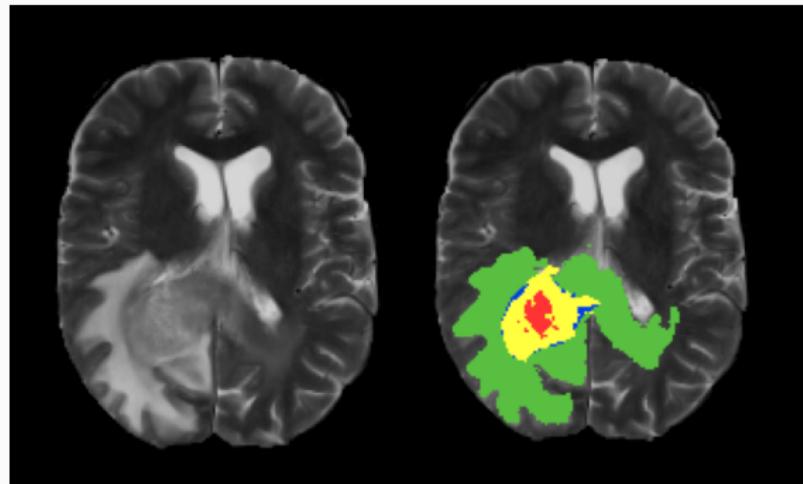


Regression vs. classification

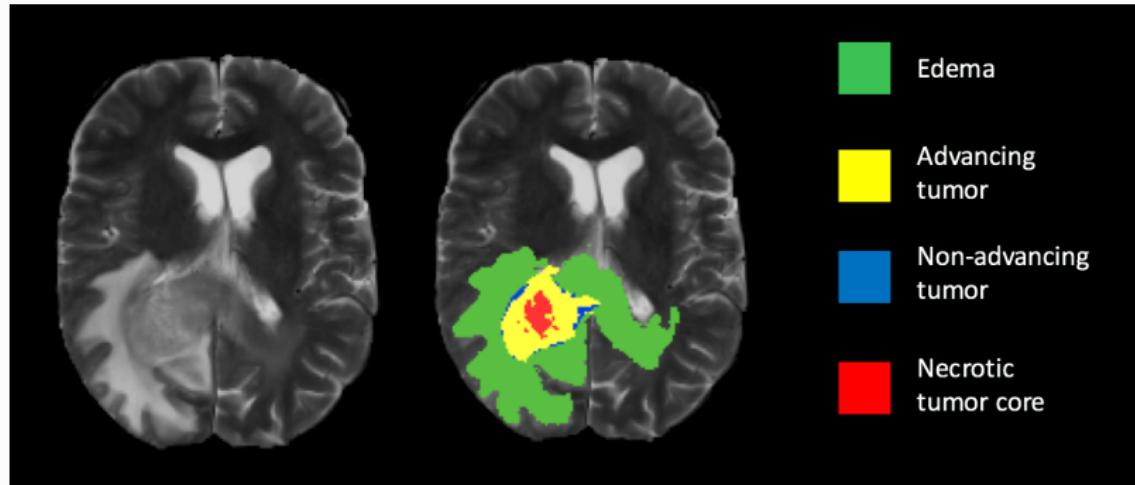
"Students taking
a machine learning
class"



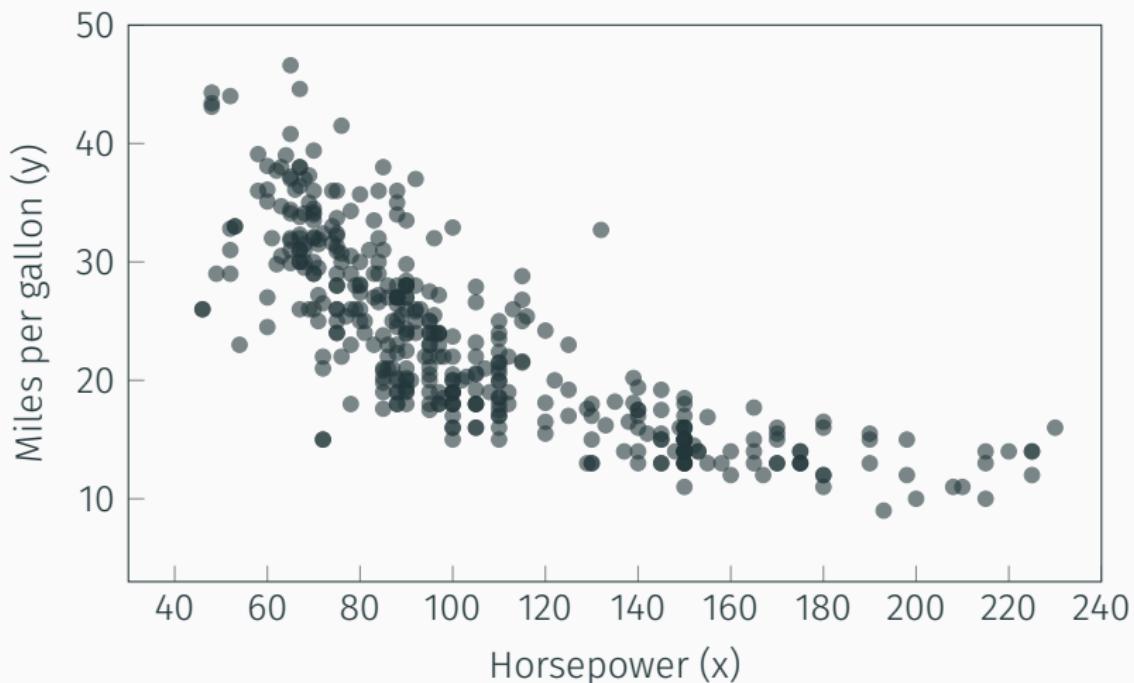
Regression vs. classification



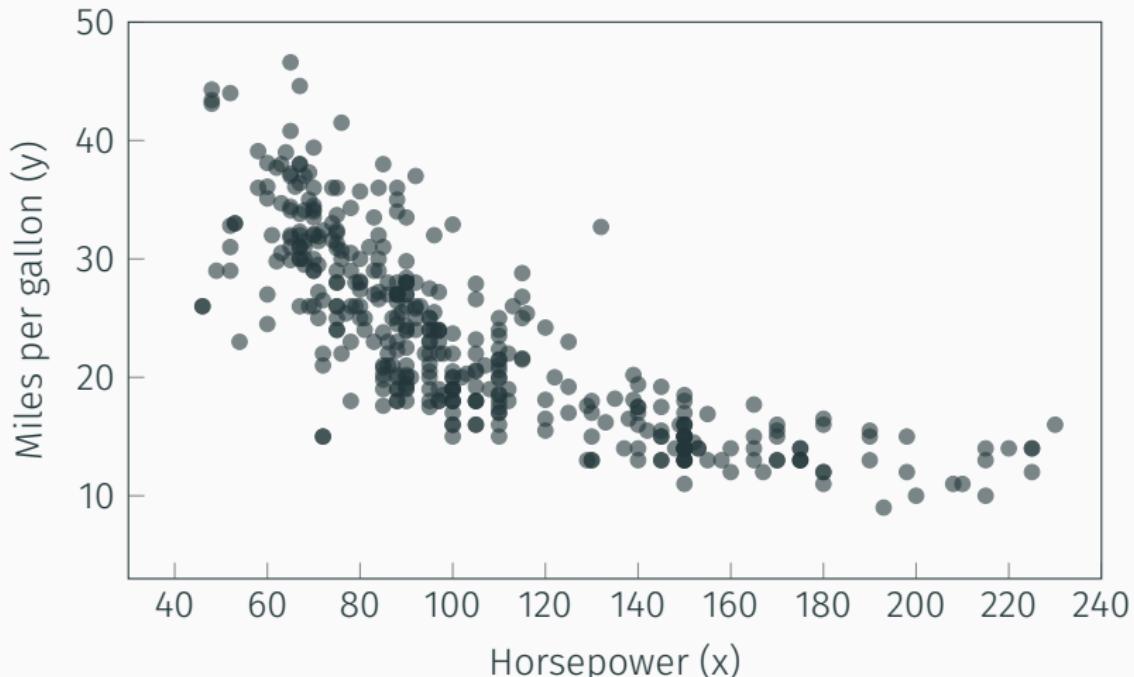
Regression vs. classification



Linear regression (via ordinary least squares)



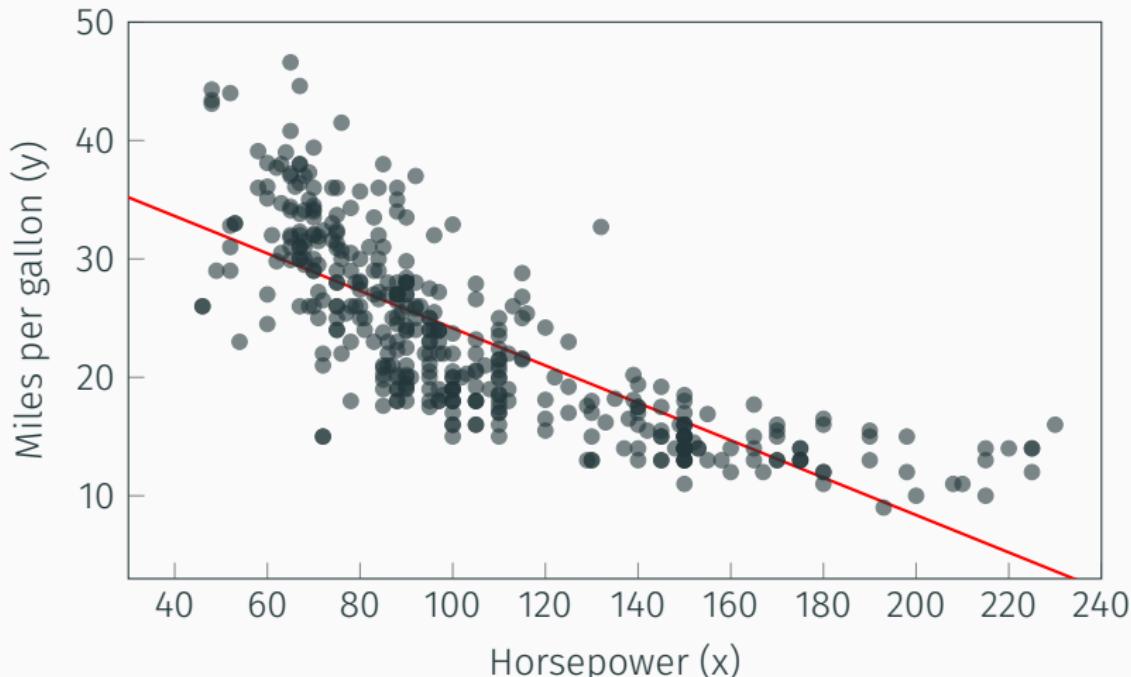
Linear regression (via ordinary least squares)



$$\hat{y} = \beta_0 - \beta_1 x$$



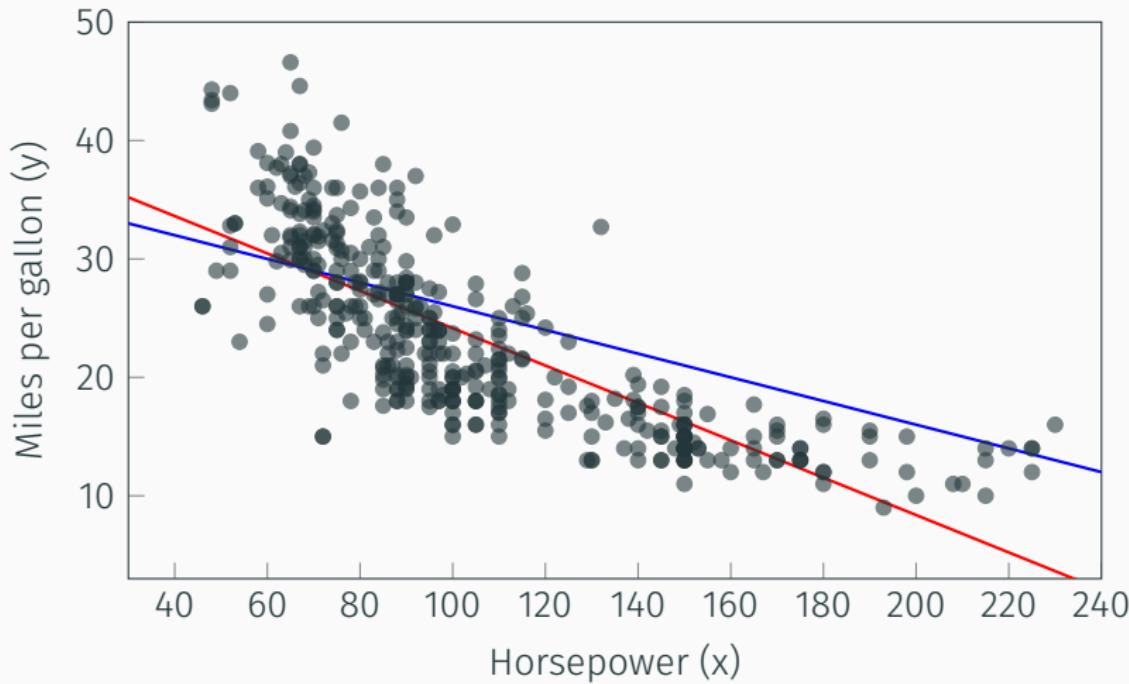
Linear regression (via ordinary least squares)



$$\hat{y} = 39.93 - 0.1578x$$



Linear regression (via ordinary least squares)

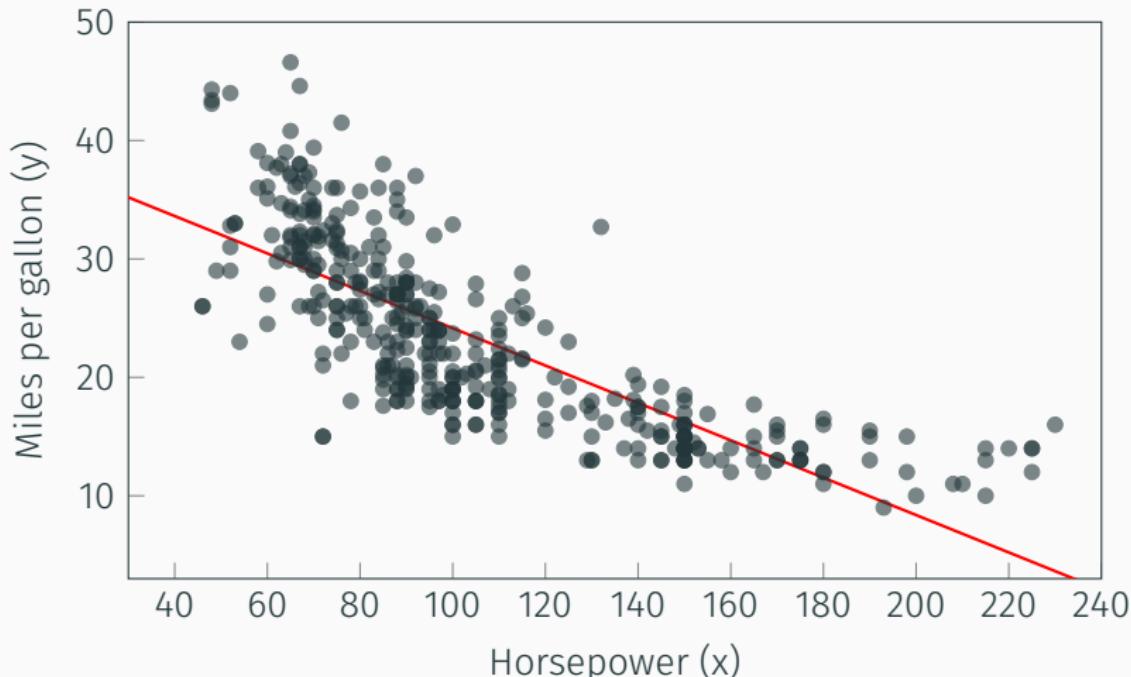


$$\hat{y} = 39.93 - 0.1578x$$

$$\hat{y} = 36 - 0.1x$$



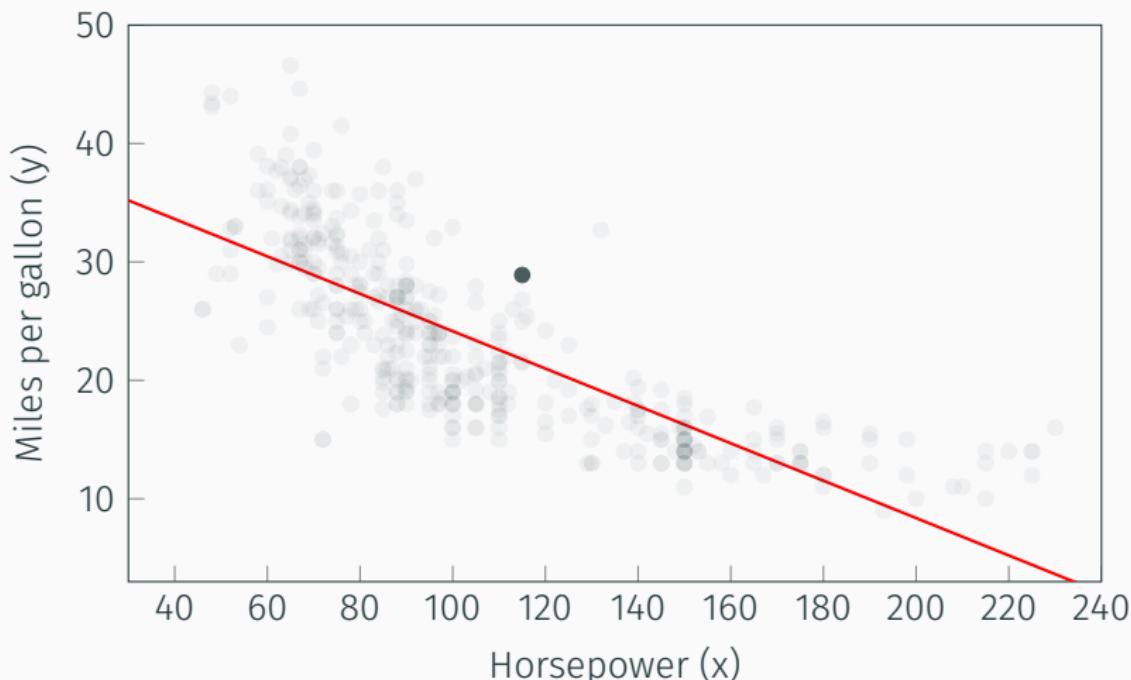
Linear regression (via ordinary least squares)



$$\hat{y} = 39.93 - 0.1578x$$



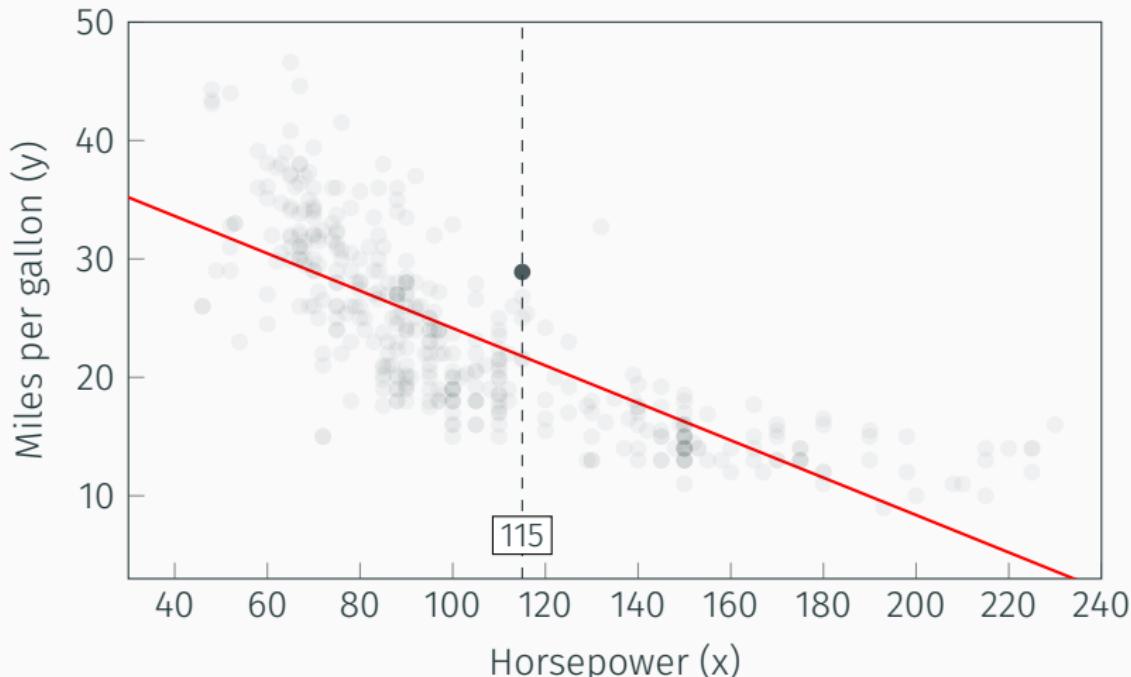
Linear regression (via ordinary least squares)



$$\hat{y} = 39.93 - 0.1578x$$



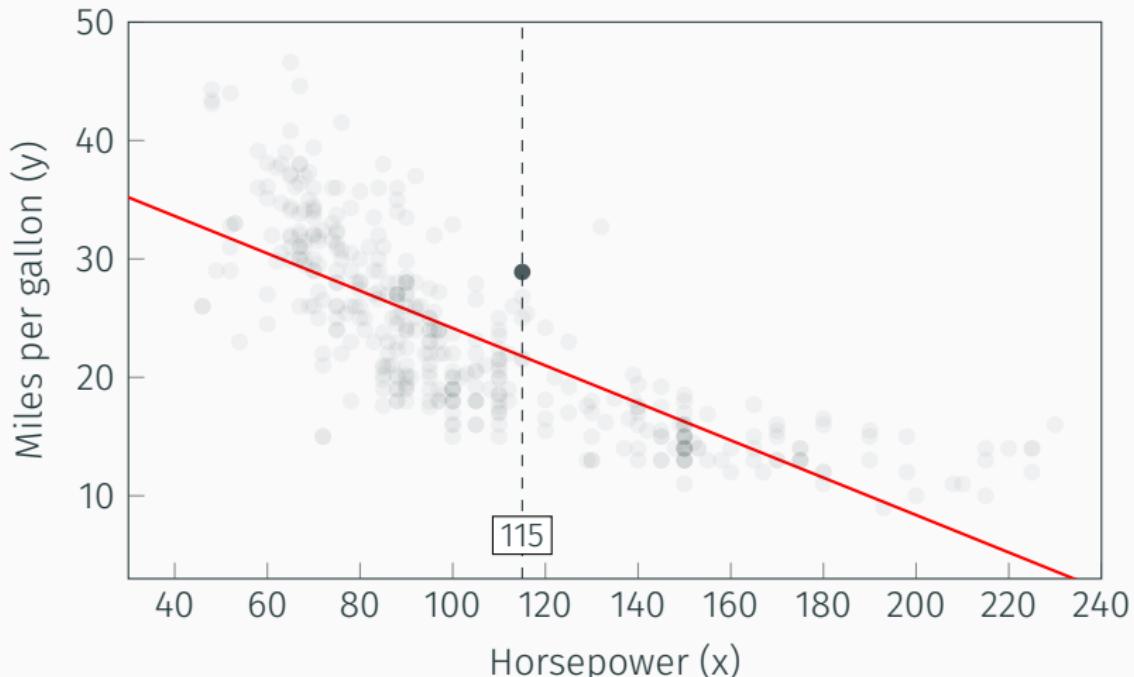
Linear regression (via ordinary least squares)



$$\hat{y} = 39.93 - 0.1578x$$



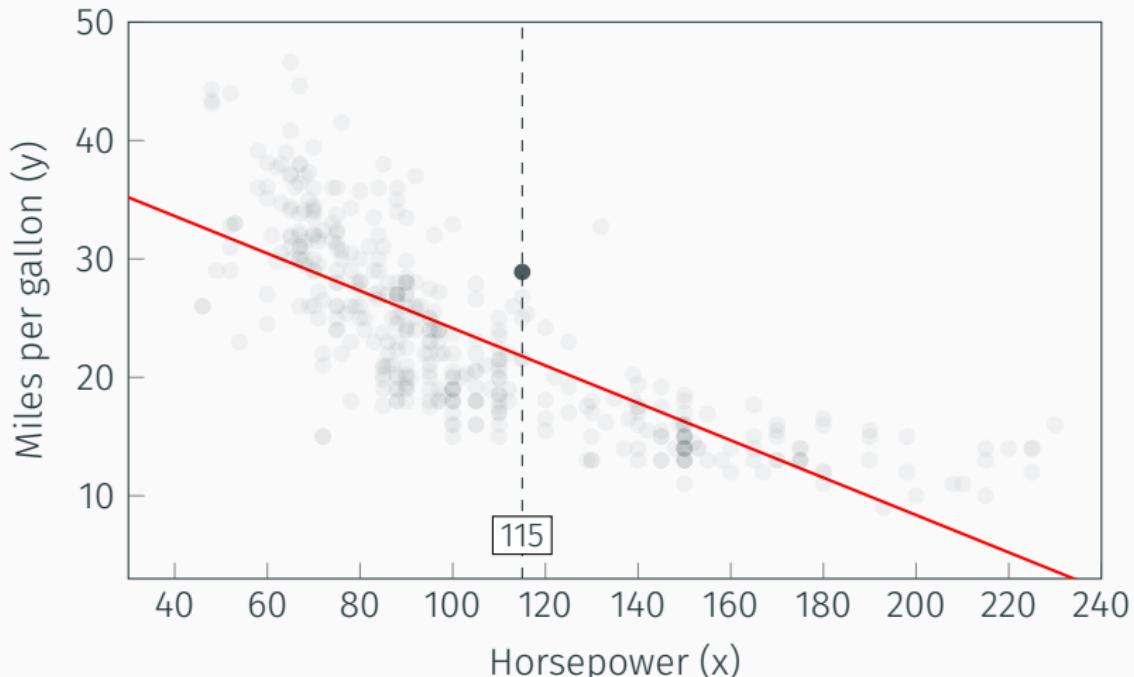
Linear regression (via ordinary least squares)



$$\hat{y} = 39.93 - 0.1578 * 115$$



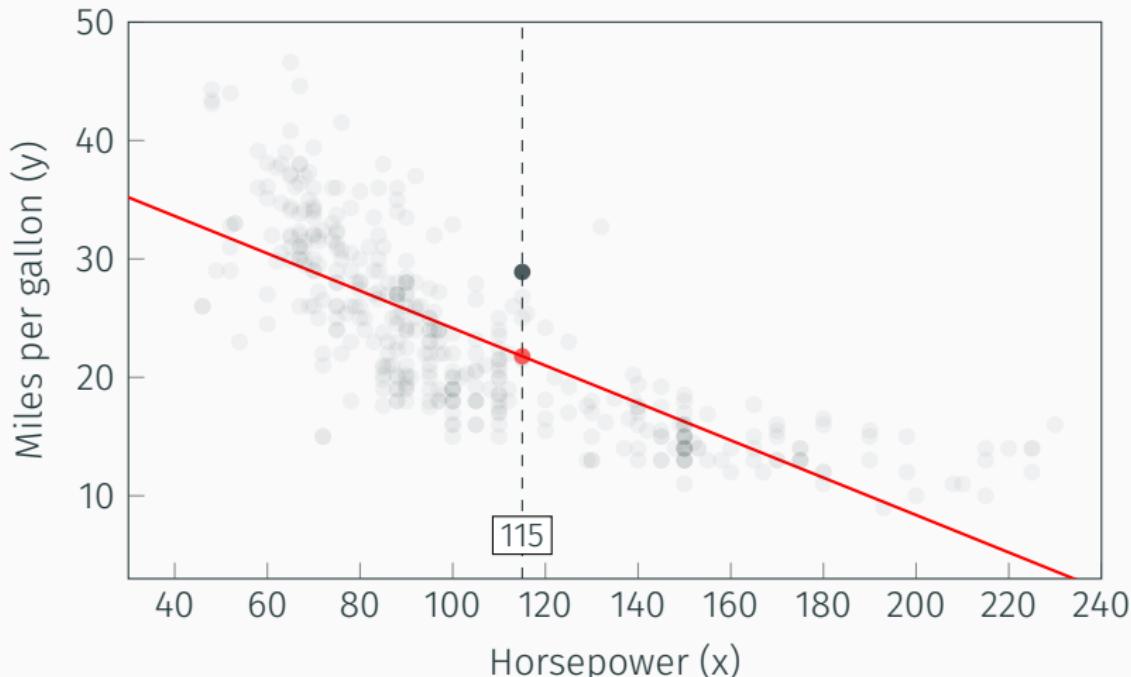
Linear regression (via ordinary least squares)



$$21.78 = 39.93 - 0.1578 * 115$$



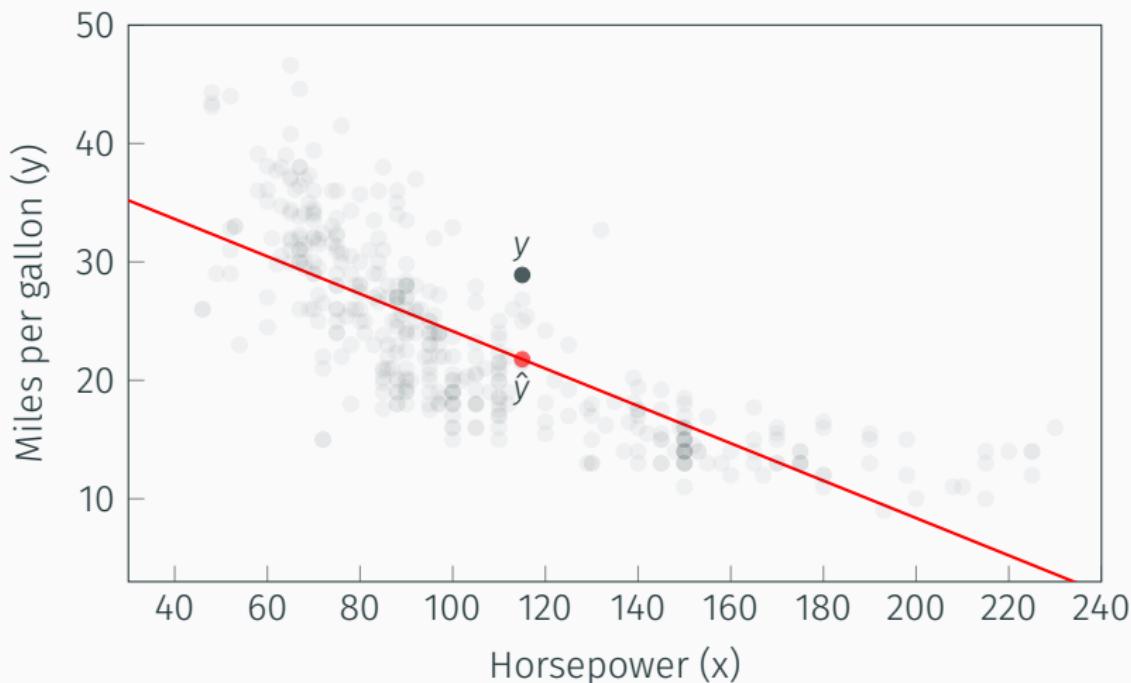
Linear regression (via ordinary least squares)



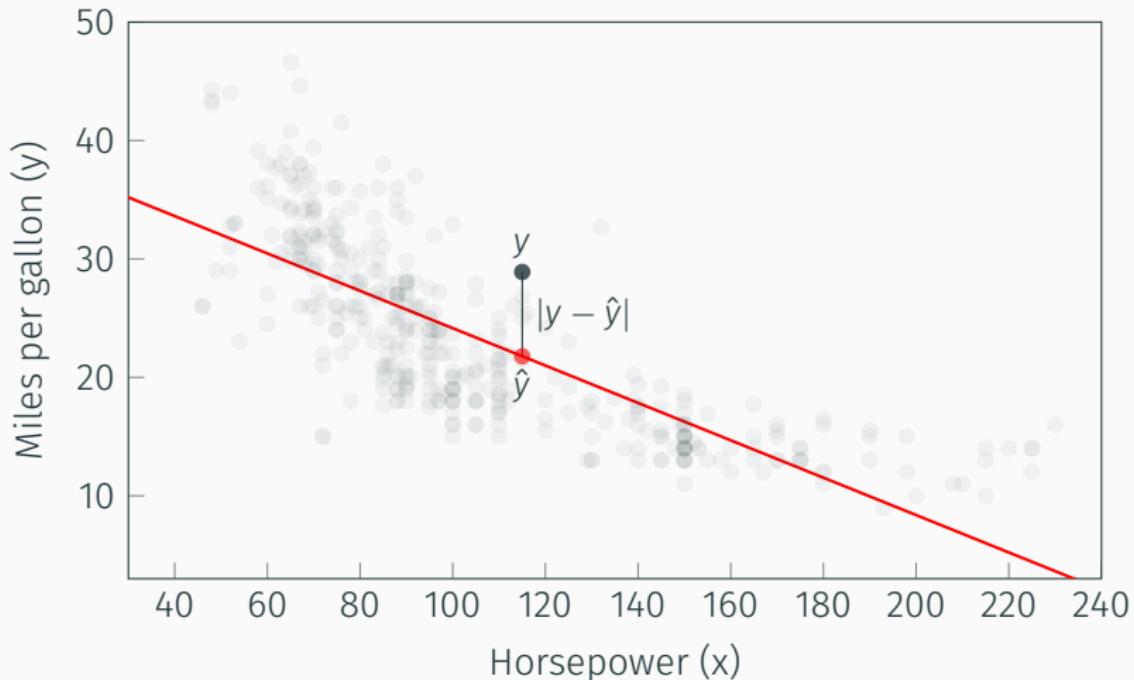
$$21.78 = 39.93 - 0.1578 * 115$$



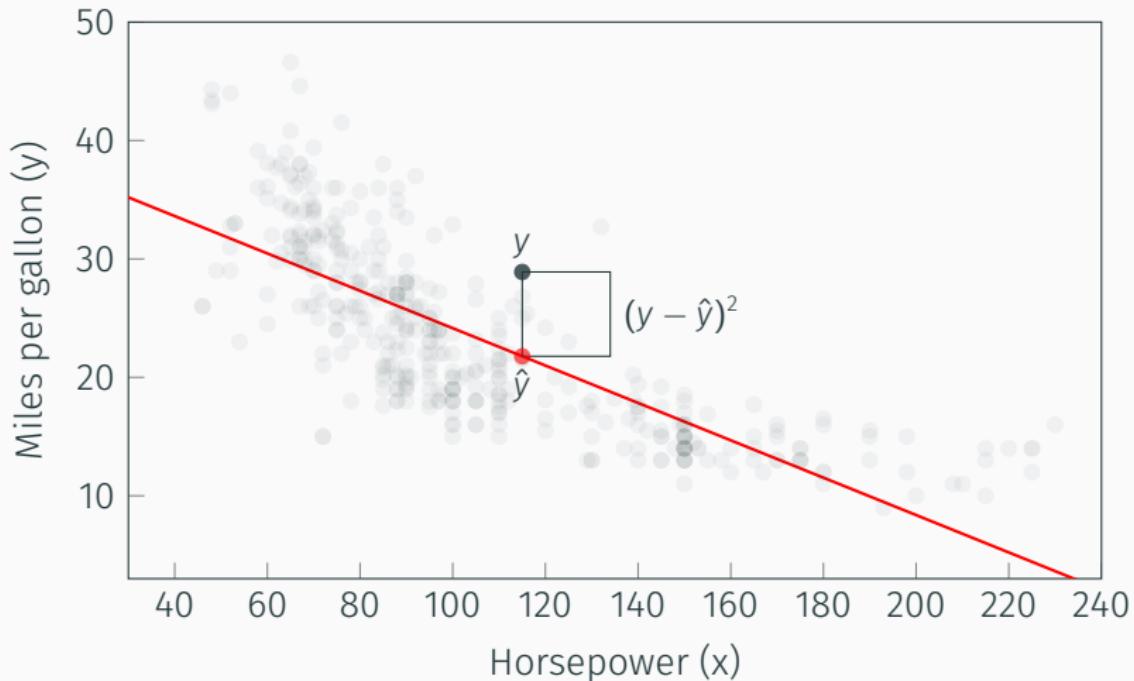
Linear regression (via ordinary least squares)



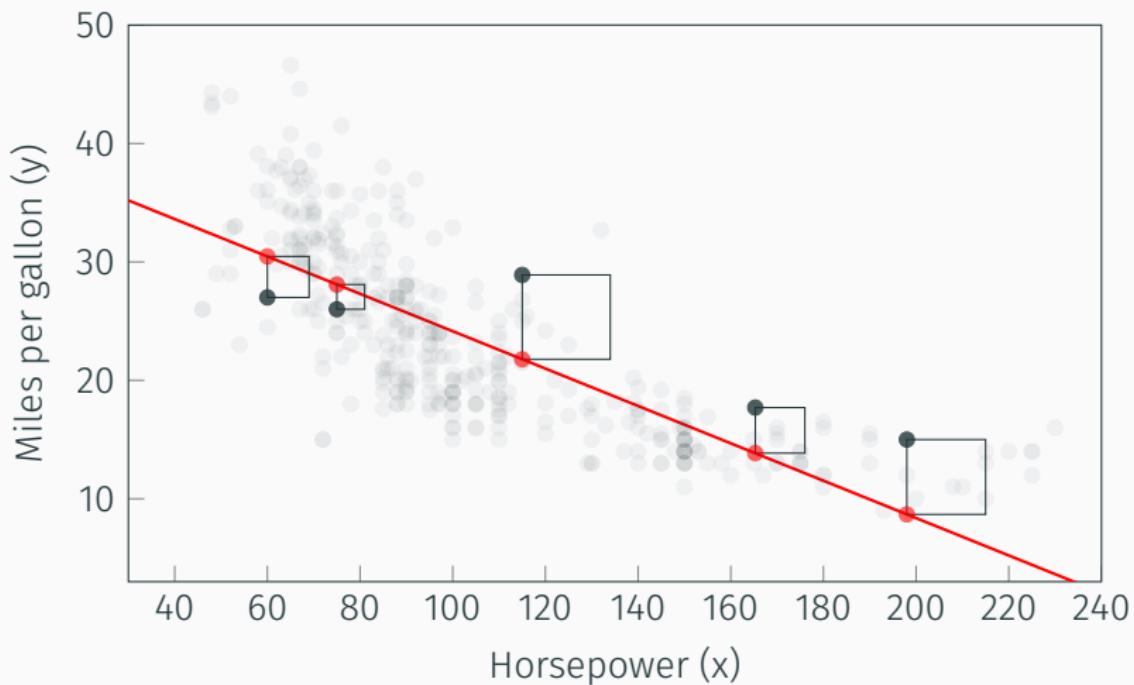
Linear regression (via ordinary least squares)



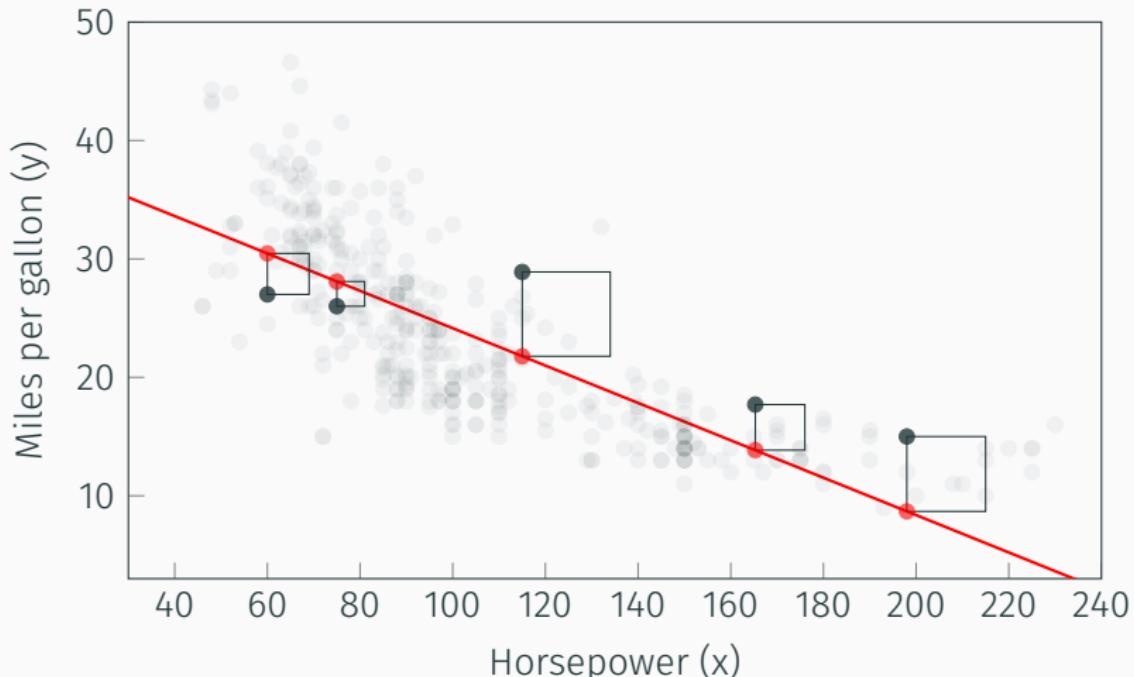
Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



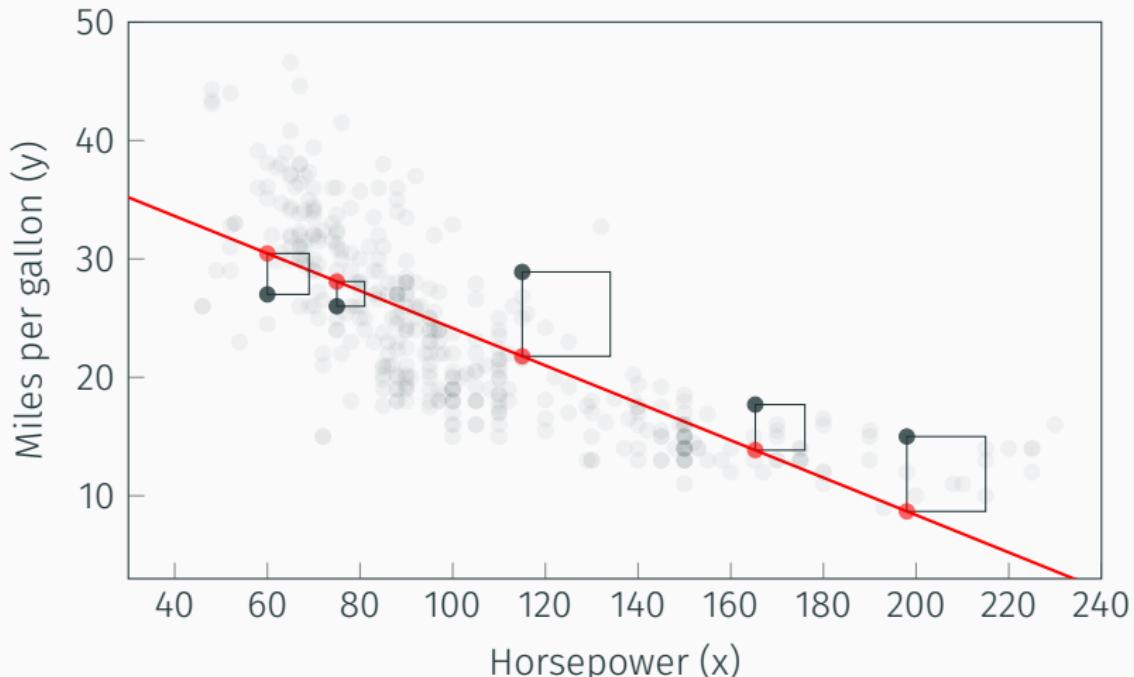
Linear regression (via ordinary least squares)



$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



Linear regression (via ordinary least squares)



$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



Linear regression (via ordinary least squares)

$$\hat{y} = \beta_0 + \beta_1 x$$

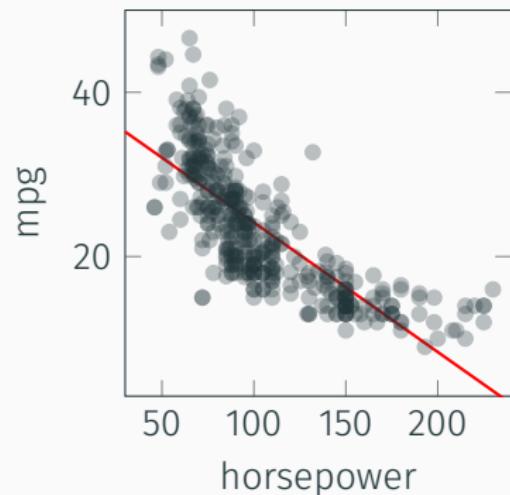


Linear regression (via ordinary least squares)

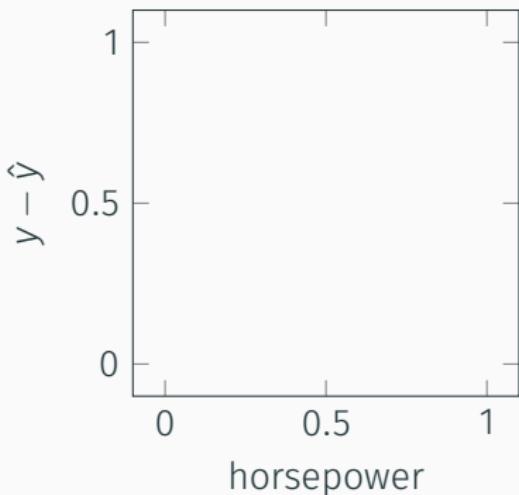
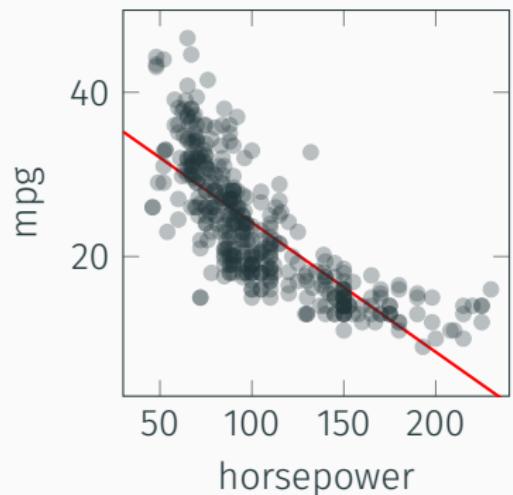
$$\hat{y} = \beta_0 + \beta_1 x$$



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)

$$\hat{y} = \beta_0 + \beta_1 x$$



Linear regression (via ordinary least squares)

$$\hat{y} = \beta_0 + \beta_1 x$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



Linear regression (via ordinary least squares)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \beta_0 + \beta_1 x_i)^2$$



Linear regression (via ordinary least squares)

$$\hat{y} = \beta_0 + \beta_1 x$$

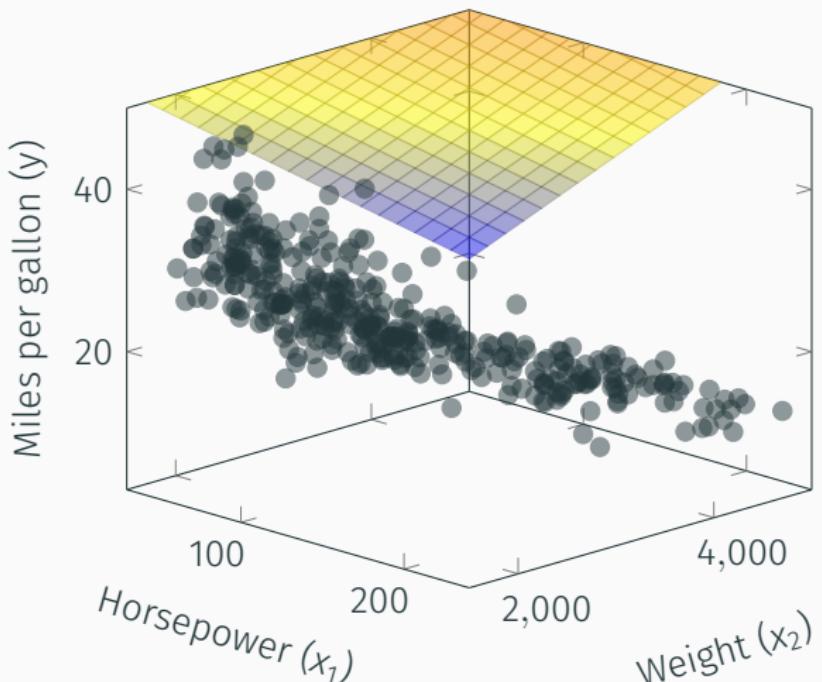


Linear regression (via ordinary least squares)

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)

mpg	manufacturer
36	Chevrolet
15	Ford
25	Chevrolet
26	Chevrolet
17	Ford
15	Ford
32	Chevrolet
14	Ford
14	Ford
28	Chevrolet

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{manufacturer}$$



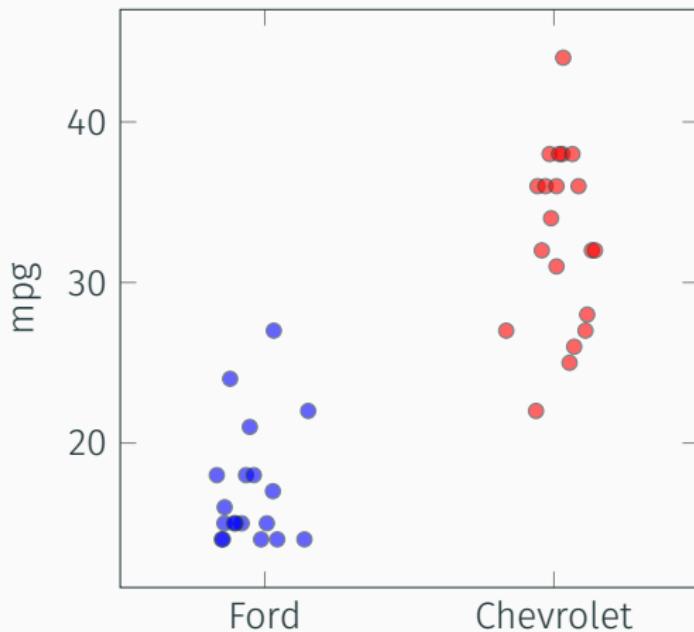
Linear regression (via ordinary least squares)

mpg	manufacturer	chevrolet
36	Chevrolet	1
15	Ford	0
25	Chevrolet	1
26	Chevrolet	1
17	Ford	0
15	Ford	0
32	Chevrolet	1
14	Ford	0
14	Ford	0
28	Chevrolet	1

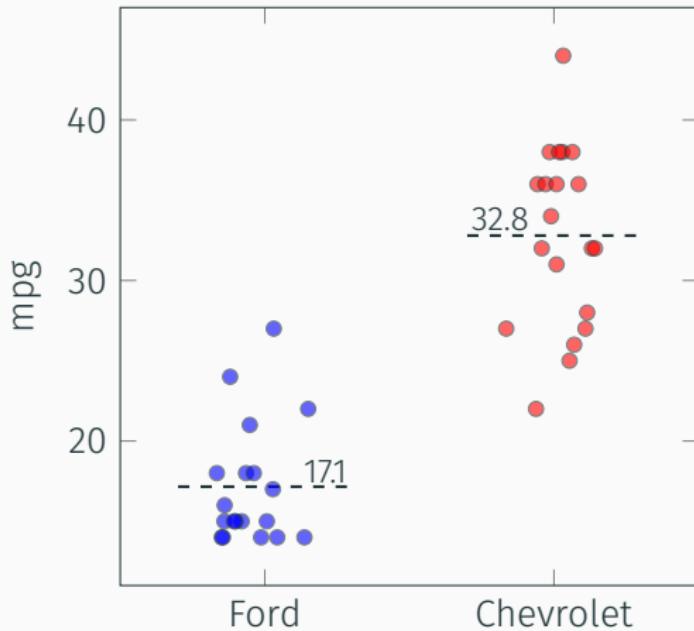
$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet}$$



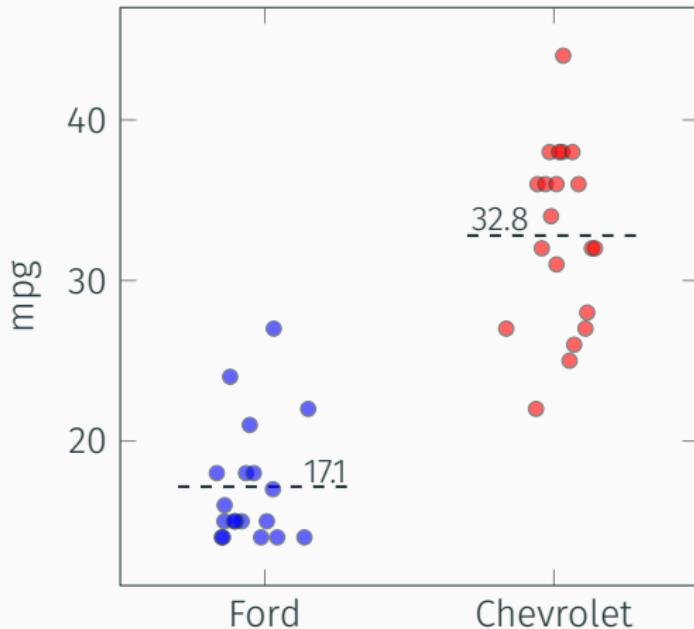
Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Blackboard!



Linear regression (via ordinary least squares)

mpg	manufacturer
36	Chevrolet
15	Ford
25	Chevrolet
26	Pontiac
17	Ford
15	Ford
32	Pontiac
14	Ford
14	Pontiac
28	Chevrolet

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{manufacturer}$$



Linear regression (via ordinary least squares)

mpg	manufacturer	chevrolet	pontiac
36	Chevrolet	1	0
15	Ford	0	0
25	Chevrolet	1	0
26	Pontiac	0	1
17	Ford	0	0
15	Ford	0	0
32	Pontiac	0	1
14	Ford	0	0
14	Pontiac	0	1
28	Chevrolet	1	0

$$\widehat{\text{mpg}} = \beta_0 + \beta_1 \times \text{chevrolet} + \beta_2 \times \text{pontiac}$$



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



k nearest neighbours



Logistic regression



Generative models

