PSY9511: Seminar 4

The basics of regression and classification

Esten H. Leonardsen

29.05.2024

UNIVERSITETET
I OSLO

Recap

What is statistical learning?

Recap

What is statistical learning?

Inferentiental 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?

Inferentiental 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.

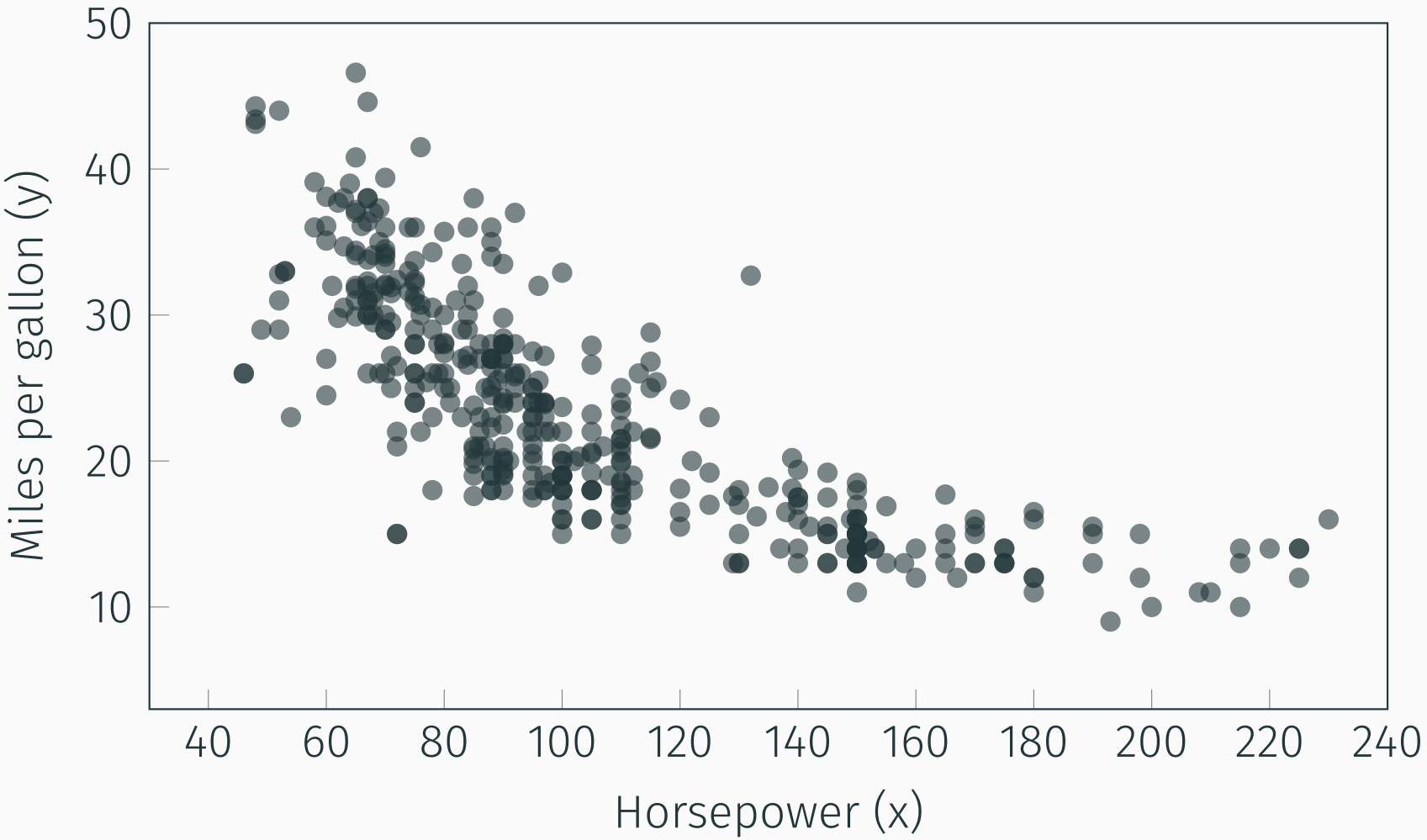
Recap

What is statistical learning?

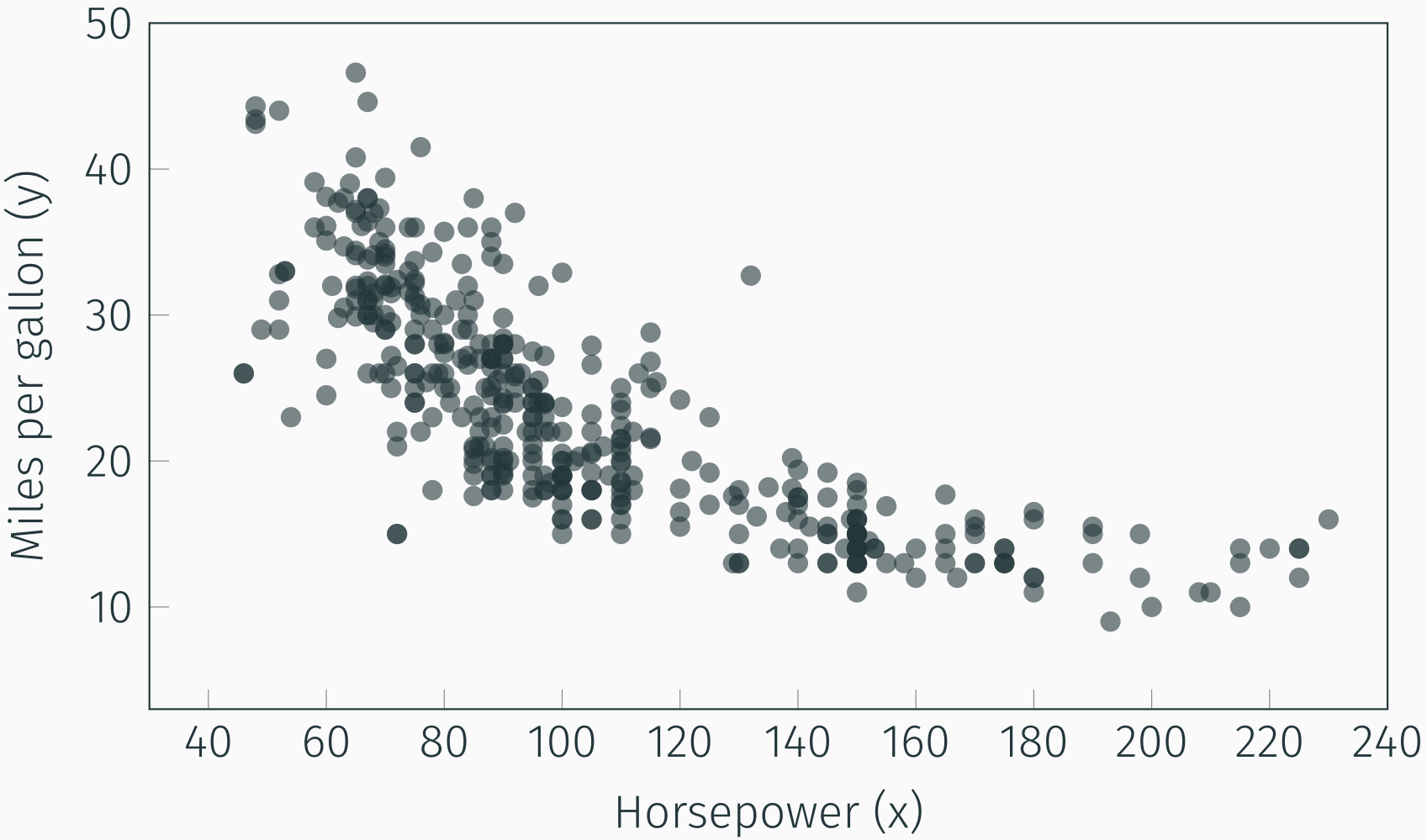
Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.

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

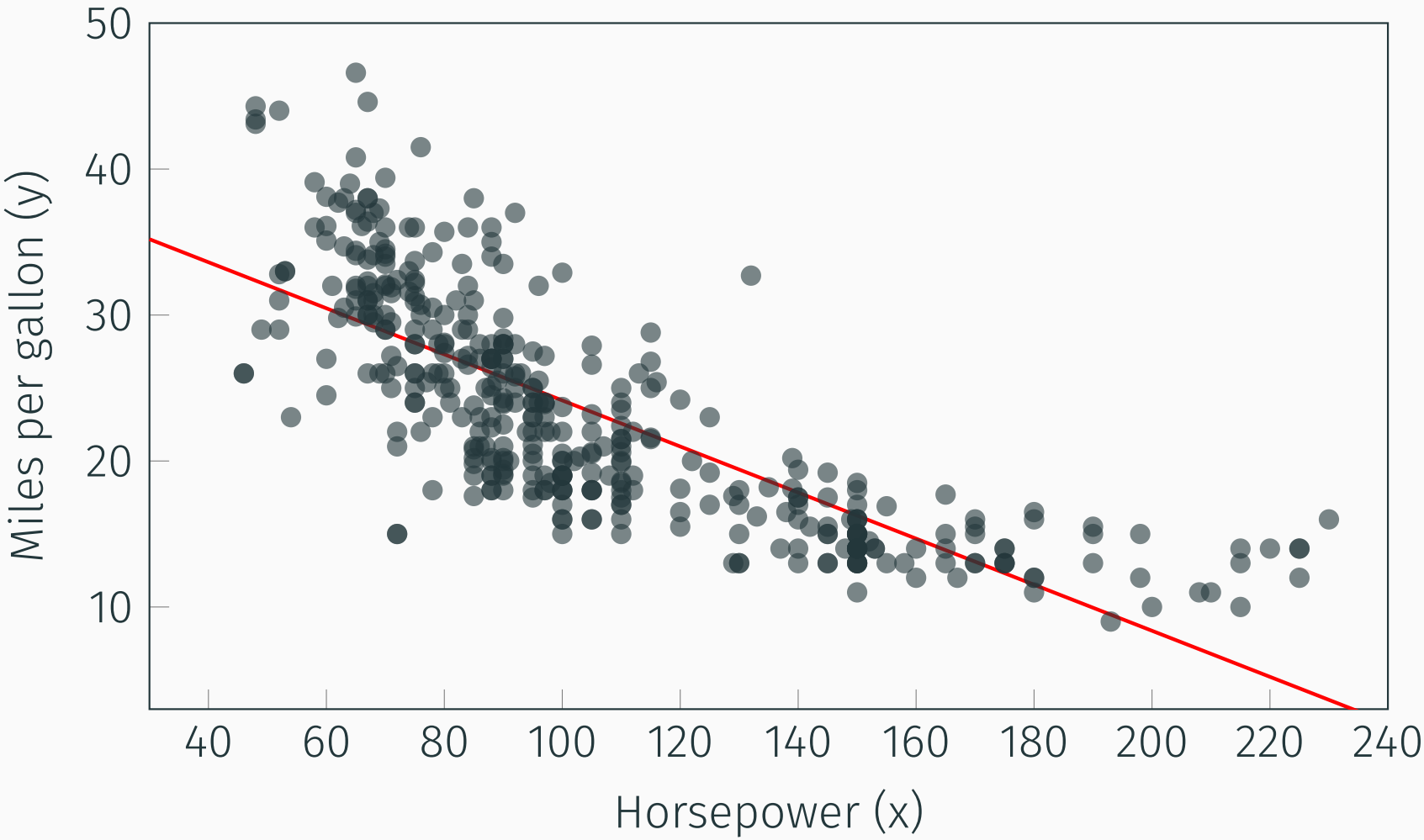

Miles per gallon (y)
50
40
30
20
10
40
60
80
100
120
140
160
180
200
220
240
Horsepower (x)

Recap



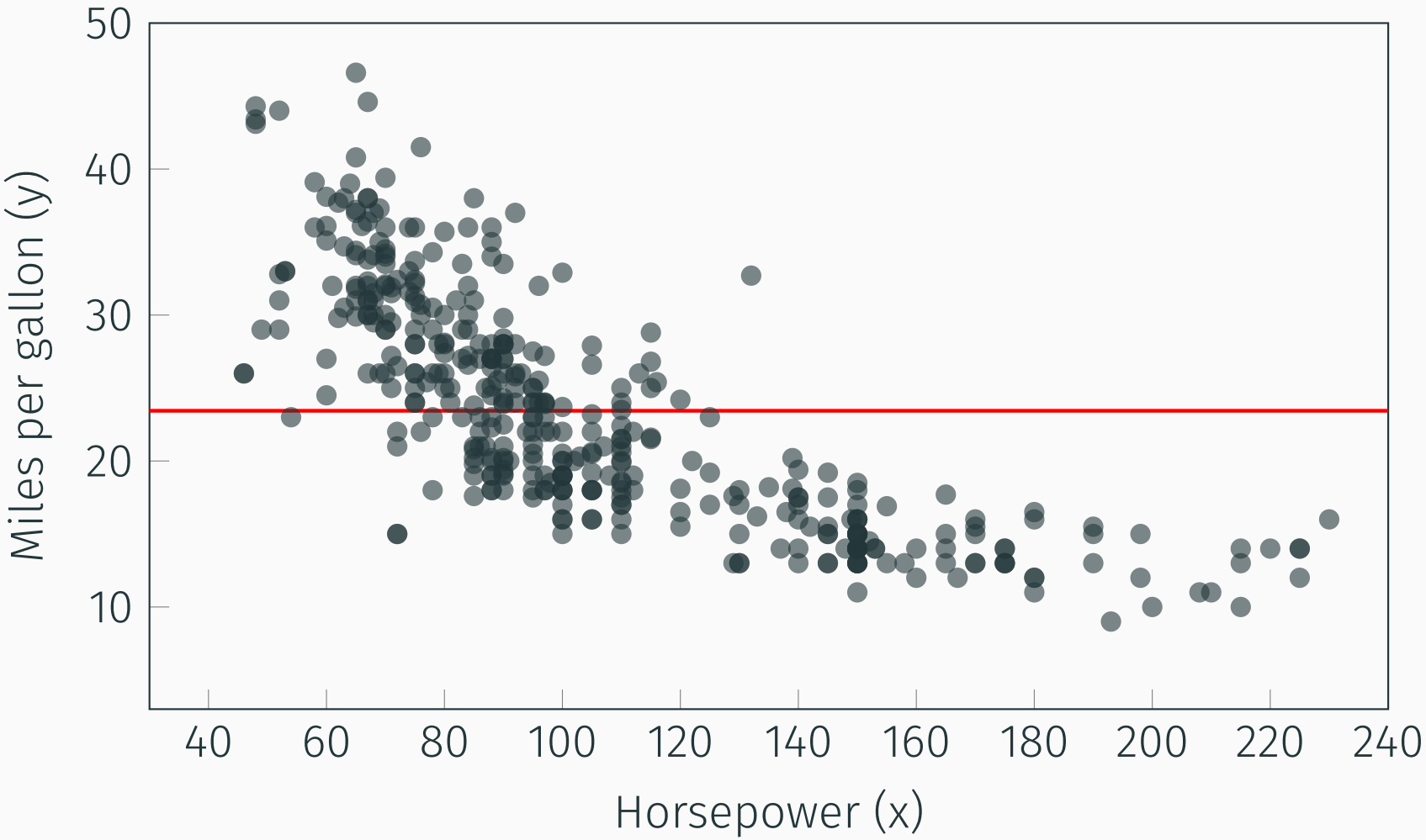
$$\hat{y} = \hat{f}(x)$$

Recap



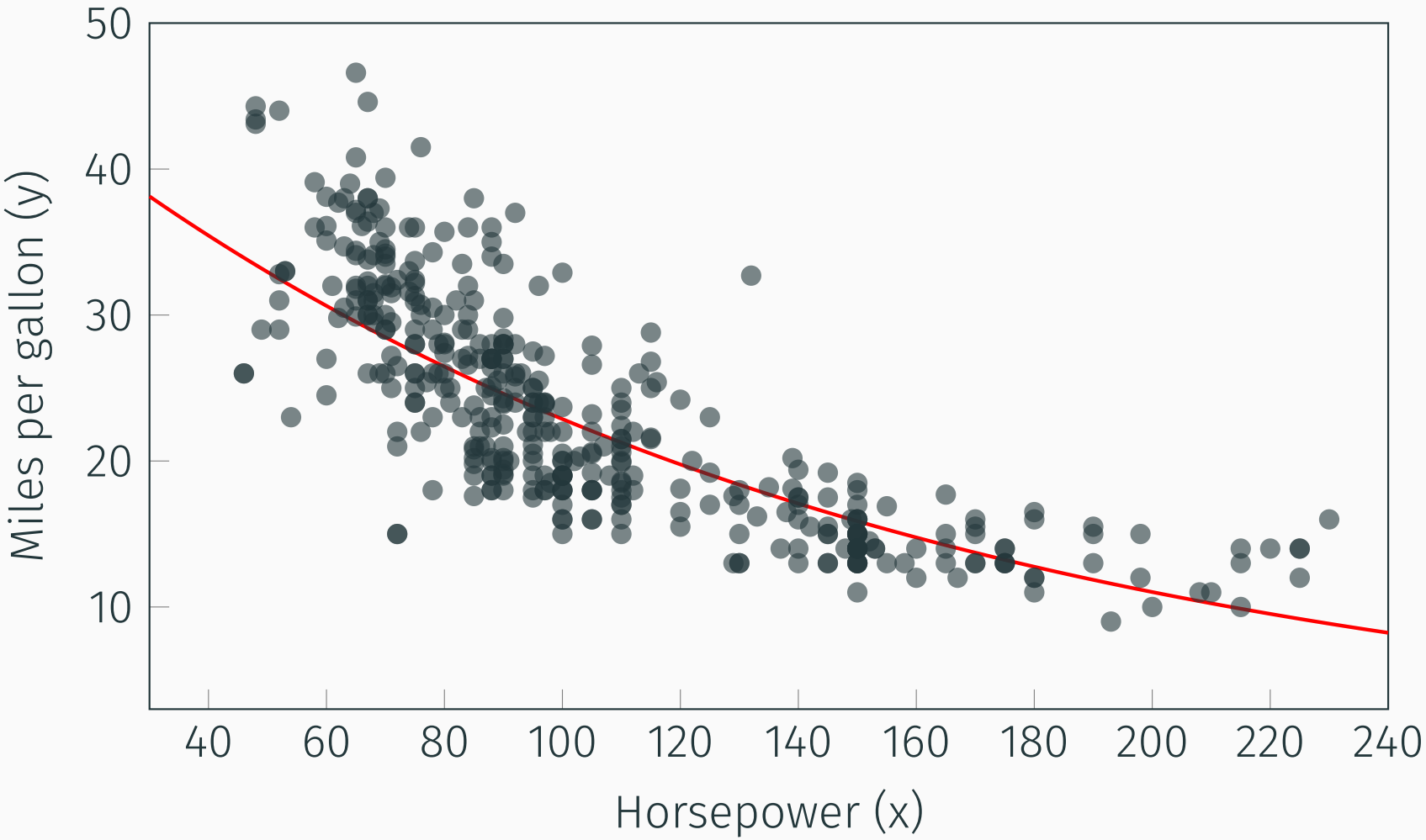
$$\hat{y} = 39.93 - 0.1578x$$

Recap



$$\hat{y} = 23.44$$

Recap



$$\hat{y} = e^{3.86 - 0.0073x}$$

Outline

Plan for the day:

Outline

Plan for the day:

- Different types of outputs y: Regression vs classification

Outline

Plan for the day:

- Different types of outputs y: Regression vs classification
- Simple solutions to regression problems

Outline

Plan for the day:

- Different types of outputs y: Regression vs classification
- Simple solutions to regression problems
  - Linear regression
  - k nearest neighbours
- Finding $\hat{f}(X)$: Training machine learning models

Outline

Plan for the day:

- Different types of outputs y: Regression vs classification
- Simple solutions to regression problems
  - Linear regression
  - k nearest neighbours
- Finding $\hat{f}(X)$: Training machine learning models
- Simple solutions to classification problems

Outline

Plan for the day:

- Different types of outputs y: Regression vs classification
- Simple solutions to regression problems
  - Linear regression
  - k nearest neighbours
- Finding $\hat{f}(X)$: Training machine learning models
- Simple solutions to classification problems
  - Logistic regression
  - Generative models

Outline

Plan for the day:

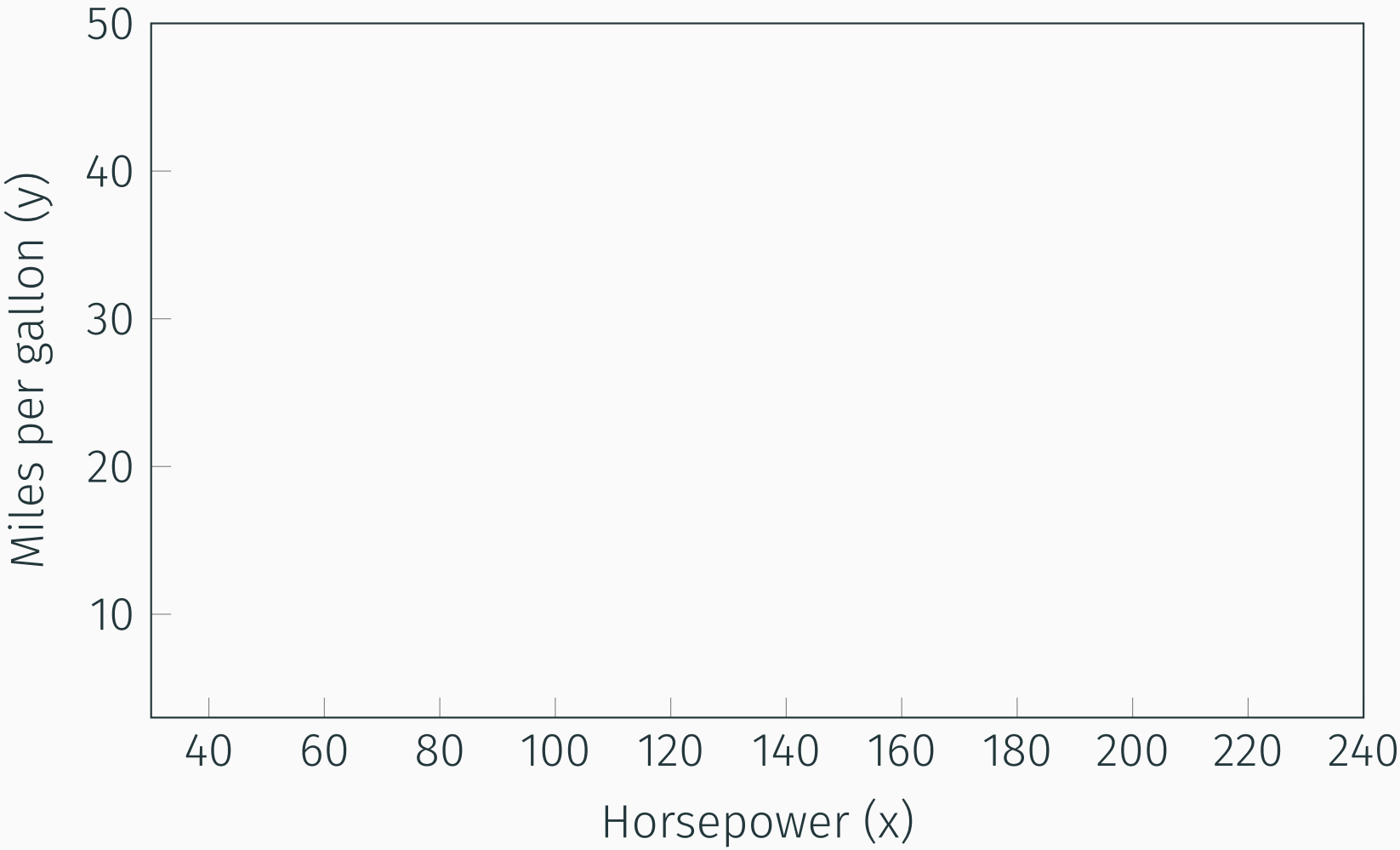
- Different types of outputs y: Regression vs classification
- Simple solutions to regression problems
  - Linear regression
  - k nearest neighbours
- Finding $\hat{f}(X)$: Training machine learning models
- Simple solutions to classification problems
  - Logistic regression
  - Generative models

Plan for future lectures:

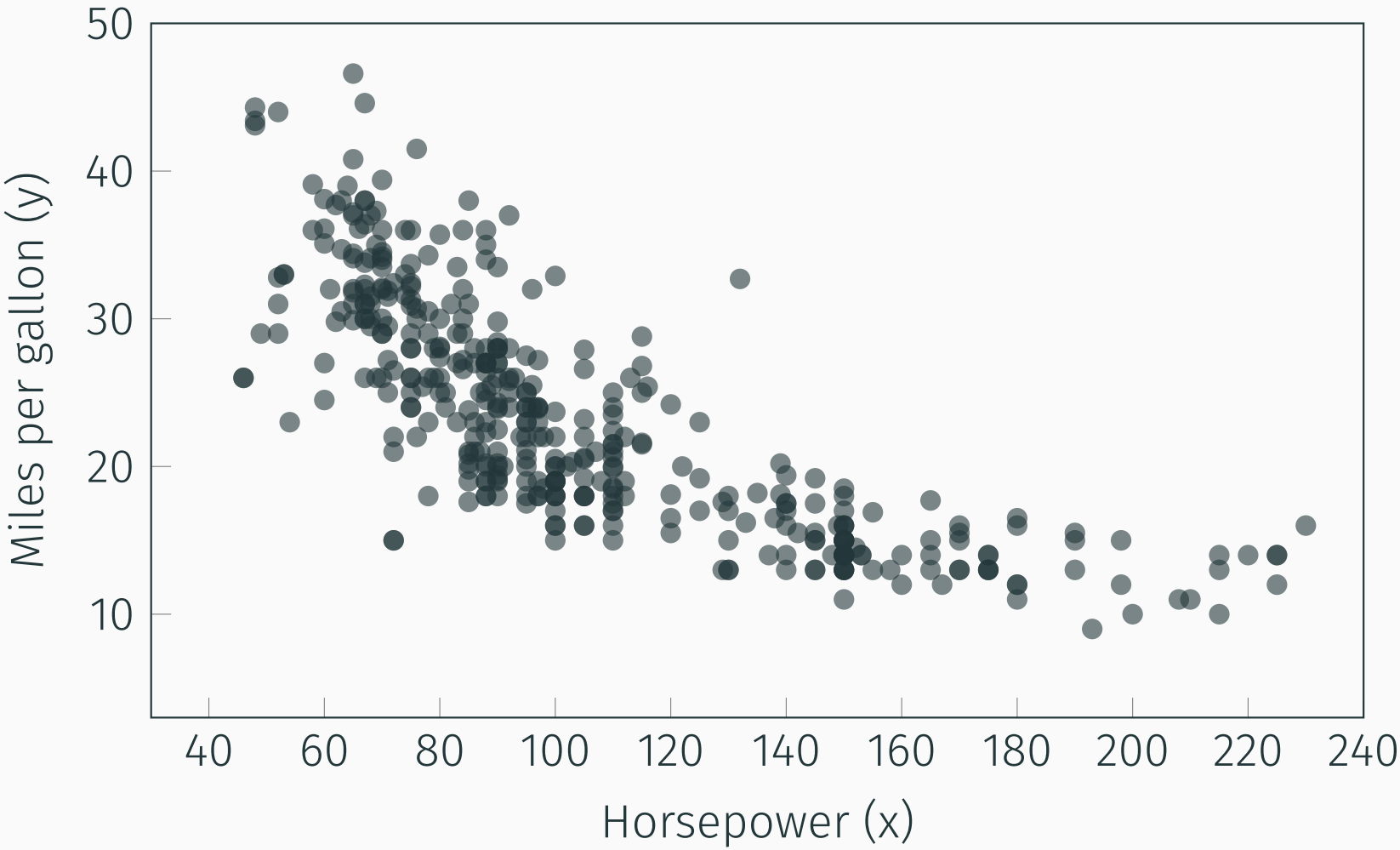
- How do we evaluate how good our models are? (Lecture 3)
- Complex solutions to regression and classification problems (Lecture 4 and onwards)

Regression vs classification

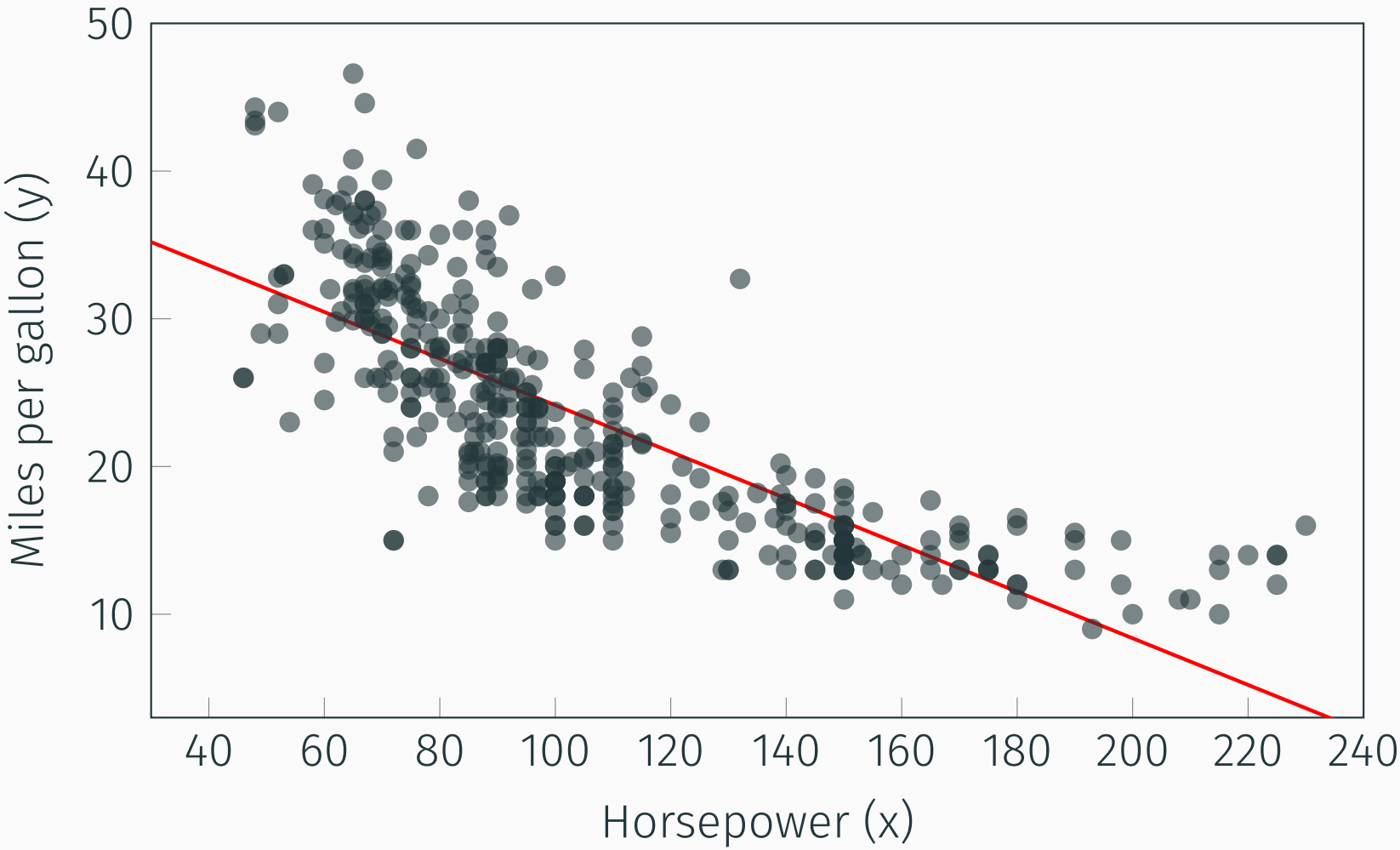
Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)

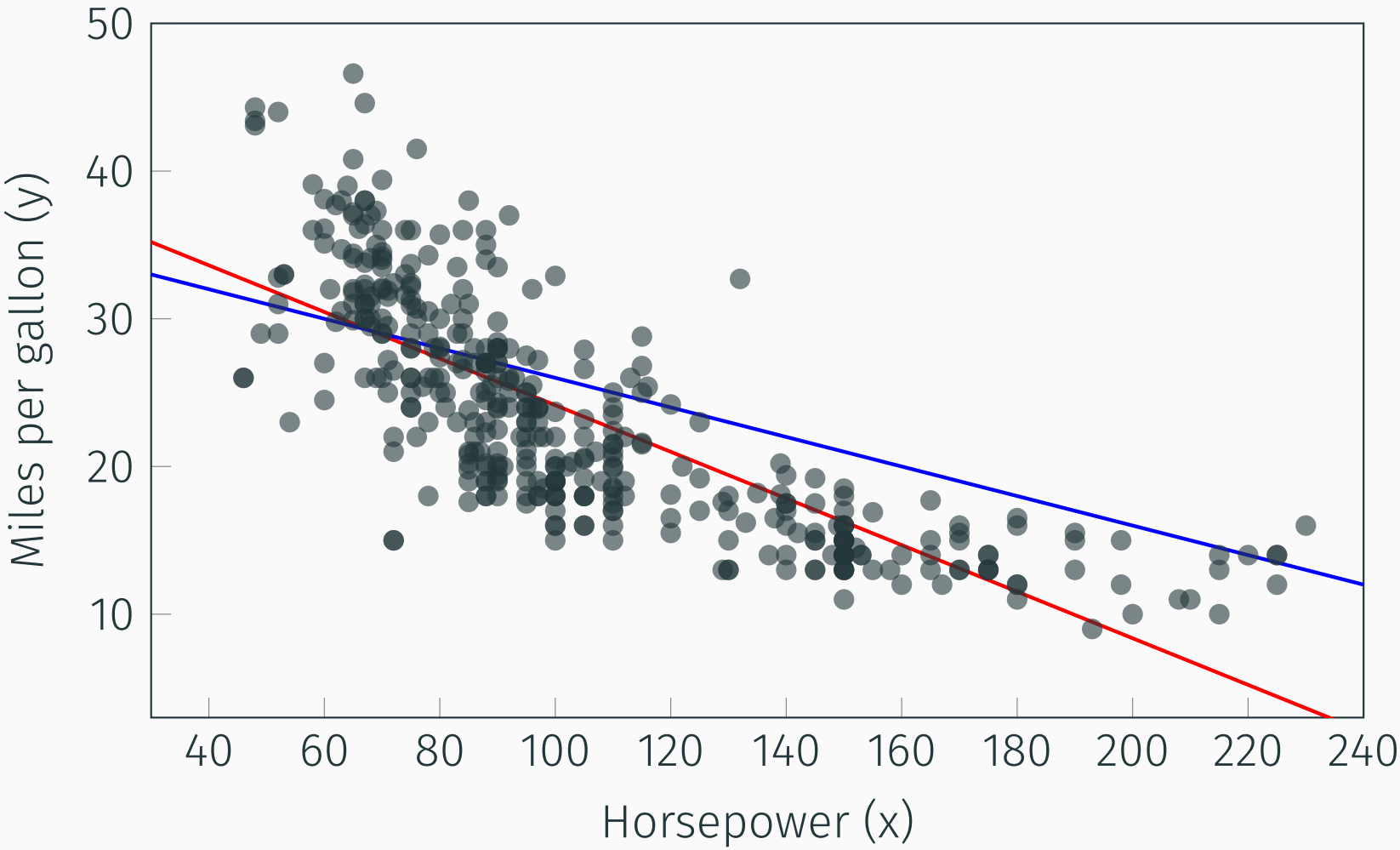


Linear regression (via ordinary least squares)



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

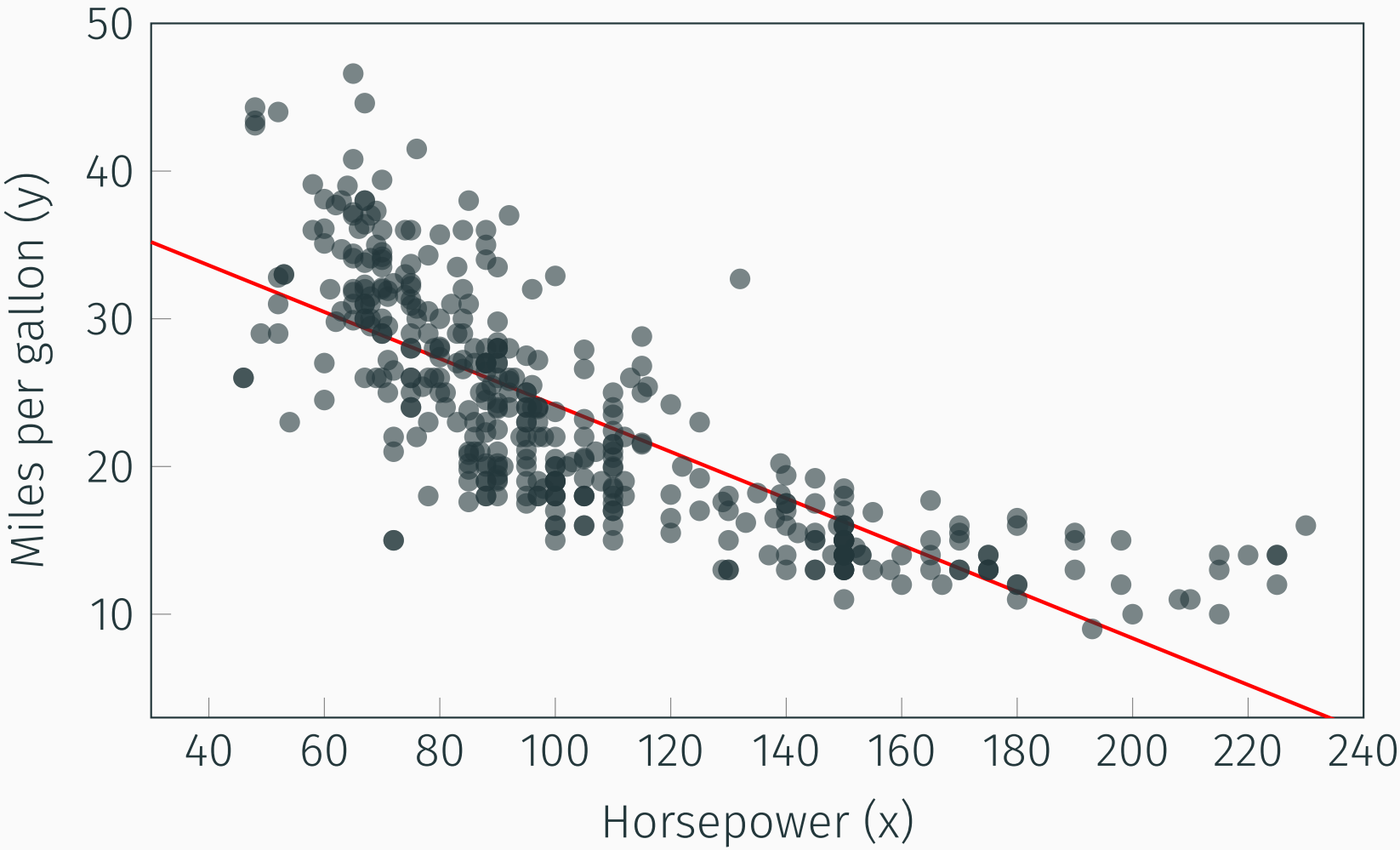
Linear regression (via ordinary least squares)



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

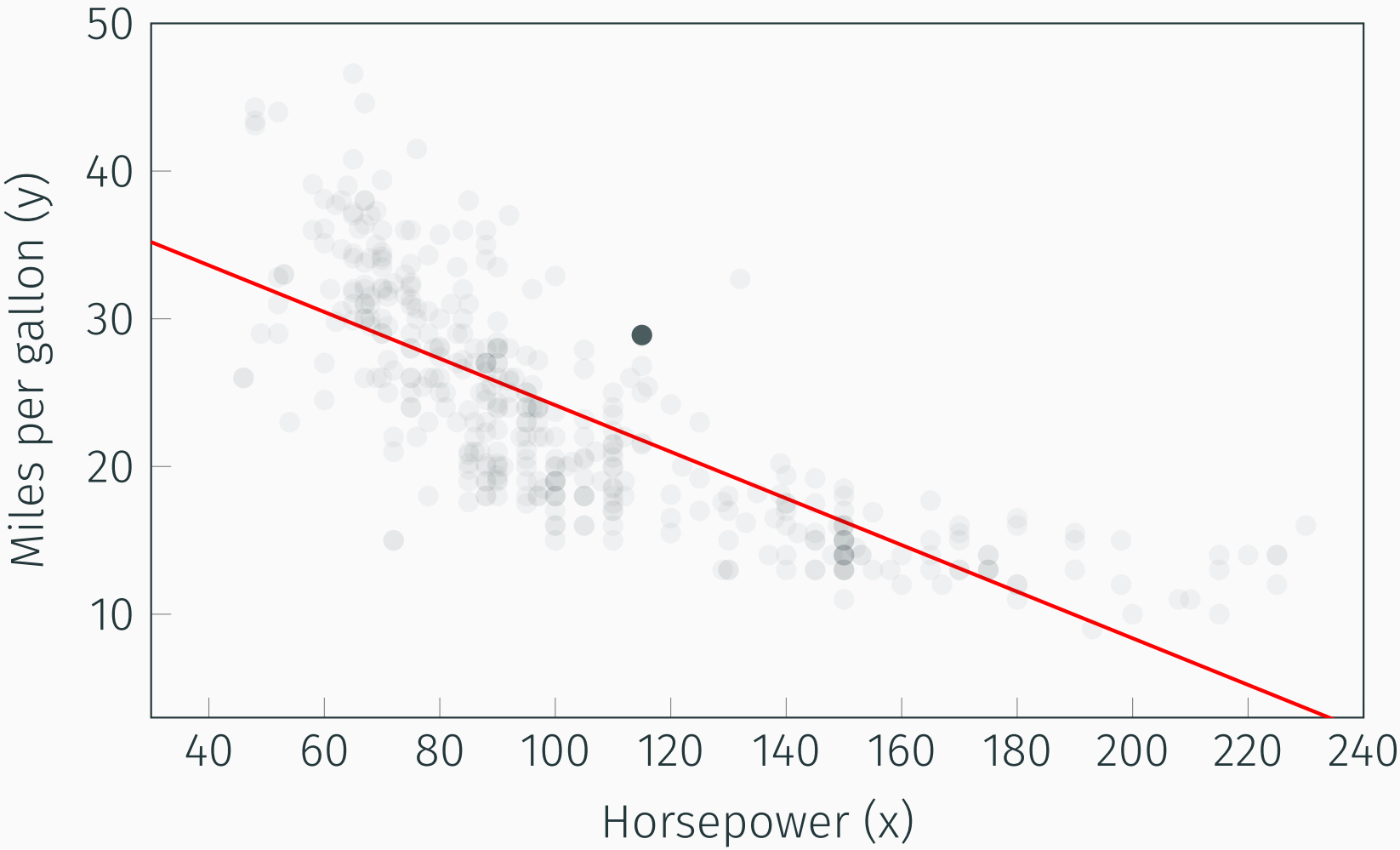
$$\hat{y} = 36 - 0.1x$$

Linear regression (via ordinary least squares)



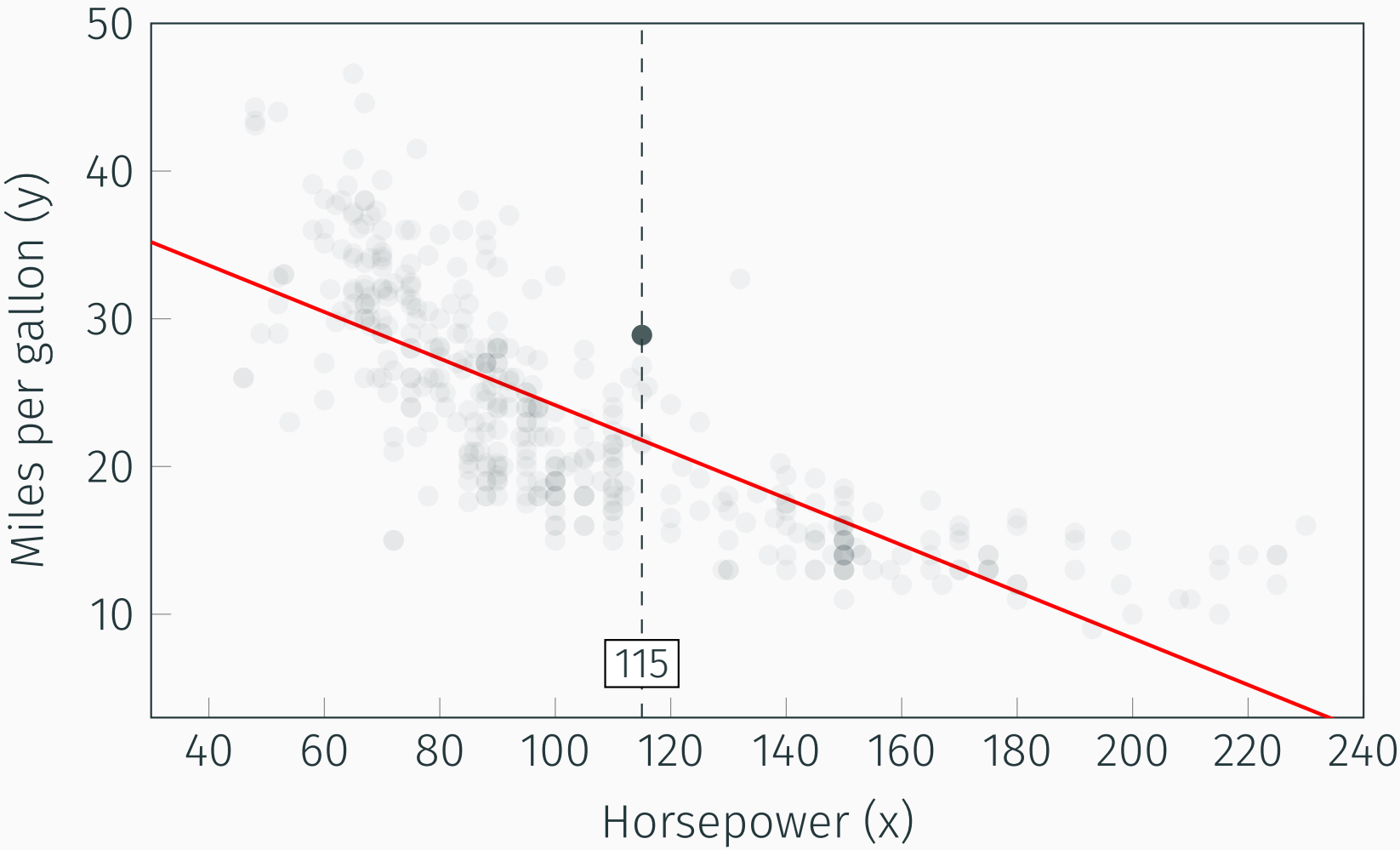
$$\hat{y} = 39.93 - 0.1578x$$

Linear regression (via ordinary least squares)



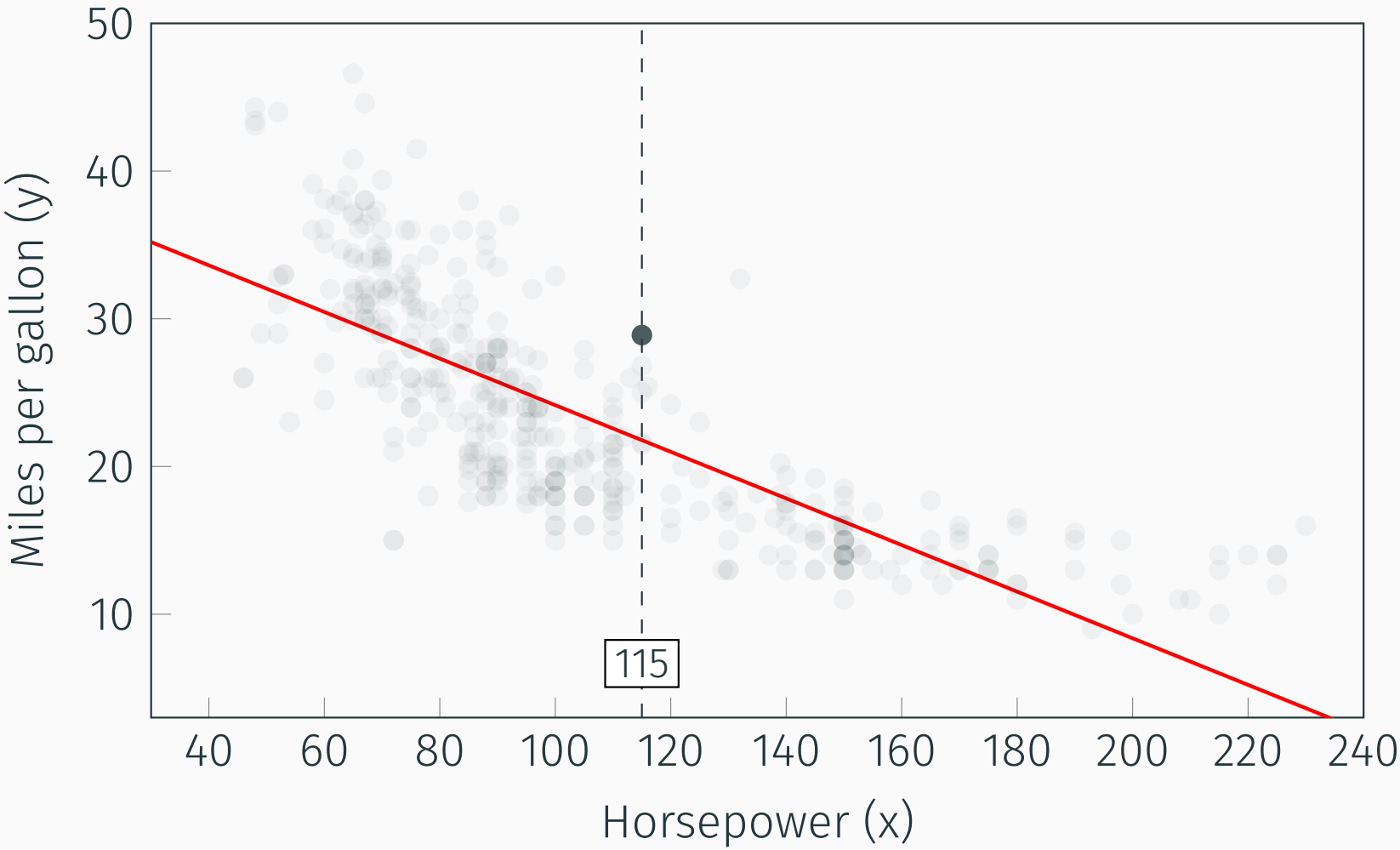
$$\hat{y} = 39.93 - 0.1578x$$

Linear regression (via ordinary least squares)



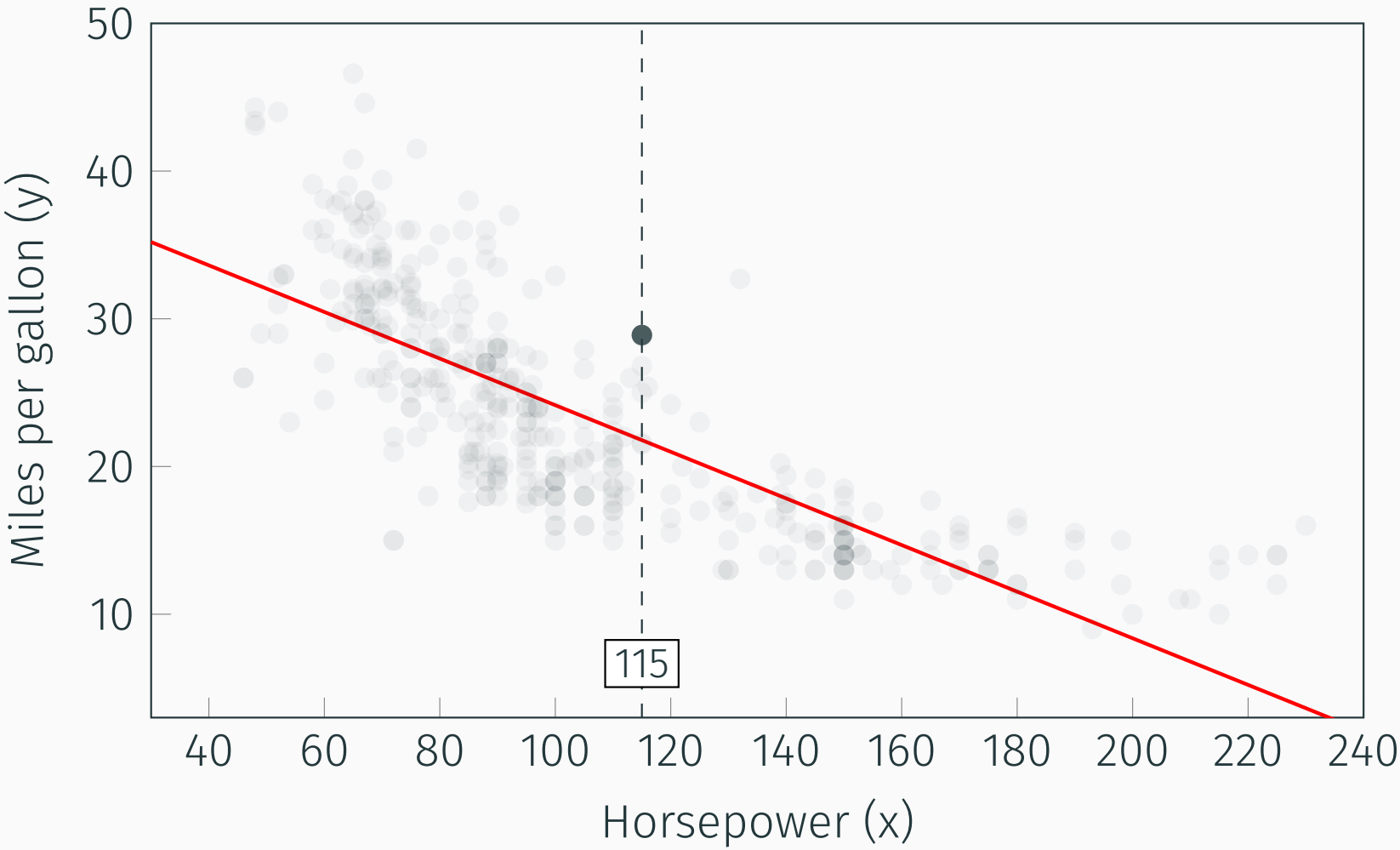
$$\hat{y} = 39.93 - 0.1578x$$

Linear regression (via ordinary least squares)



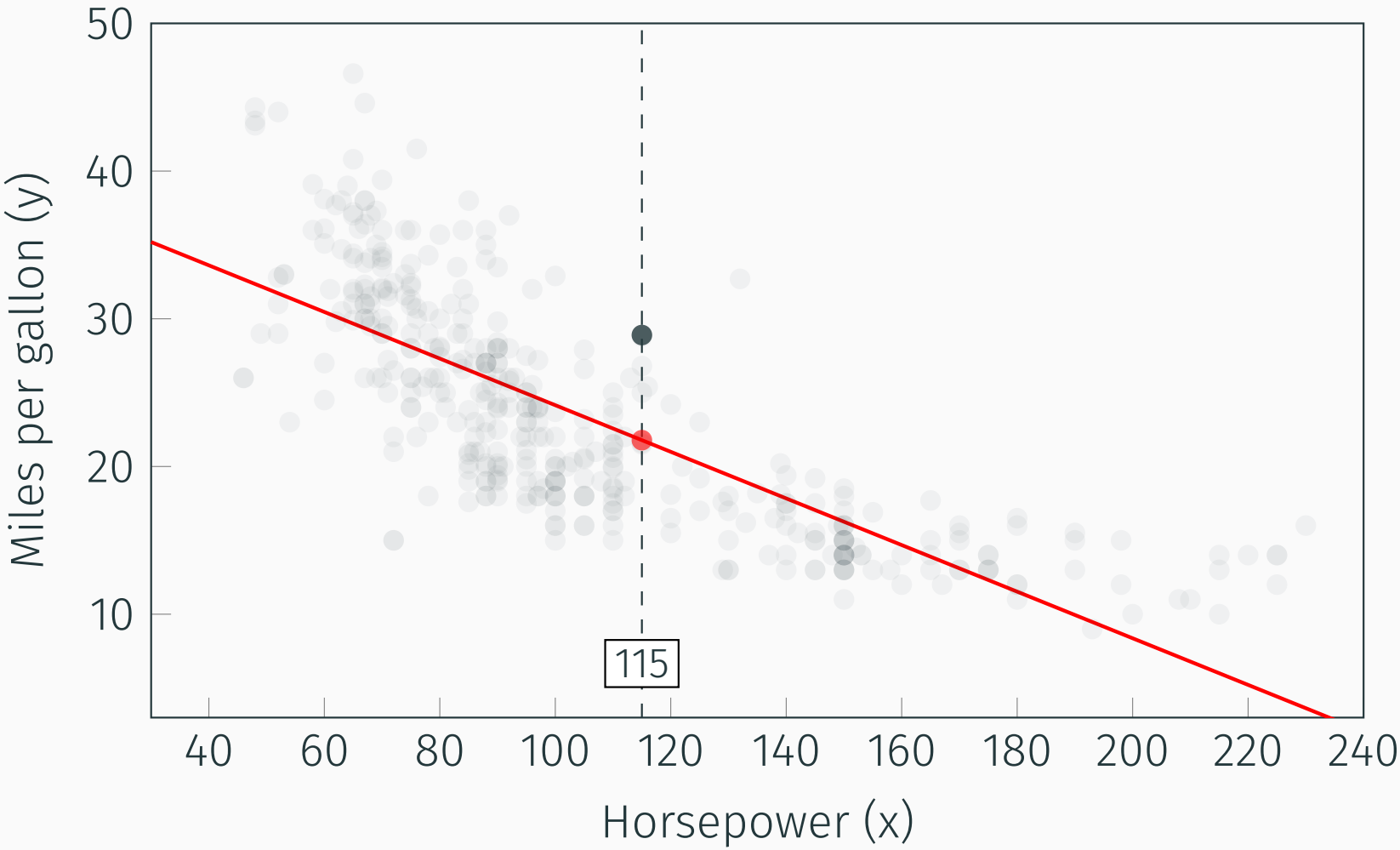
$$\hat{y} = 39.93 - 0.1578 * 115$$

Linear regression (via ordinary least squares)



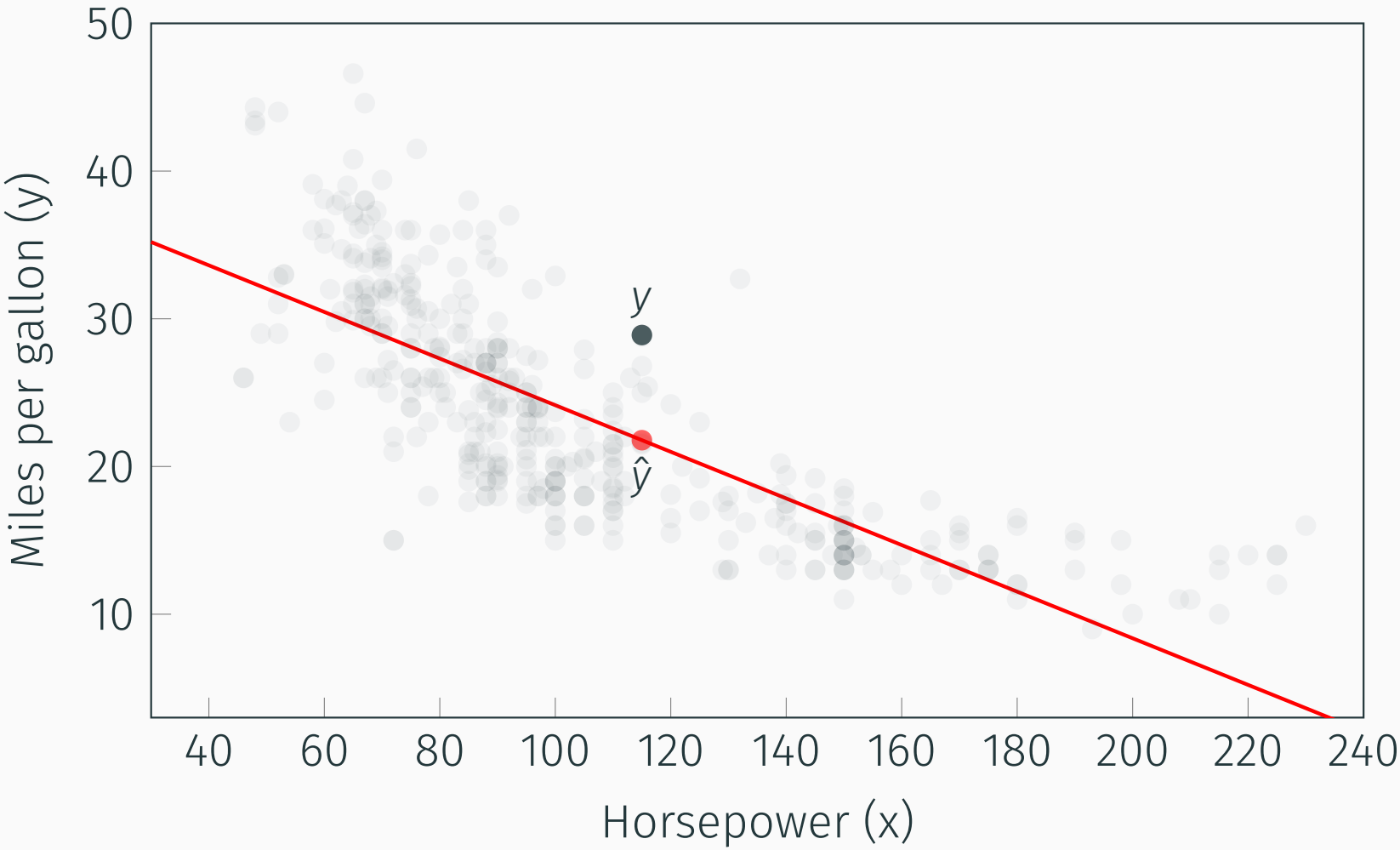
$$21.78 = 39.93 - 0.1578 * 115$$

Linear regression (via ordinary least squares)

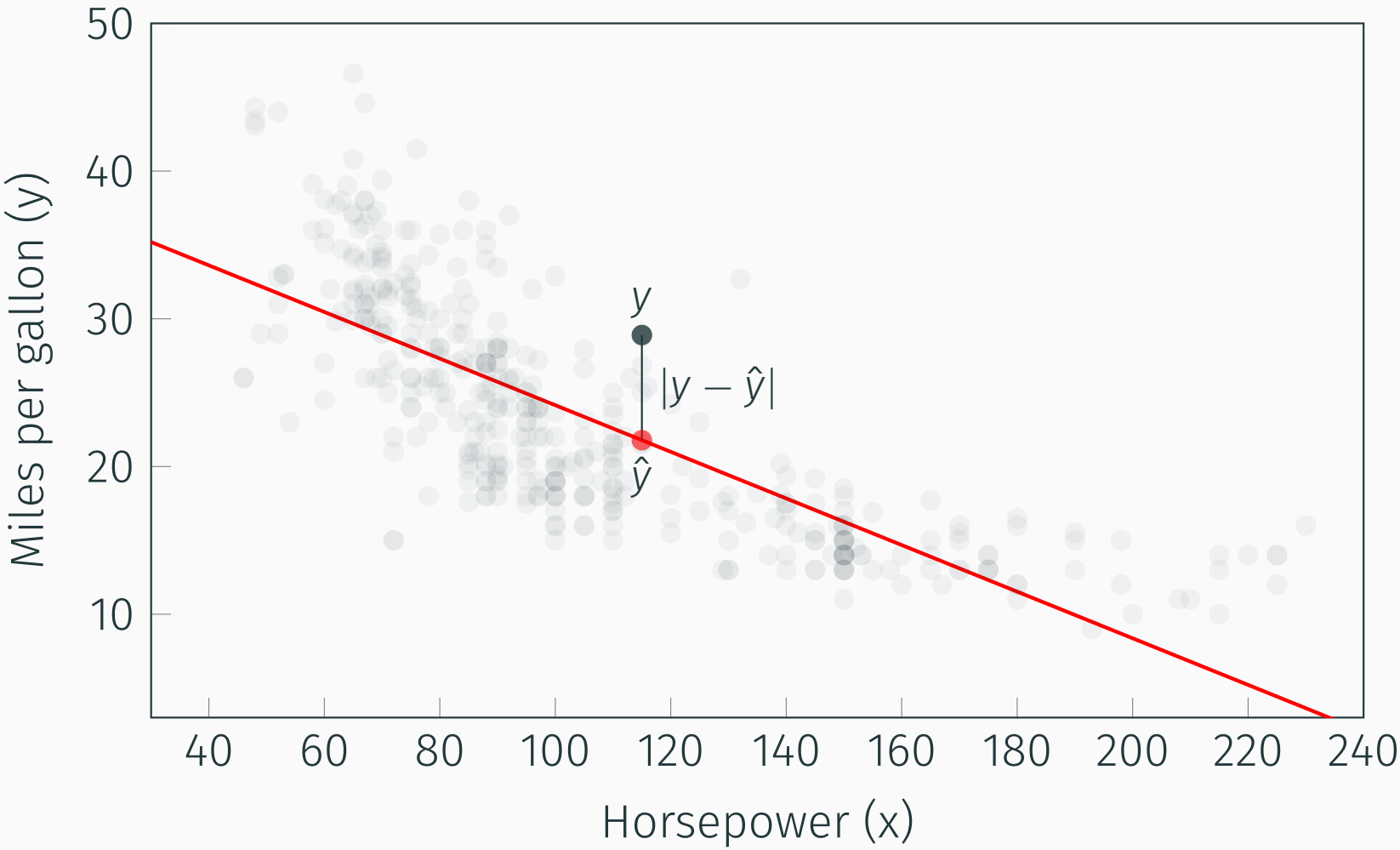


$$21.78 = 39.93 - 0.1578 * 115$$

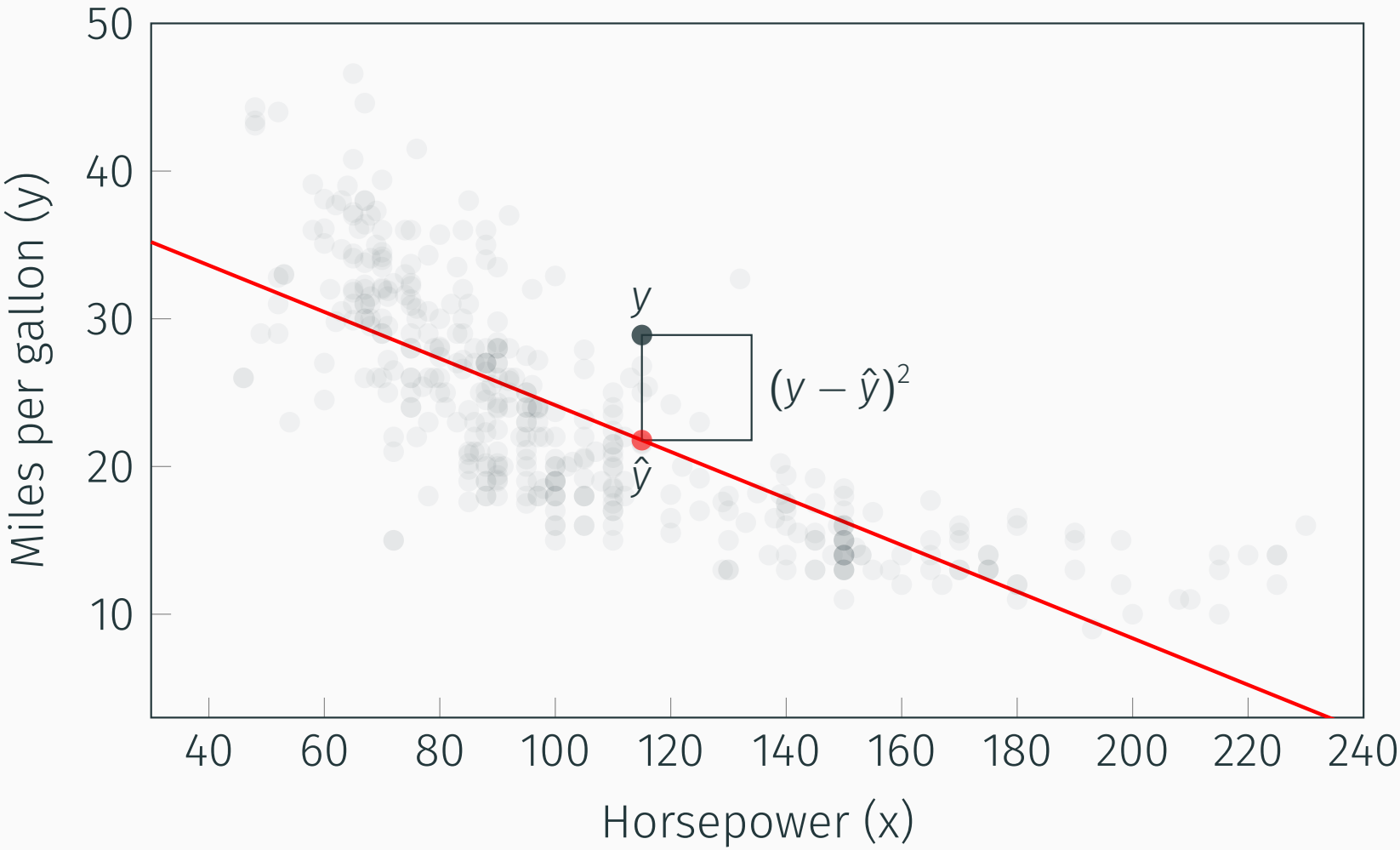
Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)



Linear regression (via ordinary least squares)