

This class is being conducted over Zoom. As the instructor, I will be **recording** this session. I have disabled the recording feature for others so that no one else will be able to record this session. I will be posting this session to the course's website.

If you have privacy concerns and **do not wish to appear in the recording**, you may turn video off (click "**stop video**") so that Zoom does not record you.

The chat box is always open for discussion and questions to the entire class. You may also send messages privately to the instructor or the TAs. Please note that Zoom saves all chat transcripts.

I create a live transcription of each session using **Otter.ai**. This means that Otter.ai will transcribe anything spoken over the Zoom audio. The transcript will be posted with the session video on the course website.

Regression

Stats 7

Mary Ryan

Aug. 11, 2020



Course website:

<https://canvas.eee.uci.edu/courses/28451>



Slides can be found at:

<https://maryryan.github.io/stats7-SS2-2020-slides/stats7-SS2-2020-regression/stats7-SS2-2020-regression>

Learning Objectives

By the end of today's lecture you should be able to:

- understand how to describe the relationship between two variables
- interpret the intercept and slope(s) of a regression model
- diagnose the appropriateness of a regression model

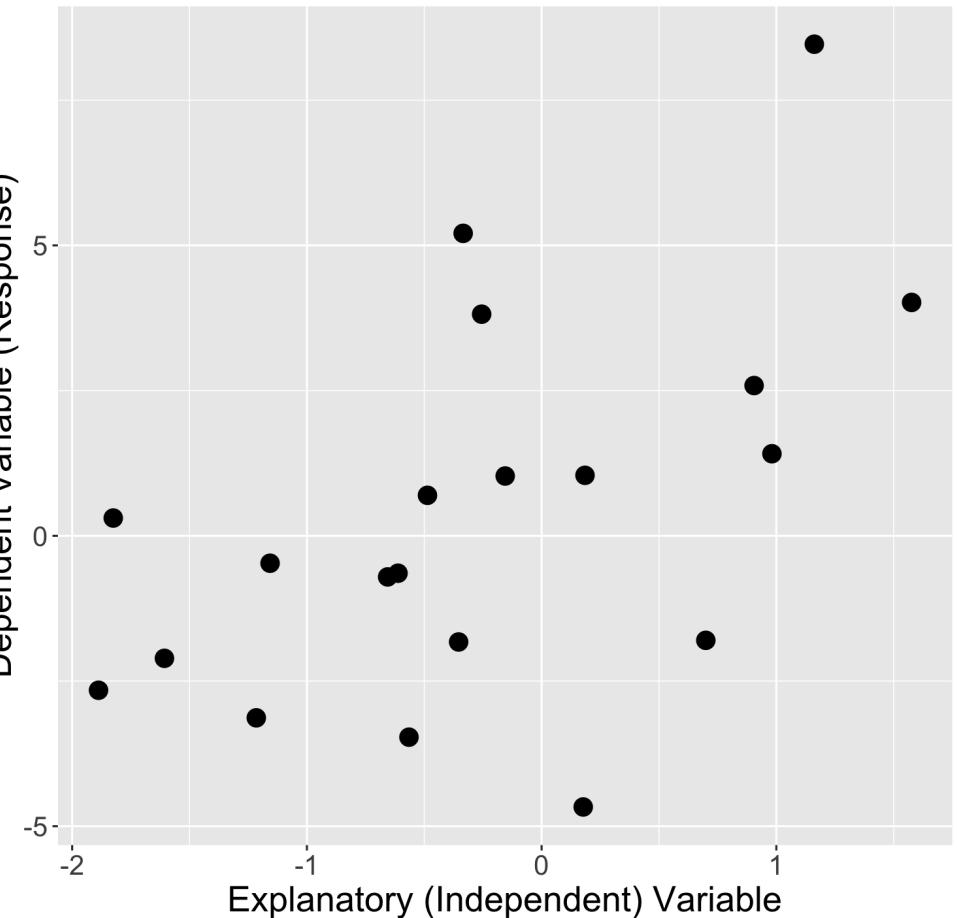
Studies

- The goal is to explore an association or relationship between two variables
 - There may be a relationship or there may not!
- How do we assess whether there is a relationship?

Scatterplots

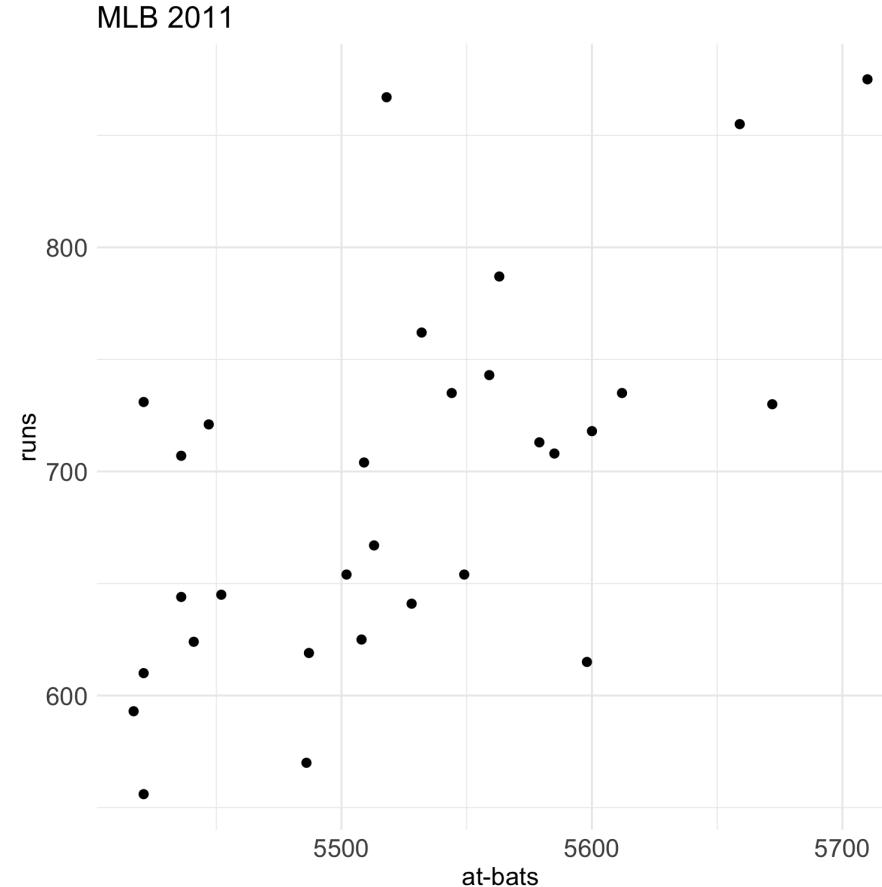
- Used to **visualize** relationship between 2 variables
- **Explanatory/Indendent variable** goes on x-axis
- **Dependent variable/response** goes on y-axis
- **Are x and y related?**
 - Shape (line, curve, mass of points)
 - Direction (positive, negative, horizontal, vertical)
 - Strength (defined/tight or scattered)
 - Outliers?

Scatterplot



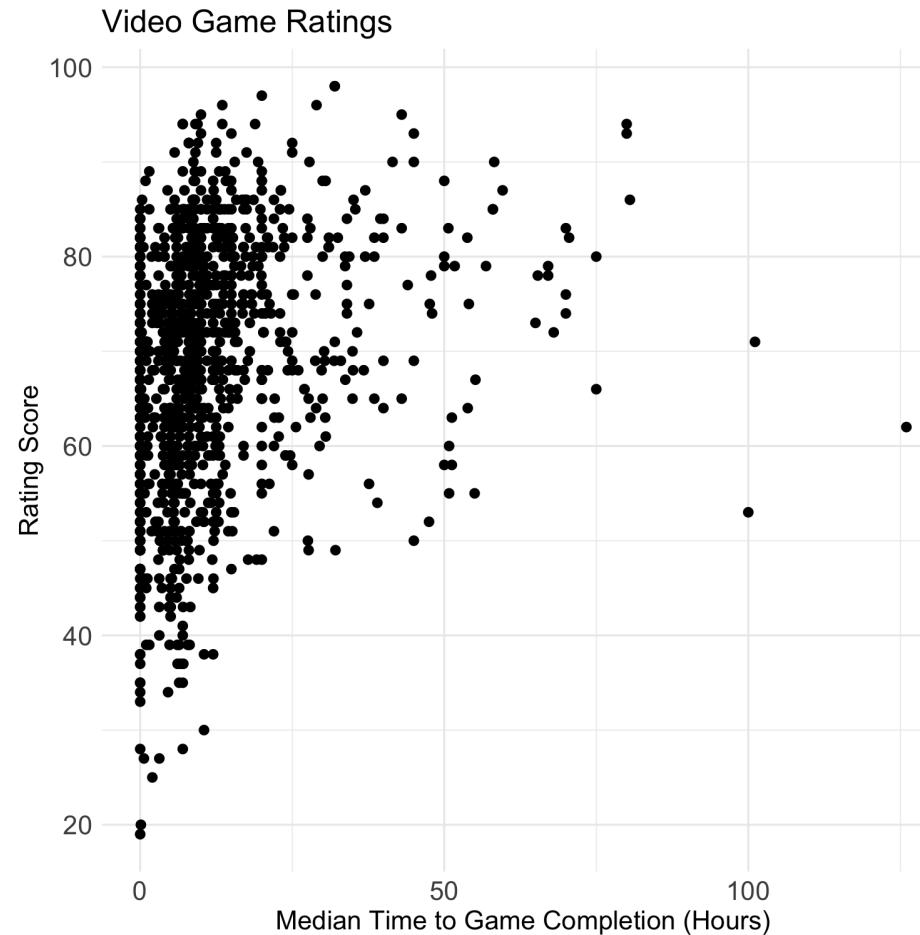
Example: MLB 2011

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?



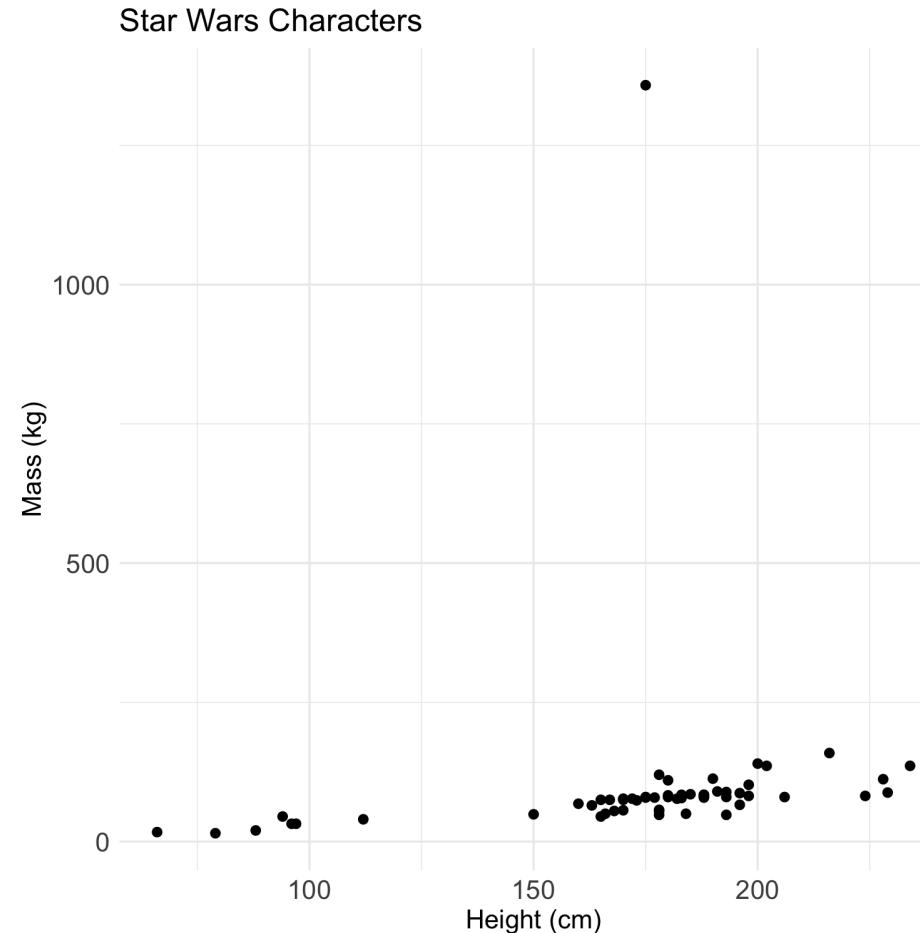
Example: Video Game Ratings

Does the time it takes to complete a handheld game affect its ratings?



Example: Mass and Height of Star Wars Characters

Is the **height** of a character related to that character's **mass**?



Correlation

- Correlation, R , tries to quantify the strength and direction of a linear relationship between two variables

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

Correlation

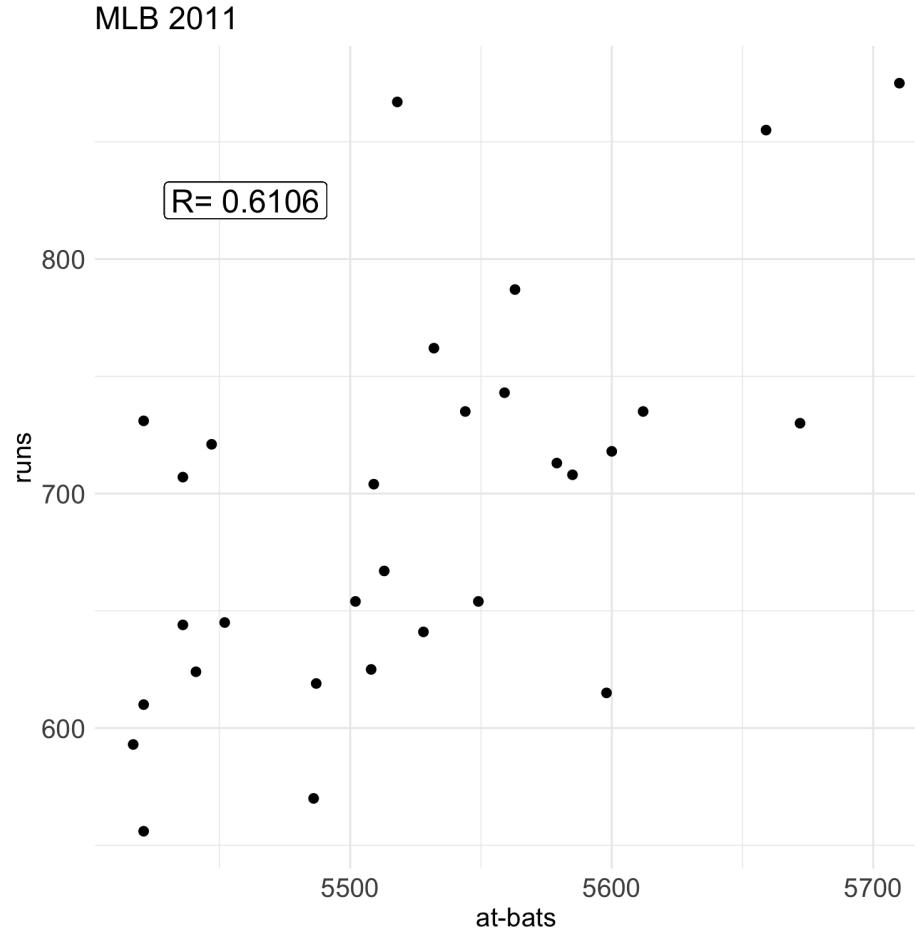
- Correlation, R , tries to quantify the strength and direction of a linear relationship between two variables

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

- Takes on values from -1 to 1 (unitless)
 - $R = 0$ means the variables have no linear relationship
 - $R = 1$ means the variables have a perfect, positive linear relationship
 - If X increases by 1 unit, Y increases by 1 unit as well
 - $R = -1$ means the variables have a perfect, negative linear relationship
 - If X increases by 1 unit, Y decreases by 1 unit

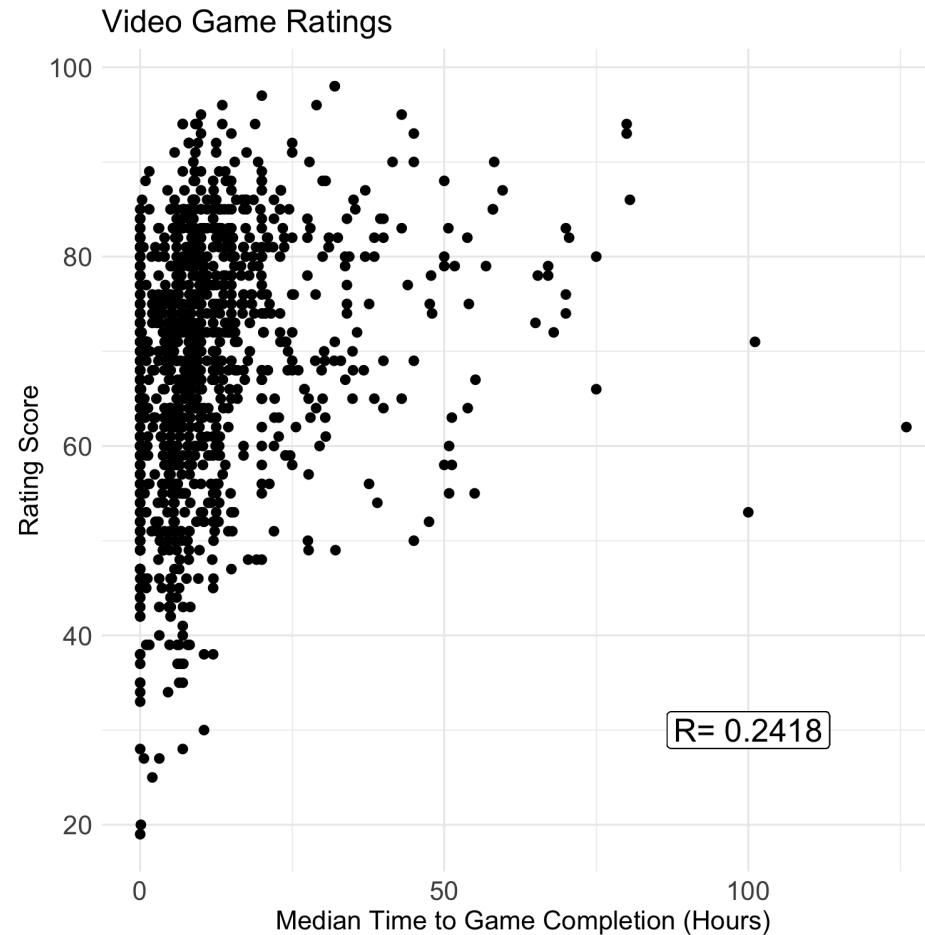
Example: MLB 2011 (pt. 2)

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?



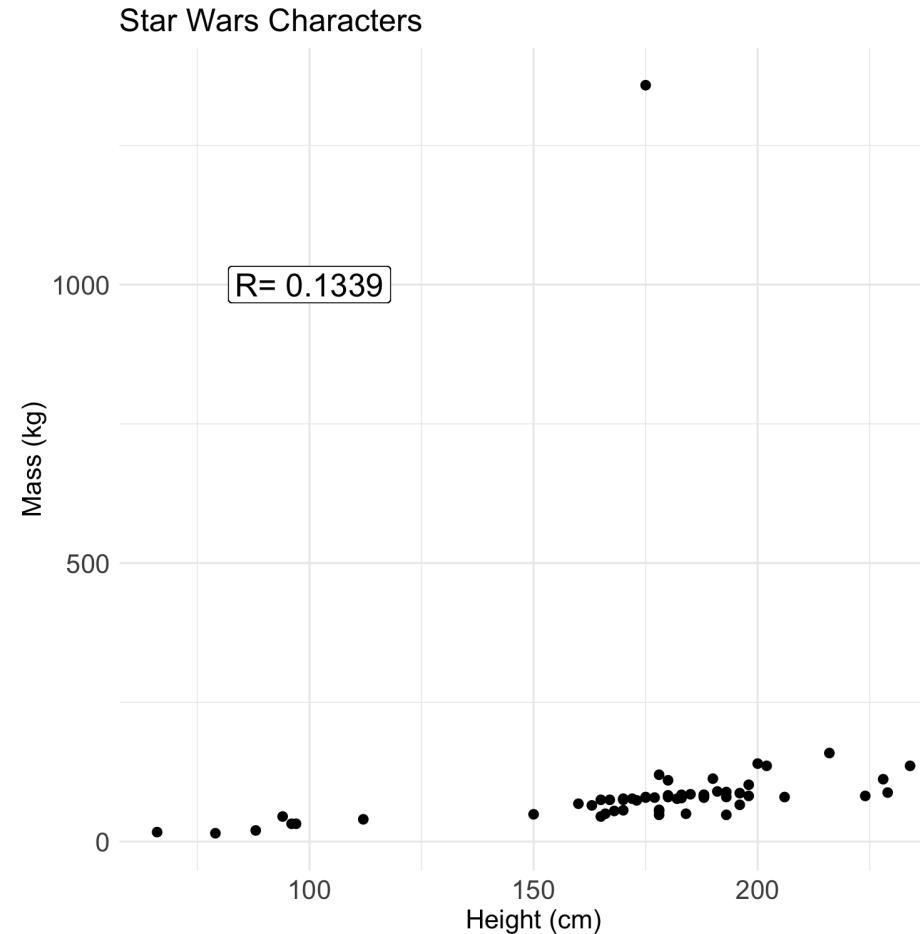
Example: Video Game Ratings (pt. 2)

Does the **time it takes to complete** a handheld game affect its **ratings**?



Example: Star Wars (pt. 2)

Is the **height** of a character related to that character's **mass**?



Correlation

- R only measures the **linear** relationship between two variables
 - Two variables could be perfectly related in a non-linear way and have an R near 0

Correlation

- R only measures the **linear** relationship between two variables
 - Two variables could be perfectly related in a non-linear way and have an R near 0
- Just because two variables *seems* related, does not mean changes in one *cause* changes in the other

Correlation does not equal causation!

Correlation doesn't equal causation example

Linear Regression

- Knowing there *is* a linear relationship is great, but I want to know *how* two variables are related

Linear Regression

- Knowing there *is* a linear relationship is great, but I want to know *how* two variables are related
- **Linear regression** is like finding the line of best fit

$$y = mx + b$$

$$\Rightarrow y = \beta_0 + \beta_1 x$$

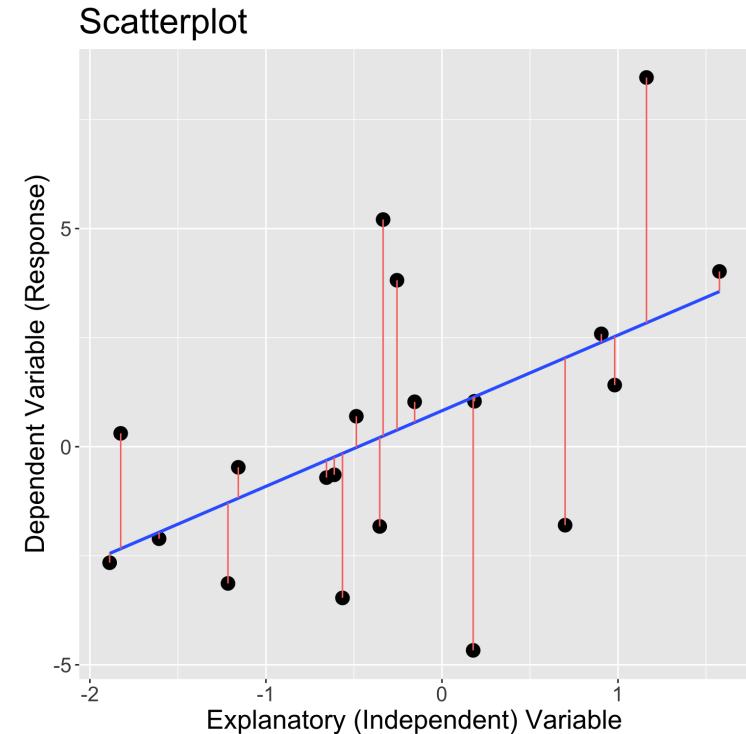
- β_0 is the **intercept**
 - What we would expect y to be if $x=0$
 - Doesn't always have a logical real-world interpretation
- β_1 is the **slope**
 - How much we expect y to change by if we increase x by 1 unit
 - Usually what we're most interested in

Regression

- What makes our regression line the "best fit"?
 - It's the line that minimizes the **sum of squared residuals**
- A **residual**, e_i , is the distance between an observed data point, and what our line would predict it to be

$$e_i = y_i - \hat{y}_i$$

- Squaring the residuals makes sure positive and negative residuals don't cancel each other out



Regression

$$\hat{\beta}_0 = \bar{y} + \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Conditions for Linear Regression

There are 4 conditions we need to meet for linear regression to be appropriate. If we fail to meet one (or multiple) of these conditions, the reported relationship between X and Y from the regression might be biased from the truth

1. Linearity

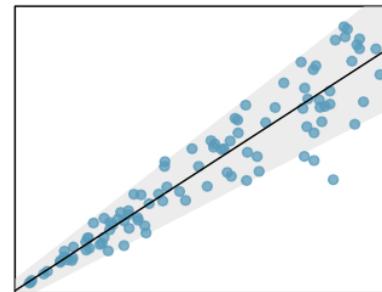
- Linear regression models the linear relationship between variables, so the data should look like some sort of linear relationship exists

2. Roughly "bell-curve" residuals

- If you make histogram of the residuals, it should look like a bell curve (symmetric, unimodal)

3. Constant variability

- The variability of points around the least squares line remains roughly constant
 - Example of non-constant variability:

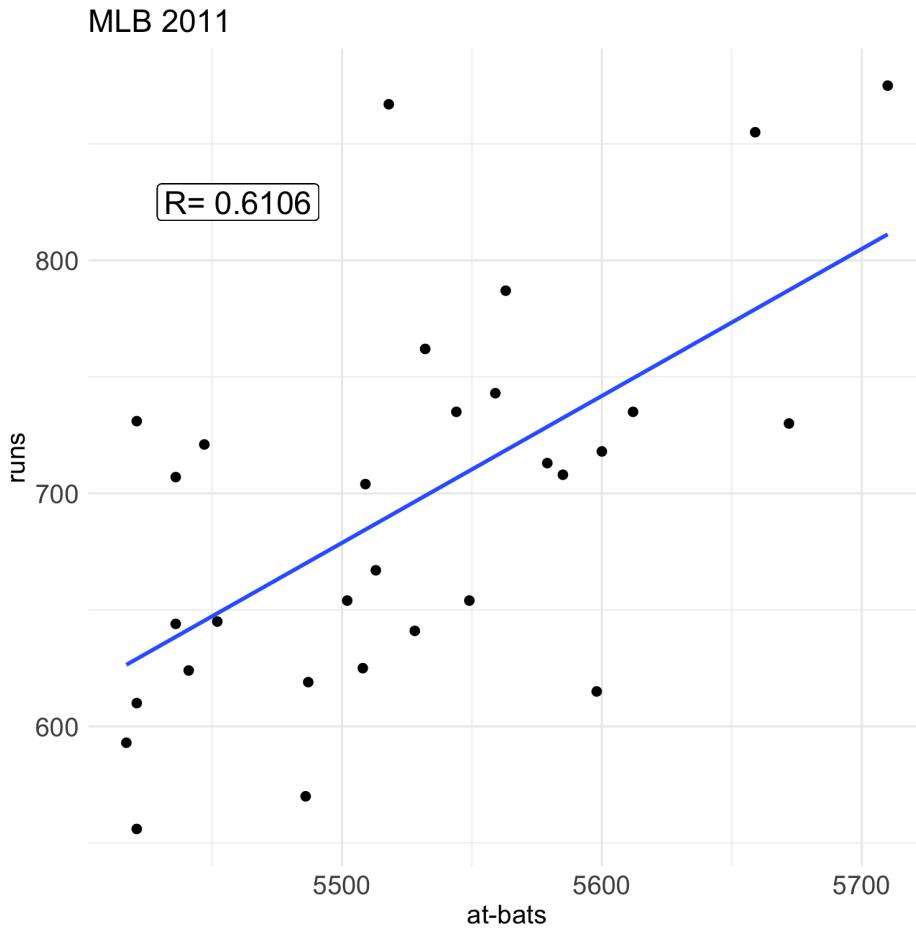


4. Independent observations

- No observation should influence the values of other observations

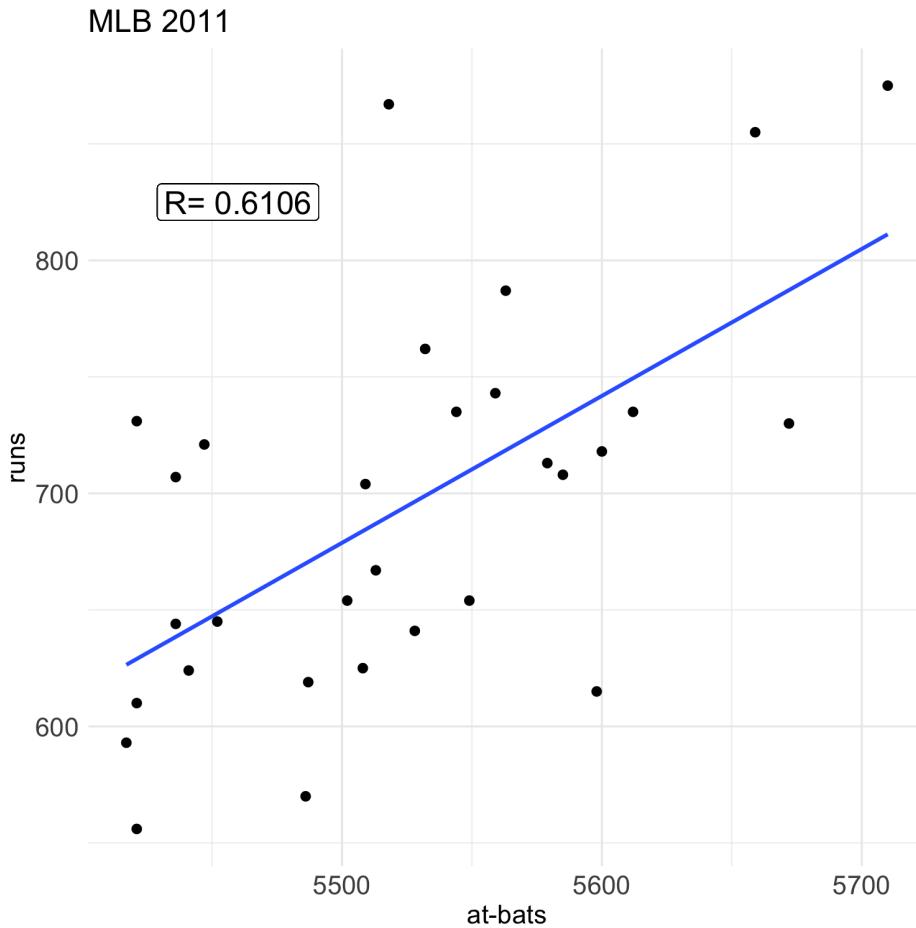
Example: MLB 2011 (pt. 3.1)

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?



Example: MLB 2011 (pt. 3.1)

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?

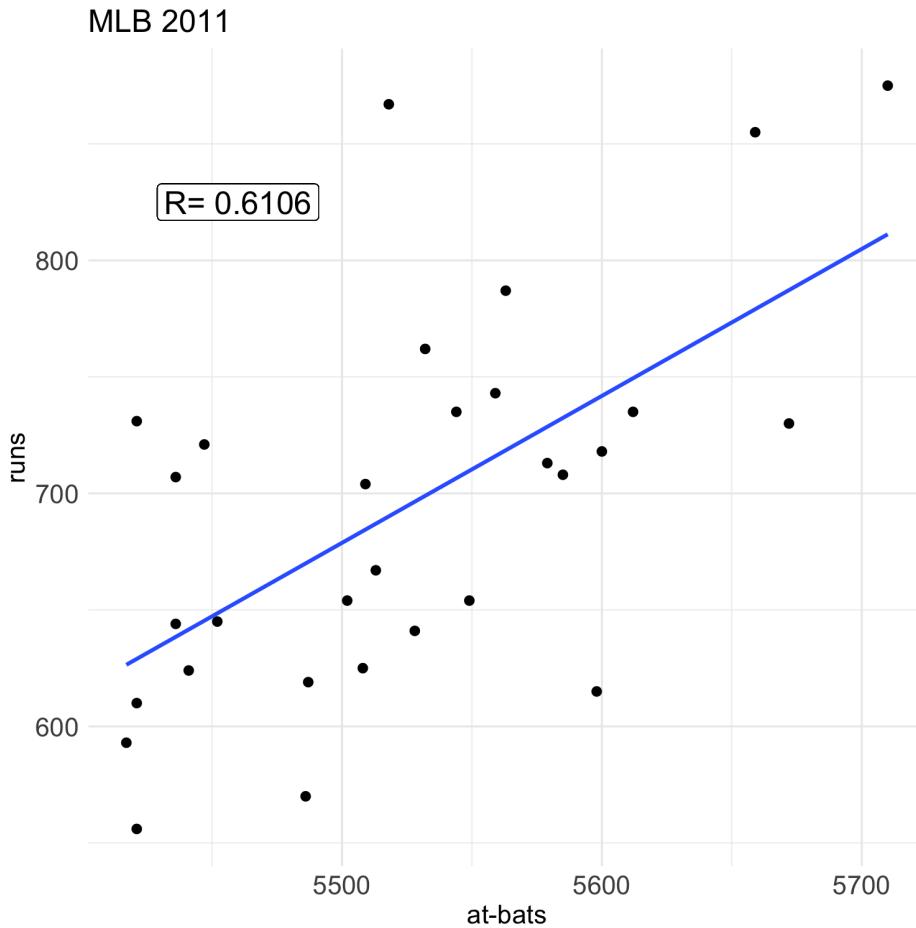


$$\text{Runs} = \beta_0 + \beta_1 (\text{at-bats})$$

$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

Example: MLB 2011 (pt. 3.1)

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?



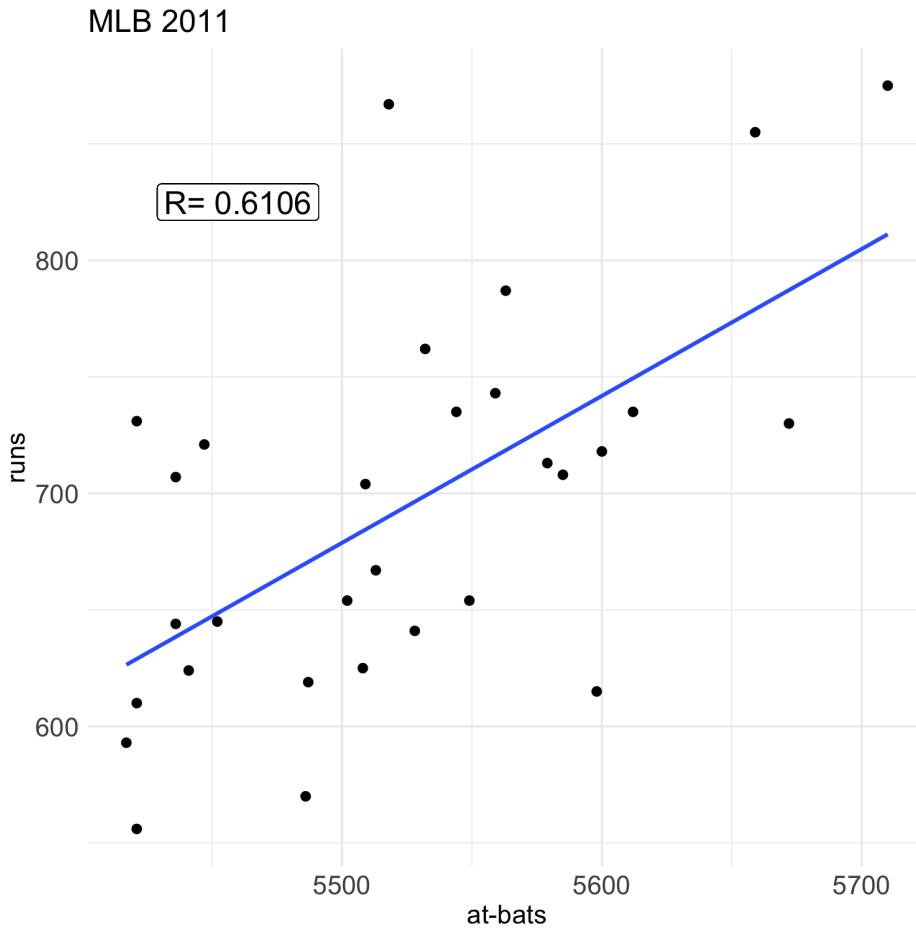
$$\text{Runs} = \beta_0 + \beta_1 (\text{at-bats})$$

$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

When a team sends no players to bat for an entire season, we would expect their runs for the season to be -2789.2429, on average.

Example: MLB 2011 (pt. 3.1)

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?



$$\text{Runs} = \beta_0 + \beta_1 (\text{at-bats})$$

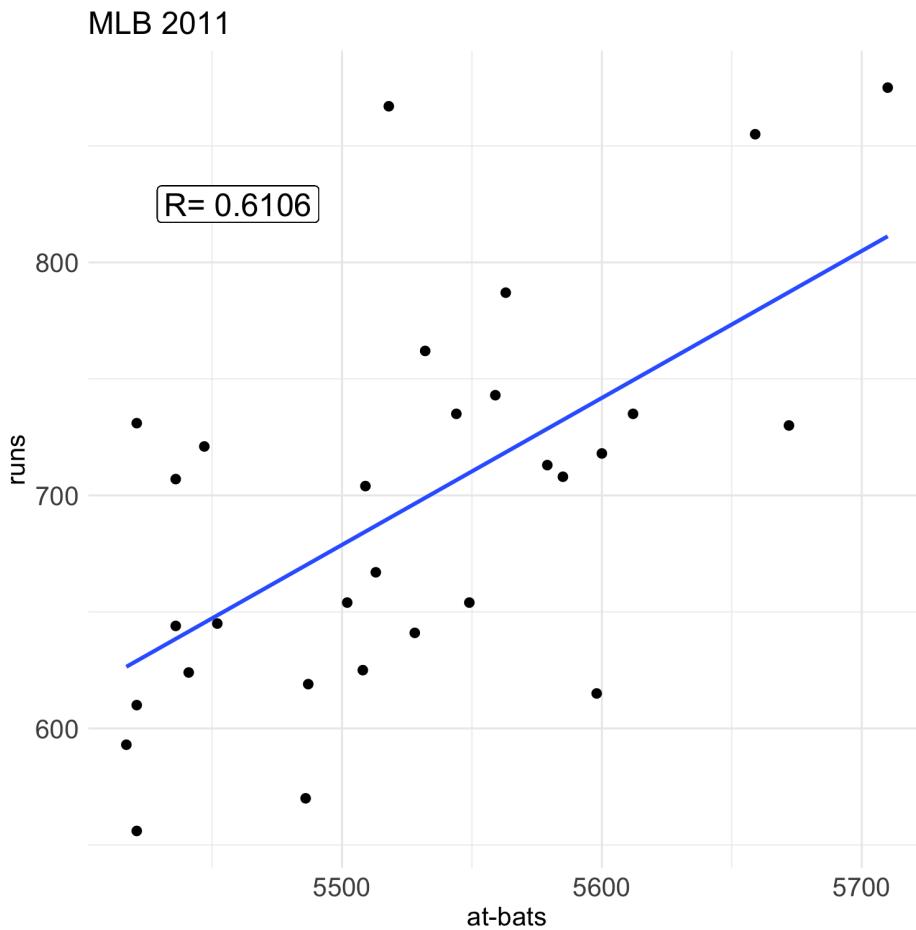
$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

When a team sends no players to bat for an entire season, we would expect their runs for the season to be -2789.2429, on average.

Doesn't make sense

Example: MLB 2011 (pt. 3.1)

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?



$$\text{Runs} = \beta_0 + \beta_1 (\text{at-bats})$$

$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

When a team sends no players to bat for an entire season, we would expect their runs for the season to be -2789.2429, on average.

Doesn't make sense

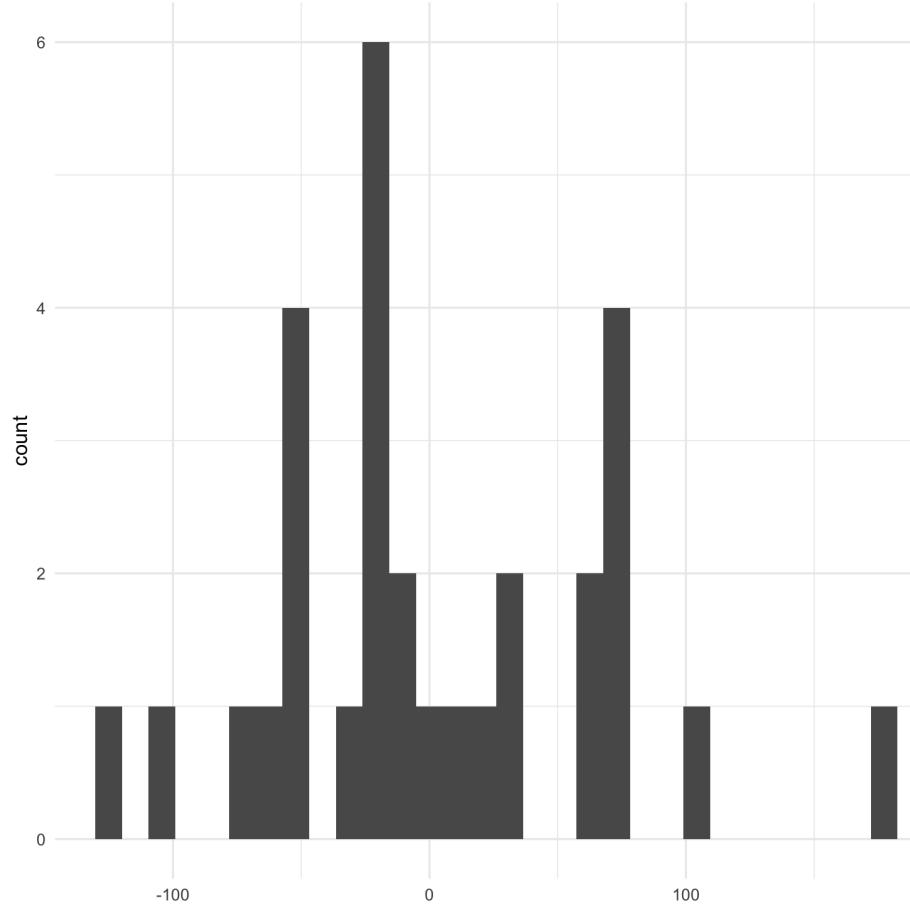
For every 1 extra player a team sends to bat during a season, their total runs increase by 0.6305, on average.

Example: MLB 2011 (pt. 3.2)

- Is linear regression appropriate for this data?

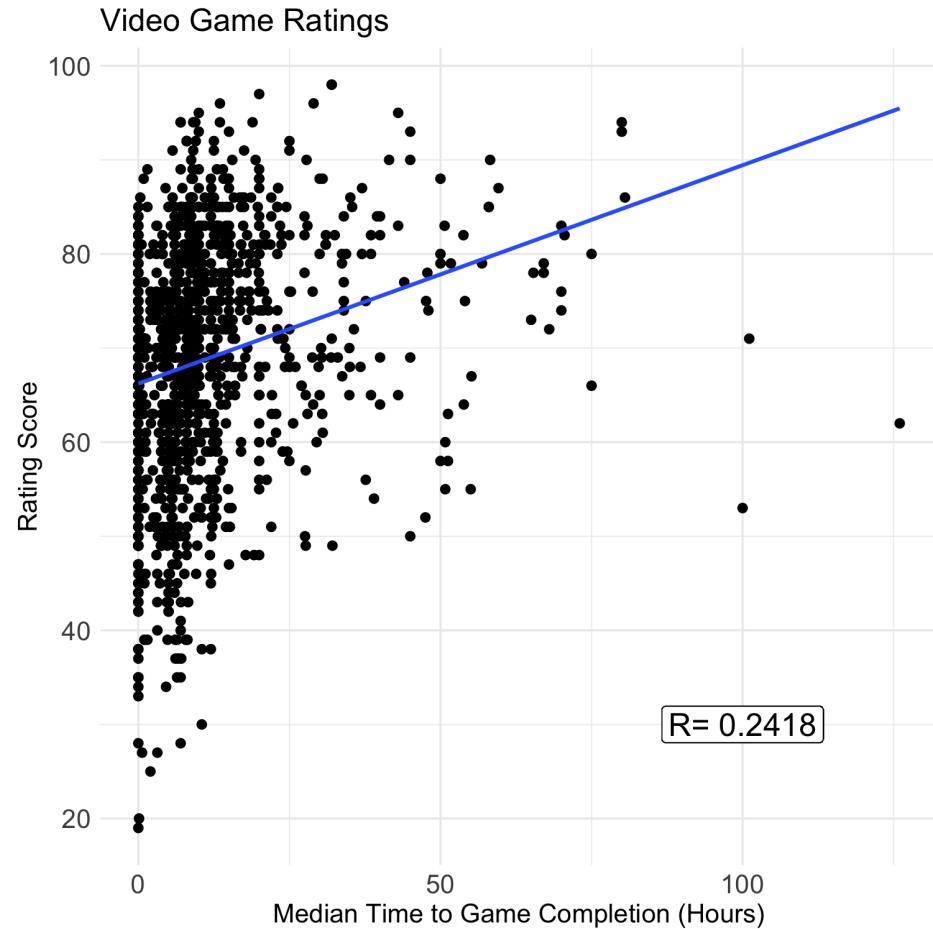
Example: MLB 2011 (pt. 3.2)

- Is linear regression appropriate for this data?



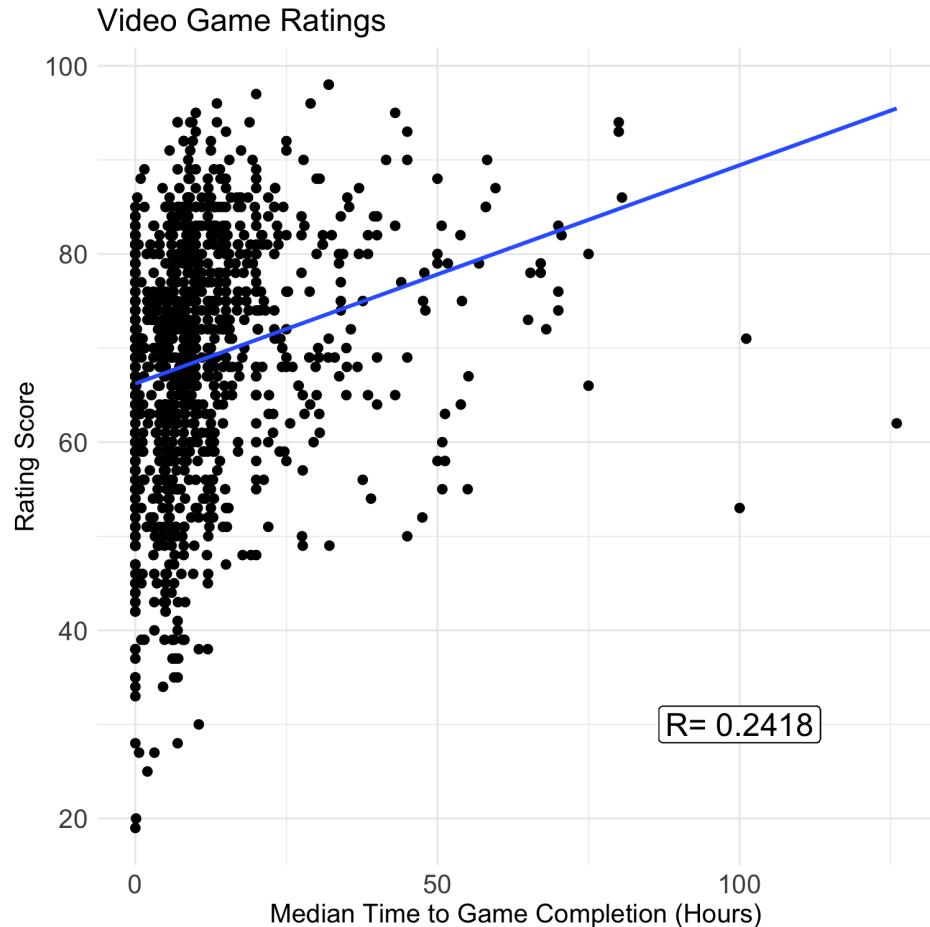
Example: Video Game Ratings (pt. 3.1)

Does the **time it takes to complete** a handheld game affect its **ratings**?



Example: Video Game Ratings (pt. 3.1)

Does the time it takes to complete a handheld game affect its ratings?

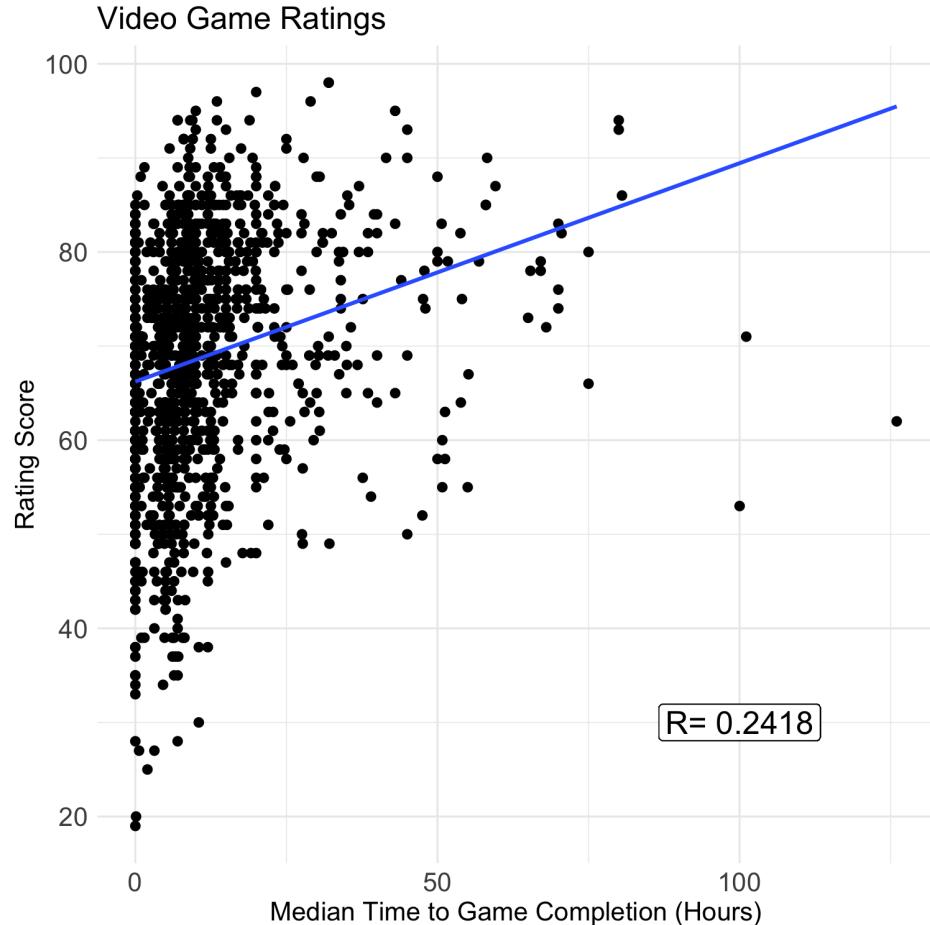


$$\text{Review Score} = \beta_0 + \beta_1 (\text{Median time to completion})$$

$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

Example: Video Game Ratings (pt. 3.1)

Does the time it takes to complete a handheld game affect its ratings?



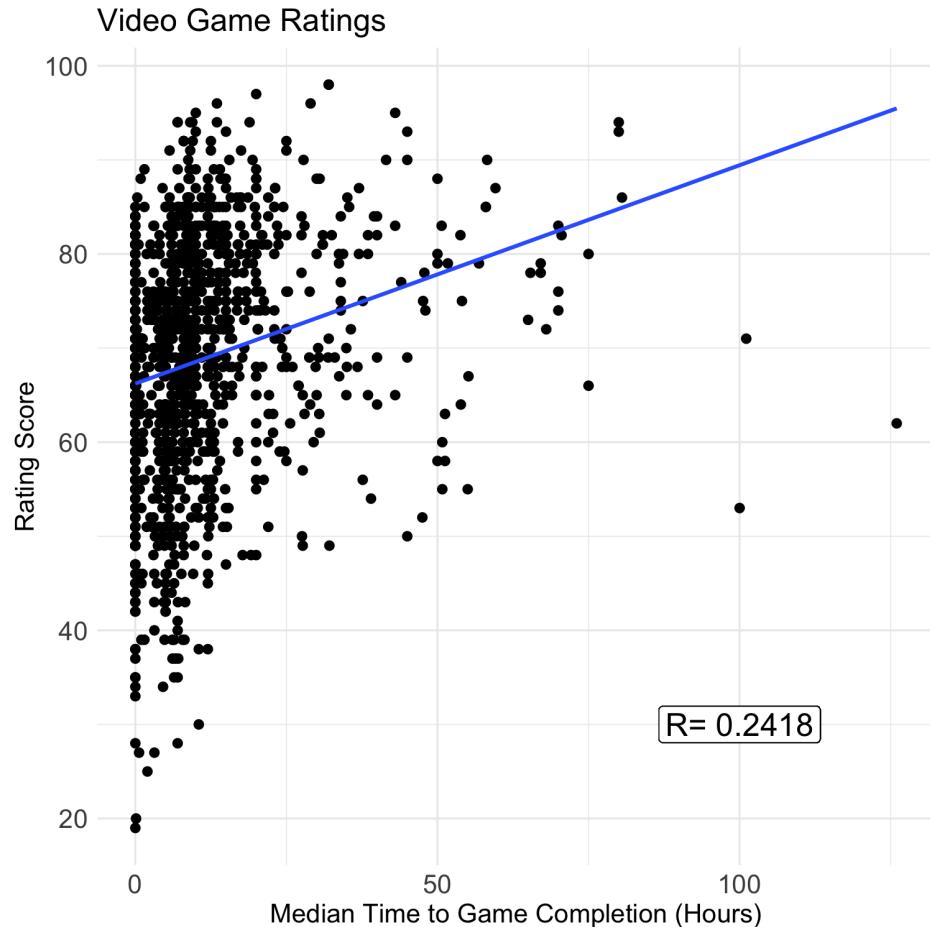
$$\text{Review Score} = \beta_0 + \beta_1 (\text{Median time to completion})$$

$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

When a game's median completion time is 0 hours, the game will receive a rating of 66.2225, on average.

Example: Video Game Ratings (pt. 3.1)

Does the time it takes to complete a handheld game affect its ratings?



$$\text{Review Score} = \beta_0 + \beta_1 (\text{Median time to completion})$$

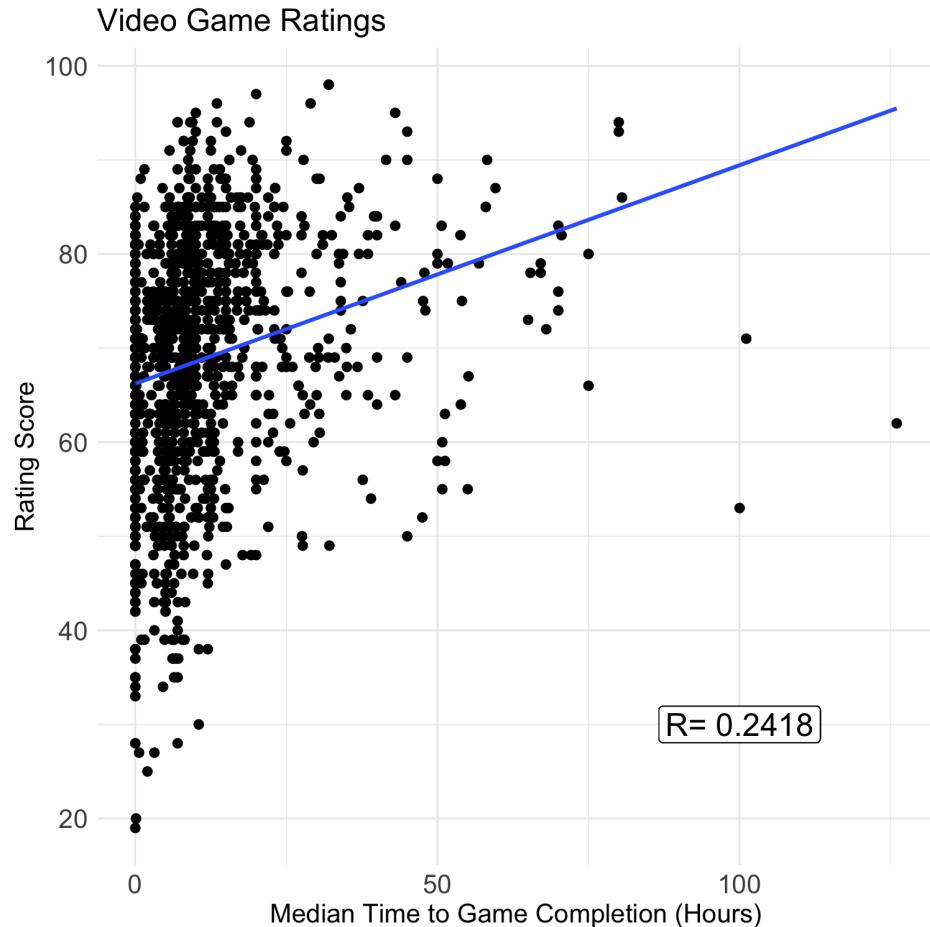
$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

When a game's median completion time is 0 hours, the game will receive a rating of 66.2225, on average.

Doesn't make sense

Example: Video Game Ratings (pt. 3.1)

Does the **time it takes to complete** a handheld game affect its **ratings**?



$$\text{Review Score} = \beta_0 + \beta_1 (\text{Median time to completion})$$

$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

When a game's median completion time is 0 hours, the game will receive a rating of 66.2225, on average.

Doesn't make sense

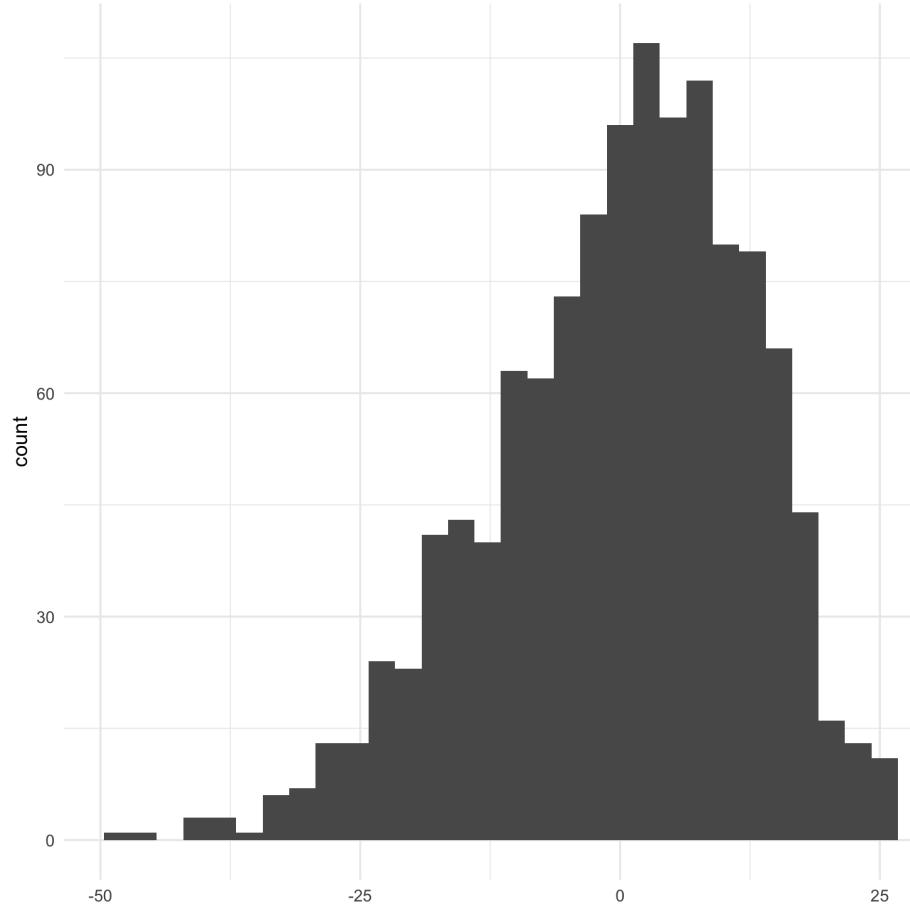
When the median completion time increases by 1 hour, a game's rating increases by 0.2321, on average.

Example: Video Game Ratings (pt. 3.2)

- Is linear regression appropriate for this data?

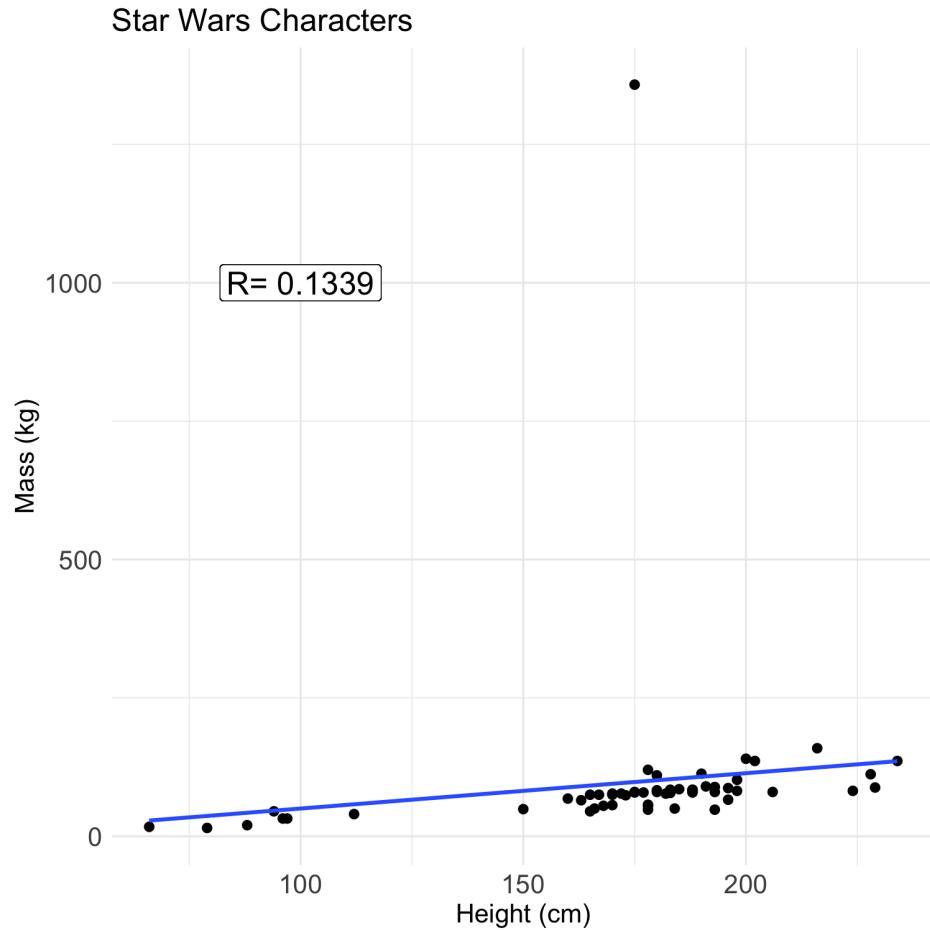
Example: Video Game Ratings (pt. 3.2)

- Is linear regression appropriate for this data?



Example: Star Wars (pt. 3.1)

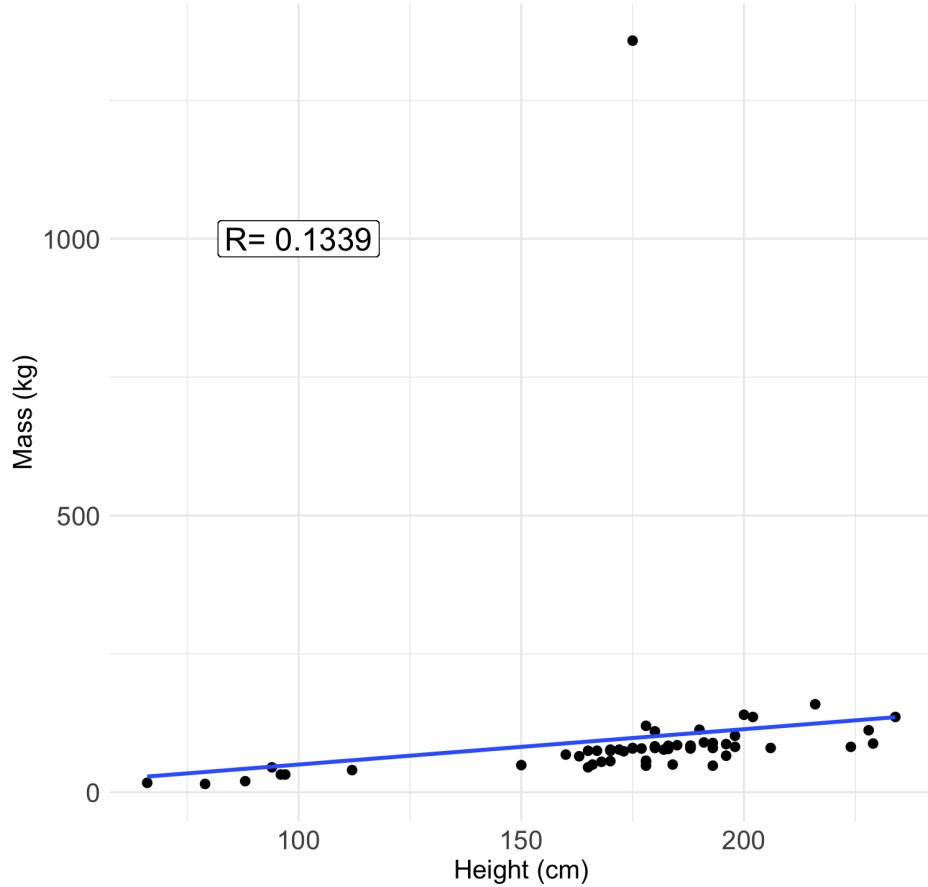
Is the **height** of a character related to that character's **mass**?



Example: Star Wars (pt. 3.1)

Is the **height** of a character related to that character's **mass**?

Star Wars Characters

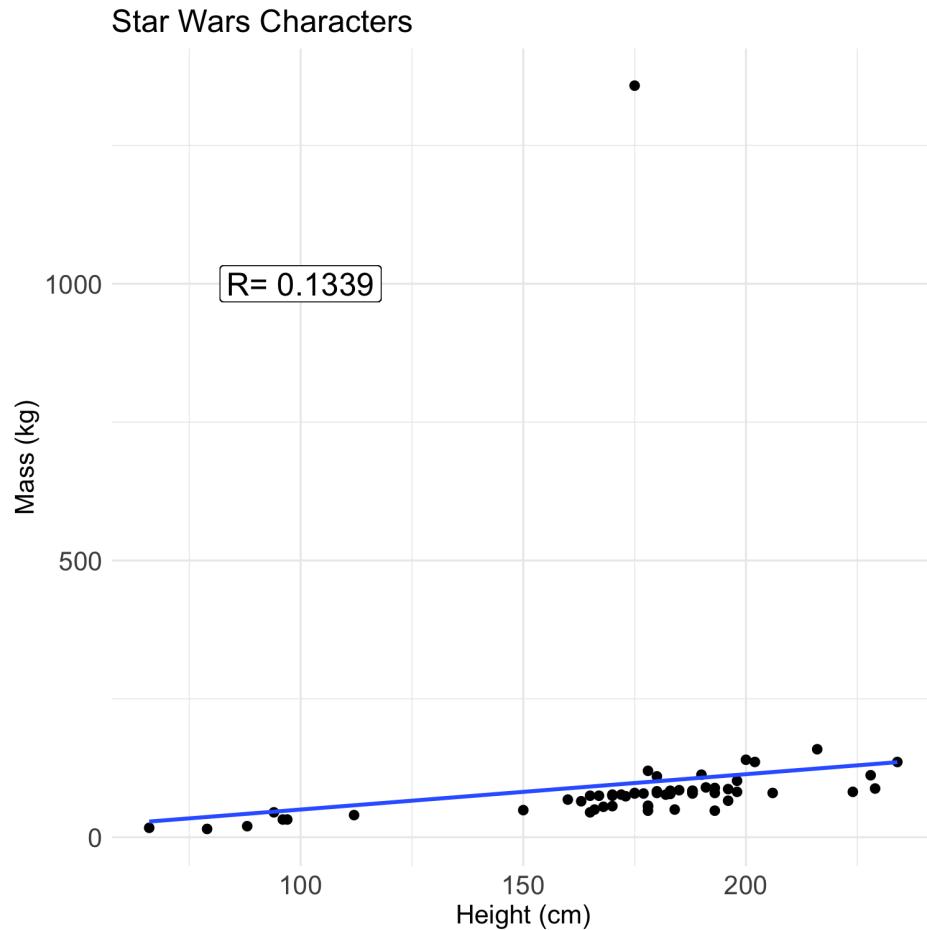


$$\text{Mass} = \beta_0 + \beta_1 (\text{height})$$

$$\text{Mass} = -13.8103 + 0.6386(\text{height})$$

Example: Star Wars (pt. 3.1)

Is the **height** of a character related to that character's **mass**?



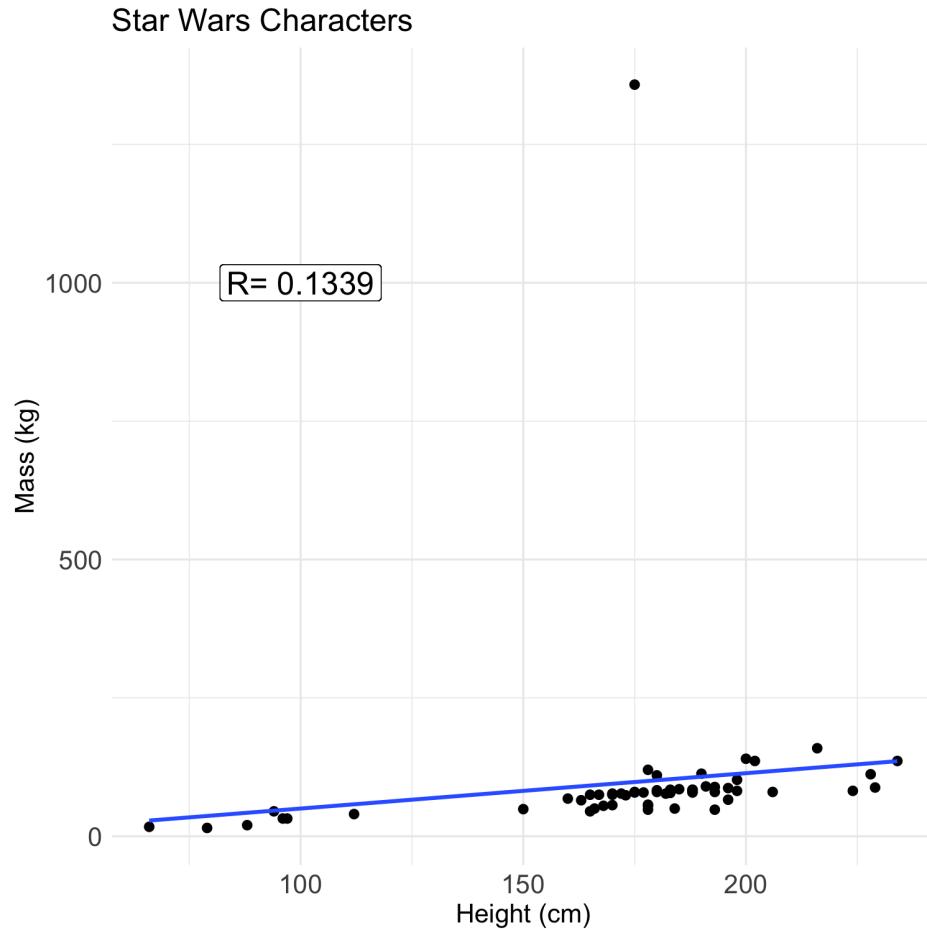
$$\text{Mass} = \beta_0 + \beta_1 (\text{height})$$

$$\text{Mass} = -13.8103 + 0.6386(\text{height})$$

When a character is 0 cm tall, we would expect their mass to be -13.8103 kg, on average.

Example: Star Wars (pt. 3.1)

Is the **height** of a character related to that character's **mass**?



$$\text{Mass} = \beta_0 + \beta_1 (\text{height})$$

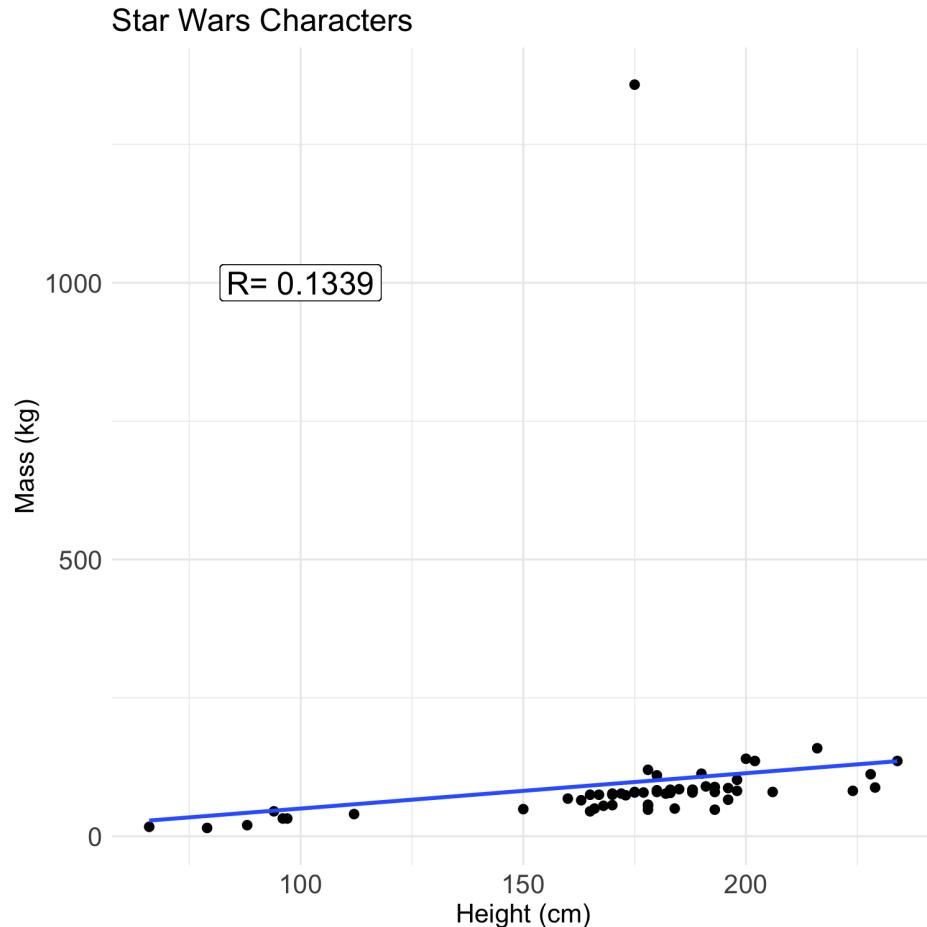
$$\text{Mass} = -13.8103 + 0.6386(\text{height})$$

When a character is 0 cm tall, we would expect their mass to be -13.8103 kg, on average.

Doesn't make sense

Example: Star Wars (pt. 3.1)

Is the **height** of a character related to that character's **mass**?



$$\text{Mass} = \beta_0 + \beta_1 (\text{height})$$

$$\text{Mass} = -13.8103 + 0.6386(\text{height})$$

When a character is 0 cm tall, we would expect their mass to be -13.8103 kg, on average.

Doesn't make sense

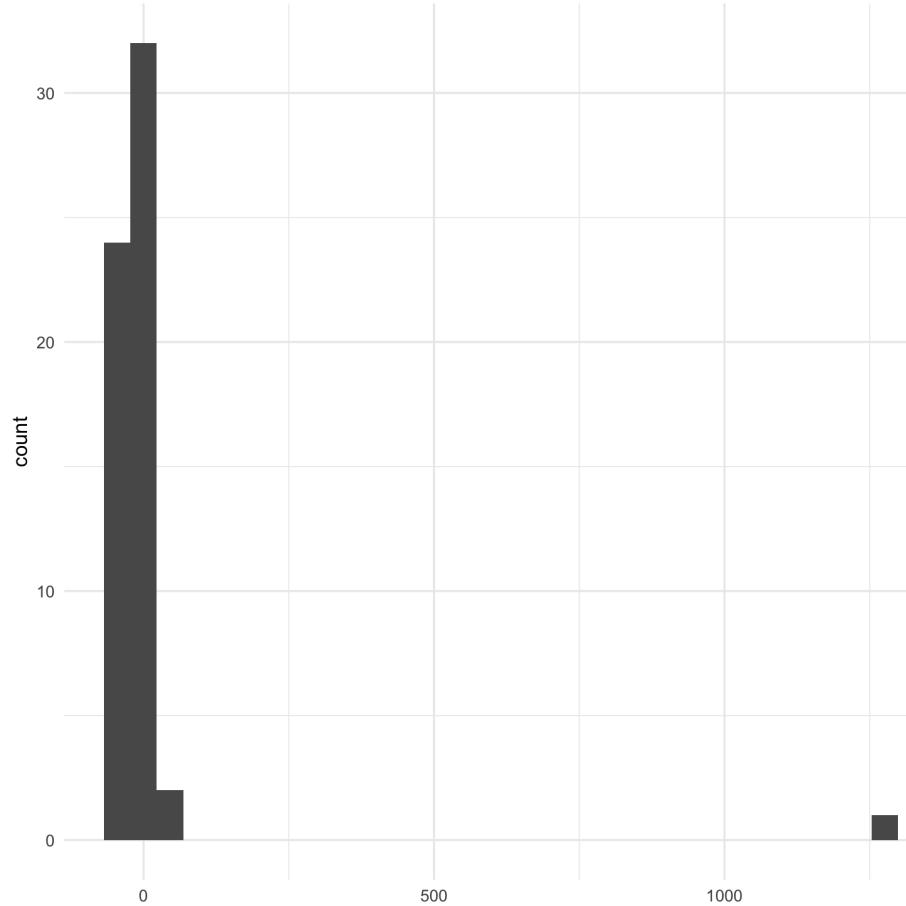
For every 1 extra centimeter taller a character is, their mass increases by 0.6386 kg, on average.

Example: Star Wars (pt. 3.2)

- Is linear regression appropriate for this data?

Example: Star Wars (pt. 3.2)

- Is linear regression appropriate for this data?



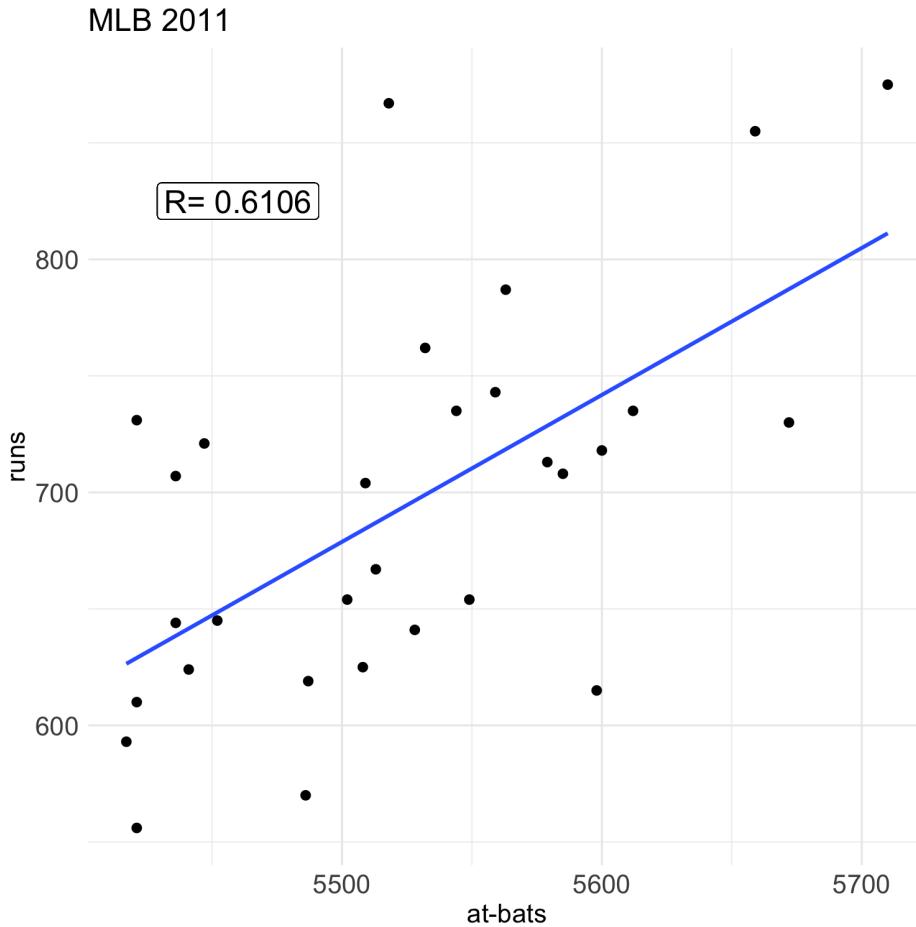
R^2

- How can we tell whether our regression model is good?
 - Regression gives us the "best fit" line, but is "best" actually good?
- R^2 quantifies how much **variation** in our response is explained by our explanatory variable
 - If our explanatory variable explains most of the variation in our response, then our regression model is probably pretty good at predicting what the response will do
 - If our explanatory variable explains little of the variation in our response, then it will not do a good job of predicting what the response will do

"Variable x explains $(R^2 \times 100)\%$ of the variation in response variable y "

Example: MLB 2011 (pt. 4)

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?



$$\text{Runs} = \beta_0 + \beta_1 (\text{at-bats})$$

$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

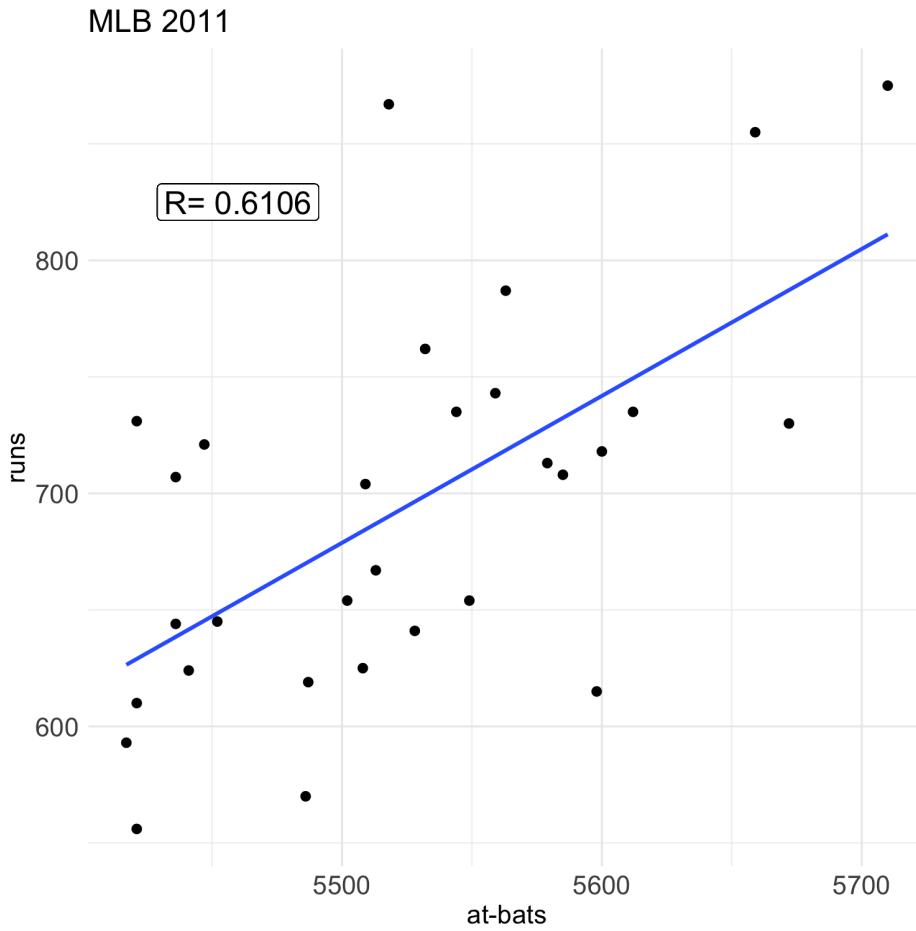
When a team sends no players to bat for an entire season, we would expect their runs for the season to be -2789.2429, on average.

Doesn't make sense

For every 1 extra player a team sends to bat during a season, their total runs increase by 0.6305, on average.

Example: MLB 2011 (pt. 4)

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?



$$\text{Runs} = \beta_0 + \beta_1 (\text{at-bats})$$

$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

When a team sends no players to bat for an entire season, we would expect their runs for the season to be -2789.2429, on average.

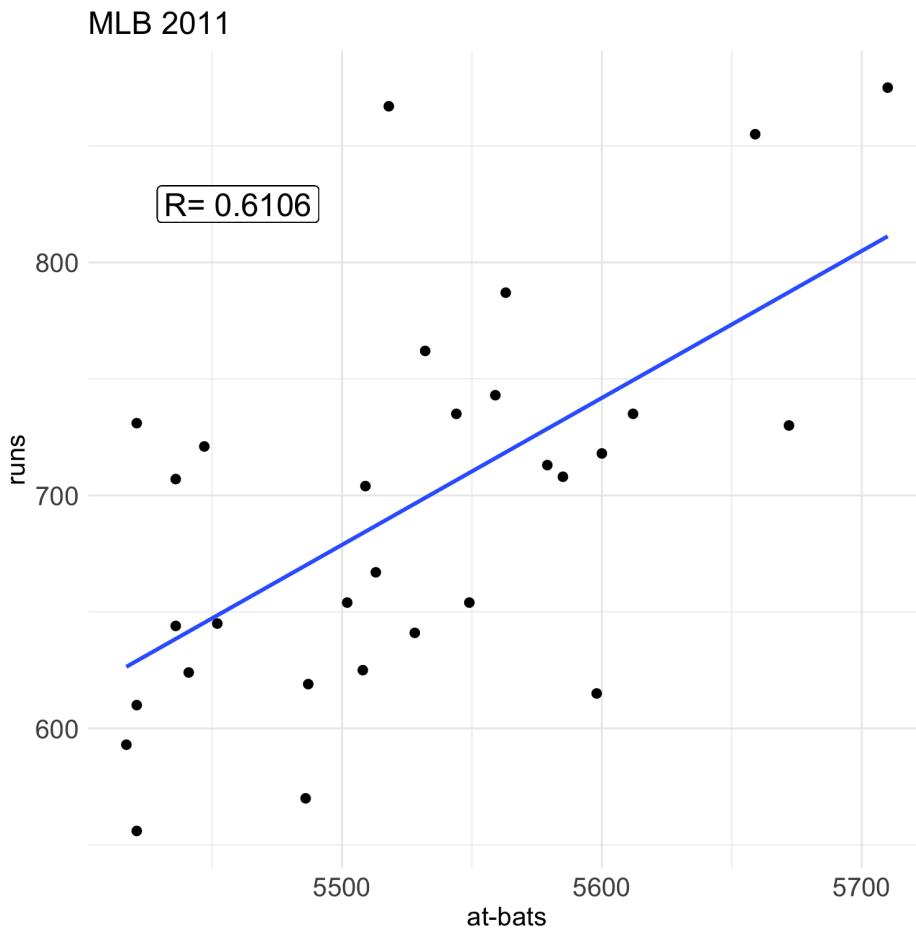
Doesn't make sense

For every 1 extra player a team sends to bat during a season, their total runs increase by 0.6305, on average.

$$R^2 = 0.3729$$

Example: MLB 2011 (pt. 4)

Is the number of times a team has a player go up to bat throughout a season (called "at-bats") related to the number of runs that teams scores?



$$\text{Runs} = \beta_0 + \beta_1 (\text{at-bats})$$

$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

When a team sends no players to bat for an entire season, we would expect their runs for the season to be -2789.2429, on average.

Doesn't make sense

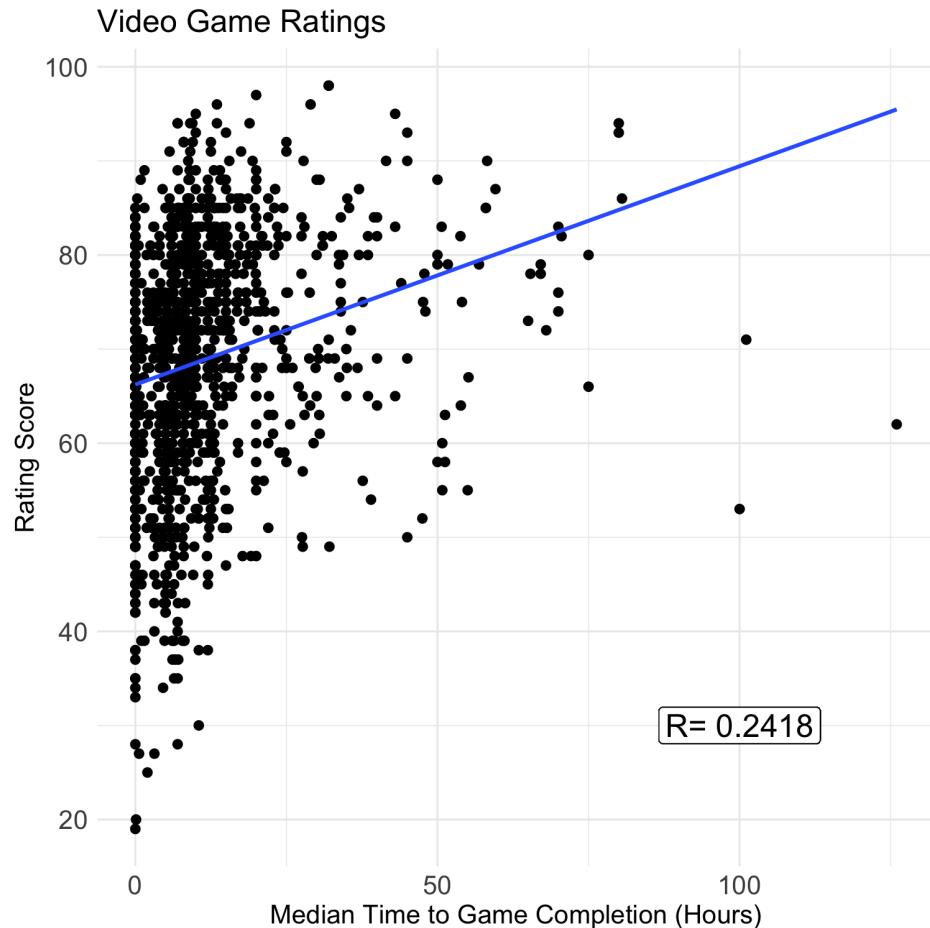
For every 1 extra player a team sends to bat during a season, their total runs increase by 0.6305, on average.

$$R^2 = 0.3729$$

37.29 % of the variability in runs is explained by at-bats.

Example: Video Game Ratings (pt. 4)

Does the **time it takes to complete** a handheld game affect its **ratings**?



$$\text{Review Score} = \beta_0 + \beta_1 (\text{Median time to completion})$$

$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

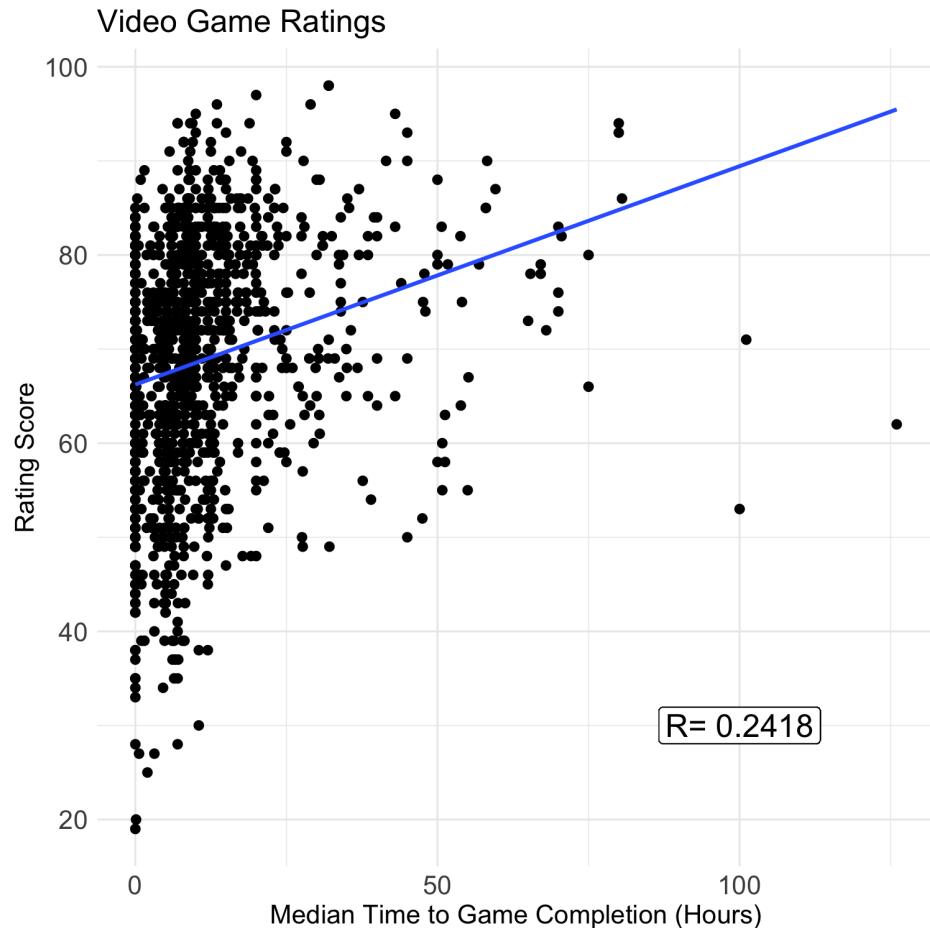
When a game's median completion time is 0 hours, the game will receive a rating of 66.2225, on average.

Doesn't make sense

When the median completion time increases by 1 hour, a game's rating increases by 0.2321, on average.

Example: Video Game Ratings (pt. 4)

Does the time it takes to complete a handheld game affect its ratings?



$$\text{Review Score} = \beta_0 + \beta_1 (\text{Median time to completion})$$

$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

When a game's median completion time is 0 hours, the game will receive a rating of 66.2225, on average.

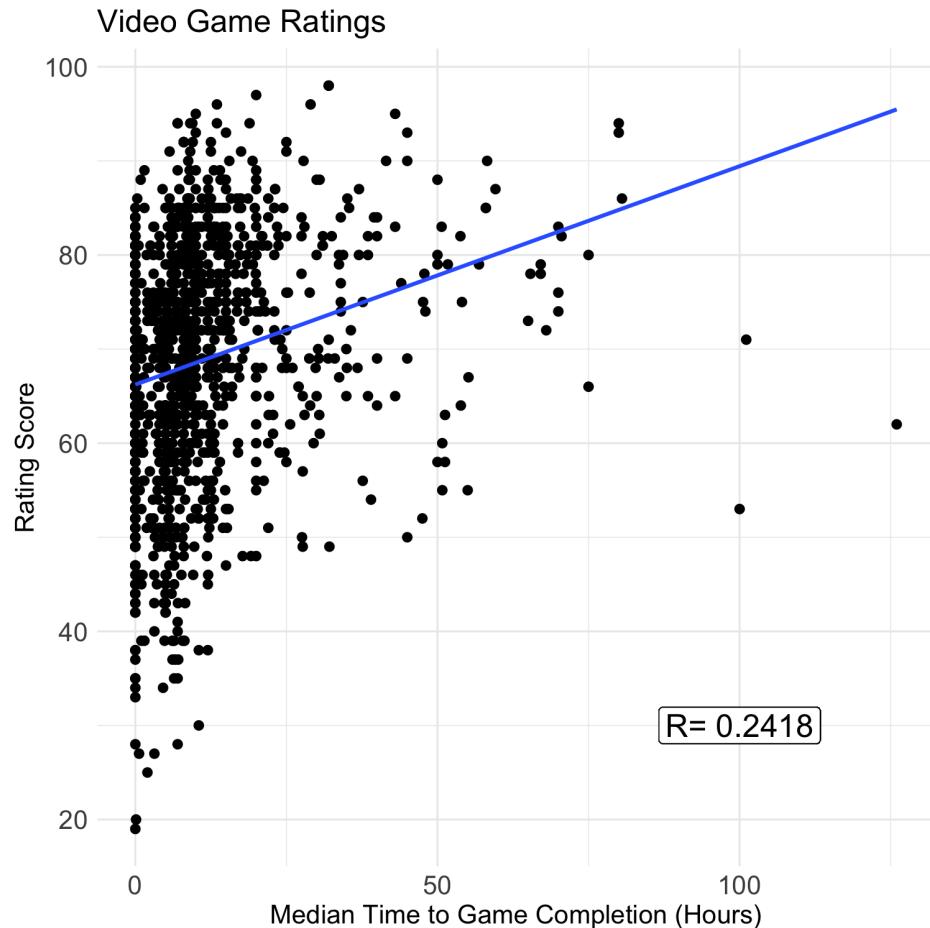
Doesn't make sense

When the median completion time increases by 1 hour, a game's rating increases by 0.2321, on average.

$$R^2 = 0.0585$$

Example: Video Game Ratings (pt. 4)

Does the time it takes to complete a handheld game affect its ratings?



$$\text{Review Score} = \beta_0 + \beta_1 (\text{Median time to completion})$$

$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

When a game's median completion time is 0 hours, the game will receive a rating of 66.2225, on average.

Doesn't make sense

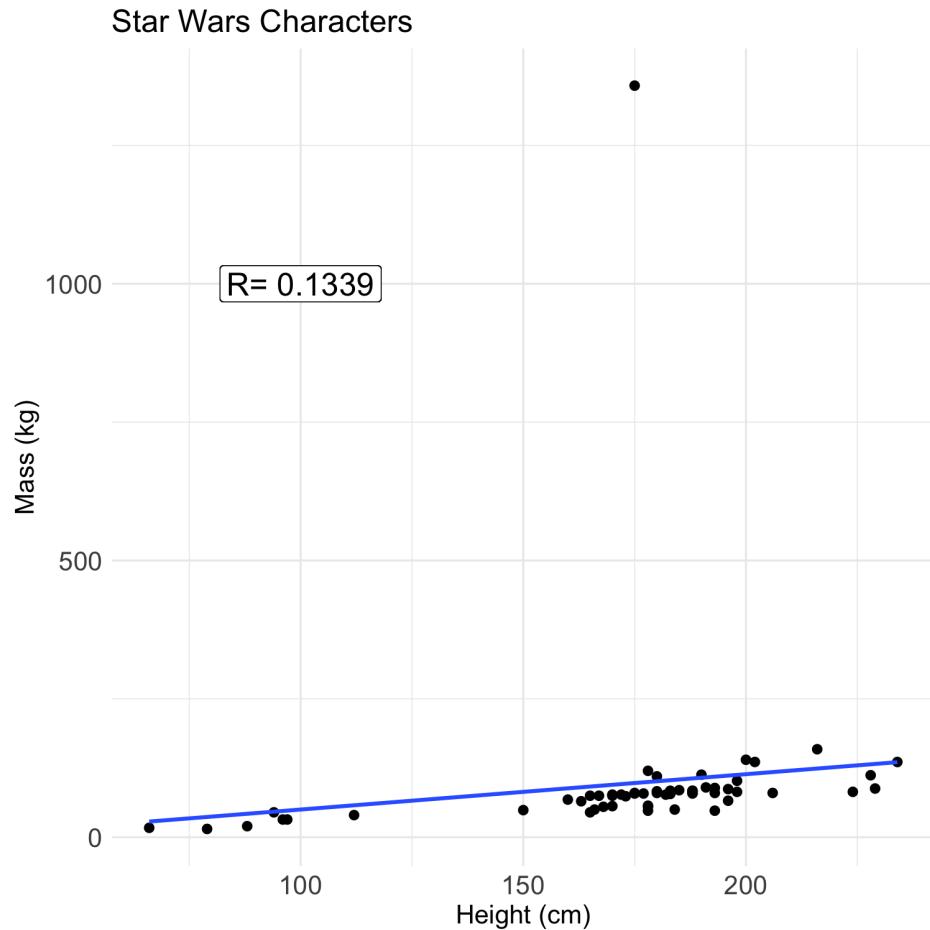
When the median completion time increases by 1 hour, a game's rating increases by 0.2321, on average.

$$R^2 = 0.0585$$

5.85 % of the variability in rating score is explained by median completion time.

Example: Star Wars (pt. 4)

Is the **height** of a character related to that character's **mass**?



$$\text{Mass} = \beta_0 + \beta_1 (\text{height})$$

$$\text{Mass} = -13.8103 + 0.6386(\text{height})$$

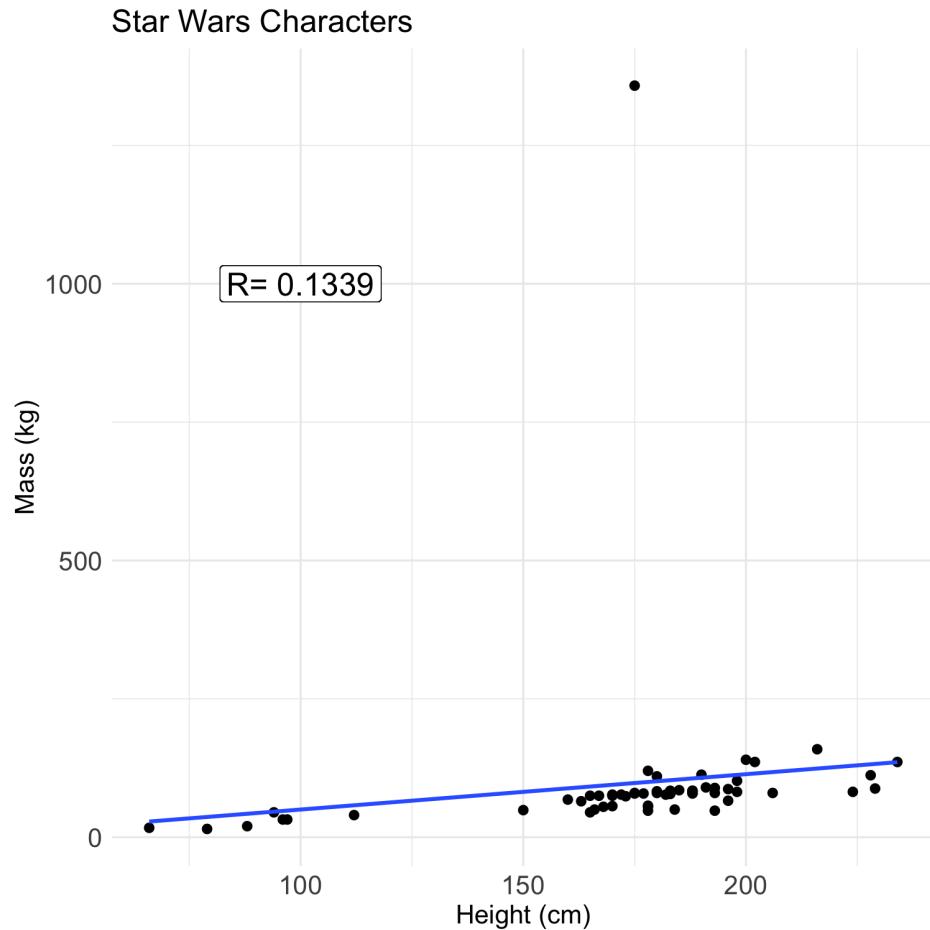
When a character is 0 cm tall, we would expect their mass to be -13.8103 kg, on average.

Doesn't make sense

For every 1 extra centimeter taller a character is, their mass increases by 0.6386 kg, on average.

Example: Star Wars (pt. 4)

Is the **height** of a character related to that character's **mass**?



$$\text{Mass} = \beta_0 + \beta_1 (\text{height})$$

$$\text{Mass} = -13.8103 + 0.6386(\text{height})$$

When a character is 0 cm tall, we would expect their mass to be -13.8103 kg, on average.

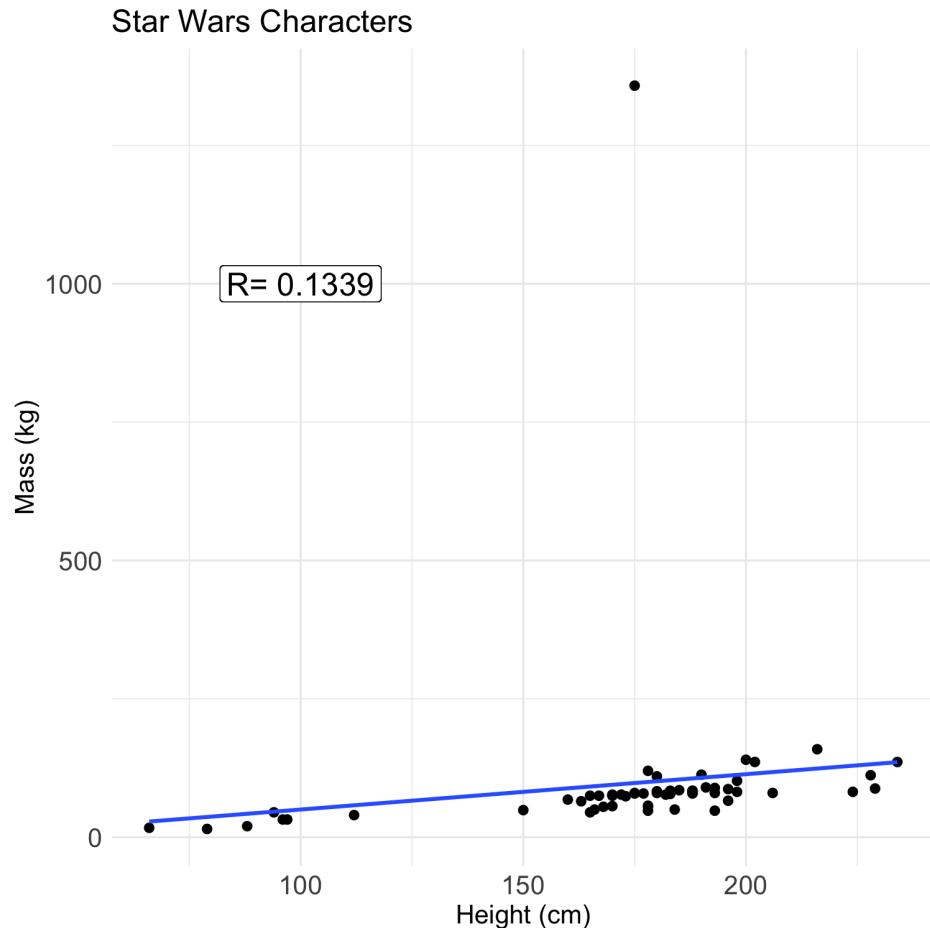
Doesn't make sense

For every 1 extra centimeter taller a character is, their mass increases by 0.6386 kg, on average.

$$R^2 = 0.0179$$

Example: Star Wars (pt. 4)

Is the **height** of a character related to that character's **mass**?



$$\text{Mass} = \beta_0 + \beta_1 (\text{height})$$

$$\text{Mass} = -13.8103 + 0.6386(\text{height})$$

When a character is 0 cm tall, we would expect their mass to be -13.8103 kg, on average.

Doesn't make sense

For every 1 extra centimeter taller a character is, their mass increases by 0.6386 kg, on average.

$$R^2 = 0.0179$$

1.79 % of the variability in mass is explained by character height.

Prediction

- We can also use the regression model to **predict** response variables given explanatory variables that aren't explicitly in our dataset!

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

Prediction

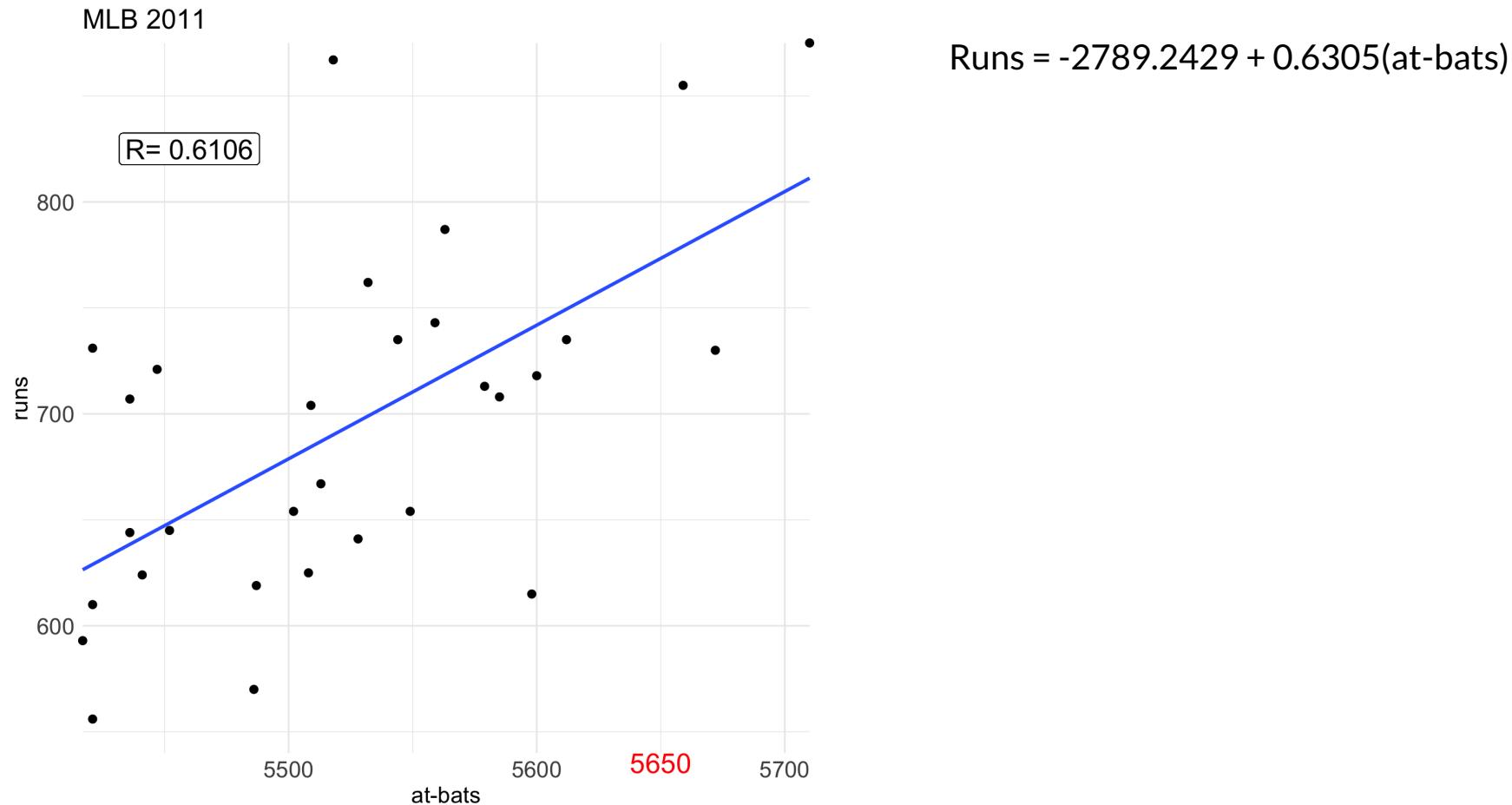
- We can also use the regression model to **predict** response variables given explanatory variables that aren't explicitly in our dataset!

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

- Must be careful of **extrapolation** though
 - Making inference wildly outside of the range of our data
 - We are only sure of the relationship between explanatory and response variables within the scope of data we have observed

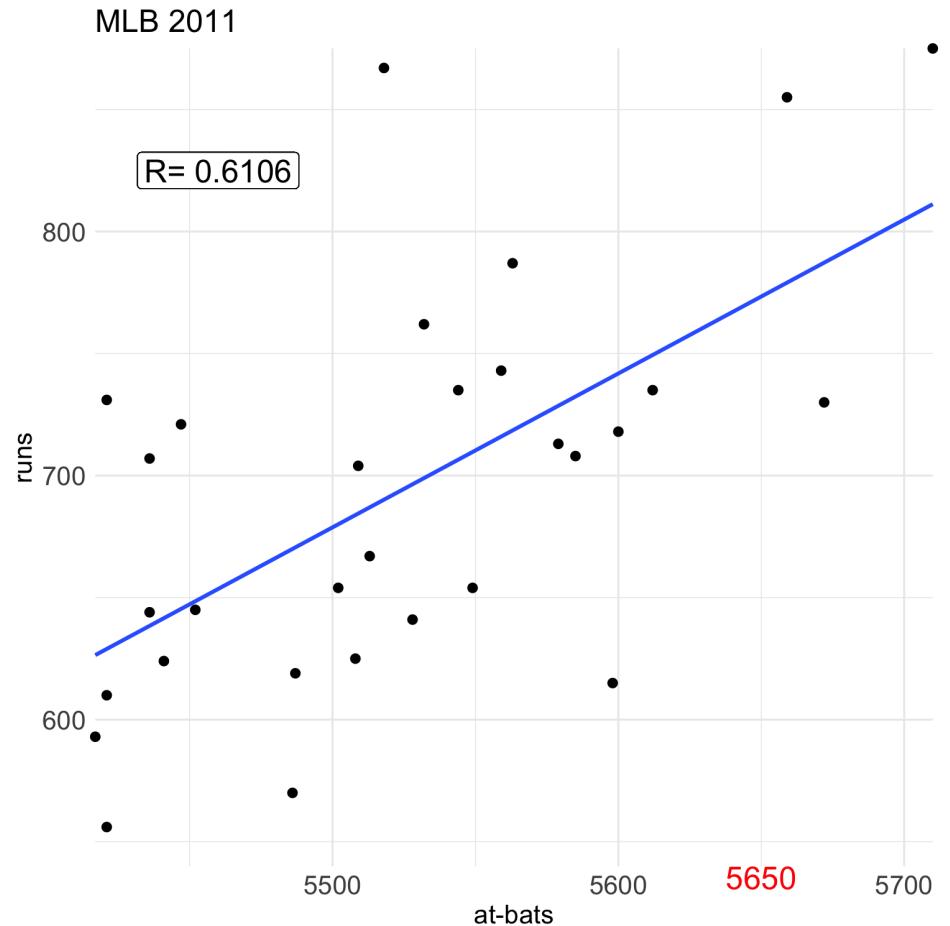
Example: MLB 2011 (pt. 5)

We want to predict the number of runs we can expect to get in a season if our team has 5,650 at-bats.



Example: MLB 2011 (pt. 5)

We want to predict the number of runs we can expect to get in a season if our team has 5,650 at-bats.

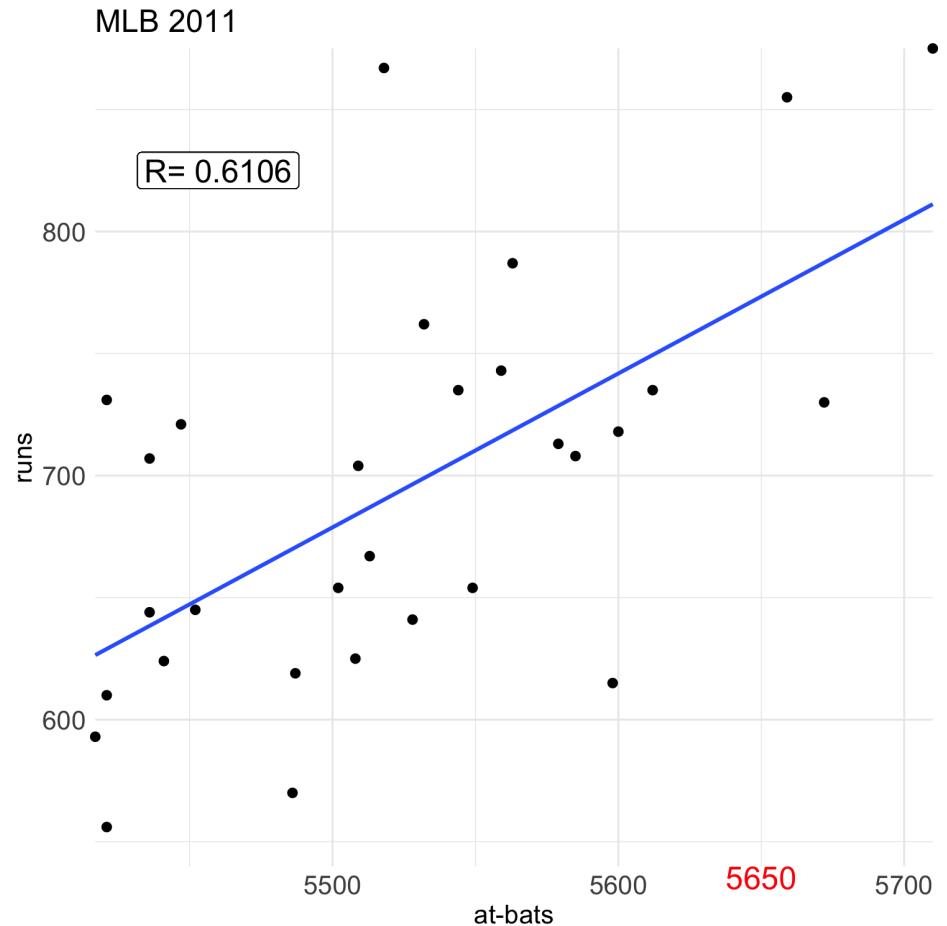


$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

$$\text{Runs}^* = -2789.2429 + 0.6305(5,650)$$

Example: MLB 2011 (pt. 5)

We want to predict the number of runs we can expect to get in a season if our team has 5,650 at-bats.



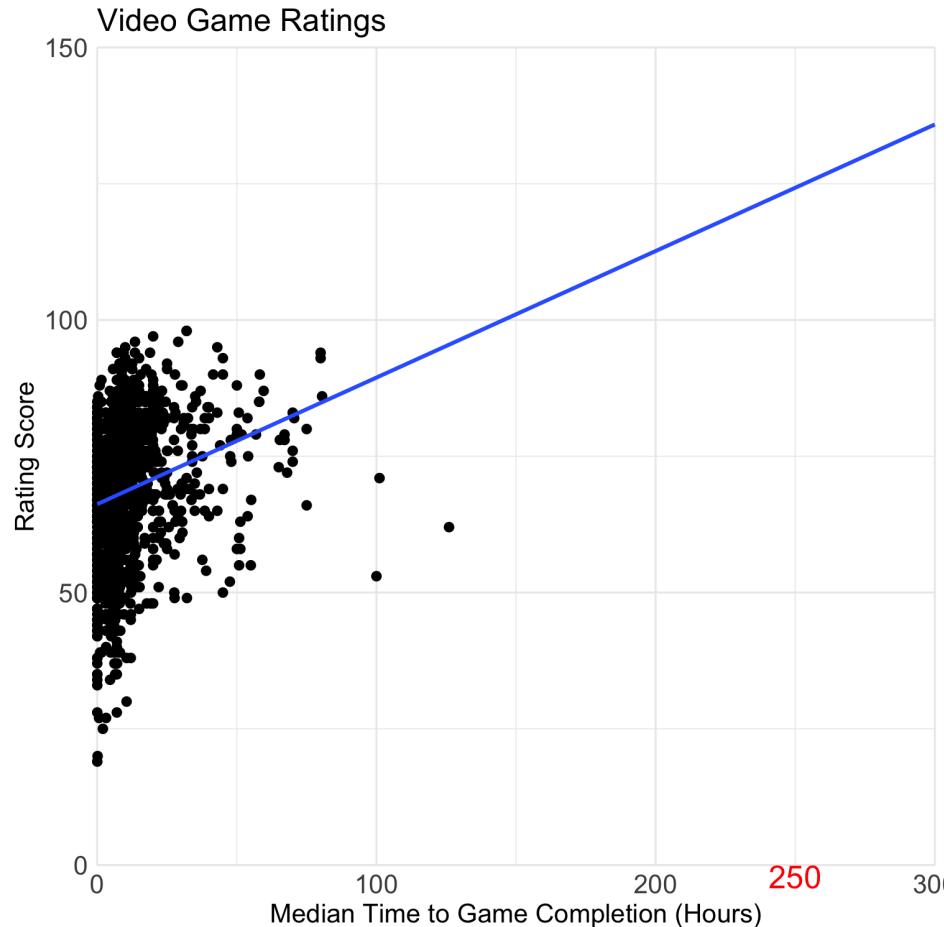
$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

$$\text{Runs}^* = -2789.2429 + 0.6305(5,650)$$

$$\text{Runs}^* = 773.0821$$

Example: Video Game Ratings, (pt. 5)

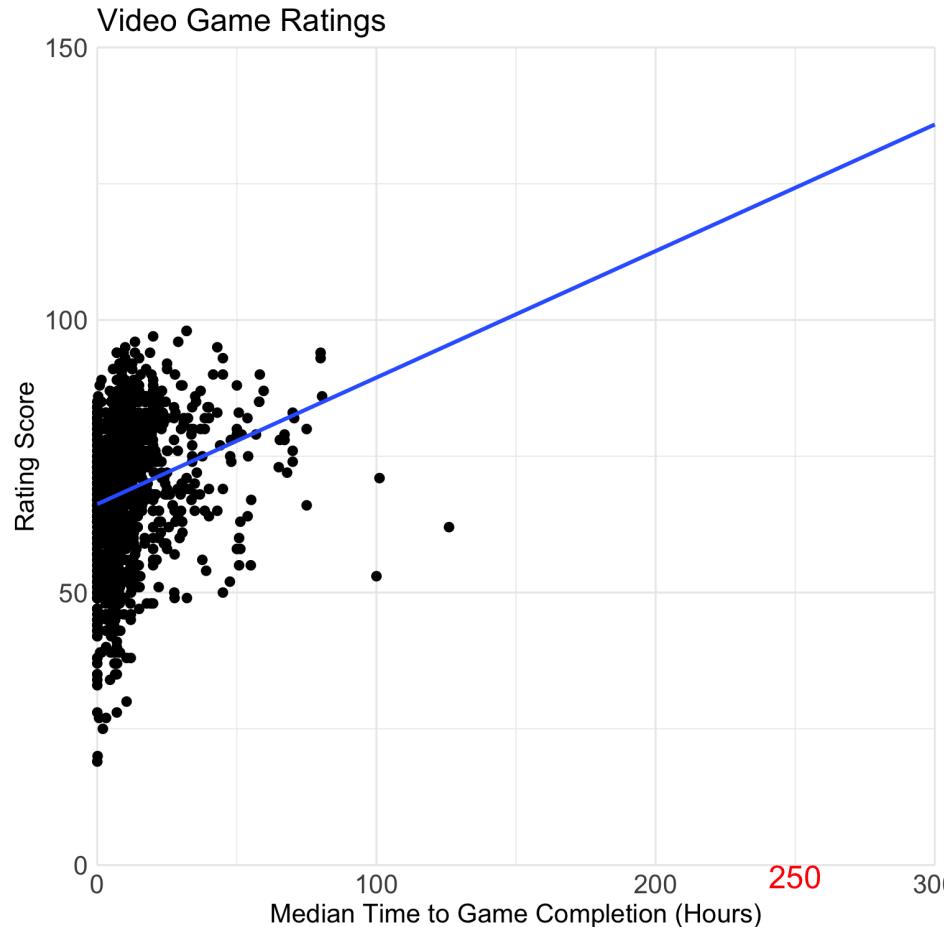
We want to predict the rating score we can expect a game to receive if the median completion time is 250 hours.



$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

Example: Video Game Ratings, (pt. 5)

We want to predict the rating score we can expect a game to receive if the median completion time is 250 hours.

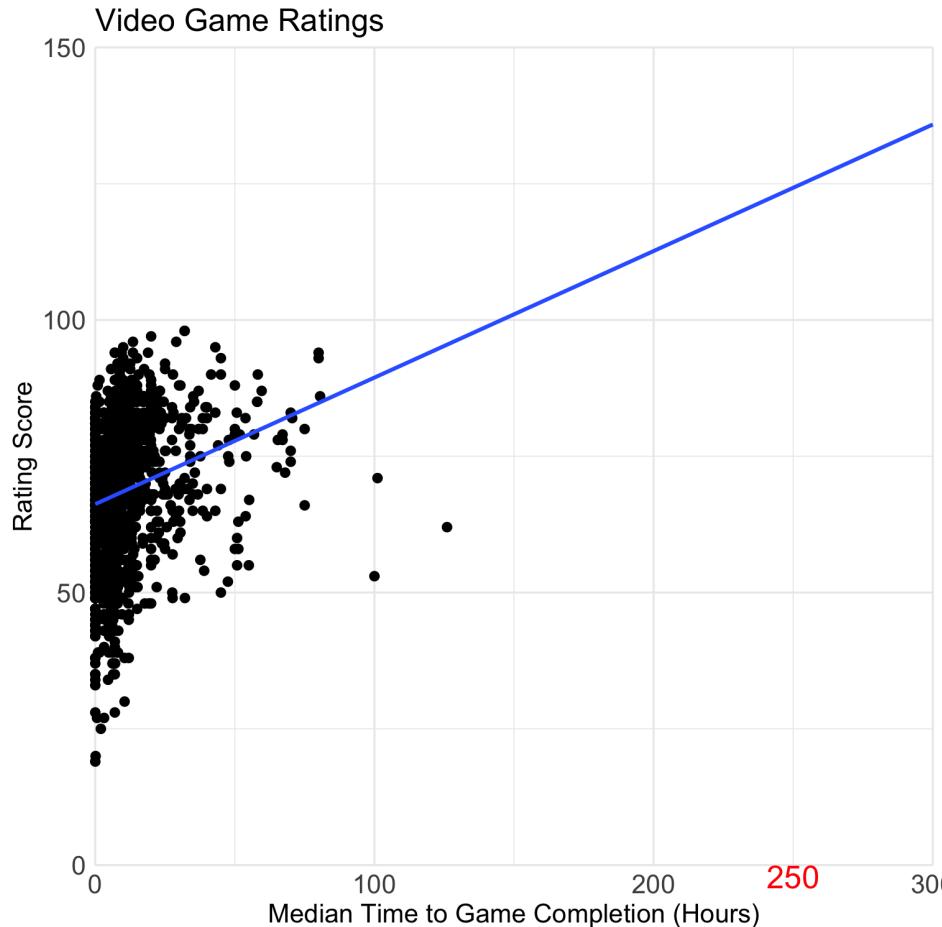


$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

$$\text{Review Score}^* = 66.2225 + 0.2321(250)$$

Example: Video Game Ratings, (pt. 5)

We want to predict the rating score we can expect a game to receive if the median completion time is 250 hours.



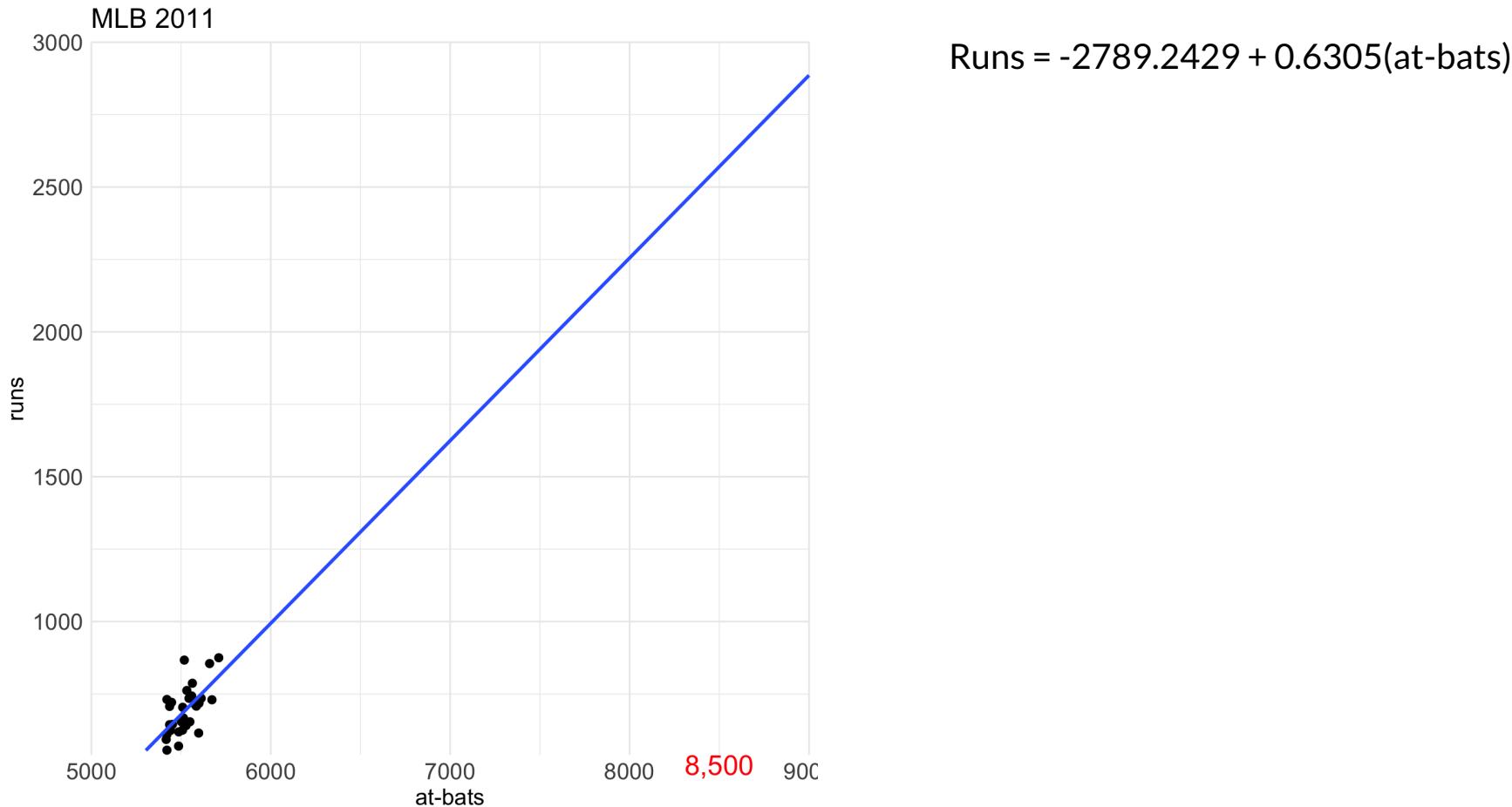
$$\text{Review Score} = 66.2225 + 0.2321(\text{Median time to completion})$$

$$\text{Review Score}^* = 66.2225 + 0.2321(250)$$

$$\text{Review Score}^* = 124.2475$$

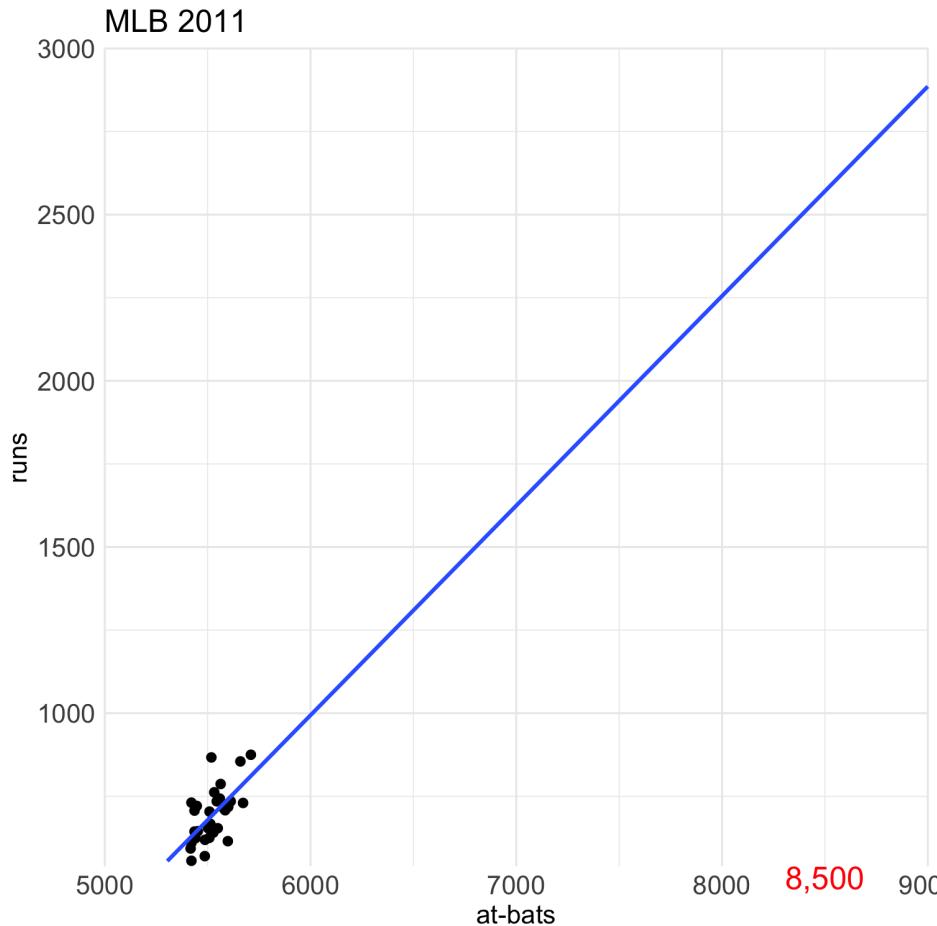
Example: MLB 2011, (pt. 6)

We want to predict the number of runs we can expect to get in a season if our team has 8,500 at-bats.



Example: MLB 2011, (pt. 6)

We want to predict the number of runs we can expect to get in a season if our team has 8,500 at-bats.

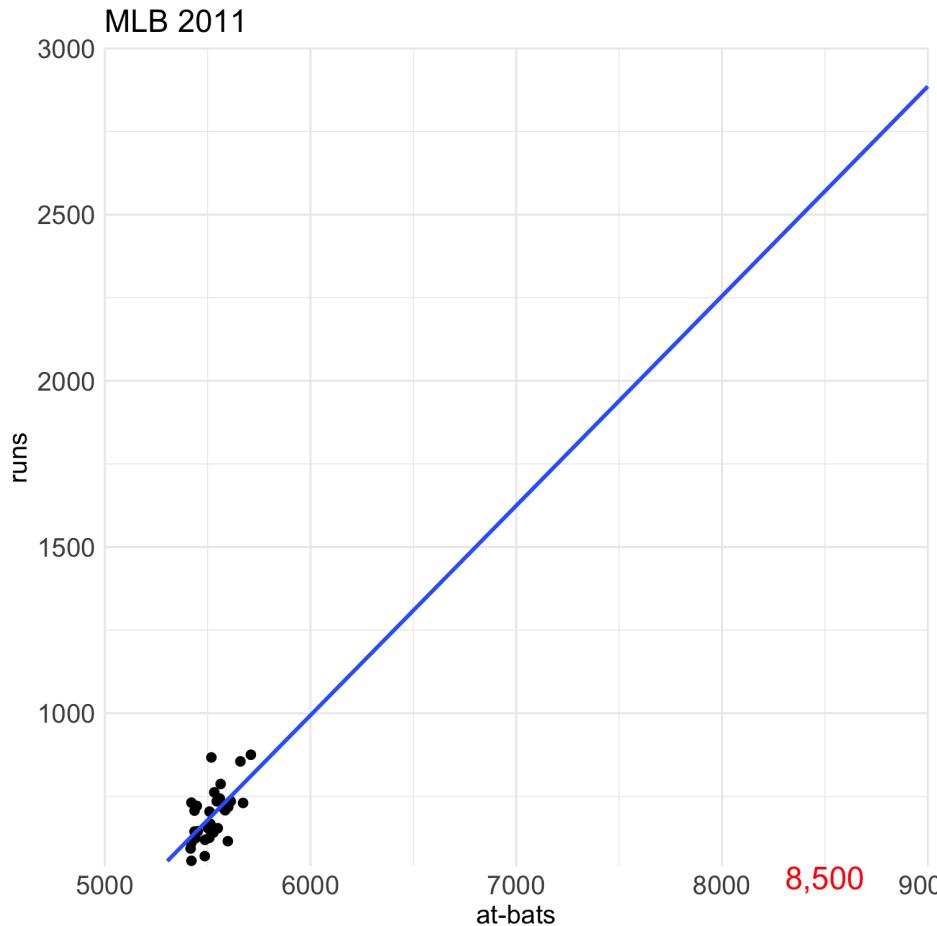


$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

$$\text{Runs}^* = -2789.2429 + 0.6305(8,500)$$

Example: MLB 2011, (pt. 6)

We want to predict the number of runs we can expect to get in a season if our team has 8,500 at-bats.



$$\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$$

$$\text{Runs}^* = -2789.2429 + 0.6305(8,500)$$

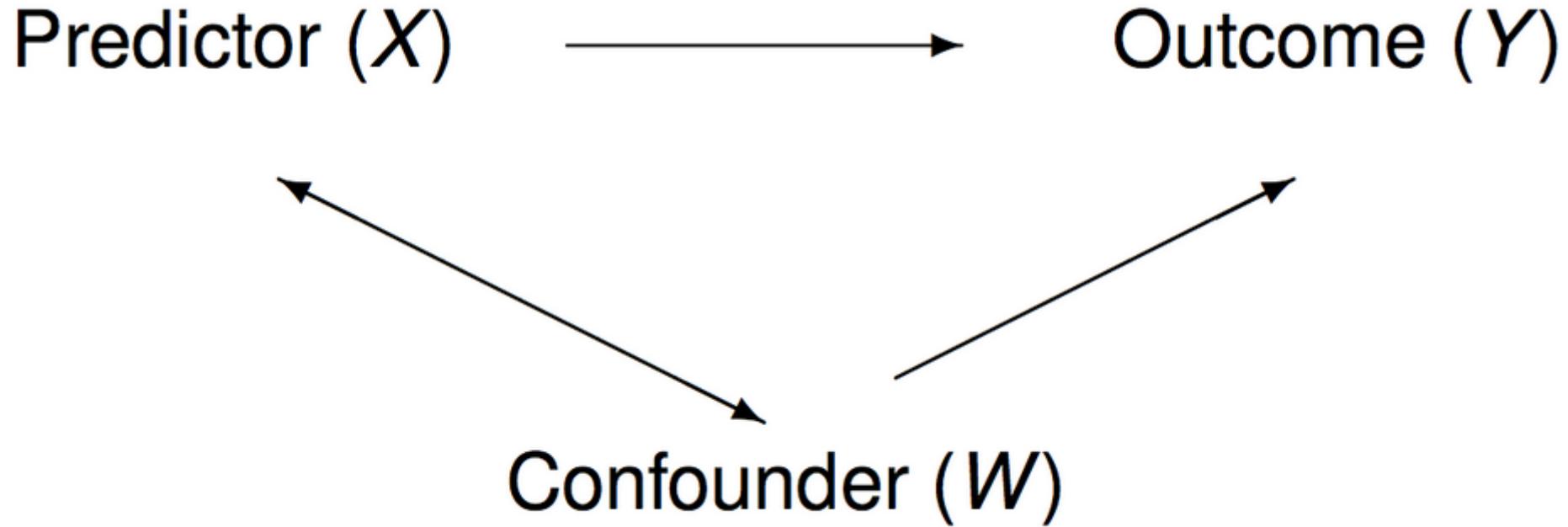
$$\text{Runs}^* = 2570.0071$$

Multiple Regression

- In the real world, one explanatory variable is rarely the sole explanation for how the response variable acts
- **Multiple regression** is an extension of simple linear regression, but instead of one explanatory variable we may have two or more
 - Still have one explanatory **variable of interest** that we care most about how it interacts with the response variable
 - Other explanatory variables in the regression are then known as **adjustment variables**, or variables that we know are related to the response variable in some way

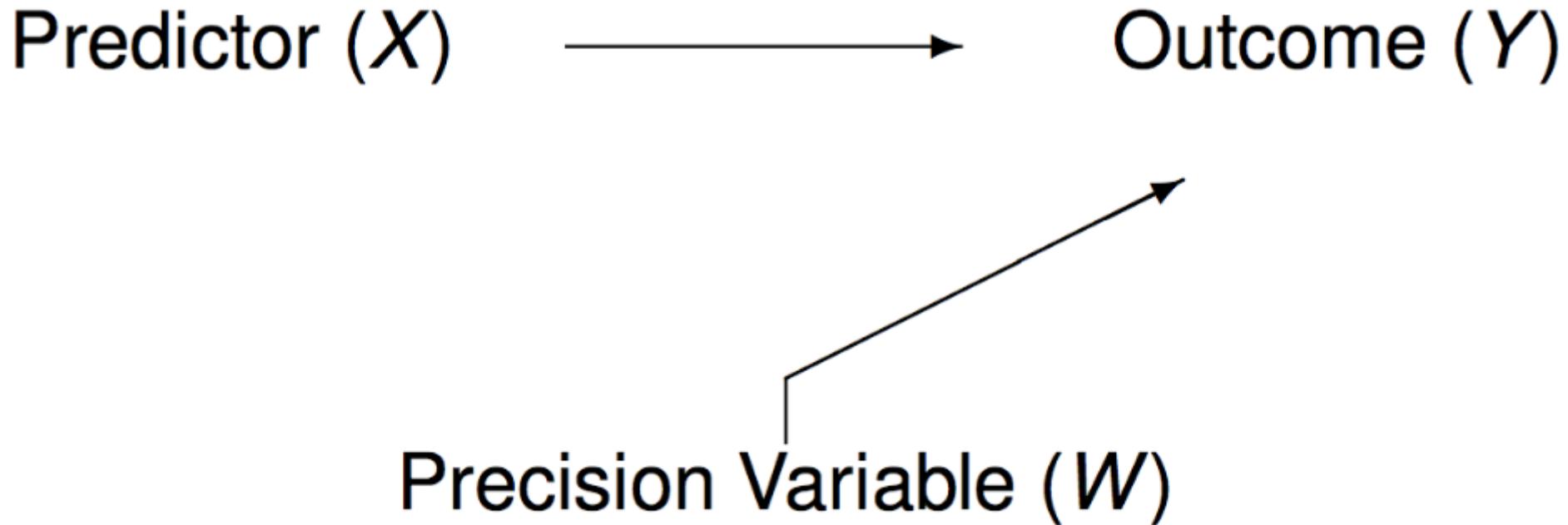
Types of Adjustment Variables

- Confounding variables
 - Related to both the explanatory variable of interest and the response variable
 - Including these in the model will adjust for this variable's confounding effect and **isolate** the effect of the explanatory variable of interest



Types of Adjustment Variables

- Precision variables
 - Just related to the response variable
 - Including these in a regression will make our model **more accurate** and give us more specific effects of the explanatory variable of interest on the response



Example: MLB 2011 (pt. 7)

What else might be related to the number of runs a team scores as well as the number of players they send to bat?

Example: MLB 2011 (pt. 7)

What else might be related to the number of runs a team scores as well as the number of players they send to bat?

- Batting average?
 - Teams with a higher batting average will probably score more runs
 - A low batting average would mean players are frequently struck out, which means less opportunity for players to bat (if 3 players strike out, the teams switch places)

Example: MLB 2011 (pt. 7)

What else might be related to the number of runs a team scores as well as the number of players they send to bat?

- Batting average?
 - Teams with a higher batting average will probably score more runs
 - A low batting average would mean players are frequently struck out, which means less opportunity for players to bat (if 3 players strike out, the teams switch places)
- Home runs?
 - Teams with more home runs score more runs
 - Teams with more home runs might mean fewer strike outs, which means more opportunities for players to bat

Example: MLB 2011 (pt. 7)

What else might be related to the number of runs a team scores as well as the number of players they send to bat?

- Batting average?
 - Teams with a higher batting average will probably score more runs
 - A low batting average would mean players are frequently struck out, which means less opportunity for players to bat (if 3 players strike out, the teams switch places)
- Home runs?
 - Teams with more home runs score more runs
 - Teams with more home runs might mean fewer strike outs, which means more opportunities for players to bat
- Wins?
 - Teams who scored more runs will have more wins
 - Don't say much about at-bats...

Example: MLB 2011 (pt. 7)

Adjusted Model:

- $\text{Runs} = \beta_0 + \beta_1(\text{at-bats}) + \beta_2(\text{batting avg}) + \beta_3(\text{homeruns})$

Example: MLB 2011 (pt. 7)

Adjusted Model:

- $\text{Runs} = \beta_0 + \beta_1(\text{at-bats}) + \beta_2(\text{batting avg}) + \beta_3(\text{homeruns})$
- $\text{Runs} = -293.9292 + -0.0287(\text{at-bats}) + 3768.6013(\text{batting avg}) + 1.2225(\text{homeruns})$

Example: MLB 2011 (pt. 7)

Adjusted Model:

- $\text{Runs} = \beta_0 + \beta_1(\text{at-bats}) + \beta_2(\text{batting avg}) + \beta_3(\text{homeruns})$
- $\text{Runs} = -293.9292 + -0.0287(\text{at-bats}) + 3768.6013(\text{batting avg}) + 1.2225(\text{homeruns})$

For every 1 extra player a team sends to bat during a season, their total runs decrease by 0.0287, on average.

Example: MLB 2011 (pt. 7)

Adjusted Model:

- $\text{Runs} = \beta_0 + \beta_1(\text{at-bats}) + \beta_2(\text{batting avg}) + \beta_3(\text{homeruns})$
- $\text{Runs} = -293.9292 + -0.0287(\text{at-bats}) + 3768.6013(\text{batting avg}) + 1.2225(\text{homeruns})$

For every 1 extra player a team sends to bat during a season, their total runs decrease by 0.0287, on average.

Old Model:

- $\text{Runs} = -2789.2429 + 0.6305(\text{at-bats})$

Example: Video Game Ratings (pt. 6)

What else might be related to a game's rating score, as well as the completion time?

Example: Video Game Ratings (pt. 6)

What else might be related to a game's rating score, as well as the completion time?

- Genre?
- ESRB Rating?

Example: Video Game Ratings (pt. 6)

Adjusted Model:

- Review Score = $\beta_0 + \beta_1$ (Median time to completion) + β_2 (ESRB rating)

Example: Video Game Ratings (pt. 6)

Adjusted Model:

- Review Score = $\beta_0 + \beta_1$ (Median time to completion) + β_2 (ESRB rating)
- Review Score = 66.671 + 0.2313(Median time to completion) - 1.8666(ESRB rating = T) - 0.1361(ESRB rating = M)

Example: Video Game Ratings (pt. 6)

Adjusted Model:

- Review Score = $\beta_0 + \beta_1$ (Median time to completion) + β_2 (ESRB rating)
- Review Score = 66.671 + 0.2313(Median time to completion) - 1.8666(ESRB rating = T) - 0.1361(ESRB rating = M)

When the median completion time increases by 1 hour, a game's rating increases by 0.2313, on average.

Example: Video Game Ratings (pt. 6)

Adjusted Model:

- Review Score = $\beta_0 + \beta_1$ (Median time to completion) + β_2 (ESRB rating)
- Review Score = 66.671 + 0.2313(Median time to completion) - 1.8666(ESRB rating = T) - 0.1361(ESRB rating = M)

When the median completion time increases by 1 hour, a game's rating increases by 0.2313, on average.

Old Model:

- Review Score = 66.2225 + 0.2321(Median time to completion)