

⚠ Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > EDA: Examining Relationships > Case Q→Q: Scatterplots > Scatterplot: Labeled

🔖 Bookmark this page

Scatterplot: Labeled

Learning Objective: Graphically display the relationship between two quantitative variables and describe: a) the overall pattern, and b) striking deviations from the pattern.

A Labeled Scatterplot

In certain circumstances, it may be reasonable to indicate different subgroups or categories within the data on the scatterplot, by labeling each subgroup differently. The result is called a **labeled scatterplot**, and can provide further insight about the relationship we are exploring. Here is an example.

Example: Hot Dogs

Recall the hot dog example from case C→Q, in which 54 major hot dog brands were examined. In this study, both the **calorie content** and the **sodium level** of each brand was recorded, as well as the **type** of hot dog: beef, poultry, and meat (mostly pork and beef, but up to 15% poultry meat). In this example, we will explore the relationship between the sodium level and calorie content of hot dogs, and we will label the three different types of hot dogs to create a labeled scatterplot.

Creating a Labeled Scatterplot



Start of transcript. Skip to the end.



The scatterplot below displays the relationship between the sodium and

calorie content of fifty four brands of hotdogs. Note that in this example there

is no clear explanatory response distinction and we decided to have sodium content as

the explanatory variable and calorie content as the response. The

Video

[Download video file](#)

Transcripts

[Download SubRip \(.srt\) file](#)

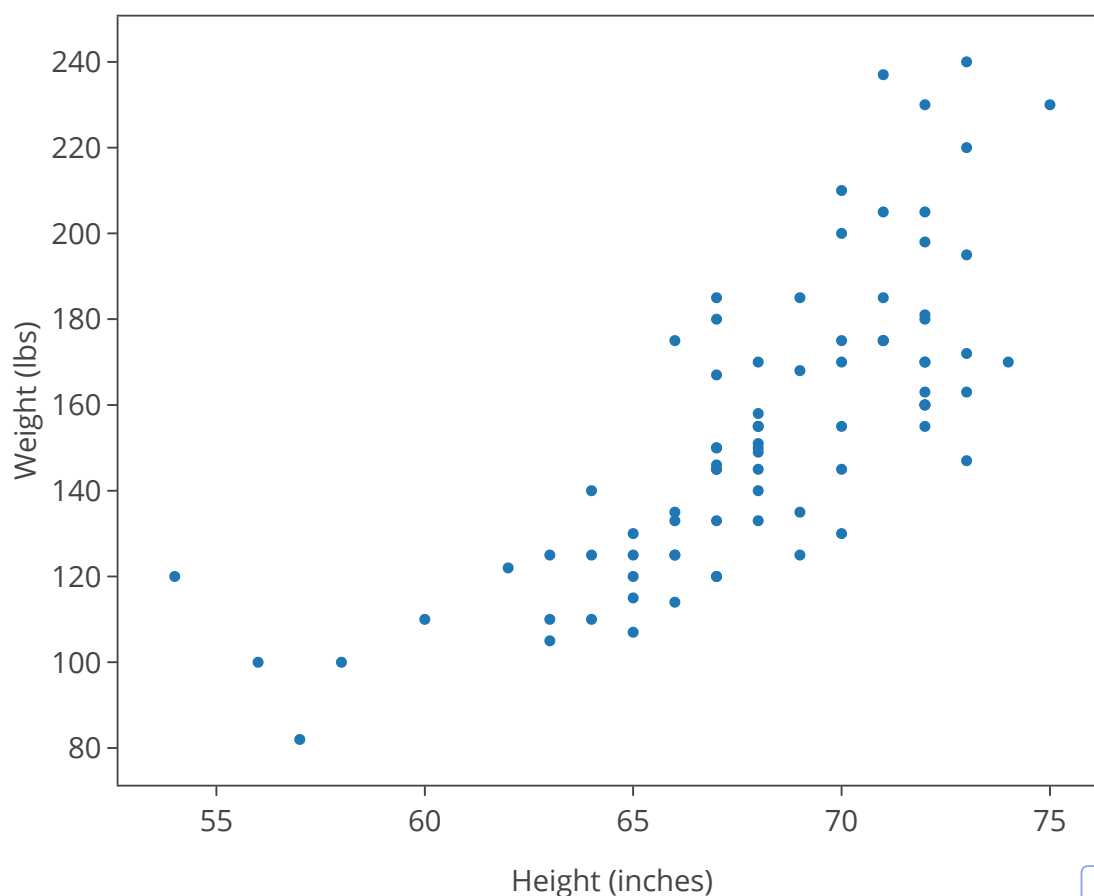
[Download Text \(.txt\) file](#)

Scenario: Height, Weight, and Gender

In this activity, we look at height and weight data that were collected from 57 males and 24 females, and use the data to explore how the weight of a person is related to (or affected by) his or her height. This implies that height will be our explanatory variable and weight will be our response variable. We will then look at gender, and see how labeling this third variable contributes to our understanding of the form of the relationship.

Here is the scatterplot to examine how weight is related to height, ignoring gender.

Height and Weight Scatterplot



Learn By Doing

1/1 point (graded)

What is the direction of the relationship between height and weight?

☒ Positive ✓☐ Negative

Answer

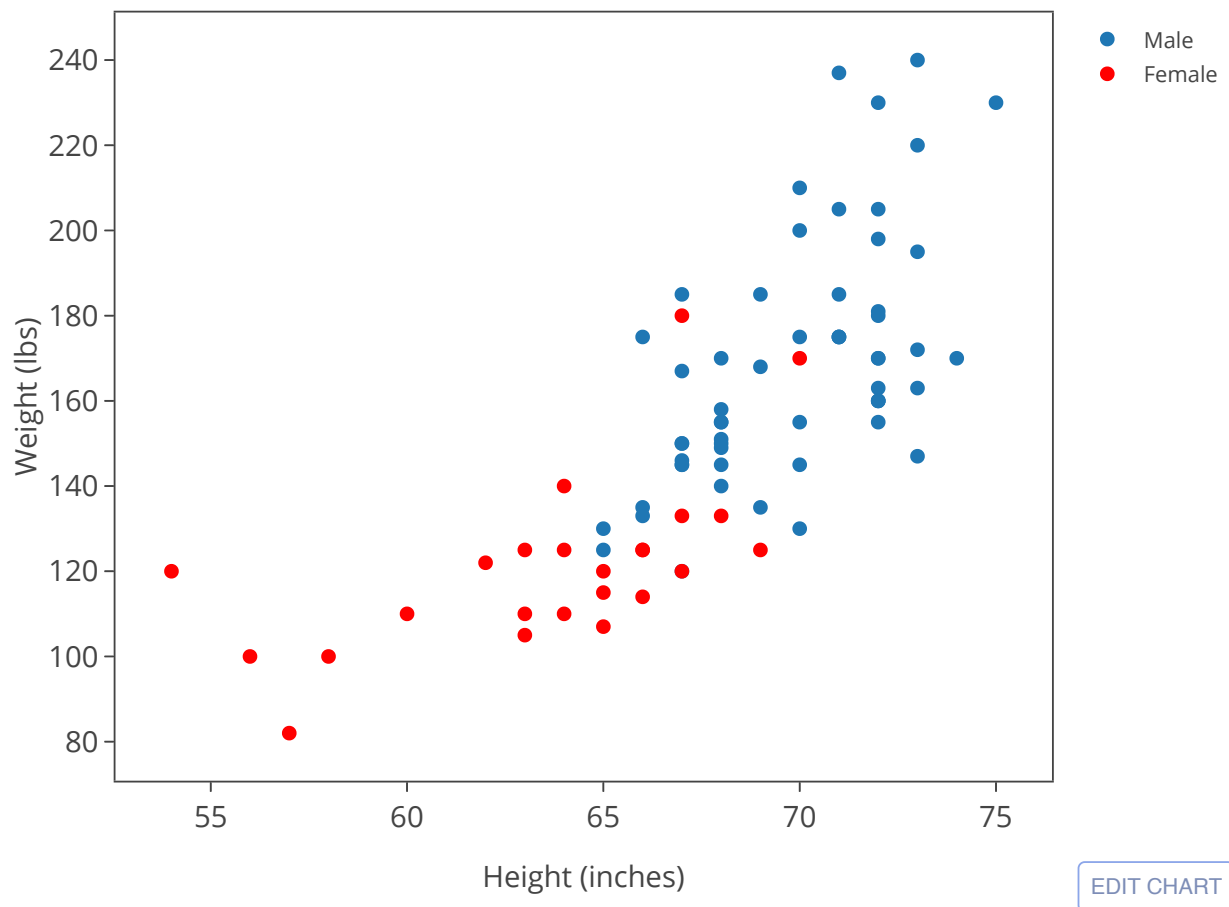
Correct:

The direction of the relationship is positive. In context, this means that individuals who are taller are heavier.

[Submit](#)

So far we have studied the relationship between height and weight for all of the males and females together. It may be interesting to examine whether the relationship between height and weight is different for males and females. To visualize the effect of the third variable, gender, we will indicate in the scatterplot which observations are males (blue) and which are females (red).

Height and Weight by Gender Scatterplot



Learn By Doing

1/1 point (graded)

True or false? The weight of females increases with an increase in height as quickly as the weight of males increases with a corresponding increase in height.

☐ True

☒ False ✓

Answer

Correct:

The main difference between males and females is that the weight of females does not appear to increase as quickly with height as the weight of males does.

Submit

Let's Summarize

- The relationship between two quantitative variables is visually displayed using the **scatterplot**, where each point represents an individual. We always plot the explanatory variable on the horizontal X axis, and the response variable on the vertical Y axis.
- When we explore a relationship using the scatterplot, we should describe the **overall pattern** of the relationship and any **deviations** from that pattern. To describe the overall pattern consider the **direction**, **form** and **strength** of the relationship. Assessing the strength just by looking at the scatterplot can be problematic; using a numerical measure to determine strength will be discussed later in this course.
- Adding labels to the scatterplot that indicate different groups or categories within the data might help us get more insight about the relationship we are exploring.

Open Learning Initiative [↗](#)



[↗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [↗](#).

© All Rights Reserved