Course > EDA: Examining Relationships > Case Q→Q: Scatterplots >
Statistics Package Exercise: Creating a Scatterplot

🔖 **Bookmark this page**

Statistics Package Exercise: Creating a Scatterplot

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

In this exercise, we will:

- learn how to create a scatterplot.

- use the scatterplot to examine the relationship between two quantitative variables.

- learn how to create a labeled scatterplot.

- use the labeled scatterplot to better understand the form of a relationship.

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.

| **R** | **StatCrunch▸** | **TI Calculator▸** | **Minitab▸** | **Excel** |

R Instructions

To open R with the dataset preloaded, right-click here and choose "Save Target As" to download the file to your computer. Then find the downloaded file and double-click it to open it in R.

The data have been loaded into the data frame

```
h
```

. Enter the command

```
h
```

to see the data. There are three variables in

```
h
```

:

```
gender
```

,

```
height
```

, and

```
weight
```

.

The variables are identified as follows:

```
gender
```

: 0 = male, 1 = female.

```
height
```

: in inches.

```
weight
```

: in pounds.

First we will create a scatterplot to examine how weight is related to height, ignoring gender.

To do that in R, copy the following command to R:

```
plot(h$height,h$weight)
```

Again, a good graphic should have labels so lets add x and y-axis labels:

```
plot(h$height,h$weight, xlab="Height (inches)", ylab="Weight
(lbs)")
```

Learn By Doing (1/1 point)

Describe the relationship between the height and weight of the subjects suggested by the data. Consider the pattern of the data—mainly direction and form—and any deviations from this pattern, such as outliers.

Your Answer:

A mostly positive correlation can be observed from the data, save for one outlier whose height is at 54 (lower than the rest) but weight is at 120 (higher than four others!)

Our Answer:

The direction of the relationship is positive. In context, this means that individuals who are taller are heavier. The form of the relationship is curvilinear. Weight seems to increase more and more rapidly with height as we consider taller individuals. We might say that the relationship is moderate in strength, because the points suggest, but do not closely follow, a curvilinear form. There do not appear to be any outliers.

Resubmit   Reset

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 and which are females.

- **R**   **StatCrunch**   **TI Calculator**   **Minitab**   **Excel**

R Instructions

To do that with R, we change the color of the data points representing females to red:

- ```
  plot(h$height,h$weight, xlab="Height (inches)",
  ylab="Weight (lbs)")
  ```

- ```
  points(h$height[h$gender==1],h$weight[h$gender==1],col="
  red")
  ```

You can make a nicer looking plot with males shown in blue, females in red, and labels telling which is which:

- ```
  plot(h$height,h$weight, xlab="Height (inches)",
  ylab="Weight
  (lbs)",col="blue")points(h$height[h$gender==1],h$weight[
  h$gender==1],col="red")legend(55,225, pch=1,
  col=c("red","blue"),legend=c("females","males"))
  ```

Note: To look up more details about

```
legend()
```

, simply type

```
?legend
```

into R and press Enter to get the help information about the function.

Learn By Doing (1/1 point)

Consider how taking gender into account adds to our understanding of the overall form of the relationship. Address any interesting differences between the clusters associated with the two genders. More specifically, address the following two questions:(A) Where do the males and females fall on the scatterplot with respect to height and weight. Explain why you think this is so. (B) Does it look like the

weight of females increases with an increase in height as quickly as the weight of males increases with a corresponding increase in height?

Your Answer:

A.) It can also be noticed that males' points are denser -- it seems that males' heights are much more varied than females'.

B.) The height and weight of males have a stronger linear relationship than those of females. So males' weight will increase higher than those of females' for every unit of increase in height.

Our Answer:

We see that the overall form of the relationship can be thought of in terms of two clusters, one for each gender. Points corresponding to the females cluster in the lower left of the scatterplot, which means that females generally have a lower height and weight than males (due to obvious biological differences between males and females). 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.

[Resubmit] [Reset]