

3.1: If something caused something else to change, what would that look like?

Before we go around determining the causes of happiness, we should prepare ourselves with some analytical tools so that we could identify what causation looks like. If we don't prepare ourselves for what we might find, then we won't know how to interpret our own data. Instead, we need to anticipate what the data could look like. Specifically, we need to know what data would look like when one thing does not cause another thing, and what data would look like when one thing does cause another thing. This chapter does some of this preparation. Fair warning: we will find out some tricky things. For example, we can find patterns that look like one thing is causing another, even when that one thing DOES NOT CAUSE the other thing. Hang in there.

Charlie and the Chocolate factory

Let's imagine that a person's supply of chocolate has a causal influence on their level of happiness. Let's further imagine that, like Charlie, the more chocolate you have the more happy you will be, and the less chocolate you have, the less happy you will be. Finally, because we suspect happiness is caused by lots of other things in a person's life, we anticipate that the relationship between chocolate supply and happiness won't be perfect. What do these assumptions mean for how the data should look?

Our first step is to collect some imaginary data from 100 people. We walk around and ask the first 100 people we meet to answer two questions:

1. how much chocolate do you have, and
2. how happy are you.

For convenience, both the scales will go from 0 to 100. For the chocolate scale, 0 means no chocolate, 100 means lifetime supply of chocolate. Any other number is somewhere in between. For the happiness scale, 0 means no happiness, 100 means all of the happiness, and in between means some amount in between.

Here is some sample data from the first 10 imaginary subjects.

subject	chocolate	happiness
1	1	1
2	1	1
3	2	2
4	2	4
5	4	5
6	4	5
7	7	5
8	8	5
9	8	6
10	9	6

We asked each subject two questions so there are two scores for each subject, one for their chocolate supply, and one for their level of happiness. You might already notice some relationships between amount of chocolate and level of happiness in the table. To make those relationships even more clear, let's plot all of the data in a graph.

Scatter plots

When you have two measurements worth of data, you can always turn them into dots and plot them in a scatter plot. A scatter plot has a horizontal x-axis, and a vertical y-axis. You get to choose which measurement goes on which axis. Let's put chocolate supply on the x-axis, and happiness level on the y-axis. The plot below shows 100 dots for each subject.

```
library(ggplot2)
subject<-1:100
chocolate<-round(1:100*runif(100,.5,1))
happiness<-round(1:100*runif(100,.5,1))
the_df_CC<-data.frame(subject,chocolate,happiness)
ggplot(the_df_CC,aes(x=chocolate,y=happiness))+
  geom_point()+
  theme_classic()
```

run

restart

restart & run all

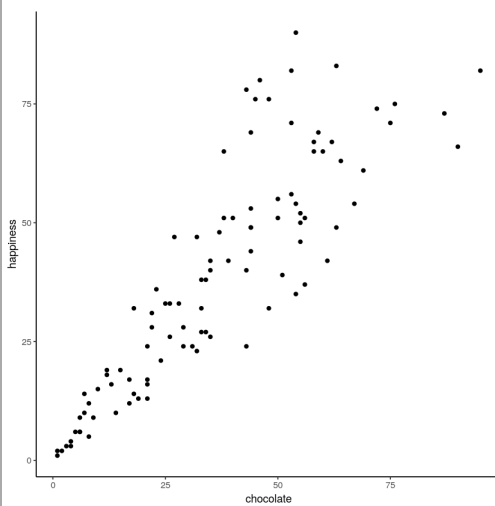


Figure \(\PageIndex{1}\): Imaginary data showing a positive correlation between amount of chocolate and amount happiness.

You might be wondering, why are there only 100 dots for the data. Didn't we collect 100 measures for chocolate, and 100 measures for happiness, shouldn't there be 200 dots? Nope. Each dot is for one subject, there are 100 subjects, so there are 100 dots.

What do the dots mean? Each dot has two coordinates, an x-coordinate for chocolate, and a y-coordinate for happiness. The first dot, all the way on the bottom left is the first subject in the table, who had close to 0 chocolate and close to zero happiness. You can look at any dot, then draw a straight line down to the x-axis: that will tell you how much chocolate that subject has. You can draw a straight line left to the y-axis: that will tell you how much happiness the subject has.

Now that we are looking at the scatter plot, we can see many things. The dots are scattered around a bit aren't they, hence scatter plot. Even when the dot's don't scatter, they're still called scatter plots, perhaps because those pesky dots in real life have so much scatter all the time. More important, the dots show a relationship between chocolate supply and happiness. Happiness is lower for people with smaller supplies of chocolate, and higher for people with larger supplies of chocolate. It looks like the more chocolate you have the happier you will be, and vice-versa. This kind of relationship is called a positive correlation.

Positive, Negative, and No-Correlation

Seeing as we are in the business of imagining data, let's imagine some more. We've already imagined what data would look like if larger chocolate supplies increase happiness. We'll show that again in a bit. What do you imagine the scatter plot would look like if the relationship was reversed, and larger chocolate supplies decreased happiness. Or, what do you imagine the scatter plot would look like if there was no relationship, and the amount of chocolate that you have doesn't do anything to your happiness. We invite your imagination to look at these graphs:

```
library(ggplot2)
subject_x<-1:100
chocolate_x<-round(1:100*runif(100,.5,1))
happiness_x<-round(1:100*runif(100,.5,1))
df_positive<-data.frame(subject_x,chocolate_x,happiness_x)
subject_x<-1:100
chocolate_x<-round(1:100*runif(100,.5,1))
happiness_x<-round(100:1*runif(100,.5,1))
df_negative<-data.frame(subject_x,chocolate_x,happiness_x)
subject_x<-1:100
chocolate_x<-round(runif(100,0,100))
happiness_x<-round(runif(100,0,100))
df_random<-data.frame(subject_x,chocolate_x,happiness_x)
all_data<-rbind(df_positive,df_negative,df_random)
all_data<-cbind(all_data,correlation=rep(c("positive","negative","random"),each=100))
ggplot(all_data,aes(x=chocolate_x,y=happiness_x))+
  geom_point()+
  theme_classic()+
  facet_wrap(~correlation)+
  xlab("chocolate supply")+
  ylab("happiness")
```

run restart restart & run all

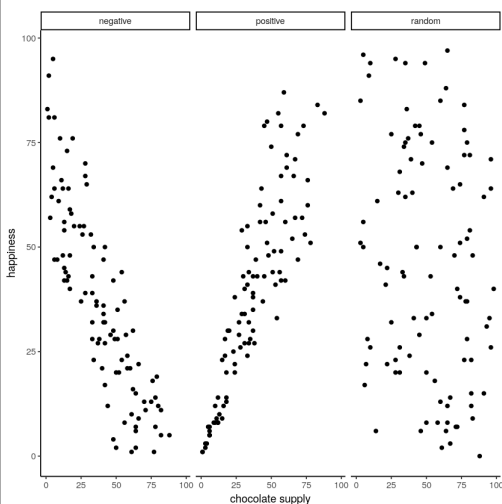


Figure \(\backslash\{PageIndex\{2\}\}): Three scatterplots showing negative, positive, and zero correlation.

The first panel shows a negative correlation. Happiness goes down as chocolate supply increases. Negative correlation occurs when one thing goes up and the other thing goes down; or, when more of X is less of Y, and vice-versa. The second panel shows a positive correlation. Happiness goes up as chocolate supply increases. Positive correlation occurs when both things go up together, and go down together: more of X is more of Y, and vice-versa. The third panel shows no correlation. Here, there doesn't appear to be any obvious relationship between chocolate supply and happiness. The dots are scattered all over the place, the truest of the scatter plots.

Note

We are wading into the idea that measures of two things can be related, or correlated with one another. It is possible for the relationships to be more complicated than just going up, or going down. For example, we could have a relationship that where the dots go up for the first half of X, and then go down for the second half.

Zero correlation occurs when one thing is not related in any way to another things: changes in X do not relate to any changes in Y, and vice-versa.

This page titled 3.1: If something caused something else to change, what would that look like? is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Matthew J. C. Crump via source content that was edited to the style and standards of the LibreTexts platform.