

## MIT BLOSSOMS Lesson Teacher Guide

### Do Credit Cards Make You Gain Weight? What is Correlation, and How to Distinguish It from Causation

By Kamilya Tazhibayeva, Postdoctoral Researcher, MIT

Hello, and thank you for your interest in this lesson. As its title suggests, this lesson introduces the concepts of correlation and causation, and discusses the difference between the two. The main objective is to help the students become aware of this difference, and show them how they can critically analyze the informational value of a correlation. The goal is to make the students comfortable with questioning how meaningful a correlation is, and how much actual truth is behind it.

Another objective is to help the students see that a correlation is not an abstract concept. Rather, there are plenty of correlations in our everyday life. The goal is to show the students that a correlation can sometimes be a useful tool for finding out more information. By the end of the lesson the students will hopefully learn to approach correlations with healthy skepticism, neither taking them on faith nor dismissing them outright.

Topics covered in this lesson include:

- Direct and indirect causation
- Omitted variables behind a correlation
- Spurious correlation
- Reverse causation

Students will practice identifying and distinguishing between different types of correlation and causation during suggested activities.

By the end of the lesson, the students should understand that:

- Correlations are a part of life
- Correlations can sometimes be misleading
- Correlations can sometimes be informative when properly analyzed
- How to evaluate a correlation's plausibility
- What an informative correlation does and does not tell us
  - It does point us towards possible explanations
  - It does not prove or imply causation

The lesson includes an optional two-page colored handout that can be used during Segment 2 of the video. It includes brief definitions and graphic examples of different types of correlations that are introduced in Segment 2. It should be helpful for students to have a printout of these visual examples.

Most of the suggested activities involve student discussions in small groups of two-three people. This is done so that each student has to voice an opinion, which is generally hard to achieve in a class-wide discussion. Having to speak up prompts a student to organize and present their thoughts and arguments more coherently than would be the case if they only had to think about a question for themselves.

For most of these discussions, it would generally be beneficial to get the ball rolling by suggesting a few sample questions to the class at the beginning of the activity, or perhaps to take a few suggestions from the class if time permits. The purpose is to help the students get started with their discussions, although care should be taken to do this in a way that encourages the students to explore, rather than prescribes them a limited range of “acceptable” questions. Below, I suggest some prompting questions for each activity, though these might need to be adjusted a bit to the realities the students are familiar with. If the classroom’s layout permits, it is a good idea for the students to move around and switch pairs between different activities.

While the students are engaged in their discussions, it might be helpful to pace between the students, keeping an eye on how the discussions are going, and noting any widespread concerns or misunderstandings – these could be then addressed and explained to the whole class. Some groups of students might need a bit of personal attention to help them with their discussions.

The rest of this document briefly goes over each suggested activity.

**A note to the teacher:** There are statistical tests that allow us to quantify the direction and strength of a correlation between two-variables. So, one does not only look at scatter plots to “see” relationships that appear to suggest correlation, but one can measure it and validate it with statistical tests.

First, we want to measure the strength of any relationship and its direction, positive or negative. It is the **Correlation Coefficient** that we are after. Its value can be anywhere between +1 and -1, where +1 or -1 indicates one-for-one perfect movement between the two variables, positively or negatively, and 0 indicates no relationship whatsoever. To complicate matters, there are several different types of correlation coefficients. The most widely used for linear relationships, and the one pertaining to this BLOSSOMS lesson, is called the Pearson  $r$  correlation coefficient. Its relatively simple formula and illustrative applications are given in several public web sites. We suggest the YouTube video “How to Use Excel-The PEARSON Function,” <https://www.youtube.com/watch?v=JO-Gc5bEG70> , where students can do the work effortlessly using Microsoft Excel.

Second, we can do statistical tests of significance, involving the “T statistic” and its “p value.” Such “tests for significance” can be difficult for beginning students (and for all of us!) due to the double negative setting of statistical hypothesis tests. A good introductory tutorial of the theory is here: <https://www.leansigmacorporation.com/correlation-coefficient-minitab/> The computations can be done easily by students in Microsoft Excel, with the students first watching this YouTube 4.5-minute tutorial: <https://www.youtube.com/watch?v=Ev86DMtLXOk> We suggest going very light on the hypothesis-testing theory and moving directly to the Microsoft Excel computation.

A final note of caution: Whatever the results of these computations, we retain our golden rule: ***“Correlation does NOT imply causation.”***

## Activity 1

This activity is done in small groups of 2-3 students. The purpose of this activity is to prompt the students to actively discuss and to question a real-life correlation. The example used in this activity demonstrates the presence of correlations in something so routine as grocery shopping. At the same time, it illustrates the sometimes provocative and unexpected nature of correlations and the relationships they seem to suggest.

Possible prompting questions are:

- When do you usually use a credit card?
- When are you more likely to pay in cash?
- When are people more likely to buy junk food?

**Recommended time:** 2-3 minutes

## Activity 2

This is the most interactive and the longest activity during this lesson. It will illustrate visually how the height of students in the class and the length of their hair correlate, by graphing one against the other. The idea is that boys tend to have shorter hair than girls and are on average taller than girls. One goal of this activity is to get the students actively engaged, which generally helps with retaining new information better. Another goal is to demonstrate how widespread correlations are, that the students' own lives can be described by them. The third goal is to get the students to analyze actively what they expect the correlation to be, and why. This lets them experience in practice what goes on behind a correlation.

This activity requires a blackboard (or a whiteboard or similar) visible to the class. Alternatively, a large, poster-size piece of paper on an easel or taped to a wall can be used. One part of the board will be used for the graph, and another part to note discussion results. If the board space is limited, the teacher can ask one of the students to note down the main discussion results in their notebook and then read them out loud (more on this below).

First, the teacher distributes blank index cards (or blank sheets of paper) to all students. Each student writes down on the card their height and hair length. While most students are likely to know their height, they are less likely to know the length of their hair. To address this, and also to make graphing easier, the students can just choose one of several ranges for their hair length. The teacher should write down these ranges on the blackboard or somewhere similar, where the students will be able to see it. Here is an example with five categories of hair length, and it can be adjusted based on the class composition, or converted to measurement units students are most familiar with:

- < 2 (shorter than 2 inches)
- 2-4 inches
- 4-6 inches
- 6-8 inches
- >8 (longer than 8 inches)

The following rules will be displayed on the video screen during this activity:

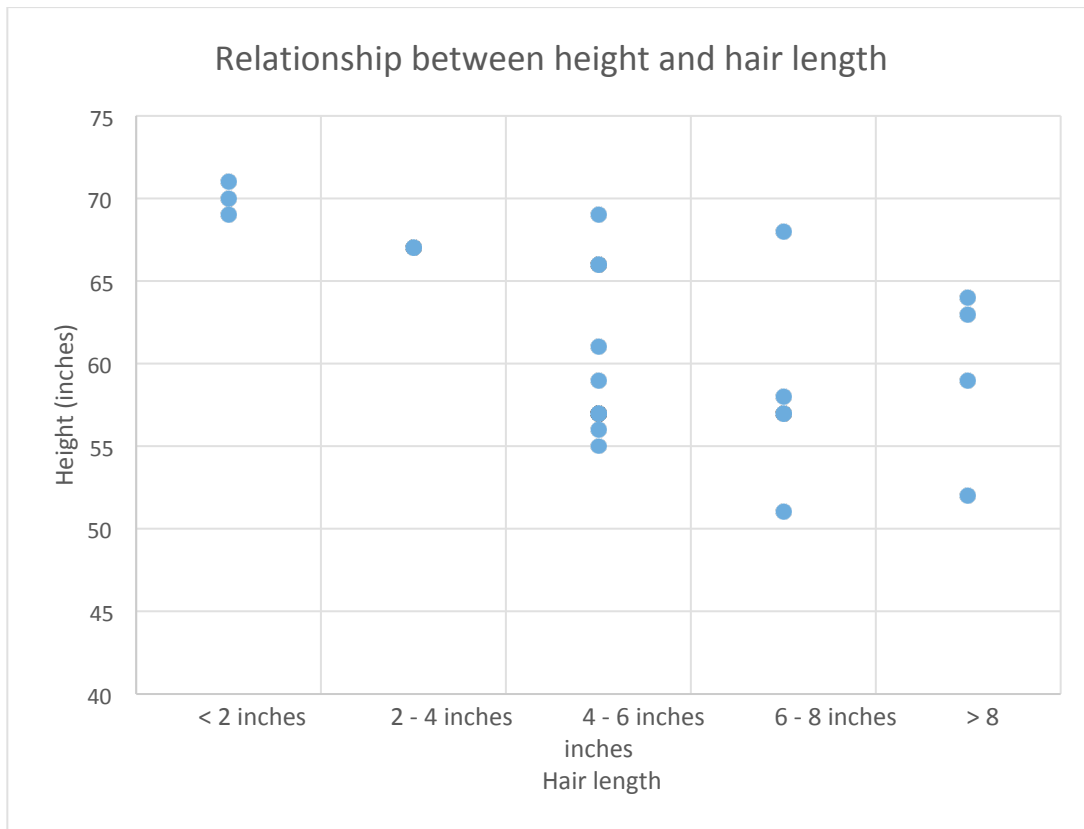
- Write down your height in cm/inches
- Write down your hair length as instructed by your teacher

- If you don't know the exact value, make your best guess
- Don't put any other information on the card

The students then return the index cards to the teacher, without showing them to each other.

While the students are filling out the cards, the teacher prepares the axis for the graph on the blackboard, with hair length intervals on the x-axis and height on the y-axis. To make the visual illustration of the correlation more prominent, it is better to start the height axis at 40 inches (rather than 0).

Once everyone turns in their cards, the teacher asks one student to graph the entries from the cards on the blackboard, using prepared axis. For each card, the student should put a mark (a "x" or a "o") in the corresponding spot on the graph. The resulting graph will look something like this:



While the student is doing this, the teacher engages the class in a short discussion:

- Do the students think these two characteristics – height and hair length – are connected in any way?
- Do the students expect them to be correlated? Why or why not? If yes, would they expect this correlation to be positive or negative? Strong or weak?
- If some students expect there to be a correlation, do they think it is based on causation? Why or why not? If yes, is this causation direct or indirect?

The teacher writes down on the board the main conclusions as the class reaches them after discussing each point (e.g., yes/no on correlation, positive or negative, etc.)

When the graph is finished, the teacher together with the class compares their discussion results with the graph – do they see the results they have expected? Which points did they anticipate correctly? Which points didn't they get right? What might be the reasons for that?

Next, if time permits, it would be good to have a short discussion with the class about the usefulness of the information provided in the graph:

- How precise is this information?
- Can they think of an example of how to use it in real life? After letting the students think about it for a moment on their own, the teacher can lead them with an example: Who would find such information useful? Hairdressers? Wig makers? No one?

**Recommended time:** about 15 minutes. 5 minutes for card collection, 5 minutes for the first discussion while the graph is being drawn, 5 and 5 minutes for the second, post-graphing, discussion.

### Activity 3

This is a paired activity. In this activity the students examine possible logical connections between two seemingly unrelated events, and evaluate how valid these connections are. By now, the students should have a general idea of how to approach such a discussion. If the class seems confused, the following prompting questions can be used to start the discussion and lead it in the right direction:

- Does it seem likely that these two events are connected directly?
  - Can ice cream sales cause people to commit more crimes?
  - Can higher crime rates make people eat more ice cream?
- Can these two events be connected indirectly?
  - What does ice cream associate with?
  - When we think of high crime rates, what do we tend to imagine, what types of crimes? When and where do these generally seem to take place?

It is important to keep a balance between encouraging the discussion and at the same time not giving away the answers, so that the students get to practice their analytical skills.

**Recommended time:** 3-4 minutes

### Activity 4

This is a paired activity.

**Recommended time:** 4 minutes in total, so each student gets 2 minutes. To ensure that each student gets the same amount of time to make their case, it is a good idea to notify the class once 2 minutes elapse and it is time for the students in each pair to switch turns.

### Activity 5

This is a paired activity.

**Recommended time:** 4 minutes in total, so each student gets 2 minutes. To ensure that each student gets the same amount of time to make their case, it is a good idea to notify the class once 2 minutes elapse and it is time for the students in each pair to switch turns.

**Side note:** Segment 4 discusses in detail the case that the students worked on in the preceding Activity 3. Depending on time availability and on how easy or confusing the students have found

Activity 3, it might be fruitful to start Activity 4 by first having a quick class-wide discussion on what the students have thought of the explanations provided in Segment 4, whether or not they agree with the proposed conclusion, and why.

### **Activity 6**

This is a paired activity. The goal of this activity is for the students to relate the lesson's concepts to their own experience. It might be helpful to open this activity with a relatable example of your own. It could be from your life, or your niece's or nephew's, or your friend's – whatever you find most appealing and effective.

Here is an example from my life: Last fall I was participating in a weekly evening seminar. It took place from 7 to 8 pm every Wednesday for several months. At one point I noticed that I was particularly tired on Wednesday evenings. I couldn't understand why. The only thing I did differently on Wednesdays from other week days was the seminar. However, I really enjoyed the seminar, and it wasn't tiring. Then I realized that, because the seminar went from 7 to 8 in the evening, and it took me half an hour to get to and from it, I ended up eating dinner much later on Wednesdays than on other days. I thought that that might explain my tiredness – late dinner meant I didn't have enough energy left from lunch, and so I was exhausting myself. I started snacking right before the seminar, to get some extra energy to last me until dinner, and my tiredness went away.

**Recommended time:** 7 minutes in total: 3 minutes for the students to think of an example, 2 minutes for each student to share their experience. To ensure that each student gets the same amount of time to make their case, it is a good idea to notify the class once 2 minutes elapse and it is time for the students in each pair to switch turns.