Data is Beautiful - Quiz

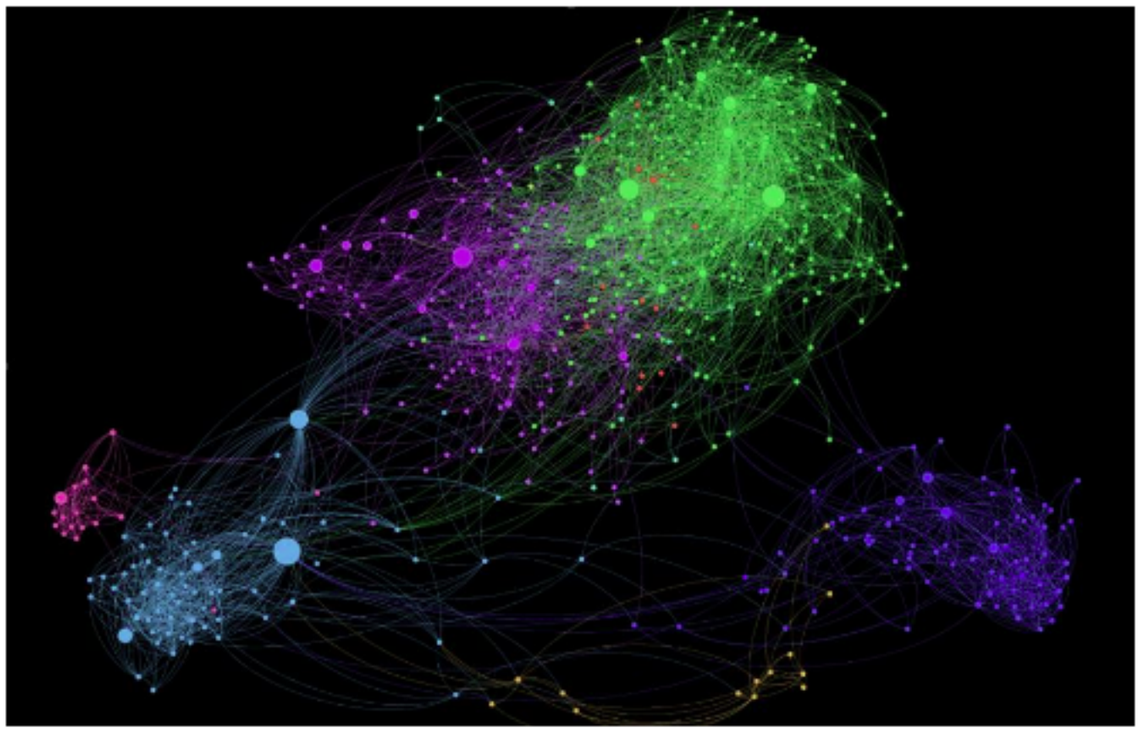
Bookmark this page

Finger Exercises due Sep 14, 2020 19:30 EDT **Past Due**

Question 1

1 point possible (graded)

Which of the following pieces of information are depicted by the network diagram discussed in class? (Check all that apply)



- [ ] Existence of links between people in the network, indicated by connecting lines of any color ✔
- [ ] How connections were made such as working together or attending school together, as indicated by the color of the nodes ✔
- [ ] Population of each village/neighborhood in the Eastleigh, indicated by the size of node
- [ ] Importance of each person in the network, indicated by the size of the node ✔

Explanation

Professor Duflo provides an example of data that is presented in a way that is both beautiful as well as informative. As discussed in class, each person or "node" in the network is represented by a circle, where the size of the circle depicts the relative importance of that node. Connections between nodes are depicted by the connecting lines. Different coloring is used to depict distinct groupings of nodes in the network. This image provides a prime example where the most interesting elements of a dataset are depicted in a way that is both informative and beautiful. For more information, see "What can we learn about Somalis from their Facebook networks?" by Kimo Quaintance.

Submit You have used 0 of 2 attempts

Show Answer

Answers are displayed within the problem

Question 2

1 point possible (graded)

Using the Facebook example, which of the following are discussed as ways you might define “importance” in the social network? (Check all that apply)

- [x] Number of connections to others in the network ✔
- [ ] When someone created their account
- [x] Number of links to others who are important in the network ✔
- [ ] Intensity of use of the network (e.g. the number of pieces of content posted).

Explanation

Number of connections to others within a network and the number of links to others who are important in the network are both discussed as possible ways to define “importance” in a social network. Professor Duflo brings up the example of the original search algorithm used by Google, where pages were ranked highly in part if they were linked to from pages that in turn were linked to by a large number of pages, known as a measure of “eigenvector centrality.”

Submit

You have used 0 of 2 attempts

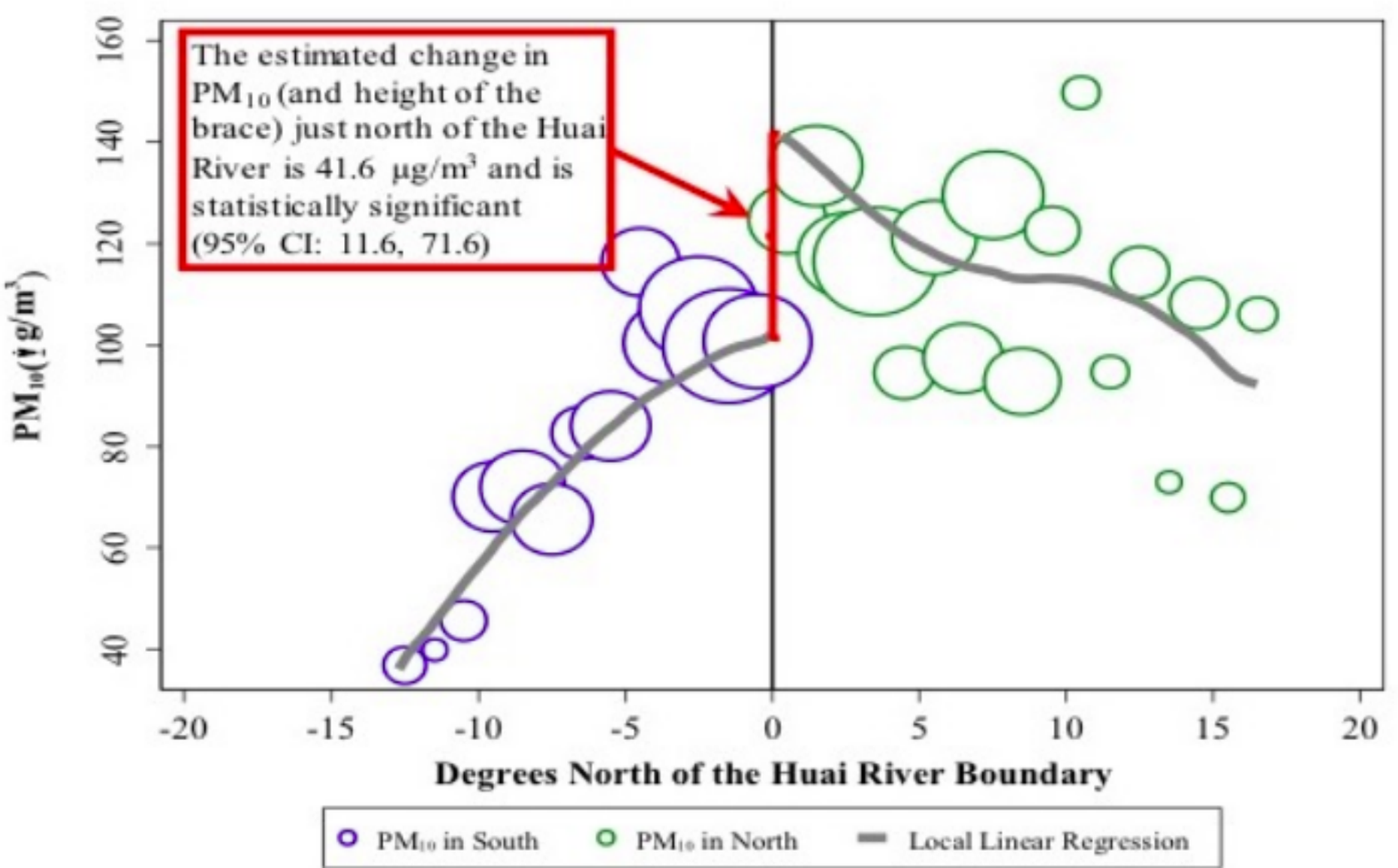
Show Answer

Question 1

0.0/3.0 points (graded)

Fill in the blank, based on the two-dimensional depiction of the difference in pollution levels to the North and South of Huai River shown below.

Figure 2
Particulate Matter Levels (PM_{10}) South and North of the Huai River Boundary



The Huai River boundary itself is depicted by _____.

Select an option **Answer:** black vertical line

Particular matter everywhere levels north of the Huai River are depicted by _____.

Select an option **Answer:** all green bubbles

Particular matter everywhere levels to the South of the Huai River are depicted by _____.

Select an option **Answer:** all blue bubbles

Explanation

The diagram shown in class provides a wealth of insightful and relevant information in a relatively small space. The river itself is represented by the vertical line at 0 degrees of the Huai River Boundary. The blue bubbles to the left of the vertical line represent average levels of PM_{10} to the South of the Huai River Boundary, while the green bubbles represent average levels of PM_{10} to the North of the Huai River Boundary. In each case, the size of the bubble represents the size of the population in each location. The grey lines represent the relationship between distance from the Huai River boundary and particulate matter, with the only assumption about the structure of their relationship being that there may be a difference in the structure of the relationship among points to the North and points to the south of the river boundary

Submit You have used 0 of 2 attempts

Show Answer

Answers are displayed within the problem

Question 2

0.0/1.0 point (graded)

Suppose your hypothesis is that there is **no** significant difference in Particulate Matter Levels (PM_{10}) between the areas north and south of the Huai River. Based on the information presented in Figure 2, would you **reject** or **not reject** this hypothesis?

- ◯ Reject ✔
- ◯ Not reject

Explanation

Reject. The graph of Particulate Matter Levels (PM_{10}) comparing areas to the North and to the South of the Huai River shows a distinct jump. Therefore, you would **reject** that there is no significant difference in PM_{10} levels. Hypothesis testing will be covered in more depth in Module 6!

Submit | You have used 0 of 1 attempt | Show Answer

ℹ Answers are displayed within the problem

Finger Exercises due Sep 14, 2020 19:30 EDT **Past Due**

Question 1

0.0/1.0 point (graded)

Which of the following are identified as the primary problems in the original pollution audits system in Gujarat that was discussed in class?

- ◯ Concentration of market power among a small group of auditors
- ◯ Conflict of interest and perverse incentives ✔
- ◯ Mismeasurement and lack of standard metrics among auditing firms
- ◯ Lack of clear and consistent reporting requirements for auditing firms

Explanation

In the original pollution audits system in Gujurat, firms were able to select and hire private auditors to produce reports that indicated whether they were in compliance with the pollution regulations or not. This introduced a perverse incentive for auditing firms to produce audit reports that would show the firm was in compliance, whether or not that was the case.

Submit You have used 0 of 2 attempts

Show Answer

Question 2

0.0/1.0 point (graded)

What solution did the researchers propose to potentially fix this problem? (See Question 1)

- ◯ The government would hire auditors directly and firms would select from this pool of auditors
- ◯ Firms would be audited by three auditors, and the government would intervene if the three reports did not match
- ◯ Firms would select the auditors and the government would pay for the audit fees
- ◯ Firms would pay into a central pool, auditors from the central pool are randomly assigned to firms ✔

Explanation

The researchers identified that the conflict of interest arose from the fact that firms were directly hiring auditors. The researchers proposed to sever the link between the firms and the auditors by setting up a system whereby firms would pay into a central pool, and then auditors would be randomly assigned from the pool to audit each firm. This reduced the incentive for firms to select and pay auditors to deliver a positive report.

Submit | You have used 0 of 2 attempts | Show Answer

ℹ Answers are displayed within the problem

Question 1

0.0/1.0 point (graded)

Out of the pool of audit-eligible firms, one portion of firms was assigned to the treatment group (where firms would pay into a central pool and auditors would be assigned from that pool to firms) and one portion was assigned to the control group (business as usual). How was this assignment done, and why?

- ◯ Randomly, to make the experiment less complicated
- ◯ Randomly, to ensure that there was (in expectation) no systematic differences in the characteristics of firms in treatment and firms in the control groups ✔
- ◯ Firms were allowed to choose, with the idea being that firms would be more likely to comply with the researchers if they were allowed to choose
- ◯ Researchers chose which would be in the treatment and which would be in the control, based on which firms were most likely to benefit from the treatment

Explanation

Firms were **randomly** assigned to either the treatment or control groups, which was a crucial element of the research design. Random assignment was important to ensure that there were no systematic differences between the firms that were assigned to either the control or treatment groups, so that any difference in outcomes among firms in the treatment versus control group can be attributed to the treatment itself. Randomization is an important topic which we will revisit as the course goes on.

Question 2

0.0/1.0 point (graded)

In general, what information does a histogram depict?

- ◯ A histogram depicts a treatment effect
- ◯ A histogram depicts changes in an outcome measurement over time
- ◯ A histogram depicts the number (or fraction) of observations that falls within certain ranges of a particular outcome measurement ✔
- ◯ A histogram compares original measurements to back check measurements

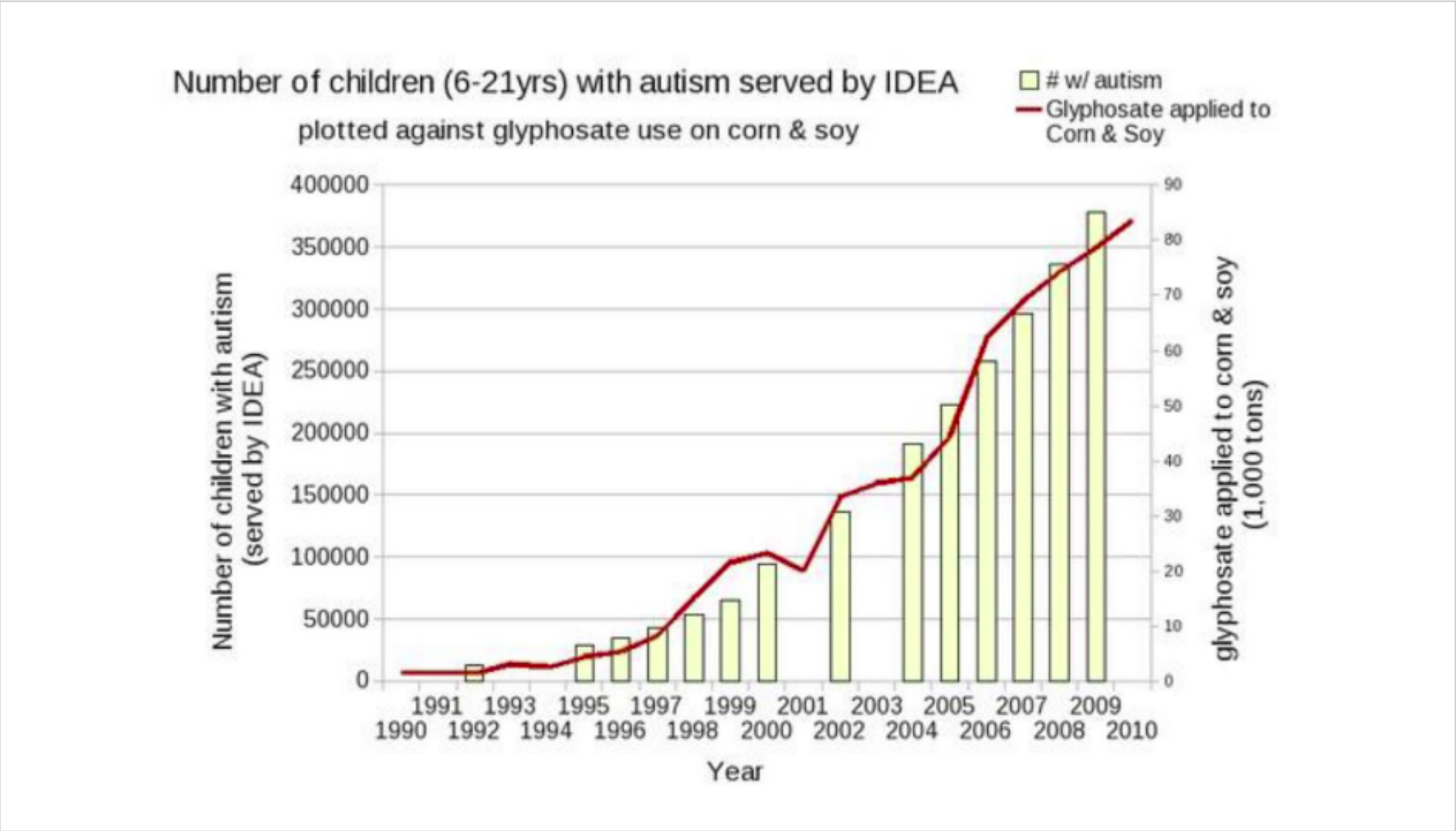
Explanation

Generally speaking, histograms depict the number of observations that falls within certain ranges, or "bins," of a particular outcome measurement. They are a helpful way for visually describing data. We will discuss histograms further in the coming lectures and provide tools for creating your own histograms using R.

Submit You have used 0 of 2 attempts

Show Answer

True or false: The chart below **proves** that increased glyphosate use over time has led to an increase in the number of cases of autism among children aged 6-21 years.



○ True

○ False ✔

Explanation

False. The chart does seem to show that use of glyphosate and the number of cases of autism have both increased over time. However, based on this evidence we cannot confidently conclude that increased glyphosate use has caused an increase in autism. There could be many other factors not captured in this chart that have led to the increase in glyphosate use over time and to the increase in cases of autism over time.

Submit

You have used 0 of 1 attempt

Show Answer

Question 2

0.0/1.0 point (graded)

What point or note of caution does Professor Duflo make using the examples given in this lecture segment? (Check all that apply)

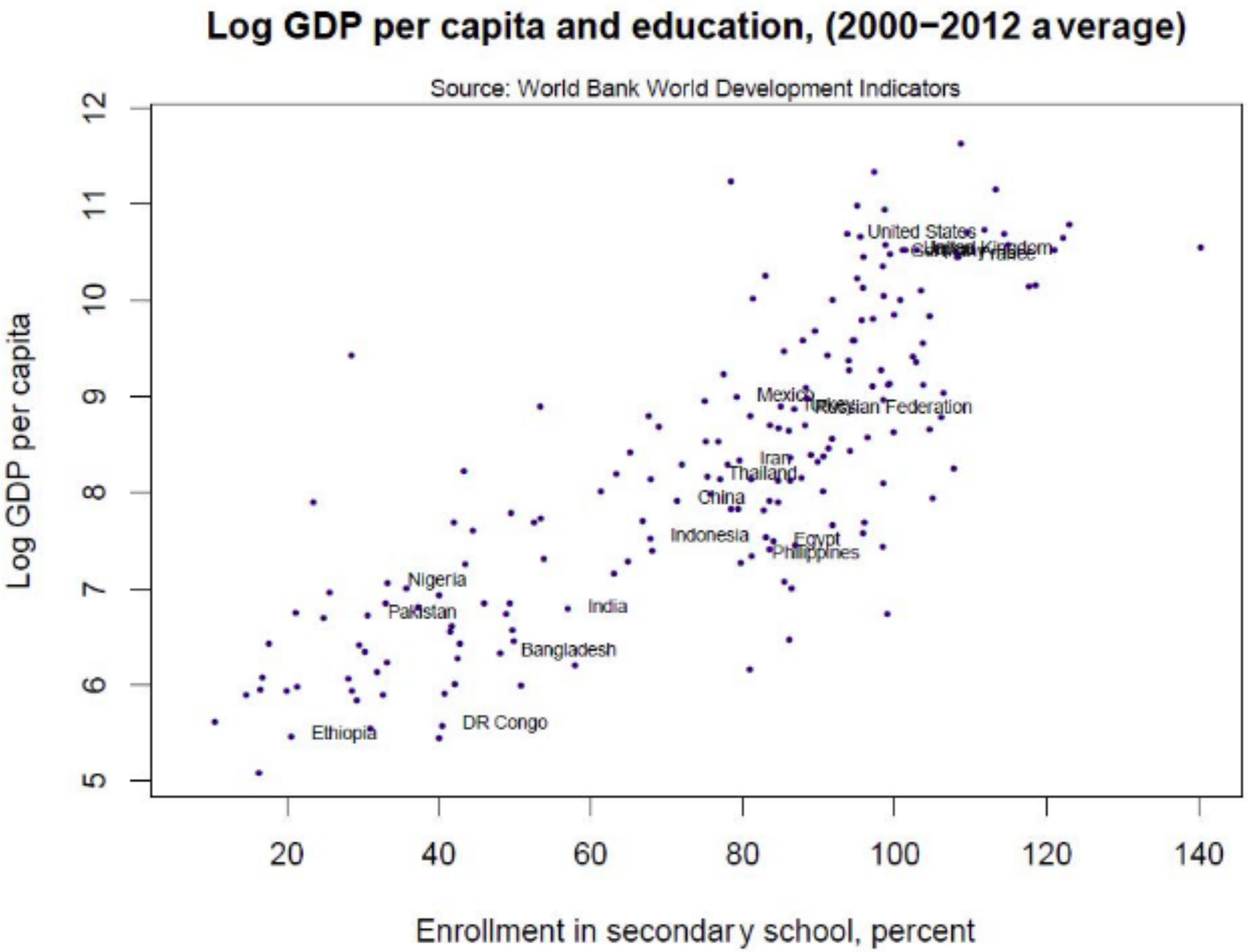
- [ ] That data can sometimes be used in a misleading way, in order to support whichever viewpoint the researcher wants to support ✔
- [ ] That data should always be taken at face value
- [ ] That there is a fine balance between letting the data speak and using theory to inform the choice of what to look for in the data ✔
- [ ] That data should be collected and interpreted carefully and thoughtfully ✔

Explanation

Professor Duflo presents various examples that demonstrate that caution should be used in collecting, presenting, and interpreting data, since it can sometimes be deceitful. One chart shows some kind of relationship between glyphosate and the number of cases of autism, while another shows some kind of relationship between organic food sales and cases of autism. Professor Duflo makes the point that researchers could focus on or drop certain portions of data in order to show support for whatever explanation or story they have in mind.

Submit You have used 0 of 2 attempts

The graph below shows that GDP per capita and education levels are highly correlated. One interpretation of this could be that higher education causes higher GDP per capita. Which of the following are discussed as reasons that we may be hesitant to conclude that education causes higher incomes? (Check all that apply)



- [ ] Hidden or omitted variables: There is a third factor that increases education and income which is not accounted for here ✔
- [ ] Measurement error: One of the variables is inaccurately measured, or measured in such a way that is not comparable across countries
- [ ] Reverse causality: Higher incomes cause higher education, rather than the other way around ✔
- [ ] Heterogeneity: Countries differ on many dimensions and we cannot compare outcomes like GDP per capita or education levels across countries.

Explanation

Reverse causality and hidden/omitted variables are discussed in class as reasons that we should use caution before concluding that higher education leads to higher GDP. There could be other third factors that contribute to higher incomes as well as higher education levels which are not included in this simplistic model. There could also be some reverse correlation at play, where it is not necessarily the case that higher education leads to higher income, but rather than higher levels of income lead to higher education levels. In this example of outcomes as complex as national GDP per capita and education levels, there are likely many interrelated factors and interactions at play beyond what is included in this simplistic model.

Submit You have used 0 of 2 attempts

Show Answer

Answers are displayed within the problem

Correlation versus Causation - Quiz

Bookmark this page

Finger Exercises due Sep 14, 2020 19:30 EDT **Past Due**

Question 1

0.0/1.0 point (graded)

True or false: Controlling for variables other than treatment and outcome could better isolate the relationship that we are interested in.

- ◯ True ✔
- ◯ False

Explanation

True. If we are able to control for a wide range relevant variables, this should allow us to better isolate the relationship of interest. However the difficulty is to be sure we have controlled for all the relevant variables.

Submit You have used 0 of 1 attempt

Show Answer

Correlation versus Causation - Quiz

Bookmark this page

Finger Exercises due Sep 14, 2020 19:30 EDT **Past Due**

Question 1

0.0/1.0 point (graded)

True or false: Controlling for variables other than treatment and outcome could better isolate the relationship that we are interested in.

- ◯ True ✔
- ◯ False

Explanation

True. If we are able to control for a wide range relevant variables, this should allow us to better isolate the relationship of interest. However the difficulty is to be sure we have controlled for all the relevant variables.

Submit You have used 0 of 1 attempt

Show Answer

Question 2

0.0/1.0 point (graded)

True or false: by hunting for correlations in the data, one can often find entirely spurious patterns.

- ◯ True ✔
- ◯ False

Explanation
True. When running many regressions in a data set, there is the risk of "overfitting": finding patterns by random chance in a particular data set that would not be found elsewhere because they are not real. We will see later in the course how this can be avoided, with a model (the position traditional econometrics take) or with statistical techniques (the position of Machine Learning)

Submit You have used 0 of 1 attempt

Show Answer

Answers are displayed within the problem