

What type of graph should I use?

Learning how to best visualize your data is almost as important as your statistical analysis. Science communication is increasingly important – you have to learn how to communicate your results to the world. One of the first steps in good science communication is understanding which types of graphs to use based on the type of data you have. This practice exercise includes (1) resources to refresh your memory on types of data and visualization; (2) practice with recognizing the basics; and (3) practice applying your data visualization skills in a real-world context.

Resources:

- <https://www.ck12.org/statistics/graphs-for-discrete-and-for-continuous-data/lesson/Basic-Graph-Types-BSC-PST/>
- <http://www.statisticshowto.com/types-graphs/>
- <http://blog.visme.co/types-of-graphs/> (note that there are far more examples here than what you'd need to know for this class, but it's useful to see the variety that's out there)
- <https://www.socialresearchmethods.net/kb/measlevl.php> (Review: types of variables; ignore ratio)

Practice with the Basics:

1. Complete the following chart, assuming you have an interval/continuous DV:

	Type of Graph	Description and Example	Example Stata Code
Nominal IV, Interval DV		Plot the mean of the DV for each value of the IV. If your IV was political party (Republican, Democrat) and your DV was contributions to the Sierra Club (\$0, \$100, ...), you could plot the average contribution amount among Democrats in one bar and the average contribution amount among Republicans in the other bar.	
Ordinal IV, Interval DV	Bar Chart		<code>graph bar (mean) DVname, over(IVname)</code>
Interval IV, Interval DV			<code>scatter DVname IVname</code>

2. In the “Colbert Bump” reading, Professor Fowler tested whether coming on the *Colbert Report* caused an increase in campaign contributions.
 - a. What is the Independent Variable?
 - b. What is the Dependent Variable?
 - c. What type of variable is the IV?
 - i. Nominal
 - ii. Ordinal
 - iii. Interval/Continuous

- d. What type of variable is the DV?
 - i. Nominal
 - ii. Ordinal
 - iii. Interval/Continuous
- e. What kind of graph would you use to visualize the IV?
 - i. Scatterplot
 - ii. Bar graph
 - iii. Histogram
 - iv. Other
- f. What kind of graph would you use to visualize the DV?
 - i. Scatterplot
 - ii. Bar graph
 - iii. Histogram
 - iv. Other
- g. What kind of graph would you use to visualize the relationship between the IV and the DV?
 - i. Scatterplot
 - ii. Bar graph
 - iii. Histogram
 - iv. Other

Practice Application:

The following is an excerpt from the New York Times

(<https://www.nytimes.com/2017/11/07/world/americas/mass-shootings-us-international.html>)

“But there is one quirk that consistently puzzles America’s fans and critics alike. Why, they ask, does it experience so many mass shootings?”

Perhaps, some speculate, it is because American society is unusually violent. Or its racial divisions have frayed the bonds of society. Or its citizens lack proper mental care under a health care system that draws frequent derision abroad.

These explanations share one thing in common: Though seemingly sensible, all have been debunked by research on shootings elsewhere in the world. Instead, an ever-growing body of research consistently reaches the same conclusion.

The only variable that can explain the high rate of mass shootings in America is its astronomical number of guns.”

One way the article proposes to understand this correlation is to gather data on gun ownership and the number of mass shooters. They collect this information for a number of countries using a survey run by the University of Alabama in 2015. The survey has data on the number of guns owned per country (for example, the U.S. owns 270 million) and the number of mass shooters per country from 1966-2012 (for example, the U.S. has had 90 mass shooters during this period).

1. What is the IV?
2. What is the DV?

3. What type of variable (nominal/ordinal/interval) is the IV?
4. What type of variable (nominal/ordinal/interval) is the DV?
5. In order to visualize the relationship between the IV and DV what type of graph might you use? Sketch a rough idea of how the graph would look.

While this correlation is interesting, the above article doesn't tell us anything about *who* chooses to own a gun. Luckily, the Pew Research Center collected this information in 2017 (<http://www.pewsocialtrends.org/2017/06/22/the-demographics-of-gun-ownership/>). One aspect of gun ownership they are interested in is partisanship. They expect that self-identified Democrats will be less likely to own guns, given that the party has historically been a proponent of gun control. They collect information by asking people to identify their partisanship (respondents can choose Democrat, Republicans, and Independent) and whether or not they or someone in their household owns a gun (respondents can respond personally own a gun, don't own a gun but someone in their household owns a gun, or no one in their household owns a gun). They want to show how the percentage of people who own a gun varies by party affiliation.

1. What is the IV?
2. What is the DV?
3. What type of variable (nominal/ordinal/interval) is the IV?
4. What type of variable (nominal/ordinal/interval) is the DV?
5. In order to visualize the relationship between the IV and DV what type of graph might you use? Sketch a rough idea of how the graph would look.

Answer Key

Practice with the Basics:

1. Complete the following chart, assuming you have an interval/continuous DV:

	Type of Graph	Description and Example	Example Stata Code
Nominal IV, Interval DV	Bar Chart	<p>Plot the mean of the DV for each value of the IV.</p> <p>If your IV was political party (Republican, Democrat) and your DV was contributions to the Sierra Club (\$0, \$100, ...), you could plot the average contribution amount among Democrats in one bar and the average contribution amount among Republicans in the other bar.</p>	<code>graph bar (mean) DVname, over(IVname)</code>
Ordinal IV, Interval DV	Bar Chart	<p>Plot the mean of the DV for each value of the IV</p> <p>If your IV was ideology (Very liberal, liberal, moderate, conservative, very conservative) and your DV was contributions to the Sierra Club (\$0, \$100, ...), you could plot the mean contribution amount made by those who are “very liberal,” the mean contribution amount made by those who are “liberal,” the mean contribution amount made by those who are “moderate,” etc.</p>	<code>graph bar (mean) DVname, over(IVname)</code>
Interval IV, Interval DV	Scatterplot	<p>Plot the IV on the X-Axis and the DV on the Y-Axis</p> <p>If your IV was age (18 years, 20 years, 50 years, 65 years, etc.) and your DV was contributions to the Sierra Club (\$0, \$100, \$1000, ...), you would plot the X-Y coordinates for each observation. For instance, if you had an 18 year old who contributed \$50, you’d have a point at X=18, Y=50.</p>	<code>scatter DVname IVname</code>

2. In the “Colbert Bump” reading, Professor Fowler tested whether coming on the *Colbert Report* caused an increase in campaign contributions.
 - a. What is the Independent Variable?
 - i. **Going on the *Colbert Report* (yes, no)**
 - b. What is the Dependent Variable?

- i. Campaign contributions (\$0, \$10, \$1000, ...)
- c. What type of variable is the IV?
 - i. Nominal
 - ii. Ordinal
 - iii. Interval/Continuous
- d. What type of variable is the DV?
 - i. Nominal
 - ii. Ordinal
 - iii. Interval/Continuous
- e. What kind of graph would you use to visualize the relationship between the IV and the DV?
 - i. Scatterplot
 - ii. Bar graph
 - iii. Histogram
 - iv. Other

Practice Application:

The following is an excerpt from the New York Times

(<https://www.nytimes.com/2017/11/07/world/americas/mass-shootings-us-international.html>)

“But there is one quirk that consistently puzzles America’s fans and critics alike. Why, they ask, does it experience so many mass shootings?

Perhaps, some speculate, it is because American society is unusually violent. Or its racial divisions have frayed the bonds of society. Or its citizens lack proper mental care under a health care system that draws frequent derision abroad.

These explanations share one thing in common: Though seemingly sensible, all have been debunked by research on shootings elsewhere in the world. Instead, an ever-growing body of research consistently reaches the same conclusion.

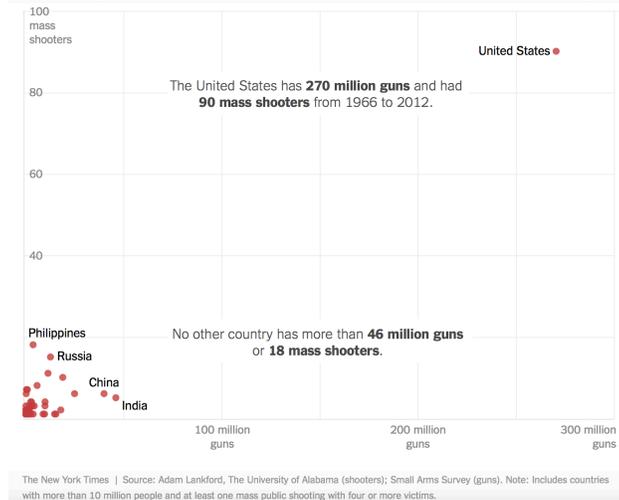
The only variable that can explain the high rate of mass shootings in America is its astronomical number of guns.”

One way the article proposes to understand this correlation is to gather data on gun ownership and the number of mass shooters. They collect this information for a number of countries using a survey run by the University of Alabama in 2015. The survey has data on the number of guns owned per country (for example, the U.S. owns 270 million) and the number of mass shooters per country from 1966-2012 (for example, the U.S. has had 90 mass shooters during this period).

1. What is the IV? **The number of guns owned (per country)**
2. What is the DV? **The number of mass shootings (per country)**
3. What type of variable (nominal/ordinal/interval) is the IV? **Interval**
4. What type of variable (nominal/ordinal/interval) is the DV? **Interval**
5. In order to visualize the relationship between the IV and DV what type of graph might you use? Sketch a rough idea of how the graph would look.

Because both variables are continuous, I can use a scatterplot. I can plot the number of guns owned (my IV) on the X-axis and the number of mass shootings (my DV) on the Y-axis. Scatterplots are helpful for visualizing the relationship between continuous variables because

each “dot” or unit of observation is one country. By looking at each country “dot”, I can see both the country’s number of guns and their number of mass shootings.



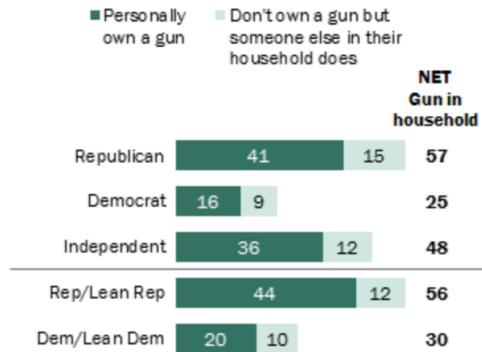
While this correlation is interesting, the above article doesn’t tell us anything about *who* chooses to own a gun. Luckily, the Pew Research Center collected this information in 2017 (<http://www.pewsocialtrends.org/2017/06/22/the-demographics-of-gun-ownership/>). One aspect of gun ownership they are interested in is partisanship. They expect that self-identified Democrats will be less likely to own guns, given that the party has historically been a proponent of gun control. They collect information by asking people to identify their partisanship (respondents can choose Democrat, Republicans, and Independent) and whether or not they or someone in their household owns a gun (respondents can respond personally own a gun, don’t own a gun but someone in their household owns a gun, or no one in their household owns a gun). They want to show how the percentage of people who own a gun varies by party affiliation.

1. What is the IV? **Partisanship**
2. What is the DV? **Whether someone in the household owns a gun**
3. What type of variable (nominal/ordinal/interval) is the IV? **Nominal**
4. What type of variable (nominal/ordinal/interval) is the DV? **Nominal**
5. In order to visualize the relationship between the IV and DV what type of graph might you use? Sketch a rough idea of how the graph would look.

Because the independent variable is nominal, I can only display the data by groups. That means I need to calculate a common statistic for each group separately. One way to look at how the groups differ in their gun ownership is to calculate the percentage of gun owners for each category separately. I can calculate the percentage of Republicans who own a gun and compare it the percentage of Democrats who own a gun. The best way to display these percentages side-by-side is a bar graph. In the bar graph, each bar is a category of the IV and the Y-axis is a summary statistic of the DV (in this case percentage, but in other examples it could also be a mean, median, etc.)

Gun ownership is far more common among Republicans than Democrats

% saying they ...



Note: Figures may not add to subtotals indicated due to rounding.
Source: Survey of U.S. adults conducted March 13-27 and April 4-18, 2017.

"America's Complex Relationship With Guns"

PEW RESEARCH CENTER
