

1. Open the file vgsales.xlsx you worked on in the last Exercise in Excel and create a new worksheet called “Answers, E3.” Write your answers to the following questions in that sheet.
2. Think of some questions you could use this data set to measure.
  - Pick one of your questions. What are two different ways you could measure it?
3. Is there anything within the data set that you’re curious about? For instance, why are some numbers the way they are? How might sales figures differ because of geographical region? Write down any thoughts you’d want to investigate further.
4. What are the data set’s dimensions?
5. What are the columns, and what do they mean?
6. Which columns have missing data? How much data is missing?
7. What are some questions you’d be unable to answer with this data set?
8. Categorize each column as quantitative/qualitative, and discrete/continuous or binary/nominal/ordinal. Keep in mind that the sales columns—EU sales, JP sales and NA sales are measuring *sold units*, but they have been rescaled, so that they show the number in millions. You check this out again in your project brief. Hint: have a really good think about what this column is counting!
9. What does an observation represent within this data set?
10. What are some of the characteristics within this data set? What characteristics aren’t here that you expected or would have liked to see?
11. What kind of bias could be implicit in this data? You can refer to the “Data Set” section of your project brief for help with this step. Here, you’ll find links to the data source and data collection method.
12. What populations could this sample be used to make generalizations about?