Exploring Big Data Sets

Giving students large data sets can be a challenging and motivating experience. One way to engage students in the data science process is to use the following structure (Burrill & Dick, 2022):

- Define a problem,
- Consider/collect the data, including how, why, by whom, and for what purpose the data were collected), paying attention to the source
- Manage the data, including dealing with outliers, missing data, entering the data in a spreadsheet, which involves making decisions about which data goes into the rows and which into the columns, labels that will work with graphs in the software etc.
- Process the data, including filtering to look at subgroups, and considering how to deal with different units and magnitudes using some form of transformation, all of which may involve some coding)
- Analyzing and visualizing, including different representations, ways to represent multiple variables in a plot, and unique plots
- Creating a model, typically a model that involves variability and several variables
- Communicating and making a decision or recommendation, using non-technical language with evidence supporting the decision.

Teaching Tips

1) Getting started: Students typically do not do well with broad open questions such as "describe the gender wage gap"; they need specific tasks that can help them focus on salient features. Because some try to do too much or investigate a question not appropriate for the data, having students submit a statement describing what they are going to investigate can prevent them from trying to do something that will not work leading to frustration. Be sure the data and question match.
2) Check-points: Students without much experience working with data or using mathematical action technology to do so often create graphs that do not make sense; typically, someone thinks bar graphs are the only way to go but often don't know how to create them on their device. Rather than instruct the whole class before they start (then they all do bar graphs), have check-points throughout the activity where students have to share what they are doing to be sure they are on a good track. Students often confuse categorical and quantitative data and try things such as finding a mean for categorical data, making a box plot for a single variable ignoring time, or making a scatter plot with categories on the horizontal axis.
3) Students have difficulty communicating using both words and numbers, typically using only numerical arguments because that is what much of the mathematics they have learned is about, showing the calculations for means and standard deviations without tying them to the context, or writing the equation of a regression line without words to indicate what story these tell about the context.

4) If students add information or visual representations they find from another source to their work, they should give the source.
5) Peer review: Having students review each other's work and then revising their work before they turn it in is very useful in building community and giving students agency in the work. However, to avoid feedback that is not useful ("it was a really nice report"; " I liked what you did"), give students specific questions with the goal to improve the response. For example, students critique using sentence stems such as: The work a) is understandable and makes sense because ...; b) might have more details or clarification about ...; c) uses appropriate mathematical/statistical language, but I was unsure when it said …; d) left me a bit confused about ...
6) At the end of the activity, after their work is turned in, ask students to reflect on what they learned doing the activity - this is where the teacher can prompt those who learned something worthwhile to share (i.e., how to make a bar graph on TI-Nspire™).

Burrill, G., & Dick, T. (2022). Connecting mathematics to the world: Engaging students with data science. In J. Morska, J. & A. Rogerson (Eds.). Building on the Past to Prepare for the Future, *Proceedings of the 16th International Conference of The Mathematics Education for the Future Project*, King's College, Cambridge.