ISSUE 29 MARCH 2020

CAMBRIDGE √Mathematics

ESPRESSO

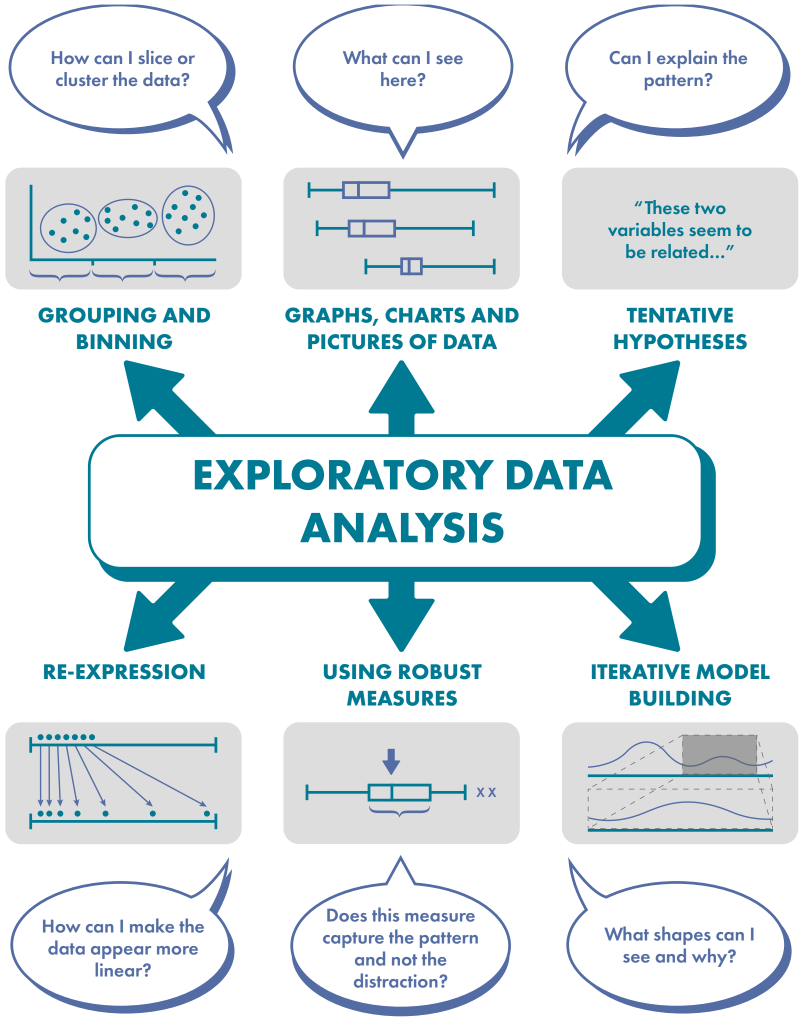
RESEARCH, FILTERED BY CAMBRIDGE MATHEMATICS

TALKING POINT:

WHAT DOES RESEARCH SUGGEST ABOUT TEACHING STATISTICS USING EXPLORATORY DATA ANALYSIS (EDA)?

IN SUMMARY

- EDA is an informal and exploratory approach to statistics characterised by drawing different representations, searching for patterns and considering "what is going on here?"
- Statistics education research suggests using EDA at any age, and waiting until later to use CDA
- EDA allows students to be "data detectives"; using creative and varied approaches to statistics which may provide a richer, more coherent experience
- Creating and moving between different representations, as well as using proportional reasoning, are key skills that can connect students' mathematical experiences across topics
- Students can use any useful representations (and invent their own) but should consider how robust these measures are
- Students should have access to data sets containing multiple data types so that they get used to independently sifting and selecting what to focus on
- Using research-supported technology such as TinkerPlots, Fathom and CODAP supports EDA by allowing students to test ideas and focus on data interpretation



1

Exploratory data analysis (EDA) is a way of approaching statistics where the aim is to look at data from as many angles as possible in order to observe interesting features.[1] It is much more informal, non-prescriptive and investigative than its counterpart, confirmatory data analysis (CDA). It is characterised by: exploring a data set with the very broad question "what is going on here?"; graphic representations (graphs, charts, sketches, pictures) of data; and flexibility regarding which methods to apply.[2] It also aims to uncover underlying structures, isolate important variables, detect outliers and other anomalies, and suggest suitable models for conventional statistics.[3] It can be done with students over a range of ages and attainment levels, as it begins with simple but powerful statistical ideas such as counting, grouping, finding the centre, and looking at the shape (distribution) of the data.[4] EDA is recommended as more appropriate to students' initial experience of statistics education than CDA, which comes later.[2,5]

IMPLICATIONS: EDA is about exploring data sets to look for patterns, structure and shape, and is recommended as a more appropriate place to begin statistics education than CDA, which comes later

Statistics education research supports the use of EDA in the classroom as a useful way to begin exploring statistical ideas at any age

CAMBRIDGE UNIVERSITY PRESS
UNIVERSITY OF CAMBRIDGE Faculty of Mathematics
UNIVERSITY OF CAMBRIDGE Faculty of Education
Cambridge Assessment

2

Teachers might be familiar with the how of EDA – such as drawing stem-and-leaf or box plots and using the median and quartiles – but not the why, which is that students are empowered and motivated to explore, become "data detectives" and find stories in the data for themselves, ideas which are much more akin to the way statistics is used outside schools.[6] Using images and diagrams to find and show different features of data is a key skill, and students being able to create and invent their own representations, critique them, and flexibly move between them is particularly important as it extends to all maths learning.[7] Using EDA changes the teaching paradigm for statistics education because it allows for better continuity and coherence across students' experience, providing both a more meaningful context for students and a richer, more authentic experience of statistics.[5]

IMPLICATIONS: EDA is creative and empowering for students, allowing them to be "data detectives," which is similar to the way statistics is used in problem-solving in the real world

Creating and moving between different representations is a key skill that can connect students' mathematical experiences across topics

Using EDA in the classroom allows for a richer, more coherent experience for students

3

The EDA approach suggests that drawing lots of pictures, graphs and charts is a good place to start,[5] not only as a way to communicate, but also a way to explore data.[8] There is no set "toolkit" of techniques for EDA, but some of the suggested ways to start are: stem-and-leaf, scatter, box plot, shapes of distributions, quartiles, median, binning and slicing, and re-expression (see infographic).[7] It is important to use *robust measures*: statistical measures that are *resistant* (not too sensitive to outliers or blips), *smooth* (not too affected by bad data) and *broad* (applicable to a wide range of situations).[8] EDA is both strongly underpinned by and supports proportional reasoning, as students are encouraged to group, bin, re-express and slice data using ideas about relationships between variables.[7] It is important to start with a rich data set with varied data types, so students get used to independently sifting and selecting what to focus on.[6]

IMPLICATIONS: Students can use any useful representations (and invent their own), but should consider how robust these measures are and be used to working with varied data sets

Teachers can help students to connect EDA meaningfully to ideas elsewhere in mathematics through the use of proportional reasoning

4

The use of technology – often dynamic statistical software – is woven into EDA because it is efficient when moving flexibly between representations, allowing for a shift in focus towards interpretation and empowering students to make meaningful decisions for themselves.[9,10] Software programs such as TinkerPlots, Fathom and CODAP have been recommended by researchers as part of EDA as they help students develop statistical thinking, allowing them to test ideas frequently and quickly change boundaries when slicing or binning to see immediate effects on data visualisation.[8]

IMPLICATIONS: Using technology supports EDA by allowing students to test ideas and focus on data interpretation, giving them more autonomy and freedom

Research suggests TinkerPlots, Fathom and CODAP as examples of software designed to support EDA

"there is no single "plot of the data" but rather only one of many possible plots"[2]

Behrens, 1997

"the greatest value of a picture is when it forces us to notice what we never expected to see"[7]

Tukey, 1977

"You have designed your experiment, collected the data, and are now confronted with a tangled mass of information that must be analyzed... Turning this heap of raw spaghetti into an elegant *fettucine alfredo* will be immensely easier if you can visualize the message buried in your data"[8]

Ellison, 1993

Lucy Rycroft-Smith & Darren Macey, 2020

REFERENCES

1. Morgenthaler, S. (2009). Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 1*(1), 33–44.
2. Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*(2), 131–160.
3. Hinterberger, H. (2009). Exploratory data analysis. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems*. Springer US.
4. Morrison, P. S. (1983). Teaching exploratory data analysis. *New Zealand Journal of Geography, 75*(1), 2–6.
5. Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. CD Beta Press.
6. Pfannkuch, M. (2018). Reimagining curriculum approaches. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 387–413). Springer International Publishing.
7. Tukey, J. W. (1977). *Exploratory data analysis* (Vol.2). Addison-Wesley Publishing Company.
8. Ellison, A. M. (1993). Exploratory data analysis and graphic display. In S. M. Scheiner & J. Gurevitch (Eds.), *Design and analysis of ecological experiments* (pp. 14–43). Chapman & Hall.
9. Mojica, G. F., Barker, H., & Azmy, C. N. (2019, February 21–24). *Instrumented learning in a CODAP-enabled learning environment* [Paper presentation]. Third International Virtual Congress on Statistical Education.
10. Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1–2), 127–155.

Cambridge University Press
University of Cambridge Faculty of Mathematics
University of Cambridge Faculty of Education
Cambridge Assessment