

# DATA PATTERNS WITH DR ERIC CHI

## TALKING POINTS

### KNOWLEDGE:

1. What is a data cube?

### COMPREHENSION:

2. Why are data cubes useful?
3. Why is clustering a useful technique?

### ANALYSIS:

4. Why has data science become more widespread (and increasingly needed) in recent decades?

### SYNTHESIS:

5. Eric talks about his work analysing voting patterns in the US Senate. How do you think his techniques could be useful in, say, predicting the outcome of the next presidential election?
6. Simplifying datasets means that information is inevitably lost. How do you think Eric's team might minimise the loss of potentially useful information?
7. Eric's work has helped a broad range of disciplines. Can you think of scenarios where Eric's techniques might help a) epidemiologists, b) meteorologists, and c) social scientists?

## ACTIVITIES YOU CAN DO AT HOME OR IN THE CLASSROOM

Imagine you are a data scientist. You are approached by people with the following problems. For each issue:

- Firstly, interpret the problem they have. They have plenty of data – why is it difficult to draw conclusions from it?
- Consider how the methods explored in the article could be used.
- Think about how the collaborators could draw conclusions that are useful in the real world. How could they benefit society?
- A plant epidemiologist is tracking the spread of a disease that affects wheat. They have data from across the world about where the disease has been found, how prevalent it is at each site, when it was discovered at each site, the varieties of wheat infected, how it's affected each crop, and weather records for each region.
- A social media platform is being paid to target its users with advertisements for various clothing brands. They have plenty of information on the brands, such as the type of person they are marketed towards, and plenty of information on their users, such as age, country of residence, and other interests.
- A political thinktank wants to predict the results of an upcoming election. They have the information on how different constituencies voted for past candidates, all of whom had varying degrees of differing stances on social, economic and environmental issues.
- A researcher is attempting to create a treatment for a new virus. They have a vast database on the properties of similar viruses, the properties of many pharmaceuticals, and how these viruses and pharmaceuticals interact.

## MORE RESOURCES

Eric recommends the following websites for resources and tutorials on data science computing skills that you can learn:

<http://shinstitute.org/webinar-series/>

<https://rstudio.com/resources/training/>

- In this TEDx talk, Rebecca Nugent talks about how data science is 'the people's science' and its role in education:

<https://www.youtube.com/watch?v=YMnqPTLoj7o&t=13s>

- This article talks about how big data is being used for good, such as in attempts to eradicate malaria in Zambia: <https://www.hpe.com/us/en/insights/articles/data-for-good-wonderful-ways-data-is-making-the-world-better-1810.html>