

5.1: Why Summarise Data?

When we summarise data, we are essentially throwing away information, and one may object to this. As an example, let's go back to the PURE study that we discussed in Chapter 1. You may think, what happens to participant information beyond what was summarised in the dataset? What about the specific details of how the data were collected such as the time of day or whether it was on a weekend or a weekday? What about the mood of the participant? All of these details are lost when we summarise the data.

We summarise data because it allows us to *describe* and *compare*. Two actions that we do all the time in everyday life.

Curiously, students often complain that statistics is confusing and irrelevant. But, the very same students are the ones comparing their GPAs with their classmates. They may even be calculating the grade they need for the next assessment in order to pass. Most people don't realise this, but we are cognizant of **descriptive statistics** because we use numbers to summarise information on a daily basis.

For example, GPA is often used by universities to assess high school students' academic potential. In Australia, high school students are given a ranking called the Australian Tertiary Admission Rank (ATAR) as the primary criterion to enter undergraduate courses. A student with an ATAR of 95 is clearly a stronger student than someone in the same class with an ATAR of 80.^[1] Using Charles Wheelan's words from *Naked Statistics*,^[2] GPA (or in our case, ATAR) makes a nice descriptive statistic: it's easy to calculate, it's easy to understand, and it's easy to compare across students. However, it is important to note that it is not *perfect* as it often does not reflect the difficulty of the subjects that different students may have taken.

We also summarise data because it enables us to *generalise*. That is, to make general statements that extend beyond specific observations. Psychologists have long recognised the importance of generalisation, including the process of categorisation. For example, we can easily recognise different types of birds, despite their differing surface features. We know that an ostrich, a robin and a chicken belong to the "bird category", but we understand that these birds individually differ from one another. Importantly, generalisation allows us to make predictions. In the case of birds, we can predict that they can fly and eat worms and that they probably can't drive a car or speak English. These predictions won't always be right, but they are often useful in the real world.

Chapter attribution

This chapter contains taken and adapted material from *Statistical thinking for the 21st Century* by Russell A. Poldrack, used under a CC BY-NC 4.0 licence.

1. It's a percentile ranking between 0.00 and 99.95. For more information:
https://en.wikipedia.org/wiki/Australian_Tertiary_Admission_Rank ↩
2. Wheelan, C. (2013). *Naked statistics: Stripping the dread from the data*. WW Norton & Company. ↩

This page titled 5.1: Why Summarise Data? is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by Klaire Somoray (Council of Australian University Librarians Initiative).

- 4.1: Why Summarize Data? by Russell A. Poldrack is licensed CC BY-NC 4.0. Original source:
<https://statstinking21.github.io/statstinking21-core-site>.