ONTARIO NONPROFIT NETWORK

# Data 101: The Language of Data

# Examples

## Introduction

This document showcases examples of how quantitative data can be described or used. You can find more information on what data is and how nonprofits use it in our resource, The Language of Data.

## Sample data set

We're going to work with a fictional sample data set from Abbey Crossing Homework Help Program, which provides after-school homework help at the local school:

| Name | Address | Age | Gender Identity | Regularly attended Homework Club | Final GPA |
|---|---|---|---|---|---|
| John Lemon | 123 Mulberry Street | 11 | M | Yes | 4 |
| Deepika Sharma | 456 Twopence Road | 12 | W | No | 2 |
| Thandiwe Abara | 789 London Circle | 13 | M | Yes | 3 |
| Jimin Kang | 987 Beatle Blvd | 13 | NB/NC | No | 3 |
| Ana Yoko | 246 Record Road | 12 | W | Yes | 4 |

## Pro tip:

If you are collecting information on gender identity, questions should be carefully phrased so respondents see their identity(ies) reflected and included. For this example, we used Man/Woman/Non-Binary/Non-Conforming, but other choices may include Transgender, Cisgender, Gender-binary, Genderqueer, Select all that Apply, Prefer Not to Answer. These can be presented as answer choices or as an open-ended question. Resources such as this one can help your organization learn more guidelines on asking questions about gender identity and sexuality.

# Anonymized data

Nonprofits use data regularly, including looking at census data to better understand the communities they serve or adjusting their programs and services to match seasonal trends.

| Participant | Age | Gender Identity | Regularly Attended Homework Club | Final GPA |
|---|---|---|---|---|
| 1 | 11 | M | Yes | 4 |
| 2 | 12 | M | No | 2 |
| 3 | 13 | M | Yes | 3 |
| 4 | 13 | NB/NC | No | 3 |
| 5 | 12 | W | Yes | 4 |

We took out the names and addresses, but anything that may be used to identify participants (ie. physical descriptions, phone numbers, etc.) should be taken out to ensure privacy.

# Mean/Average and Median

Mean/Average Grade = Sum of all grades/number of grades
For example: (4+2+3+3+4)/5= 16/5 = **3.5**

Median Grade = Middle value in ordered list (from smallest to largest)
For example: 2,3,3,4,4, Median = **3**

The small difference between the mean/average and the median tells us that our numbers are evenly distributed. It also means that there are no outliers that are skewing our data.

# Aggregation and Disaggregation

As we look at the information behind the numbers, we can see who attended our homework club regularly by particular variables, such as age. To do this, we collapse our table rows by age:

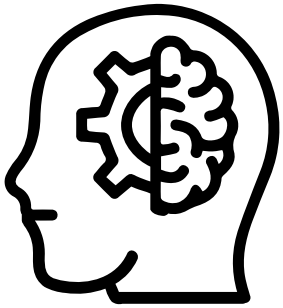
| Age | Number of Participants in Age Range |
|---|---|
| 11 | 1 |
| 12 | 2 |
| 13 | 2 |

Now, we build out our table adding in attendance:

| Age | Number of participants in age range | Regularly Attended Homework Club | Did not Regularly Attend Homework Club |
|---|---|---|---|
| 11 | 1 | 1 | 0 |
| 12 | 2 | 1 | 1 |
| 13 | 2 | 1 | 1 |

We have aggregated the data (by age), and then disaggregated it to showcase the information we are interested in. The table above might help us understand who is using our programs and may help us understand their needs. It can also help us think about how different people access services and whether different outreach (or programming) can help increase access.

Because our data sample is so small, we want to be cautious of whether aggregation/disaggregation by a particular variable may reveal identifiers and compromise anonymity and safety. In our example, disaggregating by gender, where we only have one participant that identifies as non-binary/non-conforming in a sample of five participants may mean that we do not want to use gender as a way to share this data. Equity-seeking groups often find themselves absent or under-represented, and it is important to think through how nonprofit organizations can share their stories while maintaining their confidentiality and dignity.



# How can this data be used?

While this sample set has limited use, we may want to use it in a couple of ways:

- Maybe we want to try targeted outreach to the age/gender identity to increase our numbers, and track the information over time, producing longitudinal attendance data.
- We might want an in-depth conversation with a participant or distribute a survey to understand any barriers to participation, producing qualitative data.
- We might pull out last year's data and compare the attendance to see if there are trends or commonalities.
- We might share the anonymized data with other service providers for comparison or to expand the program elements.

**Even a small amount of data can support a nonprofit organization in making program decisions, reviewing their mandates, and furthering their strategic plans.**

# How not to use this data

Data is powerful, and we must be careful of how we use it. There are a couple of things that we absolutely cannot do:

- We cannot draw conclusions or definitive statements about whether homework club attendance is an indicator of good grades.
- We cannot collect this data without being explicit with our participants about why we are collecting it or how we are using it.
- We cannot share this data with identifiers (unless we have actually requested permission from participants beforehand).
- We cannot present the data as is, without adding context to it. We have to acknowledge the limitations of our data upfront before we use it.
- We cannot generalize the experiences of the participants or expand their experience as being representative of a particular demographic.

# Skill-building activity

If you're still building your data skills and would like to try mimicking what we did above, look at the sample set below and answer our questions. Our next sample data comes from the Friendly Food Drive. They tracked individual food donations. Think about how Friendly Food Drive can use this information to direct their fundraising efforts towards individual donors.

| Name | Age | Gender Identity | Type of Food Donated | Type of Food Donated 2 | Previous Donations |
|---|---|---|---|---|---|
| Bellamy Brokkoli | 34 | W | Vegetable | Vegetable | No |
| Pierre Pêche | 36 | M | Fruit | Vegetable | Yes |
| Krish McClucky | 19 | M | Protein | Canned goods | Yes |
| Cathy Catsup | 32 | W | Canned goods | Vegetable | Yes |
| Jerome Junkphood | 36 | NB/NC | Canned goods | None | No |
| Jevad Baigan | 37 | Prefer not to answer | Vegetable | None | No |
| Alina Aloo | 39 | NB/NC | Vegetable | Vegetable | No |

# Practice Questions:

1. How would you anonymize this data set?
2. What are some variables you can use to aggregate this data?
3. Are there any outliers?
4. What is the average age of the donors? What is the median age?
5. Does this data set give you enough information to direct fundraising? What else might you need?
6. How might you share this data - in a report, on social media, etc.?