Critical Data Literacy

Critical Data Literacy

Strategies to Effectively Interpret and Evaluate Data Visualizations

NORA MULVANEY; AUDREY WUBBENHORST; AND AMTOJ KAUR

JOANNA ANNEKE RUMMENS

RYERSON UNIVERSITY
TORONTO

This book was produced with Pressbooks (https://pressbooks.com) and rendered with Prince.

Funding Acknowledgement

This project is made possible with funding by the Government of Ontario and through eCampusOntario's support of the Virtual Learning Strategy. To learn more about the Virtual Learning Strategy visit: https://vls.ecampusontario.ca.

Accessibility Statement

Accessibility features of the web version of this resource

The web version of Critical Data Literacy: Strategies to Effectively Interpret and Evaluate Data Visualizations has been designed with accessibility in mind by incorporating the following features to ensure it has been optimized for users of screen-reader technology:

- All content can be navigated using a keyboard
- Links, headings, and tables are formatted to work with screen readers
- Images have text descriptions (and data sources for graphs have been shared)
- Information is not conveyed by colour alone
- The option to increase font size (see tab on top right of screen) is available

Other file formats available

In addition to the web version, this book is available in a number of file formats including PDF, EPUB (for eReaders), MOBI (for Kindles), and various editable files. You can download this book in another file format (look for the "Download this book" drop-down menu to select the file type you want).

Those using a print copy of this resource can find the URLs for any websites mentioned in this resource in the footnotes.

Known accessibility issues and areas for improvement

While we aim to ensure that this book is as accessible as possible, we may not always get it right. There may be some supplementary third-party materials, or content not created by the authors of this book, which are not fully accessible. This may include videos that do not have closed captioning or accurate closed captioning, inaccessible PDFs, etc. Any issues we identify will be listed below. There are currently no known issues.

Accessibility standards

The web version of this resource has been designed to meet Web Content Accessibility Guidelines 2.0, level AA. In addition, it follows all guidelines in Accessibility Toolkit: Checklist for Accessibility. The development of this toolkit involved working with students with various print disabilities who provided their personal perspectives and helped test the content.

Let us know if you are having problems accessing this guide

We are always looking for ways to make our resources more accessible. If you have problems accessing this resource, please contact us to let us know, so we can fix the issue.

Please include the following information:

- The location of the problem by providing a web address or

page title

- A description of the problem
- The computer, software, browser, and any assistive technology you are using that can help us diagnose and solve your issue, e.g., Windows 10, Google Chrome (Version 65.0.3325.181), NVDA screen reader

If you are having problems accessing any content within the book, please contact: nmulvaney@ryerson.ca.

This statement was last updated on February 27, 2022

About This Course

Why We Made This Course

This short course began its life as a Library workshop developed by Nora Mulvaney, which aimed to teach strategies to critically assess data sources and visualizations and slow the spread of misinformation. The driving force behind developing the workshop was the glut of confusing data visualizations that emerged throughout 2020 as the world tried to make sense of the COVID-19 pandemic.

While the pandemic made line graphs and dashboards ubiquitous, we believe that critically interpreting data has now become an essential skill for everyday life. The ability to critically interpret data visualization will be crucial–across disciplines–for jobs of the future and the rapidly evolving labour market.

We are hopeful that the value of this short course will outlive the pandemic that inspired it.

Purpose of This Course

This short course is on the topic of data literacy that will enable you to become a more discerning and critical user of data, graphs, charts and infographics. The purpose is to deepen your critical engagement with data visualizations and develop your data literacy skills.

In Module 1, we provide a quick introduction to key concepts and theory related to data literacy and data visualizations. In Module 2, we investigate how to critically assess data sources and visualizations. In Module 3, you will learn how to find and

evaluate credible sources of data. Finally, in Module 4, we cover considerations for creating your own data visualizations and infographics.

Throughout the course, we provide examples and activities designed to test your ability to critically analyze and evaluate key elements of data visualizations and identify design choices that lead to misinformation.

What You Need

We suggest the following for this short course:

- An estimated 30 minutes per module. Depending on how many of the external links you explore, it may take you more time.
- Access to a computer and the Internet.

How to Navigate the Modules

We have designed this short course to be done at your own pace. You can navigate to the parts you find most interesting or relevant to your current information needs.

This version of the course is hosted in Pressbooks (a WordPress-based online platform). Pressbooks is used to host open courses and textbooks. Please see the video below on how to navigate Pressbooks.

One or more interactive elements has been excluded from this version of the text. You can view them online here:
https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=474#oembed-1

Learning Outcomes

These are the learning outcomes for the overall course. Each module will include a list of objectives specific to that module to guide you through your learning.

Learning Outcomes

By the end of this short course, you should be able to:

1. Discuss the different stages of the data journey
2. Explain why data visualizations are effective
3. Differentiate between different types of data visualizations
4. Analyze and break down a data visualization
5. Identify misleading features in a data visualization
6. Evaluate a data visualization using data literacy strategies

7. Assess the credibility of a data visualization's underlying data
8. Explain how to "fact check" a data source
9. Explain some of the key considerations for telling responsible stories with data
10. Outline some effective ways to communicate data

Not Covered in This Course

This course concentrates on critically interpreting data visualizations and the theory behind responsible design choices. We do not cover how to create data visualizations from scratch. We also do not cover other stages of the data journey, such as collecting or cleaning data.

MODULE 1: AN INTRODUCTION TO DATA LITERACY AND DATA VISUALIZATION

Introduction

The need to understand **data visualizations** has never been more important. Every day we are inundated with more data, graphs and charts. Some of these data visualizations are well-designed and easy to understand, and others are confusing and misleading.

Data literacy is often framed as a set of skills for data professionals, but we believe data literacy is for everyone. Everyone can benefit from improving their understanding of how data is created and their ability to analyze and interpret data.

In this module, we will introduce the key stages in the data journey, discuss why data visualizations can be so effective at conveying information, and provide an overview of different ways of presenting data and different types of data visualizations.

Learning Objectives

By the end of this module, you should be able to:

1. Discuss the different stages of the data journey.
2. Explain why data visualizations are effective.
3. Differentiate between different types of data visualizations.

The Data Journey

To get started, let's consider the data visualization[1] in Figure 1.1 below.

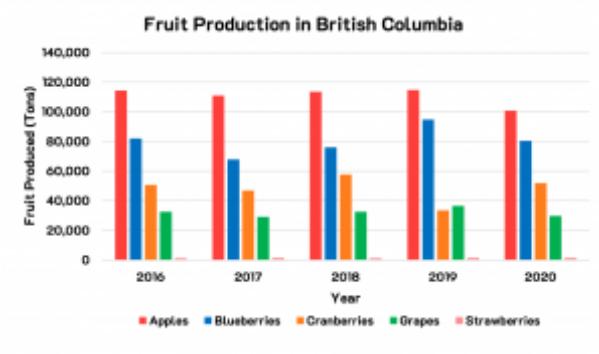

Figure 1.1. Production of apples, blueberries, cranberries, graphs, and strawberries in British Columbia, 2016-2020.

The underlying raw data went through many stages before it was presented to you in this data visualization. The information had to be:

- Collected via surveys
- Inputted into a database
- Stored on secure servers
- Cleaned for accuracy and consistency
- Analyzed to understand the trends
- Presented as a bar graph

1. Statistics Canada. Table 32-10-0364-01 Area, production and farm gate value of marketed fruits. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210036401-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

It is likely that multiple people, with complementary areas of expertise, worked on turning the raw dataset into a well-presented data visualization.

Deeper Dive

Watch this 4min video from Statistics Canada to learn more about the data journey: The Data Journey: What you need to know for successful navigation

Methods of Data Presentation

Once data has been collected and analyzed, there are many different ways you can communicate those results:

- Write a report describing your results
- Organize your results into a table
- Display your results visually in a chart or infographic

For example, if a university wanted to find out if their students preferred virtual or in-person classes, they might conduct a survey. That survey might ask students if they prefer attending their classes virtually or in-person, but would probably also include an option for students who wanted to do a mix of both and students who didn't have a strong preference.

When the data is ready to be presented, it could be written in text, perhaps as part of a report:

"Of the 1260 students surveyed, 560 students stated a preference for attending their classes virtually, 440 preferred in-person, 1006 stated that they would prefer a mix of both virtual and in-person classes and 194 did not have a strong preference."

This is a simple way of presenting the data, but does require reading through quite a bit of text, and doesn't allow for easy comparison between the results.

To really focus in on the numbers, this information could also be presented in a table, as in Table 1.1 below.

Table 1.1. Student Preferences for Attending Class

Preferences for Attending Class	Number of Students
Virtually	560
In-Person	440
Mix of Virtual and In-Person	1006
No Preference	194

The advantage of presenting this data in a table is that it's all about the numbers. This table makes it clear that most students prefer a mix of virtual and in-person classes. However, it still doesn't allow for immediate comparison between all of the results. When you are looking to compare more than two numbers, it is often better to present the data in a more visual way.

To make it easier to compare the results, this information could also be presented visually in a bar graph[1] (see Figure 1.2 below).

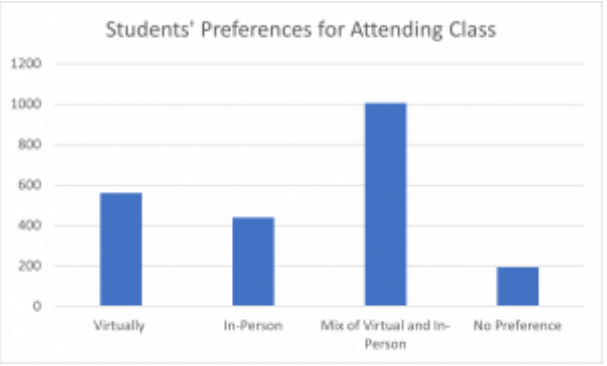

Figure 1.2. Student preferences for attending class.

From looking at this graph, it is clear straight away that most students prefer a mix of virtual and in-person classes. This illustrates how powerful data visualizations can be; you can tell a clear story without using words or numbers.

Key Takeaways

1. Author: Nora Mulvaney, License CC0 1.0 https://creativecommons.org/publicdomain/zero/1.0/

There are three main methods of data presentation:

- Textual
- Tabular
- Visual

Different Types of Data Visualizations

If you've decided that you want to represent your data visually, you'll also need to think about which type of data visualization would best represent your data. Choosing the right kind of visualization depends on many factors:

- What kind of data do you have?
- Are there patterns in the data? Do they change over time?
- What is the purpose of your data visualization?
- Who is your intended audience?

Keep reading for an introduction to some basic types of data visualizations.

Bar Chart

A bar chart (also known as a bar graph or a column graph) is a good option for making comparisons between different groups of things and for identifying patterns. Typically, a bar chart uses either vertical or horizontal bars to show numerical comparisons across categories. One axis of the chart shows the categories being compared and the other axis displays the value scale of the numbers that are being compared. Bar charts with a large number of bars can become difficult to read

as the labels for the categories will be more difficult to display clearly. (See Figure 1.3[1] below for an example.)

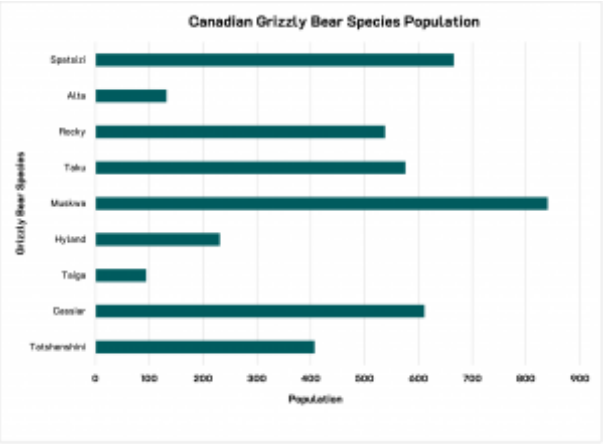

Figure 1.3. Populations of various Canadian grizzly bear species.

Key Takeaways

A bar graph is good for making comparisons and identifying patterns.

1. Government of British Columbia. Grizzly Bear Population Units. Retrieved January 8th, 2022. URL: https://open.canada.ca/data/en/dataset/caa22f7a-87df-4f31-89e0-d5295ec5c725 Government of British Columbia, Contains information licensed under the Open Government Licence – British Columbia. https://www2.gov.bc.ca/gov/content/data/open-data/open-government-licence-bc

Pie Chart

A pie chart is a good option for displaying the composition of a whole or the proportional distribution of the data. A pie chart consists of a circle that is divided into proportional segments, with the full circle representing the total sum of the data. (See Figure 1.4[2] below for an example.)

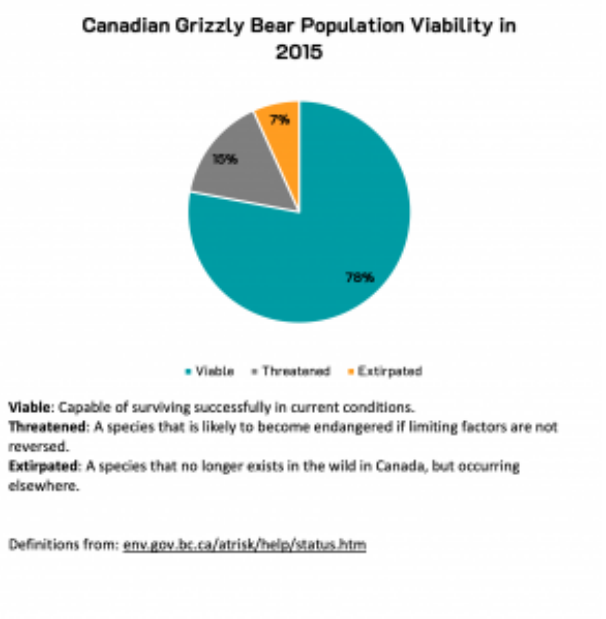

Figure 1.4. Viability of Canadian grizzly bear species in 2015.

2. Government of British Columbia. Grizzly Bear Population Units. Retrieved January 8th, 2022. URL: https://open.canada.ca/data/en/dataset/caa22f7a-87df-4f31-89e0-d5295ec5c725 Government of British Columbia, Contains information licensed under the Open Government Licence – British Columbia. https://www2.gov.bc.ca/gov/content/data/open-data/open-government-licence-bc

Key Takeaways

A pie chart is: good for making comparisons, indicating proportional representation and showing part-to-a-whole

Line Graph

A line graph is used to display quantitative values over a time period or other continuous interval, and is usually used to display trends and portray change over time. A line graph can also be used for comparison when grouped with other lines; this can, however, become cluttered and difficult to read if you are using more than 3-4 lines per graph. (See Figure 1.5[3] below for an example.)

3. Statistics Canada. Table 32-10-0356-01 Area, production and sales of mushrooms. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210035601-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

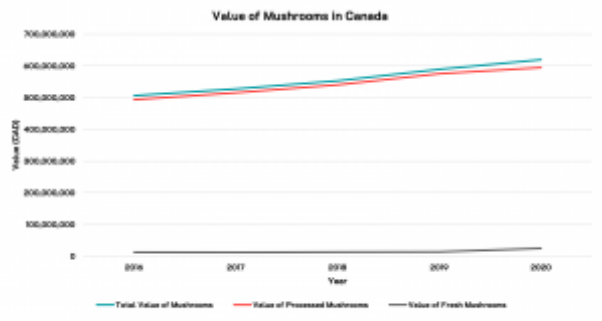

Figure 1.5. Value of processed and fresh mushrooms, plus the total value of those mushrooms together in Canadian Dollars (CAD), 2016-2020.

Key Takeaways

A line graph is good for identifying patterns, showing data over time and making comparisons (when the lines are grouped in a single graph).

Scatterplot

A scatter plot (also known as a scatter graph, scatter chart, point graph or X-Y plot) uses a collection of points to display values from two variables. By displaying the values in each axis,

you can detect if a correlation or relationship exists between the two variables. (See Figure 1.6[4] below for an example.)

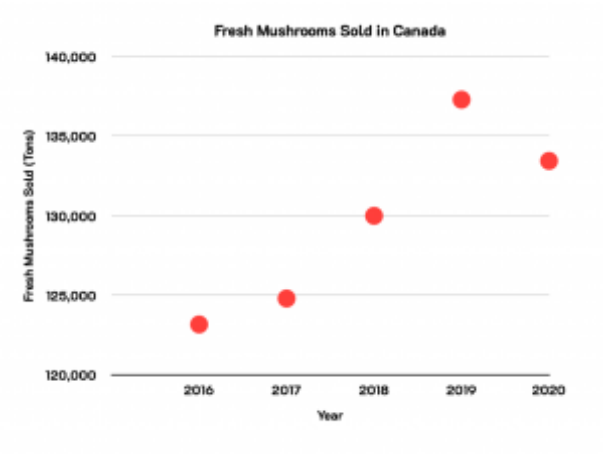

Figure 1.6. Fresh mushrooms sold in Canada by year, 2016-2020.

Key Takeaways

A scatterplot is good for identifying patterns and relationships.

4. Statistics Canada. Table 32-10-0356-01 Area, production and sales of mushrooms. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210035601-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

Maps

A map can display data in many different ways. Maps can display divided geographical areas that are coloured, shaded or patterned in relation to a data variable. Maps can also be used to detect spatial patterns or the distribution of data over a geographical region using a gradient colour scheme. (See Figure 1.7[5] below for an example, where each province is a different colour representing its raspberry production.)

Figure 1.7. The amount of raspberries produced in Canada by province in 2020.

Key Takeaways

5. Statistics Canada. Table 32-10-0364-01 Area, production and farm gate value of marketed fruits. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210036401-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

A map is good for mapping distribution, identifying patterns, and making comparisons.

Infographic

An infographic consists of several data visualizations, and combines information and graphics (as the name suggests) to tell a clear data story. (See Figure 1.8[6] below for an example.)

6. Statistics Canada. Table 32-10-0364-01 Area, production and farm gate value of marketed fruits. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210036401-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

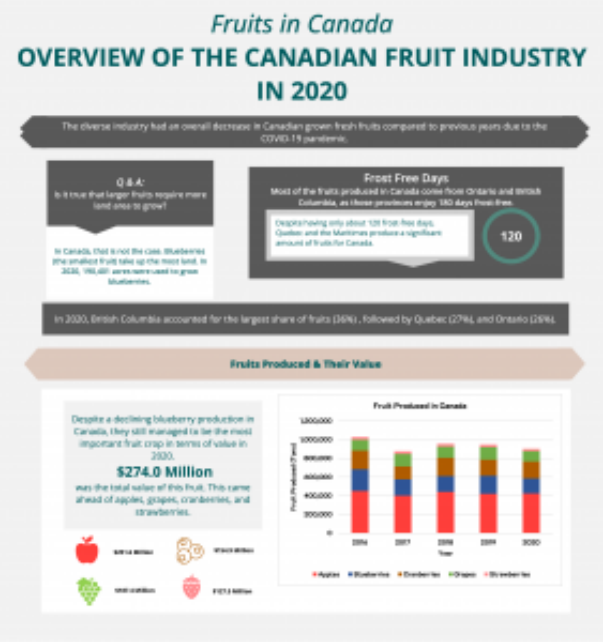

Figure 1.8. Overview of the Canadian fruit industry in 2020 by labour statistics, types of fruit produced and their comparative value.

Key Takeaways

An infographic is good for telling a comprehensive data story.

Dashboard

A dashboard is when several data visualizations are put

together. Dashboards can use tables, charts and graphics to display information, often to inform business decisions. Dashboards are usually updated regularly and show changes over time. (See Figure 1.9[7] below for an example, which includes a pie chart and a bar chart presented under a header that describes their relation to each other.)

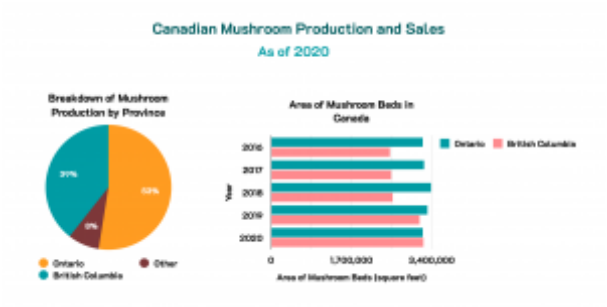

Figure 1.9. The dashboard displays Canadian mushroom production and sales for 2020 using a pie chart and a bar chart.

Key Takeaways

A dashboard is very versatile and is often used in a variety of ways to inform decision making.

7. Statistics Canada. Table 32-10-0356-01 Area, production and sales of mushrooms. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210035601-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

Exercises

Take a moment to test your understanding of the different types of data visualization. In the activity below, you'll find a series of scenarios. Match the type of data visualization to the most appropriate use.

An interactive H5P element has been excluded from this version of the text. You can view it online here: https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=263#h5p-13

Deeper Dive

Want to learn more about different types of data visualizations? Check out the Data Visualization Catalogue.

The Power of Data Visualizations to Tell Impactful Stories

Data visualizations are powerful because they help us identify patterns, make sense of relationships between variables, understand trends … and because they enable us to more readily tell a story.

Florence Nightingale

One of the most famous examples of the power of data visualizations to persuade involves Florence Nightingale. She is probably best known as the founder of modern nursing, but she was also a statistician and early pioneer of data visualizations.

Nightingale tended to soldiers during the Crimean War (1853 to 1856) where she observed that more soldiers were dying from diseases caught in camp than on the actual battlefield. For two years, she meticulously gathered mortality data and arranged them in very detailed tables that contained several columns and rows upon rows of numbers.

She subsequently conveyed the patterns she found in her data in what has come to be known as the Nightingale Rose

Chart (see Figure 1.10[1], below), also known as a polar area diagram or a 'coxcomb' chart.

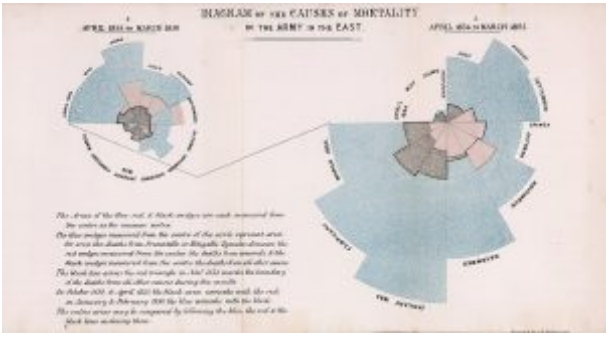

Figure 1.10. Diagram of the causes of mortality in the army in the East (1858) by Florence Nightingale, April 1854-March 1855 and April 1855-March 1856.

This visualization might look a bit like a pie chart, but it was designed to show change over time with each triangular wedge indicating the month of the year. The different colours convey the cause of death: blue represents soldier deaths from infections and preventable diseases, red were those from wounds, while black indicates deaths from accidents and other causes.

Nightingale used data visualization to convey meaningful patterns in her data and to communicate them to others in a very compelling way. Her "Diagram of the causes of mortality in the army in the East" was published in Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army and sent to Queen Victoria in 1858.

Shared with leading influencers and decision makers, her

1. Florence Nightingale, Public domain, circa 1858 via Wikimedia Commons https://commons.wikimedia.org/wiki/File:Nightingale-mortality.jpg

work sparked a revolution in hygiene and health care that saved countless lives, both in battlefield and public hospitals around the world.

Nightingale's story teaches us that clearly conveyed data can inform actions that have tremendous impact.

Deeper Dive

Interested in more about Florence Nightingale's use of data visualizations?

Listen to episode of the 99% Invisible podcast: Episode 433 – Florence Nightingale: Data Viz Pioneer.

Summary

In this module, you learned about the stages in the data journey and the different ways data can be presented, different types of data visualizations, were introduced to some basic types of data visualizations and learned about the power of data visualizations to tell impactful and persuasive stories.

Now that we understand the basics of data visualization, we can dive deeper into analyzing and breaking down data visualizations in order to identify any misleading features.

Next up, in Module 2, we are going to learn how to critically assess data visualizations and identify misleading features.

MODULE 2: HOW TO CRITICALLY ANALYZE AND INTERPRET DATA VISUALIZATIONS

Introduction

As we learned in the first module, data visualizations are an effective way to present information because they allow us to discover insights relatively quickly and easily–at least compared to sifting through the original data ourselves.

But are data visualizations always easy to understand? Can they always be interpreted quickly?

Not necessarily. Some data visualizations will require more time to analyze and understand, either because they are complex or confusing or just not what we were expecting.

In this module, we will discuss strategies and tools for critically assessing data visualizations and will identify features of data visualizations that can be confusing or manipulative.

Learning Objectives

By the end of this module, you should be able to:

1. Analyze and break down a data visualization.
2. Identify misleading features in a data visualization.
3. Evaluate a data visualization using data literacy strategies.

Strategies for Critically Assessing Data Visualizations

Critically assessing data visualizations can be tricky, so we've broken it down into several steps:

- Step 1: Stop and Slow Down
- Step 2: Separate the Scaffolding from the Visual Encoding
- Step 3: Focus on Scaffolding
- Step 4: Focus on Content and Visual Encoding
- Step 5: Step Back to Spot Patterns and Relationships
- Step 6: Examine Your Own Biases

Keep reading to learn more about each step.

Step 1: Stop and Slow Down

Stop.

The first step is easy in principle, but often hard to do. Before you jump to any conclusions about the data visualization in front of you, stop, slow down and take your time to understand it.

As we've discussed, data visualizations can be very powerful, persuasive and sometimes overwhelming. When you come across a chart, graph or other visualization you haven't seen before, don't feel pressured to understand it straight away. Take the time to understand what the visualization is trying to say by analyzing its features and contents, as any patterns, relationships or outliers that you notice. More on this in the next steps!

Step 2: Separate the Scaffolding From the Visual Encoding

Now that we aren't feeling rushed, it may be helpful to break a data visualization down into its scaffolding and content:

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=538#h5p-14

Exercises

Take a moment to test your understanding of the different features of a data visualization. In the activity below, you'll see an example of a data visualization. Click on the hotspots to learn more about which features of the data visualization are the scaffolding and which are the content.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=538#h5p-9

Step 3: Focus on Scaffolding

Type of Data Visualization

To get the background information you need to interpret a data visualization, start by focusing on its scaffolding and considering some of the following questions:

- Is the chart appropriate for the data?
- Does the format enable clear comparison of the data?

The type of chart will determine the questions you need to ask, so there may be more questions to consider.

For example, review Figure 2.1[1] below. What type of data visualization is this? Focus on the content highlighted in the graph area.

1. Statistics Canada. Table 37-10-0079-01 Registered apprenticeship training, registrations by major trade groups and sex. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/3710007901-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

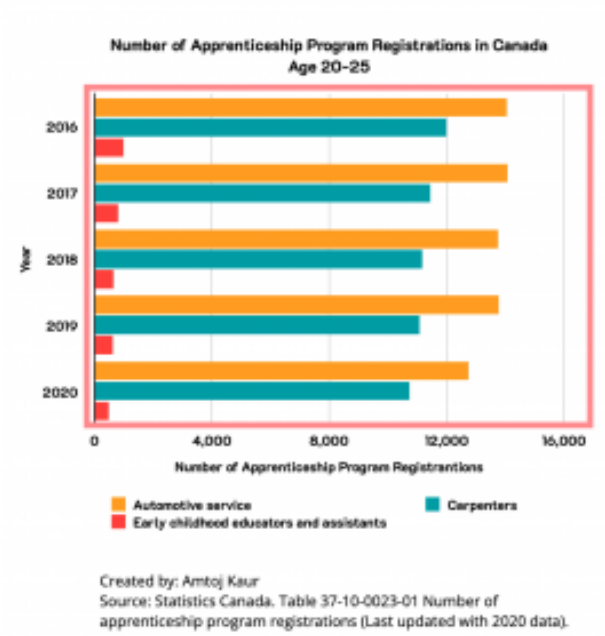

Figure 2.1. Number of apprenticeship program registrations in Canada (age 20-25). The content of the graph is highlighted in pink to show that this is a bar graph.

Title

Next, consider some of the following questions as you review the data visualization's title:

- Is there a title?
- Does the title tell you what to expect from the chart?
- Does the chart support the title?

For example, review the title in Figure 2.2 below (found above the graph area).

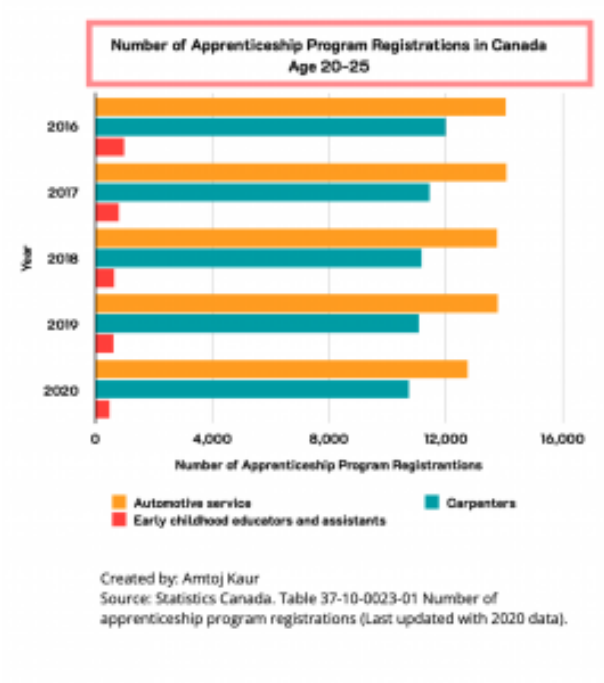

Figure 2.2. The title of the bar graph (Number of Apprenticeship Program Registrations in Canada, Age 20-25) is highlighted in pink.

Key/Legends

Next, consider some of the following questions as you review the data visualization's key/legends:

- Is there a key?
- Does it explain the features of the data visualization (colours, shapes, etc.)?
- Are data points labelled?

For example, review Figure 2.3 below for the information in its key/legend (found below the graph area).

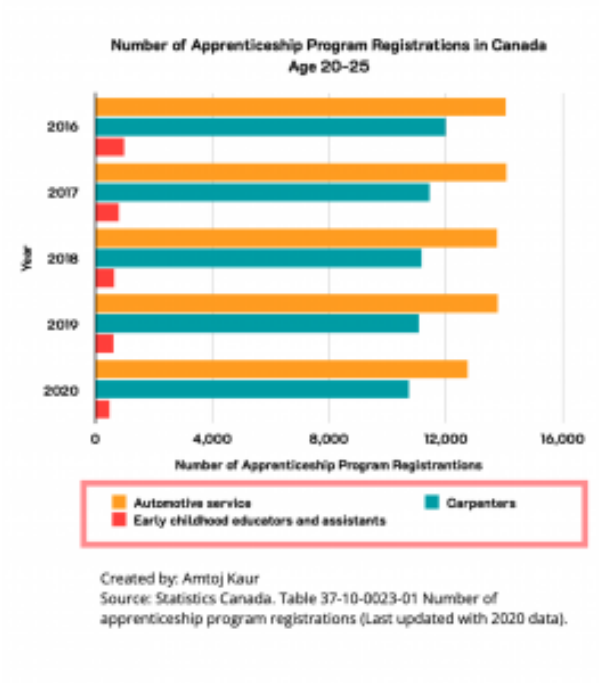

Figure 2.3. The legend in the bar graph is highlighted in pink and includes automotive service, carpenters and early childhood educators and assistants.

Scales

Next, consider some of the following questions as you review the data visualization's scales:

- What are the scales?
- Do they start at zero? [Note: there are lots of good reasons to not start the scale at zero, but pay attention to it!]
- Are the scales complete?
- Are the axes appropriate for the data?

For example, review information in the scales for Figure 2.4 below (which run across each of the two axes of the graph).

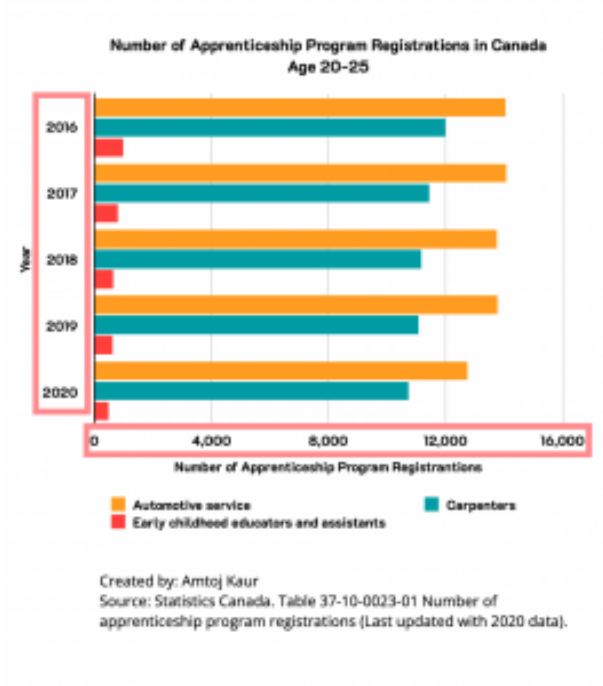

Figure 2.4. The scales in the bar graph are highlighted in pink and include dates for the year (2016-2020) and number of apprentices hip program registration s (0-16,000).

Bylines

Next, consider some of the following questions as you review the data visualization's byline:

- Who made the data visualization?
- Did they also collect and analyze the underlying data?

For example, review information in the byline for Figure 2.5 below (which is found below the key/legend for the graph).

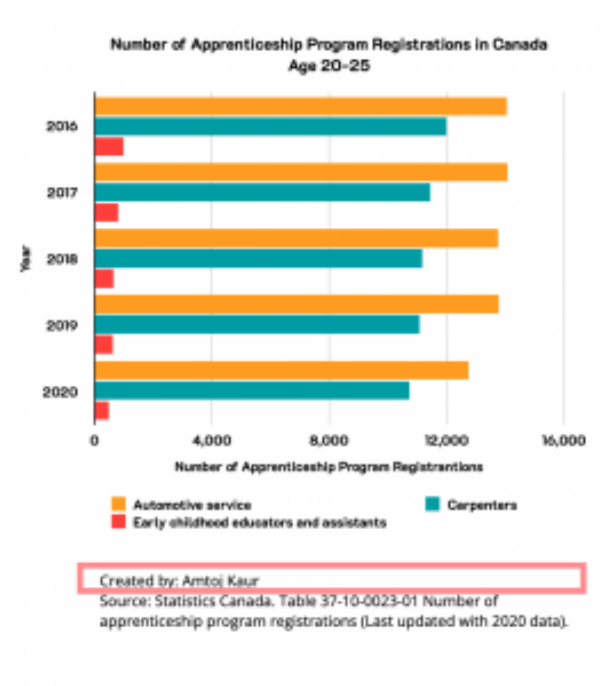

Figure 2.5. The byline in the bar graph (created by: Amtoj Kaur) is highlighted in pink.

Sources

Next, consider some of the following questions as you review the data visualization for its source:

- Is it clear where the data come from?
- Is the source reputable? (More on this in the next module!)

For example, review information in the source for Figure 2.6 below (which is found below the source information).

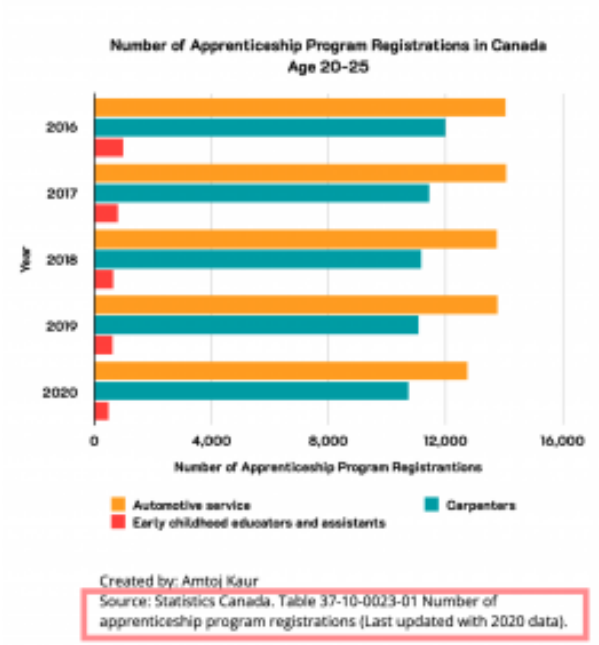

Figure 2.6. The source in the bar graph is highlighted in pink, and is shared as "Source: Statistic Canada. Table 37-10-0023-01. Number of apprenticeship program registrations (Last updated with 2020 data)".

After reviewing the data visualizations scaffolding, you should have a better sense of what the visualization is trying to convey and are ready to dive into analyzing the content.

Step 4: Focus on Content and Visual Encoding

The most important element of most data visualizations is their content and how that content is **visually encoded**.

Essentially, data visualizations are always built with symbols that follow a set of rules. This might sound a bit complicated, so let's break it down using Figure 2.7[1] below, a standard bar chart of monthly average retail prices for two food categories in Canada from January to July 2021.

1. Statistics Canada. Table 18-10-0002-01 Monthly average retail prices for food and other selected products. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/1810000201-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

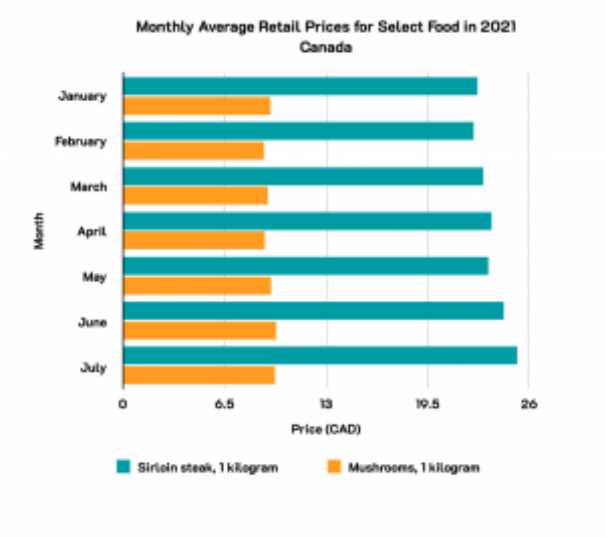

Figure 2.7. A standard bar graph of monthly average retail prices in Canada, comparing the cost of 1 kg of steak vs. 1 kg of mushrooms from January to July 2021.

Let's focus on the visual encoding, and the "rules" that this chart is following. Click on each method below to learn more about the "rules" that are being followed in Figure 2.7.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=542#h5p-18

This may all seem obvious, but it is important to understand the "rules" of the data visualization that you are looking at to interpret it correctly. This will also help you recognize when the rules are being broken to manipulate the data or alter the way that you perceive the data.

Exercises

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=542#h5p-16

Step 5: Step Back to Spot Patterns and Relationships

Now that you've spent some time identifying the features that support the data visualization (the scaffolding) and how the content itself is represented (visual encoding), you're ready to step back and look for patterns, relationships and trends in the data. Figure 2.7[1] below displays an increase in data.

1. Statistics Canada. Table 37-10-0079-01 Registered apprenticeship training, registrations by major trade groups and sex. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/3710007901-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

Figure 2.8. A scatterplot shows the increase in apprenticeship registrations (2,000-11,000) for several trades over the course of 4 years: stationary engineers and power plant operators, landscapers, and refrigeration and air conditioning mechanics.

Some patterns or relationships may be immediately obvious, and others may take more time to make sense of.

Again, take your time.

Step 6: Examine Your Own Biases

Sometimes data visualizations tell us what we want to hear. Or so we think.

In general, are more likely to believe things that line up with what we already believe and less likely to believe things that contradict our beliefs. This is an example of **confirmation bias.**

Reflection

Keep the issue of confirmation bias in mind when you're trying to make sense of a graph, chart or other data visualization.

Ask yourself: what am I expecting this data visualization to say or what am I hoping it will prove? Then take a minute to consider whether or not the content of the data visualization supports your expectations.

Misleading Data Visualizations

Now that we know how to analyze and break down a data visualization, let's go through a few examples of design choices (and mistakes!) that can create confusion.

Using the Wrong Type of Data Visualization

As we learned in Module 1, some types of data visualizations work well for communicating specific types of information, but not others. For example, pie charts are good for making comparisons between a few different categories, but are not great for identifying patterns or showing data over time.

Data visualizations can be confusing and misleading when the designer has picked a format that isn't well suited to the data they are analyzing.

Review Figure 2.9[1] below, a pie chart of Ontario television viewing in 2004. There are 12 categories of television and similar colours used in the graph, as well as white font over the bright colours, making this hard to read.

1. Statistics Canada. Table 22-10-0097-01 Television viewing time of all television stations, by province, content and type of programme. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/2210009701-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

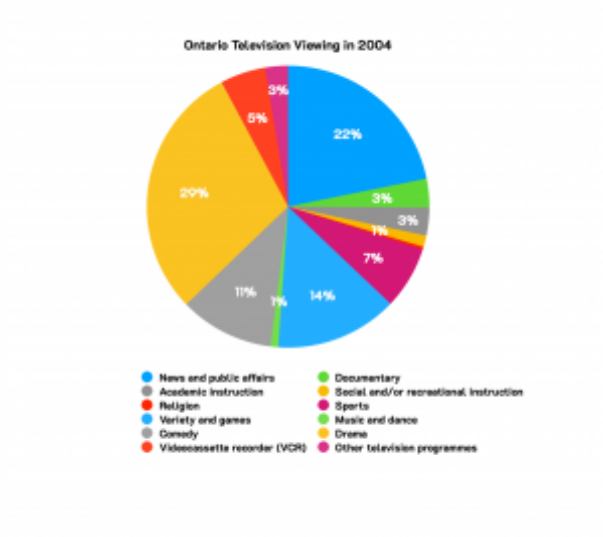

Figure 2.9. A pie chart displaying 12 categories of television viewing in Ontario in 2004 provides too much visual information, making it hard to read.

False Causation

Correlation does not imply causation.

If you've ever taken a statistics or data analysis course, you have almost certainly come across this common phrase. It means that, just because two trends seem to fluctuate alongside each other, it doesn't prove that one causes the other or that they are related in a meaningful way.

Review Figure 2.10[23] below, which shows a line graph of the

2. Statistics Canada. Table 37-10-0079-01 Registered apprenticeship training, registrations by major trade groups and sex. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/3710007901-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence
3. Statistics Canada. Table 32-10-0364-01 Area, production and farm gate

decrease of Canadian automotive apprenticeship registrations and nectarine production. What do these two things have to do with each other? They are unrelated quantities that appear to decrease at the same rate over a similar time period.

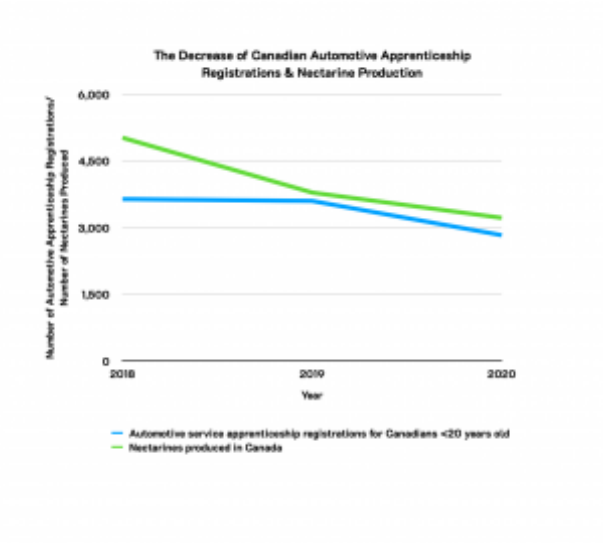

Figure 2.10. A line graph of the number of nectarines produced and automotive apprenticeship registrations in Canada appears to show a relation, as they both begin to decrease in 2019, where there is none.

Inconsistent or Manipulated Scale

It's important to examine the scales of a data visualization carefully. Compressing or expanding the scale of a graph can make the changes between data points seem either more or less significant than they really are.

value of marketed fruits. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210036401-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

Review Figure 2.11[4] below, which shows the cost of sugar in Canada from January to July 2021. Because of the expanded scale on the line graph, there does not appear to be much fluctuation in the cost of sugar in Canada. This makes the data appear less significant than it could really be (see Figure 2.12 below for a more compressed scale).

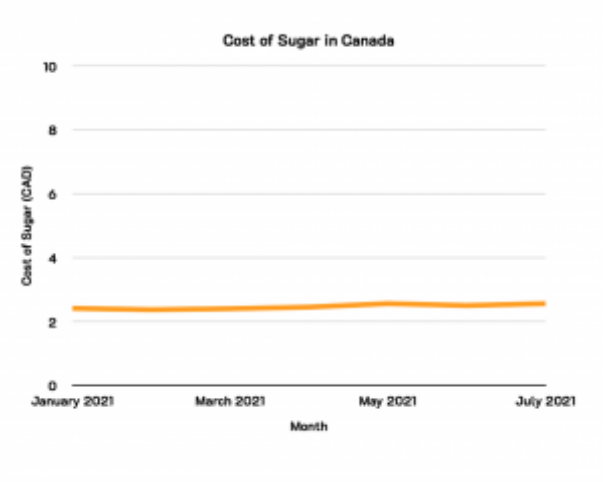

Figure 2.11. A line graph showcasing the price of sugar in Canada with an expanded scale.

4. Statistics Canada. Table 18-10-0002-01 Monthly average retail prices for food and other selected products. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/1810000201-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

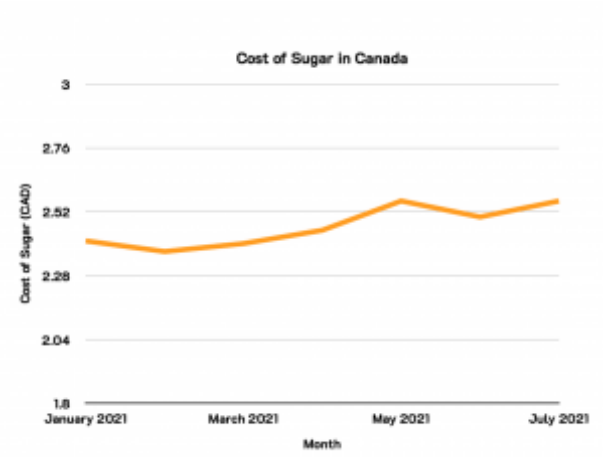

Figure 2.12. A line graph showcasing the price of sugar in Canada with a more compressed scale.

Cherry-picking or Omitting Data

The term "cherry-picking" refers to only presenting the best data, and omitting data points which are less favourable, in order to reinforce a particular narrative. This can create a false impression of the data. For example, showing an upward sales trend over the first few months of a year, while omitting the data that showed sales declined for the rest of the year.

Review Figure 2.13[5] below, which shows a downward trend on gasoline prices in Canada from May 2019 to February 2020. Because of the carefully selected timeframe (i.e., short

5. Statistics Canada. Table 18-10-0002-01 Monthly average retail prices for food and other selected products. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/1810000201-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

timeframe), it appears that the gasoline prices in Canada are decreasing.

Figure 2.13. A line graph showing a downward trend on gasoline prices in Canada over a short timeframe (May 2019-February 2020).

Now review Figure 2.14[6] below, which shows an overall upward trend on gasoline prices in Canada from May 2019 to November 2021. When looking at the full timeline (i.e., long timeframe), the reader can see that gasoline prices are increasing in Canada.

6. Statistics Canada. Table 18-10-0002-01 Monthly average retail prices for food and other selected products. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/1810000201-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

Figure 2.14. A line graph showing an overall upward trend on gasoline prices in Canada over a longer timeframe (May 2019-November 2021).

3D Distortion or Occlusion

Three-dimensional (3D) data visualizations may look visually appealing, but they often make it more difficult to interpret the data and spot patterns within them. Two common issues are: distortion and occlusion.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=198#h5p-17

Review Figure 2.15[7] below, which is a 3D bar graph of the

7. Statistics Canada. Table 22-10-0097-01 Television viewing time of all

percentage of Canadian vs. foreign television programmes watched in Saskatchewan from 2000 to 2003. Because of the tilt of the 3D bar graph, the bars in the front hide the bars in the back, making it hard to read. The reader cannot pinpoint the exact percentage of Canadian vs. foreign programmes by the year it is presented.

television stations, by province, content and type of programme. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/2210009701-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

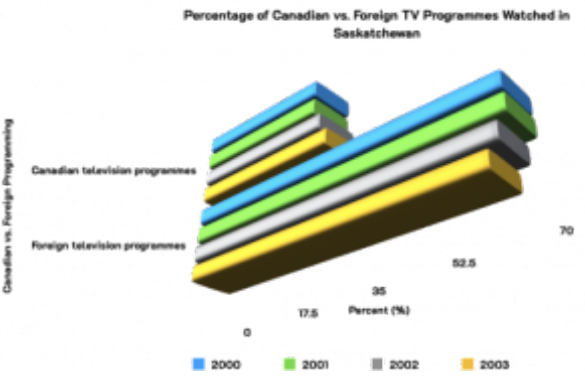

Figure 2.15. A 3D bar graph comparing the percentage of Canadian television programmes vs. foreign television programmes in Saskatchewan over 2000-2003. It is hard to read, as the reader cannot pinpoint the exact percentage of Canadian vs. foreign programmes by the year they are presented.

The Colour Scale

When used thoughtfully, colour can make it easier to spot trends and relationships in a data visualization. However, colour can also cause confusion.

Some common issues include: using too many colours, using colours with minimal contrast, using colours that aren't safe for colourblind viewers and using colours in unconventional

ways. Review Figure 2.16[8] below, which is a line graph of the percentage of Canadian vs. foreign television programmes watched in New Brunswick from 2000 to 2004. Because of the similar colours of the lines, it is difficult for the reader to understand which line graph corresponds to which colour from the legend.

8. Statistics Canada. Table 22-10-0097-01 Television viewing time of all television stations, by province, content and type of programme. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/2210009701-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

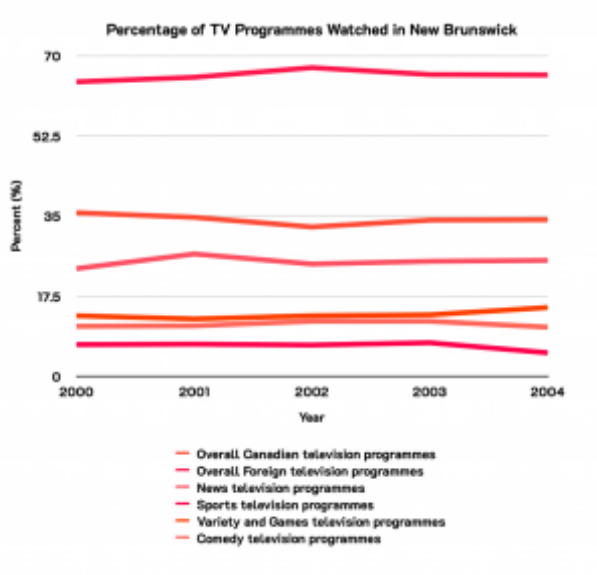

Figure 2.16. A line graph presents the percentage of Canadian, foreign, news, sports, variety and games, and comedy TV programmes watched in New Brunswick over the years 2000-2004. Similar colours used in the line graph make it difficult for the reader to understand which line graph corresponds to which colour in the legend.

Data Visualizations and Misinformation

We've all probably encountered a data visualization that was confusing and difficult to understand. It probably wasn't intentional, but, rather, the consequence of poor design choices, not fully understanding the dataset, using the wrong dataset for their purposes, or just rushing and making sloppy errors.

Usually misleading data visualizations fall into the realm of **misinformation** rather than **disinformation**, but both can cause harm.

Because data visualizations can be very persuasive and invoke strong emotional reactions, it is important to go through the steps outlined earlier in the module to make sure you fully understand what it is saying before you decide to share it with someone else.

Exercises

Take a moment to test your ability to interpret a data visualization. In the activity below, you'll see an example of a data visualization. Following the steps outlined earlier in the module, analyze this visualization and answer the questions that follow.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=204#h5p-11

Reflection

After analyzing this data visualization, would you feel confident sharing it with your friends and colleagues?

Summary

In this module, you learned critical strategies to analyze data visualizations, common design and data issues that result in confusing and misleading data visualizations and how misleading data visualizations can contribute to the spread of misinformation.

Now that you understand how to break down the features of a data visualization, you can dive deeper into finding reliable and trustworthy data.

Next up, in Module 3, we will learn about finding and evaluating credible sources of data.

MODULE 3: ASSESSING DATA CREDIBILITY

Introduction

Data visualizations can only be effective and meaningful if they are based on high quality data. It is important to source credible, current data sources. This module will provide some guidance as to how to select quality data for your visualizations.

Learning Objectives

By the end of this short course, you should be able to:

1. Assess the credibility of a data visualization's underlying data.
2. Explain how to "fact check" a data source.

Where to Find Data

Many leading institutions (governments, media, universities and corporations) share their data. This is done to ensure that there is transparency in decision-making and reporting. Open data is a relatively new phenomenon and refers to data sources that are available free to be downloaded. Data sets are open to anyone to modify, reuse and share. Here are some open data web sites:

- Google Public Data Set
- Five Thirty-Eight – Web site for data-driven journalism and story-telling
- Government of Canada Open Data Sets – Data sets on a wide variety of topics
- City of Toronto Open Data Resources – Variety of data sets regularly updated by the City of Toronto

Start With Good Quality Data

When creating a data visualization, you need to ensure that your data source is legitimate and credible. Choosing your data source is critical. Given the glut of information available online, it is important to sift through it to find high quality information. Ask yourself the following questions:

- Is this source trustworthy? Is the data verifiable?
- Can I share this data? Are there privacy concerns or other harms to be aware of?
- Is this data or information appropriate for the story and message?
- Does it reveal something that makes me uncomfortable? How will I make sure unconscious biases are not affecting my interpretation of it?

Deeper Dive

A good lens to use is the CRAAP test. Watch this video, Evaluating Sources, to learn more.

The CRAAP test asks you to consider: currency, relevance, accuracy, authority and purpose. Browse through the list below for a definition of each.

An interactive H5P element has been excluded from this version of the text.

You can view it online here:

https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=271#h5p-6

Beware of Bias

Be aware that some data sources may be biased such as:

- Organizations reporting on themselves
- Data that is generated by interest groups
- Data that is self-reported where they may be room for embellishment or incentives to inaccurately report (e.g. individuals reporting their own salary data)

Review the data sets you are using and make sure that it makes sense. Review how the data is collected and how terms are defined. Some knowledge and research on the topic will help. Consider: new sets of data against past years, data series that shows drastic changes should be investigated and understood before it is presented. It may not be the quality of the data that needs to be considered but how it is presented.

Examples

Data may not be biased exactly but may be socially constructed. For instance, here is a map showing racial change in Hartford, Connecticut from 1900-2018[1]. Over

1. "Steven Manson, Jonathan Schroeder, David Van Riper, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 14.0 [Database]. Minneapolis, MN: IPUMS. 2019. DOI:

time, definitions of race have changed and new terminology has emerged and become commonplace. In developing illustrations to visualize this data, you would want to be careful to acknowledge these changes. The explanation at the bottom of the graph helps to explain this as accurately as possible. There is not necessarily one correct way to display this data. When developing the visualization, clearly explain your choices and limitations.

How to Recognize Bad Data

As much as possible, try to recognize bad data. The following could be red flags:

- **Empty/blank cells:** Ask if the respondents did not answer this information or if it is simply incomplete.
- **Data that doesn't make sense:** For instance, dates should be in a date format. Postal codes should be written as Letter/Number/Letter Space Number/Letter/Number.

Many open data sets come with source notes. Take the time to review the notes to understand how the data was collected and what it does (and doesn't) represent.

http://doi.org/10.18128/D050.V14.0. Retrieved February 14, 2022 from https://ontheline.github.io/otl-racial-change/index-caption.html"

Look Closely at Your Data

Watch out for Simpson's Paradox! Edward Simpson described this phenomenon in 1951. It is often associated with an analysis of potential gender discrimination at University of California, Berkeley (UC Berkeley). In the 1970s, there was a concern about gender discrimination with regard to student acceptance. At an aggregate level, it appeared as though there may be an issue.

However, once the data was further broken down into groups and their acceptance rates to investigate the situation further, it turned out that the gender bias was for women instead of against them in four departments, while there was no bias in the other departments.

What had led to the data coming across as the reverse of what was actually happening was due to Simpson's Paradox as a result of most women actually applying to departments with lower acceptance rates in comparison to men. This situation tells us to look at data from various angles and break it down further if possible to avoid something like Simpson's Paradox situation arising from any hidden variables in the data. And not only to just look at charts and numbers when making decisions, but also to take the time to disaggregate the data as required.

Ensure Your Comparisons Are Meaningful

Data visualization helps to compare information. You hear the term “apples to apples” comparisons. This means comparing data sets that are alike. Two sets of statistics that are being used to derive a comparison need to be meaningful comparators.

Common Descriptive Statistics

Browse through the list below for commonly used descriptive statistics:

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=278#h5p-7

You may also need to adjust for time. For example, when looking at household income data from 1970 to 2021, you will need to adjust for inflation. [See the Bank of Canada’s Inflation Calculator.] Often, economists distinguish between nominal data (unadjusted) versus real data (adjusted over time). See

Figure 3.1[1] for a line graph example of how nominal and real data differ.

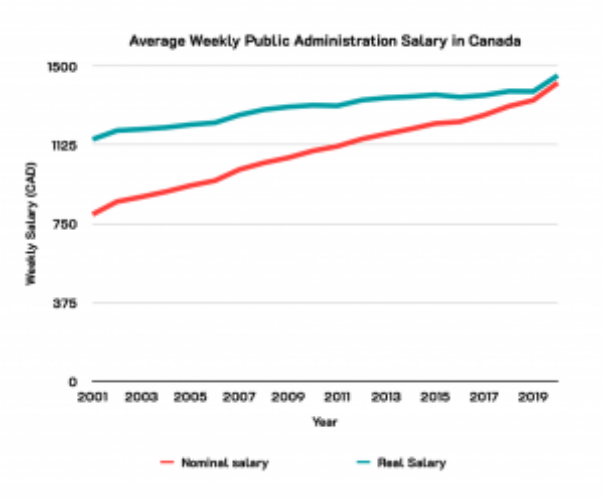

Figure 3.1. The "real" salary in this graph considers inflation as of 2021 by showing what the 2021 equivalent of the "nominal" salary is.

Check your knowledge using the line graph in Figure 3.2[2] with the activity that follows.

1. Statistics Canada. Table 14-10-0204-01 Average weekly earnings by industry, annual. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 9th, 2022. DOI: https://doi.org/10.25318/14100204401-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence
2. Statistics Canada. Table 32-10-0364-01 Area, production and farm gate value of marketed fruits. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210036401-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

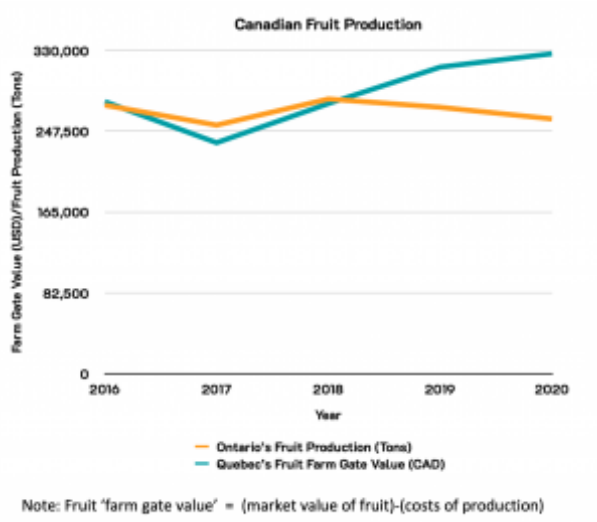

Figure 3.2. Canadian fruit production shows how much fruit is produced in Ontario and the change in Quebec's fruit farm gate value, 2016-2020.

An interactive H5P element has been excluded from this version of the text. You can view it online here:
https://pressbooks.library.ryerson.ca/criticaldataliteracy/?p=278#h5p-8

Avoid Confirmation Bias

When looking for data, we may be seeking to confirm what we already believe. While this may be human nature, it is important to find objective sources of data and let the analysis speak for itself. Reviewing good quality data can help to

challenge our assumptions and ensure that we retain as much objectivity as possible.

Citing Your Data

Although you are likely creating your own, original data visualizations, they are based on external data sources. Any reader who is looking at your data visualization should be able to find its original source. Don't forget to cite your data source that you used to create your visualization. Review some tips on citing data from the Ryerson University library.

Summary

In this module, you learned how to find good quality data, examine a data source for bias, ensure comparisons are meaningful and cite your data appropriately.

Now that you understand how to find and evaluate credible sources of data, you can dive deeper into responsible storytelling with data.

Next up, in Module 4, we will learn some considerations for creating your own responsible data visualizations.

MODULE 4: RESPONSIBLE DATA STORYTELLING

Introduction

In addition to using good quality data, you should also consider how you are presenting the data. It is important to communicate the data effectively and accurately to ensure that you are not misleading your audience.

Learning Objectives

By the end of this module, you should be able to:

1. Explain some of the key considerations for telling responsible stories with data.
2. Outline some effective ways to communicate data.

Advantages of Data Visualization

Why visualize data? Data can help to convey important information in a meaningful and persuasive way. Visualizing data has many advantages:

- Complex data can be easily understood.
- A simple visual representation of outliers, target audiences and futures markets can be created.
- Storytelling may be more effective with dashboards and animations.
- Data can be explored through interactive visualizations.

Data Visualization Expert Edward Tufte

Statistician and artist Edward Tufte wrote a book The Visual Display of Quantitative Information which provides a history of data graphics. Although the book was published in 1983, it is still considered one of the key books in this area. Tufte uses the term **chartjunk** to describe poor data visualizations that either don't inform at all or providing misleading information due to the way the data is presented. He also recommends a strong **data-ink ratio** where the data is presented as is and not distorted to provide misinformation or a form of editorializing of the data.

Considerations

In this section, we'll explore the various considerations for creating your own data visualization. These include:

- Cognitive load
- Accessibility
- Colour
- Fonts
- Gestalt's principles:
    - Proximity
    - Similarity
    - Continuity
    - Closure
    - Figure-ground

Cognitive Load

Cognitive scientist James Bruner suggested that we are 22 times more likely to remember information if it's presented as a story. When drafting data visualizations, we want to manage our audience's cognitive load. Cognitive load is the mental effort needed to take in new information. Think of it as humans' ability to process. In designing a data visualization, we want to make the best use of human processing power and direct the audience to what they need to know.

Accessibility

To ensure your infographic is accessible, make sure that you design for accessibility. Below are some tips:

- **Use contrasting colours.** A bar graph with five shades of blue for example may be too difficult to read. If possible, review your data visualizations through colourblind eyes. Try the Colour Contrast Checker which is a tool for checking foreground and background colours. Figure 4.1-4.3[1] below display the same data with different background colours.

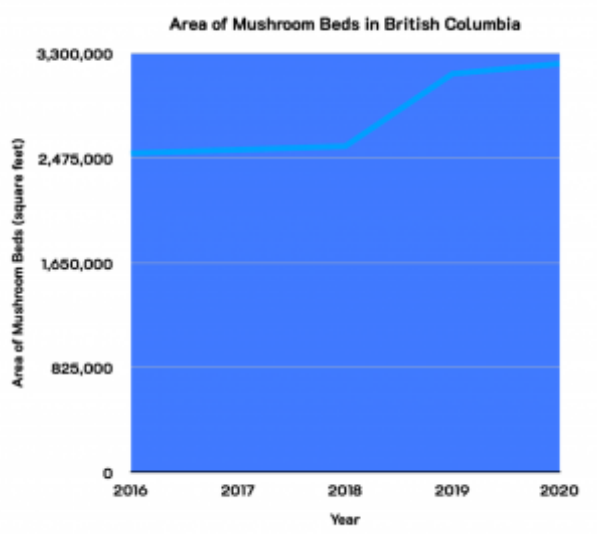

Figure 4.1. Line graph showing how the area (in square feet) of mushroom beds in British Columbia has changed, 2016-2020.

1. Statistics Canada. Table 32-10-0356-01 Area, production and sales of mushrooms. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 8th, 2022. DOI: https://doi.org/10.25318/3210035601-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

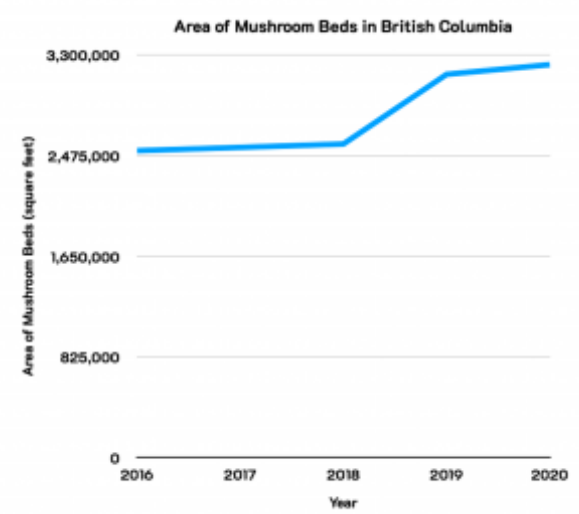

Figure 4.2. Line graph showing how the area (in square feet) of mushroom beds in British Columbia has changed, 2016-2020.

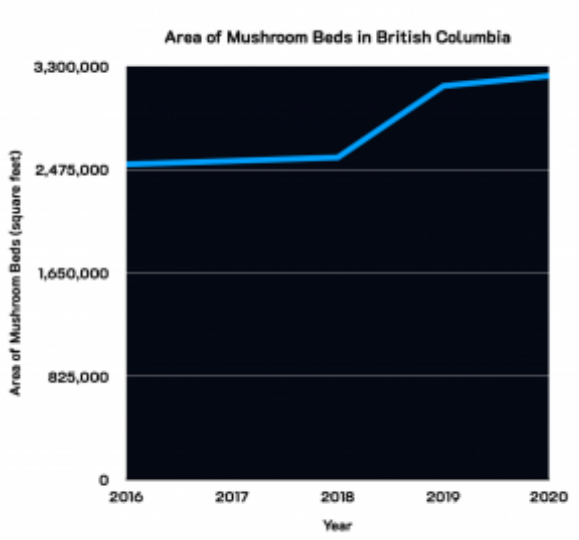

Figure 4.3. Line graph showing how the area (in square feet) of mushroom beds in British Columbia has changed, 2016-2020.

Reflection

Which graph grabs your attention from the ones above?

- **Provide an alternative format or brief description** for images for individuals using screen readers.

Deeper Dive

You can learn more about creating accessible infographics through this resource from the Universal Design Center.

Colour

Visual communication can overcome certain communication barriers such as language. Data visualizations have the ability to communicate more universally. However, visual communication is not universal. Our ways of looking are strongly influenced and possibly determined by the visual

culture we inhabit. We learn certain ways of looking at the world, and what we see is governed by the ways of looking that we have acquired culturally. Colours may also have cultural connotations. When developing visualizations for external audiences, you may want to look at David McCandless' work on colour meaning across cultures.

Fonts

When selecting fonts, here are a few suggestions:

- Strategically use typefaces to communicate varying feelings, moods, style, and emphasis.
- Serif fonts have a traditional feel and style and are often preferred in print.

Figure 4.4. Serif font contains a stroke at the end of a letter's stem.

- San Serif fonts have a modern feel/style and are better for headlines.

Figure 4.5. San serif font does not have a stroke at the end of a letter's stem.

Questions to ask yourself about selecting fonts:

- Is your message evident from the visuals and text?
- Have you used creative, relevant visuals in your designs?
- Is the text appropriate and tailored to your audience?
- How do your creative pieces stand out?

Gestalt's Principles

In the 1900s, the Gestalt School of Psychology defined some basic principals of visual perception that are still widely accepted today and can be applied as a framework in developing data visualizations.

Proximity

Proximity refers to the closeness of visual elements. The separate design entities come together to create a "unified whole" due to their distance/space from one another. The closer the entities appear, the stronger the relationship. See Figure 4.1[1] below for an example of how proximity in a scatterplot defines a relationship.

1. Statistics Canada. Table 32-10-0364-01 Area, production and farm gate value of marketed fruits. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210036401-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

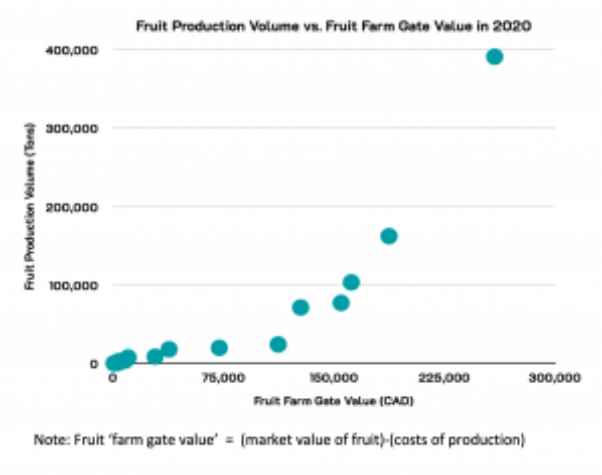

Figure 4.1- The relationship between fruit production volume and fruit farm gate value based on various fruits in 2020.

Similarity

Similarity refers to unity and wholeness (e.g. shapes, text, colours). Elements that look alike are seen as belonging to the same group or creating a pattern to form a singular unit. See Figure 4.2[2] for an example of how repeating colours represent similarity.

2. Statistics Canada. Table 32-10-0364-01 Area, production and farm gate value of marketed fruits. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 9th, 2022. DOI: https://doi.org/10.25318/3210036401-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

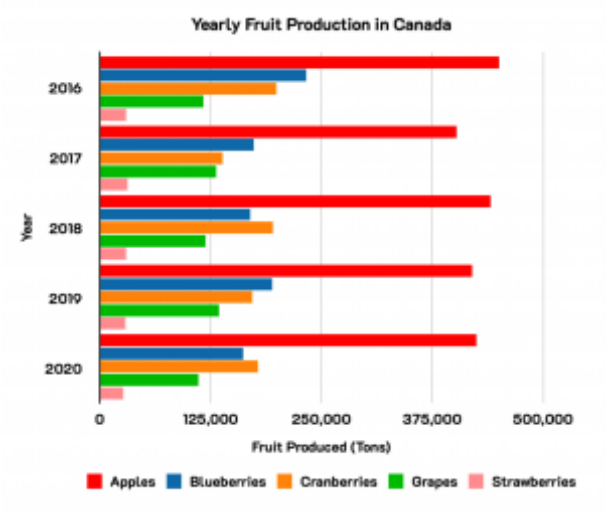

Figure 4.2- Yearly fruit production in Canada based on 5 fruits over the timespan of 2016-2020.

Continuity

Continuity refers to our eyes continuing the design of a path, line, or curve , though it may extend beyond the page. The mind will automatically fill in the gap to “go with the flow”. See Figure 4.3[3] for an example of continuity.

3. Statistics Canada. Table 32-10-0356-01 Area, production and sales of mushrooms. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 8th, 2022. DOI: https://doi.org/10.25318/3210035601-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

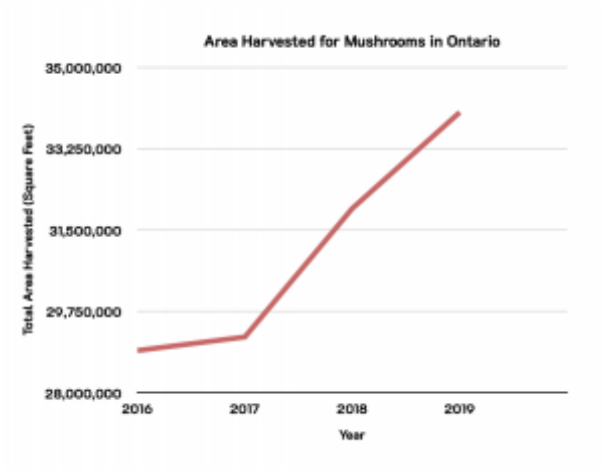

Figure 4.3- Ontario area (in square feet) used to harvest mushrooms over the years.

Closure

Closure refers to our mind completing missing portions of a design. There must be enough parts available for the image to be “filled in”; if the image is too abstract, there are minimal reference points for the mind to complete it. See Figure 4.4[4] for an example of how our mind automatically imagine a line connecting the 2 broken ones.

4. Statistics Canada. Table 18-10-0002-01 Monthly average retail prices for food and other selected products. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved February 2nd, 2022. DOI: https://doi.org/10.25318/1810000201-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

Figure 4.4- Line graph of the cost of round steak from January-July of 2021, with missing data from the months of March and April, misleading the reader to think that the cost was increasing in March and April as well.

Figure-Ground

Figure-Ground refers to the design's focal point (figure) and background (ground) details. Using figure-ground will allow the audience to automatically find the areas to focus upon. Figure 4.4[5] displays good contrast, whereas Figure 4.5 displays poor contrast with the same data.

5. Statistics Canada. Table 32-10-0356-01 Area, production and sales of mushrooms. Data is reproduced and distributed on an "as is" basis with the permission of Statistics Canada. Retrieved January 8th, 2022. DOI: https://doi.org/10.25318/3210035601-eng. Statistics Canada Open Licence: https://www.statcan.gc.ca/en/reference/licence

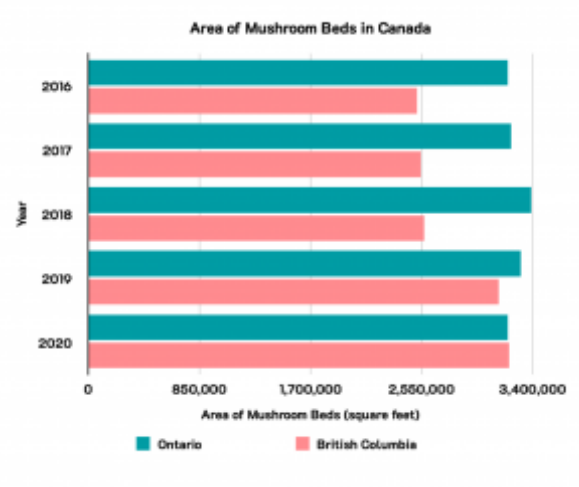

Figure 4.5- Bar graph with good contrast presents how the area (in square feet) of mushroom beds in Ontario and British Columbia has changed over the years.

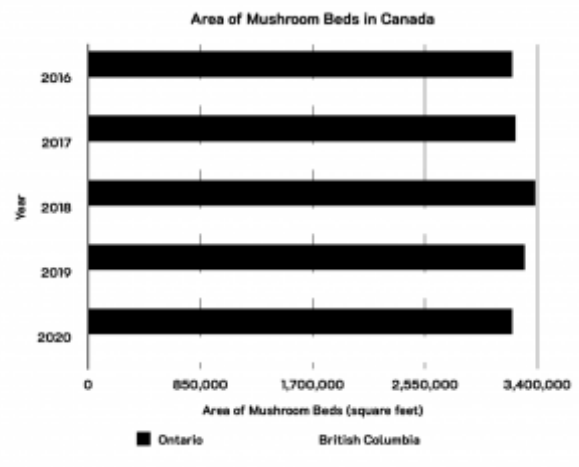

Figure 4.6- Bar graph with poor contrast presents how the area (in square feet) of mushroom beds in Ontario and British Columbia has changed over the years.

Developing Your Data Visualization

Consider your audience first. Be specific, e.g., students in their first-year of university. Broad terms like “stakeholders” may be appropriate but try to narrow down your audience even further such as employees or patients. If you have multiple stakeholders with a wide variety of interests, that may lend itself to multiple data visualizations.

Questions to ask yourself about your audience include:

1. To whom are you communicating?
2. What do you want your audience to know or do?
3. How can you use data to make your point? (Nussbaumer Knaflic, p 20-21)

When building a good data story, consider the following:

- Keep it relevant. The data should be meaningful to your audience.
- Analyze good quality data from reliable sources.
- Develop a clear narrative.
- Use visuals to augment your story in a compelling way.

Data Visualization Checklist

Below is a checklist of considerations when creating your data visualization:

- It has a coherent narrative that tells a story
- Information is from reliable sources
- Charts and diagrams selected are appropriate for data
- Data visualizations have clear titles and labels
- Sources are references as needed
- It is easy to understand at a glance
- Key messages are clear and supported by visuals
- It is clutter free

Summary

In this module, you learned some guidelines for telling an effective data story and considerations for creating your own data visualizations, such as cognitive load, accessibility, colour, fonts and Gestalt's Principles.

Combined with lessons you learned in earlier modules, you should now have the skills you need to avoid misinformation caused by data visualizations.

In the next section, you can review the overall learning objectives of this short course to see how many you achieved, as well as learn more about this project.

Conclusion

Congratulations on completing this short course on critical data literacy!

We hope you have come away from this short course with a greater understanding of data visualizations and design decisions that can create confusion and lead to the spread of misinformation. We hope that you feel more confident analyzing and interpreting data, graphs, charts and infographics.

Learning Outcomes

Below are the learning outcomes for this course that we shared with you in the About this Course section and at the beginning of each module. Please take some time to review the full list and see how many outcomes you achieved.

By completing this short course, you should now be able to:

1. Discuss the different stages of the data journey
2. Explain why data visualizations are effective
3. Differentiate between different types of data visualizations
4. Analyze and break down a data visualization
5. Identify misleading features in a data visualization

6. Evaluate a data visualization using data literacy strategies
7. Assess the credibility of a data visualization's underlying data
8. Explain how to "fact check" a data source
9. Explain some of the key considerations for telling responsible stories with data
10. Outline some effective ways to communicate data

About the Authors

Nora Mulvaney

Nora is an Academic Librarian at Ryerson University who has been working in higher education institutions for over 10 years with a keen interest in data literacy and teaching critical evaluation of information.

Audrey Wubbenhorst

Audrey has been teaching communications for over five years. Prior to that, she held roles in communications in the corporate sector. She has worked with students to develop their skills in writing, presentations and health communications.

Amtoj Kaur

Amtoj is a B.Sc student at Ryerson University studying Mathematics with an interest in data, consulting and student mental health. She will be working in the financial services sector in a business analyst role upon graduation.

Acknowledgements

Project Support

This project was supported by the University Library and Centre for Excellence in Learning and Teaching at Ryerson University. In particular, we would like to thank the following individuals for their support and expertise: Anneke Rummens, Mariam Ahmed, Sally Goldberg Powell, Adam Chaboryk, Ann Ludbrook and Amy Poon.

Book cover design by Emma Seston.

Special thanks to Tanya Pobuda for shepherding this project from (almost) the beginning and for her endless enthusiasm and Pressbooks expertise.

Special thanks to Amtoj Kaur, our brilliant Research Assistant, who created the example data visualizations throughout this resource.

Student Consultation and Feedback

On February 4th, 2022, we held a feedback session with over 50 students from Ryerson University, where we introduced our project, gave a brief overview of the 4 modules and requested feedback in the following areas: content organization, clarity and effectiveness of the graphs and overall look and feel of the course.

We are grateful for the detailed feedback provided by the students, which has improved the quality and cohesiveness of the modules.

Glossary

Chartjunk

Term coined by statistician Edward Tufte used to describe a chart or graph that misleads or distorts data.

Cognitive load

The mental effort needed to take in new information.

Confirmation bias

The tendency to search for, interpret, prefer, and remember information in a way that confirms one's own pre-existing beliefs and positions.

Data literacy

The ability to derive meaningful information from data.

Data visualization

The graphical or pictorial representation of information and data

Data-ink ratio

Term coined by statistician Edward Tufte. It underscores the importance of providing data as is and avoiding any possible distortion or editorializing.

Disinformation

Intentionally biased or misleading information.

Distortion

When objects in the foreground appear larger than objects in the background, which appear smaller.

Infographic

Combination of information and graphics that tells a data story.

Misinformation

False information that is spread, regardless of intent to mislead.

Occlusion

When one 3D graphic partially blocks another one.

Scaffolding

The features that surround and support the content of a data visualization.

Visual encoding

Representing data with a visual element or symbol.