Data State of Mind

PREFACE

Being data literate has taken on a whole new meaning with the influx of information you interact with daily. Data has such power and influence when it is applied effectively to address a public or behavioral health issue. With this easy-to-use and action-oriented manual, you will learn to question and critique existing data, as well as identify, collect, and analyze data. The objective of this training manual is to help you bring public and behavioral health data to life so you can become a responsible data citizen, using data to create positive change in the communities you serve.

You may have already worked with data to some degree when doing community assessment or evaluating programs. What advantages can this manual provide to you?

First, you will notice that the manual focuses on public health and behavioral health data. Often data manuals are written to include examples of data that are not pertinent to the work you do every day. This manual includes real data examples that pertain to your field and are specific to Colorado.

Second, the manual meets you where you are on your data literacy journey. You will see basic content and additional resources that allow you to expand your knowledge with flexibility.

Third, the manual emphasizes how to communicate data and sustain data efforts, not just how to find data or analyze it. In today's world, being able to sustain data efforts and communicate data to the media and community stakeholders is crucial to raise awareness and change policy.

SEOW | Colorado State Epidemiological Outcomes Workgroup

ACKNOWLEDGEMENTS

This manual and its training modules are the result of the cumulative efforts of many skilled contributors. The Colorado Department of Human Services, Office of Behavioral Health are extremely grateful to all of them and the hard work they put into the chapters that constitute the manual and accompanying training modules. Thank you!

The contents of the manual were prepared by the University of Oklahoma Outreach, Southwest Prevention Center for the Colorado State Epidemiological Outcomes Workgroup (SEOW) project managed by the Office of Behavioral Health.

The Southwest Prevention Center thanks everyone at the Colorado Office of Behavioral Health, particularly Sharon Liu and Shaina Riciputi, and members of the Colorado State Epidemiology Outcomes Training Workgroup, Jeffrey Proctor and Laurie Jevons. Their support throughout the entire process was vital in transforming this manual into a living, breathing resource and guide to data literacy. As always, the State of Colorado provides support to the community and prevention and public health practitioners across the state to build capacity around data-based decision-making in the field.

The Southwest Prevention Center and Colorado Office of Behavioral Health would also like to express their gratitude to the following reviewers for their thoughtful and constructive comments: Samantha Barron, Gina Olberding, Bill Pollock, Gregor Rafal, Shaina Riciputi, Allison Rosenthal, Ari Stillman, Lisa Thomason, and Heather Tolle.

The Office of Behavioral Health thanks everyone at the Rocky Mountain Public Health Training Center at the Center for Public Health Practice – Colorado School of Public Health in taking the contents of the manual and developing the training modules. It was an incredible effort to translate the manual and design and visualize it into several training modules.

Finally, a special thanks goes out to The Evaluation Center, School of Education & Human Development, University of Colorado Denver for managing the SEOW project and supporting multiple agencies.

HOW TO USE THIS MANUAL

Who should use this manual?

The ideal audience for this manual is community members working in public or behavioral health who are just beginning to use data in their work or participate in workgroups that utilize data for decision-making.

What is needed to use the manual?

This manual needs your time and motivation! Any software referenced is typically available with the Microsoft Office package. You will need a computer for some of the activities if you decide to practice the exercises in the companion guide. The manual does require some familiarity with Microsoft Excel. If you are new to Excel, consider taking some online tutorials available for free on the Microsoft website. It might also be helpful if you work through the manual with a colleague so you can discuss the content together for further learning.

How should I use the manual?

It is important to go through the manual in the order presented, as the concepts build on each other. The manual also includes multiple opportunities to practice concepts in the companion document with activities that are more challenging. The manual is designed to be flexible for learning. For maximum learning, you can work through each section at your own pace to allow time for exploration of additional resources.

What does the manual do for me?

The manual provides a safe space for you to practice data literacy skills in a private setting at your own pace. It aims to increase your confidence in working with data and gaining new skills to help you make data-informed decisions in your work. You can take the skills you learn to your community or organization. When communities increase capacity for data literacy, it places communities and organizations in a better position for future grant opportunities and improves your ability to monitor programs and outcomes.

1

Getting The Concepts Down

WHAT IS DATA & DATA LITERACY?

*"We don't just need citizen data scientists—we need more data citizens."**

In public and behavioral health, the word data is used a lot, but what does it mean? Data is facts that are used as a basis for reasoning, discussion, or calculation. Raw data is the base level of information gathered before applying any analysis or interpretation. With the right skills, raw data can be analyzed and interpreted to help make the information meaningful. The analysis and interpretation then become information that can be used to make better-informed decisions about public or behavioral health issues that are useful to the communities you serve.

DATA LITERACY REFERS TO THE ABILITY TO COLLECT, MANAGE, EVALUATE, AND CRITICALLY APPLY DATA.[1]

Data literacy is being able to take data in the form of numbers and statistics, turn it into information that helps address community problems, and communicate it in a useful and compelling way to others.

*Brent Dykes

KEY TERMS TO GET YOU STARTED:

Data: Factual information used as a basis for reasoning, discussion, or calculation.

Raw Data: Data files that hold information that has not been summarized.

Data Literacy: The ability to collect, manage, evaluate, and critically apply data.

Data Citizen: A person who relies on data to make decisions.

Conceptual Framework: A theoretical structure of assumptions, principles, and rules that holds together the ideas comprising a broad concept.

KEY POINT:

Data is the "raw" stuff. When you start analyzing and summarizing, you turn data into usable information. In order to make it usable, you need to acquire knowledge, skills, and abilities that help you work with and interpret data in meaningful ways.

Becoming a Data Citizen

THERE ARE MANY BENEFITS TO HAVING A STRONG DATA CITIZEN COMMUNITY.

A strong data citizen community:

- Makes effective decisions with data;
- Helps communities succeed, measuring progress and improvements in health outcomes; and
- Provides opportunities for community engagement with data to mobilize for change.

DATA CITIZENS—YES, THAT'S YOU!

Public and behavioral health practitioners have more data available than ever before, but finding the right data and doing the right thing with that data has become a critical skill for practitioners. The term "data citizen" describes a person who relies on data to make decisions.

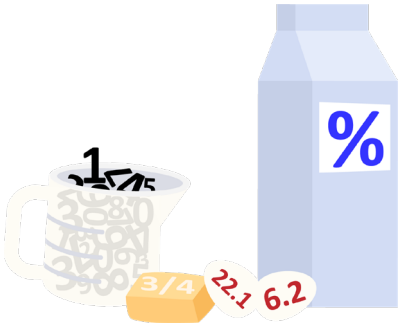
What if I'm Scared of Data?

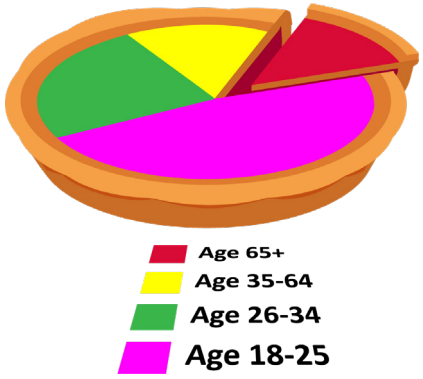
People fear data for different reasons. Some people fear data related to confidentiality and data security. You may have personal experiences that were not positive. As you learn more about how to keep your data safe and protect confidentiality, your fears around this will decrease. Some people also have memories of math class being difficult in school. As you become more data literate, your confidence will increase, and data will become an asset. This manual takes the approach that data and statistics are tools and skills you need to do good work for others. It models ways to make data understandable and fun so that when you are working with others in your community, they will have good experiences with data and statistics too.

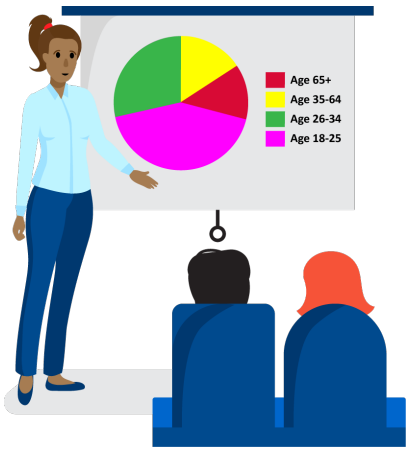
KEY POINT:

Strong data citizen communities can better achieve desirable public and behavioral health outcomes.

The Path to Data Citizenship







THIS MANUAL IS A ROAD MAP FOR DATA LITERACY.

ALONG THE ROAD YOU WILL LEARN:

- Defining data literacy
- Developing a foundation
- Collecting data
- Managing data
- Evaluating data
- Applying data.

Throughout this manual, you will develop skills in each of these areas in order to become data literate.

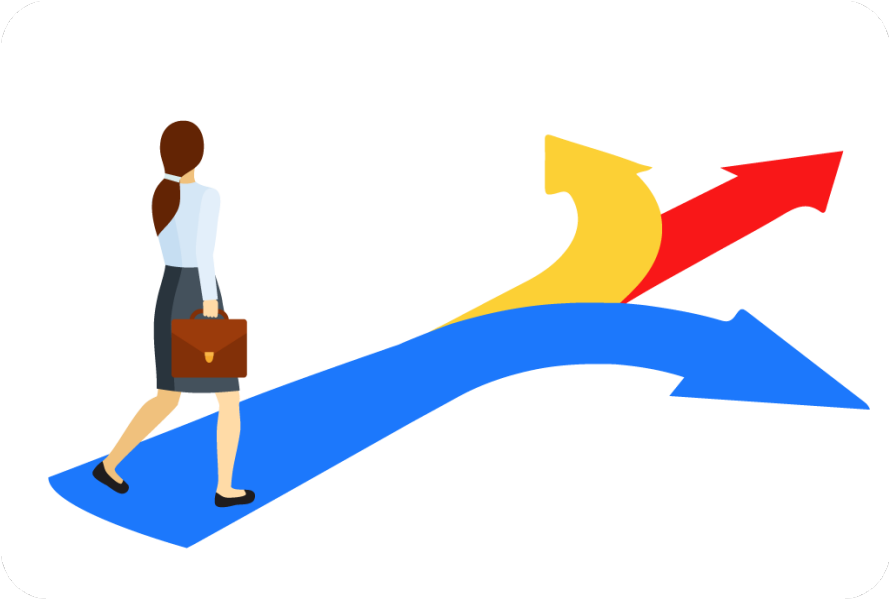
KEY POINT:

Data literacy is achieved when you can collect, manage, evaluate, and apply data in a critical manner.

FURTHER LEARNING

Qlik Continuous Classroom: A Culture of Data Literacy
https://qcc.qlik.com/course/view.php?id=723

Qlik Data Literacy Assessment: What's Your Data Personality?
http://www.dataliteracy.info/quiz



REFERENCES

1. Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report.
Retrieved from http://dataliteracy.ca/wp-content/uploads/2016/04/Strategies-and-Best-Practices-for-Data-Literacy-Education.pdf.

2

The Land of Literature

THE LAND OF LITERATURE

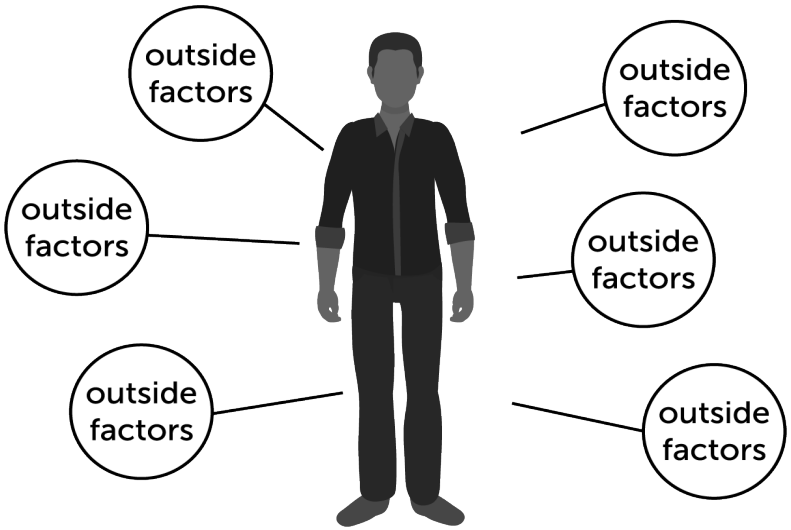
*"Most of the world will make decisions by either guessing or using their gut. They will be either lucky or wrong."**

This chapter will prepare you to explore the wide world of research and literature so you can make better decisions for your programs and organizations. This is called data-based decision-making. You will learn the types of data, what to look for in a good research article, key tips for examining research, and some basic information about analyzing data. You will get comfortable with reading research articles and evaluating what is high-quality research and what is not.

While conducting a community assessment, public health and behavioral health practitioners typically research a problem or may conduct a formal literature review. Before beginning the process of a community assessment, you need to understand the different types of studies and key things to examine, such as credibility of the research, and its relevancy to the issue. This requires a basic understanding of terms and even parts of a typical research article. You will need to think critically and determine if the conclusions drawn are accurate, or if they may be explained by outside factors. You will work through some critical steps to examine research to build your data citizen skills.

Study Type ▼

- case control studies
- case series
- cohort studies
- meta-analysis
- randomized control trial
- systematic review



*Suhail Doshi

THE LAND OF LITERAURE

As a data citizen, you can take the following steps to evaluate a research article:

1

Does the research collect numbers or words?

When you see the term quantitative data[2], researchers are referring to data that involves information that can be measured. Sometimes you will see quantitative data collected as a score on a survey, assessment, or screener tool. In contrast, qualitative data[3] does not have a numerical value. You may use qualitative data to explore an idea or when you want to learn more from program participants about a problem that is difficult to capture using quantitative data. Qualitative data is recorded as text and words and can be collected during observations, interviews, or focus groups.

2

Did the authors collect the data themselves or use existing data?

Data is classified as primary or secondary. You are using primary data when you analyze data for its original intention. With secondary data, you are using the data for a different purpose than its original intention.

3

What is the Evidence?

Often in public and behavioral health, the highest level of evidence may be a Level III since you may not have the ability to do a randomized control trial. You can also access evidence-based registries for assistance with selecting effective interventions. Registries have subject-matter experts review the literature for you and summarize the evidence of effectiveness based on a ranking or scale. For more information on selecting evidence-based strategies, refer to the Further Learning section at the end of this chapter.

4

Using the Abstract.

The abstract summarizes the purpose, design, results, and conclusions of the study. You can use the abstract to quickly review the overall content of the paper. The abstract will usually indicate if the article is qualitative or quantitative. If you do not see the method listed in the abstract, look for the method or methodology section of the article.

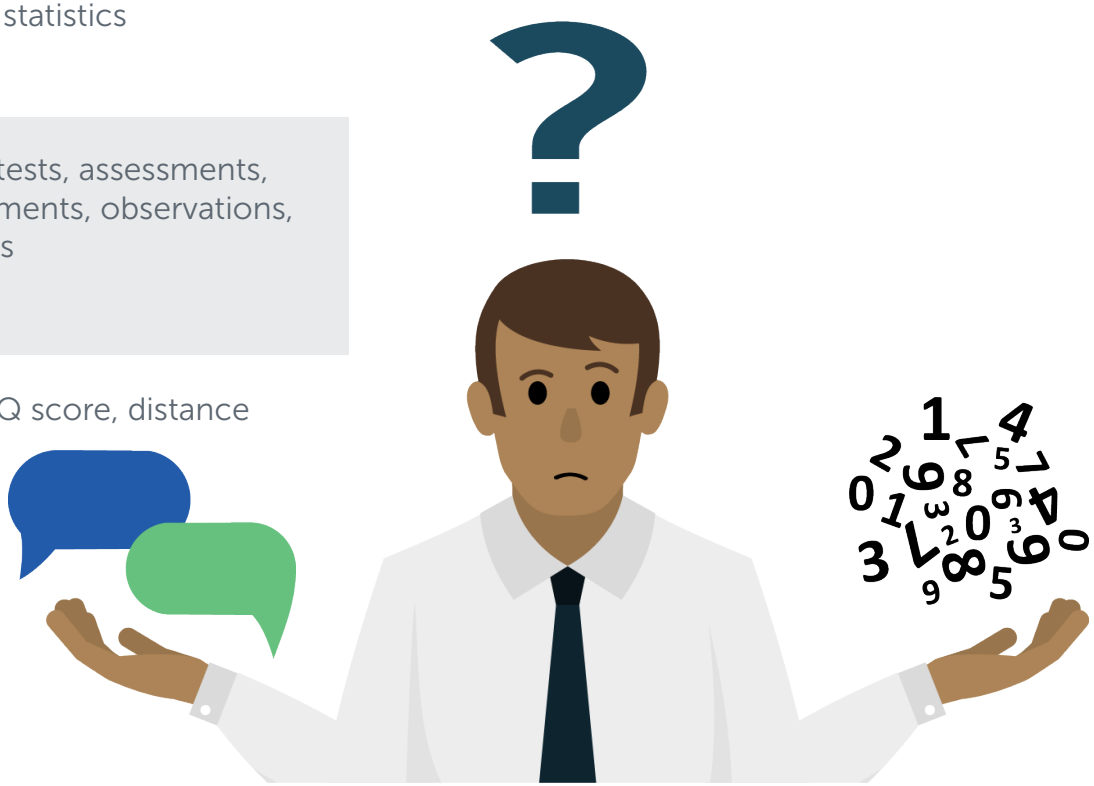
5

When the Abstract Isn't Enough—Looking at the Method, Results, and Discussion Sections of an Article.

The Method and Results sections are next in line to help you. Beyond the abstract, the method often has clues as to what type of data the researchers collected and how they collected it. You then have what you need to determine the level of evidence in most cases. If not, you can also examine the next parts of a research article, the Results and Discussion sections.

STEP 1. DOES THE RESEARCH COLLECT NUMBERS OR WORDS?

	Qualitative Data	Quantitative Data
What is it?	Data that consists of descriptive information that cannot be expressed as a number.	Data that can be measured and expressed numerically
What are the types of data?	Text-based, images, behaviors, objects	Number, statistics
How is it collected?	Interviews, focus groups, written documents, observations, case studies, records, open-ended survey questions	Surveys, tests, assessments, measurements, observations, interviews
What are some examples?	Favorite food, gender, quotes, race/ethnicity	Weight, IQ score, distance



KEY POINT:

The main difference between primary and secondary data is that you collect primary data yourself, and secondary data already exists.

STEP 2. DID THE AUTHORS COLLECT DATA THEMSELVES OR USE EXISTING DATA?

Data is classified as primary or secondary. You are using primary data when you analyze data for its original intention. So, for example, if you collect survey data on youth risk and protective factors and plan to examine the relationship to substance use, you then analyze the data for that purpose. With secondary data, you are using the data for a different purpose than its original intention. Using the example above, another researcher might obtain your data from the risk and protective factor survey but analyze the data to predict academic success in college. Then it would be considered secondary data. Below you will see some advantages and disadvantages of using primary and secondary data.[4]

Primary Data		Secondary Data	
Advantages	**Disadvantages**	**Advantages**	**Disadvantages**
Can collect data that is specific to the study	Expensive to obtain	Have access to large data sets	May be less reliable
Have more control over data collection	Constraints due to time, place and participants	Save time and money	May not fully answer the research question
May be considered more trustworthy		May not have restrictions due to time, place and participants	Often requires advanced statistical technique to analyze and can be difficult to determine how data was collected

STEP 3: WHAT IS THE EVIDENCE?

Types of Studies

Social scientists often conduct what is called human subjects' research. When you obtain, use, study, or analyze data from a living person, you are conducting human subjects' research. Because public and behavioral health practitioners deal with humans, there are many outside factors that you cannot control. If you are not able to control certain conditions in a study, it may affect your results and will limit your ability to conclude that a variable, such as receiving a substance use intervention, resulted in reduced substance use rates. Outside factors affecting control might be previous substance use, family history, etc. Below is a graphic that classifies different methods used in the research, based on the strength of the evidence, from a Level I with multiple randomized control trials, to a Level VII that focuses on an expert opinion. Often in public and behavioral health, the highest level of evidence may be a Level III since you may not have the ability to do a randomized control trial.

ULTIMATELY YOU WANT TO STRIVE FOR THE HIGHEST LEVEL OF EVIDENCE POSSIBLE.

Level of Evidence[5]	Description
Level I	Quantitative evidence from multiple randomized control trials (RCT)
Level II	Quantitative evidence from one RCT
Level III	Quantitative evidence obtained from research with controls (not randomized)
Level IV	Quantitative evidence from analyzing groups that already exist (comparing or observing over time)
Level V	Qualitative evidence from more than one study
Level VI	Qualitative evidence from one study
Level VII	Expert opinion

KEY POINT:

Research articles can include quantitative, qualitative, or mixed methods research.

STEP 4. USING THE ABSTRACT

Reviewing research to gather more information about an issue sometimes seems like an insurmountable task. First, try looking at the article's Abstract. The abstract summarizes the purpose, design, results, and conclusions of the study. You can use the abstract to quickly review the overall content of the paper. The abstract will usually indicate if the article is qualitative or quantitative. If you do not see the method listed in the abstract, look for the method or methodology section of the article. This section will provide a detailed description of the methods used in the research study.[6] Some tips to identify each type of study are mentioned below.

Quantitative Tip: When an article uses quantitative methods, you will see some of the following key words or clues: statistical significance, hypothesis, survey, control, larger number of participants.

Qualitative Tip: When an article uses qualitative methods, you will see some of the following key words or clues: explore, interviews, focus groups, grounded theory, smaller number of participants.

Mixed Methods Tip: Often in the social sciences and public health world, researchers employ more than one method for a well-rounded study. Social scientists, public health practitioners, and educators typically view research not as quantitative or qualitative, but on a continuum from one to the other. Some studies lean more toward quantitative, some more toward qualitative. There are compelling arguments for the view that mixed methods (i.e., the integration of both quantitative and qualitative data and analytical techniques) are essential for solving complex problems that involve policy, vulnerable groups, and research methods that are tailored and culturally appropriate.[7] You may see these tips or clues: data triangulation, mixed method, mixed methodology, multimethod, integrating methods, qualitative, and quantitative.

KEY POINT:

The abstract often gives a lot of information about the method and results of research. If you need more information to determine the quality of the method and results, you can examine the method and results sections of the article.

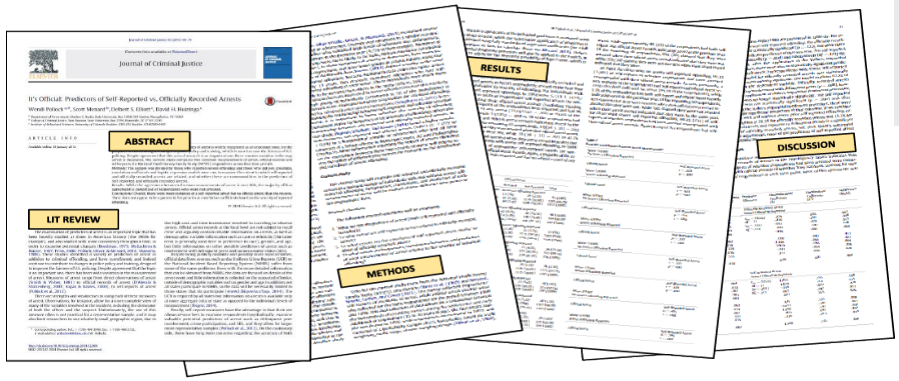

Journal of Criminal Justice

It's Official: Predictors of Self-Reported vs. Officially Recorded Arrests

ABSTRACT

STEP 4. WHEN THE ABSTRACT ISN'T ENOUGH—LOOKING AT THE METHOD, RESULTS, AND DISCUSSION SECTIONS OF AN ARTICLE

As your skills as a data citizen grow, you may need to go beyond the abstract of an article to find the information you need. The **Method** and **Results** sections are next in line to help you. Beyond the abstract, the method often has clues as to what type of data the researchers collected and how they collected it. You then have what you need to determine the level of evidence in most cases. If not, you can also examine the next parts of a research article, the Results and **Discussion** sections. To the right you will see the typical order of a research article.



Parts of a Research article

Abstract
The abstract is a short, one-paragraph summary of the major parts of the article

Literature Review
The introduction establishes the context of the research paper by reviewing existing, relevant literature. Researchers also use this section to state their hypothesis and/or research question(s).

Method
The method details how the research was carried out, including the research design, protocol or procedure, measurement instruments, and any other materials used

Results
The results include key results without interpretation. This section may contain tables and figures to help illustrate the results

Discussion
The discussion provides an interpretation of results as they relate to the existing research and the hypothesis and/or research

WHAT DO THESE TERMS MEAN?

When you are reviewing methods, results, and discussion sections, there are some common terms you will encounter.

SAMPLING

The sample simply is those who participate in the study; the population is all the possible individuals who might be selected for the study. The goal of sampling is to have participants closely mirror the population you are studying. This is called being representative of the population.[8]

THERE ARE TWO MAIN APPROACHES TO SAMPLING—PROBABILITY SAMPLING AND NONPROBABILITY SAMPLING.

Probability sampling is when all members of a population have the same opportunity and chance of being included in the study.

Nonprobability sampling occurs when participants are selected based on availability. This is often called convenience sampling. You can utilize convenience sampling when it is difficult to obtain participants, or you do not have access to all members of a population. You may need to use a convenience sample when the budget of a youth community assessment may not be adequate for random sampling, or you may not be able to determine the entire population when researching a substance use problem. In this case, you could use an existing group of students or youth summer camp participants to collect survey data.

WHAT DO THESE TERMS MEAN?

DESCRIPTIVE STATISTICS

You can use descriptive statistics when you want to summarize the data and find patterns. The box to the right highlights descriptive statistics you are most likely to see in your work.

You can use descriptive statistics when the research is limited to the sample and does not need to be generalized to a larger population.

For example, if you are comparing the percentage of children vaccinated in two different rural communities, then you are reviewing descriptive statistics. Descriptive statistics focuses on comparisons of values, such as percentages, on a single variable.

KEY POINT:

Some common descriptive statistics in public health and prevention include the mean, median, mode, range, percentages, and frequencies.

COMMON DESCRIPTIVE STATISTICS

Mean
Numerical average of a set of values. Calculation = Sum divided by the number of responses

Median
Midpoint of a set of numerical values. Calculation = Order values from least to greatest. Locate the middle number. Tip: if it is an even number of responses you average the two middle numbers to get the median

Mode
Most common value among a set of values. Calculation= Value that occurs most frequently in a dataset

Percentage
Used to express how a value or group of respondents within the data relates to a larger value or group of respondents. Calculation = The number of responses for a certain value (e.g. yes or no) divided by the total number of responses, multiplied by 100

Range
The space between the highest and lowest values. Calculation= Highest value minus the lowest value in your dataset

WHAT DO THESE TERMS MEAN?

INFERENTIAL STATISTICS

Often, researchers collect data on a sample of their population. Next, they generalize the results to the entire population or target group.

Inferential statistics generalize results and make predictions about a larger population.

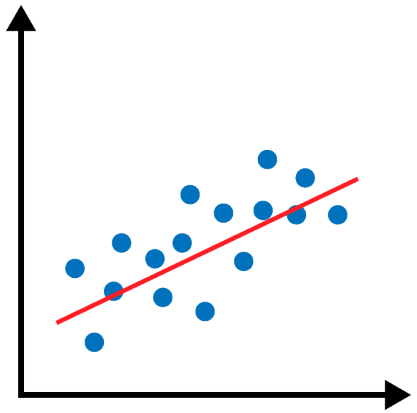
You use inferential statistics when your analysis needs to go beyond percentages or averages of a single variable, and you want to understand the relationships between variables.

KEY POINT:

When examining articles, you can look at the level of evidence to determine how generalizable the results might be to the population.

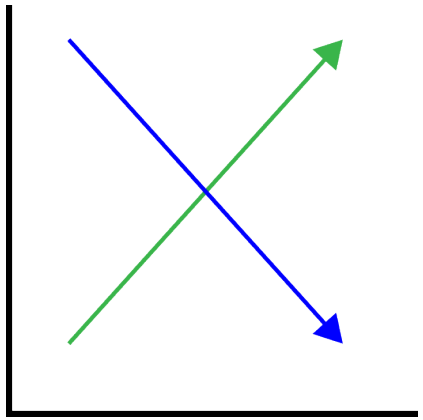
WHAT DO THESE TERMS MEAN?

COMMON INFERENTIAL STATISTICS



Correlation: This describes the relationship between two variables. When two variables are correlated, it means as one variable goes up, the values on the other variable will increase (or decrease) in a similar pattern. This similar pattern between the two variables could indicate a connection between the two variables. However, two variables being correlated does not indicate that one causes the other variable to change. There could be another unknown factor that influences both variables. For example, taller people tend to have a higher weight. Height and weight are correlated with each other. However, this doesn't necessarily mean that one variable causes the other (e.g., gaining weight doesn't cause people to grow taller).

Regression: This shows the relationship between two or more variables. For example, regression can help you estimate someone's weight based on their height. Regression allows for the prediction of one variable based on other variables.



T-test: This makes a comparison between two groups, for example, an experimental group and a control group. You can use the t-test to compare two group means and test for statistical significance. For example, you may use a t-test to determine if you found a difference in appointment no-shows between a group who received a reminder call and those who did not.

Analysis of variance (ANOVA): You can use analysis of variance to compare differences of means among more than two groups. For example, with analysis of variance, you can measure the difference in weight loss between a group who received a nutrition program, a group who received an exercise program, and a group who received a combination of the nutrition and exercise programs.

WHAT DO THESE TERMS MEAN?

QUALITATIVE DATA ANALYSIS METHODS

Several methods are available to analyze qualitative data. You can refer to the more common methods you may see in articles below:

Content analysis: Content analysis is used more frequently than other qualitative analysis methods. You can analyze information in text, media, or even physical items. Most often, you conduct content analysis to examine interview responses.

Narrative analysis: With this method, you analyze content from various sources, such as interviews, observations, or surveys. You focus on using the stories and experiences shared by people to answer your research question.

Discourse analysis: Like narrative analysis, you can use discourse analysis to analyze interactions with people. You are analyzing the social context in which the communication between the researcher and the respondent occurred. Discourse analysis also looks at the respondent's day-to-day environment and uses that information during analysis.

Grounded theory: With this method, you analyze qualitative data to explain why something happened. You study a variety of similar situations in different settings and use the data to come up with explanations. As the study unfolds and you collect more data, you alter the explanations or create new ones until you arrive at an explanation that fits all cases.

APPENDIX

KEY TERMS

Abstract: a short, one-paragraph summary of a research article to establish the context of the research paper by reviewing existing, relevant literature.

Analysis of Variance (ANOVA): a statistical method used to compare differences of means among more than two groups.

Correlation: a statistical method used to compare the relationship between two variables.

Data-Based Decision Making: making data-informed choices related to programs or strategies.

Descriptive Statistics: using statistics to summarize data and find patterns.

Inferential Statistics: collecting data on a sample of a population, then generalizing results and making predictions about a larger population.

Mixed Methods: the integration of quantitative and qualitative data.

Nonprobability Sampling: when research participants are selected based on availability. Also, called Convenience Sampling.

Primary Data: data that is analyzed for its original intention.

Probability Sampling: when all members of a population have the same opportunity and chance of being included in the study.

Qualitative Data: data that includes text or words that can be categorized into themes.

Quantitative Data: data that involves numbers and can be measured.

Regression: a statistical method used to compare the relationship between two or more variables.

Sample: those who participate in the study.

Secondary Data: data that is used for a purpose other than its original intention.

T-test: a statistical method used to make a comparison between two groups.

APPENDIX

FURTHER LEARNING

Evidence-Based Practice for Health Professionals: Levels of Evidence
https://libguides.nvcc.edu/c.php?g=361218&p=2439383

Selecting Best-fit Programs and Practices: Guidance for Substance Misuse Prevention Practitioners
https://www.samhsa.gov/sites/default/files/ebp_prevention_guidance_document_241.pdf

Introduction to Statistics & Data Analysis in Public Health
https://www.coursera.org/learn/introduction-statistics-data-analysis-public-health

Sample Size Calculator
https://stattrek.com/survey-sampling/sample-size-calculator.aspx?Tutorial=samp

REFERENCES

2. Creswell, John & Guetterman, Timothy. (2018). Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, 6th Edition.
3. Ibid.
4. Koziol, N., & Arthur, A. (n.d.) An Introduction to Secondary Data Analysis. Retrieved from http://r2ed.unl.edu/presentations/2011/RMS/120911_Koziol/120911_Koziol.pdf
5. Winona State University (2019). Evidence-based practice toolkit. Retrieved from https://libguides.winona.edu/c.php?g=11614&p=61584.
6. Eastern Michigan University (2017). Quantitative and qualitative methods. Retrieved from https://guides.emich.edu/c.php?g=188042&p=1242147
7. Gomez, A. (2014). New developments in mixed methods with vulnerable groups. Journal of Mixed Methods Research, 8(3), 317-320.
8. McGuigan, F. J. (1997). Experimental psychology: Methods of research. Seventh edition. Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

3

Becoming a Data Skeptic

DATA SKEPTIC

*"The job of the data scientist is to ask the right questions."**

When you hear the word "skeptic," you might think of that word in a negative way. However, being a skeptic when it comes to data just means that you question or have doubt about something, and asking questions is never a bad thing! Think of it as being a detective. In this chapter, you will learn how to ask the right questions. You will develop skills as a good data citizen so you can think critically about how research or data results are reported and what conclusions the authors are making.

Too often society accepts data at face value. After reading this chapter, you will be able to examine the original source of the statistics being summarized to determine if the findings are being interpreted correctly, as well as how the data was collected and analyzed. You will develop skills to examine data critically to benefit your organization and community.

In this chapter, you will explore ways you can become a competent data skeptic in your community in Colorado.

AS YOU REVIEW DATA OR RESEARCH YOU WILL ANSWER FOUR QUESTIONS TO EVALUATE **METHOD, CREDIBILITY, BIAS, AND TRUTH** OF DATA AND PUBLICATIONS.

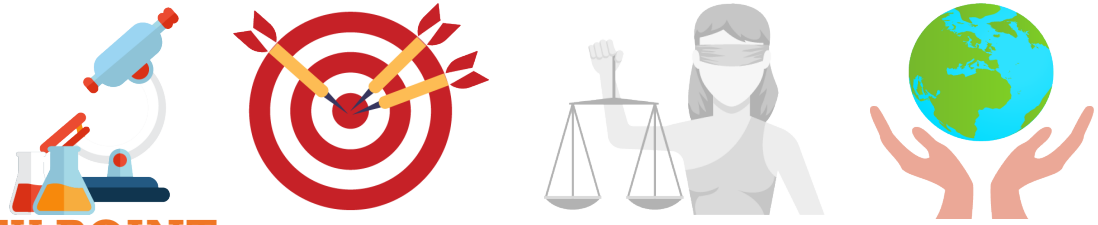
*Hilary Mason

Questions to Ask

1 COULD THE METHOD OR WAY IN WHICH THE DATA WAS COLLECTED INFLUENCE THE RESULTS?

How data is collected (in-person or online, survey or interview), as well as the procedure, can have an impact on the accuracy of the data in an article or report.[9]

If data is not gathered in a scientific, consistent and ethical way, negative consequences could include: participants may not be able to answer the questions on a survey accurately, it may be difficult to replicate the study, resources may be wasted on ineffective policies or interventions, other researchers may waste time doing further studies, and there could be harm to the public or participants involved in the study.



KEY POINT:

When reviewing articles and data, practice using the four questions (question the Method, Credibility, Bias, and Truth). Thoroughly evaluate your potential data sources and articles.

QUESTION METHOD

- ✓ Is the data qualitative or quantitative?
- ✓ What methods did the authors use to collect the data (interview, survey, focus group, existing health data)?
- ✓ What procedures did the authors follow during data collection to ensure confidentiality and security of data?
- ✓ Did the authors follow consistent procedures if data was collected at multiple sites?
- ✓ If data was collected by others, were protocols written down? Did the researchers provide data collection training?
- ✓ Was the data collected in-person, over the phone, online, through the mail?
- ✓ Did participants drop out of the study? Why?

2 HOW CREDIBLE OR RELIABLE IS THE SOURCE OF THE DATA OR RESEARCH?

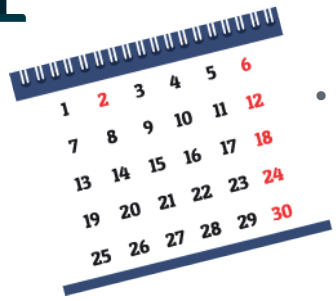
In public and behavioral health, often the author is a government organization. Many national data sources from the government have extensive reviews before being published (e.g., CDC, SAMHSA, Colorado state government sources).

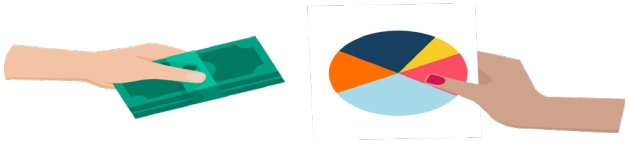
If you are not using government data sources, use online scholarly databases such as InfoTrac, PsychINFO, Google Scholar, PubMed, LexisNexis, and EBSCO, which provide access to the latest research in hundreds of areas. Local librarians are also a great resource if you would like one-on-one assistance finding credible research.

Before you decide to use data for a report, publication, or assessment, make sure you examine the credibility of the article by asking some important questions about the author, reference, and publication history.

KEY POINT:

For credible data sources, utilize local, state, and national government data sources, as well as database sources reviewed by experts (also called peer review) including InfoTrac, PubMed, PsychINFO, LexisNexis, Google Scholar, and EBSCO.





QUESTION CREDIBILITY

Author:

- Who is the author?
- What else have they written?
- What communities and contexts does the author have expertise?

References and Reviews:

- What sources does the study cite?
- Are there any published reviews, responses, or rebuttals to the study?

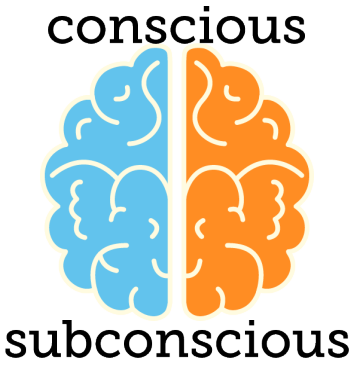
Publishing:

- Were there barriers to publication?
- When was the article first published?
- What edition of the source are you using?
- Has anything changed in the field of study since publication?

3 IS THERE POTENTIAL BIAS FROM EITHER THE DATA PRODUCER, OR YOU AS THE CONSUMER?

QUESTION BIAS

- ✓ Why was this source created?
- ✓ Does it have an economic value for the author, publisher, or consumer?
- ✓ Is it an educational resource? Persuasive?
- ✓ What (research) questions does it attempt to answer?
- ✓ Does it strive to be objective?
- ✓ Does it fill any other personal, professional, or societal needs?
- ✓ Who is the intended audience?



Your next step is to determine potential bias. You will need to ask about the purpose of the report or article. Bias can be conscious or subconscious.

As a data citizen, utilize the questions to the right to assess bias.[10] Look for transparency when you examine articles for bias. Bias is not always a bad thing, but it is something a researcher should state (such as a financial conflict of interest) so that it can be acknowledged by the reader.

Examine articles to see if the authors clearly state any bias and what efforts they made to address the bias. To give you an example, imagine your coalition is planning to conduct an evaluation of a local alcohol policy change. You decide as a coalition that with limited funding available for the project, you need to utilize your own staff to conduct the evaluation. In your report, you would need to acknowledge that your staff was involved, and what procedures you put in place to limit the bias as much as possible.

4 IS THE DATA BEING MANIPULATED?

Is the data being manipulated in a way—intentionally or inadvertently—that misrepresents its true meaning?

Your last task is to question the "truth" of the data or study. Make sure to ask yourself the questions to the right.

For example, if the study method was a focus group in a particular rural community in Colorado, you would not assume that the same or similar results occur in a resort community in Colorado. Or, if researchers conducted a quantitative study and found a high correlation between teen cigarette smoking rates and vaping, the authors should not state that smoking causes vaping, only that there is a significant relationship between the two activities.

KEY POINT:

Ask questions about the data. Remember that the job of the data citizen is to ask the right questions. Reach out to others who can help you find credible sources of information (epidemiologists, data analysts, and librarians).

QUESTION TRUTH

- Does the author discuss the limitations of the study?
- How large was the sample?
- Does the author make assumptions that could affect the results?
- Does the author discuss both statistically significant and non-significant results?
- Are the results generalizable to a larger population?
- How do the results compare to other studies on the same topic?
- Are the methods clear and reproducible?
- Was the study peer-reviewed?
- Is correlation being confused with causation?

APPENDIX

KEY TERMS

Bias: potential preconceptions that may affect the research findings.

Credibility: an assessment of whether research findings are reliable.

Method: the way in which data is gathered.

FURTHER LEARNING

Step-by-Step Guide & Research Rescue: Evaluating Credibility
https://guides.lib.byu.edu/c.php?g=216340&p=1428399

Assessing Research Quality
https://www.researchconnections.org/content/childcare/understand/research-quality.html

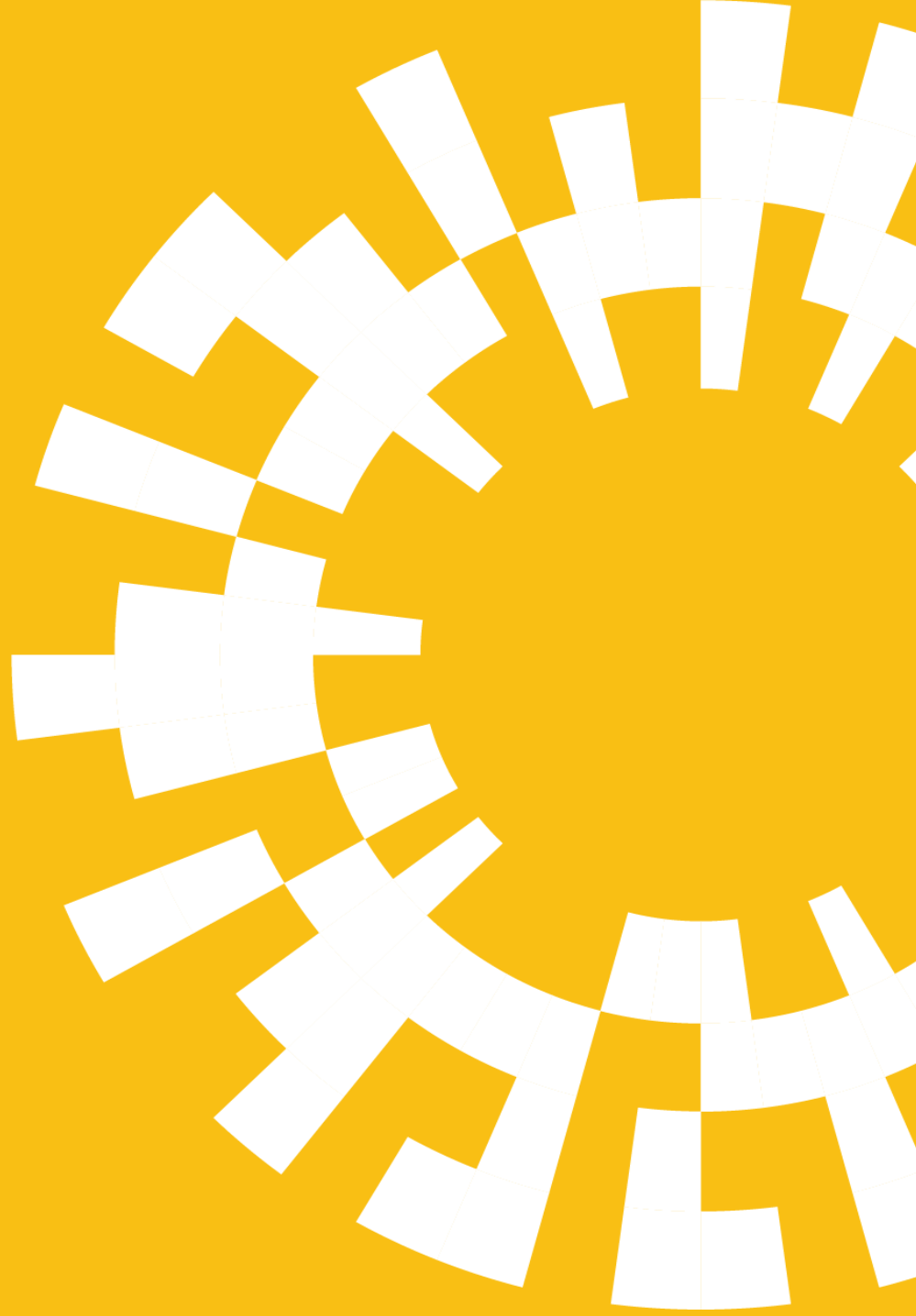
9 Types of Research Bias and How to Avoid Them
https://www.quirks.com/articles/9-types-of-research-bias-and-how-to-avoid-them

REFERENCES

9. Ponto J. (2015). Understanding and Evaluating Survey Research. Journal of the Advanced Practitioner in Oncology, 6(2), 168–171.
10. Protogerou, C., & Hagger, M. S. (2019). A Case For a Study Quality Appraisal in Survey Studies in Psychology. Frontiers in Psychology, 9, 2788.

4

Data, Data, Everywhere

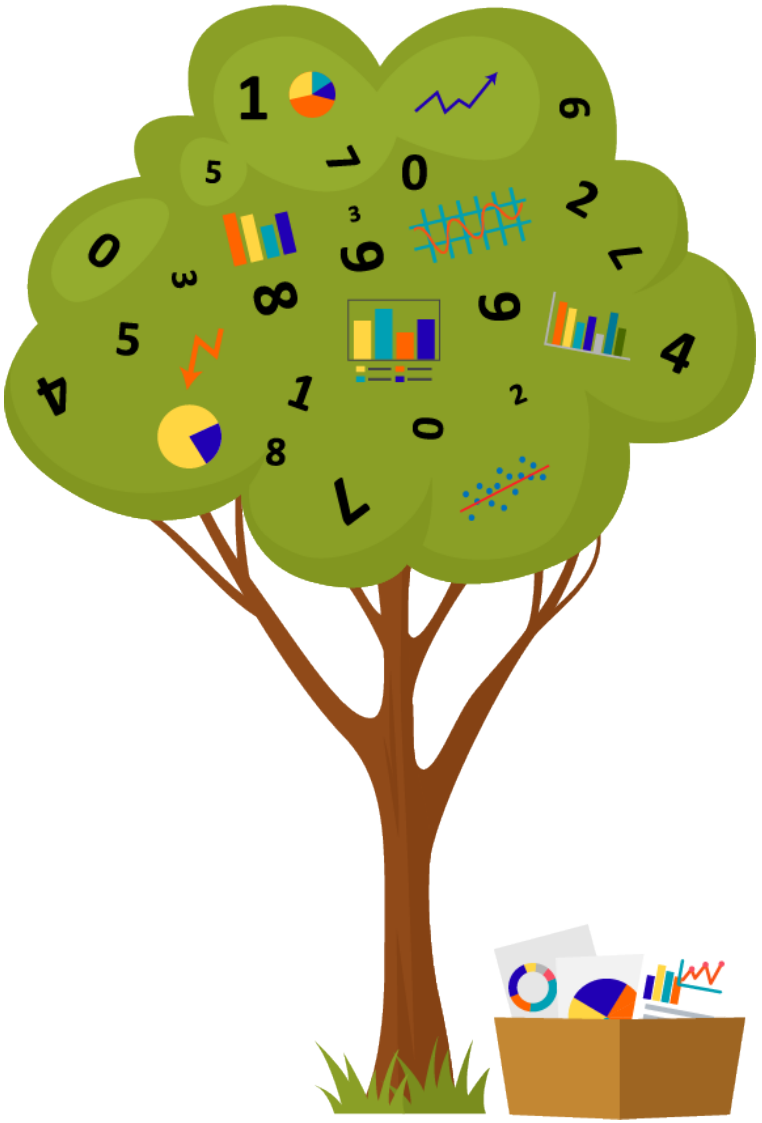


DATA EVERYWHERE

*"Some of the best theorizing comes after collecting data because then you become aware of another reality."**

Now that you know how to be a skeptic and ask the right questions with other people's data, you will learn how to collect data when conducting your own research, evaluating a community program or policy, or working to gather data from multiple institutions to address a public health issue. If you are working in a community on a grant, this may be your community assessment phase. You are gathering information on the data sources, location, contacts, gaps, and needs in the community. If you are working with an epidemiological workgroup, you might be going through a similar process, but may be working with multiple government agencies at the state or regional level to gather existing data and share it with each other to maximize its potential.

Whatever your circumstances might be, there are some simple tips and steps to help you through the process of working with existing data, requesting data, and collecting primary data. You will emerge from this data collection phase with a deeper awareness of another reality—how collecting data opens you to other perspectives.



*Robert J. Shiller

USING EXISTING DATA

Public health and behavioral health practitioners often work with existing data, or secondary data as you learned in Chapter 2. When you work with existing data, you do not directly collect the data yourself. However, it can be just as complicated to work with data in other ways. You will need to know information about how the data was collected, the intent of the data, the steps to request data, and how to respect the data confidentiality concerns of other organizations.

To help you identify some existing local, state, and national data sources in Colorado that may be helpful, you can consult the table of data sources in the companion guide to this manual. You can use this table to help you contact different agencies, explore data websites, and identify the types of data other agencies may already be collecting.

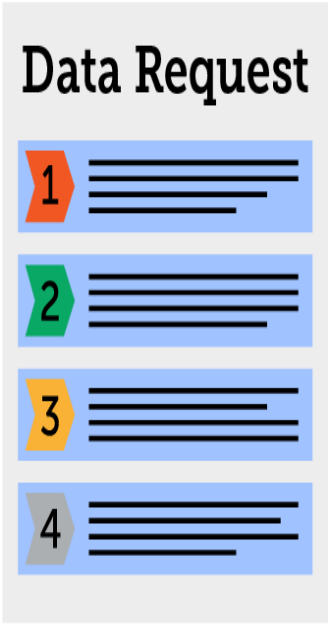
USING EXISTING DATA IS USUALLY MORE COST-EFFECTIVE, BUT IT DOES INVOLVE A CERTAIN LEVEL OF EXPERTISE IN NEGOTIATING WITH OTHER ORGANIZATIONS, AS WELL AS SOME STATISTICAL EXPERTISE TO ANALYZE THE DATA FOR THE ISSUE YOU ARE ADDRESSING.

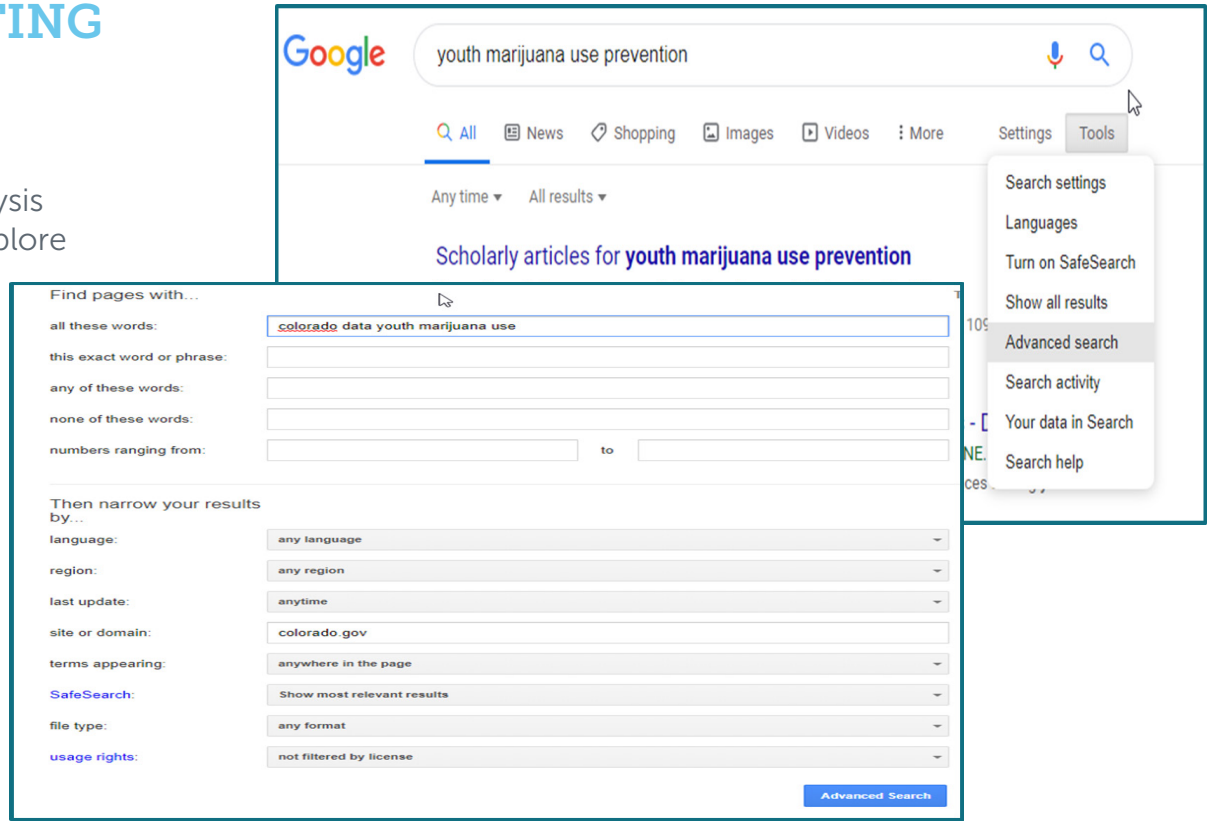
KEY POINT:

Assess the existing data before jumping right in to collect data on your own. Use regional data instead of county-level data or find a community with similar demographics that has existing data on your research issue to use as a proxy to save time and resources. Evaluate the practicality of collecting data on your own by exhausting other avenues first.

TIPS FOR REQUESTING EXISTING DATA FROM AGENCIES

You can find value in reports, not just during analysis and reporting, but also when you are trying to explore what data an organization collects that may be useful. Data reports or products from institutions provide clues as to what data you can collect or access from other organizations. If you already know that a certain agency has the data you are looking for, you can go to their website and navigate to reports.





If you are not sure what data sources exist on your topic, an easy way to do this is to use Google Advanced Search. You can limit your search to file formats, like .pdf, Word, Excel, or certain domains. For Colorado government data, you can specify formats like "Colorado.gov" to retrieve government report documents. To the right is an example of using Google Advanced Search to look for Colorado data on youth marijuana use. Select settings and Advanced search to filter results as shown.

ANALYSIS DATA SET

When you request data from other agencies, you must address data ethics.[11] Other agencies want to know that when they share data with you, your agency will properly protect the information of the clients or participants in their database. If you do not need the identifying information of participants or clients (often called Personally Identifiable Health Information, Protected Health Information, or PHI), it is in your best interest, and the interest of the agency you are getting data from, to obtain what is called an **analysis data set**.

In an analysis data set, you remove personal identifiers before sharing data to decrease the risk or threat to confidentiality. If you are trying to obtain data from an agency that may have concerns about how the data will be stored and managed, think about whether you really need the entire dataset before you make the request. Be sure to properly cite the original source of the data and verify you have the correct citation with the agency.

Name	County	Race
[redacted]	Pitkin	Asian
[redacted]	El Paso	Multiple Races
[redacted]	Jefferson	White - Non Hispani
[redacted]	Weld	Black or African Ame
[redacted]	Montezuma	White - Hispanic
[redacted]	Eagle	White - Non Hispani
[redacted]	El Paso	American Indian or A

DECISIONS TO MAKE PRIOR TO MAKING A DATA REQUEST:

1. Purpose of the data request

2. Types of demographic variables needed (age, sex, race, ethnicity, income)

3. Level of detail for analysis (percentages, frequencies, scale scores, rates (health), performance level (education))

4. Format for report (Excel spreadsheet, PDF, Word document, etc.)

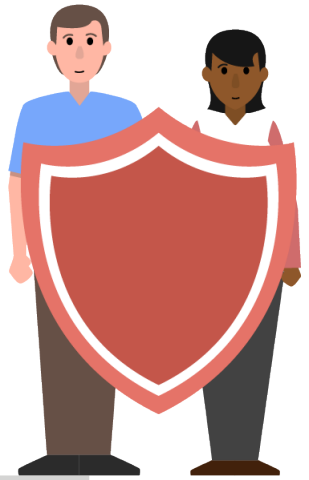
5. Are you requesting PHI (Personally Identifiable Health Information)? If so, additional requirements for data sharing may need to be in place.

Now that you know some tips for finding existing data sources, you can begin to assemble your community assessment data resources and develop a contact list. Before contacting agencies and requesting data though, you need to have a basic understanding of data ethics.

IRB APPROVAL

If you will be conducting research, check your agency's policy on obtaining Institutional Review Board (IRB) approval. IRBs protect the rights and welfare of research participants recruited for studies affiliated with an organization. You may have an internal IRB, or you may need to utilize an external IRB. To search for IRBs in your area, you can go to the Department of Health and Human Services, Office for Human Research Protections (https://ohrp.cit.nih.gov/search/irbsearch.aspx?styp=bsc).

You can discuss your research proposal with the IRB before completing the application process. Often evaluation or community assessment is exempt from review, but you should check with your organization before moving forward, especially if you will be working with identifiable data or you are conducting primary data collection with protected groups, such as youth or prison inmates.



KEY POINT:

You can increase chances of success with an IRB by following these best practices[12]

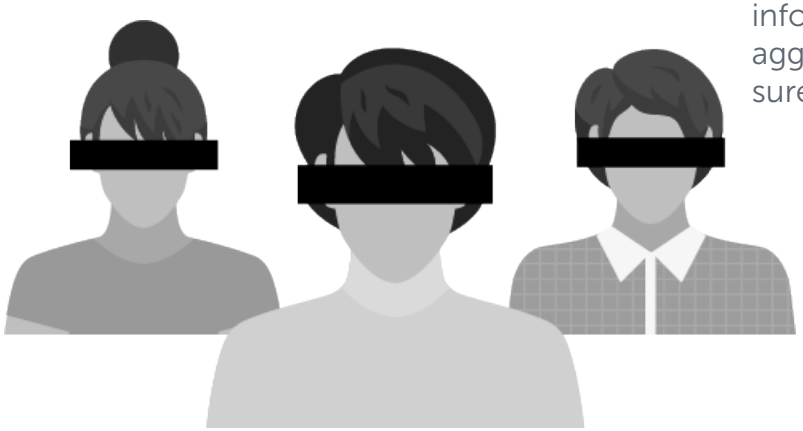
- Build a positive relationship with the IRB. Engage in dialogue through in-person meetings or phone calls about your research.
- Follow the IRB's guidelines and submit all required paperwork.
- Demonstrate knowledge and expertise.
- Ask questions and be proactive.

DATA PRIVACY

You also need to consider data ethics and HIPAA, or the Health Insurance Portability and Accountability Act. HIPAA Privacy and Security Rules protect the privacy and security of individually identifiable health information. If you would like more information regarding HIPAA, there are some free trainings through the federal government listed in the Further Learning section at the end of this chapter. Your organization may also have a HIPAA compliance officer that could connect you with trainings or answer any questions you might have.

If you are working with student data, you will also need to consider FERPA, or the Family Educational Rights and Privacy Act. FERPA protects the privacy of student education records, including surveys, assessments, test, and evaluations, and ensures schools get written parental consent before students are required to participate.

If the data you need is publicly available for download on a website, you can assume you have permission to use the data without making a data request and that data analysts have removed all potentially identifiable information. Often the data you find publicly available has been aggregated, or summarized, so that individuals cannot be identified. Make sure to properly cite the source of the data.



COLLECTING DATA

WHAT IF EXISTING DATA DOES NOT MEET YOUR NEEDS? OR WHAT IF IT GIVES YOU ONLY PART OF THE STORY?

You then need to collect data on your own to fill in the gaps. This often happens in prevention or public health, especially if you are trying to assess local problems. You may have access to some national data, or even state and regional data, but it may not be relevant to your community.

TAKE AN EXAMPLE OF PLANTING A GARDEN.

> You find your vegetables are not growing well. You might look online and find various reasons why the plants may not be growing well in the Farmer's Almanac. You may even locate a Colorado state guide for growing vegetables and ways to grow healthy productive plants. Regionally, you gather information on soil types, temperature, and average rainfall, but this still does not give you the detail you need to fix the problem. You may decide to measure the pH of the soil where things are growing well in your yard, and the pH of the soil where things are not growing well. You also measure the rainfall and record the amount of time the plants spend in the shade and the sun each day. Now you are collecting primary data that is locally sourced! This will help you accurately diagnose the conditions needed to produce healthy vegetables in your yard.

Starting with national, state, or regional data is a best practice, and then you can examine the issue with local data collection. By doing so, you will gain a comprehensive assessment of the health issue without the expense of collecting primary data. In addition, you can use national and state data sources to establish credibility.

In some cases, you may be able to answer your research questions adequately using existing data, especially if data is available at the county or zip code level.

OFTEN YOU HAVE MORE QUESTIONS AFTER LOOKING AT EXISTING DATA AND NEED TO COLLECT MORE INFORMATION.

For example, after seeing youth marijuana rates for your state, you may decide that you want to look at your county's marijuana rates. Perhaps you notice lack of parental monitoring in your county is correlated highly with youth marijuana use, but this does not help you understand the local context for why this is occurring. You may decide to conduct a focus group with youth at local school districts who report students are smoking marijuana after school before their parents come home. It is the combination of utilizing existing quantitative data (secondary) and new local qualitative data (primary) that provide you with a comprehensive assessment to inform programs and strategies.

NEEDS ASSESSMENT

"A needs assessment is a systematic approach to identifying community needs and determining program capacity to address the needs of the population being served."[13] *– SAMHSA*

Before you begin collecting data, make sure you address each of the steps below to ensure a well-defined assessment. *Following these steps can help you be strategic with the* ***who, what, where, when, and why*** *of data collection and best utilize your resources.*

1. Define the purpose and goals for the assessment

2. Identify the target populations for the assessment of needs and services

3. Determine how data will be collected and used

4. Determine the timeline for the process

5. Determine the strategic use of the findings

ONCE YOU HAVE OUTLINED YOUR NEEDS ASSESSMENT PROCESS AND DETERMINED YOUR DATA COLLECTION METHODS, YOU CAN MOVE INTO IMPLEMENTING DATA COLLECTION.

In the previous example of youth marijuana use, you collected additional local data using focus groups, but there are other ways to gather primary data. You will now explore the different methods of collecting primary data.

It is a best practice to use both quantitative and qualitative methods when conducting a needs assessment. For a more in-depth review on how to implement these methods, refer to the Further Learning section at the end of this chapter.

WAYS TO COLLECT DATA

There are many ways to collect data, but the following examples are the most common methods you will encounter in public health or behavioral health.

FOCUS GROUPS

In a focus group, you engage members on a specific discussion or topic. Focus groups are qualitative in nature and strategically designed with the intent of encouraging members to speak openly. You use a discussion guide to obtain detailed information from participants. It is helpful to work with a trained facilitator and evaluator to effectively conduct the focus group and analyze the data. You can also use an external evaluator (outside your organization) to help elicit full and honest responses from participants.

KEY INFORMANT INTERVIEWS

Interviews, also a qualitative method, can be thought of as a conversation with a purpose. If you need more information in your community about the perceptions of key informants (community experts), this is a great way to collect data. The researcher works one on one with the key informant or expert to garner in-depth information on a particular topic.

OBSERVATION

During observation, you watch or participate in the activities of people, places, or conditions where they exist. Observation is also qualitative in nature.

SURVEYS

A survey is typically a questionnaire that asks about a community's needs. You analyze the results and then use the data to guide further action on a problem. You collect mostly quantitative data on a survey, but you can include a few open-ended responses to collect qualitative data.

SERVICE RECORDS

You can also collect data on access to services, depending on the problem you are exploring. For example, if you are starting a new syringe exchange program, you will want to collect data on who (demographics such as race, age, gender, education level, etc.) is accessing the program, where, when, and how often. This method is mostly quantitative in nature.

SOFTWARES FOR DATA COLLECTION

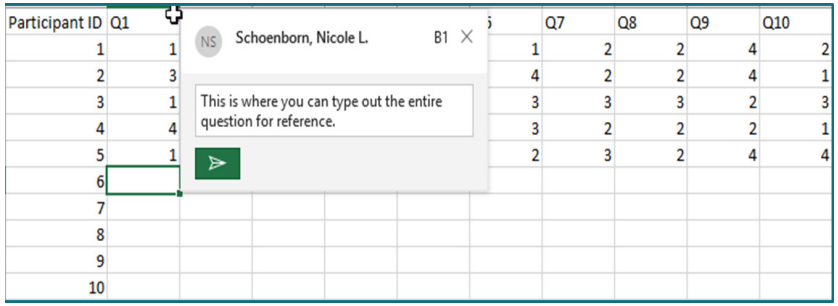
YOU DON'T HAVE TO SPEND A LOT OF TIME AND MONEY USING FANCY DATA COLLECTION TOOLS. YOU CAN USE **MICROSOFT EXCEL, ACCESS**, AND **WORD** FOR MOST SMALLER-SCALE QUANTITATIVE AND QUALITATIVE STUDIES. IF YOU HAVE THE MICROSOFT OFFICE PACKAGE, YOU ALREADY HAVE ACCESS TO THIS SOFTWARE.

To set up an Excel spreadsheet to record data, simply open Excel and select "New Workbook." You will use your first row to label each column with a variable name to distinguish each data point you will collect from the participants.

Variable names should be descriptive. Remember to include a participant ID variable to keep track of participants. Each row will be the responses given from a participant, with one respondent's answers per row.

To the right is an example of how this might look. You can also insert a comment in the column header for each question and type in the entire question for easy reference later, which is helpful for making sure you are on the right question. In this example, survey data has been entered for five participants for questions 1-10. This works well if you are collecting paper surveys in-person or through the mail.

The University of Wisconsin Extension Office has a great resource that walks you step-by-step through the survey process in Excel. Check the Further Learning section for a more in-depth tutorial.



QUALITATIVE DATA COLLECTION

For qualitative data, you can record data in **Microsoft Excel** or **Word**. After you have recorded the notes or had audio recordings transcribed, you can create a document with participants' responses to questions. You can write a summary of a focus group, gather notes, and type transcriptions all in Word or Excel!

KEY POINT:

Use tools already available on your computer for data collection, including Microsoft Excel and Word.

LOCAL DATA COLLECTION CHALLENGES

DESPITE THE RICHNESS LOCAL DATA CAN PROVIDE, COMMUNITIES OFTEN FACE MANY CHALLENGES WITH OBTAINING LOCAL DATA.

With existing data, you may find national and state databases have missing, invalid, unreliable, or suppressed data at the county level.[14] To overcome this problem, you may use data for a nearby geographic area, which may be regional level data. By looking at regional data for Colorado, you can get a better picture than just looking at state level data. You then collect local data to complement your findings. You can spend a lot of money and time doing primary data collection, so thoroughly evaluate the utility of collecting more local-level data before implementing. Refer to the question box to guide your decision.

Before collecting local level data on your own, ASK:

1. How will you recruit?

2. Will people be willing to participate?

3. Do you have the resources, both human and financial, to carry out data collection?

4. What is the value added to the data that already exists?

5. Are there similar communities (demographics) in your area that have available data that you could use as a proxy?

6. Do you have access to expertise later for analysis?

APPENDIX

FURTHER LEARNING

HIPAA Training and Resources
https://www.hhs.gov/hipaa/for-professionals/training/index.html
https://www.colorado.gov/pacific/hcpf/health-insurance-portability-and-accountability-act-hipaa-0
http://www.ucdenver.edu/research/ORC/HIPAA/Pages/PHI.aspx

Research Ethics and Compliance Training
https://about.citiprogram.org/en/homepage/

HIPAA Training and Resources
https://www.hhs.gov/hipaa/for-professionals/training/index.html

Data Collection Methods Semi-Structured Interviews and Focus Groups
https://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR718.pdf

Coursera: Qualitative Methods
https://www.coursera.org/learn/qualitative-research

Conducting Focus Groups
https://ctb.ku.edu/en/table-of-contents/assessment/assessing-community-needs-and-resources/conduct-focus-groups/main

University of Wisconsin-Extension Using Excel for Analyzing Questionnaires
https://learningstore.uwex.edu/assets/pdfs/G3658-14.pdf

KEY TERMS

Focus Group: a discussion group designed to gather qualitative data about a specific topic.

Key Informant Interviews: working one-on-one with an informant or expert to garner in-depth information on a topic.

Needs Assessment: a systematic approach to identifying community needs and determining program capacity to address the needs of population being served.

Observation: collecting data through observing behavior.

Service Records: data collected when a person is accessing a service.

Surveys: a questionnaire used to collect data.

REFERENCES

11. Ross, M. W., Iguchi, M. Y., & Panicker, S. (2018). Ethical aspects of data sharing and research participant protections. American Psychologist, 73(2), 138-145.

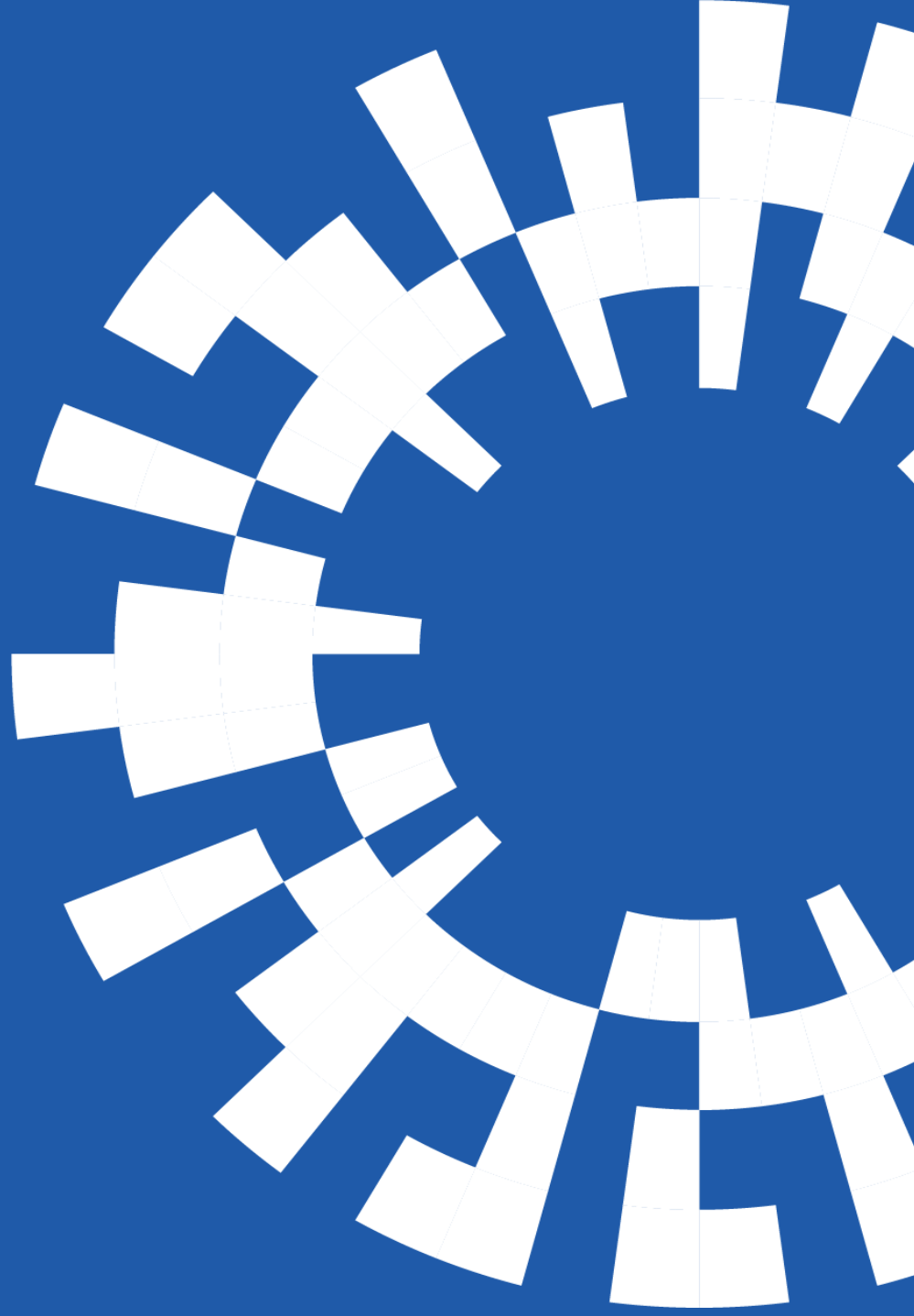
12. Cartwright, J. C., Hickman, S. E., Nelson, C. A., & Knafl, K. A. (2013). Investigators' successful strategies for working with Institutional Review Boards. Research in Nursing & Health, 36(5), 478–486.

13. Substance Abuse and Mental Health Services Administration (2019). How States Can Conduct a Needs Assessment. Retrieved from https://www.samhsa.gov/section-223/certification-resource-guides/conduct-needs-assessment.

14. Gurganus, K. M., Butt, A. L., Kirchenbauer, C. M., Melkvik, C., Piatt, J., Hawkins, J., U'Ren, S., & Onuorah, Y. (2014). Implementing the Regional Epidemiological Outcomes Workgroup (REOW) in the State of Oklahoma for Substance Abuse Prevention: An ODMHSAS Project. Community Mental Health Journal,51(5), 535-539.

5

Working with Data



CLEANING DATA

*"Without a systematic way to start and keep data clean, bad data will happen."**

Many data sources have purposes other than analysis. Therefore, the data is not usually standardized. This means the data may not have specific field lengths, appropriate spelling or grammar, or pre-defined options to choose from in a data field. For example, if you work in a patient waiting area and you are gathering information to check patients in for their appointment, you are entering data from a registration form and into an electronic medical record (EMR) system. You are not entering data designed for research. You are focused on getting people in to see the doctor for their presenting medical issue.

When you access this data, you may notice patient addresses are in multiple fields and different formats. In the name column, some may have the first, middle, and last, and some have just the first, middle initial, and last name. Some of the birthdates may not be in the same format. Some have the format "mm/dd/yy" and some "month day, year." This makes analysis very difficult. Computers cannot analyze data when formats do not match up!

KEY POINT:

Before you begin analyzing data, make sure you familiarize yourself with the data and examine it for errors. Clean the data before analyzing and interpreting data.

Patient Name	Street Address	City	Zip Code	Date of Birth	Gender
Alena Vasilieva				Aug 01 95	Female
Desiree Pearl Perea	1346 Clover Drive Arvada 80004			04/26/1983	Female
Frank Andrew Goodson	4941 McKinley Ave.	Denver	80209	February 6th, 1993	Male
Gregory Fabien Sault	558 Sweetwood Drive	Denver, 80213		03/05/78	Male
Guadalupe J Lopez	4793 Scheuvront Dr, Unit 3	Arvada	80002	3/18/1964	Female
Kyou Morishita	4239 Stark Hollow Road	Denver	80216	May 17 1974	Female
Pamela G Campos	357 Victoria Street, Denver, 80			10/15/1968	Female

*Donato Diorio

CLEANING DATA

Follow these tips to help you clean up messy data:[15]

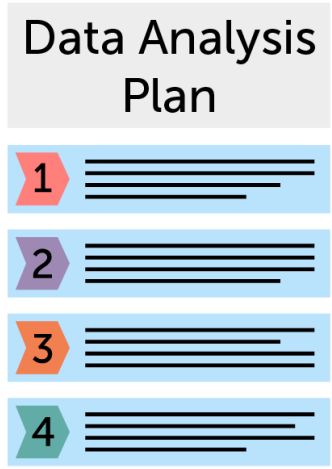
1 **Explore the data** and get familiar with it. You can do this by using Excel and sorting the data by columns, filtering the data, or creating frequency tables of the categorical variables (such as gender). A frequency table would just have the total number of each category of gender in this case. A categorical variable is just a variable, like gender, that has different values. Others include race, ethnicity, age group, or education.

2 **Change the data to match** a specific format. For example, if you have gender as a variable, but it is entered in different formats when you receive the data, you can change "Male" and "M" to a value of 1, "Female" and "F" to 2, and "Transgender" and "T" to 3 for your gender variable. This is called recoding data.

3 **Perform basic calculations** in Excel to check for errors in numerical data. This may highlight outliers or errors in data entry. Outliers are data that have unusual values compared to the rest of your data.

4 **Explore** using OpenRefine, which is a free software that helps you clean up messy data. For large datasets, using a data cleaning software or program may save you time. You can learn more about cleaning your data in the Further Learning section at the end of this chapter.

5 **Refer to the data dictionary.** The data dictionary, or codebook, will help you decide what are valid and invalid values for a field or variable before you make changes. It also specifies missing values and how they are coded. If there is no data dictionary, you can contact the data owner and ask for assistance.



ANALYZING DATA

"Data analysis is the process of ***systematically*** examining data with the purpose of spotlighting ***useful*** information. Data analysis is the foundation of scientific research. Conducting a complete analysis allows the ***impact*** of the work to be determined, the ***quality*** of programming to be assessed, and the results to be ***communicated*** to stakeholders."[16]

As you go through the steps of data analysis, ask yourself:

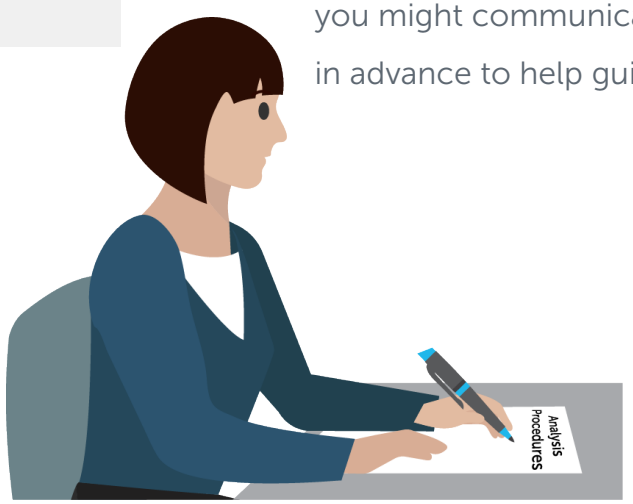
1. **AM I BEING *SYSTEMATIC?*** Make sure you are documenting your analysis procedures so that someone can else can replicate the procedures. From the beginning, make a plan for how you will analyze the data.

2. **IS THIS *USEFUL?*** Think about how others can utilize the results. Use analysis procedures that will have practical application for stakeholders.

3. **DOES IT HAVE *IMPACT?*** Consider how findings may influence public policy or programs. The way you report results can have both positive and negative impacts on programs.

4. **DOES IT AFFECT THE *QUALITY*** of a program or strategy? Assess how results might influence the quality of public health or prevention programming in Colorado—what improvements might be needed?

5. **CAN IT BE EFFECTIVELY *COMMUNICATED*** to stakeholders? Examine how you might communicate the data analysis methods and results to stakeholders in advance to help guide your decisions.



KEY POINT:

With these questions in mind, you can begin your data analysis process!

Step 1.

WHAT IS THE PROBLEM YOU ARE TRYING TO SOLVE WITH YOUR DATA?

Remember to go back to the assessment steps in Section 4 and clearly outline the purpose of the analysis. You don't want to analyze data without a clear purpose.

FILL IN THE BLANKS BELOW TO HELP YOU OUTLINE YOUR PURPOSE FOR ANALYSIS.

Purpose Statement for Analysis

"With this data, I intend to _____

so that _____."

Step 2.

Common Software for Data Analysis[17]

Common Types of Software for Data Analysis[17]

	• Excel is part of the Microsoft Office software package • Easy-to-use program for basic data management and analysis • Provides users with basic descriptive statistical functions • Useful for an evaluation plan that does not include a rigorous experimental design • For more information on Microsoft Office products visit the official Microsoft Office site.
	• Access is part of the Microsoft Office software package • Can analyze multiple fields at a time • Can perform a variety of functions, including descriptive statistics and t-tests • For more information on Microsoft Office products visit the official Microsoft Office site.
	• Statistical Package for the Social Sciences (SPSS) is designed specifically for social science statistical analysis • An advanced program that may require training • Can be used to store large amounts of data • Can perform complex statistical functions including regression analysis and correlation and variance analysis • Suitable for experimental, longitudinal designs or basic evaluation needs • For more information on SPSS, visit the SPSS site.
	• SAS is a statistical program designed specifically for social science statistical analysis • Advanced statistical program with a wide range of functions, similar to SPSS • Includes survey data analysis, regression analysis, and categorical data analysis • For more information on SAS, visit the SAS site.
	• R is a free statistical software that integrates a wide variety of types of statistical analyses • Advanced statistical program that can be extended with readily available functions and extensions • Includes complex statistical functions, including both linear and non-linear modeling • For more information on R, visit the R site.

WHAT TYPE OF DATA DO YOU HAVE?

Think about the type of data you have in your spreadsheet or data file. Ask these questions:

IS IT QUALITATIVE OR QUANTITATIVE?

Qualitative data is represented by words, and quantitative data is represented by numbers. In most datasets, you will have both types of data.

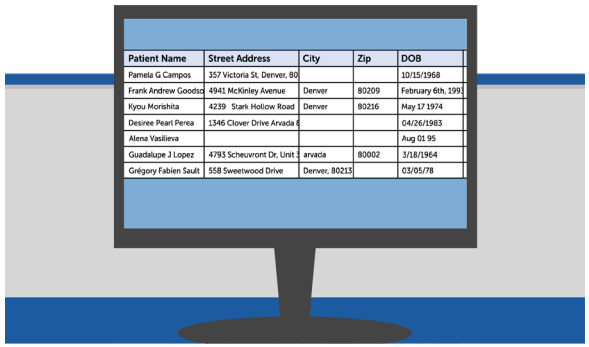
WHAT IS THE FORMAT OF THE DATA FILE?

Your data file may be in Excel (.csv or .xls), Word (.rtf or .doc), DBase (.dbf), Access (.mdb), Adobe (.pdf), or many more! Some formats import easily into statistical software and others do not.

WILL YOU NEED TO IMPORT THE DATA INTO ANOTHER SOFTWARE PROGRAM TO ANALYZE IT?

Check the help menu of your analysis software to determine if you have the data in the appropriate format before importing. If you already have your data in Excel and plan on using Excel for analysis, you will not need to worry about this step. Sometimes you need to perform more advanced statistics than Excel can run, but Excel files can be imported into analysis packages.

Step 3.



KEY POINT:

Develop a plan for analyzing the data.

Document your procedures, so others know how you arrived at your conclusions, or so they can replicate the analysis.

WHAT STATISTICAL TESTS WILL YOU USE?

Calculating Summary Statistics

IN YOUR COMMUNITY WORK, YOU WILL CALCULATE SUMMARY STATISTICS MORE OFTEN THAN INFERENTIAL STATISTICS.

You covered most of these (mean, median, mode, and range) in Chapter 2 when you evaluated results in research articles. You can calculate these statistics using formulas in Excel. These statistics are often referred to as measures of central tendency.

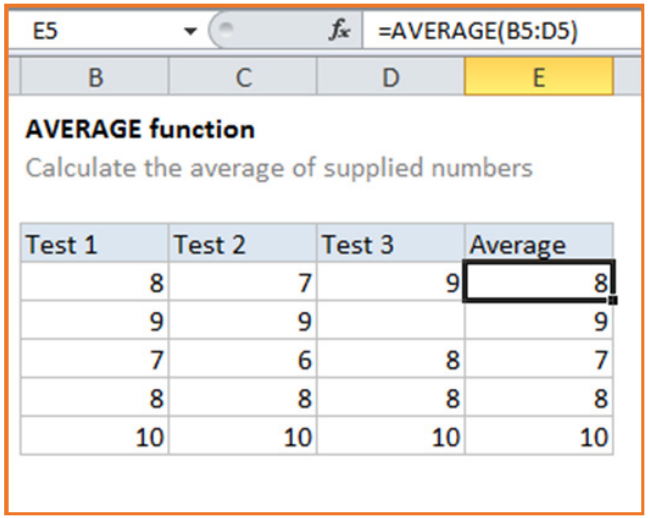
Other important functions for summary statistics in Excel are SUM, COUNT, and calculating a percentage. You can create simple formulas in Excel to run these calculations.

mean	**median**
mode	**percentage**
frequency	**range**

WHAT STATISTICAL TESTS WILL YOU USE?

Using Functions in Excel

In the example below, you will use the function AVERAGE in Excel to calculate the mean. You can use SUM, COUNT, MEDIAN, MODE, MIN and MAX for a cell range in the same way. You simply click the cell where you want the calculation, type an "=" sign, and that prompts Excel to know you are going to use a formula. You then type the cell range or highlight it with your mouse in parentheses after the function name, as seen below. Use your Excel help menu to explore other functions for calculations.



KEY POINT:

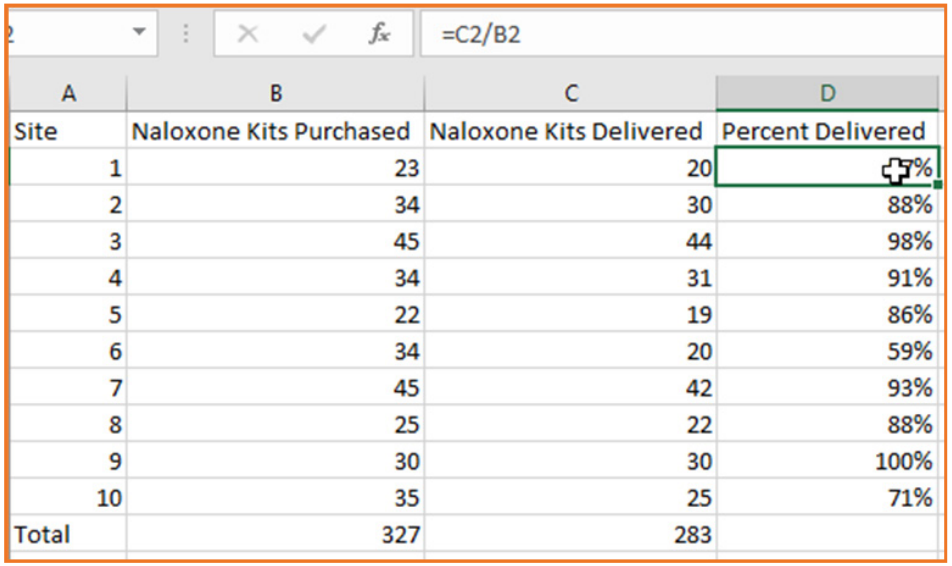
Use Excel to perform summary statistics.

If you need to do more advanced statistical analysis, Excel can be easily imported into statistical packages like SPSS or SAS.

WHAT STATISTICAL TESTS WILL YOU USE?

Calculating Percent

In Excel, you can calculate percentages easily by typing an "=" in the cell where you want to do the calculation. Then click on the cell for the numerator, type "/" for divide, and then click on the cell for the denominator. Using the example below, you can calculate the following: of the naloxone kits purchased, what percent were actually delivered to the site? You can view the formula below, and then this can be copied for each row to calculate the percent by site.



	A	B	C	D
	Site	Naloxone Kits Purchased	Naloxone Kits Delivered	Percent Delivered
	1	23	20	87%
	2	34	30	88%
	3	45	44	98%
	4	34	31	91%
	5	22	19	86%
	6	34	20	59%
	7	45	42	93%
	8	25	22	88%
	9	30	30	100%
	10	35	25	71%
	Total	327	283	

WHAT STATISTICAL TESTS WILL YOU USE?

Calculating Percent of a Total

If you want the denominator to be constant and not change cell locations, you can specify an absolute value. For example, think back to your naloxone kit data. You need to calculate the percentage of the total naloxone kits purchased. You type an "=" sign in any empty cell, click on the naloxone kits purchased for site 1, and then divide by the Total Naloxone Kits purchased. For each of these site percentages, the denominator will not change so you can put a "$" in the formula before both the column letter and row number in the formula, so this value does not change as we copy the formula. In this example, the absolute value is cell "B12" below.

=B2/B12

A	B	C	D	E
Site	Naloxone Kits Purchased	Naloxone Kits Delivered	Percent Delivered	Percent of Total Purchased
1	23	20	87%	7%
2	34	30	88%	10%
3	45	44	98%	14%
4	34	31	91%	10%
5	22	19	86%	7%
6	34	20	59%	10%
7	45	42	93%	14%
8	25	22	88%	8%
9	30	30	100%	9%
10	35	25	71%	11%
Total	327	283		

Step 3.

Analyzing Epidemiological Data

In public and behavioral health, you are often given summary tables that have counts of the number of people who responded in a defined way. From the counts, you can calculate a percent to determine how many answered a certain way out of the total number of participants.

The total number of participants is often stated as N = [Number of participants] on the chart or table. Sometimes N is the total population when you work with epidemiological data. Below is an example using Census Data from 2010 for Durango, Colorado.[18] Notice the percent is calculated for you. This column is equal to the Number column divided by the Total Population for each category. This also highlights a great resource for demographic data, American Fact Finder, managed by the US Census Bureau. You can also refer to this data source in the companion guide and run reports for your city, county, state, and national data.

Geography: Durango city, Colorado ▼

Subject	Number	Percent
SEX AND AGE		
Total population	16,887	100.0
Under 5 years	791	4.7
5 to 9 years	735	4.4
10 to 14 years	659	3.9
15 to 19 years	1,603	9.5
20 to 24 years	2,662	15.8
25 to 29 years	1,773	10.5
30 to 34 years	1,210	7.2
35 to 39 years	1,072	6.3
40 to 44 years	863	5.1
45 to 49 years	967	5.7
50 to 54 years	1,115	6.6
55 to 59 years	1,033	6.1
60 to 64 years	766	4.5
65 to 69 years	499	3.0
70 to 74 years	326	1.9
75 to 79 years	250	1.5
80 to 84 years	225	1.3
85 years and over	338	2.0

KEY POINT:

Look for relationships and patterns. Numbers alone have no meaning. Look for patterns over time and relationships between data.

Step 3.

Inferential Statistics

You may need to analyze data using more advanced statistical methods.

The table to the right provides some basic types of inferential statistics and how they are used in practice.

You may come across inferential statistics when conducting a literature review, or when you are evaluating the evidence of effectiveness for a particular strategy. When you publish, you will likely analyze data using one of these statistical tests. If you need to run inferential statistics, you can use SPSS or SAS to analyze your data. You can download a free limited version of SAS on their website (refer to the Further Learning Section). Some advanced statistical expertise is required when analyzing inferential statistics.

UTILIZING A LOCAL EVALUATOR OR STATISTICIAN MAY BE NEEDED TO HELP YOU RUN ANALYSES AND INTERPRET RESULTS FROM INFERENTIAL STATISTICS.

TYPES OF INFERENTIAL STATISTICS TESTS[19]

Type of Test	Use
Correlational: these tests look for an association between variables	
Pearson Correlation	Tests for the strength of the association between two continuous variables
Spearman Correlation	Tests for the strength of the association between two ordinal variables (does not rely on the assumption of normally distributed data)
Chi-Square	Tests for the strength of the association between two categorical variables
Comparison of Means: these tests look for the difference between the means	
Paired T-Test	Tests for the difference between two variables from the same population (e.g., a pre- and post-test score)
Independent T-Test	Tests for the difference between the same variable from different populations (e.g., comparing boys to girls)
ANOVA	Tests for the difference between group means after any other variance in the outcome variable is accounted for (e.g., controlling for sex, income, or age)
Regression: these tests assess if change in one variable predicts change in another	
Simple Regression	Tests how change in the predictor variable predicts the level of change in the outcome variable
Multiple Regression	Tests how changes in the combination of two or more predictor variables predict the level of change in the outcome variable
Non-Parametric: these tests are used when data does not meet assumptions required for parametric tests	
Wilcoxon Rank-Sum Test	Tests for the difference between two independent variables; takes into account magnitude and direction of difference
Wilcoxon Sign-Rank Test	Tests for the difference between two related variables; takes into account the magnitude and direction of difference
Sign Test	Tests if two related variables are different; ignores the magnitude of change—only takes into account direction

Step 4.

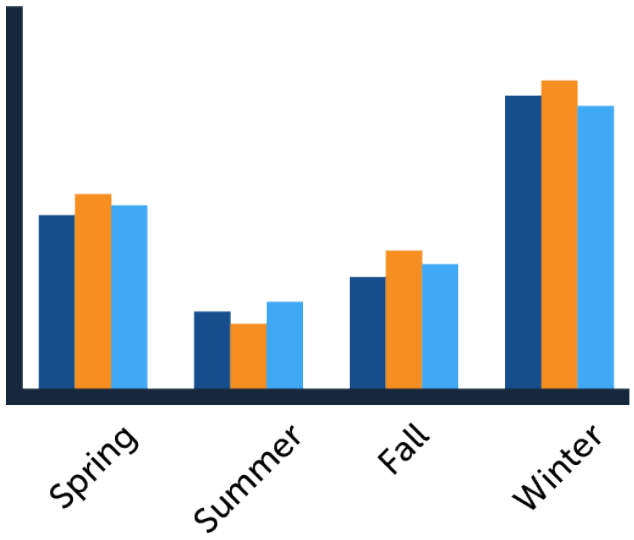
WHAT PATTERNS OR TRENDS DO YOU NOTICE IN THE DATA?

Incidence and Prevalence

When you are interpreting epidemiological data, a first and simple step is to count the cases in the population of interest.

ALWAYS CHECK WHETHER DATA SOURCES ARE PROVIDING INCIDENT (NEW EVENTS AMONG THE POPULATION) OR PREVALENT (AN EXISTING EVENT AT A SPECIFIC POINT IN TIME) CASES.

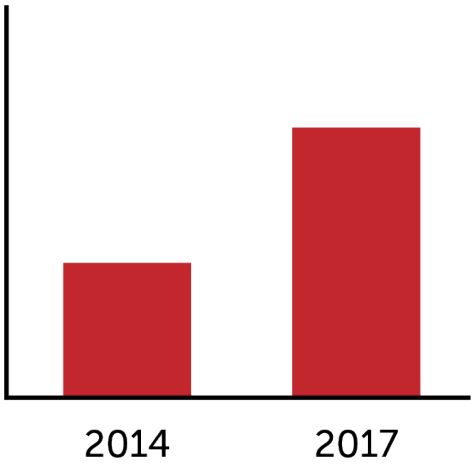
For incident cases, specify the period during which the cases occurred. This count of incident cases over time in a population is called incidence. The counts of incident or prevalent cases can be compared with their historical norm or another expected or target value. These case counts are valid for epidemiologic comparisons only when they come from a similar population or approximately the same size.



WHAT PATTERNS OR TRENDS DO YOU NOTICE IN THE DATA?

Making Comparisons

Comparing your data to other, existing data provides a context for understanding your assessment or evaluation results. For example, survey data from your high school may seem, at a glance, to reveal high smoking rates among 9th-graders. However, a comparison of these data to statewide Colorado data might show that your school's smoking rates are much lower than the statewide average.



WHEN EXAMINING DATA, YOU CAN MAKE COMPARISONS, SUCH AS:

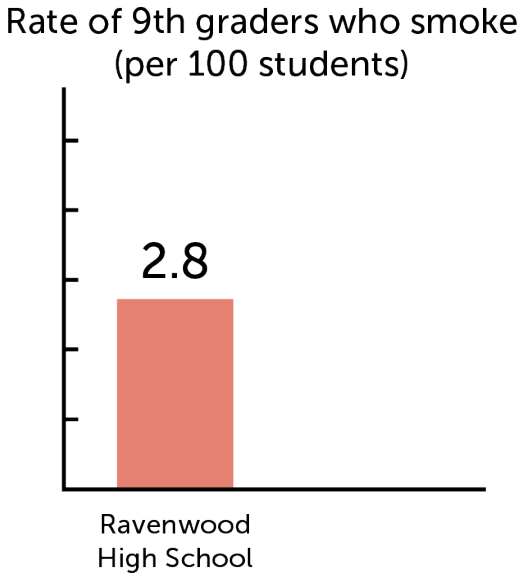
- Between the community now and the community in the past
- Between the community and other similar communities
- Among different population groups in the community, including different age groups and genders
- Between the community and the state
- Compare current 30-day use to past 30-day use and find out if alcohol use is higher or lower than it once was.
- Compare usage rates of males and females and discover if use among males is higher or lower than among females; or compare 14-year-olds and 17-year-olds, and see how their rates compare.
- Compare current use in your community to average rates across the state of Colorado and find out if alcohol use in your community is lower or higher than the Colorado state average.

WHAT PATTERNS OR TRENDS DO YOU NOTICE IN THE DATA?

Rates

EPIDEMIOLOGY USES RATES TO STANDARDIZE DATA IN ORDER TO BE ABLE TO COMPARE DATA ACROSS POPULATIONS THAT ARE DIFFERENT SIZES.

A rate is a measure of the frequency with which a health event occurs in a defined population over a period of time. For example, if a town with a population of 30,000 has 500 arrests in a given year for driving under the influence (DUIs), you can divide 500 by 30,000 to get a rate of .017 arrests per person per year. To make the rate easier to understand, you can multiply by 1,000 and say that the rate of DUIs for the town is 17 per 1,000 people per year. A neighboring town of 22,000 people with 367 DUI arrests would have the same rate (divide 367 by 22,000 and then multiply by 1,000). For national data sources or larger population sizes, you can calculate the rate with 10,000 or 100,000 population size. Multiply by whatever size makes sense for your community.



Calculation for rate per 1,000 people per time period:

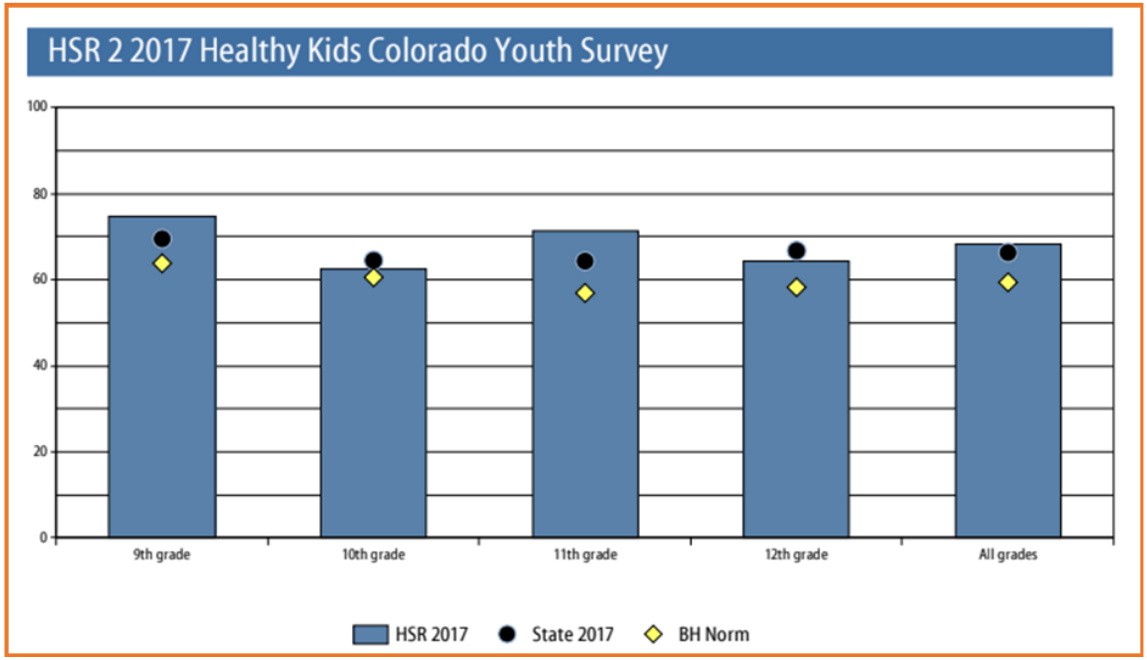
$$\text{Rate} = \frac{\text{Number of cases}}{\text{Population over time period}} \times 1{,}000$$

WHAT PATTERNS OR TRENDS DO YOU NOTICE IN THE DATA?

Benchmarking

YOU CAN ALSO REFER TO USING COMPARISONS AS BENCHMARKING.

When you use data from another source or geographic region to set a target or compare how you are doing to a standard, you are using a benchmark. In Colorado, you can go to the Communities that Care Risk and Protective Factor Profiles and use the cut-off point as a benchmark for whether your area is above or below the national average (BH Norm) for that particular risk or protective factor. Next, you can compare it to the state average, as depicted in the graph to the right.[20]



In this example, the data in **BLUE** represents the regional data or local data, the **BLACK DOT** represents the state of Colorado average, and the **YELLOW DIAMOND** represents the behavioral health norm or national average. You can evaluate if you are doing better or worse compared to other standards by using these benchmarks. This can help you identify which risk or protective factors are a concern in your local area.

WHAT PATTERNS OR TRENDS DO YOU NOTICE IN THE DATA?

Small Numbers

When looking at data and making comparisons, look closely at sample size. Small numbers can sometimes be misleading. For example, perhaps you notice the data indicates that between 2014 and 2017, motor vehicle accidents increased by 100%. However, when you look at the data more closely, only one motor vehicle accident occurred in 2014 and two in 2017.This is still an increase of 100%, but one that is not so alarming. In this case, it is important to look at data both in terms of percentages and as actual numbers.

Keep in mind the following:

- When small numbers are used to calculate rates or percentages, **differences may be exaggerated.**

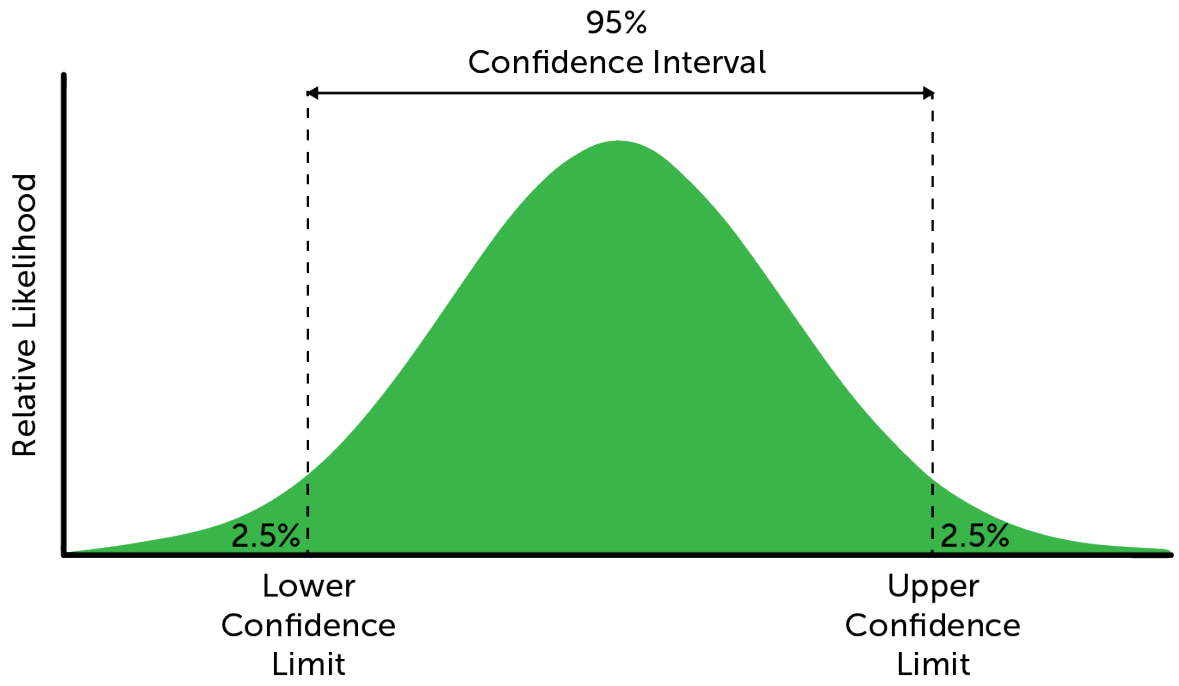
- **Low percentages of some behaviors should not be dismissed as unimportant,** especially if the behavior has severe consequences.

WHAT PATTERNS OR TRENDS DO YOU NOTICE IN THE DATA?

Confidence Intervals

WHEN YOU ARE ANALYZING DATA OR REVIEWING DATA FROM AN ARTICLE, IT IS IMPORTANT TO LOOK AT THE LEVEL OF CONFIDENCE IN RESULTS, AS WELL AS TESTS THAT SHOW IF YOUR INTERVENTION OR STRATEGY MADE A DIFFERENCE.

A confidence interval, often set at 95%, means that 95% of the time, you can be confident the mean would fall within the range. There is typically an upper limit and lower limit specified. In other words, you are confident that 95 times out of 100 the mean will fall within the upper and lower limits. If the range is larger, you should be less confident that your results are precise. If the range is smaller, you can be more confident in your data.



WHAT PATTERNS OR TRENDS DO YOU NOTICE IN THE DATA?

Significance

STATISTICAL SIGNIFICANCE IS ANOTHER TERM YOU MAY COME ACROSS IN A LITERATURE REVIEW OR AS YOU ARE ANALYZING DATA.

Statistical significance is often labeled with a p for probability. If p is less than .05, the results are typically considered statistically significant. This means the results are not due to chance alone. So, if you found statistical significance for the effectiveness of a prevention strategy or policy, you could conclude that the intervention was having an impact on behavior, and the differences are not random.

You also may hear the term **practical significance**. At times, results may not be statistically significant based on p values when you analyze data; however, the results are "practically" significant. Practical significance asks whether the effect is large enough to care about. You can use statistical analyses to determine statistical significance and subject-area expertise to assess practical significance.

WHAT PATTERNS OR TRENDS DO YOU NOTICE IN THE DATA?

For Example

Practical Significance

For example, you may have some preliminary data that shows an after-school intervention that builds self-esteem is showing improvements in students' GPAs. You have a pretty small sample size since the program has not been in place for long. You decide to run t-tests comparing GPAs of students who received and did not receive the after-school intervention. Your p value did not indicate statistical significance. However, a large percentage of the students saw increases in GPA after 2 months of being in the program. You would consider this practically significant—it is making a difference in real life, and it makes sense to continue the program and gather more data. The more data you collect, the more likely you are to find statistical significance because you gain what is called statistical power. For more information on statistical power and how to calculate it, refer to the Further Learning resources.

Important Considerations for Interpreting Existing Data

GETTING TO KNOW YOUR EXISTING DATA

WHEN YOU FIRST LOOK AT DATA FROM AN EXISTING SOURCE, IT CAN BE OVERWHELMING TO INTERPRET AT FIRST GLANCE. HERE ARE SOME QUESTIONS TO HELP YOU GET TO KNOW THE DATA.

1. Title & labels: Is the table or chart titled and labeled descriptively and clearly?
2. Time frame: What is the date range(s) for the data being presented?
3. Data source: Do you know where the data came from?
4. Unit(s) of measurement: Do you clearly understand what the metrics in the tables or charts represent?
5. Scales: Are the scales of the graph axes clear and effective?
6. Calculated metric(s): For ratios, rates, and other formulas, do you have a clear understanding of how they are calculated?
7. Dimensions: Are the dimensions or categories used to organize or segment the data clear and meaningful?
8. Filters: Is it clear whether any specific filters have been applied to the data set (e.g., all patients vs new patients)?
9. Sorting: If different values have been sorted or ranked, is it clear what criteria was used?
10. Targets: If goals or targets have been added to the charts, is it clear what they represent?

KEY POINT:

Remember that not all data is equal. Some data is more valid and reliable than others. Examine data for sample size, statistical and practical significance, and how the data compares with other findings.

DRAWING MEANING FROM EXISTING DATA

ONCE YOU HAVE GOTTEN TO KNOW YOUR EXISTING DATA, YOU THEN HAVE TO FIGURE OUT WHAT IT IS TRYING TO TELL YOU—HOW DO YOU INTERPRET DATA THAT YOU DIDN'T COLLECT? HERE ARE SOME QUESTIONS TO ASK THAT WILL HELP YOU INTERPRET EXISTING DATA.

1. Trends: What direction are the trends going (up, down, flat)?
2. Patterns: What repeatable patterns or cycles are you seeing in the data (e.g., seasonality)?
3. Gaps: Are there any obvious gaps or omissions in the dataset?
4. Clusters: Are some values bunched closely together in certain areas?
5. Skewness: Are values noticeably concentrated or skewed more to one side than another?
6. Outliers: Is there a data point that is detached or far removed from the rest of the data?
7. Focus: Has something in the chart or table been emphasized to draw attention to it? Is it obvious why part of the data was highlighted?
8. Noise: Is there any extraneous data included that detracts from the main message of the chart?
9. Logical: Does the data help to answer a specific question? Does the data support a proposed conclusion or argument?

Whether you are working with data you have collected (primary), or you are working with data you have to get to know better (secondary), the important thing is to think through your process of analyzing and interpreting data, so you make accurate and valid conclusions. Remember, the data you report has implications for public programs or policy. When you arrive at conclusions based on appropriate and thorough analysis, you can be confident the results are more accurate and meaningful for those you serve.

APPENDIX

FURTHER LEARNING

The CDC Field Epidemiology Manual: Describing Epidemiologic Data

https://www.cdc.gov/eis/field-epi-manual/chapters/Describing-Epi-Data.html

OpenRefine: A free, open source, powerful tool for working with messy data

http://openrefine.org/

Open Source Epidemiologic Statistics for Public Health

http://www.openepi.com/Menu/OE_Menu.htm

Free Class on Coursera: Introduction to Statistics & Data Analysis in Public Health

https://www.coursera.org/learn/introduction-statistics-data-analysis-public-health

Free SAS Software for University Students

https://www.sas.com/en_us/software/university-edition.html

Statistical Power: What it is, How to Calculate it

https://www.statisticshowto.datasciencecentral.com/statistical-power/

Power and Sample Size .com Free, Online, Easy-to-Use Power and Sample Size Calculators

https://powerandsamplesize.com/

KEY TERMS

Categorical Variables:

data that fits into categories, such as gender or race.

Data Dictionary:

a codebook to help decipher how data values are coded.

Frequency tables:

organizing data values in ascending or descending order, with the number of times each value occurs.

Measures of Central Tendency:

measures used to describe a dataset, including mean, median, mode, and range.

Recoding:

converting data from one code to another.

Standardized:

data put into a common format.

REFERENCES

15. Gray, J., Bounegru, L., & Chambers, L. (2012). The Data Journalism Handbook: How Journalists Can Use Data to Improve the News. Sebastopol, CA: O'Reilly Media, Inc.

16. Regents of the University of Minnesota (2019). What is Data Analysis? Retrieved from https://cyfar.org/what-data-analysis.

17. Regents of the University of Minnesota (2019). Statistical Software. Retrieved from https://cyfar.org/statistical-software.

18. United States Census Bureau (2019). American Fact Finder: Community Facts. Retrieved from https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml?.

19. Regents of the University of Minnesota (2019). Types of Statistical Tests. Retrieved from https://cyfar.org/types-statistical-tests.

20. Colorado Department of Public Health and Environment (2019). Communities That Care Risk and Protective Factor Profile Reports. Retrieved from https://www.colorado.gov/pacific/cdphe/communities-care-risk-and-protective-factor-profile-reports.

6

Telling Your Data Story

DATA STORY

*"Data, I think, is one of the most powerful mechanisms for telling stories. I take a huge pile of data and I try to get it to tell stories."**

Now that you have explored your data, analyzed it, and interpreted it, you can begin to share your findings. You might find this to be the most challenging part of becoming data literate, but also the most meaningful. You will learn how to disseminate evidence-based information to make positive changes in the community. You need to think critically, not just about the data itself, but about the stakeholders who will be using it—their communication styles, characteristics, needs, and values. Learning how to promote and communicate your findings is an art, as well as a science. If you can master this skill, you will be a powerful asset to your community. You are about to gain some higher-level skills on your data citizen journey, but it is a realistic goal with the tips and tools you will cover in this chapter. Although you can think of communicating results as sharing a story, remember that the story is based on facts and science.

To begin telling your data story, review the chronology using key events and cases. You can tell the story in multiple ways. The stories that emerge from data can be compelling and promote public health or prevention messages. However, you will need to continually evaluate and improve the data story, as well as adapt it for each audience you are engaging.

*Steven Levitt

Dissemination Planning

A FIRST STEP TO TELLING YOUR DATA STORY IS TO THINK OF DISSEMINATION—how will you disseminate the results to multiple partners or stakeholders? What methods best suit their needs? A good way to plan out your communication of the results is to establish a dissemination plan.

A dissemination plan outlines how you will disseminate research findings or data products to stakeholders. It should be a part of the process you discuss with your data workgroup or community coalition early on; it will help you get the message out whether it is products that contain results of research, epidemiological findings, or a successful public health or prevention intervention.

EFFECTIVE STORYTELLERS

For data citizens to be effective storytellers, your communication of the results must be:

- INTENTIONAL
- PURPOSEFUL
- VISUALLY AESTHETICALLY PLEASING
- INDEPENDENTLY INTERPRETABLE

KEY POINT:

Dissemination plans are extremely important to encourage buy-in of stakeholders and ensure your data will be useful to those you are serving. Think about dissemination before even collecting and analyzing data.

SOME KEY QUESTIONS TO ADDRESS IN A DISSEMINATION PLAN

1. What findings are you going to disseminate? What products will you create?

2. Who will use the data in practice? How will you reach them?

3. Who are your dissemination partners? How can they help you reach your audience and get your message out?

4. How will you communicate your results?

5. How will you measure if stakeholders are accessing your products?

Communication Plan

Background

Faucibus in ornare quam viverra orci sagittis eu volutpat. Ultrices gravida dictum fusce ut placerat orci. Viverra nam libero justo laoreet sit amet cursus sit. Purus ut faucibus pulvinar elementum integer enim neque. Pretium viverra suspendisse potenti nullam ac tortor vitae. Sed faucibus turpis in eu mi bibendum neque. Nunc eget lorem dolor sed viverra ipsum nunc aliquet. Ut aliquam purus sit amet luctus venenatis. Magna fringilla urna porttitor rhoncus dolor purus non. Condimentum id venenatis a condimentum vitae sapien pellentesque habitant.

Strategy

Phasellus vestibulum lorem sed risus ultricies tristique nulla aliquet enim. Fermentum et sollicitudin ac orci phasellus egestas tellus rutrum tellus. Aliquam ultrices sagittis orci a scelerisque purus semper eget duis. Nullam ac tortor vitae purus faucibus ornare. Vel quam elementum pulvinar etiam non quam lacus suspendisse. Suscipit adipiscing bibendum est ultricies integer. Nascetur ridiculus mus mauris vitae ultricies leo integer. Duis ultricies lacus sed turpis tincidunt id. Dictum non consectetur a erat nam at lectus urna. Suspendisse ultrices gravida dictum fusce ut placerat orci nulla pellentesque.

Nulla facilisi nullam vehicula ipsum. Nunc id cursus metus aliquam eleifend. Vitae auctor eu augue ut. Cras sed felis eget velit aliquet sagittis. Vel fringilla est ullamcorper eget nulla facilisi. Tellus mauris a diam maecenas sed enim ut sem. Rhoncus urna neque viverra justo nec ultrices dui sapien. Et netus et malesuada fames ac turpis egestas maecenas pharetra. Venenatis lectus magna fringilla urna porttitor. Ut venenatis tellus in metus vulputate eu. Viverra nam libero justo laoreet sit. Facilisi cras fermentum odio eu feugiat.

Deliverables

Product	Audience	Delivery	Target Date	Lead Contributors	Status

Telling Your Story with Data

You want to ensure your data story is useful and actionable by stakeholders.

AS A DATA CITIZEN, YOU CAN HELP MAKE DATA ACTIONABLE BY USING NEW AND INNOVATIVE WAYS TO ILLUSTRATE DATA FINDINGS.

Visualization of your findings can increase understanding and message retention among your intended audiences. A reporting approach that can prompt your audience to use your data story is **action-oriented reporting.**[21] To adopt an action-oriented approach to data storytelling:

1. Use reports that are purposely shorter and incorporate action steps stakeholders can implement.

2. Call attention to key findings.

3. Make visual displays of data that are simple to understand.

4. Tailor your reports to the audience you are trying to reach.

5. Use a variety of written, verbal, and electronic forms of reporting, depending on your intended users.

BEFORE YOU DEVELOP A DATA PRODUCT, ASK YOURSELF:

1. Who is the intended audience?
2. What is the key message?
3. What data and information should be emphasized?

"A well-designed action-oriented report saves time by calling attention to important findings and possible next steps."

- Centers for Disease Control and Prevention

Telling Your Story With Data

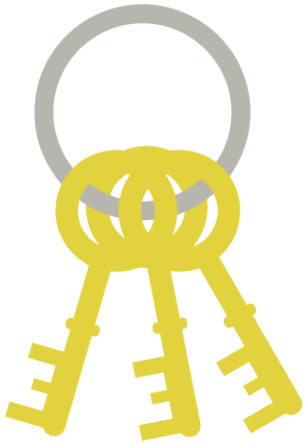
Next, you will review a few examples of some of the easiest forms of written, verbal, and electronic reporting that promote innovation and creativity. Below are some questions to critically think about before developing any product.

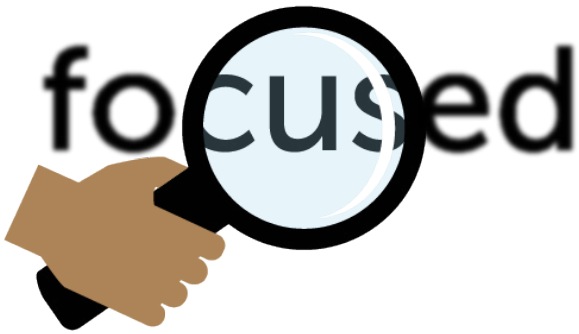
WRITTEN PRODUCTS

Infographics

Infographics are visual representations of data, information, or knowledge. You can incorporate graphics, data, and text in a creative way that presents complex information or concepts clearly and quickly. You can engage the viewer with take-home messages that are highlighted in a visually appealing manner.

Even if you have very limited resources and cannot purchase additional software to create infographics, you can use Microsoft Word to create appealing infographics. In addition, there are several free online resources available that can be used to create infographics, such as Piktochart, Easel.ly, and Venngage, to name a few.





KEY POINT:

Before creating a data product, remember to ask yourself:

- Who is the intended audience?
- What is the key message?
- What data and information should be emphasized?

WRITTEN PRODUCTS

Infographics

The infographic below was created using Piktochart and displays an appealing way to create an evaluation report from a behavioral health conference.

1. How satisfied were you with the overall quality of this conference?

2. I expect this event to benefit me professionally.

BEFORE USING AN INFOGRAPHIC TO SHARE YOUR DATA FINDINGS, FOLLOW THESE TIPS TO BE EFFECTIVE:

1. Make sure your infographic is short and simple.

2. Tell a consistent and focused story about your key findings and conclusions.

3. Use one to three key concepts to tell your story, as opposed to a dozen.

4. Create a draft outline to make sure the concepts can be clearly laid out.

5. Choose attractive color schemes and graphics to garner interest among your viewers.

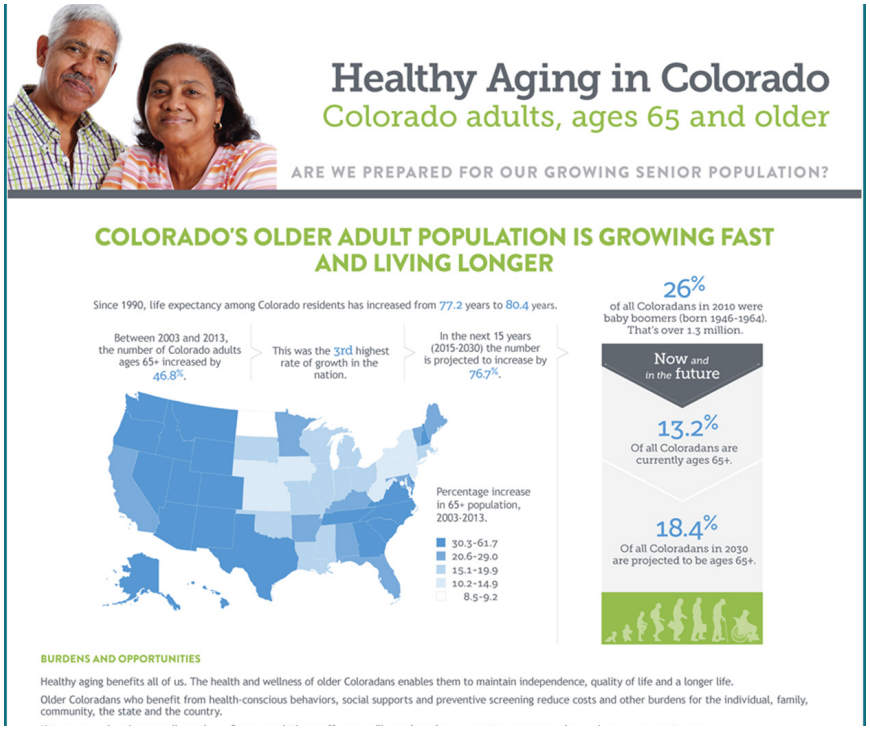
6. Proofread and edit your draft infographic to ensure clarity.

WRITTEN PRODUCTS

Infographics

Here is an example of storytelling using an appealing infographic from the Colorado Department of Public Health and Environment on Healthy Aging.[22]

THIS INFOGRAPHIC USES LARGER TEXT TO HIGHLIGHT KEY MESSAGES AND STATISTICS. NOTICE HOW THE PERCENTAGES GRAB YOUR ATTENTION BEFORE YOU READ THE TEXT BELOW.

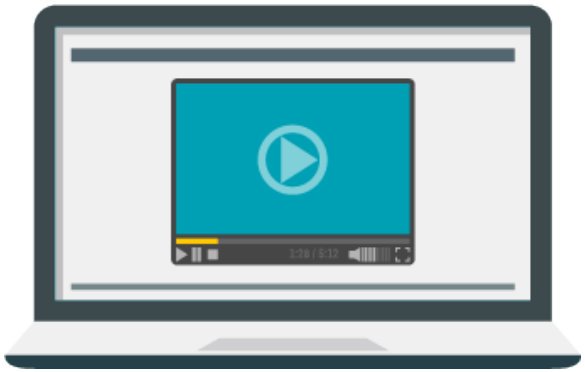


Color Accessible

Not Color Accessible

KEY POINT:

Data reporting mediums include written, verbal, and electronic forms. Some ways to be innovative and creative with each of these forms is to use infographics, video, and interactive dashboards.



Here are some tips for creating effective data storytelling videos that "show," not just "tell."

1. Be clear about the purpose and obtain permission from any participants before videotaping.

2. Balance qualitative and quantitative information.

3. Consider the intended audience when determining the ideal video length.

4. Post the video on both your website and on YouTube to help expand your viewing audience.

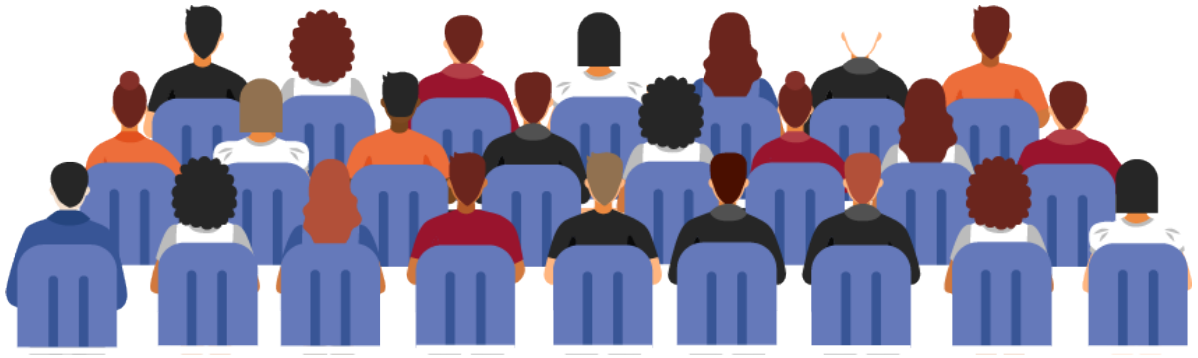
VERBAL PRODUCTS

Videos

Video is another innovative tool for enhancing communication of your data findings. Video reporting can convey a lot of information in a short amount of time and can be more memorable than posters and verbal presentations for many users. Use of video can reach more people than text and is an innovative way to affect an audience's attitudes and opinions about your program, thus encouraging action. Video has a broad reach and would be beneficial for program staff and other users, such as public health practitioners, policymakers, and evaluators.

VIDEO IS BECOMING MORE ACCESSIBLE IN TERMS OF COST AND RESOURCES. YOU MAY ONLY NEED A CAMERA, TRIPOD, AND MICROPHONE, OR JUST YOUR PHONE, DEPENDING ON YOUR DESIRED PRODUCT!

A great example of the use of video to illustrate social determinants of health in Colorado can be found on the Children's Hospital Colorado website (childrenscolorado.org/community/community-health). You will notice the video is short, keeps your attention, uses simple graphics, and is posted both on the Children's Hospital Colorado website and YouTube. Also, notice the title ("What is Population Health?") is very simple and concise.

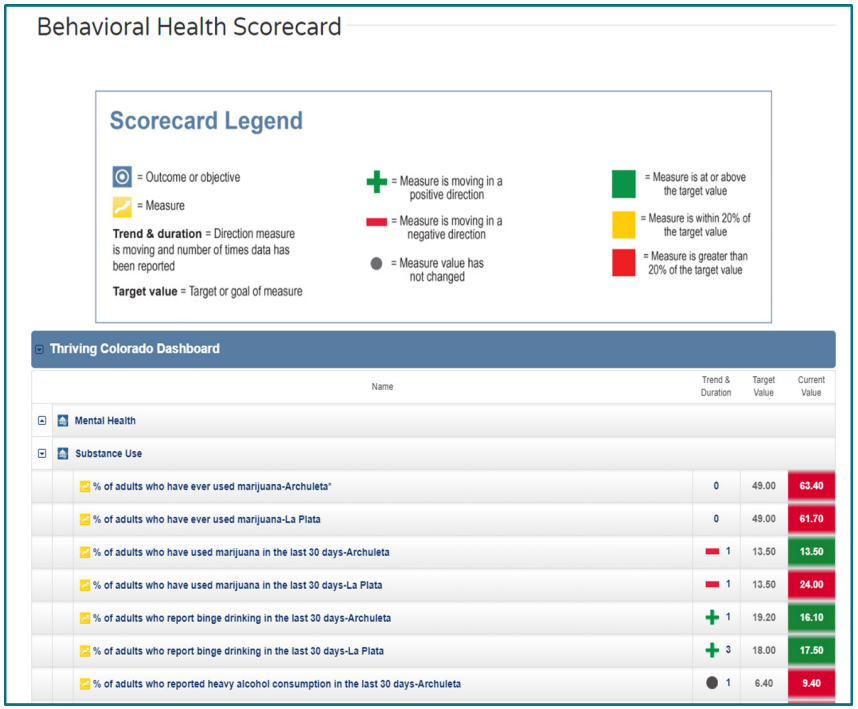


ELECTRONIC PRODUCTS

Interactive Webpages

INTERACTIVE WEBPAGES RANGE IN COMPLEXITY, BUT THEY ARE BECOMING INCREASINGLY COMMON.

While a static webpage will probably get your point across, it may not effectively grab and retain your reader's attention. Interactive webpages greatly enhance the user experience and may inspire users to employ your evaluation results. Here is an example of a data dashboard on the San Juan Basin Public Health website that serves La Plata and Archuleta counties. By clicking on the behavioral health scorecard, dashboard visitors can quickly see if mental health and substance use/misuse measures are trending in a positive or negative direction or remaining stable.[23]

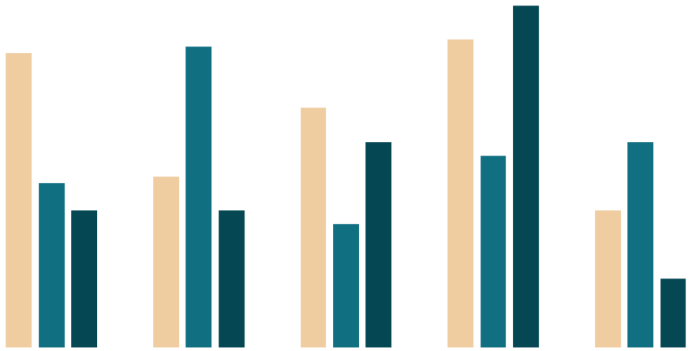


INTERACTIVE WEBPAGES REPRESENT ANOTHER BROAD-REACHING FORM OF DATA: EVALUATION REPORTING. Program staff, public health practitioners, policymakers, evaluators, and others would benefit from this reporting method. Many elements go into interactive webpages, but the combination of design and functionality is paramount to making your webpage appealing. If you decide that this medium is useful for your data reporting, make sure there is a designated content manager to ensure that the information is accurate. Also, consider what data security measures are required for your organization. Lastly, this may be the most budget-intensive reporting medium presented in this chapter, so, make sure that you have the resources and budget within your organization to create, maintain, and disseminate your interactive webpage. Tableau Public is a free software option that you can explore for creating data dashboards. You can refer to the tutorials and links for Tableau in the Further Learning section.

COMMON CHARTS, TABLES, AND MAPS USED IN REPORTING

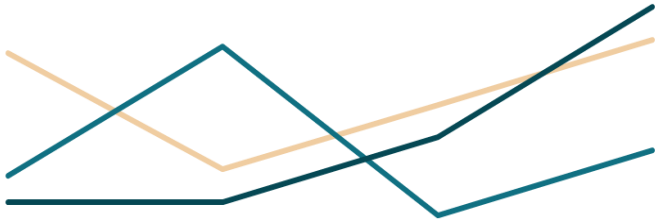
Bar Chart

BAR CHARTS ARE OFTEN USED FOR DISPLAYING QUALITATIVE OR CATEGORICAL DATA (e.g., age group, race/ethnicity), although they can also be used for quantitative data if the number of unique scores in the data set is not large. For example, if you want to display the number of children with different numbers of Adverse Childhood Experiences (ACEs), you can use a bar chart to group the data and display the frequency of children with 1, 2, 3, 4 or 5+ ACEs (see above). If you do not group all scores above 5, the bar chart would get too large, and counts in higher ACE scores would be very small.



Line Graph

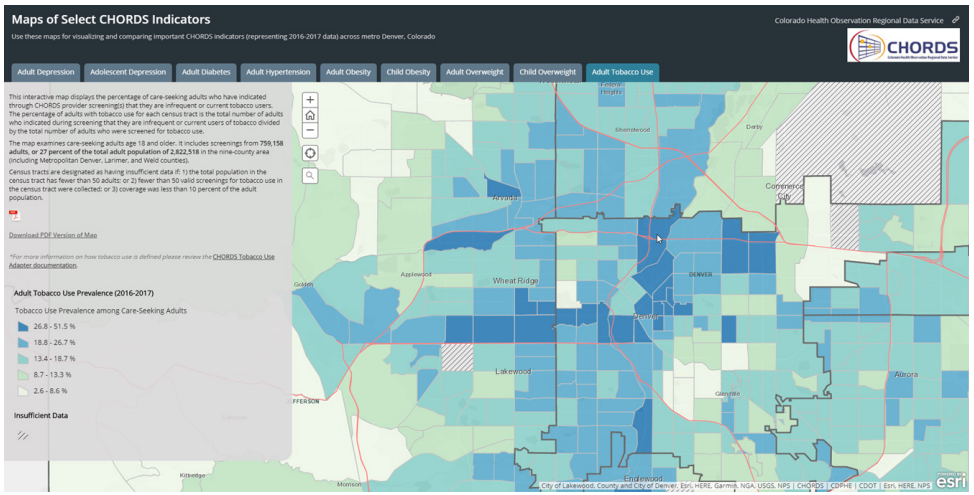
A LINE GRAPH IS COMMONLY USED TO DISPLAY CHANGE OVER TIME as a series of data points connected by straight line segments on two axes. The line graph, therefore, helps to determine the relationship between two sets of values, with one data set always being dependent on the other set. Line graphs work well to display trends over time, such as prevalence or incidence rates for substance use or mental health issues, or they could be used to look at substance use consumption and consequences over time.



COMMON CHARTS, TABLES, AND MAPS USED IN REPORTING

Pie Chart

A PIE CHART IS A DIVIDED CIRCLE, IN WHICH EACH SLICE OF THE PIE REPRESENTS A PART OF THE WHOLE. The categories that each slice represents are mutually exclusive and exhaustive. Data with negative values cannot be displayed as a pie chart. Pie charts are helpful when you have demographic data or other categorical data where respondents cannot mark more than one item. Pie charts should not be used when you have a large number of categories to be displayed. For example, you could use a pie chart to display the percentage of students in each grade level that responded to your middle school survey.



Interactive Mapping

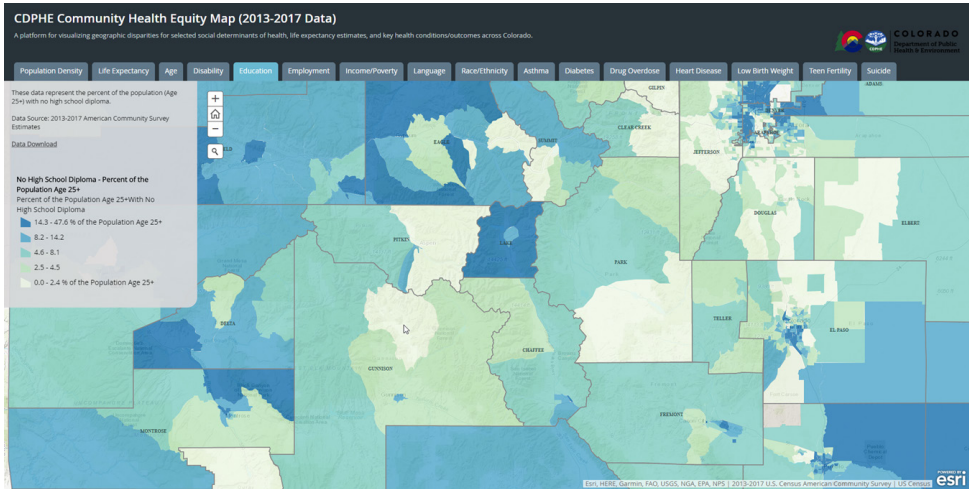
Interactive mapping involves using maps that allow zooming in and out, panning around, identifying specific features, querying underlying data by topic, or a specific indicator (e.g., socioeconomic status), generating reports, and other means of using or visualizing select information in the map. Above is an illustration of the Denver metro area CHORDS interactive map showing different rates of adult tobacco use.[24] THE MAP ALLOWS YOU TO ZOOM IN AND OUT, AND YOU CAN ALSO CLICK ON DIFFERENT DATA INDICATORS AT THE TOP TO CHANGE THE MAP.

COMMON CHARTS, TABLES, AND MAPS USED IN REPORTING

GIS Mapping

Maps will typically display one data variable or indicator, often using color-coding to indicate the density, frequency, or percentage in a given region, allowing quick comparison between regions.

Color-coding can also be used to show changes over time (e.g., red for regions where an indicator got worse between two specific years, blue for areas where improvement was seen, and so on). Other information can be overlaid on the map, for example, specific sites (e.g., hospitals or distribution centers) and text.



HERE IS AN EXAMPLE OF GIS MAPPING FROM THE COLORADO DEPARTMENT OF PUBLIC HEALTH AND ENVIRONMENT COMMUNITY HEALTH EQUITY MAP.[25] By selecting Education, site visitors can view by geographic area where you are likely to see more individuals over 25 who do not have a high school diploma. Notice the color-coding to the left that indicates the percentage. As the shaded areas get darker, the percentages increase.

KEY POINT:

Data visualization does not have to be expensive or complicated. Using Microsoft Excel for tables and graphs, Tableau for maps and dashboards, Piktochart for infographics, and even your phone for videos are all affordable and relatively easy ways to be creative and innovative with your data stories.

APPENDIX

KEY TERMS

Benchmark: to set a target or standard to which other scores are compared.

Confidence Interval: a range of plausible values for an unknown parameter.

Incidence: the count of incident cases over time in a population.

Incident: new events among a population.

Prevalent: an existing event at a specific point in time.

Practical Significance: to determine whether the result is useful.

Rate: a measure of the frequency with which an event occurs over a period.

Statistical Significance: a number that expresses the probability that the result or relationship is caused by something other than chance.

RESOURCES

Betterevaluation.org Blog
https://www.betterevaluation.org/

Free Data Visualization Software Resources
https://piktochart.com/
https://venngage.com/
https://infogram.com/
https://public.tableau.com/en-us/s/

Free Training from Tableau on a variety of topics, including basic, filtering and sorting data, calculations, and mapping
https://www.tableau.com/learn/training

Colorado Department of Public Health and Environment Open Data
https://data-cdphe.opendata.arcgis.com/

FURTHER LEARNING

David McCandless TED Talk. David turns complex data sets (like worldwide military spending, media buzz, Facebook status updates) into beautiful, simple diagrams that tease out unseen patterns and connections.
https://www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization/up-next

Coursera Data Visualization Course (offered by University of Illinois at Urbana-Champaign)
https://www.coursera.org/learn/datavisualization

Coursera Data Visualization and Communication with Tableau Course (offered by Duke University)
https://www.coursera.org/lecture/analytics-tableau/delivering-your-data-story-Tledh

The Extreme Presentation(tm) Method: Extremely effective communication of complex information
https://extremepresentation.typepad.com/blog/

Storytelling With Data
http://www.storytellingwithdata.com/

Using Microsoft Excel to Create Charts and Tables
https://support.office.com/en-us/article/video-create-a-chart-4d95c6a5-42d2-4cfc-aede-0ebf01d409a8
https://support.office.com/en-us/article/video-create-pie-bar-and-line-charts-ce4beacc-7e25-4574-a8b9-a865dbb3e3d2?ui=en-US&rs=en-US&ad=US
https://support.office.com/en-us/article/video-create-a-table-bf0ce08b-d012-42ec-8ecf-a2259c9faf3f

Colorado College: GIS Video Tutorials
https://www.coloradocollege.edu/academics/dept/sociology/people/gis-video-tutorials.html

Tableau: 5 Best Practices for Telling Great Stories
https://www.tableau.com/5_best_practices

REFERENCES

21. Centers for Disease Control and Prevention (2013). Evaluation Reporting: A Guide to Help Ensure Use of Evaluation Findings. Atlanta, GA: US Dept of Health and Human Services.

22. Colorado Department of Public Health and Environment (2019). The Health of Colorado's Older Adult Population: Data Infographic. Retrieved from https://www.cohealthdata.dphe.state.co.us/chd/Age/Healthy-Aging-in-Colorado-Infographic.html.

23. San Juan Basin Public Health (2019). Thriving Colorado Dashboard. Retrieved from http://sjbpublichealth.org/tcdashboard.

24. Colorado Health Observation Regional Data Service (2019). Maps of Select CHORDS Indicators. Retrieved from https://www.cohealthmaps.dphe.state.co.us/chords_maps/chords_webmap/.

25. Colorado Department of Public Health and Environment (2019). CDPHE Community Health Equity Map (2013-2017 Data). Retrieved from http://www.cohealthmaps.dphe.state.co.us/cdphe_community_health_equity_map/

7

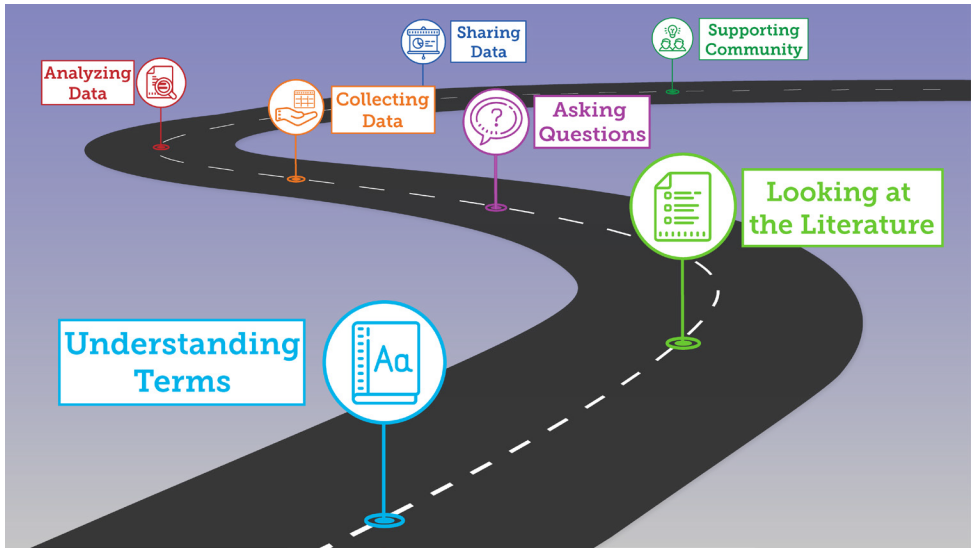
Sustaining Your Knowledge

SUSTAINING YOUR KNOWLEDGE

*"Data is a precious thing and will last longer than the systems themselves."**

You are now at the end of this piece of your data citizen journey. Throughout this manual, you have gained skills and knowledge in data literacy foundations, data collection, data management, analysis, and evaluation. To ensure your continued learning and reflection, in this chapter, you will focus on the sustainability of your data literacy professional development, as well as the systems and data collection mechanisms you have put into place. You will learn some simple steps you can take to continue your data citizen journey.

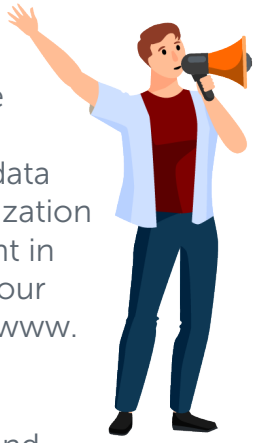
By developing a sustainable learning plan for yourself or your agency staff, you will keep your data literacy skills fresh, especially as new ways of communicating and analyzing data are developed all the time. When you get back to your day-to-day work, sometimes learning is pushed to the side and is not prioritized. **GO BACK TO YOUR ORGANIZATION EMPOWERED WITH DATA LITERACY KNOWLEDGE, BE A CHAMPION TO OTHERS, AND HONOR YOUR LEARNING BY CONTINUING TO IMPROVE.**



*Tim Berners-Lee

Steps You Can Take

HERE ARE SOME WAYS YOU CAN WORK TO ACCOMPLISH THESE GOALS:



- **Develop a Sustainable Learning Plan.** Think about how you can bring the knowledge from this manual to your organization. Utilize the activities at coalition meetings or conduct in-person training on one or more of the data literacy topics. Prioritize learning opportunities for yourself or your organization to continue to build capacity around data literacy. Use the self-assessment in Chapter 1 to determine what types of resources may be most helpful in your learning or staff development. The self-assessment can be accessed via: www.thedataliteracyproject.org/assessment.

- **Review Data Literacy Learning and Progress.** Update learning materials and improve them as you learn more about data literacy. Supervisors can also add data literacy as a skill for improvement in performance management processes.

- **Become a Data Champion or Mentor.** You can use this manual to provide support and assistance to others in your community and share the resources you find helpful.

- **Build a Data Literacy Library.** Organize training materials and toolkits for staff and community members to further their learning.

- **Create an Online Data Literacy Learning Community.** Form a learning community or a group of people committed to sharing information and cooperating around a particular goal to improve outcomes for the community. In your online learning community, you can share resources, trainings, and ideas about data literacy and build a network of people and resources.

KEY POINT:

Expand your data literacy knowledge by developing a learning plan, bringing your knowledge to others as a mentor, and networking with others around data literacy (e.g., data literacy learning community).

Steps You Can Take

You cannot rush change in an organization, and it is the same for promoting and sustaining data literacy. Start simply by promoting data presentations at meetings or providing data reports—these are great ways to improve the culture of an organization around data. You can even highlight current or past data-based decision-making your community achieved with success. You can take this manual to your next meeting or give an overview of the content.

Highlight the benefits of becoming data-literate instead of stressing more training, both professionally and within your organization. Begin weaving this manual's content into your organizational processes slowly to start those system-level changes.

HERE ARE SOME WAYS YOU CAN WORK TO ACCOMPLISH THESE GOALS:

- **Maintain and Track Local, State, and National Data Sources.** Continue to nurture data partnerships for data access in the future and document how data sources can be used.
- **Document and Maintain Data Management Plans.** Use the template in the Further Learning section at the end of this chapter to create your data management plan.
- **Celebrate Success with Data Literacy.** Highlight community members or staff on social media and provide a story about their success with data literacy.

Steps You Can Take

HERE ARE SOME WAYS YOU CAN WORK TO ACCOMPLISH THESE GOALS:

- **Establish a Data Workgroup for Product Development.** Establish a data workgroup at your organization that leads product development. This group can build community capacity by involving stakeholders in data literacy activities and asking for feedback on the effectiveness of data products.

- **Encourage Leadership to Show Support.** Obtain support from leadership to change systems at your organization. To build enthusiasm from leadership, create data products that show the value of your learning and how they can document and improve outcomes for your organization or community.

- **Institutionalize Data Literacy into Organizational Processes.** Include aspects of data literacy in job announcements, onboarding, professional development opportunities, performance management processes, and even staff meetings. Ask for presentations at staff meetings around different data literacy topics. Emphasize the importance of data in all aspects of the organization.

KEY POINT:

Promote and institutionalize data literacy within your organization, garner support from leadership, demonstrate the utility and strength of data to improve programs, and include data literacy in work processes and systems throughout your organization.

NEXT SETPS

Below is an action plan to outline some steps you or your organization are going to take to begin implementing your data literacy knowledge. Think about the next 30 days. What are some things you can do to bring this knowledge back to your community? Complete this template and take it to your next meeting with your coalition, organization, or director.

Goal	Strategy	Who?	Task	Materials or Resources	Due Date

It is your turn to take the next steps! *You have some data literacy knowledge and tools, now go out and be data citizens! As a data citizen, you can begin advocating for your learning and the learning of others around data literacy. When others see you have outlined steps for action, they know you mean business!*

KEY POINT:

Start simple and slow with systems change. Weave data literacy concepts throughout existing processes, so staff begin to see the benefits. Provide examples of data-based decision-making that have benefited the organization in the past.

RESOURCES

Data Management Plan Tool: Build Your Data Management Plan.
https://dmp.cdlib.org/

Qlik Course: A Culture of Data Literacy
https://qcc.qlik.com/course/view.php?id=723

Data literacy: A critical skill for the 21st century
https://www.tableau.com/about/blog/2018/9/data-literacy-critical-skill-21st-century-94221