Data Management and Visualization Training

Stanford Center for Open Solutions, National Center for Ecological Analysis and Synthesis

2021-12-17

Contents

Chapter 1

Introduction to Course

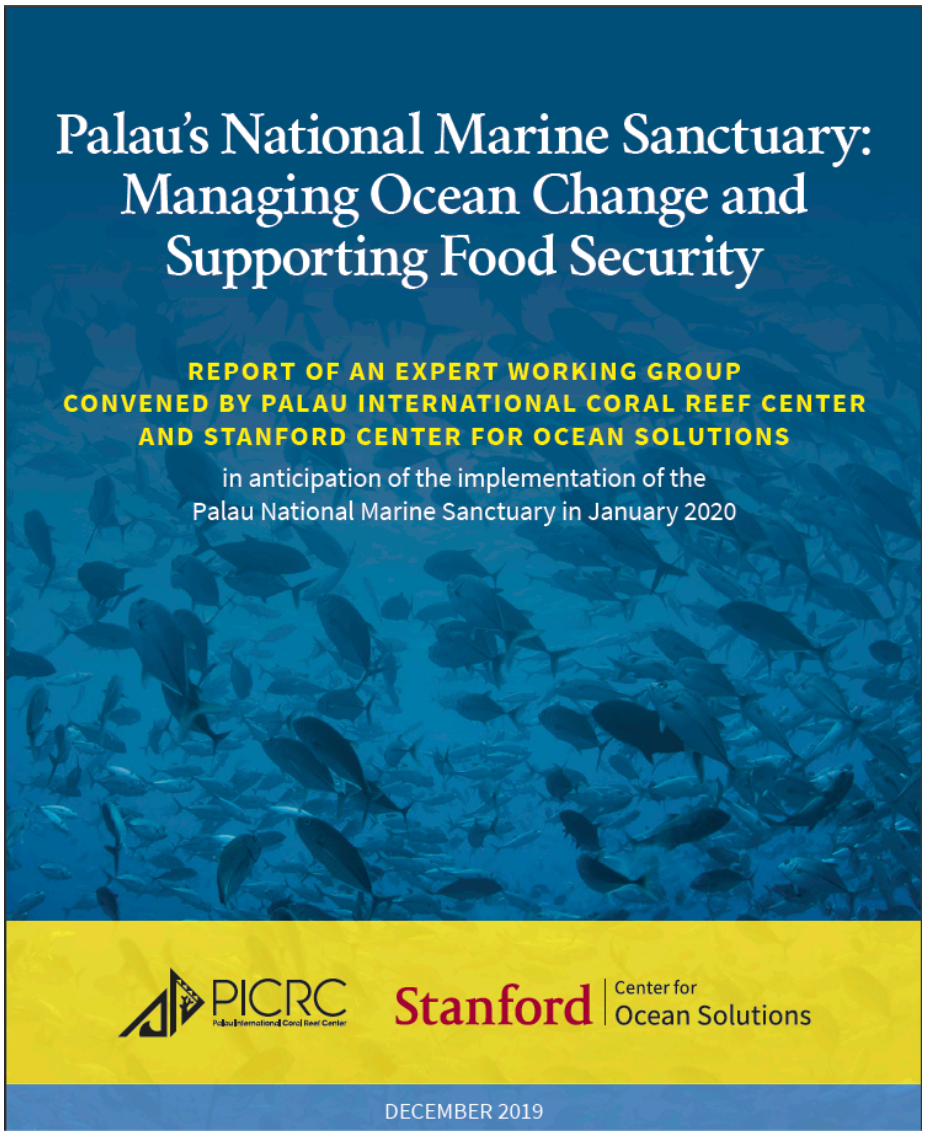
In this Session

1. Training Overview
2. Structural Approach for Using the Material

1.1 Training Overview

1. ***Relevance:*** As environmental issues increase in scale and scope, there will be an increasing need to "tell the story" of monitoring and conservation efforts. To address this need, data enthusiasts - such as you - have an opportunity to expand their knowledge base in working with complex data sets to creatively convey impactful, data-driven messages to a variety of audiences. This compilation of training materials is intended to provide an accessible, digestible resource for anyone interested in any aspect of data management or visualization.

2. ***Background:*** This compilation of data management and visualization training materials stems from a multi-year engagement amongst the Palau International Coral Reef Center (PICRC), the Stanford Center for Ocean Solutions (COS), and the National Center for Ecological Analysis and Synthesis (NCEAS) through financial support by Future Earth. The funding supported a collaborative working group requested by former Palauan President, Tommy Remengesau, Jr. to investigate major implications of the full implementation of the Palau National Marine Sanctuary and compile a report of the findings (see PICRC page or COS page). One key lesson from that experience was the importance of enhancing the capacity for managing and visualizing data driven messages to more effectively communicate significant management findings.

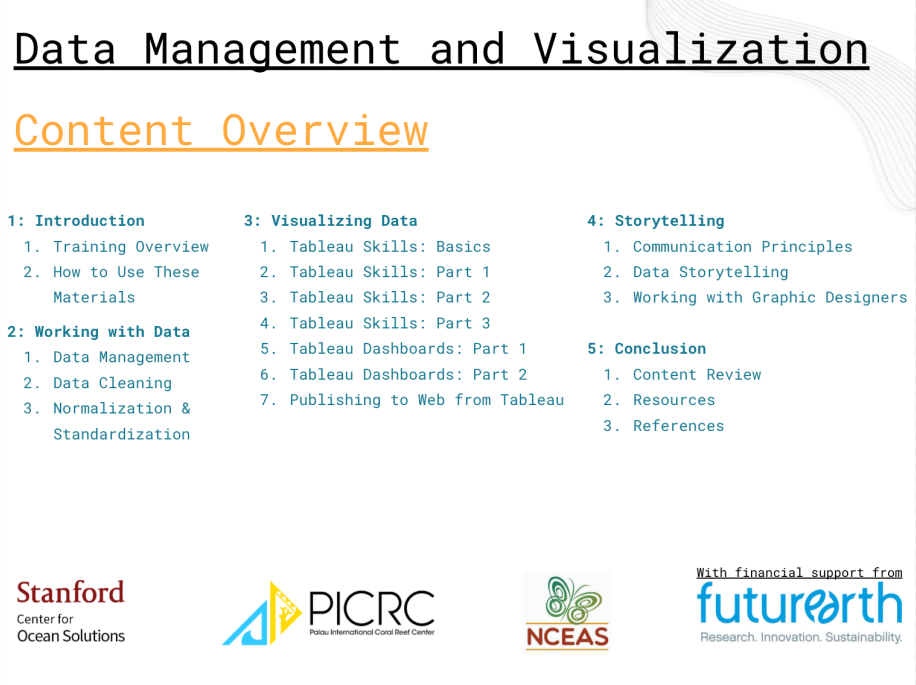


3. ***Content:*** In this training package, you will learn about three core elements of data management and visualization. There are five modules, including this introductory module, each broken into several sessions. The core focus of this training is in Modules 2 and 3 with additional brief resources for storytelling or communication needs in Module 4.

    - *Module 2: Working with Data* - including management of data resources, cleaning datasets, and normalizing & standardizing data sets and databases.
    - *Module 3: Visualizing Data* - including a focus on using Tableau

software to build or enhance visualization skills and resources.

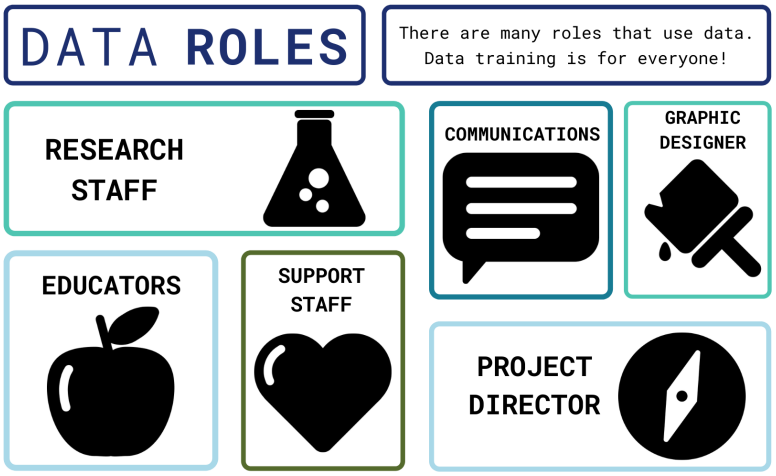
- *Module 4: Storytelling* - including core principles in overall data communication, lessons in building and telling stories using your data sets, and working with external partners on visuals (such as graphic designers).
- The concluding module (*Module 5*) includes a collection of final messages and resources to aid you in your path forward.



4. ***Intended audience:*** In recognizing the wide range of roles and responsibilities individuals may have with respect to data use within an organization, this training is designed to be completed either in a sequence of all material from beginning to end or compartmentalized into modules and sessions to focus on specific skills. Whether you are a data novice interested in learning about all aspects of management and visualization, a data nerd who wants to learn more about visualization and communication, or a communications wizard who wants to be more familiar with data details or skills - this training package includes material to advance your understanding. This content is intended to equip you to better utilize and communicate data stories. In preparation for this course, it may help to briefly consider which data role(s) from the non-exhaustive list below best describes your interests and which lessons best suit your position and needs.

5. ***Required tools:*** To take this training you will need:

- An internet connection, either to download the training booklet and supplementary documents or to view the training through a web browser. Watching the videos in Module 3 on Tableau requires an internet connection.
- Access to the software Tableau for Module 3. Details on how to install Tableau are in Module 3, Session 1: Tableau Basics and this website provides details on the necessary operating system requirements to install Tableau.
- The example data sets for exercises in Modules 2 and 3 are provided as Microsoft Excel files.



5. ***Authors:*** This Data Management and Visualization Training package is a collaborative effort amongst researchers and staff from PICRC, NCEAS, and COS. The core contributors to the existing content are recognized below.

1.2 How to Use These Materials

1. ***Overview:*** The content in the three core modules of this training package targets an audience of researchers or communicators seeking to build their skill set. Each module contains links to additional resources for those that are interested in deeper topical dives.

2. ***Materials:*** Some sessions include distilled handouts or briefs as well as short exercises to help you practice the skills while you are developing them.

    - Briefs: Documents that can be utilized as a standalone resource - compiling core messages from a session.

- Exercises: Self-guided practical work to facilitate learning and enable exploration from a more conceptual framework.
- Video Tutorials: Recorded lessons (primarily in Module 4: Tableau) providing stepwise guidance to the interface and content.

3. ***Iteration:*** As previously mentioned, the content in this training package is intended to be revisited on an iterative basis as needed. You can proceed through the content in sequence or focus on the modules or sessions of most interest to you. Regardless of your specific data-related role in an organization, there are opportunities to gain greater mastery through iterative practice and collaborative learning.

4. ***Conclusion:*** The concluding module (Module 5) includes a compilation of resources for "at-a-glance" skill reminders and references for further investigation.

Chapter 2

Working with Data

2.1 Data Management

In this session

1. Learning Objectives
2. Value Statement
3. File naming
4. Creating a Working Folder Structure
5. Metadata
6. Version control
7. Summary

2.1.1 Data Management: Learning objectives

1. Learn good practices to store, name and keep track of different versions of your data files.

2. Understand the importance of metadata and use a template to create your own format.

2.1.2 Data Management: Value statement

1. Learning how to properly store, name, and version control your files is the basis to developing systematic procedures to analyze and visualize data.

2. Creating metadata files is essential to capturing key information about the context and content of a data file. Every data file should be accompanied by its metadata, which provides data about data. By doing this, you can enable others to better understand your data and work more efficiently.

2.1.3 File naming

1. While there is not a unique way to properly name your files, it is important to be consistent within your organization to help your colleagues recognize which data each file contains. Some of the best practices to do this include:
    1. Use descriptive names that would make sense to any reader. For example, don't use acronyms or abbreviations.
    2. Avoid the use of special characters, such as !@#$%^&*() and others, these are sometimes not supported by some software.
    3. Make consistent use of capitalization. Are you going to use capital letters in every word? Maybe only at the beginning of the column name? For example: My_File1.xlsx vs. my_file1.xslx. While none of them is wrong, it is good to choose one format and stick to it
    4. Use underscores instead of blank spaces, just as in the example above.
    5. If you want to keep track of versions, don't use endings such as "final", "v5", etc. Instead, we suggest that you either use the date in format YYYYMMDD, or use a version control system, as we will see in a later section.
    6. Don't start a file name with a number, though it is fine to use numbers elsewhere in the name, as in the examples above.

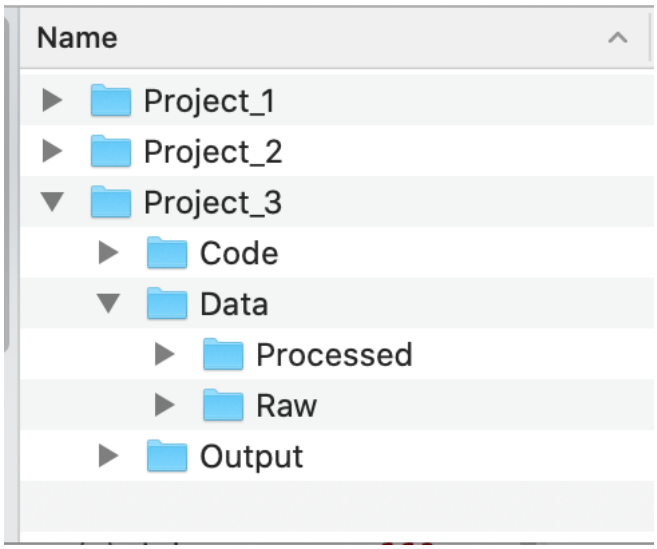
2.1.4 Creating a working folder structure

1. It is important to create a folder structure that will allow you and your colleagues to easily organize and retrieve any files that you are working with.

2. There are some key principles to creating an efficient folder structure:

    1. Store your raw data in its own folder. Every time that you get new data, the first thing you should do is to create a backup by adding it to your raw data folder. This will allow you to always be able to come back to the original file and look for processing errors.
    2. Create a separate folder for "working" files. These are the files where you will change column names, clean data errors and merge multiple files when necessary. For example, if your team performs annual assessments of reef fish biomass, you may want to store separate files for each year, but combine them in a "master" file, which should be located in your "working" files folder.
    3. Having a consistent file structure across projects, such as folders for script outputs or figures, can make it easier for colleagues and collaborators to understand where to find the files they need.

3. In the example file structure below, each project has its own folder and within each project folder, there are folders for code, data, and output files. The Code folder could store scripts for data cleaning or analysis, the Data folder has subfolders for the raw dataset and a version that has been tidied, and the Output folder could store plots or other results.



2.1.5 Metadata

1. Metadata are documentation describing the content, context, and structure of data to enable future interpretation and reuse of the data. Generally, metadata describe who collected the data, what data were collected, when and where they were collected, and why they were collected so that anyone who looks at the data can understand what they mean.

2. Along with this class, we provided you with a metadata template, downloadable here. Please notice that this template can be adjusted to reflect your organization's needs, therefore some of the requested information might change accordingly. Below, you will find an example of a filled out metadata template for the data that we will be using in Module 3.

3. Make sure to always share the metadata along with your data files. Some recommendations to do this include:

    1. If working in Excel, create a new tab named "metadata" and put the information there
    2. If working in other data formats, create a new file with the metadata, store it in the same folder, and name it exactly the same as the data file, adding "_metadata" at the end.

	A	B
1	**Metadata template**	
2	**Dataset name:**	*Palau Island Sites - Reef Transects*
3	**Date of last update:**	*02/09/2021 (DD/MM/YYYY)*
4	**Project title:**	*Palau Reef Survey 2021*
5	**Abstract:**	*This dataset is of preliminary data for our 2021 Palau Reef Health Survey*
6	**Key words:**	*Palau, Reefs, Marine Health*
7	**Lead investigator:**	*Singe Hideos*
8	**Contact information:**	*singe@picrc.org*
9	**Data Manager:**	*Meli Tuisova*
10	**Usage Rights & Copyrights:**	*This dataset is for public use within the nation of Palau, for international use permission must be granted by PICRC administrators*
11	**How to cite (Authors in correct order):**	*Hideos, 2021, 2021, Palau International Coral Reef Center, Version 1, www.picrc.org/2021-reef-transects*
12	**Geographic Area:**	*Palau, Palau coral reefs*
13	**Coordinate system:**	*Latitude-longitude*
14	**Collection dates:**	*04-01-2016 to 02/09/2021 (Present)*
15	**Methods of collection:**	*Transects*
16	**Additional Notes:**	*This data was collected in collaboration with local fisheries. We appreciate all our local partners for helping us complete these transects.*
17		
18	**Tab name:**	*Fish populations*
19	**Fields:**	**Description:**
20	Fish name	Name/species of the specified fish
21	Number of fish	Average number of this specified fish recorded over one transect
22	...	...
23		
24		
25	**Related resources:**	www.picrc.org, www.oceanmarinehealth.org, www.palauoceanwarriors.org
26	*(Indicate resources that have used*	

2.1.6 Version control

1. Version control is the process of tracking and organizing the changes you make to files over time. Throughout the lifetime of a project, using a system to keep track of different versions of your files and how and why you changed them can make it easier to revert to earlier versions of files or work collaboratively on the same files. Software like Git and the associated interface GitHub provide a version control framework, particularly useful for code scripts. If you are interested in learning more about Git and GitHub, this lesson and these modules (Git collaboration, Git branches from a course by NCEAS, focusing on using Git with R, are some starting points. For a broader overview of reproducible science, including the utility of version control, see this paper by Lowndes and colleagues.

2.1.7 Data Management: Summary

1. Well done! Having a good folder structure, file naming practices, and version control capabilities are valuable tools to enable easy collaboration and to streamline workflows.

2. Remember that using metadata is extremely important, this is the only way to make sure that you and your team know exactly what is included in a data file. Never share a database without its corresponding metadata.

2.2 Data Cleaning

In this session

1. Learning objectives
2. Value statement

3. Benefits of clean data
4. Data types
5. Best practices
6. Common cleaning procedures
7. Summary

2.2.1 Data Cleaning: Learning objectives

1. Learn the benefits of having clean data sets and best practices for organizing and naming in data sheets.

2. Learn best practices for cleaning data files.

2.2.2 Data Cleaning: Value statement

1. Maintaining clean data sets reduces errors, facilitates searching and analysis, and eases reuse.

2.2.3 Benefits of clean data

1. Let's start by defining what we mean by "clean data". Clean data are as consistent, complete, and accurate as possible. Cleaning data is the process of reviewing data sets to standardize formatting, remove duplication, and fix or remove inaccuracies before using or analyzing the data set. Creating clean data sets involves steps in setting up your data organization and in reviewing data before analysis.

2. Having clean data has many benefits:

    - Reduces errors from redundant updates
    - Enforces data integrity
    - Helps you and future researchers handle large, complex datasets
    - Enables powerful search filtering

3. Much has been written on effective data management to enable reuse of data. The following two papers offer words of wisdom:

    - Some simple guidelines for effective data management. Borer et al. 2009. Bulletin of the Ecological Society of America.
    - Nine simple ways to make it easier to (re)use your data. White et al. 2013. Ideas in Ecology and Evolution 6.

2.2.4 Data types

1. Before we get into cleaning data, let's go over the different ways computers can classify data. The primary data types are

    - Numeric: integers, decimals

  - Any kind of numeric field, except Boolean (binary or true/false statements). Examples include integer, double, floating, etc.

- Text: strings, including special characters
- Yes/no, true/false
  - Also known as logical or boolean
- Spatial Location: Coordinates, Country Name, City Name, ZIP code
- Date and time

2.2.5 Data Cleaning: Best practices

Data sheet set-up, column names, and data values

1. In brief, some of the best practices to follow are:

   - Design your tables to add new observations as rows, not columns. A column should be only one variable and a row should be only one observation.

     For example, Table 1 below does not follow this design. The highlighted box has information for two types of variables in it and the Observations column has many rows with multiple observations listed. In Table 2, the data has been reentered so that each column has only one data type - see the moved note in orange - and Species column now has only one observation per row.

Table 1:

Site	Transect	Observations
Angaur	1	Bolbometopon muricatum, Scarus ghobban, Lutjanus gibbus
Angaur	2	Lutjanus gibbus, unusually large fish
Airai	1	Scarus xanthopleura, Lutjanus gibbus
Airai	2	Epinephelus polyphekadion, Chlorurus sordidus

Table 2:

Site	Transect	Species	Notes
Angaur	1	Bolbometopon muricatum	
Angaur	1	Scarus ghobban	
Angaur	1	Lutjanus gibbus	
Angaur	2	Lutjanus gibbus	unusually large fish
Airai	1	Scarus xanthopleura	
Airai	1	Lutjanus gibbus	
Airai	2	Epinephelus polyphekadion	
Airai	2	Epinephelus polyphekadion	

- Name columns in a clear way that is easy for both people and computers to interpret.
  - Use descriptive column names that would make sense to any reader. For example, don't use acronyms or abbreviations.
  - Avoid the use of special characters, such as !@#$%^&*() and others
  - Make consistent use of capitalization. Are you going to use capital letters in every word? Maybe only at the beginning of the column name?
  - Use underscores instead of blank spaces.

    For example: Number of fish -> Number_of_fish
- When entering data, make sure there is only one type of value in each column, for example all numeric values or all character strings.
- As a good practice, we recommend that you and your organization agree on a format and follow it diligently.

2.2.6 Common cleaning procedures

1. When analyzing data, it is important to make sure that there has been a thorough check of the records to ensure that they are as clean and ready as possible. While there is no straightforward way to make sure that all data is clean (given the unique nature of each dataset), there are some guidelines that can be followed.
2. This first set of recommendations deals with the overall structure of the database:
   - Verify that your column names follow the conventions stated in the section above.
   - Verify that there are no empty columns in your dataset, and if there are, eliminate them
   - If you have a column with the date, verify that it is in the right format and that your software (e.g., Excel, Access, Tableau) is reading it accordingly.
     - Word of caution: when working with international collaborators, check the format of dates. Many places use DD/MM/YYYY as the standard date format. The US uses MM/DD/YYYY as the standard date format.
     - Tableau can recognize different date formats. Sometimes it is easier to capture data in 3 separate columns, (year, month, and day), which can then be linked together directly in Tableau.

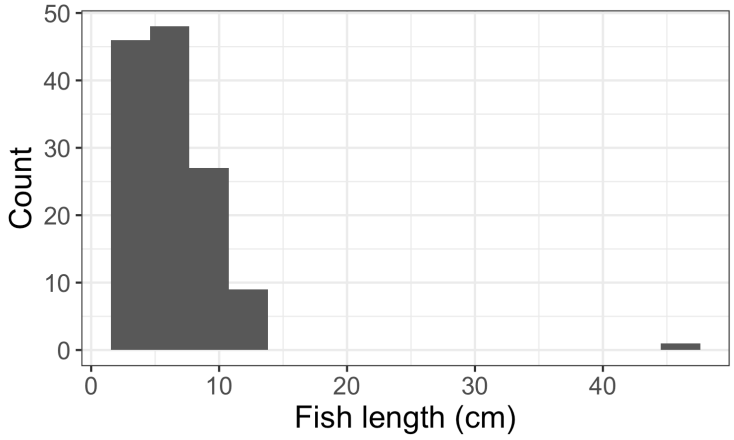
- Verify that your database is in long format, as opposed to wide format.
  - Wide format is a condensed way of presenting data but contains observations in both the column headers and the entered values. Also called block format, wide format is sometimes how data is collected in the field and can be an efficient way of presenting data in a table in a report but is less useful for analysis.
  - In long format, columns only include the type of data and not any information about your actual observations. Having information about your observations in the column headers makes it hard to use that information when analyzing data in Tableau or coding software like R.
  - The same data set often has more rows and fewer columns when in long format compared to wide format, as you can see in the example below. Wide format presents the data efficiently - we can see the whole table - but contains information about the type of meal and type of food in the column headers, which is moved to the rows in long format.

Day	Breakfast_fruit	Lunch_fruit	Dinner_fruit	Breakfast_grain	Lunch_grain	Dinner_grain
Monday	banana	mango	apple	oatmeal	bread	rice
Tuesday	banana	mango	apple	oatmeal	bread	rice
Wednesday	strawberry	orange	apple	cereal	bread	rice
Thursday	raspberry	orange	apple	cereal	bread	bread
Friday	banana	orange	apple	oatmeal	bread	rice
Saturday	pineapple	pineapple	pineapple	pancakes	bread	pasta
Sunday	pineapple	pineapple	pineapple	waffles	bread	pasta

Day	Meal	Food Type	Selection
Monday	breakfast	fruit	banana
Monday	breakfast	grain	oatmeal
Monday	lunch	fruit	mango
Monday	lunch	grain	bread
Monday	dinner	fruit	apple
Monday	dinner	grain	rice
Tuesday	breakfast	fruit	banana
Tuesday	breakfast	grain	oatmeal

3. Next, let's explore some common methods to verify your individual data records:
   - If the column is numeric:

  - Verify that there are no text values
  - Plot a histogram with binned values in the x-axis and frequency in the y-axis. If there are very few observations for very low/high values, those are likely outliers or mistakes in the data, like the single high value in the example below. Find them in your raw data and decide whether it is something that you can fix



- If the column is text:
  - Verify that there are no numeric values
  - Create a list with all the unique elements and sort them out in alphabetical order. From this list, verify that there are no leading or trailing spaces, misspellings, or typos. Get back to your raw data and correct accordingly.
- Make sure that you distinguish between values that are zero and data that was not collected. Zeros should be used for values that were measured and found to be zero, for example the number of fish on a transect you swam but saw no fish. Data that was not collected, for example the number of fish you saw at a site you skipped sampling, should be recorded as NA or another symbol you use consistently throughout your data set.

2.2.7 Data Cleaning: Summary

1. Congratulations! You've now cleaned your data set, storing and presenting it in a way that will be easier to handle, reduce errors, and be easier for others to interpret - both future collaborators and those within your organization. Now that you've cleaned your data set you are ready to start analyzing it.

2.3 Normalization and Standardization

In This Session

1. Learning objectives
2. Value statement
3. Introduction to tidy data
4. Normalization
5. Data organization
6. Standardization
7. Exercise
8. Summary

2.3.1 Normalization and Standardization: Learning objectives

1. Understand the basics of tidy data, particularly normalization and standardization (the processes of normalization and standardization we refer to here are specific to data management and distinct from the other meanings involving rescaling and recentering data as used in statistics.
2. Learn how to design and create effective data tables.

2.3.2 Normalization and Standardization: Value statement

1. Normalizing and standardizing data eases analysis and reduces errors. It is also the basis for having datasets that can be easily updated as more data becomes available.

2.3.3 Introduction to tidy data

1. Tidy data describes a method of organizing databases and data tables that makes data analysis, searching, and reuse easier.
2. The tidy data approach is also known as the relational data model, which is used by relational databases like mySQL, Microsoft Access, and Oracle to store and organize data. If you'd like a more detailed look at the relational data model, see this lesson by NCEAS.
3. You don't have to have a large and complex data set or be using a relational database, however, to see the benefits of having tidy data!

2.3.4 Normalization

1. One characteristic of tidy data is normalization - a process of streamlining data structures and reducing redundancy. Normalized data are essentially data in the long format we discussed in the previous session: each column

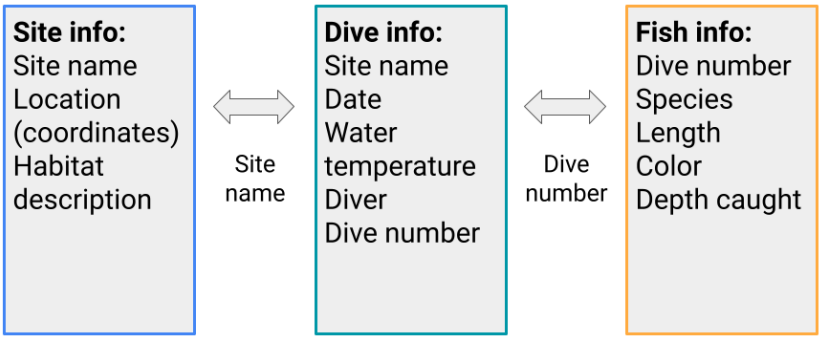
contains only one type of measurement, each row contains one observation, and data are stored in different tables such that each data point only appears once in the database.

2. Three practices to keep in mind when designing your tables are:

   - Add rows, not columns, for new observations.
   - Put values for only one variable in each column.
   - Record each piece of data only once in the database.

3. We covered the first two points in the previous session when discussing best practices for designing data tables. The third point addresses how you store your data in different tables and how they interact. To record each piece of data only once in your database, you need to have distinct collections of data stored in different tables that then link to one another.

   - As an example of separately storing distinct collections of data, say you visit several sites and do multiple dives at each site. On each dive, you record the fish you see and information about each fish, such as species and length. In this case, you would store information about the sites in one table, information about each dive in another table, and information about the fish recorded on each dive in a third table.
     - Storing data only once in a database makes it less likely that there will be errors, especially if you have to alter data later on. For example, if you want to change the format of site coordinates or if the taxonomy of one of the species you sample is updated, you will only have to change entries in one place rather than across multiple tables.
     - After you have normalized your data and identified the relevant tables (e.g., sites, species and dives), there needs to be an identifier that serves as the link between each set of tables. In the example below, the *Dive number* links the Fish info table and the Dive info table - to get information about the dive when analyzing fish, you can pull the data from the Dive info table for the dive of the correct number. Similarly, *Site name* links the Site info table and the Dive info table. The schematic below shows these tables and how they relate.



2.3.5 Data organization

1. Once you have your data tables normalized, make sure that they are arranged in a way that a computer can easily interpret.
    - Make sure that each table is stored separately. For example, if you are using a spreadsheet software, each table should hve its own sheet or tab. Multiple tables on the same sheet are hard for software programs like Tableau to read in and understand.
    - Do calculations, like statistics or marginal sums, outside of your data spreadsheet. They are not observations, like the other rows, but rather data analysis.

2.3.6 Standardization

1. Once you have your data organized in a tidy fashion, it's time to make sure the actual data values are standardized. Standardizing data involves many of the data cleaning steps from the previous session:
    - Checking that you use consistent entries for the same data values ("Yes" vs. "Y" vs. "yes" → choose just one).
    - Checking for data points that are outside the realm of possibility and are likely to be errors, like a dog recorded as weighing 500kg.

- Verifying that all values within a column are the same data type.

2.3.7 Exercise

1. Time to practice with an untidy data set! Use what you've learned about data cleaning to normalize and standardize the data set here. See the tidied version here.

2.3.8 Normalization and Standardization: Summary

1. Well done! Here is a checklist summarizing what we've learned about data organizing and tidying in this module. Now that you've learned about best practices for managing and storing data, it's time to move onto Tableau.

Chapter 3

Visualizing Data

3.1 Tableau Basics

In this Session

1. Learning Objectives
2. Why Tableau?
3. Installing Tableau
4. User Interface
5. Loading Data
6. Opening a File
7. Checking Data Types
8. Sheets
9. Summary

3.1.1 Tableau Basics: Learning Objectives

1. Get familiar with the user interface of Tableau.
2. Learn how to connect data sets to Tableau interface.

3.1.2 Why Tableau?

1. We are focusing on Tableau for two primary reasons.
    - First, Tableau can read a variety of data types, as you will learn in this session. This means that many different software applications and assets can be combined in a single visualization to present to your audience.
    - Second, Tableau is extremely intuitive. Once you get a grasp of the basic skills, you can create compelling visualizations of data in half of the time it would take in other coding software applications.

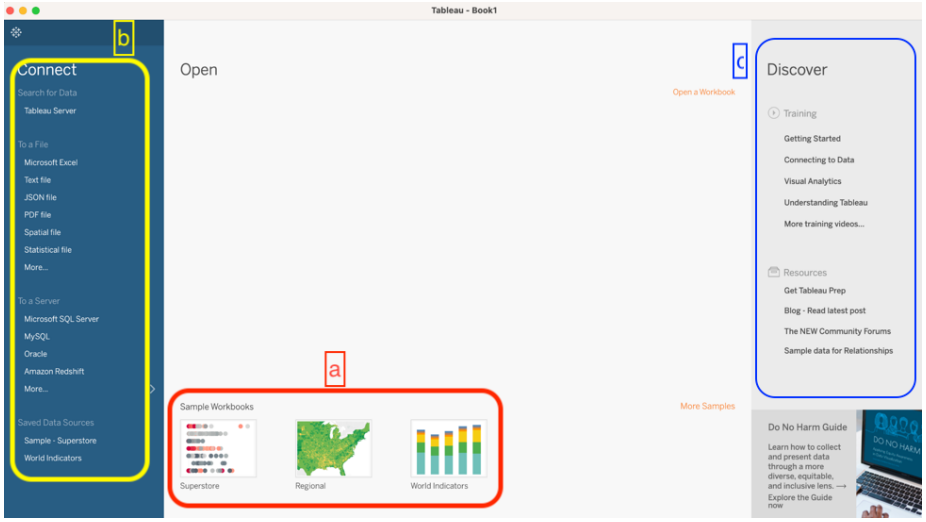
- Note: For this training, all the exercises and examples were done with Tableau version 2021.1. The menus may look slightly different in other versions of Tableau.

3.1.3 Installing Tableau

1. Let's get started with installing the software, to do so, go to https://www.tableau.com/ and select "Try Tableau for Free."
2. Register your email address and username and download the software.
3. Once installed, input your key (you get one when you buy the software) and you are ready to go.
4. **Important note:** Tableau offers free 1-year licenses for educational purposes. If you belong to an academic institution, get a free license by going to https://www.tableau.com/community/academic. Once there, click on "Free student license". You will need to provide information about your academic institution to validate your educational license.

3.1.4 User Interface

1. Let's open Tableau! The image below is what you will see the first time you open Tableau. Once you start to save workbooks, they will populate here under **Open**.



(a) The examples under **Sample Workbooks** are a great resource for inspiration to see the potential of Tableau.
(b) This column on the left is where you will load data.
(c) The **Training** and **Resources tabs** are very useful to help you keep learning new skills. If you are interested in investigating or expanding

your skills, we recommend starting with the materials in these tabs, rather than searching for other sources on the internet. The materials in these two tabs are very well thought out and provide step-by-step instructions on how to learn new Tableau skills.

3.1.5 Loading Data

1. The most important section of this interface is the **Connect** menu, where you will load data. There are three main categories of sources that you can load into Tableau.
   - *Tableau Server:* Data you have saved through your Tableau account. We will not cover that process in this course because most of the users of this training will probably not have a server setup explicitly for Tableau.
   - *Connecting to a local file:* Files from your computer can be Excel files, text files, statistical files, PDFs, and spatial files
     - Text files include .csv .doc .log
       - You are going to be using **Text file** frequently if you have .csv (comma separated value) files.
     - Spatial files are commonly referred to as shape files or GIS files. If you have ever dealt with shape files, you might remember that shape files are composed of a few different files – some containing the coordinates, metadata, and a table. Tableau has made the adjustments necessary to load one shape file and all the associated files.
       - As we will show this later in the course, you can load a shape file and Tableau will display it in a map with just a few clicks.
   - *Connecting to a server:* You can connect to services that are online like Google Sheets, or Microsoft Office 365.
     - One of the benefits of using this approach is that when you modify the original file, the modifications will be directly updated in Tableau.
       - Let's say you created a report with nice graphs in 2020 because that's the data that you had when you created the report and then in 2021 you edited your Google sheet. The next time you open your Tableau workbook, all the data will be updated and contain the new data from 2021.
2. It's very important to make sure that you organize and manage your files such that you know exactly where your files are. Keep in mind that if you connect to a server, any changes that you make in the original file will have an impact on how your visualization looks in Tableau.

3.1.6 Opening a file

1. Create a new folder on your desktop. Name the folder "Tableau Data Training."

2. Download this file: CRM_Fish_edit and save it in the "Tableau Data Training" folder.

3. Follow along with this video to connect the new data file to Tableau.

4. Progress check-in. So far, you have:
   - Downloaded an excel file and connected it to Tableau
   - In Tableau, you learned about drag and drop. As Alfredo always says, "Tableau is all about drag and drop."
   - You connected the two tabs CRM_Fish and CRM_Site within the CRM_fish_edit Excel file—now in the Tableau interface. There is an orange line showing the connection.

3.1.7 Checking Data Types

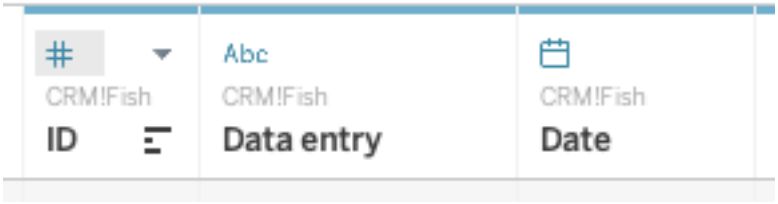
1. When you load a data file, you have to check that Tableau properly labels the data types. Here, we will discuss what those data types are and how they are assigned in Tableau.
   - ***Dimensions and Measurements***
     - **Dimension** (also referred to as factor) is any information that is defined a priori of your actual sampling and helps to describe or categorize your results. Whenever you can call something a category it is a factor. Depending on your data, factors are sometimes written as numbers, like year or site ID.
       * Example: Site name, species, habitat
     - **Measurements** are the actual results of your sampling. They are usually numbers that you went into the field to measure.
       * Example: Fish size, population count, temperature
     - Tricky Case: Coordinates

        * Coordinates are a factor when you already know the coordinates where you are going to take your sample. Example: You have ten sites that you visit each year to count the number of whales. The coordinates of each site are a factor because they are predetermined. You know where you are going to gather data each time.
        * Coordinates are a measurement when it is part of the information that you are recording and it is not predefined. Example: you put a gps tracker on a whale and every time the whale surfaces, you record the coordinates. You did not know ahead of time where the whale would surface and the location is what the study is measuring.
    - Tricky case: Is depth a factor or a measurement?
        * Depth is a factor when indicating a transect location. Example: If you are always putting transects at either 3m or 30m you can categorize depth into shallow or deep.
        * Depth is a measurement when it is recorded with every observation. Example: If you are swimming for 50 minutes to record visual observations of fish while also noting the depth at each observation, then depth is also a measurement. You measured and recorded your depth as you recorded the observation of the fish.
    - Practice exercise: Time
        * SCENARIO 1: Let's say that you are birdwatching and you want to know what time of day birds are more active. You count the birds and you write down the time that you saw each bird. Is time a measurement or a factor?
            · Answer: Measurement because you are registering the time every time you see a bird.
        * SCENARIO 2: You want to know whether there are more birds in the morning, the afternoon, or in the evening.
            · Answer: In this case, time is a factor because you go out three times a day, you count all the birds you see at that time and categorize by morning, afternoon, or evening.
- ***Types of Data***
    - Different ways computers classify data:
        * Numeric - integers, decimals
            · Any kind of numeric field, except Boolean (binary or true/false statements). Examples include integer, double, floating, etc.
        * Text - strings, including special characters
        * Yes/no, true/false - Also known as logical or boolean
        * Spatial Location - Coordinates, Country Name, City Name, ZIP code
        * Date and time. Tableau can recognize different date formats. One best practice is to separate fields for year, month, and

day, which can be linked together directly in Tableau.

- ***Discrete and Continuous Data***
  - **Discrete data** refer to variables that can only be conceptualized in a unitary manner. *It is quite similar to the difference between using many vs much. In this case, discrete data would be for things described as many.*
    - Example for text – color, name, city
    - Example for numbers – number of transects conducted, position in a ranking
  - **Continuous data** refer to variables that cannot be counted by units, but instead as continuous amounts. *Following the example above, continuous data would be for things described as much.*
    - Example: number of fish recorded, total biomass, height, depth

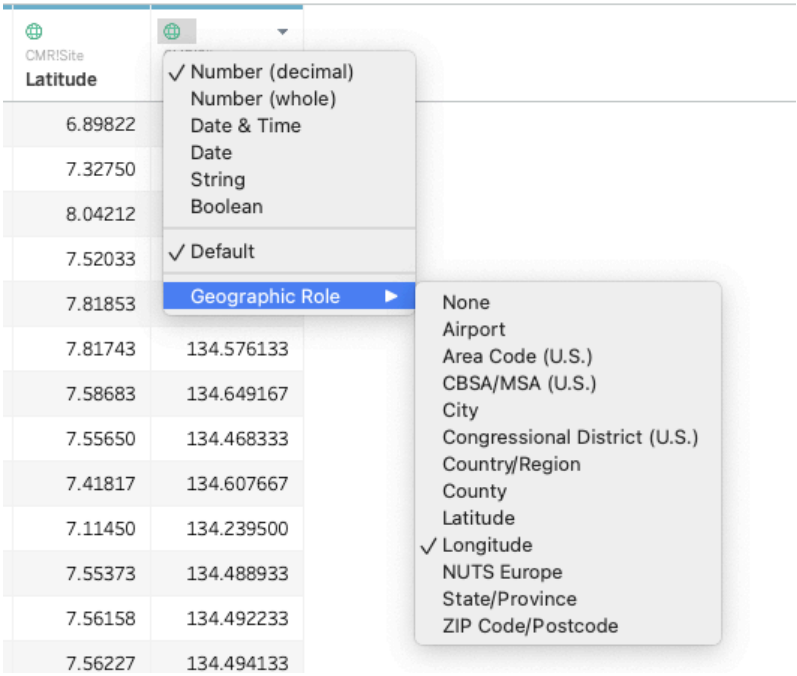
2. You should go through the preview of the data and ensure that each column is labeled correctly.
   - *Note: Never skip this step even if you have worked with the data set in Tableau before!*
3. The type of data is shown as symbols above the column title.



4. As you can see, the **ID** column is correctly labeled as a number; **Data entry** is a string (or text); and **Date** is a date (little calendar symbol).
   - *Note: Remember string is a synonym for text in Tableau.*
5. To change the label, click on the symbol and a drop-down menu will appear.
   - The **CRM_Fish** table should look like this:

   - The **CMR_Site** table should look like this:

   - Notice that Latitude and Longitude are spatial data. If you open the drop-down, you will see that you can specify the type of spatial data provided under Geographic Role.



3.1.8 Sheets

1. Sheets Cheat Sheet
   - Discrete vs. Continuous
     - DISCRETE is Blue
     - CONTINUOUS is Green
   - Factors vs. Measures
     - Above the grey line is a FACTOR
     - Below the grey line is a MEASURE

3.1.9 Tableau Basics Summary

1. Great job! You have completed your first in depth Tableau session. Now, you know how to explore Tableau's user interface, to connect a data file, check data types, and open sheets.

3.2 Tableau Skills Part 1: Basic plots with single variables

In this Session

1. Learning Objectives
2. Value Statement

3. Creating Your First Graph
4. Basic Plots using the Marks Menu
5. Line Plot
6. Area Plot
7. Map
8. Bubble Plot
9. Scatter Plot
10. Editing and Formatting the Axes
11. Saving my Files
12. Homework Exercise
13. Summary

3.2.1 Basic Plots: Learning Objectives

1. Explore the user interface.

2. Learn how to make basic plots and customize them.

3. Learn how to save and share workbooks.

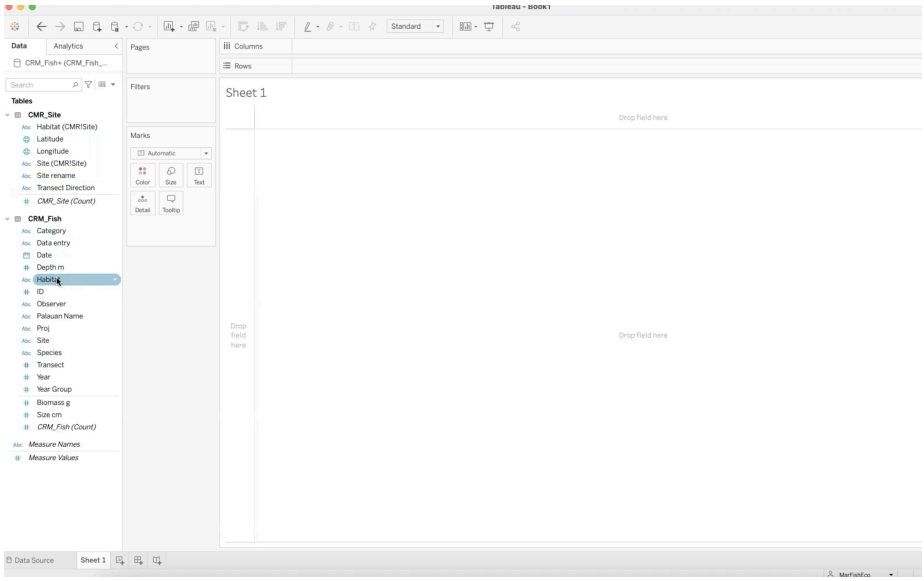
3.2.2 Basic Plots: Value Statement

1. Learning how to make basic plots in Tableau (e.g., bars, lines, pies, etc.) and how to take full control of the appearance will prepare you to take full advantage of Tableau's more advanced skills.

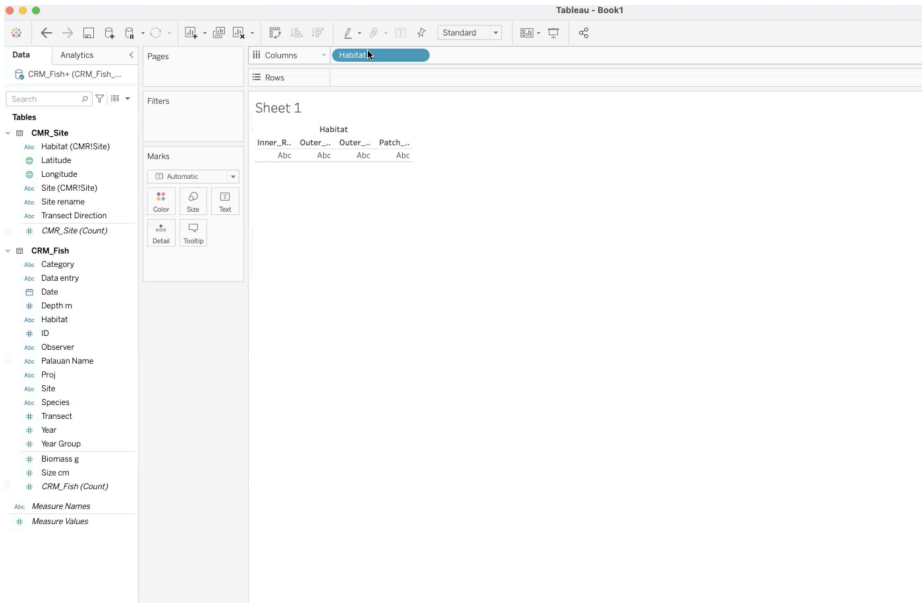
2. *For this session we recommend that, if possible, you have two screens open—one with Tableau and one with the course material—so you can follow along with the exercises.*

3.2.3 Creating your first graph

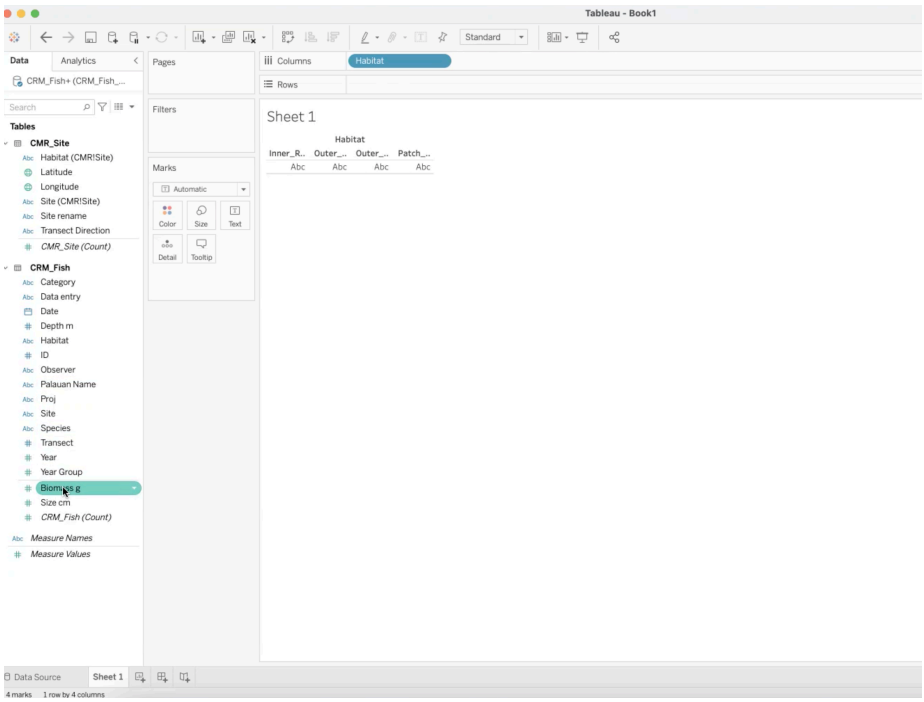
1. In the menu on the left, find **Habitat** under CRM_Fish



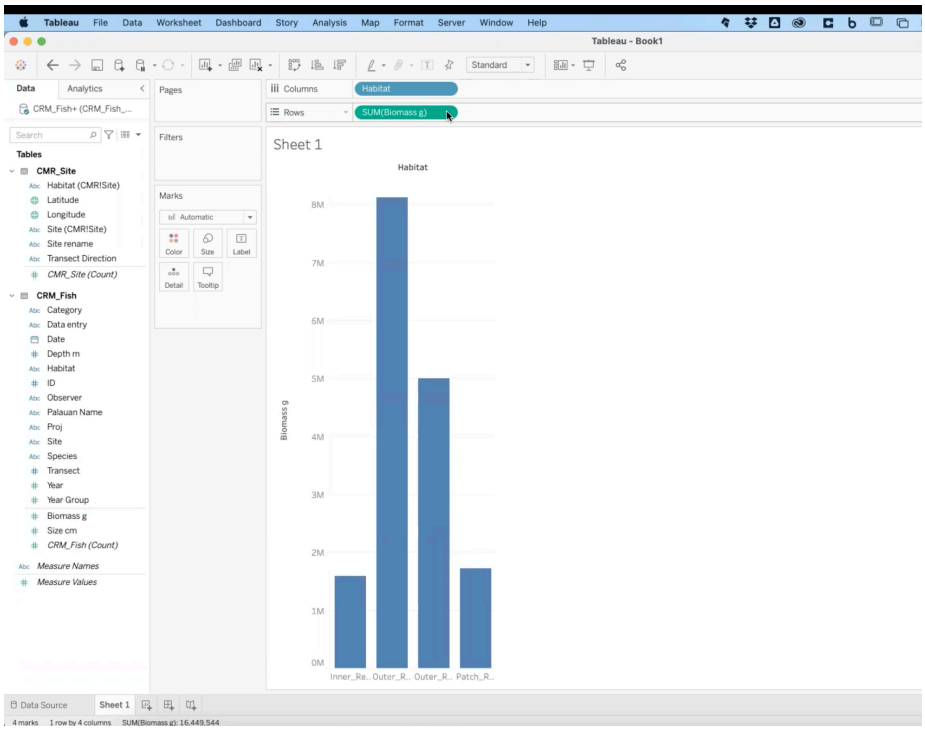
2. Drag and drop **Habitat** into **Columns**
    - This will place **Habitat** in the x-axis of the graph



3. Find **Biomass** under CRM_Fish



4. Drag and drop **Biomass** into **Rows**
    - This will place **Biomass** in the y-axis.

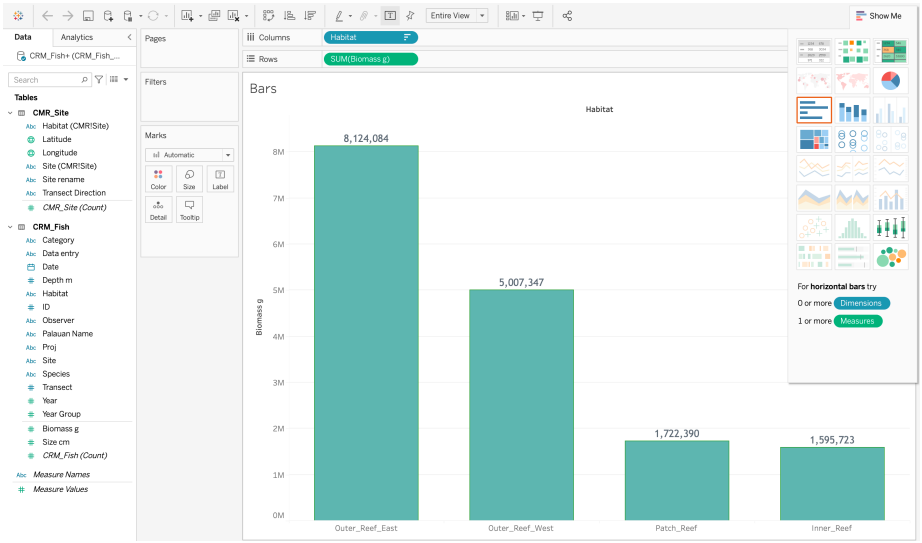


5. Congratulations! You just made your first bar plot in Tableau!

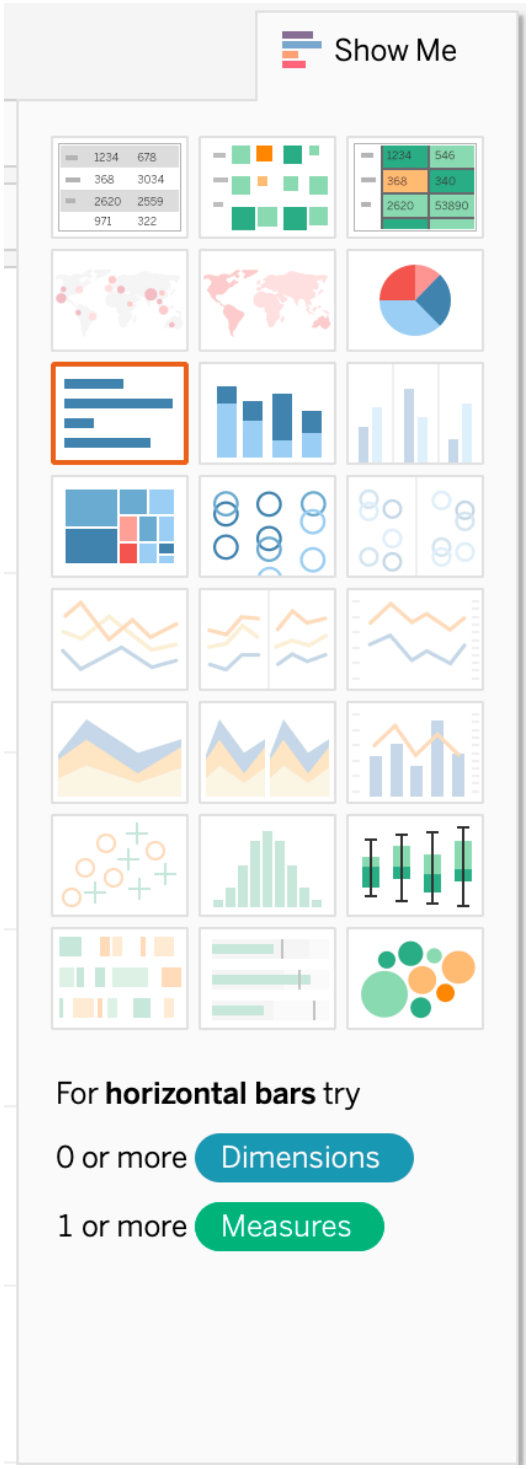
Entire View and Sorting

Show Me Menu

1. The Show Me menu may pop up when you create a new graph. It looks like the image below.



2. If the Show Me menu obscures your view, you can simply click on **Show Me** at the top and the menu will collapse.


Show Me
1234 678
368 3034
2620 2559
971 322
1234 546
368 340
2620 53890
For horizontal bars try
0 or more Dimensions
1 or more Measures

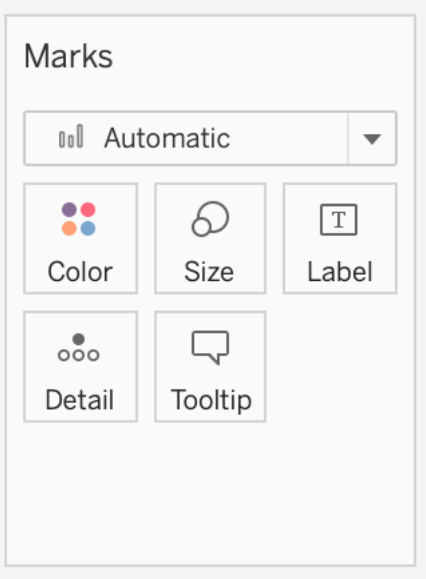
3. How to use the **Show Me** menu

Troubleshooting: Switching Rows and Columns

1. Lilli has a question that came up when she was exploring the Show Me menu. When she tried to go back to a bar chart, the chart was flipped! Watch Alfredo's answer:

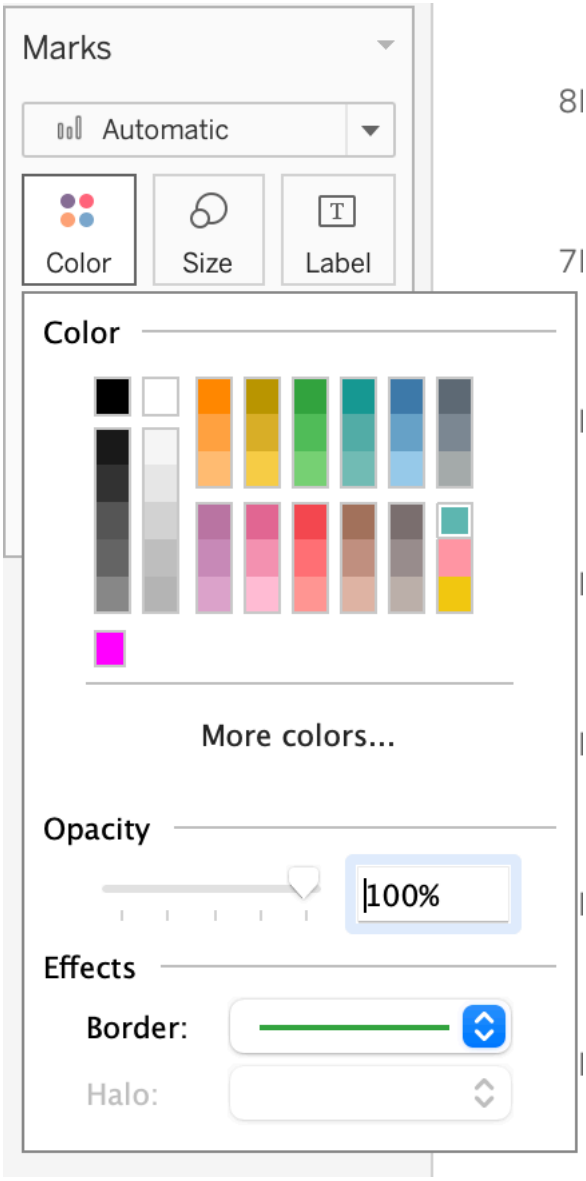
3.2.4 Basic plots using the Marks menu

1. The Show Me menu is a great tool where Tableau can suggest automated graphs, but we want to teach you how to take full control of Tableau. In order to do that, we will utilize the Marks menu.
    - The Marks menu is located to the left of your bar graph:



2. Color

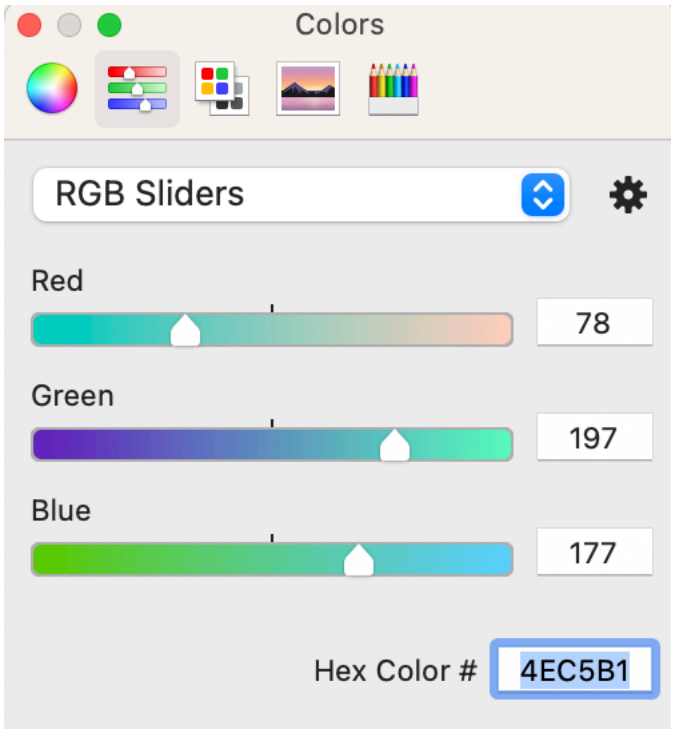
    - Here you can customize the color, opacity, and border of your graph.



- Click on **More colors...** and you will see even more options to customize your color palette.

- Practice: Insert a Hex Color

  - Find the Hex Color field under **More colors...**
    - For example, here is the "RGB Sliders" window on a Mac.



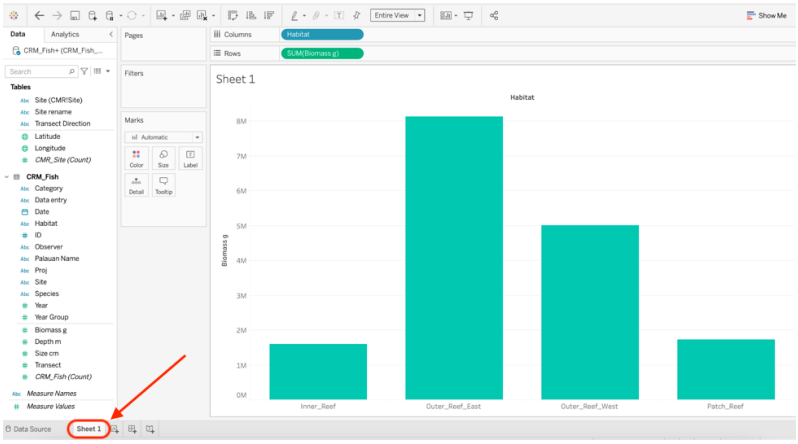
* Type or paste **4EC5B1** in the Hex Color #

* Now, your graph should be this color:

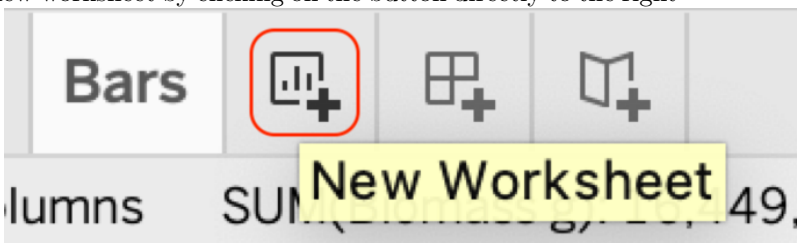
3. Size

- If you click on Size, a slider will appear. In bar graphs, the slider changes the width of each bar. Let's see how Size affects a pie chart.
  - First, name this sheet "Bars." REMEMBER TO ALWAYS NAME YOUR SHEETS

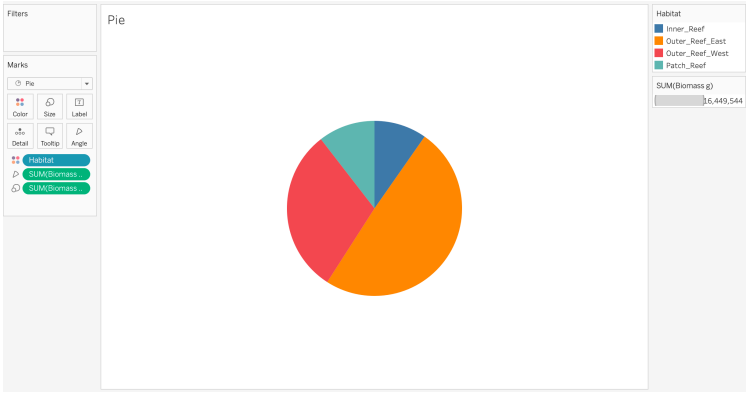
- To edit, double-click on the bottom where it says "Sheet 1"



- Create a new worksheet by clicking on the button directly to the right of "Bars".



- Name this sheet "Pie".
- In the new sheet, create exactly the same graph. Drag **Habitat** into **Columns** and **Biomass** into **Rows**.
- Go to **Show Me** and choose Pie Chart.
  - The pie chart will look very small, so change the view to Entire View. Your current sheet should look like the image below:



- Try using the **Size** function in the **Marks** menu. Notice that when you change the size of this graph, you are changing the size of the whole pie chart.

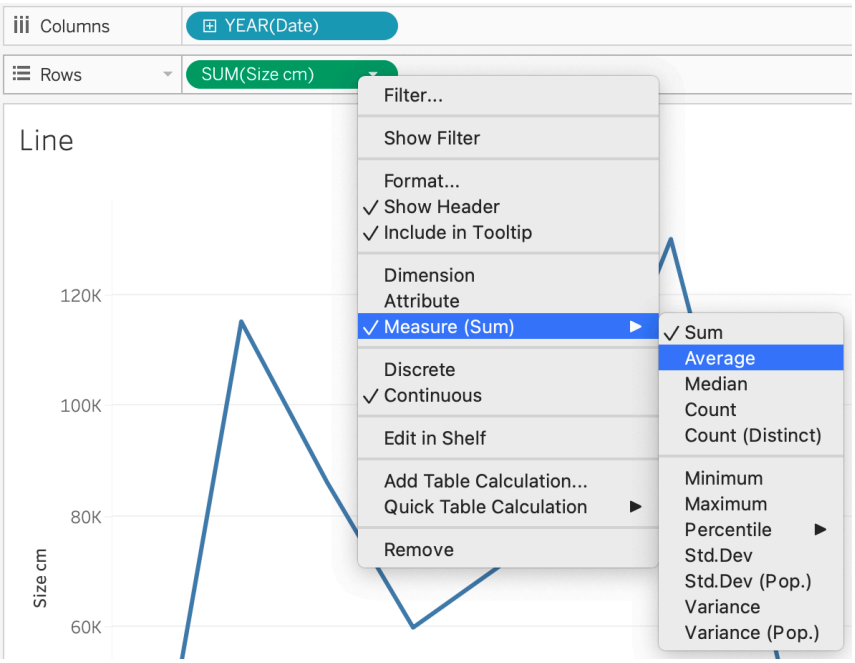
- When you are changing something like the size, you have to experiment

a bit to see what the software will do. Each graph type will respond differently to changing the size.

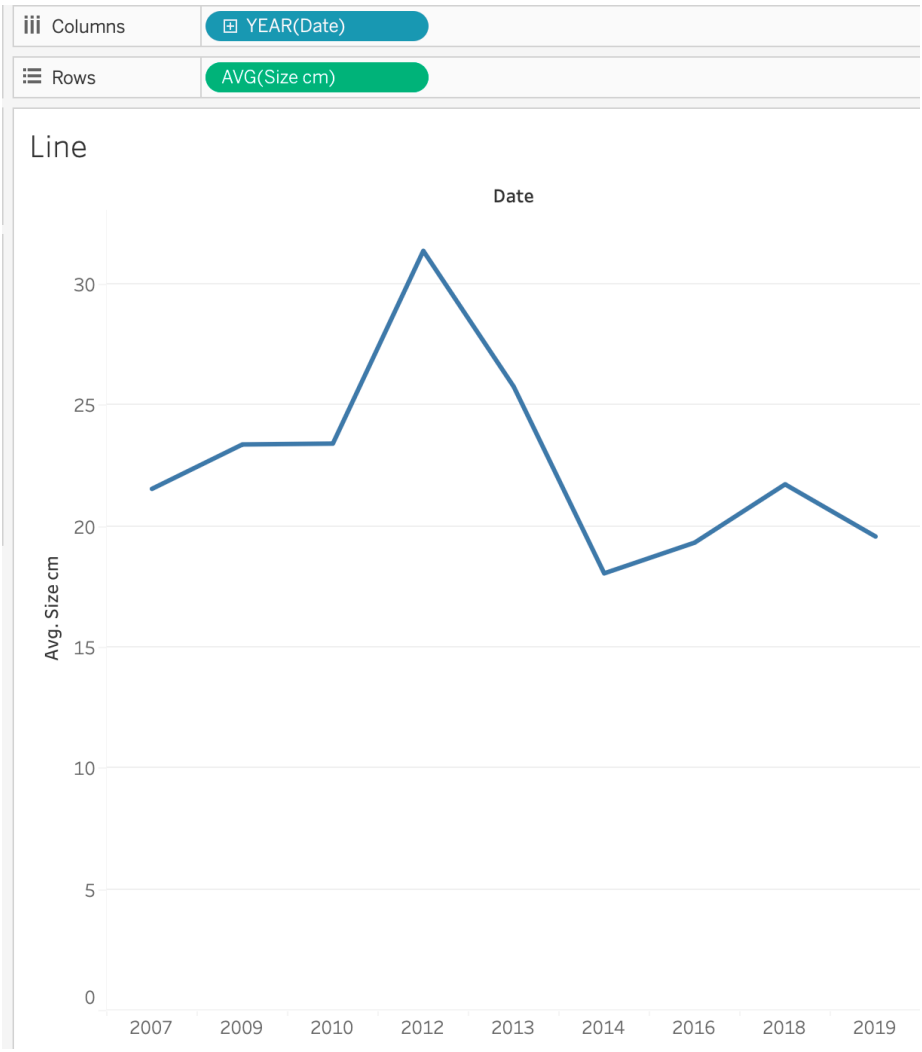
4. Labels
5. Details
    - We will cover the Details marks in a later session.
6. Tooltip

3.2.5 Line Plot

1. Practice Exercise: Create and modify a **Line Plot**
    - Create a new sheet and name it "Line".
    - Drag **Year** into **Columns** and drag **Size** into **Rows**.
    - For a line plot, summing the size of all the fish doesn't make much sense. In this case, showing the average size of the fish per year makes more sense, so let's do that.
    - Right click on the field where it says **SUM (Size cm)** and a drop down menu will appear. Choose **Measure**, then click on **Average**.

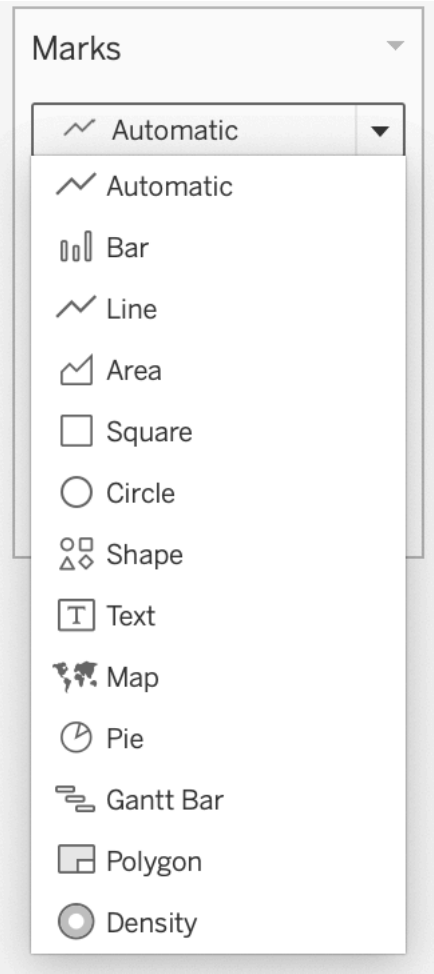


- Does your plot now look like the image below? If so, you can move on to the next steps.



3.2.6 Area plot

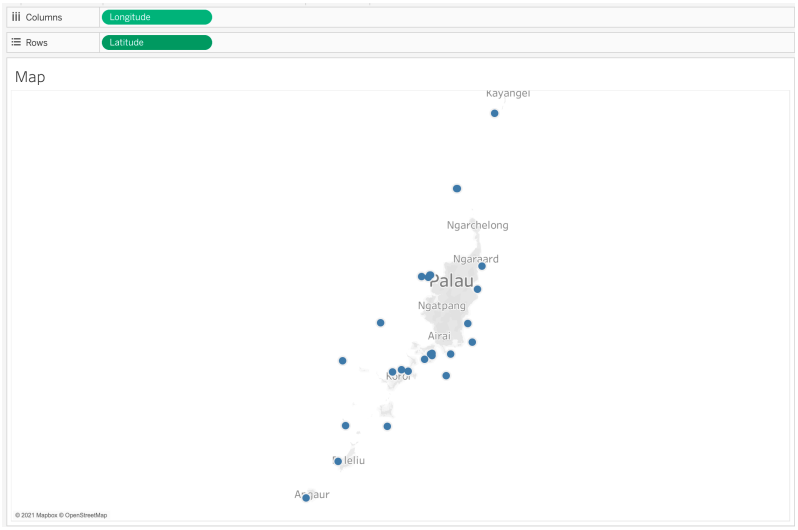
1. Good job! From here, it is easy to change your **Line** plot to an **Area** plot. Under the **Marks** menu, open the dropdown menu where it says **Automatic** and click on **Area**.



2. As you can see, there are many options in this dropdown menu. This is where you can manually change the type of plot displayed.

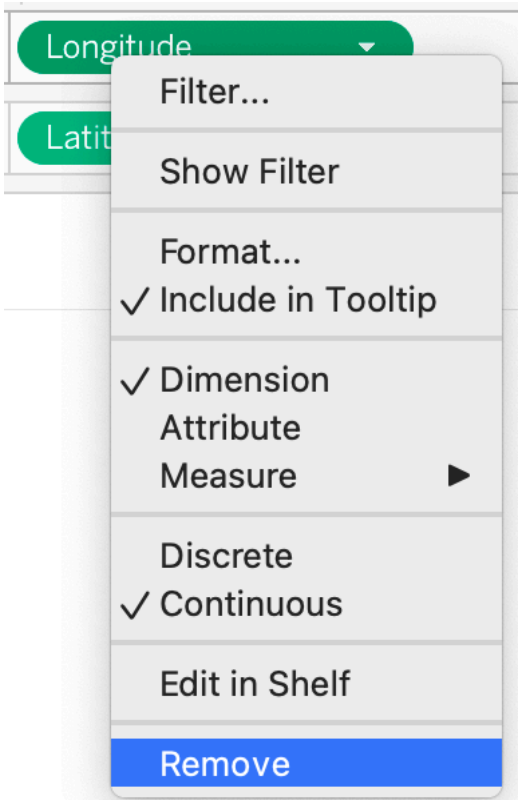
3.2.7 Map

1. There are two ways to create a map.
2. First, on a new sheet you can drag and drop **Longitude** to **Columns** and **Latitude** to **Rows**.
    - *Remember that Columns are the x-axis and Rows are the y-axis.*

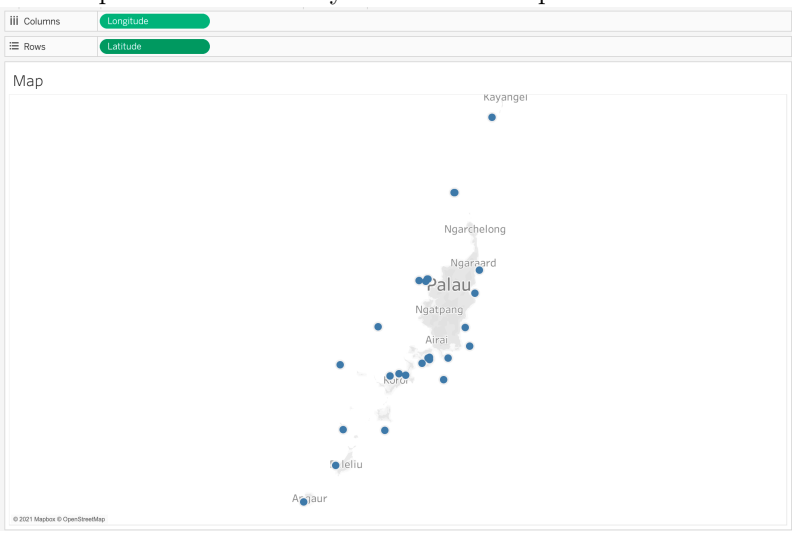


- Your map should look like this:

3. Next, clear the **Columns** and **Rows** fields. You can do this either by right-clicking and then choosing **Remove** or you can drag **Longitude** and **Latitude** back into the left column.

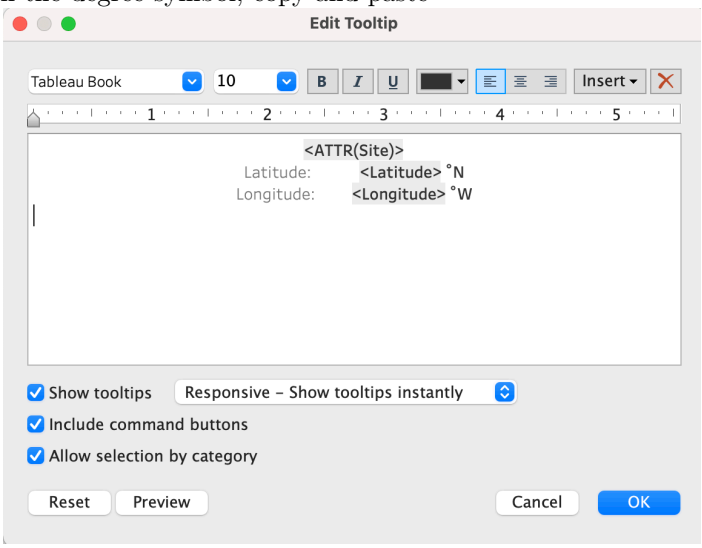


4. Second, there is an even easier way to plug in geographic spatial data. Double-click **Latitude** and then double-click **Longitude**. If you assigned the data properly (i.e. Dimension, Geographic Role: Latitude), Tableau will know exactly how to create this map for you.
    - Your map should look exactly the same as the previous visualization:

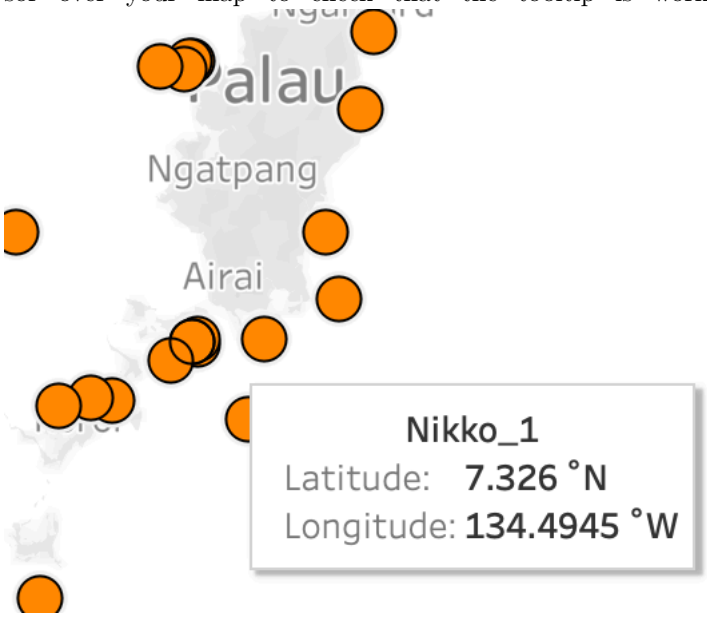


5. Name this sheet "Map"
6. Customizing the Map using the **Marks** menu
    - Using the **Color** tool, change the color of the elements to a preferred color and add a black border.
    - Using the **Size** tool, make the elements a bit larger so you can see them better.
        - Remember that sliding to the right makes the elements larger and to the left makes them smaller.
    - Do you notice anything missing? If we were to show this map to a collaborator, they would only see the Longitude and Latitude, but not the Site name. Let's add the **Site** name to the **Tooltip** as dynamic text (or text that will be updated as the user interacts with the visualization).
        - Find **Site** in the left column under CRM_Fish. Drag and drop **Site** into **Tooltip** in the Marks menu.
    - Edit **Tooltip** to clarify the map for users.
        - Click on **Tooltip** so we can edit the text. We want the name of the site to be at the top, so cut and paste "Site: <ATTR(Site)>" to move it above Latitude and Longitude.
        - Delete the static text "Site:" leaving just "<ATTR(Site)>" at the top.
        - Select all the text and change the alignment to center.
        - Next, add "°N" to the right of <Latitude> and "°W" to the right of <Longitude>

* Mac shortcut: "Option + Shift + 8"
* Windows shortcut: ""Alt + 0176" (with Num Lock on)
* If you have trouble with the degree symbol, copy and paste from this text: ˚N ˚W



– Click **OK** to close the editing window and hover your cursor over your map to check that the tooltip is working.



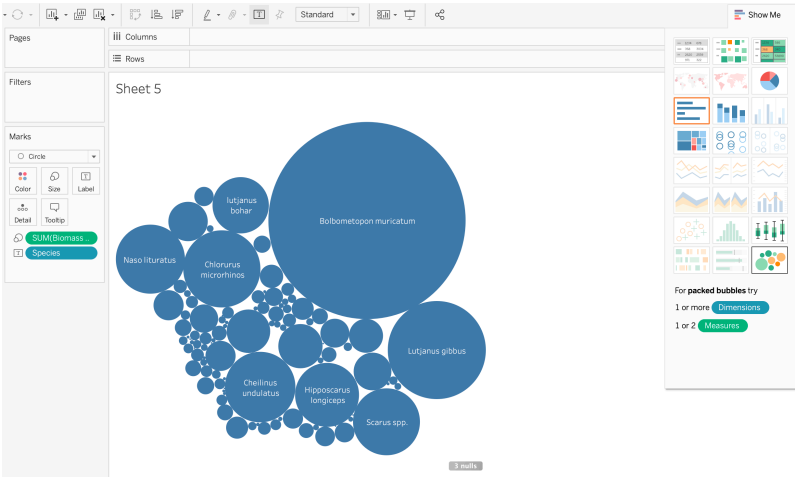
7. You have just created an interactive map showing the various sites around Palau where data was collected.
   - Take a moment to explore the functionality of the map you created by moving your cursor around the map and clicking on a few of the

data points.

3.2.8 Bubble plot

1. To create a Bubble plot, you always start with a Bar plot.
    - In a new sheet, drag and drop **Species** into **Columns** and **Biomass** into **Rows**.

- Then, click on **Show Me** and choose **Bubble**.



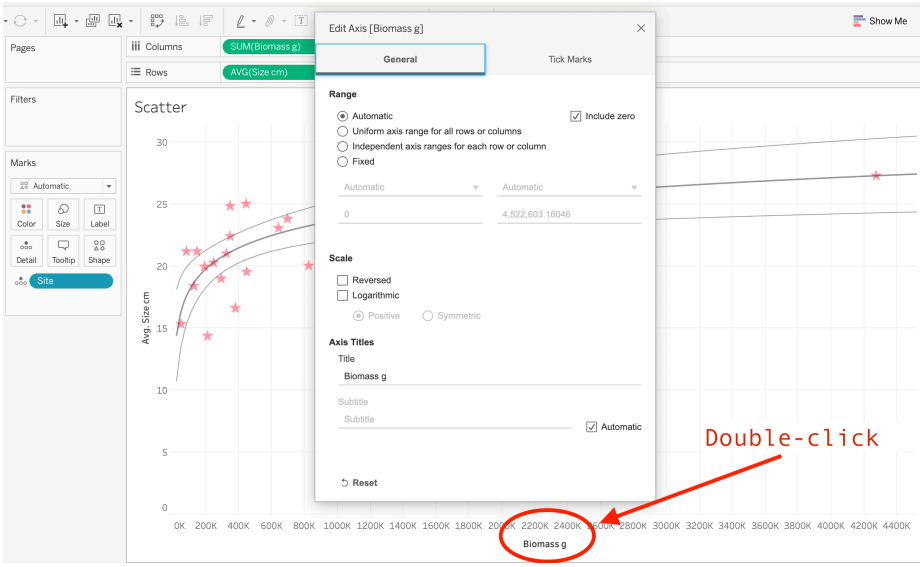
- E Voilá! A Bubble plot!
- Name this sheet "Bubble."

3.2.9 Scatter plot

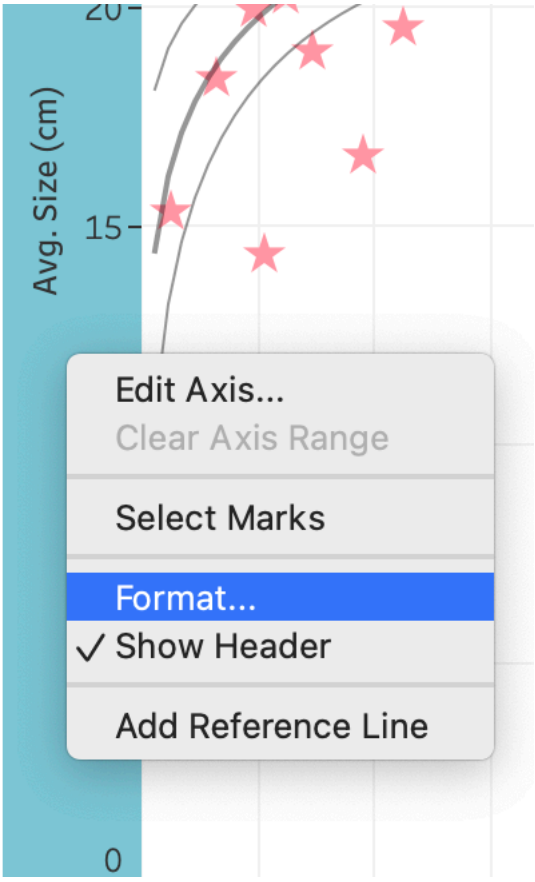
1. Finally, let's create a Scatter plot. Scatter plots are tricky; follow along with this video to learn about scatter plots in Tableau.

3.2.10 Editing and Formatting the Axes

1. "Biomass g" is not a very descriptive title for the x-axis. Let's change it to give more information to a viewer. Double-click on the axis and an editing window will pop up.

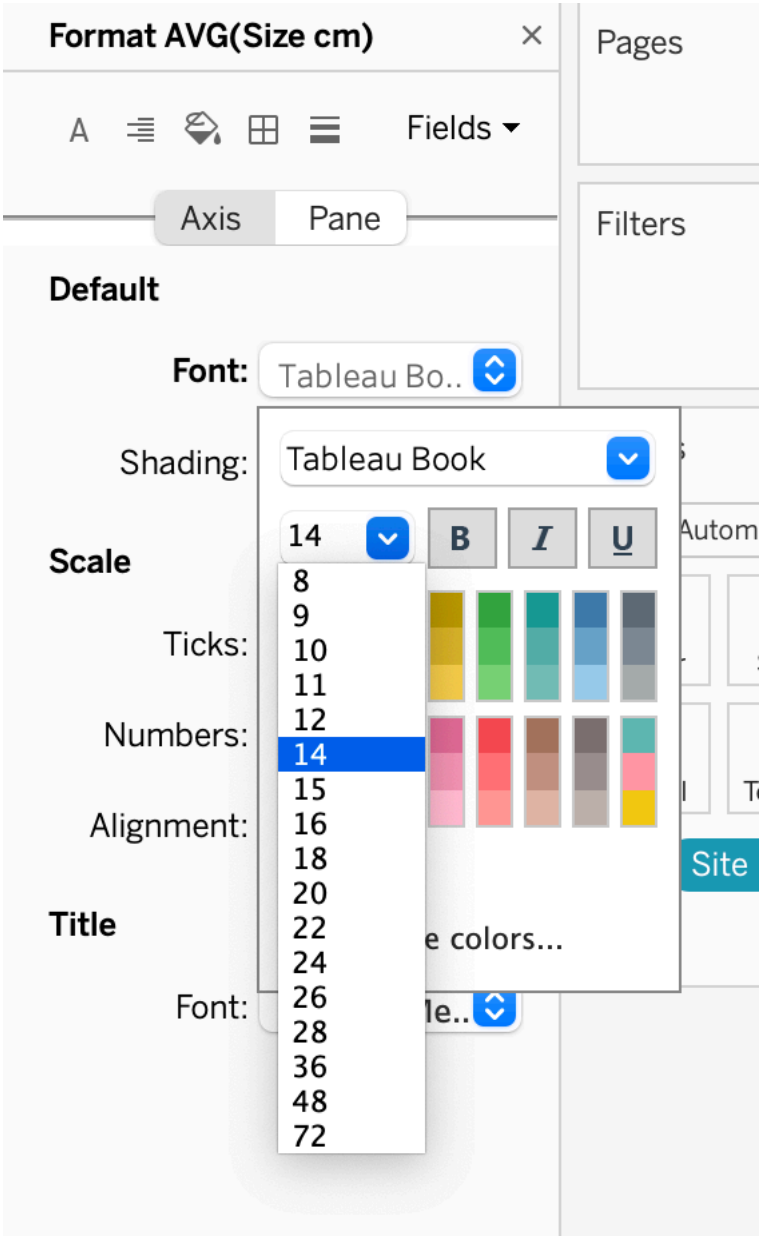


2. Under Title delete “Biomass g” and type “Total biomass (g).”
3. Do the same for the y-axis. Replace the title with “Avg. Size (cm).”
4. Now, we want to increase the font size to make the axes titles easier to read.



- Right-click on the y-axis and choose Format.
- Open the dropdown menu that says Font: and change the font size to

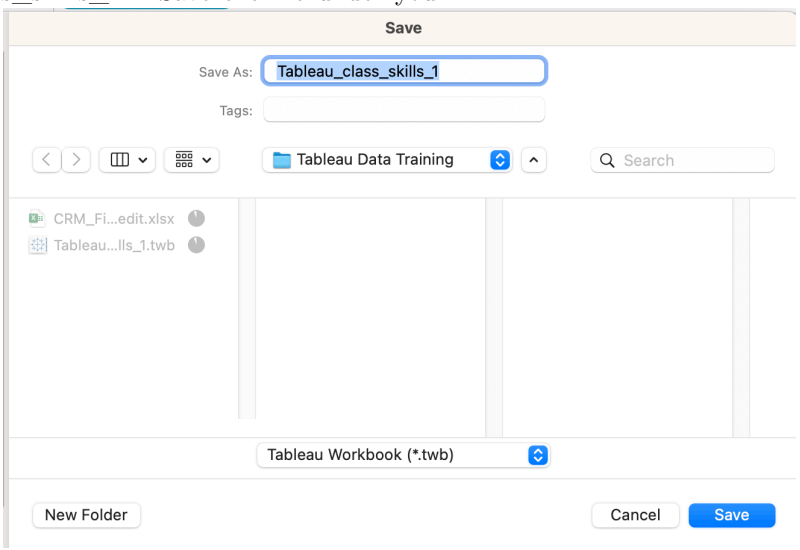
14.



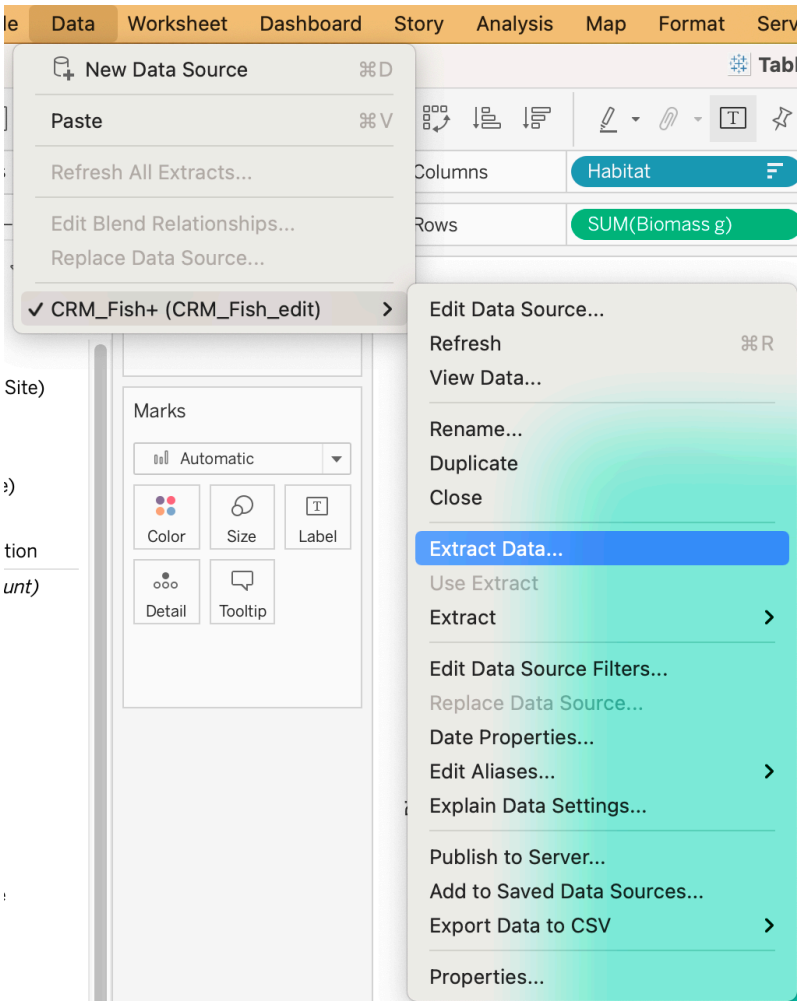
5. Do the same for the x-axis.
6. When you are done, click on the X in the right corner of the Format window to exit.

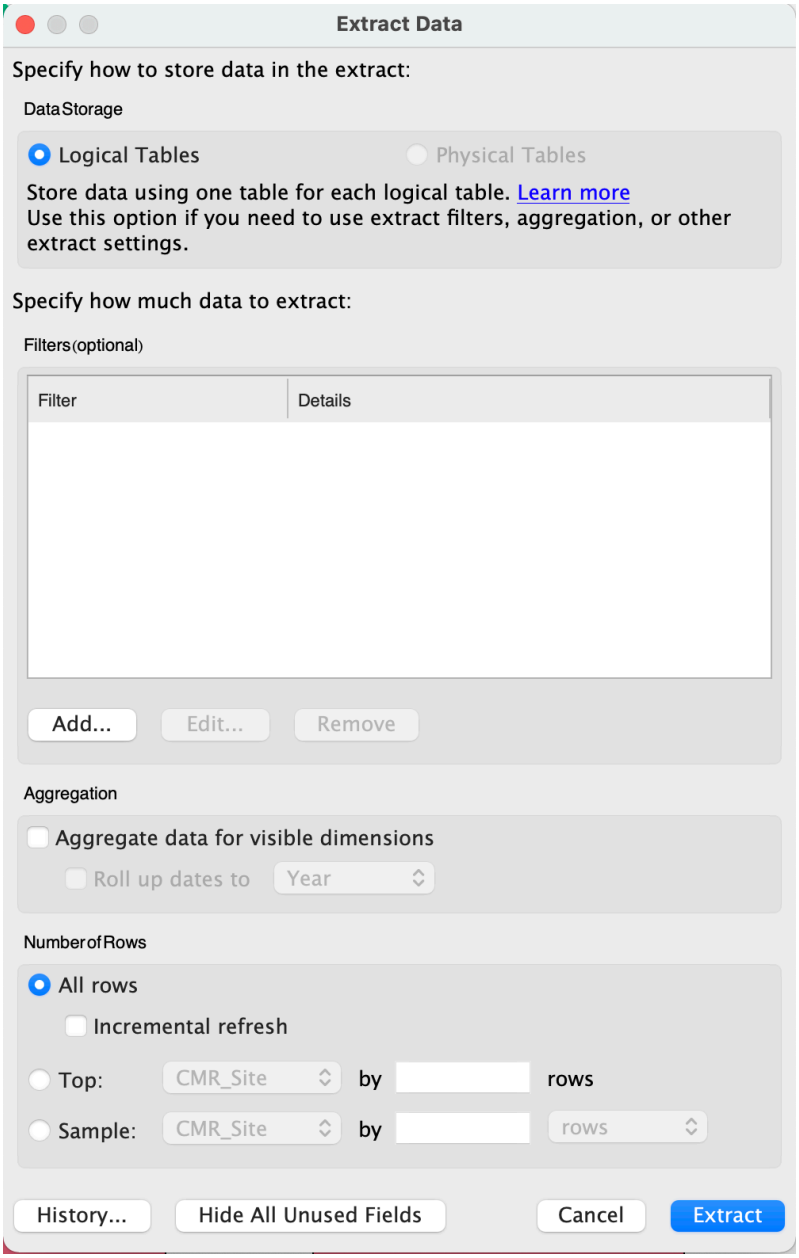
3.2.11 Saving my Files

1. There are two primary ways to save your work in Tableau. The first is similar to saving any other type of file on your computer.
   - Click on **File > Save As** on the menu bar at the top of your screen. Name this file "Tableau_class_skills_1". Save the file under your Tableau Data Training folder.



   - Let's take a look at the files. If you look at the Tableau Data Training folder, you'll see that the Excel file CRM_Fish is 2.1 megabytes while the Tableau Workbook (.twb) that you just saved is 226 kilobytes. That's significantly smaller than the original Excel file because your Tableau Workbook file doesn't contain the data, it just has the instructions for what to do with the data. Therefore if you save .twb files (Tableau Workbook file) and you want to share them with a colleague, you also have to share the original data file.
2. Packaged Workbooks
   - The second way to save a file saves a compressed version of both the data file and the Tableau Workbook. This is called a Packaged Workbook.
   - First, you have to create a data extract, which puts your data in Tableau format. On the menu bar at the top of your screen click on **Data > CRM_Fish > Extract Data**


Data
Worksheet
Dashboard
Story
Analysis
Map
Format
New Data Source ⌘D
Paste ⌘V
Refresh All Extracts...
Edit Blend Relationships...
Replace Data Source...
CRM_Fish+ (CRM_Fish_edit)
Columns
Habitat
Rows
SUM(Biomass g)
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Edit Data Source...
Refresh ⌘R
View Data...
Rename...
Duplicate
Close
Extract Data...
Use Extract
Extract
Edit Data Source Filters...
Replace Data Source...
Date Properties...
Edit Aliases...
Explain Data Settings...
Publish to Server...
Add to Saved Data Sources...
Export Data to CSV
Properties...



- Click extract, name this file CRM_Fish_Extract, and save it in your Tableau Data Training folder.
- Now that you have created your extract, you will create a packaged file. In order to do that, go to **File > Export Packaged Workbook > Save**
- If you look at the folder in your finder, you will see that the new

file type is **.twbx** which means it contains both the data and the instructions (Tableau Packaged Workbook file).
- If you want to share your work with colleagues who also have Tableau installed, you will share Packaged Workbooks (.twbx) with them.

3.2.12 Homework Exercise

1. Open data file Excel file CRM_Fish in Tableau, check field dimensions vs measure, data types, discrete vs. continuous.
2. Create a bar plot, a pie chart, line, map, scatter plot.
3. Make presentable for an audience. Ensure that:
   - Axes are legible size and descriptive titles
   - Marks utilize nice and legible colors
   - Tooltips provide useful information
   - Name all worksheets
4. Create a packaged workbook (.twbx).
5. Check your work by comparing it with this file.

3.2.13 Tableau Skill Part 1: Summary

1. Congratulations! You are now equipped to create basic plots in Tableau and take full control of their design.
2. You also know that when sharing your work with colleagues, you should always include the data, either as a separate file, or included in your workbook as a data extract.

3.3 Tableau Skills Part 2: Multiple factors and dynamic tables

In this session

1. Learning objectives
2. Value statement
3. Multiple factor plots using the Marks menu
   - Create a map with multiple factors
   - Create a bar plot with multiple factors
   - Create a timeseries with multiple factors
   - Create a scatter plot with multiple factors
   - Create a bar plot using continuous data
4. Maps
   - Practice exercise: Search and reset on a Map.
5. Filters
6. Dynamic tables
7. Save your work
8. Conclusions

9. Homework

3.3.1 Multiple Factors and Dynamic Tables: Learning objectives

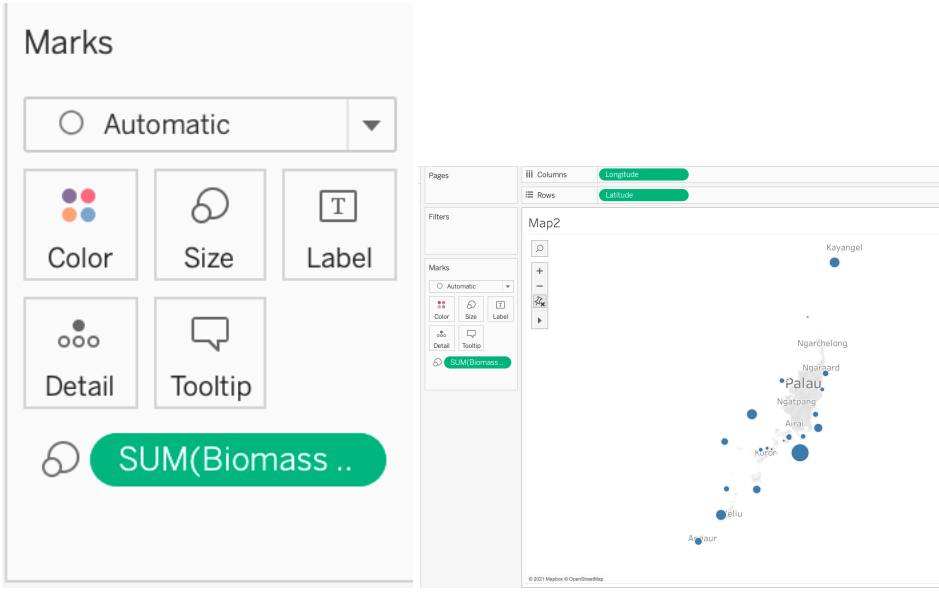
1. Learn how to create plots with multiple factors (e.g., colors, mark size, shapes).
2. Explore Tableau's mapping functionalities and learn to customize them.
3. Understand Tableau's powerful filtering capabilities.

3.3.2 Multiple Factors and Dynamic Tables: Value statement

1. In this session you will take your skills to the next level and gain full control of how your plots and maps look and how to use filters to display only the information that you are interested in. By adding multiple factors, you can communicate more information through color, size, and shapes. These are the building blocks that make a full interactive visualization!

3.3.3 Multiple factor plots using the Marks menu

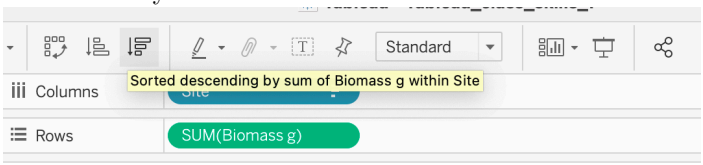
1. So far we have learned how to make simple plots, meaning that the plot has one dimension and one measure and you can modify the Marks. But what really makes Tableau awesome is that you can make interactive plots very quickly that look very appealing. Let's start with an example.
2. Create a map with multiple factors
   - In a new sheet, drag **Longitude** into **Columns** and **Latitude** into **Rows**. Name the sheet "Map2"
     - Now, let's change the size of the circles to represent the total biomass of that site. In order to make that change, drag and drop the **Biomass** into the **Size** in the Marks menu.

  - Result: 

- Troubleshooting: What if I drag something into the wrong Mark?
  - *Don't worry, there are 2 easy solutions. First, you can drag the dimension or measure that shows at the bottom of the Mark menu into the correct Mark. Second, you can simply drag the item back into the left column to delete it from the Mark menu.*
- Exploring the Legend
  - In the upper right corner of your Tableau, you should see a legend titled "SUM(Biomass g)". If you can't see it, minimize the Show Me menu.
  - If you double-click the legend, you will be able to modify even more aspects of the size marks. Go ahead and play with the ranges to explore.
- Now, let's say you want to change the color based on the habitat to show the different habitats at different sites on the map. In order to do that, drag and drop Habitat into the Color field of the Marks menu.
- Double-click the Habitat legend on the right to explore the many color palettes that Tableau offers.
  - There are two methods for choosing colors. The first is to just click **Assign Palette** and Tableau will automatically change the color of each Data Item. The second is to change each color manually. Click on the Data Item, for example **Inner_Reef**, so it is highlighted and then click on the specific color you'd like to choose.

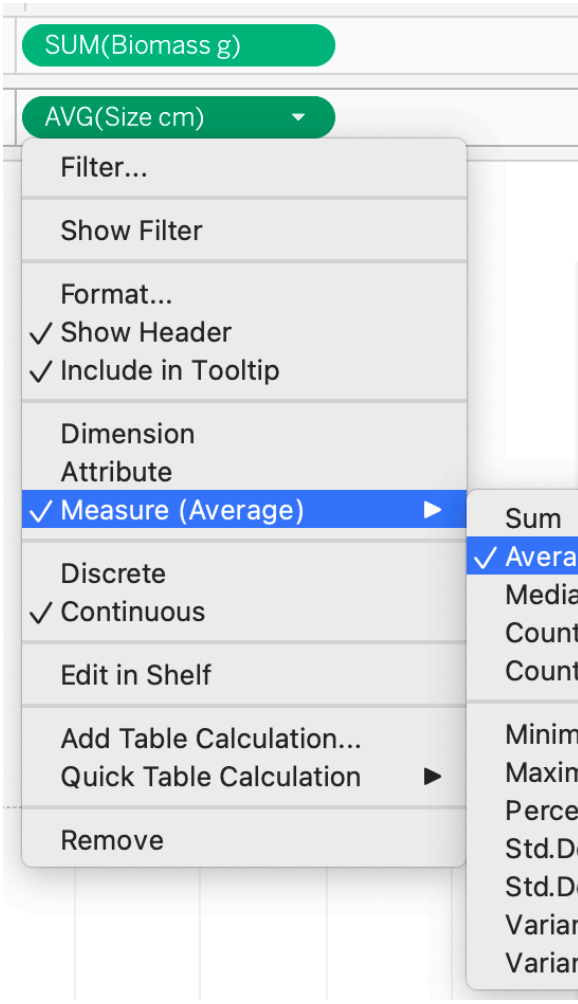
3. Create a bar plot with multiple factors
   - In a new sheet, drag **Site** into **Columns** and **Biomass** into **Rows**.
     - Name this sheet "Bars2"

  - Sort the bars in descending order. There are two potential approaches:
    - 1) Right-click the **Site** item in the Rows field. Click sort and then choose descending and close the pop out menu.
  - 2) At the top of Tableau, there is a shortcut menu with many symbols. An explanation of each function should show when you hover over the button. Click this one:



    - The arrow pointing down represents descending order.
  - Drag **Habitat** into the **Color** mark.
- You can see that this is the same representation with a bit more detail for a viewer.
- What if we want to know how different species contribute to the overall biomass of each site?
  - Remove **Habitat** from the Marks menu by dragging it to the left column.
- This is a fun example of the many possible options you have by adding dimensions and measures into the Marks menu.

4. Create a timeseries with multiple factors
5. Create a scatter plot with multiple factors
   - In a new sheet, drag **Biomass** into **Columns** and **Size** into **Rows**. Name this sheet "Scatter2"
     - Change the Size from Sum to Average by right-clicking **Size**,

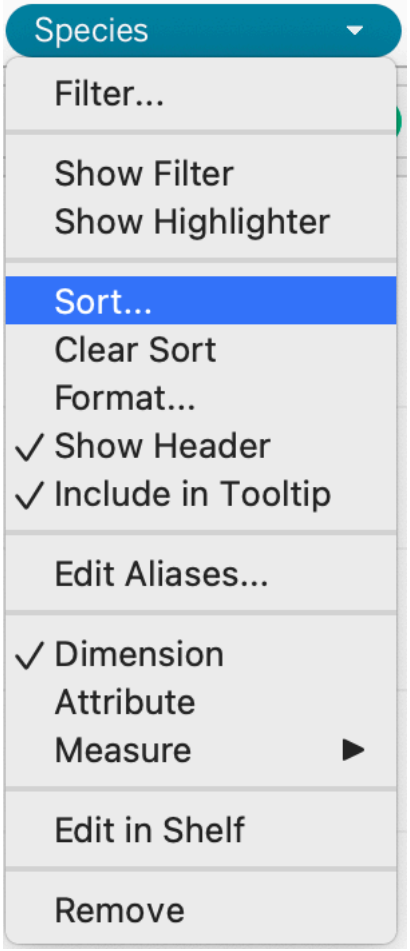


choose **Measure** and choose **Average**:

- Then, drag **Site** into **Details** in the Marks menu.
- This plot still doesn't give us much information. Let's add more information by changing the shapes of each point on the scatter plot to represent Habitat. Drag **Habitat** into **Shapes** in the Marks menu.
- *Just as you did with Colors, you can double-click Shapes and explore different shape palettes.*
- Double-click **Shapes**. Open the drop-down menu under **Select Shape Palette**: and choose **Filled**. Then click **Assign Palette** and **OK**.
- What if we want to add even more information? Let's also make each Habitat a different color. To do that, drag a new Habitat field from the left sidebar into **Color** in the Marks menu.
  - *Notice that you have two habitat fields in the Marks menu.*
- Do you notice that there is blank space at the bottom of the plot

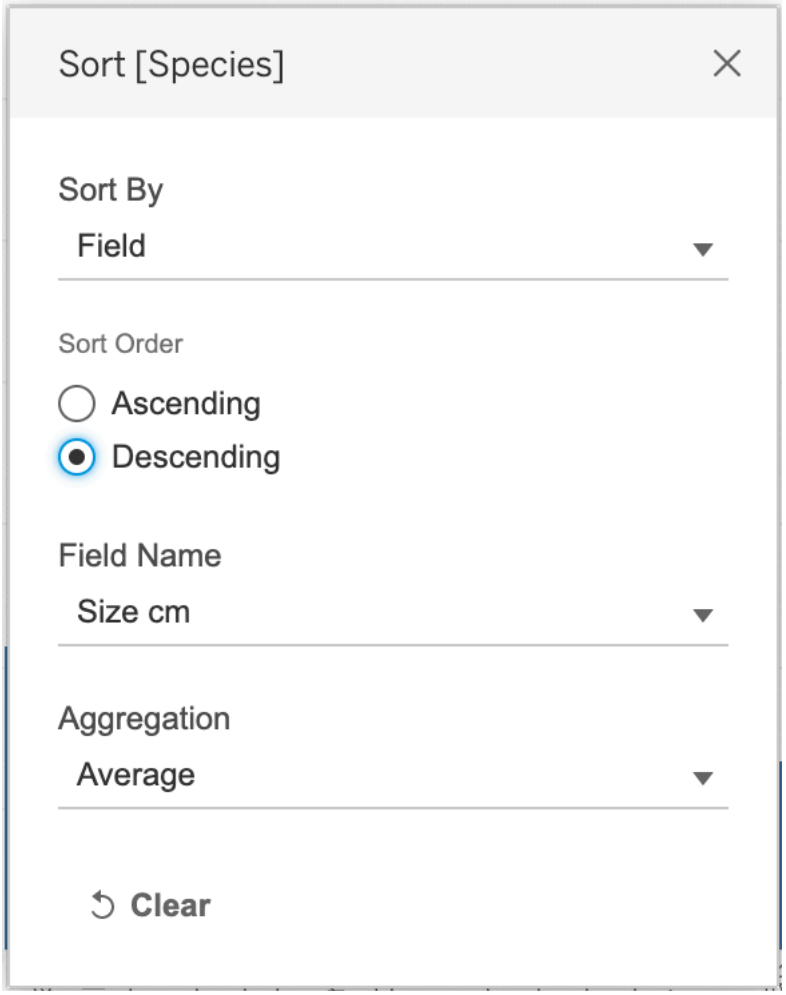
from zero to twelve? We don't really need that dead space so let's get rid of it.
  - Double-click on the **y-axis**. Uncheck **Include zero**. Exit the pop-up screen and you will see that now the plot is much clearer.

6. Create a bar plot using continuous data
   - So far we have been using discrete data in the Marks menu to customize our plots. What happens when we use continuous data? Let's find out together.
   - In a new sheet, create a bar plot with **Species** on the x-axis and average **Size** on the y-axis.
     - *Remember that you have to right-click the Size field in Rows to change it from sum to average.*
   - Then sort Species from highest to lowest Size. Right-click the **Species** field in Columns and choose **Sort**.

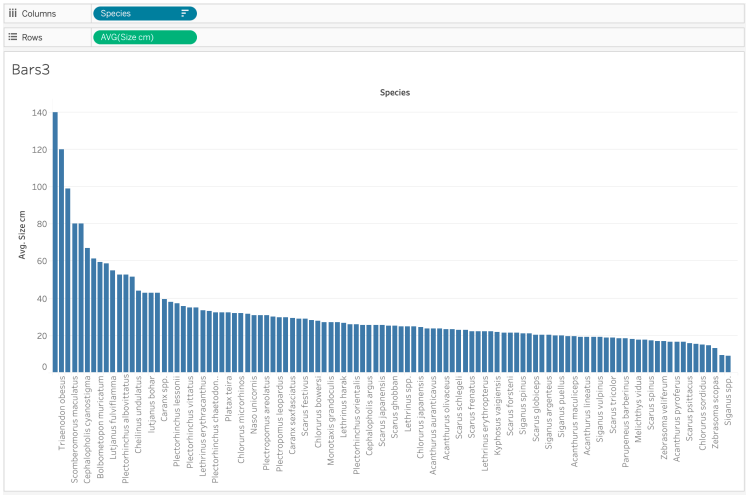


   - In the pop-up menu, open the drop-down menu under Sort

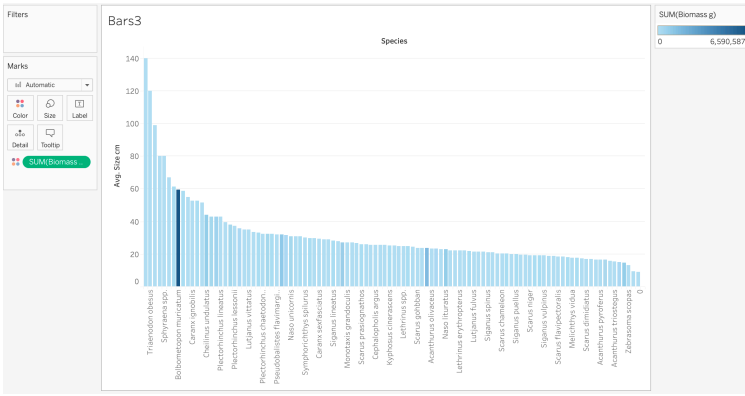
By, choose **Field**, and choose **Descending** under Sort Order.



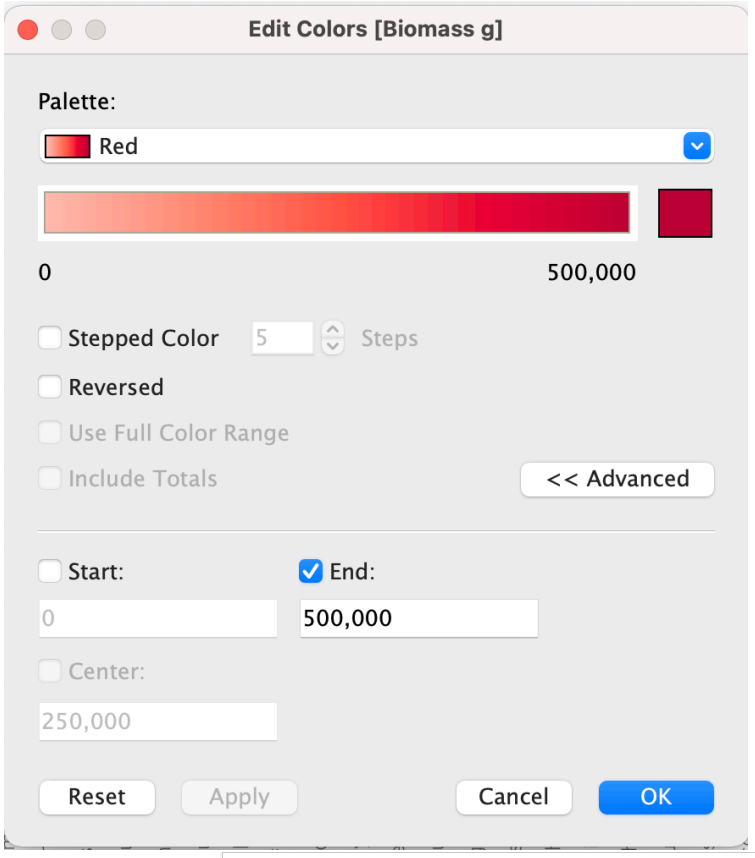
– When you close the pop-up menu, you should see this:



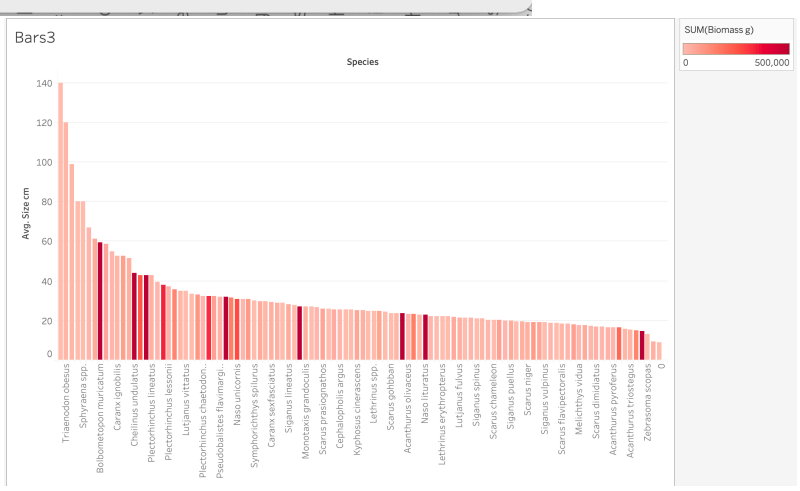
- What if now we want to know which of these species had the highest observed biomass? One way to do that is to color code the bars by biomass. So drag **Biomass** to **Color** in the Marks menu.



  - There's a huge outlier. Do you see it?
- Notice that Tableau created a color scale. In previous plots that we made using color to represent discrete data, Tableau automatically created palettes with a set number of colors depending on the number of categories in the data. In this case, Biomass is continuous data so Tableau shows a spectrum of colors to represent the spectrum of continuous data.
- To change the color palette, double-click on the legend in the upper right corner and explore the palettes that Tableau provides.
- Right now, the outlier is making it difficult for us to see the difference between other colors. Let's fix that.
  - Double-click on the legend and go to **Advanced »**. Check **End** and change the end field to **500,000**. Click **OK**.

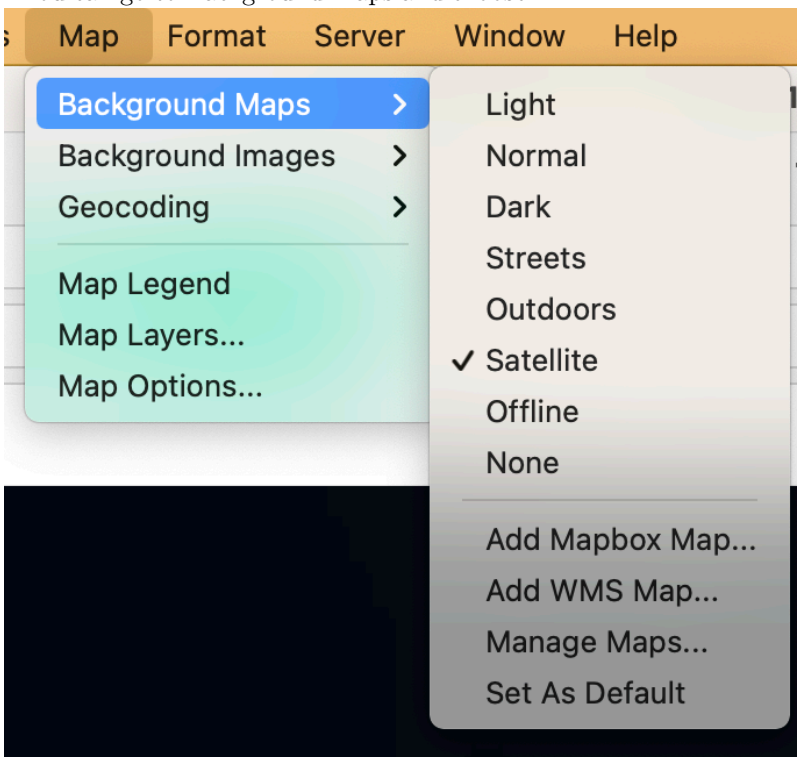


- Here is the final result:

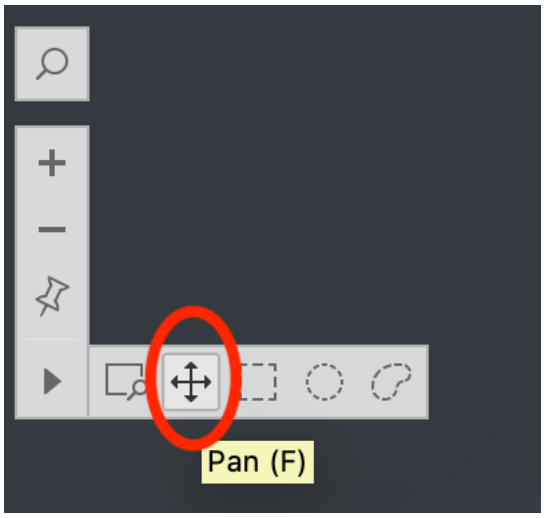


3.3.4 Maps

1. Let's explore the capabilities of maps. In a new sheet, double-click **Longitude** and **Latitude** to create a new map. Name this sheet "Map3"
   - Change the color of the points. In the Marks menu, click **Colors** and choose **orange**.
   - On the bar at the very top of your screen, you will see that there is a whole menu for Maps. You can go to Background Maps and choose from an array of styles.

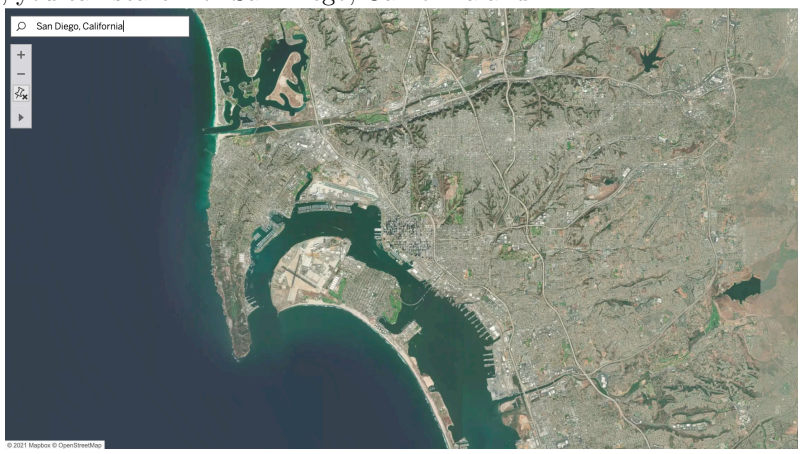


   - Open the Map Layers menu. Under **Maps**, click on **Map Layers...** and a new menu will appear on the left sidebar.
     - Under Background choose **Style: Satellite**.
     - Change **Washout** to **20%**.
       - *You can either move the slider right or you can type in the white box.*
     - *The information provided unter Data Layers only applies to the United States.*
   - The map will normally try to occupy the space and show the whole geographic scope of your data as a default. In this case, we see Palau from Kayangel to Angaur.
   - To explore the whole globe by moving the map around, hover over the triange in the upper left corner and choose **Pan**.



- Then, you can click and drag to move the map.
- You can also choose the options to the right of Pan to use selection tools that allow you to choose a group of points to highlight.
- You can also zoom in and out using the plus and minus buttons.

2. Practice exercise: Search and reset on a Map.
   - If you click on the magnifying glass, you can type in different locations to search. For example, you can search for San Diego, California and the map will show you:



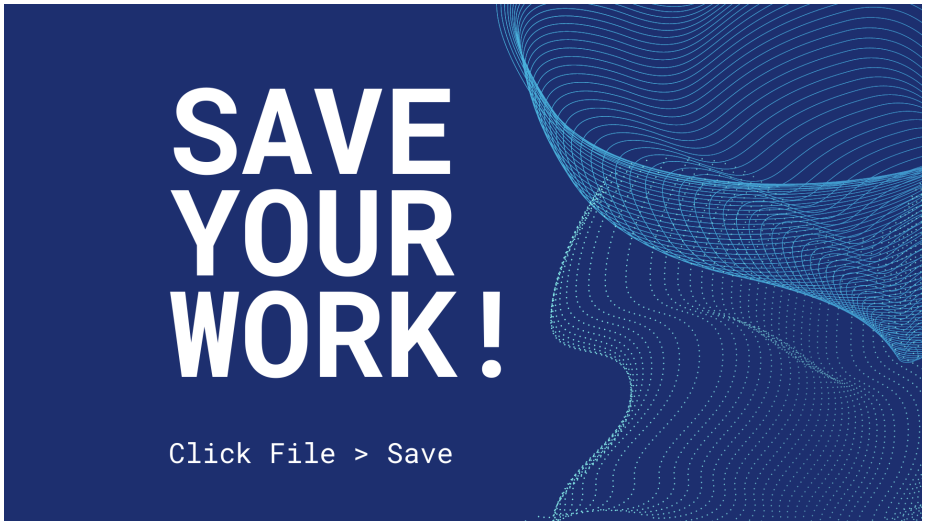
     - You can also search for the postal code of some countries.
   - Now, your turn. **Search for your hometown**. How far can you **zoom in**?
   - Say we want to return to the original view. Click on the button that looks like a thumbtack to **Reset Map**.
3. Another interesting feature of maps in Tableau is that you are not limited to coordinates. For example, your data set can include country names instead.

3.3.5 Filters

3.3.6 Dynamic Tables

1. The last thing we are going to learn in this session is how to create and use tables.
2. There are three conditions to keep in mind when you create a table in Tableau:
   - Data should be dragged into Rows because we are following a long format.
     - Refer to Module 2 to learn more about long versus block format
   - Everything has to be discrete because the software will always attempt to plot continuous data as a chart, not as text.
   - You can use filters, sort, and all the other functionalities that you learned for other types of plots. Tables in Tableau are similar to dynamic tables in Excel.

3.3.7 Save your work



3.3.8 Conclusions

1. Congratulations! You have completed one of the most important sections of this training to be able to fully customize your visualizations in Tableau. By putting together all the pieces that you have learned so far, you are ready to dive into how to make interactive visualizations to share with others. Yet before that step, our next session will teach you how to create new calculated fields to enhance the insights that you can get from your data.

3.3.9 Homework

1. Create a bar plot where x-axis is year, y-axis is total biomass, and color is habitat.
2. Create a map where size of bubbles represents avg size of fish and color of bubble represents total biomass.
3. Create a scatter plot where x-axis is total biomass, y-axis is avg size. Dots represent sites. Color and shape are the year.
4. Challenge: Pie chart where color represents habitat, size of slice represents total biomass, 1 pie chart in the same plot for each year 2012, 2013, 2014.
5. Create filters for species that allow you to select 1 species at a time in a single dropdown menu for any of these graphs.
    - Note: Make sure all axes, labels, tooltips are legible and look presentable for an audience.
6. Create a packaged workbook (.twbx).
7. Check your work by comparing it with this file.

3.4 Tableau Skills Part 3: Operations and calculated fields

In this session

1. Learning objectives
2. Value statement
3. Convert units using Calculated Fields
4. Manually calculate average
5. More functions
6. Table calculations: percentage of total and running total
7. Conclusions

3.4.1 Operations and calculated fields: Learning objectives

1. Learning the different ways to calculate custom fields in Tableau.
2. Using new calculations to extract new insights from data.

3.4.2 Operations and calculated fields: Value statement

1. In this session you will learn how to create new calculated fields in tableau, ranging from simple sums and averages, to more complex operations such as running totals and percentages.

3.4.3 Convert units using Calculated Fields

1. So far we have mostly used measures to create plots by using SUM and AVERAGE. In certain situations, however, we might need to perform some

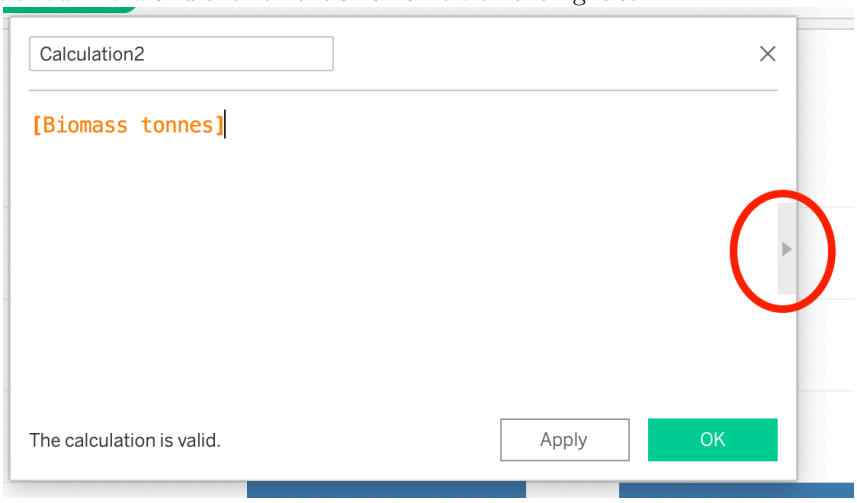
operations by using one or more variables or calculate a new field. In the example below you will see how these calculated fields can be used to change a value from grams to tonnes.

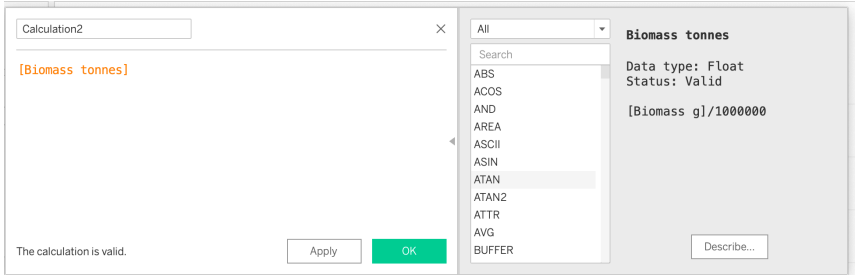
3.4.4 Manually calculate average

1. While Tableau provides an easy way to calculate the AVERAGE through using the right click functionalities on a field that is being used in an active chart, there are other ways to estimate it manually, and with more control over what the software is doing. The video below shows how to do this and how it compares to Tableau's automatic method.

3.4.5 More functions

1. While we have covered the basics of how to create calculated fields, there is a lot more to explore.
2. Open a new **Calculated Field** and click on the small arrow on the right to expand the menu.





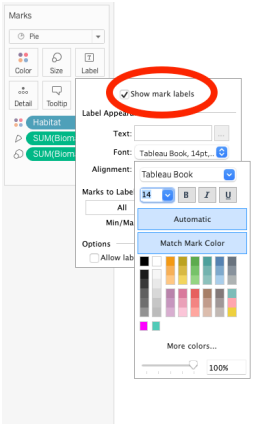
3. If you scroll through the list, you will see all the functions that Tableau is capable of performing. These functions are very similar to operations and columns in Excel.
4. We recommend that you check out Tableau's free online resources to learn more about calculated fields.

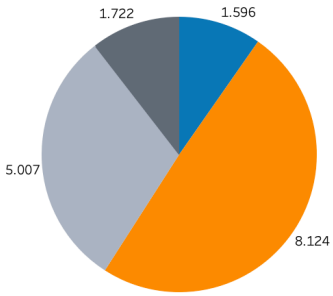
- Tableau free training videos
- Or if you are looking for tips not provided by Tableau, make sure that you are searching for operations in SQL language.

5. One common calculation in ecology is finding the average biomass at a site that has been sampled multiple times, such as by several divers each swimming transects. You can use a calculated field, yet the details of your sampling and data arrangement affect how you set it up. See an example of setting up a calculated field in the sample data set below.

3.4.6 Table calculations: percentage of total and running total

1. One additional way to perform custom calculations is by using Table calculations, which are affected explicitly by whatever is being displayed in the active chart.
2. In order to understand these better, let's do an example. Create a new pie chart.
   - Create a new sheet and name it "Pie4."
   - Drag **Habitat** into **Columns** and **Biomass tonnes** into **Rows**.
     - *A bar chart will appear* .
   - Open the **Show Me** menu and choose **pie charts**.
   - Change the view to Entire View.
   - Activate labels by clicking on **Label** in the Marks menu and check **Show mark labels** and change the font size to 14pt.





3. Next, follow along as Alfredo shows you how to calculate and display percentages of total.
4. Just like with Calculated Fields, Table Calculations are a powerful tool in Tableau's repository. If you want to explore more resources, go to:
   - https://help.tableau.com/current/pro/desktop/en-us/calculations_tablecalculations.htm

3.4.7 Operations and calculated fields: Conclusions

1. Great work! You have completed the section on how to create your own calculations and use them to enhance your understanding of the data through visualizations. You are now ready to put all of these elements together and dive right into the creation of dashboards.

3.5 Tableau Dashboards Part 1: Basic dashboards

In this session

1. Learning objectives
2. Value statement
3. Introduction to dashboards
4. Floating dashboards
5. Actions: filter and highlight
6. Conclusions
7. Homework

3.5.1 Basic dashboards: Learning objectives

1. Learn the basic windows and menus to create dashboards.
2. Explore how to make different charts interact with each other through dashboard actions.

3.5.2 Basic dashboards: Value statement

1. In this session, you will jump from creating individual charts, to creating dashboards where you can combine multiple charts and make them interact with each other.
2. You will also learn the basics for how to make your visualizations more visually appealing through the customization of colors, positions, legends, and other visual elements.

3.5.3 Introduction to dashboards

1. The very first thing to learn about dashboards is how to add different elements to it. Not everything is about the interactive charts, but also the text, symbols and other elements that help you convey a story for the reader. In this video, you will learn how to set up your canvas and add elements to it.

3.5.4 Floating dashboards

1. Another important skill to master is the correct positioning of the elements in a dashboard. While Tableau tries to help with a tiled arrangement (as in the video above), it is important that you know how to gain full control of the exact size and position of each item that you add. The video below will help you get familiar with these important skills.
2. Remember that when working with graphic designers, managing sizes and positions is extremely important. Make sure to practice these skills.

3.5.5 Actions: filter and highlight

1. Once you have added various elements, and charts, to your dashboards, it is time to make them interact with each other. The video below will show you two methods (filtering and highlighting) to use one chart to control the items in another chart. We will also cover some common troubleshooting situations.

3.5.6 Basics dashboards: Conclusions

1. Congratulations! You have completed the most important step in moving from single charts, to fully interactive data visualizations. The next step is to enhance these visuals a little more by learning how to incorporate graphic design elements.

3.5.7 Basic dashboards: Homework

1. Create a map.
    - Remember that you know how to change the background and edit colors of the dots.
2. Create a scatter plot where the x-axis is average size, y-axis is total biomass and each mark within the scatterplot is a different species.
3. Now, create a dashboard with the map and the scatterplot. Create a filter so that when you click on the map it filters the scatterplot.
    - Remember to include title and text that describes the purpose of the dashboard.
    - Feel free to create a canvas size that fits your taste. If you want a place to start, try 1000x800 (WxH).
    - Pay special attention to the positioning of your elements by using the 'layout' menu. Look at the answer key to see the exact positioning that the instructor used.
    - Remember that you can also change the font size and color of any item in the dashboard (including the color legend) by clicking on the particular element and selecing 'format'.
    - If you have trouble making the filter work, use the answer key to open up the available action in the Dashboard/Actions menu. Make sure you understand everything that is going on within the action.

    - Finally, remember that you can right click on the titles of each chart and hide them. This might be helpful to streamline the number of titles in the dashboard.
4. Try a few adjustments to practice getting the aesthetic feel just the way you want.

3.6 Tableau Dashboards Part 2: Graphic design elements

In this session

1. Learning objectives
2. Value statement
3. Recap: Tooltip customization and exact colors
4. Background images
5. Custom shapes
6. Filters in dashboards
7. Dynamic text elements
8. Creating a template
9. Conclusions

3.6.1 Graphic design: Learning objectives

1. Learn how to add custom graphic design elements to your visualizations.
2. Learn how to add custom shapes to your plots.
3. Learn how to use text interactively.
4. Create your own template for amazing-looking visualizations.

3.6.2 Graphic design: Value statement

1. You already have the tools to create visualizations and communicate your data. Let's take it to the next level by adding custom shapes, images, and other graphic design elements!

3.6.3 Recap: Tooltip customization and exact colors

1. Before we get started, let's remind ourselves of the importance of using good tooltips and the perfect colors. When working with graphic designers, ask them for the specific color code, so that you can reproduce their vision accurately.

3.6.4 Background images

1. When creating dashboards, tableau usually prefers solid colors. However, you can add your own images to make it more lively. Remember to use

PNGs with transparent backgrounds whenever possible. The video below shows how to do so.

3.6.5 Custom shapes

1. Wouldn't it be cool if you could have a little image of each fish species in your database show up on top of their respective bar plot or instead of the little circles in a scatter plot? This is your lucky day, that is exactly what the video below will show you how to do.
2. Before you begin this section, download this folder of images and save it to your desktop folder "Tableau Data Training."
3. Alfredo's tips for custom shapes:
   - *Save the files as .png with transparent background.*
   - *Save all the .png images in one folder.*

3.6.6 Filters in dashboards

1. So far, we have focused on using sheets to filter each other when together in a dashboard. However, sometimes we might want to use a field (not a sheet) to filter the information in the whole dashboard.
2. Below, you will find an example of how to do this.

3.6.7 Dynamic text elements

1. Another important aspect of using interactive elements in Tableau is that we not only have the possibility to filter charts, but also tables. Remember, Tableau is a table manager in the end. While this might not seem too helpful, the video below will show you how powerful this concept can be.

3.6.8 Creating a template

1. We strongly recommend that you develop a template with a graphic designer when possible before you create dashboards. This is the best way to enhance the quality of your dashboards. See Module 4 Session 3 for tips on working with a graphic designer.

3.6.9 Graphic design: Conclusions

1. Congratulations! You have completed all of the technical Tableau skill sessions of this training. You have come a long way, and while you may not realize it yet, you are ready to create some of those amazing dashboards that you have seen published on the web or that we saw in our introductory session to Tableau. The video below will help you notice how simple elements can add up to amazing visualizations.

2. Check out https://public.tableau.com/app/profile/datamares to explore a wide range of dashboards that you can create using Tableau.

3.7 Publishing to the web from Tableau

In this session

1. Learning objectives
2. Value statement
3. Saving your work and your datasets
4. Sharing your work
5. Embedding into a website
6. Summary

3.7.1 Publishing to the Web: Learning objectives

1. Learn how to share your content with colleagues.
2. Publish your work online.
3. Embed your dashboards in your website.

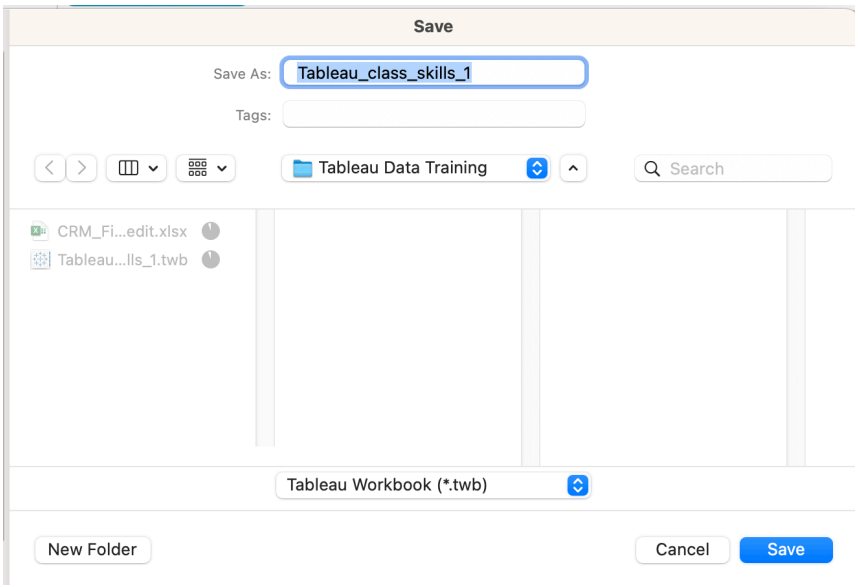
3.7.2 Publishing to the Web: Value statement

1. Tableau is a powerful tool to visualize and analyze data, however, its real value comes from being able to share this information with colleagues and the general public through the many options available.

3.7.3 Saving your work and your datasets

1. As we learned in Session 2, there are two primary ways to save your work in Tableau. The first is similar to saving any other type of file on your computer.

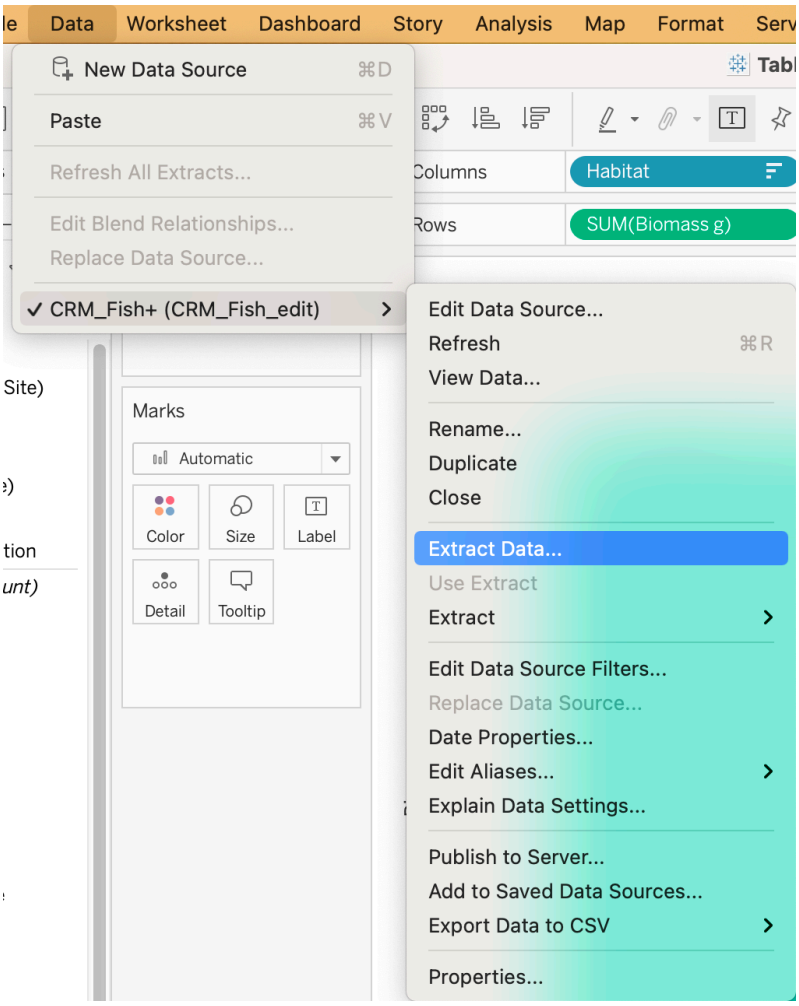
    - Click on **File > Save As** on the menu bar at the top of your screen. Name this file "Tableau_class_skills_1". Save the file under your Tableau Data Training folder.

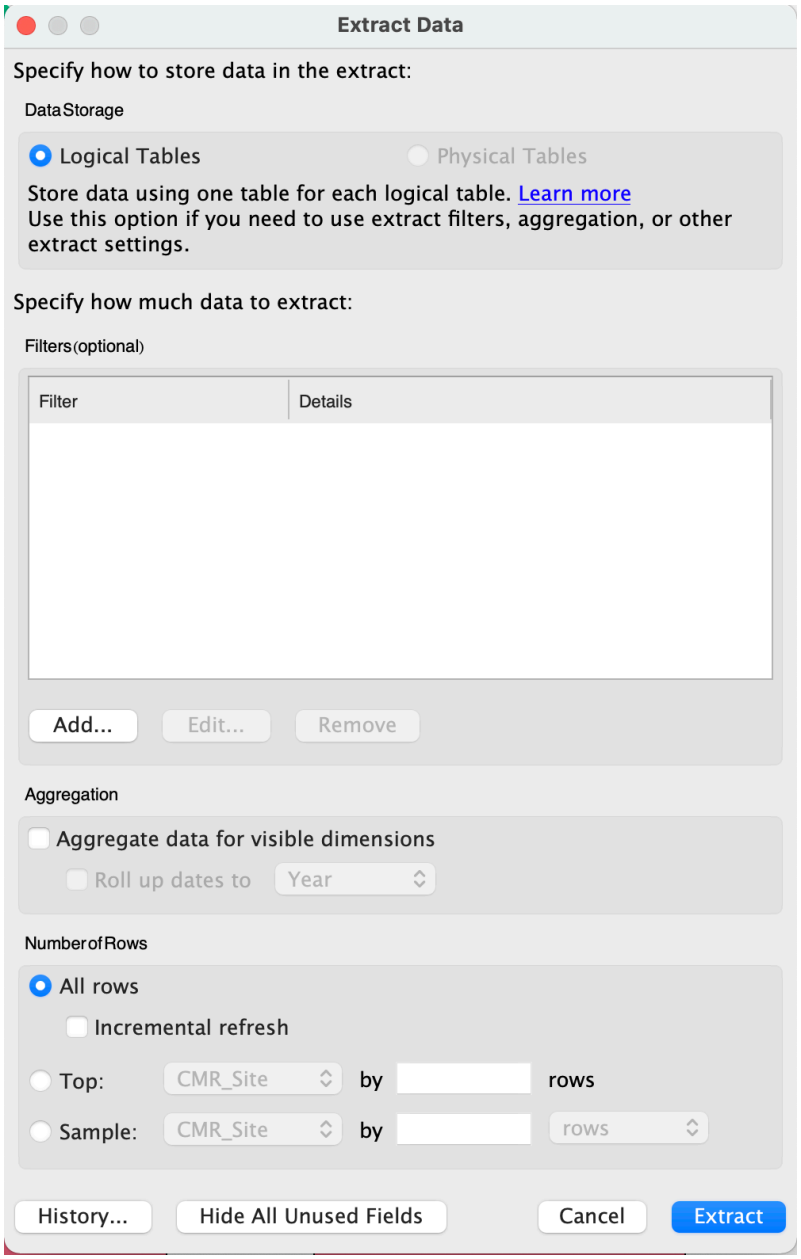


- Let's take a look at the files. If you look at the Tableau Data Training folder, you'll see that the Excel file CRM_Fish is 2.1 megabytes while the Tableau Workbook (.twb) that you just saved is 226 kilobytes. That's significantly smaller than the original Excel file because your Tableau Workbook file doesn't contain the data, it just has the instructions for what to do with the data. Therefore if you save **.twb** files (Tableau Workbook file) and you want to share them with a colleague, you also have to share the original data file.

2. Packaged Workbooks

- The second way to save a file saves a compressed version of both the data file and the Tableau Workbook. This is called a Packaged Workbook.
- First, you have to create a data extract, which puts your data in Tableau format. On the menu bar at the top of your screen click on **Data > CRM_Fish > Extract Data**


Data
Worksheet
Dashboard
Story
Analysis
Map
Format
New Data Source ⌘D
Paste ⌘V
Refresh All Extracts...
Edit Blend Relationships...
Replace Data Source...
CRM_Fish+ (CRM_Fish_edit)
Columns
Habitat
Rows
SUM(Biomass g)
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Edit Data Source...
Refresh ⌘R
View Data...
Rename...
Duplicate
Close
Extract Data...
Use Extract
Extract
Edit Data Source Filters...
Replace Data Source...
Date Properties...
Edit Aliases...
Explain Data Settings...
Publish to Server...
Add to Saved Data Sources...
Export Data to CSV
Properties...



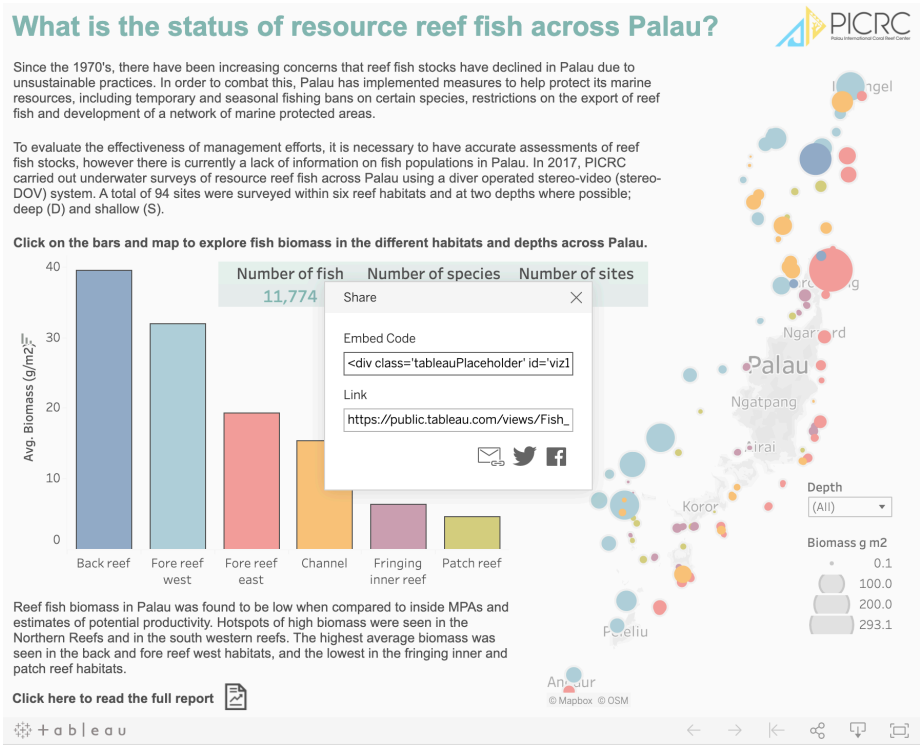
- Click extract, name this file CRM_Fish_Extract, and save it in your Tableau Data Training folder.
- Now that you have created your extract, you will create a packaged file. In order to do that, go to **File > Export Packaged Workbook > Save**
- If you look at the folder in your finder, you will see that the new

file type is **.twbx** which means it contains both the data and the instructions (Tableau Packaged Workbook file).
- If you want to share your work with colleagues who also have Tableau installed, you will share Packaged Workbooks (.twbx) with them.
- Remember that packaged workbooks will not be automatically updated when your datasets get more information. In order to do this, you will need to reconnect your file to have a "live connection" to your data source. To do this, go to **Data > CRM_Fish > Data Source** and select **Live** in the upper right corner. If the software cannot automatically locate the file, it will ask you to manually select it. Once you have selected the file and verified that your new data records are reflected in the visualization, go ahead and create a new data extract and packaged workbook with the updated information.

3.7.4 Sharing your work

1. Exporting images
   - One way to share your results is through presentations and printed reports. In these cases, you cannot use the interactive dashboard, but instead use fixed images. Once you are happy with your dashboard or worksheet's design, you can export it as an image. To export an image, click on **Worksheet or Dashboard -> Export image**. A menu will pop-up that will allow you to select the output format and directory.
2. Publishing workbooks online
   - Another way to share results is by saving your workbook and sharing it directly with another Tableau user. The limitation is that you can only share workbooks with people who have Tableau installed on their computers. Another solution is to publish it to a server and send the link to anyone you would like to look at the results. There are two kinds of services that Tableau provides to share information online: through a Public server (Tableau Public) or through a Private server (Tableau Server).
   - To publish to Tableau Public, click on **Server -> Tableau Public -> Save to Tableau Public**. This will pop-up a menu that allows you to login to your Tableau Public profile (create a free user in case you don't have one yet), write the name of the workbook and add a description.
   - To publish to a Private server (Tableau Server), click on **Server -> Publish Workbook**. This will pop-up a menu that allows you to log in to your Tableau Server profile (your institution should provide you with a username and password), select the project inside your server, write the name for the workbook and add a description.
   - To find the hyperlink that can be shared with colleagues, go to your public tableau profile, navigate to the visualization of interest and click on **Share** in the lower right corner of the visualization. Copy

and paste the provided link and share it with colleagues!



3.7.5 Embedding into a website

1. Using the embed code
   - From the same **Share** menu where you would get the link to share with colleagues, there is another option to get an **Embed Code**. This code is useful when you are trying to incorporate your visualization directly into your website. For example, if you are using an html-based website, the only thing you need to do is to copy and paste this embed code into your html code, and the interactive visualization will display in your website. (Note: WordPress is an html platform, thus you could add a new section with the Tableau Embed Code).
   - Remember that when creating the dashboard, you selected a size for your canvas. This size will determine the size of the embedded visualization, so make sure to be in constant communication with your website manager to know the ideal width and height of your visualizations.

3.7.6 Publishing to the Web: Summary

1. After completing this session you are equipped to save your datasets in Tableau format, your workbooks, and share them with colleagues who have Tableau either with .twb or .twbx files.
2. You are also ready to publish your visualizations to different services online, either through a public website or a private server. Remember to create your Tableau Public account to get access to these services. While not covered in this session, we encourage you to explore the different settings available, including how to set some of your visualizations as private even though published in the public service.
3. Finally, you are ready to make these visualizations part of your own website! Remember to be in constant communication with your website manager to make sure that you follow the right format.

Chapter 4

Storytelling

4.1 Communication Principles

In this Session

1. Learning Objectives
2. Prioritizing Engagement
3. Salience, Credibility and Legitimacy
4. Crafting your Message
5. The Message Box
6. Practice Exercise: Fill Out The Message Box
7. Resources

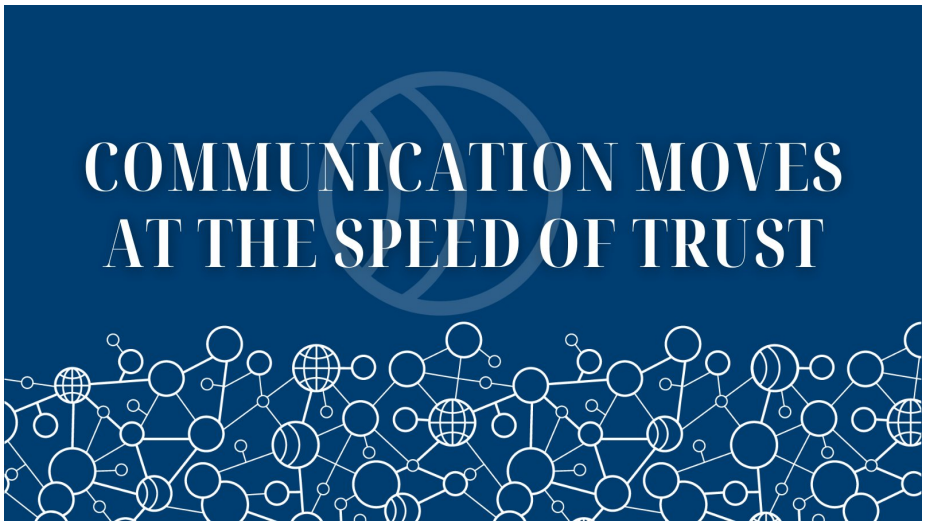
4.1.1 Communication: Learning Objectives

1. These three storytelling sessions are additional resources intended for communication professionals or those interested in expanding their skill sets beyond data management and visualization.
2. Approach science communication from an engagement framework.
3. Utilize tools like the "Message Box" for developing and crafting your message (practice exercise).

4.1.2 Prioritizing Engagement

1. Science communication is more than simply providing information to your audience. Rather, a successful communication approach will include learning what your audience desires and needs as well as building a relationship with the community you are trying to reach in order to anticipate future interests.

2. In this module we present an engagement model of communication. In a deficit model of communication, information flows one way. The scientist presents new information to an audience. On the other hand, an engagement model relies on including members of your audience as active collaborators in knowledge-building. Successful engagement means building trust and long-term relationships with the audience you hope to reach.



Source: Twitter @TheComNetwork

4.1.3 Salience, Credibility and Legitimacy: Cornerstones of Effective Communication

1. A guiding principle in effective communication to inform decision making is to consider the perceived salience, credibility, and legitimacy of the information from your audience's perspective (Cash et. al. 2002). Another way to think of these terms could be the "when," "who," and "how" of the information provided.
    - Salience: "… the relevance of information for an actor's decisions."
        - Ensure that the content you are providing is relevant to the decisions that your audience members encounter in a time frame that meets their needs. For example, there is limited value in informing a tourist who traveled to a beach that the water quality was unhealthy a week after they visited. Rather, aim to anticipate the needs of your audience members such that you can provide digestible information right when they need it.
    - Credibility: "…whether an actor perceives information as scientifically plausible and technically adequate."
        - The reputation of an organization or individual is of paramount importance when providing actionable information. Ensuring that "honest brokers of unbiased information" are the leading communicators enhances the perceived credibility of the information. This "credibility" consideration extends to every member of an organization or collaborative working group involved. Furthermore, after a period of time, audience members may desire to reflect on the accuracy of previous statements to gain an understanding of "how often have those providing the information

been right?"
- Legitimacy: "…whether an actor perceives the process in a system as unbiased as well as politically and procedurally fair."
  - In many cases, audience members strongly consider how the information was collected, analyzed and conveyed - as well as the process for engaging communities that are most closely tied to the content. This requires transparency in the process and ensuring that perspectives of impacted stakeholders are integrated into the overall message.

4.1.4 Crafting your message

1. In order to create messages that stick, brainstorm your call to action and draft your core message.
   - Messages are intended to **lead your audience to action.**
     - *Key term:* **Call to Action (CTA)**
     - What is the action you want to motivate people to do?
   - Draft before you write.

   "Consider and articulate what your story is really about. Not the noun, the verb. It's not enough to say your story is about, say, salmon. Is it a story about bears that eat salmon? Salmon that eat bears?" - Michelle Nijhuis in The Science Writers' Handbook: Everything You Need to Know to Pitch, Publish, and Prosper in the Digital Age (2013)

2. ***Good messages …***

   - Invite your audience into conversation

     - Avoid jargon
     - Instead, use descriptive language that your audience would be familiar with
     - When you can't find a perfect synonym for a jargon term, define the term or provide an analogy to something they are familiar with.
     - Consider these examples of "jargony" terms and their "audience-friendly" counterparts:
       * Anthropogenic → human-caused
       * Ecosystem services → benefits from nature
       * Harmful algal blooms → outbreak of toxic algae
       * Submerged aquatic vegetation → underwater grasses
     - We want to make our messages accessible by leveraging our audience's existing knowledge and framing the language to be socially and culturally relevant. Note that this approach is distinct from simply "dumbing down" the content.

   - Share more details about yourself to connect to the audience's values,

interests, and concerns.

  - By sharing details about yourself and your background you can build trust with your audience and relate to them on a deeper level. Personal stories and connections can aid in establishing credibility and an attachment to the issue.

- Help people see themselves in your story.

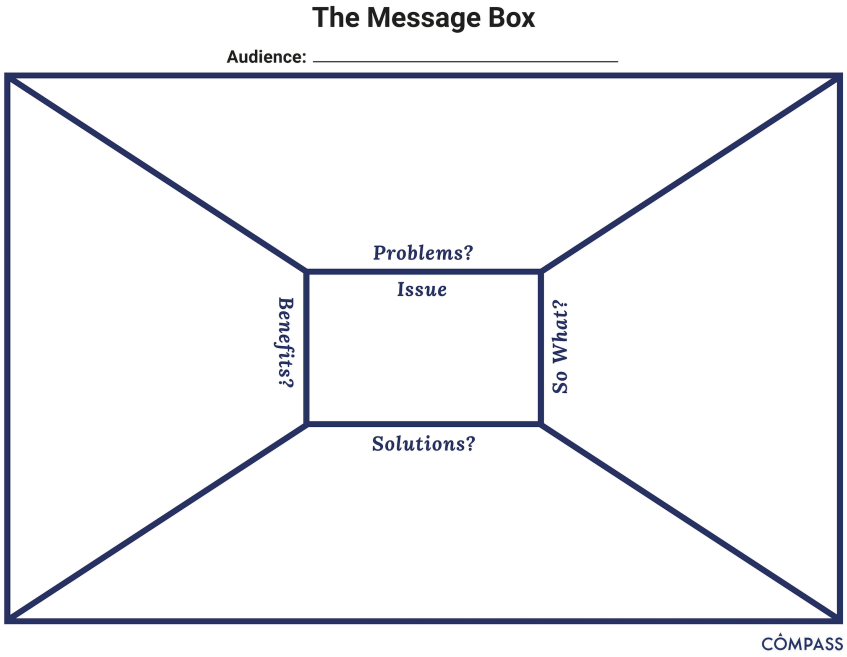
  - Find compelling hooks. A hook is the first part of a story that pulls people in. It conveys the significance and entices readers to keep reading.
    * For example: Why is it important now? What makes this story timely?
    * Can I tell this message through a story? People gravitate to stories and narrative structures. A good story captivates an audience's attention and makes the message memorable.

4.1.5 The Message Box

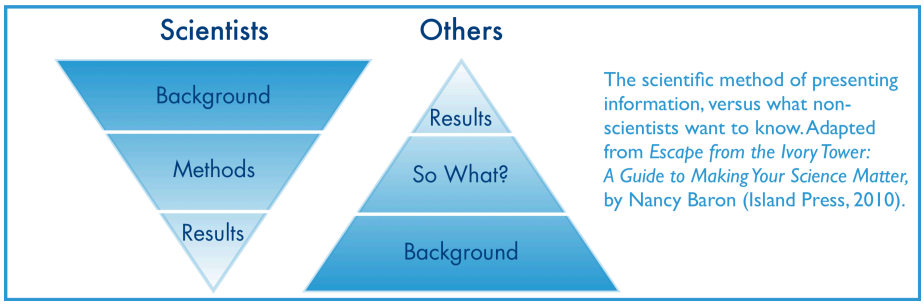
1. In preparing to communicate your scientific information, consider the use of targeted tools or practices to best focus your messages. One of the most popular tools in the science communication field is the Message Box - developed and shared by COMPASS.

2. The Message Box is a tool that helps researchers take the information they hold about their research and communicate it in a way that resonates with the chosen audience. It can be used to help prepare for interviews with journalists or employers, plan for a presentation, outline a paper or lecture, prepare a grant proposal, or clearly, and with relevance, communicate your work to others. While the message box *can* be used in all these ways, you must first identify the audience for your communication as that will drive your messaging. You would have a very different delivery for a group of scientists compared to a group of school children even with the same core message.

The Message Box comprises five sections to help you sort and distill your knowledge in a way that will resonate with your (chosen) audience. How we communicate with other scientists (through scholarly publications) is not how the rest of the world typically communicates.



In a scientific paper, we establish credibility in the introduction and methods, provide detailed data and results, and then share the significance of our work in the discussion and conclusions. But the rest of the world leads with the impact, the take-home message. A quick glance of newspaper headlines demonstrates this.



The five sections of the Message Box are provided below. For a detailed explanation of the sections and guidance on how to use the Message Box, work through the Message Box Workbook.

3. Message Box Sections

The Issue

- The "Issue" section in the center of the box identifies and describes the

overarching issue or topic that you're addressing in broad terms. It's the big-picture context of your work. This should be very concise and clear; no more than a short phrase. You might find you revisit the issue after you've filled out your Message Box to see if your thinking on the overarching topic has changed since you started.

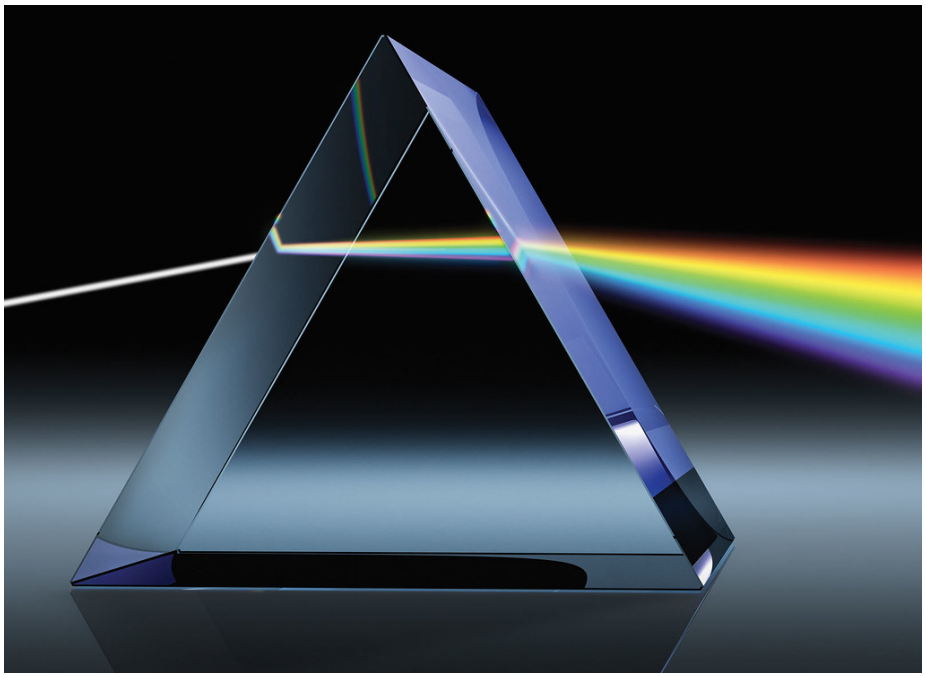
The Problem

- The "Problem" is the chunk of the broader issue that you're addressing in your area of expertise. It's your piece of the pie, reflecting your work and expert knowledge. Think about your research questions and which aspect of the specific problem you're addressing would matter to your audience. The Problem is also where you set up the "So What" and describe the situation you see and want to address.

The So What

- The crux of the Message Box, and the critical question the COMPASS team seeks to help scientists answer, is "So what?" Why should your audience care? What about your research or work is important for them to know? Why are you talking to them about it? The answer to this question may change from audience to audience, and you'll want to be able to adjust based on their interests and needs.

- We like to use the analogy of putting a message through a prism that clarifies the importance to different audiences. Each audience will be interested in different facets of your work, and you want your message to reflect their interests and accommodate their needs. The prism below includes a spectrum of audiences you might want to reach, and some of the questions they might have about your work.



The Solution

- The Solution section outlines the options for solving the problem you identified. When presenting possible solutions, consider whether they are something your audience can influence or act upon. And remind yourself of your communication goals: Why are you communicating with this audience? What do you want to accomplish?

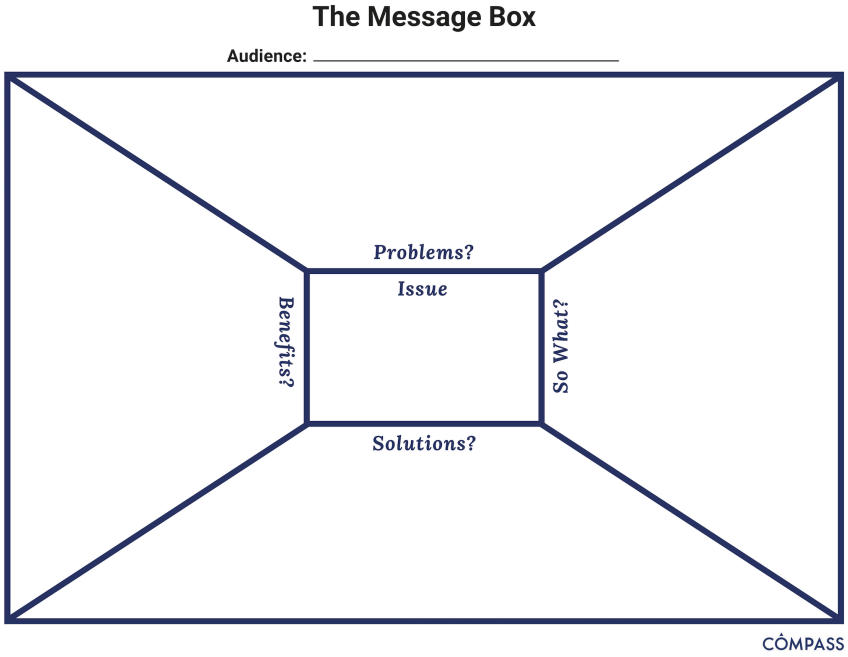
The Benefit

- In the Benefit section, you list the benefits of addressing the Problem — all the good things that could happen if your Solution section is implemented. This ties into the So What of why your audience cares, but focuses on the positive results of taking action (the "So What?" may be a negative thing — for example, inaction could lead to consequences that your audience cares about). If possible, it can be helpful to be specific here — concrete examples are more compelling than abstract. Who is likely to benefit, and where, and when?

- In addition to the Message Box Workbook, COMPASS provides resources on how to increase the impact of your message (include important statistics, draw comparisons, reduce jargon, use examples), exercises for practicing and refining your message and published examples.

4.1.6 Practice Exercise: Fill Out The Message Box

1. For your own practice, fill out a sample Message Box with an example audience from your own work (e.g. a local school, government agency, or even internal leadership). You can download a blank message box or draft your own.
    - Fill in your audience at the top of the Message Box.
    - Fill out the Message Box's 5 sections: "Issue", "Problems", "So What?", "Solutions", and "Benefits".
    - Practice providing your talking points for your latest project with a colleague.



4.1.7 Communication Resources

1. If you'd like to learn more, watch DataONE Webinar: Communication Strategies to Increase Your Impact from DataONE on Vimeo.

4.2 Data Storytelling

In This Session

1. Learning Objectives
2. Value Statement

3. New Terms
4. Receive Your Data
5. Identifying Your Audience
6. Understand Key Insights
7. Connect Stories to Key Insights
8. Utilize Engaging Visuals
9. Share Your Data Stories
10. Summary

4.2.1 Storytelling: Learning Objectives

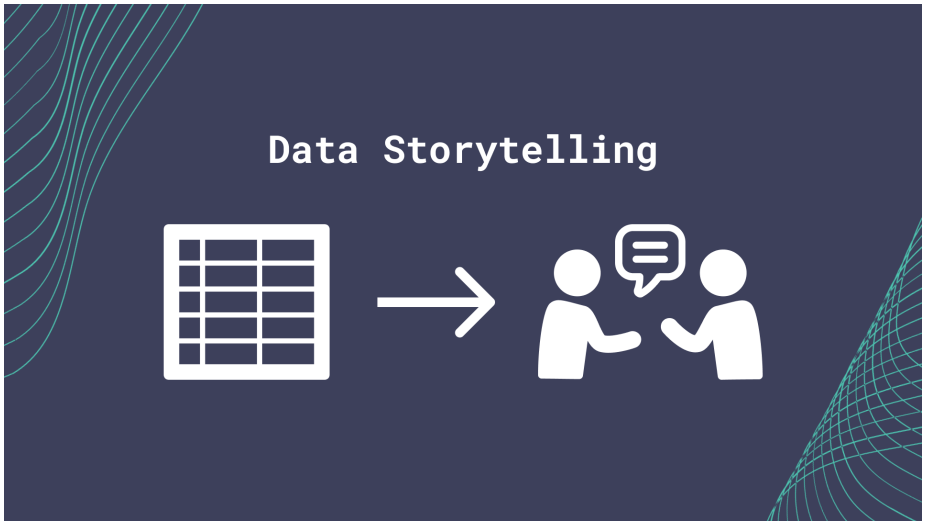
1. To learn the importance and potential of data storytelling.
    - ***Data storytelling:*** *The process of translating data sets into stories, or messages that your target audience can understand.*
2. To learn a sample of approaches to data storytelling, including connecting stories and visuals to data for maximum effectiveness.

4.2.2 Storytelling: Value Statement

1. Do not let your data go to waste! Learning how to effectively share engaging stories from data will allow your data, research, and synthesis to reach an increased range of audiences.
2. Not everyone knows how to derive key insights and stories from your data, so increasing your data storytelling capacity will allow you to increase understanding and engagement from both internal and external organizational partners.

4.2.3 New Terms

1. ***Data storytelling:*** The process of translating data sets into stories, or messages that your target audience can understand.



4.2.4 Receive Your Data

1. The first step in effective data storytelling is to ensure you are working with clean, standardized data:
    - This section assumes that the data you have to tell stories from are clean and standardized already. If your data are still raw and not cleaned, please refer to *Module 2* in this training curriculum, titled *"Data"* to learn more about how to clean and standardize your data set.

4.2.5 Identify Your Audience

1. Knowing your audience is one of the primary steps in effective communication.
2. Crafting good messages is about building connections. The science you are representing will be more accessible to the communities who need it if messaging is customized for each audience.

3. Take time to think about who your audience is for this project:
   - Is this for local school kids, or for large international agencies?
   - What are your audience's needs or goals?
   - How can members of your audience help co-create these materials?
4. To help facilitate your thought process on who your audience is, we have included a handout and self-guided practice exercise below for you to use as a tool:
   - ***"Identify Your Audience" Handout + Exercise***
5. What considerations are most relevant as we create these data stories? Drafting a short list of these considerations can be a helpful starting point when initiating a data visualization process.

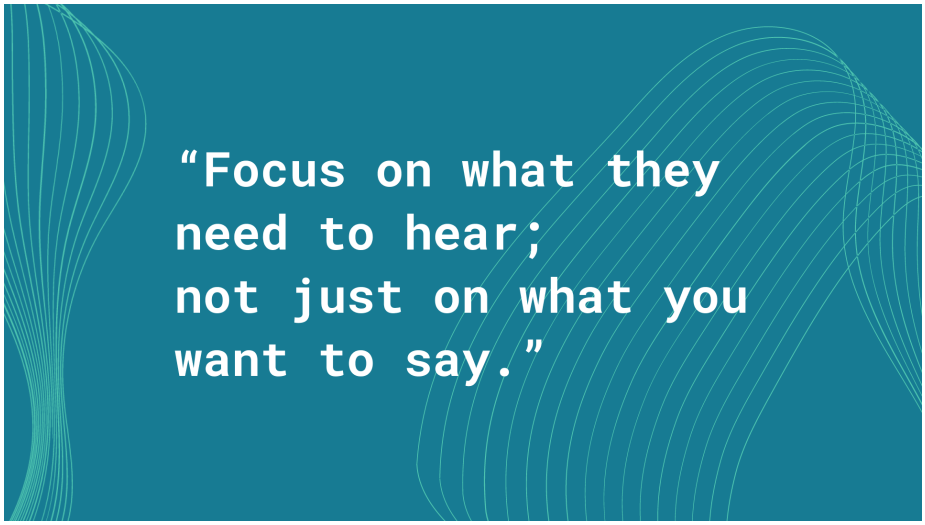
4.2.6 Understand Key Insights

1. After receiving your clean, standardized data, and identifying your audience, reserve time to review your data to discover key data insights that would be highly relevant to your audience:
   - *For example, if you are creating material for your internal conservation agency, look and see if there are any trends or relationships between your conservation practices and your measured conservation performance indicators.*

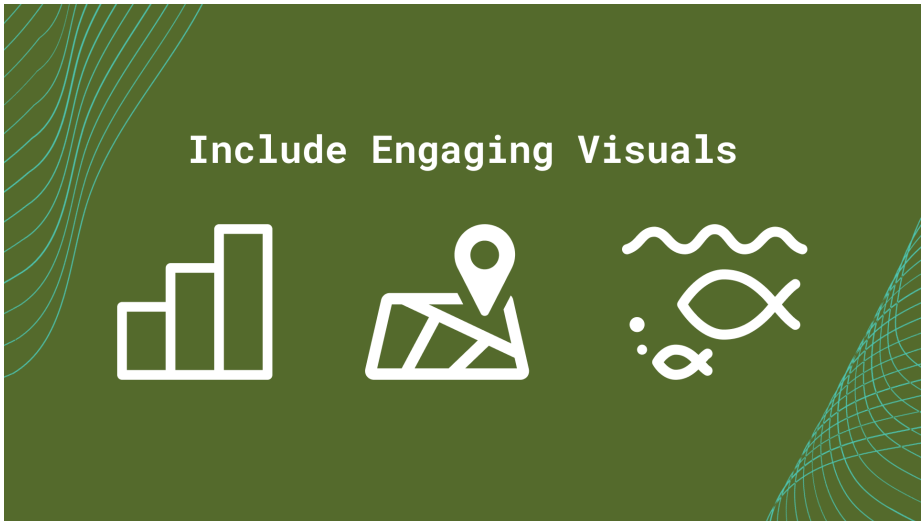
2. Determine insights that will draw the attention of your audience because these can amplify both the relevance and impact these insights will have.

4.2.7 Connect Stories To Key Insights

1. Even if you have convincing graphs or numbers behind these key insights, it can still be hard for audience members to relate personally to these messages.
2. Use relatable stories to bring key insights to life:
    - *For example, if you are sharing data insights on conservation efforts working with local fisheries, consider including a story of how local fishery members are reacting to recent conservation policy. Perhaps multiple stakeholders note the growing pains of the transition yet later begin to receive and recognize the value that results from the revised actions. The different perspectives noting the implications of the policy on their daily lives—in the short and long term—would make the material more approachable and engaging.*
3. Include these relatable, or personal, stories and anecdotes to more fully engage your audience in the story of your data.
4. If you have difficulty selecting which stories to share, think about your target audience, and choose stories that they will likely relate to, or resonate with. Focus on what they need to hear; not just on what you want to say.



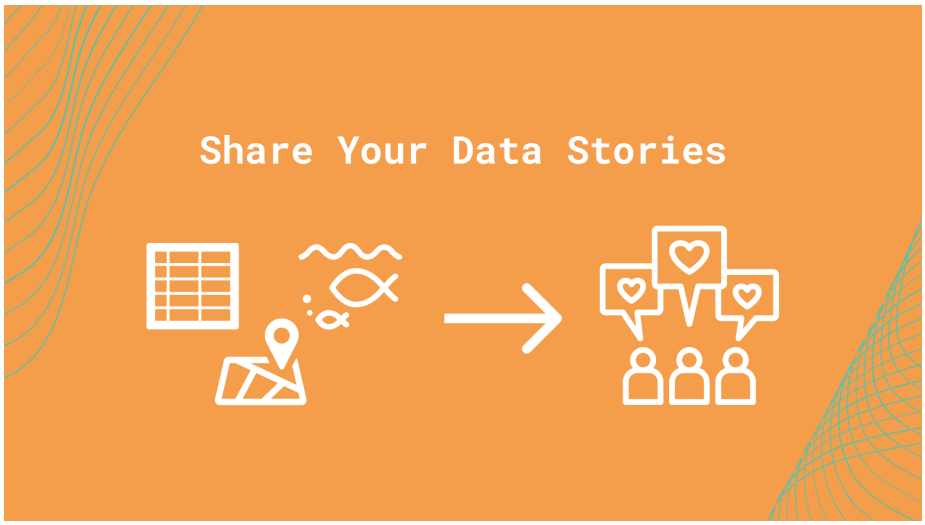
4.2.8 Utilize Engaging Visuals



1. Furthermore, some audience members might only become more fully engaged with your data if you include engaging visuals—like graphs, maps, or other data visualizations—in your data storytelling.
2. Data can be hard to visualize mentally, so having graphs or maps showcasing data for an audience can alleviate a lot of mental labor your audience will have to do:
    - *For example, audience members may not remember all the places a certain fish species lives, but having a map that shows the picture of the fish next to each of its local habitats can easily illustrate where this particular fish species is known to exist.*
3. Remember to focus on visuals or aesthetic details that will resonate with your audience.

4.2.9 Share Your Data Stories

1. The last step is to share your data stories with your audience.



2. In whatever way is most appropriate, share your data stories with your target audience. Whether you made a recorded Zoom presentation, or an Instagram post, share your data stories so that all this work does not go to waste:
    - *For example, imagine you recorded a short Instagram live video to share a data story on national conservation efforts for mangroves. If your goal is to increase general public awareness, then you could send that video along to local community organizations like schools, work places, or any other frequented community space.*
3. The main point in data storytelling is to encourage you to develop and share your data stories with the people who need them! Do not let the effort you and your colleagues have put into your research remain as untold stories.

4.2.10 Data Storytelling: Summary

Distilled details with the key messages from this module are in this downloadable brief, to be used as a reference guide.

1. Receive Your Data
2. Identify Your Audience
3. Understand Key Insights
4. Connect Stories To Key Insights
5. Utilize Engaging Visuals
6. Share Your Data Stories

4.3 Working with Graphic Designers

In This Session

1. Learning Objectives
2. Value Statement
3. New Terms
4. Before You Start
5. Know Your Audience
6. Define Your Vision
7. Leave Execution to The Designer
8. Continue to Iterate
9. Finalize and Distribute Your Design
10. Summary

4.3.1 Working with Graphic Designers: Learning Objectives

1. Learn how to work effectively and collaboratively with a designer in the context of a targeted design project.
2. Learn how to facilitate a successful design project from start to finish with a designer.

4.3.2 Working with Graphic Designers: Value Statement

1. Working effectively with graphic designers (designers for short in this session) will save you and your organization time and resources, particularly with consistent communication and clear expectations.
2. Learning how to work with designers will allow you to produce effective, meaningful and compelling visual content for your internal organization, or for external audiences.
3. Utilizing designers elevates the effectiveness of your communication through visual storytelling expertise.

4.3.3 Working with Graphic Designers: New Terms

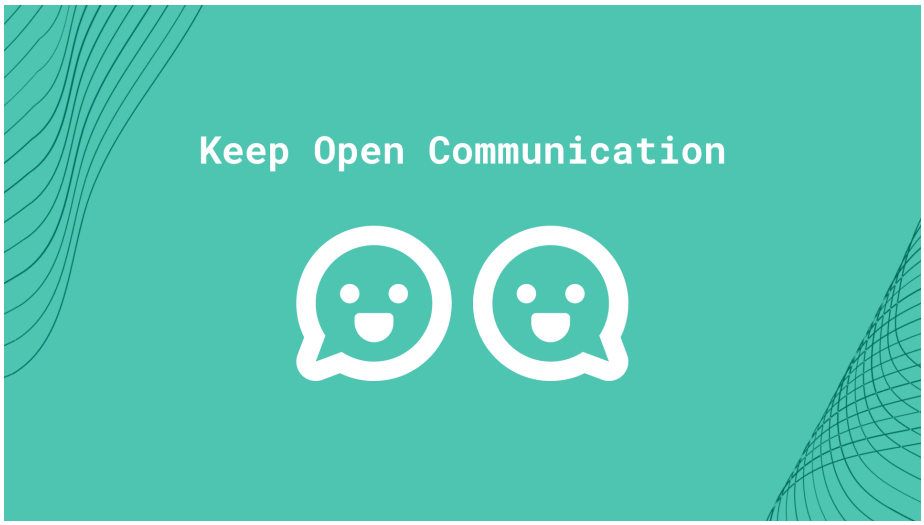
1. ***Deliverable:*** A tangible or intangible good or service produced as a result of a project that is intended to be delivered to a project manager/customer (internal or external).
2. ***Design assets:*** In terms of web design and development, "assets" typically refer to the text content, graphics, photographs, videos, audio files, and databases (for example, accessing an organization's photo bank and icon libraries can be very beneficial for designers in producing work that has support permissions and image rights). Past design projects can also be useful resources.
3. ***Writing norms:*** A document any staff can refer to for an organization's guidelines for written materials.
4. ***HEX code (digital):*** a six-digit, three-byte hexadecimal number used in web and digital design to represent colors (for example, the HEX code for a shade of dark blue is: #177B93).

5. ***RGB (digital):*** Stands for "Red Green Blue." RGB refers to three hues of light that can be mixed together to create different colors. Combining red, green, and blue light is the standard method of producing color images on screens, such as TVs, computer monitors, and smartphone screens (for example, the RGB for a shade of purple is: R: 132, G: 17, B: 170).
6. ***CMYK (printing):*** Stands for "Cyan Magenta Yellow Black." These are the four basic colors used for printing color images. Unlike RGB (red, green, blue), which is used for creating images on your computer screen, CMYK colors are for pigments (for example, the CMYK for a dark green is: C: 100, M: 30, Y: 73, K: 52).
7. ***Iterative design:*** A circular design process that models, evaluates and improves designs based on the results of testing.

4.3.4 Before You Start

1. Establish common expectations with the designer at the beginning of the project. For example, some common expectations you would need to establish with the designer include:
   - What do you need them to do for this project? Be as specific as possible while creating space for creative energy.
   - What deliverables and timelines are you expecting?
     - *__Deliverable__: A tangible or intangible good or service produced as a result of a project that is intended to be delivered to a project manager/customer (internal or external).
     - For example, a deliverable from your graphic designer could be a logo design, or educational graphics for a slide presentation.*
   - When and how often will you meet to complete the project?
   - Are there visuals or details that may be sensitive to the audience that the designer should be aware of?

2. Introduce your designer to your organization's current design assets:
    - ***Design assets:*** *In terms of web design and development, "assets" typically refer to the text content, graphics, photographs, videos, audio files, and databases (for example, accessing an organization's photo bank and icon libraries can be very beneficial for designers in producing work that has support permissions and image rights).*
    - Let the designer know about designs or design assets that your organization already has to utilize.
        - *For example, show the designer where to find your organization's current logo and flyer designs.*
        - *Another example of a design asset you can share with a designer is your organization's writing norms.*
        - ***Writing norms:*** *A document any staff can refer to for an organization's guidelines for written materials - to aid in copy editing processes.*
    - Helping a designer know what your organization already has in terms of design material allows them to better design future material, as well as not duplicate efforts that have already been completed.
3. Keep open communication with the designer throughout the project:
    - Communicate with the designer regularly at scheduled times, or as needed, not just at the beginning and end of a project.
    - These meetings can even be as short (~10 minutes) or long (60+ minutes) in duration according to your and the designer's needs.



4. Give the designer sufficient time and space to do their best work:
    - After giving sufficient direction to the designer, let them take the design execution from there and do what they know best: design!
    - Give the designer time and space to work unsupervised and trust in their specific skill set and ability to execute the design.

- Let the designer adapt your direction to their unique style and vision.

4.3.5 Communicate Audience Needs

1. Communicate your target audience and their needs to the designer:
   - Ensure you and the designer are on the same page for whichever audience you are trying to design for.
   - Share with the designer essential information about your target audience. This can include:
     - Audience: age, gender, organization, occupation, geographic location, nationality, economic needs, work mission, personal goals and more.
   - These pieces of audience information can be very helpful in determining what is appropriate for this design project.

4.3.6 Define Your Vision

1. Given your audience and their needs, envision what you want designed for this project.
   - For example, would your audience prefer materials as slides, posters or videos?
   - What colors, or other sources of inspiration, would they take a liking to?
     - *Tip: When communicating color choices with a designer, sharing a color's specific code (such as a HEX code for digital designs) is extremely helpful. It allows the designer to see exactly what type color you are envisioning.*
     - Below are 3 types of color codes (two are digital, and one is for physical printing):

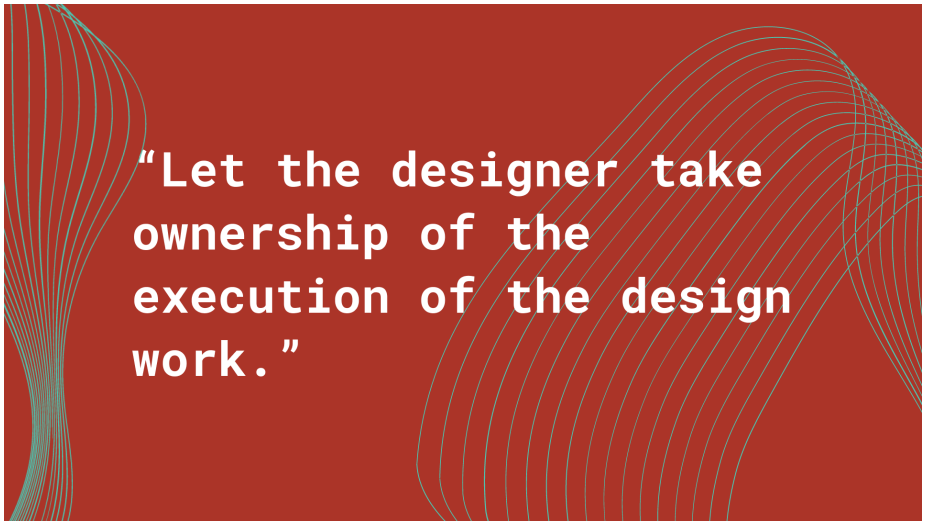
    - **HEX code (digital):** a six-digit, three-byte hexadecimal number used in web and digital design to represent colors (for example, the HEX code for a shade of dark blue is: #177B93).
    - **RGB (digital):** Stands for "Red Green Blue." RGB refers to three hues of light that can be mixed together to create different colors. Combining red, green, and blue light is the standard method of producing color images on screens, such as TVs, computer monitors, and smartphone screens (for example, the RGB for a shade of purple is: R: 132, G: 17, B: 170).
    - **CMYK (printing):** Stands for "Cyan Magenta Yellow Black." These are the four basic colors used for printing color images. Unlike RGB (red, green, blue), which is used for creating images on your computer screen, CMYK colors are for pigments (for example, the CMYK for a dark green is: C: 100, M: 30, Y: 73, K: 52).
  - If your audience may contain people affected by color vision deficiency, consider reviewing your images using a color-analyzing tool, such as COBLIS.
2. Bring your ideas together in a slide deck, vision board, collage etc. so you can communicate your specific vision to your designer.

3. After compiling your vision materials, communicate your specific vision to your designer to get them on the same page.
    - This vision does not need to be final or entirely complete, yet the vision must be specific enough to communicate the general direction and idea to the designer so that they can start to execute, or iterate on, the vision effectively.

- ***Iterative design:*** *A circular design process that models, evaluates and improves designs based on the results of testing.*

4. In addition, your designer will likely also have some helpful ideas for the vision of the project. Remain open to innovative ideas as they come up!
   - Whether new ideas come from you, the designer, your client etc., follow the ideas that most effectively meet the needs of your target audience.
     - *For example, perhaps when meeting with a client, the client themselves comes up with a great idea for their organization's upcoming slide presentation. It is fine for you to take that idea, cite the client in an appropriate way, and execute it if it seems like the idea will meet the organization's training needs.*
   - If someone has a better idea than you, embrace it! Support that person and their energy in pursuing that idea. Humility is a beneficial trait to have in brainstorming and multidisciplinary teamwork environments.

4.3.7 Leave Execution To The Designer



1. After communicating your project vision to the designer, let the designer take ownership of the execution of the design work.
   - Trust the designer's process and artistic judgement.
   - Let the designer execute the design, from draft to final product.
     - *For example, let the designer handle all formatting specifics, hex codes for colors, margins etc. Those are things you can leave to your designer, unless you have specific requirements requested by your target audience.*
2. Continue to have regular check-ins with your designer to assess project progress and evaluate current designs.

3. If you do have concerns about the designs, continue to communicate those openly with your designer.
    - Resolve design issues and problems together when they arise.
        - *For example, one issue that could arise is that the designer may always be delivering designs late. Communicate openly with your designer to see why this might be, and set new achievable goals and deadlines together.*

4.3.8 Continue To Iterate

1. As your designer continues to give you scheduled updates, ensure you give honest, actionable feedback.
    - Honest, actionable feedback can be a great support in improving designs, and focusing designs on target audience needs.
        - *For example, one example of actionable feedback can be: "I think we should use colors that are familiar to this school audience for this outreach flyer. Perhaps we can use their school's colors."*

Bad feedback:
"This flyer looks horrible. Can you change it to look better?"

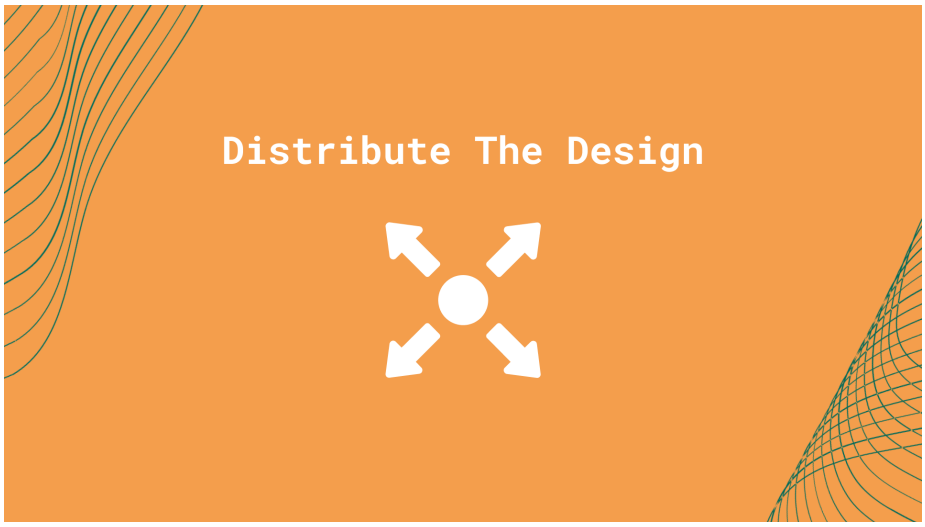
Actionable feedback:
"These colors seem to clash too much for this flyer. Can we try using more monochrome ones?

Bad feedback:
"These logos look good. I just wish they were better, you know?"

Actionable feedback:
"I like this initial logo direction. Maybe we can include local motifs to up cultural appeal?"

2. Continue to iterate off each other's ideas.
   - ***Iterative design:*** *A circular design process that models, evaluates and improves designs based on the results of testing.*
   - There can be many drafts of designs over the course of a project.
   - Continue to be open to new and better ideas as you make progress in your design project.
   - Embrace evolution! Pivoting or refining a design for a project is a normal, frequent occurrence.

4.3.9 Finalize and Distribute Your Design



1. Once you and your designer both feel comfortable with the final design

products, move forward with distributing the design to your audience.
2. Distribute your design to your target audience in the appropriate method. Distributing your design could look like:
    - *Emailing your slide presentation to your target environmental organization's leadership team.*
    - *Printing out your posters and delivering the posters to local elementary school classrooms.*
3. As a last step, take note of how your target audience reacts to or uses your materials.
    - This feedback can help you and the designer in future projects, and also be used to improve long-term relationships with these organizations or people that you are working with.

4.3.10 Working with Graphic Designers: Summary

Distilled details contained in this brief

1. *Before You Start - Set Expectations*
2. *Know Your Audience*
3. *Define Your Vision*
4. *Leave Execution to The Designer*
5. *Continue to Iterate*
6. *Finalize and Distribute Your Design*

Chapter 5

Conclusion

In this Session

1. Content Review
2. Resources
3. References

5.1 Content Review

1. Congratulations on completing the training materials. The content from this training package is intended to be an additional compilation of resources to aid you in your personal and professional endeavors to convey compelling visual stories from well-managed data sets. If you have any feedback or comments on this training that you'd like to share, please use this feedback survey.
2. While the materials can be valuable for an individual, greater value can come from collaborating over these materials with colleagues with related roles throughout your organization. Remember that there are many roles in data science communication and the information passes through many phases and hands before ultimately getting to a target audience. Familiarity with your "data role," and the role of your work partners will ultimately aid in the full data storytelling process.
3. While this training can be an informative leap into data sciences, the greatest experience comes with repetitive practice on new topic areas. Consider these resources as a reference point to be reviewed intermittently when you need a reminder about key steps or best practices in managing, visualizing, or communicating data messages.

5.2 Resources

1. The table below provides direct links to all of the files from each Module and Session. The full zipped package can be accessed and downloaded using this link.

Module	Session	Relevant Resources
2. Data	1. Data Management	• Metadata template
	2. Data Cleaning	• Data cleaning checklist
	3. Normalization and Standardization	• Data Cleaning Exercise and Key
3. Tableau	1. Tableau Basics	• CRM_Fish_edit
	2. Tableau Skills: Part 1	• M3S2 Progress Check Workbook
	3. Tableau Skills: Part 2	• M3S3 Progress Check Workbook
	4. Tableau Skills: Part 3	• Biomass as a Calculated Field Slides
	5. Tableau Dashboards: Part 1	• Map Making Homework File
	6. Tableau Dashboards: Part 2	• M3S6 Progress Check Workbook
	7. Publishing to the web	
4. Storytelling	1. Communication Principles	• Message Box Workbook

	• Message Box Blank
2. Data Storytelling	• Identify Your Audience Handout and Exercise • Data Storytelling Brief
3. Working with Graphic Designers	• Working with Graphic Designers Brief

5.3 References

- Cash, David, William C. Clark, Frank Alcock, Nancy M. Dickson, Noelle Eckley, Jill Jäger. "Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making," KSG Working Papers Series, 2003. http://nrs.harvard.edu/urn-3:HUL.InstRepos:32067415
- Lowndes,Julia S. Stewart, Benjamin D. Best, Courtney Scarborough, Jamie C. Afflerbach, Melanie R. Frazier, Casey C. O'hara, Ning Jiang, Benjamin S. Halpern. "Our Path to Better Science in Less Time Using Open Data Science Tools," Nature Ecology & Evolution 1, 0160 (2017) https://doi.org/10.1038/s41559-017-0160
- Borer, Elizabeth T., Eric W. Seabloom, Matthew B. Jones, Mark Schildhauer. "Some Simple Guidelines for Effective Data Management," Bulletin of the Ecological Society of America 90, no. 2 (2009): 205–214. https://doi.org/10.1890/0012-9623-90.2.205
- White, Ethan P., Elita Baldridge, Zachary T. Brym, Kenneth J. Locey, Daniel J McGlinn, Sarah R. Supp. "Nine Simple Ways to Make it Easier to (Re)use Your Data," Ideas in Ecology and Evolution 6, no. 2 (2013): 1–10. https://doi:10.4033/iee.2013.6b.6.f
- Hayden, Thomas, and Michelle Nijuis. The Science Writers' Handbook: Everything You Need t oKnow to Pitch, Publish, and Prosper in the Digital Age. Da Capo Lifelong Books, 2013.
- "COMPASS homepage," accessed September 29, 2021, https://www.compasscicomm.org/