Aligning Course Assignments to Fulfill IS2020 Competencies

Jonathan P. Leidig
jonathan.leidig@gvsu.edu
School of Computing
Grand Valley State University
Allendale, MI 49401

Abstract

Educators are tasked with continually updating course objectives, content, assignments, and assessment to meet model curriculum guidelines. IS2020 proposes program level outcomes for required and elective areas. Two elective areas in IS2020 are Data and Business Analytics and Data and Information Visualization. IS2020 details 14 program level competencies (organized within knowledge elements and skills) that are then integrated into individual course-level design. This work presents a set of laboratory exercises to fulfill the competencies of both elective areas. The set of exercises have been taught in the classroom over several years and have been refined to evaluate coverage of the 14 program competencies. The exercises begin with step-by-step tutorials that build student capabilities with software. Advanced exercises propose open challenges to solve. These resources provide IS programs with a draft of potential exercises to include in courses and a framework for covering program-level objectives.

Keywords: learning objectives, assignments, data analytics, visualization

1. INTRODUCTION

The IS2020 model curriculum articulates requisite and optional program competencies for Information Systems (IS) programs (Leidig, Salmela, Anderson, Babb, de Villiers, Gardner, Nunamaker, Scholtz, Shankararaman, Sooriamurthi, & Thouin, 2021). Meeting the report's knowledge and skills recommendations is an open challenge for IS programs. Covering a comprehensive set of knowledge elements and skills requires programs to design learning objectives and materials across multiple courses.

Data and information management is one of the six high-level competency realms in IS2020. In this realm, the required data management area is supplemented by two elective areas (data / business analytics and data / information visualization), see Table 1. To fulfill the IS2020 guidelines, a program must fulfill the data management area but is not required to offer the data / business analytics or data / information visualization areas. The two optional areas build upon the requisite data management area's knowledge elements, e.g., data manipulation, database joins, and non-relational models.

Area	Example courses
Data / Info. Management (Required)	Databases, information search and retrieval, and knowledge management
Data / Bus. Analytics (Elective)	Artificial intelligence, business intelligence, data mining, and machine learning
Data / Info. Visualization (Elective)	Information visualization and visual analytics

Table 1: The three areas within the IS2020 data and information management realm.

The data / business analytics area covers a broad set of knowledge elements from a variety of domains including business intelligence, data mining, data science, etc. The knowledge elements and skills from multiple courses must be coordinated in order for students to recognize and internalize connections between the full set of program competencies and course competencies.

A regional school in the Midwest offers an IS degree that includes an optional track in data analysis. The curriculum for the degree with this track requires four courses in database management, data analytics tools, information visualization, and either AI or data mining. Informal student feedback suggests that the logical connections between topics taught in separate courses is not always obvious to students. As these courses are also utilized by several academic programs and taught by multiple faculty members, additional considerations are required for students to integrate the set of related competencies. Courses in the data analysis track are partly coordinated on several of the required knowledge elements, materials, case studies, software laboratories, and projects. A coordinated series of lab assignments and projects were designed to cover technical competencies while focusing on problem solving in the memorable domain of marine ecology. This domain was found to provide engaging datasets, familiar case studies for students from varied backgrounds, intuitive patterns to recognize, and insights to uncover.

The contribution of this work is an organized set of knowledge elements, skills, datasets, and assignments (concentrated in the data management realm) that has been refined in light of IS2020. Appendices A-C contain proposed exercises and are summarized in Table 2. Appendices A and C have been tested in the classroom over the last four years.

Appendix	Marine Ecology-Based Resources
A	A case study, datasets, and tasks for data analytics of tabular data with Python
B	A case study, datasets, and tasks for machine learning of unstructured multimedia data
C	A step-by-step tutorial for interactive visual analytics for quantitative data with Tableau

Table 2: Overview of the Appendices.

The remaining sections of the paper present an approach to fulfilling competencies in the two elective areas of the IS2020 model. Section 2 covers the seven competencies that comprise the IS2020 elective area of Data / Business Analytics. In the first part, the first four of the seven competencies are discussed in light of tabular data exercises. In the second part, the last three of the seven competencies are discussed in light of semi-structured and unstructured data exercises. Similarly, Section 3 covers the seven competencies that comprise the IS2020 elective area of Data / Information Visualization. In the first part, the first five of the seven competencies are discussed in light of descriptive visualizations. In the second part, the remaining two of seven competencies are discussed in light of interaction and discovery.

2. DATA ANALYTICS

IS2020 details the elective area of Data / Business Analytics within section A3.2.2. Programs that offer content in this optional area build student competencies in handling and analyzing large, diverse datasets. Students must be able to inform business problems with actionable solutions by leveraging large underlying data. Analytics tasks involve techniques covered across several business, computing, math, and statistics courses. In particular, the fields of data mining and machine learning cover topics related to classification, clustering, modeling, prediction, optimization, and recommendation. IS2020 recommends seven competencies to be covered via analytics courses.

In the author's IS program, the seven competencies for the data analytics area are intentionally covered within the four-course track. In these courses, much of the course content is aligned with common textbooks and traditional assignments that teach students to perform common analytical tasks on tabular datasets. As an example, competency #7 requires the use of big data tools on real-world case studies via Hadoop, Spark, and the map-reduce framework. In an information management course, students use map-reduce queries to filter, aggregate, and analyze collections of semi-structured JSON files. Students also use HBase (Hadoop) or Google BigTable to store and retrieve social media data in scalable cloud-based columnar databases. Instead of focusing on course materials that have already been widely adopted, this section focuses on the less traditional datasets and exercises that have been incorporated into the data analytics track to specifically meet the IS2020 Bloom's cognitive level recommendation for the seven data analytics competencies.

Tabular and geo-temporal data exercises

Tabular data is used to cover the first four out of the seven Data / Business Analytics competencies, see Table 3. Competencies #1-4 require students to formulate and perform data-driven analytical tasks. The remaining three competencies (#5-7) will be covered with unstructured data in the next subsection.

Competency	Description
#1	apply the principles of computational thinking, abstraction, pattern recognition, etc.
#2	analyze data science problems
#3	express business problems as data problems
#4	perform data analysis using descriptive statistics and visualization

Table 3: IS2020 competencies in the elective field of Data / Business Analytics.

These competencies can be evaluated in a pipeline of skills that requires students to perform data analytics tasks with quantitative datasets. Tasks in the pipeline include managing data; performing extract, transform, and load (ETL) tasks; analyzing data; uncovering insights; presenting findings in descriptive charts; and suggesting solutions to problems. These tasks provide students with experiential learning in analyses, interpretations, and recommendations. After exposure to diverse analytical techniques in a variety of pre-requisite courses, a graded activity in an upper-level course is used to challenge students to solve several problems related to marine ecology. Students are given a brief background primer on toxic harmful algal blooms (HABs) and hypoxia (a stratified water column with insufficient oxygen to support fish and other aquatic life). The scientific observations and series of analytical tasks are used to challenge students to recognize HABs in field datasets and identify hypoxic events in the Great Lakes region and Gulf of Mexico.

For these competencies, students are given background information on the marine ecology problem of hypoxia as detailed in Appendix A (Louisiana Universities Marine Consortium, 2018). They are then given scientific observations that inform problem solving efforts. Specific problems are paired with appropriate data sources including governmental warning systems (GLERL, 2022), national buoy sensor networks (NDBC, 2021), merchant ships that report real-time water readings (NDBC, 2022), and academic datasets (GLERL, n.d.). In reviewing the metadata and field descriptions for the datasets, students learn about sensors, standard units of measurement (e.g., microSiemens/cm, mg/L, and NTUs), valid data ranges, calibration challenges for the hardware sensors, maintenance service intervals, biofouling and algae effects on sensors, prevalent dirty data, the constant degradation of oxygen sensor membranes (which affects accuracy), sensor failures, prior data cleaning steps, data provenance, etc. The case study requires students to review data descriptions and footnotes in order to self-evaluate their understanding of the data and its context before performing analysis tasks. Students have to take these factors into consideration as they develop Python scripts to clean observations of nonsensical measurements and perform imputation on missing values. Students then solve analytics tasks by developing Python code to manipulate and integrate cleaned data, produce descriptive statistics, and create charts that focus on a specific problem. Specific tasks are updated each semester and might include:

- Visualize hypoxia events in Lake Erie based on a high-dimension dataset. Create a histogram or similar visualization that details when the dissolved oxygen falls below 2 mg/L. Generate at least one chart for each buoy location, covering one year of data. Consider using matplotlib, ggplot2, or Tableau charts.
- Visualize events in Lake Erie by integrating multiple datasets for buoys and shore-based research stations. Create a geo-spatial map that either uses color or size to display the observations for multiple POI locations. Select one attribute (temperature, DO, turbidity, etc.) and generate charts that display all of the readings from multiple buoys at a similar timestamp. Note: the timestamps might not be aligned for the entire set of locations. Consider using a Python mapping library or Tableau maps.
- Visualize the locations of ships that are reporting scientific observations on Lake Erie. You may use either the current set of active ships or else a historic dataset. Create a geo-spatial map that plots the coordinates of each ship as well as field attributes (air temperature, water temp, air visibility, wind direction, wave height, etc.). Generate a dashboard or map showing all active ships for a given timestamp (e.g., over a one-hour window). Consider using a Python mapping library or Tableau maps.

The results of these tasks are then explored, interpreted, and written up in documentation. Students interpret their results to determine the potential for hypoxia (dissolved oxygen readings under 2 mg/L) or HABs (algal blooms with high readings from chlorophyll and phycocyanin sensors). Students reflect on whether HABs or hypoxia conditions were identified, identify geo-temporal patterns of detected events, and identify any potential need for governmental responses. Students finish the laboratory exercises by modifying their scripts to best present their findings to decision makers on the respective problem. See Appendix A for further details regarding the publicly available dataset and proposed laboratory exercises.

Competencies #1-4 are successfully evaluated by these guided laboratory assignments that require students to reflect on a new domain, investigate a range of available datasets, perform analytics, and formulate their own data-driven solutions.

Multimedia data exercises

Data / Business Analytics competencies #5-7 require exercises for semi-structured and unstructured datasets. Multimedia data is used to cover these last three of the seven Data / Business Analytics competencies, see Table 4.

Competency	Description
#5	explain the principles of classification, clustering, optimization, and recommendation
#6	articulate the potential of big data given volume, velocity, and variety
#7	demonstrate the use of big data tools on real world case studies

Table 4: IS2020 competencies in the elective field of Data / Business Analytics.

Convolutional neural networks (CNNs) and other machine learning models are now widely used to classify unstructured data such as multimedia. See Appendix B for background information regarding two marine ecology problems that require machine learning algorithms to calculate biodiversity and coral cover (Reef Renewal Bonaire, 2021). A new oceanic dataset comprised of coral images was developed to support these machine learning tasks and education (Leidig, 2022). A sabbatical leave period was utilized to travel to The Bahamas, Bonaire, and Florida to capture and organize coral image collections. The dataset contains millions of underwater images with known classes (coral genus and species labels). These new collections provide large training datasets for 40+ species of coral. See Figure 1 for coral images from the collections.

With these new collections, exercises can now be developed for competencies #5-7. Specifically, laboratory tasks cover techniques for managing massive datasets, manipulating unstructured data, executing parallel code, and training machine learning classification models. The Python with PyTorch framework (Paszke, Gross, Massa, Lerer, Bradbury, Chanan, & Chintala, 2019) is used by students to train CNNs using these datasets. See Appendix B for additional details regarding the publicly available collections and exercises. Competencies #5-7 are met by guided laboratory assignments that require students to access large datasets, classify two or more species using common CNN tools, and evaluate the classification performance.

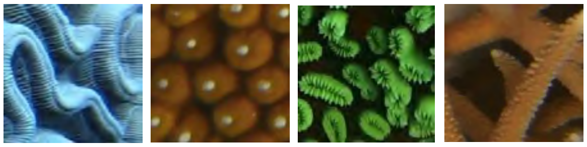


Figure 1 Four coral classes (boulder brain, great star, smooth flower, and staghorn).

3. DATA VISUALIZATION

IS2020 details the elective area of Data / Information Visualization within section A3.2.3. Programs that offer content in this area build student competencies in manipulating datasets, exploring, and presenting insights. Students must be able to inform business decisions by storytelling with clear narratives built upon underlying data. Visualization tasks involve the techniques covered in computing and statistical domains as well as the fields of art and media, cogitative sciences, psychology, etc. In particular, important topics are related to data integration, manipulation, visual encoding, interaction, human perception, and software tools. IS2020 recommends seven competencies to be covered via visualization courses. This section details datasets and exercises that have been incorporated into the data analytics track in light of IS2020's visualization competencies.

Descriptive visualization exercises

In Data / Information Visualization competencies #1-5, students design descriptive visualizations that guide analysts to insights, see Table 5. Competency #1 requires additional emphasis on human cognition. Students are tasked with intentionally considering their intended insight given a message to convey, data to encode, potential charts, and anticipated audience. In addition to the actual analytical tasks performed, assignments related to competency #1 require students to provide written reflections. For each chart, students perform three additional tasks. 1) Write in a sentence: what is the overarching theme of the chart, title, caption, and insight that should stick out to the user? E.g., what are the temporal patterns of HAB events that occur in Lake Erie? 2) Design multiple charts with different plot types and ranges of context. Each chart must encode data with a single, clear takeaway. Respond: for consumer-facing charts, were pre-attentive cognition and the Gestalt Laws of Psychology utilized to make the takeaway apparent to the chart consumer? Did the chart's title and captions reinforce the insight to guide the consumer in interpreting the chart? 3) Write in a paragraph: Reflect how your visualization follows information visualization theory and principles. Consider discussing: accurate mapping from data to the generated graph; how static infographics show scale, comparison, or patterns; ease for the user to 'unmap' the graphic and understand the insight; ease for users to forage and find information; high ratios of data per pixel; presence of chart junk; adherence to Cleveland's rules; use of the Gestalt Laws of Grouping; interaction techniques; scalability; and use of overviews and navigation. In addition to generating accurate visualizations, reflection forces students to focus their attention on human perception and competency #1.

The scientific datasets and related set of Python-based analytical exercises (described above in the data / business analytics competencies #1-4 subsection) are also used to evaluate data / information visualization competencies #2-5.

Competency	Description
#1	understand human perception
#2	effectively display quantitative datasets
#3	present analyzed datasets and insights using visualization theory
#4	develop scripts to manipulate datasets
#5	design and generate visualizations

Table 5: IS2020 competencies in the elective field of data visualization.

Discovery visualization exercises

Within the seven data / information visualization competencies, #6 and 7 require students to design interactive visualizations that allow consumers to uncover insights and form compelling, data-driven narratives, see Table 6.

Due to the high learning curve of web-development using D3.js, laboratory exercises leverage the intuitive Tableau software package, see Appendix C. With hundreds of thousands of fish surveys dating back several decades, the Reef Environmental Education Foundation (REEF) provides quantitative datasets that track hundreds of fish species on specific reefs across the globe (REEF, 2022). The geo-temporal organization of multi-dimension survey reports allows for hierarchical overviews that group reefs into larger combined reef tracks, countries, and world regions as well as hierarchical time components (month, season, year, and decade). The fish species themselves are also organized by hierarchical taxa (e.g., genus and species). The datasets lend themselves to exploration of the relations between fish species and geo-temporal distribution patterns.

Competency	Description
#6	explore datasets using interactive visual analytics
#7	express compelling discoveries and insights

Table 6: IS2020 competencies in the elective field of data visualization.

To evaluate competencies #6 and 7, students complete a series of step-by-step laboratory exercises that show students how to conduct visual analytics using Tableau. Students are given REEF data monitoring 547 fish species through 26,737 survey reports (dating back to 1993) for the 168 coral reef locations surrounding the country of Bonaire. Tutorials guide the students to learn how to import data, join and integrate datasets, filter observations, design charts, and compose interactive dashboards. The dashboards then serve as a basis for reasoning with descriptive statistics, probability, apparent paradoxes, and pattern interpretation. Students are prompted to answer questions by exploring the dataset using the interactive charts they design. Relevant views of the dataset are identified and interpreted in narratives. As examples, students identify differences in the survey reports to comment on data validation issues when comparing expert surveyors to novice surveyors. Fish distributions inform questions on probabilistic encounters with various wildlife. At least two laboratory exercises are needed for students to become familiar with the software as well as engage in visual analytics.

See Appendix C for further details regarding the publicly available REEF survey datasets and related laboratory exercises. Pairing this high-dimensional dataset with Tableau supports competencies #6 and 7. Specifically, several laboratory reports evaluate skills, hypothesis generation, interactive visualizations, narratives, visual exploration, and visual explanation.

4. SUMMARY

The IS2020 guidelines led IS programs to consider the extent to which they cover the 21 competencies in the Data / Information Realm. This paper contributes a coordinated set of datasets and exercises used to teach and evaluate knowledge elements and skills. Adopting these materials (or curating alternative collections) would benefit other IS programs aiming to cover the 14 elective Data / Business Analytics and Data / Information Visualization competencies. These data-driven exercises require students to synthetize knowledge elements, skills, and analytical techniques taught across multiple courses in order to demonstrate program competencies in their upper-level courses.

The datasets and exercises presented here have been leveraged over the last four years to provide and evaluate coverage of IS2020's 14 elective competencies. The exercises in Appendix A have been utilized over five semesters to introduce students to data analytics in Python. The exercises in Appendix B have only been tested so far by five graduate students – who individually desired to learn about deep learning with CNNs and Python. These graduate students were able to produce classification models with ~96% accuracy, depending on the coral species being recognized. As the coral classification training dataset was produced in 2022, another year is needed to finalize and evaluate step-by-step lab exercises at a level appropriate for undergraduate students. Appendix C exercises have been utilized in 10 sections of visualization courses over the last four years. Students commonly express their appreciation for the step-by-step tutorials that introduce them to Tableau and performing interactive visual analytics.

This set of resources supports all 14 competencies in the IS2020 elective areas. The exercises were designed to provide IS programs with memorable case studies, provide intuitive problems, and expose students to problem solving within a new domain. Marine ecology was selected as the targeted domain as existing quantitative datasets were well described (with sufficient metadata) and minimal background knowledge was required to begin problem solving (e.g., brief overviews of biodiversity, HABs, hypoxia, and the use of reef photomosaics were sufficient coverage of the domain). The challenges of identifying HABs, hypoxia, coral species, and coral reef growth are familiar, engaging, attention grabbing, and memorable for students regardless of personal or academic backgrounds. The real-world coral case study is an active, open challenge for coral restoration groups that monitor the degradation of coral reefs. Concentrating analyses in a single applied domain (across multiple IS courses) provides a structure for students to recall, reflect, and integrate previous laboratory experiences.

5. REFERENCES

GLERL. (2022). *Experimental Lake Erie Hypoxia Forecast*. NOAA - Great Lakes Environmental Research Laboratory. https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/hypoxiaWarningSystem.html

GLERL. (n.d.). *International Field Years on Lake Erie (IFYLE)*. NOAA - Great Lakes Environmental Research Laboratory. https://www.glerl.noaa.gov/res/projects/ifyle

Leidig, J. (2022). *Coral Reef Image Collections for Machine Learning, Mapping, and Monitoring*. In Proceedings of the 33rd IEEE OCEANS Conference 2022.

Leidig, P., Salmela H., Anderson G., Babb J., de Villiers C., Gardner L., Nunamaker J.F., Scholtz B., Shankararaman V., Sooriamurthi R., & Thouin M. (2021). *IS2020 A Competency Model for Undergraduate Programs in Information Systems*. Association for Computing Machinery and Association of Information Systems. DOI: 10.1145/3460863.

Louisiana Universities Marine Consortium. (2018). *About Hypoxia*. Gulf of Mexico Hypoxia. https://gulfhypoxia.net/about-hypoxia

NDBC. (2021, August). *National Data Buoy Center*. NOAA. https://www.ndbc.noaa.gov

NDBC. (2022, March). *Ship Observations Report*. NOAA. www.ndbc.noaa.gov/ship_obs.php

Paszke, A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., and Chintala S. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library.* In Advances in Neural Information Processing Systems 32 (pp. 8024–8035). Curran Associates, Inc. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

REEF. (2022). *Reef Environmental Education Foundation Volunteer Fish Survey Project Database*. World Wide Web electronic publication. Retrieved July 15, 2022, from www.REEF.org

Reef Renewal Bonaire. (2021, June 8). *Measuring our Coral Reef Restoration Success*. Reef Renewal Bonaire. Retrieved July 15, 2022, from https://www.reefrenewalbonaire.org/news/measuring-our-restoration-success

APPENDIX A
Tabular and geo-temporal data exercises

Scientific sensors produce large, longitudinal, and semi-structured datasets. Due to the hardware involved, water buoys provide students with challenges to mitigate (dirty data, erroneous readings, missing values, unit conversions, etc.). Students can gain experiences in checking assumptions, skew, outliers, anomalies, descriptive statistics, and clustering. The data also contains geo-temporal patterns to mine.

Content:
- Buoy and ship water sample data. Latent insights include hypoxic events and HABs.

Requirements:
- Less than 5MB of storage space

Recommended tech stack and platform:
- Python environment with common libraries, Tableau (free academic licenses for personal devices and/or cloud access)

Sample Laboratory Exercise

Part 1: Data manipulation (Pandas), analysis (SciPy), and visualization (Matplotlib, NumPy, and Tableau)

Description

SciPy is a set of widely used packages for managing, analyzing, and visualization large-scale content (see https://scipy.org). It consists of libraries for data structures such as DataFrames and analysis (pandas), N-D arrays (NumPy), 2D plotting (Matplotlib), scientific analysis (SciPy), etc. In addition to matplotlib, other popular libraries include *ggplot (any R fans in the class?), plotly, seaborn, pygal, bokeh, geoplotlib, etc.* These packages differ by their customization, expected data input structures, charts available, export formats (e.g., svg), interactivity, dependencies (web integrated, Pyglet OOP interface), etc.

Tableau is a leading software tool for visual analytics and rapidly creating interactive visualizations (see https://www.tableau.com). Download the tool from tableau.com or else login to the GVSU Blade server for Bioinformatics. Students (worldwide) get renewable year-long free licenses (with a valid university email address). If you require brief tutorials (outside of the in-class training) regarding the GUI and its functionality, see https://www.tableau.com/learn/training. Tableau connects with a variety of underlying dataset formats: offline dataset file, an online data server, or a default source (e.g., try out the World Indicators dataset). Note how the columns from the dataset are available on the left pane, you can drag columns to the middle pane to set the x-position y-position color size (which automatically updates the chart graphic), dragging columns to the filter pane generates dynamic filters within the visualization, and the right pane allows you to change the chart type depending on the columns already selected. Users explore their dataset by creating a sheet and pairing various combinations of dimensions and measures. For each combination of columns that you explore, switch the visualization to several options (including Tableau's recommended visualization which is highlighted on the list of possible charts on the right pane). One goal of this lab is to become familiar with Tableau, how to create graphs (sheets), and how to create interactive dashboards.

Learning objectives

- Become familiar with data manipulation and analysis using commonly used programming libraries
- Open and parse a (potentially) large input file

- Apply some transformation/manipulation to the content as needed
- Integrate existing databases and datasets with analytical processes
- Load cleaned data into an appropriate database, staged data, or tool
- Integrate quantitative analysis/modeling with visualization
- Develop visualizations using common libraries and interactive visualizations for exploration and decision making

Step #1 Case Study - Visual Analytics

With a partner (optional), review the following background material on hypoxic environments. Note: this is currently an open, real-world challenge, not a solved problem. You can expect any existing datasets to be less than ideal, dirty, missing values, misaligned with current real-world situation, etc.

Gather background information:

- https://gulfhypoxia.net/about-hypoxia
- http://coastal.ohiodnr.gov/portals/coastal/pdfs/owc/tech/owc_techbull3_Hypoxia.pdf
  Question 1: Define hypoxia. What is the dissolved oxygen concentration threshold that identifies hypoxia? ______
  Question 2: Which species are affected? How are these species affected? ______

Review the animation of hypoxia forecast data from Lake Erie's 2018-2020 seasons:

- https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/hypoxiaWarningSystem.html
- https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/archive/dissolvedoxygen2018.gif
- https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/archive/dissolvedoxygen2019.gif
- https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/archive/dissolvedoxygen2020.gif
  Question 3: Where and when is hypoxia likely to occur in Lake Erie? Either describe the general locations or provide an annotated map. ______

Remote monitoring:

- https://www.ndbc.noaa.gov
  Question 4: Are current NOAA buoys and sensors located in appropriate locations to monitoring and surveillance of hypoxia? If so, which buoy/station IDs are in pertinent locations? _______
- https://www.ndbc.noaa.gov/ship_obs.php (click show observations for the last 12 hours)
  Question 5: Are there shipping lanes through the area that might provide sporadically-sampled monitoring data? Skip this question for now (not graded).

Historic data 2004-2007:

- https://www.glerl.noaa.gov/res/projects/ifyle
- https://www.glerl.noaa.gov/res/projects/ifyle/data.html
- https://www.glerl.noaa.gov/res/projects/ifyle/data/data.mooring.html
- https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/data.ysi.html
- https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/ysi/eriemap.html
  Review these details on the GLERL collaboration which has attempted to capture and monitor datasets related to hypoxia for the last 15 years in Lake Erie. Select a buoy location that appears to experience hypoxia events throughout the year (e.g., station Y18). Review the metadata and challenges of collecting this type of data.
- https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/ysi/ysi_metadata.txt
  Question 6: These datasets contain very useful information (sensor depth, dissolved oxygen concentrations, temperature, turbidity, and chlorophyll) but were not always accurate (e.g., 2007

data). Describe the errors that are known to exist within the dataset. ______

Step #2 Setup a Python, R, or Tableau visualization environment

Connect to the Stratus' Bioinformatics server (Python and Tableau already installed), EOS (already installed), or else install one of R Studio, Python IDE/Interpreter, or Tableau on your laptop. Note: you likely will not have full permission on EOS or the Stratus Blade server to install new software.

Installation Resources

Stratus - https://www.gvsu.edu/hpc/stratus-2.htm
EOS - http://www.cis.gvsu.edu/eosarchitecture-labs
https://www.python.org/about/gettingstarted
https://docs.python.org/3/tutorial
https://www.tableau.com
There is a year-long free license of Tableau Desktop for students:
https://www.tableau.com/community/academic
There is a 14-day free trial of Tableau Desktop: https://www.tableau.com/products/trial

Step #3 Visualization Challenges

Select and complete one of the following challenges using Tableau, Python, or R. Each option will require a dataset to be extracted/scrapped from a website, manipulated, cleaned, and visualized. The exact approach, software, libraries, and solution is left up to the students, in keeping with the spirit of the assignment. For one of the following options, *turn in* a one paragraph writeup containing a description of your methodology/workflow (e.g., data source, how downloaded, how cleaned, and how visualized), a final visualization, and any code/workbooks developed.

Option A

Visualize hypoxia events in Lake Erie based on a high-dimension dataset. Create a histogram or similar visualization that details when the dissolved oxygen falls below 2 mg/L. Generate at least one chart for one location, covering data from the whole year at that location. Consider using matplotlib, ggplot2, or Tableau charts.
Example data:
https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/ysi/2007/Y18.txt
where day/time `206.0208 =` 30 minutes, between day 205/365 to day 288/365

Option B

Visualize events in Lake Erie by integrating datasets for multiple buoys/stations. Create a geo-spatial map that either uses color or size to display the observations for multiple POI locations. Select one attribute (temperature, DO, turbidity, etc.) and generate a chart that displays all of the readings from multiple buoys at a similar timestamp. Note: the timestamps might not be exactly identical and need to be rounded. Generate at least one chart (for the entire set of locations) at one timestamp. Consider using a python/R mapping library or Tableau maps.
Example data:
https://www.glerl.noaa.gov/res/projects/ifyle/data/Mooring/ysi/eriemap.html

Option C

Visualize the locations of ships that are reporting scientific observations on Lake Erie. You may use either the current, active ships or else a historic dataset. Create a geo-spatial map that plots the coordinates of each ship as well as one attribute (air temperature, water temp, air visibility, wind direction, wave height etc.). Generate at least one chart/map showing all active ships for a given timestamp (e.g., a one-hour window). Consider using a python/R mapping library or Tableau maps. Example data: https://www.ndbc.noaa.gov/ship_obs.php

Deliverables:

Turn in a Word/text file containing your answers to Step 1 questions and Step 3 text + visualization. Also, include ANY computing code (e.g., R/Python scripts), Tableau workbooks (.twbx file), intermediate datasets (e.g., cleaned CSV files if small in size), etc. Upload a softcopy of these documents.

APPENDIX B
Coral Multimedia Datasets and Exercises

In marine ecology, two common tasks are in estimating species biodiversity and the percent of the seabed that is covered in coral. Thousands of images are captured (looking down at the seabed) and then stitched together to form a single large photomosaic image that covers several square kilometers of the ocean floor. Researchers manually analyze the image to calculate biodiversity and coral cover (Reef Renewal Bonaire, 2021). These two features are calculated by randomly picking pixels from the image and labeling which coral species (if any) was at that location. These human-generated labels are then aggregated to determine biodiversity and coral cover. However, advancements in machine learning, specifically convolutional neural networks (CNNs), promise to semi-automate these human-intensive tasks. Large training datasets of underwater images could be used to develop CNNs that classify the genus and species of a coral found in a top-down view image. With CNNs to classify multiple coral species, the entire photomosaic could be classified at each pixel location - without efforts by human experts.

Additional details on the coral collections, field data collection methodology, data processing, and provenance can be found in the full conference paper documenting the collections release (Leidig, 2022). 5.28 million images of coral organized into 44 separate species collections (see Table 7). Each JPG image is 256x256 pixels with RGB color channels.

Requirements:

- Binary classification exercise: 20GB of storage is required for two species with 1,000 training images each, see example images in Figure 2.
- Multiclass classification exercise: 95GB of storage is required for ten species with 1,000 training images each.
- Full-dataset: 1TB storage is required for all training data over 44 coral species.
- Exercises require access to free software installed on a cloud or local HPC resources (ideally with GPUs). At the time of writing, Google Collaborate, Juypter Notebooks, Python, and PyTorch are available (at no cost) to students and faculty for educational purposes.

Recommended tech stack and platform:

- Python, PyTorch, and computing resources (Google Collab, personal student computers, or lab).

Sequential tasks:

- Practical computing laboratory steps demonstrate how to access and manage the organized multimedia datasets.
- Laboratory steps expose students to CNNs model building u6ing template code, optimizing a small number of hyperparameters.
- Laboratory steps provide background information and libraries for transfer learning with ResNet and similar image-based neural networks.
- Students perform binary classification (distinguishing between two species) with model training and performance evaluation.
- Students are challenged with multi-species classification.

Classes of coral:

	Common Name	Scientific Name	Number of Training Images
1	Staghorn	Acropora cervicornis	1125040
2	Lobed Star	Orbicella annularis	571920
3	Mountainous Star	Orbicella faveolata	461600
4	Boulder Brain	Colpophyllia natans	367720
5	Yellow Pencil	Madracis auretenra	286000
6	Great Star	Montastraea cavernosa	249000
7	Symmetrical Brain	Pseudodiploria strigosa	190200

8	Massive Starlet	Siderastrea siderea	172800
9	Blade Fire	Millepora complanata	168360
10	Whitestar Sheet	Agaricia lamarcki	163800
11	Elliptical Star	Dichocoenia stokesi	136800
12	Smooth Flower	Eusmilia fastigiata	127800
13	Bushy Black	Antipathes caribbeana	123000
14	ClubtipFinger	Porites porites	116760
15	Boulder Star	Orbicella franksi	114600
16	Pillar	Dendrogyra cylindrus	112200
17	Grooved Brain	Diploria labyrinthiformis	107280
18	Maze	Meandrina meandrites	99000
19	Elkhorn	Acropora palmata	90720
20	Mustard Hill	Porites astreoides	63000
21	Spiny Flower	Mussa angulosa	57360
22	Lettuce	Agaricia agaricites	51600
23	Rough Cactus	Mycetophyllia ferox	41400
24	Ridged Fire	Millepora striata	39600
25	Blushing Star	Stephanocoenia intersepta	33600
26	Branching Fire	Millepora alcicornis	31800
27	Thin Leaf Lettuce	Agaricia tenuifolia	27600
28	Six Ray Star	Madracis senaria	20400
29	Wire	Stichopathes luetkeni	19200
30	Ridged Cactus	Mycetophyllia lamarckiana	18000
31	Ten Ray Finger	Madracis carmabi	16200
32	Ten Ray Star	Madracis decactis	13800
33	Rough Star	Isophyllia rigida	9000
34	Lesser Starlet	Siderastrea radians	8640
35	Dimpled Sheet	Agaricia grahamae	7800
36	Knobby Cactus	Mycetophyllia aliciae	7800
37	Butter Print Rose	Meandrina danae	7000
38	Branching Finger	Porites furcata	6600
39	Golfball	Favia fragum	5400
40	Rose Lace	Millepora complanata	4200
41	Knobby Brain	Pseudodiploria clivosa	3480
42	Thin Finger	Porites divaricata	2240

Table 7: Field data collection sizes of ML training datasets (sub-images).

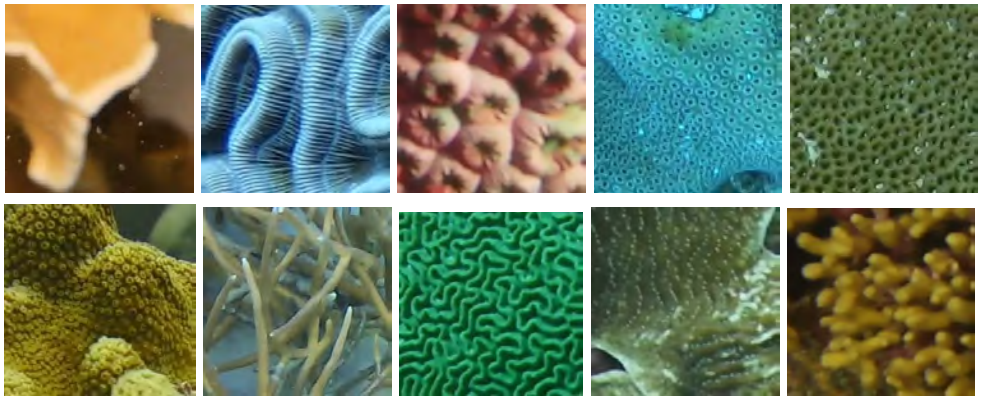


Figure 2 Top 10 classes (top-row) blade fire, boulder brain, great star, lobed star, massive starlet, (bottom-row) mountainous star, staghorn, symmetrical brain, whitestar sheet, and yellow pencil.

APPENDIX C
Fish Survey Datasets and Exercises

Reef Environmental Education Foundation (REEF) collects and aggregates hundreds of thousands of fish surveys from SCUBA-diving volunteers. Additional details on the REEF collections, roving field data collection methodology, data processing, and provenance can be found in (*REEF, 2022)*. These datasets are high-dimensional (100s of columns), with geospatial and temporal components. The many quantitative features, locations, and time components lend themselves to visual exploration in a student-friendly domain.

Requirements:

- Less than 5MB of data, extracted from reef.org

Recommended tech stack and platform:

- Excel and Tableau (Desktop or cloud-based)

Tasks:

- Load data, inner joins, chart generation, filtering, and exploration

Sample Laboratory Exercise

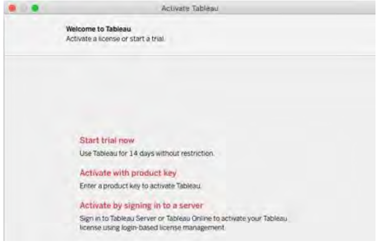
Part 1: An introduction to visual analytics, interactivity, dashboards

Tableau is a leading software tool for visual analytics and rapidly creating interactive visualizations. Students (worldwide) get renewable year-long free licenses (with a valid university email address). If you require brief tutorials (outside of the in-class training) regarding the GUI and its functionality, see https://www.tableau.com/learn/training. The goal of this lab series is to become familiar with Tableau, how to create sheets, exploring data, and how to create interactive dashboards. Lets download the tool **Tableau Desktop** from tableau.com, not the online or public tools.

1. Free License: Visit https://www.tableau.com/academic/students, select the “Free Student License” button, and apply for a free year license. It may take a few hours to a day to receive your key. However, you could likely complete our Tableau labs just using the 14-day free trial and skip the free license step.

2. Install a 14-day free Tableau trial: Visit https://www.tableau.com/products/trial. You can later enter your registration key later if you haven’t received it yet. Click the “start a free trial” icon and setup an acount or sign into your existing account. Install the software according to the instructions for your Mac or Windows.

3. Start your Tableau software.You can simply use the *start trial now* option until you receive your activation key. Enter in the email you used to register online, if prompted.



Activate Tableau

Welcome to Tableau
Activate a license or start a trial.

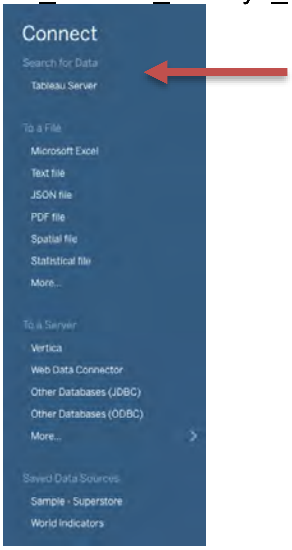
Start trial now
Use Tableau for 14 days without restriction.

Activate with product key
Enter a product key to activate Tableau.

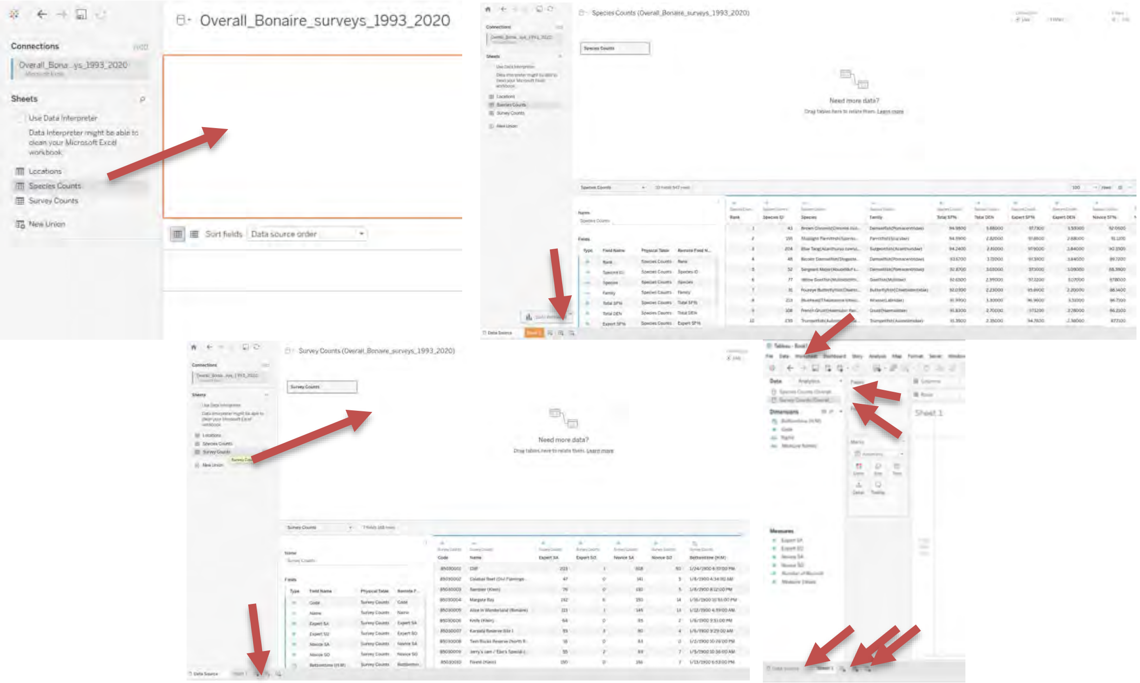
Activate by signing in to a server
Sign in to Tableau Server or Tableau Online to activate your Tableau license using login-based license management.

Bonaire is a Carribean island forming part of the Dutch Antilles (with Aruba and Curacao) located a few miles off the coast of Venezula. It is generally well known for its healthy coral reefs, pink salt industry, and tourism based on SCUBA diving and water sports. Let's use Tableau to explore Bonaire's REEF data (www.reef.org).

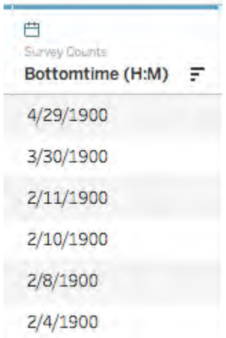
4. Download the files on Blackboard to your computer. You can open the file in Excel to check the data.
5. Import the file in Tableau - a collection of fish surveys by experts and novices. Use the Connect panel on the left to connect to a Microsoft Excel file, find and open your downloaded file: Overall_bonaire_surveys_1993_2020.xlsx. There are two sheets of data for us to import.

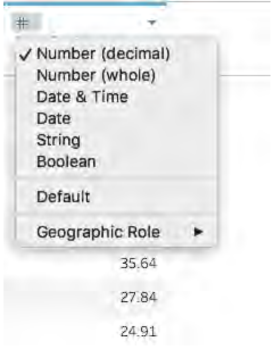


6. We will add the two sheets of data, one at a time. First, select the *Species Counts* Excel worksheet and drag it to the pane on the right. Then click on Sheet 1 on the bottom menu. A *sheet* is the Tableau word for a single chart. Under the Data menu at the top of your screen, select a New Data Source, find your excel file again, and then add the *Survey Counts* worksheet by dragging it to the right pane. Go back to the Sheet 1 chart. After this step, both sheets (*Species Counts* and *Survey Counts*) should be visible in the top left Data menu. Ignore the *Locations* datasheet from the Excel file. The data can always be reviewed in the Data Source tab on the bottom-left corner. Each chart will be created in a new Sheet.

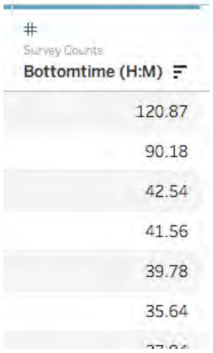

Connections
Overall_Bona...ys_1993_2020
Microsoft Excel
Sheets
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Locations
Species Counts
Survey Counts
New Union
Overall_Bonaire_surveys_1993_2020
Sort fields
Data source order
Species Counts (Overall_Bonaire_surveys_1993_2020)
Need more data?
Drag tables here to relate them. Learn more
Survey Counts (Overall_Bonaire_surveys_1993_2020)
Data Source
Sheet 1
Tableau - Book1
Data
Analytics
Dimensions
Measures
Pages
Columns
Rows
Filters
Marks
Sheet 1

a. How many rows and columns does the Species Counts Data Source contain? Hint: use the Data Source tab.
b. How many rows and columns does the Survey Counts Data Source contain? Hint: use the Data Source tab.
c. When importing the Survey Counts sheet from Excel, note that the variable Bottomtime is a number of hours and minutes (which Tableau mistakes for a date). In the data source pane, let's switch that back to Number(decimal).


Survey Counts
Bottomtime (H:M)
4/29/1900
3/30/1900
2/11/1900
2/10/1900
2/8/1900
2/4/1900



35.64
27.84
24.91


Survey Counts
Bottomtime (H:M)
120.87
90.18
42.54
41.56
39.78
35.64

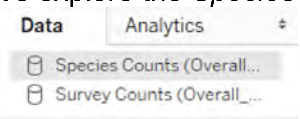
d. For more details, checkout: https://www.reef.org/db/reports/geo?end_date=2020-02-

10&format_type=chart&group_type=species&language=common&region_code=TWA&start_date=1993-01-01&zone_map=0&zones=8503

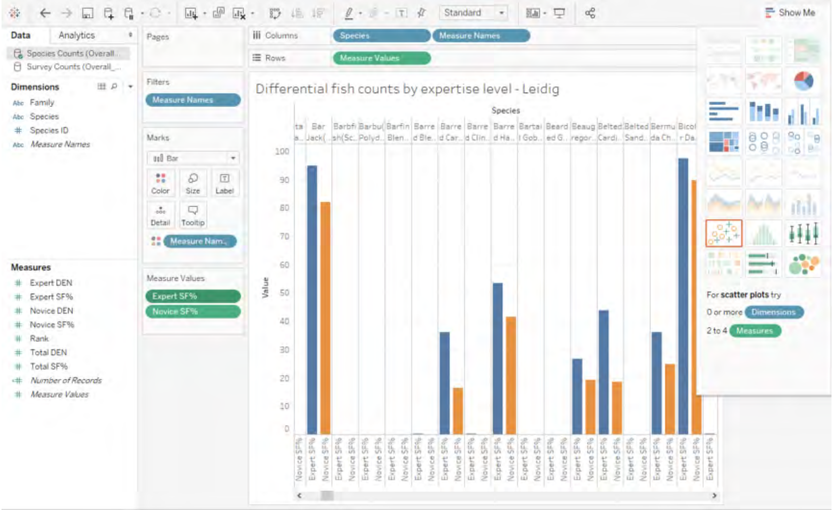
7. Use *Sheet 1*. Note how the columns from the dataset are available on the left pane, you can drag columns to the middle pane to set the x-position/y-position/color/size (which automatically updates the chart graphic), dragging columns to the filter pane generates dynamic filters within the visualization, and the right pane allows you to change the chart type depending on the columns already selected.
8. First, explore your dataset by creating a **sheet** and pairing various combinations of dimensions and measures. For each combination of columns that you explore, switch the visualization to several options (including Tableau's recommended visualization which is highlighted on the list of possible charts on the right pane).

Next use Tableau to explore and answer the following questions. For each question, create a sheet (visualization) that answers the question. Turn in a word/pdf file with a **written answer to the question** and **a screenshot of your sheet** that determined the answer. It is ok if your one screenshot does not fully answer the question; I realize that clicking and scrolling is probably needed within your sheet answer some questions.

First, let's explore the *Species Counts* data.

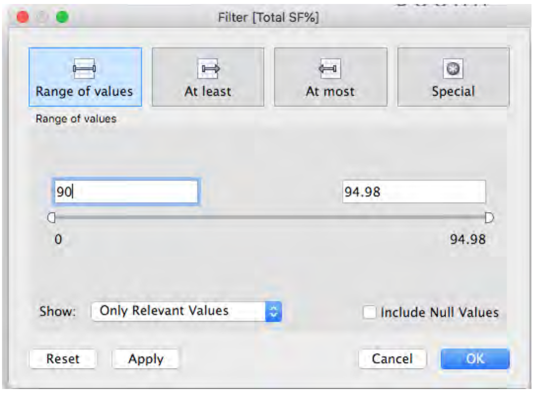

Data Analytics
Species Counts (Overall...
Survey Counts (Overall_...

9. Compare *novice sighting frequency* to *expert sighting frequency* for each *species*. **List 5-10 species** that experts seem to identify more frequently than novices. Hint: try a side-by-side bar chart.
    e. Represent the insight in the simplest visualization that gets the story (key takeaway) across. Write a short title that clarifies what the key takeaway is for the reader. **Add your name to the title to get credit for YOUR work**, e.g., "Differential fish counts by expertise level - Leidig".


Columns: Species, Measure Names
Rows: Measure Values
Filters: Measure Names
Marks: Bar
Measure Values: Expert SF%, Novice SF%
Differential fish counts by expertise level - Leidig

10. Using the same chart, are there any species that *novices* seem to identify more *frequently* than *experts*? **List any**.

11. Some species are very prevalent and are observed on almost every fish survey. Which species are seen on at least 90% (or more) of total surveys? **Show all of the ~11 species**. Hint: drag the field *Total SF%* to the Filters pane and then only show species found at least 90% of the time.


Filter [Total SF%]

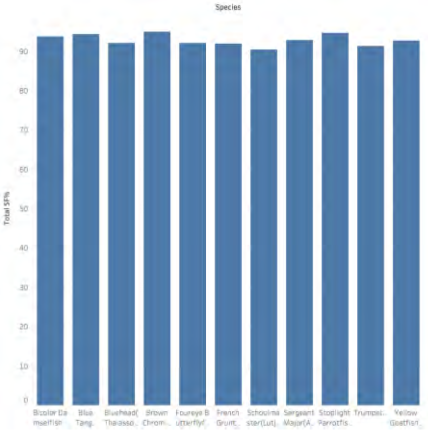
Range of values

90 — 94.98

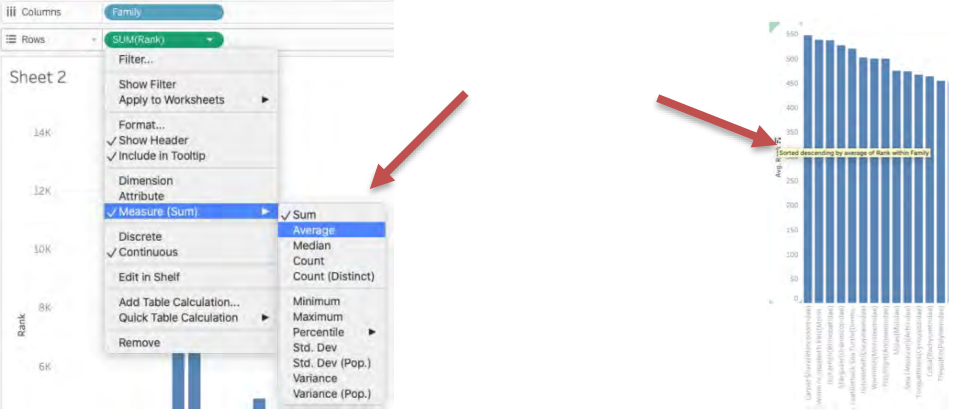
0 — 94.98

Show: Only Relevant Values

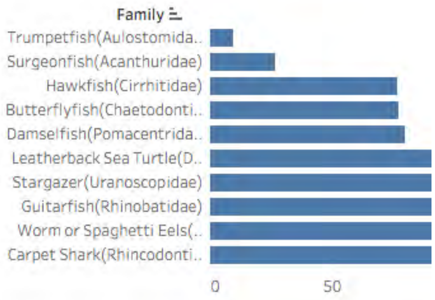
☐ Include Null Values



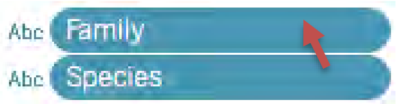
12. Recall from Biology, multiple species are organized into a family of similar organisms (Homo sapiens fall under the Hominidae family). E.g., *Brown Chromis* and *Bicolor Damselfish* are both grouped into the *Damselfish* family. The *species* with a *rank* of 1 is seen most frequently and the *species* with a *rank* of 547 is seen the least frequently.
    f. Generate a **new sheet**/graph that shows the *average* (avg) *rank* of the high-level *family* groups. Hint: you can sort the data within your charts to make it easier to answer the questions.

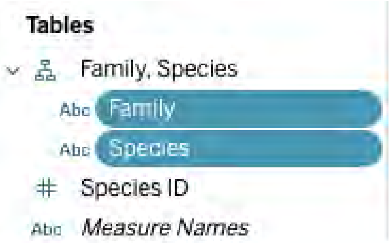


    g. Hint: generate a graph of *family* compared to Avg(*Rank*). Tableau defaults to calculating the Measure (Sum) instead of Avg – but that can be changed in the dropdown from *SUM(Rank)*.
    h. What are the **top 5 most likely families to be observed** - with the lowest Avg(Rank)?
    i. What are the **bottom 5 least likely families to be observed** - with the highest Avg(Rank)?

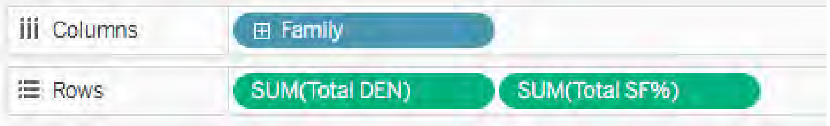


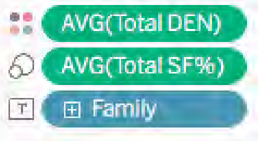
13. Dragging the *Species* dimension on top of the *Family* dimension in the left pane will create a hierarchy. Thus, the *Family* category will be the parent to the *Species* sub-category (Family,Species).

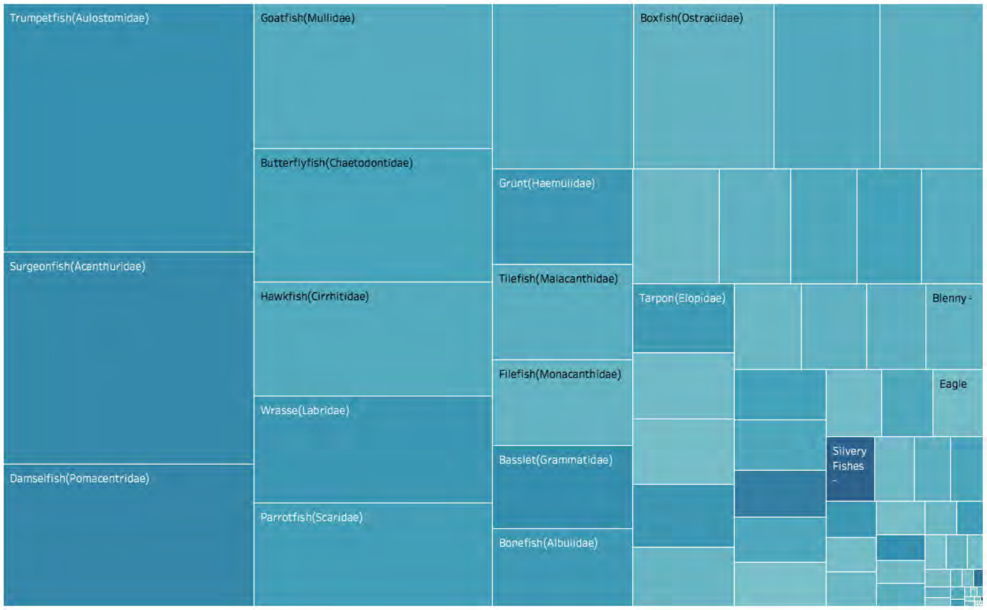




j. Generate a new Treemap sheet showing the average *total sighting frequency* and average *total density* for each *family,species* using size & color. It may be helpful to set *SF* to area and *DEN* to color.


Columns: Family
Rows: SUM(Total DEN) SUM(Total SF%)

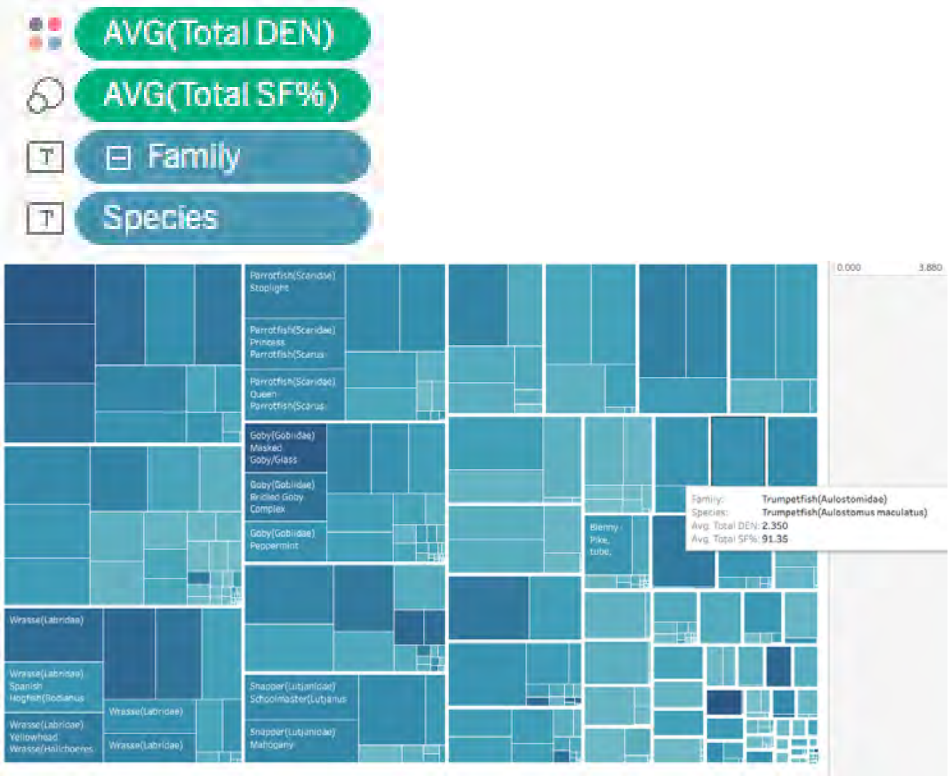

AVG(Total DEN)
AVG(Total SF%)
Family



k. Which *families* have the **highest *densities***?

l. Which *families* have the **highest *sighting freq***?
m. **Patterns**: Does *density* seem to be roughly correlated with *sighting frequency*? E.g., *families* that are seen in higher abundance are also reported more frequently?
n. **Outliers**: Are any *families* observed with low frequency but high density? E.g., a *families* that is rarely seen, but when it is seen there are large numbers?
o. **Outliers**: Are any *families* observed with high frequency but low density? E.g., a *families* that is often observed but generally observed all on its alone.
p. Side note: Observe how Trumpetfish have the highest sighting frequency at the *family* level.

14. From the same style chart, click to expand the *family* category to show the breakdown for each *species*.


AVG(Total DEN)
AVG(Total SF%)
⊟ Family
Species
Parrotfish(Scaridae) Stoplight
Parrotfish(Scaridae) Princess Parrotfish(Scarus
Parrotfish(Scaridae) Queen Parrotfish(Scarus
Goby(Gobiidae) Masked Goby/Glass
Goby(Gobiidae) Bridled Goby Complex
Goby(Gobiidae) Peppermint
Wrasse(Labridae)
Wrasse(Labridae) Spanish Hogfish(Bodianus
Wrasse(Labridae) Yellowhead Wrasse(Halichoeres
Wrasse(Labridae)
Wrasse(Labridae)
Snapper(Lutjanidae) Schoolmaster(Lutjanus
Snapper(Lutjanidae) Mahogany
Blenny - Pike, tube,
0.000
3.880

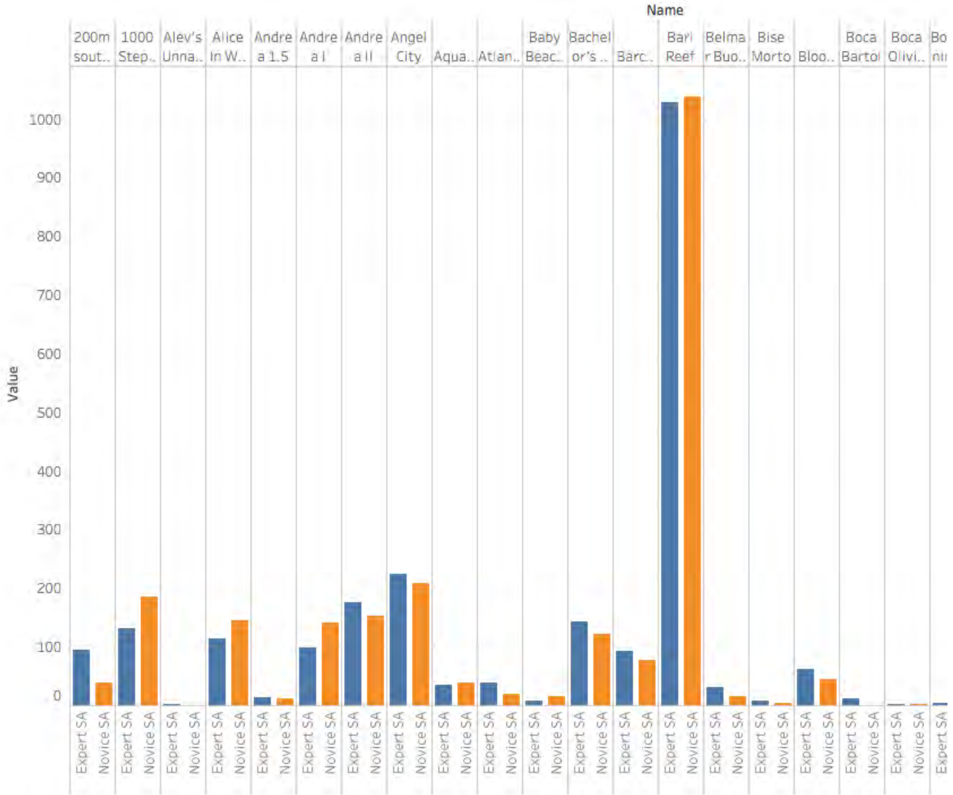
Family: Trumpetfish(Aulostomidae)
Species: Trumpetfish(Aulostomus maculatus)
Avg. Total DEN: 2.350
Avg. Total SF%: 91.35

q. Find the Trumpetfish sub-category – it may take some hunting! Let's compare the Trumpetfish *family* to the Damselfish *family*. There is only one *species* of Trumpetfish in our data while there are 14 Damselfish *species*. Note from the prior questions that the Trumpetfish *family* is observed on 91% of surveys while the average Damselfish *family* member is only observed on 52% of surveys. In this view which shows all 547 species, the set of 14 *species* of Damselfish seem to be sighted much more frequently than the Trumpetfish!
r. In a few sentences, interpret these two charts to **explain this paradox**. Suppose a tourist is going snorkeling/SCUBA diving at this location and is confused if they will see any Damselfish or Trumpetfish underwater. *Describe to the tourist what they are likely to observe underwater given these last two charts.*

Next, let's explore the *Survey Counts* dataset.

15. **SA** columns indicate that the surveyor reported the ***S**pecies and their **A**bundance*, while **SO** columns indicate that the surveyor ***O**nly reported the **S**pecies* they saw and not the abundance level. **Name** refers to the name of a coral reef that was surveyed.

s. Let’s figure out whether experts or novices are providing data for each reef. Make a **side-by-side bar plot** that compares the total (Sum) number of Expert SA reports to the total (Sum) number of Novice SA reports for each reef site.

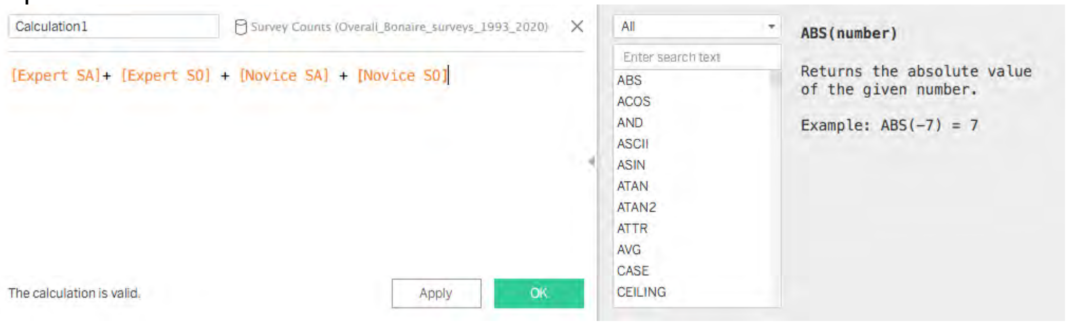


t. Let’s assume that experts provide more accurate survey data than novices. Are there any reef sites that have significantly **more data from novice surveys than expert surveys**? If so, analysis regarding these sites may be incomplete or misleading! List any such reef names (hint: ~3-5 locations).

u. Side note: On average, experts report 80.86 species per survey and novices report 55.64 species.

16. Let’s create a calculated field that counts how many total surveys are available per location, regardless of the person. To do this, we will add a new measure that simply adds the ExpertSA, ExpertSO, NoviceSA, and Novice SO columns.

v. From the top menu, select Analysis >> Create calculated field, and add the equation.



Calculation1

Survey Counts (Overall_Bonaire_surveys_1993_2020)

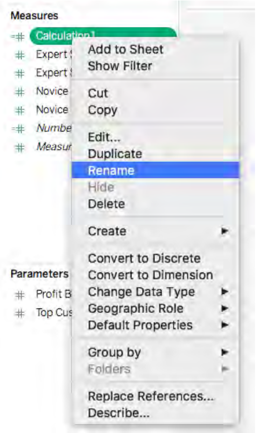
```
[Expert SA]+ [Expert SO] + [Novice SA] + [Novice SO]
```

The calculation is valid.

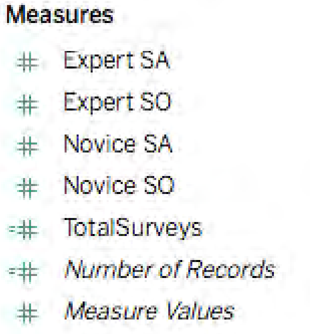
ABS(number)

Returns the absolute value of the given number.

Example: ABS(-7) = 7

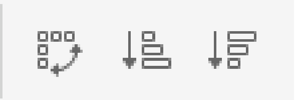


w. Rename Calculation1 to TotalSurveys.

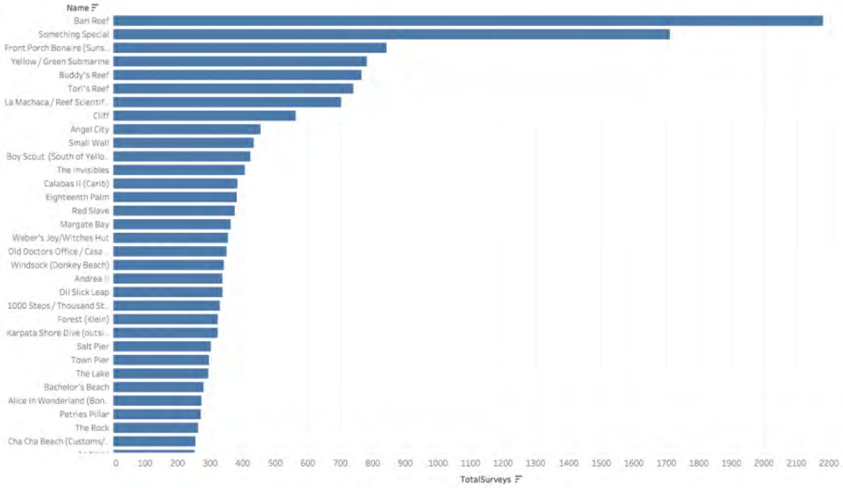


x. Create a sheet of the *total surveys* per reef *name* using a chart type of your choice, e.g., bar.

y. Sort the reefs by deceasing numbers of TotalSurveys in the top menu. This menu option can sort ascending, descending, or transpose (flip the rows and columns).

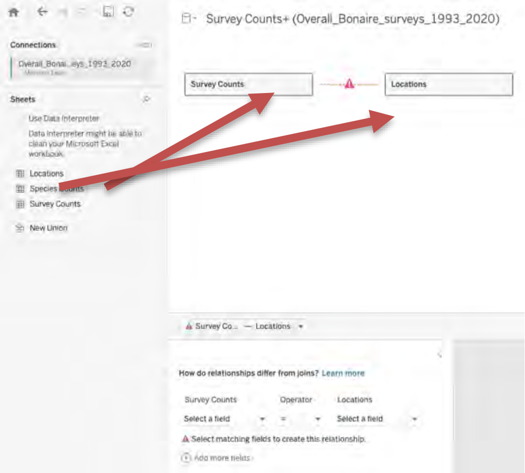


z. What are the **names of the top seven reefs** by survey count? Were any of these locations identified as being **possibly biased** due to too many novices from question 10?



Next, let's place these datapoints on a geospatial map.

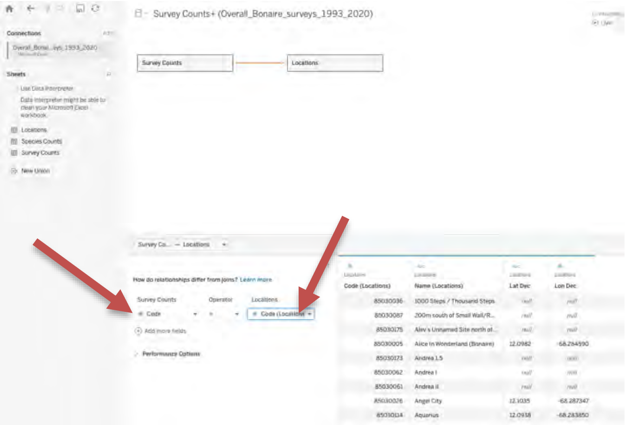
17. Add a new datasource out of the Excel file's '*Locations'* worksheet. Open the Excel file as a new data source. Drag both *Survey Counts* and *Locations* worksheets into the top pane.


Survey Counts+ (Overall_Bonaire_surveys_1993_2020)
Connections
Overall_Bonai...eys_1993_2020
Microsoft Excel
Sheets
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Locations
Species Counts
Survey Counts
New Union
Survey Counts
Locations
Survey Co... — Locations

How do relationships differ from joins? Learn more

Survey Counts	Operator	Locations
Select a field	=	Select a field

Select matching fields to create this relationship.


Survey Counts+ (Overall_Bonaire_surveys_1993_2020)
Connections
Overall_Bonai...eys_1993_2020
Sheets
Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.
Locations
Species Counts
Survey Counts
New Union
Survey Counts
Locations
Survey Co... — Locations

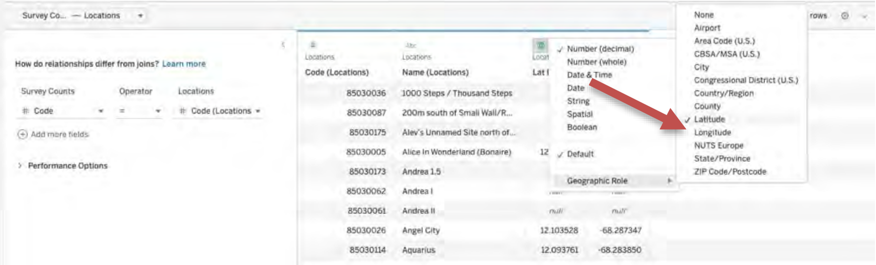
How do relationships differ from joins? Learn more

Survey Counts	Operator	Locations
Code	=	Code (Locations

Code (Locations)	Name (Locations)	Lat Dec	Lon Dec
85030036	1000 Steps / Thousand Steps	null	null
85030087	200m south of Small Wall/R...	null	null
85030175	Alev's Unnamed Site north of...	null	null
85030005	Alice In Wonderland (Bonaire)	12.0982	-68.284590
85030173	Andrea 1.5	null	null
85030062	Andrea I	null	null
85030061	Andrea II	null	null
85030026	Angel City	12.1035	-68.287347
85030114	Aquarius	12.0938	-68.283850

18. Rows from two separate datasources can be merged into a larger row as long as they both contain an identical identifier that can be matched. We are going to join the survey data for each reef location (a row in the *Survey Counts* table) with its actual Lat/Lon coordinates (from the *Locations* table). Edit the relationship to force Tableau to only match rows from the *Survey Counts* and *Location* datasets together if they discuss the same location *Code*. Each reef site gets its own unique *code* and *name*.

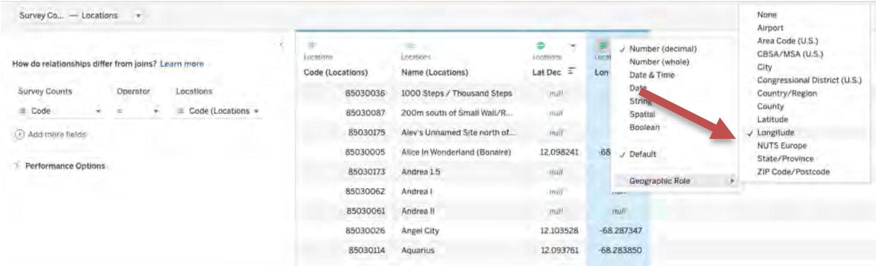
19. Edit both *Lat Dec* and *Lon Dec* columns so that Tableau interprets these as decimal numbers and as latitude and longitude data points. Also, select to use them as columns with actual Lat/Lon coordinates.



How do relationships differ from joins? Learn more

Survey Counts	Operator	Locations
Code	=	Code (Locations

Code (Locations)	Name (Locations)	Lat I	
85030036	1000 Steps / Thousand Steps		
85030087	200m south of Small Wall/R...		
85030175	Alev's Unnamed Site north of...		
85030005	Alice In Wonderland (Bonaire)	12	
85030173	Andrea 1.5		
85030062	Andrea I		
85030061	Andrea II	null	null
85030026	Angel City	12.103528	-68.287347
85030114	Aquarius	12.093761	-68.283850



Survey Co... — Locations

How do relationships differ from joins? Learn more

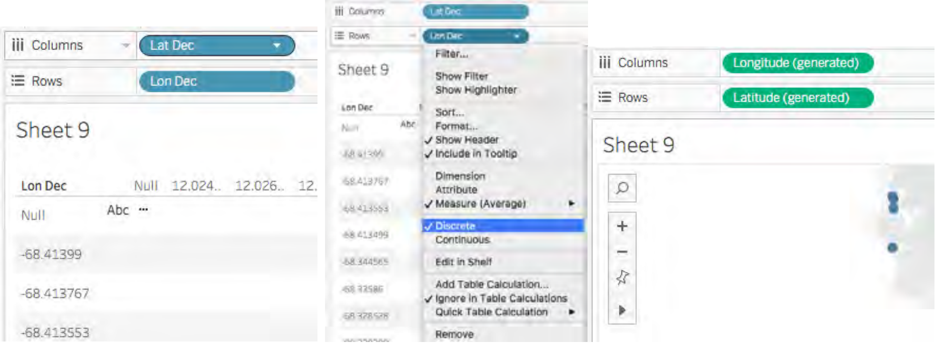
Survey Counts	Operator	Locations
Code	=	Code (Locations)

Add more fields

Performance Options

Code (Locations)	Name (Locations)	Lat Dec	Lon
85030036	1000 Steps / Thousand Steps	null	
85030087	200m south of Small Wall/R...	null	
85030175	Alev's Unnamed Site north of...	null	
85030005	Alice In Wonderland (Bonaire)	12.098241	-68
85030173	Andrea 1.5	null	
85030062	Andrea I	null	
85030061	Andrea II	null	null
85030026	Angel City	12.103528	-68.287347
85030114	Aquarius	12.093761	-68.283850

20. Create a new sheet for our map from our new *Survey Counts*+ dataset. Drag *Lat Dec* and *Lon Dec* to the chart. You may have to edit the two pills to make them both Discrete.



Columns: Lat Dec
Rows: Lon Dec

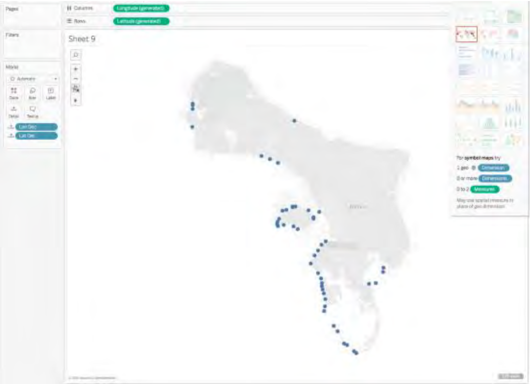
Sheet 9

Lon Dec	Null	12.024..	12.026..	12.
Null	Abc	···		
-68.41399				
-68.413767				
-68.413553				

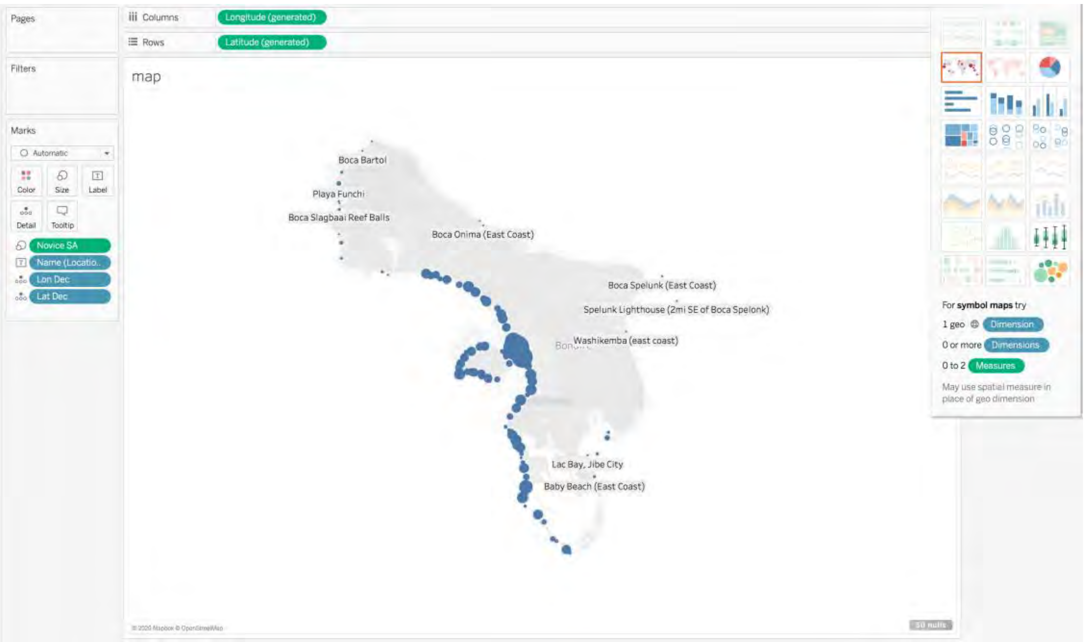
Columns: Longitude (generated)
Rows: Latitude (generated)

Sheet 9

21. If Tableau interprets your lat/lon coordinates correctly, it will generate a map with a point for each row.



22. Add *Expert SA* to set the size of each point and *Name* to each point's label. Hint: On the left Tables pane, you may have to select the *Lat Dec* and *Lon Dec* dropdown menus and *Convert them to Discrete*. When you add *Expert SA* to the *Marks* pane, you may have to change *Measure(Sum)* to *Attribute*.
23. **Where do experts** seem to conduct the most surveys?
24. Reflect+answer: Why might the Southwest coast of the island be **heavily surveyed and not** the East coast? Could this **bias the type and abundance** of species within the surveys of this trusted dataset?



Columns: Longitude (generated)

Rows: Latitude (generated)

Marks: Automatic

Novice SA

Name (Locatio..

Lon Dec

Lat Dec

25. Save and turn in your final Tableau project (.twbx file). Upload a Word or PDF file that contains your answers to each question along with screenshots of your Tableau charts. For full credit, text and screenshots are required for all questions above.