# MIDWEST COLLEGIATE COMPUTING CONFERENCE

March 21-22, 2024 www.mwc3.org



### **Contest: Data Analysis and Visualization**

### PLEASE DO NOT OPEN UNTIL INSTRUCTED TO DO SO

#### **IMPORTANT:**

Any solution received that includes any **IDENTIFYING** information will not be eligible for anything beyond "Honorable Mention."

Examples include any of the following:

- Names of the individuals or school names in the solution
- Names of the individuals or school in the "properties" of the file
  - Each program, Adobe, Microsoft, etc. contains metadata about the document. Remove ALL of these from your document. If you do not know how please ask someone!

Any solution that does not provide accurate citing of professional resources will be removed from consideration.

Examples include any of the following:

- Copying and pasting diagrams and images from a website
- Using descriptions and product data verbatim from source
- When in doubt, cite your source

## **Data Analysis and Visualization Contest**

#### **Contest Statement:**

Common techniques in *data analytics and visualization* include data acquisition, data cleaning, handling missing values, wrangling, data integration, simple data analysis (e.g., outlier detection, identifying collinearity between variables, dimension reduction, summary statistics, and determining skewed attributes), statistical model building, finding patterns, finding clusters, plotting geo-temporal points on a map, data visualization, visual analytics, problem solving, making recommendations, and providing quantitative analysis to support decision making.

Each team may consist of one or two individuals and must have at least one computer to complete your project. The contest is written from the perspective of guiding someone who is new to using Tableau. Appendix A contains a quick-start guide to Tableau's interface. You are welcome to use other software and skip the Tableau focused tutorial instructions found in Appendix B – you are only required to turn in your answers and a visualization (using any software of your choice) for the required tasks.

#### **Contest - Deliverables**

Solve the tasks in an individual or two-person team – sharing completed solutions with other teams is not allowed. You are welcome to utilize existing software to manipulate/analyze data, write your own code, or follow any other analytical methodology you prefer to arrive at a solution to each task. To complete this challenge, create and turn in a document that answers as many tasks as you can. Also, turn in any other supporting files such as computing code (e.g., Python scripts), Tableau workbooks (.twbx file), intermediate datasets (e.g., CSV files), etc. to showcase your work. Do not upload any part of this contest or your deliverables to online locations (e.g., GitHub).

#### **Resources - Software recommendations**

It is recommended to download Tableau Desktop (free year-long license) or else use online Tableau Cloud for the contest (see: https://www.tableau.com/academic/students). Alternative analytics or visualization software that students may find useful include Excel, R, Python, SAS, SPSS, Weka, and/or Power BI. Students are encouraged to use any software and online resources they wish, including Google, to learn and complete the challenges.







If you choose to use **Python** for this contest, SciPy is a set of widely used packages for managing, analyzing, and visualization large-scale content. It consists of libraries for data structures such as DataFrames and analysis (pandas), N-D arrays (NumPy), 2D graph plotting (Matplotlib), scientific analysis (SciPy), etc. In addition to matplotlib, other popular python libraries include *ggplot (any R fans in the group), 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), etc. See:

https://www.python.org/about/gettingstarted https://docs.python.org/3/tutorial https://scipy.org https://matplotlib.org

Tableau and PowerBI are leading software tools for visual analytics and rapidly generate interactive

visualizations. You may download Tableau Desktop from tableau.com with a 14-day free trial, or perhaps Tableau Cloud. Students (worldwide) get free renewable year-long licenses for Tableau Desktop (with a valid university email address), see https://www.tableau.com/academic/students. Getting the license key will likely only take a few seconds. If you require brief tutorials on how to use the GUI and its functionality, see the APPENDIX A below, or visit https://www.tableau.com/learn/training. Do NOT post and make your work visible for others to see. Tableau connects with a variety of underlying dataset formats: offline data file on your hard drive, online databases, an online data server, or the default practice datasets (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 chart choice which is highlighted with a red box on the list of possible charts on the right pane). You might use this contest as an opportunity to become familiar with Tableau, how to connect to a dataset, how to create charts (sheets), and how to create interactive dashboards. Some of the work for each task can be done in **Excel**.

#### **Contest - Rubric**

The following weighted rubric details how the results will be evaluated. The final submission should be organized in a single Word or PDF file (with legible screenshots if needed). Provide solutions that include all the required deliverables for each task that you are able to complete. Partial solutions will be given up to half credit. Sort your solved tasks A-G in their proper order. Clearly indicate which tasks you attempted and which tasks you skipped. The earliest timestamp for the final pdf answer submission will be used to break any ties.

Task A through Task G (equally weighted for each task solved correctly)

#### Contest - Case Study Background

Bonaire is an island in the Caribbean Netherlands (next to Aruba and Curacao) located a few miles off the coast of Venezula. It is well-known for its healthy coral reefs, salt industry, natural park, and tourism based on SCUBA diving and water sports. There are over 100 coral reefs locationed around the coast of Bonaire. Let's use visually explore marine biology data from this island (provided by <u>www.reef.org</u>). Over the past several decades, the REEF organization collected underwater surveys from SCUBA divers at each of these locations. The surveyors were marine biology experts as well as novice citizens scientists.

In the dataset we will use, **Name** refers to a geographic area of coral reef that was surveyed. The corresponding **Code** is a unique number that was generated for each reef location. Recall from your basic biology courses that the Taxa for family and genus are broken down into individual species. On these survey reports, **SA** columns indicate that the surveyor reported the *Species and the species' Abundance level*, while **SO** columns indicate that the surveyor *reported the Species Only* and not the abundance level of the species. The sighting frequency (**SF%**) refers to the percent of reports in which surveyors reported that species. E.g., experts report seeing the Blue Tang species on 97.9% of surveys. **DEN** refers to the density (abundance) of the species when it is observed. E.g., 1 means the species was solitary and only one was observed, while 4 means a very large school of this fish species was observed.

Given the Overall\_bonaire\_surveys\_1993\_2020.xlsx file, solve as many of the following tasks as you can. Create visualizations (e.g., a sheet within Tableau) that solve the following tasks. Turn in a Word or

PDF file with a **written answer** to each task and a **screenshot** of your visualization that visually supports your answer. The optional Appendix B provides a very helpful walkthrough tutorial for using Tableau to load the data, understand the case study, and answer these types of questions. If you are new to Tableau, I highly recommend you begin by reading the **Context Tasks** section, and then proceed through Appendix B, which will prepare you to solve the contest's actual tasks.

#### Contest - Tasks to solve

Task A: Novice and expert surveyors are able to identify different sets of fish, which results in differing reports. What is the top **species** that novices over-reported most frequently compared to experts? I.e., in comparing the SF% for species that are more often reported by novices, which species has the largest difference between the two groups?

Task B: Which **species** is ranked last with the highest Avg(Rank) (i.e., least likely to actually be identified and reported)?

Task C: What are the top most likely **family** to be observed - with the lowest Avg(Rank)?

Task D: One hotel (cleverly) paid a group of expert marine biologists to survey the reef closest to their hotel. That reef is now considered one of the most biodiverse reefs in the Caribbean due to scientists' expertise and the hotel's ploy. What **reef name** had the largest number of total expert surveys completed (Expert SA and Expert SO surveys)? How many total expert surveys were completed?

Task E: What **reef name** had more novice surveys completed at that location compared to expert surveys? How many **more** novices completed surveys at that location than experts?

Task F: What is the northernmost location (i.e., **name** of the location)?

Task G: Based on this dataset and your general background knowledge on island life, 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?





#### Appendix B: Optional Walkthrough for Solving These Tasks in Tableau

**NOTE**: This is a *teaching guide* to increase your familiarity with Tableau. Don't turn in the results from Appendix B! Your actual tasks to complete for the contest are in the <u>Contest – Tasks to solve</u> section.

Bonaire is an island in the Caribbean Netherlands (next to Aruba and Curacao) located a few miles off the coast of Venezula. It is well-known for its healthy coral reefs, salt industry, natural park, and tourism based on SCUBA diving and water sports. Let's use Tableau to explore Bonaire's Marine Biology data (from www.reef.org).

- 1. Download the Overall\_bonaire\_surveys\_1993\_2020.xlsx file to your computer (email <u>leidijon@gvsu.edu</u> to receive a copy of the dataset if you do not have access to the files). You can open the file in Excel to check the data.
- 2. Import the file in Tableau a collection of fish surveys reported 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 three sheets of data for us to import.

Connect Search for Data Tableau Server To a File Microsoft Excel Text file JSON file PDF file Spatial file	File     Data     Worksheet     Dashboard     Analysis     Map     Format     Help     New Workbook       Connect to Data     Connect to Data     Connect to the data you need to visualize.     Learn more       Data     Connect to the data you need to visualize.     Learn more       On This Site     Files     Connectors     Accelerators
Statistical file More To a Server Vertica Web Data Connector Other Databases (JDBC) Other Databases (JDBC) More > Saved Data Sources Sample - Superstore World Indicators	On Tableau Cloud Drag and drop afile or Upload from computer

3. We will start by adding 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 a second time, and then add the other *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/Analytics menu. Ignore the *Locations* datasheet from the Excel file for now. The data can always be reviewed in the Data Source tab on the bottom-left corner. Every Tableau chart is designed in a separate new Sheet.

$\ast \hspace{0.2cm} \leftarrow \hspace{0.2cm} \rightarrow \hspace{0.2cm} \blacksquare \hspace{0.2cm} \bigcirc \hspace{0.2cm}$	B- Overall_Bonaire_surveys_1993_2020	$ \begin{array}{c} \bigstar  \leftarrow \ \Rightarrow \ \hookrightarrow \ \bigtriangledown \ \Diamond \\ \hline \cr \  \  \  \  \  \  \  \  \  \  \  \  \$	B · Species Counts	(Overall_Bonaire_surveys_195	93_2020)						Connection (® LMe C	Edmant	Filters ©   Add
Connections Add Overall_Bona_ys_1993_2020 Mercent fore beta Interpreter Data Interpreter Data Interpreter might be able to clear-your Mercesht Excel workbook.		Control Review, 49, 2782, 2003     Control Review, 49, 2782, 2003     Control Review     Contro     Control Review     Control Review     Control Review     Con	Species Counts				Need more Drag tables here to relate	data? Item. Laco mace					
I Species Counts			Species Counts	· 10 fields \$47 rows								300	- revs. 0
I Survey Counts			Name		e Species Court. Rank	e Species Dourts Species ID	Ann Tgarcins Counts <b>Tgarcins</b>	Ale Species Courts Family	e Species Counts Tablel SP16	Recto Courts	Species Courts Expert SF16	Barries Courts Expert DEN	* Species Courts S Nonice SP16 8
Rew Union	🚥 💷 Sant Kalda Data any sa andar		speces courts		1	43	Brown Chromis/Chromis mut	Danselfsh(Pomacentridae)	94,9800	3.88000	977300	3.93000	92,0600
	i Sort nerds Data source order		Fields		2	155	Stopight Parothish(Spariss	Panotish(Scaridae)	94,9900	2.82000	978600	2.88000	91.1300
			Type Field Name	Physical Table Remote Field N		204	Bue targ/campurus coetu	Surgeon technology	94,2400	122000	173000	2,84000	90,3500
			<ul> <li>Rask</li> <li>Standard D</li> </ul>	Species Counts Rank Species Counts Species (D)		52	Sergeant Mejor(Abudeldul s	Damselfish(Pomacentridae)	92,8700	3.03000	973000	1.09000	88.3800
			in Species of	Species Courts Species	6	77	Nelow Gostfish(Mulloidichth	Gostfish(Wulldar)	92.6500	2,99000	97.2200	3.07000	878000
			An Family	Species Courts Family	7	31	Foursye Butterflyfish(Chaeto	Butterflyfish(Chantadortidae)	92.0300	2.23000	95.6900	2.20000	88.3400
		1	Total SP%	Species Counts Total SP%		253	Buehead(Thalassome bifasc	Wrasse(Labridae)	95,9900	3.30000	96.9600	3.50000	86.7300
		IL In Worksheet	<ul> <li>Total DEN</li> </ul>	Species Counts Total DEN	9	108	French Grunt(Haemulon Kav	Grunt(Haemulidae)	91.8300	2,70000	971200	2,78000	86,2300
		O Data Source Source De By By D	Expert SP%	Species Counts Expert SP1s	10	239	ranger any abbitorius ha	( and a construction of the	1.1500	2,35000	94,900	2,96000	-7/00

★         ⇒	Survey Counts (Ove     Survey Counts	eral_Bonaire_surveys_199:	3_2020)							$\begin{array}{c c c c c c c c c c c c c c c c c c c $	
Data Understehr might be able to celan your Vicrosoft Deat workbook El Goardines El Sorcies Counts El Sorcies Counts Envery Counts Envery Counts				Drag tab	Need mon ies here to relat	e data? In them. Learn.mr	008			Data/analytics r	nenu
g recom	Server Courts -	7 faith 168 may								A data fan de la constante de	
										Expert SA     Drop	
	Name		Survey Courts	Burney Courts	Survey Counts	Burvey Courts	Survey Courts	Survey Counts	Survey Counts	Expert SO field here	
	Survey Counts		Code	Kane	EpertSA	Expert 50	Novice SA	Novice 50	Boltománe (HM)	Novice SO	
	Excel.		85030002	Calabas Reef (Divi Farrings	00		308	50	5 1/8/2900-434:00	<ul> <li>Number of Records</li> </ul>	
	Tute Field Name	Physical Table Remote F	85030003	Sampler (Kiein)	76	0	130		5 1/8/2900 812:00	Data Source tab	
	a (ata	Survey County Code	85030004	Margate Bay	292	6	190	34	4 1/26/2900 32:36:0		
	In Same	Survey Counts Name	85030005	Alice In Wanderland (Bonaire)	113	1	145	13	3 1/12/1900 439:00		
	<ul> <li>Extent SA</li> </ul>	Survey County Expert SA	85030006	Knile (Klein)	64		95	2	2 1/6/2900 9:51:00		
	Expert SO	Survey Counts Expert 50	85030007	Karpata Reserve Site 1	59	3	90		4 1/6/2900 9/29:00	/ / Now choots /ch	arte
	Novice SA	Survey Counts Novice SA	85030008	Twin Rocks Reserve (North R_	26	0	63	0	1/2/290010.26:0	/ New sneets/cha	1115
	Novice 50	Survey Counts Novice SO	85030009	Jarry's Jam / Ebo's Special (	55	2	69	7	7 1/5/290010.36.00		
<b>•</b>	C. Bottantine (1.M)	Survey Counts Buttontim	85030010	Forest (Klein)	150	0			7 1/13/1900 653:00	Read Barriel Contraction of the State	

On Tableau Cloud, the process is the same, the top menu simply looks slightly different.

File	Data	Worksnee	t Da	ishboard	Analysis	Мар	Format	Help				New W	orkbook
$\leftarrow$	New	Data Source					ļ	Σ.	<u>@</u> *	<b>T</b>	Z	<u> </u>	8 <u>.</u> 11 ×
Data	Spec	cies Counts (C	werall_	Bonaire_su	irveys_1993_	2020) 🖡	iii Col	umns					
🖯 Spr	Surv	ey Counts (O	/erall_E	Bonaire_sur	veys_1993_2	2020) 🕨	I Roy	vs					
🖯 Sur	vey Cou	unts (Overall		~ Filter	'S								
							She	et 1					
, p s	earch	γ.	*	~ Mark	is								
Tabl	es			T Aut	omatic	Ŧ							
=	Rottor	ntime (H·M)											

Answer these questions to yourself but DO NOT turn in:

- 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.

When importing the Survey Counts sheet from Excel, note that the variable Bottomtime is a number of hours and minutes (which Tableau sometimes mistakes for a date). In the data source pane, let's switch that back to Number(decimal).

Survey Counts Bottomtime (H:M) =	# ✓ Number (decimal) Number (whole)	# Survey Counts Bottomtime (H:M) =
4/29/1900	Date	120.8
3/30/1900	Boolean	90.18
2/11/1900	Default	42.54
2/10/1900	Geographic Role	41.56
2/10/1900	35.64	39.78
2/8/1900	27.84	35.64
2/4/1900	24.91	07.0

- c. For more details on the data source, checkout: <u>https://www.reef.org/db/reports/geo?end\_date=2020-02-</u> <u>10&format\_type=chart&group\_type=species&language=common&region\_code=TWA&start\_dat</u> <u>e=1993-01-01&zone\_map=0&zones=8503</u>
- 4. Use *Sheet 1*. 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 interactive graphic), dragging columns to the filter pane generates dynamic filters within the visualization (allowing you to filter out undesired data), and the right pane allows you to change the chart type depending on the columns already selected.
- 5. 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 by a red square on the list of possible charts on the right pane depending on the columns you selected).



Next, use Tableau to explore the dataset. I recommend making a new visualization for each task, that way you can go back to earlier steps.

First, let's explore the Species Counts data. In sheet 1, select the Species Counts dataset you have imported.



- 6. Compare *novice sighting frequency* to *expert sighting frequency* for each *species*. Find *5-10 species* that experts seem to identify more frequently than novices. Hint: try a side-by-side bar chart and show individual datapoints (click Analysis in the top menu and then unselect Aggregate Measures).
  - d. You would ideally 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. Remember to add titles to your work, e.g., "Differential fish counts by expertise level - Leidig".



7. Using the new chart, how much more do *experts* seem to identify these species more *frequently* than *novices*? A calculated field would make this much easier to solve. Right click in the left pane within the Tables section and select *Add Calculated Field…* Write a simple calculation to generate a new column (*DifferenceInSF*) for each row in the data table that subtracts Novice SF% from Expert SF%, calculating how much better experts are at sighting each species than a novice. It looks like experts identify and report Rainbow Wrasse more often than novices.

Species Counts (Overall     Survey Counts+ (Overall     Filters											
Search $\rho$ $\nabla$ $\blacksquare$ $\bullet$		Data	Analytics <	Pages			iii Column	s I	Species	F	
		🗟 Spec	ies Counts (Overall				i≣ Rows	(	SUM(Diffe	renceInSF)	
<ul> <li>Locations</li> <li>Code (Locations)</li> <li>Lat Dec</li> <li>Marks</li> </ul>		Search Tables	P Counts+ (Overall	Filters			Do exp	perts re	eport m	iore sighti	ngs?
Lon Dec	T	Abc Fam Abc Sper	nily cies	Marks			50				
# Locations (Count)	ex	# Sper Abc Mea -# Diffe	cies ID as <i>ure Names</i> erenceInSF	nl A	utomatic	*	40		<b>.</b>		
<ul> <li>Survey Counts</li> <li>Bottomtime (H:M)</li> </ul>		# Expe # Expe # Novi	ert DEN ert SF% ice DEN	Color coo Detail	Size C Tooltip	Label	h. 30				
# Code Abc Name		# Novi # Rani # Tota	ice SF% k al DEN				rencelnSF :				
#         E-N         Create Calculated Field           #         Expe         Create Parameter           #         Expe         Create Folder (use group by folder)	DifferenceInSF   Species Counts (Overall_Bonaire_surveys_1993_2020) X	# Tota # Spev # Mea	al SF% cies Counts (Count) asure Values				Diffe				
## N-E Group by Folder     # Novic ✓ Group by Data Source Table	[Expert SF%]-[Novice SF%]						10 -				
<ul> <li>** Novic Sort by Name</li> <li>*# Total Sort by Data Source Order</li> </ul>	>						0				
Hide All Unused Fields Abc Measur Show Hidden Fields								e(Halichoer snatholepis ty(Acanthe	/(Elacatinu alichoeres phopterus molex (Brid	(Elacatinus . y(Coryphop. ass Goby(C. (Elacatinus .	sh(Scarus is. Ish(Stegast. Ifish/Neoni
# Measur Expand All Collapse All	The calculation is valid. Apply OK							Rainbow Wrassi Goldspot Goby(( Spinyhead Blenr	Yellownose Goor Clown Wrasse(F Pallid Goby(Cory Bridled Goby Cor	Yellowline Goby Peppermint Gob Masked Goby/GI Sharknose Goby/	Stripad Parrotfi Longfin Damself Londiaw Souirre

Create a new sheet (Sheet 2). I recommend making new sheet and chart for each question. 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: create a chart for *species* and *TotalSF*%, drag the field *Total SF*% to the Filters pane, and then only show species found at least 90% of the time.

Filter [Total SF%]	Species
Range of values At least At most Special	Sort the data
90 94.98 0 94.98	60 12 45 19 40 20
Show: Only Relevant Values 🔉 🗌 Include Null Values	20
Reset Apply Cancel OK	9. Brown Steplight Blue Bictor Da Sargast Yellow Fourys & Bluehad French Trumpet. Schoolma Chrom. Farretti. Targ. mailfild. Major(A. Gastribu. Litter/fr. Thalasso. Enviro. mt 2014)

- 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.
  - e. Generate a new sheet/chart that shows the average (avg) rank of the high-level family groups.

iii Columns ⊞ Rows Filter.. Sheet 2 Show Filte Apply to Worksheets Format... √ Show Header √ Include in Tooltip 14K Dimension 128 Attribute √ Sum Discrete Median 10K Continuous Edit in Shelf Count (Distinct) Add Table Calculation. Quick Table Calculation Minimum Maximum Percentile Sank Remove Std. Dev Std. Dev (Pop.) Variance Variance (Pop.)

Hint: you can sort the data within your charts to make it easier to answer the questions.

- f. 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)*.
- g. What are the top 5 most likely families to be observed with the lowest Avg(Rank)?



10. 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).

	Tables
	✓ 品 Family, Species
	Abc Family
the Family	Abc Species
Ade Fairing	# Species ID
Abc Species	Abc Measure Names

h. 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.

iii Columns	E Family		1	<b>::</b> A ର A	VG(Total DEN) VG(Total SF%)
⊞ Rows	UM(Total DEN) SU	M(Total SF%)		T (	] Family
Trumpetfish(Aulostomidae)	Goatfish(Mullidae)		Boxfish(Ostraciidae)		
Surgeonfish(Acanthuridae)	Butterflyfish(Chaetodontidae)	Grunt(Haemulidae)			
	Hawkfish(Cirrhitidae)	Tilefish(Malacanthidae)	Tarpon(Elopidae)		Blenny -
Damselfich(Pomacentridae)	Wrasse(Labridae)	Basslet(Grammatidae)		Silv	ery les
	Parrotfish(Scaridae)	Bonefish(Albulidae)			

- i. Which families have the highest densities? A schooling family like silvery bait fishes perhaps?
- 11. From the same style chart, click *family* to expand the category to show the breakdown for its *species*.



j. 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*. Pretty neat!

Next, let's explore the Survey Counts dataset. Create a new sheet and select the Survey Counts dataset.



- 12. **SA** columns indicate that the surveyor reported the *Species and their Abundance*, while **SO** columns indicate that the surveyor *Only reported the Species* they saw and not the abundance level. **Name** refers to the name of a coral reef that was surveyed. Code is a unique number for each reef site.
  - k. 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.



- Let's assume that experts provide more accurate survey data than novices. Are there any reef sites that appear to 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).
- m. Side note: On average, experts report 80.86 species per survey and novices report 55.64 species this could bias the overall data.
- 13. 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.
  - n. From the top menu, select Analysis >> Create calculated field, and add the equation.



- o. Rename Calculation1 to TotalSurveys.
- p. Create a sheet of the total surveys per reef name using a chart type of your choice, e.g., bar.
- q. Sort the reefs by deceasing numbers of TotalSurveys in the top menu. This menu option can

†8

<u>,</u>

sort ascending, descending, or transpose (flip the rows and columns).

r. 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.

14. 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.

Add Connections     Add      Overail_Bonal_reys_1993_2020     Sheets     P     Good Interpreter     Dea Interpreter     Dea Interpreter     Dea Interpreter     Ban Interpreter     Survey Counts     Since Connts     Since Union     Sheets     Net Union	Survey Counts+ (Overall_Bonaire_surveys_1993_2020)         Survey Counts         Locations	Connections Add Oversit, Borkateys, 1993,2020 Universit faces  P Universit faces P Universit faces P Universit faces University University University University University P New Union	Survey Counts + (Overall_Bonaire_surveys_199	3_2020)			Constitut Live
			Survey Co – Locations 🔹	# Locations	Abc Locations	Abo Locations	÷ Locations
	▲ Survey Co — Locations ▼		Survey Counts Operator Locations	85030036	1000 Steps / Theurand Steps	oull	auli
			Code v = v # Code (Locations	85030087	200m south of Small Wall/R	oull	oull
	<		Add more fields	85030175	Alev's Unnamed Site north of	null	null
	How do relationships differ from joins? Learn more			85030005	Alice In Wonderland (Bonaire)	12.0982	-68.284590
	Survey Counts Operator Locations		> Performance Options	85030173	Andrea 1.5	null	null
	Select a field x = x Select a field x			85030062	Andrea I	null	null
				85030061	Andrea II	null	null
	A select matching fields to create this relationship.			85030026	Angel City	12.1035	-68.287347
	Add more fields			85030114	Aquarius	12.0938	-68.283850

- 15. 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*. In the database world, this is called in inner join, natural join, and equijoin.
- 16. 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. You also may have to convert them between discrete/continuous/measure/dimension.

Survey Co Locations -						None	t	rows 🛞
<ul> <li>How do relationships differ from joins? Learn more</li> </ul>	# Locations Code (Locations)	Abc Locations Name (Locations)	Locat Lat I	✓ Numb Numb Date 8	er (decimal) er (whole) a Time	Area C CBSA City Congr	Code (U.S.) /MSA (U.S.)	
Survey Counts Operator Locations	85030036	1000 Steps / Thousand Steps		Date		Count	ry/Region	
# Code * = * # Code (Locations *	85030087	200m south of Small Wall/R		Spatia	1	Count	y de	
Add more fields	85030175	Alev's Unnamed Site north of		Boolea	n	Longit	tude	
	85030005	Alice In Wonderland (Bonaire)	12	✓ Defaul	t	NUTS	Europe (Province	
> Performance Options	85030173	Andrea 1.5				ZIP Co	de/Postcode	
	85030062	Andrea I	110	Geogr	aphic Role	•		
	85030061	Andrea II	nu	Alf .	null			
	85030026	Angel City	12.10	03528	-68.287347			
	85030114	Aquarius	12.05	93761	-68 283850			
			12.00	55701	00100000			
Survey Co Locations +							None	
Survey Co Locations •  C  C  W do relationships differ from joins? Learn more	# Locations Code (Locations)	Alz Locations Name (Locations)	Decatione Lat Dec	5 5 6 E	O Loost Locat Locat Date & T	(decimal) (whole) lime	None Airport Area Code (L CBSA/MSA / City Congression	J.S.) (U.S.) al District (L
Survey Co — Locations v c ow do relationships differ from joins? Learn more Survey Counts Operator Locations	# Locations Code (Locations) 85030036	Ate Locations Name (Locations) 1000 Steps / Thousand Steps	Decation Lat Dec	5 c E ///	Control → Number Locat Number Date & T Date	(decimal) (whole) ime	None Airport CBSA/MSA i City Congression Country/Reg	J.S.) (U.S.) al District (L
Survey Co Locations •         ow do relationships differ from joins? Learn more        Survey Counts     Operator     Locations       # Code • = • • # Code (Locations •	+ Locations Code (Locations) 85030036 85030087	Atc Locations Name (Locations) 1000 Steps / Thousand Steps 200m south of Small Wall/R	Costione Lat Dec nui nui	s c = //	Content of the second s	(decimal) (whole) ime	None Airport Area Code (L CBSA/MSA i Coty Congression Country/Reg County Latitude	J.S.) (U.S.) al District (U jion
Survey Co Locations • < w do relationships differ from joins? Learn more Survey Counts Operator Locations B Code • = • • • • • • Code (Locations • Add more fields	* Locations Code (Locations) 85030036 85030087 85030175	An- Lootors Name (Lootons) 1000 Steps / Thousand Steps 200m south of Small Wal/R Alev's Unnamed Step orth of	Locations Locations nux nux	s c E d/	Content of the second s	(decimal) (whole) lime	None Airport Area Code (L CBSA/MSA i City Congression County/Reg County Latitude ✓ Longitude	J.S.) (U.S.) al District (L
Survey Co Locations •  w do relationships differ from joins? Learn more  Survey Counts Operator Locations  © Code • = • @ Code (Locations •  Add more fields	tocations     Code (Locations)         85030036         85030087         85030087         85030075         85030075         85030005	Are Lostons Name (Locations) 1000 Steps / Thousand Steps 200m south of Small Wal/R. Alex's Unnamed Site north of. Alex in Worderland (Bonaire)	Lat Dec Lat Dec nui nui 12.09	5 c E /// // // // // // //	Lon Date & T String Spatial Boolean	(decimal) (whole) ime	None Arport Area Code (L CBSA/MAA City Congression Country/Reg Cou	J.S.) (U.S.) al District (L cion
Survey Co Locations • w do relationships differ from joins? Learn more Survey Counts Operator Locations © Odd more fields Performance Options	Code (Locations) 85030036 85030087 85030075 85030005 85030005	Ale Loatons Name (Locations) 1000 Steps / Thousand Steps 200m south of Small Wal/R Alex I fundamed Site north of Alice In Wonderland (Bonaire) Andrea 1.5	Locations Lat Dec nux nux 12.09 nux	s c = // // // // // // //	Lon Date & T Date & T Spatial Boolean -68    Default	(decimal) (whole) ime	None Arport Arsa Code (L CBSA/MSA I Congression Country/Reg Country/Reg Country Latitude V Longitude NUTS Europi State/Prowi ZIP Code/Po	J.S.) (U.S.) al District (L jon e scce sstcode
Survey Co Locations • w do relationships differ from joins? Learn more Survey Counts Operator Locations • Sode • • • • • • • • • Code (Locations • Add more fields Performance Options	© Loatons Code (Locations) 85030067 85030075 85030075 85030073 85030072	Adve Locations Name (Locations) 1000 Steps / Thousand Steps 200m south of Small Wal/R Alex Is Woodertand (Bonaire) Andrea 1.5 Andrea 1	Locations Lat Dec num num 12.09 num	s c E /// // 98241 // //	Control of the second of the	(decimal) (whole) Ime hic Role	None Airport Area Code (U C B3A/MSA City Congression Country/Reg Country Latitude V Longitude NUTS Europs State/Provin ZIP Code/Por	J.S.) JU.S.) al District (L tion stoce stcode
Survey Co Locations • ow do relationships differ from joins? Learn more Survey Counts © Operator © Code © Code	+ Luctions Code (Locations) 8503005 8503005 8503005 8503005 8503005 8503005 8503005	Annea (Locations) Locations Name (Locations) 1000 Steps / Thousand Steps 200m south of Small Wall/R Alice In Wonderland (Bonaire) Andrea 1 Andrea 1 Andrea 1	Lucations Lat Dec nuu nuu 12.09 nuu nuu nuu nuu	s c E d/ d/ d/ d/ b8241 d/ d/ d/ d/ d/ d/ d/ d/ b/ d/ b/ d/ b/ b/ b/ b/ b/ b/ b/ b/ b/ b/ b/ b/ b/	Control of the second	(decimal) (whole) Ime hic Role	None Arport Aracode (L CBSA/MSA City Congression Country/Reg Country/Reg Country/Neg Country/Neg Country/Reg Latitude Latitude V LongItude NUTS Lungo State/Provin	J.S.) (U.S.) al District (U jion a stoce stcode

On Tableau Cloud, the menu may be very slightly different. You must make both fields decimal numbers and also assign them a latitude or longitude geographic role.



17. 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 convert them both to Discrete if the map is not correct. Tableau generates an error, reaggregates the data, and then works fine if you try again.

	III Columns	Lat Dec	<ul> <li>          Locations</li></ul>	Marks		
iii Columns V Lat Dec V	≣ Rows ~	Lon Dec 🔹	Lat Dec     Lon Dec	Add to Sheet		
E Rows Lon Dec	Sheet 9	Show Filter Show Highlighter	# Locations (Count)	Rename Hide	III Columns	Longitude (generated)
Sheet 9	Lon Dec Null Abc	Sort Format ✓ Show Header ✓ Include in Tooltip	<ul> <li>Survey Counts</li> <li>Bottomtime (H:M)</li> <li>Code</li> <li>Name</li> </ul>	Aliases Create  Transform Convert to Continuous	Sheet 9	Lattude (generated)
Lon Dec Null 12.024 12.026 12.	-68.413767 -68.413553	Dimension Attribute ✓ Measure (Average)	=# E-N # Expert SA # Expert SO =# N-E	Convert to Measure Change Data Type Default Properties	Q	2
-68.41399	-68.413499	Discrete     Continuous	# Novice SA # Novice SO	Geographic Role  Group by  Folders	+	•
-68.413767	-68.33586	Add Table Calculation	TotalNumberOfExpert     Survey Counts (Count)	Hierarchy Replace References	12	
-68.413553	-68.328528	Quick Table Calculation  Remove	Abc Measure Names # Measure Values		P	

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



- 19. 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*.
- 20. Where do experts seem to conduct the most surveys?
- 21. 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?

