William Norfolk MADA Course Project-Water Quality

2019-09-11

# Summary/Abstract

*Write a summary of your project.*

# Introduction

## General Background Information

Water quality assessment is a core component of experimental design in a wide range of scientific disciplines. Water condition is imperative to environmental and human health, both as a direct concern such as in the case of aquatic organisms or indirectly such as in the case of irrigation of crops. The utility of water quality data is due to the fact that key parameters can provide information on the baseline health condition of a water system at a relatively low cost in sampling methodology. The efficacy and cost of water sampling has allowed the techniques to become ubiquitous across the field of environmental science and has become a core component of ecosystem health assessment. The relative ease of sampling and low cost of materials is particularly useful for research requiring a large number of samples, such as in ecosystem monitoring.

Ecosystem monitoring is the measurement of changes that occur within an ecosystem over time. Monitoring studies gather data on specific health indicators of interest to a locale and use these data to assess long and short-term changes within an ecosystem. Water quality is one of the most common monitoring indicators due to the high level of ecosystem response to changes in conditions and the accessibility of methods and equipment. Worldwide programs have been established to gather water quality data though the power of citizen science and outreach. Though a vast amount of data has been collected from various studies, much of the analysis done has been conducted at a state or country scale. Numerous data sets exist at smaller scales which can provide useful information on the local microhabitats of various water systems that may be overlooked when assessed at a larger scale. Here we assess the marine water quality conditions of Key Largo, Florida from 2016 to 2019 through the use of citizen science collected data.

Key Largo, Florida is the northernmost island in the Florida Keys archipelago, and the self-proclaimed “Diving Capital of the World.” The waters surrounding Key Largo support three major aquatic habitats: seagrass beds, mangrove forests, and coral reefs. Though distinct in community structure, these three ecosystems exist in delicate balance with one another through the use of water-mediating ecosystem functions. Key Largo is surrounded on all sides by two major bodies of water: the Florida bay which rests on the Gulf of Mexico side of the island, and the Atlantic Ocean. The Florida Bay is a relatively small body of water that extends from the end of mainland Florida and boarders the coast of the Upper Florida Keys. The bayside is a smaller, relatively shallow body, and enclosed body of water with a generally dynamic range of abiotic conditions favorable to seagrass and mangrove habitats. The oceanside boasts a substantially deeper and larger body of water with a relatively stable range of aboitic conditions favorable to coral reef habitats.

#######Add one more paragraph about Largo specific monitoring/this data to transition to the description of data and objectives/hypothesis.###

## Description of data and data source

These data are water quality measurements collected in Key Largo, Florida by the Marine Resources Development Foundation from 2016 to 2019. The Marine Resources Development Foundation is an environmental education non-profit that provides an immersive experience into the field of marine science for students ranging from fourth grade to undergraduates. Marinelab students take a variety of courses to educate them about the local ecosystems, and complement their laboratory and classroom time with daily field trips to the ecosystem of interest. Many courses within the Marinelab curriculum contain integrative data collection programs which task students with the collection of citizen science data on the health of local ecosystems. All data is collected in the field on paper data sheets and is entered into a master raw database by a Marinelab staff members. Specific subsets of citizen science data collected through the programs are passed onto other agencies for further processing based on individual need and interest.

These data are raw water quality data collected from various sampling sites frequented by Marinelab vessels. Water quality data is characterized by 11 distinct variables: date, time, location, instructor name, group name, pH, ammonia, dissolved oxygen, water temperature, salinity, and equipment. The variables: date, time, location, instructor name, group name, and equipment are all clerical data which provides information on the measuring techniques and site characteristics of a sample. The variables: pH, temperature, dissolved oxygen, salinity, and ammoinia are water quality paramaters used to assess the abiotic conditions of the sample site. The Marine Resources Development Foundation has a desire to learn the large-scale patterns of the local water quality to better educate students enrolled in the program. Though this data has been collected for some time, no formal analysis of the data has ever been conducted at a large-scale with the master data.

## Questions/Hypotheses to be addressed

*Question 1: Ocean Verses Bay*

What are the major differences in water quality paramaters between Oceanside and Bayside site locations?

*Question 2: Change Over Time*

Have the water quality conditions of highly visited sites changed over the three years of observation, and can we see impacts of hurricane Irma on the expected conditions?

*Question 3: Seasonal Change*

Are there visible seasonal changes in the abiotic conditions of the water?

*Objective 4: Giving Back to Marinelab*

Develop a script for the immediate processing of data collected by groups actively in the Marinelab program. The goal of this objective is to write a ready-to-use script that will produce scatter plots of the five water quality parameters to compare and contrast oceanside and bayside locations. These figures will then be used to show students the results of their data collection over the course of their time at Marinelab. The script will be tailored to accept a specifically formatted .xlsx file to ensure the data can be run with minimal to no cleaning required. Detailed instructions for data entry into the .xlsx file, and instructions for loading and running the script will be included in a README.md file in the folder.

# Methods and Results

*In most research papers, results and methods are separate. You can combine them here if you find it easier. You are also welcome to structure things such that those are separate sections.*

## Data aquisition

*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*

## Data import and cleaning

*Write code that reads in the file and cleans it so it’s ready for analysis. Since this will be fairly long code for most datasets, it might be a good idea to have it in one or several R scripts. If that is the case, explain here briefly what each file does. The files themselves should be commented well so everyone can follow along.*

## Univariate analysis

*Use a combination of text/tables/figures to explore and describe your data. You should produce plots or tables or other summary quantities for most of your variables. You definitely need to do it for the important variables, i.e. if you have main exposure or outcome variables, those need to be explored. Depending on the total number of variables in your dataset, explore all or some of the others.*

## Bivariate analysis

*Create plots or tables and compute simple statistics (e.g. t-tests, simple regression model with 1 predictor, etc.) to look for associations between your outcome(s) and each individual predictor variable*

## Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You can then load the results produced by this code*

# Discussion

## Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

# References