William Norfolk MADA Course Project-Water Quality

2019-10-10

# 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 abiotic 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 programs, 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 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 by means 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. 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 relatively shallow enclosed body of water with a 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 abiotic conditions favorable to coral reef habitats.

The health of these three major aquatic ecosystems is dependent on the stability of the abiotic conditions of associated waters. Though each of these aquatic habitats are closely associated to the island of Key Largo, chemical and geographical conditions differ greatly between systems and thus support considerably different community structures. Adequate monitoring of these coastal water systems is imperative to the continued environmental and economic health of the island. Citizen science data collection enables researchers to collect information at a substantially larger scale than would be possible alone. The analysis of these data will be used to establish a baseline health condition for various water bodies associated with Key Largo; and will produce data visuals to enrich future citizen science programs.

## 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 ammoina are water quality parameters 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 parameters 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: Promoting Future Citizen Science*

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

All data was collected by students or instructors of the Marinelab environmental education program. All collectors are required to complete a one-hour training program to familiarize themselves with the testing materials and data recording before they are permitted to work in the field. Data is collected in the field at various sampling locations dictated by the needs of the program, accessibility of location, and weather conditions. On site, a grab sample of water is collected and brought onto the vessel for testing. Abiotic conditions are measured immediately following collection and scored onto a paper record sheet. All measures are taken using semiquantative testing equipment or with a YSI Sonde. Temperature is measured in Celsius using a standard thermometer attached to a string to ensure body heat does not skew the data. Salinity is measured with a standard hydrometer in ppt. Dissolved Oxygen is measured using a colormetric ampoule analysis in mg/L. Ammonia and pH are both measured using standardized test kits contiaing test strips and reagent solution respectively. All data sheets are collected and entered into the master database upon return to shore.

Data will be analyzed using R Studio software and the following packages: #####LIST ALL PACKAGES######. Univariate analysis will be used to produce a profile for each of the major sampling sites frequented throughout the study. This profile will detail the average value and range of the five abiotic conditions of interest to be compared in bivariate analysis. Bivariate analysis will be used to compare different water quality parameters between sample sites. In particular, comparisons will be used to establish the difference in the abiotic conditions of oceanside and bayside sampling locations.

## Data Acquisition

These data were acquired from the Marinelab citizen science master database. The database is privately maintained by the Marine Resources Development Foundation and is not readily available online. Data is primarily used as a tool to enrich scientific education in the Marinelab program; however data may be distributed for analysis at the discretion of the director of the program. This analysis was conducted under the permission of the current Marinelab director Sarah Egner, and the resulting analysis will be used for the enrichment of future programs.

## Data import and cleaning

A detailed description of data import, cleaning, and exploratory analysis can be found in the supplementary files as follows. Data import and cleaning is located in folder titled code, the subfolder processing code, and the file WQprocessing. Exploratory analysis is located in the folder code, subfolder analysis code, and is split into two files Exploratory\_Data\_Analysis\_Location which explores variables island\_side and site\_type and Exploratory\_Data\_Analysis\_Seasonal which explores data seasonality and changes across time. All processing and anlysis files are available in both Rmd and docx formats.

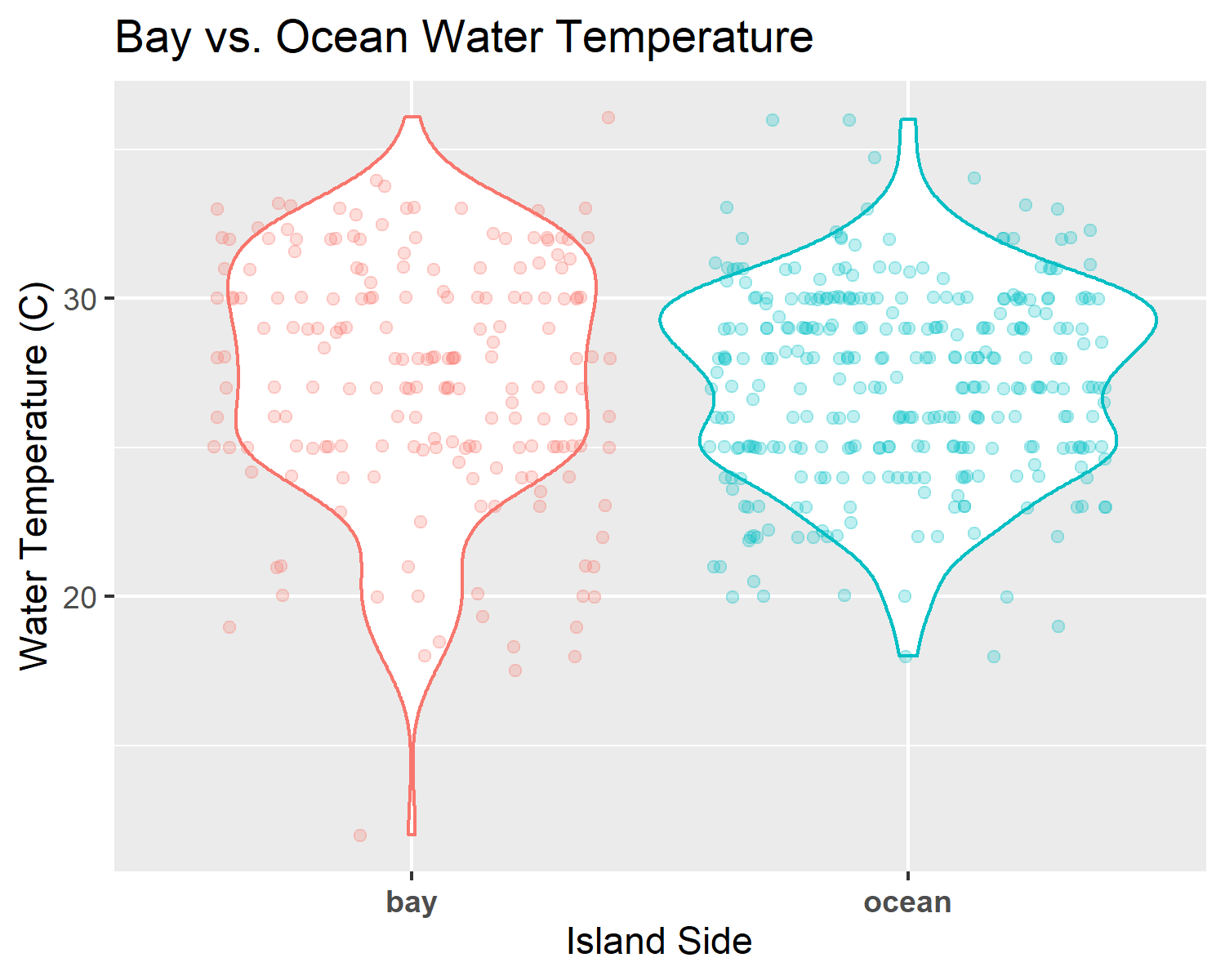
## Univariate analysis

Exploratory analysis of the cleaned water quality dataset shows general trends across all parameters measured and identifies specific aspects of the code that warrent deeper analysis.

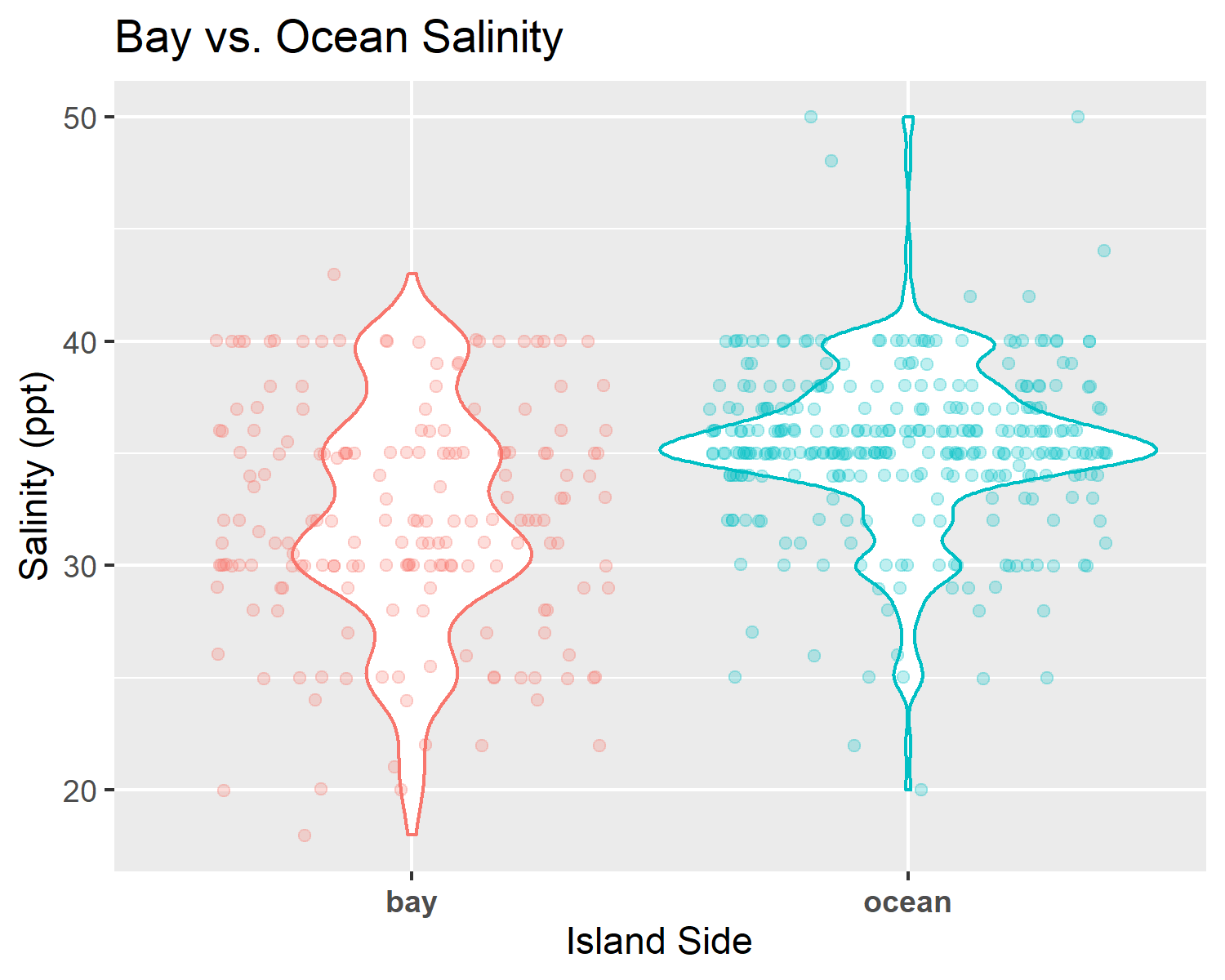
The first variables explored will be collectively refrred to as the location variables, as they provide site specific information of the sampling locations that is understandable to inividuals that are not familar with the specific sites by name. Loaction variables explored consisted of island\_side and site\_type. Island\_side refers to the specific side of the island the sampling site is located and consists of two categories: oceanside (Atlantic Ocean) and bayside (Gulf of Mexico). Site\_type defines the specific type of ecosystem the site represents and consists of three categories: coral reef, seagrass/mangroves, and path reef/hardbottom. The thee site categories are distinguished by distance from shore, seagrass/mangrove represent near shore sites, path reef/hardbottom are mid-shore, and coral reefs are offshore locations.

Notable location patterns were observed in both salinity and temperature parameters for both island\_side and site\_type. Oceanside sites typically showed a reduced range of temperature and salinity measures, whereas bayside sites were distinctly dynamic. In terms of site\_type, seagrass/mangrove sites generally exhibited a wider range of temperature and salinity, whereas coral reefs and patch reefs/hardbottom sites were less variable. The following trends cna be seen in figures 1-4 below.

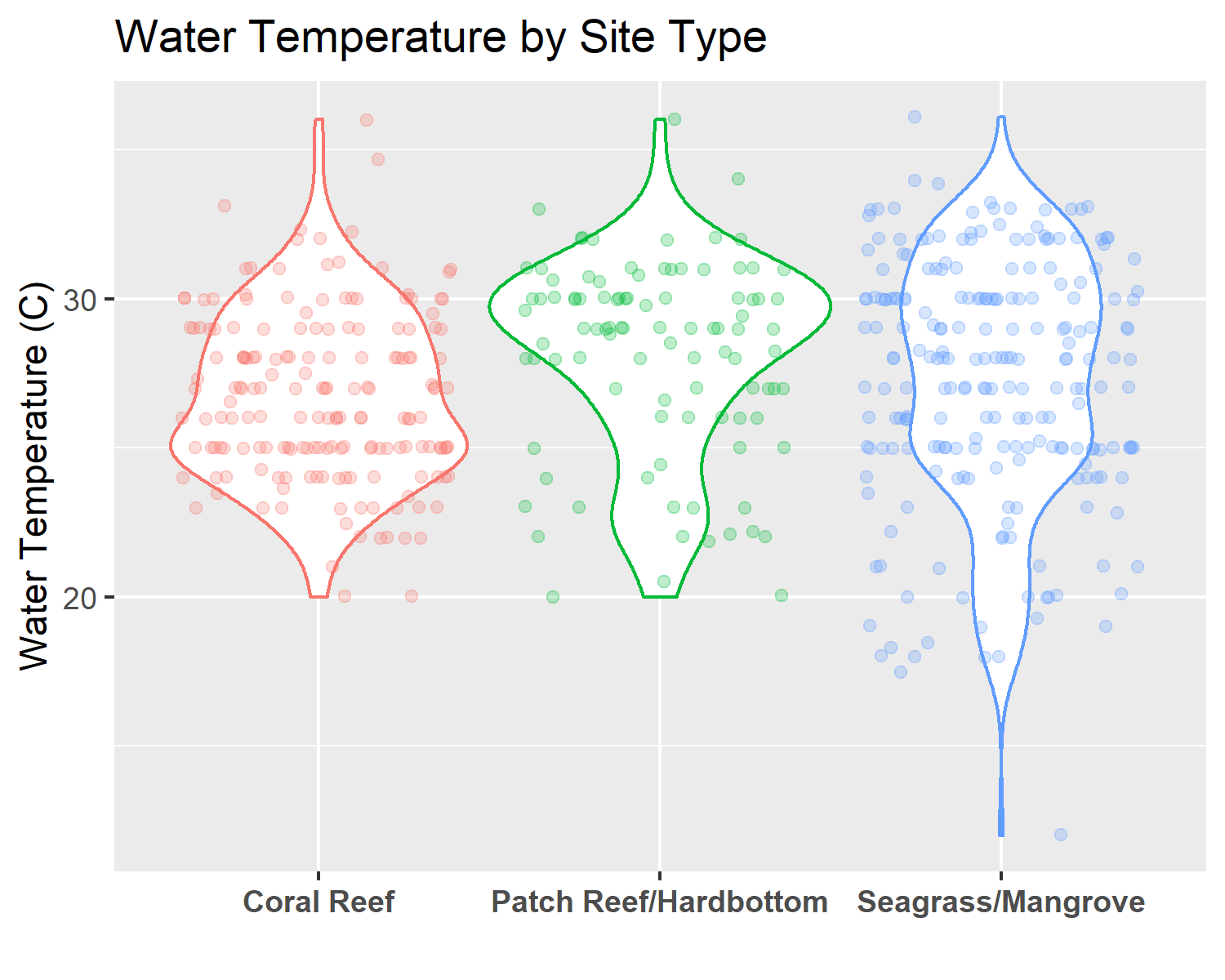
knitr::include\_graphics("../../results/bay\_v\_ocean\_water\_temp.png")



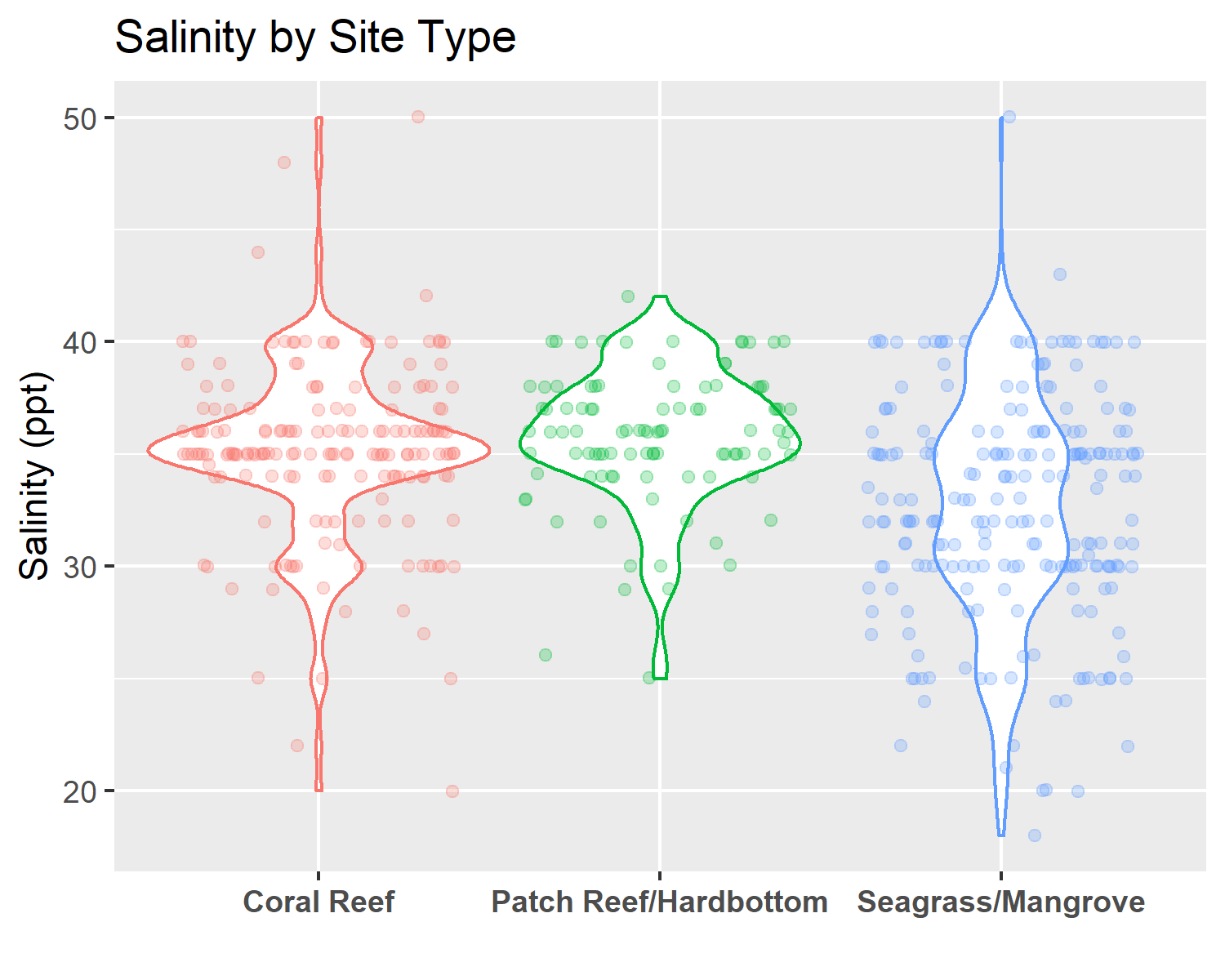
knitr::include\_graphics("../../results/bay\_v\_ocean\_salinity.png")



knitr::include\_graphics("../../results/site\_type\_water\_temp.png")



knitr::include\_graphics("../../results/site\_type\_salinity.png")



## 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*

##Results

## 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