Nathan Greenslit Data Analysis Project

Andreas Handel

2/9/23

This uses MS Word as output format. [See here](https://quarto.org/docs/output-formats/ms-word.html) for more information. You can switch to other formats, like html or pdf. See [the Quarto documentation](https://quarto.org/) for other formats.

# 1. Summary/Abstract

*Write a summary of your project.*

# 2. Introduction

## 2.1 General Background Information

Each year, plumes of Saharan dust travel across the Atlantic via easterly trade winds. These plumes are then deposited in the surface waters of the mid-Atlantic, Caribbean, and Gulf of Mexico. These dust aerosols harbor a wide range of fungi, bacteria, virus-like-particles, minerals, and nutrients (NO3, Fe, PO4). Addition of these nutrients to otherwise oligotrophic settings can result in large and rapid blooms of potentially harmful microbes, presenting a danger to both marine and human health.

Bacteria belonging to the genus Vibrio are marine opportunistic heterotrophs that are ubiquitous in nature and are among the first to respond to the influx of nutrients. Species of Vibrio are known to cause a variety of diseases in humans ranging from cholera (V. cholerae), to ear and wound infections (V. alginolyticus), and in some extreme cases necrotizing fasciitis and septicemia (V. vulnificus). Vibrio have also been shown to impact marine health, causing coral bleaching (V. coralliiyticus, V. shiloi, V. alginolyticus), and shellfish and fish mortality (V. harveyi, V. parahaemolyticus, V. vulnificus). Vibrio population dynamics are primarily driven by temperature and salinity, with prime conditions between 25-30°C and salinities ranging from 20-35. While there is evidence of these blooms occurring in low-nutrient settings like the Florida Keys, less is known regarding microbial response to dust input in areas with *higher baseline nutrients*.

## 2.2 Description of data and data source

Data was collected at 3 sites in Corpus Christi, TX before, during, and after a Saharan dust event (starting on July 7th and ending on July 19th, 2022). Sites represent a gradient in background nutrient levels (Blind Oso and Canals are high, the Gulf is low). Water samples were collected for inorganic nutrients, dissolved and particulate organic matter, and microbial analysis. Vertical profiles of salinity, temperature, pH, and dissolved oxygen were obtained using a YSI ProPlus sonde.

Variables will include Temperature (°C), Salinity, Nitrate, Phosphate, Copies/mL produced using quantitative PCR. Provides gene copies of Vibrio bacteria (I am hoping to also get data about specific species), and Dust Concentration (AOT) from the naval research laboratory satellite data.

## 2.3 Questions/Hypotheses to be addressed

How do Saharan dust events influence Vibrio populations in high nutrient coastal waters?

*What I am looking for* I expect to see a more dramatic growth response in the Gulf (low nutrient), whereas the higher nutrient sites may exhibit a dampened growth response since they already have high baseline nutrient levels. If we still see a growth response despite already having supportive background nutrient levels at these sites, this may suggest that there are other constituents in the dust that can elicit a growth response. Nutrients, dust input, salinity, and temperature will be the primary factors of interest.

*How I will analyze it* - Look at Vibrio growth over time series (estimated from qPCR) - Examine relationships between dust input and Vibrio growth - Examine influence of site-specific environmental parameters (Temperature, Salinity, Nutrients) on growth response - Run linear models on factors like: dust x growth and nutrients x growth - NMDS to see which parameters have the strongest influence on growth

To cite other work (important everywhere, but likely happens first in introduction), make sure your references are in the bibtex file specified in the YAML header above (here dataanalysis\_template\_references.bib) and have the right bibtex key. Then you can include like this:

# 3. Methods

*Describe your methods. That should describe the data, the cleaning processes, and the analysis approaches. You might want to provide a shorter description here and all the details in the supplement.*

## 3.1 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.*

## 3.2 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 kind of cleaning/processing you do, and provide more details and well documented code somewhere (e.g. as supplement in a paper). All materials, including files that contain code, should be commented well so everyone can follow along.*

## 3.3 Statistical analysis

*Explain anything related to your statistical analyses.*

# 4. Results

## 4.1 Exploratory/Descriptive analysis

*Use a combination of text/tables/figures to explore and describe your data. Show the most important descriptive results here. Additional ones should go in the supplement. Even more can be in the R and Quarto files that are part of your project.*

**?@tbl-summarytable** shows a summary of the data.

Note the loading of the data providing a **relative** path using the ../../ notation. (Two dots means a folder up). You never want to specify an **absolute** path like C:\ahandel\myproject\results\ because if you share this with someone, it won’t work for them since they don’t have that path. You can also use the here R package to create paths. See examples of that below.

**?(caption)**

## 4.2 Basic statistical analysis

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. simple models with 1 predictor) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

**?@fig-result** shows a scatterplot figure produced by one of the R scripts.

## 4.3 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 want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

Example **?@tbl-resulttable2** shows a summary of a linear model fit.

**?(caption)**

# 5. Discussion

## 5.1 Summary and Interpretation

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

## 5.2 Strengths and Limitations

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

## 5.3 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*

This paper (Leek & Peng, 2015) discusses types of analyses.

These papers (McKay, Ebell, Billings, et al., 2020; McKay, Ebell, Dale, Shen, & Handel, 2020) are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# 6. References

Leek, J. T., & Peng, R. D. (2015). Statistics. What is the question? *Science (New York, N.Y.)*, *347*(6228), 1314–1315. <https://doi.org/10.1126/science.aaa6146>

McKay, B., Ebell, M., Billings, W. Z., Dale, A. P., Shen, Y., & Handel, A. (2020). Associations Between Relative Viral Load at Diagnosis and Influenza A Symptoms and Recovery. *Open Forum Infectious Diseases*, *7*(11), ofaa494. <https://doi.org/10.1093/ofid/ofaa494>

McKay, B., Ebell, M., Dale, A. P., Shen, Y., & Handel, A. (2020). Virulence-mediated infectiousness and activity trade-offs and their impact on transmission potential of influenza patients. *Proceedings. Biological Sciences*, *287*(1927), 20200496. <https://doi.org/10.1098/rspb.2020.0496>