Exploratory Data Analysis-Seasonal

William Norfolk

10/10/2019

Now that we have dissected the data by island side and site type we will look into seasonal differences.

Load the required libraries for exploratory analysis.

library(readxl)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyverse)

## -- Attaching packages ------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v readr 1.3.1  
## v tibble 2.1.3 v purrr 0.3.2  
## v tidyr 0.8.3 v stringr 1.4.0  
## v ggplot2 3.2.1 v forcats 0.4.0

## -- Conflicts ---------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(forcats)  
library(ggthemes)  
library(plotly)

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

library(knitr)  
library(naniar)  
library(broom)  
library(gridExtra)

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library(zoo)

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

Load the processed data from the RDS. Then take a look!

WQ\_clean\_data <- readRDS("../../data/processed\_data/processeddata.rds")  
  
glimpse(WQ\_clean\_data)

## Observations: 522  
## Variables: 15  
## $ Month <chr> "01", "02", "02", "02", "02", "02", "02", "02...  
## $ Day <chr> "08", "08", "08", "08", "09", "09", "09", "12...  
## $ Year <chr> "16", "16", "16", "16", "16", "16", "16", "16...  
## $ military\_time <dbl> 1415, 1515, 1550, 1555, 1001, 1015, 1022, 103...  
## $ location <chr> "Boat Ramp", "Grecian Dry Rocks", "Grecian Dr...  
## $ instructor\_name <chr> "Katy, Sarah, Driver", "Chelsea", "Katy, Tomm...  
## $ group\_name <chr> "NA", "McLean High School", "McLean High Scho...  
## $ ph <dbl> 8.0, 8.4, 8.2, 8.4, 8.0, 8.0, 8.0, 8.0, 8.4, ...  
## $ ammonia <dbl> 0.00, 0.00, 0.00, 0.00, 0.25, 0.00, 0.00, 0.0...  
## $ dissolved\_oxygen <dbl> 5.0, 4.0, 4.0, 6.0, 8.0, 4.0, 5.0, 6.0, 6.0, ...  
## $ water\_temp <dbl> NA, 23.5, 21.0, 36.0, 18.0, 18.0, 18.0, 18.3,...  
## $ salinity <dbl> 36, 40, 44, 35, 33, 30, 33, 35, 40, 30, 35, 2...  
## $ equipment <chr> "kit", "kit", "kit", "kit", "kit", "kit", "ki...  
## $ island\_side <chr> "ocean", "ocean", "ocean", "ocean", NA, "bay"...  
## $ site\_type <chr> "Seagrass/Mangrove", "Coral Reef", "Coral Ree...

island\_side\_filter <- filter(WQ\_clean\_data, !is.na(island\_side))  
  
site\_type\_filter <- filter(WQ\_clean\_data, !is.na(site\_type))

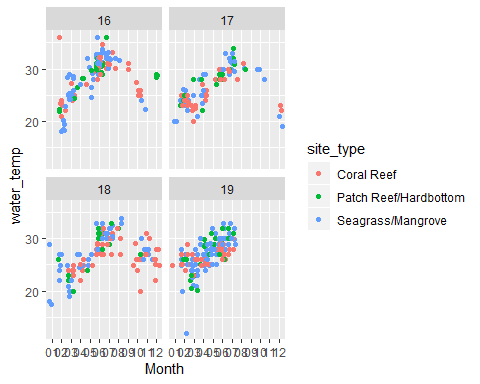
We will reload our filters for ease of access.

Let’s start with some monthly plots of each parameter across the four years of the study. We will indicate site type by color.

Looking at temperature it appears that standard seasonal fluctions are at play. It is also clear that MarineLab is closed from mid-August to September 1st due to the lack of observations during this time frame.

site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = water\_temp, color = site\_type)) + facet\_wrap(~Year)

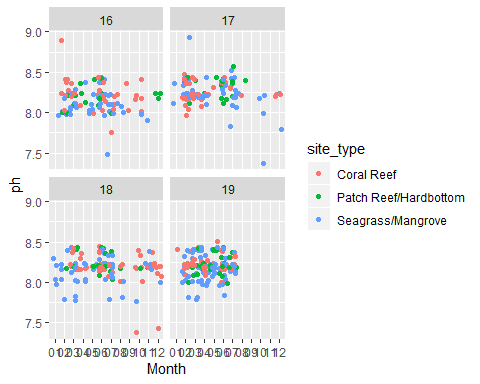
## Warning: Removed 12 rows containing missing values (geom\_point).



pH looks reasonably consistant across all sites and years of measurement.

site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = ph, color = site\_type)) + facet\_wrap(~Year)

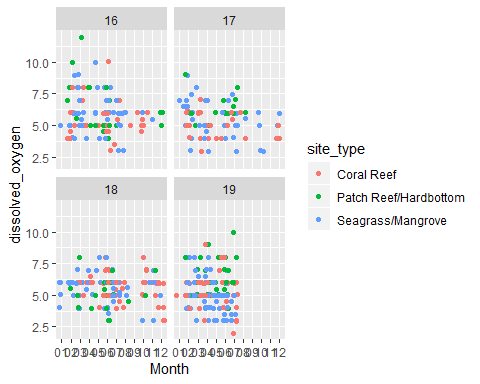
## Warning: Removed 6 rows containing missing values (geom\_point).



Dissolved oxygen shows a decently wide spread across all four years. Deeper analysis may reveal further seasonal details.

site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = dissolved\_oxygen, color = site\_type)) + facet\_wrap(~Year)

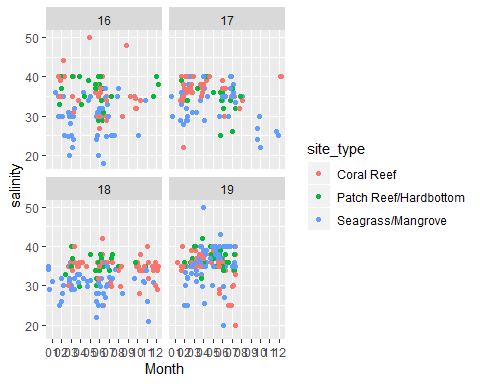
## Warning: Removed 4 rows containing missing values (geom\_point).



Salinity is reasonably consistent across all years. There is a notable drop in late 2017 (around Hurricane Irma timeline) which may be interesting to look at further. It should be noted that there are fewer observations from that time frame due to the inability to sample sites via boat until water were safe.

site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = salinity, color = site\_type)) + facet\_wrap(~Year)

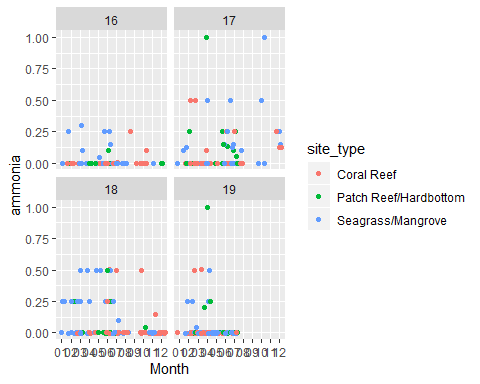
## Warning: Removed 15 rows containing missing values (geom\_point).



Lastly, ammonia looks reasonably consistent as well across the years.

site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = ammonia, color = site\_type)) + facet\_wrap(~Year)

## Warning: Removed 10 rows containing missing values (geom\_point).



Next let’s overlay some geom\_smooth plots with jitter plots to see general trends. We will need the Month variable to be numeric to properly plot.

WQ\_clean\_data$Month <- as.numeric(as.character(WQ\_clean\_data$Month))

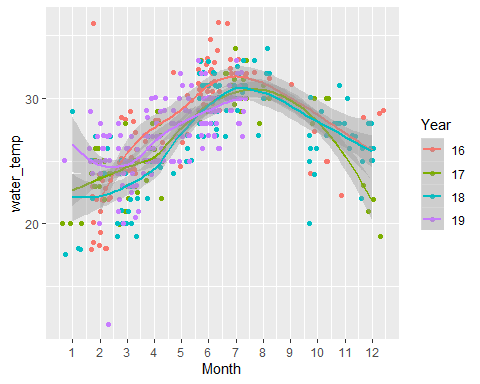
As expected, temperature appears fairly seasonal with cooler water temp in the winter and warmer in the summer months. It looks like 2016 was the warmest year, which is consistent with previous work and coral bleaching events in the Keys.

WQ\_clean\_data %>% ggplot(aes(x = Month, y = water\_temp, color = Year)) + geom\_jitter() + geom\_smooth() + scale\_x\_continuous(breaks = c(1:12))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 12 rows containing non-finite values (stat\_smooth).

## Warning: Removed 12 rows containing missing values (geom\_point).



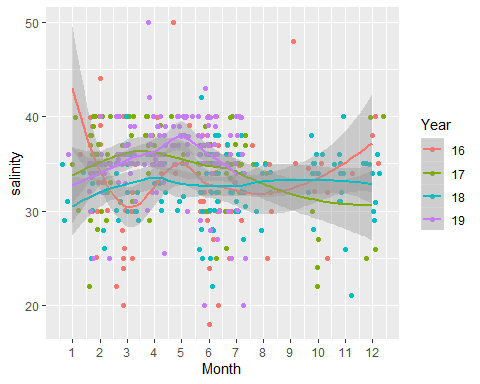
Salinity shows a bit of swing occasionally, but seems to rest pretty consistently at approximately 35 ppt. Though a bit low in late 2017 and early 2018.

WQ\_clean\_data %>% ggplot(aes(x = Month, y = salinity, color = Year)) + geom\_jitter() + geom\_smooth() + scale\_x\_continuous(breaks = c(1:12))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 16 rows containing non-finite values (stat\_smooth).

## Warning: Removed 16 rows containing missing values (geom\_point).



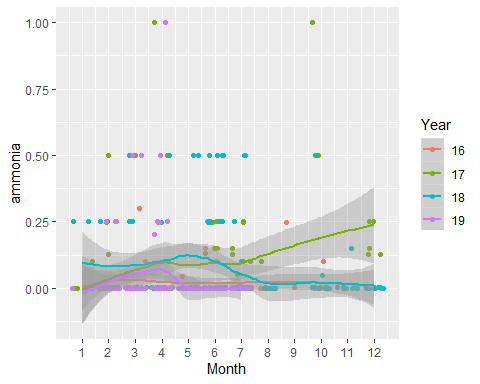
Ammonia is pretty consistent across all years except 2017. 2017 spikes upward toward the end of the year which may be attributable to Hurricane Irma nutient influx however, it appears the spike begins earlier (July) while the hurricane was in September.

WQ\_clean\_data %>% ggplot(aes(x = Month, y = ammonia, color = Year)) + geom\_jitter() + geom\_smooth() + scale\_x\_continuous(breaks = c(1:12))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 10 rows containing non-finite values (stat\_smooth).

## Warning: Removed 10 rows containing missing values (geom\_point).



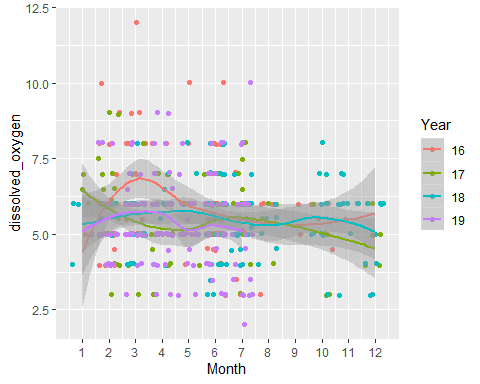
LAstly, dissolved oxygen reamins consistent across the years. There is a curious bump in early 2016.

WQ\_clean\_data %>% ggplot(aes(x = Month, y = dissolved\_oxygen, color = Year)) + geom\_jitter() + geom\_smooth() + scale\_x\_continuous(breaks = c(1:12))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 4 rows containing non-finite values (stat\_smooth).

## Warning: Removed 4 rows containing missing values (geom\_point).

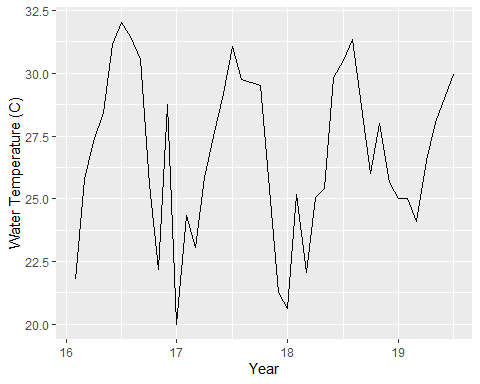


For the last part of exploratory analysis, we will prepares some plots of the entire time scale covered in the study. This is to view seasonal fluctuation in each parameter, and search for Hurricane Irma impacts. To view, we must first take the average of each month data was recorded then plot the timescale agains the parameter.

Temperature measures across the four years show a consistent seasonal fluctuation.

mean\_water\_temp <- aggregate(water\_temp ~ Month + Year, WQ\_clean\_data, mean)   
  
  
  
mean\_water\_temp$Date <- as.yearmon(paste(mean\_water\_temp$Month, mean\_water\_temp$Year, sep = "."), format = "%m.%Y")  
  
mean\_water\_temp\_plot <- mean\_water\_temp %>% ggplot(aes(x = Date, y = water\_temp)) + geom\_line() + xlab("Year") + ylab("Water Temperature (C)")  
  
mean\_water\_temp\_plot

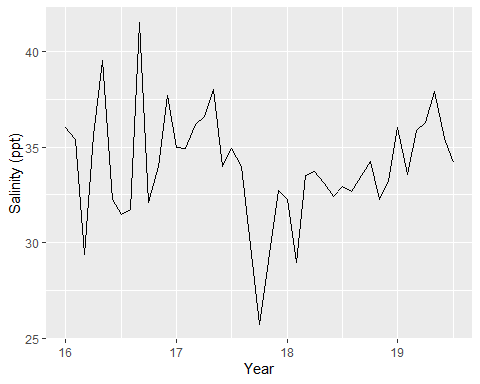
## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.



Salinity shows moderate seasonality likely strongly influenced by bayside sites. There is a distinct drop in salinity around Hurricane Irma timeframe which may be attributable to the large-scale mixing of brackish and salt water, as well as intense rainfall during and after the storm event. Caution should be taken still with the reduced observations from that time frame.

mean\_salinity <- aggregate(salinity ~ Month + Year, WQ\_clean\_data, mean)   
  
mean\_salinity$Date <- as.yearmon(paste(mean\_salinity$Month, mean\_salinity$Year, sep = "."), format = "%m.%Y")  
  
mean\_salinity\_plot <- mean\_salinity %>% ggplot(aes(x = Date, y = salinity)) + geom\_line() + xlab("Year") + ylab("Salinity (ppt)")  
  
mean\_salinity\_plot

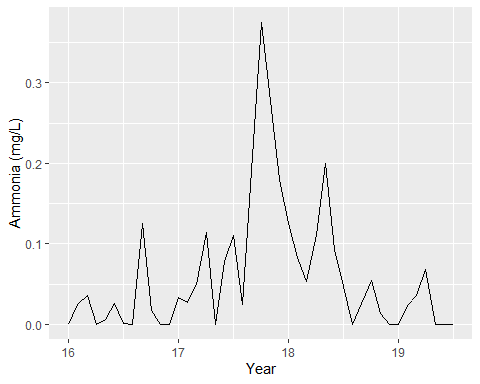
## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.



Ammonia remains fairly low on average with the exception of a major spike in late 2017. This spike is very likely a result of hurricane-induced nutrient influx during the storm and afterwards.

mean\_ammonia <- aggregate(ammonia ~ Month + Year, WQ\_clean\_data, mean)   
  
mean\_ammonia$Date <- as.yearmon(paste(mean\_ammonia$Month, mean\_ammonia$Year, sep = "."), format = "%m.%Y")  
  
mean\_ammonia\_plot <- mean\_ammonia %>% ggplot(aes(x = Date, y = ammonia)) + geom\_line() + xlab("Year") + ylab("Ammonia (mg/L)")  
  
mean\_ammonia\_plot

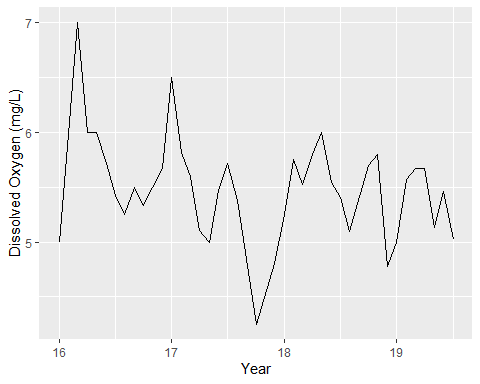
## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.



Dissolved oxygen shows a relatively seasonal trend with the addition of a sharp plunge in concentration in late September 2017. It is probable this is an additional effect of Hurricane Irma damage.

mean\_dissolved\_oxygen <- aggregate(dissolved\_oxygen ~ Month + Year, WQ\_clean\_data, mean)   
  
mean\_dissolved\_oxygen$Date <- as.yearmon(paste(mean\_dissolved\_oxygen$Month, mean\_dissolved\_oxygen$Year, sep = "."), format = "%m.%Y")  
  
mean\_dissolved\_oxygen\_plot <- mean\_dissolved\_oxygen %>% ggplot(aes(x = Date, y = dissolved\_oxygen)) + geom\_line() + xlab("Year") + ylab("Dissolved Oxygen (mg/L)")  
  
mean\_dissolved\_oxygen\_plot

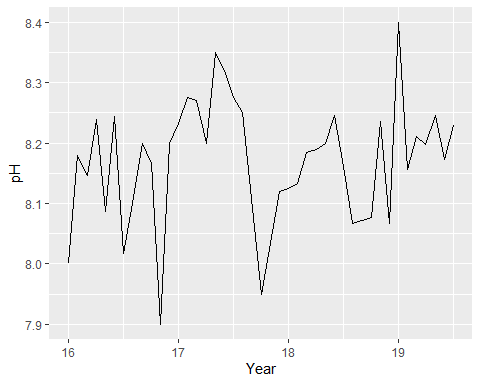
## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.



pH shows a standard range with no directly apparent seasonality. There is a visible dip in levels around the time of Hurrican Irma, however this drop does not appear to be distinctly different from previously measured levels of pH in the Keys (note 2016). It is interesting to see the general upward trend in pH which may possibly be attributable to ocean acidification. Though not part of our original questions, this may be an interesting analysis.

mean\_ph <- aggregate(ph ~ Month + Year, WQ\_clean\_data, mean)   
  
mean\_ph$Date <- as.yearmon(paste(mean\_ph$Month, mean\_ph$Year, sep = "."), format = "%m.%Y")  
  
mean\_ph\_plot <- mean\_ph %>% ggplot(aes(x = Date, y = ph)) + geom\_line() + xlab("Year") + ylab("pH")  
  
mean\_ph\_plot

## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.



grid.arrange(mean\_ammonia\_plot, mean\_ph\_plot, mean\_salinity\_plot, mean\_dissolved\_oxygen\_plot, mean\_water\_temp\_plot, nrow = 2)

## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.  
## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.  
## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.  
## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.  
## Don't know how to automatically pick scale for object of type yearmon. Defaulting to continuous.

