Exploratory Data Analysis

William Norfolk

10/9/2019

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)

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...

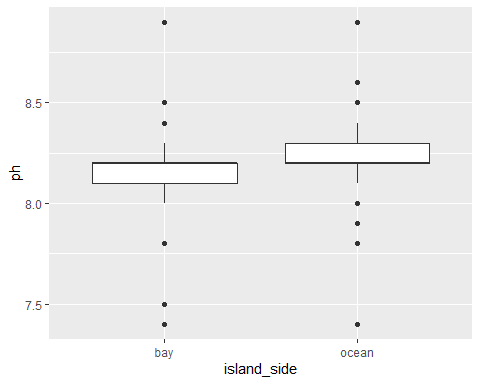
island\_side\_filter <- filter(WQ\_clean\_data, !is.na(island\_side))  
  
site\_type\_filter <- filter(WQ\_clean\_data, !is.na(site\_type))

Now lets generate some plots to compare ocean vs bay.

Looks like pH is pretty consistent between the ocean and bayside.

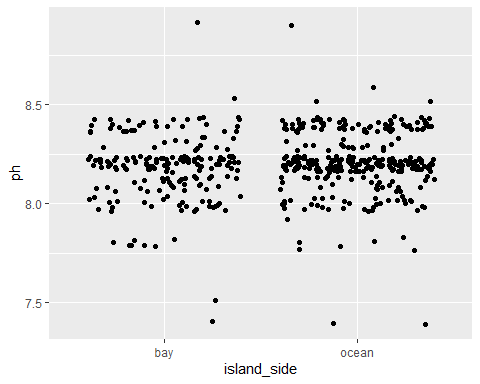
island\_side\_filter %>% ggplot() +   
 geom\_boxplot(aes(x = island\_side, y = ph))

## Warning: Removed 6 rows containing non-finite values (stat\_boxplot).



island\_side\_filter %>% ggplot() +   
 geom\_jitter(aes(x = island\_side, y = ph))

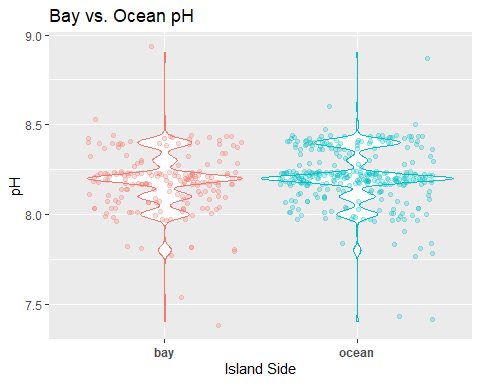
## Warning: Removed 6 rows containing missing values (geom\_point).



bay\_v\_ocean\_ph <- ggplot(island\_side\_filter, aes(x = island\_side, y = ph, color = island\_side)) + geom\_violin(width = 1) + geom\_jitter(alpha = 0.25) + ggtitle("Bay vs. Ocean pH") + xlab("Island Side") + ylab("pH") + theme(legend.position = "none", axis.text.x = element\_text(face = "bold"))  
  
bay\_v\_ocean\_ph

## Warning: Removed 6 rows containing non-finite values (stat\_ydensity).

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



ggsave(filename = "../../results/bay\_v\_ocean\_ph.png",plot = bay\_v\_ocean\_ph)

## Saving 5 x 4 in image

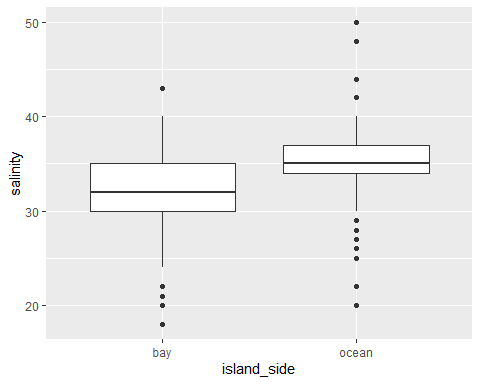
## Warning: Removed 6 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 6 rows containing missing values (geom\_point).

We will save the violin + jitter plots for use later.

Looks like salinity may be a little more interesting. Ocean side appears to have a narrower range of values compared to the bayside, which is expected due to the size of the body of water.

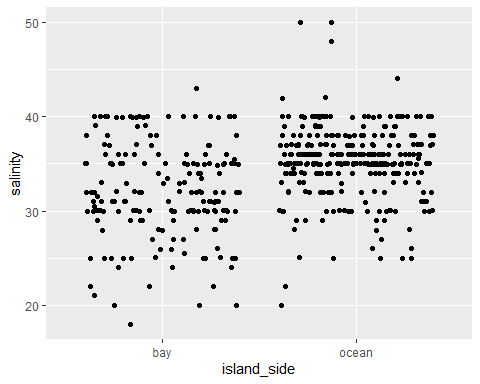
island\_side\_filter %>% ggplot() +   
 geom\_boxplot(aes(x = island\_side, y = salinity))

## Warning: Removed 15 rows containing non-finite values (stat\_boxplot).



island\_side\_filter %>% ggplot() +   
 geom\_jitter(aes(x = island\_side, y = salinity))

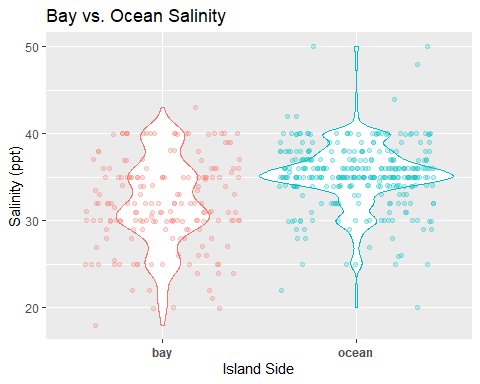
## Warning: Removed 15 rows containing missing values (geom\_point).



bay\_v\_ocean\_salinity <- ggplot(island\_side\_filter, aes(x = island\_side, y = salinity, color = island\_side)) + geom\_violin(width = 1) + geom\_jitter(alpha = 0.25) + ggtitle("Bay vs. Ocean Salinity") + xlab("Island Side") + ylab("Salinity (ppt)") + theme(legend.position = "none", axis.text.x = element\_text(face = "bold"))  
  
bay\_v\_ocean\_salinity

## Warning: Removed 15 rows containing non-finite values (stat\_ydensity).

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



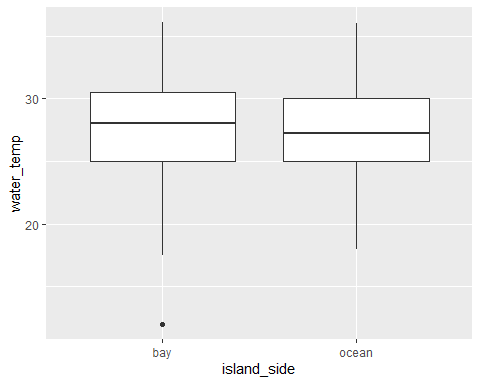
ggsave(filename = "../../results/bay\_v\_ocean\_salinity.png",plot = bay\_v\_ocean\_salinity)

## Saving 5 x 4 in image

## Warning: Removed 15 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 15 rows containing missing values (geom\_point).

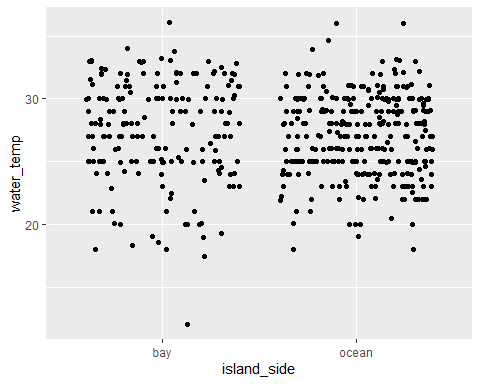
island\_side\_filter %>% ggplot() +   
 geom\_boxplot(aes(x = island\_side, y = water\_temp))

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



island\_side\_filter %>% ggplot() +   
 geom\_jitter(aes(x = island\_side, y = water\_temp))

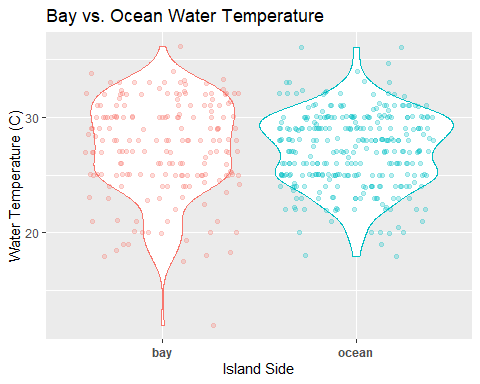
## Warning: Removed 12 rows containing missing values (geom\_point).



bay\_v\_ocean\_water\_temp <- ggplot(island\_side\_filter, aes(x = island\_side, y = water\_temp, color = island\_side)) + geom\_violin(width = 1) + geom\_jitter(alpha = 0.25) + ggtitle("Bay vs. Ocean Water Temperature") + xlab("Island Side") + ylab("Water Temperature (C)") + theme(legend.position = "none", axis.text.x = element\_text(face = "bold"))  
  
bay\_v\_ocean\_water\_temp

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

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



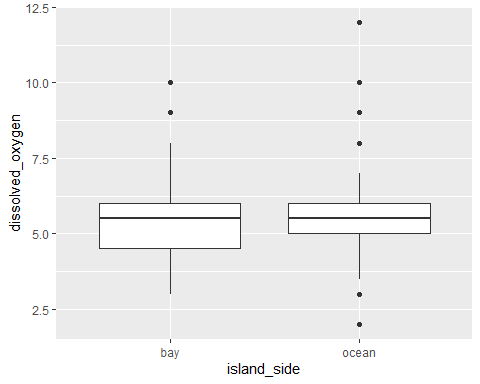
ggsave(filename = "../../results/bay\_v\_ocean\_water\_temp.png",plot = bay\_v\_ocean\_water\_temp)

## Saving 5 x 4 in image

## Warning: Removed 12 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 12 rows containing missing values (geom\_point).

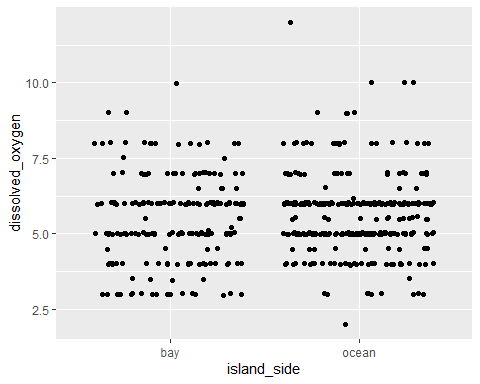
island\_side\_filter %>% ggplot() +   
 geom\_boxplot(aes(x = island\_side, y = dissolved\_oxygen))

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



island\_side\_filter %>% ggplot() +   
 geom\_jitter(aes(x = island\_side, y = dissolved\_oxygen))

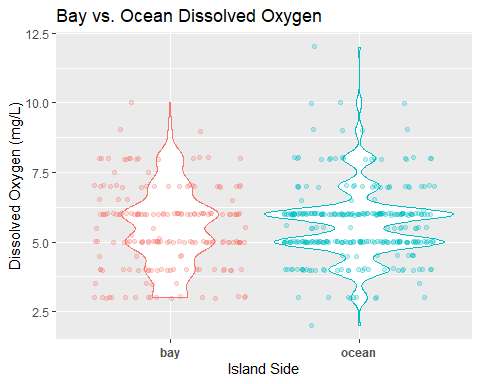
## Warning: Removed 4 rows containing missing values (geom\_point).



bay\_v\_ocean\_dissolved\_oxygen <- ggplot(island\_side\_filter, aes(x = island\_side, y = dissolved\_oxygen, color = island\_side)) + geom\_violin(width = 1) + geom\_jitter(alpha = 0.25) + ggtitle("Bay vs. Ocean Dissolved Oxygen") + xlab("Island Side") + ylab("Dissolved Oxygen (mg/L)") + theme(legend.position = "none", axis.text.x = element\_text(face = "bold"))  
  
bay\_v\_ocean\_dissolved\_oxygen

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

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



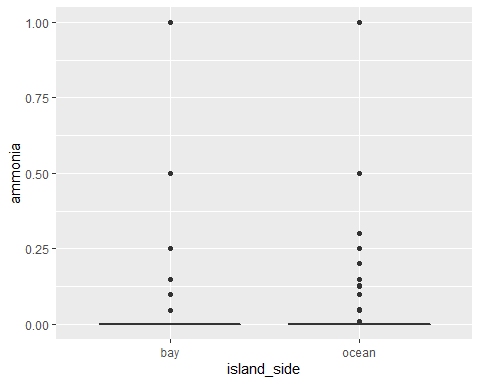
ggsave(filename = "../../results/bay\_v\_ocean\_dissolved\_oxygen.png",plot = bay\_v\_ocean\_dissolved\_oxygen)

## Saving 5 x 4 in image

## Warning: Removed 4 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 4 rows containing missing values (geom\_point).

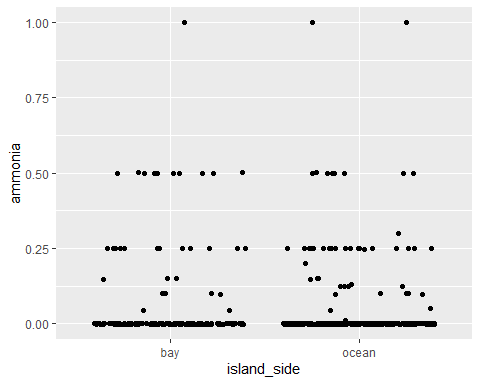
filter(WQ\_clean\_data, !is.na(island\_side)) %>% ggplot() +   
 geom\_boxplot(aes(x = island\_side, y = ammonia))

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



filter(WQ\_clean\_data, !is.na(island\_side)) %>% ggplot() +   
 geom\_jitter(aes(x = island\_side, y = ammonia))

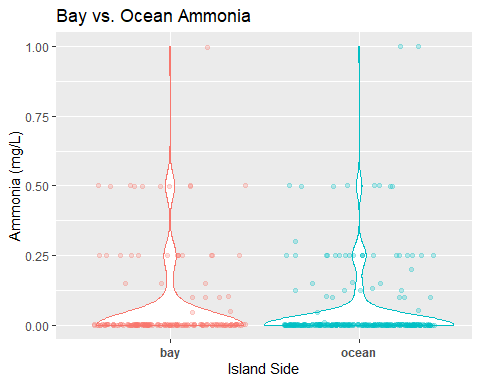
## Warning: Removed 10 rows containing missing values (geom\_point).



bay\_v\_ocean\_ammonia <- ggplot(island\_side\_filter, aes(x = island\_side, y = ammonia, color = island\_side)) + geom\_violin(width = 1) + geom\_jitter(alpha = 0.25) + ggtitle("Bay vs. Ocean Ammonia") + xlab("Island Side") + ylab("Ammonia (mg/L)") + theme(legend.position = "none", axis.text.x = element\_text(face = "bold"))  
  
bay\_v\_ocean\_ammonia

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

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



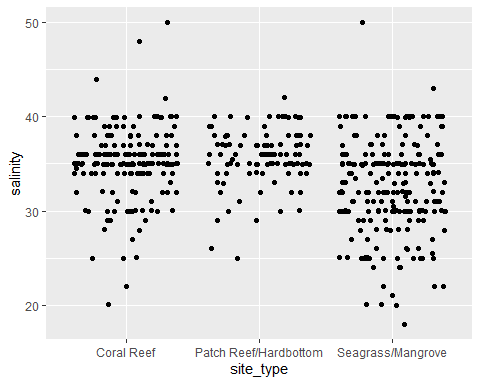
ggsave(filename = "../../results/bay\_v\_ocean\_ammonia.png",plot = bay\_v\_ocean\_ammonia)

## Saving 5 x 4 in image

## Warning: Removed 10 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 10 rows containing missing values (geom\_point).

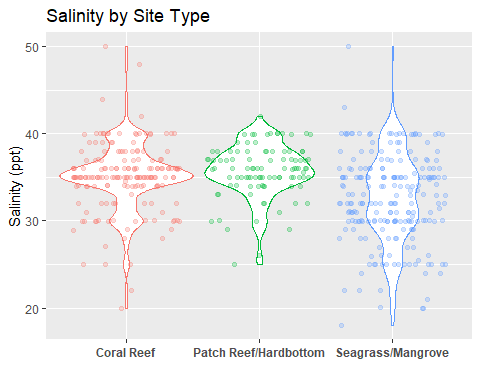
site\_type\_filter %>% ggplot() +   
 geom\_jitter(aes(x = site\_type, y = salinity))

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



site\_type\_salinity <- ggplot(island\_side\_filter, aes(x = site\_type, y = salinity, color = site\_type)) + geom\_violin(width = 1) + geom\_jitter(alpha = 0.25) + ggtitle("Salinity by Site Type") + xlab("") + ylab("Salinity (ppt)") + theme(legend.position = "none", axis.text.x = element\_text(face = "bold"))  
  
site\_type\_salinity

## Warning: Removed 15 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 15 rows containing missing values (geom\_point).



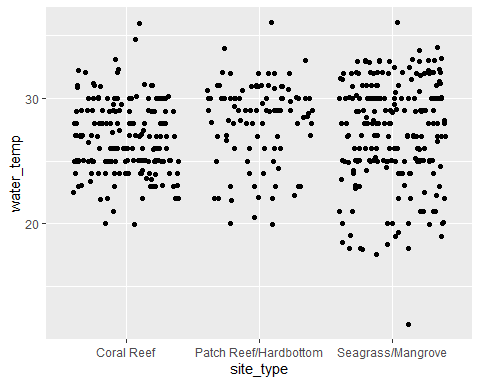
ggsave(filename = "../../results/site\_type\_salinity.png",plot = site\_type\_salinity)

## Saving 5 x 4 in image

## Warning: Removed 15 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 15 rows containing missing values (geom\_point).

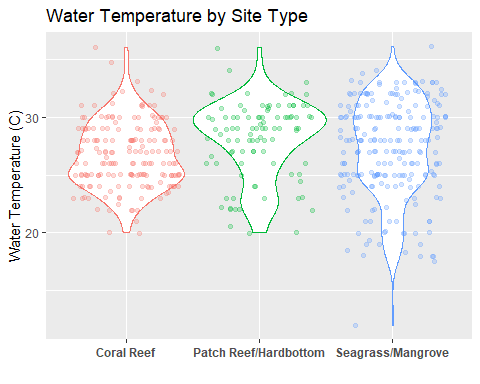
site\_type\_filter %>% ggplot() +   
 geom\_jitter(aes(x = site\_type, y = water\_temp))

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



site\_type\_water\_temp <- ggplot(island\_side\_filter, aes(x = site\_type, y = water\_temp, color = site\_type)) + geom\_violin(width = 1) + geom\_jitter(alpha = 0.25) + ggtitle("Water Temperature by Site Type") + xlab("") + ylab("Water Temperature (C)") + theme(legend.position = "none", axis.text.x = element\_text(face = "bold"))  
  
site\_type\_water\_temp

## Warning: Removed 12 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 12 rows containing missing values (geom\_point).



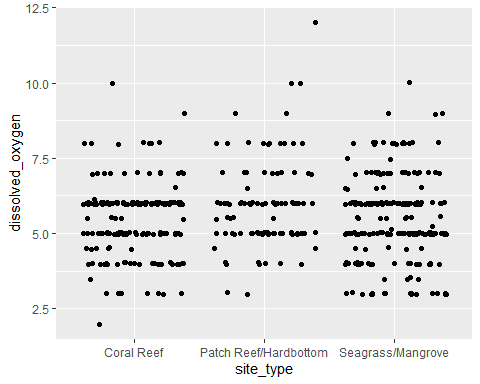
ggsave(filename = "../../results/site\_type\_water\_temp.png",plot = site\_type\_water\_temp)

## Saving 5 x 4 in image

## Warning: Removed 12 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 12 rows containing missing values (geom\_point).

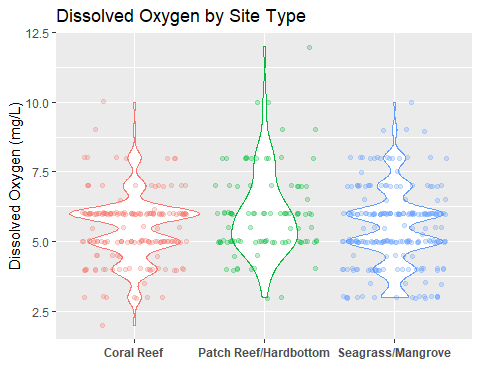
site\_type\_filter %>% ggplot() +   
 geom\_jitter(aes(x = site\_type, y = dissolved\_oxygen))

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



site\_type\_dissolved\_oxygen <- ggplot(island\_side\_filter, aes(x = site\_type, y = dissolved\_oxygen, color = site\_type)) + geom\_violin(width = 1) + geom\_jitter(alpha = 0.25) + ggtitle("Dissolved Oxygen by Site Type") + xlab("") + ylab("Dissolved Oxygen (mg/L)") + theme(legend.position = "none", axis.text.x = element\_text(face = "bold"))  
  
site\_type\_dissolved\_oxygen

## Warning: Removed 4 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 4 rows containing missing values (geom\_point).



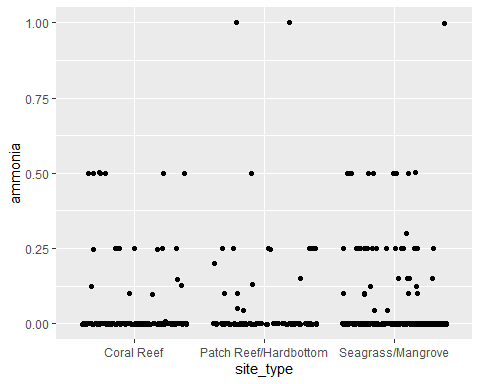
ggsave(filename = "../../results/site\_type\_dissolved\_oxygen.png",plot = site\_type\_dissolved\_oxygen)

## Saving 5 x 4 in image

## Warning: Removed 4 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 4 rows containing missing values (geom\_point).

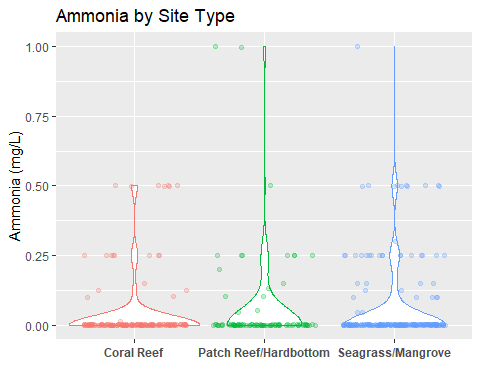
site\_type\_filter %>% ggplot() +   
 geom\_jitter(aes(x = site\_type, y = ammonia))

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



site\_type\_ammonia <- ggplot(island\_side\_filter, aes(x = site\_type, y = ammonia, color = site\_type)) + geom\_violin(width = 1) + geom\_jitter(alpha = 0.25) + ggtitle("Ammonia by Site Type") + xlab("") + ylab("Ammonia (mg/L)") + theme(legend.position = "none", axis.text.x = element\_text(face = "bold"))  
  
site\_type\_ammonia

## Warning: Removed 10 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 10 rows containing missing values (geom\_point).



ggsave(filename = "../../results/site\_type\_ammonia.png",plot = site\_type\_ammonia)

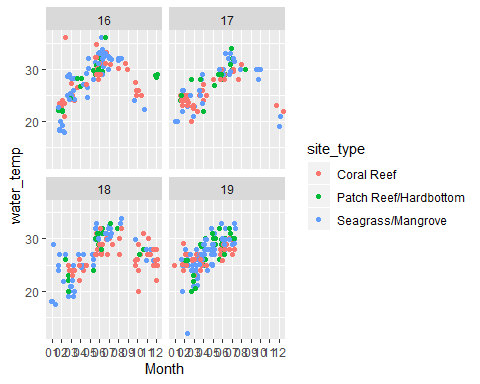
## Saving 5 x 4 in image

## Warning: Removed 10 rows containing non-finite values (stat\_ydensity).  
  
## Warning: Removed 10 rows containing missing values (geom\_point).

Now that we have baseline visuals of the differet island sides and site types, lets start to look at some seasonal changes and see if we can identify the presence of Hurricane Irma in the data.

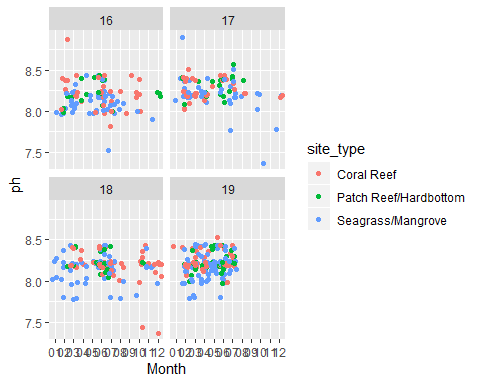
site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = water\_temp, color = site\_type)) + facet\_wrap(~Year)

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



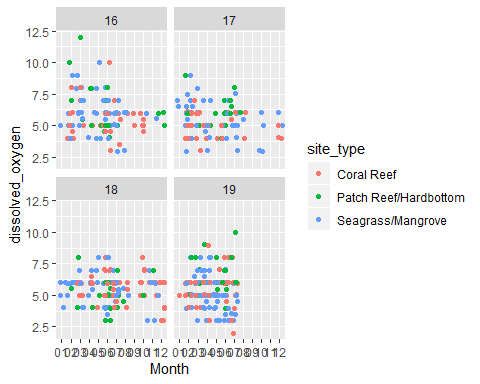
site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = ph, color = site\_type)) + facet\_wrap(~Year)

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



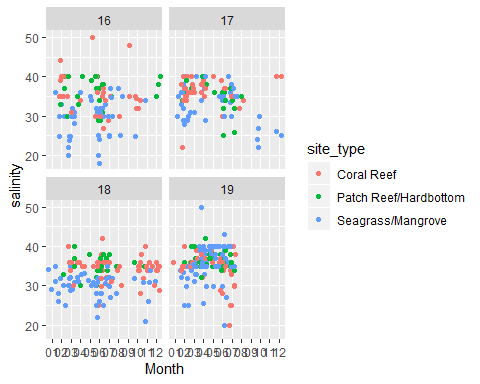
site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = dissolved\_oxygen, color = site\_type)) + facet\_wrap(~Year)

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



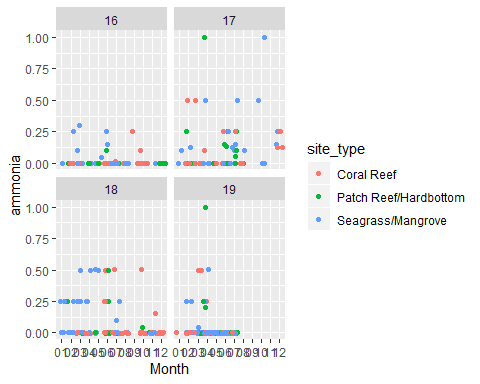
site\_type\_filter %>% ggplot() +  
 geom\_jitter(aes(x = Month, y = salinity, color = site\_type)) + facet\_wrap(~Year)

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



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).



To Do: 1. Make Bay vs Ocean plots for each paremeter 2. Make a plot of site type for each parameter 3. geom\_line for view hurricane Irma