multi-variate

Getting Started

Libraries

Data

```
dust<- read_csv(here("data", "processed_data","dust_master.csv")) #Different Dust attribut
```

```
Rows: 32 Columns: 9
-- Column specification ----------------------------------------------------
Delimiter: ","
dbl  (8): t1, t7, t13, t19, tsum, tavg, t7sum, t7avg
date (1): date

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
data<- read_csv(here("data", "processed_data", "tx_master.csv")) #Master Data
```

```
New names:
Rows: 1005 Columns: 26
-- Column specification
--------------------------------------------------------- Delimiter: "," chr
(3): hr_cst, site, Type dbl (22): ...1, SUM, dust, chl, sal, temp, do_mgl,
do_per, pH, secchi, amm,... date (1): date
i Use `spec()` to retrieve the full column specification for this data. i
Specify the column types or set `show_col_types = FALSE` to quiet this message.
* `` -> `...1`
```

Cleaning

Clean and Trim

```r
dust<- dust %>%
  filter(!is.na(date),
         date %in% as.Date("2022-07-04"):as.Date("2022-07-19"))


data<- data %>%
  filter(date %in% as.Date("2022-07-04"):as.Date("2022-07-19"))
```

Make into one data set

```r
master<-
  dust %>% full_join(data) %>%
  filter(hr_cst %in% "01") #While this is technically not for hour 1, this gets ride of al
```

```
Joining, by = "date"
```

```r
master<- master %>%
  select(site, sal, temp, copies_mL,t7sum, date, orthop,nn, don,doc) %>%
  mutate(t7sum = log10(t7sum))
```

Site-specific Data sets and Clean df

```r
b<- master %>%
  filter(site %in% "Gulf" | date %in% as.Date('2022-07-05') : as.Date('2022-07-06'))
b<- b %>%
mutate(
  lag =Lag(b$t7sum, shift  = 1)) %>% #Create 1 day lag
filter(between(date, as.Date('2022-07-07'), as.Date('2022-07-19'))) %>%
select(!c(date,t7sum,site)) %>%
  rename("log10dust" = "lag")
```

Models

0. Null Model:

Create 5 Fold

```
fold <- vfold_cv(b, v = 5, repeats = 5) #Data was too small to stratify (<20)
```

Make Models

Get Metrics for Null

RMSE

```
null_met <-
  collect_metrics(null_fit)
null_met
```

```
# A tibble: 2 x 6
  .metric .estimator    mean      n std_err .config
  <chr>   <chr>        <dbl>  <int>   <dbl> <chr>
1 rmse    standard    50159.    25   2560. Preprocessor1_Model1
2 rsq     standard      NaN      0      NA Preprocessor1_Model1
```

```
rm0<- 50771.79
rs0<- "NA"
```

AIC

```
null_mod<-
  lm(copies_mL ~ 1, data = b)
aic0<-AIC(null_mod) %>% round(digits = 2)
```

1. Dust Model

Model

```r
#Create Recipe for Growth
growth_recipe1<- recipe(copies_mL ~ log10dust, data = b)

#Workflow that adds recipe to model
Growth_wflow1<-
  workflow() %>%
  add_model(lm_mod) %>%
  add_recipe(growth_recipe1)

#Use workflow to fit model to  data set
growth_fit1<- Growth_wflow1 %>%
  fit(data = b)

#View as Tibble
growth_fit1 %>%
  extract_fit_parsnip() %>%
  tidy()
```

```
# A tibble: 2 x 5
  term         estimate std.error statistic  p.value
  <chr>           <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  135834.    29387.      4.62 0.000947
2 log10dust    101960.    36063.      2.83 0.0179
```

Metrics

```r
aug_test1 <- augment(growth_fit1, b)
rmse1 <- aug_test1 %>% rmse(truth = copies_mL, .pred)
rsq1 <- aug_test1 %>% rsq(truth = copies_mL, .pred)

mod1<- lm(copies_mL ~ log10dust, b)
aic1<-AIC(mod1) %>% round(digits = 2)

metrics1<- full_join(rmse1, rsq1)
```

```
Joining, by = c(".metric", ".estimator", ".estimate")
```

```r
metrics1
```

```
# A tibble: 2 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 rmse    standard      36132.
2 rsq     standard       0.444
```

2. Water Chemistry Model

Model

```
#Create Recipe for Growth
growth_recipe2<- recipe(copies_mL ~ sal+temp, data = b)

#Workflow that adds recipe to model
Growth_wflow2<-
  workflow() %>%
  add_model(lm_mod) %>%
  add_recipe(growth_recipe2)

#Use workflow to fit model to  data set
growth_fit2<- Growth_wflow2 %>%
  fit(data = b)

#View as Tibble
growth_fit2 %>%
  extract_fit_parsnip() %>%
  tidy()
```

```
# A tibble: 3 x 5
  term        estimate std.error statistic p.value
  <chr>          <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)  659851.   826558.     0.798   0.445
2 sal            9688.    19373.     0.500   0.629
3 temp         -33450.    11891.    -2.81    0.0203
```

Metrics

```
aug_test2 <- augment(growth_fit2, b)
rmse2 <- aug_test2 %>% rmse(truth = copies_mL, .pred)
rsq2 <- aug_test2 %>% rsq(truth = copies_mL, .pred)
```

5

```r
mod2<- lm(copies_mL ~ sal+temp, b)
aic2<- AIC(mod2)%>% round(digits = 2)

metrics2<- full_join(rmse2, rsq2)
```

```
Joining, by = c(".metric", ".estimator", ".estimate")
```

```r
metrics2
```

```
# A tibble: 2 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 rmse    standard      34417.
2 rsq     standard        0.496
```

3. Chemistry + Dust

Model

```r
#Create Recipe for Growth
growth_recipe3<- recipe(copies_mL ~ sal+temp + log10dust, data = b)

#Workflow that adds recipe to model
Growth_wflow3<-
  workflow() %>%
  add_model(lm_mod) %>%
  add_recipe(growth_recipe3)

#Use workflow to fit model to  data set
growth_fit3<- Growth_wflow3 %>%
  fit(data = b)

#View as Tibble
growth_fit3 %>%
  extract_fit_parsnip() %>%
  tidy()
```

```
# A tibble: 4 x 5
```

```
  term          estimate std.error statistic p.value
  <chr>            <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)   972372.   787547.     1.23     0.252
2 sal            -5385.    20210.    -0.266    0.797
3 temp          -23395.    12646.    -1.85     0.101
4 log10dust      70187.    43844.     1.60     0.148
```

Metrics

```r
aug_test3 <- augment(growth_fit3, b)
rmse3 <- aug_test3 %>% rmse(truth = copies_mL, .pred)
rsq3 <- aug_test3 %>% rsq(truth = copies_mL, .pred)

mod3<- lm(copies_mL ~ sal+temp + log10dust, b)
aic3<- AIC(mod3)%>% round(digits = 2)

metrics3<- full_join(rmse3, rsq3)
```

```
Joining, by = c(".metric", ".estimator", ".estimate")
```

```r
metrics3
```

```
# A tibble: 2 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 rmse    standard      29952.
2 rsq     standard       0.618
```

4. Model 3 + Nutrients

Model

```r
#Create Recipe for Growth
growth_recipe4<- recipe(copies_mL ~ sal+temp+orthop+log10dust+nn, data = b)

#Workflow that adds recipe to model
Growth_wflow4<-
  workflow() %>%
  add_model(lm_mod) %>%
```

```
    add_recipe(growth_recipe4)

  #Use workflow to fit model to  data set
  growth_fit4<- Growth_wflow4 %>%
    fit(data = b)

  #View as Tibble
  growth_fit4 %>%
    extract_fit_parsnip() %>%
    tidy()
```

```
# A tibble: 6 x 5
  term          estimate std.error statistic p.value
  <chr>            <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept) 1110980.    821260.      1.35   0.225
2 sal           -15150.    22242.     -0.681   0.521
3 temp          -17676.    13802.     -1.28    0.248
4 orthop         42986.   260706.      0.165   0.874
5 log10dust      92110.    51339.      1.79    0.123
6 nn            286931.   467807.      0.613   0.562
```

Metrics

```
  aug_test4 <- augment(growth_fit4, b)
  rmse4 <- aug_test4 %>% rmse(truth = copies_mL, .pred)
  rsq4 <- aug_test4 %>% rsq(truth = copies_mL, .pred)

  mod4<- lm(copies_mL ~ sal+temp+orthop+log10dust+nn, b)
  aic4<- AIC(mod4)%>% round(digits = 2)

  metrics4<- full_join(rmse4, rsq4)
```

```
Joining, by = c(".metric", ".estimator", ".estimate")
```

```
  metrics4
```

```
# A tibble: 2 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
```

```
1 rmse    standard   26604.
2 rsq     standard      0.699
```

5. Module 4 + DOM

Model

```r
#Create Recipe for Growth
growth_recipe5<- recipe(copies_mL ~ sal+temp+log10dust+orthop+nn+don+doc, data = b)

#Workflow that adds recipe to model
Growth_wflow5<-
  workflow() %>%
  add_model(lm_mod) %>%
  add_recipe(growth_recipe5)

#Use workflow to fit model to  data set
growth_fit5<- Growth_wflow5 %>%
  fit(data = b)

#View as Tibble
growth_fit5 %>%
  extract_fit_parsnip() %>%
  tidy()
```

```
# A tibble: 8 x 5
  term          estimate std.error statistic p.value
  <chr>            <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)   783691.  1007915.     0.778   0.480
2 sal           -11374.    25053.    -0.454   0.673
3 temp          -10370.    17575.    -0.590   0.587
4 log10dust      84206.    59840.     1.41    0.232
5 orthop        185666.   354271.     0.524   0.628
6 nn            214439.   596064.     0.360   0.737
7 don            -4314.     5160.    -0.836   0.450
8 doc             -149.      272.    -0.546   0.614
```

Metrics

```
aug_test5 <- augment(growth_fit5, b)
rmse5 <- aug_test5 %>% rmse(truth = copies_mL, .pred)
rsq5 <- aug_test5 %>% rsq(truth = copies_mL, .pred)

mod5<- lm(copies_mL ~ sal+temp+log10dust+orthop+nn+don+doc, b)
aic5<-AIC(mod5)%>% round(digits = 2)

metrics5<- full_join(rmse5, rsq5)
```

Joining, by = c(".metric", ".estimator", ".estimate")

```
metrics5
```

```
# A tibble: 2 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 rmse    standard   23880.
2 rsq     standard       0.757
```

Tables

Label Metrics

```
#R2
rs1<- rsq1$.estimate %>% round(digits = 2)
rs2<- rsq2$.estimate %>% round(digits = 2)
rs3<- rsq3$.estimate %>% round(digits = 2)
rs4<- rsq4$.estimate %>% round(digits = 2)
rs5<- rsq5$.estimate %>% round(digits = 2)

#RMSE
rm1<- rmse1$.estimate
rm2<- rmse2$.estimate
rm3<- rmse3$.estimate
rm4<- rmse4$.estimate
rm5<- rmse5$.estimate
```

Make Table

Table 1: Gulf: Comparative linear regression analysis of Vibrio copies per mL between 6 operational models.

Model	R2	RMSE	AIC
Null Model	NA	50771.79	296.98
Model 1: Dust	0.44	36132.11	291.93
Model 2: Water Chemistry	0.5	34417.21	292.77
Model 3: Chemistry + Dust	0.62	29952.44	291.43
Model 4: Model 3 + Nutrients	0.7	26604.11	292.59
Model 5: Model 4 + DOM	0.76	23880.38	293.99

```
kable(data.frame(Model = c("Null Model", "Model 1: Dust", "Model 2: Water Chemistry", "Mod
           R2 = c(rs0,rs1,rs2,rs3,rs4,rs5),
           RMSE = c(rm0,rm1,rm2,rm3,rm4,rm5),
           AIC = c(aic0, aic1,aic2,aic3,aic4,aic5)),
       caption = "Gulf: Comparative linear regression analysis of Vibrio copies per mL betw
```