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

```
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

```
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"
```

```
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

```
b<- master %>%
  filter(site %in% "Canals" | date %in% as.Date('2022-07-05') : as.Date('2022-07-06'))
b<- b %>%
mutate(
  lag =Lag(b$t7sum, shift  = 2)) %>% #Create 2 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_bo <- 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   7773.    25    648. Preprocessor1_Model1
2 rsq     standard    NaN      0     NA  Preprocessor1_Model1
```

```
rm0<- 8082.466
rs0<- "NA"
```

AIC

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

1. Dust Model

Model

```
#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)   31954.     4942.      6.47 0.0000721
2 log10dust     20859.     5861.      3.56 0.00519
```

Metrics

```
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")
```

```
metrics1
```

```
# A tibble: 2 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 rmse    standard    5191.
2 rsq     standard       0.559
```

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)  119302.   286782.     0.416   0.687
2 sal            3593.     2483.     1.45    0.182
3 temp          -8174.    10957.    -0.746   0.475
```

Metrics

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

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

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

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

```
metrics2
```

```
# A tibble: 2 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 rmse    standard    7025.
2 rsq     standard       0.192
```

3. Chemistry + Dust

Model

```
#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)   85730.   222324.     0.386  0.710
2 sal             958.     2164.     0.443  0.670
3 temp          -3055.     8698.    -0.351  0.735
4 log10dust     19429.     7331.     2.65   0.0292
```

Metrics

```
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")
```

```
metrics3
```

```
# A tibble: 2 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 rmse    standard    5127.
2 rsq     standard       0.570
```

4. Model 3 + Nutrients

Model

```
#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)    3772.   248974.    0.0152  0.988
2 sal            -833.     2590.   -0.321   0.759
3 temp           1880.    10285.    0.183   0.861
4 orthop       -33479.    24604.   -1.36    0.223
5 log10dust     19036.     7460.    2.55    0.0434
6 nn            30403.    20479.    1.48    0.188
```

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    4336.
2 rsq     standard       0.692
```

5. Module 4 + DOM

Model

```
#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) -46784.   444141.     -0.105   0.921
2 sal            -667.     3436.     -0.194   0.856
3 temp           2364.    15803.      0.150   0.888
4 log10dust     18286.     9447.      1.94    0.125
5 orthop       -26689.    35514.     -0.752   0.494
6 nn            26002.    29522.      0.881   0.428
7 don             380.     1049.      0.363   0.735
8 doc              10.1      65.4     0.154   0.885
```

Metrics

```r
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")
```

```r
metrics5
```

```
# A tibble: 2 x 3
  .metric .estimator .estimate
  <chr>   <chr>          <dbl>
1 rmse    standard    4262.
2 rsq     standard       0.703
```

Tables

Label Metrics

```r
#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: Canals: Comparative linear regression analysis of Vibrio copies per mL between 6 operational models.

Model	R2	RMSE	AIC
Null Model	NA	8082.466	253.19
Model 1: Dust	0.56	5191.220	245.37
Model 2: Water Chemistry	0.19	7025.411	254.63
Model 3: Chemistry + Dust	0.57	5126.525	249.07
Model 4: Model 3 + Nutrients	0.69	4335.503	249.04
Model 5: Model 4 + DOM	0.7	4262.452	252.64

```
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 = "Canals: Comparative linear regression analysis of Vibrio copies per mL be
```