Salinity_vs_Production_Rates

Nusrat Noor

Fall 2023

R Markdown

This is an R Markdown document for the Fall 2023 final project in EDA. This document explores the relationship between salinity and production rates of seagrass.

Notes

For temp/salinity, readings were taken on the same day for either summer/reference or winter/reference. We decided to only use the reference measurements to keep the data uniform.

Production rate (mass growth) readings were not taken the same days as the temp/salinity. They were usually taken the following day.

ex_week (experiment week) is the consistent variable across all the data collected. We used this to join the data frames.

Setup

```
#loading in necessary packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(here)
```

```
## here() starts at /home/guest/R/Final_Project/BeyerBolgerNoor_Env872_EDA_FinalProject
```

```
library(lubridate)

#checked directory
here()
```

```
## [1] "/home/guest/R/Final_Project/BeyerBolgerNoor_Env872_EDA_FinalProject"
```

```
#create and set theme
mytheme <- theme_grey(base_size = 14) +
  theme(axis.title = element_text(colour = "darkred"),
```

```
        title = element_text(colour = "darkred"),
        axis.text = element_text(colour = "darkblue"))
theme_set(mytheme)
```

Importing and Cleaning Data

```
#read in the temp/salinity data
temp_sal <- read.csv(here("./Data/Raw/Temperature-Salinity.csv"), stringsAsFactors = TRUE)

#set date column as date
temp_sal$date <- as.Date(temp_sal$date)

#read in production rate  data
prod_rates <- read.csv(here("./Data/Raw/Seagrass-production-rates.csv"), stringsAsFactors = TRUE)

#set date column as date
prod_rates$date <- as.Date(prod_rates$date)
```

```
#filtered out the reference treatment only
#omitted any NAs
temp_sal_clean <- temp_sal %>%
  filter(treatment == "reference") %>%
  na.omit()

#selected the desired columns
#filtered out just the reference treatment
#calculated the mean mass growth
#omitted any NAs
prod_rates_clean <- prod_rates %>%
  select("treatment", "date", "exp_week", "gr_mass") %>%
  filter(treatment == "reference") %>%
  group_by(date) %>%
  summarise(meangrowth = mean(gr_mass), exp_week = median(exp_week)) %>%
  mutate(treatment = "reference") %>%
  na.omit()
```

```
#joined the salinity and production rate data frames together by the experiment week
joined <- left_join(x = prod_rates_clean, y = temp_sal_clean,
                    join_by(exp_week ==exp_week, treatment ==treatment))

#set date column as date
joined$date.x <- as.Date(joined$date.x)
```

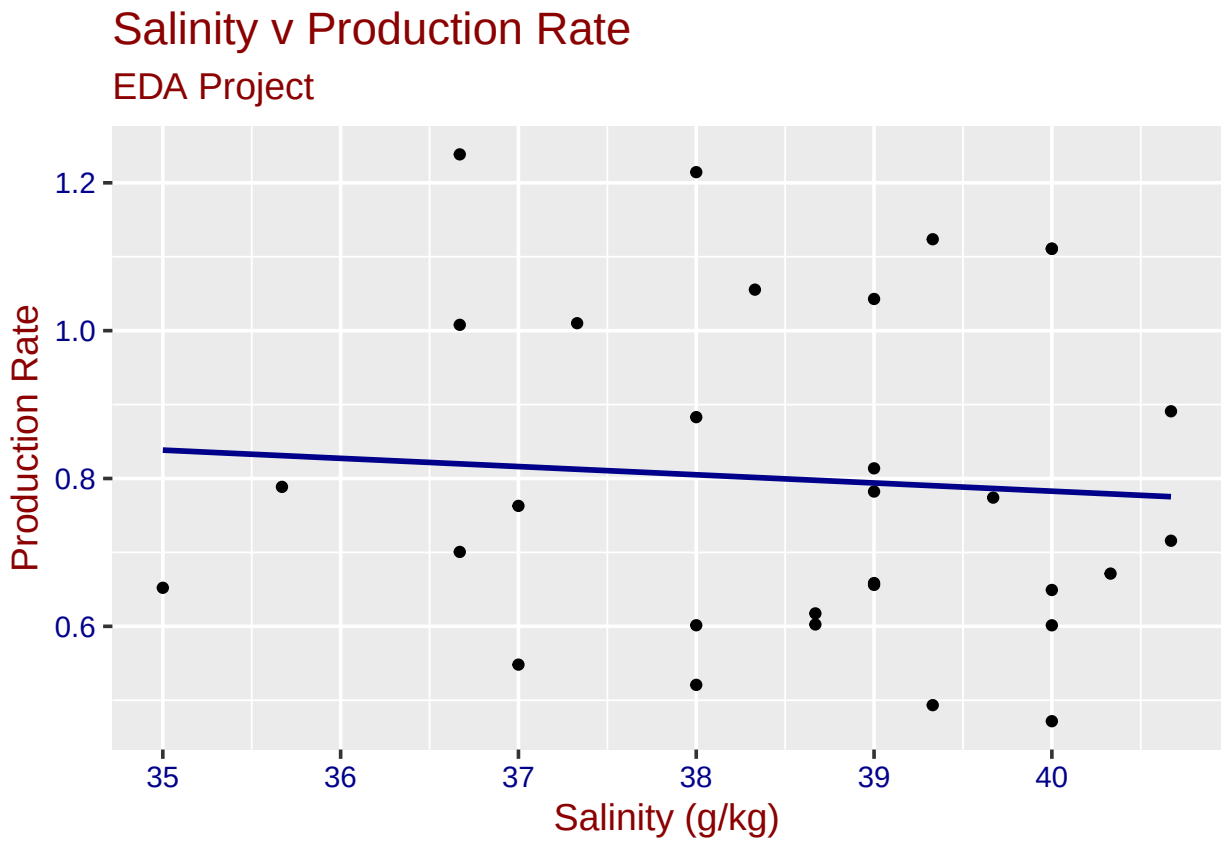
Plotting the Data

```
#plotted salinity by production rate
ggplot(joined, aes(y = meangrowth, x = salinity)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "darkblue") +
  labs(y = "Production Rate", x = "Salinity (g/kg)",
       title = "Salinity v Production Rate",
       subtitle = "EDA Project")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
```

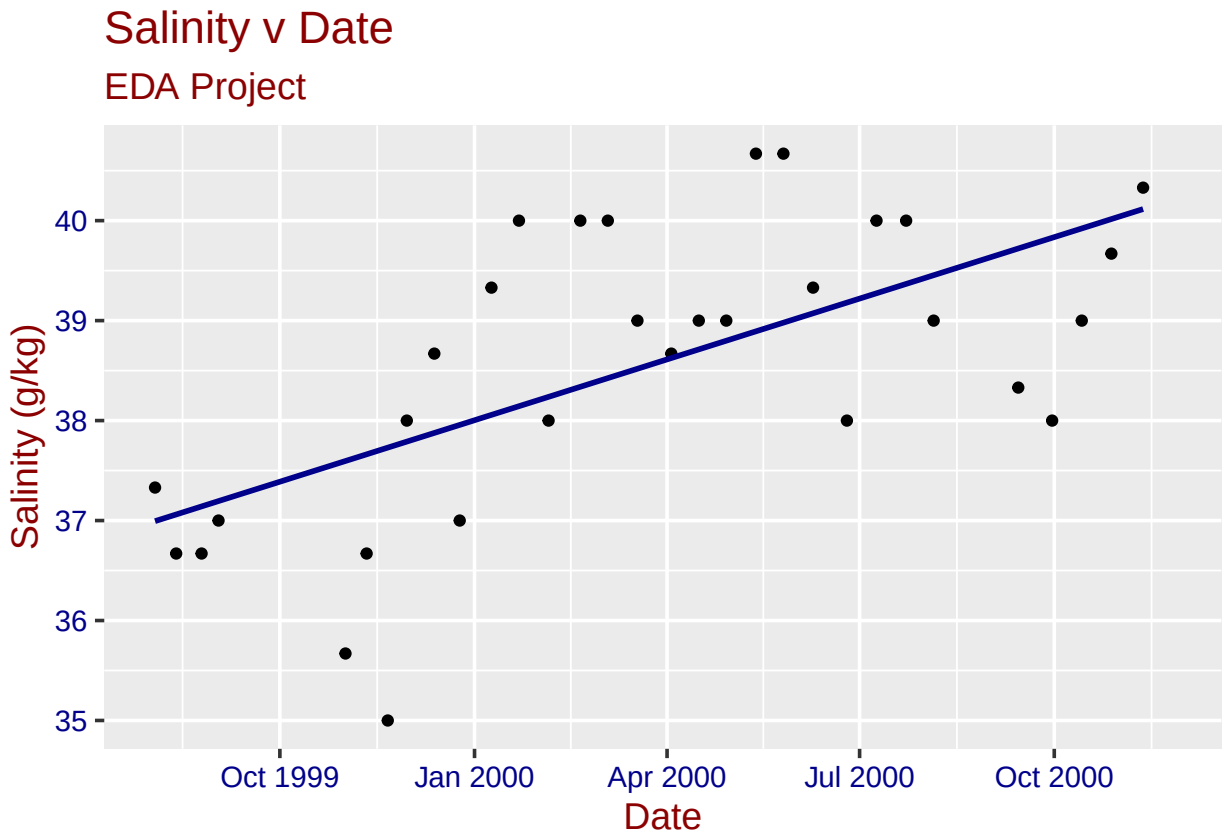
```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
```




```
#plotted salinity over time
ggplot(joined, aes(x = date.x, y = salinity)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "darkblue") +
  labs(x = "Date", y = "Salinity (g/kg)",
       title = "Salinity v Date",
       subtitle = "EDA Project")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 rows containing non-finite values (`stat_smooth()`).
## Removed 1 rows containing missing values (`geom_point()`).
```



> Results: There is a slight decrease in production rates as salinity increases but the data is fairly dispersed.

Salinity Analysis

```
#ran linear regression
sal_prod_regression <-
  lm(data = joined, salinity ~ meangrowth)
summary(sal_prod_regression)
```

```
##
## Call:
## lm(formula = salinity ~ meangrowth, data = joined)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6089 -0.8893  0.3930  1.2121  2.1762
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  38.9231     1.0108  38.509   <2e-16 ***
## meangrowth   -0.4818     1.2194  -0.395    0.696
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.496 on 29 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.005355,   Adjusted R-squared:  -0.02894
```

```
## F-statistic: 0.1561 on 1 and 29 DF,  p-value: 0.6956
```

Results: The p-value is 0.6956 and r-squared is -0.02894. There is no significant relationship between salinity and production rates of seagrass.