report for erosion

Yingmai Chen

```
library(ggplot2)
library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.1     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v lubridate 1.9.2     v tibble    3.2.1
v purrr     1.0.2     v tidyr     1.3.0
-- 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 becom
```

```
library(readr)
library(gridExtra)
```

```
Attaching package: 'gridExtra'

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

    combine
```

```
library(car)
```

```
Loading required package: carData

Attaching package: 'car'
```

```
The following object is masked from 'package:dplyr':

    recode

The following object is masked from 'package:purrr':

    some
```

```r
library(grid)
library(stats)
library(ggcorrplot)
```

read data and normalize the data

```r
stan <- function(x) {
  return((x - min(x)) / (max(x) - min(x)))
}
data1 <- read_csv("erosion.csv")# data include max wave height
```

```
Rows: 31 Columns: 8
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): Bluff
dbl (7): Orientation (deg), RR (m/yr), Max Wave Height (m), Mud (%), BaseEl ...

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.
```

```r
names(data1) <- gsub(" ", "_", names(data1))
names(data1) <- gsub("\\(", "", names(data1))
names(data1) <- gsub("\\)", "", names(data1))
names(data1) <- gsub("/", "_per_", names(data1))
names(data1) <- gsub("%", "percent", names(data1))
if ("RR (m/yr)" %in% names(data1)) {
  names(data1)[names(data1) == "RR (m/yr)"] <- "RR_m_per_yr"
}

numeric_columns <- sapply(data1, is.numeric) & !names(data1) %in% c("Bluff", "RR_m_per_yr"
```

```
data1[numeric_columns] <- lapply(data1[numeric_columns], stan)
data1$RR_m_per_yr <- stan(data1$RR_m_per_yr)
```

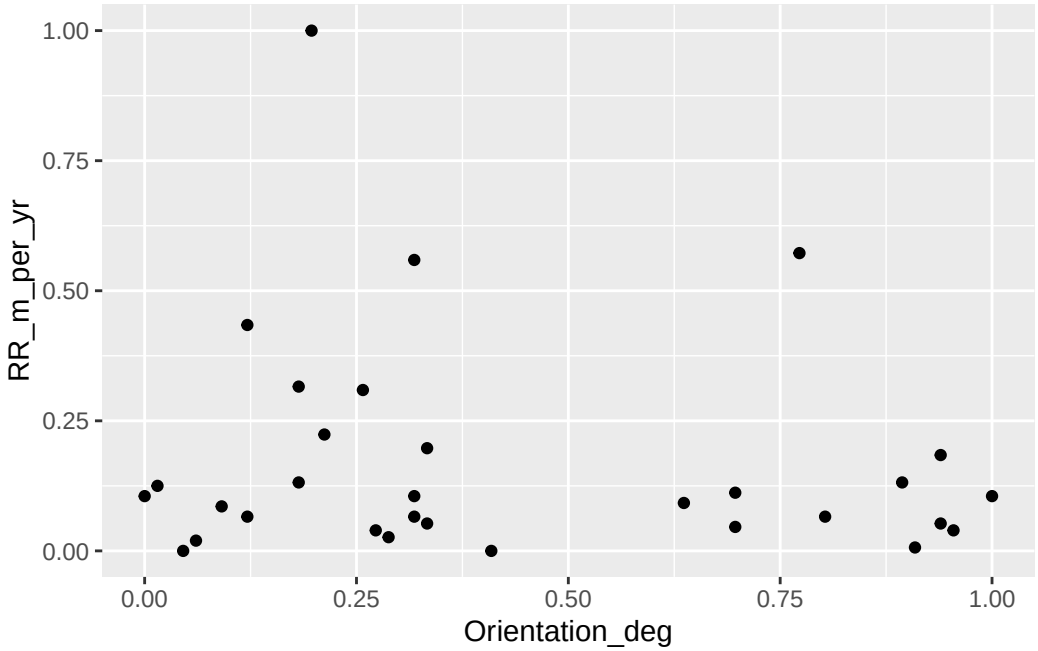
```
data2 <- read_csv("erosionnne15.csv")# data include NNE 15 m/s of max waveheight
```

```
Rows: 31 Columns: 8
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (1): Bluff
dbl (7): Orientation (deg), RR (m/yr), Wave Height for NNE wind 15 m/s (m), ...

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.
```

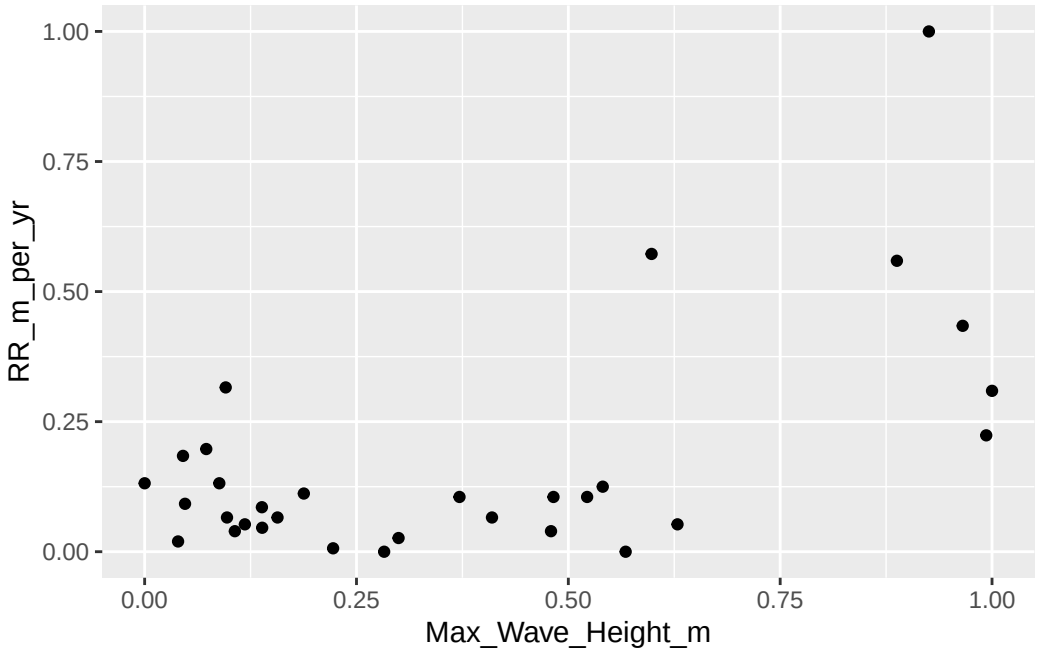
```
names(data2) <- gsub(" ", "_", names(data2))
names(data2) <- gsub("\\(", "", names(data2))
names(data2) <- gsub("\\)", "", names(data2))
names(data2) <- gsub("/", "_per_", names(data2))
names(data2) <- gsub("%", "percent", names(data2))
if ("RR (m/yr)" %in% names(data2)) {
  names(data2)[names(data2) == "RR (m/yr)"] <- "RR_m_per_yr"
}
numeric_columns <- sapply(data2, is.numeric) & !names(data2) %in% c("Bluff", "RR_m_per_yr"
data2[numeric_columns] <- lapply(data2[numeric_columns], stan)
data2$RR_m_per_yr <- stan(data2$RR_m_per_yr)
```

Scatterplot for each variable and boxplot for categorical variable seaWall

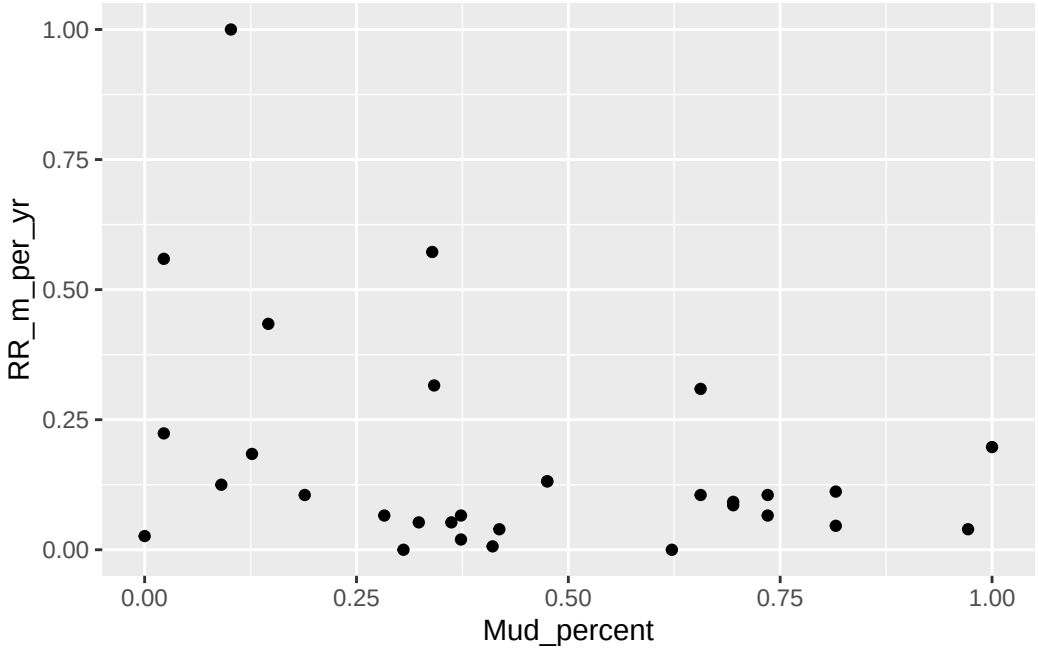
```
ggplot(data1,aes(x=Orientation_deg,y=RR_m_per_yr))+
  geom_point()
```



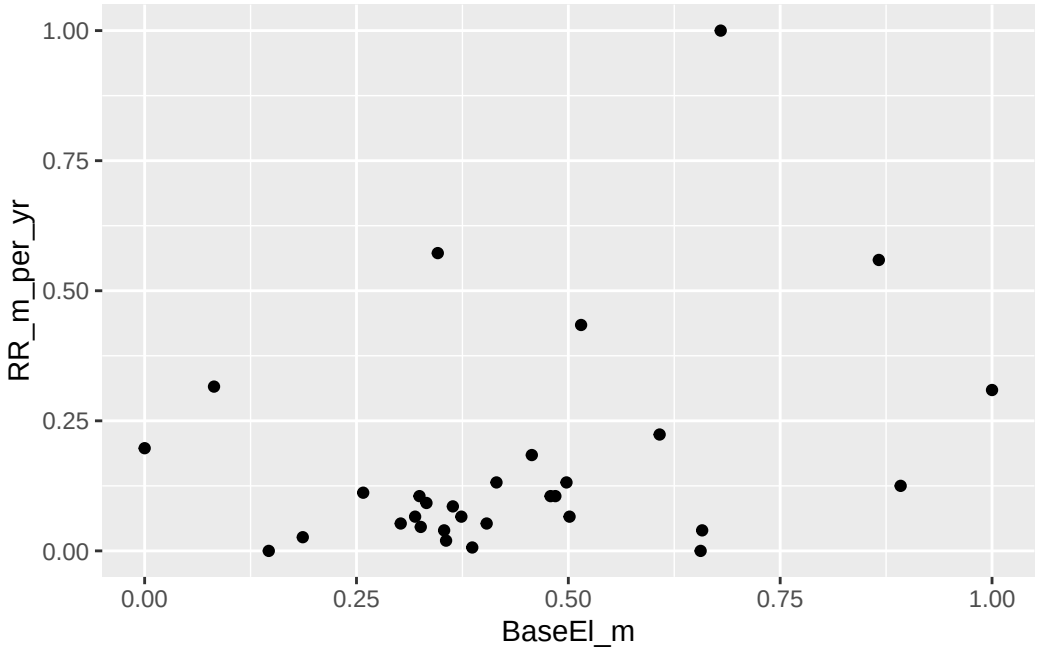

```
ggplot(data1,aes(x=Max_Wave_Height_m,y=RR_m_per_yr))+
  geom_point()
```



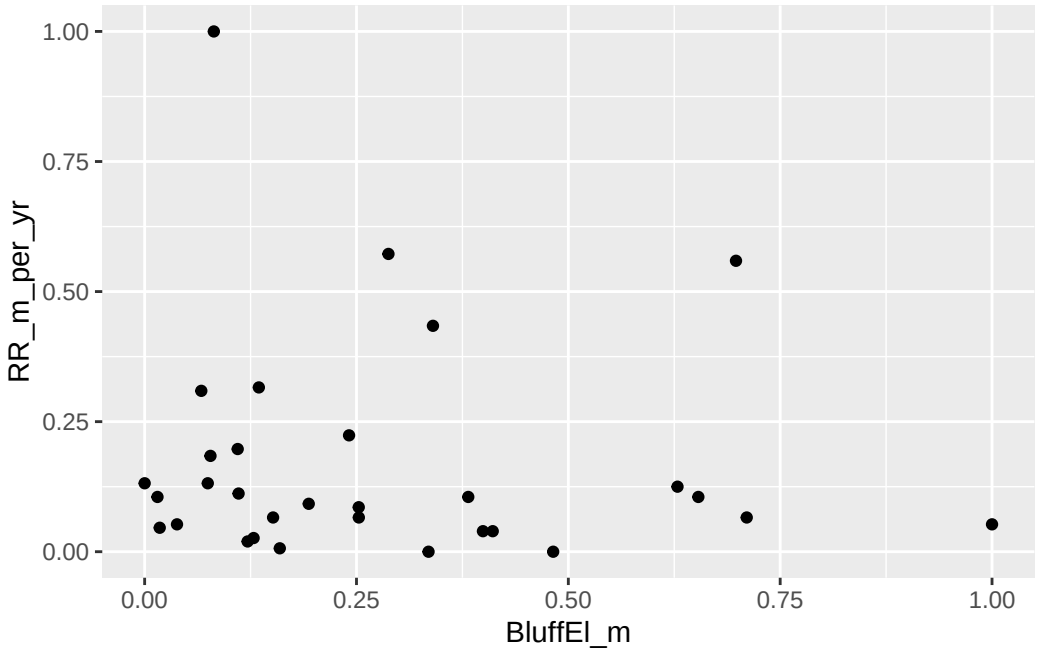
```
ggplot(data1,aes(x=Mud_percent,y=RR_m_per_yr))+
  geom_point()
```



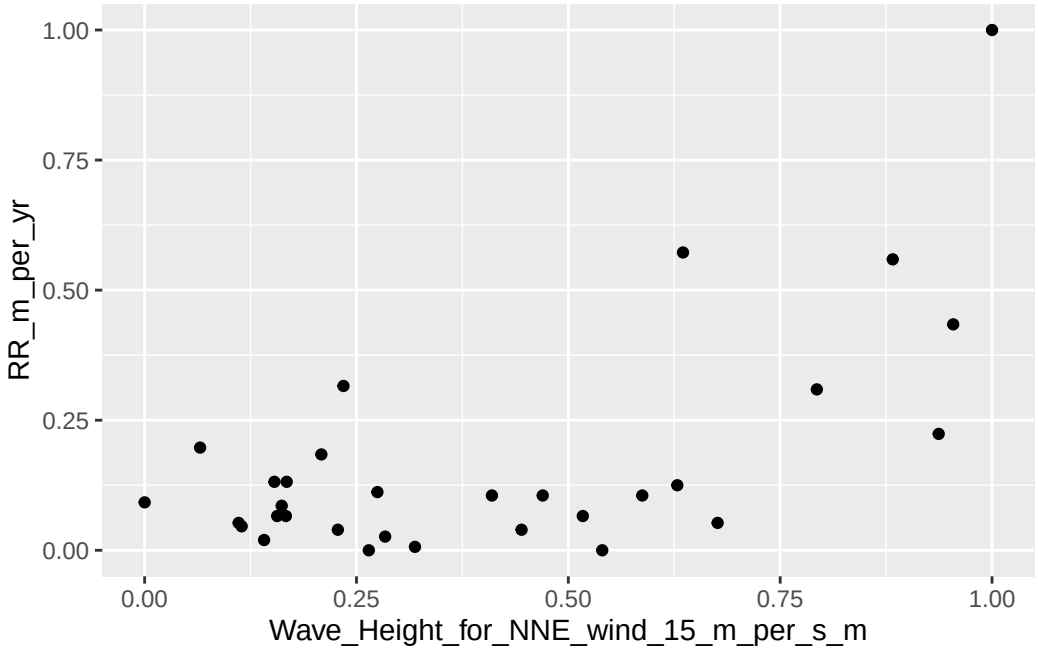

```
ggplot(data1,aes(x=BaseEl_m,y=RR_m_per_yr))+
  geom_point()
```




```
ggplot(data1,aes(x=BluffEl_m,y=RR_m_per_yr))+
  geom_point()
```



```r
ggplot(data2,aes(x=Wave_Height_for_NNE_wind_15_m_per_s_m,y=RR_m_per_yr))+
  geom_point()
```



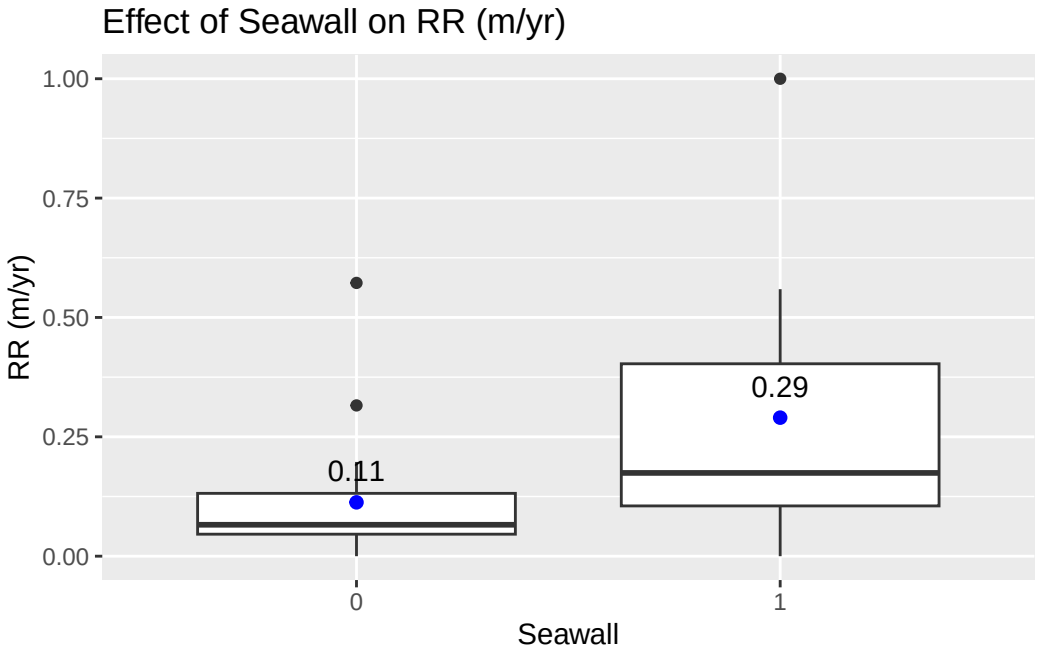

```r
data1$Seawall <- factor(data1$Seawall)
model <- lm(`RR_m_per_yr` ~ Seawall, data = data1)
summary(model)
```

```
Call:
lm(formula = RR_m_per_yr ~ Seawall, data = data1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.29013 -0.08976 -0.04699  0.01894  0.70987

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.11278    0.04392   2.568   0.0157 *
Seawall1     0.17735    0.07734   2.293   0.0293 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.2013 on 29 degrees of freedom
Multiple R-squared:  0.1535,    Adjusted R-squared:  0.1243
F-statistic: 5.259 on 1 and 29 DF,  p-value: 0.02927
```

```r
ggplot(data1, aes(x = Seawall, y = `RR_m_per_yr`)) +
  geom_boxplot() +
  stat_summary(fun = mean, geom = "point", shape = 20, size = 3, color = "blue")+
  stat_summary(fun = mean, geom = "text", aes(label = round(..y.., 2)), vjust = -1) +
  labs(title = "Effect of Seawall on RR (m/yr)", x = "Seawall", y = "RR (m/yr)")
```

```
Warning: The dot-dot notation (`..y..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(y)` instead.
```



Correlation Matrix of data1(with max wave height)

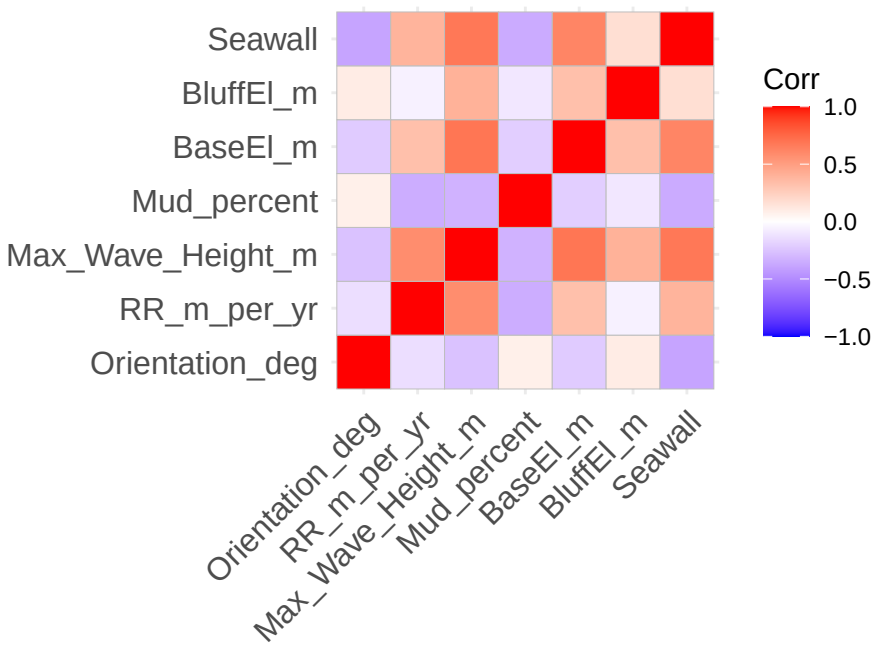
```r
correlationdata1 <- read_csv("correlationdata1.csv")
```

```
Rows: 31 Columns: 7
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl (7): Orientation_deg, RR_m_per_yr, Max_Wave_Height_m, Mud_percent, BaseE...

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.
```

```r
correlation_matrix1 <- cor(correlationdata1)
print(correlation_matrix1)
```

```
                 Orientation_deg RR_m_per_yr Max_Wave_Height_m Mud_percent
Orientation_deg        1.00000000 -0.14011080        -0.2550198  0.08098435
RR_m_per_yr           -0.14011080  1.00000000         0.5781057 -0.35498111
Max_Wave_Height_m     -0.25501981  0.57810566         1.0000000 -0.32830955
Mud_percent            0.08098435 -0.35498111        -0.3283096  1.00000000
BaseEl_m              -0.21828166  0.32952373         0.6912812 -0.21230637
BluffEl_m              0.10289157 -0.06169165         0.4016323 -0.10458668
Seawall               -0.39219328  0.39180088         0.6848326 -0.35843787
                   BaseEl_m   BluffEl_m    Seawall
Orientation_deg   -0.2182817  0.10289157 -0.3921933
RR_m_per_yr        0.3295237 -0.06169165  0.3918009
Max_Wave_Height_m  0.6912812  0.40163227  0.6848326
Mud_percent       -0.2123064 -0.10458668 -0.3584379
BaseEl_m           1.0000000  0.33197416  0.6239764
BluffEl_m          0.3319742  1.00000000  0.1661165
Seawall            0.6239764  0.16611646  1.0000000
```

```r
ggcorrplot(correlation_matrix1)
```



Correlation Matrix of data2(with max wave height on NNE 15m/s)

```
correlationdata2 <- read_csv("correlationdata2.csv")
```

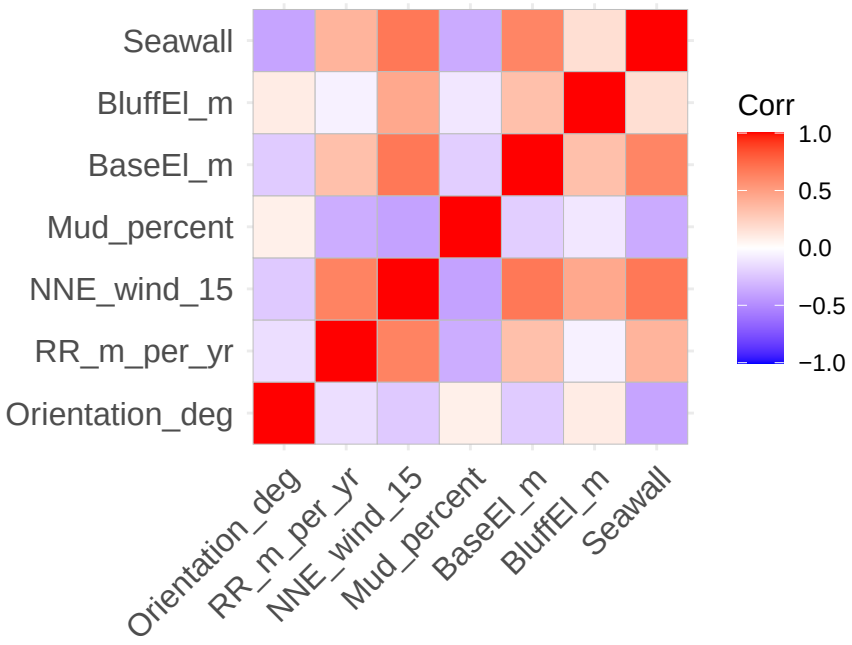
```
Rows: 31 Columns: 7
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl (7): Orientation_deg, RR_m_per_yr, NNE_wind_15, Mud_percent, BaseEl_m, B...

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.
```

```
correlation_matrix2 <- cor(correlationdata2)
print(correlation_matrix2)
```

```
                Orientation_deg RR_m_per_yr NNE_wind_15 Mud_percent   BaseEl_m
Orientation_deg      1.00000000 -0.14011080  -0.2254391  0.08098435 -0.2182817
RR_m_per_yr         -0.14011080  1.00000000   0.6301882 -0.35498111  0.3295237
NNE_wind_15         -0.22543906  0.63018825   1.0000000 -0.40007650  0.6825278
```

```
Mud_percent          0.08098435 -0.35498111  -0.4000765  1.00000000 -0.2123064
BaseEl_m            -0.21828166  0.32952373   0.6825278 -0.21230637  1.0000000
BluffEl_m            0.10289157 -0.06169165   0.4489944 -0.10458668  0.3319742
Seawall             -0.39219328  0.39180088   0.6757305 -0.35843787  0.6239764
                  BluffEl_m    Seawall
Orientation_deg  0.10289157 -0.3921933
RR_m_per_yr     -0.06169165  0.3918009
NNE_wind_15      0.44899441  0.6757305
Mud_percent     -0.10458668 -0.3584379
BaseEl_m         0.33197416  0.6239764
BluffEl_m        1.00000000  0.1661165
Seawall          0.16611646  1.0000000
```

```
ggcorrplot(correlation_matrix2)
```



linear regression for data1(with max wave height)

```
target <- "RR_m_per_yr"
predictors <- setdiff(names(data1), c("Bluff", target))
model_formula <- as.formula(paste(target, "~", paste(predictors, collapse = " + ")))

model1 <- lm(model_formula, data = data1)
print(summary(model1))
```

```
Call:
lm(formula = model_formula, data = data1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.29687 -0.08353 -0.02441  0.06200  0.48000

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.14438    0.10996   1.313  0.20159
Orientation_deg    0.04229    0.10612   0.399  0.69376
Max_Wave_Height_m  0.52523    0.16321   3.218  0.00368 **
Mud_percent       -0.14023    0.12099  -1.159  0.25785
BaseEl_m          -0.05829    0.20675  -0.282  0.78043
BluffEl_m         -0.31922    0.14579  -2.190  0.03851 *
Seawall1          -0.03885    0.10377  -0.374  0.71143
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1738 on 24 degrees of freedom
Multiple R-squared:  0.478, Adjusted R-squared:  0.3475
F-statistic: 3.662 on 6 and 24 DF,  p-value: 0.01006
```

```
predictions <- predict(model1, data1)
mse <- mean((data1[[target]] - predictions)^2)
rsquared <- summary(model1)$r.squared
cat("(MSE):", mse, "\n")
```

```
(MSE): 0.02337344
```

```r
cat("R square:", rsquared, "\n")
```

```
R square: 0.4779682
```

linear regression for data2(with max wave heigh on NNE 15m/s)

```r
target <- "RR_m_per_yr"
predictors <- setdiff(names(data2), c("Bluff", target))
model_formula <- as.formula(paste(target, "~", paste(predictors, collapse = " + ")))

model2 <- lm(model_formula, data = data2)
print(summary(model2))
```

```
Call:
lm(formula = model_formula, data = data2)

Residuals:
     Min       1Q   Median       3Q      Max
-0.30616 -0.07149 -0.03354  0.04292  0.37441

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                          0.06504    0.10148   0.641 0.527610
Orientation_deg                      0.03785    0.09512   0.398 0.694197
Wave_Height_for_NNE_wind_15_m_per_s_m 0.72374    0.16756   4.319 0.000234 ***
Mud_percent                         -0.06583    0.11157  -0.590 0.560675
BaseEl_m                            -0.10408    0.18476  -0.563 0.578423
BluffEl_m                           -0.39778    0.13477  -2.951 0.006962 **
Seawall                             -0.06072    0.09263  -0.655 0.518401
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1559 on 24 degrees of freedom
Multiple R-squared:  0.5796,    Adjusted R-squared:  0.4744
F-statistic: 5.514 on 6 and 24 DF,  p-value: 0.001042
```

```r
predictions <- predict(model2, data2)
mse <- mean((data2[[target]] - predictions)^2)
rsquared <- summary(model2)$r.squared
cat("(MSE):", mse, "\n")
```

```
(MSE): 0.01882508
```

```r
cat("R square:", rsquared, "\n")
```

```
R square: 0.5795532
```