Exploratory analysis of data

Yulong Wang

15/09/2021

1: Synopsis

The basic goal of this analysis is to explore the water age data from FVCOM simulation cases, which study water age change of different scenarios of Nature based Solution in 3 sea level rise events.

The following analysis mainly investigates the impact of severe NbS and SLR events on:

???

2: Data Processing

2.1: Set work folder.

Share same folder with the assignment 1.

```
setwd("/Users/yulong/GitHub/study_case2/ml_models")
```

2.1: Install and load packages

Load packages for data.table and ggplot2.

```
library("data.table")
library("ggplot2")
```

2.2: Data Loading

Download the raw data file and extract the data into a dataframe.Then convert to a data.table

```
waterageDF <- read.csv("data.csv")
# Converting data.frame to data.table
waterageDT <- as.data.table(waterageDF)
```

2.3: Examining Column Names

```
colnames(waterageDT)
```

```
## [1] "lat"       "lon"       "sigma_z"   "slr_hgt"   "nbs_case"  "water_age"
```

```
summary(waterageDT)
```

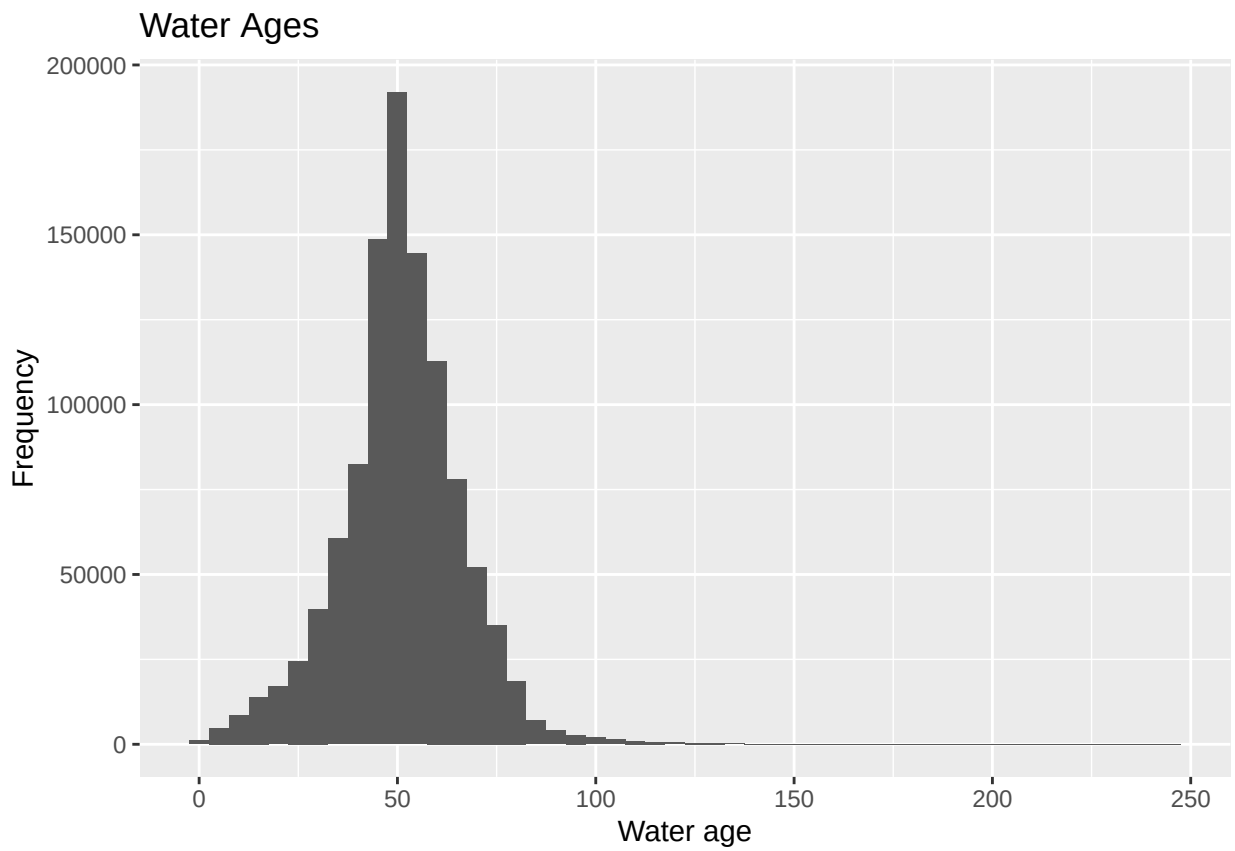
```
##       lat             lon           sigma_z        slr_hgt         nbs_case
##  Min.   :34.98   Min.   :139.6   Min.   : 1.0   Min.   :0.000   Min.   :1.000
##  1st Qu.:35.23   1st Qu.:139.8   1st Qu.: 5.0   1st Qu.:0.000   1st Qu.:1.000
##  Median :35.41   Median :139.8   Median :10.0   Median :0.300   Median :3.000
##  Mean   :35.38   Mean   :139.8   Mean   :10.5   Mean   :0.825   Mean   :2.022
##  3rd Qu.:35.53   3rd Qu.:139.9   3rd Qu.:15.0   3rd Qu.:1.000   3rd Qu.:3.000
##  Max.   :35.70   Max.   :140.1   Max.   :20.0   Max.   :2.000   Max.   :3.000
##    water_age
##  Min.   :  0.12
##  1st Qu.: 43.05
##  Median : 50.69
##  Mean   : 50.94
##  3rd Qu.: 59.72
##  Max.   :246.83
```

This data is the results of Tokyo Bay simulation, with 20 sigma layers and 4 SLR hight: 0, 0.3, 1.0, and 2.0 meters, and the Nature based Solution have 2 cases, represented by the code of "1" and "3".

The water age varies from almost 0 to 247 days, with the mean value of 51 days.

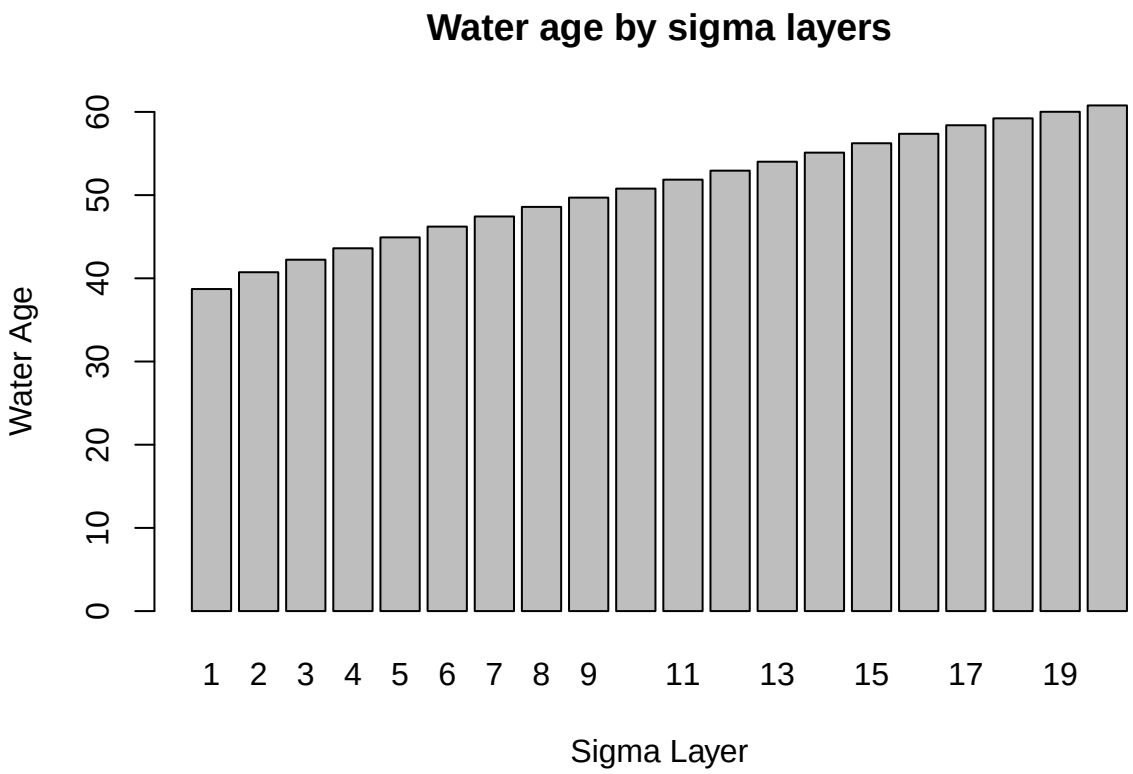
2.3: Show a histogram plot of all the water age values.

```
ggplot(waterageDT, aes(x = water_age)) +
  geom_histogram(binwidth = 5) +
  labs(title = 'Water Ages', x = 'Water age', y = 'Frequency')
```




```
MeanLayer <- waterageDT[,lapply(.SD,mean,na.rm=TRUE),
                        by=sigma_z]

barplot(MeanLayer[, water_age],
        names = MeanLayer[, sigma_z],
        xlab = "Sigma Layer", ylab = "Water Age",
        main = "Water age by sigma layers")
```


Water age by sigma layers
Water Age
0
10
20
30
40
50
60
1
2
3
4
5
6
7
8
9
11
13
15
17
19
Sigma Layer