US Dam Safety & Climate Change

https://github.com/PierreMishra/Mishra_FinalProject_ENV.872

Pierre Mishra

Rationale and Research Questions

Dams are an integral part of the US economy as they satisfy the water demands of a growing population. In addition to providing hydroelectricity and recreational services, they are crucial for regulating water supply in drought and flood-prone areas. Dams are central to storing and managing US waters. Currently, US has a network of 90,000 dams spread across its 50 states, about two-thirds of which were constructed more than 50 years ago. Aging dams can be more prone to failure and can pose risk to downstream cities. Moreover, at the time of their construction, dams might have been constructed in rural areas where a failure would not have caused any damage to life and property. However,growing urbanization has increased the vulnerability of the population living in those formerly desserted areas.

Research Question I: Which physical characteristics of a dam influence its hazard potential?

Realizing the socio-economic risks of a dam failure, the Congress passed the National Dam Inspection Act in 1972 which authorizes the US Army Corps of Engineers (USACE) to create dam inventory of the US. Since then various acts have passed to improve funding for improving data collection and interpretation efforts. The 2019 National Inventory of Dam (NID) dataset includes an inventory of 91,457 dams and classifies them as low, significant, and high-risk potential. A dam is a high safety hazard if its failure is likely to bring a significant damage to human property and loss of life in downstream areas. In addition to criteria determined by USACE, this analysis investigates several features of a dam and explores which dam characteristics can predict its hazard potential.

Research Question II: What is the spatial distribution of North Carolina dams with varying hazard potential in hydrologic units with varying temporal streamflow trends?

One of the common reasons of a dam failure is overtopping due to excessive storms or floods exceeding the capacity of dams. Such gravitation movement of water over the dams can damage the foundation of the dam thereby making it more susceptible to failure. On the other hand, low water availability risks all the services that a dam provides. This paper investigates the temporal trends of water flows in North Carolina (NC) sub-basins and determines how the dams with varying hazard potential are distributed across sub-basins of varying water flow trends.

Dataset Information

The first dataset was procured from the USACE National Inventory of Dam (NID) website at https://nid.sec.usace.army.mil/ords/f?p=105:1:::::: for gathering information on US dams. The dataset was downloaded in a Comma Separated Value (csv) format and was saved as "NID2019_U.csv". The 2019 NID dataset comprised of 91,457 dam records with 69 variables related to its identification, geography, physical characteristics, architecture, regulation, and political affiliation. This dataset was further wrangled and processed to retain certain variables of interest and saved as "dam.csv" for further processing.

The second dataset included water run-off data for US sub basins and was downloaded from the WaterWatch platform hosted by United States Geological Survey (USGS) website at https://waterwatch.usgs.gov/index.php?id=romap3&sid=w__download. The dataset was downloaded in a tab separated text file called "wy01d_col_data.txt". It contained 2,111 columns for every HUC-8 unit in the US and the first column being the years (1901 – 2019) for which water run-off was calculated. The downloaded folder also contained a PDF file ('huc_8_readme.pdf' located in Data/Raw on Github) which included detailed instructions of how water run-off was calculated for every sub-basin.

The third dataset was a shapefile of all the HUC-8 basins in North Carolina. It was downloaded from the ArcGIS application of NC Department of Environmental Quality at http://data-ncdenr.opendata.arcgis.com/datasets/8-digit-huc-subbasins?geometry=-85.940%2C33.597%2C-73.899%2C36.739&orderBy=HUC_8&orderByAsc=false.

The fourth dataset was also a shapefile of US state boundaries by the US Department of Transportation at http://osav-usdot.opendata.arcgis.com/datasets/c6717a90c9fe4f1986ba40789cbe124f_0. The shapefile was copied from the "sf" folder used in the class for the geospatial lessons. State boundary for North Carolina was used for the scope of this analysis.

Exploratory Analysis and Data Wrangling

Preparing dataset for statistical model

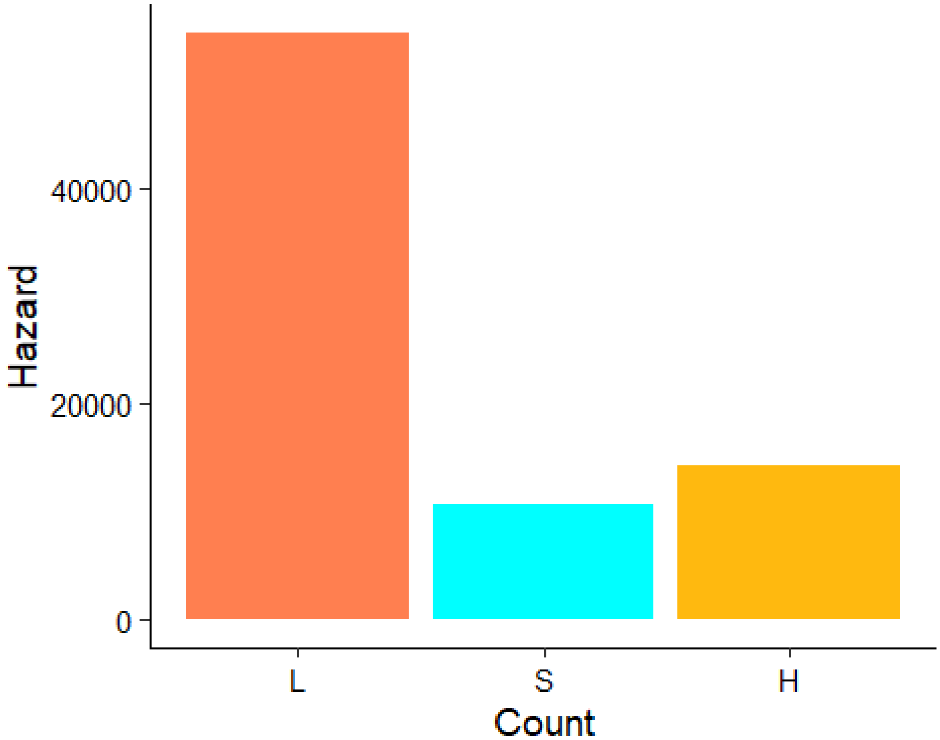
The 2019 NID dam dataset was wrangled to remove irrelevant columns and retain variables of interest. It was then corrected for class of data in each column. Since hazard potential is the response variable of our first research question, any observation with undetermined ("U") or not available ("N") hazard potential was dropped. The hazard potential was also converted into an ordinal categorical variable with the order L < S < H. Another column was added that calculated the age of each dam since the year of completed construction.

```
# Load datasets
# Dam dataset
df <- read.csv("../Data/Raw/NID2019_U.csv")

# Getting familiar with the dataset
colnames(df)

##  [1] "ï..RECORDID"             "DAM_NAME"              "OTHER_DAM_NAME"
##  [4] "DAM_FORMER_NAME"         "STATEID"               "NIDID"
##  [7] "LONGITUDE"               "LATITUDE"              "SECTION"
## [10] "COUNTY"                  "RIVER"                 "CITY"
## [13] "DISTANCE"                "OWNER_NAME"            "OWNER_TYPE"
## [16] "DAM_DESIGNER"            "PRIVATE_DAM"           "DAM_TYPE"
## [19] "CORE"                    "FOUNDATION"            "PURPOSES"
## [22] "YEAR_COMPLETED"          "YEAR_MODIFIED"         "DAM_LENGTH"
## [25] "DAM_HEIGHT"              "STRUCTURAL_HEIGHT"     "HYDRAULIC_HEIGHT"
## [28] "NID_HEIGHT"              "MAX_DISCHARGE"         "MAX_STORAGE"
## [31] "NORMAL_STORAGE"          "NID_STORAGE"           "SURFACE_AREA"
## [34] "DRAINAGE_AREA"           "HAZARD"                "EAP"
## [37] "INSPECTION_DATE"         "INSPECTION_FREQUENCY"  "STATE_REG_DAM"
## [40] "STATE_REG_AGENCY"        "SPILLWAY_TYPE"         "SPILLWAY_WIDTH"
## [43] "OUTLET_GATES"            "VOLUME"                "NUMBER_OF_LOCKS"
## [46] "LENGTH_OF_LOCKS"         "WIDTH_OF_LOCKS"        "FED_FUNDING"
## [49] "FED_DESIGN"              "FED_CONSTRUCTION"      "FED_REGULATORY"
## [52] "FED_INSPECTION"          "FED_OPERATION"         "FED_OWNER"
## [55] "FED_OTHER"               "SOURCE_AGENCY"         "STATE"
## [58] "SUBMIT_DATE"             "URL_ADDRESS"           "CONG_NAME"
## [61] "PARTY"                   "CONG_DIST"             "OTHERSTRUCTUREID"
## [64] "NUMSEPARATESTRUCTURES"   "PERMITTINGAUTHORITY"   "INSPECTIONAUTHORITY"
## [67] "ENFORCEMENTAUTHORITY"    "JURISDICTIONALDAM"     "EAP_LAST_REV_DATE"

# Lowercase column headers
colnames(df) <- tolower(colnames(df))

# Removing unnecessary columns
dam <- df %>%
  select(-c("other_dam_name","dam_former_name", "section",
```

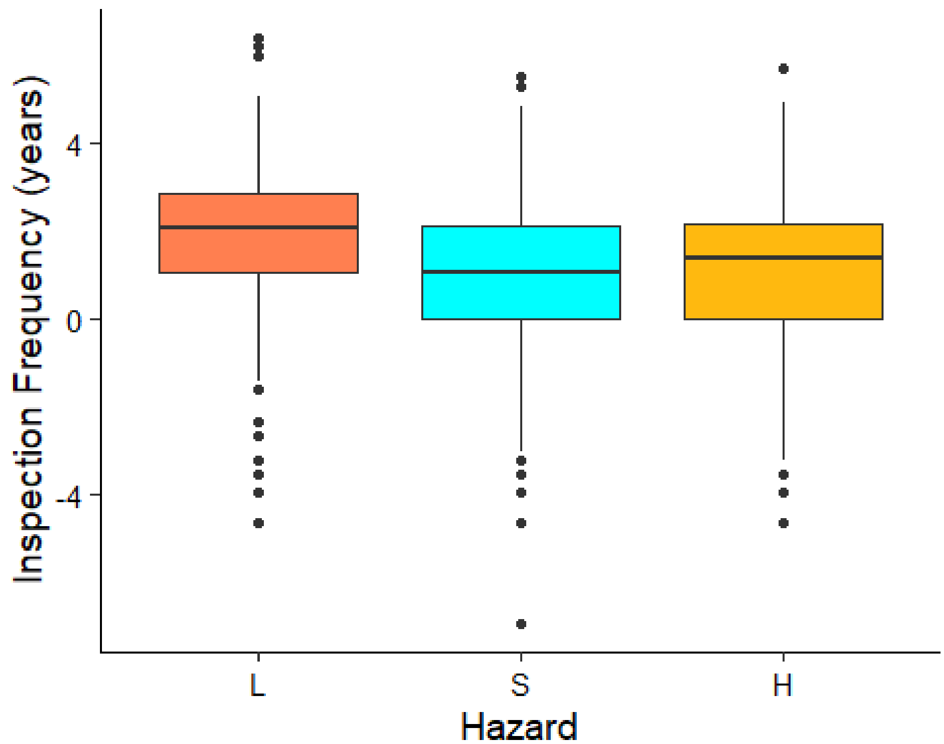
```
            "stateid", "owner_name", "dam_designer",
            "year_modified", "inspection_date",
            "state_reg_agency", "outlet_gates",
            "number_of_locks", "length_of_locks",
            "width_of_locks", "fed_funding",
            "fed_design", "fed_construction",
            "fed_regulatory", "fed_inspection",
            "fed_operation", "fed_owner",
            "fed_other", "source_agency",
            "submit_date","url_address","cong_name",
            "party", "cong_dist", "otherstructureid",
            "numseparatestructures", "permittingauthority",
            "inspectionauthority", "jurisdictionaldam",
            "eap_last_rev_date"))

# Correcting data types
dam$dam_name <- as.character(dam$dam_name)
dam$nidid <- as.character(dam$nidid)
dam$county <- as.character(dam$county)
dam$river <- as.character(dam$river)
dam$city <- as.character(dam$city)
dam$state <- as.character(dam$state)

# Removing dam records with undetermined or NA risk
dam <- droplevels(dam[!(dam$hazard=="U" | dam$hazard=="N"),])

# Making dam hazard column into ordinal categorical type
dam$hazard <- factor(dam$hazard, order = TRUE, levels = c("L", "S", "H"))

# Calculating the age of dams
dam$age <- 2019 - dam$year_completed

# Export (Uncomment to export processed data)
#write.csv(dam, "Data/Processed/dam.csv")
```

To familiarize ourselves with the response variable, the number of dams by their safety hazard were plotted along with other non-physical variables such as inspection frequency, age and distance to the nearest city.

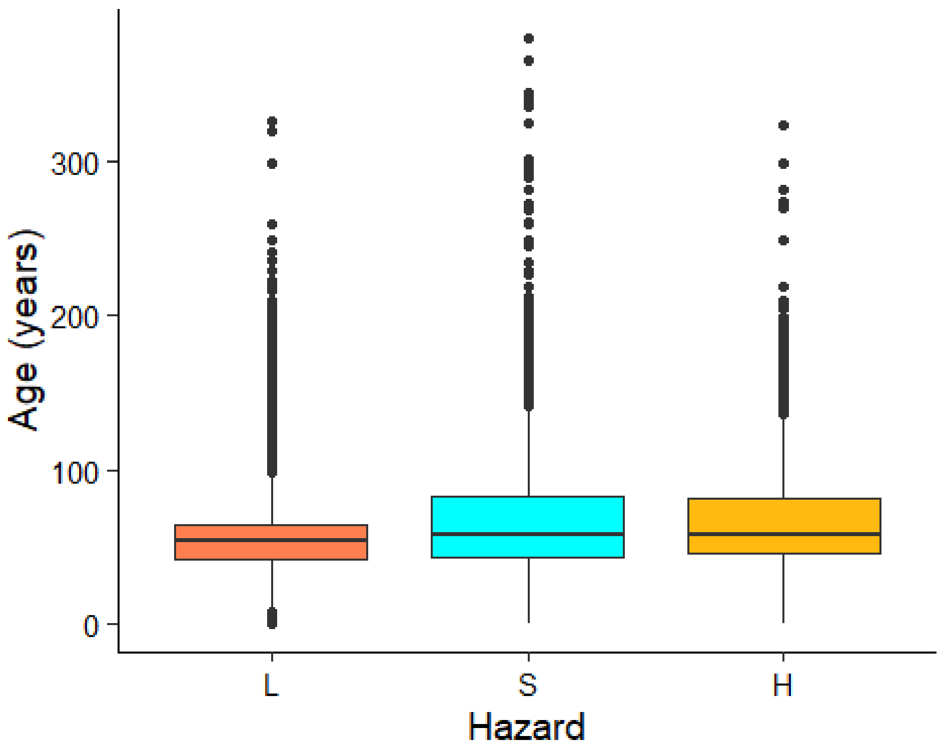
```
# Dam safety hazard
ggplot(dam, aes(x=hazard)) +
  geom_bar(fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Count", y = "Hazard") +
  peaceful.theme
```



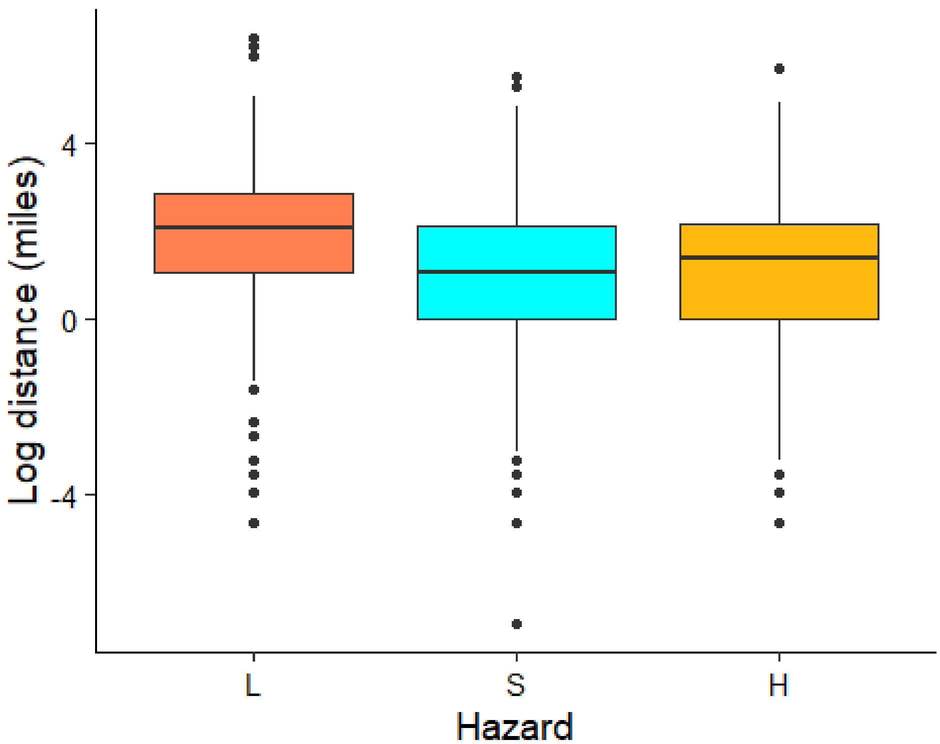

```
# Hazard Vs Inspection Frequency
ggplot(dam) +
  geom_boxplot(aes(x=hazard, y = log(distance)),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = "Inspection Frequency (years)") +
  peaceful.theme

## Warning: Removed 52323 rows containing non-finite values (stat_boxplot).
```




```
# Hazard vs Dam age
ggplot(dam) +
  geom_boxplot(aes(x=hazard, y = age),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = "Age (years)") +
  peaceful.theme

## Warning: Removed 11188 rows containing non-finite values (stat_boxplot).
```



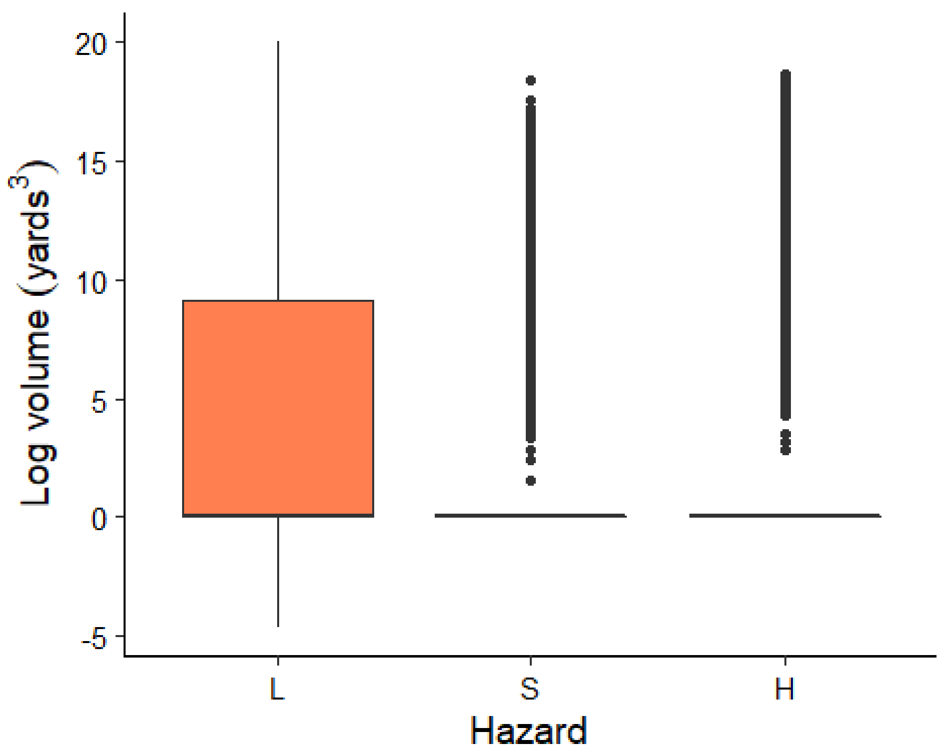

```
# Hazard vs Downstream Distance
ggplot(dam) +
  geom_boxplot(aes(x=hazard, y = log(distance)),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = "Log distance (miles)") +
  peaceful.theme

## Warning: Removed 52323 rows containing non-finite values (stat_boxplot).
```



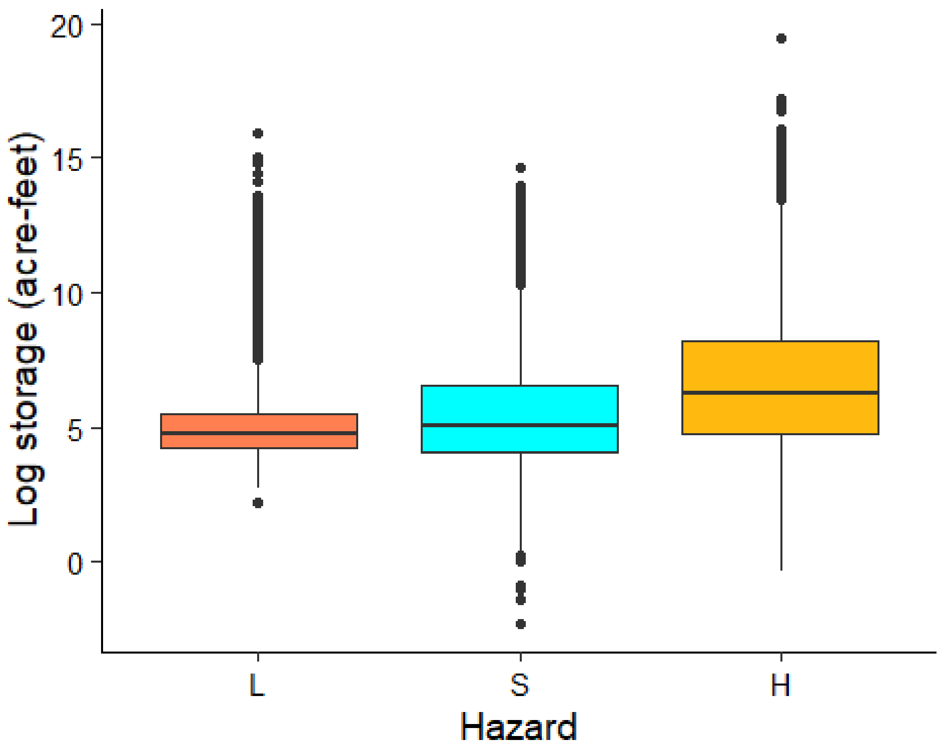
Focusing on our first research question, variables representing physical dam characterisitics were extracted and stored in a separate data frame. Based on background exploratory analysis, the variables were log transformed and stored in a new dataframe along with the response variable, hazard classification. All the variables were plotted as a boxplot. It was found that volume had a lot of zero values. The data was then exported in the processed folder to further perform statistical analysis.

```
# selecting physical characteristics of dams
dam_characteristics <- dam %>%
  select(volume, nid_storage, spillway_width,
         drainage_area, nid_height, surface_area,
         dam_length)

# Log transforming data to improve skewness
summary(is.na(dam_characteristics)) # count NAs in each column

##    volume          nid_storage      spillway_width   drainage_area
##  Mode :logical    Mode :logical    Mode :logical    Mode :logical
##  FALSE:79444      FALSE:79458      FALSE:79451      FALSE:50024
##  TRUE :14                          TRUE :7          TRUE :29434
##  nid_height       surface_area     dam_length
##  Mode :logical    Mode :logical    Mode :logical
##  FALSE:79457      FALSE:61008      FALSE:70402
##  TRUE :1          TRUE :18450      TRUE :9056

dam_characteristics[dam_characteristics==0] = 1 # to prevent log transforming
errors
```

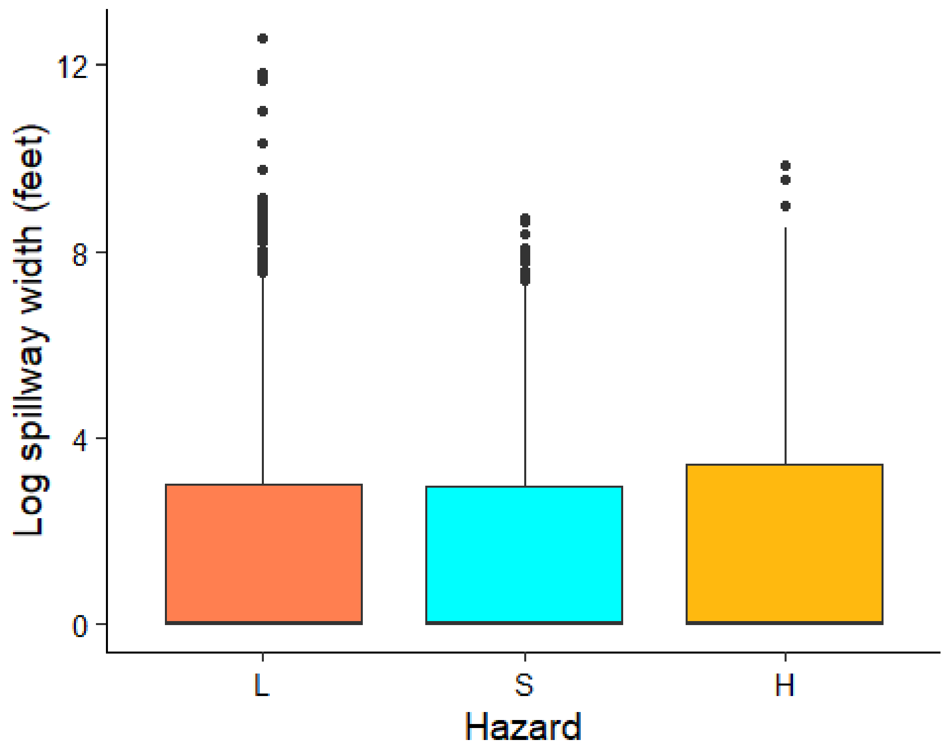
```r
log_dam_characteristics <- log(dam_characteristics)
log_dam_characteristics$hazard <- dam$hazard
#write.csv(log_dam_characteristics,
"Data/Processed/log_dam_characteristics.csv" )

# Checking logged dstribution of each potential predictor variable
ggplot(log_dam_characteristics) +
  geom_boxplot(aes(x=hazard, y = volume),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = expression(Log~volume~(yards^{3}))) +
  peaceful.theme

## Warning: Removed 14 rows containing non-finite values (stat_boxplot).
```



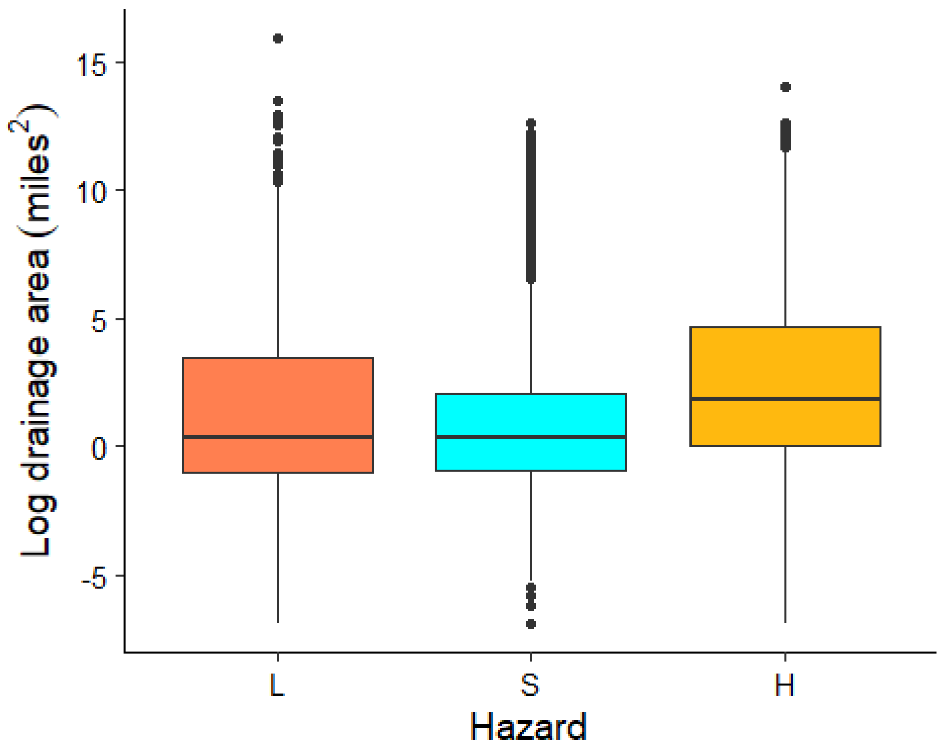

```r
ggplot(log_dam_characteristics) +
  geom_boxplot(aes(x=hazard, y = nid_storage),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = "Log storage (acre-feet)") +
  peaceful.theme
```



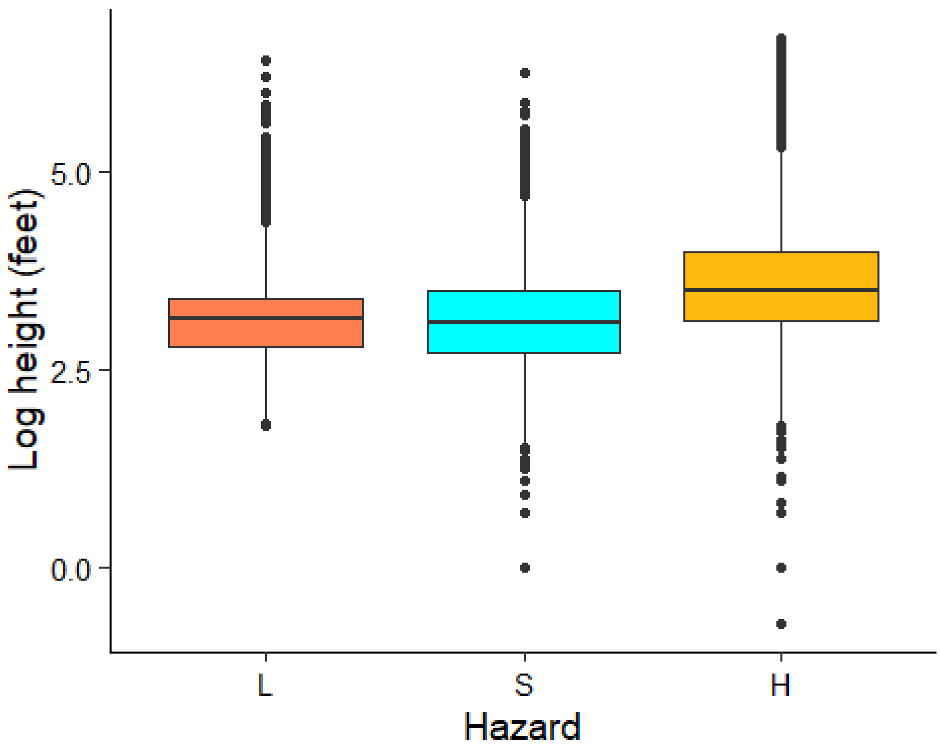

```
ggplot(log_dam_characteristics) +
  geom_boxplot(aes(x=hazard, y = spillway_width),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = "Log spillway width (feet)") +
  peaceful.theme

## Warning: Removed 7 rows containing non-finite values (stat_boxplot).
```



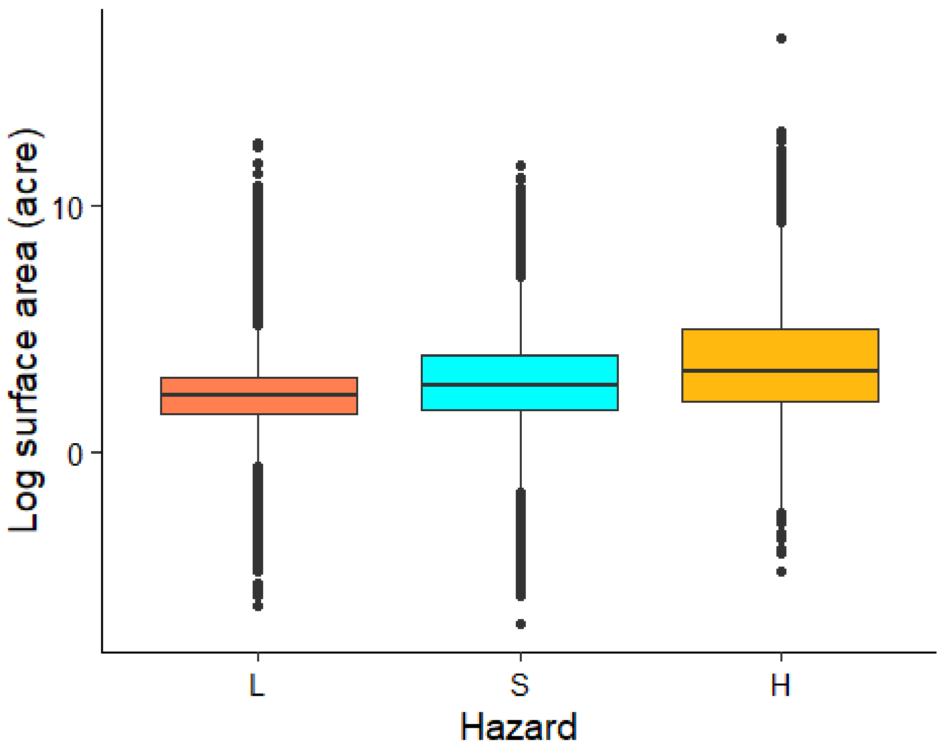

```
ggplot(log_dam_characteristics) +
  geom_boxplot(aes(x=hazard, y = drainage_area),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = expression(Log~drainage~area~(miles^{2}))) +
  peaceful.theme

## Warning: Removed 29434 rows containing non-finite values (stat_boxplot).
```



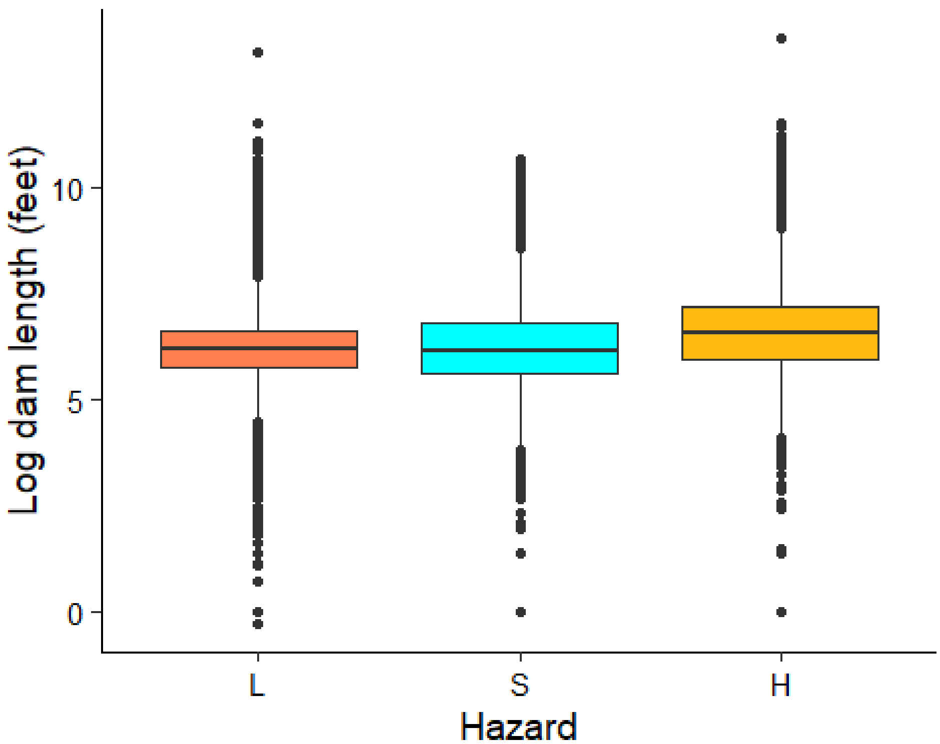

```
ggplot(log_dam_characteristics) +
  geom_boxplot(aes(x=hazard, y = nid_height),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = "Log height (feet)") +
  peaceful.theme

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




```
ggplot(log_dam_characteristics) +
  geom_boxplot(aes(x=hazard, y = surface_area),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = "Log surface area (acre)") +
  peaceful.theme

## Warning: Removed 18450 rows containing non-finite values (stat_boxplot).
```




```
ggplot(log_dam_characteristics) +
  geom_boxplot(aes(x=hazard, y = dam_length),
               fill = c("coral","cyan","darkgoldenrod1")) +
  labs (x = "Hazard", y = "Log dam length (feet)") +
  peaceful.theme

## Warning: Removed 9056 rows containing non-finite values (stat_boxplot).
```



Preparing datasets for geospatial analysis

The dataset containing water run-off time series for all US sub-basins was imported. Its columns were renamed by erasing the letter 'X' in front of HUC-8 number. A shapefile for NC sub-basins was used to extract run-off data for North Carolina HUC-8 units. The newly obtained NC gage data was stored in a separate dataframe. It was converted into a time series object for monotonic trend analysis.

```
# Importing gage time series data
huc8 <- read.csv("../Data/Raw/wy01d_col_data.txt",
                 sep = "\t")
# Importing NC sub-basins shapefile
basin_nc <- st_read("../Data/Raw/8Digit_HUC_Subbasins.shp")

## Reading layer `8Digit_HUC_Subbasins' from data source `C:\Users\Peaceful
Pierre\Documents\Academics\Spring 2020\Environmental Data
Analytics\Environmental_Data_Analytics_2020\Project\Mishra_FinalProject_ENV.8
72\Data\Raw\8Digit_HUC_Subbasins.shp' using driver `ESRI Shapefile'
## Simple feature collection with 53 features and 7 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -84.32162 ymin: 33.83437 xmax: -75.45998 ymax:
36.58841
## CRS:            4326

# renaming columns of gage dataset
colnames(huc8) <- sub("X", "", colnames(huc8))
```

```r
# selecting gage columns for North Carolina by matching with NC basin
shapefile
huc8_nc <- huc8 %>%
  select(date, one_of(as.character(basin_nc$HUC_8)))

## Warning: Unknown columns: `03020301`, `03020302`, `03040208`

ncolumns <- ncol(huc8_nc)
huc8_nc$date <- as.Date(huc8_nc$date, origin = "1901-01-01")
huc8_nc_ts <- ts(huc8_nc[2:ncolumns])

# Saving run-off time series for NC (Uncoomment to export the processed data)
#write.csv(huc8_nc, "Data/Processed/huc8_runoff_nc.csv")
```

From the 2019 NID dam dataset used above, North Carolina dams were subsetted along with their relevant geographical information. Using the ‘sf’ package it was converted from tabular data to an ‘sf’ object. Then using the US state boundary shapefile, NC boundary was stored in a new ‘sf’ object and relevant coordinate reference systems were assigned.

```r
# subsetting north carolina dams
dam_nc <- dam[which(dam$state == "NC"),]
dam_nc$river <- tolower(dam_nc$river)
dam_nc <- select(dam_nc, nidid, longitude, latitude,
                 county, river, city, hazard, state)

# importing NC boundary shapefile
state_bound_raw <- st_read("../Data/Raw/state_bounds.shp")

## Reading layer `state_bounds' from data source `C:\Users\Peaceful
Pierre\Documents\Academics\Spring 2020\Environmental Data
Analytics\Environmental_Data_Analytics_2020\Project\Mishra_FinalProject_ENV.8
72\Data\Raw\state_bounds.shp' using driver `ESRI Shapefile'
## Simple feature collection with 56 features and 17 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -10429050 ymin: -78231.9 xmax: 3407868 ymax: 6747277
## proj4string:    +proj=aea +lat_1=29.5 +lat_2=45.5 +lat_0=23 +lon_0=-96
+x_0=0 +y_0=0 +datum=NAD83 +units=m +no_defs

state_bound_raw_nc <- state_bound_raw[which(state_bound_raw$NAME == "North
Carolina"),]

# setting CRS
state_bound_nc <- state_bound_raw_nc
st_crs(state_bound_nc) <- "+proj=aea +lat_1=29.5 +lat_2=45.5 +lat_0=23
+lon_0=-96 +x_0=0 +y_0=0 +datum=NAD83 +units=m +no_defs"
state_bound_nc <- state_bound_nc %>%
  st_set_crs(5070)
```

```r
# projecting CRS for shapefiles
na_albers_proj4 <- "+proj=aea +lat_1=20 +lat_2=60 +lat_0=40 +lon_0=-96 +x_0=0
+y_0=0 +datum=NAD83 +units=m +no_defs"
na_albers_epsg <- 102008
state_bound_nc_albers <- sf::st_transform(state_bound_nc, crs =
na_albers_proj4) %>%
  st_set_crs(na_albers_epsg)

# checking CRS code
st_crs(state_bound_nc)

## Coordinate Reference System:
##   User input: EPSG:5070
##   wkt:
## PROJCS["NAD83 / Conus Albers",
##     GEOGCS["NAD83",
##         DATUM["North_American_Datum_1983",
##             SPHEROID["GRS 1980",6378137,298.257222101,
##                 AUTHORITY["EPSG","7019"]],
##             TOWGS84[0,0,0,0,0,0,0],
##             AUTHORITY["EPSG","6269"]],
##         PRIMEM["Greenwich",0,
##             AUTHORITY["EPSG","8901"]],
##         UNIT["degree",0.0174532925199433,
##             AUTHORITY["EPSG","9122"]],
##         AUTHORITY["EPSG","4269"]],
##     PROJECTION["Albers_Conic_Equal_Area"],
##     PARAMETER["standard_parallel_1",29.5],
##     PARAMETER["standard_parallel_2",45.5],
##     PARAMETER["latitude_of_center",23],
##     PARAMETER["longitude_of_center",-96],
##     PARAMETER["false_easting",0],
##     PARAMETER["false_northing",0],
##     UNIT["metre",1,
##         AUTHORITY["EPSG","9001"]],
##     AXIS["X",EAST],
##     AXIS["Y",NORTH],
##     AUTHORITY["EPSG","5070"]]

st_crs(state_bound_nc_albers)

## Coordinate Reference System:
##   User input: EPSG:102008
##   wkt:
## PROJCS["North_America_Albers_Equal_Area_Conic",
##     GEOGCS["GCS_North_American_1983",
##         DATUM["North_American_Datum_1983",
##             SPHEROID["GRS_1980",6378137,298.257222101]],
##         PRIMEM["Greenwich",0],
##         UNIT["Degree",0.017453292519943295]],
```

```
##     PROJECTION["Albers_Conic_Equal_Area"],
##     PARAMETER["False_Easting",0],
##     PARAMETER["False_Northing",0],
##     PARAMETER["longitude_of_center",-96],
##     PARAMETER["Standard_Parallel_1",20],
##     PARAMETER["Standard_Parallel_2",60],
##     PARAMETER["latitude_of_center",40],
##     UNIT["Meter",1],
##     AUTHORITY["EPSG","102008"]]

# converting tabular data to sf format
dam_nc_geo <- st_as_sf (dam_nc, coords = c("longitude",
                                      "latitude"),
                     crs = 4326, dim = "XY")
```

Analysis

Question 1: Which physical characteristics of a dam influence its safety hazard?

Since the safety hazard of dams follows an order of severity, which is High (H) > Significant (S) > Low (L), and because it is an ordered factor variable, it was determined that an ordinal logistic regression (OLR) would be the appropriate model to investigate the effects of physical dam features on hazard potential. The predictor variables chosen were a dam's volume, storage capacity, spillway width, drainage area, architectural height, surface area and dam length. 5 different OLR models were tested by selecting a combination of aforementioned variables. However, all-inclusive OLR model returned the lowest AIC value. Therefore, all the variables were included. The 'polr' function used to do OLR does not return p-values. Therefore, p-values were separately calculated by using t-value against standard normal distribution.

```
library("MASS")

## Warning: package 'MASS' was built under R version 3.6.3

##
## Attaching package: 'MASS'

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

model_fit <- polr(hazard ~ volume+nid_storage+spillway_width+
                  drainage_area+nid_height+surface_area+
                  dam_length, data = log_dam_characteristics,
                na.action = na.exclude)
summary(model_fit)

##
## Re-fitting to get Hessian

## Call:
## polr(formula = hazard ~ volume + nid_storage + spillway_width +
##     drainage_area + nid_height + surface_area + dam_length, data =
log_dam_characteristics,
##     na.action = na.exclude)
##
## Coefficients:
##                    Value Std. Error  t value
## volume         -0.106100   0.002337 -45.4008
## nid_storage     0.224487   0.012454  18.0256
## spillway_width  0.036448   0.005448   6.6904
## drainage_area  -0.076400   0.004246 -17.9916
## nid_height      1.122575   0.022317  50.3011
## surface_area    0.138145   0.012400  11.1410
## dam_length      0.001821   0.012520   0.1455
```

```
##
## Intercepts:
##     Value    Std. Error t value
## L|S   5.2973   0.0846    62.5833
## S|H   6.2907   0.0866    72.5998
##
## Residual Deviance: 69200.62
## AIC: 69218.62
## (37142 observations deleted due to missingness)
```

```
summary_table <- coef(summary(model_fit))
```

```
##
## Re-fitting to get Hessian
```

```
pval <- pnorm(abs(summary_table[, "t value"]),lower.tail = FALSE)* 2
summary_table <- cbind(summary_table, "p value" = round(pval,3))
summary_table
```

```
##                      Value  Std. Error     t value p value
## volume         -0.10609950 0.002336953 -45.4007837   0.000
## nid_storage     0.22448692 0.012453797  18.0255803   0.000
## spillway_width  0.03644793 0.005447822   6.6903658   0.000
## drainage_area  -0.07640035 0.004246450 -17.9915805   0.000
## nid_height      1.12257454 0.022317092  50.3011119   0.000
## surface_area    0.13814511 0.012399758  11.1409518   0.000
## dam_length      0.00182133 0.012520079   0.1454727   0.884
## L|S             5.29725401 0.084643268  62.5832881   0.000
## S|H             6.29068783 0.086648773  72.5998493   0.000
```

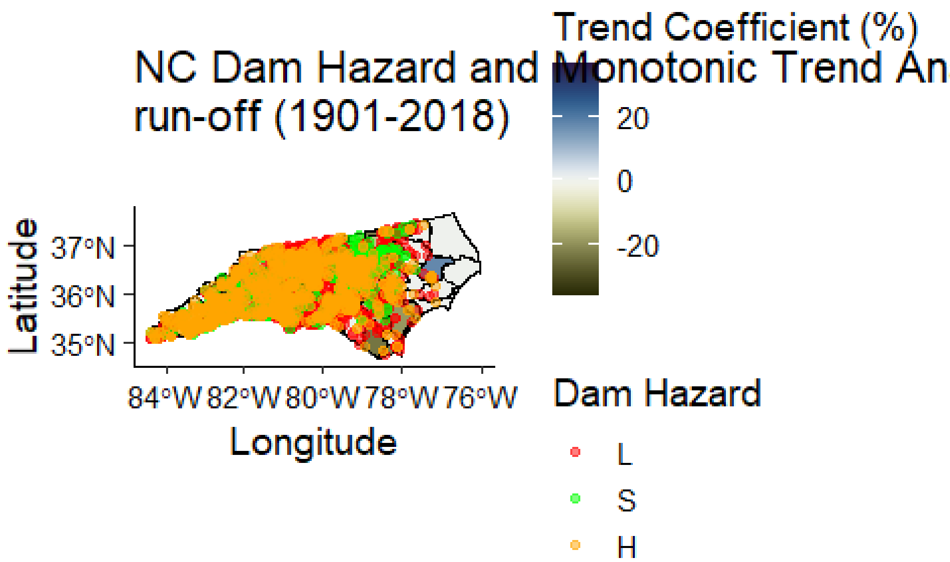
From the p-values of the model results, we can observe that except the spillway width and surface area of a dam, all other variables have a significant effect on the safety hazard of a dam. The OLR model coefficients are exponentiated below for easy interpretation. For instance, we can say that with one unit increase of volume, the log of odds of having a higher safety hazard of dams is affected by a factor of 1.03. Similarly, one unit increase in storage, drainage area, height and length affects the log of odds of increasing dam safety risk by a factor of their respective exponentiated coefficients.

Predictors	Exponentiated OLR coefficients
Volume	1.03
Storage	0.82
Drainage area	0.97
Height	0.39
Length	0.91

Question 2: What is the spatial distribution of North Carolina dams with varying hazard potential in hydrologic units with temporal streamflow trends?

Monotonic trend analysis was performed on annual water run-off (mm/year) in each NC sub-basin since 1901 - 2019 using Mann Kandall Test. The results were then stored in the NC sub-basin shapefile. All the shapefiles prepared during the data wrangling process up until now were layered over each other to get an spatial sense of how dams with various safety hazards are distributed across NC sub-basins of varying temporal trends of annual water run-off. The exact number of different risk classified dams were calculated for sub basins with increasing water run-off, decreasing water run-off or no change in water run-off over about 100 years.

```
# Mann Kendall Test for each NC sub-basin runoff time series
for (i in 1:ncol(huc8_nc_ts)){
  test_results <- mk.test(huc8_nc_ts[,i])
  if (test_results$p.value < 0.05) {
    basin_nc$trend[i] <- test_results$estimates[3]*100
  }
  else {
    basin_nc$trend[i] <- 0
  }
}

# plotting all the shapefile layers
limit <- max(abs(basin_nc$trend)) * c(-1,1)
#pdf(here("Output", "newest.pdf"), width = 11, height = 8.5)
ggplot() +
  geom_sf(data = state_bound_nc, fill = NA) +
  geom_sf(data = basin_nc, aes(fill = trend), color = "black") +
  scale_fill_scico (palette = "broc", limit = limit, direction = -1) +
  geom_sf(data = dam_nc_geo, aes(color = hazard), size = 1.5, alpha = 0.5,
show.legend = "point") +
  scale_color_manual(values = c("red", "green", "orange")) +
  labs(x = 'Longitude', y='Latitude', title = "NC Dam Hazard and Monotonic
Trend Analysis of subbasin (HUC8) annual water \nrun-off (1901-2018)\n",
       color = "Dam Hazard", fill = "Trend Coefficient (%)") +
  theme_bw() + theme(legend.key = element_blank()) +
                       peaceful.theme
```



841 low hazard dams were located in sub basins with no significant trend, 6 in sub basins with a positive trend and 484 in sub basins with a negative trend. 319 significant hazard dams were located in sub basins with no significant trend, 9 in sub basins with a positive trend and 225 in sub basins with a negative trend. Lastly, 776 significant hazard dams were located in sub basins with no significant trend, 17 in sub basins with a positive trend and 514 in sub basins with a negative trend. In general, about 84% of North Carolina has witnessed a decreasing trend in run-off.

```
huc8_dam <- st_intersection(dam_nc_geo, basin_nc)

## although coordinates are longitude/latitude, st_intersection assumes that
they are planar

## Warning: attribute variables are assumed to be spatially constant
throughout all
## geometries

l <- subset(huc8_dam, hazard == 'L')
s <- subset(huc8_dam, hazard == 'S')
h <- subset(huc8_dam, hazard == 'H')

count(l[which(l$trend == 0),]) #841
count(l[which(l$trend > 0),]) #6
count(l[which(l$trend < 0),]) # 484

count(s[which(s$trend == 0),]) #319
count(s[which(s$trend > 0),]) #9
```

```
count(s[which(s$trend < 0),]) #225

count(h[which(h$trend == 0),]) #776
count(h[which(h$trend > 0),]) #17
count(h[which(h$trend < 0),]) #514
```

Summary and Conclusions

This study focused on understanding the physical determinants of dam safety hazard in the US. In addition, it aimed to understand the temporal changes in run-off that can affect the dam operations in NC. It was found that volume, storage, drainage area, height and length have a significant effect on hazard potential of dams which might be intuitive as failure of larger dams can cause greater damages to life and property. Moreover, it was found that most dams, irrespective of their safety hazard, are located in sub-basins witnessing a decreasing trend in annual discharge. While decreasing water availability prevents a dam to operate on its full potential, they also decrease the chances of dam failures due to reduced chances of overtopping. However, one thing to keep in mind is that this project used annual run-off values to perform the trend analysis. Breaking down the analysis to monthly or daily scales could potentially change our understanding of dam overflowing risks due to extreme peak flows during storm events. Similar analysis can be performed for other states as well to focus on high risk dams situated in areas with increasing water flows and develop management strategies to prevent dam failures.