

# PLOT FUN

Ethan Allavarpu (UID: 405287603)

11/30/2020

```
training <- read.csv("training.csv", stringsAsFactors = FALSE)
dim(training)

## [1] 7242 260
any(is.na(training))

## [1] FALSE

var_types <- vapply(training, class, character(1))
names(training) [ -which(var_types %in% c("integer", "numeric"))]

## [1] "PublishedDate"

dates <- training$PublishedDate
head(dates)

## [1] "4/17/2020 10:38" "8/31/2020 9:56"  "8/16/2020 12:15" "8/22/2020 9:00"
## [5] "8/22/2020 14:18" "7/24/2020 21:16"

split_dates <- strsplit(dates, "[/:]")
head(split_dates)

## [[1]]
## [1] "4"      "17"     "2020"   "10"     "38"
##
## [[2]]
## [1] "8"      "31"     "2020"   "9"      "56"
##
## [[3]]
## [1] "8"      "16"     "2020"   "12"     "15"
##
## [[4]]
## [1] "8"      "22"     "2020"   "9"      "00"
##
## [[5]]
## [1] "8"      "22"     "2020"   "14"     "18"
##
## [[6]]
## [1] "7"      "24"     "2020"   "21"     "16"

split_dates <- lapply(split_dates, as.numeric)
years <- function(date) {
  date[3]
}

yrs <- lapply(split_dates, years)
```

```

unique(yrs) # Only 2020

## [[1]]
## [1] 2020

new_time <- function(old_date) {
  old_date <- as.numeric(old_date)
  month <- old_date[1]
  day <- old_date[2]
  hr <- old_date[4]
  minute <- old_date[5]
  month_days <- c(31, 29, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31)
  complete_months <- month - 1
  if (complete_months > 0) {
    complete_month_days <- sum(month_days[1:complete_months])
  } else {
    complete_month_days <- 0
  }
  total_days <- complete_month_days + day
  pre_hrs <- total_days * 24
  total_hrs <- pre_hrs + hr
  pre_min <- total_hrs * 60
  final_time <- pre_min + minute
  final_time
}

processed_dates <- vapply(split_dates, new_time, numeric(1))
training$PublishedDate <- processed_dates

factor_vars <- training[, 248:259]
factor_vars <- data.frame(lapply(factor_vars, factor))
training[, 248:259] <- factor_vars

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

par(mfrow = c(3, 3))
plot(growth_2_6 ~ ., data = training %>% select(-contains("hog")),
     cex = 0.5, pch = 19, col = rgb(0, 0, 0, 0.1))

```











