01 - Disturbance

Cole LaCroix

2025-09-22

```r
#install packages if not already installed
if(!require(terra)) install.packages("terra")
```

```
## Loading required package: terra
```

```
## terra 1.8.5
```

```r
if(!require(dplyr)) install.packages("dplyr")
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:terra':
##
##     intersect, union
```

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

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

```r
if(!require(ggplot2)) install.packages("ggplot2")
```

```
## Loading required package: ggplot2
```

```r
#load required packages
library(terra) #load 'terra' for raster/spatial analysis
library(dplyr) #load 'dplyr' for data manipulation
library(ggplot2) #load 'ggplot2' for plotting
library(ggridges) #load 'ggridges' for ridgeline plots (good for visualizing distributions, e.g., DOY p
library(purrr) #load 'purrr' for functional programming tools (map, map_dfr, etc. useful for batch proc
```

```r
#let's try to read the data in 2022

# Set working directory (students replace with their own folder location)
setwd("/Users/colelacroix/Downloads/h027v011")

#read the disturbance agent raster. Each pixel value represents the type of disturbance agent.
agent_2022 <- rast("./CD_027011_2022_C01/CD_027011_2022_C01_AGENT.tif")

#read the disturbance severity raster. Each pixel value indicates the severity level of the disturbance
severity_2022 <- rast("./CD_027011_2022_C01/CD_027011_2022_C01_SEVERITY.tif")

#read the disturbance time raster. Each pixel value represents the day of year of disturbance occurrenc
time_2022 <- rast("./CD_027011_2022_C01/CD_027011_2022_C01_TIME.tif")
```

```r
#check raster information
agent_2022
```

```
## class       : SpatRaster
## dimensions  : 5000, 5000, 1  (nrow, ncol, nlyr)
## resolution  : 30, 30  (x, y)
## extent      : 1484415, 1634415, 1514805, 1664805  (xmin, xmax, ymin, ymax)
## coord. ref. : AEA        WGS84
## source      : CD_027011_2022_C01_AGENT.tif
## name        : CD_027011_2022_C01_AGENT
```

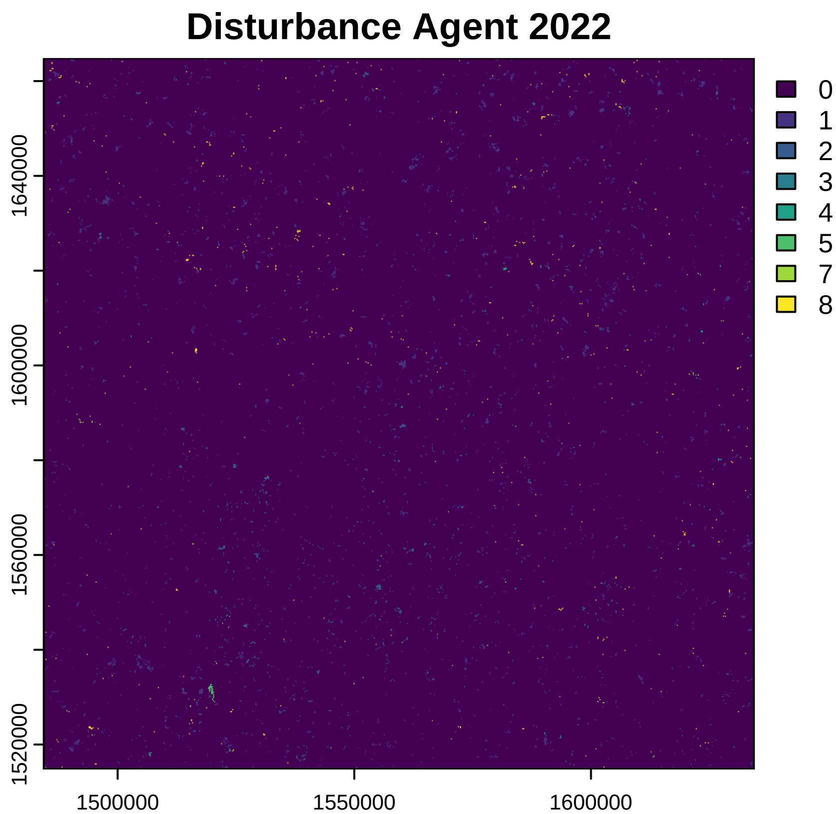
```r
severity_2022
```

```
## class       : SpatRaster
## dimensions  : 5000, 5000, 1  (nrow, ncol, nlyr)
## resolution  : 30, 30  (x, y)
## extent      : 1484415, 1634415, 1514805, 1664805  (xmin, xmax, ymin, ymax)
## coord. ref. : AEA        WGS84
## source      : CD_027011_2022_C01_SEVERITY.tif
## name        : CD_027011_2022_C01_SEVERITY
```

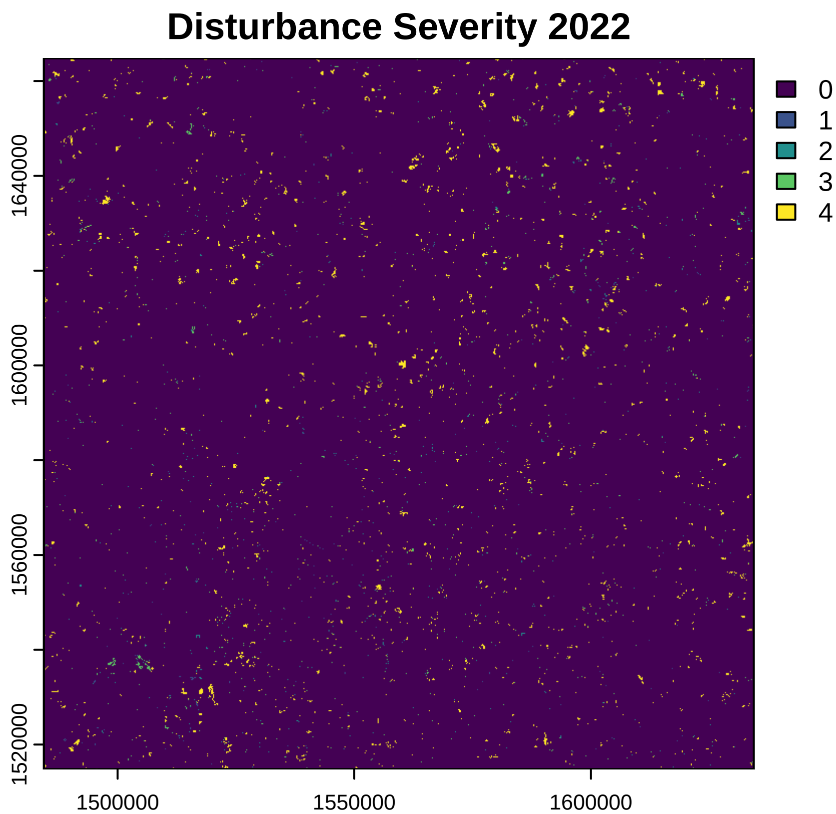
```r
time_2022
```

```
## class       : SpatRaster
## dimensions  : 5000, 5000, 1  (nrow, ncol, nlyr)
## resolution  : 30, 30  (x, y)
## extent      : 1484415, 1634415, 1514805, 1664805  (xmin, xmax, ymin, ymax)
## coord. ref. : AEA        WGS84
## source      : CD_027011_2022_C01_TIME.tif
## name        : CD_027011_2022_C01_TIME
```

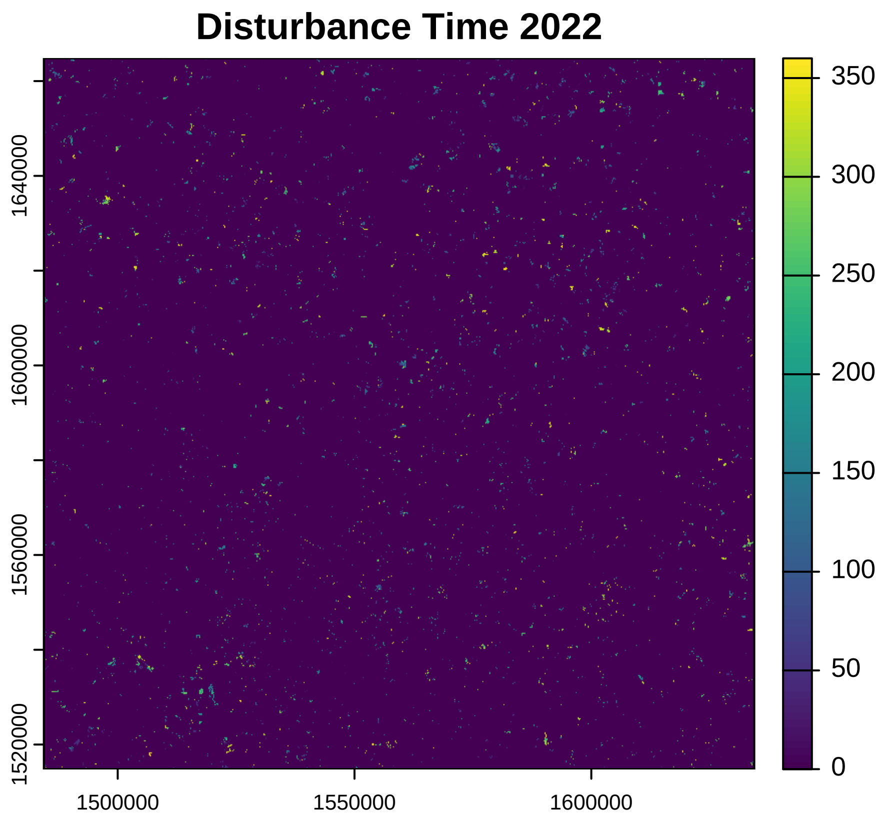
```r
#plot the disturbance agent raster for 2022
plot(agent_2022, main = "Disturbance Agent 2022")
```




```r
#plot the disturbance severity raster for 2022
plot(severity_2022, main = "Disturbance Severity 2022")
```

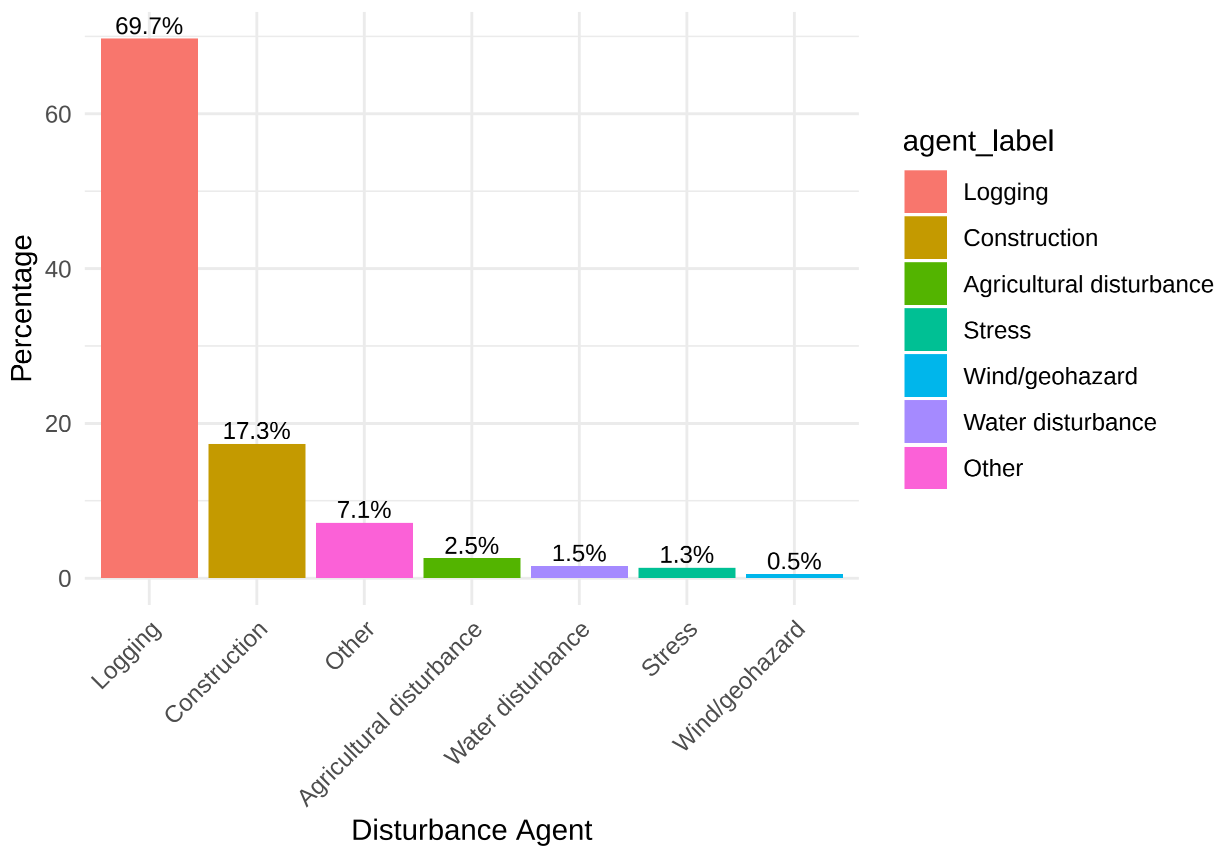



```r
#plot the disturbance timing raster for 2022
plot(time_2022, main = "Disturbance Time 2022")
```




```r
#frequency table
df_agent <- as.data.frame(freq(agent_2022))

#remove "layer" column
df_agent <- df_agent %>% select(-layer)

#rename column for clarity
colnames(df_agent) <- c("agent", "count")

#recode numeric agent values into human-readable labels
agent_labels <- c(
  "1" = "Logging",
  "2" = "Construction",
  "3" = "Agricultural disturbance",
  "4" = "Stress",
  "5" = "Wind/geohazard",
  "6" = "Fire",
  "7" = "Water disturbance",
  "8" = "Other"
)

#add new column with labeled factors (1-8 -> agent names)
df_agent <- df_agent %>%
  mutate(agent_label = factor(agent,
                              levels = 1:8,
                              labels = agent_labels[as.character(1:8)]))
```

```r
#remove Non-break (0) and Filled (255)
df_agent_filtered <- df_agent %>%
  filter(!(agent %in% c(0, 255)))

#compute proportions of each agent
df_agent_filtered <- df_agent_filtered %>%
  mutate(perc = count / sum(count) * 100)

#create a bar plot of disturbance agents by percentage
ggplot(df_agent_filtered, #use the filtered disturbance agent dataframe
       aes(x = reorder(agent_label, -perc), #put disturbance types on x-axis; #reorder so the largest %
           y = perc, #use percentage as bar height
           fill = agent_label)) +  #fill bars with colors by disturbance type
  geom_col() +  #draw bars
  geom_text(aes(label = paste0(round(perc, 1), "%")), #format numbers as percentages
            vjust = -0.3, #move labels slightly above the bar
            size = 3) + #control label text size
  labs(
       x = "Disturbance Agent", #x-axis label
       y = "Percentage") + #y-axis label
  theme_minimal() +  #use a minimal clean theme
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) #rotate x-axis labels 45 degrees
```



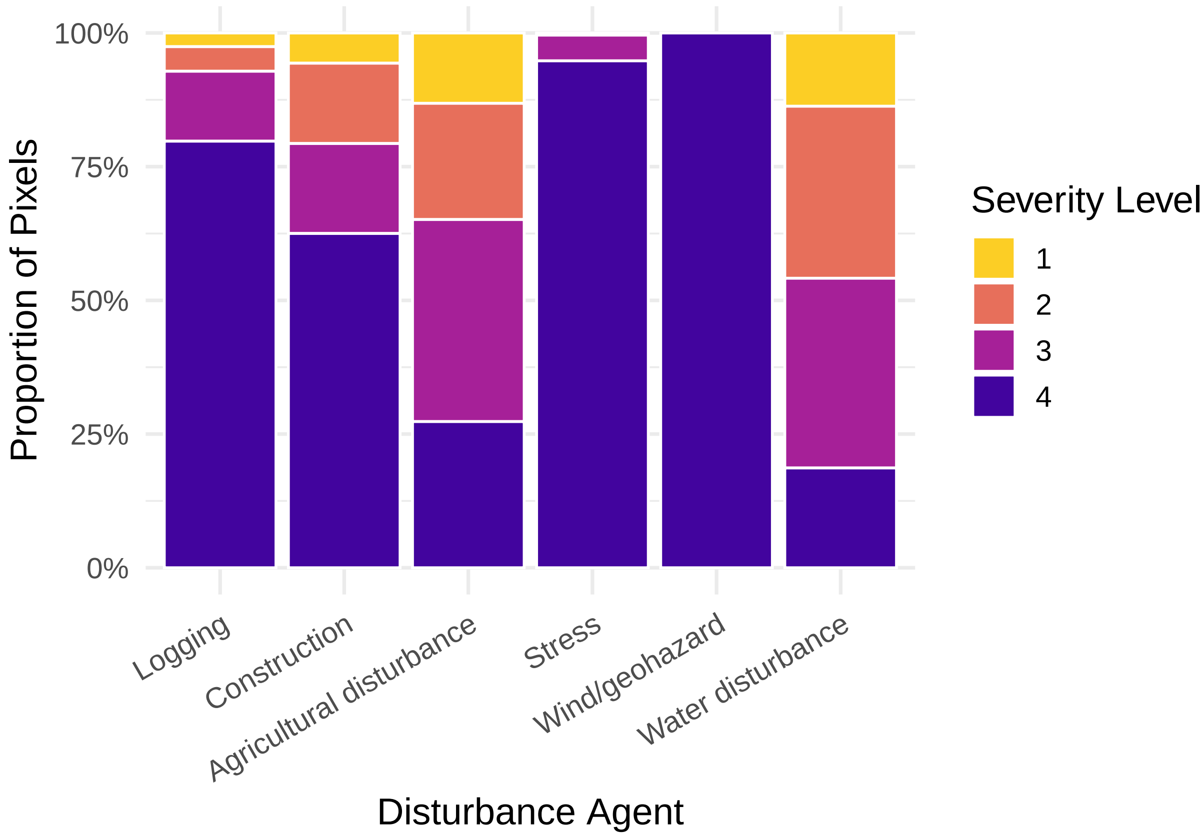
Exercise. Q1. The figure above shows the relative proportions of disturbance agents in 2022. Which disturbance agent appears most common in Raleigh–Durham, and why do you think this type of disturbance might be frequent there?

The figure above indicates that logging is far and away the leading distrubance in the area. I'm assuming that this is because there is a larger timber market here and because an area that can be logged can also be replanted and grown for another harvest 30 years later. Because of this rotation, a larger amount of area can be distrubed every year without it being a conversion of use like construction. Once construction occurs in an area, it is rarely 'reversed'.

```
#extract and clean disturbance raster values (2022)

#create a dataframe with pixel values from the 2022 rasters:
#- agent type
#- severity level
#- day of year (DOY)
vals <- data.frame(
  agent_2022_value = as.vector(values(agent_2022)), #extract agent values
  severity_2022_value = as.vector(values(severity_2022)), #extract severity values
  time_2022_value = as.vector(values(time_2022)) #extract DOY values
  )%>%
  #remove missing (NA) values from all three columns
  filter(!is.na(agent_2022_value),
         !is.na(severity_2022_value),
         !is.na(time_2022_value))

#remove "Non-break" (0) and "Filled" (255)
vals_filtered <- vals %>%
  filter(!(agent_2022_value %in% c(0, 255)))

#add new column with labeled factors (1-8 -> agent names)
vals_filtered <- vals_filtered %>%
  mutate(agent_label = factor(agent_2022_value,
                              levels = 1:8,
                              labels = agent_labels[as.character(1:8)]))

#remove pixels with severity = 0 (no severity assigned)
vals_filtered <- vals_filtered %>%
  filter(!(severity_2022_value %in% 0))
```

```
#plot: 100% stacked bar chart of severity distribution per agent
ggplot(vals_filtered, aes(x = agent_label, fill = factor(severity_2022_value))) +
  geom_bar(position = "fill", color = "white") +   #bar height = proportion; white border for clarity
  scale_y_continuous(labels = scales::percent) +  #convert y-axis to percentages (0-100%)
  #apply a color scale for severity levels (Viridis option C, reversed order)
  scale_fill_viridis_d(option = "C", begin = 0.1, end = 0.9,
                       direction = -1,
                       name = "Severity Level") +
  #add titles and axis labels
  labs(
       x = "Disturbance Agent",
       y = "Proportion of Pixels") +
  #minimal clean theme
  theme_minimal(base_size = 14) +
  theme(
    axis.text.x = element_text(angle = 30, hjust = 1), #tilt x-axis labels
    plot.title = element_text(face = "bold", hjust = 0.5), #bold + centered title
    legend.position = "right" #place legend to the right
```

```
)
```

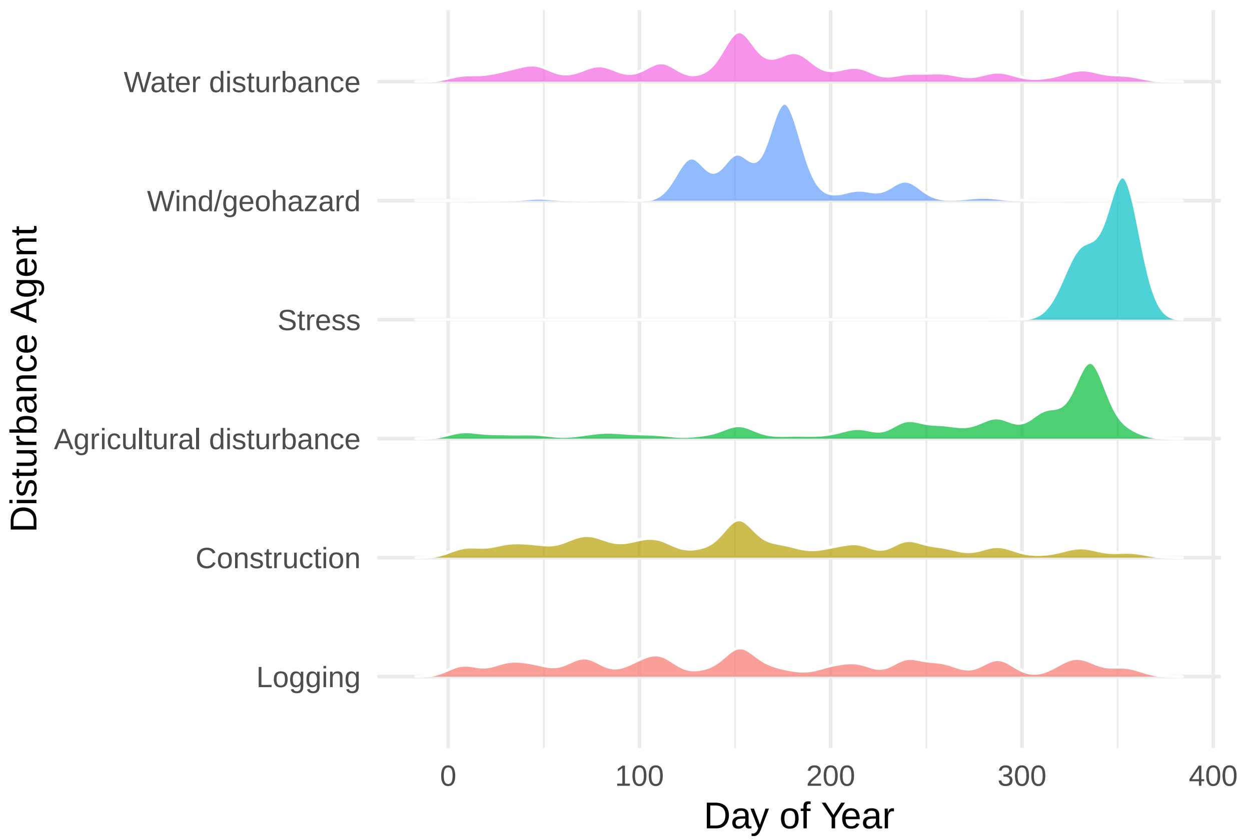


Exercise. Q2-1. The figure shows the severity distribution of different disturbance agents in 2022. Which disturbance agent tends to cause the most severe disturbances, and which shows more low-severity events? What ecological reasons might explain this difference?

It appears that the wind/geohazard category has the highest disturbance intensity level with what appears to be 100% level 4 severity. This is likely because of the type of distruction that earthquakes and tornadoes cause when they do infrequently occur. I would anticipate that the need to disturb the area happens long after the initial event as well because of the need for clean up. Interestingly, water disturbance seems to have the lowest severity rating. I guess this is because of the range of severity water distrubances can occur at. While some distrubances can be very strong (similar to an earth quake), others might simple result in ground saturation for a number of days or weeks.

```r
ggplot(vals_filtered,
       aes(x = time_2022_value, #x-axis: day of year (DOY) of disturbance
           y = agent_label, #y-axis: disturbance agents (categories)
           fill = agent_label)) +  #fill each ridge with a different color for agents
  #draw ridgeline density curves to show distribution of DOY for each agent
  # color = "white" adds a clean border around each ridge
  geom_density_ridges(alpha = 0.7, #alpha = transparency (0.7 = semi-transparent);
                      scale = 1.2, #scale = vertical spacing (1.2 = slight overlap between ridges)
                      color = "white") + #color = "white" adds a clean border around each ridge
  #add axis labels
  labs(x = "Day of Year", y = "Disturbance Agent") +
  theme_minimal(base_size = 14) + #use a clean minimal theme with larger base font size
  theme(legend.position = "none")  #remove legend
```

```
## Picking joint bandwidth of 7.99
```



Exercise. Q2-2. The figure shows the Day of Year (DOY) distribution of different disturbance agents in 2022. Which disturbance agents tend to occur early in the year, and which occur late in the year? What ecological or human processes might explain these seasonal patterns?

Water and wind/geohazard both appear to occur in the early/mid summer time. This likely indicates a season of most intense story severity. After that, stress and agricultural distrubance both occur near the end of the year. Stress might be due to a dryer portion of the year (water disturbance appears to be low for a long period before that). The agricultural disturbance is likely due to the final harvest of the year before the growing season ends.

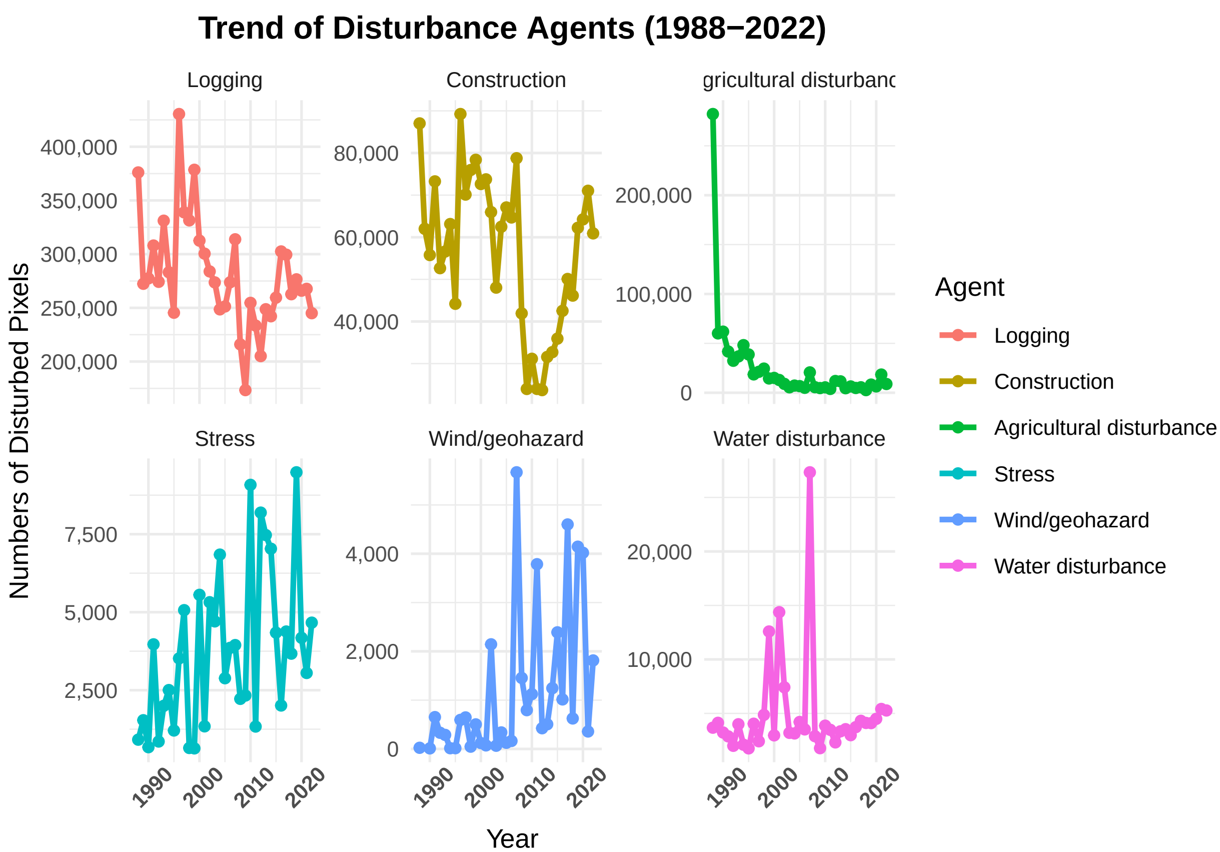
```r
#read disturbance data across multiple years (1988-2022)

#path to the parent directory containing all year subfolders (change to your data folder)
parent_dir <- "/Users/colelacroix/Downloads/h027v011"

#find all immediate subfolders (each corresponds to one year)
year_folders <- list.dirs(parent_dir, recursive = FALSE)

#function to process disturbance data for a single year
process_year <- function(folder) {

  bn   <- basename(folder)  #extract the folder name (e.g., "CD_027011_2004_C01")
  year <- stringr::str_extract(bn, "(?<=_)\\d{4}(?=_)")  #extract the 4-digit year from the folder name
```

```r
  #find the raster files inside the folder:
  #  - agent (disturbance type)
  #  - severity (disturbance severity level)
  #  - time (day of year when disturbance occurred)
  agent_file <- list.files(folder, pattern = "_AGENT.tif$", full.names = TRUE)
  sev_file   <- list.files(folder, pattern = "_SEVERITY.tif$", full.names = TRUE)
  doy_file   <- list.files(folder, pattern = "_TIME.tif$", full.names = TRUE)

  #read each raster using terra
  agent <- terra::rast(agent_file)
  severity   <- terra::rast(sev_file)
  doy   <- terra::rast(doy_file)

  #convert raster values to a dataframe
  df <- data.frame(
      year   = as.integer(year), #year of disturbance
      agent  = as.vector(terra::values(agent)), #disturbance agent values
      severity = as.vector(terra::values(severity)), #severity values
      doy    = as.vector(terra::values(doy)) #day of year values
    ) %>%
    #clean the data:
      dplyr::filter(
        !is.na(agent), #remove missing agent values
        !is.na(severity), #remove missing severity values
        !is.na(doy), #remove missing DOY values
        !(agent %in% c(0, 255)), #exclude "Non-break" (0) and "Filled" (255)
        severity > 0 #exclude pixels with no severity
      )

    return(df)} #return the cleaned dataframe for this year

#apply the function across all year folders and combine results
multi_year_data <- map_dfr(year_folders, process_year)
```

```r
#summarize disturbance pixels by year × agent
trend_df <- multi_year_data %>%
  group_by(year, agent) %>% #group data by year and agent type
  summarise(pixels = n(), .groups = "drop") %>% #count number of disturbed pixels per group
  group_by(year)  #regroup by year for later calculations

#add agent labels (convert numeric codes 1-8 into readable names)
trend_df <- trend_df %>%
  mutate(agent_label = factor(agent, #use 'agent' values
                              levels = 1:8,  #numeric codes 1-8
                              labels = agent_labels[as.character(1:8)])) #assign text labels

#plot: Faceted time series of disturbance trends by agent
ggplot(trend_df, aes(x = year, y = pixels, color = agent_label)) +
  geom_line(size = 1) + #draw trend lines
  geom_point(size = 1.5) +  #add points for each year
  facet_wrap(~ agent_label, #make small multiples (one panel per agent)
             scales = "free_y") + #allow each panel to have its own y-axis scale
  scale_y_continuous(labels = scales::comma) + #format y-axis labels with commas
```

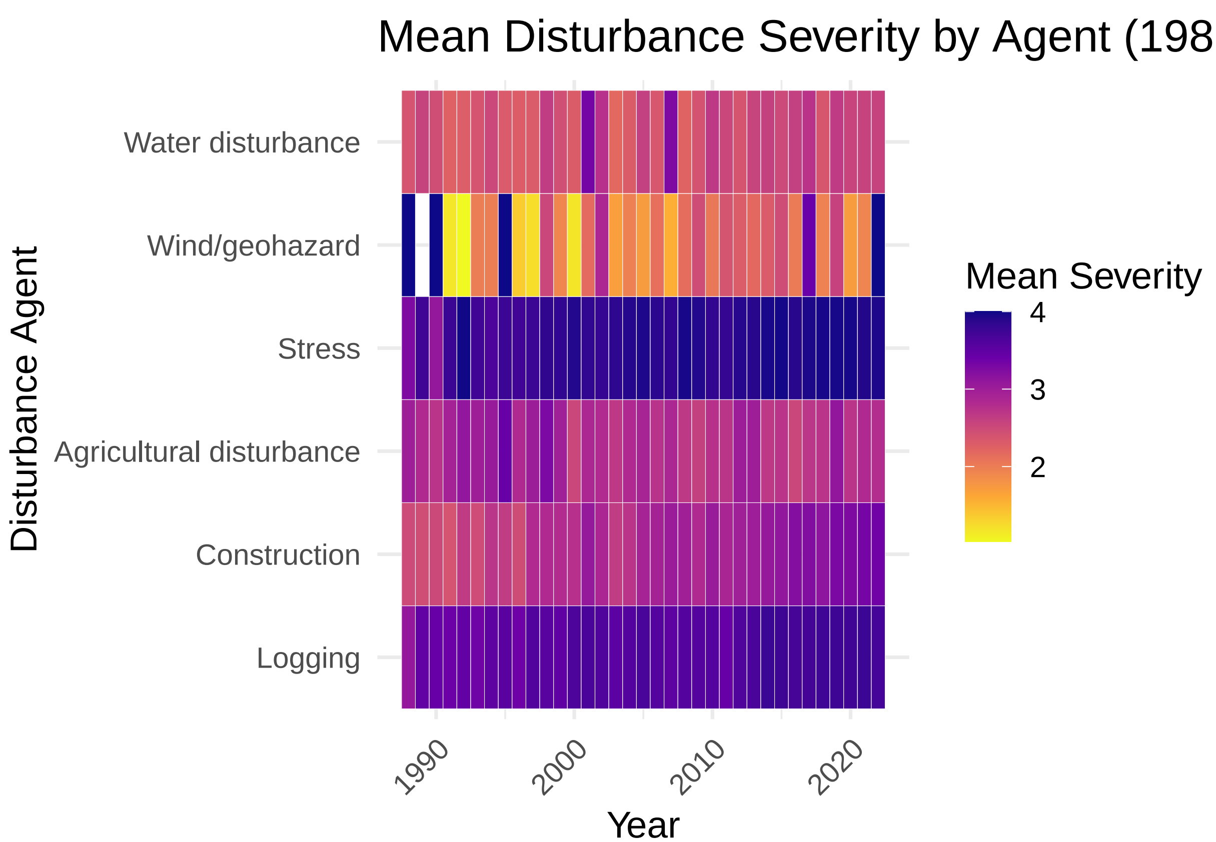
```r
  #add titles and labels
  labs(title = "Trend of Disturbance Agents (1988-2022)",
       x = "Year",
       y = "Numbers of Disturbed Pixels",
       color = "Agent") +
  theme_minimal(base_size = 10) +
  #customize appearance
  theme(
    plot.title = element_text(face = "bold", hjust = 0.5),
    legend.position = "right",
    axis.text.x = element_text(angle = 45, vjust = 0.8, hjust = 0.5, face = "bold")
  )
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```r
#compute mean severity per year × agent
sev_df <- multi_year_data %>%
  group_by(year, agent) %>%
  summarise(mean_severity = mean(severity, na.rm = TRUE), .groups = "drop") #drop grouping after summar

#add agent labels (convert numeric codes 1-8 into readable names)
```

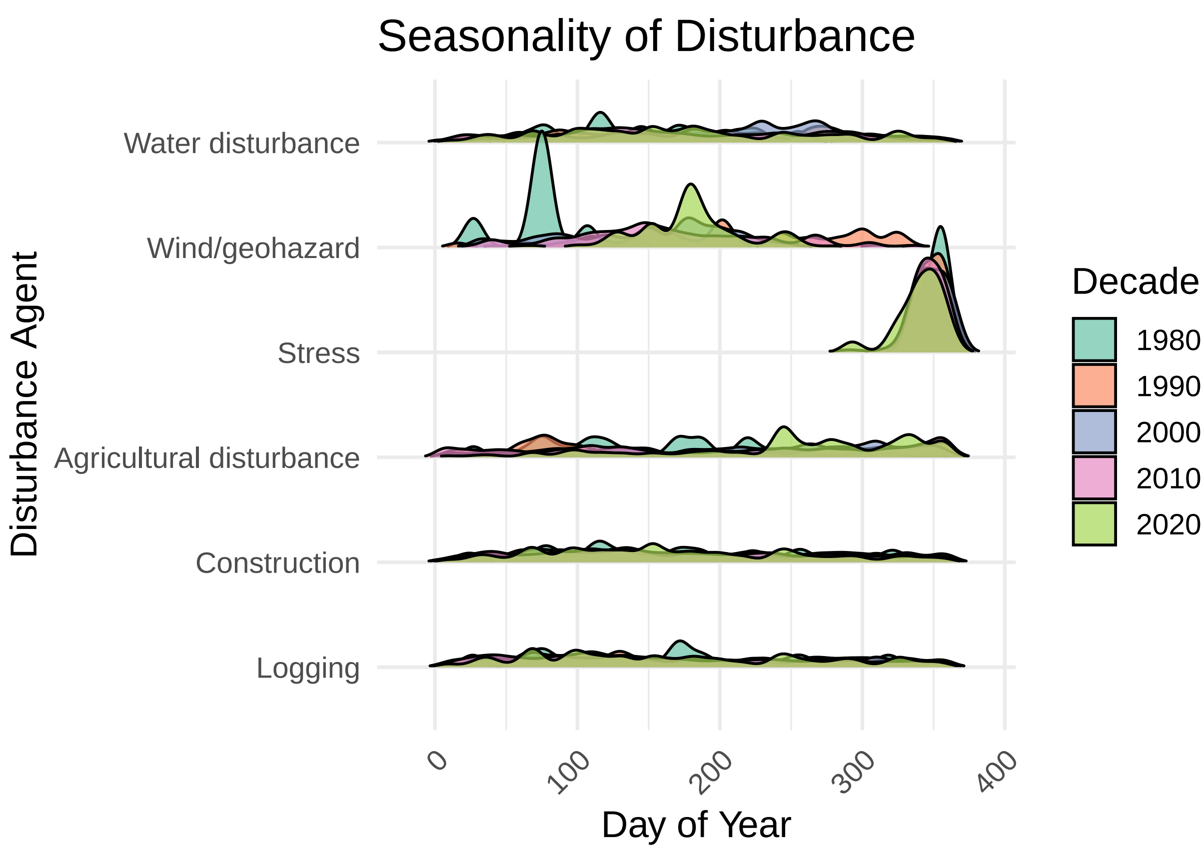
```
sev_df <- sev_df %>%
  mutate(agent_label = factor(agent, #use 'agent' values
                              levels = 1:8,  #numeric codes 1-8
                              labels = agent_labels[as.character(1:8)])) #assign text labels

#plot: Heatmap of mean severity over time by agent
ggplot(sev_df, aes(x = year, y = agent_label, fill = mean_severity)) +
  geom_tile(color = "white") +  #draw heatmap tiles with white borders
  #apply viridis color scale (plasma option, reversed direction)
  scale_fill_viridis_c(option = "plasma", name = "Mean Severity", direction = -1) +
  labs(title = "Mean Disturbance Severity by Agent (1988-2022)",
       x = "Year", y = "Disturbance Agent") +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
#prepare disturbance DOY data for seasonality analysis
doy_df <- multi_year_data %>%
  mutate(decade = floor(year / 10) * 10) %>%  #create new column for decade (1980s, 1990s, etc.)
  filter(!is.na(doy), doy > 0)  #remove missing DOY and invalid values (<= 0)

#add agent labels (convert numeric codes 1-8 into readable names)
doy_df <- doy_df %>%
  mutate(agent_label = factor(agent, #use 'agent' values
                              levels = 1:8,  #numeric codes 1-8
                              labels = agent_labels[as.character(1:8)])) #assign text labels
```

```r
#plot: Ridgeline distribution of DOY by agent × decade
ggplot(doy_df,
       aes(x = doy,
           y = agent_label,
           fill = factor(decade))) +
  #draw ridgeline density plots
  #alpha = transparency, scale = vertical spacing, rel_min_height trims tails
  geom_density_ridges(alpha = 0.7, scale = 1.2, rel_min_height = 0.01) +
  #use a qualitative color palette (Set2) for decades
  scale_fill_brewer(palette = "Set2", name = "Decade") +
  labs(title = "Seasonality of Disturbance",
       x = "Day of Year", y = "Disturbance Agent") +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
## Picking joint bandwidth of 7.08
```



Exercise. Q1. The figure above shows the relative proportions of disturbance agents in 2022. Which disturbance agent appears most common in Raleigh–Durham, and why do you think this type of disturbance might be frequent there?

The figure above indicates that logging is far and away the leading distrubance in the area. I'm assuming that this is because there is a larger timber market here and because an area that can be logged can also be replanted and grown for another harvest 30 years later. Because of this rotation, a larger amount of area can be distrubed every year without it being a conversion of use like construction. Once construction occurs in an area, it is rarely 'reversed'.

Exercise. Q2-1. The figure shows the severity distribution of different disturbance agents in 2022. Which disturbance agent tends to cause the most severe disturbances, and which shows more low-severity events? What ecological reasons might explain this difference?

It appears that the wind/geohazard category has the highest disturbance intensity level with what appears to be 100% level 4 severity. This is likely because of the type of distruction that earthquakes and tornadoes cause when they do infrequently occur. I would anticipate that the need to disturb the area happens long after the initial event as well because of the need for clean up. Interestingly, water disturbance seems to have the lowest severity rating. I guess this is because of the range of severity water distrubances can occur at. While some distrubances can be very strong (similar to an earth quake), others might simple result in ground saturation for a number of days or weeks.

Exercise. Q2-2. The figure shows the Day of Year (DOY) distribution of different disturbance agents in 2022. Which disturbance agents tend to occur early in the year, and which occur late in the year? What ecological or human processes might explain these seasonal patterns?

Water and wind/geohazard both appear to occur in the early/mid summer time. This likely indicates a season of most intense story severity. After that, stress and agricultural distrubance both occur near the end of the year. Stress might be due to a dryer portion of the year (water disturbance appears to be low for a long period before that). The agricultural disturbance is likely due to the final harvest of the year before the growing season ends.

Exercise. Q3. How do trends, severity, and seasonality together help us understand disturbance dynamics?

Trends might help us see if actions we are taking are lessening or perpetuating a distrubance type. Severity might help us see if the efforts we make to mitigate impacts is effective (ie, disincentivizing high-grading in forestry). Seasonality might help us understand the greatest source of the disturbance and whether or not it's human related. Together, we begin to understand the full nature of the disturbances, our ability to address them and forcast them.