Natural Disasters (occuring between 2000-2022)

2024-04-23

Research Questions

- Has there been a change in the prevalence of natural disasters since 2000? Particularly has there been an increase in natural disasters related to climate change i.e. flooding?

- What regions are most effected by natural disasters driven by climate change?

Data Origins

The data set comes from the **Centre for Research on the Epidemiology of Disasters (CRED)**. This organisation records every instance of natural disasters since 1900 within the EM-DAT database. This comprehensive open-source database complies data from various sources (UN agencies, government agencies, research centres, humanitarian organisations, reinsurance companies and world press agencies). For a full list of sources see the EM-DAT website.I chose to download all the information regarding natural disasters between the year 1922 and 2022 . After looking at the data it was clear that the historic record before 2000 was too sparse to stand up against the quality of data recorded by CRED since its inception in 2000. Rather than looking at changes over a century using the historic record I have decided to focus on non-historic entries of natural disasters which have occurred since 2000.

Key Variables To Know

To answer my research questions I need to know where and when different natural disasters occurred. The following variables from the EM-DAT database are relevant to my research questions:

- **Group** - The disaster subgroup:

  - Geophysical - a hazard cause by tectonic plate movement i.e., earthquakes
  - Hydrological - a hazard cause by extreme changes in distribution of water such as flooding or drought
  - Meteorological - a hazard cause by atmospheric change such as extreme temperature

  - Climatological - a hazard cause by sustained periods of hot or cold such as wildfires
  - Biological - a hazard cause by bacteria viruses etc., which effect human health i.e. an epidemic
  - Extra terrestrial - A hazard caused by asteroids, meteoroids, and comets as they pass near-earth

- **Type** - The specific type of disaster i.e., Drought or Earthquake

- **Region** - Region or continent where the disaster occurred

- **Year** - the year the disaster started

- **Climate Change Effect** - whether the natural disaster is a direct effect of climate change, indirect effect of climate change or unrelated to climate change. Categorized based on an EU report.

- Direct Effect - Extreme Temperature, Flood, Storm
- Indirect Effect - Glacial lake outburst flood, Drought, Wildfire, Mass Movement (wet and dry)
- Not related - Volcanic activity, Earthquake, Animal incident, Epidemic, Impact, Infestation

for further explanation of each variable see the codebook provided by the EM-DATA database

Data Preparation

Code

```r
# read in data
data <- here("data","emdat.csv") %>% read_csv()
# nrow(data) #16388


# select columns relevant to research question
my_data <- data %>% select(c(Historic, DisNo., Historic, `Classification Key`,
                             `Disaster Group`, `Disaster Subgroup`, `Disaster Type`,
                             `Disaster Subtype`, ISO, Country, Subregion, Region,
                             `Start Year`, `Start Month`, `Start Day`))

#check
# nrow(my_data) #16388


# tidy the names of the columns so its in a better format

#changing the names so there are no spaces or capital letter
my_data <- my_data %>% rename(id = DisNo.,
                  historic = Historic,
                  classification = `Classification Key`,
                  group = `Disaster Group`,
                  subgroup = `Disaster Subgroup`,
                  type = `Disaster Type`,
                  subtype = `Disaster Subtype`,
                  iso = ISO,
                  country = Country,
                  subregion = Subregion,
                  region = Region,
                  year = `Start Year`,
                  month = `Start Month`,
                  day = `Start Day`)

#check the class is correct for every variable
#str(my_data)


#change the class to numeric for year, month and day variable
my_data <- my_data %>% mutate(year = as.numeric(year),
                              month = as.numeric(month),
                              day = as.numeric(day))
```

```r
#look at the data
#total number of disasters per year
historical <- my_data %>%
        group_by(year) %>%
        summarise(count = n())
# Only look at first 3 rows.
first_few <- historical %>%
        head(3)

# Only look at last 3 rows.
last_few <- my_data %>%
        group_by(year) %>%
        summarise(count = n()) %>%
        tail(3)

# Combine old and new
combined_table <- rbind(first_few, last_few)


kable(combined_table,
      caption = "First 6 rows showing historic data and last 6 rows showing most recent") %>%
  kable_styling()
```

Table 1: First 6 rows showing historic data and last 6 rows showing
most recent

year	count
1922	8
1923	16
1924	9
2020	407
2021	440
2022	436

```r
#remove any Historic data - data from before 2000
updated_data <- my_data %>%
  filter(historic == "No")

# check
#nrow(updated_data) #9505

#check I have not lost any data I shouldn't have
my_data %>% nrow() - my_data %>% filter(historic == "Yes") %>% nrow() #9505
```

```
## [1] 9505
```

Now the data is set up I can create a new column with the climate change information.

```r
#make a new column for climate change effect

# Define vectors for climate change effect
direct_effects <- c("Extreme temperature",
                    "Flood",
                    "Storm")
indirect_effects <- c("Glacial lake outburst flood",
                      "Drought",
                      "Wildfire",
                      "Mass movement (wet)",
                      " Mass movement (dry)")

# Create a new column using a loop to assigns climate change condition
full_data <- updated_data %>%
  mutate(climate_change_effect = case_when(
    updated_data$type %in% direct_effects ~ "Direct effect",
    updated_data$type  %in% indirect_effects ~ "Indirect effect",
    TRUE ~ "Not related"
  ))

#set the climate_change_effect variable to a factor
full_data$climate_change_effect <- factor(full_data$climate_change_effect)

#check
#class(full_data$climate_change_effect)

#set a sensible order to aid plotting
full_data$climate_change_effect <- factor(full_data$climate_change_effect,
                                    levels = c("Direct effect",
                                               "Indirect effect",
                                               "Not related"))
#check
#levels(full_data$climate_change_effect)

#check
#full_data %>% nrow() #9505
kable(head(full_data), caption = "First 6 rows of my Dataset") %>%
  kable_styling()
```

Table 2: First 6 rows of

historic	id	classification	group	subgroup	type	subtype	iso
No	1999-9388-DJI	nat-cli-dro-dro	Natural	Climatological	Drought	Drought	DJI
No	1999-9388-SDN	nat-cli-dro-dro	Natural	Climatological	Drought	Drought	SDN
No	1999-9388-SOM	nat-cli-dro-dro	Natural	Climatological	Drought	Drought	SOM
No	2000-0002-AGO	nat-hyd-flo-riv	Natural	Hydrological	Flood	Riverine flood	AGO
No	2000-0003-BGD	nat-met-ext-col	Natural	Meteorological	Extreme temperature	Cold wave	BGD
No	2000-0008-GTM	nat-geo-vol-ash	Natural	Geophysical	Volcanic activity	Ash fall	GTM

Initial data exploration

Now my data set is ready I can look at basic summaries to see if everything is expected.

```
#view the counts of key variables
#summary of number of disasters grouped by subgroup and type of disaster
kable(full_data %>%
        group_by(subgroup, type) %>%
        summarise(count = n()) %>%
        head(), caption = "First 6 rows showing counts of disasters by subgroup and type") %>%
  kable_styling()
```

Table 3: First 6 rows showing counts of disasters by subgroup and type

subgroup	type	count
Biological	Animal incident	1
Biological	Epidemic	880
Biological	Infestation	29
Climatological	Drought	393
Climatological	Glacial lake outburst flood	3
Climatological	Wildfire	282

```
#summary of number of disasters for each region
kable(full_data %>%
        group_by(region) %>%
        summarise(count = n()) %>%
        head(), caption = "overall counts of disasters per continent") %>%
  kable_styling()
```

Table 4: overall counts of disasters per continent

region	count
Africa	2032
Americas	2180
Asia	3703
Europe	1232
Oceania	358

```
#summary of number of disaster for each year
kable(full_data %>%
        group_by(year, region) %>%
        summarise(count = n()) %>%
        head(), caption = "First 6 rows showing counts of disaster per continent each year") %>%
  kable_styling()
```

Table 5: First 6 rows showing counts of disaster per continent each year

year	region	count
2000	Africa	125
2000	Americas	101
2000	Asia	193
2000	Europe	94
2000	Oceania	12
2001	Africa	116

Visualisations

Question 1

Has there been a change in the prevalence of natural disasters since 2000? Particularly has there been an increase in natural disasters related to climate change i.e. flooding?

```r
#Overall Totals
#line graph which charts changes in prevalence over time split by climate conditions.
# graph where x is years, y is prevalence, split by climate change

# use summarise to provide total natural disaster data
overall_disaster <-
  full_data %>%
  group_by(year) %>%
  summarise(count = n())



plot1 <-
   #plot year on the x axis, add stats function to count frequency
   ggplot(full_data,
          aes(x = year, y = after_stat(count),
              color = climate_change_effect)) +
   #add line to plot the data with the total disasters
   geom_line(data = overall_disaster,
             aes(y = count, color = "Total Disasters"), size = 1) +
   #set the line statistic to count and line size to 1
   geom_line(stat = "count", size = 1) +
   #apply labels for the axis and legend
   labs(x = "Year",
        y = "Prevalence",
        color = "Climate Change Effect",
        caption = "Data source: EM-DAT") +
   #add a title
   ggtitle("Prevalence of Natural Disasters(2000 - 2022)") +
   #specify colours for each line
```

```r
    scale_color_manual(values = c("Total Disasters" = "black",
                                  "Direct effect" = "#E74C3C",
                                  "Indirect effect" = "#F39C12",
                                  "Not related" = "#616A6B")) +
    #set the basic theme for the graph
    theme_minimal() +
    #make adjustments to the theme
    theme(plot.title = element_text(size = 16),   #adjust the size of the title
          axis.text.y = element_text(size = 10),  #adjust the size of x axis scale
          axis.title.x = element_text(size = 12),  #adjust the size of the x axis title
          axis.title.y = element_text(size = 12),  #adjust the size of the y axis title
          strip.text = element_text(size = 12),  #adjust the size of the facet labels
          legend.text = element_text(size = 12),  # adjust the size of legend text
          legend.title = element_text(size = 13),  # adjust the size of legend title
          legend.key.size = unit(3, "mm"),  # adjust the size of legend colours
          panel.background = element_rect(fill = "white", colour = "white"))
#white background

interactive_plot1 <-    #makes the plot interactive
  ggplotly(plot1, tooltip = c( "y", "x")) %>%
  layout(width = NULL, height = NULL) #resize plot

interactive_plot1 <-
  layout(interactive_plot1, annotations = list(    #add a caption
  text = "Data source: EM-DAT",    #write the caption
  x = 1.425,    #set the x co-ordinate for the caption to be displayed
  y = -0.05,  #set the y co-ordinate for the caption to be displayed
  showarrow = FALSE,    #don't include an arrow
  xref = "paper",    #specifies the co-ordinates reference point
  yref = "paper"
))




#save the ggplot as a png file with a white background
ggsave("output/disaster_prevelance.png",
       plot = plot1, bg = "white", width =6, height = 4)

#save the interactive plot as a html file
saveWidget(interactive_plot1,
           file = "output/interactive_disaster_prevelance.html")
```
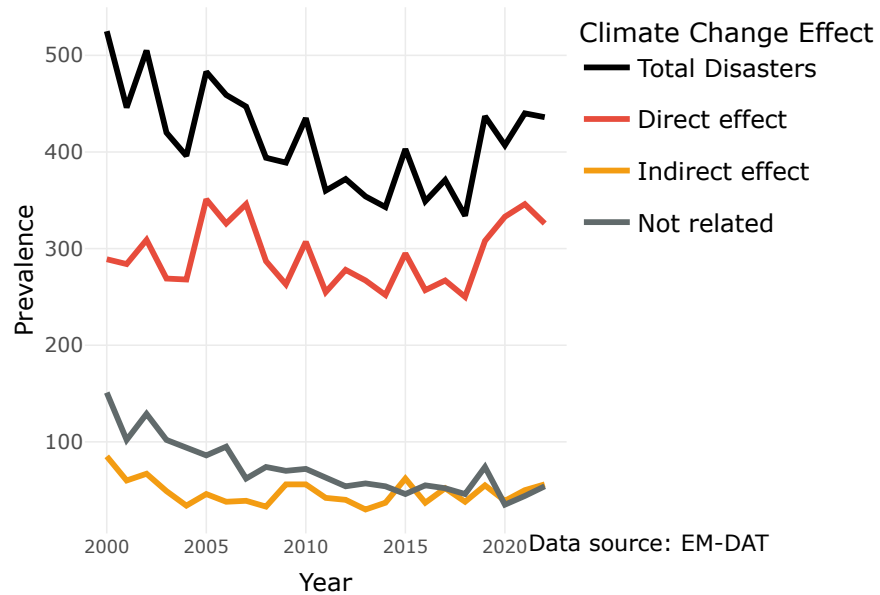
Code

```r
#display the plot
interactive_plot1
```

Prevalence of Natural Disasters(2000 - 2022)

Graph

To use the interactive graph hover over areas of the visualization you are interested in. If you would like to isolate conditions double click on the legend to select what you would like to see. If you would like to zoom in or out you can drag over an area of interest or use the magnifying glass located in the top right corner.

To reset the axis press the home button in the top right corner.

Interpretations   There is no obvious increase in natural disasters globally since 2000, however, the direct effect of climate change is stark. It is clear that natural disasters directly effected by climate change (Extreme Temperature, Flood and Storm) make up the highest number of natural disaster each year. Natural disasters which are an indirect effect of climate change (a consequence of direct natural disasters) such as drought make up a smaller number of total disasters. Natural disasters not related to climate change again account for a small proportion of total disasters since 2000.

Question 2 (Main Visualisation)

What regions are most effected by natural disasters driven by climate change?

```
#create a stacked bar chart with with the counts of disaster for each year.
#split by climate change effect and type

#sets what the hover text shows for each data point in the interactive plot
hover_text <- paste(
  "Type: ", full_data$type,
  "<br>Group: ", full_data$subgroup,
  "<br>Year: ", full_data$year
  )

plot2 <-
  #plot year on the x axis. plot the count on the y axis for climate change.
  #specify that the hover_text will be shown when the graph is interactive
  ggplot(full_data,
         aes(x = year,  y = after_stat(count),
             fill = climate_change_effect,
             text = hover_text)) +
  geom_bar(position = "stack", #plot the data as a stacked bar graph.
           alpha = 0.8,  # set the transparency to 0.8,
           colour = "white", #the line colour to white
           size = 0.1) + # thickness of the line to 0.1
  #use facet_wrap to split the graph up by region (continent)
  facet_wrap(~ region, scales = "free_x", nrow = 2) + #fix x scales and 2 graph per row.
  #specify the colours for each climate change category
  scale_fill_manual(values = c("#E74C3C", "#F39C12", "#616A6B")) +
  #add labels
  labs(title = "Prevelance of Natural Disasters per Continent",
       caption = "Data source: EM-DAT",
       x = " ",
       y = "Number of Disasters",
       fill = "Effect of Climate Change") +
  #apply basic theme
  theme_minimal() +
  #adjustments to the theme
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size = 10),  # Add a slant
        axis.text.y = element_text(size = 10),  # Adjust the size of y axis scale
        axis.title.x = element_text(size = 12),  # Adjust the size of the x axis title
```

```r
        axis.title.y = element_text(size = 12),   # Adjust the size of the y axis title
        strip.text = element_text(size = 12),   # Adjust the size of continent titles
        legend.text = element_text(size = 12),   # Adjust the size of legend text
        legend.title = element_text(size = 13),   # Adjust the size of legend title
        legend.key.size = unit(3, "mm"),   # Adjust the size of legend squares
        plot.title = element_text(size = 16),    # Adjust the size of plot title
        panel.background = element_rect(fill = "white", colour = "white"))
# set background to white




interactive_plot2<-    #make plot interactive.
  ggplotly(plot2, tooltip = c( "y", "text")) %>%  #set tooltip
layout(width = NULL, height = NULL) #set graph size




# Add caption
interactive_plot2 <-
  layout(interactive_plot2, annotations = list(    #add a caption
  text = "Data source: EM-DAT",    #caption
  x = 1.25,     #x co-ordinate
  y = -0.05,    # y co-ordinate
  showarrow = FALSE,    #don't show an arrow
  xref = "paper",    #set co-ordinate to reference entire plot
  yref = "paper"
))


#save ggplot as a pgn with a white background
ggsave("output/disaster_per_region.png",
       plot = plot2, bg = "white", width =6, height = 4)

#save interactive plot to a html file
saveWidget(interactive_plot2,
           file = "output/interactive_disaster_per_region.html")
```
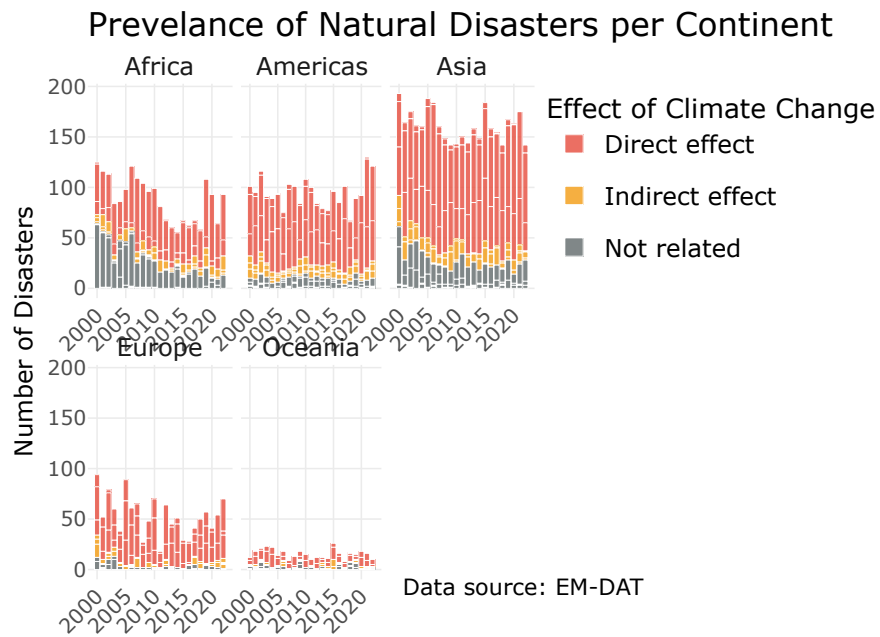
Code

```r
interactive_plot2    #show plot
```

Prevelance of Natural Disasters per Continent

Graph

To use the interactive graph hover over areas of the visualization you are interested in. If you would like to isolate conditions double click on the legend to select what you would like to see. If you would like to zoom in or out you can drag over an area of interest or use the magnifying glass located in the top right corner.

To reset the axis press the home button in the top right corner.

Interpretations When visualising natural disasters per continents we see the global trends repeating themselves. The highest proportion of total natural disaster in each continent are those which are a direct consequence of climate change. My second visualisation has more power to see which continent are worse effected by natural disasters. It is clear that Asia has experience a huge number of natural disasters which are both direct effects and indirect effects of climate change. Oceania on the other hand has experience relatively few disasters, however, this may be reflective of the comparatively smaller land mass Oceania represents. The data displayed in this graph for the continent of Africa suggests that natural disaster not effected by climate change have decreased although it is unlikely that this is a meaningful observation.

Reflection

I am pleased with the graphs I have made. I played around with lots of different graphs and ways to display my data. The dataset I originally downloaded was large with multiple interesting variables. Ultimately, I chose to focus on the variables which had the most complete data. I think adding the interactive element to my graph really enhanced my visualisation and allowed me to include more information. The key information I wanted to translate was the prevalence of disasters over time split by the effect of climate change. I felt, however, that this was not enough and was interested in creating a main visualisation that not only showed the split per continent but also the specific disaster type which actually makes up the climate change categories. I hope that this has translated well on different devices. An issue with my second visualisation in particular is that it can render rather small on some screens making it difficult to see events which were rare. Hopefully the zoom function on the interactive graph will help with readability.

If I had more time, I would delve deeper into the locations of the disaster. I did attempt at first to create a world map which showed where the disaster occurred, however, the information for some of the small islands which make up Oceania were difficult to see on a world map. It would be interesting, however, to have an individual map for each continent which shows where the disasters struck and when. It would have also been interesting to look at some of the different variables on the original dataset such as cost of disaster both to lives and the economy.

An issue with the EM-DAT database was firstly the incomplete data for the historical data points. This was initially disappointing as I would have loved to track the trends in natural disasters over the last century, however, the historical record is still being complied and will inevitably be more incomplete the further back you go. Another issue I noticed was only land disasters are recorded, Tsunamis for example were absent. The data entries which are analysed in this project are not therefore an exhaustive list of disasters.

One of the most important skills I have developed during this project is to keep making notes as I am working so I always know what the code does. I have also found that keeping a record of code which doesn't work when I am trying to figure out a problem is still helpful.

References

Raw data available from www.emdat.be. The dataset is maintained by The Centre for Research on the Epidemiology of Disasters (CRED), UCLouvain, Brussels, Belgium.

My full repository is at https://github.com/angharad00/natural_disaster_project.