AIM 5001 - Data Acquisition & Management

Jesus Olivera

Final Project - Exploring Climate Data

Introduction

Climate change poses a great risk to humanity.

Climate change poses a greater risk to Small Island Developing States (SIDS) and Least Developed Countries (LDCs).

Between 1970 and 2019, 79% of natural disasters worldwide involved weather and climate related hazards.

Scientists have high confidence that the global temperatures will continue to rise for decades to come, largely due to the greenhouse gases produced by human activity.

Research Questions

How are global temperatures affecting the weather and safety of our planet? This question can be broken down into the following sub questions:

What are the trends in global surface temperature?

How do global temperature trends affect glaciers?

How do global temperature trends affect the number of deaths by natural disasters?

Surface Temperature Anomaly

Data Source: Our World in Data

Description: The data is based on the GISTEMP analysis from the NASA Goddard Institute for Space Studies. The temperature data is measured in degrees Celsius and it is relative to the 1951 – 1980 global average temperature. The data was collected via CSV from its source.

Importing the Data:

```
read = pd.read_csv("https://raw.githubusercontent.com/j
read.head(3)
```

	Entity	Code	Year	Surface temperature anomaly
0	Afghanistan	AFG	1881	0.28
1	Afghanistan	AFG	1882	-0.35
2	Afghanistan	AFG	1883	-0.29

Data Preparation

The data was properly prepared using functions to locate and deal with null values.

We discovered that Micronesia was missing its country code and we used the “replace” function to replace the null values for the country’s code.

```
# Finding null values
is_nan = read.isnull()
row_with_nan = is_nan.any(axis=1)
nan_rows = read[row_with_nan]

# Finding unique countries with "NaN" country code
nan_rows.Entity.unique()
```

```
array(['Micronesia'], dtype=object)
```

```
# Replacing "NaN" values for Micronesia 3 digit ISO code: FSM

read['Code'] = read['Code'].replace(np.NaN, 'FSM')

# Reviewing if "NaN" values were replaced
read.isnull().sum()
```

```
Entity                          0
Code                            0
Year                            0
Surface temperature anomaly     0
dtype: int64
```

EDA - Surface Temp Anomaly

Statistical Summary

```
read.describe().round(2)
```

	Year	Surface temperature anomaly
count	23127.00	23127.00
mean	1955.10	0.14
std	37.82	0.67
min	1880.00	-4.70
25%	1925.00	-0.26
50%	1958.00	0.11
75%	1987.00	0.52
max	2017.00	3.41

```
print(read.shape)
```

```
(23127, 4)
```

Distribution

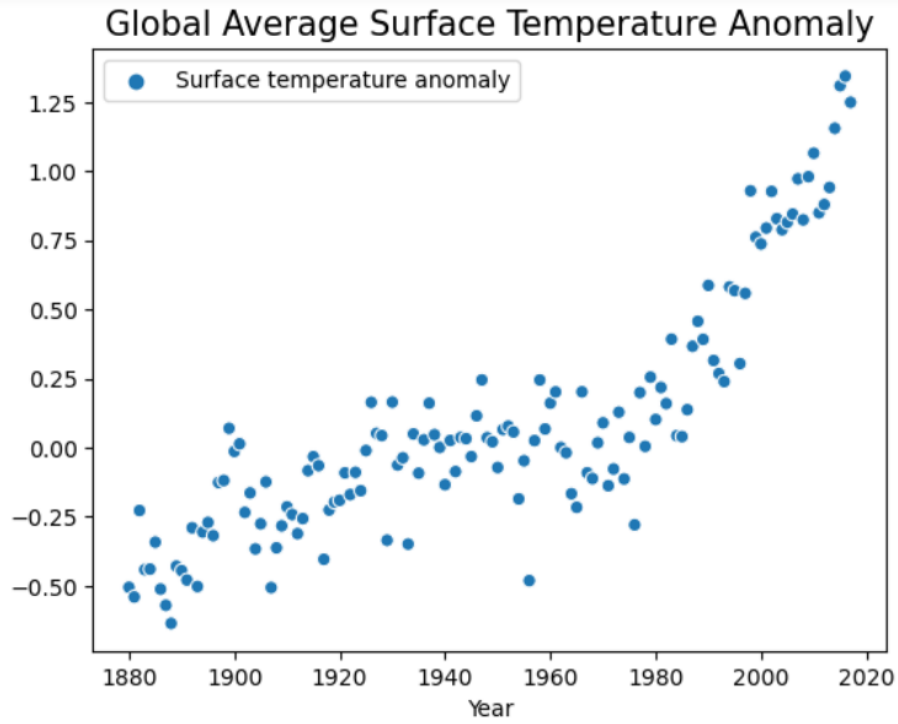
```
read['Entity'].value_counts().sort_values()
```

```
Solomon Islands          38
Oman                     44
Sao Tome and Principe    53
Timor                    60
Vanuatu                  61
                        ...
North Macedonia         138
Algeria                 138
New Zealand             138
Norway                  138
Sweden                  138
Name: Entity, Length: 198, dtype: int64
```

EDA - Surface Temp Anomaly

```
temp_country = read.groupby(['Year']).mean()

sns.scatterplot(data = temp_country)
plt.title("Global Average Surface Temperature Anomaly", size=15)
plt.show()
```



What does surface temperature mean?
Comparable to how warm or cool the earth feels if you were to touch it.

What is an anomaly?
An instance when conditions depart from the average conditions of a particular place at a given time of the year.

Graph Conclusions:
This graph shows a sharp increase in the global surface temperature over time. There is a particularly steep increase from 1980 onwards.

EDA - Deaths from Natural Disasters

Data Source: Our World in Data

Description: Deaths caused by natural disasters by country and year. The data set contains the name of the country, the country code, the year and the total deaths due to natural disasters. The data needed will be collected from this source via csv.

Importing the Data:

```
read1 = pd.read_csv("https://raw.githubusercontent.com/jaynuel
deaths_by_country = read1.rename(columns={"Deaths - Exposure t

# Printing Results
deaths_by_country.head(3)
```

	Entity	Code	Year	Total_Deaths
0	Afghanistan	AFG	1990	0.000000
1	Afghanistan	AFG	1991	1349.999434
2	Afghanistan	AFG	1992	614.000083

Data Preparation

The data was properly prepared using functions to locate and deal with null values.

We discovered that the data was not only organized by country, but it was also grouped by regions , and the null values on the code column was due to this grouping.

As part of the data transformation process we decided to drop all grouped values and only work with individual country values.

```
nan_values = deaths_by_country[deaths_by_country['Code'].isna()]
nan_values.Entity.unique()
```

```
array(['Andean Latin America', 'Australasia', 'Caribbean', 'Central Asia',
       'Central Europe',
       'Central Europe, Eastern Europe, and Central Asia',
       'Central Latin America', 'Central Sub-Saharan Africa', 'East Asia',
       'Eastern Europe', 'Eastern Sub-Saharan Africa', 'England',
       'High SDI', 'High-income', 'High-income Asia Pacific',
       'High-middle SDI', 'Latin America and Caribbean', 'Low SDI',
       'Low-middle SDI', 'Middle SDI', 'North Africa and Middle East',
       'North America', 'Northern Ireland', 'Oceania', 'Scotland',
       'South Asia', 'Southeast Asia',
       'Southeast Asia, East Asia, and Oceania', 'Southern Latin America',
       'Southern Sub-Saharan Africa', 'Sub-Saharan Africa',
       'Tropical Latin America', 'Wales', 'Western Europe',
       'Western Sub-Saharan Africa'], dtype=object)
```

```
deaths_by_country = deaths_by_country.dropna().reset_index(drop=True)
```

```
deaths_by_country.drop(deaths_by_country[deaths_by_country['Entity']== 'World'].index, inplace=T
```

EDA - Deaths from Natural Disasters

Statistical Summary

```
print(deaths_by_country.describe().round(2))
```

```
Deaths by Country
             Year  Total_Deaths
count     5460.00       5460.00
mean      2003.50        261.24
std          8.08       4990.05
min       1990.00          0.00
25%       1996.75          0.00
50%       2003.50          0.00
75%       2010.25         12.00
max       2017.00     222658.31
```

```
print(deaths_by_country.shape)
```

```
(6468, 4)
```

Distribution

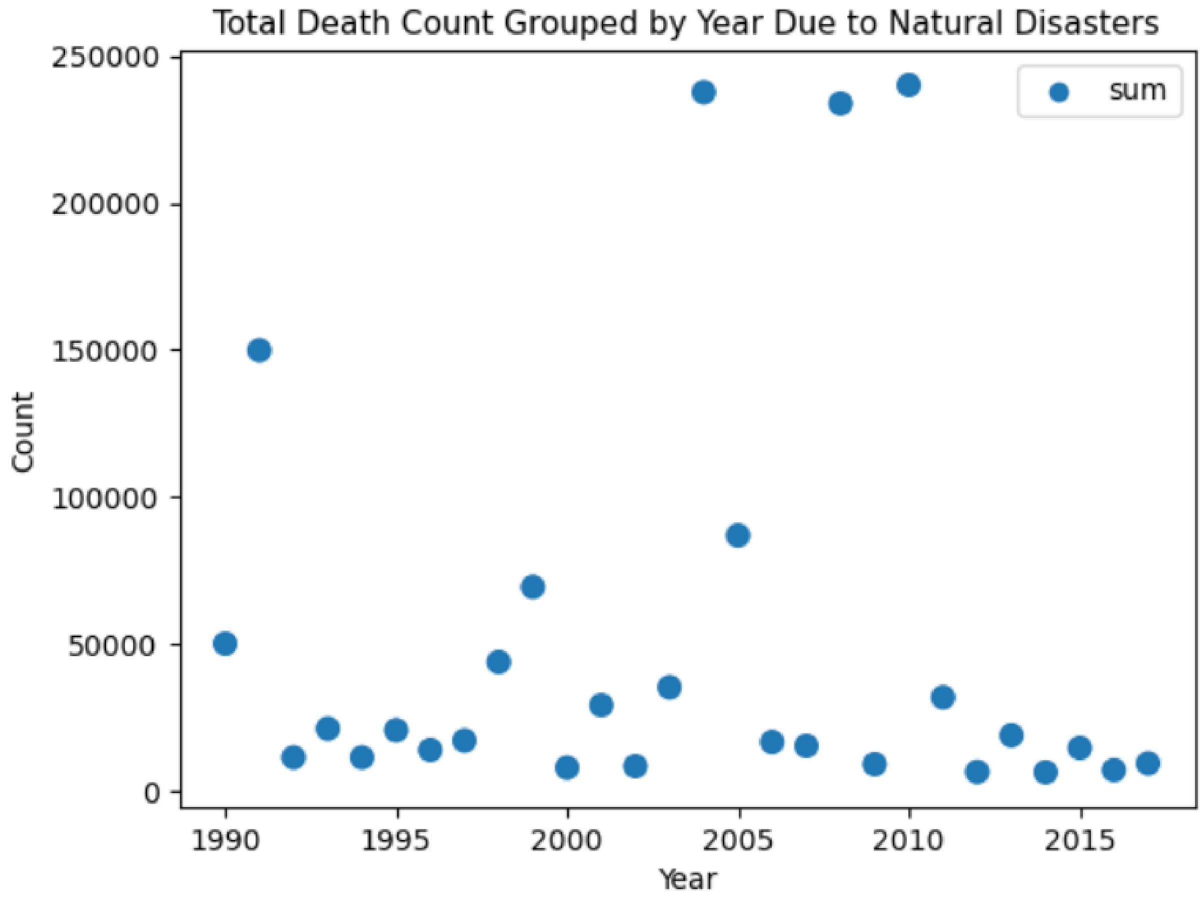
```
print(deaths_by_country['Entity'].value_counts().sort_values().head(5))
```

```
Deaths by Country
Laos         28
Malaysia     28
Sweden       28
Namibia      28
Bahrain      28
```

```
deaths_by_country.groupby('Year').Entity.agg(['count']).head(5)
```

Year	count
1990	195
1991	195
1992	195
1993	195
1994	195

EDA - Deaths from Natural Disasters




```
total_by_country = deaths_by_country.groupby('Year').Total_Deaths.agg(['sum']).round()
```

EDA - Understanding the Outliers

```
outliers_deaths = total_by_country.nlargest(4, 'sum')
```

Year	sum
2010	240034.0
2004	237601.0
2008	233796.0
1991	149891.0

```
total_by_year.nlargest(4, 'sum')
```

Entity	sum
Haiti	227488.0
Indonesia	185304.0
Bangladesh	154153.0
Myanmar	139682.0

Findings:

Haiti: 2010 Earthquake

```
     Entity Code  Year  Total_Deaths
2148  Haiti  HTI  2010   222658.3065
```

Indonesia: 2004 Indian Ocean Earthquake & Tsunami

```
         Entity Code  Year  Total_Deaths
2282  Indonesia  IDN  2004   166041.1199
```

Bangladesh: 1991 Cyclone

```
         Entity Code  Year  Total_Deaths
393  Bangladesh  BGD  1991   139252.1216
```

Myanmar: 2008 Cyclone Nargis

```
       Entity Code  Year  Total_Deaths
3350  Myanmar  MMR  2008   138366.4963
```

EDA - Glacier Mass Balance

Data Source: DataHub io

Description: Average cumulative mass balance of reference glaciers worldwide. Average cumulative balance of “reference” glaciers worldwide from 1945-2014 from US EPA and World Glacier Monitoring Service (WGMS). The values represent the average of all the glaciers that were measured. Negative values indicate a net loss of ice and snow compared with the base year of 1945. The data needed was collected from this source via Web Scraping.

Importing the Data:

```
r = requests.get('https://pkgstore.datahub.io/core/glacier-mas
data = StringIO(r)
glacier = pd.read_csv(data, sep=",")

# Printing dataframe
glacier.head(3)
```

	Year	Mean cumulative mass balance	Number of observations
0	1945	0.00	NaN
1	1946	-1.13	1.0
2	1947	-3.19	1.0

EDA - Glacier Mass Balance

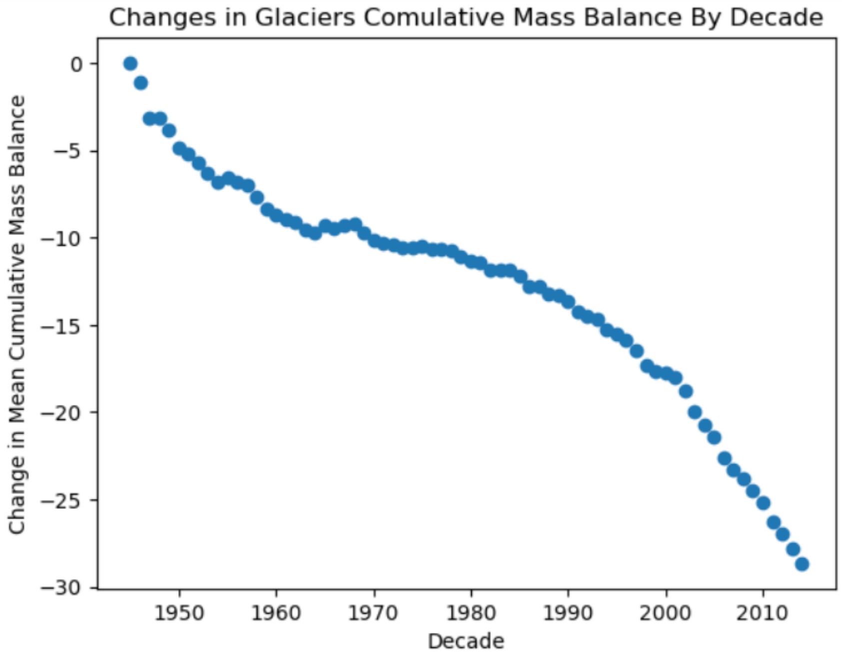
Statistical Summary

```
glacier.describe().round(2)
```

	Year	Mean cumulative mass balance	Number of observations
count	70.00	70.00	70.00
mean	1979.50	-12.84	27.36
std	20.35	6.58	13.07
min	1945.00	-28.65	0.00
25%	1962.25	-16.34	20.50
50%	1979.50	-11.22	35.50
75%	1996.75	-9.14	37.00
max	2014.00	0.00	37.00

```
print(glacier.shape)
```

```
(70, 3)
```



EDA - Climate Projections

Data Source: Climate Impact Lab

Web Address: http://www.impactlab.org/map/#usmeas=absolute&usyear=1981-2010&gmeas=absolute&gyear=1986-2005&tab=global

Description: The projections were aggregated to regional estimates by first transforming the daily min, average, or maximum temperatures at the grid scale, then aggregating to regions using a weighted average. From this source we were be able to extract necessary data regarding climate projections by region and year. The data needed was collected from this source as a CSV format.

Importing the Data:

```
r3= requests.get('https://raw.githubusercontent.com/jaynuel/AIM-5
r4 = r3.text
data = StringIO(r4)
projections = pd.read_csv(data, sep=",")

# Printing dataframe
projections.head(3)
```

	Country	Average	5th	50th	95th	5th.1	50th.1	95th.1	5th.2	50th.2	95th.2
0	Aruba	80.6	81.1	82.1	83	81.3	83.2	84.9	83.6	85.7	90.5
1	Afghanistan	74	74.9	76.7	79.9	76.5	78.7	83.2	80.5	84.1	95.6
2	Angola	68.5	69.7	70.7	71.9	70.9	72.3	74.6	73.6	75.7	83.2

Data Preparation

The data was properly prepared using functions to locate and deal with null values.

As part of the data transformation process we created 4 dataframes from the original dataframe and rename the columns for a more intuitive analysis.

```
projections[projections['Country'].isna()]
```

	Country	Average	5th	50th	95th	5th.1	50th.1	95th.1	5th.2	50th.2	95th.2
252	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
253	NaN	1986-2005	2020-2039	NaN	NaN	2040-2059	NaN	NaN	2080-2099	NaN	NaN
254	NaN	Average	5th	50th	95th	5th	50th	95th	5th	50th	95th

```
projections = projections.dropna()
print(projections.isnull().sum())
```

```
Country     0
Average     0
5th         0
50th        0
95th        0
5th.1       0
50th.1      0
95th.1      0
5th.2       0
50th.2      0
95th.2      0
dtype: int64
```

```
AVG_1986_to_2005 = projections[["Country","Average"]].rename(columns={"Average":"AVG_1986_to_200
PROJECTION_2020_to_2039 = projections[["Country","5th","50th","95th"]].reset_index(drop=True)
PROJECTION_2040_to_2059 = projections[["Country","5th.1","50th.1","95th.1"]].rename(columns={"5t
PROJECTION_2080_to_2099 = projections[["Country","5th.2","50th.2","95th.2"]].rename(columns={"5t
```

EDA - Climate Projections

```
Actual AVG from 1986 to 2005
         Country AVG_1986_to_2005
count        504              504
unique       252              238
top         Laos             79.2
freq           2                8

2020 to 2039
         Country   5th  50th  95th
count        504   504   504   504
unique       252   261   263   273
top         Laos  80.1  81.8  82.9
freq           2     7     7     8

2040 to 2059
         Country   5th  50th 95th
count        504   504   504  504
unique       252   272   272  273
top         Laos  81.3  83.2   85
freq           2     9     7    8

2080 to 2099
         Country   5th  50th  95th
count        504   504   504   504
unique       252   279   281   301
top         Laos  83.6  83.2  89.4
freq           2     7     7     7
```

```python
print(projections.shape)
```

```
(509, 11)
```

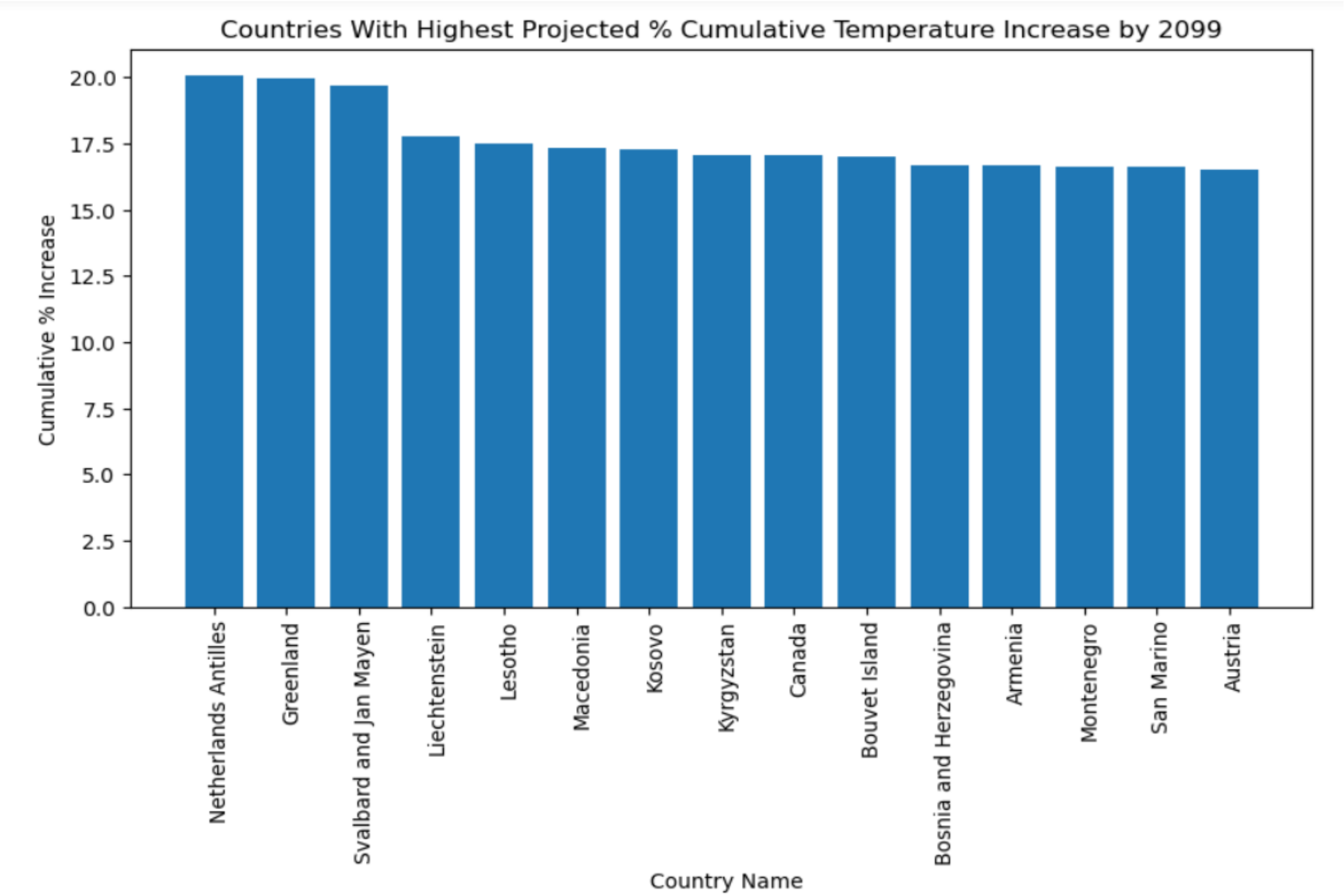
```python
pc = percentage_change.pct_change(axis='columns')*100
percentage_change = pd.concat([countries,pc],axis =1, sort=False)
```

```python
top_15 = percentage_change.nlargest(15,'Projected % Cumulative Temperature Increase by 2099')

# Printing Results
top_15.head(5)
```

	Country	1986-2005	2020-2039	2040-2059	2080-2099	Projected % Cumulative Temperature Increase by 2099
6	Netherlands Antilles	NaN	6.386861	3.945111	9.735974	20.067946
92	Greenland	NaN	5.882353	4.629630	9.439528	19.951511
199	Svalbard and Jan Mayen	NaN	4.289544	5.398458	10.000000	19.688002
131	Liechtenstein	NaN	5.585586	3.242321	8.925620	17.753526

EDA - Climate Projections



EDA - Global Glacier Reductions

Data Source: National Snow & Ice Data Center

Web Address: https://nsidc.org/glims/glaciermelt

Description: Dataset contains the percentage of total Area and the percentage of contribution to volume change from 1961 to 2003. This analysis is focused on mountain glaciers and smaller ice caps, which have a total area at least 785x103 km2. Although they make up only 4% of the total land ice area, they may have contributed to as much as 30% of sea level change in the 20th century due to rapid ice volume reduction connected with global warming. The data needed was collected from this source via Web Scraping.

Importing the Data:

```
tables2 = pd.read_html('https://nsidc.org/glims/glaciermelt')

# Retrieving Data
g2 = tables2[0]

# Printing Results
g2.head(3)
```

	Largest Contributors to Global Water Cycle and Sea Level Rise		
	Region	Percentage of Total Area	Percentage of Contribution to Volume Change 1961-2003
0	Arctic	52.7	31.5
1	High Mountain Asia	19.4	23.9
2	Alaska and Coastal Mountains	15.0	23.0

Data Preparation

The data did not contained any null value.

```
print(g2.isnull().sum())
```

```
Largest Contributors to Global Water Cycle and Sea Level Rise  Region
0
                                                               Percentage of Total Area
0
                                                               Percentage of Contribution to
Volume Change  1961-2003    0
dtype: int64
```

EDA

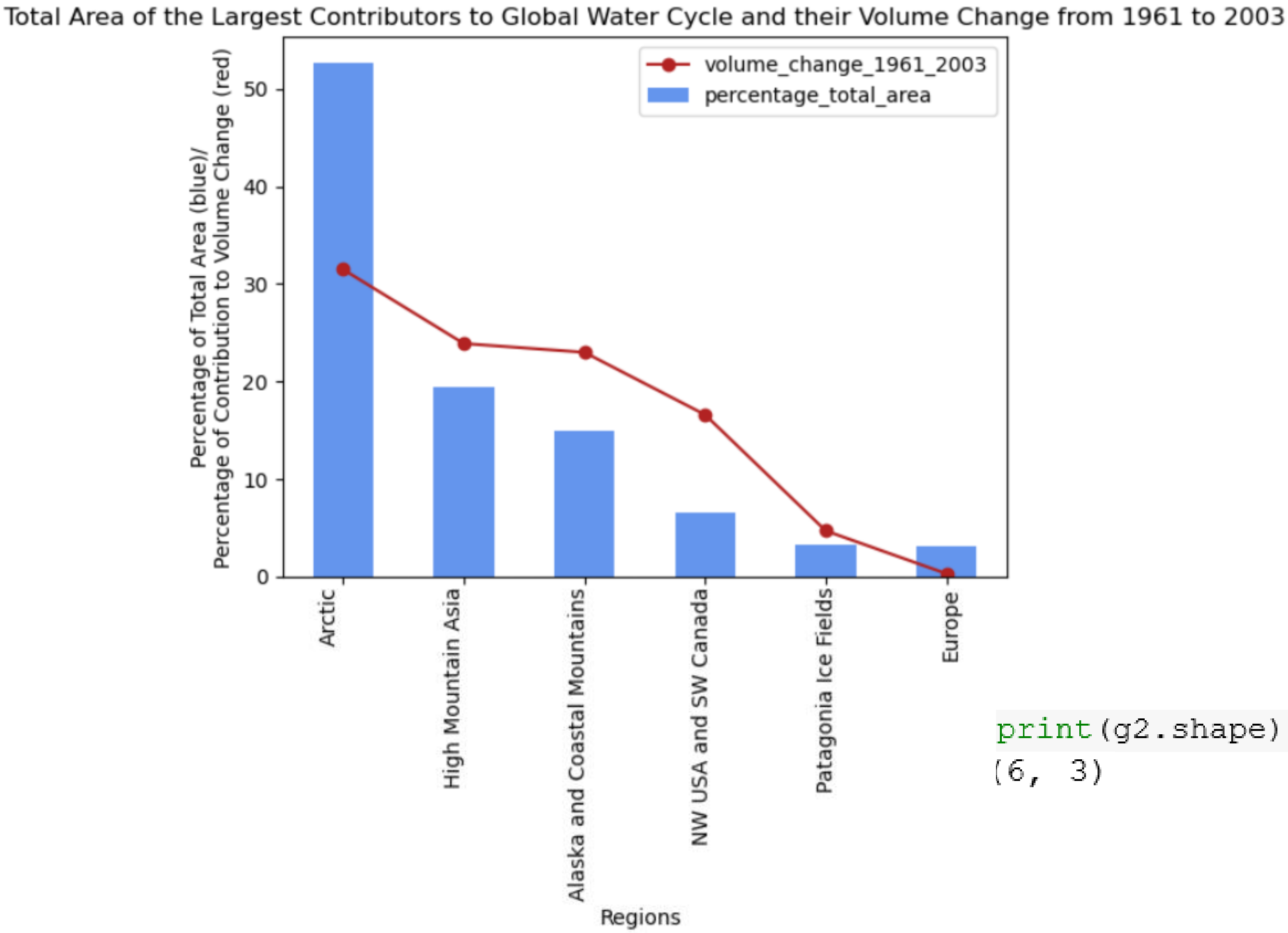
Results:

Percentage total surface area decreased significantly

Decrease in the volume change contributing to sea level

These two variables are related so this makes sense

Follows along with the Glacier Mass Balance EDA




```
print(g2.shape)
(6, 3)
```

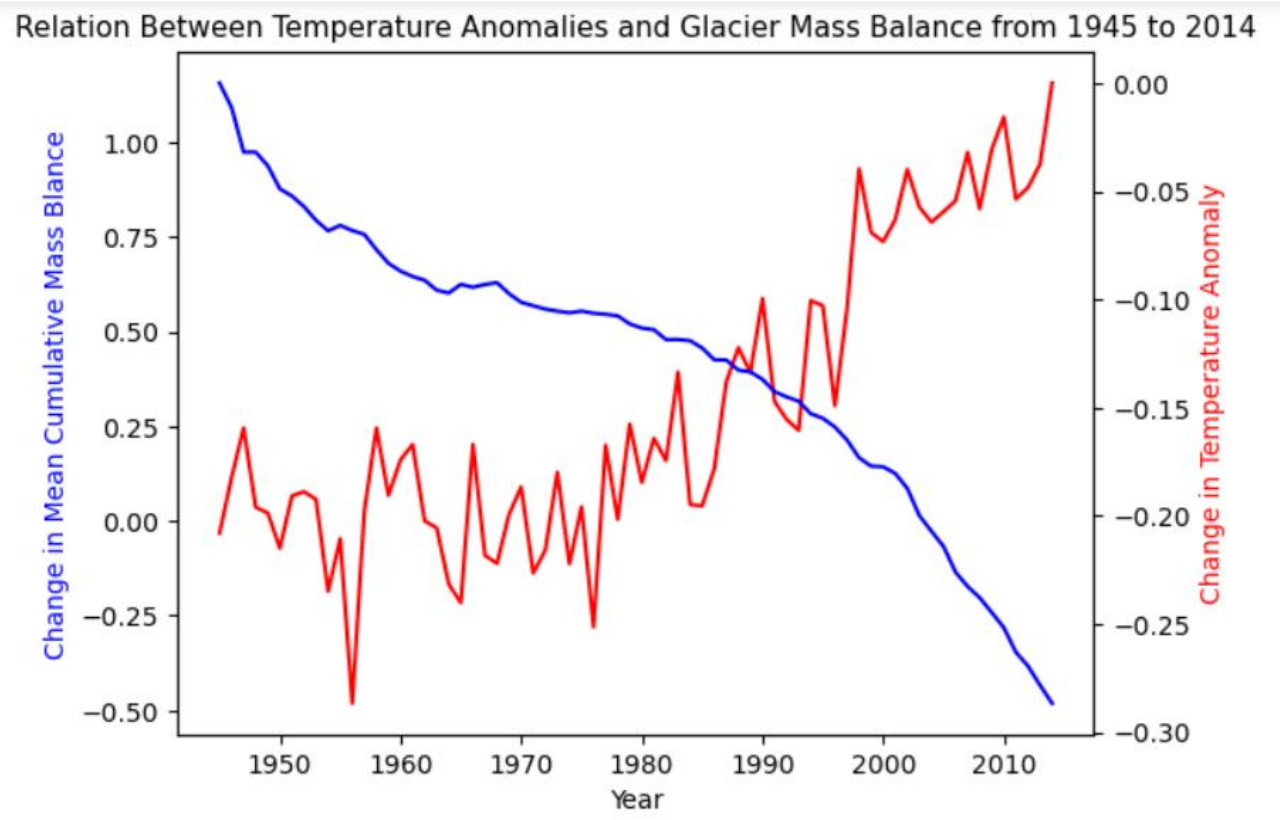
EDA: Surface Temperature Anomaly & Glaciers Mean Cumulative Mass Balance

Results:

Strong negative relationship between the average surface temperature anomaly and glaciers mass balance

Results show that as temperature increases, glacier mass balance decreases

Significant changes happen around the same time (~1990)



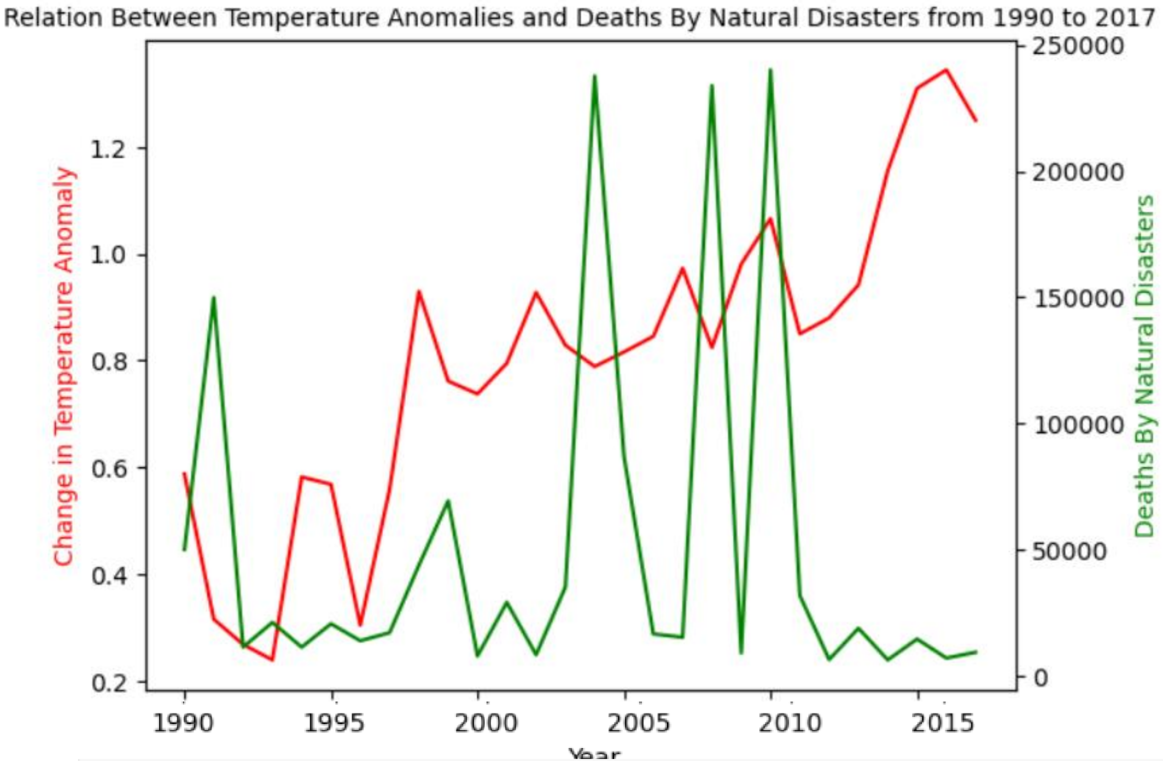

```
corr = relationship['temp_mean'].corr(relationship['glacier_mean'])
corr
-0.8613173446158233
```

Surface temperature anomaly and deaths caused by natural disasters

Results:

Weak negative relationship between global surface temperature anomalies and deaths caused by naturals disasters

As temperature raises, global deaths by naturals disasters are relatively steady with some peaks around areas of sharp temperature change




```
corr1 = relation['temp_mean'].corr(relation['total_deaths'])
corr1
```

```
-0.03557230315248272
```

GeoPandas

Python package to make working with geospatial data in Python easier

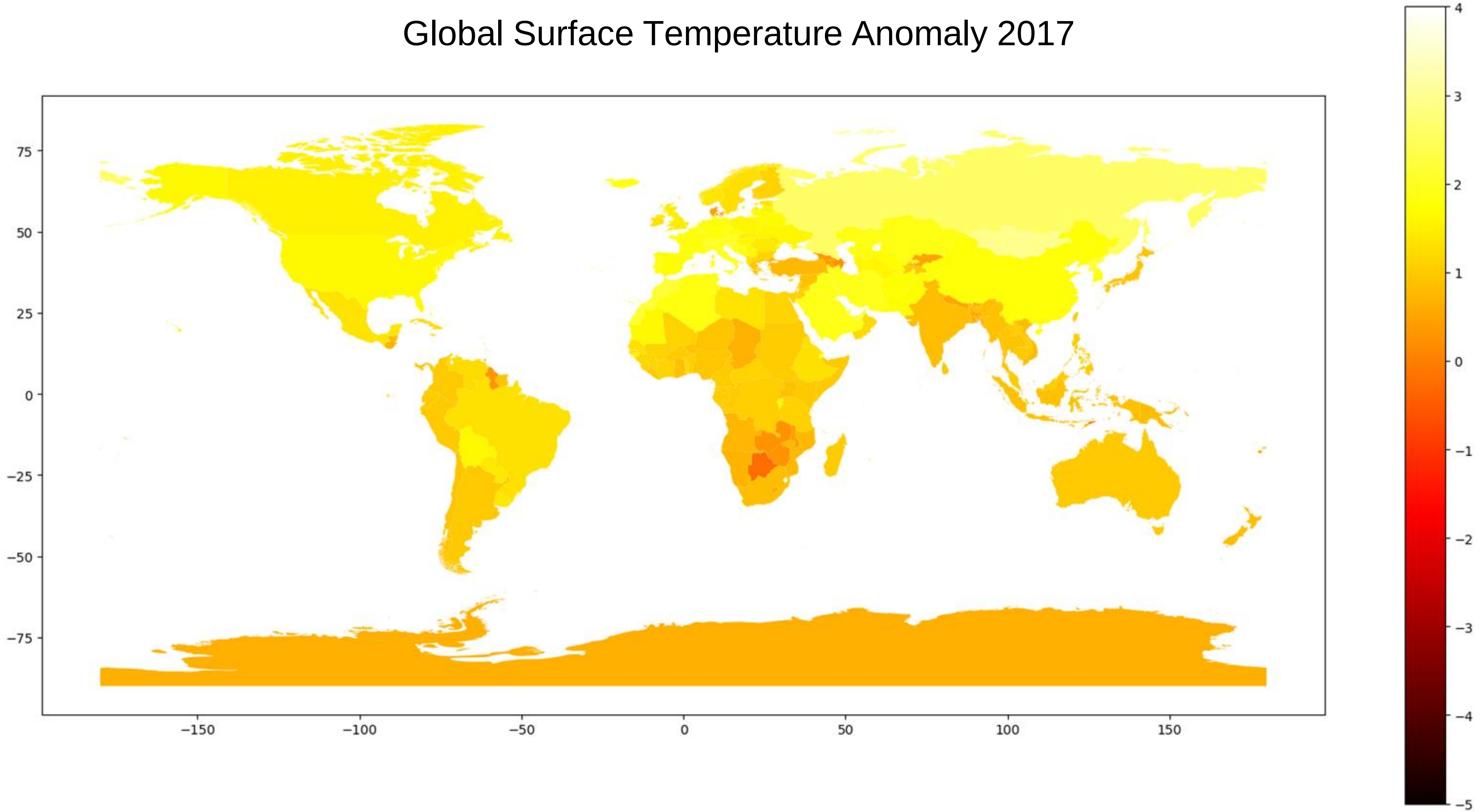
Extension of pandas to allow spatial operation on geometric types

shapefile

	NAME	ISO3	ISO2	COUNTRY	CONTINENT	EU	SQKM	geometry
0	Åland	ALA	AX	Åland	Europe	0	1.243719e+03	MULTIPOLYGON (((20.99567 60.64228, 20.99261 60...
1	Afghanistan	AFG	AF	Afghanistan	Asia	0	6.413834e+05	POLYGON ((73.27336 36.88856, 73.26195 36.89025...
2	Albania	ALB	AL	Albania	Europe	0	2.848611e+04	MULTIPOLYGON (((20.98057 40.85522, 20.98096 40...
3	Algeria	DZA	DZ	Algeria	Africa	0	2.316559e+06	MULTIPOLYGON (((-8.67387 27.29807, -8.67172 27...
4	American Samoa	ASM	AS	American Samoa	Oceania	0	2.110151e+02	MULTIPOLYGON (((-171.07492 -11.06860, -171.078...
...	...	...	...	...	...	...	...	...
260	Western Sahara	ESH	EH	Western Sahara	Africa	0	2.668299e+05	MULTIPOLYGON (((-17.05185 20.77416, -17.05445 ...
261	Yemen	YEM	YE	Yemen	Asia	0	4.198999e+05	MULTIPOLYGON (((53.10706 16.65440, 53.09564 16...
262	Zambia	ZMB	ZM	Zambia	Africa	0	7.513153e+05	POLYGON ((30.41826 -15.61757, 30.41293 -15.622...
263	Zimbabwe	ZWE	ZW	Zimbabwe	Africa	0	3.906484e+05	POLYGON ((28.84701 -21.74224, 28.83875 -21.738...
264	Paracel Islands	P--	P-	Paracel Islands	Asia	0	5.343721e+00	MULTIPOLYGON (((112.34131 16.92700, 112.34064 ...

265 rows × 8 columns

GeoPandas Maps



Results and Conclusions

1. **What are the trends in global surface temperature?**

Steady increase from 1986 to 2017

Steep incline from 1980 onwards

1. **How do global temperature trends affect glacier?**

Strong negative correlation

Fits with what we expect from the increasing surface temperature

The Arctic region is the most affected by this

1. **How do global temperature trends affect the number of deaths by natural disasters?**

Weak negative correlation

Few sharp peaks