Assignment -2

GLOBAL LANDSLIDES REPORT

Vijeta Tulsiyan_s3398979

Table of Contents

Introduction

The global landslide data from NASA gives details of landslide occurrences across countries/continents. It also provides latitude and longitudinal location of these incidents to aid researchers in understanding landslide susceptibility in various regions.

In this report, I have analysed the data to find which continents / latitude-longitude are more susceptible to landslides. Also, find the main triggers of landslides in each continent. The analysis also involves finding the frequency of types of landslides, and study of correlation between injuries and fatalities for each landslide.

Methods

The data for this report is sourced from data.nasa.gov. There are 9471 observations and 28 variables. Data is covered mainly for the period 2007 – 2016. However, 48 observations are recorded between period 1900-2004.

The software/tools used for data analysis are Excel Pivot table, and SAS Studio. The statistical test such as 'correlation' is used to find correlation between variables. Other statistical test includes finding frequencies of occurrences of various events/triggers. Also, visualisation packages such as SGplot, scatter plot etc and histogram is used for data analysis.

Results

An analysis of the percentage of the landslides in each continent is done in terms of landslide size. The results show that most of the landslides were concentrated on Asia (AS) and North America (NA). Of the total landslides captured in the data, 48.7% occurred in Asia and 34.6% in North America.

If we compare the size of landslides that took place, the most common is the medium landslide at 69.4% of the total landslides for the given data, followed by small size at 22.2% of the total landslides.

Frequency Percent	Table of conti by land_sz					
	land_sz					
conti	Large	Medium	Small	Very_Large	unknown	Total
AF	32 0.37	145 1.68	19 0.22	5 0.06	0 0.00	201 2.33
AS	382 4.42	3389 39.20	376 4.35	47 0.54	13 0.15	4207 48.66
EU	28 0.32	342 3.96	132 1.53	3 0.03	4 0.05	509 5.89
NA	94 1.09	1574 18.21	1301 15.05	15 0.17	4 0.05	2988 34.56
OC	17 0.20	195 2.26	69 0.80	1 0.01	1 0.01	283 3.27
SA	62 0.72	356 4.12	26 0.30	13 0.15	0 0.00	457 5.29
Total	615 7.11	6001 69.42	1923 22.24	84 0.97	22 0.25	8645 100.00
Frequency Missing = 826						

Figure 1: number of landslides that have occurred in each continent for each landslide size category and their percentage of the total

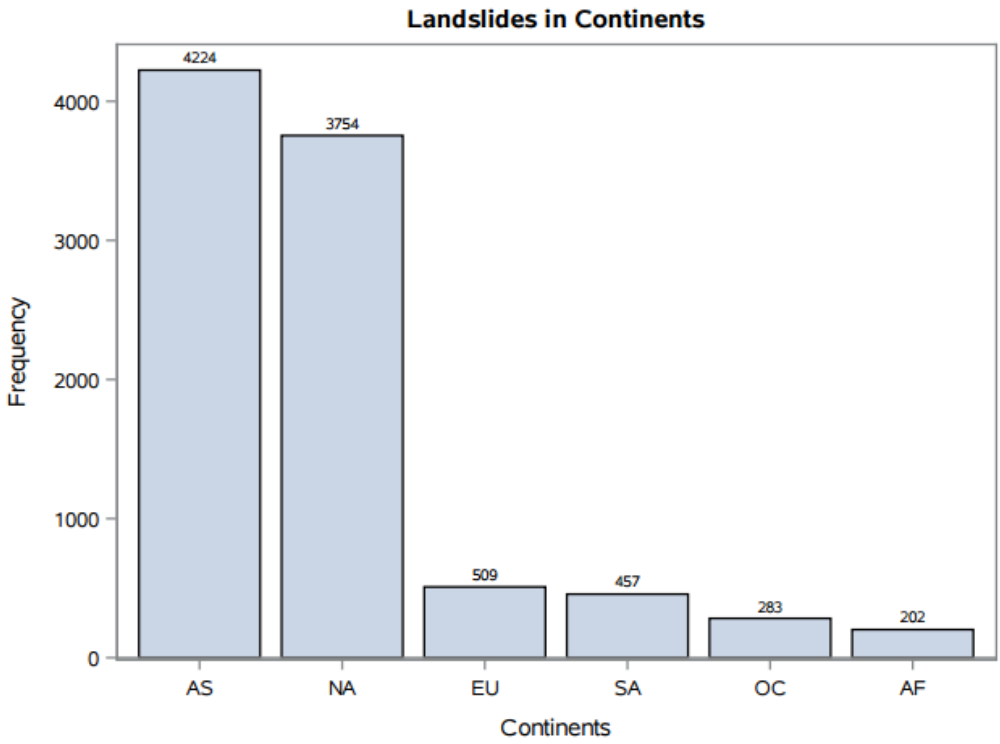

Figure 2: number of landslides that have occurred in each continent

Correlation analysis between fatalities and injuries

For the given data, the sum of the total fatalities caused by landslides is 28,832 and the sum of the injuries recorded is 3354.

The correlation between fatalities and injuries is 0.142 which is very low, suggests there is little or no significant correlation between the two variables. A visualisation of the relation between 2 variables through scatter plot also suggests the same result.

The CORR Procedure

2 Variables:	inj	fat

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
inj	3925	0.85452	8.29846	3354	0	374.00000
fat	8151	3.53723	64.90818	28832	0	5000

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	inj	fat
inj	1.00000 3925	0.14242 <.0001 3918
fat	0.14242 <.0001 3918	1.00000 8151

Figure 3: correlation between the number of injuries and the number of fatalities caused by each landslide recorded

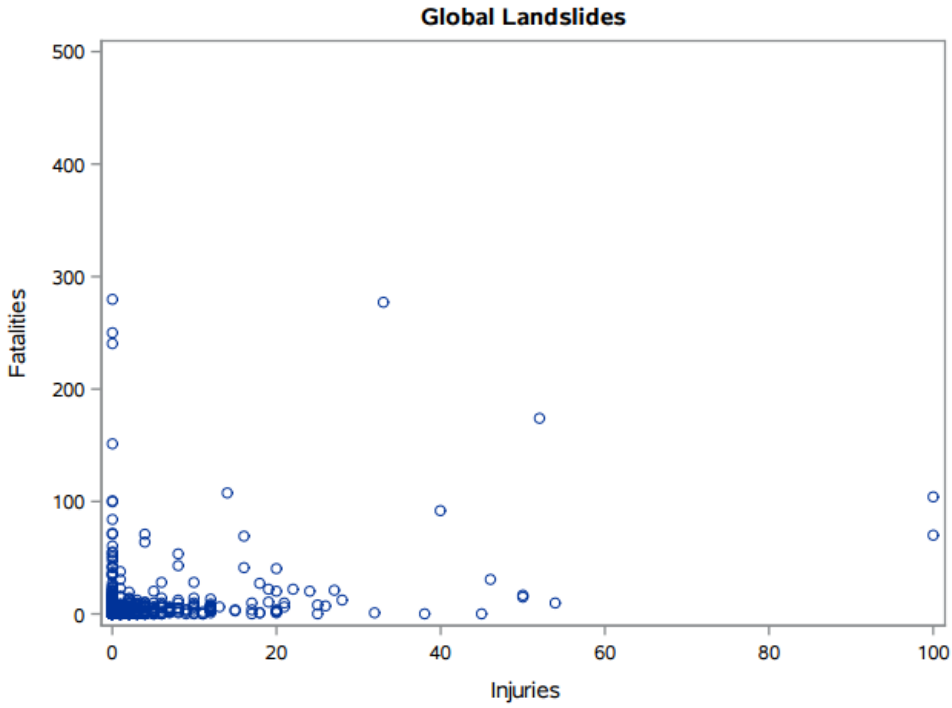

Figure 4: scatter plot of the number of injuries and the number of fatalities caused by each landslide recorded

Analysis of the occurrence of Landslide Types

The various types of the landslides are landslide, mudslide, rockfall, debris flow, riverbank collapse, earth flow, lahar etc. of these, landslide and mudslides are the most common.

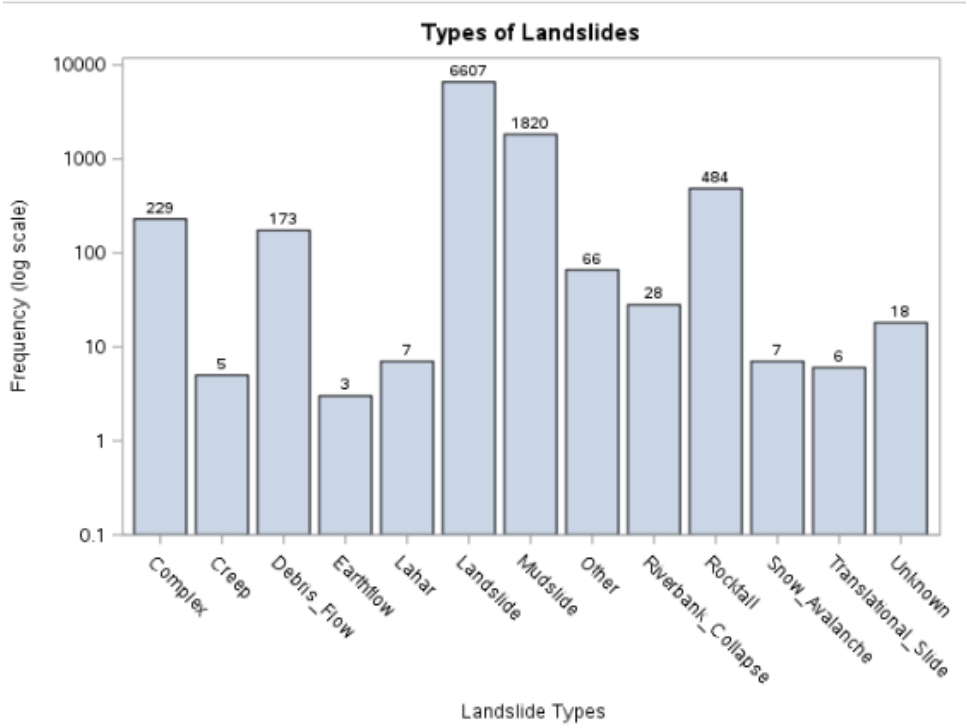

Figure 5: Frequency plot of landslide types on the log scale.

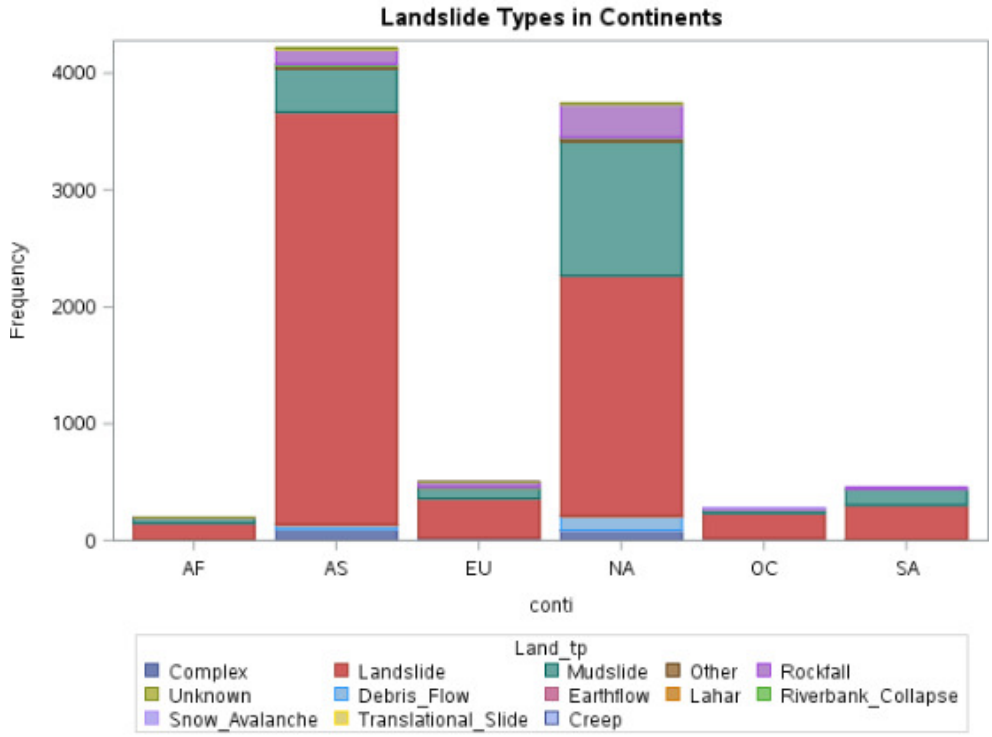

Figure 6: Bar plot of the frequencies of the landslide types in each continent.

Triggers of the landslides in various continents

A study of the triggers of landslides suggests that the main causes of landslides are downpour (51% of total landslides), rain (22% of the total landslides) and continuous rain (9% of the total landslides). Year 2010 recorded maximum number of landslides. Mining and Construction can also cause landslides, though less common.

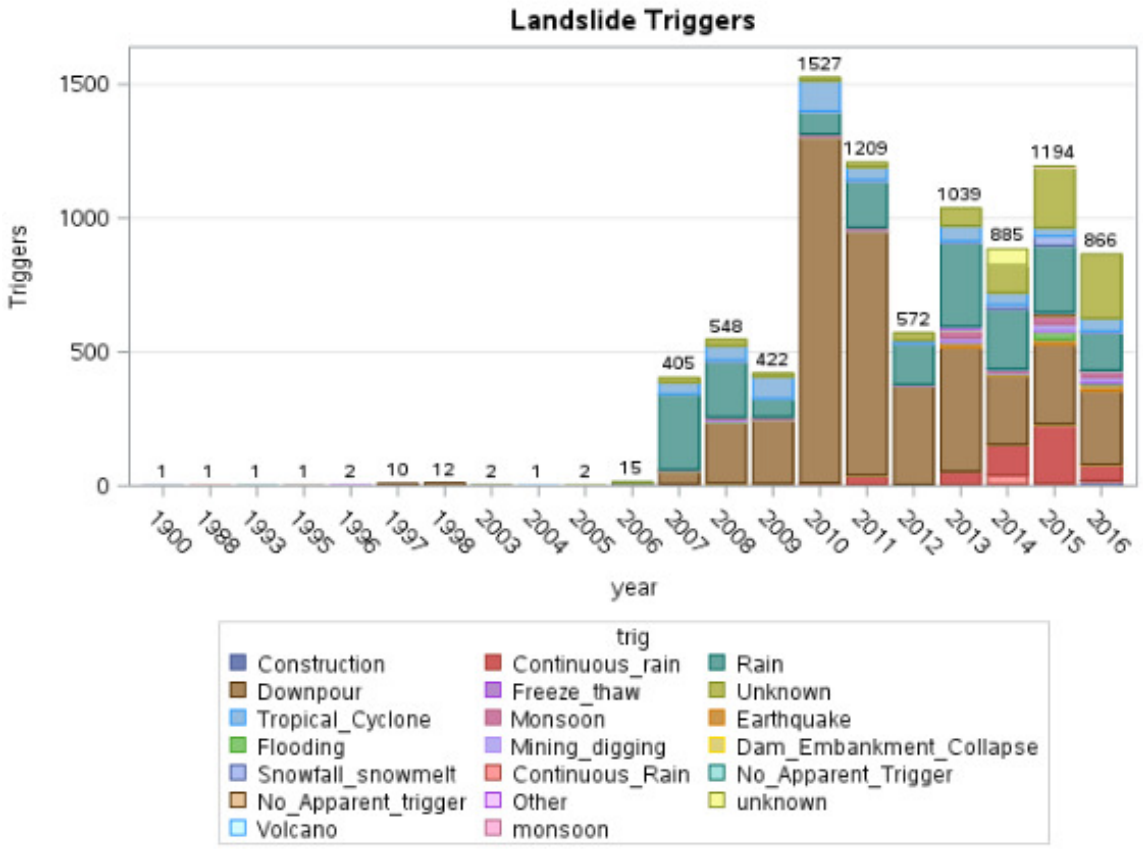

Figure 7: Bar plot of the frequencies of triggers of the landslides

Landslide Triggers	% of Total Data
Construction	1%
Continuous_rain	6%
Dam_Embankment_Collapse	0%
Downpour	51%
Earthquake	1%
Flooding	1%
Freeze_thaw	0%
Mining_digging	1%
Monsoon	1%
No_Apparent_Trigger	0%
Other	0%

Rain	22%
Snowfall_snowmelt	1%
Tropical_Cyclone	6%
Unknown	9%
Volcano	0%
(blank)	0%
Grand Total	**100%**

A study of Latitude and Longitudinal location of landslides

A graphical representation of the latitude and longitude location of landslides, using scatter plot, clearly suggests that the occurrence is mostly is Asia and North America. Also, its more prominent between latitude 0 degrees Equator to 50 degrees North of Equator.

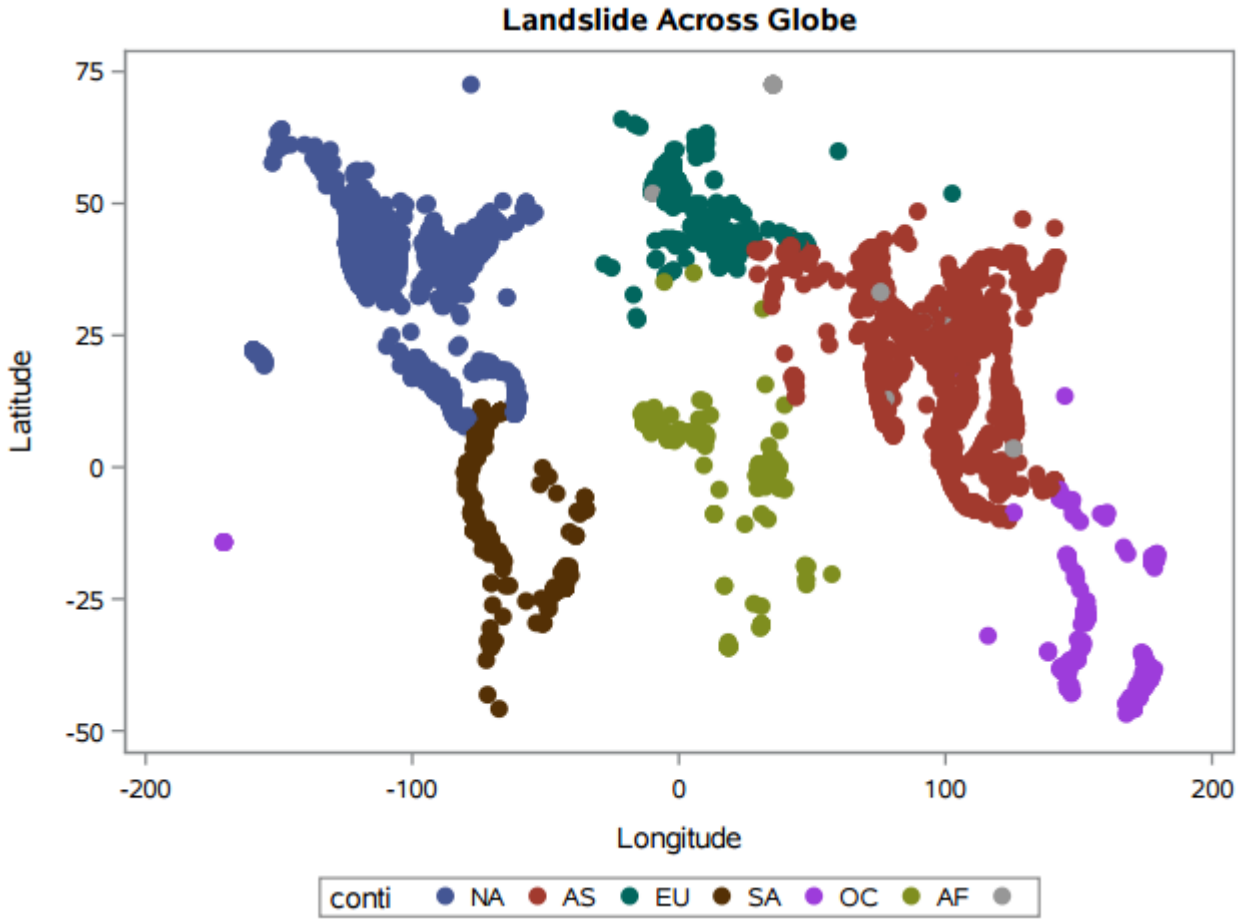

Figure 8: Scatter plot of longitude and longitude occurrences of data

Conclusion

From the above findings, we conclude that most of the landslides were concentrated in Asia and North America. Among the landslides, medium and small size are more common. Also, major triggers for landslides are downpour and continuous rain. Mining and construction, although less common, contribute to landslide occurrences. For latitude and longitudinal location, landslides are more concentrated between 0° Equator to 50°N. Also, the correlation analysis of the fatalities and injuries suggests no noticeable relation between the two variables.

Appendix

APPENDIX OF SAS CODES	
Question.1	DATA landslide; INFILE "/home/s33989790/Assgmnt_2/Global_Landslide.csv" delimiter="," firstobs=2 dsd missover; INPUT id day month year country:$100. nr_pl:$100. hz_tp:$20. Land_tp:$50. trig:$40. storm:$40. fat inj loc_des:$40. loc_km land_sz:$20. cat_src:$10. cat_id coun_name:$40. near:$40. dist admin_1:$40. admin_2:$40. population cntry_code:$10. conti:$10. ver lati longi ; run; PROC Contents data=landslide varnum; run;

Question.2	proc sort data=landslide;
	by year;
	run;
	
	proc print data=landslide (firstobs=1 obs=3);
	Title "Landslide Historic";
	run;

Question.3	proc freq data = landslide; tables conti * land_sz / nocol norow; run;

Question.4	proc corr data=landslide;
	var inj fat;
	run;
	--- 	
	proc sgplot data=landslide;
	title "Global Landslides";
	scatter x=inj y=fat;
	xaxis label="Injuries" type=log min=0 max=100;
	yaxis label="Fatalities" type=log min=1 max=500;
	run;

Question.5	proc sgplot data=landslide;
	Title "Types of Landslides";
	vbar Land_tp / baseline = 0.1 datalabel;
	xaxis label= "Landslide Types";
	yaxis label= "Frequency (log scale)"
	TYPE=LOG LOGSTYLE=LOGEXPAND LOGBASE=10
	;
	run;
	--- 	
	(Additional analysis)
	proc sgplot data = landslide;
	Title "Landslide Types in Continents";
	vbar conti / group=land_tp;
	run;

Question.6	proc sgplot data = landslide;
	Title "Landslide Triggers";
	vbar year / group=trig datalabel;
	yaxis label="Triggers" grid;
	run;
	--- 	
	proc sgplot data=landslide;
	Title "Landslide Across Globe";
	scatter x=longi y=lati / markerattrs=(symbol=circlefilled size=10px) group=conti;
	xaxis label = "Longitude";
	yaxis label="Latitude";
	run;