Assignment 2
[Global Landslide Assessment]

Table of Contents

Introduction

Landslides are the most common types of Natural hazards which are responsible for reshaping the landscapes and anything which comes in its path. Maintaining the information on occurrences of landslides worldwide is bit difficult due to the lack of centralized database system but NASA researchers have updated the global landslide Catalog with the aim of identifying rainfall triggered landslide incidents around the world regardless of the size, impact or location. The basic objective of the report is to analyse the global landslide forecasting initiatives and understanding the patterns in landslide distribution and frequency worldwide.

Methods

The data is extracted from data.nasa.gov. The data in the Catalog considers all types of mass movements for several years that are reported in the media, disaster databases, scientific reports and other sources. This will help in estimating human and economic losses, understanding the relationships between landslide occurrences, climate variations and evaluating the landslide prediction efforts based on the data.
A landslide incident occurred in the globe is stored in the form of 2 elements (viz. Primary and Secondary). Primary elements describe the event and secondary element provides the additional information about the landslide characteristics.
Primary elements:

1) Location Information: (Country, Continent, nearest places)
2) Time (Date and Time of the event)
3) Trigger (Causes of Landslides- Construction, Continuous rainfall, Flooding, snowfall, tropical cyclone etc.)

Secondary elements:

1) Type and Size of the landslides (Landslide types are categorized as complex, debris flow, riverbank collapse, rockfall, snow avalanche etc.) and size of the landslides as (Large, Medium, Small & Very large)
2) Latitude and Longitude indicating the location accuracy.
3) Impact information indicating injuries and fatalities.

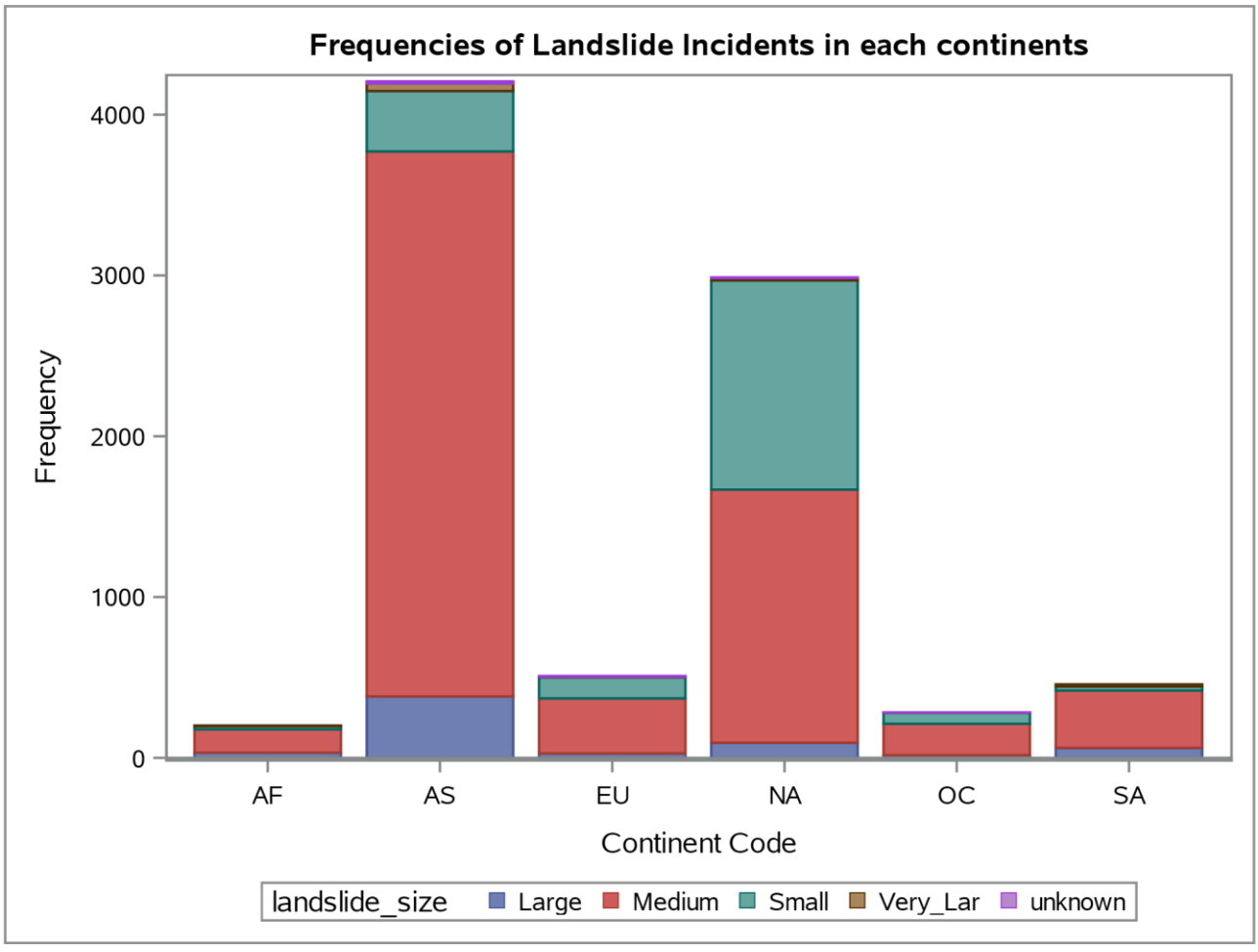
SAS programming language is used to analyse the dataset. It is a very powerful programming language which is used for statistical analysis.

Results

We can follow the below steps to analyse the dataset.

1) Overview function in SAS helps us in analysing the dataset and check whether the dataset has been read properly. There are total 9471 landslide events occurred across the globe which are included in the dataset. The first landslide event is reported in the year 1900 and the trigger for the event is due to construction.
2) PROC Freq function gives us the frequency and overall percentage of Landslides in each continent. This will give us the clear picture of number of incidents reported in each continent.

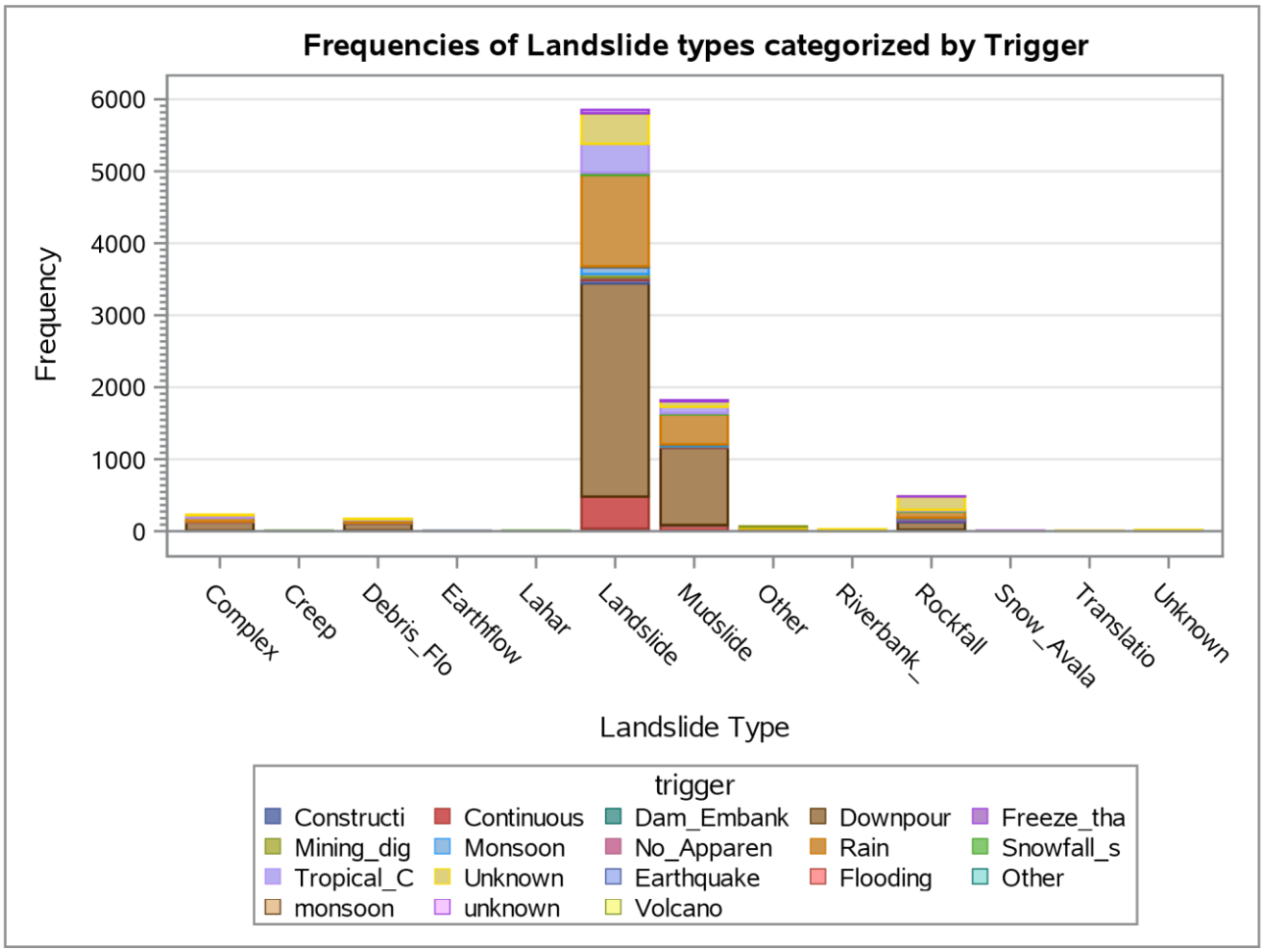
Table of continentcode by landslide_size						
	landslide_size					
continentcode	Large	Medium	Small	Very_Lar	unknown	Total
AF	32 0.37	145 1.68	19 0.22	5 0.06	0 0.00	201 2.33
AS	382 4.42	3389 39.20	376 4.35	47 0.54	13 0.15	4207 48.66
EU	28 0.32	342 3.96	132 1.53	3 0.03	4 0.05	509 5.89
NA	94 1.09	1574 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						



Maximum number of landslide events occurred in Asia which is about (4207) and North America (2988). The sizes of the landslides in these continents are basically Medium and small. Each landslide event occurred is categorized as per the types such as Landslide, Mudslide, Rockfall, Debris flow etc. The Landslides incidents occurred in other continents (Africa, Europe, Oceania & South America) are below 10 %.

3) On Further analysing the dataset we can find that each landslide type is triggered by some events such as Rain, Construction, Flooding, Downpour, Snowfall etc. From the graph, it can be seen that there are 2 main types of trigger events namely Rain and Downpour which causes landslides, Mudslides and Rockfalls, Debris Flow.

Sunday, September 17, 2017 07:29:33 PM 1



4) Landslide incidents results in human and economic losses. Correlation procedure in SAS helps us in understanding the relation between the 2 variables. We can check the relation between the number of fatalities and number of injuries reported due to global landslide events. From the data in the 2nd table we can conclude that the correlation in diagonal cells are always equal since the variable is always perfectly correlated with itself. The Test gives us the following values ($r=0.14242$, $p<0.0001$).

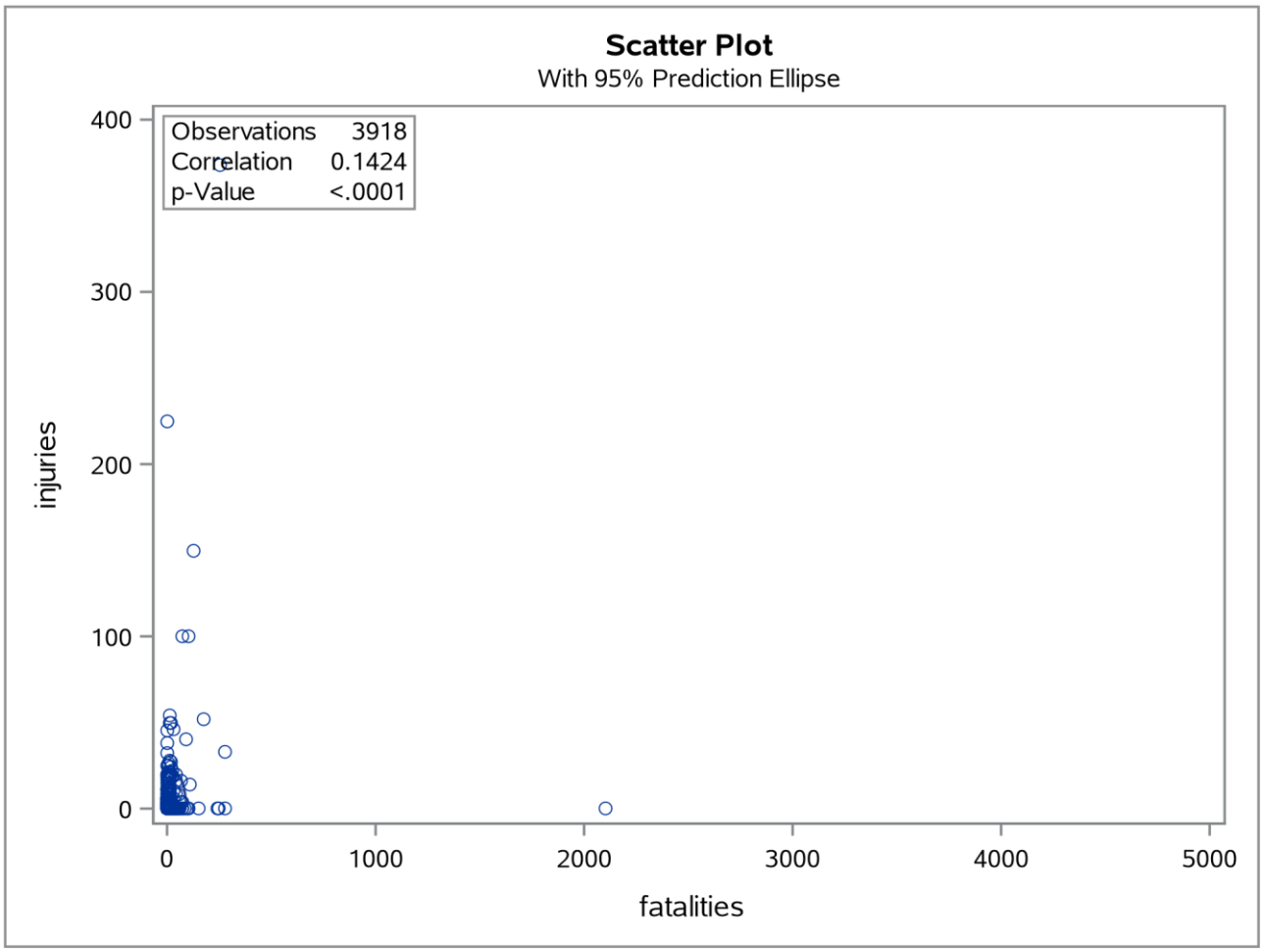
2 Variables:	fatalities injuries

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
fatalities	8151	3.53723	64.90818	28832	0	5000
injuries	3925	0.85452	8.29846	3354	0	374.00000

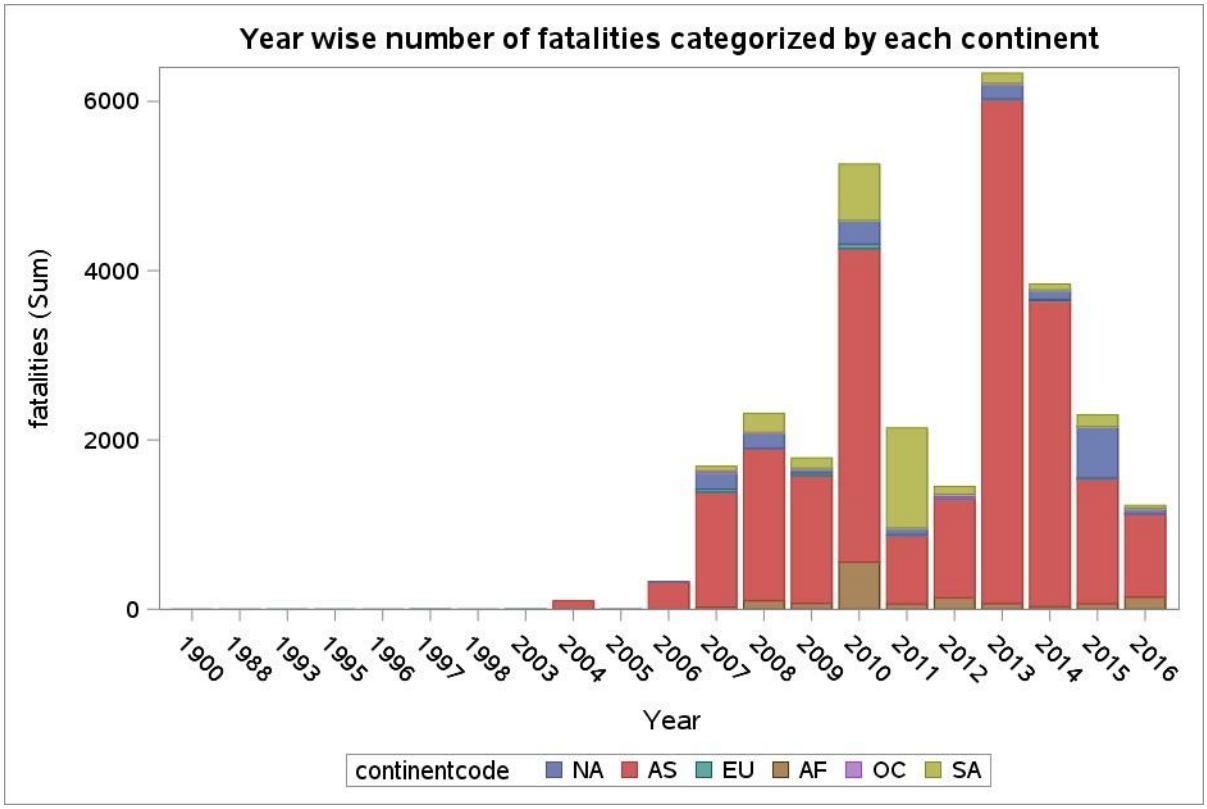
Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations		
	fatalities	injuries
fatalities	1.00000 8151	0.14242 <.0001 3918
injuries	0.14242 <.0001 3918	1.00000 3925

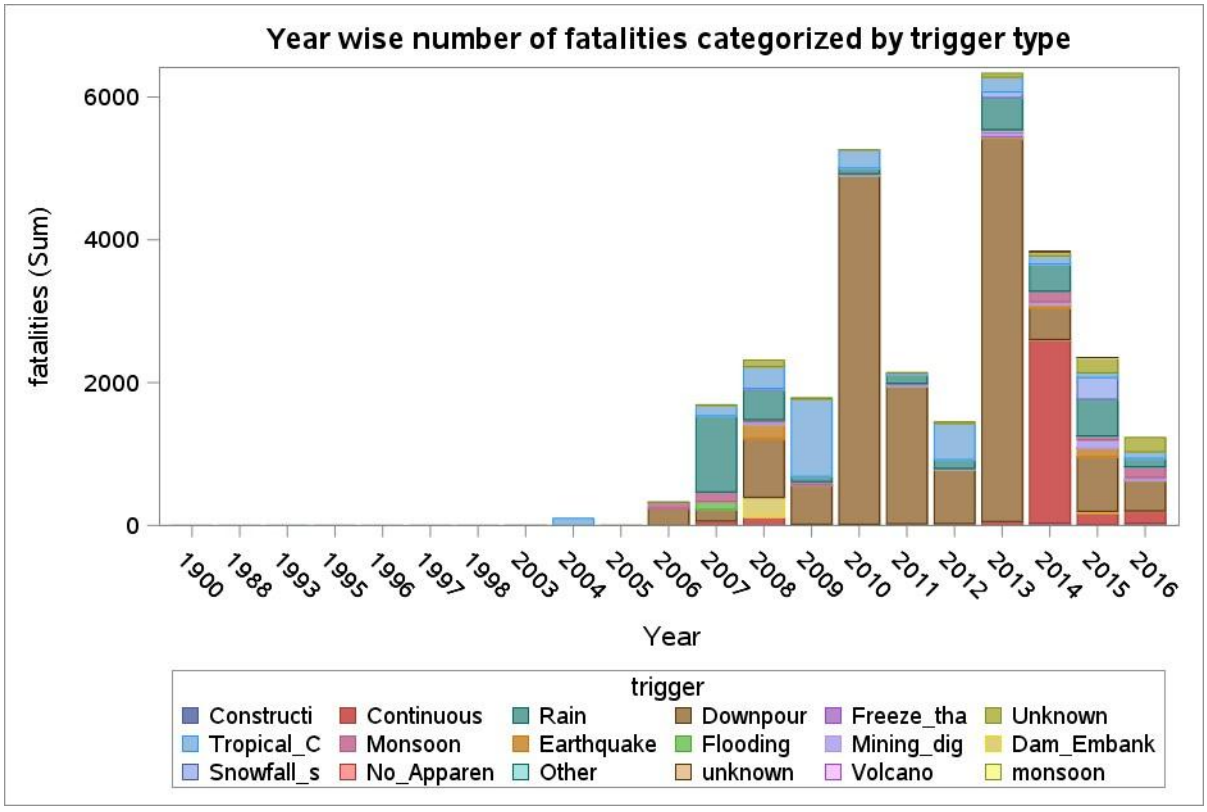
\

The CORR Procedure



Using Bar graph in SAS we can plot the year wise Number of fatalities in each continent. Number of fatalities are highest in the year 2010 and 2013 for Asia. There are no fatalities recorded from the year 1900 to 2003.





The above graph gives the pictorial view of number of fatalities recorded in each year and the trigger type of the landslides. (i.e. the reason of occurrences of landslides). Downpour and Rain are the biggest causes of landslides. The other factors which are causing the landslides to occur are construction, Continuous rain, earthquake, volcano etc.

Conclusion

Based on the analysis we can derive the following conclusion.

1) Around 83% of the total landslide incidents are recorded in Asia and North America continents.
2) 17 % of incidents are recorded in other continents such as Europe, Africa, Oceania and South America.
3) Downpour and Rain are the biggest causes of the landslides to occur in any continent. This in turn results in landslide, Mudslides and rockfalls causing number of fatalities and injuries.
4) Highest number of fatalities are recorded in the year 2010, 2013 and 2014 and are approximately equal to (5000, Above 6000, and 4000) respectively.
5) Overall 8151 fatalities and 3925 injuries are recorded due to global landslides.

References

The dataset is sourced from data.nasa.gov. More information can be extracted from the link https://data.nasa.gov/dataset/Global-Landslide-Catalog-Export/dd9e-wu2v/data.

Appendix

1) Reading of the dataset in SAS and preparing an overview.

```
data globallandslide;
infile "/home/s36772710/sasuser.v94/data/Global_Landslide.csv" delimiter="," firstobs=2
truncover dsd;
input id Day Month Year country~$20. nearest_places~$45. hazard_type~$10.
landslide_type~$10. trigger~$10. storm_name~$45. fatalities injuries
location_description~$25. location_accuracy_km landslide_size$ cat_src$ cat_id
countryname~$15. near~$40. distance adminname1~$20. adminname2~$20. population
countrycode$ continentcode$ version latitude longitude;
run;

proc contents data=globallandslide order=varnum;
run;
```

2) Code to read 3 oldest records in the dataset

```
proc contents data=globallandslide order=varnum;
run;
```

3) Code for producing a table that compares the number of landslides for each landslide size category

```
proc freq data=globallandslide;
Table continentcode*landslide_size/norow nocol;
```

```
title "Frequency & Overall Percentage of Landslides in each continents";
run;

proc sgplot data=globallandslide;
vbar continentcode / group= landslide_size;
xaxis display=all;
yaxis type=log max=7000 offsetmin=0 label='Frequency' grid;
xaxis label="Continent Code" ;
title "Frequencies of Landslide Incidents in each continents" bold;
run;
```

4) Test the correlation between the number of injuries and the number of fatalities

```
PROC CORR data=globallandslide;
var fatalities injuries;
RUN;

PROC CORR data=globallandslide plots(MAXPOINTS=9471)=scatter();
var fatalities injuries;
RUN;
```

5) To produce a plot of the frequencies of the landslide types.

```
proc sgplot data=globallandslide;
vbar landslide_type / group= trigger;
xaxis display=all;
yaxis label="Frequency" type=log logstyle=logexpand logbase=10  minor offsetmin=0.05
offsetmax=0.05 grid;
xaxis label="Landslide Type";
title "Frequencies of Landslide types categorized by Trigger" bold;
run;
```

6)

```
ods graphics / reset width=600px height=400px imagefmt=png imagename='BarChart';
title "Year wise distribution of fatalities categorized by each continent";
proc sgplot data=globallandslide;
vbar year / group=continentcode response=fatalities seglabel seglabelattrs=(size=12);
run;

ods graphics / reset width=600px height=400px imagefmt=png imagename='BarChart';
title "Year wise distribution of fatalities categorized by trigger type";
proc sgplot data=globallandslide;
vbar year / group=trigger response=fatalities seglabel seglabelattrs=(size=12);
run;
```