

Group 7

Jae Hee Hwang, Junhui Liao, Peter Darche, Gary Sztajnman, Pawel Buczak

March 8, 2016

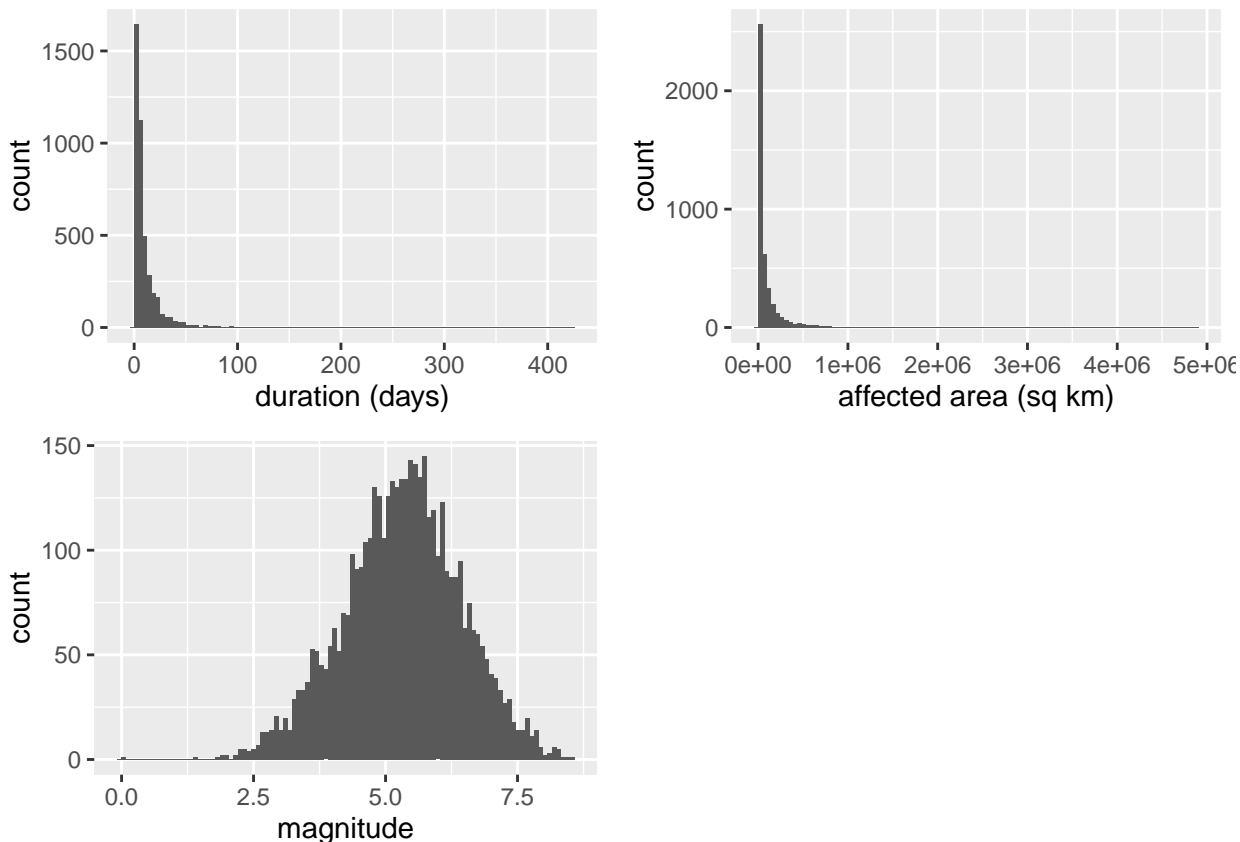
Introduction

Our analysis took place in roughly 3 parts. The first was initial analysis looking at the basic features of the flood dataset. It sought to get a picture of the how the data was distributed, what was typical, and to generate hypotheses about potential explanations for the variation in the dimensions. The second sought to examine the connection between geopotential height and flooding. Lastly, we examined the impacts of flooding and what explained the (sometimes very large) differences in the damage done by these events.

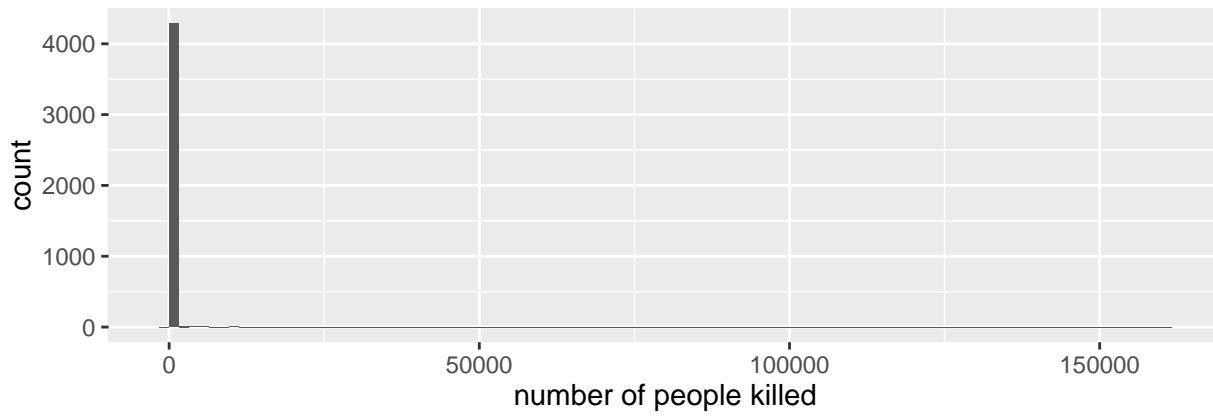
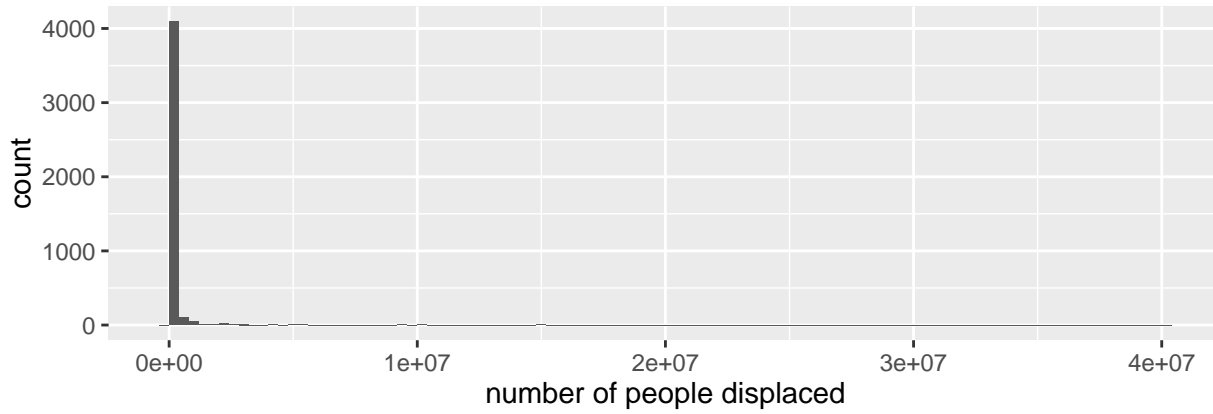
Part 1 - Understanding the Flood Data: Where Do Floods Happen and Why?

We started our analysis by trying to get a basic sense of the properties of the floods recorded in the dataset. This meant looking for things like the spatial and temporal distribution of flood, the distribution of flood size, duration, and area affected, and distributions of . We wanted to see what variation there was between floods, and if we could get any indicators of what was explaining that variation.

Distributions of flood duration, affected area, and magnitude

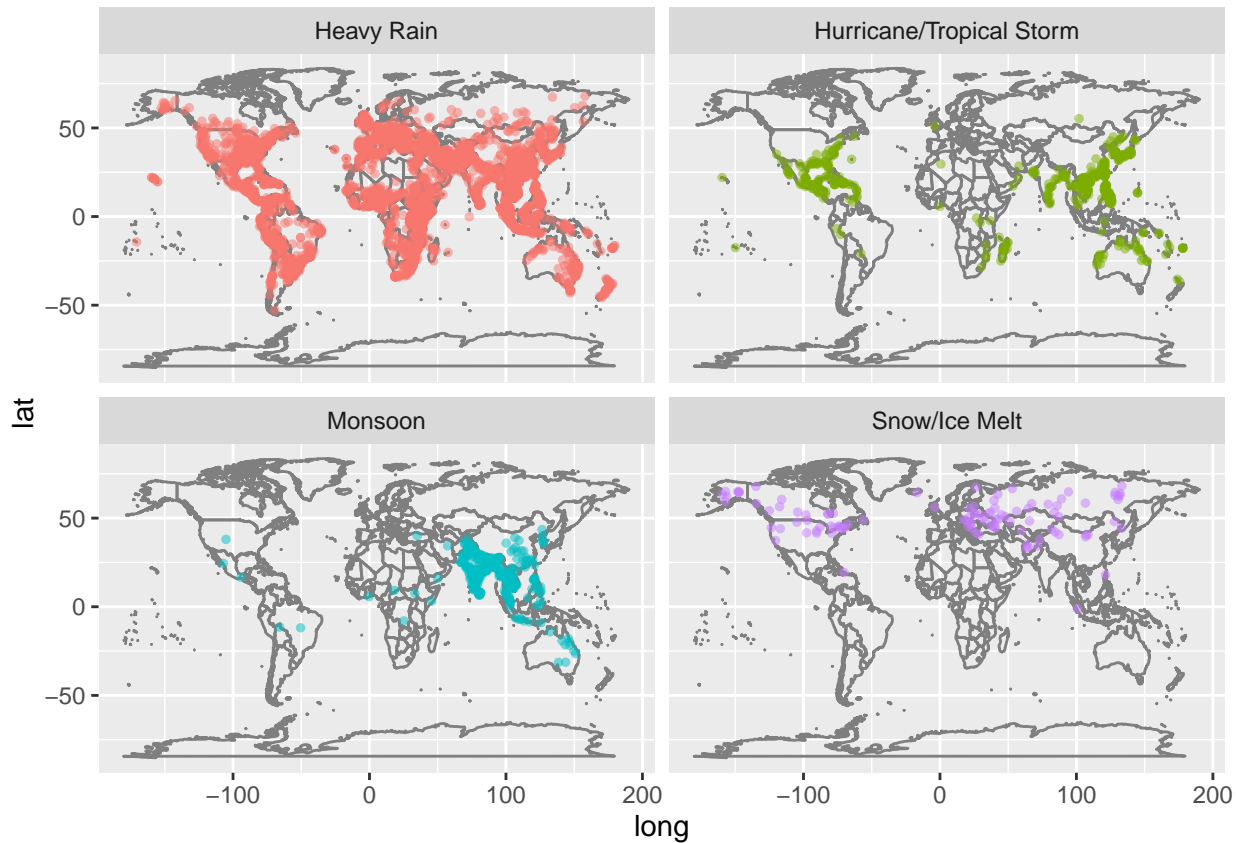


Distribution of flood consequences (people killed and displaced)



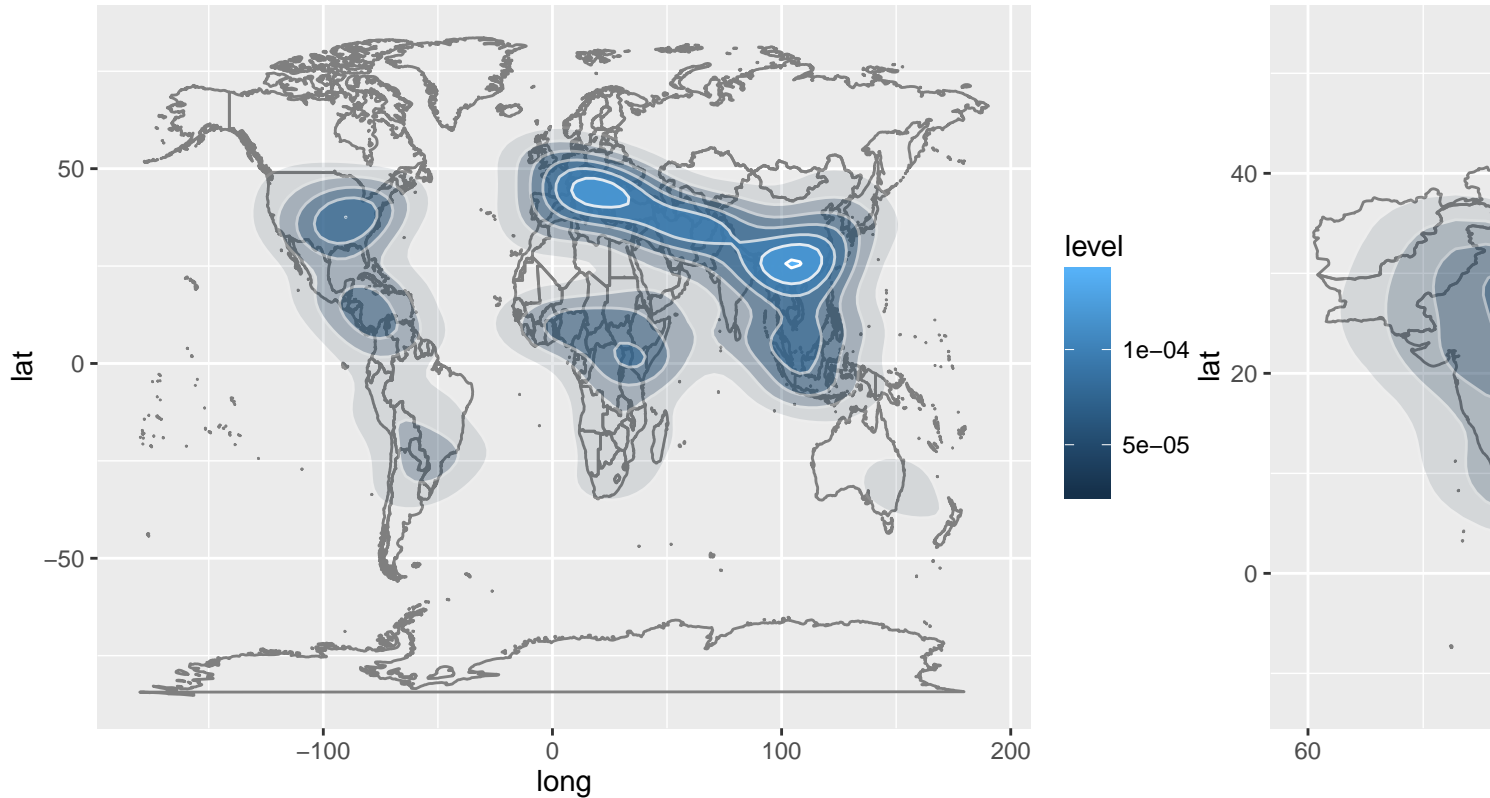
Many of the above plots mostly follow a 'power law' distribution, with the bulk of the mass residing in the first 20 percent of the distribution and with a thin but long tail of extreme outlying events. Subsetting and faceting the distributions by type, cause, severity didn't yield significantly different distributions.

Spatial distribution of floods



This plot shows the spatial distribution of floods by cause. Not surprisingly, the plot indicates that the geography has a significant effect on the type of flooding experienced in different parts of the world. Floods due to Monsoon, for example happen primarily in the Indian Ocean and Southeast Asia; floods caused by snow melt happen at more northerly mountainous regions; floods caused by hurricanes and tropical storms happen mostly along the coast in warmer, more tropical latitudes. Floods caused by heavy rainfall were present all over the world (except in expected locations like deserts). Further (and also not surprisingly), we noticed that floods occur most frequently near rivers and coastlines.

Heavy Rain Density



The density maps allow for somewhat easier interpretation of where the bulk of floods occur. As mentioned previously and as vividly indicated by the monsoon density map, densities often center near rivers and coastlines, such as the Ganges River Delta.

The animation shows every day from the beginning of June to the end of October 2007. Floods appear on the map once they have “began” according to the flood data and disappear in a similar fashion. Geopotential height is indicated by the shading, which is red for lower levels and blue for higher levels. The animation suggests that lower levels of geopotential height tend to occur before or at the beginning of floods.

The animation also points out another difficulty in predicting the severity of floods - not only does the actual rainfall (or proxy for it such as geopotential height) matter, but geography also plays a role. Note that there are many floods along the northern portion of India and into Bangladesh, despite there not being extreme values of geopotential height during this time. This, however, is related to the fact that this is where the mighty Ganges river runs, and rivers such as this can drain very large areas. Thus flooding can potentially occur long after and far away from the heavy rains.

Part 2 - Relationship between Geopotential Heigh & Flood Magnitude

This section explores the relationship between Geopotential Heigh and Flood Magnitude by two major datasets. This first one is the “NOAA_Daily_phi_500mb.nc”, which provides the geopotential heigh values and the second one is the “GlobalFloodsRecord.xls”, which provides different kinds of flood data. In addition, we focus on both dataset in 2012 to 2013 as well as 2014 to 2015 within the region of the United States.

Methology

scatter plot is utilized to glance at corresponding geopotential heigh values and flood magnitude. Besides, a simple linear regression models is used to determine the relationship between the two variables:

$$floodMagnitude = \beta_0 + \beta_1 * geopotentialHeigh$$

Data

All data has been cleaned up and exported as csv file. The first five corresponding geopotential heigh and magnitude values from 2012 to 2013 are displayed below. Since each flood appearen in a period of time, the corresponding geopotential heigh value is taking as the mean during that period of time

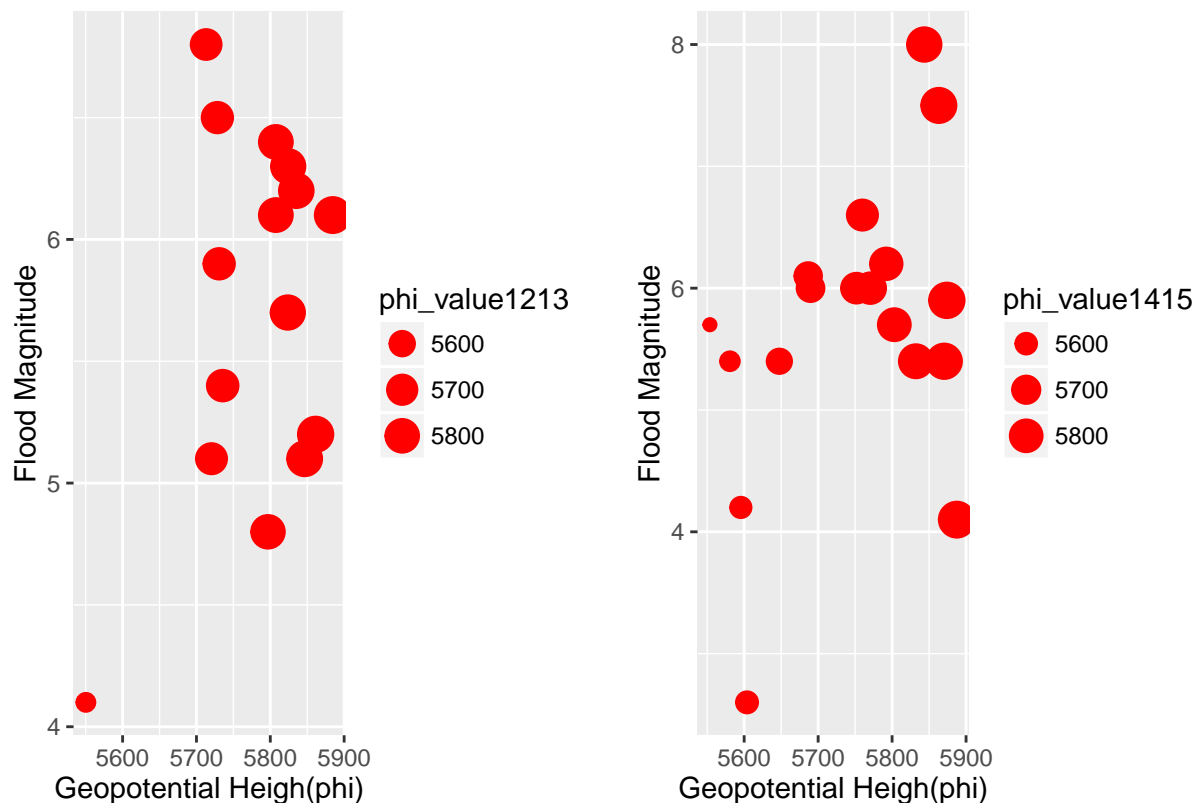
```
##  phi_value1213 magnitude1213
## 1    5730.727          5.9
## 2    5824.222          6.3
## 3    5823.667          5.7
## 4    5807.500          6.4
## 5    5807.500          6.1
## 6    5735.500          5.4
```

Below are the data values from 2014 to 2015. Same as above, the corresponding geopotential heigh value is takeing as mean during that period of time

```
##  phi_value1415 magnitude1415
## 1    5759.889          6.6
## 2    5792.154          6.2
## 3    5874.000          5.9
## 4    5870.500          5.4
## 5    5803.000          5.7
## 6    5843.536          8.0
```

Result

Scatter Geopotential Height & Magnitude



The above graphics display the relationship between geopotential heigh and flood magnitude in an Euclidean space, where the x axis represents geopotential heigh(phi) value and the y axis preresents the flood magnitude value. The size of the bubble also indicate how large the corresponding phi value is.

The correlation between these variables are also included below

geopotential heigh and flood magnitude correlation from 2012-2013:

```
## [1] 0.3698957
```

geopotential heigh and flood magnitude correlation from 2014-2015:

```
## [1] 0.4008081
```

finally a simple linear regression model is being appied to the dataset.

geopotential heigh and flood magnitude regression from 2012-2013:

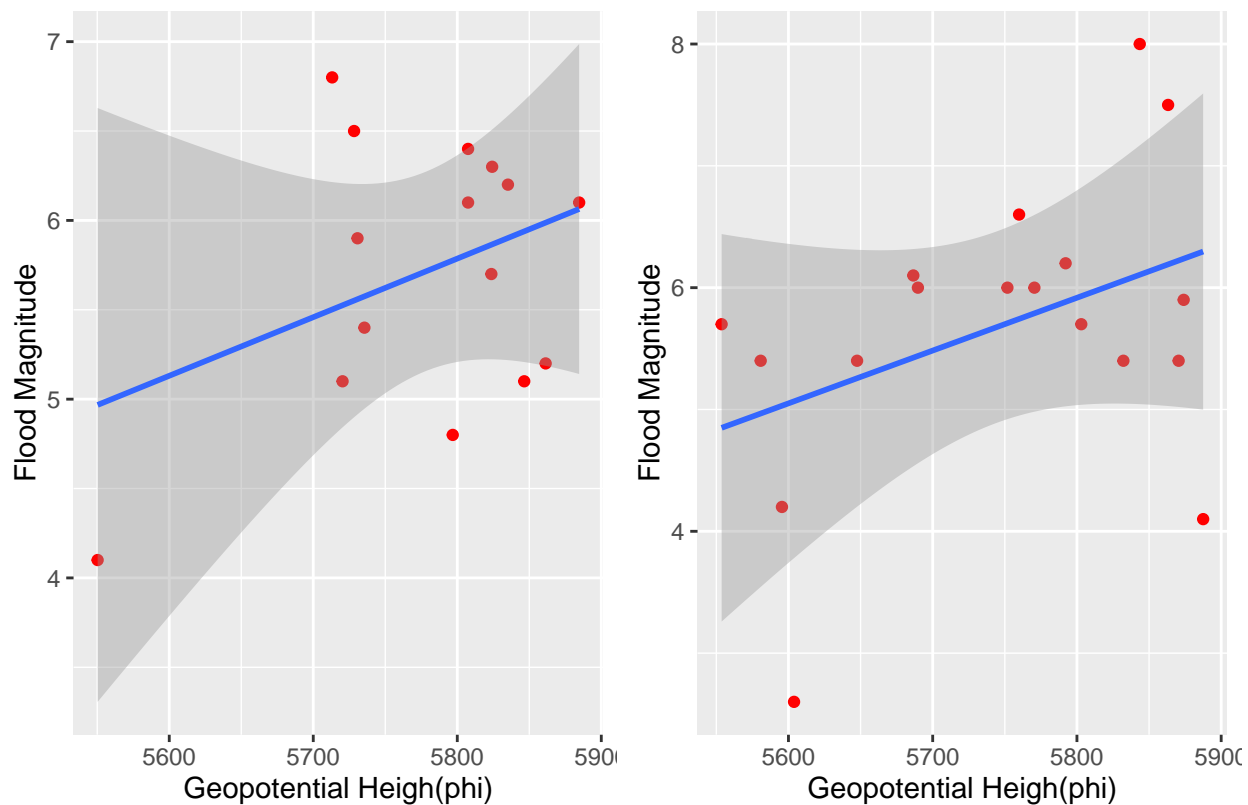
```
##
## Call:
## lm(formula = magnitude1213 ~ phi_value1213, data = phi_magni1213)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.97609 -0.60638  0.03615  0.38745  1.29840
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -13.227083  13.195664  -1.002   0.334
## phi_value1213   0.003278   0.002284   1.435   0.175
##
## Residual standard error: 0.7168 on 13 degrees of freedom
## Multiple R-squared:  0.1368, Adjusted R-squared:  0.07042
## F-statistic: 2.061 on 1 and 13 DF,  p-value: 0.1748
```

geopotential heigh and flood magnitude regression from 2014-2015:

```
##
## Call:
## lm(formula = magnitude1415 ~ phi_value1415, data = phi_magni1415)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.4670 -0.5770  0.2512  0.6461  1.8941
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -19.224277  14.232740  -1.351   0.1956
## phi_value1415   0.004335   0.002477   1.750   0.0993 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.141 on 16 degrees of freedom
## Multiple R-squared:  0.1606, Adjusted R-squared:  0.1082
## F-statistic: 3.062 on 1 and 16 DF,  p-value: 0.09928
```

visualize the result:

Geopotential Height & Magnitude Linear Regression Model



Both the shaped areas represent the 99% confident level in the regression model.

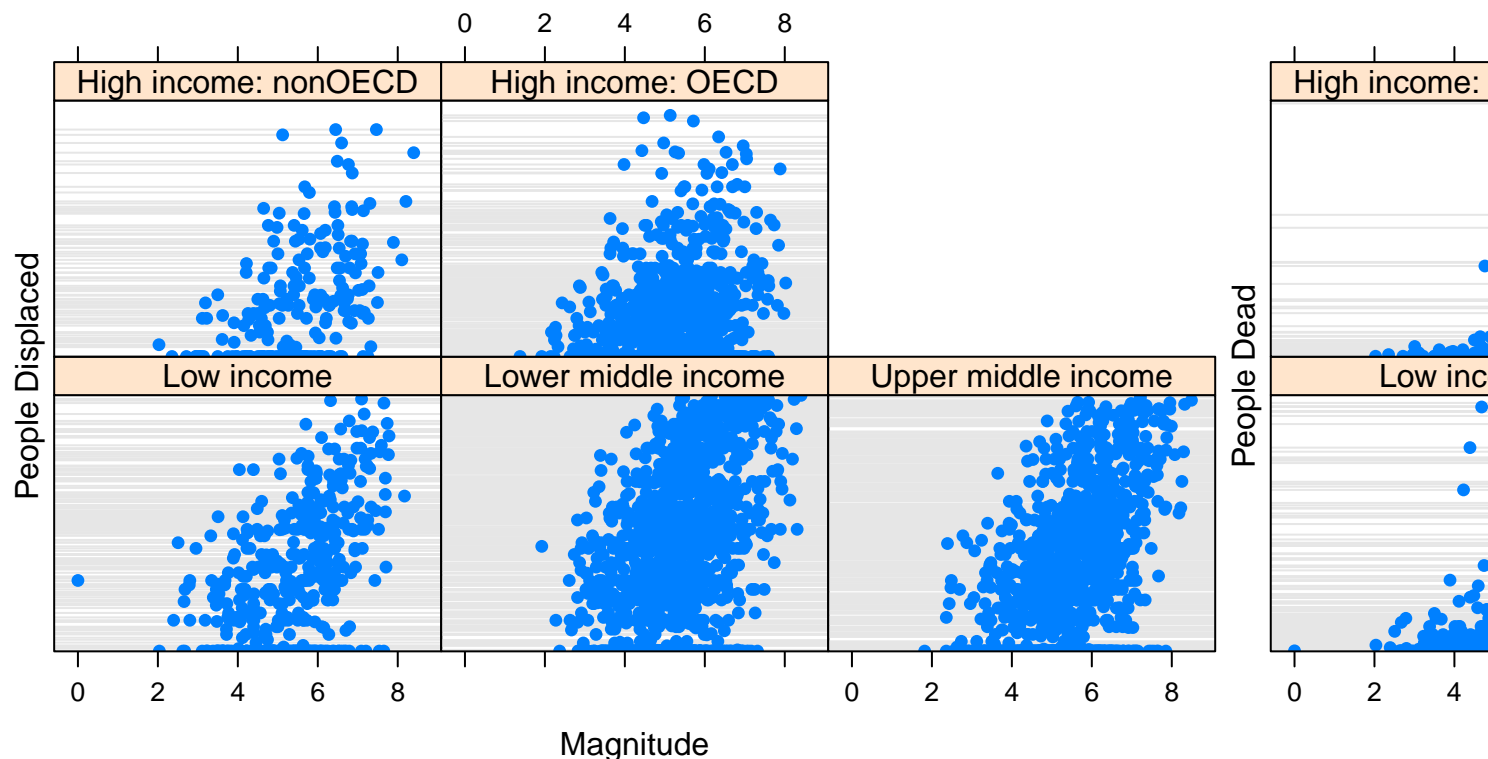
Conclusion

From the above result, we can not find a statistically significant relationship between geopotential height and flood magnitude in the United States from 2012 to 2013 and 2014 to 2015 data.

Part 3 - Understanding the Impact of Floods

Floods have a huge impact on population and economy. Heavy rain leads to destruction, death, displacement, health issues, reparations. Some countries may receive more floods based on its location (for example India) but some countries also don't have the capacity to prevent, prepare population, manage flood and repair. In this analysis we will try to analyse what characteristics influence the impact of a flood.

Displacement by Magnitude Grouped by Income Level



Displacement

The scatterplot of people displaced for high income is concentrated in the bottom, which means that High Income countries tend to not displace people much regardless of the magnitude of the flood.

In contrast, low income, lower middle income, and upper middle income countries seem to be more middle-heavy. These countries also seem to displace more people as magnitude increases.

Deaths

The scatterplot of people dead for high income is very bottom-heavy, even more so than in displacement. As income level goes down, we notice that a lot more points are plotted on top. We can quantify this finding further by running a regression analysis by country income level.

```
## Call:
##   Model: peopleDisplaced ~ magnitude | incomeGroup
##   Data: datc
##
## Coefficients:
##               (Intercept)  magnitude
## Low income              -356801.66  81579.940
## Lower middle income    -1822435.25 403441.847
## Upper middle income    -624934.98 137749.151
## High income: nonOECD   -79965.76  17940.510
## High income: OECD     -12639.72   4798.086
##
## Degrees of freedom: 4159 total; 4149 residual
## Residual standard error: 1243627
```



```
## Call:
##   Model: peopleDead ~ magnitude | incomeGroup
##   Data: datc
##
## Coefficients:
##               (Intercept)  magnitude
## Low income             -121.72851  34.940425
## Lower middle income    -642.53399 172.759060
## Upper middle income    -102.40638  52.260055
## High income: nonOECD   -334.65530  79.528348
## High income: OECD       57.91202  -7.579596
##
## Degrees of freedom: 4159 total; 4149 residual
## Residual standard error: 3654.833
```

Displacement

There is a trend of the slope coefficient on the variable “magnitude” as income level goes down. This means that the higher the magnitude, the higher number of people are displaced as country income level goes down. Intuitively, this can be interpreted as since high-income countries are better prepared for floods in all magnitude, they tend to displace only few people regardless of magnitude. However, countries with low income do not have resources to prepare for floods with large magnitude. Therefore, they tend to displace more people as magnitude intensifies.

Death

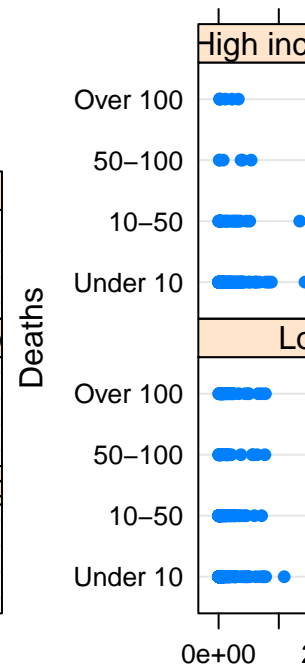
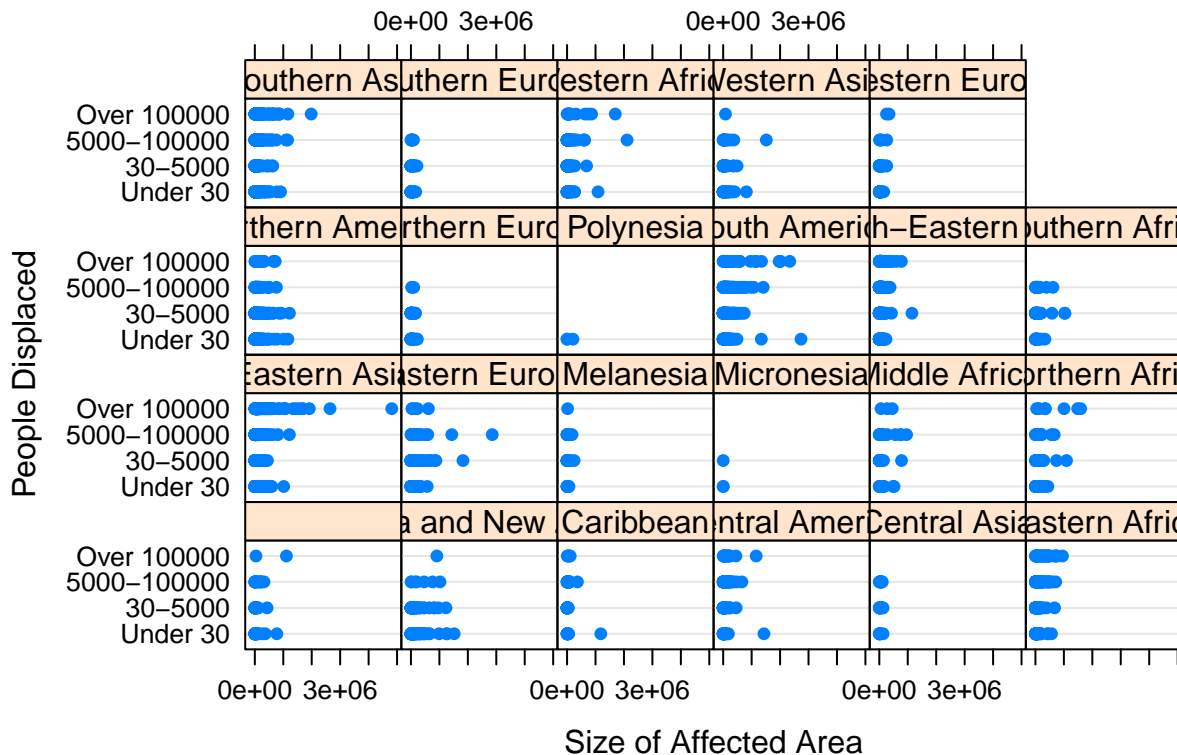
Similarly, the coefficient on magnitude to predict number of death is not even positive for high income OECD countries. However, as the income level goes down, the importance of magnitude in predicting the number of deaths increases. The reasoning is the same as displacement. Countries with high income are better prepared, and they take good measures to prevent deaths from happening, even in cases of severe magnitudes. Countries with lower incomes do not have such resources.

Now, let’s look at differing effects of affected area by country income group.

Size of Affected Area and Displacement/Death

Displacement by Size of Affected Area Grouped by Income Level

Deaths by



Displacement

Similar to magnitude, in High Income OECD countries, the scatterplot is bottom-heavy. Regardless of the size of the affected area, very few people were displaced. However for Lower middle income and upper middle income countries, the scatterplot is top-heavy, which means that the larger the affected area, the more people were displaced.

Death

The correlation between number of people dead and the area of affected region seems to be lower for deaths. Again, for High-OECD countries, the scatterplot is very much bottom-heavy.

```
## Call:
##   Model: peopleDisplaced ~ affectedSqKm | incomeGroup
##   Data: datc
##
## Coefficients:
##               (Intercept) affectedSqKm
## Low income              42097.31    0.39752791
## Lower middle income      157822.82    1.81979411
## Upper middle income       66786.57    0.43434747
## High income: nonOECD      12158.40    0.04800677
## High income: OECD         11032.64    0.01281745
##
## Degrees of freedom: 4159 total; 4149 residual
## Residual standard error: 1253774
```

```
## Call:
##   Model: peopleDead ~ affectedSqKm | incomeGroup
##   Data: datc
##
## Coefficients:
##               (Intercept)  affectedSqKm
## Low income              53.67329  1.252309e-04
## Lower middle income     257.64169  1.865733e-04
## Upper middle income     184.20074 -1.624620e-05
## High income: nonOECD     84.41207  1.539756e-04
## High income: OECD       21.09149 -2.687088e-05
##
## Degrees of freedom: 4159 total; 4149 residual
## Residual standard error: 3656.426
```

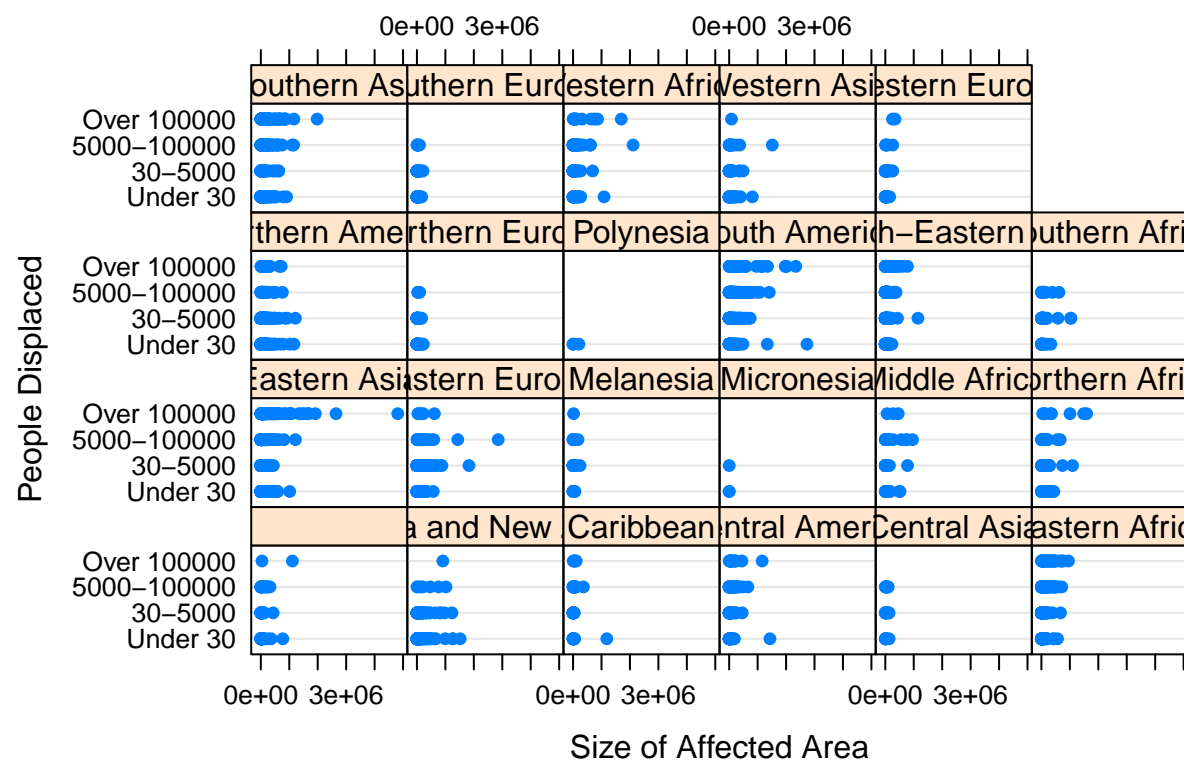
The regression analysis indicates that similar trend of strong correlation between area affected and number of people displaced as the country income level goes down.

The relationship between the size of area affected and number of deaths seems to be much weaker. Interestingly, high income OECD countries and upper middle income countries have a negative relationship between affected square kilometers and number of deaths. While the coefficient is not very large, this can mean that when there are floods that affect large areas, countries predict it and take measures to prevent deaths.

More detailed data can be found when countries are analyzed at a subregion level.

Size of Affected Area

Displacement by Size of Affected Area Grouped by Region

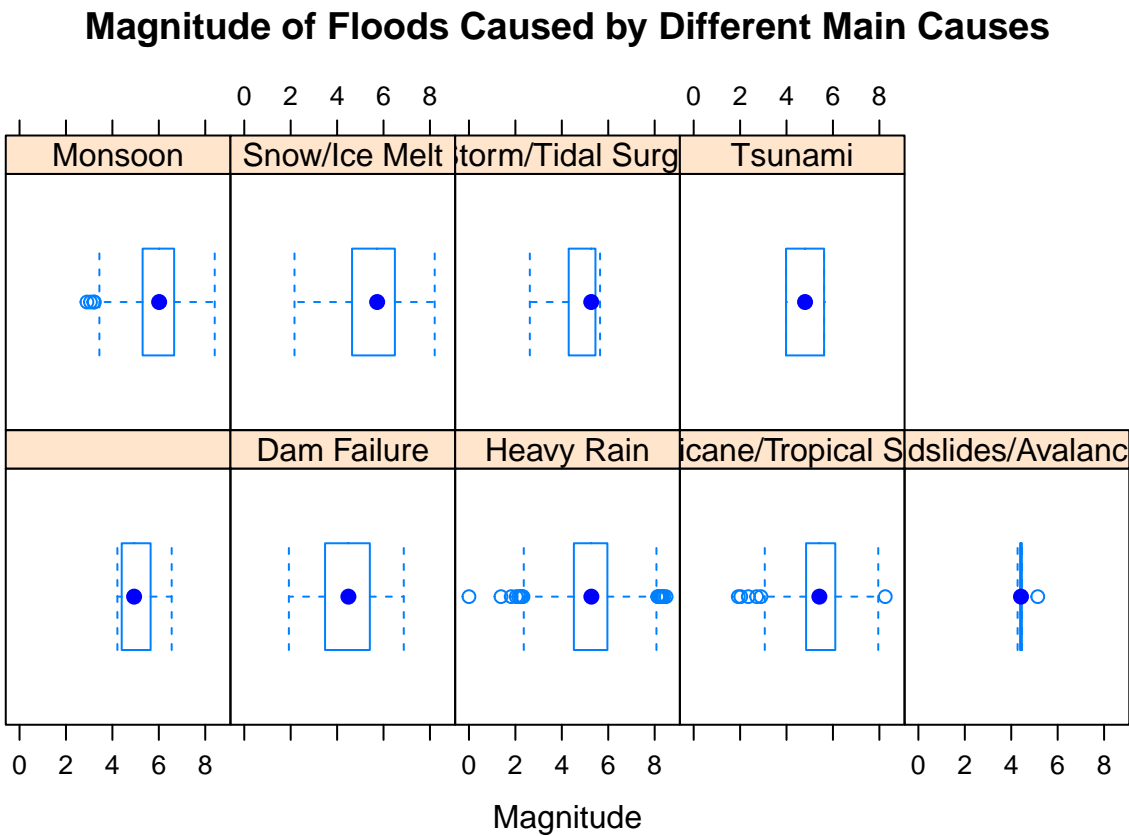


Subregions that displaced more people as the size of affected region increased were South America, Eastern Asia, and Southern Asia. Subregions that displaced very few people regardless of the size of the affected area were western Asia and Australia and New Zealand.

Subregions that experienced more deaths as the size of affected region increased were Southern Asia and Eastern

Asia Subregions that experienced few deaths regardless of the size of the affected area were Western Asia and Austria/New Zealand.

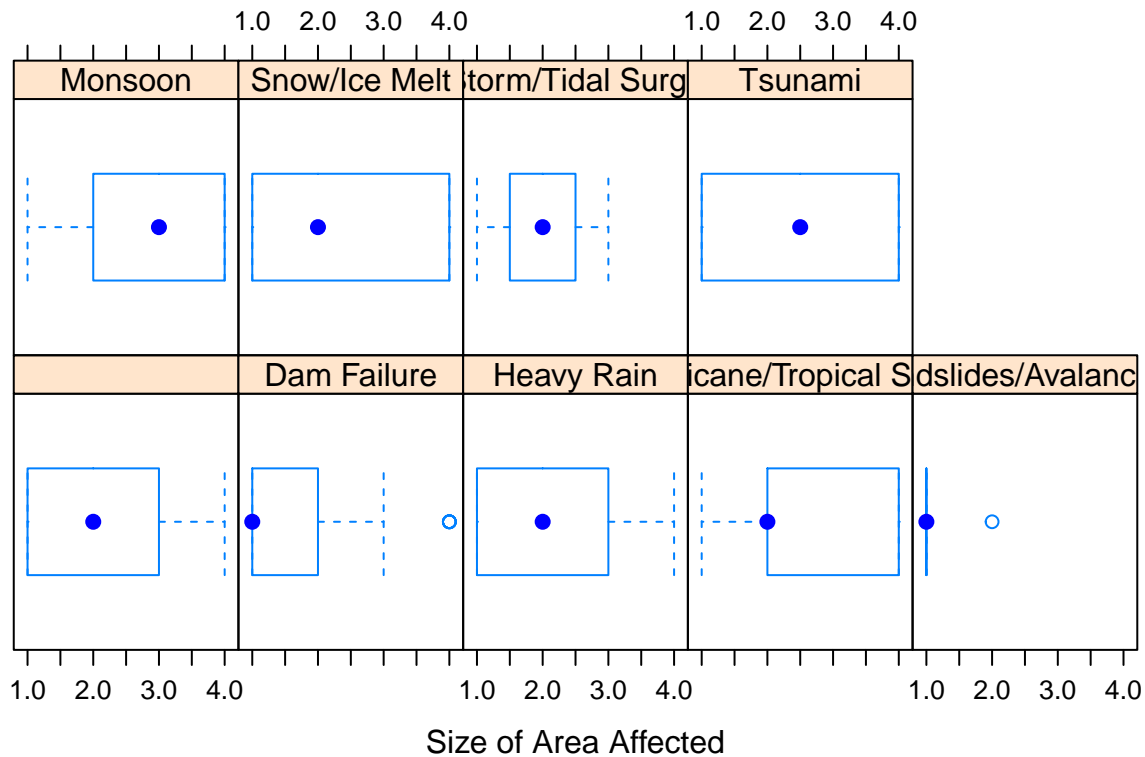
Magnitudes by Floods of Different Causes



While Landslides and Avalanches seem to cause floods only between magnitudes of 4 and 5, Heavy rain seems to cause floods of varying degrees of magnitudes.

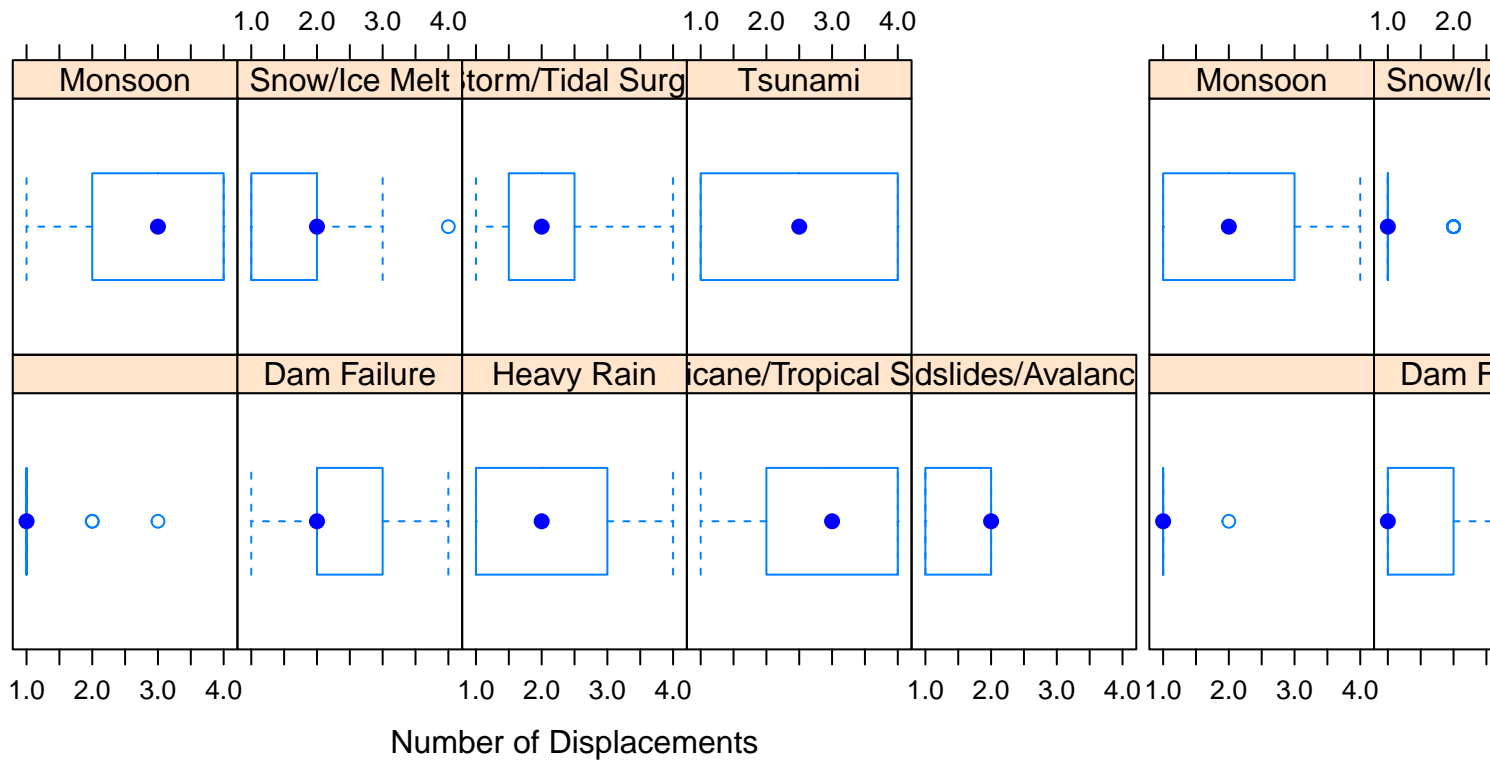
Size of Affected Areas by Floods of Different Causes

Size of Area Affected by Different Main Causes



Snow/Ice Melt and Tsunami can affect a wide array of sizes of lands. Landslides/Avalanches seem to usually affect small areas, and Monsoons seem to affect large areas.

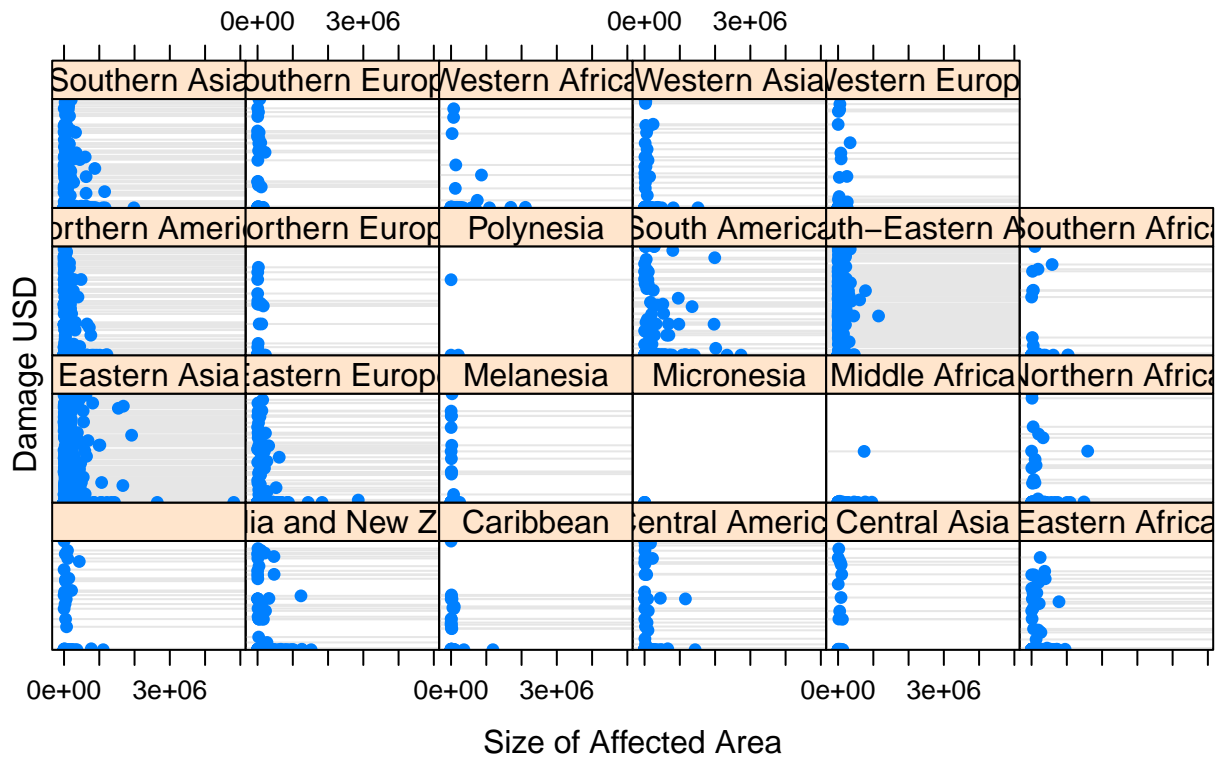
Number of Displaced People by Main Causes



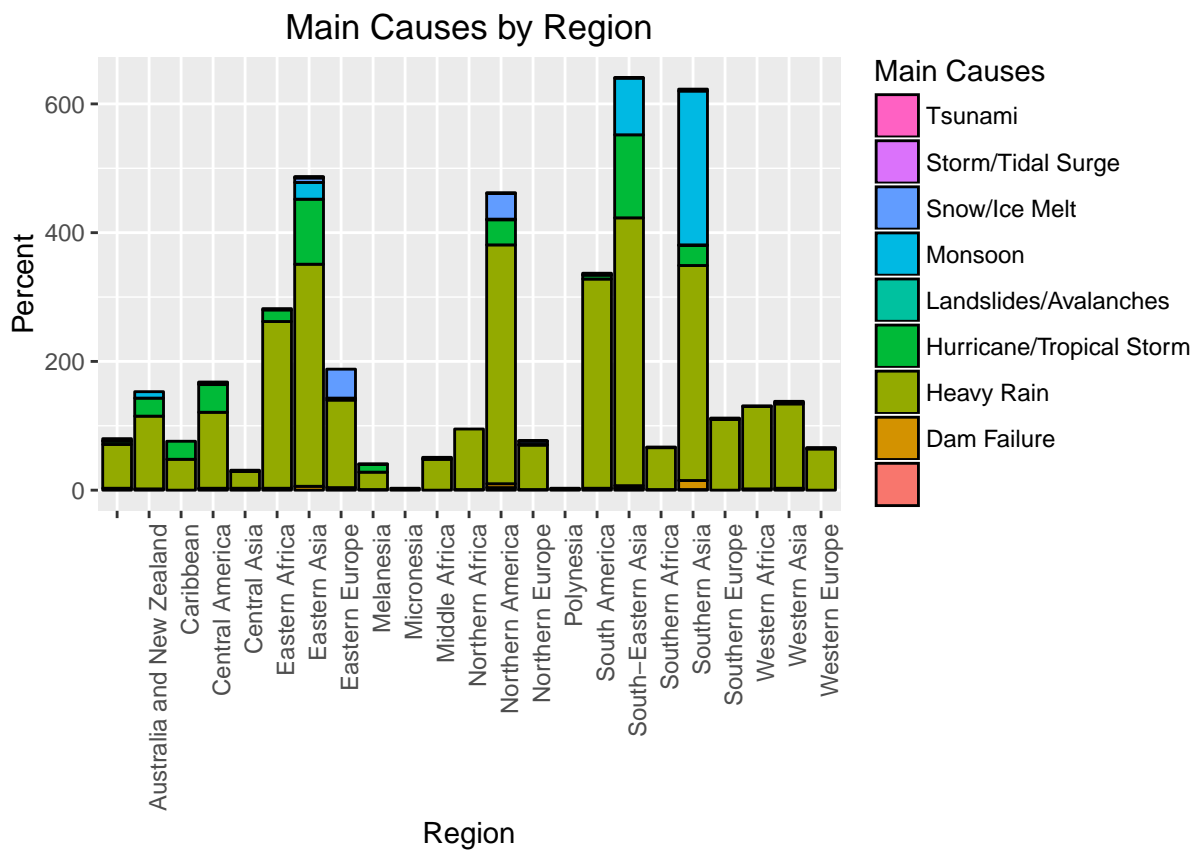
Monsoons and Hurricanes/Tropical Storm seem to displace many people while Landslides/avalanches and snow/ice melt seem to usually displace fewer people. Tsunamis always seem to kill many people, while snow/ice melt seems to always kill only few people.

Damage Caused by Floods

Damage USD by Size of Affected Area Grouped by Region



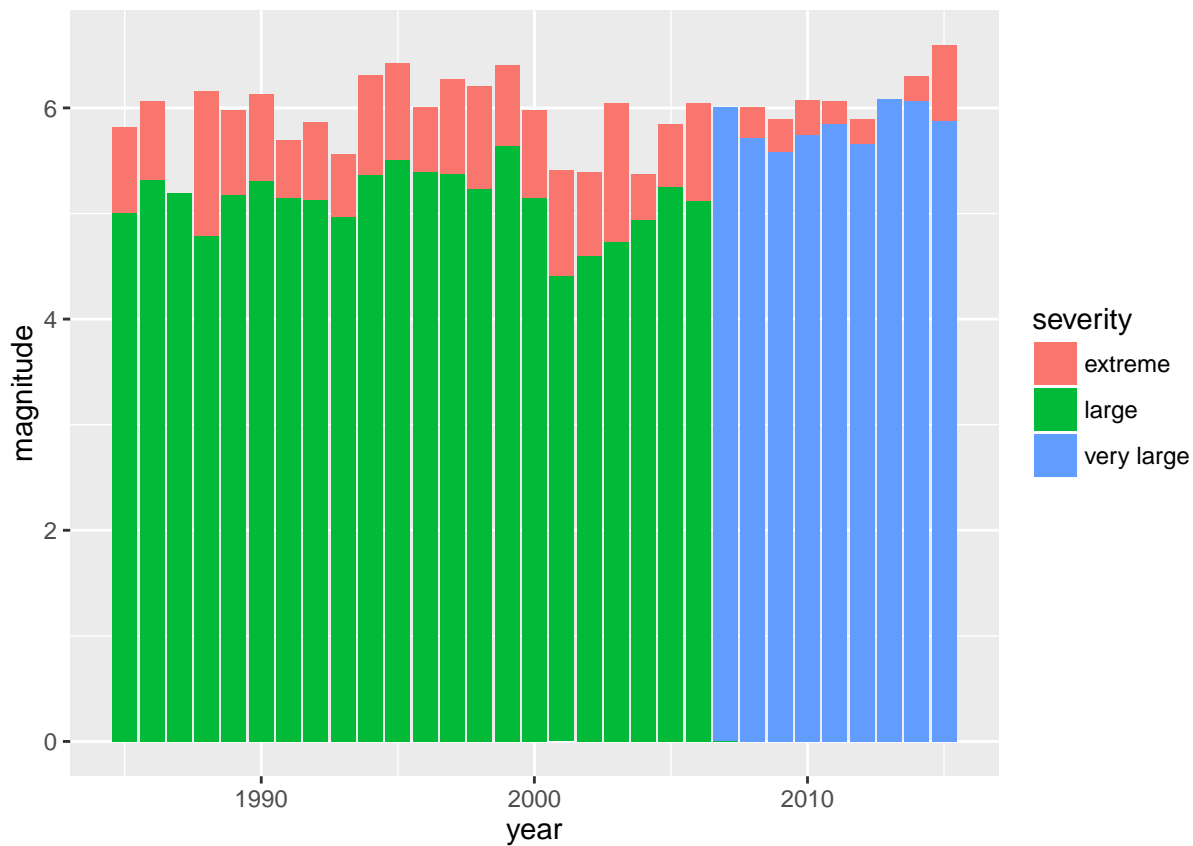
Northern America and Northern Europe seem to have very high starting point of damage in USD probably because since the GDP is high, when damage is done, it is more expensive to recover from the damage. Middle Africa seems to have the lowest starting point in terms of damage in USD.



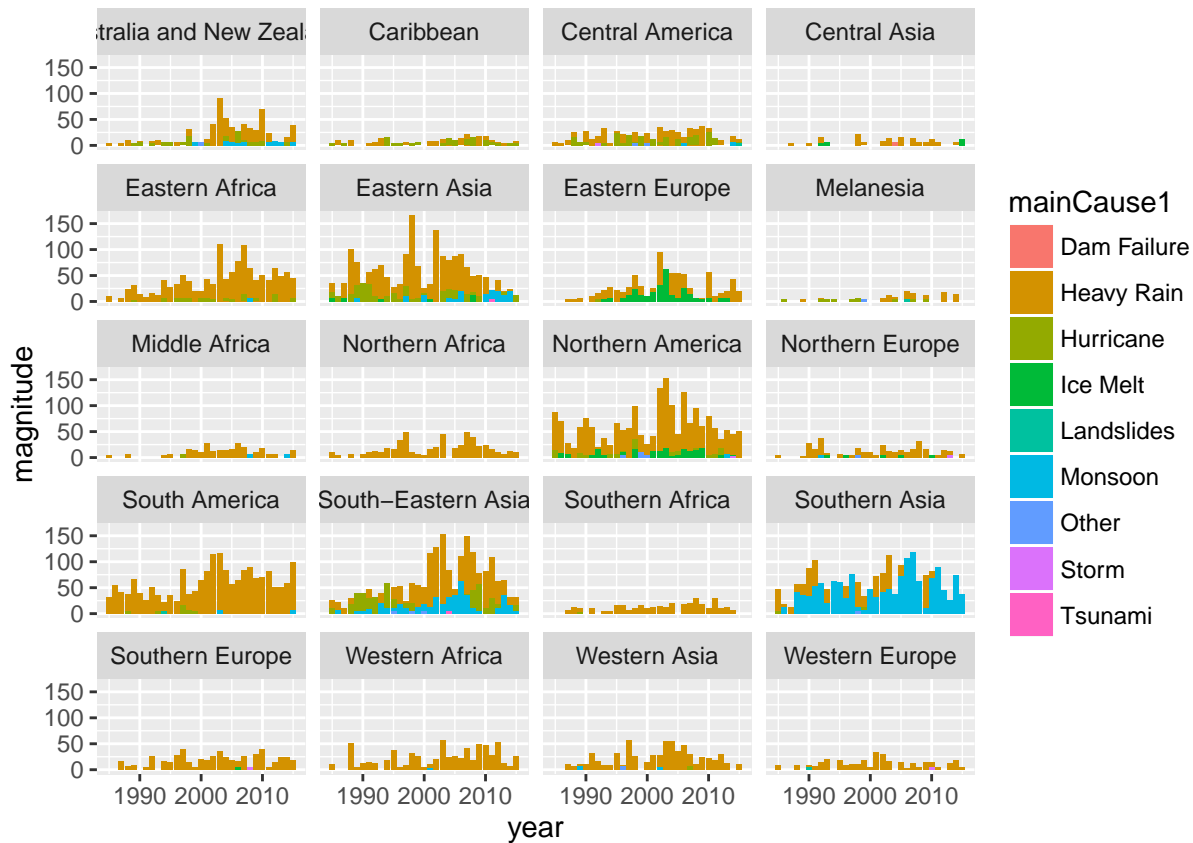
South-Eastern Asia, followed by Southern Asia and then Eastern Asia and Northern America, experience the most number of floods. For each region, over 50% or more floods are caused by heavy rain. In the case of Southern Asia, a large portion of floods are caused by Monsoon. Hurricane/Tropical Storms seem to happen quite often in South-Eastern Asia and Eastern Asia.

Impact evolution

Let's first take a look at the evolution of flood in time



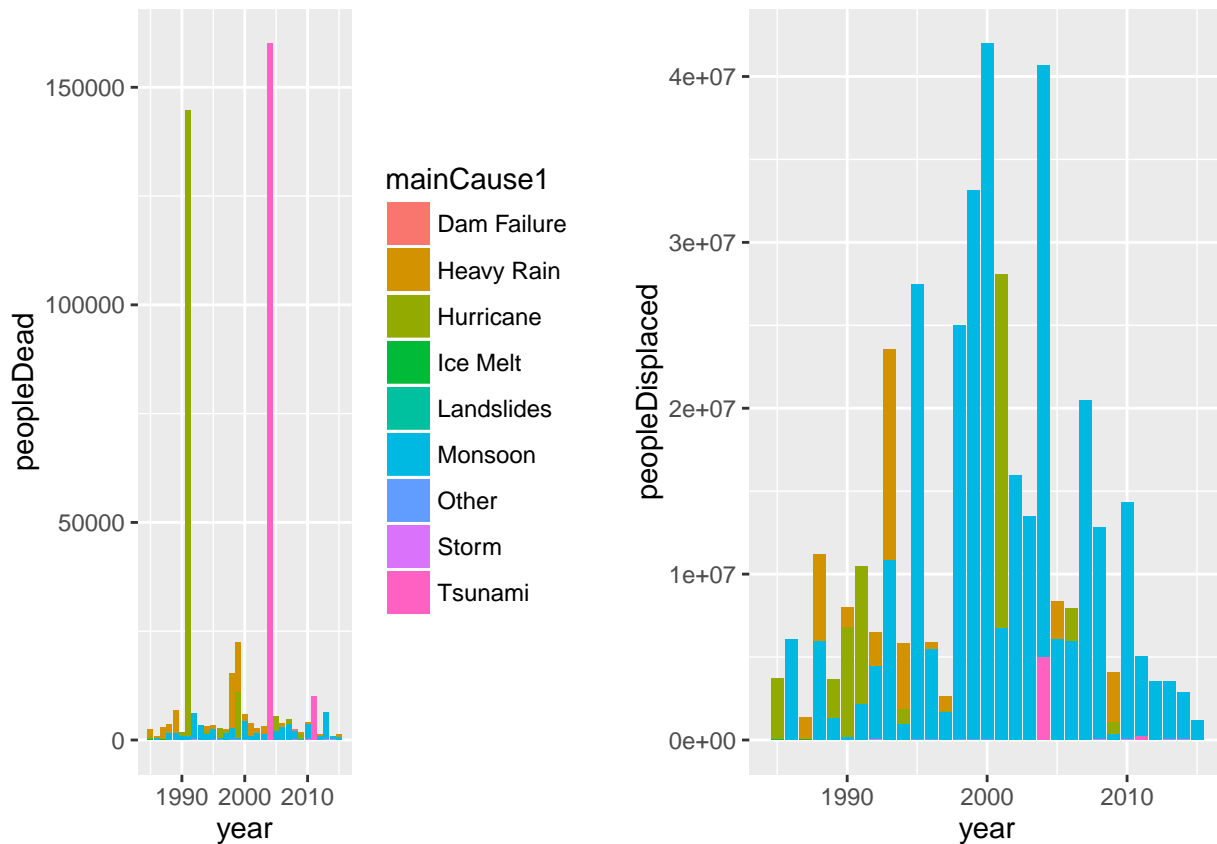
We note that there is an augmentation of the severity of flood in the last decade. But there is no specific cause that can explain it.



When we analyze the repartition of floods by region, we note that there are some region more affected and that some region are more targeted by specific type of disaster. For example:

- Eastern Europe and North American are more touched by Ice Melt
- Southern Asia is highly affected by Monsoon.
- Central America is mainly touched by hurricane.

Impact on population

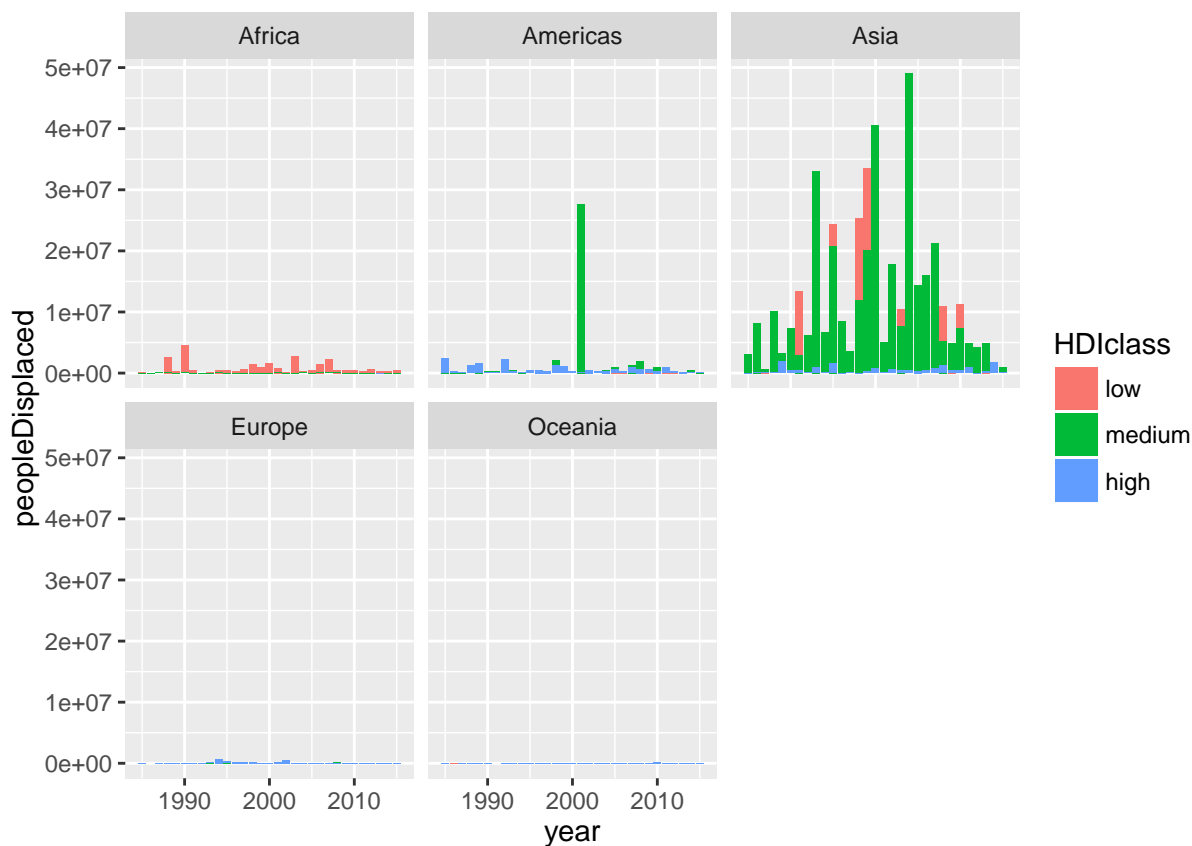


Based on these 2 graphs, we note 2 things:

- Some floods have terrible impact on human lives. For example, in Thailand, in 2004 when 160,000 people died in a tsunami. Therefore, we assume that the number of dead is not linked to the characteristic of the country.
- When we look at the number of people displaced, we see a very different pattern. It is quite stable and the main event that push people to move is Monsoon. Furthermore, we know that monsoon tend to happen in a very specific region of the world, mainly south east Asia.

Characteristics of country and floods' impact

In what type of countries does flood have a greater impact on population? To answer this question we will gather and merge data about Human development index (HDI), life expectancy, expected number of school year, Gross National Income (GNI) per capita.



When we look at people displaced, we see that it mainly affect Asian and African countries and mainly countries with low (below 0.6) or medium (between 0.6 and 0.75) HDI.

Regression Model

```
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)  -582.5722965  686.734554  -0.8483224  0.39630772
## HDI          -1532.8972850  777.478162  -1.9716274  0.04871894
## lifeExp      25.8339559   15.253336   1.6936594  0.09040560
## GNIPerCapita -0.7485036    1.219345  -0.6138573  0.53934354
```

The regression analysis confirm that country characteristics does not impact the number of dead people per flood.

```
##
## Call:
## lm(formula = peopleDisplaced ~ HDI + lifeExp + GNIPerCapita,
##     data = flood2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -607416 -184480 -116856   -3324 39738981
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -688620.7   263012.8  -2.618  0.00887 **
## HDI          -2024673.1   297766.7  -6.800  1.20e-11 ***
## lifeExp       30766.7     5841.9    5.267  1.46e-07 ***
## GNIPerCapita    808.3      467.0    1.731  0.08354 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1272000 on 4121 degrees of freedom
## Multiple R-squared:  0.01174,    Adjusted R-squared:  0.01102
## F-statistic: 16.31 on 3 and 4121 DF,  p-value: 1.541e-10
```

There is a true correlation between the stage of development of a country and the number of displaced people during floods.

Main finding

When it comes to death toll, no country is protected against a huge event as a huge tsunami or a hurricane. But facing heavy rain, developed country have better infrastructure and a greater capacity to take care of the people touched by such event. They also have the ability to quickly repair in order to make the population suffer a minimum time so that they don't have to move.