**COVID-19 Findings Summary**

**Method:**

1. Data resources:

a. COVID-19 dataset : data repository posted on GitHub from JHU,<https://github.com/CSSEGISandData/COVID-19>,

b. Weather Temperature data from Open Weather API<https://openweathermap.org/api>,

c. Population density data from NASA, Gridded Population of the World (GPW), v4,<https://sedac.ciesin.columbia.edu/data/collection/gpw-v4>

2. Statistical method: statistic summary was used to describe the demographic and socio-economic characteristics of COVID-19. weather temperature groups were defined as : “H Temp” if weather temperature is >=72 F which is 75% quantile of this weather dataset, “L Temp” if weather temperature is less than 43 F which is 25% quantile of this weather dataset, “M Temp” if weather temperature is between 43f and 72 F. population density groups were defined as : “H density” if population density of the location is greater than 75% quantile of the population density dataset, “L density” if population density of the location is less than 25% quantile, “M density ” is between 25% and 75% quantile.

the number of confirmed cases were transformed using log base 2 method before doing any analysis. The COVID-19 transmission rate for each location was calculated by applying a linear regression model to daily time series of confirmed cases in each location. one way Anova was applied to detect if there are significant differences in log2 of total number of confirmed cases and transmission rate of COVID-19 among groups if data were normally distributed, otherwise Kruskal wallis tests were applied.

**Question 1:**

**What are the world-wide statistics related to COVID-19 for confirmed cases, deaths and recovery?**

Our data for the summary statistics comes from the data repository posted on GitHub for the 2019 Novel Coronavirus Visual Dashboard operated by the Johns Hopkins University Center for Systems Science and Engineering, specifically the COVID-19 Daily Report and we have updated with the most recent data from March 20th. The data from Johns Hopkins was clean already, so that was a considerable plus.

As of March 20th, the total confirmed cases of COVID 19 world wide stands at 274,180 with 11,375 deaths. After conducting a summary analysis of Johns Hopkins’ data, we pulled the 5 countries with the highest rates of confirmed cases and produced a grouped bar chart (Image 1 below) to compare them by the number of confirmed cases as well as their number of deaths and recovered cases. Here, we noticed that Italy’s deaths were higher than China’s and wanted to take a closer look at those 2 countries.

Image 2 shows two pie charts representing the breakdown of confirmed cases for both China and Italy. We found that Italy’s percentage of deaths is more than double China’s. What is most alarming here is that you can see China’s confirmed cases consists mostly of recovered cases with only 8.35% unresolved, while Italy still has 82% of its total cases unresolved, meaning most are still sick and so the death toll has the potential to rise significantly.

Image 1:

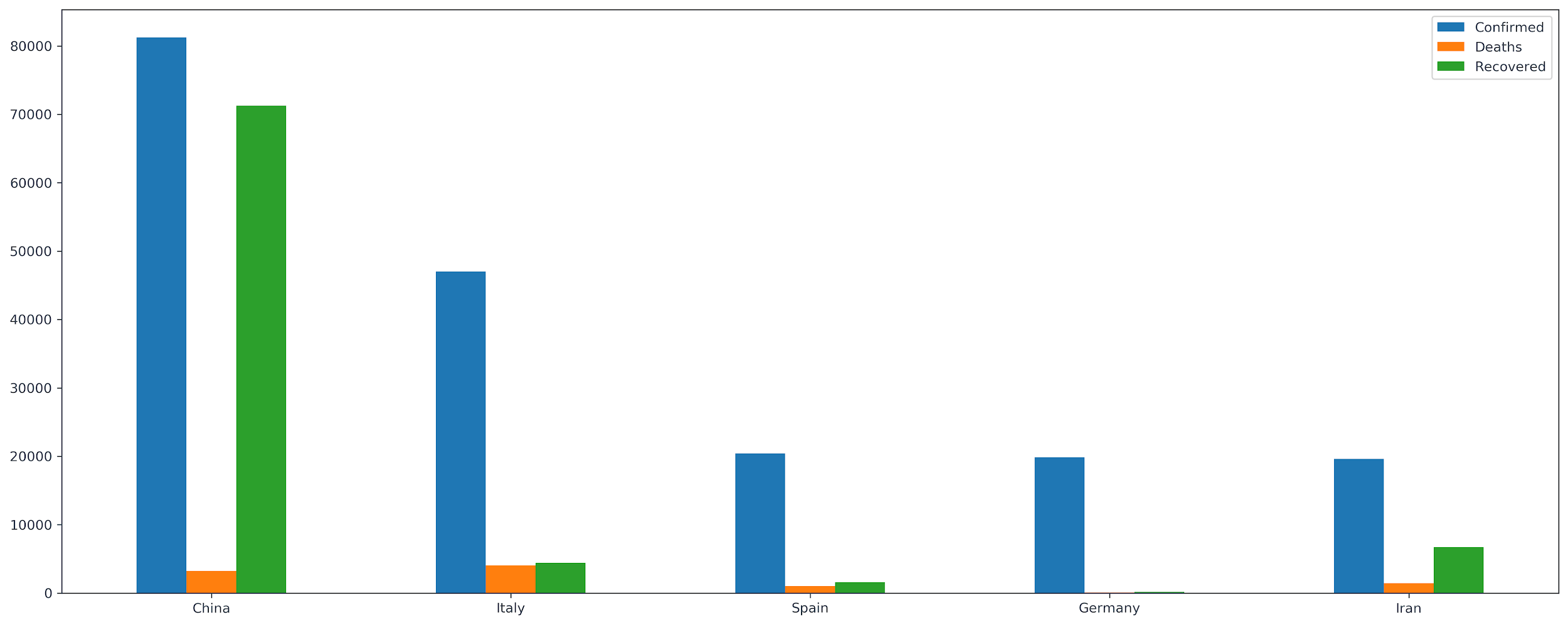
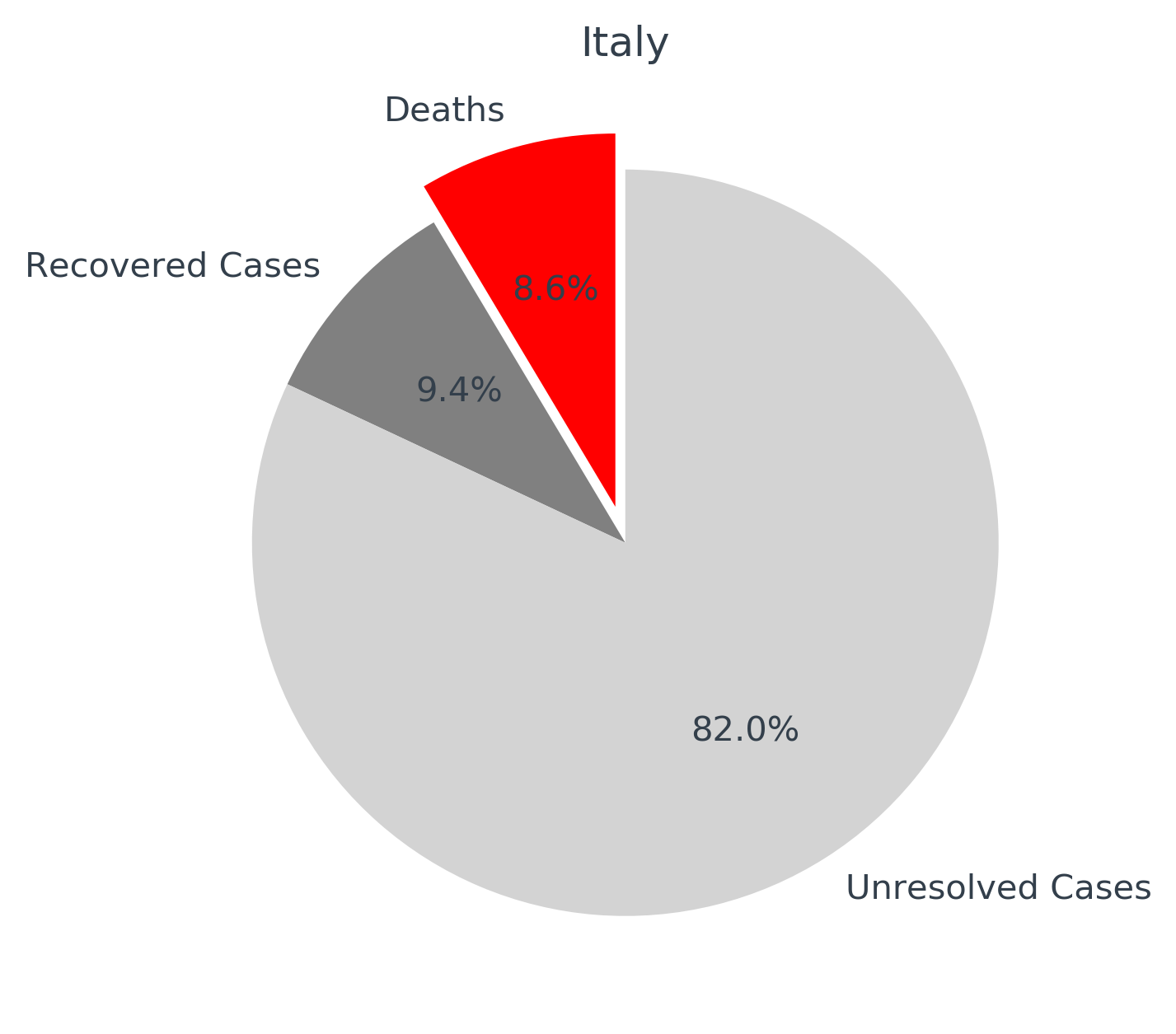
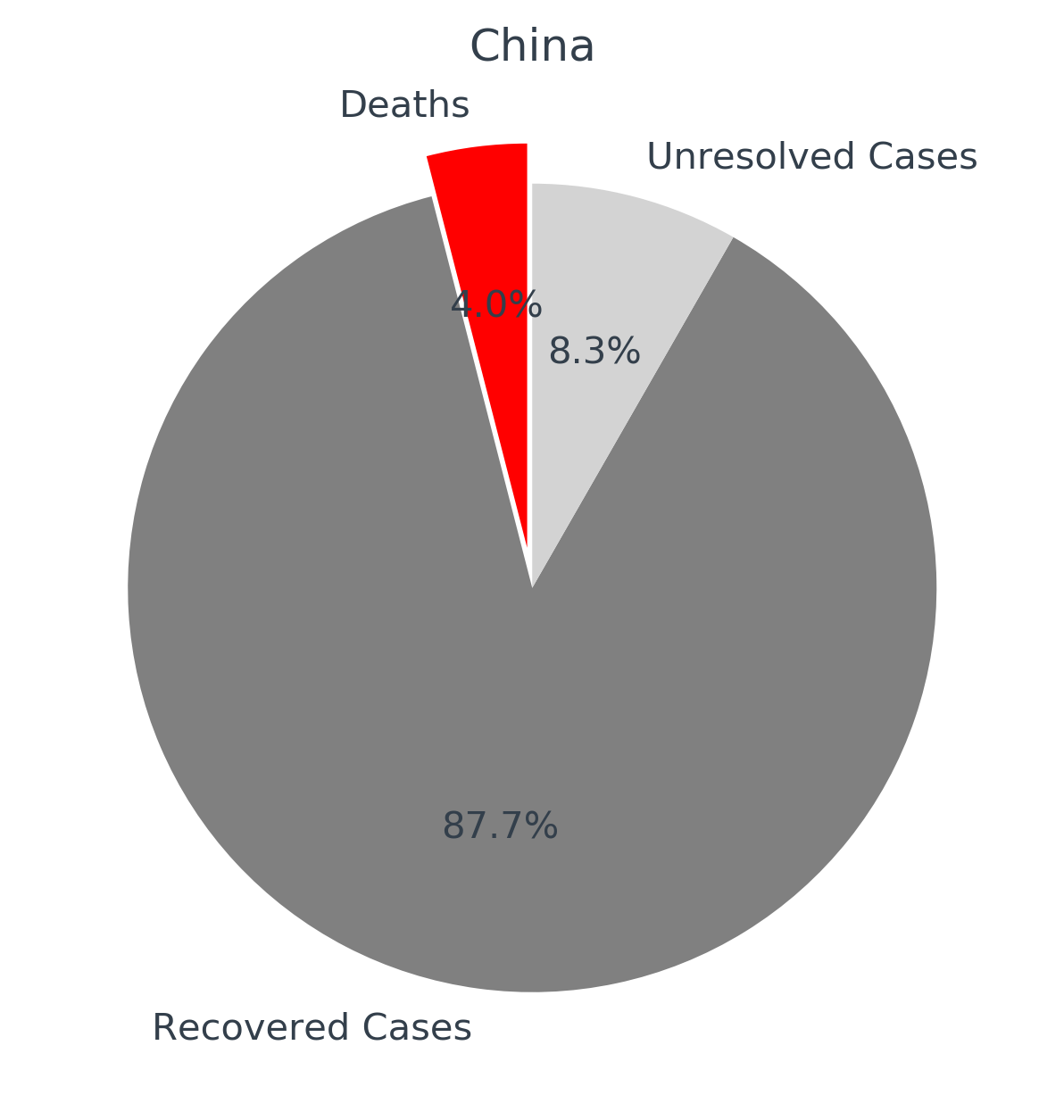
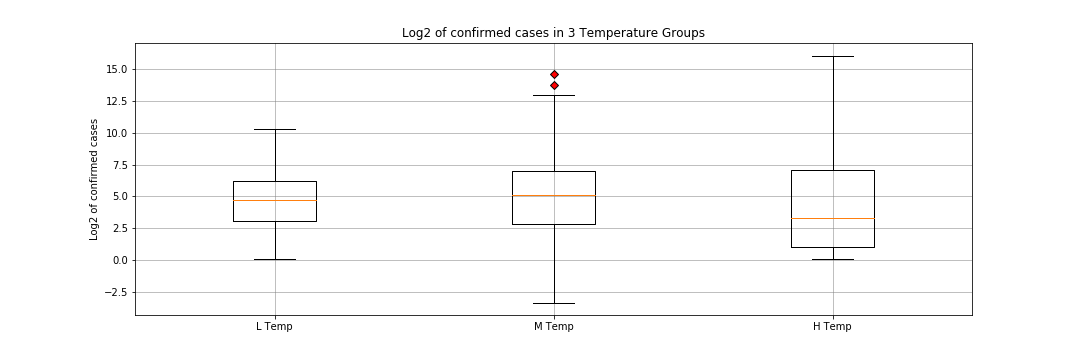


Image 2:

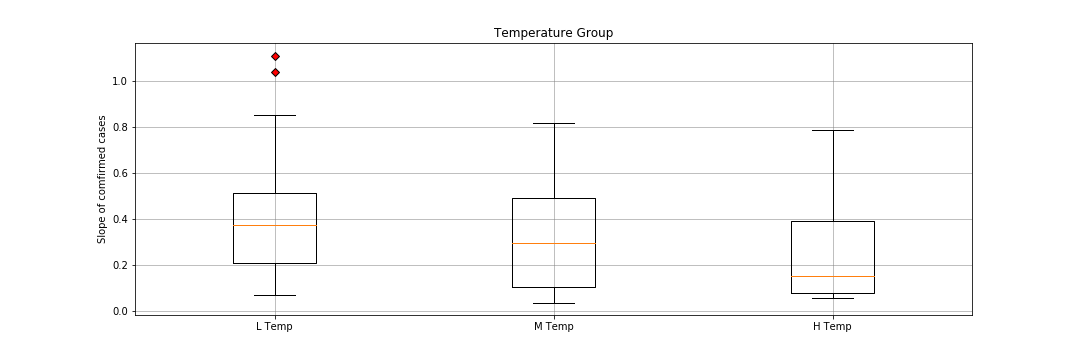


**Question 2:**

**Does weather temperature have an impact on COVID-19 :**



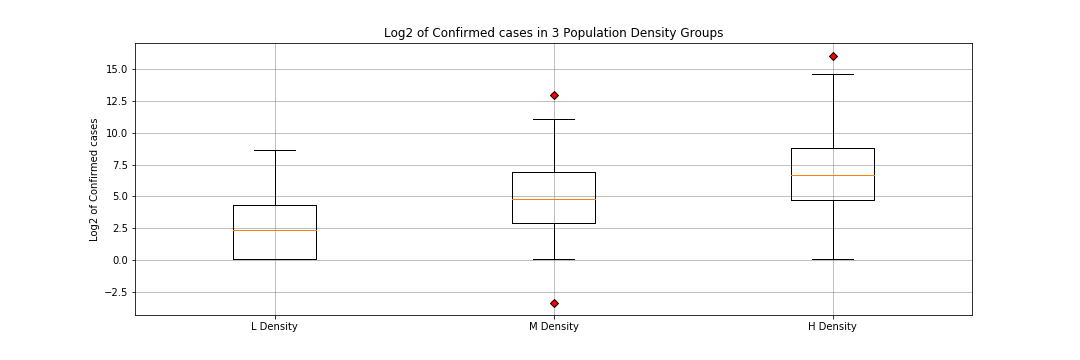
High weather temperature group (>73 F) has significantly lower confirmed cases than Intermedia and Low temperature groups, P <0.0001, Kruskal wallis test.



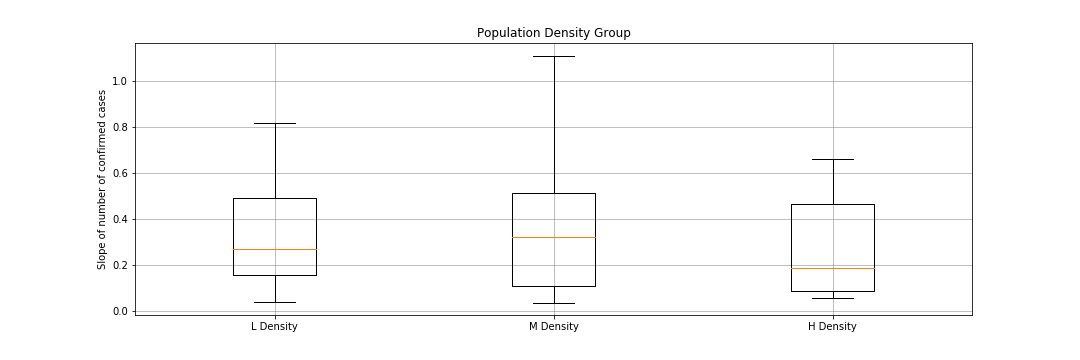
High weather temperature group (>73 F) has a significantly lower transmission rate than Intermedia and Low temperature groups, P =0.023, Kruskal wallis test.

**Question 3:**

**Does population density have an impact on COVID-19 :**



High population density group has significant higher confirmed cases than Intermedia and Low temperature groups, P <0.0001, Kruskal wallis test.



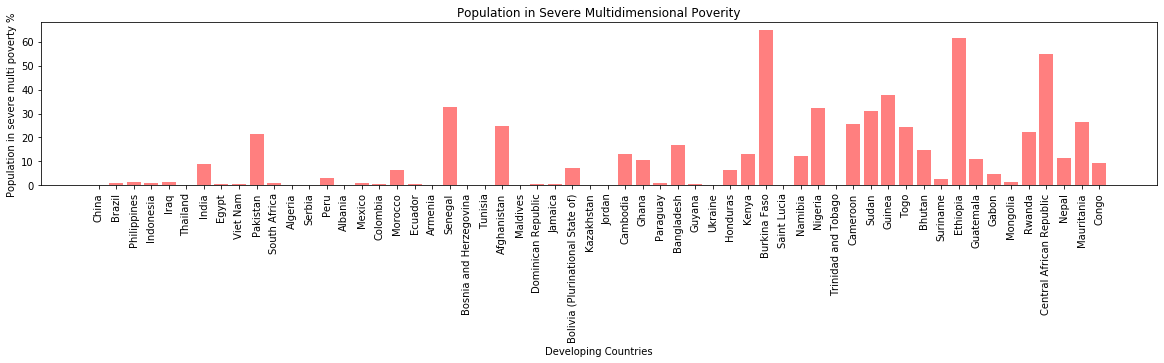
There is no significant difference found among different population density groups, P=0.3, Kruskal wallis test, possible explanation is, other factors are playing a more important role in occurrence and transmission of COVID-19, like social distance between persons, the containment methods used by individual local governments, the distribution of high risk factor like age, more sick people in the locations.

**Question 4:**

**Socio-economic background of those affected**

Our analysis showed that developed countries are the ones more affected by COVID-19 than the developing countries. 42% percent of the developing countries have not reported any confirmed cases. Only 1 of the 58 developing countries (China) is in the top 20 affected countries and it’s multidimensional poverty percentage is very small compared to many of the other developing countries. 53 % of the developed countries have reported confirmed cases.

Text(0, 0.5, 'Population in severe multi poverty %')



**Question 5:**

**Correlation between air travel and the spread of COVID-19**

Looking at the location of the confirmed cases and also looking at the location of the top 20 major world airports, there seems to be some correlation between major airports such as several around China and around France and Germany but there isn’t a significant correlation. Most of the confirmed cases in the US, for example, are on the east and west coasts versus around the major airports.

