Tokyo Olympics 2020 versus HDI Index

Pawan Kumar

11/2/2022

Roll No: MDS202224
Visualization
Email: pawan@cmi.ac.in
Chennai Mathematical Institute

Main Subject

The report examined the medal tally of developed and developing countries participating in the Tokyo Olympics 2020. Based on their(countries) HDI index, this report aims to show a relationship between the quality wise (Gold, Silver, and Bronze) or quantity wise (total no. of medals) of medals in the Tokyo 20'olympics for nations.

Abstract

Analysis of the HDI index versus the nations through various visualization processes.

The various plots reveal a positive yet not linear relationship between the HDI index and nations' medal tally. Countries in a specific range of HDI indexes, both developed and developing countries, earned the most medals in the Olympics.

Data Description

The initial dataset contains five attributes: Countries, Gold, Silver, Bronze, and Total, and includes only those nations which have won any medal in the Olympics. Then, the countries were classified as either developed or developing based on the HDI index 2021; if the HDI index of a particular country is >= 8.0, then it is classified as a Developed country, otherwise as a Developing country. There is a total of 93 records(countries) in the data. The dataset is arranged in decreasing order of the number of medals, preferred by gold, silver, and bronze in the same order. The new dataset consists of 93 countries' records stating their names,Rank, gold, silver, bronze, and total medals earned by them, their HDI index, and their classification as developed or developing countries.

- *X* : Rank of the Country based on their medal tally
- *Country.Portuguese.* : Name of countries in the Portuguese language
- *Country.english.* : Name of countries in the English language
- *Gold* : Number of gold medals bagged by the countries
- *Silver* : Number of silver medals bagged by the countries
- *Bronze* : Number of bronze medals bagged by the countries
- *Total* : Number of Total medals bagged by the countries
- *HDI* : classifies countries as Developed(D) or Developing(N)
- *HDI.index* : HDI index of a particular country

Introduction

The 2020 summer Olympics, also knowns as Tokyo Olympics 2020, was the 32nd Olympics, held in the heart of Tokyo, Japan, from 23 July 2021 to 8 August 2021 (postponed due to the Covid-19 pandemic).

203 nations, including EOR and ROC teams, participated in these games in 339 events in 33 sports(50 disciplines), Out of which only 93 countries got a chance to share the podium.

Out of these 93 nations, 51 Developed and 42 Developing countries were able to make it to the medal tally based on the HDI index 2021. China is the only Developing country(placed 2nd after the USA) able to make it to the top 10 nations in the medal tally with an HDI index of 0.768. While Bulgaria, with the highest HDI of 0.795 among the developing countries, placed 30th in the count. Likewise, Berhhda could not reach the top 62, ranked 63 with the highest HDI index of 0.981.

This report analyses developed and Developing countries from the perspective of medal tally through various visualization methods.

The report investigates the relationship between the HDI index of nations to their medal tally.

Pictorial Display

```
pawan <- read.csv(file.choose())
```

```
pawan <- pawan[rowSums(is.na(pawan)) == 0,]
```

```
library(ggplot2)
```

```
## Warning: replacing previous import 'ellipsis::check_dots_unnamed' by
## 'rlang::check_dots_unnamed' when loading 'tibble'
```

```
## Warning: replacing previous import 'ellipsis::check_dots_used' by
## 'rlang::check_dots_used' when loading 'tibble'
```

```
## Warning: replacing previous import 'ellipsis::check_dots_empty' by
## 'rlang::check_dots_empty' when loading 'tibble'
```

```
library(plyr)
```

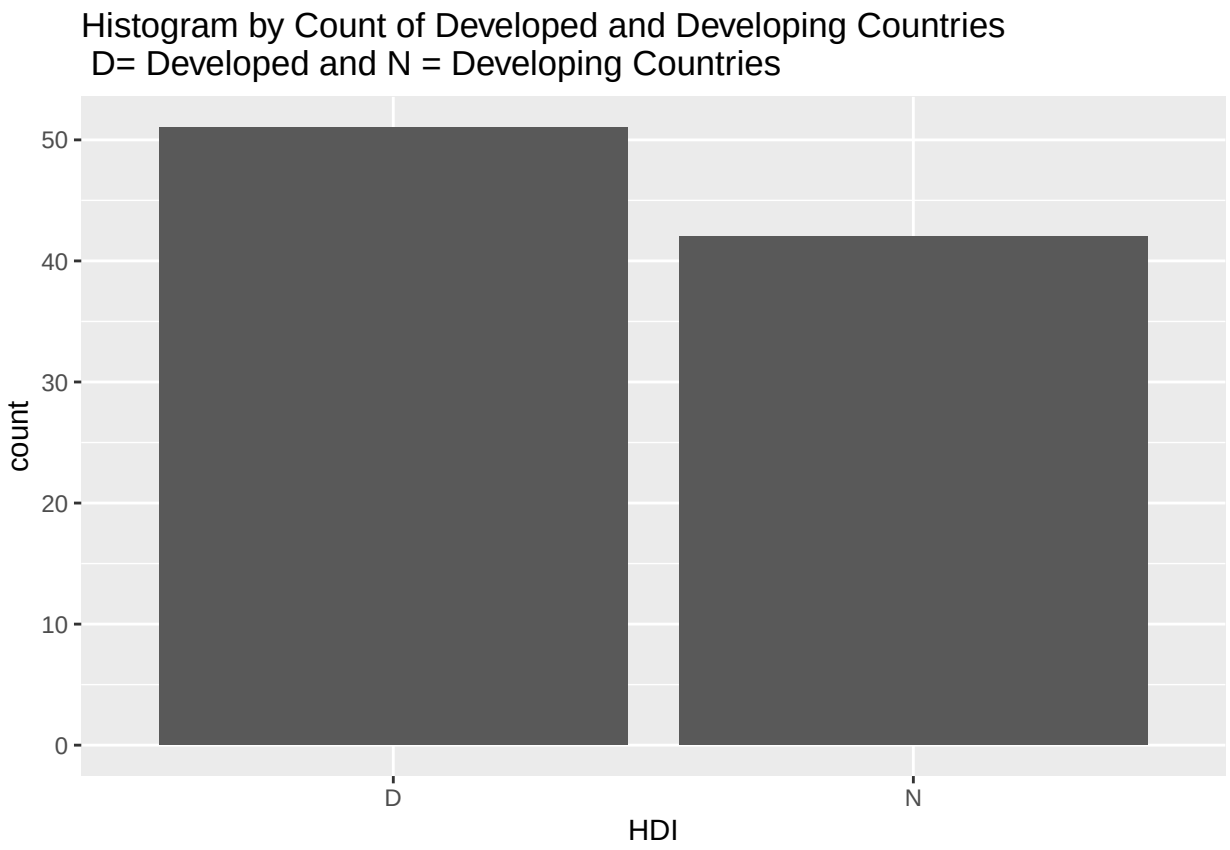
```
## Warning: package 'plyr' was built under R version 4.0.5
```

```
library(Rcpp)
```

```
mu <- ddply(pawan, "HDI", summarise, grp.mean=mean(Total))
```

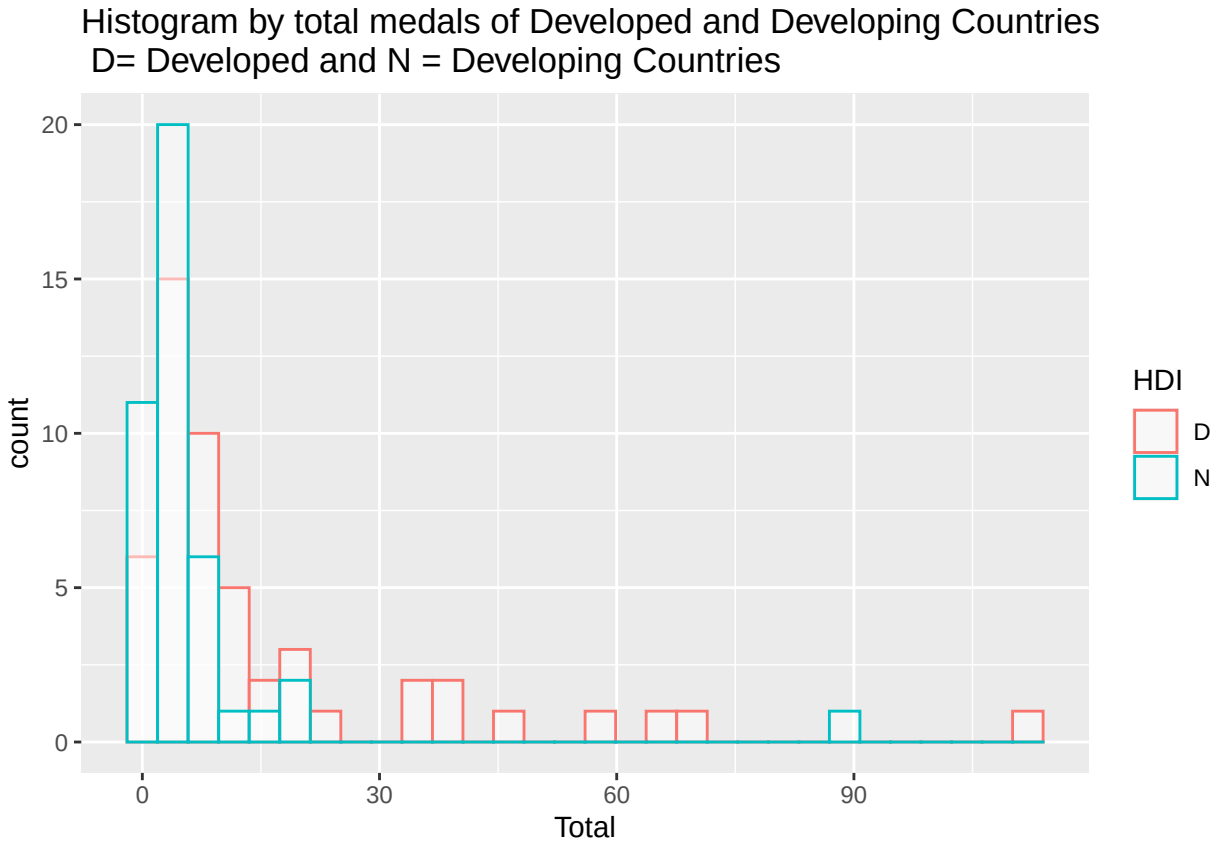
```
ggplot(pawan,aes(HDI))+geom_histogram(stat='count')+ggtitle("Histogram by Count of Developed and Develop
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

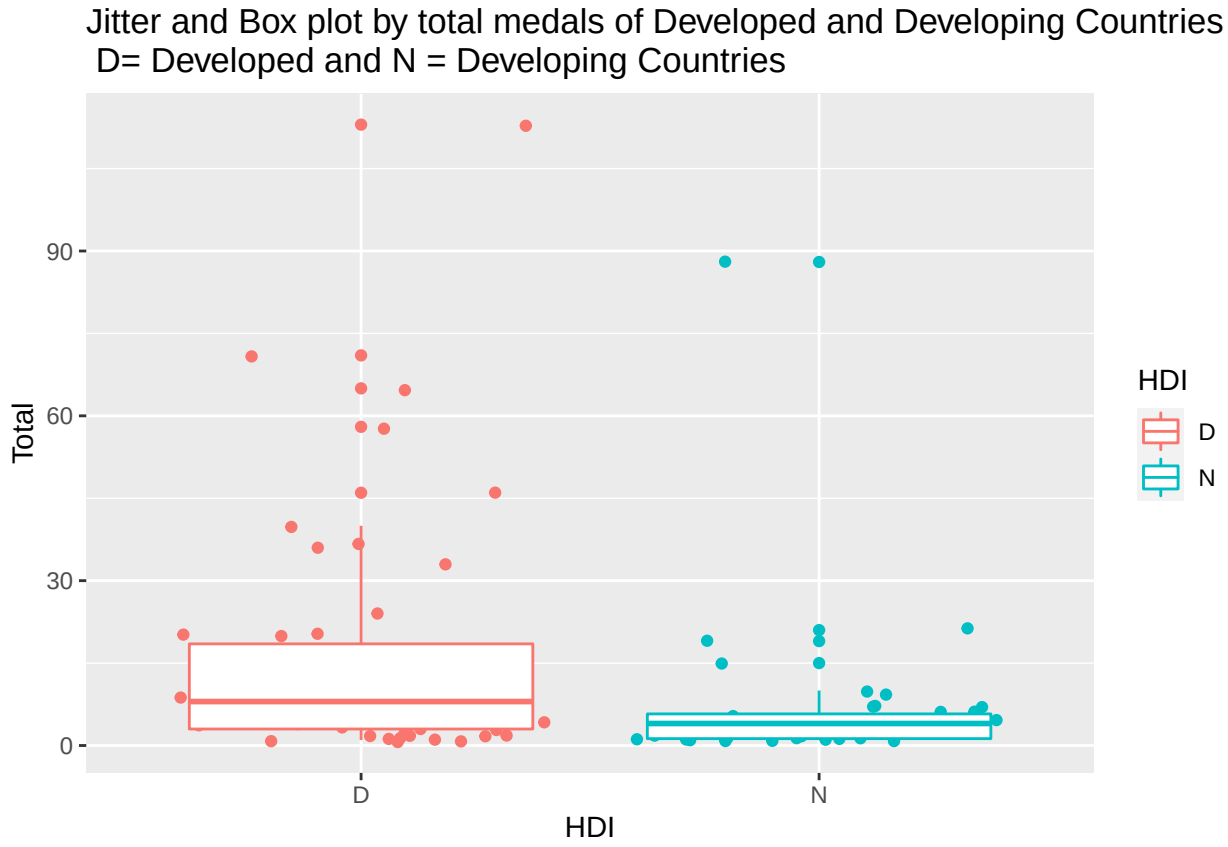



```
ggplot(pawan,aes(Total,color=HDI))+geom_histogram(fill="white", alpha=0.5, position="identity")+ggtitle
```

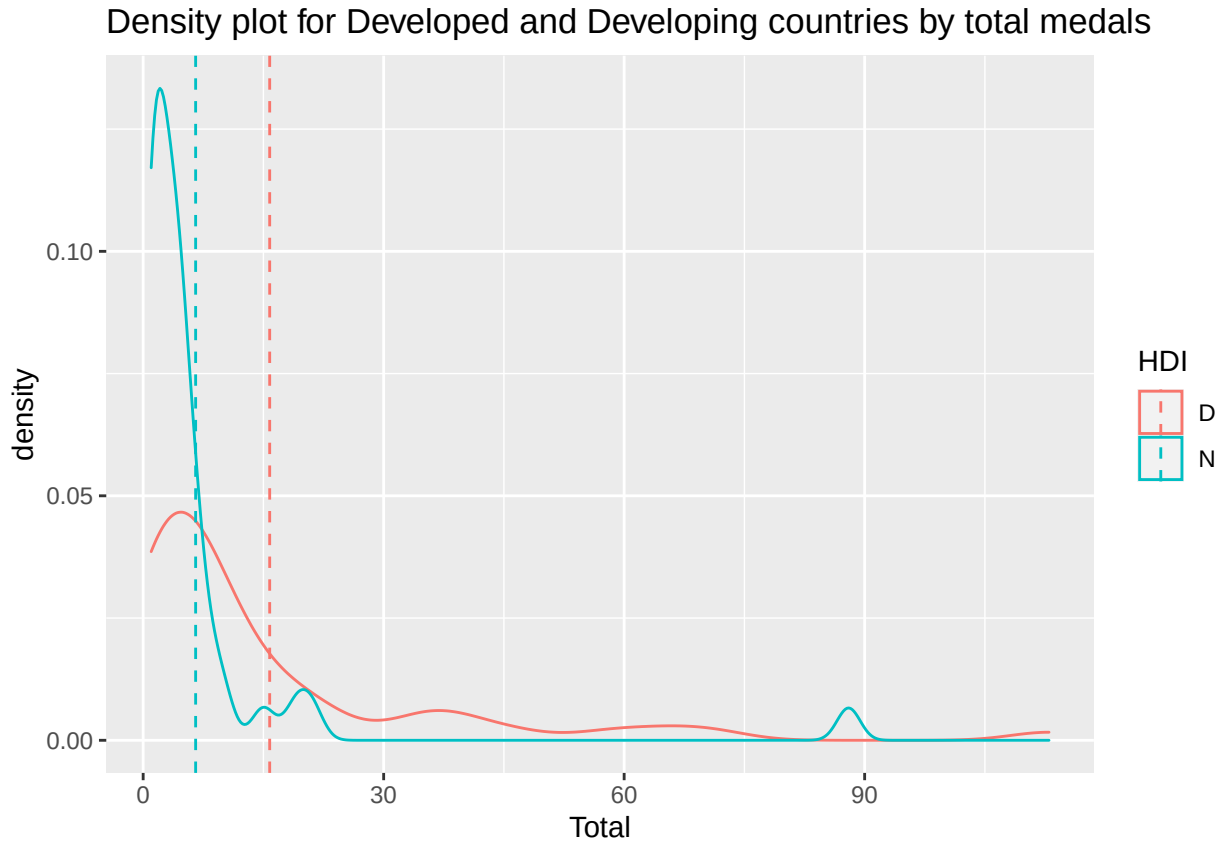
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



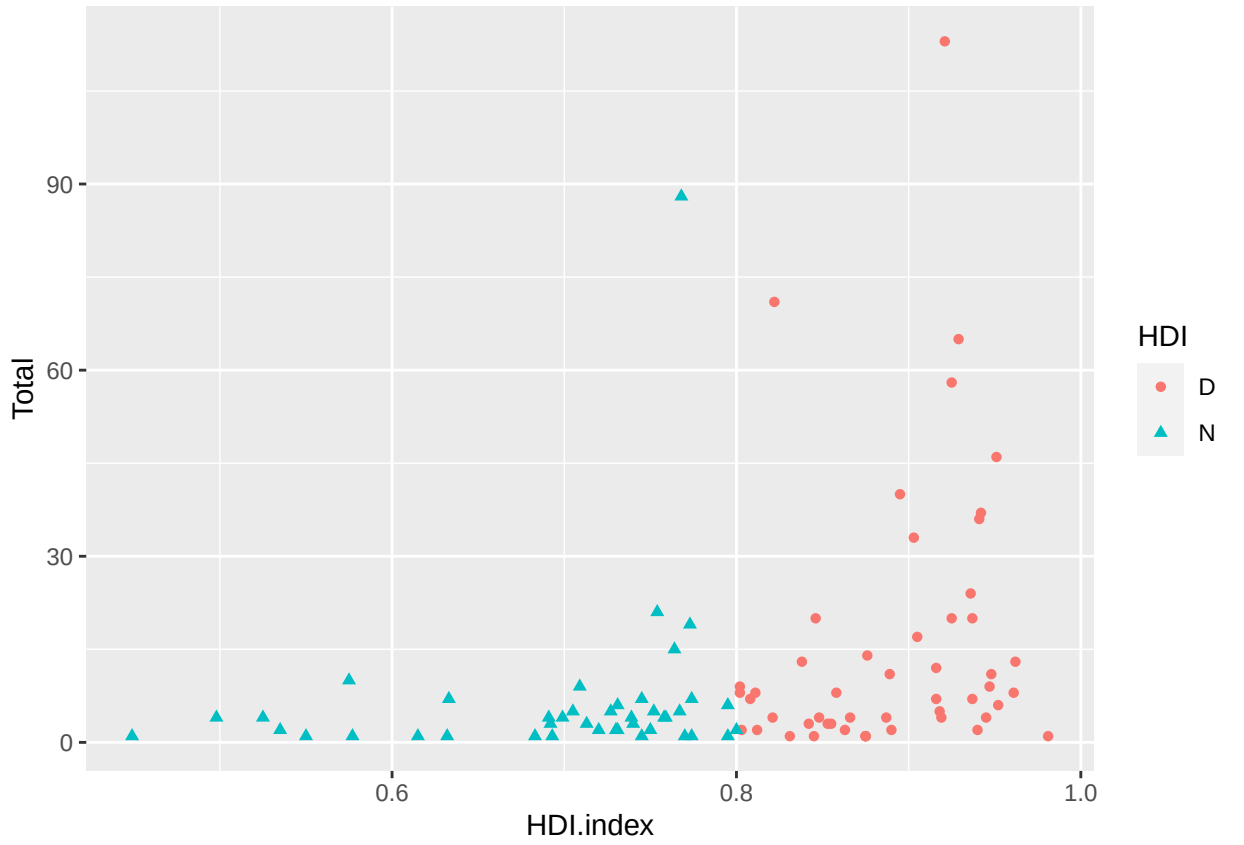

```
ggplot(pawan,aes(HDI,Total,color=HDI))+geom_jitter()+geom_boxplot()+ggtitle("Jitter and Box plot by tota
```




```
ggplot(pawan, aes(Total, color=HDI)) +geom_density()+geom_vline(data=mu, aes(xintercept=grp.mean, color=
            linetype="dashed")+ggtitle('Density plot for Developed and Developing countries by total me
```



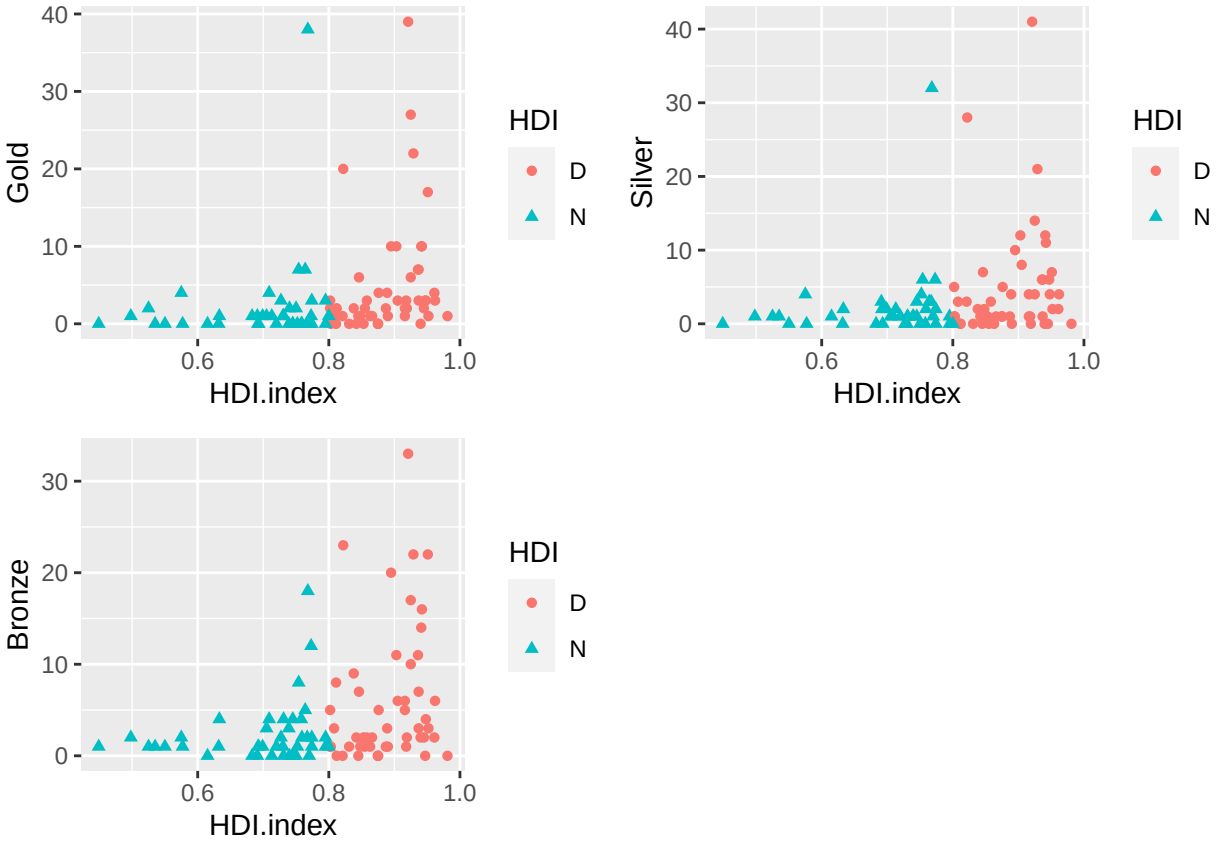

```
ggplot(pawan,aes(HDI.index,Total,color=HDI,shape=HDI))+geom_point()
```




```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.0.5
```

```
plot1 <- ggplot(pawan,aes(HDI.index,Gold,color=HDI,shape=HDI))+geom_point()
plot2 <- ggplot(pawan,aes(HDI.index,Silver,color = HDI,shape=HDI))+geom_point()
plot3 <-ggplot(pawan,aes(HDI.index,Bronze,color = HDI,shape=HDI))+geom_point()
grid.arrange(plot1,plot2,plot3,ncol =2)
```



Analysis

The Above pictorial representation shows that -

- There is not hhch difference in the number of Developed and Developing countries which share the podium. There are 51 developed and 42 developing countries in total.
- Only one country (China placed 2nd) makes it to the top ten (From medal tally table).
  - There are 9 Developed and only one developing country with a Total medal greater than

25.

*The histogram and Boxplot (Total versus Developed and Developing countries) indicates that for developing countries(represented by N) median is hhch lesser than that of developed countries(represented by D). It is also the most concentrated because the interquartile range is smaller than the other. For developed countries, the difference between the quartile 3rd and the median is hhch higher than that of developing countries, indicating the distribution is positively skewed, or data has a longer tail towards higher values.

*The density plot shows that most developing countries haven't earned many medals. However, some outliers (like china) can be observed in the curve. The blue line shows the group means for Developing countries, which equals around 6.547 with a density value of roughly 0.04. Developed countries have a curve skewed toward the right direction and do not show anomalies like the formal one. The red line shows the group means for Developing countries, which equals around 15.784 with a density value of roughly 0.017.

*Scatter plots indicate that countries within the particular HDI index range of 0.875 - 0.96 (for Developed countries) and 0.7 - 0.8 (for Developing countries) bagged the most medals(including Gold, Silver, Bronze, and Total). Also, the respective plot for gold, silver, and Bronze shows similar trends as that of the Total medal plot for particular types of countries(either Developed or Developing).

Conclusion

The above Analysis indicates that the HDI Index and nations' medals show a positive relationship. However, It doesn't be interpreted as a linear relationship between the two variables. Since, most of the countries, for example, Berhhda, despite having the highest HDI index, couldn't perform well when it came to sharing the podium and placed 63rd overall.