Analysis of road traffic deaths from 1990 to 2019

In this experiment, we are going to look at and analyse a dataset that records the count of deaths from road accidents from 1990 to 2019 across all countries and categories of countries. The data is obtained from the Global Health Data Exchange portal at https://ghdx.healthdata.org/gbd-results-tool .

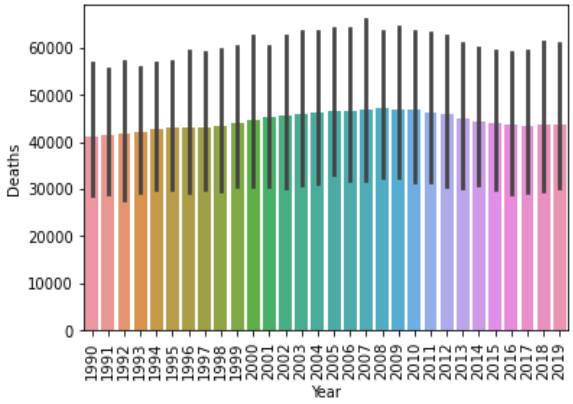
The data is further available in an extracted and pre-processed format at https://www.kaggle.com/shivkumarganesh/road-traffic-deaths-1990-to-2019.

The pre-processed dataset is first read into memory using Pandas. The dataset itself has four columns. These are - (a) Entity - The full name of the country/category of the recorded data, (b) Code - A three-letter code representing the entity, (c) Year - The year in which the data is recorded, (d) The total count of deaths from road accidents in the year in the entity.

We would assume that since the number of vehicles in the road has increased steadily over the years, the number of deaths from road accidents would also steadily increase. We will test our assumption using a bar plot that shows the global count of the number of deaths per year.

Fig 1. shows the yearly count of deaths from road accidents from 1990 to 2019. We see that the number of deaths steadily increased from 1990 to around 2005. From 2005 to 2010, there was a plateau in the number of deaths recorded, without major variation. We further see that the number of deaths from road accidents has been decreasing since 2010.

Fig 1 - Yearly count of deaths

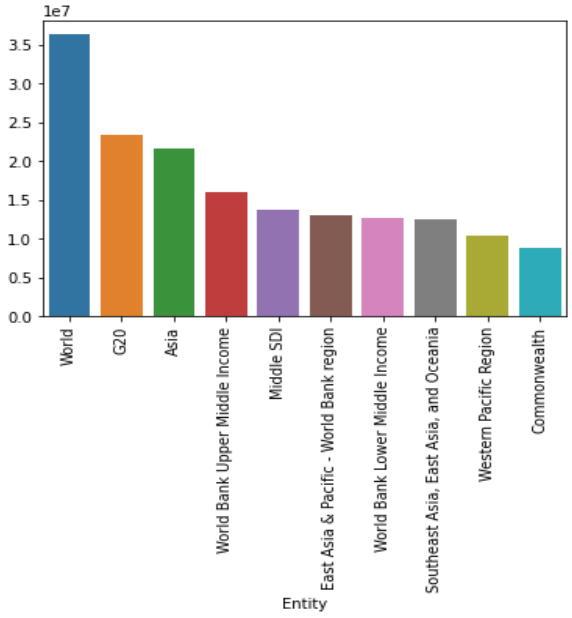


This is contradictory to our assumption that the number of deaths would be steadily rising given that the number of vehicles on the road has been rapidly rising.

This would lead us to the hypothesis that the number of deaths has been decreasing because of safer road infrastructure being built and more rigorous safety measures being undertaken. Therefore, we can assume that the more developed countries would have fewer road deaths compared to the underdeveloped countries. We would further assume that because India and China, in between them, have almost half the population of the world, we would see that most of the accidents are concentrated in these countries.

We will plot the category-wise plot to indicate the deaths in the last two decades in different categories for regions.

Fig 2 - Category-wise count of deaths

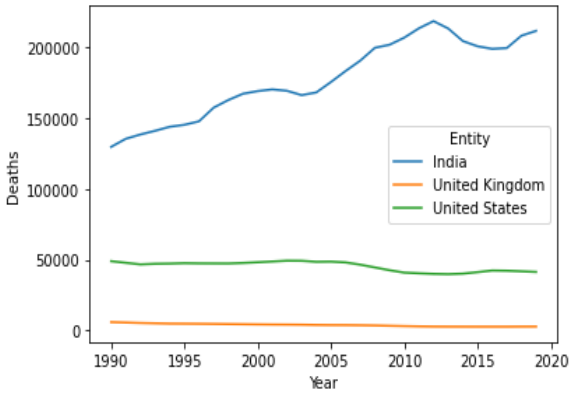


We see that G20 countries have a far higher death count compared to the countries in Asia, despite having a lower population. Also, countries categorized as 'middle-income' countries have a higher death rate compared to the more underdeveloped commonwealth countries.

This, once again, contradicts our assumption (that was hypothesized from looking at the global trend) that the more developed countries/regions would have a lower death rate in road accidents owing to having more strict regulations and better road infrastructure.

We now look at three countries in isolation - India, UK, and USA. While UK and USA are considered developed countries and have developed road infrastructure, India is a developing country where the road infrastructure, despite improving rapidly, is still not at par with the countries in the developed world. We would assume that the rate of death in India is not as controlled as in the case of the UK or USA.

Fig 3 - Comparative analysis



The comparative analysis shown in Fig. 3 confirms our assumption. We can see that countries like the UK and USA have a decreasing or stagnant death rate from road accidents. However, the rate is, in general, increasing in India.

This would imply our suspicion that the road infrastructure and safety measures in India have not been developing at the rate that would be expected in a more developed nation like the UK and the USA.

References -

1. https://www.kaggle.com/shivkumarganesh/road-traffic-deaths-1990-to-2019
2. https://ghdx.healthdata.org/gbd-results-tool