

Technical brief 1

Data Sources: [Global Burden of Disease](#) via [Our World in Data](#):

- Dataset 1: Deaths by Risk Factor. This dataset contains the estimated number of deaths for 28 risk factors from 1990 – 2019, the entity for which the measurement is based on (country or region), and iso country codes where applicable.
- Dataset 2: Disease burden by health expenditure. This dataset contains the estimated population per entity, health expenditure per capita, and Disability-Adjusted Life Years from 1990-2019, estimated population for 10,000 BC to 1990, the entity for which the measurement is based on (country or region), iso country codes where applicable, and continent.

Billionaire philanthropist Elon Musk wants to know where they should focus their charitable donations and asks for recommendations based on the above datasets.

Your report should contain statistics and visualisations that you interpret to provide the following key insights:

1. Collapsing across all years and countries/regions, what are the top three causes of death globally and how many deaths are estimated to be attributed to each of these factors?
2. Collapsing across all countries/regions, how have causes of death changed between 1990 and 2019? What are the key trends?
3. How many people are estimated to have died from smoking and alcohol related causes on each continent each year between 1990 and 2019?
4. On average, what was the health expenditure per capita for each Income Region for each year between 1995 and 2014? Note that there are multiple income categories in the dataset, you should use the ones labelled High income, Upper middle income, Lower middle income, and Low income.
5. One original insight of your own, based on your exploration of the data.

Based upon all the above, which risk factor/region/country/group should the philanthropist donate to and why?

Notes.

1. In reality, medical and technical expertise regarding possible interventions and their relative impact given cost would also inform the decision of where to donate. For the purposes of this assignment and to ensure a level playing field across all students and backgrounds, your decision should be based on the data alone.
2. The numbers provided in the dataset are estimated death numbers based on modelling conducted by the [Global Burden of Disease](#) study which is why they are not whole numbers.
3. The geographical information contained within both datasets is quite complex – take the time to review the data to ensure you know what is in each variable:
 - a. “Entity” contains geographical information. Importantly, there are several different units represented in this variable: individual countries, World Bank Income Regions, geographical regions, and political unions.
 - b. “Code” contains [ISO country codes](#). Note that only individual countries have an ISO code, the higher-level groupings of regions, unions, and income groups will not have any data in this column.

- c. [The World Bank](#) provides a list of the current Country and Lending Group income classifications.
- 4. The names of the columns are difficult to work with, but this is how the data comes so we've left it alone because it represents a very realistic example of messy data. You will want to tidy this up as quickly as possible to make your life easier.