**T5: DataSci Warriors - Lu, Bianca, Clare, and Josh**

Our research proposal is focused on the relationship/association between a country’s gross domestic product (GDP), population size, region, and air pollution - total death per million due to air pollution. We would like to examine 231 countries, their population size, and GDP over 27 years (1990 - 2017), leaving us with about 6,237 observations.

Main Question:

Are lower GDP countries more affected by the effects of air pollution?

* Is there a correlation between GDP per capita and deaths caused by pollution? Is it linear? How strong is the correlation?
* Is there a difference in means of death caused by pollution between low, mid, and high GDP per capita?

SMART:

1. Which countries have the highest and lowest air pollution? How is this related to the region?
2. Which years have the lowest and highest pollution?
3. How does air pollution increase over time? More specifically, are death rates in recent X amount of years higher than death rates from groups of X years before?
4. Is there a relationship between population size and pollution?

Data & Variables:

1. Variables:
   1. GDP, Population Size, Years, Country Code, Country Name, Death due to Air Pollution, Geographic Regions (Africa; Asia; Europe; Latin America/Caribbean; Northern America; Oceania)
2. Data:
   1. [Air pollution (death)](https://www.kaggle.com/akshat0giri/death-due-to-air-pollution-19902017)
   2. [GDP](https://www.kaggle.com/zackerym/gdp-annual-growth-for-each-country-1960-2020)
   3. [Region](https://population.un.org/wpp/DefinitionOfRegions/)

Supplemental:

* [Worldbank](https://data.worldbank.org/indicator/NY.GDP.MKTP.CD)
* [OECD](https://stats.oecd.org/)
* [UN](https://population.un.org/wpp/)

[Github](https://github.com/goebelclare/GitHub_Repo)