**Data Source (URL web address with hyperlink):**

<https://en.wikipedia.org/wiki/World_Heritage_Sites_by_country>

https://en.wikipedia.org/wiki/List\_of\_countries\_by\_GDP\_(PPP)\_per\_capita

https://en.wikipedia.org/wiki/List\_of\_countries\_and\_dependencies\_by\_area

**Context of Data and Variables:**

* Number of world heritage sites by country (Integer count) – broken down by type of site (Natural, Cultural, Mixed)
* GDP per capita (converted $US)
* Total area of country (km^2)
* “density” of heritage sites (number / area) broken down by type of site

**Output from str() function applied to the data object:**

> str(full.tib)

tibble[,11] [153 x 11] (S3: tbl\_df/tbl/data.frame)

$ country : chr [1:153] "Afghanistan" "Albania" "Algeria" "Angola" ...

$ natural\_sites : num [1:153] 0 1 0 0 0 5 0 12 1 0 ...

$ cultural\_sites : num [1:153] 2 2 6 1 1 6 2 4 9 3 ...

$ mixed\_sites : num [1:153] 0 1 1 0 0 0 1 4 0 0 ...

$ total\_sites : num [1:153] 2 4 7 1 1 11 3 20 10 3 ...

$ GDP\_per\_capita : num [1:153] 2156 14648 12020 6966 22880 ...

$ area : num [1:153] 652230 28748 2381741 1246700 442 ...

$ natural\_density : num [1:153] 0.00 3.48e-05 0.00 0.00 0.00 ...

$ cultural\_density: num [1:153] 3.07e-06 6.96e-05 2.52e-06 8.02e-07 2.26e-03 ...

$ mixed\_density : num [1:153] 0.00 3.48e-05 4.20e-07 0.00 0.00 ...

$ total\_density : num [1:153] 3.07e-06 1.39e-04 2.94e-06 8.02e-07 2.26e-03 ...

**Research Questions to be explore:**

**1. Is there a bias in allocation of cultural heritage sites to ‘richer’ countries?**

**2. Is there a relationship between area and GDP per capita?**

**3. Identify any other relationship between distribution of world heritage sites and GDP per capita/area**

**Statistical Analysis Plan**

**Population:**

* Countries with at least 1 UNESCO world heritage site

**Primary Objective:**

* Investigate the relationship between a country’s GDP per capita and the number of *cultural* world heritage sites in the country.

**Secondary Objectives:**

* Compare against the relationship between a country’s GDP per capita and number of *natural* world heritage sites in the country.
* Investigate the relationship between a country’s area and the number of world heritage sites.

**Data Collection methods:**

* Data scraping from Wikipedia (accessed 26/05/2021). Links to the pages above.
* GDP data based on 2019 information.

**Variables under consideration:**

* Number (integer count) of cultural world heritage sites in a country – **Primary outcome variable**
* Number (integer count) of natural/mixed world heritage sites in a country – **Secondary outcome variable**
* GDP per capita (converted US$) – gross domestic product per capita as of 2019 (World Bank) – **Primary explanatory variable**
* Area (km^2) – total area including bodies of water of a country – **Secondary explanatory variable**
* Density of world heritage sites (Number/Area) – **Secondary explanatory variable**

**Missing data procedures:**

* Any countries missing Area or GDP per capita are excluded from the analysis.
* Data has been scraped from several different table and then joined on country name so any that have been represented under different names have been excluded

**Numerical and graphical summaries to be presented:**

* Boxplot, mean, sd of number of heritage sites
* Boxplot, mean, sd of GDP per capita,
* Boxplot, mean, sd of Area (identify outliers such as Marshall Islands)
* Scatterplot of number of cultural heritage sites (y-axis) against GDP per capita (x-axis)
* Scatterplot of number of natural heritage sites (y-axis) against GDP per capita

**Models to be fitted:**

* Main model will be cultural heritage sites as response with GDP per capita as explanatory
* Simple linear regression model tried first
* Residual plots to check for model assumption/outliers
* Can try log(cultural sites) against GDP per capita
* Secondary model will be natural heritage sites as response with GDP per capita as explanatory to compare significance with main model
* Same analysis can be carried out using cultural/natural heritage site density as response and GDP per capita as explanatory