DAT405 Introduction to Data Science and AI

Assignment 1: Introduction to Data Science and Python

Student name	Hours spent on the tasks
Lenia Malki	10
Maële Belmont	10

Comment on the code

If the plots are not displayed when you open the notebook and the pdf, please either

- run the cells again with the year value as defined in the comment in the cell
  or
- look at the html file uploaded on Canvas.

Sorry for the inconvenience.

Setup

Python modules need to be loaded to solve the tasks.

```
In [1]:
import numpy as np
import math
import matplotlib.pyplot as plt
import pandas as pd
import plotly.express as px
```

Task 1 - Download some data related to GDP per capita and life expectancy

A. Write a Python program that draws a scatter plot of GDP per capita vs life expectancy. State any assumptions and motivate decisions that you make when selecting data to be plotted, and in combining data. [1p]

When extracting the data, we chose to not include population size as a parameter. As we are primarily interested in exploring the relationship between life expectancy and GDP, we did not consider the population size data to be contributing to this relationship. Disincluding this data makes it easier to focus and study the other two data entities.

In terms of countries, we chose to remove data points with null values in the columns of interest ('Life expectancy' and 'GDP per capita'). Among the available data, we chose only to focus on the data obtained in 2018. The reason for this lies in the assumption that recent data, in this context, is more accurate as it is more available today than it might have been in the nineteen hundreds.

In order to visualize the data in a readable manner, we chose to group countries together by their respective continents and color code these. The continents for each entity were defined only in the year 2015, thus we found a way (describe in the code) to replace NaN values in the 'Continent' column. To minimize cluttering, we used Plotly to create an interactive plot, which displays more information (country and exact value of life expectancy and GDP per capita) when the cursor is on a dot. GDP per capita (x-axis) was plotted on a log scale to avoid cluttering on the left side of the plot.

It is possible to study any data between the year 1543 and 2018. In other words, the program is not limited to only data from one specific year though only data for one year is plotted at a time.

In [2]:

```python
def GDP(csv_file, column1, column2, title, GDP):
    """
    input:
    - csv_file (string): name of csv file
    - column1 (string): name of column for data GDP per capita
    - column2 (string): name of column for the other data entity (ex: 'Life expectancy')
    - title (string): Title of the plot
    - GDP (string): 'GDP only' -> returns dataframe with GDP / 'GDP per Capita' -> returns
    output:
    - finalData (dataframe): dataframe containing data for the chosen year without NaN val
    """
    #Read the csv file containing the downloaded data
    rawData = pd.read_csv(csv_file)

    #Remove rows containing NaN value in the 'GDP per capita' or 'Life expectancy'
    noNaN = rawData.dropna(subset=[column1, column2])

    #Print list with years for which data is available
    yearList = [] #create empty list
    for i in range(len(noNaN)):
        if noNaN['Year'].values[i] not in yearList: #if the year value is not in the list
            yearList.append(noNaN['Year'].values[i]) #...it is added at the end of the lis
    yearlist = yearList.sort() #the list is sorted it in ascending order to make it easier
    print('---------------------------------------------------------------------------
    print('List of years for which data is available: \n')
    print(yearList)
    print('---------------------------------------------------------------------------

    #Create a new table with the data for year of interest chosen by the user
    year = int(input('Year of interest for the data "%s": \n' % (title)))
    print('---------------------------------------------------------------------------
    yearData = rawData.loc[rawData['Year']==year].reset_index()

    #Replace NaN values in 'Continent' by the continent value for every country
    year2015 = rawData.loc[rawData['Year']==2015].reset_index() #continents are defined or
    for i in range(len(yearData)):
        for j in range(len(year2015)):
            if yearData['Entity'][i] == year2015['Entity'][j]: #compare the 'Entity' of th
                yearData.at[i, 'Continent'] = year2015['Continent'][j] #...assign the 'Con

    # GDP per capita on x-axis
    if GDP == 'GDP per capita':
        #Create a new table with only the relevant columns
        relevantColumns = yearData.loc[:, ['Entity','Year', column1, column2, 'Continent']

    # GDP on x-axis
    elif GDP == 'GDP only':
        #Create a new column with the GDP by mulplying 'GDP per capita' by the 'Population
        yearData['GDP'] = yearData[column1]*yearData['Population (historical estimates)']

        #Create a new table with only the relevant columns
        relevantColumns = yearData.loc[:, ['Entity','Year', 'GDP', column2, 'Continent']]
```

```python
        #Create a new table without rows containing NaN values
        finalData = relevantColumns.dropna()

        return finalData
```

In [3]:
```python
def plotFigure(finalData, xlabel, ylabel, title, min_x, max_x, min_y, max_y, nbFigure):
    """
    input:
    - finalData: dataframe will be plotted
    - xlabel: label displayed under x-axis
    - ylabel: label displayed under y-axis
    - title of the plot
    - min_x, max_x, min_y, max_y (float): numbers used for adjusting the axis limit on the
    - nbFigure (int): number of the figure
    output:
    - figure with the plot
    """
    ###Plot
    try:
        #Print the number of the figure
        print('Figure', int(nbFigure),':')

        #Assign year value to a variable to use it in the title of the plot
        year = finalData['Year'].values[0]

        column1=finalData.columns[2] # GDP column
        column2=finalData.columns[3] # other data entity (ex: 'Life expectancy') column

        # Calculate axis limits that make the plot look good
        minimum_x = finalData[column1].min()-min_x
        maximum_x = finalData[column1].max()+max_x
        minimum_y = finalData[column2].min()-min_y
        maximum_y = finalData[column2].max()+max_y

        #Create Plotly figure

        fig = px.scatter(finalData, x=column1,
                    y=column2,
                    color='Continent', hover_data=['Entity'],
                    log_x=True, title=f'{title}, {year}',
                    labels={column1: xlabel, column2: ylabel},
                    trendline='ols', trendline_options=dict(log_x=True),
                        trendline_scope='overall', trendline_color_override='black',
                    range_x=[minimum_x, maximum_x], range_y=[minimum_y, maximum_y])

        #Display figure
        fig.show()

    #Print error if the year input is not available
    except:
        print('No data available for this year, please run the cell again and try with and
```

In [4]:
```python
#Plot life expectancy vs GDP per capita in 2018
plotFigure(GDP('life-expectancy-vs-gdp-per-capita.csv',
        'GDP per capita',
        'Life expectancy',
        'Life expectancy vs. GDP per capita', 'GDP per capita'), 'GDP per capita ($)',
         'Life expectancy (years)', 'Life expectancy vs. GDP per capita', 100, 10000, 1,
```

---------------------------
List of years for which data is available:

[1543, 1548, 1553, 1558, 1563, 1568, 1573, 1578, 1583, 1588, 1593, 1598, 1603, 1608, 1613,
1618, 1623, 1628, 1633, 1638, 1643, 1648, 1653, 1658, 1663, 1668, 1673, 1678, 1683, 1688,
1693, 1698, 1703, 1708, 1713, 1718, 1723, 1728, 1733, 1738, 1743, 1748, 1751, 1752, 1753,
1754, 1755, 1756, 1757, 1758, 1759, 1760, 1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768,
1769, 1770, 1771, 1772, 1773, 1774, 1775, 1776, 1777, 1778, 1779, 1780, 1781, 1782, 1783,
1784, 1785, 1786, 1787, 1788, 1789, 1790, 1791, 1792, 1793, 1794, 1795, 1796, 1797, 1798,
1799, 1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813,
1814, 1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828,
1829, 1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843,
1844, 1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858,
1859, 1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873,
1874, 1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888,
1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903,
1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918,
1919, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933,
1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948,
1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963,
1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978,
1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993,
1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008,
2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018]

Year of interest for the data "Life expectancy vs. GDP per capita":
2018

Figure 1 :

Life expectancy vs. GDP per capita, 2018
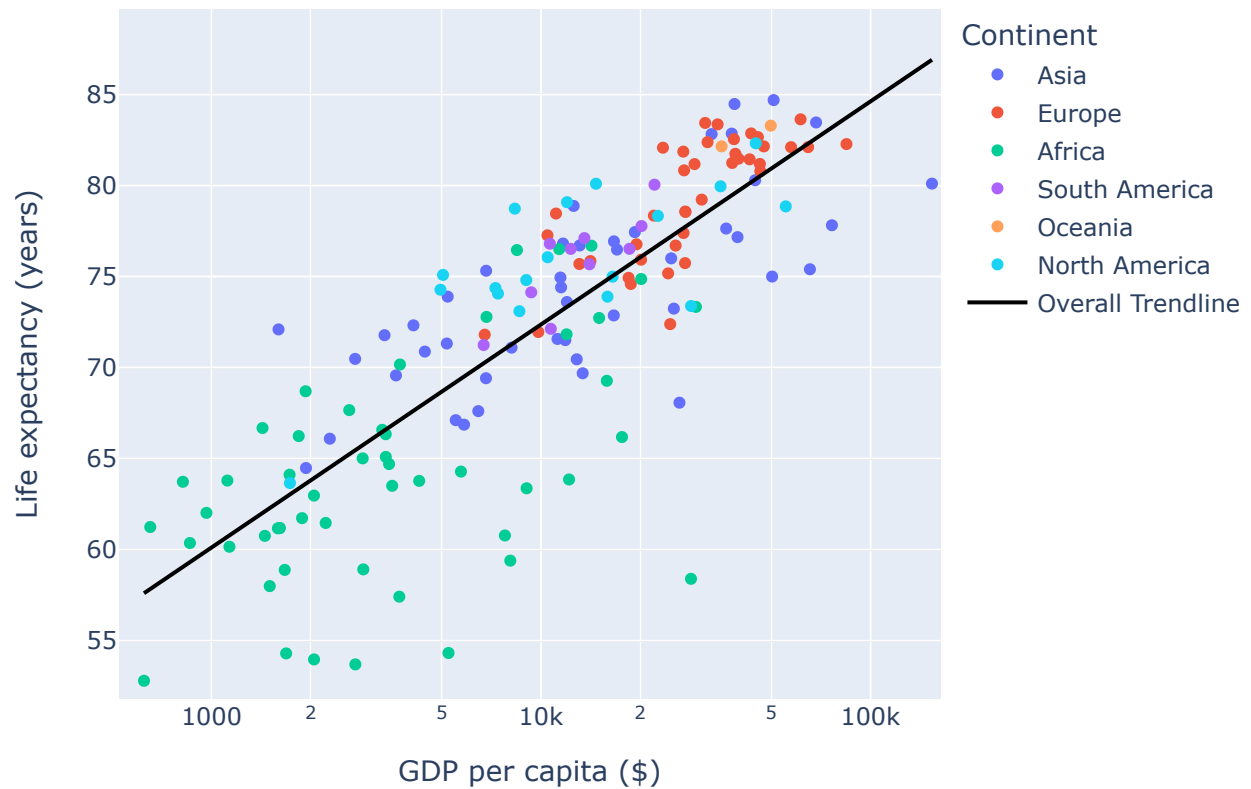
B. Consider whether the results obtained seem reasonable and discuss what might be the explanation for the results you obtained. [1p]

The results show a clear trend. Countries with a higher GDP tend to also rank higher in life expectancy. However, this does not mean that it is always true. We can for example see that Saudi Arabia has one of the higher GDP per capita at approximately $50,000 and a life expectancy of 75 years. On the same y-coordinate, we find Honduras which is one of Latin America's poorest countries with a GDP of \$5,000. Even though it seems to be a positive trend related to GDP per capita and life expectancy, deviations from the regression can reveal otherwise. Generally speaking, countries with higher GDP per capita might have the ability to provide better health care to their population as well as better living standards for the people, thus resulting in a higher life expectancy. One must however consider that other factors are at play when looking at life expectancy, not only GDP per capita.

C. Did you do any data cleaning (e.g., by removing entries that you think are not useful) for the task of drawing scatter plot(s) and the task of answering the questions d, e, f, and g? If so, explain what kind of entries that you chose to remove and why. If not, explain why you did not need to. [0.5p]

As mentioned in question 1.a, we decided to remove the population size data in order to only focus on the relationship between GDP and life expectancy. Data points with null values, such as for example missing GDP data or life expectancy score, were also removed in order to avoid outliers. Lastly, we decided to only collect and visualize the data from 2018. The reasoning behind this has to do with the assumption that the quality and availability of recent data, within this context, is better than that of much earlier years. That being said, we do not believe there to be big differences between closeby years.

D. Which countries have a life expectancy higher than one standard deviation above the mean? [0.5p]

With a standard deviation of approximately 7.747 and a mean of approximately 2.66, one standard deviation above the mean would require the life expectancy to be at a minimum of 80.41. The countries with a life expectancy of one standard deviation above the mean would be those presented in Figure 2. The data for these life expectancies are limited to that of a specific year supplied by the user. In this case, the input year was 2018.

In [5]:

```python
def aboveSD(finalData, ylabel, nbFigure):
    """
    inputs:
    - finalData (dataframe): dataframe containing the data of interest
    - ylabel (string): data topic
    - - nbFigure (int): number of the figure
    output:
    - displays dataframe with the data above the standard mean
    """

    column1=finalData.columns[2] # GDP column
    column2=finalData.columns[3] # other data entity (ex: 'Life expectancy') column

    # Useful statistical values
    mean = finalData[column2].mean() # mean

    #if the mean is NaN it means the year input is not available, so an error message is p
    if math.isnan(mean)==True:
        print('No data available for this year, please run the cell again and try with and

    else:
        sd = finalData[column2].std() # standard deviation
        oneSDAboveMean = mean + sd # one standard deviation above the mean

        # Create filter where values are 'True' when values in finalData have a life expec
        Filter = finalData[column2] >= oneSDAboveMean
        # Create table with relevant columns
```

```
            tableSDAboveMean = finalData[['Entity','Year', column2]]

            # Display figure number
            print('Figure', int(nbFigure), ':','Dataframe with countries having', ylabel ,'hig

            # Display filtered dataframe (life expectancy of one standard deviation above the
            display(tableSDAboveMean[Filter])
```

In [6]:
```
#Display dataframe with countries having a higher life expectancy than the mean in 2018
aboveSD(GDP('life-expectancy-vs-gdp-per-capita.csv',
        'GDP per capita',
        'Life expectancy',
        'Life expectancy vs. GDP per capita', 'GDP per capita'), 'Life expectancy vs. GI
```

```
--------------------------------------------------------------------------------
---------------------------
List of years for which data is available:

[1543, 1548, 1553, 1558, 1563, 1568, 1573, 1578, 1583, 1588, 1593, 1598, 1603, 1608, 1613,
1618, 1623, 1628, 1633, 1638, 1643, 1648, 1653, 1658, 1663, 1668, 1673, 1678, 1683, 1688,
1693, 1698, 1703, 1708, 1713, 1718, 1723, 1728, 1733, 1738, 1743, 1748, 1751, 1752, 1753,
1754, 1755, 1756, 1757, 1758, 1759, 1760, 1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768,
1769, 1770, 1771, 1772, 1773, 1774, 1775, 1776, 1777, 1778, 1779, 1780, 1781, 1782, 1783,
1784, 1785, 1786, 1787, 1788, 1789, 1790, 1791, 1792, 1793, 1794, 1795, 1796, 1797, 1798,
1799, 1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813,
1814, 1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828,
1829, 1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843,
1844, 1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858,
1859, 1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873,
1874, 1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888,
1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903,
1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918,
1919, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933,
1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948,
1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963,
1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978,
1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993,
1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008,
2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018]
--------------------------------------------------------------------------------
---------------------------
Year of interest for the data "Life expectancy vs. GDP per capita":
2018
--------------------------------------------------------------------------------
---------------------------
Figure 2 : Dataframe with countries having Life expectancy vs. GDP per capita higher than
the standard mean.
```

	Entity	Year	Life expectancy
14	Australia	2018	83.281
15	Austria	2018	81.434
22	Belgium	2018	81.468
39	Canada	2018	82.315
56	Cyprus	2018	80.828
60	Denmark	2018	80.784
78	Finland	2018	81.736
81	France	2018	82.541

	Entity	Year	Life expectancy
87	Germany	2018	81.180
90	Greece	2018	82.072
101	Hong Kong	2018	84.687
103	Iceland	2018	82.855
108	Ireland	2018	82.103
110	Israel	2018	82.819
111	Italy	2018	83.352
113	Japan	2018	84.470
130	Luxembourg	2018	82.102
137	Malta	2018	82.376
157	Netherlands	2018	82.143
159	New Zealand	2018	82.145
169	Norway	2018	82.271
181	Portugal	2018	81.857
203	Singapore	2018	83.458
206	Slovenia	2018	81.172
211	South Korea	2018	82.846
214	Spain	2018	83.433
219	Sweden	2018	82.654
220	Switzerland	2018	83.630
239	United Kingdom	2018	81.236

E. Which countries have high life expectancy but have low GDP (per capita)? [0.5p]

It essentially depends on how you define high life expectancy and low GDP. If we were to define low GDP as one standard deviation below the mean, it would yield a negative GDP score. This is because the standard deviation is greater than the mean, indicating a high variance between data points. Another way of defining low GDP would be to extract those data points whose GDP is below the median.

The GDP median is 12165.79 with a life expectancy median of 74.368. We can see that these countries, listed in Figure 3, are located closer to the upper left corner of the graph.

In [7]:

```python
def medianFilter(finalData, ylabel, nbFigure, aboveORbelow):
    """
    input:
    - finalData (dataframe): dataframe containing the data of interest
    - ylabel (string): name of the data other than GDP per capita (ex: Life expectancy)
    - nbFigure (int): number of the figure in the notebook
    - aboveORbelow (string):
        - 'above': displays median values + dataframe of countries with high life expectan
        - 'below': displays median values + dataframe of countries with low life expectand
        - 'none': returns only median values
    output:
    - prints median values
    - displays countries with low GDP and high "Y-axis" data (ex: Life expectancy)
```

```python
    """
    column1=finalData.columns[2] # GDP column
    column2=finalData.columns[3] # other data entity (ex: 'Life expectancy') column

    GDPmedian = finalData[column1].median() # GDP median

    #if the median is NaN it means the year input is not available, so an error message i
    if math.isnan(GDPmedian)==True:
        print('No data available for this year, please run the cell again and try with and

    else:
        Ymedian = finalData[column2].median() # other data entity median

        # Print median values
        print('GPD per capita median: %f \n' %(GDPmedian))
        print(ylabel, 'median: %f \n' %(Ymedian))
        print('----------------------------------------------------------------

        # Print dataframe with desired criterium
        if aboveORbelow == 'above' or aboveORbelow == 'below':
            # Criterium 1: low GDP and high other entity (ex: high life expectancy)
            if aboveORbelow == 'above':
                print('Figure', nbFigure,': Dataframe with countries having', ylabel ,'hig
                Filter1 = finalData[column1] < GDPmedian
                Filter2 = finalData[column2] > Ymedian
            # Criterium 2: high GDP and low other entity (ex: low life expectancy)
            elif aboveORbelow == 'below':
                print('Figure', nbFigure,': Dataframe with countries having', ylabel ,'low
                Filter1 = finalData[column1] > GDPmedian
                Filter2 = finalData[column2] < Ymedian

            # Create filter that combines the two filters
            combinedFilter = Filter1 & Filter2
            # Create dataframe with relevant columns only
            filteredColumns = finalData[['Entity','Year', column1, column2]]
            # Apply combinedFilter on the filteredColumns dataframe and display the result
            display(filteredColumns[combinedFilter])

        # Print 'nothing' when input is none
        elif aboveORbelow == 'none':
            print('\n')
        # error message if argument n°4 doesn't have the right input
        else:
            print('Argument n°4 has to be "above" or "below" or "none".\n')
```

```python
In [8]:  # Print countries with high life expectancy but low GDP in 2018
         medianFilter(GDP('life-expectancy-vs-gdp-per-capita.csv',
                 'GDP per capita',
                 'Life expectancy', 'Countries with high life expectancy and low GDP', 'GDP per c
```

```
--------------------------------------------------------------------------------
--------------------------
List of years for which data is available:

[1543, 1548, 1553, 1558, 1563, 1568, 1573, 1578, 1583, 1588, 1593, 1598, 1603, 1608, 1613,
1618, 1623, 1628, 1633, 1638, 1643, 1648, 1653, 1658, 1663, 1668, 1673, 1678, 1683, 1688,
1693, 1698, 1703, 1708, 1713, 1718, 1723, 1728, 1733, 1738, 1743, 1748, 1751, 1752, 1753,
1754, 1755, 1756, 1757, 1758, 1759, 1760, 1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768,
1769, 1770, 1771, 1772, 1773, 1774, 1775, 1776, 1777, 1778, 1779, 1780, 1781, 1782, 1783,
1784, 1785, 1786, 1787, 1788, 1789, 1790, 1791, 1792, 1793, 1794, 1795, 1796, 1797, 1798,
1799, 1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813,
1814, 1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828,
1829, 1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843,
1844, 1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858,
```

```
1859, 1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873,
1874, 1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888,
1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903,
1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918,
1919, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933,
1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948,
1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963,
1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978,
1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993,
1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008,
2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018]
--------------------------------------------------------------------------------
---------------------------
Year of interest for the data "Countries with high life expectancy and low GDP":
2018
--------------------------------------------------------------------------------
---------------------------
GPD per capita median: 12080.490800

Life expectancy median: 74.386500


--------------------------------------------------------------------------------
---------------------------
Figure 3 : Dataframe with countries having Life expectancy higher than the median and a GD
P per capita lower than the median.
```

	Entity	Year	GDP per capita	Life expectancy
2	Albania	2018	11104.1665	78.458
11	Armenia	2018	11454.4251	74.945
20	Barbados	2018	11995.1868	79.081
29	Bosnia and Herzegovina	2018	10460.5201	77.262
54	Cuba	2018	8325.6313	78.726
62	Dominica	2018	9021.1737	74.806
66	Ecuador	2018	10638.8251	76.800
100	Honduras	2018	5041.6354	75.088
114	Jordan	2018	11506.3383	74.405
151	Morocco	2018	8451.1355	76.453
191	Saint Lucia	2018	10475.3689	76.057
215	Sri Lanka	2018	11662.9064	76.812
231	Tunisia	2018	11353.8865	76.505
247	Vietnam	2018	6814.1423	75.317

F. Does every strong economy (normally indicated by GDP) have high life expectancy? [1p]

For the year 2018, we can see a steep trendline, indicating a strong relationship however, the variance between the data points is quite high. This can be seen by the spread out data. It is not always the case that countries with higher GDP also score better on life expectancy. For example, India with a GDP of around 9.2T have a life expectancy of 69.41 whilst Sao Tome and Principie has a GDP of 787.192M for approximetaly the same life expectancy. In conclusion, GDP per capita is a better indicator for the relationship of life expectancy and GDP per capita.

```
In [9]:    #Plot life expectancy vs. GDP in 2018
```

```
plotFigure(GDP('life-expectancy-vs-gdp-per-capita.csv',
          'GDP per capita',
          'Life expectancy',
          'Life expectancy vs. GDP', 'GDP only'), 'GDP ($)', 'Life expectancy',
      'Life expectancy vs. GDP', pow(10,8), pow(10,12), 5, 5, 4)
```

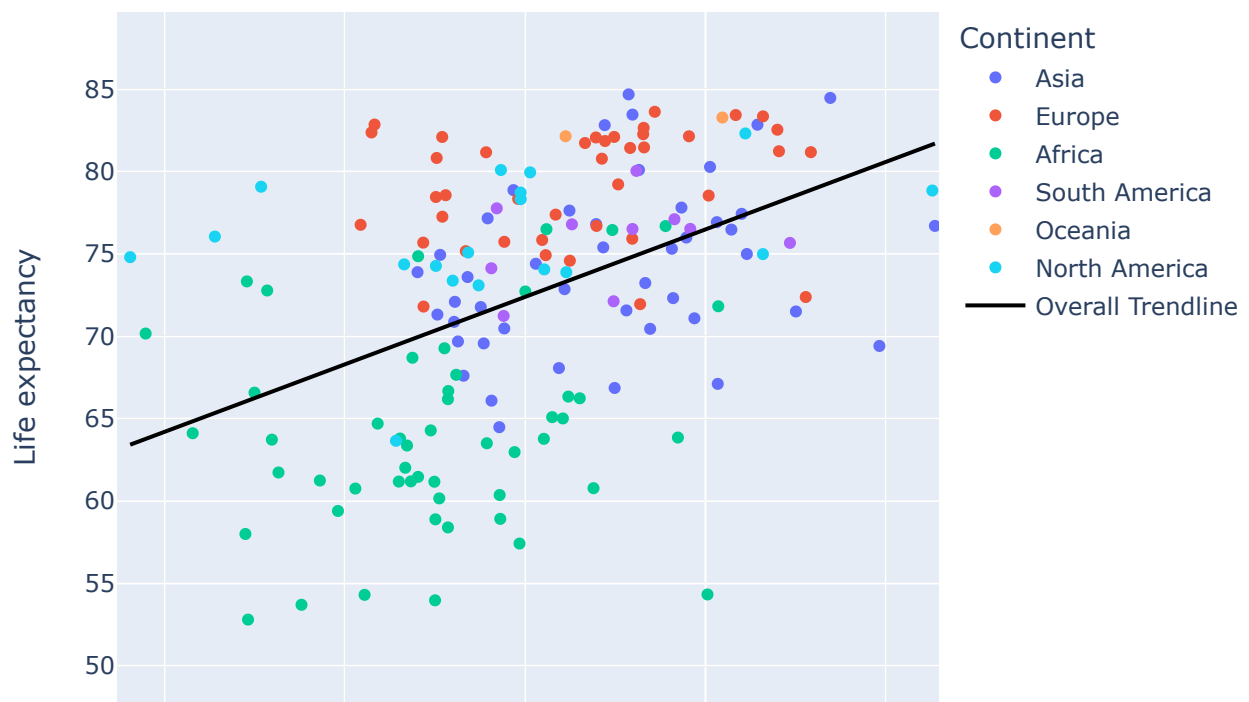List of years for which data is available:

[1543, 1548, 1553, 1558, 1563, 1568, 1573, 1578, 1583, 1588, 1593, 1598, 1603, 1608, 1613,
1618, 1623, 1628, 1633, 1638, 1643, 1648, 1653, 1658, 1663, 1668, 1673, 1678, 1683, 1688,
1693, 1698, 1703, 1708, 1713, 1718, 1723, 1728, 1733, 1738, 1743, 1748, 1751, 1752, 1753,
1754, 1755, 1756, 1757, 1758, 1759, 1760, 1761, 1762, 1763, 1764, 1765, 1766, 1767, 1768,
1769, 1770, 1771, 1772, 1773, 1774, 1775, 1776, 1777, 1778, 1779, 1780, 1781, 1782, 1783,
1784, 1785, 1786, 1787, 1788, 1789, 1790, 1791, 1792, 1793, 1794, 1795, 1796, 1797, 1798,
1799, 1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813,
1814, 1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828,
1829, 1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843,
1844, 1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858,
1859, 1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873,
1874, 1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888,
1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903,
1904, 1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918,
1919, 1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933,
1934, 1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948,
1949, 1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963,
1964, 1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978,
1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993,
1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008,
2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018]

Year of interest for the data "Life expectancy vs. GDP":
2018

Figure 4 :
```

## Life expectancy vs. GDP, 2018

|  | 1B | 10B | 100B | 1T | 10T |
|---|---|---|---|---|---|

GDP ($)

## G. Related to question f, what would happen if you use GDP per capita as an indicator of strong economy? Explain the results you obtained, and discuss any insights you get from comparing the results of g and f. [1p]

The correlation of life expectancy and GDP per capita is stronger than that of life expectancy and GDP only. This is clear when comparing the same year with each other. The variance is visabily greater in the graph from Figure 4 than that of Figure 1. There are however some similarities found in both graphs. For example, the majority of countries in Africa tend to show up below the regression line and more left of the graph. The regression line in 1.a is also steaper, indicating a stronger relationship. Using GDP per capita can thus provide a clearer indication of the country's prosperity.

# Task 2 - Download some other data sets, e.g. related to happiness and life satisfaction, trust, corruption, etc.

## A. Think of several meaningful questions that can be answered with these data, make several informative visualisations to answer those questions. State any assumptions and motivate decisions that you make when selecting data to be plotted, and in combining data. [2.5p]

## B. Discuss any observations that you make, or insights obtained, from the data visualisations. [2p]

5 datasets were downloaded for this task. We asked then answered one question per datasets and wrote a conclusion comment at the end.

In the following figures, we plotted data for the most recent year available to make analyses that are more likely to represent the current situation.

We used the same functions as in task 1 since the data downloaded has the same format as 'Life expectancy vs. GDP per capita'.

The visualization decisions are the same as in question 1.a.

### 1. Life satisfaction vs. GDP per capita

***Which countries have low life satisfaction but have high GDP per capita?***
To determine the countries with a low life satisfaction and high GDP per capita, the data was filtered to display countries with a life satisfaction below the median and a GDP per capita above the median. The countries respecting the criterium are located in the bottom right corner of the plot (Figure 5). The GPD per capita median is $18,278, while the life satisfaction median is 5.81. The results are displayed in Figure 6.

In [10]:
```
#Plot life satisfaction vs. GDP per capita in 2020
plotFigure(GDP('gdp-vs-happiness.csv',
        'GDP per capita, PPP (constant 2017 international $)',
        'Life satisfaction in Cantril Ladder (World Happiness Report 2021)',
        'Happiness vs. GDP per capita', 'GDP per capita'), 'GDP per capita ($)',
        'Life satisfaction (scale: 0-10)', 'Life satisfaction vs. GDP per capita', 30,
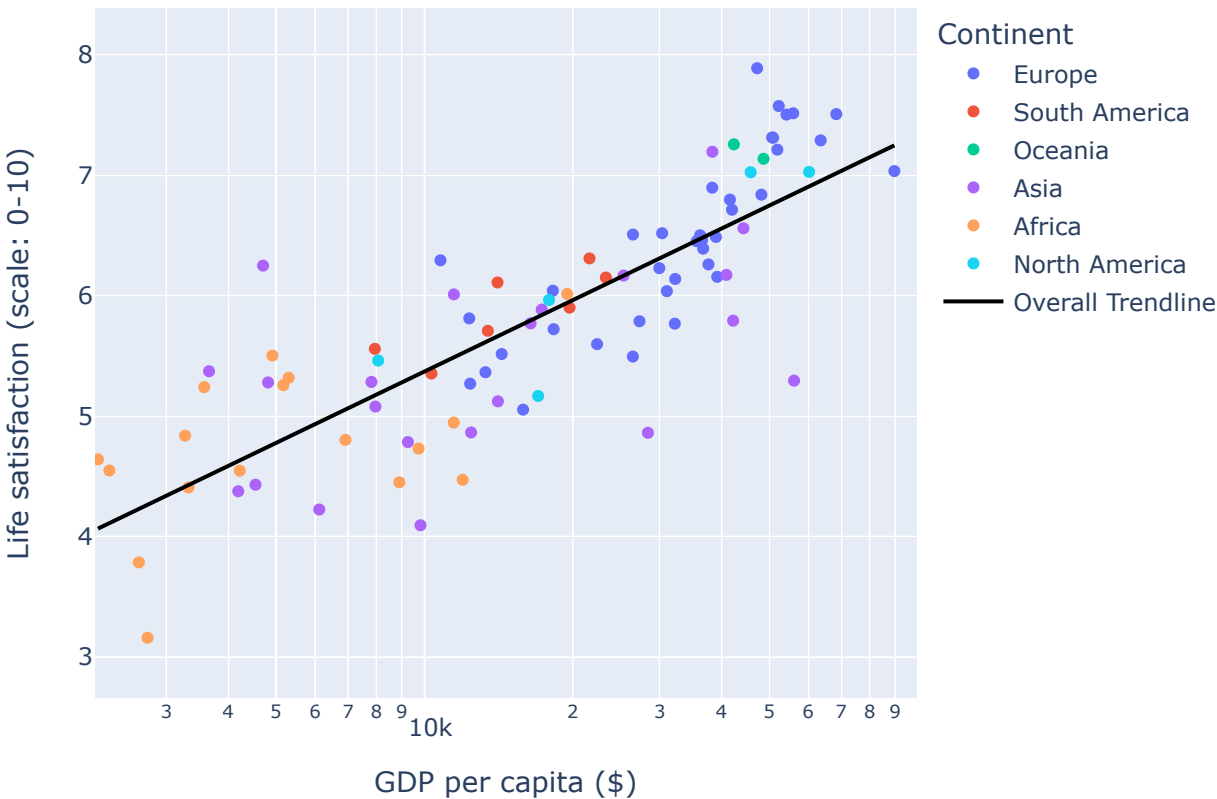```

```
------------------------------------
List of years for which data is available:

[2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019,
2020]
------------------------------------------------------------------------------------------
------------------------------------
Year of interest for the data "Happiness vs. GDP per capita":
2020
------------------------------------------------------------------------------------------
------------------------------------
Figure 5 :
```

Life satisfaction vs. GDP per capita, 2020


```
In [11]:    # Print countries with low life satisfactionbut high GDP per capita in 2020
            medianFilter(GDP('gdp-vs-happiness.csv',
                     'GDP per capita, PPP (constant 2017 international $)',
                     'Life satisfaction in Cantril Ladder (World Happiness Report 2021)',
                     'Happiness vs. GDP per capita', 'GDP per capita'), 'life satisfaction', 6, 'belo
```

```
------------------------------------------------------------------------------------------
------------------------------------
List of years for which data is available:

[2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019,
2020]
------------------------------------------------------------------------------------------
------------------------------------
Year of interest for the data "Happiness vs. GDP per capita":
2020
------------------------------------------------------------------------------------------
------------------------------------
GPD per capita median: 18278.730792
```

```
life satisfaction median: 5.812000
```


Figure 6 : Dataframe with countries having life satisfaction lower than the median and a G
DP per capita higher than the median.

	Entity	Year	GDP per capita, PPP (constant 2017 international $)	Life satisfaction in Cantril Ladder (World Happiness Report 2021)
36	Bulgaria	2020	22383.805544	5.598
98	Greece	2020	27287.083401	5.788
111	Hong Kong	2020	56153.971499	5.295
207	Portugal	2020	32181.154537	5.768
214	Russia	2020	26456.387938	5.495
242	South Korea	2020	42251.445057	5.793
264	Turkey	2020	28384.987785	4.862

2. Children per women vs. GDP per capita

Is the number of children per women positively correlated to the GDP per capita ?
The trendline has a negative slope, which implies that the number of children per women is negatively correlated to the GDP per capita. The highest numbers of children per woman are observed in the upper left corner on Figure 8, where the GDP per capita is lower than the median ($11,815). The predominant color of the dots is red in Figure 7, indicating that the countries with the most children per woman are located in Africa.

In [12]:
```python
#Plot children per women vs. GDP per capita in 2017
plotFigure(GDP('children-per-woman-by-gdp-per-capita.csv',
        'Output-side real GDP per capita (gdppc_o) (PWT 9.1 (2019))',
        'Estimates, 1950 - 2020: Annually interpolated demographic indicators - Total fe
        'Children per woman vs. GDP per capita', 'GDP per capita'), 'GDP per capita ($)'
        'Children per woman', 'Children per woman vs. GDP per capita', 20, pow(10,4), (
```


List of years for which data is available:

```
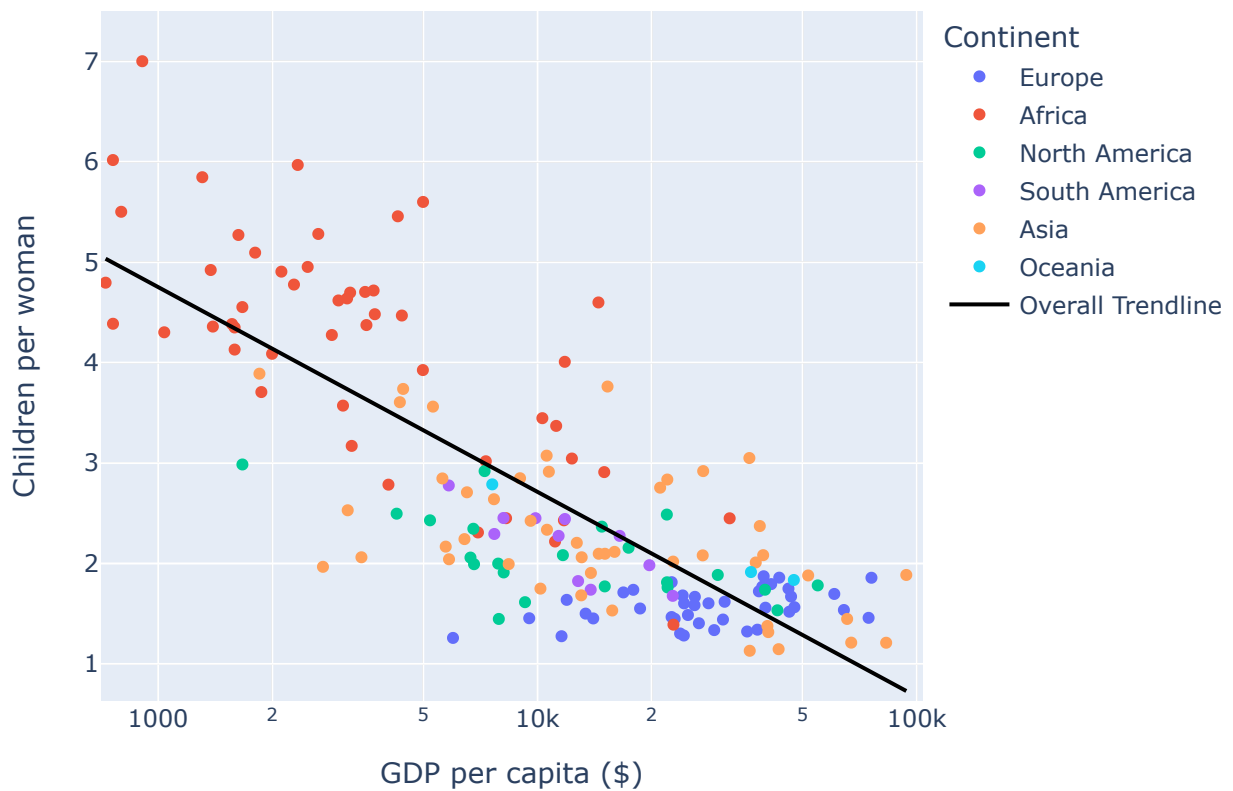[1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964,
1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979,
1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994,
1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009,
2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017]
```


Year of interest for the data "Children per woman vs. GDP per capita":
2017

Figure 7 :

Children per woman vs. GDP per capita, 2017

```
# Print median values in 2017
medianFilter(GDP('children-per-woman-by-gdp-per-capita.csv',
        'Output-side real GDP per capita (gdppc_o) (PWT 9.1 (2019))',
        'Estimates, 1950 - 2020: Annually interpolated demographic indicators - Total fe
        'Children per woman vs. GDP per capita', 'GDP per capita'), 'children per woman'
```

```
-----------------------------------------------------------------------------------
---------------------------
List of years for which data is available:

[1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964,
1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979,
1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994,
1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009,
2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017]
-----------------------------------------------------------------------------------
---------------------------
Year of interest for the data "Children per woman vs. GDP per capita":
2017
-----------------------------------------------------------------------------------
---------------------------
GPD per capita median: 11815.036000

children per woman median: 2.157000

-----------------------------------------------------------------------------------
---------------------------
```

3. Share of adults who smoke vs. GDP per capita

Is there a trend between the share of adults who smoke vs. GDP per capita ?

The data is scattered, the variance is significant, compared to datasets previously analysed. The share of adults who smoke appears to be the lowest in African countries, where the GDP per capita is the lowest (Figure 9). Countries with a high GDP per capita have a slightly higher share of smokers than countries in Africa. The countries with the highest share of adults who smoke have a GDP per capita lower than the median ($14,253). Apart from these observations, no specific conclusions can be drawn because R^2 is close to zero, which indicates that the data is in principle not suitable for regression. This also implies the correlation between the share of adults who smoke and the GDP per capita is low.

In [14]:
```
#Plot share of adults who smoke vs. GDP per capita in 2018
plotFigure(GDP('share-of-adults-who-are-smoking-by-level-of-prosperity.csv',
         'GDP per capita, PPP (constant 2017 international $)',
         'Prevalence of current tobacco use (% of adults)',
         'Share of adults who are smoking vs. GDP per capita', 'GDP per capita'), 'GDP pe
          'Adults who smoke (% of adults)', 'Share of adults who are smoking vs. GDP per
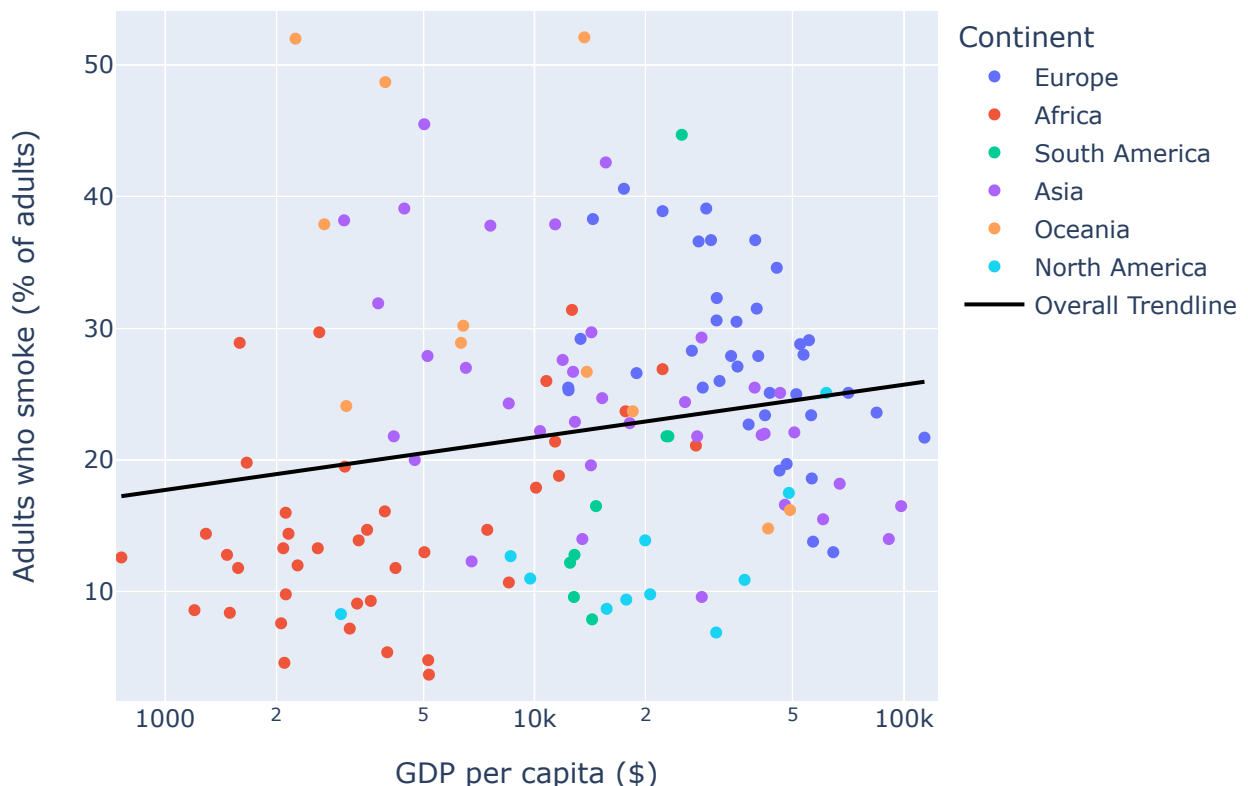          25, pow(10,4), 2, 2, 9)
```


List of years for which data is available:

[2007, 2010, 2012, 2014, 2016, 2018]

Year of interest for the data "Share of adults who are smoking vs. GDP per capita":
2018

Figure 9 :

Share of adults who are smoking vs. GDP per capita, 2018

```
#Print the GDP per capita and Share of adults who are smoking median in 2018
medianFilter(GDP('share-of-adults-who-are-smoking-by-level-of-prosperity.csv',
        'GDP per capita, PPP (constant 2017 international $)',
        'Prevalence of current tobacco use (% of adults)',
        'Share of adults who are smoking vs. GDP per capita', 'GDP per capita'),
            'Share of adults who are smoking', 10, 'none')
```

```
----------------------------------------------------------------------------
----------------------------
List of years for which data is available:

[2007, 2010, 2012, 2014, 2016, 2018]
----------------------------------------------------------------------------
----------------------------
Year of interest for the data "Share of adults who are smoking vs. GDP per capita":
2018
----------------------------------------------------------------------------
----------------------------
GPD per capita median: 14253.408986

Share of adults who are smoking median: 22.000000


----------------------------------------------------------------------------
----------------------------
```

4. Medical doctors per 1,000 people vs. GDP per capita

A interesting question to ask here is whether countries with higher GDP per capita have less or more medical doctors per 1.000 people. The graph for 2018 (Figure 11) shows several interesting facts. Up until 6.5K of GDP, the variance seems to be quite low and there seem to exist a relationship between medical doctors and GDP per capita. However, this trend is not really clear. From 10k and greater, the data points are much more spread. There are also some extremes such as Georgia and Lithuania. Overall, there seem to be a positive trend. Those countries with lower GPD and which are considered developing countries are once again located on the lower left corner of the graph. This does make intuitively sense as education is not as accessible in these countries.

```
# Plot Medical doctors per 1,000 people vs. GDP per capita in 2018
plotFigure(GDP('medical-doctors-per-1000-people-vs-gdp-per-capita.csv',
        'GDP per capita, PPP (constant 2017 international $)',
        'Physicians (per 1,000 people)',
        'Medical doctors per 1000 people vs. GDP per capita', 'GDP per capita'), 'GDP pe
        'Medical doctors (per 1,000 people)', 'Medical doctors per 1000 people vs. GDP
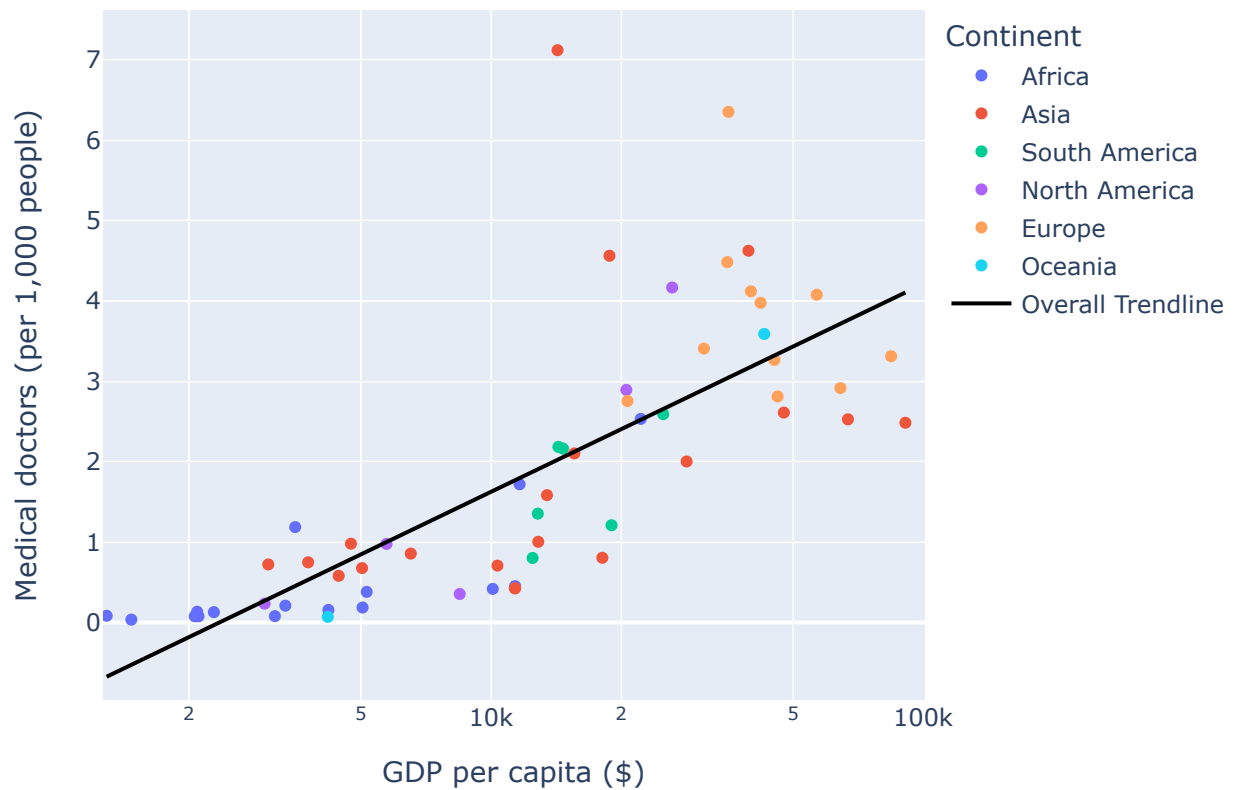        25, pow(10,4), 1, 0.5, 11)
```

```
----------------------------------------------------------------------------
----------------------------
List of years for which data is available:

[1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004,
2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018]
----------------------------------------------------------------------------
----------------------------
Year of interest for the data "Medical doctors per 1000 people vs. GDP per capita":
2018
----------------------------------------------------------------------------
----------------------------
Figure 11 :
```

Medical doctors per 1000 people vs. GDP per capita, 2018

```
# Print median values in 2018
medianFilter(GDP('medical-doctors-per-1000-people-vs-gdp-per-capita.csv',
        'GDP per capita, PPP (constant 2017 international $)',
        'Physicians (per 1,000 people)',
        'Medical doctors per 1000 people vs. GDP per capita', 'GDP per capita'),
            'Medical doctors (per 1,000 people)', 12, 'none')
```

```
-----------------------------------------------------------------------------
----------------------------
List of years for which data is available:

[1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004,
2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018]
-----------------------------------------------------------------------------
----------------------------
Year of interest for the data "Medical doctors per 1000 people vs. GDP per capita":
2018
-----------------------------------------------------------------------------
----------------------------
GPD per capita median: 12844.014426

Medical doctors (per 1,000 people) median: 1.198400


-----------------------------------------------------------------------------
----------------------------
```

5. Child mortality vs. GDP per capita

> Child mortality is defined as the number of children born alive that die before their 5th birthday.

How has child mortality evolved from 1986 to 2016 (30 years)?

First of all, we can see that the trendline starts at around 21% in 1986 (Figure 13) whilst it starts at around 9% in 2016 (Figure 14) which indicates that the overall percentage of child mortality has decreased. We can also find most of African countries on the upper left side of the graph. As we have seen earlier, Africa has many developing countries with lower GDPs (per capita). A much lower GDP per capita can be the cause of higher child mortality rates as the trendline in both graphs are quite evident. There seem to be a higher variance between the countries in 2016 as opposed to 1986, at leats for most of Africa and some of Asia. One can also argue that countries with a lower GDP and greater population size tend to have a greater child mortaility rate.

In [18]:

```
# Plot child mortality vs. GDP per capita in 1986
plotFigure(GDP('child-mortality-gdp-per-capita.csv',
        'GDP per capita',
        'Child mortality (Select Gapminder, v10) (2017)',
        'Child mortality vs. GDP per capita', 'GDP per capita'), 'GDP per capita ($)',
        'Child mortality rate (% children < 5 y.o.)', 'Child mortality vs. GDP per capi
        25, pow(10, 4), 3, 1, 13)
```


List of years for which data is available:

[1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813, 1814,
1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828, 1829,
1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843, 1844,
1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858, 1859,
1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873, 1874,
1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889,
1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903, 1904,
1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918, 1919,
1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933, 1934,
1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949,
1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964,
1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979,
1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994,
1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009,
2010, 2011, 2012, 2013, 2014, 2015, 2016]

Year of interest for the data "Child mortality vs. GDP per capita":
2016

Figure 13 :

Child mortality vs. GDP per capita, 2016

Child mortality rate (%)

4

2

0

-2

1000   2   5   10k   2   5   100k

GDP per capita ($)

In [19]:
```
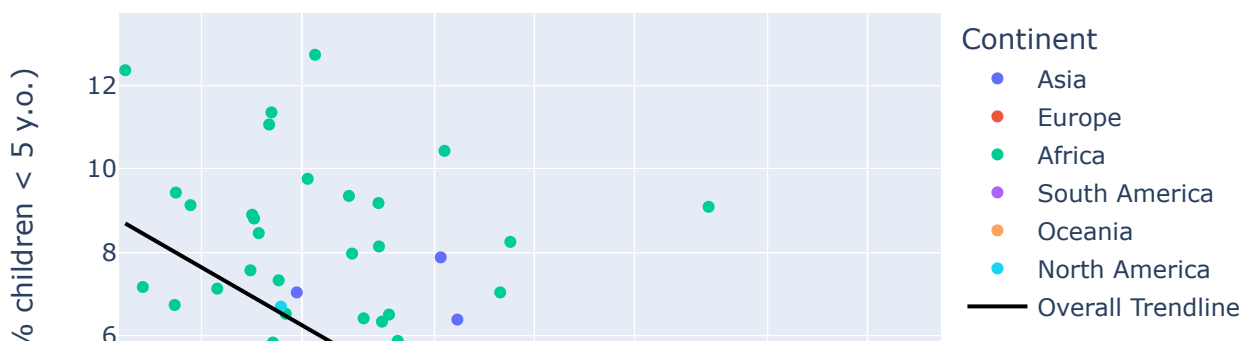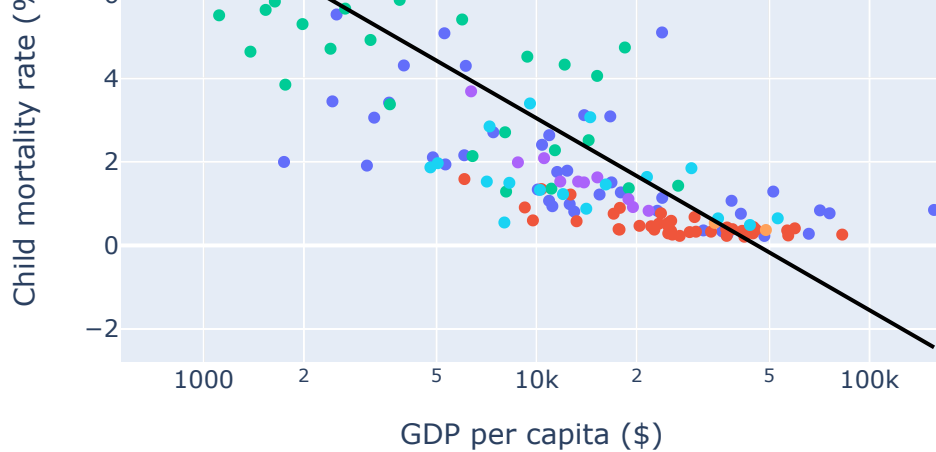# Plot child mortality vs. GDP per capita in 2016
plotFigure(GDP('child-mortality-gdp-per-capita.csv',
        'GDP per capita',
        'Child mortality (Select Gapminder, v10) (2017)',
        'Child mortality vs. GDP per capita', 'GDP per capita'), 'GDP per capita ($)',
        'Child mortality rate (% children < 5 y.o.)', 'Child mortality vs. GDP per cap
        25, pow(10, 4), 3, 1, 14)
```


List of years for which data is available:

[1800, 1801, 1802, 1803, 1804, 1805, 1806, 1807, 1808, 1809, 1810, 1811, 1812, 1813, 1814,
1815, 1816, 1817, 1818, 1819, 1820, 1821, 1822, 1823, 1824, 1825, 1826, 1827, 1828, 1829,
1830, 1831, 1832, 1833, 1834, 1835, 1836, 1837, 1838, 1839, 1840, 1841, 1842, 1843, 1844,
1845, 1846, 1847, 1848, 1849, 1850, 1851, 1852, 1853, 1854, 1855, 1856, 1857, 1858, 1859,
1860, 1861, 1862, 1863, 1864, 1865, 1866, 1867, 1868, 1869, 1870, 1871, 1872, 1873, 1874,
1875, 1876, 1877, 1878, 1879, 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889,
1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, 1900, 1901, 1902, 1903, 1904,
1905, 1906, 1907, 1908, 1909, 1910, 1911, 1912, 1913, 1914, 1915, 1916, 1917, 1918, 1919,
1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929, 1930, 1931, 1932, 1933, 1934,
1935, 1936, 1937, 1938, 1939, 1940, 1941, 1942, 1943, 1944, 1945, 1946, 1947, 1948, 1949,
1950, 1951, 1952, 1953, 1954, 1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964,
1965, 1966, 1967, 1968, 1969, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979,
1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994,
1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009,
2010, 2011, 2012, 2013, 2014, 2015, 2016]

Year of interest for the data "Child mortality vs. GDP per capita":
2016

Figure 14 :

Child mortality vs. GDP per capita, 2016

ren < 5 y.o.)

12

10

Continent

● Asia
● Europe
● Africa
● South America
● Oceania
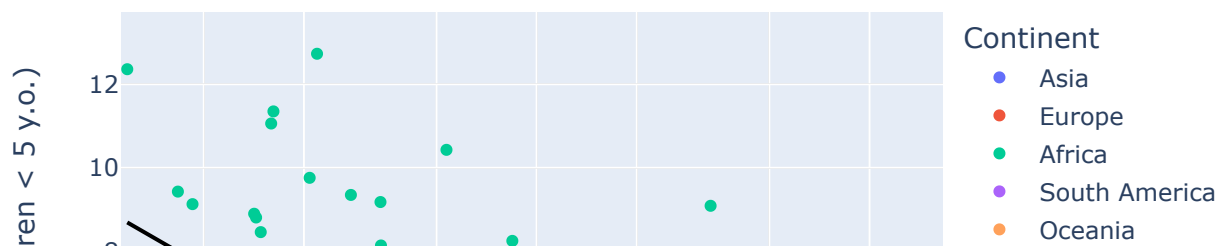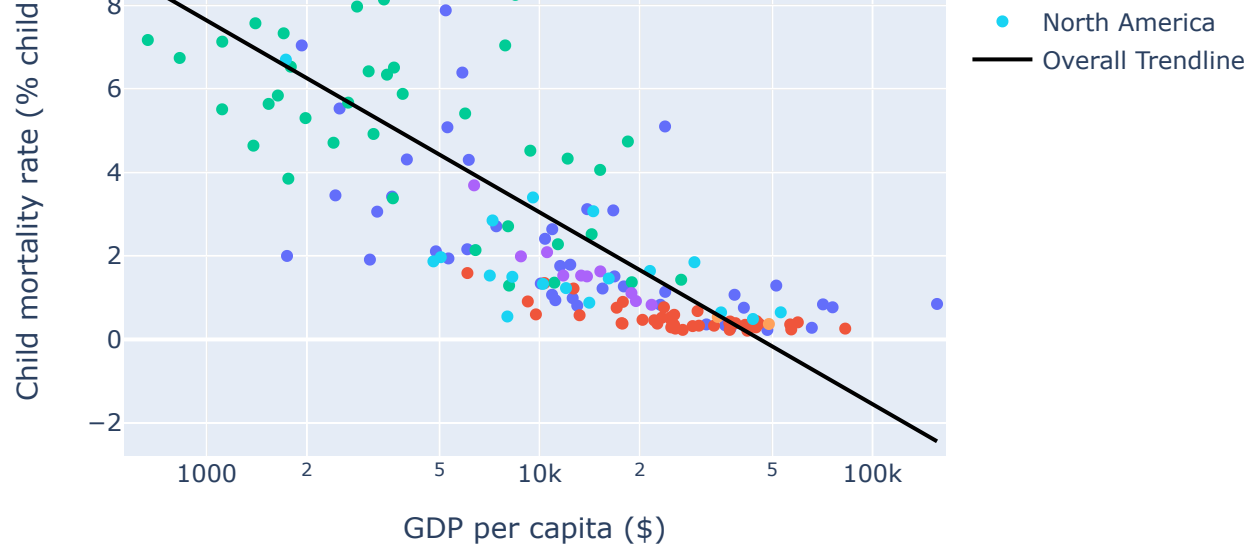
Overall, one can conclude that countries with a low GDP have more issues related to health. By relating the observations drawn from the datasets, one could argue that the life expectancy and child mortality rate could be explained by the number of doctors per 1,000 people. Countries with low GDP per capita have less doctors, a higher child mortality rate and a lower life expectancy.

References

Task 1

1. Data was compiled by Our World in Data (2022): *Life expectancy vs. GDP per capita.* [online] Available at: https://ourworldindata.org/grapher/life-expectancy-years-vs-real-gdp-per-capita-2011us [Accessed 20 January 2022].
   Based on estimates by:
   A. Life expectancy: James C. Riley (2005) – Estimates of Regional and Global Life Expectancy, 1800–2001. Issue Population and Development Review. Population and Development Review. Volume 31, Issue 3, pages 537–543, September 2005., Zijdeman, Richard; Ribeira da Silva, Filipa, 2015, "Life Expectancy at Birth (Total)", http://hdl.handle.net/10622/LKYT53, IISH Dataverse, V1, and UN Population Division (2019)
   B. GDP per capita: Bolt, Jutta and Jan Luiten van Zanden (2020), *"Maddison style estimates of the evolution of the world economy. A new 2020 update".*
2. Plotly. 2022. *Plotly Python Graphing Library.* [online] Available at: https://plotly.com/python/ [Accessed 23 January 2022].
3. Plotly. 2022. *Linear Fits.* [online] Available at: https://plotly.com/python/linear-fits/ [Accessed 23 January 2022].

Task 2

1. Data compiled by Our World in Data (2022): *Self-reported Life Satisfaction vs GDP per capita.* [online] Available at: https://ourworldindata.org/grapher/gdp-vs-happiness [Accessed 20 January 2022].
   Based on estimates by:
   A. Life satisfaction: World Happiness Report (2021) [online] Available at: https://worldhappiness.report/ed/2021/#appendices-and-data [Accessed 20 January 2022].

B. GDP per capita: International Comparison Program - World Bank, World Development Indicators - World Bank, Eurostat-OECD PPP Programme (2021) [online] Available at: http://data.worldbank.org/data-catalog/world-development-indicators [Accessed 20 January 2022].

2. Data compiled by Our World in Data. 2022. *Children per woman by GDP per capita.* [online] Available at: https://ourworldindata.org/grapher/children-per-woman-by-gdp-per-capita [Accessed 20 January 2022].
Based on estimates by:
   A. Children per woman: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects: The 2019 Revision, DVD Edition. [online] Available at: https://population.un.org/wpp2019/Download/Standard/Interpolated/ [Accessed 20 January 2022].
   B. GDP per capita: Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" American Economic Review, 105(10), 3150-3182, available for download at www.ggdc.net/pwt. PWT v9.1

3. Data compiled by Our World in Data. 2022. *Share of adults who smoke vs GDP per capita.* [online] Available at: https://ourworldindata.org/grapher/share-of-adults-who-are-smoking-by-level-of-prosperity [Accessed 23 January 2022].
Based on estimates by:
   A. Share of adults who smoke: Global Health Observatory Data Repository - World Health Organization (2021) [online] Available at: http://data.worldbank.org/data-catalog/world-development-indicators [Accessed 23 January 2022].
   B. GDP per capita: International Comparison Program - World Bank, World Development Indicators - World Bank, Eurostat-OECD PPP Programme (2021). [online] Available at: http://data.worldbank.org/data-catalog/world-development-indicators [Accessed 23 January 2022].

4. Data compiled by Our World in Data. 2022. *Medical doctors per 1,000 people vs. GDP per capita.* [online] Available at: https://ourworldindata.org/grapher/medical-doctors-per-1000-people-vs-gdp-per-capita [Accessed 23 January 2022].
Based on estimates by:
   A. Medical doctors per 1,000 people: Global Health Workforce Statistics - World Health Organization, OECD, official national sources (2021) [online] Available at: http://data.worldbank.org/data-catalog/world-development-indicators [Accessed 23 January 2022].
   B. GDP per capita: International Comparison Program - World Bank, World Development Indicators - World Bank, Eurostat-OECD PPP Programme [online] Available at: http://data.worldbank.org/data-catalog/world-development-indicators [Accessed 23 January 2022].

5. Data compiled by Our World in Data. 2022. *Child mortality vs GDP per capita.* [online] Available at: https://ourworldindata.org/grapher/child-mortality-gdp-per-capita [Accessed 20 January 2022].
Based on estimates by:
   A. Child mortality: Gapminder [online] Available at: https://www.gapminder.org/data/documentation/gd005/ [Accessed 20 January 2022].
   B. GDP per capita: The Maddison Project Database [online] Available at: https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020 [Accessed 20 January 2022].

In [ ]: