In [6]:

```python
from IPython.core.display import Image
Image('https://i.imgur.com/5BCNfD4.png')
```

Out[6]:

Note:

Pham Dang Khoa completed all the Python coding for this report, whereas Pepijn van den Berg and Sebastiaan Beschoor Plug wrote the report.

Table of contents

1. Introduction
    - A. Research question
    - B. Research Methodology
    - C. The dataset
    - D. Sub questions
2. Data Analysis
    - A. Which are the top 10 donors and top 10 recipients of ODA?
    - B. What are the percentages of received or donated funds for different countries?
    - C. Which goals receive the most and the least funding?
    - D. Are their differences among the top 5 donors in terms of donation pattern?
    - E. Does the annual ODA donation change over time?
    - F. How much ODA does each continent contribute and receive?
    - G. How much ODA does each income group contribute and receive?
    - H. How much does each country donate and receive in average?
    - I. For High Income Group: How does their GDP relate to their donation?
    - J. For Upper middle, Lower middle and Low income groups: How does their GDP relate to their received ODA?
3. Further Research
4. Conclusions
5. Sources
6. Appendix

1. INTRODUCTION

1. INTRODUCTION

According to the United Nations 'The Sustainable Development Goals are the blueprint to achieve a better and more sustainable future for all. They address the global challenges we face, including those related to poverty, inequality, climate, environmental degradation, prosperity, and peace and justice.' [1]. There are 17 different goals, which are designed in such a manner that they are interconnected, thus leaving no one behind. Furthermore, the goals are formulated as specific targets, that should be realised by the year 2030.

Official development assistance (ODA) plays an important role in reaching these targets. ODA is defined as government aid designed to promote the economic development and welfare of developing countries [2]. This excludes loans and credits for military purposes. Traditionally, developed nations who are part of the UN should devote 0.7% of their gross national income towards ODA.

The eligibility of ODA are still fundamentally based on a country's economic growth performance. However, according to the OECD, 'A growing consensus among academic, practitioner and political communities reveal that classifying countries according to their per capita income is inadequate to measure well-being or sustainability.' [3]. This means that achieving the Sustainable Development Goals is more complex than just simply investing money.

1A. Research question

In order to gain more insight into the current situation, this report shall focus on the following research question: "How do different characteristics of a country relate to its donation or reception of international aid?" Variables such as geographical (grouped by continent) and economic (GDP) shall be tested. The data used for this research will be provided by AidData [4].

1B. Research Methodology

In order to gain more insight into the relations between a country's characteristics and its spending on, or receiving of ODA, the following research methodology shall be executed:

- Performing research on ODA and the different Sustainable Development Goals.
- Forming a hypothesis and a research question.
- Identifying a suitable dataset (AidData).
- Defining different sub research questions.
- Structuring and if necessary, cleaning the dataset using python.
- Performing visual analysis on the dataset using python.
- And lastly, testing our hypothesis.

In this paper we will showcase the visualy analysis of each sub question identified below, accompanied by their respective conclusions.

1C. The dataset

This analysis is based on data provided by the AidData, a research lab at the William & Mary's Global Research Institute. Their mission is to equip policymakers and practitioners with data-based evidence to improve the way sustainable development investments are targeted, monitored and evaluated (AidData.org). We view the dataset form AidData as reliable, as most of their aid activity records originate from the Creditor Reporting System (CRS), the central database for foreign aid compiled by the OECD's Development Assistance Committee (DAC). In addition, numerous other academic peer-reviewed papers use data sets from AidData.

The dataset we downloaded is called "Financing to the SDGs Dataset, Version 1.0", a csv file created in November 2017. It tracks $1.5 trillion from 2000 to 2013 spent in 1,252,036 Official Development Assistance (ODA) projects from the AidData's project-level database. According to the website this is the "most comprehensive project-level data tracking international development finance" [5]. Each project is an entry in the csv, containing the project id, the commencement year, the donor, the recipient, the project title, short description, long description, AidData activity code, AidData purpose code, committed dollars, SDG committed dollars and then the amount of dollar committed to each of the 17 SDG's.

With financial help from the William and Flora Hewlett Foundation, the dataset first mapped the 544 activity codes developed by AidData to SDG's, which allowed them to use the projects research in an earlier dataset AidData Core Research Release, Version 3.1 [5]. The they split each project up into the different activities and subsequently connected each activity to the associated SDG goal [6].

Using this dataset as the basis for our analysis, we do have to recognize some of its limitations. Due to certain donor practices, difficulties arise is aligning the project with an SDG, either because of lack of details or vagueness of the project. The interrelatedness of SDG's does not make it any easier, nor does the fact that an ODA project usually involves multiple activities. Both of these characteristics make discrete tracking of finances difficult. In addition, the current methodology (which they are going to change in future datasets), divide the project financing equally amount the activities and the associated SDG's. However, this is not a fair representation of reality [6].

AidData themselves also report a limitation with regard to their own activities code, as it aligns better with some SDG's than others, for example SDG 3 (Health) or SDG4 (Education). By matching better, more financing could be mapped, distorting the data. In addition, for certain SDG's, limited funding is reported. Such is the case with SDG's on environment (13, 14, 15). This may reflect the low priority given to the environment by the international community, but may be the result of the specificity of the SDG's; climate change, oceans and land ecosystems. As a result, when a project has the description "General environmental Protection", they are too broad to be linked to contributing to any SDG. This could explain the low expenditure on these SDG's. This is also an issue with activities codes such as rural development and population.

In addition, we also used two datasets from World Bank [9]:

- SDGCountry: grouping the countries into 7 geographical regions (East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, North America, South Asia, Sub-Sahran Africa) and 4 income groups (High income, Upper middle income, Lower middle income, Low income).
- SDGData: this dataset contains data in a wide range of indicators. In this study, we focused on only 2 indicators in this dataset: GDP (NY.GDP.MKTP.CD) and GDP per capita (NY.GDP.PCAP.CD).

1D. Sub questions

The extensiveness of the dataset as well as the broadness of the problem, means that defining different sub questions will help to better structure the research. The sub questions are designed in such a manner that the first ones will help us to better understand the lay of the land. The sub questions thereafter will help give more specific insight into our research question. Therefore, following this structure, we shall be able to better understand the dynamics with regard to ODA and thereby answer the question how different characteristics of a country relate to its spending on, or receiving of international aid. The sub questions are as follows:

- Which are the top 10 donors and top 10 recipients of ODA?
- What are the percentages of received or donated funds for different countries?
- Which goals receive the most and the least funding?
- Are their differences among the top 5 donors in terms of donation pattern?
- Does the annual ODA donation change over time?
- How much ODA does each continent contribute and receive?
- How much ODA does each income group contribute and receive?
- How much does each country donate and receive in average?

- For High Income Group: How does their GDP relate to their donation?
- For Upper middle, Lower middle and Low income groups: How does their GDP relate to their received ODA?

We end this paper with a short summary of how the data set answered the questions above.

2. DATA ANALYSIS

We begin by importing the libraries we use, and reading the data set.

In [7]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from operator import add
import scipy.stats as sp
!pip install squarify
import squarify
%matplotlib inline
plt.style.use('ggplot')
sns.set_palette('Set3',12)
```

- Numpy is used for mathematical operations on our data
- Pandas is used for manipulating (merging, slicing, assigning,...) our dataframes
- Mathplotlib is used for plotting the graphs
- Seaborn is used to set and change the color palettes (the default palette in this report is "Set3" with 12 different colors, which is suitable for qualitative data visualization). Specific graphs that require specific palettes will be adjusted accordingly.
- Operator is used for simple mathematical operation
- Scipy is used for statistical analysis
- Squarify is used to plot the treemap (rectangular chart)

In [8]:

```
# Read the ODA data (spending broken down into donors, recipients, individual
SDG,...)
ODA_data = pd.read_csv('FinancingtotheSDGsDataset_v1.0.csv')
ODA_data[:2]
```

Out[8]:

	aiddata_id	year	donor	recipient	title	short_description	long_description	aiddata_act
0	54886623.0	2007	Germany	Brazil	Environmental education via visual media in Br...	ENVIRONMENTAL EDUCATION/TRAINING	Environmental education via visual media in Br...	
1	54886662.0	2003	Belgium	Congo, Democratic Republic of	MSF / APPUI À 14 ZONES DE SANTÉ AU SHABA, EN ...	APPUI À 14 ZONES DE SANTÉ AU SHABA, EN EQUATEUR	NaN	

2 rows × 28 columns

The data set is split up into specific donations, with a recipient and a donor. The year and a short description are provided, as well as to into what specific SDG the aid can be grouped.

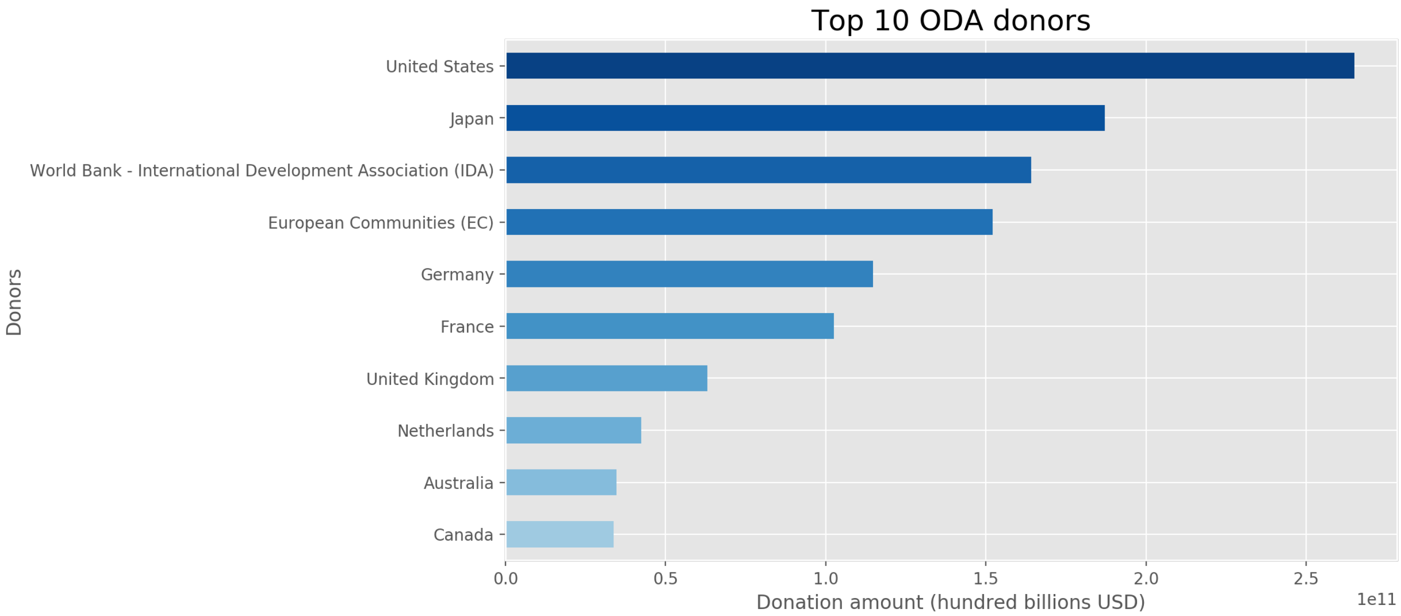
2A. Which are the top 10 donors and top 10 recipients of ODA?

In order to better understand the spending and receiving of ODA, we shall look at the top 10 donors and the top 10 recipients. This way, we shall be introduced to the different key players and learn how much they spend or receive.

Top 10 ODA Donors
The number one ODA donor is the United States of America which donated significatly more than other donors. The top ten is rounded of by a number of European countries, as well as the World Bank, Canada, Japan and Australia. All developed countries, with an HDI above 0.9 [7].

In [9]:

```
# Calculate the amount of donations for all donors, then choose the top 10 donors
donor = ODA_data.groupby('donor').sdg_commitment_amount_usd_constant.sum()
top_donor = donor.sort_values(ascending=False)[:10]

# Plot the top 10 donors
plt.figure(figsize=(10,6), dpi=200)
donor_plt = top_donor.plot.barh(color=sns.color_palette('Blues_r',15))
donor_plt.invert_yaxis()
donor_plt.set_ylabel('Donors',size=12)
donor_plt.set_xlabel('Donation amount (hundred billions USD)',size=12)
donor_plt.set_title('Top 10 ODA donors',size=18)
plt.show()
```

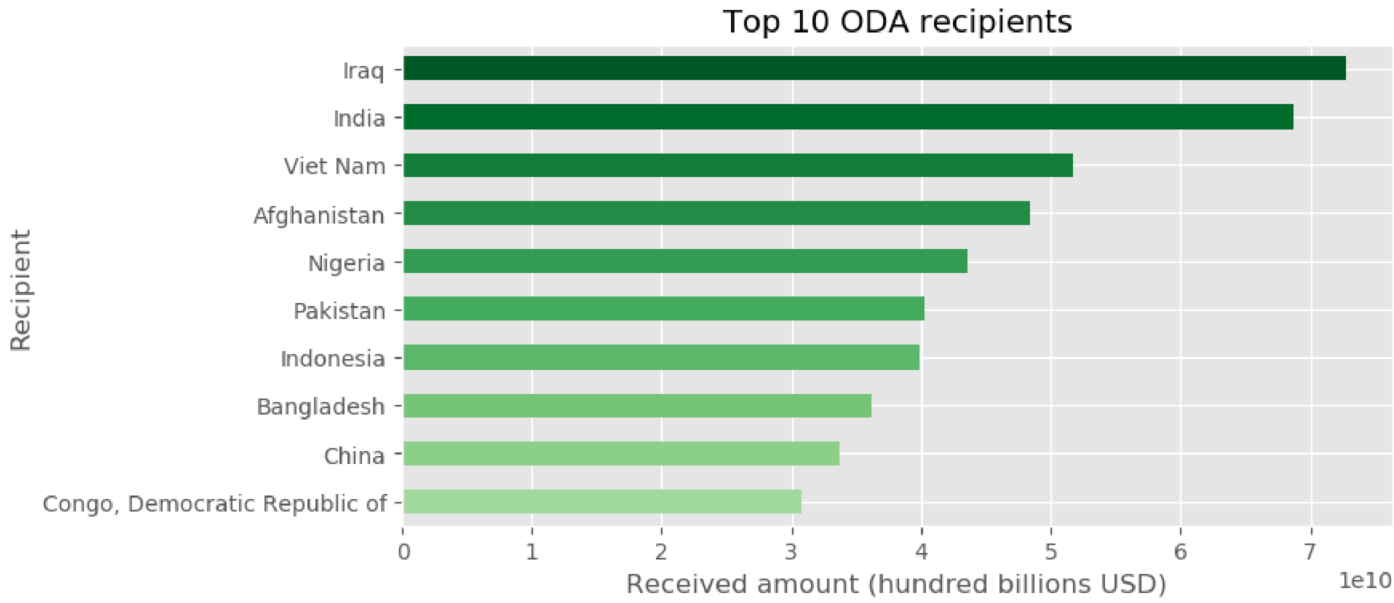


Top 10 ODA Recipients.
All countries in top 10 recipients are from Asia and Africa.

In [10]:

```
# Calculate the amount received by recipients, then choose the top 10 donors
recipient = ODA_data.groupby('recipient').sdg_commitment_amount_usd_constant.sum()
recipient = recipient[recipient.index != 'Bilateral, unspecified'] # Remove ODA pr
ojects that do not specify recipients.
top_recipient = recipient.sort_values(ascending=False)[:10]

# Plot the top 10 donors
plt.figure(figsize=(8,4), dpi=100)
```

```
recipient_plt = top_recipient.plot.barh(color=sns.color_palette('Greens_r',15))
recipient_plt.invert_yaxis()
recipient_plt.set_xlabel('Received amount (hundred billions USD)',size=12)
recipient_plt.set_ylabel('Recipient',size=12)
recipient_plt.set_title('Top 10 ODA recipients',size=14)
plt.show()
```



2B. What are the percentages of received or donated funds for different countries?

After having defined the top 10 donors and recipients, we want to compare their contribution to the total amount of ODA. We can do this by taking the percentages of total funding.

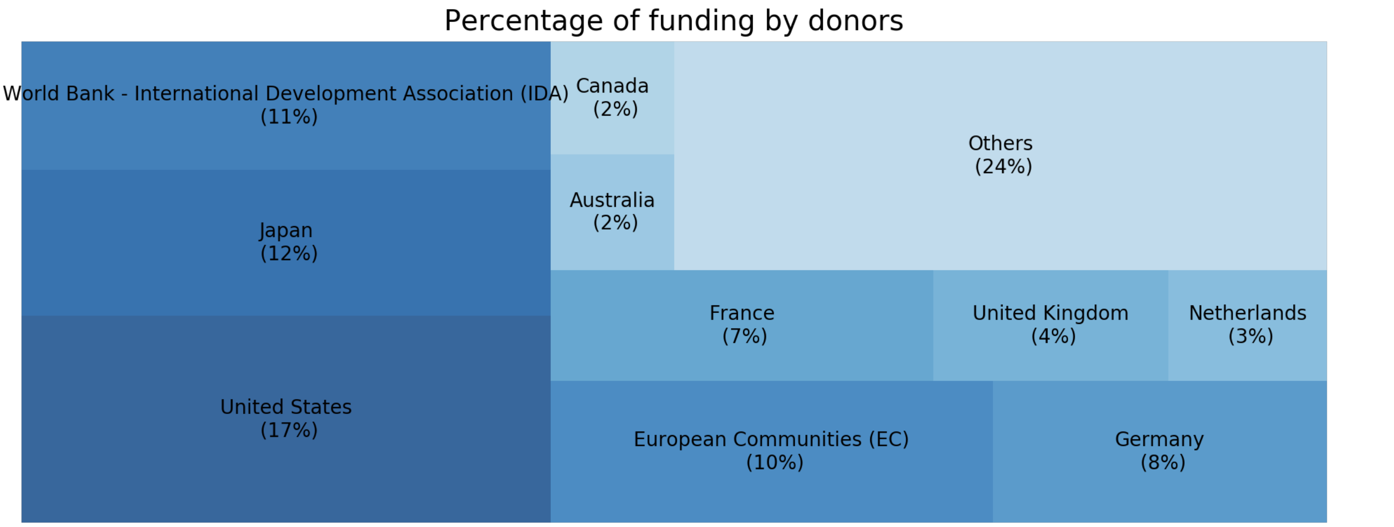
Percentage of funding by donors
When looking at the percentages it can be seen that the USA donates as much as 17% of the total amount of ODA. The other top 10 donors amount to 59%. This means that the top 10 donors are responsible for 76% of all donations.

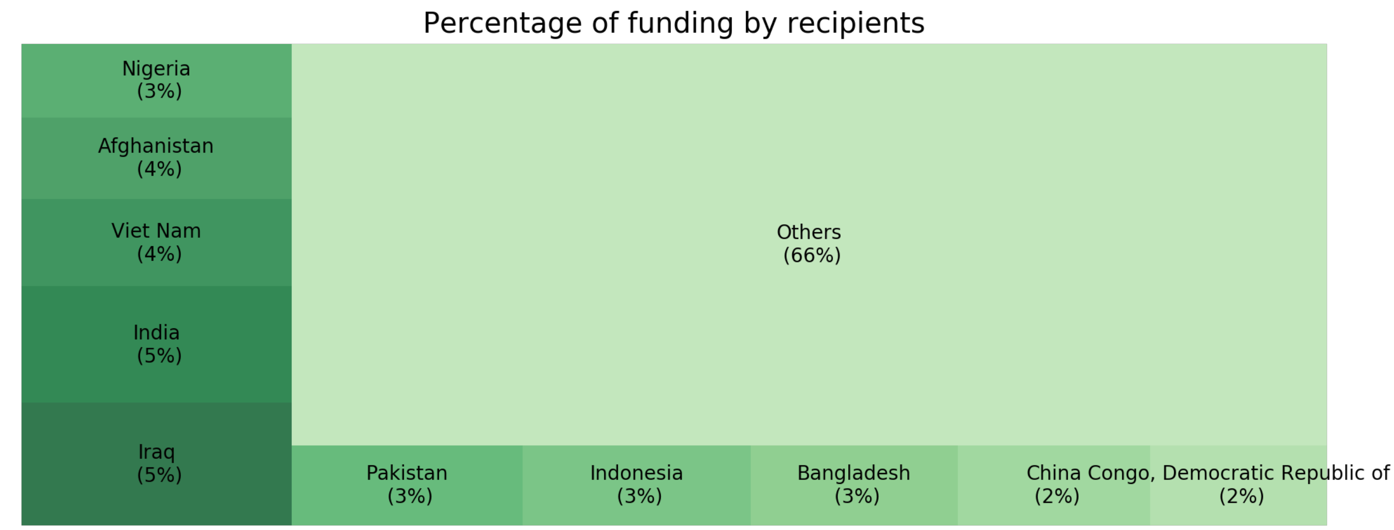
Percentage of funding by recipients
Compared to the top 10 donors, the difference among the top 10 recipients is a lot smaller. The top 10 recipients all receive between five and two percent. This means they only receive about 24% of all funding.

In [11]:

```
# Calculate percentage of top 10 donors and recipients
percent_donor = top_donor/donor.values.sum()*100
percent_recipient = top_recipient/recipient.values.sum()*100

# Calculate percentage of other countries
percent_donor.at['Others'] = 100 - percent_donor.sum()
percent_recipient.at['Others'] = 100 - percent_recipient.sum()

# Plot 2 pie chart side by side:
fig = plt.figure(figsize=(12,10),dpi=200)

# Pie number 1
ax1 = plt.subplot2grid((2,1),(0,0))
labels = percent_donor.to_frame().reset_index().apply(lambda x: str(x[0]) + "\n ("
+ str(round(x[1])) + "%)", axis=1)
squarify.plot(sizes=percent_donor, label=labels, color=sns.color_palette('Blues_r'
,15), alpha=.8)
```

```
plt.title('Percentage of funding by donors')
plt.axis('off')

# Pie number 2
ax1 = plt.subplot2grid((2,1),(1,0))
labels = percent_recipient.to_frame().reset_index().apply(lambda x: str(x[0]) +
"\n (" + str(round(x[1])) + "%)", axis=1)
squarify.plot(sizes=percent_recipient, label=labels, color=sns.color_palette('Greens_r',15), alpha=.8)
plt.title('Percentage of funding by recipients')
plt.axis('off')

plt.show()
```





Conclusion
The easy conclusion that can be drawn, looking at these graphs, is that the top 10 donors are responsible for most of the funding, whereas the recipients are much closer together.

2C. Which goals receive the most and the least funding?

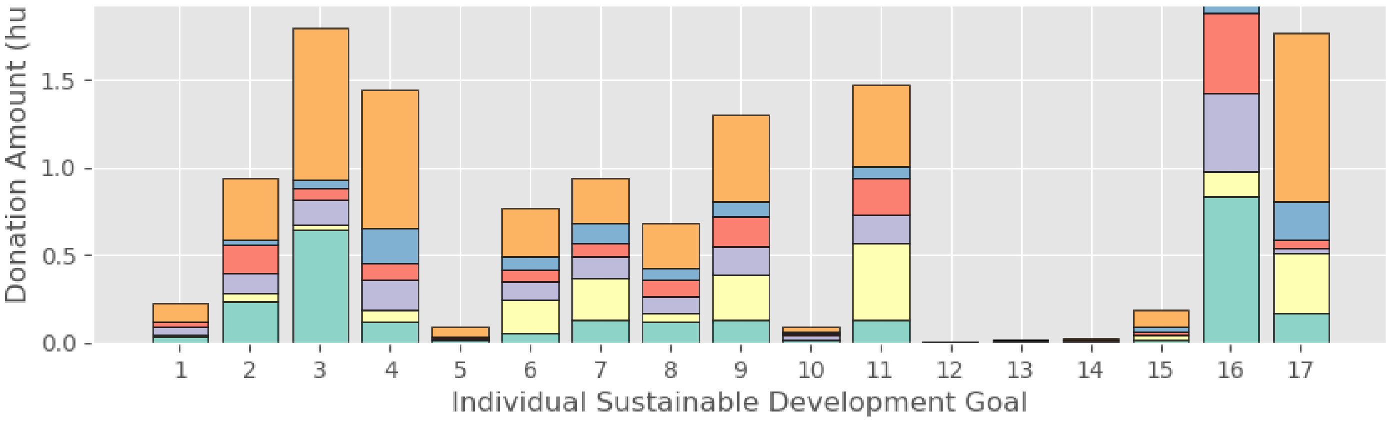
Having a better grasp on the distribution on receiving and spending between different countries, we shall now further look into the differences between the Sustainable Development Goals, grouped per donor.
On the x-axis we have the 17 different SDG. A full list of the SDG can be found in Appendix A.

In [12]:

```
# Which goal receive the most and least funding?
```

```
goal = []
for i in range(-17,0):                          # The funding data for 17 SDGs ar
in the last 17 columns
    element = ODA_data.iloc[:,i].sum()
    goal.append(element)
print('Goal %d received the most ODA funding with %.2f billion USD.' % (goal.index
(max(goal))+1, max(goal)/10**9))
print('Goal %d received the least ODA funding with %.2f billion USD.' % (goal.inde
x(min(goal))+1, min(goal)/10**9))

# How much does each country donate for each goal?
# Calculate donation for each goal of top 5 donors:
top5 = {}
for i in range (5):
    country = ODA_data [ODA_data['donor'] == top_donor.index[i]] # Obtain the data
for top 5 countries
    lst1 = []
    for j in range(-17,0):                        # The funding data for 17 SDGs a
e in the last 17 columns.
        element = country.iloc[:,j].sum()
        lst1.append(element)
    top5[top_donor.index[i]] = lst1
df = pd.DataFrame(top5).transpose()

# Calculate donation for each goal of other countries:
lst2 = []
for i in range(17):
    element = ODA_data.iloc[:,i-17].sum() - df.iloc[:,i].sum()
    lst2.append(element)
df.loc['Others',:]=lst2

# Plot the result
plt.figure(figsize=(10,5), dpi=100)

plt.bar(df.columns+1,list(df.iloc[0,:]),edgecolor='black')
plt.bar(df.columns+1,list(df.iloc[1,:]),bottom = list(df.iloc[0,:]),edgecolor='bla
ck')

btm = list(df.iloc[0,:])
for v in range(2,6):
    btm = list(map(add, btm, list(df.iloc[v-1,:])))
    plt.bar(df.columns+1,list(df.iloc[v,:]),bottom = btm,edgecolor='black')

plt.xticks(range(1,18), df.columns+1)
plt.title("Donation for each SDG (grouped by donors)",size=14)
plt.ylabel('Donation Amount (hundred billion USD)')
plt.xlabel('Individual Sustainable Development Goal')
plt.legend(list(df.index), bbox_to_anchor=(0.61, 0.7), fontsize=10, frameon = Fals
e, labelspacing=-2.5)
plt.show()
```

```
Goal 16 received the most ODA funding with 342.55 billion USD.
Goal 12 received the least ODA funding with 0.62 billion USD.
```



As can be seen in the graph, the proportions of the donors per SDG remains relatively the same, meaning that for the most part the donors have the same priorities with regard to the SDG. However, the differences between the individual SDG are quite staggering. Peace, justice and strong institutions (16) receives by far the most funding, close to 350 billion dollar. Other populair goals are partnerships (17), good health and wellbeing (3), quality education (4), industry, innovation and infrastructure (9) and sustainable cities and communities (11).

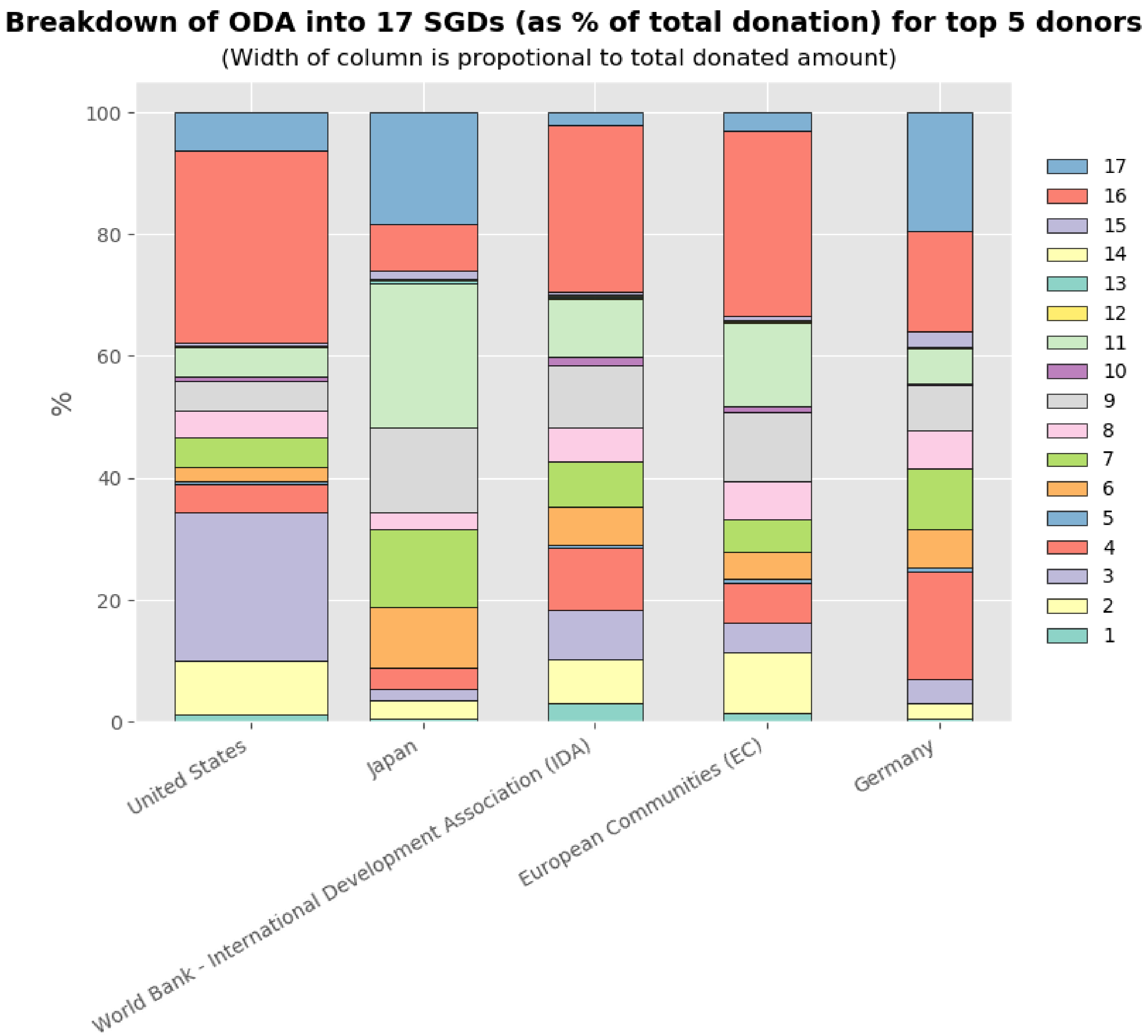
2D. Are their differences among the top 5 donors in terms of donation pattern?

In the previous sub question, the assumption was made that for the most part, the donors have very similar priorities. However, we shall further test this by looking at the differences among the top 5 donors, in terms of donation pattern.

In [13]:

```
# Calculate percentage of funding that each goal receives for top 5 donors
top5_percentage = {}
for i in range (5):
    country = ODA_data [ODA_data['donor'] == top_donor.index[i]]
    total = country.sdg_commitment_amount_usd_constant.sum()
    lst = []
    for j in range(-17,0):
        element = country.iloc[:,j].sum()/total*100
        lst.append(element)
    top5_percentage[top_donor.index[i]] = lst

df = pd.DataFrame(top5_percentage).transpose()
df.columns = range(1,18)

# Plot the result
plt.figure(figsize=(8,6), dpi=100)

plt.bar(top_donor.index[0:5],list(df[1]),edgecolor='black',width=top_donor[0:5]/3e
11)
plt.bar(top_donor.index[0:5],list(df[2]),bottom = list(df[1]),edgecolor='black',wi
dth=top_donor[0:5]/3e11)

btm = list(df[1])
for v in range(3,18):
    btm = list(map(add, btm, list(df[v-1])))
    plt.bar(top_donor.index[0:5],list(df[v]),bottom = btm,edgecolor='black',width=
top_donor[0:5]/3e11)

plt.xticks(range(0,5), top_donor.index[0:5], rotation=30,
horizontalalignment='right', size=10)
```

```
plt.suptitle('(Width of column is propotional to total donated amount)', y=0.92)
plt.title("Breakdown of ODA into 17 SGDs (as % of total donation) for top 5 donors
",
          y=1.06, size=14, weight='bold')
plt.ylabel('%', size=14)
plt.legend(list(df.columns), loc=6, bbox_to_anchor=(1.02, 0.5), labelspacing=-2.5,
frameon=False, fontsize=10)
plt.show()
```



In this graph, the differences become much clearer. For example, Japan and Germany are much more keen on building good partnerships (17). The US, World Bank and the EC spend more heavily on peace, justice and strong institutions (16), compared to Germany and Japan. Furthermore, the US spend much more on quality education (3), compared to the rest. Meanwhile, Japan is the top donor when it comes to investing in sustainable cities and communities (11). For other SDG's such as decent work and economic growth (8), the differences aren't that big. However, from the graph, it can be concluded that there are significant differences among the top 5 donors, in terms of the donation pattern.

2E. Does the annual ODA donation change over time?

After defining the top 10 donors and recipients, as well as looking into the breakdown of aid per SDG, we shall now look at how the annual ODA donations change over time. We shall do this by looking at the top 5 donors.

In [14]:

```
# Check our original dataset ODA_data is from which year to which year:
print(ODA_data['year'].min())
print(ODA_data['year'].max())
```

```
2000
2013
```

In [15]:

```
# Create a dataframe with the period 2000-2013 as first column
time_df = pd.DataFrame()
time_df.insert(0, 'Year', range(2000,2014))

# Calculate the annual donation from top 5 donors
for i in range(5):
    country = ODA_data [ODA_data['donor'] == top_donor.index[i]]
    year_total = country.groupby('year').sdg_commitment_amount_usd_constant.sum()
    time_df.insert(i+1,top_donor.index[i],year_total.values)

# Calculate the annual donation from other donors
all_donation = ODA_data.groupby('year').sdg_commitment_amount_usd_constant.sum()
others_donation = np.subtract(all_donation.values, time_df.iloc[:,1:].sum(axis=1).
values)
time_df['Others'] = others_donation
time_df
```

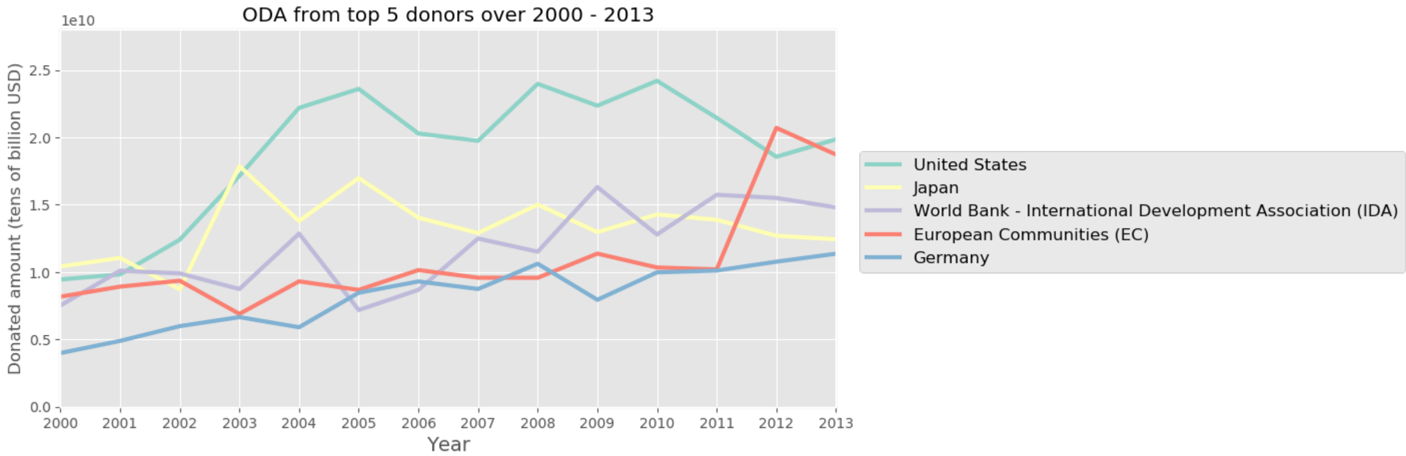
Out[15]:

	Year	United States	Japan	World Bank - International Development Association (IDA)	European Communities (EC)	Germany	Others
0	2000	9.453727e+09	1.042232e+10	7.526544e+09	8.175954e+09	3.989517e+09	2.993058e+10
1	2001	9.830561e+09	1.105521e+10	1.010353e+10	8.920531e+09	4.893227e+09	3.026127e+10
2	2002	1.240460e+10	8.727203e+09	9.901537e+09	9.362426e+09	5.987943e+09	3.923048e+10
3	2003	1.714780e+10	1.787365e+10	8.743825e+09	6.896189e+09	6.658820e+09	3.795284e+10
4	2004	2.219301e+10	1.379737e+10	1.285786e+10	9.313664e+09	5.905555e+09	3.719104e+10
5	2005	2.360352e+10	1.698432e+10	7.183938e+09	8.678043e+09	8.467866e+09	4.990110e+10
6	2006	2.029161e+10	1.404144e+10	8.682095e+09	1.015104e+10	9.305787e+09	5.507981e+10
7	2007	1.974779e+10	1.290244e+10	1.249197e+10	9.583938e+09	8.746785e+09	4.369002e+10
8	2008	2.398034e+10	1.501276e+10	1.151775e+10	9.578603e+09	1.062209e+10	4.832045e+10
9	2009	2.235625e+10	1.297443e+10	1.630937e+10	1.137294e+10	7.945197e+09	5.612636e+10
10	2010	2.420387e+10	1.427951e+10	1.279299e+10	1.034304e+10	9.982448e+09	5.505631e+10
11	2011	2.145346e+10	1.388665e+10	1.574017e+10	1.020266e+10	1.011533e+10	4.007341e+10
12	2012	1.856315e+10	1.269777e+10	1.550593e+10	2.070870e+10	1.077403e+10	5.528546e+10
13	2013	1.986029e+10	1.243444e+10	1.480629e+10	1.872875e+10	1.136627e+10	5.823973e+10

In [16]:

```
# Draw a Line graph to compare top 5 donors over the years
fig, ax = plt.subplots(1,1,figsize=(10,5),dpi=100)
plt.plot(time_df.Year.values, time_df.iloc[:,1:6].values, lw=3)
```

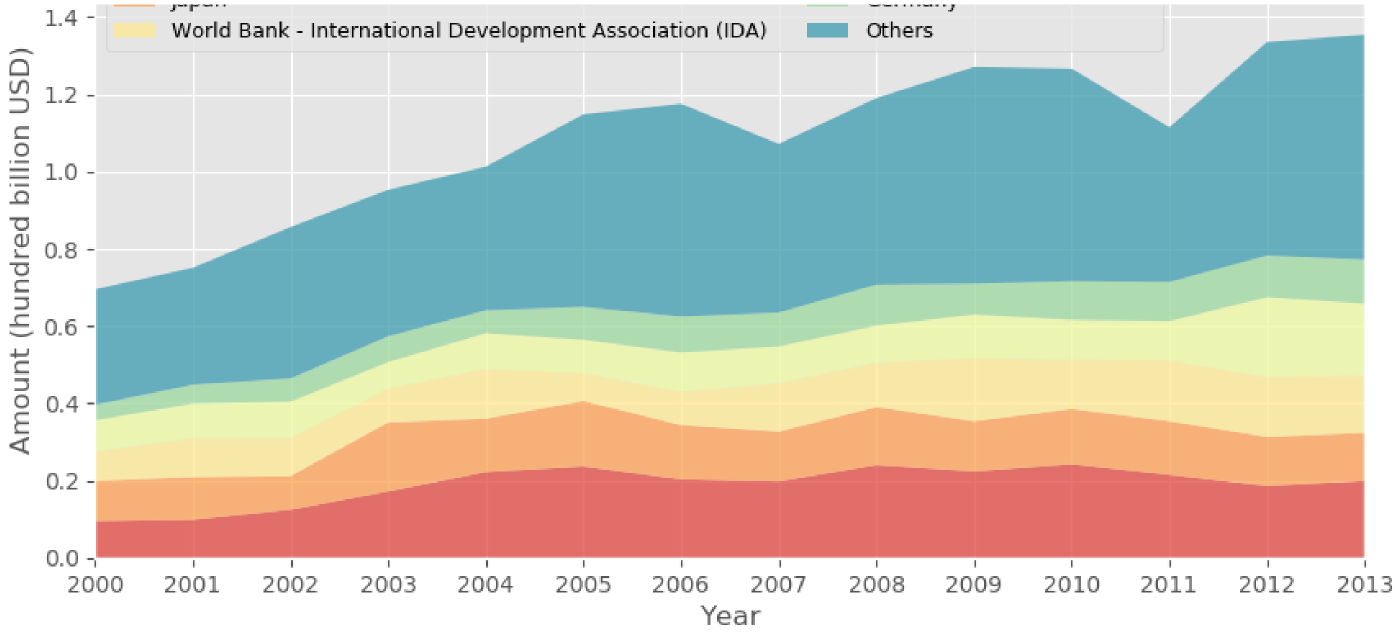
```
# Decoration
plt.legend(time_df.iloc[:,1:6].columns,bbox_to_anchor=(1.02, 0.7),fontsize=12)
plt.xticks(time_df.Year.values)
plt.xlabel('Year',size=14)
plt.ylabel('Donated amount (tens of billion USD)')
plt.xlim(2000,2013)
plt.ylim(0,2.8e10)
plt.title('ODA from top 5 donors over 2000 - 2013')
plt.show()
```



Looking at graph shown above, it can be concluded that although fluctuations can be distinguished, the overall trend from 2000-2013 is positive. In other words, the top 5 donors donate more in 2013 than they have done in 2013. Especially the US has seen a rise in the early 2000's, but has stagnated since then. The same is true for Japan, having peaked in 2003. The EC however have doubled their ODA in 2012.

In [17]:

```
# Plot a stacked area chart to visualize the time_df
fig, ax = plt.subplots(1,1,figsize=(10,5),dpi=100)
columns = time_df.columns[1:]
labs = columns.values.tolist()

# Prepare data
x  = time_df['Year'].values.tolist()
y = time_df[columns[0]].values.tolist()
for i in range(1,6):
    y_i = time_df[columns[i]].values.tolist()
    y = np.vstack([y, y_i])

# Plot the stacked area chart
ax = plt.gca()
ax.stackplot(x, y, labels=labs, colors=sns.color_palette("Spectral",6), alpha=0.8)

# Decorations
ax.set_title('Stacked area chart of ODA donation over 2000-2013', fontsize=14)
ax.legend(fontsize=9, ncol=2, loc='upper left')
plt.xticks(x, fontsize=10, horizontalalignment='center')
plt.yticks(np.arange(0,1.6e11,0.2e11), fontsize=10)
plt.xlim(x[0], x[-1])
plt.ylim(0, 1.6e11)
plt.xlabel('Year')
plt.ylabel('Amount (hundred billion USD)')
plt.show()
```



2F. How much ODA does each continent contribute and receive?

We have defined the top 10 donors and recipients, but in order to better understand the geographical relations, we have defined how much ODA different geographical regions contribute and receive.

In [18]:

```
# SDGCountry.csv classfies countries into regions and income groups
SDGCountry = pd.read_csv('SDGCountry.csv')
# Filter the countries only, excluding continents/groups/...:
country_filter = SDGCountry[pd.notnull(SDGCountry.Region)]
# Subset only 4 columns we need:
country_data = country_filter[['Country Code','Short Name','Region','Income Group'
]]
country_data[:2]
```

Out[18]:

	Country Code	Short Name	Region	Income Group
0	**ABW**	**Aruba**	**Latin America & Caribbean**	**High income**
1	**AFG**	**Afghanistan**	**South Asia**	**Low income**

In [19]:

```
# Merge 2 dataframe ODA_data and country_data by donors
# Change the column name 'Short Name' of country_data to be the same ('donor')
with ODA_data,
# then merge 2 dataframe:
country_data1 = country_data.rename(columns={'Short Name':'donor'})
ODA_data1 = pd.merge(country_data1, ODA_data, on="donor", how="outer" )

# Calculate the donation of each region and "others" (organizations, unspecified
donors,...):
region_donation = ODA_data1.groupby('Region').sdg_commitment_amount_usd_constant.s
um()
sum_donor =  ODA_data.sdg_commitment_amount_usd_constant.sum()
region_donation['Others'] = sum_donor - region_donation.sum()
region_donation
```

Out[19]:

```
Region
East Asia & Pacific            2.360290e+11
Europe & Central Asia          5.056936e+11
Latin America & Caribbean      0.000000e+00
Middle East & North Africa     1.945905e+10
North America                  2.987755e+11
South Asia                     0.000000e+00
Sub-Saharan Africa             0.000000e+00
Others                         4.595024e+11
Name: sdg_commitment_amount_usd_constant, dtype: float64
```

In [20]:

```
# Merge 2 dataframe ODA_data and country_data by recipients
# Change the column name to be the same ('recipient') with ODA_data, then merge 2 dataframe:
country_data2 = country_data.rename(columns={'Short Name':'recipient'})
ODA_data2 = pd.merge(country_data2, ODA_data, on="recipient", how="outer" )

# Calculate the funding received by each region and "others" (organizations, unspecified recipients,...):
region_recipient = ODA_data2.groupby('Region').sdg_commitment_amount_usd_constant.sum()
region_recipient['Others'] = sum_donor - region_recipient.sum()
region_recipient
```

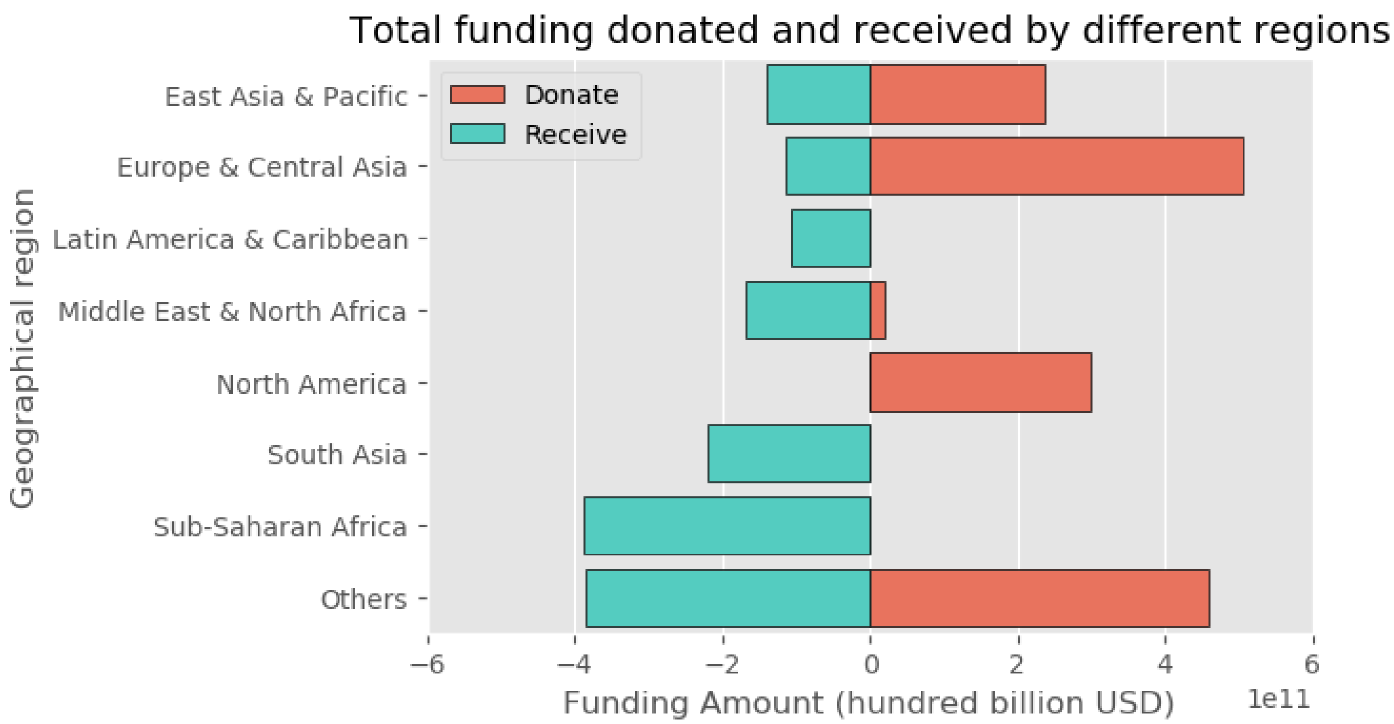
Out[20]:

```
Region
East Asia & Pacific            1.404248e+11
Europe & Central Asia          1.151183e+11
Latin America & Caribbean      1.069772e+11
Middle East & North Africa     1.671817e+11
North America                  1.695860e+05
South Asia                     2.193980e+11
Sub-Saharan Africa             3.862008e+11
Others                         3.841585e+11
Name: sdg_commitment_amount_usd_constant, dtype: float64
```

In [21]:

```
# To visualize the funding amount donated and received by each region:
# Creat dataframe of funding donated by regions:
region_df1 = pd.DataFrame(columns=['Category', 'Mode','Value'])
region_df1['Category'] = region_donation.keys()
region_df1['Mode'] = ['Donate']*len(region_donation)
region_df1['Value'] = region_donation.values

# Creat dataframe of funding received by regions:
region_df2 = pd.DataFrame(columns=['Category', 'Mode','Value'])
region_df2['Category'] = region_recipient.keys()
region_df2['Mode'] = ['Receive']*len(region_recipient)
region_df2['Value'] = -region_recipient.values

# Concatenate 2 dataframe:
region_df = pd.concat([region_df1,region_df2], ignore_index=True, axis=0, sort = False)

# Draw Plot
grp = country_data['Region'].unique()
plt.figure(figsize=(6,4), dpi=100)
group_col = 'Mode'
```

```
colors = ['tomato','turquoise']

for c, group in zip(colors, region_df[group_col].unique()):
    sns.barplot(x='Value', y='Category', data=region_df.loc[region_df[group_col]==
group, :],
                 color=c, label=group, edgecolor='black')

# Decorations
plt.xlabel("Funding Amount (hundred billion USD)")
plt.ylabel("Geographical region")
plt.yticks(fontsize=10)
plt.xlim(-6e11,6e11)
plt.title("Total funding donated and received by different regions", fontsize=14)
plt.legend(loc='upper left')
plt.show()
```

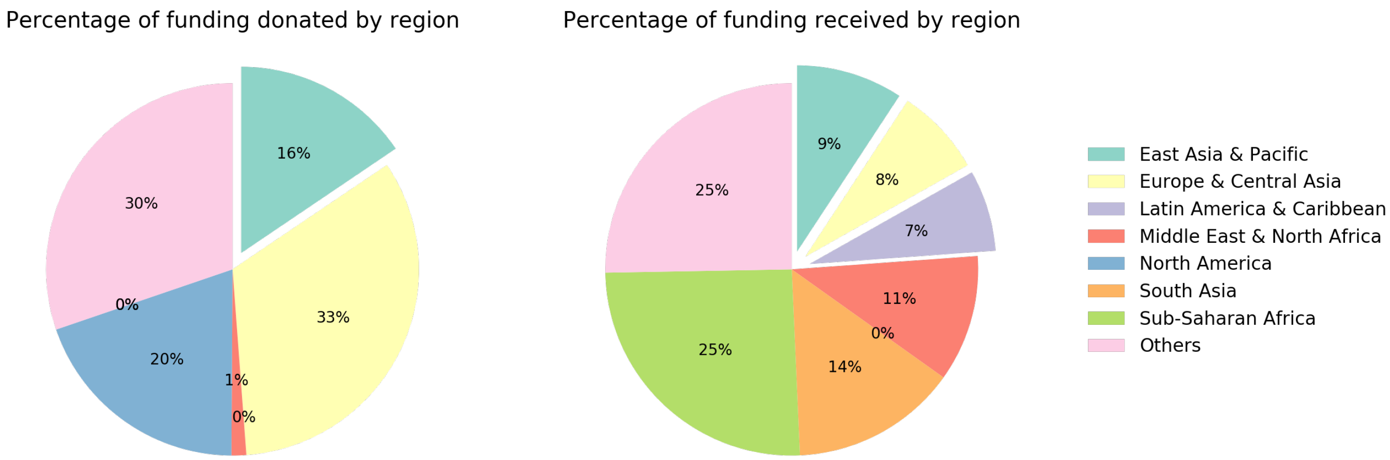


Unsurprisingly, regions that contribute a lot receive very little to none. The same goes the other way, regions who receive a lot donate very little to none. Also to no surprise, developed regions like Europe and North America can be categorised as donors, whereas underdeveloped regions, such as Sub Saharan-Africa, South Asia and Latin America and the caribbeans can be categorised as recipients. The "Others" category refer to international organizations such as World Bank that cannot be classified into a single region. In order to further prove this, the graph below shows the different percentages per region.

In [22]:

```
# Calculate the percentage of ODA donated & received by each region:
percent_donor_region = region_donation/sum_donor*100
percent_recipient_region = region_recipient/sum_donor*100

# Plot 2 pie chart side by side:
fig = plt.figure(figsize=(12,8),dpi=200)
sns.set_palette('Set3',8)

# Pie number 1
ax1 = plt.subplot2grid((1,2),(0,0))

plt.pie(percent_donor_region, startangle=90, counterclock=False, autopct='%.0f%%',
        explode=[0.1,0,0.2,0,0,0,0,0])
```

```
plt.title('Percentage of funding donated by region')

# Pie number 2
ax1 = plt.subplot2grid((1,2),(0,1))
plt.pie(percent_recipient_region, startangle=90, counterclock=False, autopct='%.0f
%%',
        explode=[0.1,0.1,0.1,0,0,0,0,0])
plt.title('Percentage of funding received by region')

# Add a legend
plt.legend(percent_donor_region.index, bbox_to_anchor=(1.1, 0.8), frameon=False, f
ontsize=12)
plt.show()
```



The conclusion that can be drawn from the graphs, is that the percentage of funding donated or received is highly dependable on the region, and whether or not this is a developing or developed region. This of course leads to the hypothesis that a country's income is a determining factor in how much they either receive or donate. With the next research question, we shall test this hypothesis.

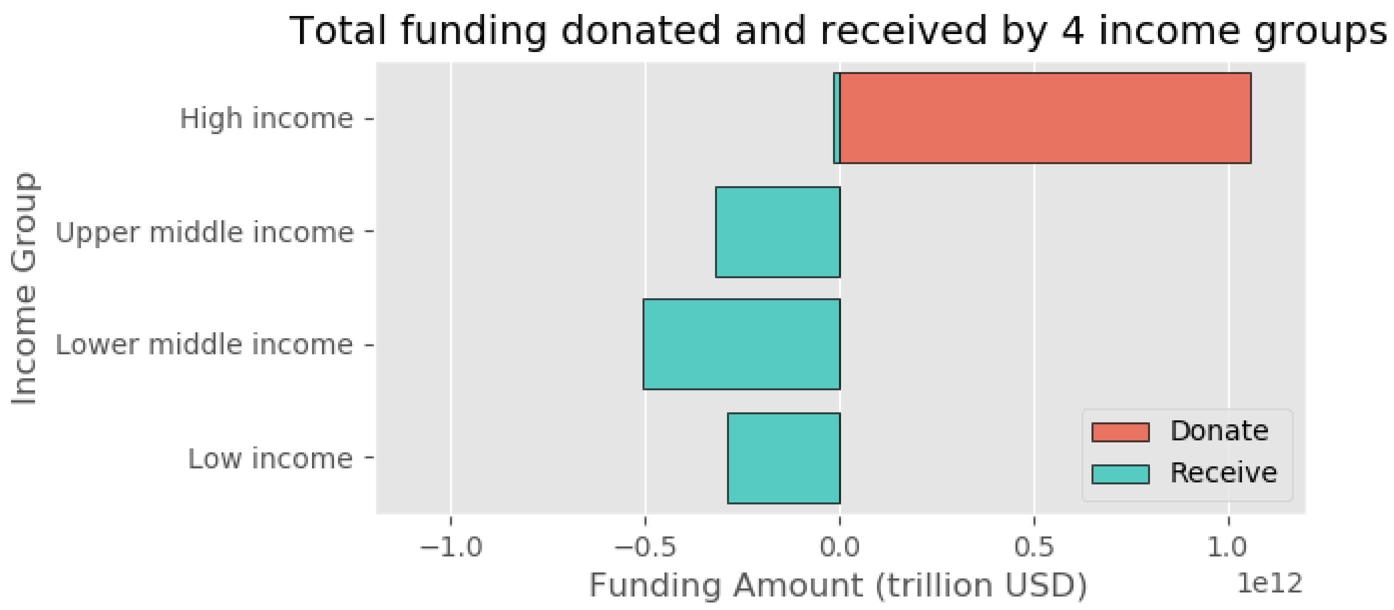
2G. How much ODA does each income group contribute and receive in total?

In order to test the hypothesis that a country's income determine how much to either contribute or receive, we have performed the following analysis. By creating a dataframe that focuses on the different income groups, we are able to plot the following graphs. We make a distinction between the total amount and the average amount.

In [23]:

```
grp = ['High income','Upper middle income','Lower middle income','Low income']

# Create dataframe for donated funding grouped by 4 income groups
income_donate = ODA_data1.groupby('Income
Group').sdg_commitment_amount_usd_constant.sum()
income_donate = income_donate.reindex(index = grp)
income_df1 = pd.DataFrame(columns=['Category', 'Mode','Value'])
income_df1['Category'] = income_donate.keys()
income_df1['Mode'] = ['Donate']*len(income_donate)
income_df1['Value'] = income_donate.values

# Create dataframe for received funding grouped by 4 income groups
income_receive = ODA_data2.groupby('Income
Group').sdg_commitment_amount_usd_constant.sum()
income_receive = income_receive.reindex(index = grp)
income_df2 = pd.DataFrame(columns=['Category', 'Mode','Value'])
```

```
income_df2['Category'] = income_receive.keys()
income_df2['Mode'] = ['Receive']*len(income_receive)
income_df2['Value'] = -income_receive.values

# Concatenate 2 dataframe
income_df = pd.concat([income_df1,income_df2], ignore_index=True, axis=0, sort = F
alse)

# Draw Plot
plt.figure(figsize=(6,3), dpi=100)
group_col = 'Mode'
colors = ['tomato','turquoise']

for c, group in zip(colors, income_df[group_col].unique()):
    sns.barplot(x='Value', y='Category', data=income_df.loc[income_df[group_col]==
group, :],
                color=c, label=group, edgecolor='black')

# Decorations
plt.xlabel("Funding Amount (trillion USD)")
plt.ylabel("Income Group")
plt.yticks(fontsize=10)
plt.xlim(-1.2e12,1.2e12)
plt.title("Total funding donated and received by 4 income groups", fontsize=14)
plt.legend(loc='lower right')
plt.show()
```



Total funding donated and received by 4 income groups
The graph makes it very clear that countries with a high income are responsible for all the donations. Surprisingly, countries with an upper middle income donate nothing and receive more than country with a low income. However, since this is the total amount of funding, it can be misleading. Therefore we shall look into the average amount of funding.

2H. How much does each country donate and receive in average?

In [24]:

```
# Show the number of countries in each income group:
country_data['Income Group'].value_counts()
```

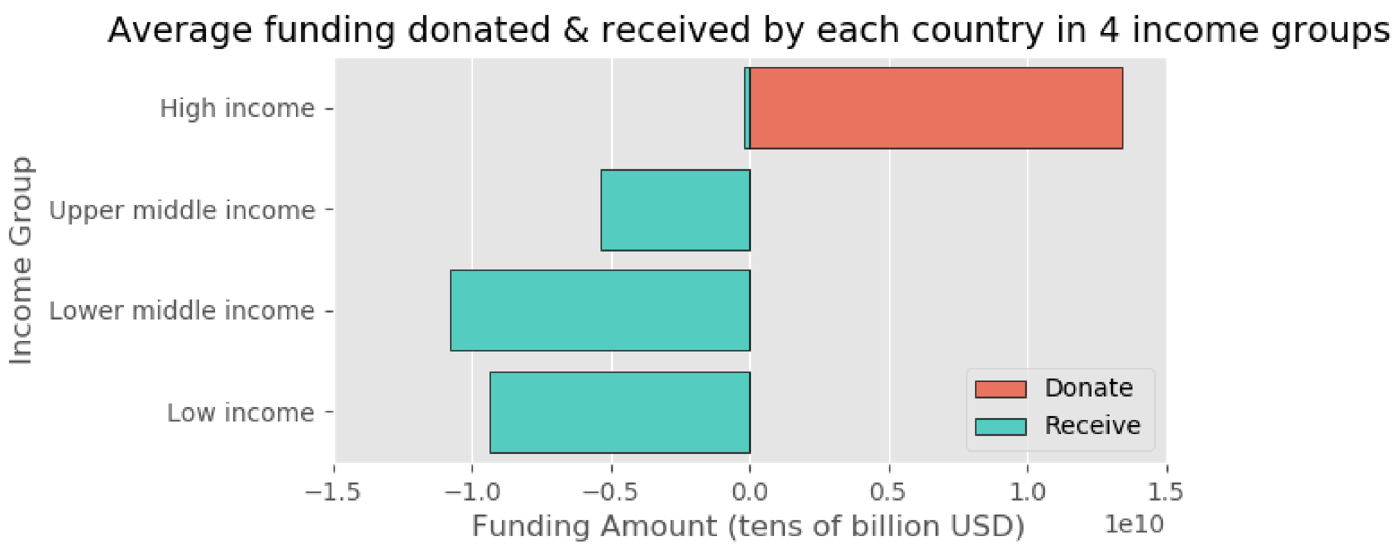
Out[24]:

```
High income            79
```

```
Upper middle income     60
Lower middle income     47
Low income              31
Name: Income Group, dtype: int64
```

In [25]:

```
# Create dataframe for donated funding grouped by 4 income groups
income_df1 = pd.DataFrame(columns=['Category', 'Mode','Value'])
income_df1['Category'] = income_donate.keys()
income_df1['Mode'] = ['Donate']*4
for i in range(0,4):
    income_df1.iloc[i,2] = income_donate[i]/len(country_data[country_data['Income 
Group']==grp[i]].index)

# Create dataframe for received funding grouped by 4 income groups
income_df2 = pd.DataFrame(columns=['Category', 'Mode','Value'])
income_df2['Category'] = income_receive.keys()
income_df2['Mode'] = ['Receive']*4
for i in range(0,4):
    income_df2.iloc[i,2] = 
-income_receive[i]/len(country_data[country_data['Income Group']==grp[i]].index)

# Concatenate 2 dataframe
income_df = pd.concat([income_df1,income_df2], ignore_index=True, axis=0, sort = F
alse)

# Draw Plot
plt.figure(figsize=(6,3), dpi= 100)
group_col = 'Mode'
colors = ['tomato','turquoise']

for c, group in zip(colors, income_df[group_col].unique()):
    sns.barplot(x='Value', y='Category', data=income_df.loc[income_df[group_col]==
group, :],
                color=c, label=group, edgecolor='black')

# Decorations
plt.xlabel("Funding Amount (tens of billion USD)")
plt.ylabel("Income Group")
plt.yticks(fontsize=10)
plt.xlim(-1.5e10,1.5e10)
plt.title("Average funding donated & received by each country in 4 income groups",
fontsize=14)
plt.legend(loc='lower right')
plt.show()
```



Comparing the average amount of funding with the total amount, we see some differences. On average the low income countries receive more than the upper middle income countries, which is not the case when looking at the total amount of funding. What is similar though, is that the high income countries are responsible for all of the funding and that the lower middle income countries receive the most.

Conclusion of relation between a country's income and its ODA
Looking at the data from the two graphs, it can be concluded that the amount a country spends on ODA, or receives it, depends on its income (GDP). Countries with a high income are responsible for the total amount of ODA. Furthermore, on average as well as in total, lower middle income countries receive the most ODA. To further explore the pattern, we will analyze the relationship between GDP and ODA.

- **For High income group: GDP (or GDP/capita) vs Donation**
- **For Middle & Low income groups: GDP (or GDP/capita) vs Reception**

2I. For High Income Group: How does their GDP relate to their donation?

In order to further examine what is proven in the previous research question, we shall look into how a country's GDP and GDP/capita affects its spending on ODA. We shall also use the Kendall tau coefficient to measure the ordinal association between two measured quantities. This way, we can prove if the GDP and ODA spending are statistically dependent.

In [26]:

```
# Read the SDGData
indicator_data = pd.read_csv('SDGData.csv')
indicator_data[:2]

# Filter only GDP (NY.GDP.MKTP.CD) and GDP/capita indicator (NY.GDP.PCAP.CD)
# We will study the data in year 2013 (because the latest data of ODA is in year 2013)
GDP_data1 = indicator_data[indicator_data['Indicator Code'] == 'NY.GDP.MKTP.CD']
GDP_data1 = GDP_data1[['Country Name','Country Code','2013']]
GDP_data1 = GDP_data1.rename(columns={'2013':'GDP'})

GDP_data2 = indicator_data[indicator_data['Indicator Code'] == 'NY.GDP.PCAP.CD']
GDP_data2 = GDP_data2[['Country Code','2013']]
GDP_data2 = GDP_data2.rename(columns={'2013':'GDPpc'})

GDP = pd.merge(GDP_data1, GDP_data2, on="Country Code", how="outer" )
GDP[:2]
```

Out[26]:

	Country Name	Country Code	GDP	GDPpc
0	Arab World	ARB	2.867265e+12	7551.282834
1	Caribbean small states	CSS	7.187461e+10	10090.027335

In [27]:

```
# Merge with ODA donor dataset 'ODA_data1'
ODA_dt1 = ODA_data1[ (ODA_data1['year']==2013) & (ODA_data1['Income Group']=='High income') ]
ODA_country_code1 = ODA_dt1.groupby('Country
```

```
Code').sdg_commitment_amount_usd_constant.sum()
GDP_donate_final = pd.merge(GDP, ODA_country_code1, on="Country Code", how="inner"
)

GDP_donate_final =
GDP_donate_final.rename(columns={'sdg_commitment_amount_usd_constant':'Value'})
GDP_donate_final = GDP_donate_final.fillna(GDP_donate_final.mean()) # Replace NaN
values (if any) with mean
GDP_donate_final[:3]
```

Out[27]:

	Country Name	Country Code	GDP	GDPpc	Value
0	Australia	AUS	1.576184e+12	68150.107041	2.802854e+09
1	Austria	AUT	4.300687e+11	50716.708706	4.406200e+08
2	Belgium	BEL	5.209255e+11	46680.389823	8.450604e+08

In [28]:

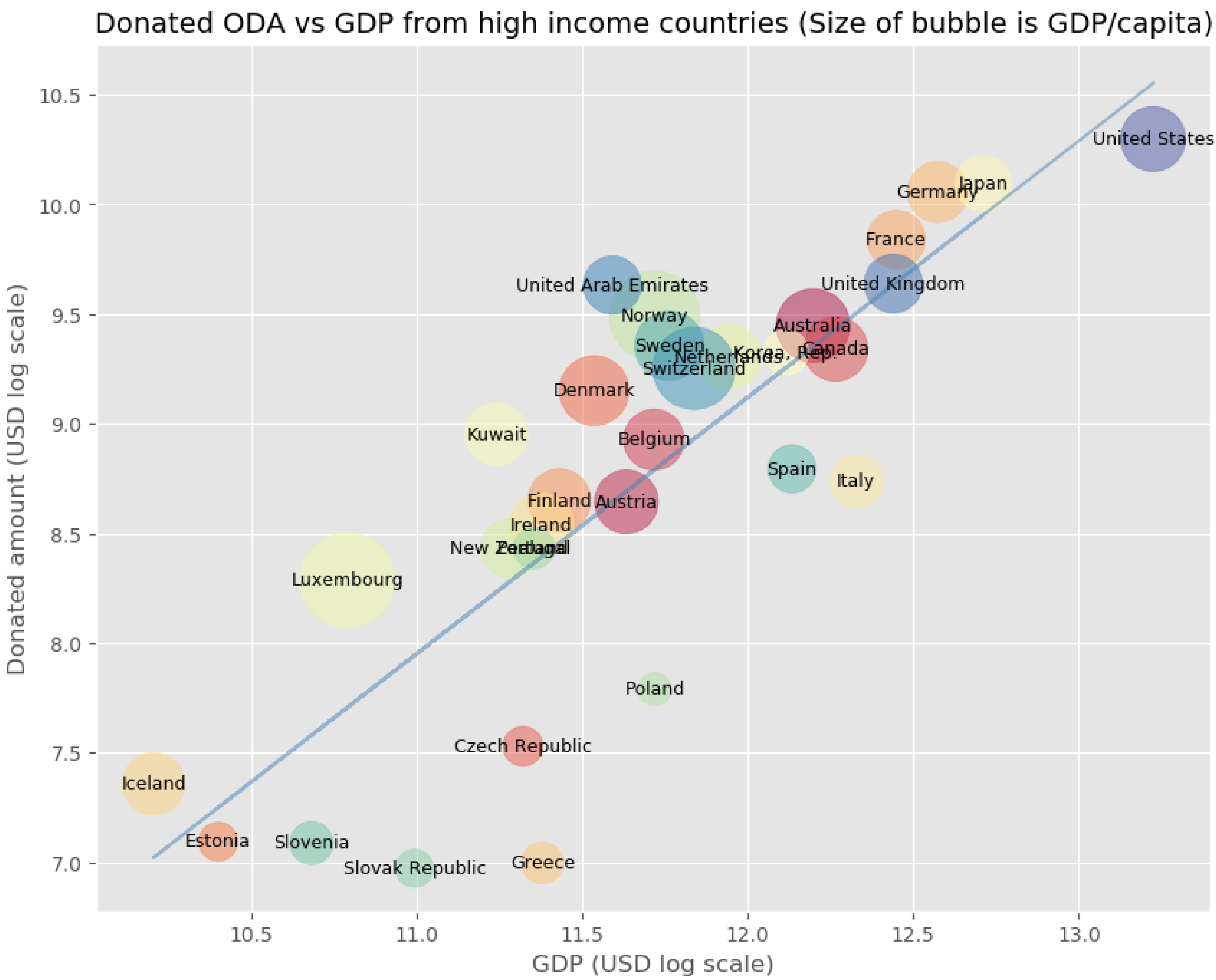
```
# Check whether our data follows normal distribution:
for i in range(2,5):
    W,p = sp.shapiro(GDP_donate_final['GDP'])
    if p < 0.05:
        print(str(GDP_donate_final.columns.values[i]) + ' is NOT normally
distributed')
    else:
        print(str(GDP_donate_final.columns.values[i]) + ' is normally distributed'
)
```

```
GDP is NOT normally distributed
GDPpc is NOT normally distributed
Value is NOT normally distributed
```

In [29]:

```
# Plot the scatter plot (Donation vs GDP) with bubble size equal to GDP/capita
area = GDP_donate_final['GDPpc']/50
x = np.log10(GDP_donate_final['GDP'])
y = np.log10(GDP_donate_final['Value'])
plt.figure(figsize=(10,8), dpi= 100)
plt.scatter(x, y, s=area, c=sns.color_palette("Spectral",len(GDP_donate_final)),
alpha=0.5)

# Decoration
plt.ylabel("Donated amount (USD log scale)")
plt.xlabel("GDP (USD log scale)")
plt.title("Donated ODA vs GDP from high income countries (Size of bubble is GDP/ca
pita)", fontsize=14)

# Show country code over the bubble
x_coords = x
y_coords = y
for i,type in enumerate(GDP_donate_final['Country Name']):
    horizontal = x_coords[i]
    vertical = y_coords[i]
    plt.text(horizontal, vertical, type, fontsize=9, horizontalalignment='center',
             verticalalignment='center')

# Draw the best-fit line:
plt.plot(x, np.poly1d(np.polyfit(x, y, 1))(x), color='steelblue', alpha=.5)
```

```
plt.show()

# Calculate tau (correlation coefficient) and p-value (significant level):
tau, p_value = sp.kendalltau(x, y) # We use Kendall method to calculate
correlation
                        # because the data is not normally distributed (cannot use
Spearson method).
print('Kendall tau is %f' % tau)
print('p-value is %f' % p_value)
```




```
Kendall tau is 0.655914
p-value is 0.000000
```

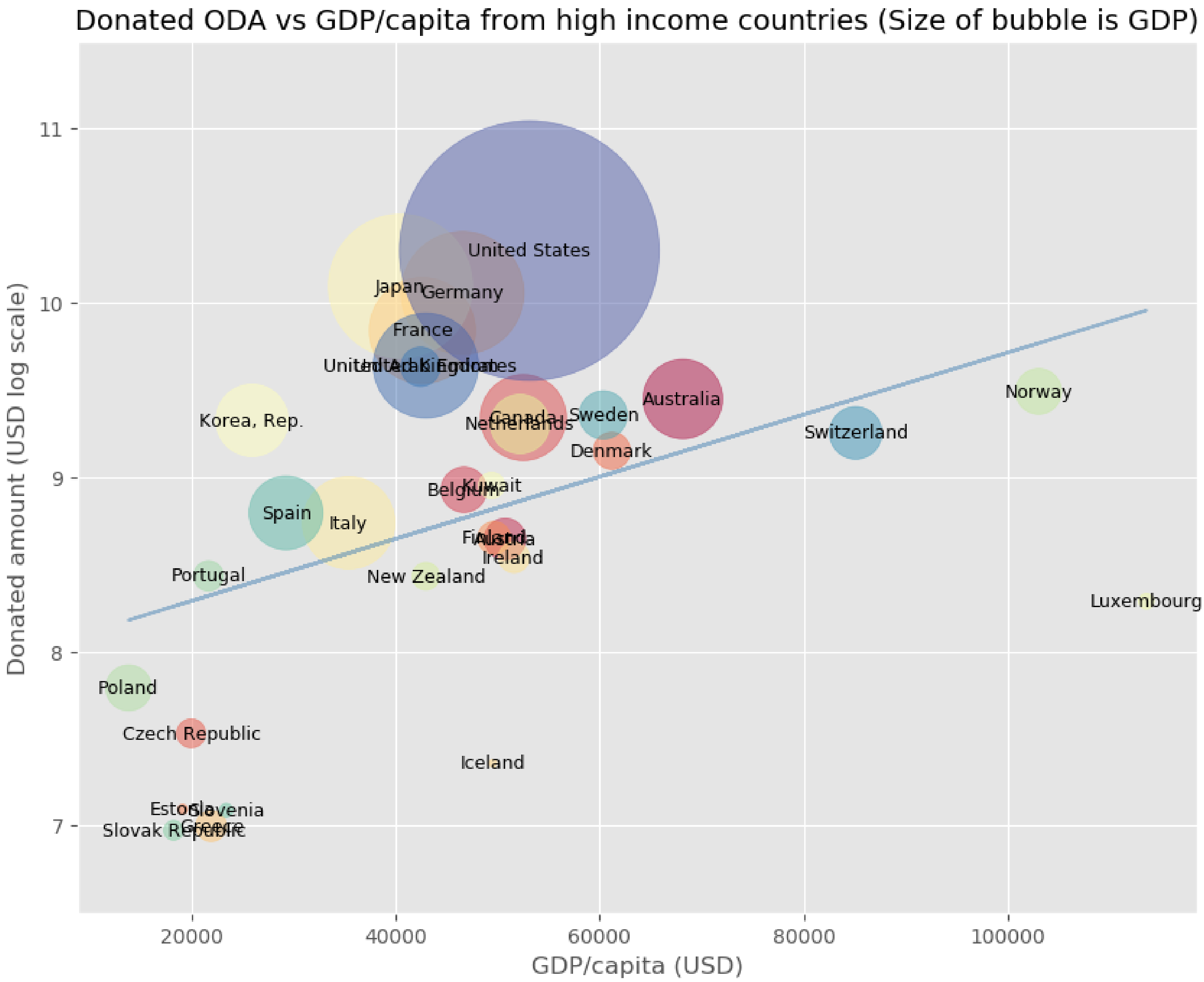
Donated ODA vs GDP from high income countries

From the data, we can conclude that the GDP is not normally distributed. However, the Kendall Tau number is 0.655914, meaning the GDP and ODA are statistically dependent.

In [30]:

```
# Plot the scatter plot (Donation vs GDP) with bubble size equal to GDP/capita
area = GDP_donate_final['GDP']/1e9
x = (GDP_donate_final['GDPpc'])
y = np.log10(GDP_donate_final['Value'])
plt.figure(figsize=(10,8), dpi= 100)
plt.scatter(x, y, s=area, c=sns.color_palette("Spectral",len(GDP_donate_final)),
alpha=0.5)

# Decoration
plt.ylabel("Donated amount (USD log scale)")
```

```
plt.xlabel("GDP/capita (USD)")
plt.title("Donated ODA vs GDP/capita from high income countries (Size of bubble is 
GDP)", fontsize=14)
plt.ylim(6.5,11.5)

# Show country code over the bubble
x_coords = x
y_coords = y
for i,type in enumerate(GDP_donate_final['Country Name']):
    horizontal = x_coords[i]
    vertical = y_coords[i]
    plt.text(horizontal, vertical, type, fontsize=9, horizontalalignment='center',
          verticalalignment='center')

# Draw the best-fit line:
plt.plot(x, np.poly1d(np.polyfit(x, y, 1))(x), color='steelblue', alpha=.5)
plt.show()

# Calculate tau (correlation coefficient) and p-value (significant level):
tau, p_value = sp.kendalltau(x, y) # We use Kendall method to calculate 
correlation
                        # because the data is not normally distributed (cannot use 
Spearson method).
print('Kendall tau is %f' % tau)
print('p-value is %f' % p_value)
```




```
Kendall tau is 0.316129
p-value is 0.012213
```

Donated ODA vs GDP/capita from high income countries

Donated ODA vs GDP/capita from high income countries

When looking at how the ODA relates to the GDP/capita we see that there is a difference with the relation between the ODA and GDP. For the per capita relation, the Kendall Tau coefficient is actually smaller, being 0.316129.

Conclusion for high income countries
We can conclude that both the GDP and the GDP/capita are statistically dependant with the ODA. However, for the GDP the relation is much stronger.

2J. For Upper middle, Lower middle and Low income groups: How does their GDP relate to their received ODA?

Having looked at the High income groups, which is previously shown as “donating” countries, we now turn our gaze to “receiving” countries, and how their GDP relates to how much ODA they receive. We follow the same methodology to create an ordinal association between countries.

In [31]:

```
# Merge with ODA recipient dataset 'ODA_data2'
ODA_dt2 = ODA_data2[ (ODA_data2['year']==2013) &
                     (ODA_data2['Income Group'].isin(['Upper middle income','Lower middle income','Low income'])) ]
ODA_country_code2 = ODA_dt2.groupby('Country Code').sdg_commitment_amount_usd_constant.sum()
GDP_receive_final = pd.merge(GDP, ODA_country_code2, on="Country Code", how="inner" )

GDP_receive_final = GDP_receive_final.rename(columns={'sdg_commitment_amount_usd_constant':'Value'})
GDP_receive_final = GDP_receive_final.fillna(GDP_receive_final.mean()) # Replace NaN values (if any) with mean
GDP_receive_final[:3]
```

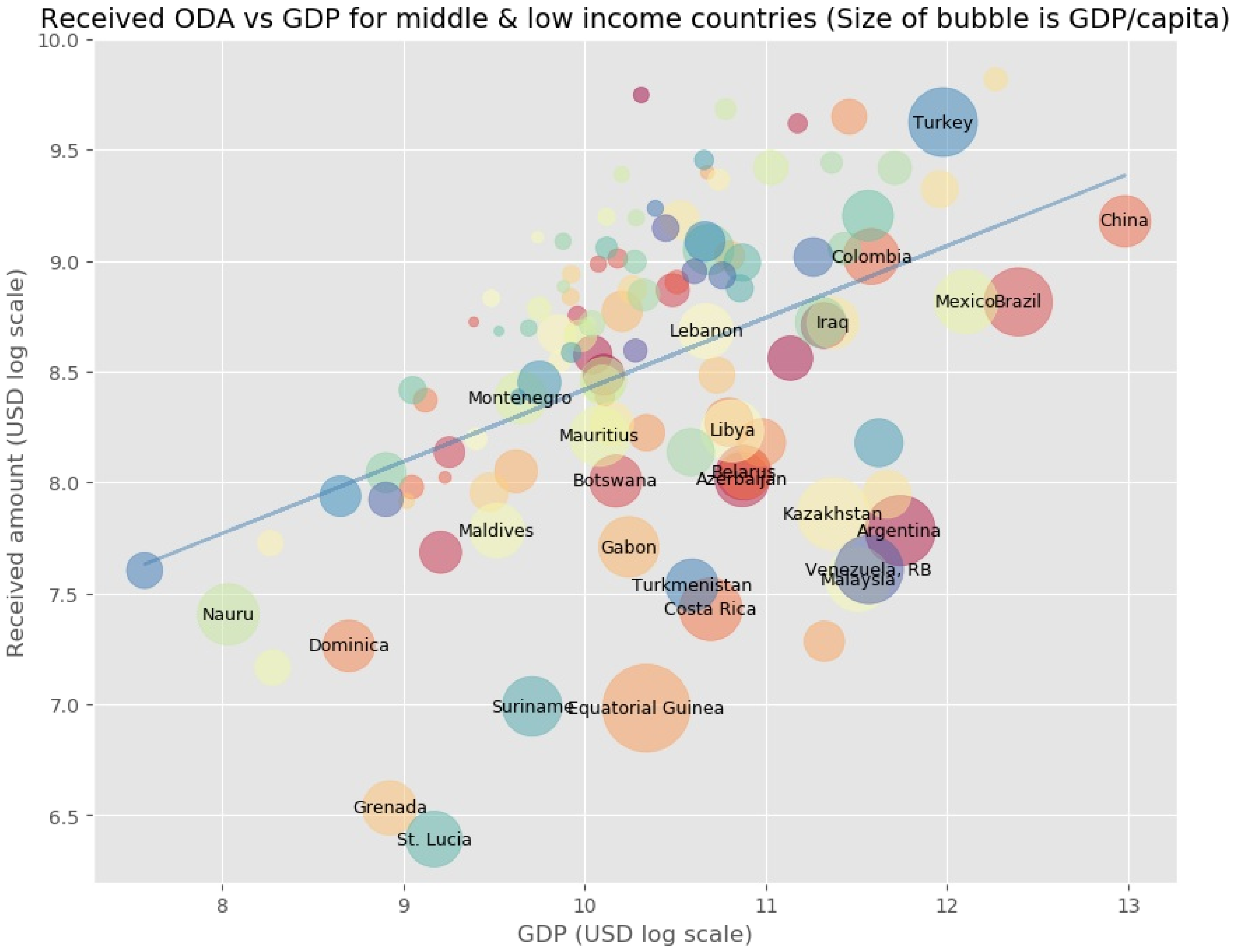
Out[31]:

	Country Name	Country Code	GDP	GDPpc	Value
0	Afghanistan	AFG	2.056105e+10	637.165044	5.581974e+09
1	Albania	ALB	1.277628e+10	4413.082887	3.057200e+08
2	Algeria	DZA	2.097550e+11	5499.587764	5.097180e+08

In [32]:

```
# Plot the scatter plot (Received ODA vs GDP) with bubble size equal to GDP/capita
area = GDP_receive_final['GDPpc']/10
x = np.log10(GDP_receive_final['GDP'])
y = np.log10(GDP_receive_final['Value'])
plt.figure(figsize=(10,8), dpi= 100)
plt.scatter(x, y, s=area, c=sns.color_palette("Spectral",len(GDP_receive_final)), alpha=0.5)

# Decoration
plt.ylabel("Received amount (USD log scale)")
plt.xlabel("GDP (USD log scale)")
plt.title("Received ODA vs GDP for middle & low income countries (Size of bubble is GDP/capita)",
          fontsize=14)
```

```
# Show country name over the bubble
x_coords = x
y_coords = y
for i,type in enumerate(GDP_receive_final['Country Name']):
    horizontal = x_coords[i]
    vertical = y_coords[i]
    if GDP_receive_final.iloc[i,3] > 7000:     # Only show country name for those w
ith GDP/capita > 7000
        plt.text(horizontal, vertical, type, fontsize=9, horizontalalignment='cente
r',
            verticalalignment='center')

# Draw the best-fit line:
plt.plot(x, np.poly1d(np.polyfit(x, y, 1))(x), color='steelblue', alpha=.5)
plt.show()

# Calculate tau (correlation coefficient) and p-value (significant level):
tau, p_value = sp.kendalltau(x, y) # We use Kendall method to calculate
correlation
                        # because the data is not normally distributed (cannot use
Spearson method).
print('Kendall tau is %f' % tau)
print('p-value is %f' % p_value)
```

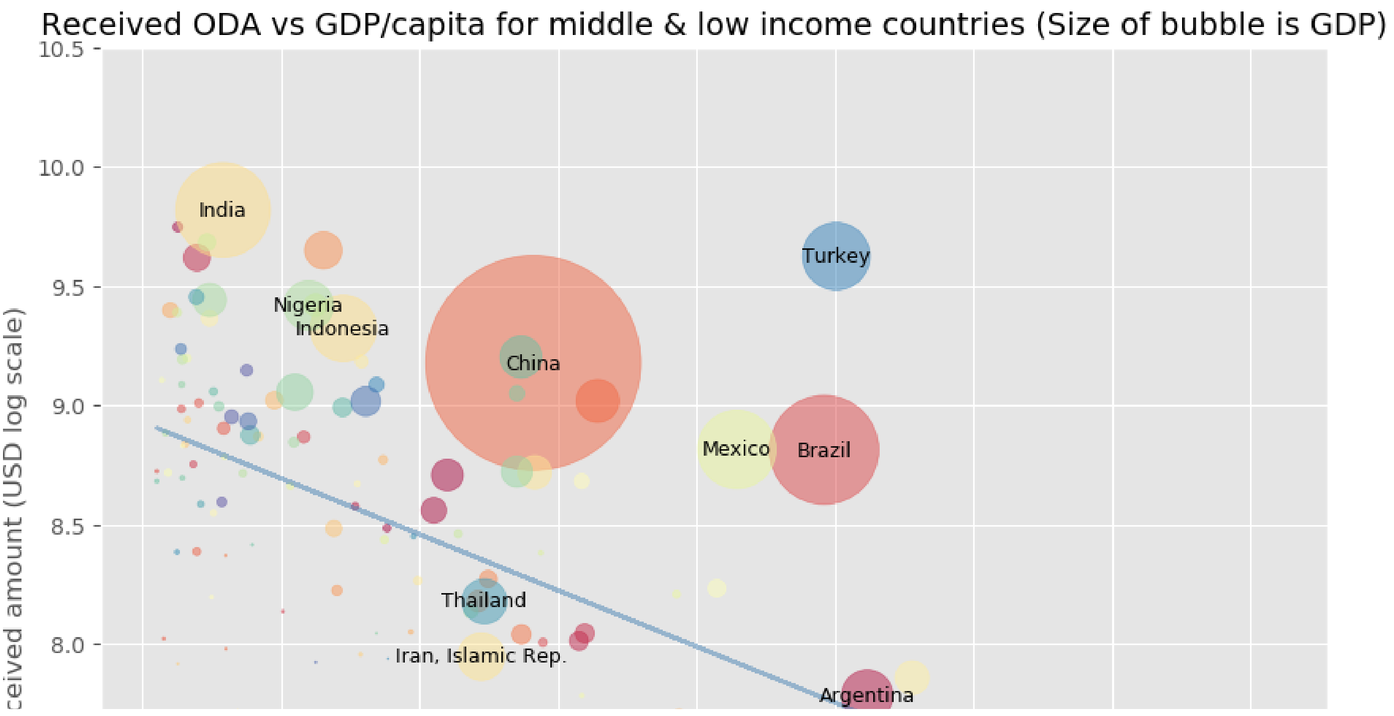


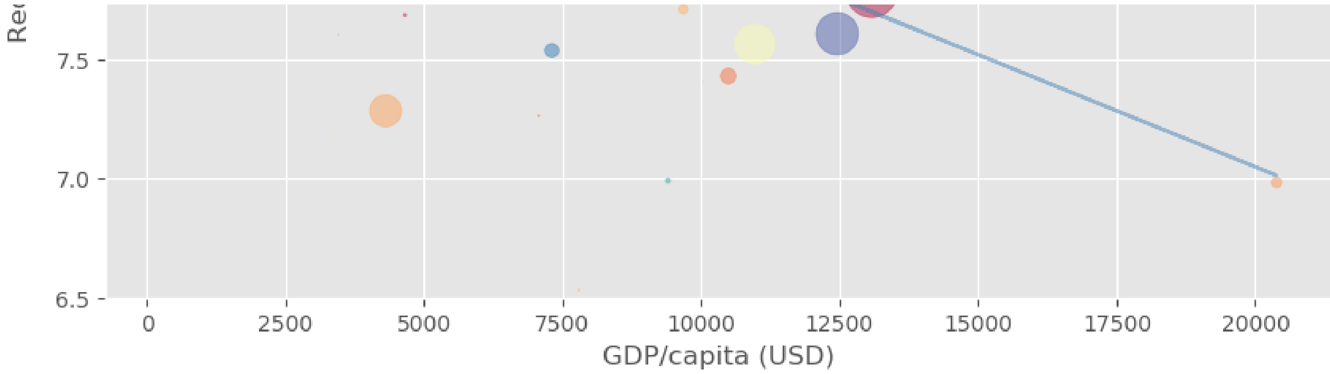

```
Kendall tau is 0.303202
p-value is 0.000001
```

Received ODA vs GDP for middle & low income countries
We analyzed the relationship between received ODA and GDP. Interestingly, we see a clear positive correlation, exemplified by a best fit line with positive slope. However, we can see that there is a lot of variation among the countries. There are clear outliers, both above and beneath the average.

In [33]:

```
# Plot the scatter plot (Received ODA vs GDP) with bubble size equal to GDP/capita
area = GDP_receive_final['GDP']/1e9
x = (GDP_receive_final['GDPpc'])
y = np.log10(GDP_receive_final['Value'])
plt.figure(figsize=(10,8), dpi= 100)
plt.scatter(x, y, s=area, c=sns.color_palette("Spectral",len(GDP_receive_final)),
alpha=0.5)

# Decoration
plt.ylabel("Received amount (USD log scale)")
plt.xlabel("GDP/capita (USD)")
plt.title("Received ODA vs GDP/capita for middle & low income countries (Size of
bubble is GDP)",
          fontsize=14)
plt.ylim(6.5,10.5)

# Show country name over the bubble
x_coords = x
y_coords = y
for i,type in enumerate(GDP_receive_final['Country Name']):
    horizontal = x_coords[i]
    vertical = y_coords[i]
    if GDP_receive_final.iloc[i,2] > 4e11:     # Only show country name for those w
ith GDP > 400 billion USD
        plt.text(horizontal, vertical, type, fontsize=9, horizontalalignment='cente
r',
            verticalalignment='center')

# Draw the best-fit line:
plt.plot(x, np.poly1d(np.polyfit(x, y, 1))(x), color='steelblue', alpha=.5)
plt.show()

# Calculate tau (correlation coefficient) and p-value (significant level):
tau, p_value = sp.kendalltau(x, y) # We use Kendall method to calculate
correlation
                              # because the data is not normally distributed (cannot use
Spearman method).
print('Kendall tau is %f' % tau)
print('p-value is %f' % p_value)
```






```
Kendall tau is -0.345502
p-value is 0.000000
```

Received ODA vs GDP/capita for middle & low income countries
When exploring the relation between the amount of received ODA and the country's GDP/capita, we see a negative correlation. This result is consistent with the general perception that underdeveloped countries should receive more funding for SDG's.

Conclusion for low & middle income countries
We hypothesized that there is a negative relationship between GDP/capita and amount of ODA received: the lower the country's GDP/capita, the higher the amount of ODA received. This is supported by the dataset.

In contrast, the received ODA amount is positively correlated to a country's total GDP. It can be explained by the fact that a country with higher GDP provides more investment opportunities, thus inviting more ODA projects.

3. Further research

Does the growth of donation correspond to the growth of GDP? Does the growth of HDI correspond to the growth of ODA received by a country? Breakdown of ODA into 17 SDG's for top 5 receivers

4. Conclusion

In this research we have looked at the financing of SDG's. Several patterns have been identified by analysing and visualizing the data:

- The biggest donors were the US, Japan, World Bank and European countries, whereas the biggest receivers were Asian and African countries.
- Sustainable Development Goal 16 "Peace, Justice and Strong Institutions" received the most funding.
- Different donors focused on different goals.
- In general, the donation amount increased gradually from 2000 to 2013.
- In terms of geographical region, Europe & Central Asia donated the most while Sub-Saharan Africa received the most funding.
- Nearly 100% of funding came from High Income group, and flowed to the other 3 groups: Upper middle, Lower middle and Low Income.
- For High Income countries, the donated amount is positively correlated to both GDP and GDP/capita, with GDP showing a stronger correlation.
- For Middle and Low Income countries, the received amount is negatively related to GDP/capita but positively related to GDP.

5. Sources

1. About the Sustainable Development Goals - United Nations Sustainable Development. (n.d.). Retrieved from https://www.un.org/sustainabledevelopment/sustainable-development-goals/.
2. Official development assistance (ODA) - Net ODA - OECD Data. (n.d.). Retrieved from https://data.oecd.org/oda/net-oda.htm.
3. Development Matters. (2018, April 27). Revising ODA in the era of SDGs. Retrieved from https://oecd-development-matters.org/2016/06/14/revising-oda-in-the-era-of-sdgs/.
4. Financing for sustainable development. (n.d.). Retrieved from https://www.aiddata.org/sdg.
5. AidData: AidData Core Research Release, Version 3.1. (2016, October 1). Retrieved from https://www.aiddata.org/data/aiddata-core-research-release-level-1-3-1.
6. DiLorenzo, M., Ghose, S., & Turner, J. (2017, November 6). Estimating Baseline Aid to the Sustainable Development Goals. Retrieved from http://docs.aiddata.org/ad4/pdfs/Estimating_Baseline_Aid_to_the_Sustainable_Development_Goals.
7. Human Development Reports. (n.d.). Retrieved from http://hdr.undp.org/en/data.
8. Sustainable Development Goals to kick in with start of new year | UN News. (n.d.). Retrieved from https://news.un.org/en/story/2015/12/519172-sustainable-development-goals-kick-start-new-year.
9. The World Bank. (2019, October 2019). Sustainable Development Goals. Retrieved from https://datacatalog.worldbank.org/dataset/sustainable-development-goals

6. Appendix A: Sustainable Development Goals

Source: United Nations News, december 2015