```
import pandas as pd

# Load the data
file_path = 'Final Data CA+FA.csv'
data = pd.read_csv(file_path)

# Display the first few rows of the dataframe to understand its
structure
data.head()

                WOS ID First Name                   Country Country
Code  \
0  WOS:000174718100007     Janice                   England
GB
1  WOS:000207062600010   Kyoungho               South Korea
KR
2  WOS:000207451700010  Elizabeth  United States of America
US
3  WOS:000207695900002    Joachim  United States of America
US
4  WOS:000207784200003      Beate                   Germany
DE

   Gender Gender Probability  Publication Year    Author Type
0  female            100.00%              2002  Corresponding
1    male            100.00%              2007  Corresponding
2  female            100.00%              2007  Corresponding
3    male            100.00%              2008  Corresponding
4  female             99.00%              2009  Corresponding

# Correctly splitting the data based on the 'Author Type'
first_authors = data[data['Author Type'] == 'First']
corresponding_authors= data[data['Author Type'] == 'Corresponding']

# Calculate counts for male and female First Authors and Corresponding
Authors
first_authors_male = first_authors[first_authors['Gender'] == 'male']
['Country'].value_counts()
first_authors_female = first_authors[first_authors['Gender'] ==
'female']['Country'].value_counts()

corresponding_authors_male =
corresponding_authors[corresponding_authors['Gender'] == 'male']
['Country'].value_counts()
corresponding_authors_female =
corresponding_authors[corresponding_authors['Gender'] == 'female']
['Country'].value_counts()

# Creating DataFrames with gender counts for top 10 countries by total
authors
```

```python
fa_counts = pd.DataFrame({
    'Male First Authors': first_authors_male,
    'Female First Authors': first_authors_female
}).fillna(0).astype(int)
fa_counts['Total First Authors'] = fa_counts.sum(axis=1)
top_10_fa = fa_counts.sort_values(by='Total First Authors',
ascending=False).head(10)

ca_counts = pd.DataFrame({
    'Male Corresponding Authors': corresponding_authors_male,
    'Female Corresponding Authors': corresponding_authors_female
}).fillna(0).astype(int)
ca_counts['Total Corresponding Authors'] = ca_counts.sum(axis=1)
top_10_ca = ca_counts.sort_values(by='Total Corresponding Authors',
ascending=False).head(10)

top_10_fa, top_10_ca
```

```
(                          Male First Authors  Female First Authors  \
 China                                   1195                   894   
 United States of America                1214                   860   
 Germany                                  362                   278   
 Japan                                    365                   134   
 England                                  222                   176   
 Italy                                    136                   202   
 Spain                                    107                   198   
 Canada                                   160                   108   
 South Korea                              140                   102   
 Sweden                                   118                   119   

                           Total First Authors  
 China                                    2089  
 United States of America                 2074  
 Germany                                   640  
 Japan                                     499  
 England                                   398  
 Italy                                     338  
 Spain                                     305  
 Canada                                    268  
 South Korea                               242  
 Sweden                                    237  ,
                           Male Corresponding Authors  \
 China                                           1717   
 United States of America                        1649   
 Germany                                          499   
 Japan                                            443   
 England                                          268   
 Italy                                            178   
 Canada                                           236   
 Spain                                            170   
 South Korea                                              223
 Australia                                                188

                           Female Corresponding Authors  \
 China                                                849
 United States of America                             724
 Germany                                              206
 Japan                                                 72
 England                                              113
 Italy                                                173
 Canada                                                83
 Spain                                                147
 South Korea                                           81
 Australia                                             75

                           Total Corresponding Authors
 China                                               2566
 United States of America                            2373
 Germany                                              705
 Japan                                                515
 England                                              381
 Italy                                                351
 Canada                                               319
 Spain                                                317
 South Korea                                          304
 Australia                                            263  )
```

```python
# Calculate top countries for both First Authors and Corresponding
Authors separately
male_first_authors = first_authors[first_authors['Gender'] == 'male']
['Country'].value_counts()
female_first_authors = first_authors[first_authors['Gender'] ==
'female']['Country'].value_counts()

male_corresponding_authors =
corresponding_authors[corresponding_authors['Gender'] == 'male']
['Country'].value_counts()
female_corresponding_authors =
corresponding_authors[corresponding_authors['Gender'] == 'female']
['Country'].value_counts()

# Create separate DataFrames for easier plotting for First Authors
(FA)
fa_top_countries = set(male_first_authors.index.tolist() +
female_first_authors.index.tolist())
fa_top_countries_list = sorted(list(fa_top_countries))

fa_counts = pd.DataFrame(index=fa_top_countries_list, columns=['Male
First Authors', 'Female First Authors'])
for country in fa_top_countries_list:
    fa_counts.loc[country, 'Male First Authors'] =
male_first_authors.get(country, 0)
    fa_counts.loc[country, 'Female First Authors'] =
female_first_authors.get(country, 0)

# Create separate DataFrames for easier plotting for Corresponding
Authors (CA)
ca_top_countries = set(male_corresponding_authors.index.tolist() +
female_corresponding_authors.index.tolist())
ca_top_countries_list = sorted(list(ca_top_countries))

ca_counts = pd.DataFrame(index=ca_top_countries_list, columns=['Male
Corresponding Authors', 'Female Corresponding Authors'])
for country in ca_top_countries_list:
    ca_counts.loc[country, 'Male Corresponding Authors'] =
male_corresponding_authors.get(country, 0)
    ca_counts.loc[country, 'Female Corresponding Authors'] =
female_corresponding_authors.get(country, 0)

# Combine and clean the DataFrames
fa_counts.fillna(0, inplace=True)
ca_counts.fillna(0, inplace=True)
fa_counts = fa_counts.astype(int)
ca_counts = ca_counts.astype(int)

fa_counts, ca_counts
```

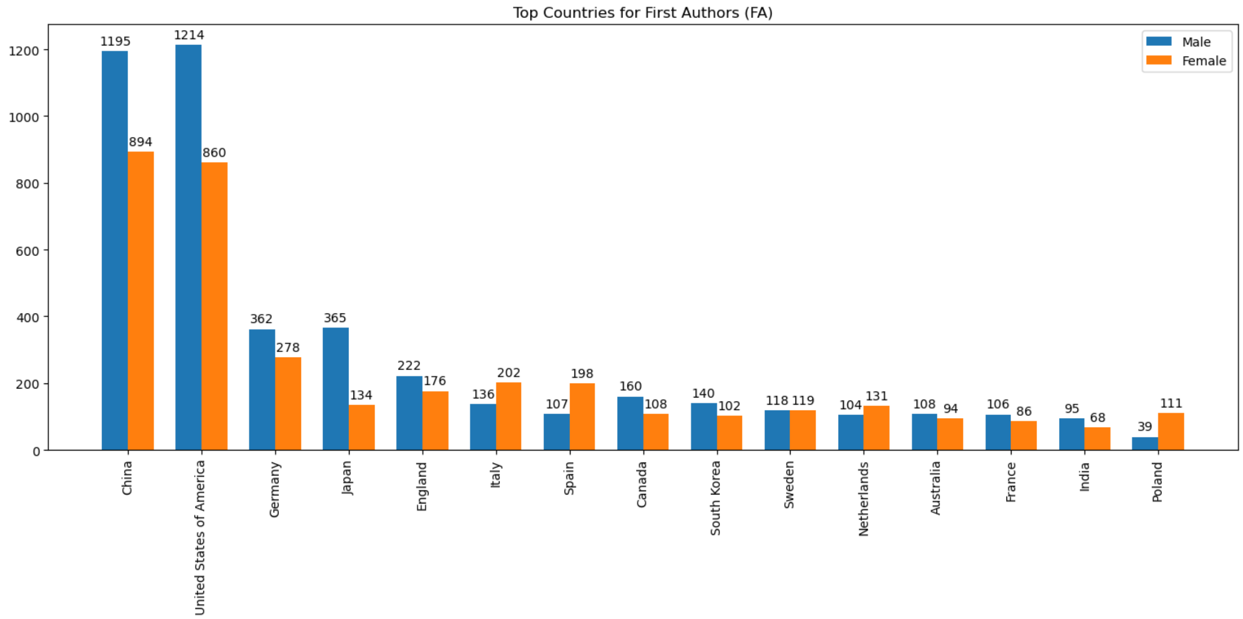
```
(                          Male First Authors  Female First Authors
 Algeria                                    2                     2
 Argentina                                  4                     8
 Armenia                                    0                     1
 Australia                                108                    94
 Austria                                   44                    22
 ...                                      ...                   ...
 United States of America                1214                   860
 Uruguay                                    3                     1
 Venezuela                                  0                     1
 Vietnam                                    0                     1
 Zimbabwe                                   0                     1

 [85 rows x 2 columns],
            Male Corresponding Authors  Female Corresponding Authors
 Argentina                           3                            10
 Australia                         188                            75
 Austria                            56                            18
 BELARUS                             1                             0
 Bahrain                             0                             1
 ...                               ...                           ...
 Uruguay                             1                             2
 Venezuela                           1                             0
 Vietnam                                    0                              1
 Wales                                      5                              2
 Zimbabwe                                   0                              1

 [89 rows x 2 columns])

# Filter the data for male and female authors separately for both
first and corresponding authors
male_first_authors = first_authors[first_authors['Gender'] == 'male']
['Country'].value_counts().head(10)
female_first_authors = first_authors[first_authors['Gender'] ==
'female']['Country'].value_counts().head(10)

male_corresponding_authors =
corresponding_authors[corresponding_authors['Gender'] == 'male']
['Country'].value_counts().head(10)
female_corresponding_authors =
corresponding_authors[corresponding_authors['Gender'] == 'female']
['Country'].value_counts().head(10)

# Create a DataFrame for easier plotting
top_countries = set(male_first_authors.index.tolist() +
female_first_authors.index.tolist() +
                    male_corresponding_authors.index.tolist() +
female_corresponding_authors.index.tolist())

top_countries_list = sorted(list(top_countries))

# Reinitialize the DataFrame with the sorted list of top countries
country_gender_counts = pd.DataFrame(index=top_countries_list,
columns=['Male First Authors', 'Female First Authors',

'Male Corresponding Authors', 'Female Corresponding Authors'])

# Fill the DataFrame with the counts for a more accurate plotting
for country in top_countries_list:
    country_gender_counts.loc[country, 'Male First Authors'] =
male_first_authors.get(country, 0)
    country_gender_counts.loc[country, 'Female First Authors'] =
female_first_authors.get(country, 0)
    country_gender_counts.loc[country, 'Male Corresponding Authors'] =
male_corresponding_authors.get(country, 0)
    country_gender_counts.loc[country, 'Female Corresponding Authors']
= female_corresponding_authors.get(country, 0)

country_gender_counts.fillna(0, inplace=True)  # Ensure there are no
NaN values
country_gender_counts = country_gender_counts.astype(int)  # Convert
counts to integers for plotting
```

```
country_gender_counts

                          Male First Authors  Female First Authors  \
Australia                                108                     0
Canada                                   160                     0
China                                   1195                   894
England                                  222                   176
Germany                                  362                   278
Italy                                    136                   202
Japan                                    365                   134
Netherlands                                0                   131
Poland                                     0                   111
South Korea                              140                     0
Spain                                      0                   198
Sweden                                   118                   119
United States of America                1214                   860

                          Male Corresponding Authors  \
Australia                                        188
Canada                                           236
China                                           1717
England                                          268
Germany                                          499
Italy                                            178
Japan                                            443
Netherlands                                        0
Poland                                             0
South Korea                                      223
Spain                                            170
Sweden                                             0
United States of America                        1649

                          Female Corresponding Authors
Australia                                            0
Canada                                              83
China                                              849
England                                            113
Germany                                            206
Italy                                              173
Japan                                                0
Netherlands                                         92
Poland                                               0
South Korea                                         81
Spain                                              147
Sweden                                             100
United States of America                           724

# calculate counts for male and female authors for both First and
Corresponding Authors for all countries
```

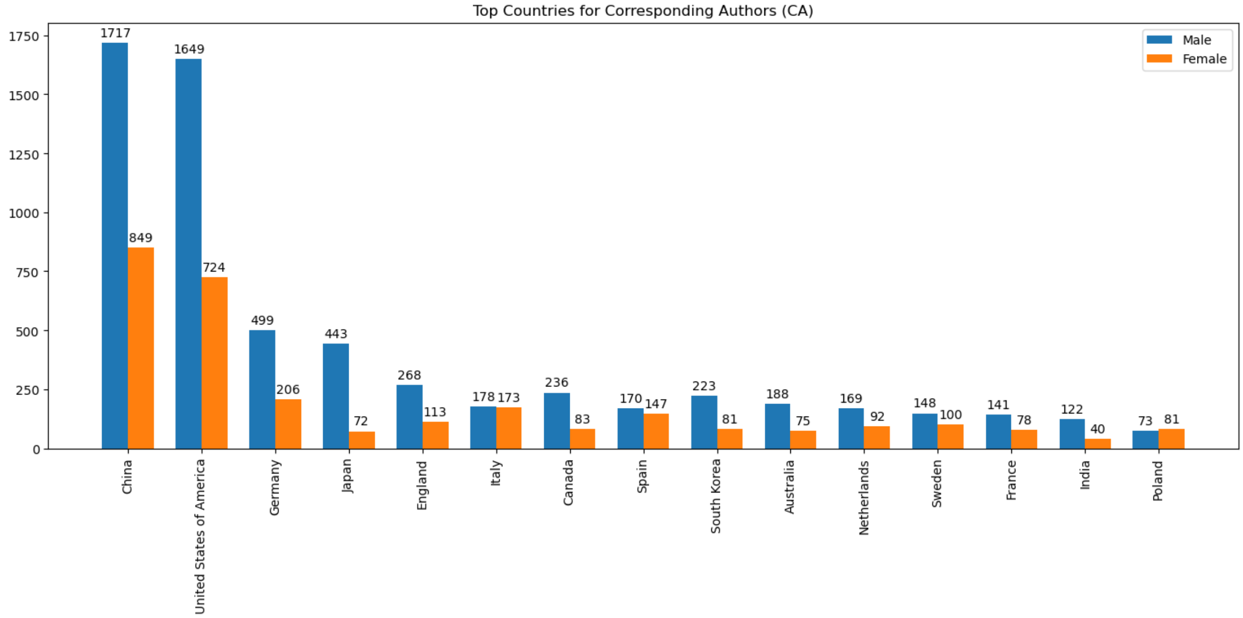
```
male_fa_counts = first_authors[first_authors['Gender'] ==
'male'].groupby('Country').size()
female_fa_counts = first_authors[first_authors['Gender'] ==
'female'].groupby('Country').size()

male_ca_counts = corresponding_authors[corresponding_authors['Gender']
== 'male'].groupby('Country').size()
female_ca_counts =
corresponding_authors[corresponding_authors['Gender'] ==
'female'].groupby('Country').size()

# Combine these counts into a new DataFrame for easier plotting and
analysis
combined_counts = pd.DataFrame({
    'Male FA': male_fa_counts,
    'Female FA': female_fa_counts,
    'Male CA': male_ca_counts,
    'Female CA': female_ca_counts
}).fillna(0).astype(int)  # Fill missing values with 0 and ensure
counts are integers

# Sort this combined data for First Authors and Corresponding Authors
separately in descending order
sorted_fa_combined = combined_counts[['Male FA', 'Female
FA']].sum(axis=1).sort_values(ascending=False).head(15)
sorted_ca_combined = combined_counts[['Male CA', 'Female
CA']].sum(axis=1).sort_values(ascending=False).head(15)

# Now retrieve the detailed counts for top countries for FA and CA
separately
top_fa_countries = combined_counts.loc[sorted_fa_combined.index]
top_ca_countries = combined_counts.loc[sorted_ca_combined.index]

top_fa_countries, top_ca_countries
```

```
(                           Male FA  Female FA  Male CA  Female CA
 Country
 China                         1195        894     1717        849
 United States of America      1214        860     1649        724
 Germany                        362        278      499        206
 Japan                          365        134      443         72
 England                        222        176      268        113
 Italy                          136        202      178        173
 Spain                          107        198      170        147
 Canada                         160        108      236         83
 South Korea                    140        102      223         81
 Sweden                         118        119      148        100
 Netherlands                    104        131      169         92
 Australia                      108         94      188         75
 France                         106         86      141         78
 India                          95         68      122         40
 Poland                         39        111       73         81,
                           Male FA  Female FA  Male CA  Female CA
 Country
 China                        1195        894     1717        849
 United States of America     1214        860     1649        724
 Germany                       362        278      499        206
 Japan                         365        134      443         72
 England                       222        176      268        113
 Italy                         136        202      178        173
 Canada                        160        108      236         83
 Spain                         107        198      170        147
 South Korea                   140        102      223         81
 Australia                     108         94      188         75
 Netherlands                   104        131      169         92
 Sweden                        118        119      148        100
 France                        106         86      141         78
 India                          95         68      122         40
 Poland                         39        111       73         81)
```

```python
import numpy as np
import matplotlib.pyplot as plt

# Adjusted function to plot side-by-side bars without internal value
labels, only on top
def plot_author_type_counts_side_by_side_adjusted(data, title,
columns, ax):
    sorted_data =
data[columns].sum(axis=1).sort_values(ascending=False)
    data_sorted = data.loc[sorted_data.index]

    ind = np.arange(len(data_sorted))  # the x locations for the
groups
    width = 0.35  # width of the bars

    # Plotting the bars side by side with adjusted colors for
colorblind-friendly visualization
    ax.bar(ind - width / 2, data_sorted[columns[0]], width,
color='tab:blue', label='Male')
    ax.bar(ind + width / 2, data_sorted[columns[1]], width,
color='tab:orange', label='Female')

    ax.set_title(title)
    ax.set_xticks(ind)
    ax.set_xticklabels(data_sorted.index, rotation='vertical')
    ax.legend()

# Function to add value labels on top of each bar
def add_value_labels_on_top(ax):
    """Add value labels on top of each bar in the given axis."""
    for bar in ax.patches:
        height = bar.get_height()
        ax.annotate(
            f'{height:.0f}',  # Format the label as an integer
            xy=(bar.get_x() + bar.get_width() / 2, height),
            xytext=(0, 3),  # Offset the text label by 3 units upward
            textcoords='offset points',
            ha='center',
            va='bottom'
        )

# Re-create figure and axes for the subplots
fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(14, 14))

# Plot for First Authors with labels only on top
plot_author_type_counts_side_by_side_adjusted(top_fa_countries, 'Top
Countries for First Authors (FA)', ['Male FA', 'Female FA'], ax1)
add_value_labels_on_top(ax1)

# Plot for Corresponding Authors with labels only on top
plot_author_type_counts_side_by_side_adjusted(top_ca_countries, 'Top
Countries for Corresponding Authors (CA)', ['Male CA', 'Female CA'],
ax2)
add_value_labels_on_top(ax2)

plt.tight_layout()
plt.show()
```



Top Countries for First Authors (FA)



Top Countries for Corresponding Authors (CA)

```
import pandas as pd

# Load the dataset
file_path = 'Final Data CA+FA.csv'
final_data = pd.read_csv(file_path)

# Filter data for only 'Corresponding' author type
final_data_corresponding = final_data[final_data['Author Type'] ==
'Corresponding']

# Get the top 20 countries by publication count in the final dataset
with 'Corresponding' author type
```

```
top_20_final_data_corresponding = final_data_corresponding['Country'].value_counts().head(20).reset_index()
top_20_final_data_corresponding.columns = ['Country', 'Publication Count']

# Display the result
print(top_20_final_data_corresponding)
```

```
                     Country  Publication Count
0                      China               2839
1   United States of America               2512
2                    Germany                721
3                      Japan                539
4                    England                410
5                      Italy                353
6                South Korea                349
7                     Canada                340
8                      Spain                321
9                  Australia                275
10               Netherlands                268
11                    Sweden                259
12                    France                223
13                     India                189
14                    Taiwan                171
15                    Poland                157
16                    Brazil                151
17               Switzerland                131
18                   Denmark                116
19                   Belgium                103
```

```
# Calculate the number of unique publications from each country using the 'WOS ID' column
unique_publications_per_country = data.groupby('Country')['WOS ID'].nunique().sort_values(ascending=False)

unique_publications_per_country
```

```
Country
United States of America    2462
China                       2254
Germany                      713
Japan                        523
England                      465
                            ... 
Liberia                        1
Mali                           1
North Korea                    1
SWITZERLAND                    1
```

```
Zimbabwe                          1
Name: WOS ID, Length: 99, dtype: int64

# Extract the top 10 countries based on the number of unique
publications
top_15_unique_publications_per_country =
unique_publications_per_country.head(15)
top_15_unique_publications_per_country

Country
United States of America    2462
China                       2254
Germany                      713
Japan                        523
England                      465
Italy                        350
Canada                       335
Spain                        317
South Korea                  288
Sweden                       270
Netherlands                  270
Australia                    262
France                       218
India                        185
Taiwan                       154
Name: WOS ID, dtype: int64

import pandas as pd

# Load the data from a CSV file
file_path = 'Final Data CA+FA.csv'

data = pd.read_csv(file_path)

# Split the data based on the 'Author Type'
first_authors = data[data['Author Type'] == 'First']
corresponding_authors = data[data['Author Type'] == 'Corresponding']

# Calculate counts for male and female authors
first_authors_male = first_authors[first_authors['Gender'] == 'male']
['Country'].value_counts()
first_authors_female = first_authors[first_authors['Gender'] ==
'female']['Country'].value_counts()
corresponding_authors_male =
corresponding_authors[corresponding_authors['Gender'] == 'male']
['Country'].value_counts()
corresponding_authors_female =
corresponding_authors[corresponding_authors['Gender'] == 'female']
['Country'].value_counts()
```

```
# Combine the data into DataFrames for easier manipulation
fa_counts = pd.DataFrame({
    'Male First Authors': first_authors_male,
    'Female First Authors': first_authors_female
}).fillna(0).astype(int)

ca_counts = pd.DataFrame({
    'Male Corresponding Authors': corresponding_authors_male,
    'Female Corresponding Authors': corresponding_authors_female
}).fillna(0).astype(int)

# Define top countries including the Netherlands and their specific
order
top_countries = [
    "United States of America", "China", "Germany", "Japan",
"England",
    "Italy", "Canada", "Spain", "South Korea", "Netherlands", "Sweden"
]

# Filter the DataFrames to include only the specified countries
fa_gender_bifurcation = fa_counts.loc[top_countries]
ca_gender_bifurcation = ca_counts.loc[top_countries]

# Output the gender bifurcation tables for First and Corresponding
Authors
print("Gender Bifurcation for First Authors (FA):")
print(fa_gender_bifurcation)
print("\nGender Bifurcation for Corresponding Authors (CA):")
print(ca_gender_bifurcation)
```

```
Gender Bifurcation for First Authors (FA):
                          Male First Authors  Female First Authors
United States of America                1214                   860
China                                   1195                   894
Germany                                  362                   278
Japan                                    365                   134
England                                  222                   176
Italy                                    136                   202
Canada                                   160                   108
Spain                                    107                   198
South Korea                              140                   102
Netherlands                              104                   131
Sweden                                   118                   119

Gender Bifurcation for Corresponding Authors (CA):
                          Male Corresponding Authors  \
United States of America                        1649
China                                           1717
Germany                                          499
Japan                                            443
```

```
England                                          268
Italy                                            178
Canada                                           236
Spain                                            170
South Korea                                      223
Netherlands                                      169
Sweden                                           148
```

```
                          Female Corresponding Authors
United States of America                           724
China                                              849
Germany                                            206
Japan                                               72
England                                            113
Italy                                              173
Canada                                              83
Spain                                              147
South Korea                                         81
Netherlands                                         92
Sweden                                             100
```