In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import dask.dataframe as dd
```

The dataset describes human trafficking in the UK during 2013-2016. I found it from Kaggle (https://www.kaggle.com/algosforgood/uk-human-trafficking-data); it was compiled from UK National Crime Agency's National Referral Mechanism Statistics.

Knowing nothing about human trafficking in the UK, I want to start by answering the following questions:

- In minors and adults, what is the gender proportion of the victims?
- What are the most common human trafficking crimes minors and adults are the victims of?
- Are there any noticeable patterns in gender ratio if we look at specific exploitation types?
- Might human trafficking be on the rise, based on these years?
- What are the most common origin countries for victims of different exploitation types?

First, I'll load in data about how much were different types of exploitation recorded during 2016.

In [26]:

```
pd.options.mode.chained_assignment = None #the warning/error r
aising is unnecessary with the operations here
exploitation_df = pd.read_csv('2016_exploitation_type.csv')
#add params and a function?
```

In [3]:

```
exploitation_df
```

Out[3]:

	Claimed Exploitation Type	Female	Male	Trans-gender	Total 2016	2015 - 2016 % Change
0	Adult - Domestic Servitude	259	67	0	326	-7.60%
1	Adult - Labour Exploitation	182	925	0	1107	23.70%
2	Adult - Organ Harvesting	1	0	0	1	-50.00%
3	Adult - Sexual Exploitation	888	58	5	951	10.20%
4	Adult - Unknown Exploitation	70	72	0	142	-17.00%
5	Minor - Domestic Servitude	67	36	0	103	49.30%
6	Minor - Labour Exploitation	68	400	0	468	62.50%
7	Minor - Sexual Exploitation (non-UK national)	118	29	0	147	31.30%
8	Minor - Sexual Exploitation (UK national)	203	12	0	215	104.80%
9	Minor - Unknown Exploitation	80	265	0	345	-14.80%
10	Total	1936	1864	5	3805	NaN

I want to look at adults and underage victims separately; splitting the data based on whether it mentions "adult" or "minor". I can then shorten the column names, since I already know which type of victim the data is about.

In [4]:

```
child_exploitation_df = exploitation_df[exploitation_df["Claimed Exploitation Type"].str.contains("Minor")]
adult_exploitation_df = exploitation_df[exploitation_df["Claimed Exploitation Type"].str.contains("Adult")]
total_exploitation = exploitation_df[exploitation_df["Claimed Exploitation Type"] == 'Total']
```

In [5]:

```
adult_exploitation_df.loc[:, 'Claimed Exploitation Type'] = adult_exploitation_df.loc[:, 'Claimed Exploitation Type'].str[7:]
child_exploitation_df.loc[:, 'Claimed Exploitation Type'] = child_exploitation_df.loc[:, 'Claimed Exploitation Type'].str[7:]
child_exploitation_df
```

Out[5]:

	Claimed Exploitation Type	Female	Male	Trans-gender	Total 2016	2015 - 2016 % Change
5	Domestic Servitude	67	36	0	103	49.30%
6	Labour Exploitation	68	400	0	468	62.50%
7	Sexual Exploitation (non-UK national)	118	29	0	147	31.30%
8	Sexual Exploitation (UK national)	203	12	0	215	104.80%
9	Unknown Exploitation	80	265	0	345	-14.80%

In [6]:

```
adult_exploitation_df
```

Out[6]:

	Claimed Exploitation Type	Female	Male	Trans-gender	Total 2016	2015 - 2016 % Change
0	Domestic Servitude	259	67	0	326	-7.60%
1	Labour Exploitation	182	925	0	1107	23.70%
2	Organ Harvesting	1	0	0	1	-50.00%
3	Sexual Exploitation	888	58	5	951	10.20%
4	Unknown Exploitation	70	72	0	142	-17.00%

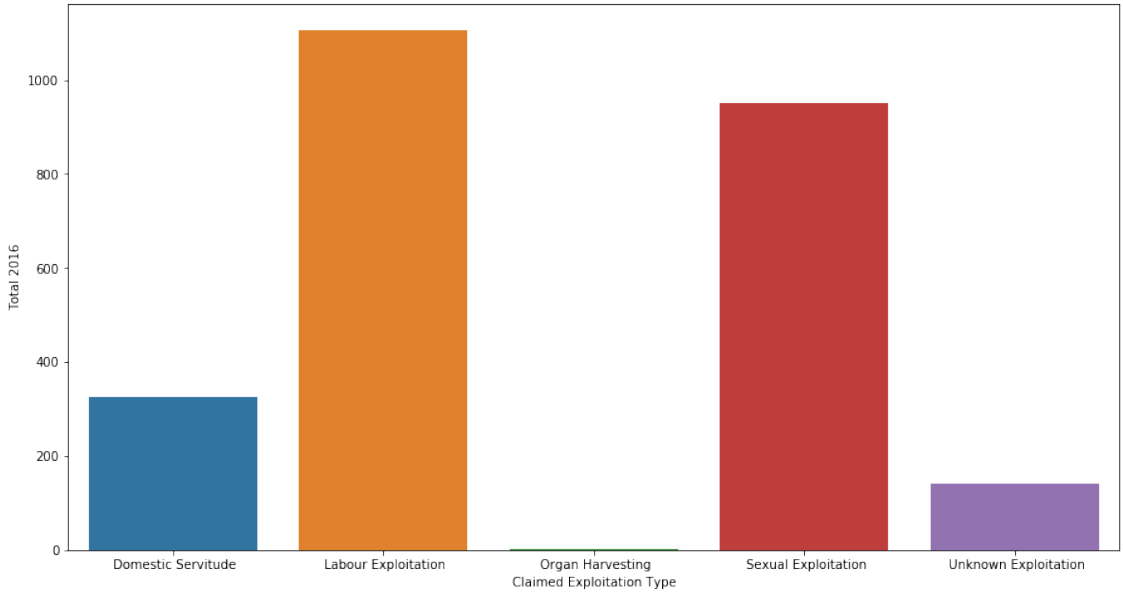
Now I can draw a plot for both adults and minors. First, let's look at exploitation types.

In [7]:

```
plt.figure(figsize=(15,8))
sns.barplot(x='Claimed Exploitation Type', y='Total 2016', dat
a=adult_exploitation_df)
```

Out[7]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c1e06
0898>
```

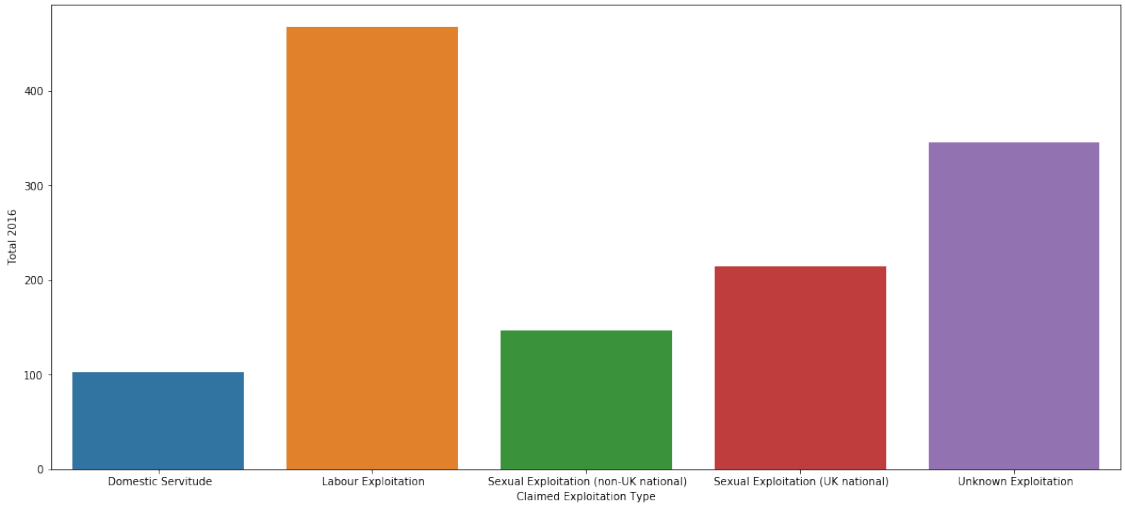


In [8]:

```
plt.figure(figsize=(18,8))
sns.barplot(x='Claimed Exploitation Type', y='Total 2016', dat
a=child_exploitation_df)
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c1e44
c198>
```



Labour exploitation and sexual exploitation are the most common types of exploitation. One adult was known to be a victim of organ harvesting during 2016. For minors, there is a large number of cases classified as Unknown Exploitation, as well. What this label can mean should be investigated.

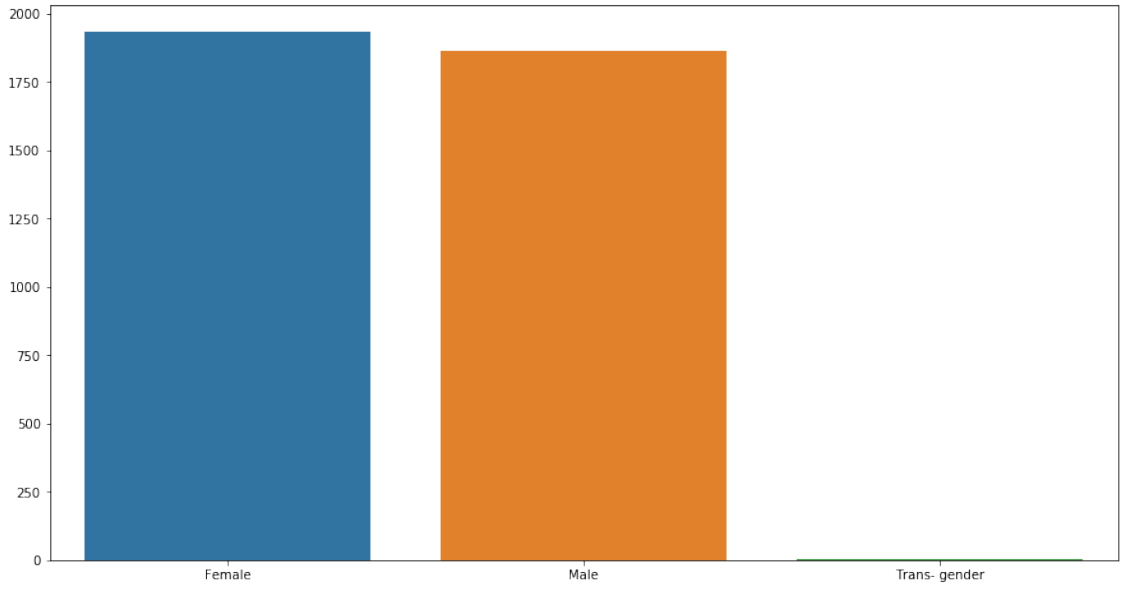
Let's look at gender ratios.

In [9]:

```
plt.figure(figsize=(15,8))
sns.barplot(data=total_exploitation[['Female', 'Male', 'Trans-
gender']])
```

Out[9]:

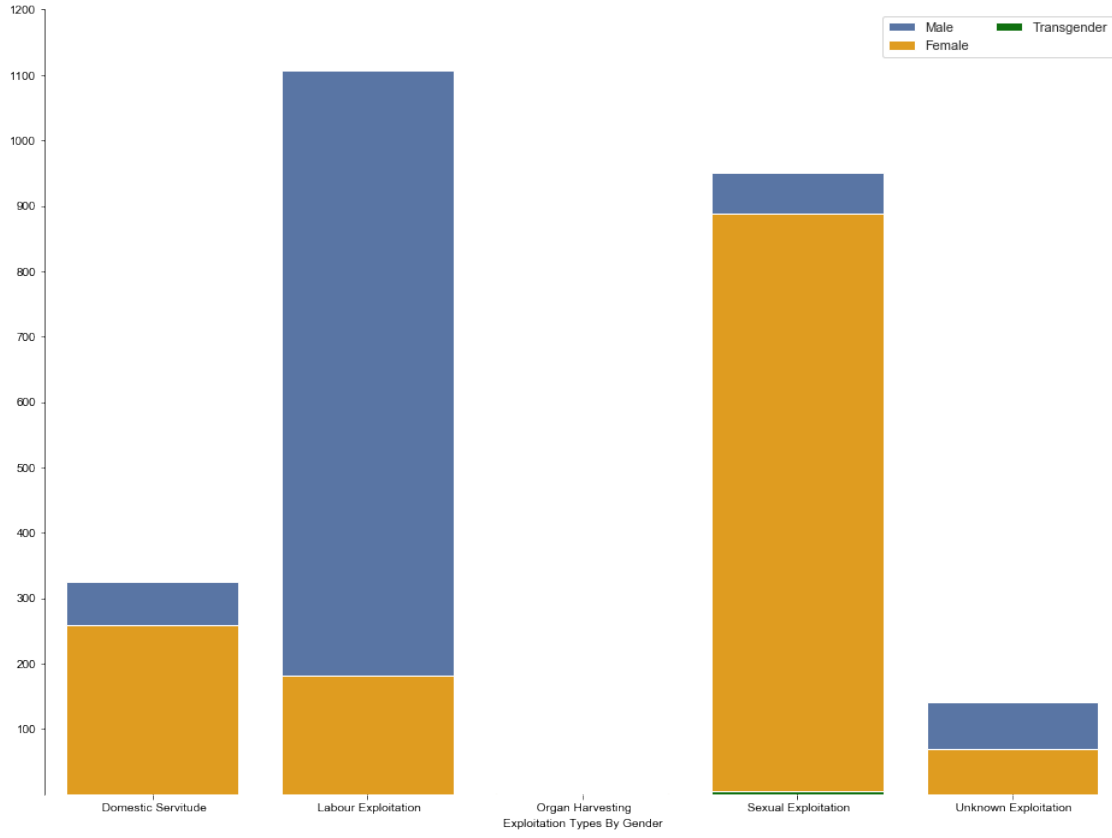
```
<matplotlib.axes._subplots.AxesSubplot at 0x1c1e94
c160>
```



Overall, the gender ratio of victims seems to be quite even. Lets dig a bit deeper and look at gender proportion by exploitation type, as well as minors and adults.

In [12]:

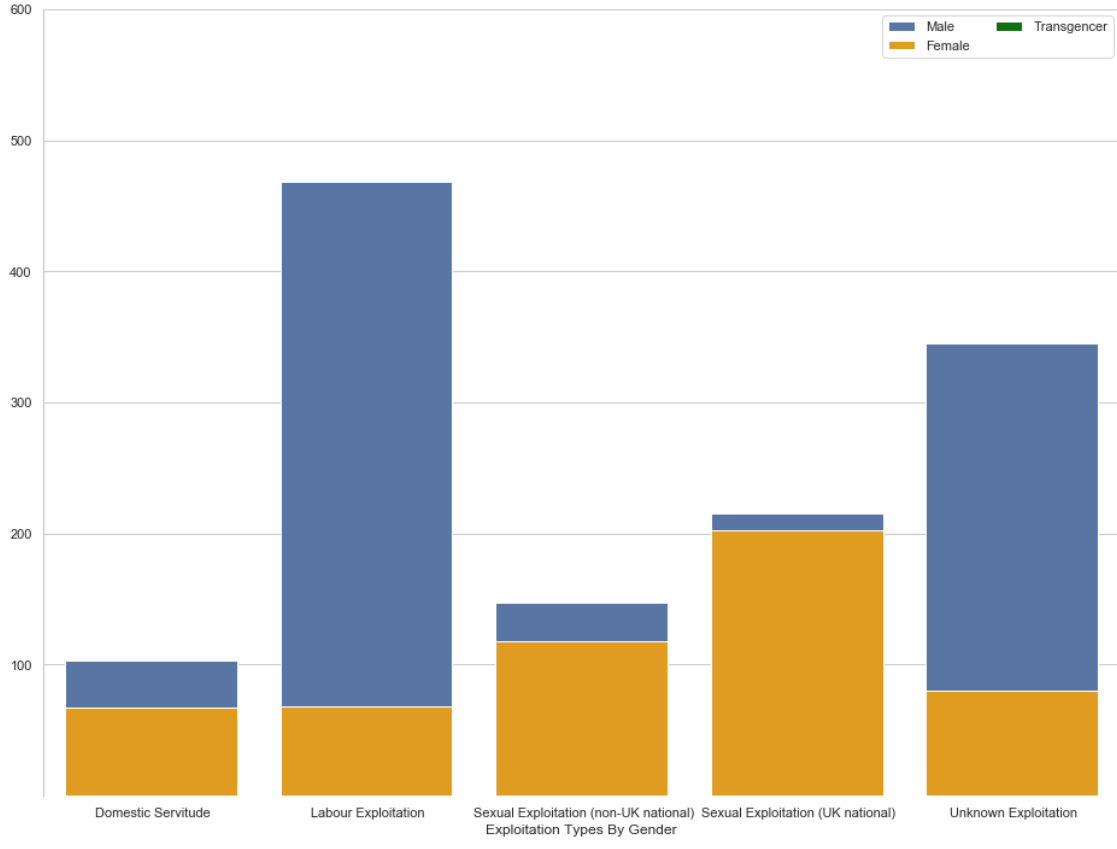
```
f, ax = plt.subplots(figsize=(16,12))
sns.set(style="whitegrid")
sns.set_palette("pastel")
sns.barplot(x="Claimed Exploitation Type", y="Total 2016", dat
a=adult_exploitation_df, label="Male", color="b")
sns.barplot(x="Claimed Exploitation Type", y="Female", data=ad
ult_exploitation_df, label="Female", color="orange")
sns.barplot(x="Claimed Exploitation Type", y="Trans- gender",
data=adult_exploitation_df, label="Transgender", color="green"
)
ax.legend(ncol=2, loc="upper right")
ax.set(ylim=(0, 1200), ylabel="", xlabel="Exploitation Types B
y Gender")
ax.set_yticks([100, 200, 300, 400, 500, 600, 700, 800, 900, 10
00, 1100, 1200])
sns.despine(bottom=True, top=True)
```



We can see that gender rate for specific crimes in adults is very different: the majority of recorded cases of sexual exploitation and domestic servitude have female victims, whereas male victims are the majority in labour exploitation. The 5 recorded cases with a transgender victim were among those of sexual exploitation.

In [13]:

```
f, ax = plt.subplots(figsize=(16,12))
sns.set(style="whitegrid")
sns.set_palette("pastel")
sns.barplot(x="Claimed Exploitation Type", y="Total 2016", dat
a=child_exploitation_df, label="Male", color="b")
sns.barplot(x="Claimed Exploitation Type", y="Female", data=ch
ild_exploitation_df, label="Female", color="orange")
sns.barplot(x="Claimed Exploitation Type", y="Trans- gender",
data=child_exploitation_df, label="Transgencer", color="green"
)
ax.legend(ncol=2, loc="upper right")
ax.set(ylim=(0, 600), ylabel="", xlabel="Exploitation Types By
Gender")
ax.set_yticks([100, 200, 300, 400, 500, 600])
sns.despine(bottom=True, top=True)
```



We can observe a roughly similar pattern with children: the majority of victims of sexual exploitation are female. Victims of labour exploitation are majorly male. Unlike with adults, the ratio in unknown exploitation is male-dominated. Investigating what this label can refer to would be an interesting follow-up question. Next, let's load in some more data and see if there are any obvious patterns when we connect our knowledge thus far with information about the victims' origin countries.

In [14]:

```
a_referral_files = ["2013_referrals_adult.csv", "2014_referrals_adult.csv", "2015_referrals_adult.csv", "2016_referrals_adult.csv"]
c_referral_files = ['2013_referrals_minor.csv', '2014_referrals_minor.csv', '2015_referrals_minor.csv', '2016_referrals_minor.csv']
a_referrals_16 = pd.read_csv('2016_referrals_adult.csv')
m_referrals_16 = pd.read_csv('2016_referrals_minor.csv')
```

In [15]:

```
adult_referral_countries_df = a_referrals_16
```

The function below reads in the specified files and cleans the data to be more easily accessible.

In [16]:

```
def concat_columns(files):
    series = []
    for file in files:
        column = pd.read_csv(file)
        column = column.rename(columns={column.columns[0]: "Nationality/Origin"})
        length = len(column.columns)
        column = column.iloc[:, [0, length - 2]]
        column.set_index('Nationality/Origin', inplace=True)
        column.iloc[:, 0] = column.iloc[:, 0].astype('str').str.replace(',', '').astype('int64')
        series.append(column)
    output_df = pd.concat(series, axis=1, sort=True)
    return output_df
```

In [17]:

```
adult_top_countries = concat_columns(a_referral_files)
adult_top_countries.columns = ['2013', '2014', '2015', '2016']
adult_top_countries.loc["Total", "2015"] = adult_top_countries.loc["Adult Total", "2015"]
adult_top_countries.drop('Adult Total', inplace=True)
```

In [18]:

```
minor_top_countries = concat_columns(c_referral_files)
minor_top_countries.columns = ['2013', '2014', '2015', '2016']
minor_top_countries.loc["Total", "2015"] = minor_top_countries
.loc["Minor Total", "2015"]
minor_top_countries.drop('Minor Total', inplace=True)
```

In [19]:

```
adult_total_values = adult_top_countries.loc['Total', :]
minor_total_values = minor_top_countries.loc['Total', :]
```

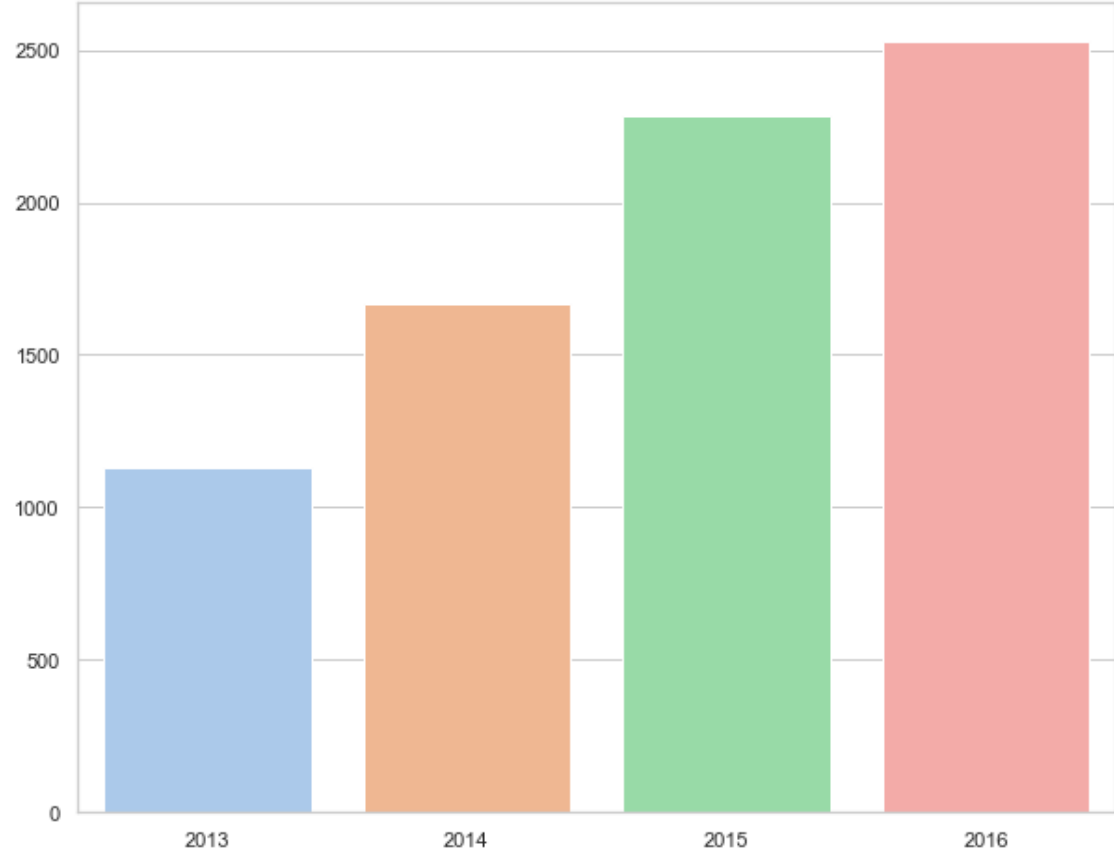
Let's look at total number of cases for all the years in this dataset.

In [20]:

```
plt.figure(figsize=(10,8))
sns.barplot(y=adult_total_values.values, x=adult_total_values.
index)
```

Out[20]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c1f64
d198>
```

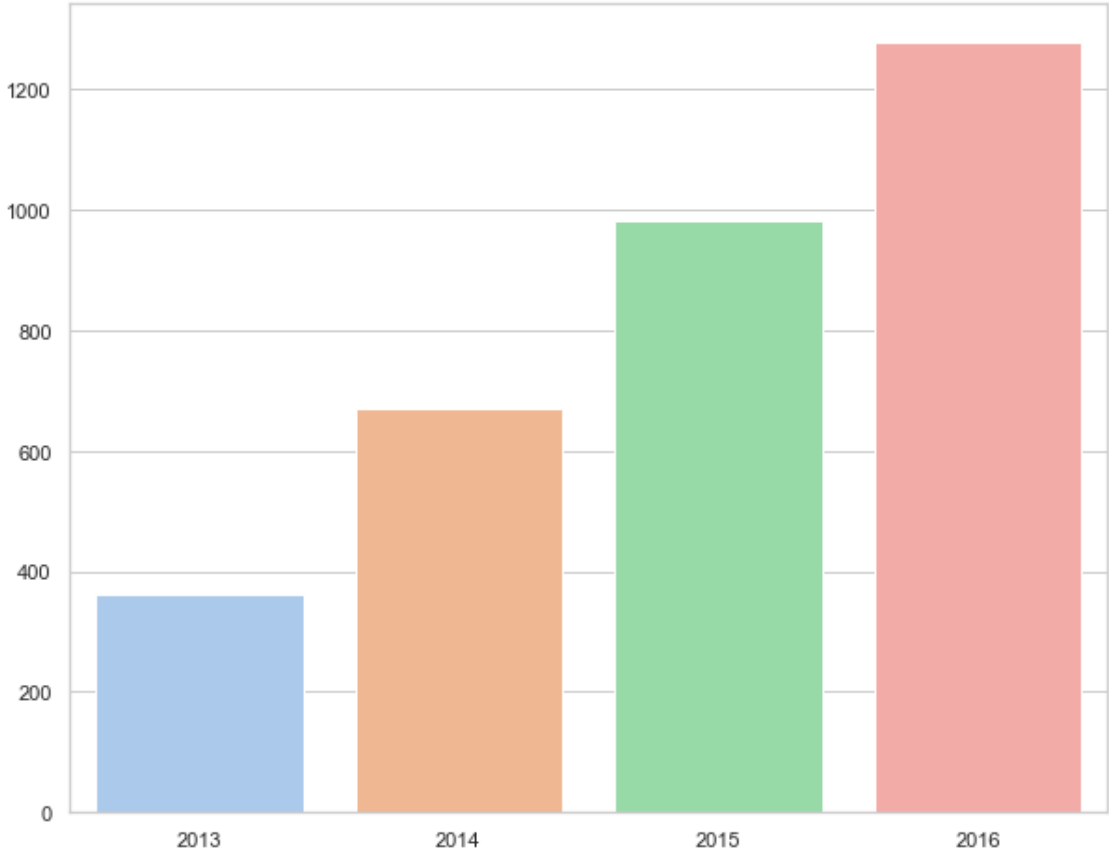


In [21]:

```
plt.figure(figsize=(10,8))
sns.barplot(y=minor_total_values.values, x=minor_total_values.
index)
```

Out[21]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c1f64
ce10>
```



Either there has been an increase in the ability to detect human trafficking, or it is on the rise with both minor and adult victims: each year from 2013-2016 saw an increase in total cases. This should be investigated further in a broader context and with subject expertise.

In [22]:

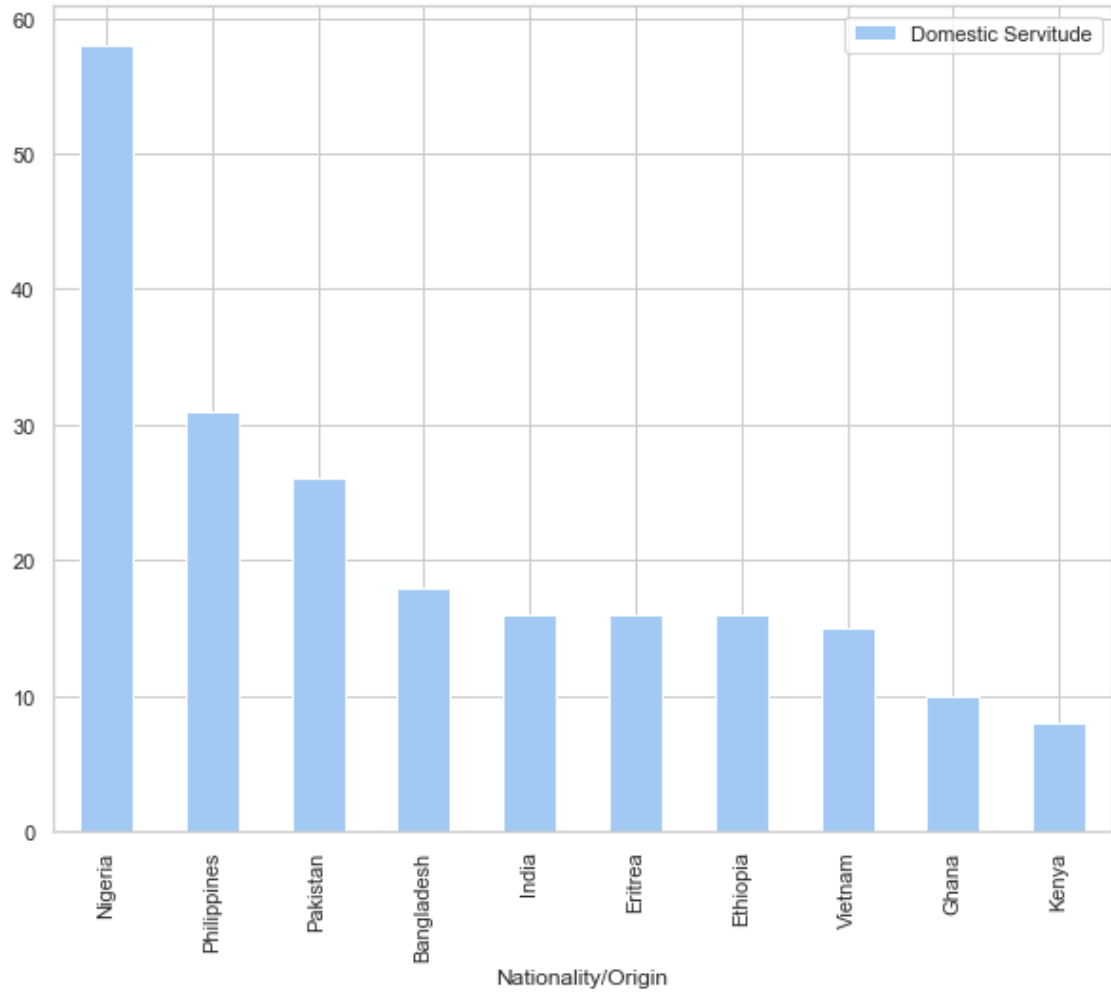
```
adult_referral_countries_df = adult_referral_countries_df.rena
me(columns={adult_referral_countries_df.columns[0]: "Nationali
ty/Origin"})
adult_referral_countries_df.iloc[:, 2:6] = adult_referral_coun
tries_df.iloc[:, 2:6].astype('int64')
adult_referral_countries_df.set_index('Nationality/Origin', in
place=True)
adult_referral_countries_df.drop(['Total'], inplace=True)
```

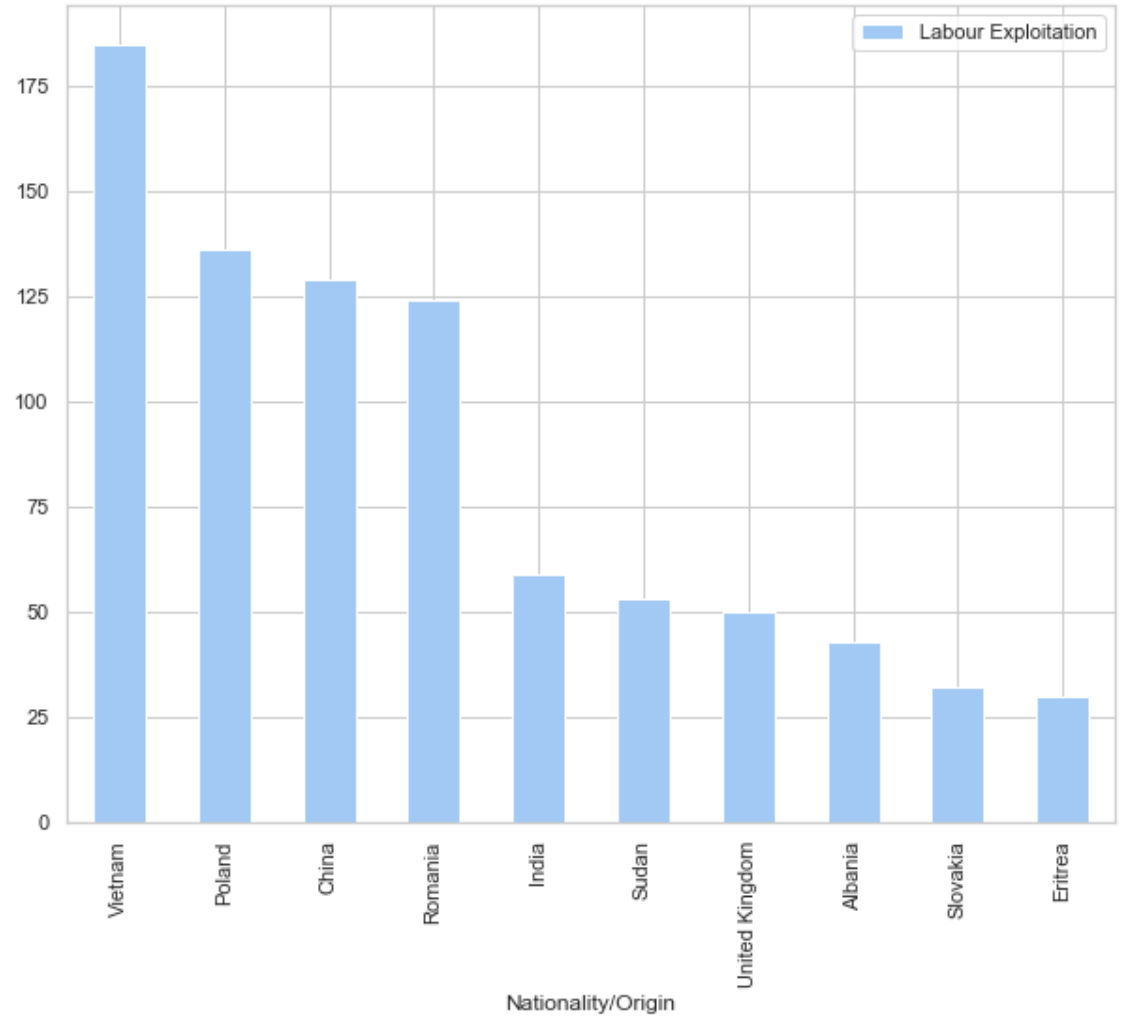
In [23]:

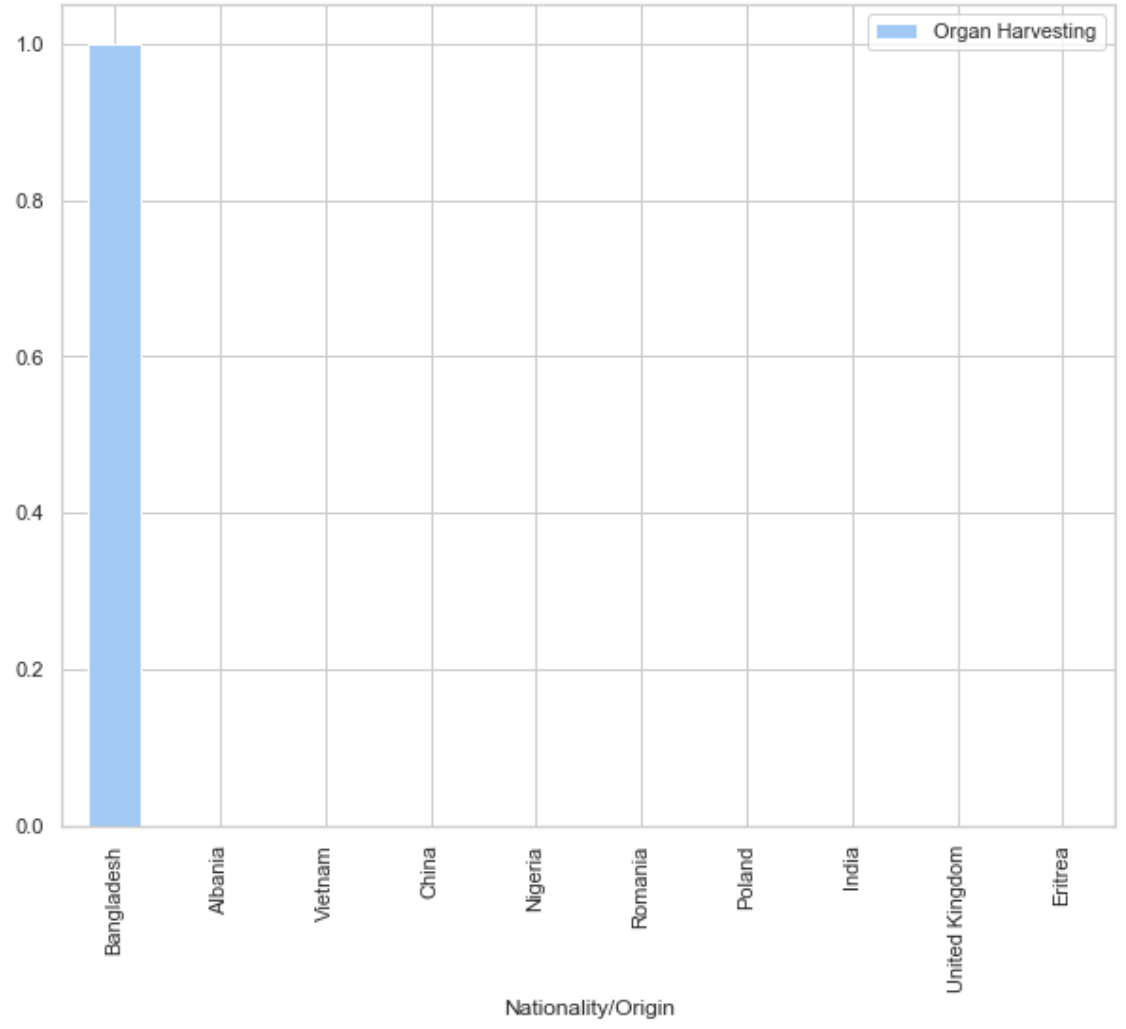
```
m_referrals_16 = m_referrals_16.rename(columns={m_referrals_16
.columns[0]: "Nationality/Origin"})
m_referrals_16.iloc[:, 2:6] = m_referrals_16.iloc[:, 2:6].asty
pe('int64')
m_referrals_16.set_index('Nationality/Origin', inplace=True)
m_referrals_16.drop(['Total'], inplace=True)
```

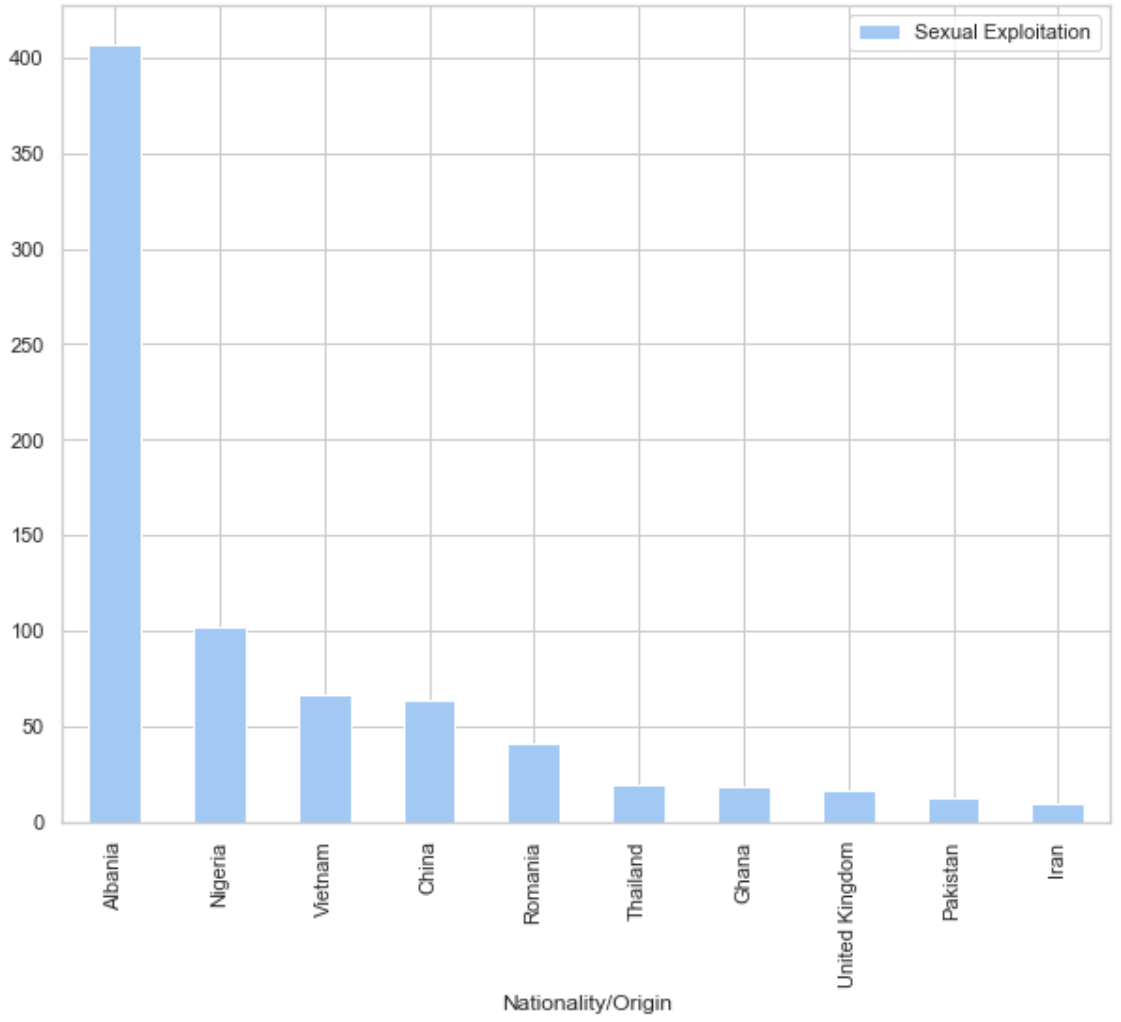
In [30]:

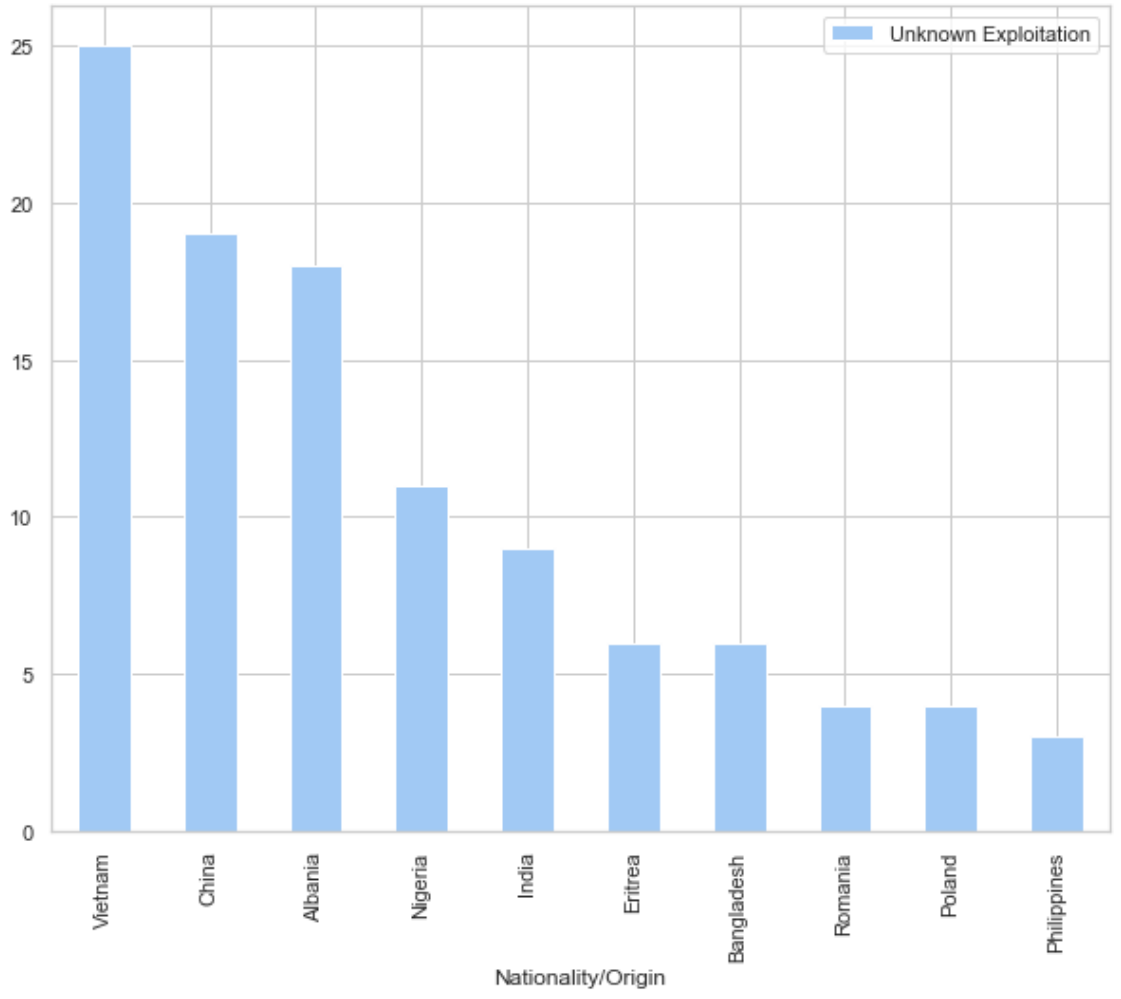
```
adult_crimes = ["Domestic Servitude", "Labour Exploitation", "
Organ Harvesting", "Sexual Exploitation", "Unknown Exploitatio
n"]
for crime in adult_crimes:
    data_16 = adult_referral_countries_df.loc[:, [crime]].nlar
gest(10, [crime])
    data_16.plot.bar(figsize=(10, 8))
```




Labour Exploitation
175
150
125
100
75
50
25
0
Vietnam
Poland
China
Romania
India
Sudan
United Kingdom
Albania
Slovakia
Eritrea
Nationality/Origin


Organ Harvesting
1.0
0.8
0.6
0.4
0.2
0.0
Bangladesh
Albania
Vietnam
China
Nigeria
Romania
Poland
India
United Kingdom
Eritrea
Nationality/Origin

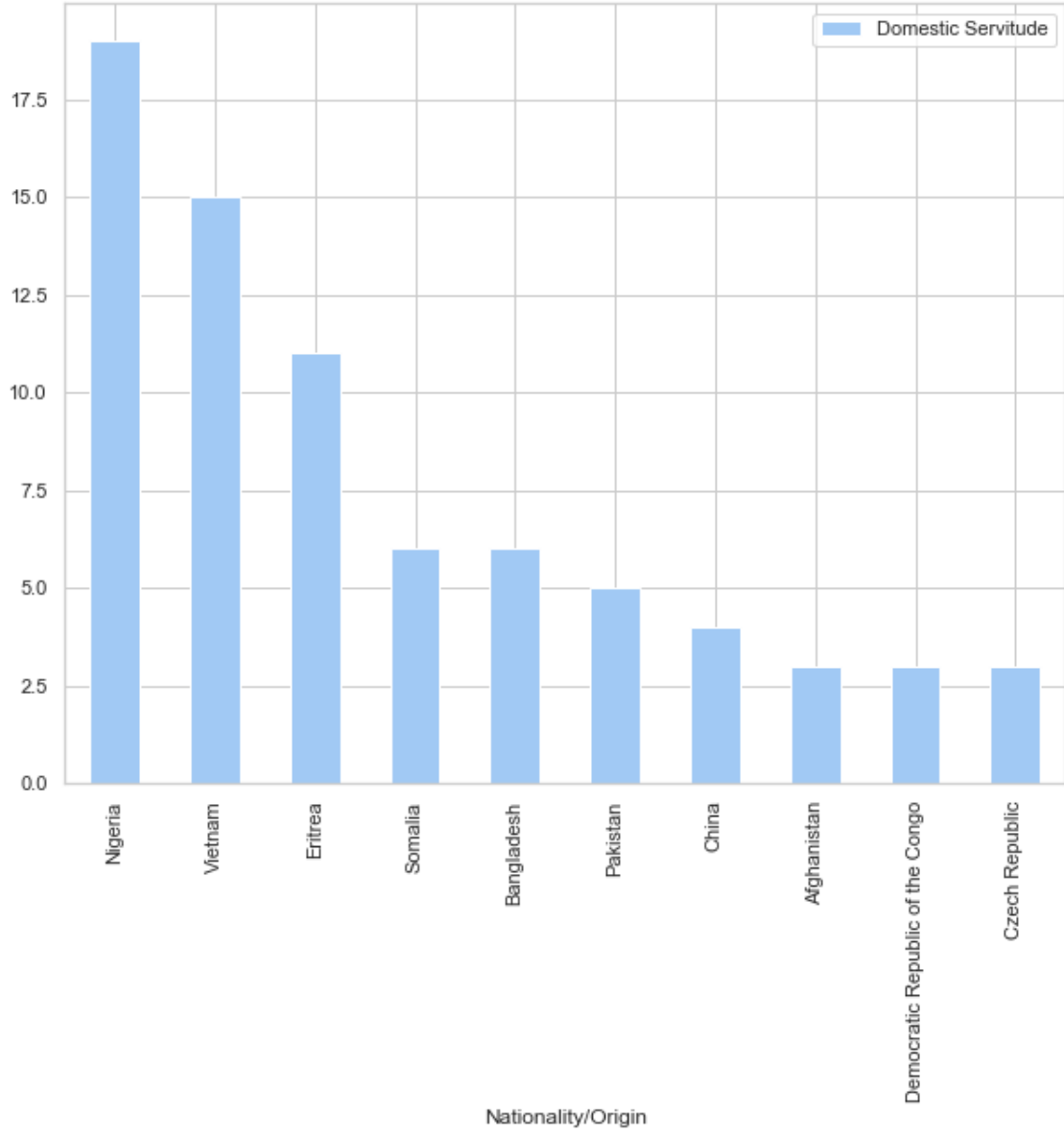

Sexual Exploitation
400
350
300
250
200
150
100
50
0
Albania
Nigeria
Vietnam
China
Romania
Thailand
Ghana
United Kingdom
Pakistan
Iran
Nationality/Origin

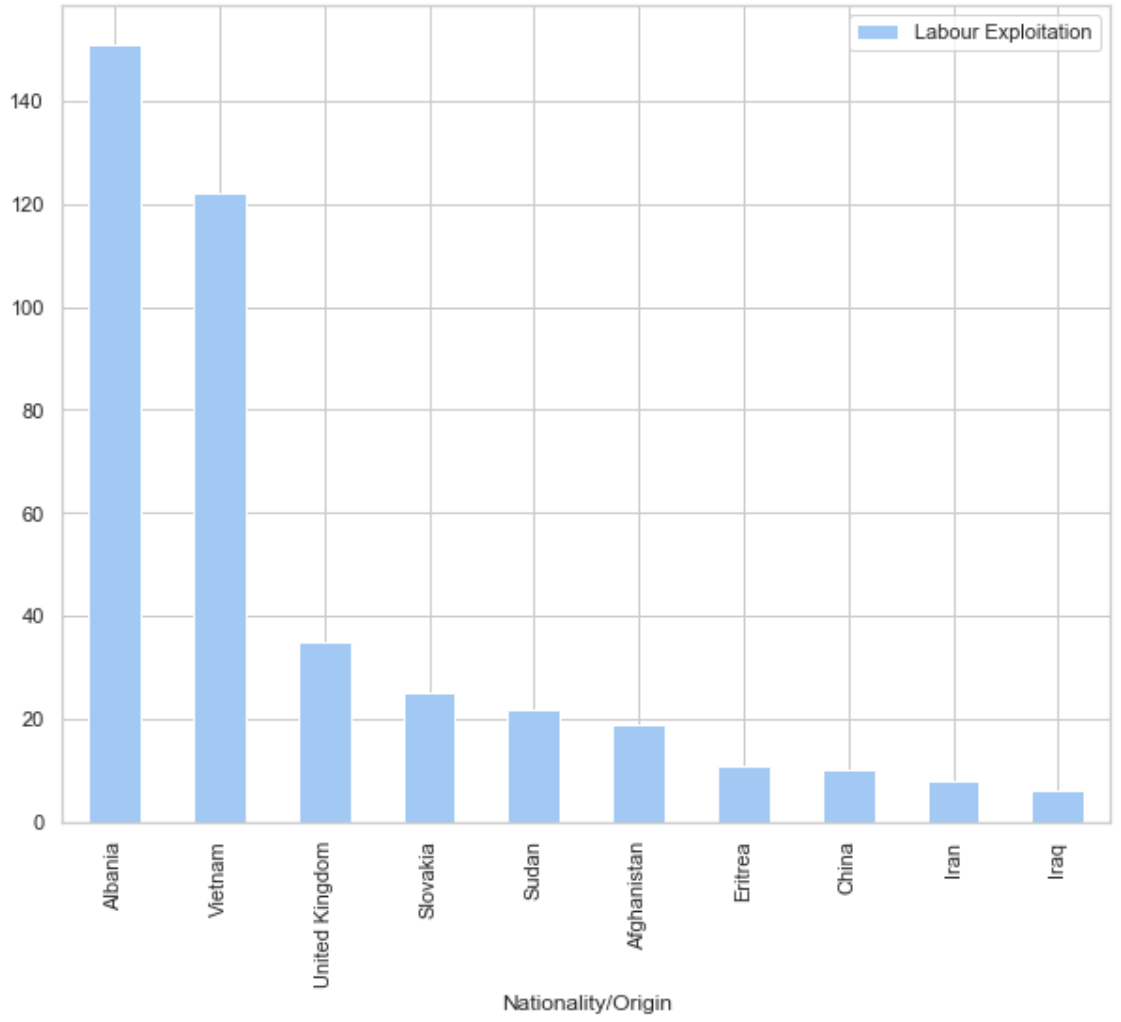

Unknown Exploitation
25
20
15
10
5
0
Vietnam
China
Albania
Nigeria
India
Eritrea
Bangladesh
Romania
Poland
Philippines
Nationality/Origin

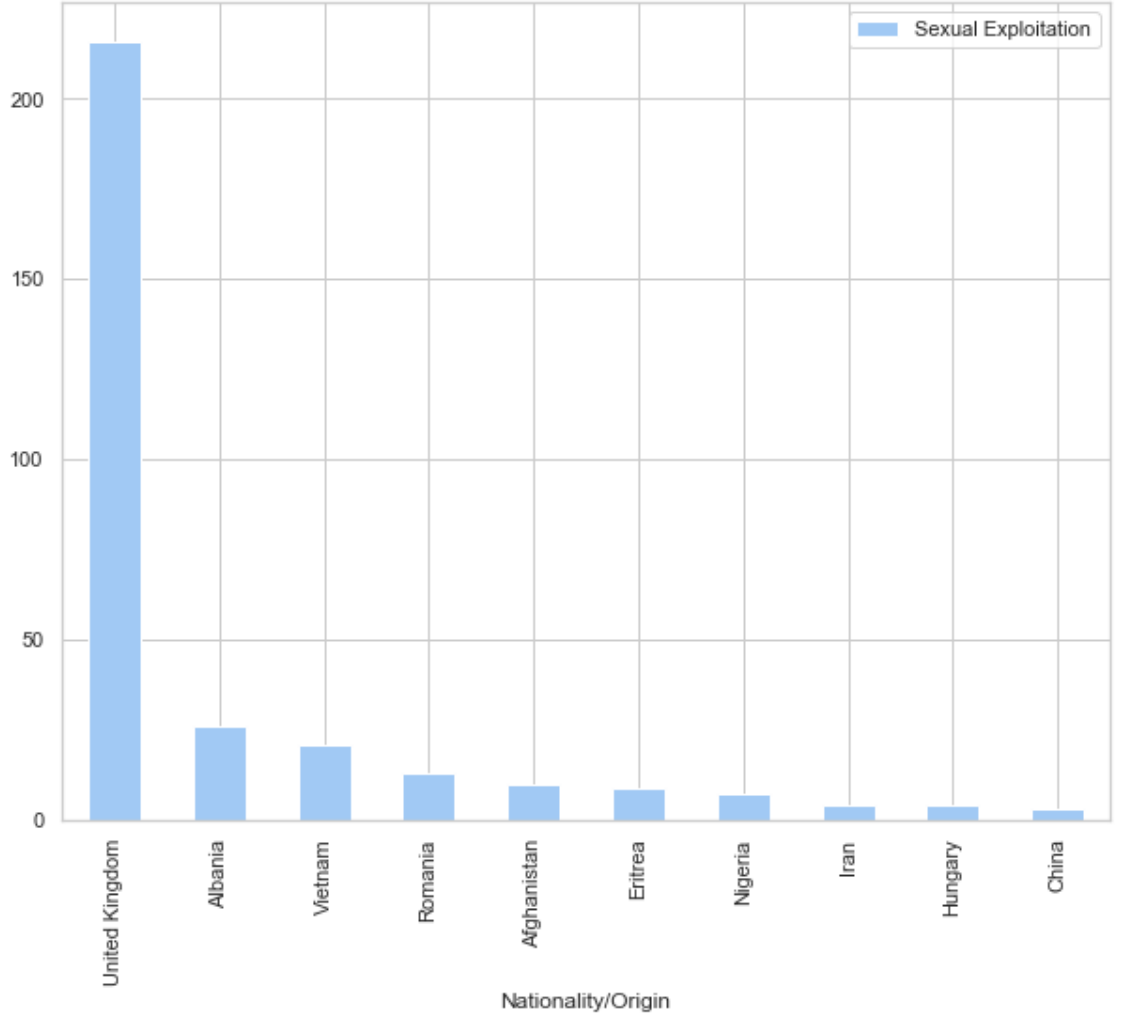
In the statistics for 2016, the victims are not distributed evenly in terms of their origin countries/nationality. In domestic servitude, most victims came from Nigeria. Most victims of labour exploitation were from Vietnam, Poland, China and Romania. The one registered victim of organ harvesting was from Bangladesh. Albania was the origin country for a significant majority of cases of sexual exploitation, with around 400 cases total, compared to 100 victims from Nigeria and below 70 victims from Vietnam, China and Romania. In the category of unknown exploitation, Vietnam, China and Albania were the origin countries for clearly more people than others.

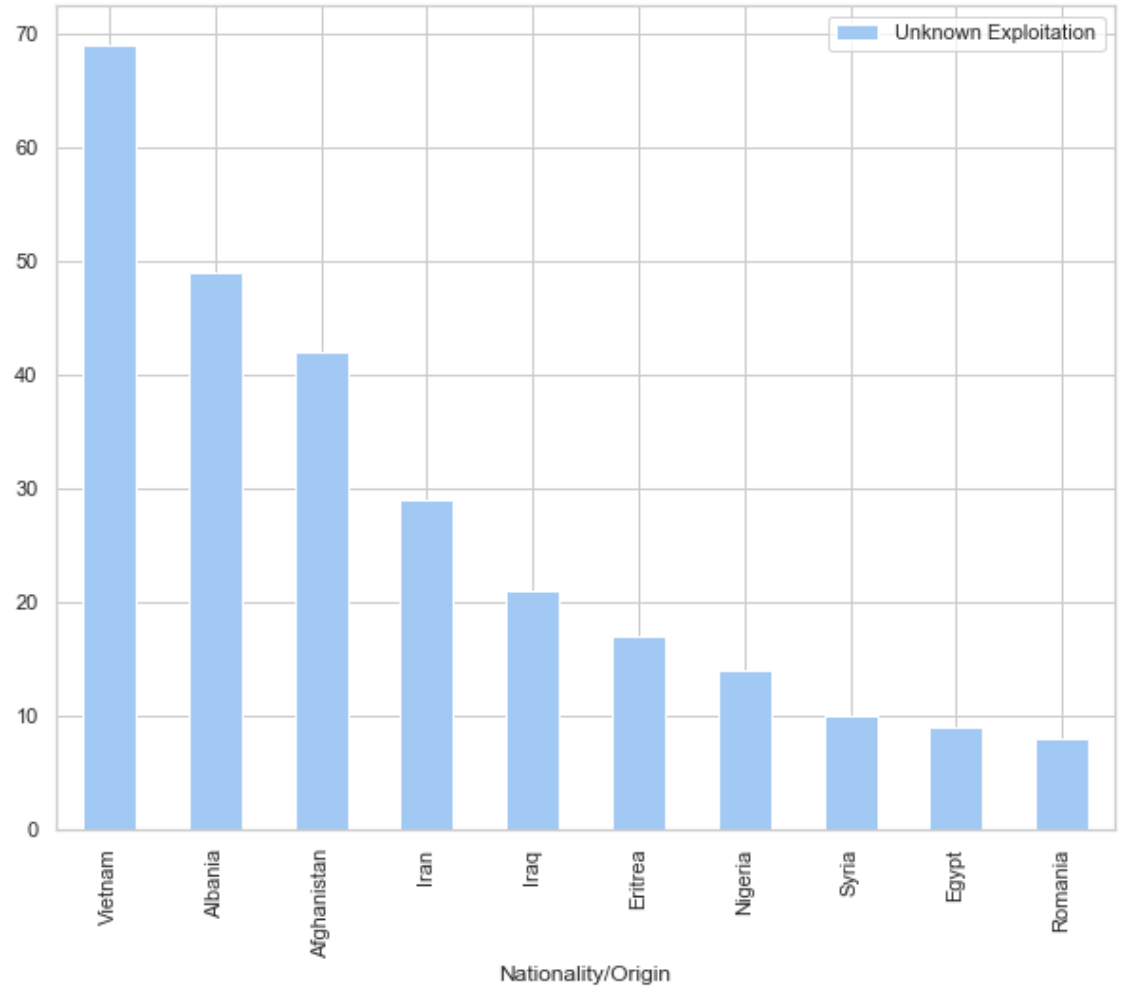
In [31]:

```
minor_crimes = ["Domestic Servitude", "Labour Exploitation", "
Sexual Exploitation", "Unknown Exploitation"]
for crime in minor_crimes:
    data_16 = m_referrals_16.loc[:, [crime]].nlargest(10, [cri
me])
    data_16.plot.bar(figsize=(10, 8))
```




Labour Exploitation
140
120
100
80
60
40
20
0
Albania
Vietnam
United Kingdom
Slovakia
Sudan
Afghanistan
Eritrea
China
Iran
Iraq
Nationality/Origin


Sexual Exploitation
200
150
100
50
0
United Kingdom
Albania
Vietnam
Romania
Afghanistan
Eritrea
Nigeria
Iran
Hungary
China
Nationality/Origin



In recorded cases involving minors during 2016, many similar patterns seem to be observable. In the category of domestic servitude, top 3 countries were Nigeria, Vietnam and Eritrea. A clear majority of victims of labour exploitation came from Albania and Vietnam, with around 150 and 120 cases. A significant majority of victims of sexual exploitation were from United Kingdom, with over 200 cases, compared to around 20 victims from Albania and Vietnam. In the category of unknown exploitation, the top 3 were Vietnam, Albania and Afghanistan, with around 70, 50 and 40 cases.

I will expand this practice analysis in the near future, but for now, by doing a few sweeps on these statistics, I know a lot more about how human trafficking has manifested in the UK during recently.

I know that most common trafficking-related exploitation types are labour exploitation and sexual exploitation, and that gender ratio of victims on the whole is quite even, but differs significantly with specific types of crimes. Most noteworthy example of this could be how most victims of sexual exploition are female but a lesser, but still larger amount of victims of labour exploitation are male. While this matches the stereotypical stories we hear about this subject, it is important to verify what is actually happening, and whether any new patterns are emerging.

I also noticed that the total amount of recorded human trafficking related cases has been on the rise during each successive year from 2013 to 2016; however, this sample is so limited that it needs to be surrounded by a more comprehensive statistical look, as well as understanding of the subject matter.

Common origin countries of victims are Nigeria, Vietnam, Albania, Poland, China, Romania, Afghanistan and United Kingdom. Cross-referencing these with specific exploitation types could yield interesting information.

In my next round of analysis, I want to answer at least these follow-up questions:

- Are the top origin countries roughly the same in 2013-15 as well?
- Are there any clear patterns in how cases from a certain country of of a certain type of exploitation did rise or fall from one year to the next?
- Can I identify any specific cases or interventions, and/or speculate whether they had any kind of impact?
- What kind of profiles could I draw of the most typical types of victims - ie. is it predictable who gets victimised? Where do they come from? What is their recorded gender? Which organization or authority finds and helps them?
- If profiling as mentioned above is viable, are there existing solutions for trying to prevent human trafficking crimes from happening? What kind of systems might be helpful in this task, and how would I go about in creating one?

In [ ]: