In [4]:
```
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [1]:
```
import urllib.request
```

In [6]:
```
urllib.request.urlretrieve("https://raw.githubusercontent.com/datasciencedojo/dat
```

Out[6]:
```
('./glo/titanic.csv', <http.client.HTTPMessage at 0x1c5c5888>)
```

In [8]:
```
titanic_df = pd.read_csv("./glo/titanic.csv")
```

In [9]:
```
titanic_df
```

Out[9]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	
...	...	...	...	...	...	...	...	...	...	...	
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	

891 rows × 12 columns

In [348]:
```
survey_df = pd.read_csv("./survey_results_public.csv")
```

In [362]:
```
survey_quest_df = pd.read_csv("./survey_results_schema.csv", index_col="qname")
```

In [372]:
```
survey_question_only = survey_quest_df["question"]
```

In [375]:
```
survey_question_only
```

Out[375]:
```
qname
S0                  <div><span style="font-size:19px;"><strong>Hel...
MetaInfo                                            Browser Meta Info
S1                  <span style="font-size:22px; font-family: aria...
MainBranch          Which of the following options best describes ...
Employment          Which of the following best describes your cur...
Country             Where do you live? <span style="font-weight: b...
US_State            <p>In which state or territory of the USA do y...
UK_Country          In which part of the United Kingdom do you liv...
S2                  <span style="font-size:22px; font-family: aria...
EdLevel             Which of the following best describes the high...
Age1stCode          At what age did you write your first line of c...
LearnCode           How did you learn to code? Select all that apply.
YearsCode           Including any education, how many years have y...
YearsCodePro        NOT including education, how many years have y...
DevType             Which of the following describes your current ...
OrgSize             Approximately how many people are employed by ...
Currency            Which currency do you use day-to-day? If your ...
CompTotal           What is your current total compensation (salar...
CompFreq             Is that compensation weekly, monthly, or yearly?
S3                  <span style="font-size:22px; font-family: aria...
Language            Which <b>programming, scripting, and markup la...
Database            Which <b>database environments </b>have you do...
Platform            Which <b>cloud platforms</b> have you done ext...
Webframe            Which <strong>web frameworks </strong><span st...
MiscTech            Which <b>other frameworks and libraries</b> ha...
ToolsTech           Which <strong>tools</strong> have you done ext...
NEWCollabTools      Which <strong>development environments</strong...
OpSys               What is the primary operating system in which ...
NEWStuck            What do you do when you get stuck on a problem...
S4                  <span style="font-size:22px; font-family: aria...
NEWSOSites          Which of the following Stack Overflow sites ha...
SOVisitFreq         How frequently would you say you visit Stack O...
SOAccount                        Do you have a Stack Overflow account?
SOPartFreq          How frequently would you say you participate i...
SOComm              Do you consider yourself a member of the Stack...
NEWOtherComms       Are you a member of any other online developer...
NEWOtherCommsNames  Please name up to 5 other online developer com...
S5                  <span style="font-size:22px; font-family: aria...
Age                                                  What is your age?
Gender              Which of the following describe you, if any? P...
Trans                                  Do you identify as transgender?
Sexuality           Which of the following describe you, if any? P...
Ethnicity           Which of the following describe you, if any? P...
Accessibility       Which of the following describe you, if any? P...
MentalHealth        Which of the following describe you, if any? P...
S6                  <span style="font-size:22px;"><strong>Final Qu...
SurveyLength        How do you feel about the length of the survey...
SurveyEase          How easy or difficult was this survey to compl...
Name: question, dtype: object
```

In [379]:
```
survey_question_only["UK_Country"]
```

Out[379]:
```
'In which part of the United Kingdom do you live? <i>This information will be k
ept private.</i>'
```

In [397]:
```
survey_df
```

Out[397]:

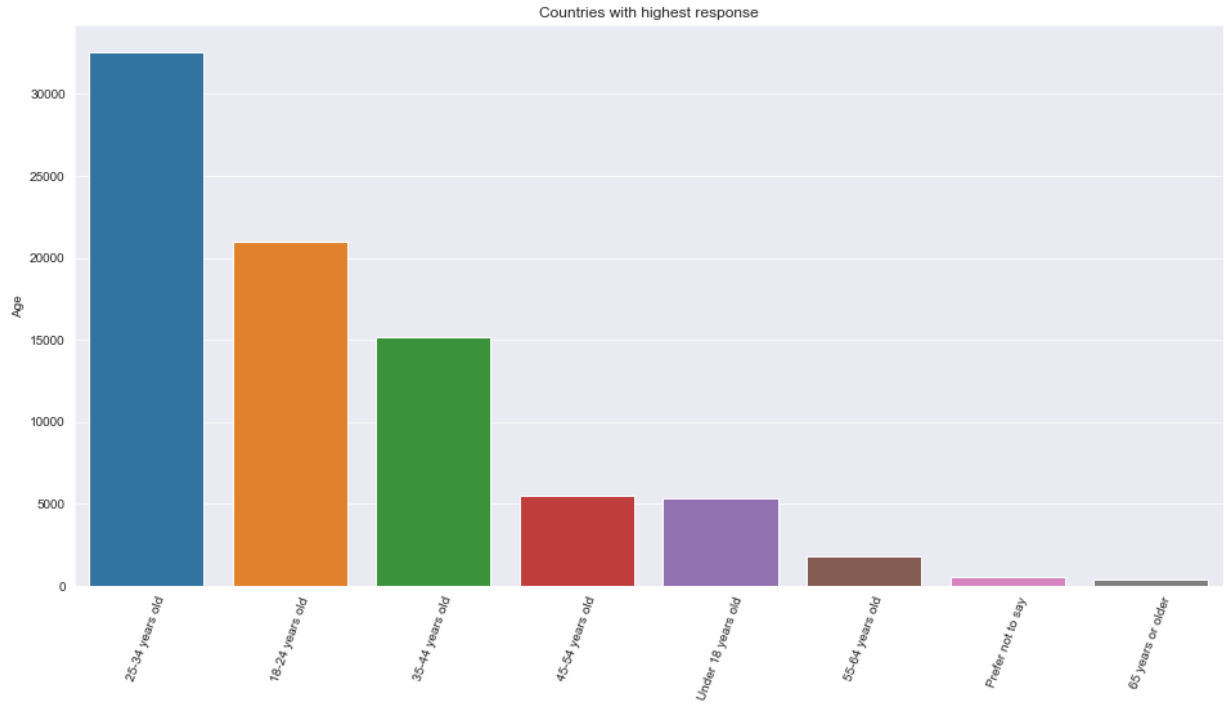
	ResponseId	MainBranch	Employment	Country	US_State	UK_Country	EdLevel
0	1	I am a developer by profession	Independent contractor, freelancer, or self-em...	Slovakia	NaN	NaN	Secondary school (e.g. American high school, G...
1	2	I am a student who is learning to code	Student, full-time	Netherlands	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)
2	3	I am not primarily a developer, but I write co...	Student, full-time	Russian Federation	NaN	NaN	Bachelor's degree (B.A., B.S., B.Eng., etc.)
3	4	I am a developer by	Employed full-time	Austria	NaN	NaN	Master's degree (M.A., M.S.,

In [440]:
```
the_age = survey_df["Age"].value_counts()
```

In [441]:
```
the_age
```

Out[441]:
```
25-34 years old       32568
18-24 years old       20993
35-44 years old       15183
45-54 years old        5472
Under 18 years old     5376
55-64 years old        1819
Prefer not to say       575
65 years or older       421
Name: Age, dtype: int64
```

In [442]:
```
sns.set_style("darkgrid")
plt.figure(figsize=(16,8))
plt.xticks(rotation=70)
plt.title("Countries with highest response")
#plt.bar(Top_response_countries.index, Top_response_countries);
sns.barplot(x=Age.index, y=Age);
```

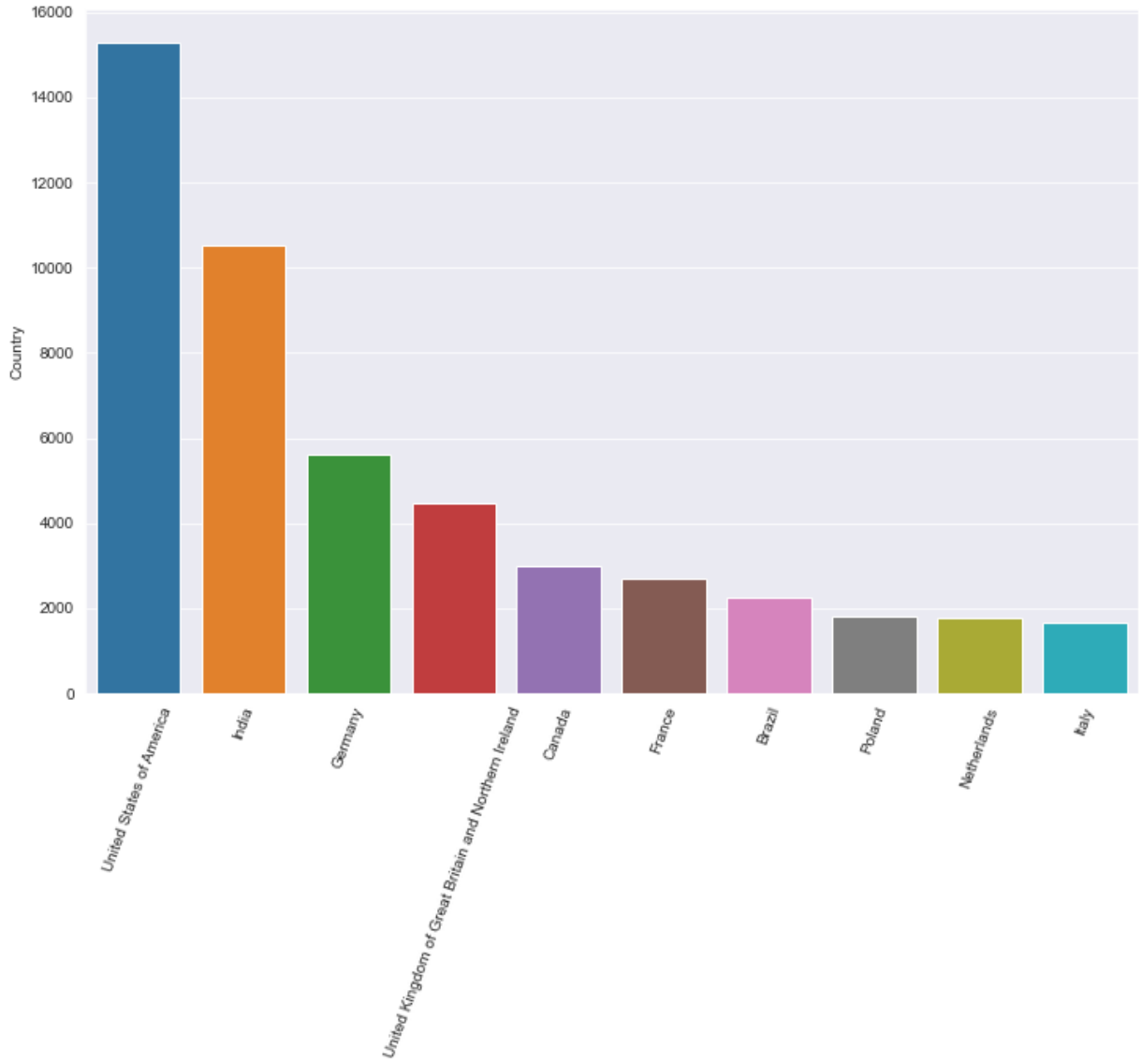


In [404]:
```
country_unique = survey_df["Country"].unique()
```

In [410]:
```
country_uniques = survey_df.Country.value_counts()
```

In [414]:
```
countries_unique = country_uniques.sort_values(ascending=False).head(10)
```

In [421]:
```
plt.figure(figsize=(12,8))
plt.xticks(rotation=70)
sns.barplot(x=countries_unique.index, y=countries_unique);
```



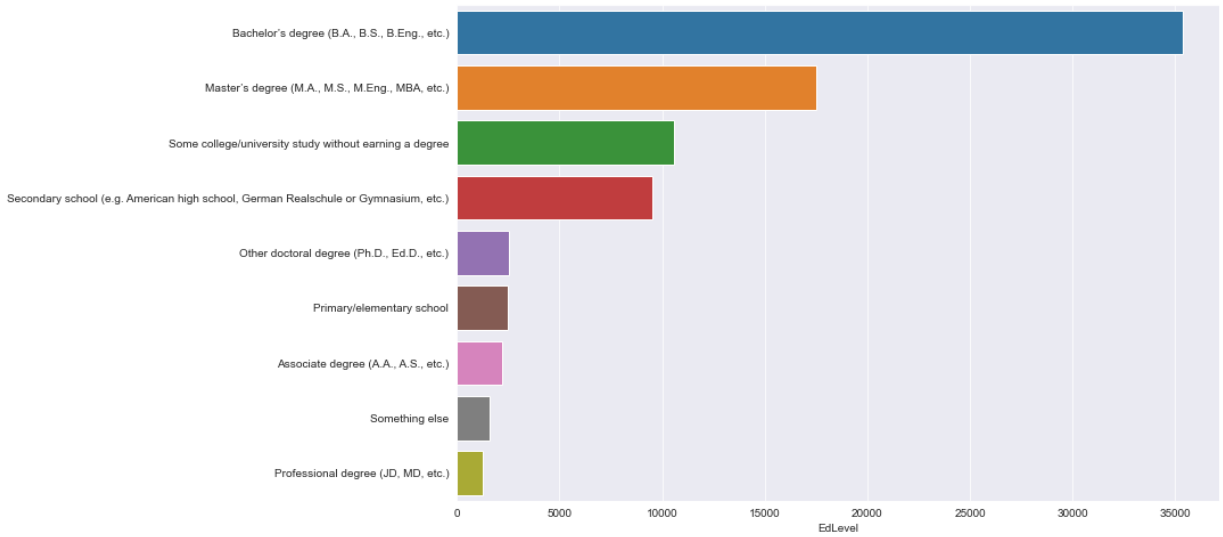
In [435]:
```
education = survey_df["EdLevel"].value_counts()
```

In [436]:
```
education
```

Out[436]:
```
Bachelor’s degree (B.A., B.S., B.Eng., etc.)
35357
Master’s degree (M.A., M.S., M.Eng., MBA, etc.)
17512
Some college/university study without earning a degree
10589
Secondary school (e.g. American high school, German Realschule or Gymnasium, et
c.)     9534
Other doctoral degree (Ph.D., Ed.D., etc.)
2567
Primary/elementary school
2479
Associate degree (A.A., A.S., etc.)
2231
Something else
1601
Professional degree (JD, MD, etc.)
1256
Name: EdLevel, dtype: int64
```

In [439]:
```
plt.figure(figsize=(12, 8))
sns.barplot(y=education.index, x=education)
```

Out[439]:
```
<AxesSubplot:xlabel='EdLevel'>
```



In [390]:
```
Nigeria_survey = survey_df[survey_df.Country == "Nigeria"]
```

In [393]:
```
Nigerians_survey = Nigeria_survey[["Country", "LanguageHaveWorkedWith"]]
```

In [396]:
```
Nigerians_survey["LanguageHaveWorkedWith"].value_counts()
```

Out[396]:
```
HTML/CSS;JavaScript                          14
HTML/CSS;JavaScript;Python                   12
HTML/CSS;JavaScript;Node.js                   9
HTML/CSS                                      9
HTML/CSS;JavaScript;PHP;SQL                   8
                                             ..
HTML/CSS;JavaScript;SQL                       1
Bash/Shell;C;C++;HTML/CSS;JavaScript          1
Dart;HTML/CSS;JavaScript;Node.js;PHP;SQL      1
APL;HTML/CSS;Java;PHP                         1
Go;Python                                     1
Name: LanguageHaveWorkedWith, Length: 270, dtype: int64
```

In [384]:
```
survey_df["LanguageHaveWorkedWith"].nunique()
```

Out[384]:
```
28690
```

In [273]:
```
survey_df.Employment.value_counts()
```

Out[273]:
```
Employed full-time                                      53584
Student, full-time                                      11781
Independent contractor, freelancer, or self-employed    8041
Not employed, but looking for work                      2961
Employed part-time                                      2461
Student, part-time                                      2051
Not employed, and not looking for work                  1228
I prefer not to say                                      890
Retired                                                  326
Name: Employment, dtype: int64
```

In [242]:
```
survey_df.Country.nunique()
```

Out[242]:
```
181
```

In [206]:
```
country_counts = survey_df.Country.value_counts()
```

In [207]:
```
country_counts
```

Out[207]:
```
United States of America                              15288
India                                                 10511
Germany                                                5625
United Kingdom of Great Britain and Northern Ireland   4475
Canada                                                 3012
                                                      ...
Saint Kitts and Nevis                                     1
Dominica                                                  1
Saint Vincent and the Grenadines                          1
Tuvalu                                                    1
Papua New Guinea                                          1
Name: Country, Length: 181, dtype: int64
```

In [219]:
```
Top_response_countries = country_counts.sort_values(ascending=False).head(10)
```

In [243]:
```
Top_response_countries
```

Out[243]:
```
United States of America                                15288
India                                                   10511
Germany                                                  5625
United Kingdom of Great Britain and Northern Ireland     4475
Canada                                                   3012
France                                                   2708
Brazil                                                   2254
Poland                                                   1805
Netherlands                                              1772
Italy                                                    1666
Name: Country, dtype: int64
```

In [138]:
```
survey_df.columns
```

Out[138]:
```
Index(['ResponseId', 'MainBranch', 'Employment', 'Country', 'US_State',
       'UK_Country', 'EdLevel', 'Age1stCode', 'LearnCode', 'YearsCode',
       'YearsCodePro', 'DevType', 'OrgSize', 'Currency', 'CompTotal',
       'CompFreq', 'LanguageHaveWorkedWith', 'LanguageWantToWorkWith',
       'DatabaseHaveWorkedWith', 'DatabaseWantToWorkWith',
       'PlatformHaveWorkedWith', 'PlatformWantToWorkWith',
       'WebframeHaveWorkedWith', 'WebframeWantToWorkWith',
       'MiscTechHaveWorkedWith', 'MiscTechWantToWorkWith',
       'ToolsTechHaveWorkedWith', 'ToolsTechWantToWorkWith',
       'NEWCollabToolsHaveWorkedWith', 'NEWCollabToolsWantToWorkWith', 'OpSys',
       'NEWStuck', 'NEWSOSites', 'SOVisitFreq', 'SOAccount', 'SOPartFreq',
       'SOComm', 'NEWOtherComms', 'Age', 'Gender', 'Trans', 'Sexuality',
       'Ethnicity', 'Accessibility', 'MentalHealth', 'SurveyLength',
       'SurveyEase', 'ConvertedCompYearly'],
      dtype='object')
```

In [ ]:
```

```

In [336]:
```
employment_code = survey_df[["Employment", "LanguageHaveWorkedWith"]]
```

In [337]:
```
employment_code
```

Out[337]:

	Employment	LanguageHaveWorkedWith
0	Independent contractor, freelancer, or self-em...	C++;HTML/CSS;JavaScript;Objective-C;PHP;Swift
1	Student, full-time	JavaScript;Python
2	Student, full-time	Assembly;C;Python;R;Rust
3	Employed full-time	JavaScript;TypeScript
4	Independent contractor, freelancer, or self-em...	Bash/Shell;HTML/CSS;Python;SQL
...	...	...
83434	Employed full-time	Clojure;Kotlin;SQL
83435	Independent contractor, freelancer, or self-em...	NaN
83436	Employed full-time	Groovy;Java;Python
83437	Employed full-time	Bash/Shell;JavaScript;Node.js;Python
83438	Employed full-time	Delphi;Elixir;HTML/CSS;Java;JavaScript

83439 rows × 2 columns

In [333]:
```
student_full_time = employment_code[employment_code.Employment == "Student, full-
```

In [339]:
```
student_full_time
```

Out[339]:

	LanguageHaveWorkedWith	Employment
1	JavaScript;Python	Student, full-time
2	Assembly;C;Python;R;Rust	Student, full-time
7	HTML/CSS;JavaScript;PHP;Ruby;SQL;TypeScript	Student, full-time
15	Bash/Shell;Python	Student, full-time
20	Java;Python	Student, full-time
...	...	...
83391	Bash/Shell;C;C#;Dart;HTML/CSS;JavaScript;Python	Student, full-time
83406	C;C++;Python	Student, full-time
83414	HTML/CSS;Python	Student, full-time
83417	Bash/Shell;C;C++;Java;SQL	Student, full-time
83418	Bash/Shell;C#;C++;Java;JavaScript;Kotlin;PHP;P...	Student, full-time

11781 rows × 2 columns

In [451]:
```
col_series = student_full_time["LanguageHaveWorkedWith"].value_counts()
```

In [452]:
```
col_series
```

Out[452]:
```
Python                                                                298
HTML/CSS;JavaScript;Python                                            135
C++;Python                                                             87
HTML/CSS;Python                                                        80
Java;Python                                                            78
                                                                     ...
C;Haskell;Java;JavaScript;Matlab;Python;SQL                             1
Bash/Shell;C;C#;Kotlin;VBA                                              1
C++;Java;JavaScript;Node.js;Python;SQL;TypeScript                       1
Assembly;C++;HTML/CSS;Java;JavaScript;Python;SQL                        1
Bash/Shell;C#;C++;Java;JavaScript;Kotlin;PHP;PowerShell;Python;SQL      1
Name: LanguageHaveWorkedWith, Length: 6465, dtype: int64
```

In [ ]:
```
survey_df.LanguageHaveWorkedWith
```

In [455]:
```
def split_multicolumn(col_series):   #function for splitting multi columns
    result_df = col_series.to_frame()
    options = []
    # Iterate over the column
    for idx, value  in col_series[col_series.notnull()].iteritems():
        # Break each value into list of options
        for option in value.split(';'):
            # Add the option as a column to result
            if not option in result_df.columns:
                options.append(option)
                result_df[option] = False
            # Mark the value in the option column as True
            result_df.at[idx, option] = True
    return result_df[options]
```

In [457]:
```
popular_language = split_multicolumn(survey_df.LanguageHaveWorkedWith)
```

In [458]:
```
popular_language
```

Out[458]:

	C++	HTML/CSS	JavaScript	Objective-C	PHP	Swift	Python	Assembly	C	R
0	True	True	True	True	True	True	False	False	False	False
1	False	False	True	False	False	False	True	False	False	False
2	False	False	False	False	False	False	True	True	True	True
3	False	False	True	False	False	False	False	False	False	False
4	False	True	False	False	False	False	True	False	False	False
...	...	...	...	...	...	...	...	...	...	...
83434	False	False	False	False	False	False	False	False	False	False
83435	False	False	False	False	False	False	False	False	False	False
83436	False	False	False	False	False	False	True	False	False	False
83437	False	False	True	False	False	False	True	False	False	False
83438	False	True	True	False	False	False	False	False	False	False

83439 rows × 38 columns

In [461]:
```
popular_language.sum().sort_values(ascending=False).head(10)
```

Out[461]:
```
JavaScript    53587
HTML/CSS      46259
Python        39792
SQL           38835
Java          29162
Node.js       27975
TypeScript    24909
C#            22984
Bash/Shell    22385
C++           20057
dtype: int64
```

In [347]:
```
student_program.sort_values(ascending=False).head(5)
```

Out[347]:
```
Python                          298
HTML/CSS;JavaScript;Python      135
C++;Python                       87
HTML/CSS;Python                  80
Java;Python                      78
Name: LanguageHaveWorkedWith, dtype: int64
```

In [ ]:
```
professionals_df = employment_code[employment_code.Employment == "Student, full-t
```

In [296]:
```
sorting_student = employment_code.groupby("LanguageHaveWorkedWith")["Employment"]
```

In [297]:
```
sorting_student
```

Out[297]:
```
LanguageHaveWorkedWith
APL                            83
APL;Assembly                    5
APL;Assembly;Bash/Shell         1
APL;Assembly;Bash/Shell;C       1
APL;Assembly;Bash/Shell;C#      1
                               ..
Swift;TypeScript                3
Swift;TypeScript;VBA            1
Swift;VBA                       1
TypeScript                     52
VBA                            46
Name: Employment, Length: 28690, dtype: int64
```

In [295]:
```
sorting_student.sort_values("Student, full-time", ascending=False).head(10)
```

```
C:\New\envs\snakes\lib\site-packages\ipykernel_launcher.py:1: FutureWarning: In
a future version of pandas all arguments of Series.sort_values will be keyword-
only
  """Entry point for launching an IPython kernel.

---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
C:\New\envs\snakes\lib\site-packages\pandas\core\generic.py in _get_axis_number
(cls, axis)
    545         try:
--> 546             return cls._AXIS_TO_AXIS_NUMBER[axis]
    547         except KeyError:

KeyError: 'Student, full-time'

During handling of the above exception, another exception occurred:

ValueError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_5316\1738020599.py in <module>
----> 1 sorting_student.sort_values("Student, full-time", ascending=False).head
(10)

C:\New\envs\snakes\lib\site-packages\pandas\util\_decorators.py in wrapper(*arg
s, **kwargs)
    309                     stacklevel=stacklevel,
    310                 )
--> 311             return func(*args, **kwargs)
    312
    313         return wrapper

C:\New\envs\snakes\lib\site-packages\pandas\core\series.py in sort_values(self,
axis, ascending, inplace, kind, na_position, ignore_index, key)
   3428         inplace = validate_bool_kwarg(inplace, "inplace")
   3429         # Validate the axis parameter
-> 3430         self._get_axis_number(axis)
   3431
   3432         # GH 5856/5853

C:\New\envs\snakes\lib\site-packages\pandas\core\generic.py in _get_axis_number
(cls, axis)
    546             return cls._AXIS_TO_AXIS_NUMBER[axis]
    547         except KeyError:
--> 548             raise ValueError(f"No axis named {axis} for object type {cl
s.__name__}")
    549
    550     @final

ValueError: No axis named Student, full-time for object type Series
```

In [275]:
```
survey_df["Employment"].value_counts()
```

Out[275]:
```
Employed full-time                                    53584
Student, full-time                                    11781
Independent contractor, freelancer, or self-employed     8041
Not employed, but looking for work                     2961
Employed part-time                                     2461
Student, part-time                                     2051
Not employed, and not looking for work                 1228
I prefer not to say                                     890
Retired                                                 326
Name: Employment, dtype: int64
```

In [263]:
```
survey_df["DevType"].value_counts()
```

Out[263]:
```
Developer, full-stack
8601
Developer, back-end
5467
Developer, front-end
2402
Developer, front-end;Developer, full-stack;Developer, back-end
2141
Developer, full-stack;Developer, back-end
1774

...
Developer, mobile;Developer, desktop or enterprise applications;Developer, full
-stack;Engineer, data
1
Developer, mobile;Developer, front-end;Developer, desktop or enterprise applica
tions;Developer, full-stack;Engineer, data;Developer, back-end;Scientist;Develo
per, QA or test;Developer, game or graphics;Developer, embedded applications or
devices;DevOps specialist;Engineering manager;Engineer, site reliability;Produc
t manager;Senior Executive (C-Suite, VP, etc.);Marketing or sales professional
1
Developer, back-end;Student;Developer, embedded applications or devices;Designe
r
1
Developer, front-end;Developer, full-stack;Engineer, data;Data scientist or mac
hine learning specialist;Developer, back-end;Database administrator;Designer;En
gineering manager;Engineer, site reliability;Product manager;Senior Executive
(C-Suite, VP, etc.);Marketing or sales professional
1
Developer, mobile;Developer, desktop or enterprise applications;Data scientist
or machine learning specialist;Developer, back-end;Engineering manager
1
Name: DevType, Length: 8626, dtype: int64
```

In [252]:
```
developers_profession = survey_df.MainBranch[survey_df.MainBranch == "I am a deve
```

In [253]:
```
total_response = survey_df["ResponseId"].count()
```

In [256]:
```
develope_percentage = developers_profession * 100 / total_response
```

In [258]:
```
import math
develope_percentage = math.ceil(develope_percentage)
```

In [259]:
```
develope_percentage
```

Out[259]:
```
70
```

In [251]:
```
survey_df.MainBranch[survey_df.MainBranch == "I code primarily as a hobby"].count
```

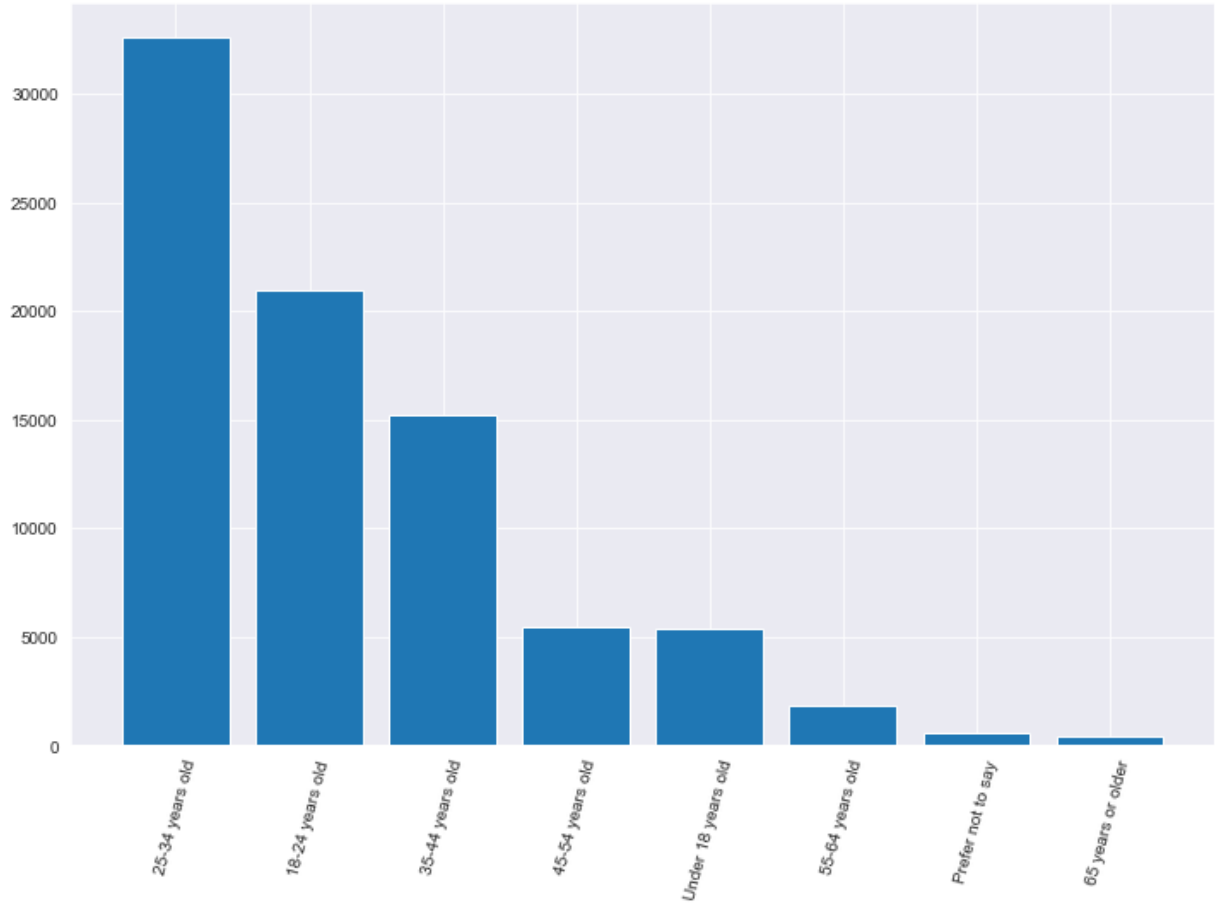
Out[251]:
```
4929
```

In [144]:
```
Age = survey_df.Age.value_counts()
```

In [149]:
```
Age
```

Out[149]:
```
25-34 years old       32568
18-24 years old       20993
35-44 years old       15183
45-54 years old        5472
Under 18 years old     5376
55-64 years old        1819
Prefer not to say       575
65 years or older       421
Name: Age, dtype: int64
```

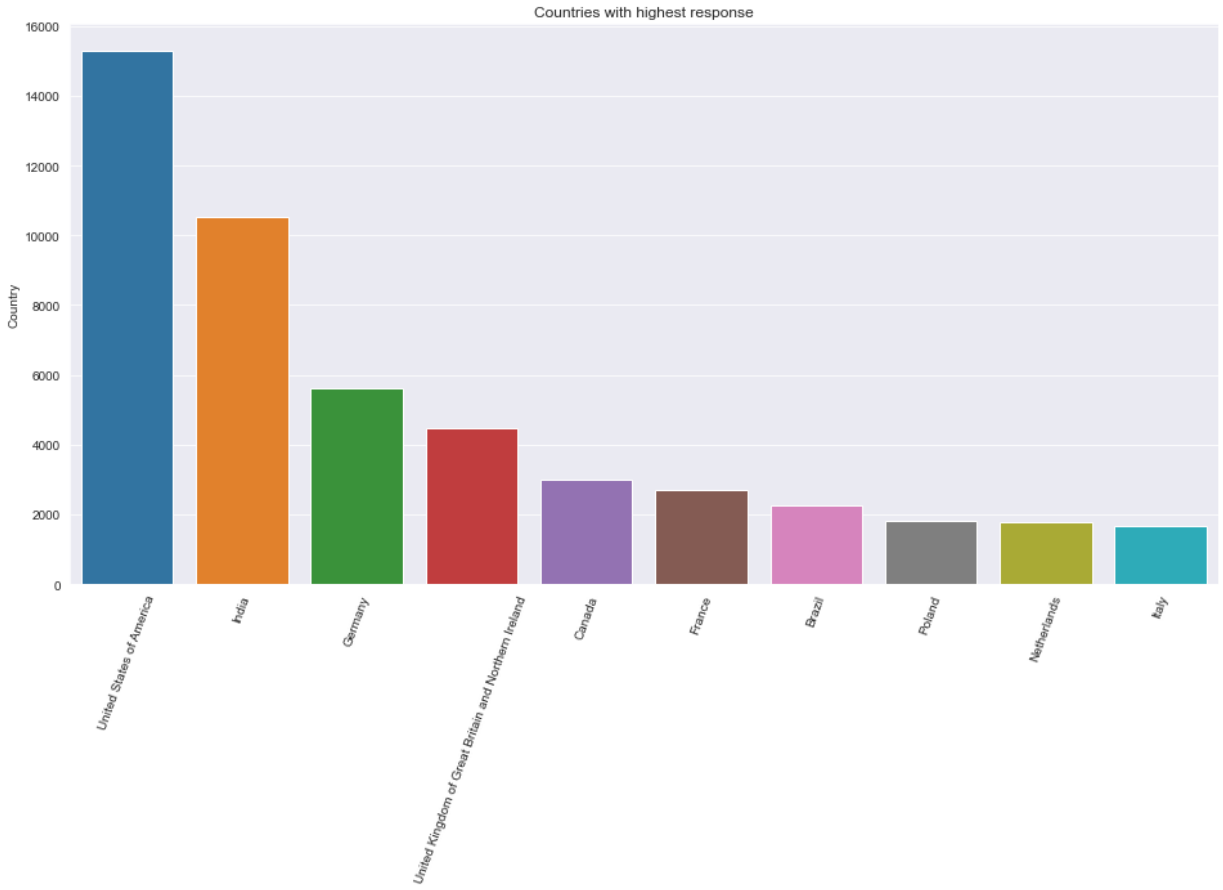
In [244]:
```
sns.set_style("darkgrid")
plt.figure(figsize=(12,8))
plt.xticks(rotation=75)
plt.bar(Age.index, Age);
```



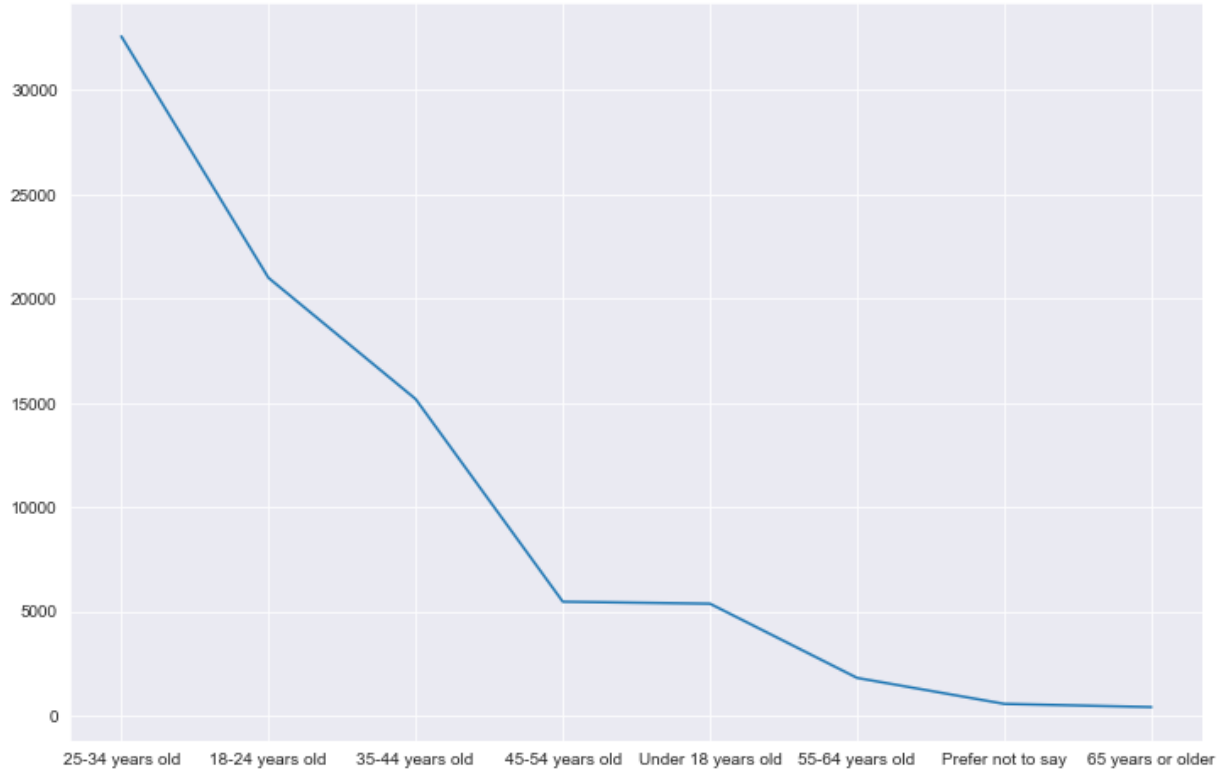
In [157]:
```
Age["25-34 years old"]
```

Out[157]: 32568

In [238]:
```
sns.set_style("darkgrid")
plt.figure(figsize=(16,8))
plt.xticks(rotation=70)
plt.title("Countries with highest response")
#plt.bar(Top_response_countries.index, Top_response_countries);
sns.barplot(x=Top_response_countries.index, y=Top_response_countries);
```



In [224]:
```
sns.set_style("darkgrid")
plt.figure(figsize=(12,8))
plt.plot(Age.index, Age);
```



In [225]:
```
country = survey_df.Country.value_counts()
```

In [226]:
```
country_above_3000 = country[country > 5]
```

In [227]:
```
country_below_3000 = country[country < 5]
```

In [228]:
```
country_below_3000.shape
```

Out[228]: (24,)

In [229]:
```
def split_multicolumn(col_series):
    result_df = col_series.to_frame()
    options = []
    # Iterate over the column
    for idx, value  in col_series[col_series.notnull()].iteritems():
        # Break each value into list of options
        for option in value.split(';'):
            # Add the option as a column to result
            if not option in result_df.columns:
                options.append(option)
                result_df[option] = False
            # Mark the value in the option column as True
            result_df.at[idx, option] = True
    return result_df[options]
```

Out[229]:
```
(150,)
```