

Introduction

This dataset contains details of various nations and their flags. With this data we are trying ## to predict the religion of a country from the shapes and the colours in its flag.

Flags Data Analysis - Part I

Import Dataset

In [2]:

```
path="C:\\Users\\user\\Assignment 2"
```

In [3]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [4]:

```
flags = pd.read_csv(path+"/flags_headers.csv")
```

Data Cleaning and Exploration

In [5]:

```
print(flags.head(5))
```

	country	landmass	zone	area	population	language	religion	b
ars \								
0	Afghanistan	5	1	648	16	10	2	
0								
1	Albania	3	1	29	3	6	6	
0								
2	Algeria	4	1	2388	20	8	2	
2								
3	American-Samoa	6	3	0	0	1	1	
0								
4	Andorra	3	1	0	0	6	0	
3								
	stripes	colours	...		saltires	quarters	sunstars	crescent
\								
0	3	5	...		0	0	1	0
1	0	3	...		0	0	1	0
2	0	3	...		0	0	1	1
3	0	5	...		0	0	0	0
4	0	3	...		0	0	0	0
	triangle	icon	animate	text	opleftcolour	botrightcolor		
0	0	1	0	0	black	green		
1	0	0	1	0	red	red		
2	0	0	0	0	green	white		
3	1	1	1	0	blue	red		
4	0	0	0	0	blue	red		

[5 rows x 30 columns]

In [6]:

```
flags.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 194 entries, 0 to 193
Data columns (total 30 columns):
country          194 non-null object
landmass         194 non-null int64
zone             194 non-null int64
area             194 non-null int64
population       194 non-null int64
language         194 non-null int64
religion         194 non-null int64
bars             194 non-null int64
stripes          194 non-null int64
colours          194 non-null int64
red              194 non-null int64
green            194 non-null int64
blue             194 non-null int64
gold             194 non-null int64
white            194 non-null int64
black            194 non-null int64
orange           194 non-null int64
dominantcolour   194 non-null object
circles          194 non-null int64
crosses          194 non-null int64
saltires        194 non-null int64
quarters         194 non-null int64
sunstars        194 non-null int64
crescent         194 non-null int64
traingle         194 non-null int64
icon             194 non-null int64
animate          194 non-null int64
text             194 non-null int64
topleftcolour    194 non-null object
botrightcolor    194 non-null object
dtypes: int64(26), object(4)
memory usage: 45.5+ KB
```

In [7]:

```
flags['religion'] = flags['religion'].astype(object)
```

In [8]:

```
flags.columns
```

Out[8]:

```
Index([u'country', u'landmass', u'zone', u'area', u'population', u'language',
      u'religion', u'bars', u'stripes', u'colours', u'red', u'green', u'blue',
      u'gold', u'white', u'black', u'orange', u'dominantcolour', u'circles',
      u'crosses', u'saltires', u'quarters', u'sunstars', u'crescent',
      u'triangle', u'icon', u'animate', u'text', u'topleftcolour',
      u'botrightcolor'],
      dtype='object')
```

In [9]:

```
flags=flags.rename(columns = {'triangle':'triangle'})
```

In [10]:

```
flags.columns
```

Out[10]:

```
Index([u'country', u'landmass', u'zone', u'area', u'population', u'language',
      u'religion', u'bars', u'stripes', u'colours', u'red', u'green', u'blue',
      u'gold', u'white', u'black', u'orange', u'dominantcolour', u'circles',
      u'crosses', u'saltires', u'quarters', u'sunstars', u'crescent',
      u'triangle', u'icon', u'animate', u'text', u'topleftcolour',
      u'botrightcolor'],
      dtype='object')
```

In [11]:

```
flags.head(5)
```

Out[11]:

	country	landmass	zone	area	population	language	religion	bars	stripes	colours
0	Afghanistan	5	1	648	16	10	2	0	3	5
1	Albania	3	1	29	3	6	6	0	0	3
2	Algeria	4	1	2388	20	8	2	2	0	3
3	American-Samoa	6	3	0	0	1	1	0	0	5
4	Andorra	3	1	0	0	6	0	3	0	3

5 rows × 30 columns

To check how many null values in each column

In [12]:

```
flags.isnull().sum()
```

Out[12]:

```
country          0
landmass         0
zone             0
area            0
population       0
language         0
religion         0
bars            0
stripes         0
colours         0
red             0
green           0
blue           0
gold           0
white          0
black          0
orange         0
dominantcolour  0
circles        0
crosses        0
saltires       0
quarters       0
sunstars       0
crescent       0
triangle       0
icon           0
animate        0
text           0
toleftcolour   0
botrightcolor  0
dtype: int64
```

In [13]:

```
#flags = flags[numeric_cols].copy()
flags= flags.drop(columns=['country','landmass','zone','area', 'population', 'language'], axis=1)
```

In [14]:

```
flags.head(2)
```

Out[14]:

	religion	bars	stripes	colours	red	green	blue	gold	white	black	...	saltires	quarter
0	2	0	3	5	1	1	0	1	1	1	...	0	
1	6	0	0	3	1	0	0	1	0	1	...	0	

2 rows × 24 columns

In [15]:

```
flags['religion'].value_counts()
```

Out[15]:

```
1    60
0    40
2    36
5    27
6    15
3     8
7     4
4     4
```

Name: religion, dtype: int64

In [16]:

```
flags['bars'].value_counts()
```

Out[16]:

```
0    159
3     21
2      7
1      6
5      1
```

Name: bars, dtype: int64

In [17]:

```
flags['stripes'].value_counts()
```

Out[17]:

0	110
3	44
2	16
5	12
1	4
9	2
14	1
13	1
11	1
7	1
6	1
4	1

Name: stripes, dtype: int64

In [18]:

```
flags['colours'].value_counts()
```

Out[18]:

3	75
4	44
2	42
5	15
6	9
7	7
8	1
1	1

Name: colours, dtype: int64

In [19]:

```
flags['red'].value_counts()
```

Out[19]:

1	153
0	41

Name: red, dtype: int64

In [20]:

```
flags['green'].value_counts()
```

Out[20]:

0	103
1	91

Name: green, dtype: int64

In [21]:

```
flags['blue'].value_counts()
```

Out[21]:

```
1    99
0    95
Name: blue, dtype: int64
```

In [22]:

```
flags['gold'].value_counts()
```

Out[22]:

```
0    103
1     91
Name: gold, dtype: int64
```

In [23]:

```
flags['white'].value_counts()
```

Out[23]:

```
1    146
0     48
Name: white, dtype: int64
```

In [24]:

```
flags['black'].value_counts()
```

Out[24]:

```
0    142
1     52
Name: black, dtype: int64
```

In [25]:

```
flags['orange'].value_counts()
```

Out[25]:

```
0    168
1     26
Name: orange, dtype: int64
```


In [26]:

```
flags['dominantcolour'].value_counts()
```

Out[26]:

```
red      71
blue     40
green    31
white    22
gold     19
black     5
orange    4
brown     2
Name: dominantcolour, dtype: int64
```

In [27]:

```
flags['circles'].value_counts()
```

Out[27]:

```
0    165
1     27
4      1
2      1
Name: circles, dtype: int64
```

In [28]:

```
flags['crosses'].value_counts()
```

Out[28]:

```
0    167
1     25
2      2
Name: crosses, dtype: int64
```

In [29]:

```
flags['saltires'].value_counts()
```

Out[29]:

```
0    176
1     18
Name: saltires, dtype: int64
```

In [30]:

```
flags['quarters'].value_counts()
```

Out[30]:

```
0    168
1     25
4      1
Name: quarters, dtype: int64
```

In [31]:

```
flags['sunstars'].value_counts()
```

Out[31]:

```
0    114
1     50
5      7
2      6
4      5
7      2
6      2
3      2
50     1
22     1
15     1
14     1
10     1
9      1
```

Name: sunstars, dtype: int64

In [32]:

```
flags['crescent'].value_counts()
```

Out[32]:

```
0    183
1     11
```

Name: crescent, dtype: int64

In [33]:

```
flags['triangle'].value_counts()
```

Out[33]:

```
0    167
1     27
```

Name: triangle, dtype: int64

In [34]:

```
flags['icon'].value_counts()
```

Out[34]:

```
0    145
1     49
```

Name: icon, dtype: int64

In [35]:

```
flags['animate'].value_counts()
```

Out[35]:

```
0    155
1     39
```

Name: animate, dtype: int64

In [36]:

```
flags['text'].value_counts()
```

Out[36]:

```
0    178
1     16
Name: text, dtype: int64
```

In [37]:

```
flags['topleftcolour'].value_counts()
```

Out[37]:

```
red      56
blue     43
white    41
green    32
black    12
gold      6
orange   4
Name: topleftcolour, dtype: int64
```

In [38]:

```
flags['botrightcolor'].value_counts()
```

Out[38]:

```
red      69
blue     47
green    40
white    17
black     9
gold      9
brown     2
orange    1
Name: botrightcolor, dtype: int64
```

Data Visualization

In [57]:

```
flags['religion1']=flags['religion']

flags['religion1'] = flags['religion1'].astype('category')

flags['religion1'] = pd.Series(pd.Categorical(flags['religion1'], categories=[0,1,2,3,4,5,6,7]))
flags['religion1'] = flags['religion1'].cat.rename_categories(['Catholic', 'Other Chris
tian', 'Muslim',
                                                             'Buddhist', 'Hindu', 'Et
hnic', 'Marxist',
                                                             'Others'])

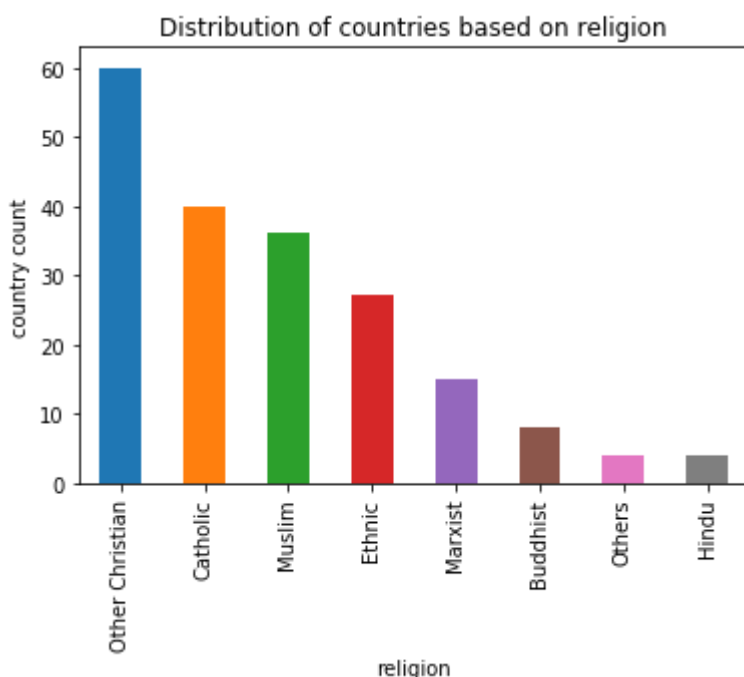
flags['religion1'].head(10)
```

Out[57]:

```
0      Muslim
1    Marxist
2      Muslim
3  Other Christian
4      Catholic
5      Ethnic
6  Other Christian
7  Other Christian
8      Catholic
9      Catholic
Name: religion1, dtype: category
Categories (8, object): [Catholic, Other Christian, Muslim, Buddhist, Hindu, Ethnic, Marxist, Others]
```

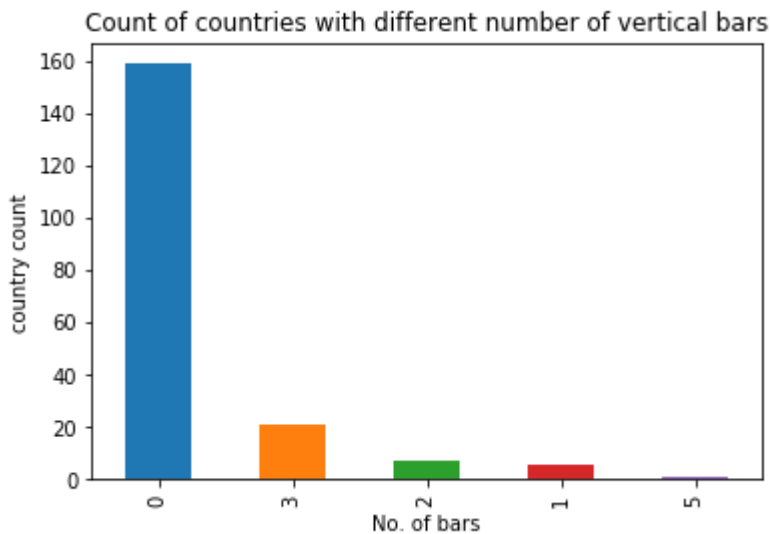
In [58]:

```
flags['religion1'].value_counts().plot(kind='bar')
plt.title('Distribution of countries based on religion')
plt.xlabel('religion')
plt.ylabel('country count')
plt.show()
```



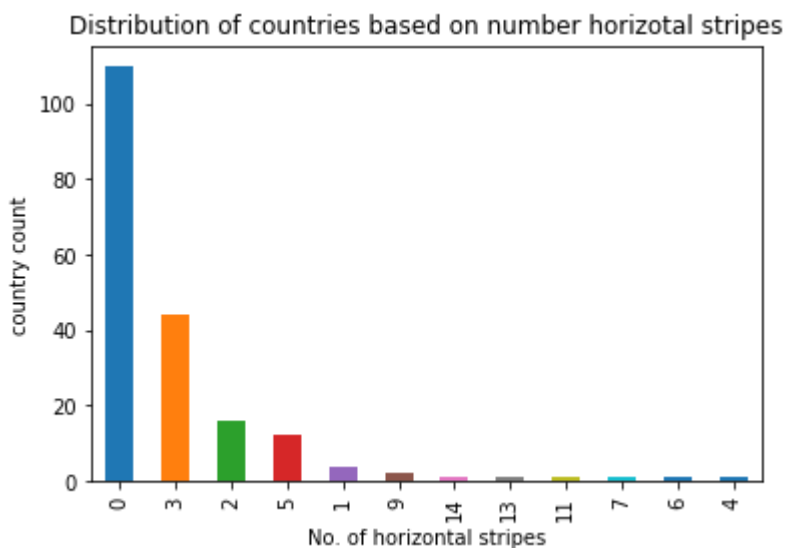
In [59]:

```
flags['bars'].value_counts().plot(kind='bar')  
plt.title('Count of countries with different number of vertical bars')  
plt.xlabel('No. of bars')  
plt.ylabel('country count')  
plt.show()
```



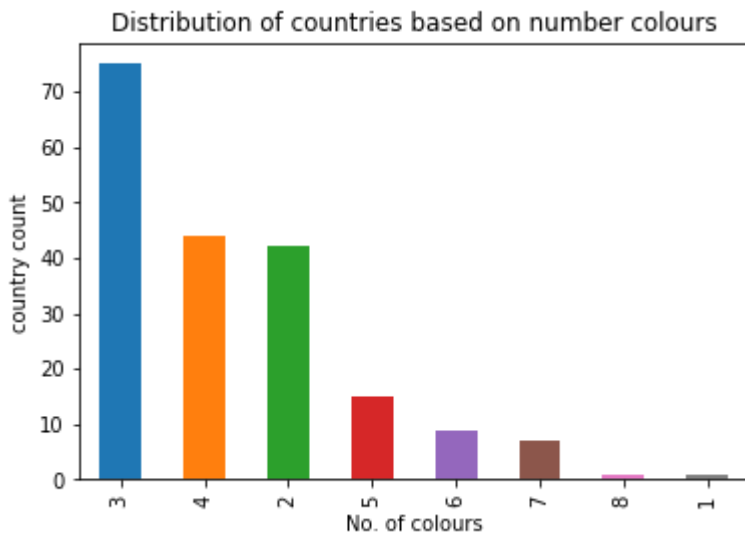
In [60]:

```
flags['stripes'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on number horizontal stripes')  
plt.xlabel('No. of horizontal stripes')  
plt.ylabel('country count')  
plt.show()
```



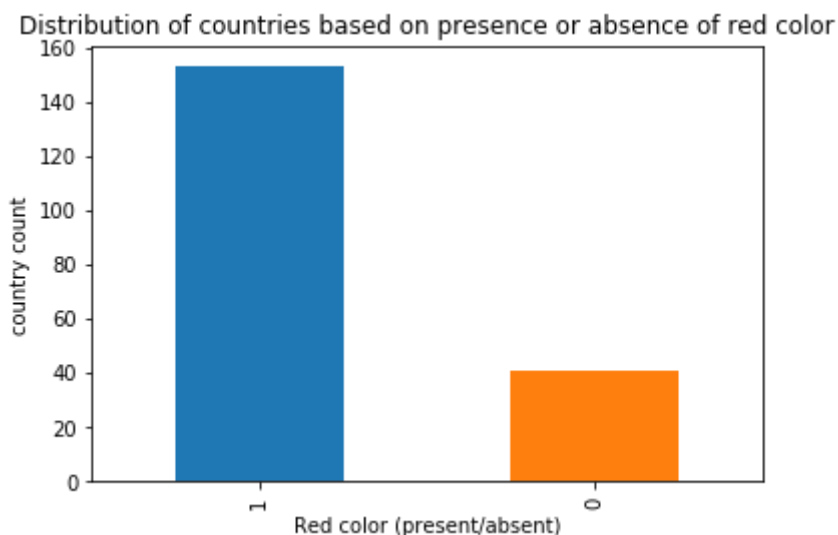
In [81]:

```
flags['colours'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on number colours')  
plt.xlabel('No. of colours')  
plt.ylabel('country count')  
plt.show()
```



In [61]:

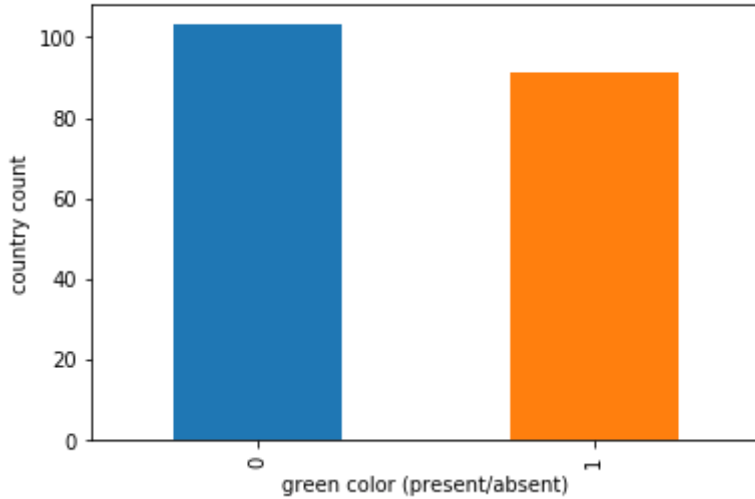
```
flags['red'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on presence or absence of red color')  
plt.xlabel('Red color (present/absent)')  
plt.ylabel('country count')  
plt.show()
```



In [62]:

```
flags['green'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on presence or absence of green color')  
plt.xlabel('green color (present/absent)')  
plt.ylabel('country count')  
plt.show()
```

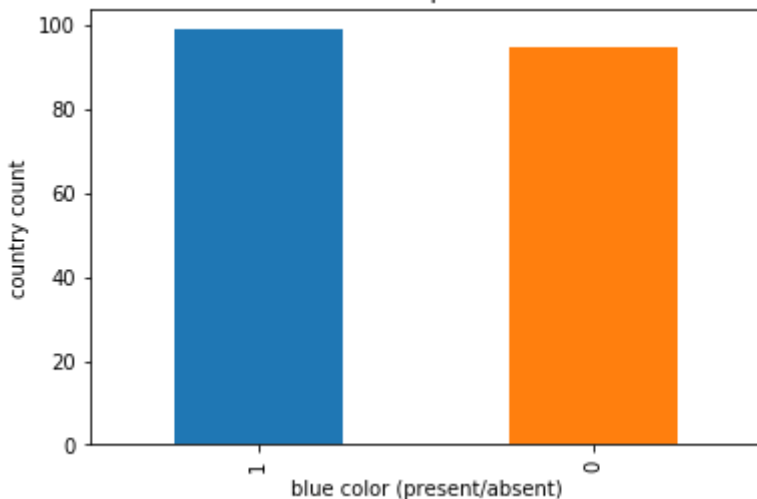
Distribution of countries based on presence or absence of green color



In [63]:

```
flags['blue'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on presence or absence of blue color')  
plt.xlabel('blue color (present/absent)')  
plt.ylabel('country count')  
plt.show()
```

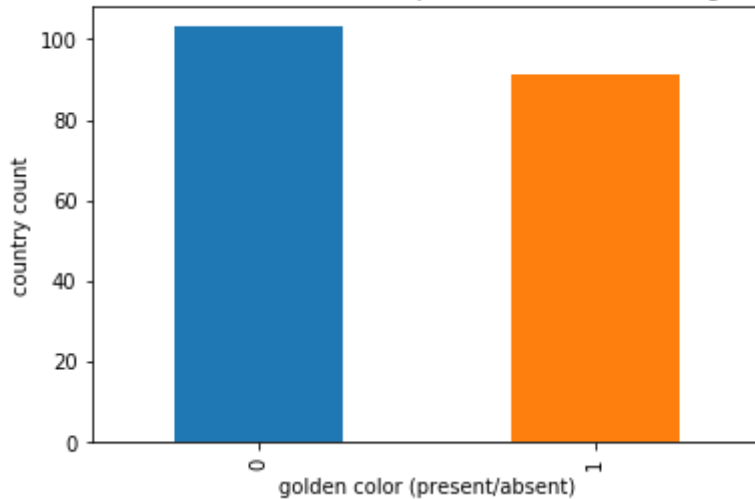
Distribution of countries based on presence or absence of blue color



In [64]:

```
flags['gold'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on presence or absence of golden color')  
plt.xlabel('golden color (present/absent)')  
plt.ylabel('country count')  
plt.show()
```

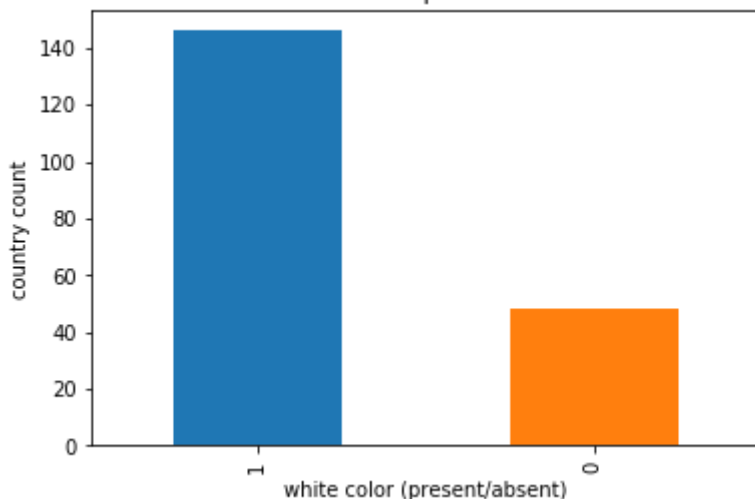
Distribution of countries based on presence or absence of golden color



In [65]:

```
flags['white'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on presence or absence of white color')  
plt.xlabel('white color (present/absent)')  
plt.ylabel('country count')  
plt.show()
```

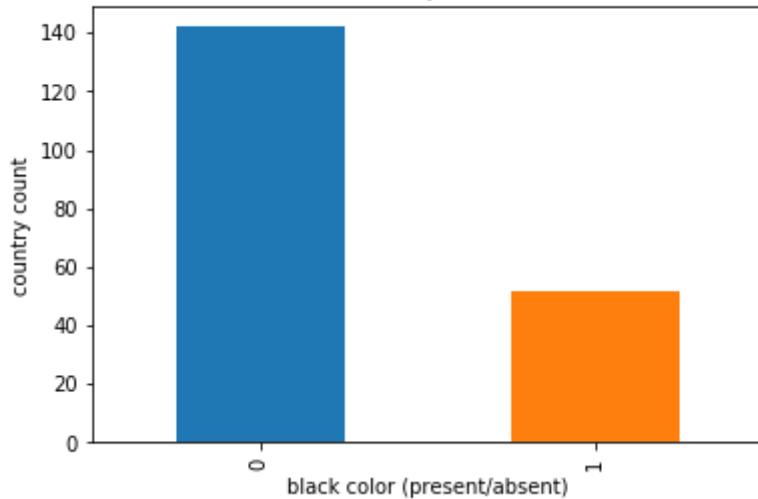
Distribution of countries based on presence or absence of white color



In [66]:

```
flags['black'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on presence or absence of black color')  
plt.xlabel('black color (present/absent)')  
plt.ylabel('country count')  
plt.show()
```

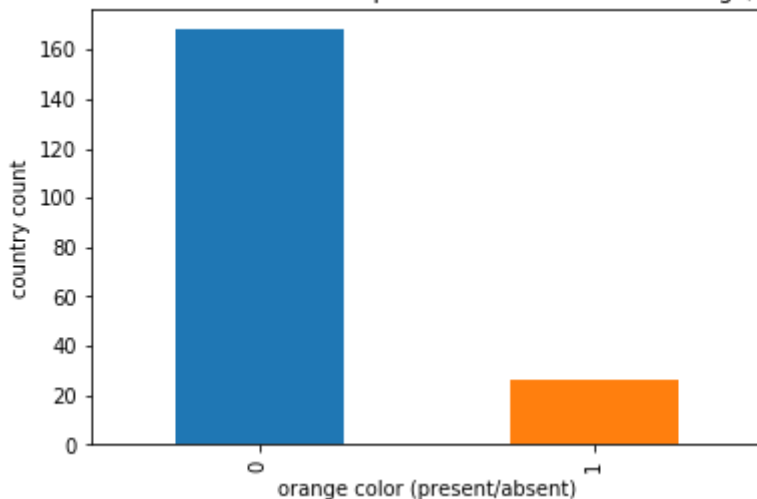
Distribution of countries based on presence or absence of black color



In [67]:

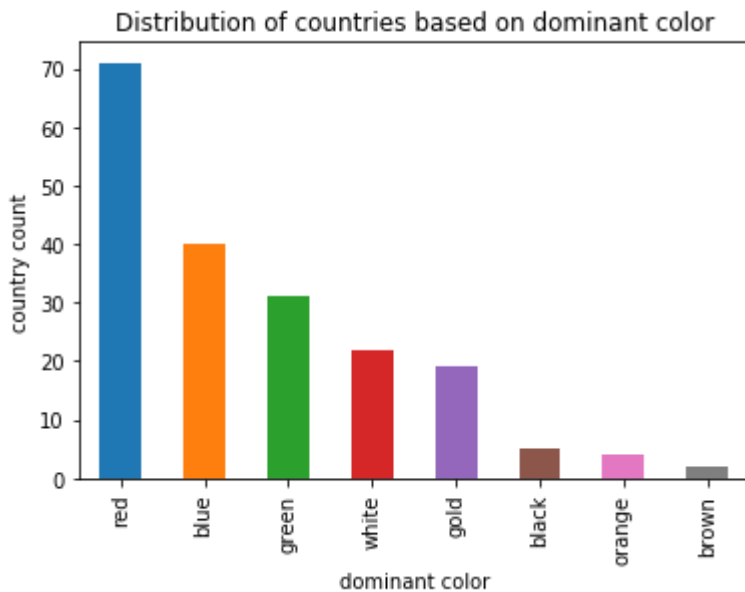
```
flags['orange'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on presence or absence of orange/brown color')  
plt.xlabel('orange color (present/absent)')  
plt.ylabel('country count')  
plt.show()
```

Distribution of countries based on presence or absence of orange/brown color



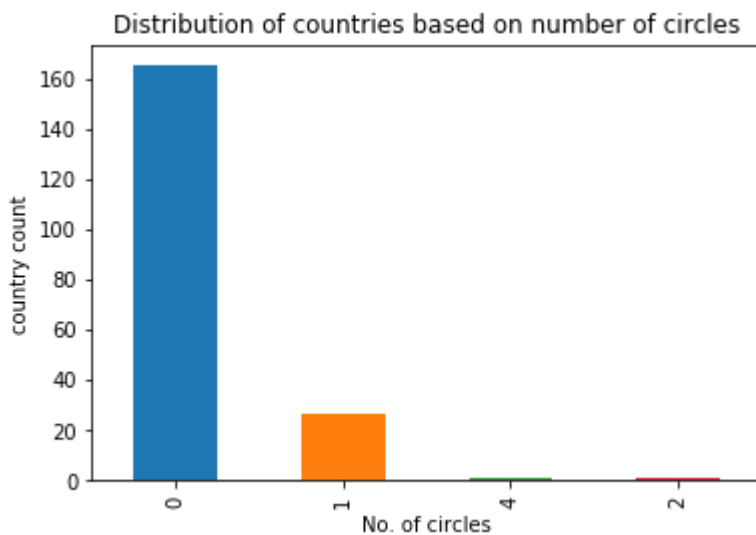
In [68]:

```
flags['dominantcolour'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on dominant color')  
plt.xlabel('dominant color')  
plt.ylabel('country count')  
plt.show()
```



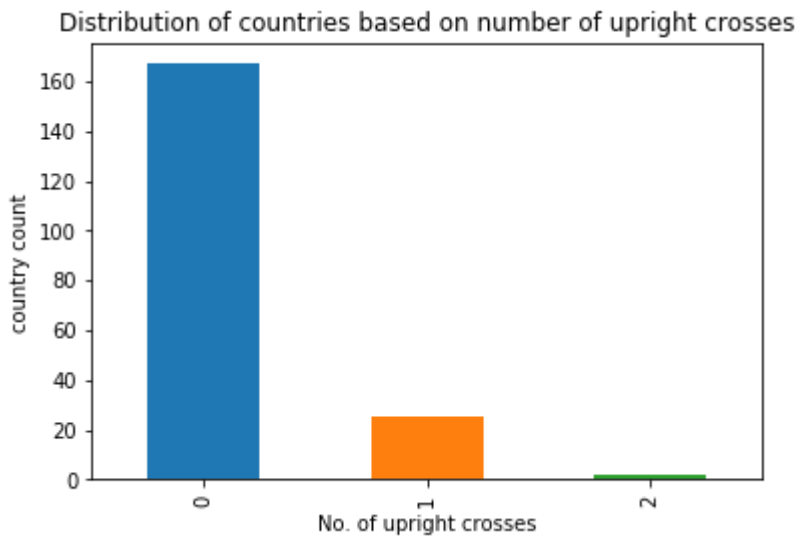
In [69]:

```
flags['circles'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on number of circles')  
plt.xlabel('No. of circles')  
plt.ylabel('country count')  
plt.show()
```



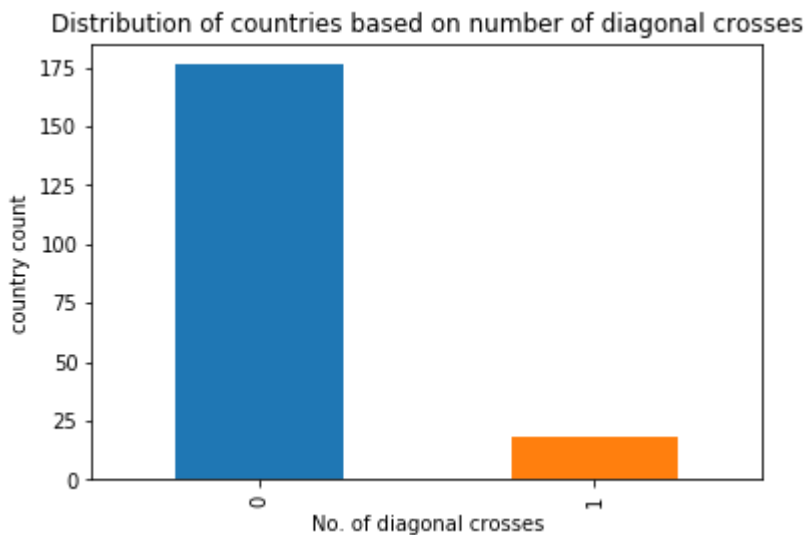
In [70]:

```
flags['crosses'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on number of upright crosses')  
plt.xlabel('No. of upright crosses')  
plt.ylabel('country count')  
plt.show()
```



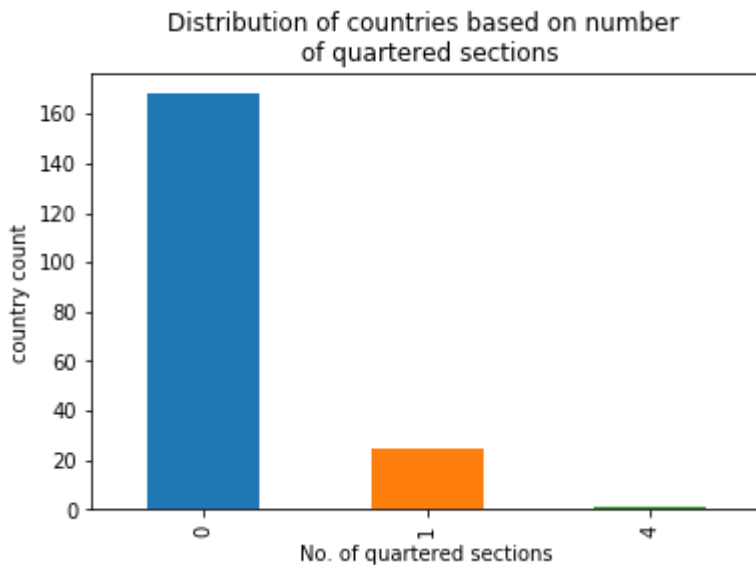
In [71]:

```
flags['saltires'].value_counts().plot(kind='bar')  
plt.title('Distribution of countries based on number of diagonal crosses')  
plt.xlabel('No. of diagonal crosses')  
plt.ylabel('country count')  
plt.show()
```



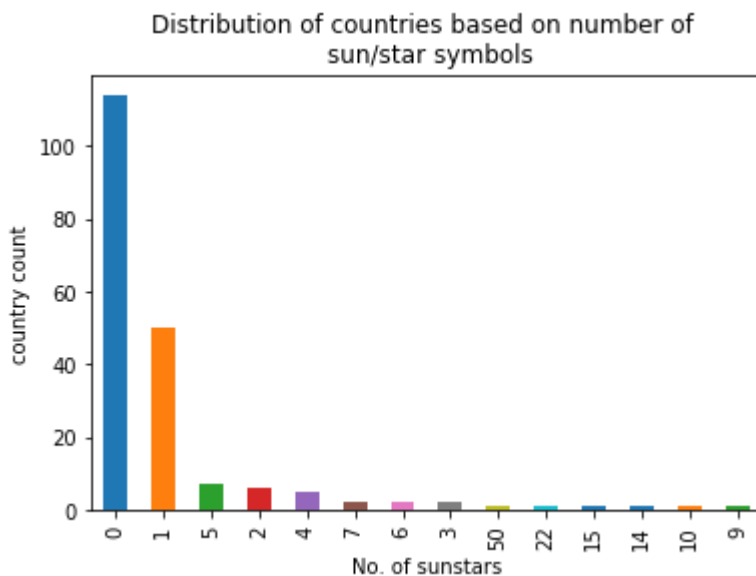
In [72]:

```
flags['quarters'].value_counts().plot(kind='bar')
plt.title('Distribution of countries based on number \n of quartered sections')
plt.xlabel('No. of quartered sections')
plt.ylabel('country count')
plt.show()
```



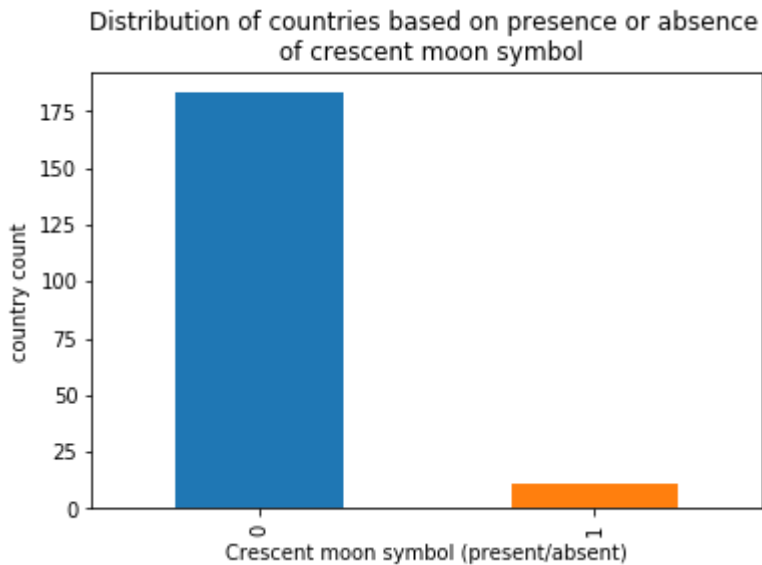
In [73]:

```
flags['sunstars'].value_counts().plot(kind='bar')
plt.title('Distribution of countries based on number of \n sun/star symbols')
plt.xlabel('No. of sunstars')
plt.ylabel('country count')
plt.show()
```



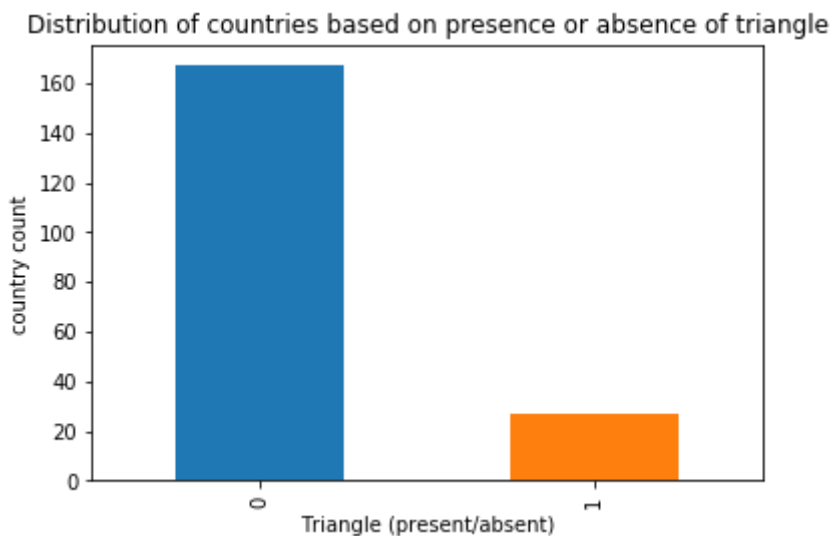
In [74]:

```
flags['crescent'].value_counts().plot(kind='bar')
plt.title('Distribution of countries based on presence or absence \n of crescent moon s
ymbol')
plt.xlabel('Crescent moon symbol (present/absent)')
plt.ylabel('country count')
plt.show()
```



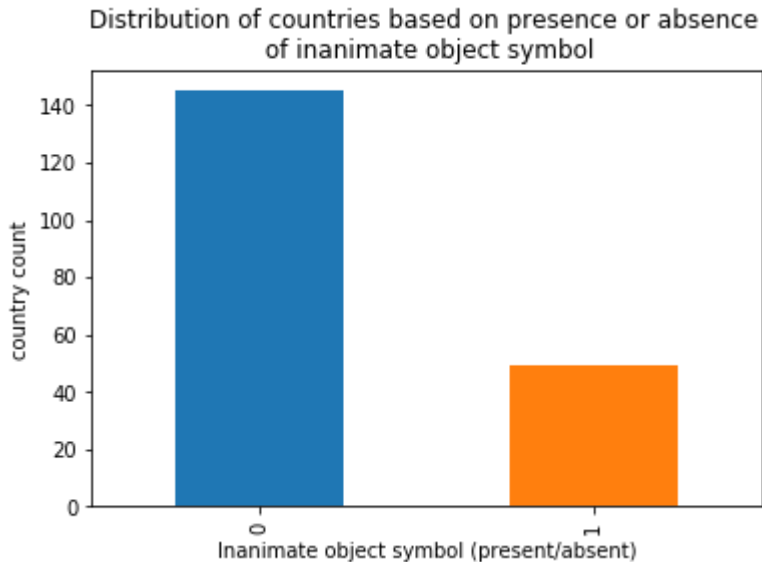
In [75]:

```
flags['triangle'].value_counts().plot(kind='bar')
plt.title('Distribution of countries based on presence or absence of triangle')
plt.xlabel('Triangle (present/absent)')
plt.ylabel('country count')
plt.show()
```



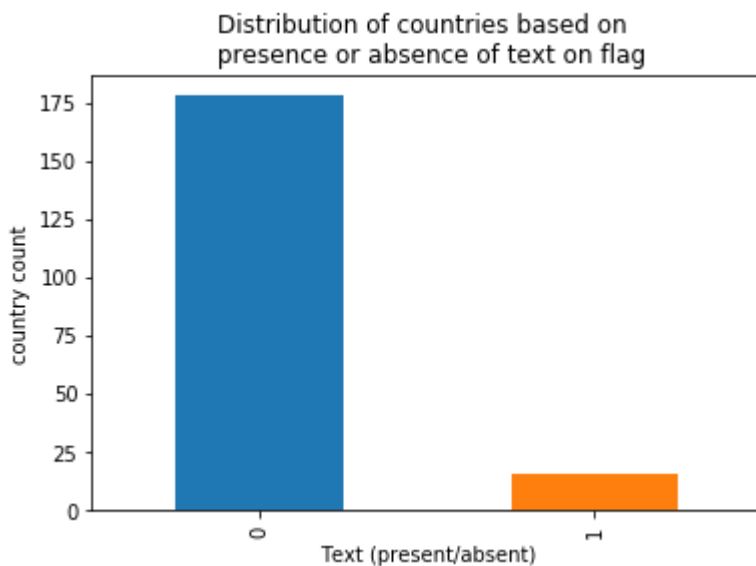
In [76]:

```
flags['icon'].value_counts().plot(kind='bar')
plt.title('Distribution of countries based on presence or absence \n of inanimate object symbol')
plt.xlabel('Inanimate object symbol (present/absent)')
plt.ylabel('country count')
plt.show()
```



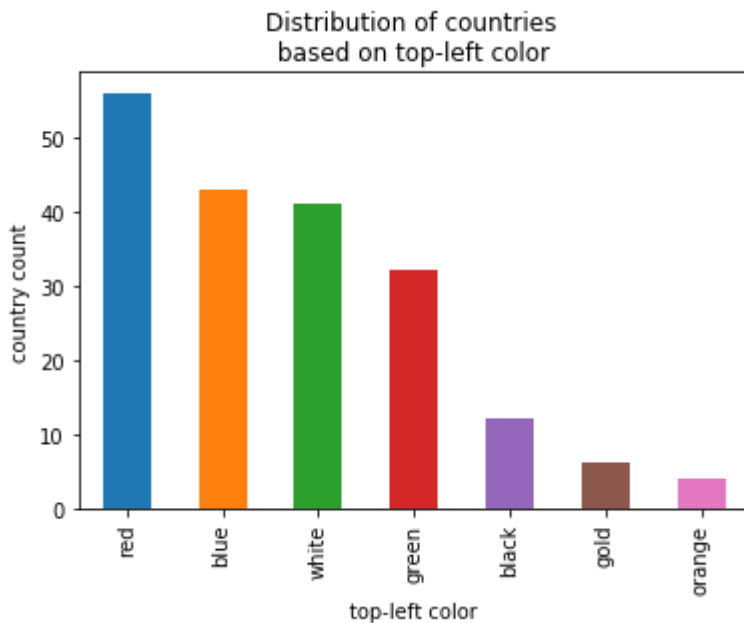
In [77]:

```
flags['text'].value_counts().plot(kind='bar')
plt.title('Distribution of countries based on \n presence or absence of text on flag')
plt.xlabel('Text (present/absent)')
plt.ylabel('country count')
plt.show()
```



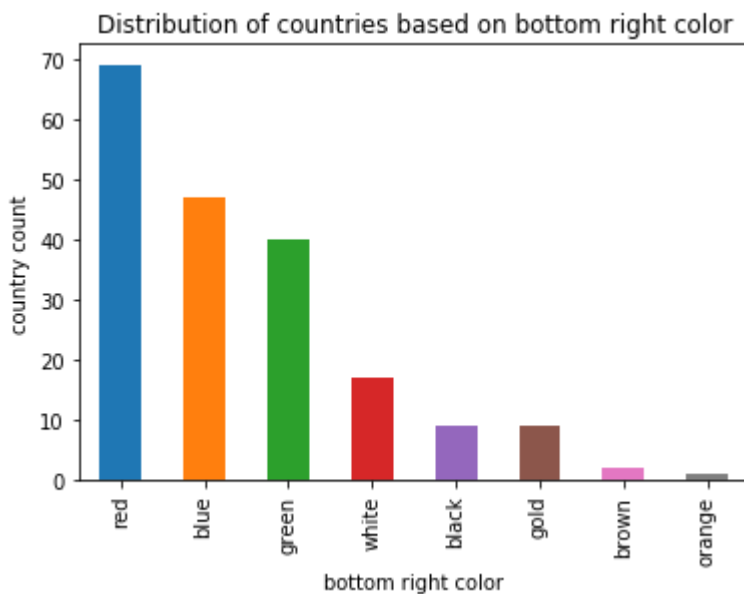
In [78]:

```
flags['topleftcolour'].value_counts().plot(kind='bar')
plt.title('Distribution of countries \n based on top-left color ')
plt.xlabel('top-left color')
plt.ylabel('country count')
plt.show()
```



In [79]:

```
flags['botrightcolor'].value_counts().plot(kind='bar')
plt.title('Distribution of countries based on bottom right color')
plt.xlabel('bottom right color')
plt.ylabel('country count')
plt.show()
```



Bivariate Relationships

In [80]:

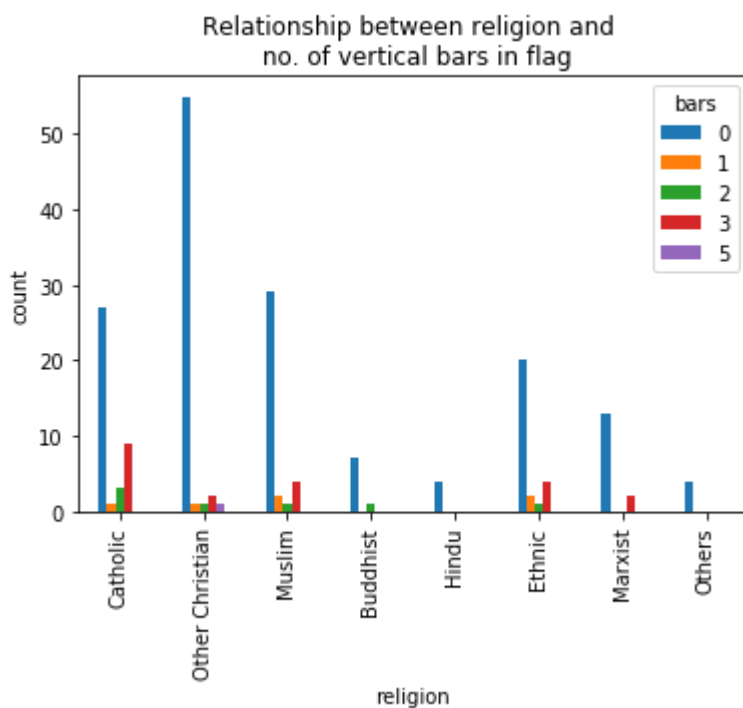
```

print (flags.groupby(['religion1', 'bars']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['bars']).plot.bar()
plt.title('Relationship between religion and \n no. of vertical bars in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()

```

bars	0	1	2	3	5
religion1					
Catholic	27	1	3	9	0
Other Christian	55	1	1	2	1
Muslim	29	2	1	4	0
Buddhist	7	0	1	0	0
Hindu	4	0	0	0	0
Ethnic	20	2	1	4	0
Marxist	13	0	0	2	0
Others	4	0	0	0	0

<matplotlib.figure.Figure at 0xdb83b00>



In [63]:

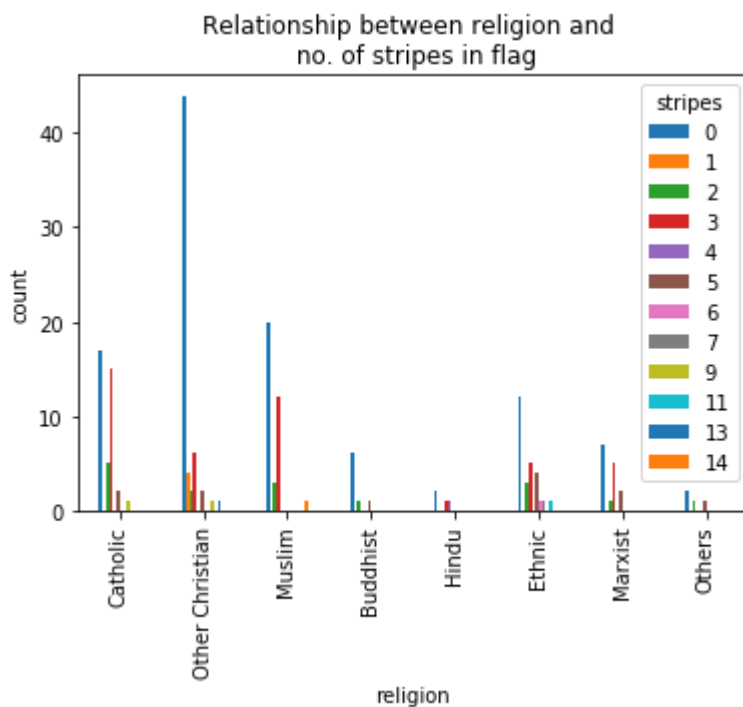
```

print (flags.groupby(['religion1', 'stripes']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['stripes']).plot.bar()
plt.title('Relationship between religion and \n no. of stripes in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()

```

stripes	0	1	2	3	4	5	6	7	9	11	13	14
religion1												
Catholic	17	0	5	15	0	2	0	0	1	0	0	0
Other Christian	44	4	2	6	0	2	0	0	1	0	1	0
Muslim	20	0	3	12	0	0	0	0	0	0	0	1
Buddhist	6	0	1	0	0	1	0	0	0	0	0	0
Hindu	2	0	0	1	1	0	0	0	0	0	0	0
Ethnic	12	0	3	5	0	4	1	1	0	1	0	0
Marxist	7	0	1	5	0	2	0	0	0	0	0	0
Others	2	0	1	0	0	1	0	0	0	0	0	0

<matplotlib.figure.Figure at 0xe7da0b8>

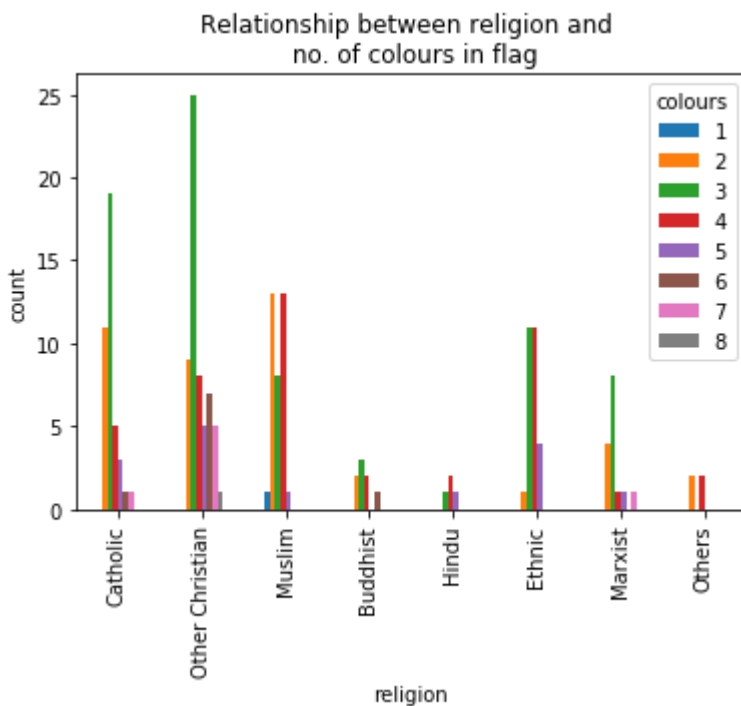


In [64]:

```
print (flags.groupby(['religion1', 'colours']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['colours']).plot.bar()
plt.title('Relationship between religion and \n no. of colours in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

colours	1	2	3	4	5	6	7	8
religion1								
Catholic	0	11	19	5	3	1	1	0
Other Christian	0	9	25	8	5	7	5	1
Muslim	1	13	8	13	1	0	0	0
Buddhist	0	2	3	2	0	1	0	0
Hindu	0	0	1	2	1	0	0	0
Ethnic	0	1	11	11	4	0	0	0
Marxist	0	4	8	1	1	0	1	0
Others	0	2	0	2	0	0	0	0

<matplotlib.figure.Figure at 0xe2df470>

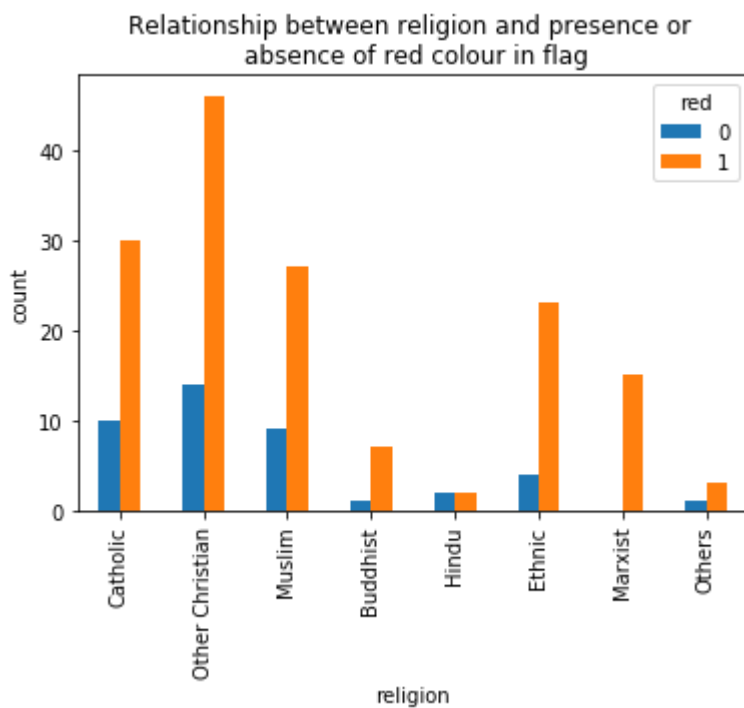


In [65]:

```
print (flags.groupby(['religion1','red']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['red']).plot.bar()
plt.title('Relationship between religion and presence or \n absence of red colour in fl
ag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

red	0	1
religion1		
Catholic	10	30
Other Christian	14	46
Muslim	9	27
Buddhist	1	7
Hindu	2	2
Ethnic	4	23
Marxist	0	15
Others	1	3

<matplotlib.figure.Figure at 0xdd61ba8>

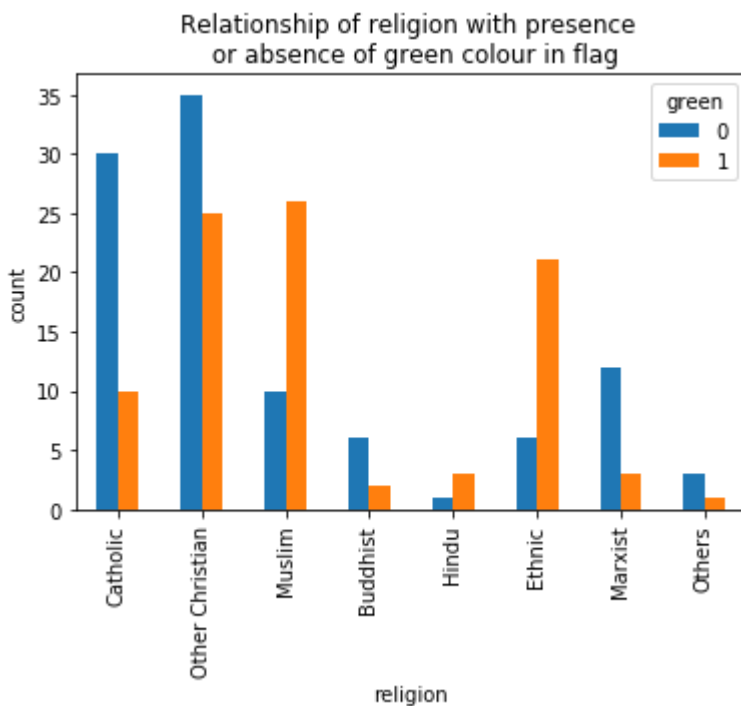


In [66]:

```
print (flags.groupby(['religion1', 'green']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['green']).plot.bar()
plt.title('Relationship of religion with presence \n or absence of green colour in fla
g')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

green	0	1
religion1		
Catholic	30	10
Other Christian	35	25
Muslim	10	26
Buddhist	6	2
Hindu	1	3
Ethnic	6	21
Marxist	12	3
Others	3	1

<matplotlib.figure.Figure at 0xd0c2550>

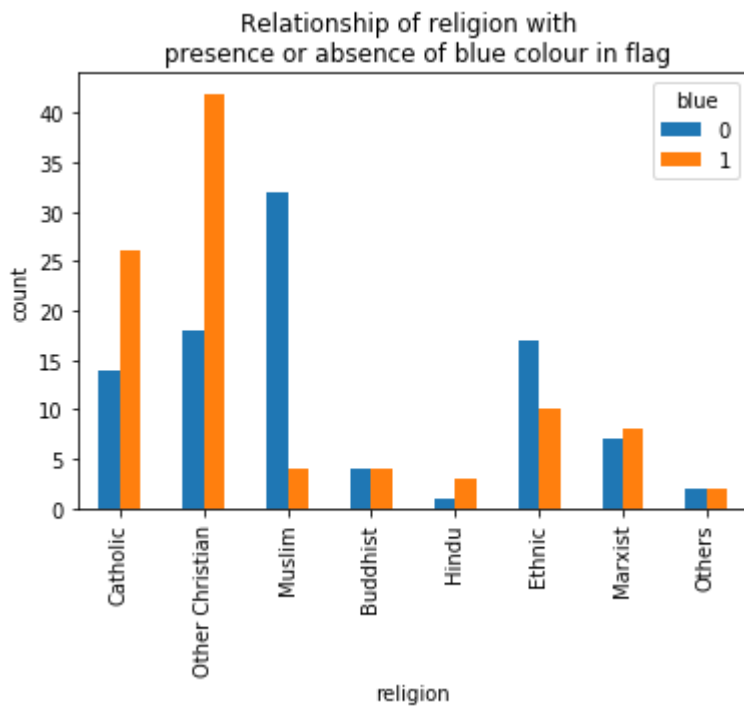


In [67]:

```
print (flags.groupby(['religion1', 'blue']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['blue']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of blue colour in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

blue	0	1
religion1		
Catholic	14	26
Other Christian	18	42
Muslim	32	4
Buddhist	4	4
Hindu	1	3
Ethnic	17	10
Marxist	7	8
Others	2	2

<matplotlib.figure.Figure at 0xcc0ad30>

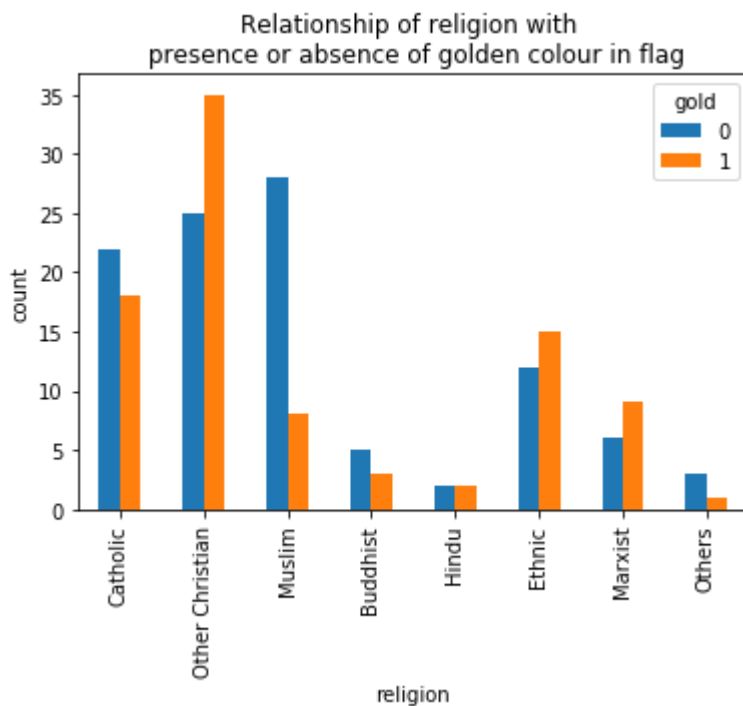


In [68]:

```
print (flags.groupby(['religion1', 'gold']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['gold']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of golden colour in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

gold	0	1
religion1		
Catholic	22	18
Other Christian	25	35
Muslim	28	8
Buddhist	5	3
Hindu	2	2
Ethnic	12	15
Marxist	6	9
Others	3	1

<matplotlib.figure.Figure at 0xcd06e48>

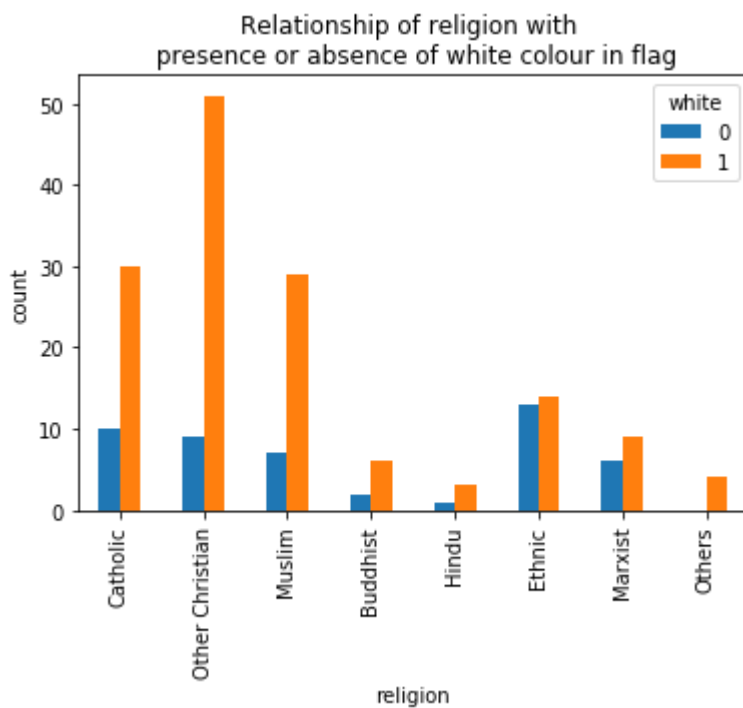


In [69]:

```
print (flags.groupby(['religion1','white']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['white']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of white colour in fla
g')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

white	0	1
religion1		
Catholic	10	30
Other Christian	9	51
Muslim	7	29
Buddhist	2	6
Hindu	1	3
Ethnic	13	14
Marxist	6	9
Others	0	4

<matplotlib.figure.Figure at 0xed923c8>

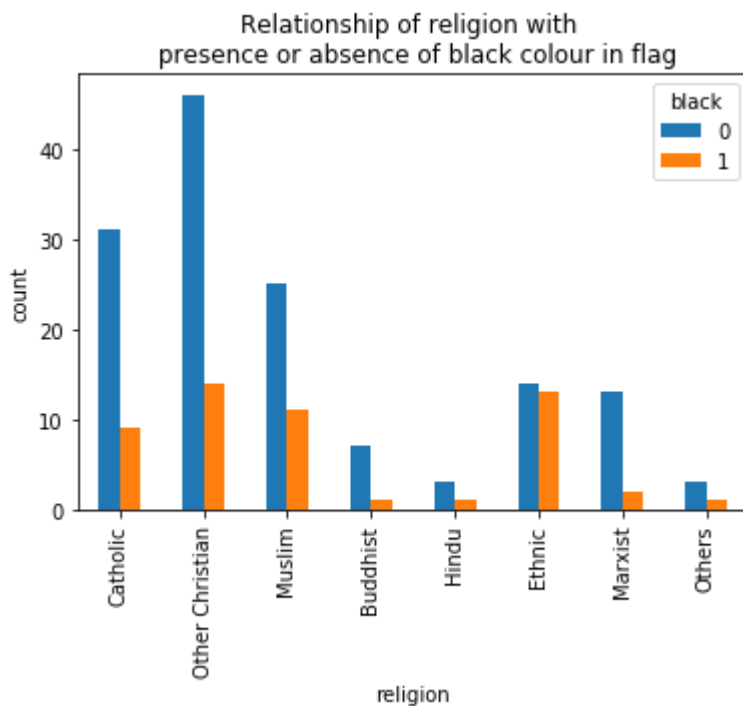


In [70]:

```
print (flags.groupby(['religion1', 'black']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['black']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of black colour in fla
g')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

black	0	1
religion1		
Catholic	31	9
Other Christian	46	14
Muslim	25	11
Buddhist	7	1
Hindu	3	1
Ethnic	14	13
Marxist	13	2
Others	3	1

<matplotlib.figure.Figure at 0xec4c358>

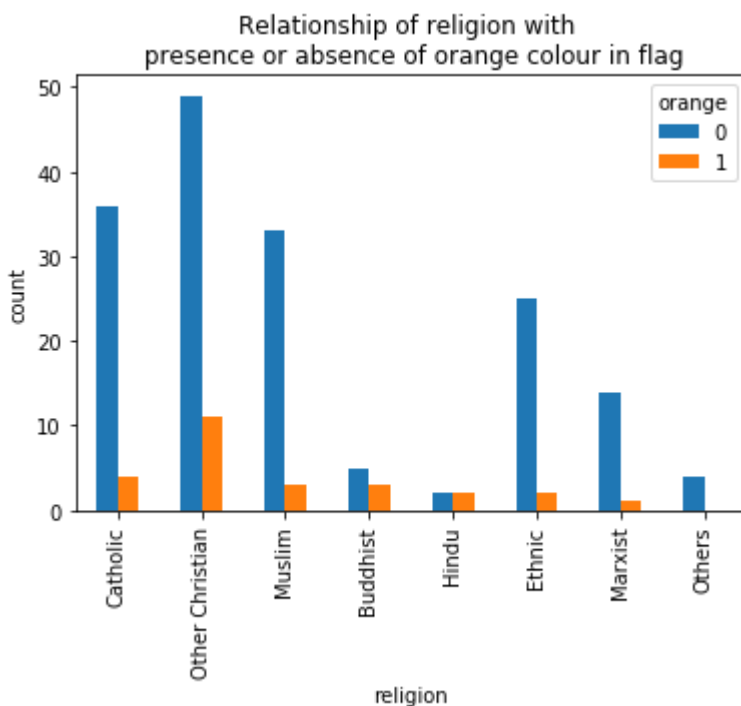


In [71]:

```
print (flags.groupby(['religion1', 'orange']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['orange']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of orange colour in fla
g')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

orange	0	1
religion1		
Catholic	36	4
Other Christian	49	11
Muslim	33	3
Buddhist	5	3
Hindu	2	2
Ethnic	25	2
Marxist	14	1
Others	4	0

<matplotlib.figure.Figure at 0xec60d30>



In [72]:

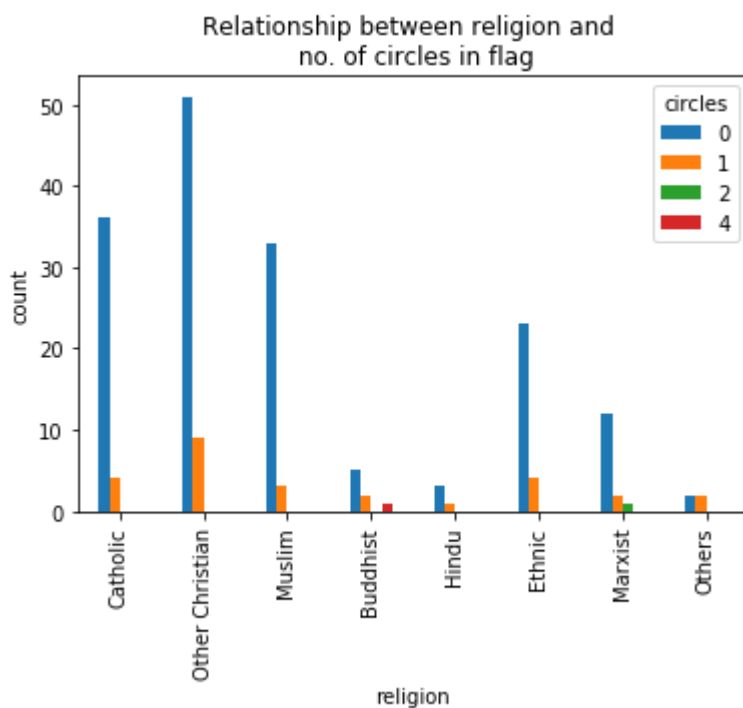
```

print (flags.groupby(['religion1', 'circles']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['circles']).plot.bar()
plt.title('Relationship between religion and \n no. of circles in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()

```

circles	0	1	2	4
religion1				
Catholic	36	4	0	0
Other Christian	51	9	0	0
Muslim	33	3	0	0
Buddhist	5	2	0	1
Hindu	3	1	0	0
Ethnic	23	4	0	0
Marxist	12	2	1	0
Others	2	2	0	0

<matplotlib.figure.Figure at 0xea33c50>

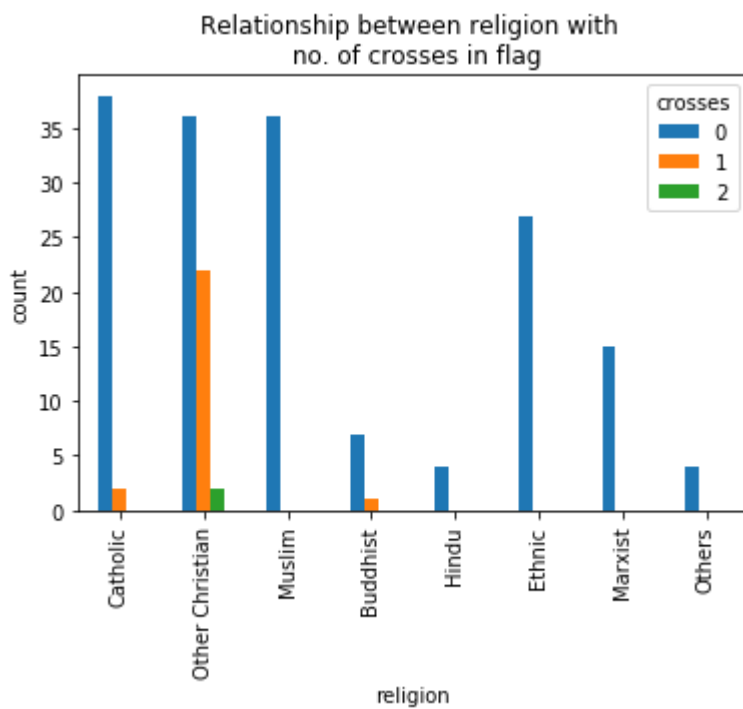


In [73]:

```
print (flags.groupby(['religion1', 'crosses']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['crosses']).plot.bar()
plt.title('Relationship between religion with \n no. of crosses in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

crosses	0	1	2
religion1			
Catholic	38	2	0
Other Christian	36	22	2
Muslim	36	0	0
Buddhist	7	1	0
Hindu	4	0	0
Ethnic	27	0	0
Marxist	15	0	0
Others	4	0	0

<matplotlib.figure.Figure at 0xdcd81d0>

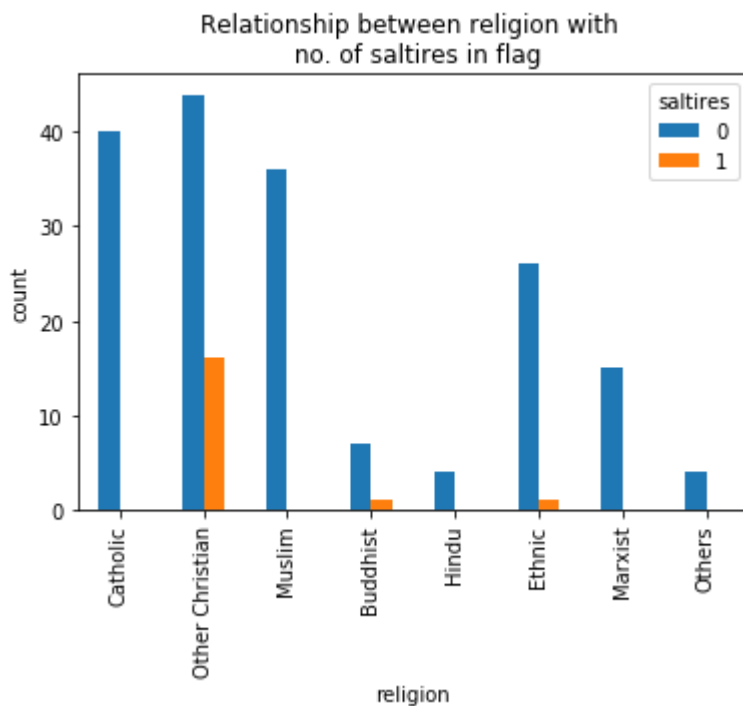


In [74]:

```
print (flags.groupby(['religion1', 'saltires']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['saltires']).plot.bar()
plt.title('Relationship between religion with \n no. of saltires in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

saltires	0	1
religion1		
Catholic	40	0
Other Christian	44	16
Muslim	36	0
Buddhist	7	1
Hindu	4	0
Ethnic	26	1
Marxist	15	0
Others	4	0

<matplotlib.figure.Figure at 0xe1b3320>

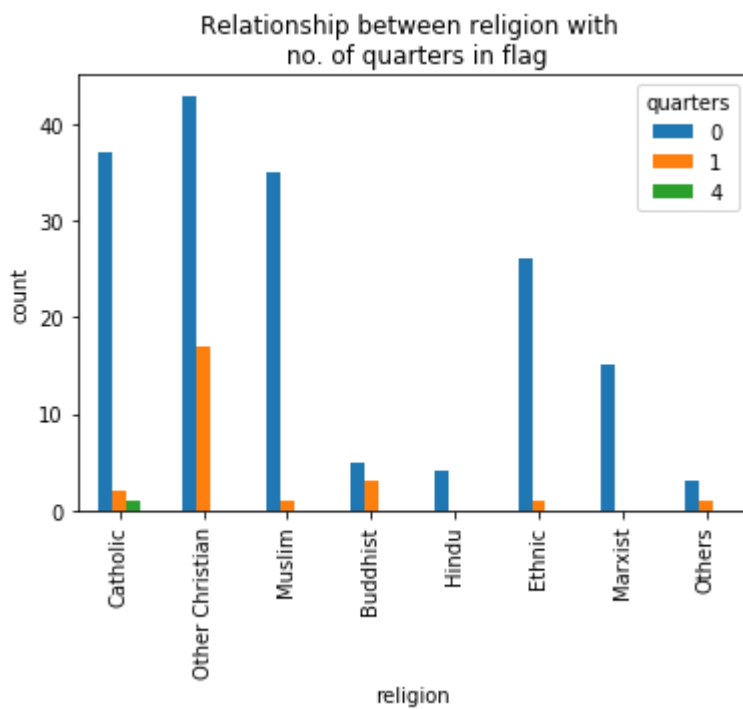


In [75]:

```
print (flags.groupby(['religion1', 'quarters']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['quarters']).plot.bar()
plt.title('Relationship between religion with \n no. of quarters in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

quarters	0	1	4
religion1			
Catholic	37	2	1
Other Christian	43	17	0
Muslim	35	1	0
Buddhist	5	3	0
Hindu	4	0	0
Ethnic	26	1	0
Marxist	15	0	0
Others	3	1	0

<matplotlib.figure.Figure at 0xe587898>

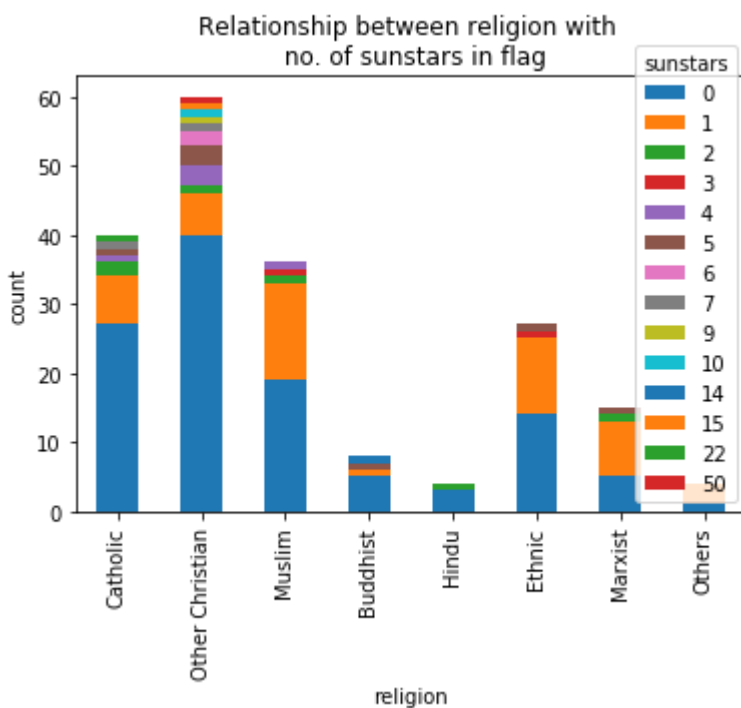


In [76]:

```
print (flags.groupby(['religion1', 'sunstars']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['sunstars']).plot.bar(stacked=True)
plt.title('Relationship between religion with \n no. of sunstars in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

sunstars	0	1	2	3	4	5	6	7	9	10	14	15	22	50
religion1														
Catholic	27	7	2	0	1	1	0	1	0	0	0	0	1	0
Other Christian	40	6	1	0	3	3	2	1	1	1	0	1	0	1
Muslim	19	14	1	1	1	0	0	0	0	0	0	0	0	0
Buddhist	5	1	0	0	0	1	0	0	0	0	1	0	0	0
Hindu	3	0	1	0	0	0	0	0	0	0	0	0	0	0
Ethnic	14	11	0	1	0	1	0	0	0	0	0	0	0	0
Marxist	5	8	1	0	0	1	0	0	0	0	0	0	0	0
Others	1	3	0	0	0	0	0	0	0	0	0	0	0	0

<matplotlib.figure.Figure at 0xdbed400>

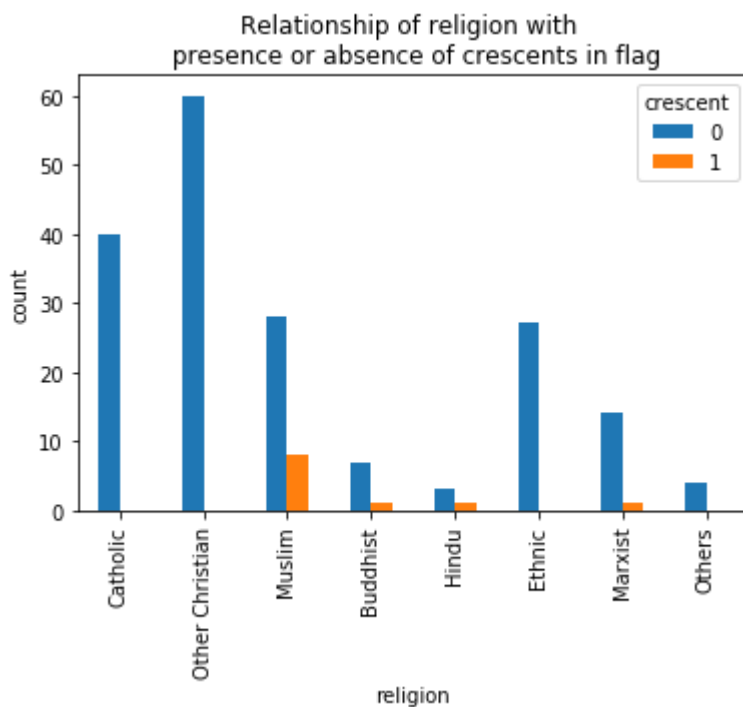


In [77]:

```
print (flags.groupby(['religion1', 'crescent']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['crescent']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of crescents in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

crescent	0	1
religion1		
Catholic	40	0
Other Christian	60	0
Muslim	28	8
Buddhist	7	1
Hindu	3	1
Ethnic	27	0
Marxist	14	1
Others	4	0

<matplotlib.figure.Figure at 0xe2f2320>

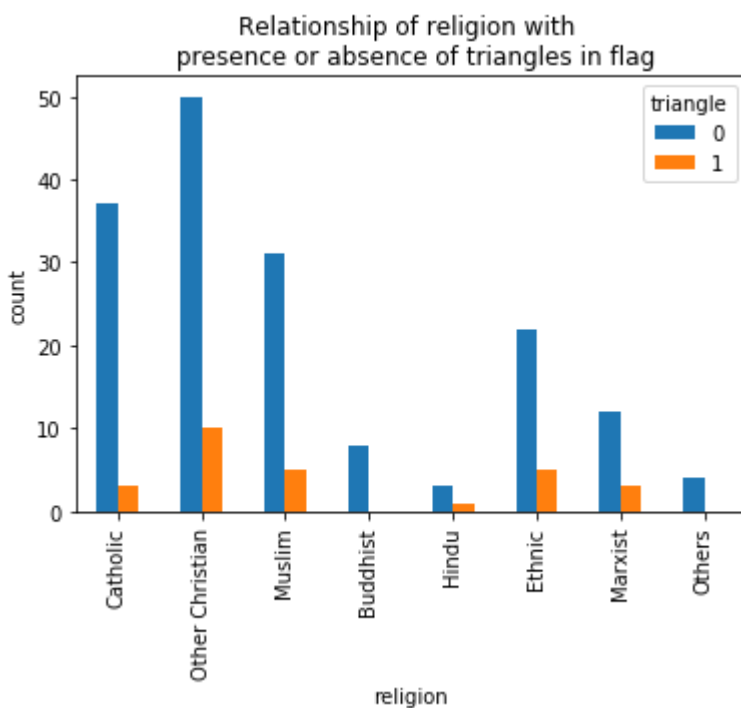


In [78]:

```
print (flags.groupby(['religion1', 'triangle']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['triangle']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of triangles in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

triangle	0	1
religion1		
Catholic	37	3
Other Christian	50	10
Muslim	31	5
Buddhist	8	0
Hindu	3	1
Ethnic	22	5
Marxist	12	3
Others	4	0

<matplotlib.figure.Figure at 0xe563ef0>

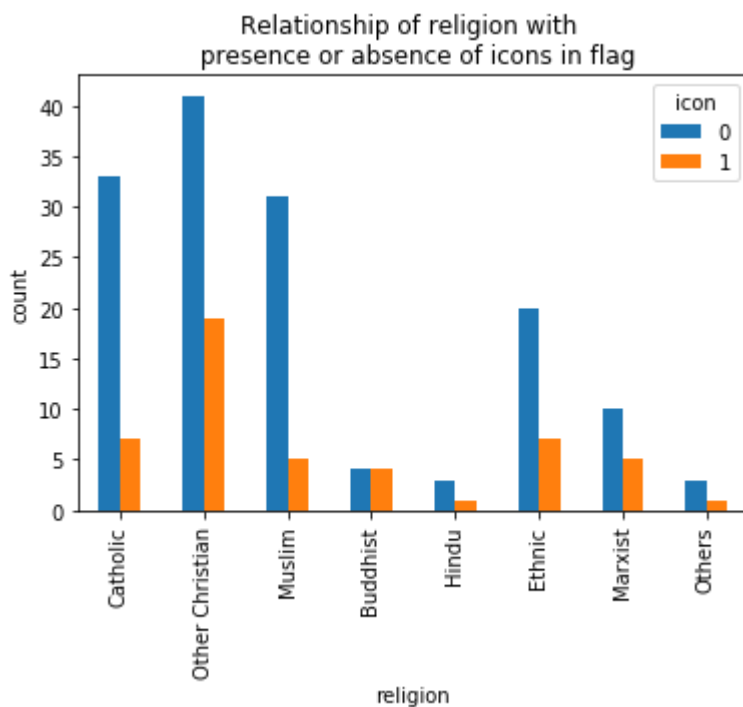


In [79]:

```
print (flags.groupby(['religion1', 'icon']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['icon']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of icons in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

icon	0	1
religion1		
Catholic	33	7
Other Christian	41	19
Muslim	31	5
Buddhist	4	4
Hindu	3	1
Ethnic	20	7
Marxist	10	5
Others	3	1

<matplotlib.figure.Figure at 0xe2f2780>

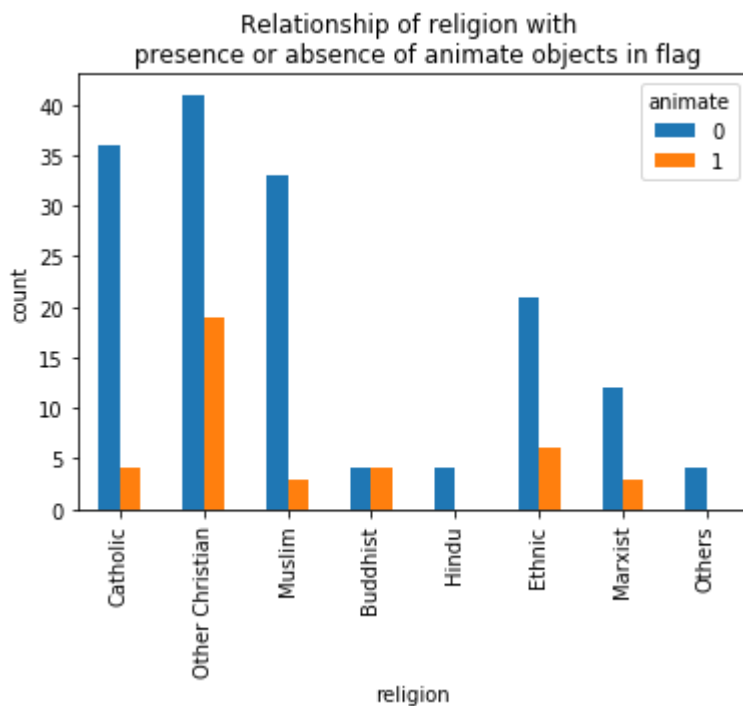


In [80]:

```
print (flags.groupby(['religion1', 'animate']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['animate']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of animate objects in f
lag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

animate	0	1
religion1		
Catholic	36	4
Other Christian	41	19
Muslim	33	3
Buddhist	4	4
Hindu	4	0
Ethnic	21	6
Marxist	12	3
Others	4	0

<matplotlib.figure.Figure at 0x98dd048>

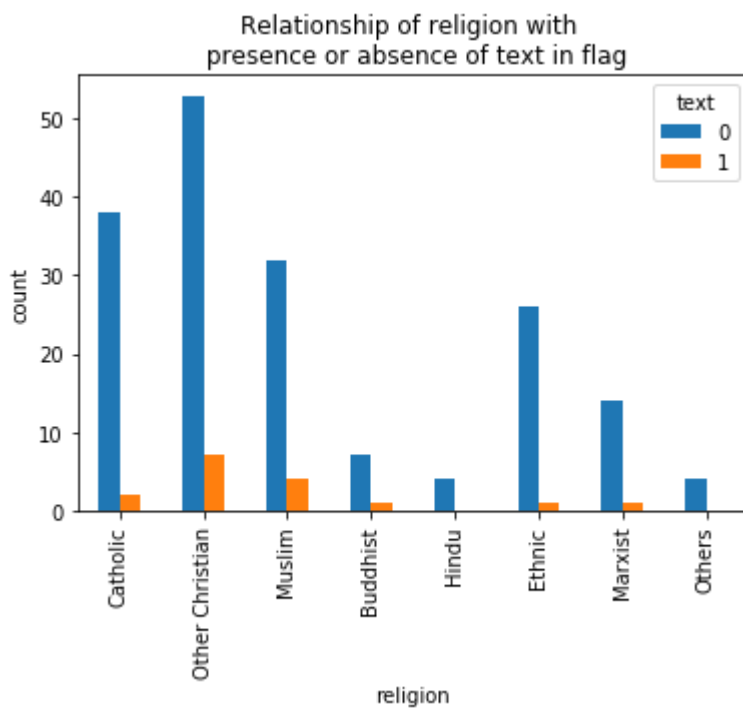


In [81]:

```
print (flags.groupby(['religion1', 'text']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['text']).plot.bar()
plt.title('Relationship of religion with \n presence or absence of text in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

text	0	1
religion1		
Catholic	38	2
Other Christian	53	7
Muslim	32	4
Buddhist	7	1
Hindu	4	0
Ethnic	26	1
Marxist	14	1
Others	4	0

<matplotlib.figure.Figure at 0xdeef6a0>

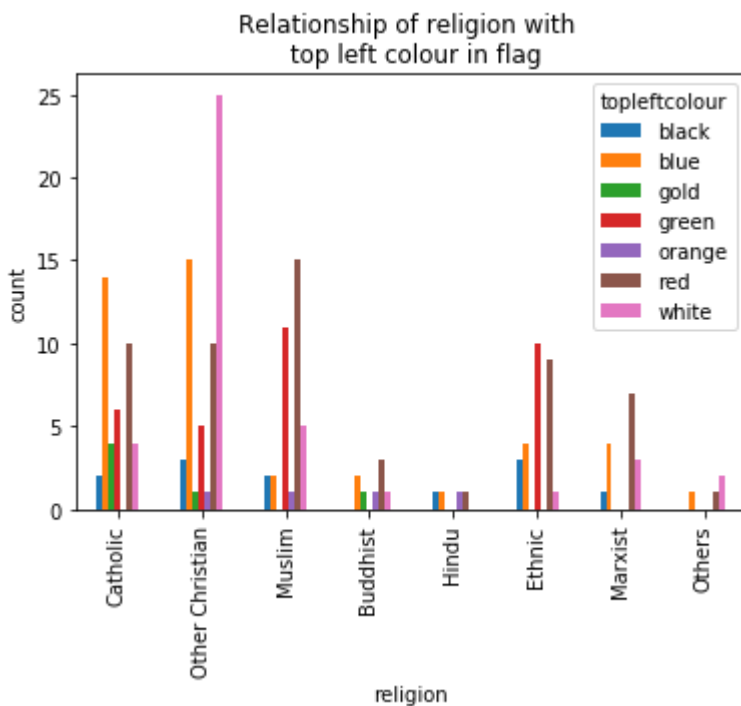


In [82]:

```
print (flags.groupby(['religion1', 'topleftcolour']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['topleftcolour']).plot.bar()
plt.title('Relationship of religion with \n top left colour in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

topleftcolour	black	blue	gold	green	orange	red	white
religion1							
Catholic	2	14	4	6	0	10	4
Other Christian	3	15	1	5	1	10	25
Muslim	2	2	0	11	1	15	5
Buddhist	0	2	1	0	1	3	1
Hindu	1	1	0	0	1	1	0
Ethnic	3	4	0	10	0	9	1
Marxist	1	4	0	0	0	7	3
Others	0	1	0	0	0	1	2

<matplotlib.figure.Figure at 0xebd98d0>

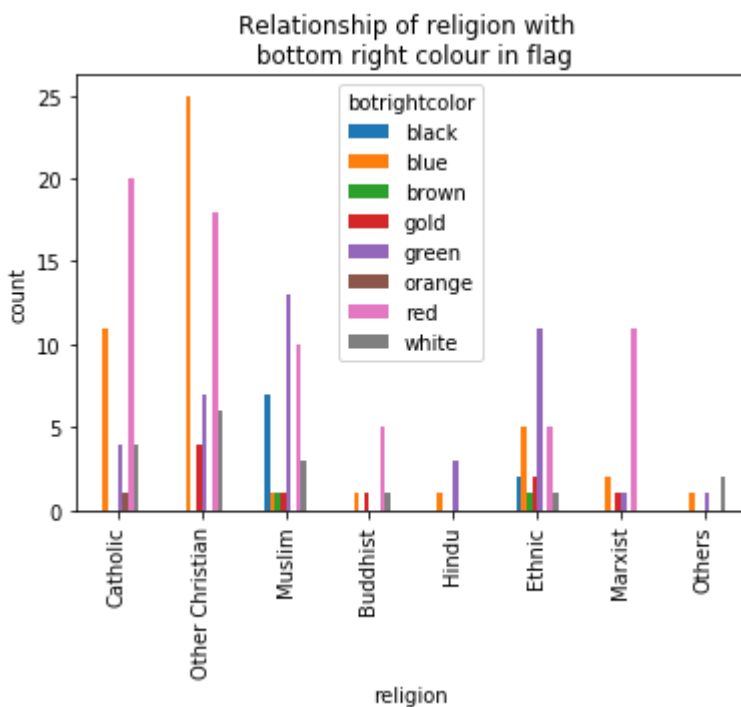


In [83]:

```
print (flags.groupby(['religion1', 'botrightcolor']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['botrightcolor']).plot.bar()
plt.title('Relationship of religion with \n bottom right colour in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

botrightcolor	black	blue	brown	gold	green	orange	red	white
religion1								
Catholic	0	11	0	0	4	1	20	4
Other Christian	0	25	0	4	7	0	18	6
Muslim	7	1	1	1	13	0	10	3
Buddhist	0	1	0	1	0	0	5	1
Hindu	0	1	0	0	3	0	0	0
Ethnic	2	5	1	2	11	0	5	1
Marxist	0	2	0	1	1	0	11	0
Others	0	1	0	0	1	0	0	2

<matplotlib.figure.Figure at 0xcc1c908>

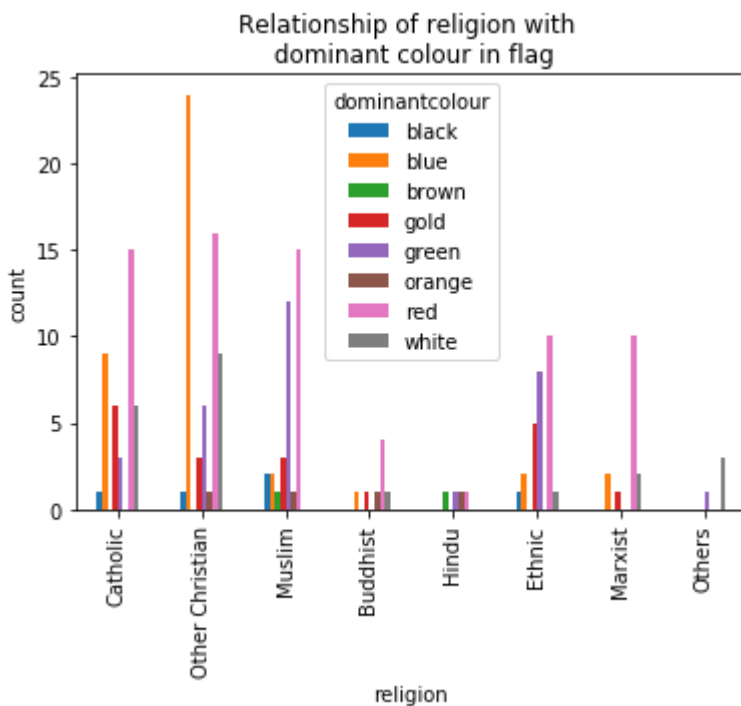


In [84]:

```
print (flags.groupby(['religion1', 'dominantcolour']).size().unstack(fill_value=0))
plt.figure(figsize=(50,20))
#fig.set_size_inches(11.7, 8.27)
pd.crosstab(flags['religion1'], flags['dominantcolour']).plot.bar()
plt.title('Relationship of religion with \n dominant colour in flag')
plt.xlabel('religion')
plt.ylabel('count')
plt.show()
```

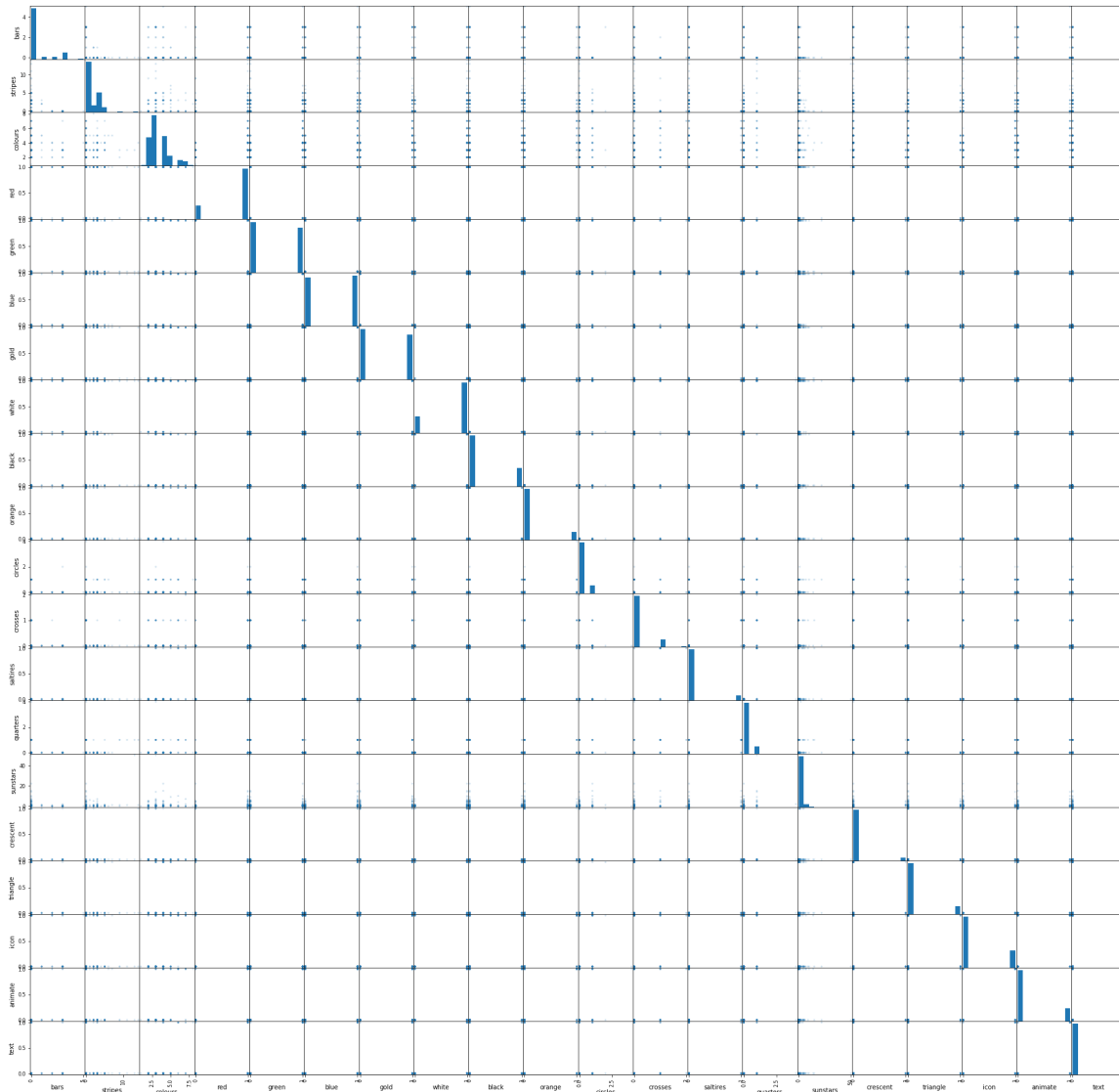
dominantcolour	black	blue	brown	gold	green	orange	red	white
religion1								
Catholic	1	9	0	6	3	0	15	6
Other Christian	1	24	0	3	6	1	16	9
Muslim	2	2	1	3	12	1	15	0
Buddhist	0	1	0	1	0	1	4	1
Hindu	0	0	1	0	1	1	1	0
Ethnic	1	2	0	5	8	0	10	1
Marxist	0	2	0	1	0	0	10	2
Others	0	0	0	0	1	0	0	3

<matplotlib.figure.Figure at 0xdc62780>



In [85]:

```
from pandas.plotting import scatter_matrix
pd.plotting.scatter_matrix
scatter_matrix(flags,alpha=0.2,figsize=(32,32),diagonal='hist')
plt.show()
```



In [86]:

```
cleanup_nums = {"topleftcolour": {"black": 1, "blue": 2, "gold": 3, "green":4, "orange":5, "red":6, "white":7},
                "botrightcolor": {"black": 1, "blue": 2, "gold": 3, "green":4, "orange":5, "red":6, "white":7, "brown":8},
                "dominantcolour": {"black": 1, "blue": 2, "gold": 3, "green":4, "orange":5, "red":6, "white":7, "brown":8}}
```

In [87]:

```
flags.replace(cleanup_nums, inplace=True)
flags.head(2)
```

Out[87]:

	religion	bars	stripes	colours	red	green	blue	gold	white	black	...	quarters	sunst
0	2	0	3	5	1	1	0	1	1	1	...	0	
1	6	0	0	3	1	0	0	1	0	1	...	0	

2 rows × 25 columns

In [88]:

```
flags = pd.concat([flags, pd.get_dummies(flags['topleftcolour'], prefix = 'topleft')],
axis=1)
del flags['topleftcolour']
```

In [89]:

```
flags = pd.concat([flags, pd.get_dummies(flags['botrightcolor'], prefix = 'botright')],
axis=1)
del flags['botrightcolor']
```

In [90]:

```
flags = pd.concat([flags, pd.get_dummies(flags['dominantcolour'], prefix = 'dom')], axis=1)
del flags['dominantcolour']
```

In [91]:

```
flags.head(2)
```

Out[91]:

	religion	bars	stripes	colours	red	green	blue	gold	white	black	...	botright_7	botr
0	2	0	3	5	1	1	0	1	1	1	...	0	
1	6	0	0	3	1	0	0	1	0	1	...	0	

2 rows × 45 columns

In [92]:

```
flags['religion'].astype('category')
```

Out[92]:

0	2
1	6
2	2
3	1
4	0
5	5
6	1
7	1
8	0
9	0
10	1
11	0
12	1
13	2
14	2
15	1
16	0
17	1
18	5
19	1
20	3
21	0
22	5
23	0
24	1
25	2
26	6
27	5
28	3
29	5
..	
164	1
165	1
166	1
167	2
168	3
169	5
170	3
171	7
172	1
173	1
174	2
175	2
176	1
177	1
178	2
179	5
180	1
181	0
182	1
183	1
184	6
185	1
186	0
187	0
188	6
189	1
190	6
191	5

```
192    5
193    5
```

```
Name: religion, Length: 194, dtype: category
Categories (8, int64): [0, 1, 2, 3, 4, 5, 6, 7]
```

In [93]:

```
flags.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 194 entries, 0 to 193
Data columns (total 45 columns):
 religion      194 non-null object
 bars          194 non-null int64
 stripes       194 non-null int64
 colours       194 non-null int64
 red           194 non-null int64
 green         194 non-null int64
 blue          194 non-null int64
 gold          194 non-null int64
 white         194 non-null int64
 black         194 non-null int64
 orange        194 non-null int64
 circles       194 non-null int64
 crosses       194 non-null int64
 saltires     194 non-null int64
 quarters     194 non-null int64
 sunstars     194 non-null int64
 crescent     194 non-null int64
 triangle     194 non-null int64
 icon         194 non-null int64
 animate      194 non-null int64
 text         194 non-null int64
 religion1     194 non-null category
 topleft_1    194 non-null uint8
 topleft_2    194 non-null uint8
 topleft_3    194 non-null uint8
 topleft_4    194 non-null uint8
 topleft_5    194 non-null uint8
 topleft_6    194 non-null uint8
 topleft_7    194 non-null uint8
 botright_1   194 non-null uint8
 botright_2   194 non-null uint8
 botright_3   194 non-null uint8
 botright_4   194 non-null uint8
 botright_5   194 non-null uint8
 botright_6   194 non-null uint8
 botright_7   194 non-null uint8
 botright_8   194 non-null uint8
 dom_1        194 non-null uint8
 dom_2        194 non-null uint8
 dom_3        194 non-null uint8
 dom_4        194 non-null uint8
 dom_5        194 non-null uint8
 dom_6        194 non-null uint8
 dom_7        194 non-null uint8
 dom_8        194 non-null uint8
dtypes: category(1), int64(20), object(1), uint8(23)
memory usage: 36.8+ KB
```

In [94]:

```
import csv  
flags.to_csv(path_or_buf=path+"/part1.csv", sep=';')
```