

In [1]:

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

Подготовка данных

Датасет содержит степень влияния разных факторов на уровень жизни в стране. В нём 156 стран. Столбцы: страна, место в рейтинге, среднее отклонение от этого места, положительный и отрицательный аффекты, соцподдержка, свобода, коррупция, щедрость, ВВП на душу население и продолжительность жизни.

In [2]:

```
happy = pd.read_csv('happy.csv', delimiter=',')
```

In [3]:

```
happy.head(15)
```

Out[3]:

	Country (region)	Ladder	SD of Ladder	Positive affect	Negative affect	Social support	Freedom	Corruption	Generosity	Log of GDP\ner capita	Healthy life\neexpectancy
0	Finland	1	4	41.0	10.0	2.0	5.0	4.0	47.0	22.0	27.0
1	Denmark	2	13	24.0	26.0	4.0	6.0	3.0	22.0	14.0	23.0
2	Norway	3	8	16.0	29.0	3.0	3.0	8.0	11.0	7.0	12.0
3	Iceland	4	9	3.0	3.0	1.0	7.0	45.0	3.0	15.0	13.0
4	Netherlands	5	1	12.0	25.0	15.0	19.0	12.0	7.0	12.0	18.0
5	Switzerland	6	11	44.0	21.0	13.0	11.0	7.0	16.0	8.0	4.0
6	Sweden	7	18	34.0	8.0	25.0	10.0	6.0	17.0	13.0	17.0
7	New Zealand	8	15	22.0	12.0	5.0	8.0	5.0	8.0	26.0	14.0
8	Canada	9	23	18.0	49.0	20.0	9.0	11.0	14.0	19.0	8.0
9	Austria	10	10	64.0	24.0	31.0	26.0	19.0	25.0	16.0	15.0
10	Australia	11	26	47.0	37.0	7.0	17.0	13.0	6.0	18.0	10.0
11	Costa Rica	12	62	4.0	87.0	42.0	16.0	58.0	75.0	67.0	28.0
12	Israel	13	14	104.0	69.0	38.0	93.0	74.0	24.0	31.0	11.0
13	Luxembourg	14	3	62.0	19.0	27.0	28.0	9.0	30.0	2.0	16.0
14	United Kingdom	15	16	52.0	42.0	9.0	63.0	15.0	4.0	23.0	24.0

In [4]:

```
happy.columns = ['country', 'ladder', 'ladder_sd', 'positive_affect', 'negative_affect',  
                 'social_support', 'freedom', 'corruption', 'generosity', 'gdp_per_capita',  
                 'healthy_life_expectancy']
```

Описательная статистика + гипотезы

1. Предположим, что влияние фактора свободы больше в более развитых странах

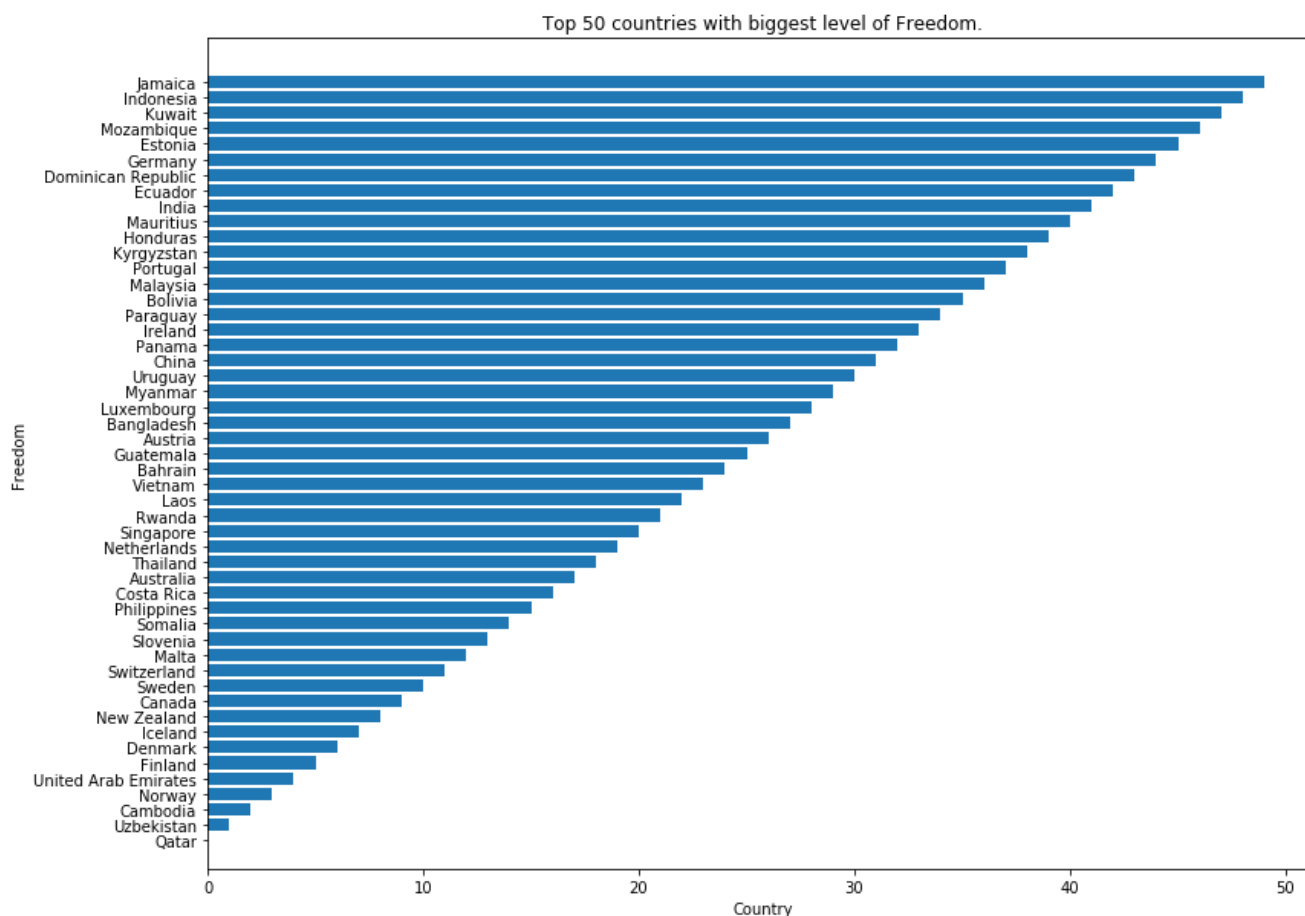
In [7]:

```
plt.figure(figsize=(13,10))  
happy.sort_values('freedom', ascending=True, inplace=True, na_position='first')
```

```
plt.barh(happy.country[:50],happy.freedom[:50])
plt.title('Top 50 countries with biggest level of Freedom.')
plt.xlabel('Country')
plt.ylabel('Freedom')
```

Out[7]:

```
Text(0, 0.5, 'Freedom')
```



1. Попробуем изобразить, как три фактора: свобода, здоровье и социальная поддержка, влияют на рейтинг страны вместе:

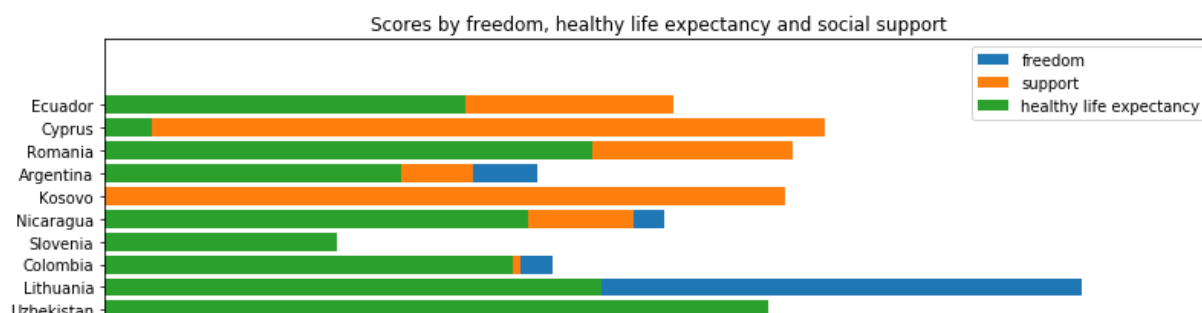
In [9]:

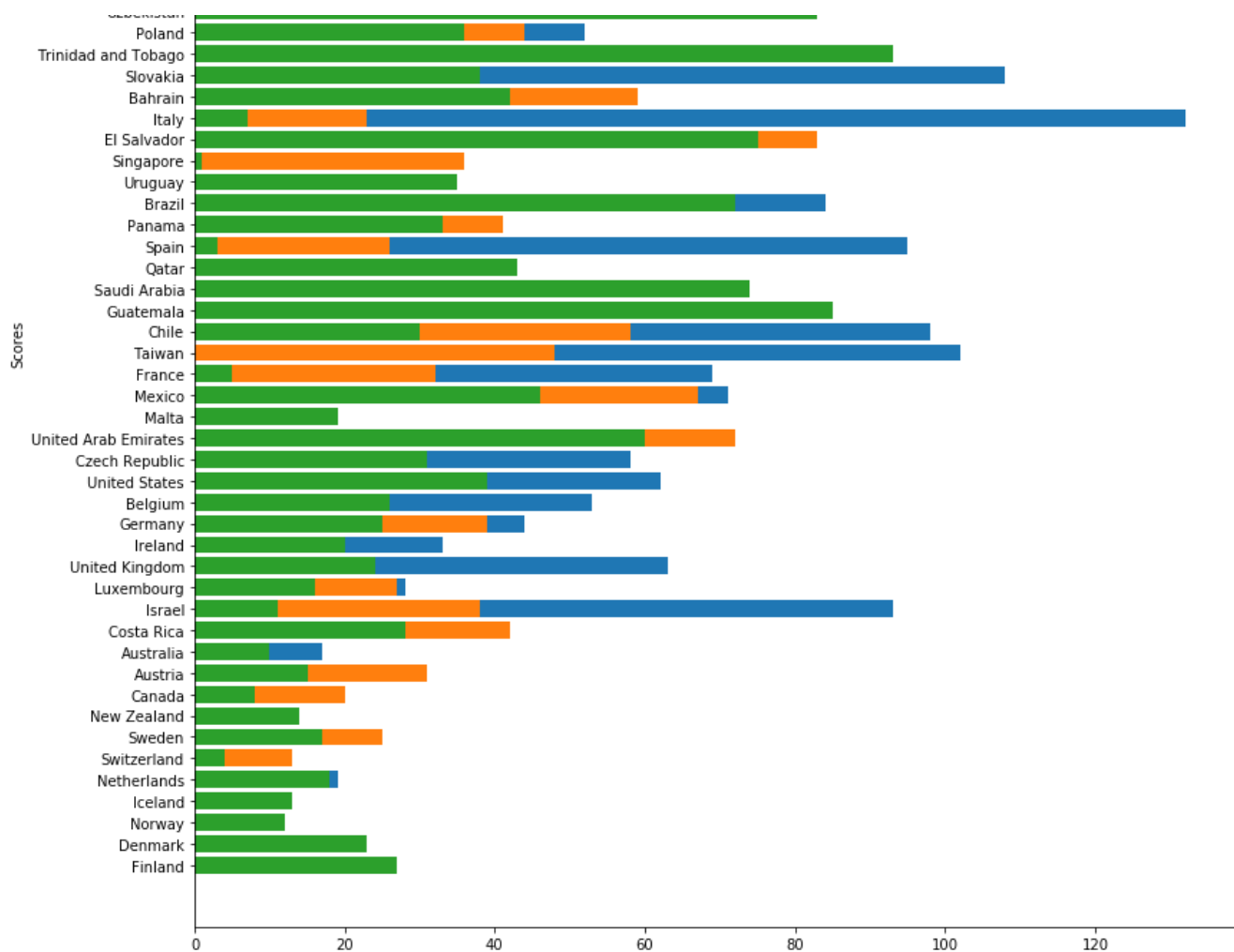
```
plt.figure(figsize=(13,15))
happy.sort_values('ladder',ascending=True, inplace=True)

p1 = plt.barh(happy.country[:50],happy.freedom[:50])
p2 = plt.barh(happy.country[:50],happy.social_support[:50])
p3 = plt.barh(happy.country[:50],happy.healthy_life_expectancy[:50])

plt.ylabel('Scores')
plt.title('Scores by freedom, healthy life expectancy and social support')
plt.yticks(np.arange(50))
plt.legend((p1[0], p2[0], p3[0]), ('freedom', 'support', 'healthy life expectancy'))

plt.show()
```





1. Предположим, что влияние коррупции на уровень меньше там, где коррупции меньше и уровень жизни выше

In [15]:

```
plt.figure(figsize=(13,15))
happy.sort_values('corruption', ascending=True, inplace=True)
plt.barh(happy.country[:50], happy.corruption[:50])
plt.title('Top 50 countries with biggest level of Corruption')
plt.xlabel('Country')
plt.ylabel('Corruption')
```

Out[15]:

Text(0, 0.5, 'Corruption')

