World Happiness Analysis

Principal Investigators: Soma Vasuthevan (SV1168) & Demetrius Beckham (DAB614)

1. Project Overview

Background and Justification for Project

Happiness is a subjective concept that people readily assume can be measured. The 6th World Happiness Report was published in 2018, to survey the science of measuring and understanding subjective well-being.

In this project we set out to further analyze the results observed in this published report. Why do certain factors impact happiness over others? Does wealth lead to health and happiness? Are there any patterns worth identifying when it comes to how people in different countries view happiness? What are some ways in which the overall global level of happiness could be improved over time?

Understanding the data

The World Happiness Report used the following factors to determine the happiness score for each country:

- Economy (GDP per Capita) - Purchasing Power Parity
- Family
- Health (Life Expectancy)
- Freedom - "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"
- Trust (Government Corruption) - "Is Corruption widespread throughout the government or not?"
- Generosity - "Have you donated money to a charity in the past month?"
- Dystopia Residual

The happiness scores and rankings use data from the Gallup World Poll. Respondents were asked to think of a ladder with the best possible life for them being a 10 and the worst being a 0. If you add all 6 factors up, the happiness score is calculated.

"Dystopia" is an imaginary coutnry that has the world's least-happy people. In this data set, it's used as a benchmark against which all other countries can be favorably compared with, in regards to each of the 6 listed variables. In other words, the lowest scores observed for the 6 key variables, characterize Dystopia. The Dystopia Residual metric is the Dystopia Happiness Score (1.85) + the Residual value, or the extent to which the six variables either over-or-under explain average life evaluations.

Methods and Expected Results

In addition to the data provided by The World Happiness Report, we were interested to see how certain other factors - crime (https://www.numbeo.com/crime/rankings_by_country.jsp?title=2015), pollution and education (https://www.rug.nl/ggdc/productivity/pwt/) - impact the overal quality of life for people.

We will be reading in data from the World Happiness Report for 2015, 2016 and 2017. These values are averaged over the 3 years to produce a mean for each country. The additional factors (crime, pollution and education) are appended to the data based on country. From there, we can analyze the different factors and there relative contribution to the happiness scores.

It is anticipated that there will be a strong corelation between some of these aforementioned indicators and the relative level of happiness a country is experiencing. In the end, we hope to obtain a better understanding of how specific economic and social factors impact the overall quality of life around the world.

2. Importing Packages and Data

In [1]:
```
import pandas as pd              # data package
import matplotlib.pyplot as plt # graphics
import requests, io             # internet and input tools
import zipfile as zf            # zip file tools
import shutil                   # file management tools
import os                       # operating system tools (check files)
import chardet
import quandl
import datetime
import numpy as np

%matplotlib inline
```

Source: https://pythonhosted.org/PyDrive/_sources/quickstart.txt (https://pythonhosted.org/PyDrive/_sources/quickstart.txt)

Drive API requires OAuth2.0 for authentication. *PyDrive* makes your life much easier by handling complex authentication steps for you.

1. Go to `APIs Console_` and make your own project.
2. Search for 'Google Drive API', select the entry, and click 'Enable'.
3. Select 'Credentials' from the left menu, click 'Create Credentials', select 'OAuth client ID'.
4. Now, the product name and consent screen need to be set -> click 'Configure consent screen' and follow the instructions. Once finished:

   a. Select 'Application type' to be *Web application*. b. Enter an appropriate name. c. Input *http://localhost:8080 (http://localhost:8080)* for 'Authorized JavaScript origins'. d. Input *http://localhost:8080/ (http://localhost:8080/)* for 'Authorized redirect URIs'. e. Click 'Save'.
5. Click 'Download JSON' on the right side of Client ID to download **client*secret*.json**.

The downloaded file has all authentication information of your application. **Rename the file to "client_secrets.json" and place it in your working directory.**

Read in World Happiness Report Data from Google Drive

In [2]:
```
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive

gauth = GoogleAuth()
gauth.LocalWebserverAuth() # Creates local webserver and auto handles authenticati
on.
drive = GoogleDrive(gauth)
```

```
Your browser has been opened to visit:

    https://accounts.google.com/o/oauth2/auth?client_id=542848275155-g692v7r9qfn
4ip8chdo77gkijg8r5ujd.apps.googleusercontent.com&redirect_uri=http%3A%2F%2Flocal
host%3A8080%2F&scope=https%3A%2F%2Fwww.googleapis.com%2Fauth%2Fdrive&access_type
=offline&response_type=code

Authentication successful.
```

In [3]:
```
#file_id of reports zip
file_id = '1jY8qFccnjhClfcAaw_J1bwYhWBBkT0Gj'
downloaded = drive.CreateFile({'id': file_id})
downloaded.GetContentFile(file_id)

zipf = zf.ZipFile(file_id)

file_list = zipf.namelist()

file_list
```

Out[3]:
```
['2015.csv', '2016.csv', '2017.csv']
```

In [4]:
```
df15  = pd.read_csv(zipf.open(zipf.namelist()[0]))
df16  = pd.read_csv(zipf.open(zipf.namelist()[1]))
df17  = pd.read_csv(zipf.open(zipf.namelist()[2]))
```

In [5]:
```
df15.head()
```

Out[5]:

	Country	Region	Happiness Rank	Happiness Score	Standard Error	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom
0	Switzerland	Western Europe	1	7.587	0.03411	1.39651	1.34951	0.94143	0.66557
1	Iceland	Western Europe	2	7.561	0.04884	1.30232	1.40223	0.94784	0.62877
2	Denmark	Western Europe	3	7.527	0.03328	1.32548	1.36058	0.87464	0.64938
3	Norway	Western Europe	4	7.522	0.03880	1.45900	1.33095	0.88521	0.66973
4	Canada	North America	5	7.427	0.03553	1.32629	1.32261	0.90563	0.63297

Adding World Bank, Crime and Education Data

To expand on the story of factors influencing happiness measurements, we read in data from the World Bank. This includes the following data points for every country in the world:

- % of primary school age children
- Air pollution
- Patent applications
- International Homicides (per 100,000 people)
- GDP per Capita
- Population
- Population Density

These factors paint a clearer picture of the overal economic and societal state of the countries we are observing.

In [6]:
```
#Read in and format World Bank data

file_id = '1hN9l_hBj54oAxIa3kosCnTRbKFW8gTxo'
downloaded = drive.CreateFile({'id': file_id})
downloaded.GetContentFile(file_id)

zipf = zf.ZipFile(file_id)

file_list = zipf.namelist()

wbdf  = pd.read_csv(zipf.open(zipf.namelist()[0]))

wbdf.head()
```

Out[6]:

	Time	Time Code	Country Name	Country Code	Adjusted net enrollment rate, primary (% of primary school age children) [SE.PRM.TENR]	PM2.5 air pollution, population exposed to levels exceeding WHO guideline value (% of total) [EN.ATM.PM25.MC.ZS]	PM2.5 air pollu mean annual expo (micrograms per c m [EN.ATM.PM25.MC
0	2016	YR2016	Afghanistan	AFG	..	100	62.8548565825039
1	2016	YR2016	Albania	ALB	95.6585464477539	100	14.6340083732708
2	2016	YR2016	Algeria	DZA	99.2312088012695	100	37.2309558775069
3	2016	YR2016	American Samoa	ASM	..	0	3.76341150774918
4	2016	YR2016	Andorra	AND	..	100	10.8794724208596

Crime is another crucial factor to consider when discussing the well-being of a country. In this case, the Crime Index of a country is inversely correlated with its Safety index. We will read in data to determine if these indices are in fact correlated to happiness score in anyway.

In [7]:
```
#Read in crime data

url_crime = "https://raw.githubusercontent.com/DBeckham96/GitHub/master/Crime%20In
dex%20for%20Country%202015.csv"

crime = pd.read_csv(url_crime)

cdf = pd.DataFrame(crime)

cdf.drop("Rank",inplace = True,axis = 1)

cdf.head()
```

Out[7]:

	Country	Crime Index	Safety Index
0	South Sudan	85.32	14.68
1	Venezuela	84.07	15.93
2	Guatemala	79.34	20.66
3	South Africa	78.44	21.56
4	Afghanistan	77.34	22.66

Last but not least, we read in data that gauges the education levels of countries around the world. This is measured by Human Capital, based on years of schooling and return to education.

In [8]:
```
# Read in data for Education levels (Human Capital)

url = "http://www.rug.nl/ggdc/docs/pwt81.xlsx"
pwt = pd.read_excel(url, sheetname='Data')

pwt_2011 = pwt[pwt.year == 2011]

pwt_hc = pwt_2011[["country","hc"]]

pwt_hc.head()
```

```
/anaconda3/lib/python3.6/site-packages/pandas/util/_decorators.py:118: FutureWar
ning: The `sheetname` keyword is deprecated, use `sheet_name` instead
  return func(*args, **kwargs)
```

Out[8]:

	country	hc
61	Angola	NaN
123	Albania	3.004226
185	Argentina	2.818635
247	Armenia	3.037621
309	Antigua and Barbuda	NaN

3. Structure and Combine Data Frames

To get the most accurate picture, we will combine and average the world happiness measures for 2015, 2016 and 2017. Further, the World Bank Data, Crime and Education measures are added to one data set.

Averaging World Happiness Data

In [9]:

```
col15 = df15.columns.values.tolist()
df15['Year'] = "2015"
df15.drop(['Standard Error'], axis=1,inplace = True)
df15 = df15[[
             'Country',
              'Year',
              'Region',
             "Happiness Rank",
              'Happiness Score',
              'Economy (GDP per Capita)',
              'Family',
              'Health (Life Expectancy)',
              'Freedom',
              'Trust (Government Corruption)',
              'Generosity',
              'Dystopia Residual',
                 ]]
df15.head()
```

Out[9]:

	Country	Year	Region	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	(Gov Co
0	Switzerland	2015	Western Europe	1	7.587	1.39651	1.34951	0.94143	0.66557	0.41
1	Iceland	2015	Western Europe	2	7.561	1.30232	1.40223	0.94784	0.62877	0.14
2	Denmark	2015	Western Europe	3	7.527	1.32548	1.36058	0.87464	0.64938	0.48
3	Norway	2015	Western Europe	4	7.522	1.45900	1.33095	0.88521	0.66973	0.36
4	Canada	2015	North America	5	7.427	1.32629	1.32261	0.90563	0.63297	0.32

In [10]:
```
col16 = df16.columns.values.tolist()
df16['Year'] = "2016"
df16.drop(['Lower Confidence Interval', 'Upper Confidence Interval'], axis=1,inpla
ce = True)
df16 = df16[[
            'Country',
             'Year',
             'Region',
            "Happiness Rank",
             'Happiness Score',
             'Economy (GDP per Capita)',
             'Family',
             'Health (Life Expectancy)',
             'Freedom',
             'Trust (Government Corruption)',
             'Generosity',
             'Dystopia Residual',
                ]]
df16.head()
```

Out[10]:

	Country	Year	Region	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	(Gov Co
0	Denmark	2016	Western Europe	1	7.526	1.44178	1.16374	0.79504	0.57941	0.44
1	Switzerland	2016	Western Europe	2	7.509	1.52733	1.14524	0.86303	0.58557	0.41
2	Iceland	2016	Western Europe	3	7.501	1.42666	1.18326	0.86733	0.56624	0.14
3	Norway	2016	Western Europe	4	7.498	1.57744	1.12690	0.79579	0.59609	0.35
4	Finland	2016	Western Europe	5	7.413	1.40598	1.13464	0.81091	0.57104	0.41

In [11]:
```
df17['Year'] = "2017"

#df17.set_index("Happiness.Rank", inplace = True)

df17.drop(['Whisker.high','Whisker.low'], axis=1,inplace = True)

col17 = df17.columns.values.tolist()

col17 = [   'Country',
            "Happiness Rank",
             'Happiness Score',
             'Economy (GDP per Capita)',
             'Family',
             'Health (Life Expectancy)',
             'Freedom',
             'Trust (Government Corruption)',
             'Generosity',
             'Dystopia Residual',
             'Year',
                ]

df17.columns = col17

df17 = df17[[
            'Country',
             'Year',
            "Happiness Rank",
             'Happiness Score',
             'Economy (GDP per Capita)',
             'Family',
             'Health (Life Expectancy)',
             'Freedom',
             'Trust (Government Corruption)',
             'Generosity',
             'Dystopia Residual',
                ]]

df17.head()
```

Out[11]:

	Country	Year	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trus (Governmen Corruption
0	Norway	2017	1	7.537	1.616463	1.533524	0.796667	0.635423	0.362012
1	Denmark	2017	2	7.522	1.482383	1.551122	0.792566	0.626007	0.355280
2	Iceland	2017	3	7.504	1.480633	1.610574	0.833552	0.627163	0.475540
3	Switzerland	2017	4	7.494	1.564980	1.516912	0.858131	0.620071	0.290549
4	Finland	2017	5	7.469	1.443572	1.540247	0.809158	0.617951	0.245483

In [12]:
```
# combine annual data
mdf = df15.append(df16, ignore_index = True).append(df17, ignore_index = True)
mdf = mdf[[
              'Country',
               'Year',
              "Happiness Rank",
               'Happiness Score',
               'Economy (GDP per Capita)',
               'Family',
               'Health (Life Expectancy)',
               'Freedom',
               'Trust (Government Corruption)',
               'Generosity',
               'Dystopia Residual',
                  ]]
mdf.shape
```

Out[12]: (470, 11)

In [13]:
```
mdf.head()
```

Out[13]:

	Country	Year	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)
0	Switzerland	2015	1	7.587	1.39651	1.34951	0.94143	0.66557	0.41978
1	Iceland	2015	2	7.561	1.30232	1.40223	0.94784	0.62877	0.14145
2	Denmark	2015	3	7.527	1.32548	1.36058	0.87464	0.64938	0.48357
3	Norway	2015	4	7.522	1.45900	1.33095	0.88521	0.66973	0.36503
4	Canada	2015	5	7.427	1.32629	1.32261	0.90563	0.63297	0.32957

In [14]:
```
#Create a df on the mean of 2015 - 2017 data
mdf_mean = mdf.groupby("Country").mean()
mdf_mean.sort_values("Happiness Score", ascending = False, inplace = True)
mdf_mean.head()
```

Out[14]:

	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Gener
Country								
Switzerland	2.333333	7.530000	1.496273	1.337221	0.887530	0.623737	0.374120	0.3148
Denmark	2.000000	7.525000	1.416548	1.358481	0.820749	0.618266	0.427793	0.3679
Iceland	2.666667	7.522000	1.403204	1.398688	0.882907	0.607391	0.255580	0.3555
Norway	3.000000	7.519000	1.550968	1.330458	0.825889	0.633748	0.361601	0.3473
Finland	5.333333	7.429333	1.379934	1.331049	0.836393	0.610227	0.356414	0.2903

World Bank Data

In [15]:
```
wbdf.drop(wbdf.columns[1], axis =1, inplace = True)
wbdf.drop(wbdf.columns[2], axis =1, inplace = True)
wbdf.drop(["Intentional homicides (per 100,000 people) [VC.IHR.PSRC.P5]"],axis = 1
,inplace = True)
wbdf.drop(["Time"],axis = 1, inplace = True)
wbdfcol = wbdf.columns.tolist()
wbdfcol
```

Out[15]:
```
['Country Name',
 'Adjusted net enrollment rate, primary (% of primary school age children) [SE.P
RM.TENR]',
 'PM2.5 air pollution, population exposed to levels exceeding WHO guideline valu
e (% of total) [EN.ATM.PM25.MC.ZS]',
 'PM2.5 air pollution, mean annual exposure (micrograms per cubic meter) [EN.ATM
.PM25.MC.M3]',
 'GINI index (World Bank estimate) [SI.POV.GINI]',
 'Patent applications, residents [IP.PAT.RESD]',
 'GDP per capita, PPP (current international $) [NY.GDP.PCAP.PP.CD]',
 'GDP, PPP (current international $) [NY.GDP.MKTP.PP.CD]',
 'Population, total [SP.POP.TOTL]',
 'Population density (people per sq. km of land area) [EN.POP.DNST]']
```

In [16]:
```
wbdf.columns = ['Country',
                'Adjusted net enrollment rate, primary (% of primary school age c
hildren) [SE.PRM.TENR]',
                'PM2.5 air pollution, population exposed to levels exceeding WHO
guideline value (% of total) [EN.ATM.PM25.MC.ZS]',
                'PM2.5 air pollution, mean annual exposure (micrograms per cubic
meter) [EN.ATM.PM25.MC.M3]',
                'GINI index (World Bank estimate) [SI.POV.GINI]',
                'Patent applications, residents [IP.PAT.RESD]',
                'GDP per capita, PPP (current international $) [NY.GDP.PCAP.PP.CD
]',
                'GDP, PPP (current international $) [NY.GDP.MKTP.PP.CD]',
                'Population, total [SP.POP.TOTL]',
                'Population density (people per sq. km of land area) [EN.POP.DNST
]']
```

In [17]:
```
mdf1 = pd.merge(mdf_mean, wbdf, left_index = True, right_on = 'Country')

mdf1.set_index("Country",inplace = True)

mdf1.head()
```

Out[17]:

	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Gener
Country								
Switzerland	2.333333	7.530000	1.496273	1.337221	0.887530	0.623737	0.374120	0.314
Denmark	2.000000	7.525000	1.416548	1.358481	0.820749	0.618266	0.427793	0.367
Iceland	2.666667	7.522000	1.403204	1.398688	0.882907	0.607391	0.255580	0.355
Norway	3.000000	7.519000	1.550968	1.330458	0.825889	0.633748	0.361601	0.347
Finland	5.333333	7.429333	1.379934	1.331049	0.836393	0.610227	0.356414	0.290

Crime Data

In [18]:
```
mdf2 = pd.merge(mdf1, cdf, left_index = True, right_on = 'Country')
mdf2.set_index("Country",inplace = True)
mdf2.head()
```

Out[18]:

	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Gener
Country								
Switzerland	2.333333	7.530000	1.496273	1.337221	0.887530	0.623737	0.374120	0.3148
Denmark	2.000000	7.525000	1.416548	1.358481	0.820749	0.618266	0.427793	0.3679
Iceland	2.666667	7.522000	1.403204	1.398688	0.882907	0.607391	0.255580	0.3555
Norway	3.000000	7.519000	1.550968	1.330458	0.825889	0.633748	0.361601	0.3473
Finland	5.333333	7.429333	1.379934	1.331049	0.836393	0.610227	0.356414	0.2903

In [19]:
```
mdf2.columns.tolist()
```

Out[19]:
```
['Happiness Rank',
 'Happiness Score',
 'Economy (GDP per Capita)',
 'Family',
 'Health (Life Expectancy)',
 'Freedom',
 'Trust (Government Corruption)',
 'Generosity',
 'Dystopia Residual',
 'Adjusted net enrollment rate, primary (% of primary school age children) [SE.P
RM.TENR]',
 'PM2.5 air pollution, population exposed to levels exceeding WHO guideline valu
e (% of total) [EN.ATM.PM25.MC.ZS]',
 'PM2.5 air pollution, mean annual exposure (micrograms per cubic meter) [EN.ATM
.PM25.MC.M3]',
 'GINI index (World Bank estimate) [SI.POV.GINI]',
 'Patent applications, residents [IP.PAT.RESD]',
 'GDP per capita, PPP (current international $) [NY.GDP.PCAP.PP.CD]',
 'GDP, PPP (current international $) [NY.GDP.MKTP.PP.CD]',
 'Population, total [SP.POP.TOTL]',
 'Population density (people per sq. km of land area) [EN.POP.DNST]',
 'Crime Index',
 'Safety Index']
```

In [20]:
```
mdf2.columns = ['Happiness Rank',
                'Happiness Score',
                'Economy',
                'Family',
                'Health',
                'Freedom',
                'Trust',
                'Generosity',
                'Dystopia Residual',
                'Primary Enrollment',
                'Air Pollution (% pop)',
                'Air Pollution (exposure)',
                'GINI',
                'Patent applications',
                'GDP per capita',
                'GDP',
                'Population',
                'Population density',
                'Crime Index',
                'Safety Index']
```

In [21]:
```
mdf2.dtypes
```

Out[21]:
```
Happiness Rank              float64
Happiness Score             float64
Economy                     float64
Family                      float64
Health                      float64
Freedom                     float64
Trust                       float64
Generosity                  float64
Dystopia Residual           float64
Primary Enrollment           object
Air Pollution (% pop)        object
Air Pollution (exposure)     object
GINI                         object
Patent applications          object
GDP per capita               object
GDP                          object
Population                   object
Population density           object
Crime Index                 float64
Safety Index                float64
dtype: object
```

In [22]:
```
mdf2 = mdf2.convert_objects(convert_numeric = True, copy = False)
```

```
/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:1: FutureWarning: c
onvert_objects is deprecated.  To re-infer data dtypes for object columns, use D
ataFrame.infer_objects()
For all other conversions use the data-type specific converters pd.to_datetime,
pd.to_timedelta and pd.to_numeric.
  """Entry point for launching an IPython kernel.
```

In [23]:
```
mdf2.dtypes
```

Out[23]:
```
Happiness Rank              float64
Happiness Score             float64
Economy                     float64
Family                      float64
Health                      float64
Freedom                     float64
Trust                       float64
Generosity                  float64
Dystopia Residual           float64
Primary Enrollment          float64
Air Pollution (% pop)       float64
Air Pollution (exposure)    float64
GINI                        float64
Patent applications         float64
GDP per capita              float64
GDP                         float64
Population                    int64
Population density          float64
Crime Index                 float64
Safety Index                float64
dtype: object
```

In [24]:
```
mdf2.head()
```

Out[24]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity	D F
Country									
Switzerland	2.333333	7.530000	1.496273	1.337221	0.887530	0.623737	0.374120	0.314872	2
Denmark	2.000000	7.525000	1.416548	1.358481	0.820749	0.618266	0.427793	0.367957	2
Iceland	2.666667	7.522000	1.403204	1.398688	0.882907	0.607391	0.255580	0.355536	2
Norway	3.000000	7.519000	1.550968	1.330458	0.825889	0.633748	0.361601	0.347301	2
Finland	5.333333	7.429333	1.379934	1.331049	0.836393	0.610227	0.356414	0.290347	2

In [25]:
```
mdf2.sort_values("GDP per capita", ascending = False).head(10)
```

Out[25]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity
Country								
Qatar	33.000000	6.453667	1.795152	1.077512	0.741553	0.603774	0.444348	0.362970
Luxembourg	18.333333	6.893333	1.667791	1.239068	0.869816	0.587053	0.338150	0.291628
Singapore	24.000000	6.703000	1.619896	1.080465	0.973977	0.526687	0.435979	0.367473
Kuwait	39.666667	6.213000	1.601437	1.101073	0.664239	0.494329	0.240357	0.179030
United Arab Emirates	23.000000	6.707333	1.542378	1.087767	0.755326	0.604022	0.367461	0.284893
Ireland	17.333333	6.941333	1.451692	1.363094	0.839888	0.576987	0.337476	0.402343
Switzerland	2.333333	7.530000	1.496273	1.337221	0.887530	0.623737	0.374120	0.314872
Norway	3.000000	7.519000	1.550968	1.330458	0.825889	0.633748	0.361601	0.347301
United States	14.000000	7.072000	1.482910	1.238284	0.805026	0.511137	0.233386	0.315820
Saudi Arabia	35.333333	6.378000	1.471855	1.072966	0.634356	0.379757	0.257645	0.188354

In [26]:
```
#drop top 5 anamolous low population countries

mdf3 = mdf2.drop(["Qatar", "Luxembourg", "Singapore", "Kuwait","United Arab Emirat
es"])

mdf3.sort_values("GDP per capita", ascending = False).head(10)
```

Out[26]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity	D F
Country									
Ireland	17.333333	6.941333	1.451692	1.363094	0.839888	0.576987	0.337476	0.402343	1
Switzerland	2.333333	7.530000	1.496273	1.337221	0.887530	0.623737	0.374120	0.314872	2
Norway	3.000000	7.519000	1.550968	1.330458	0.825889	0.633748	0.361601	0.347301	2
United States	14.000000	7.072000	1.482910	1.238284	0.805026	0.511137	0.233386	0.315820	2
Saudi Arabia	35.333333	6.378000	1.471855	1.072966	0.634356	0.379757	0.257645	0.188354	2
Iceland	2.666667	7.522000	1.403204	1.398688	0.882907	0.607391	0.255580	0.355536	2
Austria	12.666667	7.108333	1.424902	1.280272	0.837133	0.578549	0.238904	0.293530	2
Netherlands	6.666667	7.364667	1.432688	1.246076	0.838615	0.584418	0.362633	0.410974	2
Denmark	2.000000	7.525000	1.416548	1.358481	0.820749	0.618266	0.427793	0.367957	2
Sweden	9.000000	7.313000	1.425969	1.284957	0.857652	0.618301	0.410836	0.376520	2

Education Data

In [27]:
```
# Add education levels
mdf4 = pd.merge(mdf3, pwt_hc, left_index = True, right_on = 'country')
mdf4.set_index("country",inplace = True)
mdf4 = mdf4.rename(columns={'hc': 'Human Capital'})
mdf4.head()
```

Out[27]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity	D F
country									
Switzerland	2.333333	7.530000	1.496273	1.337221	0.887530	0.623737	0.374120	0.314872	2
Denmark	2.000000	7.525000	1.416548	1.358481	0.820749	0.618266	0.427793	0.367957	2
Iceland	2.666667	7.522000	1.403204	1.398688	0.882907	0.607391	0.255580	0.355536	2
Norway	3.000000	7.519000	1.550968	1.330458	0.825889	0.633748	0.361601	0.347301	2
Finland	5.333333	7.429333	1.379934	1.331049	0.836393	0.610227	0.356414	0.290347	2

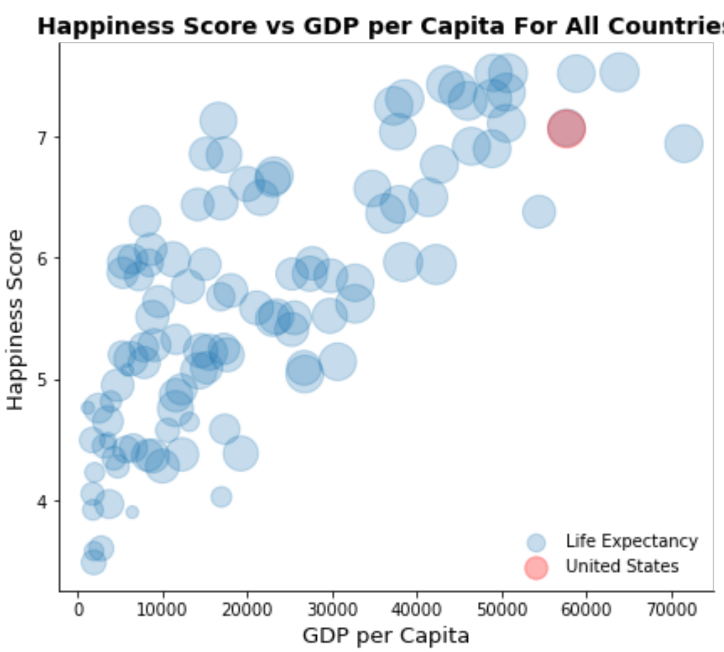
5 rows × 21 columns

4. Analysis

Does wealth lead to health and happiness?

The first comparison we set up is to see the correlation between wealth (as measured by GDP per capita) and happiness. In doing so, we aim to see how strongly correlated the two are, if at all.

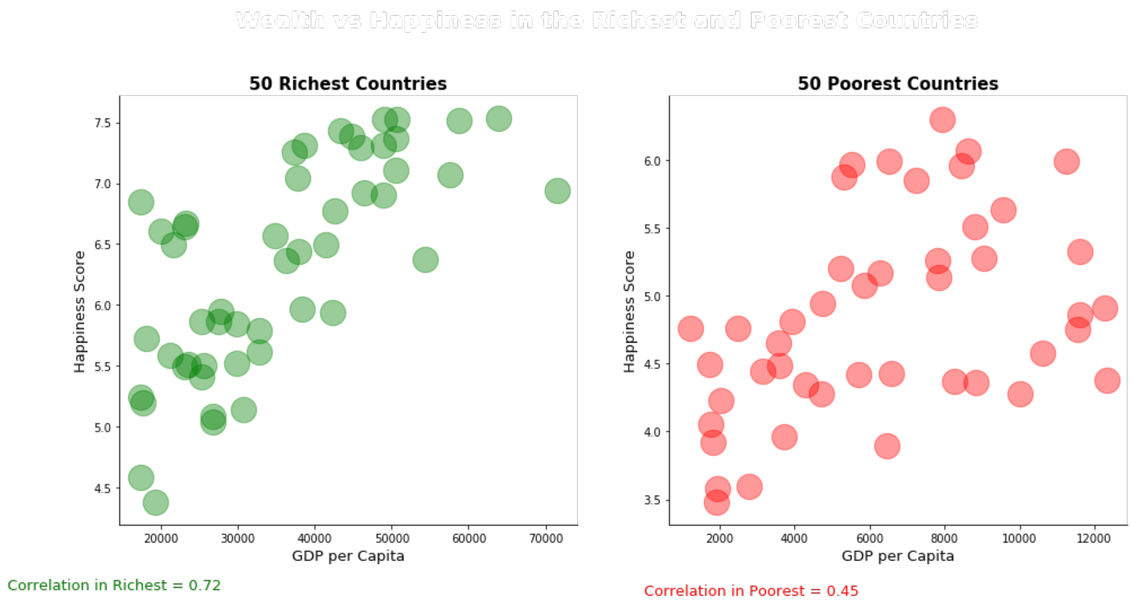
In [28]:
```
fig, ax = plt.subplots(figsize = (7,6))

ax.scatter(x = mdf3["GDP per capita"], y = mdf3["Happiness Score"], alpha = 0.25,
s = 600 * mdf3["Health"])

ax.spines["right"].set_visible(False)
ax.spines["top"].set_visible(False)

#Add title and axis labels
ax.set_title("Happiness Score vs GDP per Capita For All Countries", fontsize = 14,
fontweight = "bold")
ax.set_ylabel("Happiness Score", fontsize = 13)
ax.set_xlabel("GDP per Capita", fontsize = 13)

#Calculate and add Economy correlation text
cr = mdf3["Happiness Score"].corr(mdf3["GDP per capita"])
message = "Wealth / Happiness Correlation = " + str(round(cr,2))
ax.text(0, 2.2, message, size = 13, horizontalalignment='left')

#Calculate and add Health correlation text
cr = mdf3["Health"].corr(mdf3["GDP per capita"])
message = "Wealth / Health Correlation = " + str(round(cr,2))
ax.text(0,2.5,message, size = 13, horizontalalignment='left')

#United States highlight
ax.scatter(x = mdf3.loc['United States'][14], y = mdf3.loc['United States'][1], co
lor = 'red',alpha = 0.3, s = 500)

#Add legend
ax.legend(["Life Expectancy", "United States"],loc = 4, frameon = False, markersca
le = 0.6)

plt.show()
```



Wealth / Health Correlation = 0.73
Wealth / Happiness Correlation = 0.77

As expected, wealth has a strong positive correlation with health and happiness on a global scale. That is to say, wealthier countries tend to be healthier and happier. The United States (highlighted red) lies on the far right end of the spectrum in terms of reported happiness. Let's zoom into the richest and poorest countries to see if this relationship manifests in those groups as well.

In [29]:
```
fig, ax = plt.subplots(nrows = 1, ncols = 2, figsize = (16,7))

fig.text(0.5, 1, "Wealth vs Happiness in the Richest and Poorest Countries", horiz
ontalalignment='center', size = 20, color = "white", weight = "bold")

#plot richest countries

ax[0].scatter(y = mdf3.sort_values("GDP per capita", ascending = False).head(50)["
Happiness Score"], x = mdf3.sort_values("GDP per capita", ascending = False).head(
50)["GDP per capita"], alpha = 0.4, s = 500, c = "g")

ax[0].spines["right"].set_visible(False)
ax[0].spines["top"].set_visible(False)

ax[0].set_title("50 Richest Countries", fontsize = 15, fontweight = "bold")
ax[0].set_xlabel("GDP per Capita", fontsize = 13,)
ax[0].set_ylabel("Happiness Score", fontsize = 13,)

cr = mdf3.sort_values("GDP per capita", ascending = False).head(50)["Happiness Sco
re"].corr(mdf3.sort_values("GDP per capita", ascending = False).head(50)["GDP per
capita"])
message = "Correlation in Richest = " + str(round(cr,2))
ax[0].text(2,3.67, message, horizontalalignment='left', size = 13, color = "g")

# plot poorest countries

ax[1].scatter(y = mdf3.sort_values("GDP per capita", ascending = False).tail(50)["
Happiness Score"], x = mdf3.sort_values("GDP per capita", ascending = False).tail(
50)["GDP per capita"], alpha = 0.4, s = 500, c = "r")

ax[1].spines["right"].set_visible(False)
ax[1].spines["top"].set_visible(False)

ax[1].set_title("50 Poorest Countries", fontsize = 15, fontweight = "bold")
ax[1].set_xlabel("GDP per Capita", fontsize = 13,)
ax[1].set_ylabel("Happiness Score", fontsize = 13,)

cr = mdf3.sort_values("GDP per capita", ascending = False).tail(50)["Happiness Sco
re"].corr(mdf3.sort_values("GDP per capita", ascending = False).tail(50)["GDP per
capita"])
message = "Correlation in Poorest = " + str(round(cr,2))
ax[1].text(0, 2.8, message, horizontalalignment='left', size = 13, color = "r")

plt.show()
```



In the previous chart we observed that wealth is a good indicator of happiness and health globally.

The charts directly above tell us an interesting story about wealth and its relationship to happiness. It appears that in the richest countries (GDP per capita > ~2000), wealth has a strong positive correlation (0.72) with happiness (i.e amongst the richest countries, more wealth tends to increase reported levels of happiness)

Interestingly, amongst the poorest countries (GDP per capita < ~12000) wealth has a weaker positive correlation with happiness compared to the richest countries (0.45 vs 0.72). This suggests that at lower levels of wealth, the incremental happiness of more wealth is less than at higher levels of wealth.

We can speculate that perhaps at low levels of wealth increases in wealth are still not sufficient to meet the population's basic needs. Moreover, in the poorest countries, increases in wealth may be pocketed by corrupt government officials instead of being invested into public infrastructure. Higher levels of inequality may also lead to a disproportionate amount of the increase in wealth being funneled to the wealthy, and therefore not moving the needle on happiness for society at large. Finally, there may be other pressing societal problems such as crime, pollution and access to education that may be depressing happiness scores in the poorest countries.

What is the correlation between happiness and crime?

We suspect that some correlation between happines and crime will be found, perhaps not as strong as the one previously observed.

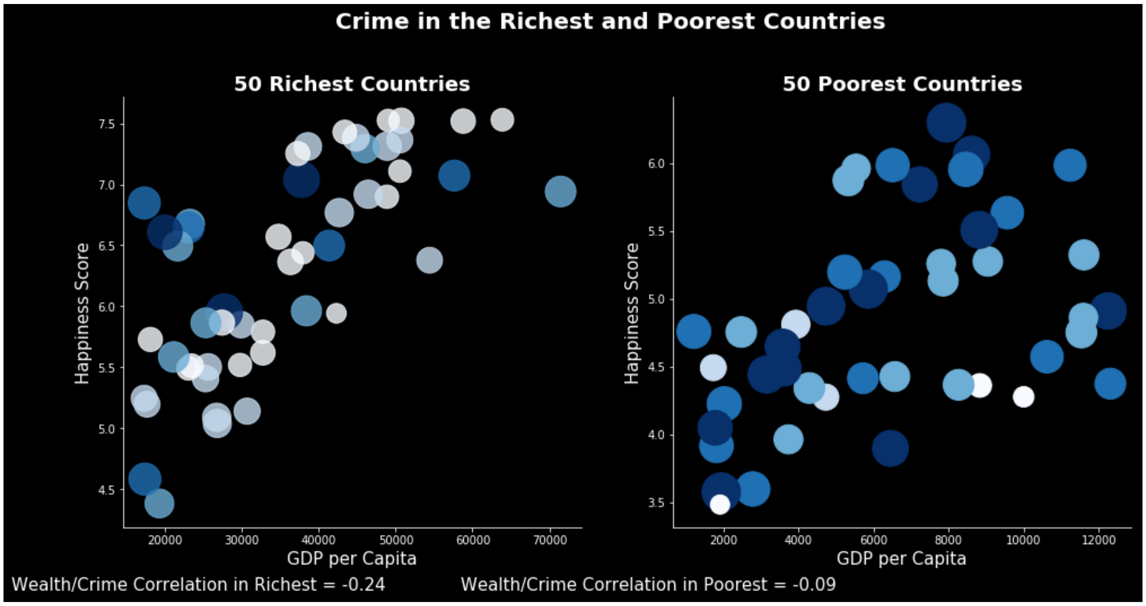
In [30]:
```
#Add column representing crime and pollution quantile to each row
mdf3["crime index quintile"] = pd.qcut(mdf3["Crime Index"], 5, labels = [.2, .4, .
6, 0.8, 1.0] , retbins=False, precision=3, duplicates='raise')
mdf3["air pollution quintile"] = pd.qcut(mdf3["Air Pollution (exposure)"], 5, labe
ls = [.2, .4, .6, 0.8, 1.0] , retbins=False, precision=3, duplicates='raise')

#split by crime / pollution index quantile
cqdf = mdf3.groupby("crime index quintile").mean()
pqdf = mdf3.groupby("air pollution quintile").mean()
```

In [31]:
```
cqdf
```

Out[31]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity	Dysto Resid
crime index quintile									
0.2	48.942029	6.149116	1.224043	1.150699	0.768640	0.475483	0.212559	0.214312	2.103
0.4	58.030303	5.913667	1.162793	1.109601	0.690894	0.410831	0.199398	0.208532	2.131
0.6	80.043478	5.337594	0.954710	0.994804	0.629961	0.412167	0.151712	0.193866	2.000
0.8	77.553030	5.393220	0.840810	1.009018	0.554461	0.427895	0.173204	0.202764	2.185
1.0	82.152174	5.291899	0.782483	0.943839	0.459759	0.400991	0.148305	0.189536	2.366

In [32]:
```
plt.style.use('dark_background')

fig, ax = plt.subplots(nrows = 1, ncols = 2, figsize = (16,7))

fig.text(0.5, 1, "Crime in the Richest and Poorest Countries", horizontalalignment
='center', size = 20, color = "white", weight = "bold")

#plot richest countries

c = mdf3.sort_values("GDP per capita", ascending = False).head(50)["crime index qu
intile"]

ax[0].scatter(y = mdf3.sort_values("GDP per capita", ascending = False).head(50)["
Happiness Score"], x = mdf3.sort_values("GDP per capita", ascending = False).head(
50)["GDP per capita"], alpha = 0.8, s = mdf3.sort_values("GDP per capita", ascendi
ng = False).head(50)["Crime Index"]* 15, c = c, cmap = "Blues")

ax[0].spines["right"].set_visible(False)
ax[0].spines["top"].set_visible(False)

ax[0].set_title("50 Richest Countries", fontsize = 18, fontweight = "bold")
ax[0].set_xlabel("GDP per Capita", fontsize = 15,)
ax[0].set_ylabel("Happiness Score", fontsize = 15,)

cr = mdf3.sort_values("GDP per capita", ascending = False).head(50)["Crime Index"]
.corr(mdf3.sort_values("GDP per capita", ascending = False).head(50)["GDP per capi
ta"])
message = "Wealth/Crime Correlation in Richest = " + str(round(cr,2))
ax[0].text(0,3.67, message, horizontalalignment='left', size = 15, color = "white"
)

# plot poorest countries

c = mdf3.sort_values("GDP per capita", ascending = False).tail(50)["crime index qu
intile"]

ax[1].scatter(y = mdf3.sort_values("GDP per capita", ascending = False).tail(50)["
Happiness Score"], x = mdf3.sort_values("GDP per capita", ascending = False).tail(
50)["GDP per capita"], alpha = 1, s = mdf3.sort_values("GDP per capita", ascending
= False).tail(50)["Crime Index"]* 15, c = c, cmap = "Blues")

ax[1].spines["right"].set_visible(False)
ax[1].spines["top"].set_visible(False)

ax[1].set_title("50 Poorest Countries", fontsize = 18, fontweight = "bold")
ax[1].set_xlabel("GDP per Capita", fontsize = 15,)
ax[1].set_ylabel("Happiness Score", fontsize = 15,)

cr = mdf3.sort_values("GDP per capita", ascending = False).tail(50)["GDP per capit
a"].corr(mdf3.sort_values("GDP per capita", ascending = False).tail(50)["Crime Ind
ex"])
message = "Wealth/Crime Correlation in Poorest = " + str(round(cr,2))
ax[1].text(0, 2.85, message, horizontalalignment='center', size = 15, color = "whi
te")

#legend = "Crime Index Quintiles:"
#fig.text(0, 0.8, legend, horizontalalignment='center', size = 13, color = "black"
, weight = "bold" )
#fig.text(0, 0.75, "Q5 - Yellow", horizontalalignment='center', size = 13, color =
"black")
#fig.text(0, 0.7, "Q4 - Green", horizontalalignment='center', size = 13, color = "
black")
#fig.text(0, 0.65, "Q3 - Teal", horizontalalignment='center', size = 13, color = "
black")
```



As clearly contrasted, the 50 richest countries on average, experience much less crime (shaded in lighter blue and white) than the 50 poorest countries that experience high crime (shaded in dark blue). Thus, we can deduce that crime is inversely correlated with happiness levels. This is certaintly expected considering the presence of violence and crime is not preferred when discussing relative levels of happiness.

What is the correlation between happiness and pollution?

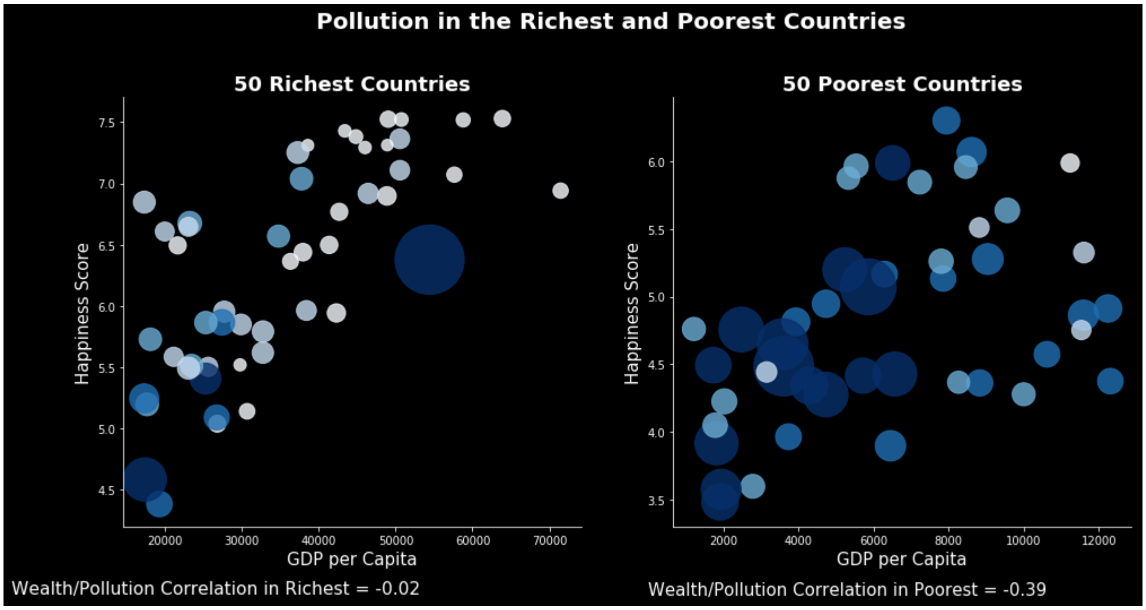
This is an especially interesting comparison in that individuals don't have any real control over the pollution rates they experience. Some of the most advanced countries are just as polluted as third world countries. As a result, the correlation between pollution and happiness might not be as strong as the previous factors.

In [34]:
```
pqdf
```

Out[34]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity	Dyst Resi
air pollution quintile									
0.2	26.956522	6.729826	1.338828	1.267135	0.828731	0.529981	0.264170	0.271026	2.22
0.4	50.636364	6.060561	1.139427	1.115243	0.724124	0.430373	0.143716	0.196450	2.31
0.6	78.391304	5.359304	0.831134	0.998676	0.536655	0.403878	0.161712	0.188280	2.23
0.8	92.984848	5.029439	0.914635	0.971093	0.566745	0.379613	0.112577	0.138110	1.94
1.0	98.101449	4.896935	0.743007	0.854410	0.449439	0.382271	0.197947	0.211807	2.05

In [35]:
```
fig, ax = plt.subplots(nrows = 1, ncols = 2, figsize = (16,7))

plt.style.use('dark_background')

fig.text(0.5, 1, "Pollution in the Richest and Poorest Countries", horizontalalign
ment='center', size = 20, color = "white", weight = "bold")


#plot richest countries

c = mdf3.sort_values("GDP per capita", ascending = False).head(50)["air pollution
quintile"]

ax[0].scatter(y = mdf3.sort_values("GDP per capita", ascending = False).head(50)["
Happiness Score"], x = mdf3.sort_values("GDP per capita", ascending = False).head(
50)["GDP per capita"], alpha = 0.8, s = mdf3.sort_values("GDP per capita", ascendi
ng = False).head(50)["Air Pollution (exposure)"]* 20, c = c, cmap = "Blues")

ax[0].spines["right"].set_visible(False)
ax[0].spines["top"].set_visible(False)

ax[0].set_title("50 Richest Countries", fontsize = 18, fontweight = "bold")
ax[0].set_xlabel("GDP per Capita", fontsize = 15,)
ax[0].set_ylabel("Happiness Score", fontsize = 15,)

cr = mdf3.sort_values("GDP per capita", ascending = False).head(50)["Air Pollution
(exposure)"].corr(mdf3.sort_values("GDP per capita", ascending = False).head(50)["
GDP per capita"])
message = "Wealth/Pollution Correlation in Richest = " + str(round(cr,2))
ax[0].text(0,3.65, message, horizontalalignment='left', size = 15.5, color = "whit
e")

# plot poorest countries

c = mdf3.sort_values("GDP per capita", ascending = False).tail(50)["air pollution
quintile"]

ax[1].scatter(y = mdf3.sort_values("GDP per capita", ascending = False).tail(50)["
Happiness Score"], x = mdf3.sort_values("GDP per capita", ascending = False).tail(
50)["GDP per capita"], alpha = 0.8, s = mdf3.sort_values("GDP per capita", ascendi
ng = False).tail(50)["Air Pollution (exposure)"]* 20, c = c, cmap = "Blues")

ax[1].spines["right"].set_visible(False)
ax[1].spines["top"].set_visible(False)

ax[1].set_title("50 Poorest Countries", fontsize = 18, fontweight = "bold")
ax[1].set_xlabel("GDP per Capita", fontsize = 15,)
ax[1].set_ylabel("Happiness Score", fontsize = 15,)

cr = mdf3.sort_values("GDP per capita", ascending = False).tail(50)["GDP per capit
a"].corr(mdf3.sort_values("GDP per capita", ascending = False).tail(50)["Air Pollu
tion (exposure)"])
message = "Wealth/Pollution Correlation in Poorest = " + str(round(cr,2))
ax[1].text(0, 2.8, message, horizontalalignment='left', size = 15, color = "white"
)

#legend = "Air Pollution Quintiles:"
#fig.text(0, 0.8, legend, horizontalalignment='center', size = 13, color = "black"
, weight = "bold" )
#fig.text(0, 0.75, "Q5 - Yellow", horizontalalignment='center', size = 13, color =
"black")
#fig.text(0, 0.7, "Q4 - Green", horizontalalignment='center', size = 13, color = "
black")
#fig.text(0, 0.65, "Q3 - Teal", horizontalalignment='center', size = 13, color = "
```



It seems like the poorest countries are much more polluted than the richest countries (Much more lighter shaded circles for richest countries versus poor countries). In fact, there is a very strong negative correlation between the amount of pollution and the country's average happiness level. The more pollution, the more displeasure there is with quality of life.

What is the correlation between happiness and education?

Last but not least, we observe how education levels affect the happiness scores in different countries. We anticipate that there will be a positive correlation between education levels and happiness because access throughought life is increased with education.

In [36]:
```
mdf4.head()
```

Out[36]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity	D F
country									
Switzerland	2.333333	7.530000	1.496273	1.337221	0.887530	0.623737	0.374120	0.314872	2
Denmark	2.000000	7.525000	1.416548	1.358481	0.820749	0.618266	0.427793	0.367957	2
Iceland	2.666667	7.522000	1.403204	1.398688	0.882907	0.607391	0.255580	0.355536	2
Norway	3.000000	7.519000	1.550968	1.330458	0.825889	0.633748	0.361601	0.347301	2
Finland	5.333333	7.429333	1.379934	1.331049	0.836393	0.610227	0.356414	0.290347	2

5 rows × 21 columns

In [37]:
```
mdf4["human capital quintile"] = pd.qcut(mdf4["Human Capital"], 5, labels = [.2, .
4, .6, 0.8, 1.0] , retbins=False, precision=3, duplicates='raise')
#split by hc index quantile
hcqdf = mdf4.groupby("human capital quintile").mean()
hcqdf.head()
```

Out[37]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity	Dysto Resid
human capital quintile									
0.2	107.009259	4.694287	0.639934	0.829850	0.419432	0.395463	0.169221	0.201759	2.038
0.4	74.777778	5.460167	0.923892	1.072408	0.568502	0.428316	0.136167	0.170930	2.159
0.6	56.814815	5.885741	1.113150	1.112627	0.680701	0.440010	0.141232	0.165839	2.232
0.8	46.981481	6.185130	1.228730	1.130346	0.777324	0.443869	0.185511	0.205582	2.213
1.0	39.037037	6.436259	1.315062	1.253145	0.797520	0.490337	0.233430	0.261075	2.085

5 rows × 21 columns

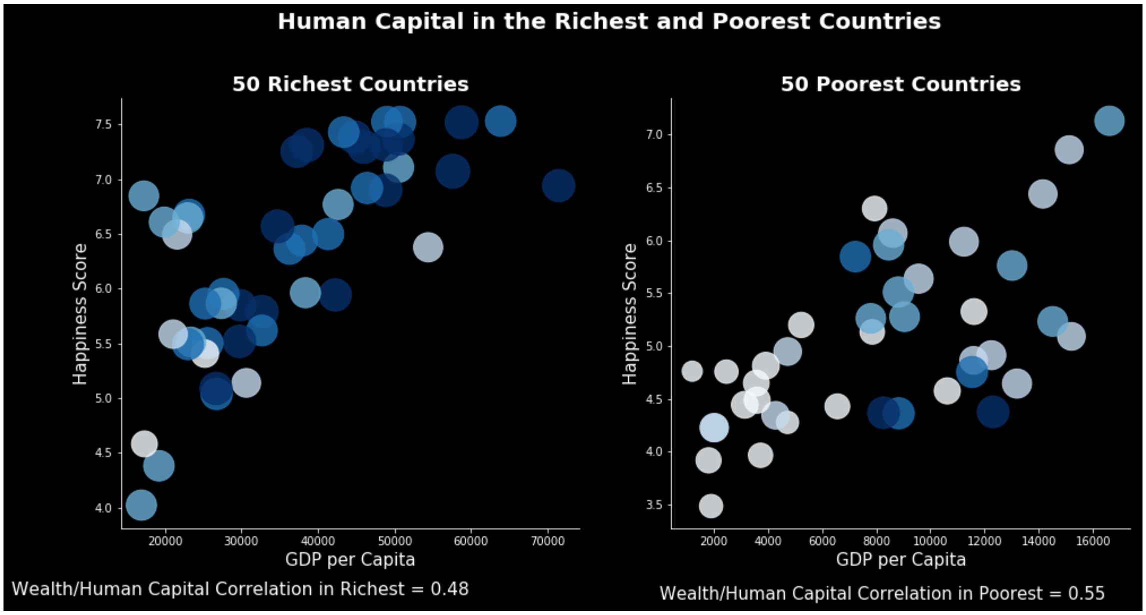
In [38]:
```
mdf4.head()
```

Out[38]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity	D F
country									
Switzerland	2.333333	7.530000	1.496273	1.337221	0.887530	0.623737	0.374120	0.314872	2
Denmark	2.000000	7.525000	1.416548	1.358481	0.820749	0.618266	0.427793	0.367957	2
Iceland	2.666667	7.522000	1.403204	1.398688	0.882907	0.607391	0.255580	0.355536	2
Norway	3.000000	7.519000	1.550968	1.330458	0.825889	0.633748	0.361601	0.347301	2
Finland	5.333333	7.429333	1.379934	1.331049	0.836393	0.610227	0.356414	0.290347	2

5 rows × 22 columns

In [39]:
```
fig, ax = plt.subplots(nrows = 1, ncols = 2, figsize = (16,7))

plt.style.use('dark_background')

fig.text(0.5, 1, "Human Capital in the Richest and Poorest Countries", horizontala
lignment='center', size = 20, color = "white", weight = "bold")


#plot richest countries

c = mdf4.sort_values("GDP per capita", ascending = False).head(50)["human capital
quintile"]

ax[0].scatter(y = mdf4.sort_values("GDP per capita", ascending = False).head(50)["
Happiness Score"], x = mdf4.sort_values("GDP per capita", ascending = False).head(
50)["GDP per capita"], alpha = 0.8, s = mdf4.sort_values("GDP per capita", ascendi
ng = False).head(50)["Human Capital"]* 250, c = c, cmap = "Blues")

ax[0].spines["right"].set_visible(False)
ax[0].spines["top"].set_visible(False)

ax[0].set_title("50 Richest Countries", fontsize = 18, fontweight = "bold")
ax[0].set_xlabel("GDP per Capita", fontsize = 15,)
ax[0].set_ylabel("Happiness Score", fontsize = 15,)

cr = mdf4.sort_values("GDP per capita", ascending = False).head(50)["Human Capital
"].corr(mdf4.sort_values("GDP per capita", ascending = False).head(50)["GDP per ca
pita"])
message = "Wealth/Human Capital Correlation in Richest = " + str(round(cr,2))
ax[0].text(0,3.2, message, horizontalalignment='left', size = 15.5, color = "white
")

# plot poorest countries

c = mdf4.sort_values("GDP per capita", ascending = False).tail(50)["human capital
quintile"]

ax[1].scatter(y = mdf4.sort_values("GDP per capita", ascending = False).tail(50)["
Happiness Score"], x = mdf4.sort_values("GDP per capita", ascending = False).tail(
50)["GDP per capita"], alpha = 0.8, s = mdf4.sort_values("GDP per capita", ascendi
ng = False).tail(50)["Human Capital"]* 250, c = c, cmap = "Blues")

ax[1].spines["right"].set_visible(False)
ax[1].spines["top"].set_visible(False)

ax[1].set_title("50 Poorest Countries", fontsize = 18, fontweight = "bold")
ax[1].set_xlabel("GDP per Capita", fontsize = 15,)
ax[1].set_ylabel("Happiness Score", fontsize = 15,)

cr = mdf4.sort_values("GDP per capita", ascending = False).tail(50)["GDP per capit
a"].corr(mdf4.sort_values("GDP per capita", ascending = False).tail(50)["Human Cap
ital"])
message = "Wealth/Human Capital Correlation in Poorest = " + str(round(cr,2))
ax[1].text(0, 2.6, message, horizontalalignment='left', size = 15, color = "white"
)

#legend = "Air Pollution Quintiles:"
#fig.text(0, 0.8, legend, horizontalalignment='center', size = 13, color = "black"
, weight = "bold" )
#fig.text(0, 0.75, "Q5 - Yellow", horizontalalignment='center', size = 13, color =
"black")
#fig.text(0, 0.7, "Q4 - Green", horizontalalignment='center', size = 13, color = "
black")
#fig.text(0, 0.65, "Q3 - Teal", horizontalalignment='center', size = 13, color = "
```



We can see that education levels are in fact directly correlated with the overall wellbeing (happiness) of the people in each country. To the left, we see that the richest countries experience an increasing level of happiness as the education level improves (more darker shaded points as happiness level increases relative to GDP). On the other side, the people belonging to countries with low human capital rates are forced to settle for the lowest levels of happiness. This is reflective of the terrible conditions these people are subjected to, with no real way to escape their circumstance due to lack of education.

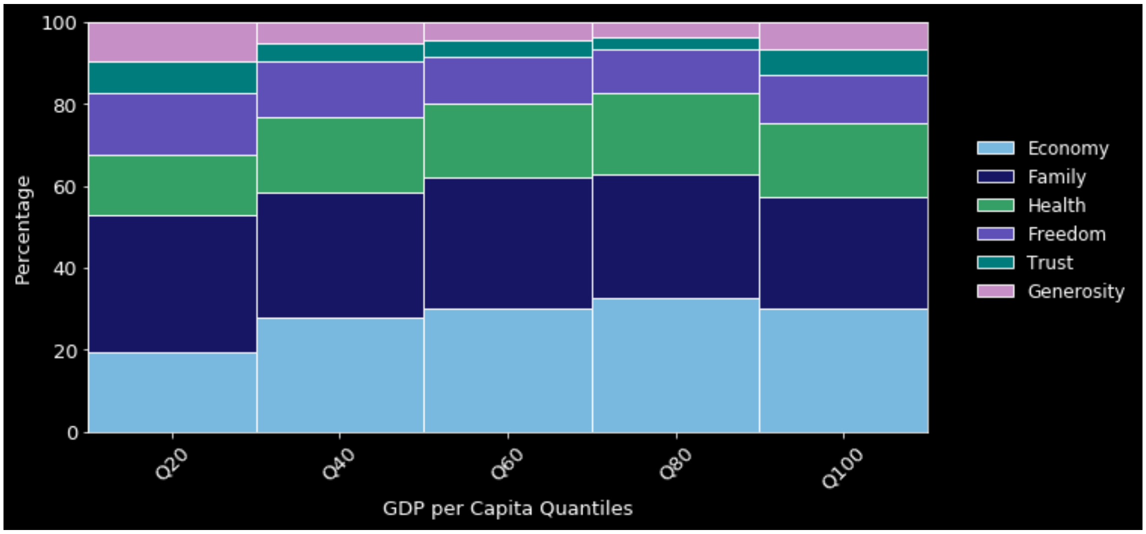
Distribution of Happiness Factors by GDP Segmentation

In [40]:
```
#Add column representing GDP per capita quantile to each row
mdf3["GDP per capita quantile"] = pd.qcut(mdf3["GDP per capita"], 5, labels = [.2, .4, .6, 0.8, 1.0] , retbins=False, precision=3, duplicates='raise')
#split by GDP per capita quantile
qdf = mdf3.groupby("GDP per capita quantile").mean()
qdf.head()
```

Out[40]:

	Happiness Rank	Happiness Score	Economy	Family	Health	Freedom	Trust	Generosity	Dyst Resi
GDP per capita quantile									
0.2	114.613636	4.513947	0.477369	0.813832	0.365620	0.367133	0.185747	0.234240	2.06
0.4	84.873016	5.189127	0.851707	0.932996	0.566006	0.413959	0.138218	0.153305	2.13
0.6	73.045455	5.518591	1.041549	1.096911	0.620880	0.394541	0.135027	0.157230	2.07
0.8	57.031746	5.868302	1.218841	1.124316	0.750324	0.392597	0.109630	0.144072	2.12
1.0	16.772727	7.007182	1.409337	1.278066	0.840379	0.549477	0.296474	0.314023	2.31

In [41]:
```
# code adapted from: https://chrisalbon.com/python/data_visualization/matplotlib_p
ercentage_stacked_bar_plot/

# Create a figure with a single subplot
f, ax = plt.subplots(1, figsize=(10,5))

# Set bar width at 1
bar_width = 1

# positions of the left bar-boundaries
bar_l = [i for i in range(len(qdf.index))]

# positions of the x-axis ticks (center of the bars as bar labels)
tick_pos = [i+(bar_width/2) for i in bar_l]

# Create the total score for each quartile
totals = [i+j+k+m+n+o for i,j,k,m,n,o in zip(qdf['Economy'],
                                 qdf['Family'],
                                 qdf['Health'],
                                 qdf['Freedom'],
                                 qdf['Trust'],
                                 qdf['Generosity'])]

# Create the percentage of the total score the Economy value for each quartile was
Econ_rel = [i / j * 100 for  i,j in zip(qdf['Economy'], totals)]

# Create the percentage of the total score the Family value for each quartile was
Fam_rel = [i / j * 100 for  i,j in zip(qdf['Family'], totals)]

# Create the percentage of the total score the Health value for each quartile was
Health_rel = [i / j * 100 for  i,j in zip(qdf['Health'], totals)]

# Create the percentage of the total score the Economy value for each quartile was
Free_rel = [i / j * 100 for  i,j in zip(qdf['Freedom'], totals)]

# Create the percentage of the total score the Trust value for each quartile was
Trust_rel = [i / j * 100 for  i,j in zip(qdf['Trust'], totals)]

# Create the percentage of the total score the Generosity value for each quartile
was
Gen_rel = [i / j * 100 for  i,j in zip(qdf['Generosity'], totals)]

# Create a bar chart in position bar_1
ax.bar(bar_l,
       # using pre_rel data
       Econ_rel,
       # labeled
       label='Economy',
       # with alpha
       alpha=0.9,
       # with color
       color='#87CEFA',
       # with bar width
       width=bar_width,
       # with border color
       edgecolor='white'
       )

# Create a bar chart in position bar_1
ax.bar(bar_l,
       # using mid_rel data
       Fam_rel,
       # with pre_rel
       bottom=Econ_rel,
```



After segmenting the countries relative to GDP per Capita, we can observe some interesting facts. Particularly, as expected countries in varying tiers of GDP value elements of well-being more or less than others.

In the Q20, generosity and trust play a more significant role in overall happiness than in all of the other quartiles. This is an interesting discovery considering the people in these countries have less, but are willing to give more it seems. This disposition of positive impact is further supported by the fact that Q20 countries place significant emphasis on family when weighting happiness.

As expected, the factor of economy and its role in determing happiness, is positively correlated with a countries GDP. Q2O place much less emphasis on economy as compared to Q40, Q60 and Q100. Further, it seems that there is a depreciation of the value of economy with regard to happiness, when the GDP of a country exceeds a certain amount. In this case, Q100 countries are concerned less with Economy than Q80 countries.

On the more individual side of well-being, all countries across the spectrum appear to place mostly equal emphasis on the importance of health and freedom in life. We can see that these factors are not the most crucial, but certaintly can not be excluded from determing happiness levels.

Conclusion

This analysis has provided us with a strong grasp on the plethora of factors that go into determing the overall happiness of countries, and the world at large. We learn that happiness is indeed difficult to control but there are certain necessities that have to be present - economy, health, and family. Others, are negotiable but contribute greatly to the overall level of happiness - generosity, freedom, and trust of government.

The most interesting and relevent takeaways include:

- At lower levels of wealth, increases in GDP are only marginally additive to happiness at best, and inconsequential at worst. This is likely due to the fact that the state of being poor requires a significant investment to properly overcome.
- Generosity and trust of government are regarded as the least crucial when determining overall happiness. Health and freedom are considered essential parts of maintining a happy life, yet still valued less than economy and family.
- Other factors like crime, pollution and education levels play a signficant role in the overall well-being of an individual. As could be expected education is positively correlated with happiness, while crime and pollution were inversely correlated.

In the end, we observe that perhaps education is one of the leading factors of why people end happy or unhappy. There was a very strong correlation to support the notion that lack of education is the leading catalyst of unhappiness around world. With this in mind, it is important that policy makers around the world prioritize increasing education efforts and accessibility to maximize the overall quality of life.