UNICEF Research

October 2, 2018

```
In [1]: import pandas as pd
        import numpy as np
        import math
        import matplotlib.pyplot as plt
```

1 Load "World Bank indicators"

```
In [2]: df = pd.read_csv("Indicators.csv")

In [3]: df.head()

Out[3]:   CountryName CountryCode                                      IndicatorName  \
        0  Arab World         ARB  Adolescent fertility rate (births per 1,000 wo...
        1  Arab World         ARB  Age dependency ratio (% of working-age populat...
        2  Arab World         ARB  Age dependency ratio, old (% of working-age po...
        3  Arab World         ARB  Age dependency ratio, young (% of working-age ...
        4  Arab World         ARB        Arms exports (SIPRI trend indicator values)

             IndicatorCode  Year         Value
        0      SP.ADO.TFRT  1960  1.335609e+02
        1      SP.POP.DPND  1960  8.779760e+01
        2   SP.POP.DPND.OL  1960  6.634579e+00
        3   SP.POP.DPND.YG  1960  8.102333e+01
        4   MS.MIL.XPRT.KD  1960  3.000000e+06

In [4]: df.groupby("IndicatorName").aggregate(np.mean)

Out[4]:                                                              Year         Value
        IndicatorName
        2005 PPP conversion factor, GDP (LCU per intern...  2005.000000  2.122996e+02
        2005 PPP conversion factor, private consumption...  2005.000000  2.499051e+02
        ARI treatment (% of children under 5 taken to a...  2003.880519  5.427963e+01
        Access to electricity (% of population)             2003.000000  7.451074e+01
        Access to electricity, rural (% of rural popula...  2003.000000  6.558007e+01
        Access to electricity, urban (% of urban popula...  2003.000000  8.644389e+01
        Access to non-solid fuel (% of population)          2003.000000  6.209638e+01
        Access to non-solid fuel, rural (% of rural pop...  2011.000000  5.644107e+01
        Access to non-solid fuel, urban (% of urban pop...  2011.000000  7.743126e+01
```

Adequacy of social insurance programs (% of tot...	2008.754286	3.298501e+01
Adequacy of social protection and labor program...	2008.778378	2.575662e+01
Adequacy of social safety net programs (% of to...	2008.844311	1.029080e+01
Adequacy of unemployment benefits and ALMP (% o...	2008.970149	1.312885e+01
Adjusted net enrolment rate, primary, both sexe...	1994.795947	8.440404e+01
Adjusted net enrolment rate, primary, female (%)	1995.156570	8.126404e+01
Adjusted net enrolment rate, primary, male (%)	1995.159403	8.498042e+01
Adjusted net national income (annual % growth)	1994.759665	3.505781e+00
Adjusted net national income (constant 2005 US$)	1994.538680	1.502742e+12
Adjusted net national income (current US$)	1993.492742	9.168502e+11
Adjusted net national income per capita (annual...	1994.759665	1.829301e+00
Adjusted net national income per capita (consta...	1994.538680	7.370960e+03
Adjusted net national income per capita (curren...	1993.492942	4.976916e+03
Adjusted net savings, excluding particulate emi...	1995.163501	7.942871e+00
Adjusted net savings, excluding particulate emi...	1995.417230	2.566986e+10
Adjusted net savings, including particulate emi...	2006.885653	7.042000e+00
Adjusted net savings, including particulate emi...	2002.345361	3.288440e+10
Adjusted savings: carbon dioxide damage (% of GNI)	1993.215880	5.830795e-01
Adjusted savings: carbon dioxide damage (curren...	1992.454119	7.024923e+08
Adjusted savings: consumption of fixed capital ...	1993.285122	1.025734e+01
Adjusted savings: consumption of fixed capital ...	1993.446493	2.408672e+10
...	...	...
Unmet need for contraception (% of married wome...	2002.808399	1.943119e+01
Urban population	1987.078884	6.240680e+07
Urban population (% of total)	1987.055908	4.903390e+01
Urban population growth (annual %)	1987.146285	3.016630e+00
Urban poverty gap at national poverty lines (%)	2007.492126	1.093756e+01
Urban poverty headcount ratio at national pover...	2006.456763	2.584106e+01
Use of IMF credit (DOD, current US$)	1993.079669	1.774858e+09
Use of insecticide-treated bed nets (% of under...	2006.478495	2.254221e+01
Value lost due to electrical outages (% of sales)	2010.146825	4.768089e+00
Vitamin A supplementation coverage rate (% of c...	2006.455142	7.282242e+01
Vulnerable employment, female (% of female empl...	2002.357379	2.703239e+01
Vulnerable employment, male (% of male employment)	2002.357379	2.656795e+01
Vulnerable employment, total (% of total employ...	2002.312133	2.736432e+01
Wage and salaried workers, female (% of females...	2001.465089	7.326770e+01
Wage and salaried workers, total (% of total em...	2001.187978	7.133710e+01
Wage and salary workers, male (% of males emplo...	2001.465089	7.047066e+01
Wanted fertility rate (births per woman)	2000.861423	3.516273e+00
Water productivity, total (constant 2005 US$ GD...	2000.060574	3.393002e+01
Wholesale price index (2010 = 100)	1991.708541	5.503397e+01
Women who believe a husband is justified in bea...	2007.424658	4.141370e+01
Women who believe a husband is justified in bea...	2007.409722	2.408958e+01
Women who believe a husband is justified in bea...	2007.429577	1.458239e+01
Women who believe a husband is justified in bea...	2007.424658	2.759247e+01
Women who believe a husband is justified in bea...	2007.424658	3.103151e+01
Women who believe a husband is justified in bea...	2007.433566	2.021958e+01
Women's share of population ages 15+ living wit...	2001.936422	4.251081e+01

```
Youth literacy rate, population 15-24 years, bo...  2001.238030  8.677694e+01
Youth literacy rate, population 15-24 years, fe...  2001.265110  8.441812e+01
Youth literacy rate, population 15-24 years, ge...  2001.265110  9.264546e-01
Youth literacy rate, population 15-24 years, ma...  2001.265110  8.920173e+01

[1344 rows x 2 columns]
```

2 Load vaccination coverage database

In [5]:
```
vac = pd.read_csv(".csv")
```

In [6]:
```
vac = vac[vac["Vaccine"]=="PCV3"]
vac = vac[vac["Year"]==2016]
vac.head()
```

Out[6]:
```
                    Unicef Region         Name  Year Vaccine  Coverage  \
136                    South Asia  Afghanistan  2016    PCV3        65
331  Eastern Europe & Central Asia      Albania  2016    PCV3        98
499     Middle East & North Africa      Algeria  2016    PCV3        61
675                Western Europe      Andorra  2016    PCV3        92
834      Eastern & Southern Africa       Angola  2016    PCV3        59

    Unvaccinated     Target                                             Source
136      379,000  1,083,000  WHO/UNICEF estimates of national immunization ...
331       <1,000     34,000  WHO/UNICEF estimates of national immunization ...
499      356,000    912,000  WHO/UNICEF estimates of national immunization ...
675         <100     <1,000  WHO/UNICEF estimates of national immunization ...
834      465,000  1,134,000  WHO/UNICEF estimates of national immunization ...
```

3 Load Vaccination Survialance % database

In [7]:
```
survive = pd.read_csv("survive_pcv3.csv")
```

In [8]:
```
survive[survive["country"]=="Kazakhstan"]
```

Out[8]:
```
   iso3     country vaccine  2016
88  KAZ  Kazakhstan    PCV3  97.0
```

4 Load Death of Pneumonia database

In [9]:
```
death = pd.read_csv("death_pneumo.csv")
```

In [10]:
```
death[death["CountryName"]=="Kazakhstan"]
```

Out[10]:
```
    ISO Code CountryName  Year  Pneumonia
87       KAZ  Kazakhstan  2016       13.1
281      KAZ  Kazakhstan  2000       18.0
```

In [11]:
```python
death.columns = ["code","country","year","death"]
```

In [12]:
```python
death.head()
```

Out[12]:

	code	country	year	death
0	AFG	Afghanistan	2016	16.1
1	ALB	Albania	2016	10.2
2	DZA	Algeria	2016	13.0
3	AND	Andorra	2016	0.0
4	AGO	Angola	2016	17.8

5 Add "Survive" column from survialance database

In [13]:
```python
a = []
for i,row in death.iterrows():
    val = None
    surv = survive[survive["country"]==row["country"]]
    if len(surv) > 0:
        val = surv.iloc[0]["2016"]
    a.append(val)
death["survive"] = a
```

In [14]:
```python
death.head()
```

Out[14]:

	code	country	year	death	survive
0	AFG	Afghanistan	2016	16.1	65.0
1	ALB	Albania	2016	10.2	98.0
2	DZA	Algeria	2016	13.0	61.0
3	AND	Andorra	2016	0.0	92.0
4	AGO	Angola	2016	17.8	59.0

6 Add "Vaccinated" column from vaccinated database

In [16]:
```python
a = []
for i,row in death.iterrows():
    val = None
    vac1 = vac[vac["Name"]==row["country"]]
    if len(vac1) > 0:
        val = vac1.iloc[0]["Coverage"]
    a.append(val)
death["vaccinated"] = a
```

In [17]:
```python
death[death["country"]=="Kazakhstan"]
```

Out[17]:

	code	country	year	death	survive	vaccinated
87	KAZ	Kazakhstan	2016	13.1	97.0	97.0
281	KAZ	Kazakhstan	2000	18.0	97.0	97.0

We see that Survialance and Vaccination coverage databases completely coincide

```
In [18]: death = death[death["year"]==2016]
```

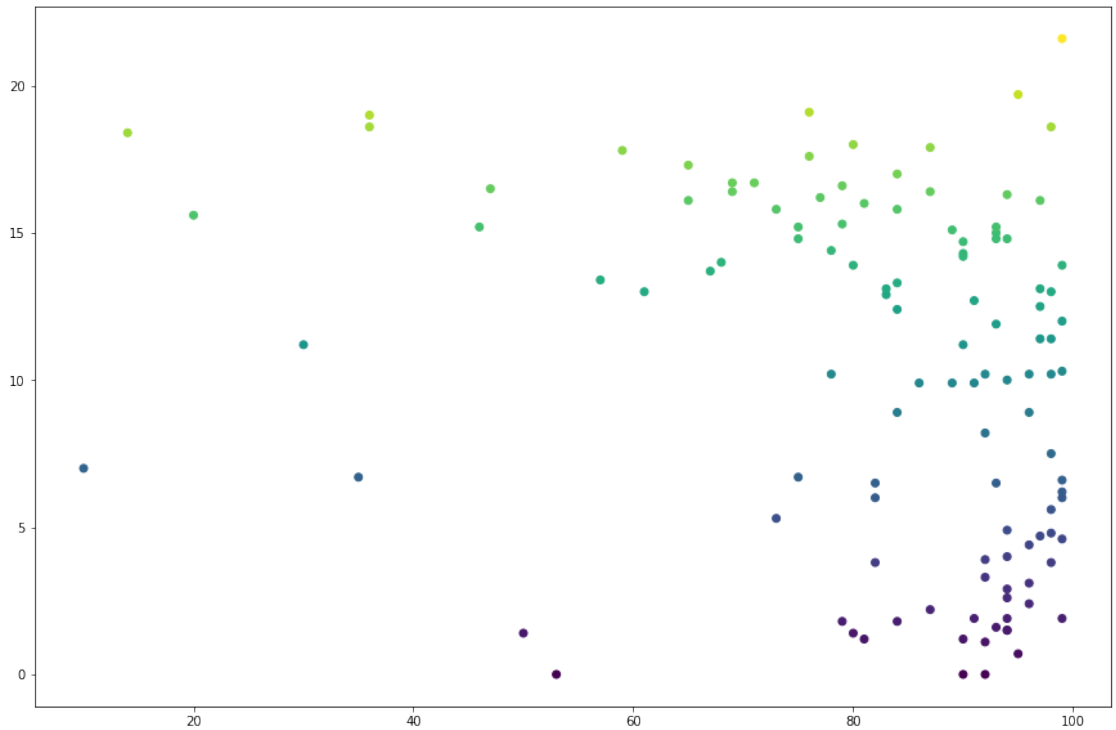
```
In [19]: death = death.dropna()
```

```
In [20]: death[death["country"]=="Kazakhstan"]
```

```
Out[20]:    code     country  year  death  survive  vaccinated
         87  KAZ  Kazakhstan  2016   13.1     97.0        97.0
```

```
In [23]: fig = plt.figure(figsize=(15,10))
         plt.scatter(death["vaccinated"],death["death"], c=death["death"])
```

```
Out[23]: <matplotlib.collections.PathCollection at 0x7fb80f141cc0>
```



We see that correlation is not high

```
In [24]: kaz = df[df["CountryName"]=="Kazakhstan"]
         kaz = kaz[kaz["Year"]==2015]
         #look at all indicators
         kaz
```

```
Out[24]:         CountryName CountryCode  \
         5649400  Kazakhstan         KAZ
         5649401  Kazakhstan         KAZ
```

5649402	Kazakhstan	KAZ
5649403	Kazakhstan	KAZ
5649404	Kazakhstan	KAZ
5649405	Kazakhstan	KAZ
5649406	Kazakhstan	KAZ
5649407	Kazakhstan	KAZ
5649408	Kazakhstan	KAZ
5649409	Kazakhstan	KAZ
5649410	Kazakhstan	KAZ
5649411	Kazakhstan	KAZ
5649412	Kazakhstan	KAZ
5649413	Kazakhstan	KAZ
5649414	Kazakhstan	KAZ
5649415	Kazakhstan	KAZ
5649416	Kazakhstan	KAZ
5649417	Kazakhstan	KAZ
5649418	Kazakhstan	KAZ
5649419	Kazakhstan	KAZ
5649420	Kazakhstan	KAZ
5649421	Kazakhstan	KAZ
5649422	Kazakhstan	KAZ
5649423	Kazakhstan	KAZ
5649424	Kazakhstan	KAZ
5649425	Kazakhstan	KAZ
5649426	Kazakhstan	KAZ
5649427	Kazakhstan	KAZ
5649428	Kazakhstan	KAZ
5649429	Kazakhstan	KAZ
...	...	...
5649449	Kazakhstan	KAZ
5649450	Kazakhstan	KAZ
5649451	Kazakhstan	KAZ
5649452	Kazakhstan	KAZ
5649453	Kazakhstan	KAZ
5649454	Kazakhstan	KAZ
5649455	Kazakhstan	KAZ
5649456	Kazakhstan	KAZ
5649457	Kazakhstan	KAZ
5649458	Kazakhstan	KAZ
5649459	Kazakhstan	KAZ
5649460	Kazakhstan	KAZ
5649461	Kazakhstan	KAZ
5649462	Kazakhstan	KAZ
5649463	Kazakhstan	KAZ
5649464	Kazakhstan	KAZ
5649465	Kazakhstan	KAZ
5649466	Kazakhstan	KAZ
5649467	Kazakhstan	KAZ

```
5649468  Kazakhstan         KAZ
5649469  Kazakhstan         KAZ
5649470  Kazakhstan         KAZ
5649471  Kazakhstan         KAZ
5649472  Kazakhstan         KAZ
5649473  Kazakhstan         KAZ
5649474  Kazakhstan         KAZ
5649475  Kazakhstan         KAZ
5649476  Kazakhstan         KAZ
5649477  Kazakhstan         KAZ
5649478  Kazakhstan         KAZ

                                              IndicatorName      IndicatorCode  \
5649400                      Bank capital to assets ratio (%)     FB.BNK.CAPA.ZS
5649401  Bank nonperforming loans to total gross loans (%)     FB.AST.NPER.ZS
5649402                              Bird species, threatened     EN.BIR.THRD.NO
5649403  Business extent of disclosure index (0=less di...     IC.BUS.DISC.XQ
5649404  Cost of business start-up procedures (% of GNI...  IC.REG.COST.PC.ZS
5649405  Depth of credit information index (0=low to 8=...     IC.CRD.INFO.XQ
5649406  Disbursements on external debt, long-term (DIS...     DT.DIS.DLXF.CD
5649407  Distance to frontier score (0=lowest performan...     IC.BUS.DFRN.XQ
5649408  Ease of doing business index (1=most business-...     IC.BUS.EASE.XQ
5649409                              Fish species, threatened     EN.FSH.THRD.NO
5649410  Improved sanitation facilities (% of populatio...        SH.STA.ACSN
5649411  Improved sanitation facilities, rural (% of ru...     SH.STA.ACSN.RU
5649412  Improved sanitation facilities, urban (% of ur...     SH.STA.ACSN.UR
5649413  Improved water source (% of population with ac...     SH.H2O.SAFE.ZS
5649414  Improved water source, rural (% of rural popul...  SH.H2O.SAFE.RU.ZS
5649415  Improved water source, urban (% of urban popul...  SH.H2O.SAFE.UR.ZS
5649416  Interest payments on external debt, long-term ...     DT.INT.DLXF.CD
5649417  Interest payments on external debt, private no...     DT.INT.DPNG.CD
5649418  Labor tax and contributions (% of commercial p...  IC.TAX.LABR.CP.ZS
5649419                  Lifetime risk of maternal death (%)     SH.MMR.RISK.ZS
5649420  Lifetime risk of maternal death (1 in: rate va...        SH.MMR.RISK
5649421                            Mammal species, threatened     EN.MAM.THRD.NO
5649422  Maternal mortality ratio (modeled estimate, pe...        SH.STA.MMRT
5649423  Methodology assessment of statistical capacity...        IQ.SCI.MTHD
5649424     Mortality rate, infant (per 1,000 live births)     SP.DYN.IMRT.IN
5649425  Mortality rate, infant, female (per 1,000 live...  SP.DYN.IMRT.FE.IN
5649426  Mortality rate, infant, male (per 1,000 live b...  SP.DYN.IMRT.MA.IN
5649427   Mortality rate, neonatal (per 1,000 live births)        SH.DYN.NMRT
5649428                  Mortality rate, under-5 (per 1,000)        SH.DYN.MORT
5649429  Mortality rate, under-5, female (per 1,000 liv...     SH.DYN.MORT.FE
...                                                    ...                ...
5649449                PPG, multilateral (INT, current US$)     DT.INT.MLAT.CD
5649450          PPG, official creditors (AMT, current US$)     DT.AMT.OFFT.CD
5649451          PPG, official creditors (DIS, current US$)     DT.DIS.OFFT.CD
5649452          PPG, official creditors (INT, current US$)     DT.INT.OFFT.CD
```

5649453	PPG, other private creditors (AMT, current US$)	DT.AMT.PROP.CD
5649454	PPG, other private creditors (DIS, current US$)	DT.DIS.PROP.CD
5649455	PPG, other private creditors (INT, current US$)	DT.INT.PROP.CD
5649456	PPG, private creditors (AMT, current US$)	DT.AMT.PRVT.CD
5649457	PPG, private creditors (DIS, current US$)	DT.DIS.PRVT.CD
5649458	PPG, private creditors (INT, current US$)	DT.INT.PRVT.CD
5649459	Principal repayments on external debt, long-te...	DT.AMT.DLXF.CD
5649460	Principal repayments on external debt, private...	DT.AMT.DPNG.CD
5649461	Private credit bureau coverage (% of adults)	IC.CRD.PRVT.ZS
5649462	Procedures to build a warehouse (number)	IC.WRH.PROC
5649463	Procedures to register property (number)	IC.PRP.PROC
5649464	Profit tax (% of commercial profits)	IC.TAX.PRFT.CP.ZS
5649465	Proportion of seats held by women in national ...	SG.GEN.PARL.ZS
5649466	Public credit registry coverage (% of adults)	IC.CRD.PUBL.ZS
5649467	Source data assessment of statistical capacity...	IQ.SCI.SRCE
5649468	Start-up procedures to register a business (nu...	IC.REG.PROC
5649469	Strength of legal rights index (0=weak to 12=s...	IC.LGL.CRED.XQ
5649470	Tax payments (number)	IC.TAX.PAYM
5649471	Time required to build a warehouse (days)	IC.WRH.DURS
5649472	Time required to enforce a contract (days)	IC.LGL.DURS
5649473	Time required to get electricity (days)	IC.ELC.TIME
5649474	Time required to register property (days)	IC.PRP.DURS
5649475	Time required to start a business (days)	IC.REG.DURS
5649476	Time to prepare and pay taxes (hours)	IC.TAX.DURS
5649477	Time to resolve insolvency (years)	IC.ISV.DURS
5649478	Total tax rate (% of commercial profits)	IC.TAX.TOTL.CP.ZS

	Year	Value
5649400	2015	1.408464e+01
5649401	2015	1.237220e+01
5649402	2015	2.600000e+01
5649403	2015	9.000000e+00
5649404	2015	1.000000e-01
5649405	2015	7.000000e+00
5649406	2015	2.417710e+08
5649407	2015	7.268000e+01
5649408	2015	4.100000e+01
5649409	2015	1.400000e+01
5649410	2015	9.750000e+01
5649411	2015	9.810000e+01
5649412	2015	9.700000e+01
5649413	2015	9.290000e+01
5649414	2015	8.560000e+01
5649415	2015	9.940000e+01
5649416	2015	5.465150e+08
5649417	2015	5.300000e+07
5649418	2015	1.120000e+01
5649419	2015	3.369779e-02

```
5649420  2015  3.000000e+03
5649421  2015  1.500000e+01
5649422  2015  1.200000e+01
5649423  2015  9.000000e+01
5649424  2015  1.260000e+01
5649425  2015  1.070000e+01
5649426  2015  1.450000e+01
5649427  2015  7.000000e+00
5649428  2015  1.410000e+01
5649429  2015  1.200000e+01
...       ...           ...
5649449  2015  2.865100e+07
5649450  2015  1.491500e+08
5649451  2015  2.417710e+08
5649452  2015  3.863800e+07
5649453  2015  0.000000e+00
5649454  2015  0.000000e+00
5649455  2015  6.790000e+05
5649456  2015  0.000000e+00
5649457  2015  0.000000e+00
5649458  2015  4.548770e+08
5649459  2015  1.491500e+08
5649460  2015  0.000000e+00
5649461  2015  8.140000e+01
5649462  2015  2.400000e+01
5649463  2015  3.000000e+00
5649464  2015  1.620000e+01
5649465  2015  2.616822e+01
5649466  2015  0.000000e+00
5649467  2015  9.000000e+01
5649468  2015  4.000000e+00
5649469  2015  4.000000e+00
5649470  2015  7.000000e+00
5649471  2015  1.540000e+02
5649472  2015  3.700000e+02
5649473  2015  8.300000e+01
5649474  2015  4.500000e+00
5649475  2015  5.000000e+00
5649476  2015  1.880000e+02
5649477  2015  1.500000e+00
5649478  2015  2.920000e+01

[79 rows x 6 columns]
```

7 Try to bind to SH.STA.ACSN

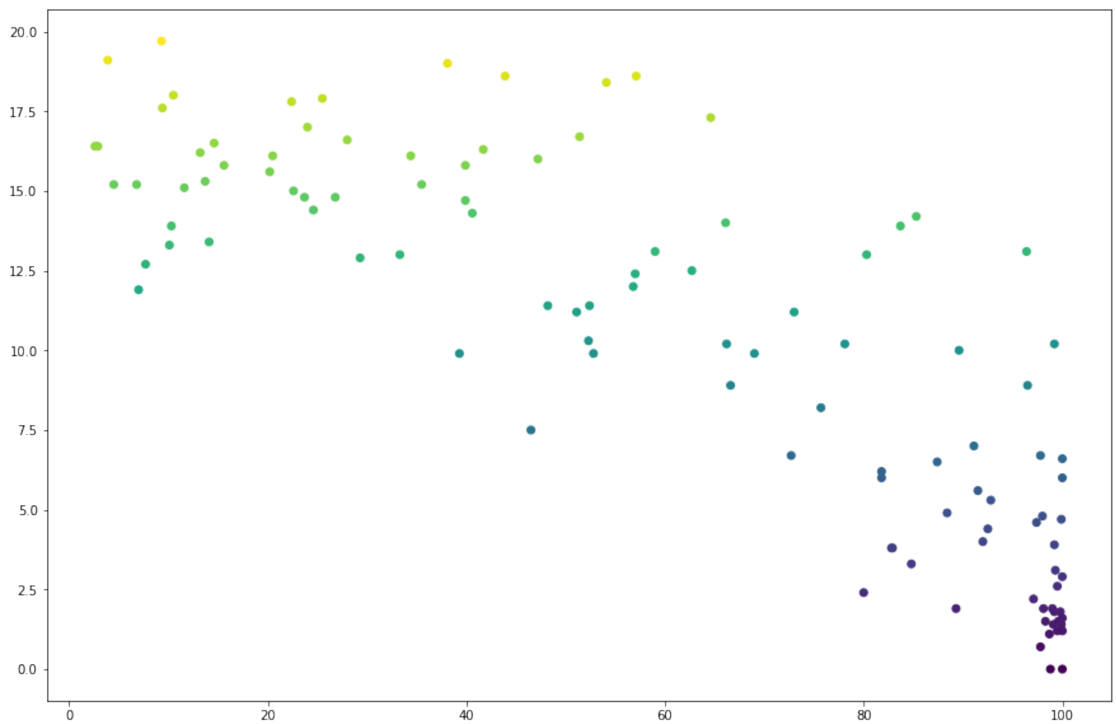
7.1 We have chosen Improved sanitation facilities field to check correelation

```
In [26]: acsn = df[df["IndicatorCode"]=="SH.STA.ACSN"]
         a = []
         for i,row in death.iterrows():
             val = None
             san = acsn[acsn["CountryName"]==row["country"]]
             if len(san) > 0:
                 val = san.iloc[0]["Value"]
             a.append(val)
         death["Sanitary"] = a
```

```
In [27]: death = death.dropna()
```

```
In [30]: fig = plt.figure(figsize=(15,10))
         plt.scatter(death["Sanitary"],death["death"],c=death["death"])
```

```
Out[30]: <matplotlib.collections.PathCollection at 0x7fb80e437f28>
```



We see a strong correlation between Improved sanitation facilities rate and Death from Pneumonia

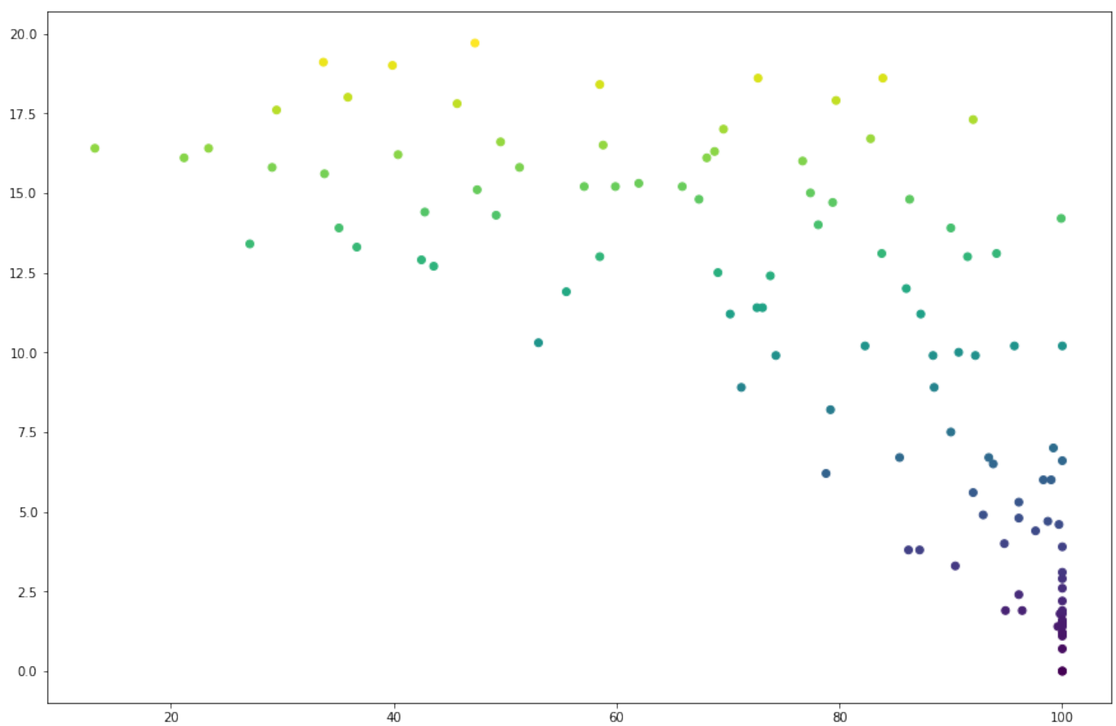
7.2 Next we have chosen Improved water source field to check correlation

```
In [32]: h2o = df[df["IndicatorCode"]=="SH.H2O.SAFE.ZS"]
         a = []
         for i,row in death.iterrows():
             val = None
             san = h2o[h2o["CountryName"]==row["country"]]
             if len(san) > 0:
                 val = san.iloc[0]["Value"]
             a.append(val)
         death["H2O"] = a
```

```
In [33]: death = death.dropna()
```

```
In [34]: fig = plt.figure(figsize=(15,10))
         plt.scatter(death["H2O"],death["death"],c=death["death"])
```

```
Out[34]: <matplotlib.collections.PathCollection at 0x7fb80e418940>
```



Here we also see a strong correlation between Improved water source and Death from Pneumonia

8 Load GDP database and check the correlation

```
In [35]: gdp = pd.read_csv("gdp.csv")
```

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

Out[36]:

	country	gdp
0	Afghanistan	20,608,089,735
1	Albania	11,335,262,227
2	Algeria	165,874,275,965
3	Andorra	2,811,617,372
4	Angola	115,143,205,131

In [37]:
```
a = []
for i,row in death.iterrows():
    val = None
    san = gdp[gdp["country"]==row["country"]]
    if len(san) > 0:
        val = san.iloc[0]["gdp"]
    a.append(val)
death["gdp"] = a
```

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

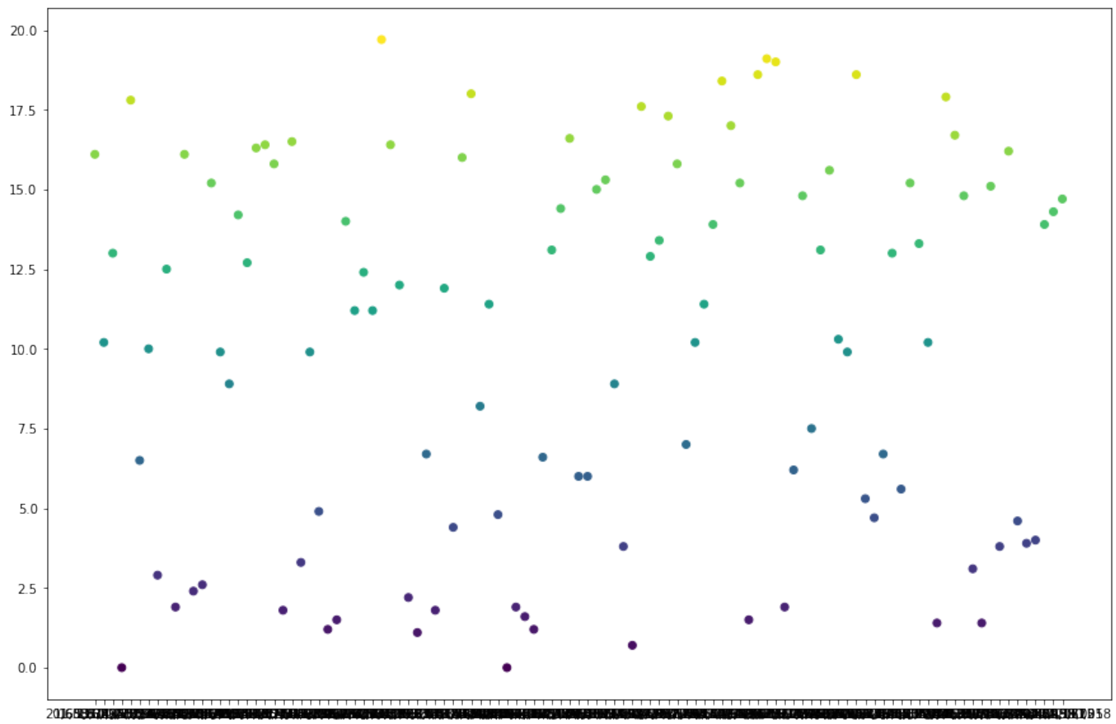
Out[38]:

	code	country	year	death	survive	vaccinated	Sanitary	H2O	\
0	AFG	Afghanistan	2016	16.1	65.0	65.0	20.5	21.2	
1	ALB	Albania	2016	10.2	98.0	98.0	78.1	95.7	
2	DZA	Algeria	2016	13.0	61.0	61.0	80.3	91.5	
3	AND	Andorra	2016	0.0	92.0	92.0	100.0	100.0	
4	AGO	Angola	2016	17.8	59.0	59.0	22.4	45.7	

	gdp
0	20,608,089,735
1	11,335,262,227
2	165,874,275,965
3	2,811,617,372
4	115,143,205,131

In [40]:
```
fig = plt.figure(figsize=(15,10))
plt.scatter(death["gdp"],death["death"],c=death["death"])
```

Out[40]: <matplotlib.collections.PathCollection at 0x7fb80b17b0b8>



We could not find any correlation between GDP and Pneumonia death