

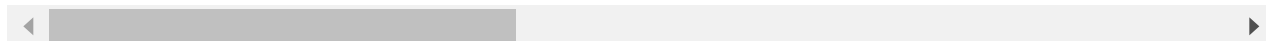
```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [2]: df = pd.read_csv('number-of-deaths-by-risk-factor.csv')
df.head()
```

Out[2]:

	Entity	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontinue breastfeeding
0	Afghanistan	1990	7554.049543	5887.747628	5412.314513	22388.49723	3221.138842	156.09755
1	Afghanistan	1991	7359.676749	5732.770160	5287.891103	22128.75821	3150.559597	151.53985
2	Afghanistan	1992	7650.437822	5954.804987	5506.657363	22873.76879	3331.349048	156.60919
3	Afghanistan	1993	10270.731380	7986.736613	7104.620351	25599.75628	4477.006100	206.83445
4	Afghanistan	1994	11409.177110	8863.010065	8051.515953	28013.16720	5102.622054	233.93057

5 rows × 31 columns



```
In [3]: df.columns
```

```
Out[3]: Index(['Entity', 'Year', 'Unsafe water source', 'Unsafe sanitation',
              'No access to handwashing facility',
              'Household air pollution from solid fuels',
              'Non-exclusive breastfeeding', 'Discontinued breastfeeding',
              'Child wasting', 'Child stunting', 'Low birth weight for gestation',
              'Secondhand smoke', 'Alcohol use', 'Drug use', 'Diet low in fruits',
              'Diet low in vegetables', 'Unsafe sex', 'Low physical activity',
              'High fasting plasma glucose', 'High total cholesterol',
              'High body-mass index', 'High systolic blood pressure', 'Smoking',
              'Iron deficiency', 'Vitamin A deficiency', 'Low bone mineral density',
              'Air pollution', 'Outdoor air pollution', 'Diet high in sodium',
              'Diet low in whole grains', 'Diet low in nuts and seeds'],
              dtype='object')
```

```
In [5]: df.shape
```

Out[5]: (6468, 31)

```
In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6468 entries, 0 to 6467
Data columns (total 31 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Entity                                6468 non-null   object
```

1	Year	6468 non-null	int64
2	Unsafe water source	6468 non-null	float64
3	Unsafe sanitation	6468 non-null	float64
4	No access to handwashing facility	6468 non-null	float64
5	Household air pollution from solid fuels	6468 non-null	float64
6	Non-exclusive breastfeeding	6468 non-null	float64
7	Discontinued breastfeeding	6468 non-null	float64
8	Child wasting	6468 non-null	float64
9	Child stunting	6468 non-null	float64
10	Low birth weight for gestation	6468 non-null	float64
11	Secondhand smoke	6468 non-null	float64
12	Alcohol use	6468 non-null	float64
13	Drug use	6468 non-null	float64
14	Diet low in fruits	6468 non-null	float64
15	Diet low in vegetables	6468 non-null	float64
16	Unsafe sex	6468 non-null	float64
17	Low physical activity	6468 non-null	float64
18	High fasting plasma glucose	6468 non-null	float64
19	High total cholesterol	1561 non-null	float64
20	High body-mass index	6468 non-null	float64
21	High systolic blood pressure	6468 non-null	float64
22	Smoking	6468 non-null	float64
23	Iron deficiency	6468 non-null	float64
24	Vitamin A deficiency	6468 non-null	float64
25	Low bone mineral density	6468 non-null	float64
26	Air pollution	6468 non-null	float64
27	Outdoor air pollution	6467 non-null	float64
28	Diet high in sodium	6468 non-null	float64
29	Diet low in whole grains	6468 non-null	float64
30	Diet low in nuts and seeds	6468 non-null	float64

dtypes: float64(29), int64(1), object(1)  
memory usage: 1.5+ MB

In [8]:

```
df['Entity'].unique()
```

```
Out[8]: array(['Afghanistan', 'Albania', 'Algeria', 'American Samoa',
'Andean Latin America', 'Andorra', 'Angola', 'Antigua and Barbuda',
'Argentina', 'Armenia', 'Australasia', 'Australia', 'Austria',
'Azerbaijan', 'Bahamas', 'Bahrain', 'Bangladesh', 'Barbados',
'Belarus', 'Belgium', 'Belize', 'Benin', 'Bermuda', 'Bhutan',
'Bolivia', 'Bosnia and Herzegovina', 'Botswana', 'Brazil',
'Brunei', 'Bulgaria', 'Burkina Faso', 'Burundi', 'Cambodia',
'Cameroon', 'Canada', 'Cape Verde', 'Caribbean',
'Central African Republic', 'Central Asia', 'Central Europe',
'Central Europe, Eastern Europe, and Central Asia',
'Central Latin America', 'Central Sub-Saharan Africa', 'Chad',
'Chile', 'China', 'Colombia', 'Comoros', 'Congo', 'Costa Rica',
'Cote d'Ivoire', 'Croatia', 'Cuba', 'Cyprus', 'Czechia',
'Democratic Republic of Congo', 'Denmark', 'Djibouti', 'Dominica',
'Dominican Republic', 'East Asia', 'Eastern Europe',
'Eastern Sub-Saharan Africa', 'Ecuador', 'Egypt', 'El Salvador',
'England', 'Equatorial Guinea', 'Eritrea', 'Estonia', 'Eswatini',
'Ethiopia', 'Fiji', 'Finland', 'France', 'Gabon', 'Gambia',
'Georgia', 'Germany', 'Ghana', 'Greece', 'Greenland', 'Grenada',
'Guam', 'Guatemala', 'Guinea', 'Guinea-Bissau', 'Guyana', 'Haiti',
'High SDI', 'High-income', 'High-income Asia Pacific',
'High-middle SDI', 'Honduras', 'Hungary', 'Iceland', 'India',
'Indonesia', 'Iran', 'Iraq', 'Ireland', 'Israel', 'Italy',
'Jamaica', 'Japan', 'Jordan', 'Kazakhstan', 'Kenya', 'Kiribati',
'Kuwait', 'Kyrgyzstan', 'Laos', 'Latin America and Caribbean',
'Latvia', 'Lebanon', 'Lesotho', 'Liberia', 'Libya', 'Lithuania',
'Low SDI', 'Low-middle SDI', 'Luxembourg', 'Madagascar', 'Malawi',
'Malaysia', 'Maldives', 'Mali', 'Malta', 'Marshall Islands',
'Mauritania', 'Mauritius', 'Mexico', 'Micronesia (country)',
```

```
'Middle SDI', 'Moldova', 'Mongolia', 'Montenegro', 'Morocco',
'Mozambique', 'Myanmar', 'Namibia', 'Nepal', 'Netherlands',
'New Zealand', 'Nicaragua', 'Niger', 'Nigeria',
'North Africa and Middle East', 'North America', 'North Korea',
'North Macedonia', 'Northern Ireland', 'Northern Mariana Islands',
'Norway', 'Oceania', 'Oman', 'Pakistan', 'Palestine', 'Panama',
'Papua New Guinea', 'Paraguay', 'Peru', 'Philippines', 'Poland',
'Portugal', 'Puerto Rico', 'Qatar', 'Romania', 'Russia', 'Rwanda',
'Saint Lucia', 'Saint Vincent and the Grenadines', 'Samoa',
'Sao Tome and Principe', 'Saudi Arabia', 'Scotland', 'Senegal',
'Serbia', 'Seychelles', 'Sierra Leone', 'Singapore', 'Slovakia',
'Slovenia', 'Solomon Islands', 'Somalia', 'South Africa',
'South Asia', 'South Korea', 'South Sudan', 'Southeast Asia',
'Southeast Asia, East Asia, and Oceania', 'Southern Latin America',
'Southern Sub-Saharan Africa', 'Spain', 'Sri Lanka',
'Sub-Saharan Africa', 'Sudan', 'Suriname', 'Sweden', 'Switzerland',
'Syria', 'Taiwan', 'Tajikistan', 'Tanzania', 'Thailand', 'Timor',
'Togo', 'Tonga', 'Trinidad and Tobago', 'Tropical Latin America',
'Tunisia', 'Turkey', 'Turkmenistan', 'Uganda', 'Ukraine',
'United Arab Emirates', 'United Kingdom', 'United States',
'United States Virgin Islands', 'Uruguay', 'Uzbekistan', 'Vanuatu',
'Venezuela', 'Vietnam', 'Wales', 'Western Europe',
'Western Sub-Saharan Africa', 'World', 'Yemen', 'Zambia',
'Zimbabwe'], dtype=object)
```

```
In [9]: df['Year'].nunique() #Printing the count of unique values in Year column
```

```
Out[9]: 28
```

```
In [10]: df['High total cholesterol'] = df['High total cholesterol'].fillna(value=df['High total
df['Outdoor air pollution'] = df['Outdoor air pollution'].fillna(value=df['Outdoor air
```

```
In [12]: df.isnull().sum()
```

```
Out[12]: Entity                                0
Year                                            0
Unsafe water source                          0
Unsafe sanitation                            0
No access to handwashing facility            0
Household air pollution from solid fuels     0
Non-exclusive breastfeeding                 0
Discontinued breastfeeding                  0
Child wasting                               0
Child stunting                              0
Low birth weight for gestation               0
Secondhand smoke                            0
Alcohol use                                 0
Drug use                                    0
Diet low in fruits                          0
Diet low in vegetables                      0
Unsafe sex                                  0
Low physical activity                        0
High fasting plasma glucose                  0
High total cholesterol                       0
High body-mass index                        0
High systolic blood pressure                 0
Smoking                                     0
Iron deficiency                             0
Vitamin A deficiency                        0
Low bone mineral density                    0
```

```

Air pollution
Outdoor air pollution
Diet high in sodium
Diet low in whole grains
Diet low in nuts and seeds
dtype: int64

```

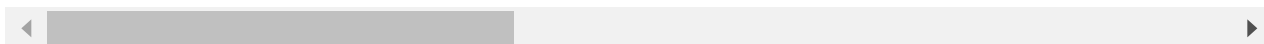
```
In [13]: clean_data = df.copy()
```

```
In [14]: clean_data.head()
```

Out[14]:

	Entity	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontinue breastfeeding
0	Afghanistan	1990	7554.049543	5887.747628	5412.314513	22388.49723	3221.138842	156.09755
1	Afghanistan	1991	7359.676749	5732.770160	5287.891103	22128.75821	3150.559597	151.53985
2	Afghanistan	1992	7650.437822	5954.804987	5506.657363	22873.76879	3331.349048	156.60919
3	Afghanistan	1993	10270.731380	7986.736613	7104.620351	25599.75628	4477.006100	206.83445
4	Afghanistan	1994	11409.177110	8863.010065	8051.515953	28013.16720	5102.622054	233.93057

5 rows × 31 columns

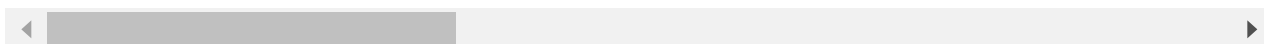


```
In [15]: clean_data.describe()
```

Out[15]:

	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontin breastfeeding
count	6468.000000	6.468000e+03	6.468000e+03	6.468000e+03	6.468000e+03	6468.000000	6468.000
mean	2003.500000	3.156632e+04	2.337436e+04	1.893305e+04	4.308421e+04	6231.427632	409.110
std	8.078372	1.527731e+05	1.144930e+05	8.981037e+04	1.877345e+05	28517.846341	1874.989
min	1990.000000	8.650193e-03	6.495981e-03	7.791357e-02	2.058533e-02	0.003816	0.000
25%	1996.750000	1.019665e+01	4.603845e+00	1.688487e+01	8.759783e+01	4.633254	0.264
50%	2003.500000	2.790317e+02	1.601965e+02	2.524991e+02	1.091671e+03	102.428307	6.619
75%	2010.250000	5.301718e+03	3.832344e+03	3.811442e+03	9.161964e+03	1367.827277	78.279
max	2017.000000	2.111659e+06	1.638021e+06	1.239519e+06	2.708905e+06	514102.351600	34850.395

8 rows × 30 columns



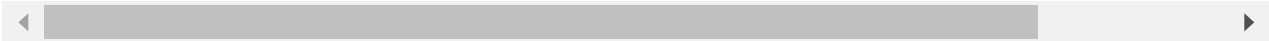
```
In [16]: clean_data.describe().transpose()
```

Out[16]:

	count	mean	std	min	25%	50%	
<b>Year</b>	6468.0	2003.500000	8.078372	1990.000000	1996.750000	2003.500000	2010.25
<b>Unsafe water source</b>	6468.0	31566.317807	152773.116467	0.008650	10.196650	279.031692	5301.71
<b>Unsafe sanitation</b>	6468.0	23374.362141	114493.039414	0.006496	4.603845	160.196536	3832.34
<b>No access to handwashing facility</b>	6468.0	18933.050500	89810.372691	0.077914	16.884869	252.499098	3811.44
<b>Household air pollution from solid fuels</b>	6468.0	43084.206901	187734.464522	0.020585	87.597828	1091.671153	9161.96
<b>Non-exclusive breastfeeding</b>	6468.0	6231.427632	28517.846341	0.003816	4.633254	102.428307	1367.82
<b>Discontinued breastfeeding</b>	6468.0	409.110423	1874.989431	0.000520	0.264366	6.619327	78.27
<b>Child wasting</b>	6468.0	43446.432828	202236.710526	0.101713	41.372448	730.346237	10234.53
<b>Child stunting</b>	6468.0	11767.717972	58248.914775	0.001401	1.863717	77.873619	1971.59
<b>Low birth weight for gestation</b>	6468.0	30948.006623	134294.632657	0.326638	144.562750	1220.716952	8708.14
<b>Secondhand smoke</b>	6468.0	24282.250536	100256.183193	2.890665	278.067695	1196.227901	5963.66
<b>Alcohol use</b>	6468.0	50203.341291	195822.608202	-2315.344758	363.952195	2803.321905	12891.26
<b>Drug use</b>	6468.0	8890.242150	35415.115589	1.240062	92.909932	408.586291	2170.84
<b>Diet low in fruits</b>	6468.0	45452.642748	183428.565074	1.578807	536.043698	2452.885952	10521.82
<b>Diet low in vegetables</b>	6468.0	28742.012912	111659.952882	0.776438	412.982809	1837.753086	7612.29
<b>Unsafe sex</b>	6468.0	26764.450110	121709.063241	1.021822	136.082958	831.822256	5948.95
<b>Low physical activity</b>	6468.0	21141.486434	82215.985896	2.416705	261.559164	1189.412372	5694.74
<b>High fasting plasma glucose</b>	6468.0	99555.714649	384033.016304	21.042632	2034.714167	7820.164595	34704.78
<b>High total cholesterol</b>	6468.0	51628.248060	131283.335535	9.527324	51628.248060	51628.248060	51628.24
<b>High body-mass index</b>	6468.0	68685.287815	268134.065820	19.998208	1141.442770	4739.652491	21601.17

7/11/2021World\_deaths\_Analysis(2017 data)

	count	mean	std	min	25%	50%	
High systolic blood pressure	6468.0	174383.185897	680991.545760	21.026071	2665.313367	10993.308535	47322.84
Smoking	6468.0	133548.348210	529931.503714	11.707478	1292.925608	5935.789171	31638.09
Iron deficiency	6468.0	1878.745701	9011.891580	0.005499	2.256209	31.990666	421.38
Vitamin A deficiency	6468.0	11908.622027	58801.648611	0.003465	1.896386	70.490245	2081.94
Low bone mineral density	6468.0	4579.055654	18884.513384	0.381232	40.602658	246.750756	1096.10
Air pollution	6468.0	95735.506099	390933.534804	8.524593	1076.836756	6125.098028	22727.35
Outdoor air pollution	6468.0	55573.127275	229786.045213	4.830000	553.747500	2242.540000	12831.64
Diet high in sodium	6468.0	54240.674047	243437.333182	2.673823	355.637308	1945.638251	9691.37
Diet low in whole grains	6468.0	53348.812853	209715.312191	9.317592	798.734879	3504.309221	14463.69
Diet low in nuts and seeds	6468.0	34967.039529	135943.192066	5.188788	553.348464	2279.157286	10038.79



```
In [20]: print("Data Visualization")
print("***__INDIA - Deaths Analysis__**")
```

Data Visualization  
\*\*\_\_INDIA - Deaths Analysis\_\_\*\*

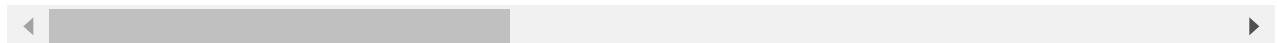
```
In [21]: india = clean_data[clean_data['Entity']=='India']
india
```

Out[21]:

	Entity	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontinued breastfeeding
2688	India	1990	807723.2259	636517.8157	430219.2519	691699.0043	122801.68150	6598.317849
2689	India	1991	811265.9069	638456.6346	430207.6328	693852.4155	120834.70210	6663.946536
2690	India	1992	809832.0949	636020.6287	428346.3225	693823.3194	118244.39090	6512.996663
2691	India	1993	798193.5484	625142.5128	421085.3844	680356.2421	113499.59350	6086.902085
2692	India	1994	782064.1082	611345.7080	411623.0162	670049.3665	108452.87970	5617.580276
2693	India	1995	770997.9855	601637.2419	404314.4473	659136.6390	103304.98030	5169.259357

	Entity	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontinued breastfeeding
<b>2694</b>	India	1996	757174.1992	590231.3711	395475.5173	653262.9633	97235.34133	4726.993619
<b>2695</b>	India	1997	756259.4763	589696.9919	393456.1289	671861.3869	92509.35738	4401.078820
<b>2696</b>	India	1998	750757.1052	585826.0562	388720.2580	668996.9391	88892.96233	4253.957251
<b>2697</b>	India	1999	741045.3527	577967.4818	381265.1672	642748.1845	86196.15713	4293.700457
<b>2698</b>	India	2000	734455.8641	572125.8089	376339.0291	629573.6696	84637.39936	4258.705212
<b>2699</b>	India	2001	718835.9458	559478.1710	367703.6314	619160.9953	80677.91805	3782.868914
<b>2700</b>	India	2002	700750.5468	544821.6410	357291.2530	604227.0178	76002.36876	3273.233842
<b>2701</b>	India	2003	683466.4598	529869.4559	346840.4285	582220.5003	71788.54824	2924.532955
<b>2702</b>	India	2004	668881.3042	515958.8538	338033.9666	552103.8981	68158.89792	2658.520948
<b>2703</b>	India	2005	660257.8942	505825.6925	333250.2428	545624.6479	64578.38255	2344.104524
<b>2704</b>	India	2006	654674.2677	497423.0770	329214.6801	546837.6094	61488.86631	1990.621725
<b>2705</b>	India	2007	650617.7678	489606.9735	325868.8083	549219.8389	58591.64913	1731.429866
<b>2706</b>	India	2008	648163.4067	481544.8127	322651.3581	548852.7310	56138.69294	1664.637589
<b>2707</b>	India	2009	651194.0599	476656.9839	321282.3525	547817.3069	54242.15683	1707.458900
<b>2708</b>	India	2010	656580.0425	473093.9055	321058.2874	550363.8975	53166.66832	1817.688552
<b>2709</b>	India	2011	645210.7307	457656.6462	313755.1790	551432.3215	52162.27505	1916.172493
<b>2710</b>	India	2012	625665.6091	434398.4699	302658.0757	563339.6875	51151.27414	2037.333723
<b>2711</b>	India	2013	604731.7691	405370.4932	289350.1457	542067.3894	48605.79226	2097.360537
<b>2712</b>	India	2014	588806.9726	377367.6562	276761.9034	525296.8903	44444.69785	2044.993931
<b>2713</b>	India	2015	573767.0403	352445.5677	267021.0850	511055.8769	41402.48083	1969.301334
<b>2714</b>	India	2016	560745.6633	332744.6846	258164.9075	500517.1481	36956.76102	1758.432018
<b>2715</b>	India	2017	569679.1687	328719.9721	257783.8570	481737.8380	33415.29006	1611.151610

28 rows × 31 columns



```
In [22]: risk_factors = [rf for rf in india.columns if rf not in ['Entity','Year']]
risk_factors
```

```
Out[22]: ['Unsafe water source',
'Unsafe sanitation',
'No access to handwashing facility',
'Household air pollution from solid fuels',
'Non-exclusive breastfeeding',
'Discontinued breastfeeding',
'Child wasting',
'Child stunting',
```

```
'Low birth weight for gestation',
'Secondhand smoke',
'Alcohol use',
'Drug use',
'Diet low in fruits',
'Diet low in vegetables',
'Unsafe sex',
'Low physical activity',
'High fasting plasma glucose',
'High total cholesterol',
'High body-mass index',
'High systolic blood pressure',
'Smoking',
'Iron deficiency',
'Vitamin A deficiency',
'Low bone mineral density',
'Air pollution',
'Outdoor air pollution',
'Diet high in sodium',
'Diet low in whole grains',
'Diet low in nuts and seeds']
```

```
In [23]: #Lets calculate the average deaths for each risk factor
average_deaths = []

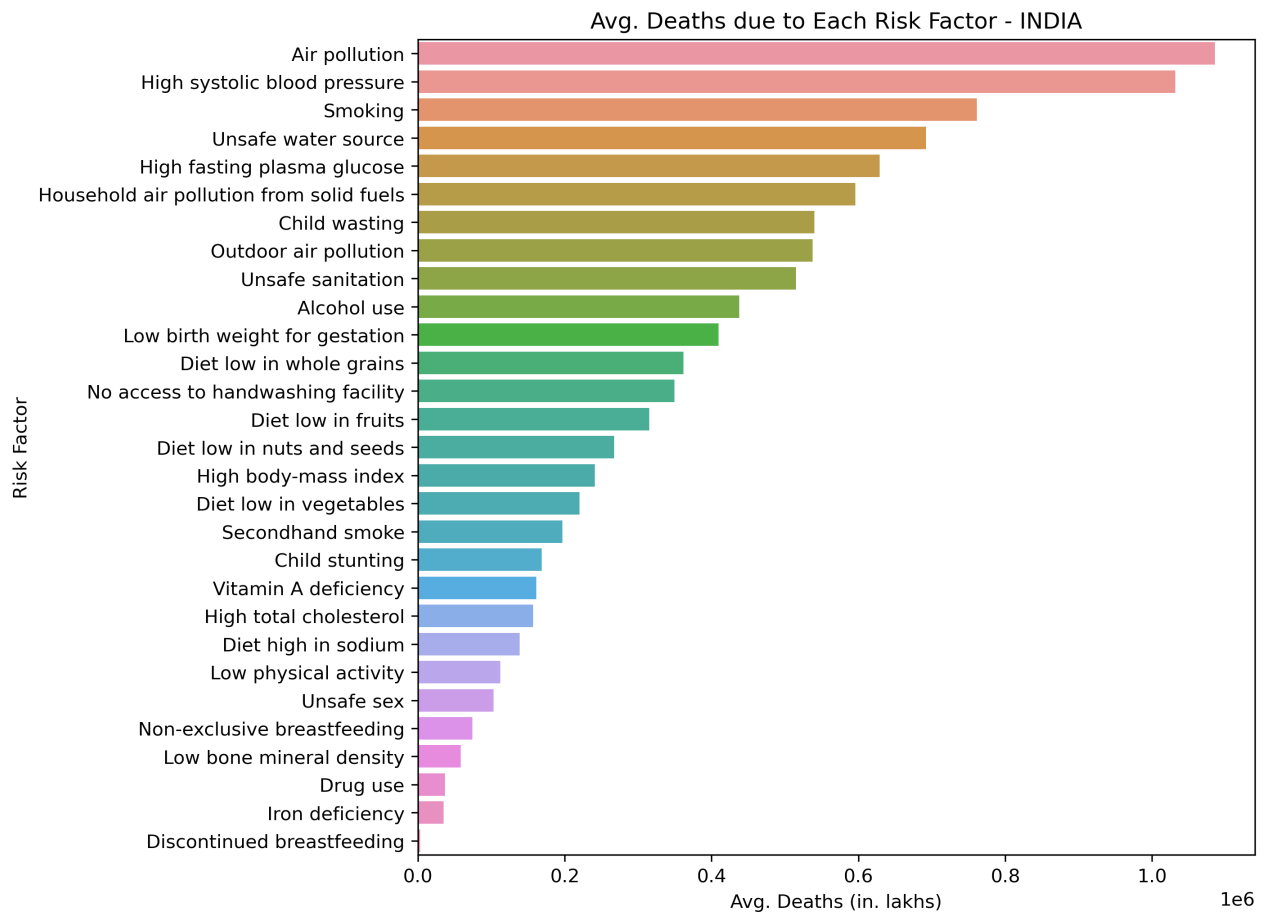
for rf in risk_factors:
    average_deaths.append(india[rf].mean())
average_deaths
```

```
Out[23]: [692207.0541607141,
515283.97529642854,
349633.6542000001,
595615.5616071428,
74627.93449249999,
3425.474342357145,
540388.6854035716,
168667.23107,
409614.6570892857,
197002.18847142858,
437495.6127892857,
37175.507973928565,
315318.88531071425,
220255.84635714287,
103100.10545499998,
112155.68383464286,
628757.6898464285,
157260.6051053682,
240996.77285214287,
1031488.4393107144,
761627.4202464285,
35397.44180964287,
161300.8784614286,
58577.2670775,
1085535.42335,
537817.9953571429,
138980.55262857143,
361639.40244285704,
267398.24519642856]
```

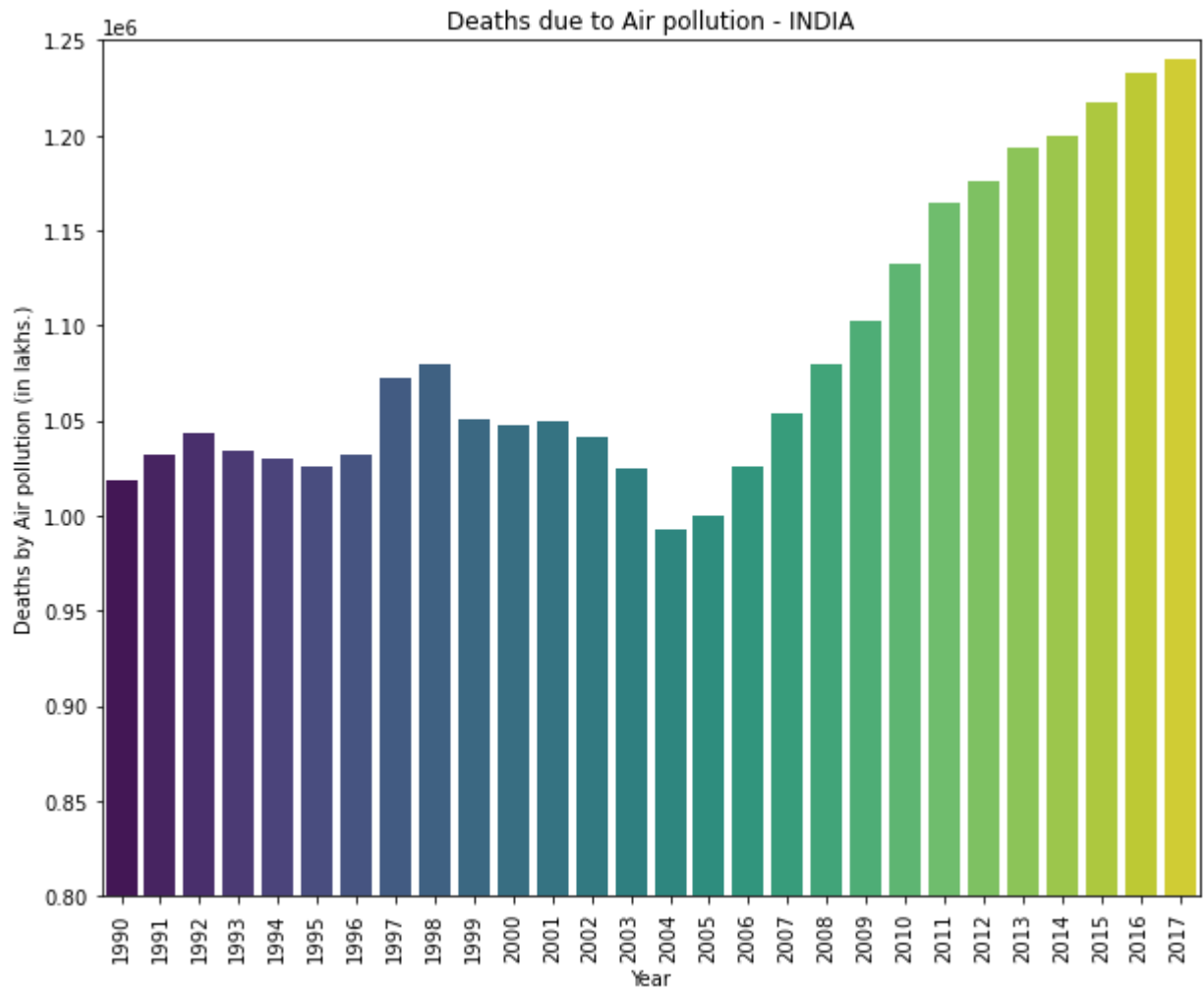
```
In [24]: #making a dataframe to store riskfactors and average deaths
df = pd.DataFrame(list(zip(risk_factors,average_deaths)),columns=['Risk Factor','Avg. D
```



```
In [26]: #Lets Plot the Graph to undestand the Deaths for each risk factor
import seaborn as sns
plt.figure(figsize=(8,8),dpi=300)
sns.barplot(y='Risk Factor',x='Avg. Deaths',data=df)
plt.title('Avg. Deaths due to Each Risk Factor - INDIA')
plt.xlabel('Avg. Deaths (in. lakhs)')
plt.show()
```

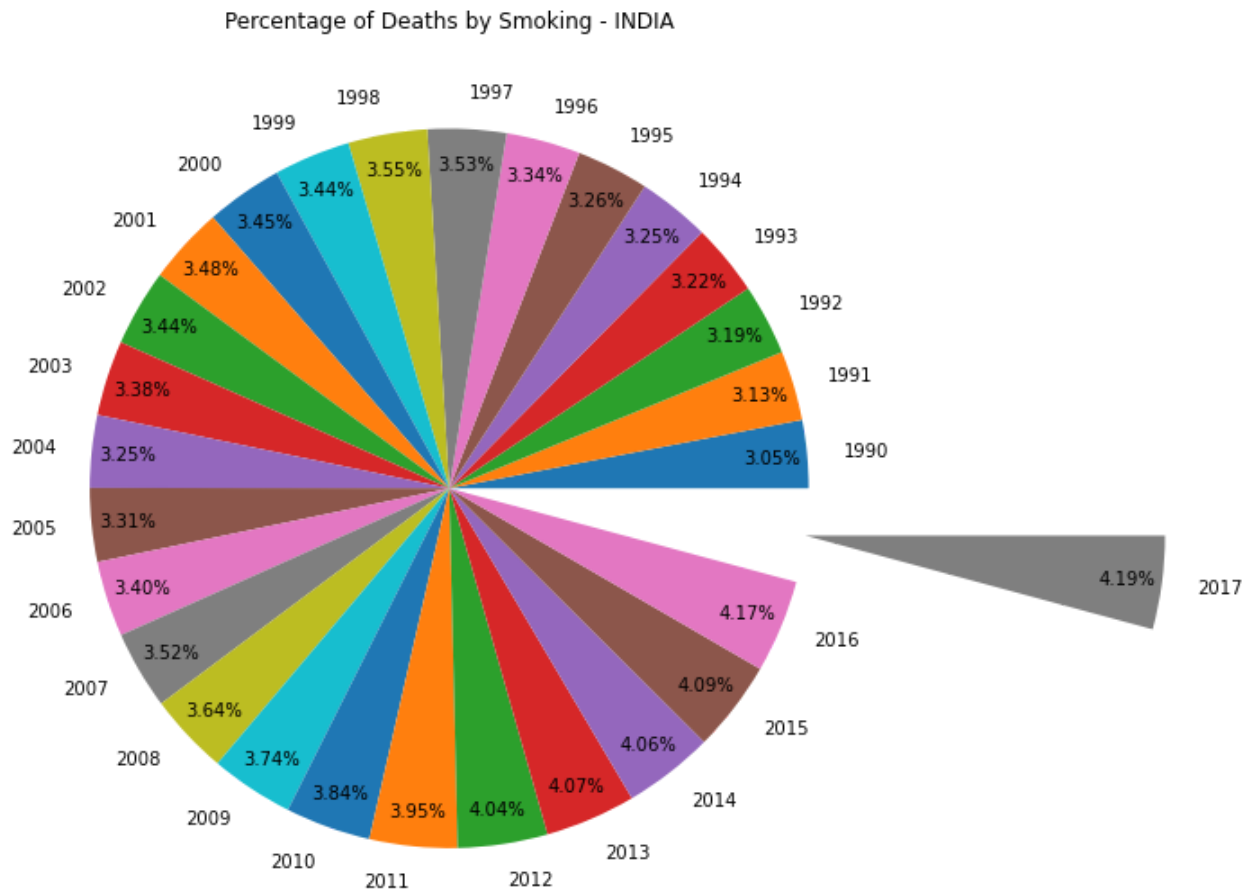


```
In [27]: #Lets check deaths in india due to Air pollution
plt.figure(figsize=(10,8))
sns.barplot(x='Year',y='Air pollution',data=india,palette='viridis')
plt.xticks(rotation=90)
plt.title('Deaths due to Air pollution - INDIA')
plt.ylim(800000,1250000)
plt.ylabel('Deaths by Air pollution (in lakhs.)')
plt.show()
```



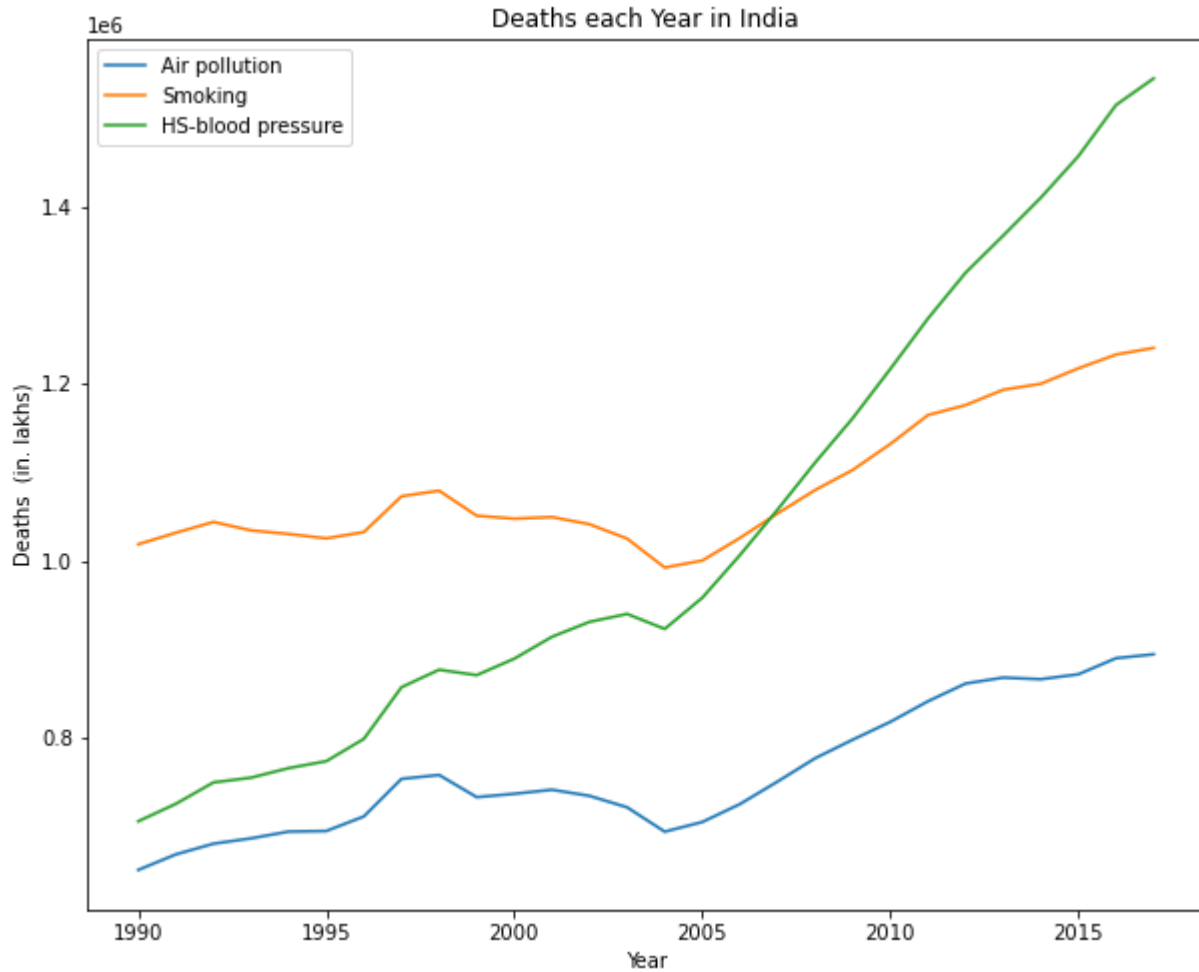
In [28]:

```
plt.figure(figsize=(10,10),tight_layout=True)
e=[0 if i<27 else 1 for i in range(28)]
plt.pie(india['Smoking'],labels=india.Year,autopct='%.2f%',pctdistance=0.9,explode=e)
plt.title('Percentage of Deaths by Smoking - INDIA')
plt.show()
```



In [29]:

```
plt.figure(figsize=(10,8))
plt.plot(india.Year,india['Smoking'],label='Air pollution')
plt.plot(india.Year,india['Air pollution'],label='Smoking')
plt.plot(india.Year,india['High systolic blood pressure'],label='HS-blood pressure')
plt.legend()
plt.xlabel('Year')
plt.ylabel('Deaths (in. lakhs)')
plt.title('Deaths each Year in India')
plt.show()
```



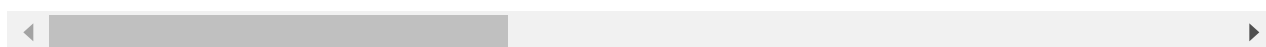
```
In [30]: world = clean_data[clean_data['Entity']=='World']
world
```

```
Out[30]:
```

	Entity	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontinued breastfeeding
<b>6356</b>	World	1990	2111659.077	1.638021e+06	1.239519e+06	2708904.820	514102.3516	34850.39553
<b>6357</b>	World	1991	2095066.505	1.622959e+06	1.230319e+06	2677805.722	502180.0259	33853.61620
<b>6358</b>	World	1992	2053362.286	1.588275e+06	1.209424e+06	2648544.390	481844.5346	32221.54775
<b>6359</b>	World	1993	2013224.699	1.553936e+06	1.188248e+06	2604092.494	462904.1330	30647.96698
<b>6360</b>	World	1994	1973406.044	1.519967e+06	1.167209e+06	2563321.579	445422.5075	29178.11928
<b>6361</b>	World	1995	1931065.508	1.484244e+06	1.144504e+06	2521547.506	425889.4485	27731.32010
<b>6362</b>	World	1996	1877778.830	1.440235e+06	1.116520e+06	2477283.308	405322.2782	26123.42414
<b>6363</b>	World	1997	1840394.520	1.409442e+06	1.096678e+06	2456653.199	386961.3670	24776.91639
<b>6364</b>	World	1998	1808119.007	1.382608e+06	1.077689e+06	2419886.895	372047.0743	23887.01349
<b>6365</b>	World	1999	1766645.451	1.347958e+06	1.052547e+06	2357044.544	357098.3394	23068.35938
<b>6366</b>	World	2000	1727905.469	1.315192e+06	1.030042e+06	2314854.156	343019.1682	22083.23036

	Entity	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontinued breastfeeding
6367	World	2001	1682357.015	1.277426e+06	1.003460e+06	2262796.548	328012.1893	20739.48575
6368	World	2002	1638128.167	1.240535e+06	9.768055e+05	2218646.043	312995.3409	19504.70007
6369	World	2003	1596385.848	1.205164e+06	9.517598e+05	2172350.595	299066.9771	18449.95506
6370	World	2004	1561987.644	1.174028e+06	9.303295e+05	2109307.172	286549.6128	17610.22055
6371	World	2005	1536826.824	1.149262e+06	9.143999e+05	2057468.406	274539.5928	16819.40560
6372	World	2006	1523674.684	1.132529e+06	9.053644e+05	2002416.138	266156.2990	16281.51882
6373	World	2007	1503844.158	1.109839e+06	8.918319e+05	1961085.424	255904.2064	15825.76529
6374	World	2008	1491461.978	1.089920e+06	8.798532e+05	1927176.022	247528.8480	15725.66256
6375	World	2009	1489789.499	1.077638e+06	8.736675e+05	1897357.343	240509.8515	15843.25268
6376	World	2010	1483731.657	1.061903e+06	8.650517e+05	1869568.125	233039.1549	15878.10448
6377	World	2011	1431184.379	1.012280e+06	8.353075e+05	1832880.494	222655.1329	14953.00717
6378	World	2012	1374545.908	9.578132e+05	8.043316e+05	1824696.213	211430.0757	14194.98613
6379	World	2013	1322082.215	9.025648e+05	7.736068e+05	1771869.513	199855.9254	13357.84569
6380	World	2014	1282539.799	8.540761e+05	7.481440e+05	1744033.091	188371.9281	12323.16812
6381	World	2015	1245322.727	8.102460e+05	7.254840e+05	1705654.026	178114.3003	11281.73997
6382	World	2016	1220145.311	7.790716e+05	7.091127e+05	1696332.020	168080.7873	10414.22949
6383	World	2017	1232368.284	7.742410e+05	7.072477e+05	1640599.784	160983.3744	10011.80923

28 rows × 31 columns



In [31]:

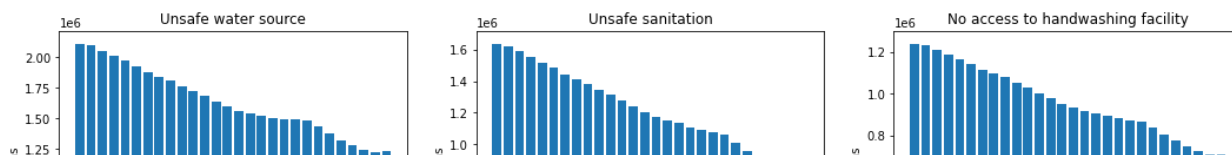
```
#Lets plot each risk factor

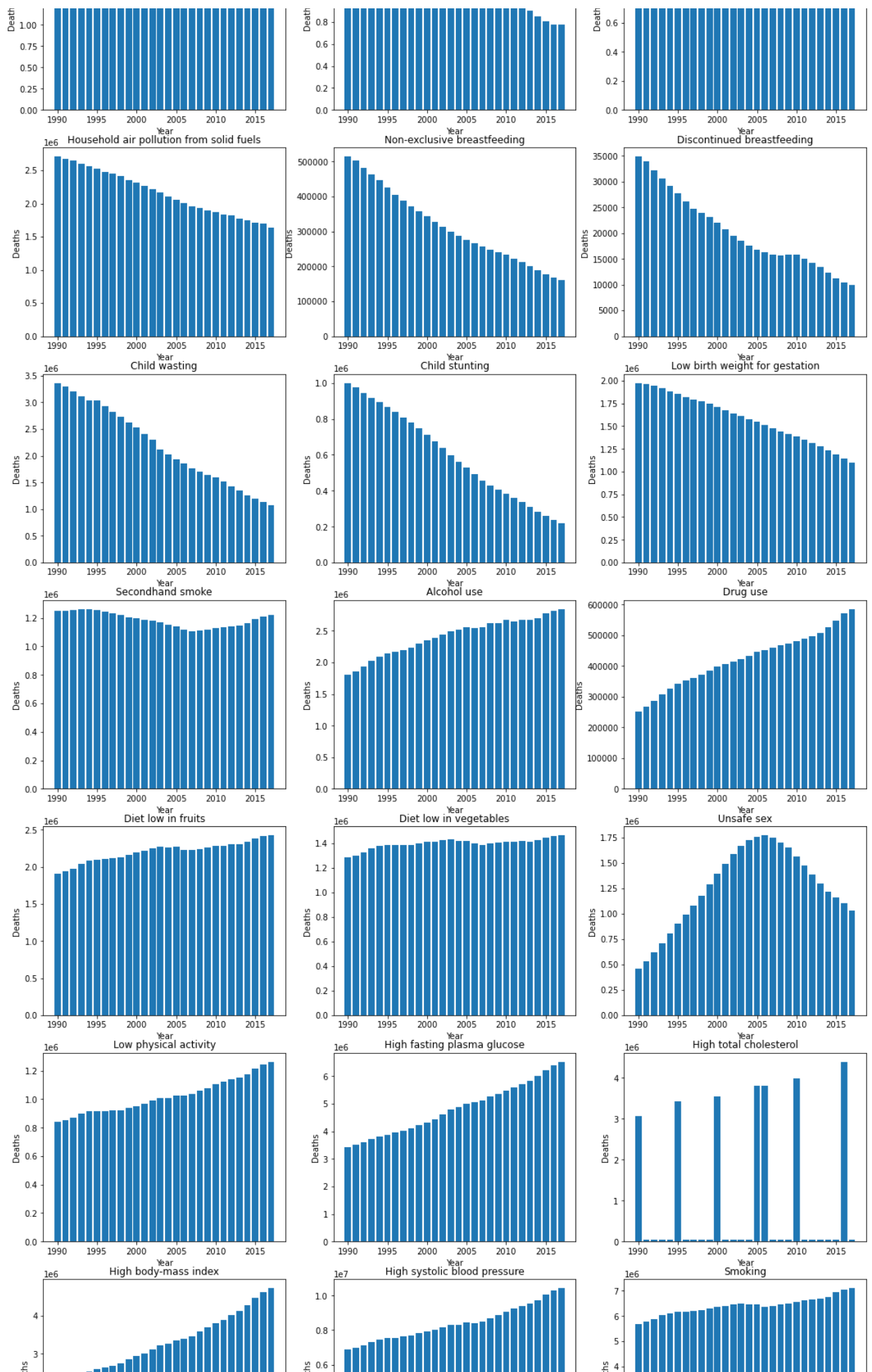
risk_factors = [rf for rf in india.columns if rf not in ['Entity','Year']]

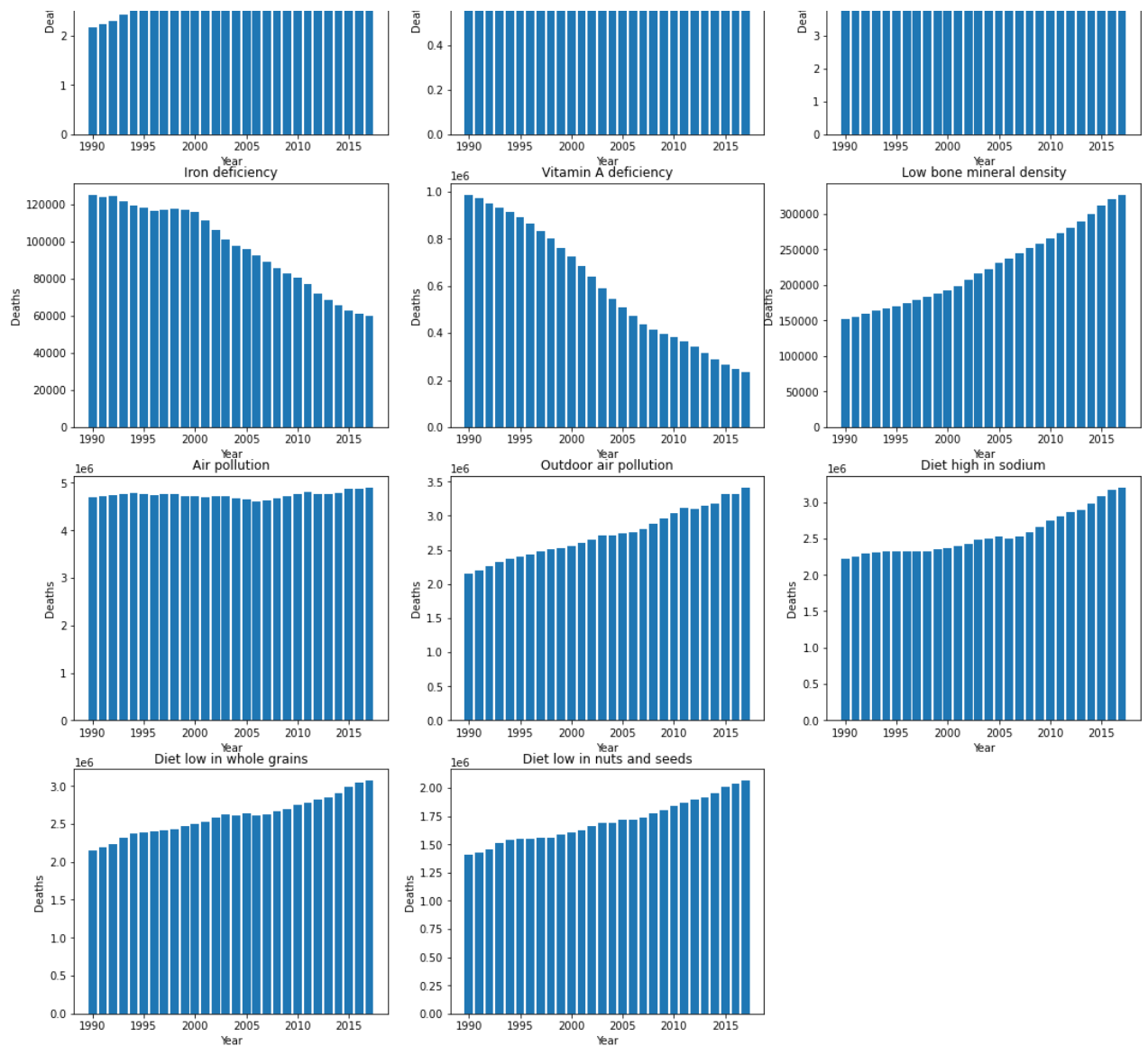
plt.figure(figsize=(18,50))

for index,rf in enumerate(risk_factors):
    plt.subplot(10,3,index+1)
    plt.bar(world['Year'],world[rf],label=rf)
    plt.title(rf)
    plt.ylabel('Deaths')
    plt.xlabel('Year')

plt.show()
```







In [32]:

#Comparison of Average Deaths in INDIA v/s World each Year

```
india_world = clean_data[(clean_data['Entity']=='India') | (clean_data['Entity']=='World')]
india_world
```

Out[32]:

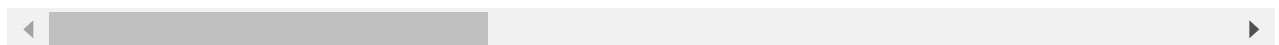
	Entity	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontinued breastfeeding
2688	India	1990	8.077232e+05	6.365178e+05	4.302193e+05	6.916990e+05	122801.68150	6598.3178
2689	India	1991	8.112659e+05	6.384566e+05	4.302076e+05	6.938524e+05	120834.70210	6663.9465
2690	India	1992	8.098321e+05	6.360206e+05	4.283463e+05	6.938233e+05	118244.39090	6512.9966
2691	India	1993	7.981935e+05	6.251425e+05	4.210854e+05	6.803562e+05	113499.59350	6086.9020
2692	India	1994	7.820641e+05	6.113457e+05	4.116230e+05	6.700494e+05	108452.87970	5617.5802
2693	India	1995	7.709980e+05	6.016372e+05	4.043144e+05	6.591366e+05	103304.98030	5169.2593
2694	India	1996	7.571742e+05	5.902314e+05	3.954755e+05	6.532630e+05	97235.34133	4726.9936

	Entity	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontinued breastfeeding
<b>2695</b>	India	1997	7.562595e+05	5.896970e+05	3.934561e+05	6.718614e+05	92509.35738	4401.0788
<b>2696</b>	India	1998	7.507571e+05	5.858261e+05	3.887203e+05	6.689969e+05	88892.96233	4253.9572
<b>2697</b>	India	1999	7.410454e+05	5.779675e+05	3.812652e+05	6.427482e+05	86196.15713	4293.7004
<b>2698</b>	India	2000	7.344559e+05	5.721258e+05	3.763390e+05	6.295737e+05	84637.39936	4258.7052
<b>2699</b>	India	2001	7.188359e+05	5.594782e+05	3.677036e+05	6.191610e+05	80677.91805	3782.8689
<b>2700</b>	India	2002	7.007505e+05	5.448216e+05	3.572913e+05	6.042270e+05	76002.36876	3273.2338
<b>2701</b>	India	2003	6.834665e+05	5.298695e+05	3.468404e+05	5.822205e+05	71788.54824	2924.5329
<b>2702</b>	India	2004	6.688813e+05	5.159589e+05	3.380340e+05	5.521039e+05	68158.89792	2658.5209
<b>2703</b>	India	2005	6.602579e+05	5.058257e+05	3.332502e+05	5.456246e+05	64578.38255	2344.1049
<b>2704</b>	India	2006	6.546743e+05	4.974231e+05	3.292147e+05	5.468376e+05	61488.86631	1990.6217
<b>2705</b>	India	2007	6.506178e+05	4.896070e+05	3.258688e+05	5.492198e+05	58591.64913	1731.4298
<b>2706</b>	India	2008	6.481634e+05	4.815448e+05	3.226514e+05	5.488527e+05	56138.69294	1664.6379
<b>2707</b>	India	2009	6.511941e+05	4.766570e+05	3.212824e+05	5.478173e+05	54242.15683	1707.4589
<b>2708</b>	India	2010	6.565800e+05	4.730939e+05	3.210583e+05	5.503639e+05	53166.66832	1817.6889
<b>2709</b>	India	2011	6.452107e+05	4.576566e+05	3.137552e+05	5.514323e+05	52162.27505	1916.1724
<b>2710</b>	India	2012	6.256656e+05	4.343985e+05	3.026581e+05	5.633397e+05	51151.27414	2037.3337
<b>2711</b>	India	2013	6.047318e+05	4.053705e+05	2.893501e+05	5.420674e+05	48605.79226	2097.3609
<b>2712</b>	India	2014	5.888070e+05	3.773677e+05	2.767619e+05	5.252969e+05	44444.69785	2044.9939
<b>2713</b>	India	2015	5.737670e+05	3.524456e+05	2.670211e+05	5.110559e+05	41402.48083	1969.3013
<b>2714</b>	India	2016	5.607457e+05	3.327447e+05	2.581649e+05	5.005171e+05	36956.76102	1758.4320
<b>2715</b>	India	2017	5.696792e+05	3.287200e+05	2.577839e+05	4.817378e+05	33415.29006	1611.1516
<b>6356</b>	World	1990	2.111659e+06	1.638021e+06	1.239519e+06	2.708905e+06	514102.35160	34850.3959
<b>6357</b>	World	1991	2.095067e+06	1.622959e+06	1.230319e+06	2.677806e+06	502180.02590	33853.6162
<b>6358</b>	World	1992	2.053362e+06	1.588275e+06	1.209424e+06	2.648544e+06	481844.53460	32221.5477
<b>6359</b>	World	1993	2.013225e+06	1.553936e+06	1.188248e+06	2.604092e+06	462904.13300	30647.9669
<b>6360</b>	World	1994	1.973406e+06	1.519967e+06	1.167209e+06	2.563322e+06	445422.50750	29178.1192
<b>6361</b>	World	1995	1.931066e+06	1.484244e+06	1.144504e+06	2.521548e+06	425889.44850	27731.3207
<b>6362</b>	World	1996	1.877779e+06	1.440235e+06	1.116520e+06	2.477283e+06	405322.27820	26123.4247
<b>6363</b>	World	1997	1.840395e+06	1.409442e+06	1.096678e+06	2.456653e+06	386961.36700	24776.9163
<b>6364</b>	World	1998	1.808119e+06	1.382608e+06	1.077689e+06	2.419887e+06	372047.07430	23887.0134
<b>6365</b>	World	1999	1.766645e+06	1.347958e+06	1.052547e+06	2.357045e+06	357098.33940	23068.3593
<b>6366</b>	World	2000	1.727905e+06	1.315192e+06	1.030042e+06	2.314854e+06	343019.16820	22083.2303



	Entity	Year	Unsafe water source	Unsafe sanitation	No access to handwashing facility	Household air pollution from solid fuels	Non-exclusive breastfeeding	Discontinued breastfeeding
6367	World	2001	1.682357e+06	1.277426e+06	1.003460e+06	2.262797e+06	328012.18930	20739.4857
6368	World	2002	1.638128e+06	1.240535e+06	9.768055e+05	2.218646e+06	312995.34090	19504.7000
6369	World	2003	1.596386e+06	1.205164e+06	9.517598e+05	2.172351e+06	299066.97710	18449.9550
6370	World	2004	1.561988e+06	1.174028e+06	9.303295e+05	2.109307e+06	286549.61280	17610.2205
6371	World	2005	1.536827e+06	1.149262e+06	9.143999e+05	2.057468e+06	274539.59280	16819.4056
6372	World	2006	1.523675e+06	1.132529e+06	9.053644e+05	2.002416e+06	266156.29900	16281.5188
6373	World	2007	1.503844e+06	1.109839e+06	8.918319e+05	1.961085e+06	255904.20640	15825.7652
6374	World	2008	1.491462e+06	1.089920e+06	8.798532e+05	1.927176e+06	247528.84800	15725.6625
6375	World	2009	1.489789e+06	1.077638e+06	8.736675e+05	1.897357e+06	240509.85150	15843.2526
6376	World	2010	1.483732e+06	1.061903e+06	8.650517e+05	1.869568e+06	233039.15490	15878.1042
6377	World	2011	1.431184e+06	1.012280e+06	8.353075e+05	1.832880e+06	222655.13290	14953.0071
6378	World	2012	1.374546e+06	9.578132e+05	8.043316e+05	1.824696e+06	211430.07570	14194.9861
6379	World	2013	1.322082e+06	9.025648e+05	7.736068e+05	1.771870e+06	199855.92540	13357.8456
6380	World	2014	1.282540e+06	8.540761e+05	7.481440e+05	1.744033e+06	188371.92810	12323.1681
6381	World	2015	1.245323e+06	8.102460e+05	7.254840e+05	1.705654e+06	178114.30030	11281.7395
6382	World	2016	1.220145e+06	7.790716e+05	7.091127e+05	1.696332e+06	168080.78730	10414.2292
6383	World	2017	1.232368e+06	7.742410e+05	7.072477e+05	1.640600e+06	160983.37440	10011.8092

56 rows × 31 columns



In [33]:

```
ad_india = []
ad_world = []

for i in ['India', 'World']:
    df = india_world[india_world['Entity']==i]

    if i=='India':
        for rf in risk_factors:
            ad_india.append(df[rf].mean())
    else:
        for rf in risk_factors:
            ad_world.append(df[rf].mean())

df1 = pd.DataFrame(list(zip(risk_factors, ad_india)), columns=['Risk Factor', 'Avg. Deaths'])
df2 = pd.DataFrame(list(zip(risk_factors, ad_world)), columns=['Risk Factor', 'Avg. Deaths'])
```

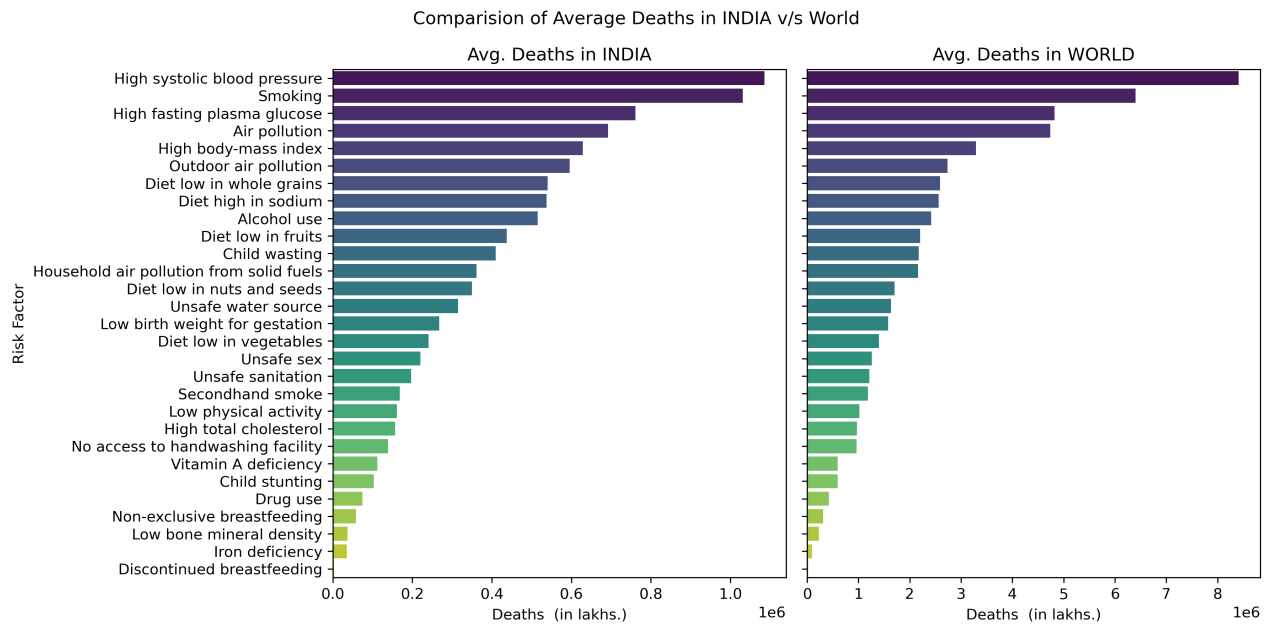
In [34]:

```
fig, axes = plt.subplots(1, 2, sharey=True, figsize=(12, 6), tight_layout=True, dpi=300)
fig.suptitle('Comparison of Average Deaths in INDIA v/s World')
sns.barplot(ax=axes[0], y='Risk Factor', x='Avg. Deaths', data=df1, palette='viridis')
```

```

axes[0].set_title('Avg. Deaths in INDIA')
axes[0].set_xlabel('Deaths (in lakhs.)')
sns.barplot(ax=axes[1],y='Risk Factor',x='Avg. Deaths',data=df2,palette='viridis')
axes[1].set_title('Avg. Deaths in WORLD')
axes[1].set_xlabel('Deaths (in lakhs.)')
axes[1].set_ylabel('')
plt.show()

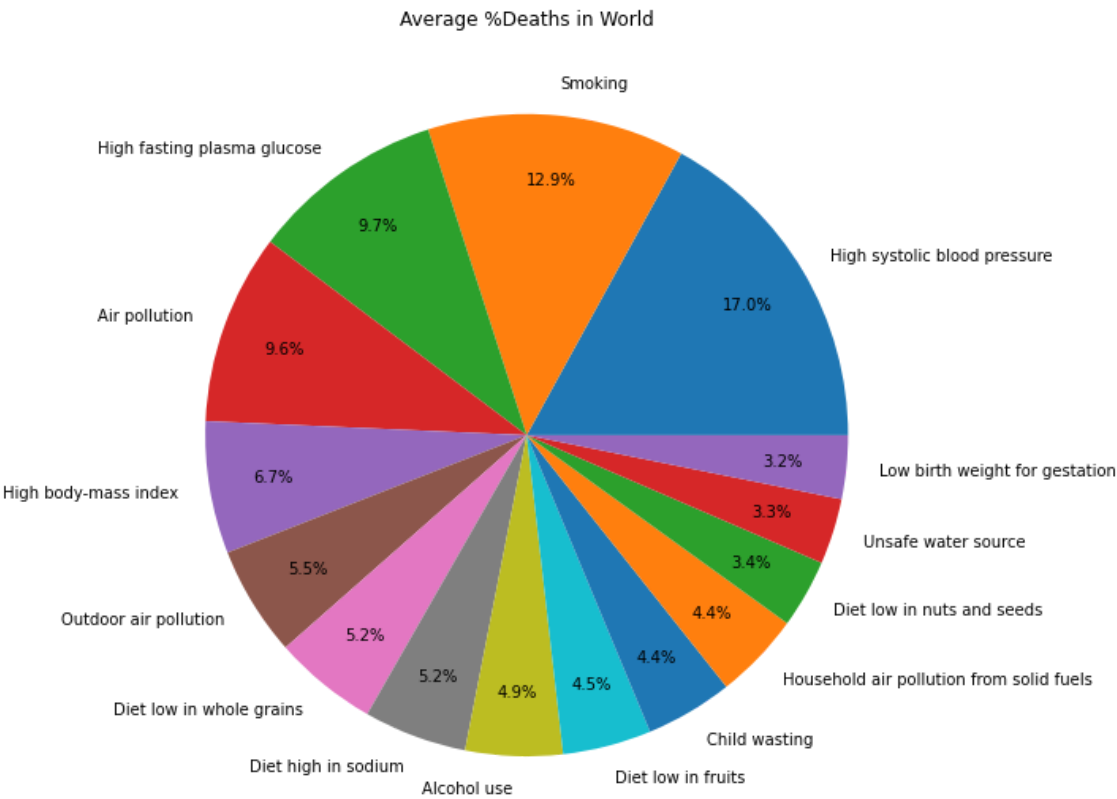
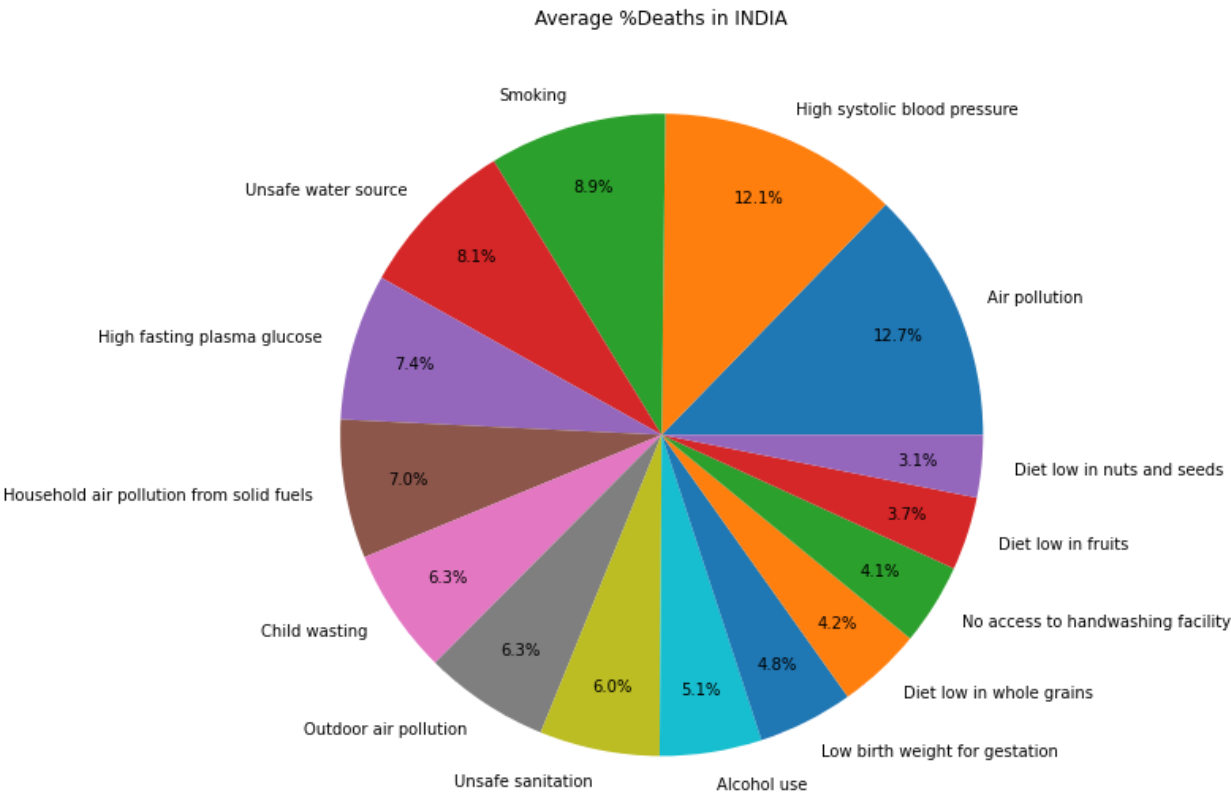
```



```

In [35]: fig, axes = plt.subplots(2,1,figsize=(15,15),tight_layout=True)
axes[0].pie(df1['Avg. Deaths'][:15],labels=df1['Risk Factor'][:15],autopct='%.1f%%',pct
axes[0].set_title('Average %Deaths in INDIA')
axes[1].pie(df2['Avg. Deaths'][:15],labels=df2['Risk Factor'][:15],autopct='%.1f%%',pct
axes[1].set_title('Average %Deaths in World')
plt.show()

```



In [ ]: