Smoking By numbers

Project Documentation for Comp. Applications Lab

Done By:

Yazan Mohammad Hassan Hussein 0161943

Alaa Wael Nazem Al Safadi 0166024

Dataset used

All dataset used in this project was taken from the annual Global Burden of Disease (GBD) study [1] conducted by World Health Organization (WHO) and the institute of Health Metrics (IHME) the csv files used in this project are:

1) `number-of-deaths-by-risk-factor.csv`

Dataset that maps different deaths from various diseases from 1990 to 2017 sorted by country

2) `share-deaths-smoking.csv`

Share of premature deaths attributed to tobacco smoking.

3) `death-rate-smoking.csv`

death rates from tobacco smoking across the world. Death rates measure the number of premature deaths from smoking per 100,000 people in each country or region.

4) `smoking-deaths-by-age.csv`

breakdown of deaths from smoking by age

5) `smoking-deaths-1990-2017.csv`

comparison of smoking death rates from 1990 to 2016 by country and income classification

6) `share-of-adults-who-smoke.csv`

The percentage of the population ages 15 years and over who currently use any tobacco product (smoked and/or smokeless tobacco) on a daily or non-daily basis. Tobacco products include cigarettes, pipes, cigars, cigarillos, waterpipes (hookah, shisha), bidis, kretek, heated tobacco products, and all forms of smokeless (oral and nasal) tobacco. Tobacco products exclude e-cigarettes (which do not contain tobacco), “e-cigars”, “e-hookahs”, JUUL and “e-pipes”. The rates are age-standardized to the WHO Standard Population.

7) `comparing-the-share-of-men-and-women-who-are-smoking.csv`

Prevalence of smoking, female is the percentage of women ages 15 and over who currently smoke any tobacco product on a daily or non-daily basis. It excludes smokeless tobacco use. The rates are age standardized.

8) `sales-of-cigarettes-per-adult-per-day.csv`

average number of cigarettes sold per adult per day Total cigarettes include both manufactured and hand-rolled cigarettes. For Latvia, total cigarettes comprise of manufactured cigarettes and Papyrosi cigarettes. In some countries the original data reports 0 for a specific year, with otherwise large, reported values for the years before and after. In these cases, we treat 0 as a mistake and report missing data for that year. The 'number of cigarettes smoked per person per day' for both males and females has been averaged across all years in which multiple estimates were provided in the ISS dataset for the United States, to arrive at one estimate for each year. The time series for Germany includes West Germany cigarette sales for the 1948-1989 period.

9) `adults-smoking-2000-2016.csv`

Prevalence of smoking is the percentage of men and women ages 15 and over who currently smoke any tobacco product on a daily or non-daily basis. It excludes smokeless tobacco use. The rates are age standardized.

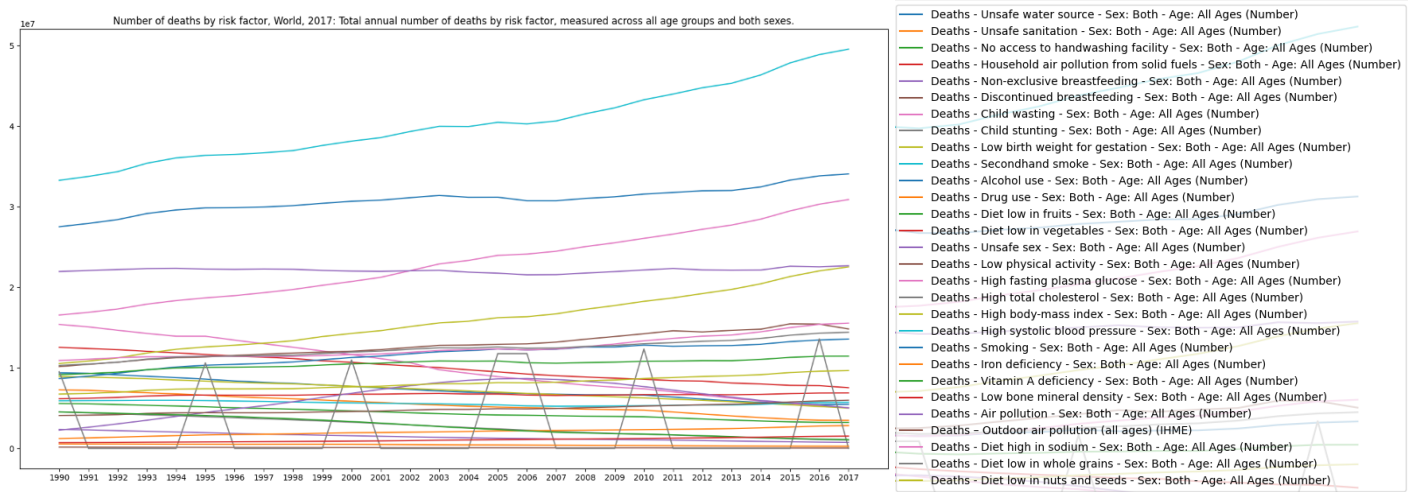
Introduction

Smoking is one of the leading risks for early deaths, approximately 8 million premature deaths are attributed directly to smoking according to the global death toll from tobacco use published by the World Health Organization (WHO) and the institute of Health Metrics (IHME) and Evaluation, We use “Tobacco use” here because other forms of consuming Tobacco exist that lead to premature death even though according to the IHME smoking tobacco accounts to 99.9% of the death cases.

WHO estimates 8 million deaths related to tobacco; 7 million of those are result of direct tobacco use and about 1.2 million are non-smokers who got exposed to second-hand smoke.

IHME Global Burden of Disease study

In the annual Global Burden of Disease (GBD) study it estimated that 8.7 million people die prematurely from tobacco use every year, 71% of those dead of smoking are men as shown in the following plot :



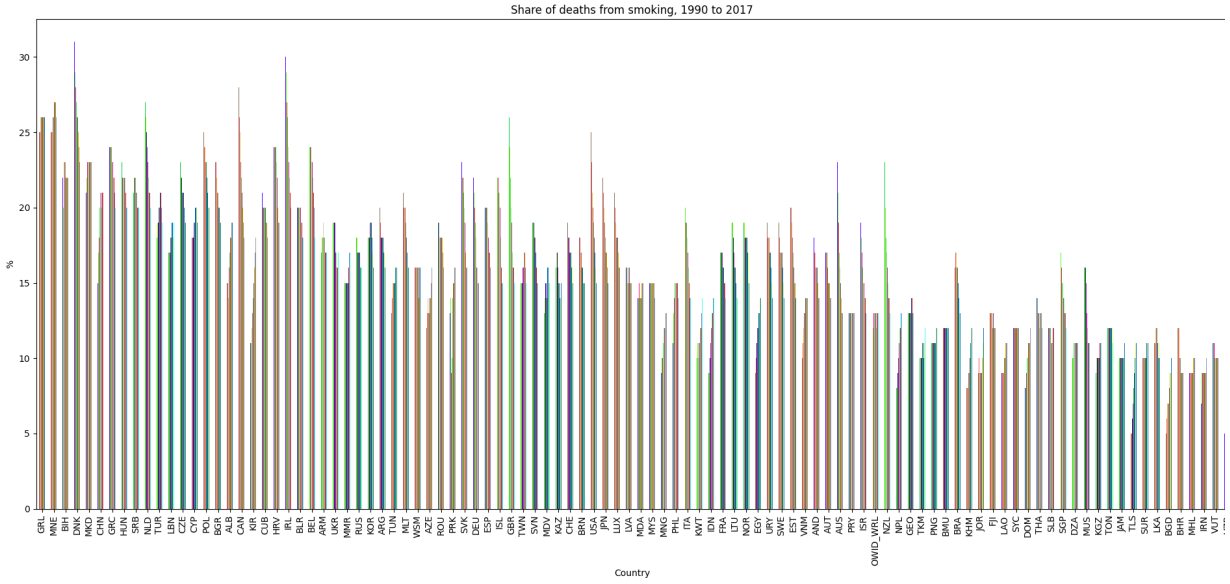
To get the plot we used the following code:

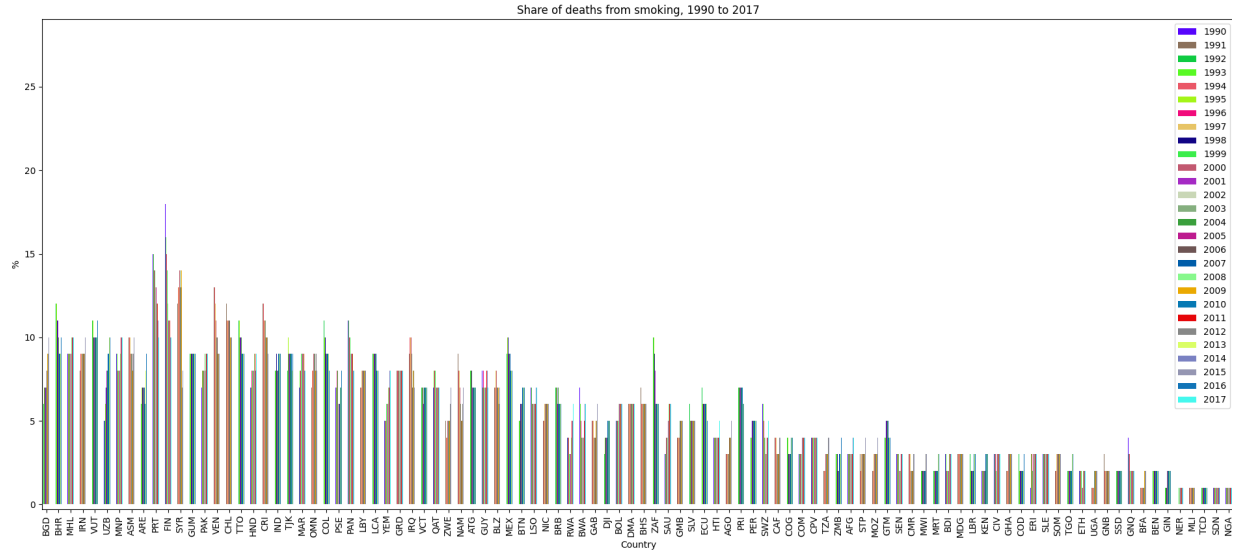
```
deaths_by_risk_factors = pd.read_csv(
    'number-of-deaths-by-risk-factor.csv').round()
(deaths_by_risk_factors.groupby(['Year']).sum()).plot(
    xticks=list(range(1990, 2018)), title='Number of deaths by risk factor, World
, 2017: Total annual number of deaths by risk factor, measured across all age gro
ups and both sexes.', color=color)
plt.show()
```

GBD estimates are presented in another publication. The Tobacco Atlas is published by the American Cancer Soceity and Vital Strategies and presents estimates for the global death toll from smoking taken from the GBD study published by the IHME.

The global distribution of smoking deaths

15% of global deaths are attributed to smoking, below we see the shares of death attributed to direct smoking across the world, unfortunatly in some countries the share reached more than1 3%, in some countries like China, Denmark, The Netherlands, Bosnia and Herzegovina and Greenland more than 1-in-5 deaths were a result of smoking as shown below : [1]



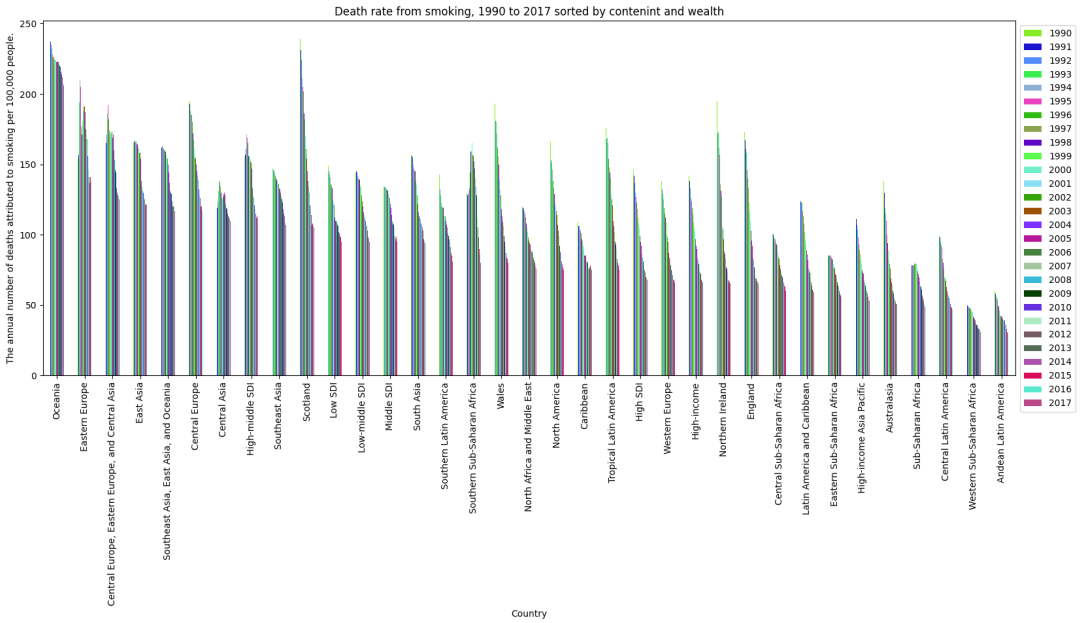


To get the plot we used the following code :

```
share_of_deaths_from_smoking = pd.read_csv(-deaths-smoking.csv').round()
share_of_deaths_from_smoking = share_of_deaths_from_smoking.dropna()
share_of_deaths_from_smoking = share_of_deaths_from_smoking.pivot(='Code', column
s='Year', values='Smoking (IHME, 2019)')
share_of_deaths_from_smoking.sort_values(by=[2017], ascending=False).plot(
    kind='bar', color=color, title='Share of deaths from smoking, 1990 to 2017',
xlabel='Country', ylabel='%')
plt.legend(bbox_to_anchor=(1.0, 1.0))
plt.show()
```

From the above plot that is ordered by the countries that have the most share of smokers as of 2017, greenland tops the chart with over 25% of the shares of deaths world wide, and the least country is nigeria with less than 4% of the world death.

Death rates from smoking are highest across Asia and Eastern Europe, in the below chart we sorted the death rate from smoking acsending from 2017, that means that the region with the biggest death rate from smoking in 2017 is Oceania, whilst the region with least death rate from smoking was Andean Latin America as shown below :

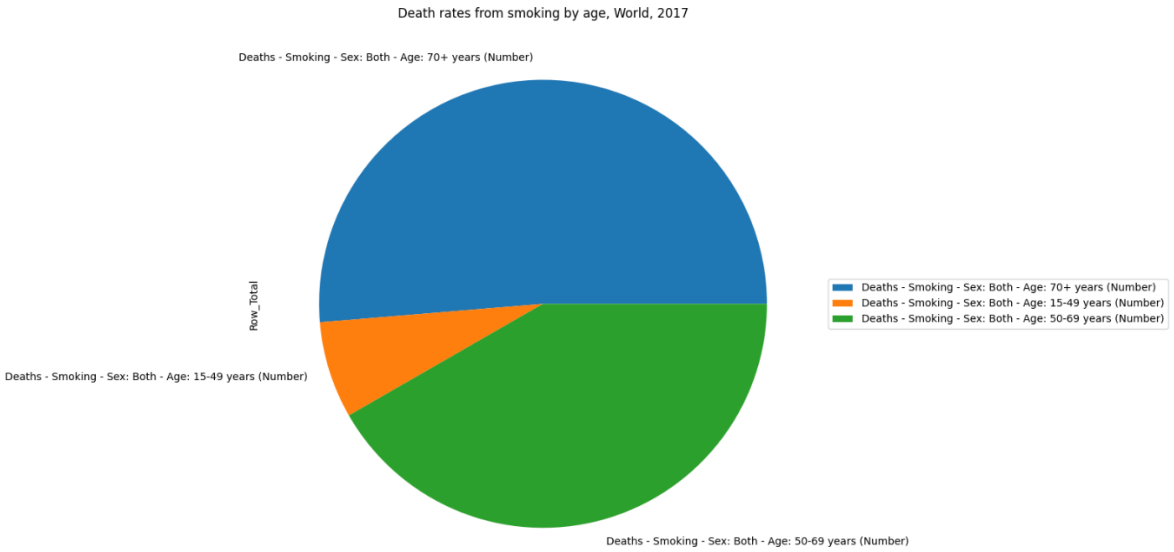


To get the following plot we used the following code:

```
death_rate_from_smoking = pd.read_csv('death-rate-smoking.csv').round()
death_rate_by_cont = death_rate_from_smoking.loc[death_rate_from_smoking['Code'].
isnull()]
death_rate_by_cont = death_rate_by_cont.drop(['Code'], axis=1)
death_rate_by_cont = death_rate_by_cont.pivot(index='Entity', columns='Year', val
ues='Deaths')
death_rate_by_cont.sort_values(by=[2017], ascending=False).plot(kind='bar', color
=color, title='Death rate from smoking, 1990 to 2017 sorted by contenint and weal
th',xlabel='Country', ylabel='The annual number of deaths attributed to smoking p
er 100,000 people.')
plt.legend(bbox_to_anchor=(1.0, 1.0))
plt.show()
```

This refutes a very common misconception people have that poorer nations smoke more, this is not true whoever, across the lowest-income countries in the world like sudan or nigeria have lower number of smokers so the death rate is about 10 times lower.

Older people are more succeptable to death from smoking, as shown in the below graph :

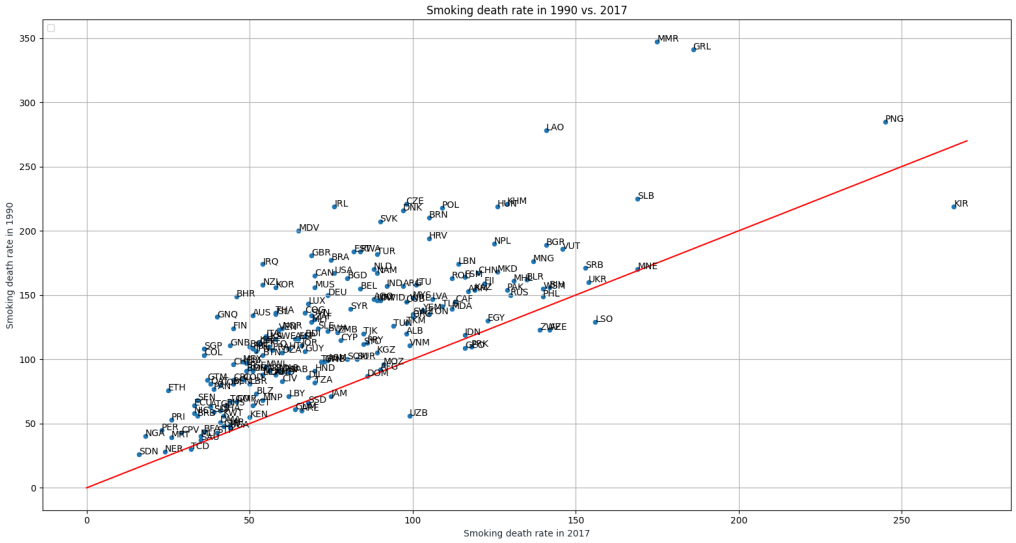


To get the plot we used the following code :

```
smoking_death_by_age = pd.read_csv(
    'smoking-deaths-by-age.csv').round()
smoking_death_by_age = smoking_death_by_age.dropna()
smoking_death_by_age = smoking_death_by_age.loc[smoking_death_by_age['Year'] == 2
017]
smoking_death_by_age = smoking_death_by_age.drop(
    ['Year', 'Code', 'Entity'], axis=1)
smoking_death_by_age = smoking_death_by_age.T
smoking_death_by_age.loc[:, 'Row_Total'] = smoking_death_by_age.sum(
    numeric_only=True, axis=1)
smoking_death_by_age.plot.pie(
    y='Row_Total', title='Death rates from smoking by age, World, 2017')
plt.legend(loc='center left', bbox_to_anchor=(1.0, 0.5))
plt.show()
```

Smoking deaths over time

Death rates from smoking are fallen, in 1990 146 people per 100,000 people were dead from smoking, in 2017 the number fell to 90, to see where exactly the rates are falling or rising we plot the numbers in 1990 vs 2017 in a scatter plot as shown below :



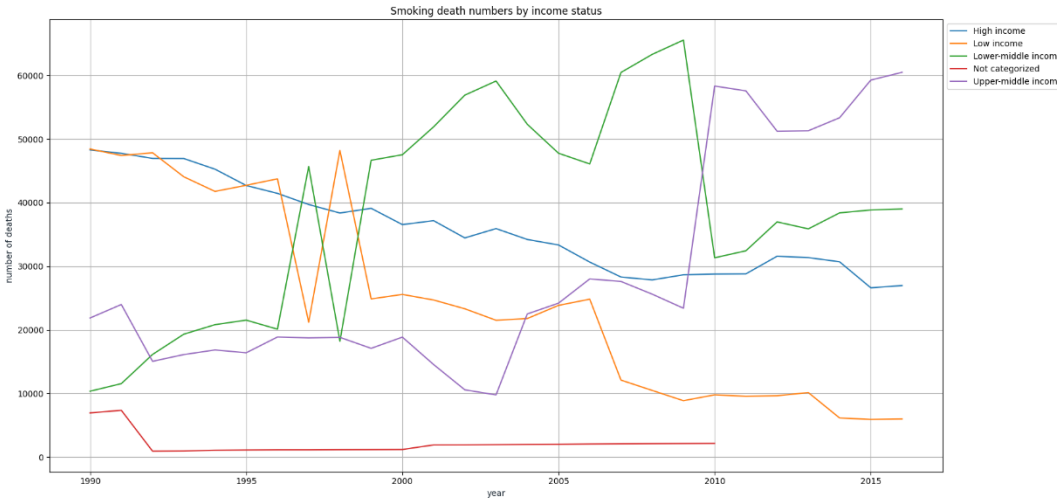
To get the plot we used the following code:

```
death_rate_from_smoking = pd.read_csv(
    'death-rate-smoking.csv').round()
death_rate_from_smoking = death_rate_from_smoking.dropna()
death_rate_from_smoking = death_rate_from_smoking.loc[
    death_rate_from_smoking['Year'].isin([1990, 2017])]
death_rate_from_smoking = death_rate_from_smoking.pivot(
    index='Code', columns='Year', values='Deaths')
death_rate_from_smoking = death_rate_from_smoking.reset_index()
ax = death_rate_from_smoking.plot(
    x=2017, y=1990, kind='scatter', figsize=(10, 10))
death_rate_from_smoking[[2017, 1990, 'Code']].apply(
    lambda x: ax.text(*x), axis=1)
x = np.random.rand(100)
y = np.random.rand(100)
t = np.arange(100)
x = np.linspace(0, 270, 100)
y = x
plt.plot(x, y, '-r')
plt.title('Smoking death rate in 1990 vs. 2017')
```

```
plt.xlabel('Smoking death rate in 2017', color='#1C2833')
plt.ylabel('Smoking death rate in 1990', color='#1C2833')
plt.legend(loc='upper left')
plt.grid()
plt.show()
```

In the scatterplot here we see the comparison of smoking death rates in 1990 (shown on the y-axis) versus the death rate in 2017 (on the x-axis). The red line is the line of parity: countries which lie along this line had equal death rates in 1990 as in 2017. Countries which lie above the grey line had higher death rates in 1990; those which lie below the red line had higher rates in 2017, almost everywhere in the world we can see the declining global trend.

Number of smoke-caused deaths are falling in rich countries but rising in poor to midde-income countries as shown below:



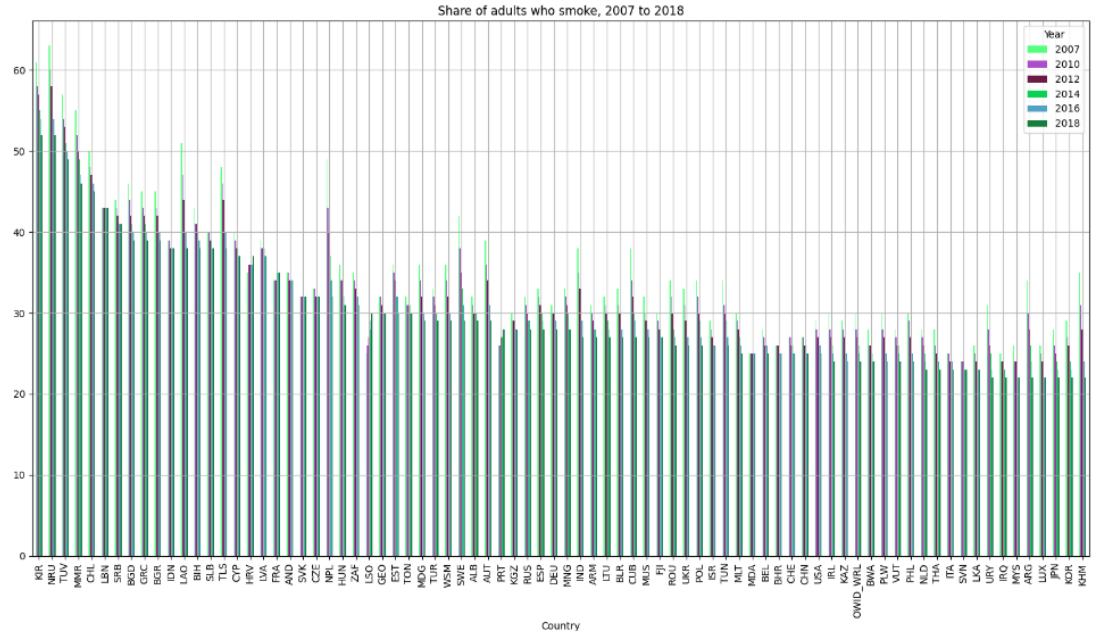
To get the plot we used the following code:

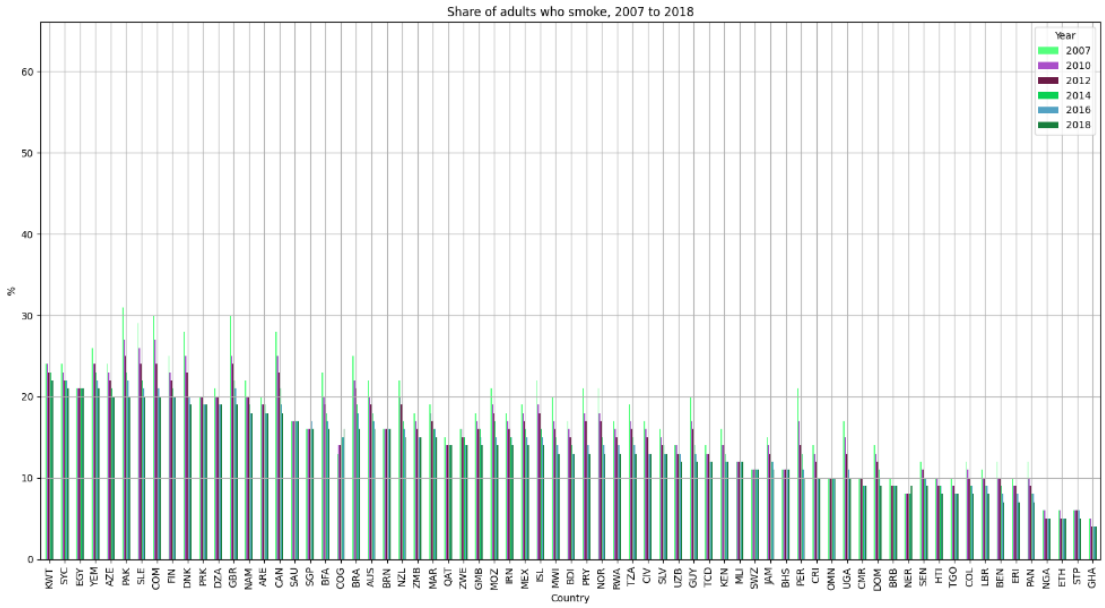
```
smoking_deaths = pd.read_csv(
    'smoking-deaths-1990-2017.csv').round()
smoking_deaths = smoking_deaths.loc[smoking_deaths['Year'].isin(
    list(range(1990, 2017)))]
smoking_deaths = smoking_deaths[smoking_deaths.Entity != 'World']
smoking_deaths = smoking_deaths.drop(
    ['Entity', 'Code', 'Year.1', 'Deaths - Smoking - Sex: Both - Age: All Ages (N
umber).1'], axis=1)
smoking_deaths = smoking_deaths.dropna()
smoking_deaths = smoking_deaths.pivot_table(index='Year', columns='Income classif
ications (World Bank (2017))',
```

```
                                        values='Deaths - Smoking - Sex: Both
- Age: All Ages (Number)')
smoking_deaths.plot()
plt.title('Smoking death numbers by income status')
plt.xlabel('year', color='#1C2833')
plt.ylabel('number of deaths', color='#1C2833')
plt.legend(bbox_to_anchor=(1.0, 1.0))
plt.grid()
plt.show()
```

Smoking across the world

About 20% of the world's population smokes tobacco, in the below chart we can see which country has the most % of adults that smoke :





To get the plot we use the following code:

```
smoking_shares = pd.read_csv(
    'share-of-adults-who-smoke.csv').round()
smoking_shares = smoking_shares.pivot_table(
    index='Code', columns='Year', values='Prevalence of current tobacco use (% of
 adults)')
smoking_shares.sort_values(by=[2018], ascending=False).plot(
    kind='bar', color=color, title='Share of adults who smoke, 2007 to 2018', xla
bel='Country', ylabel='%')
plt.grid()
plt.show()
```

We see that the share of adults aged 15 years or older who smoke tobacco, there are five countries where more than 40% of the puplation smokes, these are:

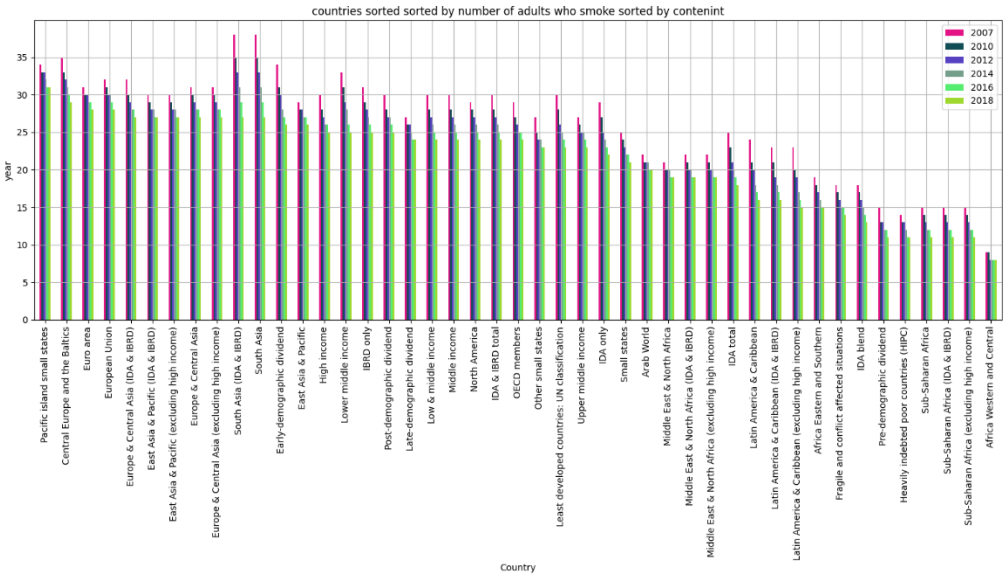
Kiribati (47%)

Montenegro (46%)

Greece (43%)

Timor (43%)

Nauru (40%)



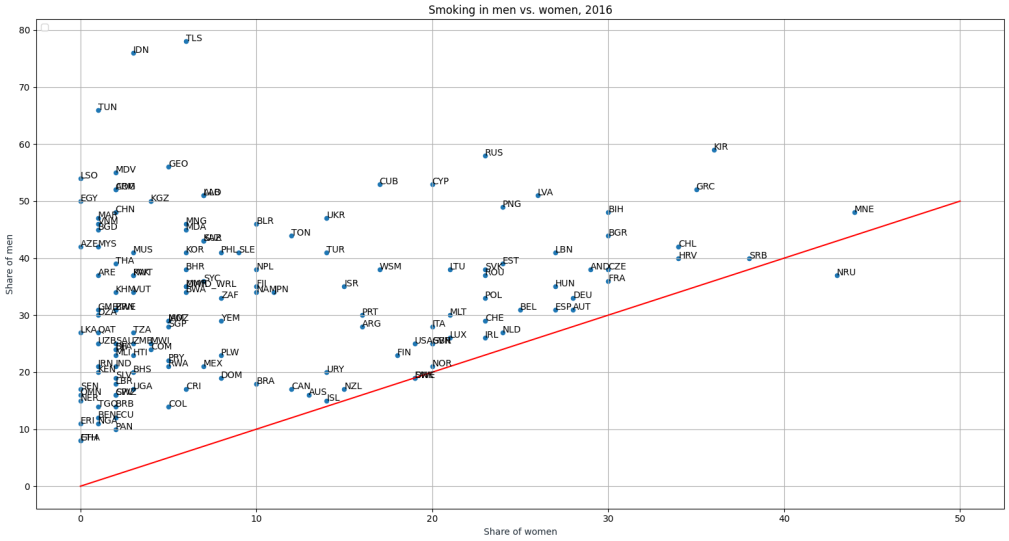
To get the following plot we used this code:

```
smoking_shares = pd.read_csv(
    'share-of-adults-who-smoke.csv').round()
smoking_shares = smoking_shares.pivot_table(
    index='Code', columns='Year', values='Prevalence of current tobacco use (% of
 adults)')
smoking_shares.sort_values(by=[2018], ascending=False).plot(
    kind='bar', color=color, title='Share of adults who smoke, 2007 to 2018', xla
bel='Country', ylabel='%')
plt.grid()
plt.show()
smoking_shares_cont = pd.read_csv(
    'share-of-adults-who-smoke.csv').round()
smoking_shares_cont = smoking_shares_cont.loc[smoking_shares_cont['Code'].isnull(
)]
smoking_shares_cont = smoking_shares_cont.drop(['Code'], axis=1)
smoking_shares_cont = smoking_shares_cont.pivot_table(
    index='Entity', columns='Year', values='Prevalence of current tobacco use (%
of adults)')
smoking_shares_cont.sort_values(by=[2018], ascending=False).plot(
    kind='bar', color=color, title='countries sorted sorted by number of adults w
ho smoke sorted by contenint', xlabel='Country', ylabel='year')
plt.legend(bbox_to_anchor=(1.0, 1.0))
plt.grid()
plt.show()
```

The places where many people smoke are clustered in two regions. South-East Asia and the Pacific islands and Europe – particularly the Balkan region – but also France (33%), Germany (31%), and Austria (30%). In some countries very few people smoke: in Ethiopia, Ghana, Peru and Honduras less than 5% smoke. In Honduras, it's every 50th person. Smoking rates are high across many countries, but we know from experience that this can change quickly. Many of today's high-income countries had much higher rates of smoking in the past, and have seen a big decline. In 2000, the UK had rates similar to Indonesia today – 38% of adults smoked. Since then, rates in the UK have fallen to 22%. The rise, peak, then decline of smoking is one we see across many countries.

Smoking and gender

It's a common fact that men smoke more than women, about 35% of men are smokers while just over 6% of women are, if we look on the below plot :



To get the plot we use the following code:

```
men_v_women = pd.read_csv(
    'comparing-the-share-of-men-and-women-who-are-smoking.csv').round()
men_v_women = men_v_women.loc[men_v_women['Year'].isin([2016])]
men_v_women = men_v_women.drop(
    ['Total population (Gapminder, HYDE & UN)', 'Continent', 'Entity'], axis=1)
men_v_women = men_v_women.dropna()
men_v_women = men_v_women.reset_index()
ax = men_v_women.plot.scatter(x='Smoking prevalence, females (% of adults)',
```

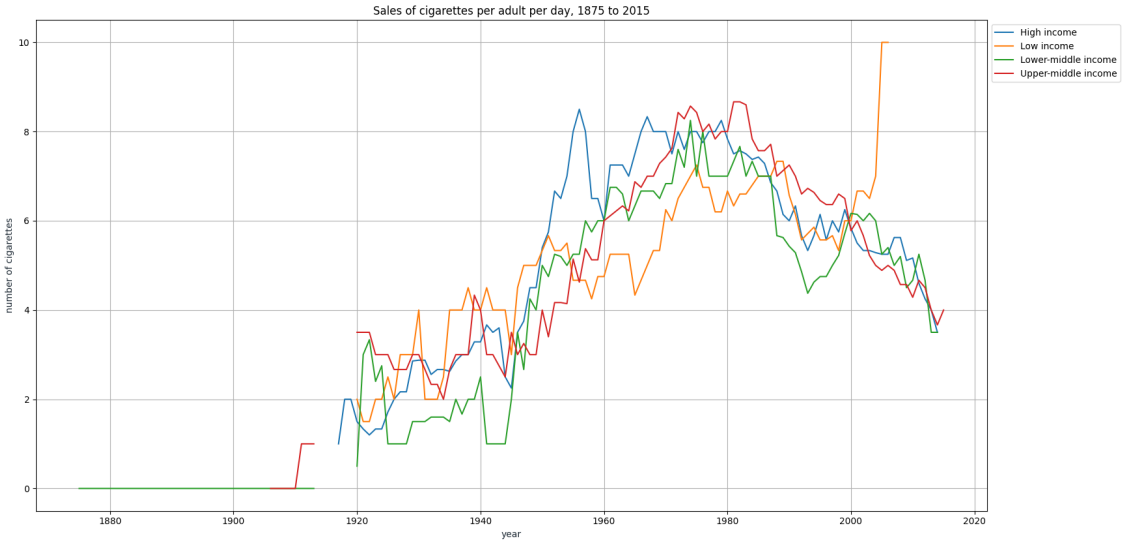
```
                              y='Smoking prevalence, males (% of adults)')
men_v_women[['Smoking prevalence, females (% of adults)', 'Smoking prevalence, ma
les (% of adults)', 'Code']].apply(
    lambda x: ax.text(*x), axis=1)
x = np.random.rand(100)
y = np.random.rand(100)
t = np.arange(100)
x = np.linspace(0, 50, 100)
y = x
plt.plot(x, y, '-r')
plt.title('Smoking in men vs. women, 2016')
plt.xlabel('Share of women', color='#1C2833')
plt.ylabel('Share of men', color='#1C2833')
plt.legend(loc='upper left')
plt.grid()
plt.show()
```

The red line in the plot represents equality in the prevalence: countries where smoking is more common in men will lie above this line; and countries where more women smoke lie below. We see that almost all countries lie above the red line, meaning a higher share of men smoke. But there are a few exceptions: in the Pacific island-state of Nauru 43% of women smoke compared to 37% of men; and smoking rates in Denmark and Sweden show almost no sex difference. In many countries – particularly across Asia and Africa – the differences are very large. We see these countries clustered on the far left, where smoking rates for women are very low – typically less than 5%. In Indonesia, 76% of men smoke but only 3% of women; in China it's 48% of men versus 2% of women; and in Egypt half of men smoke whilst almost no women (0.2%) do.

Change in prevalence of smoking over time

In the early 20th century we saw the rise of smoking cigarettes in rich countries, since then trends have gone through a decline.

In thw following chart we see the sales of cigarettes per adult per day in rich,low income and upper-middle income countries:



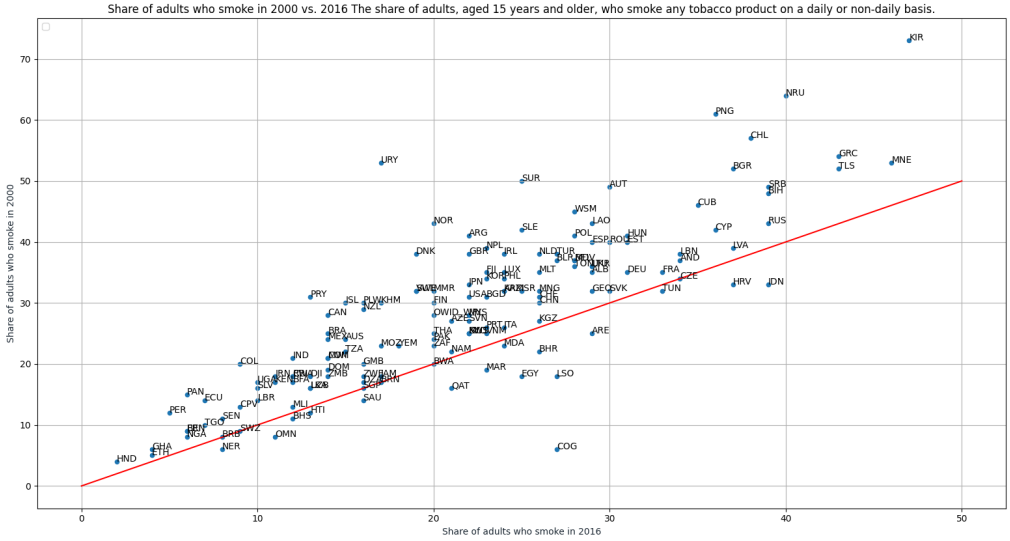
To get the plot we use the following code:

```
smoking_deaths = pd.read_csv(
    'smoking-deaths-1990-2017.csv').round()
cigarette_sales = pd.read_csv(
    'sales-of-cigarettes-per-adult-per-day.csv').round()
smoking_deaths = pd.read_csv(
    'smoking-deaths-1990-2017.csv').round()
smoking_deaths = smoking_deaths.dropna()
classi = smoking_deaths.set_index('Code').to_dict(
)['Income classifications (World Bank (2017))']
cigarette_sales = pd.read_csv(
    'sales-of-cigarettes-per-adult-per-day.csv').round()
cigarette_sales = cigarette_sales.dropna()
cigarette_sales['class'] = smoking_deaths['Code'].map(classi)
cigarette_sales = cigarette_sales.drop(['Entity', 'Code'], axis=1)
cigarette_sales = cigarette_sales.pivot_table(
    index='Year', columns='class', values='Sales of cigarettes per adult per day
(International Smoking Statistics (2017)) ')
cigarette_sales.plot()
plt.title('Sales of cigarettes per adult per day, 1875 to 2015')
plt.xlabel('year', color='#1C2833')
plt.ylabel('number of cigarettes', color='#1C2833')
plt.legend(bbox_to_anchor=(1.0, 1.0))
plt.grid()
plt.show()
```

We can see a decline in numbers in high,upper-middle,and lower-middle income while the numbers in low income is increasing.

In almost all countries smoking rates are falling, The rise, peak then decline of smoking in rich countries took around a century. A long trajectory with severe health impacts.

In the below plot we can see the share of adults who smoke in the year 2000 and 2016, The red line here shows parity: countries that lie along this line would have the same prevalence of smoking in 2000 as in 2016. Countries which lie above this line had higher smoking prevalence in 2000; those below had lower prevalence in 2000.



To get the plot we used the following code:

```
adults_smoking = pd.read_csv(
    'adults-smoking-2000-2016.csv').round()
adults_smoking = adults_smoking.loc[
    adults_smoking['Year'].isin([2000, 2016])]
adults_smoking = adults_smoking.drop(
    ['Year.1', 'Entity', 'Smoking prevalence, total (ages 15+).1', 'Continent'],
axis=1)
adults_smoking = adults_smoking.dropna()
adults_smoking = adults_smoking.pivot_table(
    index='Code', columns='Year', values='Smoking prevalence, total (ages 15+)')
adults_smoking = adults_smoking.reset_index()
print(adults_smoking)
ax = adults_smoking.plot(
    x=2016, y=2000, kind='scatter', figsize=(10, 10))
adults_smoking[[2016, 2000, 'Code']].apply(
    lambda x: ax.text(*x), axis=1)
```

```
x = np.random.rand(100)
y = np.random.rand(100)
t = np.arange(100)
x = np.linspace(0, 50, 100)
y = x
plt.plot(x, y, '-r')
plt.title('Share of adults who smoke in 2000 vs. 2016 The share of adults, aged 1
5 years and older, who smoke any tobacco product on a daily or non-daily basis.')
plt.xlabel('Share of adults who smoke in 2016', color='#1C2833')
plt.ylabel('Share of adults who smoke in 2000', color='#1C2833')
plt.legend(loc='upper left')
plt.grid()
plt.show()
```

We see that most countries lie above the grey line: this means the share of adults who smoke has declined in most countries in the world over the past 16 years. This is a surprising fact to many, since it means smoking prevalence is not only falling in high-income countries, but also at low-to-middle incomes.

5 Low-to-middle income countries have effectively 'leapfrogged' the century-long rise-peak-decline pathway of rich countries. Almost everywhere, smoking is on the decline.

Sources

[1]*GBD Results Tool | GHDx*. (2017). Global Burden of Disease. http://ghdx.healthdata.org/gbd-results-tool