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

In [2]:
```
final_df = pd.read_csv('lung_cancer_country_wise.csv')
```

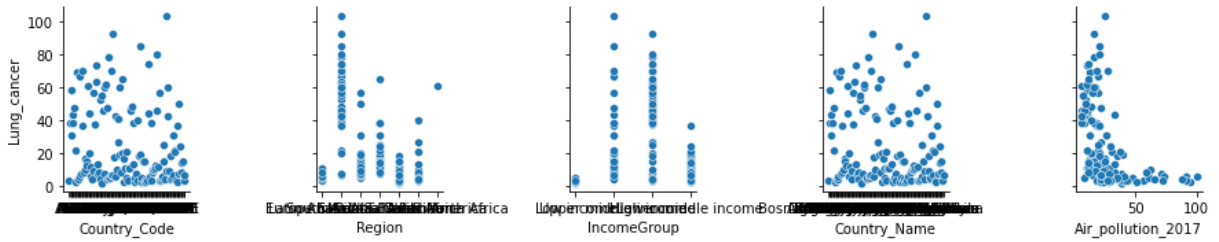
In [3]:
```
final_df
```

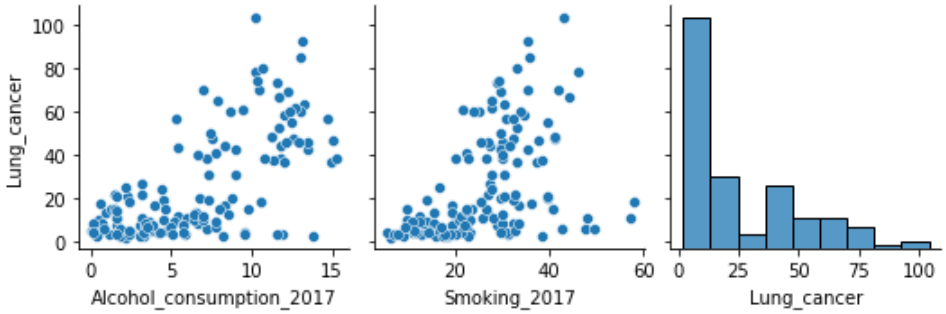
Out[3]:

	Country_Code	Region	IncomeGroup	Country_Name	Air_pollution_2017	Alcohol_consumption_2
0	AFG	South Asia	Low income	Afghanistan	56.910808	0.2
1	ALB	Europe & Central Asia	Upper middle income	Albania	18.200603	7.1
2	AND	Europe & Central Asia	High income	Andorra	10.307621	11.9
3	ARG	Latin America & Caribbean	Upper middle income	Argentina	13.311834	9.0
4	ARM	Europe & Central Asia	Upper middle income	Armenia	32.528118	5.3
...	...	...	...	...	...	
132	VUT	East Asia & Pacific	Lower middle income	Vanuatu	11.652777	1.5
133	WSM	East Asia & Pacific	Lower middle income	Samoa	11.548027	3.3
134	ZAF	Sub-Saharan Africa	Upper middle income	South Africa	25.102205	9.8
135	ZMB	Sub-Saharan Africa	Low income	Zambia	27.438035	5.0
136	ZWE	Sub-Saharan Africa	Lower middle income	Zimbabwe	22.251671	3.5

137 rows × 8 columns

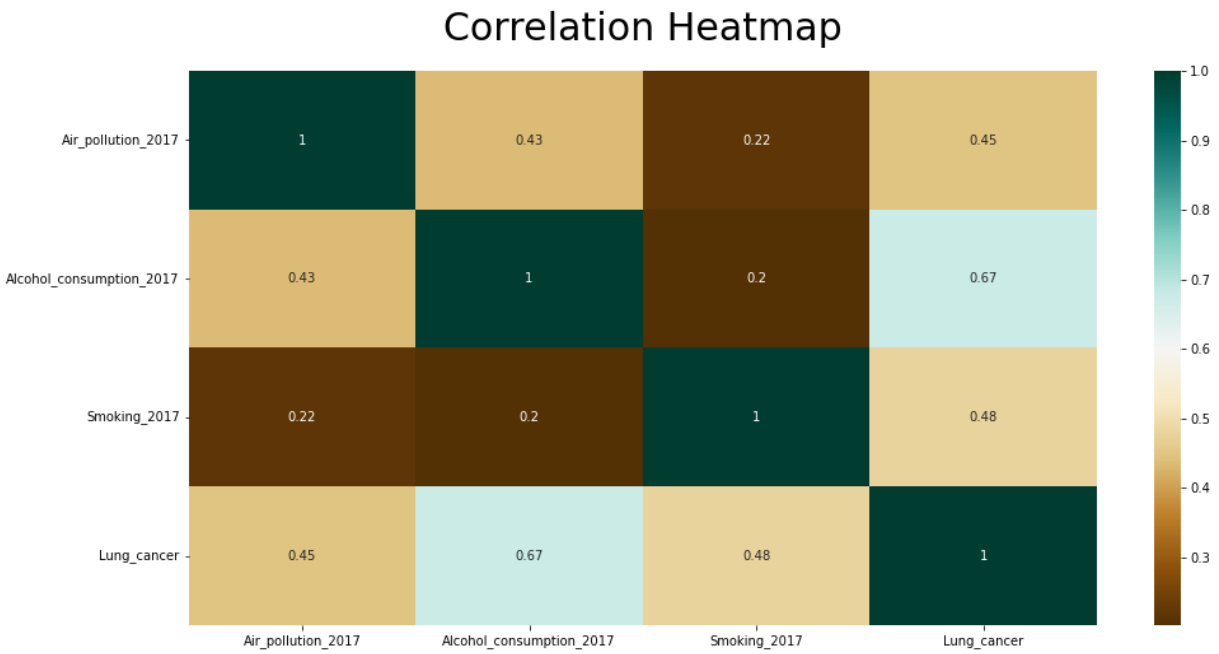
In [4]:
```
for i in range(0, len(final_df.columns), 5):
    sns.pairplot(data=final_df,
                 x_vars=final_df.columns[i:i+5],
                 y_vars=['Lung_cancer'])
```





In [5]:
```
# Correlation
plt.figure(figsize=(16, 8))
heatmap = sns.heatmap(final_df.corr().abs(),annot=True, cmap='BrBG')
heatmap.set_title('Correlation Heatmap', fontdict={'fontsize':30}, pad=24)
```

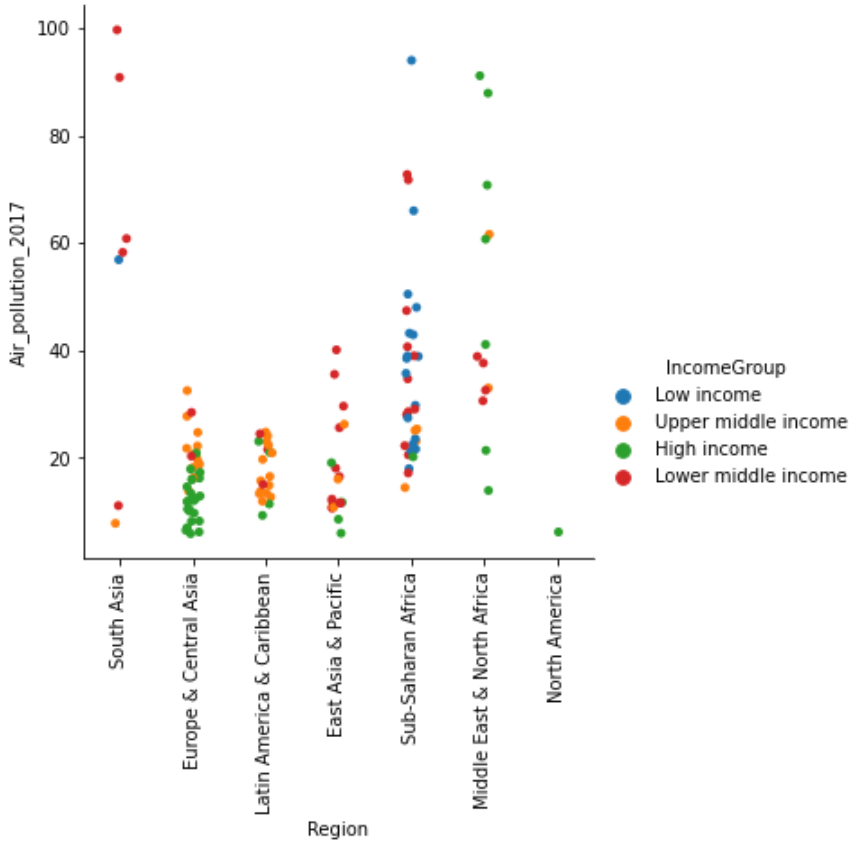
Out[5]:
```
Text(0.5, 1.0, 'Correlation Heatmap')
```

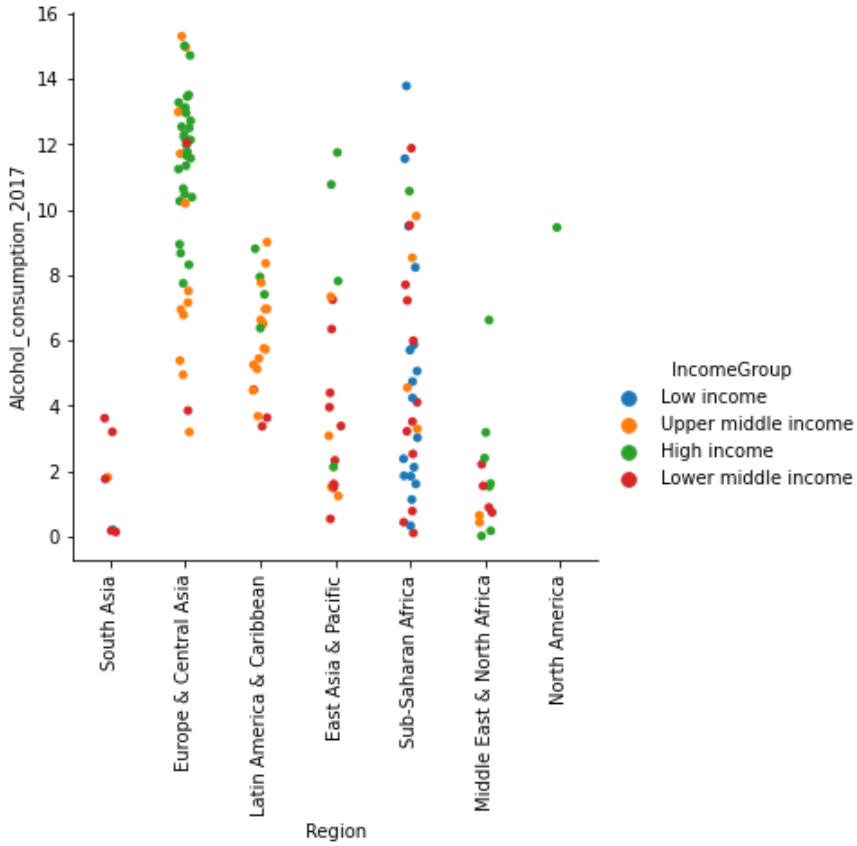


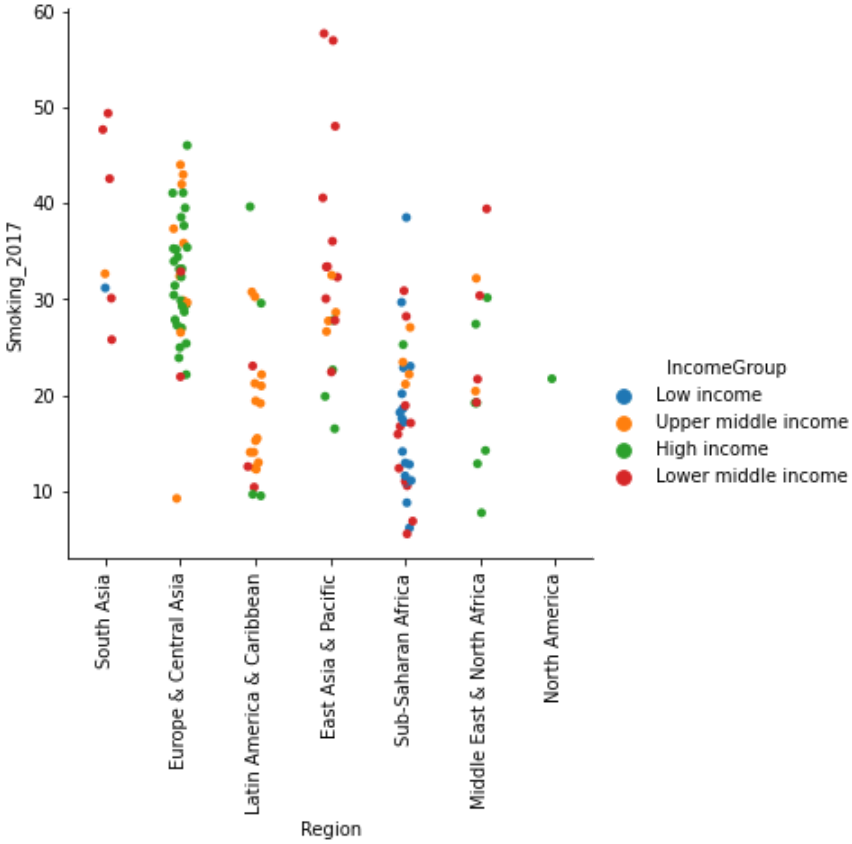
Categorical Scatterplots: Across Region

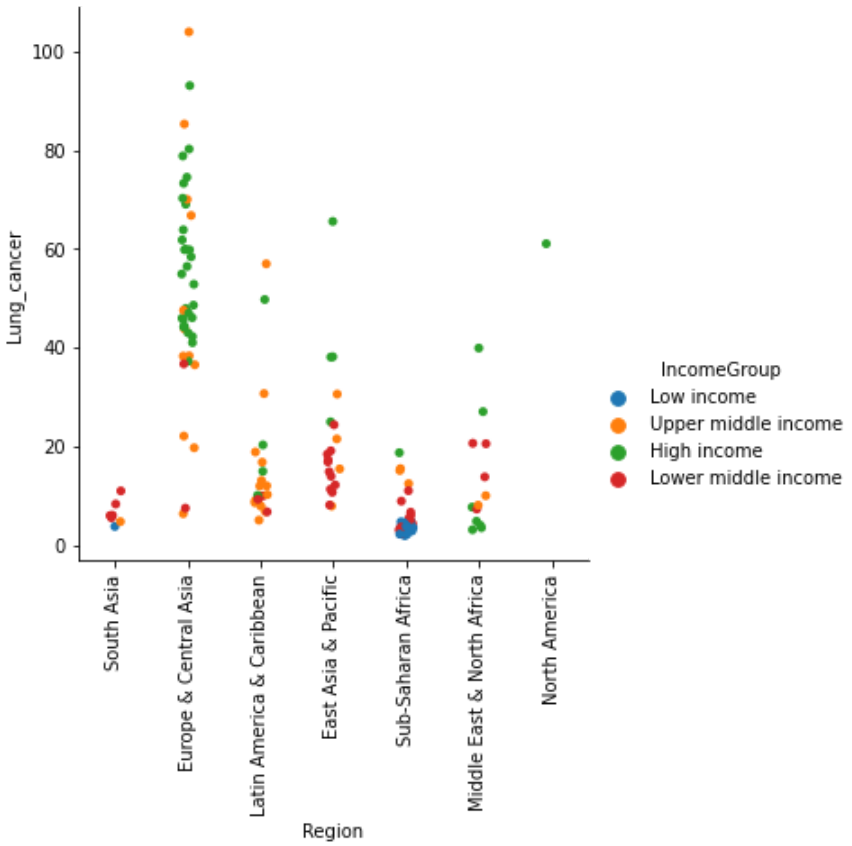
In [6]:
```
chart1 = sns.catplot(data=final_df, x="Region", y="Air_pollution_2017", hue="IncomeGro
chart1.set_xticklabels(rotation=90)

chart2 = sns.catplot(data=final_df, x="Region", y="Alcohol_consumption_2017", hue="Inc
chart2.set_xticklabels(rotation=90)

chart3 = sns.catplot(data=final_df, x="Region", y="Smoking_2017", hue="IncomeGroup", k
chart3.set_xticklabels(rotation=90)

chart4 = sns.catplot(data=final_df, x="Region", y="Lung_cancer", hue="IncomeGroup", ki
chart4.set_xticklabels(rotation=90)
```

Out[6]:
```
<seaborn.axisgrid.FacetGrid at 0x7fb03143f370>
```


Air_pollution_2017
100
80
60
40
20
South Asia
Europe & Central Asia
Latin America & Caribbean
East Asia & Pacific
Sub-Saharan Africa
Middle East & North Africa
North America
Region
IncomeGroup
Low income
Upper middle income
High income
Lower middle income


Alcohol_consumption_2017
16
14
12
10
8
6
4
2
0
South Asia
Europe & Central Asia
Latin America & Caribbean
East Asia & Pacific
Sub-Saharan Africa
Middle East & North Africa
North America
Region
IncomeGroup
Low income
Upper middle income
High income
Lower middle income



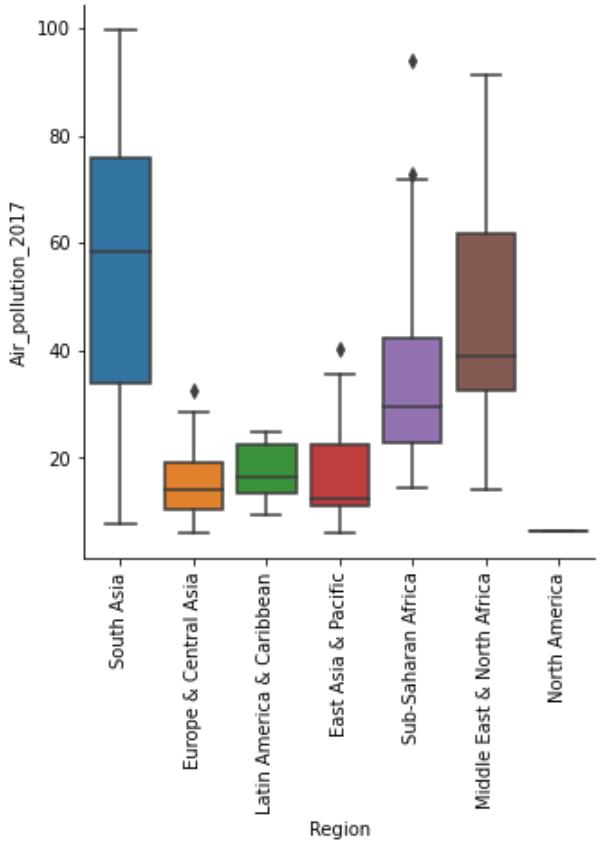


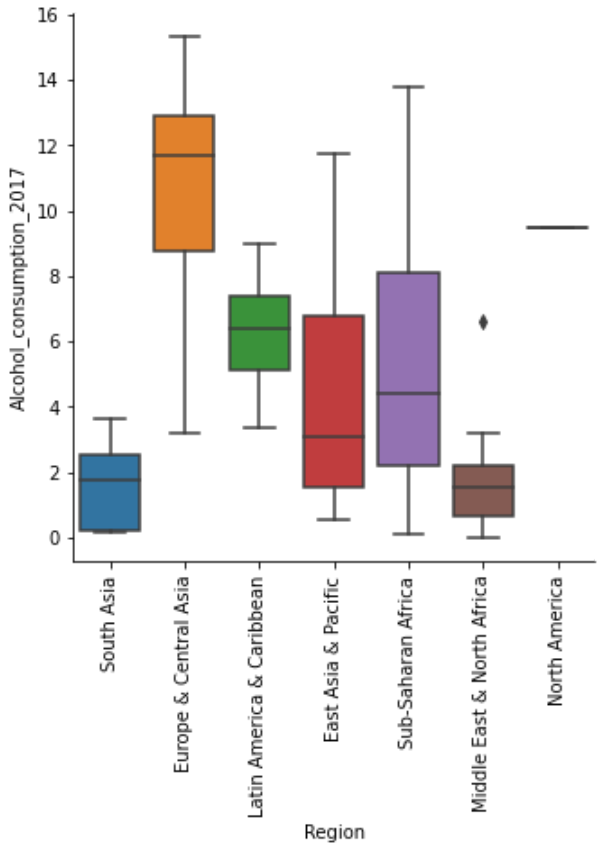
Boxplot Across region

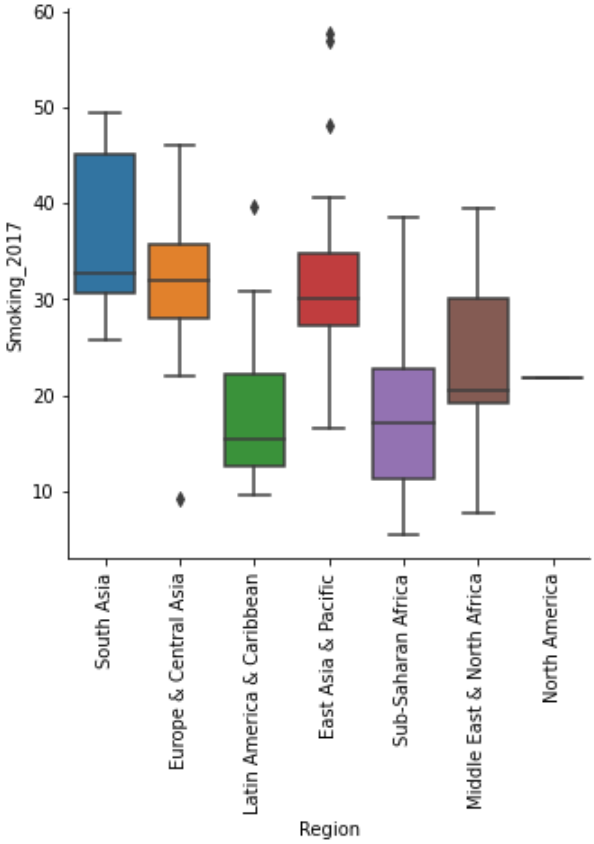
In [7]:
```
chart1 = sns.catplot(data=final_df, x="Region", y="Air_pollution_2017", kind="box")
chart1.set_xticklabels(rotation=90)

chart2 = sns.catplot(data=final_df, x="Region", y="Alcohol_consumption_2017", kind="bo
chart2.set_xticklabels(rotation=90)

chart3 = sns.catplot(data=final_df, x="Region", y="Smoking_2017", kind="box")
```

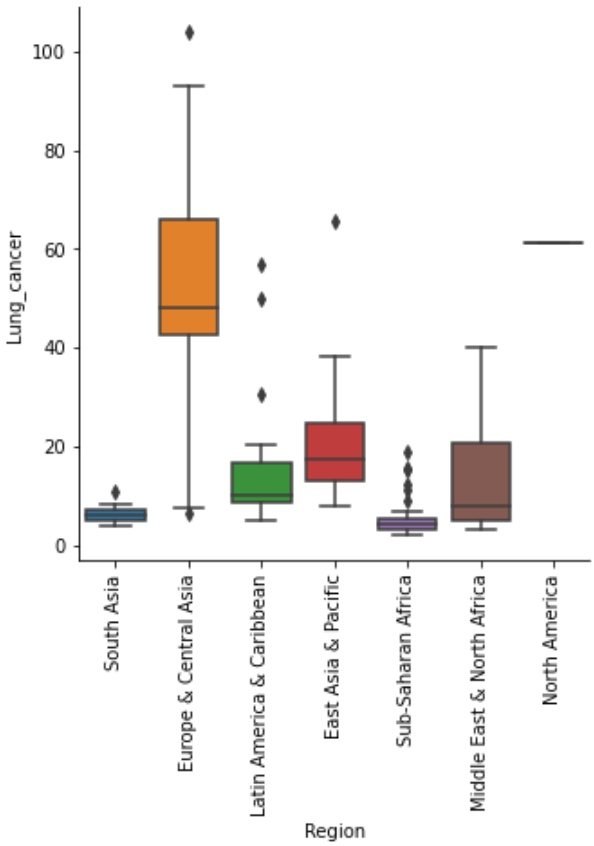
```
chart3.set_xticklabels(rotation=90)

chart4 = sns.catplot(data=final_df, x="Region", y="Lung_cancer", kind="box")
chart4.set_xticklabels(rotation=90)
```

Out[7]: <seaborn.axisgrid.FacetGrid at 0x7fb0313bf820>





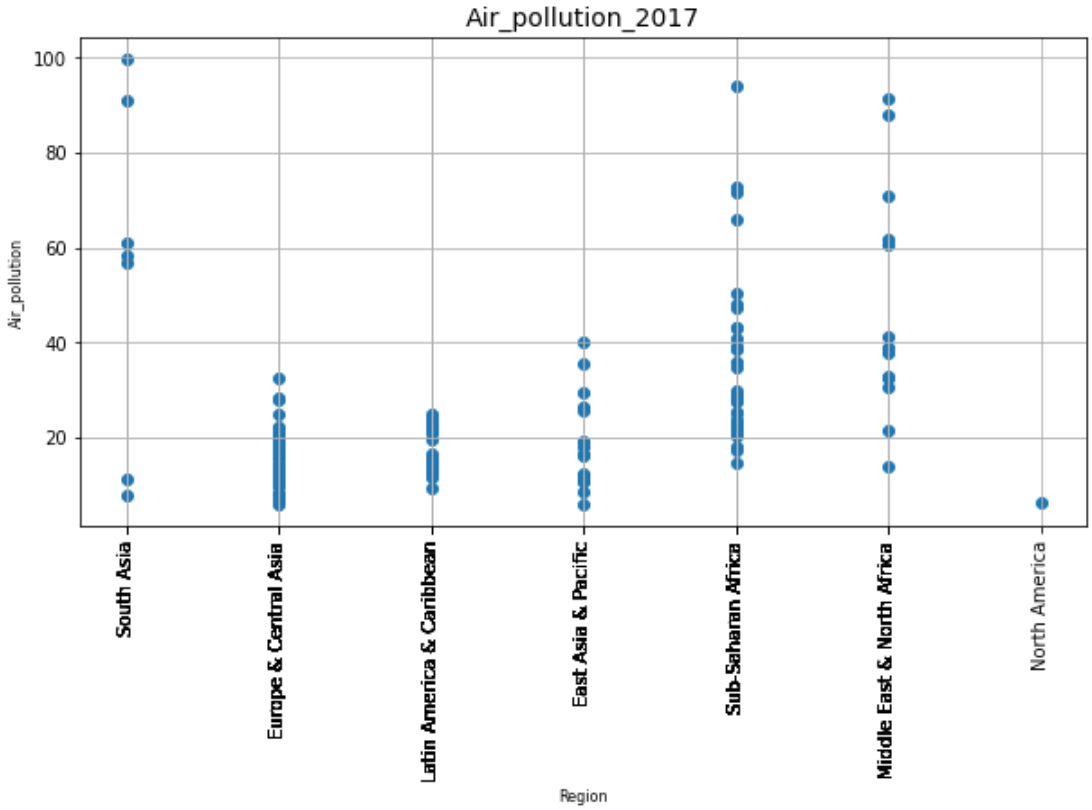


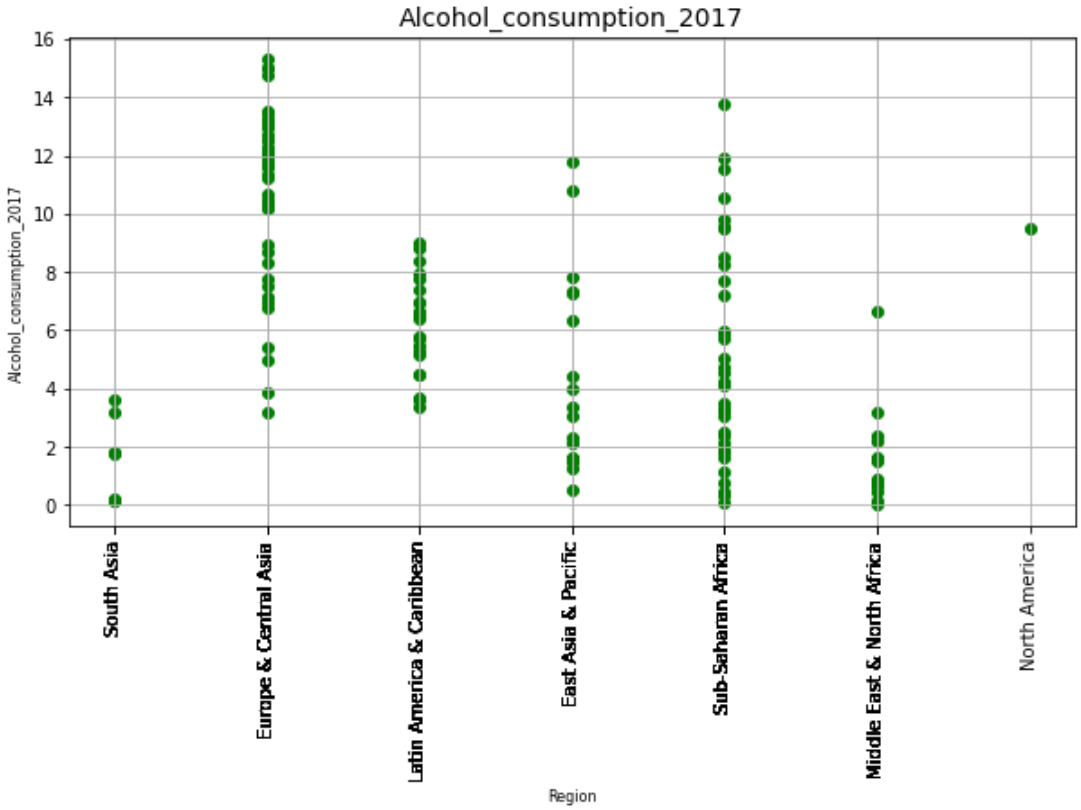


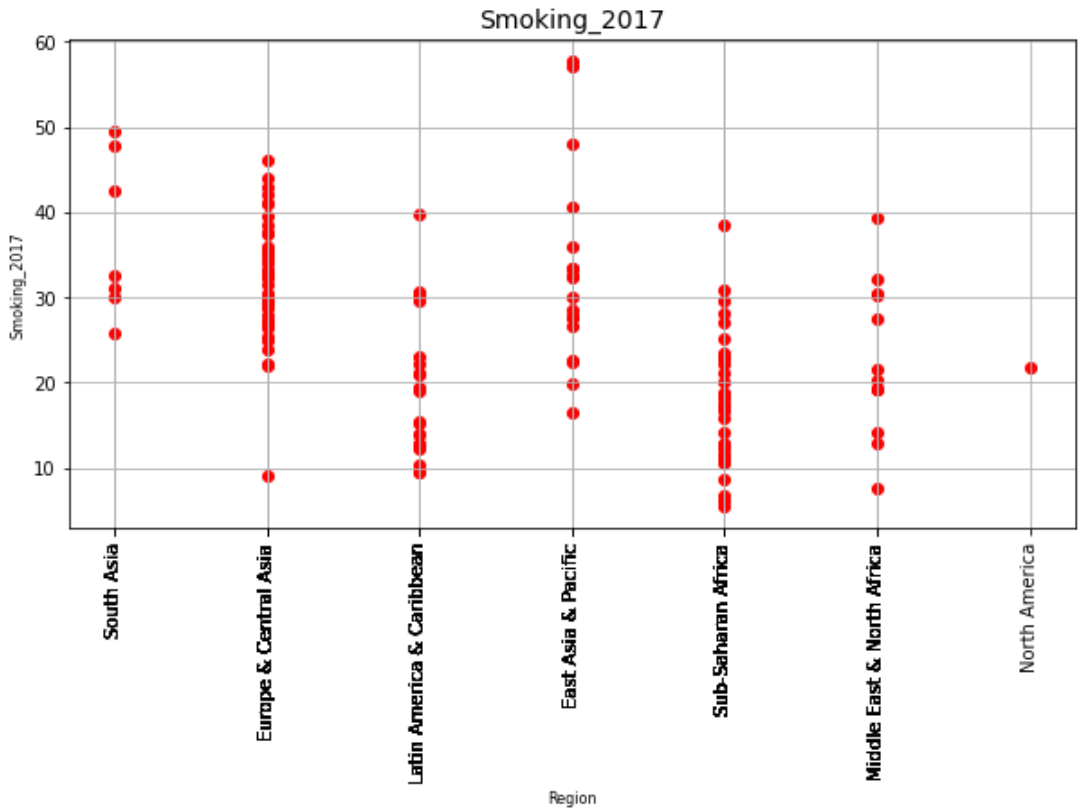
Distribution across region

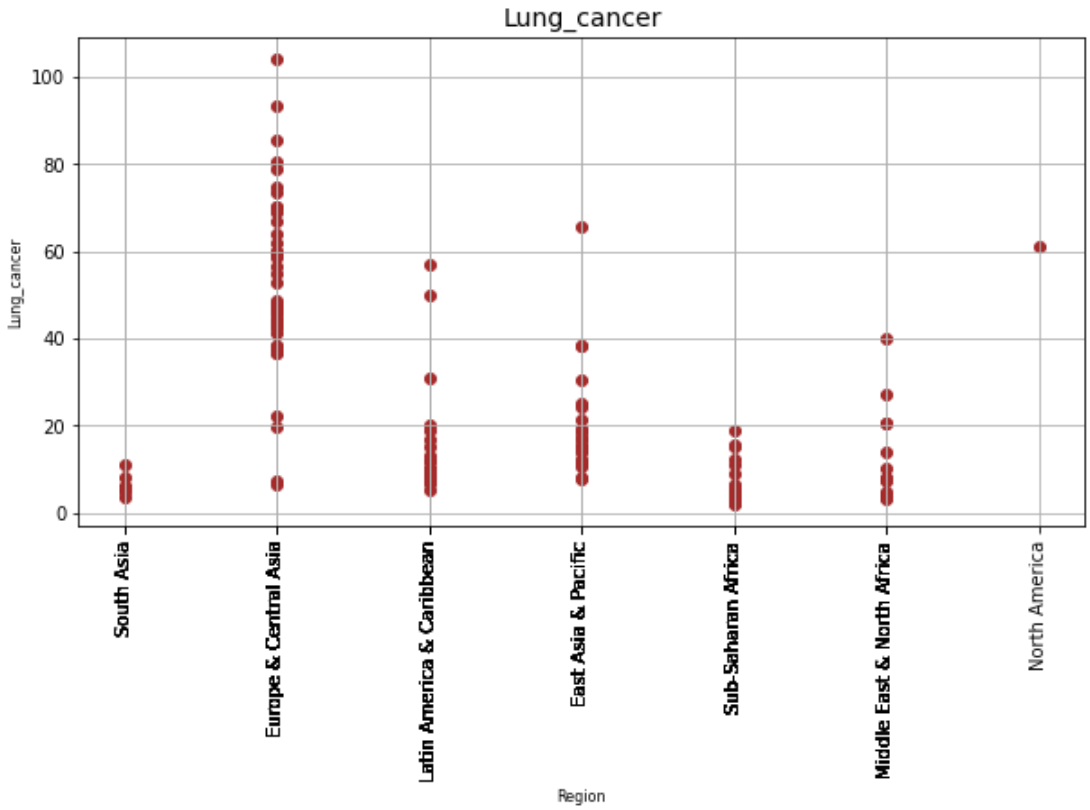
In [8]:
```
plt.figure(figsize=(10,5))
# plt.bar(final_df['Region'],final_df['gdp_per_capita_2017'],color='green') ##bar grap
plt.scatter(final_df['Region'],final_df['Air_pollution_2017'])#,color='green', alpha=0
plt.xlabel("Region", fontsize=8)
plt.ylabel("Air_pollution", fontsize=8)
plt.title("Air_pollution_2017", fontsize=14)
plt.grid(True)
```

```
plt.xticks(final_df['Region'], rotation='vertical', size=10)
plt.show();

plt.figure(figsize=(10,5))
# plt.bar(final_df['Region'],final_df['gdp_per_capita_2017'],color='green') ##bar grap
plt.scatter(final_df['Region'],final_df['Alcohol_consumption_2017'], color='green')#,
plt.xlabel("Region", fontsize=8)
plt.ylabel("Alcohol_consumption_2017", fontsize=8)
plt.title("Alcohol_consumption_2017", fontsize=14)
plt.grid(True)
plt.xticks(final_df['Region'], rotation='vertical', size=10)
plt.show();

plt.figure(figsize=(10,5))
# plt.bar(final_df['Region'],final_df['gdp_per_capita_2017'],color='green') ##bar grap
plt.scatter(final_df['Region'],final_df['Smoking_2017'], color='red')#, alpha=0.5)
plt.xlabel("Region", fontsize=8)
plt.ylabel("Smoking_2017", fontsize=8)
plt.title("Smoking_2017", fontsize=14)
plt.grid(True)
plt.xticks(final_df['Region'], rotation='vertical', size=10)
plt.show();

plt.figure(figsize=(10,5))
# plt.bar(final_df['Region'],final_df['gdp_per_capita_2017'],color='green') ##bar grap
plt.scatter(final_df['Region'],final_df['Lung_cancer'], color='brown')#, alpha=0.5)
plt.xlabel("Region", fontsize=8)
plt.ylabel("Lung_cancer", fontsize=8)
plt.title("Lung_cancer", fontsize=14)
plt.grid(True)
plt.xticks(final_df['Region'], rotation='vertical', size=10)
plt.show();
```




Alcohol_consumption_2017
Alcohol_consumption_2017
0
2
4
6
8
10
12
14
16
South Asia
Europe & Central Asia
Latin America & Caribbean
East Asia & Pacific
Sub-Saharan Africa
Middle East & North Africa
North America
Region


Smoking_2017
Smoking_2017
10
20
30
40
50
60
South Asia
Europe & Central Asia
Latin America & Caribbean
East Asia & Pacific
Sub-Saharan Africa
Middle East & North Africa
North America
Region


Lung_cancer
100
80
60
40
20
0
Lung_cancer
South Asia
Europe & Central Asia
Latin America & Caribbean
East Asia & Pacific
Sub-Saharan Africa
Middle East & North Africa
North America
Region

In [ ]: