Effectiveness of Government Policies on Smoker Percentages

Team 6 - David, Giovanni, Harrison, and Ranadeep

Question:

Are government policies effective in reducing smoker percentages? How does such government intervention account for gender? Does the effectiveness of this intervention depend on the country's per-capita GDP?

Non-Technical Executive Summary:

Smoking has been a major contributing factor in many different forms of cancer and medical diseases. However, recent information about smoking has caused an overall decline in usage, and government entities have focused on banning smoking advertisements and providing quitting support to smokers. With these recent government actions, we explore their effects on both men and women and examine the impact that GDP can have on smoking trends.

When exploring the data, we see that GDP has a positive correlation with the number of smokers in that country. While several factors could explain this trend, a closer examination of the data reveals a trend that indicates there are disproportionately more female smokers in countries with higher GDP but shows the opposite trend for men. When testing the connection between GDP and government assistance programs for smokers, we see that countries that provide more help to help smokers quit their habit also tend to be wealthier. This indicates that there is a discrepancy in attracting female smokers to quitting programs, and society can reduce their smoking populations by addressing this issue.

1. Introduction

Smoking risks like lung cancer and heart disease have been increasingly exposed, leading to more informed smoking decisions. With clear evidence to support the dangers of the habit, there has been a steady decline in the daily usage of tobacco since the early 2000s. Examining the daily sales of several countries, we see empirical observations confirm this trend (Figure 4).

When accounting for both the change in cigarette prices and government support for quitting for 2012 to 2014, there has mostly been a decrease in the percentage of the daily sales of tobacco products. Note here that government help did not change from 2012 to 2014. While Bulgaria and Lithuania present opposite sales trends, they remain outliers in the overall decrease in smoking sales. There is then a question about the actual smoking usage in this same time frame. When we look across the percentage of smokers in these same countries, there is still a clear decline in smoking usage, even when looking across males and females (Figure 5).

Note that both Lithuania and Bulgaria show negative changes to smoking percentages across men, women, and both genders. This indicates that the previous increase in tobacco sales does not provide counterevidence of a decline in smoking usage. The only country where we see a usage increase is in Slovakian females, but this appears to be an outlier to the rest of the data. In addition to the data, government intervention in smoking has remained strong within the 2000s. Both government restrictions on advertisements and support for quitting have been heightened because of the response to smoking risks, but these two factors must be explored to determine their effectiveness in reducing smoking trends.

2. Restricting Smoking Advertisements

Question: What causes advertising bans to be more effective in decreasing the percentage of people who smoke?

We split countries into three income brackets, based on per-capita GDP (PPP):
(i) [$0, $7,499]
(ii) [$7,500, $24,999]
(iii) [$25,000+]
Each group has approximately an equal number of countries, and show the effect of advertising bans within each group by comparing the degree of ban enforcement (either no bans on advertising, a ban on only TV, news, and radio ads, or a ban on both direct and indirect advertising).

In each income group, the trend of enforcement of ad bans was that only in the high-income bracket was there a trend of higher direct ad enforcement (banning on TV, radio, newspaper) leading to the lowest tobacco usage, which likely means that counties with a higher income brackets are better able to enforce a ban on indirect advertising compared to those that don’t have as much income.

Why Violin Plots?

We use violin plots as a visual representation of the spread of % smokers within each degree of ad ban enforcement for each of the 3 income groups. Violin plots are for visualization purposes only to convey the difference in distribution between the % of smokers (unlike the boxplot, the violin plots have a thickness to further describe frequency distribution). We will later conduct a T-Test to see if the differences are statistically significant. The X-axis of violin plots:

2 - no bans on tobacco advertising

4 - bans on direct advertising of tobacco (ads on TV, radio and news, etc.)

5 - bans on indirect and direct advertising of tobacco (4 plus indirect forms such as tobacco being shown on films)

Figure 1: Violin Plot of Enforcement of Bans on Tobacco Ad in low-income countries

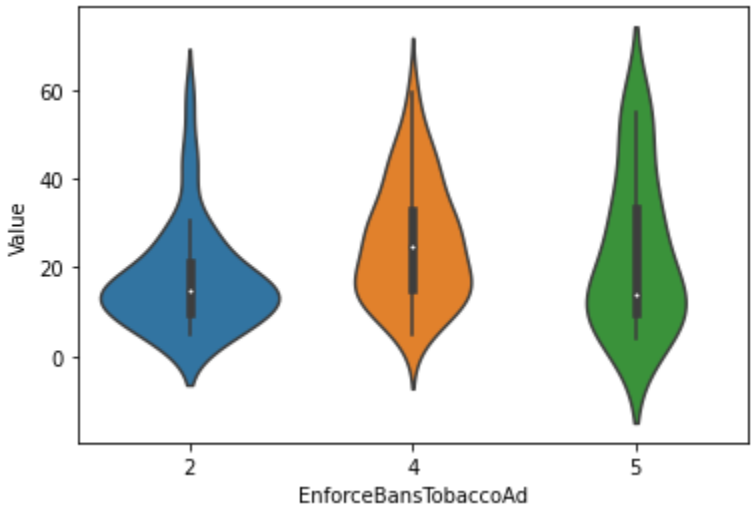


Figure 2: Violin Plot of Enforcement of Bans on Tobacco Ad in middle-income countries

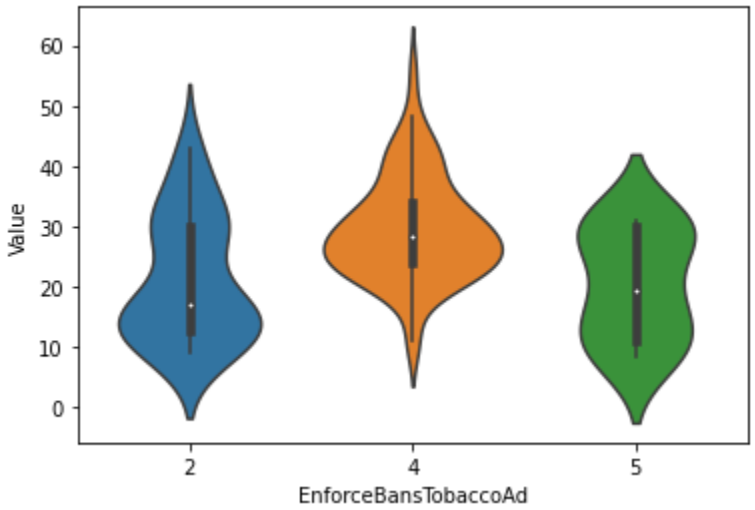
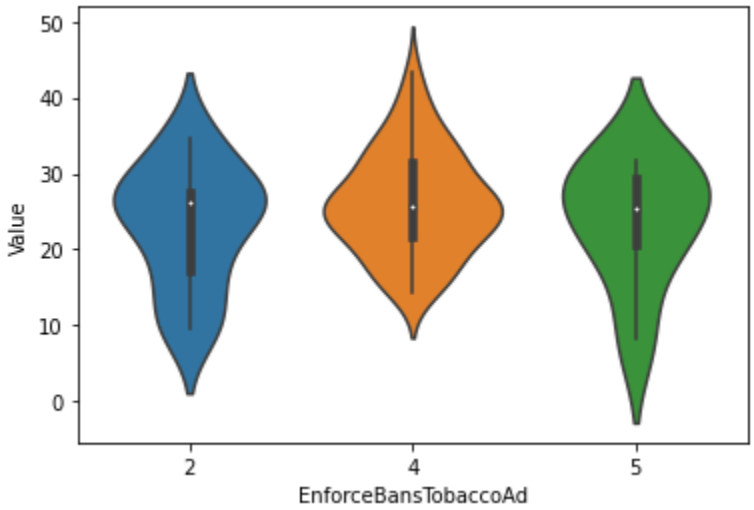


Figure 3: Violin Plot of Enforcement of Bans on Tobacco Ad in high-income countries



We see from Violin Plot Figures 1, 2, and 3 and in Technical References (1*) that the changes between groups 2 and 4 are significant for low and middle-income countries, and the changes between groups 4 and 5 are significant only for the middle-income countries.

We conclude that banning direct advertising correlates with higher cigarette use in low and middle-income countries while banning indirect advertising is more effective for middle-income countries.

3. Helping Smokers Quit

Various governments have helped smokers with quitting assistance. For countries with data about smoking percentages, there has at least been a small amount of government aid provided to help smokers quit their habit. Out of 149 countries in 2010 and 2014, none provided a “help score” of less than 3 (which indicates there were at least some cessation services for smokers). However, the data does not indicate that an increase in government help leads to a decrease in smoking percentage (Figure 6).

There is not a decrease in the value of smoking percentage when considering the help categories 3 and 4. When comparing the effectiveness of help programs on males and females, we see this irregularity also appears in females (Figure 7) but not in males (Figure 8).

Even when including enforcement ban categories in our regression, we see this same trend in females (Figure 9) and the lack of this within males (Figure 10).

With these linear regression models, quitting help appears to be beneficial for men whereas there does not appear to be any significant impact of these programs on women.

4. Social Examination

4.1 Impact of GDP on Smoking

Government intervention is not the only significant factor in smoker percentages. One contributing factor which has not yet been examined is a country's GDP. When looking at the connection between GDP and smoker percentages for both sexes in the years 2010 and 2014, we see that countries with greater GDPs tend to have more smokers in their country, ref Technical References (Figure 11).

There are several hypotheses on why we see this trend. Firstly, higher GDPs could indicate that citizens have more expendable income to spend on cigarettes, as more expendable income allows people to make purchases on cigarettes, whereas those in poorer countries do not have this liberty and instead need to direct their income to essential items such as food and rent.

Another possible explanation for the increase in smoker percentage with an increase of GDP is due to the healthcare availability of these countries. Countries with greater GDPs will be able to regularly visit healthcare physicians and stay updated on their physical health. In a country with a lower GDP, there could be reduced access to healthcare services, which means that people are more cautious with risky habits. It is then worth including GDP in our regression models from section (3) to test the impact of GDP on smoking.

Let us see if there is a significant connection between GDP and the amount of government help for smokers. When testing the three available categories of smoker help against GDP, we see that higher levels of smoker assistance tend to have higher values of GDP, ref Technical References (Figure 12). This indicates that wealthier countries will provide more help for smokers to quit their habit. This could be another explanation for the positive correlation between GDP and smoker percentages, as people in wealthier countries realize they have more opportunities to quit smoking if they ever wanted to stop.

4.2 Trends of Price and Tax with GDP

A potential reason for this can be seen with positive correlations between log(GDP) with both average price and tax percentage. Signaling an extra deterrent for those wanting to purchase cigarettes, ref. Technical References (Figure 13).

Following this analysis, we regressed log(GDP), enforcement ban, and help quit on average prices to find that the majority of the model data points were underestimating the true cost due to a significant set of residuals that overperformed/more expensive than what the model predicted. (Figure 14)

Further examination of the two groups of nations (residual above 1 and those under), gives us a significant difference in means of log(GDP) between these groups such that we can say that under 95% Confidence Interval (CI) that the group of residuals above 1 are significantly higher mean for log(GDP). (Figure 15)

4.3 Results when Including GDP

The inclusion of GDP has had a significant impact on the female smoking population in terms of government smoking assistance. When GDP is included in our model, we see that each increased level of help sees a more negative slope for the percentage of smokers (Figure 16). That is, the female population now appears to align more with our expectations of the percentage of smokers, where more help should lead to fewer smokers. Men also show a similar trend, which indicates that GDP should be included in our model for predicting the smoker percentage of a population (Figure 17).

Examining the GDP to smoker percentage trend for men and women individually yields an unexpected result. Looking at the male population, we see that increased values of GDP are negatively correlated with higher smoker percentages (Figure 18), whereas women follow the trend that was present in both sexes (Figure 19).

We find that while GDP has an overall positive correlation to the smoker percentage of a country, there seems to be a discrepancy between men and women in these same countries.

First, let us address reasons why men may not follow the general trend for both genders. Examining the plot of GDP to smoker percentages for men, we see a reverse fanning pattern, which means that countries with higher GDPs tend to have similar smoking percentages. For example, in countries around a log(GDP) of 7, we see both a smoking percentage about as high as 80% and about as low as 20%, whereas countries about log(GDP) of 11 have smoker values within a range of 15% to 40%.

It appears as if wealthier countries have a more concentrated smoking percentage. Note from before that wealthier countries also have more access to government quitting programs, which is a reasonable explanation for this trend in the male population. However, under the assumption

that men benefit from greater access to programs that help them stop smoking, we must now explain why women do not benefit in this same way.

4.4 Increased Effectiveness of Government Programs

People in wealthier countries tend to have more government services to help them quit their smoking habit. However, we have seen consistent discrepancies in the female smoking population, even within these higher GDP countries. While there are many potential explanations for this trend, it is clear that women are not impacted as much by these programs as men. This could indicate that the target audience for these programs is men or that the programs themselves are less welcoming to female smokers (either in their design or implementation).

5. Conclusion

If people want to further continue the decreasing trend in smokers, society should encourage the government to restructure their quitting assistance programs to be more open to women. If these services in wealthier countries were more available to female smokers, the focus should shift towards assisting other countries less able to provide these services because of their smaller GDPs.

6. Technical References

Figure 1:

Continuation from Violin Plot on Page 3, examining the GDP to smoker percentage trend for men and women individually yields an unexpected result. Looking at the male population, we see a low-income country that either has no bans or advertising bans on direct ads or bans on direct and indirect ads in 2014.

Figure 2:

Continuation from Violin Plot on Page 3, enforcement on Bans on Tobacco Ads vs. % of smokers in middle-income countries, where each point in a violin plot is a middle-income country that either has no bans on ads, bans on direct ads, or bans on direct and indirect ads in 2014.

Figure 3:

Continuation from Violin Plot on Page 4, enforcement on Bans on Tobacco Ads vs. % of smokers in high-income countries, where each point in a violin plot is a high-income country that either has no bans on ads, bans on direct ads or bans on direct and indirect ads in 2014.

Figure 4:

Code	Country	HelpToQuit	2012to2014CigarettePriceChange(%)	2012to2014DailySaleChange(%)
BGR	Bulgaria	4	8.51	1.72
CZE	Czechia	4	15.89	-16.98
EST	Estonia	4	0.00	-4.26
DEU	Germany	4	10.45	-2.38
ISL	Iceland	3	19.81	-14.81
JPN	Japan	4	7.43	-8.16
LVA	Latvia	4	55.68	-3.45
LTU	Lithuania	4	6.02	14.29
SVK	Slovakia	4	6.22	-7.14
SVN	Slovenia	4	18.46	-20.69
GBR	United Kingdom	5	3.90	-11.76
USA	United States	5	0.00	-8.57
ISR	Israel	5	NaN	-15.38

Our sample size is quite small, as there were not many countries that had daily sales data in 2012 and 2014. However, we are examining a random sample of countries, which means we can predict this trend will be the same worldwide. We do know that most countries have seen a decrease in smokers over this timeframe as well.

Figure 5:

	Country	2012to2014FemaleValueChange(%)	2012to2014BothSexesValueChange(%)	2012to2014MaleValueChange(%)
BGR	Bulgaria	-1.10	-2.84	-3.95
CZE	Czechia	-0.72	-1.23	-1.61
EST	Estonia	-1.20	-4.14	-5.65
DEU	Germany	-1.47	-2.35	-2.78
ISL	Iceland	-10.11	-11.35	-13.02
JPN	Japan	-4.24	-5.91	-6.17
LVA	Latvia	-0.41	-1.06	-1.35
LTU	Lithuania	-2.03	-3.33	-4.22
SVK	Slovakia	2.50	0.00	-1.50
SVN	Slovenia	-1.41	-0.85	-0.78
GBR	United Kingdom	-11.01	-9.09	-7.39
USA	United States	-3.29	-3.30	-2.71
ISR	Israel	-5.52	-2.56	-1.10

We look at the change in smoker percentages across men and women in case there were any trends specific to gender.

(1*)

We see that there appears to be a trend for banning tobacco ads within the low and middle-income groups. We perform a T-test to see if the trend is significant where within each income group the % of smokers within countries with the 3 present categories of a ban on tobacco ads are compared. Here are the p-values:

P-Value List:
Income group: low
2 and 4: 0.0032881028962124114
2 and 5: 0.13105239519988035
4 and 5: 0.5575360605663711
Income group: mid
2 and 4: 0.00013745861200658657
2 and 5: 0.8235146575597221
4 and 5: 0.0008037931921276979
Income group: high
2 and 4: 0.09215138849393427
2 and 5: 0.8884067565650492
4 and 5: 0.36592074977745104

So the difference between % smokers in countries with no ad bans and countries with a ban on direct ads is significant for low income, and the difference between % smokers in countries with

no ad bans and countries with a ban on direct ads is significant for the middle-income group as well as the difference between a ban on direct ads and ban on indirect advertising as well.

Figure 6:

OLS Regression Results

Dep. Variable:	Value	**R-squared:**	0.017
Model:	OLS	**Adj. R-squared:**	0.006
Method:	Least Squares	**F-statistic:**	1.596
Date:	Sun, 21 Nov 2021	**Prob (F-statistic):**	0.191
Time:	18:54:20	**Log-Likelihood:**	-1086.1
No. Observations:	286	**AIC:**	2180.
Df Residuals:	282	**BIC:**	2195.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	21.4737	2.492	8.615	0.000	16.567	26.380
C(HelpToQuit)[T.3]	3.0615	2.748	1.114	0.266	-2.348	8.472
C(HelpToQuit)[T.4]	4.1770	2.659	1.571	0.117	-1.056	9.410
C(HelpToQuit)[T.5]	0.7702	3.015	0.255	0.799	-5.165	6.705

Omnibus:	12.738	**Durbin-Watson:**	1.817
Prob(Omnibus):	0.002	**Jarque-Bera (JB):**	13.148
Skew:	0.500	**Prob(JB):**	0.00140
Kurtosis:	3.321	**Cond. No.**	9.31

While the three values have wide ranges for their slopes (high standard error), we see that the most help yields the smallest slope (i.e. less percentage of smokers), while a help ranking of 4 has a greater slope compared to a help ranking of 3.

Figure 7:

OLS Regression Results

Dep. Variable:	Value	**R-squared:**	0.028
Model:	OLS	**Adj. R-squared:**	0.018
Method:	Least Squares	**F-statistic:**	2.718
Date:	Sun, 21 Nov 2021	**Prob (F-statistic):**	0.0449
Time:	17:47:34	**Log-Likelihood:**	-1098.4
No. Observations:	286	**AIC:**	2205.
Df Residuals:	282	**BIC:**	2219.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	8.9947	2.603	3.456	0.001	3.872	14.118
C(HelpToQuit)[T.3]	2.3598	2.870	0.822	0.412	-3.289	8.009
C(HelpToQuit)[T.4]	5.8574	2.776	2.110	0.036	0.393	11.322
C(HelpToQuit)[T.5]	5.1882	3.149	1.648	0.101	-1.009	11.386

Omnibus:	51.628	**Durbin-Watson:**	1.936
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	79.298
Skew:	1.074	**Prob(JB):**	6.04e-18
Kurtosis:	4.429	**Cond. No.**	9.31

Female data has a trend similar to that of both sexes, where a help ranking of 3 has a smaller slope than that of ranking 4.

Figure 8:

OLS Regression Results

Dep. Variable:	Value	**R-squared:**	0.028
Model:	OLS	**Adj. R-squared:**	0.017
Method:	Least Squares	**F-statistic:**	2.690
Date:	Sun, 21 Nov 2021	**Prob (F-statistic):**	0.0466
Time:	17:47:34	**Log-Likelihood:**	-1167.7
No. Observations:	286	**AIC:**	2343.
Df Residuals:	282	**BIC:**	2358.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	33.9421	3.316	10.235	0.000	27.414	40.470
C(HelpToQuit)[T.3]	3.7863	3.657	1.035	0.301	-3.412	10.984
C(HelpToQuit)[T.4]	2.4985	3.537	0.706	0.481	-4.464	9.461
C(HelpToQuit)[T.5]	-3.6348	4.012	-0.906	0.366	-11.531	4.262

Omnibus:	5.286	**Durbin-Watson:**	1.717
Prob(Omnibus):	0.071	**Jarque-Bera (JB):**	5.310
Skew:	0.333	**Prob(JB):**	0.0703
Kurtosis:	2.953	**Cond. No.**	9.31

Males, in contrast to females, show a steady downward trend in slope with more government quitting help. However, we must explore this relationship further since our model has not included other contributing factors to smoker percentage.

Figure 9:

OLS Regression Results

Dep. Variable:	Value	**R-squared:**	0.120
Model:	OLS	**Adj. R-squared:**	0.104
Method:	Least Squares	**F-statistic:**	7.648
Date:	Sun, 21 Nov 2021	**Prob (F-statistic):**	9.35e-07
Time:	19:06:22	**Log-Likelihood:**	-1084.2
No. Observations:	286	**AIC:**	2180.
Df Residuals:	280	**BIC:**	2202.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	6.6058	2.545	2.596	0.010	1.596	11.615
C(HelpToQuit)[T.3]	0.2335	2.776	0.084	0.933	-5.231	5.698
C(HelpToQuit)[T.4]	2.5431	2.722	0.934	0.351	-2.816	7.902
C(HelpToQuit)[T.5]	1.9837	3.082	0.644	0.520	-4.083	8.050
C(EnforceBansTobaccoAd)[T.4]	7.9616	1.516	5.252	0.000	4.978	10.946
C(EnforceBansTobaccoAd)[T.5]	2.7910	2.223	1.255	0.210	-1.585	7.167

Omnibus:	46.902	**Durbin-Watson:**	1.946
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	73.708
Skew:	0.963	**Prob(JB):**	9.88e-17
Kurtosis:	4.572	**Cond. No.**	10.8

Even when including advertising restrictions, we see a similar trend as before with female smokers. We do see that the help categories appear to contribute less to smoking percentage when accounting for advertising restrictions.

Figure 10:

OLS Regression Results

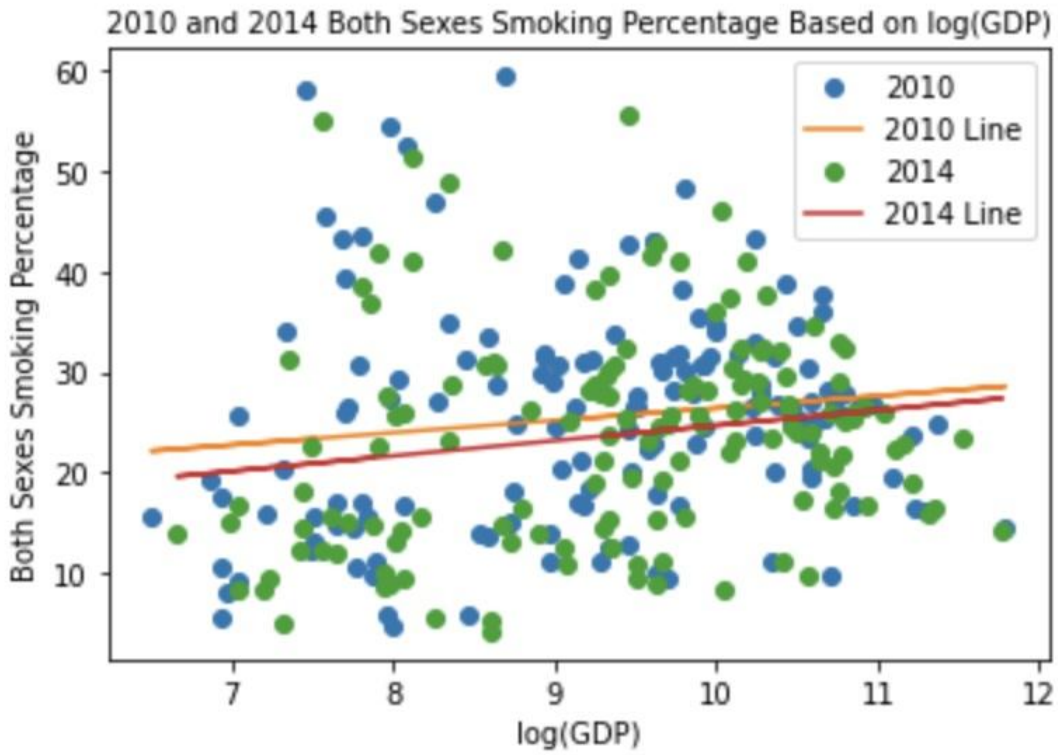
Dep. Variable:	Value	**R-squared:**	0.085
Model:	OLS	**Adj. R-squared:**	0.068
Method:	Least Squares	**F-statistic:**	5.171
Date:	Sun, 21 Nov 2021	**Prob (F-statistic):**	0.000148
Time:	19:05:59	**Log-Likelihood:**	-1159.1
No. Observations:	286	**AIC:**	2330.
Df Residuals:	280	**BIC:**	2352.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	31.1859	3.307	9.430	0.000	24.676	37.696
C(HelpToQuit)[T.3]	1.4446	3.608	0.400	0.689	-5.657	8.546
C(HelpToQuit)[T.4]	-0.8453	3.538	-0.239	0.811	-7.809	6.119
C(HelpToQuit)[T.5]	-7.1882	4.005	-1.795	0.074	-15.072	0.695
C(EnforceBansTobaccoAd)[T.4]	8.2041	1.970	4.165	0.000	4.326	12.082
C(EnforceBansTobaccoAd)[T.5]	5.6735	2.889	1.964	0.051	-0.013	11.360

Omnibus:	15.013	**Durbin-Watson:**	1.682
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	15.847
Skew:	0.547	**Prob(JB):**	0.000362
Kurtosis:	3.363	**Cond. No.**	10.8

The smoker percentage decrease appears to drop more aggressively with higher levels of help than before. When accounting for advertising restrictions, there is a more apparent decrease in the slope of quitting help within the male population.

Figure 11:



Countries with higher GDPs will see greater percentages of smokers, which is why their slopes for 2010 and 2014 are positive.

Figure 12:

OLS Regression Results

Dep. Variable:	GDP	**R-squared:**	0.412
Model:	OLS	**Adj. R-squared:**	0.406
Method:	Least Squares	**F-statistic:**	65.84
Date:	Sun, 21 Nov 2021	**Prob (F-statistic):**	2.69e-32
Time:	19:06:49	**Log-Likelihood:**	-383.16
No. Observations:	286	**AIC:**	774.3
Df Residuals:	282	**BIC:**	788.9
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	7.5311	0.213	35.284	0.000	7.111	7.951
C(HelpToQuit)[T.3]	1.0422	0.235	4.428	0.000	0.579	1.506
C(HelpToQuit)[T.4]	2.1050	0.228	9.246	0.000	1.657	2.553
C(HelpToQuit)[T.5]	2.8687	0.258	11.110	0.000	2.360	3.377

Omnibus:	3.597	**Durbin-Watson:**	1.828
Prob(Omnibus):	0.166	**Jarque-Bera (JB):**	3.015
Skew:	-0.151	**Prob(JB):**	0.221
Kurtosis:	2.598	**Cond. No.**	9.31

Note that we have log-transformed our GDP data to create a better regression model. The model shows that countries with more help for smokers appear to have higher GDP values. This would indicate that richer countries provide more help for smokers. Also, note that the slopes of these variables are significant since the p-values for each one are below a 0.05 significance level.

Figure 13:

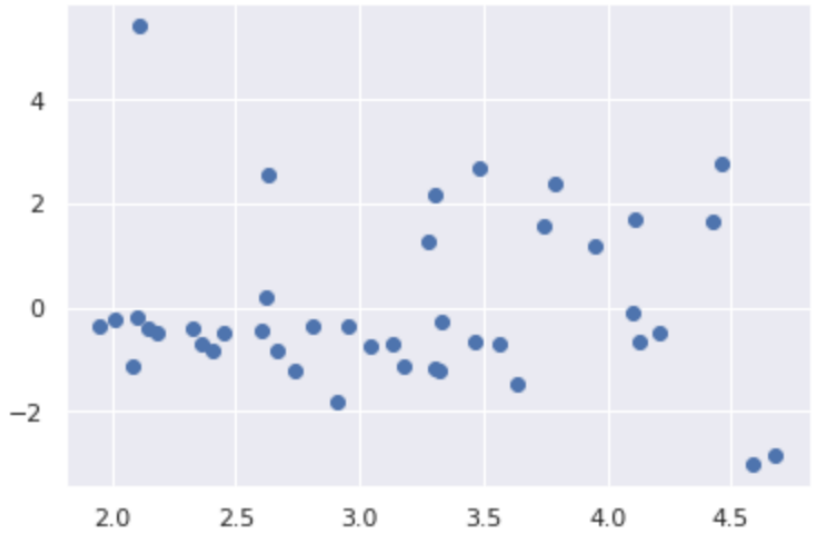
	Year	AvgCigarettePriceDollars	AvgTaxesAsPctCigarettePrice	EnforceBansTobaccoAd	HelpToQuit	Year1	GDP	LogGDP
Year	1.000000	0.030703	0.015790	0.279130	0.436648	1.000000	0.024037	0.038433
AvgCigarettePriceDollars	0.030703	1.000000	0.446901	0.082855	0.411391	0.030703	0.271861	0.369325
AvgTaxesAsPctCigarettePrice	0.015790	0.446901	1.000000	0.101067	0.350546	0.015790	0.028140	0.276539
EnforceBansTobaccoAd	0.279130	0.082855	0.101067	1.000000	0.183614	0.279130	0.043584	-0.037933
HelpToQuit	0.436648	0.411391	0.350546	0.183614	1.000000	0.436648	0.561506	0.619103
Year1	1.000000	0.030703	0.015790	0.279130	0.436648	1.000000	0.024037	0.038433
GDP	0.024037	0.271861	0.028140	0.043584	0.561506	0.024037	1.000000	0.891095
LogGDP	0.038433	0.369325	0.276539	-0.037933	0.619103	0.038433	0.891095	1.000000

Correlation Matrix between the covariates of the data frame containing GDP and log(GDP) values. Note the positive values for log(GDP) with price and tax.

Figure 14:

```
plt.scatter(clf1214price.predict(X), y - clf1214price.predict(X))
```

```
<matplotlib.collections.PathCollection at 0x7f19dea85490>
```



Residual plot of the regression on ww_the data with the addition of GDP (Avg Price ~ Enforce + Help to Quit + log(GDP)). Notice majority were pushed downwards by the group of residuals over 1.

Figure 15:

```
df_2012_2014.loc[resid_low_lst].agg(["mean", "var"])
```

	Year	AvgCigarettePriceDollars	AvgTaxesAsPctCigarettePrice	EnforceBansTobaccoAd	HelpToQuit	Year1	GDP	LogGDP
mean	2013.032258	2.183226	46.422581	3.709677	3.677419	2013.032258	1.638125e+04	9.243121
var	1.032258	0.501023	263.243140	1.012903	0.559140	1.032258	3.198068e+08	0.943257

```
df_2012_2014.loc[resid_high_lst].agg(["mean", "var"])
```

	Year	AvgCigarettePriceDollars	AvgTaxesAsPctCigarettePrice	EnforceBansTobaccoAd	HelpToQuit	Year1	GDP	LogGDP
mean	2012.909091	5.872727	63.890909	3.636364	4.181818	2012.909091	2.599154e+04	9.914156
var	1.090909	0.806302	314.892909	0.654545	0.563636	1.090909	2.748318e+08	0.706856

```
from scipy import stats

stats.ttest_ind(df_2012_2014.LogGDP.loc[resid_low_lst], df_2012_2014.LogGDP.loc[resid_high_lst])
```

```
Ttest_indResult(statistic=-2.0334487359693463, pvalue=0.048678755728859664)
```

Comparing the two groups, we test the difference of mean and see p-value < 0.05 on the difference of mean between logGDPs.

Figure 16:

OLS Regression Results

Dep. Variable:	Value	**R-squared:**	0.161
Model:	OLS	**Adj. R-squared:**	0.143
Method:	Least Squares	**F-statistic:**	8.938
Date:	Sun, 21 Nov 2021	**Prob (F-statistic):**	6.21e-09
Time:	17:47:35	**Log-Likelihood:**	-1077.4
No. Observations:	286	**AIC:**	2169.
Df Residuals:	279	**BIC:**	2194.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-12.7245	5.792	-2.197	0.029	-24.127	-1.322
C(HelpToQuit)[T.3]	-2.2508	2.797	-0.805	0.422	-7.758	3.256
C(HelpToQuit)[T.4]	-2.4639	2.988	-0.825	0.410	-8.345	3.417
C(HelpToQuit)[T.5]	-5.1352	3.577	-1.435	0.152	-12.177	1.907
C(EnforceBansTobaccoAd)[T.4]	6.9386	1.508	4.600	0.000	3.969	9.908
C(EnforceBansTobaccoAd)[T.5]	3.4472	2.182	1.580	0.115	-0.848	7.742
GDP	2.5933	0.702	3.696	0.000	1.212	3.975

Omnibus:	52.657	**Durbin-Watson:**	1.966
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	94.695
Skew:	0.997	**Prob(JB):**	2.74e-21
Kurtosis:	4.993	**Cond. No.**	90.0

Now that we have included GDP as a factor to determine the smoker percentages in the female population, we now see a downward trend of smoking values with increased government smoking help. This is the trend we would expect, which indicates that GDP should be included within our model.

Figure 17:

OLS Regression Results

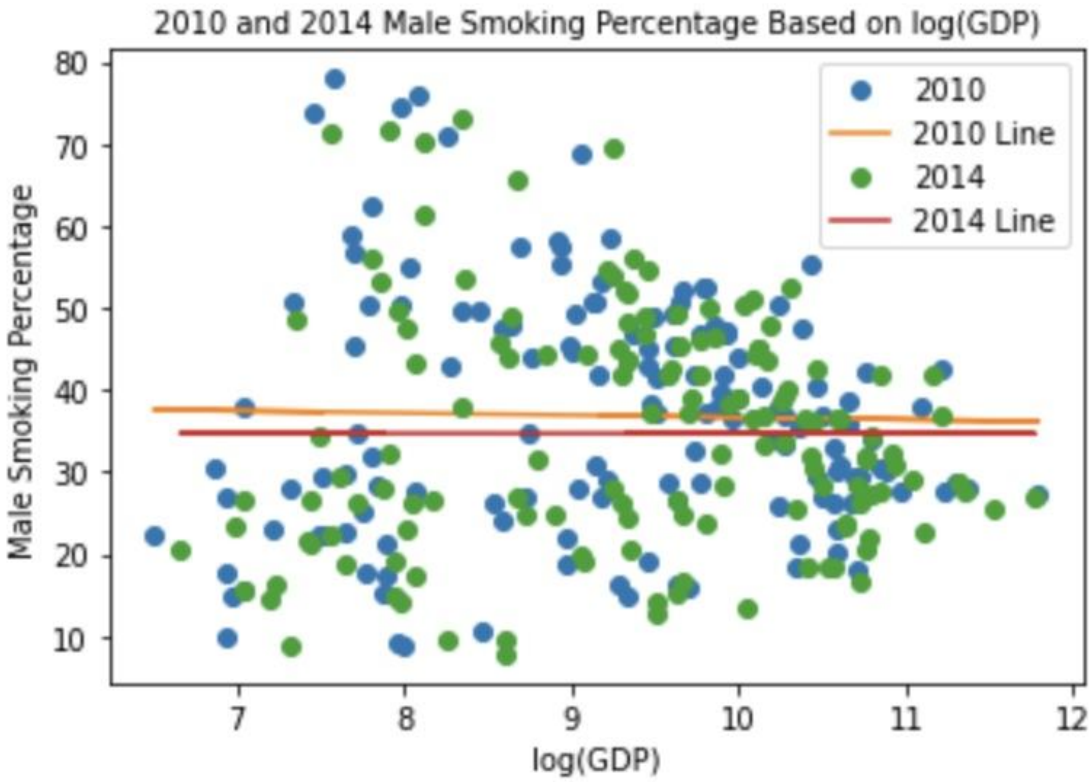
Dep. Variable:	Value	**R-squared:**	0.085
Model:	OLS	**Adj. R-squared:**	0.065
Method:	Least Squares	**F-statistic:**	4.298
Date:	Sun, 21 Nov 2021	**Prob (F-statistic):**	0.000365
Time:	17:47:35	**Log-Likelihood:**	-1159.1
No. Observations:	286	**AIC:**	2332.
Df Residuals:	279	**BIC:**	2358.
Df Model:	6		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	30.1361	7.709	3.909	0.000	14.961	45.311
C(HelpToQuit)[T.3]	1.3097	3.723	0.352	0.725	-6.019	8.638
C(HelpToQuit)[T.4]	-1.1172	3.976	-0.281	0.779	-8.945	6.710
C(HelpToQuit)[T.5]	-7.5748	4.761	-1.591	0.113	-16.947	1.797
C(EnforceBansTobaccoAd)[T.4]	8.1485	2.007	4.059	0.000	4.197	12.100
C(EnforceBansTobaccoAd)[T.5]	5.7091	2.904	1.966	0.050	-0.007	11.425
GDP	0.1408	0.934	0.151	0.880	-1.697	1.979

Omnibus:	15.360	**Durbin-Watson:**	1.679
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	16.267
Skew:	0.555	**Prob(JB):**	0.000294
Kurtosis:	3.367	**Cond. No.**	90.0

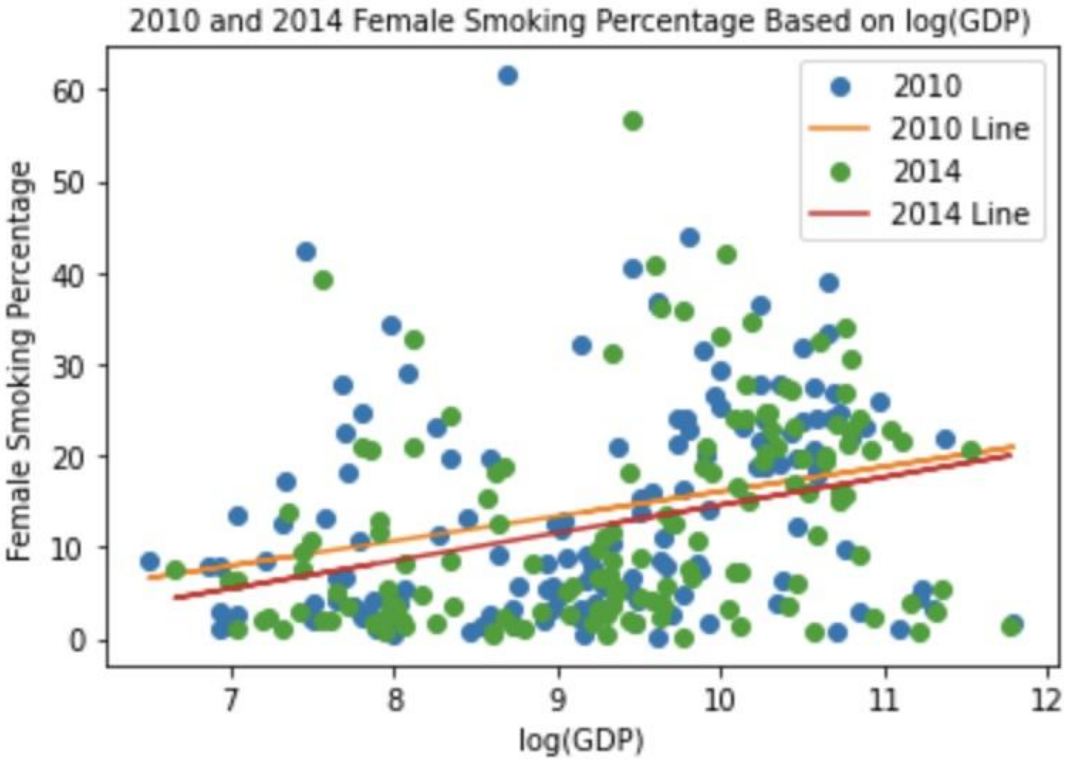
In the male population, the slopes still decrease with each increase of government help, which suggests that included GDP has not greatly impacted the trend we expect for men.

Figure 18:



In men, both 2014 and 2010 show a negative correlation, which is opposite to the correlation we see in our regression for both sexes. One of the most interesting patterns we can see is the reverse fanning of the data points. It appears that countries with higher GDPs will have similar smoking percentages for the male population.

Figure 19:



The female population follows a more aggressive positive trend than the one for both sexes. Additionally, we see two outlier countries points in four dots (since there are two years we are addressing). This would cause female smokers to follow the trend present in both sexes more closely.

External Data Sources Used

1. https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD This file contains the GDP data for each country. The data included in this file are rows of countries (where each row corresponds to a country) where each row contains values of GDP (in units of international currency) for the years 1960-2021. The table has the values of GDP that are available (so not every country has a GDP value for each year).