Deliverable 1

Question to be addressed: Does there seem to be a correlation between legalization of marijuana and drug-related deaths?

Notes:

- The purpose of this question is to gauge an initial understanding of whether the number of drug-related deaths is related to legalization of marijuana.
- If a relationship is indeed present, it will allow us to move forth with our project and investigate potential factors that contribute to the observed relationship.
- To answer this initial question, we chose to look at the state of Colorado. It was the first state to legalize retail sales of marijuana, so it has the most yearly post-legalization data available.
- Original Kaggle Dataset (unparsed) may be found here.

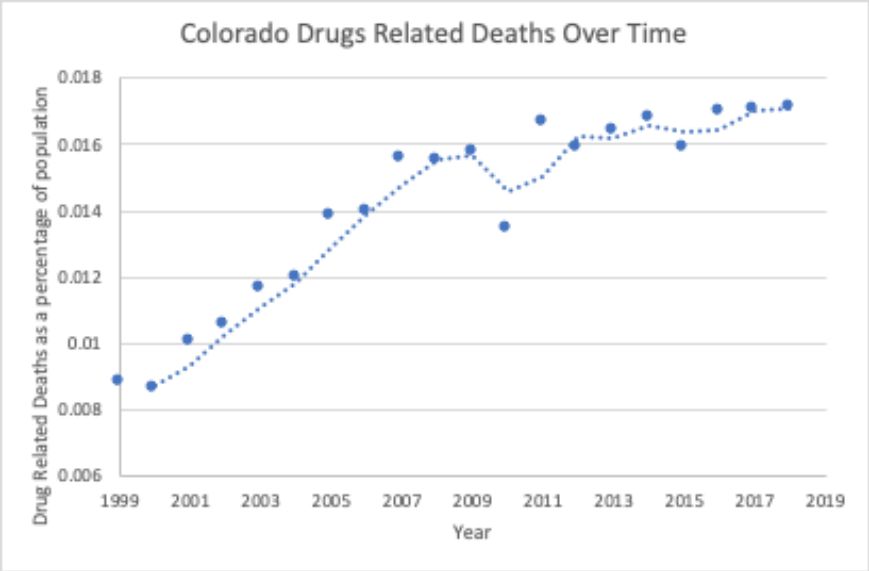
Colorado Data:

Year	Drug Deaths as a proportion of pop.	Percentage Change
1999	0.00887	
2000	0.00865	-2.480270575
2001	0.01003	15.95375723
2002	0.01056	5.284147557
2003	0.01168	10.60606061
2004	0.01198	2.568493151
2005	0.01382	15.35893155
2006	0.01398	1.157742402
2007	0.01555	11.23032904
2008	0.01554	-0.064308682
2009	0.01577	1.48005148
2010	0.01344	-14.77488903
2011	0.01665	23.88392857
2012	0.01592	-4.384384384
2013	0.0164	3.015075377
2014	0.0168	2.43902439
2015	0.01593	-5.178571429
2016	0.01701	6.779661017
2017	0.01707	0.352733686
2018	0.01711	0.234329233

Average Percentage Change from 1999-2012: 5.0963%
Average Percentage Change from 2013-2018: 1.274%

Drug-related deaths are calculated as a percentage of Colorado's population in that year (# deaths/population)

Plotting the Data:



We observe that between the years of 1999-2009, the number of drug related-deaths increased yearly at a steady rate. However, in later years, it appears that the rate of change of drug-death increase seems to slow down. We note that legalization occurred in the year 2012.

Taking a closer look at each period (pre 2012 legalization and post 2012 legalization):

Pre-2012 Legalization:

Year and Drug Deaths as a percentage of population

The REG Procedure
Model: MODEL1
Dependent Variable: drug_deaths

Number of Observations Read	14
Number of Observations Used	14

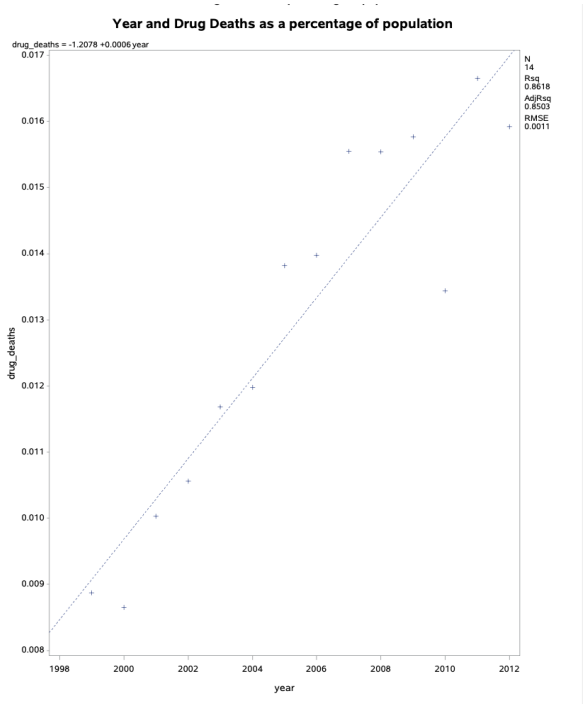
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.00008431	0.00008431	74.83	<.0001
Error	12	0.00001352	0.00000113		
Corrected Total	13	0.00009783			

Root MSE	0.00106	R-Square	0.8618
Dependent Mean	0.01303	Adj R-Sq	0.8503
Coeff Var	8.14520		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-1.20781	0.14113	-8.56	<.0001	-1.51531	-0.90031
year	1	0.00060875	0.00007037	8.65	<.0001	0.00045542	0.00076208



From the above (PRE-LEGALIZATION) data, we can make a few observations:

- The r-squared value is .8618, indicating a relatively strong relationship between year and the number of drug-related deaths.
- The F-statistic is 74.83, which is relatively large, corresponding to a low p-value of <.0001, indicating that there is a significant relationship between year and the number of drug related deaths.
- The slope of the least-squares regression line is .00060875. The model shows that before legalization of marijuana, every subsequent year, the proportion of drug-related deaths was predicted to increase by .00060875 (an average percent increase of 5.063% per year)

Post-2012 Legalization:

Year and Drug Deaths as a percentage of population

The REG Procedure
Model: MODEL1
Dependent Variable: drug_deaths

Number of Observations Read	6
Number of Observations Used	6

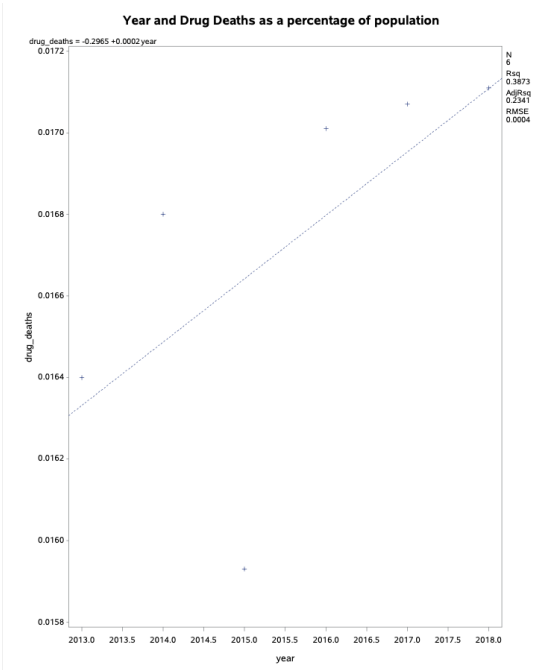
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4.227657E-7	4.227657E-7	2.53	0.1870
Error	4	6.688343E-7	1.672086E-7		
Corrected Total	5	0.00000109			

Root MSE	0.00040891	R-Square	0.3873
Dependent Mean	0.01672	Adj R-Sq	0.2341
Coeff Var	2.44564		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-0.29655	0.19701	-1.51	0.2067	-0.84354	0.25045
year	1	0.00015543	0.00009775	1.59	0.1870	-0.00011596	0.00042682



From the above (POST-LEGALIZATION) data, we can make a few observations:

- The r-squared value is .3873, indicating a relatively weak relationship between year and the number of drug-related deaths.
- The F-statistic is 2.53, which is moderate, corresponding to a p-value of .1870, indicating that there may be a significant linear relationship between year and the number of drug related deaths.
- The slope of the least-squares regression line is .00015543. The model shows that before legalization of marijuana, every subsequent year, the proportion of drug-related deaths was predicted to increase by .00015543 (an average percent increase of 1.274% per year)

Analysis:

Before legalization, the slope of the regression line was .0060875, and after legalization, the slope was .00015543 - showing that the slope decreased by about 74.467%. Furthermore, the

average percentage increase of drug deaths each year before 2012 legalization was 5.063%, compared to a much lower 1.274% after legalization. Although the F-test still appears to show a significant linear relationship between year and drug deaths post legalization, the r-squared value shows that the relationship is much weaker. Our analysis suggests that there does seem to be a relationship between legalization of marijuana and drug related deaths because the rate of change of increase of drug-related deaths decreased significantly in the years following legalization in 2012.

Limitations and Next Steps:

- This is only data for Colorado- we have noted that there is generally a relationship between legalization of marijuana drug-related deaths, and this hypothesis should be tested with data from more states.