Murder: Analysis and Description

Murder: Analysis and Description

Primary response variable: Homicide Rate

This variable demonstrates homicide rates per 100,000 people per state. State is the identifying variable.

Classification variable: High/Low Homicide

This variable is classified as a “High” or a “Low” based on whether the homicide rate is above or below the national average. In some analyses 1 identifies a “High” and 0 identifies a “Low”.

Murder: Analysis and Description

Abortion Ban

This variable shows the state's abortion ban by latest possible week of abortion. Assuming an average length of pregnancy of 40 weeks (9 months), the ban number (22, 26, or 40) depicts the last week in which a woman may get an abortion. If a state's value is 40 weeks, there is no ban.

Prison Rate

This is the number of incarcerated people per 100,000 people per state. Prisons differ from jails in that prisons are longer-term incarceration facilities to which prisoners are moved following their sentencing. In general, while people wait for trials before they're sentenced, they're put in jail. After they're sentenced ("*You are hereby sentenced to 12 months…*"), they go to prison.

Population Density

This value is calculated based on the total population of a state divided by the total land area. Higher values in this variable indicate that the state has a high number of residents who therefore live relatively close together.

Governor Party

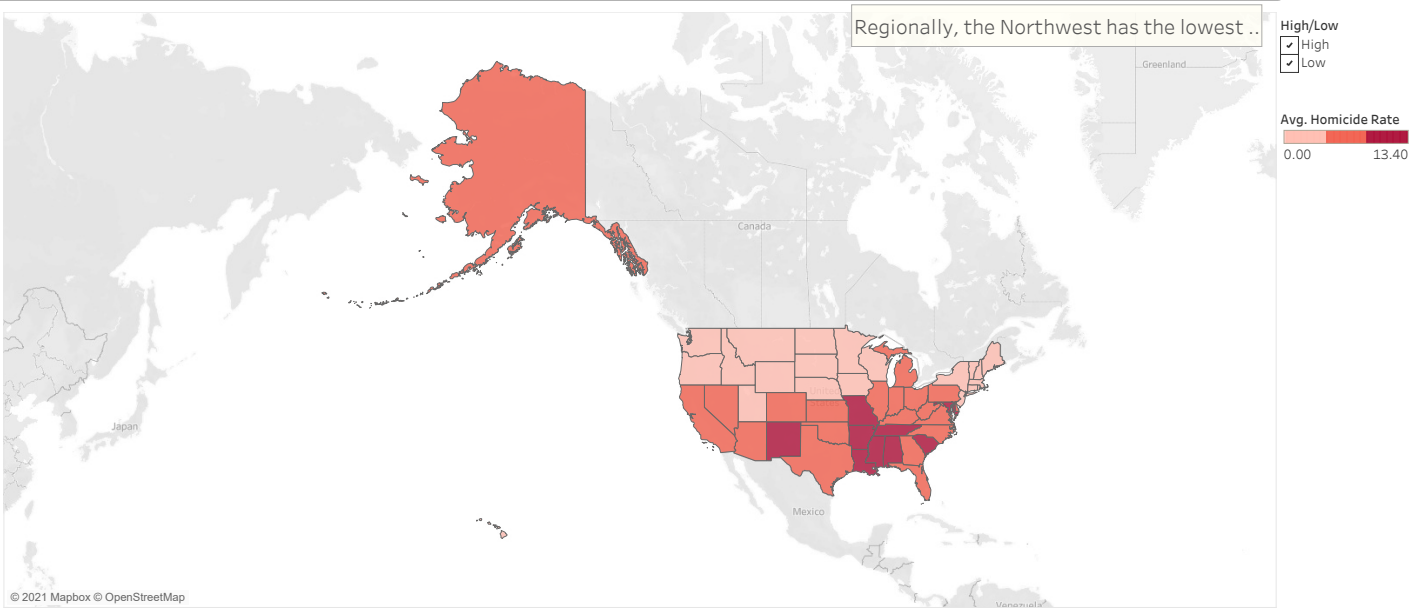
The governor party (Democratic, Independent, or Republican) is determined based on the predominant state governor from 2018. If a gubernatorial election was held during 2018 and a former governor was unseated, the governor that spent the majority of 2018—that is, at least 183 days—in the governor's seat was considered the "governor of 2018." There was only one Independent governor that year, and he was from Alaska.

Minority Population

Minority population is comprised of the groups "Black or African American", "American Indian and Alaska Native", and "Asian". There are only a few states with a minority-majority population (for example, Hawaii has more Asian-identifying than White-identifying people), therefore, this dataset considers the largest nonwhite population in the state.

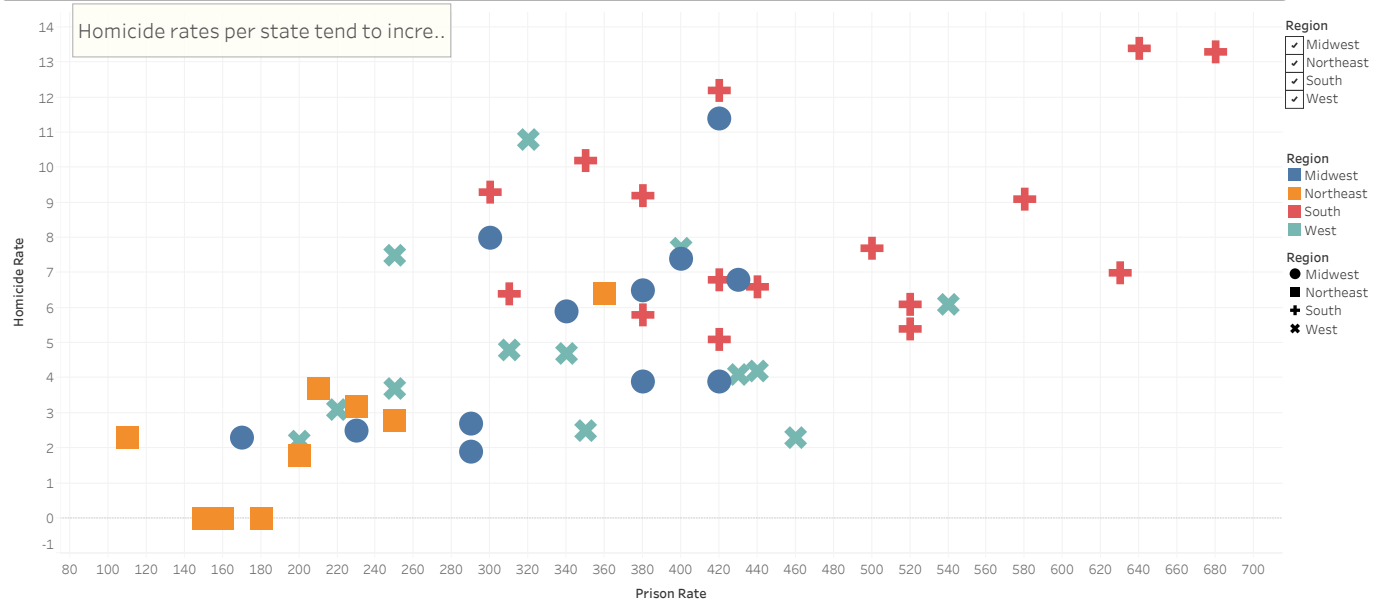
Region

Murder: Analysis and Description

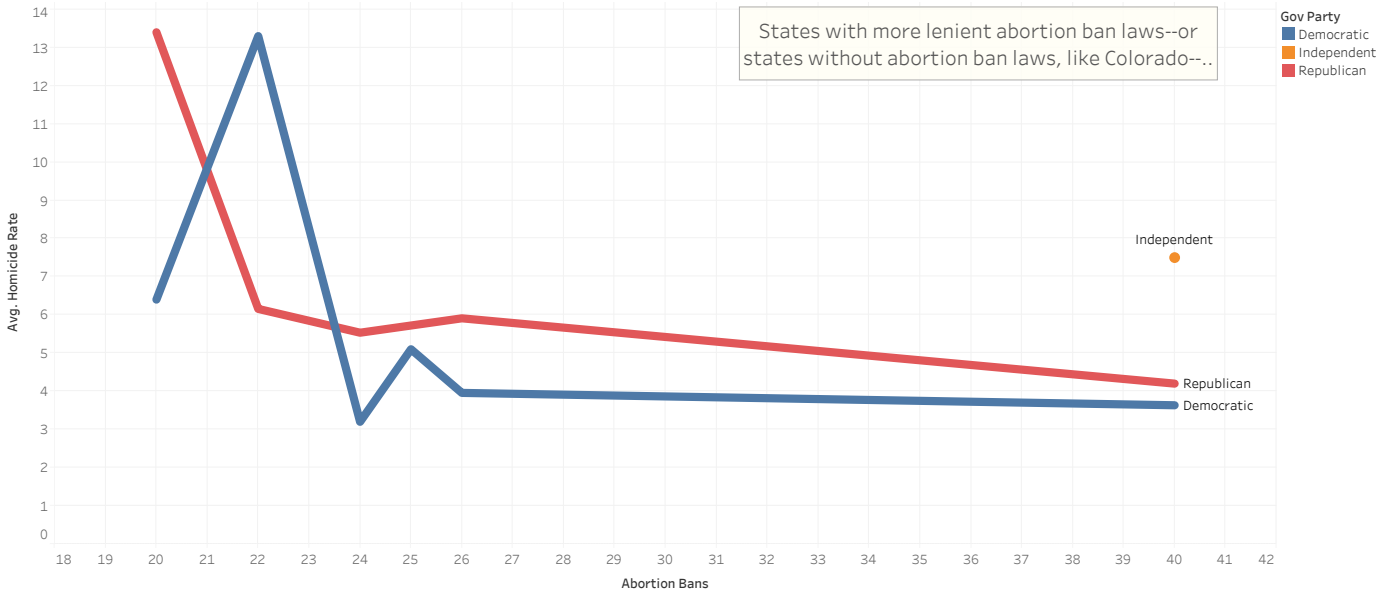


Murder: Analysis and Description

Problem Definition | Dependent Variables | Independent Variables | Visualization: Map | Visualization: Scatterplot | Visualization: Line Graph | Visualization: Point Graph | Continuous Analysis (pt.1) | Continuous Analysis (pt.2) | Categorical Analysis (pt.1) | Categorical Analysis (pt.2)



Murder: Analysis and Description



Murder: Analysis and Description

Murder: Analysis and Description

Continuous

Continuous

Model Comparison

Predictors

Measures of Fit for Homicide_Rate

Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
M4:Predicted Values	Fit Least Squares		0.5888	2.1216	1.7414	50
M5: Predicted Values	Fit Least Squares		0.5337	2.2593	1.9118	50
M6: Predicted Values	Fit Least Squares		0.5519	2.2148	1.9070	50

M4: Least Squares

Summary of Fit

RSquare	0.588753
RSquare Adj	0.469708
Root Mean Square Error	2.433685
Mean of Response	5.614
Observations (or Sum Wgts)	50

M5: Backward Stepwise

Summary of Fit

RSquare	0.533663
RSquare Adj	0.492211
Root Mean Square Error	2.381489
Mean of Response	5.614
Observations (or Sum Wgts)	50

M6: Forward Stepwise

Summary of Fit

RSquare	0.551858
RSquare Adj	0.500933
Root Mean Square Error	2.360947
Mean of Response	5.614

Here we can see the comparison of each of the Continuous models:

the model with the best RSquare value is the Fit of Least Squares (Model 4)
with a value of 58.9% which means the model can accurately predict 58.9% of the data based on the inputted values.
However, based on the adjusted RSquare, the best model is the Forward stepwise sitting at 50.1%.

Further analysis on this aspect is and model is in Continuous Analysis (pt.2)

Murder: Analysis and Description

Continuous pt2

Effect Summary

Source	LogWorth	PValue
Prison_Rate	2.171	0.00675
Region	1.124	0.07511
Minority_Pop	0.546	0.28459
Abortion_Bans	0.499	0.31732
Pop_Density	0.470	0.33850
Gov_Party	0.315	0.48413
Biggest_Cities	0.286	0.51702

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	322.21301	29.2921	4.9456
Error	38	225.06719	5.9228	Prob > F
C. Total	49	547.28020		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.733389	2.93647	-0.25	0.8041
Abortion_Bans	0.0776435	0.076624	1.01	0.3173
Prison_Rate	0.0113482	0.00396	2.87	0.0067*
Minority_Pop[American Indian and Alaska Native]	-0.246324	0.792329	-0.31	0.7576
Minority_Pop[Asian]	-0.873323	0.870849	-1.00	0.3223
Pop_Density	0.0017991	0.001856	0.97	0.3385
Biggest_Cities	-0.155247	0.237363	-0.65	0.5170
Region[Midwest]	0.2138486	0.736425	0.29	0.7731
Region[Northeast]	-2.408487	0.987829	-2.44	0.0195*
Region[South]	1.7345754	0.775518	2.24	0.0313*
Gov_Party[Democratic]	-1.295185	1.066578	-1.21	0.2321
Gov_Party[Independent]	2.0745128	1.913563	1.08	0.2851

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Abortion_Bans	1	1	6.081537	1.0268	0.3173
Prison_Rate	1	1	48.637404	8.2119	0.0067*

After selecting the best model of continuous data (Fit of Least Squares Regression) we can determine what variables are the best at predicting homicide rates.

Here we can see that only one variable passes the p-test of being less than .05 and that is Prison_Rate. This means that Prison_Rate is significant and is a better predictor of homicide rate than other variables. The next two values (variables) are Region[NorthEast] and Region[South] that are also decent but not the best predictors of homicide rate. Together the equation for this regression model is determined through the Parameter Estimates.

Murder: Analysis and Description

Categorical

Model 1: Classification Tree

		M1: Most Likely High/Low	
High/Low	Validation	0	1
0	Training	13	2
	Validation	8	2
1	Training	1	14
	Validation	2	8

Training Error: 3/(13+2+1+14) = .1
Validation Error: 4/20 = .2
Training Sensitivity: 14/15 = .933
Validation Sensitivity: 8/10 = .8

Model 2: kNN

		M2: Most Likely High/Low	
High/Low	Validation	0	1
0	Training	13	2
	Validation	8	2
1	Training	1	14
	Validation	2	8

Training Error: 3/(13+2+1+14) = .1
Validation Error: 4/20 = .2
Training Sensitivity: 14/15 = .933
Validation Sensitivity: 8/10 = .80

Model 3: Naïve Bayes

Training

Actual	Predicted Count	
High/Low	0	1
0	15	0
1	0	15

Validation

Actual	Predicted Count	
High/Low	0	1
0	9	1
1	2	8

Training Error: 0/15 = 0
Validation Error: 3/20 = .15
Training Sensitivity: 15/15 = 1
Validation Sensitivity: 8/10 =.8

Here we can see the analysis of the Categorical models:

Throughout these models, when conducting the error and sensitivity analysis, one model stands out: Naive bayes.
With a training error of 0%, a validation error of 15% & training sensitivity of 100% and validation sensitivitiy at 80%, this model is nearly perfect. The lower the error rate and the higher the sensitivity rate, the better the model. Here, Naive Bayes can correctly predict whether homicide rate will be high or low with 80% certainty.

Murder: Analysis and Description

Training

Count	Misclassification Rate	Misclassifications
30	0	0

Validation

Count	Misclassification Rate	Misclassifications
20	0.15000	3

Confusion Matrix

Training

Actual High/Low	Predicted Count 0	1
0	15	0
1	0	15

Validation

Actual High/Low	Predicted Count 0	1
0	9	1
1	2	8

Fit Details

Measure	Training	Validation	Definition
Entropy RSquare	0.9657	-0.090	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.9838	-0.177	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0238	0.7555	∑ -Log(ρ[j])/n
RASE	0.0651	0.3763	√ ∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0212	0.1847	∑ |y[j]-ρ[j]|/n
Misclassification Rate	0.0000	0.1500	∑ (ρ[j]≠ρMax)/n
N	30	20	n

As stated previously, there were 0 misclassifications within the training data set and only 3 misclassifications within the validation set. This shows the validity of Naive bayes theorem for this data set.

The Sensitivity Rate shows that the model can accurately predict 80% (validation) of the homicide rates whether they are / were high or low. Thi is very valuable to take into account for predicting future homicide rates based off of previous values.

Murder: Analysis and Description

Citations:

Ballotpedia (2018). Partisan Composition of Governors by State. Ballotpedia. Retrieved from https://ballotpedia.org/Partisan_composition_of_governors

Center for Disease Control (2018). Homicide Mortality by State. National Center for Health Statistics . Retrieved from https://www.cdc.gov/nchs/pressroom/sosmap/homicide_mortality/homicide.htm

Guttmacher Institute (2020). An Overview of Abortion Laws. State Laws and Policies . Retrieved from https://www.guttmacher.org/state-policy/explore/overview-abortion-laws

National Geographic (2021). United States Regions. National Geographic . Retrieved from https://www.nationalgeographic.org/maps/united-states-regions/

States101 (2021). U.S. States: Populations, Land Area, and Density. States101 . Retrieved from https://www.states101.com/populations

White, M. (2019). The Top 10 Largest U.S. Cities by Population. Moving.com . Retrieved from https:/..