Visualization and Causal Inference of the Mexican Drug War

Valeria Espinosa and Joseph Kelly

vespinos@fas.harvard.edu, kelly2@fas.harvard.edu

Statistics Department, Harvard University

Problem Description

The presidency of Felipe Calderón (2006-2012) has been characterized for the war against organized crime, raising many questions regarding security and violence. We attempt to visualize and analyze homicide rates at the municipality level, to answer whether **homicide rates increase significantly after a military intervention**. Visually we link this to information obtained about the association of drug cartels to municipalities.

Estimand

Let $Y_i(1)$ denote the homicide rate change in region i from one year before to one year after receiving a military intervention, and $Y_i(0)$ what it would have been if it hadn't received it (Rubin Causal Model). Our estimand is the average causal effect of the military intervention, for the regions that were intervened,

$$\tau = \overline{Y}(1) - \overline{Y}(0) = \frac{\sum_{i=1}^{I} Y_i(1) - Y_i(0)}{I}.$$

Let N_i denote the number of municipalities that correspond to region i, then

$$Y_i(1) = \sum_{j=1}^{N_i} w_{ij} Y_{ij}(1) \text{ and } Y_i(0) = \sum_{j=1}^{N_i} w_{ij} Y_{ij}(0),$$

$$\text{where } w_{ij} = \frac{\text{Pop}_{ij}}{\text{Pop}_i} \text{ and } \text{Pop}_i = \sum_{j}^{N_i} \text{Pop}_{ij}.$$

However, $Y_i(0)$ and $Y_{ij}(0)$ are missing $\forall i, j$.

Key Assumptions

- **SUTVA:No hidden values of treatments** Broad definition of treatment: at least one municipality in the region received an intervention between 2007-2010, or not ([2]).
- **SUTVA:No interference between units** Grouped close regions that received an intervention, and their neighboring municipalities to make the "no interference" assumption more reasonable.
- **Unconfoundedness** We assume we have all covariates, $\mathbf{X}$, such that given $\mathbf{X}$, treatment assignment is independent of $\mathbf{Y}$.
- **Missing Data** One covariate had missing values. We exactly matched on missingness pattern and Political Party in municipality before Calderón.
- **Appropriateness of response variable** We assume Y is an adequate measure of violence.

Estimation

The control pool consists of 2213 municipalities. There are 13 treated regions considered (205 municipalities). Propensity score matching was used to identify 5 control municipalities that look like each treated ones, and estimate $Y_{ij}(0)$ and $Y_i(0)$. Let M_{ij} be the number of municipalities matched to the jth municipality in region i, and $\text{PopM}_{ij} = \sum_{k=1}^{M_{ij}} \text{PopM}_{ijk}$ is the sum of their populations. Then,

$$\hat{Y}_{ij}(0) = \sum_{k=1}^{M_{ij}} v_{ijk} Y_{ijk}(0), \qquad \text{where } v_{ijk} = \frac{\text{PopM}_{ijk}}{\text{PopM}_{ij}}.$$

Therefore,

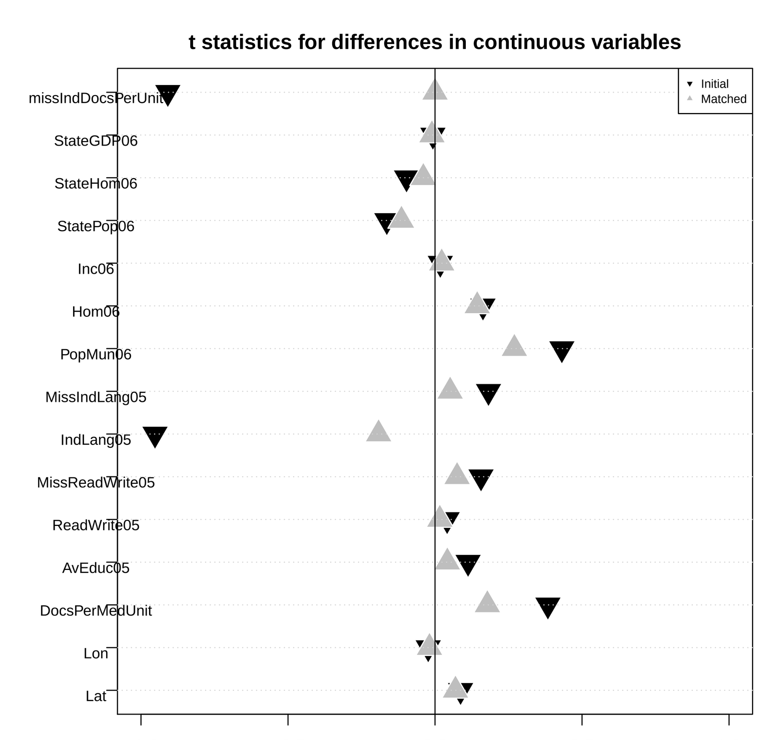
$$\hat{Y}_i(0) = \sum_{j=1}^{N_i} w_{ij} \hat{Y}_{ij}(0) = \sum_{j=1}^{N_i} w_{ij} \sum_{k=1}^{M_{ij}} v_{ijk} Y_{ijk}(0) = \sum_{j=1}^{N_i} \sum_{k=1}^{M_{ij}} \tilde{w}_{ijk} Y_{ijk}(0) \qquad \text{with } \tilde{w}_{ijk} = w_{ij} v_{ijk},$$

and

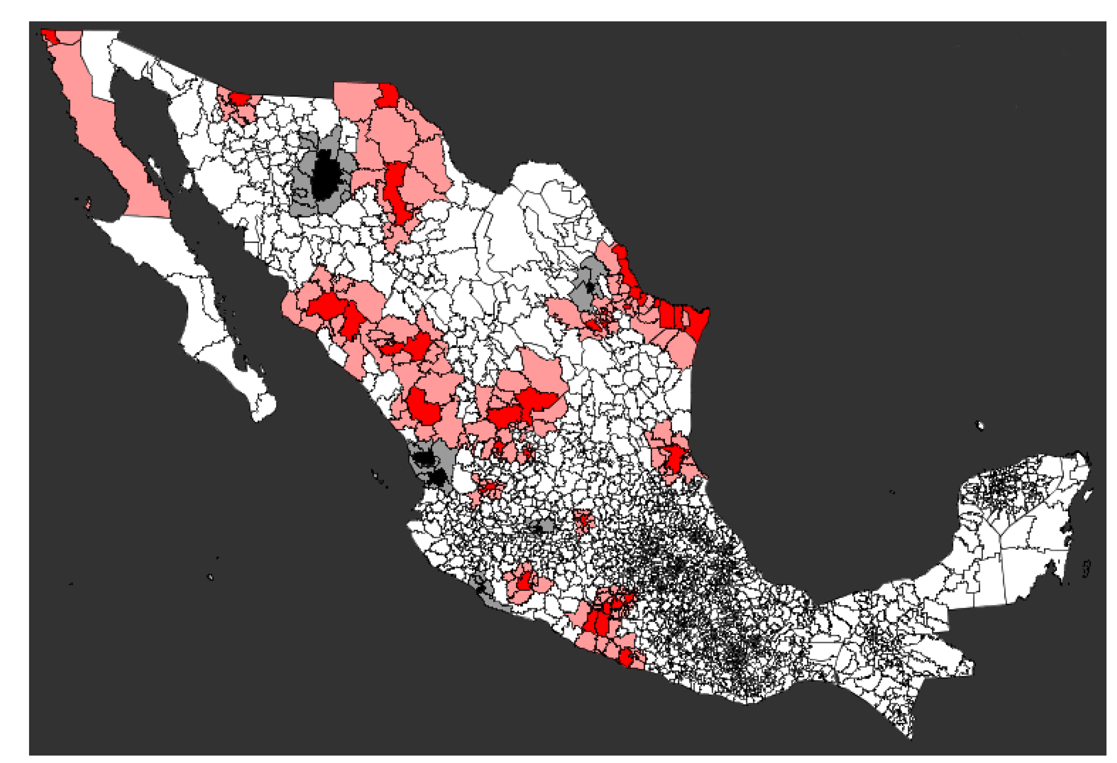
$$\hat{\tau} = \frac{\sum_j Y_j(1)}{J} - \frac{\sum_{j=1}^{J} \hat{Y}_j(0)}{J} = \overline{Y}(1) - \overline{Y}(0).$$

We know that $var(\hat{\tau}) \leq var(\overline{Y}(1)) + var(\overline{Y}(0))$, and it achieves the bound under additivity of potential outcomes. We use that over estimate and assume the homicide rates have a Poisson distribution to get confidence intervals.

Visualization

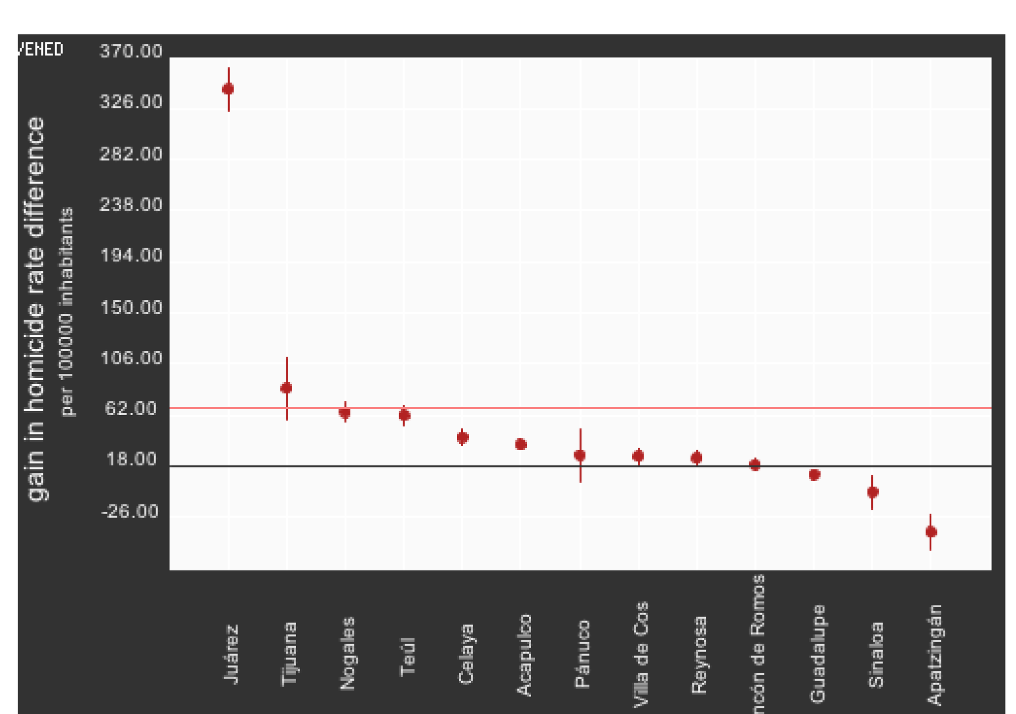


Love plot - balance checks



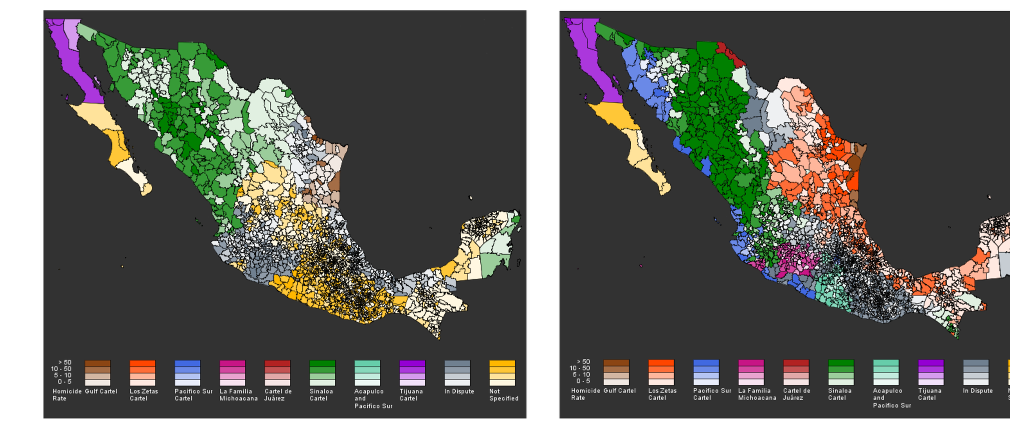
Interventions and SUTVA

Results



Region	number of municipalities	Date of first intervention	Regional causal effect ($\hat{\tau}_j$)	SD
Juárez	15	2009	147.54	4.33
Tijuana	5	2008	30.60	6.15
Nogales	5	2008	21.18	1.84
Teúl	36	2008	19.89	1.85
Celaya	29	2007	11.31	1.37
Acapulco	20	2009	8.49	0.89
Pánuco	10	2009	4.15	5.10
Villa de Cos	22	2008	3.94	1.53
Reynosa	9	2009	3.34	1.36
Rincón de Romos	24	2008	0.47	1.14
Guadalupe	14	2007	-3.32	1.02
Sinaloa	7	2008	-10.14	3.32
Apatzingán	9	2007	-25.78	3.38
$\hat{\tau}$	250	-	22.42	0.85

Visualization and Processing



(a) 2007 (b) 2010

Processing is an open source programming language that allows the creation of dynamic graphics and tables. Due to the size of the data we collected, this tool played a significant role in the display and understanding of the results. The use of dynamic graphics allows:

- The navigation through over 2400 municipalities, and view the homicide rate all within the same window via a map of Mexico.
- The presentation of further information, such as cartel presence and homicide rate over time.
- The display of matched municipalities and the corresponding causal effects for each region.

Conclusions

On average, the military interventions result in an increase of the homicide rates. However, the effect varies across the treated regions. The Juárez region is clearly an outlier with an increase of homicide rate of 147 per 100000 inhabitants. Visualizations like this one can help experts to better evaluate the quality of the matches and improve their understanding of the estimates of the causal effects. For example, in this case it is clear that the particularities of the Juárez region play a very important role in the overall average effect.

Key References

[1] Abadie, Diamond, Hainmueller *Synthetic Control Methods for Comparative Case Studies* (2009)

[2] Escalante F, *Homicidios 2008-2009 La muerte tiene permiso* (2011)

[3] Imbens G. & Rubin D.R., (2012)

[4] Rubin D.R. ,*Matched Sampling for Causal Effects*(),

Data Sources

INEGI, CIDAC, Stratfor, Nexos