Short-term test chapter

This chapter explores the relationship between state capacity and disaster-related deaths in the immediate aftermath of a disaster. This chapter builds off of the work of @lin_governing_2015, @tol2021state, and @kahn2005death, exploring the relationship between state capacity and disaster-related deaths but with a more comprehensive measure of state capacity, and finds, consistent with the existing literature that greater state capacity is associated with fewer deaths.

Additionally, this chapter further expands the existing literature by separately exploring the three main components of state capacity: administrative capacity, coercive capacity, and extractive capacity [@hendrix2010, @hanson2013leviathan,@mao2021, @hanson2018state]. As mentioned in chapter 1, administrative capacity is a state's ability to create and implement public policies - the quality of a state's bureaucracy and legislature. Coercive capacity pertains to the ability of a state to control its citizens whether through force, the rule of law, or good will. Finally, extractive capacity is a state's ability to raise revenue from its territories whether it be through taxation or state-run business enterprises. These three components of state capacity are all critical for the effective function of a state and complement one another. For instance, it is a lot easier for a state to have extensive administrative capacity when it has ample funds which are made possible through a state's coercive capacity. That being said, despite their complementary nature, the three types of state capacity are quite different and it is important to explore them as separate entities in addition to observing them as a whole.

In the following pages, I will explain the data sources that I use to test my hypotheses, describe in-depth the methods that I use for my analysis and the justifications for using them, and finally, present the results of my analyses.

Explanation of Data Sources

To test my hypothesis that greater state capacity leads to fewer disaster-related deaths and displaced people immediately after the disaster, I use the Emergency Events Database (EM-DAT). Created by the Centre for Research on the Epidemiology of Disasters (CRED), the EM-DAT disaster dataset provides information on over 22,000 disasters since the year 1900 compiled from a number of sources including the United Nations, national governments – both of the affected countries and countries providing aid – NGOs, and insurance companies, etc. For the analysis in this chapter, I use EM-DAT data to identify disasters, and to measure the results of the disasters: the number of people killed and number of displaced persons. EM-DAT also provides important control data including location data on both the country and admin 1 (province/state) level, data on the specific type of disaster, the scale of the disaster (based off of disaster-specific metrics), the specific disaster name if one exists, and whether or not the disaster was declared by the government. @lin_governing_2015, @tol2021state, and @kahn2005death, all used the EMDAT data in their analyses, meaning both that this is a generally well-regarded measure of disaster effects and that my results can be directly compared to those presented in the three papers. Importantly, EMDAT only records events that result in significant harm, meaning that this data does not include events that do not result in destruction - that being said

To measure my independent variable, state capacity, I use a metric created by @hanson2013leviathan. Their state capacity variable uses 21 different indicators[1] and Bayesian latent variable analysis to create a comprehensive measure of state capacity. The measure exists on a scale of -3 to 3 and is not finite, meaning that the values can be multiple decimal places long and quite exact. This dataset provides capacity information for 177 countries between the years of 1960 and 2015. While built from 21 different variables, this metric relies primarily on the three key components of state capacity mentioned above: administrative capacity, coercive capacity, and extraction capacity. Additionally, the @hanson2013leviathan dataset includes specific measures for each of these three components of state capacity and a number of useful control variables including GDP per capita, and anocracy or the level of democracy in a country. To measure administrative capacity, I use the "Bureaucratic Quality" metric which rates how strong a country's institutions are on a scale of 0-4 with 4 indicating bureaucracies that are able to govern consistently without policy disruptions and that experience little political intervention [Howell (2011)]. To measure coercive capacity, I use the the Military personnel percapita variable from @hanson2013leviathan which indicates the number of military-affiliated people per 1000 residents. Lastly, to measure extractive capacity, also from @DVN/NRR7MB/KQQRFO_2011, I use the Absolute Political Extraction (APE) variable. This variable measures tax, mining, and export revenue relative to GDP and is displayed as a rate. The greater the value, the greater the extractive capacity of the nation.

[1] For a detailed description of the 21 indicators please see the supplementary materials included by @hanson2013leviathan. These indicators are sourced from a number of different academic publications, NGO indicator lists, and national reports.

Methods

To explore the relationship between state capacity and disaster related deaths in the immediate aftermath of a disaster, I analyze the data in three ways: 1) testing the relationships found in @lin_governing_2015 and @tol2021state using different modeling techniques, 2) using the new state capacity data to explain the variation in deaths, and 3) breaking down state capacity into its three components and observing their relationships with total deaths separately.

Expansion and Test of the Current Literature

Because the arguments and subsequent analyses presented in this thesis are based off of those of @lin_governing_2015, @tol2021state, and @kahn2005death, an important preliminary check requires that I test the relationships presented in those papers using my models of choice. Both @lin_governing_2015 and @kahn2005death test the relationship between GDP per capita and log deaths while @tol2021state explores the relationship between taxation as a percentage of GDP and the log number of disaster-related deaths. To test these relationships, I use two simple linear models with no controls - one observes the relationship between log GDP per capita and log deaths while the other observes the relationship between taxation as a percentage of GDP and log deaths. Importantly, unlike @lin_governing_2015, I do not use log deaths per 100,000 people as a dependent variable. While an attempt to contextualize the number of deaths and the impact of the disaster within a nation, due to the highly-localized nature of natural disasters, I believe that it is unhelpful to observe deaths per capita. For example, it would be unhelpful if looking at deaths associated with a hurricane in Florida, USA to then look at those deaths as they relate to the US population as a whole. Because the hurricane only struck Florida, it doesn't make sense to contextualize the deaths as they relate to the US population as a while.

State Capacity Exploration Using @hanson2013leviathan as a measure of State Capacity

I explore the relationship between state capacity and deaths following a disaster using three critical outcome variables - log deaths, death rate, and total deaths. To explore these relationships, I run three separate regressions, one for each outcome variable. The notation used for the three models are explained in the following estimation equations:

$$ln(Total\ Deaths) = \beta_0 + \beta_1 * Capacity + \beta_2 * Disaster\ Type + \beta_3 * Magnitude + \beta_4 * ln(GDP\ percapita)$$

$$Deadliness = \frac{Total\ Deaths}{Total\ Affected} = \beta_0 + \beta_1 * Capacity + \beta_2 * Disaster\ Type + \beta_3 * Magnitude + \beta_4 * GDP\ percapita$$

$$Total\ Deaths = \beta_0 + \beta_1 * Capacity + \beta_2 * Disaster\ Type + \beta_3 * Magnitude + \beta_4 * GDP\ percapita$$

Due to the differing outcome variables, I need to use different model types for each. For the log deaths variable I use a log linear model to measure the relationship. I choose to observe deaths in their log form because there is extremely high variance in the number of deaths and through standardizing them, I am able to play close attention to the relationship. For the deaths per number of people affected model, I first create the death ratio variable by dividing the total number of deaths reported in the EMDAT data by the total number of people affected by the disaster as reported by EMDAT. While I could use the number of people affected as a control variable in the previous model, I believe that observing the death rate separately is important as shows fatality within the context of the overall scope of the disaster. While previous work has attempted to contextualize deaths as they relate to the national population, I think that it is more fruitful to observe the deaths as they relate to the areas affected by the disaster. To observe this relationship I use a simple linear model, controlling for all of the variables listed earlier.

Finally, although I believe it is important to observe the log relationship between state capacity and deaths, I also recognize that disaster-related fatalities are real people losing their lives and that this relationship should also be observed in its count form. To observe the relationship between state capacity and total deaths as a count variable I use a negative binomial model.[2] This model will report the estimated relationship between a nation's level of state capacity and the total number of people expected to perish in a

[2] While both poisson and negative binomial models are well-suited for analyzing count data — in fact a poisson model is a version of a negative binomial model in which it is assumed that the mean and variance are about equal — due to the high variance of the total death data, a poisson model would suffer from overdispersion. Overdispersion, when the variance is greater than the mean can result in inflated standard errors and thus artificially increases the significance of results. In the case of the total death data, the variance is nearly 500,000 times greater than the mean (Mean = 509.5, Variance = 251,120,213), meaning that a poisson model would most definitely suffer from overdispersion. To account for this, I use a negative binomial model. Figure 1 in the appendix shows a histogram of the log death data. It was impossible to effectively present the data in its pure count form due to the extremely high variance but the varition is still able to be seen when looking at the data in its log form.

disaster.

For all three models I control for the type of disaster, magnitude of the disaster, and the log GDP per capita of the nation in which the disaster happened. While the EMDAT dataset does include more fine-grain disaster descriptors — referred to as disaster subtypes and “subsubtypes” — these indicators are not present for all disasters and including them in the model results in significant data loss. Because there are so many different types of natural disasters included in the data, I select only those event types that resulted in greater than 1000 deaths listed in the data — floods, storms, earthquakes, extreme temperature events, wildfires, droughts, and epidemics. Importantly, I choose to observe deaths and GDP per capita in their log forms, because these values encompass a wide range and I am interested in observing whether a linear relationship exists.

The control variables included are an attempt to isolate the relationship between disaster outcomes and state capacity. Notably, log gdp per capita is not included in the 21 indicators that make the @hanson2013leviathan measure of state capacity, meaning that it can serve as a control variable without the threat of double counting. Most importantly for this analysis, I control for the type of disaster. Destruction from different types of disasters varies so it is important to control for these differences For example, extreme temperature events are extremely different from earthquakes and thus should be observed as separate entities. There is an actual disaster magnitude variable included in the model. The EM-DAT data provides a disaster magnitude value and a disaster magnitude scale for each event. Important to note, these values are all relative to their respective scales and can’t be compared between disasters — the Richter scale for earthquakes; square kilometers for droughts, floods, and wildfires; kilometers per hour for wind speeds for storms; degrees Celsius for extreme temperatures; and the number of people with relevant vaccines for epidemics. To enable whole-sample comparison of disaster magnitude, I standardize the disaster magnitude value for each disaster, the results are disaster magnitudes on a scale of -1 to 1 that can be compared across disaster types. To standardize my data I use the following equation:

$$Standardized\ Magnitude = \frac{Magnitude\ Value * Mean\ of\ Magnitudes}{Standard\ Deviation\ of\ Magnitudes}$$

Finally, because state capacity and GDP are sometimes used interchangeably (@lin_governing_2015 & @kahn2005death), I control for GDP percapita to isolate the effects of state capacity and to ensure that what I'm picking up isn't actually just the strength of the nation's economy.

Different types of state capacity

As the third and final test of the relationship between state capacity and short-term disaster deaths, I observe the relationships between disaster-related deaths and the three different components of state capacity: administrative capacity, coercive capacity, and extractive capacity. While I noted earlier that I believe that previous literature on this topic took an overly simplistic look at state capacity, reducing it to just one of these components or or even a sub-component, due to the divergent nature of these components I believe that it is important to observe them separately but also within the context of eachother. To observe this relationship, I use the same three dependent variables explained above – log deaths, the death ratio, and total deaths – but change the independent variables slightly, replacing the singular measure of state capacity with the measures for the three components. As with the previous section, I use linear models to explore the relationship between capacity and log deaths and the death ratio while using a negative binomial model to measure the total death count variable. Again, it is important to include this count data as it shows in real practical terms the relationship between state capacity and lives lost.

Results

Robustness Check for the Current Literature

Table 1 shows the output of the two log-linear models, one exploring the relationship between log GDP per capita and log disaster related deaths, and the other tax revenue as a percentage of GDP and log disaster related deaths. Model 1 finds that consistent with the findings of @lin_governing_2015 and @kahn2005death, there is a significant and negative relationship between log GDP per capita and log disaster related deaths, meaning that as GDP per capita increases it is associated with a decrease in disaster related deaths, in short, when struck by a disaster, richer countries will experience fewer disaster-related deaths than poorer countries. A one unit increase in log GDP per capita is associated with just under a 0.5 decrease in log deaths.

Similarly, the findings from Model 2 are consistent with those of @tol2021state, finding that the larger proportion tax revenue is of GDP, the fewer deaths associated with

Table 1:

	Dependent variable:	
	ln(Total Deaths)	
	(1)	(2)
ln(GDP per capita)	-0.454^{***}	
	(0.014)	
Tax Revenue		-7.609^{***}
		(0.340)
Constant	6.420^{***}	4.161^{***}
	(0.103)	(0.054)
Observations	6,362	6,176
R^2	0.147	0.075
Adjusted R^2	0.146	0.075
Residual Std. Error	1.680 (df = 6360)	1.742 (df = 6174)
F Statistic	$1{,}092.805^{***}$ (df = 1; 6360)	500.928^{***} (df = 1; 6174)
Note:		*p<0.1; **p<0.05; ***p<0.01

natural disasters. These results are also significant and show that a 1 unit increase in tax revenue as a proportion of GDP is associated with a 7 unit decrease in log disaster deaths. While the regression coefficient is larger in Model 2 than Model 1, the R-squared values for each of the models indicate that Model 1 which uses log GDP per capita as the independent variable is better able to explain the variation in the death data, indicating that log GDP per capita is a better measure than tax revenue as a percentage of GDP. Before testing the more comprehensive measure of state capacity, it was important to ensure that the literature I am basing my hypotheses off of is indeed correct.

Observing State Capacity as a Comprehensive Variable

Table 2 exclusively uses the @hanson2013leviathan measure of state capacity as the primary independent variable of interest. As explained in the Methods section earlier, Table 2 observes three different dependent variables: log disaster deaths, death rate, and the total number of deaths. Model 1 uses log deaths as the dependent variable and finds a significant negative relationship between state capacity and log deaths. Interestingly, GDP per capita is used as a control variable in this Model but while greater GDP per capita is associated with fewer log disaster deaths, the effect is much smaller here than seen in Model 1 of Table 2 above. This is likely because this model includes a state capacity measure in addition to simply

Table 2:

	Dependent variable:		
	ln(Total Deaths)	death_rate	total_deaths
	normal	*normal*	*negative binomial*
	(1)	(2)	(3)
State Capacity	-0.411^{***}	-0.038^{**}	-1.219^{***}
	(0.062)	(0.019)	(0.061)
Disaster Magnitude	0.370^{***}	-0.006	1.601^{***}
	(0.041)	(0.012)	(0.040)
GDP per capita	-0.00001^{***}	0.00000	0.00001^{***}
	(0.00000)	(0.00000)	(0.00000)
Constant	4.733^{***}	0.001	6.602^{***}
	(0.930)	(0.270)	(0.908)
Observations	2,225	1,928	2,225
Log Likelihood	−4,532.499	−1,540.714	−12,089.700
θ			0.300^{***} (0.007)
Akaike Inf. Crit.	9,086.998	3,103.429	24,201.410
Note:			*p<0.1; **p<0.05; ***p<0.01

GDP per capita and suggests that it was the relationship between GDP and state capacity that was indeed driving the results seen in Table 1.

Model 2 in Table 2 explores the fatality rate of the disaster. The "death rate" is the number of people who die as a result of the disaster divided by the number of people affected by the disaster. A higher rate means that there are a greater proportion of people dying. Model 2 finds that while not as statistically significant as the findings mentioned up to this point, greater state capacity is associated with a lower death rate. For every one unit increase in state capacity it is expected that the death rate will decrease by about .038 or nearly 4%. This is an important metric to observe because it explains the number of deaths within the context of the destruction surrounding them and allows me to directly observe how deadly a disaster is. Unile other measures that try to contextualize disaster related deaths as a part of a local or national population, by looking at deaths relative to affected people, I am able to observe how deadly a disaster is in the context of those affected by the disaster. Important to note, unlike the other models, Model 2 does not find a statistically significant relationship between the magnitude of the disaster and the outcome variable and finds no relationship (but not a statistically significant 0) between GDP per capita and the outcome variable.

Lastly, Model 3 in Table 2 serves to measure the real human relationship between state capacity and disaster related deaths. Like in Model 1, it finds a statistically significant negative relationship between state capacity and disaster related deaths. For every one unit increase in state capacity, this model predicts a 1.2 person reduction in disaster related deaths. While this relationship is statistically significant, it may not be practically significant, revealing an often overlooked shortcoming of quantitative social science methods. The @hanson2013leviathan state capacity variable ranges in value from -3 to 3 meaning that it encompasses 7 different integer values. As explained earlier, a one unit increase in state capacity is associated with a 1.2 person reduction in deaths, meaning that this model predicts that the states with the greatest state capacity can be expected to experience 8.4 fewer deaths than those with the least possible state capacity. While every life is precious, 8.4 lives is an extremely small number, suggesting that while the results of these regression models indicate statistical significance, that this relationship should be observed more closely to see whether there is a practical significance. In addition to providing an example of the need to observe regression results closely, these findings also highlight the shortcomings of solely observing outcome variables in their log form as the abstract nature of the results can convey results in a way that makes them seem more important and relevant than they are.

Observing State Capacity Components Seperately

As mentioned at the beginning of this chapter and in the Methods section, state capacity, while a singular concept, consists of three seperate components: administrative capacity, coercive capacity, and extractive capacity. Table 3 explores the relationships between these three measures of state capacity and the three outcome variables observed above. While the different outcome variables are observed in each model, I choose to break down my analysis by state capacity component rather than outcome variable.

As seen in Table 3, all three models show a statistically significant negative relationship between administrative capacity and deaths. Interestingly, Model 2 finds that a one unit increase in administrative capacity is associated with a 0.038 decrease in the death ratio, same value associated with a 1 unit increase in the general state capacity measure. This finding may suggest that administrative capacity is the largest driver of the relationship

Table 3:

	Dependent variable:		
	log_death	death_rate	total_deaths
	normal	*normal*	*negative binomial*
	(1)	(2)	(3)
Administrative Capacity	-0.320^{***}	-0.038^{***}	-0.596^{***}
	(0.056)	(0.012)	(0.054)
Coercive Capacity	-0.503^{**}	-0.077^{*}	-0.360
	(0.232)	(0.046)	(0.221)
Extractive Capacity	-1.182^{**}	0.131	-1.206^{**}
	(0.510)	(0.104)	(0.485)
Disaster Magnitude	0.320^{***}	-0.005	1.714^{***}
	(0.059)	(0.012)	(0.056)
GDP per capita	-0.00002^{**}	0.00000	-0.00001
	(0.00001)	(0.00000)	(0.00001)
Constant	4.646^{***}	-0.016	6.059^{***}
	(1.137)	(0.217)	(1.080)
Observations	1,327	1,145	1,327
Log Likelihood	$-2,716.947$	-438.514	$-7,294.677$
θ			0.313^{***} (0.010)
Akaike Inf. Crit.	5,459.894	903.028	14,615.350
Note:			*p<0.1; **p<0.05; ***p<0.01

between increased state capacity and a decreased death ratio. Model 3 finds that a one unit increase in administrative capacity is associated with a 0.6 person reduction in deaths. Again, this real-life finding suggests that while the results of this thesis may be statistically significant that they lack real-life significance.

Unlike administrative capacity where all three models found a significant and negative relationship between capacity and death, across all three models, results for coercive capacity are not all statistically significant. Models 1 and 2 find that a one unit increase in coercive capacity is associated with a 0.5 reduction in log deaths and a 7% reduction in the death rate. While these death reductions are greater than that observed for administrative capacity, due to the smaller scale of the coercive capacity measure[3] and the lesser significance of the results, these findings are less convincing. This reduced significance is supported by Model 3 which does not find a statistically significant relationship between coercive capacity and the number of people killed by a disaster. The relative insignificance of coercive capacity on deaths in the immediate aftermath of a disaster is quite intuitive. While coercive capacity is important for mobilizing a population, if a disaster strikes without warning, there is no opportunity for a government to intervene and prevent greater destruction in the immediate term. Conversely, for natural hazard events that are more long-lasting like droughts or famines, coercive capacity could play a larger role, enabling governments to compel people to travel to different places and cooperate with different relief efforts. I expect the effects of coercive capacity to be more present in the following chapters where longer-term outcomes are observed, not just short-term deaths.

Extractive capacity poses the most interesting puzzle out of the three measures of state capacity with both Models 1 and 3 suggesting larger and somewhat statistically significant relationships between extractive capacity and both log and total deaths. Interestingly, while not significant, unlike the previous capacity measures in which the sign of the relationship was consistent across models, with extractive capacity, Model 2 finds an insignificant positive relationship between extractive capacity and the death rate. While the extractive capacity variable used was created using a number of different measures of extraction as they relate to a country's GDP including taxes, mining, and export revenue, it may be that coutries suffering from resource curses [CITE] have greater extractive capacity but overall poor living conditions and limited income re-distribution. This resource curse

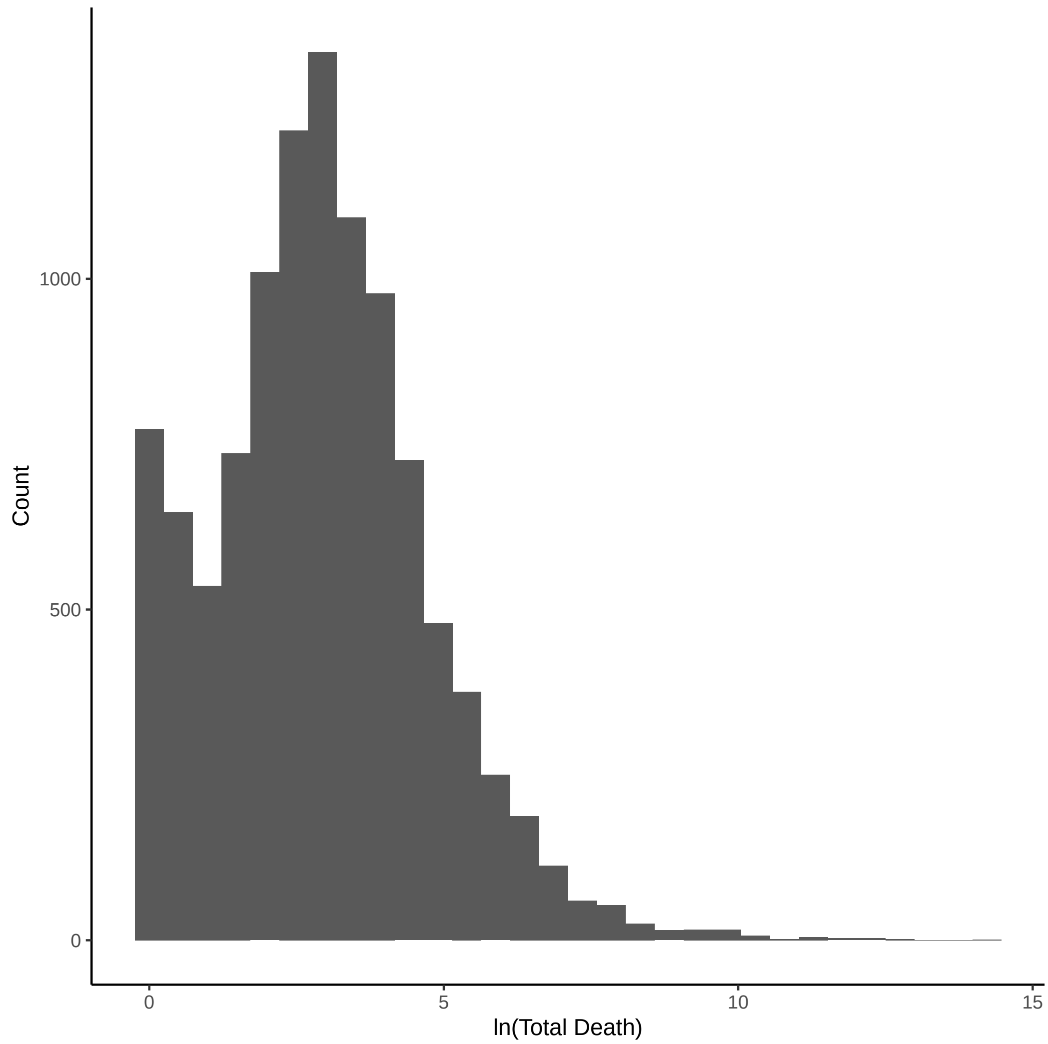
[3]Measures of coercive capacity range in value from approximately 0 to 2 while the administrative capacity measure ranged from about -4 to 5

in which countries rich in natural resources are able to fund their governments without the financial and political buy in from the people typicallty results in poor governance and social services. This poor governance and weak baseline could mean that if resource cursed states have some of the highest extractive capacity that these poor institutions actually make people more vulnerable to disasters.

Discussion

The findings presented in Chapter 1 present an interesting story about the role of state capacity and disaster related deaths. While increases in state capacity are significantly associated with reductions in log deaths, the ratio of deaths to affected peopl, and total deaths, these results may not be practically significant. Every life matters, but when models predict that increasing state capacity from it's lowest to it's highest level will only save about 8 peoples lives in the event that a disaster strikes, this relationship may not necessarily be one to focus future policy initiatives on. Despite the low practical significance of these results, this chapter revealed that while each of the three components of state capacity is associated with a reduction in deaths, that administrative capacity seems to be driving the results seen by the overall capacity variable. This relationship makes logical sense. Institutions that regulate where people can live, how they can build their homes, when they should evacuate, and also coordinate resuce efforts all fall under the umbrella of administrative capacity. In the immediate aftermath of a disaster and even sometimes in the short time before, directives and policies coming from a strong national bureaucracy are critically important for managing risk.

Appendix


Count
1000
500
0
0
5
10
15
ln(Total Death)