MERCATUS CENTER
George Mason University

POLICY BRIEF

A Policymaker's Guide to State RegData 2.0

Kofi Ampaabeng, James Broughel, Ethan Greist, Patrick A. McLaughlin, Jonathan Nelson, Walter Stover, Stephen Strosko, and Hayden Warlick

October 2020

The availability of data is critical in assessing the effects of government policy on the economy. A rich literature on the effects of government spending, taxes, and monetary policy exists in part because data are plentiful in these areas. Data pertaining to the extent of regulation are less available, even though the effects of regulation on the economy and people's day-to-day lives are considerable.

Historically, researchers trying to measure regulation often had to rely on crude page counts or highly imperfect (and often politicized) cost-benefit reports produced by federal agencies. Fortunately, technology in recent years has helped overcome this challenge. For instance, researchers at the Mercatus Center at George Mason University developed QuantGov, an open-source machine learning and text analysis platform for analyzing regulatory text. QuantGov can process large quantities of regulatory documents (and other policy documents), and it allowed researchers to create RegData,[1] a groundbreaking dataset that includes various dimensions of federal regulation, such as volume, applicability, and complexity.

RegData was created out of necessity. As of 2019, the US *Code of Federal Regulations* (CFR) was 185,984 pages long[2]—too long for any individual to read in its entirety without literally spending years on the task, but entirely within the capabilities of modern computers to analyze. Following the success of RegData, which now covers the years 1970 to 2019, Mercatus researchers created numerous spin-off datasets and data products including RegData Canada, RegData Australia, the Federal Regulation and State Enterprise (FRASE) index, and State RegData, the focus of this policy brief.

By using the QuantGov platform to analyze state administrative codes, researchers created version 1.0 of State RegData in 2019. It was the first dataset to quantify aggregate regulatory volume

as well as industry-specific regulatory volume across US states. State RegData 2.0 is an improvement over the previous iteration of the dataset, as it is built using an improved industry classifier and it includes additional estimates of complexity. These data should allow for much more rigorous academic research of state-level regulation and its impacts, in the same way that RegData permitted novel research on federal regulation's impact.[3] To date, state-level regulation has been understudied relative to federal regulation, likely, again, because of a dearth of high-quality data. With State RegData, that is changing.

ABOUT STATE REGDATA 2.0

State RegData 2.0 includes the cumulative volume of regulations found in the administrative codes of 44 states and the District of Columbia. Administrative codes contain rules written by administrative agencies, which are generally part of the executive branch of state governments and therefore under the direction of the governor, though they sometimes operate independently of gubernatorial oversight. Administrative laws written by these agencies have the force of law like any other law, but they are different from statutes written by elected representatives and signed into law by a governor in that they are primarily written by unelected, career civil servants (i.e., regulators). State RegData covers only administrative laws and does not contain data on statutes or the various policy documents, guidance documents, or memoranda that regulators issue—although it is worth keeping in mind that these other layers of law exist and that future versions of State RegData might be able to include them.

We collected state administrative codes between March and June of 2020. Several states could not be included in State RegData 2.0 because of limitations on data availability. First, four states that were not included in version 1.0 were also not included in version 2.0. These are Arkansas, Hawaii, New Jersey, and Vermont. Arkansas does not have a compiled administrative code, although the state passed a law in 2019 to create one.[4] Once Arkansas's administrative code is available online, the volume of regulation in Arkansas will be quantified in future iterations of State RegData. Hawaii's online code simply links to the individual websites of specific state departments. Therefore, writing a program to analyze Hawaii's code means writing a unique program for each department, which is unrealistic at this time. New Jersey and Vermont delegate control of their codes to a private company, LexisNexis. LexisNexis makes state rules freely available online for anyone to read. However, the company's terms and conditions prohibit the downloading of text,[5] which is necessary to analyze the state's rules using the QuantGov technology. As a result of these policies, we do not include New Jersey and Vermont in version 2.0 of State RegData but leave open the possibility of their inclusion in future releases.

We included two states in State RegData 1.0 that we do not include in version 2.0, owing to new circumstances. These states are Alaska and Connecticut. Alaska was updating its website in mid-2020, so only an incomplete catalog of rules was available during the time we collected codes for

State RegData 2.0. Connecticut's website could not be scraped using traditional web-scraping techniques, as the website would block programs after a certain number of pages were downloaded.

For each state, State RegData includes the following measures: word counts, restriction counts, complexity, and industry relevance. Regulatory restrictions are instances of the terms "shall," "must," "may not," "prohibited," and "required" in regulations.[6] Industry relevance refers to the probability that a portion of regulatory text is relevant to a particular industry. When multiplied by the number of regulatory restrictions in that portion of text, industry-relevant restrictions can be estimated. Industries in State RegData are distinguished from one another using the North American Industry Classification System (NAICS) and are calculated at the three-digit NAICS level.

Regulatory restrictions and industry relevance were components of State RegData 1.0, but complexity metrics are a new addition to version 2.0. Complexity measures include the average sentence length (in number of words) in a code or section of code; Shannon Entropy, which is a measure of the amount of information contained in text;[7] and the number of conditional terms.[8]

The data in State RegData 2.0 have numerous advantages, especially when compared with other metrics, such as page counts. Whereas page counts are not comparable across states owing to states' codes having different page and font sizes, counts of words or restrictive terms are comparable, as are numbers of industry-relevant restrictions or complexity metrics. Restriction and word counts are also more comprehensive in the sense that they can be gathered from all of the rules in place. Contrast this with, say, cost-benefit estimates, which are usually drawn from a handful of regulations that probably are not representative of the regulatory system as a whole.[9] Moreover, regulatory restrictions, words, and industry-relevant restrictions have a clear, concrete meaning, unlike the net-benefit calculations in regulatory agency cost-benefit reports, which are open to various interpretations.[10]

State RegData 2.0 is also an improvement over version 1.0, as it more clearly delineates the agencies that write particular regulations.[11] Version 2.0 also took less time to collect data for. Because each state's administrative code is structured differently, creating version 1.0 required writing a unique program for each state, so analyzing all the states' codes took more than three years. By contrast, releasing version 2.0 did not require having to write the initial programs from scratch, so codes were collected and analyzed over the course of only a few months in mid-2020. In that sense, version 2.0 more precisely pinpoints the stock of state regulation that existed at a particular moment in time across the various states.

The data here are also likely to prove useful to policymakers. Policymakers have used State RegData 1.0, sometimes citing its metrics in official reports or government directives. For example, Missouri claims to have reduced its regulatory restrictions by about 20 percent, as measured by State RegData.[12] Idaho's Board of Pharmacy used RegData metrics to measure and track progress

of its regulatory reforms.[13] Finally, State RegData statistics (or statistics inspired by State RegData) were referenced in regulatory reform legislation and several executive orders around the country.[14]

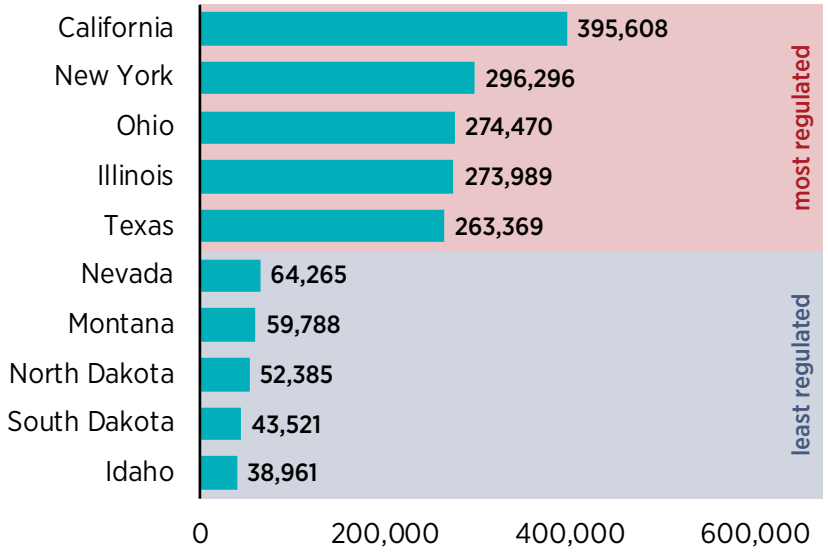
OVERVIEW OF REGDATA 2.0 STATISTICS

All told, there are 416 million words in state regulations, which contain 6.07 million regulatory restrictions. It would take about 23,000 hours, or more than 11.5 years, to read every word of every state regulatory code for the states analyzed.[15] The average state has 135,000 regulatory restrictions in its administrative rules, but the states vary hugely. California is the most regulated state, with 395,608 regulatory restrictions; Idaho is the least regulated state, with just 38,961 regulatory restrictions.

For comparison, in 2019, Canadian provinces averaged about 29,400 restrictions, and the Australian states averaged about 27,000 restrictions.[16] State regulatory codes combined are about four times the size of the CFR in word counts and five and a half times the size of the CFR in regulatory restrictions. This means there is considerably more state regulation than federal regulation. However, the federal code is still much larger than any individual state's administrative code (and many businesses may have to comply with only one state's rules).

As figure 1 shows, the top five states in restrictiveness are California, New York, Ohio, Illinois, and Texas. Meanwhile, the bottom five states in restrictiveness are Idaho, South Dakota, North Dakota, Montana, and Nevada. These results are only slightly different from those of version 1.0. One notable change, which is consistent with recent claims from the governor of Idaho,[17] is that

Figure 1. Top Five Least and Most Regulated States, by Number of Restrictions



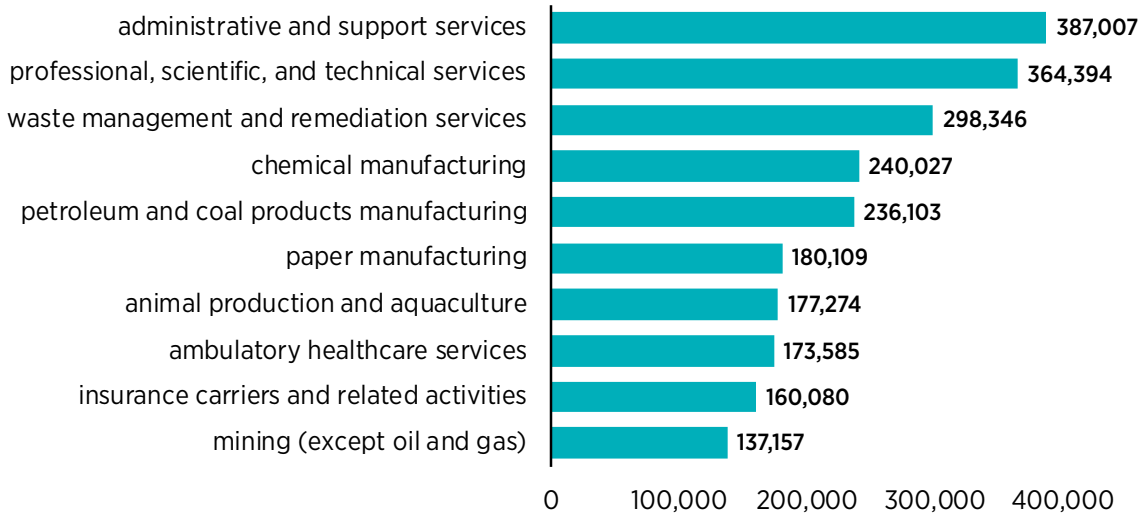
Source: Patrick A. McLaughlin et al., "State RegData 2.0" (dataset), Quantgov, accessed July 15, 2020, https://www.quantgov.org/bulk-download.

Idaho has now become the least regulated state in the nation (previously it was the fourth least regulated state).[18]

One way to classify which industries are most regulated at the state level is to aggregate industry-relevant restrictions across all states and the District of Columbia. Figure 2 presents this information. By this measure, the most regulated industry classification in the nation is NAICS code 561, administrative and support services, which includes industries such as employment services, collection agencies, and telephone call centers. Cumulatively, these industries face an estimated 387,007 restrictions across the states. Coming in second is NAICS code 541, professional, scientific, and technical services. The industries in this category face an estimated 364,393 cumulative state-level restrictions.

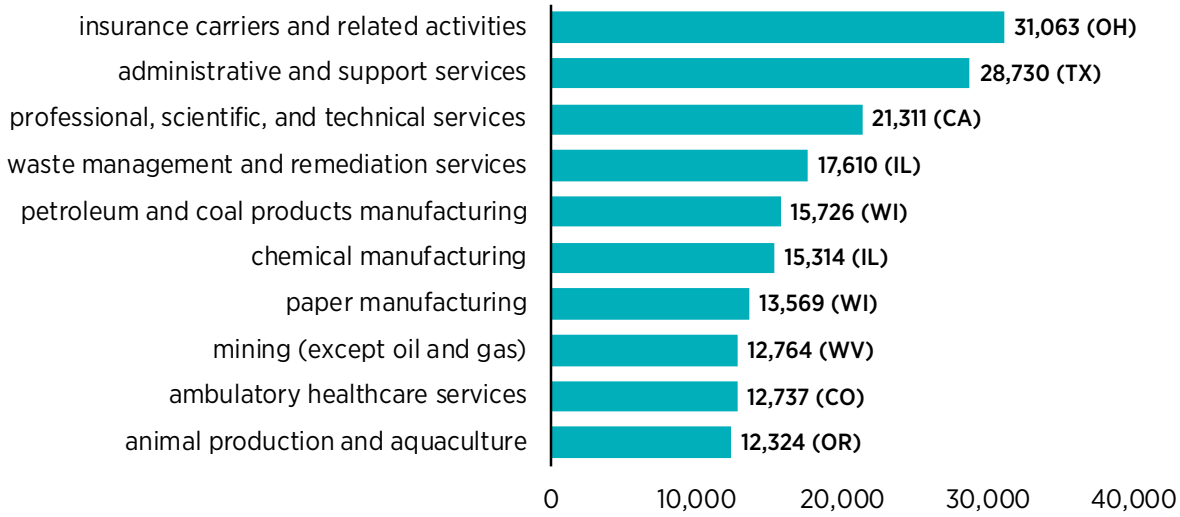
Digging a little deeper, one can identify the states that regulate each of these industries most. For example, Texas regulates administrative and support services more than any other state, targeting more than 28,700 regulatory restrictions at these industries. Ambulatory healthcare centers are regulated most in Colorado, the insurance industry is regulated most in Ohio, and mining (except oil and gas) is regulated most in West Virginia. This information is presented in figure 3.

Figure 2. Top 10 Most Regulated Industries, by Cumulative State-Level Industry Restrictions



Source: McLaughlin et al., State RegData 2.0.

Figure 3. Most Regulated Industries and the State that Regulates Them Most



Source: McLaughlin et al., State RegData 2.0.

Table 1 presents complexity metrics in state regulation. Arizona has the least complex regulations, as measured by average sentence length, with about 12 words per sentence. Meanwhile New York has the most complex regulations, by this measure, with about 42 words per sentence. South Carolina has the highest Shannon Entropy score, whereas South Dakota has the lowest. These scores suggest that South Carolina includes a lot of different kinds of information in its regulations, on average, whereas South Dakota includes less information.

Table 1. Complexity by State

STATE	AVERAGE CONDITIONALS (PER PART OF CODE)	AVERAGE SENTENCE LENGTH	SHANNON ENTROPY
Alabama	24	22	7.6
Arizona	151	12	8.6
California	117	31	7.7
Colorado	103	28	8.6
Delaware	41	27	7.8
District of Columbia	31	38	8.1
Florida	28	23	7.7
Georgia	43	32	7.4
Idaho	46	13	8.4
Illinois	48	39	8.2
Indiana	52	27	7.4
Iowa	29	22	7.7

Table 1 (continued)			
STATE	AVERAGE CONDITIONALS (PER PART OF CODE)	AVERAGE SENTENCE LENGTH	SHANNON ENTROPY
Kansas	19	19	7.2
Kentucky	9	18	7.4
Louisiana	21	19	7.7
Maine	37	22	8.2
Maryland	16	24	7.7
Massachusetts	42	29	8.1
Michigan	37	19	7.8
Minnesota	70	19	8.3
Mississippi	117	25	8.8
Missouri	27	30	7.6
Montana	84	24	8.4
Nebraska	26	26	7.5
Nevada	83	23	8.0
New Hampshire	55	31	8.1
New Mexico	20	32	7.6
New York	28	42	7.3
North Carolina	123	20	8.4
North Dakota	49	17	7.7
Ohio	89	35	8.1
Oklahoma	69	28	8.3
Oregon	41	24	8.0
Pennsylvania	43	17	8.1
Rhode Island	35	22	8.1
South Carolina	176	20	8.9
South Dakota	8	21	6.9
Tennessee	35	22	8.0
Texas	86	36	8.2
Utah	18	25	7.7
Virginia	50	16	8.0
Washington	39	22	7.7
West Virginia	32	15	8.0
Wisconsin	41	14	8.0
Wyoming	16	23	7.8

Source: McLaughlin et al., State RegData 2.0.

Table 2 presents the biggest movers in regulatory restrictions from the last iteration of State RegData. Notably, Idaho and Missouri saw the biggest reductions in regulatory restrictions in percentage

terms; both of these states have attempted to reduce red tape in recent years, and both have used State RegData to help guide their efforts.[19] Idaho saw a 37 percent reduction in regulatory restrictions, and Missouri saw a 30 percent reduction. Kentucky also instituted a red tape cutting effort under its previous governor, Matt Bevin,[20] which explains why the state saw the fourth-largest percentage reduction in the country. Nebraska, although not shown in table 2, had the sixth-largest reduction, following a regulatory reform executive order from the governor in 2017.[21]

At the other end of the spectrum, Maryland saw a 13 percent *increase* in regulatory restrictions, and New Hampshire saw an 8 percent increase. Maryland has also reviewed its regulations as part of a red tape reduction effort in recent years,[22] though Maryland's review is notable for not including many of the elements identified as key to successful regulatory reduction efforts, such as having a concrete measure regulation, estimating a baseline level of regulation, and setting a reduction target.[23] Ohio also shows up in the top five in percentage increase in regulatory restrictions. In 2019, Ohio passed regulatory reform legislation,[24] although it may not have had enough time to implement reforms that would show up in these statistics.

Table 2. Top 10 Biggest Movers by Percentage Change in Restrictions since State RegData Version 1.0

STATE	NUMBER OF REGULATORY RESTRICTIONS (VERSION 1.0)	NUMBER OF REGULATORY RESTRICTIONS (VERSION 2.0)	CHANGE IN ABSOLUTE NUMBER OF RESTRICTIONS	PERCENTAGE CHANGE IN RESTRICTIONS
Idaho	61,848	38,961	−22,887	−37.01
Missouri	134,702	93,915	−40,787	−30.28
Wyoming	99,566	71,294	−28,272	−28.40
Kentucky	127,935	116,274	−11,661	−9.11
Michigan	83,484	76,236	−7,248	−8.68
Ohio	260,977	274,470	13,493	5.17
Illinois	259,832	273,989	14,157	5.45
Pennsylvania	153,661	162,937	9,276	6.04
New Hampshire	123,423	133,592	10,169	8.24
Maryland	121,741	137,695	15,954	13.10

Source: McLaughlin et al., State RegData 2.0; McLaughlin et al., State RegData 1.0.

CONCLUSION

State RegData 2.0 is a significant improvement over version 1.0. As such, the varying regulatory landscape across states is becoming much clearer. Future research can shed light on the causes and consequences of state regulation, as well as the ways regulations implemented by different levels of government interact with one another. Whereas at one time, empirical analysis of the consequences of state regulation seemed beyond reach, that is all beginning to change with the development of modern tools like State RegData.

ABOUT THE AUTHORS

Kofi Ampaabeng is a research fellow and data scientist at the Mercatus Center at George Mason University. He specializes in curating data and generating policy-relevant insights from data. Before joining the Mercatus Center, he worked for IMPAQ International, LLC, where he evaluated the efficacy of government programs.

James Broughel is a senior research fellow at the Mercatus Center at George Mason University. Broughel has a PhD in economics from George Mason University. He is also an adjunct professor at the Antonin Scalia Law School at George Mason University.

Ethan Greist is a research associate at the Mercatus Center at George Mason University. He performs original data analysis and background research for Mercatus scholars studying regulatory and legal topics. In addition to his research work, Greist also produces short-form content for the Mercatus Center and develops the QuantGov website.

Patrick A. McLaughlin is the director of Policy Analytics and a senior research fellow at the Mercatus Center at George Mason University. His research focuses primarily on regulations and the regulatory process. McLaughlin created and leads the RegData and QuantGov projects, deploying machine learning and other tools of data science to quantify governance indicators found in federal and state regulations and other policy documents.

Jonathan Nelson is a software developer at the Mercatus Center at George Mason University. He writes code for the Policy Analytics team and has worked on RegData, Quantgov, and other related projects. He is a graduate of Grove City College, where he studied economics, mathematics, and philosophy.

Walter Stover is a Python developer with the Mercatus Center at George Mason University. Stover graduated from Allegheny College with a BA in economics and a minor in Chinese Studies. He primarily works with the Policy Analytics project at Mercatus as a software developer and data engineer, and recently completed a stay at the Institute of Public Affairs (IPA) in Melbourne, Australia, as part of a cooperative effort between Mercatus and the IPA.

Stephen Strosko is a Python developer and project coordinator for Policy Analytics at the Mercatus Center. He specializes in regulatory research and, notably, has worked on the RegData, Quantgov, FRASE, and RegData Canada projects.

Hayden Warlick is a research assistant at the Mercatus Center at George Mason University. His research interests include regulations and innovation, healthcare policy, and labor economics. Previously, he has worked in education administration and digital marketing. He received a BA in economics and a BS in marketing from California State University, Northridge.

NOTES

1. The methodology underlying RegData is described in Omar Al-Ubaydli and Patrick A. McLaughlin, "RegData: A Numerical Database on Industry-Specific Regulations for All United States Industries and Federal Regulations, 1997–2012," *Regulation & Governance* 11, no. 1 (2015): 109–23.

2. "Federal Register Statistics," Federal Register, accessed October 20, 2020, https://www.federalregister.gov/reader-aids/understanding-the-federal-register/federal-register-statistics.

3. As an example, see Bentley Coffey, Patrick A. McLaughlin, and Pietro Peretto, "The Cumulative Cost of Regulations," *Review of Economic Dynamics*, published ahead of print, April 2, 2020, https://doi.org/10.1016/j.red.2020.03.004.

4. H.B. 1429, 92nd Gen Assemb., Reg. Sess. (Ark. 2019).

5. LexisNexis terms and conditions, state that "you and your Authorized Users are prohibited from downloading, emailing, faxing, storing, reproducing, transmitting, displaying, copying, distributing, or using Materials retrieved from the Online Services." Exceptions exist that may allow for these states to be included in the future, but we have opted to exclude them for the time being. "Terms & Conditions of Use for the LexisNexis Services," LexisNexis, December 2, 2019, https://www.lexisnexis.com/en-us/terms/general/default.page.

6. Regulatory restrictions can occur in legal text for purposes other than imposing binding requirements on private parties, such as for definitional purposes. At times, restrictions may relate to government employees rather than the private sector.

7. Shannon entropy can be thought of as measuring the number of new ideas that are introduced in a document, or, alternatively, how much computational effort would be required to understand a document.

8. Conditional terms are "if," "but," "except," "provided," "when," "where," "whenever," "unless," "notwithstanding," "in the event," and "in no event."

9. James Broughel and Richard Williams, "More Information Needed on the Benefits and Costs of Regulations," *The Bridge*, August 22, 2018.

10. James Broughel, "Cost-Benefit Analysis as a Failure to Learn from the Past," *Journal of Private Enterprise* 35, no. 1 (2020): 105–13; James Broughel, "The Mighty Waves of Regulatory Reform: Regulatory Budgets and the Future of Cost-Benefit Analysis," *Business, Entrepreneurship & Tax Law Review* 3, no. 2 (2019): 206–23.

11. Better agency parsing means that users can now get data at the summary, agency, or document levels ("title-part") for the states included in version 2.0. These data can be retrieved using either the QuantGov interactive downloader or the RegCensus API. See both at "Download QuantGov Data," QuantGov, accessed July 15, 2020, https://www.quantgov.org/download-data.

12. Justin D. Smith, "Regulatory Reform at the State Level: A Guide to Cutting Red Tape for Governors and Executive Branch Officials," *Business, Entrepreneurship & Tax Law Review* 3, no. 2 (2019): 276–99.

13. Alex J. Adams, "Transitioning Pharmacy to 'Standard of Care' Regulation: Analyzing How Pharmacy Regulates Relative to Medicine and Nursing," *Research in Social and Administrative Pharmacy* 15, no. 10 (2019): 1230–35.

14. Examples include H.B. 166, 133rd Gen. Assemb., Reg. Sess. (Ohio 2019); Idaho Exec. Order 2019-02 (2019); Oklahoma Exec. Order 2020-03 (2020); Nebraska Exec. Order No. 17-04 (2017).

15. This assumes the reader reads 300 words per minute for 8 hours a day, 40 hours a week, and 50 weeks out of the year.

16. Patrick A. McLaughlin, Jonathan Nelson, and Stephen Strosko, "RegData Canada" (dataset), QuantGov, accessed October 2020, https://www.quantgov.org/regdata-canada-documentation; Patrick A. McLaughlin and Walter Stover, "RegData Australia" (dataset), QuantGov, accessed October 2020, https://www.quantgov.org/regdata-australia-documentation.

17. Idaho Office of the Governor, "Idaho Cuts and Simplifies 75 Percent of Rules in One Year, Becomes Least-Regulated State in Country," press release, December 4, 2019, https://gov.idaho.gov/pressrelease/idaho-cuts-and-simplifies-75-percent-of-rules-in-one-year-becomes-least-regulated-state-in-country/.

18. James Broughel and Krista Chavez, "Idaho Is the Least Regulated State and a Model for the Rest of the Country," *The Bridge*, January 2, 2020.

19. Smith, "Regulatory Reform at the State Level"; Idaho Exec. Order 2019-02 (2019).

20. James Broughel, "Tracking the Progress of Kentucky's Red Tape Reduction Initiative" (Mercatus Policy Brief, Mercatus Center at George Mason University, Arlington, VA, June 2019).

21. Nebraska Exec. Order No. 17-04 (2017).

22. Maryland Exec. Order No. 01.01.2015.20 (2015).

23. James Broughel, "A Step-by-Step Guide to Using Mercatus Tools to Reduce State Regulation Levels" (Mercatus on Policy, Mercatus Center at George Mason University, Arlington, VA, April 2017).

24. H.B. 166, 133rd Gen. Assemb., Reg. Sess. (Ohio 2019).