

# Census of Regulatory Restrictions

---

Kofi Ampaabeng, Patrick McLaughlin, and Thurston Powers



**MERCATUS CENTER**

**George Mason University**

3434 Washington Blvd., 4th Floor, Arlington, Virginia 22201

[www.mercatus.org](http://www.mercatus.org)

*Kofi Ampaabeng, Patrick McLaughlin, and Thurston Powers, “Census of Regulatory Restrictions,” Mercatus Research Paper, Mercatus Center at George Mason University, Arlington, VA, April 2022.*

## ABSTRACT

Ever since its introduction in 2012, RegData—the Mercatus Center’s database for quantifying federal regulation—has served as an objective tool for measuring regulations within a jurisdiction. In the decade since its launch, RegData has expanded significantly to cover several jurisdictions, both national and subnational. Using RegData, we compare the volume and other features of the regulatory landscapes of three large countries—Australia, Canada, and the United States—and their states and provinces. In addition to comparing the volume of regulations, we also discuss the complexity of the regulations, the industries that are affected by those regulations, and, in the case of the United States, the agencies that issue them.

*JEL* codes: K0, K1, K2, K23

Keywords: Regulations; RegData; United States; Australia; Canada; Regulatory Agencies; Regulated Industries

© 2022 by Kofi Ampaabeng, Patrick McLaughlin, Thurston Powers, and the Mercatus Center at George Mason University.

The views expressed in Mercatus Research are the authors’ and do not represent official positions of the Mercatus Center or George Mason University.

**R**egulations—state and federal—have always been a subject of interest because they shape the lives of many Americans. Several studies have shown the importance of regulations, yet deleterious effects have been documented. What has been missing up until this point is an objective way of measuring regulations, and RegData stepped in to fill this void. How many regulations are there in a jurisdiction? What industries do they affect? Which agencies promulgate the most regulations? How easy is it to understand and comply with regulations? These are all relevant questions that RegData seeks to answer. Launched in 2012, RegData presented an objective way of measuring regulations and helps to answer these and many other questions.

Hitherto, analysis of the effects of regulations utilized crude measures, such as page counts of regulations. RegData provides an objective alternative that is useful for quantitative analyses of regulations. RegData has changed the way regulations are measured. More than 20 peer-reviewed publications have used RegData as a primary source of data for analysis. In addition, analyses of regulatory restrictions using RegData have spurred regulatory reforms in a number of jurisdictions.

RegData has now been applied to the regulatory code of other countries, including Australia, Brazil, Canada, Denmark, and India. In addition, data exist for subnational jurisdictions in Australia, Canada, and the United States. Those data are updated annually and therefore serve as a rich resource that is useful for analyzing the evolution of the regulatory landscape across countries. The use of a common methodology to measure regulations across multiple jurisdictions enables analysts to compare the regulations in an objective and consistent manner. To that end, we introduce the Census of Regulatory Restrictions (“RegCensus”), a comprehensive database of regulatory restrictions from RegData across multiple jurisdictions, allowing for comparison and other analyses.

RegData is built using QuantGov, a library of machine learning algorithms and text analysis tools. QuantGov quantifies regulatory restrictions by counting the number of restrictions within the regulatory text. Even more important,

these algorithms allow us to determine the industry or business sectors that are likely to be affected by a unit of regulation. To date, we have produced RegData at the national level for Australia (2005–2020), Canada (2006–2020), and the United States (1970–2019). At the subnational (state or province) level, we have produced regulatory restriction counts for all but four US states (Arkansas, Hawaii, New Jersey, and Vermont<sup>1</sup>) and all Canadian provinces.

In this paper, we examine the regulatory landscape of Australia, Canada, and the United States using RegData. We examine broad patterns in the nature and growth of restrictions, the regulators, and the regulated industries across several jurisdictions, including federal and state (or province, as in Canada) levels. We use the three terms already defined—QuantGov, RegData, and RegCensus—to refer to the different tools or products.

The rest of the paper proceeds as follows. We begin with an overview of RegData and its underlying methodology. Then, we discuss the patterns in the growth of regulations at the federal level across the three countries. Next, we examine the agency-specific patterns in the stock and growth of federal regulations in the United States. After that, we discuss the nature of subnational regulatory restrictions. Finally, we conclude this overview of RegCensus and provide suggestions and ideas for further research on the impact of regulations on the economy using more vigorous statistical and economic methods.

## OVERVIEW OF REGDATA METHODOLOGY AND DATA

### History of RegData

In years past, regulation was a phenomenon that went unmeasured, rendering discussions and research related to the regulatory process qualitative and abstruse. On those rare occasions when regulation was quantified, researchers might measure it by counting pages published. The problems with measuring regulation in this way are well documented: in addition to being noisy because many pages have nothing to do with regulation, this measurement method also runs the risk of counting deregulation as an increase in regulation because deregulation requires the publication of pages. Even more important, not all regulations are created equal terms of in their effects on individuals and businesses.

---

1. RegData restrictions are not available for these states for various reasons. Because of copyright reasons, we are unable to extract and analyze Vermont and New Jersey regulatory texts. Hawaii and Arkansas regulatory texts are more difficult to extract. In addition, we are unable to include data for Connecticut after 2017 due to a change in the state's website.

One page of regulatory text can be quite different from another in content and consequence. Measuring regulation by counting pages, therefore, misses a lot of detail that could be useful in understanding the causes and effects of regulation.

RegData introduced an objective, replicable, and transparent methodology for measuring regulation (Al-Ubaydli and McLaughlin 2017). RegData improved on existing measures of regulation in two principal ways:

1. **RegData quantifies regulations on the basis of the actual content of regulatory text.** In other words, this custom-made program examines the regulatory text itself, counting the number of binding constraints or “restrictions” words that indicate an obligation to comply, such as *shall* or *must*. This function creates a more precise metric because some regulatory programs can be hundreds of pages long with few restrictions, whereas others have only a few paragraphs with a high number of restrictions.
2. **RegData quantifies regulations by industry.** RegData determines the probability that a given regulatory restriction is targeting a specific industry, thereby allowing the creation of industry-specific measures of regulation over time. RegData uses the same business sectors (or industry classes) as the North American Industrial Classification System (NAICS), which categorizes and describes each industry in the US economy.<sup>2</sup> Using industry-specific quantifications of regulation, users can examine the growth of regulation relevant to an industry over time or compare growth rates across industries.

Industry-specific measurements of regulation can be used in several ways. Both the causes and the effects of regulation can differ from one industry to the next; with quantified regulations for all industries, users can test whether industry characteristics—such as industry growth, dynamism, employment, or a penchant for lobbying—are connected to industry-specific regulation levels.

Appendix A contains a summary of the evolution of RegData, including the methodological changes between 2012 and 2020.

## Methodology and Data Sources

RegData uses text analysis and machine learning algorithms to quantify the number of restrictions in regulations in a jurisdiction. At the federal level in the

---

2. The next iteration of RegData will include industry classifiers for the Standard Industrial Classification (SIC) and the Australia and New Zealand Standard Industrial Classification (ANZSIC) and will expand to cover more jurisdictions.

United States, those regulations are reported in the Code of Federal Regulations (CFR). States and provinces have varying degrees of regulatory text available. RegData relies on QuantGov, a machine learning platform that is applicable to many settings that use text as the input data source. RegData comprises three parts: identifying restrictions, assigning probabilities that a given regulation applies to a given industry, and determining the complexity of the language of the regulatory text.

**Regulatory Restrictions.** The first part of RegData is to count the number of restrictions within a given regulatory text. Using a predefined set of terms, QuantGov identifies the number of restrictions a piece of regulation imposes (the details of this implementation are discussed in Al-Ubaydli and McLaughlin 2017). QuantGov searches regulations for the occurrences of restrictive terms such as *shall*, *must*, *may not*, *prohibited*, and *required*. The total number of occurrences of those terms is referred to as *restrictions*. In addition, RegData includes the total number of words in each regulatory text.

**Industry-Relevant Restrictions.** The second part of RegData is the estimated likelihood that a given regulation applies to a given industry. Regulatory text often does not explicitly identify the industries that would be directly affected. For a useful analysis of the impacts of regulations on industry, one must identify all affected industries. Al-Ubaydli and McLaughlin (2017) recommend and use logistic algorithms to predict the probability that a given regulation applies to an industry. In RegData, the term *industry restrictions* simply refers to the probability that a regulation applies to a given industry, multiplied by the total number of restrictions identified in the first step; it is one of the key features of RegData. Those estimates are produced at the two- and three-digit NAICS code (McLaughlin and Sherouse 2019). The terms *business sector* and *industry* are used interchangeably, and they refer to the NAICS classification of industries.

**Complexity of Regulatory Language.** RegData also includes data on the complexity of the language used in writing regulations—the ease of reading and understanding the regulatory text. When regulations are written as clearly as possible, they are easy to understand, and—all other things being equal—less costly to comply with. RegData provides three measures of language complexity: the Shannon entropy score, the average length of a sentence, and the average number of conditional clauses in a sentence.

1. **Shannon's Entropy.** Shannon's entropy is a concept in the field of information theory,<sup>3</sup> and it is useful for measuring the information content in a body of text. All other things being equal, a document with a lower Shannon's entropy score is less complex and easier to read than a document with a higher score. Conversely, a higher score means the content of the text spans a wider range of topics and concepts compared with lower scores. Katz and Bommarito (2014) adopted Shannon's entropy and other concepts to explore the complexity of legal texts.
2. **Average Length of a Sentence.** In general, long sentences are difficult to read and understand. This data element in RegData measures the average number of words in a sentence in the regulatory text.
3. **Average Number of Conditional Clauses.** The number of conditional clauses in a document measures the occurrence of the terms *if, but, except, provided, when, where, whenever, unless, notwithstanding, in the event, and in no event*. When a document has several of these conditional terms, it becomes difficult to read and understand. No generally applicable standard exists regarding the acceptable number of conditional terms in a sentence; however, this measure could be useful in comparing the readability of different documents or regulatory texts.

Table 1 presents the jurisdictions, including national and subnational, for which RegData have been produced. Broadly speaking, RegData are available in two types: panel and cross-sectional. Depending on the jurisdiction, regulatory data are available annually—that is, they are published or organized for each year the current regulations are in force. RegData are available for those jurisdictions, thus panel data are available. Panel data are available at the national level for the three countries in this study. Some states or provinces—such as Idaho, Indiana, Kentucky, Missouri, Oregon, and Washington in the United States and Ontario and British Columbia in Canada—have panel data. The remaining subnational jurisdictions do not have panel data available. However, since RegData is updated annually, there is now a short panel of data for most subnational jurisdictions.

---

3. For another application of Shannon's entropy to measure document complexity, see Mark Febrizio, Scott King, Patrick A. McLaughlin, and Oliver Sherouse. *Is Dodd-Frank the Biggest Law Ever?* Mercatus Policy Brief. *forthcoming*.

TABLE 1: AVAILABLE DATA

Administrative level	Country	Jurisdiction	Start year	End year	Years available
National	Australia	Australia	2005	2020	16
	Canada	Canada	2006	2020	15
	United States	United States	1970	2019	50
State/province	Australia	New South Wales	2019	2020	2
		Queensland	2019	2020	2
		South Australia	2019	2020	2
		Tasmania	2019	2020	2
		Victoria	2019	2020	2
		Western Australia	2019	2020	2
	Canada	Alberta	2018	2020	3
		British Columbia	2004	2020	17
		Manitoba	2018	2020	3
		New Brunswick	2018	2020	3
		Newfoundland and Labrador	2018	2020	3
		Northwest Territories	2018	2020	3
		Nova Scotia	2018	2020	3
		Nunavut	2018	2020	3
		Ontario	2004	2020	16
		Prince Edward Island	2018	2020	3
		Quebec	2018	2020	3
		Saskatchewan	2018	2020	3
		Yukon	2018	2020	3
	United States	Alabama	2019	2020	2
		Alaska	2019	2019	1
		Arizona	2017	2020	2
		California	2019	2020	2
		Colorado	2017	2020	2
		Connecticut	2017	2017	1
		Delaware	2019	2020	2
		District of Columbia	2019	2020	2
		Florida	2017	2020	2
		Georgia	2019	2020	2
		Idaho	1996	2020	22
		Illinois	2017	2020	2
		Indiana	2003	2020	17
Iowa		2017	2020	2	
Kansas		2019	2020	2	
Kentucky		2010	2020	10	
Louisiana	2019	2020	2		
Maine	2018	2020	2		

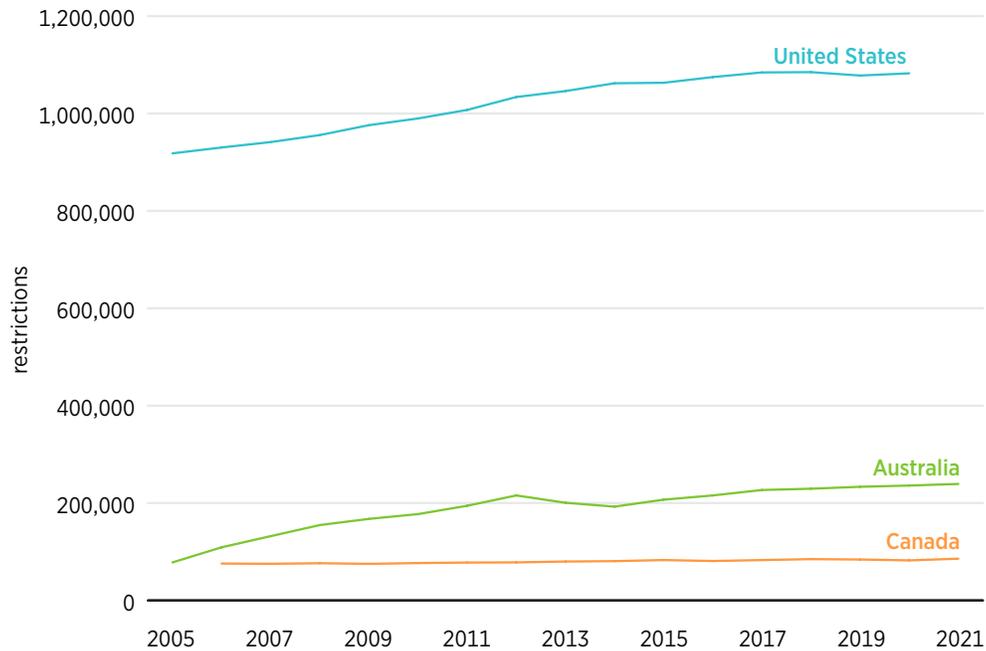
Administrative level	Country	Jurisdiction	Start year	End year	Years available
State/province	United States	Maryland	2017	2020	2
		Massachusetts	2019	2020	2
		Michigan	2017	2020	2
		Minnesota	2017	2020	2
		Mississippi	2018	2020	2
		Missouri	2001	2020	17
		Montana	2019	2020	2
		Nebraska	2017	2020	2
		Nevada	2019	2020	2
		New Hampshire	2019	2020	2
		New Mexico	2018	2020	2
		New York	2017	2020	2
		North Carolina	2017	2020	2
		North Dakota	2019	2020	2
		Ohio	2018	2020	2
		Oklahoma	2019	2020	2
		Oregon	2009	2020	11
		Pennsylvania	2017	2020	2
		Rhode Island	2019	2020	2
		South Carolina	2019	2020	2
		South Dakota	2019	2020	2
		Tennessee	2018	2020	2
		Texas	2018	2020	2
		Utah	2017	2020	2
		Virginia	2017	2020	2
		Washington	2005	2020	12
		West Virginia	2017	2020	2
		Wisconsin	2017	2020	2
		Wyoming	2018	2020	2

Source: Patrick A. McLaughlin et al., State RegData 2.0 (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2020, <https://quantgov.org/state-regdata/>.

## FEDERAL REGULATORY RESTRICTIONS—AUSTRALIA, CANADA, AND THE UNITED STATES

This section discusses the main features and trends of federal regulatory restrictions in Australia, Canada, and the United States. The total number of regulatory restrictions, the total number of words found in the regulatory code, and the relevance of a piece of regulation to an industry (using two- or three-digit NAICS codes) are discussed for Australia (2005–2020), Canada (2006–2020), and the United States (1970–2020). This section begins by examining the total number of

FIGURE 1. ACCUMULATION OF REGULATORY RESTRICTIONS



restrictions for the years with data available. To facilitate comparison, analyses for the years 2006–2020, the period in which all three countries have data, are also included.

## Accumulation of Regulatory Restrictions

Unsurprisingly, the United States has the highest number of total regulatory restrictions among the three countries (figure 1). As will be seen in figure 4, after accounting for population, Australia has the highest per capita regulatory restrictions, significantly outpacing both the United States and Canada.

**Regulatory Accumulation in Australia, Canada, and the United States.** In 1970, the United States had 400,000 regulatory restrictions. By 2019, the number of restrictions had more than doubled, to more than 1 million. Figure 1 compares the number of restrictions across the three countries between 2006 and 2020. Longer series are available on the QuantGov website.

Between 2006 and 2020, the number of regulatory restrictions in Australia more than doubled, compared with only 13 percent growth in Canada and 16 percent in the United States.

FIGURE 2. NUMBER OF WORDS IN REGULATORY TEXT

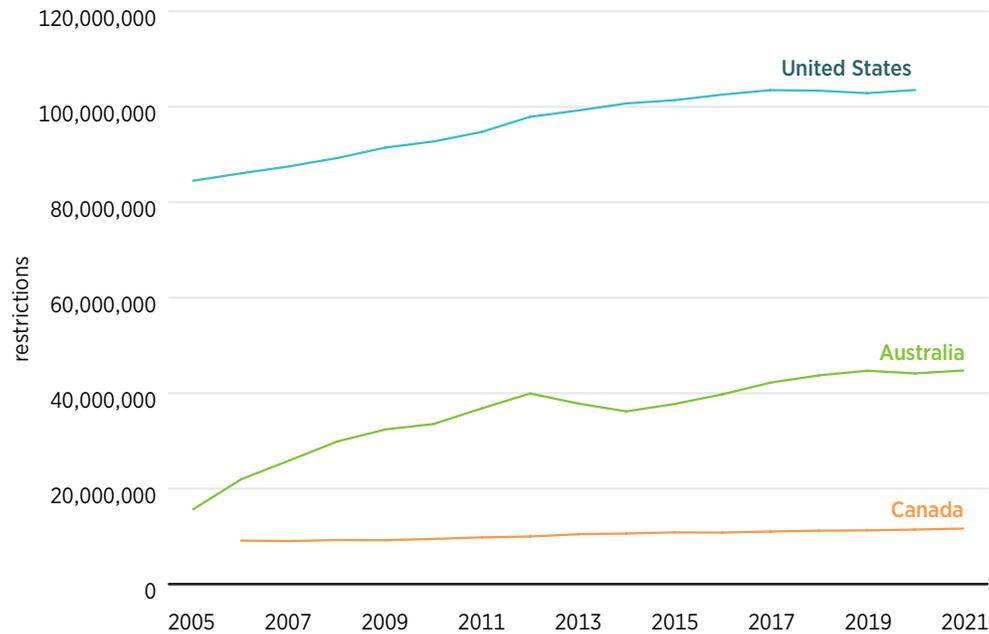


TABLE 2. CHANGES IN REGULATORY RESTRICTIONS BETWEEN 2006 AND 2019

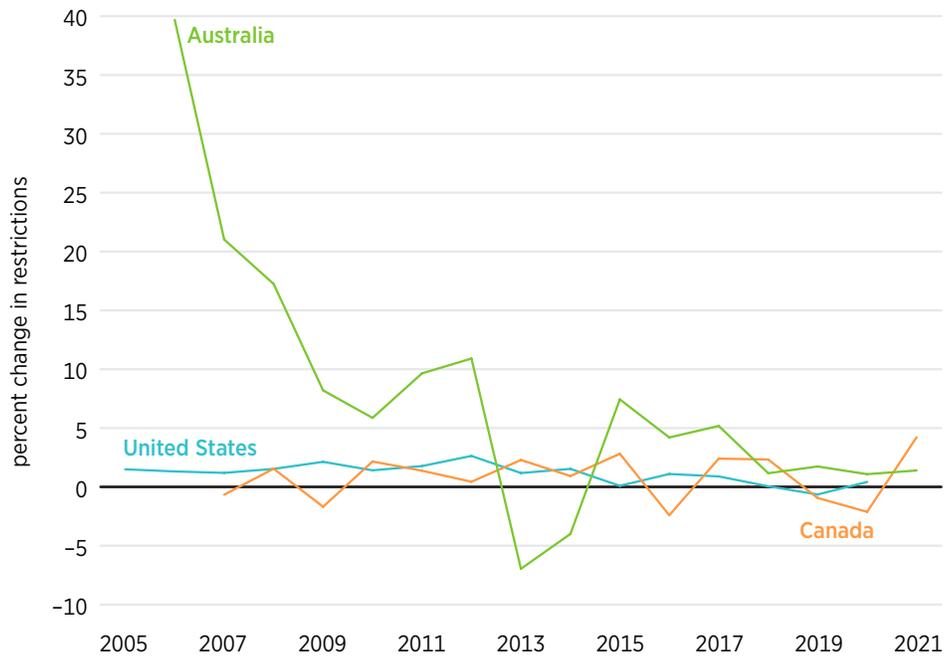
Country	2006	2010	2014	2019	Percent change (2006 to 2019)
Australia	109,087	177,308	192,595	233,400	114
Canada	77,299	78,861	83,696	87,525	13.2
United States	930,447	989,779	1,062,077	1,078,213	15.9

Sources: Patrick A. McLaughlin et al., State RegData 2.0 (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2020, <https://quantgov.org/state-regdata/>; authors' calculations.

Figure 2 shows the total number of words in regulatory text for all three countries. Table 2 shows the percentage change in regulatory restrictions between 2006 and 2019 for the three countries shown in figure 1.

**Year-over-Year Growth in Regulations.** Figure 1 shows the cumulative growth in regulatory restrictions but masks the variability in the growth over the years. Figure 3 plots the annual percentage change in the stock of regulatory restrictions for the three countries for the period 2007–2019. The growth rate of regulations in the United States over the period fluctuates within a narrow range, with a high of 4 percent in 2012, except for brief periods in the 1980s and once in the 1990s (not shown in figure 3), when the number of regulatory restrictions declined (negative growth rate). In recent years, the United States has experienced a reduction in regulatory

FIGURE 3. ANNUAL GROWTH IN REGULATORY RESTRICTIONS



accumulation. In 2019, the growth rate was negative from the previous year, following a downward trend that started in 2017. The negative growth rate of regulatory restrictions in the United States over 2017 to 2019 could be attributed to Executive Order (EO) 13771. Known as the “one-in, two-out” executive order, EO 13771 requires executive agencies to offset the costs of any new regulations by changing or eliminating existing regulations. In contrast to the United States, Canada occasionally experiences an actual reduction in the number of regulatory restrictions. For the period 2006–2020, figure 3 shows wild fluctuations in the annual growth rates of restrictions in Canada and Australia but a steady rate for the United States.

Table 3 shows the number of regulatory restrictions added each year in all three countries. In addition, it includes a five-year moving average of the annual percentage change in restrictions.

**Restrictions per Capita.** To put the number of restrictions in perspective, we considered the number of restrictions per person for the period. For display purposes, the restrictions per 1,000 people are shown in figure 4. Although the United States has more restrictions than both Australia and Canada, Australia leads with the most restrictions after normalizing by population. For example, in 2006, the data showed 5.5 restrictions for each 1,000 Australians, compared

**TABLE 3. GROWTH IN FEDERAL REGULATORY RESTRICTIONS**

Year	Australia			Canada			United States		
	New restrictions	% change	5-year MVA of % change	New restrictions	% change	5-year MVA of % change	New restrictions	% change	5-year MVA of % change
2007	22,935	21.0	0.0	-502	-0.6	0.0	10,850	1.2	0.0
2008	22,769	17.2	0.0	1,166	1.5	0.0	14,235	1.5	0.0
2009	12,694	8.2	0.0	-801	-1.0	0.0	20,516	2.1	0.0
2010	9,823	5.9	0.0	1,699	2.2	0.0	13,731	1.4	0.0
2011	17,107	9.6	12.4	1,235	1.6	0.7	17,547	1.8	1.6
2012	21,209	10.9	10.4	1,037	1.3	1.1	26,521	2.6	1.9
2013	-15,007	-7.0	5.5	1,138	1.4	1.1	12,189	1.2	1.8
2014	-8,022	-4.0	3.1	1,425	1.7	1.6	16,041	1.5	1.7
2015	14,321	7.4	3.4	2,557	3.1	1.8	996	0.1	1.4
2016	8,694	4.2	2.3	-1,824	-2.1	1.1	11,564	1.1	1.3
2017	11,168	5.2	1.2	2,008	2.4	1.3	9,691	0.9	1.0
2018	2,639	1.2	2.8	2,395	2.8	1.6	586	0.1	0.7
2019	3,983	1.7	3.9	-1,307	-1.5	0.9	-6,701	-0.6	0.3

Source: Patrick A. McLaughlin et al., State RegData 2.0 (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2020, <https://quantgov.org/state-regdata/>.

MVA = moving average.

**FIGURE 4. RESTRICTIONS PER CAPITA**

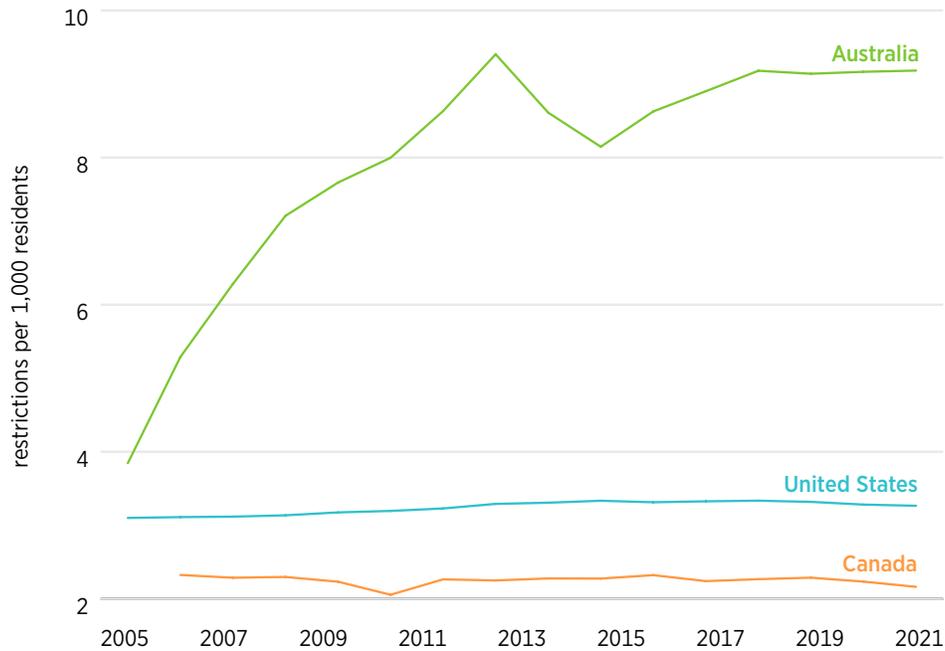


TABLE 4. REGULATORY LANGUAGE COMPLEXITY

Country	Year	Average sentence length	Shannon's entropy	Average number of conditional terms
Australia	2020	29.88	6.81	9.85
Canada	2020	68.74	7.05	21.78
United States	2019	24.68	7.86	83.35

Source: Patrick A. McLaughlin et al., State RegData 2.0 (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2020, <https://quantgov.org/state-regdata/>.

with 3.2 for Americans and 2.4 for Canadians. In addition, Australia's restrictions per capita seems to have increased significantly between 2006 and 2012, before a slight dip beginning in 2013.

### Complexity of Regulatory Language

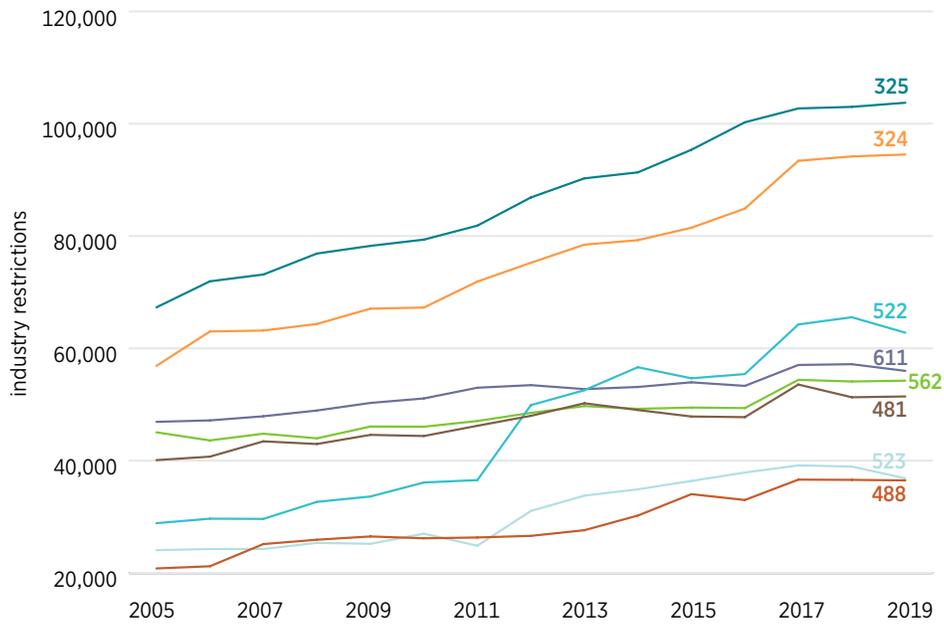
Because all three countries—Australia, Canada, and the United States—use English as the official language, one can compare the complexity of the regulatory language. Table 4 shows the average sentence length, the Shannon's entropy score, and the average number of conditional clauses found in the regulatory texts of the three countries.

US regulatory texts tend to have fewer words in a sentence (lower sentence length) than the regulatory texts of both Australia and Canada. On the other two measures of complexity, Australian and Canadian texts are less complex. The average Shannon's entropy score of a unit of US regulatory text—the CFR—is 7.86, compared with 6.81 and 7.05 for regulatory texts in Australia and Canada, respectively. Each unit of the CFR has an average of 83.35 conditional terms, compared with only 9.85 for Australian regulatory texts and 21.78 for Canadian texts.

### Industry-Specific Restrictions

Within the context of the economy, regulations affect businesses; therefore, understanding the number of regulatory restrictions borne by a business within an industry is important. This section examines the incidence—the distribution of a given regulation—of regulatory restrictions across industries, defined by the NAICS. As described earlier, estimating the probability that an industry is affected by a given regulation is a novel feature of RegData. RegData first estimates the absolute number of restrictions and then the estimated probability that an industry is affected by those restrictions. The industry relevance of a piece of regulation is the probability that an industry is affected by a given regulation multiplied

FIGURE 5. MOST REGULATED INDUSTRIES: UNITED STATES



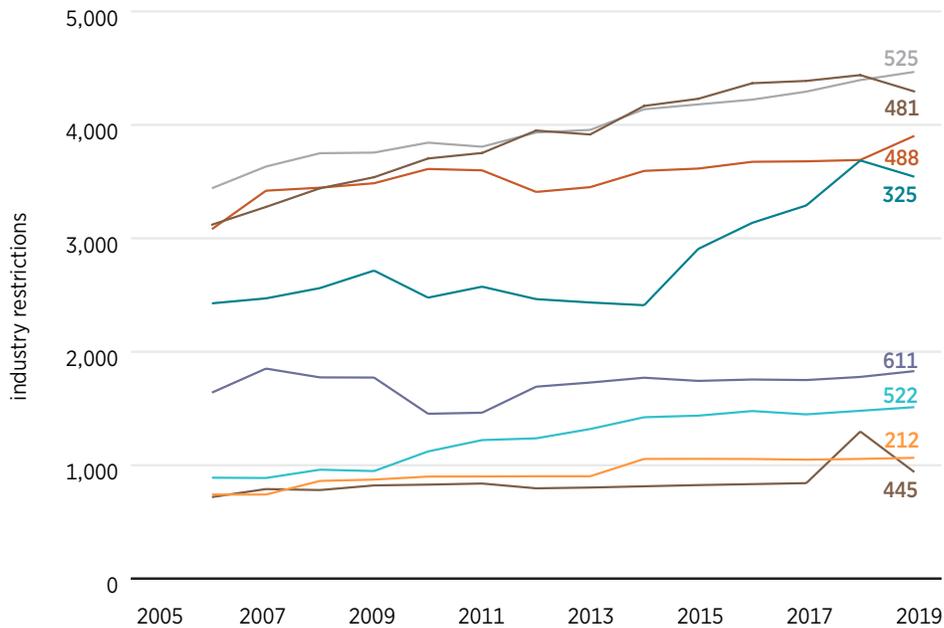
Note: 324 = Petroleum and Coal Products Manufacturing; 325 = Chemical Manufacturing; 481 = Air Transportation; 488 = Support Activities for Transportation; 522 = Credit Intermediation and Related Activities; 523 = Securities Commodity Contracts and Other Financial Investments and Related Activities; 562 = Waste Management and Remediation Services; 611 = Educational Services.

by the number of restrictions found in the first step. The terms *industry-relevant restrictions* and *industry-relevant word counts* are used to describe the regulatory restrictions assigned to the industries. As a methodological note, the sum of industry-relevant restrictions will not be equal to the total restrictions because a unit of regulation would typically apply to multiple industries.

**Industry Restriction Trends: United States.** In the United States, in line with the general secular increase in regulatory restrictions, there has been a significant increase in the number of restrictions on specific industries. For example, between 1970 and 1990, the number of restrictions on the chemical manufacturing industry increased by 631 percent. Over the next 20-year period, they increased by another 519 percent. The story is similar for the petroleum and coal products manufacturing industry. Figure 5 shows the most regulated industries in the United States from 2005 to 2019. The industries were selected on the basis of the top eight most restricted industries in 2019.

In 1980, the chemical manufacturing industry was the least restricted among these eight business sectors; by 2019, it had the most restrictions. In absolute terms,

FIGURE 6. MOST REGULATED CANADIAN INDUSTRIES



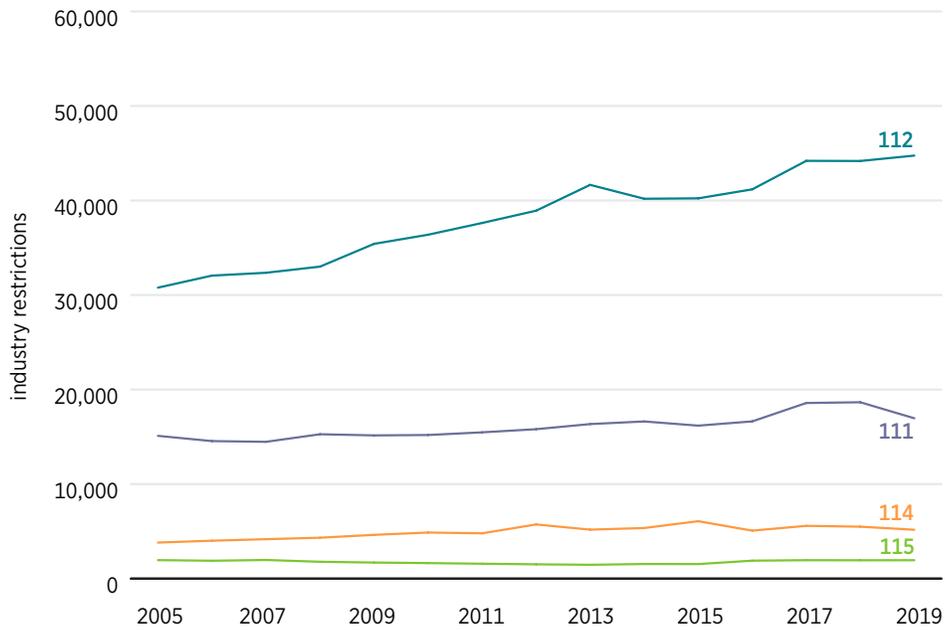
Note: 212 = Mining (except Oil and Gas); 325 = Chemical Manufacturing; 445 = Food and Beverage Stores; 481 = Air Transportation; 488 = Support Activities for Transportation; 522 = Credit Intermediation and Related Activities; 525 = Funds Trusts and Other Financial Vehicles; 611 = Educational Services.

the number of restrictions on the chemical manufacturing sector increased from 4,444 in 1980 to 70,888 in 2006 and 103,848 in 2019. Between 2006 and 2019 (the years with common data for both the United States and Canada), the restrictions on this sector increased by 31 percent. The story is similar for other industries, including the petroleum and coal products manufacturing sector.

**Industry Restriction Trends: Canada.** In contrast with the United States, the top eight most regulated industries in Canada have remained stable over the period (figure 6). The most regulated industry, air transportation, had 3,365 restrictions in 2007; that number had risen to 4,555 in 2019—an increase of 26 percent.

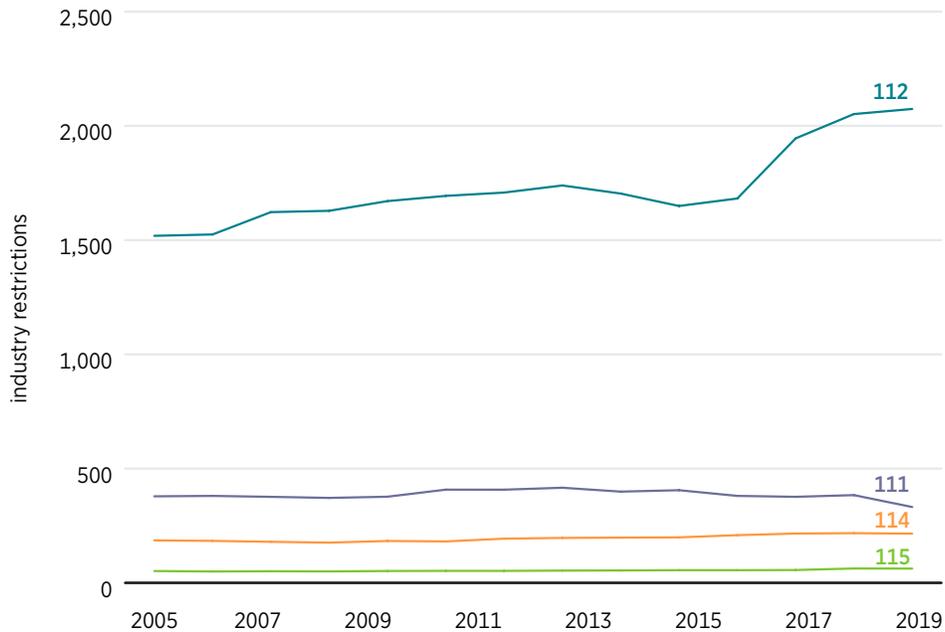
**Comparison of Industry Groups between Canada and the United States.** The next sets of figures compare industry-relevant restrictions in classes of industries. Figures 7 and 8 compare regulatory restrictions on the agricultural sectors in the United States and Canada. In both countries, the animal production and aquaculture sectors are more restricted than the other agriculture sectors. Figures 9 and 10 compare regulatory restrictions on transportation industries. Again, both countries are similar in terms of the relative number of restrictions

FIGURE 7. RESTRICTIONS ON US AGRICULTURE INDUSTRIES



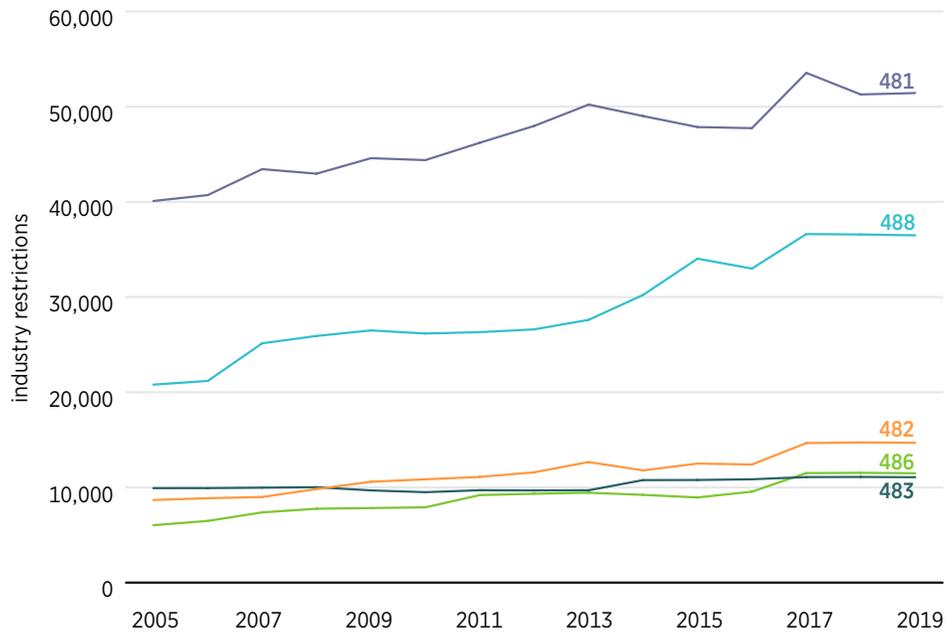
Note: 111 = Crop Production; 112 = Animal Production and Aquaculture; 114 = Fishing, Hunting, and Trapping; 115 = Support Activities for Agriculture and Forestry.

FIGURE 8. RESTRICTIONS ON CANADIAN AGRICULTURE INDUSTRIES



Note: 111 = Crop Production; 112 = Animal Production and Aquaculture; 114 = Fishing, Hunting, and Trapping; 115 = Support Activities for Agriculture and Forestry.

FIGURE 9. RESTRICTIONS ON US TRANSPORTATION INDUSTRIES



Note: 481 = Air Transportation; 482 = Rail Transportation; 483 = Water Transportation; 486 = Pipeline Transportation; 488 = Support Activities for Transportation.

on the sectors. The air transportation sector is the most restricted, followed by the support activities for transportation sector.

## CHARACTERISTICS OF REGULATING AGENCIES (US FEDERAL ONLY)

In the United States, the number of agencies that issue regulations has increased over the years. Figure 11 shows the growth in the number and types of agencies over time, starting from 1901. To the knowledge of the authors, the actual number of agencies of the US government is not clearly defined. This paper examines the agencies that have issued regulations in the CFR and classifies them into the agency types described in the following paragraphs, yielding 150 agencies. Most are not rulemaking agencies but have issued regulations.

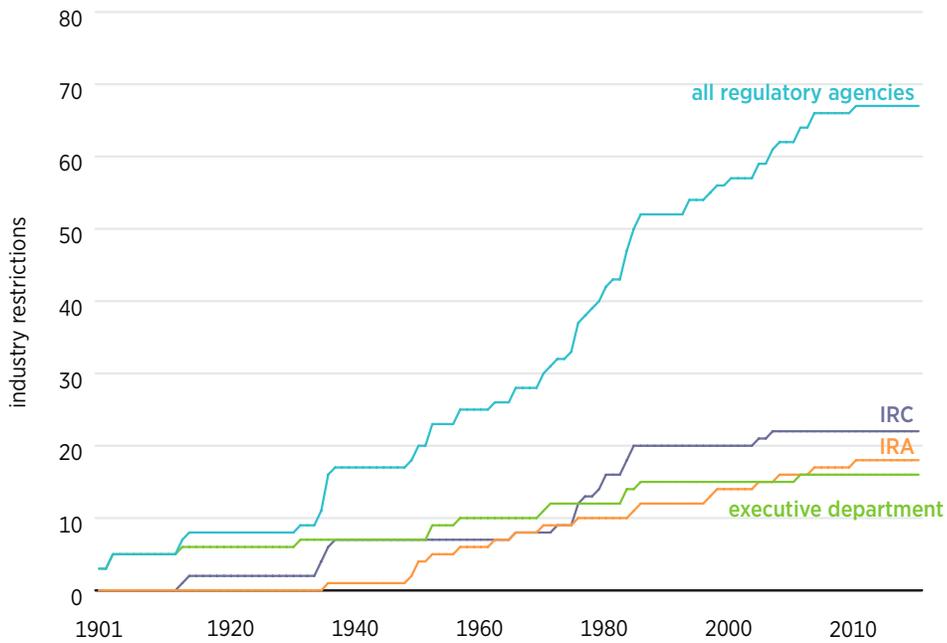
Table 5 summarizes the information about the different types of agencies, including the average and total number of regulatory restrictions issued as of 2019. In addition, table 5 shows the number of agencies in each type of agency, the year the first type of that agency was created, the most recent year in which that type of agency was created, the number of agencies that are official rulemaking agencies,

FIGURE 10. RESTRICTIONS ON CANADIAN TRANSPORTATION INDUSTRIES



Note: 481 = Air Transportation; 482 = Rail Transportation; 483 = Water Transportation; 486 = Pipeline Transportation; 488 = Support Activities for Transportation.

FIGURE 11. NUMBER OF REGULATORY AGENCIES



Note: IRA = independent regulatory agency; IRC = independent regulatory commission.

TABLE 5. SUMMARY OF RESTRICTIONS BY TYPE OF AGENCY

Agency type	Number of agencies	First created	Latest created	Number of rulemakings	Average number of leaders	Average number of restrictions	Total number of restrictions
Corporation	13	1789	2003	10	8	709	8,510
Executive department	16	1789	2002	15	1	42,464	679,420
Independent regulatory agency (IRA)	18	1934	2010	18	1	13,007	234,127
Independent regulatory commission (IRC)	22	1913	1998	22	5	5,760	126,727
IRC—nonregulatory	11	1934	2004	9	6	386	4,241
Other nonregulatory	64	1846	2013	18	8	567	19,847
Regional commission/ agency	6	1961	2012	0	8	263	1,577

Sources: Patrick A. McLaughlin et al., State RegData 2.0 (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2020, <https://quantgov.org/state-regdata/>; Jennifer L. Selin and David E. Lewis, Sourcebook of United States Executive Agencies, 2nd ed. (Administrative Conference of the United States, 2018).

and the average number of leaders for the agency. This analysis includes the Environmental Protection Agency in the list of executive departments because it has traditionally been treated as such by past administrations.

Figure 12 shows the accumulation of regulatory restrictions by agency type. Executive departments, as expected, have accumulated the most restrictions, followed by independent regulatory agencies. In addition, the independent regulatory commissions have accumulated regulations at a slower pace than have executive departments and independent regulatory agencies. The differences in the regulatory output of independent regulatory agencies on one hand and independent regulatory commissions on the other require further examination (addressed in another working paper). Among executive departments (figure 13), the Department of the Treasury has accumulated the most regulatory restrictions over the period.

## STATE/PROVINCE REGDATA

In the United States, as of 2020, California had the most regulatory restrictions, followed by New York, Ohio, Illinois, and Texas. Each of these states has more regulatory restrictions than do Australia and Canada. In fact, only 10 US states have fewer regulatory restrictions than Canada has federal regulatory restrictions. Among Canadian provinces, Ontario has the most regulatory restrictions, followed by Quebec. Similarly, for Australia, New South Wales has the most regulatory restrictions. Figure 14 shows the total number of regulatory restrictions

FIGURE 12. REGULATORY ACCUMULATION BY TYPE OF AGENCY

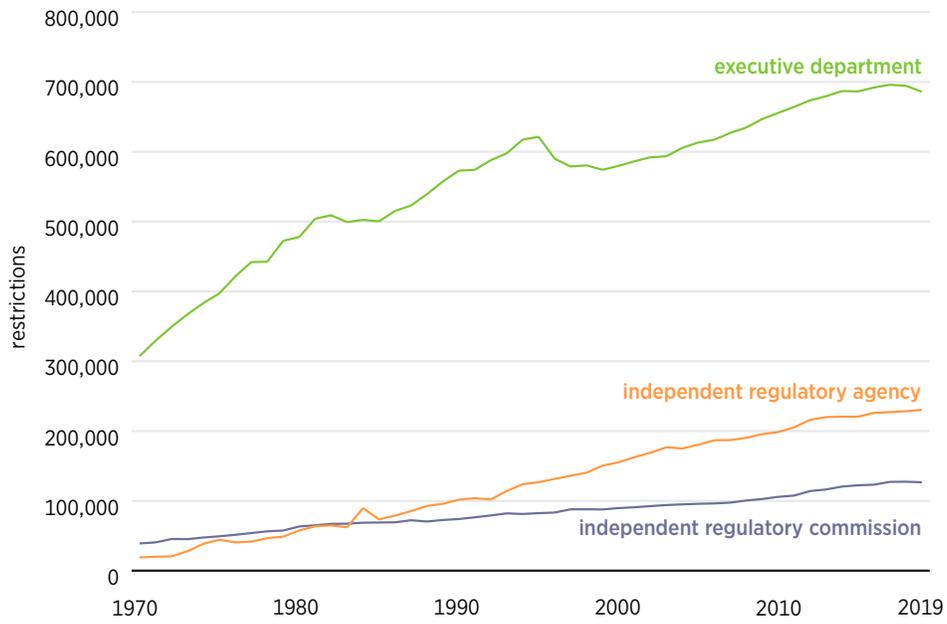


FIGURE 13. ACCUMULATION OF REGULATORY RESTRICTIONS BY SELECTED EXECUTIVE DEPARTMENTS

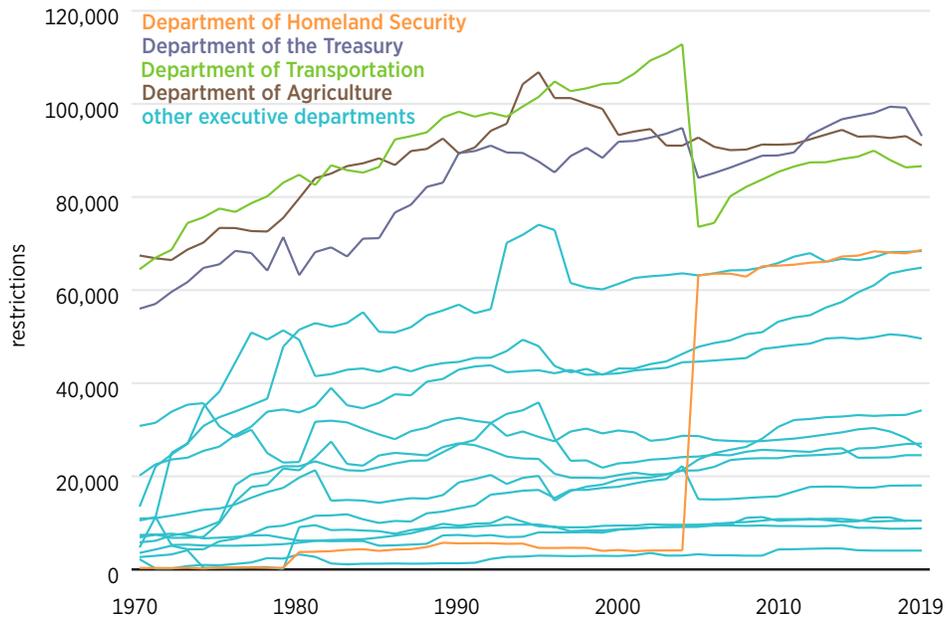
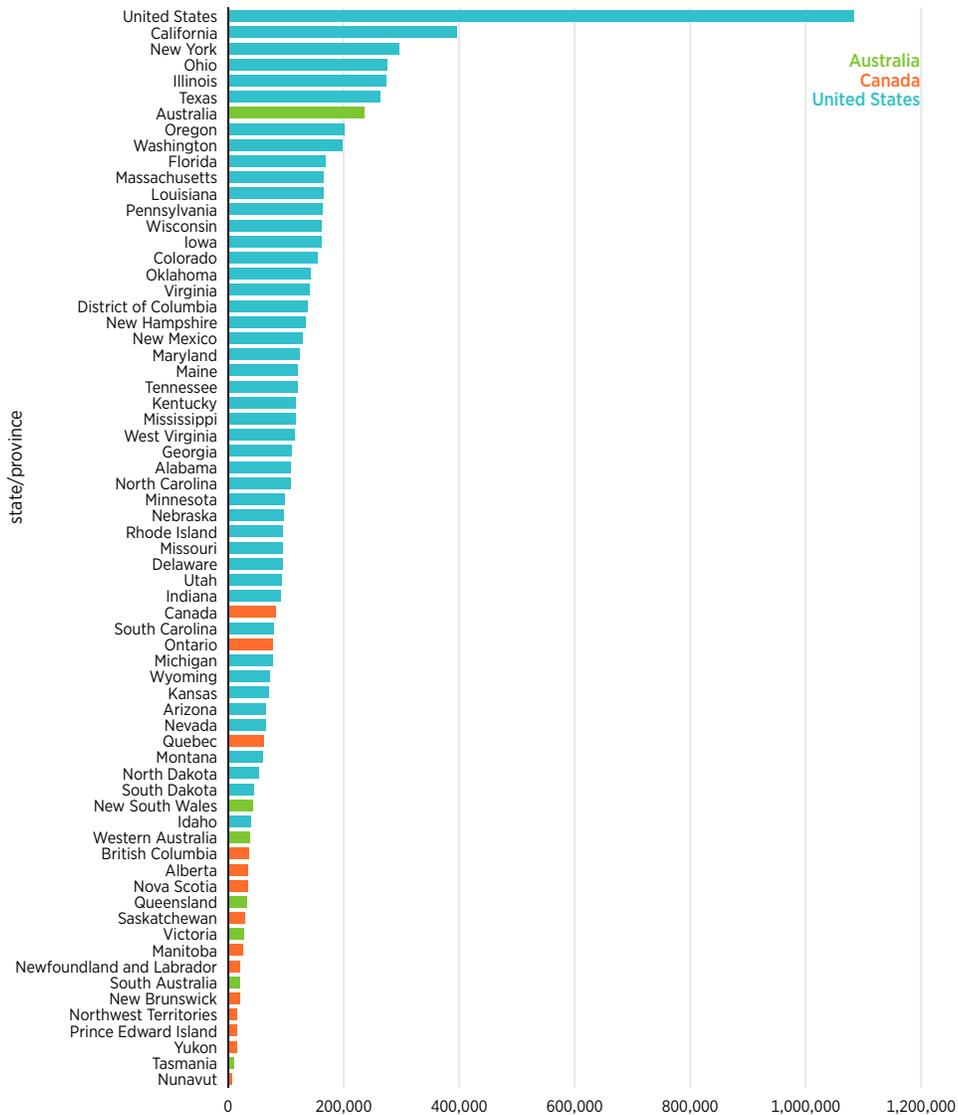


FIGURE 14. STATE/PROVINCIAL REGULATORY RESTRICTIONS

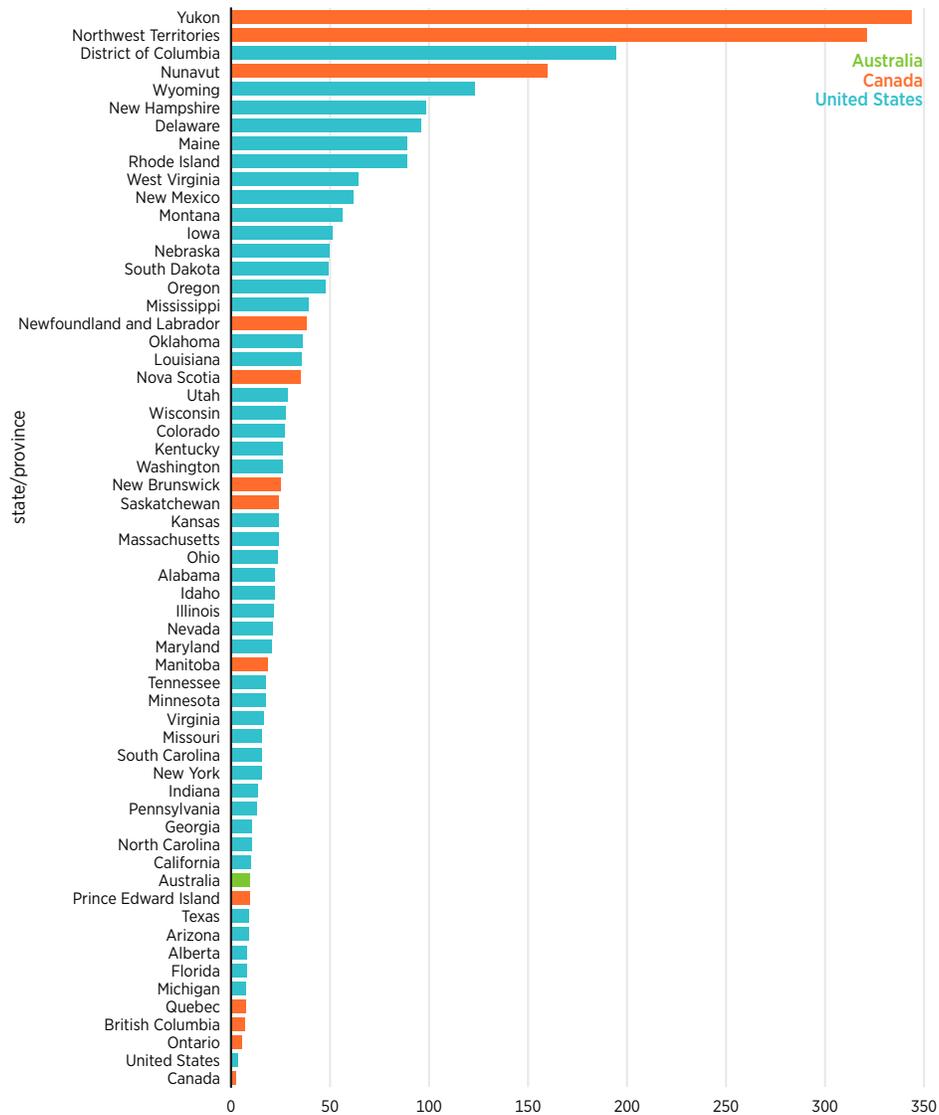


by state or province. What is common across all three countries is that the most regulated states are also major population centers. Figure 15 therefore accounts for the population in each state or province.

## Industry Characteristics

**Combined State and Federal Restrictions.** Being federal jurisdictions, industries are regulated by both state and federal agencies. This paper examines

FIGURE 15. RESTRICTIONS PER CAPITA



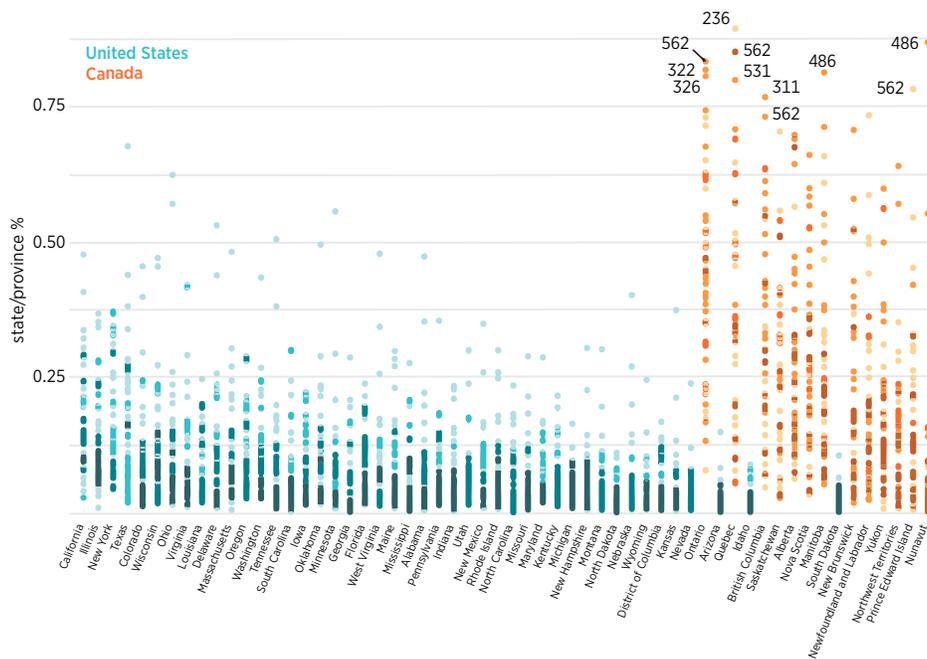
the combined number of restrictions per state, adding the number of federal land state restrictions. Table 6 and figure 16 show the proportion of industry-relevant restrictions imposed by each state. Table 6 shows the summary for all states or provinces in the country. The proportion of state restrictions on an industry is computed for each industry as the number of industry-relevant restrictions imposed by the state divided by the sum of industry-relevant restrictions imposed by both the state (or provincial) government and the federal government. Across all industries, Canadian provincial regulations, on average,

TABLE 6. STATE/PROVINCE SHARE OF TOTAL INDUSTRY RESTRICTIONS

Country	Number of states (in RegData)	Average state proportion	Median state proportion	Number of industries with state share > 25%	Number of industries (NAICS 3-digit)
Canada	13	0.25	0.19	236	46
United States	46	0.08	0.05	88	46

Source: Patrick A. McLaughlin et al., State RegData 2.0 (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2020, <https://quantgov.org/state-regdata/>.

FIGURE 16. SHARE OF TOTAL INDUSTRY REGULATORY RESTRICTIONS IMPOSED BY STATES

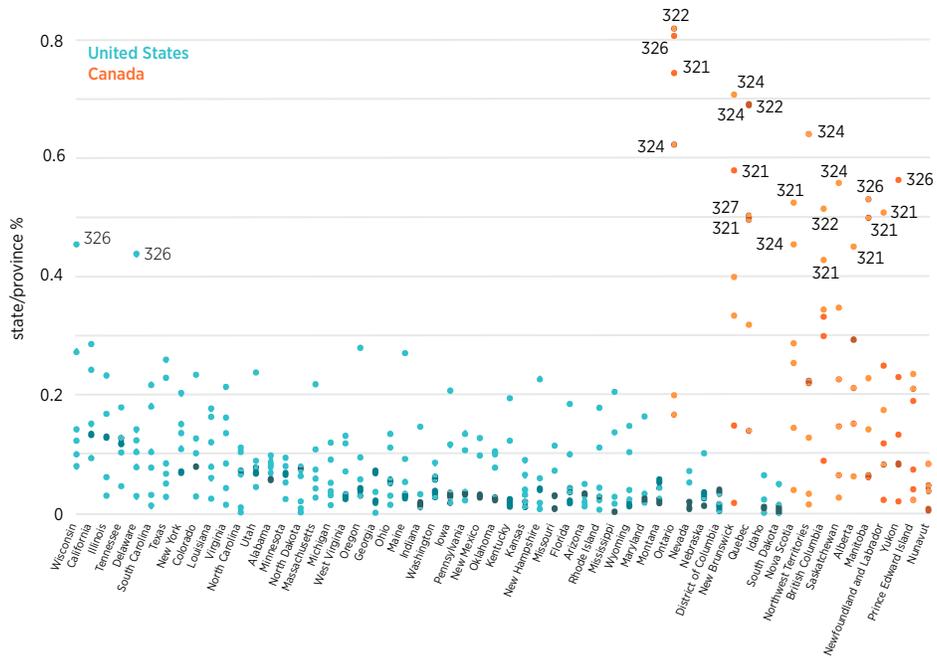


Note: 236 = Construction of Buildings; 311 = Food Manufacturing; 322 = Paper Manufacturing; 326 = Plastics and Rubber Products Manufacturing; 486 = Pipeline Transportation; 531 = Real Estate; 562 = Waste Management and Remediation Services.

constitute one-quarter of all restrictions. In contrast, the average for US states is 8 percent, and across all 46 states (and Washington, DC) with data, only 88 (out of the potential 2,070 industries) have state share of industry restrictions exceeding 25 percent compared with 236 (of 598 industries) for Canada.

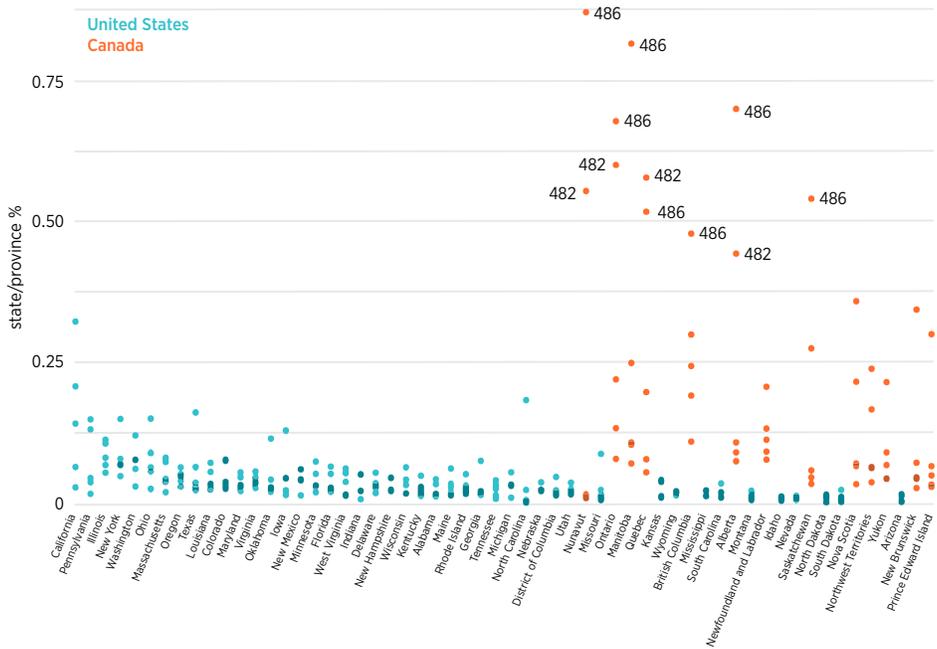
The next set of charts (figures 17–20) show the same information but for a subset of industries, namely manufacturing (NAICS major group 32), transportation (NAICS 48), healthcare and social assistance (NAICS 62), and agriculture (NAICS 11). The reader is encouraged to examine other industries on the QuantGov website. These four charts clearly show that Canada in general imposes

FIGURE 17. STATE/PROVINCE SHARE OF INDUSTRY RESTRICTIONS: MANUFACTURING



Note: 321 = Wood Product Manufacturing; 322 = Paper Manufacturing; 324 = Petroleum and Coal Products Manufacturing; 326 = Plastics and Rubber Products Manufacturing.

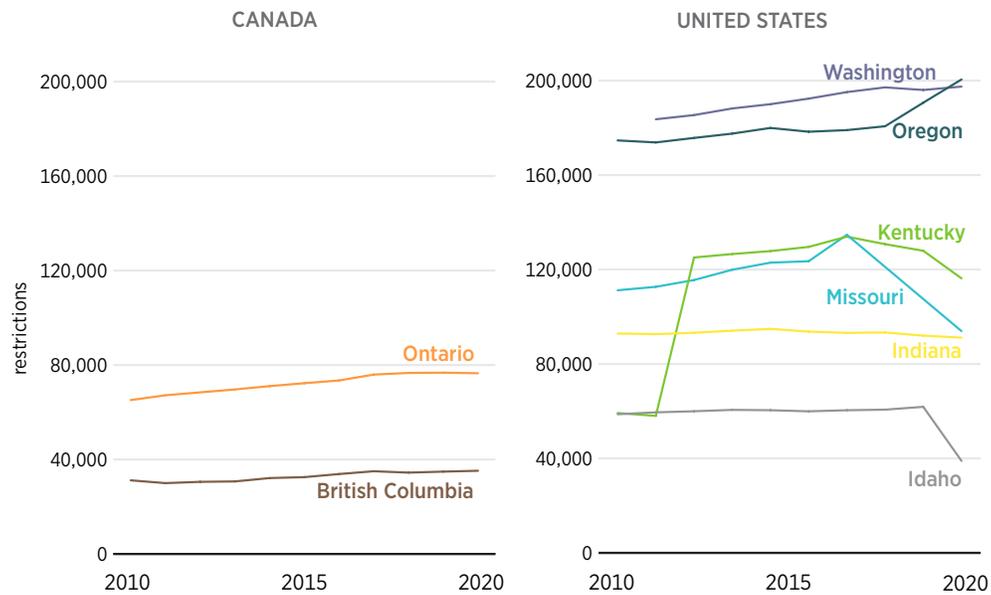
FIGURE 18. STATE/PROVINCE SHARE OF INDUSTRY RESTRICTIONS: TRANSPORTATION



Note: 482 = Rail Transportation; 486 = Pipeline Transportation.



FIGURE 21. EVOLUTION OF REGULATORY RESTRICTIONS IN SELECTED US STATES AND CANADIAN PROVINCES



more regulatory restrictions at the provincial level than do American states, a reflection of the differences between the US and Canadian constitutions in terms of the relationship between the federal and state governments.

### Evolution of Regulatory Restrictions in Selected US States and Canadian Provinces

For a few states and provinces, we can track the evolution of regulatory restrictions because their respective administrative codes are published annually in electronic format (see table 1). Those states and provinces are Idaho, Indiana, Kentucky, Missouri, Oregon, and Washington in the United States and Ontario and British Columbia in Canada. Tracking that evolution affords a unique opportunity to examine the changes in the stock of regulatory restrictions. Among those states, Idaho has the most historical data—dating back to 1995—and Oregon has the most restrictions, followed by Washington and Kentucky. All the analyses in this paper could be replicated for those states and provinces.

Figure 21 shows the evolution of regulatory restrictions in those states and provinces. As with federal regulatory restrictions, states and provinces tend to accumulate regulatory restrictions; however, Idaho, Kentucky, and Missouri

have reduced the number of regulatory restrictions over the past few years. All three states embarked upon red-tape reduction over the same period.

## CONCLUSIONS AND FURTHER RESEARCH

This research paper provided an overview of regulatory restrictions in three countries—Australia, Canada, and the United States—both at the federal and state (or province) levels. In addition, the authors showed how other indicators could be combined with RegData for analysis. Later research will attempt to establish causality, if any, in any of the correlations described. Regulations in the three countries take quite different tracks and are influenced by the legislative process. Whereas Canadian restrictions fluctuate, in the United States and Australia, they only increase year over year. Also, in the United States, a few agencies dominate in terms of restrictions issued. Major events such as financial crises, court rulings, and new laws that arise can cause dramatic changes in the trajectory of regulations for an industry or from an agency.

Several issues require further detailed research. First, the authors intend to dig deeper into the agency structure and how that affects the number and nature of regulations. Second, further exploration is needed into the relationships between regulations and other economic indicators. The results shown in this brief are contemporaneous even though the effects of restrictions may not be immediate. In addition, these relationships often are dynamic, requiring more sophisticated approaches to prove any relationships. Third, it is important to examine the combined effects of state and federal restrictions on economic outcomes; this topic was mentioned briefly, but it requires serious consideration. Most important, RegCensus can be used to examine current changes in regulations across the United States and Canada—both federal and state or provincial levels—and how those reforms are affecting, first, the stock of restrictions and second, the broader economy.

## REFERENCES

- Al-Ubaydli, Omar, and Patrick A. McLaughlin. 2017. “RegData: A Numerical Database on Industry-Specific Regulations for All United States Industries and Federal Regulations, 1997–2012.” *Regulation & Governance* 11, no. 1: 109–23.
- Katz, Daniel Martin, and Michael J. Bommarito II. 2014. “Measuring the Complexity of the Law: The United States Code.” *Artificial Intelligence and Law* 22, no. 4 (December 2014): 337–74.
- McLaughlin, Patrick A., and Oliver Sherouse. 2019. “RegData 2.2: A Panel Dataset on US Federal Regulations.” *Public Choice* 180: 43–55.

## APPENDIX A: VERSION HISTORY OF REGDATA

RegData has evolved over time and is now subsumed under the QuantGov platform. To date, five different iterations of the RegData database have been released.

### Version 1.0 (1997–2010)

Released in 2012, RegData 1.0 introduced the restrictions metric—the method of measuring regulation by counting words such as *shall* and *must* within regulatory text. It also introduced the idea of creating industry-specific measures of regulation, which were based on a human-assisted search algorithm, which involved creating a set of search terms or keywords based on the descriptions of specific industries in the North American Industry Classification System. The data in version 1.0 covered the years 1997–2010 and included two- and three-digit NAICS-coded industries.

### Version 2.0 (1997–2012)

RegData 2.0 provided the ability to quantify the regulations that specific federal regulators (including agencies, offices, bureaus, commissions, or administrations) have produced. For example, with version 2.0, a user could see how many restrictions a specific administration of the Department of Transportation (for example, the National Highway Traffic Safety Administration) has produced each year. It also added the years 2011 and 2012 to the database. Version 2.0 was bundled with a new dataset that calculated the probabilities of specific industry

### Version 2.1 (1975–2013)

RegData 2.1 introduced machine learning algorithms to the project. Whereas versions 1.0 and 2.0 had relied upon search terms devised using a scheme initially conceptualized by McLaughlin and Al-Ubaydli that created permutations of individual industry descriptions, the algorithms used in RegData 2.1 did not require humans to tell the program what specific words or phrases to search for. Instead, the authors found thousands of documents that they knew related to specific industries and used those documents to train the programs. The programs parsed the training documents and identified which words and phrases were used about specific industries. This enhancement permitted industry-specific classification of regulation to be much more accurate, primarily by reducing the number of false positives.

RegData 2.1 added several more years' data, covering 1975 to 2013, and included three-digit NAICS-coded industries. It also introduced the public law database (PLDB), which mapped specific regulations to their authorizing statutes from 1980 to 2013.

### Version 2.2 (1975–2014)

RegData 2.2 included significant refinements in the machine learning algorithm used to classify regulations by industry. The authors also expanded the machine learning-based dataset to include two- and four-digit NAICS-coded industries and added the year 2014 to both the regulations data and the PLDB so that RegData 2.2 covers 1975–2014 and the PLDB covers 1980–2014.

### Version 3.0 (1970–2016)

RegData 3.0 for the first time covers all levels of the NAICS standard, from two to six digits. This version also broadens the scope of the dataset back to 1970 and forward to 2016. In addition, the machine learning model has been further improved to better identify related industries that are regulated as a group. The PLDB has been separated out from the main release to allow for different updating schedules.

### Version 3.1 (1970–2017)

RegData 3.1 expanded the dataset an additional year to 2017, using the XML representation of the Electronic CFR. The eCFR has been annualized by choosing

the last-modified version of each title of that title’s publication date in the annual CFR publication cycle. In addition, RegData 3.1 changed the way agency names are parsed during the creation of the CFR corpus, improving both the accuracy and the speed of the creation of the corpus.

### Version 3.2 (1970–2019)

RegData 3.2 expanded the main RegData dataset two more years to 2019. In addition, RegData 3.2 included metrics for determining the complexity of regulatory text, including number of conditional terms or phrases, average sentence length, reading grade level, and Shannon Entropy.

### Version 4.0 (1970–2020)—Released May 2021

RegData 4.0 introduced a new method of counting regulatory restrictions to include obligations or prohibitions previously hidden in lists or bullet points. In addition, like other iterations, RegData 4.0 expanded the dataset an additional year to cover the years 1970 to 2020. RegData 4.0 also added some accuracy improvements by cleaning older text and better parsing agency data.

## APPENDIX B: AVAILABLE DATA ELEMENTS

The data elements listed in the table below are the different measures included in RegData and RegCensus.

Data element (series)	Description
Total: Restrictions analysis	Total number of restrictions (sum of the number of times the words <shall, must, may not, required, prohibited> show up in the regulatory text) imposed by regulations for the period in the jurisdiction
Total: Words analysis	Total number of words found in the regulatory code for the period in the jurisdiction
Total: Terms “shall” analysis	Total number of times the word <shall> occurs in the regulatory code for the period in the jurisdiction
Total: Terms “must” analysis	Total number of times the word <must> occurs in the regulatory code for the period in the jurisdiction
Total: Terms “may not” analysis	Total number of times the phrase <may not> occurs in the regulatory code for the period in the jurisdiction
Total: Terms “prohibited” analysis	Total number of times the word <prohibited> occurs in the regulatory code for the period in the jurisdiction
Total: Terms “required” analysis	Total number of times the word <required> occurs in the regulatory code for the period in the jurisdiction
Total: Restrictions by industry probability analysis	Total number of restrictions multiplied by the probability of the text being related to a specific industry
Total: Restrictions by occupational licensing probability analysis	Total occupational licensing relevant restrictions
Total: Deregulatory terms analysis	Total number of deregulatory terms that show up in the regulatory text imposed by regulations for the period in the jurisdiction
Total: Restrictions by healthcare probability analysis	Total occupational licensing relevant restrictions
Total: Restrictions by agency analysis	Total number of restrictions (sum of the number of times the words <shall, must, may not, required, prohibited> show up in the regulatory text) imposed by an agency for the period in the jurisdiction
Total: Restrictions by agency and industry analysis	Total number of industry-relevant restrictions imposed by an agency for the period in the jurisdiction
Probability: Occupational licensing analysis	Probability of text containing an occupation license
Probability: Occupational licensing by two-digit SOC analysis	Probability of text containing an occupational license of a specific occupation
Probability: Healthcare analysis	Probability of text being related to healthcare
Probability: Industry analysis	Probability of text being related to an industry
Complexity: Sentence length analysis	Average length of a sentence in a unit of regulation
Complexity: Shannon’s entropy analysis	Measures complexity by counting how much new information is introduced
Complexity: Conditionals analysis	Counts the number of “branching words,” such as “if,” “but,” and “provided” that identify logical branches in a document
Complexity: Readability consensus score analysis	Summary of all readability scores
Complexity: Sentence length by agency analysis	Average length of a sentence in a unit of regulation organized by agency
Complexity: Shannon’s entropy by agency analysis	Measures complexity of a unit of regulation organized by agency by counting how much new information is introduced

Data element (series)	Description
Complexity: Conditionals by agency analysis	Counts the number of "branching words," such as "if," "but," and "provided," which identify logical branches in a document organized by agency
Period: Significant revisions: restrictions analysis	Last time the total number of restrictions changed
Period: Significant revisions: words analysis	Last time significant revisions were made to the code; revision is considered significant if the total number of words in a year changes 2.1% or more
Period: Last updated analysis	Measures the recency of revisions to documents

Source: Patrick A. McLaughlin et al., State RegData 2.0 (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2020, <https://quantgov.org/state-regdata/>.

## ABOUT THE AUTHORS

**Kofi Ampaabeng** is a senior research fellow at the Mercatus Center at George Mason University. He is also a data scientist, specializing in the application of natural language processing methods to quantify essential features of regulatory text.

**Patrick A. McLaughlin** is the director of policy analytics and a senior research fellow at the Mercatus Center at George Mason University. He created and leads Mercatus Center's RegData and QuantGov projects, which deploy machine learning and other tools of data science to quantify regulation.

**Thurston Powers** is a database engineer at the Mercatus Center at George Mason University. He specializes in database management and data visualization and communications.

## ABOUT THE MERCATUS CENTER AT GEORGE MASON UNIVERSITY

The Mercatus Center at George Mason University is the world's premier university source for market-oriented ideas—bridging the gap between academic ideas and real-world problems.

A university-based research center, Mercatus advances knowledge about how markets work to improve people's lives by training graduate students, conducting research, and applying economics to offer solutions to society's most pressing problems.

Our mission is to generate knowledge and understanding of the institutions that affect the freedom to prosper and to find sustainable solutions that overcome the barriers preventing individuals from living free, prosperous, and peaceful lives.

Founded in 1980, the Mercatus Center is located on George Mason University's Arlington and Fairfax campuses.