

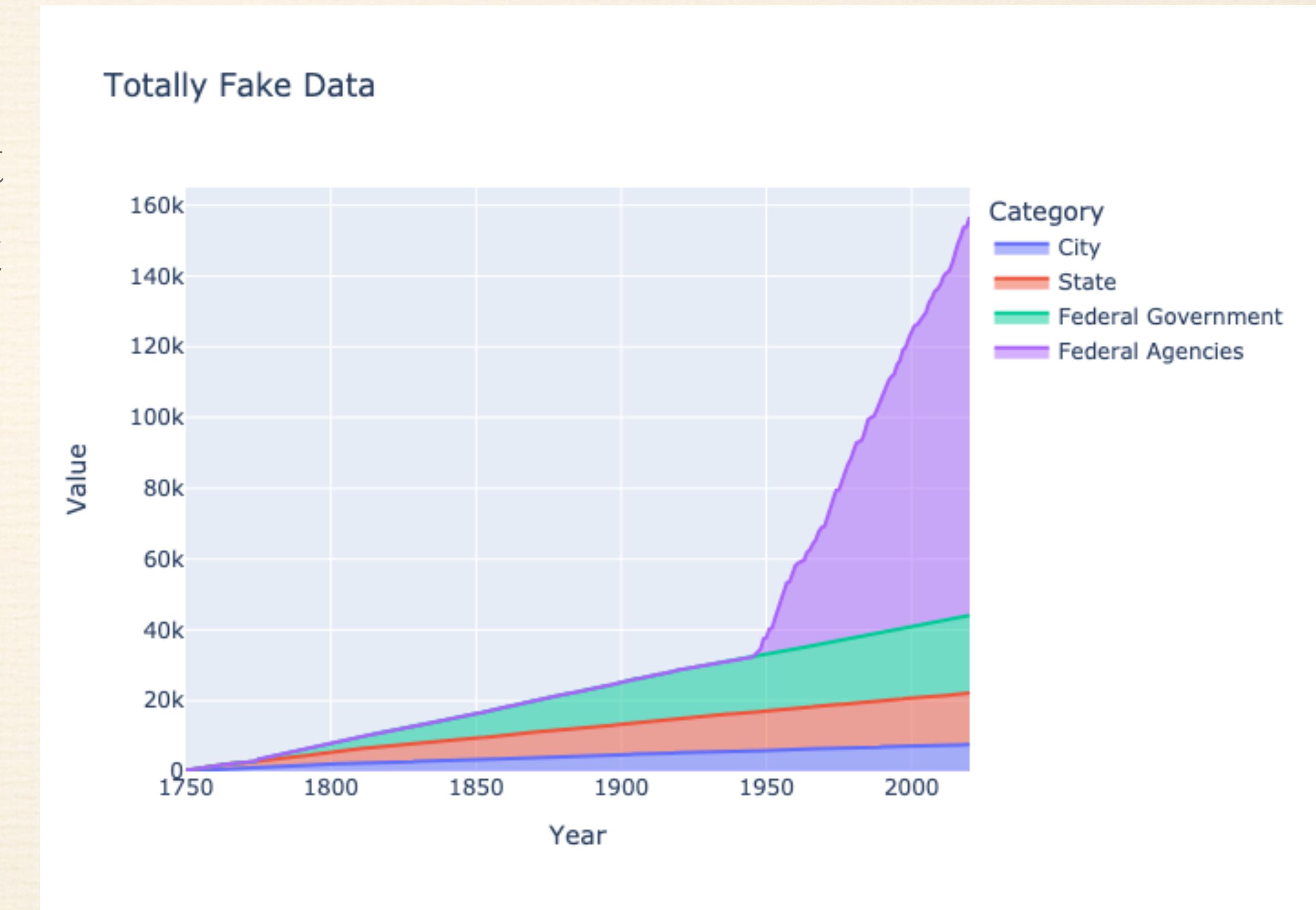
The Weight of Legislation



Jed and David Madsen

Our Initial Goal

- ❖ Understanding the total amount of law produced by the different levels of government in the US.
- ❖ Data not available. Not even close.
- ❖ Need a grant to gather data.



LegiScan

- ❖ Site that contains datasets for legislative sessions by state as well as for Congress since 2009
- ❖ Organized datasets that contain information regarding bills, their status, as well as document-size in bytes*
- ❖ All bills have a unique bill ID associated with them, which made it easier for us to perform analysis

The screenshot shows the LegiScan homepage with a dark blue header containing links for Register, About, Features, Datasets, LegiScan API, Bill Tracking, Search, VOTE!, and Log In. The main content area features the LegiScan logo and the tagline "BRINGING PEOPLE TO THE PROCESS". On the left, there's a sidebar with "LegiScan Search" and "LegiScan Info" sections. The "LegiScan Info" section lists various links such as 2024 Schedules, Price List, LegiScan API, Weekly Datasets, Documentation, Search Help, Voter Registration, Governor Deadlines & Effective Dates, Legislation 101, On Bill Numbers, and State Support. The central part of the page is titled "National Legislative Datasets" and includes a note: "... Please note you must be logged in to download the datasets. Registration is free. [Signup here.](#)". Below this is a grid of buttons for states: 2024, AK, AL, AR, AZ, CA, CO, CT, DE, FL, GA, HI, IA, ID, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NJ, NM, NV, NY, OH, OK, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WA, WI, WV, WY, DC, US. A note below the buttons says: "Weekly snapshots of session data are created each Sunday morning with updated information on an as-needed basis. Provided in simple comma-separated values files for general bill data, or the most complete form packaged as LegiScan API JSON payloads. Current year data is below, for a complete historical list of session files select an individual state above." It also includes links for Inquire, LegiScan API Info, LegiScan API Client Source, LegiScan API Client Docs, and LegiScan API Manual. At the bottom, it says "Legislative Datasets by LegiScan LLC is licensed under CC BY 4.0" and provides links for JSON and CSV files for various states and years, along with their file sizes.

State	Year	Session	Modified	Exported	API JSON	CSV Basic
Alabama	2024	Regular Session	2024-03-21	2024-03-24	JSON	CSV 1.4 MB 244 KB
Alaska	2023-2024	33rd Legislature	2024-03-24	2024-03-24	JSON	CSV 1.6 MB 226 KB
Arizona	2024	Fifty-sixth Legislature 2nd Regular	2024-03-22	2024-03-24	JSON	CSV 5.7 MB 553 KB
California	2023-2024	Regular Session	2024-03-22	2024-03-24	JSON	CSV 16.1 MB 2.3 MB
Colorado	2024	Regular Session	2024-03-22	2024-03-24	JSON	CSV 2.5 MB 321 KB

*We thought we'd be able to use this particular field to determine bill length. We were wrong...

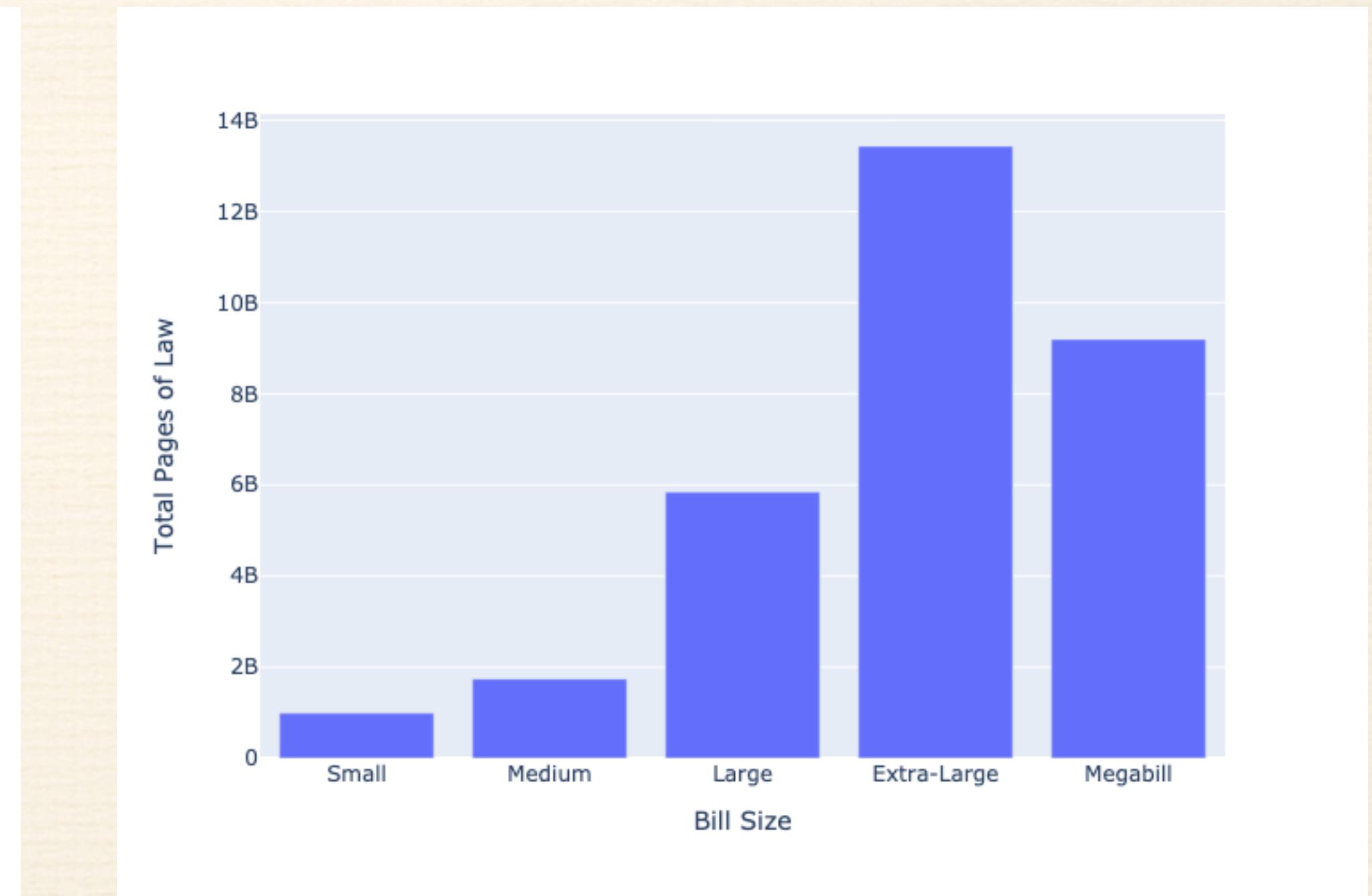
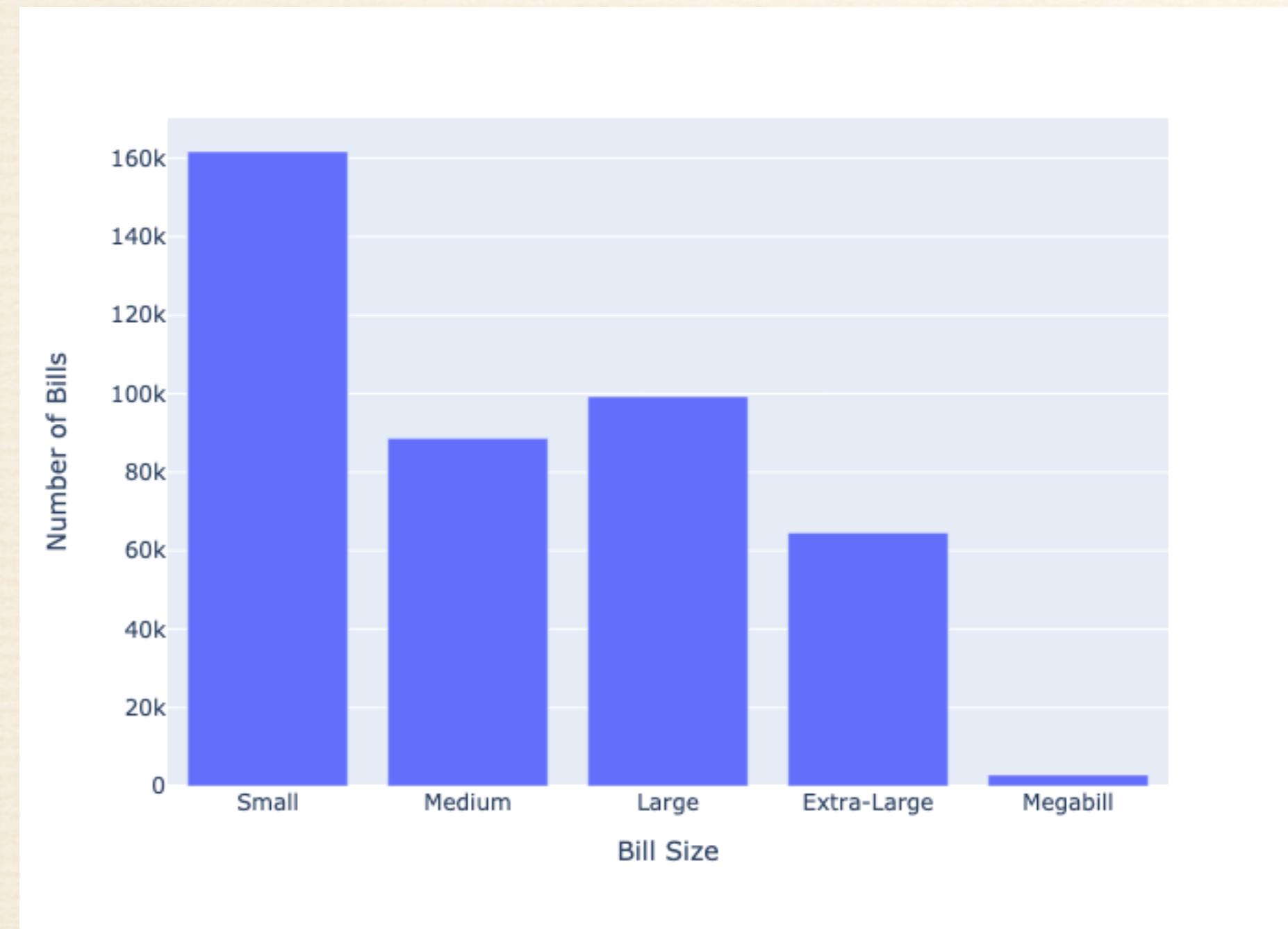
Initial Pass

❖ Document Size in bytes

	title	last_action_date	State	document_id	document_size	dc
	Metro Township Modifications	2024-03-11	UT	2956555	249303177	
	Metro Township Modifications	2024-03-11	UT	2956556	249303177	
	Court Amendments	2024-03-13	UT	2957600	222307334	
	Boards and					

Initial Pass

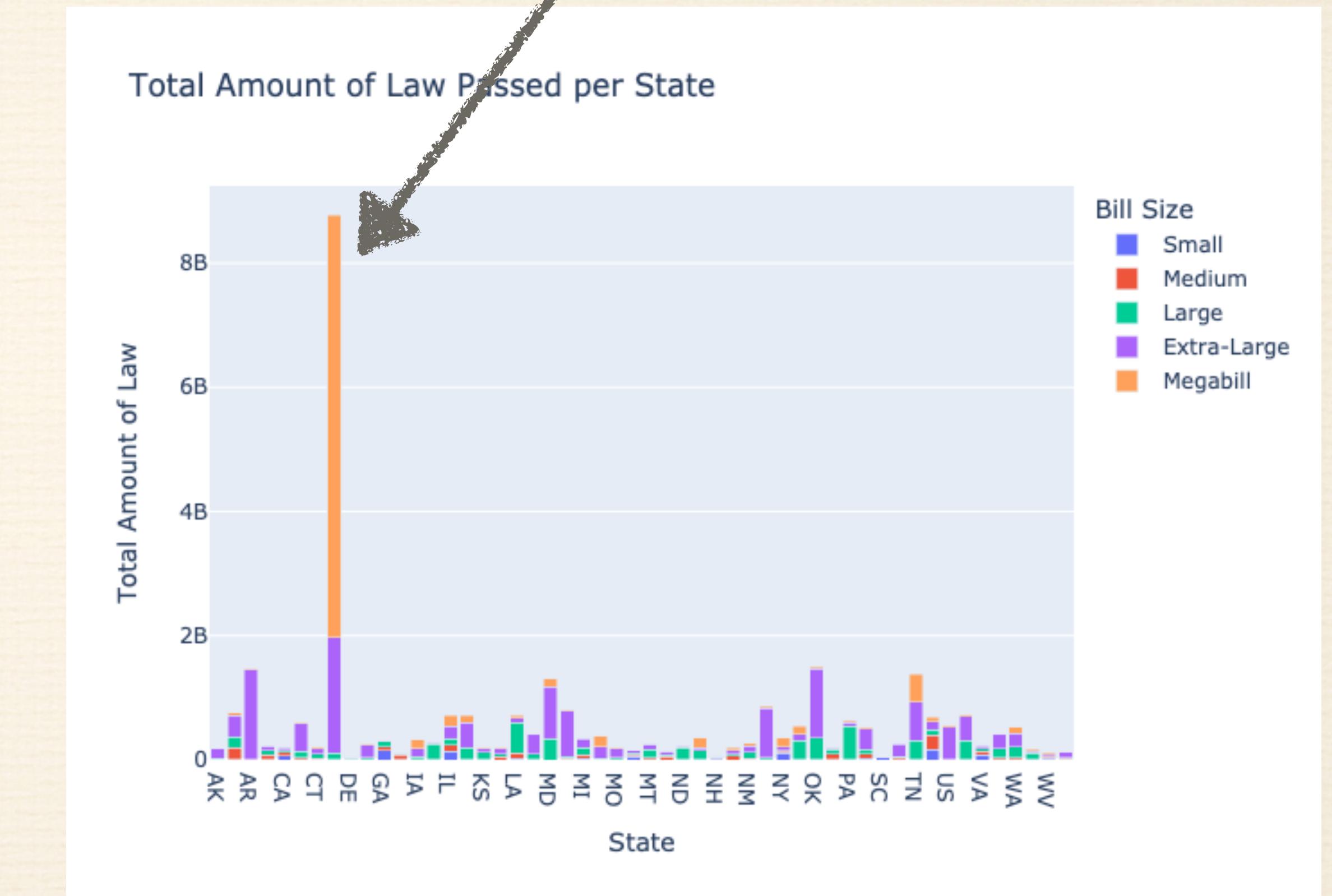
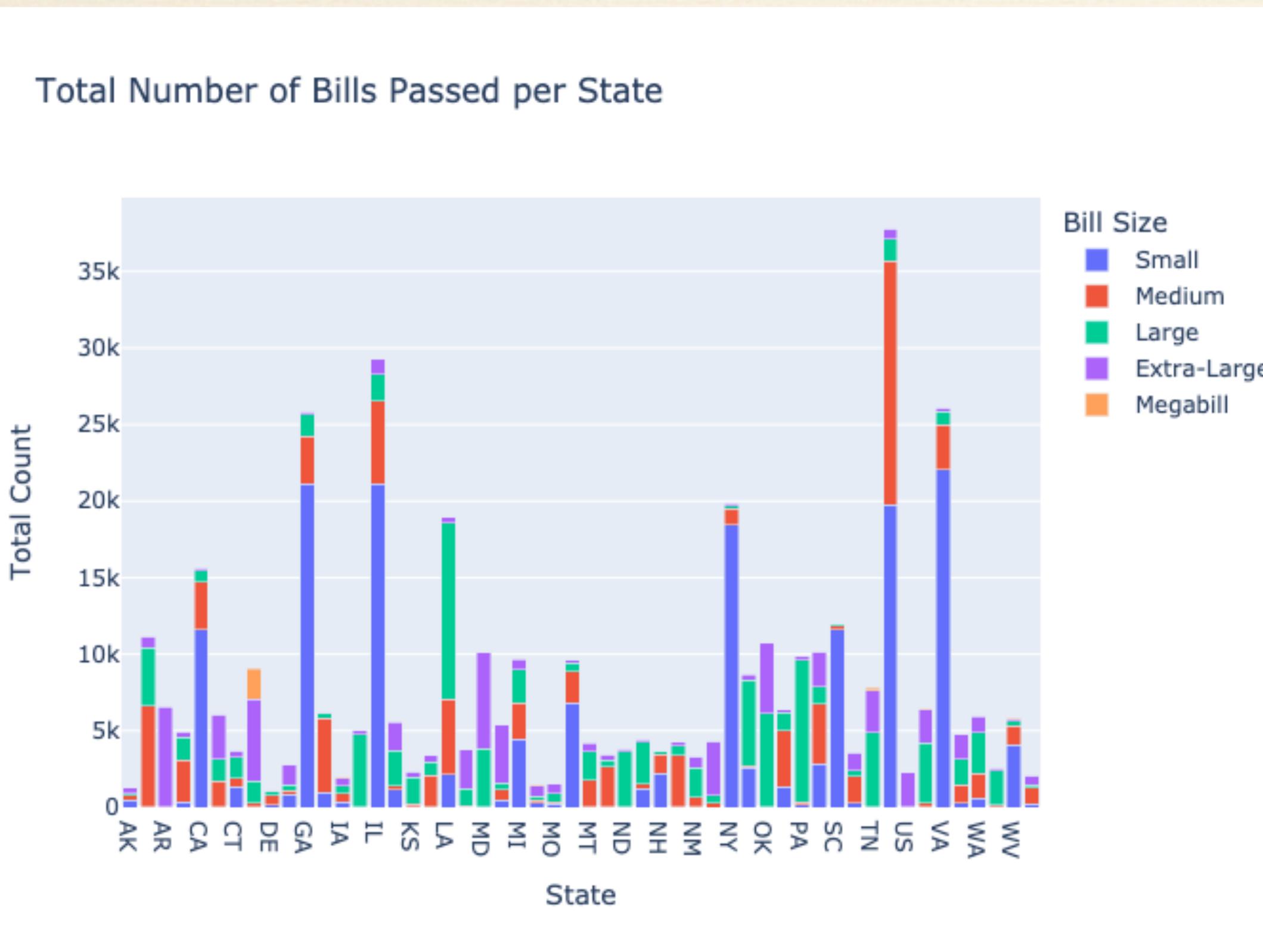
- ❖ Number of bills vs. amount of law



Initial Pass

❖ Uh-oh...

What is this?



The Great PDF Download

- ❖ Created a process to download the PDFs, count the pages and add the page count to the DataFrame



Unlimited power!

```
# Function to download PDF, get number of pages, and update DataFrame
def process_pdf(row):
    url = row['url']
    filename = os.path.join('pdfs', f'{row.name}.pdf')
    try:
        # Download PDF file
        response = requests.get(url, stream=True, verify=False)
        with open(filename, 'wb') as f:
            for chunk in response.iter_content(chunk_size=1024):
                f.write(chunk)

        # Get number of pages using PdfReader from PyPDF2
        pdf = PdfReader(filename)
        num_pages = len(pdf.pages)

        # Update 'pages' column in DataFrame
        bill_id_uniques.at[row.name, 'pages'] = num_pages

        # Delete downloaded PDF file
    except:
        os.remove(filename)
    except Exception as e:
        print("can't remove file")
    except Exception as e:
        # Write 'NA' if there is an error
        bill_id_uniques.at[row.name, 'pages'] = 'NA'
        print(f"Error processing {url}: {e}")
    try:
        os.remove(filename)
    except:
        print(f"can't remove file")

# Assuming bill_id_uniques is your pandas DataFrame containing 'url' and 'document_size' columns
# Iterate through rows in the DataFrame and process each PDF
for idx, row in two_more_df.iterrows():
    process_pdf(row)

# Write updated DataFrame to pickle file every 10000 rows
if idx % 10000 == 0:
    two_more_df.to_pickle('bill_id_uniques2.pkl', protocol=4)
```

The Great PDF Download

- ❖ Oops, lots of incorrect URLs, missing docs, bad certs etc.



Somewhat limited power

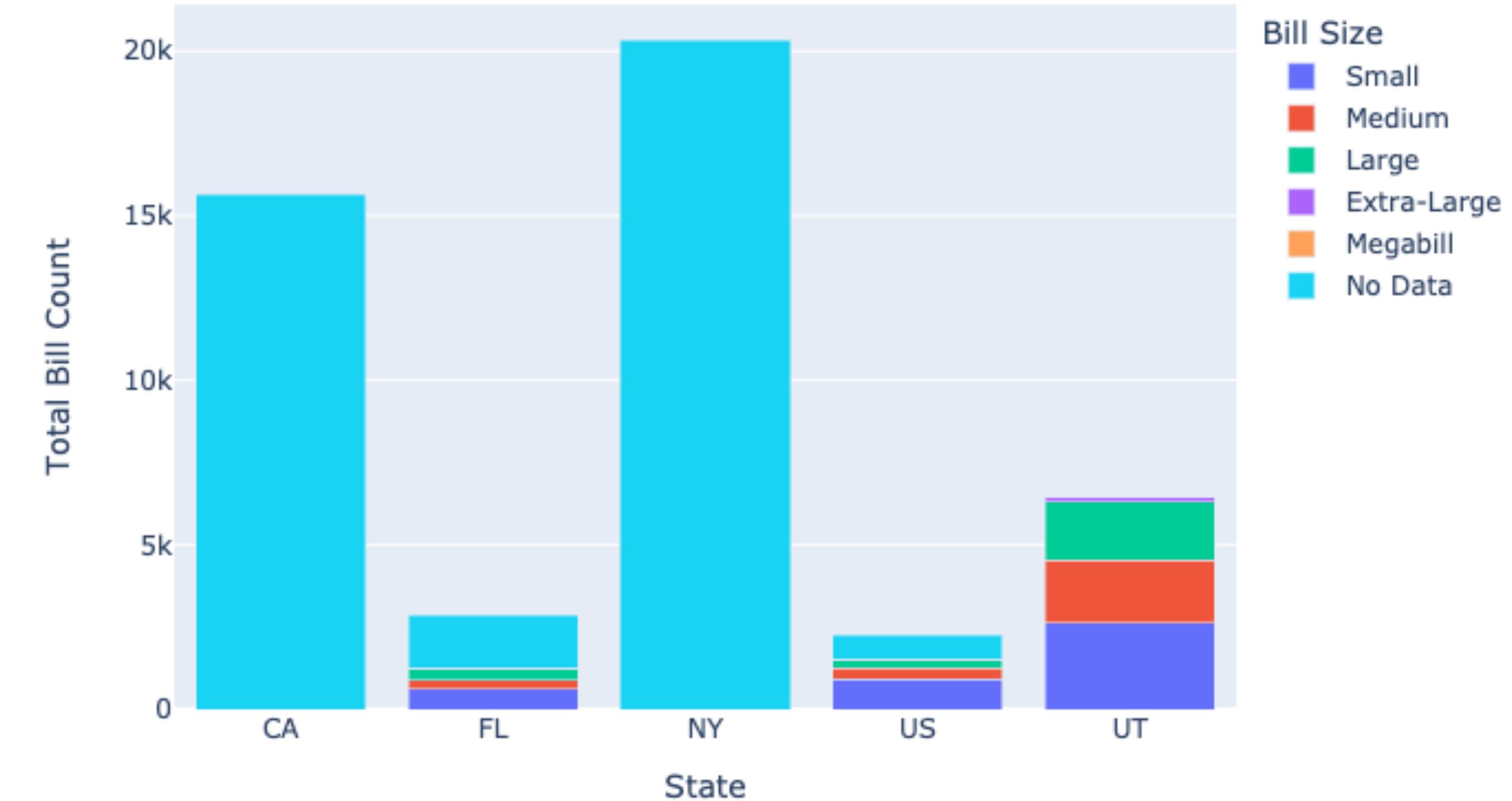
```
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05167-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>  
can't remove file  
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05168-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>  
can't remove file  
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05169-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>  
can't remove file  
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05170-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>  
can't remove file  
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05172-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>  
can't remove file  
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05175-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>  
can't remove file  
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05200-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>  
can't remove file  
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05201-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>  
can't remove file  
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05202-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>  
can't remove file  
Error processing https://www.cga.ct.gov/2022/T0B/H/PDF/2022HB-05205-R00-HB.PDF: <urlopen error [SSL: CERTIFICATE_VERIFY_FAILED] certificate verify failed: unable to get local issuer certificate (_ssl.c: 1007)>
```

Real-ish Data

- ❖ We lost much
- ❖ So much blue

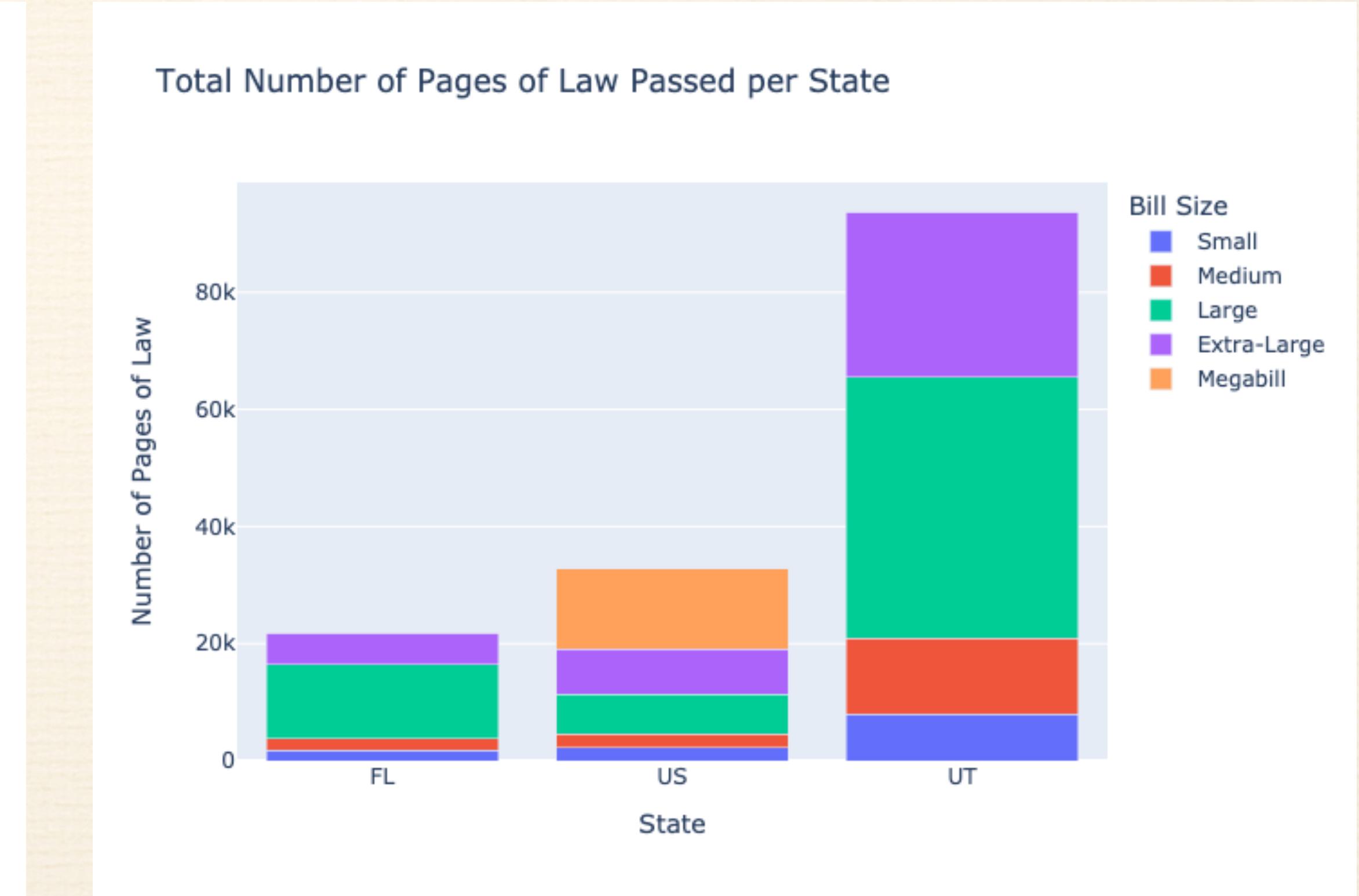
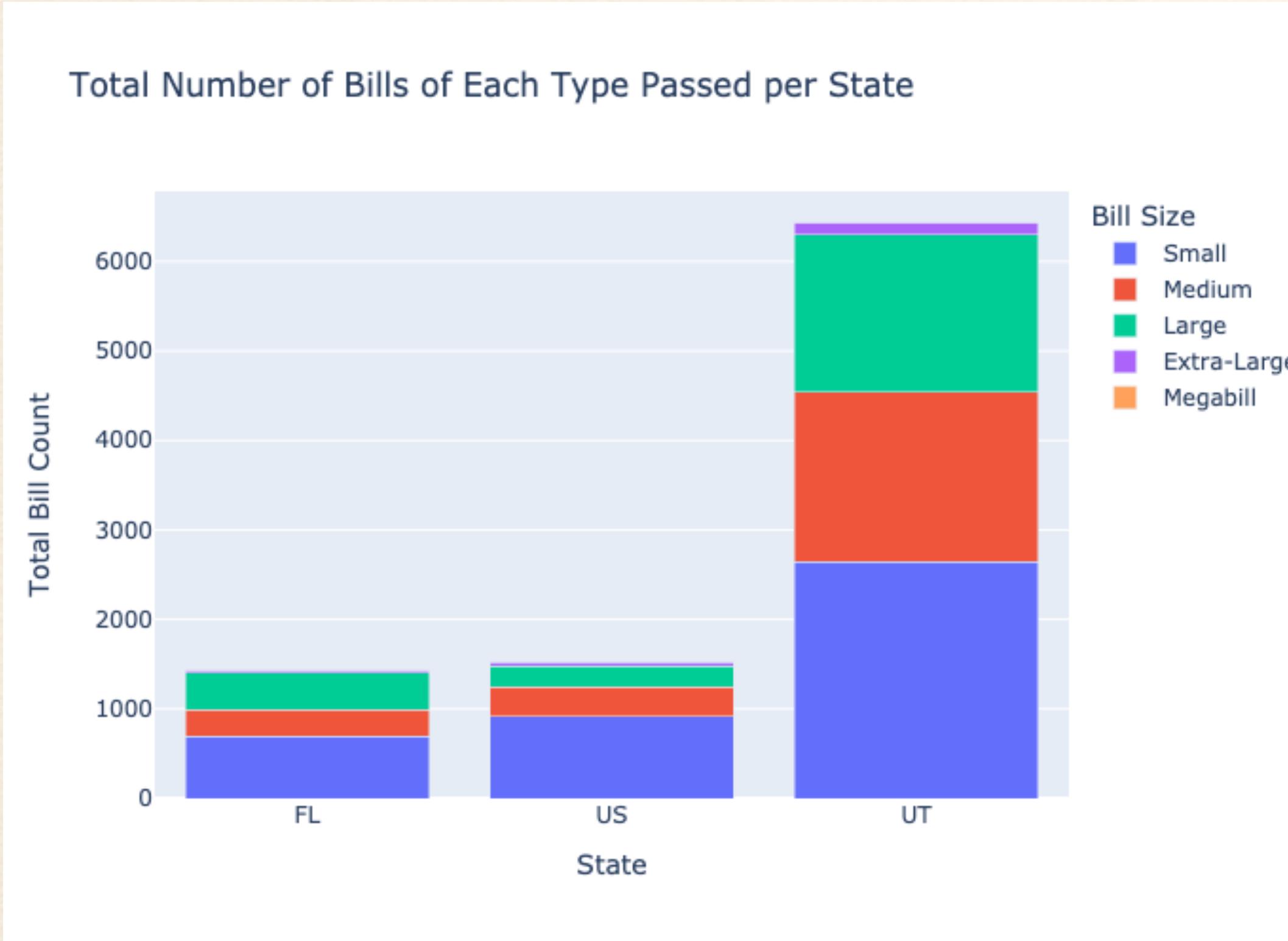


Total Number of Bills of Each Type Passed per State



Real-ish Data

- ❖ With the nulls dropped (a very questionable thing to do):

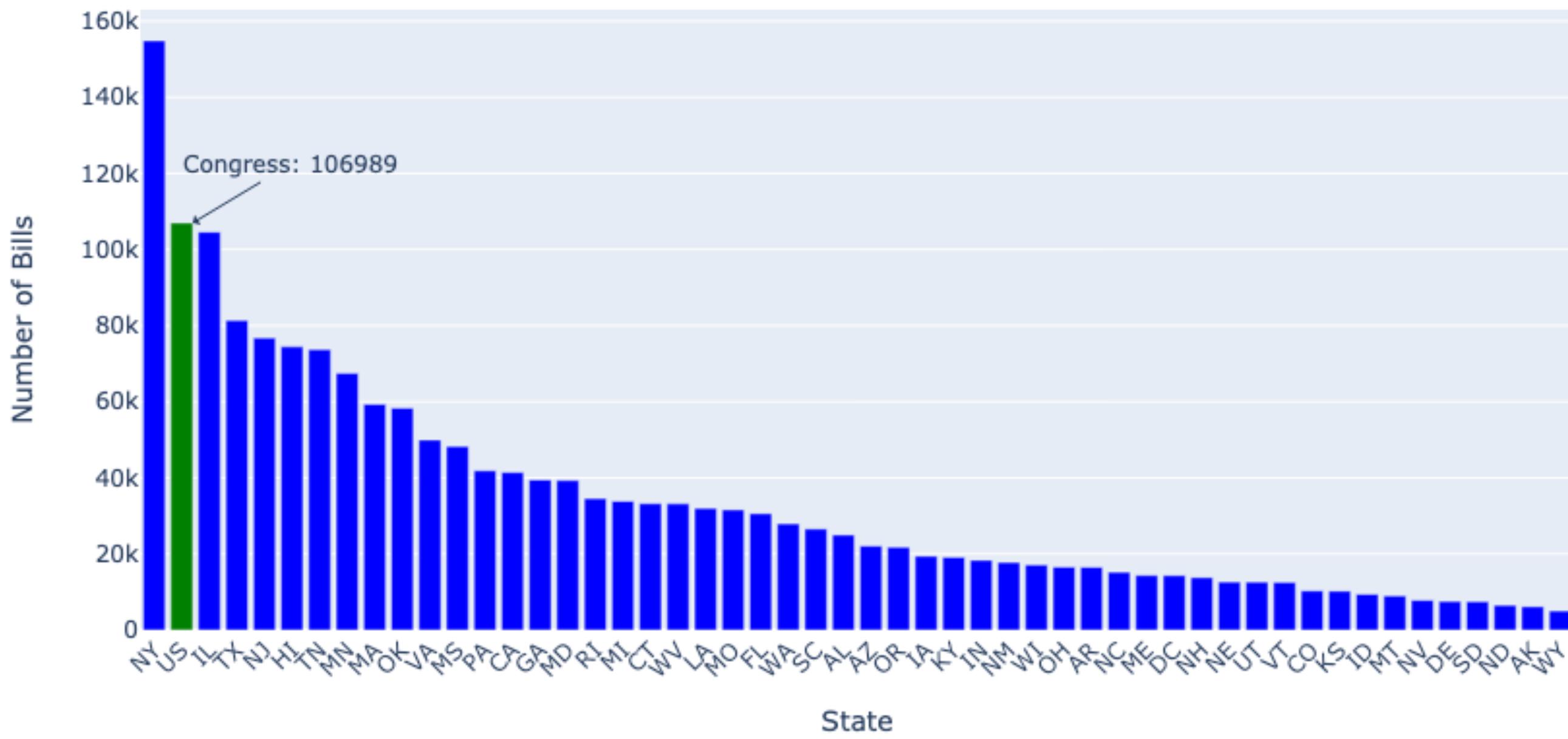


Revised Story

- ❖ After the issues we encountered with the document size, we decided to revise our story a bit using the data that we did have. Some of the questions that came out of this revision were:
 - ❖ Which states had the most legislation in the last 15 years?
 - ❖ Did bills pass at a similar rate state-by-state, or were there some states that passed bills at a higher rate than others?
 - ❖ Has more legislation been introduced year-over-year?

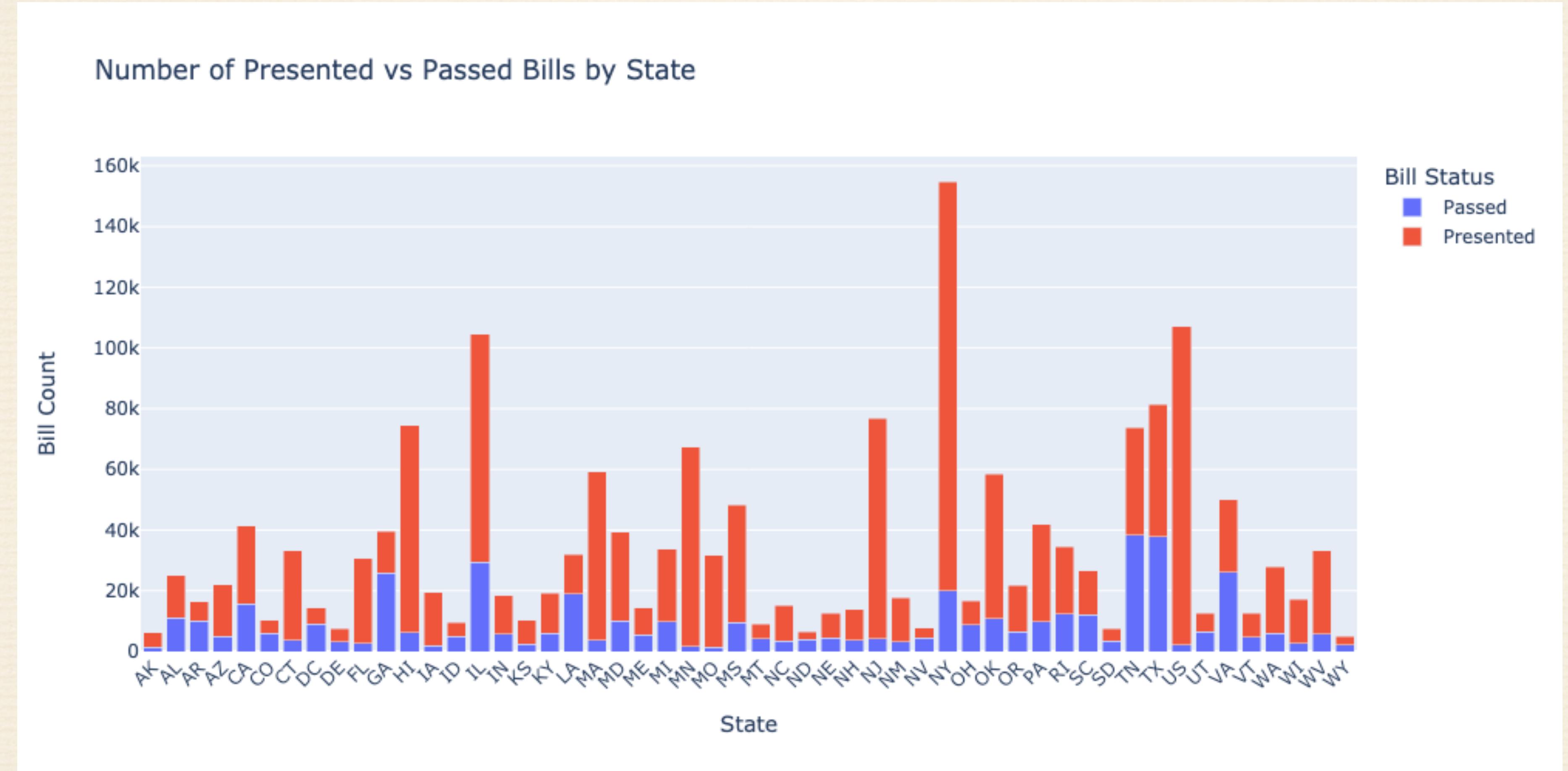
Total Legislation by State

Number of Bills Introduced by State (2009-2024)



- ❖ Over the last 15 years across all 50 states, an average of 30,429 bills have been introduced, averaging out to about 2,029 bills per state per year.
- ❖ **Start Spreading the News:** New York had a staggering 154,852 bills presented in some capacity over the last 15 years, meaning they averaged 5x more bills presented per year than the average state.
- ❖ **I Want to Be a Cowboy, Baby:** Wyoming had 5,045 bills introduced since 2009, averaging out to 336 bills per year.

Presented vs Passed Bills by State



- ❖ **Mr. T:** From a total bill standpoint, Tennessee and Texas both stand out for passing a large proportion of bills.
- ❖ **The Buck Stops Here:** US Congress, on the other hand, stands out as having a large amount of bills presented with a tiny proportion passing.

Legislation Trends over Time

Number of Bills Passed by Year (2009-2023)



Holding Steady: The number of bills passed YoY has held fairly steady since 2013, with 2020 being a noticeable outlier. However, the percent of bills passing seems to be trending downward in the last few years, from a little over 31% in 2018 to about 24% in 2023.

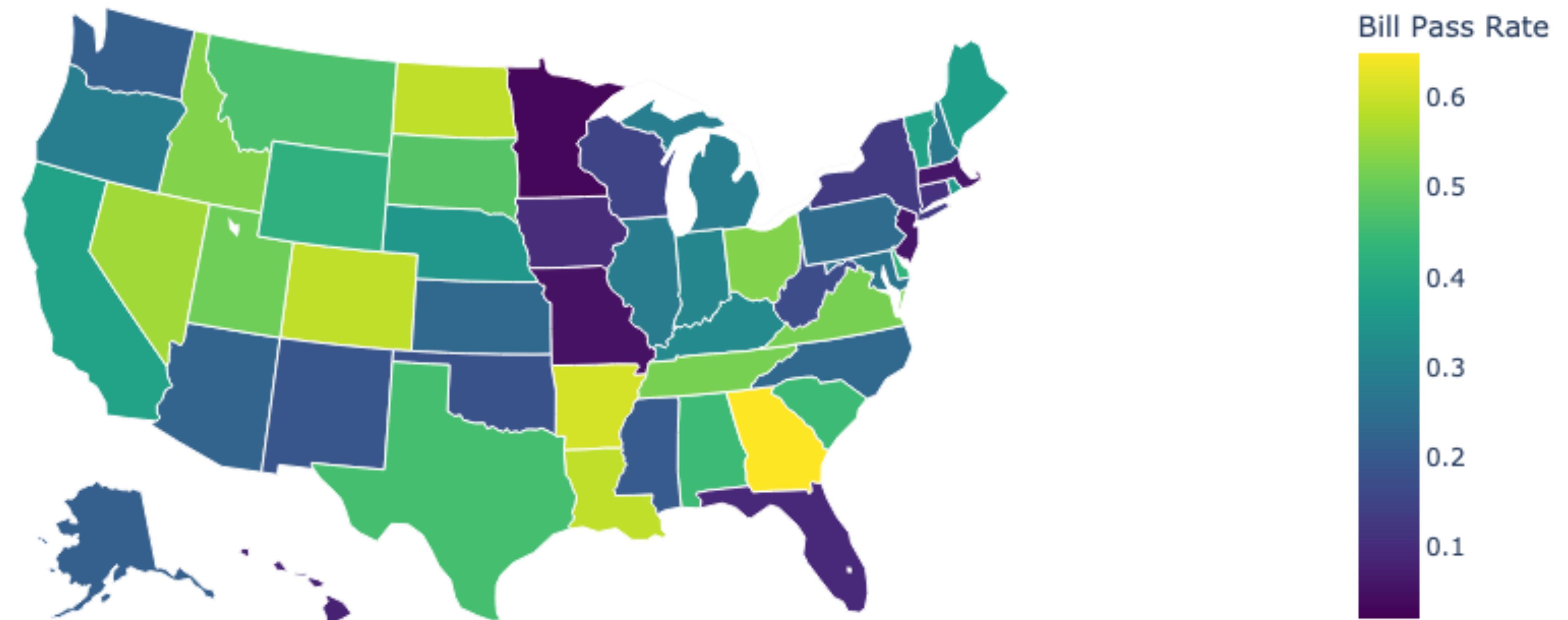
Percent of Bills Passed by Year (2009 - 2023)



Legislation Pass Ratio Visualized

- **Georgia on My Mind:** The Peach State has the highest legislation pass rate with 65% of bills being passed since 2009.
- **Gopher Cold:** Minnesota's pass rate stands at an ice cold 3%, the lowest of the 50 states.

Bill Pass Rates by State



Where To Next?

- ❖ If you corrected the URLs and had the time, with the current dataset you could...
- ❖ Get the real bills passed and bill sizes
- ❖ Find correlations between bill size and difficulty passing
- ❖ Look for other correlations (who's passing the big vs. small bills)
- ❖ With a grant you could actually produce a set of graphs showing the amount of law each person is under from each source of legislation.