How many of you know of GovTrack? How often do you use it? How many of you would call yourselves data journalists? Cover politics?

Let me start with one of my favorite quotes from a news article. This was about the debt negotiations between Congress and the President in 2011 right before the supercommittee was created. Remember how the Republicans asked for too much and President Obama "moved the goal post"? Reflecting on how the events were covered, reporter Matt Bai wrote for The New York Times:

> [A]fter the so-called grand bargain ... the two sides quickly settled into dueling, self-serving narratives of what transpired behind closed doors. ... [T]he whole debacle became the perfect metaphor for a city in which the two parties seem more and more to occupy not just opposing places on the political spectrum, but distinct realities altogether.

Politicians get away with operating in these distinct realities because of information asymmetry. Government is big. There's no one person either inside or outside of government who knows how all of government works, or knows at least how the whole legislative process works, or even just knows how the House works. I'm sure is nothing new to you. The people in power, inside of government, know government better than we do because they live it. And when we're trying to hold them accountable, we're at a disadvantage.

There are about 10,000 bills introduced in each Congress. Each Congress holds thousands of committee meetings. Those bills are revising the United States Code. There are 48,083 sections of the US Code. It's amazing we have any idea what's going on. If you're covering any of this, good on you for trying to make sense of it.

So what are we gonna do?

Andy Samberg's reaction to the system was to throw everything on the ground. I don't think we have to be that extreme. We just look for other tools. Here are some of the headlines for articles mentioning our work: Remixing government data, deep throat meets data mining, computer science in the service of democracy (that one was a little pretentious), and data mining meets city hall.

All right, let me tell you what we actually do. GovTrack.us is a free legislative tracking website. I hope you're as geeky about Congress as I am, otherwise that might sound pretty boring. We gather a LOT of data about Congress and help people understand what's going on, from what bills have been introduced to how your Member of Congress stacks up against his or her peers. It's a very data oriented website. But actually I think of what we do as a lot like journalism. Think of GovTrack as an app that teaches civics through current events.

I'll give some examples of what's on the website. Back in January when we all thought Congress might enact gun control, thousands of individuals came to GovTrack to see what these bills were really all about. Proponents wanted to know whether the bills were strong enough, opponents how much of the 2nd Amendment was being rolled back. Our users read these bills, checked which Members of Congress supported them, and many went on to track it --- which means we'll send them an email alert if a bill has any further action. That includes getting onto the House or Senate's schedule for the week ahead, getting a vote, gaining cosponsors, when the bill's text becomes available, and so on. We also write our own summary of some of the most viewed bills so our users don't have to read hundreds of pages of legalese themselves. And another way we address the problem of legalese is in our bill text comparisons. You can use this to see changes to the text of bills as bills move through the legislative process, or to compare related bills. Here we're showing that a word was deleted, a paragraph was inserted, and a date was changed from 2014 to 2015.

A few weeks ago there was the vote on the Amash amendment to end the NSA's mass data collection. You can find that vote on GovTrack and see how the vote broke down. And you can use

GovTrack to get an email every time Congress votes.

And if you're just getting started tracking Congress, you'll probably start by finding out who your representatives are. [interactive demo?] When you get to a page for a Member of Congress, you'll see a few of our own statistical analyses of Congress: ideology, leadership, and missed votes. And you can click the big Track button at the top to get email updates, or an RSS feed, of this Member's legislative activity.

It's all official, public information about Congress, and mostly information that you can find elsewhere. But we try to make that information a) clear, b) instructive about how Congress works --- how it actually works not the middle school version of civics and not the whitewashed vrsion of how goernment tells you how government works, and c) actionable like being able to track a bill. It's some of the same goals that I'm sure many of you have. In journalism, the medium that you do this through is usually the story, written or spoken. Our medium is sort of interactive exploration and data analysis.

At least for our casual users.

Journalists use GovTrack too, especially for providing context for the stories they're writing. Because we're looking at all bills all the time, our analysis can help journalists put the data points in your stories into a bigger picture.

Here's an article from the Shreveport Times from earlier in the year about Rand Paul as a possible presidential contender. "Rand Paul has sponsored 70 bills, related to federal spending cuts, the Federal Reserve, wiretaps," etc etc "None has passed the Senate." It tells you a little bit about who Rand Paul is as a senator.

We can put that into even more context. How do Paul's bills compare to the bills sponsored by other Republicans? Are they any more conservative than others, or does it just seem that way? We do an analysis that we call the ideology analysis. We feed who's sponsoring whose bills into a statistical analysis that puts every Member of Congress on a spectrum, using about ten thousand bills over the last few years. The purple triangle here is Rep. Rand Paul, and the dots are the other representatives in the House. Just look at this in terms of the left-right axis. Paul is on the right because his record of bills sponsored correlates with other right-leaning representatives. Members of Congress with similar political views will tend to cosponsor the same set of bills and inversely Members of Congress with different political views will tend to cosponsor a different set of bills. So we can see here that Paul is pretty far off to the right, but if you look within just the Republican representatives, he's not too far from the middle.

Here are two reporters using the chart. In the Atlantic Wire, "Toomey's no Tom Colburn." And on Philly.com, "Lautenberg was one of the Senate's most reliable liberals, a Democrat who nearly always voted with his party. The website GovTrack.us rated him as one of the four most liberal members of the Senate." I don't really like to see this called a rating. It gives the sense that it's a judgment, when it's more objective than that.

Let's go back to bills.

Do you remember how many bills I mentioned are introduced in every Congress? About 10,000. In the first 200 days of this year about 4,700 bills were introduced. Is that a lot or a little? What does that tell us about the gridlock on capitol hill? We did an analysis of historical data last month. And let me say, thank goodness we're not living in the '70s. Those congressmen wrote a ridiculous number of bills. In recent history, 4,700 is low. And you know that whether you want bigger government or smaller government, you still want Congress to be introducing bills to make that happen. It takes a law to repeal a law. Low doesn't mean less government, it means Congress just isn't doing anything.

And do you know what percent of bills will be enacted? Let's say even before the gridlock today started. Let's go back to the Clinton and Bush years Let's do a show of hands. Everyone put your hand in the air. If you think Congress passes at least half of the bills it introduces, put your hand down. If you think Congress passes less than half of the bills it introduces, keep your hand up. Ok, so you all think it's fewer than half. Do you think they pass a third of the bills they introduce? If you think they

pass a third, put your hand down. If you think it's less than a third keep your hand up. What about a quarter? What about just an eighth? A sixteenth? All right, well there were a few Congresses that passed as many as a sixteenth of the bills they introduced. But in most years it was about a twentieth or less, 4 or 5 percent. In 2011-2012, it was just 2%. Looking at the first 200 days of this year, we're on track for Congress enacting the fewest number of bills in modern history. The McLaughlin Group on TV reported on our coverage of this, and the Huffington Post used our data to come to similar conclusions.

This is all really surprising to most people. Until you know how few bills are enacted, you think every story about a bill in Congress is important. You get the wrong idea about what is actionable information.

On GovTrack we give every bill a prognosis, it's a percent, the likelihood that we think the bill will be enacted. And like everything else, it's based on an objective statistical analysis, that takes into account factors that have, historically, made bills more or less likely to be enacted. Let's take an immigration bill by Rep. Chaffetz. We're giving this bill a 33% chance of being enacted. That's ten times more likely than most bills. This is a bill you should probably pay attention to.

This is context for understand whether this bill is relevant, whether as citizens and journalists this is something we should take action on and put energy into, or if it's something we can ignore. The prognosis is actually also a guide to why it might be important. Because we list the factors that go into the result. The sponsor of the bill, Chaffetz, is in the majority party in the House. That's crucial. He's also on a committee that the bill was referred to. That's really important. And he has cosponsors from the minority party, which is correlated with successful bills in the past. *And*, what's often not obvious at all, the prognosis tells us there's a history to this bill. Chaffetz introduced this bill last congress and it made it out of committee last Congress. That gives the bill a huge leg up this time around.

You'd know to look for all this if you worked in Congress. You'd know the factors and you'd have an idea of what those factors meant for a bill's chance of being enacted if you were on the inside. But here's this information asymmetry again. We don't know it, and that puts us at a disadvantage when we're trying to report on government. We have to guess what the factors are and then use statistical models to tell us which matter and which don't, and of those that matter whether they are good or bad for a bill, and how much. Each factor provides a little more context about what may be happening behind closed doors, and it might be a lead to an interesting story.

The prognosis is our most cited analysis. Often when a bill is *un*likely to be enacted. In U.S. News and World Report, they wrote, "[Rep. Matt] Salmon has proposed an amendment to make term limits apply to all his colleagues .... But don't count on it going anywhere — GovTrack estimates a 0 percent chance of the bill getting past committee, and a 0 percent chance of it being enacted." And on the ever-colorful Huffington Post, they wrote that the Medicare Identity Theft Prevention Act has "only a 2 percent chance of becoming law. In other words, the Chicago Cubs have a better shot at winning the World Series."

We're also tracking votes. Do you all think missing 2% of votes is good or bad? Does anyone want to guess the median missed vote rate in the senate? This is our chart for missed votes over time. This one is for Senator Bob Menendez. Each point is the percent of missed votes, and it's over three-month periods from when Menendez entered the Senate in 2006. So here in April to June of *last* year he missed about 2% of votes. We're also showing the context. The bands are marking off percentiles. The first band at the bottom is the best 25% of senators, the second band takes you to the median which is actually right about 2%. So Menendez's record is right in the middle, not really either good or bad.

Heading into last year's primaries, Representative Connie Mack took a beating in the press for missing votes while he was running for the senate. Mack deflected by saying his opponent, Senator Bill Nelson, had missed even more votes. Politifact's Truth-o-meter looked up the numbers on GovTrack

and rated Mack's claim true.

And the Chicago Tribune ran a story on the absentee rates of the representatives in the Chicago area. I worked with the reporter on pulling the numbers she needed for the story, which showed that Chicago's representatives were among the worst in the House. The story hit a nerve in Bobby Rush, who sent an angry letter to Tribune Publishing's CEO a few weeks later complaining that a good legislator isn't one who shows up to votes.

So we work with journalists a fair amount on pulling out some custom data from what we have. Last summer The Washington Post reported Mitt Romney touting that his VP running mate had "demonstrated an ability to work across the aisle." The Post came to us to figure out if that was true, at least by the numbers. So we looked at how often Paul Ryan signed on as a cosponsor to a Democratic bill, and compared that to the average for all Republican congressmen. He co-sponsored 975 bills in his career, up to that point, and of those, 22% were Democratic bills. The median across all Republicans serving at the time was 19%, making Ryan ever-so-slightly more of an aisle-crosser than the rest of his party. So, there was some truth to Romney's claim in the numbers.

What The Post didn't end up printing was that the median was heavily skewed by the freshmen class of Republicans, who were particularly partisan. When you take them out of the picture, the median rate of cosponsoring a Democratic bill jumps up to 25%. Meaning, Ryan actually appeared *more* partisan than his peers, by the numbers.

Bloomberg picked up on the story and we worked with them on figuring out which bills made up that 22%, and whether what Ryan was cosponsoring was meaningful legislation, which it wasn't. But that's not really surprising. Now that you know that 95% of bills don't get enacted, you could guess that most bills are really not meaningful.

There are some things we *don't* do on the site. We don't do campaign finance. And we don't look for corruption. Those aren't stories that we believe can't be told primary through data, so we leave that to the real journalists. But the legislative record is often an important part of showing that conflicts of interest are systematically present.

The Washington Examiner last year ran a series of articles called "Tariff waivers are cash cows for Hill leaders." They went through thousands of bills on GovTrack looking for MTBs, which apparently stands for Miscellaneous Tariff Bills. These bills create special exemptions from tariff duties. The Examiner's staff compared the bills to MTB disclosure forms filed by congressmen to identify the companies that would benefit from these bills, and then looked for those company's employees's campaign contributions to see if the money was going full circle. It was a major research project. They found that Senators Bob Casey and Robert Menendez had introduced hundreds of these tariff break bills. Other Members of Congress, like Speaker John Boehner, had received hundreds of thousands of dollars from companies connected to MTBs. And, I think as a result of the Examiner's analysis, there have been some efforts to reform this process, though I don't know if anything has actually changed. If you're interested in knowing which bills these were, the Examiner did a great thing and published their data spreadsheet to Google Docs.

The Washington Post did a similar story around the same time but instead of tariffs, they looked at the stock portfolio's of Members of Congress. "More than 100 Members of Congress...traded stocks or bonds in companies lobbying on bills that passed through their committees." One in eight trades by a Member of Congress was in a company they regulate. If it wasn't insider trading, it was at least a conflict of interest. I think we provided the information on committee assignments, which was used to match representatives to the companies being investigated or regulated by a committee.

Both the Post and the Examiner began with a large data analysis to quantify the problem across every bill or every stock trade. And they followed-through with particular stories about individual trades and the circumstances that made it more or less of a conflict. Having the data to show a pattern is wide-spread, and then showing from particular cases that it means what you think it means, makes for a

very powerful story.

That's what we do. I want to mention a little about how we do it, and how you could do it to.

Everything on GovTrack is automated. I keep saying we, but GovTrack is me and a few part time staff. But we don't enter anything into the site by hand. It all works by automatically fetching information like bills from government websites that are already publishing this information, but usually not in a very effective way. And almost always not in a data friendly way. If you wanted to do your own analysis of a vote in Congress, you're going to hit a lot of data problems. For instance, on a recent vote to reauthorize the Violence Against Women Act, representative Mike Rogers voted both for and against the bill. Does anyone know how that's possible? Well there are two Mike Rogers's in Congress. So this is what we worry about. If you wanted to get a spreadsheet of all of the bills in Congress, something simple. Well, I asked folks in Congress about that more than 10 years ago and I'm still waiting on it. What GovTrack does is make that spreadsheet, and a lot more, until Congress comes around and realizes they should be making this data available themselves. We badger Congress to publish voting information like *this* using alphanumeric identifiers for Members of Congress so we don't have to worry about duplicate names, and similarly for information about bills, committee assignments, and so on, and all so we can take the data and go on to do reliable analysis and build applications on top of that data.

Most of our work to build this database is open source, we do a lot of the work on github which is a collaborative coding website, and we collaborate with developers from The New York Times, Yahoo News, the Sunlight Foundation especially, and others. You can download all of the raw data on GovTrack, and GovTrack also has an API if you want to query our database. The House itself has been making *some* headway on legislative transparency through data. There's a new website called docs.house.gov that has upcoming bills and committee meetings, and there's all new data on the United States Code from the House. So there are a lot of options for you to start with some raw data and find something new in it.

And if you found any of this appealing and want to get your hands dirty in this whole open government data thing, I'd say check out my book, which is a guide to the new open government data movement.