Texas Legislative Accountability and Transparency Database: A Database Design Study

Kira Lowe
EPPS 6354 Information Management
University of Texas at Dallas
Dr. Karl Ho
May 8, 2020

<u>Introduction</u>

The importance of being informed

In the internet era, information is more available than ever before. When it comes to staying informed about politics, Americans have their pick of sources ranging from official government websites to news media or even social media. Despite the wide availability of information today, many Americans are unaware of how their government works or even who represents them.

A 2018 survey from Johns Hopkins University found that fewer than twenty percent of respondents could name one of their state legislative representatives. Alarmingly, even though the average citizen doesn't know who is leading their state, about seventy percent of people say their state government is doing a better job than the federal government (Johns Hopkins University [JHU] 2018). Benjamin Ginsburg, a political scientist at Johns Hopkins who was involved in the survey, speculates that this is because the media focus is primarily on coverage of national, rather than local, politics (JHU 2018).

An additional consideration is that many Americans do not understand the importance of local government. States collect and spend roughly $1.7 trillion tax dollars annually (JHU 2018). State and local governments make decisions that impact daily life for Americans – they spend billions on everything from highways to schools. During the current COVID-19 pandemic, elected officials at the state and local level are making serious decisions about public health and the economy that will have effects lasting for many years.

The idea that a majority of Americans lack awareness of who represents them and, consequently, what that representative is doing for them can have disturbing implications during election years. Not only does a contentedly uninformed public open the door to significant potential for corruption, but also complacency which can suppress voting rates. Indeed, voter turnout is historically lower in local elections than in national elections. It is imperative that information about these

legislators is not only public and transparent, but easily accessible in a format that is customizable to individual locations and issue preferences. Providing this information in an easily digestible way will help voters make informed decisions and increase accountability of elected officials.

Existing resources

At the federal level, sites like inflenceexplorer.com and opensecrets.org focus on ensuring transparency of campaign finance for federal legislators. A very useful resource for government transparency is GovTrack.us (GovTrack), which has made information about federal representatives and legislation free in an easy-to-use format since the early 2000s (GovTrack 2020). This resource is very thorough and user-friendly.

From the homepage of GovTrack.us, a user of any knowledge or experience level can easily input their address to instantly see who their senators and congressional representatives are. They can then click through to a profile page for each legislator which provides information on the legislator's party, leadership roles, ideology, and how long they've been in office and the year of their next election, how they've been rated by advocacy groups ranging from Planned Parenthood to Americans for Prosperity. Importantly, each legislator's page also contains information about the bills they've recently authored and how they've voted on other bills. Users can also look at any bill by topic by selecting one of thirty-three categories from the home page. Additionally, GovTrack makes their data and methods of analysis public, even going as far as to encourage hackers, journalists, and researchers to download and reuse data from their GitHub repository (GovTrack 2020).

While the service provided by GovTrack is a very thorough, useful, and user-friendly resource for examining *federal* congressional activity, such a resource does not exist for the Texas state legislature. The most comprehensive existing resource, capitol.texas.gov, is cumbersome to operate and does not provide a way to get more than the most basic information about one representative on a single page. Data are siloed such that, despite offering almost as much information as GovTrack, it is impossible to

access all relevant information without visiting dozens of pages or even leaving the site altogether. For example, basic information like party or year elected are not consistently available on a legislator's profile page.

Proposed resource

In this paper, I will explore the process of migrating data from capoitol.tx.gov to a new database which will ultimately power a site similar to GovTrack but with a focus on the Texas state legislature. By setting up a streamlined database structure, information can be easily queried to increase functionality in the user interface. In a final version of the site, users will be able to search for the names of their representatives by entering their address, which will allow them to examine profile information including party, year elected, committee memberships, sponsored bills, and key votes for their representatives. Users will also be able to search topics of interest to see what legislation has been introduced or enacted that interests and impacts them. The goal of this page is to increase transparency within the Texas state legislature to help citizens develop awareness of legislative representatives and activities and hold their legislators accountable by making informed voting decisions.

Database design

Entity-relationship diagram

In the Texas state congress, each legislator serves a two-year term, so there is potential for a single house district to have five different legislators in a decade. There are sixty-eight house, senate, and joint committees on which multiple legislators will serve. Thirty-two senators and 150 representatives in the Texas state legislature author and vote on over ten thousand bills in a typical legislative year. The purpose of enumerating this is to say that the volume of data required to power this site is very large. Because the data on capitol.texas.gov were difficult to access because of limited data available for download, the current design utilizes a small subset of data for providing a proof of

concept. However, the database design decisions discussed in this section were made with the ultimate

goal of using the full range of data available on the Texas legislative website and others.

    A large volume of data requires serious consideration in designing the most efficient database

structure. An entity-relationship diagram (E-RD) was designed to graphically represent the database

structure in a conceptual framework which reflected the needs of prospective users of the site

(Silberschatz, Korth, & Sudarshan 2019). A key consideration of the E-RD shown in figure 1, below, was

to eliminate redundancy and maximize efficiency.

Figure 1: Entity-Relationship Diagram
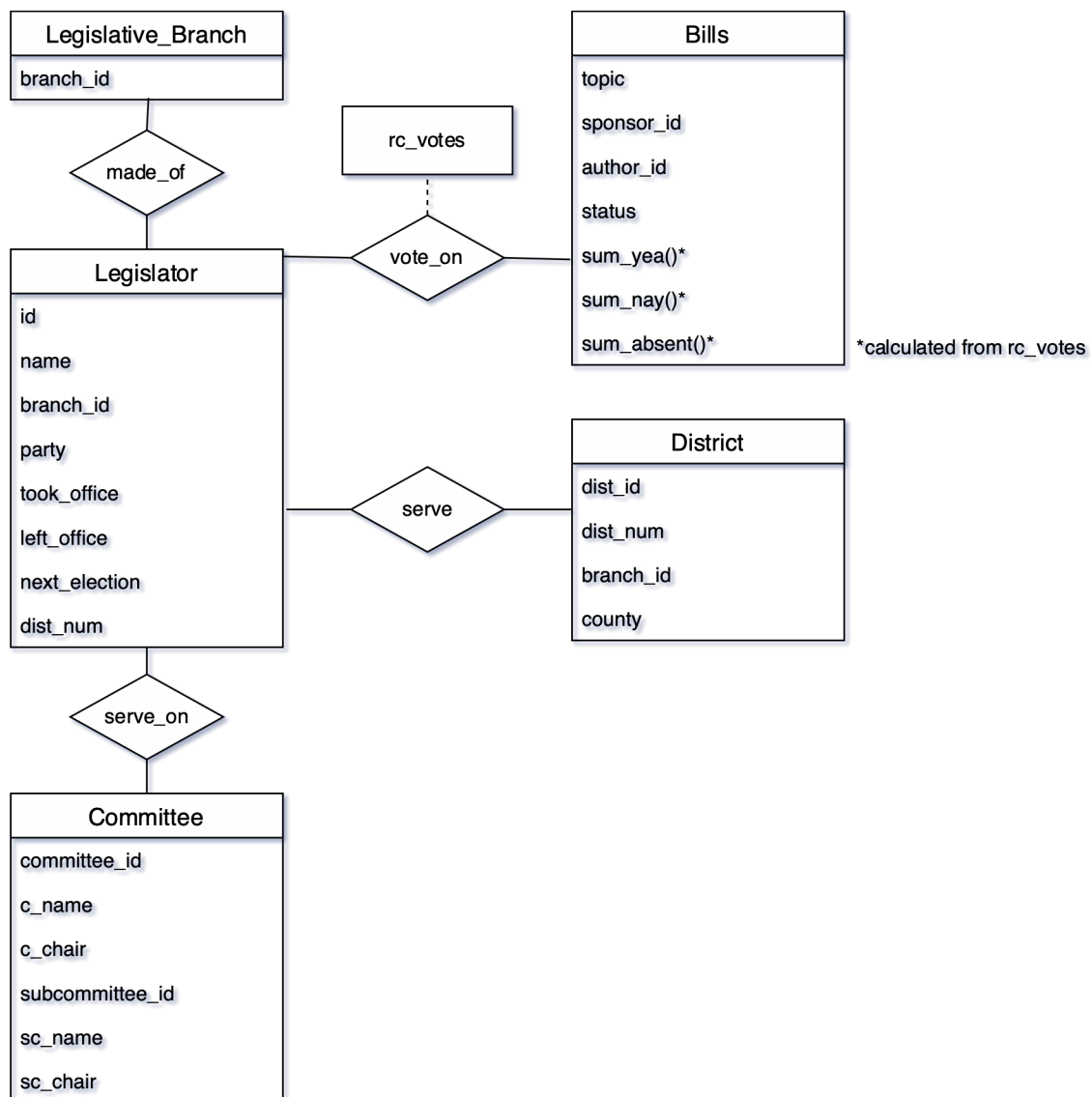
Each entity set can be connected to others via primary or foreign keys (Silberschatz, Korth, & Sudarshan 2019). For example, the "district number" attribute in the "legislator" entity serves as a foreign key connecting to the "district number" foreign key in the "district" entity. The diagram also contains the descriptive attribute "roll call votes" which describes the votes of yea, nay, and absent for bills which have reached the floor. Values from this descriptive attribute can be used to calculate the "yea," "nay," and "absent" attributes in the "bills" table.

Data collection and database build

Data were collected from the Texas legislative websites including the Texas capitol, house, and senate sites. Because certain data like party affiliation were not consistently available on these sites, some data had to be sourced externally from sources like *The Texas Tribune*. Because of limited availability of data for download, each entity table was constructed by manually pulling information from each site. For this reason, examples of data given in this paper are comprised of a limited subset of data and are intended to represent proof of concept only.

Using the model provided by the E-RD design, tables were constructed using the sources described above. Using pgAdmin 4 and PostgreSQL, table structures were created with defined column names and types like integer, character, or varying character length. Primary keys were also set using PostgreSQL code. After defining the schema, attributes were uploaded to each table from CSV files.

Using the database

PostgreSQL queries

PostgreSQL is a popular open-source object-relational database system (PostgreSQL Global Development Group 2020). Queries can be developed to view one or many attributes of a single table, or tables can be joined to display attributes from multiple tables. Using the PostgreSQL syntax and the limited, proof of concept data that was loaded into pgAdmin 4, I was able to create queries that would provide information similar to what would be on a legislator's GovTrack profile.

For example, I was able to select the attributes for legislator name, legislative branch, party affiliation, year of taking office, and year of next election for my current state representative without having to leave the "legislator" table. Currently, three of these five attributes are not available on the government profile pages of members of the Texas House of Representatives. By using a join command to connect the "legislator" and "bill" tables, I was able to see the topics of bills that my representative had authored. I was also able to see the topics of bills authored by all representatives in my county and surrounding counties by joining the "bills" and "district" tables. I can also query interesting information like the number of Republicans vs. Democrats in the Texas Senate or see all Republicans who are up for reelection in 2020.

Visualization using Falcon

Using the Falcon SQL Client, an open-source editor with visualization capabilities (Falcon by Plotly 2020), I was able to use data from my pgAdmin4 database and PostgreSQL queries to produce data visualizations like the one seen in figure 2, below. Visualizations like this one can be useful to help illustrate concepts to users that might be more difficult to see in table format. Here, I have distilled all

Figure 2
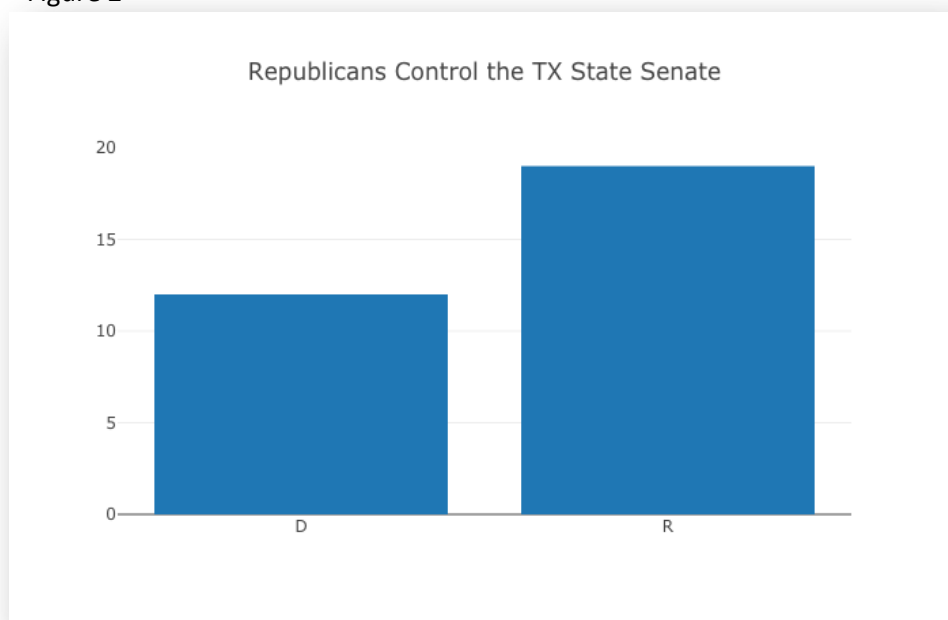
members of the Texas Senate into their parties so that a visual comparison can be made to see which party controls this legislative body.

Other useful visualizations could include comparison of the volume of bills authored by topic – as we can see in figure 3, State Senator Angela Paxton authors more bills on education than she does for healthcare (note: this is a limited data set for proof of concept – this is not an accurate representation of the work of Angela Paxton). A user who is a strong supporter of healthcare reform in Texas could use this data to make decisions about whether they support Senator Paxton's legislative focus, enabling the user to share their thoughts with their senator and to make informed decisions when deciding whether to reelect them.

Figure 3
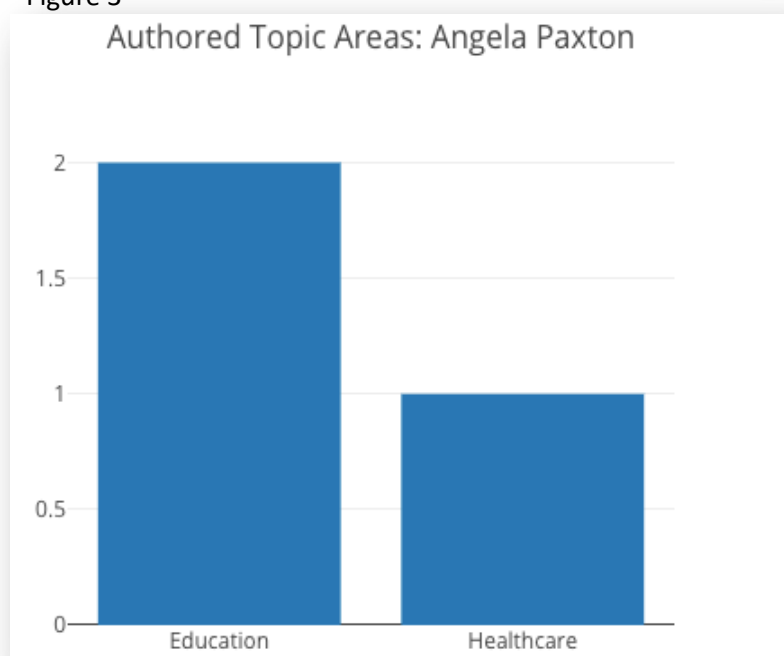
Django Integration

The next step was to create and locally host a web application using Django, a web framework for application development (Django Software Foundation 2020) which can connect to a pgAdmin 4 database. After connecting my database to the server, I used Python to create data models which supported a catalog containing various tables for display on the web. Due to technical difficulties the

web pages did not display, however I was able to set up an online form where data could be entered online and translated to the pgAdmin database.

Conclusion and next steps

Current status

Currently, the database is not ready for use by the public. Limited availability of data downloads from government or other reputable websites meant that only a small subset of data is presented in this demonstration project. Additionally, technical issues with Django prevented the display of data on the web. However, the result of queries from the database show that it is possible to present more information on a single page using this database than is currently available on existing resources. While a fully functional web application is not up and running, the PostgreSQL queries and Falcon visualizations show that the database is feasible and can function as intended.

Future work

Future versions of this project will come in several iterations. First, more data will be collected. Potential avenues for expanding the dataset could involve writing a Python script to scrape basic legislator information off the web to complete the "legislator" table. Additionally, some text mining can be performed using R to identify key words within bills. Finally, additional data from advocacy group ratings should be incorporated into a new table.

After the data collection outlined above, full data from multiple legislative years will be uploaded to the existing database in pgAdmin 4. Once full data are available in the database, more visualizations will be possible using Falcon. Importantly, a web application must be developed to convey this information to the public. Front-end development and a graphic user interface are key to the user experience. With the full scope of data available in the pgAdmin database, more information will be viewable on this site than is currently available on the existing government resources.

The final essential component to a new and improved accountability and transparency database website is to include the ability for a user to search for their representative based on their address. This component will involve a significant dataset that has yet to be explored in the current project but is essential to making the data as easily accessible as possible. It is imperative that there are as few barriers as possible between the public and information that is essential to their understanding of their government. Once fully functional, this website will take a big step toward enabling citizens to interact effectively with their elected officials and to vote responsibly.

Django Software Foundation. n.d. *Django*. https://www.djangoproject.com/ (8 May 2020).

Falcon by Plotly. 2020. Falcon Free, open-source SQL client for Windows and Mac. https://github.com/plotly/falcon (8 May 2020).

GovTrack.us. n.d. *About GovTrack.us*. https://www.govtrack.us/about (8 May 2020).

Johns Hopkins University. (2018, December). *JHU Survey: Americans Don't Know Much About State Government*. https://releases.jhu.edu/2018/12/11/jhu-survey-americans-dont-know-much-about-state-government/

PostgreSQL Global Development Group. n.d. *About PostgreSQL*. https://www.postgresql.org/about/ (8 May 2020).

Silberschatz, Abraham, Korth, Henry F. and Sudarshan, S., 2019. *Database system concepts, 7th edition*. New York: McGraw-Hill.

Texas House of Representatives. n.d. *Texas House of Representatives*. https://house.texas.gov/ (8 May 2020).

Texas State Senate. n.d. *The Texas Senate*. https://senate.texas.gov/ (8 May 2020).

Texas State Legislature. n.d. *Texas Legislature Online*. https://capitol.texas.gov/ (8 May 2020).

The Texas Tribune. n.d. *Elected Officials Directory*. https://www.texastribune.org/directory/ (8 May 2020).