govdat vignette - Interface to various APIs for government data.

govdat vignette - Interface to various APIs for government data.

About the package

`govdat` is an R package to search and retrieve data from two sets of APIs, the Sunlight Labs APIs, andn the New York Times congress API.

Returned objects from functions are simple lists. That is, you likely will want to take output objects and make data.frames, vectors, matrices, etc. In future versions of govdat, I will return data.frame's when possible as those are easy to work with in R for beginners, though advanced users probably like lists more (I'm guessing, or just the raw JSON, eh)?

The following are examples of using the Sunlight Labs API. I will add examples of using the New York Times Congress API once their site is up again; I'm doing this on 2013-08-28, just after the takedown of their site due to hackers.

Install govdat {% highlight r %} install.packages("devtools") library(devtools) install_github("govdat", "schamberlain") {% endhighlight %}

Load govdat and other dependency libraries {% highlight r %} library(govdat) {% endhighlight %}

SunlightLabs Data

Gets details (subcommittees + membership) for a committee by id.
{% highlight r %} library(govdat) key <- getOption("SunlightLabsKey") out <- sll_cg_getcommittees(id = "JSPR", key = sunlightkey) out$response$committee$members[[1]]$legislator[1:5] {% endhighlight %}

{% highlight text %} $website [1] "http://www.alexander.senate.gov"

$fax [1] "202-228-3398"

$govtrack_id [1] "300002"

$firstname [1] "Lamar"

$chamber [1] "senate" {% endhighlight %}

Gets a list of all committees that a member serves on, including subcommittes. {% highlight r %} out <- sll_cg_getcommitteesallleg(bioguide_id = "S000148", key = sunlightkey) out$response$committees[[1]] {% endhighlight %}

{% highlight text %} $committee $committee$chamber [1] "Senate"

$committee$id [1] "SSRA"

$committee$name [1] "Senate Committee on Rules and Administration" {% endhighlight %}

Get districts for a latitude/longitude. {% highlight r %} out <- sll_cg_getdistrictlatlong(latitude = 35.778788, longitude = -78.787805, key = sunlightkey) out$response$districts {% endhighlight %}

{% highlight text %} [[1]] [[1]]$district [[1]]$district$state [1] "NC"

[[1]]$district$number [1] "2" {% endhighlight %}

Get districts that overlap for a certain zip code. {% highlight r %} out <- sll_cg_getdistrictzip(zip = 27511, key = sunlightkey) out$response$districts {% endhighlight %}

{% highlight text %} [[1]] [[1]]$district [[1]]$district$state [1] “NC”

[[1]]$district$number [1] “2”

[[2]] [[2]]$district [[2]]$district$state [1] “NC”

[[2]]$district$number [1] “4”

[[3]] [[3]]$district [[3]]$district$state [1] “NC”

[[3]]$district$number [1] “13” {% endhighlight %}

Search congress people and senate members. {% highlight r %} out <- sll_cg_getlegislatorsearch(name = “Reed”, key = sunlightkey) out$response$results[[1]]$result$legislator[1:5] {% endhighlight %}

{% highlight text %} $website [1] "http://www.reed.senate.gov"

$fax [1] “202-224-4680”

$govtrack_id [1] “300081”

$firstname [1] “John”

$chamber [1] “senate” {% endhighlight %}

Search congress people and senate members for a zip code. {% highlight r %} out <- sll_cg_legislatorsallforzip(zip = 77006, key = sunlightkey) library(plyr) ldply(out$response$legislators, function(x) data.frame(x$legislator[c(“firstname”, “lastname”)])) {% endhighlight %}

{% highlight text %} firstname lastname 1 Sheila Jackson Lee 2 Ted Cruz 3 John Cornyn 4 Ted Poe {% endhighlight %}

Find the popularity of a phrase over a period of time.

Get a list of how many times the phrase "united states" appears in the Congressional Record in each month between January and June, 2010: {% highlight r %} sll_cw_timeseries(phrase = "united states", start_date = "2009-01-01", end_date = "2009-04-30", granularity = "month", key = sunlightkey) {% endhighlight %}

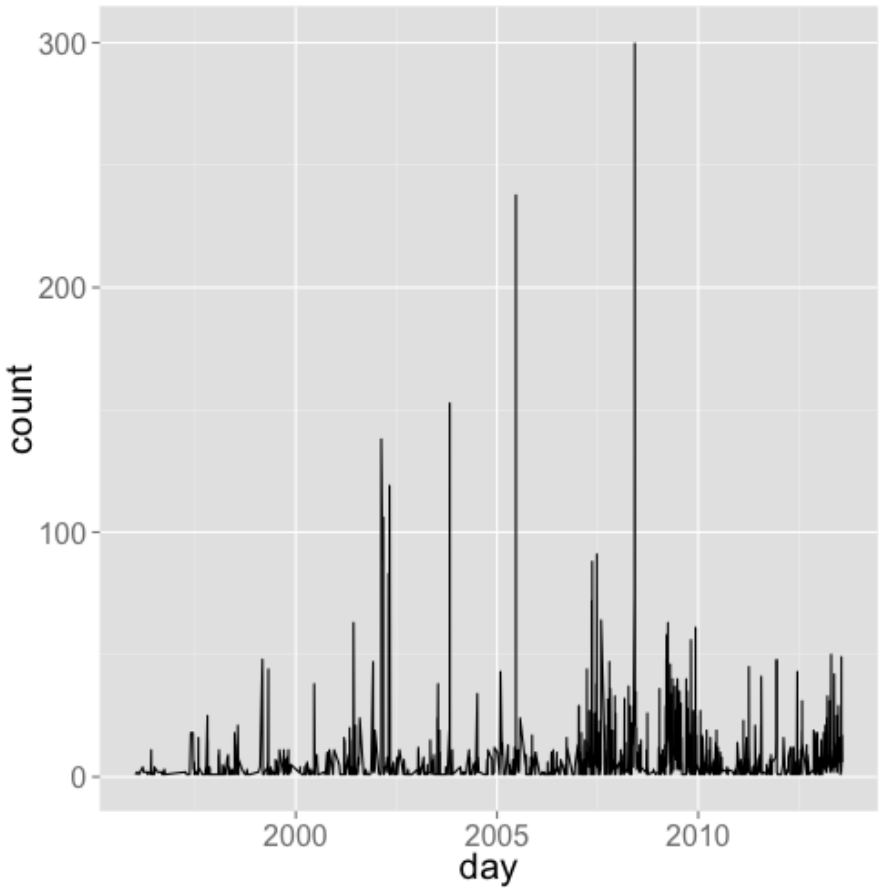
{% highlight text %} 4 records returned {% endhighlight %}

{% highlight text %} count month 1 3805 2009-01-01 2 3512 2009-02-01 3 6018 2009-03-01 4 2967 2009-04-01 {% endhighlight %}

Plot data {% highlight r %} library(ggplot2) dat <- sll_cw_timeseries(phrase = "climate change", key = sunlightkey) {% endhighlight %}

{% highlight text %} 1354 records returned {% endhighlight %}

{% highlight r %} ggplot(dat, aes(day, count)) + geom_line() + theme_grey(base_size = 20) {% endhighlight %}



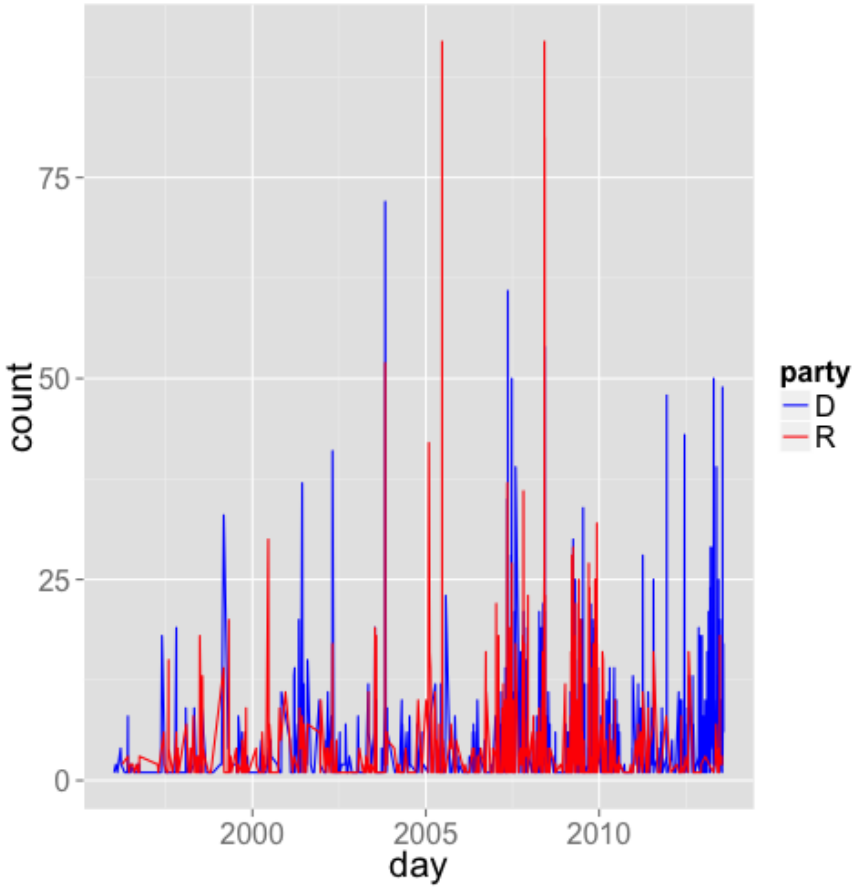
Plot more data {% highlight r %} dat_d <- sll_cw_timeseries(phrase = “climate change”, party = “D”, key = sunlightkey) {% endhighlight %}

{% highlight text %} 908 records returned {% endhighlight %}

{% highlight r %} dat_d$party <- rep(“D”, nrow(dat_d)) dat_r <- sll_cw_timeseries(phrase = “climate change”, party = “R”, key = sunlightkey) {% endhighlight %}

{% highlight text %} 623 records returned {% endhighlight %}

{% highlight r %} dat_r$party <- rep(“R”, nrow(dat_r)) dat_both <- rbind(dat_d, dat_r) ggplot(dat_both, aes(day, count, colour = party)) + geom_line() + theme_grey(base_size = 20) + scale_colour_manual(values = c(“blue”, “red”)) {% endhighlight %}



Search OpenStates bills. {% highlight r %} out <- sll_os_billsearch(terms = “agriculture”, state = “tx”, chamber = “upper”, key = sunlightkey) lapply(out, “[[”, “title”)[100:110] {% endhighlight %}

{% highlight text %} [[1]] [1] “Relating to the sale by the Brazos River Authority of certain property at Possum Kingdom Lake.”

[[2]] [1] “Proposing a constitutional amendment providing immediate additional revenue for the state budget by creating the Texas Gaming Commission, and authorizing and regulating the operation of casino games and slot machines by a limited number of licensed operators and certain Indian tribes.”

[[3]] [1] “Relating to production requirements for holders of winery permits.”

[[4]] [1] “Relating to the use of human remains in the training of search and rescue animals.”

[[5]] [1] “Relating to end-of-course assessment instruments administered to public high school students and other measures of secondary-level performance.”

[[6]] [1] “Relating to public high school graduation, including curriculum and assessment requirements for graduation and funding in support of certain curriculum authorized for graduation.”

[[7]] [1] “Relating to certain residential and other structures and mitigation of loss to those structures resulting from natural catastrophes; providing a criminal penalty.”

[[8]] [1] “Recognizing March 28, 2013, as Texas Water Conservation Day at the State Capitol.”

[[9]] [1] “Recognizing March 26, 2013, as Lubbock Day at the State Capitol.”

[[10]] [1] “In memory of Steve Jones.”

[[11]] [1] “Relating to the regulation of dangerous wild animals.” {% endhighlight %}

Search Legislators on OpenStates. {% highlight r %} out <- sll_os_legislatorsearch(state = “tx”, party = “democratic”, active = TRUE, key = sunlightkey) out[[1]][1:4] {% endhighlight %}

{% highlight text %} $last_name [1] “Naishtat”

$updated_at [1] “2013-08-29 03:03:22”

$nimsp_candidate_id [1] “112047”

$full_name [1] “Elliott Naishtat” {% endhighlight %}

Search for entities - that is, politicians, individuals, or organizations with the given name {% highlight r %} out <- sll_ts_aggregatesearch("Nancy Pelosi", key = sunlightkey) out <- lapply(out, function(x) { x[sapply(x, is.null)] <- "none" x }) ldply(out, data.frame) {% endhighlight %}

{% highlight text %} name count_given firm_income count_lobbied 1 Nancy Pelosi (D) 0 0 0 2 Nancy Pelosi for Congress 7 0 0 seat total_received state lobbying_firm count_received party 1 federal:house 14173534 CA none 10054 D 2 none 0 none 0 none total_given type id 1 0 politician 85ab2e74589a414495d18cc7a9233981 2 7250 organization afb432ec90454c8a83a3113061e7be27 non_firm_spending is_superpac 1 0 none 2 0 {% endhighlight %}

Return the top contributoring organizations, ranked by total dollars given. An organization's giving is broken down into money given directly (by the organization's PAC) versus money given by individuals employed by or associated with the organization. {% highlight r %} out <- sll_ts_aggregatetopcontribs(id = "85ab2e74589a414495d18cc7a9233981", key = sunlightkey) ldply(out, data.frame) {% endhighlight %}

{% highlight text %} employee_amount total_amount total_count 1 64000.00 101300.00 79 2 3500.00 90000.00 29 3 0 86600.00 48 4 0 85000.00 32 5 0 82500.00 37 6 0 77000.00 19 7 0 76000.00 36 8 0 72500.00 18 9 0 72500.00 21 10 8000.00 72000.00 31 name direct_count employee_count 1 Akin, Gump et al 16 63 2 American Fedn of St/Cnty/Munic Employees 25 4 3 National Assn of Realtors 48 0 4 United Auto Workers 32 0 5 National Education Assn 37 0 6 Teamsters Union 19 0 7 National Assn of Letter Carriers 36 0 8 Intl Brotherhood of Electrical Workers 18 0 9 Sheet Metal Workers Union 21 0 10 Wells Fargo 20 11 id direct_amount 1 105dcfc8c9384875a099af230dad9917 37300.00 2 fb702029157e4c7c887172eba71c66c5 86500.00 3 bb98402bd4d3471cad392a671ecd733a 86600.00 4 4d3167b97c9f48deb433aad57bb0ee44 85000.00 5 1b8fea7e453d4e75841eac48ff9df550 82500.00 6 f89c8e3ab2b44f72971f91b764868231 77000.00 7 390767dc6b4b491ca775b1bdf8a36eea 76000.00 8 b53b4ad137d743a996f4d7467700fc88 72500.00 9 425be85642b24cc2bc3d8a0bb3c7bc92 72500.00 10 793070ae7f5e42c2a76a58663a588f3d 64000.00 {% endhighlight %}