**Introduction**

During a recent negotiation of an informed consent form for use in a clinical trial, the opposing lawyer and I skirmished over the applicability of the Genetic Information Nondiscrimination Act of 2008, commonly known as GINA. Specifically, the opposing lawyer thought that [guidance issued by the U.S. Office for Human Research Protections in 2009](https://www.hhs.gov/ohrp/regulations-and-policy/guidance/guidance-on-genetic-information-nondiscrimination-act/index.html) was now outdated, in part because enforcement efforts were erratic. The argument was primarily driven by policy, rather than data.

Being a data-driven guy, I wanted to see whether the data supported the argument advanced by the other lawyer. Fortunately, the U.S. Equal Employment Opportunity Commission (EEOC), which is responsible for administering GINA complaints, maintains [statistics regarding GINA claims and resolutions](https://www.eeoc.gov/laws/types/genetic.cfm). I’m not great at making sense of numbers in a table, so I thought this presented the perfect opportunity to rvest some data!

libraries <- c("tidyverse", "rvest", "magrittr")

lapply(libraries, require, character.only = TRUE)

**Data Scraping**

In standard rvest fashion, we’ll read a url, extract the table containing the GINA enforcement statistics, and then do some data cleaning. Once we read the table and gather all of the annual results into key/pair of year/value, we get the following results:

url <- "https://www.eeoc.gov/eeoc/statistics/enforcement/genetic.cfm"

GINA.resolutions <- read\_html(url) %>%

html\_nodes("table") %>%

extract2(1) %>%

html\_table(trim = TRUE, fill = TRUE, header = TRUE)

names(GINA.resolutions)[1] <- "metric" # Top left table cell is blank, will throw errors

names(GINA.resolutions) <- gsub("FY (.+)", "\\1", names(GINA.resolutions)) # Remove FY from year so we can convert to numeric

GINA.resolutions <- GINA.resolutions %>%

filter(! metric == "") %>% # Remove percentage rows

filter(! metric == "Resolutions By Type") %>% # Remove blank line

gather(year, value, 2:9) %>% # short and wide data to tall and skinny

mutate(

year = as.integer(year),

value = gsub("[\\$\\%]", "", value)

) %>%

mutate(

value = as.numeric(value)

) %>%

as.tibble()

GINA.resolutions

## # A tibble: 88 x 3

## metric year value

##

## 1 Receipts 2010 201

## 2 Resolutions 2010 56

## 3 Settlements 2010 3

## 4 Withdrawals w/Benefits 2010 2

## 5 Administrative Closures 2010 11

## 6 No Reasonable Cause 2010 38

## 7 Reasonable Cause 2010 2

## 8 Successful Conciliations 2010 1

## 9 Unsuccessful Conciliations 2010 1

## 10 Merit Resolutions 2010 7

## # ... with 78 more rows

**Claim Numbers over Time**

Now that we have the data in a format we can use, we’ll look at the volume of claims and resolutions over time:

GINA.resolutions %>%

filter(metric == "Receipts" | metric == "Resolutions") %>%

ggplot(aes(year, value, color = metric)) +

geom\_line() +

labs(

title = "EEOC Enforcement of GINA Charges",

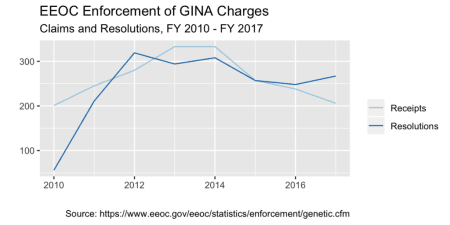
subtitle = "Claims and Resolutions, FY 2010 - FY 2017",

caption = paste0("Source: ", url),

x = "", y = ""

) +

scale\_color\_brewer("", palette = "Paired")

The number of GINA claims rose for the first few years, but then declined down to enactment-year levels. This could represent support for the opposing lawyer’s argument that enforcement is waning. However, it could just as likely be showing that the deterrent effect of the law has proven effective, and most firms subject to GINA are now aware of the law and have taken appropriate steps toward compliance. Given these competing interpretations, we’ll need to look at little deeper to see if we can derive trends from the data.

**GINA Claim Resolutions**

One of the arguments made by the opposing lawyer is that the Obama administration was pushing GINA enforcement, and that the Trump administration hates the law and won’t enforce it. We can look at the resolution types to test this hypothesis:

GINA.resolutions %>%

filter(metric != "Receipts" & metric != "Resolutions") %>%

ggplot(aes(year, value)) +

geom\_line() +

facet\_wrap(~ metric, scales = "free\_y") +

labs(

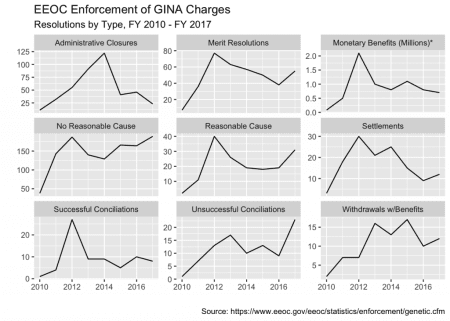
title = "EEOC Enforcement of GINA Charges",

subtitle = "Resolutions by Type, FY 2010 - FY 2017",

caption = paste0("Source: ", url),

x = "", y = ""

)

[](https://i0.wp.com/www.nathanchaney.com/wp-content/uploads/2018/07/resolutions-by-type-1.png)In 2017, the first year of the Trump administration, administrative closures were down and resolutions on the merits were up, which contradicts the opposing lawyer’s argument. While findings of no reasonable cause were up about 10-15%, findings of reasonable cause were up 50%; if anything, this also contradicts the opposing lawyer’s argument. Monetary settlements appear to be relatively flat from 2013 – 2017, and in any event a million dollars isn’t a ton of money in light of the [EEOC’s annual budget of about $376 million](https://www.eeoc.gov/eeoc/plan/2017budget.cfm) (note that the EEOC handles many other types of charges besides GINA).

The resolution type that jumped most markedly in 2017 was “unsuccessful conciliation.” A conciliation is where the EEOC “attempt[s] to achieve a just resolution of all violations found and to obtain agreement that the respondent will eliminate the unlawful employment practice and provide appropriate affirmative relief.” [29 C.F.R. § 1601.24](https://www.law.cornell.edu/cfr/text/29/1601.24). It’s unclear why this jump occurred from the summary statistics provided by the EEOC.

Finally, I thought it was useful to plot all the resolution types together to show relative numbers:

GINA.resolutions %>%

filter(metric != "Receipts" & metric != "Resolutions" & metric != "Monetary Benefits (Millions)\*") %>%

ggplot(aes(year, value, color = metric)) +

geom\_line() +

labs(

title = "EEOC Enforcement of GINA Charges",

subtitle = "Resolutions by Type, FY 2010 - FY 2017",

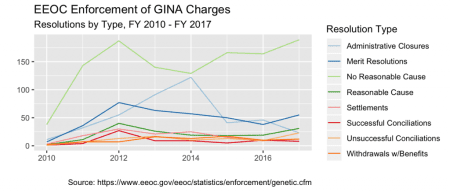
caption = paste0("Source: ", url),

x = "", y = ""

) +

# scale\_y\_sqrt() +

scale\_color\_brewer("Resolution Type", palette="Paired")

[](https://i1.wp.com/www.nathanchaney.com/wp-content/uploads/2018/07/resolutions-1.png)From this perspective, it does look like the long-term trend has been for the EEOC to dismiss a majority of GINA-related charges as unfounded (i.e., no reasonable cause). However, for the cases that do have merit, it appears the EEOC has reversed an initial trend showing a preference toward administrative closure.

**Conclusion**

In all, I didn’t find the opposing lawyer’s argument particularly compelling in light of the data from President Trump’s first year in office. However, the first month of 2017 was President Obama’s last in office, and there was a flurry of activity by many regulatory agencies. It wouldn’t surprise me if EEOC also participated in a high volume of lame-duck activity, and a lot of activity in January 2017 could haved skewed the annual results. Monthly statistics would be nice but didn’t appear to be readily available. The goal with any R project is for it to be repeatable with additional data, so it will be interesting to see what the data from FY2018 shows.

This wasn’t a particularly complicated coding project – in fact, this writeup took me longer to produce than writing the actual code and coming to conclusions about whether GINA is on its last leg or not. Despite that fact, I thought it was a good example of how data science can be used to inform solutions to simple as well as complex problems.