Amber Pasiak

Data Analysis

INFO 640

Final Project

12/14/19

Audience: Basic/introduction knowledge of data analysis

**The Future of FOIA:**

**Forecasting Requests Released Across Administrations**

**Introduction and Literary Review:**

According to the FOIA.gov website, The Freedom of Information Act (FOIA) is “often described as the law that keeps citizens in the know about their government. Federal agencies are required to disclose any information requested under the FOIA, unless it falls under one of nine exemptions which protect interests such as personal privacy, national security, and law enforcement.” (FOIA.gov) FOIA itself does not change under each new presidential administration, however the way in which agencies and departments respond to FOIA requests can be altered with each new administration. This has led many to question whether or not opposing political parties affect FOIA requests differently.

In 2016, Ben Wasike conducted a study, ““FoIA in the age of “Open. Gov”: An analysis of the performance of the Freedom of Information Act under the Obama and Bush administrations”. This study took a deep look into the FOIA performance of both former presidents Barack Obama and George W. Bush. Wasike found that contrary to what he terms, media outcry, neither administration used the amount of exemptions that had been claimed. Wasike did however, note a trend within both administrations dealing with a decline in government transparency. (Wasike, 2016) Wasike concludes his finding stating that: “Overall, the study returned mixed results for FoIA performance. While requests were processed much faster under Obama on the most part, they were also more likely to be honored in part or be denied in full. However, success rates on appeal were better under him than under Bush. Both presidents also had similar rates of invoking exemptions. However, FoIA staffers face smaller workloads under Obama meaning that requests are getting more attention from staffers. These last two findings bode well for information seekers, meaning that requests get processed faster and assumedly with more attention given the lower workload. A notable finding was that neither president has invoked national security at the rate that critics have claimed.” (Wasike, 2016) This is an extremely interesting finding due in part to the political standing that the United States was in during each Presidency. Under Bush the 9/11 attacks happened causing him to request the removal of public information from a number of government websites and tighten control over FOIA releases. Obama, on the other hand, made public claims striving to improve transparency of the government, yet the ratio for FOIA release remained comparable to that of Bush.

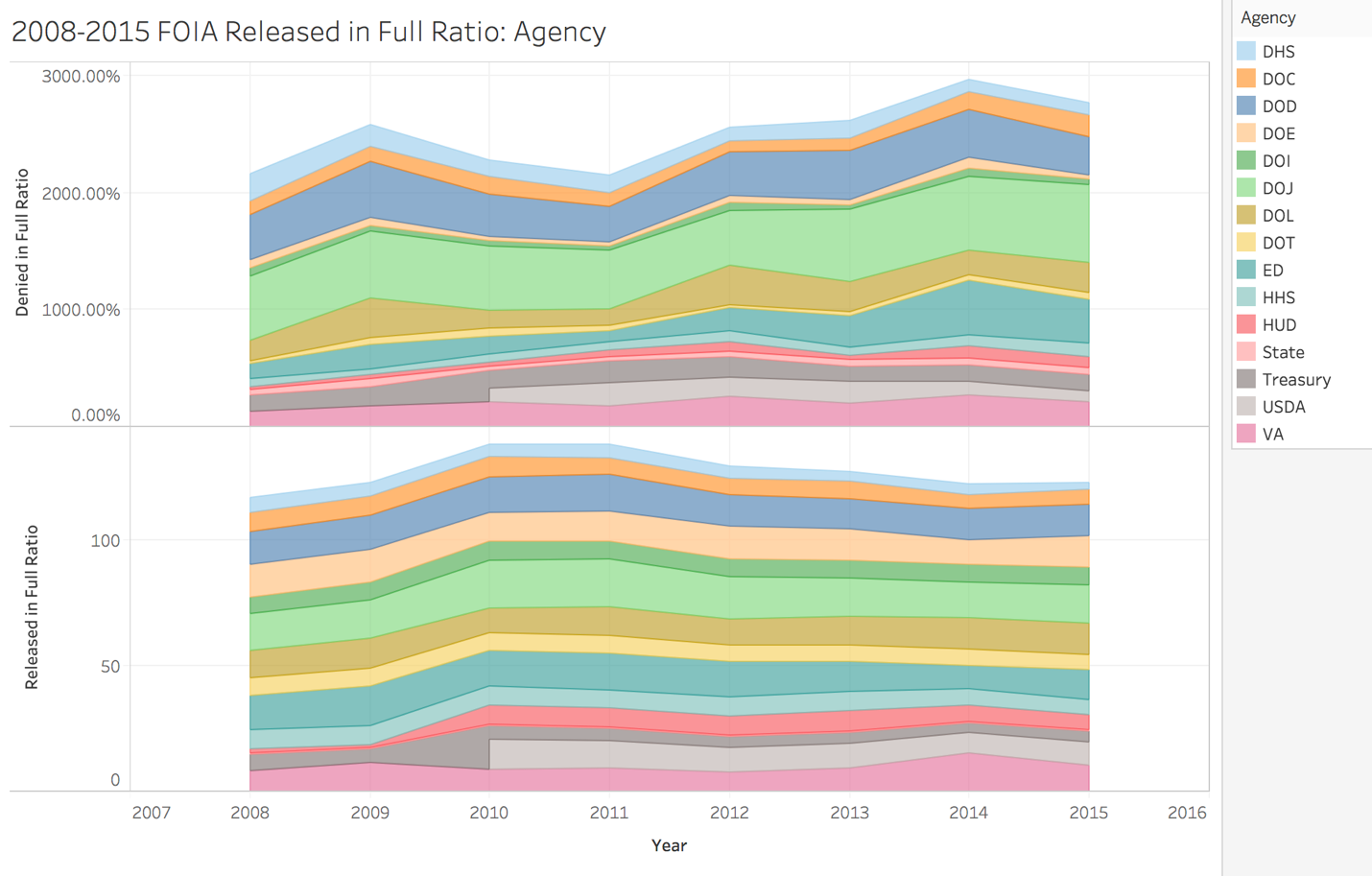
With digitally born documents growing and the OPEN act being passed into law, one would assume that government documentation would be easier for the public to access. The OPEN act, which stands for Open, Public, Electronic and Necessary was first introduced in 2017 under former President Obama and was officially enacted in 2019, under President Trump in the beginning of his first year of presidency. This act is meant to promote standards for open government data across departments and encourage evidence-based polices. Yet, in 2017, Ted Bridis, for PBS, published an article stating that during Obama’s presidency, he set the record for lawsuits in defending the refusal of FOIA releases. In a First Amendment Watch article, a similar statement has been made about President Trump, claiming that he also has increased the challenges for the public to access records. If these presidents, across different political parties, are both actively endorsing government transparency and open data why are journalists claiming that it is actually becoming harder to attain information?

Well, according to Harold Relyea in “Federal freedom of information policy: Highlights of recent developments”, FOIA was neither supported nor received well since it was enacted into law in 1966. (Relyea, 2009) Many scholars have approached the topic of declining government transparency with a multitude of different theoretical methods and looks into political parties, yet the outlook remains hopeful that with each new administration access will become easier. This project aims to look at the projection of FOIA releases over time and examine a forecast of the current trends. If FOIA has never been fully supported by law makers and the overall consensus is that more needs to be done to promote government transparency, what does that say about the US during one of the most controversial presidents? Through a data analysis in RStudio and visualizations in Tableau this project will investigate FOIA release trends and question the implications for the future of FOIA.

**Methods Statement and Data Analysis:**

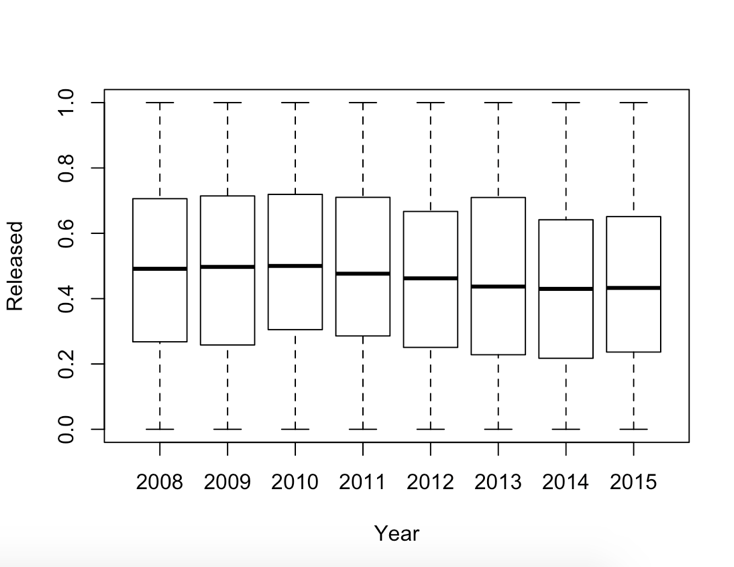
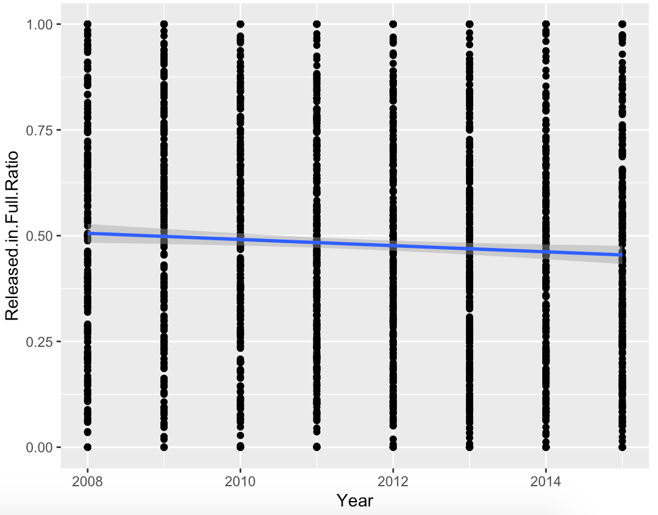
The US government has open data on FOIA dating from 2008 to 2018. This data is sectioned off by agency and then by the complication of a release. When uploaded into R, the data is all in character form and not numeric. Due to these inconstancies I will be using only an open dataset from Kaggle that ranges from 2008-2015. With this limitation in mind, a predictive data analysis, or more specifically a quantitative forecast, will be performed. As I would be unable to directly compare the Kaggle data to the FOIA.gov data of 2015-2018 because the variables are so different and the project time constraint, this project will analyze President Obama’s FOIA release time from 2008-2015 and then through R make a prediction on the future progression of release time based on a continuation of the trends. The first research question is whether or not there is an overall trend within the FOIA data available. The second question asks what a projection of the current trend would look like.

The dataset chosen has 11 variables: Agency, Department, Year, Requests pending at the start of the year, Requests pending at the end of the year, Requests received, Requests processed, Released in full, Released in part, Denied in full. The released variables are in a ratio format. I was unable to find how this was computed in the metadata, which is another limitation this dataset has. The variables that I will be taking a deeper look at include: Agency, Released in full, Denied in full, and Year.

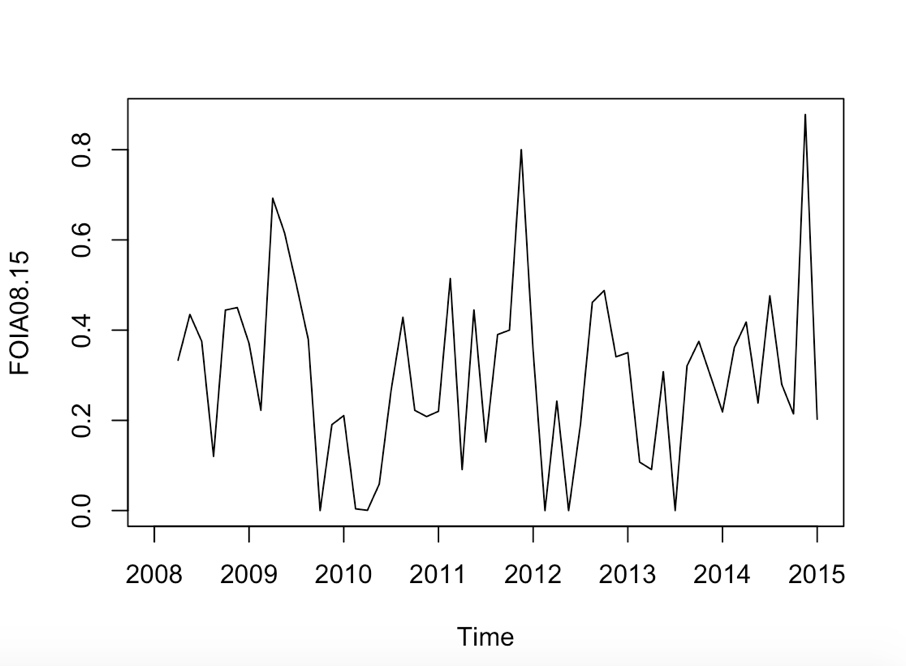


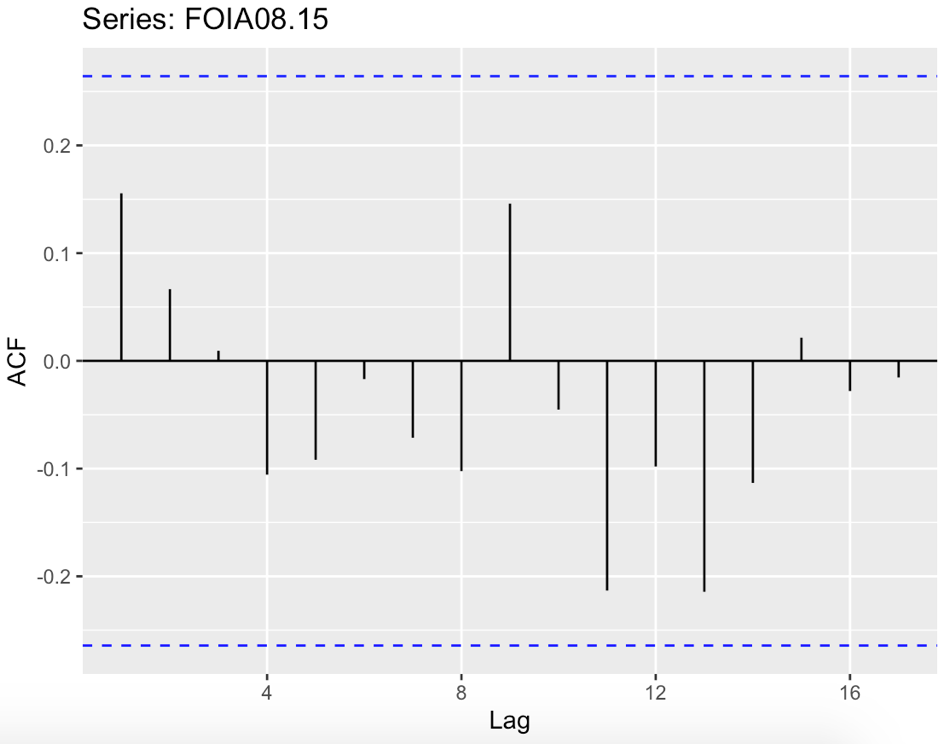
After making sure that dataset was cleaned and uniform, I ran a summary and glimpse function to better understand the information. I found that the mean for Released in Full was .4794, Denied in Full was .09465 and the median of requests received was 105 with a mean of 1854. I also found that there are over 2,000 departments in the dataset and that the average difference of requests still pending at the end of the year was 88,275. I then created a graph in Tableau to look at the differences between Released in Full and Denied in Full throughout the years. It is significant to note that an increase in denials does not one hundred percent correlate to a decrease in releases, there may be similar underlying factors that affect both, however this cannot be fund 3within this dataset. Due to this I chose to look more closely at the Released in Full variable. I next ran a regression line in RStudio to investigate any trends. I found that there was decrease over time of FOIA releases. As I do not have any metadata to associate to each year, causes for the decrease can only be speculation. I would have liked to create a separate timeline with possible reasons for this, however investigating each department on a yearly basis was not possible. It is important to note that this analysis is of the averages of all of FOIA and individual departments may be affected differently with each new regulation, political party and employees.

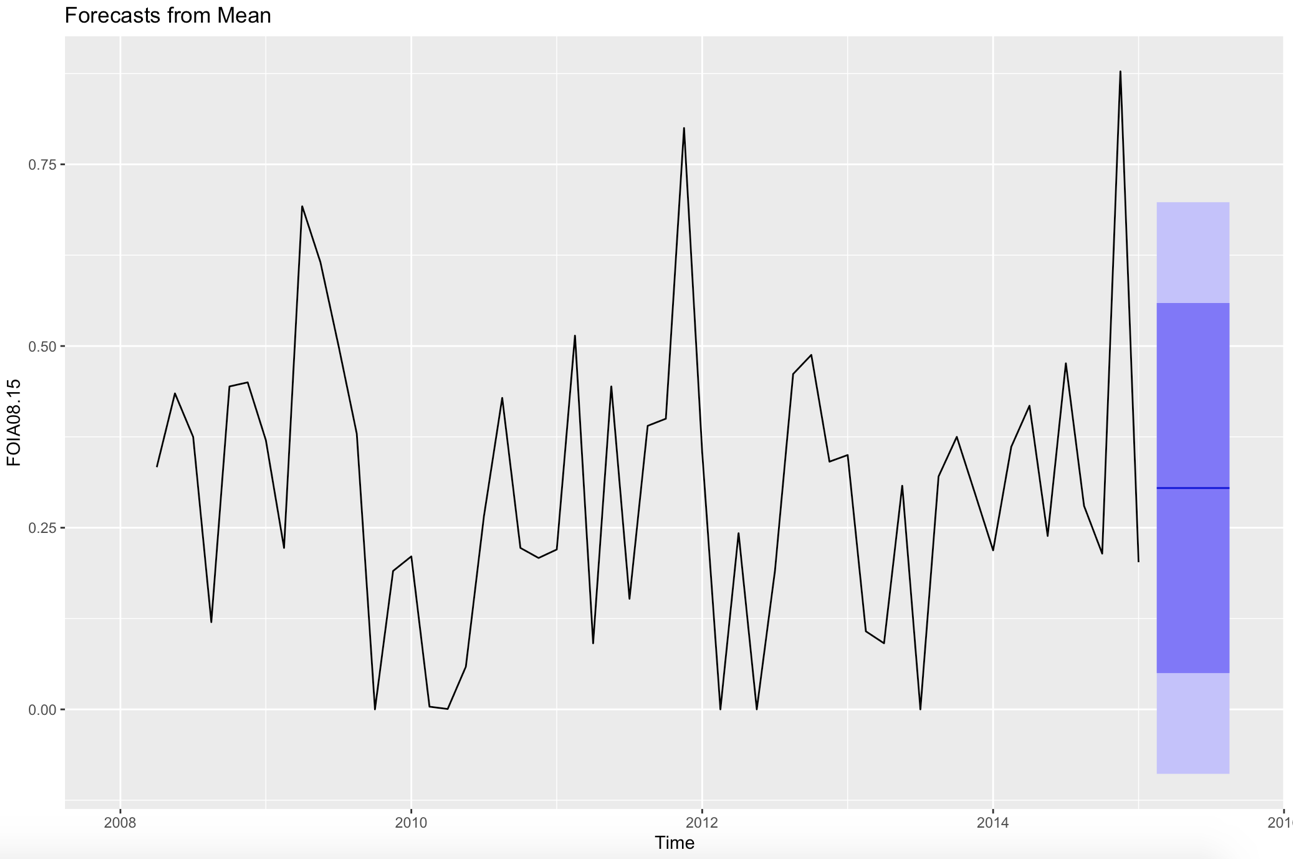
Other main observations I made through a linear model with a regression line and box plot, include a sharp change around the year 2012 in releases, with a steady decrease after. I did find information relating to a mandate signed by Obama around this time that made alterations to the FOIA defense exemption, this possibly could have influenced such a change in releases. (FOIAproject.org) Next I wanted to see if there were any outliers that could be affecting the release averages but found that the departments had a mostly consistent and steady decrease in releases with no extreme values. This can also be seen in the first graph.

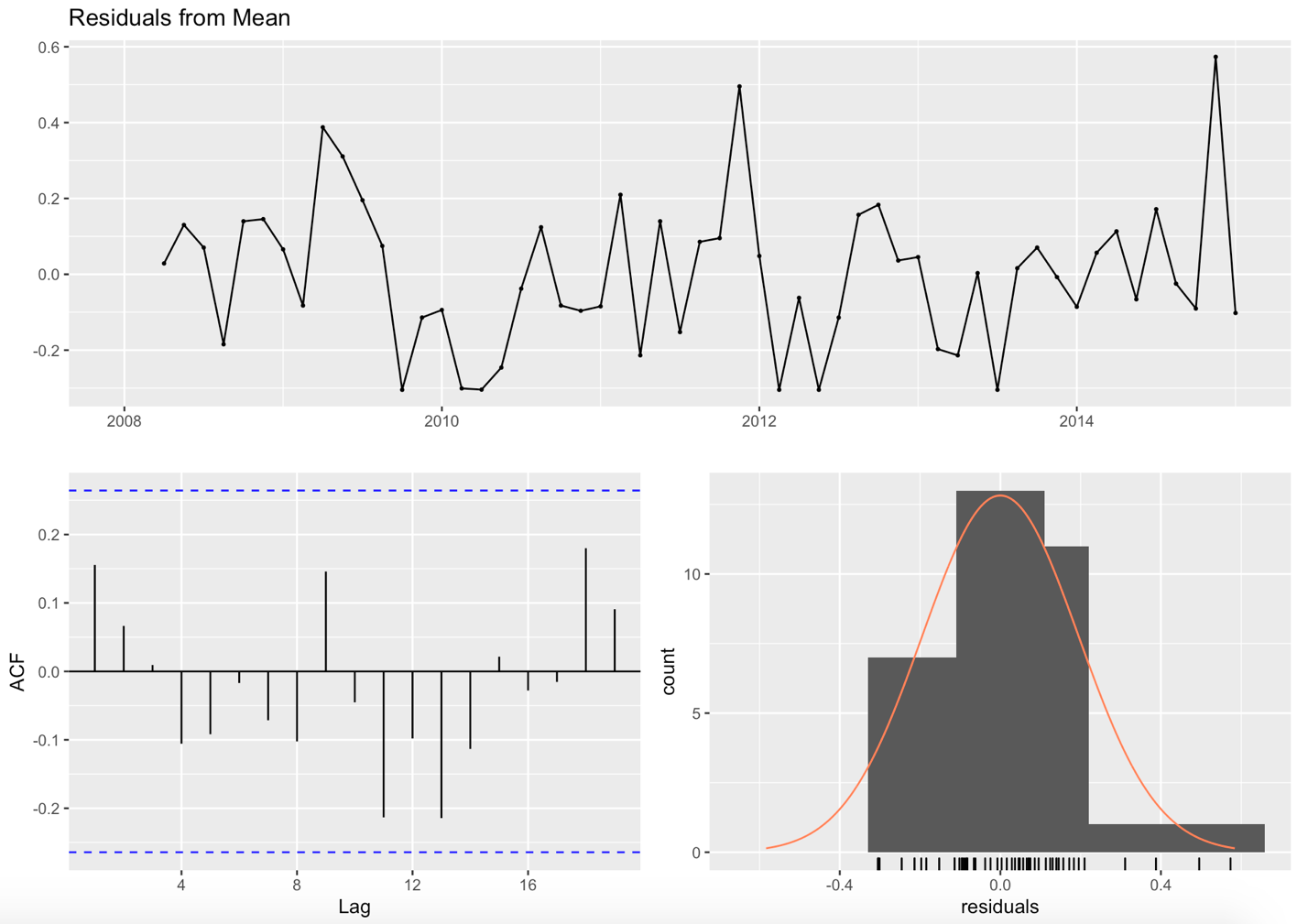


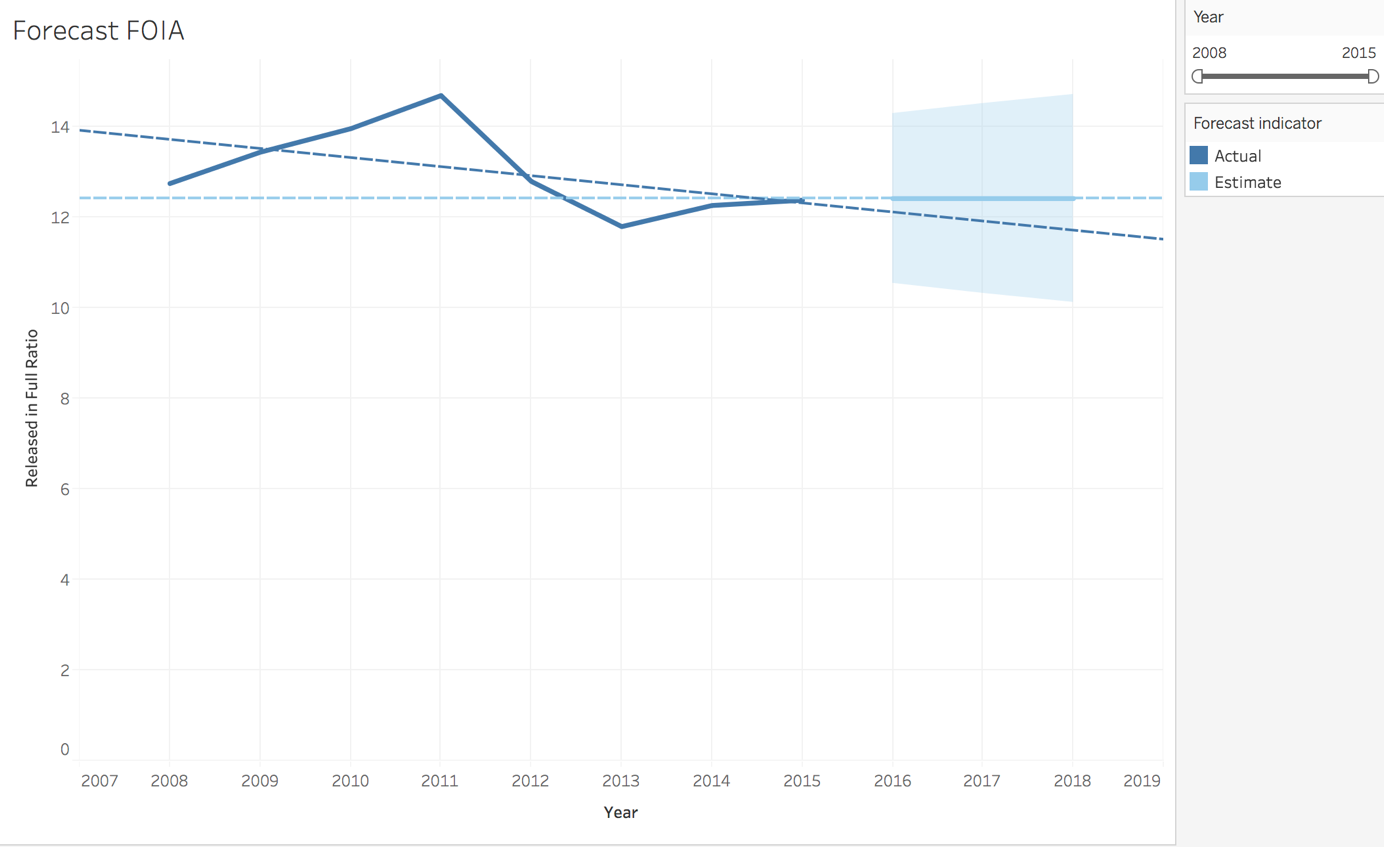
The next step was to inspect what it would look like if this trend continued. I did this two ways. The first was to create a time series in RStudio and the second was to use a Forecasting function in Tableau. Understanding fully the uncertainty of forecasting this analysis is based on the trend and pattern of decreasing releases already found in the data.

In order to create the time series, I had to find the mean, .4940, the standard deviation, .2788, and set the frequency, 8 years. The time series showed that the trend was not seasonal and had irregular fluctuations. I originally made the mistake of trying to smooth the series and use the HoltWinters function to minimize a prediction error, but quickly found out that this function works best with seasonal data. Next I ran an autocorrelation function (ACF) to measure how strong the correlation between the current series was to itself. To do this I began with the Box.Test function and found a p-value of .2486. The p-value suggests that this series is not serially correlated, and that the observed result thus far is not any more or less extreme than the alpha level. The correlations show tend to jump around without a solid pattern, and all of the lines fall within the dashed lines, indicating no statistical significance to any of the correlations.

Next I used the SMA function to smooth the forecast and create a more accurate window for the predicted average. Then I used the mean method to forecast future values. The h argument for this function is what specifies how far into the future the forecast will go, I chose 5 years. The concluding graph shows a dip in the mean with the prediction intervals being lower than the histo0rical data. After this I wanted to double check the accuracy of this plot. I ran an accuracy assessment by looking again at residuals and doing a Ljung-Box Test. The final few graphs show how the predicted performance of FOIA continues to decrease and that this prediction falls within an accurate range. The very last image is from Tableau and depicts a forecast with similar results to those found in RStudio. Both concluding that if FOIA continues to follow the trend from 2008-2015, a decrease in requests released is the most likely scenario.







**Conclusion and further research:**

In the introduction to Akhlaque Haque’s book “Surveillance, Transparency, and Democracy: Public Administration in the Information Age”, the romanticized vision of the world wide web is eloquently described as “technology and technique needs to be open to diversity of values so it can seamlessly integrate with the society it ought to serve.” (Haque, 2015). Haque goes on to describe the relationship between information and administrations, stating that there is a lack of balance between those providing policy guidance and the goals intended for the general public. This is the struggle that reporters and scholars are finding within FOIA. FOIA was mandated in order to provide facts to the public about government policy, administrations and changes.

In an era of rampant fake news accusations and collusion scandals the government could have the opportunity to provide a reliable narrative and show accountability for certain actions. FOIA provides the public with the option of finding information without any distortion. But what happens if that option is slowly being chiseled away? Through an analysis of FOIA requests released from 2008-2015, I have found the unsettling trend showing that FOIA releases are decreasing with a continuation projected in the future. When a comparison can be made between Obama and Trump and to see if the prediction was correct, another analysis would be beneficial. A look into multiple administrations could either reject the hypothesis that even in the Information Age, information is getting harder to access regardless of political party or show how different administrations actually affect FOIA.

For now, though, there is a sense that the reports are correct when they fear that the US government is using FOIA to move further away from proactive disclosure of information and access to government documents. An imbalance of power exaggerates the difficulties of a true democracy with an informed public. This project also begs the question of causation. The research into individual departments and the changing uses of exemptions would be extremely interesting to see in direct correlation to requests released. If FOIA is decreasing access across all administrations and departments, what does that say about government transparency and accountability?

**Code:**

library(lubridate)

library(dplyr)

library(tidyverse)

library(gmodels)

library (broom)

library (GGally)

library (ggplot2)

install.packages("ggplot2")

install.packages("rcompanion")

install.packages("forecast")

install.packages("broom")

install.packages("GGally")

install.packages("quadprog")

library(forecast)

library(quadprog)

FOIA <- read.csv("../amberpasiak/Desktop/2019-Pasiak-DATALIB628-FOIA/CSV/requests.csv", header = TRUE)

summary(FOIA)

glimpse(FOIA)

head(FOIA)

tail(FOIA)

summary(FOIA$Released.in.Full.Ratio)

summary(FOIA$Denied.in.Full.Ratio)

summary(FOIA$Department)

summary(FOIA$Agency)

#summary of all requests in full

Requests\_FullNONA <- na.omit(FOIA$Released.in.Full.Ratio)

head(Requests\_FullNONA)

hist(Requests\_FullNONA)

#filtering for year 2008 only

summary(FOIA$Year)

Requests\_Year <- select (FOIA, Year)

head(Requests\_Year)

FOIA %>% arrange (desc(Year))

Requests\_2008 <- FOIA %>%

filter (Year == "2008") %>%

summary(Requests\_2008)

#filtering for visual by years

ggplot(FOIA, aes (x= Year, y= Released.in.Full.Ratio)) +

geom\_point() + geom\_smooth(method = "lm")

plotmeans(Released.in.Full.Ratio ~ Year, data = FOIA, frame = FALSE)#didnt work

hist(FOIA$Released.in.Full.Ratio,

col="gray",

main="",

xlab="Released Full")

library(rcompanion)

x = FOIA$Released.in.Full.Ratio

plotDensityHistogram(x,

adjust = 1)

hist(FOIA$Year,

col="blue",

main="",

xlab="Year")

boxplot(Released.in.Full.Ratio ~ Year,

data=FOIA,

ylab="Released")

mean(FOIA$Released.in.Full.Ratio, na.rm = TRUE)

sd(FOIA$Released.in.Full.Ratio, na.rm = TRUE)

by(FOIA$Released.in.Full.Ratio, FOIA$Year, mean, na.rm= TRUE)

summary(FOIA$Year)

Requests\_Year <- select (FOIA, Year)

head(Requests\_Year)

FOIA %>% arrange (desc(Year))

Requests\_2008 <- FOIA %>%

filter (Year == "2008")

summary(Requests\_2008)

R2008 <- rnorm(329, mean= .4940, sd= .2788)

glimpse(R2008)

hist(R2008)

ggplot(FOIA %>%

filter (Year == "2008"), aes(x= FOIA$Year, y= FOIA$Released.in.Full.Ratio)) + geom\_point() + stat\_smooth(method = "lm")

#switching to predictive

mean(FOIA\_DoE$Released.in.Full.Ratio, na.rm = TRUE)

sd(FOIA\_DoE$Released.in.Full.Ratio, na.rm = TRUE)

ts(FOIA$Year, start = 2008, end = 2015,frequency = 8 )

plot(FOIA)

FOIA08.15 <- ts(FOIA$Released.in.Full.Ratio, start = 2008, end = 2015,frequency = 8 )

plot(FOIA08.15)

FOIA08.15P <- rnorm(329, mean= .6741, sd= .2225)

head(FOIA08.15P)

frequency(FOIA08.15)

Freq\_8.15 <- ts(FOIA08.15P, start= 2008, end = 2015, frequency = 8)

plot(Freq\_8.15)

plot.ts(FOIA08.15)

Box.test(FOIA08.15)

acf(log(FOIA$Released.in.Full.Ratio)

ggAcf(FOIA08.15)

ggAcf(

FOIA08.15,

lag.max = NULL,

type = c("correlation", "covariance", "partial"),

plot = TRUE,

na.action = na.contiguous,

demean = TRUE,)

install.packages("TTR")

library(TTR)

sm5 <- SMA(FOIA08.15, gausnoise, n=5)

mFOIA <- meanf(FOIA08.15, h=5)

autoplot(mFOIA)

checkresiduals(mFOIA)

**References:**

Bridis, T. (2017, March 14). In Obama’s final year, U.S. spent $36 million in FOIA lawsuits. *PBS.* Retrieved from <https://www.pbs.org/newshour/nation/obamas-final-year-u-s-spent-36-million-foia-lawsuits>

Department of Justice. (2017). Freedom of Information Act Requests. Kaggle. Retrieved from <https://www.kaggle.com/doj/foia-requests/metadata>

First Amendment Watch at New York University. The Limits of Transparency and FOIA Under Trump. Retrieved from <https://firstamendmentwatch.org/deep-dive/the-limits-of-transparency-and-foia-under-trump/>

Fischetti, T. (2018). D*ata Analysis with R*. Birmingham: Packt Publishing.

FOIA.gov. Retrieved from [www.FOIA.gov](http://www.FOIA.gov)

FOIA Project. Freeing Information through public accountability. Retrieved from http://foiaproject.org/2012/03/01/defensive-standards-hinder-foia-openness/

Haque, A. (2015). *Surveillance, Transparency, and Democracy: Public Administration in the Information Age*. Tuscaloosa: University Alabama Press. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1021673&site=eds-live&scope=site>

Nayek, J. K. (2018). Evaluation of Open Data Government Sites: A Comparative Study. *Library Philosophy & Practice*, 1–23. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=ofm&AN=133873696&site=eds-live&scope=site

The United States Department of Justice. Retrieved from <https://www.justice.gov/oip/foia-library>

Relyea, H. (2009, April). Federal freedom of information policy: Highlights of recent developments. *Government Information Quarterly,* 314-320. Retrieved from <https://doi.org/10.1016/j.giq.2008.12.001>

Wasike, B. (2016, July). FoIA in the age of “Open. Gov”: An analysis of the performance of the Freedom of Information Act under the Obama and Bush administrations. *Government Informati9on Quarterly.* 417 – 426. Retrieved from https://doi.org/10.1016/j.giq.2016.05.001