Exploring Federal Employment Data Over Presidential Terms

Jenn Le

September 17, 2017

Abstract

Add an abstract/executive summary (¼ page) that introduces the problem and highlights the major results. Be concrete with your major results (e.g., "This report shows that...").

Contents

1 Business Understanding **2**

2 Data Understanding **2**

2.1 Collecting Data 2

3 Data Preparation **3**

4 Modeling **3**

5 Evaluation **3**

1 Business Understanding

Data Source: https://archive.org/details/opm-federal-employment-data

This data contains federal employment information over the course of President George W. Bush's and President Barack Obama's terms in office.

2 Data Understanding

I examined the status data of government employees, excluding the Department of Defense, for the years 2001 - 2014. This includes Bush's entire presidency and all but the last two years of Obama's.

The main code file I used is shown below and outlines the steps I took in my investigation of the data.

../code/project-code.R

```
# Importing code stored in other files
source("collect-data.R")

# Storing the years I will be examining, split up by President
bush_years <- c(2001:2008)
obama_years <- c(2009:2014)

bush_data <- lapply(bush_years, collect)
```

2.1 Collecting Data

Below is the function that I used to collect the text data for a specific year, concatenate the files for each quarter, and convert it into a dataframe. I used the headers files that was provided to label the columns in the dataframe.

../code/collect-data.R

```
require(data.table)

collect_dat <- function(year) {
  headers <- read.csv("headers.csv", header = TRUE)
  base_url <- paste("https://archive.org/download/opm-federal-employment-data/data/1973-09-to
      ↪ -2014-06/non-dod/status/Status_Non_DoD_", year.toString, sep="")
  url_ext <- c("_03.txt", "_06.txt", "_09.txt", "_12.txt")
  urls <- paste(base_url, url_ext, sep="")

  # The data frame where we'll store the year's data
  convert_file <- function(url, test = TRUE) {
    if(test == FALSE) {
      dat_raw <- readLines(url)
```

```
    } else {
      dat_raw <- readLines(url, n = 5)
    }
    dat_curr <- t(sapply(dat_raw, FUN = function(x) trimws(substring(x, headers[,2], headers[,3]))))
    dimnames(dat_curr) <- NULL
    dat_curr <- as.data.frame(dat_curr)
    colnames(dat_curr) <- headers[,1]
    return(dat_curr)
  }

  # Convert all the text files to dataframe
  dat_list <- lapply(urls, convert_file)
  dat <- rbindlist(dat_list)

  return(dat)
}
```

3 Data Preparation

4 Modeling

5 Evaluation