Exploring Federal Employment Data Over Presidential Terms

Jenn Le

October 11, 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 necessary libraries
library(tidyverse)

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

# Storing the years I will be examining, split up by President
dat.2000 <− CollectDat(2000)
dat.2001 <− CollectDat(2001)
dat.2002 <− CollectDat(2002)
dat.2003 <− CollectDat(2003)
dat.2004 <− CollectDat(2004)
dat.2005 <− CollectDat(2005)
dat.2006 <− CollectDat(2006)
dat.2007 <− CollectDat(2007)
dat.2008 <− CollectDat(2008)
dat.2009 <− CollectDat(2009)
dat.2010 <− CollectDat(2010)
dat.2011 <− CollectDat(2011)
dat.2012 <− CollectDat(2012)
dat.2013 <− CollectDat(2013)
dat.2014 <− CollectDat(2014)

load(file = "../dat.rda")
```

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

```r
require(data.table)

CollectDat <− function(year) {
  headers <− read.csv("../data/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, sep="")
  url.ext <− if (year < 2014) c("_03.txt", "_06.txt", "_09.txt", "_12.txt") else c("_03.txt", "_06.txt")
  urls <− paste(base.url, url.ext, sep="")

  # The data frame where we'll store the year's data
  ConvertFile <− function(url, test = FALSE) {
    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, ConvertFile)
  dat <− rbindlist(dat.list)

  return(dat)
}
```

3 Data Preparation

4 Modeling

5 Evaluation