Exploring Federal Employment Data Over Presidential Terms

Jenn Le

September 25, 2017

Abstract

This is an investigation of federal employment data over the course of President George W. Bush's and President Barack H. Obama's presidencies. The data was analyzed alongside both presidents' policies and major events that happened during their terms in office to determine if there is an effect on government payroll. This report shows that major administration changes such as the establishment of the Department of Homeland Security are clearly reflected in payroll data, with large increases in hiring and specific pay plans. However, some policies such as social security, did not affect employment data in ways we would expect.

Contents

1 Business Understanding

This data contains federal employment information over the course of President George W. Bush's and President Barack H. Obama's terms in office. I will be examining the data alongside each president's policies as well as major events that occured during their terms in office and how they affect employment within our federal government. In addition, my investigation will include a comparison of each president's campaign platform and policies and the federal employment data to see how they reflect in the government's payroll. President Bush's presidency lasted two terms (2001 - 2008) and so did President Obama's (2009-2016). However, the data is only available up to the second quarter of 2014. Six years is still a significant amount of data so for President Obama, I will only focus on major events that happened prior to 2014. Aside from agencies relating to each presidents' policies, I will also be analyzing general employment data across all the agencies.

1.1 The Bush Presidency

1.1.1 2000 Presidential Campaign

The major points of President Bush's platform included economic reforms, the No Child Left Behind policy, foreign affairs, and energy reform. His economic platform centered on tax breaks for all as well as low mortage rates to encourage homeownership. In addition, he looked to expand health care, social security, and medicare. He also wanted to assist small businesses by ending the harassment of small businesses by federal agencies such as OSHA and the IRS. His education reform focused on The No Child Left Behind policy which included raising academic standards through national testing, increased funding for schools, requiring teachers to have relevant degrees, and assistance for needy families to send their kids to schools of their choosing. With regards to foreign affairs, President Bush took a non-interventionist stance and promised no nation building. Another focus of his platform was an energy reform bill, the National Energy Security Act that included incentives for domestically produced enery, especially renewable energy, and a reduction of dependence on foreign oil [1].

From this platform, we can expect to see increases in personel and funding for the Department of Education, the Centers for Medicare and Medicaid Services, the Department of Health and Human Services, the Department of Energy, and the Social Security Administration. Consequently, we can also expect reductions for the U.S. Agency for International Development, the

Internal Revenue Service, and the Environmental Protection Agency.

1.1.2 Major Events and Policies in Office

Not long after Bush entered office, the terrorist attacks on September 11, 2001 occured and the Department of Homeland Security was created in response [4]. With the Transportation Security Administration as a focus, the creation and expansion of the DHS are events that I expect to be obviously reflected in the data as well as a readjustment of government employees as a result of it.

Under Bush, foreign assistance was a large focus in order to alleviate the roots of extremism: hunger, disease, and poverty. Official development assistance more than doubled during his time in office and he also took steps to fight diseases such as HIV/AIDS, malaria, and typically neglected tropical diseases [3]. Therefore, contrary to the expectation I drew from his campaign platform, it seems like reasources for the U.S. Agency for International Development would have increased during his presidency.

In 2003, Bush declared war on Iraq and soon after reduced his tax cut plan to less than half of the original amount in order to fund the war. A few months later, however, he signs the third largest tax cut in history in an attempt to reverse umemployment [4].

In the midst of the 2008 financial crisis, Bush signed the largest bailout in history as well as bailing out GM and Chrysler. The interest rates are set at 0% to try and stimulate the economy [4].

1.2 The Obama Presidency

1.2.1 2008 Presidential Campaign

President Obama's campaign platform, like President Bush's, included economic reforms, energy, and health care. Obama proposed lowering taxes for the lower and middle class but raising them for the higher class. He also wanted more government regulation of the financial markets, claiming that too little was the cause of the 2008 financial crisis. One of his promises includes the reformation of the bankruptcy laws passed by President Bush which some believe helped cause the financial crisis. His health care platform included incentives for a strong health care workforce, increased funding for health related research including HIV/AIDS, as well as expanding Medicaid. He also promised to invest in clean energy and wanted to impose more government regulation as well as profits tax on petroleum companies.

While there are fundamental differences in the platforms of both President Bush and President Obama, the agencies they seem to focus on are similar. At the most, I expect the differences in the payroll data of their terms to be most prevalent around education and the environment seeing as how President Bush pushed for more educational reform than Obama and Obama pushed to reverse climate change. [2]

1.2.2 Major Events and Policies in Office

Obama came into office during a financial crisis so naturally, he quickly passes legislation in an attempt to provide relief to those affected by the crisis. The American Recovery and Reinvestment Act was signed into law with the purpose of saving and creating jobs [5].

In 2010, the Affordable Care Act becomes law and Obama approves drilling for oil and gas in the Gulf of Mexico and off the coast of Virginia [5]. However, only a month later, the Deepwater Horizon oil rig explodes. The cause was BP continuing to drill before a pressure seal had been completely tested and as a result, they were forced to pay reparations.

Starting in 2011, Obama froze pay raises on federal employees for two years in an attempt to get the deficit under control [5]. Because of this, we shouldn't see much, if any, increase in pay during these two years.

During his presidency, Obama also implemented the Climate Action Plan in his fight against climate change. As such, we can expect to see a change in the EPA's payroll.

2 Data Understanding

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

I examined the status data of government employees, excluding the Department of Defense, for the years 2001 - 2014. This includes Bush's entire presidency (2001 - 2008) and all but the last two years of Obama's (2009 - 2014). This data was obtained by BuzzFeed News through the Freedom of Information Act.

2.1 Collecting Data

The raw data that I looked at is stored in text files sorted by quarter and formatted the following way.

```
"000172152THOMAS,R SOLE          20010331AB00FR390000040-44 17GS1315-19 0810P064993220F2"
"000299511NEESE,DANIEL L         20010331AB00NL500000045-49 12GS1115-19 1630A044259220F2"
"000547233GREEN,MICHAEL W        20010331AB00FR705000045-49 04GS093-4   1630A034362620F2"
"000733085COOPER, DANNIE         20010331AB00PM300000055-59 04GS1010-14 1630A041505220F2"
"000806215GROSSO,RONALD A        20010331AB00IT700000050-54 13GS14< 1   0301A******215F1"
```

In order to work with this data easily, it needs to be parsed and stored in dataframes. To identify the various features in the records, I used the headers file that was provided. The file was created based on information provided in this pdf included with the data. This file also describes the data attributes stored within these lines of text.

2.2 Data Description

Table 1 below outlines the attributes of the data set that have been extracted from the raw text files according to the previously referenced pdf.

Attribute	**Description**	**Measurement Scale**	**Values**
Pseudo ID	Unique identifier for employees	Nominal	008108528, 000702278, 004390427
Name	The full name of an employee	Nominal	JOHNSON,PATRICIA A, SMITH,JAMES E
Date	The date the pay report is generated. Always the last day of the quarter	Ordinal	20010331, 20020331
Agency	A six character code representing the department where the employee worked.	Nominal	AGHE36, AGAE00, AGCM28, AGARXE
Station	The location where the employee worked	Nominal	110010001, 240130031, 426540101
Age	Age ranges of 5 years where the employee would fall under	Ordinal	50-54, 20-24, 45-49
Education	A two digit number representing the highest level of education this employee has completed	Nominal	13, 17, 04
Pay Plan	A two character variable that represents the type of work the employee does	Nominal	EE, GS, WG.

Grade	Two digit number representing the employee's basic pay rate	Nominal	12, 13, 11, 09, 07
LOS	Time spans representing the length of service of the employee in years	Ordinal	1, 1-2, 5-9, 10-19, 35+
Occupation	Classifies the employee's occupation	Nominal	0303, 0301, 0610, 1811
Category	Represents the type of work the employee does	Nominal	P, A, T, C, O, B
Pay	The employee's annual pay in U.S. dollars	Ratio	49080, 69534, 34151
Supervisory Status	Represents the employee's level of authority	Nominal	2, 4, 5, 6, 7, 8
Appointment	The expected length of the career (permanent, nonpermanent) represented by a two digit number	Nominal	10, 38, 15, 48
Schedule	Describes if the position is full-time, part-time, intermittent, or seasonal	Nominal	F, P, I , J
NSFTP	Represents if the position is seasonal or non-seasonal for a full time, permanent employee	Nominal	1, 2

Table 1: Data Attributes

When working with the data, I added an additional attribute based off of the SCTFILE for agency names to make this data more readable. Other encoded attributes that I will be focused on are education, pay plan, category, and supervisory status.

Code	**Education**
01	No formal education or some elementary school
02	Elementary school
03	Some high school
04	High school graduate or certificate of equivalency
05	Terminal occupational program - Did not complete
06	Terminal occupational program
07	Some collage - less than one year
08	One year college

09	Two years college
10	Associate degree
11	Three years college
12	Four years college
13	Bachelor's degree
14	Post-bachelor's
15	First professional
16	Post-first professional
17	Master's degree
18	Post-master's
19	Sixth-year degree
20	Post-sixth year
21	Doctorate degree
22	Post-doctorate

Table 2: Education Codes

Code	**Category**
A	Administrative
B	Blue collar
C	Clerical
O	Other white collar
P	Professional
T	Technical

Table 3: Category Codes

Code	**Supervisory Status**
2	Supervisor or manager
4	Supervisor
5	Management official
6	Leader
7	Team leader
8	All other positions

Table 4: Supervisory Status Codes

Table 2 shows the translations for the education codes while table 3

does for category and table 4 does for supervisory status. There are too many for pay plan for it to be feasible to display them here so I will provide their translations whenever they are mentioned. I will do the same for appointment and schedule but mostly because I don't expect to mention them often.

2.3 Verification of Data Quality

The data is well formatted and documented so it is fairly easy to work with. However, the dates all fall on one day of each quarter, making it more difficult to track temporal changes in the payroll data. For instance, there are various entries with the same pseudo ID and name which indicates that it is the same person. However, the age ranges are different and we cannot be sure if the person had a birthday inbetween the two record entries due to the fact that the date recorded on the entries are exactly the same. Another instance where this is an issue is in the case of multiple entries for the same person within a quarter but each entry has a different salary amount. In this situation, there is no way to tell if this person got a pay increase or reduction.

An interesting outlier I've noticed is where pay is 0. This doesn't seem to be a mistake, however, since some people could be working for the government without taking a paycheck. I chose to just leave that data as is.

3 Data Preparation

I introduce NA's into my data wherever there is an unknown indicated by asteriks, empty strings, pound signs, "UNSP" in the case of age, and in the case of names, "NAME UNKNOWN", "NAME WITHHELD BY OPM", "NAME WITHHELD BY AGENCY".

Much of the NA's, I decided to leave alone. However, there are a few duplicate ID's within each quarter. I only keep the entry with the higher pay because I'm choosing to assume that the employee had gotten a raise and that was the reason for the second entry. Pay is the most important attribute that I look at so I don't want there to be multiple entries of the same employee for any one quarter.

4 Simple Statistics

I first look at simple statistics, wherever they make sense, for all the attributes. I will however, be ignoring PseudoID since it's just a unique identifier, name since it doesn't really provide much insight, and date since I will be looking at these statistics on a by year basis. I also choose to just ignore NSFTP since it's expected that non-seasonal employees will be much more common than seasonal in government roles. Occupation is also another attribute I don't think we can get much information from so I won't be looking at its simple statistics either. Catefory encodes the same information so I won't lose out on much by ignoring occupation. I take a look at these statistics every four years to show the changes spread out.

Year	Agency	Count
2001	VATA (Vetern's Health Administration)	822, 373
	TR93 (Internal Revenue Service)	424, 763
	SZ00 (Social Security Administration)	258, 782
2005	VATA (Vetern's Health Administration)	870, 277
	TR93 (Internal Revenue Service)	387, 866
	SZ00 (Social Security Administration)	265, 078
2009	VATA (Vetern's Health Administration)	1, 057, 099
	TR93 (Internal Revenue Service)	388, 174
	SZ00 (Social Security Administration)	264, 961
2013	VATA (Vetern's Health Administration)	1, 196, 323
	TR93 (Internal Revenue Service)	355, 900
	HSBC (Transportation Security Administration)	255, 075

Table 5: Agency

An interesting thing to note in table 5 is that VHA grows significantly during Bush's second term in office and so does the TSA after its creation as a response to 9/11. In 2005, the TSA is the 4th largest agency and becomes the 3rd largest by Obama's second term. However, both the IRS and SSA slowly decline under both administrations. Both pushed for tax cuts so a reduction in the size of the IRS makes sense. However, Bush looked to privatize social security while Obama wanted to keep it public so it seems strange that under Bush, SSA grew slightly while under Obama, it declined.

Year	Station	Count

2001	110010001 (District of Columbia) 240130031 (Maryland) 426540101 (Pennsylvania)	399, 912 51, 908 48, 009
2005	110010001 (District of Columbia) 240130031 (Maryland) 426540101 (Pennsylvania)	401, 044 51, 937 51, 851
2009	110010001 (District of Columbia) 240130031 (Maryland) 426540101 (Pennsylvania)	427, 441 55, 801 53, 466
2013	110010001 (District of Columbia) 240130031 (Maryland) 426540101 (Pennsylvania)	450, 227 58, 297 49, 735

Table 6: Station

Table 6 shows that the is a large disparity between the number of federal employees in Washington D.C and elsewhere. This doesn't come as much of a surprise as it is our capitol city.

Year	Age	Count
2001	45-49 50-54 40-44	780, 945 777, 408 666, 719
2005	50-54 45-49 40-44	808, 787 767, 516 673, 168
2009	50-54 45-49 55-59	820, 687 783, 096 704, 393
2013	50-54 45-49 55-59	821, 542 765, 792 723, 270

Table 7: Age

The majority of federal employees are middle aged and older, as shown by table 7. This age ranges do fluctuate in terms of which group is the biggest and this is probably due to employees aging or retiring.

Year	Education	Count

2001	13	1, 040, 658
	04	884, 003
	17	363, 577
2005	13	1, 146, 621
	04	1, 061, 722
	17	432, 985
2009	13	1, 307, 995
	04	1, 076, 471
	17	542, 507
2013	13	1, 428, 091
	04	1, 107, 010
	17	674, 362

Table 8: Education

Refer to table 2 for education codes

Year	Pay Plan	Count
2001	GS (General Schedule)	3, 327, 921
	WG (Non-supervisory pay schedules)	180, 752
	AD (Administrative determined rates)	138, 971
2005	GS (General Schedule)	3, 469, 164
	SV (TSA employees other than executives)	229, 540
	AD (Administrative determined rates)	167, 149
2009	GS (General Schedule)	3, 611, 156
	SV (TSA employees other than executives)	241, 305
	VN (Nursing)	188, 726
2013	GS (General Schedule)	3, 753. 391
	SV (TSA employees other than executives)	254, 564
	VN (Nursing)	222, 732

Table 9: Pay Plan

Table 9 shows that after the Department of Homeland Security was established, the SV pay plan which includes non-executive TSA employees quickly rose to be the second most frequent pay plan in 2005. This clearly reflects the changes in the administration. Another interesting occurance is in 2013 when the VN pay plan that includes employees in nursing rises to be third on the list. This reflects Obama's incentives for healthcare workers.

Year	Grade	Count
2001	12	508, 981
	13	496, 745
	11	438, 985
2005	12	542, 535
	13	520, 030
	11	484, 871
2009	13	593, 962
	12	553, 307
	11	528, 426
2013	12	717, 257
	13	684, 777
	11	474, 558

Table 10: Grade

Year	Length of Service	Count
2001	10 - 14	844, 907
	15 - 19	649, 351
	5 - 9	570, 507
2005	15 - 19	755, 763
	5 - 9	676, 803
	10 - 14	579, 377
2009	5 - 9	966, 576
	20 - 24	647, 703
	1 - 2	637, 611
2013	5 - 9	1, 117, 869
	10 - 14	819, 511
	3 - 4	615, 213

Table 11: Length of Service

Table 11, like table 9 reflects the creation of the Department of Homeland Security. It is shown that the number of employees that have been in their positions for 5 - 9 years almost doubles in 2009. This indicates that there was a large amount of new hires around the time that the Department of Homeland Security was established.

Year	Category	Count

2001	A	1, 511, 170
	P	1, 095, 224
	T	912, 136
2005	A	1, 721, 387
	P	1, 171, 513
	T	1, 094, 857
2009	A	1, 953, 247
	P	1, 330, 041
	T	1, 017, 241
2009	A	2, 037, 661
	P	1, 451, 731
	T	998, 883

Table 12: Category

Refer to table 3 for education codes

Year	Min	Max	Median	Mean
2001	0	230, 628	49, 080	53, 993
2005	0	283, 042	57, 715	64, 132
2009	0	393, 411	65, 810	74, 210
2013	57	401, 589	73, 396	79, 760

Table 13: Pay

Year	Supervisory Status	Count
2001	8	3, 792, 342
	2	471, 649
	5	44, 148
2005	8	4, 094, 209
	2	527, 141
	6	35, 569
2009	8	4, 477, 273
	2	605, 145
	6	42, 647
2013	8	4, 509, 614
	2	653, 254
	6	61, 880

Table 14: Supervisory Status

Refer to 4 for education codes

Year	Appointment	Count
2001	10 (Career)	2, 689, 539
	38 (Other)	659, 613
	15 (Career-conditional)	421, 937
2005	10 (Career)	2, 711, 026
	38 (Other)	941, 949
	15 (Career-conditional)	387, 607
2009	10 (Career)	2, 714, 173
	38 (Other)	1, 113, 801
	15 (Career-conditional)	507, 921
2013	10 (Career)	2, 908, 466
	38 (Other)	1, 354, 279
	15 (Career-conditional)	408, 747

Table 15: Appointment

Year	Schedule	Count
2001	F (Full-time)	3, 910, 415
	P (Part-time)	173. 300
	I (Intermittent)	139, 299
2005	F (Full-time)	4, 257, 345
	P (Part-time)	187, 536
	I (Intermittent)	171, 451
2009	F (Full-time)	4, 688, 367
	P (Part-time)	209, 439
	I (Intermittent)	180, 573
2013	F (Full-time)	4, 846, 185
	P (Part-time)	201, 763
	I (Intermittent)	185, 924

Table 16: Schedule

5 Modeling

5.1 Single Attribute Visualization

5.1.1 Agency

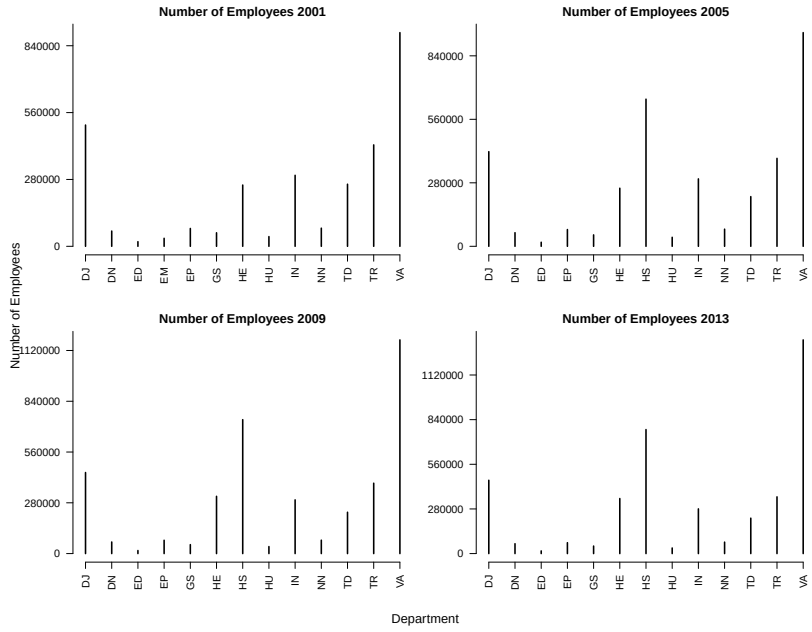

Figure 1: Number of Employees by Agency

Figure 1 shows the distribution of the number of employees in important departments in the government. I chose a barplot because it clearly shows how the size of each department compares to the others. From 2001 to 2005, we can see that the Department of Homeland Security has been created and quickly became one of the largest in my representation.

5.1.2 Pay

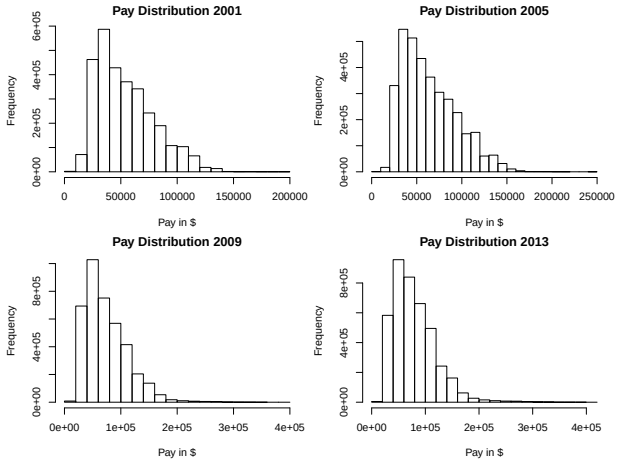

Figure 2: Pay Distribution by Year

Figure 2 shows what the pay distribution is by year. I chose to use a histogram to clearly depict what salary ranges are the most common.

5.1.3 Education

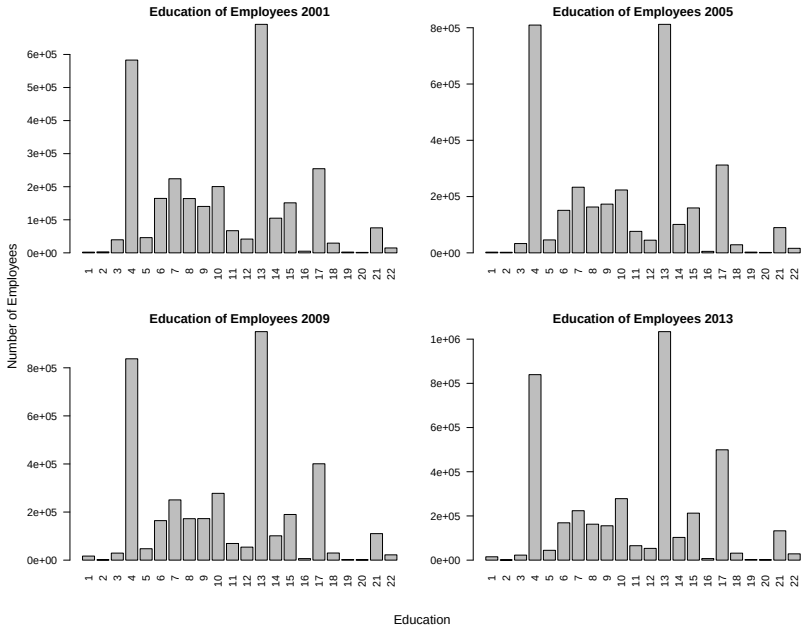

Figure 3: Distribution of Education Levels

Figure 3 shows what the common educational levels are in the government. I used a histogram to show the counts of each level and how they compare to one another. Not surprisingly, high school and bachelor's are the more common for every year.

5.1.4 Age

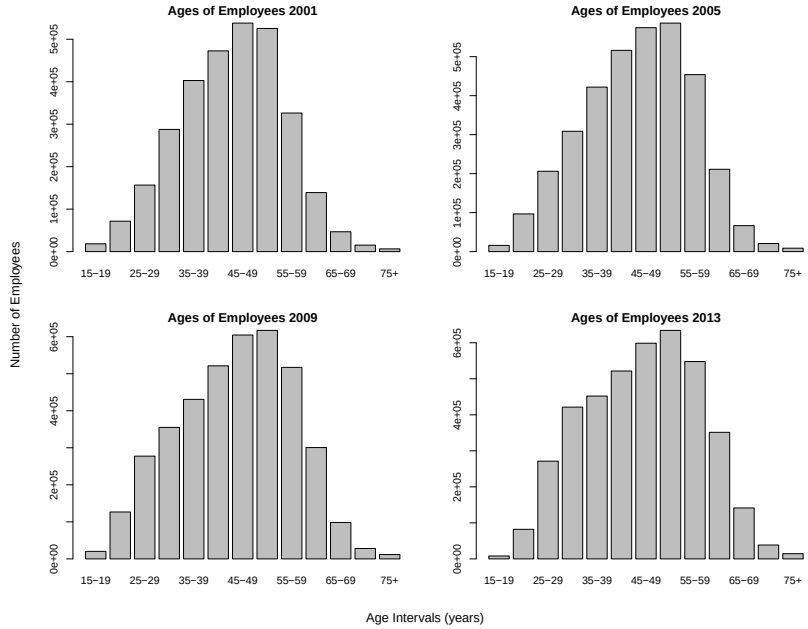

Figure 4: Distribution of Age

Figure 4 shows the counts for each of the ages. Again, I chose to use a histogram because it is best for visualizing these types of distributions.

5.1.5 Length of Service

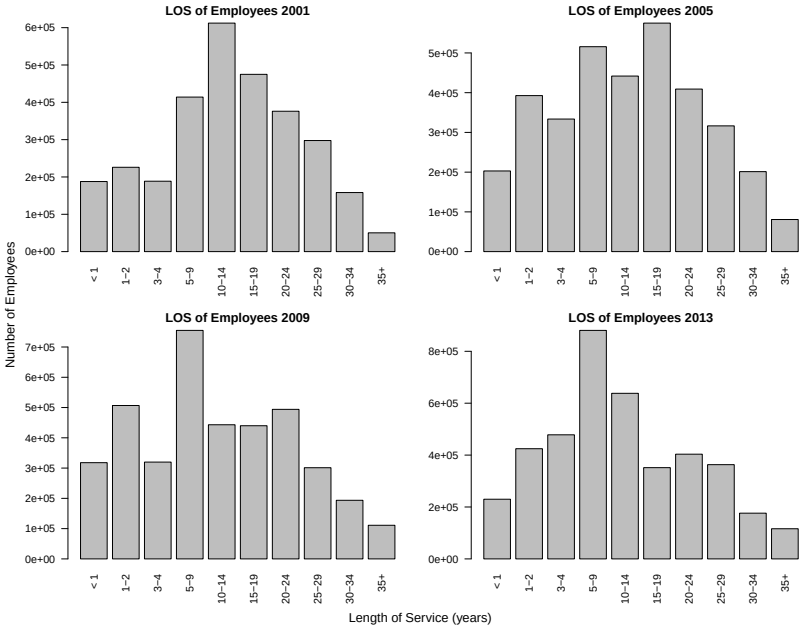

Figure 5: Distribution of Length of Service

Figure 5 shows the frequency of each length of service range. I chose to use histograms because you can clearly see the shift towards newer employees.

5.2 Relationship Visualization

5.2.1 Effect of Education on Pay

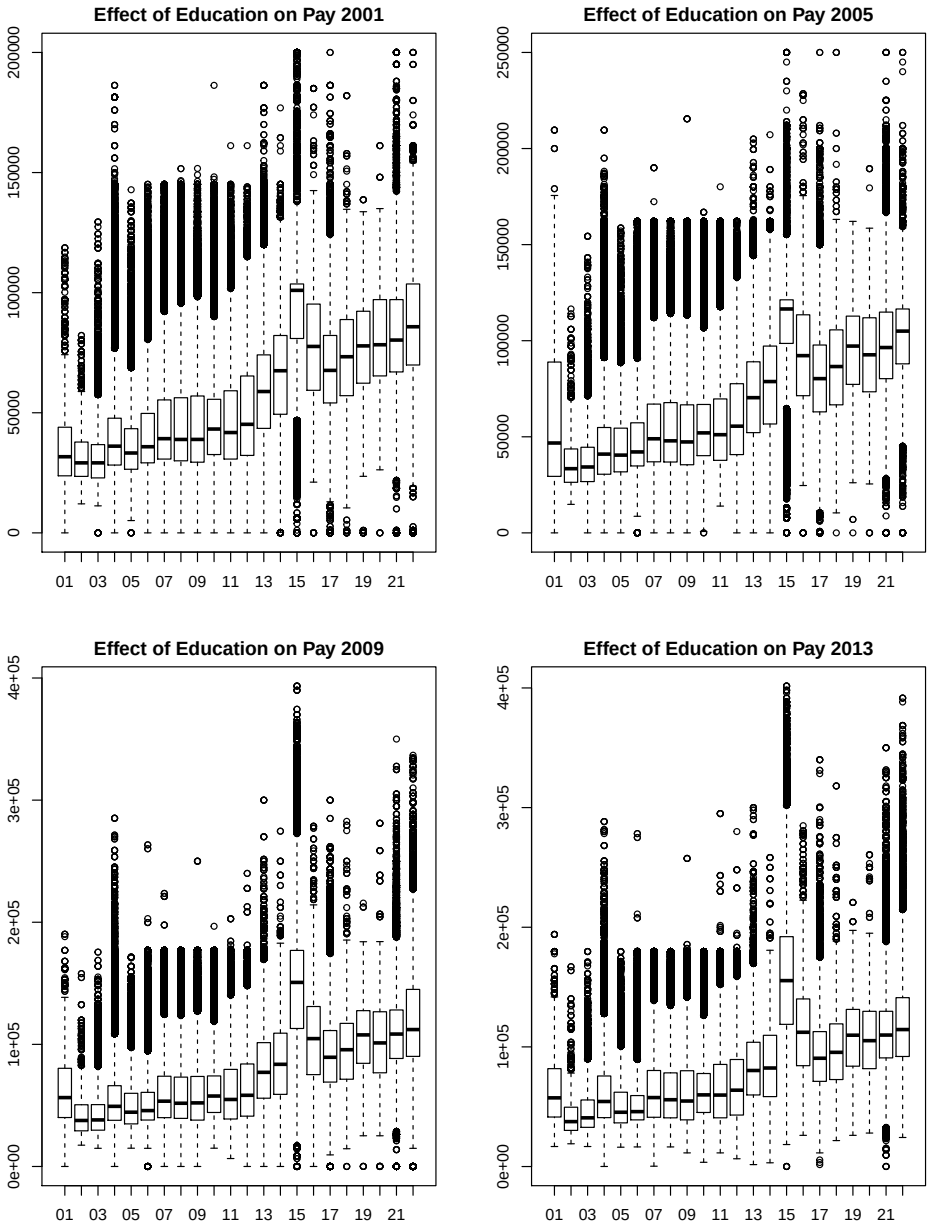

Figure 6: Boxplot of Education and Pay

Refering to table 2 for the code translations, we can see that 15 - First professional has the highest average salary but also the highest degree of variance. Figure 6 shows that there is a general trend where the higher your education level is, the more you make on average. I chose to use a boxplot to visualize this relationship in order to emphasize that education level doesn't necessarily correlate to a higher income. While it is especially so in the case of 15, there are a lot of outliers with all of the education levels and boxplots best represent that.

5.2.2 Average Pay Across States

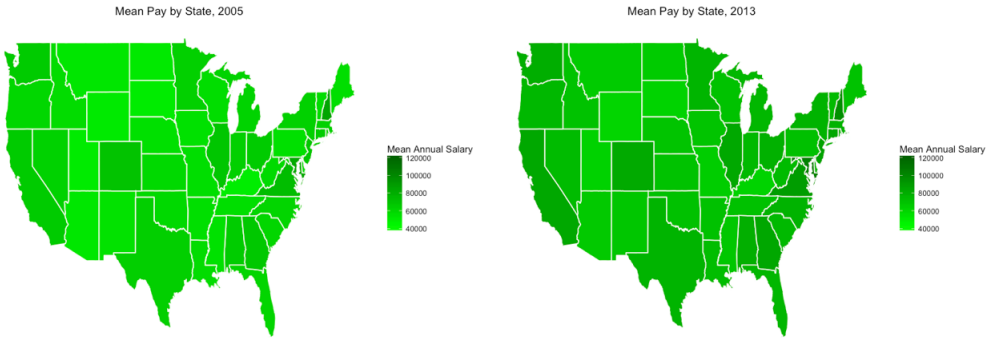

Figure 7: U.S. Averge Pay Map: Courtesy of Jake Carlson

Figure 7 shows that there is a clear correlation between how much a government employee makes and where they are stationed. Employees in larger states such as California and New York make much more than those in South Dekota.

5.2.3 Effect of Length of Service on Pay

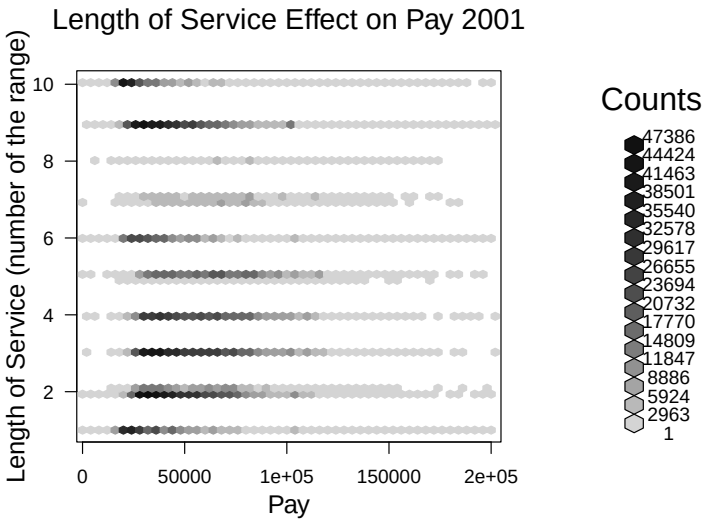

Figure 8: LOS Effect on Pay 2001

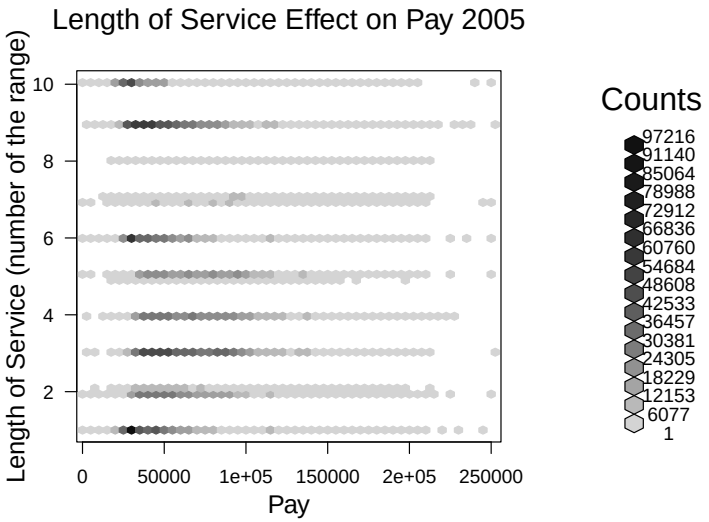

Figure 9: LOS Effect on Pay 2005

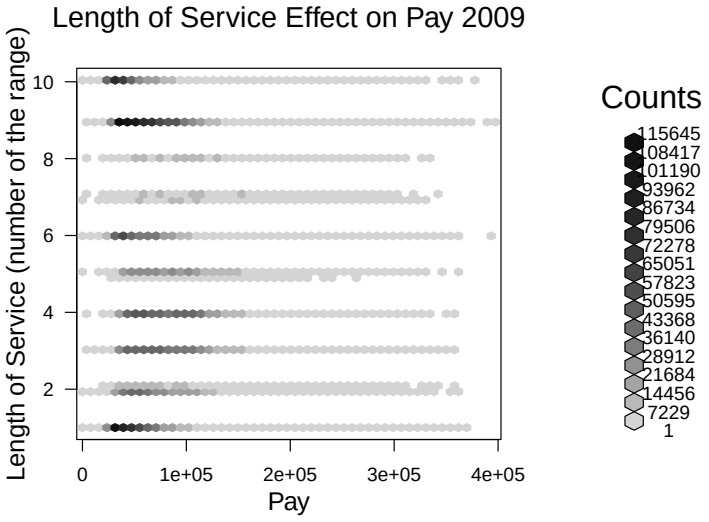

Figure 10: LOS Effect on Pay 2009

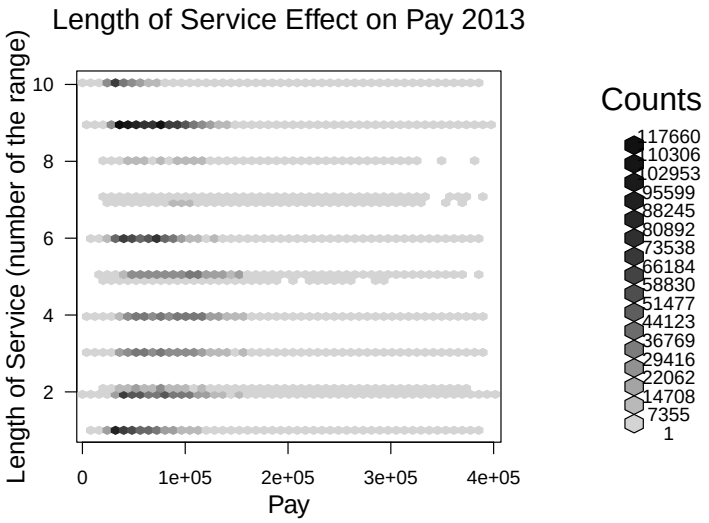

Figure 11: LOS Effect on Pay 2013

Figures 8, 9, 10, and 11 all show a fairly even distribution of pay across the length of service ranges. This can be attributed to the fact that this is looking at data across all the agencies so there are bound to be many where length of service doesn't really help get you an increase in pay. I chose to graph this relationship using hexgonal binding because it clearly shows the relationship between length of service and pay as well as the frequency for each case.

6 Evaluation

6.1 Government Policies

The trends that I had predicated based on policies and platforms were most prevalent in the data when it came to healthcare and the Department of Homeland Security.

General employment trends were also reflected in this data such as higher educational levels having higher average pays.

6.2 Further Investigation

I could further analyze changes over time with this data set in order to pinpoint events in history and see how the government responds to them. I kind of did this in my investigation but I could have shown how each agency grew or shrunk over time rather than just comparing separate years.

References

[1] *2000 Republican Party Platform*
http://www.presidency.ucsb.edu/ws/index.php?pid=25849

[2] *2008 Democratic Party Platform*
http://www.presidency.ucsb.edu/papers_pdf/78283.pdf

[3] *Policies of the Bush Administration 2001 - 2009*
https://georgewbush-whitehouse.archives.gov/infocus/bushrecord/documents/Policies_of_the_Bush_Administration.pdf

[4] *George W. Bush - Key Events*
https://millercenter.org/president/george-w-bush/key-events

[5] *Barack Obama - Key Events*
https://millercenter.org/president/barack-obama/key-events