

Data Mining - EMIS 7331 Project 1

Mohamed Elsaied

09/24/2017

Abstract

This Project discusses how federal, non-military employment was impacted by the transition from President George Bush to President Barack Obama to learn about the impact of the difference in policy between the two administrations on the federal government. In the First section the Project will look at the campaign promises from each and then look at the actual policies applied by each and the Key events to see if that affected the Federal employment strategy. After that the project explains the data set provided, what it means, how was it cleaned and what are the assumptions and Imputations. In the Third section some agencies are selected for analysis and visualization using appropriate methods. The Fourth and the last section has the conclusion of the project, and discussed the limitation to work, and how can this report taken for further research

Contents

1	Introduction:	3
1.1	Buisness Understanding	3
1.2	Candidates History	4
1.2.1	George W. Bush	4
1.2.2	Barack Obama	5
2	Data set	6
2.1	Data History, Definition, & Explanation:	6
2.2	Data Assumptions & Imputations:	6
2.2.1	Data Assumptions:	6
2.2.2	Data Imputations:	8
3	Data Analysis & Visualization	8
3.1	Agencies selected for Analysis & Visualization:	9
3.1.1	Department of Health and Human services	9
3.1.2	NASA	11
4	Limitation to work and future research	17
5	Conclusion	17

List of Figures

1	Mean Pay Salary 2000 - 2014	9
2	HHS Employees over the years	10
3	Level of Education	11
4	HHS Pay	12
5	Pay VS Education	12
6	Pay VS LOS	13
7	HHS Employment	13
8	Nasa Education	14
9	NASA Pay Over The Years	15
10	Pay VS Education	16
11	Pay VS LOS	16

List of Tables

1	Attribtes definition	7
2	Mean, Median, & Max Pay	8
3	Corelation Between Pay, Education, & LOS	11
4	NASA Employees over the years	14

1 Introduction:

On May 24, 2017. BuzzFeed news shared an enormous dataset — one that sheds light on four decades of the United States’ federal payroll. The dataset contains hundreds of millions of rows and stretches all the way back to 1973. It provides salary, title, and demographic details about millions of U.S. government employees, as well as their migrations into, out of, and through the federal bureaucracy. In many cases, the data also contains employees’ names.

They obtained the information — more than 30 gigabytes of it — from the U.S. Office of Personnel Management, via the Freedom of Information Act (FOIA).

1.1 Buisness Understanding

Understanding the Federal Budget process will help us understanding how the federal, non-military employment is affected by different administrations.

Step one of the process is The President’s budget request: The process starts when the president submits a detailed budget request for the coming fiscal year. First, it recommends the congress for the overall federal fiscal policy. The request has three components (1) how much money should be spent on public purposes. (2) How much should it take as tax revenue (3) how much of a deficit the federal government should run. Second, the president budget layout his relative priorities for federal programs (how much should be spent on each agency) Third, the president’s budget signals to congress his recommendations for spending and tax policy changes. The budget comprises different types of programs: (1) annually appropriated programs (known as “Discretionary” and it makes about one-third of all federal spending. (2) Taxes, “mandatory” or “entitlement” programs, and interest. (3) Recommendations for mandatory programs (discusses the changes to eligibility criteria and levels of individual benefits) (4) changes to tax code.[6]

Step two the congressional budget resolution: Where the congress holds hearings to question administration officials about their requests and then develop its own budget plan. [7]

Step Three Enacting Budget Legislation: Following the adaption of the budget resolution, Congress considers the annual appropriations bills that are needed to fund discretionary programs in the coming fiscal year and legislation to enact changes to mandatory spending or revenue levels as specified in the budget resolution.[8]

So from understanding the Process we know that the President is not the final only variable, the congress approval is also in the process, but definetly the President policy affects the overall outcome of the Federal budget, which affects the Fedral Employment.[9]

1.2 Candidates History

1.2.1 George W. Bush

George W. Bush Was Born July 6, 1946 in New Haven, Connecticut. He was raised in Midland and Houston, Texas. His Grand Father was a U.S. Senator from Connecticut. His father, was Ronald Regan's vice president from 1981 to 1989 and the 41th U.S. President from 1989 to 1993. He attended Yale University from 1964 to 1968, studying Arts with degree in history. He got his MBA degree from Harvard Business School. In 1994 Bush declared his candidacy for Texas gubernatorial election. And he became the Governor of Texas in 1995. His major achievement was using the budget surplus to push through Texas's largest tax-cut, \$2 billion. He extended government funding for organizations providing education of the dangers of alcohol and drug use and abuse, and helping to reduce domestic violence. Critics contended that during his tenure, Texas ranked near the bottom in environmental evaluations. Supporters pointed to his efforts to raise the salaries of teachers and improve educational test scores. In 1998, Bush won re-election with 69% to be the first Governor in Texas history to be elected to two consecutive four-year terms.[5]

Presidential Campaign

The 2000 presidential campaign was officially launched on June 14, 1999. Following the Clinton Presidency, the Bush campaign made extensive use of the compassionate conservatism concept. Bush criticized the Clinton-Gore administration for being too interventionist, he promised a humble foreign policy with no nation building. Bush promised tax breaks for all, sometimes using the slogan "whoever pays taxes gets a tax break". He proposed an act that increases funding for schools, while requiring greater accountability for results. he named the act "No Child Left Behind". For Energy: The Bush campaign supported a comprehensive energy reform bill which includes initiatives for energy conserving technologies as well as decreasing the foreign dependence on oil through increased domestic production and the use of non-fossil fuel based energy production methods.[14]

Presidency

After his won, in January 2001, Bush bans providing aid to organizations that perform abortions. In June 2001 Bush signs \$1.35 Trillion tax cut law. He also outlines a new policy on STEM cell research. On September 11 United States was attacked by a terrorist attack. January 2002 Bush a law on education reform bill "No Child Left Behind". May 2003 Bush sign 3rd largest tax cut in history. August 2005 Hurricane C=Katrina hits the gulf coast. October 2006, Bush approves 700 mile fence along US-Mexico border.[2]

1.2.2 Barack Obama

Barack Obama Was born on August 4, 1961 in Honolulu, Hawaii. He attended high school in Hawaii, after graduating he moved to Los Angeles to attend Occidental College. Later in 1981 he transferred to Columbia University where he majored in political science with a specialty in international relations and in English literature. He entered Harvard Law School in 1988. In 1991 he was a visiting law and Government fellow at the University of Chicago Law School. He was Illinois State Senator from 1997 to 2004. And he became a U.S. senator from Illinois from 2005 to 2008.[5]

Presidential Campaign

The 2008 presidential campaign had a lot of promises starting with the Energy, it Promised to end the dependance on oil from the middle east. For Education it promised to invest in early childhood education. Recruit an army of new teachers, and pay them higher salaries, and give them more support. And in exchange, we ask for higher standards and more accountability. It promised to stop giving tax break to companie that ship jobs to overseas. For healthcare it promised to lower the premiums if you have a health care. if you don't you will be able to get some kind of coverage.[15]

Presidency

After he won, In February 2009, he pulled the restriction on how much money could go to STEM research. October 2009, Obama lifts a ban preventing people with HIV & AIDS from entering the US. Janurary 2010, Obama announces \$900 million in grants for schools underperforming. April 2010, Deepwater horizon oil explodes killing 11 people. April 2010, Obama proposes NASA budget increase by \$6 billion over the next 5 years. November 2010, Obama institutes a two year pay raise freeze for federal employees.[1]

2 Data set

The Data provided by BuzzFeed [9] is in text file with column width. Each year divided into 4 quarters. So the files were broken and headers applied using the code given by Dr. Hahsler[10]. Then a new Column was added to the data by reading the translation file which matched the agency code with the agency number and was added to my Data. Then a code[11] was used to clean and group the data.

2.1 Data History, Definition, & Explanation:

The Data was obtained via the Freedom of Information Act (FOIA) Via Jeremy Singer-Vine [8] and it discusses the Payroll for the Federal government since 1974 till 2016. In our report we will be interested in the Bush and Obama administrations. The dataset contains hundreds of millions of rows and stretches all the way back to 1973. It provides salary, title, and demographic details about millions of U.S. government employees, as well as their migrations into, out of, and through the federal bureaucracy. In many cases, the data also contains employees' names. One tracks the "status" of every federal employee at the end of every fiscal quarter. The other tracks entries and exits, known in bureaucratic lingo as "accessions" and "separations." (That data only goes back to 1982.) For each employee, the "status" data includes the following information: Name, Agency, Subagency, Location (only to the state/country level in recent data, but a more detailed "duty station" before that), Occupation, Occupational category, Age (as a range), Education level, and years since degree (as a range), Adjusted basic pay, Pay plan and pay grade, Type of appointment (e.g., "career," "nonpermanent," "Schedule A," etc.), Work schedule (e.g., full time, part-time seasonal, etc.), Supervisory status. Detailed definitions can be found in Table 1

2.2 Data Assumptions & Imputations:

This section talks about the data set, and what assumptions were made to missing value, and the Imputations to fix it

2.2.1 Data Assumptions:

The Following shows the data assumptions that was done for the data set. Though the data has more than 15 attributes, the following is the important attributes that I want to focus on and clean. Age, Education, Length Of Service, Pay, and Agency Name)

1-For the Length of service, the data was given in range of years. The average of the range was taken as a value for the specific Length of service.

2-For the Age, the data was given as age groups. The average of the age group was taken as a value for the specific Age.

Table 1: Attributes definition

Attribute	Description
PseudoID	A unique identifier number for each employee(Nominal Variable)
Name	The name of Employee if provided
Date	The date of the Quarter(all Observations within the same quarter have the same date)
Agency	The departments encoded as 6 character variable
Station	Actual Location in which the Employee worked in a 9 character variable
Age	Age groups of employees(Ordinal Variable)
Education	Level of education achieved in a 2 digit number(Ordinal Variable)
PayPlan	2 Character variable that represents type of work the employee does
Grade	Used with PayPlan to determine employee basic pay
LOS	Represents the length of service of an employee in years(ordinal Variable)
Category	Represents the type of work an employee does
Pay	Annual pay of an employee(Ratio Variable)
SupervisoryStatus	Represents the level of leadership of an employee
Appointment	Represents if the position is permanent, not permanent
Schedule	Represents if the employee position is full time, part time, or seasonal

Year	Mean	Median	Maximum
2001	\$51,266	\$48,469	\$200,000
2005	\$63,014	\$57,002	\$250,000
2009	\$73,605	\$64,021	\$393,000
2013	\$79,333	\$72,722	\$401,589

Table 2: Mean, Median, & Max Pay

3-For the Pay, all the values with less than \$3500(less than 0.0004%) were asumed to be ouliers and were taken out of the data set.

4-Duplicate ID's with the same name, were assumed to be an upgrade or change in the Agency.

2.2.2 Data Imputations:

The following shows the data imputations that was done for the data set. Many of the Variables had missing values (e.g. * or **** or #####) so these values had to be changed to NA's so we can see if we want to delete them or do some imputations.

1-For all the attributes, all the missing values were replaced with NA.

2-For the Pay, all NA's were replaced with median pay.

3-For the Age, all NA's were replaced with the median age for the agency.

3 Data Analysis & Visualization

After looking at the data and understanding the buisness, we wanted to look and see if the employment of the federal government was affected with natural disasters, for example 09/11 Attacks. so the mean pay of the employee for each year and was plotted against the years 2000 to 2014. so if we look at1 the maximum mean pay(\$79,333) and the minimum mean pay(\$51266) over the years. we will find that it's exactly equal to 2% which is the inflation. so the big picture looks consistent. but if you look at the transition between 2002 & 2003 you will find that the slope of the line has changed due to the 09/11 attacks.Mean, Median, & Maximum Pay per year can be found in 2

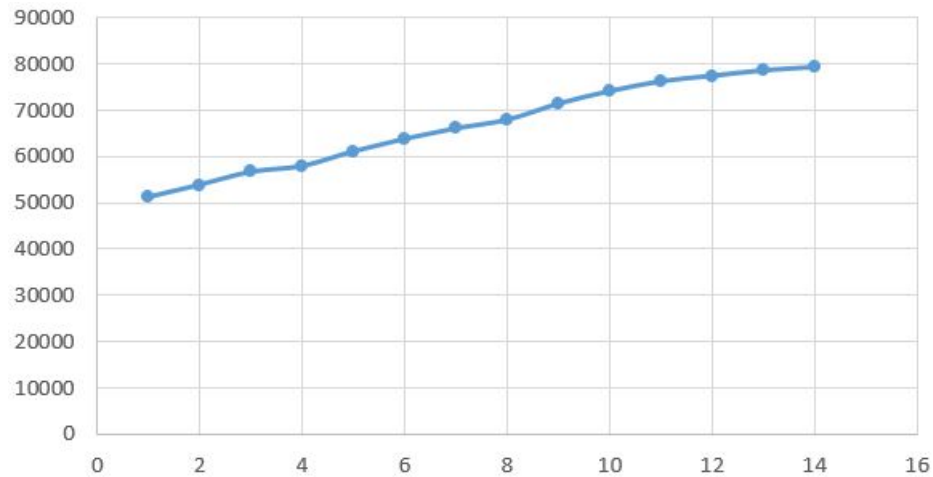


Figure 1: Mean Pay Salary 2000 - 2014

3.1 Agencies selected for Analysis & Visualization:

The time frame that we will be comparing in between different agencies is 2001, 2005, 2009, and 2013. After Comparing the differences in policies between the 2 administrations, it was observed that the main agencies that we would really see difference in Employments is (1) The department of Health and Human services(HHS).

3.1.1 Department of Health and Human services

also known as the Health Department, is a cabinet-level department of the U.S. federal government with the goal of protecting the health of all Americans and providing essential human services. Its motto is "Improving the health, safety, and well-being of America". Before the separate federal Department of Education was created in 1979, it was called the Department of Health, Education, and Welfare (HEW).

Number of Employees over the years

The change in Number of employees over the years gives us an idea of how the 2 different administrations prioritize this specific agency. Looking at 2 We can see How the Employment Dramatically changed between the 2 administrations. it's really noticable that Obama was really focusing on the HHS Department.

HHS Education

As we saw in 2 the number of employeeed in HHS increased dramatically in the Obama administration. So it would be really helpful to look at the level of

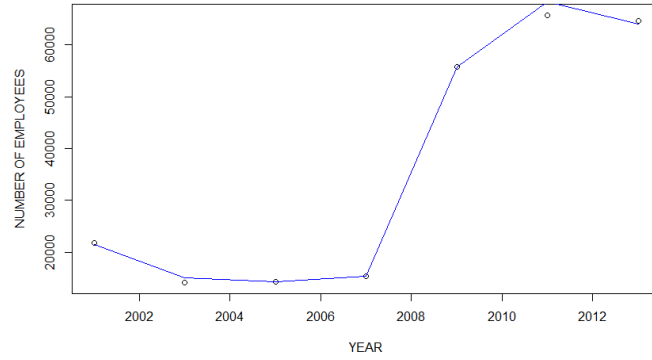


Figure 2: HHS Employees over the years

education of the new employees and compare it with the Bush administration. 3 shows us that most of the people hired were Level 1(Middle School Education) or Level 4(High School Education).

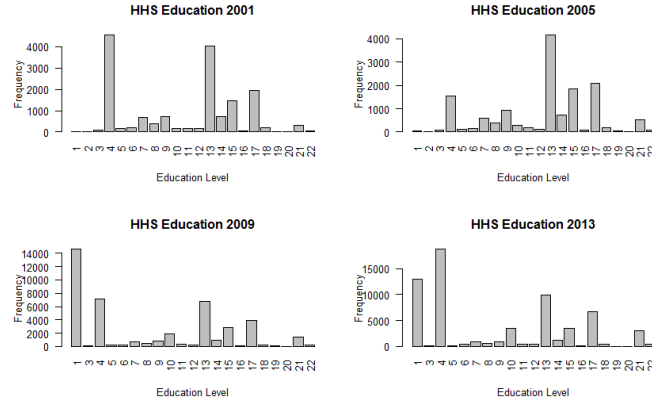


Figure 3: Level of Education

Year	Pay VS LOS	Pay vs Education
2001	0.46	0.47
2005	0.35	0.35
2009	0.49	0.45
2013	0.43	0.39

Table 3: Correlation Between Pay, Education, & LOS

HHS Median Pay

After looking at the Number of Employees and their Education we see that the Median jumped from \$55,153 in 2001 to \$68,809 in 2014 which the difference is almost the inflation (2%) so in 2014 people were paid the same. but the Median of Education dropped from 13 in 2001 to 6 in 2014.

4 Shows HHS Pay Distribution and 5 Shows the relation between Pay and Education. 6 Shows the Pay vs LOS. Also the corelation(how much the dependent variable, the LOS & Education, explains the independent variable, the Pay) between the Pay, Education, & LOS can be found in Table 2. we see that the relation between them is good in all years except 2005 for the HHS agency.

HHS Employee concentration comparison between 2001 and 2013 can be found in 7

3.1.2 NASA

The National Aeronautics and Space Administration is an independent agency of the executive branch of the United States federal government responsible for the civilian space program, as well as aeronautics and aerospace research. President Dwight D. Eisenhower established NASA in 1958 with a distinctly civilian (rather than military) orientation encouraging peaceful applications in space

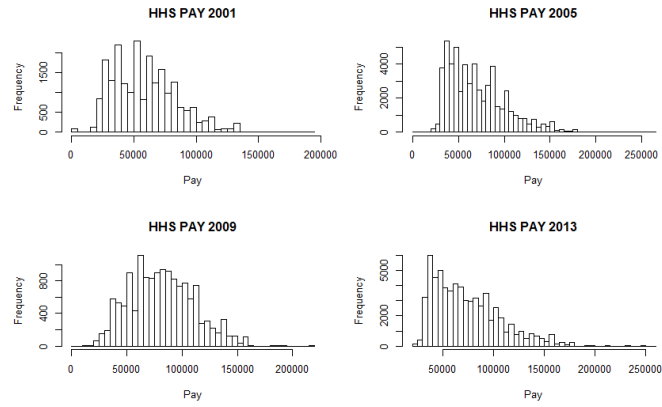


Figure 4: HHS Pay

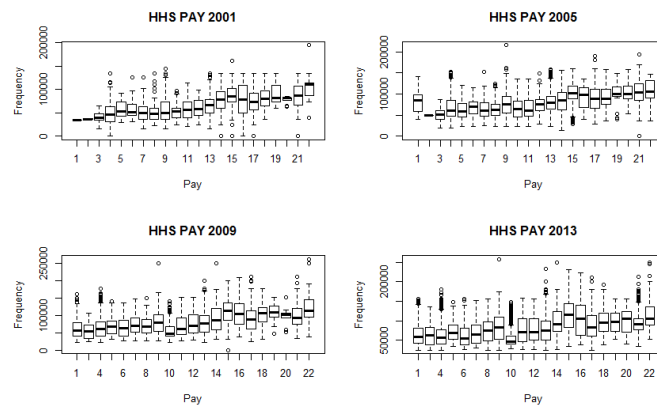


Figure 5: Pay VS Education

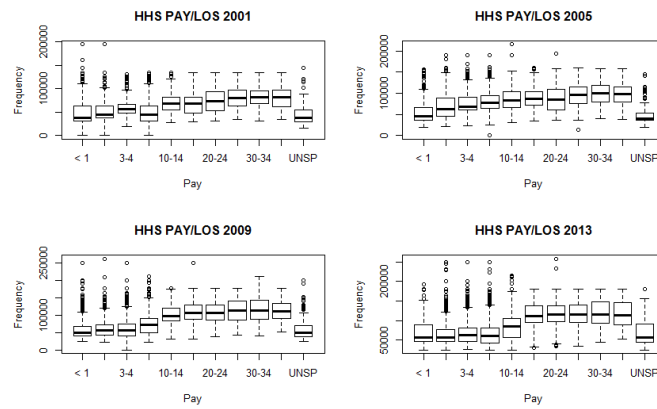


Figure 6: Pay VS LOS

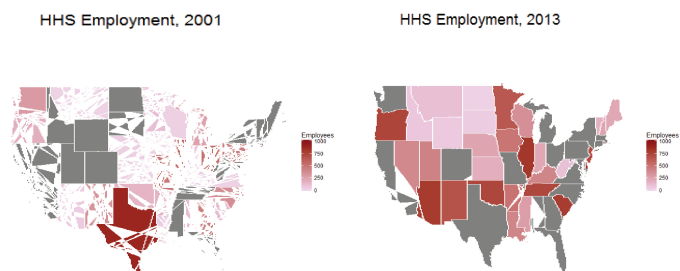


Figure 7: HHS Employment

Year	Number Of Employees
2001	5154
2005	5583
2009	6369
2013	5955

Table 4: NASA Employees over the years

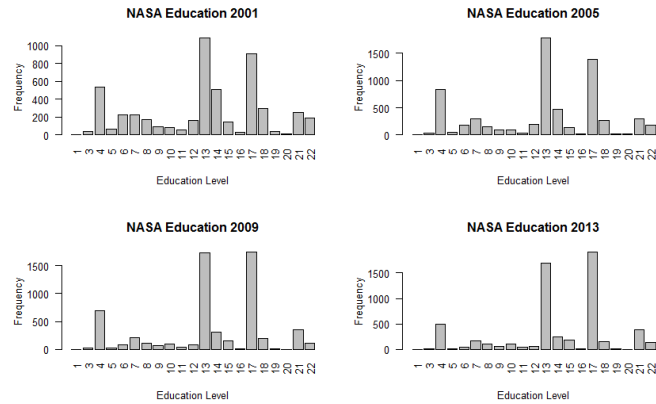


Figure 8: Nasa Education

science. The National Aeronautics and Space Act was passed on July 29, 1958, disestablishing NASA's predecessor, the National Advisory Committee for Aeronautics (NACA). The new agency became operational on October 1, 1958. Since that time, most US space exploration efforts have been led by NASA, including the Apollo Moon landing missions, the Skylab space station, and later the Space Shuttle. Currently, NASA is supporting the International Space Station and is overseeing the development of the Orion Multi-Purpose Crew Vehicle, the Space Launch System and Commercial Crew vehicles. The agency is also responsible for the Launch Services Program (LSP) which provides oversight of launch operations and countdown management for unmanned NASA launches.[13]

NASA Employees over the years:

In 4 we can see the Number of Emolyees changing over the years, the Obama adminstration has employed more people.

NASA Education

In 8we can observe as expected that most of the people has Education 13 which is a Bachelor Degree and Education 17 which is a Master's Degree

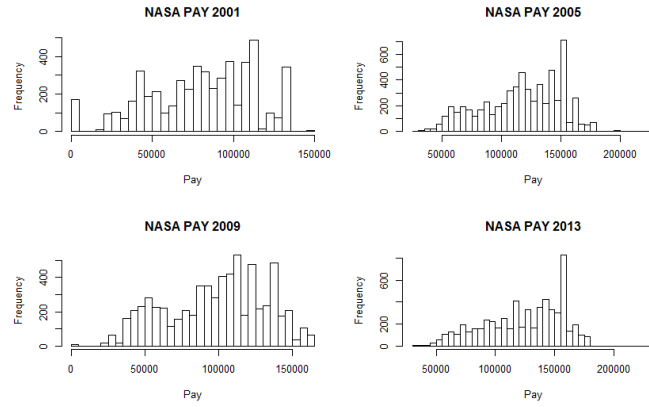


Figure 9: NASA Pay Over The Years

NASA Median Pay

After Looking at the number of Emolyees and the Pay we find that the Median Pay jumped from \$82,168 in 2001 to \$129,758 in 2013 with an increase of (4.5%) over the years. 9 Represents the Pay Distribution. In figure 10 shows the relation between the Pay and the Education. 11 Shows the relation between Pay and LOS.

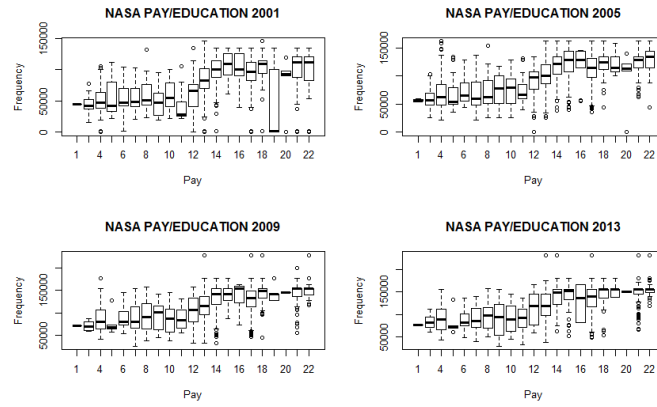


Figure 10: Pay VS Education

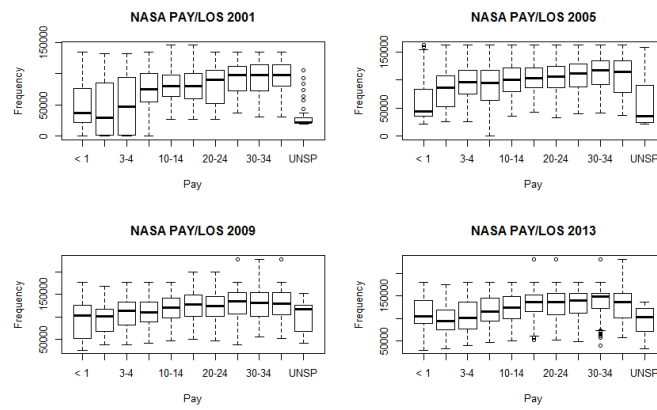


Figure 11: Pay VS LOS

4 Limitation to work and future research

In this report the search was only limited to 2 agencies HHS & NASA due to lack of time, so further look into Agencies like Health and Urban Development, IRS, Homeland Security, Transportation, Education, and Energy is recommended. Also the data definition the the Fedscope document mentioned that there is information about the ethnicity of the Employees but unfortunately we couldn't find it. so it's recommended to look at the common Last names to see the percentage of Hispanic, Black, and minorities to see the percentage difference between two administrations.

5 Conclusion

This report started with introduction discussing the business and the 2 key factors in the business(Bush & Obama) going from their history, to the presidential campaign policies to the actual presidency. After that the data provided was discussed, what does it mean, how to clean it, and how to use it for further analysis. all the data assumptions and permutation was discussed as well. After that Agencies of HHS & NASA were analyzed using different graphical and statistical methods. The most important findings was that there was change between the 2 administrations and it can be seen clearly in the HHS & NASA Department. at the end the limitation and further research to work was discussed

References

- [1] Barack Obama: Key Events<https://millercenter.org/president/barack-obama/key-events>
- [2] George W. Bush: Key Events<https://millercenter.org/president/george-h-w-bush/key-events>
- [3] Barack Obama: Early Life<https://en.wikipedia.org/wiki/BarackObama>
- [4] George W. Bush: Early Life<https://en.wikipedia.org/wiki/GeorgeW.Bush>
- [5] United States Budget Process(Obama, President Barack. "State of the Union address JanUary 28, 2014." (2014).)
- [6] BudgetResolution-USSenate(Saturno, James. "The congressional budget process: a brief overview." Congressional Research Service, Library of Congress, 2000.)
- [7] Enactment of a Law (Crain, W. Mark, and James C. Miller III. "Budget process and spending growth." Wm. Mary L. Rev. 31 (1989): 1021.)
- [8] BuzzFeed-FederalPayRecord(https://www.buzzfeed.com/jsvine/sharing-hundreds-of-millions-of-federal-payroll-records?utm_term=.fbQxPYGG6N.rk0PblJJDX) =
- [9] Federal Pay Record (<https://archive.org/details/opm-federal-employment-data>)
- [10] Dr. Hahsler Codehttp://michael.hahsler.net/SMU/EMIS7331/data/federal_employment/project1.html
- [11] Clean File <https://github.com/jakecarlson1/data-mining-projects/blob/master/project-1/project1>
- [12] Department of Health and Human services: History<https://en.wikipedia.org/wiki/UnitedStatesDepartmentofHealthandHumanServices>
- [13] Nasa: History<https://en.wikipedia.org/wiki/NASA>
- [14] Bush - Presidential Campaign: https://en.wikipedia.org/wiki/GeorgeW.Bush_presidential_campaign,2000
- [15] ObamaPresidentialCampaign<http://www.huffingtonpost.com/entry/a-look-at-how-obama-is-keeping->