Data Mining Project 1

The Impact of POTUS on the U.S. Federal Employment

A Data Mining Analysis

Student Name

➢ Executive Summary:

During election races, presidential candidates promote their policies by offering several promises to the voters. The employment by federal government is an indicator for the adopted policies of each U.S. president. It is intended to reveal the impact of each president on the federal employment based on the released data from 1972 [1] Particularly, the focus of this analysis is on the comparison of the age, income, education, working schedule, and length of service for full-time federal employees in the quarters right before re-election of the seating president during Bush and Obama administrations. Several features are selected and then individually analyzed in a comparison for both administrations. To get a deeper understating of data, the relationship of various feature of data set in also compared for both administrations. The analysis reveals the trend of the two administrations according the level of education, age, length of service, payment, and schedule of each employee. It is reveal that U.S government became large in Obama administration by hiring high-wage and high-educated employees to address administration agenda over modern technologies to tackle climate change as well health care. On the other hand, as the agenda for Bush administration was on defense and security, it required lower education level and lower payments. To expand the horizon of the analysis, relative income of employees of four agencies including Defense and Safety, Education, Environmental, and Health are considered during U.S. presidential terms since 1972.

Table of Conents

➢ Business Understanding

According to CRISP-DM business understanding is the first step of data mining. The analyst's goal of this project is to uncover crucial factors that could influence the POTUS on the U.S. Federal Employment during Bush and Obama administration. This focus of the government in Bush administration was more on agency of defense and security, while in Obama administration the focus was on health and technologies to address to address climate change. Many publications claim that the size of government was larger in Bush administration. However, since this result may be the effect of war in Bush administration. As the data regarding DoD is not provided in the presented set, it is very important to state that taking the DoD data would change the results. Thus, any indication of the size of the government must be further corrected by considering the data for DoD.

The major policy difference between the two agency is that the number non-military employee was increased in Obama term. Many of them are employed to serve in the health sector. By acknowledging the changes in climate, large amount of investment in renewable energy caused a bombastic growth in green energy sector. On the other hand, 9/11 play an important role in the agenda of Bush administration who initiated two long wars in middle east. The employment data could say a lot on the focus of the administrations to address their agenda.

The main actors of the house of representative, U.S senate, and of course "we the people." All of them has the impact on the direction of the U.S government. The analysis of employment data can help understanding that what would happen if house, senate, and oval office is governed by the same party. It is also interesting to see the confliction between the governing party and opponent in other sections of the U.S government. Government shutdown in 2013 is an example of such confrontation. For sure, the employment data could reveal that if a president was successful to advance his promises in the presidential campaign.

Not all the claim by the media would be correct. As data might indicated another direction. For example, it is frequently mentioned that the U.S. government (in addition to DoD) is increased under Obama while the following indicates that it is not necessarily true [2].

A publish topic is based on the examination of the total number of government employees (GE) as a percentage of the population (P) (GE/P). A higher ratio indicates a higher percentage of government employees relative to the total population. The ratio was 7.4% at the end of Bush term while it was 6.9% at the end of Obama term. It essentially takes a snapshot at the end of each president's term and compares it to the point when they took office. To glean the full picture, we'll need to look at the percentage increase for the population and for the number of government employees. If the population increases by a rate like the increase in government employment, that would be expected. From 1980 to 2008, the percentage increase in total government employees at the end of each of their terms (compared to when they took office), was between 6.4% and 10.2%. This is relatively close to the percentage increase in the population over the same period (5.0% to 9.8%). However, during Obama's first term, the population increased by 3.0% while the number of government employees fell by 7%. This is the reason for the decline in the GE/P Ratio shown above. This decline is largely a result of the financial crisis. With revenue in shorter supply, state governments which have a mandate to balance their budgets, cut staff.

➢ Data Set Description

The initial data file provides the raw data of U.S. Federal employment during 41 years from Sep. 1973 to Jun.,2014, where each year is divided to four quarters. Several types of data is reported which required clarification. Thus, a description of the provided data is given in Table 1. Additionally, the data type of each feature is provided, where nominal scales are used for labeling variables, without any quantitative value, ordinal scalers are used where the order of the values are important and significant, but the differences between each one is not really known and ratio scales tell us about the order, the exact value, and an absolute zero–which allows for a wide range of both descriptive and inferential statistics to be applied.

Table 1 The description of feature in the raw data

Feature	Data Type	Description
ID	Nominal	Identification number of each Employee
Name	Nominal	Name of each employee
Date	Nominal	Analysis horizon
Agency	Nominal	Category of government jobs (Department of defense, education, …)
Station	Nominal	Description of the position of employee in city, country or state
Age	Ordinal	Age of each employee
Education	Nominal	Education level of each employee (High-School, Bachelor, Master, …)
Pay Plan	Nominal	Structure payment of employees according to their responsibilities
Grade	Nominal	Federal government's system for categorizing and defining jobs
LOS	Ordinal	Length of service of each employee
Occupation	Nominal	Job description of employees which is defined by office of personal management(OPM)
Category	Nominal	Classification of employees according to their skills (Computer Technician, Physical Science Trainee, …)
Pay	Ratio	Annual income of each employee
Supervisory Status	Nominal	Description of the type of supervisory responsibilities of employees
Appointment	Nominal	Categorization of service to permanent or non-permanent
Schedule	Nominal	Classification of employees corresponding to hours of work (Full-Time, Part-Time, …)
NSFTP	Nominal	Categorization of full-time, permanent employees to non-seasonal or not non-seasonal

➢ Verifying Data Quality

The necessitate of verifying data quality is that quality of the raw data is one of the crucial factors to have reliable results. Without data preprocessing, data mistakes such as missed data, duplicated data or outliers will detract from the quality of data mining.

The main focus of this project is on the file for the third quarter of 2004 and 2014, which shows the administration of Bush and Obama, respectively.

The reason of selectin of these files is to investigate analyses on what each president did before reelection.

The quality of the data and data cleaning is disused following.

ID:
There were no missing data for ID. However, some IDs were duplicated because the frequency of same ID number was more than one. It begs the question whether it is needed to remove duplicated IDs or not. Checking the repeated IDs, showed that no mistake had happened because those repeated IDs belonged to the same people who worked in different agencies at separate times. Thus, they have been kept the data as it was.

Name:
The most frequent name was OPM. At first, it seemed like a mistake. But, checking the data showed that OPM was used for employees who wanted their name to be withheld by OPM, thus no missing data for name is reported. Moreover, there is no problem with duplication of names, because there were some employees with the same popular name which was normal.

Date:
All the employees had the same date because analysis was based on all employees of the same period. Also, there were no missing data because the frequency of date was equal to the number of all employees.

Agency:
Agency VATA had the most frequency of duplicated agency for both presidents. However, it was not a mistake because many employees can work same agency. There were also no missing data reported.

Station:
Unfortunately, the data of the most stations were missed. After the missed data, 110010001 had the most frequency for both presidents. Duplication of data of a station will not be a problem since, many employees can work at the same station. Because of the frequency of missed data, there is no deeper analysis.

Age:

There were some missing data found. However, they have been kept for statistical purposes and have been replaced with "NA". In addition, the range of the age of employees was between 17 to 72 and there was no outlier for data file of Bush. However, two outliers with value of "12" and "13" have been found for data file of Obama. Checking other features of the same employees uncovered that these values needed to be removed. First, the amount of their payment was "$5" and "$8"! which didn't make sense. In addition, their LOS was "080164" and "1224" however no one can live this much! After removing the outliers, the range of the age for this file was 17 to 72 which seemed normal. There were also duplicated data for age which couldn't be a mistake since many of the employees could have the same age.

For statistical calculation, it was needed to change the range of the age to numeric, so new variable "numeric age" have been defined and replaced each interval with a number in between.
(E.g. "15-19" →"17")

Education:
The data file contained some missed values of this feature. However, employees with this missed feature have been kept as "**" because of statistical analysis. This feature also had duplication with "bachelor" mode which was normal since employees could have same level of education.

Pay Plan:
The data file contained some missed values of this feature. However, employees with this missed feature have been kept as "**" because of statistical analysis. This feature also had duplication with mode "GS" which was normal since employees could have same pay plan.
The most frequent pay plan was "GS" for both presidents.

Grade:
The data file contained some missed values of this feature and because of the lack of information about this feature the whole feature have been removed. This feature had duplication with "13" and mode "12" for Bush and Obama, respectively, which was normal since employees could have same job category grade.

LOS:
There were some missing data found. However, they have been kept for statistical purposes and have been replaced with "NA". In addition, the range of LOS of employees was between 1 to 32 and there was no outlier for data file of Bush. However, two outliers with value of "80164"and "1224" have been found for data file of Obama. However, no one can live this much! After removing the outliers, the range of LOS for this file was 1 to 32 which seemed normal. There were also duplicated data for LOS which couldn't be a mistake since many of the employees could have the same LOS.

For statistical calculation, it was needed to change the range of the LOS to numeric, so new variable "numeric LOS" have been defined and replaced each interval with a number in between.
(E.g. "<1" →"1", "1-2" →"1.5")

Occupation:

The data file contained some missed values of this feature and because of the lack of information about this feature the whole feature have been removed. This feature had duplication, which was normal since employees could have same occupation.

Category:
The data file contained some missed values of this feature and because of the lack of information about this feature the whole feature have been removed. This feature had duplication which was normal since employees could have same category.

Pay:
In the raw data pay was given as character which was not useful for statistical purposes, so it has been changed it to numeric. There was no outlier for the data file of Bush, but there were 2 outliers "$5" and "$8" in the data file for Obama. Checking other features of the same employees uncovered that these values needed to be removed. There were some missing data that have been replaced them with" NA" and kept for statistical purposes. There were also duplicated data that was normal because different employees could be paid same.

Supervisory Status:
The data file contained some missed values of this feature. However, employees with this missed feature have been kept as "**" because of statistical analysis. This feature also had duplication with mode "2" that described as supervisor or manager. Thus, the duplicated data have been kept.

Appointment:
The data file contained some missed values of this feature and because of the lack of information about this feature the whole feature have been removed. This feature had duplication which was normal since employees could have same Appointment.

Schedule:
The data file contained some missed values of this feature. However, employees with this missed feature have been kept as "**" because of statistical analysis. This feature also had duplication with mode "F" described as full-time status. Thus, the duplicated data have been kept.

NSFTP:
The data file contained some missed values of this feature. However, employees with this missed feature have been kept as "**" because of statistical analysis. This feature also had duplication with mode "1" described as non-seasonal, full-time, permeant employees. Thus, the duplicated data have been kept.

Statistical Summary

The statistical summary of all feature of each U.S. government employee during the Bush and Obama administration are given in Table 2 and Table 3, respectively. The separation help identify the main difference in a glance. There is not much of statistics other than mode and frequency from nominal features like ID, name, date, station, agency, education, pay plan, grade, occupation, category, supervisory status, schedule, and NSFTP. Thus, other statistical measures cannot be addressed for these nominal data types. However, slight differences can be seen from the two administrations. For example, with a 30% increase in the population, the mode of the grades in Obama administration is decreased to 12 from 13 in Bush administration. **This might be an indication of massive promotion for lower grade employees in Obama administration. Another interesting observation is the 14.6% increase in the total number of employees in Obama administration which consistent to an assumption that the size government is ought be larger in a democratic administration compared to a GOP administration.**

On the other hand, as the ordinal data types, i.e. age and LOS, are transferred to ratio data types, there statistical analysis can be presented in detail. Therefore, the statistical analysis for pay, age, and LOS are provided in Table 2 and Table 3. Several indications can be seen from the presented statistical data. For example, the ratio of mean to median of payment for federal employees is a measure for the fairness of payments among the federal employees which is defined here as *fairness measure*. The fairness measure is 1.109 in Bush administration while it is decreased to 1.096 in Obama administration. **Adding this up to the increase in the number of employees in Obama administration may lead to a conclusion that the fairness among payments are well considered in Obama administration.** However, the range of payments in Obama administration is 47.5% is higher in Bush administration which is in countertradition with the previous statement on the comparison of the fairness measure. **It might an indicator for employment of a few number high-wage export in Obama administration to address certain challenging issues for the U.S government. This is indication is consistent with the idea that Obama administration invested in new technologies which was not the center of attention in Bush administration. The** median and mean for length of service is decreased 71% and 15% in Obama administration which is consistent with the statement that new employees are added to the U.S. government. Another justification for this interpretation is the similarity of statistical data for the age of government employees. While many parameters are similar to the administrations, the standard deviation is slightly increased in Obama administration. **A deeper Analysis is required to procure a valid conclusion, but it is fair to claim that those new employees which decreased the LOS are the somehow increasing the standard deviation.**

Extra Analysis: There is no free lunch in real world. As the size of government as well as the payment to government employees are increased in Obama administration, the national debt is increased by 67% in Obama administration [3]. Although, government spending is not limited to payments to federal employments it has an impact on the increase.

Based on the stated analysis of the statistical summary two nominal data types including schedule and education as well three ratio data including pay, age, and LOS are employed for further analysis in the next section.

Table 2 The statistical summary of employees features during Bush administration

Variable	Data Type	Mean	Median	Mode	Frequency	St. Dev	Variance	Min	Max	Range
ID	Nominal	--------	----------	003519082	5	------------	-------------	------	---------	---------
Name	Nominal	--------	----------	OPM	263754	------------	-------------	------	---------	---------
Date	Nominal	--------	----------	20040930	1189697	------------	-------------	------	---------	---------
Agency	Nominal	--------	----------	VATA	217990	------------	-------------	------	---------	---------
Station	Nominal	--------	----------	110010001	101089	------------	-------------	------	---------	---------
Education	Nominal	--------	----------	13	281352	-----------	-------------	------	---------	---------
Pay Plan	Nominal	--------	----------	GS	866999	------------	-------------	------	---------	---------
Grade	Nominal	--------	----------	13	135568	------------	-------------	------	---------	---------
Occupation	Nominal	--------	----------	0303	56392	------------	-------------	------	---------	---------
Category	Nominal	--------	----------	A	425141	------------	-------------	------	---------	---------
Pay	Ratio	61,572	55,472	26,699	11062	29,021.22	842,231,198	11,113	270,882	259,769
Supervisory Status	Nominal	--------	----------	2	130412	------------	-------------	------	---------	---------
Appointment	Nominal	--------	----------	10	680244	------------	-------------	------	---------	---------
Schedule	Nominal	--------	----------	F	1067919	------------	-------------	------	---------	---------
NSFTP	Nominal	--------	----------	1	1005585	------------	-------------	------	---------	---------
LOS(Numeric)	Ratio	13.56	12.0	17.0	189135	9.74179	94.90248	1.0	32.0	31.0
Age(Numeric)	Ratio	45.16	47.0	52	204391	11.13111	123.9016	17.0	72.0	55.0

Table 3 The statistical summary of employees features during Bush administration

Variable	Data Type	Mean	Median	Mode	Frequency	St. Dev	Variance	Min	Max	Range
ID	Nominal	--------	---------	008319571	5	-----------	--------------	-----	---------	---------
Name	Nominal	--------	---------	OPM	302894	-----------	--------------	-----	---------	---------
Date	Nominal	--------	---------	20120930	1350593	-----------	--------------	-----	---------	---------
Agency	Nominal	--------	---------	VATA	288957	-----------	--------------	-----	---------	---------
Station	Nominal	--------	---------	110010001	115541	-----------	--------------	-----	---------	---------
Education	Nominal	--------	---------	13	359181	-----------	--------------	-----	---------	---------
Pay Plan	Nominal	--------	---------	GS	949617	-----------	--------------	-----	--------	---------
Grade	Nominal	--------	---------	12	176307	-----------	--------------	-----	---------	---------
Occupation	Nominal	--------	---------	0301	67327	-----------	--------------	-----	---------	---------
Category	Nominal	--------	---------	A	516526	-----------	--------------	-----	---------	---------
Pay	Ratio	78,782	71,901	155,500	13001	39,068.82	1,526,372,430	15,120	398,332	383,212
Supervisory Status	Nominal	--------	---------	2	162350	-----------	--------------	-----	---------	---------
Appointment	Nominal	--------	---------	10	718746	-----------	--------------	-----	---------	---------
Schedule	Nominal	--------	---------	F	1224315	-----------	--------------	-----	---------	---------
NSFTP	Nominal	--------	---------	1	1158845	-----------	--------------	-----	---------	---------
LOS(Numeric)	Ratio	11.77	7.00	7.00	165710	9.29664	86.42752	1.00	32.0	32.0
Age(Numeric)	Ratio	45.07	47.0	52	116828	11.99606	143.9056	17.0	72.0	55.0

Analysis and Visualization of Selected Data Types

As discussed, five data sets are selected for a deeper analysis in a hope to uncover a more detailed comparison of the effect of POTUS on the U.S. federal employment.

A. Age (Ratio)

Although age was an ordinal data type, it is transferred to ratio data type to enable further analysis. The given data is represented for Bush and Obama administrations are presented in Fig. 1 and Fig. 2., respectively. As the ages can be represented by certain interval, a histogram figure is employed. The histogram is represented by 12 intervals in the range given data, where each interval represents a 5 years period.

As federal employees are getting older over time, it might be natural to see a shift in the age for a few years. However, it is not just the increase in the age of employees that reshaped the age distribution during Obama administration as the variance as well as total number of employees are increased. A lower number of very young (under 20) federal employees are shown in Obama administration. It might be indicators with the assumption that a higher number of employees with higher level of skills and education are employed. **This observation can be verified as the number of U.S federal employees in their 20s and early 30s are significantly increased in Obama administration.** This increase is also the reason for the increase the standard deviation of the age distribution of U.S federal employees in Obama administration.

Even with considering the shift in the age of employees since Bush administration, the number of employees who were in their 40s and 50s is also slightly increased in Obama administration. **This data is consistent with the statement made while discussing the summary of payments to the U.S federal employees that a few number of expert with high wage are added. It is expected that these new experts with high wage are those who are in their 40s and 50s that are shown here.**

Extra Analysis: The life expectancy of people within united states in increased from 77.49 in 2004 to 78.74 in 2012 [4]. This 1.25 years of increase might be related to the higher number of U.S federal employees in their 70s' in Obama administration.
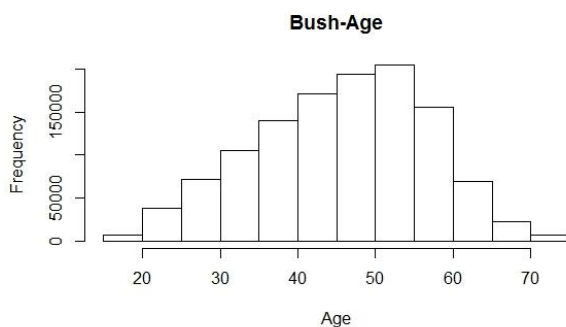
Fig. 1. The age distribution in Bush administration
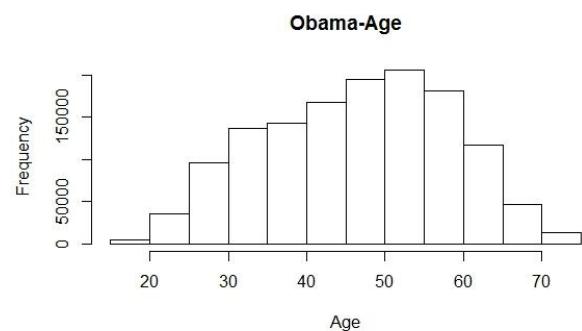
Fig. 2. The age distribution in Obama administration

B. Education (Nominal)

The education of the US. federal employees for Bush and Obama administrations, which is a nominal data type, are visualized in Fig. 3 and Fig. 4, respectively. As the presented data is nominal, the code for each type of education is presented in Table 4 to have a better understanding of the given Figures as bar since they cannot be properly aggregated.

Table 4 The description of codes provided for each level of education

Code		*	1	2	3	4
Description	No Data Reported	Invalid	NO FORMAL EDUCATION OR SOME ELEMENTARY SCHOOL - DID NOT COMPLETE	ELEMENTARY SCHOOL COMPLETED - NO HIGH SCHOOL	SOME HIGH SCHOOL - DID NOT COMPLETE	HIGH SCHOOL GRADUATE OR CERTIFICATE OF EQUIVALENCY
Code	5	6	7	8	9	10
Description	TERMINAL OCCUPATIONAL PROGRAM - DID NOT COMPLETE	TERMINAL OCCUPATIONAL PROGRAM - CERTIFICATE OF COMPLETION, DIPLOMA OR EQUIV	SOME COLLEGE - LESS THAN ONE YEAR	ONE YEAR COLLEGE	TWO YEARS COLLEGE	ASSOCIATE DEGREE
Code	11	12	13	14	15	16
Description	THREE YEARS COLLEGE	FOUR YEARS COLLEGE	BACHELOR'S DEGREE	POST-BACHELOR'S	FIRST PROFESSIONAL	POST-FIRST PROFESSIONAL
Code	17	18	19	20	21	22
Description	MASTER'S DEGREE	POST-MASTER'S	SIXTH-YEAR DEGREE	POST-SIXTH YEAR	DOCTORATE DEGREE	POST-DOCTORATE
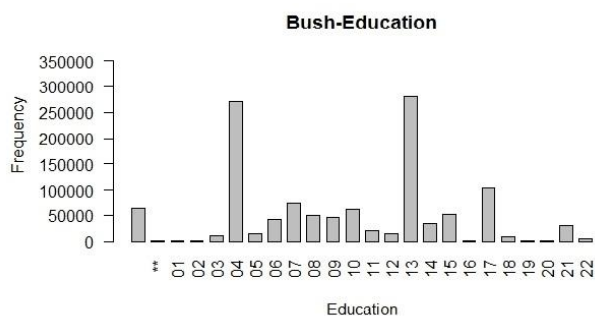
Fig. 3. The distribution of education of U.S federal employees in Bush Administration
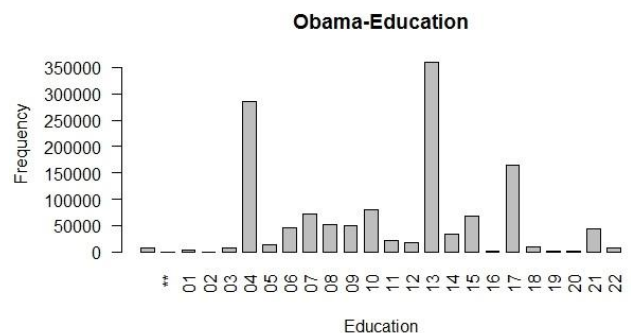
Fig. 4. The distribution of education of U.S federal employees in Obama Administration

The number of federal employees with no data on their education is decreased in Obama administration which is an indicator for transparency in the administration as well as the advancement of data recording with the advanced in the database technologies.

The number of U.S government employees with high-school diploma is slightly decreased in Obama administration compared to Bush administration. The significant increase in the number of educated employees with Bachelor degree, Master degree, PhD degree, and even Postdocs in Obama administration is interesting. This verify the previous statements that a number of new high-skill and high-wage employees are added to the U.S government. **Many of those with these degrees are those in their 20s and 30s who are added to the pool of U.S. employees as mentioned in the discussion on the age of employees in the previous section.**

Extra Analysis: The total number of master degrees offered in U.S is only increased 25.6% [5] in Obama administration while the number of U.S. federal employees with master degree is increased in much higher rate compared to Bush administration. This comparison verifies the previous claim that the Obama administration was interested in employing more educated employees with a rate higher than the supply of individuals with master degree.

C. Length of Service (Ratio)

The Length of service is was an ordinal data type which is transferred as a ratio data type for further analysis. The summary of statistics indicated that the length of service is significantly decreased in Obama administration. Thus, a histogram is employed to aggregate the presented data as presented for Bush and Obama administrations in Fig. 5 and Fig. 6, respectively.

The highest frequency for both administrations is the 0-5 LOS. It is a great indicator that each administration is willing to employ fresh staff for their own agenda. The ratio of the new employees to the total number of employees is the same for both administrations. **This does make sense as new administration are willing to employ those are in the page as their policies.**

The comparison of the two figures indicates that some employees with 15-20 LOS lost their job in Obama administration as the number of employees with around 10-15 LOS in Bush administration is higher than employees with 15-20 LOS in Obama administration. This might be related to the policies regarding the advancement in technologies. **Those employees with lower level of education are replaced by younger and less experienced employees with higher level of education as mentioned in the analysis on education.**

Extra Analysis: To check the claim that a number of federal employees are fired in the first term of Obama, a search over the media is conducted. It turns out that the presented interpretation was true according a piece from Washington Post back in 2010 [6]
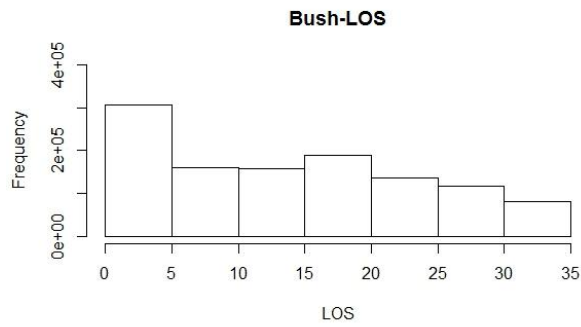
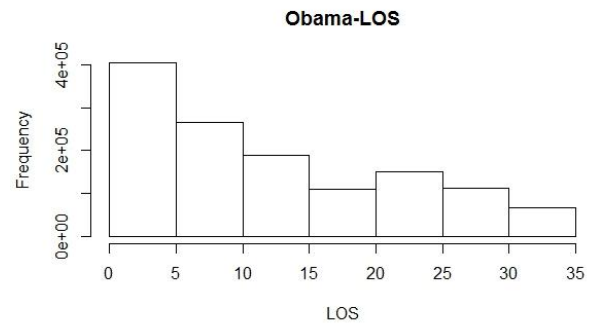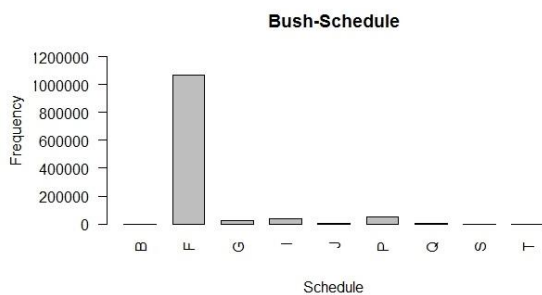Fig. 5. The distribution of LOS of U.S federal employees in Bush Administration

Fig. 6. The distribution of LOS of U.S federal employees in Obama Administration

D. Schedule (Nominal)

The working schedule of U.S federal employees is presented in bar plots as shown in Fig. 7 and Fig. 8 for Bush and Obama administrations, respectively. The codes employed to present each type of working schedule is given in Table 5.

Table 5 The description of codes provided for each working schedule

Code		*	B
Description	No Data Reported	Invalid	BAYLOR PLAN
Code	F	G	I
Description	FULL-TIME	FULL-TIME SEASONAL	INTERMITTENT
Code	J	P	Q
Description	INTERMITTENT SEASONAL	PART-TIME	PART-TIME SEASONAL
Code	R	S	T
Description	A SCHEDULE REQUIRED WHEN AN EMPLOYEE HAS ELECTED PHASED EMPLOYMENT/PHASED RE	PART-TIME JOB SHARER	PART-TIME SEASONAL JOB SHARER

Fig. 7. The distribution of working schedule of U.S federal employees in Bush Administration
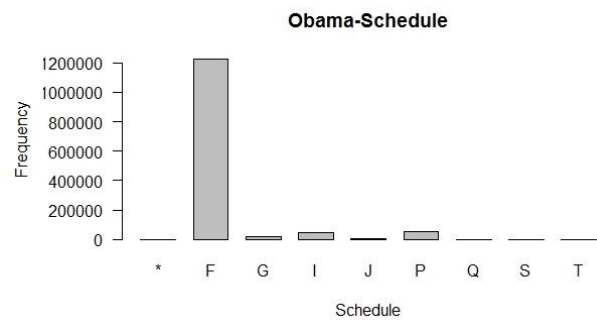
Fig. 8. The distribution of working schedule of U.S federal employees in Obama Administration

As expected a great majority of U.S employees are on full-time schedule. Also, it clearly shows the increase in the full-time employee during Obama administration. It is interesting the ratio of full-time employees to others is also increased in Obama administration.

> *Extra Analysis: A comparison of the overall employment in U.S with the employment with U.S. federal government reveal an interesting trend in federal government. As it can be conferred from the comparison of Fig. 9 and Fig. 10, in quarter of interest for Bush administration in 2004, full-time employment was around five time more popular than part-time employment for all U.S employers private or government. The same trend is true for the quarter year of interest in Obama administration in 2012.* ***Thus, it is fair to claim that the full-time to part time ratio for U.S federal government employees is totally different from the rest of U.S economy.***
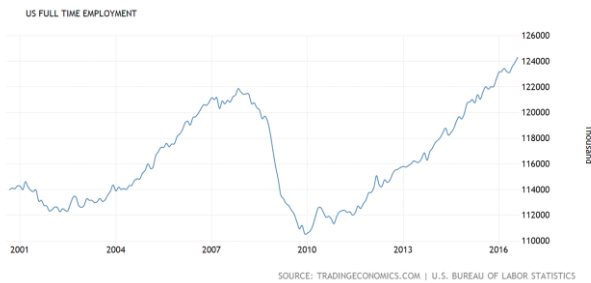
Fig. 9. The history data for the full-time employment in U.S [............]
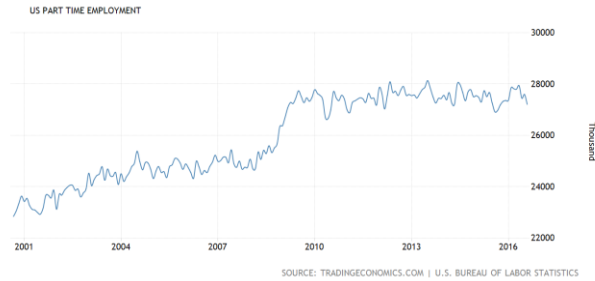
Fig. 10 The history data for the full-time employment in U.S [............]

E. Pay (Ratio)

The payment for each U.S government employee is a ratio that is suitable for histogram presentation as given for Bush and Obama administrations in Fig. 11 and Fig. 12, respectively. The range of the payments is increased in Obama administration as discussed in the summary of statistics.

It is interesting that the number of people with lower than $50k payment is decreased in Obama administration compared to Bush administration while the high-wage payment of more than $150 is
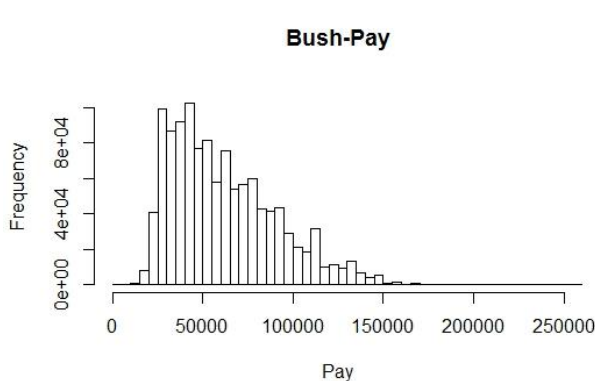

Fig. 11. The distribution of payments of U.S federal employees in Bush Administration
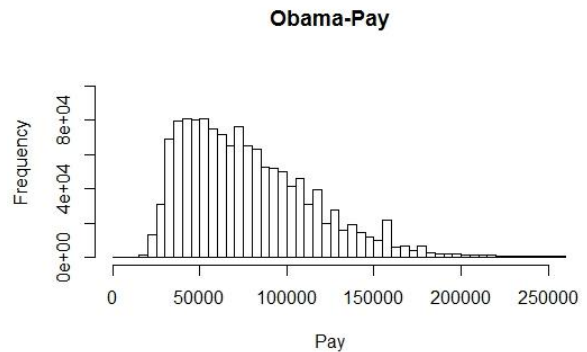

Fig. 12. The distribution of payments of U.S federal employees in Obama Administration

significantly increased. **This interpretation is consistent with the statement that many high-educated, and young employees as well as some high-skilled and high-wage older experts are added to U.S. government in Obama administration.** As predicated from the statistics in the summary, the increase in the range of payments in related to a few experts employed by Obama administration. Therefore, it is true to claim the fairness among U.S federal employees became better in Obama administration. The diversity of payments is increased in Obama administration. Although inflation rate is not considered here, the increase in the federal payments is way more than inflation rate.

Extra Analysis: ***The impact of re-election!***

One approach to clarify the data regarding the payment of U.S government to its employees is the effect of re-elections. The provided data is captured for the last quarter of the fiscal years before re-election for president Bush and President Obama. A comparison of the distribution of payments in this quarter with another two quarters two year before their re-election and two years after their re-election could reveal how much does the impact of re-election is. Considering the inflation rate, in Fig. 13 and Fig. 15 the two distributions for Bush administration is given while they are provided in fig. 14 and Fig. 16 for Obama administration. This data also could reveal that if they have changed their policies after re-election. It turns out that the distribution of payments did not change much in the year of election compared to the middle of the first term for Bush administration as revealed by comparison of Fig. 11 and Fig. 13. However, the distribution of payments is shifting for the second term. There are
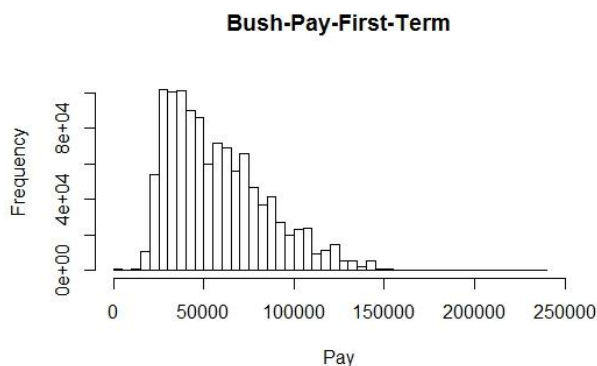

Fig. 13. The distribution of payments of U.S federal employees in Bush Administration in the middle of first term
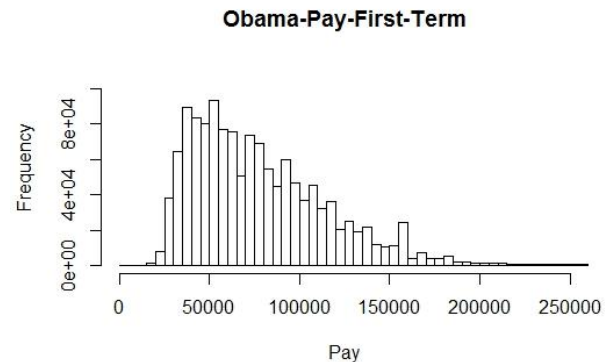

Fig. 14. The distribution of payments of U.S federal employees in Obama Administration in the middle of first term
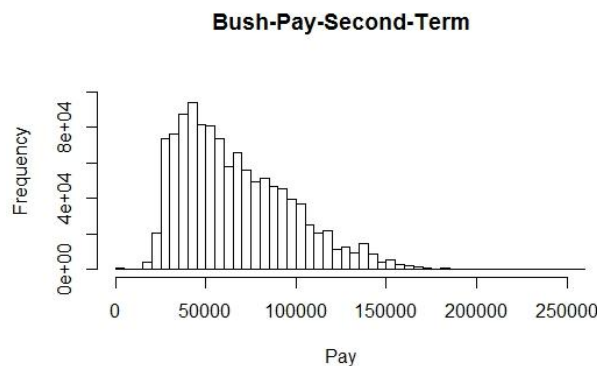

Fig. 15. The distribution of payments of U.S federal employees in Bush Administration in the middle of second term
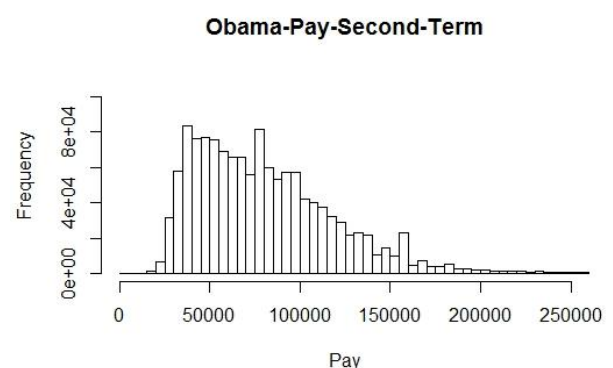

Fig. 16. The distribution of payments of U.S federal employees in Bush Administration in the middle of second term

many more over $150k payments in the second term and much less very low payments in the second term. Thus, it is fair to say that payments did not simply change in the event of election itself while they are increased after the election. The interpretation is somehow different for Obama administration. The distribution of payments in the election year is different from the middle of the first and it is more like the second term. It is fair to state that many low payment employees got the raise in the election year in Obama administration. The distribution did not change much though after the However, to be fair, the distribution of payments did change a lot in the first term of Obama administration compared to the second term of Bush administration.

➤ Exploring the Relationships among Various Data Types

As occurred many times in the analysis of selected data types, many interpretations are related to information in other aspects of data set given in other data types. For example, it is argued that Obama administration employed a larger number of educated people "and" pay them more. There are other kinds of such relationships in the presented data set which is explored in more detail in this section. The following relationship are studied.

A. Pay vs. Schedule

The payment of various employees is indeed related to their working schedule. To provide a clear picture of this relationship the median of payments for each working schedule is compared for Bush and Obama administrations as given in Fig. 17. The codes for each type of working is already given in Table 5. Additionally, to provide a fair comparison the payments for Bush administration are converted considering the inflation rate according the publicly available data [7]. In general, the median payments are higher for various types of working schedule in Obama administration which is consistent with the statement that payments were higher in Obama administration. However, the payments for the part-time jobs were higher in Bush administration. It is also interesting that most of working schedules have much lower payment compared to the full-time schedule expect one of them which is part-time job sharer.
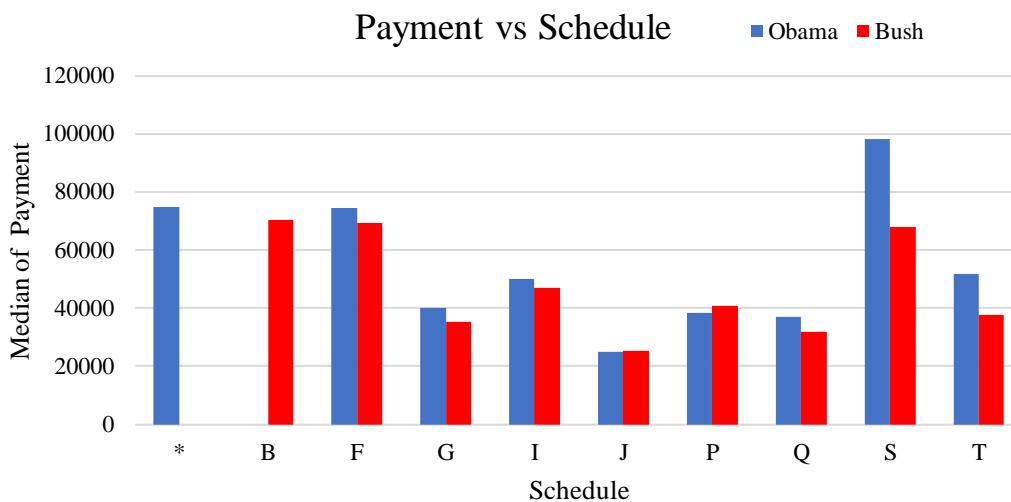

Fig. 17. The relationship among payment and working schedule of employees in Bush and Obama administrations

According to Fig. 18, it is not a surprise to see that the payment of those with higher education is higher compared to those with lower education. The codes for various level of education is given in Table 4. It is an incremental increase over the education level which is consistent in both agencies. However, for those with high school diploma the payment was notably higher in Obama admiration. Another interesting Observation is related to the individuals with post bachelor degree who has the highest payments in both administrations. These are probably those experienced employees that are going to take higher education while employed. Thus, they have the highest payments even higher than PhDs and postdocs.

It is also interesting that for highly educated employee the payments are higher in Bush administration. **This is another verification with the claim that the fairness among employees is improved in Obama administrations as the high-wages are lower compared to Bush administration while it is higher for low wages.** However, the decrease in the number of those employees with high school diploma and the increase in the number of those with Bachelor and Master degree, as discussed in description of Fig. 4, play a major role in enhancing such fairness. That is the payments for individuals with bachelor and master degree is lower in Obama administration while their population is higher in Bush administration.
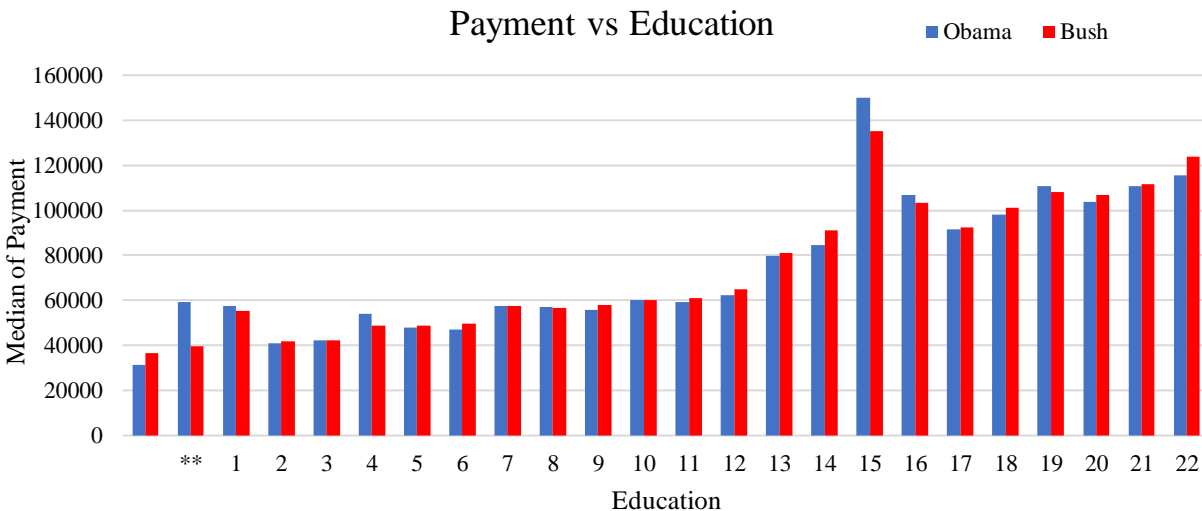

Fig. 18. The relationship among payment and education of employees in Bush and Obama administrations

C. Education vs. Length of Service

The comparison of education with LOS could reveal interesting insight from the data as given in Fig. 19. It is necessary to remember that the overall LOS is smaller in Obama administration as many new employees are added to US government. The relative relationships among the two features are almost the same for both administrations, where people with higher level of education has smaller length of service. However, there are a few exceptions to this general rule for higher educated people. With those with higher level of education than PhD degree, the median of LOS is the same. It means that in both administrations, people with PhD has similar LOS, while it seems than for the transition among different degree those with longer LOS had a better change of continuing their education in Bush

administration. It is might be a change in policies that the LOS for those in transient is decreased in Obama administration, specifically for those seeking PhD after their master.
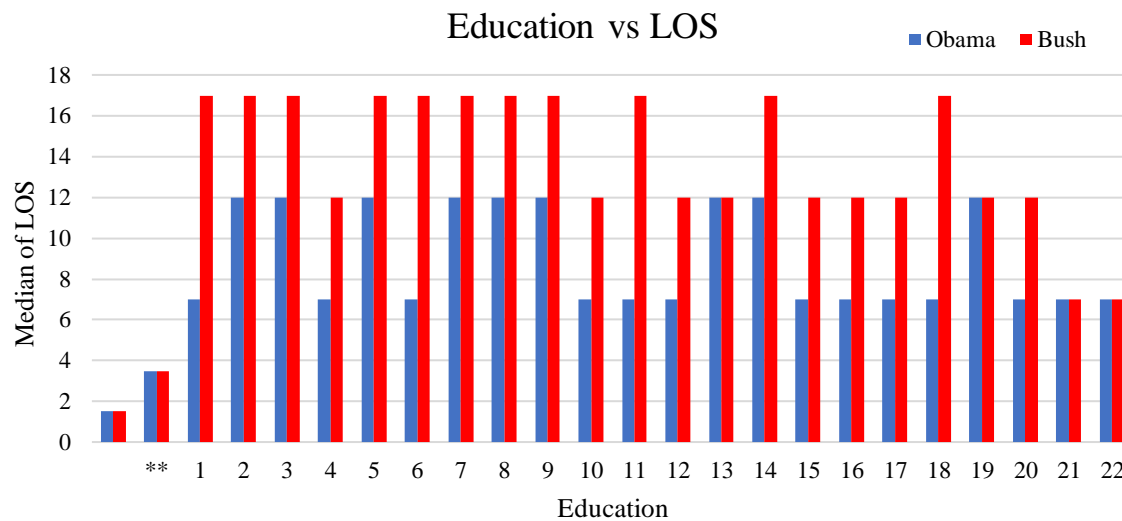
Education vs LOS

Fig. 19. The relationship among LOS and education of employees in Bush and Obama administrations

D. Pay vs. Age

These features are both ratio and can be easily interpreted in numerical analysis. With the increase in age, the median payment to employees or as called here their median income is increased to a certain level. However, such relationship is true for both administrations, the increased of payment with age reaches a saturation point in 40s. The provided data also considered the inflation rate same as subsection B of this section to ensure the validity of the comparison among the administrations.

The relationship given in Fig. 20 implies that younger employees has a higher chance of getting higher payment in Obama administration which is consistent with the previous statements on the policy of the administration to employ young high educated people. However, it also does confirm the employment of some export with high wage that might be much older. That could be the best explanation for the notable increase in the payment to employees in their 60s in Obama administration. The decline in the payment to employee in their 60s in Bush administration is also interesting. It might be impact of the increase in the number unpaid off days or the lower level of education of those employees. **It could be also another indication of general trends for democratic administrations that supposed to take care of elderly.**
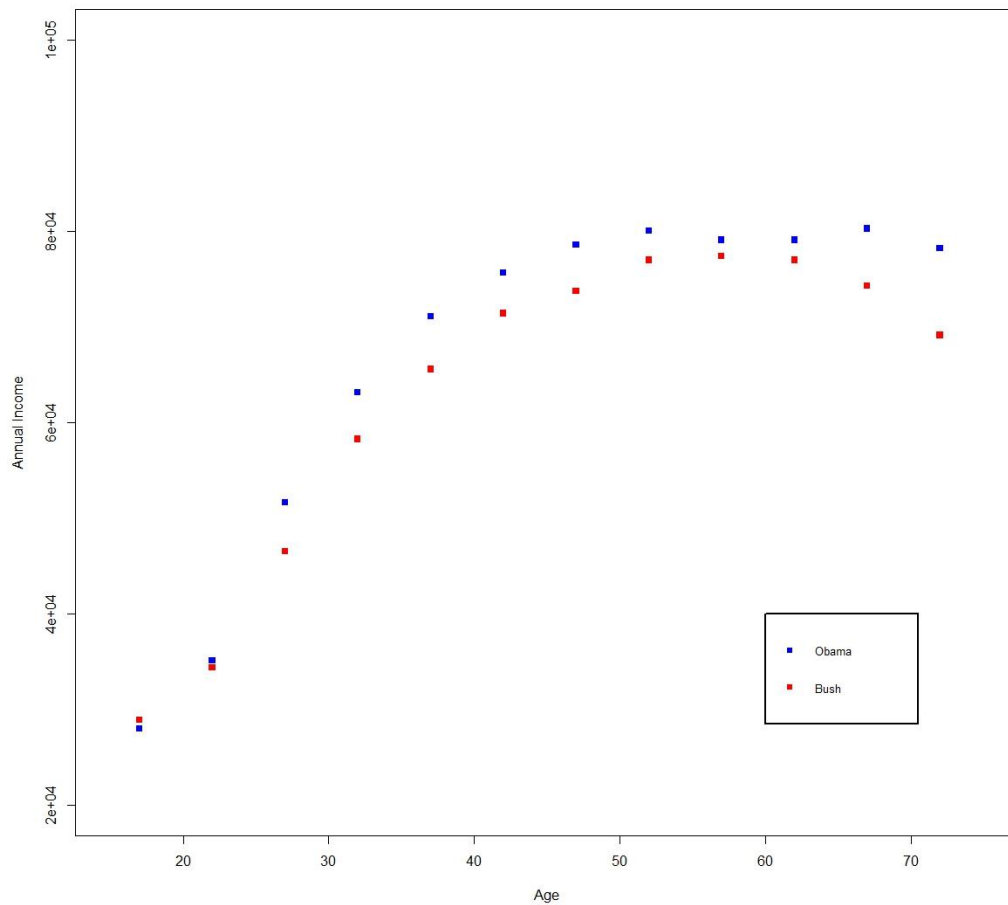
Fig. 20. The relationship among median annual income(payment) and age of employees in Bush and Obama administrations

E. Pay vs. LOS

It is expected that with the increase in the LOS the payments for each employee, its payment is also increased. The relationship among median of payments, considering the inflation rate as subsection B of this section, and LOS is given in Fig. 21.

As expected and from verified in multiple analysis, the payments in Obama administration is higher compared to Bush administration, while it is even much higher for those with shorter LOS. This is another indication for the policy of Obama administration to employ young high educated as well as older experts for the U.S government. **The best explanation for the larger gap between the two administrations for shorter LOS compared to younger employee is the employment of those experts that might be older.**

Another interesting observation is the almost exponential increase in the payment with the increase in the LOS for both administration in early years of employment. However, the relationship further become linear for both administration and then saturated in Obama administration. The smallest gap in the payments is for those with around 35 years of service.
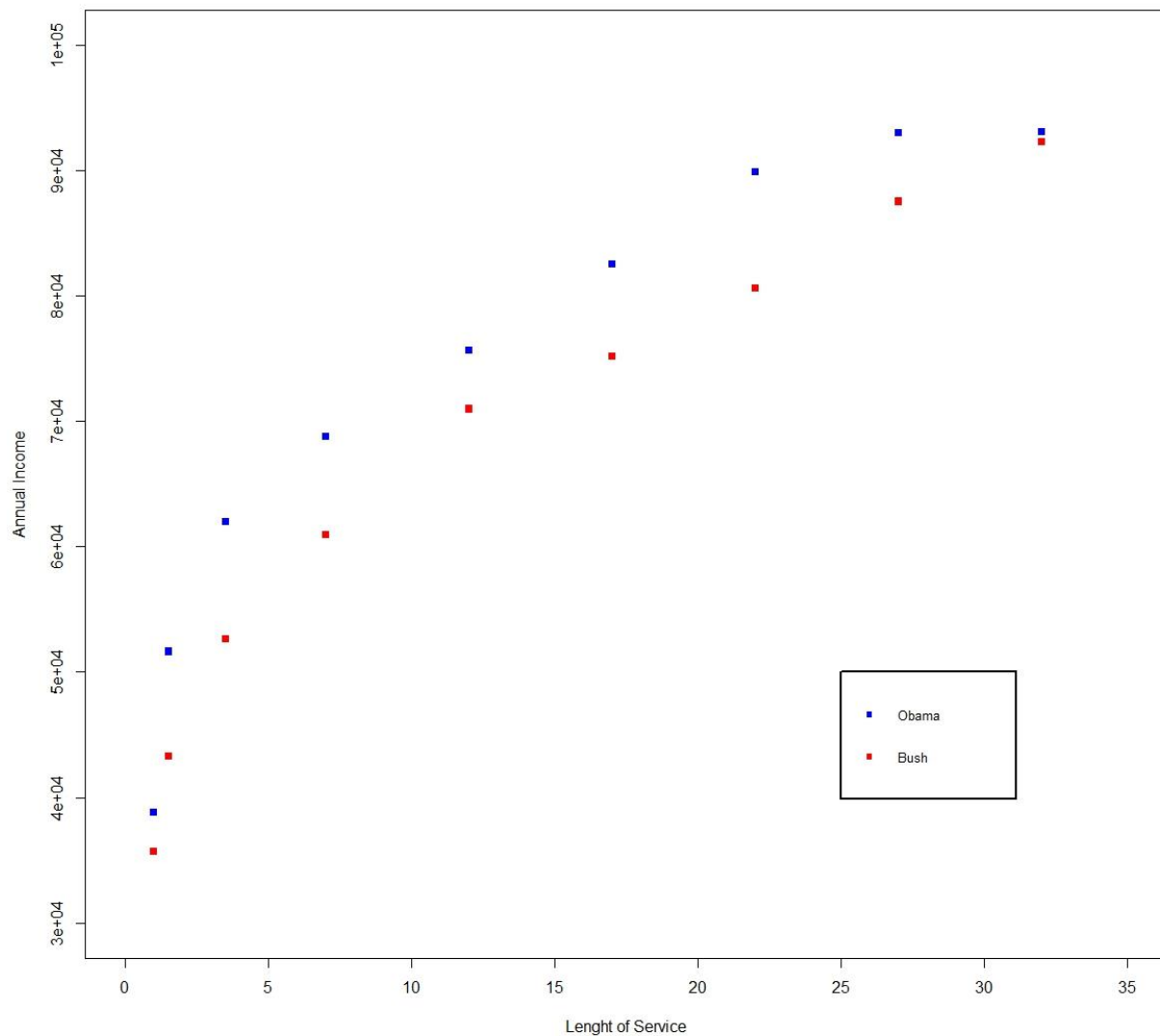
Fig. 21. The relationship among median annual income(payment) and LOS of employees in Bush and Obama administrations

➢ Exploring the Temporal Components in 40 Years of Data

The priorities of U.S. government would indeed change over time in different administrations. In this section, it is intended to analyze those changes over time. A good indicator for such change is the payment to the employees of each agency. Thus, in Fig. 22, the ratio of the median payments to the employee of various agencies to the median payments of U.S federal government employees is presented. It would render the priorities of each administration in the past 40 years. In general, the diversity of payments is increased over time.

The most balance spending is for Ford administration and then Carter administration which are both single-term administrations loosing the election to the other party. However, as the data does not include DoD, it might underestimate the spending on military which is also a major factor. With minor changes compared to Cater, Reagan administration enhance the focus on environment which become the trend for the following administrations as well. It seems that the main reason for this increase is the demand for higher educated employees to tackle the environmental issues. Thus, it is not a surprise that those with higher degree would have paid higher.

On the contrary, although Reagan administration paid higher for employees in health, such trend went the other way in the next administrations. The ratio of payments in health agencies decreased over years. Additionally, the "Law and Order" policy of Reagan administrations can be inferred from the increase in the ratio of defense and security.

During W. H. Bush administration, except heath sector, similar policies to Reagan administration are followed which is not surprise as president W. H. Bush served as VP for president Reagan.

The main changes in Clinton administration compared to W.H. Bush is increase in the environment sector which is an expected action from a democratic administration. Another very interesting and kind of surprising is the increase defense and security in Clinton administration.

Although, it is true for the twenty century that the increase in the ratio implies the increase in the attention of the government on a specific category. The understanding is totally different in the modern history of third millennium as the increase in the ratio indicated the professionalism in a field.
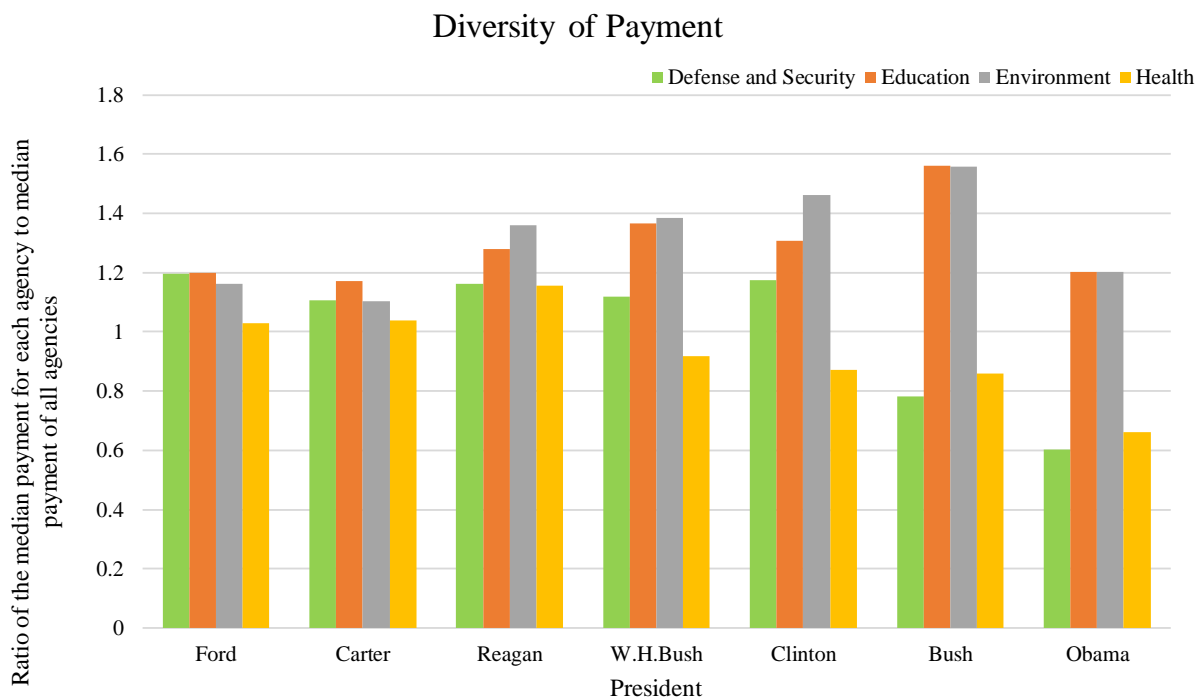

Fig. 22. The ratio of government payment to employees of various agencies in 40 years

It is very crucial to note that the increase in the ratio of the payment does not necessary mean that the center of attention is on the specific agency in the modern history of U.S. It is also an indicator for the professionalism in the agency. It is not a secret that the focus of Bush administration was on defense and

security. Thus, the sudden decrease in the ratio of payments might be misleading while it is an indicator for new recruitment who are expected to paid lower. **Therefore, the sudden decrease in the ratio of defense and security is the indicator for the new recruitments in the sector.** The data shows that the number U.S federal employees in this sector got doubled in Bush administration. That would make sense as it was the administration agenda. **The same story is true for health sector is Obama administration as health insurance was the Obama agenda, the sudden decrease in the ratio of payments in health sector is an indicator for the new recruitments.** As discussed in the last two sections, it is true that the payments would be lower for those with shorter LOS. Additionally, as the level of required education for defense and security is lower than that in health service, the decrease in the ratio of payments in Bush administration was larger as the payments to employees with lower level of education was smaller.

Extra Analysis: A very good indicator of the effect of this is data regarding the number of prisoners in U.S. prison during the same years. The impact of "Law and Order" in Reagan administration and Clinton policies has led to doubling the increase in the number of prisoners in 80s and 90s from XX to YY in 2000. However, in the first time in the last 40 years, the number of U.S prisoners was decreased in Obama administration as shown in Fig. 23.
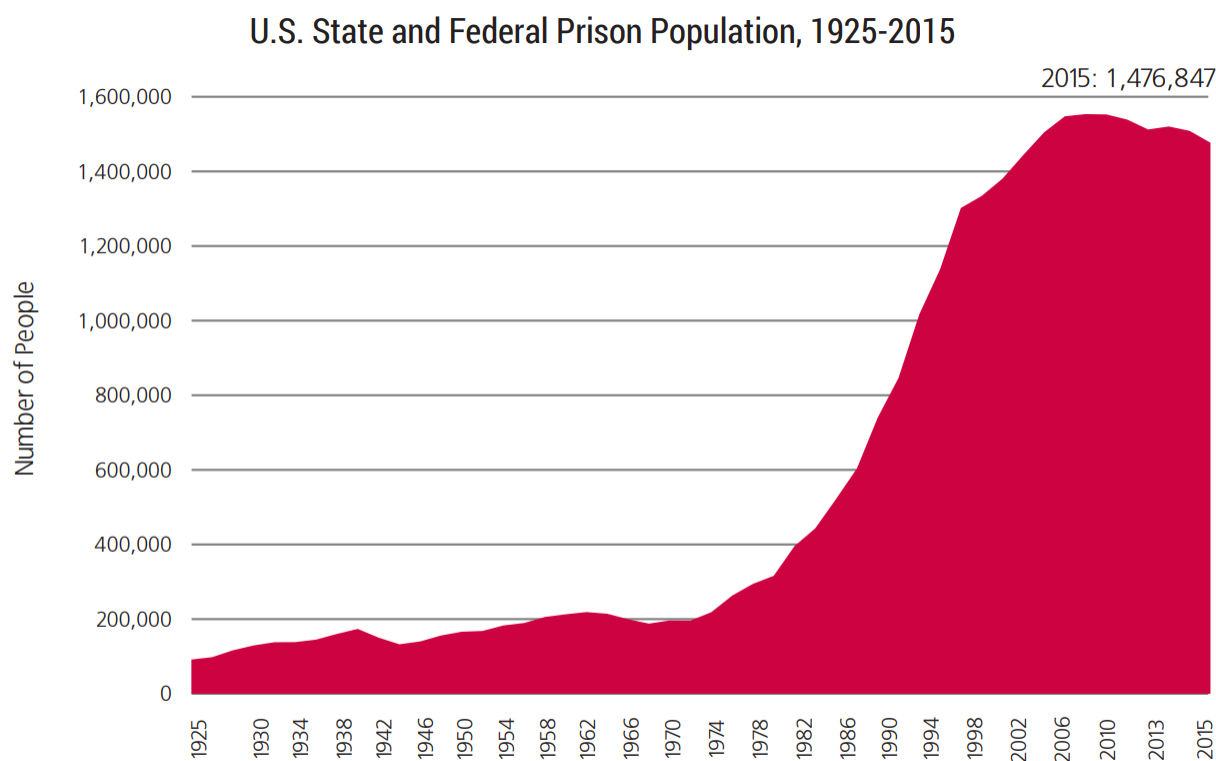
Fig. 23. The population of prisoners in US during since 1925 [8]

➢ **Conclusions**

The data of the U.S federal employee is a potential indictor for the policies of each U.S administration. In the report, an extensive analysis is performed over various feature of each U.S federal employee during Bush and Obama administrations. The following conclusion is made from mining the data in this project.

- Health insurance and modern technologies were the agenda for Obama administrations
- Security and Defense were the agenda for Bush administration
- Obama administration employed many new employees while many of them were young and high-educated, some of them were older experts.
- Obama administration paid higher wages to federal employee, particularly took care of elderly federal employees
- The fairness among employees is improved in Obama administrations
- The ratio of full-time to part-time employees is larger than average U.S economy
- Both administration hired new employees for their own agenda, while Obama hired more and paid them more as the nature of agenda required skilled staff
- Re-election did a notable impact on the Obama administration

The analysis over other additional feature would certainly clarify some perceptions provided in this report. Analysis over the number of employees over 40 year, impact of hurricane Katrina and hurricane Sandy, impact of Iraq and Afghanistan wars could be also interesting.

➢ **Reference**

[1] [https://www.buzzfeed.com/jsvine/sharing-hundreds-of-millions-of-federal-payroll-records?utm_term=.lsVY98vj8#.leWRm7nw7]

[2] [https://www.forbes.com/sites/mikepatton/2013/01/24/the-growth-of-the-federal-government-1980-to-2012/#4bf6d1f317b6]

[3] [https://www.thebalance.com/us-debt-by-president-by-dollar-and-percent-3306296]

[4] *[https://data.worldbank.org/indicator/SP.DYN.LE00.IN?name_desc=false].*

[5] *https://nces.ed.gov/programs/coe/indicator_ctb.asp*

[6] *[http://www.washingtonpost.com/wp-dyn/content/article/2010/12/03/AR2010120303160.html]*

[7] [http://www.inflation.eu/inflation-rates/united-states/historic-inflation/cpi-inflation-united-states.aspx]

[8] [http://sentencingproject.org/wp-content/uploads/2016/01/Trends-in-US-Corrections.pdf]