New York City Civil Service Jobs

Armenoush Aslanian-Persico

May 2018

# Note to the Reader

The following is a Capstone project for the Master of Science in Data Science degree at the CUNY School of Professional Studies.

Please view the interactive website for this project:

<https://armenoush.shinyapps.io/nycjobs/>

# Data

## Background

The City of New York employs over 300,000 individuals in full-time and part-time positions. These people work in over 60 city agencies, including large agencies like the Department of Transportation, and small agencies like the Department of Cultural Affairs.

Thanks to the city’s 2012 Open Data Law, city agencies have increasingly made their data public. While the public data is sometimes incomplete, intentionally censored, anonymized, or not maintained, it can still be valuable for research purposes.

This project focuses on two datasets available through NYC’s Open Data website. The first is a list of all employees and their salaries, referred to here as the payroll dataset. The second dataset is a list of job titles, title codes, and other title information, referred to here as the titles dataset.

The primary tools used for this project were R, RStudio, RMarkdown, Shiny, Plotly, ggplot2, and Github, and supporting libraries including dplyr, tidyr and data.table. The source datasets were entirely available on the NYC Open Data website, and supporting information was available on NYC.gov.

## Questions

The following questions were used to guide the data analysis for this project.

* What can we learn about titles, hiring, and salaries?
* What can we learn about city agencies?
* What are good agencies to work for?
* What are good titles to work in?
* What is the impact of unions?
* What trends are visible in the available time period?

## Payroll dataset

The payroll dataset is partially anonymized, with first initial and last name. It contains salary, year, agency name, agency code, and title code. Importantly, it does not contain the person’s job title.

The payroll dataset included only the 2014 and 2015 calendar years until late April 2018, when it was updated to include the 2016 and 2017 calendar years. This project was updated to include the additional years.

The payroll dataset includes 1,161,078 records. This has a consistent split of observations across four years: 269,838 records for 2014; 300,653 records for 2015; 278,869 records for 2016; 311,718 records for 2017. Each record represents a unique employee.

## Titles Dataset

In the titles dataset, each observation represents a unique civil service title. A basic knowledge of the civil service system is necessary to understand this dataset.

New York City government, like other municipal governments, functions according to the civil service system. This system requires each employee to be serving in a civil service title. An employee may be hired through a job posting or through a civil service exam list. Employees who are hired provisionally are potentially at risk of being terminated if a civil service exam is scheduled for their title. Employees who attain a permanent civil service title have a very low risk of being terminated. A person who takes and passes a civil service exam for a title can become permanent in that title if they score well enough and are “picked up,” or chosen, by an agency.

As a result of the civil service system, these titles significantly shape a person’s city career. Each title has a minimum and maximum salary. Therefore, people continue to take civil service exams throughout their careers to attain titles with higher salary maximums, and may also continue to apply to city jobs in different titles for the same reason.

The titles dataset required several transformations to achieve a cleaned version for efficient matching with the payroll dataset.

The titles dataset originally included multiple subtitles belonging to the same parent title. For efficient matching, multiple subtitles were combined into one title in the cleaned dataset. For example, a Sign Language Interpreter may have five levels of title, where an employee may be promoted through the levels. The count of subtitles was captured in the CountSubTitles column.

In the cleaned titles dataset, the lowest salary of the lowest title and the highest salary of the highest title was captured. This was represented by the minimum of the minimum rate, Min.MinRate, and the maximum of the maximum rate, Max.MaxRate. For some titles, both of these values were zero. The SalaryMinMaxValid column was created as an indicator column to flag these observations as invalid.

After the cleanup and consolidation of the titles dataset, there were 4,424 unique titles.

## Merged Dataset

When matched with the payroll set, the titles dataset can provide title information about city employees. This match was performed based on the title code key.

See Appendix A, Part 2 for the details of this merge.

The merged dataset contains 20 variables. The variables most relevant to this analysis are listed here.

* title - the title code, alphanumeric code referring to the civil service title.
* year - the calendar year of the payroll data, which ranges from 2014 to 2017.
* agency - the name of the city agency.
* empname - the first initial and last name of the employee.
* sal - the salary earned by the employee in the given calendar year.
* Description - the civil service title name.
* SalaryMinMaxValid - a check on if the title data is displaying a valid salary for this observation.
* PayType - a category of salary level to indicate if the salary is annual, hourly, or both. A title ranging in salary from $300 to $10,000 may be hiring both annual and hourly workers.

## Dropped Observations

Of the approximately 1.1 million records, about 300,000 resulted in incomplete observations after the merge. This was due to no match found for the title code in the titles dataset. Through the Open Data website, I inquired with the Department of Citywide Administrative Services about the incomplete titles dataset and was informed that the city does not provide title codes for “non-mayoral” agencies.

Because of the lack of title information, these incomplete observations were excluded from the analysis. The discarded observations were mostly New York Transit Authority (MTA) and City University of New York (CUNY) employees. These are considered to be non-mayoral agencies, but the employees are still included in the payroll list.

The remaining complete observations were mostly annual salaries, with valid salary values, and in active titles.

See Appendix A, Part 3 for the summary details of the complete cases.

# Scope

To accurately assess hiring trends and salaries, the scope was restricted to employees earning more than $10,000 in a year. Salaries under $10K were inspected and summarized.

## Salaries Under $10K

The list of employees with salaries under $10,000 in the dataset includes trades workers such as carpenters, painters, and mechanics, which are full-time, but are listed with an hourly salary instead of annual. This causes these employees to appear as low-wage workers when they may actually earn a mid- or high-range annual salary. For clarity, these have been removed from the primary dataset instead of attempting to calculate an equivalent annual rate.

Union representation for salaries under $10K was primarily non-managerial titles, DC37 job training participants, and seasonals. This was to be expected.

Agency representation for salaries under $10K was primarily the Department of Citywide Administrative Services (DCAS), Parks, and the Police Department (NYPD). This is in line with the seasonal and part-time hiring and explains these low salaries. Parks hires many seasonal workers for summer months to maintain the parks and beaches. DCAS hires temporary and part-time workers throughout the year for custodial and other laborer duties, and also hires part-time monitors for civil service exams.

The titles most represented in the under $10K group were Monitors, Job Training Participants, and College Assistants.

See Appendix A, Part 4 for the summary details of the sub-10K salaries.

## Salaries Over $10K

Salaries over $10,000 primarily represent full-time non-seasonal jobs. An analysis by variable highlights some of the significant subgroups.

Across all four years, the Police Officers union (Patrolmen’s Benevolent Association) has the most representation, with 24,443 members earning above $10,000 in 2017; 24,049 in 2015; 23,768 in 2016; and 22,2197 in 2014. This is followed by DC37, which represents a wide range of mostly technical titles.

The city agencies representing the most $10K+ earners across all years are the NYPD, followed by the Fire Department (FDNY), and the Human Resources Agency (HRA), which provides social services.

The civil service titles most represented in the $10K+ earners group are Police Officers, Correction Officers, which are employees at city jails, and Firefighters. As each civil service title can be represented by only one agency, union distribution and civil service title distribution in these groups are similar.

Across all four years, the 10 high and 10 low outliers by title, count in title, and average salary by title were as follows.

On the high end, there were four records for a Pension Investment Advisor, making an average salary of $318,644. This was followed by three records for a Chief Actuary, at an average of $287,512. The Chief Actuary serves as a pension advisor and administrator. This was followed by the First Deputy Mayor, City Council Chief of Staff, Mayor, Deputy Mayors, and other Executive Director roles. The 10th highest average salary was $224,168, with only 45 records at or above that amount for all four years.

On the low end, college interns, community service aides, city service aides, and city seasonal aides ranged from an average of $27,524 to $18,067. Over 2000 records were represented in this tail end of the list. These may not necessarily be full-time jobs.

See Appendix A, Part 5 for the summary details of the over-10K salaries.

## Comparing Agencies

Looking at complete cases of salaries over $10,000, there are dozens of city agencies represented. A significant group of these agencies were similarly, but not identically, named. These were the Community Boards. New York City is divided into 59 Community Boards (CBs), or districts, and each has its own office with a small staff, usually fewer than five civil service employees. Each of these CBs is considered a different agency in this dataset. Looking at comparisons between Community Boards is not statistically significant due to the small size of each agency. For the purposes of this analysis, employees working for Community Boards were removed from the dataset.

It is difficult to compare agencies with unique functions and titles. For example, employees serving in Firefighter titles work only for the FDNY. Employees serving as Sanitation Workers work only for DSNY. Therefore, comparing average headcounts, salaries, or other summary statistics would be inaccurate because of the unique titles. However, there are some titles which work at many agencies, such as clerical workers, analysts, managers, and similar titles. Focusing on titles that are present at multiple agencies prevents the distortion that would occur in a broader analysis.

The 10K+ complete cases dataset was grouped and summarized by title, agency and count to determine the titles that appeared across the most city agencies. These are also referred to here as “flexible titles.” A person in one of these titles could potentially transfer to other agencies more easily than a person in a non-transferable title. Not all employees may necessarily want one of these flexible titles. For example, somebody dedicated to working in healthcare as a medical professional would probably prefer a title relevant to the Department of Health (DOH) or another social services agency, and these titles would not be available at the Department of Transportation or the Department of Finance (DOF).

See Appendix B, Part 3 for the complete list of the top 20 titles available in the most agencies.

Community Associate is the most frequently appearing title, present in 63 agencies. This is followed by Administrative Staff Analysts, who work at 62 agencies. Community Coordinator, Principal Administrative Associate, Clerical Associate, Computer Systems Manager, Executive Agency Counsel (lawyers), Administrative Manager, Associate Staff Analyst, and Community Assistant complete the top 10 most frequent titles across agencies, ranging from 61 down to 47 agencies.

Together, these 20 most frequent titles make up 116753 of the 672578 observations across all four years for complete cases of salaries over $10K. A new dataset was created to contain this specific group of employees.

Of these select titles, the headcounts and highest average salaries were as follows for all years. Executive Agency Counsels earned an average of $125,927 and numbered 1,984; Computer Systems Managers earned an average of $122,018 and numbered 5,128; Administrative Managers earned an average of $112,526 and numbered 597; Administrative Public Information Specialists earned an average of $105,464 and numbered 603; Computer Specialists (Software) earned an average of $97,897 and numbered 5915; and Administrative Staff Analysts earned an average of $96,276 and numbered 7541.

See Appendix B, Part 4 for the detailed list of the top 20 titles, headcounts, and average salaries by year.

# Findings

The following section should be read in conjunction with the Shiny app: <https://armenoush.shinyapps.io/nycjobs/>

Supplemental graphics are also available in Appendix C.

## All Salaries

View this plot on the Shiny app.

A histogram output of the distribution of all salaries over $10,000 for the year 2017 shows most salaries are below $100,000. This is a visual representation of all salaries, not only select titles. The salaries follow a mostly normal distribution with a median around $50,000 and a right skew. Based on previous analysis, the right tail consists primarily of managerial positions including Commissioners, Deputy Commissioners, Assistant Commissioners, and some Agency Counsel and technical positions.

There is a significant spike in the histogram between $85,000 and $90,000. Further investigation uncovers the makeup of this anomalous segment.

63.4% of the 85K-90K salaries are police officers at the NYPD. The next largest group is FDNY Firefighters at 26.9%, and NYPD Sergeants at 6.27%. 54 other titles make up the remaining 3% of these salaries. The NYPD holds a sizable group of employees at a higher than average salary for city employees.

See Appendix C, Part 3 for further details and visualizations of salary distributions, including the histogram exploration.

## All Salaries, In Detail

View this plot in Appendix C.

A scatterplot of the 100 most popular titles in 2017 by average salary shows that most titles have fewer than 5,000 employees serving in the title, with the majority having fewer than 2,000 employees. However, there are some outliers. Again, NYPD is the greatest outlier, with 24,443 employees in the Police Officer title earning an average salary of $70,223 for the year 2017. The runner-up outliers on the x-axis (count in title) are Correction Officers, Firefighters, and Sanitation Workers. Outliers on the y-axis (average salary) are Sanitation Superintendents, with 273 employees earning an average salary of $210,385, Assistant Corporation Counsels, with 983 employees earning an average salary of $164,111, and FDNY Battalion Chiefs, with 369 employees earning an average salary of $156,388. Other legal titles and technical titles such as engineers are also low in count and high in average salary.

A closer view of the scatterplot, adjusting the x-axis to a maximum of 5,000 count per title, shows a cluster of salaries below $80,000, as reflected in other data summaries.

See Appendix C, Part 3 for further details and visualizations of salary distributions, including the scatterplots.

## Influencers

View these plots on the Shiny app.

The first plot shows the top 10 agencies with the greatest number of jobs with salaries above $10,000. Out of the top 10, NYPD makes up 48% of this workforce, followed by FDNY with 12%, the Department of Correction (DOC) with 11.9%, Sanitation (DSNY) with 8.63%, and the Housing Authority (NYCHA) with 7.07%. Rounding out the top 10 are the Department of Education (DOE), the Department of Parks and Recreation (DPR), the Department of Homeless Services (DHS), the Department of Transportation (DOT), and the Department of Environmental Protection (DEP).

These agencies have large workforces because of the extent of their operations and, for most, the citywide nature of their work. Out of these large agencies, Correction and Housing have the most limited scopes, as they are responsible for facilities in limited geographical areas. Correction maintains Rikers Island and other city jail facilities, while NYCHA maintains public housing projects in all boroughs, but on limited plots of land. In comparison, agencies like Parks, Homeless Services, and Environmental Protection are responsible for infrastructure or services in all neighborhoods throughout the five boroughs.

This information could be valuable to job seekers. On the surface, someone looking for a full-time job with a salary above $10,000 would have the greatest chances at NYPD. However, the workforce volume alone is not an indication of ease of hiring, as many NYPD titles have prerequisites and qualification requirements such as health evaluations. The same is true for Correction Officers, Firefighters, and Sanitation workers. These are also physically taxing jobs and mostly require working outdoors.

The second plot shows the top 10 titles with the most jobs over $10,000. Out of the top 10, NYPD again makes up the largest slice, with NYPD Officers representing 32.2% of the jobs above $10,000; Correction Officers with 13.4%; FDNY Firefighters with 11.5%, and DSNY Sanitation Workers at 8.78%. The remainder of the top 10 most common titles are Community Associate, across all agencies; NYPD School Safety Agents; NYPD Sergeants; Community Coordinators across all agencies; NYCHA Caretakers; and NYPD Detectives. Community Associate is the most common title that is not physically taxing, and appears fifth in the list in order.

These subsets of the high-earning workforce are important in understanding the city’s hiring practices. These are mostly physical jobs performing daily services and have workforces that cannot be depleted without risking the physical safety of the public. As a correlation, the workers’ unions representing most of these titles have significant political influence in negotiations and city policy.

## Flexible Titles

View this plot on the Shiny app.

The “Comparing Agencies” section of this report explains the selection of these 20 titles. This plot output shows the occurrence of these titles across multiple agencies. These would be desirable titles for job seekers looking for flexibility in being able to transfer to new jobs frequently. A combination of title availability and salary range would be good background information for employees new to civil service or looking to enter civil service.

## Flexible Titles, Best Agencies

View this plot in Appendix C.

For this group of select titles, do some agencies pay more? To determine this, the list of select titles was grouped by agency and average salary, and ordered by descending salary for the agency by each title. The count of occurrences of agencies appearing in the top five best paying agencies per title was then tallied.

Out of the 20 select titles, the NYC Police Pension Fund was one of the five highest paying agencies five times. The Financial Information Services Agency (FISA), Department of Information Technology and Telecommunications (DOITT), and District Attorney (Special Projects) also each appeared five times as a high payer. The Tax Commission, Office of Management and Budget, and the Fire Pension were also high payers, as well as the other District Attorneys’ offices in each borough.

This result indicates that financially-oriented agencies may be higher payers, and again reflects the previous result of pension offices having higher than average salaries. This is also evidence of the importance of the employee pension funds in the city budget. The pension program is likely the most attractive benefit for city employees, as most will continue to receive regular paychecks in retirement amounting to a portion of their active paychecks.

This result also potentially shows the financial influence of the District Attorneys’ offices, which have a broad scope in pursuing crimes, and also receive other funding sources such as state and federal grants. This is closely related to the influence of law enforcement offices and operations in the city overall.

## Who’s Earning

View this plot on the Shiny app.

This interactive graph allows the reader to select any of the 20 titles most common across agencies and see the trends in average salary and headcount by year.

The graph also includes a reactive description of each title, based on the most recent Notice of Examination. The city holds exams for most titles approximately every four to five years. Each exam has minimum qualification requirements which are listed in the Notice of Examination. After the exam, candidates are ranked in order of their scores and receive extra points for selective certifications, military veteran, and disability status. Agencies who want to hire in these titles must select from the established list. The exams have a significant impact on hiring. In anticipation of the exam, hiring in that given title usually decreases. Between the time of the exam and the time the list is established one to two years later, hiring in that title drops significantly as agencies can only hire provisionally, or temporarily, until the list is established. Agencies are unlikely to want to hire provisionals knowing that these individuals could have their jobs terminated if they do not score highly enough on the upcoming exams. Therefore, information on the date of the most recent exam and the date of the next upcoming exam are useful to determine the current likelihood of available postings in the title.

Some examples of this phenomenon can be seen in the graphs for the Administrative Manager, Administrative Staff Analyst, and Associate Staff Analyst titles, which experience a steep headcount drop after 2015, the year of their most recent exam. Similarly, Clerical Associate and Secretary have steep headcount drops leading up to 2018, the year of the most recent exam.

## Who’s Hiring

View this plot on the Shiny app.

This interactive graph allows the reader to select any of the agencies hiring the 20 most common titles and see the trends in average salary and headcount by year.

The line plots for hiring by agency and hiring by title mostly trend up from 2014 to 2017. Significant drops in agencies occur more frequently in agencies with small headcounts, where a drop from 70 to 50 employees may look severe in terms of percentage, but the raw value can be explained by routine attrition. The y-axis in the reactive plot automatically adjusts to fit the data and must be considered before making inferences about trendlines.

## Pay Raises

View this plot on the Shiny app.

For the select group of 20 titles, the Administrative Managers had the largest percent salary increase over this time period, at over 13% between 2014 and 2017. This could be in part due to the 2016 exam, as exams prompt job changes and promotional opportunities. Administrative Staff Analysts, Executive Agency Counsels, Principal Administrative Associates, and Associate Staff Analysts all had their average salaries increase between 11% and 12% over these years. Salary increases are not necessarily an indication of individuals being promoted or given raises in their title. These increases could also indicate new hires being hired at higher salaries than incumbent employees, and therefore pulling up the average for all employees serving in the title.

## Growing Agencies

View this plot on the Shiny app.

For the select group of 20 titles, three of the top 10 fastest growing agencies were District Attorneys’ Offices. The largest by far was the Department of Correction, with an 86% headcount increase for these 20 titles combined. This may be surprising considering the city’s recently announced plan of closing Rikers Island and decreasing the number of incarcerated individuals in New York City. This workforce almost doubled from 278 to over 500 employees in this time period. DCAS, DOH, DOITT, and Housing Preservation and Development (HPD) were other large agencies with significant headcount increases for the select titles.

## Promotions

View this plot on the Shiny app.

This box plot of all individual employee salaries in the 20 common titles is illuminating. It shows large salary ranges for some titles and limited ranges for others. This information is valuable when choosing or pursuing employment in specific titles. Computer Systems Manager, a technical title requiring programming knowledge, has one of the broadest ranges with one of the highest average salaries. A person hired in this or a similarly broad-range title could have many opportunities for increases while remaining in the title. This also suggests that there are different skill and managerial levels within the same title, another indication of the title’s flexibility. Similar ranges with a large number of high-salary outliers are evident for Executive Agency Counsel and Administrative Staff Analyst.

An employee looking for choices in many agencies with room for growth could benefit from choosing one of these three titles. Again, each title has educational and experience prerequisites for hiring as listed in job postings and in the Notices of Examination. A job seeker with a long term plan to be hired by the City in one of these titles could make themselves a stronger candidate by gaining the appropriate required experience and education.

# Tools

The following are some special tools and functions that were necessary throughout the project.

## General Practices

* Commenting all code and creating headings for all chunks. This is especially valuable in a project spanning multiple months where it is not easy to remember variable names and steps.
* Dividing the project into multiple R files to allow for faster processing and fewer file dependencies.
* Exporting and importing dataframes to allow for faster processing when creating filtered subsets; setting eval=FALSE after the first export.
* Choosing and refining an appropriate scope of data to make analysis feasible within the provided time period.
* Creating new files for subset data to prevent the storage and calling of large files.
* Maintaining an organized folder hierarchy with versions and backups. Using multiple machines while also performing backups was challenging. The best solution may be working off of one external drive and regularly backing up the drive. Running the same code on multiple machines was not seamless and required updates or changes in file locations.
* Doing sampling and checks outside of R. Exporting select datasets and doing summary analysis in Excel allowed for data verification throughout the project.
* Running regular checks on structure and summary, to ensure that dataframes were not being corrupted or overwritten. Comparing numbers on paper with expected values. This was especially challenging when restarting the project in late April when a new dataset was released by the city.
* Making the code structure as flexible as possible to allow for updated datasets to be processed; avoiding the hard-coding of information and variables as much as possible.

## Data Manipulation

* df %>% mutate\_if(is.factor, as.character); forces factors into character formats.
* as.numeric(gsub('[$,]', '',...) ; converts a salary format to numeric, strips dollar symbols.
* - grep("search term", dataframe$columnname) ; removes all observations containing the search term
* read.csv(stringsAsFactors = false) ; prevents strings from being converted to factors when reading a csv
* trim <- function (x) gsub("^\\s+|\\s+$", "", x) ; removes whitespace, for example when whitespace is preventing accurate grouping of variables
* Pipe, mutate, gather, and spread functions.

## Plots

* Shiny tutorials via R and DataCamp.
* marker=list(color='#82E0AA') to fill colors in bar charts. (Shiny)
* layout(tickangle = 90) to rotate axis labels by 90 degrees. (Shiny)
* add\_pie(hole = 0.6) to create the hole in donut charts. (Shiny)
* xlim=c(0,200000) in R plots to prevent axes from being shorter than the data range.
* plot\_ly(dataframe, y = ~var1, color = ~var2, type = "box") to create a box plot (plotly)

## Shiny App

* Determining allowable image locations. Displaying hosted images in Shiny is inconsistent. Images hosted on Github and Google would not load, while other URLs such as imgur would load.
* splitLayout(cellWidths = c("50%", "50%")...) to create side-by-side plot displays.
* Using the reactive server logic to load data from multiple separate sources based on the same input value. This was accomplished with a slice approach using a filter.
* Converting year values to factors for successful time series plots.
* Clearly commenting and labeling all sections of the Shiny app to clarify the relationship between the corresponding server and UI chunks.

# Insights

## Implications

The volume of city jobs earning above $10,000 is an indication of the strength of the city workforce and its influence in sustaining a middle class in New York City. However, comparing these salaries to private sector salaries and rents in the city also highlights the difficulty in maintaining a viable lifestyle with just one city salary in a household.

The stability and growth of the city workforce for the past four years is also connected to the strength of the many workers’ unions. Unions negotiate contractual salary increases and benefits for each title, and larger unions have more of a bargaining impact. In salary negotiations, smaller unions are sometimes absorbed into larger unions or have to wait longer for negotiations to begin.

The volume of jobs earning below $10,000 is also significant, as these are positions that could potentially lead to permanent or full-time employment. At the same time, this segment of the workforce also reflects a large population without advanced or technical skills, including a significant group of job training participants receiving public assistance. One interpretation of this payroll subset is that the city has a responsibility to maintain these jobs to provide economic stability and growth opportunities to a disadvantaged population.

## Technical Challenges

Some of the technical challenges encountered in this project were as follows.

As mentioned, the dataset owners released an updated dataset in late April 2018, which contained twice as many observations as the original dataset. Much of the writeup, and some of the code, had to be reworked to accommodate the new data.

The datasets had significant cleanup issues, including: errors; messy variable names and value names with whitespaces, parentheses and other punctuations; and many abbreviations and truncations which required a close knowledge of city titles to allow for successful interpretation.

The prediction aspect of this project was only partially feasible with plot trends. Since there are no specific hire dates, the hiring necessarily had to be aggregated by year, resulting in only four data points per title or agency, which was not robust enough to build predictions and split off a test dataset. Additionally, new hire salaries are not available until one or more years after the hire, and the salary ranges are dictated by the civil service titles. Due to the required salary range in the majority of current postings, predicting a specific 2018 or 2019 salary based on postings was not feasible.

The payroll dataset made available by the city should be combined with the titles dataset, and the missing title values for non-mayoral agencies should be included. The technical work necessary to match datasets and create complete cases is not conducive to an easy public understanding of this information. Additionally, the jargon, abbreviations, and truncated character values throughout the data are obstacles to public access. The city should provide a data dictionary for all of these datasets to better satisfy the intent of the Open Data Law.

## Further Research

Potential expansions of this project for future researchers could include:

* Investigating salary trends for all titles, not just select titles across all agencies, and expanding the interactive app to include all title information.
* Investigating possible correlations between hiring and agency budgets, both in absolute budgets and budgets per capita, to see how agencies spend their money proportionally on personnel compared to other assets.
* Measuring the impact of scheduled exams and lists, as discussed above.
* Investigating possible correlations with prerequisite education and degrees for select titles, and determining if college degrees, and which specific fields of degrees, result in greater salaries. This is partially evident in the analyst and computer science titles discussed above, and is also evident for the technical science-oriented positions such as engineers.

In theory, New York City is entering the age of greater data transparency. In practice, the volume of raw data means that the public cannot easily access or synthesize the information. This leads to agencies and municipalities being able to hand-pick their preferred statistics and metrics to present a positive picture. Investigations into raw data using analytical tools are crucial for public knowledge and participation. Familiarity with city hiring data can result in a better understanding of the consequences of certain city programs and practices. Under the current mayoral administration, there has been an emphasis on social services, including a goal of decriminalizing small offenses and reducing the jail population. Violent crime has decreased in the past decade, and the NYPD has grown larger in headcount. Further analysis of hiring and budgets in the social services agencies may determine if these publicized initiatives are reflected in the reality of current operations.

# Sources

NYC Payroll Dataset

<https://data.cityofnewyork.us/City-Government/Civil-List/ye3c-m4ga>

NYC Titles Dataset

<https://data.cityofnewyork.us/City-Government/NYC-Civil-Service-Titles/nzjr-3966>

NYC Archive of Open Competitive Examinations

<http://www.nyc.gov/html/dcas/html/work/oldexams.shtml>

NYC Open Data Law

<https://www1.nyc.gov/site/doitt/initiatives/open-data-law.page>

Github Repository for This Project

<https://github.com/spsstudent15/capstone>

# Appendix

Appendix A: Code and Output

<http://rpubs.com/aapsps/capstone-appendix-a>

Appendix B: Code and Output

<http://rpubs.com/aapsps/capstone-appendix-b>

Appendix C: Code and Output

<http://rpubs.com/aapsps/capstone-appendix-c>