NYC report

Introduction/Objectives

At $3.1 trillion dollars, the United States has the largest external fiscal deficit in the world. In light of the need for budget cuts caused by the coronavirus, public institutions have increasingly needed to justify their budgets. Concurrently, the Black Lives Matter movement of Summer 2020 engendered a national racial and social reckoning, calling the racial neutrality and political impartiality of public institutions into question.

Given the pressures from the political right and the political left, it is germane to investigate if NYC public institutions are using their budgets efficiently and distributing resources equitably across all races and sexes. In this report, we analyze data from 2014-2020 to:

A. Ascertain if race or gender affects employee's total salary, probability of receiving a salary raise, and probability to leave their positions;
B. Ascertain if certain departments or boroughs are more likely to be understaffed;
C. Ascertain if certain NYC government agencies are more likely to take longer to respond to complaints and feedback from the public. Identifying these agencies could be useful to keep in mind when thinking about how to most efficiently make budget cuts.

Methods

We only analyze long-term employees, i.e. those whose pay basis is designated as "per annum" in contrast with "per day" or "per hour" etc.. While the dynamics of short term employment are interesting, they are less material to the objective of discerning if NYC departments are efficient and equitable since the time of their disengagement is fixed. While we acknowledge that employees ceasing employment is not necessarily a bad thing as employees are sometimes fired, retired, we normatively treat turnover as a feature that generally should be minimized to minimize disruptions at the department and borough level.

Because the 311 dataset is very large, we pulled two different weeks from the 311 dataset to analyze how complaints were dealt with. We chose the same week, the 6th of February to the 13th of February for 2017 and 2020 to minimize the influence of external factors such as the weather in complaint time response. To see if certain government agencies take longer to resolve cases than others, we used multinomial logistic regression and ordinary least squares regression models.

Our rationale was that a week would be a representative sample (590210) cases in the data, and I wanted to pick a week before the arrival of the coronavirus as I believe that lockdown measures probably affected the data in a way that will not last after the vaccine is implemented. I created a feature called case length to represent how long a case was left open for, which was the closing date of the case minus the starting date. I also created a variable using the Payroll dataset called total_borough_spending. This variable was calculated using the payroll dataset and gives the total amount of money spent for operations across all agencies in a borough.

Since the goal was to see if certain agencies were more or less efficient than others in resolving cases, case length was used as an outcome variable and agency was used as a predictor in the first model. If a certain agency takes a long time to resolve a case, that means it is probably spending more money or could be inefficient. Obviously certain types of cases take longer to resolve than others, but in this case our job is solely to identify interesting data points.

I also made a model that uses status as an outcome variable, and uses agency as a predictor. This model was made using multinomial logistic regression as status is a categorical variable and isn't binary. The idea was to see if certain agencies were more likely than others to have pending or unresolved cases. I also used case length as a predictor and the total borough spending variable from the payroll dataset. The rationale behind including this variable was to adjust for the cost of completing a case in a more remote area of the city. For example, NYPD might take longer to respond to a call in a remote area of Queens than in midtown Manhattan. I thought that using total dollars spent in a given borough across all agencies was justified because certain boroughs likely have higher costs associated with conducting operations there. If a certain agency spends a lot of time operating in a certain area, I wanted this to be accounted for in the model.

To predict gender, we use the gender package in R that predicts gender from historical data. For unisex names, the function chooses the gender with the higher number of males or females historically named.

To assign the likely race of employees, we pulled data on common American surnames from the 2010 US census. While the data is a bit old, we argue that the most popular surnames in the US have not changed in the past decade or so. Each surname has a percent of White, Black, Asian, Hispanic, American Indian, and two or more races. We ignore two or more races since people included in this category would also be in two or more of the other categories. For each surname, we count the highest category as the likely race.

Feature 1: We created a feature that measures the **percentage of overtime hours** worked as a percentage of total hours worked. We hypothesize that the higher this aggregate percentage per borough and per neighborhood, the more likely the associated borough or neighborhood has a labor shortage. We cross-check this expectation by analyzing if this aggregate percentage is correlated with complaint response time. This was useful to 'joining the two datasets'

Feature 2: We created a feature that measures the average yearly growth rate of an individual. This feature was created by looping through the data, subsetting it and then calculating the growth rate for individuals with multiple years of data. For individuals with single years only or with multiple values from the same year, values of NA were assigned. The intent was to use this feature to calculate if promotions in different boroughs and departments were related to race and gender. Additionally, this feature could be an effective measure of attrition by testing if there was an association between pay growth rate and an employee resigning.

Because of the difficulties of using this feature, we used the **hourly** pay and **duration** instead. We were only able to find ten employees in the dataset that had multiple years entered, and thus were not able to calculate growth rates for the majority of the employees. The hourly pay measures how much regular, OT, and other pay an employee has divided by the total number of hours worked whereas duration measures length of time with company. This was useful in predicting an employee's likelihood of leaving.

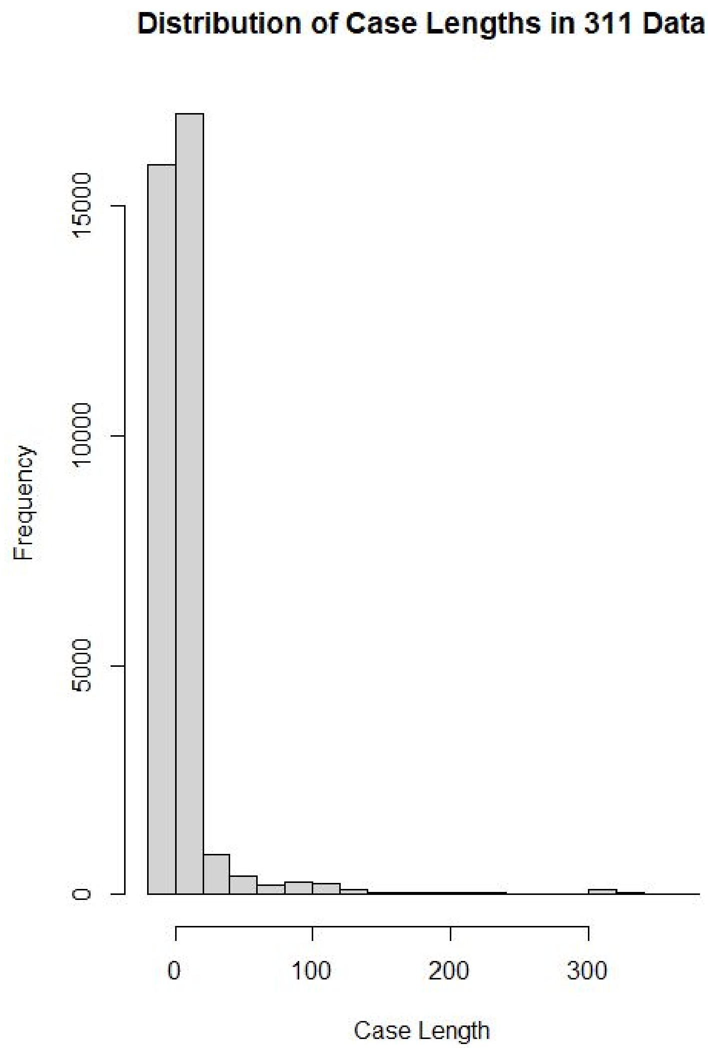
Feature 3: A case length feature was created to help measure government department efficiency in resolving cases. This feature was made by subtracting case end date from start date. This will be used as a predictor in some of the regression models.

<u>**Results**</u>

Are certain agencies more or less likely to resolve cases on time?

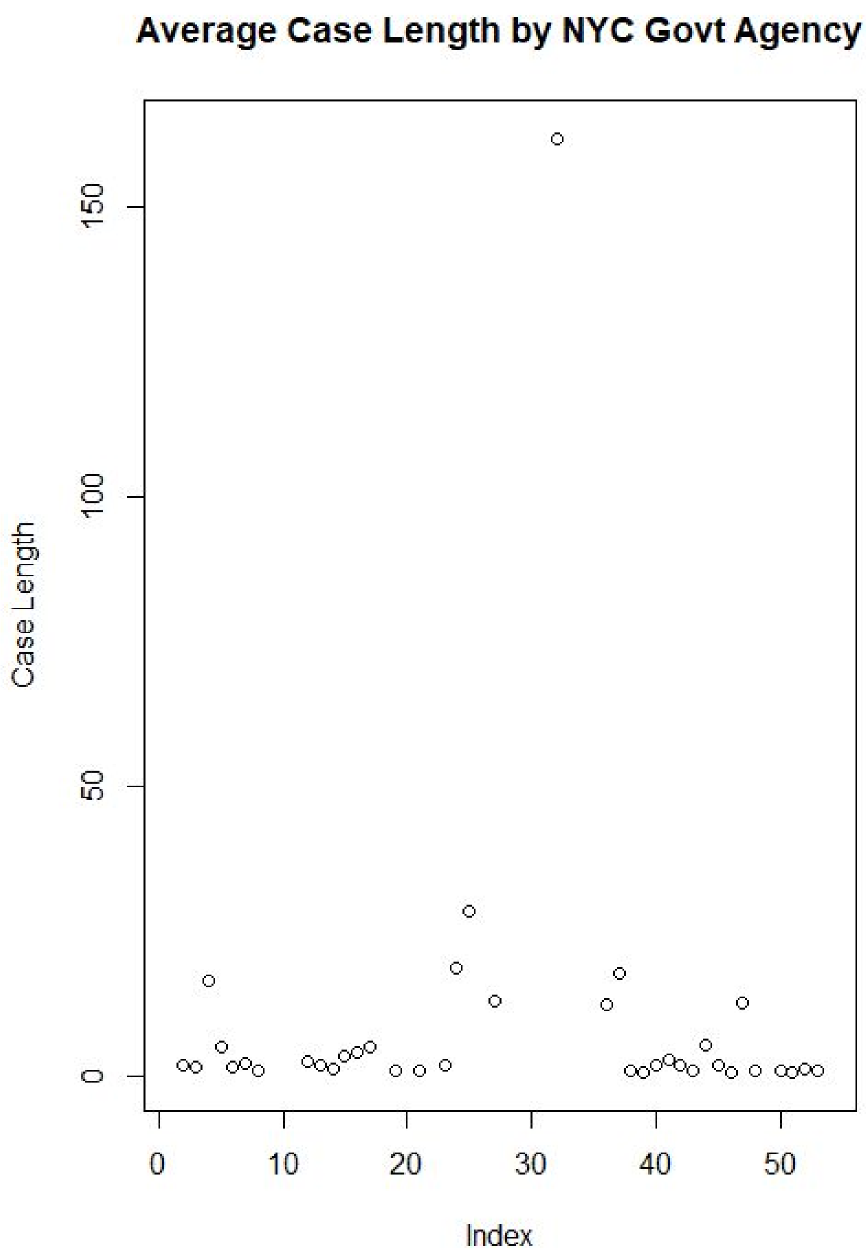
Within the dataset, most of the case lengths were between 1-4 days (across all agencies), however there were some outliers of over 300 days.

Figure 1: Distribution of Case Lengths in 311 Data



This next graph gives the average case length by agency (each agency was a separate index in the vector). It appears that there are three "levels" of agencies, ones that have average lengths of 0-1 days, those that are between 2-25 and 1 agency that has an average length of above 150:

Figure 2: Average Case Length by Agency



To measure if there was an association between case length and agency, I made a simple linear regression with case length as the outcome variable and agency as the predictor variable. Below are the outputs of the regression:

Table 1: Linear Regression Output

```
=====================================================
                         Dependent variable:
                      --------------------------
                            case_length
```

agencyDEP	-119.281***
	(2.112)
agencyDFTA	-95.849***
	(19.614)
agencyDHS	-122.781***
	(2.565)
agencyDOB	-95.702***
	(2.154)
agencyDOE	-110.907***
	(5.148)
agencyDOHMH	-116.647***
	(2.370)
agencyDOITT	57.151***
	(10.623)
agencyDOT	-114.720***
	(2.074)
agencyDPR	-77.380***
	(2.258)
agencyDSNY	-120.444***
	(2.140)
agencyEDC	-109.413***
	(3.413)
agencyHPD	-113.718***
	(2.063)
agencyNYPD	-122.240***
	(2.046)
agencyTLC	-70.413***
	(2.391)
Constant	122.849***
	(2.028)

Observations	28,065
R2	0.217
Adjusted R2	0.217
Residual Std. Error	27.590 (df = 28050)
F Statistic	555.876*** (df = 14; 28050)
Note:	p<0.1; **p<0.05; **p<0.01

RMSE: 0.07300499

Figure 1: Residual Plot of Linear Regression

Table 2: Multinomial Regression Output

	Dependent variable:			
	Closed	Open	Pending	Started
	(1)	(2)	(3)	(4)

caselength	0.090**	-1.137*	-3.064*	0.077**
	(0.000)	(0.000)	(0.000)	(0.000)
total_borough_spending	0.000**	-0.000	0.000*	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDEP	0.611**	-0.029*	-0.104*	0.072**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDFTA	0.00004**	-0.00000*	-0.00002*	-0.00000**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDHS	0.173**	-0.013*	-0.128*	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDOB	-0.906**	1.049*	-0.462*	-0.032**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDOE	0.032**	-0.004*	-0.023*	-0.0004**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDOHMH	0.461**	-0.0005*	-0.300*	-0.004**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDOITT	0.0001**	-0.00000*	-0.00001*	-0.00001**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDOT	0.086**	-0.772*	1.422*	-0.059**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDPR	0.279**	0.016*	-0.058*	-0.004**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyDSNY	0.384**	0.055*	0.099*	-0.030**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyEDC	0.047**	-0.006*	-0.037*	-0.0002**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyHPD	1.125**	0.053*	-0.945*	-0.017**
	(0.000)	(0.000)	(0.000)	(0.000)
agencyNYPD	2.082**	0.092*	-0.261*	-0.039**
	(0.000)	(0.000)	(0.000)	(0.000)

agencyTLC	0.022**	-0.002*	-0.015*	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	4.540**	0.435*	-0.881*	-0.121**
	(0.000)	(0.000)	(0.000)	(0.000)
Akaike Inf. Crit.	786.114	786.114	786.114	786.114

Figure 2: Predicted vs Actual Model Values After Test/Train Split

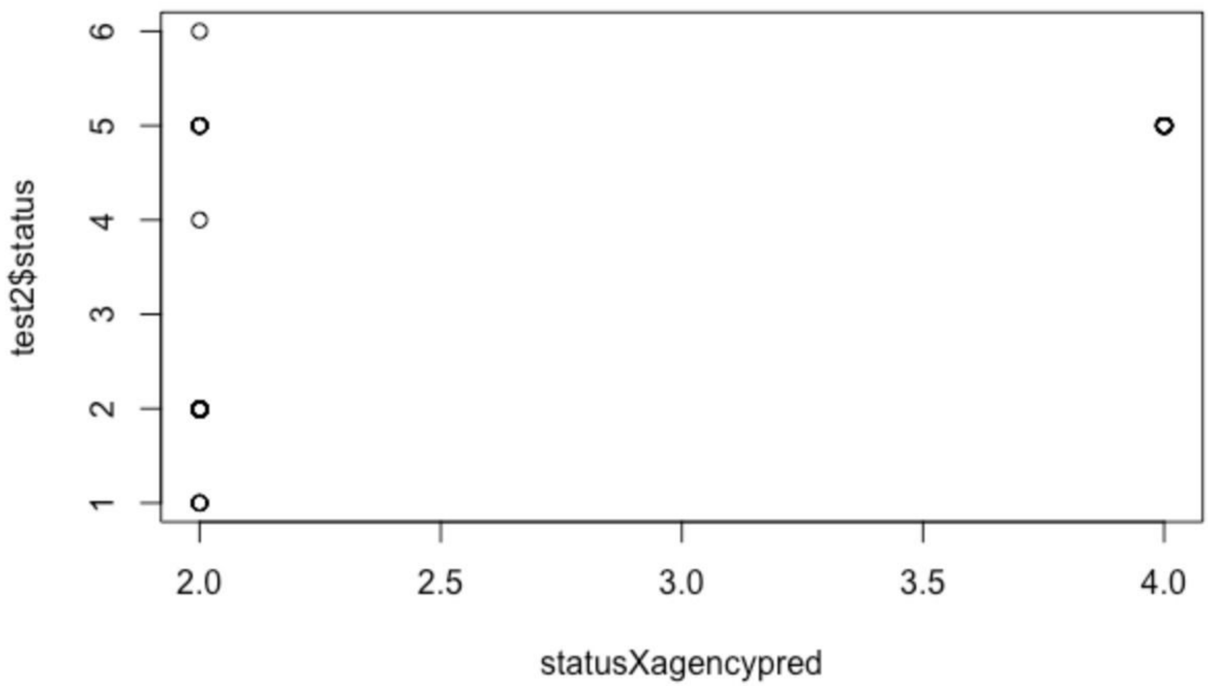
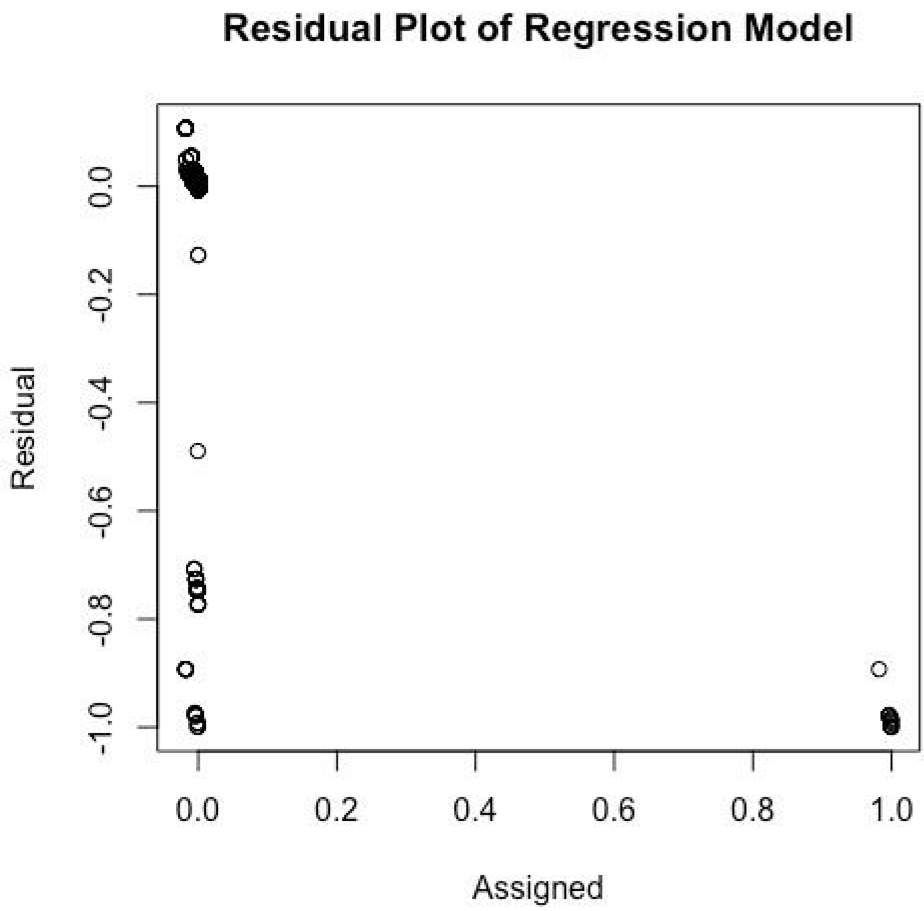
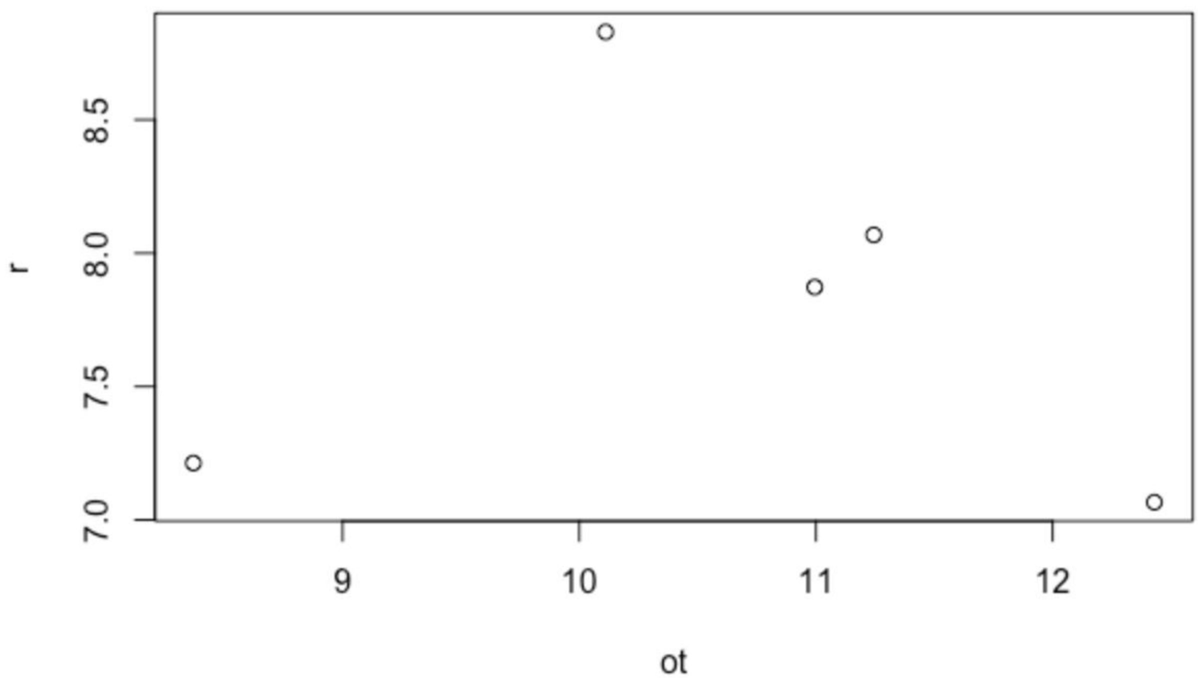


Figure 3: Residual Plot of Multinomial Regression



A. Are departments understaffed?

A1. Do different boroughs and neighborhoods have higher aggregate mean percentages of overtime hours? (None, simple data wrangling). How does this relate to response time? Simple geographic visualization.



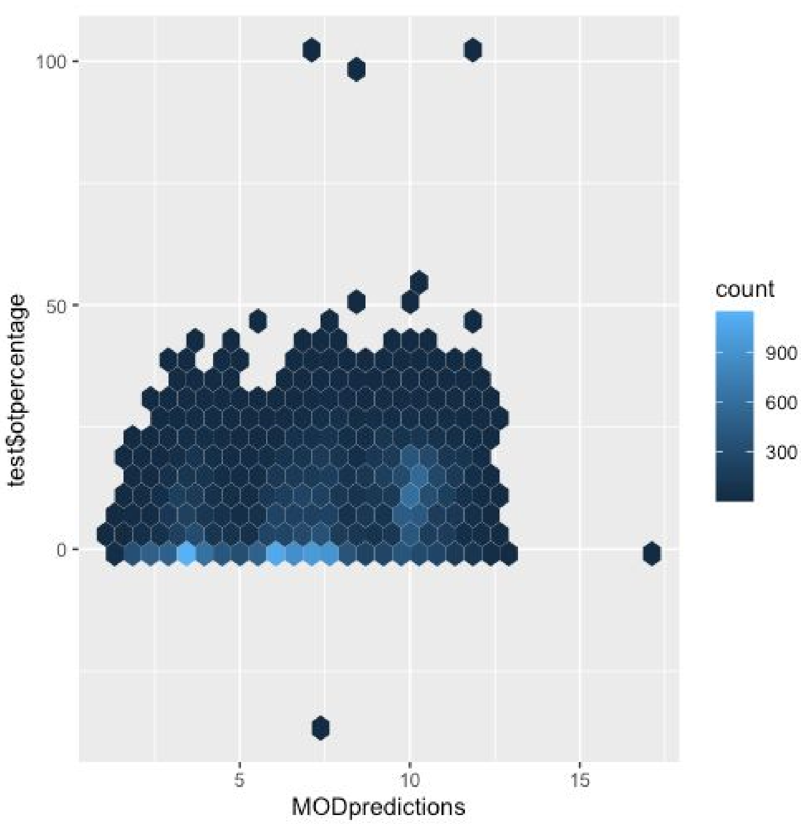
A2. Are there total salary, hourly rate, and percentage of pay from overtime differences across race and gender? We use titles as a control. (Multiple linear regression: the dependent variable is continuous.)

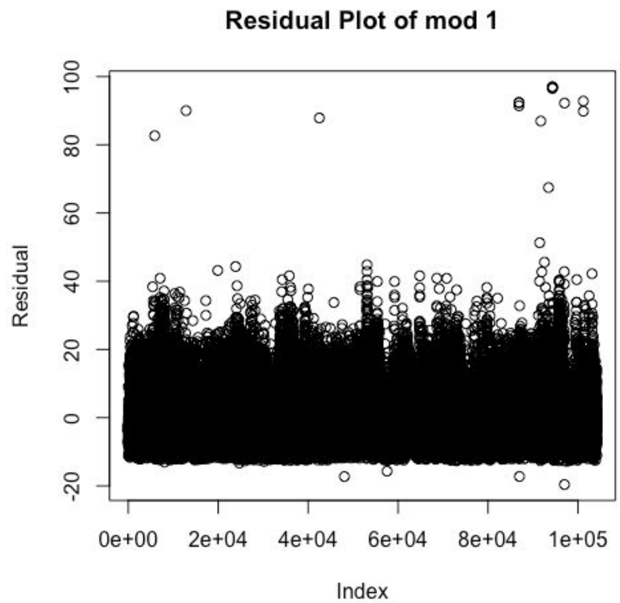
	Dependent variable:
	otpercentage
loc	0.023
	(0.024)
likelyrace	-0.004
	(0.017)
hourly	0.001***
	(0.0002)
duration	0.039***
	(0.003)
gender	-1.969***
	(0.024)
title	0.005***

	(0.0001)
Constant	3.143***
	(0.093)
Observations	104,173
R2	0.107
Adjusted R2	0.107
Residual Std. Error	7.787 (df = 104166)
F Statistic	2,081.374*** (df = 6; 104166)
Note:	*p<0.1; **p<0.05; ***p<0.01

RMSE MAE
7.784013 6.214628

Prediction error rate: 1.111114



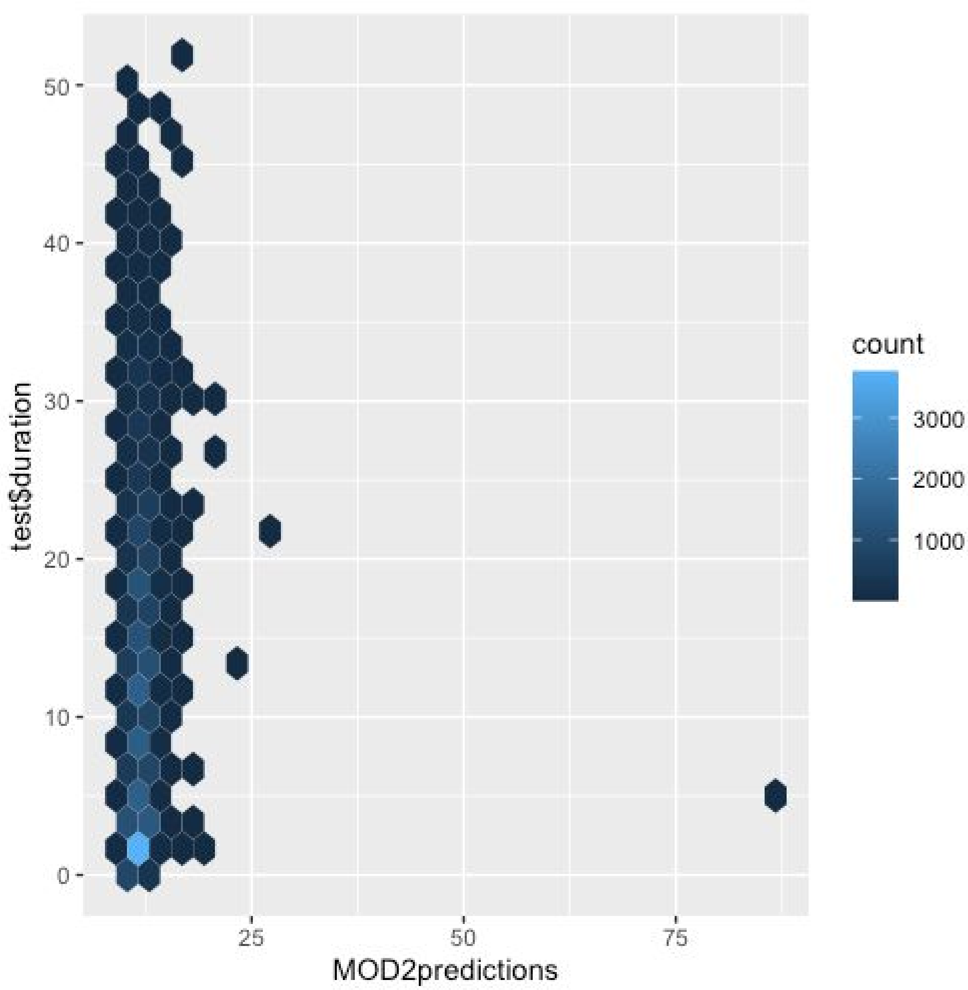


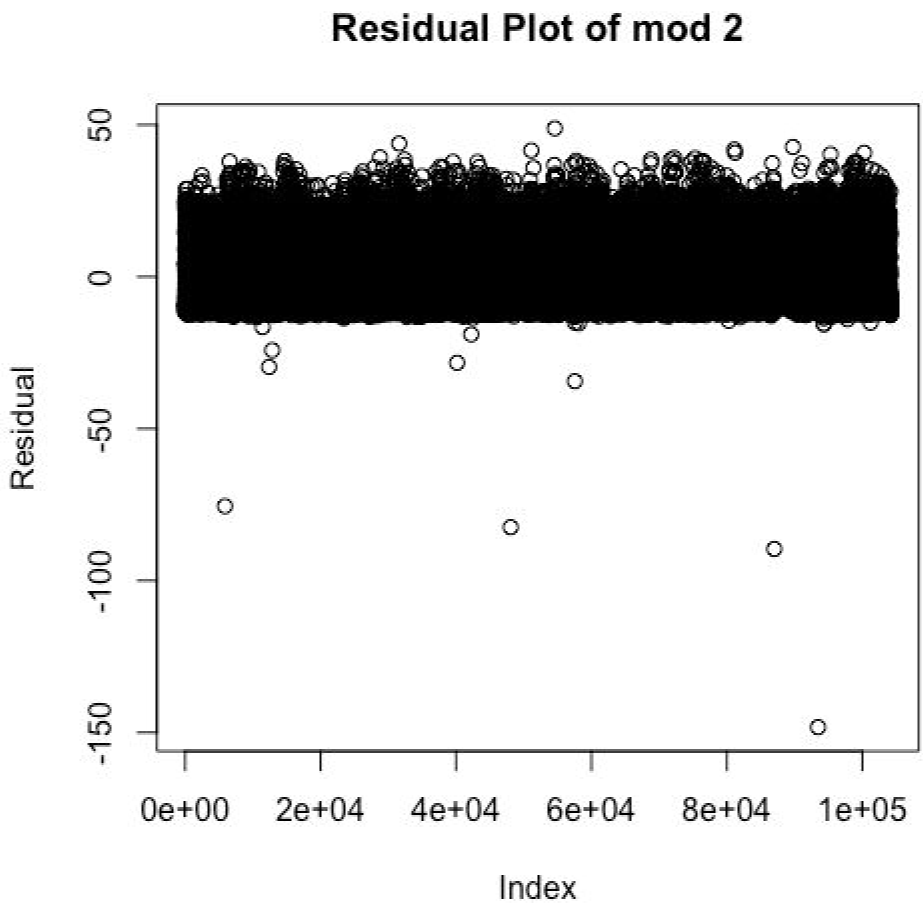
	Dependent variable: duration
loc	-0.085***
	(0.029)
likelyrace	0.318***
	(0.020)
hourly	0.008***
	(0.0003)
otpercentage	0.057***
	(0.004)
gender	0.199***
	(0.030)
title	0.001***
	(0.0001)
Constant	9.895***
	(0.109)

Observations	104,173
R2	0.012
Adjusted R2	0.012
Residual Std. Error	9.404 (df = 104166)
F Statistic	217.341*** (df = 6; 104166)
Note:	*p<0.1; **p<0.05; ***p<0.01

RMSE MAE
9.324185 7.67903

Prediction error rate: 0.8214627



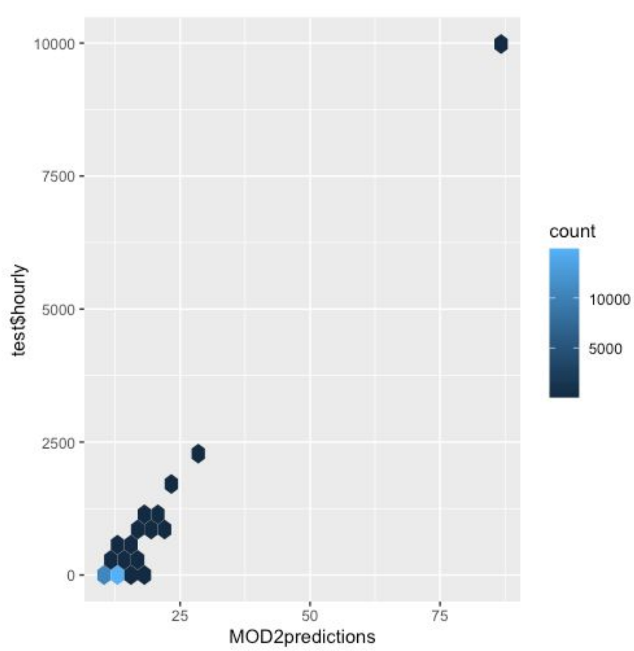


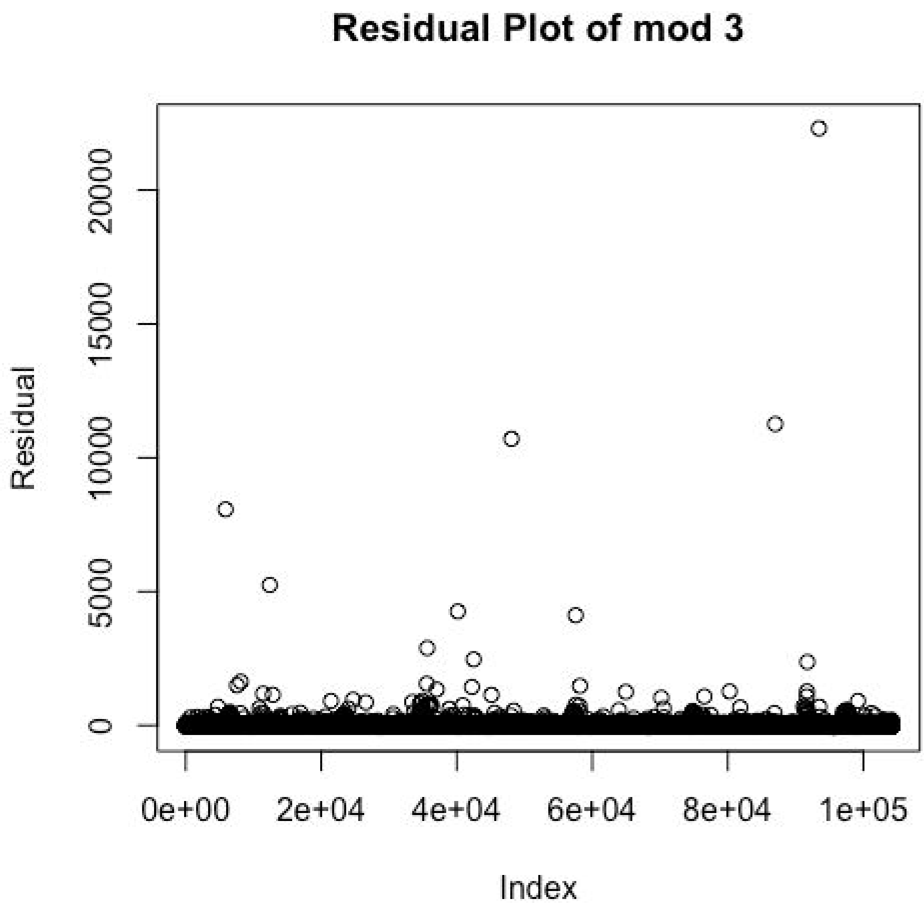
	Dependent variable:
	hourly
loc	0.988***
	(0.314)
likelyrace	1.111***
	(0.220)
duration	0.894***
	(0.033)

otpercentage	0.168***
	(0.040)
gender	-3.158***
	(0.328)
title	-0.009***
	(0.001)
Constant	33.003***
	(1.212)
Observations	104,173
R2	0.009
Adjusted R2	0.009
Residual Std. Error	101.372 (df = 104166)
F Statistic	161.579*** (df = 6; 104166)
Note:	*p<0.1; **p<0.05; ***p<0.01

RMSE MAE
74.25041 16.99334

Prediction error rate: 1.654946





B. Are departments providing optimal salary incentive structures to minimize worker turnover and maximize efficiency?

B1. After controlling for title, does race and gender influence an employee's salary and salary progression? (Multiple linear regression: the dependent variable is continuous.)

	Dependent variable:
	survived
hourly	-0.005***
	(0.0002)
loc	-0.084***
	(0.013)
duration	-0.004***
	(0.001)
race	-0.027***
	(0.009)
gender	-0.067***
	(0.012)
Constant	3.113***
	(0.040)
Observations	100,749
Log Likelihood	-25,321.650
Akaike Inf. Crit.	50,655.300
Note:	*p<0.1; **p<0.05; ***p<0.01

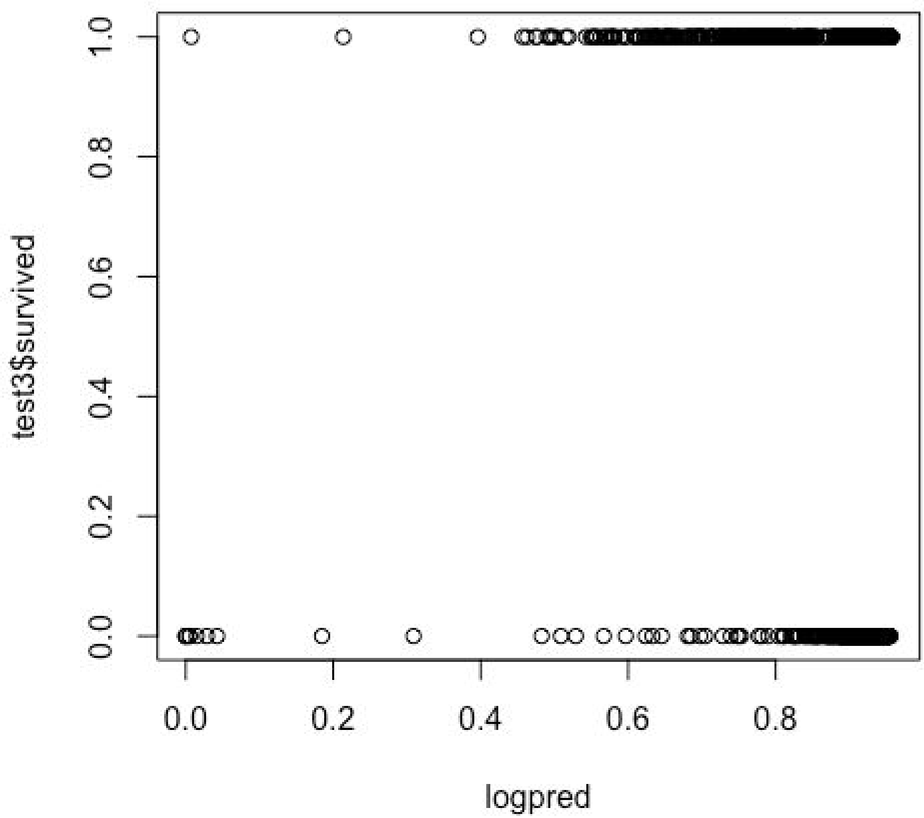
RMSE MAE
0.2558424 0.1300837

Prediction error rate= 0.274762

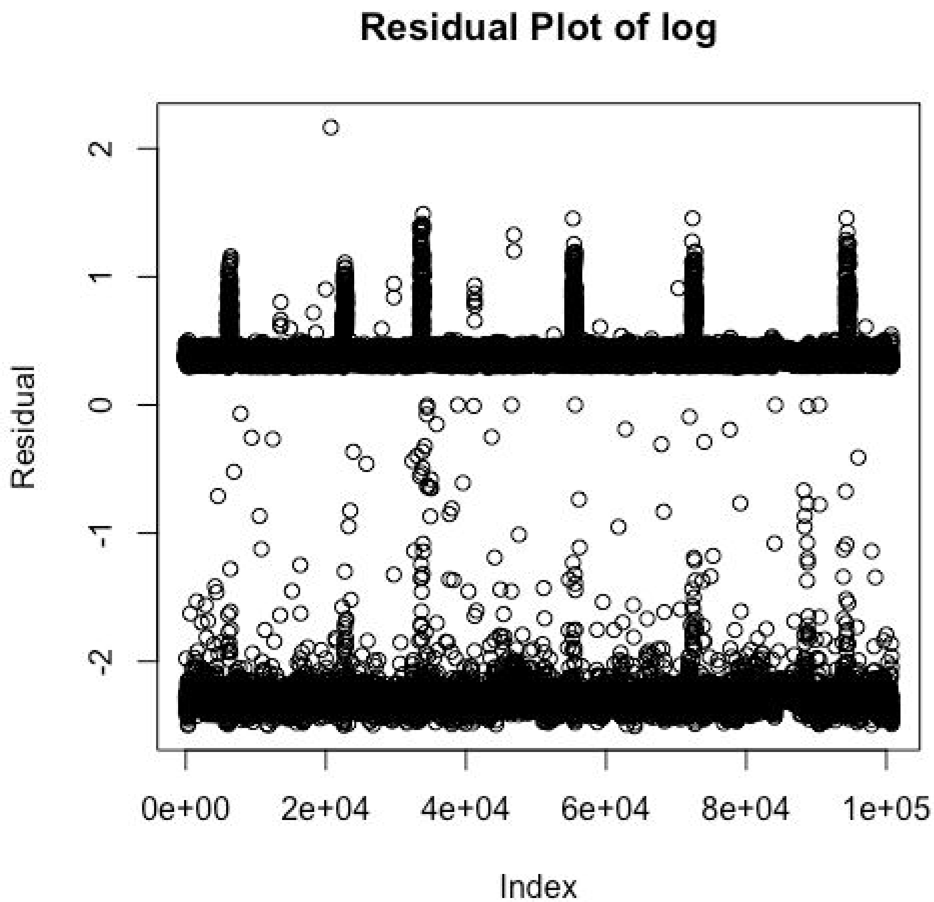
Classification Threshold = 0.5

	y=0	y=1
yhat=0	65	39
yhat=1	7073	93572

Percent Correctly Predicted = 92.94%
Percent Correctly Predicted = 0.9106%, for y = 0
Percent Correctly Predicted = 99.96% for y = 1
Null Model Correctly Predicts 92.92%
[1] 92.9408729 0.9106192 99.958338

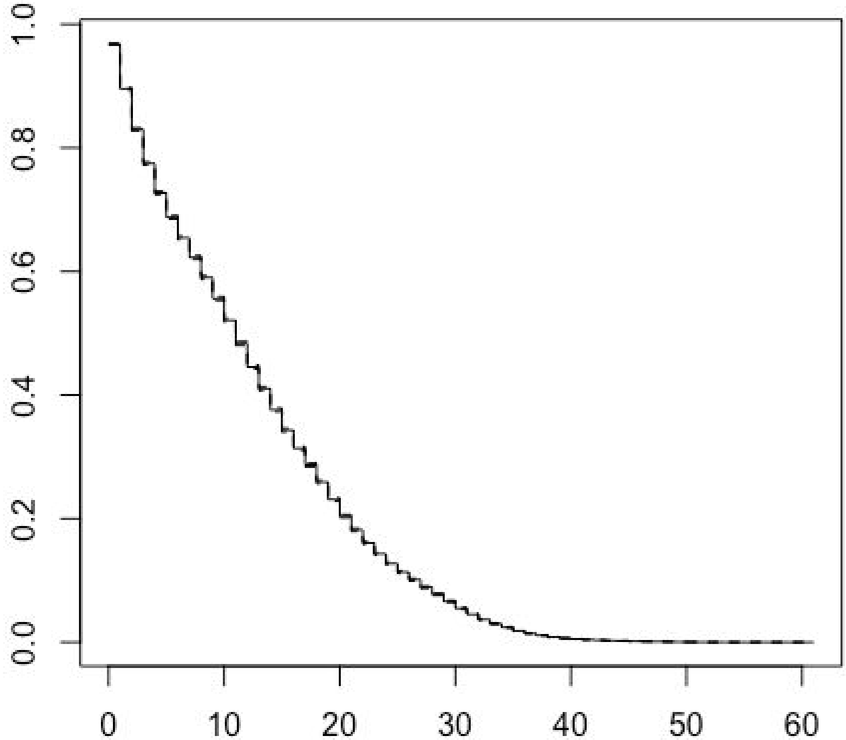


test3$survived
logpred
0.0
0.2
0.4
0.6
0.8
1.0

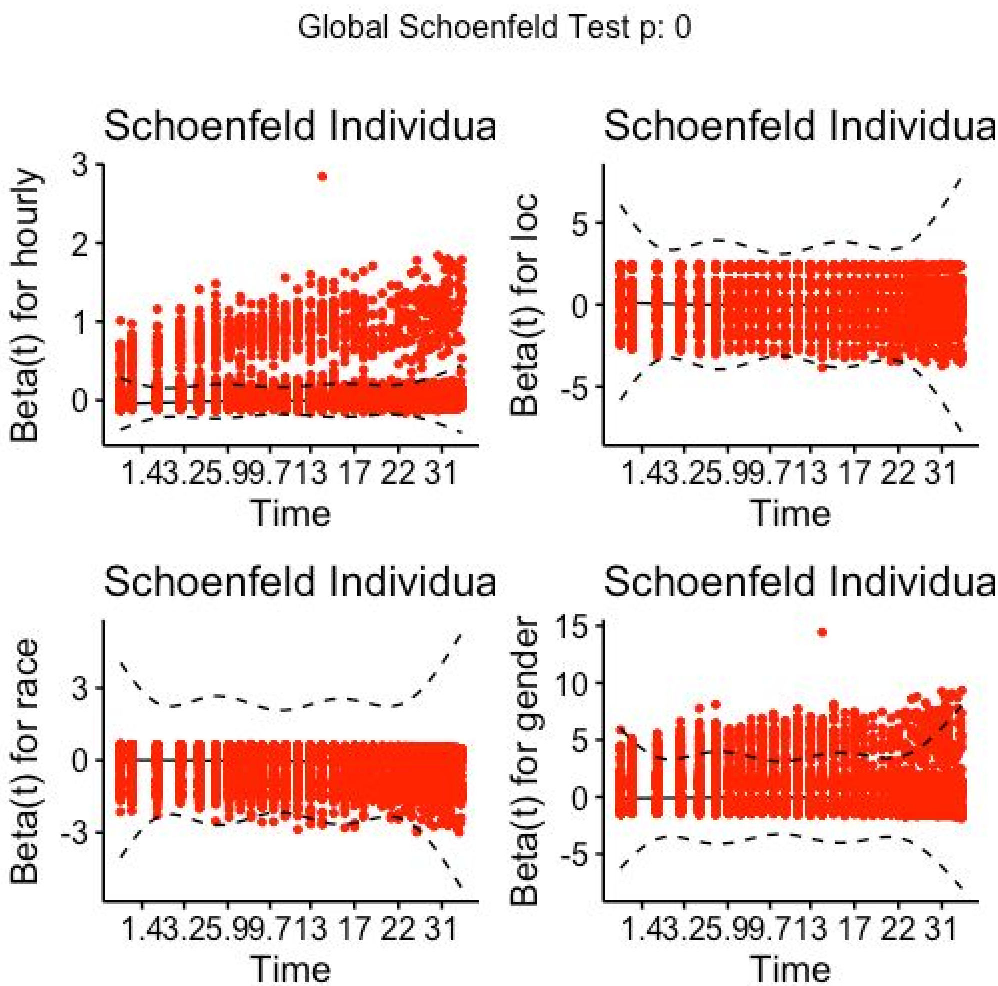


B2. How much does race, gender, department, hourly pay, duration of time with company, overtime, and salary progression affect an employee's probability of quitting? We call this model 'turnover risk'.
(We use a logit model and Cox proportional hazards model. We then compare the results of the three models after. We use these models because the dependent variable, quitting, is categorical: 'quit' or 'employed'. While the logit model calculates the probability that somewhat quits, the Cox model explores how the explanatory variables affect turnover rate. Hence, while the logit model is ideal for making predictions and calculating future turnover, the Cox model explains how NYC can deal with the explanatory variables to pre-empt quitting.)

	Dependent variable:
	duration
hourly	-0.010***
	(0.0002)
loc	0.015***
	(0.003)
race	-0.026***
	(0.002)
gender	-0.070***
	(0.003)
Observations	100,749
R^2	0.052
Max. Possible R^2	1.000
Log Likelihood	-983,607.700
Wald Test	3,315.560*** (df = 4)
LR Test	5,419.933*** (df = 4)
Score (Logrank) Test	622.898*** (df = 4)
Note:	*p<0.1; **p<0.05; ***p<0.01



1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60



Because the plots do not appear time independent, we deduced that the data contravenes the assumptions of the Cox proportional hazards model, and is thus not useful.

Relevant academic literature:

1. https://journals.sagepub.com/doi/full/10.1177/1078087418771224?casa_token=lMrJQS6O2NIAAAAA%3A-Tg9nEul2B0pPMkBNpgvq9naQOPsdeX5otpQ3THden-qUK5LmvXpr_XB9OEze_woprXBx6YCRzZs
    a. Gentrification is happening in NYC to blacks and Latinos
    b. As income rises -> % black in neighboorhood decreases, as % black increases -> rate of gentrification falls
    c. This is because they are poorer

    d. Implications: Our no discrimination NYC salary results are interesting, as black clearly aren't doing well financially in NYC
    e. Could be because city government doesn't make a high enough proportion of the work force to meaningfully affect gentrification
2. NYC gender wage gap:
    a. https://comptroller.nyc.gov/reports/gender-wage-gap/inside-the-gender-wage-gap/inside-the-gender-wage-gap-part-ii-earnings-of-latinas-in-new-york-city/
        i. Latina women make $0.49 for every $1.00 a white man makes in NYC
    b. https://comptroller.nyc.gov/reports/gender-wage-gap/inside-the-gender-wage-gap/inside-the-gender-wage-gap-part-i-earnings-of-black-women-in-new-york-city/
        i. Black women make $0.59 for every dollar a white man makes in NYC
    c. There IS a gender wage gap in NYC. This again goes against what we observed that there is no gender wage gap in NYC payroll data.
3. NYC gov't inefficiency
    a. Scalar dumping phenomenon: idea that government services burden pushed from feds to local governments which increases inefficiencies as local governments compete with each other on taxes so they finance services with debt:
        i. https://journals.sagepub.com/doi/full/10.1177/0308518X18796511
    b. https://journals.sagepub.com/doi/full/10.1177/0308518X19844794
    c.

Conclusion/Recommendations:

Efficiency

Our analysis does suggest that certain NYC government agencies take longer to resolve cases than others. Figure 2 demonstrates that there is one department (which happens to be the department of Telecommunications) that has an average case length of 161 days. Table 1 gives the output of the regression of agency onto case length. This model had R^2 of 0.217 for 28,065 observations, indicating that it does not have strong predictive accuracy. Additionally, all of the departments had coefficients from +/- 70-100, with the exception being the department of telecommunications that I mentioned above. All coefficients were statistically significant ($p < 0.01$), confirming my above assumption that the DOITT takes longer than the other 11 government departments to resolve cases.

Interestingly, our analysis does not seem to suggest that there is an association between NYC government agency and case resolution status. Identifying which agencies tend to not resolve their cases could be useful when designing future city budgets. Literature suggests that pragmatic municipalism can stem the effects of austerity measures when kept in mind while introducing budget cuts.[1] While all coefficients where statistically significant in the multinomial regression, the caselength variable was the only variable that had significant weight. Other notable variables include the Departments of Transportation and Housing Preservation & Development, however these variables were not as significant as case length, which had a coefficient of -3.064 ($p < 0.1$) for a status of pending. There appears to be no logic behind the case length coefficients however, as some are positive and some are negative. Our total borough spending variable had coefficients of 0, suggesting that borough does not have an effect on the resolution of a case. Overall, I do not think any useful conclusions can be drawn from this model.

Racial Equity

Gender Equity

Workload split (primary):
Payroll: Bernadette
311: Nikhil

[1] https://journals.sagepub.com/doi/full/10.1177/0308518X19844794