Effect of Base Salary on Overtime

Shaun Stearns

8/21/2019

# Load Necessary Libraries

library(psych)  
library(QuantPsyc)

## Loading required package: boot

##   
## Attaching package: 'boot'

## The following object is masked from 'package:psych':  
##   
## logit

## Loading required package: MASS

##   
## Attaching package: 'QuantPsyc'

## The following object is masked from 'package:base':  
##   
## norm

library(mixlm)

##   
## Attaching package: 'mixlm'

## The following object is masked from 'package:psych':  
##   
## fparse

## The following objects are masked from 'package:stats':  
##   
## glm, lm

library(ez)  
library(readr)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following object is masked from 'package:mixlm':  
##   
## tally

## The following object is masked from 'package:MASS':  
##   
## select

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(stats)  
library(DescTools)

##   
## Attaching package: 'DescTools'

## The following objects are masked from 'package:QuantPsyc':  
##   
## Kurt, Skew

## The following objects are masked from 'package:psych':  
##   
## AUC, ICC, SD

library(ggplot2)

##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

library(ggthemes)  
library(extrafont)

## Registering fonts with R

library(jtools)

##   
## Attaching package: 'jtools'

## The following object is masked from 'package:DescTools':  
##   
## %nin%

library(knitr)

# Set directory

#setwd("Desktop/CSU Global Data Analytics/MIS500/Portfolio Project")

# Load file and replace 0 with NA

NYC\_Fisc\_Clean <- read\_csv("NYC\_Fisc\_Clean.csv",   
 col\_types = cols(`Base Salary` = col\_number(),   
 `OT Hours` = col\_number(), `Regular Gross Paid` = col\_number(),   
 `Regular Hours` = col\_number(), `Total OT Paid` = col\_number(),   
 `Total Other Pay` = col\_number()),  
 na="0")

# Check File

### View(NYC\_Fisc\_Clean)

### Descriptives for Borough and Over Time Hours. Determine Boroughs with the highest sample.

describeBy(NYC\_Fisc\_Clean$`OT Hours`, NYC\_Fisc\_Clean$`Work Location Borough`)

##   
## Descriptive statistics by group   
## group: ALBANY  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 16 143.66 122.9 122 135.36 138.81 9.75 393.75 384 0.58  
## kurtosis se  
## X1 -0.93 30.72  
## --------------------------------------------------------   
## group: BRONX  
## vars n mean sd median trimmed mad min max range  
## X1 1 29091 244.99 200.85 212.25 223.33 210.13 -0.25 1738.25 1738.5  
## skew kurtosis se  
## X1 1.09 1.83 1.18  
## --------------------------------------------------------   
## group: BROOKLYN  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 53638 234.45 197.54 204.5 212.19 200.89 -8 3045 3053 1.47  
## kurtosis se  
## X1 5.2 0.85  
## --------------------------------------------------------   
## group: DELAWARE  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 129 111.85 110.07 85.5 94.06 91.37 1.2 504.22 503.02 1.49  
## kurtosis se  
## X1 2.16 9.69  
## --------------------------------------------------------   
## group: DUTCHESS  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 35 44.06 37.46 28.25 41.46 32.25 1 120.25 119.25 0.51  
## kurtosis se  
## X1 -1.26 6.33  
## --------------------------------------------------------   
## group: GREENE  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 15 140.5 83.82 136.75 138 104.89 33.25 280.25 247 0.27  
## kurtosis se  
## X1 -1.41 21.64  
## --------------------------------------------------------   
## group: MANHATTAN  
## vars n mean sd median trimmed mad min max range  
## X1 1 82716 217.58 208.16 167.08 190.99 213.24 -6.58 2330.5 2337.08  
## skew kurtosis se  
## X1 1.34 3.1 0.72  
## --------------------------------------------------------   
## group: NASSAU  
## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 35 25.67 45.44 6.5 15.61 8.9 0.25 219 218.75 2.68 7.55  
## se  
## X1 7.68  
## --------------------------------------------------------   
## group: ORANGE  
## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 2 597.5 160.51 597.5 597.5 168.28 484 711 227 0 -2.75  
## se  
## X1 113.5  
## --------------------------------------------------------   
## group: PUTNAM  
## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 58 54.66 68.38 26.62 44.07 37.44 1 343 342 2.05 5.17  
## se  
## X1 8.98  
## --------------------------------------------------------   
## group: QUEENS  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 62367 274.58 245.91 221.5 240.91 236.1 -3 2017.5 2020.5 1.32  
## kurtosis se  
## X1 2.24 0.98  
## --------------------------------------------------------   
## group: RICHMOND  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 9610 269.26 223.3 232.5 243.44 222.61 0.25 1488 1487.75 1.26  
## kurtosis se  
## X1 2.41 2.28  
## --------------------------------------------------------   
## group: SCHOHARIE  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 41 83.41 112.66 36 59.55 47.44 1 495.37 494.37 1.96  
## kurtosis se  
## X1 3.61 17.6  
## --------------------------------------------------------   
## group: SULLIVAN  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 146 82.01 96.26 51.15 63.85 59.34 1 519.75 518.75 2.21  
## kurtosis se  
## X1 5.49 7.97  
## --------------------------------------------------------   
## group: ULSTER  
## vars n mean sd median trimmed mad min max range skew kurtosis  
## X1 1 255 69.55 98.95 27 47.9 34.1 0.5 584 583.5 2.38 6.55  
## se  
## X1 6.2  
## --------------------------------------------------------   
## group: WASHINGTON DC  
## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 1 9 NA 9 9 0 9 9 0 NA NA NA  
## --------------------------------------------------------   
## group: WESTCHESTER  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 625 141.23 170.65 83 109.44 103.78 0.2 1115.05 1114.85 2.38  
## kurtosis se  
## X1 7.66 6.83

# Remove Rows with NA Values for Over Time Hours, Base Salary, and Total Other Pay

NYC\_Fisc\_Clean<-NYC\_Fisc\_Clean[complete.cases(NYC\_Fisc\_Clean$`OT Hours`),]  
NYC\_Fisc\_Clean<-NYC\_Fisc\_Clean[complete.cases(NYC\_Fisc\_Clean$`Base Salary`),]  
NYC\_Fisc\_Clean<-NYC\_Fisc\_Clean[complete.cases(NYC\_Fisc\_Clean$`Total Other Pay`),]

# Split file based on Borough Location

NYC\_Split <- split(NYC\_Fisc\_Clean, NYC\_Fisc\_Clean$`Work Location Borough`)

# Create datasets based on Borough location with the largest sample sizes

BRONX<-NYC\_Split[[2]]  
BROOKLYN<-NYC\_Split[[3]]  
MANHATTAN<-NYC\_Split[[7]]  
QUEENS<-NYC\_Split[[11]]  
RICHMOND<-NYC\_Split[[12]]  
WESTCHESTER<-NYC\_Split[[16]]

#Combine Split files into a single file with chosen Boroughs  
BigBoroughs <- rbind(BRONX, BROOKLYN, MANHATTAN, QUEENS, RICHMOND, WESTCHESTER)  
#Overtime Hours  
describeBy(BigBoroughs$`OT Hours`, BigBoroughs$`Work Location Borough`)

##   
## Descriptive statistics by group   
## group: BRONX  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 27404 256.75 199.66 226.66 236.68 202.98 0.25 1738.25 1738 1.07  
## kurtosis se  
## X1 1.88 1.21  
## --------------------------------------------------------   
## group: BROOKLYN  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 50204 244.54 196.39 218 223.85 196.07 -8 3045 3053 1.45  
## kurtosis se  
## X1 5.3 0.88  
## --------------------------------------------------------   
## group: MANHATTAN  
## vars n mean sd median trimmed mad min max range  
## X1 1 76500 230.06 208.53 186.75 205.28 221.65 -6.58 2330.5 2337.08  
## skew kurtosis se  
## X1 1.29 3.08 0.75  
## --------------------------------------------------------   
## group: QUEENS  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 59077 282.89 244.74 232.25 250.66 233.14 -3 2017.5 2020.5 1.3  
## kurtosis se  
## X1 2.22 1.01  
## --------------------------------------------------------   
## group: RICHMOND  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 9110 278.12 222.76 242 253.21 218.66 0.25 1488 1487.75 1.26  
## kurtosis se  
## X1 2.46 2.33  
## --------------------------------------------------------   
## group: WESTCHESTER  
## vars n mean sd median trimmed mad min max range skew  
## X1 1 589 145.62 173.12 88.08 113.7 106.87 0.2 1115.05 1114.85 2.35  
## kurtosis se  
## X1 7.41 7.13

#Base Salary  
describeBy(BigBoroughs$`Base Salary`, BigBoroughs$`Work Location Borough`)

##   
## Descriptive statistics by group   
## group: BRONX  
## vars n mean sd median trimmed mad min max  
## X1 1 27404 57917.46 26284.22 54341 58724.78 30422.95 9.48 152534  
## range skew kurtosis se  
## X1 152524.5 -0.22 -0.08 158.78  
## --------------------------------------------------------   
## group: BROOKLYN  
## vars n mean sd median trimmed mad min max range  
## X1 1 50204 56624.16 27212 53270 57646.35 27995.94 9.48 193836 193826.5  
## skew kurtosis se  
## X1 -0.19 -0.05 121.45  
## --------------------------------------------------------   
## group: MANHATTAN  
## vars n mean sd median trimmed mad min max  
## X1 1 76500 60570.42 28892.25 58208 61173.5 29382.17 9.21 193836  
## range skew kurtosis se  
## X1 193826.8 -0.15 -0.23 104.46  
## --------------------------------------------------------   
## group: QUEENS  
## vars n mean sd median trimmed mad min max  
## X1 1 59077 58482.34 28050.65 59589 60382.53 27334.7 9.21 192301  
## range skew kurtosis se  
## X1 192291.8 -0.45 -0.1 115.41  
## --------------------------------------------------------   
## group: RICHMOND  
## vars n mean sd median trimmed mad min max  
## X1 1 9110 57601.64 29411.68 60592 59124.76 25847.65 9.62 133546  
## range skew kurtosis se  
## X1 133536.4 -0.5 -0.23 308.15  
## --------------------------------------------------------   
## group: WESTCHESTER  
## vars n mean sd median trimmed mad min max  
## X1 1 589 60652.64 21025.74 59497 61346.05 13389.36 16.77 130000  
## range skew kurtosis se  
## X1 129983.2 -0.59 2.15 866.35

#Total Other Pay  
describeBy(BigBoroughs$`Total Other Pay`, BigBoroughs$`Work Location Borough`)

##   
## Descriptive statistics by group   
## group: BRONX  
## vars n mean sd median trimmed mad min max  
## X1 1 27404 7857.9 6914.93 6156.74 7371.59 7132.49 -67261.56 45695.32  
## range skew kurtosis se  
## X1 112956.9 0.16 2.62 41.77  
## --------------------------------------------------------   
## group: BROOKLYN  
## vars n mean sd median trimmed mad min max  
## X1 1 50204 7717.75 6873.23 5976.65 7201.14 6816.17 -60664.75 49820.5  
## range skew kurtosis se  
## X1 110485.2 0.28 1.96 30.68  
## --------------------------------------------------------   
## group: MANHATTAN  
## vars n mean sd median trimmed mad min max  
## X1 1 76500 7355.39 7012.18 5136.36 6770.71 6291.15 -87992.37 49370.86  
## range skew kurtosis se  
## X1 137363.2 0.25 3.33 25.35  
## --------------------------------------------------------   
## group: QUEENS  
## vars n mean sd median trimmed mad min max  
## X1 1 59077 7831.71 6889.83 5957.91 7381.91 6867.14 -95896.13 68742.42  
## range skew kurtosis se  
## X1 164638.5 -0.01 5.14 28.35  
## --------------------------------------------------------   
## group: RICHMOND  
## vars n mean sd median trimmed mad min max  
## X1 1 9110 8371.35 7610.59 6374.04 7888.83 7278.71 -75865.03 52076.38  
## range skew kurtosis se  
## X1 127941.4 -0.02 3.49 79.74  
## --------------------------------------------------------   
## group: WESTCHESTER  
## vars n mean sd median trimmed mad min max range  
## X1 1 589 4357.8 3732.87 3668.85 3913.34 2749.2 8.25 28543.23 28534.98  
## skew kurtosis se  
## X1 2.87 13.47 153.81

# Create Non-transformed dataset for visualizations

BigBoroughs2 <- BigBoroughs  
#View(BigBuoroughs2)

# Log Transform to normalize data

BigBoroughs$`OT Hours` = log10(BigBoroughs$`OT Hours`)

## Warning: NaNs produced

BigBoroughs$`Base Salary` = log10(BigBoroughs$`Base Salary`)  
BigBoroughs$`Total Other Pay` = log10(BigBoroughs$`Total Other Pay`)

## Warning: NaNs produced

View(BigBoroughs)

# Analysis of Variance with Unadjusted and adjusted post hoc tests

anova1 <- aov(BigBoroughs$`OT Hours` ~ BigBoroughs$`Work Location Borough`)  
summary(anova1, digits = 4)

## Df Sum Sq Mean Sq F value Pr(>F)  
## BigBoroughs$`Work Location Borough` 5 1022 204.39 499.7 <2e-16  
## Residuals 222873 91162 0.41   
##   
## BigBoroughs$`Work Location Borough` \*\*\*  
## Residuals   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## 5 observations deleted due to missingness

#Post hoc unadjusted: Similar to a t-test but takes into account overall error  
Post<-TukeyHSD(anova1)  
Post

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = BigBoroughs$`OT Hours` ~ BigBoroughs$`Work Location Borough`)  
##   
## $`BigBoroughs$`Work Location Borough``  
## diff lwr upr p adj  
## BROOKLYN-BRONX -0.029436156 -0.043124712 -0.01574760 0.0000000  
## MANHATTAN-BRONX -0.141063539 -0.153894494 -0.12823259 0.0000000  
## QUEENS-BRONX 0.005521566 -0.007799087 0.01884222 0.8460090  
## RICHMOND-BRONX 0.028014754 0.005973182 0.05005633 0.0039660  
## WESTCHESTER-BRONX -0.376831299 -0.452730835 -0.30093176 0.0000000  
## MANHATTAN-BROOKLYN -0.111627384 -0.122095756 -0.10115901 0.0000000  
## QUEENS-BROOKLYN 0.034957721 0.023894574 0.04602087 0.0000000  
## RICHMOND-BROOKLYN 0.057450909 0.036695565 0.07820625 0.0000000  
## WESTCHESTER-BROOKLYN -0.347395143 -0.422931181 -0.27185911 0.0000000  
## QUEENS-MANHATTAN 0.146585105 0.136602619 0.15656759 0.0000000  
## RICHMOND-MANHATTAN 0.169078293 0.148878265 0.18927832 0.0000000  
## WESTCHESTER-MANHATTAN -0.235767760 -0.311153102 -0.16038242 0.0000000  
## RICHMOND-QUEENS 0.022493188 0.001978619 0.04300776 0.0219939  
## WESTCHESTER-QUEENS -0.382352865 -0.457823098 -0.30688263 0.0000000  
## WESTCHESTER-RICHMOND -0.404846053 -0.482332486 -0.32735962 0.0000000

#Scheffe Post Hoc: Statistical adjustment based on multiple comparisons  
PostSchf <- ScheffeTest(anova1, conf.level=NA)  
PostSchf

##   
## Posthoc multiple comparisons of means : Scheffe Test   
##   
## $`BigBoroughs$`Work Location Borough``  
## BRONX BROOKLYN MANHATTAN QUEENS RICHMOND  
## BROOKLYN 4.6e-07 - - - -   
## MANHATTAN < 2e-16 < 2e-16 - - -   
## QUEENS 0.925 5.0e-16 < 2e-16 - -   
## RICHMOND 0.022 4.2e-12 < 2e-16 0.082 -   
## WESTCHESTER < 2e-16 < 2e-16 1.1e-15 < 2e-16 < 2e-16

anova2 <- aov(BigBoroughs$`Base Salary` ~ BigBoroughs$`Work Location Borough`)  
summary(anova2)

## Df Sum Sq Mean Sq F value Pr(>F)  
## BigBoroughs$`Work Location Borough` 5 215 43.01 64.12 <2e-16  
## Residuals 222878 149500 0.67   
##   
## BigBoroughs$`Work Location Borough` \*\*\*  
## Residuals   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Post hoc unadjusted: Similar to a t-test but takes into account overall error  
Post<-TukeyHSD(anova2)  
Post

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = BigBoroughs$`Base Salary` ~ BigBoroughs$`Work Location Borough`)  
##   
## $`BigBoroughs$`Work Location Borough``  
## diff lwr upr p adj  
## BROOKLYN-BRONX -0.045242184 -0.062771466 -0.027712902 0.0000000  
## MANHATTAN-BRONX -0.005547433 -0.021978488 0.010883622 0.9298283  
## QUEENS-BRONX -0.056536049 -0.073594169 -0.039477929 0.0000000  
## RICHMOND-BRONX -0.124860783 -0.153086862 -0.096634704 0.0000000  
## WESTCHESTER-BRONX 0.101740339 0.004544626 0.198936052 0.0339672  
## MANHATTAN-BROOKLYN 0.039694751 0.026289274 0.053100229 0.0000000  
## QUEENS-BROOKLYN -0.011293865 -0.025460965 0.002873236 0.2056418  
## RICHMOND-BROOKLYN -0.079618599 -0.106197516 -0.053039683 0.0000000  
## WESTCHESTER-BROOKLYN 0.146982523 0.050252311 0.243712735 0.0002158  
## QUEENS-MANHATTAN -0.050988616 -0.063771831 -0.038205401 0.0000000  
## RICHMOND-MANHATTAN -0.119313351 -0.145181143 -0.093445559 0.0000000  
## WESTCHESTER-MANHATTAN 0.107287772 0.010750537 0.203825007 0.0192141  
## RICHMOND-QUEENS -0.068324735 -0.094595299 -0.042054170 0.0000000  
## WESTCHESTER-QUEENS 0.158276388 0.061630448 0.254922327 0.0000449  
## WESTCHESTER-RICHMOND 0.226601122 0.127373255 0.325828990 0.0000000

#Scheffe Post Hoc: Statistical adjustment based on multiple comparisons  
PostSchf <- ScheffeTest(anova2, conf.level=NA)  
PostSchf

##   
## Posthoc multiple comparisons of means : Scheffe Test   
##   
## $`BigBoroughs$`Work Location Borough``  
## BRONX BROOKLYN MANHATTAN QUEENS RICHMOND  
## BROOKLYN 2.0e-10 - - - -   
## MANHATTAN 0.96834 5.8e-14 - - -   
## QUEENS < 2e-16 0.39657 < 2e-16 - -   
## RICHMOND < 2e-16 2.6e-14 < 2e-16 1.4e-10 -   
## WESTCHESTER 0.11321 0.00214 0.07439 0.00058 5.0e-08

anova3 <- aov(BigBoroughs$`Total Other Pay` ~ BigBoroughs$`Work Location Borough`)  
summary(anova3)

## Df Sum Sq Mean Sq F value Pr(>F)  
## BigBoroughs$`Work Location Borough` 5 649 129.72 315.2 <2e-16  
## Residuals 220172 90617 0.41   
##   
## BigBoroughs$`Work Location Borough` \*\*\*  
## Residuals   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## 2706 observations deleted due to missingness

#Post hoc unadjusted: Similar to a t-test but takes into account overall error  
Post<-TukeyHSD(anova3)  
Post

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = BigBoroughs$`Total Other Pay` ~ BigBoroughs$`Work Location Borough`)  
##   
## $`BigBoroughs$`Work Location Borough``  
## diff lwr upr p adj  
## BROOKLYN-BRONX -0.014827169 -0.028648387 -0.00100595 0.0271338  
## MANHATTAN-BRONX -0.107465330 -0.120412592 -0.09451807 0.0000000  
## QUEENS-BRONX 0.009961739 -0.003486547 0.02341003 0.2814623  
## RICHMOND-BRONX 0.048576979 0.026266359 0.07088760 0.0000000  
## WESTCHESTER-BRONX -0.207193132 -0.283338121 -0.13104814 0.0000000  
## MANHATTAN-BROOKLYN -0.092638161 -0.103202324 -0.08207400 0.0000000  
## QUEENS-BROOKLYN 0.024788908 0.013616337 0.03596148 0.0000000  
## RICHMOND-BROOKLYN 0.063404147 0.042386686 0.08442161 0.0000000  
## WESTCHESTER-BROOKLYN -0.192365963 -0.268142142 -0.11658978 0.0000000  
## QUEENS-MANHATTAN 0.117427069 0.107355730 0.12749841 0.0000000  
## RICHMOND-MANHATTAN 0.156042308 0.135588965 0.17649565 0.0000000  
## WESTCHESTER-MANHATTAN -0.099727802 -0.175349458 -0.02410615 0.0023634  
## RICHMOND-QUEENS 0.038615239 0.017841120 0.05938936 0.0000018  
## WESTCHESTER-QUEENS -0.217154871 -0.292863917 -0.14144583 0.0000000  
## WESTCHESTER-RICHMOND -0.255770110 -0.333543930 -0.17799629 0.0000000

#Scheffe Post Hoc: Statistical adjustment based on multiple comparisons  
PostSchf <- ScheffeTest(anova3, conf.level=NA)  
PostSchf

##   
## Posthoc multiple comparisons of means : Scheffe Test   
##   
## $`BigBoroughs$`Work Location Borough``  
## BRONX BROOKLYN MANHATTAN QUEENS RICHMOND  
## BROOKLYN 0.096 - - - -   
## MANHATTAN < 2e-16 < 2e-16 - - -   
## QUEENS 0.486 1.5e-07 < 2e-16 - -   
## RICHMOND 3.0e-07 1.6e-14 < 2e-16 3.5e-05 -   
## WESTCHESTER 1.1e-11 4.6e-10 0.015 4.7e-13 < 2e-16

# Multiple regression on split files to determine what variables predict increased overtime

#BRONX  
BRONXlm<-lm(BRONX$`OT Hours` ~ BRONX$`Base Salary`)  
BRONXmlm<-lm(BRONX$`OT Hours` ~ BRONX$`Base Salary` +  
 BRONX$`Total Other Pay`, data = BRONX)  
summ(BRONXlm, digits = 4)

## MODEL INFO:  
## Observations: 27404  
## Dependent Variable: BRONX$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(1,27402) = 1519.5527, p = 0.0000  
## R² = 0.0525  
## Adj. R² = 0.0525   
##   
## Standard errors: OLS  
## ----------------------------------------------------------------  
## Est. S.E. t val. p  
## ------------------------- ---------- -------- --------- --------  
## (Intercept) 155.9119 2.8409 54.8821 0.0000  
## BRONX$`Base Salary 0.0017 0.0000 38.9814 0.0000  
## ----------------------------------------------------------------

lm.beta(BRONXlm)

## BRONX$`Base Salary`   
## 0.2292171

summ(BRONXmlm, digits = 4)

## MODEL INFO:  
## Observations: 27404  
## Dependent Variable: BRONX$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(2,27401) = 1018.1230, p = 0.0000  
## R² = 0.0692  
## Adj. R² = 0.0691   
##   
## Standard errors: OLS  
## --------------------------------------------------------------------  
## Est. S.E. t val. p  
## ----------------------------- ---------- -------- --------- --------  
## (Intercept) 164.0026 2.8395 57.7576 0.0000  
## BRONX$`Base Salary 0.0010 0.0001 16.7731 0.0000  
## BRONX$`Total Other Pay 0.0048 0.0002 22.1269 0.0000  
## --------------------------------------------------------------------

lm.beta(BRONXmlm)

## BRONX$`Base Salary` BRONX$`Total Other Pay`   
## 0.1254687 0.1655164

#BROOKLYN  
BROOKLYNlm<-lm(BROOKLYN$`OT Hours` ~ BROOKLYN$`Base Salary`)  
BROOKLYNmlm<-lm(BROOKLYN$`OT Hours` ~ BROOKLYN$`Base Salary` +  
 BROOKLYN$`Total Other Pay`, data = BROOKLYN)  
summ(BROOKLYNlm, digits = 4)

## MODEL INFO:  
## Observations: 50204  
## Dependent Variable: BROOKLYN$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(1,50202) = 1902.4565, p = 0.0000  
## R² = 0.0365  
## Adj. R² = 0.0365   
##   
## Standard errors: OLS  
## -------------------------------------------------------------------  
## Est. S.E. t val. p  
## ---------------------------- ---------- -------- --------- --------  
## (Intercept) 166.4577 1.9863 83.8048 0.0000  
## BROOKLYN$`Base Salary 0.0014 0.0000 43.6172 0.0000  
## -------------------------------------------------------------------

lm.beta(BROOKLYNlm)

## BROOKLYN$`Base Salary`   
## 0.1910821

summ(BROOKLYNmlm, digits = 4)

## MODEL INFO:  
## Observations: 50204  
## Dependent Variable: BROOKLYN$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(2,50201) = 1766.0091, p = 0.0000  
## R² = 0.0657  
## Adj. R² = 0.0657   
##   
## Standard errors: OLS  
## -----------------------------------------------------------------------  
## Est. S.E. t val. p  
## -------------------------------- ---------- -------- --------- --------  
## (Intercept) 166.0503 1.9560 84.8949 0.0000  
## BROOKLYN$`Base Salary 0.0006 0.0000 16.2433 0.0000  
## BROOKLYN$`Total Other Pay 0.0058 0.0001 39.6245 0.0000  
## -----------------------------------------------------------------------

lm.beta(BROOKLYNmlm)

## BROOKLYN$`Base Salary` BROOKLYN$`Total Other Pay`   
## 0.08292237 0.20228420

#MANHATTAN  
MANHATTANlm<-lm(MANHATTAN$`OT Hours` ~ MANHATTAN$`Base Salary`)  
MANHATTANmlm<-lm(MANHATTAN$`OT Hours` ~ MANHATTAN$`Base Salary` +  
 MANHATTAN$`Total Other Pay`, data = MANHATTAN)  
summ(MANHATTANlm, digits = 4)

## MODEL INFO:  
## Observations: 76500  
## Dependent Variable: MANHATTAN$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(1,76498) = 3356.8430, p = 0.0000  
## R² = 0.0420  
## Adj. R² = 0.0420   
##   
## Standard errors: OLS  
## --------------------------------------------------------------------  
## Est. S.E. t val. p  
## ----------------------------- ---------- -------- --------- --------  
## (Intercept) 140.4281 1.7140 81.9308 0.0000  
## MANHATTAN$`Base Salary 0.0015 0.0000 57.9383 0.0000  
## --------------------------------------------------------------------

lm.beta(MANHATTANlm)

## MANHATTAN$`Base Salary`   
## 0.2050288

summ(MANHATTANmlm, digits = 4)

## MODEL INFO:  
## Observations: 76500  
## Dependent Variable: MANHATTAN$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(2,76497) = 5291.6720, p = 0.0000  
## R² = 0.1215  
## Adj. R² = 0.1215   
##   
## Standard errors: OLS  
## --------------------------------------------------------------------  
## Est. S.E. t val. p  
## ----------------------------- ---------- -------- --------- --------  
## (Intercept) 151.1049 1.6463 91.7821 0.0000  
## MANHATTAN$`Base Salary 0.0001 0.0000 2.1063 0.0352  
## MANHATTAN$`Total Other 0.0102 0.0001 83.2031 0.0000  
## Pay   
## --------------------------------------------------------------------

lm.beta(MANHATTANmlm)

## MANHATTAN$`Base Salary` MANHATTAN$`Total Other Pay`   
## 0.008697635 0.343576065

#QUEENS  
QUEENSlm<-lm(QUEENS$`OT Hours` ~ QUEENS$`Base Salary`)  
QUEENSmlm<-lm(QUEENS$`OT Hours` ~ QUEENS$`Base Salary` +  
 QUEENS$`Total Other Pay`, data = QUEENS)  
summ(QUEENSlm, digits = 4)

## MODEL INFO:  
## Observations: 59077  
## Dependent Variable: QUEENS$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(1,59075) = 2018.0069, p = 0.0000  
## R² = 0.0330  
## Adj. R² = 0.0330   
##   
## Standard errors: OLS  
## -----------------------------------------------------------------  
## Est. S.E. t val. p  
## -------------------------- ---------- -------- --------- --------  
## (Intercept) 190.1529 2.2895 83.0527 0.0000  
## QUEENS$`Base Salary 0.0016 0.0000 44.9222 0.0000  
## -----------------------------------------------------------------

lm.beta(QUEENSlm)

## QUEENS$`Base Salary`   
## 0.1817463

summ(QUEENSmlm, digits = 4)

## MODEL INFO:  
## Observations: 59077  
## Dependent Variable: QUEENS$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(2,59074) = 2191.1164, p = 0.0000  
## R² = 0.0691  
## Adj. R² = 0.0690   
##   
## Standard errors: OLS  
## ---------------------------------------------------------------------  
## Est. S.E. t val. p  
## ------------------------------ ---------- -------- --------- --------  
## (Intercept) 184.2528 2.2499 81.8940 0.0000  
## QUEENS$`Base Salary 0.0006 0.0000 16.3204 0.0000  
## QUEENS$`Total Other Pay 0.0077 0.0002 47.8139 0.0000  
## ---------------------------------------------------------------------

lm.beta(QUEENSmlm)

## QUEENS$`Base Salary` QUEENS$`Total Other Pay`   
## 0.07442694 0.21804785

#RICHMOND  
RICHMONDlm<-lm(RICHMOND$`OT Hours` ~ RICHMOND$`Base Salary`)  
RICHMONDmlm<-lm(RICHMOND$`OT Hours` ~ RICHMOND$`Base Salary` +  
 RICHMOND$`Total Other Pay`, data = RICHMOND)  
summ(RICHMONDlm, digits = 4)

## MODEL INFO:  
## Observations: 9110  
## Dependent Variable: RICHMOND$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(1,9108) = 286.4758, p = 0.0000  
## R² = 0.0305  
## Adj. R² = 0.0304   
##   
## Standard errors: OLS  
## -------------------------------------------------------------------  
## Est. S.E. t val. p  
## ---------------------------- ---------- -------- --------- --------  
## (Intercept) 201.9357 5.0539 39.9564 0.0000  
## RICHMOND$`Base Salary 0.0013 0.0001 16.9256 0.0000  
## -------------------------------------------------------------------

lm.beta(RICHMONDlm)

## RICHMOND$`Base Salary`   
## 0.1746255

summ(RICHMONDmlm, digits = 4)

## MODEL INFO:  
## Observations: 9110  
## Dependent Variable: RICHMOND$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(2,9107) = 155.5899, p = 0.0000  
## R² = 0.0330  
## Adj. R² = 0.0328   
##   
## Standard errors: OLS  
## -----------------------------------------------------------------------  
## Est. S.E. t val. p  
## -------------------------------- ---------- -------- --------- --------  
## (Intercept) 199.4093 5.0738 39.3015 0.0000  
## RICHMOND$`Base Salary 0.0011 0.0001 12.7985 0.0000  
## RICHMOND$`Total Other Pay 0.0017 0.0003 4.8971 0.0000  
## -----------------------------------------------------------------------

lm.beta(RICHMONDmlm)

## RICHMOND$`Base Salary` RICHMOND$`Total Other Pay`   
## 0.14850298 0.05682121

#WESTCHESTER  
WESTCHESTERlm<-lm(WESTCHESTER$`OT Hours` ~ WESTCHESTER$`Base Salary`)  
WESTCHESTERmlm<-lm(WESTCHESTER$`OT Hours` ~ WESTCHESTER$`Base Salary` +  
 WESTCHESTER$`Total Other Pay`, data = WESTCHESTER)  
summ(WESTCHESTERlm, digits = 4)

## MODEL INFO:  
## Observations: 589  
## Dependent Variable: WESTCHESTER$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(1,587) = 0.4143, p = 0.5200  
## R² = 0.0007  
## Adj. R² = -0.0010   
##   
## Standard errors: OLS  
## -----------------------------------------------------------------------  
## Est. S.E. t val. p  
## ------------------------------- ---------- --------- --------- --------  
## (Intercept) 158.8790 21.8056 7.2862 0.0000  
## WESTCHESTER$`Base Salary -0.0002 0.0003 -0.6437 0.5200  
## -----------------------------------------------------------------------

lm.beta(WESTCHESTERlm)

## WESTCHESTER$`Base Salary`   
## -0.02655871

summ(WESTCHESTERmlm,digits = 4)

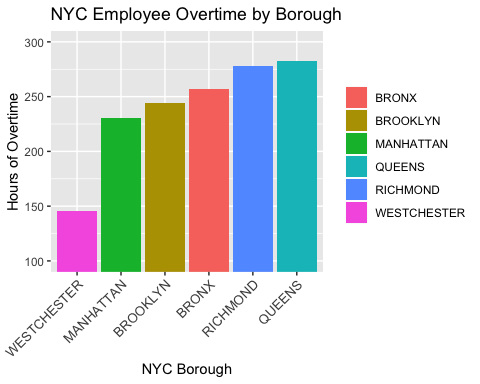
## MODEL INFO:  
## Observations: 589  
## Dependent Variable: WESTCHESTER$`OT Hours`  
## Type: OLS linear regression   
##   
## MODEL FIT:  
## F(2,586) = 3.9694, p = 0.0194  
## R² = 0.0134  
## Adj. R² = 0.0100   
##   
## Standard errors: OLS  
## -----------------------------------------------------------------------  
## Est. S.E. t val. p  
## ------------------------------- ---------- --------- --------- --------  
## (Intercept) 128.6142 24.3324 5.2857 0.0000  
## WESTCHESTER$`Base Salary -0.0001 0.0003 -0.2870 0.7742  
## WESTCHESTER$`Total Other 0.0053 0.0019 2.7422 0.0063  
## Pay   
## -----------------------------------------------------------------------

lm.beta(WESTCHESTERmlm)

## WESTCHESTER$`Base Salary` WESTCHESTER$`Total Other Pay`   
## -0.01187709 0.11347485

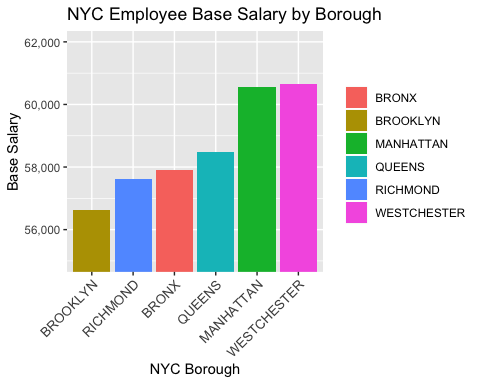
# Plots: Bar Chart (Overtime by Boroughs)

A <- ggplot(data=BigBoroughs2, aes(x=reorder(BigBoroughs2$`Work Location Borough`, BigBoroughs2$`OT Hours`),  
 y=BigBoroughs2$`OT Hours`, fill=`Work Location Borough`)) +  
 geom\_bar(stat="summary", fun.y = "mean") +   
 xlab("NYC Borough") +   
 ylab("Hours of Overtime") +  
 coord\_cartesian(ylim = c(100, 300)) +  
 ggtitle("NYC Employee Overtime by Borough")  
A + scale\_y\_continuous(labels = scales::comma) +   
 theme(axis.text.x =  
 element\_text(size = 10, angle = 45, hjust = 1, vjust = 1)) +  
 guides(fill=guide\_legend(title=NULL))



# Plots: Bar Chart (Basesalary by Borough)

B <- ggplot(data=BigBoroughs2, aes(x=reorder(BigBoroughs2$`Work Location Borough`, BigBoroughs2$`Base Salary`),   
 y=BigBoroughs2$`Base Salary`, fill=`Work Location Borough`)) +  
 geom\_bar(stat="summary", fun.y = "mean", aes(fill=`Work Location Borough`)) +  
 xlab("NYC Borough") +   
 ylab("Base Salary") +   
 coord\_cartesian(ylim = c(55000, 62000)) +  
 ggtitle("NYC Employee Base Salary by Borough")  
B + scale\_y\_continuous(labels = scales::comma) +  
 theme(axis.text.x =  
 element\_text(size = 10, angle = 45, hjust = 1, vjust = 1)) +  
 guides(fill=guide\_legend(title=NULL))



# Plots: Bar Chart (Other Pay by Borough)

C <- ggplot(data=BigBoroughs2, aes(x=reorder(BigBoroughs2$`Work Location Borough`, BigBoroughs2$`Total Other Pay`),   
 y=BigBoroughs2$`Total Other Pay`, fill=`Work Location Borough`)) +  
 geom\_bar(stat="summary", fun.y = "mean", aes(fill=`Work Location Borough`)) +  
 xlab("NYC Borough") +   
 ylab("Total Other Pay") +   
 coord\_cartesian(ylim = c(4000, 8300)) +  
 ggtitle("NYC Employee Total Other Pay by Borough")  
C + scale\_y\_continuous(labels = scales::comma) +   
 theme(axis.text.x =  
 element\_text(size = 10, angle = 45, hjust = 1, vjust = 1)) +  
 guides(fill=guide\_legend(title=NULL))

