STAT 488 Project 1: Current Employees of the City of Chicago

Charles Hwang

4/3/2022

The dataset I chose for this project is the "Current Employee Names, Salaries, and Position Titles" dataset from the Chicago Data Portal. I chose this dataset because I am more confident that this data is more accurate than an average dataset, due to it being public data and thus potentially being required to be reported (e.g., through FOIA requests). This dataset is updated every six months (with the last update on January 21, 2022) and contains eight variables: the (1) name, (2) job title, (3) department, (4-5) statuses (full- vs. part-time and salary vs. hourly), (6-7) hours per week and hourly wage (if hourly), and (8) annual salary (if salaried) of every current employee working for the City of Chicago.

I first shortened the variable names, coded the status variables (FTPT and S) as factors, and estimated annual salary for hourly workers by multiplying 50 by hours per week (HpW) by hourly wage (DpH), where 50 is the number of workweeks in a calendar year when excluding city holidays[1]. A summary of the variables is in **Figure 1**. There are 249 employee names that appear more than once (**Figure 2**), which could be from an employee holding multiple positions or multiple employees having the same first and last name. There are up to 1078 unique[2] jobs (**Figure 3**), with "Police Officer" being the most common (8876) and a strong statistical outlier. There are 36 departments (**Figure 4**) which match those listed on the city's website (https://www.chicago.gov/city/en/depts.html).

I created a two-way table to compare statuses (**Figure 5**) and as expected, there are no part-time salaried employees. I also created tables for hours per week (**Figure 6**) and saw that there are only four levels: 40, 35, 20, and 10. A large majority of hourly employees work 40 hours per week. I also created boxplots to visualize annual salaries (**Figure 7**) and we can see that salaried employees generally make more than hourly employees per year and full-time employees generally make more than part-time employees per year, as expected. The boxplots for salaried employees and full-time employees are very similar because most full-time employees are salaried (as seen in **Figure 6**).

I created tables (**Figure 8**) for hourly employees and interestingly, there are 238 employees listed as being paid below minimum wage (15 DpH[3] in Chicago). There is one employee each with hourly wages of 12 and 14 DpH, which could be from when the minimum wage was lower or an error with data entry. There are 236 employees listed as being paid 3 DpH, which could be a placeholder for interns, unpaid positions with stipends, volunteers, etc. The boxplots (**Figure 9**) comparing hourly wages by status and workload show all 236 are working 20 hours per week. We can also see that full-time employees are generally paid more than part-time employees, as expected.

With salary (Sal) as the response variable, I ran one GLM with the two factor variables (FTPT and S) as explanatory variables and another with the two quantitative variables (HpW and DpH) as explanatory variables (**Figure 10**). The intercept and both variables were statistically significant in both models. We can see that part-time employees made approximately 69343.47 per year less than full-time employees and salaried employees made approximately 13678.66 per year more than hourly employees (both while fixing the other variable). In the other model, salary increased by approximately 1426.20 per year for every additional hour worked per week when fixing hourly wage.

[1] According to the city's website (https://www.chicago.gov/city/en/narr/misc/city-holidays.html), there are 12 holidays on which offices are closed. Since there are approximately 52 weeks in a year, subtracting 12 days results in *roughly* 50 weeks.

[2] I say "up to" because there may be jobs that are entered under different names, causing additional levels to be created.

[3] Unfortunately, RMarkdown does not allow me to (easily) type dollar signs as it interprets them as the start of a LaTeX equation, but the units for the hourly wage (DpH) variable are all in United States dollars per hour.

Q-Q plots (**Figure 11**) showed that none of the three quantitative variables are normally distributed, which is expected with the right skew seen in the boxplots. Subsequent `sqrt` and `log` transformations (not shown) on affected variables are insufficient for normalizing the data. After printing a scatterplot matrix and histograms (**Figure 12**) for the three variables using the `mvn()` function, I conducted a Box-Cox transformation and produced a chi-square Q-Q plot which showed the data were still not normal, but perhaps closer to being normal than before with there being only two visual outliers with unduly large Mahalanobis distances.

I also calculated the means and variance matrix of hours per week (HpW) and hourly wages (DpH) for hourly employees (excluding Sal because it's a proxy variable). Even though the normality assumption is clearly violated, I decided to compute Hotelling's T^2 (**Figure 13**) to see what the results would have been and to conduct some multivariate analysis. Since there are so many values in the data (6814), any deviation from the mean in any variable leads to very high values for T^2 and strong evidence against nearly any nontrivial matrix $\mu_0 = \begin{bmatrix} x_{HpW} \\ x_{DpH} \end{bmatrix}$ used for H_0. The corresponding F-distribution[4] is $\frac{(n-1)p}{n-p} F_{p,n-p} \approx 2.000294 F_{2,6810}$, and the critical value at $\alpha = 0.05$ is only 5.99498. Because of this, I decided to instead see what values for each individual variable would lead to rejection. We can see that an 11 minute 44 second (or greater) increase in mean weekly workload *or* a 31-cent (or greater) increase in mean hourly wage is significant at the $\alpha = 0.05$ level. Evidently, if we were to calculate differences using both variables, even less deviation from the mean would be needed to be statistically significant.

I attempted to do PCA on the same data (**Figure 14**), but since there were only two variables it produced limited results (something I discuss in the conclusion). It was able to produce variance proportions and graphing parameters, but the scree plot was soundly underwhelming. So in an 11th-hour bid to do some meaningful multivariate analysis, I leaned on my knowledge from Predictive Analytics and grew a decision tree (**Figure 15**) and random forest (**Figure 16**) to predict salary (Sal) on the data for hourly employees. It was hard to interpret anything useful from the decision tree due to the Name variable confounding the results, but we can see the error accounted for from each variable in the complexity parameter graph and the table of predictions. We can also see the cross-validation error is relatively low at 0.2615813.

The plot of log-transformed error shows the random forest was properly fit after around 200 trees. We can also see from the variable importance plot that and hourly wages (DpH) is the most important variable, as expected. This is followed by workload (HpW), status (S), and department (Dept) which makes sense intuitively. Different departments in the City of Chicago have different pay grades due to public pensions, labor unions, staffing/hiring levels, etc.

Job, Name, and status (which is trivial because this subset is entirely hourly employees) round out the explanatory variables in the random forest. The importance of the Job variable has a similar reasoning as department (Dept), but in future analysis I would be interested to see if the Name variable has more importance or influence. Specifically, seeing if certain "sounding" names imply a certain race or gender which may lead to inherent bias in hiring practices (and thus salary). There was no Race or Sex variable so this would likely require a deep-dive into the Name variable or using some other prediction/machine learning function to subset it. This is one of the few things I would explore in future analysis if given the time.

Another thing I would do if I had additional time would be to clean the Job variable. It was mentioned in footnote 2 that there may be duplicates and we saw already (**Figure 3**) that there are two mentions each of police officers and firefighter (EMT) in different roles just in the ten most common positions. Reclassifying entries listed in positions with only a few employees to more common positions or categorizing positions by profession/occupation, type of work, etc. could also be helpful. This seems to be easier than the deep-dive into the Name variable but still a challenge with 1078 levels.

In hindsight, better review and selection of data and better understanding of the project would have produced more thorough and robust analysis. At the time of choosing this dataset, it did not occur to me that there were only three quantitative variables and that they would all have missing values (due to employees either being salaried or hourly). Given the history of Chicago public affairs, it seems this dataset may be more used for looking up details on specific employees (i.e., politicians, lobbyists, etc.) rather than any statistial analysis. The normality assumption was also badly violated (**Figures 11-12**) as the quantitative variables

[4] We can see that $(n\text{-}1)p/(n\text{-}p)$ converges to p for large n, as is the case here.

were all right-skew, so some analyses may not be valid. It seemed sufficient after reviewing the sample size and variables and seeing it was the most viewed and downloaded dataset in the Chicago Data Portal, but I did not do the best job of vetting the dataset for the analysis needed. However, I believe the project was a success overall given the short timeline in applying multivariate analyses to real-world data.

Appendix

Figure 1

```
rm(list=ls())
cdp<-read.csv("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/Curren
names(cdp)<-c("Name","Job","Dept","FTPT","S","HpW","Sal","DpH")
cdp$Dept<-as.factor(cdp$Dept) # 36 departments
cdp$FTPT<-as.factor(cdp$FTPT)
cdp$S<-as.factor(cdp$S)
cdp$Sal[is.na(cdp$Sal)==TRUE]<-50*cdp$DpH[is.na(cdp$DpH)==FALSE]*cdp$HpW[is.na(cdp$HpW)==FALSE]
summary(cdp)
```

```
##      Name               Job                          Dept         FTPT
##  Length:31101       Length:31101       POLICE       :12537   F:30018
##  Class :character   Class :character   FIRE         : 4801   P: 1083
##  Mode  :character   Mode  :character   STREETS & SAN: 2004
##                                        AVIATION     : 1887
##                                        WATER MGMNT  : 1826
##                                        TRANSPORTN   : 1091
##                                        (Other)      : 6955
##       S              HpW             Sal              DpH
##  Hourly: 6814   Min.   :10.00   Min.   :  3000   Min.   :  3.00
##  Salary:24287   1st Qu.:40.00   1st Qu.: 80360   1st Qu.: 34.55
##                 Median :40.00   Median : 95586   Median : 39.25
##                 Mean   :36.48   Mean   : 93339   Mean   : 38.85
##                 3rd Qu.:40.00   3rd Qu.:107200   3rd Qu.: 49.30
##                 Max.   :40.00   Max.   :275004   Max.   :134.40
##                 NA's   :24287                    NA's   :24287
```

Figure 2

```
head(sort(table(cdp$Name),decreasing=TRUE),32)
```

```
##
##  HERNANDEZ,  JUAN C   ROMERO,  MIGUEL A      DELGADO,  JUAN  FLORES,  MICHAEL A
##                   5                   4                   3                   3
##  GARCIA,  ALEJANDRO    GARCIA,  GABRIEL      GARCIA,  JULIO     GARCIA,  LUIS A
##                   3                   3                   3                   3
##     GONZALEZ,  JOSE  GONZALEZ,  RICARDO  HERNANDEZ,  DANIEL   HERNANDEZ,  RUBEN
##                   3                   3                   3                   3
##  JOHNSON,  NICHOLAS      JONES,  DANIEL   KELLY,  MICHAEL J      LOPEZ,  ROBERT
##                   3                   3                   3                   3
##       NUNEZ,  JESUS      PEREZ,  CARLOS      PEREZ,  JOSE A      RIOS,  ALFREDO
##                   3                   3                   3                   3
##    RIVERA,  MICHAEL    RIVERA,  RICARDO     RIVERA,  SANDRA    RODRIGUEZ,  JOSE
##                   3                   3                   3                   3
##  RODRIGUEZ,  JOSE L    SANCHEZ,  DANIEL    SANCHEZ,  JOSE L TORRES,  JACQUELINE
##                   3                   3                   3                   3
```

```
##      VEGA,  GERARDO  ANDERSON,  DAVID C ANDERSON,  RHONDA M  ARROYO,  FRANCISCO
##                   3                    2                    2                    2
```

Figure 3

```
head(sort(table(cdp$Job),decreasing=TRUE),0.01*length(table(cdp$Job)))
```

```
## 
##                          POLICE OFFICER                          FIREFIGHTER-EMT
##                                    8876                                     1362
##                                SERGEANT POLICE OFFICER (ASSIGNED AS DETECTIVE)
##                                    1178                                     1064
##                      MOTOR TRUCK DRIVER                     POOL MOTOR TRUCK DRIVER
##                                     921                                      917
##               FIREFIGHTER-EMT (RECRUIT)                   CONSTRUCTION LABORER
##                                     736                                      438
##                          LIEUTENANT-EMT                       SANITATION LABORER
##                                     401                                      401
```

Figure 4

```
sort(table(cdp$Dept),decreasing=TRUE)
```

```
## 
##            POLICE               FIRE       STREETS & SAN           AVIATION
##             12537               4801                2004               1887
##       WATER MGMNT         TRANSPORTN      PUBLIC LIBRARY               DAIS
##              1826               1091                1018                967
##              OEMC             HEALTH    FAMILY & SUPPORT            FINANCE
##               864                554                 552                478
##      CITY COUNCIL                LAW PUBLIC SAFETY ADMIN          BUILDINGS
##               359                331                 294                232
##   BUSINESS AFFAIRS  HOUSING & ECON DEV                COPA  BOARD OF ELECTION
##               167                152                 122                109
##     MAYOR'S OFFICE      INSPECTOR GEN          CITY CLERK        PROCUREMENT
##                99                 93                  81                 78
##           HOUSING    HUMAN RESOURCES       ANIMAL CONTRL    CULTURAL AFFAIRS
##                73                 69                  62                 59
##      ADMIN HEARNG      BUDGET & MGMT           TREASURER       DISABILITIES
##                35                 33                  27                 21
##    HUMAN RELATIONS    BOARD OF ETHICS        POLICE BOARD  LICENSE APPL COMM
##                15                  8                   2                  1
```

Figure 5

```
xtabs(~FTPT+S,data=cdp)[,c("Salary","Hourly")]
```

```
##     S
## FTPT Salary Hourly
##    F  24287   5731
##    P      0   1083
```

Figure 6

```
table(cdp$HpW)[sort(names(table(cdp$HpW)),decreasing=TRUE)] # HpW (overall)
```

```
##
##   40   35   20   10
## 5703   30  857  224
```

```
table(cdp$HpW[cdp$FTPT=="F"])[sort(names(table(cdp$HpW)),decreasing=TRUE)] # HpW (FT)
```

```
##
##   40   35   20   10
## 5701   27    1    2
```

```
table(cdp$HpW[cdp$FTPT=="P"])[sort(names(table(cdp$HpW)),decreasing=TRUE)] # HpW (PT)
```

```
##
##  40  35  20  10
##   2   3 856 222
```

Figure 7

```
par(mfrow=c(1,3))
boxplot(cdp$Sal,ylab="Salary (in US$ per year)",main="(Prorated) Annual Salaries")
boxplot(cdp$Sal[cdp$S=="Salary"],cdp$Sal[cdp$S=="Hourly"],names=c("Salary","Hourly"),main="(Prorated) Sa
boxplot(cdp$Sal[cdp$FTPT=="F"],cdp$Sal[cdp$FTPT=="P"],names=c("Full-Time","Part-Time"),main="(Prorated)
```

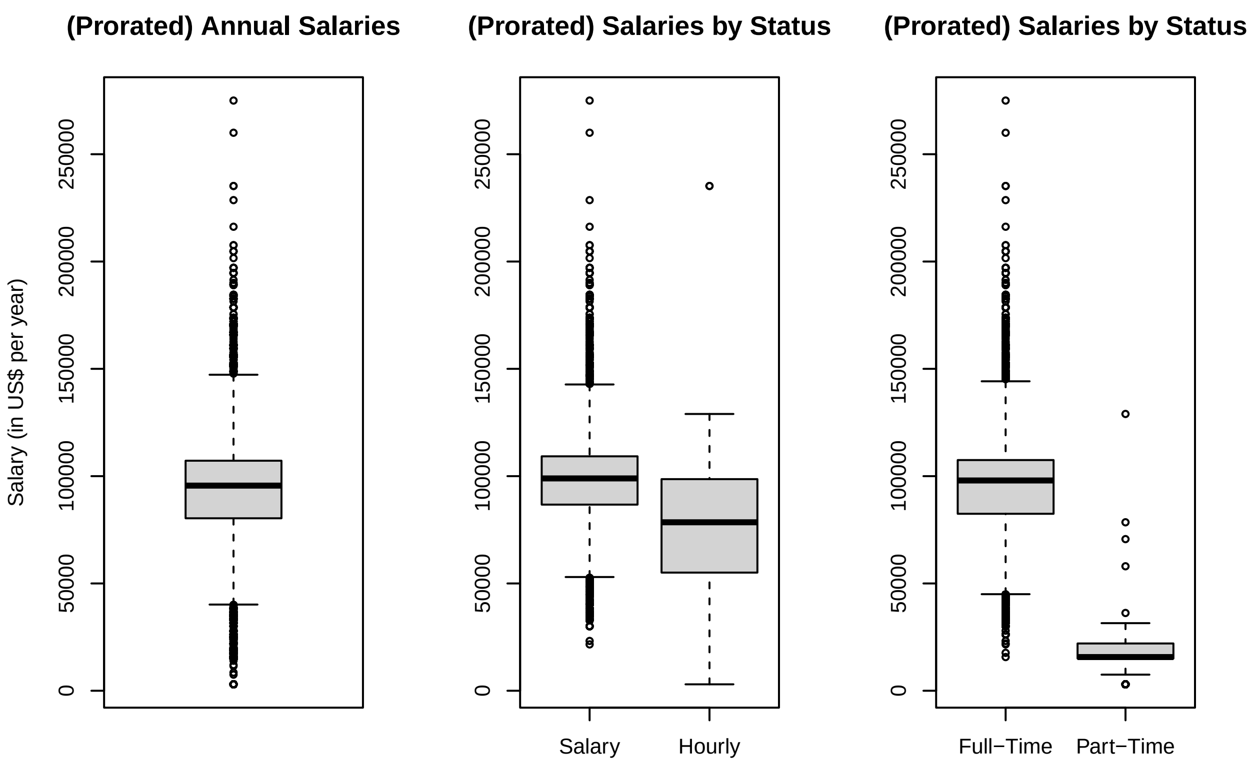


Figure 8

```
head(sort(table(cdp$DpH),decreasing=TRUE))
```

```
##
## 39.25  45.9 40.18    51    15     3
##  1538   747   390   296   257   236
```

```
head(table(cdp$DpH)[as.character(sort(as.numeric(names(table(cdp$DpH)))))])
```

```
##
##    3   12   14   15 15.8   16
##  236    1    1  257   15   13
```

Figure 9

```
par(mfrow=c(1,3))
boxplot(cdp$DpH,ylab="Wages (in US$ per hour)",main="Hourly Wages")
abline(h=15,lty=2) # Minimum wage is $15.00 per hour
boxplot(cdp$DpH[cdp$FTPT=="F"],cdp$DpH[cdp$FTPT=="P"],names=c("Full-Time","Part-Time"),xlab="Status",ma
abline(h=15,lty=2)
boxplot(subset(cdp$DpH,cdp$HpW==40),subset(cdp$DpH,cdp$HpW==35),subset(cdp$DpH,cdp$HpW==20),subset(cdp$
abline(h=15,lty=2)
```

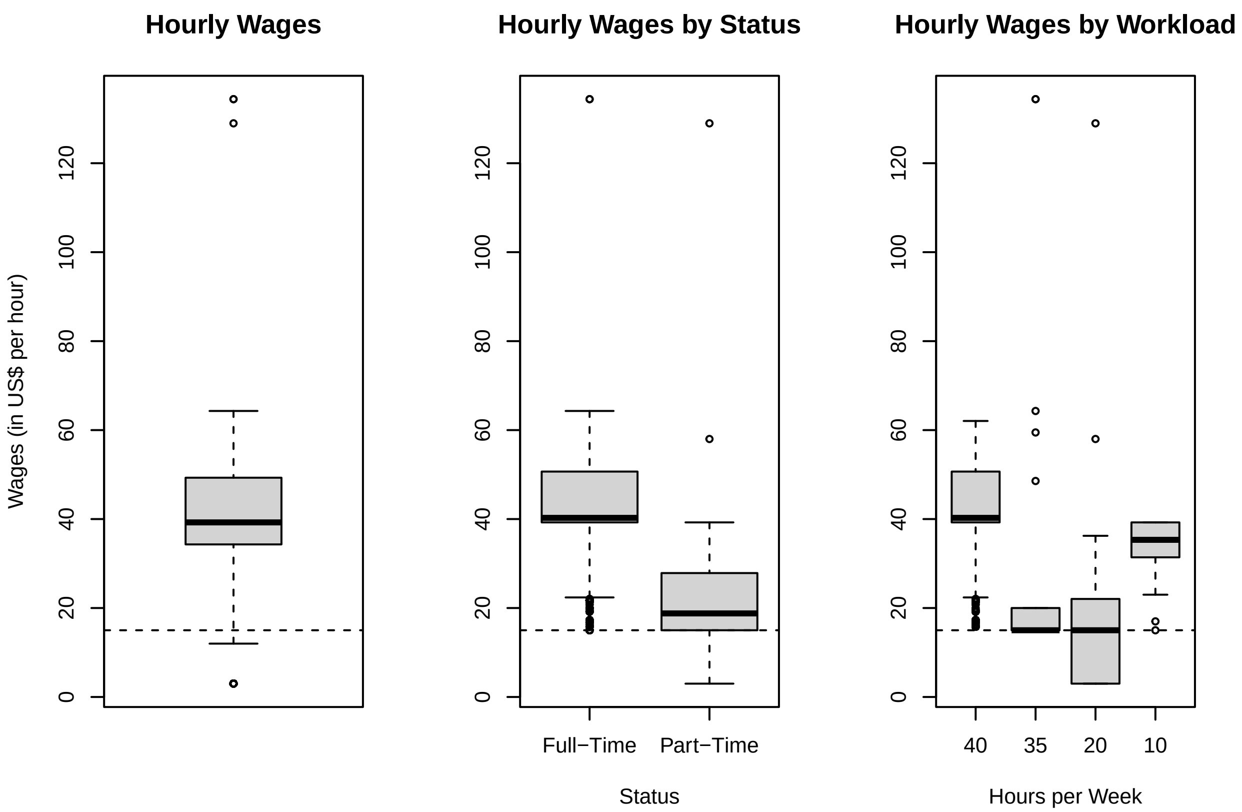


Figure 10

```
summary(glm(Sal~FTPT+S,data=cdp)) # Factor variables
```

```
##
## Call:
## glm(formula = Sal ~ FTPT + S, data = cdp)
##
## Deviance Residuals:
```

```
##    Min      1Q  Median      3Q     Max  
## -77151   -9399     189   10485  176253  
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  85072.2      283.7  299.86   <2e-16 ***
## FTPTP       -69343.5      711.6  -97.44   <2e-16 ***
## SSalary      13678.7      315.4   43.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for gaussian family taken to be 461281713)
## 
##     Null deviance: 2.1971e+13  on 31100  degrees of freedom
## Residual deviance: 1.4345e+13  on 31098  degrees of freedom
## AIC: 708716
## 
## Number of Fisher Scoring iterations: 2
```

```
summary(glm(Sal~HpW+DpH,data=cdp)) # Quantitative variables
```

```
## 
## Call:
## glm(formula = Sal ~ HpW + DpH, data = cdp)
## 
## Deviance Residuals: 
##    Min      1Q  Median      3Q     Max  
## -74295   -1252     -25    2336   13200  
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -43807.832    249.088  -175.9   <2e-16 ***
## HpW           1426.203      8.281   172.2   <2e-16 ***
## DpH           1694.573      5.246   323.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for gaussian family taken to be 20263959)
## 
##     Null deviance: 6.4738e+12  on 6813  degrees of freedom
## Residual deviance: 1.3802e+11  on 6811  degrees of freedom
##   (24287 observations deleted due to missingness)
## AIC: 133983
## 
## Number of Fisher Scoring iterations: 2
```

Figure 11

```
par(mfrow=c(1,3))
qqnorm(cdp$HpW)
qqline(cdp$HpW)
qqnorm(cdp$Sal,ylab="")
qqline(cdp$Sal)
qqnorm(cdp$DpH,ylab="")
```

```
qqline(cdp$DpH)
```

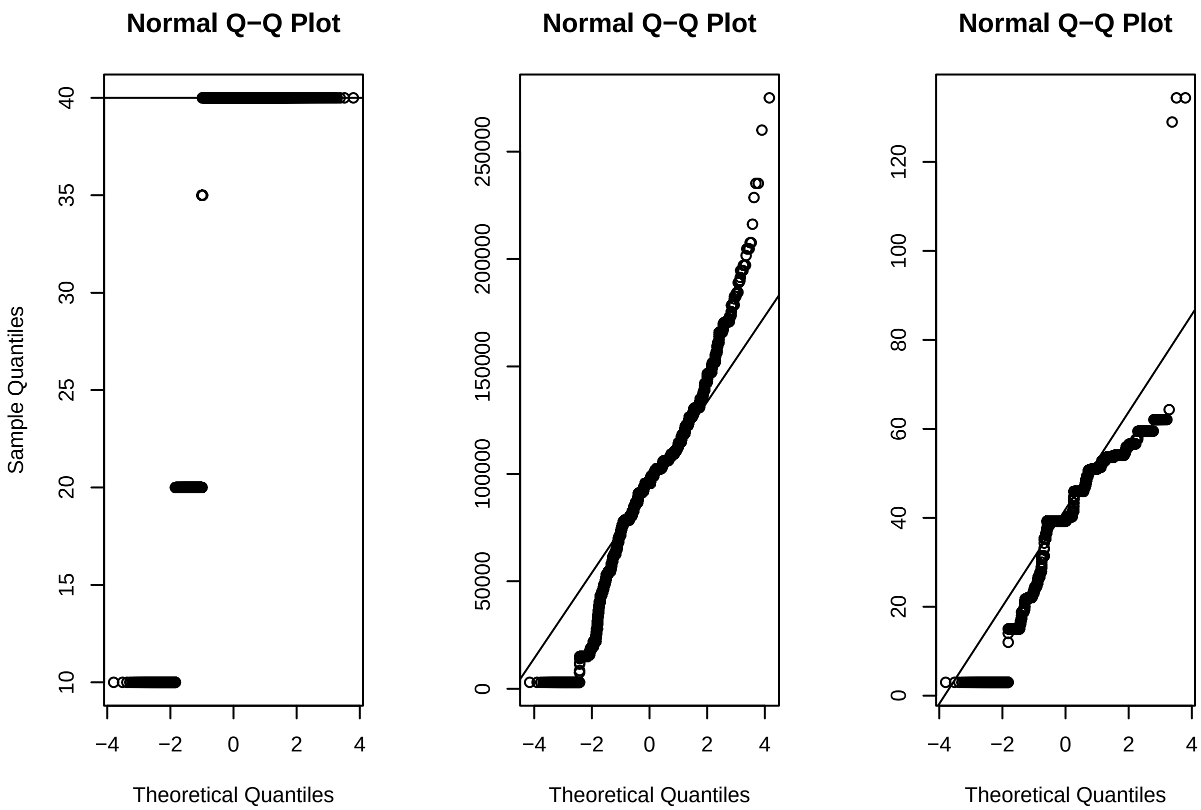
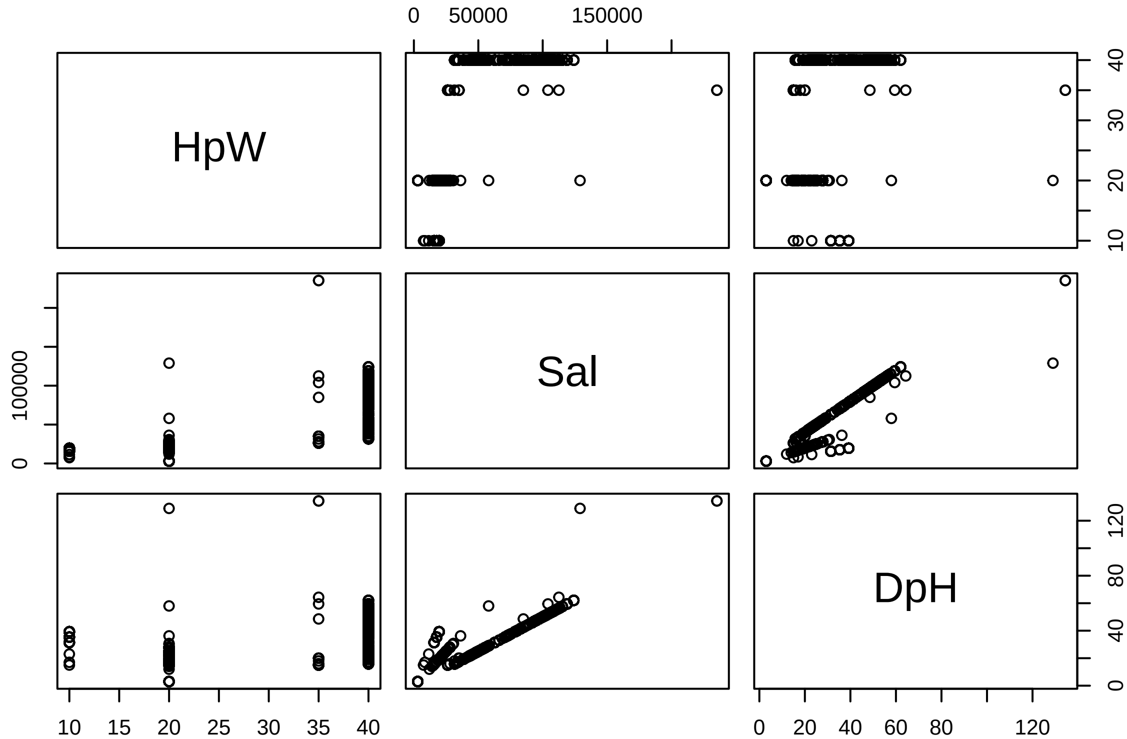


Figure 12

```
cdpn<-cdp[,c("HpW","Sal","DpH")]
library(MVN)
mvn(cdpn,univariatePlot="scatter")$Descriptives
```



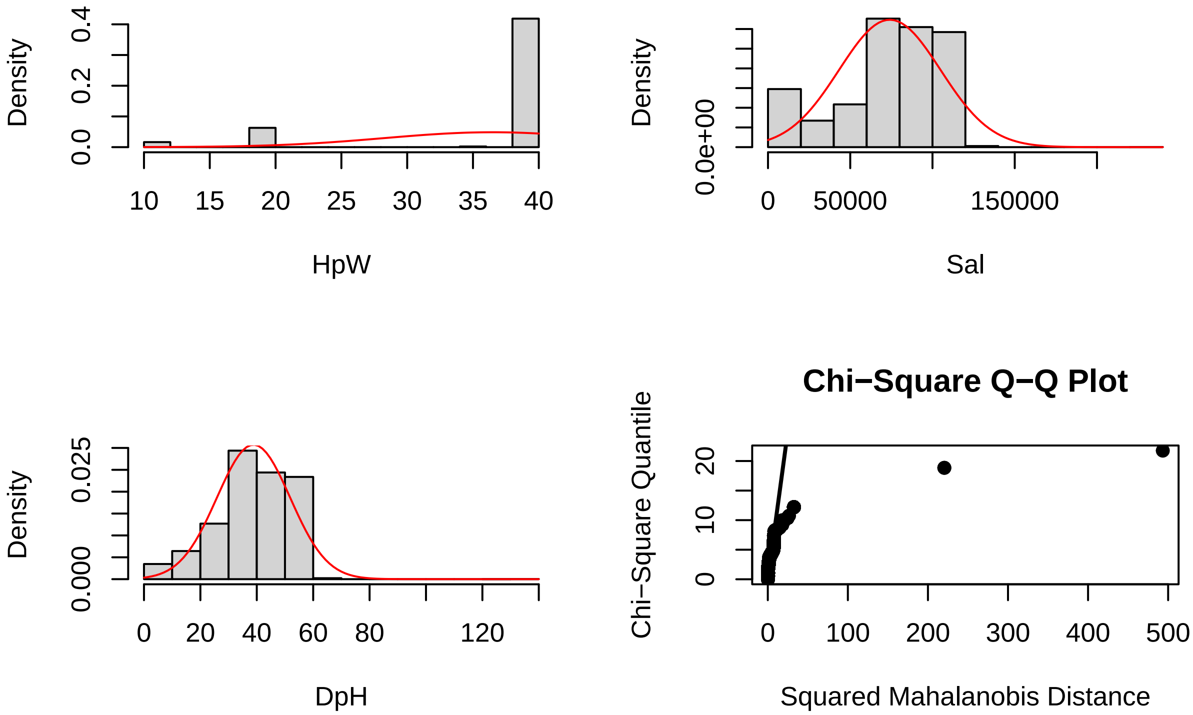

```
##         n        Mean      Std.Dev   Median  Min      Max       25th    75th
## HpW 6814    36.47637     8.221823    40.00   10     40.0    40.0000    40.0
## Sal 6814 74050.90439 30825.512998 78500.00 3000 235200.0 55060.0000 98600.0
## DpH 6814    38.85110    12.978368    39.25    3    134.4    34.5475    49.3
##           Skew    Kurtosis
## HpW -2.0441109  2.59178100
## Sal -0.8704816 -0.08249501
## DpH -0.7498379  1.47327915
```

```
mvn(cdpn,univariatePlot="histogram")$Descriptives
```

```
##         n        Mean      Std.Dev   Median  Min      Max       25th    75th
## HpW 6814    36.47637     8.221823    40.00   10     40.0    40.0000    40.0
## Sal 6814 74050.90439 30825.512998 78500.00 3000 235200.0 55060.0000 98600.0
## DpH 6814    38.85110    12.978368    39.25    3    134.4    34.5475    49.3
##           Skew    Kurtosis
## HpW -2.0441109  2.59178100
## Sal -0.8704816 -0.08249501
## DpH -0.7498379  1.47327915
```

```
mvn(cdpn,multivariatePlot="qq",bc=TRUE)$BoxCoxPowerTransformation
```

```
## Warning in log(((nr - 1)/nr) * det(var(qr.resid(xqr, w * fam(Y, lam,
## jacobian.adjusted = TRUE))))): NaNs produced
```




```
##  HpW  Sal  DpH
## 6.77 0.69 1.18
```

Figure 13

```
x<-colMeans(cdpn[is.na(cdpn$HpW)==FALSE,c("HpW","DpH")])
x
```

```
##      HpW      DpH
## 36.47637 38.85110
```

```
cor(cdpn[is.na(cdpn$HpW)==FALSE,c("HpW","DpH")])
```

```
##          HpW      DpH
## HpW 1.000000 0.598658
## DpH 0.598658 1.000000
```

```
vm<-cov(cdpn[is.na(cdpn$HpW)==FALSE,c("HpW","DpH")])
vm
```

```
##          HpW       DpH
## HpW 67.59837  63.88031
## DpH 63.88031 168.43804
```

```
n<-dim(cdpn[is.na(cdpn$HpW)==FALSE,c("HpW","DpH")])[1]
p<-length(cdpn[,c("HpW","DpH")])
(n-1)*p/(n-p)*qf(1-0.05,p,n-p)
```

```
## [1] 5.99498
```

```
n*t(x-round(x))%*%solve(vm)%*%(x-round(x))
```

```
##           [,1]
## [1,] 45.50302
```

```r
n*t(matrix(c(11/60+44/60/60,0)))%*%solve(vm)%*%(matrix(c(11/60+44/60/60,0)))
```

```
##          [,1]
## [1,] 6.008085
```

```r
n*t(matrix(c(0,0.31)))%*%solve(vm)%*%(matrix(c(0,0.31)))
```

```
##          [,1]
## [1,] 6.059199
```

Figure 14

```r
library(purrr)
prcomp(cdpn[is.na(cdpn$HpW)==FALSE,c("HpW","DpH")],scale=TRUE)
```

```
## Standard deviations (1, .., p=2):
## [1] 1.2643805 0.6335156
##
## Rotation (n x k) = (2 x 2):
##            PC1        PC2
## HpW -0.7071068  0.7071068
## DpH -0.7071068 -0.7071068
```

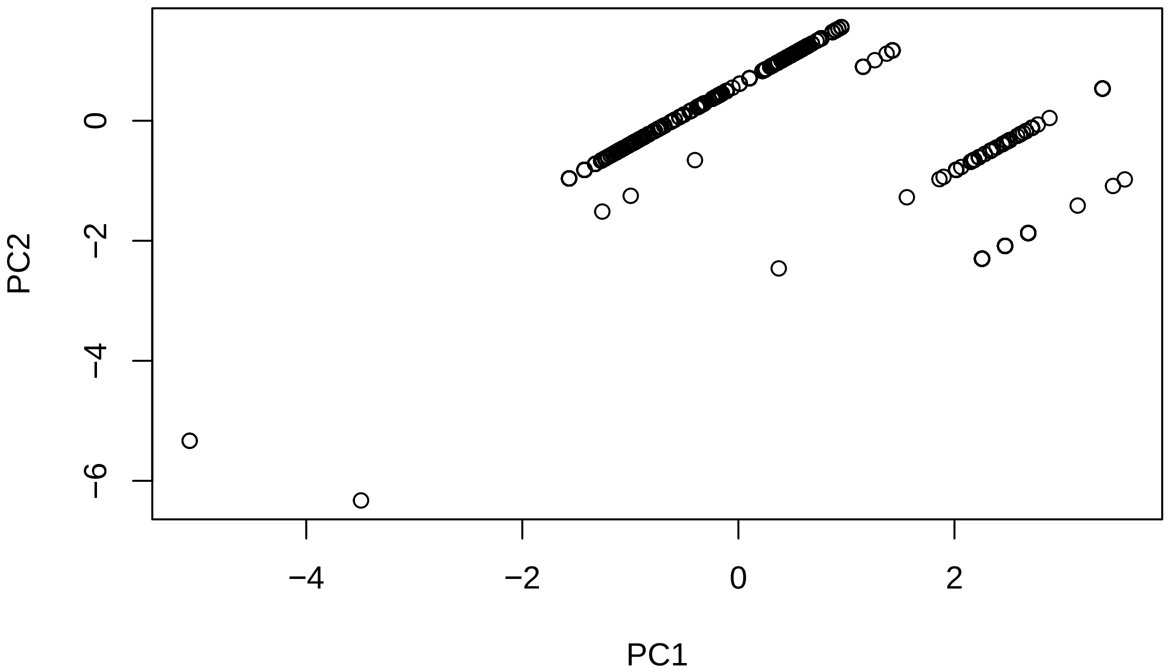
```r
summary(prcomp(cdpn[is.na(cdpn$HpW)==FALSE,c("HpW","DpH")],scale=TRUE))
```

```
## Importance of components:
##                           PC1    PC2
## Standard deviation     1.2644 0.6335
## Proportion of Variance 0.7993 0.2007
## Cumulative Proportion  0.7993 1.0000
```

```r
as.data.frame(map(prcomp(cdpn[is.na(cdpn$HpW)==FALSE,c("HpW","DpH")],scale=TRUE),sd))
```

```
##        sdev  rotation   center    scale         x
## 1 0.4460889 0.7071068 1.679186 3.363385 0.9999633
```

```r
plot(predict(prcomp(cdpn[is.na(cdpn$HpW)==FALSE,c("HpW","DpH")],scale=TRUE)))
```




```
screeplot(prcomp(cdpn[is.na(cdpn$HpW)==FALSE,c("HpW","DpH")],scale=TRUE),type="lines")
```

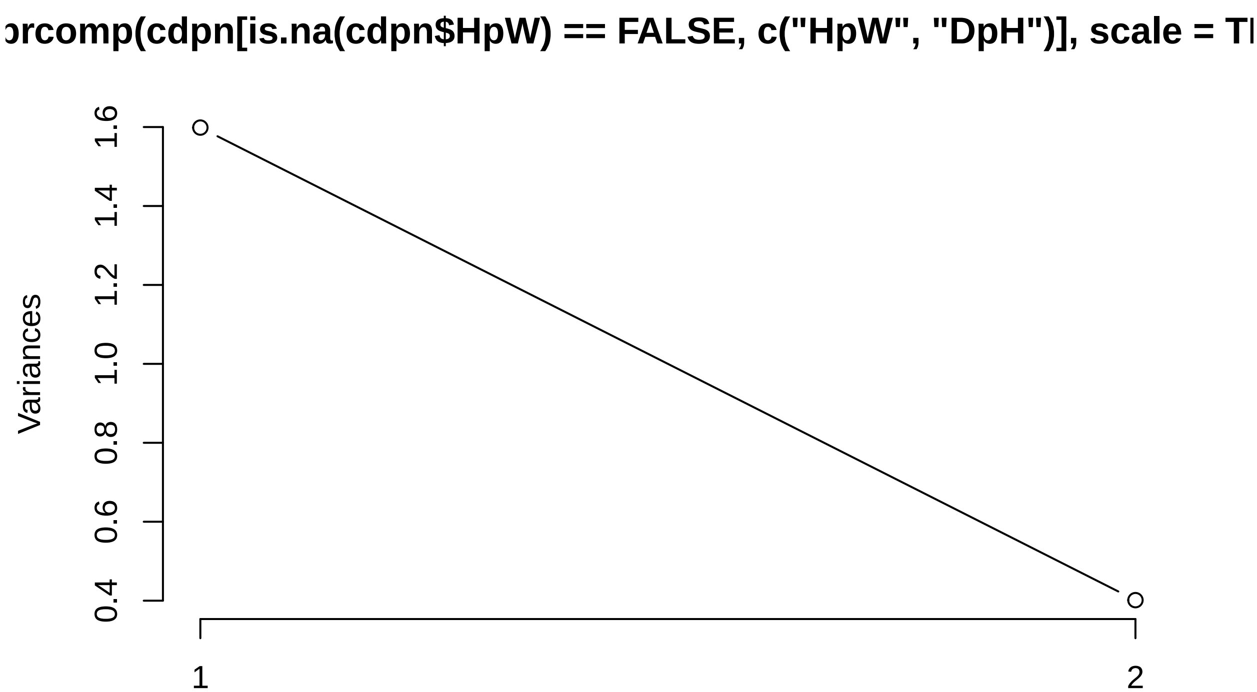
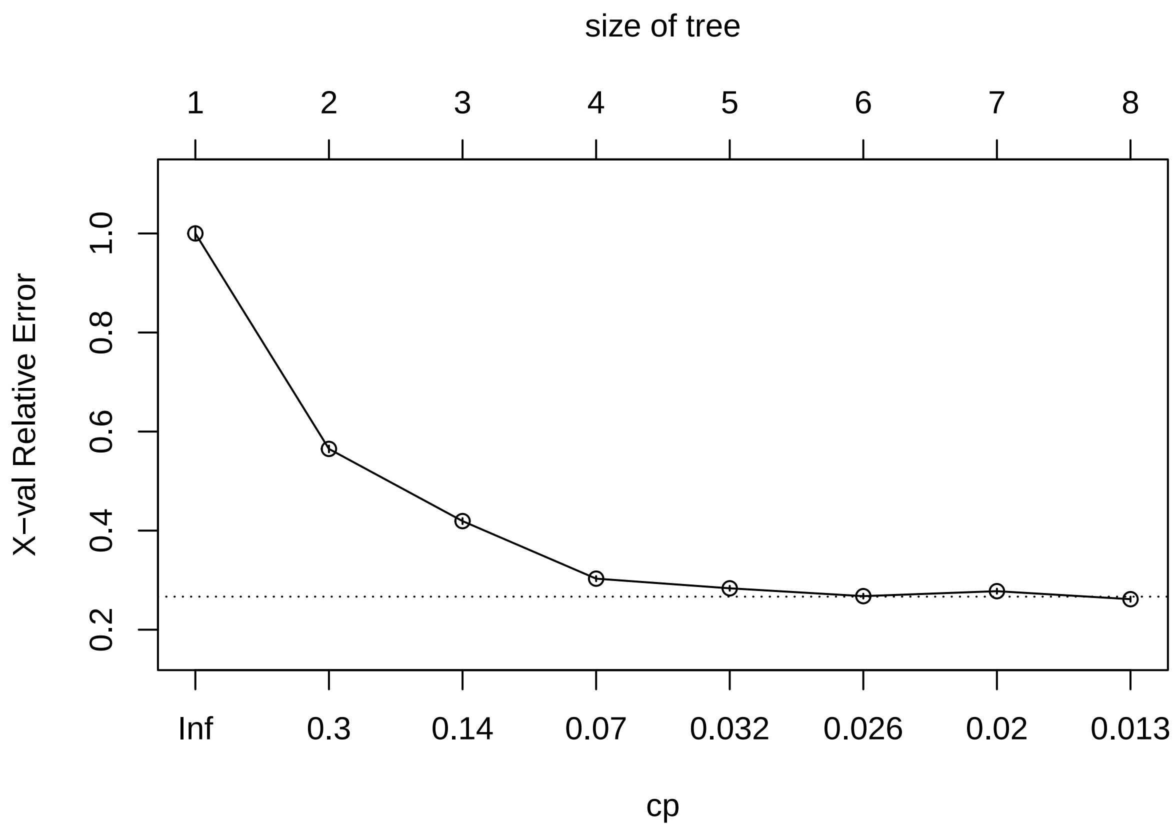


Figure 15

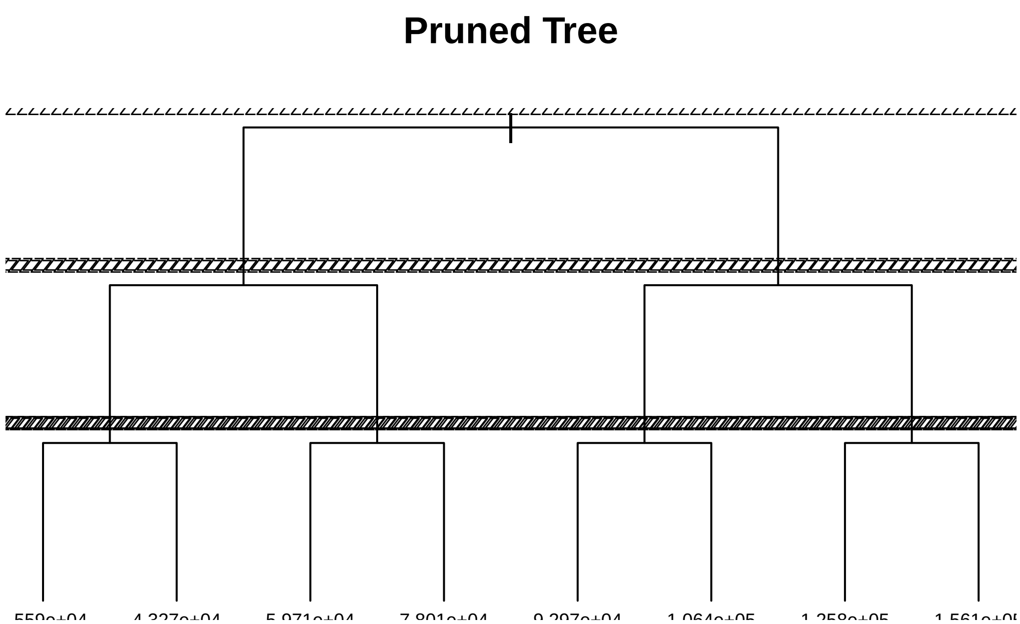
```
library(rpart)
set.seed(3422)
tree<-rpart(Sal~.,data=cdp)
ptree<-prune(tree,cp=tree$cptable[which.min(tree$cptable[,"xerror"]),"CP"])
printcp(ptree)
```

```
##
## Regression tree:
## rpart(formula = Sal ~ ., data = cdp)
##
## Variables actually used in tree construction:
## [1] Name
##
## Root node error: 2.1971e+13/31101 = 706447084
##
## n= 31101
##
##         CP nsplit rel error  xerror      xstd
## 1 0.560677      0  1.000000 1.00005 0.0113064
## 2 0.155520      1  0.439323 0.56499 0.0068044
## 3 0.134493      2  0.283803 0.41926 0.0059595
## 4 0.036133      3  0.149311 0.30303 0.0051153
## 5 0.028428      4  0.113178 0.28356 0.0049247
## 6 0.024608      5  0.084750 0.26773 0.0048262
## 7 0.016608      6  0.060142 0.27771 0.0050902
## 8 0.010000      7  0.043533 0.26158 0.0050950
```

```
plotcp(ptree)
```




```
plot(ptree,uniform=TRUE,main="Pruned Tree")
text(ptree,cex=0.6)
```




```r
table(predict(ptree))
```

```
##
## 15591.1800554017       43272.1952 59710.6694151329 78007.4134089954
##             1083              850             2445             4758
## 92971.1363801673 106356.573247676 125772.944966525  156086.04126021
##             8249             9573             3286              857
```

```r
min(tree$cptable[,"xerror"]) # Cross-validation error
```

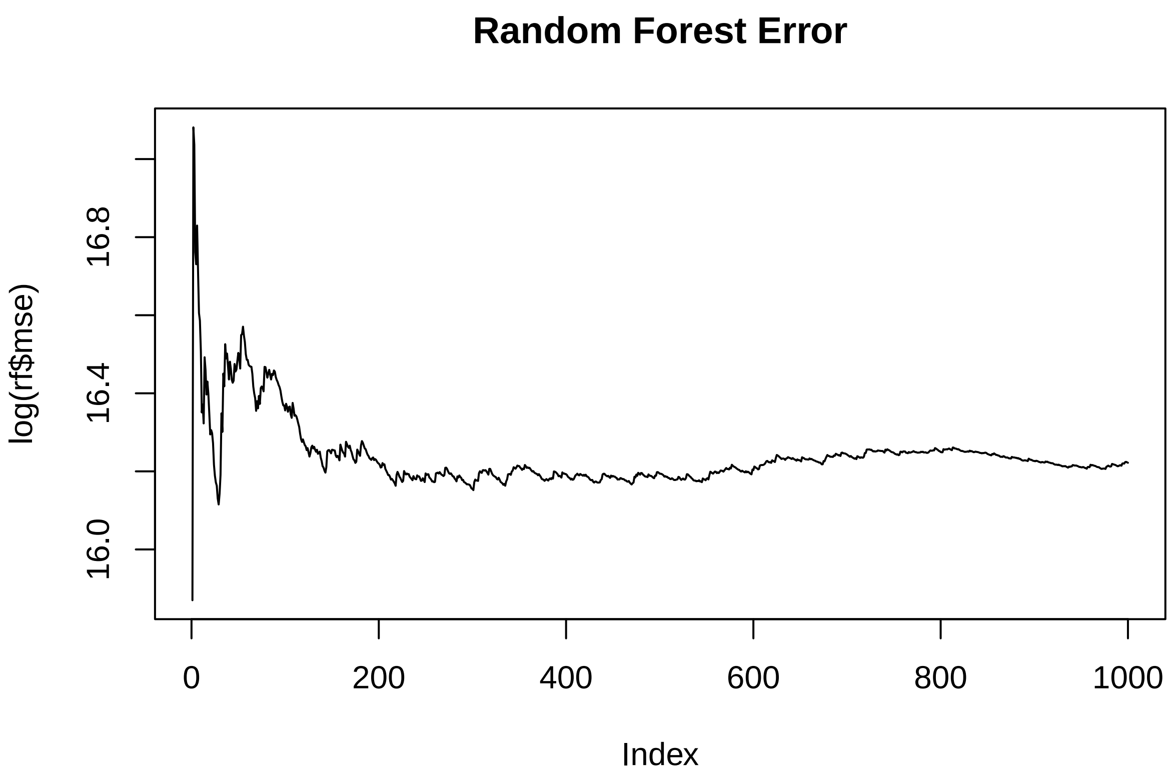
```
## [1] 0.2615813
```

Figure 16

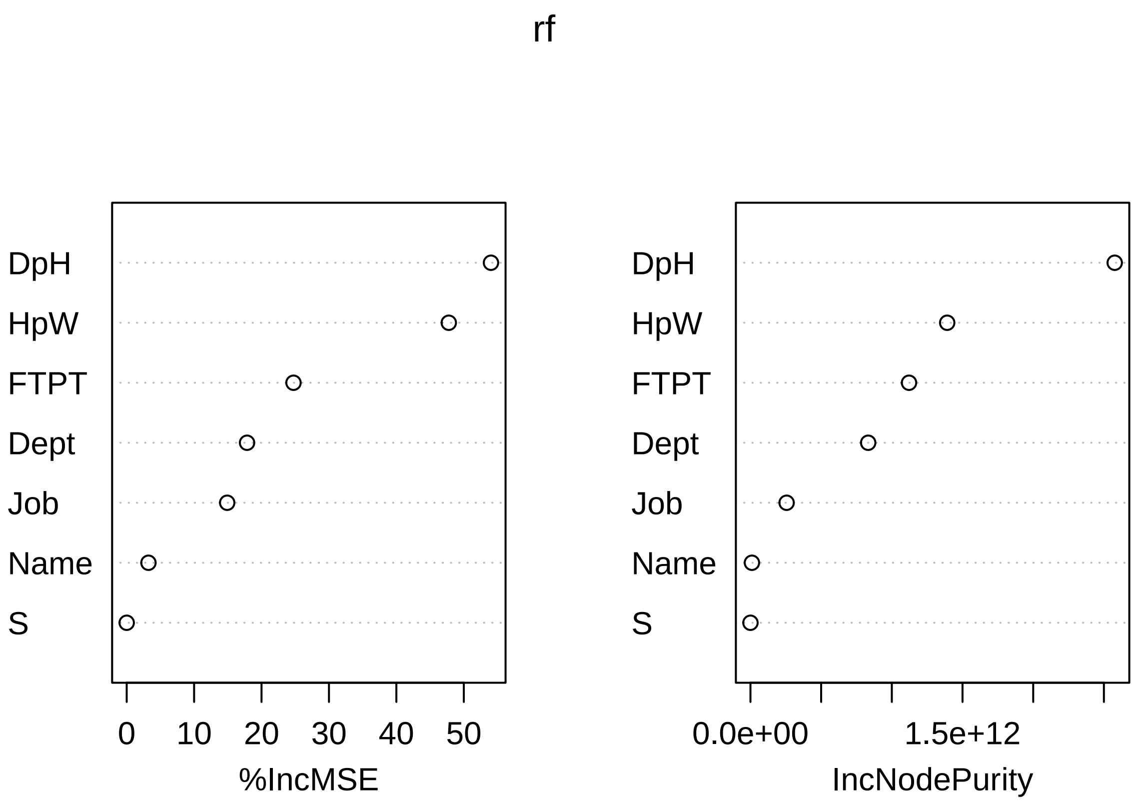
```r
library(randomForest)
set.seed(3422)
rf<-randomForest(Sal~.,data=cdp[!is.na(cdp$HpW),],ntree=1000,importance=TRUE)
rf
```

```
##
## Call:
##  randomForest(formula = Sal ~ ., data = cdp[!is.na(cdp$HpW), ],      ntree = 1000, importance = TRUE)
##                Type of random forest: regression
##                      Number of trees: 1000
## No. of variables tried at each split: 2
##
##           Mean of squared residuals: 11094092
##                     % Var explained: 98.83
```

```r
plot(log(rf$mse),type="l",main="Random Forest Error")
```




```
varImpPlot(rf)
```



Works Cited

"Current Employee Names, Salaries, and Position Titles." *Chicago Data Portal*, 27 Sept. 2011, https://data.cityofchicago.org/Administration-Finance/Current-Employee-Names-Salaries-and-Position-Title/xzkq-xp2w. Accessed 3 April 2022.