

# HR Data Analysis

```
# Clear variables in memory
```

```
rm(list=ls())
```

```
# Raw Data: HRDataset_v14.csv
```

```
# Read csv file
```

```
HR <- read.csv("HRSub2.csv", head = TRUE, sep = ",")
```

```
head(HR)
```

```
## MarriedID MaritalDesc Sex EmploymentStatus Department PerfScoreID
## 1 0 Single M Active Production 4
## 2 1 Married M Voluntarily Terminated IT/IS 3
## 3 1 Married F Voluntarily Terminated Production 3
## 4 1 Married F Active Production 3
## 5 0 Divorced F Voluntarily Terminated Production 3
## 6 0 Single F Active Production 4
## RecruitmentSource Salary Position State Age CitizenDesc
## 1 LinkedIn 62506 Production Technician I MA 38 US Citizen
## 2 Indeed 104437 Sr. DBA MA 46 US Citizen
## 3 LinkedIn 64955 Production Technician II MA 33 US Citizen
## 4 Indeed 64991 Production Technician I MA 33 US Citizen
## 5 Google Search 50825 Production Technician I MA 32 US Citizen
## 6 LinkedIn 57568 Production Technician I MA 44 US Citizen
## RaceDesc HispanicLatino EmployedYear ManagerName EngagementSurvey
## 1 White No 10 Michael Albert 4.60
## 2 White No 1 Simon Roup 4.96
## 3 White No 1 Kissy Sullivan 3.02
## 4 White No 13 Elijah Gray 4.84
## 5 White No 5 Webster Butler 5.00
## 6 White No 9 Amy Dunn 5.00
## EmpSatisfaction SpecialProjectsCount DaysLateLast30 Absences
## 1 5 0 0 1
## 2 3 6 0 17
## 3 3 0 0 3
## 4 5 0 0 15
## 5 4 0 0 2
## 6 5 0 0 15
```

```
#Import Libraries
```

```
library(glmnet)
```

```
## Warning: package 'glmnet' was built under R version 4.1.2
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-3
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.1.2
```

```
## Loading required package: ggplot2
## Loading required package: lattice
library(car)

## Loading required package: carData
library(CombMSC)

## Warning: package 'CombMSC' was built under R version 4.1.2
##
## Attaching package: 'CombMSC'
## The following object is masked from 'package:car':
##
##     subsets
## The following object is masked from 'package:stats':
##
##     BIC
library(ggplot2)
```

## Preprocessing Data

```
# Correct Errors
HR$HispanicLatino[188]="Yes"
HR$HispanicLatino[98]="No"

# Factorization
HR$MarriedID = as.factor(HR$MarriedID)
HR$MaritalDesc = as.factor(HR$MaritalDesc)
HR$Sex = as.factor(HR$Sex)
HR$EmploymentStatus = as.factor(HR$EmploymentStatus)
HR$Department = as.factor(HR$Department)
HR$RecruitmentSource = as.factor(HR$RecruitmentSource)
HR$Position = as.factor(HR$Position)
HR$State = as.factor(HR$State)
HR$CitizenDesc = as.factor(HR$CitizenDesc)
HR$RaceDesc = as.factor(HR$RaceDesc)
HR$HispanicLatino = as.factor(HR$HispanicLatino)
HR$ManagerName = as.factor(HR$ManagerName)
```

## Split the Data set in training and testing data

```
set.seed(123)
train_index = createDataPartition(HR$Salary, p = 0.7, list = FALSE, times = 1)
HRtrain = HR[train_index,]
HRtest = HR[-train_index,]

# Deal with the data absence in training dataset

# Department
```

```

test_unique = unique(HRtest$Department)
train_unique = unique(HRtrain$Department)
for (i in test_unique){
  if ((i %in% train_unique)==FALSE){
    add_row = HRtest[HRtest$Department == i,]
    HRtrain = rbind(HRtrain, add_row)
  }
}

# Position
test_unique = unique(HRtest$Position)
train_unique = unique(HRtrain$Position)
for (i in test_unique){
  if ((i %in% train_unique)==FALSE){
    add_row = HRtest[HRtest$Position == i,]
    HRtrain = rbind(HRtrain, add_row)
  }
}

# State
test_unique = unique(HRtest$State)
train_unique = unique(HRtrain$State)
for (i in test_unique){
  if ((i %in% train_unique)==FALSE){
    add_row = HRtest[HRtest$State == i,]
    HRtrain = rbind(HRtrain, add_row)
  }
}

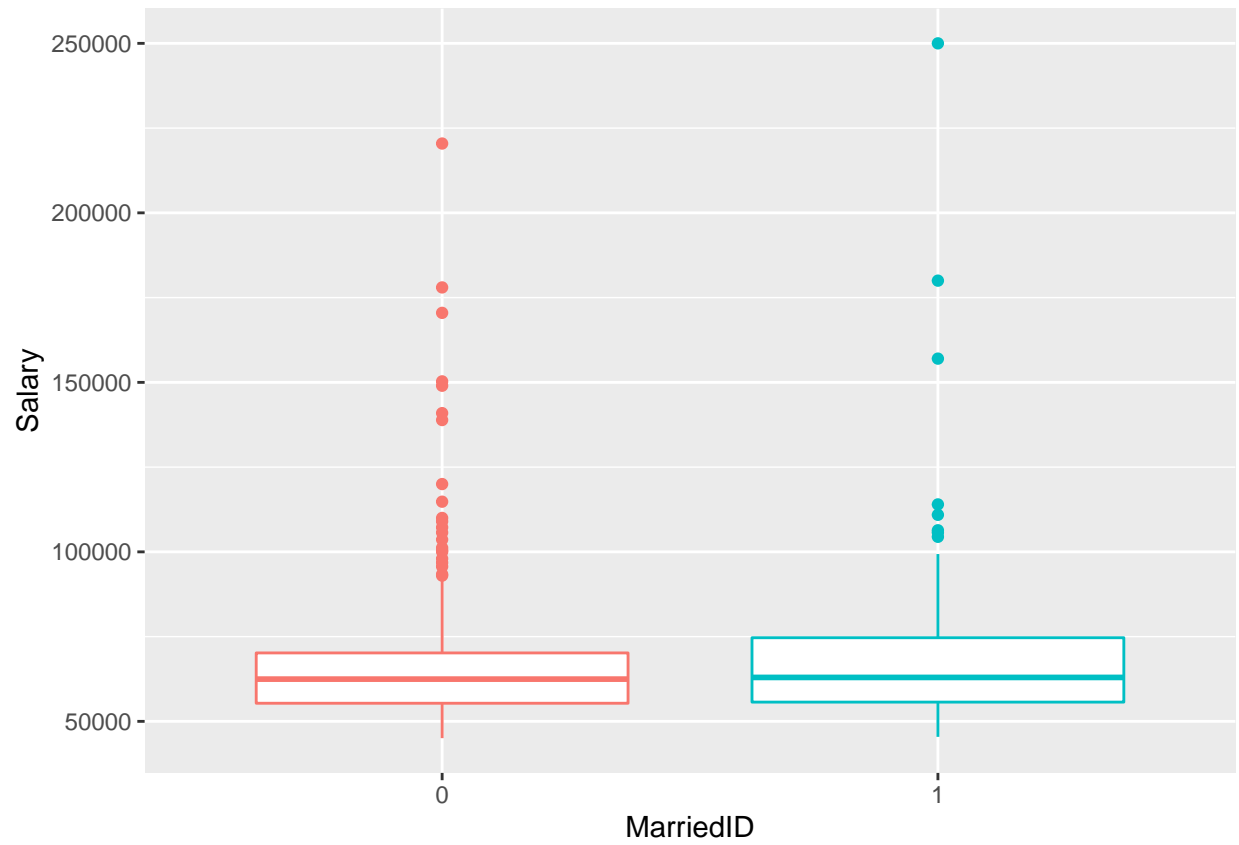
```

## Exploratory Data Analysis - Boxplots

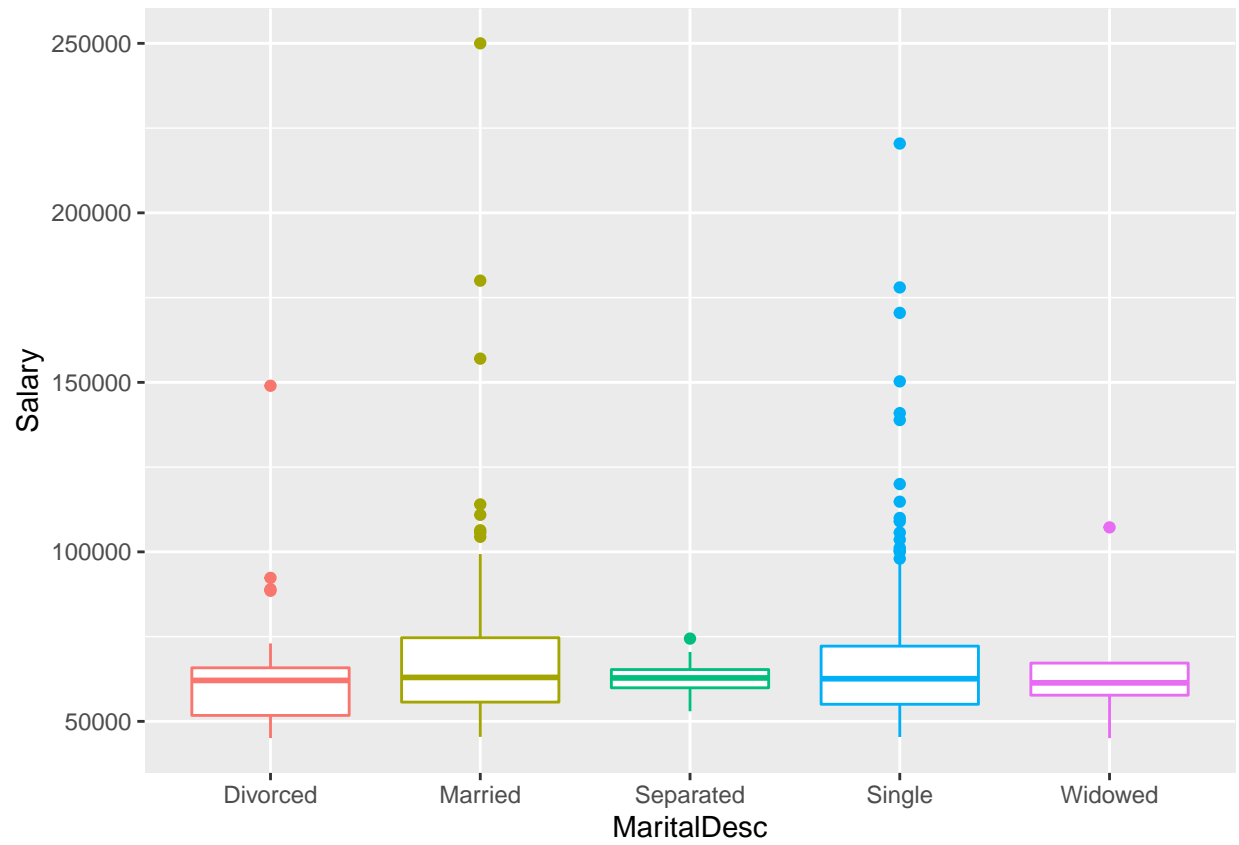
```

ggplot(HR, aes(x=MarriedID, y=Salary, color=MarriedID)) +
  geom_boxplot() + theme(legend.position="none")

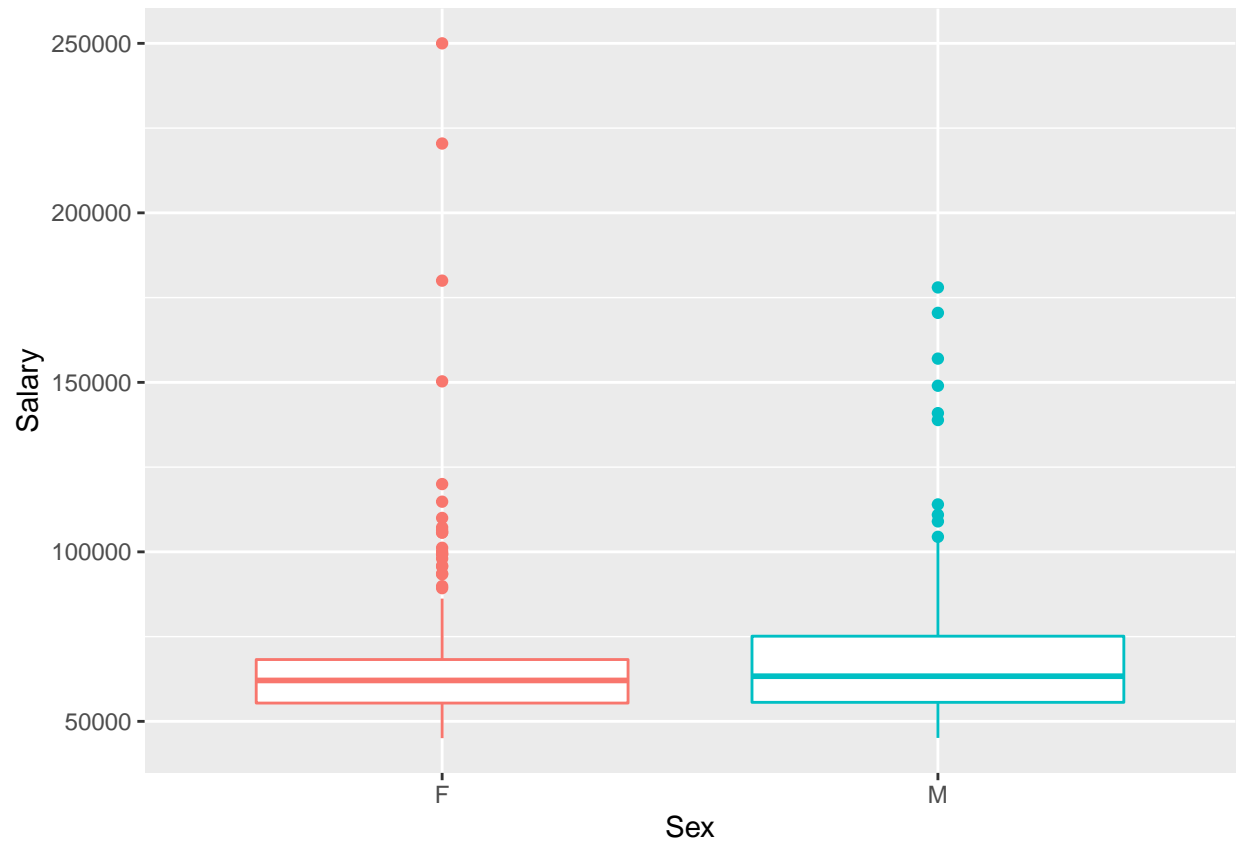
```



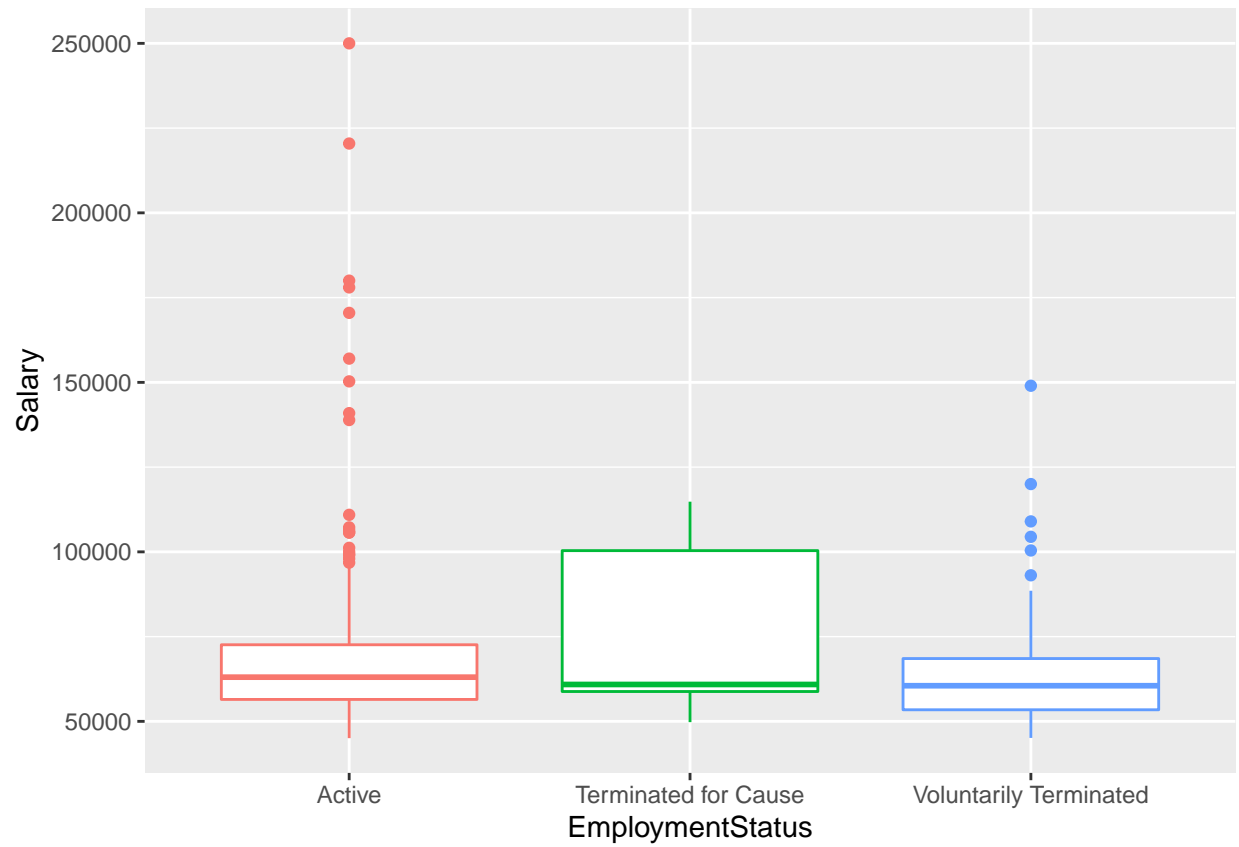
```
ggplot(HR, aes(x=MaritalDesc, y=Salary, color=MaritalDesc)) +  
  geom_boxplot()+ theme(legend.position="none")
```



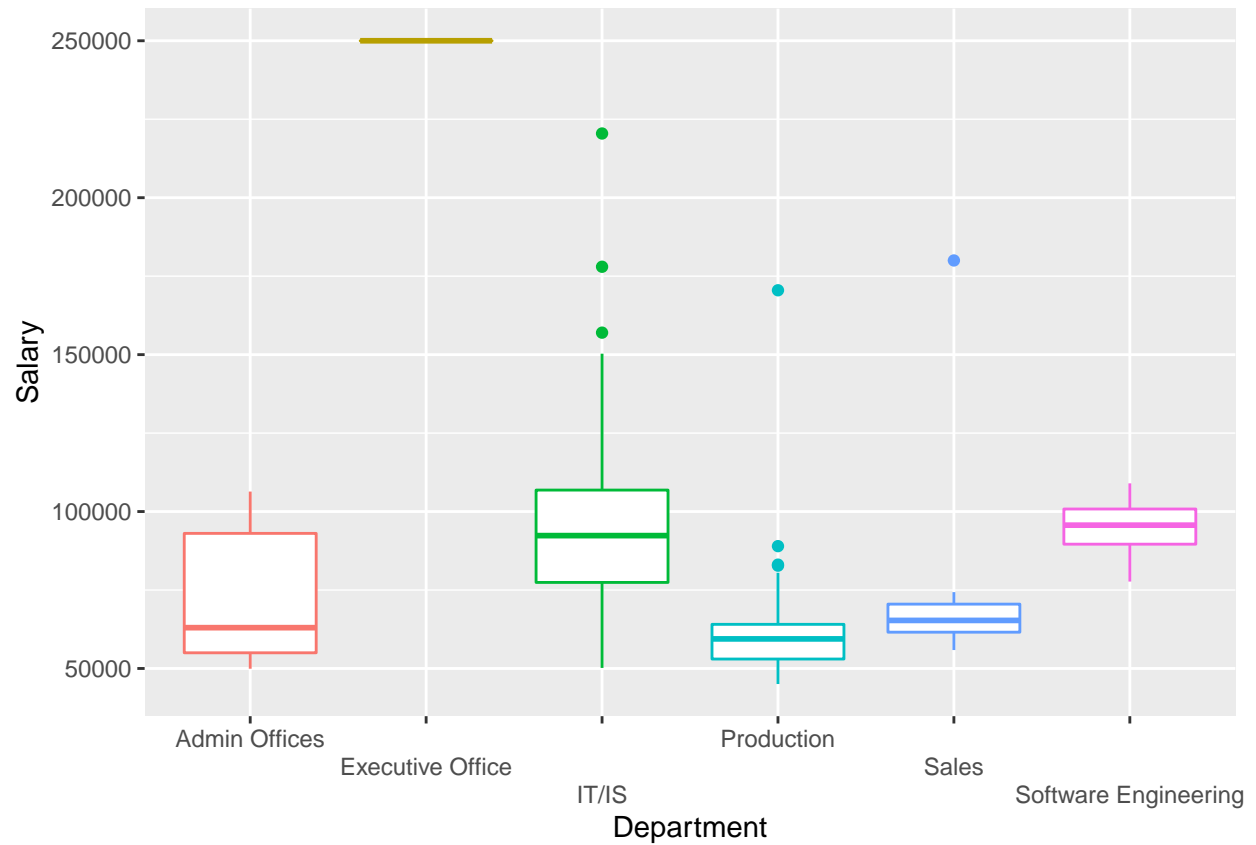
```
ggplot(HR, aes(x=Sex, y=Salary, color=Sex)) +  
  geom_boxplot()+ theme(legend.position="none")
```



```
ggplot(HR, aes(x=EmploymentStatus, y=Salary, color=EmploymentStatus)) +  
  geom_boxplot()+ theme(legend.position="none")
```

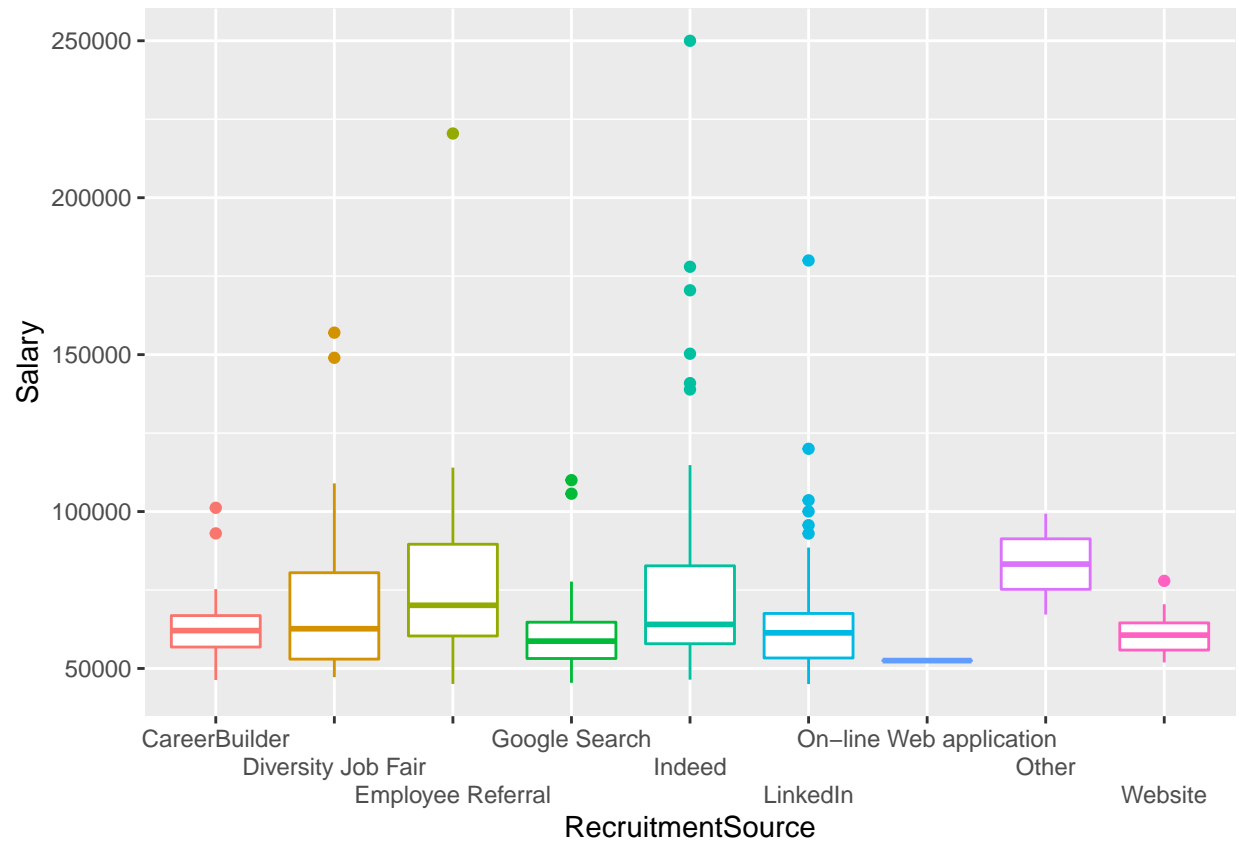


```
ggplot(HR, aes(x=Department, y=Salary, color=Department)) +  
  geom_boxplot() + scale_x_discrete(guide = guide_axis(n.dodge=3)) + theme(legend.position="none")
```

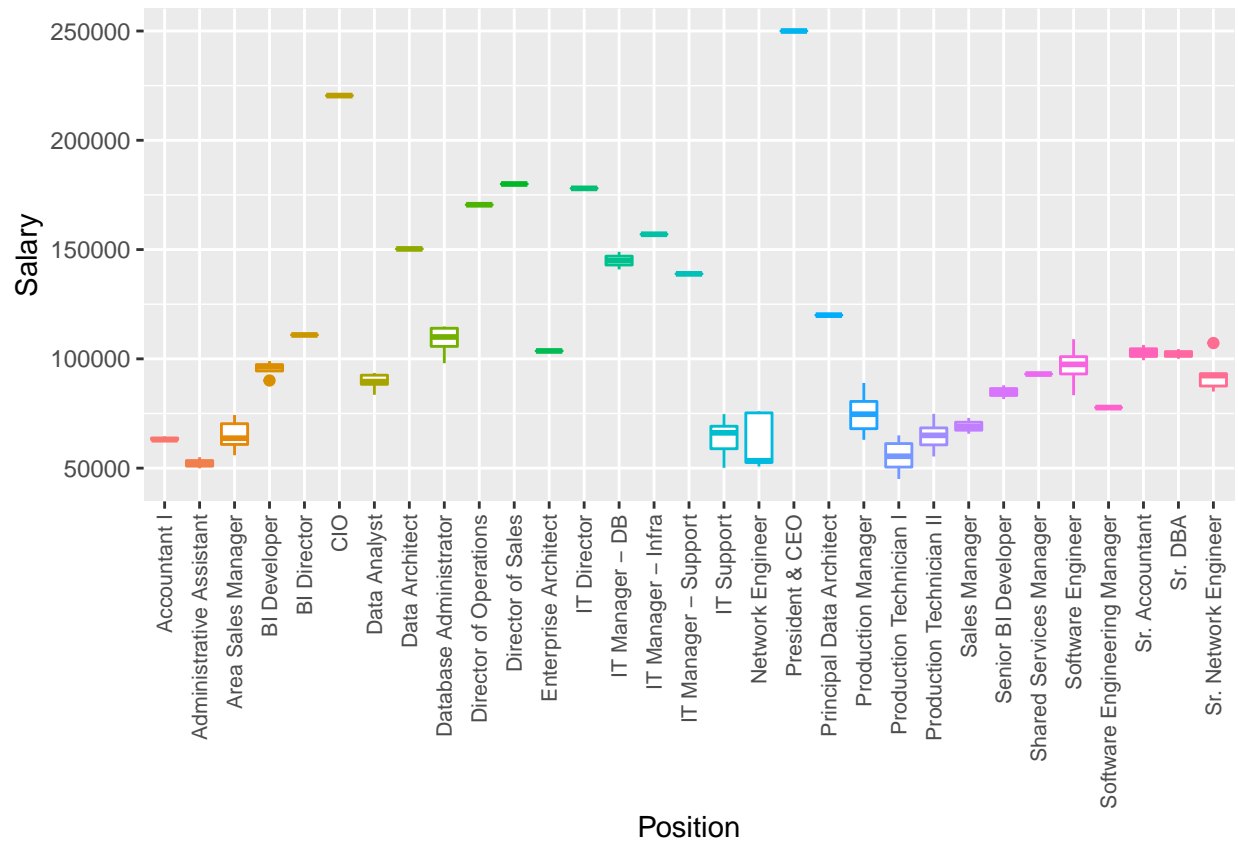


```
ggplot(HR, aes(x=RecruitmentSource, y=Salary, color=RecruitmentSource)) +
  geom_boxplot() + scale_x_discrete(guide = guide_axis(n.dodge=3)) + theme(legend.position="none")
```

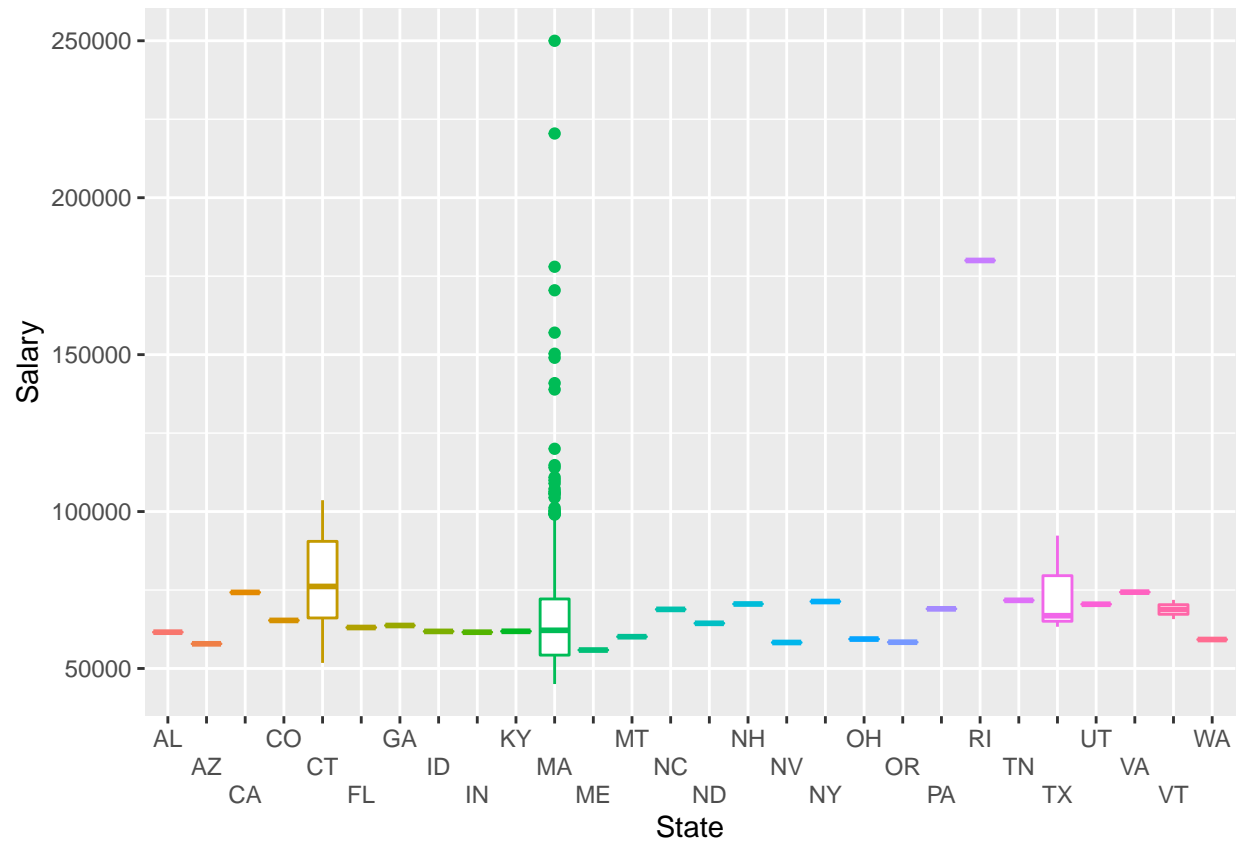




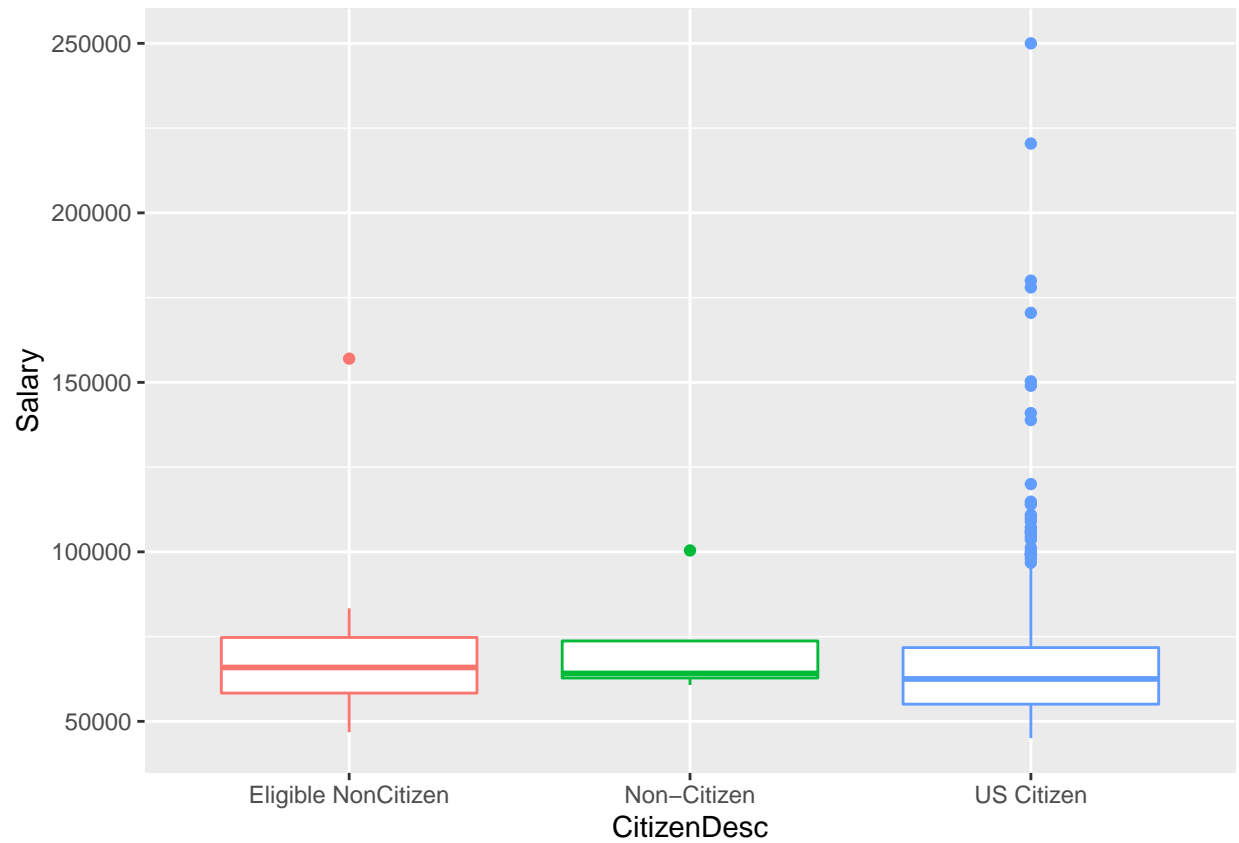
```
ggplot(HR, aes(x=Position, y=Salary, color=Position)) +
  geom_boxplot() + theme(axis.text.x = element_text(size = 8, angle = 90, vjust = 0.5, hjust=1)) + theme
```



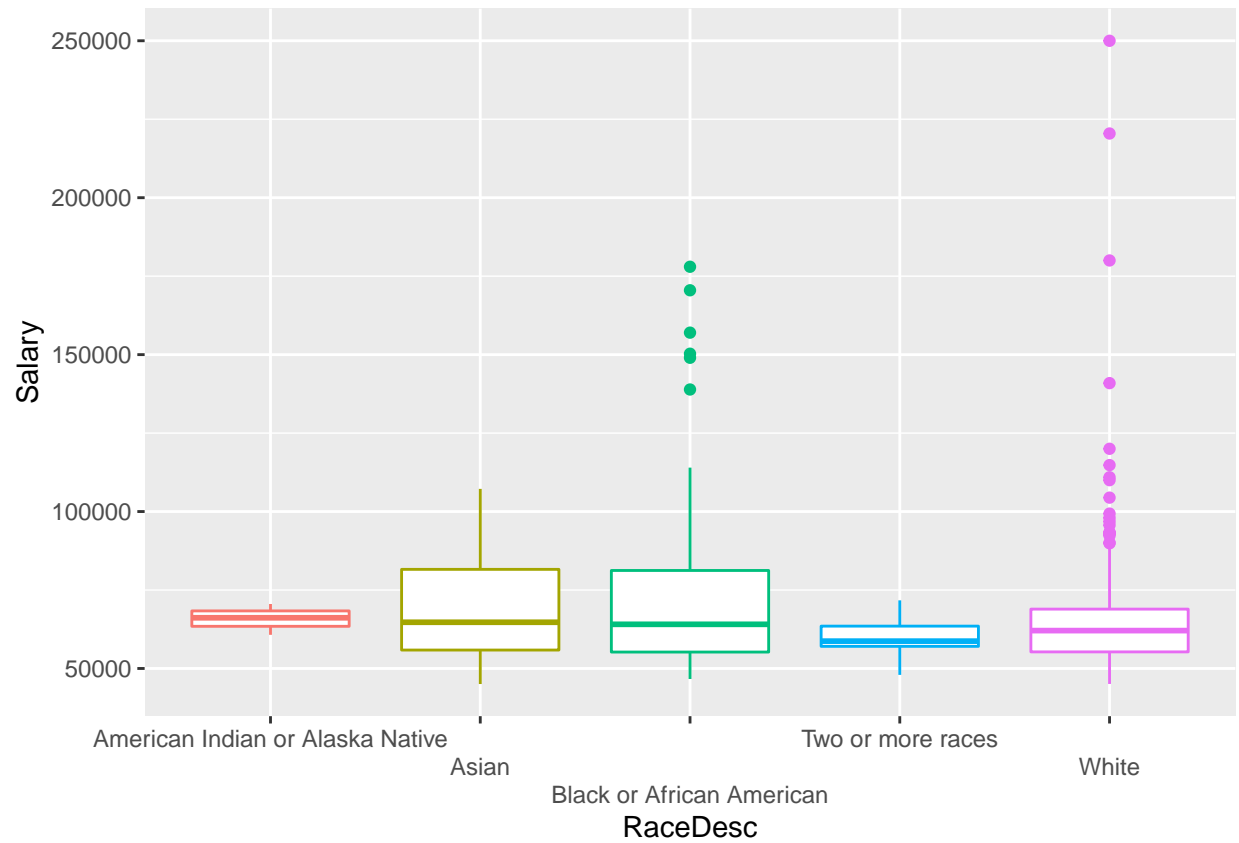
```
ggplot(HR, aes(x=State, y=Salary, color=State)) +
  geom_boxplot() + scale_x_discrete(guide = guide_axis(n.dodge=3)) + theme(legend.position="none")
```



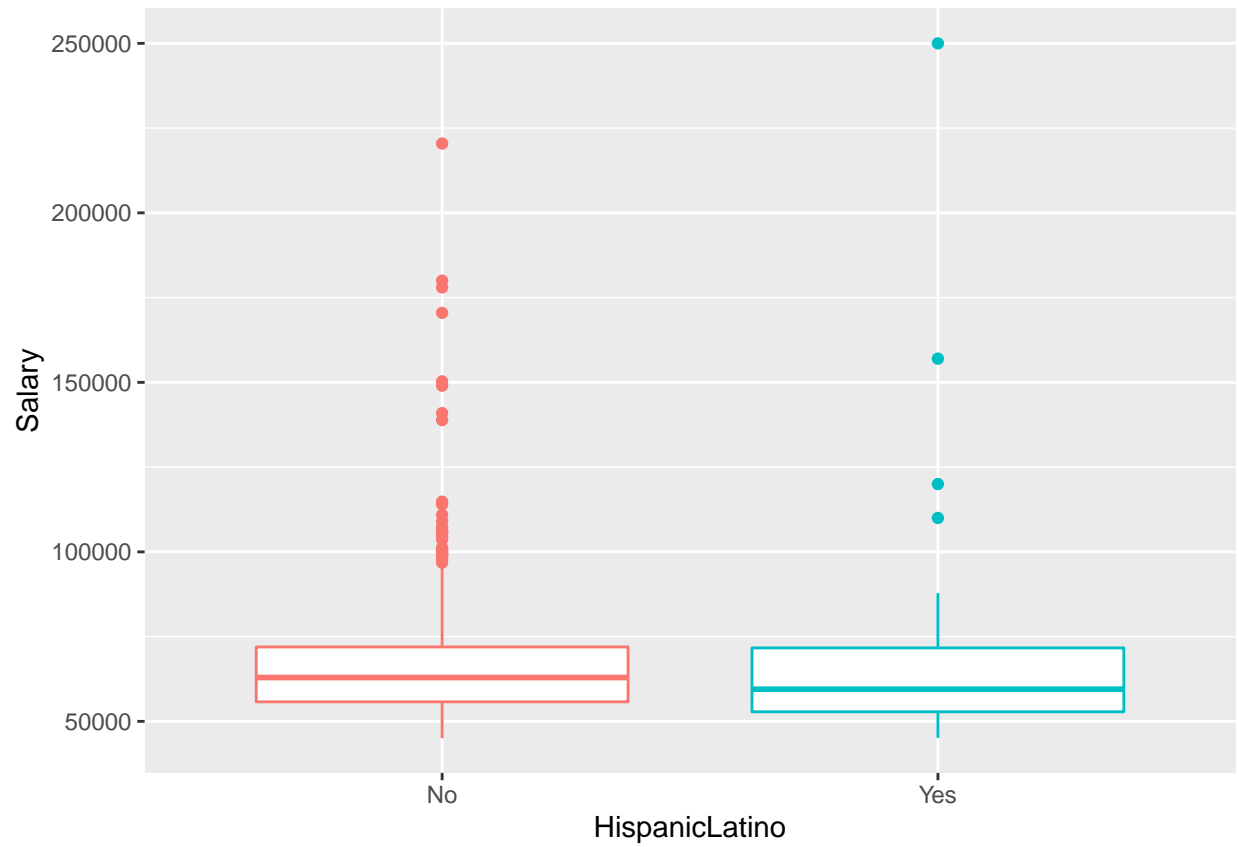
```
ggplot(HR, aes(x=CitizenDesc, y=Salary, color=CitizenDesc)) +
  geom_boxplot() + theme(legend.position="none")
```



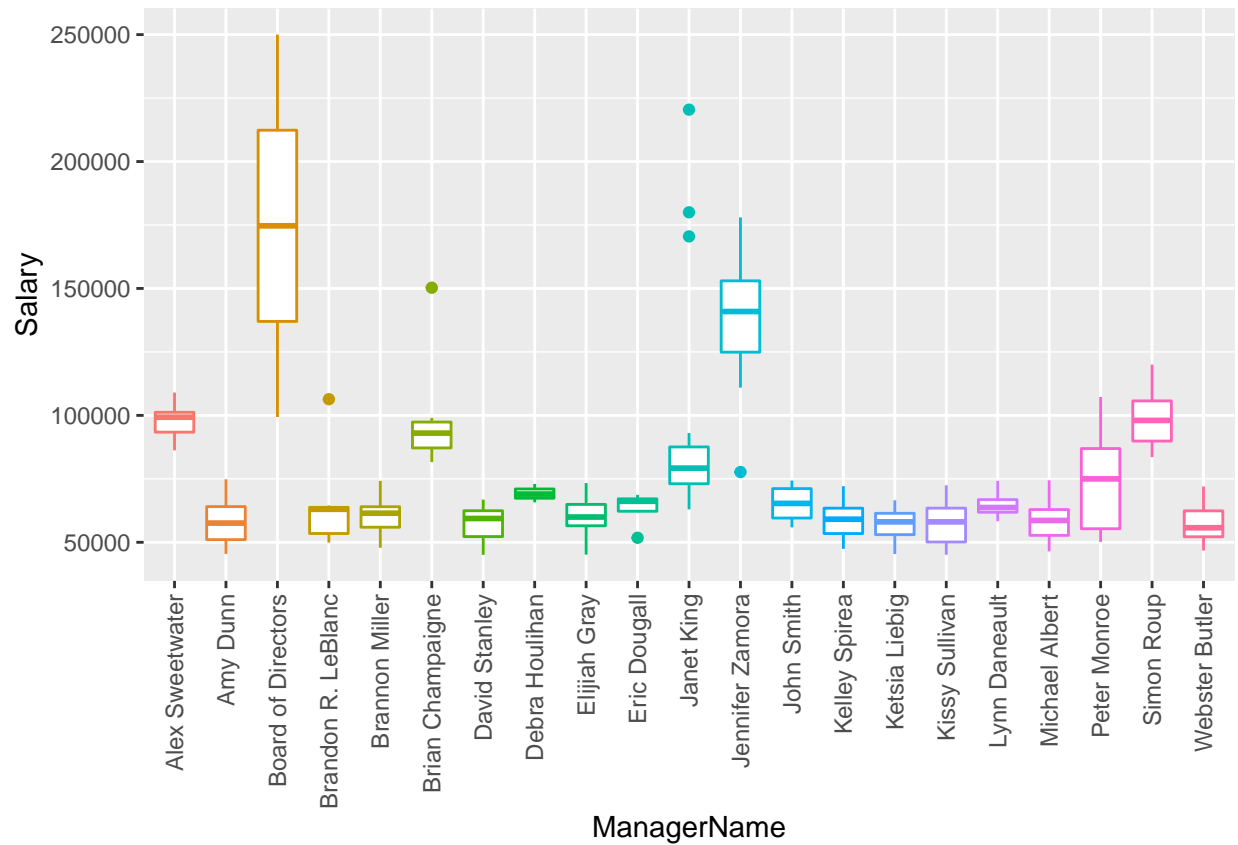
```
ggplot(HR, aes(x=RaceDesc, y=Salary, color=RaceDesc)) +  
  geom_boxplot() + scale_x_discrete(guide = guide_axis(n.dodge=3)) + theme(legend.position="none")
```



```
ggplot(HR, aes(x=HispanicLatino, y=Salary, color=HispanicLatino)) +
  geom_boxplot() + theme(legend.position="none")
```



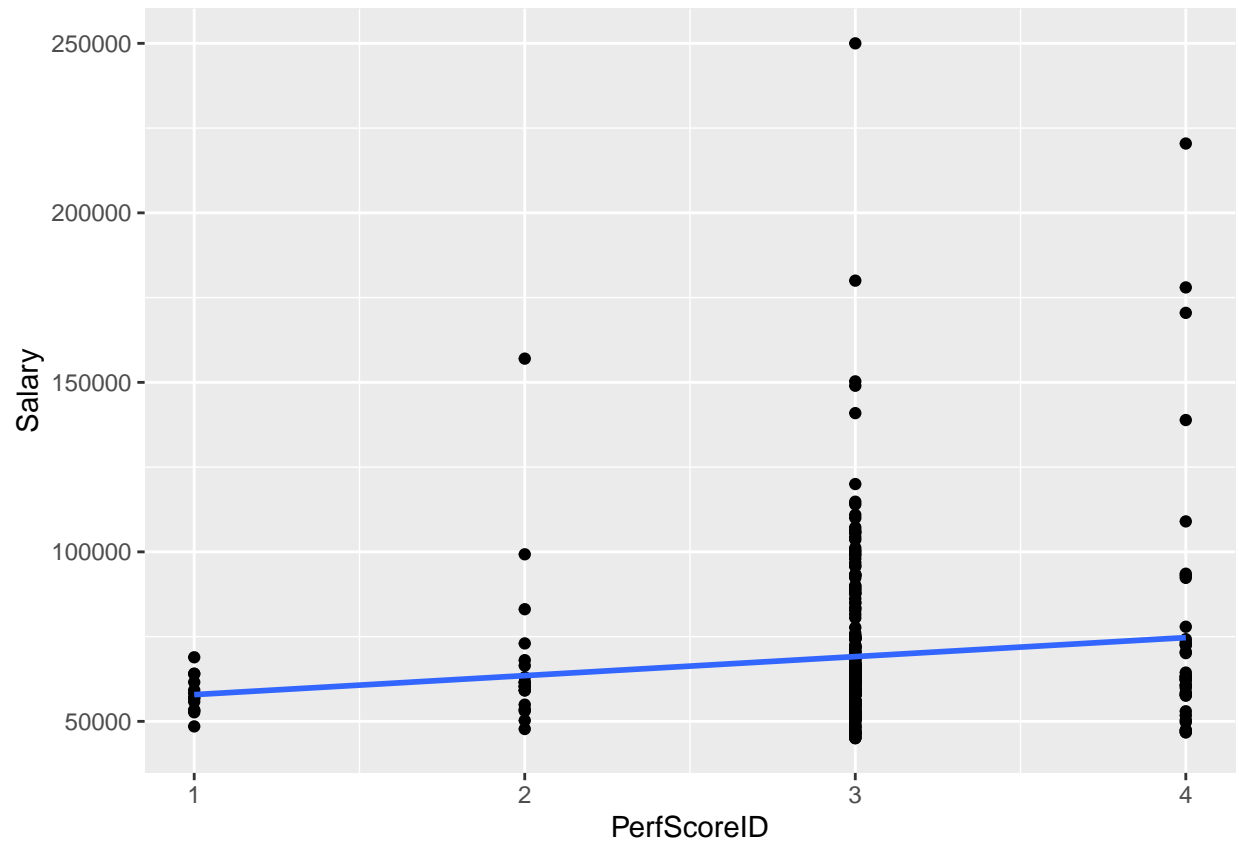
```
ggplot(HR, aes(x=ManagerName, y=Salary, color=ManagerName)) +  
  geom_boxplot()+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) + theme(legend.position = "bottom")
```



## Exploratory Data Analysis - Scatter plots

```
ggplot(HR, aes(x=PerfScoreID, y=Salary)) +
  geom_point() +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

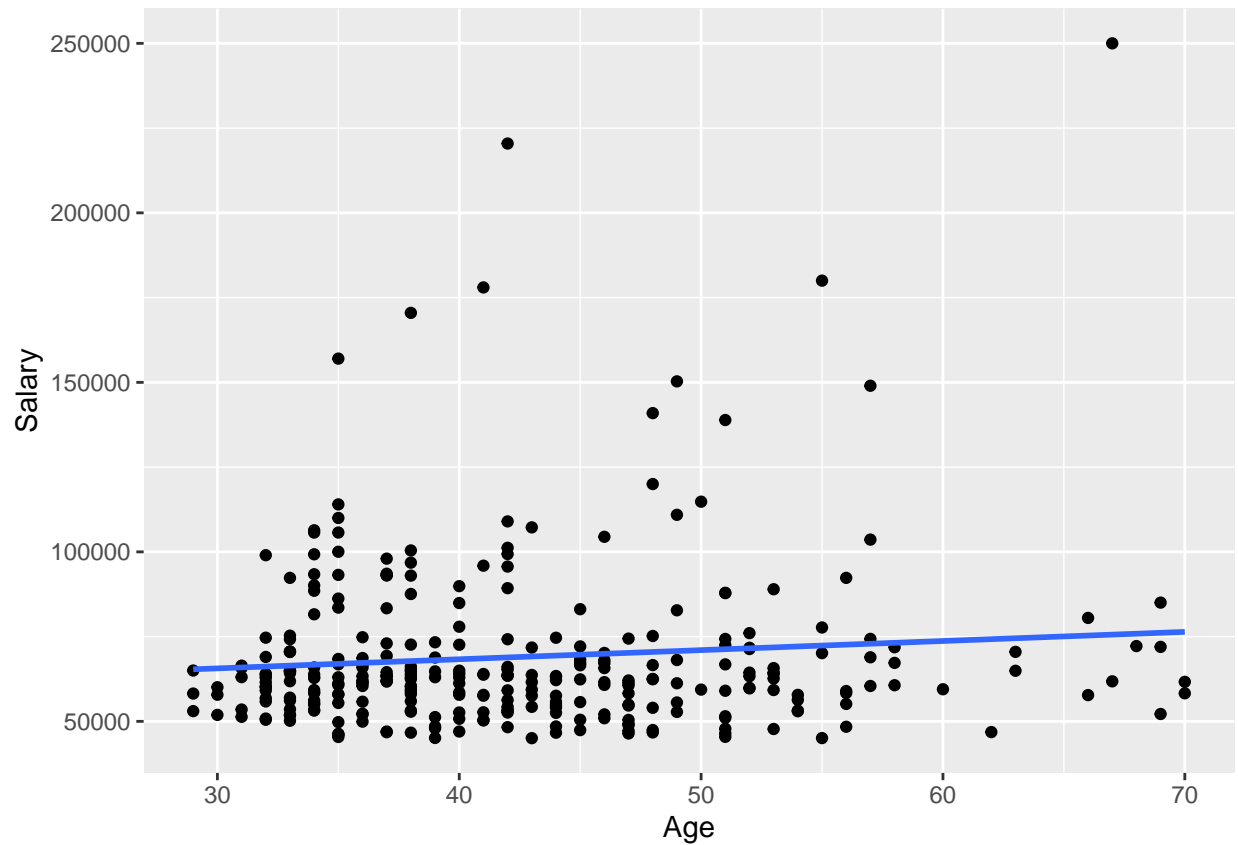
```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(HR, aes(x=Age, y=Salary)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

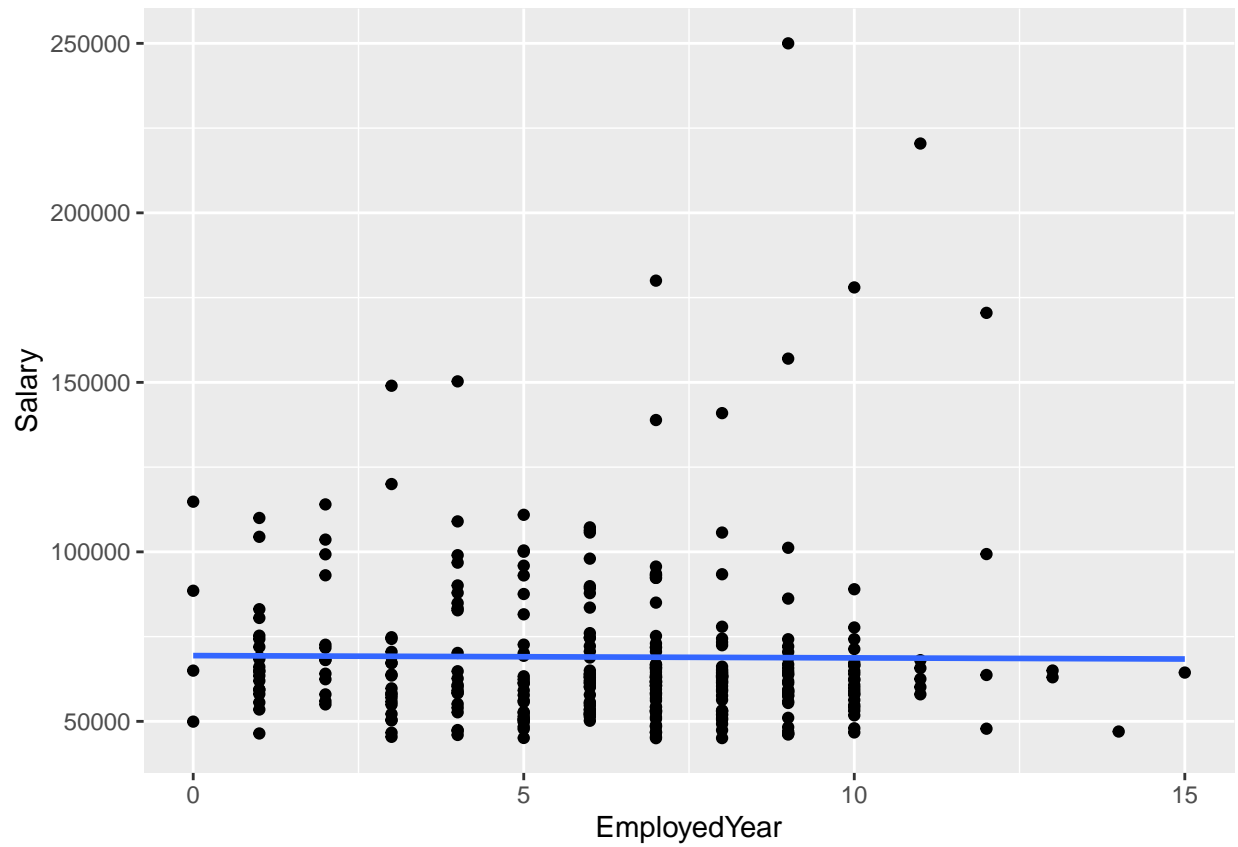
```
## `geom_smooth()` using formula 'y ~ x'
```





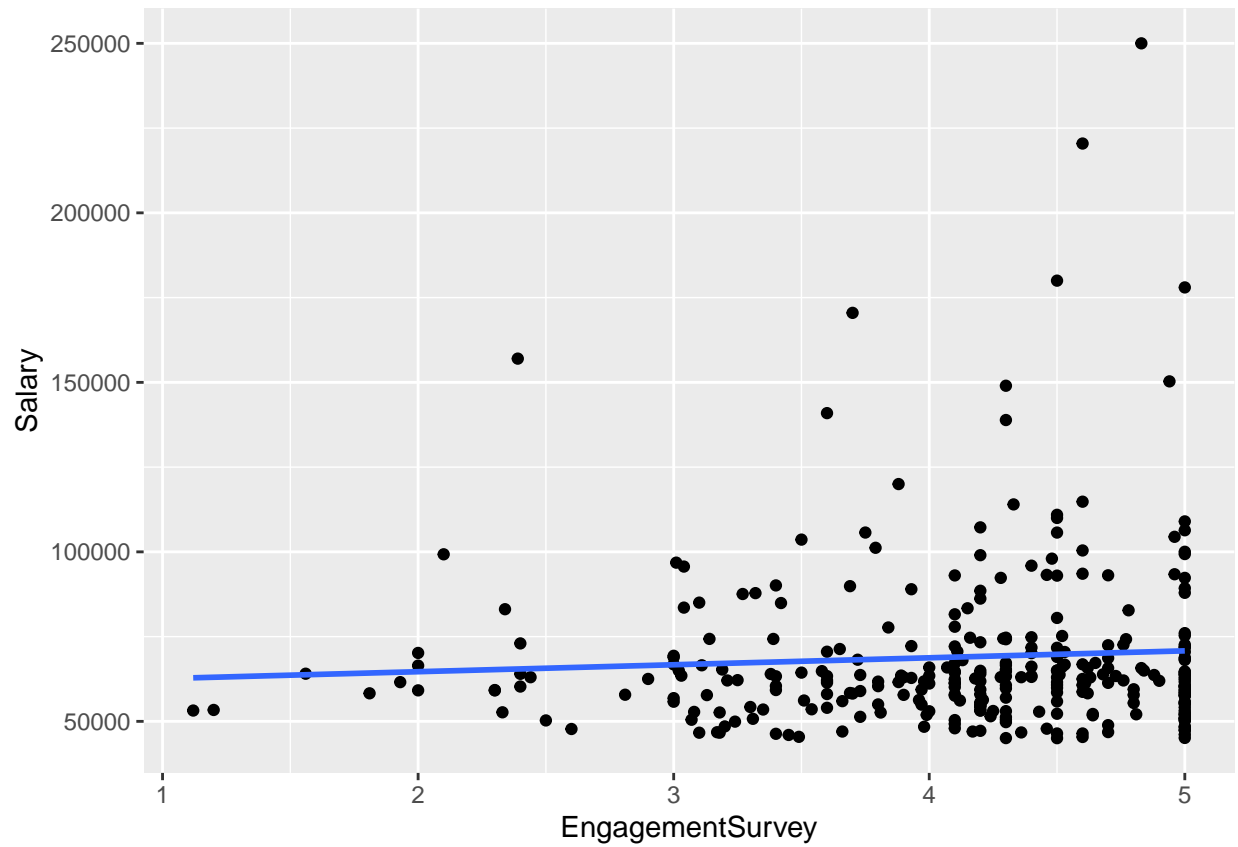
```
ggplot(HR, aes(x=EmployedYear, y=Salary)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



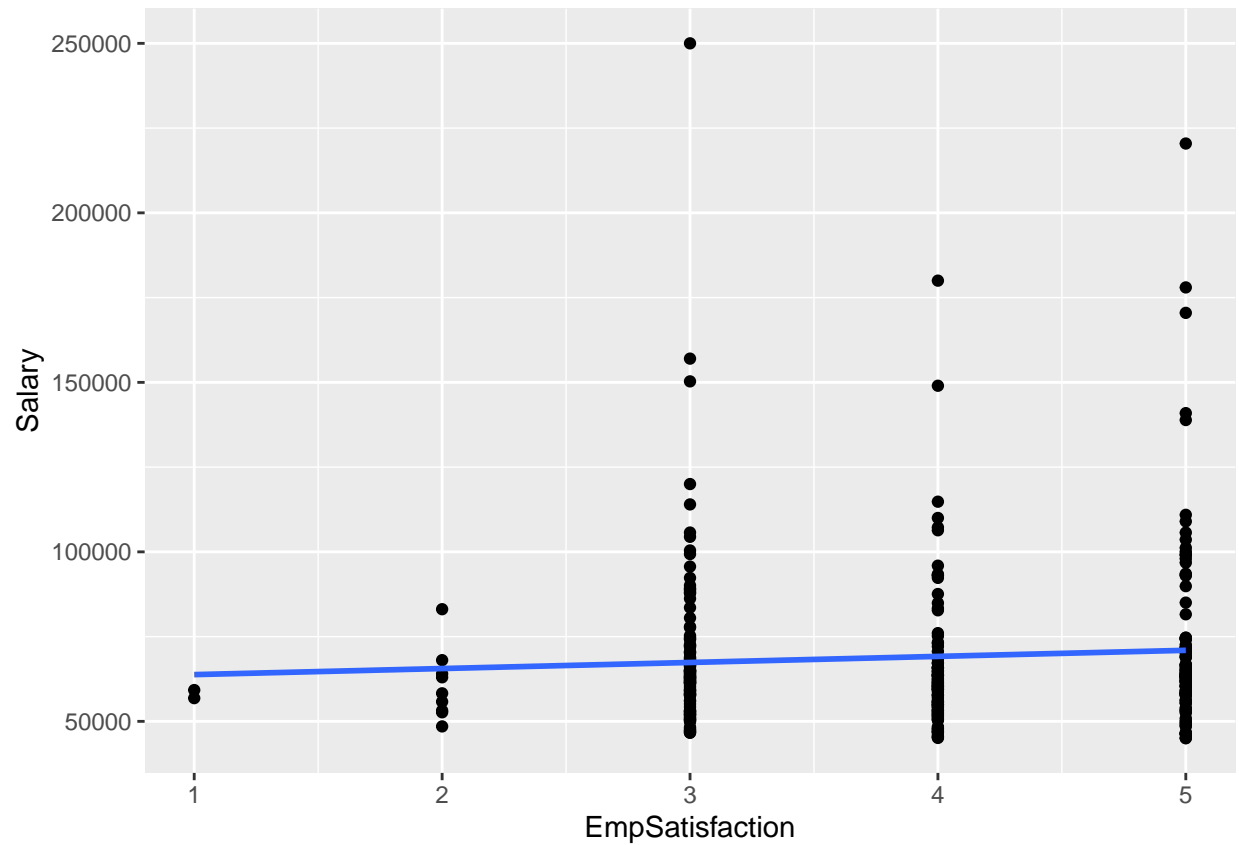
```
ggplot(HR, aes(x=EngagementSurvey, y=Salary)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



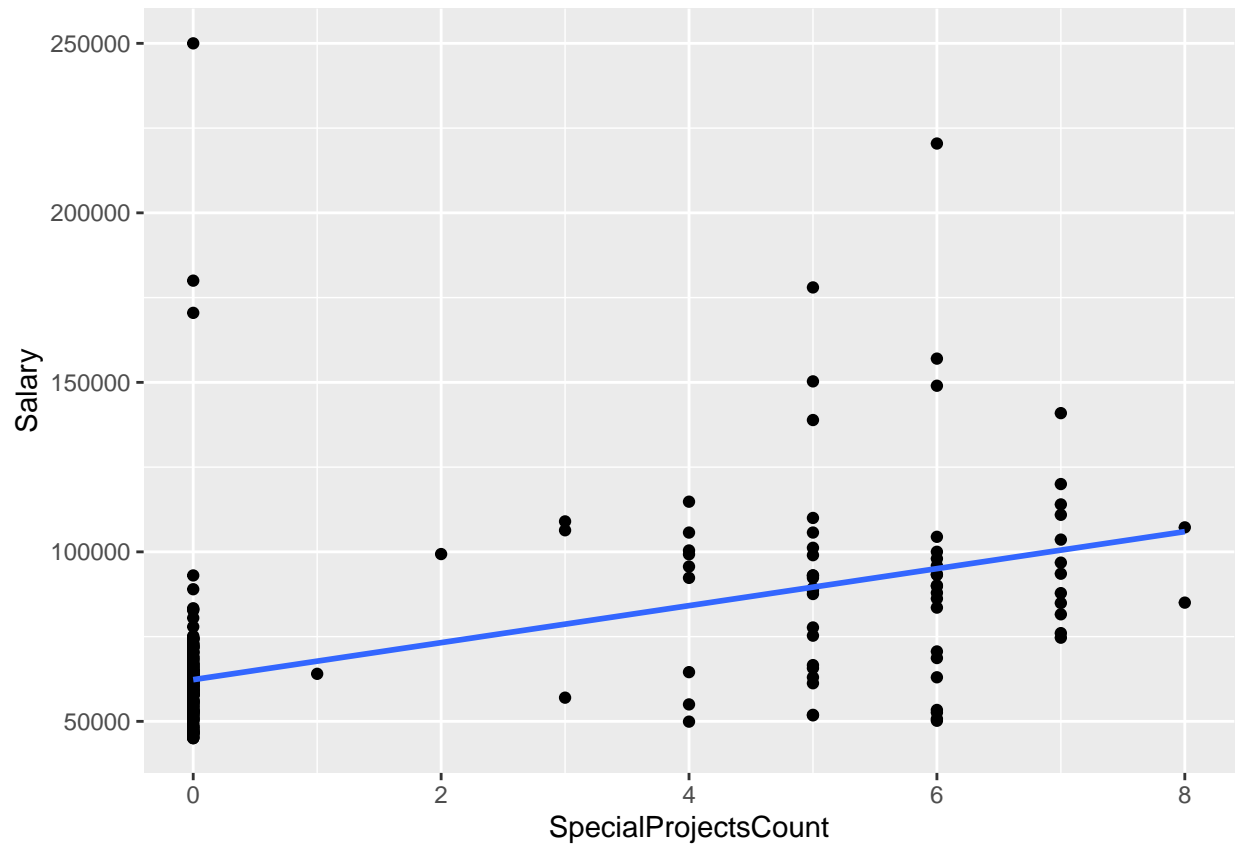
```
ggplot(HR, aes(x=EmpSatisfaction, y=Salary)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



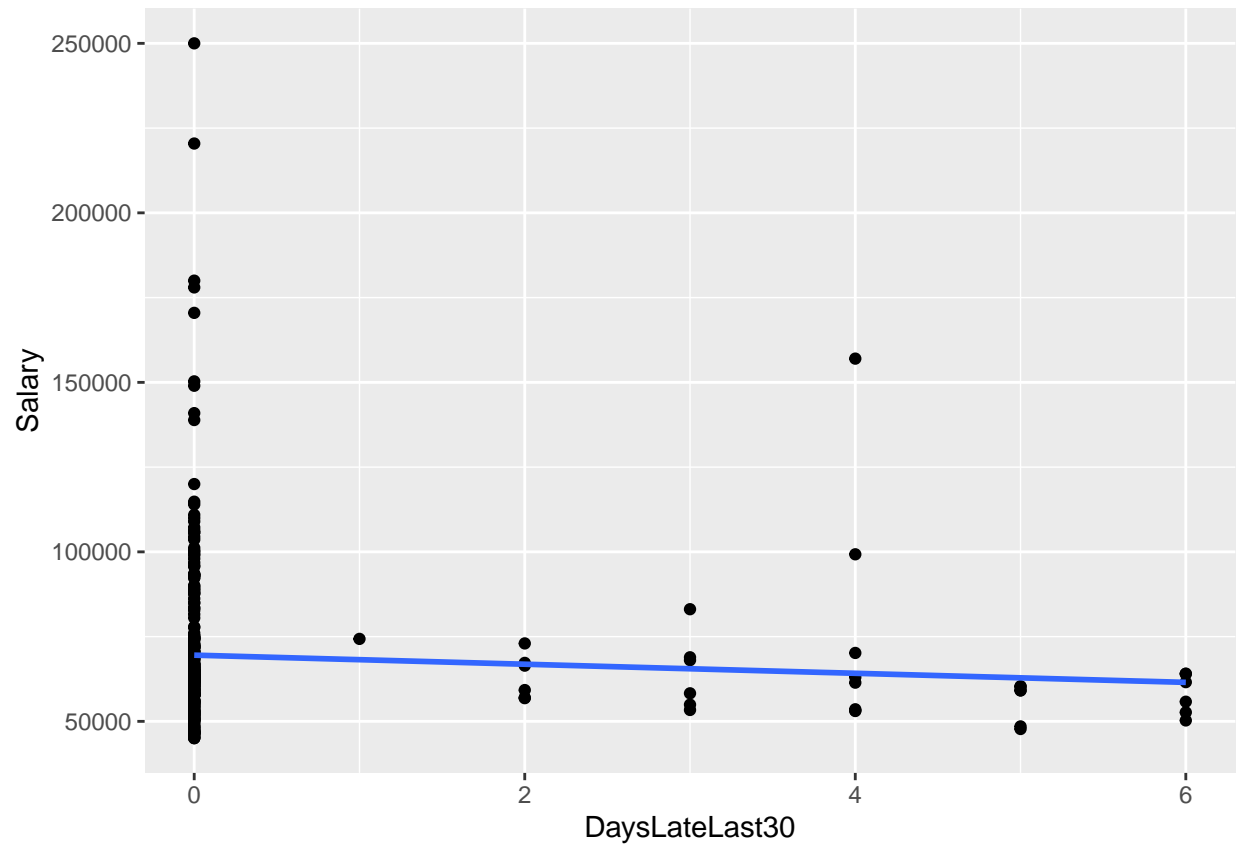
```
ggplot(HR, aes(x=SpecialProjectsCount, y=Salary)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



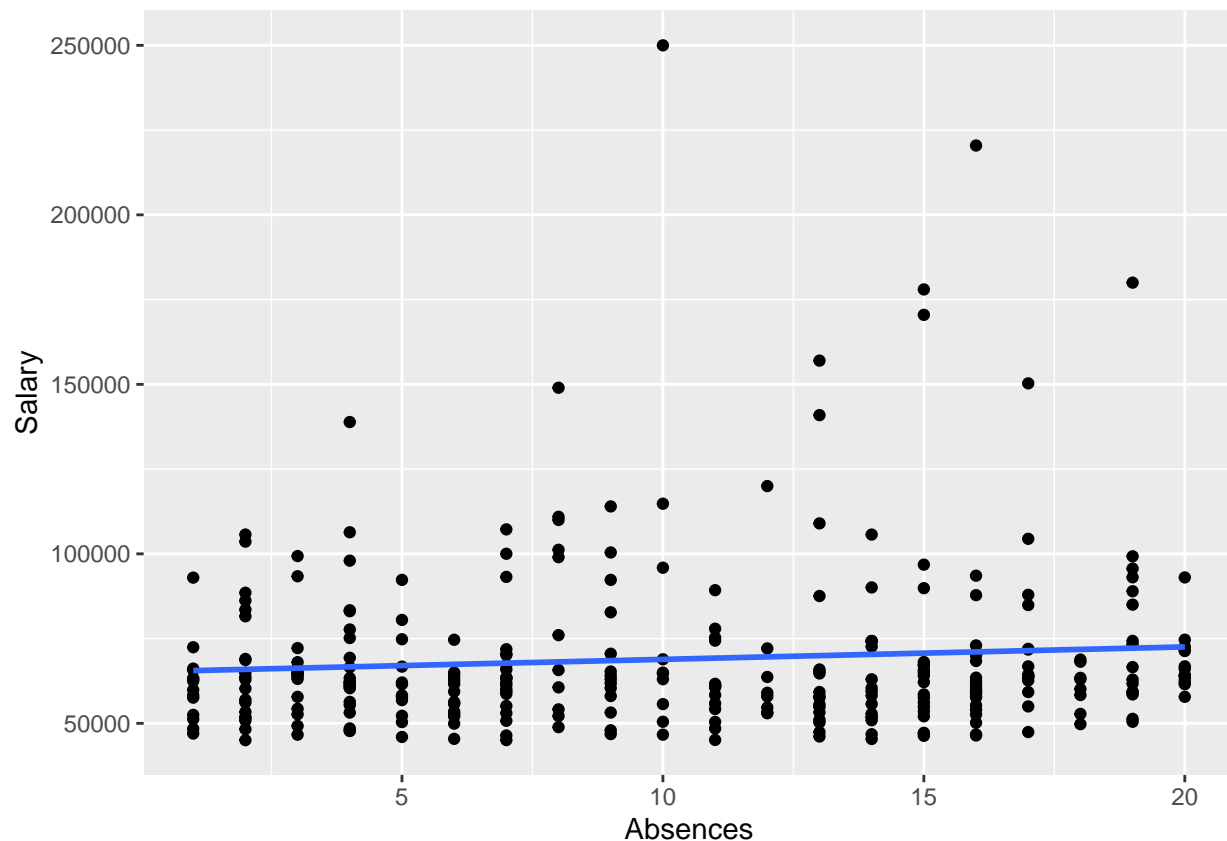
```
ggplot(HR, aes(x=DaysLateLast30, y=Salary)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(HR, aes(x=Absences, y=Salary)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



#### 4. Fitting multiple linear regression model

```
model_full = lm(Salary~., data = HRtrain)
summary(model_full)
```

```
##
## Call:
## lm(formula = Salary ~ ., data = HRtrain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12622  -2976         0    2552   14399
##
## Coefficients: (17 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    55366.43   14407.74   3.843  0.000188
## MarriedID1     -4019.09    1967.99  -2.042  0.043120
## MaritalDescMarried      NA         NA      NA      NA
## MaritalDescSeparated  -4298.44    3076.28  -1.397  0.164672
## MaritalDescSingle    -2257.31    2018.56  -1.118  0.265479
## MaritalDescWidowed     901.80    4024.41   0.224  0.823040
## SexM            -1148.62    1084.22  -1.059  0.291356
## EmploymentStatusTerminated for Cause    2319.14    3046.13   0.761  0.447811
## EmploymentStatusVoluntarily Terminated  -502.85    2025.02  -0.248  0.804272
## DepartmentExecutive Office    188703.22    9811.72  19.232 < 2e-16
```

## DepartmentIT/IS	33519.41	6145.41	5.454	2.34e-07
## DepartmentProduction	-508.20	6757.24	-0.075	0.940163
## DepartmentSales	3319.06	12018.50	0.276	0.782855
## DepartmentSoftware Engineering	18592.05	7588.33	2.450	0.015592
## PerfScoreID	-1297.39	1255.53	-1.033	0.303333
## RecruitmentSourceDiversity Job Fair	-328.25	2822.80	-0.116	0.907603
## RecruitmentSourceEmployee Referral	1853.98	2801.93	0.662	0.509331
## RecruitmentSourceGoogle Search	-1848.36	2341.71	-0.789	0.431341
## RecruitmentSourceIndeed	360.94	2272.72	0.159	0.874060
## RecruitmentSourceLinkedIn	-1137.82	2230.91	-0.510	0.610886
## RecruitmentSourceOn-line Web application	-4900.59	7131.07	-0.687	0.493152
## RecruitmentSourceOther	-6152.01	9416.82	-0.653	0.514699
## RecruitmentSourceWebsite	-66.92	3628.90	-0.018	0.985314
## PositionAdministrative Assistant	-10164.70	5419.19	-1.876	0.062907
## PositionArea Sales Manager	4369.69	8930.83	0.489	0.625454
## PositionBI Developer	353.21	5871.08	0.060	0.952118
## PositionBI Director	17383.47	7750.87	2.243	0.026579
## PositionCIO	134696.72	12165.97	11.072	< 2e-16
## PositionData Analyst	-6705.53	5384.64	-1.245	0.215225
## PositionData Architect	50616.53	8047.09	6.290	4.31e-09
## PositionDatabase Administrator	14623.15	5820.64	2.512	0.013200
## PositionDirector of Operations	117463.69	11659.20	10.075	< 2e-16
## PositionDirector of Sales	130646.14	16100.20	8.115	2.98e-13
## PositionEnterprise Architect	13755.84	8135.56	1.691	0.093229
## PositionIT Director	79393.36	8211.38	9.669	< 2e-16
## PositionIT Manager - DB	48144.48	6483.24	7.426	1.24e-11
## PositionIT Manager - Infra	68587.35	9076.81	7.556	6.17e-12
## PositionIT Manager - Support	42326.11	8094.08	5.229	6.50e-07
## PositionIT Support	-23914.80	5805.62	-4.119	6.67e-05
## PositionNetwork Engineer	-31375.42	5938.09	-5.284	5.09e-07
## PositionPresident & CEO	NA	NA	NA	NA
## PositionPrincipal Data Architect	30791.79	8077.83	3.812	0.000211
## PositionProduction Manager	27171.40	9578.78	2.837	0.005279
## PositionProduction Technician I	-10021.35	1322.85	-7.576	5.56e-12
## PositionProduction Technician II	NA	NA	NA	NA
## PositionSales Manager	NA	NA	NA	NA
## PositionSenior BI Developer	-9511.49	7756.28	-1.226	0.222271
## PositionShared Services Manager	42742.68	11071.79	3.861	0.000176
## PositionSoftware Engineer	15946.97	7113.14	2.242	0.026637
## PositionSoftware Engineering Manager	NA	NA	NA	NA
## PositionSr. Accountant	43071.39	7844.92	5.490	1.98e-07
## PositionSr. DBA	11078.68	7797.73	1.421	0.157745
## PositionSr. Network Engineer	NA	NA	NA	NA
## StateAZ	-4881.38	9260.76	-0.527	0.599006
## StateCA	5854.83	9778.30	0.599	0.550361
## StateCO	-2776.61	9530.70	-0.291	0.771254
## StateCT	-1133.89	8903.97	-0.127	0.898860
## StateFL	773.44	9992.58	0.077	0.938421
## StateGA	-1860.74	9200.97	-0.202	0.840046
## StateID	-2315.86	9888.72	-0.234	0.815199
## StateIN	-8172.85	9669.81	-0.845	0.399533
## StateKY	-6348.05	10171.75	-0.624	0.533648
## StateMA	5384.69	8926.34	0.603	0.547387
## StateME	-12102.07	10259.82	-1.180	0.240297



## StateMT	-9214.54	9789.78	-0.941	0.348301
## StateNC	254.51	10021.32	0.025	0.979777
## StateND	679.11	10299.78	0.066	0.947530
## StateNH	6509.68	10614.81	0.613	0.540757
## StateNV	-4127.29	9989.23	-0.413	0.680149
## StateNY	4809.30	10018.25	0.480	0.631984
## StateOH	-14750.10	11180.00	-1.319	0.189343
## StateOR	-12184.46	9668.69	-1.260	0.209821
## StatePA	9076.84	13126.84	0.691	0.490483
## StateRI	NA	NA	NA	NA
## StateTN	7352.60	9447.74	0.778	0.437821
## StateTX	2556.98	9640.81	0.265	0.791250
## StateUT	3666.18	10286.15	0.356	0.722096
## StateVA	8694.92	10445.44	0.832	0.406679
## StateVT	2346.94	9637.00	0.244	0.807970
## StateWA	-3867.80	10642.73	-0.363	0.716871
## Age	36.32	65.28	0.556	0.578936
## CitizenDescNon-Citizen	4423.59	5897.54	0.750	0.454545
## CitizenDescUS Citizen	435.22	3235.64	0.135	0.893205
## RaceDescAsian	1949.97	5236.97	0.372	0.710231
## RaceDescBlack or African American	2680.00	5098.22	0.526	0.599996
## RaceDescTwo or more races	-1761.26	5667.60	-0.311	0.756474
## RaceDescWhite	713.04	5002.67	0.143	0.886878
## HispanicLatinoYes	-3794.63	2096.07	-1.810	0.072515
## EmployedYear	118.82	330.62	0.359	0.719879
## ManagerNameAmy Dunn	4234.03	2527.19	1.675	0.096224
## ManagerNameBoard of Directors	NA	NA	NA	NA
## ManagerNameBrandon R. LeBlanc	NA	NA	NA	NA
## ManagerNameBrannon Miller	4199.45	2483.28	1.691	0.093179
## ManagerNameBrian Champaigne	NA	NA	NA	NA
## ManagerNameDavid Stanley	2549.37	2597.93	0.981	0.328237
## ManagerNameDebra Houlihan	NA	NA	NA	NA
## ManagerNameElijah Gray	3812.96	2601.61	1.466	0.145130
## ManagerNameEric Dougall	-5784.19	6324.27	-0.915	0.362068
## ManagerNameJanet King	-12982.35	9028.69	-1.438	0.152829
## ManagerNameJennifer Zamora	NA	NA	NA	NA
## ManagerNameJohn Smith	NA	NA	NA	NA
## ManagerNameKelley Spirea	2660.81	2431.12	1.094	0.275738
## ManagerNameKetsia Liebig	-483.80	2489.93	-0.194	0.846237
## ManagerNameKissy Sullivan	2967.80	2333.70	1.272	0.205710
## ManagerNameLynn Daneault	NA	NA	NA	NA
## ManagerNameMichael Albert	2016.44	2696.38	0.748	0.455891
## ManagerNamePeter Monroe	NA	NA	NA	NA
## ManagerNameSimon Roup	NA	NA	NA	NA
## ManagerNameWebster Butler	NA	NA	NA	NA
## EngagementSurvey	863.88	819.21	1.055	0.293572
## EmpSatisfaction	584.00	557.75	1.047	0.296985
## SpecialProjectsCount	-486.80	1017.48	-0.478	0.633134
## DaysLateLast30	-399.66	604.46	-0.661	0.509645
## Absences	131.52	90.05	1.461	0.146519
##				
## (Intercept)	***			
## MarriedID1	*			
## MaritalDescMarried				

```

## MaritalDescSeparated
## MaritalDescSingle
## MaritalDescWidowed
## SexM
## EmploymentStatusTerminated for Cause
## EmploymentStatusVoluntarily Terminated
## DepartmentExecutive Office      ***
## DepartmentIT/IS                 ***
## DepartmentProduction
## DepartmentSales
## DepartmentSoftware Engineering  *
## PerfScoreID
## RecruitmentSourceDiversity Job Fair
## RecruitmentSourceEmployee Referral
## RecruitmentSourceGoogle Search
## RecruitmentSourceIndeed
## RecruitmentSourceLinkedIn
## RecruitmentSourceOn-line Web application
## RecruitmentSourceOther
## RecruitmentSourceWebsite
## PositionAdministrative Assistant .
## PositionArea Sales Manager
## PositionBI Developer
## PositionBI Director             *
## PositionCIO                     ***
## PositionData Analyst
## PositionData Architect          ***
## PositionDatabase Administrator  *
## PositionDirector of Operations  ***
## PositionDirector of Sales       ***
## PositionEnterprise Architect    .
## PositionIT Director             ***
## PositionIT Manager - DB         ***
## PositionIT Manager - Infra      ***
## PositionIT Manager - Support    ***
## PositionIT Support              ***
## PositionNetwork Engineer        ***
## PositionPresident & CEO
## PositionPrincipal Data Architect ***
## PositionProduction Manager      **
## PositionProduction Technician I  ***
## PositionProduction Technician II
## PositionSales Manager
## PositionSenior BI Developer
## PositionShared Services Manager ***
## PositionSoftware Engineer       *
## PositionSoftware Engineering Manager
## PositionSr. Accountant          ***
## PositionSr. DBA
## PositionSr. Network Engineer
## StateAZ
## StateCA
## StateCO
## StateCT

```

```

## StateFL
## StateGA
## StateID
## StateIN
## StateKY
## StateMA
## StateME
## StateMT
## StateNC
## StateND
## StateNH
## StateNV
## StateNY
## StateOH
## StateOR
## StatePA
## StateRI
## StateTN
## StateTX
## StateUT
## StateVA
## StateVT
## StateWA
## Age
## CitizenDescNon-Citizen
## CitizenDescUS Citizen
## RaceDescAsian
## RaceDescBlack or African American
## RaceDescTwo or more races
## RaceDescWhite
## HispanicLatinoYes .
## EmployedYear
## ManagerNameAmy Dunn .
## ManagerNameBoard of Directors
## ManagerNameBrandon R. LeBlanc
## ManagerNameBrannon Miller .
## ManagerNameBrian Champaigne
## ManagerNameDavid Stanley
## ManagerNameDebra Houlihan
## ManagerNameElijah Gray
## ManagerNameEric Dougall
## ManagerNameJanet King
## ManagerNameJennifer Zamora
## ManagerNameJohn Smith
## ManagerNameKelley Spirea
## ManagerNameKetsia Liebig
## ManagerNameKissy Sullivan
## ManagerNameLynn Daneault
## ManagerNameMichael Albert
## ManagerNamePeter Monroe
## ManagerNameSimon Roup
## ManagerNameWebster Butler
## EngagementSurvey
## EmpSatisfaction

```

```
## SpecialProjectsCount
## DaysLateLast30
## Absences
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6170 on 132 degrees of freedom
## Multiple R-squared:  0.9712, Adjusted R-squared:  0.9503
## F-statistic: 46.4 on 96 and 132 DF,  p-value: < 2.2e-16
```

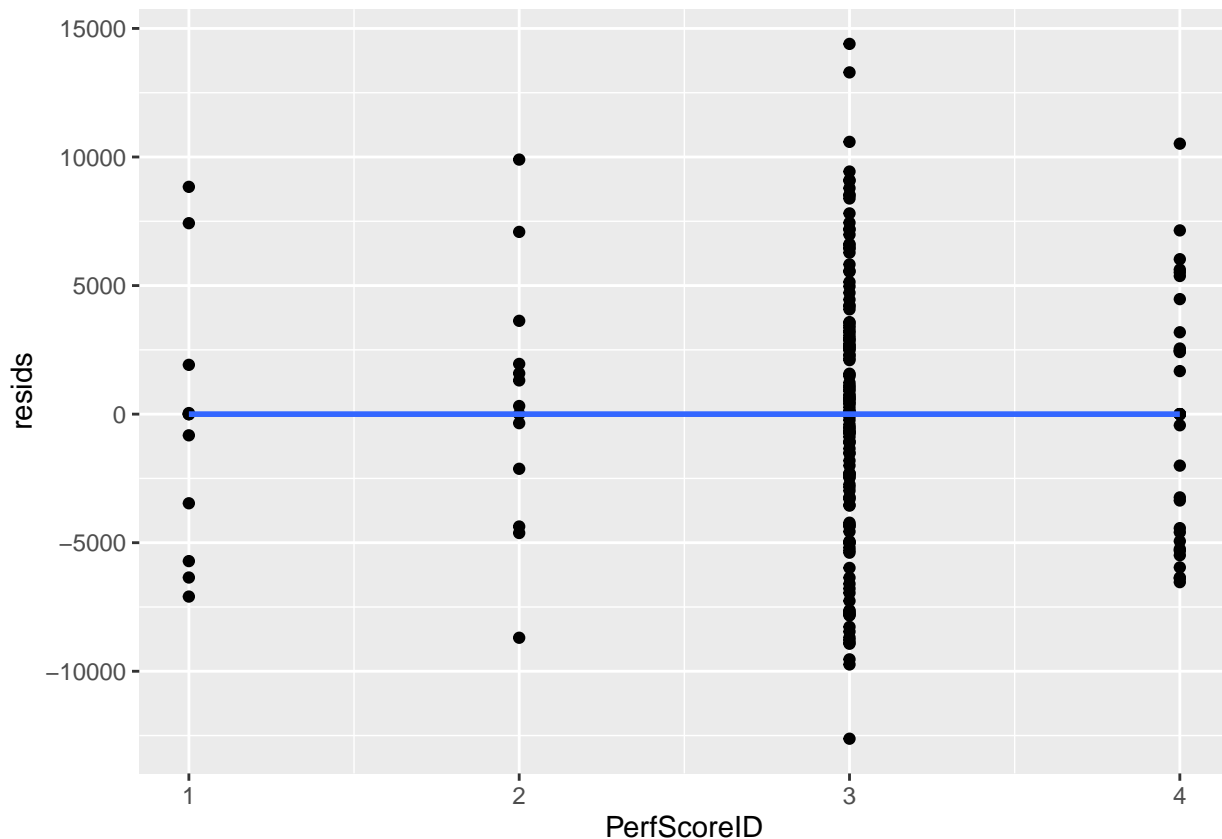
## 5. Testing model assumptions

```
#plot(Salary~PerfScoreID+Age+EmployedYear+EngagementSurvey+EmpSatisfaction+SpecialProjectsCount+DaysLateLast30)
resids = model_full$residuals
```

```
# Linearity
par(mfrow=c(2,2))

ggplot(HRtrain, aes(x=PerfScoreID, y=resids)) +
  geom_point() +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

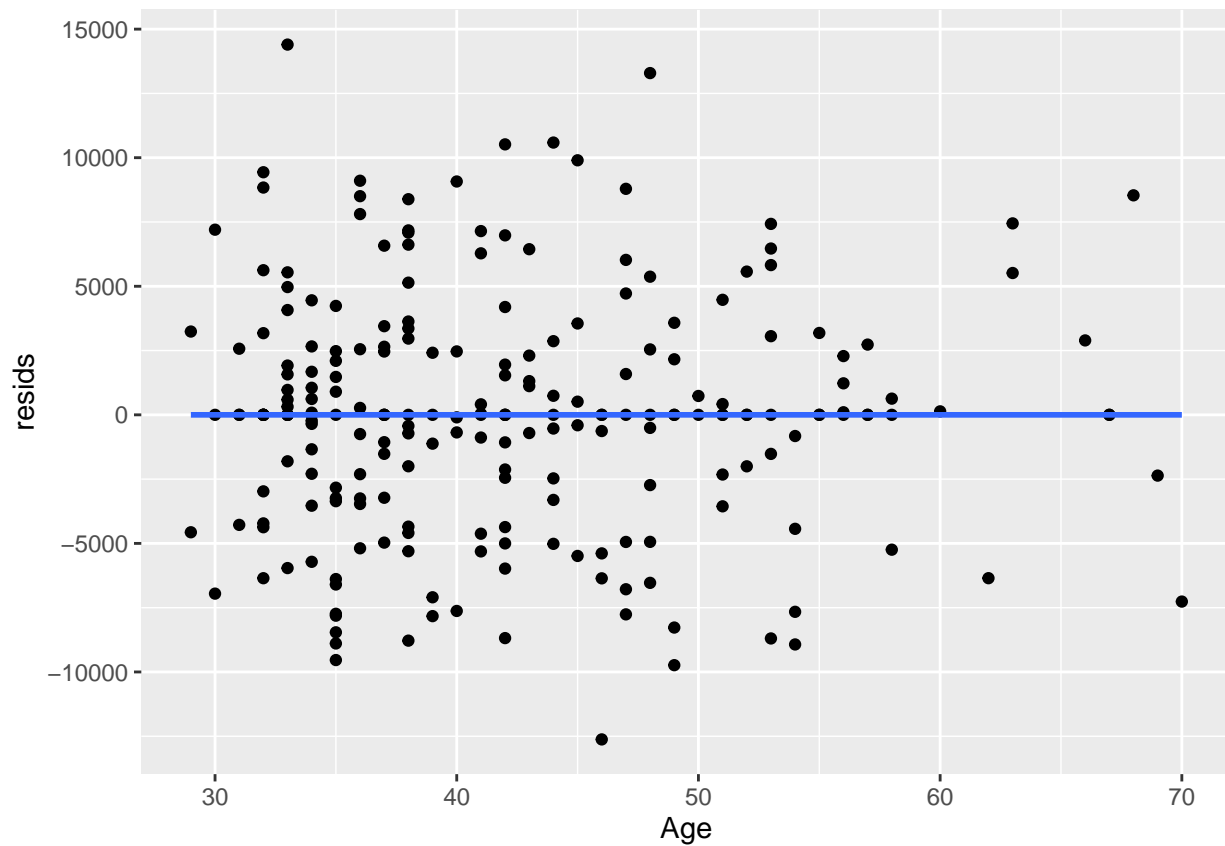
```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(HRtrain, aes(x=Age, y=resids)) +
  geom_point() +
```

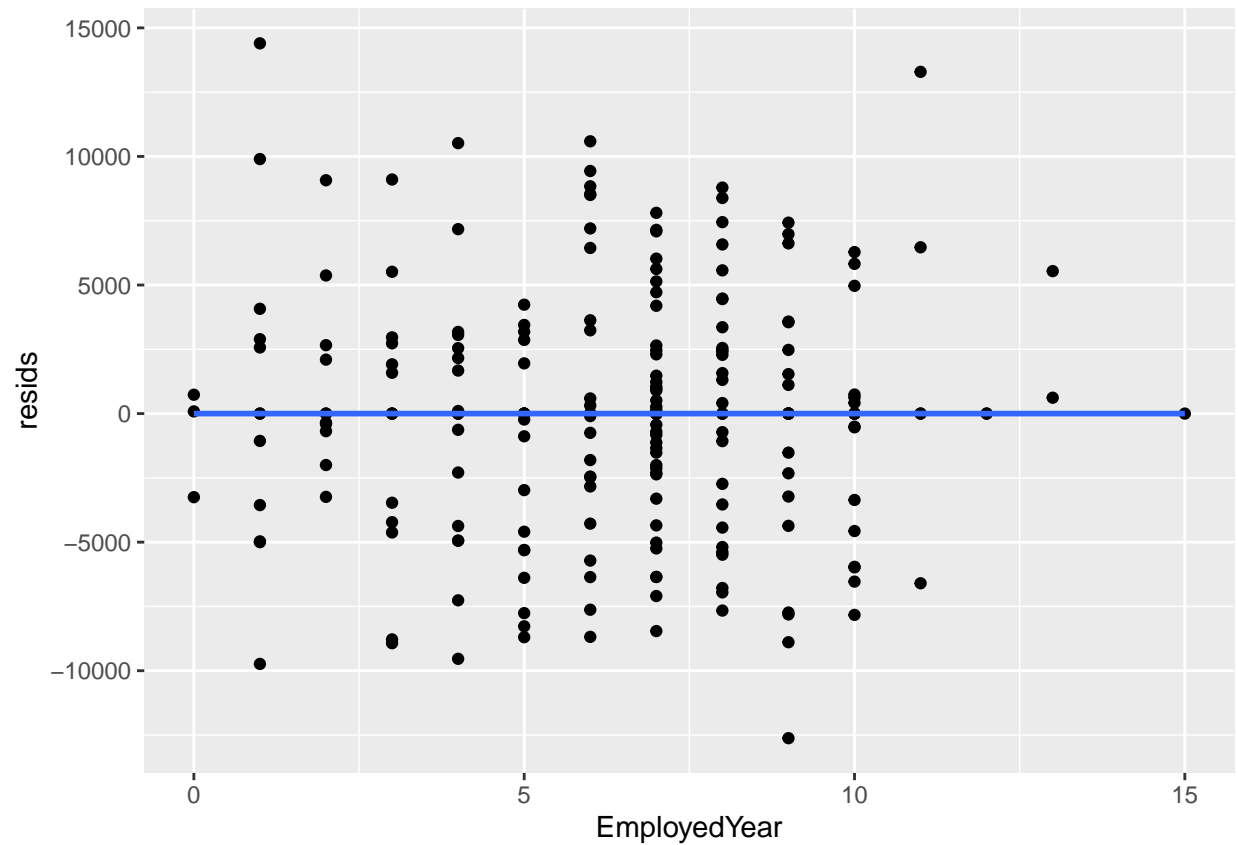
```
geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



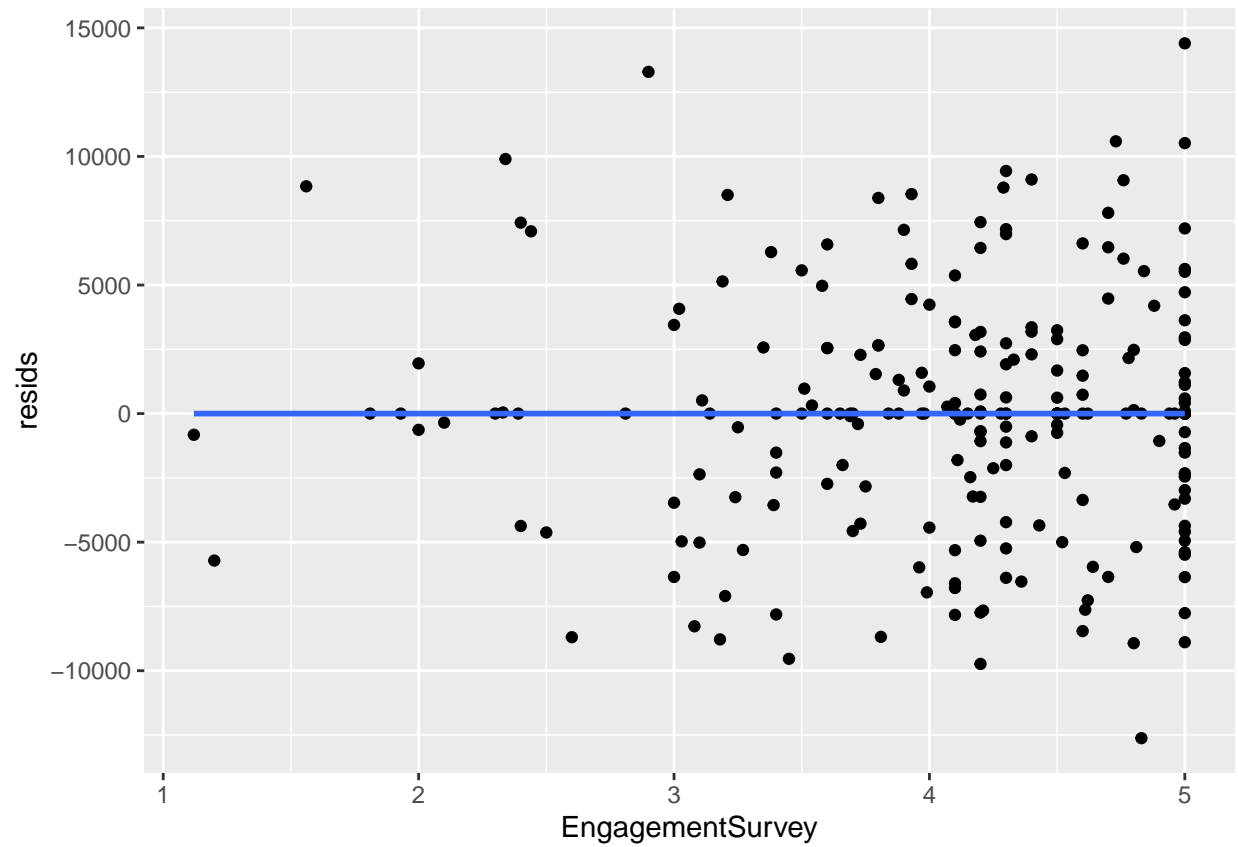
```
ggplot(HRtrain, aes(x=EmployedYear, y=resids)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



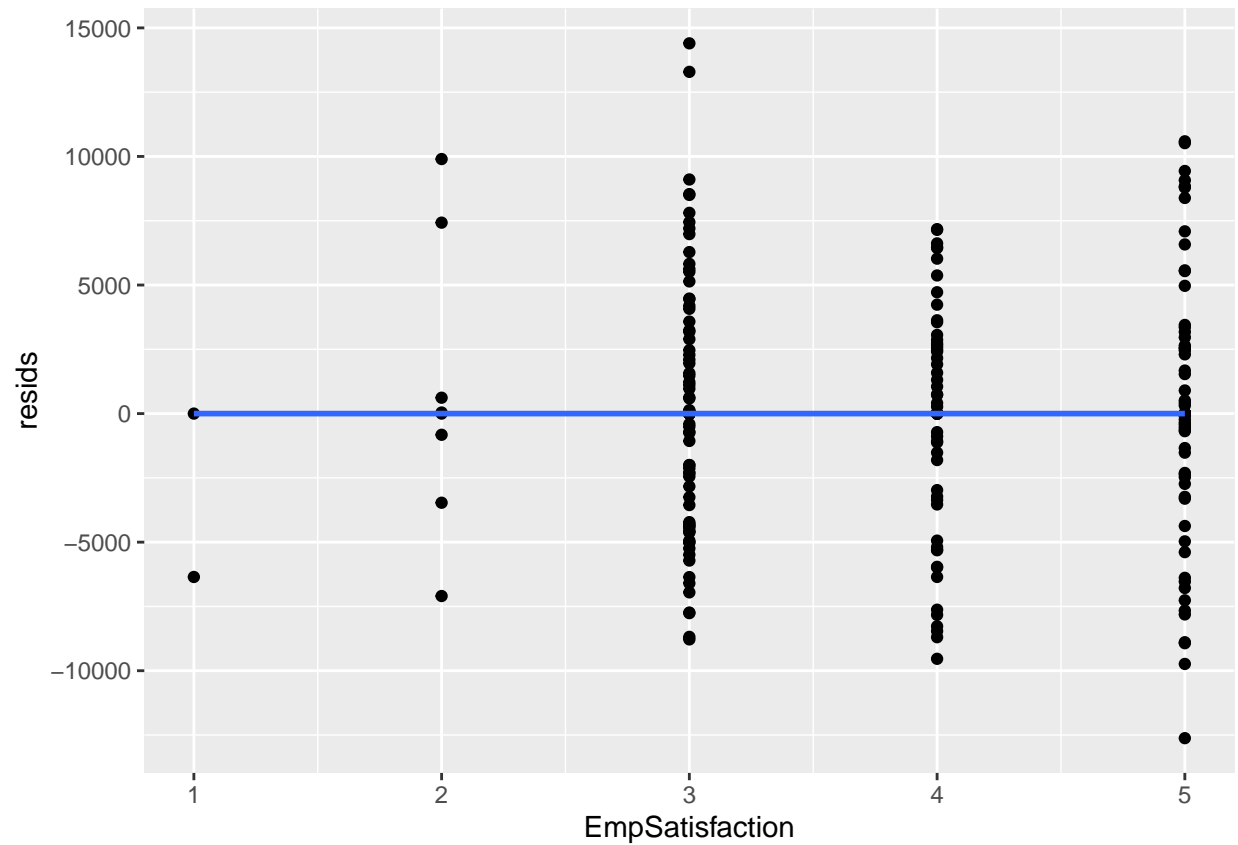
```
ggplot(HRtrain, aes(x=EngagementSurvey, y=resids)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(HRtrain, aes(x=EmpSatisfaction, y=resids)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

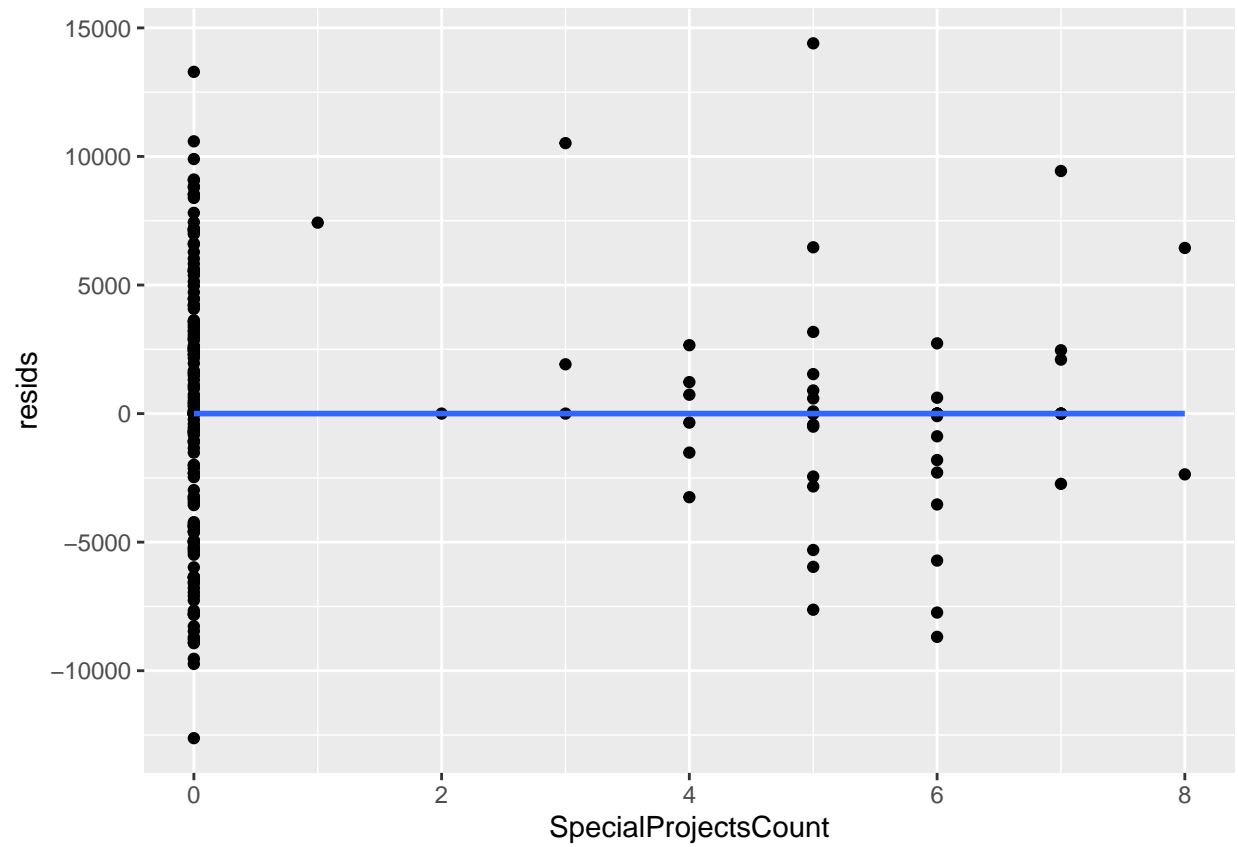
```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(HRtrain, aes(x=SpecialProjectsCount, y=resids)) +
  geom_point() +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

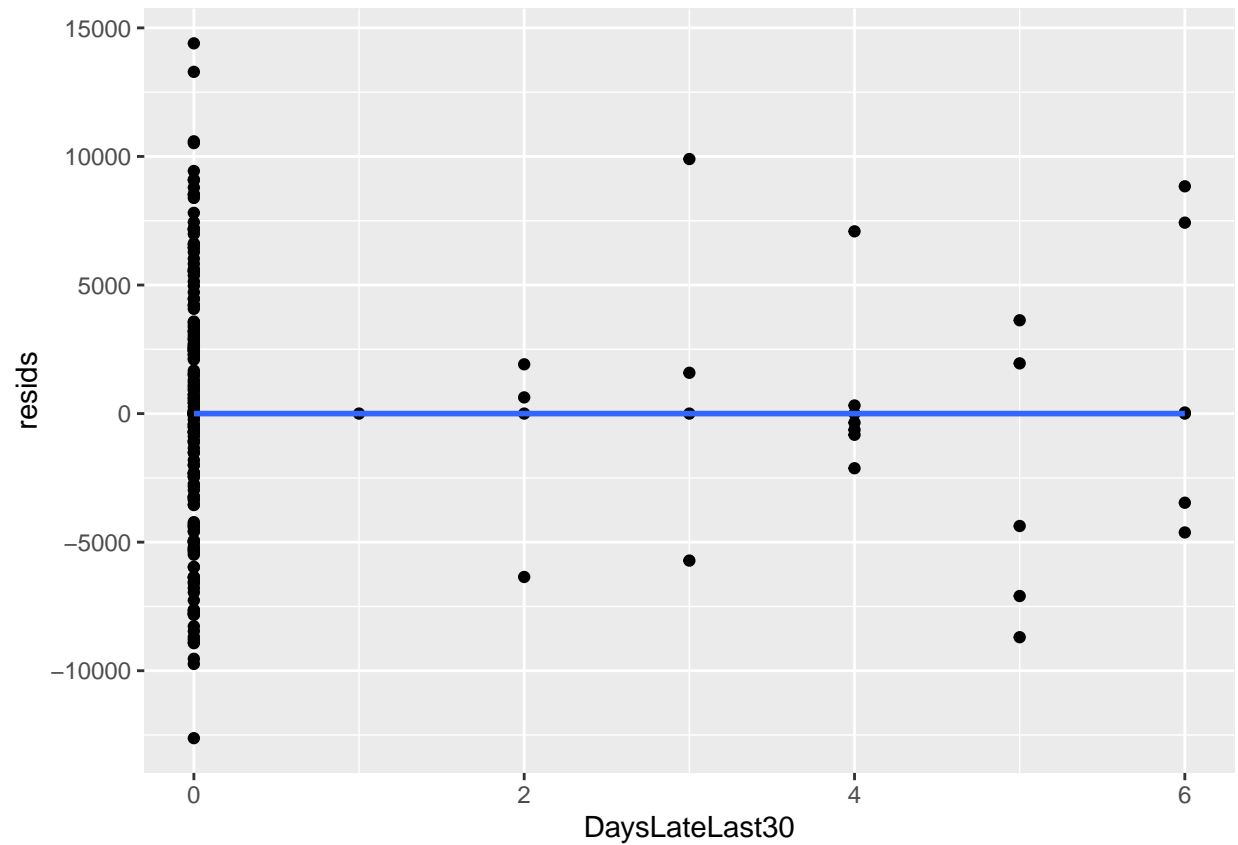
```
## `geom_smooth()` using formula 'y ~ x'
```





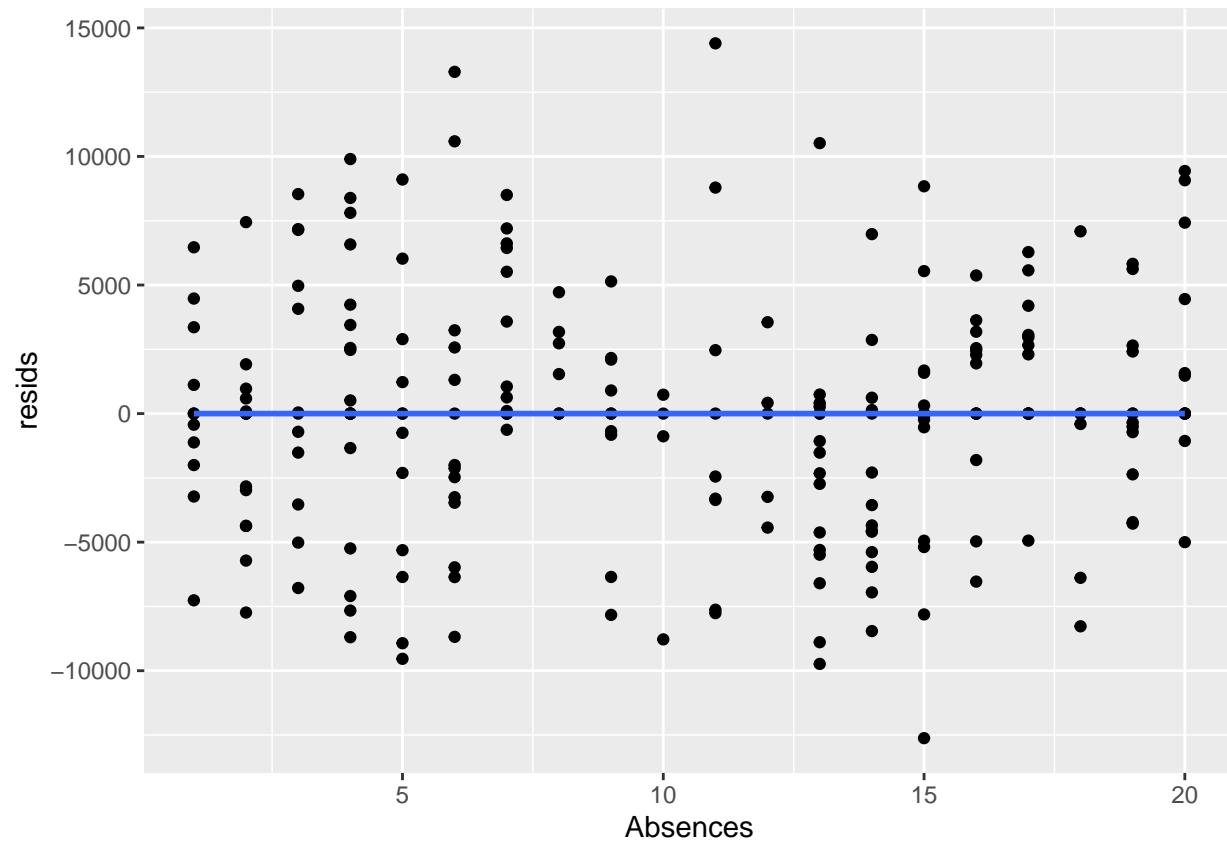
```
ggplot(HRtrain, aes(x=DaysLateLast30, y=resids)) +  
  geom_point() +  
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(HRtrain, aes(x=Absences, y=resids)) +
  geom_point() +
  geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
```

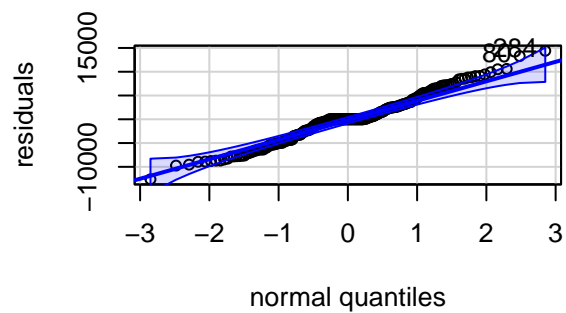
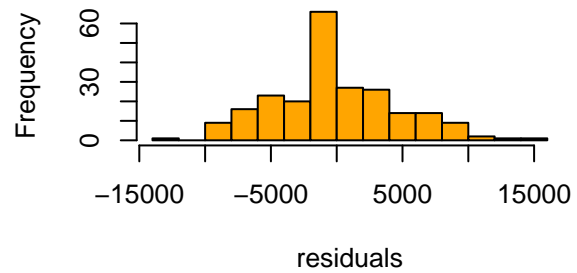
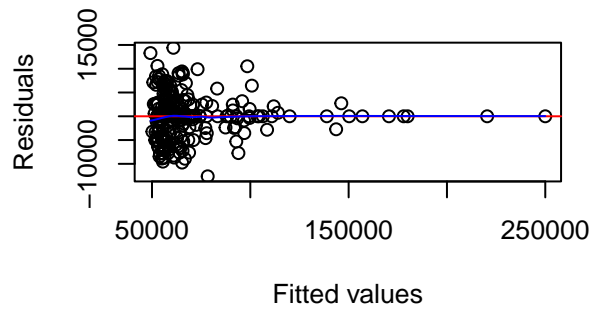
```
## `geom_smooth()` using formula 'y ~ x'
```



```
# Constant Variance
plot(fitted(model_full),resids,xlab="Fitted values",ylab="Residuals")
abline(h=0, col="red")
lines(lowess(model_full$fitted.values, resids), col='blue')

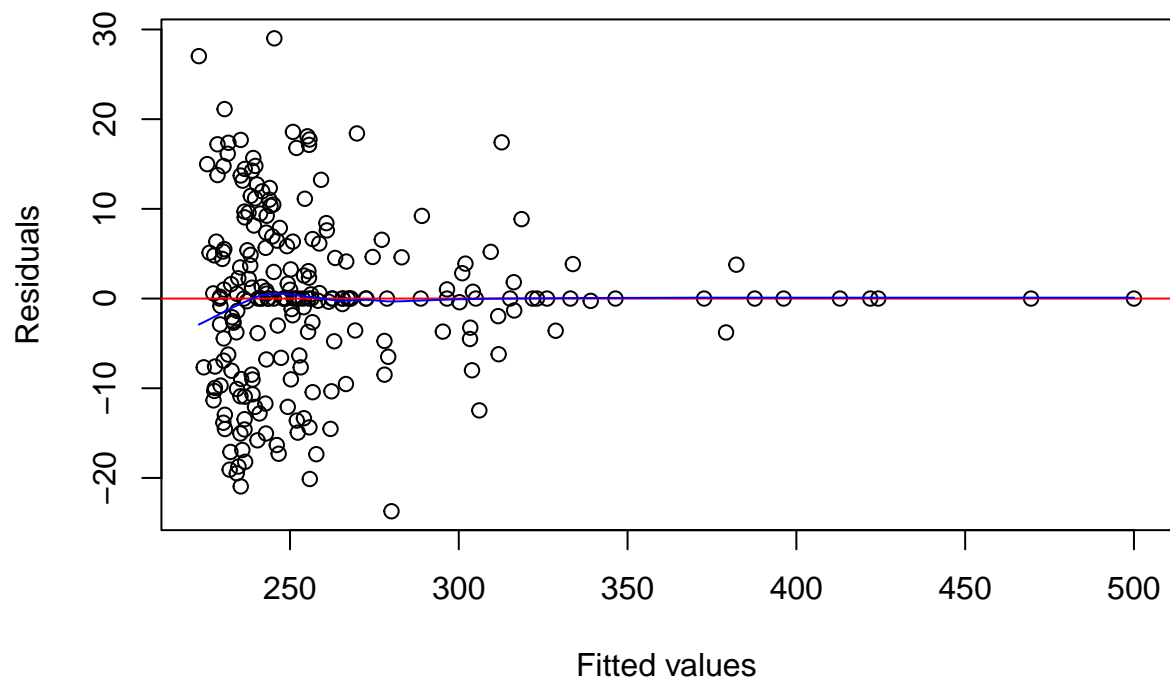
# Normality
hist(resids, xlab="residuals", col="orange",main=NULL, nclass=15)
qqPlot(resids, xlab="normal quantiles", ylab="residuals")
```

```
## 284 80
## 201 56
```

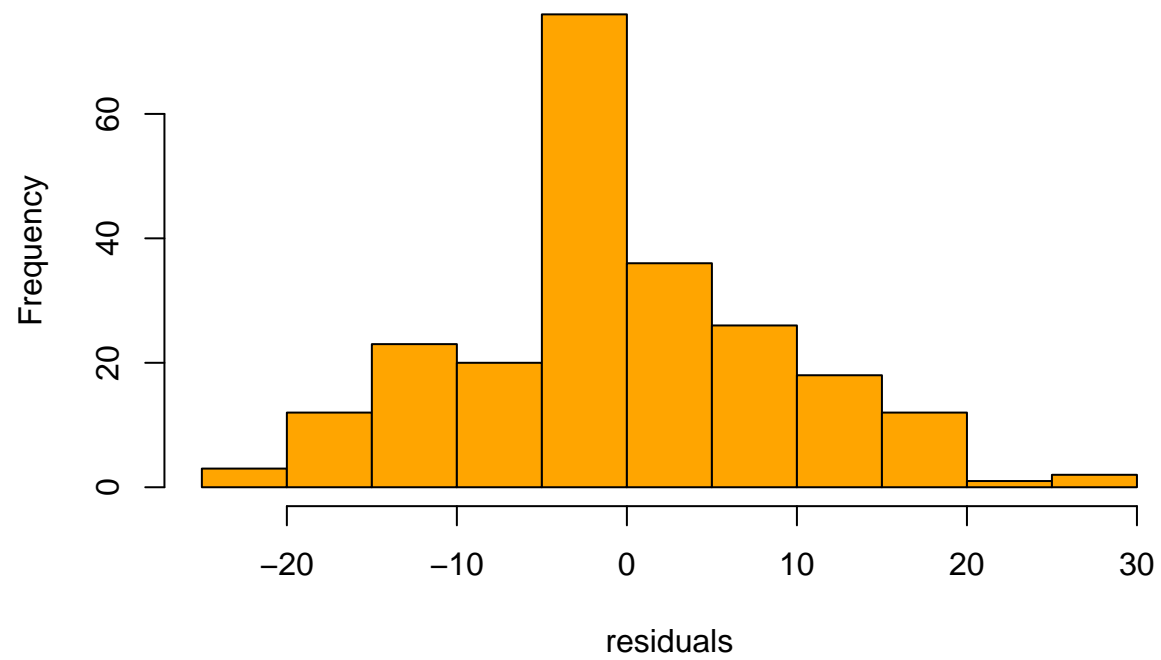


## Transformation Try Outs

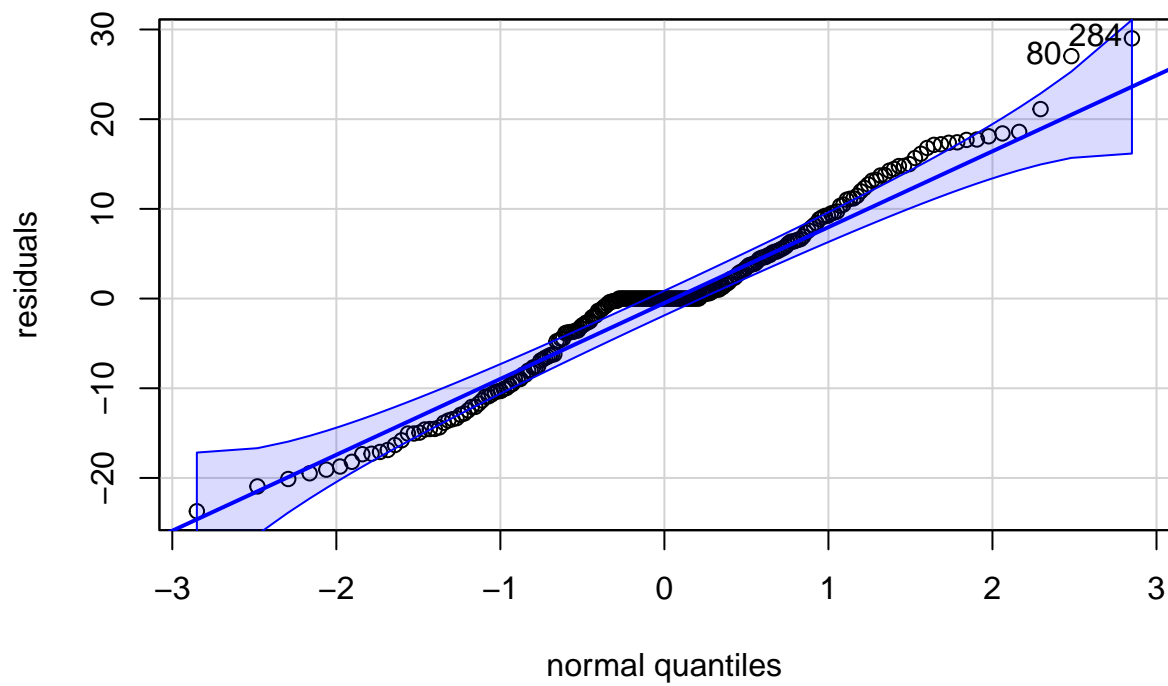
```
# Sqrt Transformation
model_full_tr = lm(sqrt(Salary)~., data = HRtrain)
# Variance
resids_tr = model_full_tr$residuals
plot(fitted(model_full_tr),resids_tr,xlab="Fitted values",ylab="Residuals")
abline(h=0, col="red")
lines(lowess(model_full_tr$fitted.values, resids_tr), col='blue')
```



```
# Normality  
hist(resids_tr, xlab="residuals", col="orange", main=NULL, nclass=15)
```

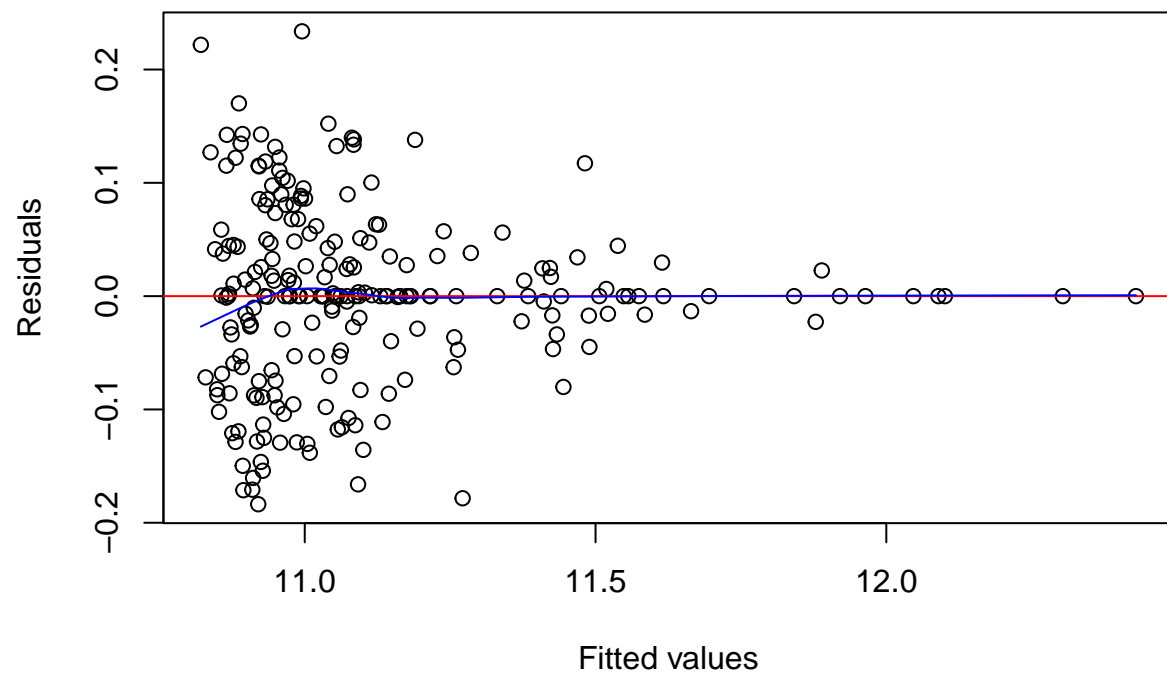


```
qqPlot(resids_tr, xlab="normal quantiles", ylab="residuals")
```



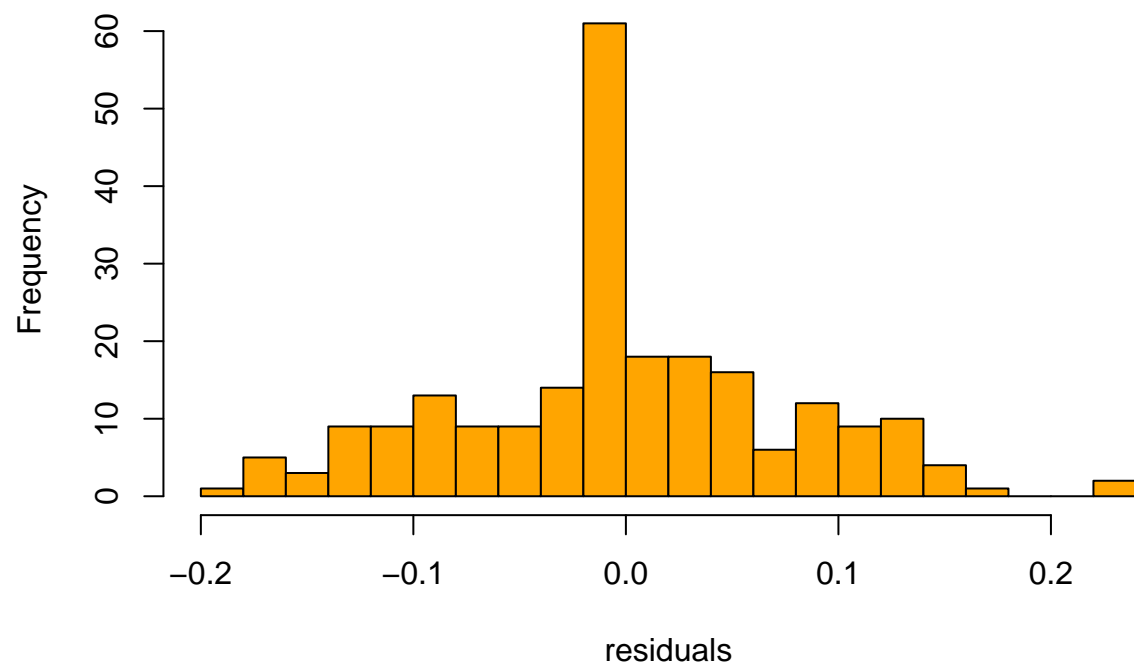
```
## 284 80
## 201 56

# Log Transformation
model_full_log = lm(log(Salary)~., data = HRtrain)
# Variance
resids_log = model_full_log$residuals
plot(fitted(model_full_log),resids_log,xlab="Fitted values",ylab="Residuals")
abline(h=0, col="red")
lines(lowess(model_full_log$fitted.values, resids_log), col='blue')
```

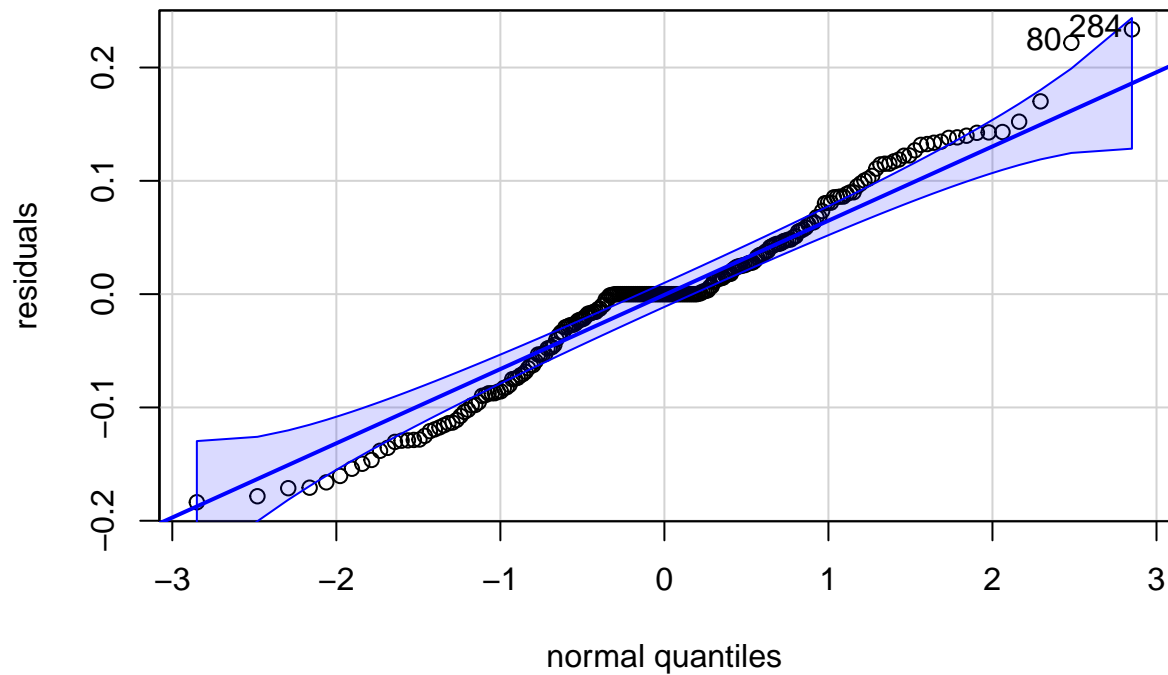


```
# Normality  
hist(resids_log, xlab="residuals", col="orange", main=NULL, nclass=15)
```





```
qqPlot(resids_log, xlab="normal quantiles", ylab="residuals")
```



```
## 284 80
## 201 56
```

## Model Selection

```
#Forward, Stepwise Regression
reduced_model = lm(Salary~1, data = HRtrain)
full_model = lm(Salary~., data = HRtrain)
step(reduced_model, scope = list(lower= reduced_model, upper=full_model), direction="forward")
```

```
## Start: AIC=4685.51
```

```
## Salary ~ 1
```

```
##
```

	Df	Sum of Sq	RSS	AIC
## + Position	30	1.6708e+11	7.5203e+09	4025.3
## + ManagerName	20	1.0905e+11	6.5547e+10	4501.2
## + Department	5	8.6971e+10	8.7626e+10	4537.6
## + SpecialProjectsCount	1	4.2231e+10	1.3237e+11	4624.1
## + Age	1	5.3622e+09	1.6923e+11	4680.4
## + RecruitmentSource	8	1.4606e+10	1.5999e+11	4681.5
## + PerfScoreID	1	3.7197e+09	1.7088e+11	4682.6
## <none>			1.7460e+11	4685.5
## + Absences	1	1.5107e+09	1.7309e+11	4685.5
## + EmpSatisfaction	1	1.4749e+09	1.7312e+11	4685.6
## + EngagementSurvey	1	1.4353e+09	1.7316e+11	4685.6

```

## + DaysLateLast30      1 1.2974e+09 1.7330e+11 4685.8
## + HispanicLatino      1 9.6722e+08 1.7363e+11 4686.2
## + EmploymentStatus    2 2.2524e+09 1.7234e+11 4686.5
## + Sex                  1 1.3346e+08 1.7446e+11 4687.3
## + EmployedYear         1 1.0951e+08 1.7449e+11 4687.4
## + MarriedID            1 6.2478e+07 1.7453e+11 4687.4
## + RaceDesc             4 4.1408e+09 1.7046e+11 4688.0
## + CitizenDesc          2 4.7506e+08 1.7412e+11 4688.9
## + MaritalDesc          4 6.7424e+08 1.7392e+11 4692.6
## + State                27 1.3895e+10 1.6070e+11 4720.5
##
## Step: AIC=4025.34
## Salary ~ Position
##
##              Df Sum of Sq      RSS      AIC
## + EmpSatisfaction  1 174393871 7345885562 4022.0
## + Absences         1 164124350 7356155083 4022.3
## + HispanicLatino   1  89532825 7430746608 4024.6
## + <none>              7520279433 4025.3
## + MarriedID        1  54001322 7466278111 4025.7
## + SpecialProjectsCount 1  52814205 7467465228 4025.7
## + Age              1  49637668 7470641766 4025.8
## + PerfScoreID      1  43649723 7476629710 4026.0
## + EngagementSurvey  1  39353574 7480925859 4026.1
## + Sex              1  37172035 7483107399 4026.2
## + DaysLateLast30   1  14813751 7505465682 4026.9
## + EmployedYear      1  1596822 7518682611 4027.3
## + MaritalDesc       4 172353147 7347926287 4028.0
## + CitizenDesc       2  34558739 7485720694 4028.3
## + EmploymentStatus  2  17518614 7502760819 4028.8
## + RaceDesc          4 115362524 7404916909 4029.8
## + RecruitmentSource  8 275209295 7245070139 4032.8
## + ManagerName       12 518554393 7001725040 4033.0
## + State             26 962543385 6557736048 4046.0
##
## Step: AIC=4021.96
## Salary ~ Position + EmpSatisfaction
##
##              Df Sum of Sq      RSS      AIC
## + Absences         1 132069340 7213816222 4019.8
## + HispanicLatino   1  74327927 7271557635 4021.6
## + <none>              7345885562 4022.0
## + Age              1  59611884 7286273678 4022.1
## + SpecialProjectsCount 1  48534335 7297351227 4022.4
## + MarriedID        1  47210469 7298675093 4022.5
## + Sex              1  28948936 7316936626 4023.1
## + EngagementSurvey  1  16944407 7328941155 4023.4
## + PerfScoreID      1  9455305 7336430257 4023.7
## + DaysLateLast30   1  1024462 7344861099 4023.9
## + EmployedYear      1  279920 7345605641 4024.0
## + MaritalDesc       4 151620317 7194265244 4025.2
## + EmploymentStatus  2  17564476 7328321086 4025.4
## + CitizenDesc       2  17292404 7328593158 4025.4
## + RaceDesc          4 116276144 7229609418 4026.3

```

```

## + RecruitmentSource      8 298358627 7047526935 4028.5
## + ManagerName           12 532107463 6813778099 4028.7
## + State                  26 830713800 6515171761 4046.5
##
## Step: AIC=4019.81
## Salary ~ Position + EmpSatisfaction + Absences
##
##              Df Sum of Sq      RSS      AIC
## + HispanicLatino    1  95501851 7118314371 4018.8
## + Age                1  73513282 7140302940 4019.5
## + MarriedID         1  67308592 7146507630 4019.7
## <none>                                7213816222 4019.8
## + SpecialProjectsCount 1  43835940 7169980282 4020.4
## + Sex                1  31721858 7182094364 4020.8
## + EngagementSurvey    1  17916113 7195900109 4021.2
## + PerfScoreID         1   6301469 7207514753 4021.6
## + DaysLateLast30      1   1050524 7212765698 4021.8
## + EmployedYear        1     31998 7213784225 4021.8
## + MaritalDesc         4 166144246 7047671976 4022.5
## + CitizenDesc         2  18619891 7195196331 4023.2
## + EmploymentStatus    2  16244615 7197571608 4023.3
## + RaceDesc            4   90053971 7123762251 4024.9
## + ManagerName        12 530220322 6683595900 4026.3
## + RecruitmentSource    8 288390527 6925425695 4026.5
## + State                26 800942960 6412873262 4044.9
##
## Step: AIC=4018.76
## Salary ~ Position + EmpSatisfaction + Absences + HispanicLatino
##
##              Df Sum of Sq      RSS      AIC
## + MarriedID         1  96619366 7021695005 4017.6
## + Age                1  61905738 7056408633 4018.8
## <none>                                7118314371 4018.8
## + SpecialProjectsCount 1  44558161 7073756210 4019.3
## + Sex                1  25355950 7092958421 4019.9
## + EngagementSurvey    1  22629416 7095684955 4020.0
## + PerfScoreID         1   5989249 7112325122 4020.6
## + DaysLateLast30      1   1301144 7117013227 4020.7
## + EmployedYear        1    227252 7118087119 4020.7
## + MaritalDesc         4 168276797 6950037574 4021.3
## + CitizenDesc         2  16242187 7102072184 4022.2
## + EmploymentStatus    2  12770444 7105543926 4022.3
## + ManagerName        12 555637028 6562677343 4024.1
## + RaceDesc            4   80303711 7038010660 4024.2
## + RecruitmentSource    8 287669271 6830645100 4025.3
## + State                26 831057367 6287257004 4042.3
##
## Step: AIC=4017.63
## Salary ~ Position + EmpSatisfaction + Absences + HispanicLatino +
##      MarriedID
##
##              Df Sum of Sq      RSS      AIC
## <none>                                7021695005 4017.6
## + SpecialProjectsCount 1  56726871 6964968133 4017.8

```

```

## + Age          1  53502816 6968192189 4017.9
## + Sex          1  29300674 6992394331 4018.7
## + EngagementSurvey 1  14348768 7007346237 4019.2
## + PerfScoreID  1  5234175 7016460830 4019.5
## + DaysLateLast30 1  2035095 7019659910 4019.6
## + EmployedYear  1  1409152 7020285853 4019.6
## + CitizenDesc   2  19602698 7002092307 4021.0
## + ManagerName   12 601578787 6420116218 4021.1
## + MaritalDesc   3  71657431 6950037574 4021.3
## + EmploymentStatus 2 10360781 7011334224 4021.3
## + RaceDesc      4  82289389 6939405615 4022.9
## + RecruitmentSource 8 302883314 6718811691 4023.5
## + State         26 864931192 6156763813 4039.5

##
## Call:
## lm(formula = Salary ~ Position + EmpSatisfaction + Absences +
##     HispanicLatino + MarriedID, data = HRtrain)
##
## Coefficients:
##              (Intercept)      PositionAdministrative Assistant
##                   59675.3                      -10431.0
##      PositionArea Sales Manager      PositionBI Developer
##                   1054.3                      31874.0
##      PositionBI Director                      PositionCIO
##                   47432.1                      154215.8
##      PositionData Analyst      PositionData Architect
##                   26402.4                      85475.7
##      PositionDatabase Administrator      PositionDirector of Operations
##                   49005.6                      104428.4
##      PositionDirector of Sales      PositionEnterprise Architect
##                   115505.2                      39656.0
##      PositionIT Director      PositionIT Manager - DB
##                   111928.4                      80015.5
##      PositionIT Manager - Infra      PositionIT Manager - Support
##                   97379.6                      74605.6
##      PositionIT Support      PositionNetwork Engineer
##                   2338.5                      -1706.6
##      PositionPresident & CEO      PositionPrincipal Data Architect
##                   190867.5                      59106.1
##      PositionProduction Manager      PositionProduction Technician I
##                   14645.5                      -7928.1
##      PositionProduction Technician II      PositionSales Manager
##                   1424.7                      3314.6
##      PositionSenior BI Developer      PositionShared Services Manager
##                   23106.7                      28388.5
##      PositionSoftware Engineer      PositionSoftware Engineering Manager
##                   31335.2                      14992.2
##      PositionSr. Accountant      PositionSr. DBA
##                   41281.0                      41058.8
##      PositionSr. Network Engineer      EmpSatisfaction
##                   29125.4                      791.3
##      Absences      HispanicLatinoYes
##                   162.7                      -3107.2

```

```

##                               MarriedID1
##                               -1436.1
#Backward, Stepwise Regression
step(full_model,scope = list(lower=reduced_model,upper=full_model),direction="backward")

## Start:  AIC=4065.02
## Salary ~ MarriedID + MaritalDesc + Sex + EmploymentStatus + Department +
##       PerfScoreID + RecruitmentSource + Position + State + Age +
##       CitizenDesc + RaceDesc + HispanicLatino + EmployedYear +
##       ManagerName + EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##       DaysLateLast30 + Absences
##
##
## Step:  AIC=4065.02
## Salary ~ MarriedID + MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
##       RecruitmentSource + Position + State + Age + CitizenDesc +
##       RaceDesc + HispanicLatino + EmployedYear + ManagerName +
##       EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##       DaysLateLast30 + Absences
##
##
## Step:  AIC=4065.02
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
##       RecruitmentSource + Position + State + Age + CitizenDesc +
##       RaceDesc + HispanicLatino + EmployedYear + ManagerName +
##       EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##       DaysLateLast30 + Absences
##
##
##           Df  Sum of Sq      RSS    AIC
## - State      25 8.8616e+08 5.9115e+09 4052.2
## - RecruitmentSource    7 1.5335e+08 5.1787e+09 4057.9
## - CitizenDesc      2 2.5962e+07 5.0513e+09 4062.2
## - RaceDesc         4 1.2120e+08 5.1465e+09 4062.5
## - ManagerName     10 3.9915e+08 5.4244e+09 4062.5
## - EmploymentStatus    2 3.5528e+07 5.0608e+09 4062.6
## - EmployedYear        1 4.9172e+06 5.0302e+09 4063.2
## - SpecialProjectsCount 1 8.7142e+06 5.0340e+09 4063.4
## - Age              1 1.1782e+07 5.0371e+09 4063.6
## - DaysLateLast30      1 1.6643e+07 5.0419e+09 4063.8
## - PerfScoreID        1 4.0651e+07 5.0660e+09 4064.9
## - EmpSatisfaction     1 4.1738e+07 5.0670e+09 4064.9
## - EngagementSurvey    1 4.2335e+07 5.0676e+09 4064.9
## - Sex              1 4.2727e+07 5.0680e+09 4065.0
## <none>                                5.0253e+09 4065.0
## - Absences          1 8.1211e+07 5.1065e+09 4066.7
## - HispanicLatino      1 1.2477e+08 5.1501e+09 4068.6
## - MaritalDesc        4 2.6945e+08 5.2947e+09 4069.0
## - Position          20 2.9791e+10 3.4817e+10 4468.3
##
## Step:  AIC=4052.21
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
##       RecruitmentSource + Position + Age + CitizenDesc + RaceDesc +
##       HispanicLatino + EmployedYear + ManagerName + EngagementSurvey +
##       EmpSatisfaction + SpecialProjectsCount + DaysLateLast30 +

```

```

##      Absences
##
##      Df  Sum of Sq      RSS      AIC
## - RecruitmentSource      7 1.9857e+08 6.1100e+09 4045.8
## - ManagerName           11 4.6206e+08 6.3735e+09 4047.4
## - EmploymentStatus       2 3.8123e+06 5.9153e+09 4048.4
## - RaceDesc               4 1.0949e+08 6.0210e+09 4048.4
## - CitizenDesc            2 4.0795e+07 5.9523e+09 4049.8
## - DaysLateLast30         1 3.9710e+03 5.9115e+09 4050.2
## - SpecialProjectsCount   1 1.1915e+06 5.9127e+09 4050.3
## - EmployedYear           1 2.3845e+06 5.9139e+09 4050.3
## - PerfScoreID            1 6.6387e+06 5.9181e+09 4050.5
## - Sex                    1 1.1411e+07 5.9229e+09 4050.7
## - Age                    1 1.4223e+07 5.9257e+09 4050.8
## - MaritalDesc            4 1.7298e+08 6.0844e+09 4050.8
## - EngagementSurvey       1 3.7358e+07 5.9488e+09 4051.7
## <none>                    5.9115e+09 4052.2
## - EmpSatisfaction        1 1.0073e+08 6.0122e+09 4054.1
## - Absences               1 1.0282e+08 6.0143e+09 4054.2
## - HispanicLatino         1 1.0982e+08 6.0213e+09 4054.4
## - Position               21 3.7480e+10 4.3392e+10 4466.7
##
## Step:  AIC=4045.78
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
##      Position + Age + CitizenDesc + RaceDesc + HispanicLatino +
##      EmployedYear + ManagerName + EngagementSurvey + EmpSatisfaction +
##      SpecialProjectsCount + DaysLateLast30 + Absences
##
##      Df  Sum of Sq      RSS      AIC
## - ManagerName           12 5.5242e+08 6.6625e+09 4041.6
## - RaceDesc              4 1.0747e+08 6.2175e+09 4041.8
## - EmploymentStatus       2 1.7785e+07 6.1278e+09 4042.4
## - DaysLateLast30         1 6.4880e+03 6.1100e+09 4043.8
## - EmployedYear           1 1.1265e+04 6.1101e+09 4043.8
## - SpecialProjectsCount   1 1.2223e+06 6.1113e+09 4043.8
## - PerfScoreID            1 1.5311e+06 6.1116e+09 4043.8
## - Sex                    1 3.1824e+06 6.1132e+09 4043.9
## - CitizenDesc            2 5.7115e+07 6.1672e+09 4043.9
## - Age                    1 2.2595e+07 6.1326e+09 4044.6
## - EngagementSurvey       1 2.3411e+07 6.1335e+09 4044.7
## - MaritalDesc            4 1.9048e+08 6.3005e+09 4044.8
## <none>                    6.1100e+09 4045.8
## - EmpSatisfaction        1 7.9926e+07 6.1900e+09 4046.8
## - HispanicLatino         1 1.2183e+08 6.2319e+09 4048.3
## - Absences               1 1.3984e+08 6.2499e+09 4049.0
## - Position               22 5.0448e+10 5.6558e+10 4511.4
##
## Step:  AIC=4041.6
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
##      Position + Age + CitizenDesc + RaceDesc + HispanicLatino +
##      EmployedYear + EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##      DaysLateLast30 + Absences
##
##      Df  Sum of Sq      RSS      AIC

```

```

## - RaceDesc          4 6.9690e+07 6.7322e+09 4036.0
## - EmploymentStatus  2 3.5792e+07 6.6983e+09 4038.8
## - CitizenDesc       2 3.8285e+07 6.7007e+09 4038.9
## - MaritalDesc       4 1.6655e+08 6.8290e+09 4039.3
## - PerfScoreID       1 4.4511e+05 6.6629e+09 4039.6
## - DaysLateLast30    1 6.6983e+06 6.6692e+09 4039.8
## - Sex               1 1.4694e+07 6.6772e+09 4040.1
## - EmployedYear       1 1.5617e+07 6.6781e+09 4040.1
## - EngagementSurvey   1 2.9059e+07 6.6915e+09 4040.6
## - Age               1 3.8262e+07 6.7007e+09 4040.9
## - SpecialProjectsCount 1 5.8077e+07 6.7205e+09 4041.6
## <none>              6.6625e+09 4041.6
## - EmpSatisfaction    1 7.0380e+07 6.7328e+09 4042.0
## - HispanicLatino     1 7.4518e+07 6.7370e+09 4042.1
## - Absences           1 1.4853e+08 6.8110e+09 4044.7
## - Position           30 1.1205e+11 1.1871e+11 4641.2
##
## Step: AIC=4035.98
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
##           Position + Age + CitizenDesc + HispanicLatino + EmployedYear +
##           EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##           DaysLateLast30 + Absences
##
##           Df Sum of Sq      RSS      AIC
## - EmploymentStatus  2 3.1479e+07 6.7636e+09 4033.1
## - CitizenDesc       2 4.1903e+07 6.7741e+09 4033.4
## - MaritalDesc       4 1.6930e+08 6.9014e+09 4033.7
## - PerfScoreID       1 4.2204e+05 6.7326e+09 4034.0
## - DaysLateLast30    1 6.5068e+06 6.7387e+09 4034.2
## - EmployedYear       1 1.5317e+07 6.7475e+09 4034.5
## - Sex               1 1.9127e+07 6.7513e+09 4034.6
## - EngagementSurvey   1 2.8148e+07 6.7603e+09 4034.9
## - SpecialProjectsCount 1 4.7753e+07 6.7799e+09 4035.6
## - Age               1 5.5018e+07 6.7872e+09 4035.8
## <none>              6.7322e+09 4036.0
## - EmpSatisfaction    1 6.8492e+07 6.8006e+09 4036.3
## - HispanicLatino     1 7.9771e+07 6.8119e+09 4036.7
## - Absences           1 1.8490e+08 6.9171e+09 4040.2
## - Position           30 1.1253e+11 1.1926e+11 4634.2
##
## Step: AIC=4033.05
## Salary ~ MaritalDesc + Sex + PerfScoreID + Position + Age + CitizenDesc +
##           HispanicLatino + EmployedYear + EngagementSurvey + EmpSatisfaction +
##           SpecialProjectsCount + DaysLateLast30 + Absences
##
##           Df Sum of Sq      RSS      AIC
## - CitizenDesc       2 3.6627e+07 6.8003e+09 4030.3
## - MaritalDesc       4 1.6518e+08 6.9288e+09 4030.6
## - PerfScoreID       1 7.0624e+05 6.7643e+09 4031.1
## - EmployedYear       1 8.3380e+05 6.7645e+09 4031.1
## - DaysLateLast30    1 7.6032e+06 6.7712e+09 4031.3
## - Sex               1 1.9473e+07 6.7831e+09 4031.7
## - EngagementSurvey   1 2.1823e+07 6.7855e+09 4031.8
## - Age               1 4.8383e+07 6.8120e+09 4032.7

```



```

## - SpecialProjectsCount  1 5.8088e+07 6.8217e+09 4033.0
## <none>                                6.7636e+09 4033.1
## - EmpSatisfaction        1 6.9826e+07 6.8335e+09 4033.4
## - HispanicLatino         1 8.2848e+07 6.8465e+09 4033.8
## - Absences               1 1.7912e+08 6.9428e+09 4037.0
## - Position               30 1.1261e+11 1.1937e+11 4630.4
##
## Step: AIC=4030.29
## Salary ~ MaritalDesc + Sex + PerfScoreID + Position + Age + HispanicLatino +
##      EmployedYear + EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##      DaysLateLast30 + Absences
##
##              Df Sum of Sq      RSS      AIC
## - MaritalDesc    4 1.5639e+08 6.9566e+09 4027.5
## - PerfScoreID    1 2.1766e+05 6.8005e+09 4028.3
## - EmployedYear    1 8.4629e+05 6.8011e+09 4028.3
## - DaysLateLast30  1 4.1764e+06 6.8044e+09 4028.4
## - EngagementSurvey 1 1.5689e+07 6.8159e+09 4028.8
## - Sex             1 2.3692e+07 6.8239e+09 4029.1
## - Age             1 5.0660e+07 6.8509e+09 4030.0
## - SpecialProjectsCount 1 5.1421e+07 6.8517e+09 4030.0
## <none>                                6.8003e+09 4030.3
## - EmpSatisfaction    1 8.6299e+07 6.8866e+09 4031.2
## - HispanicLatino     1 8.6596e+07 6.8869e+09 4031.2
## - Absences           1 1.7909e+08 6.9794e+09 4034.2
## - Position           30 1.1294e+11 1.1974e+11 4627.1
##
## Step: AIC=4027.5
## Salary ~ Sex + PerfScoreID + Position + Age + HispanicLatino +
##      EmployedYear + EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##      DaysLateLast30 + Absences
##
##              Df Sum of Sq      RSS      AIC
## - PerfScoreID      1 5.7022e+05 6.9572e+09 4025.5
## - EmployedYear      1 2.1762e+06 6.9588e+09 4025.6
## - DaysLateLast30    1 4.7459e+06 6.9614e+09 4025.7
## - Sex               1 2.1569e+07 6.9782e+09 4026.2
## - EngagementSurvey  1 2.4490e+07 6.9811e+09 4026.3
## - SpecialProjectsCount 1 4.7463e+07 7.0041e+09 4027.1
## - Age               1 5.8618e+07 7.0153e+09 4027.4
## <none>                                6.9566e+09 4027.5
## - HispanicLatino    1 8.4798e+07 7.0414e+09 4028.3
## - EmpSatisfaction    1 9.8047e+07 7.0547e+09 4028.7
## - Absences           1 1.5980e+08 7.1164e+09 4030.7
## - Position           30 1.1313e+11 1.2008e+11 4619.8
##
## Step: AIC=4025.52
## Salary ~ Sex + Position + Age + HispanicLatino + EmployedYear +
##      EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##      DaysLateLast30 + Absences
##
##              Df Sum of Sq      RSS      AIC
## - EmployedYear      1 2.1076e+06 6.9593e+09 4023.6
## - DaysLateLast30    1 4.5937e+06 6.9618e+09 4023.7

```

```

## - Sex                1 2.1836e+07 6.9791e+09 4024.2
## - EngagementSurvey    1 2.7485e+07 6.9847e+09 4024.4
## - SpecialProjectsCount 1 4.6900e+07 7.0041e+09 4025.1
## <none>                6.9572e+09 4025.5
## - Age                1 6.1310e+07 7.0185e+09 4025.5
## - HispanicLatino      1 8.5141e+07 7.0424e+09 4026.3
## - EmpSatisfaction     1 1.0379e+08 7.0610e+09 4026.9
## - Absences            1 1.6309e+08 7.1203e+09 4028.8
## - Position            30 1.1455e+11 1.2151e+11 4620.5
##
## Step: AIC=4023.58
## Salary ~ Sex + Position + Age + HispanicLatino + EngagementSurvey +
##      EmpSatisfaction + SpecialProjectsCount + DaysLateLast30 +
##      Absences
##
##              Df Sum of Sq      RSS      AIC
## - DaysLateLast30  1 5.6303e+06 6.9650e+09 4021.8
## - Sex              1 2.1251e+07 6.9806e+09 4022.3
## - EngagementSurvey  1 2.7238e+07 6.9866e+09 4022.5
## - SpecialProjectsCount 1 4.5330e+07 7.0047e+09 4023.1
## <none>              6.9593e+09 4023.6
## - Age              1 6.2261e+07 7.0216e+09 4023.6
## - HispanicLatino    1 8.4314e+07 7.0436e+09 4024.3
## - EmpSatisfaction   1 1.0798e+08 7.0673e+09 4025.1
## - Absences          1 1.6446e+08 7.1238e+09 4026.9
## - Position          30 1.1534e+11 1.2229e+11 4620.0
##
## Step: AIC=4021.77
## Salary ~ Sex + Position + Age + HispanicLatino + EngagementSurvey +
##      EmpSatisfaction + SpecialProjectsCount + Absences
##
##              Df Sum of Sq      RSS      AIC
## - Sex              1 1.9188e+07 6.9841e+09 4020.4
## - EngagementSurvey  1 2.2576e+07 6.9875e+09 4020.5
## - SpecialProjectsCount 1 4.6881e+07 7.0118e+09 4021.3
## <none>              6.9650e+09 4021.8
## - Age              1 6.2528e+07 7.0275e+09 4021.8
## - HispanicLatino    1 8.3523e+07 7.0485e+09 4022.5
## - EmpSatisfaction   1 1.0325e+08 7.0682e+09 4023.1
## - Absences          1 1.6379e+08 7.1287e+09 4025.1
## - Position          30 1.1535e+11 1.2231e+11 4618.0
##
## Step: AIC=4020.4
## Salary ~ Position + Age + HispanicLatino + EngagementSurvey +
##      EmpSatisfaction + SpecialProjectsCount + Absences
##
##              Df Sum of Sq      RSS      AIC
## - EngagementSurvey  1 2.2263e+07 7.0064e+09 4019.1
## - SpecialProjectsCount 1 5.3913e+07 7.0381e+09 4020.2
## <none>              6.9841e+09 4020.4
## - Age              1 6.2695e+07 7.0468e+09 4020.4
## - HispanicLatino    1 8.8914e+07 7.0731e+09 4021.3
## - EmpSatisfaction   1 1.0812e+08 7.0923e+09 4021.9
## - Absences          1 1.6174e+08 7.1459e+09 4023.6

```

```

## - Position          30 1.1534e+11 1.2232e+11 4616.0
##
## Step: AIC=4019.13
## Salary ~ Position + Age + HispanicLatino + EmpSatisfaction +
##   SpecialProjectsCount + Absences
##
##           Df Sum of Sq      RSS      AIC
## - SpecialProjectsCount  1 5.0003e+07 7.0564e+09 4018.8
## <none>                                7.0064e+09 4019.1
## - Age                  1 6.7350e+07 7.0738e+09 4019.3
## - HispanicLatino       1 8.4136e+07 7.0905e+09 4019.9
## - EmpSatisfaction      1 1.3006e+08 7.1365e+09 4021.3
## - Absences             1 1.6060e+08 7.1670e+09 4022.3
## - Position            30 1.1573e+11 1.2273e+11 4614.8
##
## Step: AIC=4018.76
## Salary ~ Position + Age + HispanicLatino + EmpSatisfaction +
##   Absences
##
##           Df Sum of Sq      RSS      AIC
## <none>                                7.0564e+09 4018.8
## - Age                  1 6.1906e+07 7.1183e+09 4018.8
## - HispanicLatino       1 8.3894e+07 7.1403e+09 4019.5
## - EmpSatisfaction      1 1.3295e+08 7.1894e+09 4021.0
## - Absences             1 1.6549e+08 7.2219e+09 4022.1
## - Position            30 1.5785e+11 1.6491e+11 4680.4
##
## Call:
## lm(formula = Salary ~ Position + Age + HispanicLatino + EmpSatisfaction +
##   Absences, data = HRtrain)
##
## Coefficients:
##           (Intercept)      PositionAdministrative Assistant
##                56692.0                -10822.8
##   PositionArea Sales Manager      PositionBI Developer
##                476.7                30888.2
##   PositionBI Director                PositionCIO
##                45476.4                154225.8
##   PositionData Analyst      PositionData Architect
##                25968.5                85182.1
##   PositionDatabase Administrator      PositionDirector of Operations
##                48230.7                104691.6
##   PositionDirector of Sales      PositionEnterprise Architect
##                113327.4                38558.9
##   PositionIT Director                PositionIT Manager - DB
##                111995.1                79325.4
##   PositionIT Manager - Infra      PositionIT Manager - Support
##                95948.0                73919.4
##   PositionIT Support                PositionNetwork Engineer
##                1531.5                -2222.9
##   PositionPresident & CEO      PositionPrincipal Data Architect
##                187313.4                58250.3
##   PositionProduction Manager      PositionProduction Technician I

```

```
##              13372.2              -8509.7
## PositionProduction Technician II PositionSales Manager
##              878.4              3950.9
## PositionSenior BI Developer PositionShared Services Manager
##              22682.1              27398.6
## PositionSoftware Engineer PositionSoftware Engineering Manager
##              30992.0              14189.9
## PositionSr. Accountant PositionSr. DBA
##              40115.1              39525.6
## PositionSr. Network Engineer Age
##              27823.6              65.5
## HispanicLatinoYes EmpSatisfaction
##              -2523.9              864.3
## Absences
##              153.8
```

*#Both direction, Stepwise Regression*

```
step(full_model,scope = list(lower=reduced_model,upper=full_model),direction="both")
```

```
## Start: AIC=4065.02
```

```
## Salary ~ MarriedID + MaritalDesc + Sex + EmploymentStatus + Department +
## PerfScoreID + RecruitmentSource + Position + State + Age +
## CitizenDesc + RaceDesc + HispanicLatino + EmployedYear +
## ManagerName + EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
## DaysLateLast30 + Absences
```

```
##
##
```

```
## Step: AIC=4065.02
```

```
## Salary ~ MarriedID + MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
## RecruitmentSource + Position + State + Age + CitizenDesc +
## RaceDesc + HispanicLatino + EmployedYear + ManagerName +
## EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
## DaysLateLast30 + Absences
```

```
##
##
```

```
## Step: AIC=4065.02
```

```
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
## RecruitmentSource + Position + State + Age + CitizenDesc +
## RaceDesc + HispanicLatino + EmployedYear + ManagerName +
## EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
## DaysLateLast30 + Absences
```

```
##
```

	Df	Sum of Sq	RSS	AIC
## - State	25	8.8616e+08	5.9115e+09	4052.2
## - RecruitmentSource	7	1.5335e+08	5.1787e+09	4057.9
## - CitizenDesc	2	2.5962e+07	5.0513e+09	4062.2
## - RaceDesc	4	1.2120e+08	5.1465e+09	4062.5
## - ManagerName	10	3.9915e+08	5.4244e+09	4062.5
## - EmploymentStatus	2	3.5528e+07	5.0608e+09	4062.6
## - EmployedYear	1	4.9172e+06	5.0302e+09	4063.2
## - SpecialProjectsCount	1	8.7142e+06	5.0340e+09	4063.4
## - Age	1	1.1782e+07	5.0371e+09	4063.6
## - DaysLateLast30	1	1.6643e+07	5.0419e+09	4063.8
## - PerfScoreID	1	4.0651e+07	5.0660e+09	4064.9
## - EmpSatisfaction	1	4.1738e+07	5.0670e+09	4064.9

```

## - EngagementSurvey      1 4.2335e+07 5.0676e+09 4064.9
## - Sex                   1 4.2727e+07 5.0680e+09 4065.0
## <none>                  5.0253e+09 4065.0
## - Absences              1 8.1211e+07 5.1065e+09 4066.7
## - HispanicLatino        1 1.2477e+08 5.1501e+09 4068.6
## - MaritalDesc           4 2.6945e+08 5.2947e+09 4069.0
## - Position              20 2.9791e+10 3.4817e+10 4468.3
##
## Step: AIC=4052.21
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
## RecruitmentSource + Position + Age + CitizenDesc + RaceDesc +
## HispanicLatino + EmployedYear + ManagerName + EngagementSurvey +
## EmpSatisfaction + SpecialProjectsCount + DaysLateLast30 +
## Absences
##
##           Df Sum of Sq      RSS      AIC
## - RecruitmentSource      7 1.9857e+08 6.1100e+09 4045.8
## - ManagerName            11 4.6206e+08 6.3735e+09 4047.4
## - EmploymentStatus        2 3.8123e+06 5.9153e+09 4048.4
## - RaceDesc                4 1.0949e+08 6.0210e+09 4048.4
## - CitizenDesc             2 4.0795e+07 5.9523e+09 4049.8
## - DaysLateLast30          1 3.9710e+03 5.9115e+09 4050.2
## - SpecialProjectsCount    1 1.1915e+06 5.9127e+09 4050.3
## - EmployedYear            1 2.3845e+06 5.9139e+09 4050.3
## - PerfScoreID             1 6.6387e+06 5.9181e+09 4050.5
## - Sex                     1 1.1411e+07 5.9229e+09 4050.7
## - Age                     1 1.4223e+07 5.9257e+09 4050.8
## - MaritalDesc             4 1.7298e+08 6.0844e+09 4050.8
## - EngagementSurvey        1 3.7358e+07 5.9488e+09 4051.7
## <none>                    5.9115e+09 4052.2
## - EmpSatisfaction         1 1.0073e+08 6.0122e+09 4054.1
## - Absences                1 1.0282e+08 6.0143e+09 4054.2
## - HispanicLatino          1 1.0982e+08 6.0213e+09 4054.4
## + State                   25 8.8616e+08 5.0253e+09 4065.0
## - Position                21 3.7480e+10 4.3392e+10 4466.7
##
## Step: AIC=4045.78
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
## Position + Age + CitizenDesc + RaceDesc + HispanicLatino +
## EmployedYear + ManagerName + EngagementSurvey + EmpSatisfaction +
## SpecialProjectsCount + DaysLateLast30 + Absences
##
##           Df Sum of Sq      RSS      AIC
## - ManagerName            12 5.5242e+08 6.6625e+09 4041.6
## - RaceDesc                4 1.0747e+08 6.2175e+09 4041.8
## - EmploymentStatus        2 1.7785e+07 6.1278e+09 4042.4
## - DaysLateLast30          1 6.4880e+03 6.1100e+09 4043.8
## - EmployedYear            1 1.1265e+04 6.1101e+09 4043.8
## - SpecialProjectsCount    1 1.2223e+06 6.1113e+09 4043.8
## - PerfScoreID             1 1.5311e+06 6.1116e+09 4043.8
## - Sex                     1 3.1824e+06 6.1132e+09 4043.9
## - CitizenDesc             2 5.7115e+07 6.1672e+09 4043.9
## - Age                     1 2.2595e+07 6.1326e+09 4044.6
## - EngagementSurvey        1 2.3411e+07 6.1335e+09 4044.7

```

```

## - MaritalDesc          4 1.9048e+08 6.3005e+09 4044.8
## <none>                  6.1100e+09 4045.8
## - EmpSatisfaction      1 7.9926e+07 6.1900e+09 4046.8
## - HispanicLatino       1 1.2183e+08 6.2319e+09 4048.3
## - Absences             1 1.3984e+08 6.2499e+09 4049.0
## + RecruitmentSource     7 1.9857e+08 5.9115e+09 4052.2
## + State                 25 9.3139e+08 5.1787e+09 4057.9
## - Position              22 5.0448e+10 5.6558e+10 4511.4
##
## Step: AIC=4041.6
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
##           Position + Age + CitizenDesc + RaceDesc + HispanicLatino +
##           EmployedYear + EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##           DaysLateLast30 + Absences
##
##           Df Sum of Sq      RSS      AIC
## - RaceDesc          4 6.9690e+07 6.7322e+09 4036.0
## - EmploymentStatus  2 3.5792e+07 6.6983e+09 4038.8
## - CitizenDesc        2 3.8285e+07 6.7007e+09 4038.9
## - MaritalDesc        4 1.6655e+08 6.8290e+09 4039.3
## - PerfScoreID        1 4.4511e+05 6.6629e+09 4039.6
## - DaysLateLast30     1 6.6983e+06 6.6692e+09 4039.8
## - Sex                1 1.4694e+07 6.6772e+09 4040.1
## - EmployedYear        1 1.5617e+07 6.6781e+09 4040.1
## - EngagementSurvey    1 2.9059e+07 6.6915e+09 4040.6
## - Age                1 3.8262e+07 6.7007e+09 4040.9
## - SpecialProjectsCount 1 5.8077e+07 6.7205e+09 4041.6
## <none>                6.6625e+09 4041.6
## - EmpSatisfaction     1 7.0380e+07 6.7328e+09 4042.0
## - HispanicLatino      1 7.4518e+07 6.7370e+09 4042.1
## - Absences            1 1.4853e+08 6.8110e+09 4044.7
## + ManagerName         12 5.5242e+08 6.1100e+09 4045.8
## + RecruitmentSource    8 2.8894e+08 6.3735e+09 4047.4
## + State                26 9.8737e+08 5.6751e+09 4056.9
## - Position            30 1.1205e+11 1.1871e+11 4641.2
##
## Step: AIC=4035.98
## Salary ~ MaritalDesc + Sex + EmploymentStatus + PerfScoreID +
##           Position + Age + CitizenDesc + HispanicLatino + EmployedYear +
##           EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##           DaysLateLast30 + Absences
##
##           Df Sum of Sq      RSS      AIC
## - EmploymentStatus    2 3.1479e+07 6.7636e+09 4033.1
## - CitizenDesc          2 4.1903e+07 6.7741e+09 4033.4
## - MaritalDesc          4 1.6930e+08 6.9014e+09 4033.7
## - PerfScoreID          1 4.2204e+05 6.7326e+09 4034.0
## - DaysLateLast30      1 6.5068e+06 6.7387e+09 4034.2
## - EmployedYear         1 1.5317e+07 6.7475e+09 4034.5
## - Sex                  1 1.9127e+07 6.7513e+09 4034.6
## - EngagementSurvey     1 2.8148e+07 6.7603e+09 4034.9
## - SpecialProjectsCount 1 4.7753e+07 6.7799e+09 4035.6
## - Age                  1 5.5018e+07 6.7872e+09 4035.8
## <none>                6.7322e+09 4036.0

```

```

## - EmpSatisfaction      1 6.8492e+07 6.8006e+09 4036.3
## - HispanicLatino      1 7.9771e+07 6.8119e+09 4036.7
## - Absences            1 1.8490e+08 6.9171e+09 4040.2
## + RaceDesc            4 6.9690e+07 6.6625e+09 4041.6
## + ManagerName        12 5.1464e+08 6.2175e+09 4041.8
## + RecruitmentSource    8 2.6540e+08 6.4668e+09 4042.8
## + State              26 9.3855e+08 5.7936e+09 4053.6
## - Position           30 1.1253e+11 1.1926e+11 4634.2
##
## Step: AIC=4033.05
## Salary ~ MaritalDesc + Sex + PerfScoreID + Position + Age + CitizenDesc +
##      HispanicLatino + EmployedYear + EngagementSurvey + EmpSatisfaction +
##      SpecialProjectsCount + DaysLateLast30 + Absences
##
##              Df Sum of Sq      RSS      AIC
## - CitizenDesc      2 3.6627e+07 6.8003e+09 4030.3
## - MaritalDesc      4 1.6518e+08 6.9288e+09 4030.6
## - PerfScoreID      1 7.0624e+05 6.7643e+09 4031.1
## - EmployedYear      1 8.3380e+05 6.7645e+09 4031.1
## - DaysLateLast30    1 7.6032e+06 6.7712e+09 4031.3
## - Sex              1 1.9473e+07 6.7831e+09 4031.7
## - EngagementSurvey  1 2.1823e+07 6.7855e+09 4031.8
## - Age              1 4.8383e+07 6.8120e+09 4032.7
## - SpecialProjectsCount 1 5.8088e+07 6.8217e+09 4033.0
## <none>              6.7636e+09 4033.1
## - EmpSatisfaction   1 6.9826e+07 6.8335e+09 4033.4
## - HispanicLatino    1 8.2848e+07 6.8465e+09 4033.8
## + EmploymentStatus  2 3.1479e+07 6.7322e+09 4036.0
## - Absences          1 1.7912e+08 6.9428e+09 4037.0
## + ManagerName      12 5.3080e+08 6.2328e+09 4038.3
## + RaceDesc          4 6.5377e+07 6.6983e+09 4038.8
## + RecruitmentSource  8 2.9307e+08 6.4706e+09 4038.9
## + State             26 9.1195e+08 5.8517e+09 4051.9
## - Position         30 1.1261e+11 1.1937e+11 4630.4
##
## Step: AIC=4030.29
## Salary ~ MaritalDesc + Sex + PerfScoreID + Position + Age + HispanicLatino +
##      EmployedYear + EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##      DaysLateLast30 + Absences
##
##              Df Sum of Sq      RSS      AIC
## - MaritalDesc      4 1.5639e+08 6.9566e+09 4027.5
## - PerfScoreID      1 2.1766e+05 6.8005e+09 4028.3
## - EmployedYear      1 8.4629e+05 6.8011e+09 4028.3
## - DaysLateLast30    1 4.1764e+06 6.8044e+09 4028.4
## - EngagementSurvey  1 1.5689e+07 6.8159e+09 4028.8
## - Sex              1 2.3692e+07 6.8239e+09 4029.1
## - Age              1 5.0660e+07 6.8509e+09 4030.0
## - SpecialProjectsCount 1 5.1421e+07 6.8517e+09 4030.0
## <none>              6.8003e+09 4030.3
## - EmpSatisfaction   1 8.6299e+07 6.8866e+09 4031.2
## - HispanicLatino    1 8.6596e+07 6.8869e+09 4031.2
## + CitizenDesc      2 3.6627e+07 6.7636e+09 4033.1
## + EmploymentStatus  2 2.6203e+07 6.7741e+09 4033.4

```

```

## - Absences          1 1.7909e+08 6.9794e+09 4034.2
## + RaceDesc          4 6.9110e+07 6.7311e+09 4036.0
## + RecruitmentSource  8 3.0015e+08 6.5001e+09 4036.0
## + ManagerName       12 5.0609e+08 6.2942e+09 4036.6
## + State             26 9.1229e+08 5.8880e+09 4049.3
## - Position          30 1.1294e+11 1.1974e+11 4627.1
##
## Step: AIC=4027.5
## Salary ~ Sex + PerfScoreID + Position + Age + HispanicLatino +
##      EmployedYear + EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##      DaysLateLast30 + Absences
##
##              Df Sum of Sq      RSS      AIC
## - PerfScoreID    1 5.7022e+05 6.9572e+09 4025.5
## - EmployedYear    1 2.1762e+06 6.9588e+09 4025.6
## - DaysLateLast30  1 4.7459e+06 6.9614e+09 4025.7
## - Sex            1 2.1569e+07 6.9782e+09 4026.2
## - EngagementSurvey 1 2.4490e+07 6.9811e+09 4026.3
## + MarriedID       1 9.4084e+07 6.8626e+09 4026.4
## - SpecialProjectsCount 1 4.7463e+07 7.0041e+09 4027.1
## - Age            1 5.8618e+07 7.0153e+09 4027.4
## <none>                6.9566e+09 4027.5
## - HispanicLatino  1 8.4798e+07 7.0414e+09 4028.3
## - EmpSatisfaction  1 9.8047e+07 7.0547e+09 4028.7
## + MaritalDesc     4 1.5639e+08 6.8003e+09 4030.3
## + CitizenDesc     2 2.7833e+07 6.9288e+09 4030.6
## - Absences        1 1.5980e+08 7.1164e+09 4030.7
## + EmploymentStatus 2 2.4114e+07 6.9325e+09 4030.7
## + RaceDesc        4 7.2125e+07 6.8845e+09 4033.1
## + RecruitmentSource 8 3.0732e+08 6.6493e+09 4033.1
## + ManagerName     12 4.8438e+08 6.4723e+09 4035.0
## + State           26 8.3896e+08 6.1177e+09 4050.1
## - Position        30 1.1313e+11 1.2008e+11 4619.8
##
## Step: AIC=4025.52
## Salary ~ Sex + Position + Age + HispanicLatino + EmployedYear +
##      EngagementSurvey + EmpSatisfaction + SpecialProjectsCount +
##      DaysLateLast30 + Absences
##
##              Df Sum of Sq      RSS      AIC
## - EmployedYear    1 2.1076e+06 6.9593e+09 4023.6
## - DaysLateLast30  1 4.5937e+06 6.9618e+09 4023.7
## - Sex            1 2.1836e+07 6.9791e+09 4024.2
## + MarriedID       1 9.4168e+07 6.8630e+09 4024.4
## - EngagementSurvey 1 2.7485e+07 6.9847e+09 4024.4
## - SpecialProjectsCount 1 4.6900e+07 7.0041e+09 4025.1
## <none>                6.9572e+09 4025.5
## - Age            1 6.1310e+07 7.0185e+09 4025.5
## - HispanicLatino  1 8.5141e+07 7.0424e+09 4026.3
## - EmpSatisfaction  1 1.0379e+08 7.0610e+09 4026.9
## + PerfScoreID     1 5.7022e+05 6.9566e+09 4027.5
## + MaritalDesc     4 1.5674e+08 6.8005e+09 4028.3
## + CitizenDesc     2 2.7465e+07 6.9298e+09 4028.6
## + EmploymentStatus 2 2.4476e+07 6.9327e+09 4028.7

```



```

## - Absences          1 1.6309e+08 7.1203e+09 4028.8
## + RaceDesc          4 7.2136e+07 6.8851e+09 4031.1
## + RecruitmentSource  8 3.0545e+08 6.6518e+09 4031.2
## + ManagerName       12 4.8452e+08 6.4727e+09 4033.0
## + State             26 8.3337e+08 6.1239e+09 4048.3
## - Position          30 1.1455e+11 1.2151e+11 4620.5
##
## Step: AIC=4023.58
## Salary ~ Sex + Position + Age + HispanicLatino + EngagementSurvey +
##      EmpSatisfaction + SpecialProjectsCount + DaysLateLast30 +
##      Absences
##
##              Df Sum of Sq      RSS      AIC
## - DaysLateLast30  1 5.6303e+06 6.9650e+09 4021.8
## - Sex             1 2.1251e+07 6.9806e+09 4022.3
## - EngagementSurvey 1 2.7238e+07 6.9866e+09 4022.5
## + MarriedID       1 9.0833e+07 6.8685e+09 4022.6
## - SpecialProjectsCount 1 4.5330e+07 7.0047e+09 4023.1
## <none>              6.9593e+09 4023.6
## - Age             1 6.2261e+07 7.0216e+09 4023.6
## - HispanicLatino   1 8.4314e+07 7.0436e+09 4024.3
## - EmpSatisfaction  1 1.0798e+08 7.0673e+09 4025.1
## + EmployedYear     1 2.1076e+06 6.9572e+09 4025.5
## + PerfScoreID      1 5.0159e+05 6.9588e+09 4025.6
## + MaritalDesc      4 1.5804e+08 6.8013e+09 4026.3
## + CitizenDesc      2 2.7668e+07 6.9317e+09 4026.7
## - Absences         1 1.6446e+08 7.1238e+09 4026.9
## + EmploymentStatus 2 1.3806e+07 6.9455e+09 4027.1
## + RaceDesc         4 7.2227e+07 6.8871e+09 4029.2
## + RecruitmentSource 8 3.0412e+08 6.6552e+09 4029.4
## + ManagerName      12 4.8576e+08 6.4736e+09 4031.0
## + State            26 8.3500e+08 6.1243e+09 4046.3
## - Position         30 1.1534e+11 1.2229e+11 4620.0
##
## Step: AIC=4021.77
## Salary ~ Sex + Position + Age + HispanicLatino + EngagementSurvey +
##      EmpSatisfaction + SpecialProjectsCount + Absences
##
##              Df Sum of Sq      RSS      AIC
## - Sex             1 1.9188e+07 6.9841e+09 4020.4
## - EngagementSurvey 1 2.2576e+07 6.9875e+09 4020.5
## + MarriedID       1 9.4842e+07 6.8701e+09 4020.6
## - SpecialProjectsCount 1 4.6881e+07 7.0118e+09 4021.3
## <none>              6.9650e+09 4021.8
## - Age             1 6.2528e+07 7.0275e+09 4021.8
## - HispanicLatino   1 8.3523e+07 7.0485e+09 4022.5
## - EmpSatisfaction  1 1.0325e+08 7.0682e+09 4023.1
## + DaysLateLast30  1 5.6303e+06 6.9593e+09 4023.6
## + EmployedYear     1 3.1442e+06 6.9618e+09 4023.7
## + PerfScoreID      1 7.0046e+05 6.9643e+09 4023.7
## + MaritalDesc      4 1.5807e+08 6.8069e+09 4024.5
## + CitizenDesc      2 2.5034e+07 6.9399e+09 4024.9
## - Absences         1 1.6379e+08 7.1287e+09 4025.1
## + EmploymentStatus 2 1.6498e+07 6.9485e+09 4025.2

```

```

## + RaceDesc          4 7.1627e+07 6.8933e+09 4027.4
## + RecruitmentSource  8 2.9240e+08 6.6726e+09 4027.9
## + ManagerName       12 4.9103e+08 6.4739e+09 4029.0
## + State             26 8.3975e+08 6.1252e+09 4044.3
## - Position          30 1.1535e+11 1.2231e+11 4618.0
##
## Step:  AIC=4020.4
## Salary ~ Position + Age + HispanicLatino + EngagementSurvey +
##      EmpSatisfaction + SpecialProjectsCount + Absences
##
##              Df  Sum of Sq      RSS    AIC
## - EngagementSurvey  1 2.2263e+07 7.0064e+09 4019.1
## + MarriedID         1 9.2304e+07 6.8918e+09 4019.4
## - SpecialProjectsCount 1 5.3913e+07 7.0381e+09 4020.2
## <none>              6.9841e+09 4020.4
## - Age              1 6.2695e+07 7.0468e+09 4020.4
## - HispanicLatino   1 8.8914e+07 7.0731e+09 4021.3
## + Sex              1 1.9188e+07 6.9650e+09 4021.8
## - EmpSatisfaction  1 1.0812e+08 7.0923e+09 4021.9
## + DaysLateLast30   1 3.5670e+06 6.9806e+09 4022.3
## + EmployedYear      1 2.2452e+06 6.9819e+09 4022.3
## + PerfScoreID      1 1.8256e+05 6.9840e+09 4022.4
## + MaritalDesc       4 1.5577e+08 6.8284e+09 4023.2
## + CitizenDesc       2 2.8435e+07 6.9557e+09 4023.5
## - Absences         1 1.6174e+08 7.1459e+09 4023.6
## + EmploymentStatus  2 1.6333e+07 6.9678e+09 4023.9
## + RaceDesc         4 7.7145e+07 6.9070e+09 4025.9
## + RecruitmentSource  8 2.7487e+08 6.7093e+09 4027.2
## + ManagerName      12 5.0187e+08 6.4823e+09 4027.3
## + State            26 8.0862e+08 6.1755e+09 4044.2
## - Position         30 1.1534e+11 1.2232e+11 4616.0
##
## Step:  AIC=4019.13
## Salary ~ Position + Age + HispanicLatino + EmpSatisfaction +
##      SpecialProjectsCount + Absences
##
##              Df  Sum of Sq      RSS    AIC
## + MarriedID         1 1.0009e+08 6.9063e+09 4017.8
## - SpecialProjectsCount 1 5.0003e+07 7.0564e+09 4018.8
## <none>              7.0064e+09 4019.1
## - Age              1 6.7350e+07 7.0738e+09 4019.3
## - HispanicLatino   1 8.4136e+07 7.0905e+09 4019.9
## + EngagementSurvey  1 2.2263e+07 6.9841e+09 4020.4
## + Sex              1 1.8874e+07 6.9875e+09 4020.5
## + PerfScoreID      1 5.2234e+06 7.0012e+09 4021.0
## + DaysLateLast30   1 1.7791e+06 7.0046e+09 4021.1
## + EmployedYear      1 7.8920e+05 7.0056e+09 4021.1
## - EmpSatisfaction  1 1.3006e+08 7.1365e+09 4021.3
## + MaritalDesc       4 1.6653e+08 6.8399e+09 4021.6
## - Absences         1 1.6060e+08 7.1670e+09 4022.3
## + CitizenDesc       2 2.2691e+07 6.9837e+09 4022.4
## + EmploymentStatus  2 1.1007e+07 6.9954e+09 4022.8
## + RaceDesc         4 7.4920e+07 6.9315e+09 4024.7
## + ManagerName      12 4.9688e+08 6.5095e+09 4026.3

```

```

## + RecruitmentSource      8 2.6492e+08 6.7415e+09 4026.3
## + State                  26 8.0689e+08 6.1995e+09 4043.1
## - Position               30 1.1573e+11 1.2273e+11 4614.8
##
## Step: AIC=4017.83
## Salary ~ Position + Age + HispanicLatino + EmpSatisfaction +
##     SpecialProjectsCount + Absences + MarriedID
##
##           Df Sum of Sq      RSS      AIC
## - Age      1 5.8649e+07 6.9650e+09 4017.8
## <none>                                6.9063e+09 4017.8
## - SpecialProjectsCount 1 6.1873e+07 6.9682e+09 4017.9
## + Sex      1 2.1552e+07 6.8848e+09 4019.1
## - MarriedID 1 1.0009e+08 7.0064e+09 4019.1
## + EngagementSurvey 1 1.4480e+07 6.8918e+09 4019.4
## - HispanicLatino 1 1.1333e+08 7.0197e+09 4019.6
## - EmpSatisfaction 1 1.1518e+08 7.0215e+09 4019.6
## + PerfScoreID 1 5.2146e+06 6.9011e+09 4019.7
## + EmployedYear 1 2.9652e+06 6.9034e+09 4019.7
## + DaysLateLast30 1 2.8856e+06 6.9034e+09 4019.7
## + CitizenDesc 2 2.7332e+07 6.8790e+09 4020.9
## + EmploymentStatus 2 8.6172e+06 6.8977e+09 4021.5
## + MaritalDesc 3 6.6443e+07 6.8399e+09 4021.6
## - Absences 1 1.9031e+08 7.0966e+09 4022.1
## + RaceDesc 4 7.8459e+07 6.8279e+09 4023.2
## + ManagerName 12 5.2205e+08 6.3843e+09 4023.8
## + RecruitmentSource 8 2.7560e+08 6.6307e+09 4024.5
## + State     26 8.4082e+08 6.0655e+09 4040.1
## - Position  30 1.1583e+11 1.2273e+11 4616.8
##
## Step: AIC=4017.77
## Salary ~ Position + HispanicLatino + EmpSatisfaction + SpecialProjectsCount +
##     Absences + MarriedID
##
##           Df Sum of Sq      RSS      AIC
## - SpecialProjectsCount 1 5.6727e+07 7.0217e+09 4017.6
## <none>                                6.9650e+09 4017.8
## + Age      1 5.8649e+07 6.9063e+09 4017.8
## + Sex      1 2.1805e+07 6.9432e+09 4019.1
## + EngagementSurvey 1 1.7715e+07 6.9473e+09 4019.2
## - EmpSatisfaction 1 1.0728e+08 7.0722e+09 4019.3
## - MarriedID 1 1.0879e+08 7.0738e+09 4019.3
## + PerfScoreID 1 1.0073e+07 6.9549e+09 4019.4
## + EmployedYear 1 3.9580e+06 6.9610e+09 4019.6
## + DaysLateLast30 1 3.7645e+06 6.9612e+09 4019.6
## - HispanicLatino 1 1.2791e+08 7.0929e+09 4019.9
## + CitizenDesc 2 2.7385e+07 6.9376e+09 4020.9
## + MaritalDesc 3 6.9143e+07 6.8958e+09 4021.5
## + EmploymentStatus 2 7.7289e+06 6.9572e+09 4021.5
## - Absences 1 1.7913e+08 7.1441e+09 4021.6
## + RaceDesc 4 9.5866e+07 6.8691e+09 4022.6
## + ManagerName 12 5.4704e+08 6.4179e+09 4023.0
## + RecruitmentSource 8 2.8206e+08 6.6829e+09 4024.3
## + State     26 8.4867e+08 6.1163e+09 4040.0

```

```

## - Position          30 1.2238e+11 1.2935e+11 4626.8
##
## Step:  AIC=4017.63
## Salary ~ Position + HispanicLatino + EmpSatisfaction + Absences +
##      MarriedID
##
##              Df  Sum of Sq      RSS      AIC
## <none>                7.0217e+09 4017.6
## + SpecialProjectsCount 1 5.6727e+07 6.9650e+09 4017.8
## + Age                  1 5.3503e+07 6.9682e+09 4017.9
## + Sex                  1 2.9301e+07 6.9924e+09 4018.7
## - MarriedID           1 9.6619e+07 7.1183e+09 4018.8
## + EngagementSurvey     1 1.4349e+07 7.0073e+09 4019.2
## - EmpSatisfaction      1 1.1130e+08 7.1330e+09 4019.2
## + PerfScoreID          1 5.2342e+06 7.0165e+09 4019.5
## + DaysLateLast30       1 2.0351e+06 7.0197e+09 4019.6
## + EmployedYear         1 1.4092e+06 7.0203e+09 4019.6
## - HispanicLatino       1 1.2481e+08 7.1465e+09 4019.7
## + CitizenDesc          2 1.9603e+07 7.0021e+09 4021.0
## + ManagerName          12 6.0158e+08 6.4201e+09 4021.1
## + MaritalDesc           3 7.1657e+07 6.9500e+09 4021.3
## + EmploymentStatus     2 1.0361e+07 7.0113e+09 4021.3
## - Absences             1 1.8300e+08 7.2047e+09 4021.5
## + RaceDesc             4 8.2289e+07 6.9394e+09 4022.9
## + RecruitmentSource     8 3.0288e+08 6.7188e+09 4023.5
## + State                26 8.6493e+08 6.1568e+09 4039.5
## - Position             30 1.6379e+11 1.7081e+11 4688.5
##
## Call:
## lm(formula = Salary ~ Position + HispanicLatino + EmpSatisfaction +
##      Absences + MarriedID, data = HRtrain)
##
## Coefficients:
##              (Intercept)      PositionAdministrative Assistant
##              59675.3                -10431.0
##      PositionArea Sales Manager      PositionBI Developer
##              1054.3                31874.0
##      PositionBI Director                PositionCIO
##              47432.1                154215.8
##      PositionData Analyst      PositionData Architect
##              26402.4                85475.7
##      PositionDatabase Administrator      PositionDirector of Operations
##              49005.6                104428.4
##      PositionDirector of Sales      PositionEnterprise Architect
##              115505.2                39656.0
##      PositionIT Director                PositionIT Manager - DB
##              111928.4                80015.5
##      PositionIT Manager - Infra      PositionIT Manager - Support
##              97379.6                74605.6
##      PositionIT Support                PositionNetwork Engineer
##              2338.5                -1706.6
##      PositionPresident & CEO      PositionPrincipal Data Architect
##              190867.5                59106.1

```

```
##           PositionProduction Manager           PositionProduction Technician I
##                               14645.5                               -7928.1
##      PositionProduction Technician II           PositionSales Manager
##                               1424.7                               3314.6
##           PositionSenior BI Developer           PositionShared Services Manager
##                               23106.7                               28388.5
##           PositionSoftware Engineer PositionSoftware Engineering Manager
##                               31335.2                               14992.2
##           PositionSr. Accountant           PositionSr. DBA
##                               41281.0                               41058.8
##      PositionSr. Network Engineer           HispanicLatinoYes
##                               29125.4                               -3107.2
##           EmpSatisfaction           Absences
##                               791.3                               162.7
##           MarriedID1
##                               -1436.1
```

## Result of model comparison using training dataset

- Forward: `lm(formula = Salary ~ Position + EmpSatisfaction + Absences + HispanicLatino + MarriedID, data = HRtrain)` / AIC:4017.63
- Backward: `lm(formula = Salary ~ Position + Age + HispanicLatino + EmpSatisfaction + Absences, data = HRtrain)` / AIC:4018.76
- Both: `lm(formula = Salary ~ Position + HispanicLatino + EmpSatisfaction + Absences + MarriedID, data = HRtrain)` # same as forward / AIC:4017.63

## Result of Stepwise Regression

```
model_forward = lm(formula = Salary ~ Position + EmpSatisfaction + Absences + HispanicLatino + MarriedID, data = HRtrain)
summary(model_forward)
```

```
##
## Call:
## lm(formula = Salary ~ Position + EmpSatisfaction + Absences +
##     HispanicLatino + MarriedID, data = HRtrain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13567  -4046       0    4276  14585
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      59675.31    3879.29  15.383 < 2e-16 ***
## PositionAdministrative Assistant -10430.98    4922.83  -2.119  0.03537 *
## PositionArea Sales Manager      1054.27    3679.73   0.287  0.77480
## PositionBI Developer      31874.00    4949.79   6.439 9.21e-10 ***
## PositionBI Director      47432.13    7001.67   6.774 1.45e-10 ***
## PositionCIO      154215.78    6995.10  22.046 < 2e-16 ***
## PositionData Analyst      26402.39    4601.77   5.737 3.65e-08 ***
## PositionData Architect      85475.69    6992.64  12.224 < 2e-16 ***
## PositionDatabase Administrator   49005.56    4924.62   9.951 < 2e-16 ***
## PositionDirector of Operations  104428.43    6990.33  14.939 < 2e-16 ***
```

```
## PositionDirector of Sales      115505.20    7004.05    16.491 < 2e-16 ***
## PositionEnterprise Architect    39655.96    6996.47     5.668 5.17e-08 ***
## PositionIT Director            111928.43    6990.33    16.012 < 2e-16 ***
## PositionIT Manager - DB        80015.53    5512.86    14.514 < 2e-16 ***
## PositionIT Manager - Infra     97379.58    7186.40    13.551 < 2e-16 ***
## PositionIT Manager - Support   74605.65    6987.29    10.677 < 2e-16 ***
## PositionIT Support             2338.50    4351.19     0.537 0.59158
## PositionNetwork Engineer      -1706.58    4931.97    -0.346 0.72970
## PositionPresident & CEO         190867.55    7187.14    26.557 < 2e-16 ***
## PositionPrincipal Data Architect 59106.13    7142.26     8.276 2.03e-14 ***
## PositionProduction Manager     14645.49    4024.21     3.639 0.00035 ***
## PositionProduction Technician I -7928.06    3533.05    -2.244 0.02596 *
## PositionProduction Technician II 1424.70    3598.81     0.396 0.69263
## PositionSales Manager          3314.63    5518.15     0.601 0.54876
## PositionSenior BI Developer    23106.69    6992.64     3.304 0.00113 **
## PositionShared Services Manager 28388.54    7011.18     4.049 7.43e-05 ***
## PositionSoftware Engineer      31335.22    4169.83     7.515 2.05e-12 ***
## PositionSoftware Engineering Manager 14992.21    6964.98     2.153 0.03259 *
## PositionSr. Accountant         41281.01    5540.25     7.451 2.98e-12 ***
## PositionSr. DBA                41058.80    6999.14     5.866 1.89e-08 ***
## PositionSr. Network Engineer   29125.39    4601.77     6.329 1.67e-09 ***
## EmpSatisfaction                791.28     451.24     1.754 0.08108 .
## Absences                      162.66      72.34     2.249 0.02566 *
## HispanicLatinoYes             -3107.16    1673.23    -1.857 0.06483 .
## MarriedID1                    -1436.11     878.97    -1.634 0.10391
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 6016 on 194 degrees of freedom
```

```
## Multiple R-squared:  0.9598, Adjusted R-squared:  0.9527
```

```
## F-statistic: 136.2 on 34 and 194 DF,  p-value: < 2.2e-16
```

```
model_backward = lm(formula = Salary ~ Position + Age + HispanicLatino + EmpSatisfaction + Absences, data = HRtrain)
summary(model_backward)
```

```
##
```

```
## Call:
```

```
## lm(formula = Salary ~ Position + Age + HispanicLatino + EmpSatisfaction + Absences, data = HRtrain)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -13991  -4216         0    3906  15361
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    56691.98    4314.39   13.140 < 2e-16 ***
## PositionAdministrative Assistant -10822.83    4926.94   -2.197 0.029229 *
## PositionArea Sales Manager        476.66    3711.73    0.128 0.897949
## PositionBI Developer       30888.24    4928.33    6.267 2.32e-09 ***
## PositionBI Director       45476.36    7026.80    6.472 7.73e-10 ***
## PositionCIO             154225.83    7016.15   21.982 < 2e-16 ***
## PositionData Analyst       25968.46    4614.11    5.628 6.31e-08 ***
## PositionData Architect     85182.13    7035.10   12.108 < 2e-16 ***
## PositionDatabase Administrator  48230.69    4932.53    9.778 < 2e-16 ***
```

```
## PositionDirector of Operations      104691.57      7003.13      14.949 < 2e-16 ***
## PositionDirector of Sales           113327.39      7080.67      16.005 < 2e-16 ***
## PositionEnterprise Architect         38558.89      7094.56       5.435 1.63e-07 ***
## PositionIT Director                 111995.08      7008.89      15.979 < 2e-16 ***
## PositionIT Manager - DB              79325.38      5590.05      14.190 < 2e-16 ***
## PositionIT Manager - Infra           95948.02      7161.79      13.397 < 2e-16 ***
## PositionIT Manager - Support         73919.37      7046.12      10.491 < 2e-16 ***
## PositionIT Support                   1531.51      4366.85       0.351 0.726184
## PositionNetwork Engineer            -2222.90      4935.28     -0.450 0.652918
## PositionPresident & CEO               187313.38      7358.87      25.454 < 2e-16 ***
## PositionPrincipal Data Architect     58250.32      7199.71       8.091 6.34e-14 ***
## PositionProduction Manager           13372.18      4075.93       3.281 0.001227 **
## PositionProduction Technician I      -8509.67      3557.98     -2.392 0.017725 *
## PositionProduction Technician II       878.37      3624.53       0.242 0.808773
## PositionSales Manager                3950.94      5530.00       0.714 0.475805
## PositionSenior BI Developer           22682.14      7045.95       3.219 0.001507 **
## PositionShared Services Manager       27398.57      7012.73       3.907 0.000129 ***
## PositionSoftware Engineer            30992.04      4180.01       7.414 3.70e-12 ***
## PositionSoftware Engineering Manager  14189.92      7041.50       2.015 0.045266 *
## PositionSr. Accountant               40115.10      5519.04       7.268 8.68e-12 ***
## PositionSr. DBA                     39525.62      7021.48       5.629 6.27e-08 ***
## PositionSr. Network Engineer         27823.63      4685.25       5.939 1.31e-08 ***
## Age                                  65.50        50.20       1.305 0.193578
## HispanicLatinoYes                   -2523.93      1661.89     -1.519 0.130463
## EmpSatisfaction                      864.27       452.05       1.912 0.057367 .
## Absences                            153.75        72.08       2.133 0.034177 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6031 on 194 degrees of freedom
## Multiple R-squared:  0.9596, Adjusted R-squared:  0.9525
## F-statistic: 135.5 on 34 and 194 DF,  p-value: < 2.2e-16
```

## Check regression assumptions for model\_forward, model\_backward

```
#plot(fitted(model_forward),residuals(model_forward),xlab="Fitted values",ylab="Residuals")
#abline(h=0, col="red")
#lines(lowess(model_forward$fitted.values, residuals(model_forward)), col='blue')

#hist(residuals(model_forward), xlab="residuals", col="orange", main=NULL, nclass=15)
#qqPlot(residuals(model_forward), xlab="normal quantiles", ylab="residuals")

#plot(fitted(model_backward),residuals(model_backward),xlab="Fitted values",ylab="Residuals")
#abline(h=0, col="red")
#lines(lowess(model_backward$fitted.values, residuals(model_backward)), col='blue')

# Normality
#hist(residuals(model_backward), xlab="residuals", col="orange", main=NULL, nclass=15)
#qqPlot(residuals(model_backward), xlab="normal quantiles", ylab="residuals")

#Regularized Regression
```

```

X_pred = cbind(HRtrain$MarriedID, HRtrain$MaritalDesc, HRtrain$Sex, HRtrain$EmploymentStatus, HRtrain$D

#Lasso Regression
smodel.cv = cv.glmnet(X_pred, HRtrain$Salary, alpha = 1, nfolds = 10)
smodel = glmnet(X_pred, HRtrain$Salary, alpha = 1, nlambda=100)

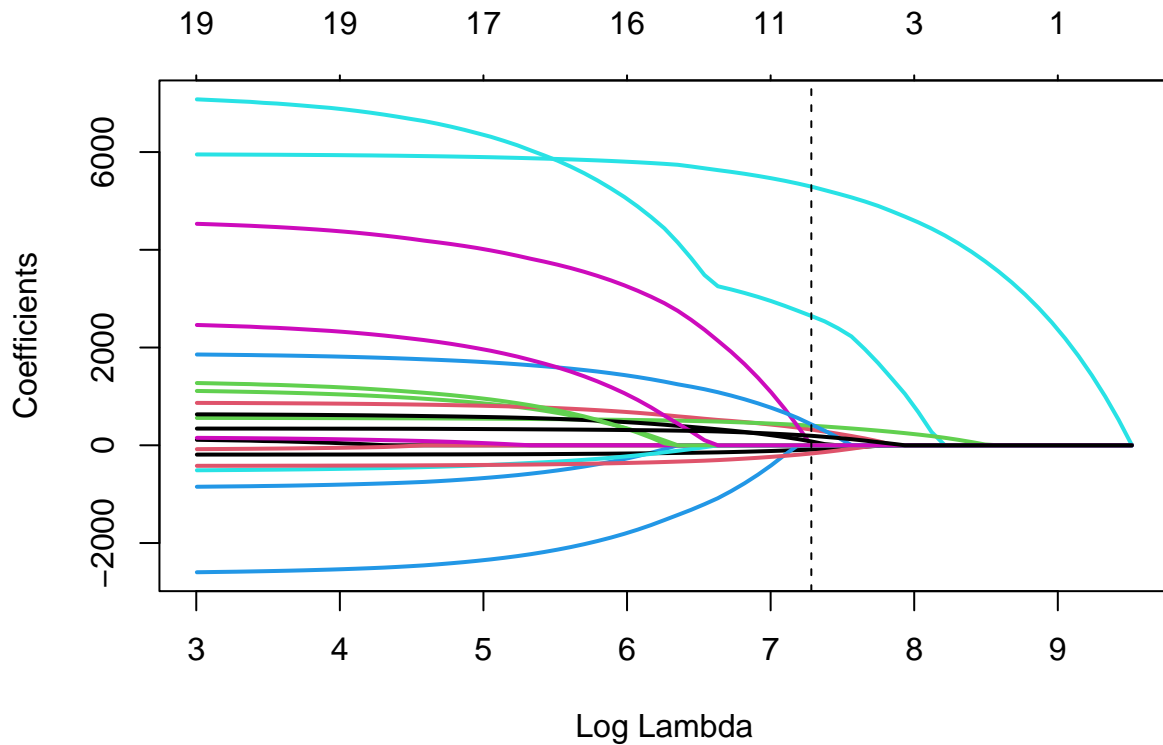
coef(smodel, s = smodel.cv$lambda.min)

## 21 x 1 sparse Matrix of class "dgCMatrix"
##              s1
## (Intercept) 35059.46413
## V1          .
## V2          .
## V3          .
## V4          .
## V5          .
## V6          2645.67778
## V7          .
## V8          -88.41895
## V9          322.11524
## V10         400.75993
## V11         .
## V12         .
## V13         .
## V14          78.39102
## V15        -167.65175
## V16         .
## V17          427.09593
## V18         5292.39994
## V19         .
## V20         193.83659

plot(smodel, xvar="lambda", lwd=2) + abline(v=log(smodel.cv$lambda.min), col="black", lty=2)

```





```
## integer(0)
```

## Result of Lasso

```
model_lasso = lm(formula = Salary ~ PerfScoreID + Position + State + Age + EmployedYear + ManagerName +
summary(model_lasso)
```

```
##
## Call:
## lm(formula = Salary ~ PerfScoreID + Position + State + Age +
##     EmployedYear + ManagerName + EmpSatisfaction + SpecialProjectsCount +
##     Absences, data = HRtrain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13917  -2974         0    3370   14353
##
## Coefficients: (10 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    50276.41   10274.10   4.894 2.46e-06 ***
## PerfScoreID         6.05     774.74   0.008  0.99378
## PositionAdministrative Assistant -10649.55   5206.11  -2.046  0.04249 *
## PositionArea Sales Manager      5476.69   7643.14   0.717  0.47473
## PositionBI Developer      31886.08   5176.27   6.160 6.00e-09 ***
## PositionBI Director    47544.52   7527.05   6.316 2.70e-09 ***
```

## PositionCIO	170433.35	11374.78	14.983	< 2e-16	***
## PositionData Analyst	26992.01	4871.15	5.541	1.26e-07	***
## PositionData Architect	85699.92	7274.28	11.781	< 2e-16	***
## PositionDatabase Administrator	49008.33	5216.64	9.395	< 2e-16	***
## PositionDirector of Operations	117304.98	9899.52	11.850	< 2e-16	***
## PositionDirector of Sales	136126.74	12737.34	10.687	< 2e-16	***
## PositionEnterprise Architect	48624.05	8843.47	5.498	1.55e-07	***
## PositionIT Director	112391.83	7199.94	15.610	< 2e-16	***
## PositionIT Manager - DB	80956.34	5988.39	13.519	< 2e-16	***
## PositionIT Manager - Infra	93847.12	7225.70	12.988	< 2e-16	***
## PositionIT Manager - Support	74595.62	7261.40	10.273	< 2e-16	***
## PositionIT Support	4869.09	5210.74	0.934	0.35153	
## PositionNetwork Engineer	-1560.16	5190.37	-0.301	0.76413	
## PositionPresident & CEO	189950.55	11981.11	15.854	< 2e-16	***
## PositionPrincipal Data Architect	57439.61	7573.25	7.585	2.88e-12	***
## PositionProduction Manager	26295.72	7888.12	3.334	0.00107	**
## PositionProduction Technician I	-13715.80	6214.83	-2.207	0.02879	*
## PositionProduction Technician II	-4546.13	6289.69	-0.723	0.47090	
## PositionSales Manager	1328.22	11633.94	0.114	0.90925	
## PositionSenior BI Developer	23789.89	7375.66	3.225	0.00153	**
## PositionShared Services Manager	40499.86	9836.31	4.117	6.21e-05	***
## PositionSoftware Engineer	33217.42	4411.68	7.529	3.92e-12	***
## PositionSoftware Engineering Manager	13987.58	7233.15	1.934	0.05496	.
## PositionSr. Accountant	42584.50	7392.94	5.760	4.39e-08	***
## PositionSr. DBA	40911.13	7453.34	5.489	1.62e-07	***
## PositionSr. Network Engineer	32920.25	5448.66	6.042	1.09e-08	***
## StateAZ	-4689.93	8963.52	-0.523	0.60157	
## StateCA	10670.11	9090.92	1.174	0.24231	
## StateCO	1578.91	8895.82	0.177	0.85936	
## StateCT	2099.71	8007.92	0.262	0.79351	
## StateFL	2806.73	8930.65	0.314	0.75373	
## StateGA	2120.12	8940.00	0.237	0.81285	
## StateID	-842.00	8922.36	-0.094	0.92494	
## StateIN	-2391.23	8990.25	-0.266	0.79061	
## StateKY	-898.22	8925.37	-0.101	0.91997	
## StateMA	9781.05	7985.12	1.225	0.22247	
## StateME	-6531.63	8911.10	-0.733	0.46468	
## StateMT	-3501.51	8965.76	-0.391	0.69667	
## StateNC	4764.88	8940.07	0.533	0.59481	
## StateND	1419.66	9112.56	0.156	0.87640	
## StateNH	8013.52	8923.89	0.898	0.37059	
## StateNV	-2788.16	8736.52	-0.319	0.75005	
## StateNY	5814.76	8954.69	0.649	0.51707	
## StateOH	-2660.75	8946.80	-0.297	0.76656	
## StateOR	-5653.21	8916.89	-0.634	0.52702	
## StatePA	12099.44	12546.58	0.964	0.33637	
## StateRI	NA	NA	NA	NA	
## StateTN	7025.72	9008.67	0.780	0.43665	
## StateTX	1590.52	8893.21	0.179	0.85829	
## StateUT	7850.96	8937.05	0.878	0.38105	
## StateVA	9393.67	8992.10	1.045	0.29781	
## StateVT	8169.76	8899.30	0.918	0.36003	
## StateWA	-2445.89	8883.12	-0.275	0.78342	
## Age	57.83	60.41	0.957	0.33994	

```
## EmployedYear          105.37      194.85    0.541  0.58944
## ManagerNameAmy Dunn   3113.48     2286.08    1.362  0.17520
## ManagerNameBoard of Directors -7977.34    8893.19   -0.897  0.37110
## ManagerNameBrandon R. LeBlanc      NA         NA      NA      NA
## ManagerNameBrannon Miller    4595.41    2325.25    1.976  0.04989 *
## ManagerNameBrian Champaign      NA         NA      NA      NA
## ManagerNameDavid Stanley    2247.37    2383.91    0.943  0.34729
## ManagerNameDebra Houlihan      NA         NA      NA      NA
## ManagerNameElijah Gray     4722.65    2360.18    2.001  0.04714 *
## ManagerNameEric Dougall    -2961.62    5860.52   -0.505  0.61403
## ManagerNameJanet King    -15323.54    8125.15   -1.886  0.06117 .
## ManagerNameJennifer Zamora      NA         NA      NA      NA
## ManagerNameJohn Smith      NA         NA      NA      NA
## ManagerNameKelley Spirea    2766.16    2294.09    1.206  0.22974
## ManagerNameKetsia Liebig      904.16    2310.53    0.391  0.69610
## ManagerNameKissy Sullivan    2598.42    2190.08    1.186  0.23726
## ManagerNameLynn Daneault      NA         NA      NA      NA
## ManagerNameMichael Albert    2308.66    2436.32    0.948  0.34481
## ManagerNamePeter Monroe      NA         NA      NA      NA
## ManagerNameSimon Roup      NA         NA      NA      NA
## ManagerNameWebster Butler      NA         NA      NA      NA
## EmpSatisfaction        564.66     532.20    1.061  0.29034
## SpecialProjectsCount    -574.63     968.30   -0.593  0.55375
## Absences              143.45      80.92    1.773  0.07823 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6166 on 155 degrees of freedom
## Multiple R-squared:  0.9662, Adjusted R-squared:  0.9503
## F-statistic: 60.78 on 73 and 155 DF,  p-value: < 2.2e-16
```

## Check regression assumptions for model\_lasso

```
#plot(fitted(model_lasso),residuals(model_lasso),xlab="Fitted values",ylab="Residuals")
#abline(h=0, col="red")
#lines(lowess(model_lasso$fitted.values, residuals(model_lasso)), col='blue')

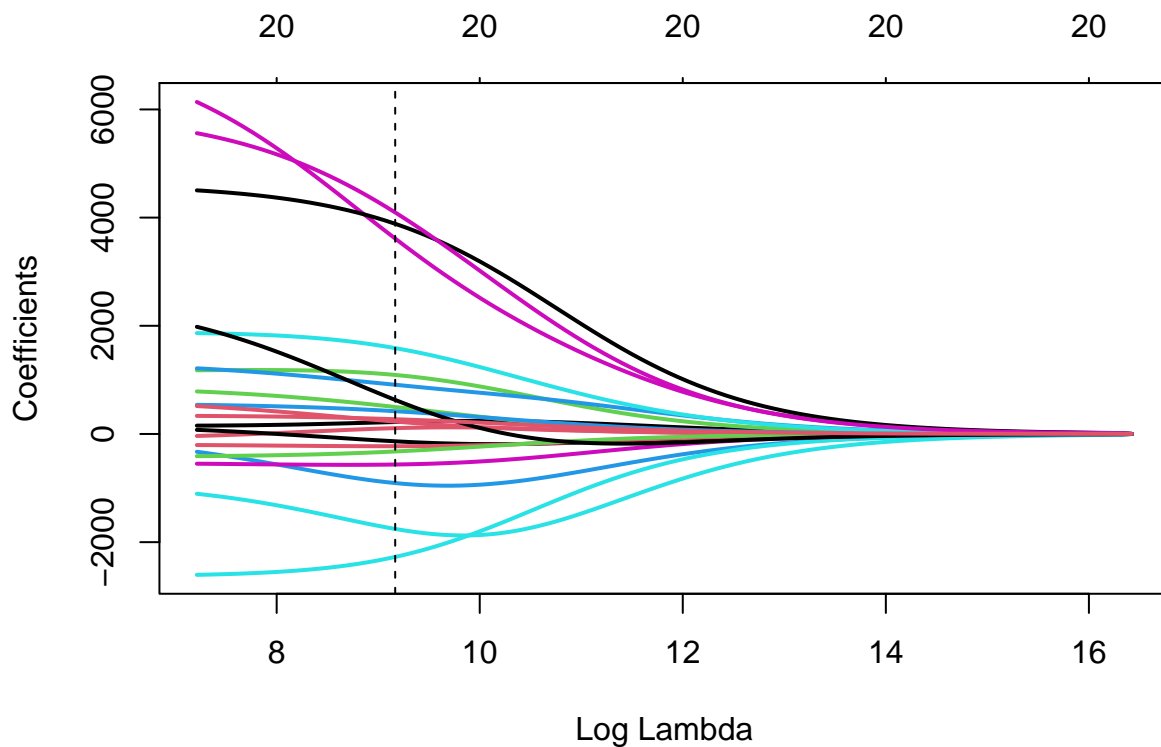
# Normality
#hist(residuals(model_lasso), xlab="residuals", col="orange",main=NULL, nclass=15)
#qqPlot(residuals(model_lasso), xlab="normal quantiles", ylab="residuals")

# Ridge Regression
smodel2.cv = cv.glmnet(X_pred, HRtrain$Salary, alpha = 0, nfolds = 10)
smodel2 = glmnet(X_pred, HRtrain$Salary, alpha = 0, nlambda=100)
coef(smodel2, s = smodel2.cv$lambda.min)

## 21 x 1 sparse Matrix of class "dgCMatrix"
##              s1
## (Intercept) 36411.4559
## V1          220.5822
## V2          107.5987
## V3          1091.1768
## V4          -907.9381
```

```
## V5      -1756.0234
## V6      3612.5031
## V7      -135.6206
## V8      -226.3703
## V9       498.5359
## V10     420.5696
## V11    -2273.1762
## V12     -564.6980
## V13     3885.8846
## V14      229.1098
## V15     -323.0585
## V16      909.7769
## V17     1587.1954
## V18     4095.0793
## V19      629.4255
## V20      272.5741
```

```
plot(smodel2,xvar="lambda",lwd=2) + abline(v=log(smodel2.cv$lambda.min),col="black",lty=2)
```



```
## integer(0)
```

```
model_ridge = glmnet(X_pred, HRtrain$Salary, alpha = 0,family="gaussian")
summary(model_ridge)
```

```
##      Length Class      Mode
## a0      100  -none-    numeric
## beta    2000 dgCMatrix S4
## df       100  -none-    numeric
```

```
## dim          2  -none-   numeric
## lambda      100 -none-   numeric
## dev.ratio   100 -none-   numeric
## nulldev      1  -none-   numeric
## npasses      1  -none-   numeric
## jerr         1  -none-   numeric
## offset       1  -none-   logical
## call         5  -none-   call
## nobs         1  -none-   numeric
```

Ridge Regression creates full model

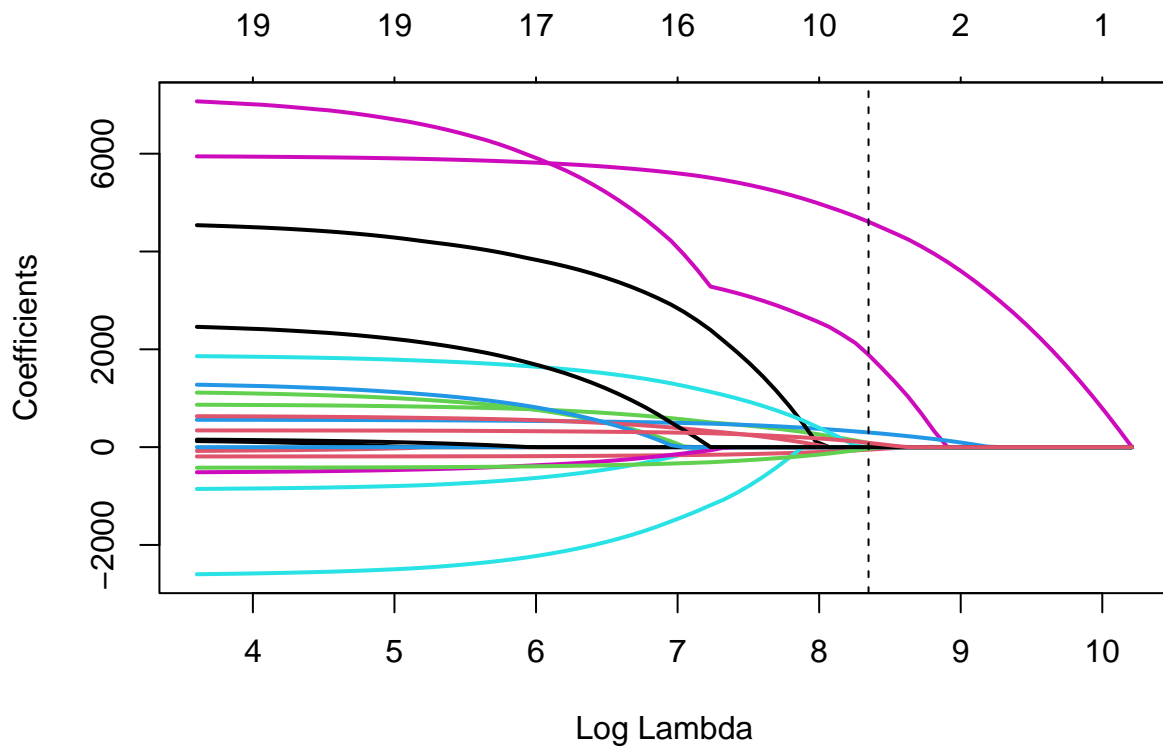
```
# Elastic Net Regression
```

```
smodel3.cv = cv.glmnet(X_pred, HRtrain$Salary, alpha = 0.5, nfolds = 10)
smodel3 = glmnet(X_pred, HRtrain$Salary, alpha = 0.5, nlambda=100)
coef(smodel3, s = smodel3.cv$lambda.min)
```

```
## 21 x 1 sparse Matrix of class "dgCMatrix"
```

```
##              s1
## (Intercept) 46013.40459
## V1          .
## V2          .
## V3          .
## V4          .
## V5          .
## V6          1889.57289
## V7          .
## V8         -51.56577
## V9          83.76383
## V10         299.55245
## V11         .
## V12         .
## V13         .
## V14         .
## V15        -30.22105
## V16         .
## V17         .
## V18         4608.15514
## V19         .
## V20         87.70521
```

```
plot(smodel3, xvar="lambda", lwd=2) + abline(v=log(smodel3.cv$lambda.min), col="black", lty=2)
```



```
## integer(0)
```

```
#Result of Elastic net
```

```
model_net = lm(formula = Salary ~ PerfScoreID + Position + State + Age + EmployedYear + EmpSatisfaction
summary(model_net)
```

```
##
```

```
## Call:
```

```
## lm(formula = Salary ~ PerfScoreID + Position + State + Age +
##     EmployedYear + EmpSatisfaction + Absences, data = HRtrain)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -13694  -3308         0    3332  14814
```

```
##
```

```
## Coefficients: (1 not defined because of singularities)
```

```
##
```

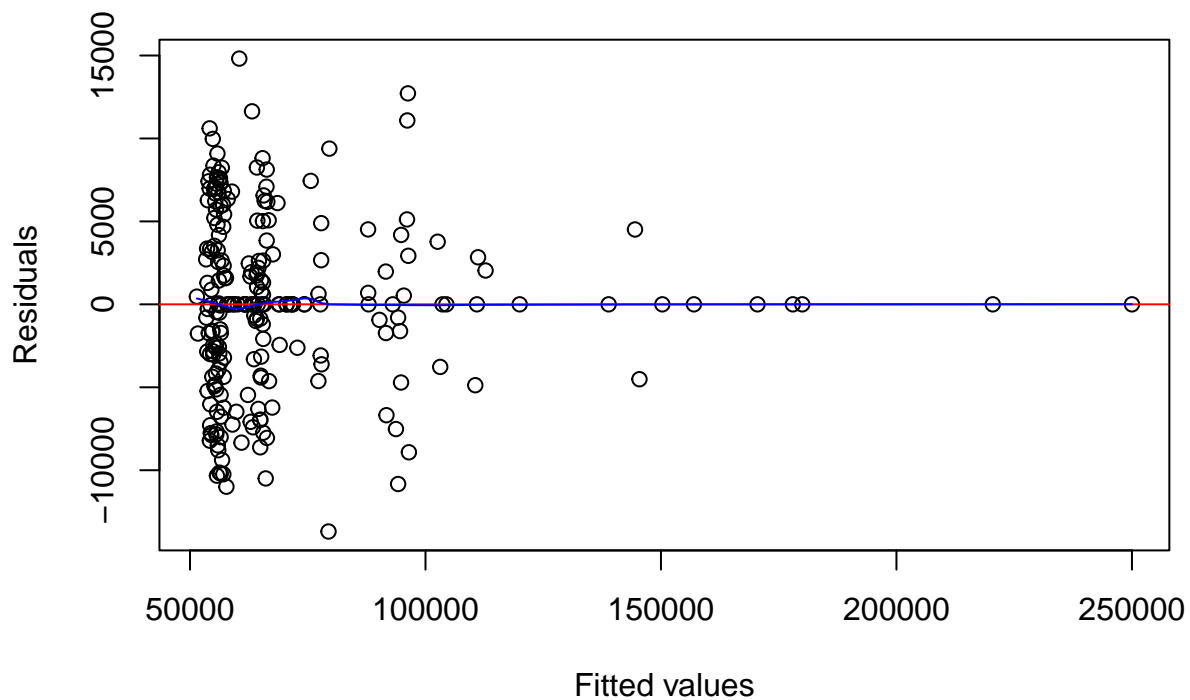
	Estimate	Std. Error	t value	Pr(> t )	
## (Intercept)	46752.60	9111.10	5.131	7.92e-07	***
## PerfScoreID	-32.11	757.41	-0.042	0.966233	
## PositionAdministrative Assistant	-10357.43	5168.02	-2.004	0.046672	*
## PositionArea Sales Manager	8502.13	5991.35	1.419	0.157744	
## PositionBI Developer	31471.24	5106.11	6.163	5.16e-09	***
## PositionBI Director	46169.09	7232.26	6.384	1.64e-09	***
## PositionCIO	154617.36	7203.88	21.463	< 2e-16	***
## PositionData Analyst	26497.92	4781.96	5.541	1.15e-07	***
## PositionData Architect	85388.81	7256.62	11.767	< 2e-16	***
## PositionDatabase Administrator	48609.04	5174.41	9.394	< 2e-16	***

## PositionDirector of Operations	105018.64	7200.44	14.585	< 2e-16	***
## PositionDirector of Sales	123903.18	10755.97	11.519	< 2e-16	***
## PositionEnterprise Architect	47981.41	8288.63	5.789	3.43e-08	***
## PositionIT Director	112472.36	7189.85	15.643	< 2e-16	***
## PositionIT Manager - DB	79792.60	5770.39	13.828	< 2e-16	***
## PositionIT Manager - Infra	93226.35	7165.78	13.010	< 2e-16	***
## PositionIT Manager - Support	74367.84	7247.82	10.261	< 2e-16	***
## PositionIT Support	3390.85	4565.14	0.743	0.458664	
## PositionNetwork Engineer	-2112.21	5126.29	-0.412	0.680843	
## PositionPresident & CEO	184211.40	7358.40	25.034	< 2e-16	***
## PositionPrincipal Data Architect	55948.91	7248.24	7.719	1.02e-12	***
## PositionProduction Manager	13556.36	4241.99	3.196	0.001668	**
## PositionProduction Technician I	-8468.85	3660.77	-2.313	0.021919	*
## PositionProduction Technician II	1000.66	3755.47	0.266	0.790218	
## PositionSales Manager	4803.62	10599.67	0.453	0.651004	
## PositionSenior BI Developer	22866.32	7270.59	3.145	0.001966	**
## PositionShared Services Manager	28068.36	7206.21	3.895	0.000142	***
## PositionSoftware Engineer	31489.35	4305.89	7.313	1.04e-11	***
## PositionSoftware Engineering Manager	13550.02	7218.47	1.877	0.062244	.
## PositionSr. Accountant	39933.53	5640.34	7.080	3.81e-11	***
## PositionSr. DBA	40041.99	7327.75	5.464	1.66e-07	***
## PositionSr. Network Engineer	32382.39	5164.26	6.270	2.97e-09	***
## StateAZ	-4191.78	8947.45	-0.468	0.640046	
## StateCA	11271.42	9065.96	1.243	0.215512	
## StateCO	1937.82	8879.19	0.218	0.827506	
## StateCT	1661.23	7935.21	0.209	0.834430	
## StateFL	3151.54	8916.02	0.353	0.724182	
## StateGA	2589.31	8921.67	0.290	0.772003	
## StateID	-1170.18	8902.21	-0.131	0.895579	
## StateIN	-1819.67	8967.33	-0.203	0.839443	
## StateKY	-384.34	8907.11	-0.043	0.965634	
## StateMA	10309.94	7904.10	1.304	0.193899	
## StateME	-6052.53	8895.26	-0.680	0.497179	
## StateMT	-2945.41	8947.57	-0.329	0.742428	
## StateNC	5189.65	8920.52	0.582	0.561509	
## StateND	1516.00	9076.93	0.167	0.867559	
## StateNH	8500.62	8905.77	0.955	0.341207	
## StateNV	-2875.14	8728.00	-0.329	0.742254	
## StateNY	6061.48	8936.36	0.678	0.498524	
## StateOH	-2658.04	8923.99	-0.298	0.766185	
## StateOR	-5672.81	8901.42	-0.637	0.524807	
## StatePA	12051.71	12524.65	0.962	0.337321	
## StateRI	NA	NA	NA	NA	
## StateTN	7165.31	8983.03	0.798	0.426206	
## StateTX	1892.75	8877.02	0.213	0.831416	
## StateUT	8278.96	8921.87	0.928	0.354778	
## StateVA	9444.87	8969.66	1.053	0.293871	
## StateVT	8169.50	8881.84	0.920	0.359005	
## StateWA	-2242.07	8869.43	-0.253	0.800744	
## Age	76.75	58.85	1.304	0.193988	
## EmployedYear	91.98	183.71	0.501	0.617254	
## EmpSatisfaction	496.61	519.29	0.956	0.340295	
## Absences	136.27	79.14	1.722	0.086950	.
## ---					

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6161 on 167 degrees of freedom
## Multiple R-squared:  0.9637, Adjusted R-squared:  0.9504
## F-statistic: 72.67 on 61 and 167 DF,  p-value: < 2.2e-16
```

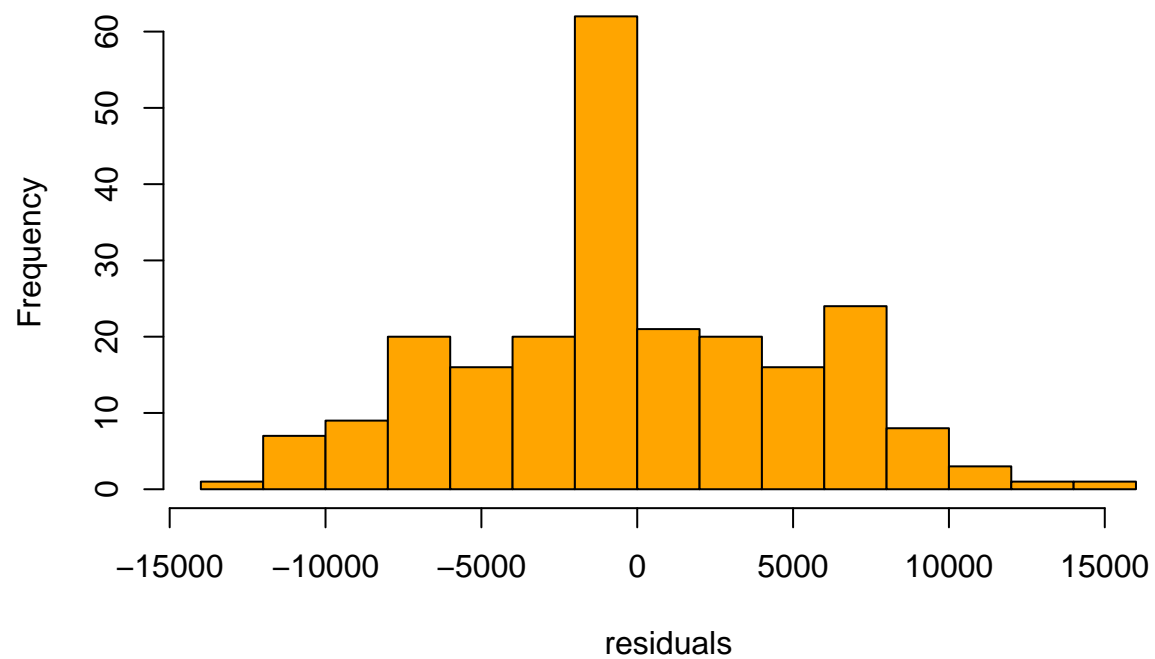
## Elastic net model assumptions test

```
# Constant Variance / Uncorrelated errors
plot(fitted(model_net),residuals(model_net),xlab="Fitted values",ylab="Residuals")
abline(h=0, col="red")
lines(lowess(model_net$fitted.values, residuals(model_net)), col='blue')
```

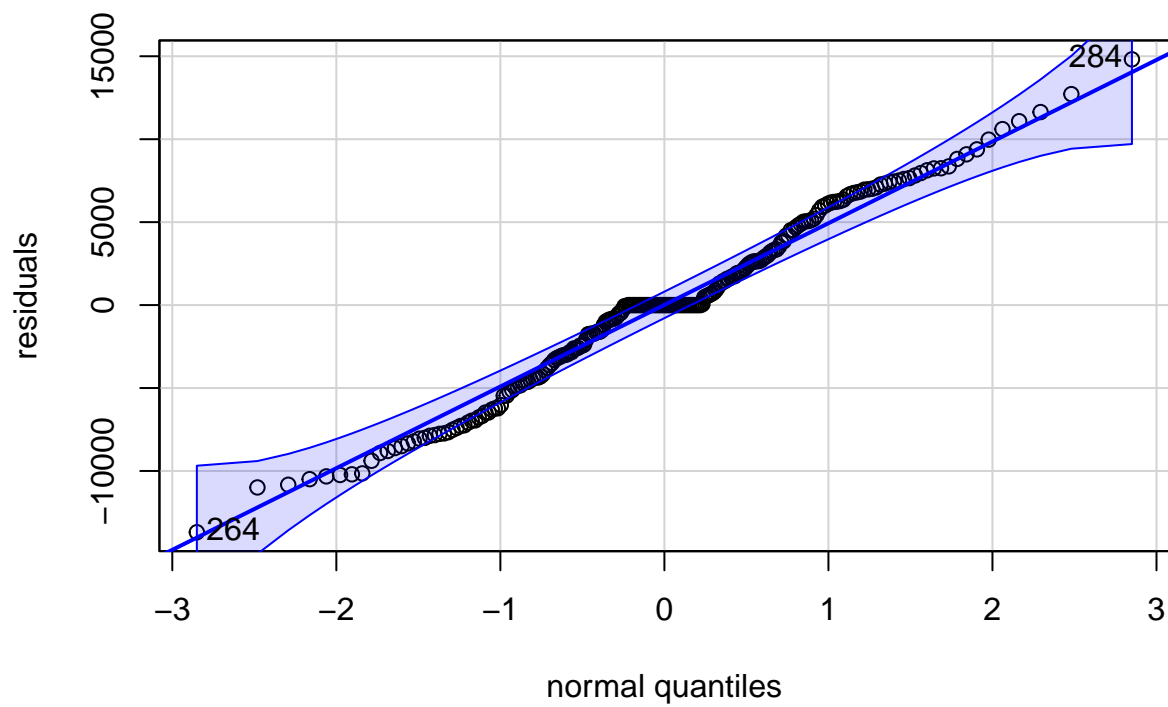


```
# Normality
hist(residuals(model_net), xlab="residuals", col="orange",main=NULL, nclass=15)
```





```
qqPlot(residuals(model_net), xlab="normal quantiles", ylab="residuals")
```



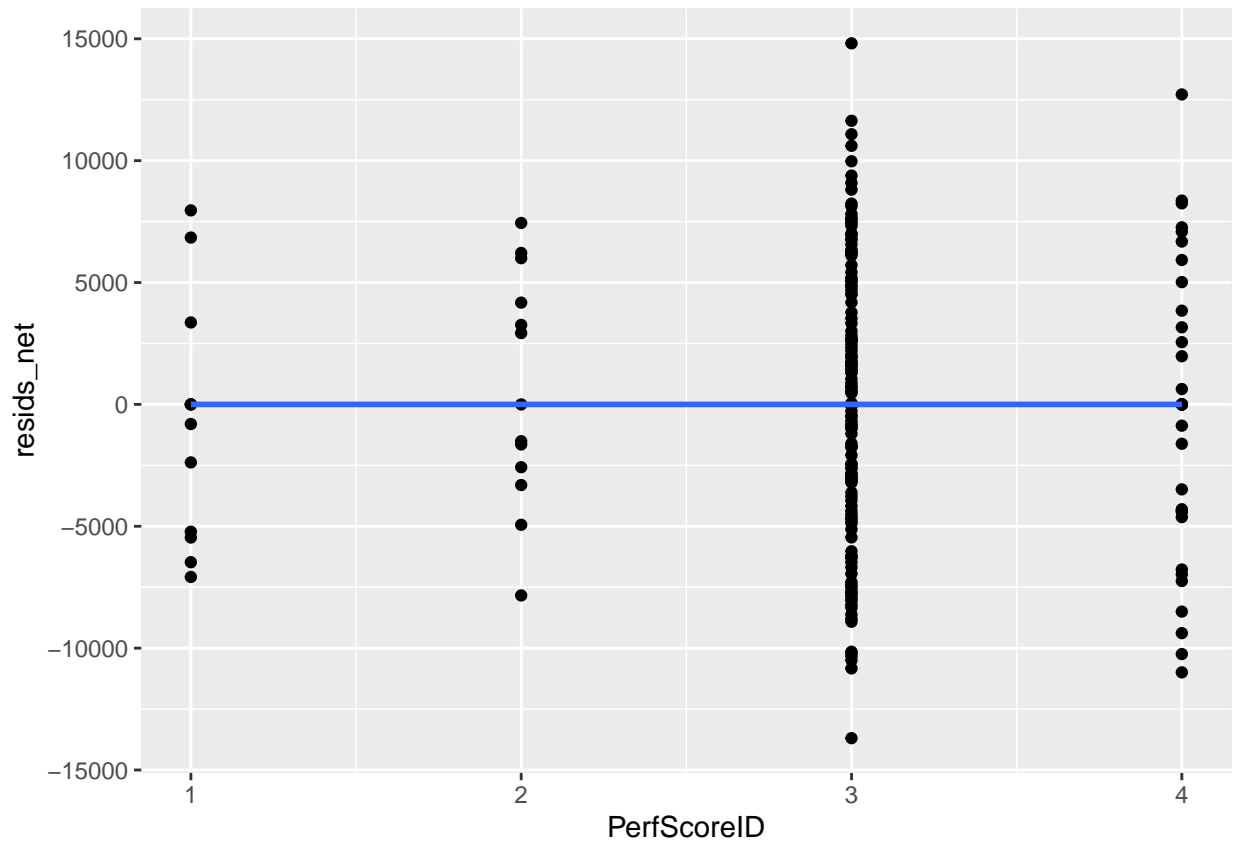
```
## 284 264
## 201 188
```

```
# Linearity
```

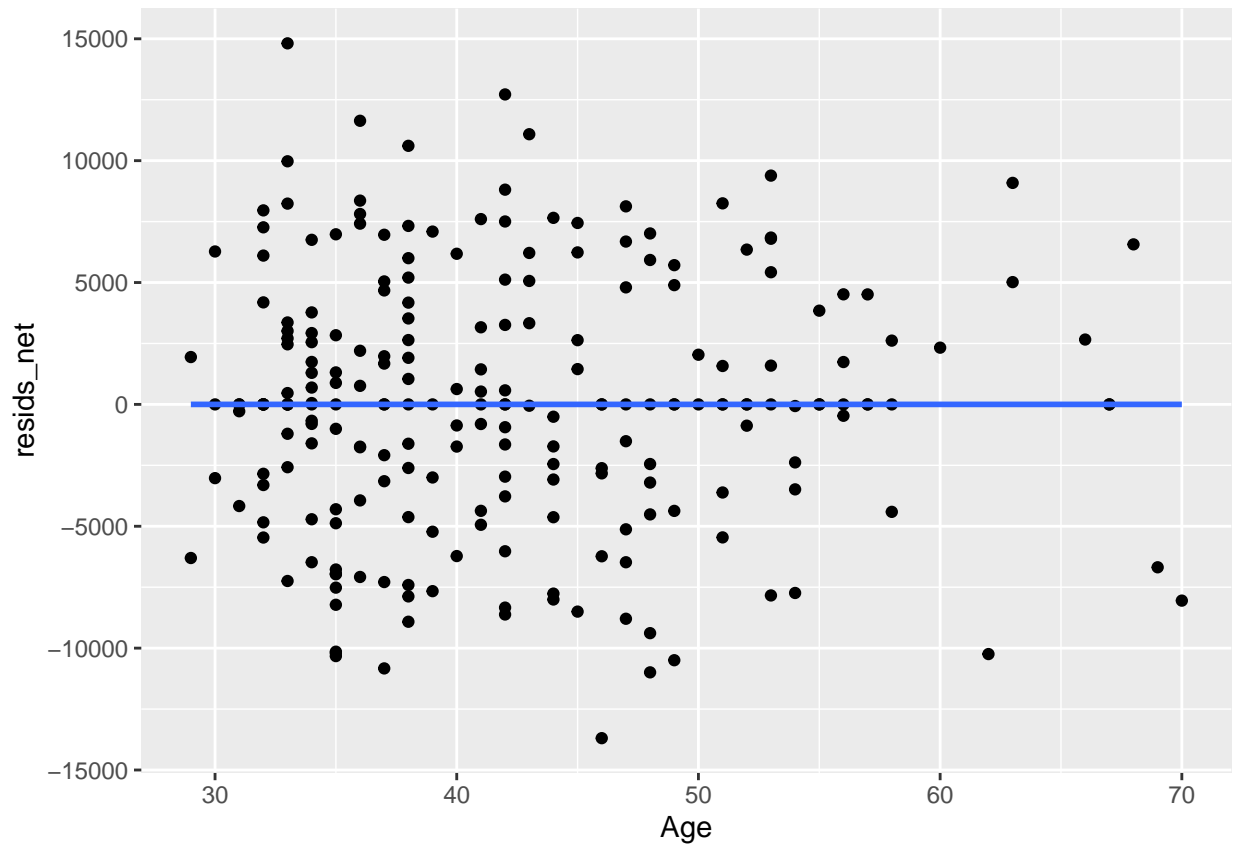
```
resids_net = model_net$residuals
```

```
ggplot(HRtrain, aes(x=PerfScoreID, resids_net)) + geom_point() + geom_smooth(method=lm, se=FALSE, fullr
```

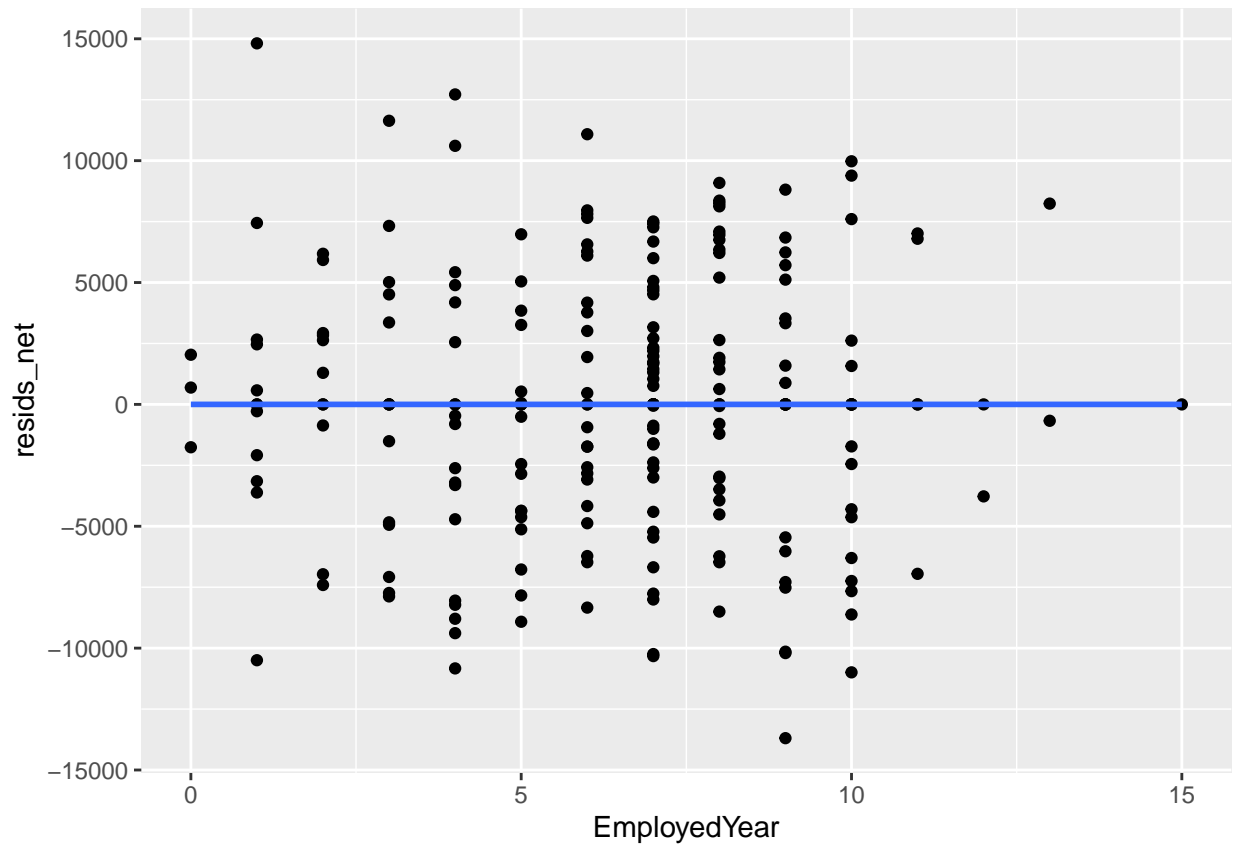
```
## `geom_smooth()` using formula 'y ~ x'
```



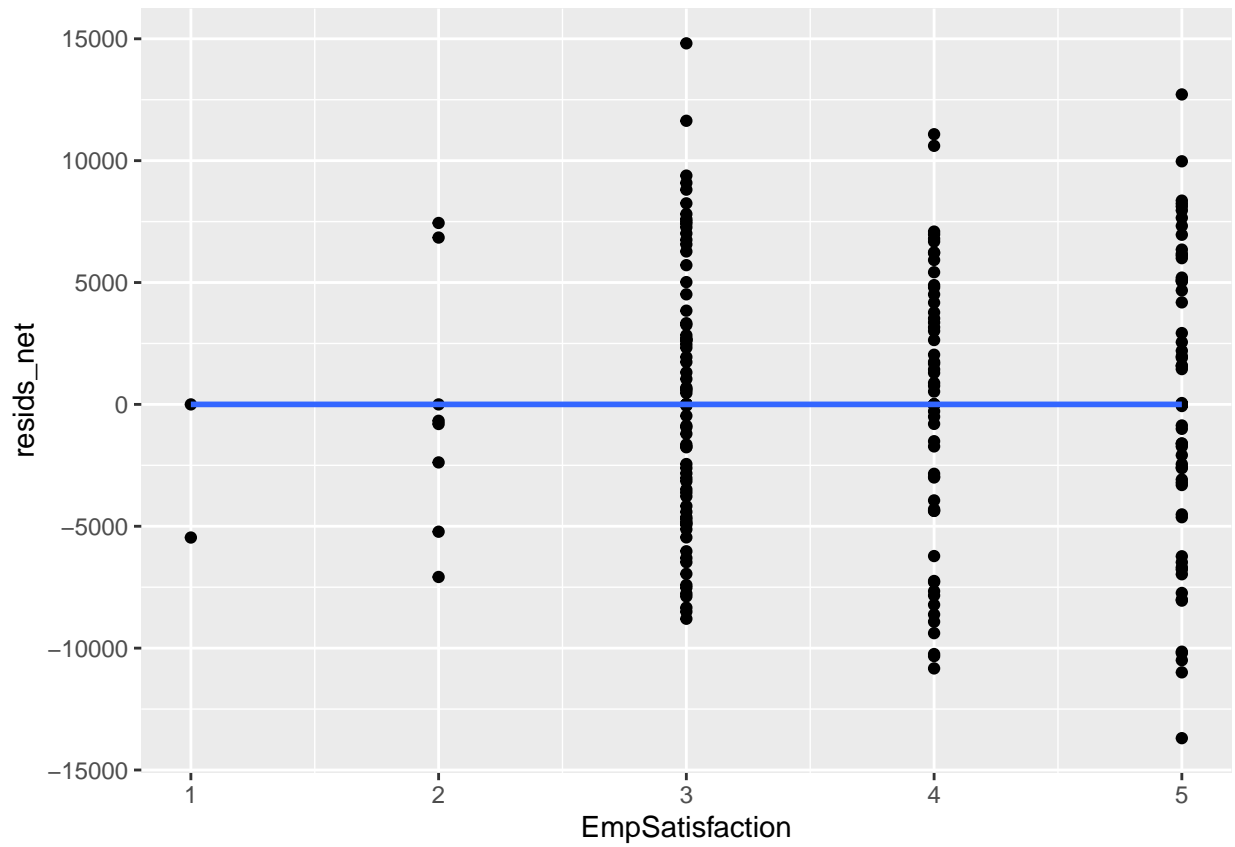
```
ggplot(HRtrain, aes(x=Age, resids_net)) + geom_point() + geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
## `geom_smooth()` using formula 'y ~ x'
```



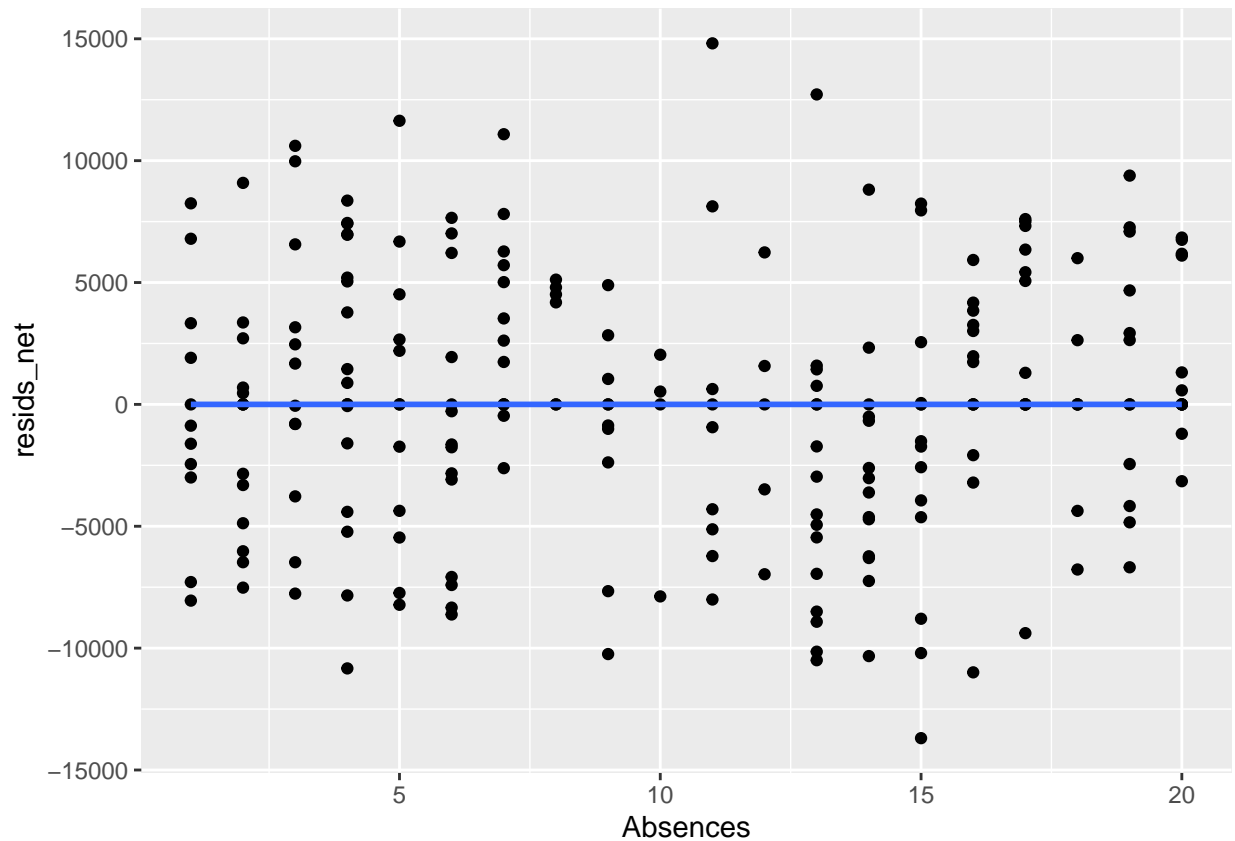
```
ggplot(HRtrain, aes(x=EmployedYear, residuals_net)) + geom_point() + geom_smooth(method=lm, se=FALSE, full.
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(HRtrain, aes(x=EmpSatisfaction, resids_net)) + geom_point() + geom_smooth(method=lm, se=FALSE, f  
## `geom_smooth()` using formula 'y ~ x'
```



```
ggplot(HRtrain, aes(x=Absences, resid_net)) + geom_point() + geom_smooth(method=lm, se=FALSE, fullrange=TRUE)
## `geom_smooth()` using formula 'y ~ x'
```



## Variable Selection Criteria

```

model_position = lm(Salary~Position,data=HRtrain)
n=nrow(HRtrain)

no=c(
  nrow(summary(model_full)$coefficients),
  nrow(summary(model_backward)$coefficients),
  nrow(summary(model_forward)$coefficients),
  nrow(summary(model_lasso)$coefficients),
  nrow(summary(model_net)$coefficients),
  nrow(summary(model_position)$coefficients)
)

rsquared=c(
  summary(model_full)$adj.r.squared,
  summary(model_backward)$adj.r.squared,
  summary(model_forward)$adj.r.squared,
  summary(model_lasso)$adj.r.squared,
  summary(model_net)$adj.r.squared,
  summary(model_position)$adj.r.squared
)

cp=c(

```

```

Cp(model_full, S2= summary(model_full)$sigma^2),
Cp(model_backward, S2= summary(model_full)$sigma^2),
Cp(model_forward, S2= summary(model_full)$sigma^2),
Cp(model_lasso, S2= summary(model_full)$sigma^2),
Cp(model_net, S2= summary(model_full)$sigma^2),
  Cp(model_position, S2= summary(model_full)$sigma^2)
)
aic=c(
  AIC(model_full,k=2),
  AIC(model_backward,k=2),
  AIC(model_forward,k=2),
  AIC(model_lasso,k=2),
  AIC(model_net,k=2),
  AIC(model_position,k=2)
)
bic=c(
  AIC(model_full,k=log(n)),
  AIC(model_backward,k=log(n)),
  AIC(model_forward,k=log(n)),
  AIC(model_lasso,k=log(n)),
  AIC(model_net,k=log(n)),
  AIC(model_position,k=log(n))
)
criteria = data.frame( "Number of Predictors"=no,"Adjusted R-Squared"=rsquared ,"Mellow Cp"=cp, AIC=aic,
row.names(criteria)=c("Full Model" , "Stepwise Backward", "Stepwise Forward", "Lasso", "Elastic Net","Position")
criteria

```

##	Number.of.Predictors	Adjusted.R.Squared	Mellow.Cp	AIC
## Full Model	97	0.9502852	131.00000	4716.897
## Stepwise Backward	35	0.9525015	26.35127	4670.631
## Stepwise Forward	35	0.9527352	25.43945	4669.501
## Lasso	74	0.9503487	93.80191	4707.386
## Elastic Net	62	0.9504364	63.49194	4700.058
## Position	31	0.9504017	30.53580	4677.211

##	BIC
## Full Model	5053.401
## Stepwise Backward	4794.245
## Stepwise Forward	4793.115
## Lasso	4964.916
## Elastic Net	4916.382
## Position	4787.090

#Prediction / Testing

```
predfull = predict.lm(model_full,HRtest)
```

```
## Warning in predict.lm(model_full, HRtest): prediction from a rank-deficient fit
## may be misleading
```

```
predbackward = predict.lm(model_backward,HRtest)
predforward = predict.lm(model_forward,HRtest)
predlasso = predict.lm(model_lasso,HRtest)
```

```
## Warning in predict.lm(model_lasso, HRtest): prediction from a rank-deficient fit
## may be misleading
```



```

prednet = predict.lm(model_net,HRtest)

## Warning in predict.lm(model_net, HRtest): prediction from a rank-deficient fit
## may be misleading

predpos = predict.lm(model_position,HRtest)

#MSPE
MSPE=c(
  mean( (HRtest$Salary - predfull)^2 ),
  mean( (HRtest$Salary - predbackward)^2 ),
  mean( (HRtest$Salary - predforward)^2 ),
  mean( (HRtest$Salary - predlasso)^2 ),
  mean( (HRtest$Salary - prednet)^2 ),
  mean( (HRtest$Salary - predpos)^2 )
)

#MAE
MAE=c(
  mean( abs(HRtest$Salary - predfull) ),
  mean( abs(HRtest$Salary - predbackward) ),
  mean( abs(HRtest$Salary - predforward) ),
  mean( abs(HRtest$Salary - predlasso) ),
  mean( abs(HRtest$Salary - prednet) ),
  mean( abs(HRtest$Salary - predpos) )
)

#MAPE
MAPE=c(
  mean( abs(HRtest$Salary - predfull)/ HRtest$Salary),
  mean( abs(HRtest$Salary - predbackward)/ HRtest$Salary),
  mean( abs(HRtest$Salary - predforward)/ HRtest$Salary),
  mean( abs(HRtest$Salary - predlasso)/ HRtest$Salary),
  mean( abs(HRtest$Salary - prednet)/ HRtest$Salary),
  mean( abs(HRtest$Salary - predpos)/ HRtest$Salary)
)

#PM
PM =c(
  sum( (HRtest$Salary - predfull)^2 ) / sum( (HRtest$Salary - mean(HRtest$Salary) )^2 ),
  sum( (HRtest$Salary - predbackward)^2 ) / sum( (HRtest$Salary - mean(HRtest$Salary) )^2 ),
  sum( (HRtest$Salary - predforward)^2 ) / sum( (HRtest$Salary - mean(HRtest$Salary) )^2 ),
  sum( (HRtest$Salary - predlasso)^2 ) / sum( (HRtest$Salary - mean(HRtest$Salary) )^2 ),
  sum( (HRtest$Salary - prednet)^2 ) / sum( (HRtest$Salary - mean(HRtest$Salary) )^2 ),
  sum( (HRtest$Salary - predpos)^2 ) / sum( (HRtest$Salary - mean(HRtest$Salary) )^2 )
)

performance = data.frame( MSPE=MSPE ,MAE=MAE, MAPE=MAPE, PM=PM)
row.names(performance)=c("Full Model" , "Stepwise Backward", "Stepwise Forward", "Lasso", "Elastic Net")
performance

##              MSPE      MAE      MAPE      PM
## Full Model      55078189 5456.196 0.09313196 0.06679844
## Stepwise Backward 39726001 4816.832 0.08141588 0.04817941

```

```
## Stepwise Forward 43125080 5047.964 0.08492801 0.05230179
## Lasso            39829696 4756.452 0.08021371 0.04830517
## Elastic Net      39309436 4686.834 0.07868069 0.04767420
## Position         40701550 5132.405 0.08545892 0.04936255
```

## Check Multicollinearity

```
df <- data.frame(HRtrain$PerfScoreID, HRtrain$Age ,HRtrain$EmployedYear, HRtrain$EmpSatisfaction, HRtrain$Absences, HRtrain$Salary)
cor(df)
```

```
##              HRtrain.PerfScoreID HRtrain.Age HRtrain.EmployedYear
## HRtrain.PerfScoreID              1.00000000  0.09855999          0.115053177
## HRtrain.Age                    0.09855999  1.00000000          -0.021298729
## HRtrain.EmployedYear            0.11505318 -0.02129873          1.000000000
## HRtrain.EmpSatisfaction          0.29842552 -0.05334735          0.009218585
## HRtrain.Absences                0.07994592 -0.03712606          0.002687228
## HRtrain.Salary                  0.14596085  0.17524732          0.025044339
##              HRtrain.EmpSatisfaction HRtrain.Absences HRtrain.Salary
## HRtrain.PerfScoreID              0.298425521      0.079945922      0.14596085
## HRtrain.Age                    -0.053347349      -0.037126061      0.17524732
## HRtrain.EmployedYear            0.009218585      0.002687228      0.02504434
## HRtrain.EmpSatisfaction          1.000000000      0.089681478      0.09190903
## HRtrain.Absences                0.089681478      1.000000000      0.09301789
## HRtrain.Salary                  0.091909031      0.093017889      1.00000000
cat("VIF Threshold for model_net:", max(10, 1/(1-summary(model_net)$r.squared)), "\n")
```

```
## VIF Threshold for model_net: 27.54581
```

```
HRmulti<- HRtrain[-c(221),]
model_multi= lm(formula = Salary ~ PerfScoreID + State + Position + Age + EmployedYear + EmpSatisfaction + Absences, data = HRmulti)
summary(model_multi)
```

```
##
## Call:
## lm(formula = Salary ~ PerfScoreID + State + Position + Age +
##     EmployedYear + EmpSatisfaction + Absences, data = HRmulti)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13694  -3353       0    3339   14814
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    46752.60     9111.10   5.131 7.92e-07 ***
## PerfScoreID      -32.11       757.41  -0.042 0.966233
## StateAZ        -4191.78     8947.45  -0.468 0.640046
## StateCA        11271.42     9065.96   1.243 0.215512
## StateCO         1937.82     8879.19   0.218 0.827506
## StateCT         1661.23     7935.21   0.209 0.834430
## StateFL         3151.54     8916.02   0.353 0.724182
## StateGA         2589.31     8921.67   0.290 0.772003
## StateID        -1170.18     8902.21  -0.131 0.895579
## StateIN        -1819.67     8967.33  -0.203 0.839443
## StateKY        -384.34     8907.11  -0.043 0.965634
```

```

## StateMA          10309.94    7904.10    1.304 0.193899
## StateME          -6052.53    8895.26   -0.680 0.497179
## StateMT          -2945.41    8947.57   -0.329 0.742428
## StateNC           5189.65    8920.52    0.582 0.561509
## StateND           1516.00    9076.93    0.167 0.867559
## StateNH           8500.62    8905.77    0.955 0.341207
## StateNV          -2875.14    8728.00   -0.329 0.742254
## StateNY           6061.48    8936.36    0.678 0.498524
## StateOH          -2658.04    8923.99   -0.298 0.766185
## StateOR          -5672.81    8901.42   -0.637 0.524807
## StatePA          12051.71   12524.65    0.962 0.337321
## StateTN           7165.31    8983.03    0.798 0.426206
## StateTX           1892.75    8877.02    0.213 0.831416
## StateUT           8278.96    8921.87    0.928 0.354778
## StateVA           9444.87    8969.66    1.053 0.293871
## StateVT           8169.50    8881.84    0.920 0.359005
## StateWA          -2242.07    8869.43   -0.253 0.800744
## PositionAdministrative Assistant -10357.43   5168.02   -2.004 0.046672 *
## PositionArea Sales Manager      8502.13   5991.35    1.419 0.157744
## PositionBI Developer      31471.24   5106.11    6.163 5.16e-09 ***
## PositionBI Director      46169.09   7232.26    6.384 1.64e-09 ***
## PositionCIO      154617.36   7203.88   21.463 < 2e-16 ***
## PositionData Analyst      26497.92   4781.96    5.541 1.15e-07 ***
## PositionData Architect      85388.81   7256.62   11.767 < 2e-16 ***
## PositionDatabase Administrator  48609.04   5174.41    9.394 < 2e-16 ***
## PositionDirector of Operations 105018.64   7200.44   14.585 < 2e-16 ***
## PositionEnterprise Architect   47981.41   8288.63    5.789 3.43e-08 ***
## PositionIT Director   112472.36   7189.85   15.643 < 2e-16 ***
## PositionIT Manager - DB      79792.60   5770.39   13.828 < 2e-16 ***
## PositionIT Manager - Infra   93226.35   7165.78   13.010 < 2e-16 ***
## PositionIT Manager - Support  74367.84   7247.82   10.261 < 2e-16 ***
## PositionIT Support      3390.85   4565.14    0.743 0.458664
## PositionNetwork Engineer   -2112.21   5126.29   -0.412 0.680843
## PositionPresident & CEO    184211.40   7358.40   25.034 < 2e-16 ***
## PositionPrincipal Data Architect 55948.91   7248.24    7.719 1.02e-12 ***
## PositionProduction Manager   13556.36   4241.99    3.196 0.001668 **
## PositionProduction Technician I -8468.85   3660.77   -2.313 0.021919 *
## PositionProduction Technician II 1000.66   3755.47    0.266 0.790218
## PositionSales Manager      4803.62   10599.67    0.453 0.651004
## PositionSenior BI Developer   22866.32   7270.59    3.145 0.001966 **
## PositionShared Services Manager 28068.36   7206.21    3.895 0.000142 ***
## PositionSoftware Engineer   31489.35   4305.89    7.313 1.04e-11 ***
## PositionSoftware Engineering Manager 13550.02   7218.47    1.877 0.062244 .
## PositionSr. Accountant      39933.53   5640.34    7.080 3.81e-11 ***
## PositionSr. DBA      40041.99   7327.75    5.464 1.66e-07 ***
## PositionSr. Network Engineer 32382.39   5164.26    6.270 2.97e-09 ***
## Age              76.75     58.85    1.304 0.193988
## EmployedYear      91.98    183.71    0.501 0.617254
## EmpSatisfaction     496.61    519.29    0.956 0.340295
## Absences           136.27     79.14    1.722 0.086950 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6161 on 167 degrees of freedom

```

```
## Multiple R-squared:  0.961, Adjusted R-squared:  0.947
## F-statistic: 68.58 on 60 and 167 DF,  p-value: < 2.2e-16
```

```
summary(model_multi)$r.squared
```

```
## [1] 0.9609993
```

```
vif(model_multi)
```

```
##              GVIF Df GVIF^(1/(2*Df))
## PerfScoreID    1.419725  1      1.191522
## State          239.370545 26      1.111095
## Position       235.676296 29      1.098758
## Age            1.641773  1      1.281317
## EmployedYear    1.612118  1      1.269692
## EmpSatisfaction 1.456007  1      1.206651
## Absences       1.331421  1      1.153872
```

## Outlier Analysis

```
#Cook's Distance Analysis
```

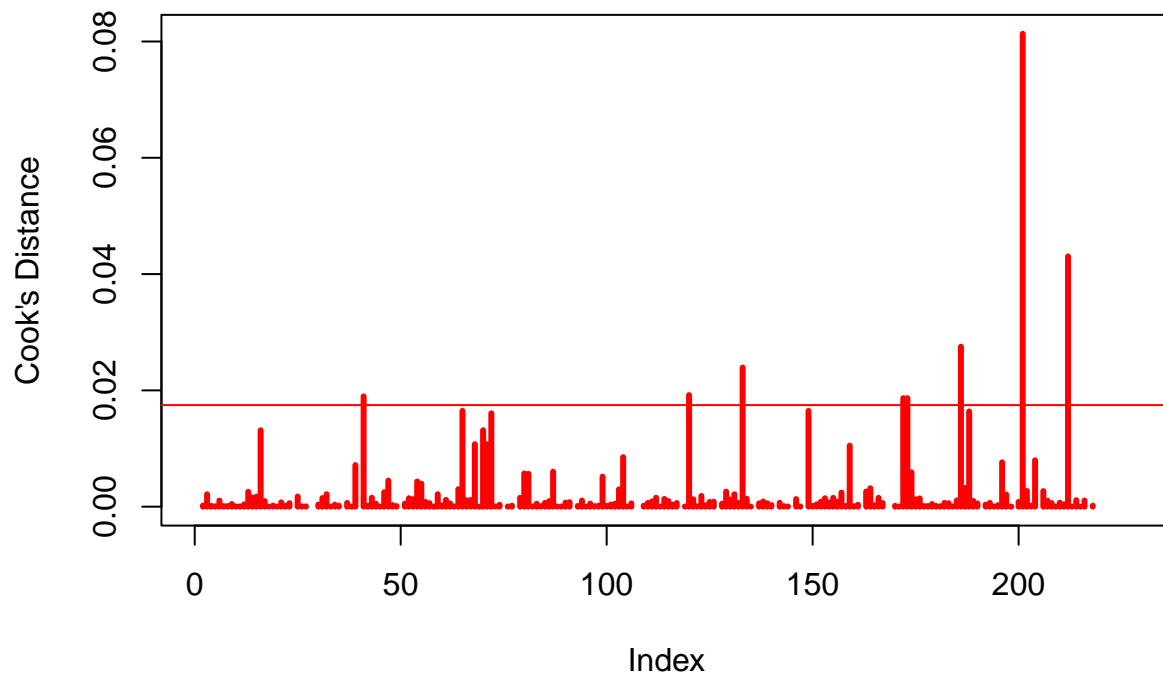
```
cook = cooks.distance(model_multi)
```

```
#Rule of Thumb
```

```
alarm = 4/nrow(HRtrain)
```

```
plot(cook,type="h",lwd=3,col="red", ylab = "Cook's Distance")
```

```
abline(h = alarm,col="red")
```



```

#Rule of Thumb
alarm = 4/nrow(HRtrain)
cat("Observation", which(cook>alarm), "has a cook's distance that is greater than", alarm)

## Observation 41 120 133 172 173 186 201 212 has a cook's distance that is greater than 0.01746725
# Check Results from Outlier Removal
HRtrain[201,]

##      MarriedID MaritalDesc Sex      EmploymentStatus Department PerfScoreID
## 284           1    Married   M Voluntarily Terminated      IT/IS           3
##      RecruitmentSource Salary      Position State Age CitizenDesc RaceDesc
## 284      CareerBuilder 75281 Network Engineer    MA 33 US Citizen    White
##      HispanicLatino EmployedYear ManagerName EngagementSurvey EmpSatisfaction
## 284                No           1 Peter Monroe                5                3
##      SpecialProjectsCount DaysLateLast30 Absences
## 284                    5                0         11

model_net_o = lm(formula = Salary ~ PerfScoreID + Position + State + Age + EmployedYear + EmpSatisfaction,
summary(model_net)

##
## Call:
## lm(formula = Salary ~ PerfScoreID + Position + State + Age +
##      EmployedYear + EmpSatisfaction + Absences, data = HRtrain)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13694  -3308       0    3332   14814
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    46752.60    9111.10   5.131 7.92e-07 ***
## PerfScoreID      -32.11     757.41  -0.042 0.966233
## PositionAdministrative Assistant -10357.43    5168.02  -2.004 0.046672 *
## PositionArea Sales Manager      8502.13    5991.35   1.419 0.157744
## PositionBI Developer      31471.24    5106.11   6.163 5.16e-09 ***
## PositionBI Director    46169.09    7232.26   6.384 1.64e-09 ***
## PositionCIO    154617.36    7203.88  21.463 < 2e-16 ***
## PositionData Analyst    26497.92    4781.96   5.541 1.15e-07 ***
## PositionData Architect    85388.81    7256.62  11.767 < 2e-16 ***
## PositionDatabase Administrator  48609.04    5174.41   9.394 < 2e-16 ***
## PositionDirector of Operations 105018.64    7200.44  14.585 < 2e-16 ***
## PositionDirector of Sales    123903.18   10755.97  11.519 < 2e-16 ***
## PositionEnterprise Architect   47981.41    8288.63   5.789 3.43e-08 ***
## PositionIT Director    112472.36    7189.85  15.643 < 2e-16 ***
## PositionIT Manager - DB      79792.60    5770.39  13.828 < 2e-16 ***
## PositionIT Manager - Infra   93226.35    7165.78  13.010 < 2e-16 ***
## PositionIT Manager - Support  74367.84    7247.82  10.261 < 2e-16 ***
## PositionIT Support      3390.85    4565.14   0.743 0.458664
## PositionNetwork Engineer   -2112.21    5126.29  -0.412 0.680843
## PositionPresident & CEO    184211.40    7358.40  25.034 < 2e-16 ***
## PositionPrincipal Data Architect  55948.91    7248.24   7.719 1.02e-12 ***
## PositionProduction Manager   13556.36    4241.99   3.196 0.001668 **
## PositionProduction Technician I -8468.85    3660.77  -2.313 0.021919 *

```

```

## PositionProduction Technician II      1000.66    3755.47    0.266 0.790218
## PositionSales Manager                  4803.62   10599.67    0.453 0.651004
## PositionSenior BI Developer            22866.32    7270.59    3.145 0.001966 **
## PositionShared Services Manager        28068.36    7206.21    3.895 0.000142 ***
## PositionSoftware Engineer              31489.35    4305.89    7.313 1.04e-11 ***
## PositionSoftware Engineering Manager   13550.02    7218.47    1.877 0.062244 .
## PositionSr. Accountant                 39933.53    5640.34    7.080 3.81e-11 ***
## PositionSr. DBA                       40041.99    7327.75    5.464 1.66e-07 ***
## PositionSr. Network Engineer           32382.39    5164.26    6.270 2.97e-09 ***
## StateAZ                               -4191.78    8947.45   -0.468 0.640046
## StateCA                               11271.42    9065.96    1.243 0.215512
## StateCO                                1937.82    8879.19    0.218 0.827506
## StateCT                                1661.23    7935.21    0.209 0.834430
## StateFL                                3151.54    8916.02    0.353 0.724182
## StateGA                                2589.31    8921.67    0.290 0.772003
## StateID                               -1170.18    8902.21   -0.131 0.895579
## StateIN                               -1819.67    8967.33   -0.203 0.839443
## StateKY                                -384.34    8907.11   -0.043 0.965634
## StateMA                               10309.94    7904.10    1.304 0.193899
## StateME                               -6052.53    8895.26   -0.680 0.497179
## StateMT                               -2945.41    8947.57   -0.329 0.742428
## StateNC                                5189.65    8920.52    0.582 0.561509
## StateND                                1516.00    9076.93    0.167 0.867559
## StateNH                                8500.62    8905.77    0.955 0.341207
## StateNV                               -2875.14    8728.00   -0.329 0.742254
## StateNY                                6061.48    8936.36    0.678 0.498524
## StateOH                               -2658.04    8923.99   -0.298 0.766185
## StateOR                               -5672.81    8901.42   -0.637 0.524807
## StatePA                               12051.71   12524.65    0.962 0.337321
## StateRI                                NA          NA          NA          NA
## StateTN                                7165.31    8983.03    0.798 0.426206
## StateTX                                1892.75    8877.02    0.213 0.831416
## StateUT                                8278.96    8921.87    0.928 0.354778
## StateVA                                9444.87    8969.66    1.053 0.293871
## StateVT                                8169.50    8881.84    0.920 0.359005
## StateWA                               -2242.07    8869.43   -0.253 0.800744
## Age                                    76.75      58.85      1.304 0.193988
## EmployedYear                           91.98     183.71      0.501 0.617254
## EmpSatisfaction                         496.61     519.29      0.956 0.340295
## Absences                               136.27      79.14      1.722 0.086950 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6161 on 167 degrees of freedom
## Multiple R-squared:  0.9637, Adjusted R-squared:  0.9504
## F-statistic: 72.67 on 61 and 167 DF,  p-value: < 2.2e-16

```

```
summary(model_net_o)
```

```

##
## Call:
## lm(formula = Salary ~ PerfScoreID + Position + State + Age +
##     EmployedYear + EmpSatisfaction + Absences, data = HRtrain[-201,
##     ])
##
##

```

```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -14023  -3158         0    3160   13116
##
## Coefficients: (1 not defined because of singularities)
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    45885.92    8893.98   5.159 7.00e-07 ***
## PerfScoreID      -296.87     744.00  -0.399  0.69040
## PositionAdministrative Assistant -9868.01    5044.84  -1.956  0.05214 .
## PositionArea Sales Manager      8617.25    5845.75   1.474  0.14235
## PositionBI Developer    31823.86    4983.24   6.386 1.64e-09 ***
## PositionBI Director    46261.17    7056.41   6.556 6.69e-10 ***
## PositionCIO    154709.73    7028.73  22.011 < 2e-16 ***
## PositionData Analyst    26867.67    4667.21   5.757 4.05e-08 ***
## PositionData Architect    85791.69    7081.33  12.115 < 2e-16 ***
## PositionDatabase Administrator  49028.43    5050.40   9.708 < 2e-16 ***
## PositionDirector of Operations 105058.53    7025.32  14.954 < 2e-16 ***
## PositionDirector of Sales   124616.42   10496.92  11.872 < 2e-16 ***
## PositionEnterprise Architect  48453.73    8088.49   5.990 1.26e-08 ***
## PositionIT Director   112633.02    7015.17  16.056 < 2e-16 ***
## PositionIT Manager - DB    79879.31    5630.11  14.188 < 2e-16 ***
## PositionIT Manager - Infra  93062.86    6991.69  13.310 < 2e-16 ***
## PositionIT Manager - Support 74501.28    7071.67  10.535 < 2e-16 ***
## PositionIT Support      3614.08    4454.70   0.811  0.41836
## PositionNetwork Engineer  -9618.10    5566.90  -1.728  0.08590 .
## PositionPresident & CEO   183982.78    7179.81  25.625 < 2e-16 ***
## PositionPrincipal Data Architect 56362.21    7073.22   7.968 2.45e-13 ***
## PositionProduction Manager   13810.04    4139.64   3.336  0.00105 **
## PositionProduction Technician I -8370.60    3571.87  -2.343  0.02029 *
## PositionProduction Technician II 1220.11    3664.83   0.333  0.73961
## PositionSales Manager     5228.30   10342.79   0.506  0.61388
## PositionSenior BI Developer  23251.29    7094.85   3.277  0.00128 **
## PositionShared Services Manager 28494.87    7032.31   4.052 7.78e-05 ***
## PositionSoftware Engineer   31656.59    4201.51   7.535 3.00e-12 ***
## PositionSoftware Engineering Manager 13268.22    7043.49   1.884  0.06135 .
## PositionSr. Accountant    39843.80    5503.23   7.240 1.59e-11 ***
## PositionSr. DBA          40693.16    7152.67   5.689 5.64e-08 ***
## PositionSr. Network Engineer 32615.22    5039.22   6.472 1.04e-09 ***
## StateAZ      -3581.33    8732.09  -0.410  0.68224
## StateCA     11984.08    8848.50   1.354  0.17746
## StateCO      2438.31    8664.76   0.281  0.77875
## StateCT      1953.98    7742.79   0.252  0.80107
## StateFL     3683.65    8700.89   0.423  0.67258
## StateGA     3020.17    8705.80   0.347  0.72909
## StateID     -843.13    8686.34  -0.097  0.92279
## StateIN     -980.76    8753.49  -0.112  0.91093
## StateKY       65.03    8691.69   0.007  0.99404
## StateMA    10919.91    7714.41   1.416  0.15879
## StateME    -5511.89    8680.69  -0.635  0.52633
## StateMT    -2451.12    8731.42  -0.281  0.77927
## StateNC     5936.96    8706.95   0.682  0.49627
## StateND     1671.56    8856.31   0.189  0.85053
## StateNH     9097.62    8691.33   1.047  0.29674
## StateNV    -2830.71    8515.72  -0.332  0.74000

```

```

## StateNY          6426.09    8719.81    0.737    0.46219
## StateOH         -1745.90    8712.00   -0.200    0.84141
## StateOR         -5044.74    8687.33   -0.581    0.56223
## StatePA         12542.08   12221.07    1.026    0.30626
## StateRI              NA          NA          NA          NA
## StateTN          8120.43    8770.06    0.926    0.35583
## StateTX          2626.01    8664.39    0.303    0.76221
## StateUT          8936.37    8707.50    1.026    0.30625
## StateVA         10266.95    8755.59    1.173    0.24263
## StateVT          8513.76    8666.54    0.982    0.32735
## StateWA         -2001.90    8654.06   -0.231    0.81735
## Age              85.70      57.49    1.491    0.13796
## EmployedYear      165.79     180.84    0.917    0.36059
## EmpSatisfaction    549.99     506.96    1.085    0.27955
## Absences          121.81      77.36    1.574    0.11728
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6011 on 166 degrees of freedom
## Multiple R-squared:  0.9656, Adjusted R-squared:  0.953
## F-statistic: 76.49 on 61 and 166 DF,  p-value: < 2.2e-16

```