

DDS_Case_Study_2

Dawson

7/26/2020

a. Deal with the Data: EDA (Job Specific)

Data summary

Name	CaseStudy
Number of rows	1470
Number of columns	41

Column type frequency:

character	2
factor	13
numeric	26

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
EducationField	0	1.0	5	16	0	6	0
Attrition	300	0.8	2	3	0	2	0

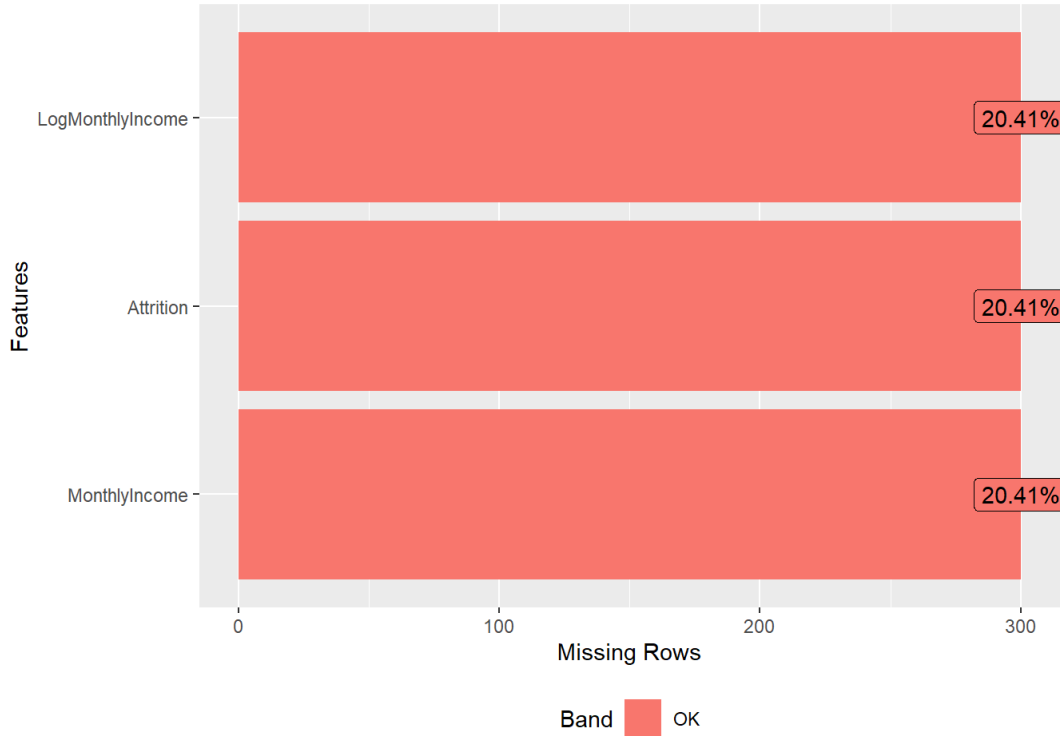
Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
BusinessTravel	0	1	FALSE	3	Tra: 1043, Tra: 277, Non: 150
Department	0	1	FALSE	3	Res: 961, Sal: 446, Hum: 63
Gender	0	1	FALSE	2	Mal: 882, Fem: 588
JobInvolvement	0	1	FALSE	4	3: 868, 2: 375, 4: 144, 1: 83
JobRole	0	1	FALSE	9	Sal: 326, Res: 292, Lab: 259, Man: 145
JobSatisfaction	0	1	FALSE	4	4: 459, 3: 442, 1: 289, 2: 280
MaritalStatus	0	1	FALSE	3	Mar: 673, Sin: 470, Div: 327
Over18	0	1	FALSE	1	Y: 1470
OverTime	0	1	FALSE	2	No: 1054, Yes: 416
PerformanceRating	0	1	FALSE	2	3: 1244, 4: 226
RelationshipSatisfaction	0	1	FALSE	4	3: 459, 4: 432, 2: 303, 1: 276
StockOptionLevel	0	1	FALSE	4	0: 631, 1: 596, 2: 158, 3: 85
WorkLifeBalance	0	1	FALSE	4	3: 893, 2: 344, 4: 153, 1: 80

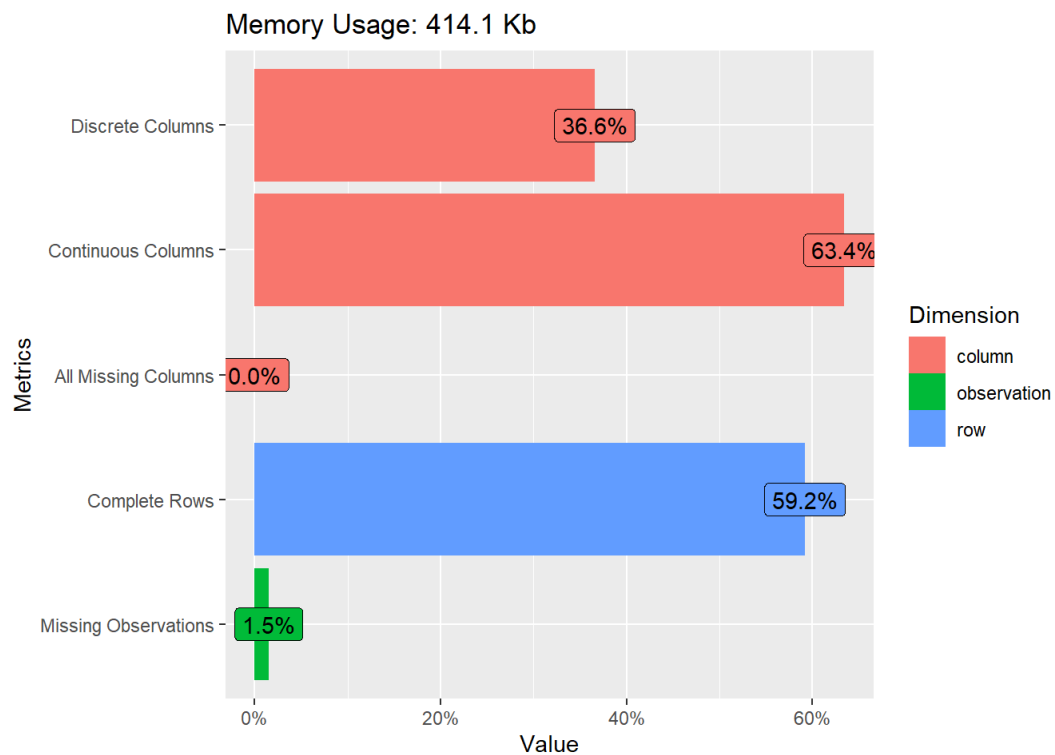
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
ID	0	1.0	735.50	424.50	1.00	368.25	735.50	1102.75	1470.00	
Age	0	1.0	36.92	9.14	18.00	30.00	36.00	43.00	60.00	
DailyRate	0	1.0	802.49	403.51	102.00	465.00	802.00	1157.00	1499.00	
DistanceFromHome	0	1.0	9.19	8.11	1.00	2.00	7.00	14.00	29.00	

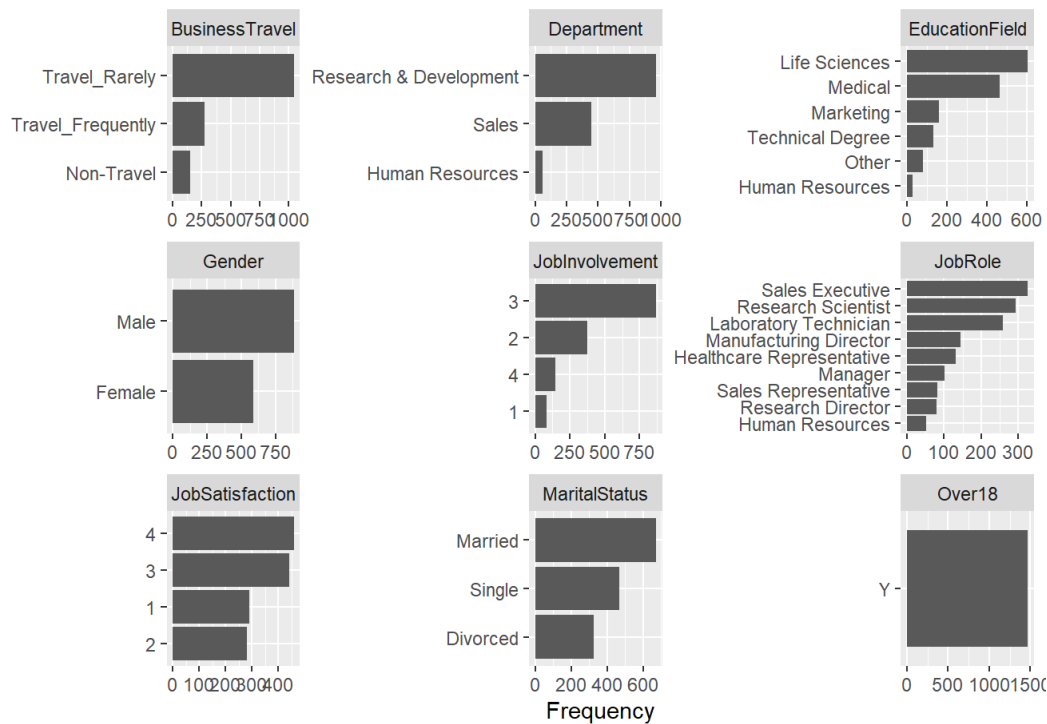
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
Education	0	1.0	2.91	1.02	1.00	2.00	3.00	4.00	5.00	
EmployeeCount	0	1.0	1.00	0.00	1.00	1.00	1.00	1.00	1.00	
EmployeeNumber	0	1.0	1024.87	602.02	1.00	491.25	1020.50	1555.75	2068.00	
EnvironmentSatisfaction	0	1.0	2.72	1.09	1.00	2.00	3.00	4.00	4.00	
HourlyRate	0	1.0	65.89	20.33	30.00	48.00	66.00	83.75	100.00	
JobLevel	0	1.0	2.06	1.11	1.00	1.00	2.00	3.00	5.00	
MonthlyRate	0	1.0	14313.10	7117.79	2094.00	8047.00	14235.50	20461.50	26999.00	
NumCompaniesWorked	0	1.0	2.69	2.50	0.00	1.00	2.00	4.00	9.00	
PercentSalaryHike	0	1.0	15.21	3.66	11.00	12.00	14.00	18.00	25.00	
StandardHours	0	1.0	80.00	0.00	80.00	80.00	80.00	80.00	80.00	
TotalWorkingYears	0	1.0	11.28	7.78	0.00	6.00	10.00	15.00	40.00	
TrainingTimesLastYear	0	1.0	2.80	1.29	0.00	2.00	3.00	3.00	6.00	
YearsAtCompany	0	1.0	7.01	6.13	0.00	3.00	5.00	9.00	40.00	
YearsInCurrentRole	0	1.0	4.23	3.62	0.00	2.00	3.00	7.00	18.00	
YearsSinceLastPromotion	0	1.0	2.19	3.22	0.00	0.00	1.00	3.00	15.00	
YearsWithCurrManager	0	1.0	4.12	3.57	0.00	2.00	3.00	7.00	17.00	
MonthlyIncome	300	0.8	6573.12	4794.96	1081.00	2872.00	4998.50	8579.50	19999.00	
LogMonthlyIncome	300	0.8	8.56	0.67	6.99	7.96	8.52	9.06	9.90	
LogDistanceFromHome	0	1.0	1.74	1.06	0.00	0.69	1.95	2.64	3.37	
LogJobLevel	0	1.0	0.59	0.52	0.00	0.00	0.69	1.10	1.61	
LogPercentSalaryHike	0	1.0	2.70	0.23	2.40	2.48	2.64	2.89	3.22	
NewRelationshipSatisfaction	0	1.0	0.81	0.39	0.00	1.00	1.00	1.00	1.00	



```
##           [,1]
## rows      1470
## columns    41
## discrete_columns  15
## continuous_columns 26
## all_missing_columns 0
## total_missing_values 900
## complete_rows    870
## total_observations 60270
## memory_usage    424008
```



```
## starting httpd help server ... done
```



Page 1

```
##
##
## processing file: report.rmd
```

```
##
|
|
|
|..
## inline R code fragments
##
##
|
|...
## label: global_options (with options)
## List of 1
## $ include: logi FALSE
##
##
|
|.....
## ordinary text without R code
##
##
|
|.....
## label: introduce
##
|
|.....
## ordinary text without R code
##
##
|
|.....
## label: plot_intro
```

```

##
|
|.....| 17%
## ordinary text without R code
##
##
|
|.....| 19%
## label: data_structure
##
|
|.....| 21%
## ordinary text without R code
##
##
|
|.....| 24%
## label: missing_profile

```

```

##
|
|.....| 26%
## ordinary text without R code
##
##
|
|.....| 29%
## label: univariate_distribution_header
##
|
|.....| 31%
## ordinary text without R code
##
##
|
|.....| 33%
## label: plot_histogram

```

```

##
|
|.....| 36%
## ordinary text without R code
##
##
|
|.....| 38%
## label: plot_density
##
|
|.....| 40%
## ordinary text without R code
##
##
|
|.....| 43%
## label: plot_frequency_bar

```

```

##
|
|.....| 45%
## ordinary text without R code
##
##
|
|.....| 48%
## label: plot_response_bar
##
|
|.....| 50%
## ordinary text without R code
##
##
|
|.....| 52%
## label: plot_with_bar
##
|
|.....| 55%
## ordinary text without R code
##
##
|
|.....| 57%
## label: plot_normal_qq

```

```

##
|
|.....| 60%
## ordinary text without R code
##
##
|
|.....| 62%
## label: plot_response_qq

```

```

##
|
|.....| 64%
## ordinary text without R code
##
##
|
|.....| 67%
## label: plot_by_qq
##
|
|.....| 69%
## ordinary text without R code
##
##
|
|.....| 71%
## label: correlation_analysis

```

```
##
|
|.....| 74%
## ordinary text without R code
##
##
|
|.....| 76%
## label: principal_component_analysis
```

```
##
|
|.....| 79%
## ordinary text without R code
##
##
|
|.....| 81%
## label: bivariate_distribution_header
##
|
|.....| 83%
## ordinary text without R code
##
##
|
|.....| 86%
## label: plot_response_boxplot
```

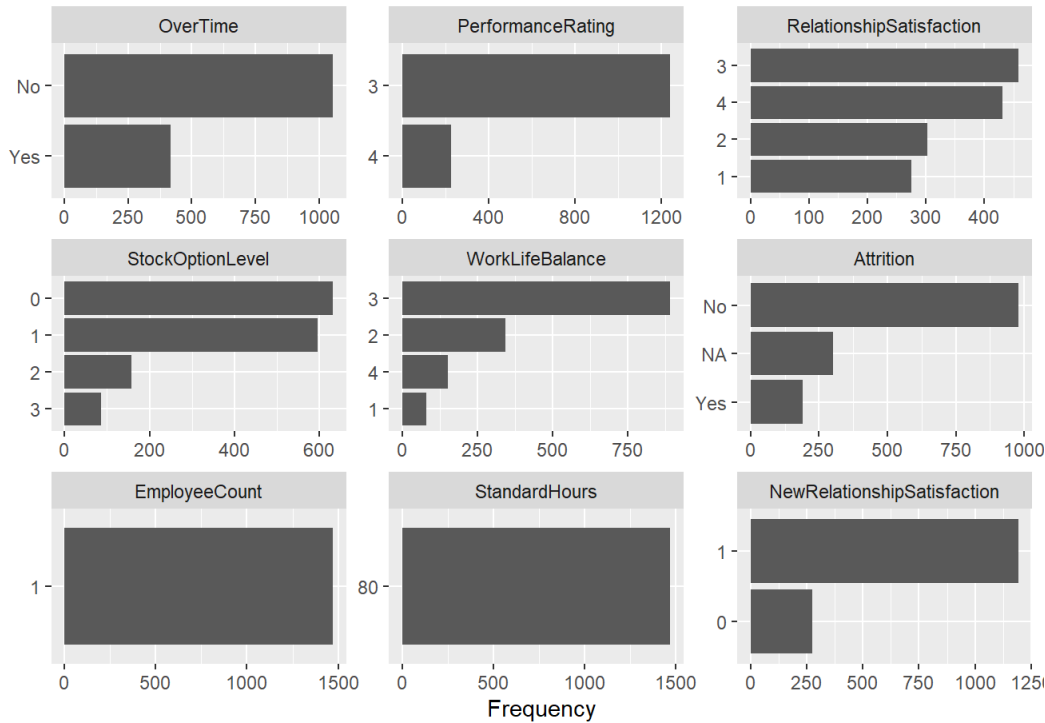
```
##
|
|.....| 88%
## ordinary text without R code
##
##
|
|.....| 90%
## label: plot_by_boxplot
##
|
|.....| 93%
## ordinary text without R code
##
##
|
|.....| 95%
## label: plot_response_scatterplot
```

```
##
|
|.....| 98%
## ordinary text without R code
##
##
|
|.....| 100%
## label: plot_by_scatterplot
```

```
## output file: C:/Users/Dawson/Documents/GitHub/Analytics/CaseStudy2DDS/report.knit.md
```

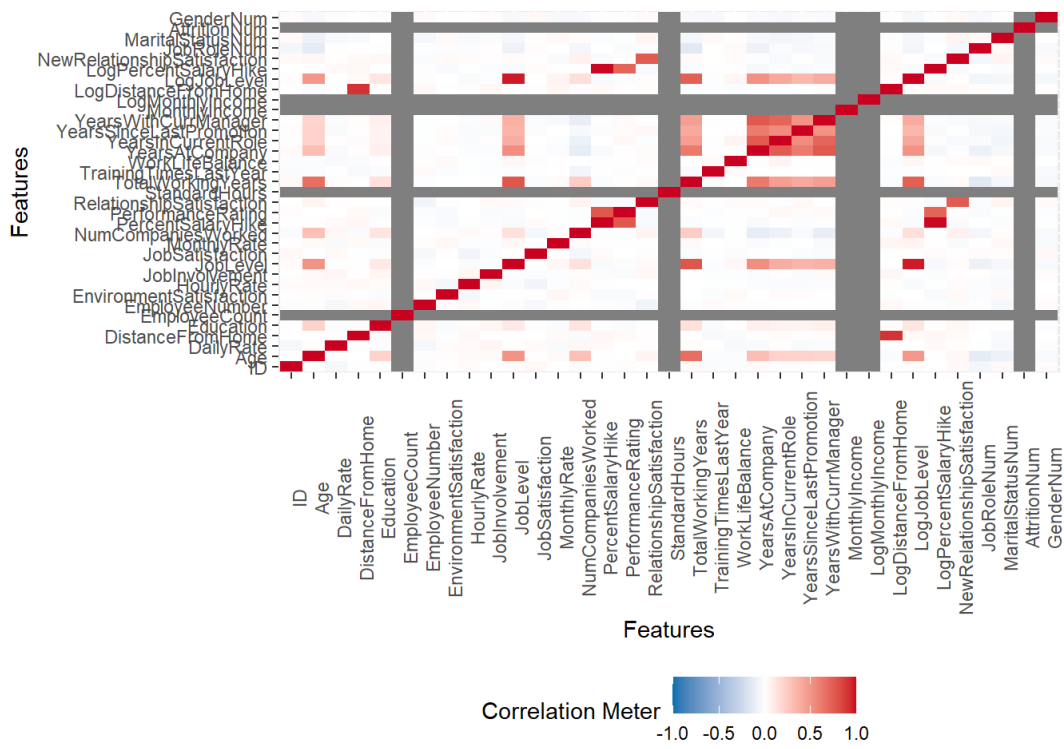
```
## "C:/Program Files/RStudio/bin/pandoc/pandoc" +RTS -K512m -RTS "C:/Users/Dawson/Documents/GitHub/Analytics/CaseStudy2DD
S/report.utf8.md" --to html4 --from markdown+autolink_bare_uris+tex_math_single_backslash --output pandoc349861d7194.html
--email-obfuscation none --self-contained --standalone --section-divs --table-of-contents --toc-depth 6 --template "C:\Use
rs\Dawson\Documents\R\win-library\4.0\rmarkdown\rmd\h\default.html" --no-highlight --variable highlightjs=1 --variable "th
eme:yeti" --include-in-header "C:/Users/Dawson\AppData\Local\Temp\Rtmpe2a8Ct\rmarkdown-str349822573dbc.html" --mathjax --v
ariable "mathjax-url:https://mathjax.rstudio.com/latest/MathJax.js?config=TeX-AMS-MML_HTMLorMML" --lua-filter "C:/Users/Da
wson/Documents/R/win-library/4.0/rmarkdown/rmd/lua/pagebreak.lua" --lua-filter "C:/Users/Dawson/Documents/R/win-library/4.
0/rmarkdown/rmd/lua/latex-div.lua"
```

```
##
## Output created: report.html
```

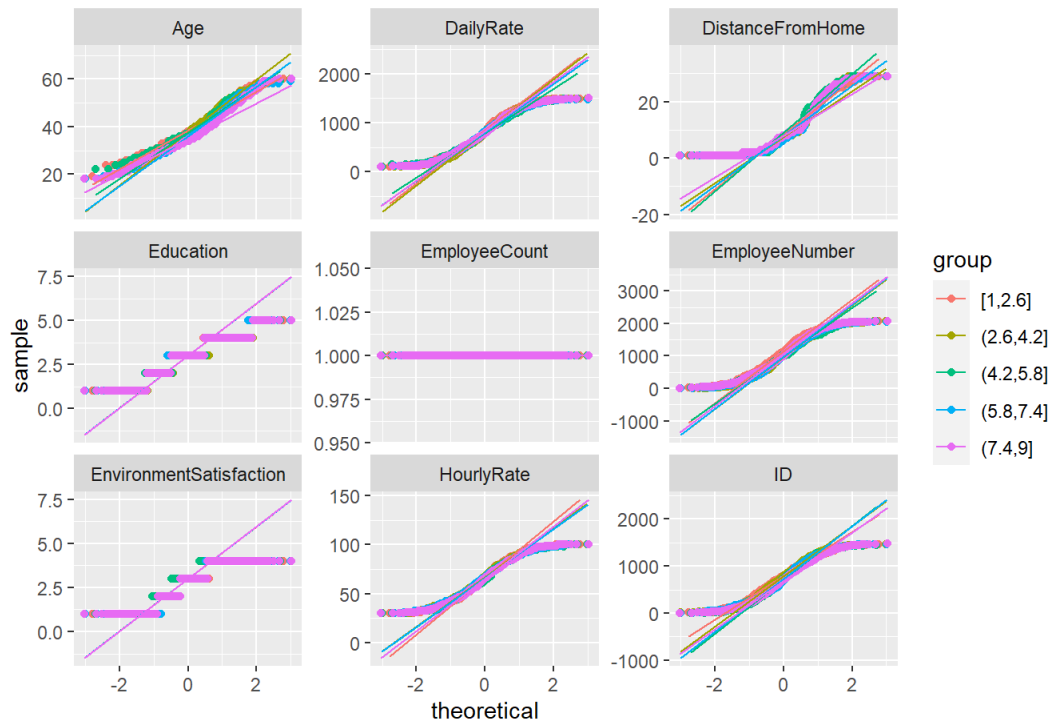
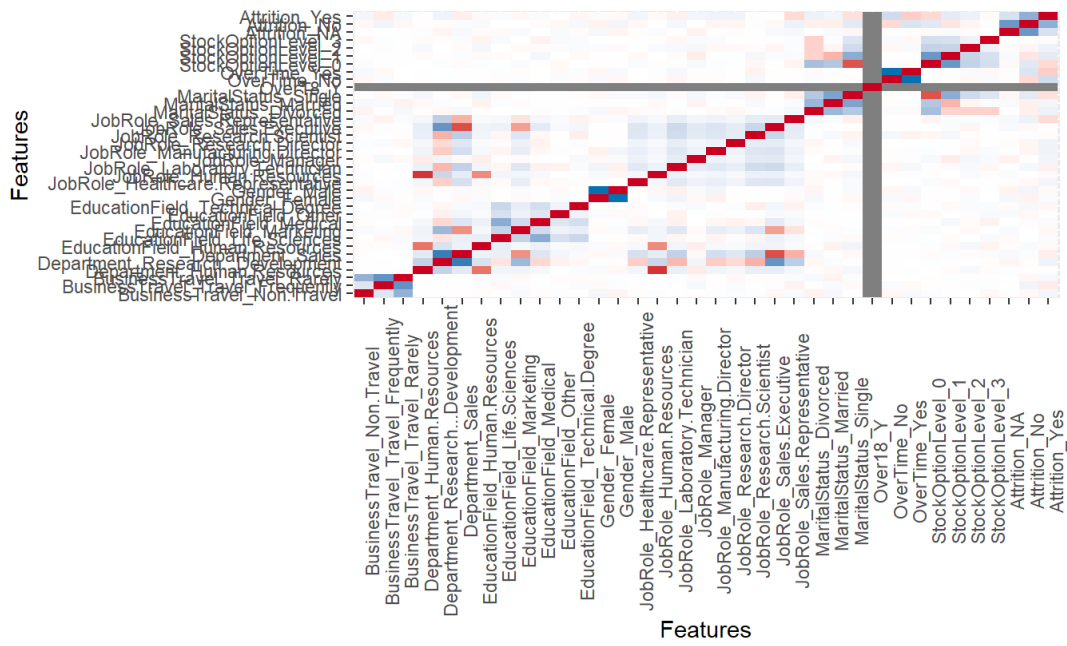


Page 2

```
## Warning in cor(x = structure(list(ID = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, : the
## standard deviation is zero
```

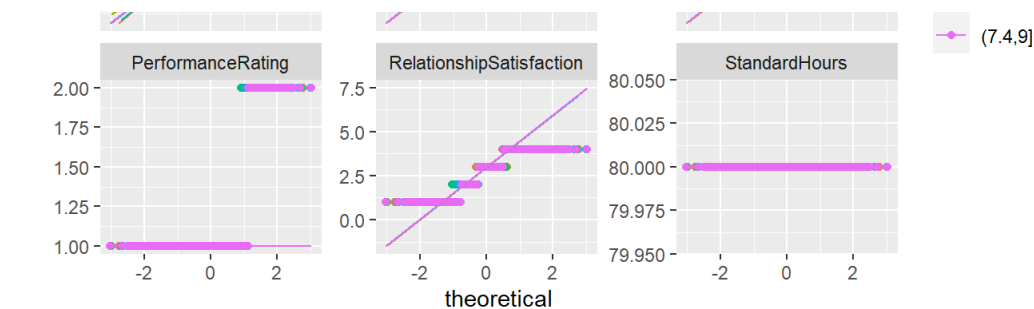



```
## Warning in cor(x = structure(list(BusinessTravel_Non.Travel = c(0L, 0L, : the
## standard deviation is zero
```



Page 1

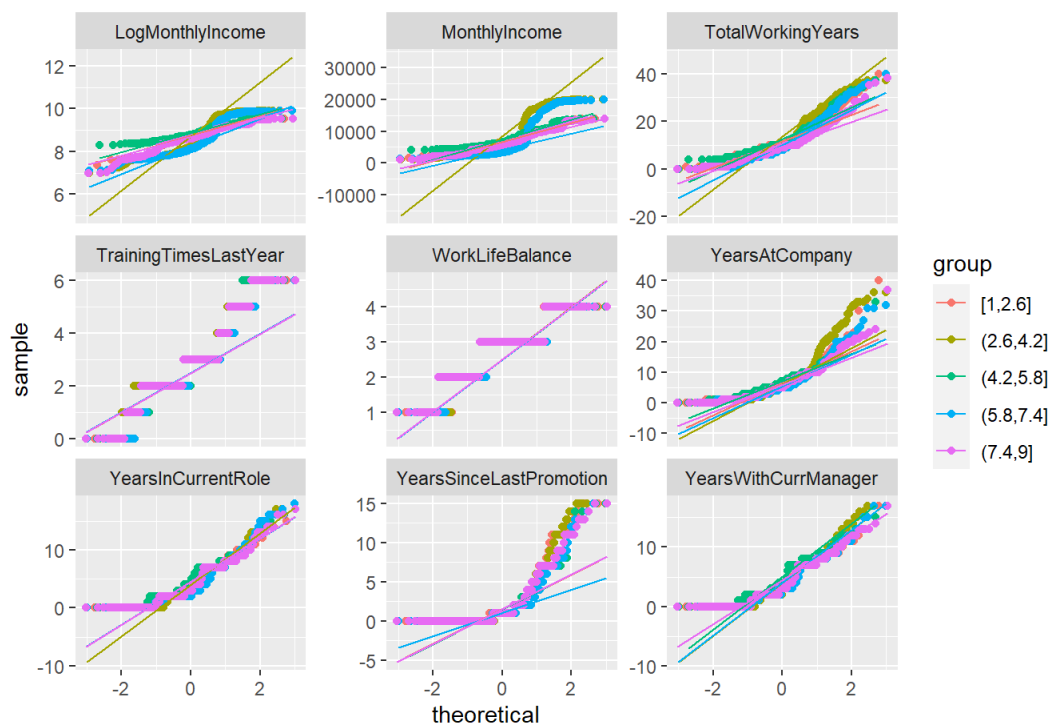




Page 2

```
## Warning: Removed 600 rows containing non-finite values (stat_qq).
```

```
## Warning: Removed 600 rows containing non-finite values (stat_qq_line).
```



Page 3

```
## Warning: Removed 300 rows containing non-finite values (stat_qq).
```

```
## Warning: Removed 300 rows containing non-finite values (stat_qq_line).
```

```
## Warning in cor(CaseStudyJobRoleNum, use = "complete.obs", method = "pearson"):  
## the standard deviation is zero
```

##	ID	Age	DailyRate
## ID	1.000000000	-0.040097596	-3.290228e-02
## Age	-0.040097596	1.000000000	1.273669e-02
## DailyRate	-0.032902279	0.012736692	1.000000e+00
## DistanceFromHome	0.066138779	0.006647112	1.494454e-02
## Education	-0.049538273	0.217290207	-1.453569e-02
## EmployeeCount	NA	NA	NA
## EmployeeNumber	-0.022352503	0.010470589	-2.973441e-02
## EnvironmentSatisfaction	-0.012725264	-0.007449882	-1.338386e-02
## HourlyRate	0.006259530	0.052328259	4.849597e-02
## JobInvolvement	-0.050285102	0.016073296	6.011801e-02
## JobLevel	-0.039260585	0.479423564	1.837626e-03
## JobSatisfaction	0.040917532	-0.022363414	-2.716745e-03
## MonthlyRate	-0.004748305	0.071740854	-2.764077e-02
## NumCompaniesWorked	-0.018549395	0.293521877	5.155983e-02
## PercentSalaryHike	0.017039305	-0.030804192	2.787530e-02
## PerformanceRating	0.016405803	-0.041084685	-2.407272e-02
## RelationshipSatisfaction	-0.041604748	-0.005340777	1.364324e-02
## StandardHours	NA	NA	NA
## TotalWorkingYears	-0.042010558	0.652613137	-7.402129e-03
## TrainingTimesLastYear	0.030408578	-0.045281792	-1.058980e-02
## WorkLifeBalance	0.010458196	-0.012229885	-2.986922e-02
## YearsAtCompany	0.006726733	0.289538716	-3.714277e-02
## YearsInCurrentRole	-0.068856386	0.208349696	-6.759731e-04
## YearsSinceLastPromotion	0.003515130	0.217191680	-5.855873e-02
## YearsWithCurrManager	-0.038176461	0.185383795	-2.434683e-02
## MonthlyIncome	-0.051639025	0.484288301	8.790339e-05
## LogMonthlyIncome	-0.040350094	0.473524929	2.307987e-02
## LogDistanceFromHome	0.073347959	0.003561597	1.775388e-02
## LogJobLevel	-0.032529996	0.453675395	1.381252e-02
## LogPercentSalaryHike	0.011019326	-0.027419490	3.420546e-02
## NewRelationshipSatisfaction	-0.031132999	-0.020229058	2.054073e-02
## JobRoleNum	-0.069017184	-0.105969885	6.317723e-03
## MaritalStatusNum	-0.019070591	-0.079557838	-2.539210e-02
## AttritionNum	0.047266995	-0.149383577	-3.379312e-02
## GenderNum	-0.003111656	-0.044228011	-7.403226e-03
##	DistanceFromHome	Education	EmployeeCount
## ID	0.066138779	-0.049538273	NA
## Age	0.006647112	0.217290207	NA
## DailyRate	0.014944536	-0.014535694	NA
## DistanceFromHome	1.000000000	0.049512940	NA
## Education	0.049512940	1.000000000	NA
## EmployeeCount	NA	NA	1
## EmployeeNumber	-0.004641466	0.016010904	NA
## EnvironmentSatisfaction	-0.038700752	-0.036527898	NA
## HourlyRate	0.067040911	0.006750312	NA
## JobInvolvement	-0.004277505	0.032255238	NA
## JobLevel	0.021727054	0.130385908	NA
## JobSatisfaction	-0.021218476	0.013111079	NA
## MonthlyRate	-0.007839039	-0.020183862	NA
## NumCompaniesWorked	-0.048909244	0.160918010	NA
## PercentSalaryHike	0.054293690	-0.002387120	NA
## PerformanceRating	0.033190328	-0.028051113	NA
## RelationshipSatisfaction	0.036234469	-0.025717762	NA
## StandardHours	NA	NA	NA
## TotalWorkingYears	0.004186191	0.146044538	NA
## TrainingTimesLastYear	-0.038949194	-0.063134803	NA
## WorkLifeBalance	-0.013820292	0.009824289	NA
## YearsAtCompany	-0.020830094	0.061974487	NA
## YearsInCurrentRole	-0.008407782	0.061381620	NA
## YearsSinceLastPromotion	-0.016595666	0.068329974	NA
## YearsWithCurrManager	-0.017700612	0.088446276	NA
## MonthlyIncome	-0.006667155	0.127131394	NA
## LogMonthlyIncome	0.014025136	0.146085491	NA
## LogDistanceFromHome	0.910392216	0.062110809	NA
## LogJobLevel	0.040015896	0.144187010	NA
## LogPercentSalaryHike	0.055001344	0.005435199	NA

## NewRelationshipSatisfaction	0.039223375	0.027542628	NA
## JobRoleNum	-0.012799586	0.063075019	NA
## MaritalStatusNum	-0.104479164	-0.016955467	NA
## AttritionNum	0.087136293	-0.049442357	NA
## GenderNum	-0.012074449	-0.009137821	NA
##	EmployeeNumber	EnvironmentSatisfaction	
## ID	-2.235250e-02	-1.272526e-02	
## Age	1.047059e-02	-7.449882e-03	
## DailyRate	-2.973441e-02	-1.338386e-02	
## DistanceFromHome	-4.641466e-03	-3.870075e-02	
## Education	1.601090e-02	-3.652790e-02	
## EmployeeCount	NA	NA	
## EmployeeNumber	1.000000e+00	2.771656e-02	
## EnvironmentSatisfaction	2.771656e-02	1.000000e+00	
## HourlyRate	9.330399e-03	-3.418928e-02	
## JobInvolvement	9.614957e-05	-3.419785e-05	
## JobLevel	2.761933e-02	1.545152e-04	
## JobSatisfaction	-6.369716e-02	-1.841649e-02	
## MonthlyRate	3.561085e-02	5.927864e-02	
## NumCompaniesWorked	8.988969e-03	1.294575e-02	
## PercentSalaryHike	-3.363243e-02	1.139159e-03	
## PerformanceRating	-2.382039e-02	-4.524833e-03	
## RelationshipSatisfaction	-6.012968e-02	4.550669e-03	
## StandardHours	NA	NA	
## TotalWorkingYears	2.370148e-02	-1.995653e-02	
## TrainingTimesLastYear	5.613245e-03	-1.450192e-02	
## WorkLifeBalance	1.093866e-02	8.412080e-02	
## YearsAtCompany	2.517627e-02	-1.510054e-02	
## YearsInCurrentRole	1.318767e-02	2.249029e-02	
## YearsSinceLastPromotion	2.709173e-02	9.176822e-03	
## YearsWithCurrManager	2.240429e-02	-2.329230e-02	
## MonthlyIncome	3.420305e-02	-1.687744e-02	
## LogMonthlyIncome	3.126918e-02	-2.760362e-02	
## LogDistanceFromHome	9.142798e-03	-2.059574e-02	
## LogJobLevel	2.446772e-02	-1.239649e-03	
## LogPercentSalaryHike	-3.161632e-02	-1.344839e-03	
## NewRelationshipSatisfaction	-5.197159e-03	7.845944e-03	
## JobRoleNum	-5.199126e-02	-5.191753e-02	
## MaritalStatusNum	-3.119574e-02	3.032765e-02	
## AttritionNum	-2.279383e-02	-7.732540e-02	
## GenderNum	-3.729314e-03	2.173554e-02	
##	HourlyRate	JobInvolvement	JobLevel
## ID	6.259530e-03	-5.028510e-02	-0.0392605846
## Age	5.232826e-02	1.607330e-02	0.4794235642
## DailyRate	4.849597e-02	6.011801e-02	0.0018376264
## DistanceFromHome	6.704091e-02	-4.277505e-03	0.0217270544
## Education	6.750312e-03	3.225524e-02	0.1303859075
## EmployeeCount	NA	NA	NA
## EmployeeNumber	9.330399e-03	9.614957e-05	0.0276193318
## EnvironmentSatisfaction	-3.418928e-02	-3.419785e-05	0.0001545152
## HourlyRate	1.000000e+00	7.036098e-02	-0.0080700744
## JobInvolvement	7.036098e-02	1.000000e+00	-0.0158750522
## JobLevel	-8.070074e-03	-1.587505e-02	1.0000000000
## JobSatisfaction	-8.904064e-02	-5.294946e-02	-0.0503060650
## MonthlyRate	-1.603517e-02	-1.827518e-02	0.0744538290
## NumCompaniesWorked	1.412012e-02	-5.613845e-03	0.1408348284
## PercentSalaryHike	-2.270814e-02	1.432785e-02	-0.0622671489
## PerformanceRating	3.182237e-03	1.168864e-02	-0.0386966885
## RelationshipSatisfaction	1.537703e-02	1.685454e-02	-0.0048204622
## StandardHours	NA	NA	NA
## TotalWorkingYears	2.947671e-02	-1.398641e-02	0.7807523526
## TrainingTimesLastYear	5.822839e-03	-2.241602e-02	-0.0458643163
## WorkLifeBalance	-2.548651e-02	7.736011e-03	0.0302838677
## YearsAtCompany	4.990148e-05	-4.159642e-02	0.5205835070
## YearsInCurrentRole	2.159320e-04	1.317048e-02	0.3913048192
## YearsSinceLastPromotion	8.053699e-03	-3.043052e-02	0.3304357051
## YearsWithCurrManager	-2.493375e-03	1.088728e-02	0.3665674724
## MonthlyIncome	2.391151e-03	4.473792e-04	0.9516400488

## LogMonthlyIncome	-5.613674e-03	3.213698e-03	0.9224616799
## LogDistanceFromHome	5.598718e-02	1.259118e-02	0.0177281964
## LogJobLevel	-1.214814e-02	-1.930007e-02	0.9684504624
## LogPercentSalaryHike	-2.398964e-02	1.129736e-02	-0.0655237434
## NewRelationshipSatisfaction	1.825655e-02	3.269000e-03	-0.0058025918
## JobRoleNum	-1.820483e-02	-1.142614e-02	-0.0742994460
## MaritalStatusNum	-5.933557e-02	-1.497949e-02	-0.0654707405
## AttritionNum	3.655418e-02	-1.877934e-01	-0.1621364440
## GenderNum	8.523982e-03	2.308382e-02	-0.0690784585
##	JobSatisfaction	MonthlyRate	NumCompaniesWorked
## ID	0.040917532	-0.004748305	-0.018549395
## Age	-0.022363414	0.071740854	0.293521877
## DailyRate	-0.002716745	-0.027640771	0.051559831
## DistanceFromHome	-0.021218476	-0.007839039	-0.048909244
## Education	0.013111079	-0.020183862	0.160918010
## EmployeeCount	NA	NA	NA
## EmployeeNumber	-0.063697159	0.035610845	0.008988969
## EnvironmentSatisfaction	-0.018416488	0.059278643	0.012945750
## HourlyRate	-0.089040641	-0.016035168	0.014120123
## JobInvolvement	-0.052949460	-0.018275178	-0.005613845
## JobLevel	-0.050306065	0.074453829	0.140834828
## JobSatisfaction	1.000000000	0.025139100	-0.075753415
## MonthlyRate	0.025139100	1.000000000	0.016753747
## NumCompaniesWorked	-0.075753415	0.016753747	1.000000000
## PercentSalaryHike	0.011967235	0.001642760	-0.022434259
## PerformanceRating	-0.004641900	-0.003288565	-0.029306651
## RelationshipSatisfaction	-0.033860142	-0.024004655	0.042047644
## StandardHours	NA	NA	NA
## TotalWorkingYears	-0.049281026	0.057889875	0.258099959
## TrainingTimesLastYear	-0.029577808	-0.013358464	-0.070591665
## WorkLifeBalance	-0.026470304	0.006510198	0.021316529
## YearsAtCompany	0.028363964	-0.018039461	-0.142025883
## YearsInCurrentRole	-0.003471690	0.025218599	-0.103814233
## YearsSinceLastPromotion	-0.021795589	0.007491151	-0.071218393
## YearsWithCurrManager	0.007359588	-0.021310286	-0.122974681
## MonthlyIncome	-0.053148530	0.064594067	0.155894318
## LogMonthlyIncome	-0.037564731	0.081252155	0.158823008
## LogDistanceFromHome	-0.029846482	0.011348090	-0.030590267
## LogJobLevel	-0.041654226	0.091297583	0.129499558
## LogPercentSalaryHike	0.012146664	0.003771186	-0.018110672
## NewRelationshipSatisfaction	-0.009287233	-0.039550299	0.070270708
## JobRoleNum	0.008906795	-0.013817053	-0.008968013
## MaritalStatusNum	0.018129989	-0.016250983	-0.011599836
## AttritionNum	-0.107520935	-0.043232173	0.061018887
## GenderNum	0.021121567	-0.032817781	-0.010620939
##	PercentSalaryHike	PerformanceRating	
## ID	0.017039305	0.016405803	
## Age	-0.030804192	-0.041084685	
## DailyRate	0.027875304	-0.024072724	
## DistanceFromHome	0.054293690	0.033190328	
## Education	-0.002387120	-0.028051113	
## EmployeeCount	NA	NA	
## EmployeeNumber	-0.033632429	-0.023820392	
## EnvironmentSatisfaction	0.001139159	-0.004524833	
## HourlyRate	-0.022708145	0.003182237	
## JobInvolvement	0.014327853	0.011688638	
## JobLevel	-0.062267149	-0.038696689	
## JobSatisfaction	0.011967235	-0.004641900	
## MonthlyRate	0.001642760	-0.003288565	
## NumCompaniesWorked	-0.022434259	-0.029306651	
## PercentSalaryHike	1.000000000	0.775053243	
## PerformanceRating	0.775053243	1.000000000	
## RelationshipSatisfaction	-0.045721608	-0.029980063	
## StandardHours	NA	NA	
## TotalWorkingYears	-0.064304461	-0.041377258	
## TrainingTimesLastYear	0.001525276	-0.007174679	
## WorkLifeBalance	0.007475735	0.017234301	
## YearsAtCompany	-0.057480309	-0.026617880	

## YearsInCurrentRole	-0.022935654	0.008802684
## YearsSinceLastPromotion	-0.074040285	-0.036530370
## YearsWithCurrManager	-0.045232326	-0.004942028
## MonthlyIncome	-0.053865939	-0.043130996
## LogMonthlyIncome	-0.054692887	-0.043068808
## LogDistanceFromHome	0.038457424	0.016568050
## LogJobLevel	-0.056932169	-0.030733506
## LogPercentSalaryHike	0.993829811	0.722937450
## NewRelationshipSatisfaction	-0.037705801	-0.020825633
## JobRoleNum	-0.004581115	-0.030687047
## MaritalStatusNum	0.016925180	0.020583452
## AttritionNum	0.015328069	0.015333837
## GenderNum	0.003694383	-0.014933022
##	RelationshipSatisfaction	StandardHours
## ID	-0.041604748	NA
## Age	-0.005340777	NA
## DailyRate	0.013643236	NA
## DistanceFromHome	0.036234469	NA
## Education	-0.025717762	NA
## EmployeeCount	NA	NA
## EmployeeNumber	-0.060129683	NA
## EnvironmentSatisfaction	0.004550669	NA
## HourlyRate	0.015377027	NA
## JobInvolvement	0.016854535	NA
## JobLevel	-0.004820462	NA
## JobSatisfaction	-0.033860142	NA
## MonthlyRate	-0.024004655	NA
## NumCompaniesWorked	0.042047644	NA
## PercentSalaryHike	-0.045721608	NA
## PerformanceRating	-0.029980063	NA
## RelationshipSatisfaction	1.000000000	NA
## StandardHours	NA	1
## TotalWorkingYears	-0.017575299	NA
## TrainingTimesLastYear	0.015752447	NA
## WorkLifeBalance	0.037105402	NA
## YearsAtCompany	0.008377778	NA
## YearsInCurrentRole	-0.001097767	NA
## YearsSinceLastPromotion	0.032135460	NA
## YearsWithCurrManager	-0.016131001	NA
## MonthlyIncome	-0.003884678	NA
## LogMonthlyIncome	-0.025328547	NA
## LogDistanceFromHome	0.038160335	NA
## LogJobLevel	-0.015338256	NA
## LogPercentSalaryHike	-0.048172268	NA
## NewRelationshipSatisfaction	0.774553995	NA
## JobRoleNum	-0.021511456	NA
## MaritalStatusNum	0.070692715	NA
## AttritionNum	-0.039646611	NA
## GenderNum	0.013253439	NA
##	TotalWorkingYears	TrainingTimesLastYear
## ID	-0.042010558	0.030408578
## Age	0.652613137	-0.045281792
## DailyRate	-0.007402129	-0.010589800
## DistanceFromHome	0.004186191	-0.038949194
## Education	0.146044538	-0.063134803
## EmployeeCount	NA	NA
## EmployeeNumber	0.023701482	0.005613245
## EnvironmentSatisfaction	-0.019956530	-0.014501924
## HourlyRate	0.029476711	0.005822839
## JobInvolvement	-0.013986410	-0.022416021
## JobLevel	0.780752353	-0.045864316
## JobSatisfaction	-0.049281026	-0.029577808
## MonthlyRate	0.057889875	-0.013358464
## NumCompaniesWorked	0.258099959	-0.070591665
## PercentSalaryHike	-0.064304461	0.001525276
## PerformanceRating	-0.041377258	-0.007174679
## RelationshipSatisfaction	-0.017575299	0.015752447
## StandardHours	NA	NA

## TotalWorkingYears	1.000000000	-0.041912467
## TrainingTimesLastYear	-0.041912467	1.000000000
## WorkLifeBalance	0.018724536	0.021735800
## YearsAtCompany	0.636083586	0.020192712
## YearsInCurrentRole	0.486590295	-0.016429957
## YearsSinceLastPromotion	0.445838540	-0.042382617
## YearsWithCurrManager	0.461485375	0.003914247
## MonthlyIncome	0.778511205	-0.039014567
## LogMonthlyIncome	0.744885313	-0.043736817
## LogDistanceFromHome	-0.013188398	-0.025205248
## LogJobLevel	0.736640558	-0.047614528
## LogPercentSalaryHike	-0.066509032	0.003857571
## NewRelationshipSatisfaction	-0.005663439	0.008584993
## JobRoleNum	-0.133122913	-0.034165342
## MaritalStatusNum	-0.077064995	0.045612729
## AttritionNum	-0.167206122	-0.062726088
## GenderNum	-0.072374947	0.001091037
##	WorkLifeBalance	YearsAtCompany
## ID	0.0104581961	6.726733e-03
## Age	-0.0122298854	2.895387e-01
## DailyRate	-0.0298692207	-3.714277e-02
## DistanceFromHome	-0.0138202924	-2.083009e-02
## Education	0.0098242894	6.197449e-02
## EmployeeCount	NA	NA
## EmployeeNumber	0.0109386611	2.517627e-02
## EnvironmentSatisfaction	0.0841208021	-1.510054e-02
## HourlyRate	-0.0254865102	4.990148e-05
## JobInvolvement	0.0077360108	-4.159642e-02
## JobLevel	0.0302838677	5.205835e-01
## JobSatisfaction	-0.0264703041	2.836396e-02
## MonthlyRate	0.0065101982	-1.803946e-02
## NumCompaniesWorked	0.0213165286	-1.420259e-01
## PercentSalaryHike	0.0074757347	-5.748031e-02
## PerformanceRating	0.0172343009	-2.661788e-02
## RelationshipSatisfaction	0.0371054018	8.377778e-03
## StandardHours	NA	NA
## TotalWorkingYears	0.0187245356	6.360836e-01
## TrainingTimesLastYear	0.0217358001	2.019271e-02
## WorkLifeBalance	1.0000000000	2.947258e-02
## YearsAtCompany	0.0294725801	1.000000e+00
## YearsInCurrentRole	0.0821049336	7.761164e-01
## YearsSinceLastPromotion	0.0416530997	6.440298e-01
## YearsWithCurrManager	0.0183780385	7.652262e-01
## MonthlyIncome	0.0208048887	4.913790e-01
## LogMonthlyIncome	0.0214772089	4.981861e-01
## LogDistanceFromHome	0.0004870109	-2.294682e-02
## LogJobLevel	0.0361190825	5.118268e-01
## LogPercentSalaryHike	0.0055466255	-6.228181e-02
## NewRelationshipSatisfaction	0.0363452033	-3.132587e-02
## JobRoleNum	0.0259822969	-9.122394e-02
## MaritalStatusNum	0.0111959316	-6.373189e-02
## AttritionNum	-0.0897897090	-1.287541e-01
## GenderNum	-0.0240378277	-8.181910e-02
##	YearsSinceLastPromotion	YearsWithCurrManager
## ID	0.0035151302	-0.038176461
## Age	0.2171916798	0.185383795
## DailyRate	-0.0585587308	-0.024346829
## DistanceFromHome	-0.0165956662	-0.017700612
## Education	0.0683299744	0.088446276
## EmployeeCount	NA	NA
## EmployeeNumber	0.0270917312	0.022404287
## EnvironmentSatisfaction	0.0091768217	-0.023292298
## HourlyRate	0.0080536995	-0.002493375
## JobInvolvement	-0.0304305184	0.010887282
## JobLevel	0.3304357051	0.366567472
## JobSatisfaction	-0.0217955893	0.007359588
## MonthlyRate	0.0074911511	-0.021310286
## NumCompaniesWorked	-0.0712183925	-0.122974681

## PercentSalaryHike	-0.0740402851	-0.045232326
## PerformanceRating	-0.0365303704	-0.004942028
## RelationshipSatisfaction	0.0321354602	-0.016131001
## StandardHours	NA	NA
## TotalWorkingYears	0.4458385396	0.461485375
## TrainingTimesLastYear	-0.0423826167	0.003914247
## WorkLifeBalance	0.0416530997	0.018378039
## YearsAtCompany	0.6440297779	0.765226243
## YearsInCurrentRole	0.5537100485	0.709461217
## YearsSinceLastPromotion	1.0000000000	0.508935431
## YearsWithCurrManager	0.5089354308	1.0000000000
## MonthlyIncome	0.3159115759	0.328487453
## LogMonthlyIncome	0.3199843020	0.379438998
## LogDistanceFromHome	-0.0179479960	-0.004369923
## LogJobLevel	0.3221550934	0.393859746
## LogPercentSalaryHike	-0.0817007976	-0.047220461
## NewRelationshipSatisfaction	0.0003609958	-0.044723637
## JobRoleNum	-0.0439791112	-0.048441140
## MaritalStatusNum	-0.0591018190	-0.032554537
## AttritionNum	-0.0045733208	-0.146782245
## GenderNum	-0.0420229129	-0.090619427
##	MonthlyIncome	LogMonthlyIncome
## ID	-5.163903e-02	-0.040350094
## Age	4.842883e-01	0.473524929
## DailyRate	8.790339e-05	0.023079873
## DistanceFromHome	-6.667155e-03	0.014025136
## Education	1.271314e-01	0.146085491
## EmployeeCount	NA	NA
## EmployeeNumber	3.420305e-02	0.031269177
## EnvironmentSatisfaction	-1.687744e-02	-0.027603625
## HourlyRate	2.391151e-03	-0.005613674
## JobInvolvement	4.473792e-04	0.003213698
## JobLevel	9.516400e-01	0.922461680
## JobSatisfaction	-5.314853e-02	-0.037564731
## MonthlyRate	6.459407e-02	0.081252155
## NumCompaniesWorked	1.558943e-01	0.158823008
## PercentSalaryHike	-5.386594e-02	-0.054692887
## PerformanceRating	-4.313100e-02	-0.043068808
## RelationshipSatisfaction	-3.884678e-03	-0.025328547
## StandardHours	NA	NA
## TotalWorkingYears	7.785112e-01	0.744885313
## TrainingTimesLastYear	-3.901457e-02	-0.043736817
## WorkLifeBalance	2.080489e-02	0.021477209
## YearsAtCompany	4.913790e-01	0.498186073
## YearsInCurrentRole	3.618405e-01	0.411676170
## YearsSinceLastPromotion	3.159116e-01	0.319984302
## YearsWithCurrManager	3.284875e-01	0.379438998
## MonthlyIncome	1.000000e+00	0.943215392
## LogMonthlyIncome	9.432154e-01	1.0000000000
## LogDistanceFromHome	-1.148026e-02	0.013589064
## LogJobLevel	8.888257e-01	0.930114585
## LogPercentSalaryHike	-5.645669e-02	-0.057709630
## NewRelationshipSatisfaction	6.878337e-03	-0.017739822
## JobRoleNum	-7.516553e-02	-0.054566670
## MaritalStatusNum	-8.103534e-02	-0.083514800
## AttritionNum	-1.549150e-01	-0.198619642
## GenderNum	-5.515709e-02	-0.069778474
##	LogJobLevel	LogPercentSalaryHike
## ID	-0.032529996	0.011019326
## Age	0.453675395	-0.027419490
## DailyRate	0.013812517	0.034205459
## DistanceFromHome	0.040015896	0.055001344
## Education	0.144187010	0.005435199
## EmployeeCount	NA	NA
## EmployeeNumber	0.024467719	-0.031616322
## EnvironmentSatisfaction	-0.001239649	-0.001344839
## HourlyRate	-0.012148140	-0.023989642
## JobInvolvement	-0.019300066	0.011297364

## JobLevel	0.968450462	-0.065523743
## JobSatisfaction	-0.041654226	0.012146664
## MonthlyRate	0.091297583	0.003771186
## NumCompaniesWorked	0.129499558	-0.018110672
## PercentSalaryHike	-0.056932169	0.993829811
## PerformanceRating	-0.030733506	0.722937450
## RelationshipSatisfaction	-0.015338256	-0.048172268
## StandardHours	NA	NA
## TotalWorkingYears	0.736640558	-0.066509032
## TrainingTimesLastYear	-0.047614528	0.003857571
## WorkLifeBalance	0.036119083	0.005546626
## YearsAtCompany	0.511826758	-0.062281811
## YearsInCurrentRole	0.416152551	-0.027811370
## YearsSinceLastPromotion	0.322155093	-0.081700798
## YearsWithCurrManager	0.393859746	-0.047220461
## MonthlyIncome	0.888825727	-0.056456686
## LogMonthlyIncome	0.930114585	-0.057709630
## LogDistanceFromHome	0.034772747	0.038329473
## LogJobLevel	1.000000000	-0.060631690
## LogPercentSalaryHike	-0.060631690	1.000000000
## NewRelationshipSatisfaction	-0.015993321	-0.039154872
## JobRoleNum	-0.054193091	-0.003348475
## MaritalStatusNum	-0.061655839	0.017553872
## AttritionNum	-0.189484919	0.010051703
## GenderNum	-0.081179711	0.009398543
##	NewRelationshipSatisfaction	JobRoleNum
## ID	-0.0311329988	-0.069017184
## Age	-0.0202290584	-0.105969885
## DailyRate	0.0205407252	0.006317723
## DistanceFromHome	0.0392233749	-0.012799586
## Education	0.0275426283	0.063075019
## EmployeeCount	NA	NA
## EmployeeNumber	-0.0051971592	-0.051991261
## EnvironmentSatisfaction	0.0078459438	-0.051917531
## HourlyRate	0.0182565532	-0.018204832
## JobInvolvement	0.0032689996	-0.011426141
## JobLevel	-0.0058025918	-0.074299446
## JobSatisfaction	-0.0092872334	0.008906795
## MonthlyRate	-0.0395502992	-0.013817053
## NumCompaniesWorked	0.0702707079	-0.008968013
## PercentSalaryHike	-0.0377058009	-0.004581115
## PerformanceRating	-0.0208256330	-0.030687047
## RelationshipSatisfaction	0.7745539953	-0.021511456
## StandardHours	NA	NA
## TotalWorkingYears	-0.0056634392	-0.133122913
## TrainingTimesLastYear	0.0085849932	-0.034165342
## WorkLifeBalance	0.0363452033	0.025982297
## YearsAtCompany	-0.0313258676	-0.091223937
## YearsInCurrentRole	-0.0319177287	-0.035873768
## YearsSinceLastPromotion	0.0003609958	-0.043979111
## YearsWithCurrManager	-0.0447236373	-0.048441140
## MonthlyIncome	0.0068783369	-0.075165525
## LogMonthlyIncome	-0.0177398218	-0.054566670
## LogDistanceFromHome	0.0404676071	-0.005261720
## LogJobLevel	-0.0159933209	-0.054193091
## LogPercentSalaryHike	-0.0391548719	-0.003348475
## NewRelationshipSatisfaction	1.0000000000	-0.054924505
## JobRoleNum	-0.0549245050	1.0000000000
## MaritalStatusNum	0.0575664124	0.049760110
## AttritionNum	-0.0547409841	0.090538602
## GenderNum	0.0479653236	-0.017882143
##	MaritalStatusNum	AttritionNum
## ID	-0.01907059	0.047266995
## Age	-0.07955784	-0.149383577
## DailyRate	-0.02539210	-0.033793125
## DistanceFromHome	-0.10447916	0.087136293
## Education	-0.01695547	-0.049442357
## EmployeeCount	NA	NA

## EmployeeNumber	-0.03119574	-0.022793834	-0.003729314
## EnvironmentSatisfaction	0.03032765	-0.077325405	0.021735539
## HourlyRate	-0.05933557	0.036554178	0.008523982
## JobInvolvement	-0.01497949	-0.187793409	0.023083817
## JobLevel	-0.06547074	-0.162136444	-0.069078459
## JobSatisfaction	0.01812999	-0.107520935	0.021121567
## MonthlyRate	-0.01625098	-0.043232173	-0.032817781
## NumCompaniesWorked	-0.01159984	0.061018887	-0.010620939
## PercentSalaryHike	0.01692518	0.015328069	0.003694383
## PerformanceRating	0.02058345	0.015333837	-0.014933022
## RelationshipSatisfaction	0.07069271	-0.039646611	0.013253439
## StandardHours	NA	NA	NA
## TotalWorkingYears	-0.07706499	-0.167206122	-0.072374947
## TrainingTimesLastYear	0.04561273	-0.062726088	0.001091037
## WorkLifeBalance	0.01119593	-0.089789709	-0.024037828
## YearsAtCompany	-0.06373189	-0.128754060	-0.081819101
## YearsInCurrentRole	-0.09440178	-0.156215707	-0.094931110
## YearsSinceLastPromotion	-0.05910182	-0.004573321	-0.042022913
## YearsWithCurrManager	-0.03255454	-0.146782245	-0.090619427
## MonthlyIncome	-0.08103534	-0.154914955	-0.055157089
## LogMonthlyIncome	-0.08351480	-0.198619642	-0.069778474
## LogDistanceFromHome	-0.07650610	0.068682895	0.005539853
## LogJobLevel	-0.06165584	-0.189484919	-0.081179711
## LogPercentSalaryHike	0.01755387	0.010051703	0.009398543
## NewRelationshipSatisfaction	0.05756641	-0.054740984	0.047965324
## JobRoleNum	0.04976011	0.090538602	-0.017882143
## MaritalStatusNum	1.00000000	0.049987703	0.047681065
## AttritionNum	0.04998770	1.000000000	0.025250336
## GenderNum	0.04768106	0.025250336	1.000000000

```
##
##
## processing file: report.rmd
```

```
##
|
|
|
|..
## inline R code fragments
##
##
|
|...
## label: global_options (with options)
## List of 1
## $ include: logi FALSE
##
##
|
|.....
## ordinary text without R code
##
##
|
|.....
## label: introduce
##
|
|.....
## ordinary text without R code
##
##
|
|.....
## label: plot_intro
```

```

##
|
|.....| 17%
## ordinary text without R code
##
##
|
|.....| 19%
## label: data_structure
##
|
|.....| 21%
## ordinary text without R code
##
##
|
|.....| 24%
## label: missing_profile

```

```

##
|
|.....| 26%
## ordinary text without R code
##
##
|
|.....| 29%
## label: univariate_distribution_header
##
|
|.....| 31%
## ordinary text without R code
##
##
|
|.....| 33%
## label: plot_histogram

```

```

##
|
|.....| 36%
## ordinary text without R code
##
##
|
|.....| 38%
## label: plot_density
##
|
|.....| 40%
## ordinary text without R code
##
##
|
|.....| 43%
## label: plot_frequency_bar

```

```

##
|
|.....| 45%
## ordinary text without R code
##
##
|
|.....| 48%
## label: plot_response_bar
##
|
|.....| 50%
## ordinary text without R code
##
##
|
|.....| 52%
## label: plot_with_bar
##
|
|.....| 55%
## ordinary text without R code
##
##
|
|.....| 57%
## label: plot_normal_qq

```

```

##
|
|.....| 60%
## ordinary text without R code
##
##
|
|.....| 62%
## label: plot_response_qq

```

```

##
|
|.....| 64%
## ordinary text without R code
##
##
|
|.....| 67%
## label: plot_by_qq
##
|
|.....| 69%
## ordinary text without R code
##
##
|
|.....| 71%
## label: correlation_analysis

```

```
##
|
|.....| 74%
## ordinary text without R code
##
##
|
|.....| 76%
## label: principal_component_analysis
```

```
##
|
|.....| 79%
## ordinary text without R code
##
##
|
|.....| 81%
## label: bivariate_distribution_header
##
|
|.....| 83%
## ordinary text without R code
##
##
|
|.....| 86%
## label: plot_response_boxplot
```

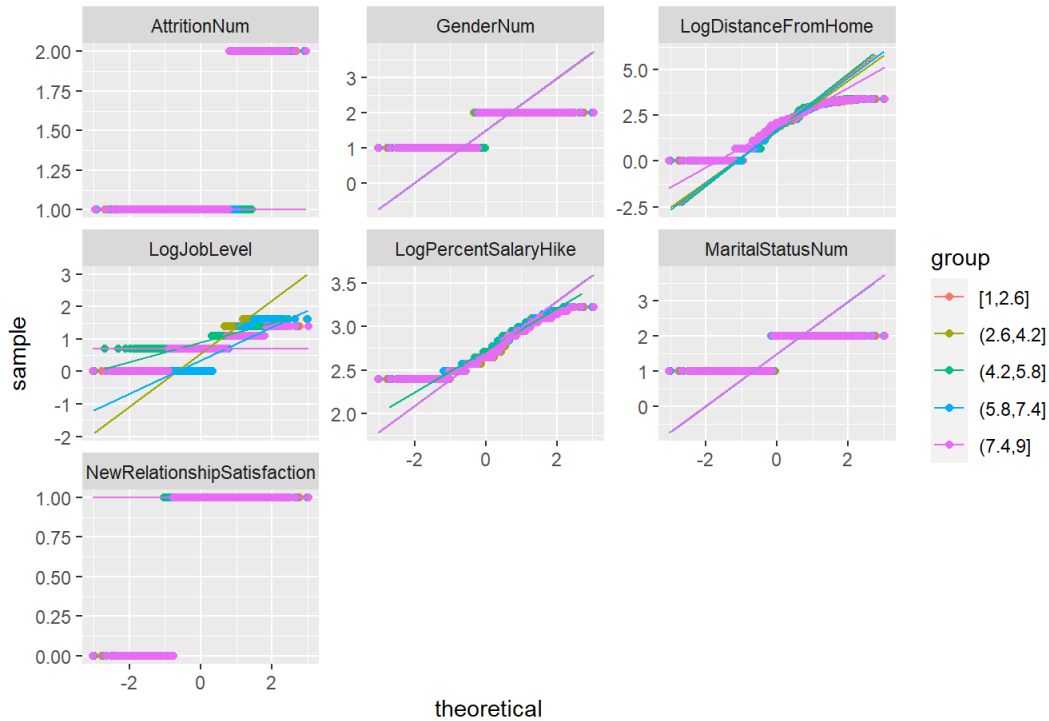
```
##
|
|.....| 88%
## ordinary text without R code
##
##
|
|.....| 90%
## label: plot_by_boxplot
##
|
|.....| 93%
## ordinary text without R code
##
##
|
|.....| 95%
## label: plot_response_scatterplot
```

```
##
|
|.....| 98%
## ordinary text without R code
##
##
|
|.....| 100%
## label: plot_by_scatterplot
```

```
## output file: C:/Users/Dawson/Documents/GitHub/Analytics/CaseStudy2DDS/report.knit.md
```

```
## "C:/Program Files/RStudio/bin/pandoc/pandoc" +RTS -K512m -RTS "C:/Users/Dawson/Documents/GitHub/Analytics/CaseStudy2DD
S/report.utf8.md" --to html4 --from markdown+autolink_bare_uris+tex_math_single_backslash --output pandoc349866e852c0.html
--email-obfuscation none --self-contained --standalone --section-divs --table-of-contents --toc-depth 6 --template "C:/Use
rs/Dawson/Documents/R/win-library/4.0/rmarkdown/rmd/h/default.html" --no-highlight --variable highlightjs=1 --variable "th
eme:yeti" --include-in-header "C:/Users/Dawson/AppData/Local/Temp/Rtmpe2a8Ct/rmarkdown-str34987ed25e77.html" --mathjax --v
ariable "mathjax-url:https://mathjax.rstudio.com/latest/MathJax.js?config=TeX-AMS-MML_HTMLorMML" --lua-filter "C:/Users/Da
wson/Documents/R/win-library/4.0/rmarkdown/rmd/lua/pagebreak.lua" --lua-filter "C:/Users/Dawson/Documents/R/win-library/4.
0/rmarkdown/rmd/lua/latex-div.lua"
```

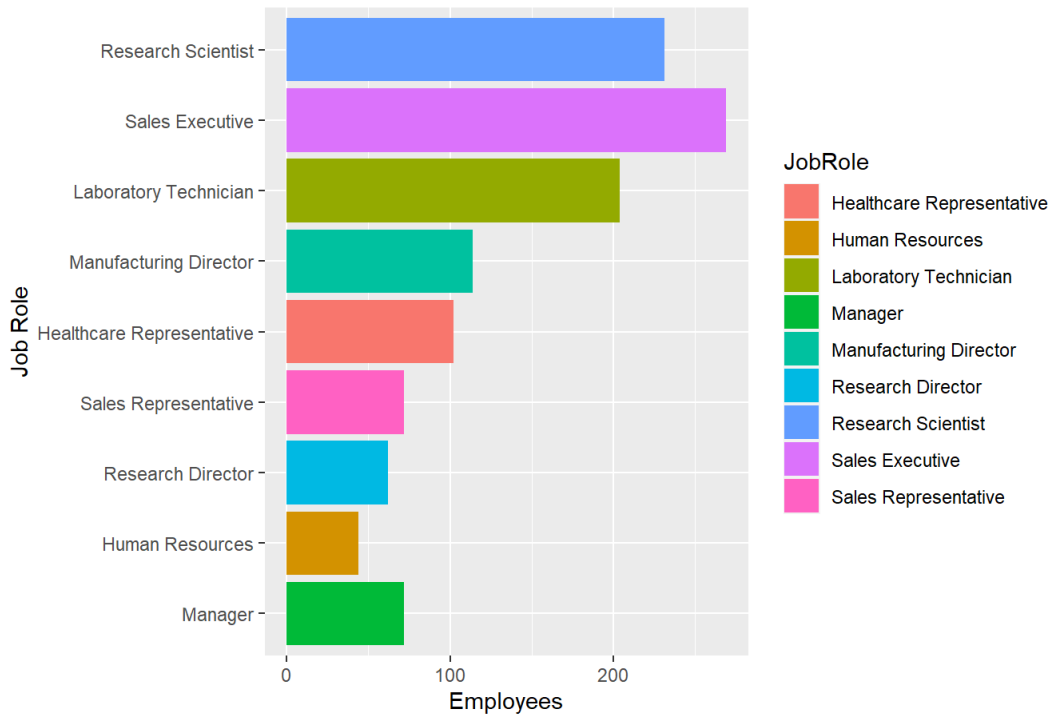
```
##
## Output created: report.html
```



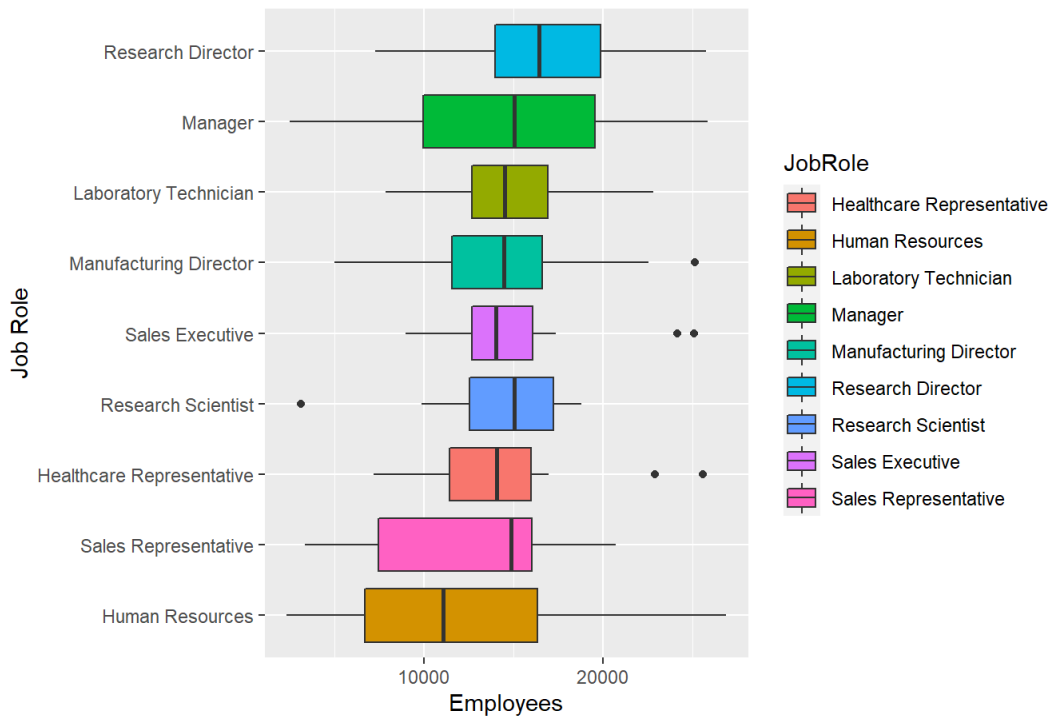
Page 4

```
## `summarise()` regrouping output by 'Department', 'JobRole', 'Attrition', 'JobLevel', 'Gender' (override with `.groups`
argument)
```

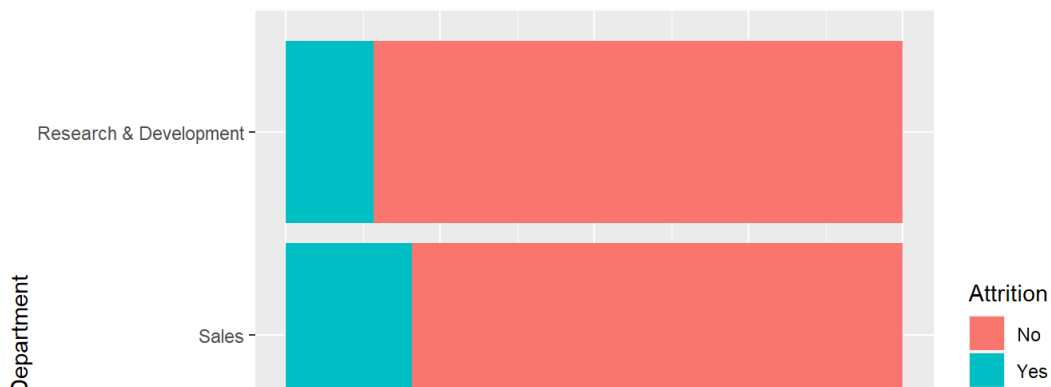
of Employees by Job Role



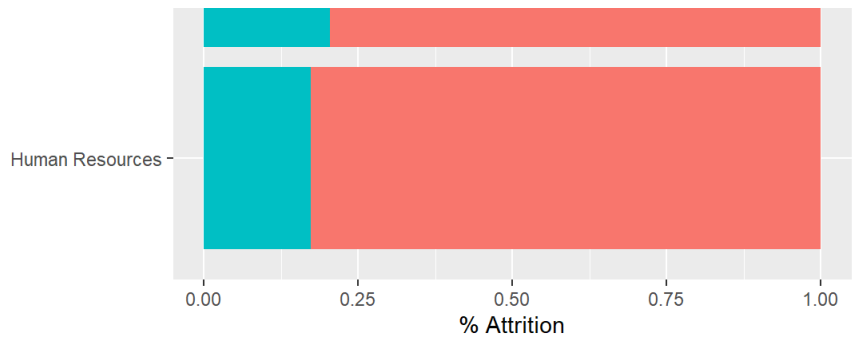
Monthly Rate (Mean) by Job Role



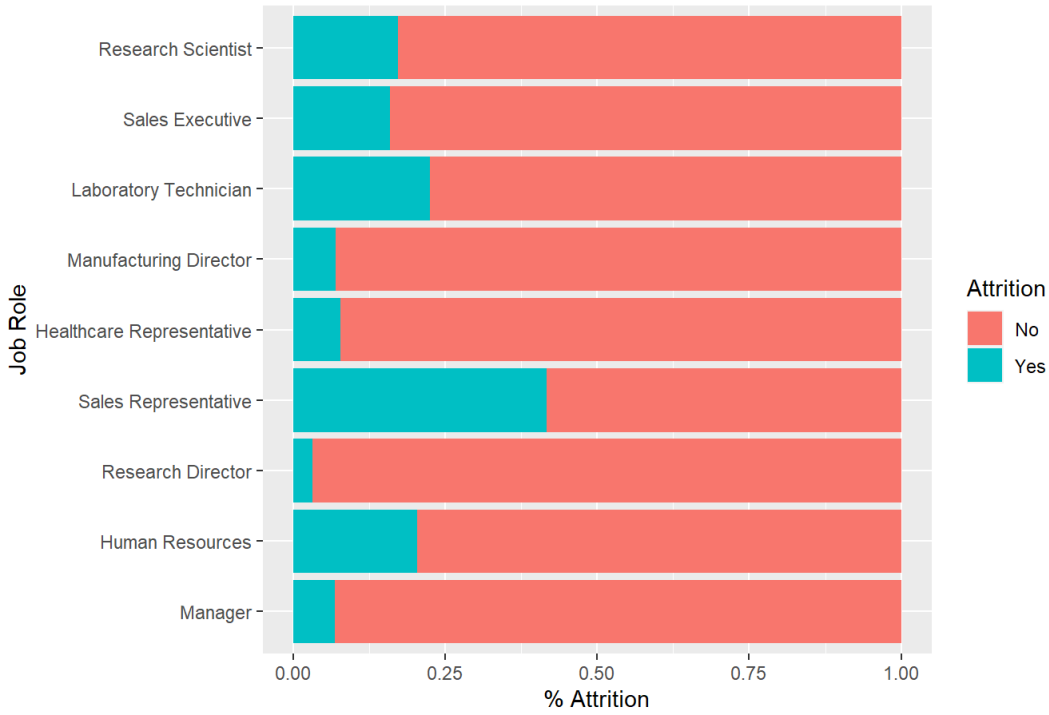
%Attrition Rate by Department



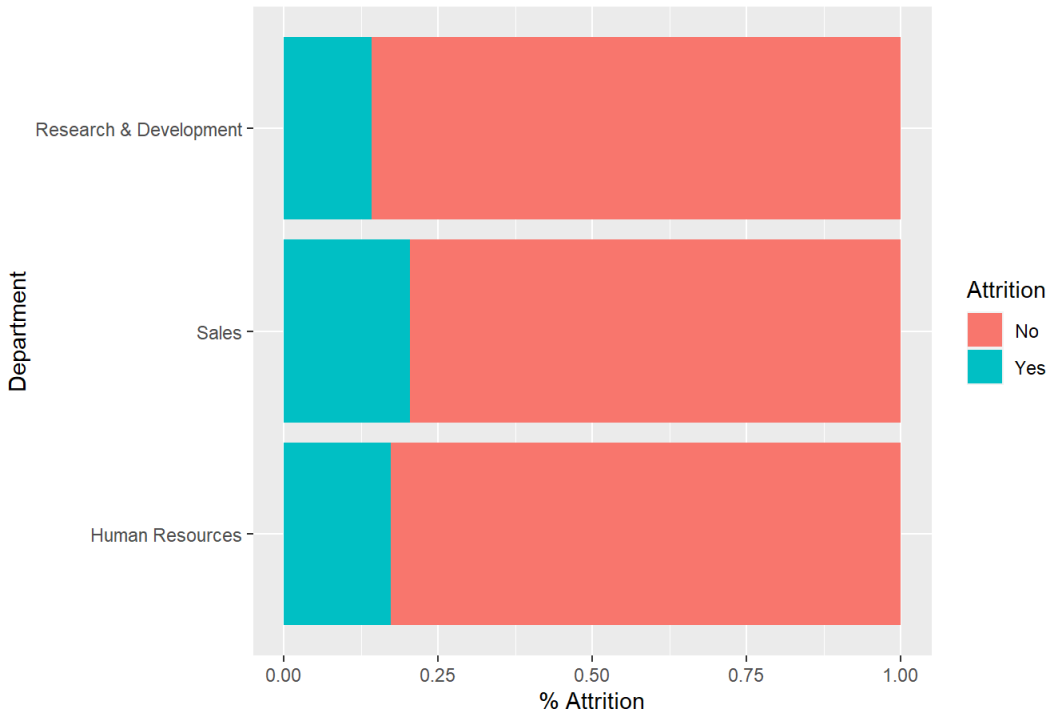
L



%Attrition Rate by Job Role

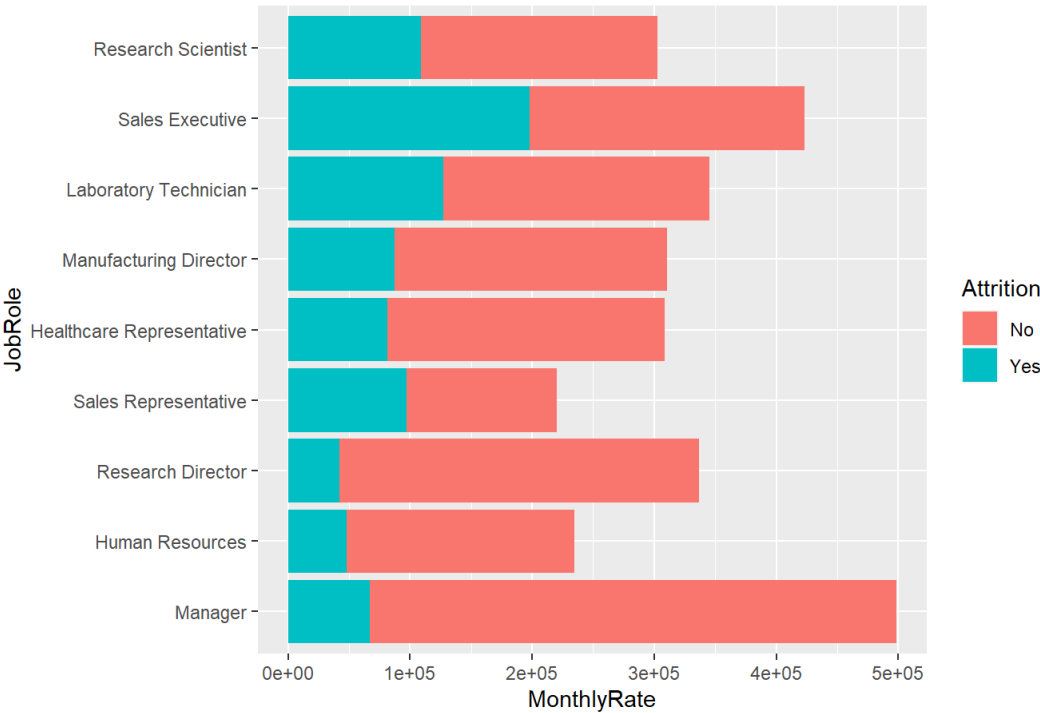


%Attrition Rate by Department

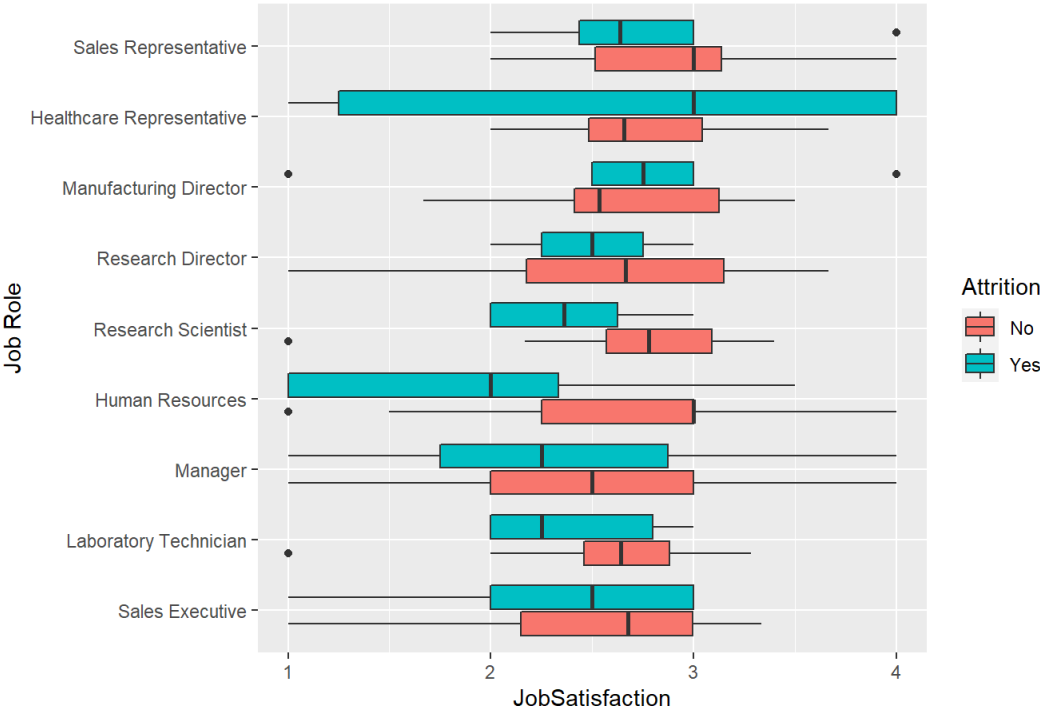


Warning: Ignoring unknown parameters: stat

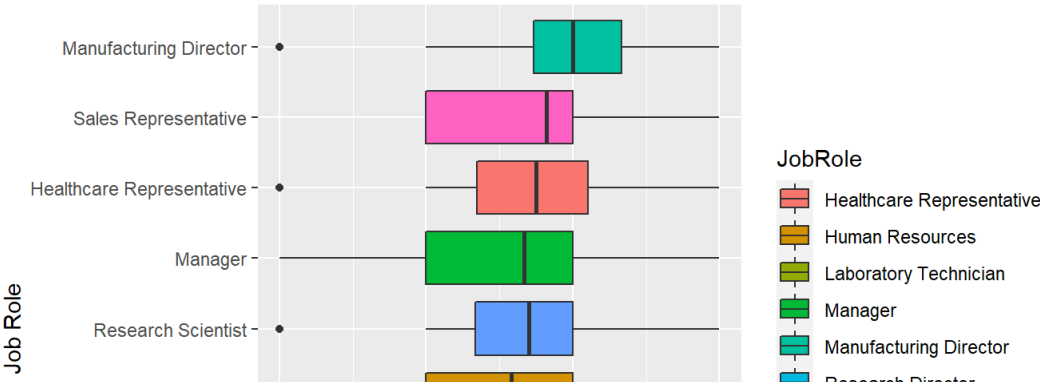
Monthly Rate by JobRole by Attrition

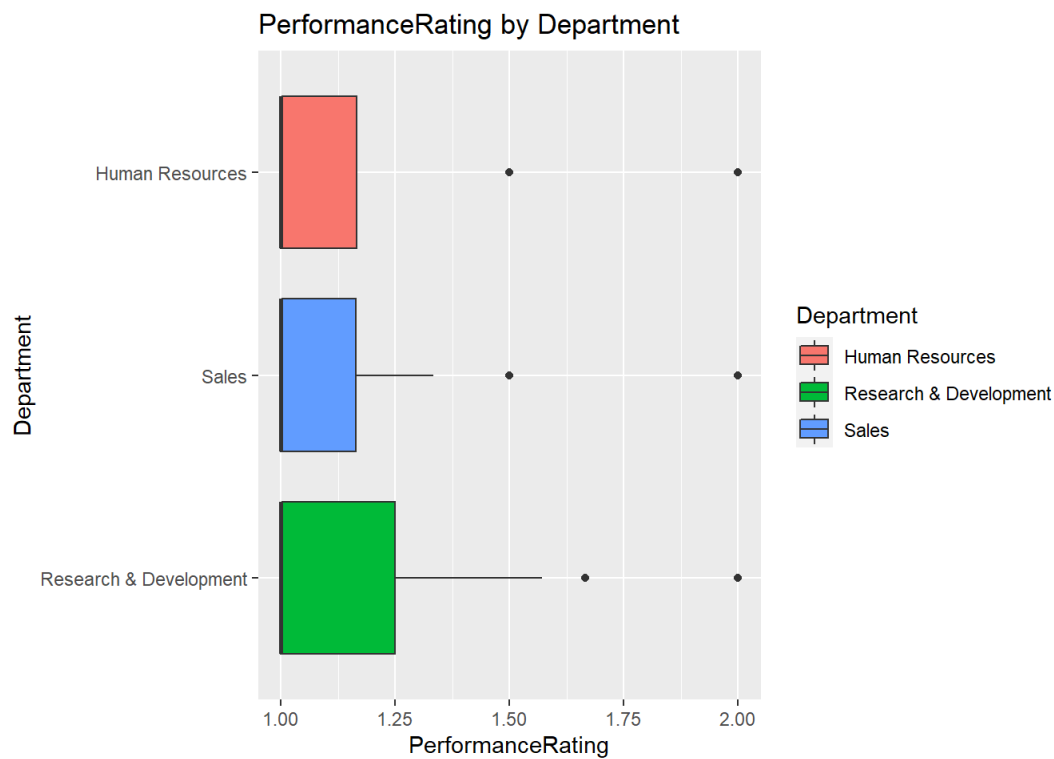
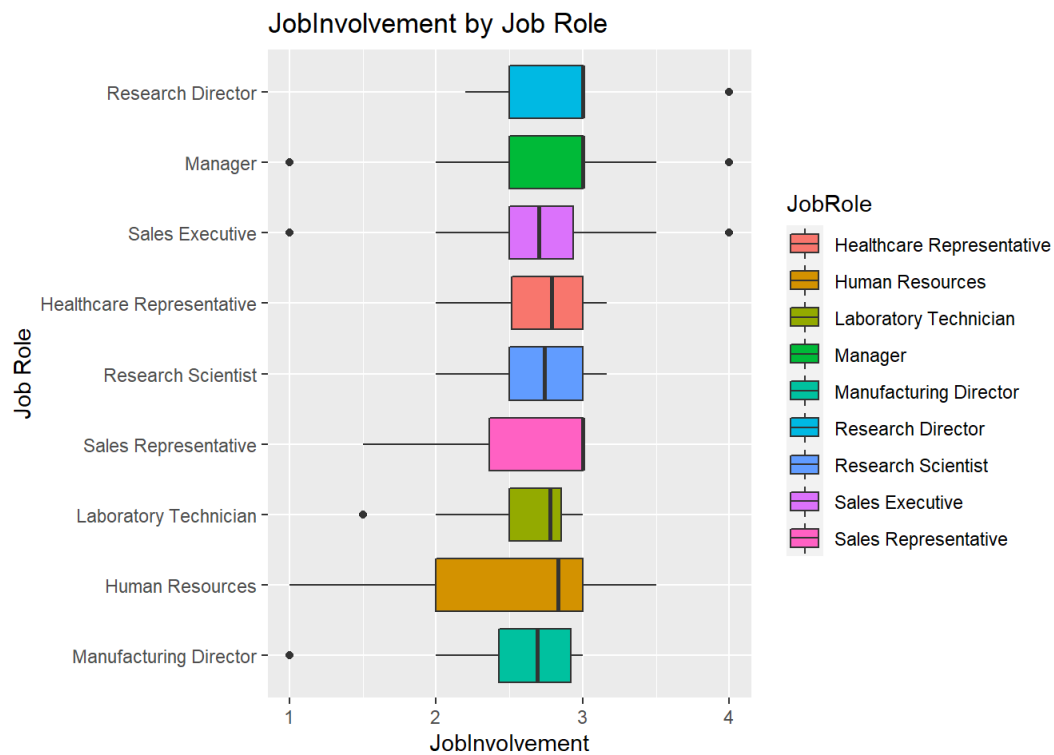
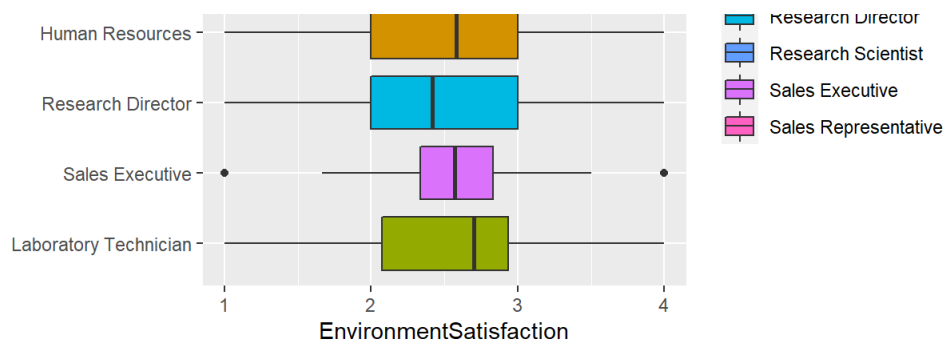


JobSatisfaction by Job Role

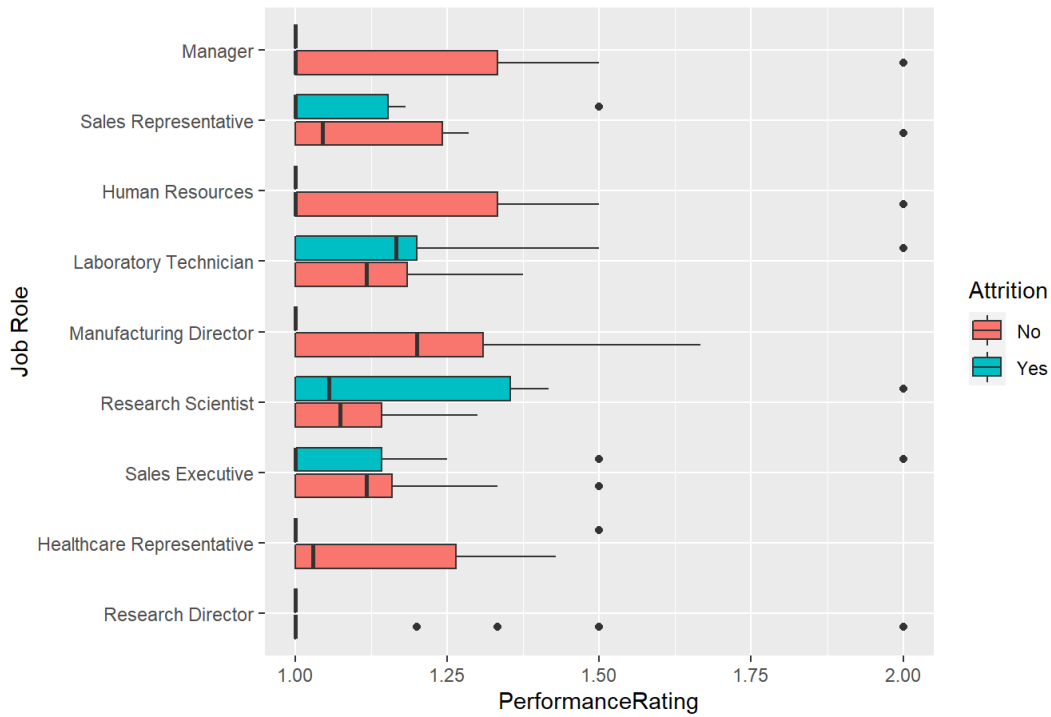


EnvironmentSatisfaction by Job Role

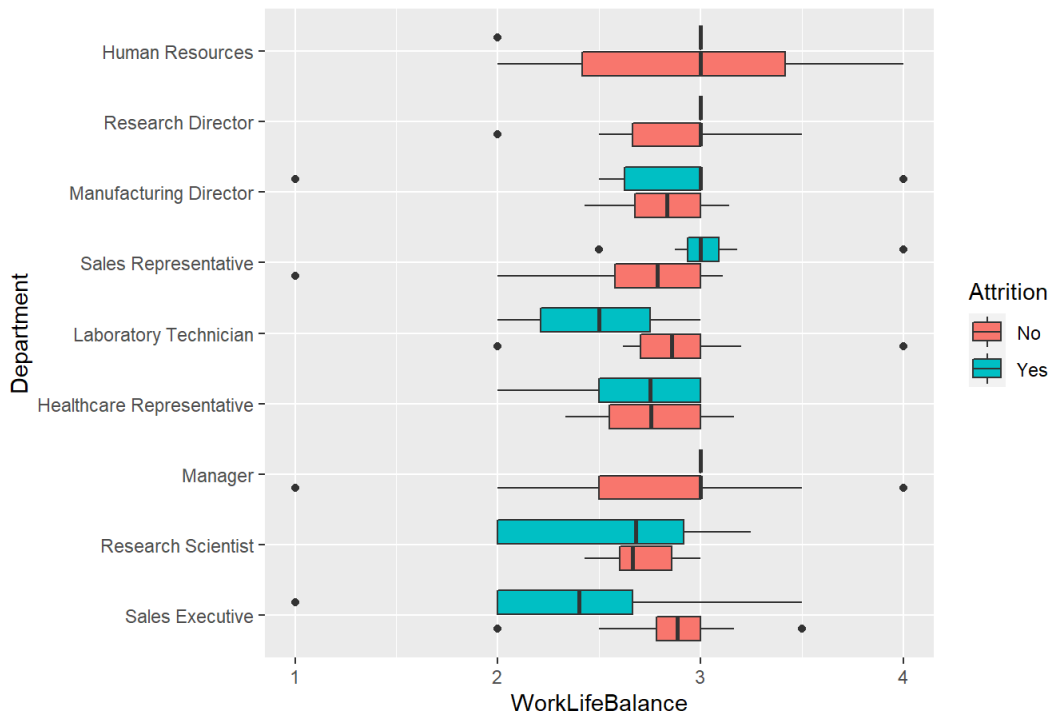




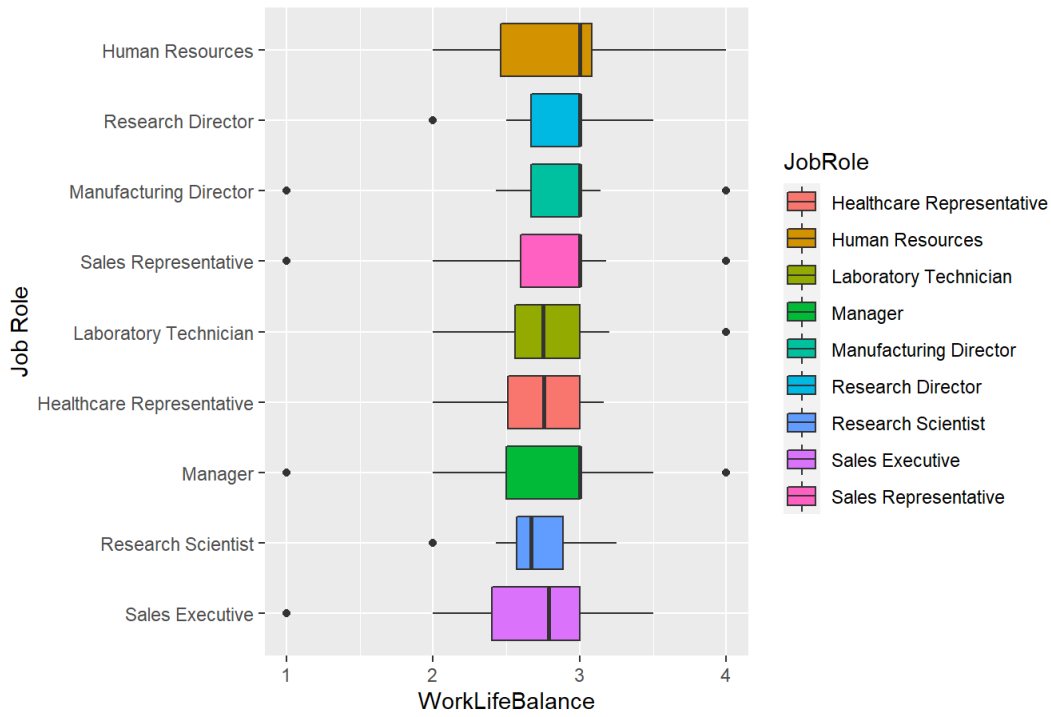
PerformanceRating by Job Role



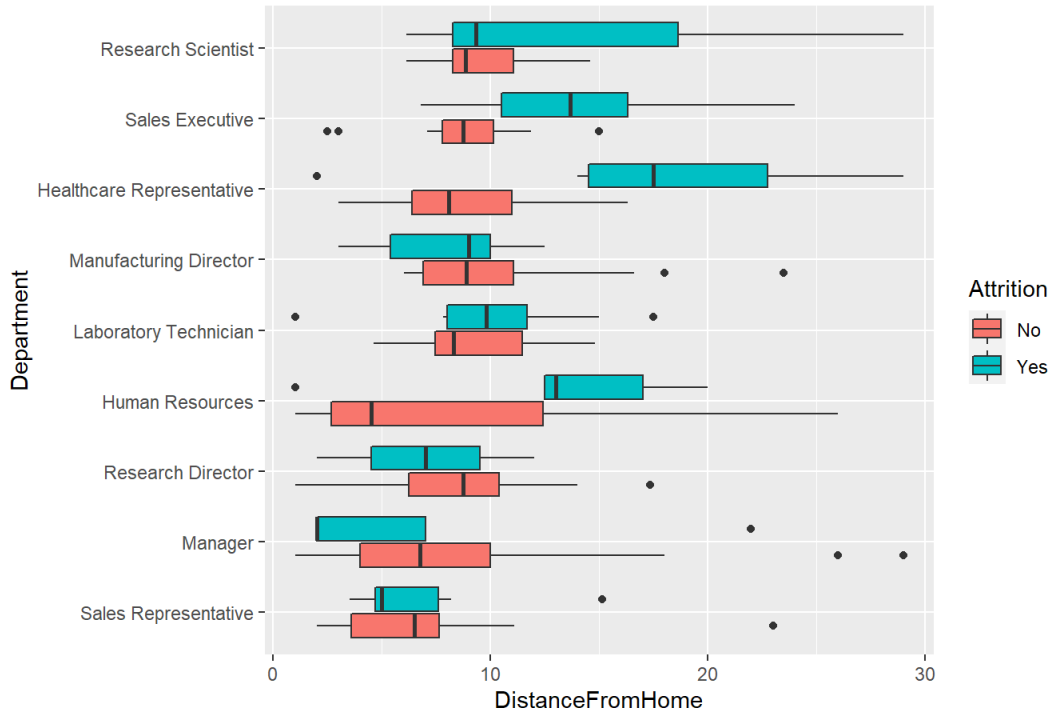
WorkLifeBalance & Attrition by JobRole



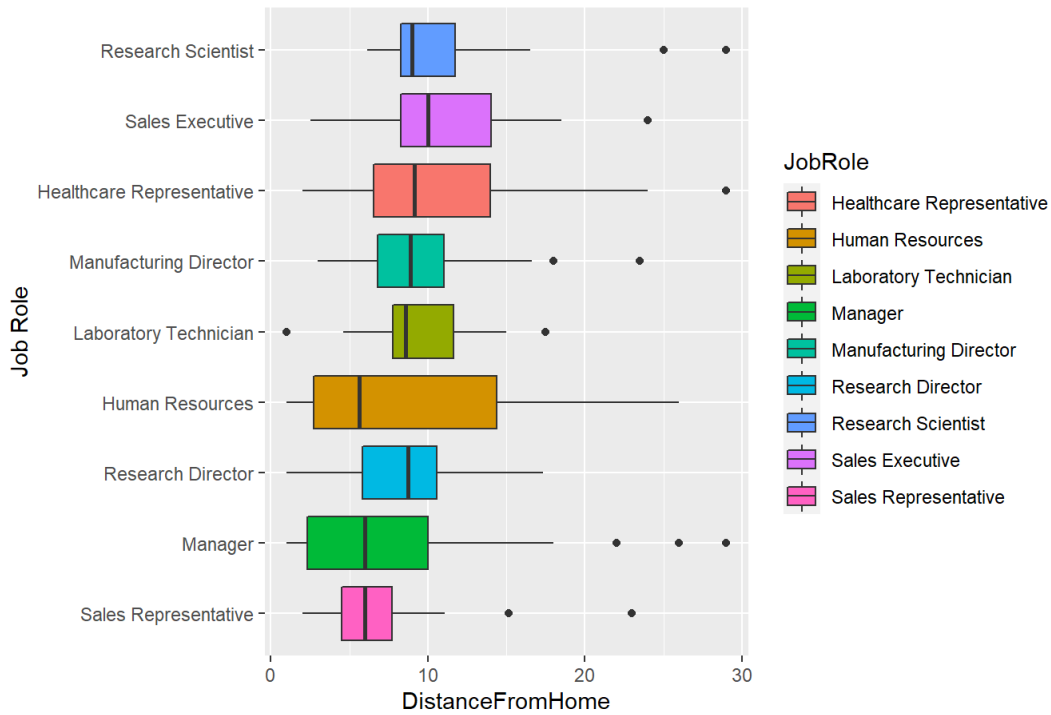
WorkLifeBalance by Job Role



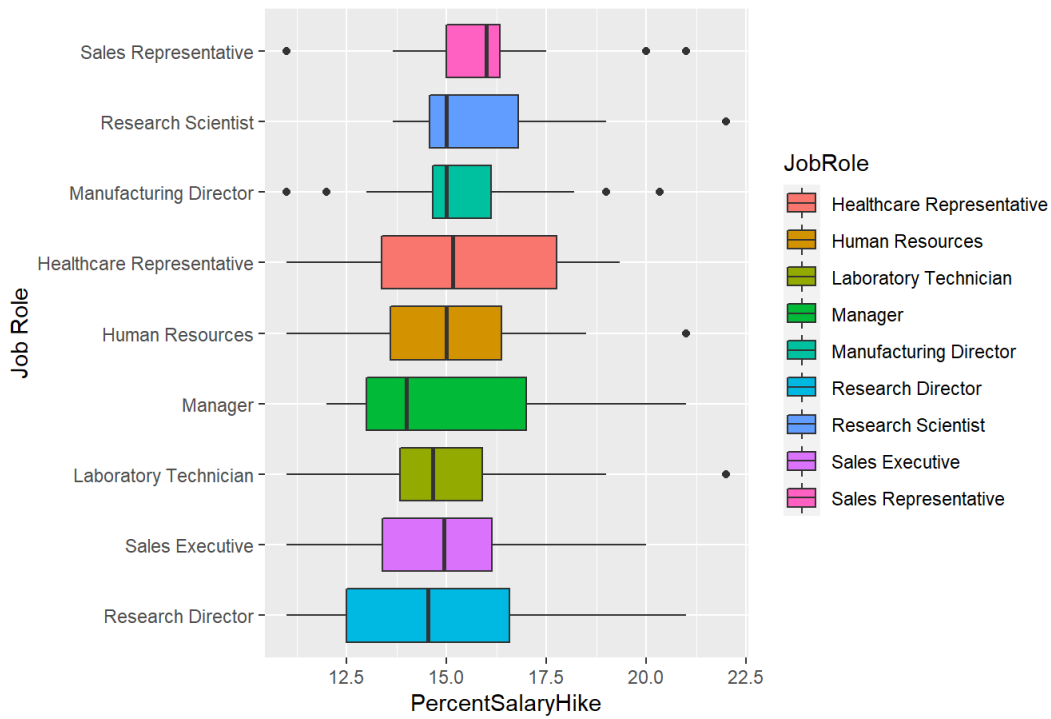
DistanceFromHome & Attrition by JobRole



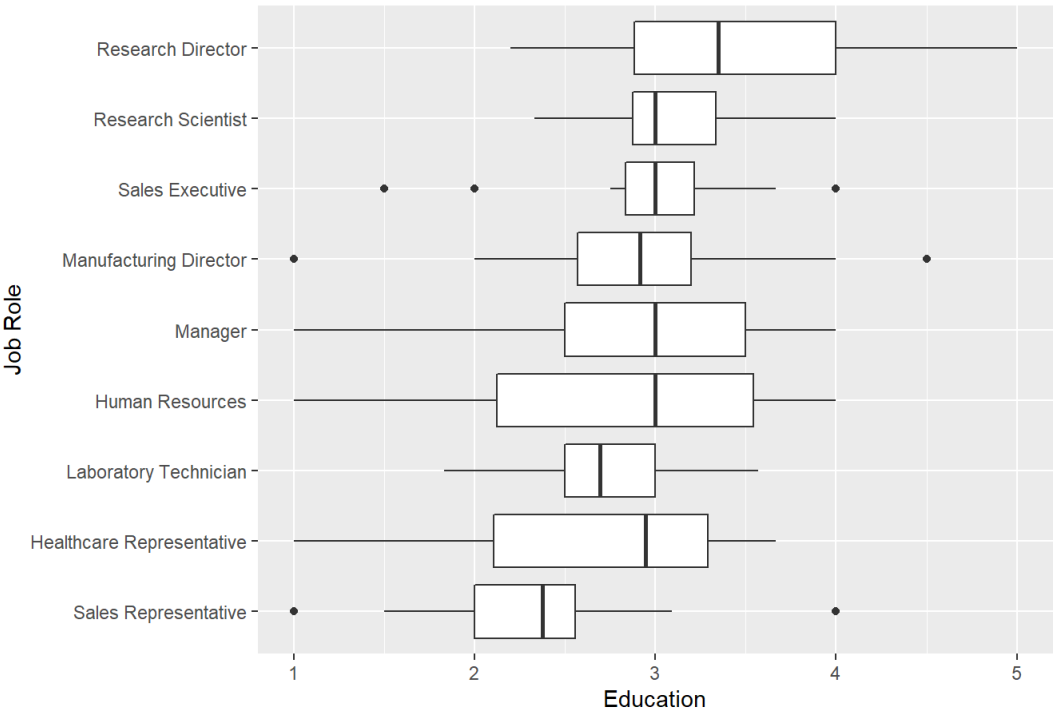
DistanceFromHome by Job Role



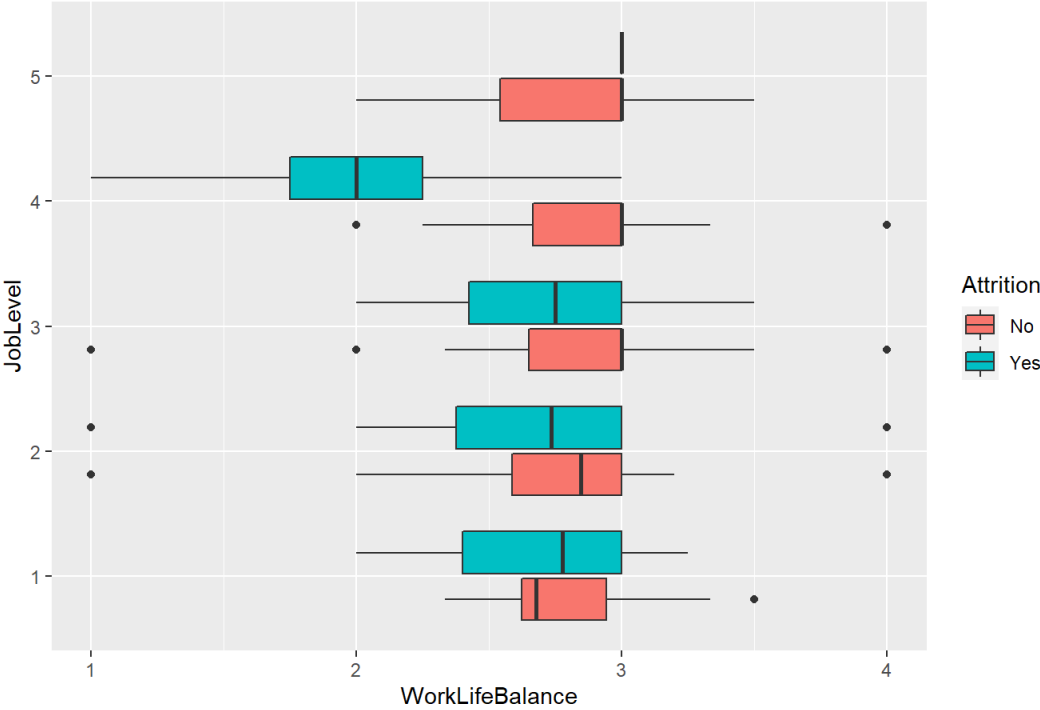
PercentSalaryHike by Job Role



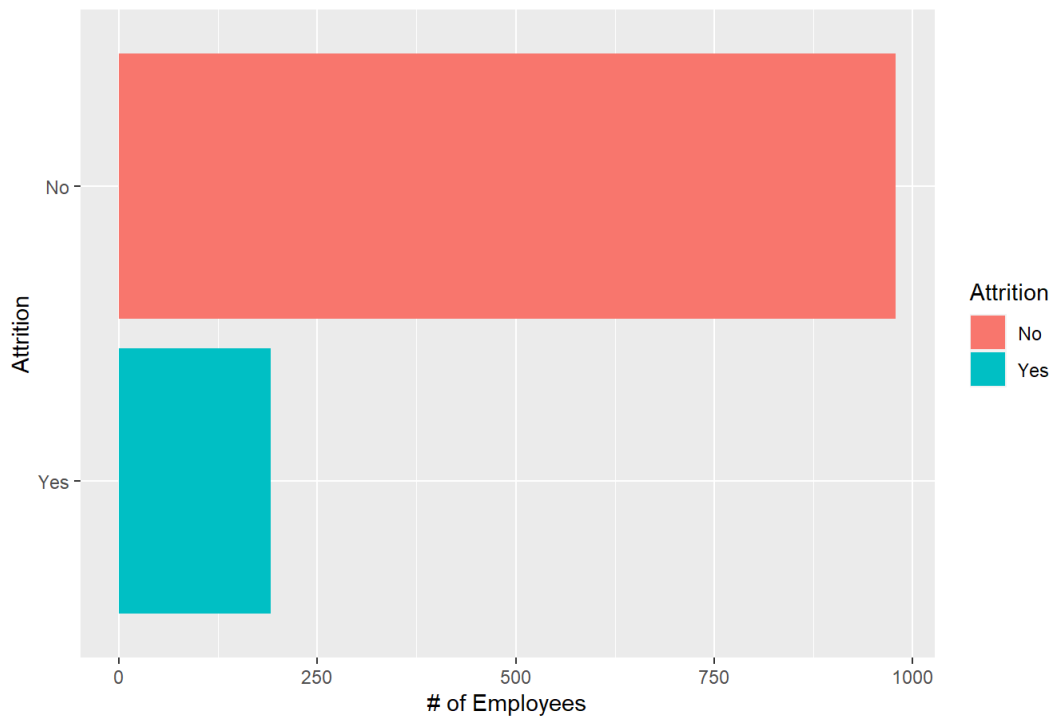
Education by Job Role



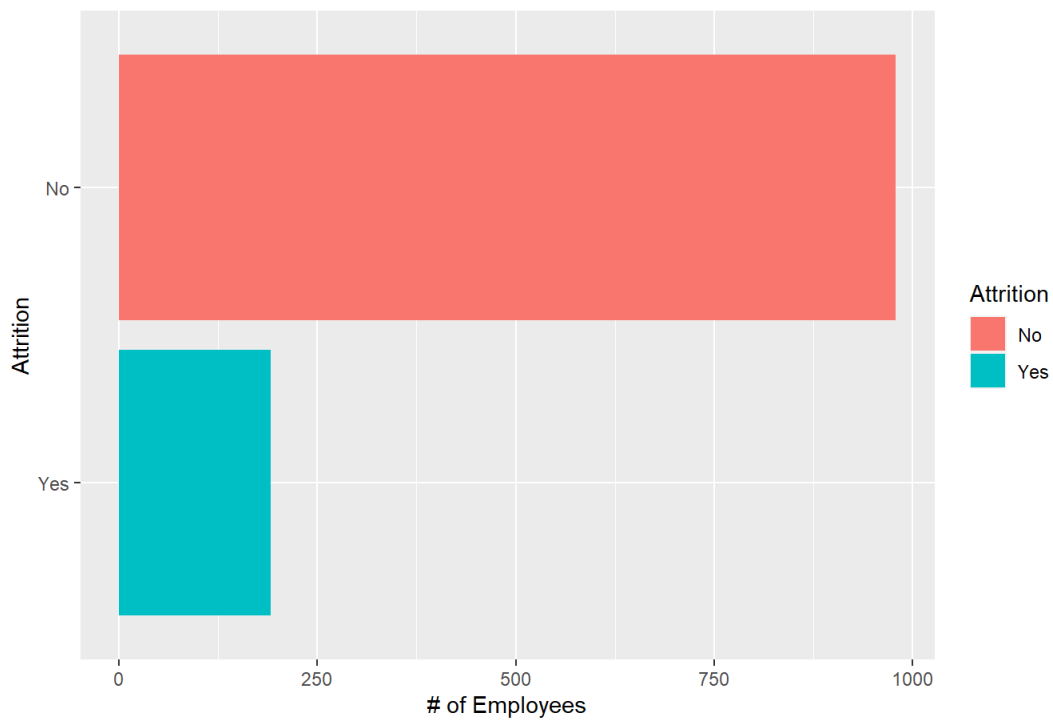
WorkLifeBalance & Attrition by JobLevel



of Employees by Attrition



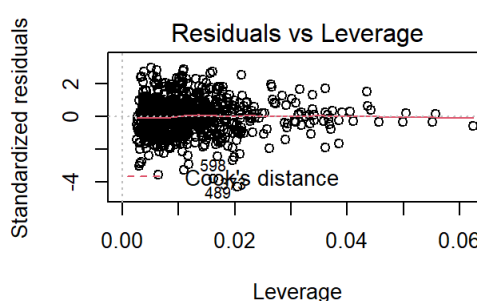
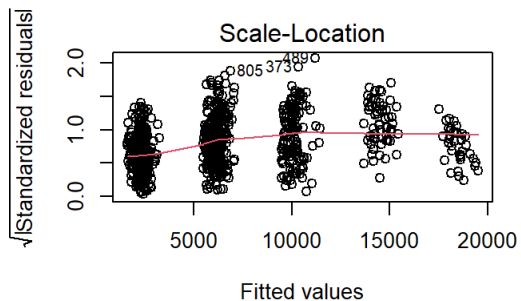
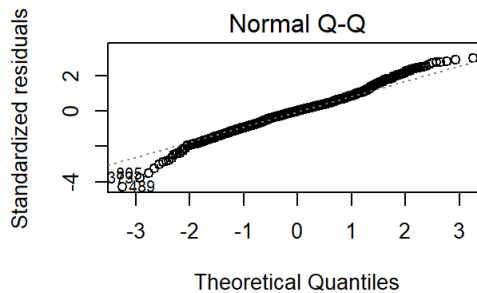
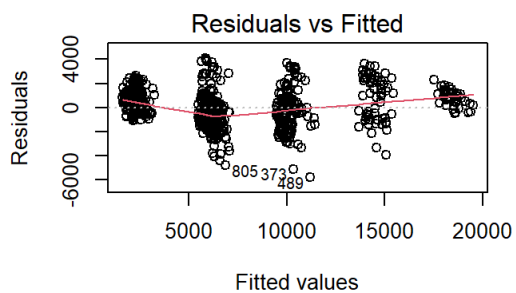
of Employees by Attrition



1. Multiple Linear Regression

Model for Predicting Monthly Income

```
##
## Call:
## lm(formula = MonthlyIncome ~ Age + JobLevel + TotalWorkingYears +
##     Department + YearsAtCompany + NewRelationshipSatisfaction +
##     PerformanceRating + Attrition, data = CaseStudyIncome)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5818.5  -851.0    22.8   746.9  4096.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1404.329    345.283   -4.067 5.20e-05 ***
## Age             -2.802      7.071   -0.396  0.69204
## JobLevel       3785.154    70.150   53.958 < 2e-16 ***
## TotalWorkingYears  65.769    13.132    5.008 6.67e-07 ***
## DepartmentResearch & Development -304.959    238.915   -1.276  0.20215
## DepartmentSales  -724.580    246.788   -2.936  0.00341 **
## YearsAtCompany  -31.031     10.275   -3.020  0.00260 **
## NewRelationshipSatisfaction  113.720    116.588    0.975  0.32964
## PerformanceRating4  -84.631    129.754   -0.652  0.51442
## AttritionYes      92.461     129.801    0.712  0.47645
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1370 on 860 degrees of freedom
## (300 observations deleted due to missingness)
## Multiple R-squared:  0.9121, Adjusted R-squared:  0.9112
## F-statistic: 991.3 on 9 and 860 DF, p-value: < 2.2e-16
```



```
##
## Call:
## lm(formula = MonthlyIncome ~ Age + JobLevel + TotalWorkingYears +
##     Department + YearsAtCompany + NewRelationshipSatisfaction +
##     PerformanceRating + Attrition, data = CaseStudyIncome)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5818.5  -851.0    22.8   746.9  4096.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1404.329    345.283   -4.067 5.20e-05 ***
## Age             -2.802      7.071   -0.396  0.69204
## JobLevel       3785.154    70.150   53.958 < 2e-16 ***
## TotalWorkingYears    65.769    13.132    5.008 6.67e-07 ***
## DepartmentResearch & Development  -304.959    238.915   -1.276  0.20215
## DepartmentSales   -724.580    246.788   -2.936  0.00341 **
## YearsAtCompany    -31.031     10.275   -3.020  0.00260 **
## NewRelationshipSatisfaction    113.720    116.588    0.975  0.32964
## PerformanceRating4    -84.631    129.754   -0.652  0.51442
## AttritionYes       92.461     129.801    0.712  0.47645
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1370 on 860 degrees of freedom
## (300 observations deleted due to missingness)
## Multiple R-squared:  0.9121, Adjusted R-squared:  0.9112
## F-statistic: 991.3 on 9 and 860 DF, p-value: < 2.2e-16
```

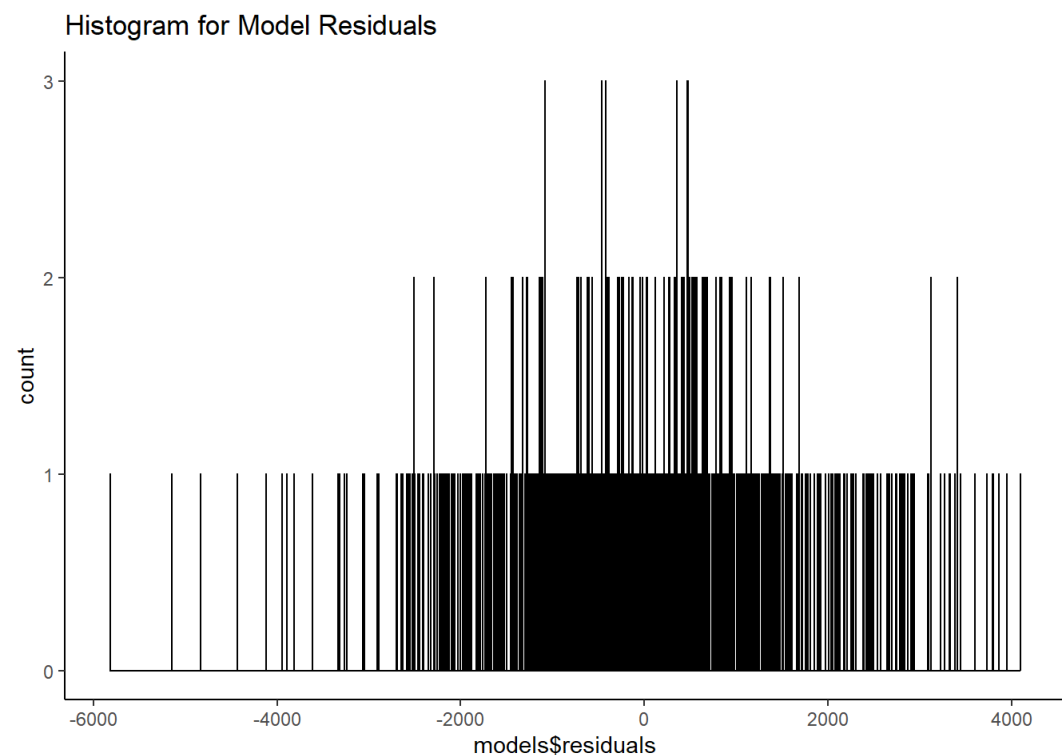
```
##              2.5 %    97.5 %
## (Intercept)   -2082.02500 -726.63380
## Age            -16.68040   11.07700
## JobLevel       3647.46799 3922.84015
## TotalWorkingYears    39.99394   91.54349
## DepartmentResearch & Development  -773.88467  163.96618
## DepartmentSales  -1208.95611 -240.20332
## YearsAtCompany   -51.19815  -10.86359
## NewRelationshipSatisfaction  -115.11043  342.54946
## PerformanceRating4  -339.30221  170.03961
## AttritionYes    -162.30228  347.22363
```

```
## [1] 1615114982
```

```
## [1] 1856454
```

```
## [1] 1362.518
```

```
## [1] 1878041
```



```
##          fit          lwr          upr
## 871  5593.179  2883.6665  8302.691
## 872  2219.160  -481.7527  4920.073
## 873 14052.656 11335.5311 16769.782
## 874  2177.782  -525.4661  4881.029
## 875  2243.384  -459.4729  4946.242
## 876  6756.888  4039.2976  9474.477
```

2. Knn/Bayes for categorizing Attrition

Data summary

```
Name          Attrition
Number of rows 1170
Number of columns 12
```

Column type frequency:

```
factor      1
numeric     11
```

```
Group variables  None
```

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
Attrition	0	1	FALSE	2	No: 979, Yes: 191

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
WorkLifeBalance	0	1	2.76	0.71	1.00	2.00	3.00	3.00	4.00	
JobInvolvement	0	1	2.73	0.70	1.00	2.00	3.00	3.00	4.00	
MaritalStatus	0	1	2.09	0.73	1.00	2.00	2.00	3.00	3.00	
DistanceFromHome	0	1	1.75	1.04	0.00	0.69	1.95	2.64	3.37	
OverTime	0	1	1.28	0.45	1.00	1.00	1.00	2.00	2.00	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
LogAge	0	1	3.57	0.25	2.89	3.40	3.56	3.74	4.09	
LogYearsWithCurrManager	0	1	1.34	0.80	0.00	1.10	1.39	2.08	2.83	
LogJobLevel	0	1	0.58	0.51	0.00	0.00	0.69	1.10	1.61	
PerformanceRating	0	1	1.15	0.36	1.00	1.00	1.00	1.00	2.00	
EducationField	0	1	3.26	1.32	1.00	2.00	3.00	4.00	6.00	
NumCompaniesWorked	0	1	2.73	2.53	0.00	1.00	2.00	4.00	9.00	

```
##
## Call:
## glm(formula = Attrition == "No" ~ ., data = Attrition)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08764  -0.04927   0.09731   0.20261   0.60449
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.368775    0.183321   2.012 0.044489 *
## WorkLifeBalance 0.030619    0.013841   2.212 0.027149 *
## JobInvolvement 0.072328    0.014002   5.166 2.82e-07 ***
## MaritalStatus  -0.077296    0.013593  -5.687 1.64e-08 ***
## DistanceFromHome -0.028525    0.009414  -3.030 0.002500 **
## OverTime       -0.212153    0.021856  -9.707 < 2e-16 ***
## LogAge         0.175772    0.047297   3.716 0.000212 ***
## LogYearsWithCurrManager 0.046415    0.013222   3.511 0.000464 ***
## LogJobLevel    0.067729    0.023175   2.923 0.003539 **
## PerformanceRating -0.001267    0.027144  -0.047 0.962791
## EducationField -0.003968    0.007424  -0.534 0.593148
## NumCompaniesWorked -0.016754    0.004147  -4.040 5.69e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1118626)
##
##      Null deviance: 159.82  on 1169  degrees of freedom
## Residual deviance: 129.54  on 1158  degrees of freedom
## AIC: 771.39
##
## Number of Fisher Scoring iterations: 2
```

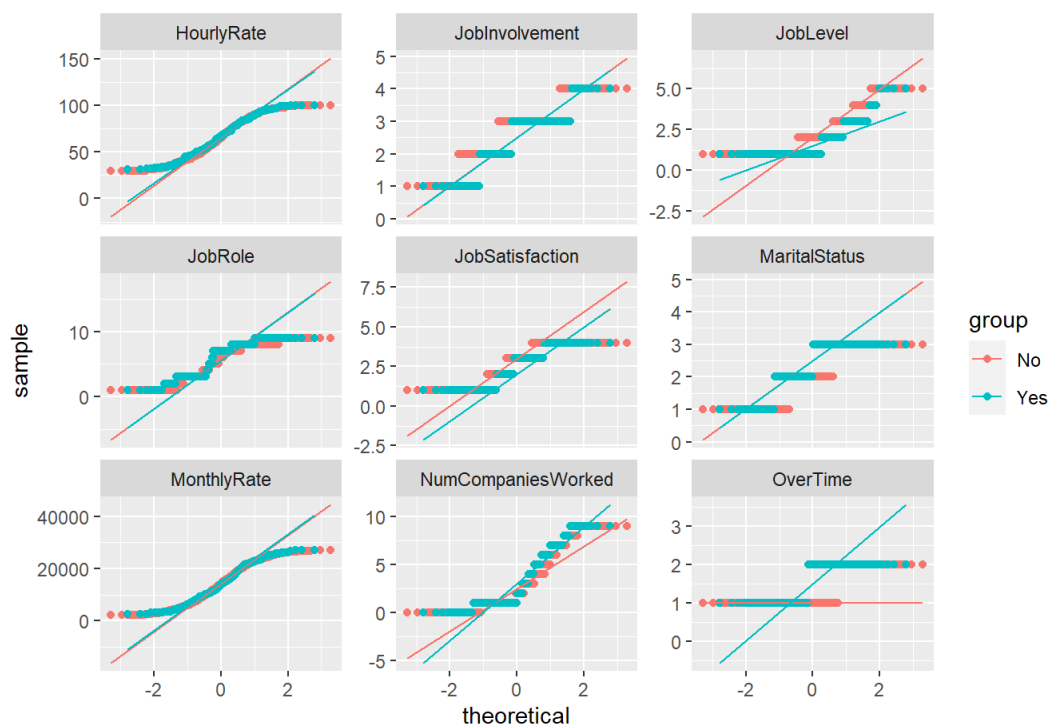
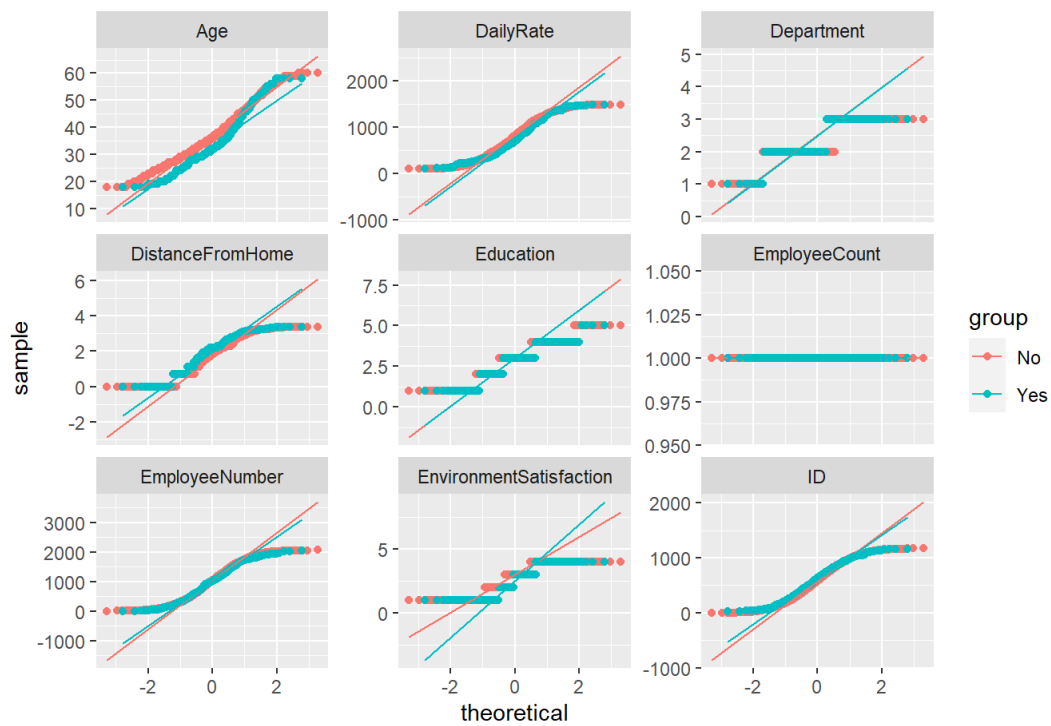
```
## Rows: 1,170
## Columns: 12
## $ WorkLifeBalance      <int> 2, 4, 3, 3, 3, 2, 2, 3, 3, 2, 3, 3, 4, 3, 4...
## $ JobInvolvement       <int> 3, 2, 3, 3, 3, 3, 4, 2, 3, 2, 2, 3, 3, 3, 3...
## $ MaritalStatus        <int> 1, 3, 3, 2, 3, 1, 2, 1, 2, 2, 2, 1, 3, 2, 3...
## $ DistanceFromHome     <dbl> 2.5649494, 2.6390573, 2.8903718, 0.0000000,...
## $ OverTime             <int> 1, 1, 1, 1, 2, 1, 2, 2, 2, 1, 1, 1, 1, 1, 1...
## $ LogAge               <dbl> 3.465736, 3.688879, 3.555348, 3.465736, 3.1...
## $ LogYearsWithCurrManager <dbl> 1.386294, 2.302585, 1.098612, 2.079442, 1.3...
## $ LogJobLevel          <dbl> 0.6931472, 1.6094379, 1.0986123, 1.0986123,...
## $ PerformanceRating    <int> 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 2, 1, 1...
## $ EducationField       <int> 2, 4, 2, 3, 6, 2, 4, 2, 2, 6, 4, 2, 5, 6, 2...
## $ NumCompaniesWorked   <dbl> 2, 1, 2, 1, 1, 1, 2, 2, 1, 1, 7, 1, 3, 1, 6...
## $ Attrition            <fct> No, No, No, No, No, No, No, No, No, No, No, No,...
```

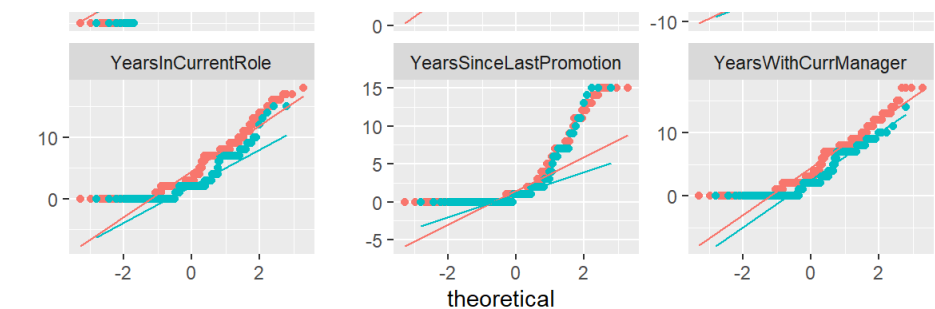
```
##
## classifications  No Yes
##                No 211 40
##                Yes 29 12
```

```
## Confusion Matrix and Statistics
##
##
## classifications      No  Yes
##                No  211  40
##                Yes   29  12
##
##                Accuracy : 0.7637
##                95% CI : (0.7107, 0.8112)
##      No Information Rate : 0.8219
##      P-Value [Acc > NIR] : 0.9952
##
##                Kappa : 0.1199
##
##  McNemar's Test P-Value : 0.2286
##
##                Sensitivity : 0.8792
##                Specificity : 0.2308
##                Pos Pred Value : 0.8406
##                Neg Pred Value : 0.2927
##                Prevalence : 0.8219
##                Detection Rate : 0.7226
##      Detection Prevalence : 0.8596
##      Balanced Accuracy : 0.5550
##
##      'Positive' Class : No
##
```

```
## [1] 76.36986
```

```
##
## classifications      No      Yes
##                No  0.72260274 0.13698630
##                Yes 0.09931507 0.04109589
```





Page 3

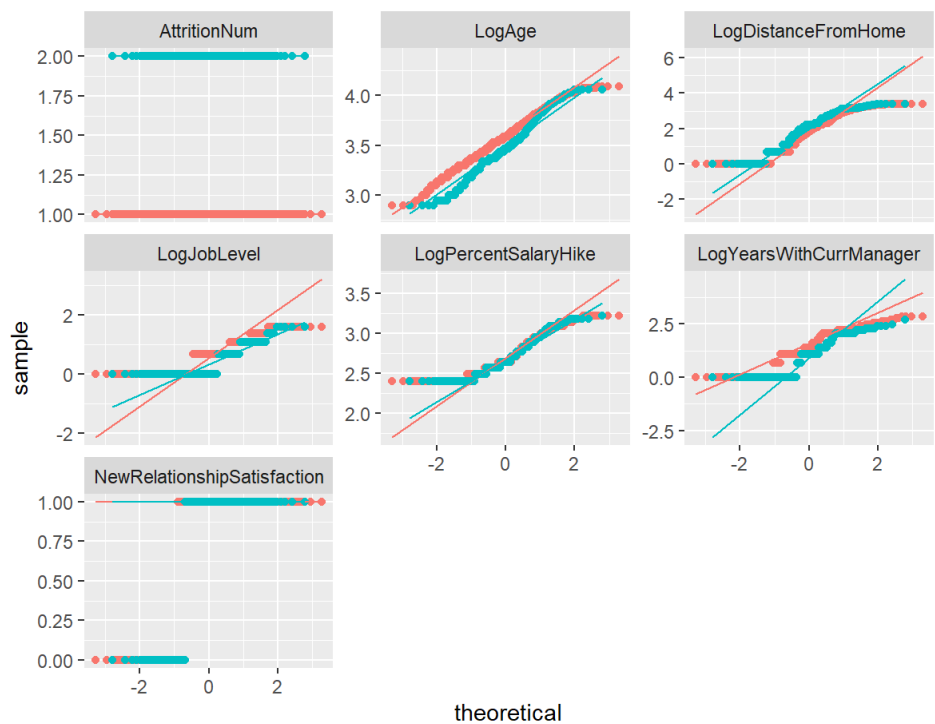


Table: Data summary

Page 4

Name AttritionNum
Number of rows 1170
Number of columns 42

Column type frequency:
character 1
factor 7
numeric 34

Group variables None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
EducationField	0	1	5	16	0	6	0



Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
BusinessTravel	0	1	FALSE	3	Tra: 832, Tra: 220, Non: 118
Gender	0	1	FALSE	2	Mal: 687, Fem: 483
Over18	0	1	FALSE	1	Y: 1170
PerformanceRating	0	1	FALSE	2	3: 990, 4: 180

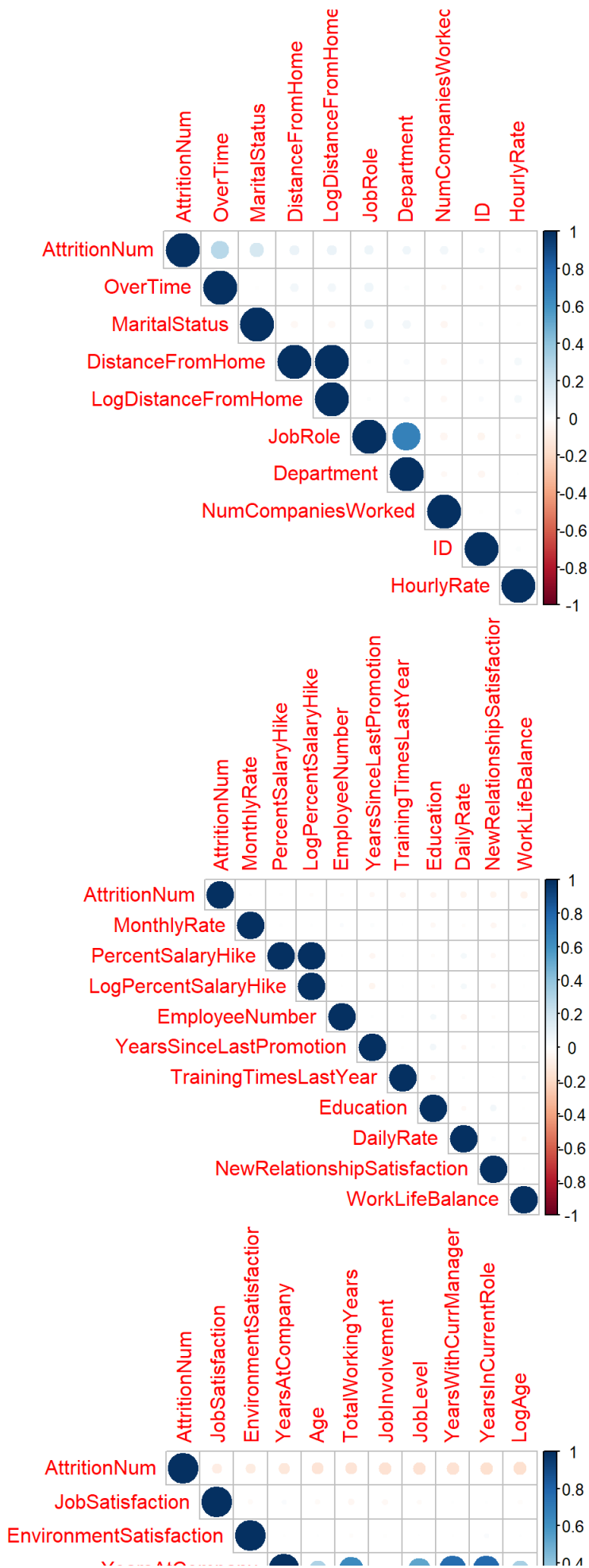
skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
RelationshipSatisfaction	0	1	FALSE	4	3: 349, 4: 343, 2: 249, 1: 229
StockOptionLevel	0	1	FALSE	4	0: 496, 1: 484, 2: 122, 3: 68
Attrition	0	1	FALSE	2	No: 979, Yes: 191

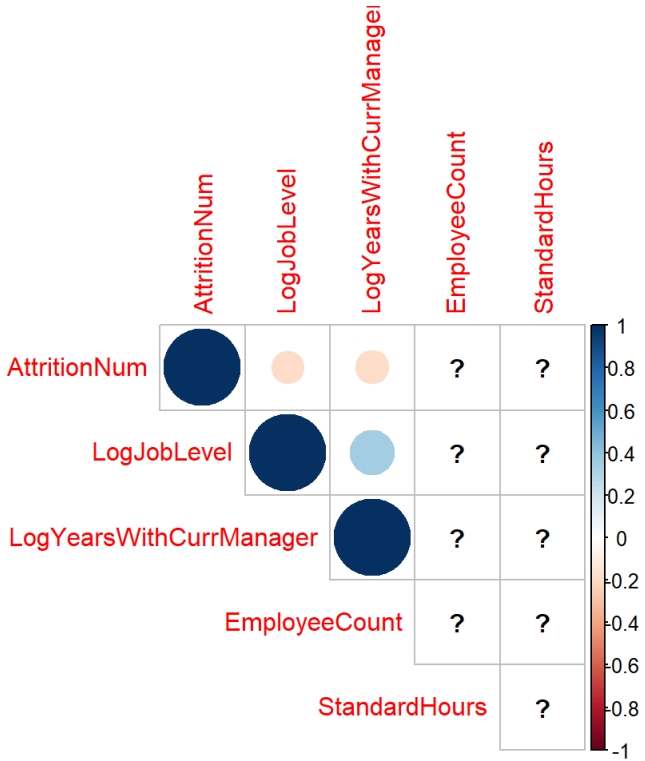
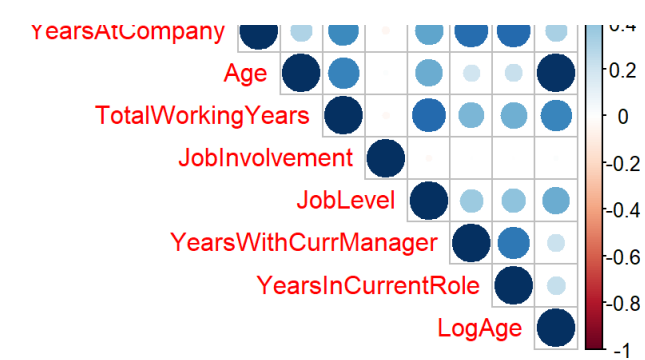
Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
ID	0	1	585.50	337.89	1.00	293.25	585.50	877.75	1170.00	
Age	0	1	36.68	9.00	18.00	30.00	35.00	42.00	60.00	
DailyRate	0	1	807.03	403.40	103.00	466.25	805.00	1167.75	1499.00	
Department	0	1	2.27	0.53	1.00	2.00	2.00	3.00	3.00	
DistanceFromHome	0	1	1.75	1.04	0.00	0.69	1.95	2.64	3.37	
Education	0	1	2.90	1.03	1.00	2.00	3.00	4.00	5.00	
EmployeeCount	0	1	1.00	0.00	1.00	1.00	1.00	1.00	1.00	
EmployeeNumber	0	1	1025.88	606.70	1.00	477.25	1029.50	1562.75	2068.00	
EnvironmentSatisfaction	0	1	2.72	1.10	1.00	2.00	3.00	4.00	4.00	
HourlyRate	0	1	65.85	20.37	30.00	48.00	66.00	84.00	100.00	
JobInvolvement	0	1	2.73	0.70	1.00	2.00	3.00	3.00	4.00	
JobLevel	0	1	2.03	1.07	1.00	1.00	2.00	3.00	5.00	
JobRole	0	1	5.51	2.48	1.00	3.00	6.00	8.00	9.00	
JobSatisfaction	0	1	2.72	1.11	1.00	2.00	3.00	4.00	4.00	
MaritalStatus	0	1	2.09	0.73	1.00	2.00	2.00	3.00	3.00	
MonthlyRate	0	1	14265.34	7113.29	2094.00	8009.75	13993.50	20461.50	26999.00	
NumCompaniesWorked	0	1	2.73	2.53	0.00	1.00	2.00	4.00	9.00	
OverTime	0	1	1.28	0.45	1.00	1.00	1.00	2.00	2.00	
PercentSalaryHike	0	1	15.22	3.67	11.00	12.00	14.00	18.00	25.00	
StandardHours	0	1	80.00	0.00	80.00	80.00	80.00	80.00	80.00	
TotalWorkingYears	0	1	10.98	7.54	0.00	6.00	9.50	15.00	40.00	
TrainingTimesLastYear	0	1	3.83	1.29	1.00	3.00	4.00	4.00	7.00	
WorkLifeBalance	0	1	2.76	0.71	1.00	2.00	3.00	3.00	4.00	
YearsAtCompany	0	1	6.88	5.90	0.00	3.00	5.00	9.00	40.00	
YearsInCurrentRole	0	1	4.20	3.65	0.00	2.00	3.00	7.00	18.00	
YearsSinceLastPromotion	0	1	2.16	3.17	0.00	0.00	1.00	3.00	15.00	
YearsWithCurrManager	0	1	4.06	3.48	0.00	2.00	3.00	7.00	17.00	
LogDistanceFromHome	0	1	1.75	1.04	0.00	0.69	1.95	2.64	3.37	
LogJobLevel	0	1	0.58	0.51	0.00	0.00	0.69	1.10	1.61	
LogPercentSalaryHike	0	1	2.70	0.23	2.40	2.48	2.64	2.89	3.22	
NewRelationshipSatisfaction	0	1	0.80	0.40	0.00	1.00	1.00	1.00	1.00	
AttritionNum	0	1	1.16	0.37	1.00	1.00	1.00	1.00	2.00	

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
LogAge	0	1	3.57	0.25	2.89	3.40	3.56	3.74	4.09	
LogYearsWithCurrManager	0	1	1.34	0.80	0.00	1.10	1.39	2.08	2.83	

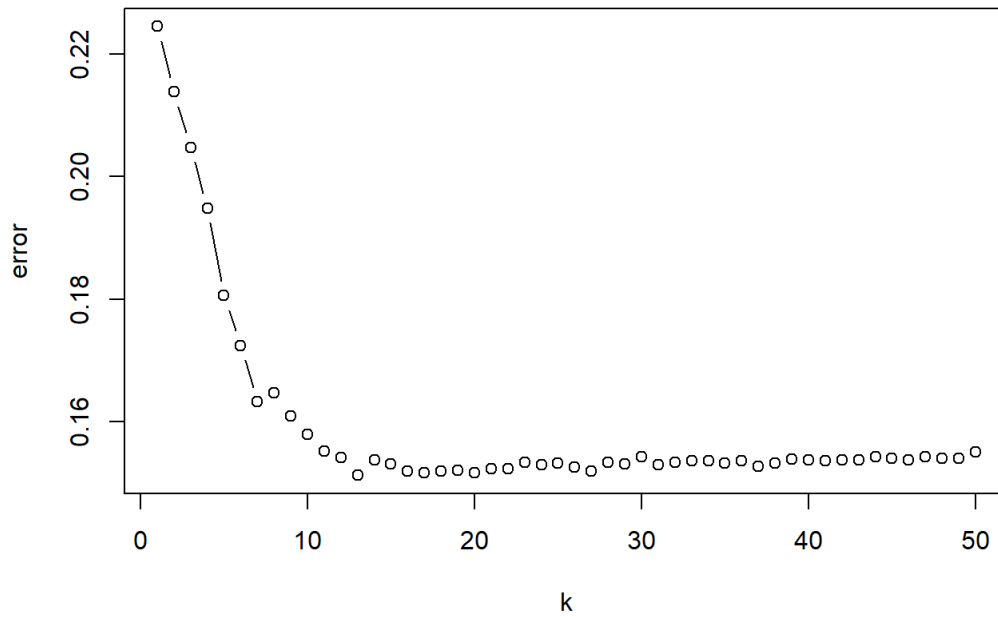
```
## Warning in cor(., use = "pair"): the standard deviation is zero
```





```
##
## Parameter tuning of 'knn.wrapper':
##
## - sampling method: bootstrapping
##
## - best parameters:
##   k
##  13
##
## - best performance: 0.1512565
##
## - Detailed performance results:
##   k      error dispersion
## 1  1 0.2244618 0.01326521
## 2  2 0.2137879 0.02007164
## 3  3 0.2047474 0.01589288
## 4  4 0.1948669 0.01201163
## 5  5 0.1805387 0.01652793
## 6  6 0.1724497 0.01600826
## 7  7 0.1632681 0.01694238
## 8  8 0.1646931 0.01210658
## 9  9 0.1609343 0.01555660
## 10 10 0.1579737 0.01810345
## 11 11 0.1552261 0.01778738
## 12 12 0.1541915 0.01629118
## 13 13 0.1512565 0.01653256
## 14 14 0.1537789 0.01426329
## 15 15 0.1531550 0.01455935
## 16 16 0.1519151 0.01575512
## 17 17 0.1516850 0.01641045
## 18 18 0.1518750 0.01639725
## 19 19 0.1521074 0.01593968
## 20 20 0.1516801 0.01547485
## 21 21 0.1523184 0.01590033
## 22 22 0.1523123 0.01795997
## 23 23 0.1533741 0.01785168
## 24 24 0.1529860 0.01735299
## 25 25 0.1531792 0.01671755
## 26 26 0.1525573 0.01650787
## 27 27 0.1519314 0.01600842
## 28 28 0.1533882 0.01648265
## 29 29 0.1531698 0.01676772
## 30 30 0.1542269 0.01573432
## 31 31 0.1529799 0.01621397
## 32 32 0.1534016 0.01622725
## 33 33 0.1536128 0.01625341
## 34 34 0.1536060 0.01620819
## 35 35 0.1531854 0.01613385
## 36 36 0.1536038 0.01562876
## 37 37 0.1527759 0.01750564
## 38 38 0.1532045 0.01695247
## 39 39 0.1538320 0.01741032
## 40 40 0.1538193 0.01821573
## 41 41 0.1536068 0.01769055
## 42 42 0.1538169 0.01742052
## 43 43 0.1538113 0.01793410
## 44 44 0.1542288 0.01788603
## 45 45 0.1540227 0.01810263
## 46 46 0.1538121 0.01822763
## 47 47 0.1542347 0.01843780
## 48 48 0.1540246 0.01834149
## 49 49 0.1540274 0.01850095
## 50 50 0.1550798 0.01733889
```

Performance of `knn.wrapper`



3. Moving to Alternative to Knn ->

SVM

```
##
## Call:
## svm(formula = Attrition ~ ., data = Attrition)
##
##
## Parameters:
##   SVM-Type:  C-classification
##   SVM-Kernel: radial
##     cost:  1
##
## Number of Support Vectors:  483
##
## ( 300 183 )
##
## Number of Classes:  2
##
## Levels:
##   No Yes
```

```
##
## Call:
## svm.default(x = x, y = y)
##
##
## Parameters:
##   SVM-Type:  C-classification
##   SVM-Kernel: radial
##         cost: 1
##
## Number of Support Vectors: 483
##
## ( 300 183 )
##
##
## Number of Classes: 2
##
## Levels:
## No Yes
```

```
##   user  system elapsed
##   0.07   0.00   0.11
```

```
##      y
## pred  No Yes
##   No  974 142
##   Yes   5  49
```

```
## [1] 0.9948927
```

```
## [1] 0.2565445
```

```
##
## Parameter tuning of 'svm':
##
## - sampling method: 10-fold cross validation
##
## - best parameters:
##   cost gamma
##     1    0.5
##
## - best performance: 0.1606838
```

```
##
## Call:
## svm(formula = Attrition ~ ., data = Attrition, kernel = "radial",
##      cost = 1, gamma = 0.5)
##
##
## Parameters:
##   SVM-Type:  C-classification
## SVM-Kernel:  radial
##      cost:   1
##
## Number of Support Vectors: 1001
##
## ( 810 191 )
##
##
## Number of Classes: 2
##
## Levels:
## No Yes
```

```
##   user  system elapsed
##  0.16   0.00   0.22
```

```
##      y
## pred  No Yes
##   No  979 40
##   Yes   0 151
```

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

```
## [1] 0.7905759
```

```
## 1 2 3 4 5 6
## No No No No No No
## Levels: No Yes
```