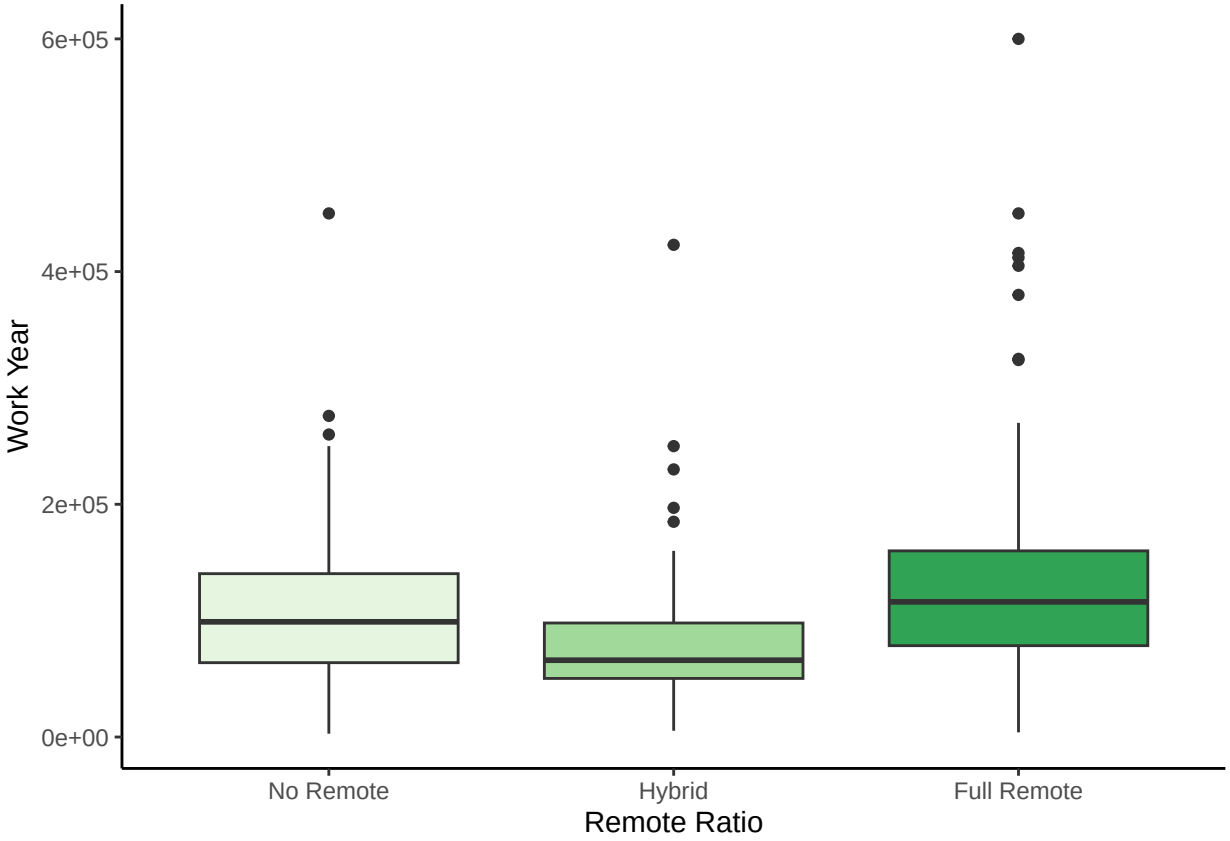
data_analysis

Wafiakmal Miftah

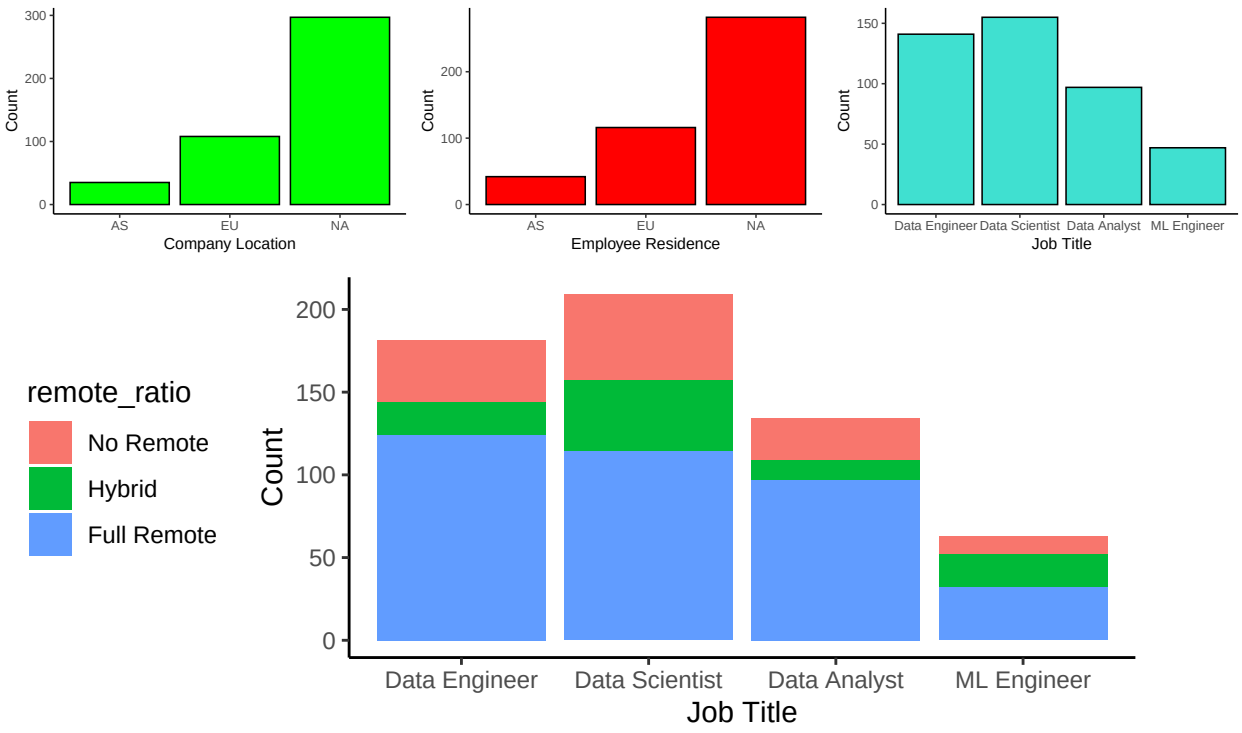
2022-11-28




```
## Warning in geom_histogram(stat = "count", color = "Black", fill = "Green"):
## Ignoring unknown parameters: ‘binwidth‘, ‘bins‘, and ‘pad‘

## Warning in geom_histogram(stat = "count", color = "Black", fill = "Red"):
## Ignoring unknown parameters: ‘binwidth‘, ‘bins‘, and ‘pad‘

## Warning in geom_histogram(stat = "count", color = "Black", fill = "turquoise"): Ignoring unknown para
## Ignoring unknown parameters: ‘binwidth‘, ‘bins‘, and ‘pad‘
```




```
## Warning in chisq.test(table(train$remote_ratio, train$experience_level), : Chi-
## squared approximation may be incorrect
```

```
## Warning in chisq.test(table(train$remote_ratio, train$employment_type), : Chi-
## squared approximation may be incorrect
```

```
## Warning in model.matrix.default(Terms, m, contrasts): the response appeared on
## the right-hand side and was dropped
```

```
## Warning in model.matrix.default(Terms, m, contrasts): problem with term 5 in
## model.matrix: no columns are assigned
```

Likelihood ratio tests of Multinomial Models

Response: remote_ratio Model 1 +employment_type + job_title + salary_in_usd + employee_residence + remote_ratio + company_location + company_size 2 experience_level + employment_type + job_title + salary_in_usd + employee_residence + company_location + company_size Resid. df Resid. Dev Test Df LR stat. Pr(Chi) 1 852 673.3576
2 846 671.1218 1 vs 2 6 2.235815 0.8967838

```
## Warning in model.matrix.default(Terms, m, contrasts): the response appeared on
## the right-hand side and was dropped
```

```
## Warning in model.matrix.default(Terms, m, contrasts): problem with term 5 in
## model.matrix: no columns are assigned
```

Likelihood ratio tests of Multinomial Models

y.level	term	estimate	std.error	statistic	p.value	conf.low	conf.high
Hybrid	(Intercept)	0.20	0	1.932173e+10	0.00	0.20	0.20
Hybrid	experience_levelEX	0.52	0	6.358290e+11	0.00	0.52	0.52
Hybrid	experience_levelMI	-0.51	0	-1.296392e+11	0.00	-0.51	-0.51
Hybrid	experience_levelSE	-0.31	0	-7.256449e+10	0.00	-0.31	-0.31
Hybrid	employment_typeFL	1.36	0	2.985698e+13	0.00	1.36	1.36
Hybrid	employment_typeFT	0.61	0	5.923650e+10	0.00	0.61	0.61
Hybrid	employment_typePT	16.34	0	2.602455e+14	0.00	16.34	16.34
Hybrid	job_titleData Scientist	0.37	0	7.291616e+10	0.00	0.37	0.37
Hybrid	job_titleData Analyst	0.14	0	1.025995e+11	0.00	0.14	0.14
Hybrid	job_titleML Engineer	1.19	0	6.445310e+11	0.00	1.19	1.19
Hybrid	salary_in_usd	0.00	0	-2.020000e+00	0.04	0.00	0.00
Hybrid	employee_residenceEU	-1.28	0	-3.299547e+11	0.00	-1.28	-1.28
Hybrid	employee_residenceNA	-1.93	0	-3.490826e+11	0.00	-1.93	-1.93
Hybrid	company_locationEU	2.00	0	5.232579e+11	0.00	2.00	2.00
Hybrid	company_locationNA	1.36	0	2.390618e+11	0.00	1.36	1.36
Hybrid	company_sizeM	-2.10	0	-7.895615e+11	0.00	-2.10	-2.10
Hybrid	company_sizeS	-1.11	0	-7.585687e+11	0.00	-1.11	-1.11
Full Remote	(Intercept)	14.05	0	2.767455e+12	0.00	14.05	14.05
Full Remote	experience_levelEX	0.28	0	8.163384e+11	0.00	0.28	0.28
Full Remote	experience_levelMI	-0.50	0	-3.327624e+11	0.00	-0.50	-0.50
Full Remote	experience_levelSE	-0.11	0	-3.170791e+10	0.00	-0.11	-0.11
Full Remote	employment_typeFL	-13.57	0	-8.685335e+14	0.00	-13.57	-13.57
Full Remote	employment_typeFT	-13.37	0	-2.632982e+12	0.00	-13.37	-13.37
Full Remote	employment_typePT	1.04	0	1.654802e+13	0.00	1.04	1.04
Full Remote	job_titleData Scientist	-0.48	0	-2.401768e+11	0.00	-0.48	-0.48
Full Remote	job_titleData Analyst	0.12	0	1.066898e+11	0.00	0.12	0.12
Full Remote	job_titleML Engineer	-0.06	0	-9.163913e+10	0.00	-0.06	-0.06
Full Remote	salary_in_usd	0.00	0	2.320000e+00	0.02	0.00	0.00
Full Remote	employee_residenceEU	-1.73	0	-1.378158e+12	0.00	-1.73	-1.73
Full Remote	employee_residenceNA	-2.22	0	-4.512828e+11	0.00	-2.22	-2.22
Full Remote	company_locationEU	2.10	0	1.652336e+12	0.00	2.10	2.10
Full Remote	company_locationNA	2.99	0	6.113973e+11	0.00	2.99	2.99
Full Remote	company_sizeM	-0.18	0	-3.523330e+10	0.00	-0.18	-0.18
Full Remote	company_sizeS	0.01	0	2.156095e+10	0.00	0.01	0.01

Response: remote_ratio Model 1 experience_level + job_title + salary_in_usd + employee_residence + remote_ratio + company_location + company_size 2 experience_level + employment_type + job_title + salary_in_usd + employee_residence + company_location + company_size Resid. df Resid. Dev Test Df LR stat. Pr(Chi) 1 852 682.0829
2 846 671.1218 1 vs 2 6 10.9611 0.08958627

```
## Warning in model.matrix.default(Terms, m, contrasts): the response appeared on
## the right-hand side and was dropped
```

```
## Warning in model.matrix.default(Terms, m, contrasts): problem with term 5 in
## model.matrix: no columns are assigned
```

Likelihood ratio tests of Multinomial Models

Response: remote_ratio Model 1 experience_level + employment_type + salary_in_usd + employee_residence + remote_ratio + company_location + company_size 2 experience_level + employment_type + job_title + salary_in_usd + employee_residence + company_location + company_size

y.level	term	estimate	std.error	statistic	p.value	conf.low	conf.high
Hybrid	(Intercept)	1.22	0	1.932173e+10	0.00	1.22	1.22
Hybrid	experience_levelEX	1.68	0	6.358290e+11	0.00	1.68	1.68
Hybrid	experience_levelMI	0.60	0	-1.296392e+11	0.00	0.60	0.60
Hybrid	experience_levelSE	0.73	0	-7.256449e+10	0.00	0.73	0.73
Hybrid	employment_typeFL	3.90	0	2.985698e+13	0.00	3.90	3.90
Hybrid	employment_typeFT	1.85	0	5.923650e+10	0.00	1.85	1.85
Hybrid	employment_typePT	12445653.93	0	2.602455e+14	0.00	12445653.93	12445653.93
Hybrid	job_titleData Scientist	1.45	0	7.291616e+10	0.00	1.45	1.45
Hybrid	job_titleData Analyst	1.15	0	1.025995e+11	0.00	1.15	1.15
Hybrid	job_titleML Engineer	3.28	0	6.445310e+11	0.00	3.28	3.28
Hybrid	salary_in_usd	1.00	0	-2.020000e+00	0.04	1.00	1.00
Hybrid	employee_residenceEU	0.28	0	-3.299547e+11	0.00	0.28	0.28
Hybrid	employee_residenceNA	0.15	0	-3.490826e+11	0.00	0.15	0.15
Hybrid	company_locationEU	7.36	0	5.232579e+11	0.00	7.36	7.36
Hybrid	company_locationNA	3.92	0	2.390618e+11	0.00	3.92	3.92
Hybrid	company_sizeM	0.12	0	-7.895615e+11	0.00	0.12	0.12
Hybrid	company_sizeS	0.33	0	-7.585687e+11	0.00	0.33	0.33
Full Remote	(Intercept)	1259191.55	0	2.767455e+12	0.00	1259191.55	1259191.55
Full Remote	experience_levelEX	1.32	0	8.163384e+11	0.00	1.32	1.32
Full Remote	experience_levelMI	0.61	0	-3.327624e+11	0.00	0.61	0.61
Full Remote	experience_levelSE	0.90	0	-3.170791e+10	0.00	0.90	0.90
Full Remote	employment_typeFL	0.00	0	-8.685335e+14	0.00	0.00	0.00
Full Remote	employment_typeFT	0.00	0	-2.632982e+12	0.00	0.00	0.00
Full Remote	employment_typePT	2.83	0	1.654802e+13	0.00	2.83	2.83
Full Remote	job_titleData Scientist	0.62	0	-2.401768e+11	0.00	0.62	0.62
Full Remote	job_titleData Analyst	1.13	0	1.066898e+11	0.00	1.13	1.13
Full Remote	job_titleML Engineer	0.94	0	-9.163913e+10	0.00	0.94	0.94
Full Remote	salary_in_usd	1.00	0	2.320000e+00	0.02	1.00	1.00
Full Remote	employee_residenceEU	0.18	0	-1.378158e+12	0.00	0.18	0.18
Full Remote	employee_residenceNA	0.11	0	-4.512828e+11	0.00	0.11	0.11
Full Remote	company_locationEU	8.14	0	1.652336e+12	0.00	8.14	8.14
Full Remote	company_locationNA	19.94	0	6.113973e+11	0.00	19.94	19.94
Full Remote	company_sizeM	0.84	0	-3.523330e+10	0.00	0.84	0.84
Full Remote	company_sizeS	1.01	0	2.156095e+10	0.00	1.01	1.01

Resid. df Resid. Dev Test Df LR stat. Pr(Chi) 1 852 685.9285
2 846 671.1218 1 vs 2 6 14.80672 0.02181445

```
## Warning in model.matrix.default(Terms, m, contrasts): the response appeared on
## the right-hand side and was dropped
```

```
## Warning in model.matrix.default(Terms, m, contrasts): problem with term 4 in
## model.matrix: no columns are assigned
```

Likelihood ratio tests of Multinomial Models

Response: remote_ratio Model 1 experience_level + job_title + employee_residence + remote_ratio + company_location + company_size 2 experience_level + employment_type + job_title + salary_in_usd + employee_residence + company_location + company_size Resid. df Resid. Dev Test Df LR stat. Pr(Chi)
1 854 685.2540
2 846 671.1218 1 vs 2 8 14.13218 0.0783847

Table 1: Regression Result

	Dependent variable:	
	Hybrid	Full Remote
	(1)	(2)
experience_levelEX	0.52*** (0.00)	0.28*** (0.00)
experience_levelMI	−0.51*** (0.00)	−0.50*** (0.00)
experience_levelSE	−0.31*** (0.00)	−0.11*** (0.00)
employment_typeFL	1.36*** (0.00)	−13.57*** (0.00)
employment_typeFT	0.61*** (0.00)	−13.37*** (0.00)
employment_typePT	16.34*** (0.00)	1.04*** (0.00)
job_titleData Scientist	0.37*** (0.00)	−0.48*** (0.00)
job_titleData Analyst	0.14*** (0.00)	0.12*** (0.00)
job_titleML Engineer	1.19*** (0.00)	−0.06*** (0.00)
salary_in_usd	−0.0000** (0.0000)	0.0000** (0.0000)
employee_residenceEU	−1.28*** (0.00)	−1.73*** (0.00)
employee_residenceNA	−1.93*** (0.00)	−2.22*** (0.00)
company_locationEU	2.00*** (0.00)	2.10*** (0.00)
company_locationNA	1.36*** (0.00)	2.99*** (0.00)
company_sizeM	−2.10*** (0.00)	−0.18*** (0.00)
company_sizeS	−1.11*** (0.00)	0.01*** (0.00)
Constant	0.20*** (0.00)	14.05*** (0.00)
Akaike Inf. Crit.	976.83	976.83
Note:		*p<0.1; **p<0.05; ***p<0.01

type	group	term	contrast	estimate	std.error	statistic	p.value	conf.low	c
probs	No Remote	job_title	Data Scientist - Data Engineer	0.06	0	53.55	0	0.05	
probs	No Remote	job_title	Data Analyst - Data Engineer	-0.02	0	-57.17	0	-0.02	
probs	No Remote	job_title	ML Engineer - Data Engineer	-0.03	0	-65.01	0	-0.03	
probs	Hybrid	job_title	Data Scientist - Data Engineer	0.07	0	64.89	0	0.07	
probs	Hybrid	job_title	Data Analyst - Data Engineer	0.00	0	68.81	0	0.00	
probs	Hybrid	job_title	ML Engineer - Data Engineer	0.13	0	62.81	0	0.13	
probs	Full Remote	job_title	Data Scientist - Data Engineer	-0.12	0	-59.28	0	-0.13	
probs	Full Remote	job_title	Data Analyst - Data Engineer	0.01	0	54.51	0	0.01	
probs	Full Remote	job_title	ML Engineer - Data Engineer	-0.11	0	-62.26	0	-0.11	

Items	Value
Accuracy	0.65
Kappa	0.20
AccuracyLower	0.61
AccuracyUpper	0.68
AccuracyNull	0.63
AccuracyPValue	0.16
McnemarPValue	0.00

```
## Warning in model.matrix.default(Terms, m, contrasts): the response appeared on
## the right-hand side and was dropped

## Warning in model.matrix.default(Terms, m, contrasts): problem with term 4 in
```

```
## model.matrix: no columns are assigned
```

Likelihood ratio tests of Multinomial Models

Response: remote_ratio Model 1 experience_level + job_title + salary_in_usd + remote_ratio + company_location + company_size 2 experience_level + employment_type + job_title + salary_in_usd + employee_residence + company_location + company_size Resid. df Resid. Dev Test Df LR stat. Pr(Chi)
1 856 686.9642
2 846 671.1218 1 vs 2 10 15.8424 0.1042336

```
## Warning in model.matrix.default(Terms, m, contrasts): the response appeared on
## the right-hand side and was dropped
```

```
## Warning in model.matrix.default(Terms, m, contrasts): problem with term 5 in
## model.matrix: no columns are assigned
```

Likelihood ratio tests of Multinomial Models

Response: remote_ratio Model 1 experience_level + job_title + salary_in_usd + employee_residence + remote_ratio + company_size 2 experience_level + employment_type + job_title + salary_in_usd + employee_residence + company_location + company_size Resid. df Resid. Dev Test Df LR stat. Pr(Chi)
1 856 690.3713
2 846 671.1218 1 vs 2 10 19.24956 0.03720502

```
## Warning in model.matrix.default(Terms, m, contrasts): the response appeared on
## the right-hand side and was dropped
```

```
## Warning in model.matrix.default(Terms, m, contrasts): problem with term 5 in
## model.matrix: no columns are assigned
```

Likelihood ratio tests of Multinomial Models

Response: remote_ratio Model 1 experience_level + job_title + salary_in_usd + employee_residence + remote_ratio + company_location 2 experience_level + employment_type + job_title + salary_in_usd + employee_residence + company_location + company_size Resid. df Resid. Dev Test Df LR stat. Pr(Chi)
1 856 719.2197
2 846 671.1218 1 vs 2 10 48.09794 5.956273e-07

```
## Likelihood ratio tests of Multinomial Models
##
## Response: remote_ratio
##
## 1                experience_level + employment_type + job_title + salary_in_usd + employee_residence
## 2 experience_level + employment_type + job_title + salary_in_usd + employee_residence + company_loca
##   Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
## 1       850   712.4610
## 2       846   671.1218 1 vs 2     4 41.33917 2.286459e-08
```

```
## Call:
## multinom(formula = remote_ratio ~ experience_level + employment_type +
##     job_title + salary_in_usd + employee_residence + company_location,
##     data = train, trace = FALSE)
##
```

```
## Coefficients:
##             (Intercept) experience_levelEX experience_levelMI
## Hybrid       -0.9718764          0.6077138         -0.5616949
## Full Remote  13.3927538          0.1848310         -0.2951606
##             experience_levelSE employment_typeFL employment_typeFT
## Hybrid              -0.6607227         0.8236287         0.5484802
## Full Remote         -0.2278289       -13.1215378       -12.8737933
##             employment_typePT job_titleData Scientist job_titleData Analyst
## Hybrid             15.4915767               0.8399116             0.6943165
## Full Remote        -0.3722455              -0.1616398             0.4658540
##             job_titleML Engineer salary_in_usd employee_residenceEU
## Hybrid                 1.6591688 -2.424504e-06           0.4541506
## Full Remote           -0.1101084  1.141724e-06          -1.9408026
##             employee_residenceNA company_locationEU company_locationNA
## Hybrid                -0.3147699          0.2119898         -0.1799808
## Full Remote           -2.0402946          2.0324016          2.8073034
##
## Std. Errors:
##              (Intercept) experience_levelEX experience_levelMI
## Hybrid      1.123412e-11       1.145038e-12       3.639429e-12
## Full Remote 5.313338e-12       4.649171e-13       1.396824e-12
##             experience_levelSE employment_typeFL employment_typeFT
## Hybrid            4.873124e-12      6.412531e-14      1.115786e-11
## Full Remote       3.678284e-12      2.166571e-14      5.295621e-12
##             employment_typePT job_titleData Scientist job_titleData Analyst
## Hybrid          1.219105e-14            5.319803e-12          1.451573e-12
## Full Remote     1.219102e-14            2.019548e-12          8.143862e-13
##             job_titleML Engineer salary_in_usd employee_residenceEU
## Hybrid              2.203639e-12  1.462085e-06         4.316562e-12
## Full Remote         7.841697e-13  9.022475e-07         1.377671e-12
##             employee_residenceNA company_locationEU company_locationNA
## Hybrid              6.076296e-12       4.156218e-12       6.193690e-12
## Full Remote         5.058153e-12       1.338059e-12       5.077995e-12
##
## Residual Deviance: 712.461
## AIC: 772.461

## Likelihood ratio tests of Multinomial Models
##
## Response: remote_ratio
##
## 1                                                                                job_title + company_loca
## 2 experience_level + employment_type + job_title + salary_in_usd + employee_residence + company_loca
##   Resid. df Resid. Dev   Test    Df LR stat.   Pr(Chi)
## 1       864   693.6361
## 2       846   671.1218 1 vs 2    18  22.5143 0.2099499

## Call:
## multinom(formula = remote_ratio ~ job_title + company_location +
##     company_size, data = train, trace = FALSE)
##
## Coefficients:
##             (Intercept) job_titleData Scientist job_titleData Analyst
## Hybrid       0.06413182               0.5481136             0.7473942
```

```
## Full Remote  0.31661496              -0.1894828             0.4481260
##             job_titleML Engineer company_locationEU company_locationNA
## Hybrid                 1.3390607          0.7554523         -0.7168916
## Full Remote           -0.1529826          0.2262414          0.9864969
##             company_sizeM company_sizeS
## Hybrid        -2.03168063    -0.4632989
## Full Remote   -0.09551458     0.5419590
##
## Std. Errors:
##             (Intercept) job_titleData Scientist job_titleData Analyst
## Hybrid        0.5980496               0.4399003             0.5636573
## Full Remote   0.4925914               0.2892217             0.3529384
##             job_titleML Engineer company_locationEU company_locationNA
## Hybrid                 0.5747712          0.5510674          0.5419637
## Full Remote            0.4587180          0.4899853          0.4491222
##             company_sizeM company_sizeS
## Hybrid          0.4062267     0.5438868
## Full Remote     0.2818427     0.4664760
##
## Residual Deviance: 693.6361
## AIC: 725.6361
```

```
##              Reference
## Prediction    No Remote Hybrid Full Remote
##   No Remote           0      0           0
##   Hybrid             10     36          22
##   Full Remote        84     35         253
```

```
##  Accuracy
## 0.6568182
```

```
##                    Sensitivity Specificity
## Class: No Remote     0.0000000   1.0000000
## Class: Hybrid        0.5070423   0.9132791
## Class: Full Remote   0.9200000   0.2787879
```

```
## Setting levels: control = FALSE, case = TRUE
```

```
## Setting direction: controls < cases
```

```
##
## Call:
## roc.default(response = (train$remote_ratio == "No Remote"), predictor = predprobs[,     1], percent =
##
## Data: predprobs[, 1] in 346 controls ((train$remote_ratio == "No Remote") FALSE) < 94 cases ((train$r
## Area under the curve: 62.97%
```

```
## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases
```

```
##
## Call:
```

```
## roc.default(response = (train$remote_ratio == "Hybrid"), predictor = predprobs[,     2], percent = T
##
## Data: predprobs[, 2] in 369 controls ((train$remote_ratio == "Hybrid") FALSE) < 71 cases ((train$remo
## Area under the curve: 87.24%

## Setting levels: control = FALSE, case = TRUE
## Setting direction: controls < cases

##
## Call:
## roc.default(response = (train$remote_ratio == "Full Remote"),     predictor = predprobs[, 3], percen
##
## Data: predprobs[, 3] in 165 controls ((train$remote_ratio == "Full Remote") FALSE) < 275 cases ((tra
## Area under the curve: 71.99%
```