R Notebook

```r
df <- readRDS("../Data/MentalHealthCLEAN.rds")
```

I am going to drop state from the dataset as moore than 50& of the data is not populated and commetns becasue it is a raw text field

```r
df = df[,!(names(df) %in% c("state", "comments", "Timestamp"))]
```

```r
model1 <- glm(treatment ~. , data =df, family=binomial )
summary(model1)
```

```
##
## Call:
## glm(formula = treatment ~ ., family = binomial, data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7714  -0.3396   0.0002   0.5686   3.2066
##
## Coefficients: (4 not defined because of singularities)
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -3.09771    1.04464  -2.965  0.00302 **
## Age                            0.02469    0.01319   1.873  0.06114 .
## Genderfemale_trans            -0.78443    1.18646  -0.661  0.50851
## Genderfluid                   -1.33262    1.30302  -1.023  0.30644
## Gendergenderqueer              0.86327    1.22662   0.704  0.48157
## Gendermale_cis                -0.72309    0.24765  -2.920  0.00350 **
## Genderunknown                 14.92775 3956.18038   0.004  0.99699
## CountryAustria               -16.92006 1686.48454  -0.010  0.99200
## CountryBelgium                -1.88575    1.62581  -1.160  0.24610
## CountryBosnia and Herzegovina -18.57211 3956.18045  -0.005  0.99625
## CountryBrazil                  0.09070    1.43829   0.063  0.94972
## CountryBulgaria                3.39288    2.77564   1.222  0.22156
## CountryCanada                  0.06110    0.74220   0.082  0.93439
## CountryChina                 -16.99848 3956.18042  -0.004  0.99657
## CountryColombia              -17.63015 2787.57956  -0.006  0.99495
## CountryCosta Rica            -13.69869 3956.18045  -0.003  0.99724
## CountryCroatia                15.66563 2693.69750   0.006  0.99536
## CountryCzech Republic        -19.61162 3956.18042  -0.005  0.99604
## CountryDenmark                16.39814 2797.44210   0.006  0.99532
## CountryFinland                 0.70654    2.27197   0.311  0.75582
## CountryFrance                  0.58902    1.43364   0.411  0.68118
## CountryGeorgia               -17.30065 3956.18044  -0.004  0.99651
## CountryGermany                 0.13147    0.78832   0.167  0.86754
## CountryGreece                -12.33523 2796.35019  -0.004  0.99648
## CountryHungary               -18.15358 3956.18043  -0.005  0.99634
## CountryIndia                   1.01801    1.32441   0.769  0.44210
## CountryIreland                 0.31307    0.82631   0.379  0.70478
## CountryIsrael                -16.44156 1470.75487  -0.011  0.99108
## CountryItaly                  -1.29926    1.37548  -0.945  0.34487
## CountryJapan                  16.01293 3956.18044   0.004  0.99677
## CountryLatvia                -12.58501 3956.18045  -0.003  0.99746
## CountryMexico                  1.66411    2.40986   0.691  0.48985
```

```
## CountryMoldova                   13.71463 3956.18047   0.003  0.99723
## CountryNetherlands               -0.38902    0.86216  -0.451  0.65184
## CountryNew Zealand               -0.06124    1.12583  -0.054  0.95662
## CountryNigeria                  -12.89073 3956.18044  -0.003  0.99740
## CountryNorway                   -14.68685 3956.18045  -0.004  0.99704
## CountryPhilippines              -17.52557 3956.18043  -0.004  0.99647
## CountryPoland                     0.16648    1.12600   0.148  0.88246
## CountryPortugal                 -13.93438 2638.90236  -0.005  0.99579
## CountryRomania                  -11.98156 3956.18046  -0.003  0.99758
## CountryRussia                   -16.10508 2043.24302  -0.008  0.99371
## CountrySingapore                 -1.03925    1.42275  -0.730  0.46512
## CountrySlovenia                  18.30652 3956.18043   0.005  0.99631
## CountrySouth Africa               1.67653    1.86628   0.898  0.36901
## CountrySpain                    -13.54726 3956.18046  -0.003  0.99727
## CountrySweden                    -1.67983    1.31209  -1.280  0.20045
## CountrySwitzerland                1.09826    1.42672   0.770  0.44143
## CountryThailand                 -14.58199 3956.18043  -0.004  0.99706
## CountryUnited Kingdom             0.69497    0.67866   1.024  0.30582
## CountryUnited States              0.13292    0.65596   0.203  0.83942
## CountryUruguay                  -12.41302 3956.18046  -0.003  0.99750
## self_employedYes                 -0.26467    0.35074  -0.755  0.45048
## family_historyYes                 0.99716    0.19089   5.224 1.75e-07 ***
## work_interfereNo Response         -2.48064    0.56810  -4.367 1.26e-05 ***
## work_interfereOften               3.62867    0.37227   9.747  < 2e-16 ***
## work_interfereRarely              2.55924    0.30435   8.409  < 2e-16 ***
## work_interfereSometimes           2.95602    0.27060  10.924  < 2e-16 ***
## no_employees100-500               0.28337    0.40880   0.693  0.48820
## no_employees26-100                0.33895    0.37193   0.911  0.36213
## no_employees500-1000              0.22839    0.56037   0.408  0.68359
## no_employees6-25                  0.14977    0.34898   0.429  0.66780
## no_employeesMore than 1000       -0.03769    0.41006  -0.092  0.92676
## remote_workYes                   -0.19293    0.21615  -0.893  0.37208
## tech_companyYes                  -0.12374    0.25339  -0.488  0.62533
## benefitsNo                        0.23253    0.28817   0.807  0.41971
## benefitsYes                       0.76529    0.28251   2.709  0.00675 **
## care_optionsNot sure             -0.16421    0.24652  -0.666  0.50535
## care_optionsYes                   0.76389    0.25339   3.015  0.00257 **
## wellness_programNo               -0.15946    0.31955  -0.499  0.61777
## wellness_programYes              -0.50279    0.37525  -1.340  0.18029
## seek_helpNo                      -0.64607    0.26770  -2.413  0.01580 *
## seek_helpYes                     -0.81009    0.32871  -2.464  0.01372 *
## anonymityNo                      -0.10367    0.42665  -0.243  0.80802
## anonymityYes                      0.54541    0.24527   2.224  0.02617 *
## leaveSomewhat difficult           0.47296    0.32989   1.434  0.15166
## leaveSomewhat easy               -0.26829    0.24303  -1.104  0.26962
## leaveVery difficult               0.30820    0.37080   0.831  0.40586
## leaveVery easy                    0.13042    0.29818   0.437  0.66184
## mental_health_consequenceNo      -0.08527    0.25886  -0.329  0.74185
## mental_health_consequenceYes     -0.19693    0.26549  -0.742  0.45824
## phys_health_consequenceNo         0.07394    0.24574   0.301  0.76349
## phys_health_consequenceYes        0.01874    0.47989   0.039  0.96885
## coworkersSome of them             0.39425    0.25792   1.529  0.12638
## coworkersYes                      1.06341    0.37818   2.812  0.00492 **
## supervisorSome of them           -0.41463    0.25564  -1.622  0.10482
```

```
## supervisorYes                       -0.32886      0.29640   -1.110  0.26721
## mental_health_interviewNo            0.33856      0.30822    1.098  0.27202
## mental_health_interviewYes           1.00992      0.67652    1.493  0.13548
## phys_health_interviewNo              0.14836      0.21480    0.691  0.48975
## phys_health_interviewYes             0.40874      0.30438    1.343  0.17932
## mental_vs_physicalNo                -0.12645      0.24205   -0.522  0.60139
## mental_vs_physicalYes               -0.06918      0.25711   -0.269  0.78787
## obs_consequenceYes                   0.28943      0.28374    1.020  0.30771
## continentAmericas                         NA           NA       NA       NA
## continentAsia                             NA           NA       NA       NA
## continentEurope                           NA           NA       NA       NA
## continentOceania                          NA           NA       NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1709.16  on 1232  degrees of freedom
## Residual deviance:  832.88  on 1139  degrees of freedom
## AIC: 1020.9
##
## Number of Fisher Scoring iterations: 16
```

```r
model2 <- glm(treatment~ Age+Gender+ family_history+ work_interfere+family_history+benefits+care_options
summary(model2)
```

```
##
## Call:
## glm(formula = treatment ~ Age + Gender + family_history + work_interfere +
##     family_history + benefits + care_options + anonymity + coworkers,
##     family = binomial, data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4423  -0.3526   0.2089   0.6116   3.2290
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -2.86067    0.51261  -5.581 2.40e-08 ***
## Age                        0.02308    0.01142   2.021 0.043305 *
## Genderfemale_trans        -0.51100    1.16244  -0.440 0.660233
## Genderfluid               -0.68534    1.23932  -0.553 0.580264
## Gendergenderqueer          0.65882    1.10910   0.594 0.552502
## Gendermale_cis            -0.75196    0.22689  -3.314 0.000919 ***
## Genderunknown             10.71894  535.41126   0.020 0.984027
## family_historyYes          0.98156    0.17544   5.595 2.21e-08 ***
## work_interfereNo Response -2.35093    0.55073  -4.269 1.97e-05 ***
## work_interfereOften        3.38019    0.33485  10.095  < 2e-16 ***
## work_interfereRarely       2.44761    0.28372   8.627  < 2e-16 ***
## work_interfereSometimes    2.82398    0.24920  11.332  < 2e-16 ***
## benefitsNo                -0.06002    0.23089  -0.260 0.794915
## benefitsYes                0.48226    0.24500   1.968 0.049027 *
## care_optionsNot sure      -0.15089    0.22230  -0.679 0.497299
## care_optionsYes            0.53490    0.22423   2.385 0.017056 *
## anonymityNo               -0.02996    0.36986  -0.081 0.935449
```

```
## anonymityYes                 0.30273    0.21200   1.428 0.153297
## coworkersSome of them         0.21571    0.20477   1.053 0.292147
## coworkersYes                  0.84316    0.28387   2.970 0.002975 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1709.16  on 1232  degrees of freedom
## Residual deviance:  913.07  on 1213  degrees of freedom
## AIC: 953.07
##
## Number of Fisher Scoring iterations: 12
```

```
model3 <- glm(treatment~ work_interfere + benefits+ care_options+  anonymity , data =df, family=binomial
summary(model3)
```
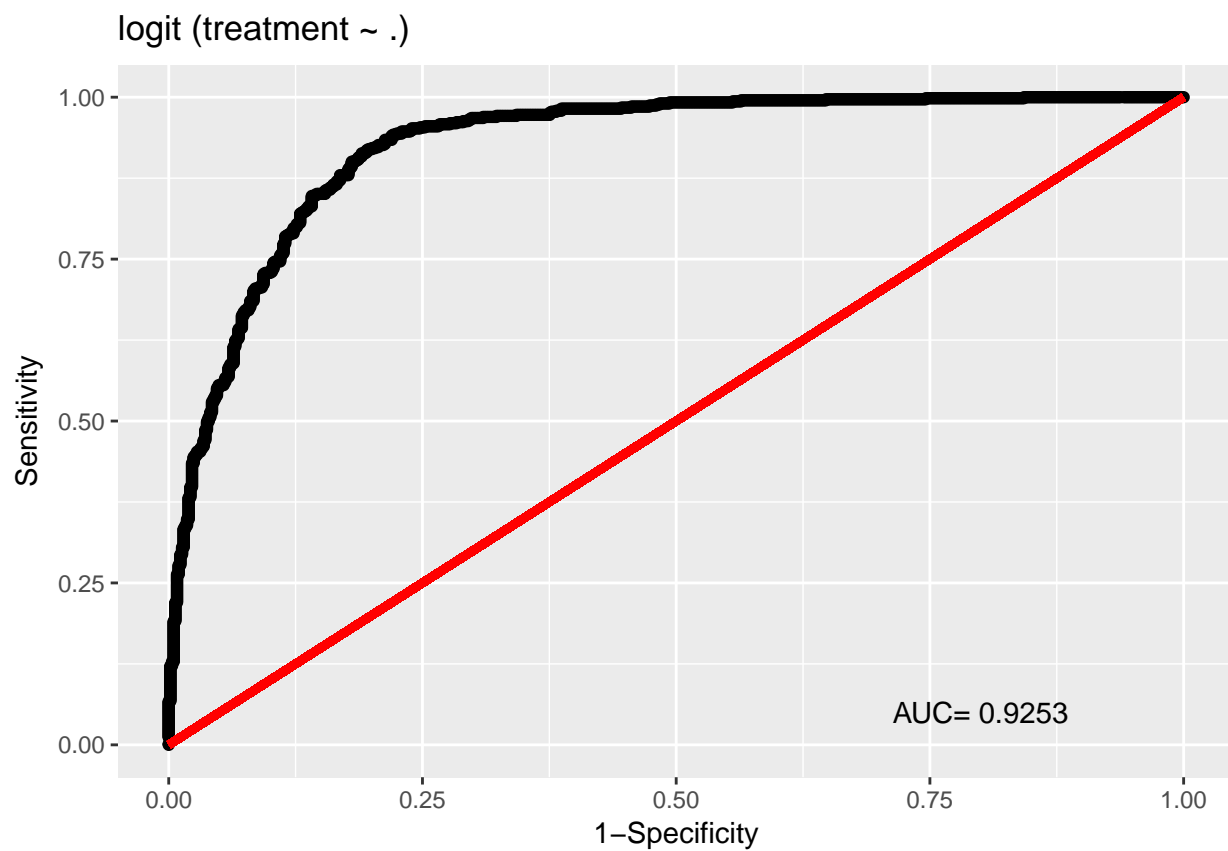
```
##
## Call:
## glm(formula = treatment ~ work_interfere + benefits + care_options +
##     anonymity, family = binomial, data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2774  -0.4251   0.3339   0.6861   3.0960
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -2.2549     0.2625  -8.592  < 2e-16 ***
## work_interfereNo Response -2.3615     0.5453  -4.331 1.49e-05 ***
## work_interfereOften        3.5817     0.3206  11.173  < 2e-16 ***
## work_interfereRarely       2.6143     0.2705   9.665  < 2e-16 ***
## work_interfereSometimes    3.0312     0.2372  12.779  < 2e-16 ***
## benefitsNo                -0.1679     0.2202  -0.763  0.44565
## benefitsYes                0.6218     0.2323   2.677  0.00743 **
## care_optionsNot sure      -0.1033     0.2127  -0.486  0.62703
## care_optionsYes            0.5672     0.2156   2.631  0.00852 **
## anonymityNo               -0.0199     0.3481  -0.057  0.95440
## anonymityYes               0.3432     0.2028   1.692  0.09056 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1709.16  on 1232  degrees of freedom
## Residual deviance:  974.55  on 1222  degrees of freedom
## AIC: 996.55
##
## Number of Fisher Scoring iterations: 6
```

```
library(Deducer)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: JGR
```
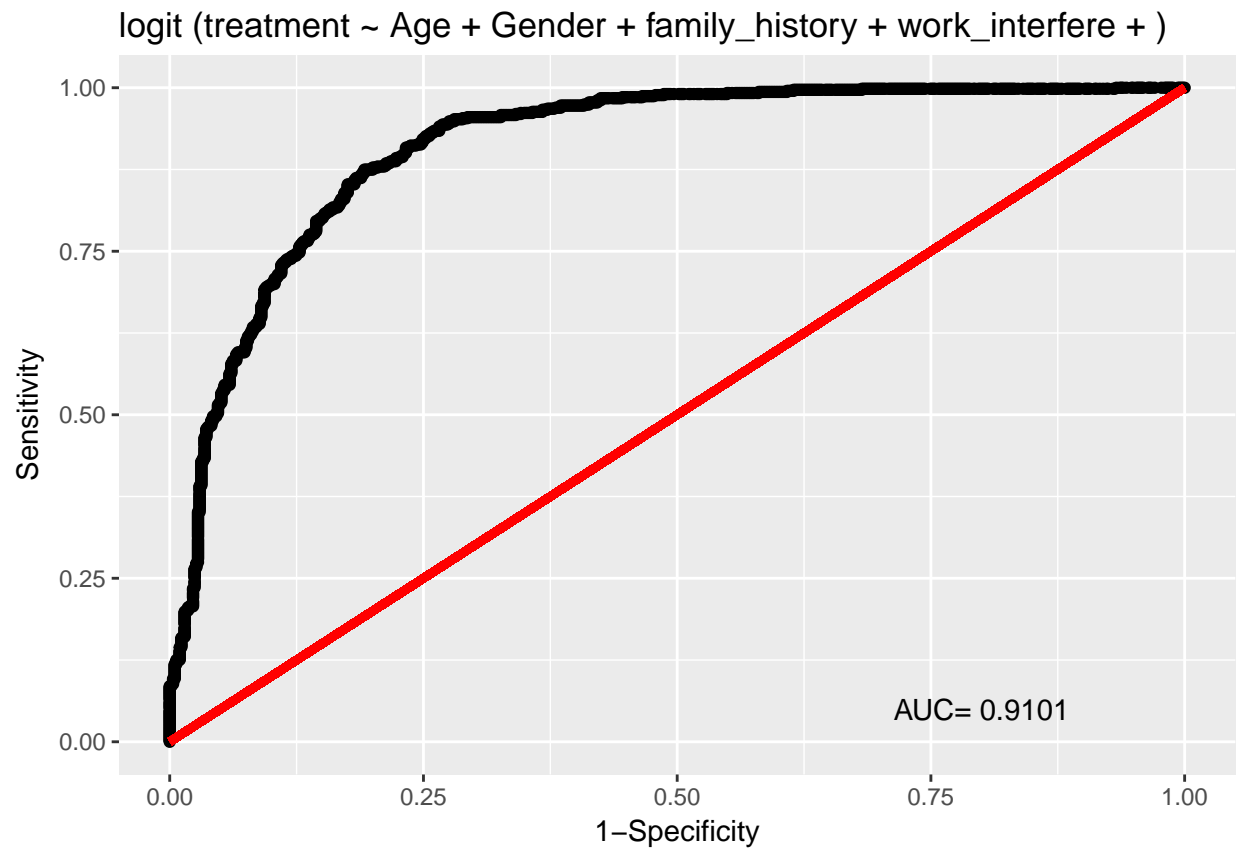
```
## Loading required package: rJava

## Loading required package: JavaGD

##
## Please type JGR() to launch console. Platform specific launchers (.exe and .app) can also be obtained

## Loading required package: car

## Loading required package: MASS

##
##
## Note Non-JGR console detected:
##  Deducer is best used from within JGR (http://jgr.markushelbig.org/).
##  To Bring up GUI dialogs, type deducer().
```

```
rocplot(model1)
```

logit (treatment ~ .)


```
rocplot(model2)
```

logit (treatment ~ Age + Gender + family_history + work_interfere +)

AUC= 0.9101

```
rocplot(model3)
```

logit (treatment ~ work_interfere + benefits + care_options + anonymity)

AUC= 0.8913