Booklike Summary of E-Waste Work

Rayhan Rashed

Contents

Chapter 1

Do People Miss Their Device??

```
backward.model <- step(full.model,direction = "backward",trace = 0)

backward.model %>% summary()
```

```
##
## Call:
## glm(formula = miss_dev ~ dump_within + did_with_device_econ +
##     memory_dev + dump_reason_break + dump_reason_theft + dump_reason_slow,
##     family = binomial, data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9330  -0.9671  -0.5300   0.9400   2.0160
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -1.77365    0.36676  -4.836 1.32e-06 ***
## dump_within1          -0.48466    0.27502  -1.762  0.07802 .
## dump_within2          -0.27973    0.12457  -2.246  0.02473 *
## dump_within3          -0.22649    0.07637  -2.966  0.00302 **
## dump_within4          -0.09543    0.06711  -1.422  0.15502
## dump_within5           0.08528    0.09866   0.864  0.38735
## did_with_device_econ1  0.55128    0.32105   1.717  0.08596 .
## memory_dev1            1.45145    0.24657   5.887 3.94e-09 ***
## dump_reason_break1     1.02806    0.25616   4.013 5.99e-05 ***
## dump_reason_theft1     1.11285    0.47485   2.344  0.01910 *
## dump_reason_slow1      0.90000    0.38924   2.312  0.02077 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 485.33  on 350  degrees of freedom
## Residual deviance: 415.57  on 340  degrees of freedom
## AIC: 437.57
##
## Number of Fisher Scoring iterations: 4
```

This is the fitted model description to investigate what are the factors that predict whether people miss their discarded devices. In the electronic waste field, this can play a large part due to the fact that, 50% of our participant responded that they keep the devices in their home, because they feel connected with the devices, miss them, and seem them to be valuable.

The AIC Value was Initially 477. ***After Stepwise Regression, it came down to 437***

The 10 fold cross validation accuracy is below:

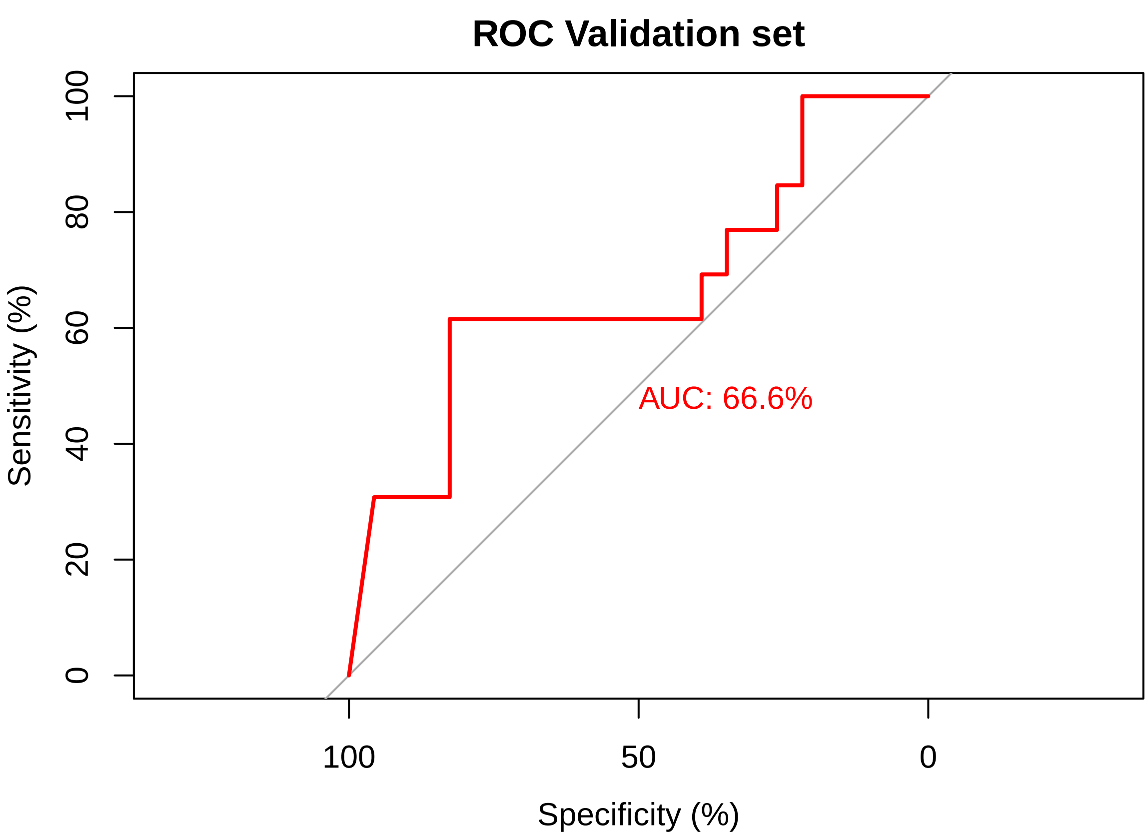
```
suppressWarnings( cross_validated_model <- boot::cv.glm(df,glmfit = backward.model,K = 10))
print((1-cross_validated_model$delta[1])*100)
```

```
## [1] 77.8717
```

Lets also look at the ROC curve for the the fitted logistic regression model: ***(A better fit than model without interaction)***

```
null.model <- glm(miss_dev~1,family = binomial(),data = df)
suppressWarnings( Generate_RoC(full_model = full.model,null_model = null.model,df = df,direction = "B",outc
```

```
## [1] 351
```



With a good fit in the model, lets diagnosis our model parameters for multicolineariy. If the ***VIF*** (Variance Inflation Factor) is > 10 for any predictor, we might be in trouble.

```
cat("MIN IVF: ",min(vif(backward.model)))
```

```
## MIN IVF:  1
```

```
cat("MAX IVF: ",max(vif(backward.model)))
```

```
## MAX IVF:  5
```

```
cat("MEAN IVF: ",mean(vif(backward.model)))
```

```
## MEAN IVF:  1.278175
```

Values > 0 means odd > 1 So, that accounts more for YES than NO.

Final Fitted Model:

Description of Variables:

dump_within - helmert coded. So, the coefficients curate (level[k] - avg of levels upto[1..k]). So, we can get an estimate how each duration is important. *dump_within1* means average feeling of missing devices during first month of discarding device. Similarly *dump_within2* means average feeling of missing devices during 1-6 of discarding device. *dump_within3* means average feeling of missing devices during 6 months - 1 year of discarding device. As we see, log odds keeps increasing *(-.48 < -0.28 < -0.23)* during this period. Which means, people tend to miss devices more and more as time passes. [The result becomes more and more significant during this period.] But, this doesnt go unbound. No significant effect for feeling miss_device for time beyond that [This corelates with our guts. That people doesnt feel that much bad after a certain period of time.] But, this is important to note that, not significant, but still implicative that the log-odd keeps increasing as time passes by. Which means, the more time passes, the more people miss their devices.

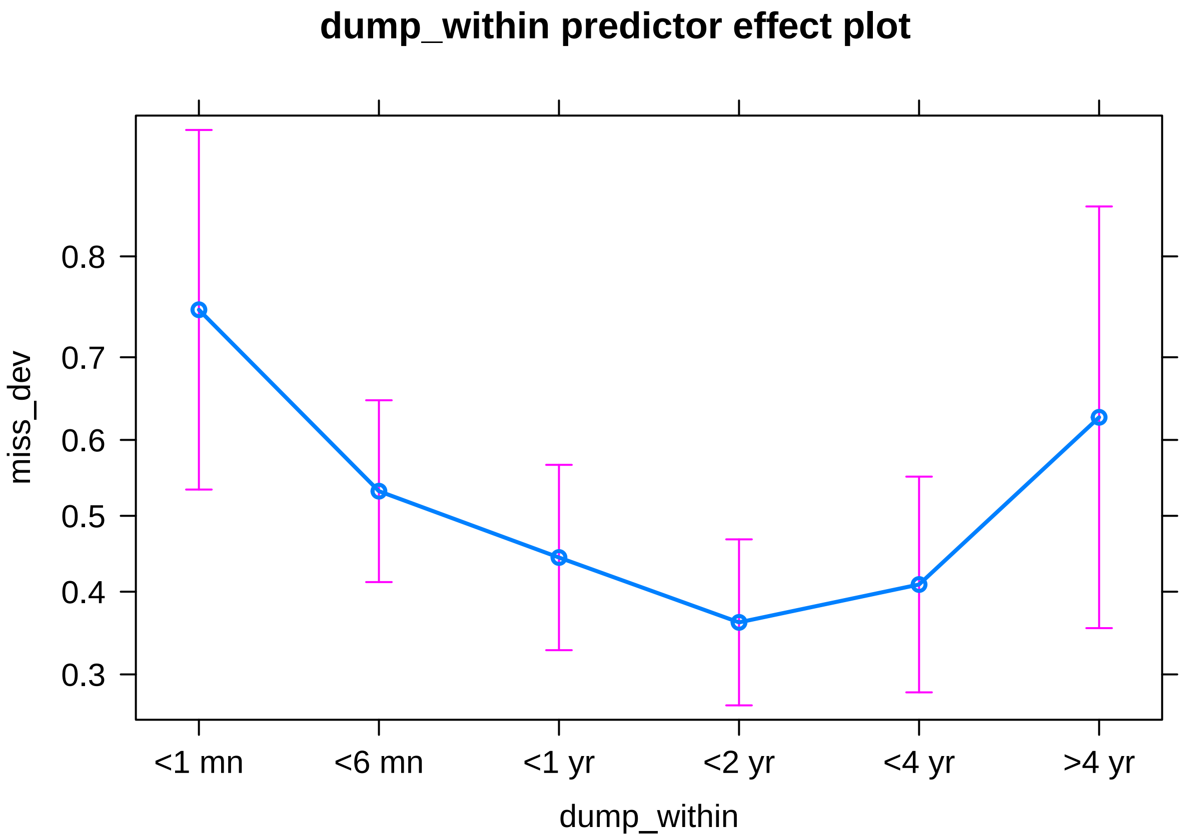
did_with_device_Y - whether the device is (kept home/dustbin) vs (sold/parts sold/sold to recycler). So, whether any economic or non-economic activity. Simple binary variable. With $p<0.1$ predicts that if economic activity was done, then device is *43%* less likely to be missed later (log odd -0.55)

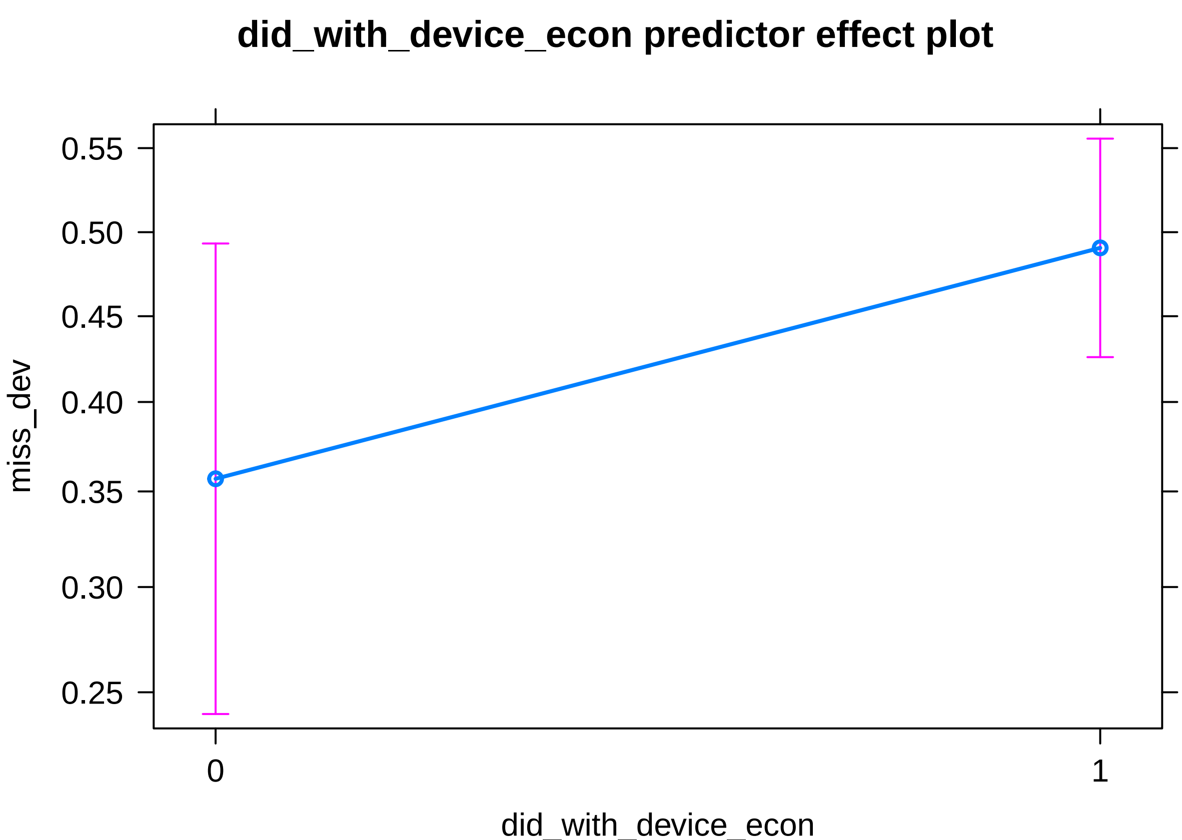
memory_with_device - whether the participant could write a memory with the device s/he used. This was a qualitative field. Ability to writing a memory with the device increases the odd of missing device * \math{326% (log odd 1.45)}*.

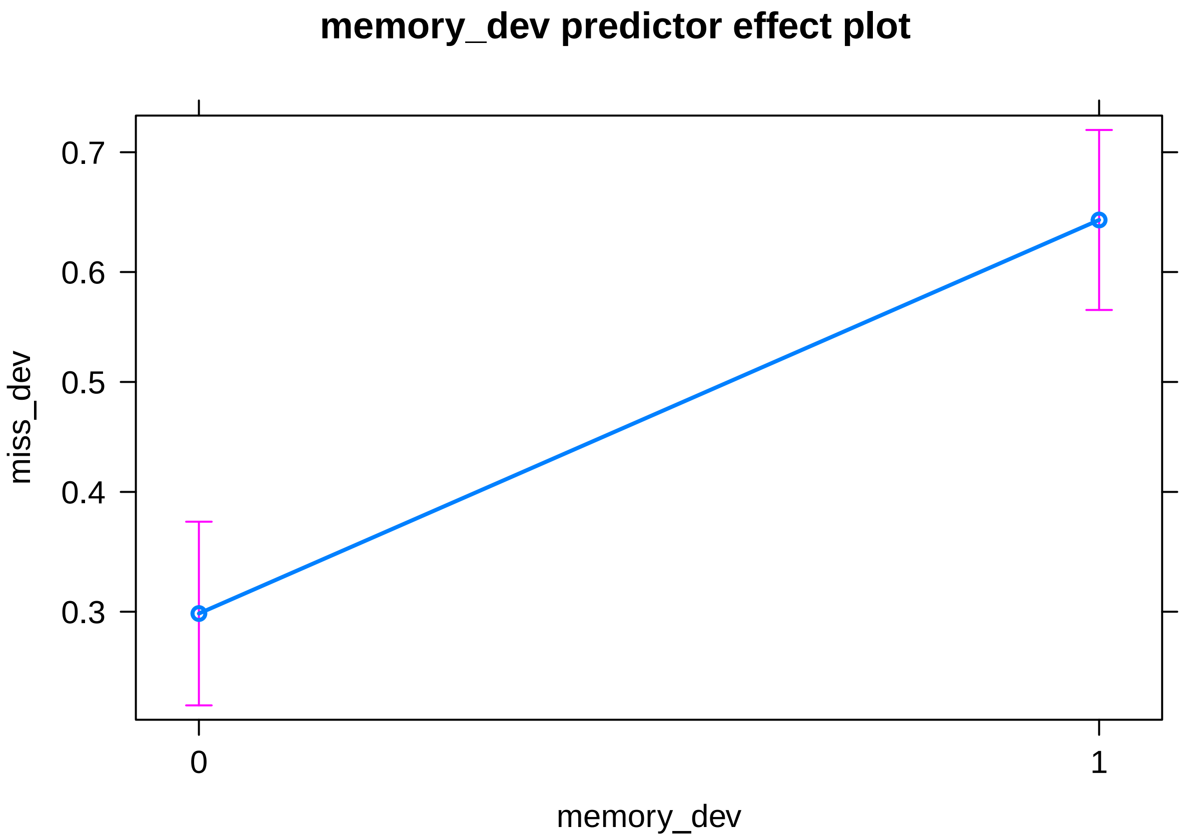
dump_reason_X/Y/Z - reason (theft/break/slow) and whether the device is being missed. The order is important. Theft indicates, the device is still probably being used (just in the state the user was using that). This has the highest log-odd among reasons *(1.1)*. Then comes device broken unfortunately/somehow. Although, the device is not usable, the device is unusable [suddenly from a usable state]. This has a lower odd ration than ***theft***. Which can be explained because the utility was diminished not by some random thief, but the owner. So, although, odd ratio is *>1*, it is not as much as theft. Finally, when the device has grown slow/unusable, the device is still being missed, *significantly*, but the odd ration is the lowest in the lot.

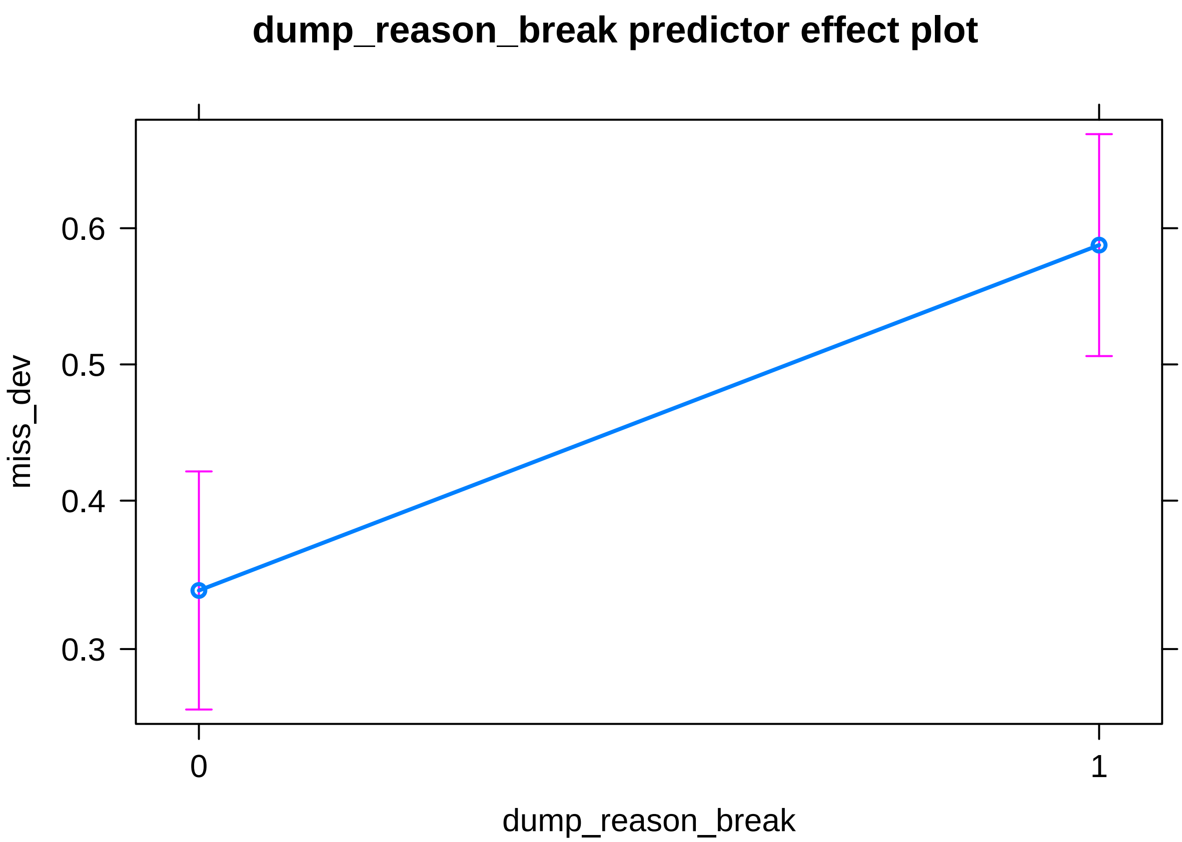
Lets look at the effects plot to better understand the individual effects of each predictor:

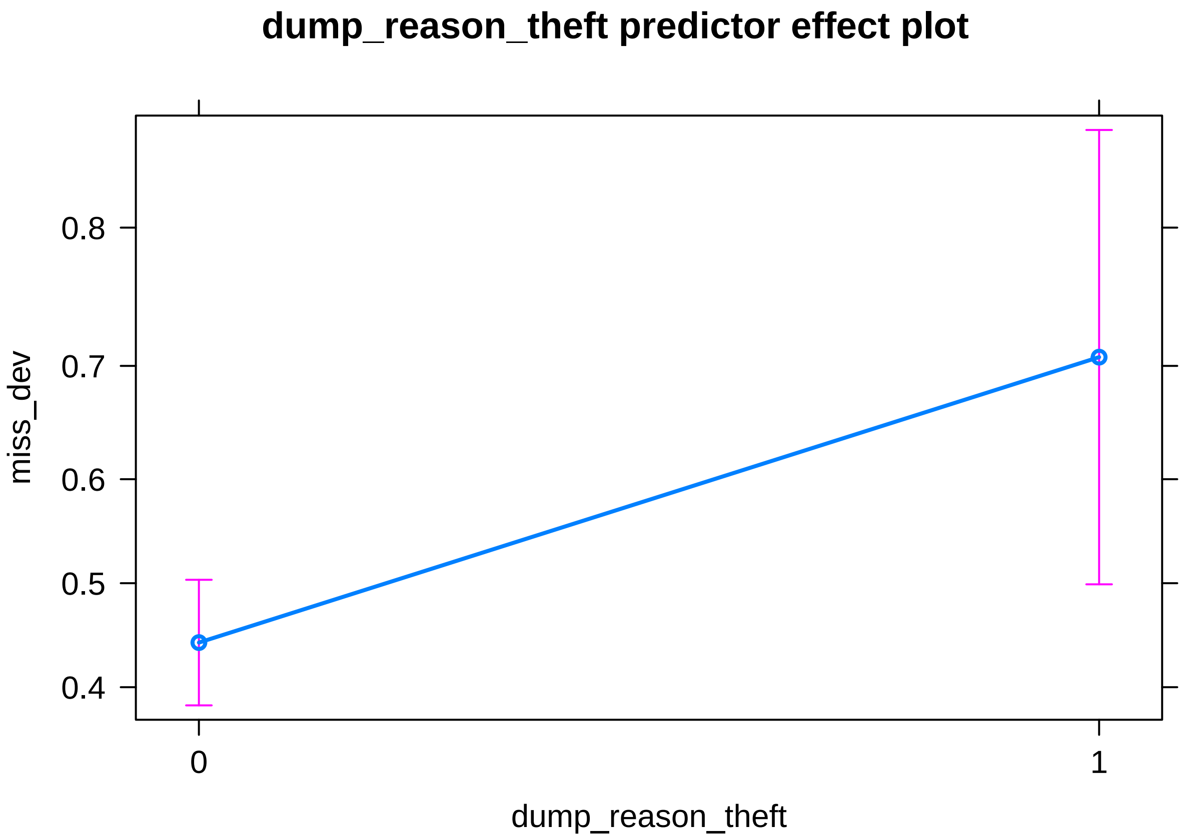
```
for (n in c(
  "dump_within"
  ,"did_with_device_econ"
  ,"memory_dev"
  ,"dump_reason_break"
  ,"dump_reason_theft"
  ,"dump_reason_slow"
)){
  #print(n)
  print(plot(effects::predictorEffect(n,backward.model)))
}
```


dump_within predictor effect plot
miss_dev
0.3
0.4
0.5
0.6
0.7
0.8
<1 mn
<6 mn
<1 yr
<2 yr
<4 yr
>4 yr
dump_within


did_with_device_econ predictor effect plot
miss_dev
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0
1
did_with_device_econ


memory_dev predictor effect plot
0.7
0.6
0.5
0.4
0.3
miss_dev
0
1
memory_dev


dump_reason_break predictor effect plot
0.6
0.5
0.4
0.3
miss_dev
0
1
dump_reason_break



Initial Regression Before Running Backward Stepwise Method:

```
glm(formula = miss_dev ~ gender + age +
      division + edu + occupation +
      device_count_5_yr + last_dumped_device +
      dump_within + did_with_device_econ +
      did_with_data_Y + memory_dev +
      miss_another + dump_reason_break +
      dump_reason_old + dump_reason_new +
      dump_reason_theft + dump_reason_slow +
      dump_reason_lag + rprd_usage_chlng_No +
      rprd_usage_chlng_fault +
      rprd_usage_chlng_reluc,
    family = binomial, data = df)
```

Chapter 2

Can We Dig Further?? [Possibly No]

We have introduced All Pairwise Interactions Here *from the best fitted model of Chapter 1.* Lets see how many survives upon the finishing of ***Backward Stepwise Regression Model***

```
suppressWarnings( backward.model <- step(full.model,
direction = "backward",trace = 0))

backward.model %>% summary()
```

```
##
## Call:
## glm(formula = miss_dev ~ dump_within + did_with_device_econ +
##     memory_dev + dump_reason_break + dump_reason_theft + dump_reason_slow +
##     dump_within:did_with_device_econ + did_with_device_econ:dump_reason_break +
##     did_with_device_econ:dump_reason_theft + memory_dev:dump_reason_slow +
##     dump_reason_break:dump_reason_theft, family = binomial, data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7734  -0.9236  -0.2387   0.8752   2.3921
##
## Coefficients:
##                                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)                             -5.93029  144.96528  -0.041 0.967369
## dump_within1                            -1.12301    0.61015  -1.841 0.065687
## dump_within2                            -0.05492    0.31648  -0.174 0.862238
## dump_within3                            -0.80549    0.25775  -3.125 0.001778
## dump_within4                            -0.52595    0.23603  -2.228 0.025861
## dump_within5                            -2.86470  144.96356  -0.020 0.984234
## did_with_device_econ1                    4.76928  144.96535   0.033 0.973755
## memory_dev1                              1.41187    0.26887   5.251 1.51e-07
## dump_reason_break1                       2.67847    0.88178   3.038 0.002385
## dump_reason_theft1                       4.25635    1.21115   3.514 0.000441
## dump_reason_slow1                        0.00594    0.52889   0.011 0.991039
## dump_within1:did_with_device_econ1       0.83392    0.69316   1.203 0.228953
## dump_within2:did_with_device_econ1      -0.20704    0.34786  -0.595 0.551719
## dump_within3:did_with_device_econ1       0.66524    0.27185   2.447 0.014401
## dump_within4:did_with_device_econ1       0.47258    0.24810   1.905 0.056815
## dump_within5:did_with_device_econ1       3.06522  144.96360   0.021 0.983130
## did_with_device_econ1:dump_reason_break1 -1.61778    0.92249  -1.754 0.079481
## did_with_device_econ1:dump_reason_theft1 -2.57704    1.43737  -1.793 0.072991
```

```
## memory_dev1:dump_reason_slow1              3.43606    1.41273   2.432 0.015007
## dump_reason_break1:dump_reason_theft1     -3.53091    1.42828  -2.472 0.013431
##
## (Intercept)
## dump_within1                             .
## dump_within2
## dump_within3                             **
## dump_within4                             *
## dump_within5
## did_with_device_econ1
## memory_dev1                              ***
## dump_reason_break1                       **
## dump_reason_theft1                       ***
## dump_reason_slow1
## dump_within1:did_with_device_econ1
## dump_within2:did_with_device_econ1
## dump_within3:did_with_device_econ1       *
## dump_within4:did_with_device_econ1       .
## dump_within5:did_with_device_econ1
## did_with_device_econ1:dump_reason_break1 .
## did_with_device_econ1:dump_reason_theft1 .
## memory_dev1:dump_reason_slow1            *
## dump_reason_break1:dump_reason_theft1    *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 485.33  on 350  degrees of freedom
## Residual deviance: 379.93  on 331  degrees of freedom
## AIC: 419.93
##
## Number of Fisher Scoring iterations: 14
```

The AIC Value was Initially 477.

After Stepwise Regression, it came down to 437

After Adding the Interaction elements, it is now 419.9321948

The 10 fold cross validation accuracy is below:

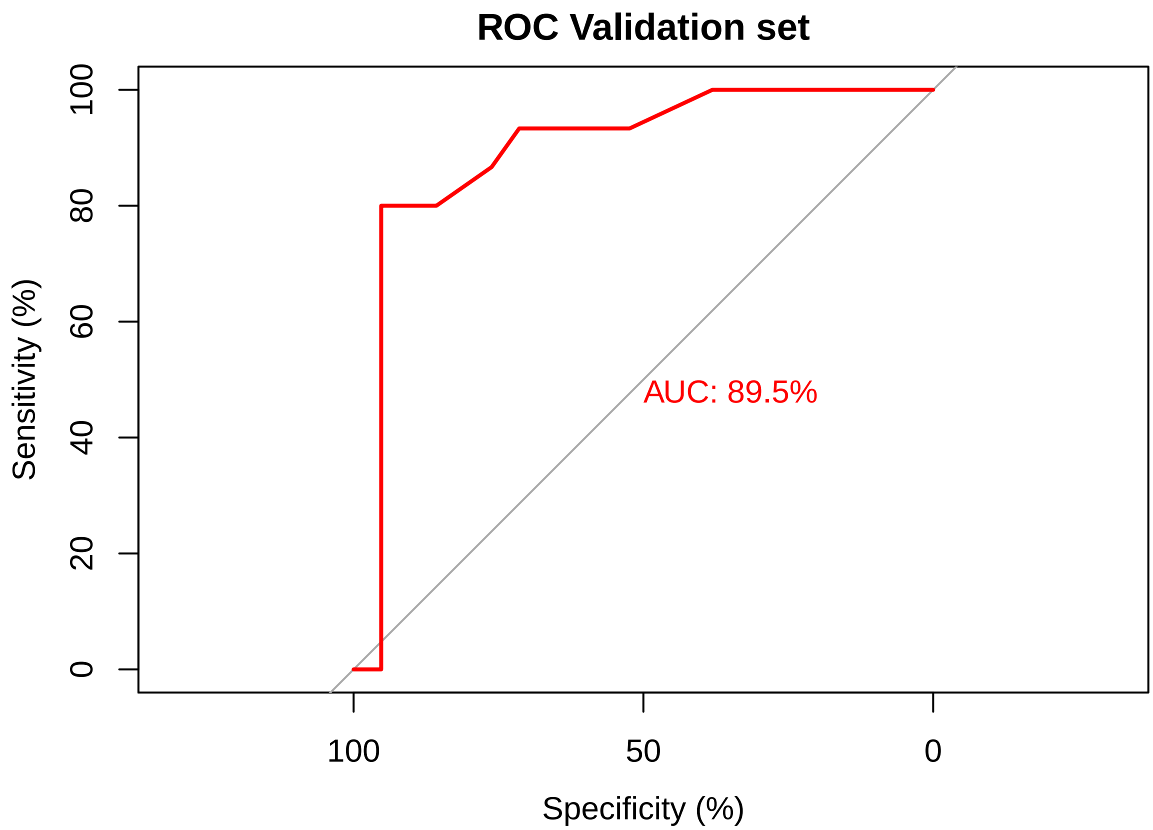
```r
suppressWarnings( cross_validated_model <- boot::cv.glm(df,glmfit = backward.model,K = 10))
print((1-cross_validated_model$delta[1])*100)
```

```
## [1] 78.89273
```

Lets also look at the ROC curve for the the fitted logistic regression model: ***(A better fit than model without interaction)***

```r
null.model <- glm(miss_dev~1,family = binomial(),data = df)
suppressWarnings(Generate_RoC(full_model = full.model,
null_model = null.model,df = df,direction = "B",outcome = "miss_dev"))
```

```
## [1] 351
```



With a good fit in the model, lets diagnosis our model parameters for multicolineariy. If the ***VIF*** (Variance Inflation Factor) is > 10 for any predictor, we might be in trouble.

```
cat("MIN IVF: ",min(vif(backward.model)))
```

```
## MIN IVF:  1
```

```
cat("MAX IVF: ",max(vif(backward.model)))
```

```
## MAX IVF:  7706858769
```

```
cat("MEAN IVF: ",mean(vif(backward.model)))
```

```
## MEAN IVF:  457406685
```

Adding Interaction introduces Multicolinearity among variables and makes the model unstable. We should not do that for this model. Lets go to chapter 3????

Chapter 3

Do People Try To Self Repair??

Model Before Applying Stepwise Regression. Not the final product. Go to the next page for an understandeable version.....

```
full.model %>% summary()
```

```
##
## Call:
## glm(formula = repair_try ~ age + gender + edu + device_count_5_yr +
##     miss_dev + dump_reason_new + dump_reason_old + dump_reason_break +
##     dump_reason_theft + dump_reason_slow + dump_reason_lag +
##     slt_lack_tu + slt_lack_lang + slt_lack_notint + slt_lack_fear +
##     slt_lack_parts + slt_lack_repairer + rprd_usage_chlng_No +
##     rprd_usage_chlng_reluc + rprd_usage_chlng_dur + rprd_usage_chlng_fault +
##     rpr_missing_trait_behave + rpr_missing_trait_ineff + rpr_missing_trait_harm +
##     rpr_missing_trait_hard + rpr_missing_trait_wage + rpr_missing_trait_trust +
##     rpr_missing_trait_gender + bad_rep_exp + dev_tknto_rec_Y +
##     dev_rec_chlng_fair_price + dev_rec_chlng_usable + dev_rec_chlng_datasec +
##     dev_rec_chlng_env_poll + dev_rec_chlng_hard_find + will_dev_rec +
##     did_with_data_Y + did_with_device_econ, family = binomial,
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5975  -0.8572   0.4249   0.7828   1.9465
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -3.81453  176.55050  -0.022 0.982762
## age1                       3.44412  176.55067   0.020 0.984436
## age2                       2.41671  176.54929   0.014 0.989078
## age3                       3.32647  176.54948   0.019 0.984967
## age4                       3.15159  176.55392   0.018 0.985758
## genderWoman               -0.76601    0.33177  -2.309 0.020952 *
## eduBachelors              -0.45058    0.32919  -1.369 0.171072
## eduMasters                -1.07114    0.55062  -1.945 0.051733 .
## device_count_5_yr          0.11875    0.06172   1.924 0.054337 .
## miss_dev1                  0.27342    0.29558   0.925 0.354952
## dump_reason_new1           0.07219    0.37264   0.194 0.846401
## dump_reason_old1          -0.42721    0.36414  -1.173 0.240712
## dump_reason_break1        -0.57940    0.34560  -1.676 0.093644 .
```

```
## dump_reason_theft1         -0.04497    0.55333  -0.081 0.935222
## dump_reason_slow1          -0.25820    0.47821  -0.540 0.589242
## dump_reason_lag1            0.22845    0.34634   0.660 0.509502
## slt_lack_tu1                0.85707    0.33681   2.545 0.010937 *
## slt_lack_lang1              0.74359    0.39371   1.889 0.058936 .
## slt_lack_notint1           -0.31912    0.35723  -0.893 0.371688
## slt_lack_fear1             -0.32464    0.30748  -1.056 0.291053
## slt_lack_parts1             0.70065    0.30417   2.304 0.021251 *
## slt_lack_repairer1          0.72305    0.33431   2.163 0.030556 *
## rprd_usage_chlng_No1        0.09627    0.47985   0.201 0.840985
## rprd_usage_chlng_reluc1    -0.24016    0.34020  -0.706 0.480238
## rprd_usage_chlng_dur1       0.19056    0.36570   0.521 0.602310
## rprd_usage_chlng_fault1    -0.70492    0.36279  -1.943 0.052012 .
## rpr_missing_trait_behave1  -0.07719    0.40245  -0.192 0.847906
## rpr_missing_trait_ineff1   -0.25274    0.30859  -0.819 0.412776
## rpr_missing_trait_harm1    -0.19489    0.30877  -0.631 0.527936
## rpr_missing_trait_hard1     0.41141    0.31345   1.313 0.189349
## rpr_missing_trait_wage1     0.16748    0.29675   0.564 0.572492
## rpr_missing_trait_trust1    0.65292    0.31573   2.068 0.038646 *
## rpr_missing_trait_gender1   0.80964    0.49728   1.628 0.103495
## bad_rep_expN               -0.15473    0.41822  -0.370 0.711399
## bad_rep_expY                0.33971    0.39820   0.853 0.393596
## dev_tknto_rec_Y1            0.60860    0.32634   1.865 0.062194 .
## dev_rec_chlng_fair_price1   0.09209    0.28958   0.318 0.750469
## dev_rec_chlng_usable1      -0.18539    0.30092  -0.616 0.537841
## dev_rec_chlng_datasec1      0.21768    0.32591   0.668 0.504180
## dev_rec_chlng_env_poll1     0.19792    0.32119   0.616 0.537756
## dev_rec_chlng_hard_find1   -1.14618    0.45119  -2.540 0.011075 *
## will_dev_recN               0.02867    0.36626   0.078 0.937615
## will_dev_recY              -0.35005    0.46936  -0.746 0.455780
## did_with_data_Y1            0.28373    0.30745   0.923 0.356089
## did_with_device_econ1       1.31402    0.37878   3.469 0.000522 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 448.21  on 350  degrees of freedom
## Residual deviance: 346.64  on 306  degrees of freedom
## AIC: 436.64
##
## Number of Fisher Scoring iterations: 13
```

```r
backward.model <- step(full.model,direction = "backward",trace = 0)
backward.model %>% summary()
```

```
##
## Call:
## glm(formula = repair_try ~ gender + device_count_5_yr + dump_reason_break +
##     slt_lack_tu + slt_lack_lang + slt_lack_parts + slt_lack_repairer +
##     rprd_usage_chlng_fault + rpr_missing_trait_trust + rpr_missing_trait_gender +
##     dev_tknto_rec_Y + dev_rec_chlng_hard_find + did_with_device_econ,
##     family = binomial, data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4999  -0.9607   0.5115   0.8135   1.7348
```

```
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -1.25724    0.48891  -2.572  0.01013 *
## genderWoman               -0.76450    0.29607  -2.582  0.00982 **
## device_count_5_yr          0.11214    0.05442   2.061  0.03932 *
## dump_reason_break1        -0.40389    0.26534  -1.522  0.12796
## slt_lack_tu1               0.90456    0.30414   2.974  0.00294 **
## slt_lack_lang1             0.75308    0.37256   2.021  0.04324 *
## slt_lack_parts1            0.73106    0.27123   2.695  0.00703 **
## slt_lack_repairer1         0.68465    0.30933   2.213  0.02688 *
## rprd_usage_chlng_fault1   -0.66105    0.29823  -2.217  0.02665 *
## rpr_missing_trait_trust1   0.64316    0.27960   2.300  0.02143 *
## rpr_missing_trait_gender1  0.70075    0.45174   1.551  0.12085
## dev_tknto_rec_Y1           0.53534    0.28816   1.858  0.06320 .
## dev_rec_chlng_hard_find1  -1.03492    0.39741  -2.604  0.00921 **
## did_with_device_econ1      1.01172    0.32764   3.088  0.00202 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 448.21  on 350  degrees of freedom
## Residual deviance: 367.01  on 337  degrees of freedom
## AIC: 395.01
##
## Number of Fisher Scoring iterations: 5
```

$$
\begin{aligned}
\log\left[\frac{P(repair\_try=1)}{1-P(repair\_try=1)}\right] = & -1.26 \\
& -0.76(gender_{woman}) \\
& +0.11(device\_count\_during\_5\_yr) \\
& -0.4(device\_dump\_reason:\ break) \\
& +0.9(self\_repair\_challenge:lack\_of\_tutorial) \\
& +0.75(self\_repair\_challenge:language\_barrier\_of\_tutorial) \\
& +0.73(self\_repair\_challenge:lack\_of\_parts) \\
& +0.68(feel\_the\_lack\_of\_good\_repairer) \\
& -0.66(fear\_that\_repaired\_device's\_faulty) \\
& +0.64(repairers\_lack\_trust)+0.7(gender\_adversary\_of\_repairers) \\
& +0.54(devices\_are\_given\_to\_RECYCLERS) \\
& -1.03(response:Recyclers\_are\_hard\_to\_find) \\
& +1.01(does\_economic\_activity\_with\_device)+\epsilon
\end{aligned}
$$

The AIC Value was Initially 436. After Stepwise Regression, it came down to 395.

The 10 fold cross validation accuracy is below:

```
cross_validated_model <- boot::cv.glm(df,glmfit = backward.model,K = 10)
print((1-cross_validated_model$delta[1])*100)
```

```
## [1] 80.60291
```

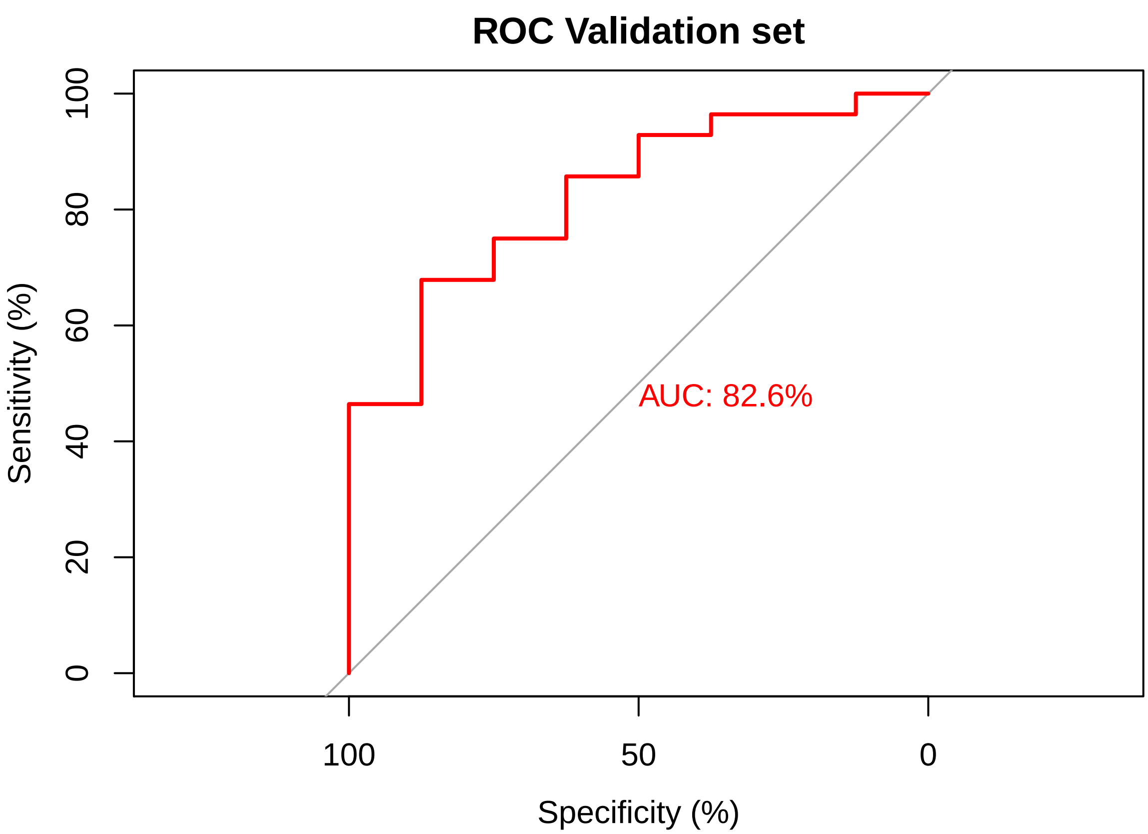
Lets also look at the ROC curve for the the fitted logistic regression model:

```
null.model <- glm(repair_try~1,family = binomial(),data = df)
Generate_RoC(full_model = full.model,null_model = null.model,
             df = df,direction = "B",outcome = "repair_try")
```

```
## [1] 351
```

```
## Setting levels: control = 0, case = 1
```

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



With a good fit in the model, lets diagnosis our model parameters for multicolineariy. If the ***VIF*** (Variance Inflation Factor) is > 10 for any predictor, we might be in trouble.

```
cat("MIN IVF: ",min(vif(backward.model)))
```

```
## MIN IVF:  1.034216
```

```
cat("MAX IVF: ",max(vif(backward.model)))
```

```
## MAX IVF:  1.242712
```

```
cat("MEAN IVF: ",mean(vif(backward.model)))
```

```
## MEAN IVF:  1.124022
```

Values > 0 means odd>1 So, that accounts more for YES than NO.

gender_woman: -0.76: Typicaly women are less inclined toward repair.

device_count_5_yr: 0.11 The more number of devices you have had during the last 5 years, the more you are inclined to try repair your own devices.

lack_of_tutorial: 0.9 If you feel that there is lack of tutorial out there, then there is a solid chance that you at least tried repair your own, but probably did/not succeed due to the lack of it.

language_barrier_of_tutorial: 0.75 This also predicts very well you feel that there is lack of tutorial ***In Your Own Language***, but for the same reason stated above, there is solid chance that you at least tried repair your own, giving odd ratio of 2.12.

lack_of_parts: 0.73 Similar reason.

fear that trying to repair will lead to faulty device -0.66 The negative log odd says this fear will lead to less amount of self repair trial.

repairers_lack_trust: 0.64 When you dont trust the repairers out there in the market, you try to do it on your own.

gender_adversary_of_repairers: 0.70 Gender Adversary makes you more likely to try your own luck.)

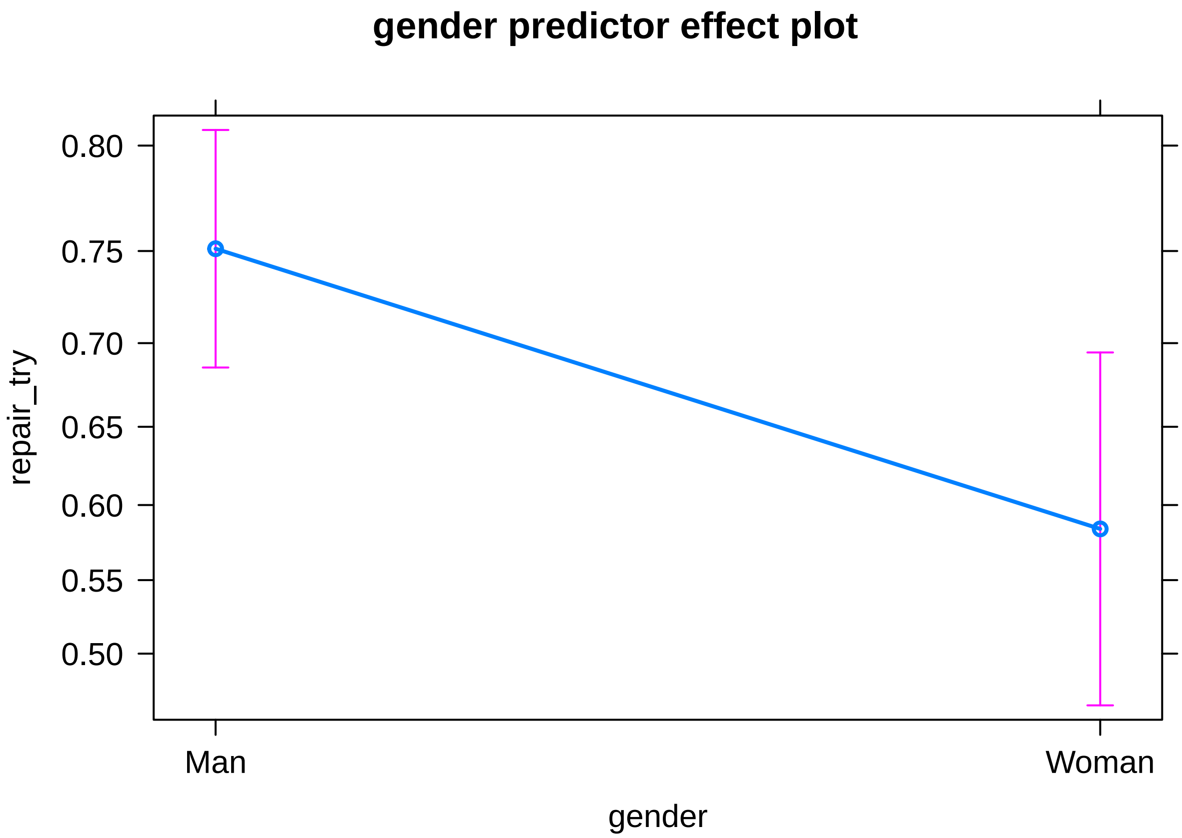
device ever taken to RECYCLER: 0.54 If you know the recycler community, and used to give them your devices/sell to them, then you care less about the safety and usefulness, and overcome fear to try your own.

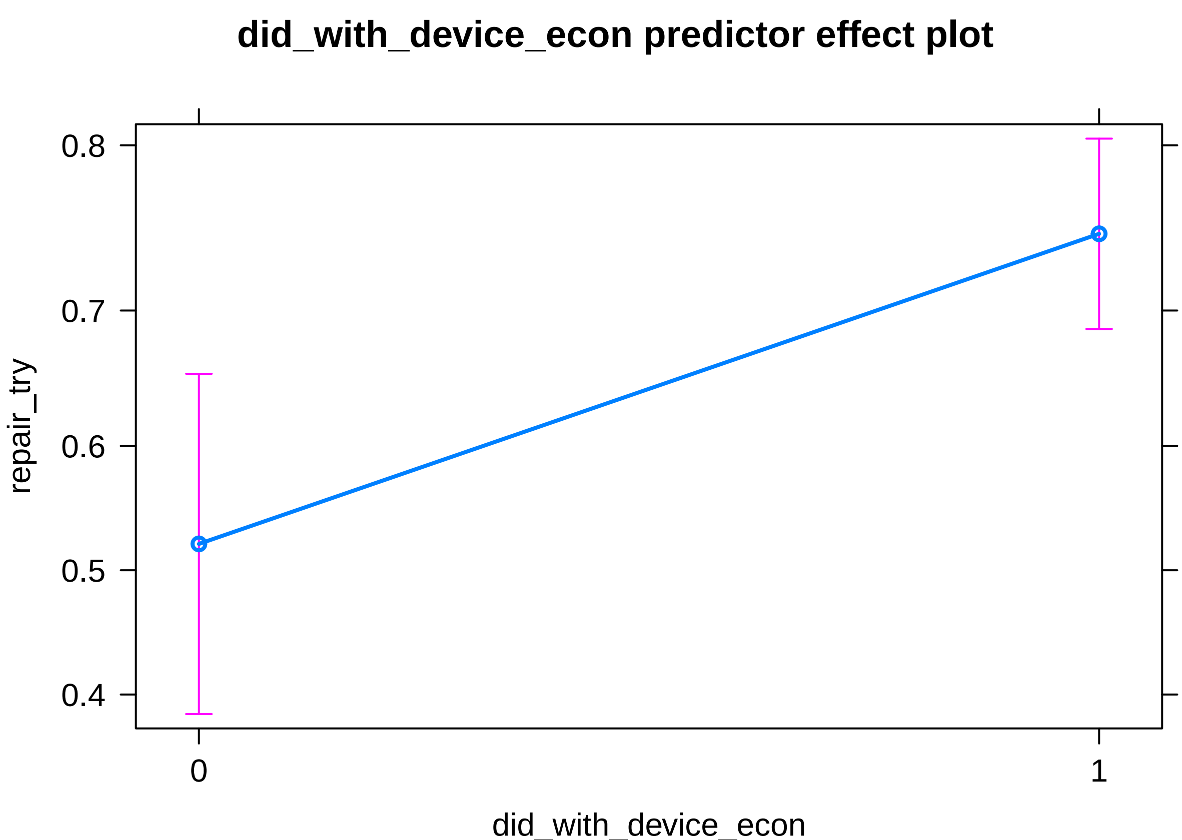
Hard to Find RECYCLER: $ -1.03$ The above effect is better described with understanding from this. When you find it hard to find any recycler, your fear of losing the device utility (log odd -0.66) intensifies. And, so you do not try repair your devices on your own.

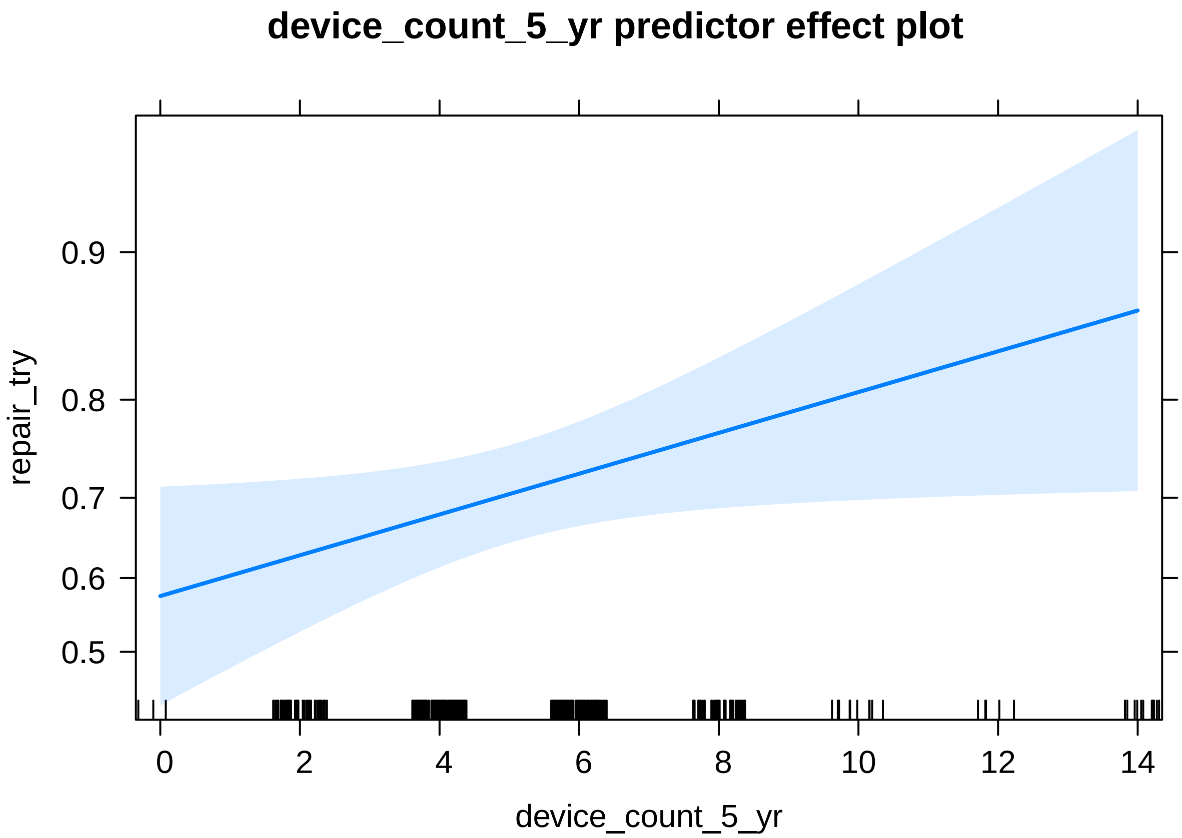
Have you did anything economic with your device? (selling/parts selling etc) 1.01. If you know your device still has some monetary value, you try to do your own. [Probably a bit contradictory! How do we fix that ? :P]

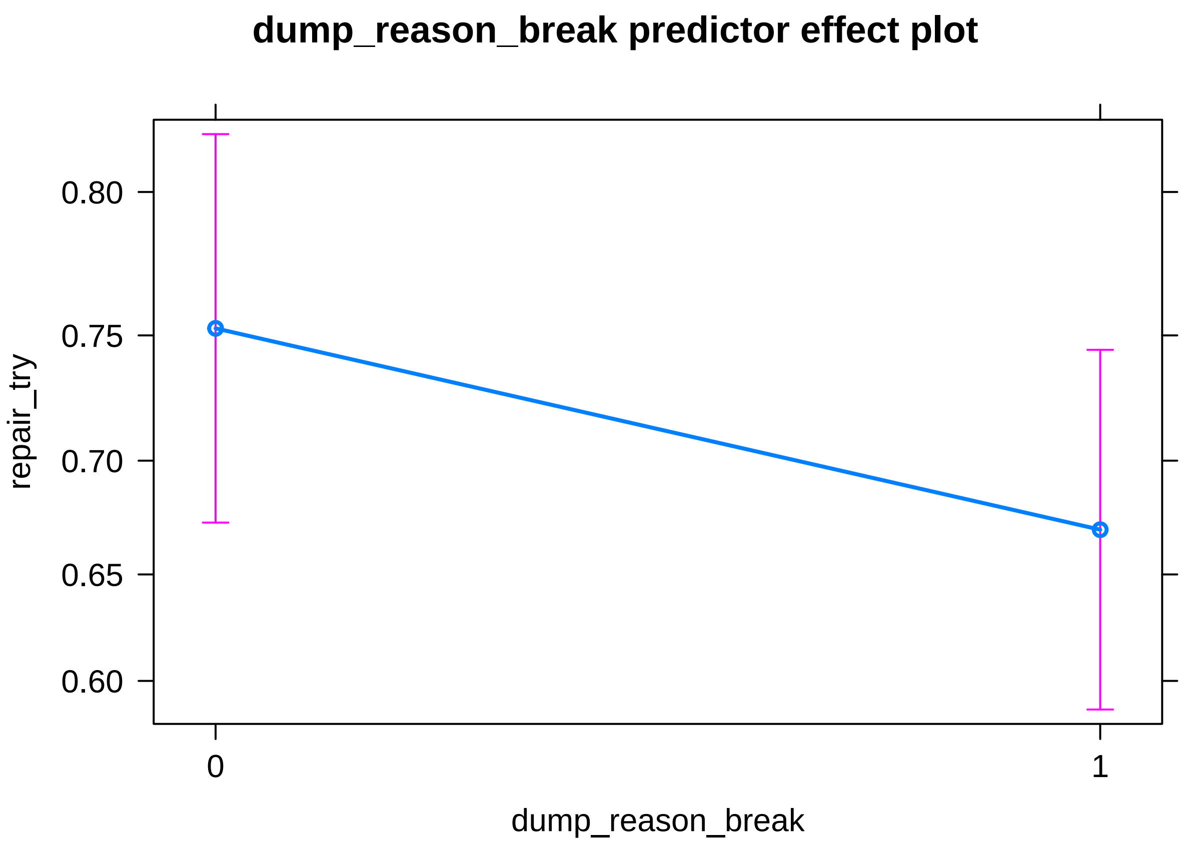
Lets look at the effects plot to better understand the individual effects of each predictor:

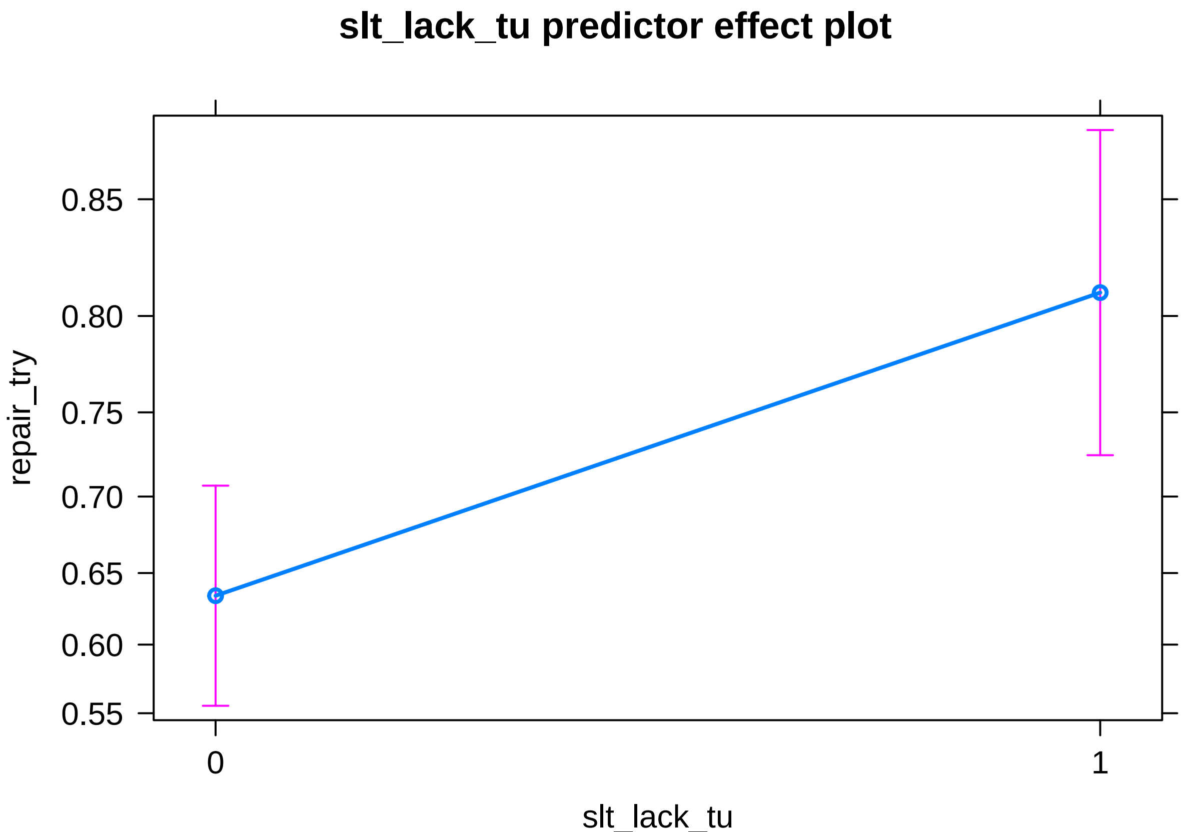
```r
#librarian::shelf(effects)

for (n in c(
"gender"
,"did_with_device_econ"
,"device_count_5_yr"
,"dump_reason_break"
,"slt_lack_tu"
,"slt_lack_lang"
,"slt_lack_parts"
,"slt_lack_repairer"
,"rprd_usage_chlng_fault"
,"rpr_missing_trait_trust"
,"rpr_missing_trait_gender"
,"dev_tknto_rec_Y"
,"dev_rec_chlng_hard_find"
)){
  #print(n)
  print(plot(effects::predictorEffect(n,backward.model)))
}
```

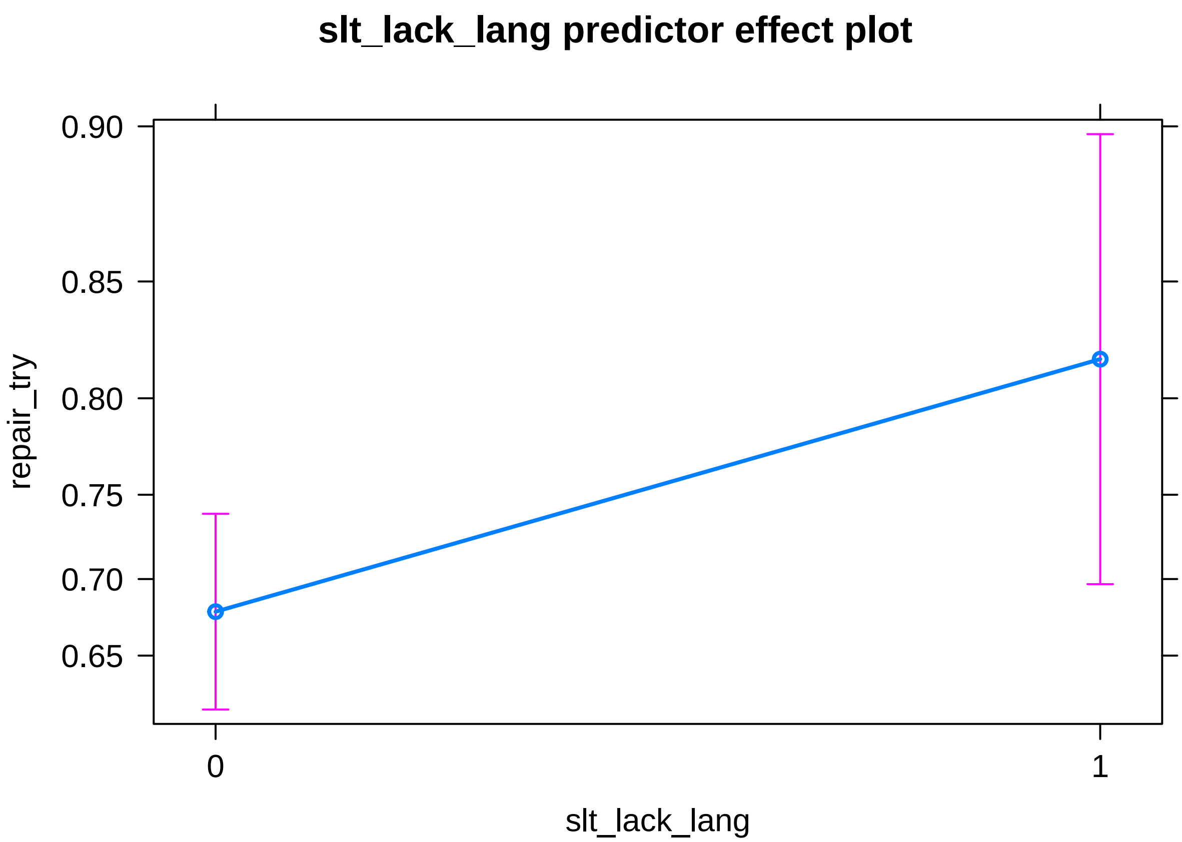

gender predictor effect plot
repair_try
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Man
Woman
gender

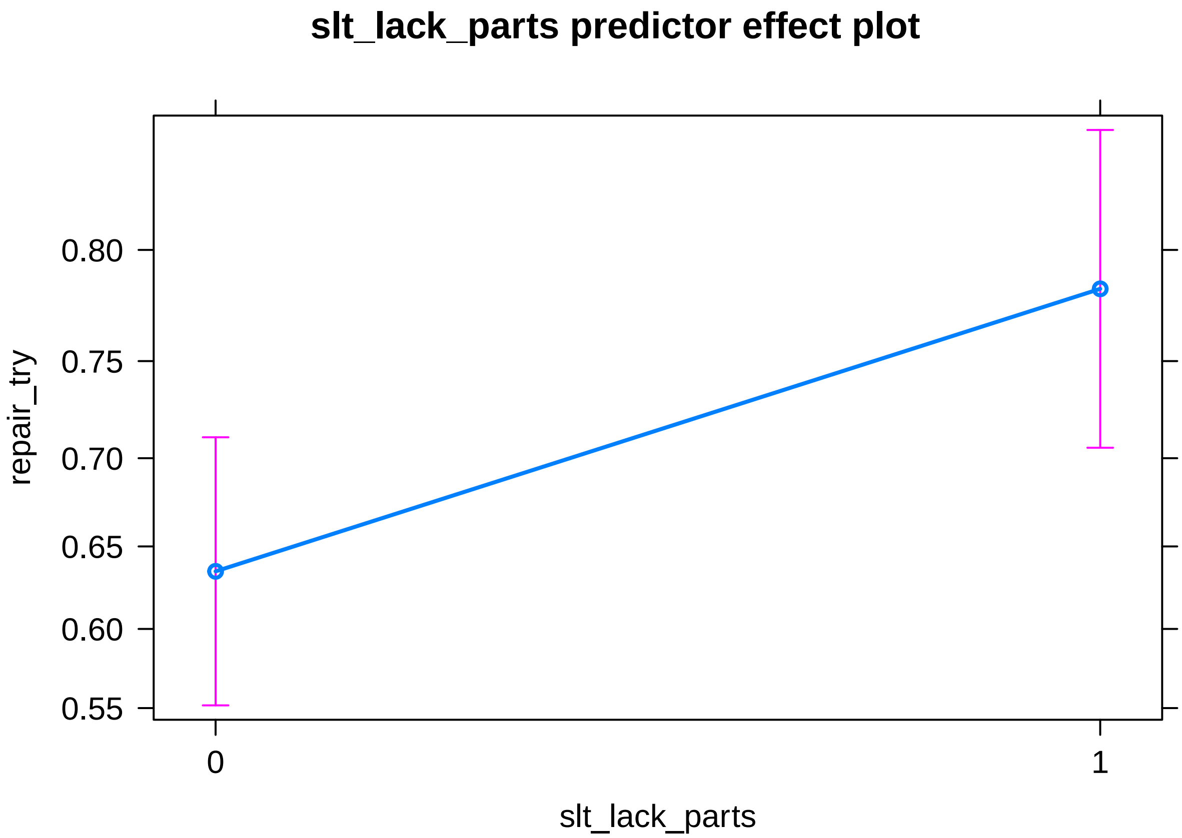

did_with_device_econ predictor effect plot
repair_try
0.8
0.7
0.6
0.5
0.4
0
1
did_with_device_econ

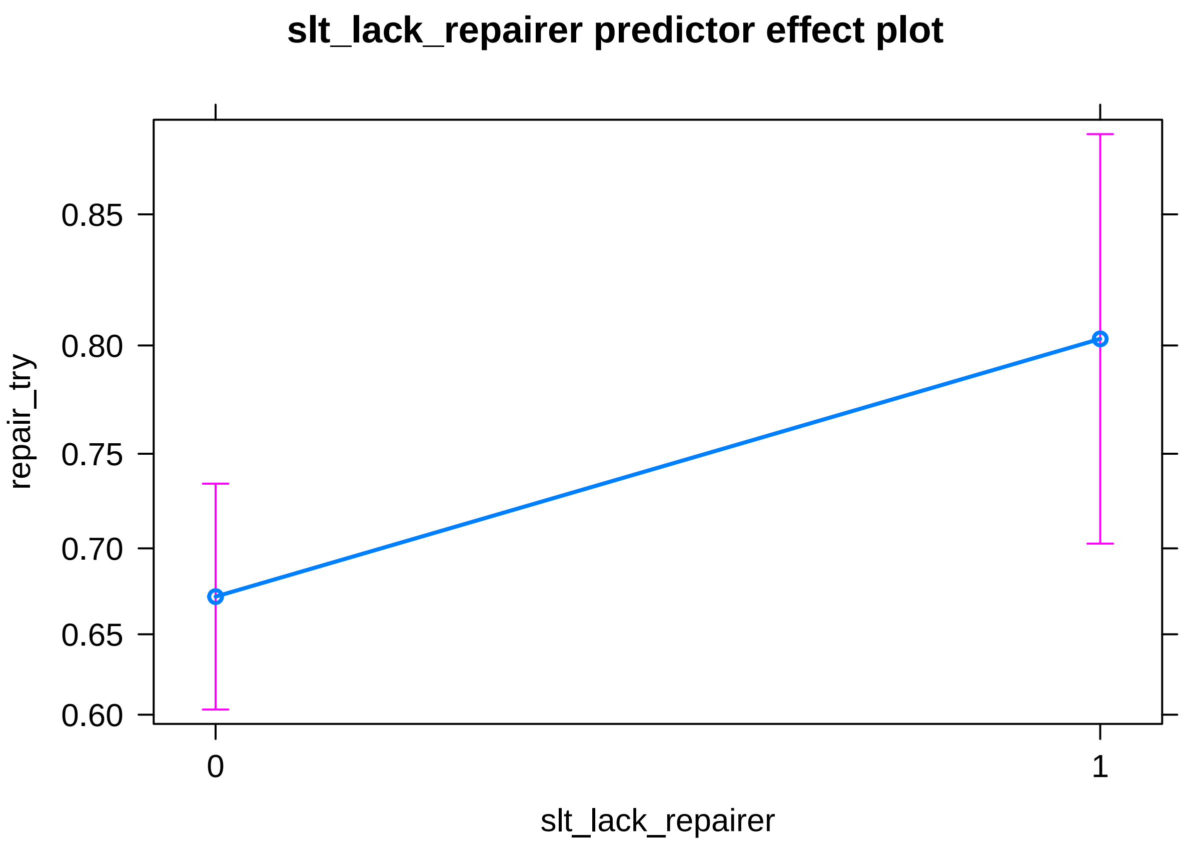

device_count_5_yr predictor effect plot
repair_try
0.5
0.6
0.7
0.8
0.9
0
2
4
6
8
10
12
14
device_count_5_yr

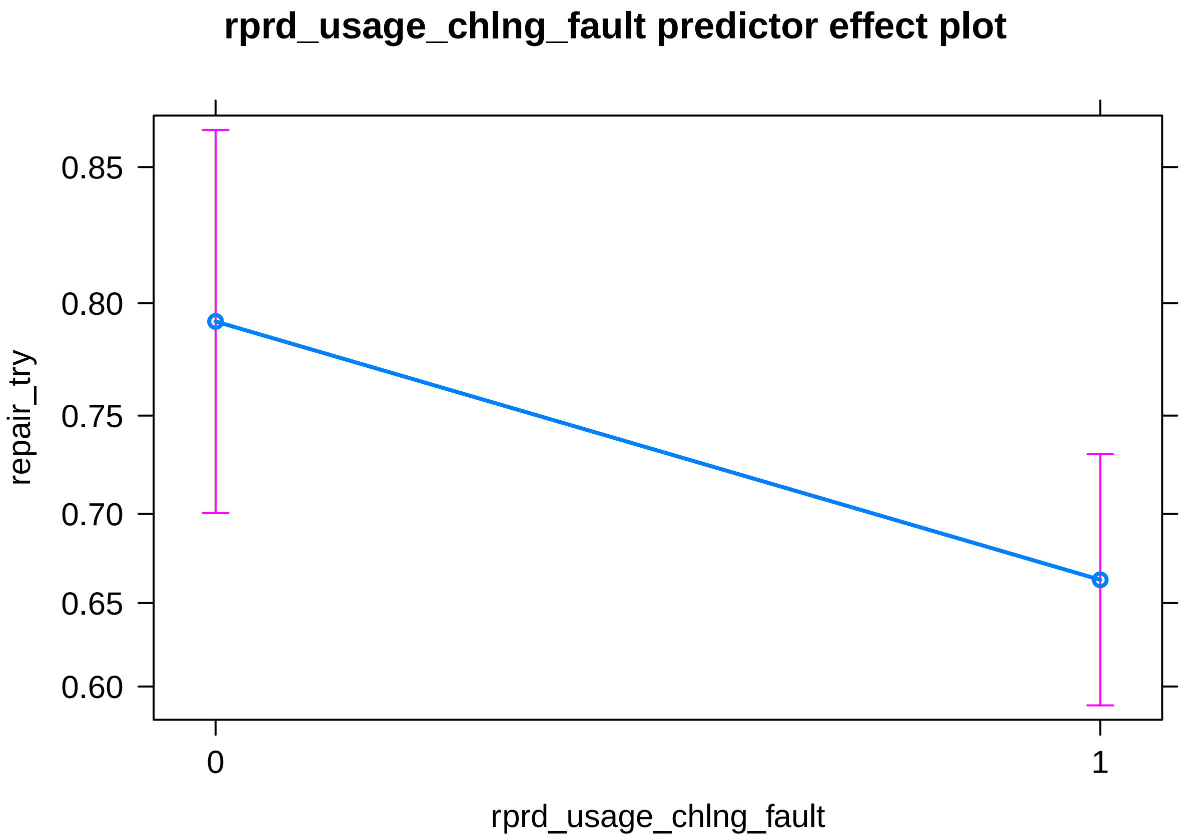

dump_reason_break predictor effect plot
repair_try
0.60
0.65
0.70
0.75
0.80
0
1
dump_reason_break


slt_lack_tu predictor effect plot
repair_try
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0
1
slt_lack_tu

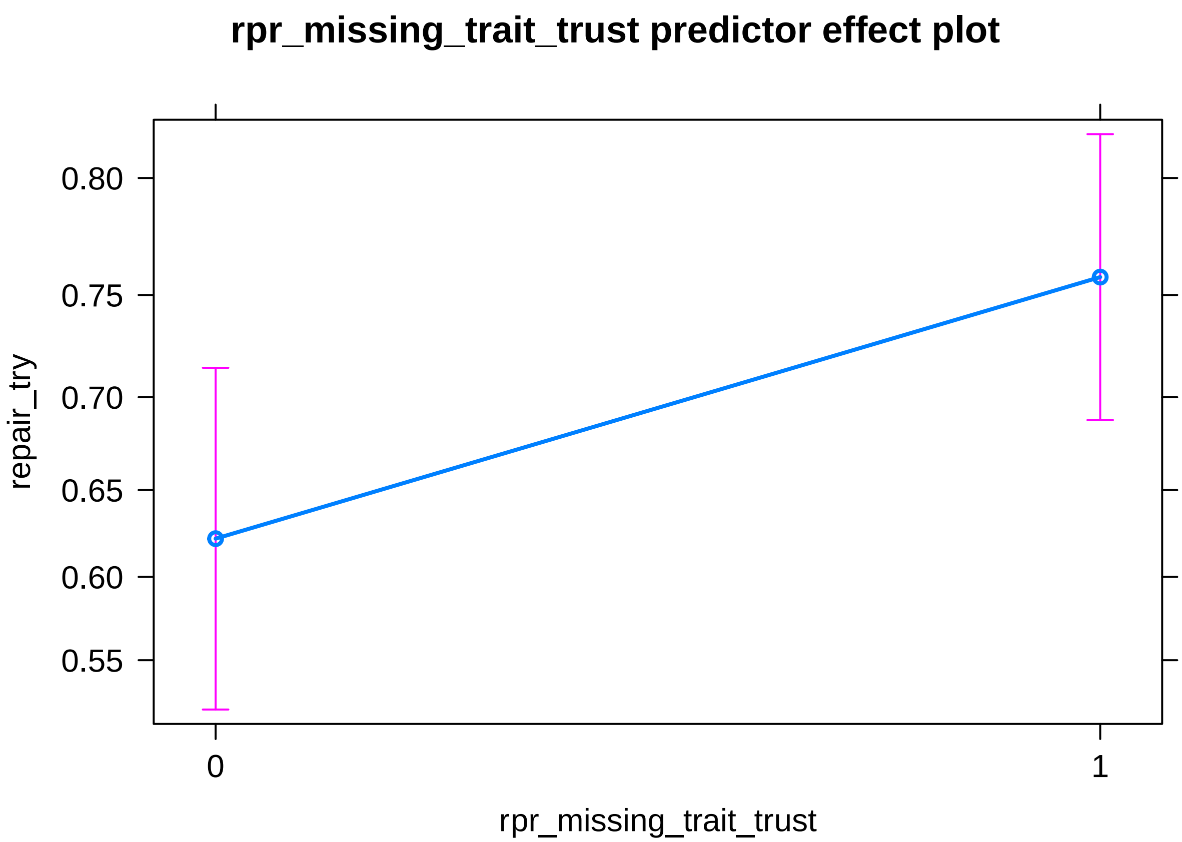

slt_lack_lang predictor effect plot
repair_try
0.90
0.85
0.80
0.75
0.70
0.65
0
1
slt_lack_lang


slt_lack_parts predictor effect plot
repair_try
0.80
0.75
0.70
0.65
0.60
0.55
0
1
slt_lack_parts

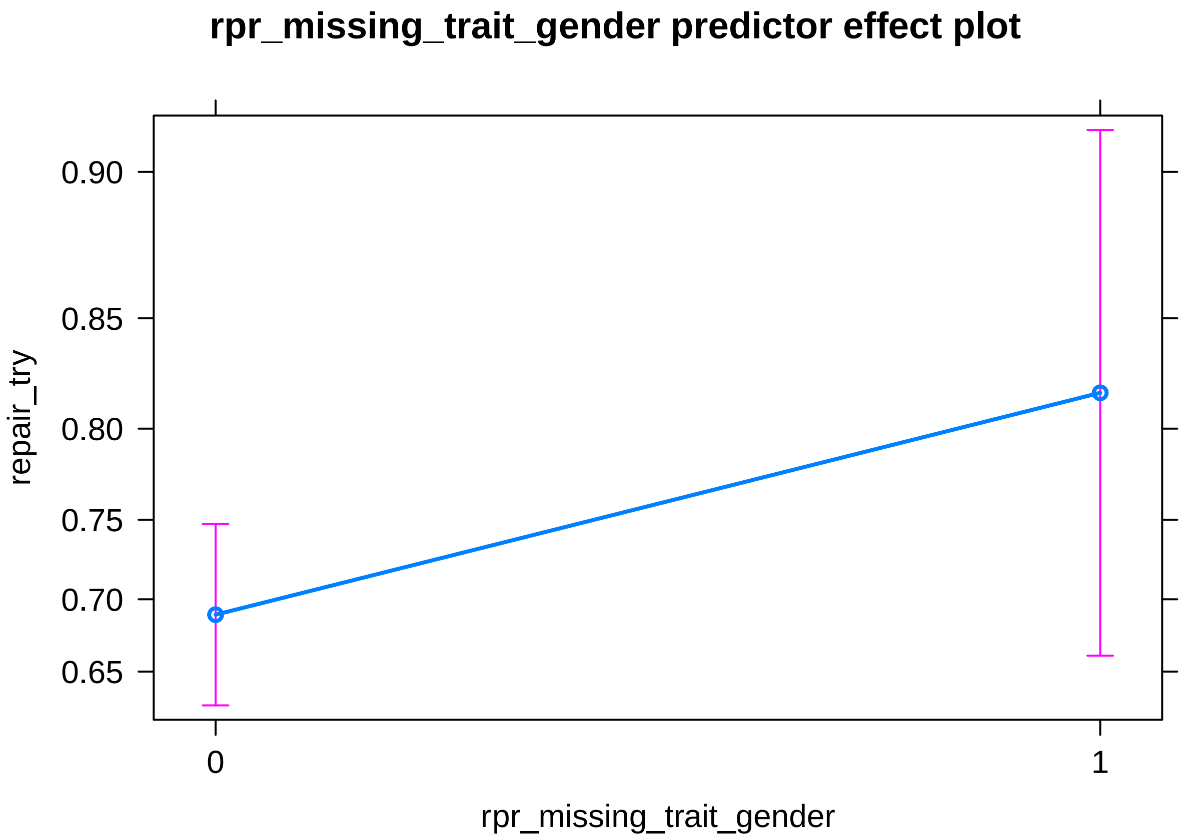

slt_lack_repairer predictor effect plot
repair_try
0.85
0.80
0.75
0.70
0.65
0.60
0
1
slt_lack_repairer

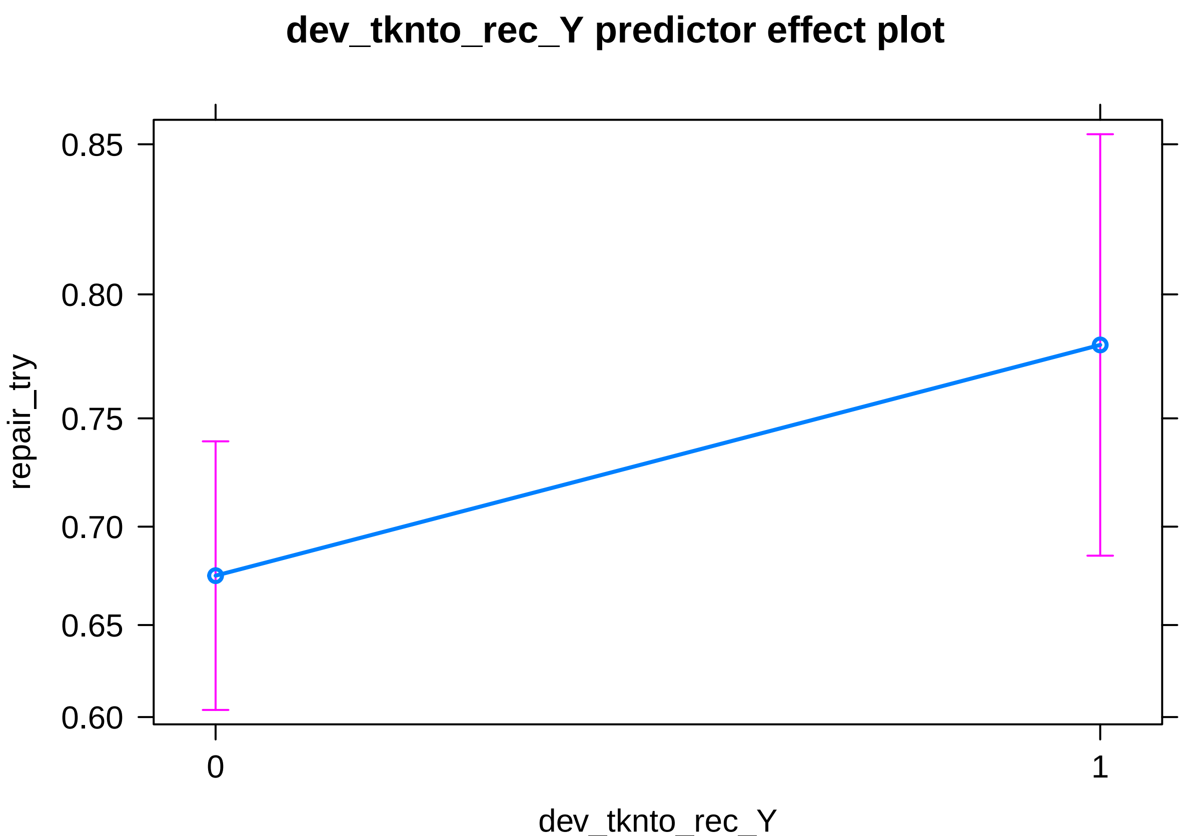


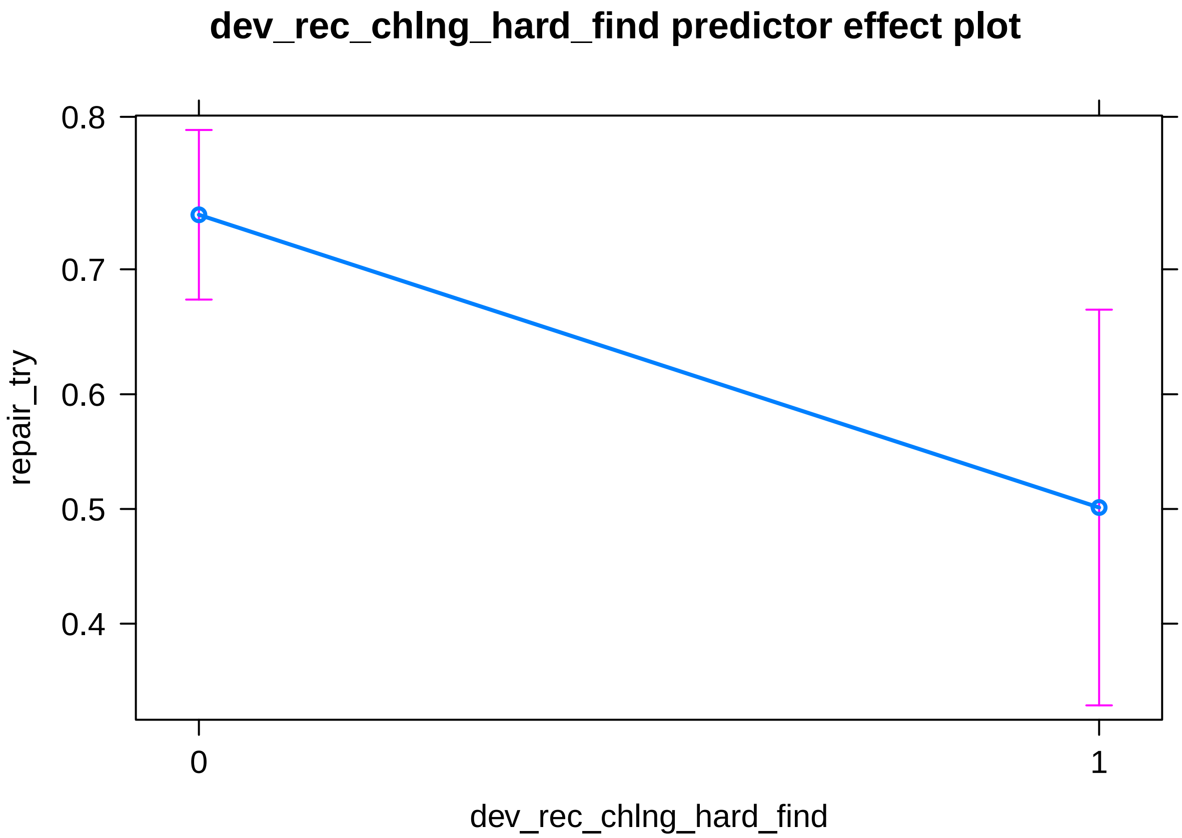
rprd_usage_chlng_fault predictor effect plot
repair_try
0.85
0.80
0.75
0.70
0.65
0.60
0
1
rprd_usage_chlng_fault



rpr_missing_trait_trust predictor effect plot
repair_try
0.80
0.75
0.70
0.65
0.60
0.55
0
1
rpr_missing_trait_trust


rpr_missing_trait_gender predictor effect plot
repair_try
0.90
0.85
0.80
0.75
0.70
0.65
0
1
rpr_missing_trait_gender


dev_tknto_rec_Y predictor effect plot
repair_try
0.85
0.80
0.75
0.70
0.65
0.60
0
1
dev_tknto_rec_Y


dev_rec_chlng_hard_find predictor effect plot
repair_try
0.8
0.7
0.6
0.5
0.4
0
1
dev_rec_chlng_hard_find

Chapter 4

Can We Go Beyond Simple Predictors? [Possibly Yes]

This time we have added the following Interactions:

device_count_5_yr:rprd_usage_chlng_fault

device_count_5_yr:dev_tknto_rec_Y

device_count_5_yr:did_with_device_econ

slt_lack_tu:dev_tknto_rec_Y

slt_lack_lang:rpr_missing_trait_gender

slt_lack_parts:rpr_missing_trait_gender

rpr_missing_trait_gender:dev_rec_chlng_hard_find

rpr_missing_trait_trust:dev_rec_chlng_hard_find

rpr_missing_trait_gender:did_with_device_econ

```r
backward.model <- step(full.model,direction = "backward",trace = 0)

backward.model %>% summary()
```

```
##
## Call:
## glm(formula = repair_try ~ gender + device_count_5_yr + dump_reason_break +
##     slt_lack_tu + slt_lack_lang + slt_lack_parts + slt_lack_repairer +
##     rprd_usage_chlng_fault + rpr_missing_trait_trust + rpr_missing_trait_gender +
##     dev_tknto_rec_Y + dev_rec_chlng_hard_find + did_with_device_econ +
##     device_count_5_yr:rprd_usage_chlng_fault + device_count_5_yr:dev_tknto_rec_Y +
##     slt_lack_tu:dev_tknto_rec_Y + rpr_missing_trait_gender:did_with_device_econ,
##     family = binomial, data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4798  -0.9052   0.3998   0.7900   1.8345
##
## Coefficients:
##                                               Estimate Std. Error z value
## (Intercept)                                    0.08008    0.60611   0.132
## genderWoman                                   -0.76520    0.30643  -2.497
```

```
## device_count_5_yr                              -0.10481    0.08528  -1.229
## dump_reason_break1                             -0.39397    0.27349  -1.441
## slt_lack_tu1                                    0.60696    0.34801   1.744
## slt_lack_lang1                                  0.66268    0.38551   1.719
## slt_lack_parts1                                 0.82522    0.28030   2.944
## slt_lack_repairer1                              0.64544    0.31311   2.061
## rprd_usage_chlng_fault1                        -2.08754    0.63395  -3.293
## rpr_missing_trait_trust1                        0.72791    0.28796   2.528
## rpr_missing_trait_gender                       -0.55054    0.79032  -0.697
## dev_tknto_rec_Y1                               -1.44058    0.80486  -1.790
## dev_rec_chlng_hard_find1                       -1.06221    0.41026  -2.589
## did_with_device_econ1                           0.74643    0.35720   2.090
## device_count_5_yr:rprd_usage_chlng_fault1       0.28921    0.11153   2.593
## device_count_5_yr:dev_tknto_rec_Y1              0.31305    0.14984   2.089
## slt_lack_tu1:dev_tknto_rec_Y1                   1.61188    0.73681   2.188
## rpr_missing_trait_gender:did_with_device_econ1  1.68981    0.96570   1.750
##                                                Pr(>|z|)
## (Intercept)                                    0.894891
## genderWoman                                    0.012520 *
## device_count_5_yr                              0.219066
## dump_reason_break1                             0.149713
## slt_lack_tu1                                   0.081142 .
## slt_lack_lang1                                 0.085620 .
## slt_lack_parts1                                0.003239 **
## slt_lack_repairer1                             0.039267 *
## rprd_usage_chlng_fault1                        0.000992 ***
## rpr_missing_trait_trust1                       0.011478 *
## rpr_missing_trait_gender                       0.486053
## dev_tknto_rec_Y1                               0.073476 .
## dev_rec_chlng_hard_find1                       0.009622 **
## did_with_device_econ1                          0.036644 *
## device_count_5_yr:rprd_usage_chlng_fault1      0.009512 **
## device_count_5_yr:dev_tknto_rec_Y1             0.036686 *
## slt_lack_tu1:dev_tknto_rec_Y1                  0.028696 *
## rpr_missing_trait_gender:did_with_device_econ1 0.080146 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 448.21  on 350  degrees of freedom
## Residual deviance: 349.44  on 333  degrees of freedom
## AIC: 385.44
##
## Number of Fisher Scoring iterations: 5
```

$$
\begin{aligned}
\log\left[\frac{P(repair\_try=1)}{1-P(repair\_try=1)}\right] = {} & 0.08 - 0.77(gender_{Woman}) - 0.1(device\_count\_5\_yr) \\
& - 0.39(dump\_reason\_break) + 0.61(slt\_lack\_tu) + 0.66(slt\_lack\_lang) \\
& + 0.83(slt\_lack\_parts) + 0.65(slt\_lack\_repairer) - 2.09(rprd\_usage\_chlng\_fault) \\
& + 0.73(rpr\_missing\_trait\_trust) - 0.55(rpr\_missing\_trait\_gender) \\
& - 1.44(dev\_tknto\_rec\_Y) \ - 1.06(dev\_rec\_chlng\_hard\_find) \\
& + 0.75(did\_with\_device\_econ_1) + 0.29(device\_count\_5\_yr \times rprd\_usage\_chlng\_fault) \\
& + 0.31(device\_count\_5\_yr \times dev\_tknto\_rec\_Y) + 1.61(slt\_lack\_tu \times dev\_tknto\_rec\_Y) \\
& + 1.69(rpr\_missing\_trait\_gender \times did\_with\_device\_econ) \ + \epsilon
\end{aligned}
$$

The AIC Value was Initially 436.

After Stepwise Regression, it came down to 395.

After Adding the Interaction elements, it is now 385

The 10 fold cross validation accuracy is below:

```
cross_validated_model <- boot::cv.glm(df,glmfit = backward.model,K = 10)
print((1-cross_validated_model$delta[1])*100)
```

```
## [1] 80.93028
```

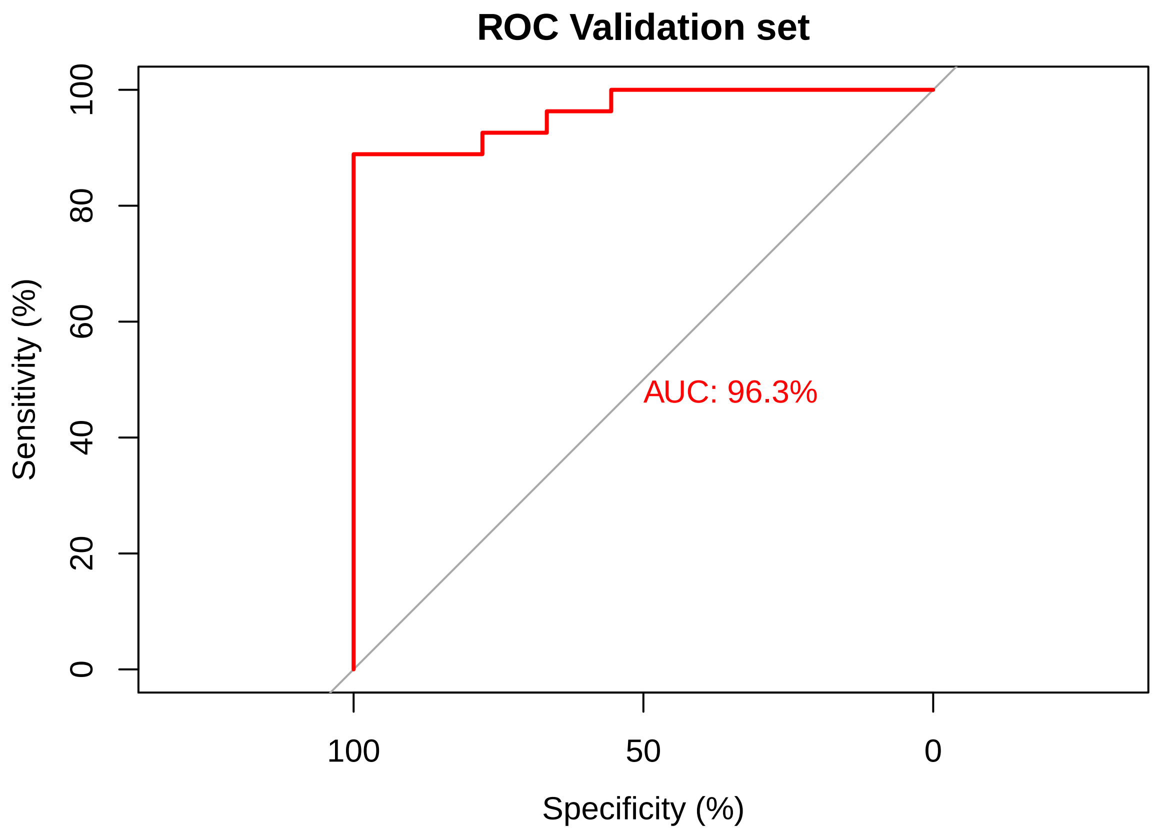
Lets also look at the ROC curve for the the fitted logistic regression model: ***(A better fit than model without interaction)***

```
null.model <- glm(repair_try~1,family = binomial(),data = df)
Generate_RoC(full_model = full.model,null_model = null.model,
             df = df,direction = "B",outcome = "repair_try")
```

```
## [1] 351
```

```
## Setting levels: control = 0, case = 1
```

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



With a good fit in the model, lets diagnosis our model parameters for multicolineariy. If the ***VIF*** (Variance Inflation Factor) is > 10 for any predictor, we might be in trouble.

```
cat("MIN IVF: ",min(vif(backward.model)))
```

```
## MIN IVF:  1.054682
```

```
cat("MAX IVF: ",max(vif(backward.model)))
```

```
## MAX IVF:  6.98304
```

```
cat("MEAN IVF: ",mean(vif(backward.model)))
```

```
## MEAN IVF:  2.6841
```

Values > 0 means odd> 1 So, that accounts more for YES than NO.

device_count_5_yr : fear of faulty device

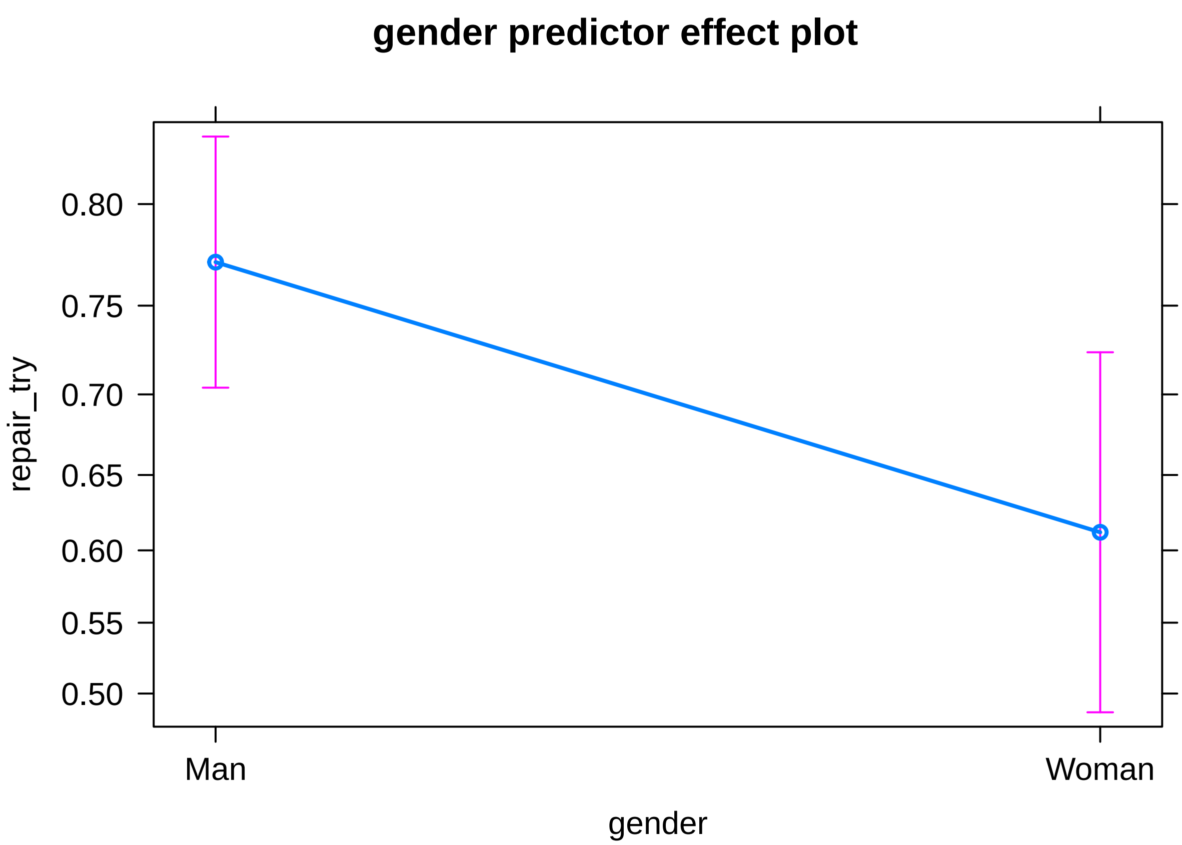
device_count_5_yr: device ever taken to RECYCLER

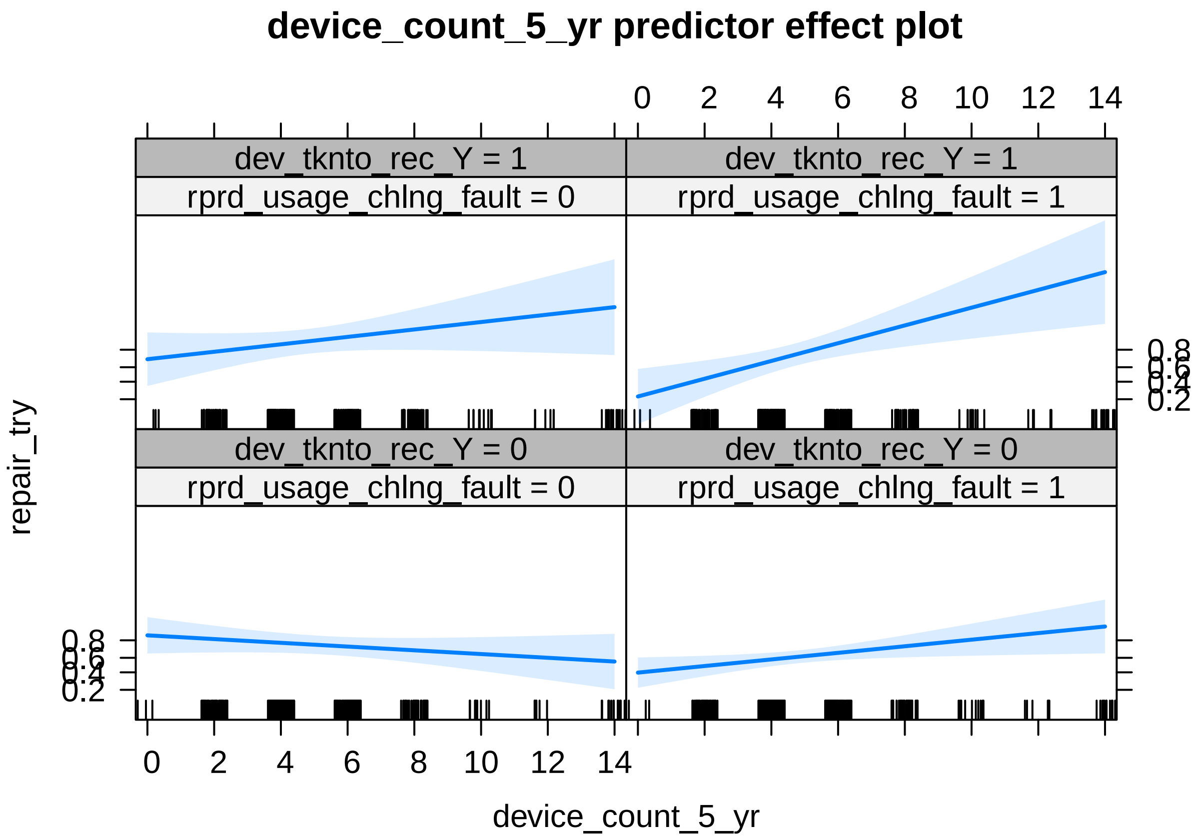
lack_of_tutorial : device ever taken to RECYCLER

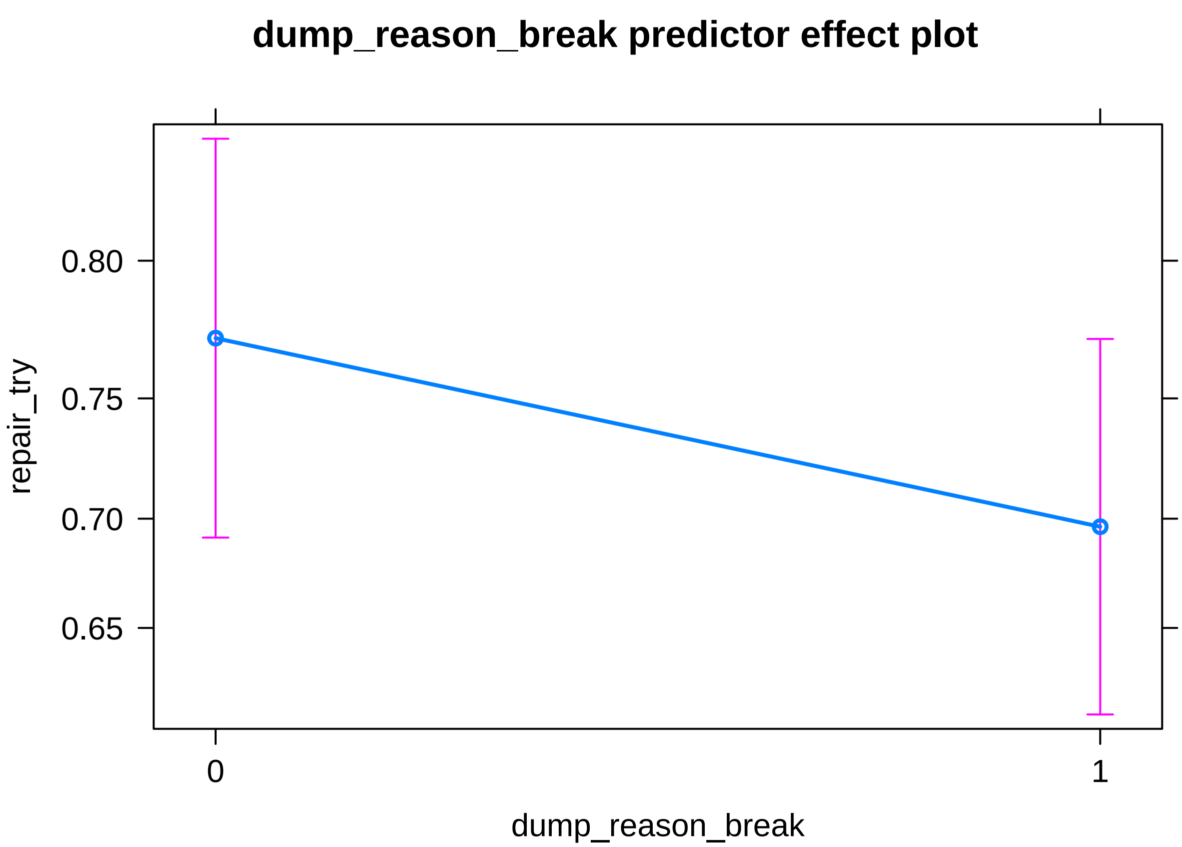
gender_adversary_of_repairers :Have you did anything economic with your device

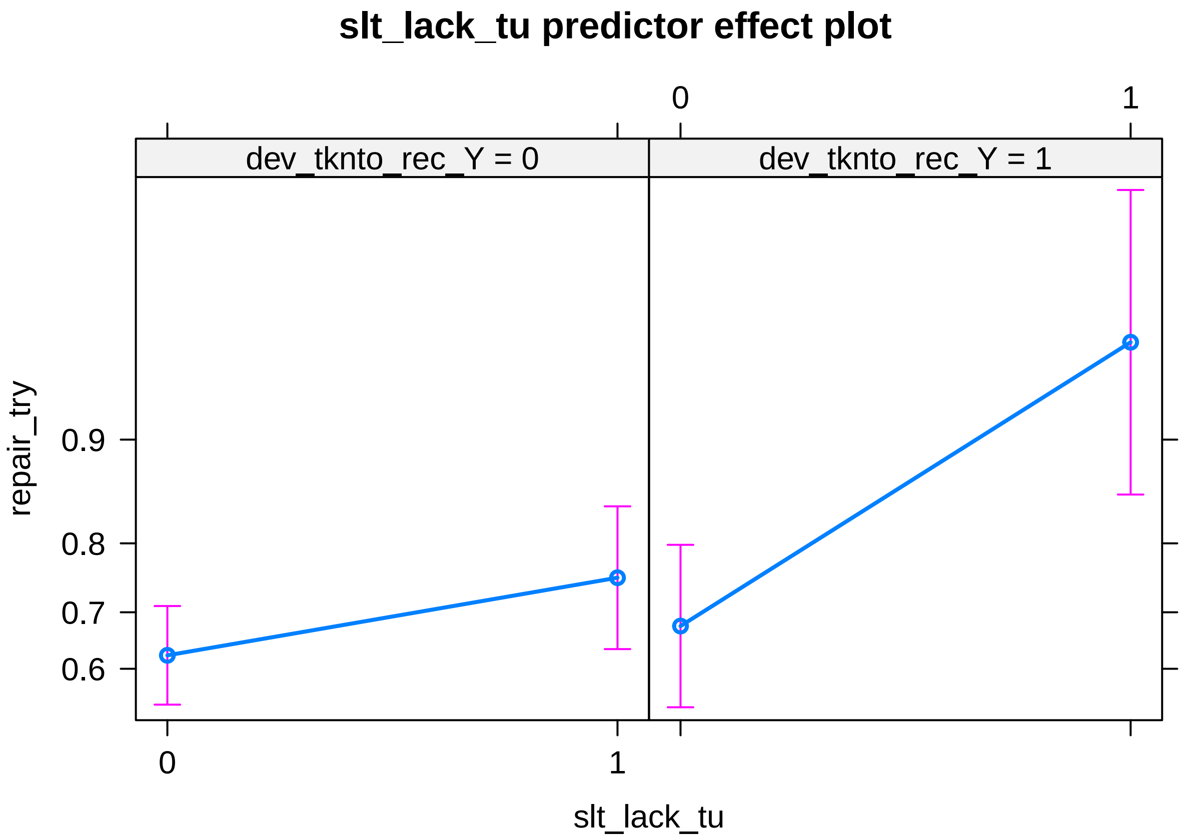
Lets look at the effects plot to better understand the individual effects of each predictor:

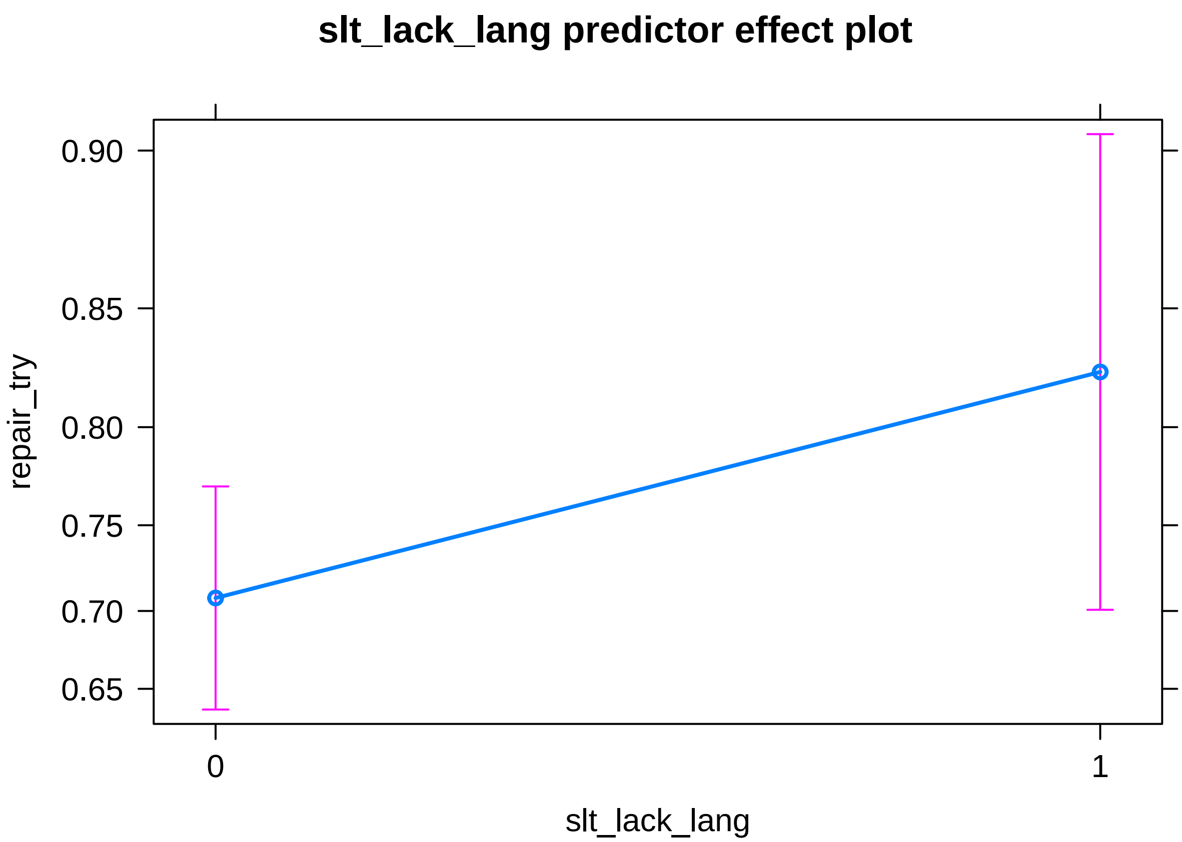
```
#librarian::shelf(effects)

for (n in c(
"gender",
"device_count_5_yr"
,"dump_reason_break"
,"slt_lack_tu"
,"slt_lack_lang"
,"slt_lack_parts"
,"slt_lack_repairer"
,"rprd_usage_chlng_fault"
,"rpr_missing_trait_trust"
,"rpr_missing_trait_gender"
,"dev_rec_chlng_hard_find",
"did_with_device_econ"
,"dev_tknto_rec_Y"
)){
  if(n=="dev_tknto_rec_Y")
  {
    print(par(mfrow=c(2,5)))
    print(plot(effects::predictorEffect(n,backward.model),
               lines=list(multiline=TRUE)))
      break()
  }
  #print(n)
  print(plot(effects::predictorEffect(n,backward.model)))
}
```

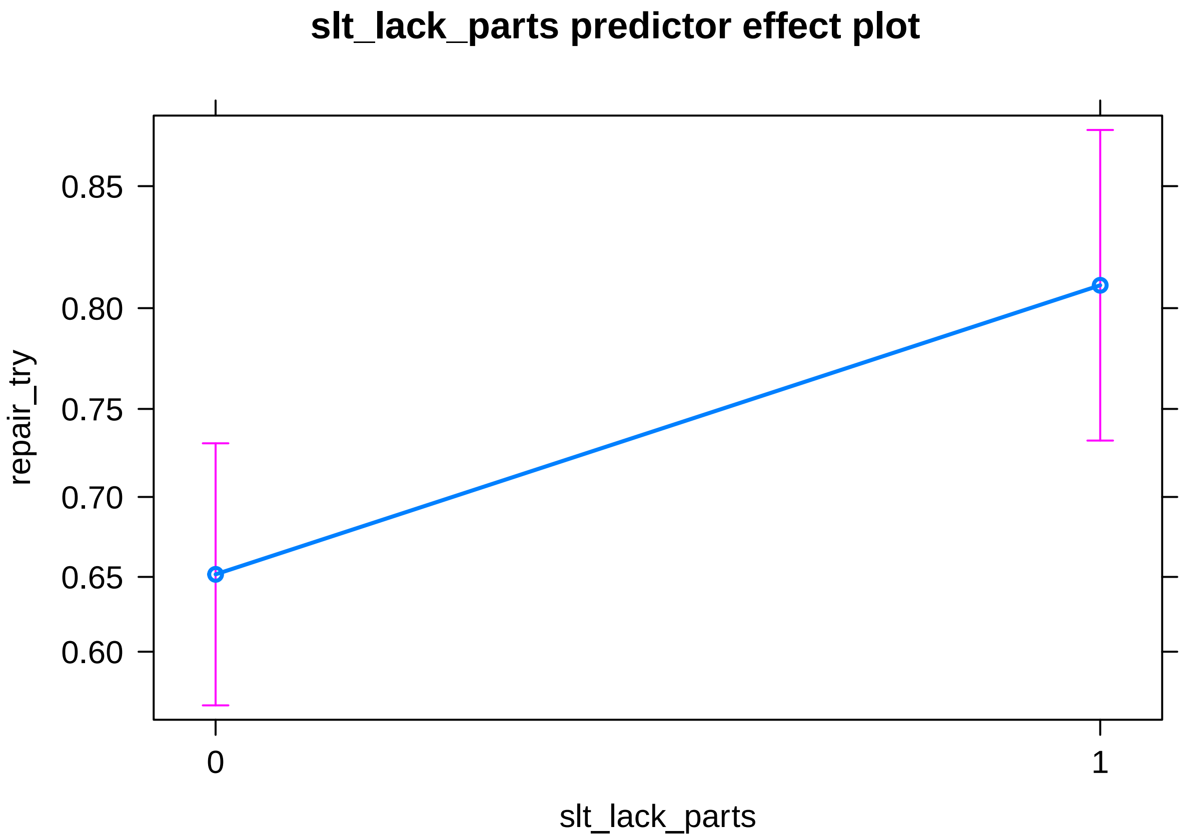


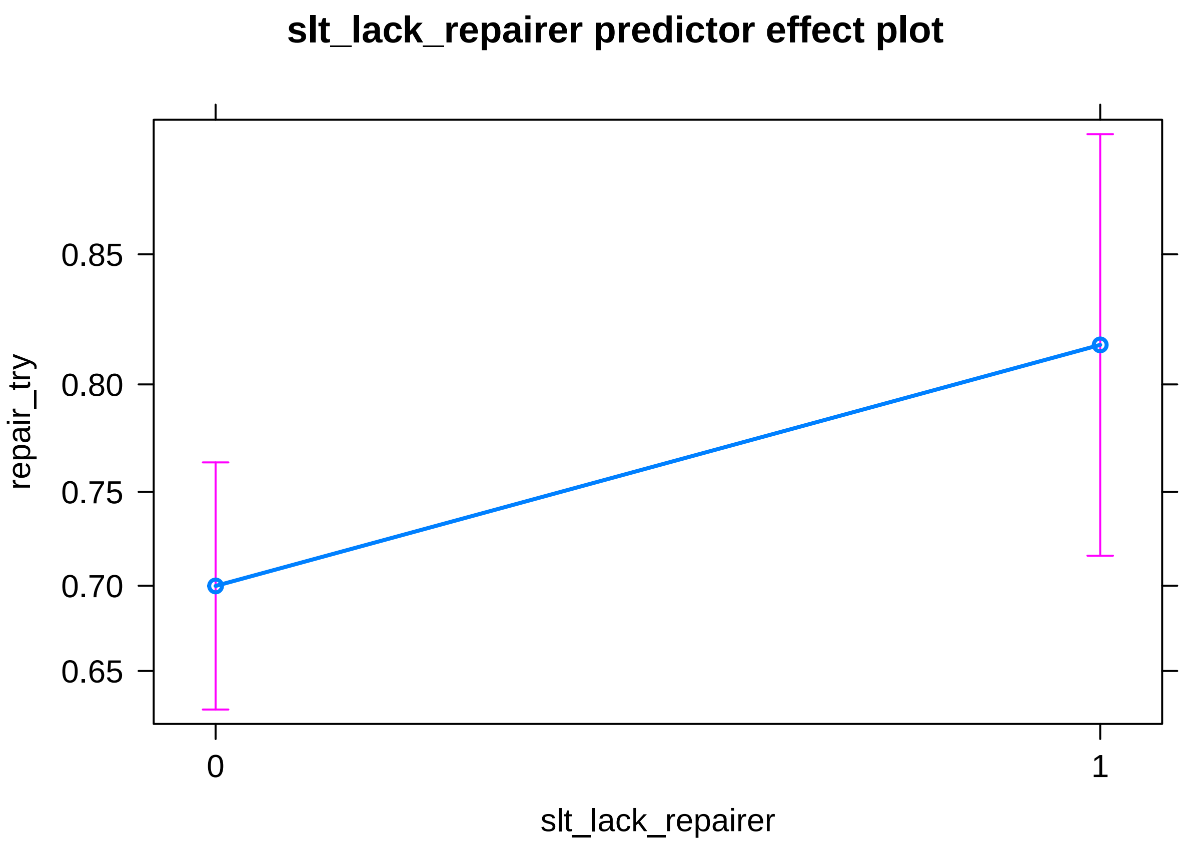

device_count_5_yr predictor effect plot
dev_tknto_rec_Y = 1
rprd_usage_chlng_fault = 0
dev_tknto_rec_Y = 1
rprd_usage_chlng_fault = 1
dev_tknto_rec_Y = 0
rprd_usage_chlng_fault = 0
dev_tknto_rec_Y = 0
rprd_usage_chlng_fault = 1
repair_try
0.2
0.4
0.6
0.8
0
2
4
6
8
10
12
14
device_count_5_yr

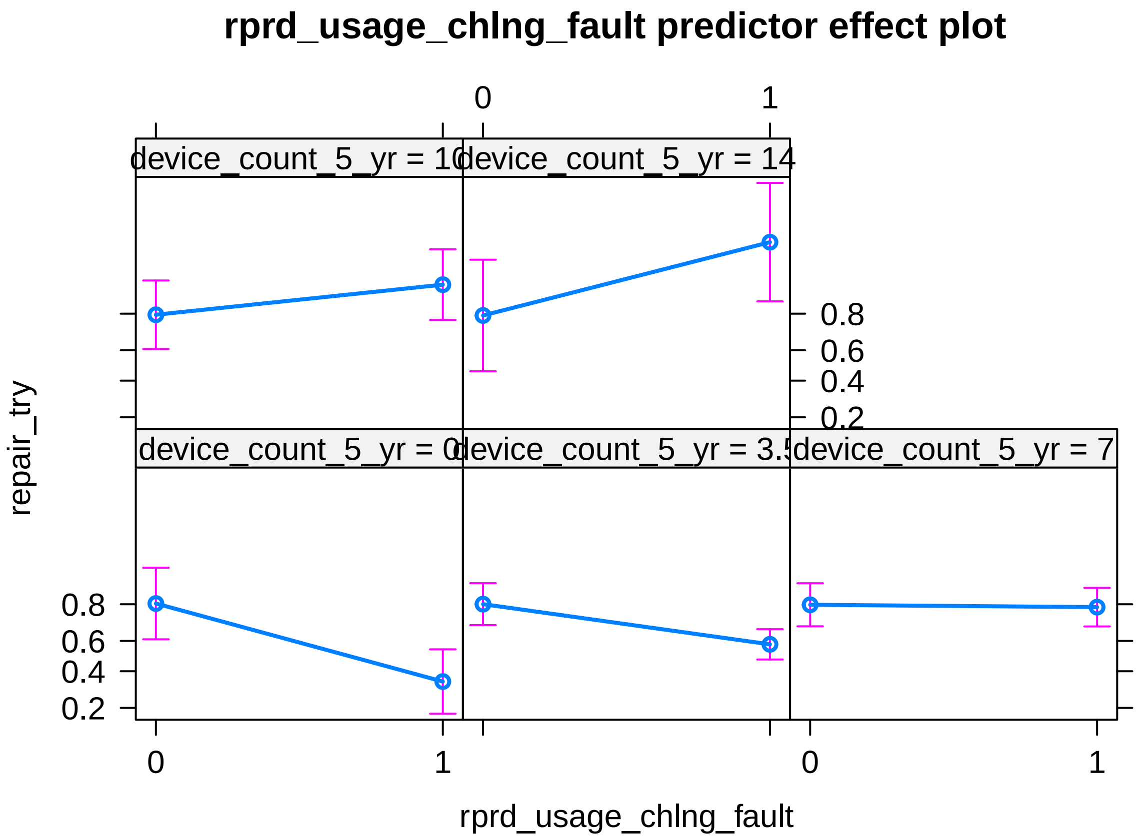

dump_reason_break predictor effect plot
repair_try
0.65
0.70
0.75
0.80
0
1
dump_reason_break

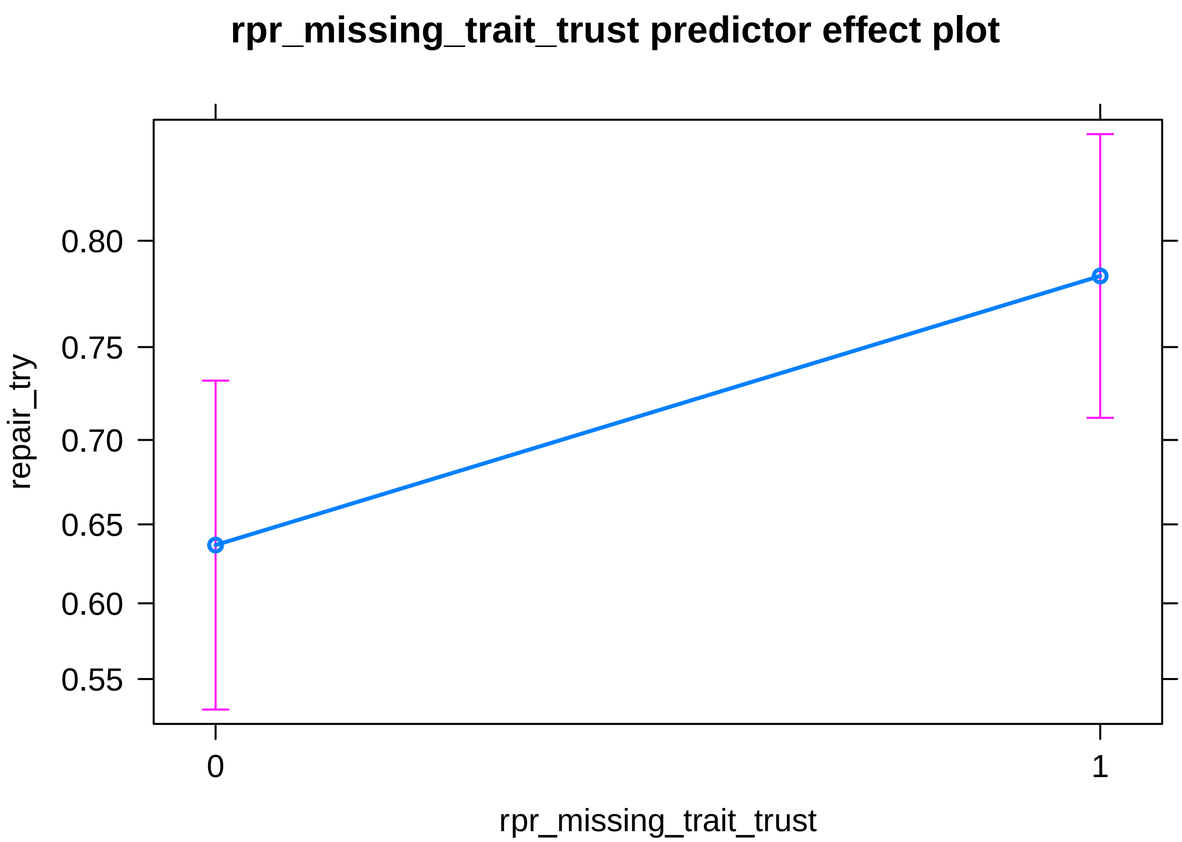

slt_lack_tu predictor effect plot
dev_tknto_rec_Y = 0
dev_tknto_rec_Y = 1
repair_try
0.9
0.8
0.7
0.6
0
1
slt_lack_tu

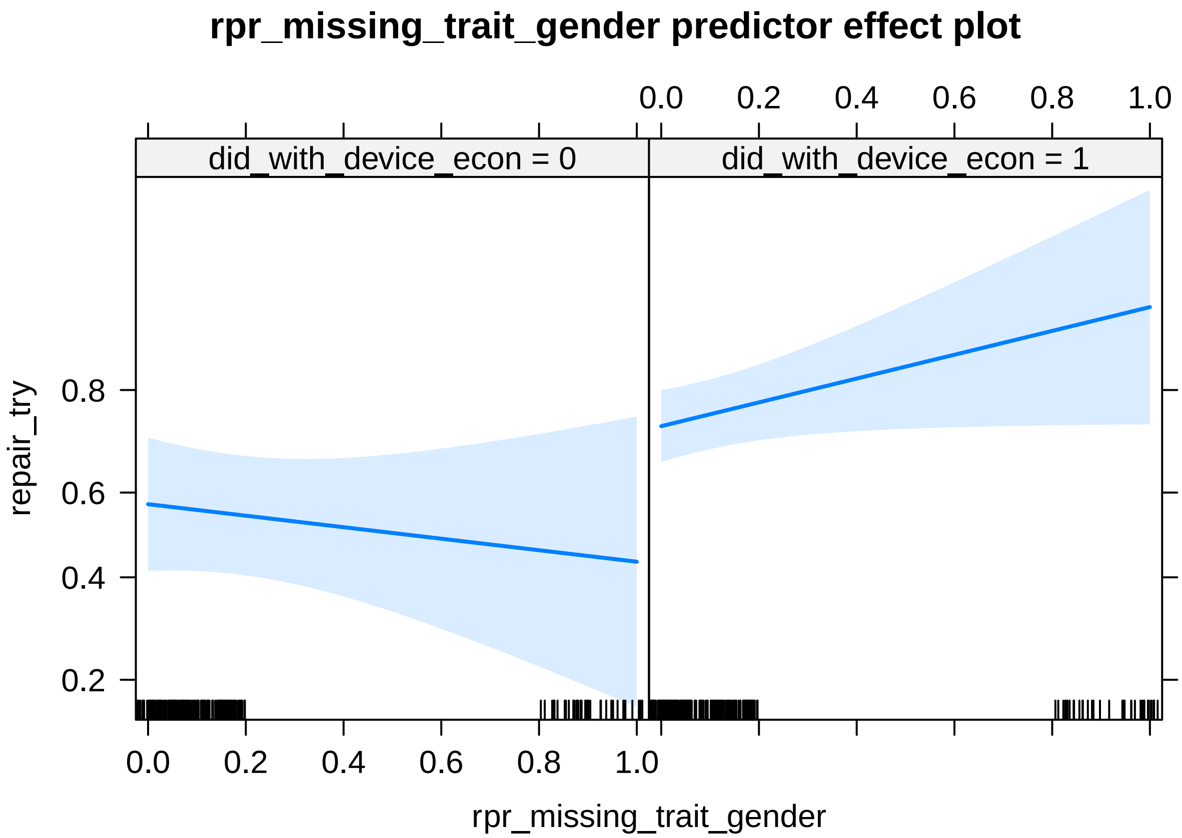

slt_lack_lang predictor effect plot
repair_try
0.90
0.85
0.80
0.75
0.70
0.65
0
1
slt_lack_lang

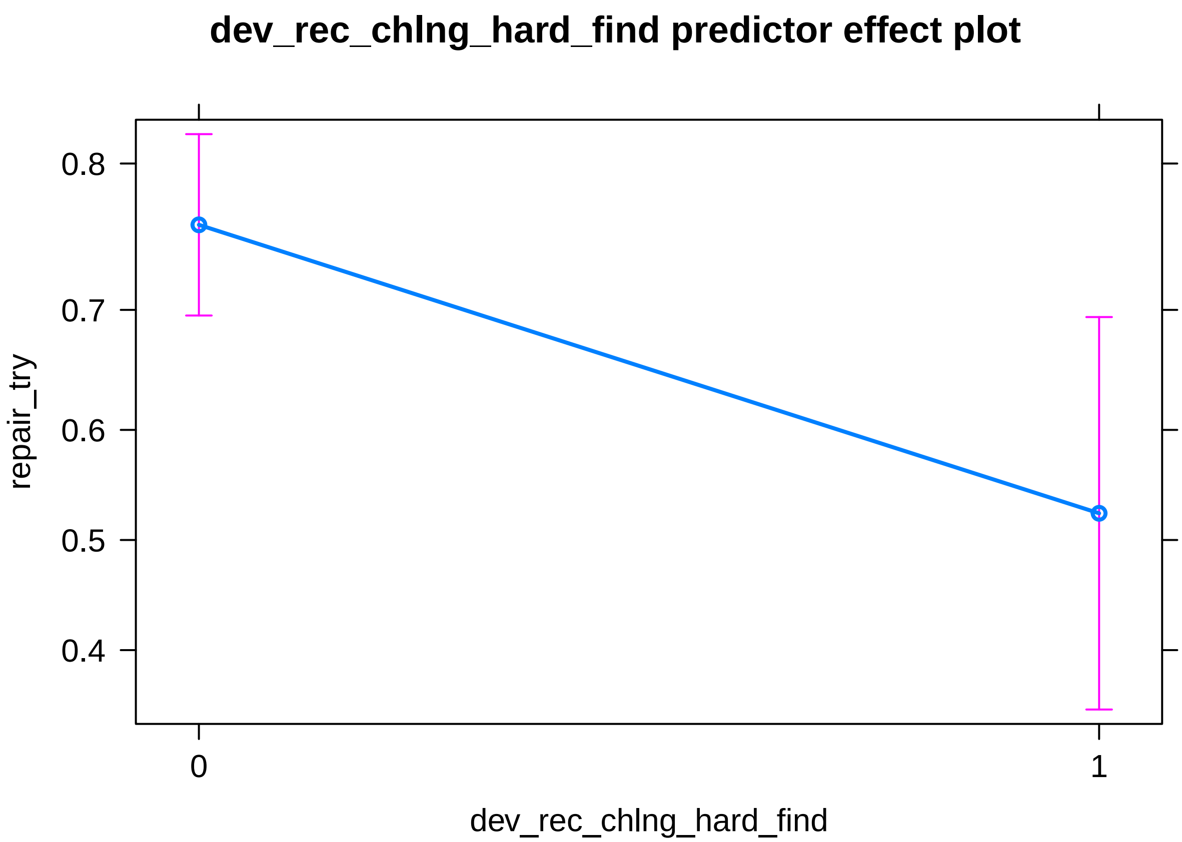

slt_lack_parts predictor effect plot
repair_try
0.85
0.80
0.75
0.70
0.65
0.60
0
1
slt_lack_parts

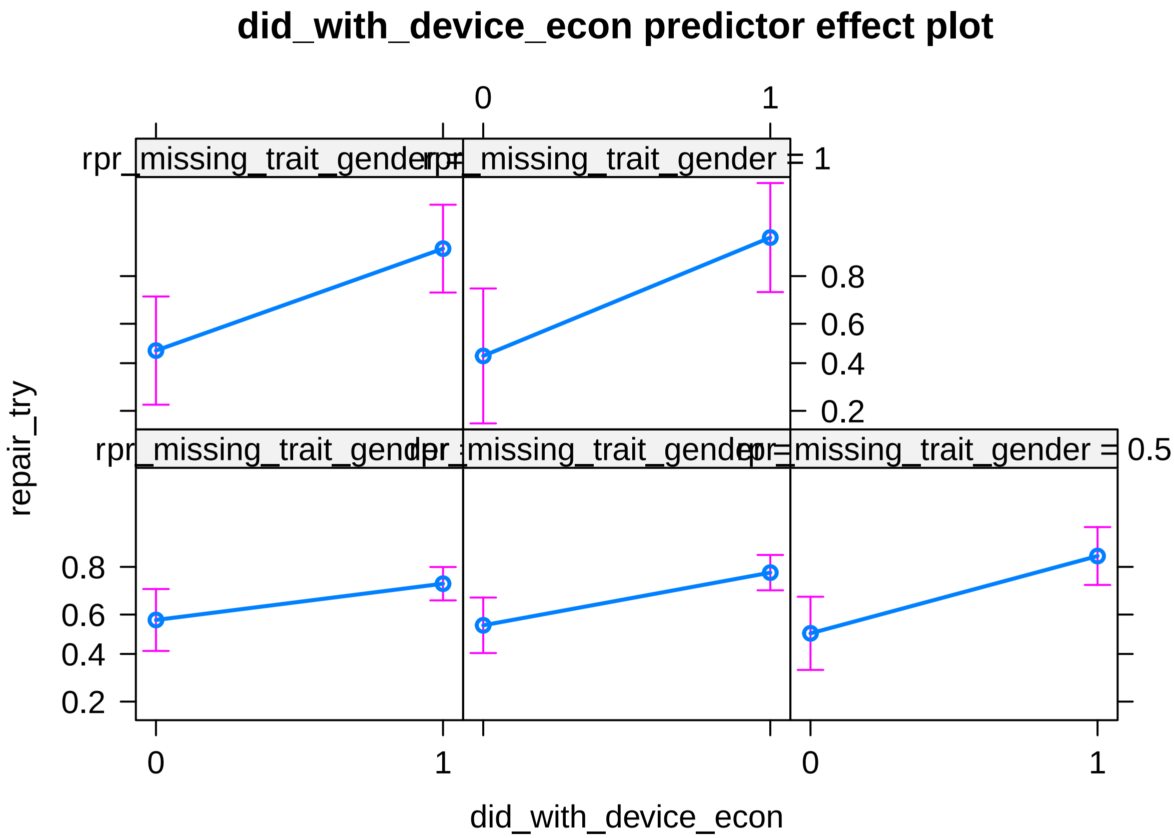

slt_lack_repairer predictor effect plot
repair_try
0.85
0.80
0.75
0.70
0.65
0
1
slt_lack_repairer


rprd_usage_chlng_fault predictor effect plot
device_count_5_yr = 10
device_count_5_yr = 14
device_count_5_yr = 0
device_count_5_yr = 3.5
device_count_5_yr = 7
0.8
0.6
0.4
0.2
0
1
repair_try
rprd_usage_chlng_fault

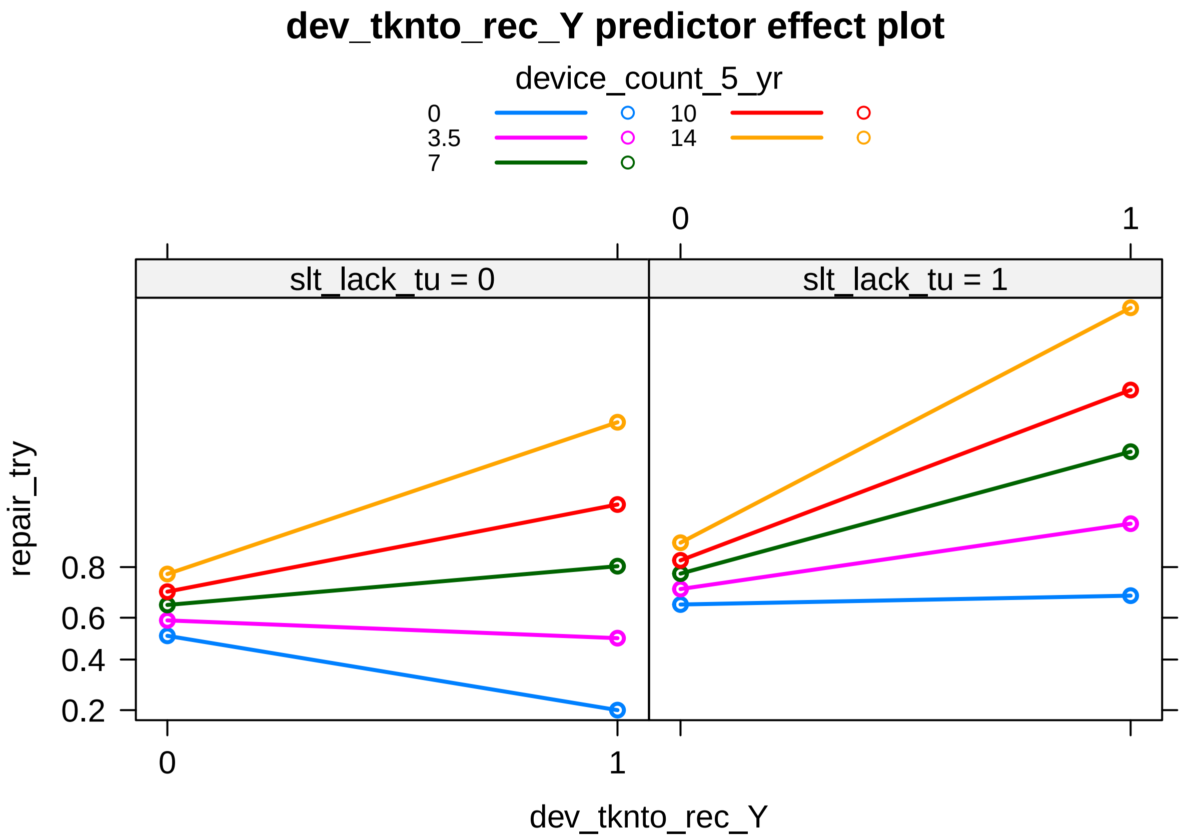

rpr_missing_trait_trust predictor effect plot
0.80
0.75
0.70
0.65
0.60
0.55
0
1
repair_try
rpr_missing_trait_trust


rpr_missing_trait_gender predictor effect plot
0.0
0.2
0.4
0.6
0.8
1.0
did_with_device_econ = 0
did_with_device_econ = 1
repair_try
0.8
0.6
0.4
0.2
rpr_missing_trait_gender


dev_rec_chlng_hard_find predictor effect plot
repair_try
0.8
0.7
0.6
0.5
0.4
0
1
dev_rec_chlng_hard_find


did_with_device_econ predictor effect plot
rpr_missing_trait_gender = 1
0.8
0.6
0.4
0.2
0
1
repair_try
did_with_device_econ
rpr_missing_trait_gender = 0.5

```
## $mfrow
## [1] 1 1
```


dev_tknto_rec_Y predictor effect plot
device_count_5_yr
0
3.5
7
10
14
slt_lack_tu = 0
slt_lack_tu = 1
repair_try
0.8
0.6
0.4
0.2
0
1
dev_tknto_rec_Y

Chapter 5

Necessary Documents and Papers.

5.1 Documents Prepared:

5.1.1 Initial Summary Doc I prepared to understand the survey results. I find it helpful to get a good grasp of our data.

5.1.2 Possible Directions for papers, I have figured out for this survey data

5.2 Paper - We can build on:

This Paper by *Sara Behdad* is an ideation of the line, we are pushing forward our work in.] Discusses whether customers whether 1)Buy device from same brand, 2) Recommend others, based on demograhy, repairability and other predictors.

RQ1: "If you successfully repaired a product, are you more likely to buy new products from the same company in the future?"

RQ2: "Have your experiences fixing your own products impacted the purchasing recommendations you give to your friends?"

5.3 Other Helpful Papers:

5.3.1 Paper 2

An investigation of used electronics return flows: A data-driven approach to capture and predict consumers storage and utilization behavior: Paper Link

Journal : Waste Management Link

CiteScore: 6.15 ; Impact Factor: 5.431

Summary: They analyzed the return record of ~10K Hard Disk drives which were given back to a remanufacturing facility located in Chicago, IL, USA during 2011–2013 . The main reason for choosing HDD is having access to the life cycle characteristics such as manufacturing year and the last time that the computer was used.

Analysis mostly deals with how long disks have been stored after last use, their yearly variance, how these characteristic (usage, life span, return ratio etc) vary from brand to brand. Two usage categories: household and corporate. How these two type impact usage and return and whether the return of devices can be predicted using ML models from these data. So many redundant graphs for manufacturer-wise HDD capacity, number, etc.

5.3.2 Paper 3

The current status of the consumer electronics repair industry in the U.S.: A survey based study: Paper Link

Journal : Resources, Conservation and Recycling

CiteScore: 6.82 ; Impact Factor: 7.044

Summary: a look into the repair industry through an analysis of a survey conducted by a third-party repair service provider. 2170 repair technicians have participated in a survey consisting of 23 questions about repair challenges in their profession. Also, a demand-based repair service pricing framework is introduced. The optimal pricing levels are found deduced from consumers' repair demand. Finally, other aspects of repair businesses, e.g. repairability degree of consumer electronics and consumer expectations of repair services, are thoroughly investigated to improve the formula.

US only study: survey of 2170 repair workers conducted by ifixit. Also, The most common products that repair businesses could not repair them together with the reasons. Clustering of consumer electronics based on difficulties in repair processes. Consumers' satisfaction of repair services from the perspective of repair businesses. 9. Business lessons learned by repair service providers

5.3.3 Paper 4

Consumer decisions to repair mobile phones and manufacturer pricing policies: The concept of value leakage Paper Link

Journal : Resources, Conservation and Recycling

CiteScore: 6.82 ; Impact Factor: 7.044

Summary: A group of 208 mobile phone users has been surveyed to capture consumer's time-dependent willingness-to-pay for repair services. **The user group is student. Because they are less inclined to repair (!). The sample size is smaller than ours**

The findings is biased with the point that the sample of consumers who participated in the survey mainly used big brands of cellphones. On the other hand, instead of promoting repair services, big brands adopt other marketing strategies (i.e., offering trade-in rebate) to foster brand loyalty, thereby decreasing phones longevity.

the repair cost is not the only factor that affects consumers' willingnessto-pay for repair services. The impact of other socio-demographic factors such as income and education level should be studied to better explain attitudes.

5.3.4 Emotional Attachment with Device :

"Consumer's emotional attachment to the current owned product affects their propensity to repair it (Page, 2014). This inclination is correlated with Eco-conscientious (Page, 2014), personal lifestyle traits [e.g. frugality (Bayus, 1991), product-retention tendency (Haws et al., 2012), product-care attentivity (Boyd and McCCNOCHA, 1996), and repairable products shopping tendency (Spack et al., 2012)], and finally socio-demographics factors such as age (McCollough, 2010), income (Bayus, 1991), education (Bayus, 1991), and gender (Hills and Worthing, 2006)."

5.3.5 Paper 6

Mining consumer experiences of repairing electronics: Product design insights and business lessons learned Paper Link

Journal of Cleaner Production; (2016)
DOI:10.1016/j.jclepro.2016.07.144

Understanding the factors contributing to unprofessional repair practices is a necessity. They investigated 4210 break and fix narratives reported by consumers of electronic devices in a survey conducted by iFixit.com. Regression analyses have been employed to examine the possible links between consumer experiences of repairing electronics and their future purchase behaviors. determine the most frequent products failed, component failure causes and the most common repair practices. a dataset of consumer stories have been used to connect previous repair experiences to future product purchase decisions in order to clarify the business outcomes of product repairability for manufactures.

Question asked in IFIXIT: What is your craziest fix story? (A crazy break story is OK as well)

(1) Have your experiences fixing your own products impacted the purchasing recommendations you give to your friends?

(2) If you successfully repaired a product, are you more likely to buy new products from the same company in the future?

Keyword: Text Mining.

5.3.6 Paper 7

Managing consumer behavior toward on-time return of the waste electrical and electronic equipment: A game theoretic approach Paper Link

Journal : International Journal of Production Economics CiteScore: 7.13 ; Impact Factor: 4.998

Summary: The consumers' decision about when to return the End-of-Use/Life (EoU/L) products and manufacturer's decision for the amount of incentive offered to consumers are incorporated into a theoretic game framework. Not quite related to our analysis.

5.3.7 A list of Journals similar to where Sara Publishes

Chapter 6

Final Words