

Student Level Analysis: SI on One Year Retention

Faith Fatchen

4/22/2022

Initial Model

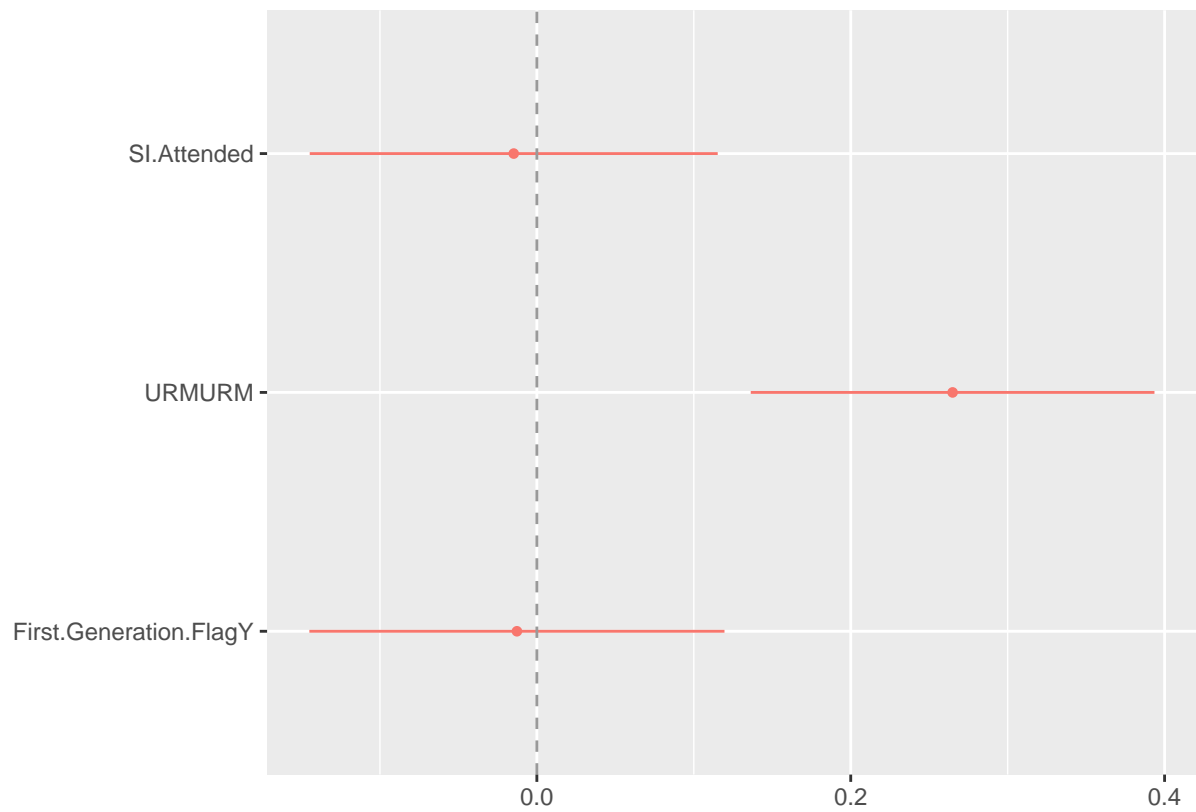
```
fit1<-glm(One.Year.Retention ~ Cohort.Term.Year + SI.Attended + Student.Class.Unit.Attempted + Student.  
summary(fit1)
```

```
##  
## Call:  
## glm(formula = One.Year.Retention ~ Cohort.Term.Year + SI.Attended +  
##      Student.Class.Unit.Attempted + Student.Class.Unit.Passed +  
##      URM + Gender.Code + First.Generation.Flag + Academic.Level +  
##      Major.1.STEM.Flag + Academic.Standing.Status + HS.GPA + Inst.MD.Persn.Onl.Othr +  
##      Course.Fee.Exist.Flag + GE.Class.Flag, family = binomial(link = "logit"),  
##      data = d)  
##  
## Deviance Residuals:  
##      Min        1Q    Median        3Q        Max  
## -3.2096   0.1647   0.2472   0.3849   3.7439  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept)      -1.106870    0.665611  -1.663   0.0963  
## Cohort.Term.Year2017    0.202192    0.102136   1.980   0.0477  
## Cohort.Term.Year2018   -0.125127    0.102458  -1.221   0.2220  
## Cohort.Term.Year2019   -0.579090    0.101165  -5.724 1.04e-08  
## Cohort.Term.Year2020   -1.082191    0.134345  -8.055 7.93e-16  
## Cohort.Term.Year2021   -8.138094    0.518192 -15.705 < 2e-16  
## SI.Attended          -0.014749    0.066320  -0.222   0.8240  
## Student.Class.Unit.Attempted -0.076773    0.173902  -0.441   0.6589  
## Student.Class.Unit.Passed    0.210638    0.019014  11.078 < 2e-16  
## URMURM                0.264928    0.065635   4.036 5.43e-05  
## Gender.CodeM          -0.045478    0.066403  -0.685   0.4934  
## Gender.CodeN          -5.734425   196.968366  -0.029   0.9768  
## First.Generation.FlagY   -0.012717    0.067485  -0.188   0.8505  
## Academic.LevelJunior    1.394835    0.156345   8.922 < 2e-16  
## Academic.LevelSenior    1.046228    0.202841   5.158 2.50e-07  
## Academic.LevelSophomore  1.539264    0.107162  14.364 < 2e-16  
## Major.1.STEM.FlagY     -0.336792    0.068999  -4.881 1.05e-06  
## Academic.Standing.StatusGood Standing  3.668284    0.317591  11.550 < 2e-16  
## Academic.Standing.StatusNo Value    3.863948    0.261376  14.783 < 2e-16  
## Academic.Standing.StatusProbation    2.728098    0.263216  10.364 < 2e-16  
## HS.GPA               -0.009804    0.080922  -0.121   0.9036  
## Inst.MD.Persn.Onl.OthrOnline    0.842629    0.099742   8.448 < 2e-16
```

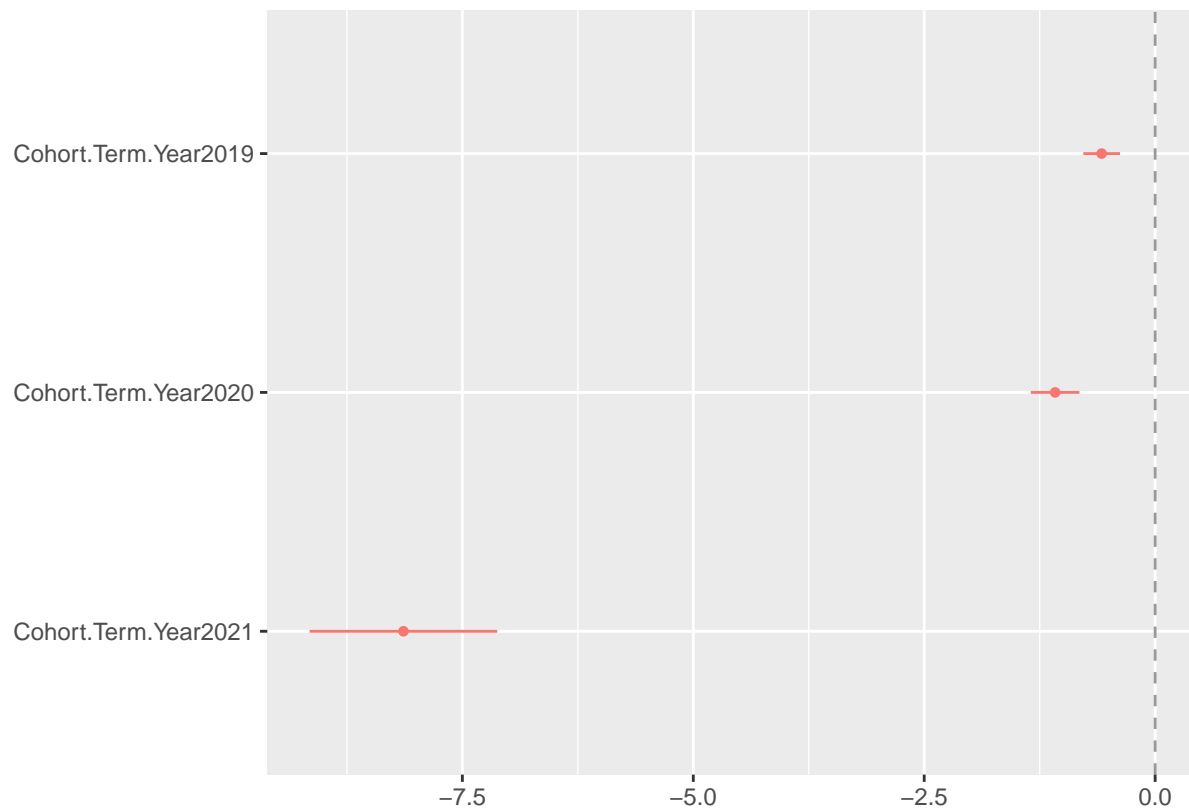
```

## Course.Fee.Exist.FlagY          -0.024791  0.162599 -0.152  0.8788
## GE.Class.FlagY                  -0.574643  0.087451 -6.571  5.00e-11
##
## (Intercept)                      .
## Cohort.Term.Year2017             *
## Cohort.Term.Year2018
## Cohort.Term.Year2019             ***
## Cohort.Term.Year2020             ***
## Cohort.Term.Year2021             ***
## SI.Attended
## Student.Class.Unit.Attempted
## Student.Class.Unit.Passed        ***
## URMURM                           ***
## Gender.CodeM
## Gender.CodeN
## First.Generation.FlagY
## Academic.LevelJunior             ***
## Academic.LevelSenior             ***
## Academic.LevelSophomore          ***
## Major.1.STEM.FlagY               ***
## Academic.Standing.StatusGood Standing ***
## Academic.Standing.StatusNo Value ***
## Academic.Standing.StatusProbation ***
## HS.GPA
## Inst.MD.Persn.Onl.OthrOnline      ***
## Course.Fee.Exist.FlagY
## GE.Class.FlagY                   ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 12190.6  on 19651  degrees of freedom
## Residual deviance:  7978.9  on 19628  degrees of freedom
##    (4 observations deleted due to missingness)
## AIC: 8026.9
##
## Number of Fisher Scoring iterations: 10
m1<-tidy(fit1) %>%
  filter(term == "SI.Attended" | term == "URMURM" | term == "First.Generation.FlagY")
dwplot(m1) +
  geom_vline(xintercept = 0, colour = "grey60", linetype = 2)

```



```
m2<-tidy(fit1) %>%
  filter(term == "Cohort.Term.Year2019" | term == "Cohort.Term.Year2020" | term == "Cohort.Term.Year2021")
dwplot(m2) +
  geom_vline(xintercept = 0, colour = "grey60", linetype = 2)
```



```
fit2<-glm(One.Year.Retention ~ Cohort.Term.Year + SI.Visit.Num + Student.Class.Unit.Attempted + Student
summary(fit2)
```

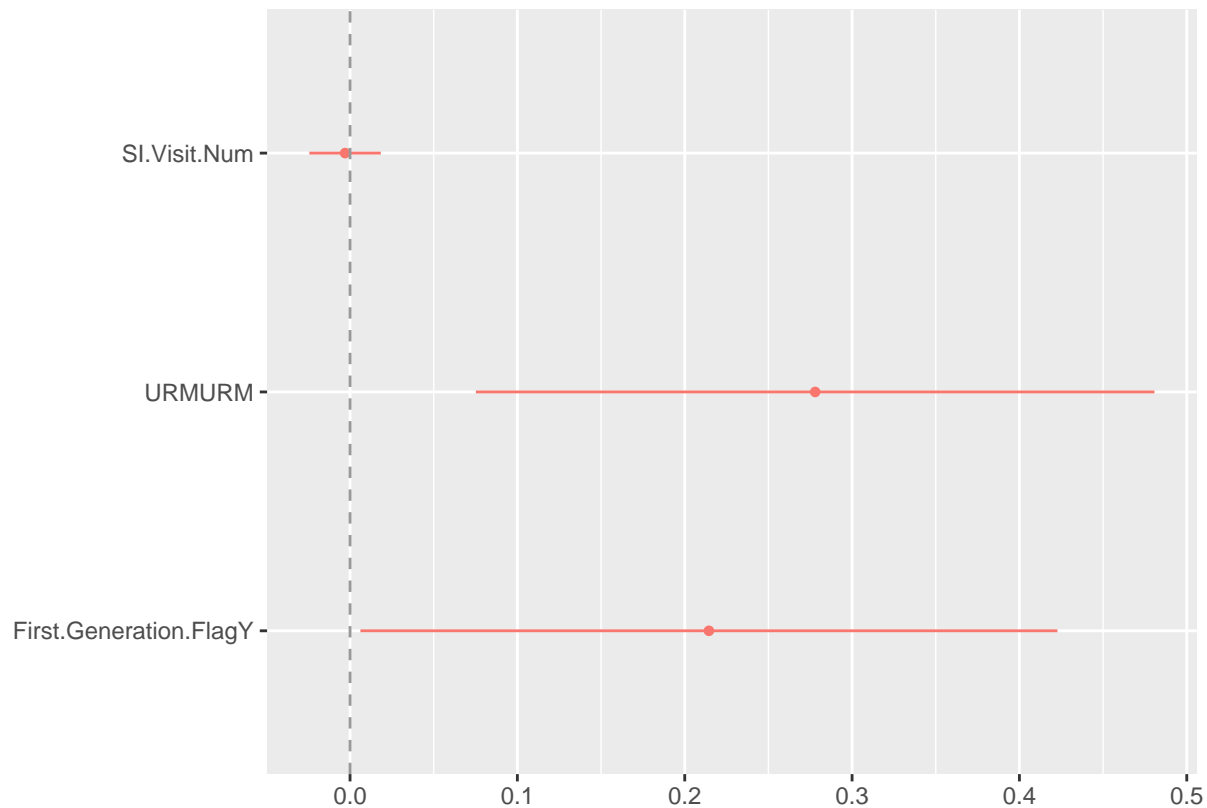
```
##
## Call:
## glm(formula = One.Year.Retention ~ Cohort.Term.Year + SI.Visit.Num +
##       Student.Class.Unit.Attempted + Student.Class.Unit.Passed +
##       URM + Gender.Code + First.Generation.Flag + Academic.Level +
##       Major.1.STEM.Flag + Academic.Standing.Status + HS.GPA + Inst.MD.Persn.Onl.Othr +
##       Course.Fee.Exist.Flag + GE.Class.Flag, family = binomial(link = "logit"),
##       data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3327   0.1690   0.2482   0.3538   3.3165
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.946786    1.152833  -1.689 0.091277
## Cohort.Term.Year2017    0.144708    0.152336   0.950 0.342151
## Cohort.Term.Year2018   -0.142839    0.155619  -0.918 0.358685
## Cohort.Term.Year2019   -0.882887    0.149155  -5.919 3.23e-09
## Cohort.Term.Year2020   -1.566676    0.248552  -6.303 2.92e-10
## Cohort.Term.Year2021   -7.935422    1.028119  -7.718 1.18e-14
## SI.Visit.Num    -0.002956    0.010892  -0.271 0.786054
## Student.Class.Unit.Attempted    0.186617    0.327932   0.569 0.569307
## Student.Class.Unit.Passed    0.210054    0.033391   6.291 3.16e-10
## URMURM          0.277953    0.103451   2.687 0.007214
```

```

## Gender.CodeM -0.065493 0.106670 -0.614 0.539232
## Gender.CodeN -6.132622 196.970353 -0.031 0.975162
## First.Generation.FlagY 0.214455 0.106252 2.018 0.043554
## Academic.LevelJunior 1.253514 0.214619 5.841 5.20e-09
## Academic.LevelSenior 1.086190 0.291797 3.722 0.000197
## Academic.LevelSophomore 1.378521 0.149916 9.195 < 2e-16
## Major.1.STEM.FlagY -0.236876 0.111328 -2.128 0.033359
## Academic.Standing.StatusGood Standing 3.873658 0.547184 7.079 1.45e-12
## Academic.Standing.StatusNo Value 3.696305 0.389802 9.483 < 2e-16
## Academic.Standing.StatusProbation 2.598705 0.396364 6.556 5.51e-11
## HS.GPA -0.006194 0.123217 -0.050 0.959907
## Inst.MD.Persn.Onl.OthrOnline 1.033565 0.204884 5.045 4.54e-07
## Course.Fee.Exist.FlagY -0.134589 0.310785 -0.433 0.664971
## GE.Class.FlagY -0.437617 0.131639 -3.324 0.000886
##
## (Intercept) .
## Cohort.Term.Year2017
## Cohort.Term.Year2018
## Cohort.Term.Year2019 ***
## Cohort.Term.Year2020 ***
## Cohort.Term.Year2021 ***
## SI.Visit.Num
## Student.Class.Unit.Attempted
## Student.Class.Unit.Passed ***
## URMURM **
## Gender.CodeM
## Gender.CodeN
## First.Generation.FlagY *
## Academic.LevelJunior ***
## Academic.LevelSenior ***
## Academic.LevelSophomore ***
## Major.1.STEM.FlagY *
## Academic.Standing.StatusGood Standing ***
## Academic.Standing.StatusNo Value ***
## Academic.Standing.StatusProbation ***
## HS.GPA
## Inst.MD.Persn.Onl.OthrOnline ***
## Course.Fee.Exist.FlagY
## GE.Class.FlagY ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 4486.1 on 9216 degrees of freedom
## Residual deviance: 3433.5 on 9193 degrees of freedom
## (10439 observations deleted due to missingness)
## AIC: 3481.5
##
## Number of Fisher Scoring iterations: 10
m3<-tidy(fit2) %>%
  filter(term == "SI.Visit.Num" | term == "URMURM" | term == "First.Generation.FlagY")
dwplot(m3) +

```

```
geom_vline(xintercept = 0, colour = "grey60", linetype = 2)
```



With some interactions

```
fit3<-glm(One.Year.Retention~ Cohort.Term.Year + SI.Attended + Student.Class.Unit.Attempted + Student.C
summary(fit3)
```

```
##
## Call:
## glm(formula = One.Year.Retention ~ Cohort.Term.Year + SI.Attended +
##     Student.Class.Unit.Attempted + Student.Class.Unit.Passed +
##     URM + Gender.Code + First.Generation.Flag + Academic.Level +
##     Major.1.STEM.Flag + Academic.Standing.Status + HS.GPA + Inst.MD.Persn.Onl.Othr +
##     Course.Fee.Exist.Flag + GE.Class.Flag + SI.Attended:URM +
##     SI.Attended:First.Generation.Flag, family = binomial(link = "logit"),
##     data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2493   0.1631   0.2456   0.3842   3.7673
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.14095    0.66626  -1.712  0.08681 .
## Cohort.Term.Year2017    0.20389    0.10221   1.995  0.04606 *
## Cohort.Term.Year2018   -0.11871    0.10272  -1.156  0.24781
## Cohort.Term.Year2019   -0.57593    0.10127  -5.687 1.29e-08 ***
```

```
## Cohort.Term.Year2020          -1.06601    0.13460   -7.920 2.38e-15 ***
## Cohort.Term.Year2021          -8.14833    0.51843  -15.717 < 2e-16 ***
## SI.Attended                   -0.15477    0.09564   -1.618 0.10560
## Student.Class.Unit.Attempted  -0.05328    0.17422   -0.306 0.75976
## Student.Class.Unit.Passed      0.21078    0.01903   11.076 < 2e-16 ***
## URMURM                        0.25277    0.08485    2.979 0.00289 **
## Gender.CodeM                  -0.04796    0.06641   -0.722 0.47024
## Gender.CodeN                  -5.64142   196.96837  -0.029 0.97715
## First.Generation.FlagY         -0.14650    0.08733   -1.677 0.09345 .
## Academic.LevelJunior           1.40094    0.15650    8.952 < 2e-16 ***
## Academic.LevelSenior           1.04997    0.20287    5.176 2.27e-07 ***
## Academic.LevelSophomore        1.54241    0.10719   14.390 < 2e-16 ***
## Major.1.STEM.FlagY            -0.33952    0.06906   -4.916 8.82e-07 ***
## Academic.Standing.StatusGood Standing  3.66507    0.31777   11.534 < 2e-16 ***
## Academic.Standing.StatusNo Value  3.86666    0.26149   14.787 < 2e-16 ***
## Academic.Standing.StatusProbation  2.72911    0.26337   10.362 < 2e-16 ***
## HS.GPA                        -0.01139    0.08110   -0.140 0.88834
## Inst.MD.Persn.Onl.OthrOnline    0.84563    0.09977    8.476 < 2e-16 ***
## Course.Fee.Exist.FlagY        -0.02346    0.16273   -0.144 0.88537
## GE.Class.FlagY                -0.57350    0.08743   -6.559 5.40e-11 ***
## SI.Attended:URMURM             0.03309    0.13045    0.254 0.79974
## SI.Attended:First.Generation.FlagY  0.31633    0.13347    2.370 0.01779 *
```

```
## ---
```

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

```
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
```

```
## Null deviance: 12190.6 on 19651 degrees of freedom
```

```
## Residual deviance: 7972.5 on 19626 degrees of freedom
```

```
## (4 observations deleted due to missingness)
```

```
## AIC: 8024.5
```

```
##
```

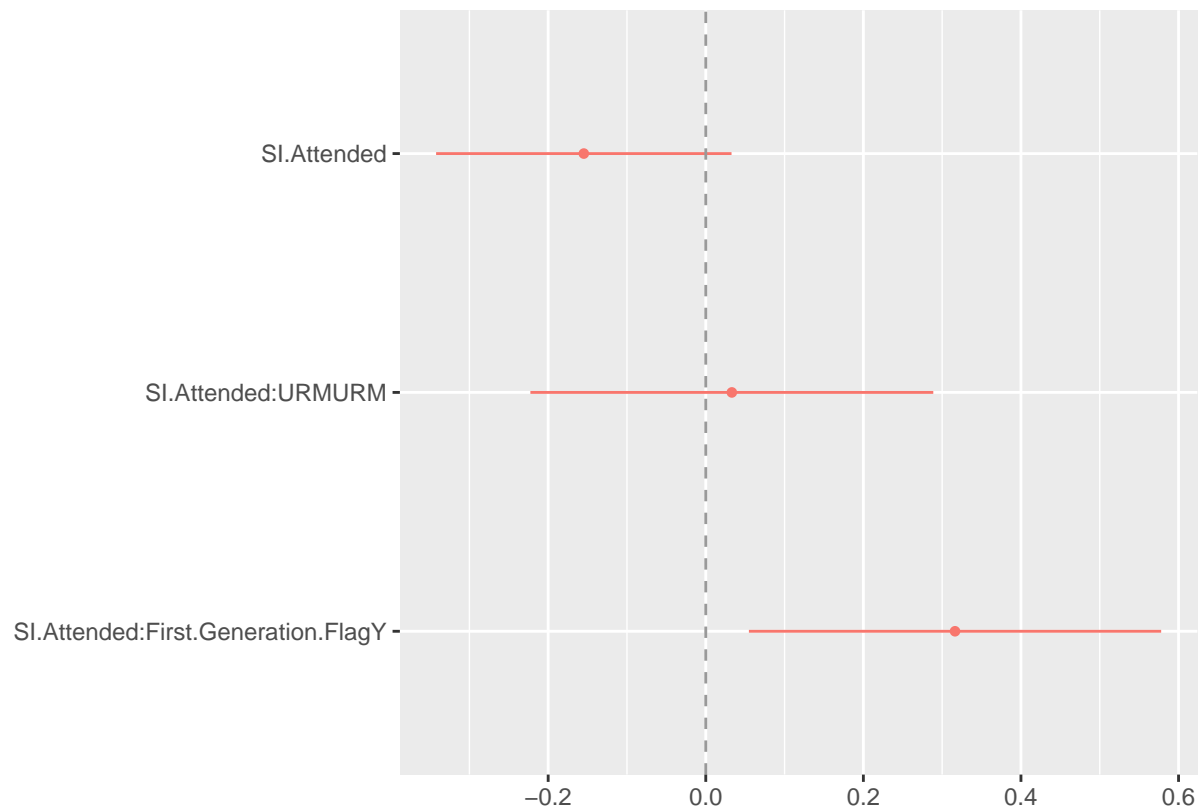
```
## Number of Fisher Scoring iterations: 10
```

```
m4<-tidy(fit3)%>%
```

```
  filter(term == "SI.Attended" | term == "SI.Attended:URMURM" | term == "SI.Attended:First.Generation.F")
```

```
dwplot(m4) +
```

```
  geom_vline(xintercept = 0, colour = "grey60", linetype = 2)
```



```
fit4<-glm(One.Year.Retention~ Cohort.Term.Year + SI.Visit.Num + Student.Class.Unit.Attempted + Student.
summary(fit4)
```

```
##
## Call:
## glm(formula = One.Year.Retention ~ Cohort.Term.Year + SI.Visit.Num +
##       Student.Class.Unit.Attempted + Student.Class.Unit.Passed +
##       URM + Gender.Code + First.Generation.Flag + Academic.Level +
##       Major.1.STEM.Flag + Academic.Standing.Status + HS.GPA + Inst.MD.Persn.Onl.Othr +
##       Course.Fee.Exist.Flag + GE.Class.Flag + SI.Visit.Num:URM +
##       SI.Visit.Num:First.Generation.Flag, family = binomial(link = "logit"),
##       data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3411   0.1688   0.2472   0.3526   3.3263
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.971840    1.153625  -1.709  0.087404
## Cohort.Term.Year2017    0.146517    0.152379   0.962  0.336285
## Cohort.Term.Year2018   -0.143731    0.155661  -0.923  0.355821
## Cohort.Term.Year2019   -0.883849    0.149493 -5.912 3.37e-09
## Cohort.Term.Year2020   -1.564661    0.249037 -6.283 3.32e-10
## Cohort.Term.Year2021   -7.937781    1.028416 -7.718 1.18e-14
## SI.Visit.Num      0.005315    0.016901   0.314  0.753146
## Student.Class.Unit.Attempted    0.180232    0.328223   0.549  0.582926
## Student.Class.Unit.Passed    0.210057    0.033389   6.291 3.15e-10
```



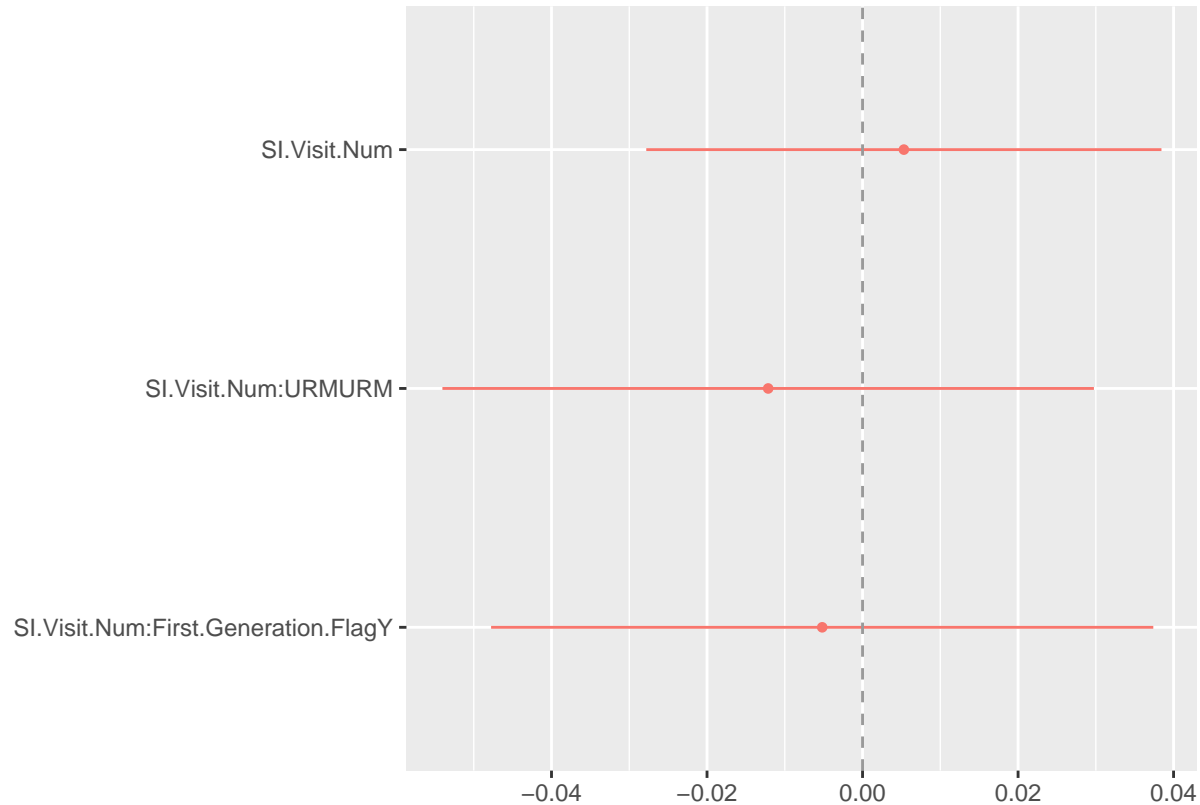
```

## URMURM                0.339549    0.151442    2.242 0.024954
## Gender.CodeM           -0.062670    0.106762   -0.587 0.557198
## Gender.CodeN          -6.097729  196.970363   -0.031 0.975303
## First.Generation.FlagY    0.240214    0.156840    1.532 0.125623
## Academic.LevelJunior     1.253815    0.214642    5.841 5.18e-09
## Academic.LevelSenior     1.083634    0.291821    3.713 0.000205
## Academic.LevelSophomore   1.378038    0.149936    9.191 < 2e-16
## Major.1.STEM.FlagY       -0.236797    0.111292   -2.128 0.033361
## Academic.Standing.StatusGood Standing  3.879565    0.546928    7.093 1.31e-12
## Academic.Standing.StatusNo Value      3.704094    0.389549    9.509 < 2e-16
## Academic.Standing.StatusProbation      2.604648    0.396090    6.576 4.84e-11
## HS.GPA                  -0.004808    0.123410   -0.039 0.968924
## Inst.MD.Persn.Onl.OthrOnline  1.032036    0.205261    5.028 4.96e-07
## Course.Fee.Exist.FlagY     -0.139299    0.310939   -0.448 0.654157
## GE.Class.FlagY            -0.435963    0.131673   -3.311 0.000930
## SI.Visit.Num:URMURM       -0.012137    0.021374   -0.568 0.570143
## SI.Visit.Num:First.Generation.FlagY    -0.005187    0.021725   -0.239 0.811272
##
## (Intercept)              .
## Cohort.Term.Year2017
## Cohort.Term.Year2018
## Cohort.Term.Year2019      ***
## Cohort.Term.Year2020      ***
## Cohort.Term.Year2021      ***
## SI.Visit.Num
## Student.Class.Unit.Attempted
## Student.Class.Unit.Passed      ***
## URMURM                      *
## Gender.CodeM
## Gender.CodeN
## First.Generation.FlagY
## Academic.LevelJunior      ***
## Academic.LevelSenior      ***
## Academic.LevelSophomore   ***
## Major.1.STEM.FlagY        *
## Academic.Standing.StatusGood Standing ***
## Academic.Standing.StatusNo Value      ***
## Academic.Standing.StatusProbation      ***
## HS.GPA
## Inst.MD.Persn.Onl.OthrOnline      ***
## Course.Fee.Exist.FlagY
## GE.Class.FlagY            ***
## SI.Visit.Num:URMURM
## SI.Visit.Num:First.Generation.FlagY
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 4486.1 on 9216 degrees of freedom
## Residual deviance: 3433.1 on 9191 degrees of freedom
## (10439 observations deleted due to missingness)
## AIC: 3485.1
##

```

```
## Number of Fisher Scoring iterations: 10
```

```
m5<-tidy(fit4)%>%
  filter(term == "SI.Visit.Num" | term == "SI.Visit.Num:URMURM" | term == "SI.Visit.Num:First.Generation.FlagY")
dwplot(m5) +
  geom_vline(xintercept = 0, colour = "grey60", linetype = 2)
```



DoD

```
dod_data <- d %>% group_by(URM, SI.Attended)
```

Modeling the Difference of Differences

```
dod_mod <- glm(One.Year.Retention ~ SI.Attended*URM, family = binomial(link = "logit"), data=dod_data)
coef(dod_mod)
```

```
##      (Intercept)      SI.Attended      URMURM SI.Attended:URMURM
##      2.02961992      0.58260076     -0.02749662      0.09998787
```

Writing contrasts

- b0 = Intercept
- b1 = SI
- b2 = URM
- b3 = SI * URM

First I define each of the 4 groups, and then calculate the gap within SI and for no SI.

```

# URM = No & SI = Yes
URM.no_SI.yes <- matrix(c(1, 1, 0, 0), 1)
# URM = Yes & SI = Yes
URM.yes_SI.yes <- matrix(c(1, 1, 1, 1), 1)
(URM.gap_SI.yes <- URM.no_SI.yes - URM.yes_SI.yes)

##      [,1] [,2] [,3] [,4]
## [1,]    0    0   -1   -1

# URM = No & SI = No
URM.no_SI.no <- matrix(c(1, 0, 0, 0), 1)
# URM = Yes & SI = No
URM.yes_SI.no <- matrix(c(1, 0, 1, 0), 1)
(URM.gap_SI.no <- URM.no_SI.no - URM.yes_SI.no)

##      [,1] [,2] [,3] [,4]
## [1,]    0    0   -1    0

difference of differences: URM gap for No SI - URM gap for SI
(K <- URM.gap_SI.no - URM.gap_SI.yes)

##      [,1] [,2] [,3] [,4]
## [1,]    0    0    0    1

mind.the.gap <- glht(dod_mod, linfct = K)
summary(mind.the.gap)

##
## Simultaneous Tests for General Linear Hypotheses
##
## Fit: glm(formula = One.Year.Retention ~ SI.Attended * URM, family = binomial(link = "logit"),
## data = dod_data)
##
## Linear Hypotheses:
## Estimate Std. Error z value Pr(>|z|)
## 1 == 0  0.09999    0.10385   0.963   0.336
## (Adjusted p values reported -- single-step method)

confint(mind.the.gap)

##
## Simultaneous Confidence Intervals
##
## Fit: glm(formula = One.Year.Retention ~ SI.Attended * URM, family = binomial(link = "logit"),
## data = dod_data)
##
## Quantile = 1.96
## 95% family-wise confidence level
##
## Linear Hypotheses:
## Estimate lwr      upr
## 1 == 0  0.09999 -0.10356  0.30353

```

The difference between URM yes and URM does not significantly differ between SI yes and SI no.

Plots

```
facet_names<-c(`0` = "Did Not Attend SI", `1` = "Attended SI")
ggplot(dod_data) +
  geom_bar(aes(x=One.Year.Retention, fill=URM),
    position = "dodge") +
  facet_wrap(~SI.Attended, labeller = as_labeller(facet_names)) + xlab("One Year Retention")+ scale_fill_
```

