EDA 141 final

Kyle A. Slaughter – 906148802

2025-05-21

SUMMER SESSIONS ENROLLMENT OF NON-UCLA COLLEGE STUDENTS

These are records for non-UCLA college-age enrolled students in a summer term. There is one record per student. Two characteristics are marked for the students (having an in-state address and taking all online courses). Otherwise, the data are counts of the courses they took in the term, by type (i.e. lower division count, online course count) and course counts by division oLering the courses – a proxy for course subject matter.

• What seems to predict or to characterize taking only online courses during summer term? • What kinds of courses are more popular for the students from outside of California? • What seems to be the greatest differences between students taking only one course and students taking two or more courses? • What advice could be given to Summer Sessions about how to increase course enrollments of these students, given what is found in this dataset?

```
library(readxl)
df <- read_excel("Mr Wahl data spring 25.xlsx")
df$californian <- as.logical(df$californian)
df$all_online <- as.logical(df$all_online)
```

```
head(df)
```

```
## # A tibble: 6 x 14
##   casenumber californian online_course_ct lower_division_course_ct arts_ct
##        <dbl> <lgl>                  <dbl>                     <dbl>   <dbl>
## 1          1 TRUE                       1                         1       0
## 2          2 TRUE                       0                         1       0
## 3          3 TRUE                       1                         1       0
## 4          4 TRUE                       0                         2       0
## 5          5 TRUE                       0                         0       0
## 6          6 TRUE                       1                         1       0
## # i 9 more variables: engineering_ct <dbl>, humanities_ugeduc_ct <dbl>,
## #   life_sci_ct <dbl>, phys_sci_ct <dbl>, soc_sci_intl_inst_ct <dbl>,
## #   mgmt_ct <dbl>, other_prof_ct <dbl>, count_of_courses_taken <dbl>,
## #   all_online <lgl>
```

```
for (i in 2:14) {
  cat(colnames(df)[i], "summary:\n")
  print(summary(df[[i]]))
  cat("\n")
}
```

```
## californian summary:
##    Mode   FALSE    TRUE
## logical     871     520
##
```

```
## online_course_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.2308  0.0000  4.0000
##
## lower_division_course_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  1.0000  1.0000  0.9597  1.0000  4.0000
##
## arts_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.4407  1.0000  4.0000
##
## engineering_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1179  0.0000  3.0000
##
## humanities_ugeduc_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1833  0.0000  3.0000
##
## life_sci_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    0.00    0.11    0.00    2.00
##
## phys_sci_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.1761  0.0000  4.0000
##
## soc_sci_intl_inst_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.4191  1.0000  2.0000
##
## mgmt_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.07692 0.00000 2.00000
##
## other_prof_ct summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.00000 0.00000 0.03523 0.00000 1.00000
##
## count_of_courses_taken summary:
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.559   2.000   4.000
##
## all_online summary:
##    Mode   FALSE    TRUE
## logical    1150     241
```

```
cor(df[,2:14])
```

```
##                           californian online_course_ct lower_division_course_ct
## californian                1.000000000      -0.08504306               0.03751925
## online_course_ct          -0.085043064       1.00000000               0.21293693
## lower_division_course_ct   0.037519254       0.21293693               1.00000000
## arts_ct                   -0.105111022      -0.10448157              -0.19645421
```

```
## engineering_ct            0.043036289       0.05516446               0.06168746
## humanities_ugeduc_ct     -0.024255483      -0.04532961               0.36364515
## life_sci_ct               0.116643745      -0.04005745               0.19018886
## phys_sci_ct               0.095145479       0.22673588               0.12743336
## soc_sci_intl_inst_ct      0.130786204      -0.03000108               0.01066445
## mgmt_ct                   0.005218642      -0.10163174              -0.04511858
## other_prof_ct             0.053859752       0.02943169               0.02181138
## count_of_courses_taken    0.124883334      -0.05471624               0.19340445
## all_online               -0.133853580       0.85957071               0.03397044
##                              arts_ct engineering_ct humanities_ugeduc_ct
## californian              -0.10511102     0.04303629          -0.02425548
## online_course_ct         -0.10448157     0.05516446          -0.04532961
## lower_division_course_ct -0.19645421     0.06168746           0.36364515
## arts_ct                   1.00000000    -0.15390799          -0.15967662
## engineering_ct           -0.15390799     1.00000000          -0.09548030
## humanities_ugeduc_ct     -0.15967662    -0.09548030           1.00000000
## life_sci_ct              -0.13702252    -0.09408706          -0.04244228
## phys_sci_ct              -0.15493273    -0.08784445          -0.06827278
## soc_sci_intl_inst_ct     -0.30083040    -0.18772409          -0.18222990
## mgmt_ct                  -0.09707628    -0.07890666          -0.05971250
## other_prof_ct            -0.08203985    -0.05580981          -0.01722587
## count_of_courses_taken    0.48601414     0.01751385           0.13507991
## all_online               -0.13674887     0.03569622          -0.08963261
##                          life_sci_ct phys_sci_ct soc_sci_intl_inst_ct
## californian               0.11664374  0.09514548          0.130786204
## online_course_ct         -0.04005745  0.22673588         -0.030001082
## lower_division_course_ct  0.19018886  0.12743336          0.010664449
## arts_ct                  -0.13702252 -0.15493273         -0.300830398
## engineering_ct           -0.09408706 -0.08784445         -0.187724095
## humanities_ugeduc_ct     -0.04244228 -0.06827278         -0.182229898
## life_sci_ct               1.00000000 -0.07567833         -0.044999113
## phys_sci_ct              -0.07567833  1.00000000         -0.117888940
## soc_sci_intl_inst_ct     -0.04499911 -0.11788894          1.000000000
## mgmt_ct                  -0.04266148 -0.07310100         -0.011701126
## other_prof_ct            -0.01587111 -0.05827468         -0.009759655
## count_of_courses_taken    0.12647218  0.15985747          0.178728186
## all_online               -0.05289575  0.07086068         -0.061885374
##                              mgmt_ct other_prof_ct count_of_courses_taken
## californian               0.005218642   0.053859752             0.12488333
## online_course_ct         -0.101631738   0.029431690            -0.05471624
## lower_division_course_ct -0.045118584   0.021811383             0.19340445
## arts_ct                  -0.097076281  -0.082039850             0.48601414
## engineering_ct           -0.078906659  -0.055809808             0.01751385
## humanities_ugeduc_ct     -0.059712499  -0.017225872             0.13507991
## life_sci_ct              -0.042661475  -0.015871108             0.12647218
## phys_sci_ct              -0.073101005  -0.058274678             0.15985747
## soc_sci_intl_inst_ct     -0.011701126  -0.009759655             0.17872819
## mgmt_ct                   1.000000000  -0.037927403             0.11318454
## other_prof_ct            -0.037927403   1.000000000             0.05058587
## count_of_courses_taken    0.113184539   0.050585869             1.00000000
## all_online               -0.103661821   0.036170539            -0.25373697
##                           all_online
## californian              -0.13385358
## online_course_ct          0.85957071
```

```
## lower_division_course_ct  0.03397044
## arts_ct                  -0.13674887
## engineering_ct            0.03569622
## humanities_ugeduc_ct     -0.08963261
## life_sci_ct              -0.05289575
## phys_sci_ct               0.07086068
## soc_sci_intl_inst_ct     -0.06188537
## mgmt_ct                  -0.10366182
## other_prof_ct             0.03617054
## count_of_courses_taken   -0.25373697
## all_online                1.00000000
```

What seems to predict or to characterize taking only online courses during summer term?

We can run logistic regression model; one for all_online regressed on each predictor

```
log_reg_1 <- glm(all_online ~ californian, data = df, family = binomial)
summary(log_reg_1)
```

```
##
## Call:
## glm(formula = all_online ~ californian, family = binomial, data = df)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -1.31052    0.08284 -15.819  < 2e-16 ***
## californianTRUE -0.80401    0.16393  -4.905 9.37e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1256.2  on 1389  degrees of freedom
## AIC: 1260.2
##
## Number of Fisher Scoring iterations: 4
```

```
log_reg_2 <- glm(all_online ~ online_course_ct, data = df, family = binomial)
summary(log_reg_2)
```

```
##
## Call:
## glm(formula = all_online ~ online_course_ct, family = binomial,
##     data = df)
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -7.010      1.000  -7.008 2.41e-12 ***
## online_course_ct    8.563      1.014   8.447  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.54  on 1390  degrees of freedom
```

```
## Residual deviance:  249.43  on 1389  degrees of freedom
## AIC: 253.43
##
## Number of Fisher Scoring iterations: 9
```

```
log_reg_3 <- glm(all_online ~ lower_division_course_ct, data = df, family = binomial)
summary(log_reg_3)
```

```
##
## Call:
## glm(formula = all_online ~ lower_division_course_ct, family = binomial,
##     data = df)
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -1.68505    0.12148 -13.871   <2e-16 ***
## lower_division_course_ct  0.12478    0.09856   1.266    0.205
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1280.9  on 1389  degrees of freedom
## AIC: 1284.9
##
## Number of Fisher Scoring iterations: 4
```

```
log_reg_4 <- glm(all_online ~ arts_ct, data = df, family = binomial)
summary(log_reg_4)
```

```
##
## Call:
## glm(formula = all_online ~ arts_ct, family = binomial, data = df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.33921    0.07812 -17.143  < 2e-16 ***
## arts_ct     -0.72272    0.14718  -4.911 9.08e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1247.1  on 1389  degrees of freedom
## AIC: 1251.1
##
## Number of Fisher Scoring iterations: 5
```

```
log_reg_5 <- glm(all_online ~ engineering_ct, data = df, family = binomial)
summary(log_reg_5)
```

```
##
## Call:
## glm(formula = all_online ~ engineering_ct, family = binomial,
```

```
##     data = df)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.59015    0.07439 -21.376   <2e-16 ***
## engineering_ct  0.21202    0.16002   1.325    0.185
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1280.9  on 1389  degrees of freedom
## AIC: 1284.9
##
## Number of Fisher Scoring iterations: 4
```

```
log_reg_6 <- glm(all_online ~ humanities_ugeduc_ct, data = df, family = binomial)
summary(log_reg_6)
```

```
##
## Call:
## glm(formula = all_online ~ humanities_ugeduc_ct, family = binomial,
##     data = df)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -1.46346    0.07464 -19.608   <2e-16 ***
## humanities_ugeduc_ct -0.70096    0.21474  -3.264   0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1269.2  on 1389  degrees of freedom
## AIC: 1273.2
##
## Number of Fisher Scoring iterations: 5
```

```
log_reg_7 <- glm(all_online ~ life_sci_ct, data = df, family = binomial)
summary(log_reg_7)
```

```
##
## Call:
## glm(formula = all_online ~ life_sci_ct, family = binomial, data = df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.51740    0.07347 -20.652   <2e-16 ***
## life_sci_ct -0.49104    0.25195  -1.949   0.0513 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1278.2  on 1389  degrees of freedom
## AIC: 1282.2
##
## Number of Fisher Scoring iterations: 4
```

```
log_reg_8 <- glm(all_online ~ phys_sci_ct, data = df, family = binomial)
summary(log_reg_8)
```

```
##
## Call:
## glm(formula = all_online ~ phys_sci_ct, family = binomial, data = df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.63641    0.07773 -21.051  < 2e-16 ***
## phys_sci_ct  0.36836    0.14136   2.606  0.00916 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1276.2  on 1389  degrees of freedom
## AIC: 1280.2
##
## Number of Fisher Scoring iterations: 4
```

```
log_reg_9 <- glm(all_online ~ soc_sci_intl_inst_ct, data = df, family = binomial)
summary(log_reg_9)
```

```
##
## Call:
## glm(formula = all_online ~ soc_sci_intl_inst_ct, family = binomial,
##     data = df)
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -1.45255    0.08337 -17.424   <2e-16 ***
## soc_sci_intl_inst_ct -0.28653    0.12466  -2.298   0.0215 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1277.0  on 1389  degrees of freedom
## AIC: 1281
##
## Number of Fisher Scoring iterations: 4
```

```
log_reg_10 <- glm(all_online ~ mgmt_ct, data = df, family = binomial)
summary(log_reg_10)
```

```
##
```

```
## Call:
## glm(formula = all_online ~ mgmt_ct, family = binomial, data = df)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.48696    0.07176 -20.720  < 2e-16 ***
## mgmt_ct     -1.93702    0.58427  -3.315 0.000915 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1260.6  on 1389  degrees of freedom
## AIC: 1264.6
##
## Number of Fisher Scoring iterations: 6
```

```
log_reg_11 <- glm(all_online ~ other_prof_ct, data = df, family = binomial)
summary(log_reg_11)
```

```
##
## Call:
## glm(formula = all_online ~ other_prof_ct, family = binomial,
##     data = df)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.58109    0.07256 -21.789   <2e-16 ***
## other_prof_ct  0.45508    0.34004   1.338    0.181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1280.9  on 1389  degrees of freedom
## AIC: 1284.9
##
## Number of Fisher Scoring iterations: 4
```

```
log_reg_12 <- glm(all_online ~ count_of_courses_taken, data = df, family = binomial)
summary(log_reg_12)
```

```
##
## Call:
## glm(formula = all_online ~ count_of_courses_taken, family = binomial,
##     data = df)
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)               0.6199     0.2285   2.713  0.00667 **
## count_of_courses_taken   -1.6301     0.1827  -8.921  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.5  on 1390  degrees of freedom
## Residual deviance: 1155.9  on 1389  degrees of freedom
## AIC: 1159.9
##
## Number of Fisher Scoring iterations: 6
```

significnt predictors: 1. californian 2. online_course_ct 3. arts_ct 4. humanities_ugeduc_ct 5. phys_sci_ct 6. soc_sci_intl_inst_ct 7. mgmt_ct 8. count_of_courses_taken

If we want, we can elaborate about the log-odds. This can be good for understanding what makes one more likely to take only online classes in the summer. Now, we'll attempt to find a full logistic regression model to predict all_online:

```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-8
```

```
X <- model.matrix(all_online ~ californian + online_course_ct +
                    lower_division_course_ct + arts_ct + engineering_ct +
                    humanities_ugeduc_ct + life_sci_ct + phys_sci_ct +
                    soc_sci_intl_inst_ct + mgmt_ct + other_prof_ct +
                    count_of_courses_taken, data = df)[, -1]
y <- df$all_online
cv_fit <- cv.glmnet(X, y, family = "binomial", alpha = 0)
coef(cv_fit, s = "lambda.min")
```

```
## 13 x 1 sparse Matrix of class "dgCMatrix"
##                                      s1
## (Intercept)                -1.329453359
## californianTRUE            -0.347339935
## online_course_ct            3.840464884
## lower_division_course_ct   -0.181203876
## arts_ct                    -0.296391277
## engineering_ct             -0.003248642
## humanities_ugeduc_ct       -0.375854460
## life_sci_ct                -0.209002735
## phys_sci_ct                -0.338333553
## soc_sci_intl_inst_ct       -0.272276554
## mgmt_ct                    -0.439823874
## other_prof_ct               0.182566376
## count_of_courses_taken     -0.926185928
```

```
null_model <- glm(all_online ~ 1,
                  data = df[, !(colnames(df) == "casenumber")],
                  family = binomial
)

full_scope <- formula(
  paste("all_online ~", paste(setdiff(names(df), c("all_online", "casenumber")), collapse = " + "))
)

step_model <- step(null_model,
```

```
                  scope = full_scope,
                  direction = "forward"
)
```

```
## Start:  AIC=1284.54
## all_online ~ 1
##
##                            Df Deviance     AIC
## + online_course_ct          1   249.43  253.43
## + count_of_courses_taken    1  1155.94 1159.94
## + arts_ct                   1  1247.10 1251.10
## + californian               1  1256.15 1260.15
## + mgmt_ct                   1  1260.59 1264.59
## + humanities_ugeduc_ct      1  1269.21 1273.21
## + phys_sci_ct               1  1276.19 1280.19
## + soc_sci_intl_inst_ct      1  1276.95 1280.95
## + life_sci_ct               1  1278.17 1282.17
## <none>                         1282.54 1284.54
## + other_prof_ct             1  1280.87 1284.87
## + engineering_ct            1  1280.88 1284.88
## + lower_division_course_ct  1  1280.95 1284.95
##
## Step:  AIC=253.43
## all_online ~ online_course_ct
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
##                            Df Deviance    AIC
## + count_of_courses_taken    1     0.00   6.00
## + lower_division_course_ct  1   147.28 153.28
## + phys_sci_ct               1   191.75 197.75
## + humanities_ugeduc_ct      1   222.03 228.03
## + californian               1   223.22 229.22
## + arts_ct                   1   227.19 233.19
## + engineering_ct            1   242.59 248.59
## <none>                          249.43 253.43
## + mgmt_ct                   1   247.86 253.86
## + soc_sci_intl_inst_ct      1   247.93 253.93
## + life_sci_ct               1   248.53 254.53
## + other_prof_ct             1   248.85 254.85
```

```
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
##
## Step:  AIC=6
## all_online ~ online_course_ct + count_of_courses_taken
```

```
## Warning: glm.fit: algorithm did not converge
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

##                             Df   Deviance AIC
## <none>                          2.9482e-08   6
## + lower_division_course_ct  1 2.9148e-08   8
## + phys_sci_ct               1 2.9155e-08   8
## + arts_ct                   1 2.9376e-08   8
## + humanities_ugeduc_ct      1 2.9462e-08   8
## + soc_sci_intl_inst_ct      1 2.9482e-08   8
## + life_sci_ct               1 2.9485e-08   8
## + other_prof_ct             1 2.9488e-08   8
## + engineering_ct            1 2.9495e-08   8
## + californian               1 2.9510e-08   8
## + mgmt_ct                   1 2.9543e-08   8
```

We have convergence issues with estimating coeffcients. This is likely due to the strong correlation between the predictor variable online_course_ct and the response variable. This is because we have perfect separation (anyone with more than 2 online course are taking only online courses)

```
table(df$all_online, df$online_course_ct)
```

```
##
##            0    1    2    3    4
##   FALSE 1107   42    1    0    0
##   TRUE     0  210   27    3    1
```

However, we've still landed on a model (all_online ~ online_course_ct + count_of_courses_taken). We can use penalized step-wise regression to get more stable estimates.

```r
library(glmnet)
X <- model.matrix(all_online ~ online_course_ct + count_of_courses_taken, data = df)[, -1]
y <- df$all_online

cv_fit <- cv.glmnet(X, y, family = "binomial", alpha = 0)
coef(cv_fit, s = "lambda.min")
```

```
## 3 x 1 sparse Matrix of class "dgCMatrix"
##                                s1
## (Intercept)            -1.622526
## online_course_ct        3.808057
## count_of_courses_taken -1.167525
```

We've landed on this model. Here

- Each additional online course a student takes in the summer increases the odds that they're only taking online classes by 45.0627967.
- On the other hand, each additional class (online or not) a student takes in the summer decreases the odda that they're only taking online classes by 0.3111361