Project 6: Randomization and Matching

Introduction

In this project, you will explore the question of whether college education causally affects political participation. Specifically, you will use replication data from Who Matches? Propensity Scores and Bias in the Causal Effects of Education on Participation by former Berkeley PhD students John Henderson and Sara Chatfield. Their paper is itself a replication study of Reconsidering the Effects of Education on Political Participation by Cindy Kam and Carl Palmer. In their original 2008 study, Kam and Palmer argue that college education has no effect on later political participation, and use the propensity score matching to show that pre-college political activity drives selection into college and later political participation. Henderson and Chatfield in their 2011 paper argue that the use of the propensity score matching in this context is inappropriate because of the bias that arises from small changes in the choice of variables used to model the propensity score. They use genetic matching (at that point a new method), which uses an approach similar to optimal matching to optimize Mahalanobis distance weights. Even with genetic matching, they find that balance remains elusive however, thus leaving open the question of whether education causes political participation.

You will use these data and debates to investigate the benefits and pitfalls associated with matching methods. Replication code for these papers is available online, but as you'll see, a lot has changed in the last decade or so of data science! Throughout the assignment, use tools we introduced in lab from the tidyverse and the MatchIt packages. Specifically, try to use dplyr, tidyr, purrr, stringr, and ggplot instead of base R functions. While there are other matching software libraries available, MatchIt tends to be the most up to date and allows for consistent syntax.

Data

The data is drawn from the Youth-Parent Socialization Panel Study which asked students and parents a variety of questions about their political participation. This survey was conducted in several waves. The first wave was in 1965 and established the baseline pre-treatment covariates. The treatment is whether the student attended college between 1965 and 1973 (the time when the next survey wave was administered). The outcome is an index that calculates the number of political activities the student engaged in after 1965. Specifically, the key variables in this study are:

- **college**: Treatment of whether the student attended college or not. 1 if the student attended college between 1965 and 1973, 0 otherwise.
- **ppnscal**: Outcome variable measuring the number of political activities the student participated in. Additive combination of whether the student voted in 1972 or 1980 (student_vote), attended a campaign rally or meeting (student_meeting), wore a campaign button (student_button), donated money to a campaign (student_money), communicated with an elected official (student_communicate), attended a demonstration or protest (student_demonstrate), was involved with a local community event (student_community), or some other political participation (student_other)

Otherwise, we also have covariates measured for survey responses to various questions about political attitudes. We have covariates measured for the students in the baseline year, covariates for their parents in the baseline year, and covariates from follow-up surveys. **Be careful here**. In general, post-treatment covariates will be clear from the name (i.e. student_1973Married indicates whether the student was married in the 1973 survey). Be mindful that the baseline covariates were all measured in 1965, the treatment occurred between 1965 and 1973, and the outcomes are from 1973 and beyond. We will distribute the Appendix from Henderson and Chatfield that describes the covariates they used, but please reach out with any questions if you have questions about what a particular variable means.

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
##  cobalt (Version 4.5.5, Build Date: 2024-04-02)
##
##
## Attaching package: 'cobalt'
##
##
## The following object is masked from 'package:MatchIt':
##
##     lalonde
##
##
##
## Attaching package: 'gridExtra'
##
##
## The following object is masked from 'package:dplyr':
##
##     combine
##
##
##
## Attaching package: 'reshape2'
##
##
## The following object is masked from 'package:tidyr':
##
##     smiths
##
##
## Rows: 1254 Columns: 174
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (174): interviewid, college, student_vote, student_meeting, student_othe...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 6 x 174
##   interviewid college student_vote student_meeting student_other student_button
##         <dbl>   <dbl>        <dbl>           <dbl>         <dbl>          <dbl>
## 1           1       1            1               0             0              0
## 2           2       1            1               1             1              1
## 3           3       1            1               0             0              1
## 4           4       0            0               0             0              0
## 5           5       1            1               1             0              0
## 6           6       1            1               0             0              0
## # i 168 more variables: student_money <dbl>, student_communicate <dbl>,
## #   student_demonstrate <dbl>, student_community <dbl>, student_ppnscal <dbl>,
## #   student_PubAff <dbl>, student_Newspaper <dbl>, student_Radio <dbl>,
## #   student_TV <dbl>, student_Magazine <dbl>, student_FamTalk <dbl>,
## #   student_FrTalk <dbl>, student_AdultTalk <dbl>, student_PID <dbl>,
## #   student_SPID <dbl>, student_GovtOpinion <dbl>, student_GovtCrook <dbl>,
## #   student_GovtWaste <dbl>, student_TrGovt <dbl>, student_GovtSmart <dbl>, ...
```

Randomization

Matching is usually used in observational studies to to approximate random assignment to treatment. But could it be useful even in randomized studies? To explore the question do the following:

1. Generate a vector that randomly assigns each unit to either treatment or control
2. Choose a baseline covariate (for either the student or parent). A binary covariate is probably best for this exercise.
3. Visualize the distribution of the covariate by treatment/control condition. Are treatment and control balanced on this covariate?
4. Simulate the first 3 steps 10,000 times and visualize the distribution of treatment/control balance across the simulations.

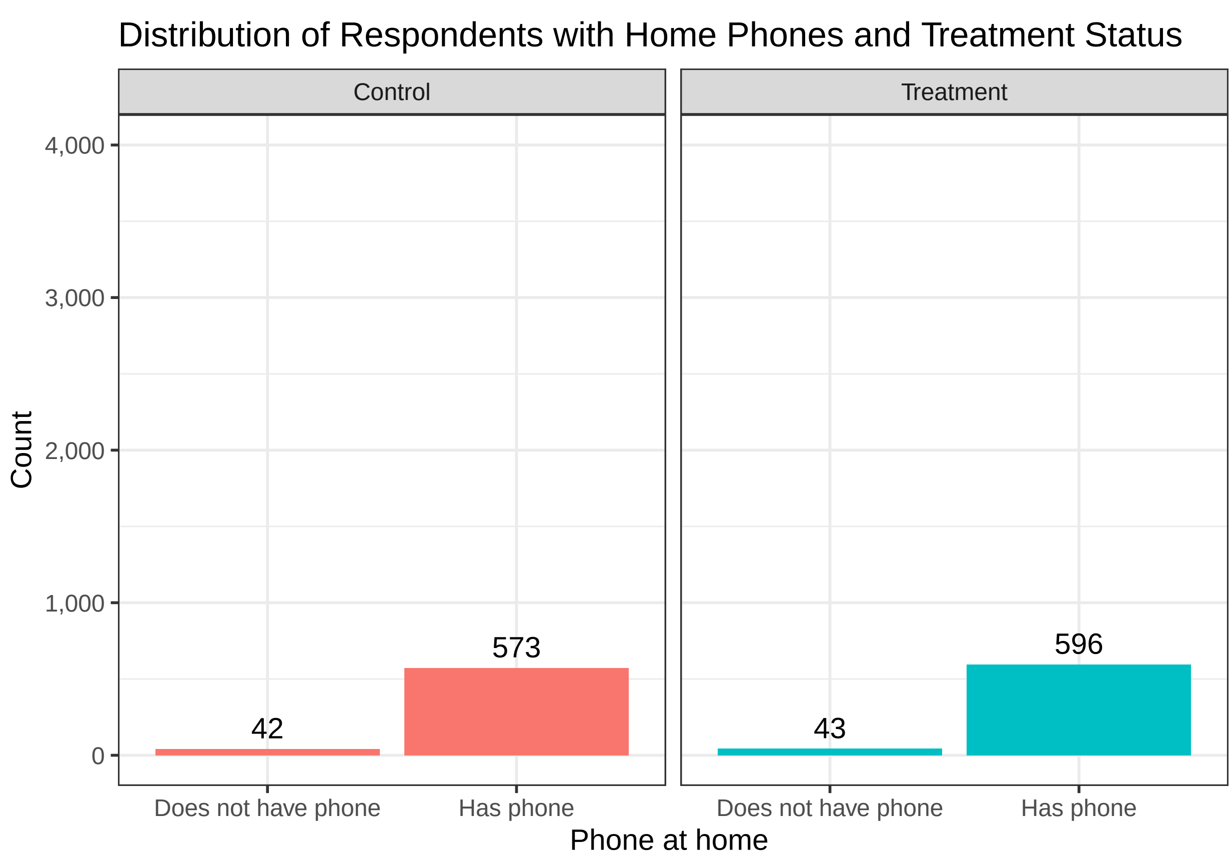
```
##   [1] "college"                        "interviewid"
##   [3] "parent_ActFrq"                  "parent_Button"
##   [5] "parent_CCamp"                   "parent_ChurchOrg"
##   [7] "parent_CivicOrg"                "parent_CLOrg"
##   [9] "parent_ClubLev"                 "parent_Court"
##  [11] "parent_EducHH"                  "parent_EducW"
##  [13] "parent_Employ"                  "parent_FarmGr"
##  [15] "parent_FDR"                     "parent_FInc"
##  [17] "parent_FPlans"                  "parent_FratOrg"
##  [19] "parent_Gen"                     "parent_GLuck"
##  [21] "parent_Govern"                  "parent_Govt4All"
##  [23] "parent_GovtCrook"               "parent_GovtOpinion"
##  [25] "parent_GovtSmart"               "parent_GovtWaste"
##  [27] "parent_GPHighSchoolPlacebo"     "parent_HHCollegePlacebo"
##  [29] "parent_HHInc"                   "parent_InfClub"
##  [31] "parent_Knowledge"               "parent_LifeWish"
##  [33] "parent_Magazine"                "parent_MChange"
##  [35] "parent_MiscClub"                "parent_Money"
##  [37] "parent_NeighClub"               "parent_Newspaper"
##  [39] "parent_OthAct"                  "parent_OthFair"
```

```
##  [41] "parent_OthHelp"               "parent_OwnHome"
##  [43] "parent_Participate1"          "parent_Participate2"
##  [45] "parent_Persuade"              "parent_PID"
##  [47] "parent_PolClub"               "parent_ProOrg"
##  [49] "parent_Race"                  "parent_Radio"
##  [51] "parent_Rally"                 "parent_Senate"
##  [53] "parent_SPID"                  "parent_SportClub"
##  [55] "parent_StrOpinion"            "parent_Tito"
##  [57] "parent_TrGovt"                "parent_TrOthers"
##  [59] "parent_TV"                    "parent_Vote"
##  [61] "parent_WinArg"                "parent_WomenClub"
##  [63] "student_1973Busing"           "student_1973ChurchAttend"
##  [65] "student_1973CollegeDegree"    "student_1973CollegeYears"
##  [67] "student_1973CurrentCollege"   "student_1973CurrentSituation"
##  [69] "student_1973Drafted"          "student_1973FutureSituation"
##  [71] "student_1973GovChange"        "student_1973GovtEfficacy"
##  [73] "student_1973GovtNoSay"        "student_1973HelpMinority"
##  [75] "student_1973HHInc"            "student_1973Ideology"
##  [77] "student_1973IncSelf"          "student_1973Knowledge"
##  [79] "student_1973Luck"             "student_1973Married"
##  [81] "student_1973Military"         "student_1973Newspaper"
##  [83] "student_1973NoEmployers"      "student_1973NoResidences"
##  [85] "student_1973OwnHome"          "student_1973PartyID"
##  [87] "student_1973PubAffairs"       "student_1973SureAboutLife"
##  [89] "student_1973ThermBlack"       "student_1973ThermDems"
##  [91] "student_1973ThermMcgovern"    "student_1973ThermMilitary"
##  [93] "student_1973ThermNixon"       "student_1973ThermRadical"
##  [95] "student_1973ThermRep"         "student_1973ThermWhite"
##  [97] "student_1973Trust"            "student_1973Unemployed"
##  [99] "student_1973VietnamApprove"   "student_1973VietnamRight"
## [101] "student_1973VoteMcgovern"     "student_1973VoteNixon"
## [103] "student_1982button"           "student_1982College"
## [105] "student_1982communicate"      "student_1982community"
## [107] "student_1982demonstrate"      "student_1982HHInc"
## [109] "student_1982IncSelf"          "student_1982meeting"
## [111] "student_1982money"            "student_1982other"
## [113] "student_1982vote76"           "student_1982vote80"
## [115] "student_AdultTalk"            "student_button"
## [117] "student_CCamp"                "student_ClubLev"
## [119] "student_communicate"          "student_community"
## [121] "student_Court"                "student_Cynic"
## [123] "student_demonstrate"          "student_EgoA"
## [125] "student_EgoB"                 "student_FamTalk"
## [127] "student_FDR"                  "student_FPlans"
## [129] "student_FrTalk"               "student_Gen"
## [131] "student_GLuck"                "student_Govern"
## [133] "student_Govt4All"             "student_GovtCrook"
## [135] "student_GovtOpinion"          "student_GovtSmart"
## [137] "student_GovtWaste"            "student_GPA"
## [139] "student_Hobby"                "student_Knowledge"
## [141] "student_LifeWish"             "student_Magazine"
## [143] "student_MChange"              "student_meeting"
## [145] "student_MiscClub"             "student_money"
## [147] "student_NeighClub"            "student_Newspaper"
```

```
## [149] "student_NextSch"              "student_OccClub"
## [151] "student_other"                "student_OthFair"
## [153] "student_OthHelp"              "student_Phone"
## [155] "student_PID"                  "student_ppnscal"
## [157] "student_PubAff"               "student_Race"
## [159] "student_Radio"                "student_RelClub"
## [161] "student_SchClub"              "student_SchOfficer"
## [163] "student_SchPublish"           "student_Senate"
## [165] "student_SPID"                 "student_StrOpinion"
## [167] "student_Tito"                 "student_TrGovt"
## [169] "student_TrOthers"             "student_Trust"
## [171] "student_TV"                   "student_vote"
## [173] "student_WinArg"               "student_YouthOrg"
## [175] "treatment"
```

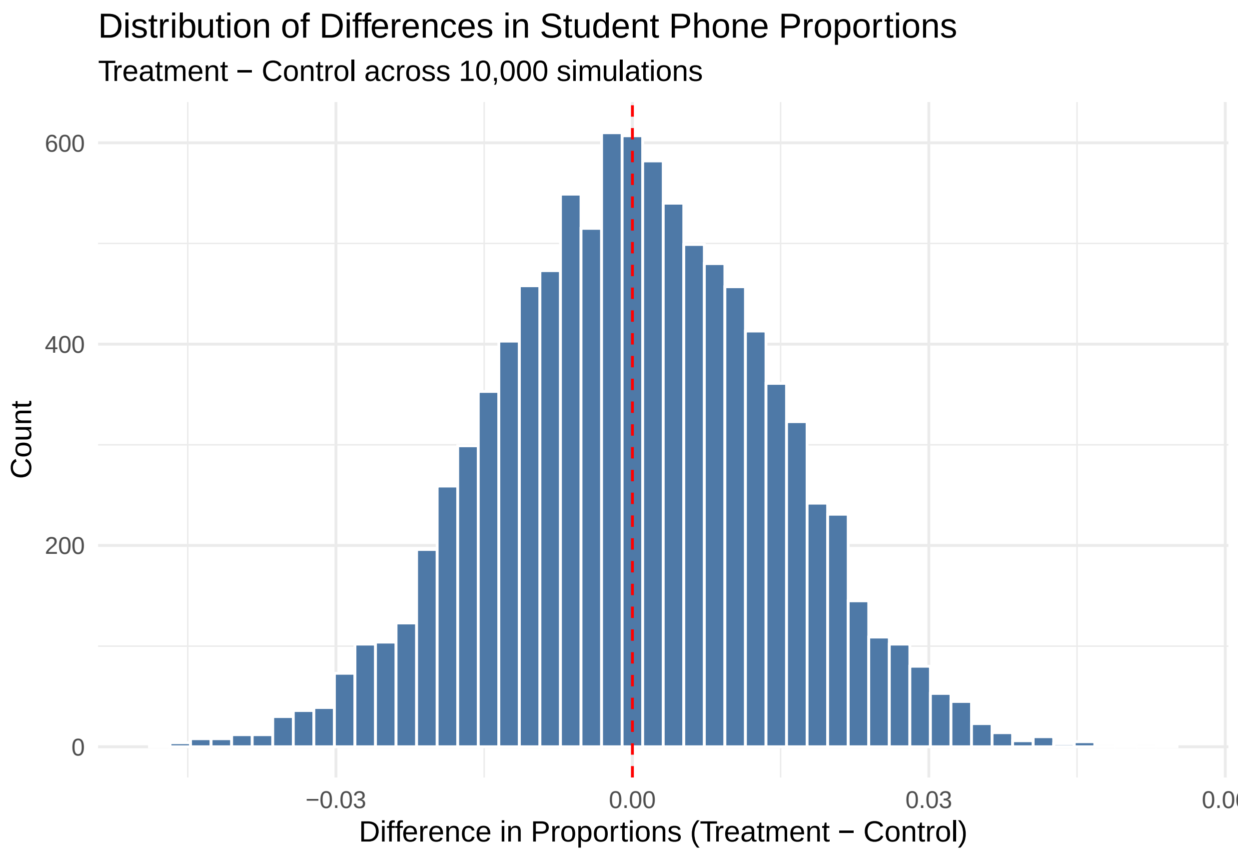
```
## [1] 1 0
```

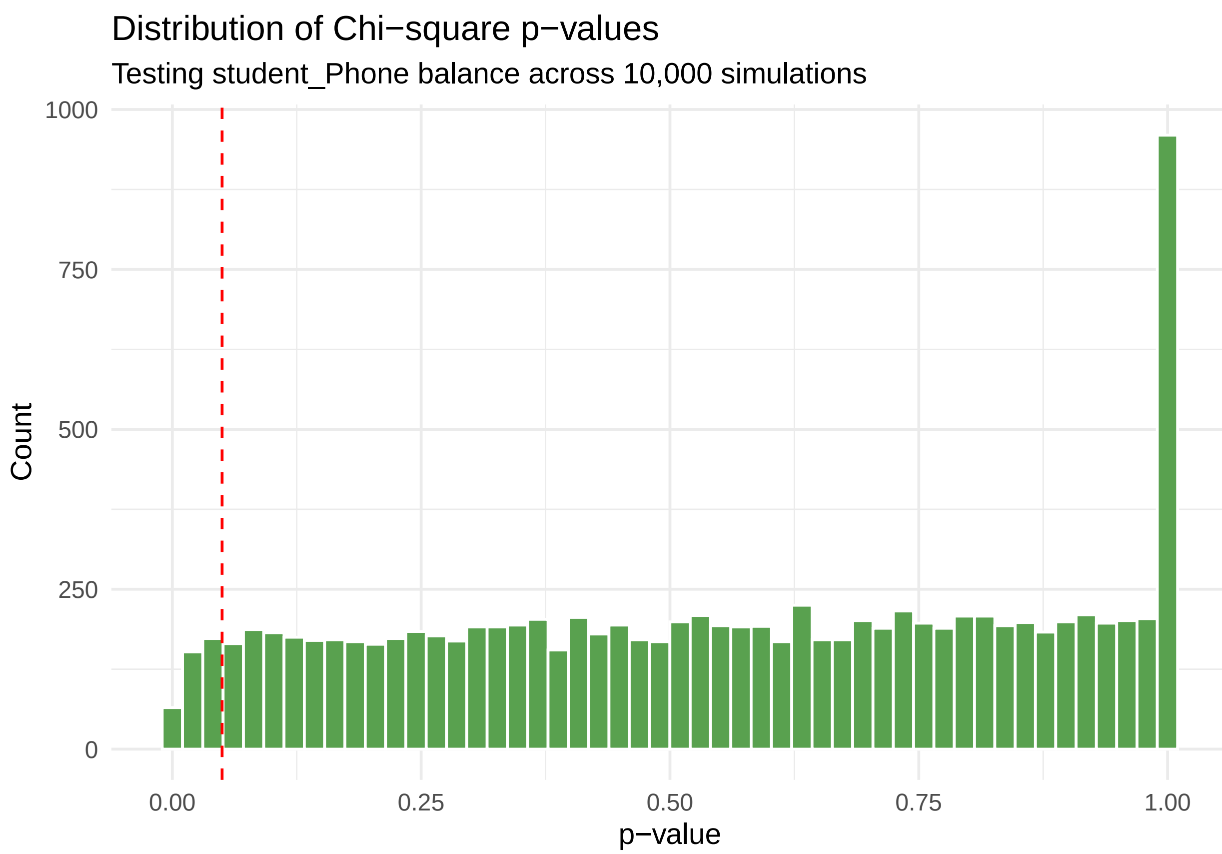
```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(count)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```
## Proportion of simulations with significant imbalance (p < 0.05): 3.82%
```

```
## Average absolute difference in proportions: 0.0113

## Standard deviation of differences: 0.0141
```





Questions

1. **What do you see across your simulations? Why does independence of treatment assignment and baseline covariates not guarantee balance of treatment assignment and baseline covariates?**

The histogram above shows the distribution fo differences in student phone proportions between treatment and control groups across all 10,000 simulations. The histogram shows a ~ normal distribution centered around 0 with differences randing from approximately -0.03 to 0.03. The second chart shows hte distribution of chi square p values from testing student phone balance across the simulations. It is mostly uniform, with a spike at 1, which indicates perfect matches. However, even with perfect randomization, some of my simulations (3.66%) still showed statistically significant imbalances (<0.05) between treatment and control groups. This shows that while randomization ensures independence in the assignment process, it does not guarantee perfect balance and by chance alone we can expect approximately 5% of properly randomized experiments to have "statistically significant" imbalanced covariates.

Propensity Score Matching

One Model

Select covariates that you think best represent the "true" model predicting whether a student chooses to attend college, and estimate a propensity score model to calculate the Average Treatment Effect on the Treated (ATT). Plot the balance of the top 10 (or fewer if you select fewer covariates). Report the balance of the p-scores across both the treatment and control groups, and using a threshold of standardized mean difference of p-score $\leq .1$, report the number of covariates that meet that balance threshold.

```
##   [1] "college"                      "interviewid"
##   [3] "parent_ActFrq"                "parent_Button"
##   [5] "parent_CCamp"                 "parent_ChurchOrg"
##   [7] "parent_CivicOrg"              "parent_CLOrg"
##   [9] "parent_ClubLev"               "parent_Court"
##  [11] "parent_EducHH"                "parent_EducW"
##  [13] "parent_Employ"                "parent_FarmGr"
##  [15] "parent_FDR"                   "parent_FInc"
##  [17] "parent_FPlans"                "parent_FratOrg"
##  [19] "parent_Gen"                   "parent_GLuck"
##  [21] "parent_Govern"                "parent_Govt4All"
##  [23] "parent_GovtCrook"             "parent_GovtOpinion"
##  [25] "parent_GovtSmart"             "parent_GovtWaste"
##  [27] "parent_GPHighSchoolPlacebo"   "parent_HHCollegePlacebo"
##  [29] "parent_HHInc"                 "parent_InfClub"
##  [31] "parent_Knowledge"             "parent_LifeWish"
##  [33] "parent_Magazine"              "parent_MChange"
##  [35] "parent_MiscClub"              "parent_Money"
##  [37] "parent_NeighClub"             "parent_Newspaper"
##  [39] "parent_OthAct"                "parent_OthFair"
##  [41] "parent_OthHelp"               "parent_OwnHome"
##  [43] "parent_Participate1"          "parent_Participate2"
##  [45] "parent_Persuade"              "parent_PID"
##  [47] "parent_PolClub"               "parent_ProOrg"
##  [49] "parent_Race"                  "parent_Radio"
##  [51] "parent_Rally"                 "parent_Senate"
##  [53] "parent_SPID"                  "parent_SportClub"
##  [55] "parent_StrOpinion"            "parent_Tito"
##  [57] "parent_TrGovt"                "parent_TrOthers"
##  [59] "parent_TV"                    "parent_Vote"
##  [61] "parent_WinArg"                "parent_WomenClub"
##  [63] "student_1973Busing"           "student_1973ChurchAttend"
##  [65] "student_1973CollegeDegree"    "student_1973CollegeYears"
##  [67] "student_1973CurrentCollege"   "student_1973CurrentSituation"
##  [69] "student_1973Drafted"          "student_1973FutureSituation"
##  [71] "student_1973GovChange"        "student_1973GovtEfficacy"
##  [73] "student_1973GovtNoSay"        "student_1973HelpMinority"
##  [75] "student_1973HHInc"            "student_1973Ideology"
##  [77] "student_1973IncSelf"          "student_1973Knowledge"
##  [79] "student_1973Luck"             "student_1973Married"
##  [81] "student_1973Military"         "student_1973Newspaper"
##  [83] "student_1973NoEmployers"      "student_1973NoResidences"
##  [85] "student_1973OwnHome"          "student_1973PartyID"
##  [87] "student_1973PubAffairs"       "student_1973SureAboutLife"
##  [89] "student_1973ThermBlack"       "student_1973ThermDems"
##  [91] "student_1973ThermMcgovern"    "student_1973ThermMilitary"
##  [93] "student_1973ThermNixon"       "student_1973ThermRadical"
##  [95] "student_1973ThermRep"         "student_1973ThermWhite"
##  [97] "student_1973Trust"            "student_1973Unemployed"
##  [99] "student_1973VietnamApprove"   "student_1973VietnamRight"
## [101] "student_1973VoteMcgovern"     "student_1973VoteNixon"
```

```
## [103] "student_1982button"           "student_1982College"
## [105] "student_1982communicate"      "student_1982community"
## [107] "student_1982demonstrate"      "student_1982HHInc"
## [109] "student_1982IncSelf"          "student_1982meeting"
## [111] "student_1982money"            "student_1982other"
## [113] "student_1982vote76"           "student_1982vote80"
## [115] "student_AdultTalk"            "student_button"
## [117] "student_CCamp"                "student_ClubLev"
## [119] "student_communicate"          "student_community"
## [121] "student_Court"                "student_Cynic"
## [123] "student_demonstrate"          "student_EgoA"
## [125] "student_EgoB"                 "student_FamTalk"
## [127] "student_FDR"                  "student_FPlans"
## [129] "student_FrTalk"               "student_Gen"
## [131] "student_GLuck"                "student_Govern"
## [133] "student_Govt4All"             "student_GovtCrook"
## [135] "student_GovtOpinion"          "student_GovtSmart"
## [137] "student_GovtWaste"            "student_GPA"
## [139] "student_Hobby"                "student_Knowledge"
## [141] "student_LifeWish"             "student_Magazine"
## [143] "student_MChange"              "student_meeting"
## [145] "student_MiscClub"             "student_money"
## [147] "student_NeighClub"            "student_Newspaper"
## [149] "student_NextSch"              "student_OccClub"
## [151] "student_other"                "student_OthFair"
## [153] "student_OthHelp"              "student_Phone"
## [155] "student_PID"                  "student_ppnscal"
## [157] "student_PubAff"               "student_Race"
## [159] "student_Radio"                "student_RelClub"
## [161] "student_SchClub"              "student_SchOfficer"
## [163] "student_SchPublish"           "student_Senate"
## [165] "student_SPID"                 "student_StrOpinion"
## [167] "student_Tito"                 "student_TrGovt"
## [169] "student_TrOthers"             "student_Trust"
## [171] "student_TV"                   "student_vote"
## [173] "student_WinArg"               "student_YouthOrg"

## Selected covariates for propensity score model: student_PubAff student_Newspaper student_Radio studen

##
## Call:
## matchit(formula = college ~ student_PubAff + student_Newspaper +
##     student_Radio + student_TV + student_Magazine + student_FamTalk +
##     student_FrTalk + student_AdultTalk + student_GovtOpinion +
##     student_GovtCrook + student_GovtWaste + student_TrGovt +
##     student_GovtSmart + student_Govt4All + student_LifeWish +
##     student_GLuck + student_FPlans + student_WinArg + student_StrOpinion +
##     student_MChange + student_TrOthers + student_OthHelp + student_OthFair +
##     student_SchOfficer + student_SchPublish + student_Hobby +
##     student_SchClub + student_OccClub + student_NeighClub + student_RelClub +
##     student_YouthOrg + student_MiscClub + student_vote + student_meeting +
##     student_other + student_button + student_money + student_communicate +
##     student_demonstrate + student_SPID + student_Knowledge +
##     student_NextSch + student_GPA + student_Phone + student_Gen +
```

```
##     student_Race + parent_Newspaper + parent_Radio + parent_TV +
##     parent_Magazine + parent_GovtOpinion + parent_GovtCrook +
##     parent_GovtWaste + parent_TrGovt + parent_GovtSmart + parent_Govt4All +
##     parent_SPID + parent_Employ + parent_EducHH + parent_EducW +
##     parent_FInc + parent_HHInc + parent_OwnHome + parent_Race,
##     data = analysis_data, method = "nearest", distance = "glm",
##     link = "logit", discard = "control", replace = TRUE, caliper = 0.2,
##     ratio = 1)
##
## Summary of Balance for All Data:
##                     Means Treated Means Control Std. Mean Diff. Var. Ratio
## distance                   0.8002        0.3558          2.0593     0.6058
## student_PubAff             0.9465        0.9490         -0.0113          .
## student_Newspaper          1.9427        2.2328         -0.2209     0.8638
## student_Radio              2.6401        2.7761         -0.0769     0.9925
## student_TV                 2.2167        2.1308          0.0672     0.9957
## student_Magazine           1.6139        2.0133         -0.4578     0.8563
## student_FamTalk            1.8431        2.1131         -0.2932     0.7770
## student_FrTalk             1.9938        2.4169         -0.4468     0.7057
## student_AdultTalk          2.9614        2.9978         -0.0350     0.9310
## student_GovtOpinion        1.5965        1.7361         -0.2193     0.9144
## student_GovtCrook          2.1021        2.1419         -0.0593     0.9806
## student_GovtWaste          1.9377        1.9690         -0.0472     0.9965
## student_TrGovt             1.6152        1.7184         -0.1682     0.9564
## student_GovtSmart          1.2503        1.3193         -0.1055     0.8078
## student_Govt4All           2.7061        2.7494         -0.0641     1.1020
## student_LifeWish           2.1320        2.2993         -0.1688     1.0821
## student_GLuck              1.2279        1.4102         -0.3244     0.5485
## student_FPlans             1.5567        1.8204         -0.2966     0.8305
## student_WinArg             2.1370        2.3880         -0.2782     1.1269
## student_StrOpinion         1.7796        1.8869         -0.1110     0.9669
## student_MChange            1.4408        1.4013          0.0494     1.0588
## student_TrOthers           1.6413        1.8204         -0.1929     0.8962
## student_OthHelp            1.7335        1.7450         -0.0122     0.9698
## student_OthFair            1.4309        1.5543         -0.1517     0.8525
## student_SchOfficer         2.0212        2.2661         -0.2956     0.9193
## student_SchPublish         1.7111        1.6098          0.0883     1.1830
## student_Hobby              1.3873        1.3747          0.0143     1.0345
## student_SchClub            1.7771        1.4390          0.3182     1.4143
## student_OccClub            1.7161        1.9024         -0.1678     0.8650
## student_NeighClub          1.7696        1.6009          0.1483     1.2445
## student_RelClub            2.5567        2.3969          0.1403     0.9893
## student_YouthOrg           1.7111        1.4745          0.2155     1.4039
## student_MiscClub           0.3313        0.1840          0.3128          .
## student_vote               0.8070        0.5721          0.5952          .
## student_meeting            0.2989        0.0909          0.4543          .
## student_other              0.1756        0.0599          0.3042          .
## student_button             0.3362        0.2284          0.2283          .
## student_money              0.2192        0.0953          0.2993          .
## student_communicate        0.4060        0.1574          0.5061          .
## student_demonstrate        0.2154        0.0466          0.4108          .
## student_SPID               1.7360        1.7716         -0.0369     0.8865
## student_Knowledge          0.6752        0.4634          0.9300     0.9403
## student_NextSch            0.9577        0.6452          1.5515          .
```

```
## student_GPA               2.2740        2.6231         -0.5332     0.9101
## student_Phone             0.9614        0.8803          0.4211          .
## student_Gen               0.5479        0.4324          0.2322          .
## student_Race              1.0822        1.0931         -0.0359     1.0438
## parent_Newspaper          3.3674        2.7938          0.4612     0.5821
## parent_Radio              2.4770        2.3038          0.0963     0.9456
## parent_TV                 3.2130        3.2639         -0.0417     0.9538
## parent_Magazine           0.6600        0.4678          0.4057          .
## parent_GovtOpinion        1.6887        1.8514         -0.2389     0.9117
## parent_GovtCrook          2.0498        1.9579          0.1283     1.0894
## parent_GovtWaste          1.5741        1.7228         -0.2498     0.8561
## parent_TrGovt             2.0349        1.9845          0.0804     0.9772
## parent_GovtSmart          1.5567        1.5078          0.0559     1.0612
## parent_Govt4All           2.4969        2.4878          0.0111     0.9756
## parent_SPID               2.0100        1.9202          0.0930     0.8759
## parent_Employ             0.6961        0.6120          0.1830          .
## parent_EducHH             3.3238        2.1685          0.7880     1.7174
## parent_EducW              3.1071        2.2993          0.6663     1.4055
## parent_FInc               7.7920        6.5366          0.6299     0.9205
## parent_HHInc              7.2864        5.9047          0.6445     0.9501
## parent_OwnHome            0.8443        0.7428          0.2801          .
## parent_Race               1.0797        1.1153         -0.1197     0.6858
##                    eCDF Mean eCDF Max
## distance              0.3886   0.6256
## student_PubAff        0.0026   0.0026
## student_Newspaper     0.0580   0.1240
## student_Radio         0.0272   0.0449
## student_TV            0.0147   0.0452
## student_Magazine      0.1331   0.2108
## student_FamTalk       0.0675   0.1018
## student_FrTalk        0.0846   0.1702
## student_AdultTalk     0.0157   0.0436
## student_GovtOpinion   0.0465   0.0977
## student_GovtCrook     0.0133   0.0290
## student_GovtWaste     0.0104   0.0165
## student_TrGovt        0.0344   0.0776
## student_GovtSmart     0.0230   0.0357
## student_Govt4All      0.0144   0.0319
## student_LifeWish      0.0558   0.0853
## student_GLuck         0.0608   0.0966
## student_FPlans        0.0879   0.1339
## student_WinArg        0.0837   0.1453
## student_StrOpinion    0.0358   0.0560
## student_MChange       0.0132   0.0246
## student_TrOthers      0.0597   0.0907
## student_OthHelp       0.0070   0.0162
## student_OthFair       0.0411   0.0676
## student_SchOfficer    0.0816   0.1873
## student_SchPublish    0.0253   0.0509
## student_Hobby         0.0059   0.0094
## student_SchClub       0.0845   0.1765
## student_OccClub       0.0466   0.0815
## student_NeighClub     0.0422   0.0711
## student_RelClub       0.0399   0.0637
```

```
## student_YouthOrg       0.0591   0.1005
## student_MiscClub       0.1472   0.1472
## student_vote           0.2349   0.2349
## student_meeting        0.2080   0.2080
## student_other          0.1157   0.1157
## student_button         0.1079   0.1079
## student_money          0.1238   0.1238
## student_communicate    0.2485   0.2485
## student_demonstrate    0.1689   0.1689
## student_SPID           0.0254   0.0353
## student_Knowledge      0.1815   0.3678
## student_NextSch        0.3124   0.3124
## student_GPA            0.0698   0.2328
## student_Phone          0.0811   0.0811
## student_Gen            0.1156   0.1156
## student_Race           0.0080   0.0174
## parent_Newspaper       0.1147   0.1669
## parent_Radio           0.0346   0.0527
## parent_TV              0.0161   0.0479
## parent_Magazine        0.1922   0.1922
## parent_GovtOpinion     0.0543   0.0966
## parent_GovtCrook       0.0306   0.0676
## parent_GovtWaste       0.0496   0.0971
## parent_TrGovt          0.0168   0.0291
## parent_GovtSmart       0.0163   0.0249
## parent_Govt4All        0.0030   0.0081
## parent_SPID            0.0224   0.0420
## parent_Employ          0.0842   0.0842
## parent_EducHH          0.1925   0.3590
## parent_EducW           0.1346   0.2858
## parent_FInc            0.1255   0.2776
## parent_HHInc           0.1382   0.2613
## parent_OwnHome         0.1015   0.1015
## parent_Race            0.0119   0.0298
##
## Summary of Balance for Matched Data:
##                     Means Treated Means Control Std. Mean Diff. Var. Ratio
## distance                   0.8002        0.7899          0.0474     0.9469
## student_PubAff             0.9465        0.9863         -0.1770          .
## student_Newspaper          1.9427        1.7098          0.1773     1.1153
## student_Radio              2.6401        2.2864          0.2000     1.3577
## student_TV                 2.2167        2.1046          0.0878     1.6172
## student_Magazine           1.6139        1.3798          0.2684     1.2826
## student_FamTalk            1.8431        1.8941         -0.0554     1.3492
## student_FrTalk             1.9938        2.0797         -0.0908     0.9284
## student_AdultTalk          2.9614        2.5293          0.4154     1.0625
## student_GovtOpinion        1.5965        1.3611          0.3696     1.1171
## student_GovtCrook          2.1021        2.0100          0.1372     1.2292
## student_GovtWaste          1.9377        1.9626         -0.0377     1.8002
## student_TrGovt             1.6152        1.4209          0.3165     0.9675
## student_GovtSmart          1.2503        1.1719          0.1200     1.2100
## student_Govt4All           2.7061        2.8344         -0.1897     1.4952
## student_LifeWish           2.1320        2.4134         -0.2839     1.0393
## student_GLuck              1.2279        1.1582          0.1241     1.0528
```

```
## student_FPlans             1.5567        1.4296          0.1428     1.0319
## student_WinArg             2.1370        1.7472          0.4320     0.7932
## student_StrOpinion         1.7796        1.5193          0.2691     1.0735
## student_MChange            1.4408        1.2528          0.2349     1.3034
## student_TrOthers           1.6413        1.5081          0.1436     0.9948
## student_OthHelp            1.7335        1.4707          0.2783     1.0957
## student_OthFair            1.4309        1.3724          0.0719     0.9744
## student_SchOfficer         2.0212        1.7298          0.3517     0.8347
## student_SchPublish         1.7111        2.4919         -0.6802     0.5739
## student_Hobby              1.3873        1.2491          0.1569     1.2177
## student_SchClub            1.7771        2.4670         -0.6494     0.5297
## student_OccClub            1.7161        1.5268          0.1704     1.0059
## student_NeighClub          1.7696        1.5442          0.1981     1.0402
## student_RelClub            2.5567        2.7733         -0.1903     1.3595
## student_YouthOrg           1.7111        2.2391         -0.4809     0.8667
## student_MiscClub           0.3313        0.1619          0.3598          .
## student_vote               0.8070        0.8493         -0.1073          .
## student_meeting            0.2989        0.1469          0.3319          .
## student_other              0.1756        0.4321         -0.6743          .
## student_button             0.3362        0.1980          0.2926          .
## student_money              0.2192        0.5106         -0.7044          .
## student_communicate        0.4060        0.2540          0.3094          .
## student_demonstrate        0.2154        0.1096          0.2575          .
## student_SPID               1.7360        1.5915          0.1496     0.9996
## student_Knowledge          0.6752        0.7219         -0.2051     0.5761
## student_NextSch            0.9577        0.9502          0.0371          .
## student_GPA                2.2740        2.2516          0.0342     1.2191
## student_Phone              0.9614        0.9751         -0.0711          .
## student_Gen                0.5479        0.2839          0.5305          .
## student_Race               1.0822        1.0573          0.0817     1.5076
## parent_Newspaper           3.3674        3.1544          0.1712     1.1203
## parent_Radio               2.4770        2.9315         -0.2527     0.9988
## parent_TV                  3.2130        3.1083          0.0858     1.0717
## parent_Magazine            0.6600        0.7347         -0.1577          .
## parent_GovtOpinion         1.6887        1.8829         -0.2852     1.1865
## parent_GovtCrook           2.0498        1.6600          0.5438     0.7815
## parent_GovtWaste           1.5741        1.4309          0.2405     0.8688
## parent_TrGovt              2.0349        2.3998         -0.5822     0.7142
## parent_GovtSmart           1.5567        2.0847         -0.6034     0.6779
## parent_Govt4All            2.4969        1.9539          0.6637     0.6197
## parent_SPID                2.0100        2.0336         -0.0245     1.2140
## parent_Employ              0.6961        0.7235         -0.0596          .
## parent_EducHH              3.3238        2.8082          0.3517     2.1887
## parent_EducW               3.1071        3.1507         -0.0360     1.0265
## parent_FInc                7.7920        8.0050         -0.1068     1.2284
## parent_HHInc               7.2864        7.6563         -0.1725     1.1712
## parent_OwnHome             0.8443        0.8381          0.0172          .
## parent_Race                1.0797        1.0573          0.0754     1.4333
##                      eCDF Mean eCDF Max Std. Pair Dist.
## distance                0.0368   0.3313          0.0519
## student_PubAff          0.0399   0.0399          0.2987
## student_Newspaper       0.0466   0.1034          0.8828
## student_Radio           0.1106   0.1594          0.9308
## student_TV              0.0631   0.1320          0.8775
```

```
## student_Magazine       0.0780   0.1196          0.7709
## student_FamTalk        0.0607   0.1469          0.9100
## student_FrTalk         0.0172   0.0710          1.0193
## student_AdultTalk      0.1111   0.2441          1.0787
## student_GovtOpinion    0.0785   0.1968          0.9288
## student_GovtCrook      0.0465   0.1158          0.9236
## student_GovtWaste      0.0756   0.1258          0.8246
## student_TrGovt         0.0648   0.1731          1.0105
## student_GovtSmart      0.0262   0.0423          0.5925
## student_Govt4All       0.0428   0.0685          0.6096
## student_LifeWish       0.0938   0.1407          0.9975
## student_GLuck          0.0232   0.0648          0.5939
## student_FPlans         0.0423   0.0685          0.7478
## student_WinArg         0.1299   0.2603          1.0807
## student_StrOpinion     0.0868   0.1357          0.9208
## student_MChange        0.0627   0.1121          0.7981
## student_TrOthers       0.0444   0.0722          0.8360
## student_OthHelp        0.0876   0.1469          0.9351
## student_OthFair        0.0195   0.0311          0.7576
## student_SchOfficer     0.0971   0.1993          1.0941
## student_SchPublish     0.1952   0.2740          1.2312
## student_Hobby          0.0346   0.0697          0.6573
## student_SchClub        0.1725   0.2976          1.2426
## student_OccClub        0.0473   0.1071          0.8858
## student_NeighClub      0.0564   0.1133          0.9139
## student_RelClub        0.0953   0.1868          1.0388
## student_YouthOrg       0.1389   0.2852          1.1069
## student_MiscClub       0.1694   0.1694          0.8573
## student_vote           0.0423   0.0423          0.6942
## student_meeting        0.1519   0.1519          0.8379
## student_other          0.2565   0.2565          0.9819
## student_button         0.1382   0.1382          0.9094
## student_money          0.2914   0.2914          1.0416
## student_communicate    0.1519   0.1519          0.9586
## student_demonstrate    0.1059   0.1059          0.7118
## student_SPID           0.0629   0.1706          1.0754
## student_Knowledge      0.0557   0.2516          1.0000
## student_NextSch        0.0075   0.0075          0.3092
## student_GPA            0.0314   0.0897          0.8332
## student_Phone          0.0137   0.0137          0.2909
## student_Gen            0.2640   0.2640          1.0960
## student_Race           0.0083   0.0162          0.4247
## parent_Newspaper       0.0720   0.2864          0.8421
## parent_Radio           0.0909   0.1519          1.0241
## parent_TV              0.0578   0.1930          0.9069
## parent_Magazine        0.0747   0.0747          0.8307
## parent_GovtOpinion     0.0681   0.1993          1.0384
## parent_GovtCrook       0.1299   0.2827          1.2214
## parent_GovtWaste       0.0486   0.1445          0.9893
## parent_TrGovt          0.1216   0.3064          1.2141
## parent_GovtSmart       0.1760   0.2814          1.1669
## parent_Govt4All        0.1810   0.2877          1.3213
## parent_SPID            0.0520   0.0922          1.0046
## parent_Employ          0.0274   0.0274          0.8881
```

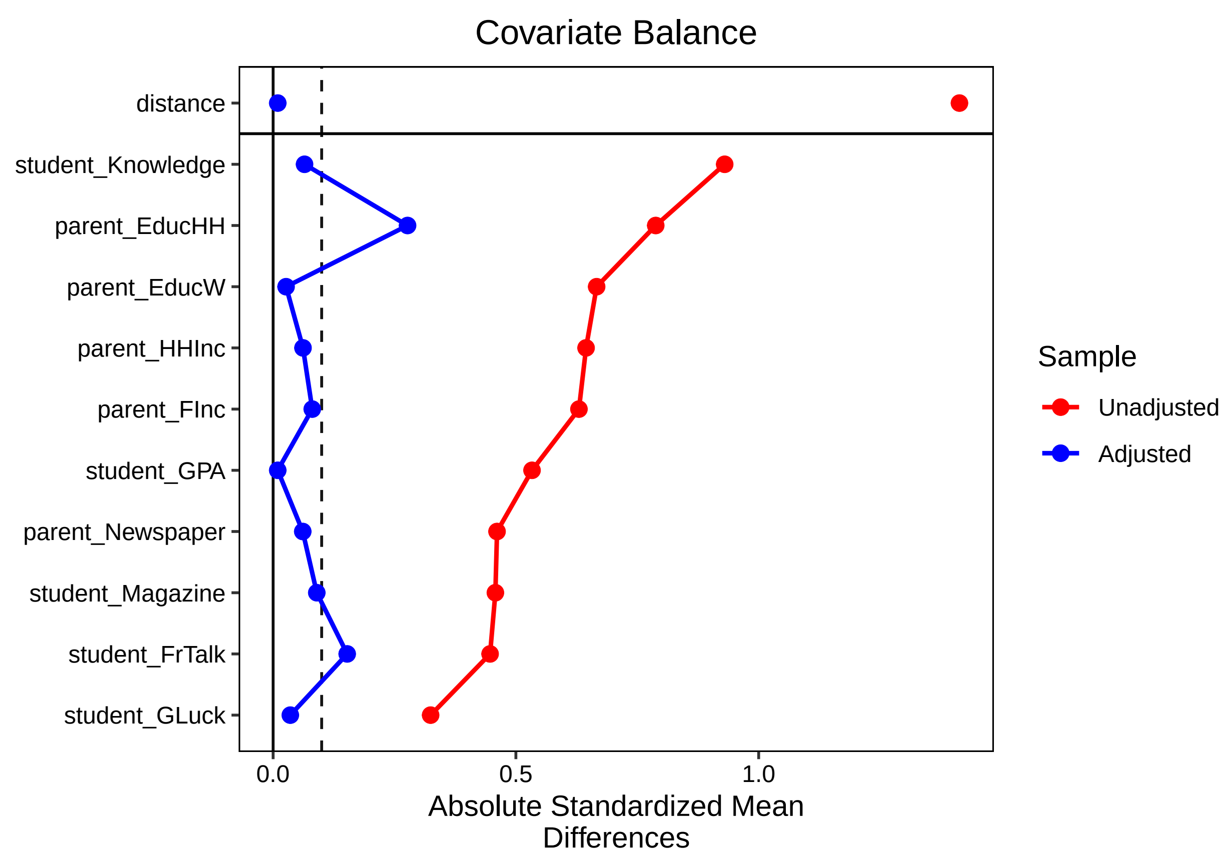
```
## parent_EducHH          0.1046   0.3263          0.9344
## parent_EducW           0.0492   0.1183          0.9440
## parent_FInc            0.0529   0.1582          0.8553
## parent_HHInc           0.0651   0.1457          0.8835
## parent_OwnHome         0.0062   0.0062          0.6698
## parent_Race            0.0075   0.0149          0.4272
##
## Sample Sizes:
##               Control Treated
## All            451.       803
## Matched (ESS)    8.04     803
## Matched        173.       803
## Unmatched      220.         0
## Discarded       58.         0
```

```
##
## Call:
## lm(formula = student_ppnscal ~ college + student_PubAff + student_Newspaper +
##     student_Radio + student_TV + student_Magazine + student_FamTalk +
##     student_FrTalk + student_AdultTalk + student_GovtOpinion +
##     student_GovtCrook + student_GovtWaste + student_TrGovt +
##     student_GovtSmart + student_Govt4All + student_LifeWish +
##     student_GLuck + student_FPlans + student_WinArg + student_StrOpinion +
##     student_MChange + student_TrOthers + student_OthHelp + student_OthFair +
##     student_SchOfficer + student_SchPublish + student_Hobby +
##     student_SchClub + student_OccClub + student_NeighClub + student_RelClub +
##     student_YouthOrg + student_MiscClub + student_vote + student_meeting +
##     student_other + student_button + student_money + student_communicate +
##     student_demonstrate + student_SPID + student_Knowledge +
##     student_NextSch + student_GPA + student_Phone + student_Gen +
##     student_Race + parent_Newspaper + parent_Radio + parent_TV +
##     parent_Magazine + parent_GovtOpinion + parent_GovtCrook +
##     parent_GovtWaste + parent_TrGovt + parent_GovtSmart + parent_Govt4All +
##     parent_SPID + parent_Employ + parent_EducHH + parent_EducW +
##     parent_FInc + parent_HHInc + parent_OwnHome + parent_Race,
##     data = match_ps_att_data, weights = weights)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0954 -0.2886 -0.1254  0.3693  1.2212
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.2384955  0.2868739   0.831   0.4060
## college              0.0408400  0.0429944   0.950   0.3424
## student_PubAff      -0.0460861  0.0684743  -0.673   0.5011
## student_Newspaper    0.0196988  0.0122008   1.615   0.1068
## student_Radio       -0.0056988  0.0085776  -0.664   0.5066
## student_TV           0.0098087  0.0127642   0.768   0.4424
## student_Magazine     0.0183934  0.0183579   1.002   0.3166
## student_FamTalk     -0.0352285  0.0175560  -2.007   0.0451 *
## student_FrTalk      -0.0231950  0.0161284  -1.438   0.1507
## student_AdultTalk   -0.0105346  0.0151027  -0.698   0.4856
## student_GovtOpinion  0.0089027  0.0252407   0.353   0.7244
```

```
## student_GovtCrook     0.0080829  0.0244130   0.331   0.7407
## student_GovtWaste     0.0080515  0.0248509   0.324   0.7460
## student_TrGovt       -0.0002727  0.0268461  -0.010   0.9919
## student_GovtSmart     0.0129327  0.0237430   0.545   0.5861
## student_Govt4All     -0.0188161  0.0253101  -0.743   0.4574
## student_LifeWish     -0.0170072  0.0152503  -1.115   0.2651
## student_GLuck         0.0283480  0.0266037   1.066   0.2869
## student_FPlans       -0.0285708  0.0173223  -1.649   0.0994 .
## student_WinArg        0.0112249  0.0165436   0.679   0.4976
## student_StrOpinion   -0.0269298  0.0166139  -1.621   0.1054
## student_MChange      -0.0043229  0.0195933  -0.221   0.8254
## student_TrOthers      0.0131033  0.0180559   0.726   0.4682
## student_OthHelp       0.0050979  0.0172003   0.296   0.7670
## student_OthFair       0.0118414  0.0206816   0.573   0.5671
## student_SchOfficer   -0.0309204  0.0183627  -1.684   0.0925 .
## student_SchPublish   -0.0087561  0.0134397  -0.652   0.5149
## student_Hobby         0.0330026  0.0172872   1.909   0.0566 .
## student_SchClub       0.0200147  0.0142754   1.402   0.1612
## student_OccClub       0.0067181  0.0140550   0.478   0.6328
## student_NeighClub     0.0267918  0.0135792   1.973   0.0488 *
## student_RelClub      -0.0040665  0.0137474  -0.296   0.7674
## student_YouthOrg      0.0130251  0.0139843   0.931   0.3519
## student_MiscClub      0.0043951  0.0323085   0.136   0.8918
## student_vote          0.9820472  0.0383279  25.622   <2e-16 ***
## student_meeting       1.1601408  0.0381279  30.428   <2e-16 ***
## student_other         0.9823127  0.0438527  22.400   <2e-16 ***
## student_button        1.0474938  0.0353400  29.640   <2e-16 ***
## student_money         1.0127830  0.0386259  26.220   <2e-16 ***
## student_communicate   1.1130532  0.0320098  34.772   <2e-16 ***
## student_demonstrate   1.0921241  0.0384874  28.376   <2e-16 ***
## student_SPID         -0.0187987  0.0155663  -1.208   0.2275
## student_Knowledge     0.0007500  0.0701937   0.011   0.9915
## student_NextSch      -0.0457984  0.0714998  -0.641   0.5220
## student_GPA           0.0063719  0.0257106   0.248   0.8043
## student_Phone        -0.0603777  0.0802518  -0.752   0.4520
## student_Gen           0.0320439  0.0339372   0.944   0.3453
## student_Race          0.0667719  0.1306348   0.511   0.6094
## parent_Newspaper      0.0223949  0.0127517   1.756   0.0794 .
## parent_Radio          0.0175036  0.0082724   2.116   0.0346 *
## parent_TV            -0.0036240  0.0126031  -0.288   0.7738
## parent_Magazine      -0.0587044  0.0340425  -1.724   0.0850 .
## parent_GovtOpinion   -0.0075244  0.0248340  -0.303   0.7620
## parent_GovtCrook     -0.0052826  0.0235589  -0.224   0.8226
## parent_GovtWaste     -0.0011262  0.0285367  -0.039   0.9685
## parent_TrGovt        -0.0231321  0.0270513  -0.855   0.3927
## parent_GovtSmart     -0.0310970  0.0182242  -1.706   0.0883 .
## parent_Govt4All      -0.0375862  0.0220101  -1.708   0.0880 .
## parent_SPID           0.0011836  0.0156494   0.076   0.9397
## parent_Employ         0.0179698  0.0325549   0.552   0.5811
## parent_EducHH         0.0135767  0.0129219   1.051   0.2937
## parent_EducW         -0.0162276  0.0152886  -1.061   0.2888
## parent_FInc           0.0344494  0.0164183   2.098   0.0362 *
## parent_HHInc         -0.0337643  0.0148295  -2.277   0.0230 *
## parent_OwnHome        0.0110147  0.0407902   0.270   0.7872
```

```
## parent_Race          0.1184724  0.1336996   0.886   0.3758
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4297 on 910 degrees of freedom
## Multiple R-squared:  0.9522, Adjusted R-squared:  0.9488
## F-statistic: 278.9 on 65 and 910 DF,  p-value: < 2.2e-16

## [1] 0.04084004

##                          Type     Diff.Un   Diff.Adj        M.Threshold
## distance             Distance 2.059323678 0.04738871     Balanced, <0.1
## student_PubAff         Binary 0.002551408 0.03985056     Balanced, <0.1
## student_Newspaper     Contin. 0.220886032 0.17731475 Not Balanced, >0.1
## student_Radio         Contin. 0.076864977 0.19995884 Not Balanced, >0.1
## student_TV            Contin. 0.067229599 0.08775282     Balanced, <0.1
## student_Magazine      Contin. 0.457791084 0.26837950 Not Balanced, >0.1
## student_FamTalk       Contin. 0.293165642 0.05544059     Balanced, <0.1
## student_FrTalk        Contin. 0.446840653 0.09075398     Balanced, <0.1
## student_AdultTalk     Contin. 0.034981141 0.41542295 Not Balanced, >0.1
## student_GovtOpinion   Contin. 0.219250263 0.36958240 Not Balanced, >0.1
## student_GovtCrook     Contin. 0.059255547 0.13723766 Not Balanced, >0.1
## student_GovtWaste     Contin. 0.047206028 0.03765461     Balanced, <0.1
## student_TrGovt        Contin. 0.168166561 0.31653718 Not Balanced, >0.1
## student_GovtSmart     Contin. 0.105525115 0.12002262 Not Balanced, >0.1
## student_Govt4All      Contin. 0.064104981 0.18970941 Not Balanced, >0.1
## student_LifeWish      Contin. 0.168808918 0.28393237 Not Balanced, >0.1
## student_GLuck         Contin. 0.324392154 0.12409270 Not Balanced, >0.1
## student_FPlans        Contin. 0.296588359 0.14284608 Not Balanced, >0.1
## student_WinArg        Contin. 0.278225433 0.43199843 Not Balanced, >0.1
## student_StrOpinion    Contin. 0.111001289 0.26914829 Not Balanced, >0.1
## student_MChange       Contin. 0.049363822 0.23490501 Not Balanced, >0.1
## student_TrOthers      Contin. 0.192941013 0.14358477 Not Balanced, >0.1
## student_OthHelp       Contin. 0.012192238 0.27829828 Not Balanced, >0.1
## student_OthFair       Contin. 0.151715294 0.07193784     Balanced, <0.1
## student_SchOfficer    Contin. 0.295563312 0.35168478 Not Balanced, >0.1
## student_SchPublish    Contin. 0.088263717 0.68015449 Not Balanced, >0.1
## student_Hobby         Contin. 0.014272526 0.15689437 Not Balanced, >0.1
## student_SchClub       Contin. 0.318225555 0.64943174 Not Balanced, >0.1
## student_OccClub       Contin. 0.167809793 0.17043524 Not Balanced, >0.1
## student_NeighClub     Contin. 0.148283895 0.19809447 Not Balanced, >0.1
## student_RelClub       Contin. 0.140288146 0.19026913 Not Balanced, >0.1
## student_YouthOrg      Contin. 0.215458028 0.48087333 Not Balanced, >0.1
## student_MiscClub       Binary 0.147222307 0.16936488 Not Balanced, >0.1
## student_vote           Binary 0.234911764 0.04234122     Balanced, <0.1
## student_meeting        Binary 0.207970112 0.15193026 Not Balanced, >0.1
## student_other          Binary 0.115724569 0.25653798 Not Balanced, >0.1
## student_button         Binary 0.107857729 0.13823163 Not Balanced, >0.1
## student_money          Binary 0.123834401 0.29140722 Not Balanced, >0.1
## student_communicate    Binary 0.248549646 0.15193026 Not Balanced, >0.1
## student_demonstrate    Binary 0.168878899 0.10585305 Not Balanced, >0.1
## student_SPID          Contin. 0.036891500 0.14957884 Not Balanced, >0.1
## student_Knowledge     Contin. 0.930012273 0.20509583 Not Balanced, >0.1
## student_NextSch        Binary 0.312425964 0.00747198     Balanced, <0.1
```

```
## student_GPA         Contin. 0.533230496 0.03424033     Balanced, <0.1
## student_Phone        Binary 0.081128694 0.01369863     Balanced, <0.1
## student_Gen          Binary 0.115572700 0.26400996 Not Balanced, >0.1
## student_Race        Contin. 0.035858544 0.08167780     Balanced, <0.1
## parent_Newspaper    Contin. 0.461170461 0.17121723 Not Balanced, >0.1
## parent_Radio        Contin. 0.096294816 0.25272745 Not Balanced, >0.1
## parent_TV           Contin. 0.041746212 0.08578398     Balanced, <0.1
## parent_Magazine      Binary 0.192175683 0.07471980     Balanced, <0.1
## parent_GovtOpinion  Contin. 0.238947605 0.28518546 Not Balanced, >0.1
## parent_GovtCrook    Contin. 0.128276664 0.54383034 Not Balanced, >0.1
## parent_GovtWaste    Contin. 0.249803187 0.24051910 Not Balanced, >0.1
## parent_TrGovt       Contin. 0.080401209 0.58219393 Not Balanced, >0.1
## parent_GovtSmart    Contin. 0.055881583 0.60338226 Not Balanced, >0.1
## parent_Govt4All     Contin. 0.011101358 0.66370525 Not Balanced, >0.1
## parent_SPID         Contin. 0.092974650 0.02450178     Balanced, <0.1
## parent_Employ        Binary 0.084166085 0.02739726     Balanced, <0.1
## parent_EducHH       Contin. 0.787999221 0.35166291 Not Balanced, >0.1
## parent_EducW        Contin. 0.666274681 0.03595188     Balanced, <0.1
## parent_FInc         Contin. 0.629859221 0.10683819 Not Balanced, >0.1
## parent_HHInc        Contin. 0.644522314 0.17252150 Not Balanced, >0.1
## parent_OwnHome       Binary 0.101539957 0.00622665     Balanced, <0.1
## parent_Race         Contin. 0.119727784 0.07539173     Balanced, <0.1
```



Report the overall balance and the proportion of covariates that meet the balance threshold:

The top 10 covariates plotted were the covariates with teh largest absolute SMD after matching.The plot shows improvement in that all (100%) the unadjusted covariates were above the 0.1 threshold line, but after adjusting/matching, all but two covariates (or 80%) were below the 0.1 threshold.

Simulations

Henderson/Chatfield argue that an improperly specified propensity score model can actually *increase* the bias of the estimate. To demonstrate this, they simulate 800,000 different propensity score models by choosing different permutations of covariates. To investigate their claim, do the following:

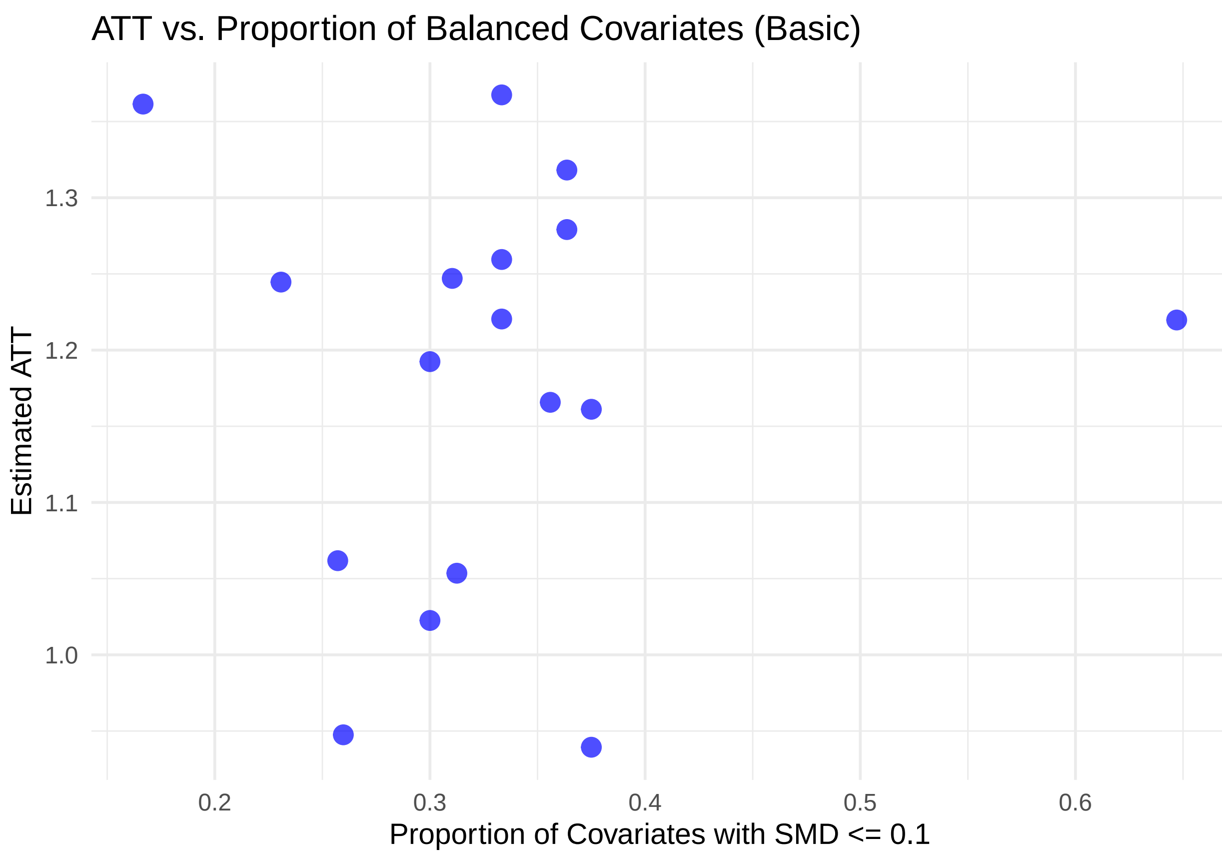
- Using as many simulations as is feasible (20-30 should be ok, more is better!), randomly select the number of and the choice of covariates for the propensity score model.
- For each run, store the ATT, the proportion of covariates that meet the standardized mean difference $\leq .1$ threshold, and the mean percent improvement in the standardized mean difference. You may also wish to store the entire models in a list and extract the relevant attributes as necessary.
- Plot all of the ATTs against all of the balanced covariate proportions. You may randomly sample or use other techniques like transparency if you run into overplotting problems. Alternatively, you may use plots other than scatterplots, so long as you explore the relationship between ATT and the proportion of covariates that meet the balance threshold.
- Finally choose 10 random models and plot their covariate balance plots (you may want to use a library like gridExtra to arrange these)

Note: There are lots of post-treatment covariates in this dataset (about 50!)! You need to be careful not to include these in the pre-treatment balancing. Many of you are probably used to selecting or dropping columns manually, or positionally. However, you may not always have a convenient arrangement of columns, nor is it fun to type out 50 different column names. Instead see if you can use dplyr 1.0.0 functions to programatically drop post-treatment variables (here is a useful tutorial).

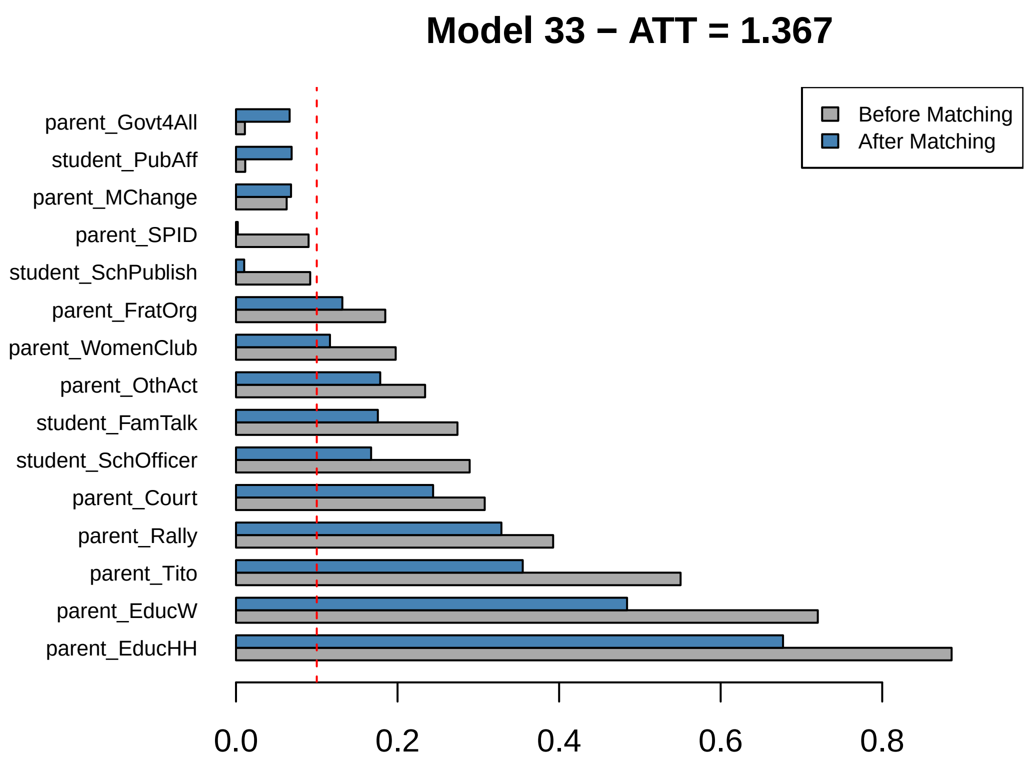
```
## Running simulation 1 of 50
##   - Valid result with ATT = 1.318 and prop_balanced = 0.364
## Running simulation 2 of 50
##   - Valid result with ATT = 1.279 and prop_balanced = 0.364
## Running simulation 3 of 50
##   - Invalid result, skipping
## Running simulation 4 of 50
##   - Invalid result, skipping
## Running simulation 5 of 50
##   - Invalid result, skipping
## Running simulation 6 of 50
##   - Invalid result, skipping
## Running simulation 7 of 50
##   - Valid result with ATT = 1.192 and prop_balanced = 0.3
## Running simulation 8 of 50
##   - Valid result with ATT = 0.939 and prop_balanced = 0.375
## Running simulation 9 of 50
##   - Invalid result, skipping
## Running simulation 10 of 50
##   - Invalid result, skipping
## Running simulation 11 of 50
##   - Invalid result, skipping
## Running simulation 12 of 50
```

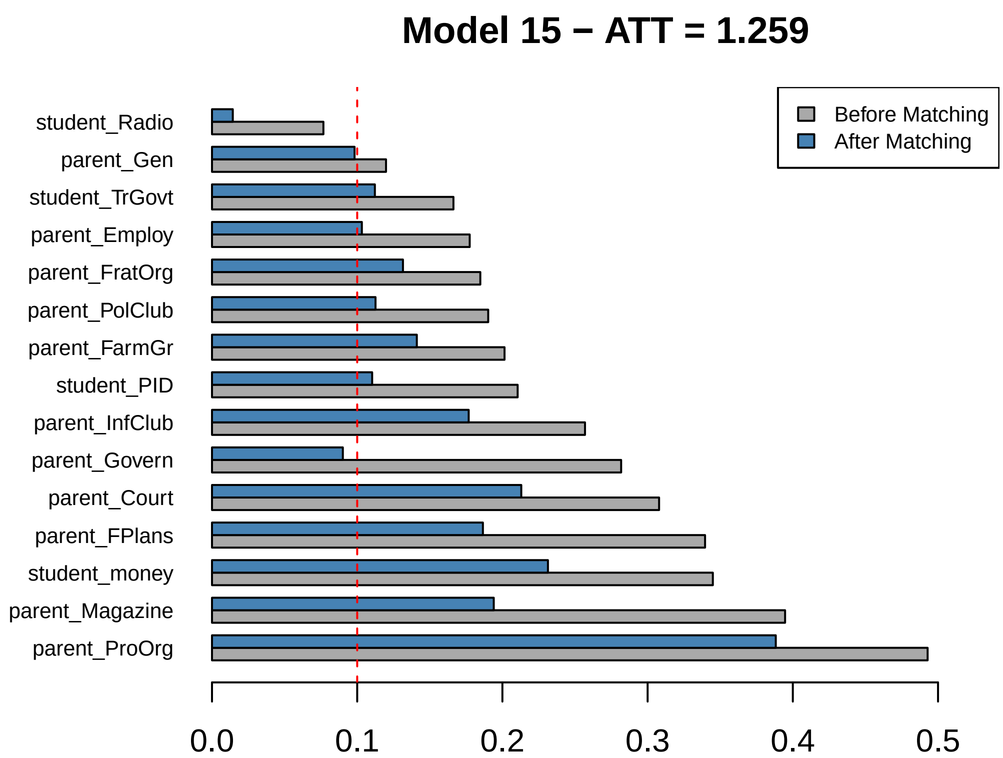
```
##   - Invalid result, skipping
## Running simulation 13 of 50
##   - Invalid result, skipping
## Running simulation 14 of 50
##   - Invalid result, skipping
## Running simulation 15 of 50
##   - Valid result with ATT = 1.259 and prop_balanced = 0.333
## Running simulation 16 of 50
##   - Valid result with ATT = 1.22 and prop_balanced = 0.647
## Running simulation 17 of 50
##   - Valid result with ATT = 1.161 and prop_balanced = 0.375
## Running simulation 18 of 50
##   - Valid result with ATT = 1.062 and prop_balanced = 0.257
## Running simulation 19 of 50
##   - Invalid result, skipping
## Running simulation 20 of 50
##   - Invalid result, skipping
## Running simulation 21 of 50
##   - Invalid result, skipping
## Running simulation 22 of 50
##   - Invalid result, skipping
## Running simulation 23 of 50
##   - Invalid result, skipping
## Running simulation 24 of 50
##   - Valid result with ATT = 0.948 and prop_balanced = 0.26
## Running simulation 25 of 50
##   - Valid result with ATT = 1.361 and prop_balanced = 0.167
## Running simulation 26 of 50
##   - Invalid result, skipping
## Running simulation 27 of 50
##   - Valid result with ATT = 1.245 and prop_balanced = 0.231
## Running simulation 28 of 50
##   - Invalid result, skipping
## Running simulation 29 of 50
##   - Invalid result, skipping
## Running simulation 30 of 50
##   - Valid result with ATT = 1.22 and prop_balanced = 0.333
## Running simulation 31 of 50
##   - Invalid result, skipping
## Running simulation 32 of 50
##   - Invalid result, skipping
## Running simulation 33 of 50
##   - Valid result with ATT = 1.367 and prop_balanced = 0.333
## Running simulation 34 of 50
##   - Invalid result, skipping
## Running simulation 35 of 50
##   - Invalid result, skipping
## Running simulation 36 of 50
##   - Valid result with ATT = 1.247 and prop_balanced = 0.31
## Running simulation 37 of 50
##   - Invalid result, skipping
## Running simulation 38 of 50
##   - Invalid result, skipping
## Running simulation 39 of 50
```

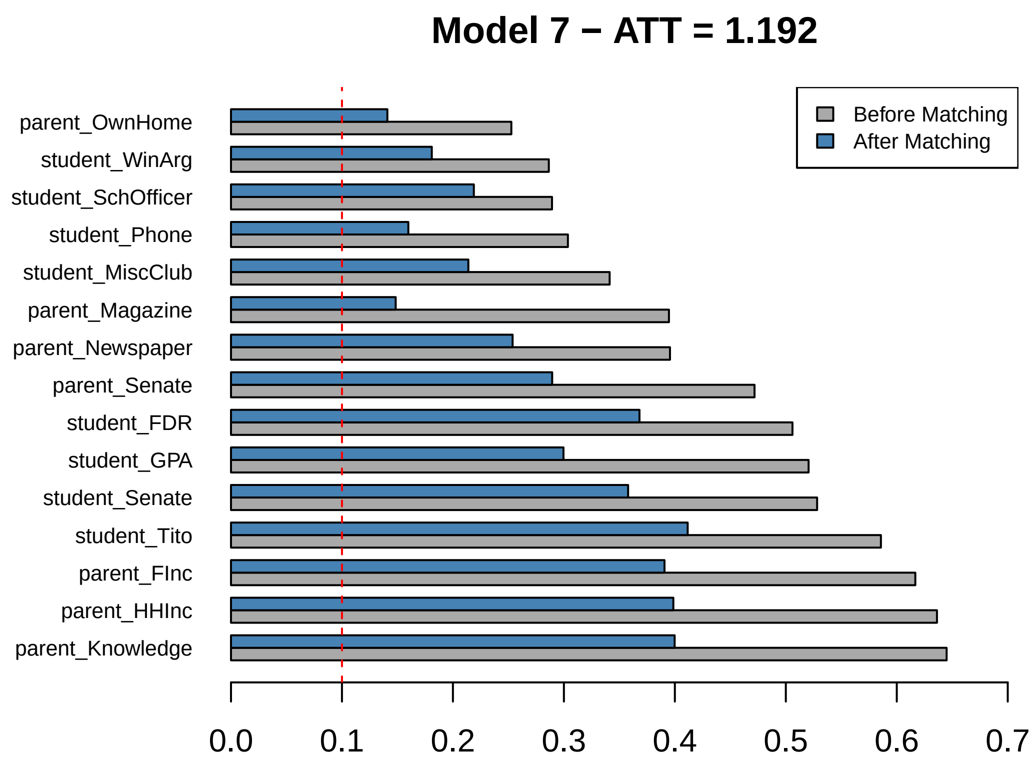
```
##   - Invalid result, skipping
## Running simulation 40 of 50
##   - Invalid result, skipping
## Running simulation 41 of 50
##   - Invalid result, skipping
## Running simulation 42 of 50
##   - Valid result with ATT = 1.023 and prop_balanced = 0.3
## Running simulation 43 of 50
##   - Invalid result, skipping
## Running simulation 44 of 50
##   - Invalid result, skipping
## Running simulation 45 of 50
##   - Invalid result, skipping
## Running simulation 46 of 50
##   - Invalid result, skipping
## Running simulation 47 of 50
##   - Valid result with ATT = 1.166 and prop_balanced = 0.356
## Running simulation 48 of 50
##   - Invalid result, skipping
## Running simulation 49 of 50
##   - Valid result with ATT = 1.054 and prop_balanced = 0.312
## Running simulation 50 of 50
##   - Invalid result, skipping

## Created results_df with 17 valid rows

##    sim_id       att prop_balanced pct_improvement num_covariates
## 1       1 1.3181621     0.3636364        8.157357             33
## 2       2 1.2791052     0.3636364       36.064244             11
## 7       7 1.1924349     0.3000000       26.002814             40
## 8       8 0.9393426     0.3750000       27.374756             80
## 15     15 1.2594674     0.3333333       22.617636             18
## 16     16 1.2197458     0.6470588       23.092722             17

## [1] "Creating ATT vs. Balance plot..."

## [1] "Diagnostic information for plotting:"

## [1] "Number of rows in clean_results_df: 17"

## [1] "Head of clean_results_df:"
##    sim_id       att prop_balanced num_covariates
## 1       1 1.3181621     0.3636364             33
## 2       2 1.2791052     0.3636364             11
## 7       7 1.1924349     0.3000000             40
## 8       8 0.9393426     0.3750000             80
## 15     15 1.2594674     0.3333333             18
## 16     16 1.2197458     0.6470588             17
## [1] "Range of ATT values: 0.939342633663096 to 1.36749396793273"
## [1] "Range of prop_balanced values: 0.166666666666667 to 0.647058823529412"
```

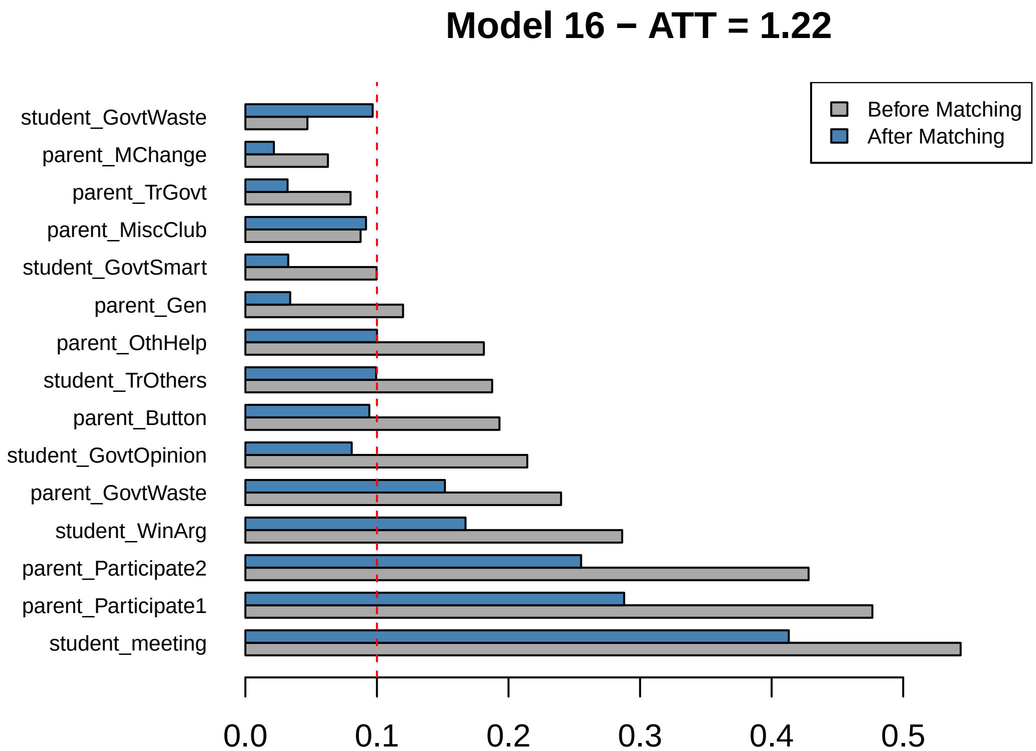


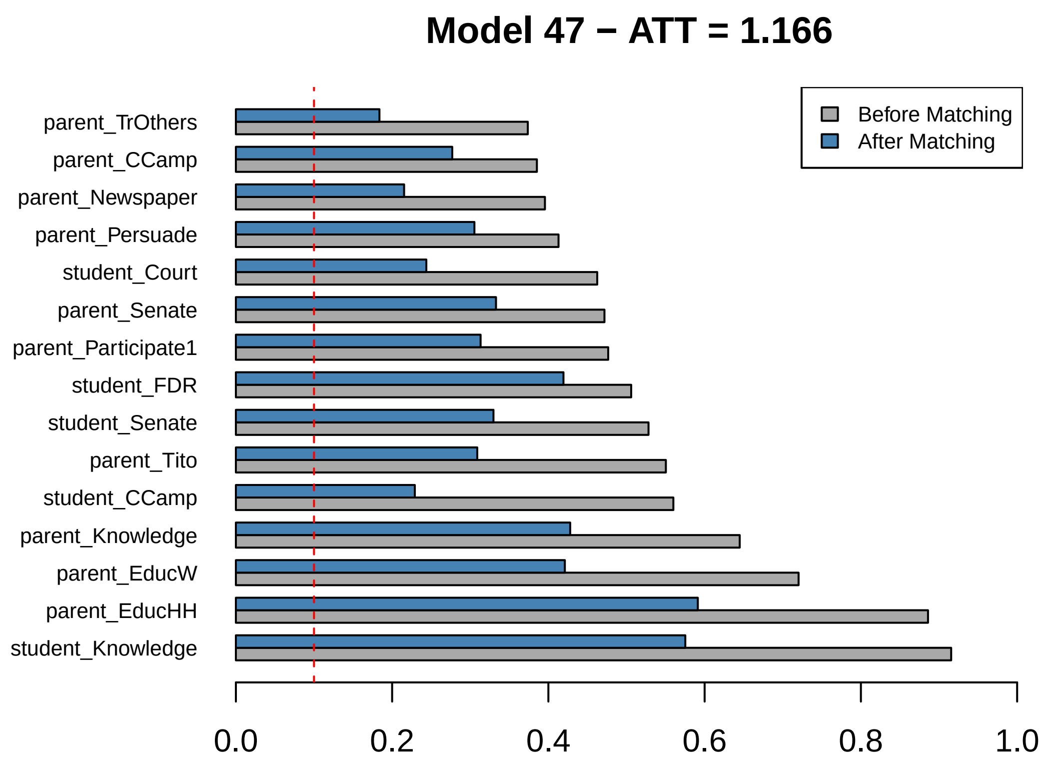

```
## Found 17 valid models out of 50
```


Model 33 − ATT = 1.367
Before Matching
After Matching
parent_Govt4All
student_PubAff
parent_MChange
parent_SPID
student_SchPublish
parent_FratOrg
parent_WomenClub
parent_OthAct
student_FamTalk
student_SchOfficer
parent_Court
parent_Rally
parent_Tito
parent_EducW
parent_EducHH
0.0
0.2
0.4
0.6
0.8


Model 15 − ATT = 1.259
Before Matching
After Matching
student_Radio
parent_Gen
student_TrGovt
parent_Employ
parent_FratOrg
parent_PolClub
parent_FarmGr
student_PID
parent_InfClub
parent_Govern
parent_Court
parent_FPlans
student_money
parent_Magazine
parent_ProOrg
0.0
0.1
0.2
0.3
0.4
0.5


Model 7 − ATT = 1.192
Before Matching
After Matching
parent_OwnHome
student_WinArg
student_SchOfficer
student_Phone
student_MiscClub
parent_Magazine
parent_Newspaper
parent_Senate
student_FDR
student_GPA
student_Senate
student_Tito
parent_FInc
parent_HHInc
parent_Knowledge
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7


Model 16 − ATT = 1.22
Before Matching
After Matching
student_GovtWaste
parent_MChange
parent_TrGovt
parent_MiscClub
student_GovtSmart
parent_Gen
parent_OthHelp
student_TrOthers
parent_Button
student_GovtOpinion
parent_GovtWaste
student_WinArg
parent_Participate2
parent_Participate1
student_meeting
0.0
0.1
0.2
0.3
0.4
0.5



Questions

1. **How many simulations resulted in models with a higher proportion of balanced covariates? Do you have any concerns about this?** Your Answer: Of the simulations that were able to run (17), most of the simulations have a proportion of balanced covariates between 0.2 and 0.4, with a few models reaching higher proportions (around 0.5-0.6). This is relatively low, but it seems that randomly selecting covariates doesn't lead to well-balanced models, which is maybe not too concerning. This seems to reinforce Henderson and Chatfield's claim that how we specify our models matters/ I need to be careful about which covariates to include.

2. **Analyze the distribution of the ATTs. Do you have any concerns about this distribution?** Your Answer: ATT estimates range from 0.9 to 1.4, which is pretty wide. This also aligns with Henderson adn Chatfield because it shows tht model specification can affect the results/findings. Without knowing the true treatment effect, it's hard to determine which estimates are more accurate but because there isn't really a convergence around one value, the results show that model specification really matters.

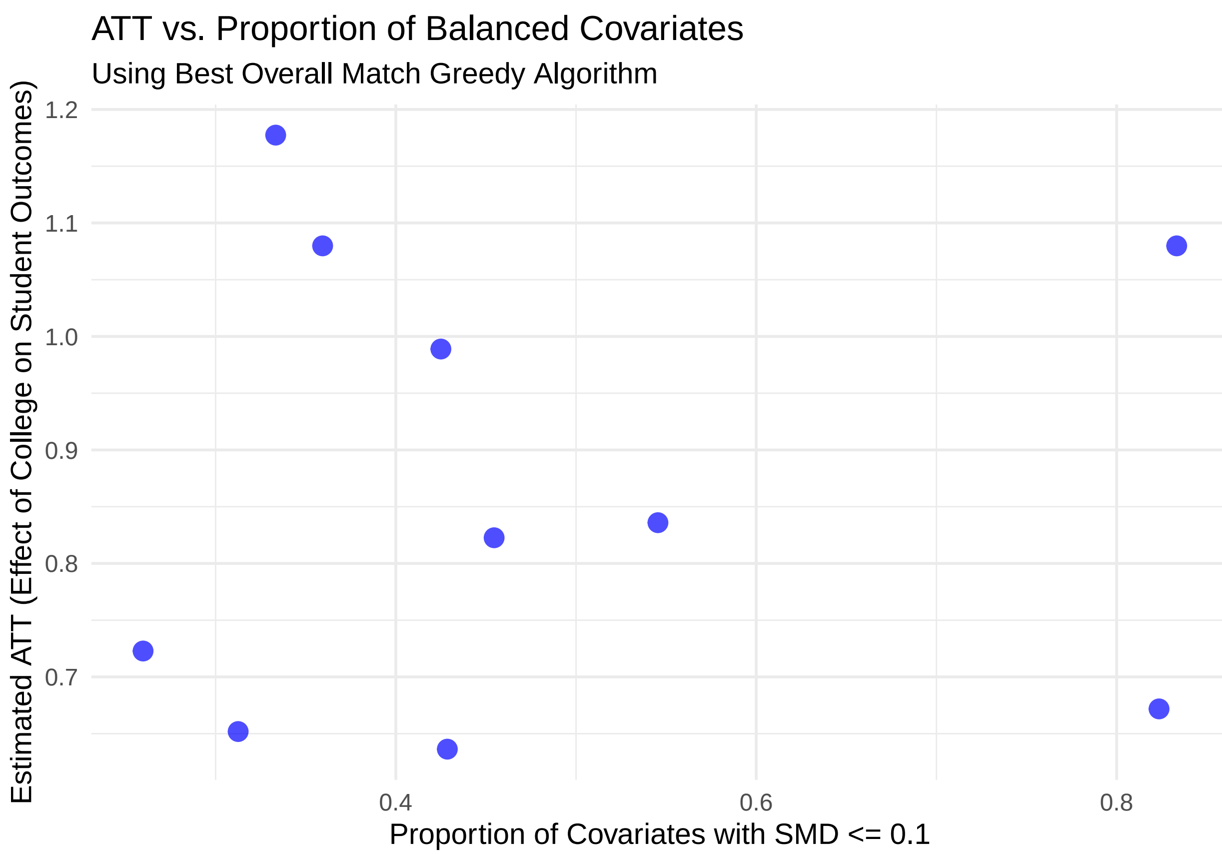
3. **Do your 5 randomly chosen covariate balance plots produce similar numbers on the same covariates? Is it a concern if they do not?** Your Answer: The different models used different sets of covariates becuase my simulation randomly selected covariates for each model and they also show different numbers for the same covariates (for example parentEducW and parentEducHH appear across multiple models and have different numbers). I don't think it's a concern if they do not have the same numbers because we set up the simulation to randomly select different covariates for each model and this confirms Henderson and Chatfield's claim that how we specify our models matters. This is informative because it shows which covariates are consistently imbalanced before matching.

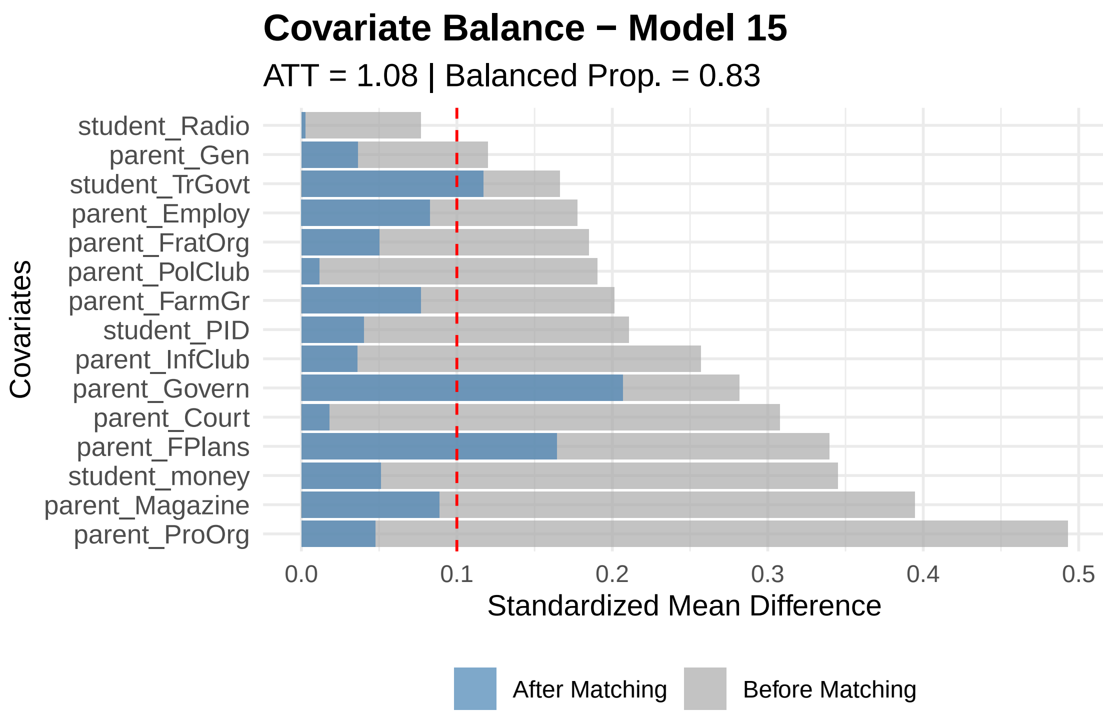
Matching Algorithm of Your Choice

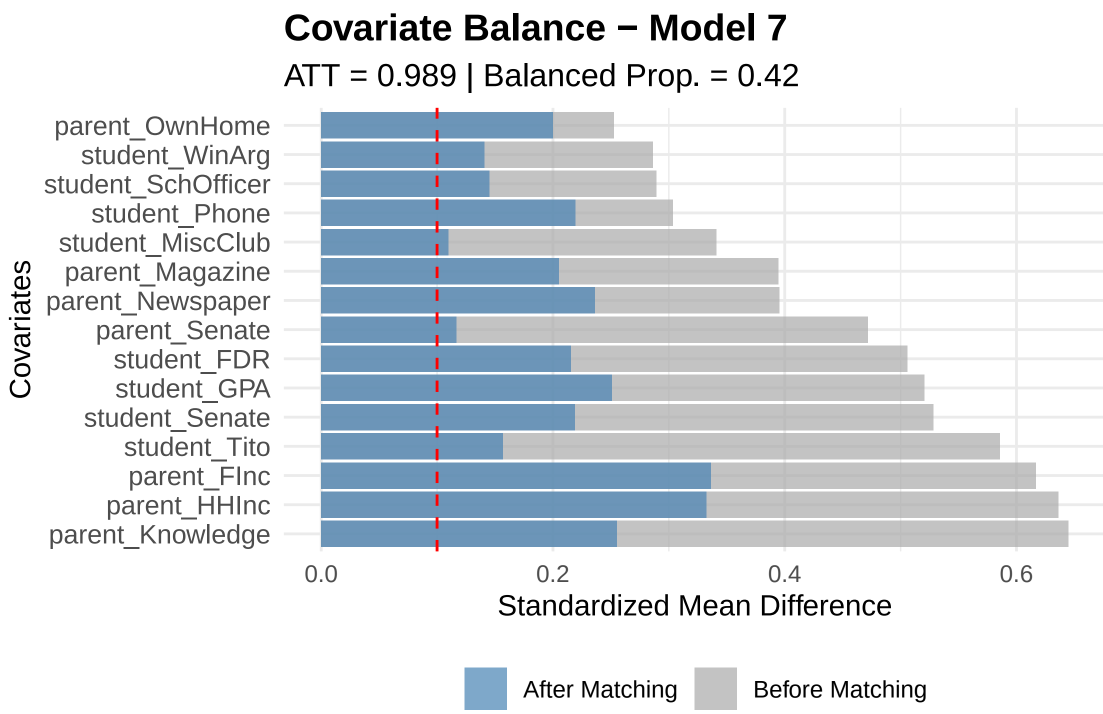
Simulate Alternative Model

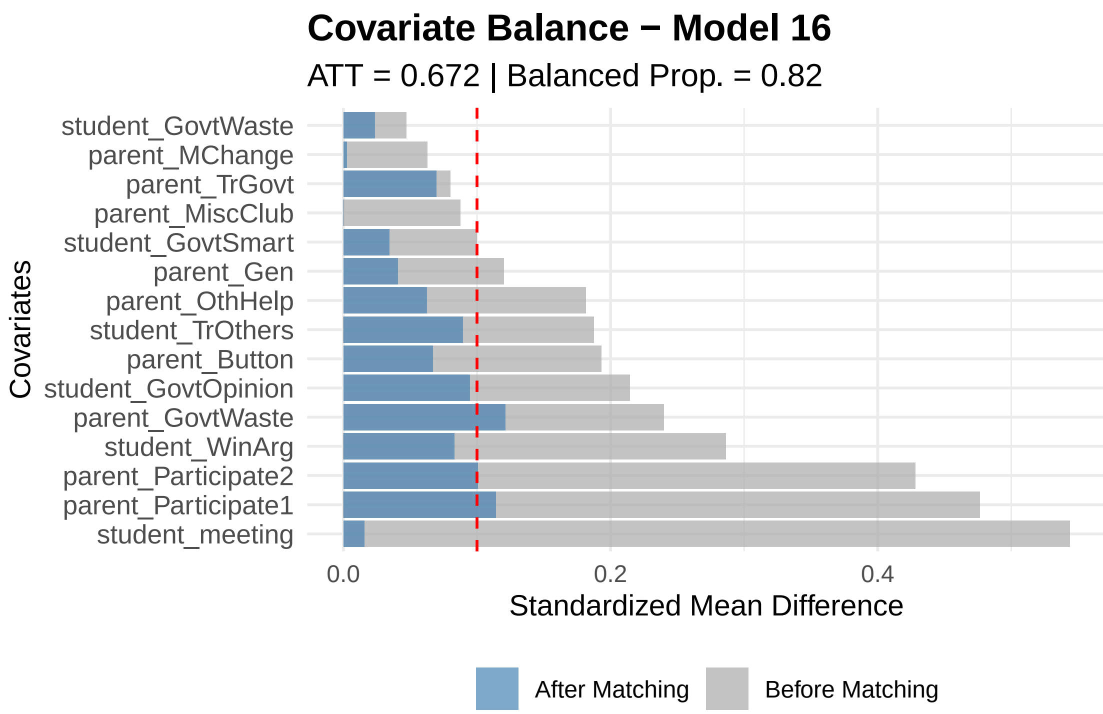
Henderson/Chatfield propose using genetic matching to learn the best weights for Mahalanobis distance matching. Choose a matching algorithm other than the propensity score (you may use genetic matching if you wish, but it is also fine to use the greedy or optimal algorithms we covered in lab instead). Repeat the same steps as specified in Section 4.2 and answer the following questions:

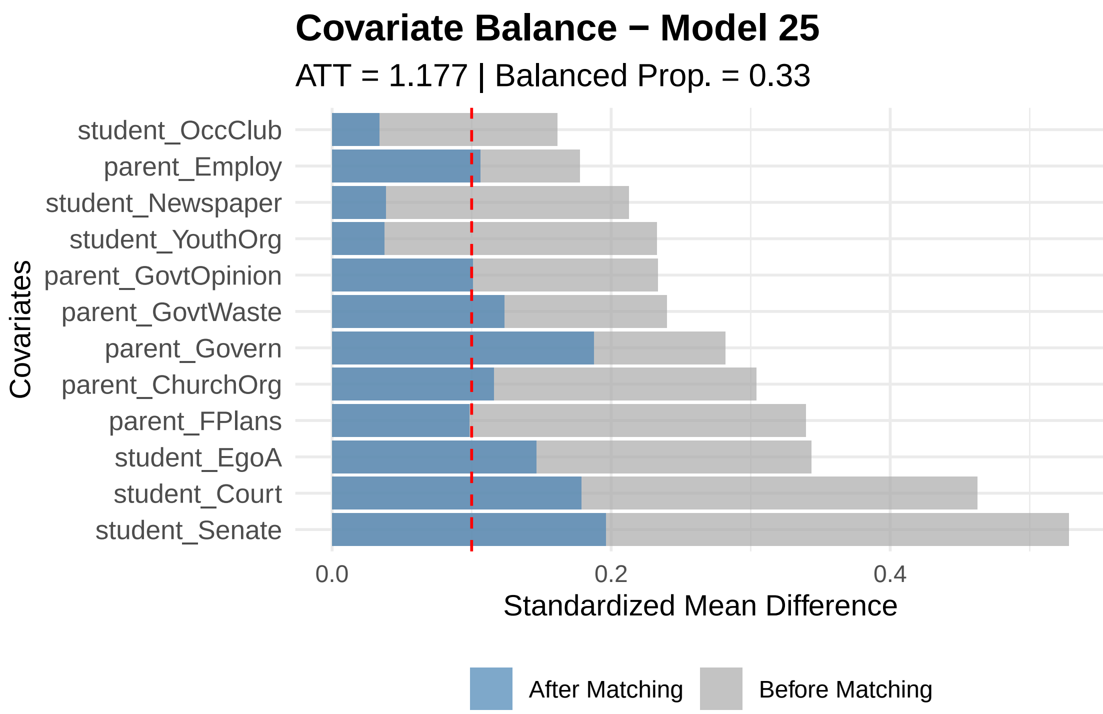
```
## Running simulation 1 of 25 using best overall match
## Running simulation 2 of 25 using best overall match
## Running simulation 3 of 25 using best overall match
## Running simulation 4 of 25 using best overall match
## Running simulation 5 of 25 using best overall match
## Running simulation 6 of 25 using best overall match
## Running simulation 7 of 25 using best overall match
## Running simulation 8 of 25 using best overall match
## Running simulation 9 of 25 using best overall match
## Running simulation 10 of 25 using best overall match
## Running simulation 11 of 25 using best overall match
## Running simulation 12 of 25 using best overall match
## Running simulation 13 of 25 using best overall match
## Running simulation 14 of 25 using best overall match
## Running simulation 15 of 25 using best overall match
## Running simulation 16 of 25 using best overall match
## Running simulation 17 of 25 using best overall match
## Running simulation 18 of 25 using best overall match
## Running simulation 19 of 25 using best overall match
## Running simulation 20 of 25 using best overall match
## Running simulation 21 of 25 using best overall match
## Running simulation 22 of 25 using best overall match
## Running simulation 23 of 25 using best overall match
## Running simulation 24 of 25 using best overall match
## Running simulation 25 of 25 using best overall match

## Created results_df with 10 valid rows

##    sim_id       att prop_balanced pct_improvement num_covariates
## 1       1 0.8226164     0.4545455        30.66677             33
## 2       2 0.8359202     0.5454545        60.38961             11
## 7       7 0.9889135     0.4250000        43.47125             40
## 8       8 0.6518847     0.3125000        37.47280             80
## 15     15 1.0798226     0.8333333        52.18268             18
## 16     16 0.6718404     0.8235294        45.61239             17
```


ATT vs. Proportion of Balanced Covariates
Using Best Overall Match Greedy Algorithm
Estimated ATT (Effect of College on Student Outcomes)
1.2
1.1
1.0
0.9
0.8
0.7
0.4
0.6
0.8
Proportion of Covariates with SMD <= 0.1

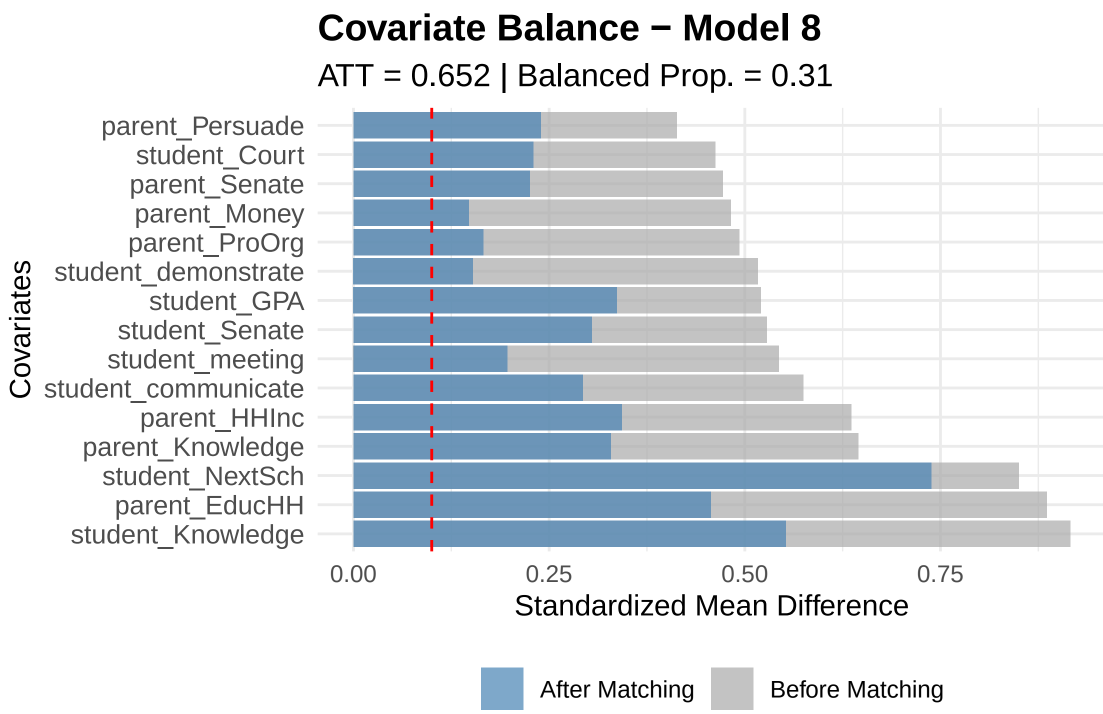

Covariate Balance − Model 15
ATT = 1.08 | Balanced Prop. = 0.83
Covariates
student_Radio
parent_Gen
student_TrGovt
parent_Employ
parent_FratOrg
parent_PolClub
parent_FarmGr
student_PID
parent_InfClub
parent_Govern
parent_Court
parent_FPlans
student_money
parent_Magazine
parent_ProOrg
0.0
0.1
0.2
0.3
0.4
0.5
Standardized Mean Difference
After Matching
Before Matching


Covariate Balance − Model 7
ATT = 0.989 | Balanced Prop. = 0.42
Covariates
parent_OwnHome
student_WinArg
student_SchOfficer
student_Phone
student_MiscClub
parent_Magazine
parent_Newspaper
parent_Senate
student_FDR
student_GPA
student_Senate
student_Tito
parent_FInc
parent_HHInc
parent_Knowledge
0.0
0.2
0.4
0.6
Standardized Mean Difference
After Matching
Before Matching


Covariate Balance − Model 16
ATT = 0.672 | Balanced Prop. = 0.82
Covariates
student_GovtWaste
parent_MChange
parent_TrGovt
parent_MiscClub
student_GovtSmart
parent_Gen
parent_OthHelp
student_TrOthers
parent_Button
student_GovtOpinion
parent_GovtWaste
student_WinArg
parent_Participate2
parent_Participate1
student_meeting
0.0
0.2
0.4
Standardized Mean Difference
After Matching
Before Matching



Covariate Balance − Model 25
ATT = 1.177 | Balanced Prop. = 0.33
Covariates
student_OccClub
parent_Employ
student_Newspaper
student_YouthOrg
parent_GovtOpinion
parent_GovtWaste
parent_Govern
parent_ChurchOrg
parent_FPlans
student_EgoA
student_Court
student_Senate
0.0
0.2
0.4
Standardized Mean Difference
After Matching
Before Matching




```
## [1] "Summary Statistics for Best Overall Match:"

## $att_mean
## [1] 0.8667406
##
## $att_sd
## [1] 0.2008484
##
## $prop_balanced_mean
## [1] 0.4775383
##
## $prop_balanced_sd
## [1] 0.2016682
##
## $pct_improvement_mean
## [1] 45.81257
##
## $pct_improvement_sd
## [1] 9.846005
```

Questions

1. **Does your alternative matching method have more runs with higher proportions of balanced covariates?** Your Answer: Looking at the two plots/descriptives, I see that this greedy

approach does have more runs with higher proportions of balanced covariates compared to the original approach. It shows a wider spread of balanced covariate proportions, with more points (than the previous approach) at higher proportion values (reaching up to 0.8). The mean improvement in the proportion of balanced covariates is 0.458.

2. **Use a visualization to examine the change in the distribution of the percent improvement in balance in propensity score matching vs. the distribution of the percent improvement in balance in your new method. Which did better? Analyze the results in 1-2 sentences.** Your Answer: Based on the two scatter plots, my propensity score plot shows the baseline propensity score matching method with most points concentrated between 0.2-0.4 proportion of balanced covariates. My Greedy Algorithm plot shows that my alternative method has points distributed across higher proportions, reaching up to 0.8. The greedy matching method outperformed standard propensity score matching with a mean improvement of 45.8

Optional: Looking ahead to the discussion questions, you may choose to model the propensity score using an algorithm other than logistic regression and perform these simulations again, if you wish to explore the second discussion question further.

Discussion Questions

1. **Why might it be a good idea to do matching even if we have a randomized or as-if-random design?** Your Answer: Matching can still be beneficial in randomized designs because even with randomization, chance imbalances in covariates may occur (as I found in the earlier part of this project), especially in small samples. Matching helps ensure similar covariate distributions between treatment and control groups, potentially reducing standard errors and increasing power. It could also provide robustness checks and increase the credibility of causal claims by demonstrating consistent results across different estimation approaches.

2. **The standard way of estimating the propensity score is using a logistic regression to estimate probability of treatment. Given what we know about the curse of dimensionality, do you think there might be advantages to using other machine learning algorithms (decision trees, bagging/boosting forests, ensembles, etc.) to estimate propensity scores instead?** Your Answer: I think machine learning methods for propensity score estimation would offer significant advantages over logistic regression, especially if we are working with high-dimensional data. During this project, it occurred to me that it would be helpful to use some of these methods to address the concern about model specification. Unlike logistic regression, which suffers from the curse of dimensionality and requires explicit specification of functional forms, methods like random forests and gradient boosting could capture non-linear relationships and interactions between covariates. From our previous classes, we learned that these algorithms can handle many variables efficiently, detect important predictors automatically, and are less prone to overfitting when properly tuned. I would imagine that machine learning-based propensity scores would help achieve better covariate balance, especially if the treatment assignment mechanism is complex. However, interpretability might then be a concern (especially for certain audiences such as the policy space).