EDA

Statistics 139 Teaching Staff

```
# load Packages
require(pacman)
```

```
## Loading required package: pacman
```

```
p_load(haven, dplyr, corrplot, vioplot, effects, sjPlot, emmeans, tidyverse,
       ggthemes, performance, boot, visreg, vtree,
       glmmTMB, rstanarm, lme4, ggpubr,
       install = TRUE)
# load data
data <- read_dta("sec matched teachers & students.dta")

# overview
# str(data)
table(data$source)
```

```
##
## master_secondary_compact       master_teacher.dta
##                    32561                     1167
```

```
table(data$version)
```

```
##
##     4   4.5     5   5.5
## 19256   905  7537  4863
```

```
table(data$year)
```

```
##
##  2007  2008  2009
## 13903 17657  1001
```

```
# subset data
perception <- data[data$source == "master_secondary_compact" & data$version == 4,]
```

For this project, I will focus on `master_secondary_compact` data from version 4 (19256 observations).

Some broad characteristics of the subsetted survey dataset: All surveys were conducted on paper. Most observations are from 2007 (13,000 observations), some are from 2008 (6000 observations), and a few from 2009 (250 observations). The data includes 51 schools from 10 states: AZ, CA, IL, MA, MO, NJ, NM, NY, OH, OR.

The data is at the student level (can be aggregated at the class, school, and state levels). The `a` variables are about how the students perceive themselves in class. The `b` variables are about how the students perceive the teachers in class. The `c` variables are about the demographic and social life of students; some are about how race influences the experience of the students. The `t` variables are from the perspectives of teachers.

```
table(perception$type)
```

```
##
## Paper
## 19256
```

```
table(perception$year)
```

```
##
##  2007  2008  2009
## 12998  6001   257
```

```
table(perception$state)
```

```
##
##   AZ   CA   IL   MA   MO   NJ   NM   NY   OH   OR
## 1424 3287  501  290  472 1096  228 8083 1846 2029
```

```
length(unique(perception$schoolid))
```

```
## [1] 51
```

```
# colSums(is.na(perception))

# student data
student <- perception[,1:310]
student_subset <- cbind(student[c("year", "state", "schoolid", "classid", "studentid")], student[ ,grepl

colSums(is.na(student_subset))
```

```
##          year         state      schoolid       classid     studentid
##             0             0             0             0          2215
##          m_a1          m_a2          m_a3          m_a4          m_a5
##           594           624           820           830           902
##          m_a6          m_a7          m_a8          m_a9         m_a10
##          1071           949           878          1053           898
##         m_a11         m_a12         m_a13         m_a14         m_a15
##           841          1008          1048          1004          1238
##         m_a16         m_a17         m_a18         m_a19         m_a20
##          1120          1023          1158          1170          1201
##         m_a21         m_a22         m_a23         m_a24         m_a25
##          1281          1197          1288          1200          1206
##         m_a26         m_a27         m_a28         m_a29         m_a30
##          1153          1274          1292          1511          1344
##         m_a31         m_a32         m_a33         m_a34         m_a35
```

```
##          1178          1310          1432          1378          1458
##         m_a36         m_a37         m_a38         m_a39         m_a40
##          1466          1495          1348          1534          1441
##         m_a41         m_a42         m_a43         m_a44         m_a45
##          1463          1820          1803         19256         19256
##         m_a46         m_a47          m_b1          m_b2          m_b3
##         19256         19256          1506          1713          1673
##          m_b5          m_b6          m_b8          m_b9         m_b10
##          1903          1768          1865          2014          1886
##         m_b11         m_b12         m_b13         m_b14         m_b15
##          2000          1988          2008          2129          2214
##         m_b16         m_b17         m_b18         m_b19         m_b21
##          2098          2191          2161          2282          2320
##         m_b22         m_b23         m_b24         m_b26         m_b27
##          2391          2413          2303          2483          2531
##         m_b28         m_b29         m_b30         m_b31         m_b32
##          2512          2420          2668          3070          2453
##         m_b33         m_b34         m_b35         m_b36         m_b37
##          2684          2570          2534          2572          2582
##         m_b39         m_b40         m_b41         m_b42         m_b43
##          2628          2679          2651          2794          2501
##         m_b44         m_b45         m_b46         m_b47         m_b48
##          2490          2608          2635          2731          2734
##         m_b49         m_b50         m_b51         m_b52         m_b54
##          2359          1980          1406          1560          5987
##         m_b55         m_b56         m_b58         m_b59         m_b60
##          1032          1731          2262          2719          2818
##         m_b61         m_b62         m_b63         m_b64         m_b65
##          3017          2327          1994         19256         19256
##         m_b66       m_books          m_c1         m_c2a         m_c2b
##         19256          1166           852           496           502
##         m_c2c         m_c2d         m_c2e         m_c2f         m_c2g
##           505           498           504           503           501
##          m_c3          m_c4         m_c5a         m_c5b         m_c5c
##           987           943          1959          1962          1958
##         m_c5d         m_c5e         m_c5f         m_c5g         m_c5h
##          1955          1958          1955          1955          3144
##         m_c5i         m_c5j          m_c6          m_c7 m_c8_a_highest
##          3141          3147          1023          1043          2279
## m_c8_b_highest        m_c12a        m_c12b        m_c12c        m_c12d
##          2754          6350          6354          6349          6364
##        m_c12e        m_c12f        m_c12g        m_c12h        m_c12i
##          6342          6334          6336          6357          6344
##        m_c12j        m_c12k        m_c12l        m_c12m        m_c12n
##          6346          6330          6341          6354          6348
##        m_c12o        m_c12p        m_c12q        m_c12r        m_c12s
##          6335          6333          6348          6244          6207
##        m_c12t        m_c12u        m_c12v        m_c12w        m_c12x
##          6346          6328         19256         19256         19256
##         m_c13         m_c14         m_c15         m_c16         m_c17
##          2157          2101         19254          2204         19254
##       m_c18_a       m_c18_b       m_c18_c       m_c19_a       m_c19_b
##          2324          2429         19256          2696          2764
##       m_c19_c       m_c19_e       m_c19_f       m_c19_h       m_c19_i
```

```
##         2748         2950         2887         2941         3175
##       m_c19_j      m_c19_k      m_c19_l      m_c19_m      m_c19_n
##         3219        18693        19256        19256        19256
##       m_c20_a     m_c20_ad     m_c20_ae     m_c20_af     m_c20_ag
##         2531         3064         3362         3584         3877
##      m_c20_ah     m_c20_ai     m_c20_aj     m_c20_ak     m_c20_al
##        18700        18702        18702        18713        19256
##      m_c20_am      m_c20_c      m_c20_d      m_c20_e      m_c20_f
##        19256         2810         2899         2909         2948
##       m_c20_h      m_c20_i      m_c20_k      m_c20_l      m_c20_m
##         3539         3082         3168         3128         3173
##       m_c20_n      m_c20_o      m_c20_p      m_c20_q      m_c20_s
##         3241         3205         3289         3229         3555
##       m_c20_t      m_c20_u      m_c20_v      m_c20_w      m_c20_x
##         3350         3274         3288         3339         3318
##       m_c20_y      m_c20_z      m_c22_a      m_c22_b      m_c23_a
##         3297         3748         2654         3424         4288
##       m_c23_b       m_c24        m_c25        m_c26      m_compu
##         6738        18736        18734        18706         1053
##       m_fthed      m_grade    m_highsch     m_hmlang       m_male
##         7323          581            0         1259         1001
##       m_mthed      m_partot     m_parvar    m_raceall        m_sibs
##         6000            0        14713            0         1150
```

Question 1: Impact of "Acting White" on GPA Existing research finds that minority students might intentionally perform worse in class if excelling is equated to "acting white" and if slacking academically helps them fit in with their peer group better. I want to verify this claim using this dataset.

The relevant variables are * `m_c20_k`: my friends think it's important to work hard to get high grades * `m_c20_s`: at this school, people like me get accused of acting white

Question 2:

Question 3:

Question 4: