==========

ANALYSIS CODE IN R

==========

EXPLORATORY DATA ANALYSIS

```
# load data
library(haven)
df1_raw <- read_dta('/Users/graceyang/Google Drive/_UCLA 420 causal/420 project/j_youth.dta')
```

TREATMENT i.e. frequency of technology / social media usage:

```
cat('Treatment option: \nHow many hours do you spend chatting or interacting with friends
    + through social web-sites on a normal school day?
    + (1 None; 2 Less than an hour; 3 1-3 hours; 4 4-6 hours; 5 7 or more hours)')
```

```
## Treatment option:
## How many hours do you spend chatting or interacting with friends
##     + through social web-sites on a normal school day?
##     + (1 None; 2 Less than an hour; 3 1-3 hours; 4 4-6 hours; 5 7 or more hours)
```

```
table(df1_raw$j_ypnetcht)
```

```
##
##  -9   1   2   3   4   5
## 522  98 839 736 243  68
```

```
round(prop.table(table(df1_raw$j_ypnetcht[df1_raw$j_ypnetcht > 0])), 2)
```

```
##
##    1    2    3    4    5
## 0.05 0.42 0.37 0.12 0.03
```

6 possible options for OUTCOME indicating child's mental health state. For each of the 6 questions, the response section shows 7 different emoticon faces where '1' is most happy with big smile, and '7' is most unhappy with big frown.

```
cat("Outcome option 1: \nHow do you feel about your SCHOOL WORK?")
```

```
## Outcome option 1:
## How do you feel about your SCHOOL WORK?
```

```
table(df1_raw$j_yphsw)
```

```
##
##  -9   1   2   3   4   5   6   7
##  19 429 868 648 306 127  69  40
```

```
round(prop.table(table(df1_raw$j_yphsw[df1_raw$j_yphsw > 0])), 2)
```

```
##
##    1    2    3    4    5    6    7
## 0.17 0.35 0.26 0.12 0.05 0.03 0.02
```

```
cat("\n")
```

```r
cat("Outcome option 2: \nHow do you feel about your APPEARANCE?")
```

```
## Outcome option 2:
## How do you feel about your APPEARANCE?
```

```r
table(df1_raw$j_yphap)
```

```
##
##  -9   1   2   3   4   5   6   7
##  12 518 716 630 314 162 102  52
```

```r
round(prop.table(table(df1_raw$j_yphap[df1_raw$j_yphap > 0])), 2)
```

```
##
##    1    2    3    4    5    6    7
## 0.21 0.29 0.25 0.13 0.06 0.04 0.02
```

```r
cat("\n")
```

```r
cat("Outcome option 3: \nHow do you feel about your FAMILY?")
```

```
## Outcome option 3:
## How do you feel about your FAMILY?
```

```r
table(df1_raw$j_yphfm)
```

```
##
##   -9    1    2    3    4    5    6    7
##   14 1549  555  241   87   32   18   10
```

```r
round(prop.table(table(df1_raw$j_yphfm[df1_raw$j_yphfm > 0])), 2)
```

```
##
##    1    2    3    4    5    6    7
## 0.62 0.22 0.10 0.03 0.01 0.01 0.00
```

```r
cat("\n")
```

```r
cat("Outcome option 4: \nHow do you feel about your FRIENDS?")
```

```
## Outcome option 4:
## How do you feel about your FRIENDS?
```

```r
table(df1_raw$j_yphfr)
```

```
##
##   -9    1    2    3    4    5    6    7
##   18 1144  851  304  118   43   14   14
```

```r
round(prop.table(table(df1_raw$j_yphfr[df1_raw$j_yphfr > 0])), 2)
```

```
##
##    1    2    3    4    5    6    7
## 0.46 0.34 0.12 0.05 0.02 0.01 0.01
```

```r
cat("\n")
```

```r
cat("Outcome option 5: \nHow do you feel about the SCHOOL YOU GO TO?")
```

```
## Outcome option 5:
## How do you feel about the SCHOOL YOU GO TO?
```

```r
table(df1_raw$j_yphsc)
```

```
##
##  -9   1   2   3   4   5   6   7
##   9 690 745 517 283 128  56  78
```

```r
round(prop.table(table(df1_raw$j_yphsc[df1_raw$j_yphsc > 0])), 2)
```

```
##
##    1    2    3    4    5    6    7
## 0.28 0.30 0.21 0.11 0.05 0.02 0.03
```

```r
cat("\n")
```

```r
cat("Outcome option 6: \nWhich best describes how you feel about your LIFE AS A WHOLE?")
```

```
## Outcome option 6:
## Which best describes how you feel about your LIFE AS A WHOLE?
```

```r
table(df1_raw$j_yphlf)
```

```
##
##  -9   1   2   3   4   5   6   7
##  16 752 864 504 220  80  49  21
```

```r
round(prop.table(table(df1_raw$j_yphlf[df1_raw$j_yphlf > 0])), 2)
```

```
##
##    1    2    3    4    5    6    7
## 0.30 0.35 0.20 0.09 0.03 0.02 0.01
```

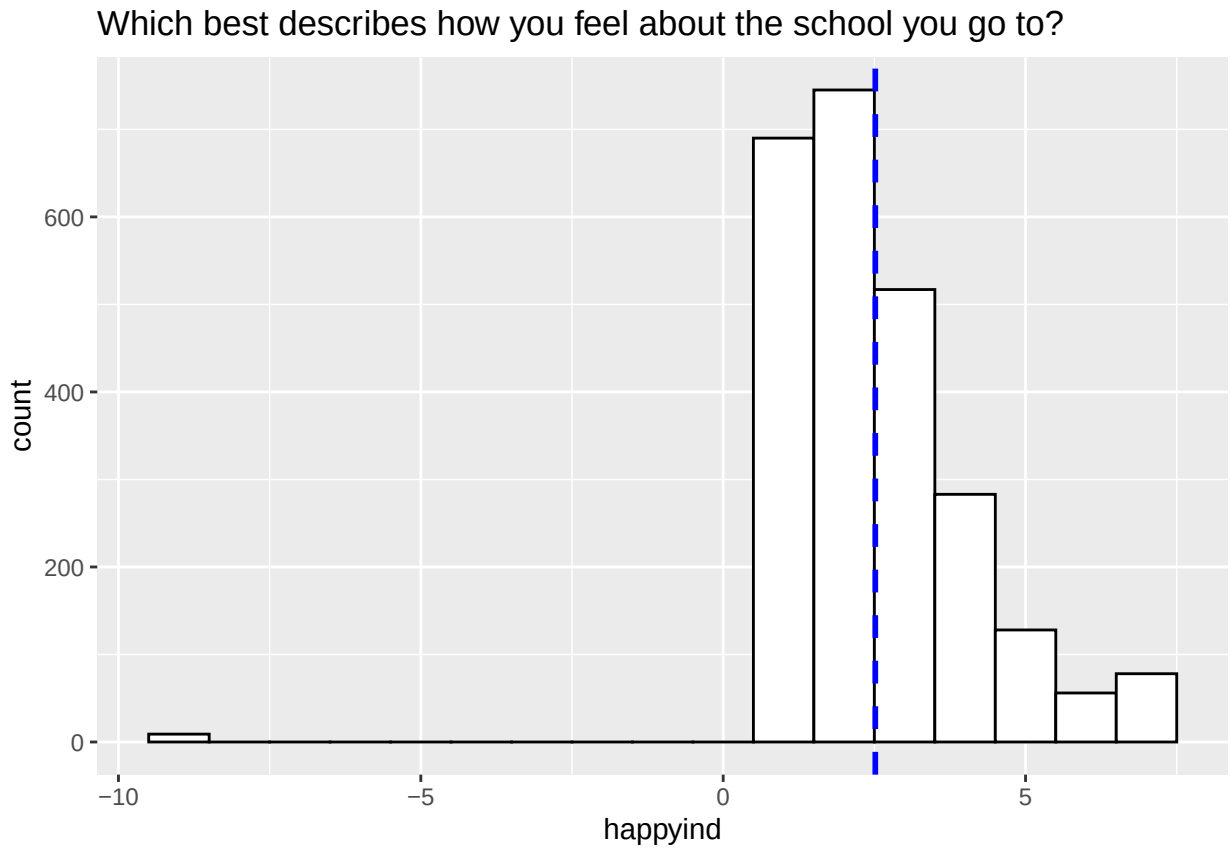
Distributions of 6 mental health indicators:

```r
## Pick 1 of 6 to review distribution
# df1_raw$happyind <- df1_raw$j_yphsw  ### school work
# df1_raw$happyind <- df1_raw$j_yphap  ### appearance
# df1_raw$happyind <- df1_raw$j_yphfm  ### family
# df1_raw$happyind <- df1_raw$j_yphfr  ### friends
df1_raw$happyind <- df1_raw$j_yphsc  ### school you go to
# df1_raw$happyind <- df1_raw$j_yphlf  ### life as a whole

library(ggplot2)
hist_happy <- ggplot(df1_raw, aes(x=happyind)) +
  geom_histogram(binwidth=1, color="black", fill="white")

hist_happy +
  geom_vline(aes(xintercept = mean(happyind)),
             color = "blue", linetype = "dashed", size = 1) +
  labs(title="Which best describes how you feel about the school you go to?")
```

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. Defau
```




```
mean(df1_raw$happyind)
```

```
## [1] 2.515563
```

```
sd(df1_raw$happyind)
```

```
## [1] 1.642729
```

```
sd(df1_raw$happyind) / sqrt(length(df1_raw$happyind))
```

```
## [1] 0.03281523
```

Possible confounders: age, gender, and socioeconomic status. I've assumed as proxy for the child's family socioeconomic status, the parents' years of education (higher of either) and household income, which were available from the survey responses.

==========

EXTRACT, TRANSFORM, CLEAN DATA

```
### Select the variables we want to analyze
myvars = c("j_hidp", "j_ypnetcht",
           "j_yphsw", "j_yphap", "j_yphfm", "j_yphfr", "j_yphsc", "j_yphlf",
           "j_ypdoby", "j_ypsex")

df1 = as.data.frame(df1_raw[myvars])

colnames(df1) <- c("j_hidp", "j_ypnetcht",
                   "hap_schoolwork", "hap_appearance", "hap_family",
```

```
                  "hap_friends", "hap_schooloverall", "hap_lifeoverall",
                  "birthyear", "gender"
                  )

### Create age from date of birth
df1$age <- NA
df1$age <- 2018 - df1$birthyear
df1$age[df1$age == 9] <- 10
df1$age[df1$age == 16] <- 15
table(df1$birthyear)
```

```
##
## 2002 2003 2004 2005 2006 2007 2008 2009
##  105  329  429  438  397  442  283   83
```

```
table(df1$age)
```

```
##
##  10  11  12  13  14  15
## 366 442 397 438 429 434
```

```
### RE-CODE (re-grouping) levels of j_ypnetcht treatment variable
### Set control group as those who average < 1 hour of net chat per school day
df1$netchat <- NA
df1$netchat[df1$j_ypnetcht == 1 | df1$j_ypnetcht == 2] <- 0

### For treatment, toggle between those who average 1-3 hours and 4+ hours
# df1$netchat[df1$j_ypnetcht == 3] <- 1                          ### 1-3 hours ###
df1$netchat[df1$j_ypnetcht == 4 | df1$j_ypnetcht == 5] <- 1   ### 4+ hours ###

### Review netchat treatment and control group sizes
table(df1$j_ypnetcht)
```

```
##
##  -9   1   2   3   4   5
## 522  98 839 736 243  68
```

```
table(df1$netchat)
```

```
##
##   0   1
## 937 311
```

```
### Drop the data points not within treatment or control groups
df1 <- na.omit(df1)
summary(df1)
```

```
##      j_hidp            j_ypnetcht    hap_schoolwork   hap_appearance
##  Min.   :6.819e+07   Min.   :1.000   Min.   :-9.000   Min.   :-9.000
##  1st Qu.:3.461e+08   1st Qu.:2.000   1st Qu.: 2.000   1st Qu.: 2.000
##  Median :7.508e+08   Median :2.000   Median : 2.000   Median : 3.000
##  Mean   :7.908e+08   Mean   :2.474   Mean   : 2.627   Mean   : 2.761
##  3rd Qu.:1.226e+09   3rd Qu.:2.000   3rd Qu.: 3.000   3rd Qu.: 4.000
##  Max.   :1.638e+09   Max.   :5.000   Max.   : 7.000   Max.   : 7.000
##    hap_family      hap_friends     hap_schooloverall hap_lifeoverall
##  Min.   :-9.000   Min.   :-9.000   Min.   :-9.000    Min.   :-9.00
##  1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.: 1.000    1st Qu.: 1.00
```

```
##  Median : 1.000   Median : 2.000   Median : 2.000    Median : 2.00
##  Mean   : 1.612   Mean   : 1.857   Mean   : 2.634    Mean   : 2.29
##  3rd Qu.: 2.000   3rd Qu.: 2.000   3rd Qu.: 3.000    3rd Qu.: 3.00
##  Max.   : 7.000   Max.   : 7.000   Max.   : 7.000    Max.   : 7.00
##    birthyear        gender           age           netchat
##  Min.   :2002   Min.   :1.000   Min.   :10.00   Min.   :0.0000
##  1st Qu.:2004   1st Qu.:1.000   1st Qu.:11.00   1st Qu.:0.0000
##  Median :2005   Median :2.000   Median :13.00   Median :0.0000
##  Mean   :2005   Mean   :1.509   Mean   :12.74   Mean   :0.2492
##  3rd Qu.:2007   3rd Qu.:2.000   3rd Qu.:14.00   3rd Qu.:0.0000
##  Max.   :2009   Max.   :2.000   Max.   :15.00   Max.   :1.0000
```

Get monthly income from adult questionnaire data set:

```
### Load individual adult questionnaire responses on monthly income
dfincome_raw <-
  read_dta('/Users/graceyang/Google Drive/_UCLA 420 causal/420 project/j_income.dta')

### Select latest monthly income for each adult
dfincome_raw <- dfincome_raw[c("j_hidp", "j_frmnthimp_dv")]

### Get total monthly income for each household
dfincome_raw$income <- round(dfincome_raw$j_frmnthimp_dv * 1, 0)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
dfincome <-
  dfincome_raw %>%
  group_by(j_hidp) %>% summarise(incomeHH = sum(income))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
summary(dfincome)
```

```
##      j_hidp             incomeHH
##  Min.   :6.801e+07   Min.   :    0
##  1st Qu.:3.475e+08   1st Qu.:  425
##  Median :6.877e+08   Median : 1090
##  Mean   :7.713e+08   Mean   : 1343
##  3rd Qu.:1.159e+09   3rd Qu.: 1844
##  Max.   :1.638e+09   Max.   :24286
```

Get highest educational qualifications from adult questionnaire data set:

```
# ### Load individual adult questionnaire responses on education level
# dfeduc_raw <-
#   read_dta('/Users/graceyang/Google Drive/_UCLA 420 causal/420 project/j_indresp.dta')
#
# ### Select educational qualifications each adult attained
```

```
# dfeduc_raw <- dfeduc_raw[c("j_hidp", "j_qfhigh_dv")]
#
# ### Save into separate file
# write.csv(dfeduc_raw, '/Users/graceyang/Google Drive/_UCLA 420 causal/420 project/dfeduc_raw.csv')
# rm(dfeduc_raw)


### Load separate file with educational qualifications data on adults
dfeduc_raw <-
  read.csv('/Users/graceyang/Google Drive/_UCLA 420 causal/420 project/dfeduc_raw.csv')

### Select highest educational qualifications in each household attained
table(dfeduc_raw$j_qfhigh_dv)
```

```
##
##   -9   -8    1    2    3    4    5    6    7    8    9   10   11   12   13   14
##   33 4656 3474 4854 2037  476  609   56 2634   25   24  472  306   78 6253  906
##   15   16   96
##  401  556 6468
```

```
dfeduc_raw$educ <- dfeduc_raw$j_qfhigh_dv
dfeduc_raw$educ[dfeduc_raw$educ < 1] <- 999
dfeduc_raw$educ[dfeduc_raw$educ == 96] <- 999
table(dfeduc_raw$educ)
```

```
##
##     1     2     3     4     5     6     7     8     9    10    11    12    13
##  3474  4854  2037   476   609    56  2634    25    24   472   306    78  6253
##    14    15    16   999
##   906   401   556 11157
```

```
library(tidyverse)
dfeduc <-
  dfeduc_raw %>% group_by(j_hidp) %>% summarise(educHH_cat = min(educ))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
table(dfeduc$educHH_cat)
```

```
##
##    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16
## 2997 3466 1456  317  433   41 1478   10   13  226  187   51 2959  404  208  282
##  999
## 5128
```

```
### Map into number of years of education (highest in each household)
dfeduc$educHH <- 6
dfeduc$educHH[dfeduc$educHH_cat <= 14] <- 10
dfeduc$educHH[dfeduc$educHH_cat <= 9] <- 12
dfeduc$educHH[dfeduc$educHH_cat <= 5] <- 14
dfeduc$educHH[dfeduc$educHH_cat == 2] <- 15
dfeduc$educHH[dfeduc$educHH_cat == 1] <- 17

### Review education data after grouping
table(dfeduc$educHH)
```

```
##
```

```
##    6   10   12   14   15   17
## 5618 3827 1542 2206 3466 2997
```

```
round(prop.table(table(dfeduc$educHH)), 2)
```

```
##
##    6   10   12   14   15   17
## 0.29 0.19 0.08 0.11 0.18 0.15
```

Number of households in different surveys and number of kids:

```
### count of unique households
NROW(unique(df1_raw$j_hidp))   ### from youth survey
```

```
## [1] 1867
```

```
NROW(unique(dfeduc_raw$j_hidp))   ### from adult survey
```

```
## [1] 19656
```

```
NROW(unique(dfincome_raw$j_hidp))   ### from household income survey
```

```
## [1] 15194
```

```
### count of children in youth survey
NROW(unique(df1_raw$pidp))
```

```
## [1] 2506
```

Join household monthly income and highest educational level to youth main data set:

```
df1 <- merge(x = df1, y = dfincome, by = "j_hidp", all.x = TRUE)
df1 <- merge(x = df1, y = dfeduc, by = "j_hidp", all.x = TRUE)
```

Drop the data points with NA:

```
df1$educHH_cat <- NULL
df1 <- na.omit(df1)
summary(df1)
```

```
##      j_hidp            j_ypnetcht    hap_schoolwork  hap_appearance
##  Min.   :6.819e+07   Min.   :1.000   Min.   :-9.00   Min.   :-9.000
##  1st Qu.:3.461e+08   1st Qu.:2.000   1st Qu.: 2.00   1st Qu.: 2.000
##  Median :7.491e+08   Median :2.000   Median : 2.00   Median : 3.000
##  Mean   :7.845e+08   Mean   :2.502   Mean   : 2.64   Mean   : 2.757
##  3rd Qu.:1.225e+09   3rd Qu.:4.000   3rd Qu.: 3.00   3rd Qu.: 4.000
##  Max.   :1.638e+09   Max.   :5.000   Max.   : 7.00   Max.   : 7.000
##    hap_family      hap_friends     hap_schooloverall hap_lifeoverall
##  Min.   :-9.000   Min.   :-9.000   Min.   :-9.000    Min.   :-9.000
##  1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.: 2.000    1st Qu.: 1.000
##  Median : 1.000   Median : 2.000   Median : 2.000    Median : 2.000
##  Mean   : 1.616   Mean   : 1.869   Mean   : 2.671    Mean   : 2.307
##  3rd Qu.: 2.000   3rd Qu.: 2.000   3rd Qu.: 4.000    3rd Qu.: 3.000
##  Max.   : 7.000   Max.   : 7.000   Max.   : 7.000    Max.   : 7.000
##    birthyear        gender          age           netchat
##  Min.   :2002   Min.   :1.00   Min.   :10.00   Min.   :0.0000
##  1st Qu.:2004   1st Qu.:1.00   1st Qu.:11.00   1st Qu.:0.0000
##  Median :2005   Median :2.00   Median :13.00   Median :0.0000
##  Mean   :2005   Mean   :1.52   Mean   :12.78   Mean   :0.2602
##  3rd Qu.:2007   3rd Qu.:2.00   3rd Qu.:14.00   3rd Qu.:1.0000
```

```
##  Max.   :2009   Max.   :2.00   Max.   :15.00   Max.   :1.0000
##     incomeHH          educHH
##  Min.   :   0.0   Min.   : 6.00
##  1st Qu.: 149.0   1st Qu.:10.00
##  Median : 636.5   Median :12.00
##  Mean   : 906.6   Mean   :12.09
##  3rd Qu.:1355.0   3rd Qu.:15.00
##  Max.   :7526.0   Max.   :17.00
```

Average score on 1-7 scale for each of 6 questions for CONTROL group:

```
mean(df1$hap_schoolwork[df1$netchat == 0])
```

```
## [1] 2.487395
```

```
mean(df1$hap_appearance[df1$netchat == 0])
```

```
## [1] 2.608643
```

```
mean(df1$hap_family[df1$netchat == 0])
```

```
## [1] 1.521008
```

```
mean(df1$hap_friends[df1$netchat == 0])
```

```
## [1] 1.82473
```

```
mean(df1$hap_schooloverall[df1$netchat == 0])
```

```
## [1] 2.440576
```

```
mean(df1$hap_lifeoverall[df1$netchat == 0])
```

```
## [1] 2.128451
```

==========

EXPLORATORY MODELS

Do a quick naive, bivariate estimated ATE using regression:

```
library(sandwich)
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(estimatr)

### Pick 1 of 6 to estimate ATE
# df1$happyind <- df1$hap_schoolwork      ### school work 0.59
# df1$happyind <- df1$hap_appearance      ### appearance 0.57
# df1$happyind <- df1$hap_family          ### family 0.37
# df1$happyind <- df1$hap_friends         ### friends 0.17
df1$happyind <- df1$hap_schooloverall   ### school you go to 0.88
```

```r
# df1$happyind <- df1$hap_lifeoverall     ### life as a whole 0.69

lm_ate_obs = lm_robust(happyind ~ netchat, data = df1)

knitr::kable(t(summary(lm_ate_obs)$coefficients[2, c(1, 2, 4, 5)]), digits = 2)
```

Estimate	Std. Error	Pr(>|t|)	CI Lower
0.88	0.12	0	0.65

Let's see how similar / different the hours of social media chatting (treatment) are, based on each of the covariates.

```r
print("Treatment vs gender")
```

```
## [1] "Treatment vs gender"
```

```r
print("Higher % of girls (gender = 2) self-reported spending 4+ hours")
```

```
## [1] "Higher % of girls (gender = 2) self-reported spending 4+ hours"
```

```r
table(df1$gender, df1$netchat)
```

```
##
##       0   1
##   1 435 105
##   2 398 188
```

```r
round(prop.table(table(df1$gender, df1$netchat), margin = 1), 2)
```

```
##
##        0    1
##   1 0.81 0.19
##   2 0.68 0.32
```

```r
cat("\n")
```

```r
print("Treatment vs age")
```

```
## [1] "Treatment vs age"
```

```r
print("Higher % of kids self-reported spending 4+ hours as they got older")
```

```
## [1] "Higher % of kids self-reported spending 4+ hours as they got older"
```

```r
table(df1$age, df1$netchat)
```

```
##
##        0   1
##   10 123   6
##   11 146  16
##   12 151  35
##   13 161  58
##   14 122  85
##   15 130  93
```

```r
round(prop.table(table(df1$age, df1$netchat), margin = 1), 2)
```

```
##
```

```
##         0    1
##   10 0.95 0.05
##   11 0.90 0.10
##   12 0.81 0.19
##   13 0.74 0.26
##   14 0.59 0.41
##   15 0.58 0.42
```

```
cat("\n")
```

```
print("Treatment vs years of parents' education")
```

```
## [1] "Treatment vs years of parents' education"
```

```
print("No clear trend in % of kids spending 4+ hours as their parents got more years of educ")
```

```
## [1] "No clear trend in % of kids spending 4+ hours as their parents got more years of educ"
```

```
table(df1$educHH, df1$netchat)
```

```
##
##        0   1
##   6  147  59
##   10 174  79
##   12 100  33
##   14  88  39
##   15 163  59
##   17 161  24
```

```
round(prop.table(table(df1$educHH, df1$netchat), margin = 1), 2)
```

```
##
##         0    1
##   6  0.71 0.29
##   10 0.69 0.31
##   12 0.75 0.25
##   14 0.69 0.31
##   15 0.73 0.27
##   17 0.87 0.13
```

```
cat("\n")
```

```
print("Treatment vs years of parents' education (check with regression)")
```

```
## [1] "Treatment vs years of parents' education (check with regression)"
```

```
print("Negligible trend in % of kids spending 4+ hours as their parents' got more years of education")
```

```
## [1] "Negligible trend in % of kids spending 4+ hours as their parents' got more years of education"
```

```
lm_educHH <- lm_robust(netchat ~ educHH, data = df1)
summary(lm_educHH)
```

```
##
## Call:
## lm_robust(formula = netchat ~ educHH, data = df1)
##
## Standard error type:  HC2
##
## Coefficients:
```

```
##             Estimate Std. Error t value  Pr(>|t|) CI Lower  CI Upper   DF
## (Intercept)  0.38759   0.044674   8.676 1.417e-17  0.29993  0.475241 1124
## educHH      -0.01053   0.003415  -3.085 2.088e-03 -0.01723 -0.003834 1124
##
## Multiple R-squared:  0.007961 ,  Adjusted R-squared:  0.007078
## F-statistic: 9.515 on 1 and 1124 DF,  p-value: 0.002088
```

```
cat("\n")
```

```
print("Treatment vs household income")
```

```
## [1] "Treatment vs household income"
```

```
print("No stat significant trend in % of kids spending 4+ hours as household income got higher")
```

```
## [1] "No stat significant trend in % of kids spending 4+ hours as household income got higher"
```

```
lm_incomeHH <- lm_robust(netchat ~ incomeHH, data = df1)
summary(lm_incomeHH)
```

```
##
## Call:
## lm_robust(formula = netchat ~ incomeHH, data = df1)
##
## Standard error type:  HC2
##
## Coefficients:
##              Estimate Std. Error   t value  Pr(>|t|)   CI Lower  CI Upper   DF
## (Intercept) 2.601e-01  1.761e-02 14.771594 2.927e-45  2.256e-01 2.947e-01 1124
## incomeHH    1.139e-07  1.302e-05  0.008751 9.930e-01 -2.543e-05 2.566e-05 1124
##
## Multiple R-squared:  5.809e-08 , Adjusted R-squared:  -0.0008896
## F-statistic: 7.657e-05 on 1 and 1124 DF,  p-value: 0.993
```

Let's estimate partial regression coeffs for the 4 covariates. The covariates show stat significant differences for gender (+0.1308 for girls), age (+0.08244 per year), and parents' years of education (-0.008979 per year of education i.e. -0.135 if the parents had bachelor's degree i.e. 15 years of education). These coeffs are expressed as the increase or decrease in % of kids responding that they spent an average of 4+ hours on a normal school day chatting with friends on social media.

```
print("Treatment vs 4 covariates")
```

```
## [1] "Treatment vs 4 covariates"
```

```
lm_cov <- lm_robust(netchat ~ gender + age + educHH + incomeHH, data = df1)
summary(lm_cov)
```

```
##
## Call:
## lm_robust(formula = netchat ~ gender + age + educHH + incomeHH,
##     data = df1)
##
## Standard error type:  HC2
##
## Coefficients:
##               Estimate Std. Error t value  Pr(>|t|)   CI Lower   CI Upper   DF
## (Intercept) -8.729e-01  1.045e-01 -8.3530 1.945e-16 -1.078e+00 -6.679e-01 1121
## gender       1.308e-01  2.451e-02  5.3352 1.154e-07  8.267e-02  1.788e-01 1121
```

```
## age          8.244e-02  6.935e-03 11.8880 8.850e-31  6.883e-02  9.605e-02 1121
## educHH      -8.979e-03  3.177e-03 -2.8263 4.793e-03 -1.521e-02 -2.745e-03 1121
## incomeHH    -1.207e-05  1.241e-05 -0.9723 3.311e-01 -3.642e-05  1.229e-05 1121
##
## Multiple R-squared:  0.1231 ,    Adjusted R-squared:   0.12
## F-statistic: 43.36 on 4 and 1121 DF,  p-value: < 2.2e-16
```

Let's check the covariate balance in this data set.

```
library(Matching)
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
## ##
## ##  Matching (Version 4.9-7, Build Date: 2020-02-05)
## ##  See http://sekhon.berkeley.edu/matching for additional documentation.
## ##  Please cite software as:
## ##   Jasjeet S. Sekhon. 2011. ``Multivariate and Propensity Score Matching
## ##   Software with Automated Balance Optimization: The Matching package for R.''
## ##   Journal of Statistical Software, 42(7): 1-52.
## ##
```

```
library(ebal)
```

```
## ##
## ## ebal Package: Implements Entropy Balancing.
```

```
## ## See http://www.stanford.edu/~jhain/ for additional information.
```

```
library(foreign)
library(MASS)

balance_formula = netchat ~ age + gender + educHH + incomeHH

match_check_obj = MatchBalance(balance_formula,
                               print.level = 0,
                               data = df1)

baltest_obs = baltest.collect(match_check_obj,
                              var.names = all.vars(balance_formula)[-1],
                              after = FALSE)[, c("mean.Tr", "mean.Co",
                                                 "T pval", "KS pval")]

knitr::kable(baltest_obs, digits = 2)
```

	mean.Tr	mean.Co	T pval	KS pval
age	13.63	12.48	0.00	0.00
gender	1.64	1.48	0.00	NA
educHH	11.53	12.29	0.00	0.00
incomeHH	906.98	906.47	0.99	0.25

The balance is bad for 3 out of 4 covariates; household income is the only one that's balanced for treatment. There's no reason to believe a causal claim made on this data as-is.

==========

FINAL MODEL: MATCHED COVARIATES

Let's rebalance on gender, age, and parents' years of education since these 3 covariates appeared to differ for treatment. Household income didn't show impact on treatment, and I've dropped this covariate from the balancing requirement.

```
library(Matching)
library(ebal)
library(foreign)
library(MASS)

# balance on 3 covariates: gender, age, parents' years of education
balance_formula = netchat ~ gender + age + educHH

# Extract variable names
variable_names = all.vars(balance_formula)[-1]

### Exact matches for gender and age; closest for parents' years of education
exact_matches = c(TRUE, TRUE, FALSE)

# Do the matching
matched_obj =
  Match(
    Y = df1$happyind,
    Tr = df1$netchat,
    X = df1[, variable_names],
    M = 1,
    exact = exact_matches,
    BiasAdjust = TRUE,
    estimand = "ATT"
  )

# ATT estimate, standard error of estimate, and t-statistic
matched_obj$est
```

```
##          [,1]
## [1,] 0.685338
```

```
matched_obj$se
```

```
## [1] 0.1359128
```

```
result_matched <- round(c(matched_obj$est, matched_obj$se, matched_obj$est / matched_obj$se), 2)
```

And the updated balance table:

```
match_balance = MatchBalance(
  balance_formula,
  match.out = matched_obj,
  data = df1,
  print.level = 0
```

```
)

balance_table_updated = baltest.collect(match_balance,
                                        var.names = variable_names,
                                        after = TRUE)[, c("mean.Tr", "mean.Co", "T pval", "KS pval")]

knitr::kable(balance_table_updated, digits = 3)
```

	mean.Tr	mean.Co	T pval	KS pval
gender	1.642	1.642	1	NA
age	13.635	13.635	1	1
educHH	11.532	11.532	1	1

==========

FINAL MODEL: CONTROLLING COVARIATES

Let's fit a naive, bivariate regression model and compare it with a second model that has more controls:

```
### Model 1 (bivariate on treatment of netchat hours spent)
model1 <- lm(happyind ~ netchat, data = df1, na.action = na.omit)

### Model 2 (add controlling for 3 covariates)
model2 <- lm(happyind ~ netchat + gender + age + educHH,
             data = df1,
             na.action = na.omit)

### Create a table
### ovb_minimal_reporting() didn't knit well for me
library(sensemakr)
```

```
## See details in:
```

```
## Carlos Cinelli and Chad Hazlett (2020). Making Sense of Sensitivity: Extending Omitted Variable Bias
```

```
table_res <-
  as.data.frame(sensitivity_stats(model1, treatment = "netchat")[c(1:7, 9)],
                col.names = names(x))

table_res[2,] <-
  sensitivity_stats(model2, treatment = "netchat")[c(1:7, 9)]

### Format table
library(formattable)
```

```
##
## Attaching package: 'formattable'
```

```
## The following object is masked from 'package:MASS':
##
##     area
```

```
table_res[, c(2, 3, 4)] <- round(table_res[, c(2, 3, 4)], 2)
table_res[, 5] <- percent(table_res[, 5])
```

```r
table_res[, 6] <- percent(table_res[, 6])
table_res[, 7] <- percent(table_res[, 7])
table_res[, 1] <- c("model 1: Naive bivariate", "model 2: Control for covariates")
names(table_res)[1:8] <-
  c("Treatment", "Est.", "SE", "t-stat", "R-sq(Y~D/X)", "RV", "RV(alpha=0.05)", "df")

### Print table
table_res
```

```
##                          Treatment Est.   SE t-stat R-sq(Y~D/X)     RV
## 1        model 1: Naive bivariate 0.88 0.11   8.02       5.42% 21.24%
## 2 model 2: Control for covariates 0.71 0.12   6.07       3.19% 16.57%
##   RV(alpha=0.05)   df
## 1         16.52% 1124
## 2         11.55% 1121
```

The above table shows that:

- Model 1 (simple bivariate model) estimates that there's an average increase of 0.88 units in unhappiness scale in the treatment group i.e. children doing 4 or more hours of netchats on an normal school day.
- Model 2 estimates a smaller increase of 0.71 units after accounting for possibly confounding factors like child's age, gender, and parents' educational level (highest attained).
- The estimated ATE seems to be most affected by the amount of hours spent on net chats. When we added in covariates in model 2, the estimated ATE seems to become more "diluted" going from 0.88 model 1 to 0.71 model 2.
- The estimated effect for model 1 appears more robust than model 2 for unobserved confounding: the robustness value (RV) tells us that if there was a confounder that explains 21.2% of the residual variance in netchat frequency and happiness/unhappiness scale, that will be sufficient to erase the model 1 estimated effect completely. The RV is lower for model 2, where it'll take an unobserved confounder being able to explain 16.6% of residual variance to eliminate the estimated effect to nothing.
- The RV_alpha=0.05 tells us that the confounding needs to have strength of 16.5% of model 1 residual variance to reduce the estimated effect to the boundary of statistical significance at $\alpha = 0.05$ level. Model 2 has a much lower strength hurdle at 11.5%, which makes it slightly easier to think up some small confounders that can eliminate any estimated effect on the child's self-rated number on the happiness/unhappiness scale. Model 2 is LESS robust to potential unobserved confounding.
- The value of R2YDX can represent an "extreme scenario" analysis: if an unobserved confounder explains 100% of the remaining outcome variation, such a confounder would need to explain only 3-5% of the residual variation in the violence treatment in order to reduce the estimated effect to zero for models 1 and 2.

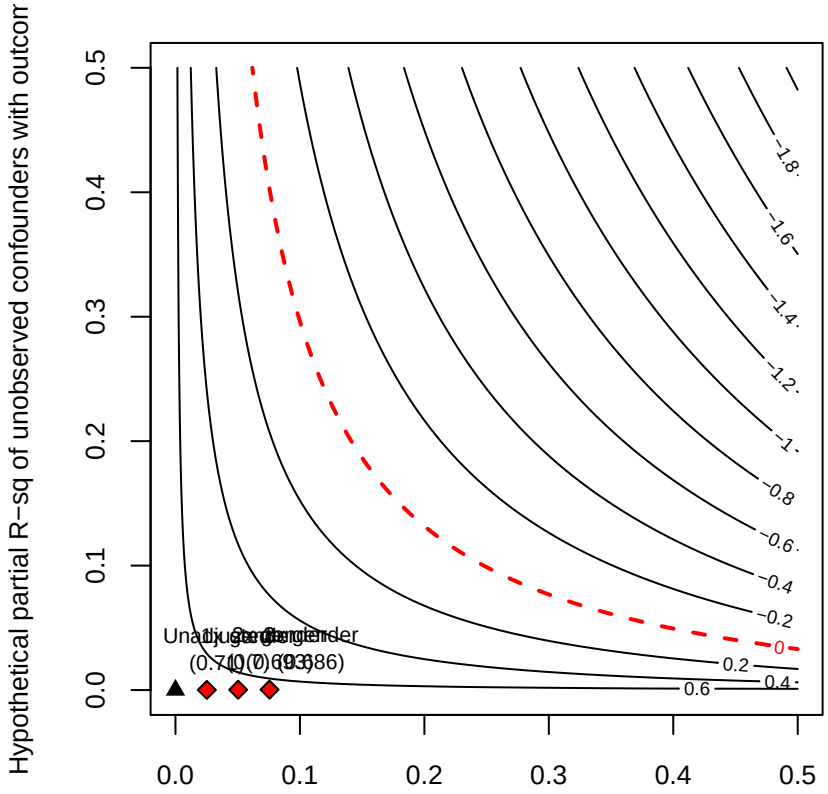
==========

SENSITIVITY ANALYSIS: CONTOUR PLOTS

How robust is the regression results to unobserved confounding factors? Would any unobserved and unnamed confounders likely exist? Let's do some sensitivity analysis on the model. Using gender as benchmark:

```r
### Sensitivity analysis - Model 2, for effect of netchat hours spent.
library(sensemakr)
sense.model2 <- sensemakr(model2,
                          treatment = "netchat",
                          benchmark = "gender",
                          kd = 1:3)
```

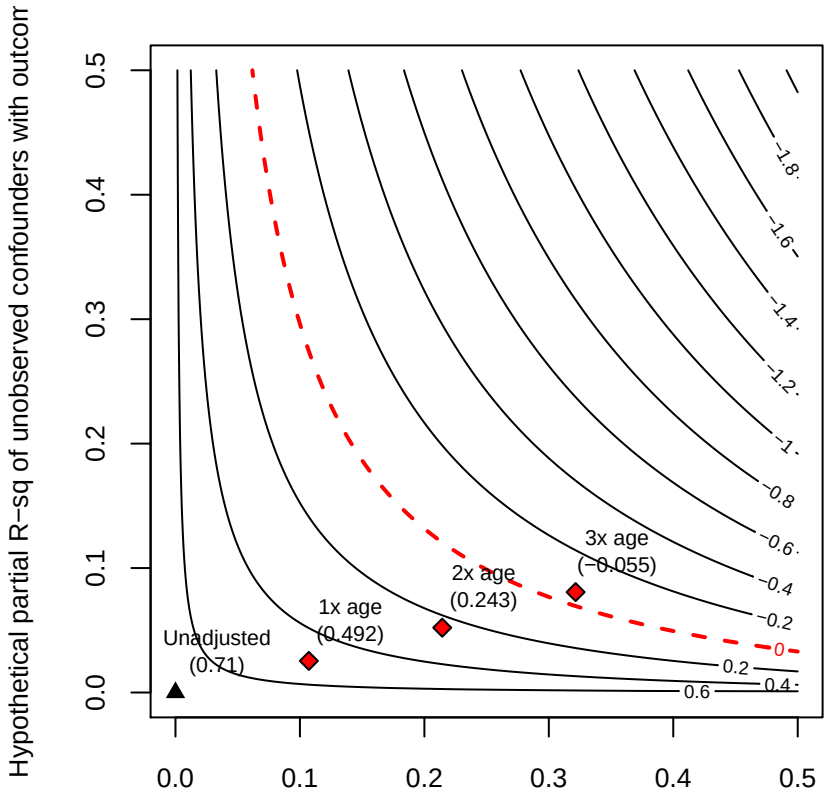
```
### Create contour plot showing sensitivity of point estimate
### to hypothesized confounder: gender, age, parents' educ level
ovb_contour_plot(
  sense.model2, lim = 0.5, lim.y = 0.5,
  xlab = "Hypothetical partial R-sq of unobserved confounders with treatment",
  ylab = "Hypothetical partial R-sq of unobserved confounders with outcome",
)
```



Using age as benchmark:

```
### Sensitivity analysis - Model 2, for effect of netchat hours spent.
library(sensemakr)
sense.model2 <- sensemakr(model2,
                          treatment = "netchat",
                          benchmark = "age",
                          kd = 1:3)

### Create contour plot showing sensitivity of point estimate
### to hypothesized confounder: gender, age, parents' educ level
ovb_contour_plot(
  sense.model2, lim = 0.5, lim.y = 0.5,
  xlab = "Hypothetical partial R-sq of unobserved confounders with treatment",
  ylab = "Hypothetical partial R-sq of unobserved confounders with outcome",
)
```



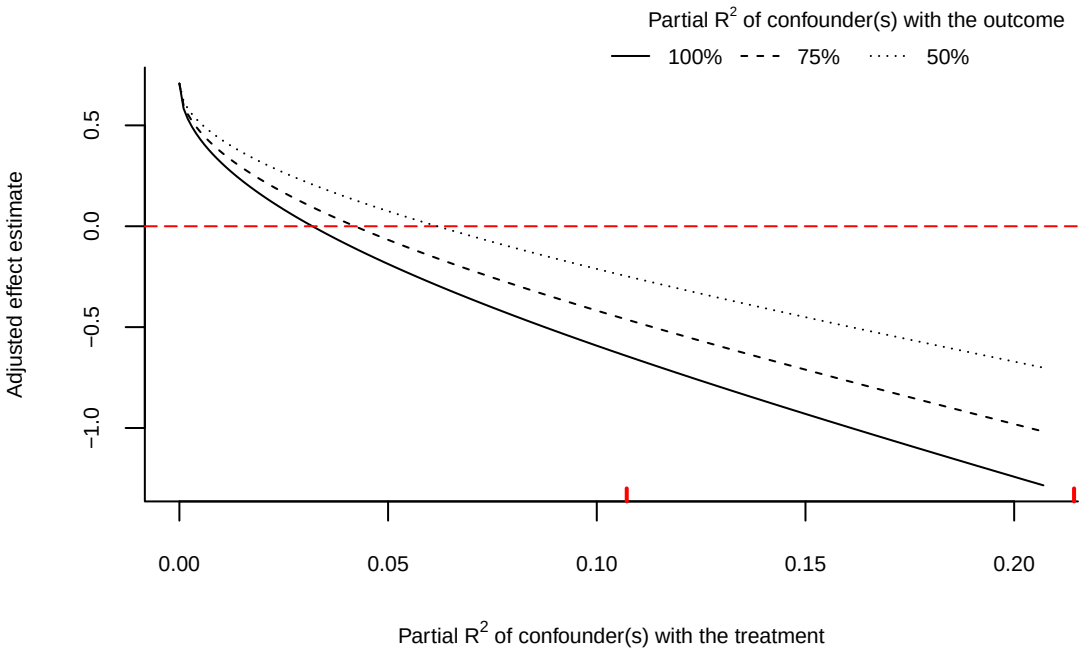

```
ovb_minimal_reporting(sense.model2, format = "latex")
```

```
## \begin{table}[!h]
## \centering
## \begin{tabular}{lrrrrrr}
## \multicolumn{7}{c}{Outcome: \textit{happyind}} \\
## \hline \hline
## Treatment: & Est. & S.E. & t-value & $R^2_{Y \sim D |{\bf X}}$ & $RV_{q = 1}$ & $RV_{q = 1, \alpha =
## \hline
## \textit{netchat} & 0.707 & 0.116 & 6.075 & 3.2\% & 16.6\% & 11.5\% \\
## \hline
## df = 1121 & & \multicolumn{5}{r}{ \small \textit{Bound (1x age)}: $R^2_{Y\sim Z| {\bf X}, D}$ = 2.5\
## \end{tabular}
## \end{table}
```

```
### this table didn't knit well for me
```

```
plot(sense.model2, type = "extreme")
```

```
## Warning in rug(x = r2dz.x, col = "red", lwd = 2): some values will be clipped
```


Partial R2 of confounder(s) with the outcome
100%
75%
50%
Adjusted effect estimate
0.5
0.0
−0.5
−1.0
0.00
0.05
0.10
0.15
0.20
Partial R2 of confounder(s) with the treatment