ANALY\_descriptives.Rmd

Niamh MacSweeney

2022-08-31

##SET UP

#Load data

pds\_y1 <- rio::import("/Volumes/GenScotDepression/users/niamh/puberty\_ABCD/ABCD\_puberty\_depression/data/pds\_timing\_y1\_R4.0.rds")  
   
dep\_vars <- rio::import("/Volumes/GenScotDepression/users/niamh/puberty\_ABCD/ABCD\_puberty\_depression/data/CBCL\_dep\_all\_yrs\_R4.0.rds")  
   
covs <- rio::import("/Volumes/GenScotDepression/users/niamh/puberty\_ABCD/ABCD\_puberty\_depression/data/covs\_main\_R4.0.rds")

#Tidy data

pds\_y1 <- pds\_y1 %>%   
 select (-c("pds\_m\_pt\_valid", "pds\_f\_pt\_valid", "age\_years")) #remove extra age column too.

## Merge

#inspect vars before merge  
colnames(covs)

## [1] "src\_subject\_id"   
## [2] "sex"   
## [3] "site\_id"   
## [4] "site\_id\_6m"   
## [5] "site\_id\_y1"   
## [6] "site\_id\_18m"   
## [7] "site\_id\_y2"   
## [8] "site\_id\_y3"   
## [9] "interview\_age\_y1"   
## [10] "interview\_age\_y2"   
## [11] "interview\_age\_y3"   
## [12] "parent\_dep\_y0"   
## [13] "parent\_dep\_y2"   
## [14] "parent\_dep\_y3"   
## [15] "rel\_family\_id"   
## [16] "acs\_raked\_propensity\_score\_y0"   
## [17] "acs\_raked\_propensity\_score\_y1"   
## [18] "bmi\_y0"   
## [19] "bmi\_y1"   
## [20] "bmi\_y2"   
## [21] "race\_y0"   
## [22] "race\_6m"   
## [23] "race\_y1"   
## [24] "race\_18m"   
## [25] "race\_y2"   
## [26] "dti\_mean\_motion\_y0"   
## [27] "dti\_mean\_motion\_y2"   
## [28] "scanner\_id\_y2"   
## [29] "WBV"   
## [30] "ICV"   
## [31] "interview\_date.baseline\_year\_1\_arm\_1"  
## [32] "household\_income\_y0"   
## [33] "parent\_high\_ed\_y0"   
## [34] "age\_y1"   
## [35] "age\_y2"   
## [36] "age\_y3"

colnames(dep\_vars)

## [1] "src\_subject\_id" "sex" "interview\_date\_y0"  
## [4] "interview\_date\_y1" "interview\_date\_y2" "interview\_date\_y3"  
## [7] "cbcl\_withdep\_y0" "cbcl\_withdep\_y1" "cbcl\_withdep\_y2"   
## [10] "cbcl\_withdep\_y3" "cbcl\_anxdep\_y0" "cbcl\_anxdep\_y1"   
## [13] "cbcl\_anxdep\_y2" "cbcl\_anxdep\_y3"

colnames(pds\_y1)

## [1] "src\_subject\_id" "sex"   
## [3] "pds\_tot\_m" "pds\_tot\_f"   
## [5] "pds\_tot\_all" "pds\_1\_p"   
## [7] "pds\_2\_p" "pds\_3\_p"   
## [9] "pds\_m5\_p" "pds\_m4\_p"   
## [11] "pds\_f4\_p" "pds\_f5b\_p"   
## [13] "pds\_p\_ss\_female\_category" "pds\_p\_ss\_male\_category"   
## [15] "gonad\_avg" "adren\_avg"   
## [17] "pt\_m" "pt\_f"   
## [19] "gonad\_f" "gonad\_m"   
## [21] "adren\_f" "adren\_m"   
## [23] "gonad\_pt\_m" "gonad\_pt\_f"   
## [25] "adren\_pt\_m" "adren\_pt\_f"

#merge by src\_subject\_id and sex as they are common to all dataframes   
  
#merge 1  
df <- covs %>%   
 left\_join(y=dep\_vars, by= c("src\_subject\_id", "sex"))  
   
#merge 2  
df <- df %>%   
 left\_join(y=pds\_y1, by= c("src\_subject\_id", "sex"))  
  
#check if merge worked okay  
colnames(df)

## [1] "src\_subject\_id"   
## [2] "sex"   
## [3] "site\_id"   
## [4] "site\_id\_6m"   
## [5] "site\_id\_y1"   
## [6] "site\_id\_18m"   
## [7] "site\_id\_y2"   
## [8] "site\_id\_y3"   
## [9] "interview\_age\_y1"   
## [10] "interview\_age\_y2"   
## [11] "interview\_age\_y3"   
## [12] "parent\_dep\_y0"   
## [13] "parent\_dep\_y2"   
## [14] "parent\_dep\_y3"   
## [15] "rel\_family\_id"   
## [16] "acs\_raked\_propensity\_score\_y0"   
## [17] "acs\_raked\_propensity\_score\_y1"   
## [18] "bmi\_y0"   
## [19] "bmi\_y1"   
## [20] "bmi\_y2"   
## [21] "race\_y0"   
## [22] "race\_6m"   
## [23] "race\_y1"   
## [24] "race\_18m"   
## [25] "race\_y2"   
## [26] "dti\_mean\_motion\_y0"   
## [27] "dti\_mean\_motion\_y2"   
## [28] "scanner\_id\_y2"   
## [29] "WBV"   
## [30] "ICV"   
## [31] "interview\_date.baseline\_year\_1\_arm\_1"  
## [32] "household\_income\_y0"   
## [33] "parent\_high\_ed\_y0"   
## [34] "age\_y1"   
## [35] "age\_y2"   
## [36] "age\_y3"   
## [37] "interview\_date\_y0"   
## [38] "interview\_date\_y1"   
## [39] "interview\_date\_y2"   
## [40] "interview\_date\_y3"   
## [41] "cbcl\_withdep\_y0"   
## [42] "cbcl\_withdep\_y1"   
## [43] "cbcl\_withdep\_y2"   
## [44] "cbcl\_withdep\_y3"   
## [45] "cbcl\_anxdep\_y0"   
## [46] "cbcl\_anxdep\_y1"   
## [47] "cbcl\_anxdep\_y2"   
## [48] "cbcl\_anxdep\_y3"   
## [49] "pds\_tot\_m"   
## [50] "pds\_tot\_f"   
## [51] "pds\_tot\_all"   
## [52] "pds\_1\_p"   
## [53] "pds\_2\_p"   
## [54] "pds\_3\_p"   
## [55] "pds\_m5\_p"   
## [56] "pds\_m4\_p"   
## [57] "pds\_f4\_p"   
## [58] "pds\_f5b\_p"   
## [59] "pds\_p\_ss\_female\_category"   
## [60] "pds\_p\_ss\_male\_category"   
## [61] "gonad\_avg"   
## [62] "adren\_avg"   
## [63] "pt\_m"   
## [64] "pt\_f"   
## [65] "gonad\_f"   
## [66] "gonad\_m"   
## [67] "adren\_f"   
## [68] "adren\_m"   
## [69] "gonad\_pt\_m"   
## [70] "gonad\_pt\_f"   
## [71] "adren\_pt\_m"   
## [72] "adren\_pt\_f"

#REDUCE TO SAMPLE NEEDED

For this RR, our final sample size requires participants to have complete pubertal data (i.e., complete PDS total score). Participants also need to have attended year 3 follow up. Due to the partial data release of year 3 follow up data in Release 4.0, only ~50% of depression data is available. For participants that attended year 3 follow up but do not have complete depression outcome measures, we will impute the depression scores (see PREP\_imputation script for more details)

###Total N = 5727

#get summary first  
df %>%   
 select(src\_subject\_id, pds\_tot\_all, interview\_date\_y3) %>%   
summary()

## src\_subject\_id pds\_tot\_all interview\_date\_y3   
## Length:11876 Min. : 5.000 Length:11876   
## Class :character 1st Qu.: 7.000 Class :character   
## Mode :character Median : 8.000 Mode :character   
## Mean : 9.032   
## 3rd Qu.:11.000   
## Max. :20.000   
## NA's :1476

#get number of complete cases for pds total and interview date y3  
sum(complete.cases(df$pds\_tot\_all)) #N = 10,400

## [1] 10400

sum(complete.cases(df$interview\_date\_y3)) #6251

## [1] 6251

final\_df <- df %>%   
 filter(!is.na(pds\_tot\_all)) %>% #N=10,400  
 filter(!is.na(interview\_date\_y3)) #N = 5727

#all vars including aux vars  
  
main\_vars <- c("src\_subject\_id", #character variable  
 "age\_y1",  
 "age\_y2",  
 "age\_y3",  
 "sex", #character variable   
 "site\_id",   
 "parent\_dep\_y2", #continuous score  
 "acs\_raked\_propensity\_score\_y0" ,   
 "bmi\_y1",   
 "household\_income\_y0", #character variable  
 "race\_y0",   
 "dti\_mean\_motion\_y2", #continuous score   
 "scanner\_id\_y2", #character variable  
 "cbcl\_withdep\_y3", #continuous score  
 "pds\_tot\_all",  
 "pt\_f",  
 "pt\_m", #continuous score  
 "gonad\_avg",  
 "adren\_avg",  
 "pds\_p\_ss\_female\_category",  
 "pds\_p\_ss\_male\_category"  
 )  
  
# main\_vars <- c("Study ID" = "src\_subject\_id", #character variable  
# "Sex", #character variable   
# "Site ID" = "site\_id",   
# "Parent depressive symptoms" = "parent\_dep\_y0",   
# "Population stratification weight" = "acs\_raked\_propensity\_score\_y0",  
# "BMI" = "bmi\_y0",  
# "Household income" = "household\_income\_y0",   
# "Race/ethnicity" = "race\_y0",   
# "DTI mean FD" = "dti\_mean\_motion\_y2", #continuous score   
# "Scanner ID" = "scanner\_id\_y2",  
# "Age" = "age\_years", #continuous score  
# "Youth depressive symptoms" = "cbcl\_withdep\_y3", #continuous score  
# "PDS total score" = "pds\_tot\_all",  
# "PT females" = "pt\_f",  
# "PT males" = "pt\_m", #continuous score  
# )

#Extract main variables   
  
main\_df <- final\_df[,main\_vars]  
  
#reorder columns for table   
  
main\_df <- main\_df %>%   
 select("src\_subject\_id",   
 "sex",  
 "age\_y1",  
 "age\_y2",  
 "age\_y3",  
 "pds\_tot\_all",  
 "cbcl\_withdep\_y3",   
 "race\_y0",  
 "bmi\_y1",   
 "household\_income\_y0",  
 "dti\_mean\_motion\_y2",  
 "parent\_dep\_y2",  
 "acs\_raked\_propensity\_score\_y0" ,   
 "site\_id",   
 "scanner\_id\_y2",  
 "pt\_f",  
 "pt\_m",  
 "gonad\_avg",  
 "adren\_avg",  
 "pds\_p\_ss\_female\_category",  
 "pds\_p\_ss\_male\_category"  
 )  
  
#tidy household income levels   
str(main\_df$household\_income\_y0)

## int [1:5727] 4 9 9 9 9 9 5 NA 9 9 ...

summary(main\_df$household\_income\_y0)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1.000 7.000 8.000 7.604 9.000 10.000 356

#make factor   
main\_df$household\_income\_y0 <- as.factor(main\_df$household\_income\_y0)  
main\_df$household\_income\_y0 <- recode\_factor(main\_df$household\_income\_y0,   
 "1" = "<$5000",  
 "2" = "$5,000-$11,999",  
 "3" = "$12,000-$15,999",  
 "4" = "$16,000-$24,999",  
 "5" = "$25,000-$34,999",  
 "6" = "$35,000-$49,999",  
 "7" = "$50,000-$74,999",  
 "8" = "$75,000-$99,999",  
 "9" = "$100,000-$199,999",  
 "10" = ">$200,000")

#MAKE TABLE

Make descriptive table

des\_table <- main\_df %>%   
 select(-c(src\_subject\_id, site\_id, scanner\_id\_y2, pt\_f, pt\_m, acs\_raked\_propensity\_score\_y0, gonad\_avg, adren\_avg,   
 pds\_p\_ss\_female\_category, pds\_p\_ss\_male\_category)) %>%   
 tbl\_summary(  
 by = sex,  
 statistic = list(all\_continuous() ~ c("{mean} ({sd}) ({min} - {max})"),  
 all\_categorical() ~ c("{n} / {N} ({p}%)")),  
 digits = list(all\_continuous() ~ c(2,2,0,0),  
 all\_categorical() ~ c(0,0,2)),  
 label = c(cbcl\_withdep\_y3 ~ "Youth depressive symptoms", pds\_tot\_all ~ "PDS total score",  
 age\_y1 ~ "Age (Y1)",  
 age\_y2 ~ "Age (Y2)",  
 age\_y3 ~ "Age (Y3)",  
 race\_y0 ~ "Race/ethnicity",  
 parent\_dep\_y2 ~ "Parent depressive symptoms",  
 bmi\_y1 ~ "BMI (Y1)",  
 household\_income\_y0 ~ "Household income",  
 dti\_mean\_motion\_y2 ~ "DTI mean FD"),  
 missing\_text = "Missing",  
 type = list (all\_continuous() ~ "continuous"),  
 missing = "ifany"  
 ) %>%   
 bold\_labels () %>%   
 italicize\_levels() %>%   
 add\_p() %>%   
 modify\_table\_body(  
 dplyr::mutate,  
 label = ifelse(label == "Missing",  
 "Missing (N)",  
 label))  
  
des\_table

## Table printed with `knitr::kable()`, not {gt}. Learn why at  
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html  
## To suppress this message, include `message = FALSE` in the code chunk header.

**Characteristic**

**F**, N = 2,726

**M**, N = 3,001

**p-value**

**Age (Y1)**

10.98 (0.63) (10 - 12)

11.00 (0.64) (10 - 12)

0.082

**Age (Y2)**

11.96 (0.64) (11 - 14)

11.98 (0.65) (11 - 14)

0.2

*Missing (N)*

34

35

**Age (Y3)**

12.87 (0.64) (11 - 14)

12.91 (0.65) (11 - 14)

0.021

**PDS total score**

10.37 (3.09) (5 - 20)

7.72 (2.12) (5 - 20)

<0.001

**Youth depressive symptoms**

1.49 (2.19) (0 - 15)

1.30 (1.90) (0 - 13)

0.012

*Missing (N)*

46

46

**Race/ethnicity**

0.072

*White*

1,876 / 2,696 (69.58%)

2,175 / 2,975 (73.11%)

*Black*

296 / 2,696 (10.98%)

278 / 2,975 (9.34%)

*Asian*

70 / 2,696 (2.60%)

73 / 2,975 (2.45%)

*AIAN/NHPI*

23 / 2,696 (0.85%)

17 / 2,975 (0.57%)

*Other*

105 / 2,696 (3.89%)

111 / 2,975 (3.73%)

*Mixed*

326 / 2,696 (12.09%)

321 / 2,975 (10.79%)

*Missing (N)*

30

26

**BMI (Y1)**

19.43 (4.24) (12 - 36)

19.43 (4.10) (11 - 36)

0.8

*Missing (N)*

51

43

**Household income**

0.6

*<$5000*

47 / 2,558 (1.84%)

66 / 2,813 (2.35%)

*$5,000-$11,999*

63 / 2,558 (2.46%)

69 / 2,813 (2.45%)

*$12,000-$15,999*

55 / 2,558 (2.15%)

43 / 2,813 (1.53%)

*$16,000-$24,999*

89 / 2,558 (3.48%)

111 / 2,813 (3.95%)

*$25,000-$34,999*

142 / 2,558 (5.55%)

133 / 2,813 (4.73%)

*$35,000-$49,999*

212 / 2,558 (8.29%)

233 / 2,813 (8.28%)

*$50,000-$74,999*

364 / 2,558 (14.23%)

388 / 2,813 (13.79%)

*$75,000-$99,999*

404 / 2,558 (15.79%)

438 / 2,813 (15.57%)

*$100,000-$199,999*

848 / 2,558 (33.15%)

964 / 2,813 (34.27%)

*>$200,000*

334 / 2,558 (13.06%)

368 / 2,813 (13.08%)

*Missing (N)*

168

188

**DTI mean FD**

1.20 (0.43) (1 - 8)

1.24 (0.53) (1 - 8)

0.14

*Missing (N)*

566

476

**Parent depressive symptoms**

3.96 (3.59) (0 - 24)

3.95 (3.54) (0 - 23)

>0.9

*Missing (N)*

35

38

#MAKE DISTRIBUTION FIGURES

##Puberty figs

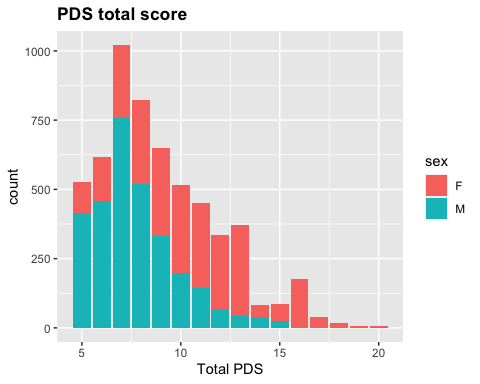
#convert sex variable to factor  
  
colnames(main\_df)

## [1] "src\_subject\_id" "sex"   
## [3] "age\_y1" "age\_y2"   
## [5] "age\_y3" "pds\_tot\_all"   
## [7] "cbcl\_withdep\_y3" "race\_y0"   
## [9] "bmi\_y1" "household\_income\_y0"   
## [11] "dti\_mean\_motion\_y2" "parent\_dep\_y2"   
## [13] "acs\_raked\_propensity\_score\_y0" "site\_id"   
## [15] "scanner\_id\_y2" "pt\_f"   
## [17] "pt\_m" "gonad\_avg"   
## [19] "adren\_avg" "pds\_p\_ss\_female\_category"   
## [21] "pds\_p\_ss\_male\_category"

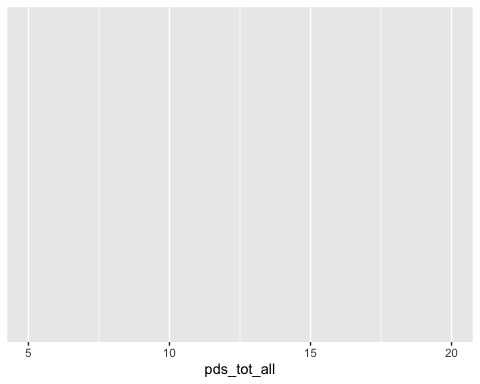
str(main\_df)

## 'data.frame': 5727 obs. of 21 variables:  
## $ src\_subject\_id : chr "NDAR\_INV030W95VP" "NDAR\_INV03BDCNWM" "NDAR\_INV03CV8RDT" "NDAR\_INV03NW0RKL" ...  
## $ sex : chr "F" "F" "M" "F" ...  
## $ age\_y1 : num 11.8 10.5 11.6 10.8 10.2 ...  
## $ age\_y2 : num 13 11.8 12.7 12.2 11.2 ...  
## $ age\_y3 : num 13.7 12.7 13.4 12.9 12 ...  
## $ pds\_tot\_all : num 20 12 6 9 10 16 9 7 5 6 ...  
## $ cbcl\_withdep\_y3 : num 5 5 0 0 2 4 3 1 0 1 ...  
## $ race\_y0 : Factor w/ 6 levels "White","Black",..: 1 1 1 1 6 1 5 2 1 1 ...  
## $ bmi\_y1 : num 24.9 19.8 13.9 16.2 24.2 ...  
## $ household\_income\_y0 : Factor w/ 10 levels "<$5000","$5,000-$11,999",..: 4 9 9 9 9 9 5 NA 9 9 ...  
## $ dti\_mean\_motion\_y2 : num 1.438 NA 1.103 1.232 0.904 ...  
## $ parent\_dep\_y2 : int 8 4 1 1 1 2 2 3 1 2 ...  
## $ acs\_raked\_propensity\_score\_y0: num 1562 531 430 502 161 ...  
## $ site\_id : chr "site19" "site13" "site09" "site17" ...  
## $ scanner\_id\_y2 : chr "HASH5ac2b20b" NA "HASH4d1ed7b1" "HASHdb2589d4" ...  
## $ pt\_f : num 7.92 2.28 NA -1.31 NA ...  
## $ pt\_m : num NA NA -2.32 NA 2.83 ...  
## $ gonad\_avg : Named num 4 2.33 1.33 2 1.67 ...  
## ..- attr(\*, "names")= chr [1:5727] NA NA NA NA ...  
## $ adren\_avg : Named num 4 2.5 1 1.5 2.5 3 2 1 1 1.5 ...  
## ..- attr(\*, "names")= chr [1:5727] NA NA NA NA ...  
## $ pds\_p\_ss\_female\_category : int 5 3 NA 3 NA 4 3 2 NA 2 ...  
## $ pds\_p\_ss\_male\_category : int NA NA 1 NA 2 NA NA 1 1 NA ...

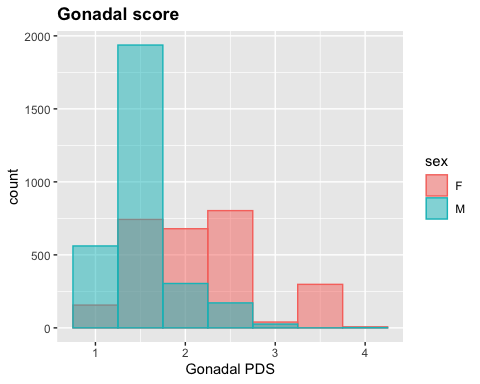
main\_df$sex <- as.factor(main\_df$sex)  
  
#plot pds total  
pds\_tot <- ggplot(main\_df, aes(pds\_tot\_all))  
  
pds\_tot + geom\_bar(aes(fill = sex)) +  
 ggtitle("PDS total score") +  
 labs (x = "Total PDS") +  
 theme(plot.title = element\_text(face = "bold"))



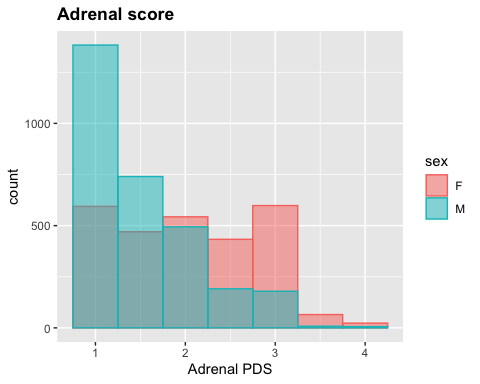
pds\_tot



#make histogram of gonadal and adrenal development within each group   
gonad\_hist <- main\_df %>%  
 ggplot(aes(x=gonad\_avg, color=sex, fill=sex)) +  
 geom\_histogram(bins = 7, alpha = 0.5, position = "identity") +  
 ggtitle("Gonadal score") +  
 labs (x = "Gonadal PDS") +  
 theme(plot.title = element\_text(face = "bold"))  
gonad\_hist



adren\_hist <- main\_df %>%  
 ggplot(aes(x=adren\_avg, color=sex, fill=sex)) +  
 geom\_histogram(bins = 7, alpha = 0.5, position = "identity") +  
 ggtitle("Adrenal score") +  
 labs (x = "Adrenal PDS") +  
 theme(plot.title = element\_text(face = "bold"))  
adren\_hist



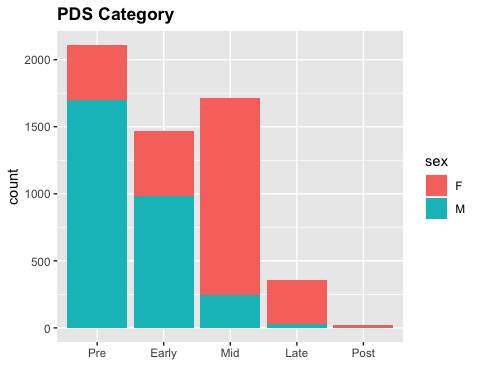
#PDS category data - there seems to be some missing data in these variables provided by ABCD.   
#let's check  
summary(main\_df$pds\_p\_ss\_female\_category)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 1.000 2.000 3.000 2.655 3.000 5.000 3018

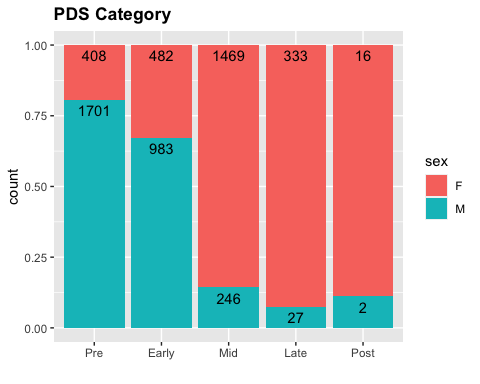
#tidy pds category data  
cat\_df <- main\_df %>% mutate(pds\_cat = coalesce(pds\_p\_ss\_female\_category, pds\_p\_ss\_male\_category)) %>%  
 select(src\_subject\_id, pds\_cat, sex)  
  
str(cat\_df)

## 'data.frame': 5727 obs. of 3 variables:  
## $ src\_subject\_id: chr "NDAR\_INV030W95VP" "NDAR\_INV03BDCNWM" "NDAR\_INV03CV8RDT" "NDAR\_INV03NW0RKL" ...  
## $ pds\_cat : int 5 3 1 3 2 4 3 2 1 2 ...  
## $ sex : Factor w/ 2 levels "F","M": 1 1 2 1 2 1 1 2 2 1 ...

cat\_df$pds\_cat <- as.factor(cat\_df$pds\_cat) # N = 60 missing  
  
#get levels in factor  
levels <- cat\_df %>%   
 group\_by(pds\_cat) %>%  
 summarise(no\_rows = length(pds\_cat))  
  
#rename factor levels   
cat\_df$pds\_cat <- recode\_factor(cat\_df$pds\_cat,   
 `1` = "Pre",  
 `2` = "Early",  
 `3` = "Mid",  
 `4` = "Late",  
 `5` = "Post",  
 `NA` = "NA"  
)  
  
cat\_hist <- cat\_df %>%  
 drop\_na(pds\_cat) %>%   
 ggplot(aes(x = pds\_cat, fill = sex)) +  
 geom\_bar() +  
 ggtitle("PDS Category") +  
 labs(x = "") +  
 theme(plot.title = element\_text(face = "bold"))  
cat\_hist



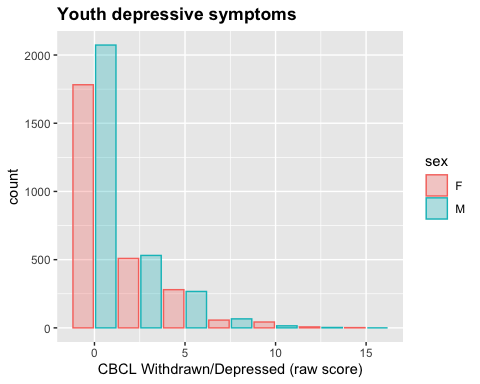
#with count scores  
cat\_hist2 <- cat\_df %>%  
 drop\_na(pds\_cat) %>%   
 ggplot(aes(x = pds\_cat, fill = sex)) +  
 geom\_bar(position = "fill") +   
 geom\_text(stat='count', aes(label=..count..), position = "fill", vjust=1.5) +  
 ggtitle("PDS Category") +  
 labs(x = "") +  
 theme(plot.title = element\_text(face = "bold"))  
cat\_hist2



## Depression figs

dep\_hist <- main\_df %>%  
 ggplot(aes(x=cbcl\_withdep\_y3, color=sex, fill=sex)) +  
 geom\_histogram(bins = 7, alpha = 0.3, position = "dodge2") +  
 ggtitle("Youth depressive symptoms") +  
 labs (x = "CBCL Withdrawn/Depressed (raw score)") +  
 theme(plot.title = element\_text(face = "bold"))  
dep\_hist

## Warning: Removed 92 rows containing non-finite values (stat\_bin).



####END OF MAIN SCRIPT #######

Variables:

(split by sex) - PDS total -CBCL depression - BMI

(overall sample) - Income - Race

## Density plots

# #create separate df with variables we want to plot  
# plot\_df <- main\_df %>%   
# select(src\_subject\_id, sex, pds\_tot\_all, cbcl\_withdep\_y3, bmi\_y0, household\_income\_y0, race\_y0)  
#   
# summary(plot\_df)  
#   
# #PDS TOTAL   
#   
# #get mean and save as tmp variable mu  
# mu <- ddply(plot\_df, "sex", summarise, grp.mean=mean(pds\_tot\_all))  
# head(mu)  
#   
# (pds\_plot <- ggplot(main\_df, aes(x=pds\_tot\_all, fill=sex, group=sex)) +  
# geom\_density(adjust=4, alpha=.4) +  
# geom\_vline(data=mu, aes(xintercept=grp.mean, color=sex),  
# linetype="dashed") +  
# theme\_classic()) +  
# labs(title=" Perceived pubertal development density curve",x="Pubertal Development Scale (total score)", y = "density") +  
# scale\_fill\_brewer(palette="Dark2")   
#   
# #CBCL depression  
# mu <- ddply(plot\_df, "sex", summarise, grp.mean=mean(cbcl\_withdep\_y3))  
# head(mu)  
#   
# (dep\_plot <- ggplot(main\_df, aes(x=cbcl\_withdep\_y3, fill=sex, group=sex)) +  
# geom\_density(adjust=4, alpha=.4) +  
# geom\_vline(data=mu, aes(xintercept=grp.mean, color=sex),  
# linetype="dashed") +   
# xlim(0,15) +  
# theme\_classic()) +  
# labs(title="Depressive symptoms density curve",x="CBCL Withdrawn Depressed (raw score)", y = "density") +  
# scale\_fill\_brewer(palette="Dark2")  
#   
# #BMI   
# mu <- ddply(plot\_df, "sex", summarise, grp.mean=mean(bmi\_y0))  
# head(mu)  
#   
# (bmi\_plot <- ggplot(main\_df, aes(x=bmi\_y0, fill=sex, group=sex)) +  
# geom\_density(adjust=4, alpha=.4) +  
# geom\_vline(data=mu, aes(xintercept=grp.mean, color=sex),  
# linetype="dashed") +  
# theme\_classic()) +  
# xlab("Body Mass Index (BMI)") +  
# ylab("density") +  
# labs(title="BMI density curve",x="Body Mass Index (BMI)", y = "density") +  
# scale\_fill\_brewer(palette="Dark2")  
#   
#   
# des\_fig\_2 <-   
# #display plots together  
# (des\_fig <- ggarrange(  
# pds\_plot,   
# ggarrange(dep\_plot, bmi\_plot, ncol= 2, labels = c("B", "C")),  
# nrow = 2,  
# labels = "A")  
# )  
#   
# ggarrange(  
# lp, # First row with line plot  
# # Second row with box and dot plots  
# ggarrange(bxp, dp, ncol = 2, labels = c("B", "C")),   
# nrow = 2,   
# labels = "A" # Label of the line plot  
# )

##Waffle charts for Race and income - waffle charts too complicated.

Stick to bar chats

# # Data transformation  
# plot\_df$race\_y0 <- as.character(plot\_df$race\_y0)  
#   
# race <- plot\_df %>%   
# select(race\_y0) %>% #select race variable only to get correct freq.   
# dplyr::group\_by(race\_y0) %>%   
# count() %>%   
# dplyr::ungroup() %>%   
# mutate(perc = `freq` / sum(`freq`)) %>%  
# arrange(perc) %>%  
# mutate(labels = scales::percent(perc))  
#   
# race\_chart <- ggplot(race, aes(x = "", y = perc, fill = (race\_y0)) +  
# geom\_col() +  
# geom\_text(aes(label = ""),  
# position = position\_stack(vjust = 0.5)) +  
# coord\_polar(theta = "y") +  
# theme\_void()  
# )  
#   
#   
#   
#   
#   
# ggplot(race, aes(x = "", y = perc, fill = (race\_y0)) +  
# geom\_bar(width = 1, stat = "identity") +  
# coord\_polar("y", start = 0) +  
# geom\_label\_repel(aes(label = perc), size=5, show.legend = F, nudge\_x = 1) +  
# guides(fill = guide\_legend(title = "Group"))  
# )  
#   
#   
# ###code working above this####  
#   
#   
# ggplot(race, aes(values = perc, fill = race\_y0)) +  
# geom\_waffle(n\_rows = 8, size = 0.33, colour = "white") +  
# scale\_fill\_manual(name = NULL,  
# values = perc,  
# labels = label) +  
# coord\_equal() +  
# theme\_void()  
#   
# geom\_text(aes(label = labels),  
# position = position\_stack(vjust = 0.5)) +  
# coord\_polar(theta = "y")