Dac Summary

Dac summary

First run on: 20201008 Spreadsheet:

This script goes through data that has been returned from the DAC and summarizes it

```r
#preprocessing - because these are sent

#Load original MS database from melissa
#load("/Users/eballer/BBL/msdepression/data/database_2019-03-04.RData")

#read in melissa's xls: n=30915
melissa_xls <- read.csv("/Users/eballer/BBL/msdepression/data/mri_file_upd_w_empi.csv", header = TRUE, s

#load our df
data <- read.csv("/Users/eballer/BBL/msdepression/data/dac/investigatingdepressioninpatientswithmsmaskda

data2 <- read.csv("/Users/eballer/Documents/Penn T32/MS_Depression_Ted_Taki/Documents/investigatingdepr
#subset EMPI, accession codes, F-codes, Phq2 and phq9 #n = 29226, with 21378 unique accession codes


#drawn from Users/eballer/BBL/msdepression/data/dac/investigatingdepressioninpatientswithmsmaskdataadde
data3_p1 <- read.csv("/Users/eballer/BBL/msdepression/data/dac/20201029_msdep_pt1.csv", header = TRUE, s

data3_p2 <- read.csv("/Users/eballer/BBL/msdepression/data/dac/20201029_msdep_pt2.csv", header = TRUE, s

data3 <- rbind(data3_p1, data3_p2)
write.csv(x = data3, file = "/Users/eballer/BBL/msdepression/data/dac/msdep_20201029_merged.csv")

data4_20210119 <- read.csv("/Users/eballer/BBL/msdepression/data/dac/investigatingdepressioninmspatienta

#set data_empi_acc_f_phq to data4
data_empi_acc_f_phq <- data4_20210119#subset(data, select = c("EMPI", "ACCESSION_NUM", "PHQ.2", "PHQ.9",

#number of unique in melissa's
m_access_n <- length(unique(melissa_xls$ACCESSION_NUM)) #n = 29183

#number of unique in my group
eb_access_n <- length(unique(data_empi_acc_f_phq$ACCESSION_NUM)) #n = 21027

#number of overlap

melissa_peeps_in_eb_group <- sum(unique(melissa_xls$ACCESSION_NUM) %in% unique(data_empi_acc_f_phq$ACCES

#make a column that indicates whether or not accession number in melissa's is in my group
melissa_xls$overlap <- ((melissa_xls$ACCESSION_NUM) %in% (data_empi_acc_f_phq$ACCESSION_NUM))

#grab unique accession numbers that are NOT in the overlap group
unique_access_nums_only_melissas_group <- melissa_xls$ACCESSION_NUM[which(!melissa_xls$overlap)]
names(unique_access_nums_only_melissas_group) <- c("count", "accession_num")
```

```r
#write table, n = 14785
write.table(file = "/Users/eballer/BBL/msdepression/data/dac/unique_access_nums_in_melissas_group_from_r

eb_peeps_in_melissa_group <- sum(unique(data_empi_acc_f_phq$ACCESSION_NUM) %in% unique(melissa_xls$ACCES

print(paste0("Number of Melissa's people in EB group : ", melissa_peeps_in_eb_group, " out of ", m_acces
```

```
## [1] "Number of Melissa's people in EB group : 25857 out of 29183 unique accession numbers"
```

```r
print(paste0("Number of EB's people in Melissa's group : ", eb_peeps_in_melissa_group, " out of ", eb_ac
```

```
## [1] "Number of EB's people in Melissa's group : 25857 out of 42840 unique accession numbers"
```

```r
overlap_group <- data_empi_acc_f_phq[which(unique(data_empi_acc_f_phq$ACCESSION_NUM) %in% unique(melissa

############################
### some preprocessing ###
############################

#change accession name for ease of merging with Rdata if using
#names(data_empi_acc_f_phq) <- gsub(pattern = "ACCESSION_NUM", replacement = "AccessionNumber", names(da


#make accession # and lab findings into integer type, contained in row 28:42, Margin=2 to work on colum
data_empi_acc_f_phq$ACCESSION_NUM <- as.integer(data_empi_acc_f_phq$ACCESSION_NUM)
data_empi_acc_f_phq[28:42] <- apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN = 2)
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
```

```
## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion

## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion

## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion

## Warning in apply(X = data_empi_acc_f_phq[28:42], FUN = as.integer, MARGIN =
## 2): NAs introduced by coercion
data_empi_acc_f_phq$sex_binarized <- ifelse(data_empi_acc_f_phq$SEX == "MALE", 1, 2)
data_empi_acc_f_phq$race_binarized <- ifelse(data_empi_acc_f_phq$RACE == "WHITE", 1, 2)

#get empis of people who have wacky deidentified empis
db_empi_acc <- subset(melissa_xls, select = c("ACCESSION_NUM", "EMPI"))

#change EMPI to "patient ID"
names(db_empi_acc) <- gsub(pattern = "EMPI", replacement = "PatientID", names(db_empi_acc))
                    #subset(database, select = c("AccessionNumber", "PatientID"))

#merge with our database, PatientID is De-identified EMPI, n = 30799
data_empi_acc_f_phq_ptID <- merge(data_empi_acc_f_phq, db_empi_acc, by = "ACCESSION_NUM")

# number of unique scanners and scans
print(paste0("Number of unique procedure names (ie. MRI Head w/out contrast; MRI Source Images): ",  len

## [1] "Number of unique procedure names (ie. MRI Head w/out contrast; MRI Source Images): 24"
print(paste0("Number of unique department names (ie. RAD HUP PCAM Ground): ",  length(unique(data_empi_

## [1] "Number of unique department names (ie. RAD HUP PCAM Ground): 3"
print(paste0("Number of unique modalities (ie. HUP PCAM GR MR4 3T; MR6 3T): ",  length(unique(data_empi_

## [1] "Number of unique modalities (ie. HUP PCAM GR MR4 3T; MR6 3T): 11"
#phq-2;
print(paste0("N PHQ-2 = ", sum(!is.na(data_empi_acc_f_phq$PHQ.2))))

## [1] "N PHQ-2 = 13820"
#print histogram of phq-2s for people

hist(data_empi_acc_f_phq$PHQ.2[which(!is.na(data_empi_acc_f_phq$PHQ.2))], main = paste0("PHQ-2 Histo : 
```
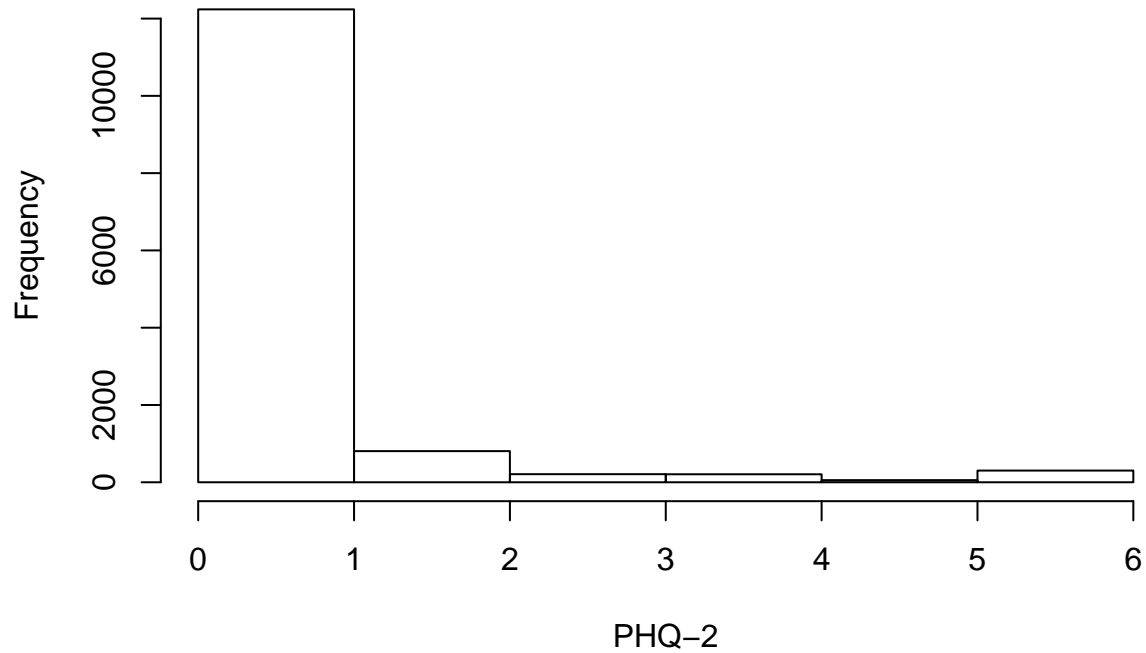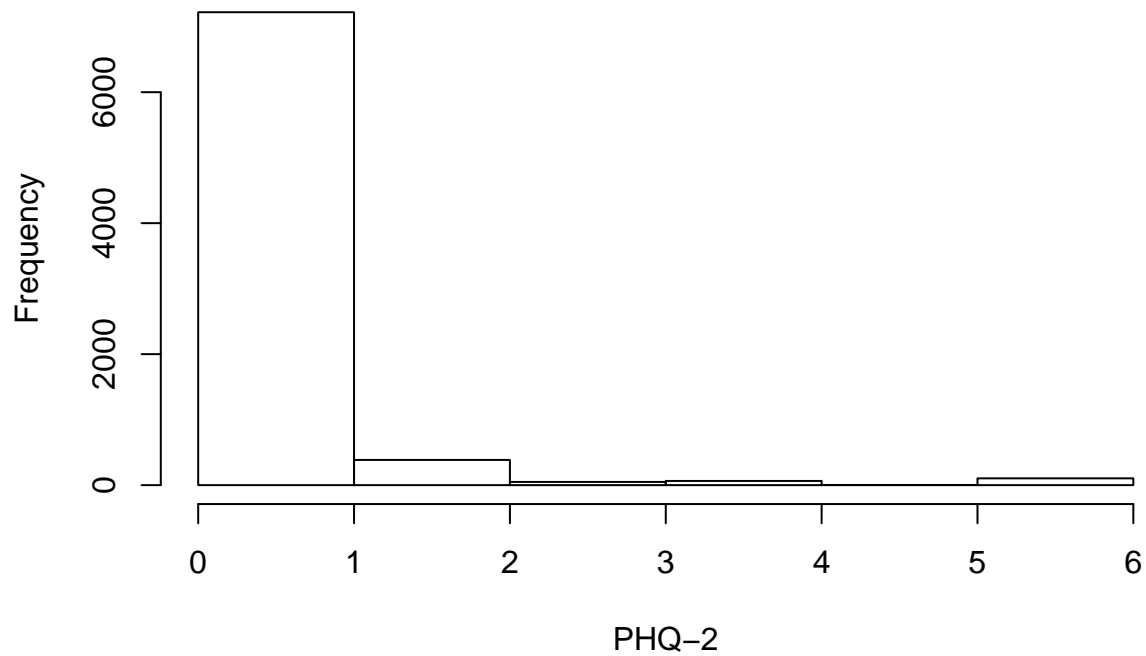
PHQ−2 Histo : n = 13820


```r
hist(data_empi_acc_f_phq_ptID$PHQ.2[which(!is.na(data_empi_acc_f_phq_ptID$PHQ.2))], main = paste0("PHQ-
```

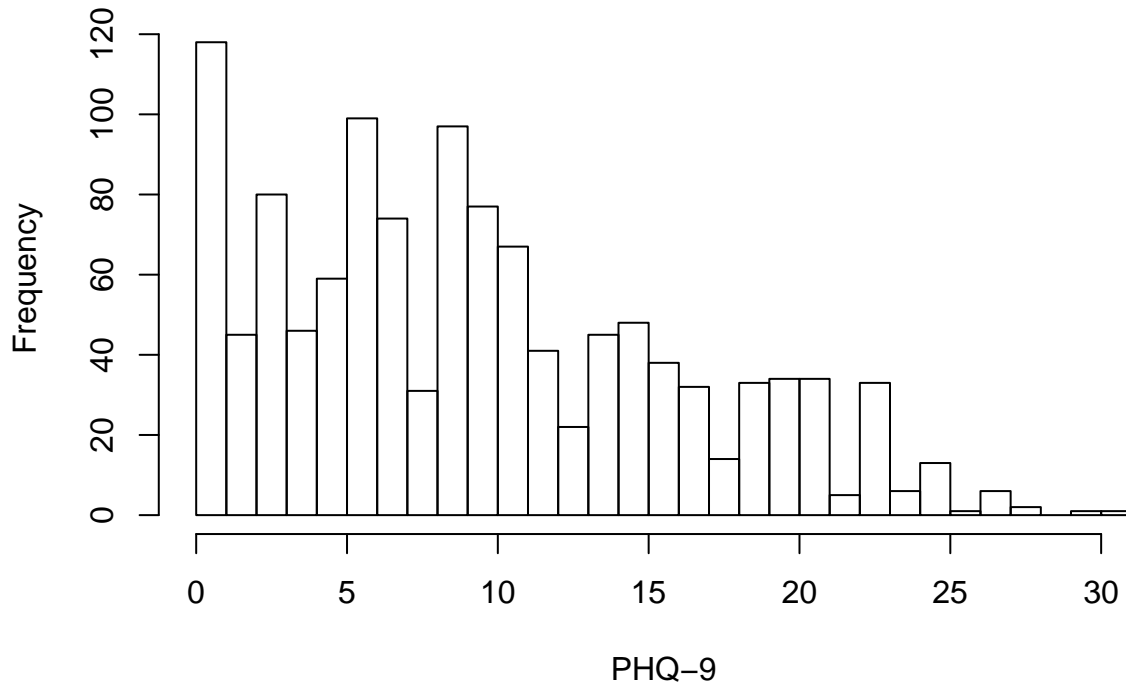## PHQ−2 Histo in Melissa's subgroup #: n = 7825


```r
#phq-9
print(paste0("N PHQ-9 = ", sum(!is.na(data_empi_acc_f_phq$PHQ.9))))
```

```
## [1] "N PHQ-9 = 1202"
```

```
#print histogram of phq-9s for people
hist(data_empi_acc_f_phq$PHQ.9[which(!is.na(data_empi_acc_f_phq$PHQ.9))], main = paste0("PHQ-9 Histo : n
```
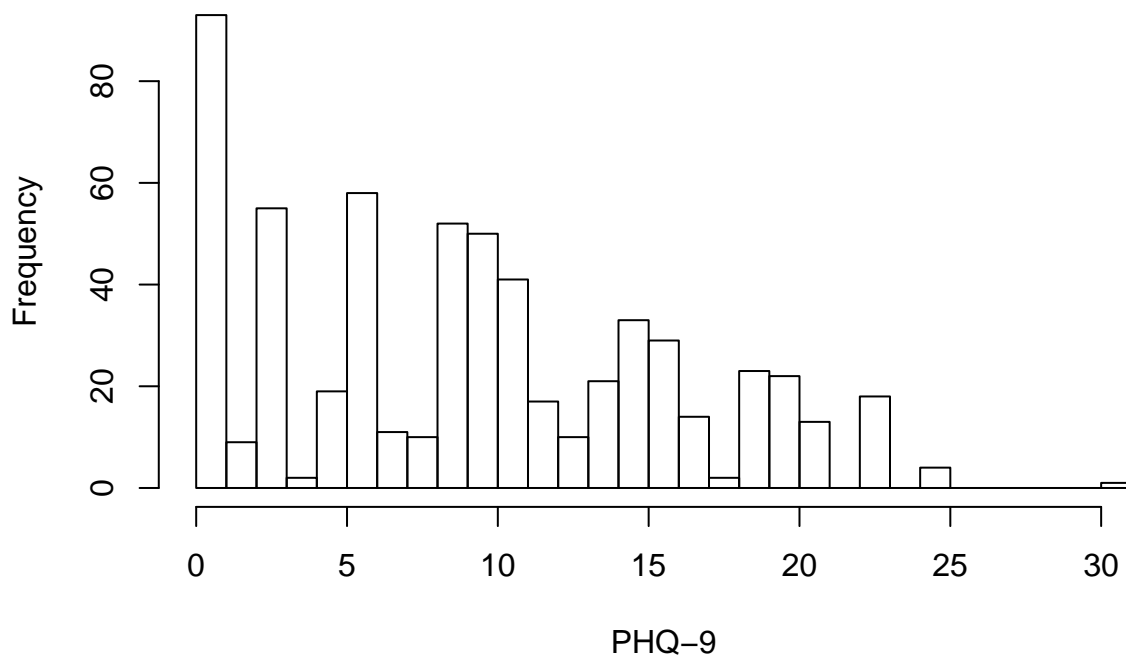
PHQ−9 Histo : n = 1202


```
hist(data_empi_acc_f_phq_ptID$PHQ.9[which(!is.na(data_empi_acc_f_phq_ptID$PHQ.9))], main = paste0("PHQ-9
```

PHQ−9 Histo In Melissa's group: n = 607

```
#both phq2 and 9

num_both_phq2_and_9 <- length(which(!is.na(data_empi_acc_f_phq$PHQ.2) & !is.na(data_empi_acc_f_phq$PHQ.9
print(paste0("num people with both phq2 and 9 = ", num_both_phq2_and_9))
```

[1] "num people with both phq2 and 9 = 0"

```
#looking at just people with phq-2 and 9
phq2_subset <- data_empi_acc_f_phq[which(!is.na(data_empi_acc_f_phq$PHQ.2)),] #n=5855
phq9_subset <- data_empi_acc_f_phq[which(!is.na(data_empi_acc_f_phq$PHQ.9)),] #n=606

#checking logic for order of subsetting - unique EMPI then ph2/9 == phq2/9 then EMPI subsetting - this i
unique_phq2_empis <- length(unique(phq2_subset$EMPI)) #n=1005
unique_phq9_empis <- length(unique(phq9_subset$EMPI)) #n=70


#age histogram
hist(data_empi_acc_f_phq$MRI_ENC_AGE[which(!is.na(data_empi_acc_f_phq$MRI_ENC_AGE))], main = paste0("Age
```
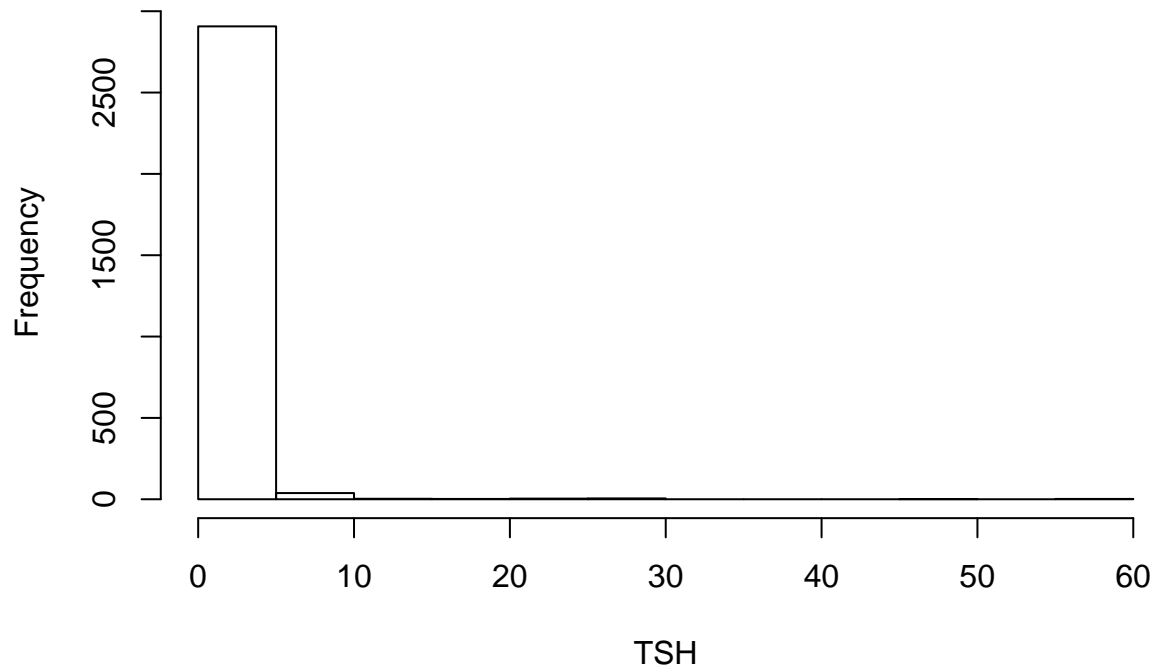
Age Histo : n = 42842

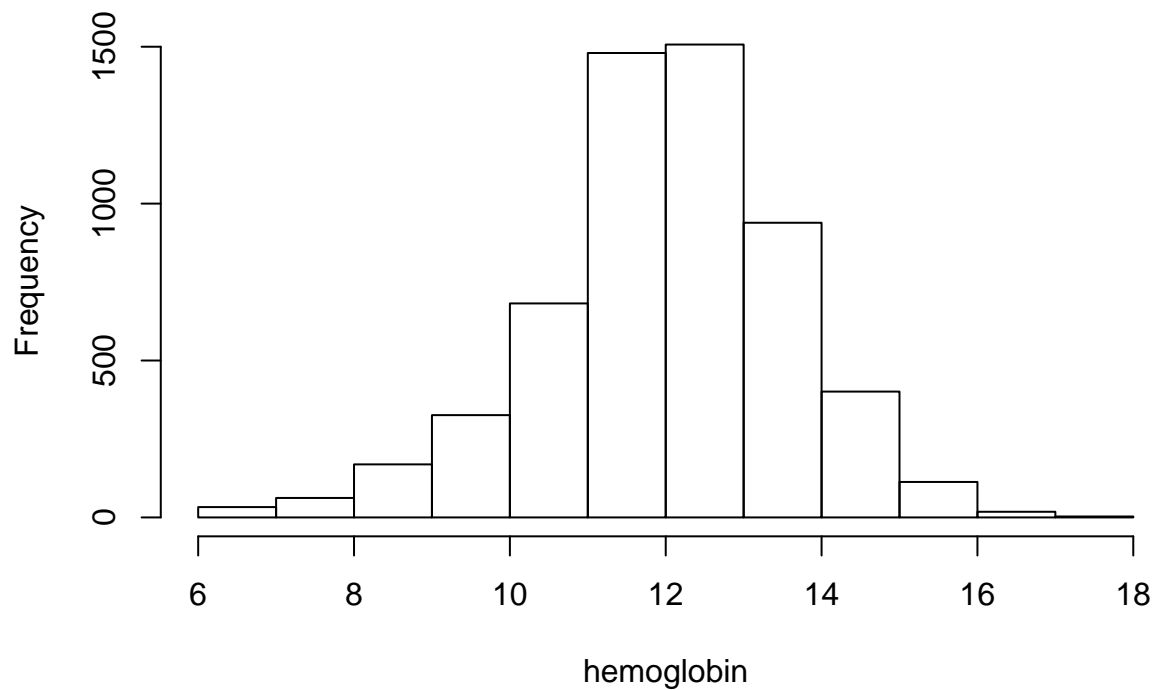AGE AT MRI ENCOUNTER

```
#tsh
hist(data_empi_acc_f_phq$TSH[which(!is.na(data_empi_acc_f_phq$TSH))], main = paste0("TSH Histo : n = ",
```
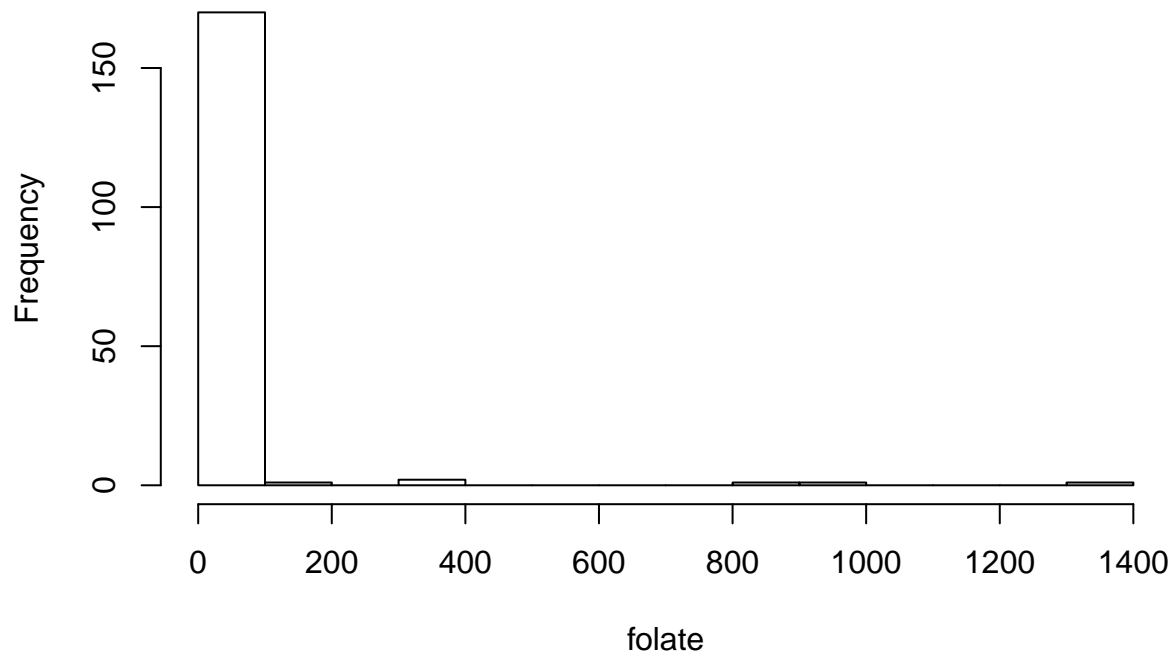
TSH Histo : n = 2962


```r
#hemoglobin
hist(data_empi_acc_f_phq$hemoglobin[which(!is.na(data_empi_acc_f_phq$hemoglobin))], main = paste0("Hemog
```

Hemoglobin Histo : n = 5733


```r
#folate
hist(data_empi_acc_f_phq$FOLATE[which(!is.na(data_empi_acc_f_phq$FOLATE))], main = paste0("Folate Histo
```
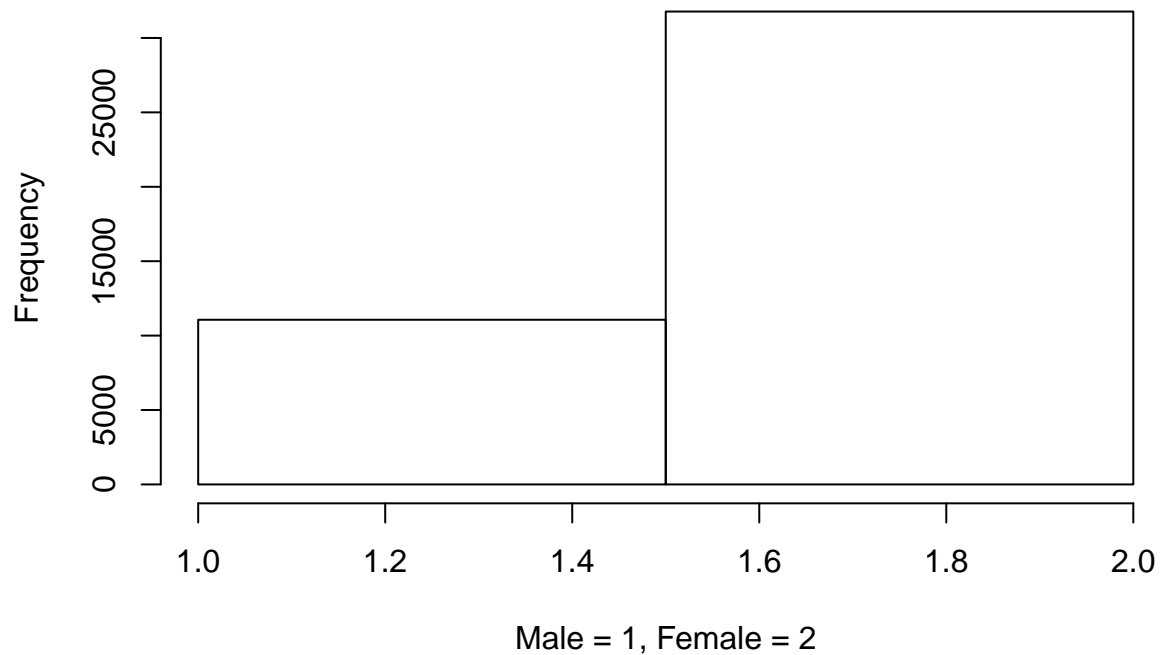
Folate Histo : n = 176


```
#sex
hist(data_empi_acc_f_phq$sex_binarized[which(!is.na(data_empi_acc_f_phq$sex_binarized))],
    main = paste0("Sex Histo : n = ", length(which(!is.na(data_empi_acc_f_phq$sex_binarized))), "; M =
    xlab = "Male = 1, Female = 2",
    breaks = 2)
```
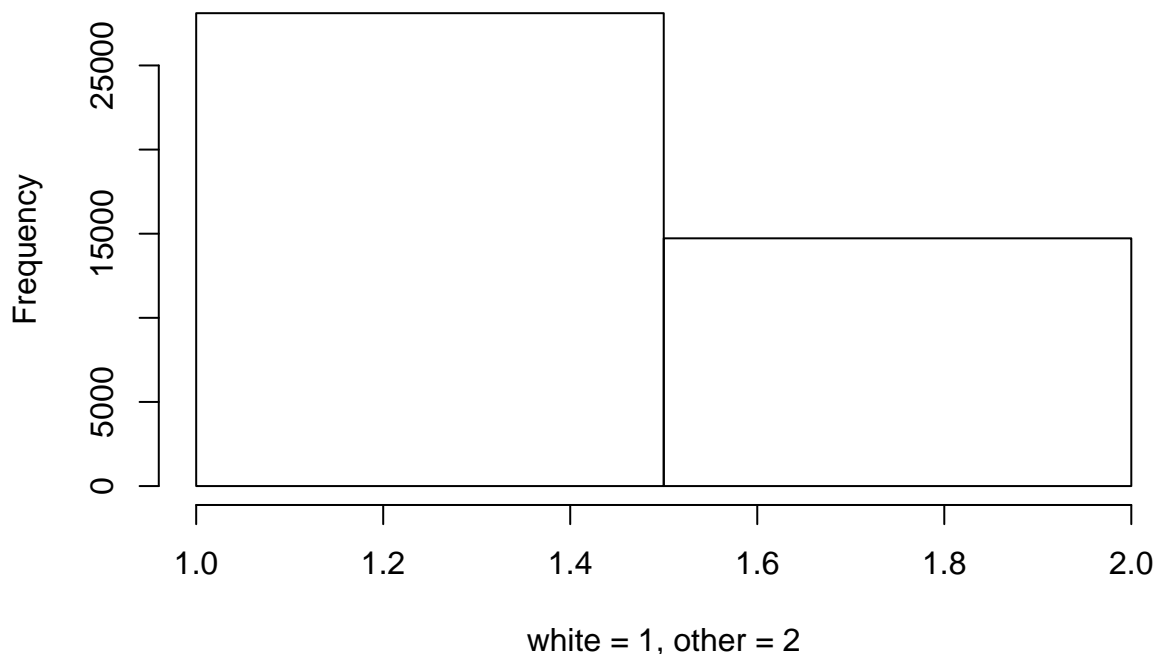
Sex Histo : n = 42842; M = 11065; F = 31777

```r
#race
hist(data_empi_acc_f_phq$race_binarized[which(!is.na(data_empi_acc_f_phq$race_binarized))],
    main = paste0("Race Histo : n = ", length(which(!is.na(data_empi_acc_f_phq$race))), "; White = ",
    xlab = "white = 1, other = 2",
    breaks = 2)
```

Race Histo : n = 42833; White = 28108; NW = 14725

white = 1, other = 2

```r
#f-codes - people usually have BOTH icd9/10 or neither
num_depression <- length(which(!is.na(data_empi_acc_f_phq$ICD9))) #n=11131
print(paste0("N with ICD 9/10 codes for depression = ", num_depression))
```

```
## [1] "N with ICD 9/10 codes for depression = 42842"
```

```r
num_unique_EMPI_and_dep <- length(which(!is.na(data_empi_acc_f_phq$ICD9) & unique(data_empi_acc_f_phq$EI
```

```
## Warning in !is.na(data_empi_acc_f_phq$ICD9) & unique(data_empi_acc_f_phq
## $EMPI): longer object length is not a multiple of shorter object length
```

```r
print(paste0("N with ICD 9/10 codes for depression and unique EMPI = ", num_unique_EMPI_and_dep))
```

```
## [1] "N with ICD 9/10 codes for depression and unique EMPI = 42842"
```

```r
num_phq2_and_dep <- length(which(!is.na(data_empi_acc_f_phq$ICD9) & !is.na(data_empi_acc_f_phq$PHQ.2)))
print(paste0("N with ICD 9/10 codes for depression and phq2 = ", num_phq2_and_dep))
```

```
## [1] "N with ICD 9/10 codes for depression and phq2 = 13820"
```

```r
num_phq9_and_dep <- length(which(!is.na(data_empi_acc_f_phq$ICD9) & !is.na(data_empi_acc_f_phq$PHQ.9)))
print(paste0("N with ICD 9/10 codes for depression and phq9 = ", num_phq9_and_dep))
```

```
## [1] "N with ICD 9/10 codes for depression and phq9 = 1202"
```

```r
#fcode histos
hist(data_empi_acc_f_phq$PHQ.2[which(!is.na(data_empi_acc_f_phq$ICD9) & !is.na(data_empi_acc_f_phq$PHQ.2
```

```
        main = paste0("PHQ2 and Dep Histo : n = ", length(which(!is.na(data_empi_acc_f_phq$ICD9) & !is.na(d
```

PHQ2 and Dep Histo : n = 13820
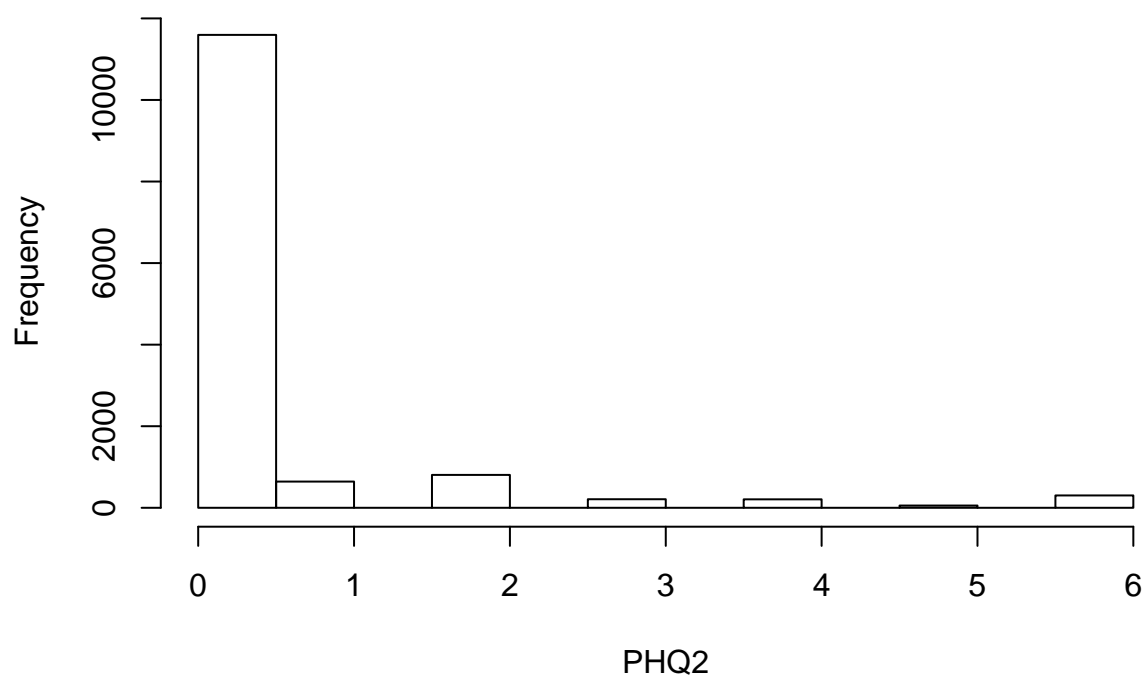
PHQ2

```
hist(data_empi_acc_f_phq$PHQ.9[which(!is.na(data_empi_acc_f_phq$ICD9) & !is.na(data_empi_acc_f_phq$PHQ.9
        main = paste0("PHQ9 and Dep Histo : n = ", length(which(!is.na(data_empi_acc_f_phq$ICD9) & !is.na(d
```
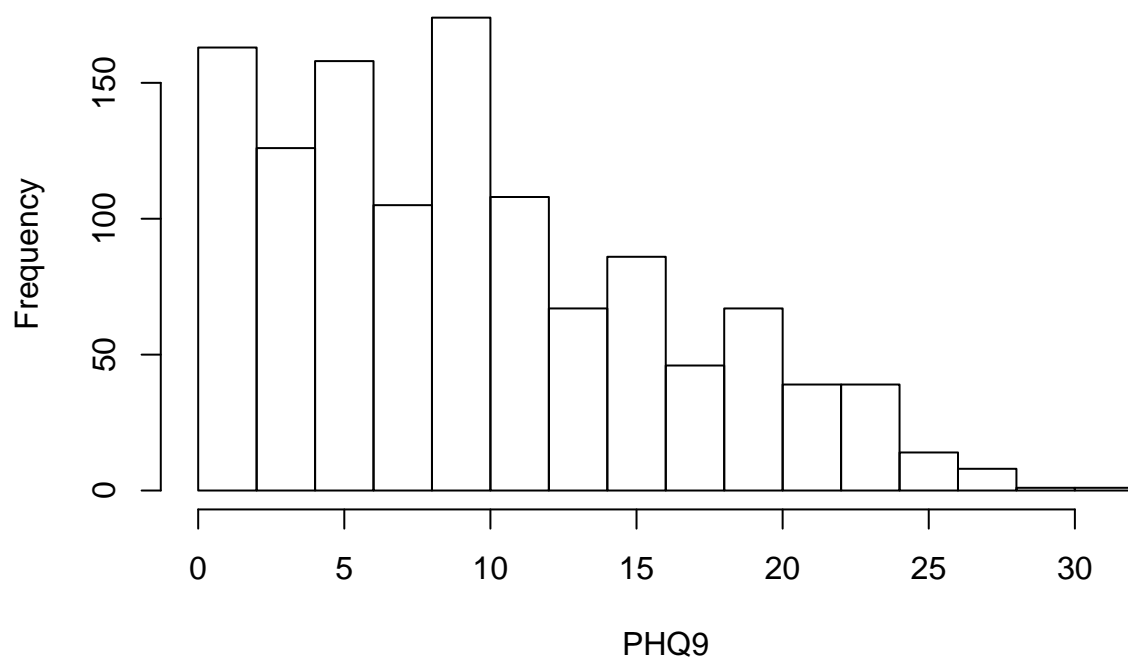
PHQ9 and Dep Histo : n = 1202

PHQ9

```
############ 11/5/2020 #############
#########################
##### overlap group #####
#########################

#make accession # and lab findings into integer type, contained in row 28:42, Margin=2 to work on colum
overlap_group$ACCESSION_NUM <- as.integer(overlap_group$ACCESSION_NUM)
overlap_group[28:42] <- apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2)
```

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

Warning in apply(X = overlap_group[28:42], FUN = as.integer, MARGIN = 2):
NAs introduced by coercion

```
overlap_group$sex_binarized <- ifelse(overlap_group$SEX == "MALE", 1, 2)
overlap_group$race_binarized <- ifelse(overlap_group$RACE == "WHITE", 1, 2)

#remove people who do not have phq2 or 9
overlap_group_with_phq2 <- overlap_group[which(!is.na(overlap_group$PHQ.2)),] #4864
overlap_group_with_phq9 <- overlap_group[which(!is.na(overlap_group$PHQ.9)),] #376

#subset people with unique scans (these are one time scanners)
overlap_group_no_duplicate_EMPI <- overlap_group[which(!duplicated(overlap_group$EMPI)),] #2675

overlap_group_with_phq2_no_duplicate_EMPI <- overlap_group_with_phq2[which(!duplicated(overlap_group_wit

overlap_group_with_phq9_no_duplicate_EMPI <- overlap_group_with_phq9[which(!duplicated(overlap_group_wit

#####take duplicated people and take first instance
```

```r
#get duplicated people
overlap_group_with_duplicate_EMPI <- overlap_group[which(duplicated(overlap_group$EMPI)),]

overlap_group_with_phq2_w_duplicate_EMPI <- overlap_group_with_phq2[which(duplicated(overlap_group_with_

overlap_group_with_phq9_w_duplicate_EMPI <- overlap_group_with_phq9[which(duplicated(overlap_group_with_


#get match - i.e. first instance of each
overlap_group_with_duplicated_EMPI_first_instance <- overlap_group_with_duplicate_EMPI[match(unique(over


overlap_group_with_phq2_duplicated_EMPI_first_instance <- overlap_group_with_phq2_w_duplicate_EMPI[match

overlap_group_with_phq9_duplicated_EMPI_first_instance <- overlap_group_with_phq9_w_duplicate_EMPI[match


#combine groups
overlap_group_unique_EMPI_combined <- rbind(overlap_group_no_duplicate_EMPI, overlap_group_with_duplicat

overlap_group_phq2_unique_empi_combined <- rbind(overlap_group_with_phq2_no_duplicate_EMPI, overlap_grou

overlap_group_phq9_unique_empi_combined <- rbind(overlap_group_with_phq9_no_duplicate_EMPI, overlap_grou

#summarize
print(paste0("N unique EMPIs in EB and Melissa overlap group = ", length(unique(overlap_group$EMPI)), "
```

```
## [1] "N unique EMPIs in EB and Melissa overlap group = 5475 out of 25857"
```

```r
print(paste0("N unique EMPIs in EB and Melissa overlap group w/phq2 = ", dim(overlap_group_phq2_unique_
```

```
## [1] "N unique EMPIs in EB and Melissa overlap group w/phq2 = 2620 out of 25857"
```
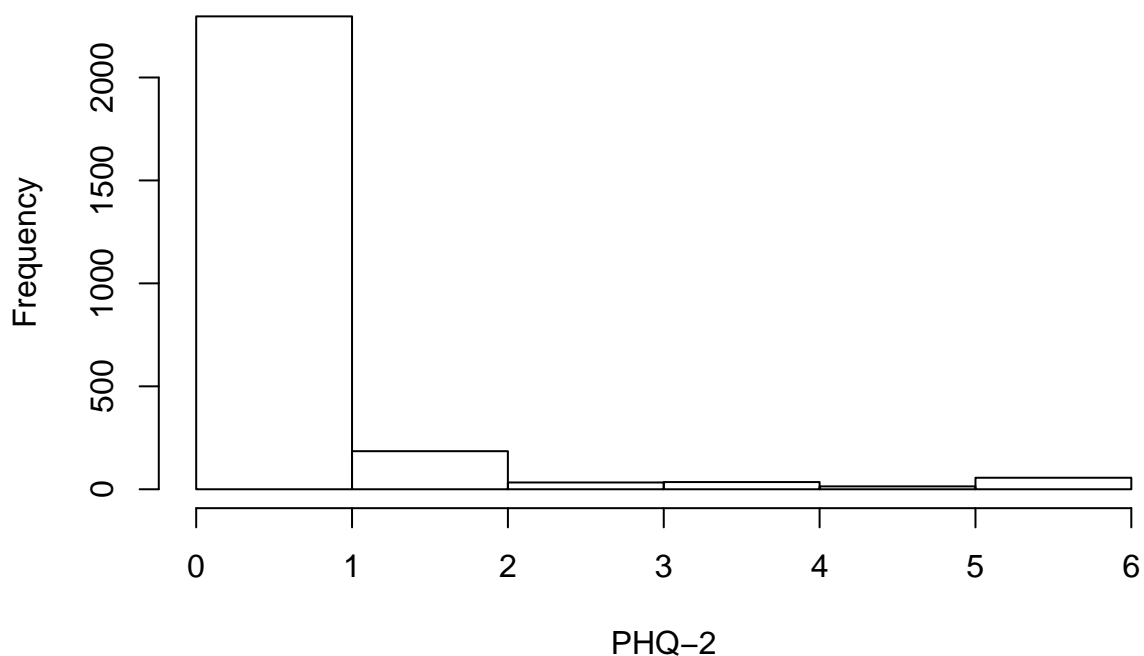
```r
print(paste0("N unique EMPIs in EB and Melissa overlap group w/phq9 = ", dim(overlap_group_phq9_unique_
```

```
## [1] "N unique EMPIs in EB and Melissa overlap group w/phq9 = 244 out of 25857"
```

```r
#histos
hist(overlap_group_phq2_unique_empi_combined$PHQ.2, main = paste0("PHQ-2 Histo combined group w/unique
```
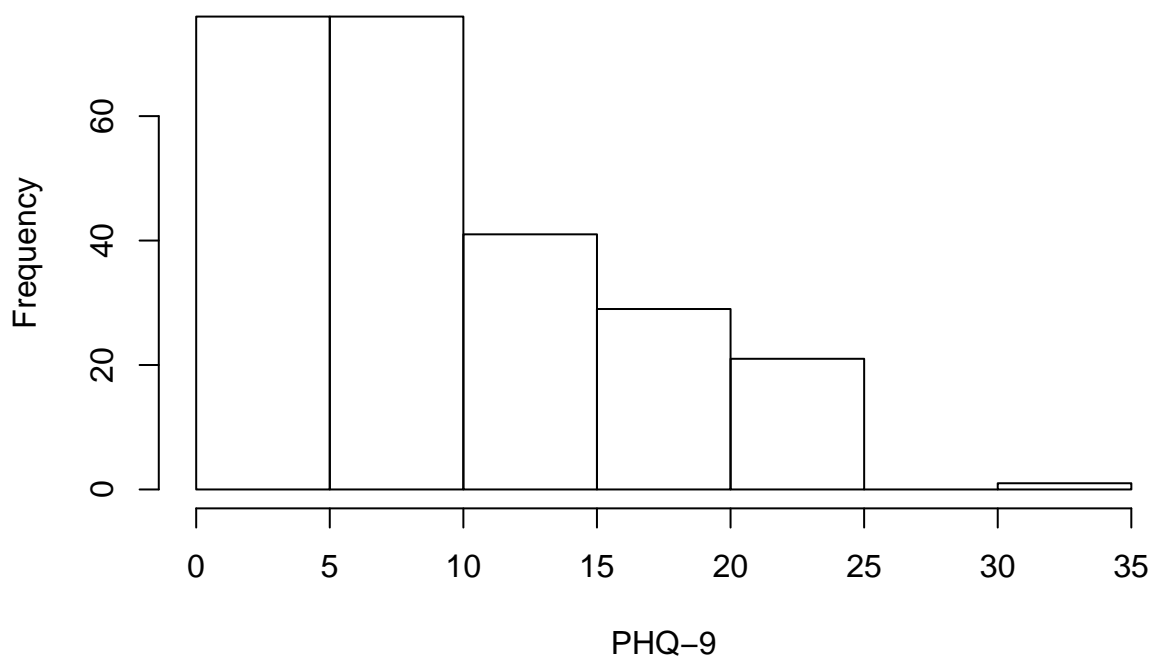
PHQ−2 Histo combined group w/unique EMPI vs all EMPI: n = 2620/77

PHQ−2

```
#histos
hist(overlap_group_phq9_unique_empi_combined$PHQ.9, main = paste0("PHQ-9 Histo combined group w/unique
```

PHQ−9 Histo combined group w/unique EMPI vs all EMPI: n = 244/60

PHQ−9

```
#########################
### Erica's group  ####
#########################
```

```
data_e <- data3

#make accession # and lab findings into integer type, contained in row 28:42, Margin=2 to work on colum
data_e$ACCESSION_NUM <- as.integer(data_e$ACCESSION_NUM)
data_e[28:42] <- apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2)
```

```
## Warning in apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2): NAs
## introduced by coercion
```

```
## Warning in apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2): NAs
## introduced by coercion
```

```
## Warning in apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2): NAs
## introduced by coercion
```

```
## Warning in apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2): NAs
## introduced by coercion
```

```
## Warning in apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2): NAs
## introduced by coercion
```

```
## Warning in apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2): NAs
## introduced by coercion
```

```
## Warning in apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2): NAs
## introduced by coercion
```

```
## Warning in apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2): NAs
## introduced by coercion
```

```
## Warning in apply(X = data_e[28:42], FUN = as.integer, MARGIN = 2): NAs
## introduced by coercion
```

```
data_e$sex_binarized <- ifelse(data_e$SEX == "MALE", 1, 2)
data_e$race_binarized <- ifelse(data_e$RACE == "WHITE", 1, 2)

#remove people who do not have phq2 or 9
data_e_with_phq2 <- data_e[which(!is.na(data_e$PHQ.2)),] #4864
data_e_with_phq9 <- data_e[which(!is.na(data_e$PHQ.9)),] #376

#subset people with unique scans (these are one time scanners)
data_e_no_duplicate_EMPI <- data_e[which(!duplicated(data_e$EMPI)),]

data_e_with_phq2_no_duplicate_EMPI <- data_e_with_phq2[which(!duplicated(data_e_with_phq2$EMPI)),]

data_e_with_phq9_no_duplicate_EMPI <- data_e_with_phq9[which(!duplicated(data_e_with_phq9$EMPI)),]

#####take duplicated people and take first instance

#get duplicated people
data_e_with_duplicate_EMPI <- data_e[which(duplicated(data_e$EMPI)),]

data_e_with_phq2_w_duplicate_EMPI <- data_e_with_phq2[which(duplicated(data_e_with_phq2$EMPI)),]

data_e_with_phq9_w_duplicate_EMPI <- data_e_with_phq9[which(duplicated(data_e_with_phq9$EMPI)),]
```

```r
#get match - i.e. first instance of each
data_e_with_duplicated_EMPI_first_instance <- data_e_with_duplicate_EMPI[match(unique(data_e_with_duplic

data_e_with_phq2_duplicated_EMPI_first_instance <- data_e_with_phq2_w_duplicate_EMPI[match(unique(data_e

data_e_with_phq9_duplicated_EMPI_first_instance <- data_e_with_phq9_w_duplicate_EMPI[match(unique(data_e


#combine groups
data_e_unique_EMPI_combined <- rbind(data_e_no_duplicate_EMPI, data_e_with_duplicated_EMPI_first_instan

data_e_phq2_unique_empi_combined <- rbind(data_e_with_phq2_no_duplicate_EMPI, data_e_with_phq2_duplicate

data_e_phq9_unique_empi_combined <- rbind(data_e_with_phq9_no_duplicate_EMPI, data_e_with_phq9_duplicate

#summarize
#print(paste0("N unique EMPIs in EB group = ", dim(data_e_unique_EMPI_combined)[1], " out of ", dim(dat

print(paste0("N unique EMPIs in EB group = ", length(unique(data_e$EMPI)), " out of ", dim(data_e)[1]))
```

```
## [1] "N unique EMPIs in EB group = 3985 out of 28774"
```

```r
print(paste0("N unique EMPIs in EB group w/phq2 = ", dim(data_e_phq2_unique_empi_combined)[1], " out of
```

```
## [1] "N unique EMPIs in EB group w/phq2 = 1868 out of 28774"
```

```r
print(paste0("N unique EMPIs in EB group w/phq9 = ", dim(data_e_phq9_unique_empi_combined)[1], " out of
```

```
## [1] "N unique EMPIs in EB group w/phq9 = 134 out of 28774"
```

```r
print(paste0("N in EB group w/phq2 = ", length(which(!is.na(data_e_no_duplicate_EMPI$PHQ.2))), " out of
```

```
## [1] "N in EB group w/phq2 = 1003 out of 28774"
```

```r
print(paste0("N in EB group w/phq9 = ", length(which(!is.na(data_e_no_duplicate_EMPI$PHQ.9))), " out of
```

```
## [1] "N in EB group w/phq9 = 70 out of 28774"
```

```r
print(paste0("N unique EMPIs in Melissa's group = ", length(unique(melissa_xls$EMPI)), " out of ", dim(m
```

```
## [1] "N unique EMPIs in Melissa's group = 4411 out of 30915"
```

```r
print(paste0("N in EB + Melissa overlap group w/depression ICD 9/10 = ", length(which(!is.na(overlap_gro
```

```
## [1] "N in EB + Melissa overlap group w/depression ICD 9/10 = 25857 out of 25857"
```

```r
print(paste0("N in EB & Melissa's group with dep ICD 9/10 and phq2 = ", length(which(!is.na(overlap_grou
```

```
## [1] "N in EB & Melissa's group with dep ICD 9/10 and phq2 = 7766 out of 25857"
```

```r
print(paste0("N in EB & Melissa's group with dep ICD 9/10 and phq9 = ", length(which(!is.na(overlap_grou
```

```
## [1] "N in EB & Melissa's group with dep ICD 9/10 and phq9 = 604 out of 25857"
```

```r
#histos
hist(data_e_phq2_unique_empi_combined$PHQ.2, main = paste0("PHQ-2 Histo EB group w/unique EMPI vs all EM
```

PHQ−2 Histo EB group w/unique EMPI vs all EMPI: n = 1868/9415
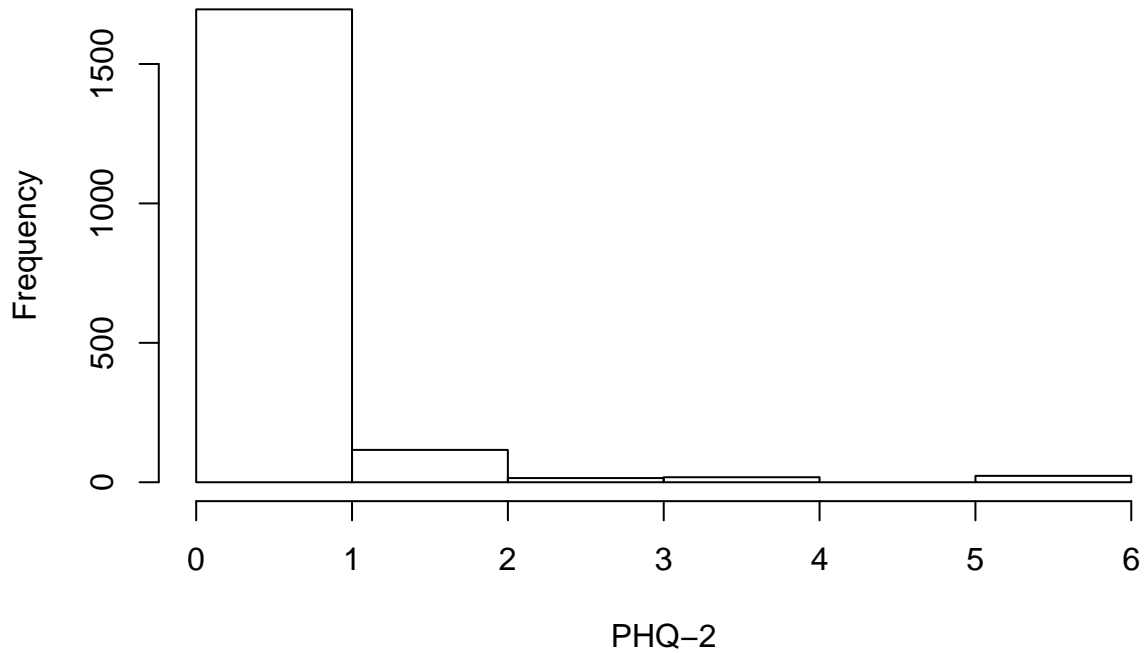

```
#histos
hist(data_e_phq9_unique_empi_combined$PHQ.9, main = paste0("PHQ-9 Histo EB group w/unique EMPI vs all E
```

PHQ−9 Histo EB group w/unique EMPI vs all EMPI: n = 134/798