

Exploring and Analysing BRFSS data

Setup

Load packages ggplot2 and dplyr Load data provided 'brfss2013' ### Load packages

```
library(ggplot2)
library(dplyr)
```

Load data

```
load("brfss2013.RData")
```

Part 1: Data

The Behavioral Risk Factor Surveillance System (BRFSS) is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. The interviewers collect data from randomly selected adult in a household in conducting BRFSS landline telephone survey and from an adult who participates by using a cellular telephone and resides in a private residence or college housing. BRFSS data was collected through the random phone call surveys for non-institutionalized population., therefore the findings are generalized to US population that has either a landline telephone or cellular phone. This study is not generalized to those institutionalized or those without telephones. Since all the data was collected (observed), this study is not based on any random experiment, hence there is no causal inference.

```
str(brfss2013)
```

```
## 'data.frame':    491775 obs. of  330 variables:
## $ X_state : Factor w/ 55 levels "0","Alabama",...: 2 2 2 2 2 2 2 2 2 ...
## $ fmonth : Factor w/ 12 levels "January","February",...: 1 1 1 1 2 3 3 3 4 4 ...
## $ idate : int 1092013 1192013 1192013 1112013 2062013 3272013 3222013 3042013 4242013 4242013 ...
## $ imonth : Factor w/ 12 levels "January","February",...: 1 1 1 1 2 3 3 3 4 4 ...
## $ iday : Factor w/ 31 levels "1","2","3","4",...: 9 19 19 11 6 27 22 4 24 24 ...
## $ iyear : Factor w/ 2 levels "2013","2014": 1 1 1 1 1 1 1 1 1 1 ...
## $ dispcode : Factor w/ 2 levels "Completed interview",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ seqno : int 2013000580 2013000593 2013000600 2013000606 2013000608 2013000630 2013000634 2013000644 2013001305 2013001338 ...
## $ X_psu : int 2013000580 2013000593 2013000600 2013000606 2013000608 2013000630 2013000634 2013000644 2013001305 2013001338 ...
## $ ctelenum : Factor w/ 1 level "Yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ pvtresd1 : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
## $ colghous : Factor w/ 1 level "Yes": NA NA NA NA NA NA NA NA NA ...
## $ stateres : Factor w/ 1 level "Yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ cellfon3 : Factor w/ 1 level "Not a cellular phone": 1 1 1 1 1 1 1 1 1 1 ...
## $ ladult : Factor w/ 2 levels "Yes, male respondent",...: NA NA NA NA NA NA NA NA NA ...
## $ numadult : Factor w/ 19 levels "1","2","3","4",...: 2 2 3 2 2 1 2 1 5 2 ...
## $ nummen : Factor w/ 14 levels "0","1","2","3",...: 2 2 3 2 2 1 2 1 5 2 ...
## $ numwomen : Factor w/ 12 levels "0","1","2","3",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ genhlth : Factor w/ 5 levels "Excellent","Very good",...: 4 3 3 2 3 2 4 3 1 3 ...
## $ physhlth : int 30 0 3 2 10 0 1 5 0 0 ...
## $ menthlth : int 29 0 2 0 2 0 15 0 0 0 ...
## $ poorhlth : int 30 NA 0 0 0 NA 0 10 NA NA ...
## $ hlthpln1 : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
## $ persdoc2 : Factor w/ 3 levels "Yes, only one",...: 1 1 1 1 1 1 2 1 1 1 ...
## $ medcost : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
## $ checkup1 : Factor w/ 5 levels "Within past year",...: 1 1 1 2 4 1 1 1 1 1 ...
## $ sleptim1 : int NA 6 9 8 6 8 7 6 8 8 ...
## $ bphigh4 : Factor w/ 4 levels "Yes","Yes, but female told only during pregnancy",...: 1 3 3 3 1 1 1 1 3 3 ...
## $ bpmeds : Factor w/ 2 levels "Yes","No": 1 NA NA NA 2 1 1 1 NA NA ...
## $ bloodcho : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
## $ cholchk : Factor w/ 4 levels "Within past year",...: 1 1 4 1 2 1 1 1 1 1 ...
## $ toldhi2 : Factor w/ 2 levels "Yes","No": 1 2 2 1 2 1 2 1 1 2 ...
## $ cvdinf4 : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
## $ cvdcrhd4 : Factor w/ 2 levels "Yes","No": NA 2 2 2 2 2 2 1 2 2 ...
## $ cvdstrk3 : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
## $ asthma3 : Factor w/ 2 levels "Yes","No": 1 2 2 2 1 2 2 2 2 2 ...
## $ asthnow : Factor w/ 2 levels "Yes","No": 1 NA NA NA 2 NA NA NA NA NA ...
## $ chcsncr : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
## $ chcocnrc : Factor w/ 2 levels "Yes","No": 2 2 2 2 1 2 2 2 2 2 ...
## $ chccopd1 : Factor w/ 2 levels "Yes","No": 1 2 2 2 2 2 2 2 2 2 ...
## $ havarth3 : Factor w/ 2 levels "Yes","No": 1 2 1 2 2 2 1 1 1 2 ...
## $ addepev2 : Factor w/ 2 levels "Yes","No": 1 1 1 2 2 2 2 2 2 2 ...
## $ chckidny : Factor w/ 2 levels "Yes","No": 1 2 2 2 2 2 2 2 2 2 ...
## $ diabete3 : Factor w/ 4 levels "Yes","Yes, but female told only during pregnancy",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ veteran3 : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
## $ marital : Factor w/ 6 levels "Married","Divorced",...: 2 1 1 1 2 1 3 1 1 1 ...
## $ children : int 0 2 0 0 0 0 1 0 1 0 ...
## $ educa : Factor w/ 6 levels "Never attended school or only kindergarten",...: 6 5 6 4 6 6 4 5 6 4 ...
## $ employ1 : Factor w/ 8 levels "Employed for wages",...: 7 1 1 7 7 1 1 7 7 5 ...
## $ income2 : Factor w/ 8 levels "Less than $10,000",...: 7 8 8 7 6 8 NA 6 8 4 ...
## $ weight2 : Factor w/ 570 levels "", "b", "100",...: 154 30 63 31 169 128 9 1 139 73 ...
## $ height3 : int 507 510 504 504 600 503 500 505 602 505 ...
## $ numhhol2 : Factor w/ 2 levels "Yes","No": 1 2 2 2 2 1 2 2 2 2 ...
## $ numphon3 : Factor w/ 6 levels "1 residential telephone number",...: 2 NA NA NA NA 1 NA NA NA NA ...
## $ cpdemo1 : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
## $ cpdemo4 : int 10 70 70 75 0 70 40 1 60 50 ...
## $ internet : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
## $ renthom1 : Factor w/ 3 levels "Own","Rent","Other arrangement": 1 1 1 1 1 1 1 2 1 1 ...
## $ sex : Factor w/ 2 levels "Male","Female": 2 2 2 2 1 2 2 2 1 2 ...
## $ pregnant : Factor w/ 2 levels "Yes","No": NA NA NA NA NA NA 2 NA NA NA ...
## $ qlactlm2 : Factor w/ 2 levels "Yes","No": 1 2 1 2 2 2 1 1 2 2 ...
## $ useequip : Factor w/ 2 levels "Yes","No": 1 2 2 2 2 2 2 2 2 2 ...
## $ blind : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
## $ decide : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
## $ diffwalk : Factor w/ 2 levels "Yes","No": 1 2 1 2 2 2 2 1 2 2 ...
## $ diffdres : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
## $ diffalor : Factor w/ 2 levels "Yes","No": 1 2 2 2 2 2 2 2 2 2 ...
## $ smoke100 : Factor w/ 2 levels "Yes","No": 1 2 1 2 1 2 1 1 2 2 ...
## $ smokday2 : Factor w/ 3 levels "Every day","Some days",...: 3 NA 2 NA 3 NA 3 1 NA NA ...
## $ stopsmk2 : Factor w/ 2 levels "Yes","No": NA NA 1 NA NA NA NA 2 NA NA ...
## $ lastsmk2 : Factor w/ 8 levels "Within the past month",...: 7 NA NA NA 1 NA 5 NA NA NA ...
## $ usenow3 : Factor w/ 3 levels "Every day","Some days",...: 3 3 3 3 3 3 3 3 1 3 ...
## $ alcdays : int 201 0 220 208 210 0 201 202 101 0 ...
## $ avedrnk2 : int 2 NA 4 2 2 NA 1 1 1 NA ...
## $ drnk3ge5 : int 0 NA 20 0 0 NA 0 0 0 NA ...
## $ maxdrnks : int 2 NA 10 2 3 NA 1 1 2 NA ...
## $ fruitju1 : int 304 305 301 202 0 205 320 0 0 202 ...
## $ fruit1 : int 104 301 203 306 302 206 325 320 101 202 ...
## $ fvbeans : int 303 310 202 202 101 0 330 360 202 203 ...
## $ fvgreen : int 310 203 202 310 310 203 315 315 203 201 ...
## $ fvorang : int 303 202 310 305 303 0 310 325 0 201 ...
## $ vegetab1 : int NA 203 330 204 101 207 310 308 101 203 ...
## $ exerany2 : Factor w/ 2 levels "Yes","No": 2 1 2 1 2 1 1 1 1 1 ...
## $ exract11 : Factor w/ 75 levels "Active Gaming Devices (Wii Fit, Dance, Dance revolution)",...: NA 64 NA 64 NA 6 64 64 7 64 ...
## $ exeroft1 : int NA 105 NA 205 NA 102 220 102 102 220 ...
## $ exerhmm1 : int NA 20 NA 30 NA 15 100 15 100 30 ...
## $ exract21 : Factor w/ 76 levels "Active Gaming Devices (Wii Fit, Dance, Dance revolution)",...: NA 71 NA 75 NA 18 75 75 75 18 ...
## $ exeroft2 : int NA 101 NA NA NA 102 NA NA NA 101 ...
## $ exerhmm2 : int NA 10 NA NA NA 30 NA NA NA 100 ...
## $ strength : int 0 0 0 0 0 205 0 102 0 ...
## $ lmtjoin3 : Factor w/ 2 levels "Yes","No": 1 NA 1 NA NA NA 2 1 2 NA ...
```

```
## $ arthdis2 : Factor w/ 2 levels "Yes","No": 1 NA 1 NA NA NA 1 2 2 NA ...
## $ arthsoc1 : Factor w/ 3 levels "A lot","A little",...: 1 NA 2 NA NA NA 3 1 3 NA ...
## $ joinpain : int 7 NA 5 NA NA NA 3 8 4 NA ...
## $ seatbelt : Factor w/ 6 levels "Always","Nearly always",...: 1 1 1 1 1 1 1 2 1 ...
## $ flushot6 : Factor w/ 2 levels "Yes","No": 2 1 1 2 2 1 2 1 1 2 ...
## $ flshmt2 : Factor w/ 26 levels "January 2012",...: NA 10 13 NA NA NA NA 10 10 NA ...
## $ tetanus : Factor w/ 4 levels "Yes, received Tdap",...: 4 1 1 4 4 4 4 1 4 ...
## $ pneuvac3 : Factor w/ 2 levels "Yes","No": 1 2 2 2 2 1 2 2 2 2 ...
## [list output truncated]
```

```
dim(brfss2013)
```

```
## [1] 491775 330
```

```
names(brfss2013)
```

```
## [1] "X_state" "fmonth" "idate" "imonth" "iday"
## [6] "iyear" "dispcode" "seqno" "X_psu" "ctelenum"
## [11] "pvtresd1" "colghous" "stateres" "cellfon3" "ladult"
## [16] "numadult" "nummen" "numwomen" "genhlth" "physhlth"
## [21] "menthlth" "poorhlth" "hlthpln1" "persdoc2" "medcost"
## [26] "checkup1" "sleptim1" "bphigh4" "bpmeds" "bloodcho"
## [31] "cholchk" "toldhi2" "cvdinfr4" "cvdcrhd4" "cvdstrk3"
## [36] "asthma3" "asthnow" "chcscncr" "chcocncr" "chcccpd1"
## [41] "havarth3" "addepev2" "chckidny" "diabete3" "veteran3"
## [46] "marital" "children" "educa" "employ1" "income2"
## [51] "weight2" "height3" "numhhol2" "numphon2" "cpdemo1"
## [56] "cpdemo4" "internet" "renthom1" "sex" "pregnant"
## [61] "qlactlm2" "useequip" "blind" "decide" "diffwalk"
## [66] "differes" "diffalon" "smoke100" "smokday2" "stopsmk2"
## [71] "lastsmk2" "usenow3" "alcdays" "avedrnk2" "drnk3ge5"
## [76] "maxdrnks" "fruitju1" "fruit1" "fvbeans" "fvgreen"
## [81] "fvorang" "vegetabl1" "exerany2" "extract11" "exeroft1"
## [86] "exerhmm1" "extract21" "exeroft2" "exerhmm2" "strength"
## [91] "lmtjoin3" "arthdis2" "arthsoc1" "joinpain" "seatbelt"
## [96] "flushot6" "flshmt2" "tetanus" "pneuvac3" "hivst6"
## [101] "hivtstd3" "whrtst10" "pdiabtst" "prediab1" "diabage2"
## [106] "insulin" "bldsugar" "feetchk2" "doctdiab" "chkhemo3"
## [111] "feetchk" "eyeexam" "diabeye" "diabedu" "painact2"
## [116] "qlmentl2" "qlstres2" "qlhlth2" "medicare" "hlthcvrg"
## [121] "delaymed" "dlyother" "nocov121" "lstcovrg" "drvisits"
## [126] "medscost" "carercvd" "medbills" "ssbsugar" "ssbfrut2"
## [131] "wtchsalt" "longwtch" "dradvise" "asthmage" "asattack"
## [136] "aservist" "asdrvist" "asrchkup" "asactlim" "asymptom"
## [141] "asnoslep" "asthmed3" "asinhalt" "harehab1" "strehab1"
## [146] "cvdasprn" "aspunsaf" "rlivpain" "rduchart" "rducstrk"
## [151] "arttoday" "arthwgt" "arthexer" "arthedu" "imfvplac"
## [156] "hpvadvc2" "hpvadsh" "hadmam" "howlong" "profexam"
## [161] "lengexam" "hadpap2" "lastpap2" "hadhyst2" "bldstool"
## [166] "lstblsds3" "hadsigm3" "hadsgeo1" "lastsig3" "pcpsaad2"
## [171] "pcpsadi1" "pcpsare1" "psatest1" "psatime" "pcpsars1"
## [176] "pcpsade1" "pcdmdecn" "rrrclass2" "rrrcognt2" "rrratwrk2"
## [181] "rrhcare3" "rrrphysm2" "rrremtsm2" "misnervs" "mishopl5"
## [186] "misrstls" "misdeprd" "miseffrt" "miswtles" "misnwrk"
## [191] "mistmnt" "mistrhlp" "misphlpf" "scntmony" "scntmeal"
## [196] "scntpaid" "scntwrk1" "scntlpad" "scntlwk1" "scntvot1"
## [201] "rcsgendr" "rcsr1tn2" "casthdx2" "casthno2" "emtsuprt"
## [206] "lsatisfy" "ctelnum1" "cellfon2" "cadult" "pvtresd2"
## [211] "cclghous" "cstate" "landline" "pctcell" "qstver"
## [216] "qstlang" "mscode" "X_ststr" "X_strwt" "X_rawrake"
## [221] "X_wt2rake" "X_imprace" "X_impnp" "X_impeduc" "X_impmt1"
## [226] "X_imphone" "X_chispnc" "X_crace1" "X_impceage" "X_impccrac"
## [231] "X_impcesex" "X_cllcpwt" "X_dualuse" "X_dualcor" "X_llcpwt2"
## [236] "X_llcpwt" "X_rfh1th" "X_hcvu651" "X_rfhype5" "X_cholchk"
## [241] "X_rfcho1" "X_lstasth1" "X_asthml1" "X_asthms1" "X_drxdar1"
## [246] "X_prace1" "X_mrce1" "X_hispanc" "X_race" "X_raceg21"
## [251] "X_racegr3" "X_race_g1" "X_age65yr" "X_age65yr" "X_age_g"
## [256] "htin4" "htm4" "wtkg3" "X_bmi5" "X_bmi5cat"
## [261] "X_rfbmi5" "X_chldcnt" "X_educag" "X_incomg" "X_smoker3"
## [266] "X_rfsmok3" "drnkany5" "drocdy3_" "X_rfbing5" "X_drnkdy4"
## [271] "X_drnkmo4" "X_rfdrrhv4" "X_rfdrrmn4" "X_rfdrrwm4" "ftjuda1_"
## [276] "frutda1_" "beanday_" "grenday_" "orngday_" "vegeda1_"
## [281] "X_misfrtn" "X_misvegn" "X_frtresp" "X_vegresp" "X_frutsum"
## [286] "X_vegesum" "X_frtlt1" "X_vegl1t1" "X_frt16" "X_veg23"
## [291] "X_fruitex" "X_vegetex" "X_totinda" "metv111_" "metv121_"
## [296] "maxvo2_" "fc60_" "actin11_" "actin21_" "padur1_"
## [301] "padur2_" "pafreq1_" "pafreq2_" "X_minac11" "X_minac21"
## [306] "strfreq_" "pamiss1_" "pamin11_" "pamin21_" "palmin_"
## [311] "pavig11_" "pavig21_" "palvigm_" "X_pacat1" "X_paindx1"
## [316] "X_pa150r2" "X_pa300r2" "X_pa30021" "X_pastrng" "X_parec1"
## [321] "X_pastae1" "X_lmtact1" "X_lmtwrk1" "X_lmtsc11" "X_rfseat2"
## [326] "X_rfseat3" "X_f1shot6" "X_pneumo2" "X_aidtst3" "X_age80"
```

Part 2: Research questions

Research question 1: What is the relationship between education and income level?

Research question 2: What is the impact of number of hours slept on general health of males and females?

Research question 3: In year 2013, what percent of males and females got the flushots in the month of November?

Part 3: Exploratory data analysis

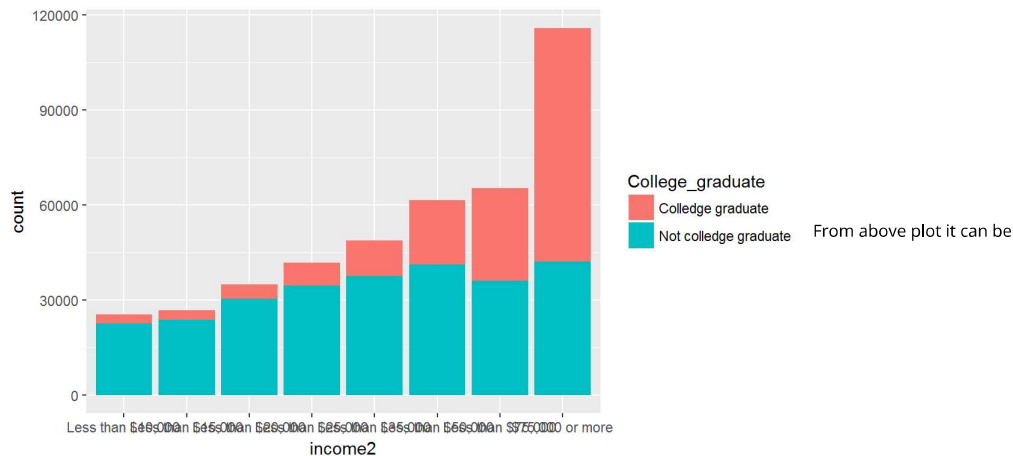
Research question 1:

```
study1<-brfss2013 %>%  
  select(income2, educa)
```

```
study1_NAreMOVED <- study1 %>%  
  filter(!is.na(educa)) %>%  
  filter(!is.na(income2))
```

```
study1_NAreMOVED <- study1_NAreMOVED %>%  
  mutate(College_graduate = ifelse(educa == "College 4 years or more (College graduate)", "College graduate", "Not college graduate"))
```

```
ggplot(study1_NAreMOVED, aes(income2, fill = College_graduate))+geom_bar(position = "stack")
```



configured that income is dependent on the level of education. Those who are College graduate they fall in the higher income range.

Research question 2:

```
study2<-brfss2013 %>%  
  select(sleptim1, physhlth, sex)
```

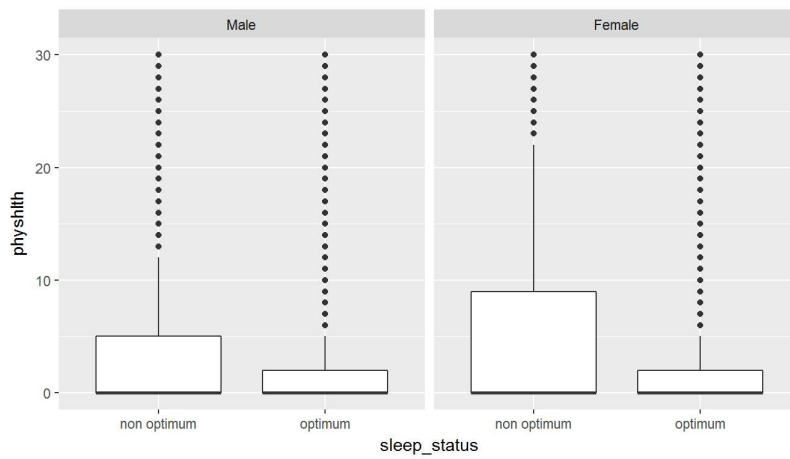
```
study2_NAreMOVED <- study2 %>%  
  filter(!is.na(sleptim1)) %>%  
  filter(!is.na(physhlth)) %>%  
  filter(!is.na(sex))
```

```
study2_NAreMOVED <- study2_NAreMOVED %>%  
  mutate(sleep_status = ifelse(sleptim1 >= 7 & sleptim1 <= 10, "optimum", "non optimum"))
```

```
study2_NAreMOVED %>%  
  group_by(sleep_status, sex) %>% summarise(mean_physhlth = mean(physhlth), sd_physhlth = sd(physhlth), n = n())
```

```
## Source: local data frame [4 x 5]  
## Groups: sleep_status [?]  
##  
##   sleep_status sex mean_physhlth sd_physhlth    n  
##   <chr> <fctr>      <dbl>      <dbl> <int>  
## 1 non optimum Male    5.451177  9.978425 66659  
## 2 non optimum Female  6.483813 10.371695 94427  
## 3 optimum Male    3.064919  7.552099 128714  
## 4 optimum Female  3.615448  7.942115 184659
```

```
ggplot(study2_NAreMOVED, aes(x = sleep_status, y = physhlth)) + geom_boxplot() + facet_wrap(~sex)
```



The above plots indicate that those

who have optimum hours of sleep more than or equal to 7 hours and less than 10 hours, their physical health is very good for both males and females.

Research question 3:

```
study3 <- brfss2013 %>%
  select(imonth, sex, flushot6)
```

```
study3_NAremoved <- study3 %>%
  filter(!is.na(imonth)) %>%
  filter(!is.na(sex)) %>%
  filter(!is.na(flushot6))
```

```
study3_November <- study3_NAremoved %>%
  filter(imonth == "November")
```

```
study3_November %>%
  count(flushot6, sex)
```

```
## Source: local data frame [4 x 3]
## Groups: flushot6 [?]
##
##   flushot6  sex    n
##   <fctr> <fctr> <int>
## 1     Yes  Male  6554
## 2     Yes Female 10324
## 3     No   Male  9227
## 4     No  Female 12009
```

```
6554 / (6554 + 9227)
```

```
## [1] 0.4153095
```

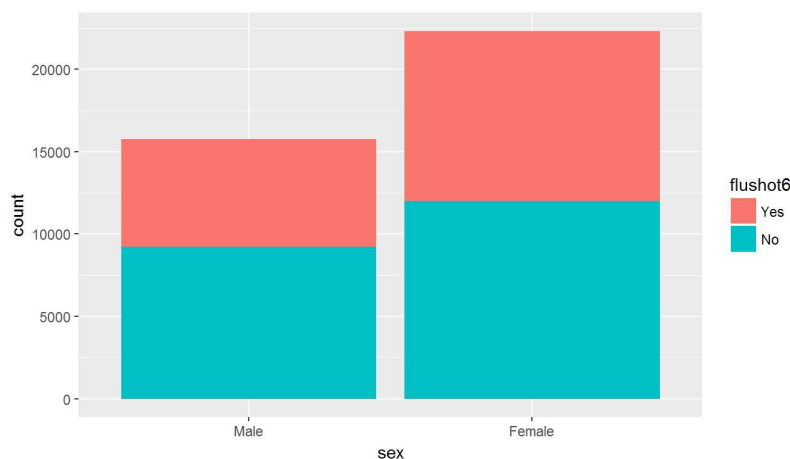
Approx. 41% males got the flushots in November 2013.

```
10324 / (10324 + 12009)
```

```
## [1] 0.4622756
```

Approx. 46% females got the flushots in November 2013.

```
ggplot(study3_November, aes(sex, fill = flushot6)) + geom_bar()
```



This plot and the calculation

indicates that in November 2013, more females got the flushots than males.