FirstAttempy

Evan Yacek

Friday, August 05, 2016

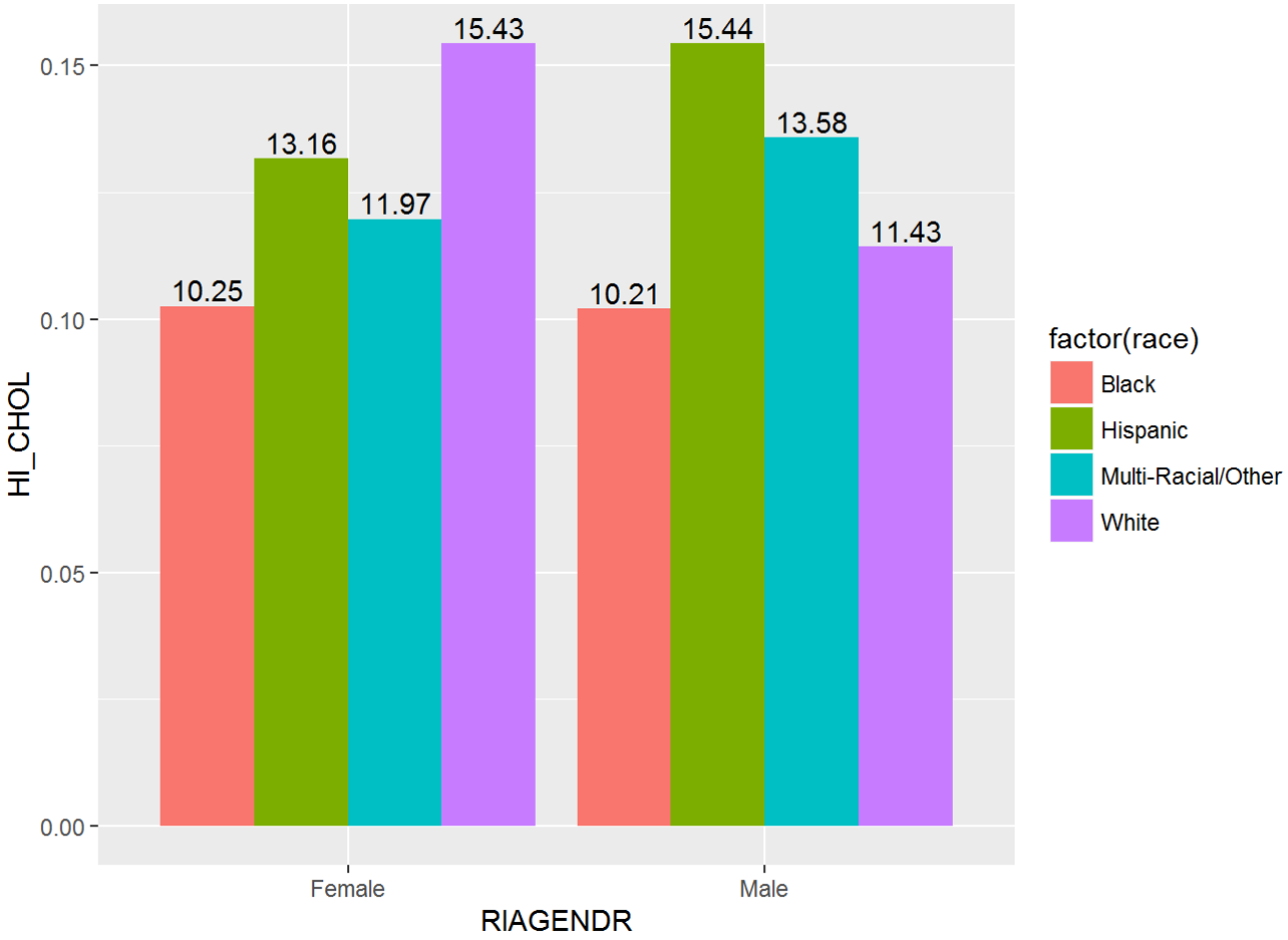
```
#MERGE data sets
NHANES2007DATA<-merge(DEMO2007,chol2007 , All = TRUE)
NHANES2007DATA<-merge(NHANES2007DATA,Diet2007, All = TRUE)
NHANES2007DATA<-merge(NHANES2007DATA,cons2007, All = TRUE)
NHANES2007DATA<-merge(NHANES2007DATA,OCP_E2007, All = TRUE)
NHANES2007DATA<-merge(NHANES2007DATA,health2007, All = TRUE)


NHANES2009DATA<-merge(DEMO2009,chol2009, All = TRUE)
NHANES2009DATA<-merge(NHANES2009DATA,Diet2009, All = TRUE)
NHANES2009DATA<-merge(NHANES2009DATA,cons2009, All = TRUE)
NHANES2009DATA<-merge(NHANES2009DATA,OCP_F2009, All = TRUE)
NHANES2009DATA<-merge(NHANES2009DATA,health2009, All = TRUE)


mergedSET <- rbind(NHANES2009DATA, NHANES2007DATA)

mergedSET <-subset(mergedSET, RIDSTATR %in% 2 | RIDSTATR %in% 1)
mergedSET<-transform(mergedSET,  HI_CHOL = ifelse(LBXTC>=240,1,0))
mergedSET<-transform(mergedSET, race=c( 3 , 3 , 1 , 2 , 4 )[RIDRETH1])
mergedSET$weightedTotal <- 1
mergedSET$WTMEC4YR <- mergedSET$WTMEC2YR / 2
mergedSET$WTINT4YR <- mergedSET$WTINT2YR / 2
# recode the RIDRETH1 variable as:
    # mexican american and other hispanic -> 3
    # non-hispanic white -> 1
    # non-hispanic black -> 2
    # other race including multi-racial -> 4
mergedSET<-transform(mergedSET, agecat=cut(RIDAGEYR,c(0,19,39,59, Inf)))


mergedSET$CBQ020[mergedSET$CBQ020== 77] <- NA   #RECODE FOR DATA THAT CANT BE ANALYZED
mergedSET$CBQ020[mergedSET$CBQ020== 99] <- NA
mergedSET$CBQ030[mergedSET$CBQ030== 77] <- NA
mergedSET$CBQ030[mergedSET$CBQ030== 99] <- NA
mergedSET$CBQ040[mergedSET$CBQ040== 77] <- NA
mergedSET$CBQ040[mergedSET$CBQ040== 99] <- NA
mergedSET$CBQ060[mergedSET$CBQ060== 77] <- NA
mergedSET$CBQ060[mergedSET$CBQ060== 99] <- NA
mergedSET$DBQ700[mergedSET$DBQ700== 7] <- NA
mergedSET$DBQ700[mergedSET$DBQ700== 9] <- NA
mergedSET$DBD895[mergedSET$DBD895== 5555] <- NA
mergedSET$DBD895[mergedSET$DBD895== 9999] <- NA
mergedSET$HSD010[mergedSET$HSD010== 9] <- NA
mergedSET$CBD160[mergedSET$CBD160 == 777] <- NA
mergedSET$CBD160[mergedSET$CBD160 == 999] <- NA


NHANES2009DATA<-subset(NHANES2009DATA, RIDSTATR %in% 2)
NHANES2009DATA<-transform(NHANES2009DATA,  HI_CHOL = ifelse(LBXTC>=240,1,0))
```

```
NHANES2009DATA<-transform(NHANES2009DATA, race=c( 3 , 3 , 1 , 2 , 4 )[RIDRETH1])
NHANES2009DATA$weightedTotal <- 1
# recode the RIDRETH1 variable as:
    # mexican american and other hispanic -> 3
    # non-hispanic white -> 1
    # non-hispanic black -> 2
    # other race including multi-racial -> 4

NHANES2009DATA<-transform(NHANES2009DATA, agecat=cut(RIDAGEYR,c(0,19,39,59, Inf)))
```

ALL RECODING DONE, BUILD MODELS

```
designA <- svydesign(id=~SDMVPSU,
strata=~SDMVSTRA,nest=TRUE,weights=~WTMEC2YR,data=NHANES2009DATA)

designB <- svydesign(id=~SDMVPSU,
strata=~SDMVSTRA,nest=TRUE,weights=~WTINT4YR,data=mergedSET)

popage<-c(55901,77670,72816,45364)

racegender<-as.data.frame(svytable(~race+RIAGENDR,designA))
racegenderage<-expand.grid(race=1:4,RIAGENDR=1:2,agecat=levels(NHANES2009DATA$agecat))
racegenderage$Freq<- as.vector(outer(racegender$Freq, popage/sum(popage)))

racegender2<-as.data.frame(svytable(~race+RIAGENDR,designB))
racegenderage2<-expand.grid(race=1:4,RIAGENDR=1:2,agecat=levels(mergedSET$agecat))
racegenderage2$Freq<- as.vector(outer(racegender$Freq, popage/sum(popage)))


####REPLACE AND REWEIGHT NA

somedata<-subset(NHANES2009DATA, !is.na(LBXTC) )
somedata2 <- subset(mergedSET, !is.na(DBQ700) & !is.na(DBD895) & !is.na(DBD900))
somedata3 <- subset(mergedSET, !is.na(CBQ020) & !is.na(CBQ030) & !is.na(CBQ040) & !is.na(CBQ060) & !is.na(CB
D160))
somedata4 <- subset(mergedSET, !is.na(HSD010))


design1 <- svydesign(id=~SDMVPSU,
strata=~SDMVSTRA,nest=TRUE,weights=~WTMEC2YR,data=somedata)

design2 <- svydesign(id=~SDMVPSU,
strata=~SDMVSTRA,nest=TRUE,weights=~WTINT4YR,data=somedata2)

design3 <- svydesign(id=~SDMVPSU,
strata=~SDMVSTRA,nest=TRUE,weights=~WTINT4YR,data=somedata3)

design4 <- svydesign(id=~SDMVPSU,
strata=~SDMVSTRA,nest=TRUE,weights=~WTINT4YR,data=somedata4)


cholEX <- svyby(~HI_CHOL,~race+RIAGENDR,design=subset(postStratify(design1,~race+RIAGENDR+agecat,racegendera
ge),RIDAGEYR>=20),svymean,na.rm=TRUE)
cholEX$race <- c("White", "Black", "Hispanic", "Multi-Racial/Other", "White", "Black", "Hispanic", "Multi-Ra
cial/Other")
cholEX$RIAGENDR <- c("Male", "Male","Male","Male","Female","Female","Female","Female")
cholEX$HI_CHOLP <- round(cholEX$HI_CHOL, 4)*100
# recode the RIDRETH1 variable as:
    # mexican american and other hispanic -> 3
    # non-hispanic white -> 1
    # non-hispanic black -> 2
    # other race including multi-racial -> 4
```

```
ggplot(cholEX, aes(x=RIAGENDR, y=HI_CHOL, fill=factor(race)))+
  geom_bar(position="dodge", stat="identity") + geom_text(aes(label=HI_CHOLP),
position=position_dodge(width=0.9), vjust=-0.25)
```



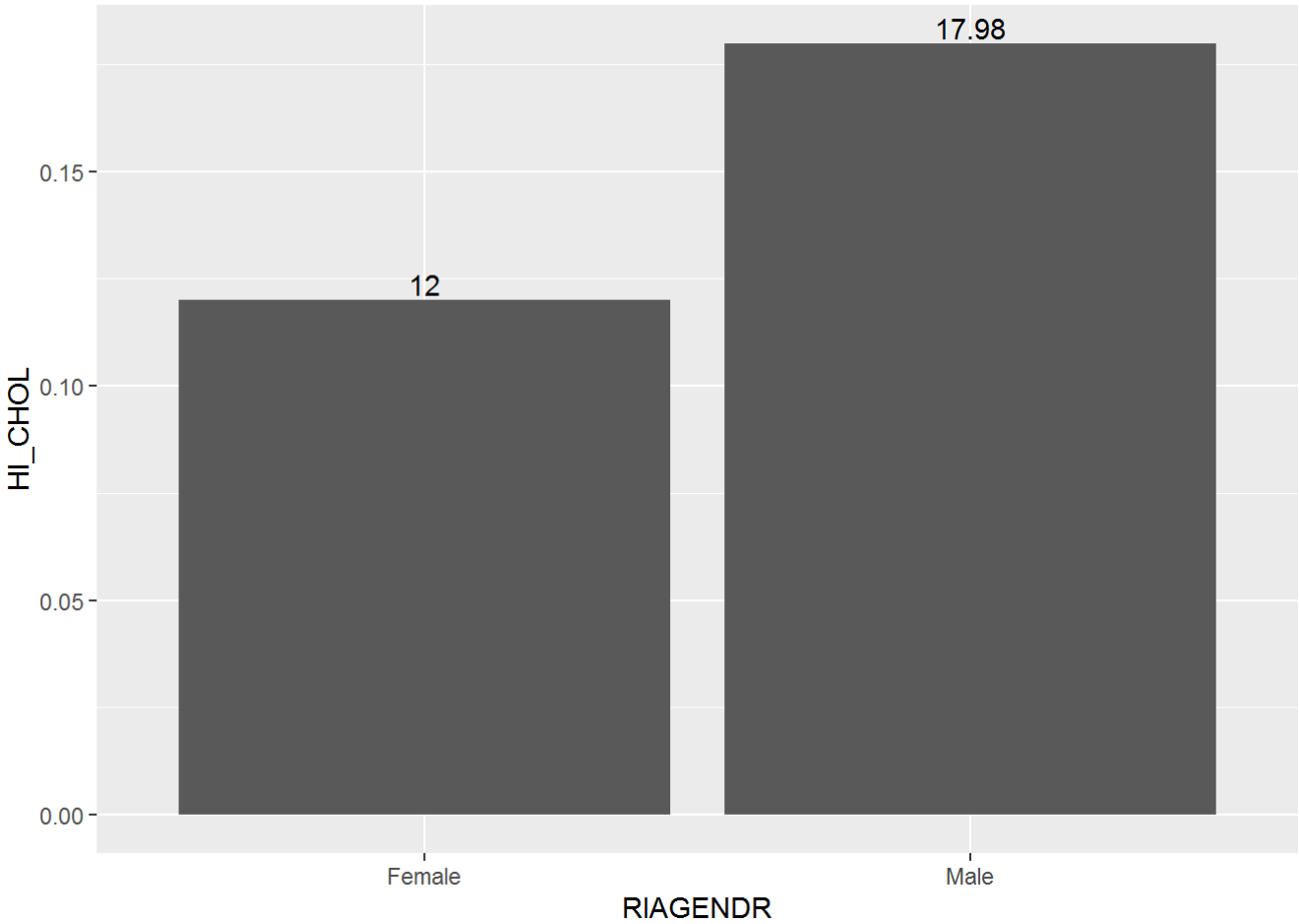

```
#ABOVE CHART Corresponds to http://www.cdc.gov/nchs/data/databriefs/db92_fig1.png

teachersD1 <- subset(postStratify(design1,~race+RIAGENDR+agecat,racegenderage),RIDAGEYR>=20 & OCD231 == 15 &
 OCD241 == 8)
teachersD2 <- subset(postStratify(design2,~race+RIAGENDR+agecat,racegenderage2),RIDAGEYR>=20 & OCD231 == 15
& OCD241 == 8)
teachersD3 <- subset(postStratify(design3,~race+RIAGENDR+agecat,racegenderage2),RIDAGEYR>=20 & OCD231 == 15
& OCD241 == 8)
teachersD4 <- subset(postStratify(design4,~race+RIAGENDR+agecat,racegenderage2),RIDAGEYR>=20 & OCD231 == 15
& OCD241 == 8)

#NOTE that this is 1 year data N = 134 for teachers
teachcholEX <- svyby(~HI_CHOL,~RIAGENDR,design= teachersD1,svymean,na.rm=TRUE)
teachcholEX
```

```
##   RIAGENDR   HI_CHOL         se
## 1        1 0.1797654 0.08611458
## 2        2 0.1199541 0.03104876
```

```
cholCompare <- svyby(~HI_CHOL,~RIAGENDR,design= subset(postStratify(design1,~race+RIAGENDR+agecat,racegender
age),RIDAGEYR>=20),svymean,na.rm=TRUE)
#Compared to NHANES2009
cholCompare
```

```
##   RIAGENDR   HI_CHOL          se
## 1        1 0.1204007 0.008613865
## 2        2 0.1422642 0.007895680
```

```
teachcholEX$RIAGENDR <- c("Male", "Female")
teachcholEX$HI_CHOLP <- round(teachcholEX$HI_CHOL, 4)*100
# recode the RIDRETH1 variable as:
    # mexican american and other hispanic -> 3
    # non-hispanic white -> 1
    # non-hispanic black -> 2
    # other race including multi-racial -> 4
ggplot(teachcholEX, aes(x=RIAGENDR, y=HI_CHOL))+
  geom_bar(position="dodge", stat="identity") + geom_text(aes(label=HI_CHOLP),
position=position_dodge(width=0.9), vjust=-0.25)
```



```
# Number of Meals not prepared at home N = 244
teachMeanMeals <- svyby(~DBD895,~RIAGENDR,design= teachersD2,svymean,na.rm=TRUE)
meanMeals <- svyby(~DBD895,~RIAGENDR,design= subset(postStratify(design2,~race+RIAGENDR+agecat,racegenderage
2),RIDAGEYR>=20),svymean,na.rm=TRUE)
teachMeanMeals
```

```
##   RIAGENDR   DBD895        se
## 1        1 5.492258 0.5886663
## 2        2 3.822003 0.2884262
```

```
meanMeals
```

```
##   RIAGENDR   DBD895         se
## 1        1 5.210445 0.09475403
## 2        2 3.847865 0.05515648
```

```
# Number of Meals from a fastfood place a week N = 244
teachFastMeals <- svyby(~DBD900,~RIAGENDR,design= teachersD2,svymean,na.rm=TRUE)
fastMeals <- svyby(~DBD900,~RIAGENDR,design= subset(postStratify(design2,~race+RIAGENDR+agecat,racegenderage
2),RIDAGEYR>=20),svymean,na.rm=TRUE)
teachFastMeals
```

```
##   RIAGENDR   DBD900        se
## 1        1 2.104578 0.4871250
## 2        2 1.452956 0.1813128
```

```
fastMeals
```

```
##   RIAGENDR   DBD900        se
## 1        1 2.882372 0.5376737
## 2        2 1.646069 0.0460875
```

```
#Number of times someone cooked dinner at home N = 275
teachDinMeals <- svyby(~CBD160, ~RIAGENDR, design = teachersD3, svymean, na.rm = TRUE)
dinMeals <- svyby(~CBD160, ~RIAGENDR, design = subset(postStratify(design3,~race+RIAGENDR+agecat,racegendera
ge2),RIDAGEYR>=20) , svymean, na.rm = TRUE)
teachDinMeals
```

```
##   RIAGENDR   CBD160        se
## 1        1 4.652224 0.3210291
## 2        2 4.888584 0.1821728
```

```
dinMeals
```

```
##   RIAGENDR   CBD160         se
## 1        1 4.987265 0.05075695
## 2        2 5.084036 0.05074969
```

```
#HOW OFTEN IS FRUIT AVAILABLE(weighted teachers, weighted pop, unwtd teach ct), N = 275
teachfruitQ <- svyby(~weightedTotal,~CBQ020,design= teachersD3,svytotal,na.rm=TRUE)
teachfruitQ$FREQ <- (teachfruitQ$weightedTotal / sum(teachfruitQ$weightedTotal)) * 100

fruitQ <- svyby(~weightedTotal,~CBQ020,design=subset(postStratify(design3,~race+RIAGENDR+agecat,racegenderage2),RIDAGEYR>=20),svytotal,na.rm=TRUE)####MUST be older than 20
fruitQ$FREQ <- (fruitQ$weightedTotal / sum(fruitQ$weightedTotal)) * 100

svyby(~weightedTotal,~CBQ020,design=subset(postStratify(design3,~race+RIAGENDR+agecat,racegenderage2),RIDAGEYR>=20),unwtd.count,na.rm=TRUE)####MUST be older than 20
```

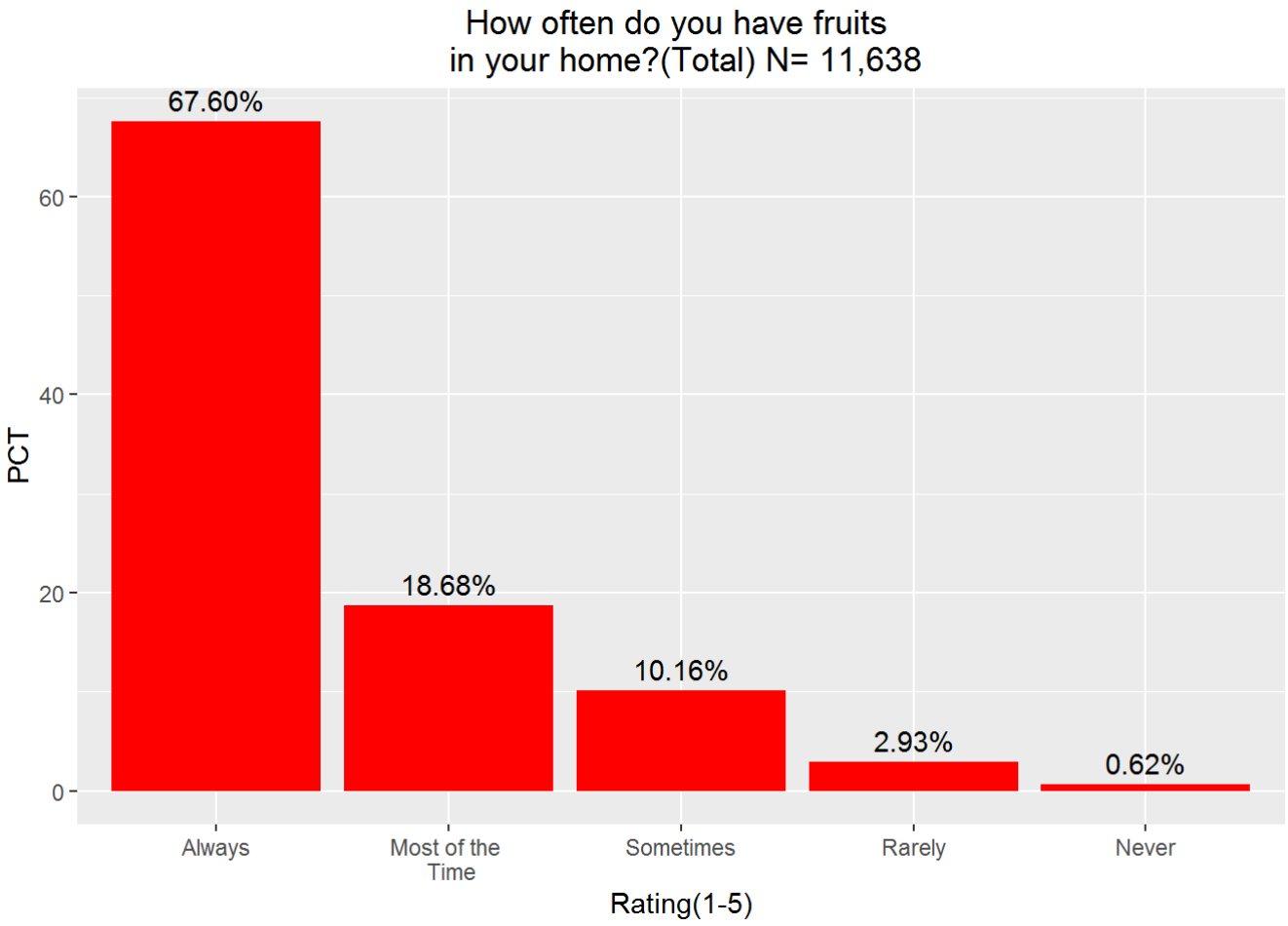
```
##   CBQ020 counts se
## 1      1   7699  0
## 2      2   2230  0
## 3      3   1290  0
## 4      4    329  0
## 5      5     75  0
```

```
svyby(~weightedTotal,~CBQ020,design= teachersD3,unwtd.count,na.rm=TRUE)
```

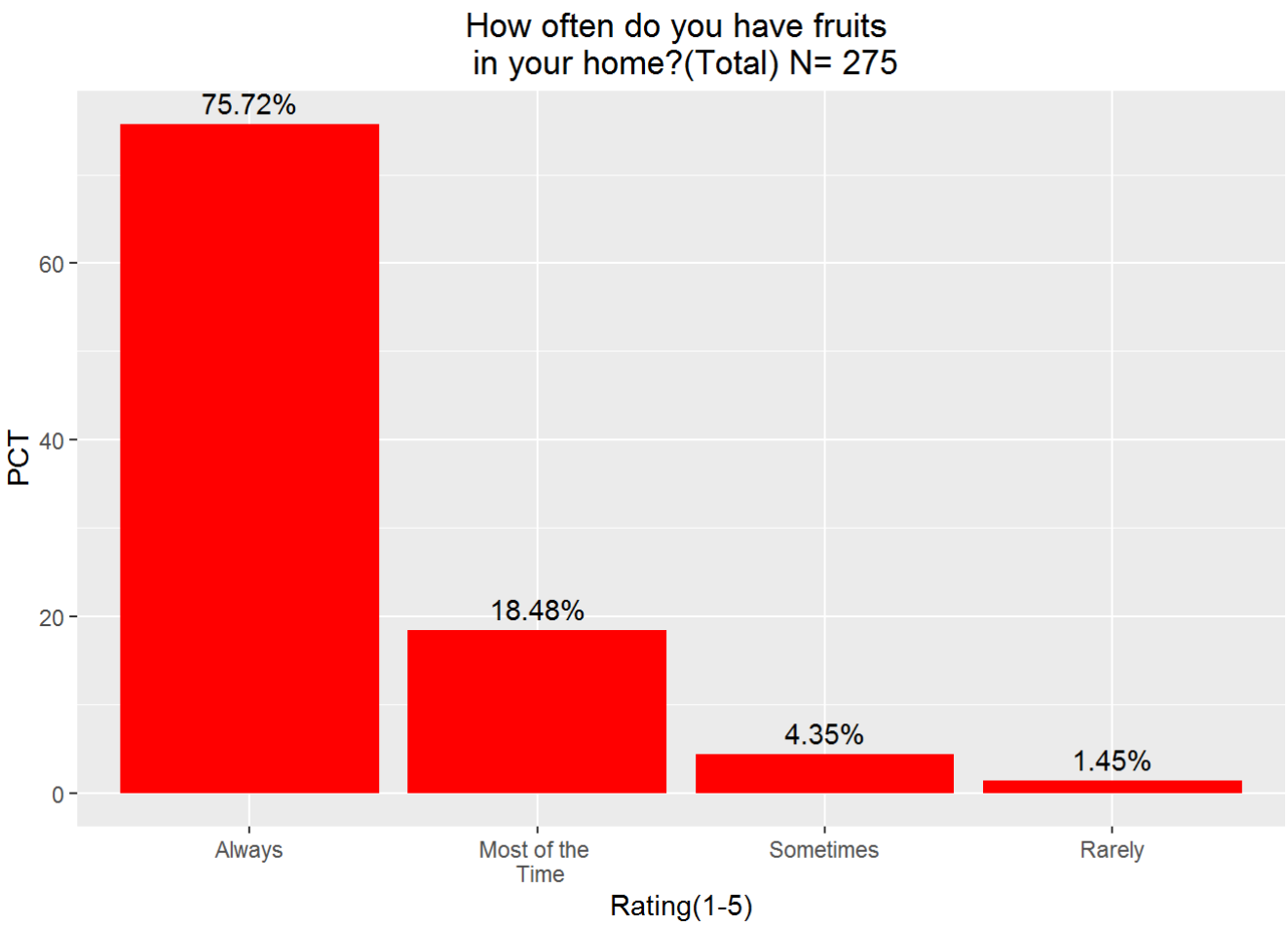
```
##   CBQ020 counts se
## 1      1    210  0
## 2      2     52  0
## 3      3     11  0
## 4      4      2  0
```

```
teachvegQ <- svyby(~weightedTotal,~CBQ030,design= teachersD3,svytotal,na.rm=TRUE)
teachvegQ$FREQ <- (teachvegQ$weightedTotal / sum(teachvegQ$weightedTotal)) * 100

vegQ <-
svyby(~weightedTotal,~CBQ030,design=subset(postStratify(design3,~race+RIAGENDR+agecat,racegenderage2),RIDAGE
YR>=20),svytotal,na.rm=TRUE)####MUST be older than 20
vegQ$FREQ <- (vegQ$weightedTotal / sum(vegQ$weightedTotal)) * 100

teachsaltQ <- svyby(~weightedTotal,~CBQ040,design= teachersD3,svytotal,na.rm=TRUE)
teachsaltQ$FREQ <- (teachsaltQ$weightedTotal / sum(teachsaltQ$weightedTotal)) * 100

saltQ <- svyby(~weightedTotal,~CBQ040,design=subset(postStratify(design3,~race+RIAGENDR+agecat,racegenderage
2),RIDAGEYR>=20),svytotal,na.rm=TRUE)####MUST be older than 20
saltQ$FREQ <- (saltQ$weightedTotal / sum(saltQ$weightedTotal)) * 100

teachsoftQ <- svyby(~weightedTotal,~CBQ060,design= teachersD3,svytotal,na.rm=TRUE)
teachsoftQ$FREQ <- (teachsoftQ$weightedTotal / sum(teachsoftQ$weightedTotal)) * 100

softQ <- svyby(~weightedTotal,~CBQ060,design=subset(postStratify(design3,~race+RIAGENDR+agecat,racegenderage
2),RIDAGEYR>=20),svytotal,na.rm=TRUE)####MUST be older than 20
softQ$FREQ <- (softQ$weightedTotal / sum(softQ$weightedTotal)) * 100

teachDBQ <- svyby(~weightedTotal,~DBQ700,design= teachersD2,svytotal,na.rm=TRUE)
teachDBQ$FREQ <- (teachDBQ$weightedTotal / sum(teachDBQ$weightedTotal)) * 100

DBQ <-
svyby(~weightedTotal,~DBQ700,design=subset(postStratify(design2,~race+RIAGENDR+agecat,racegenderage2),RIDAGE
YR>=20),svytotal,na.rm=TRUE)####MUST be older than 20
DBQ$FREQ <- (DBQ$weightedTotal / sum(DBQ$weightedTotal)) * 100

###N = 250 for gen health
teachGenHealth <- svyby(~weightedTotal,~HSD010,design= teachersD4,svytotal,na.rm=TRUE)
teachGenHealth$FREQ <- (teachGenHealth$weightedTotal / sum(teachGenHealth$weightedTotal)) * 100

genHealth <-svyby(~weightedTotal,~HSD010,design= subset(postStratify(design4,~race+RIAGENDR+agecat,racegende
rage2),RIDAGEYR>=20),svytotal,na.rm=TRUE)
genHealth$FREQ <- (genHealth$weightedTotal / sum(genHealth$weightedTotal)) * 100
```

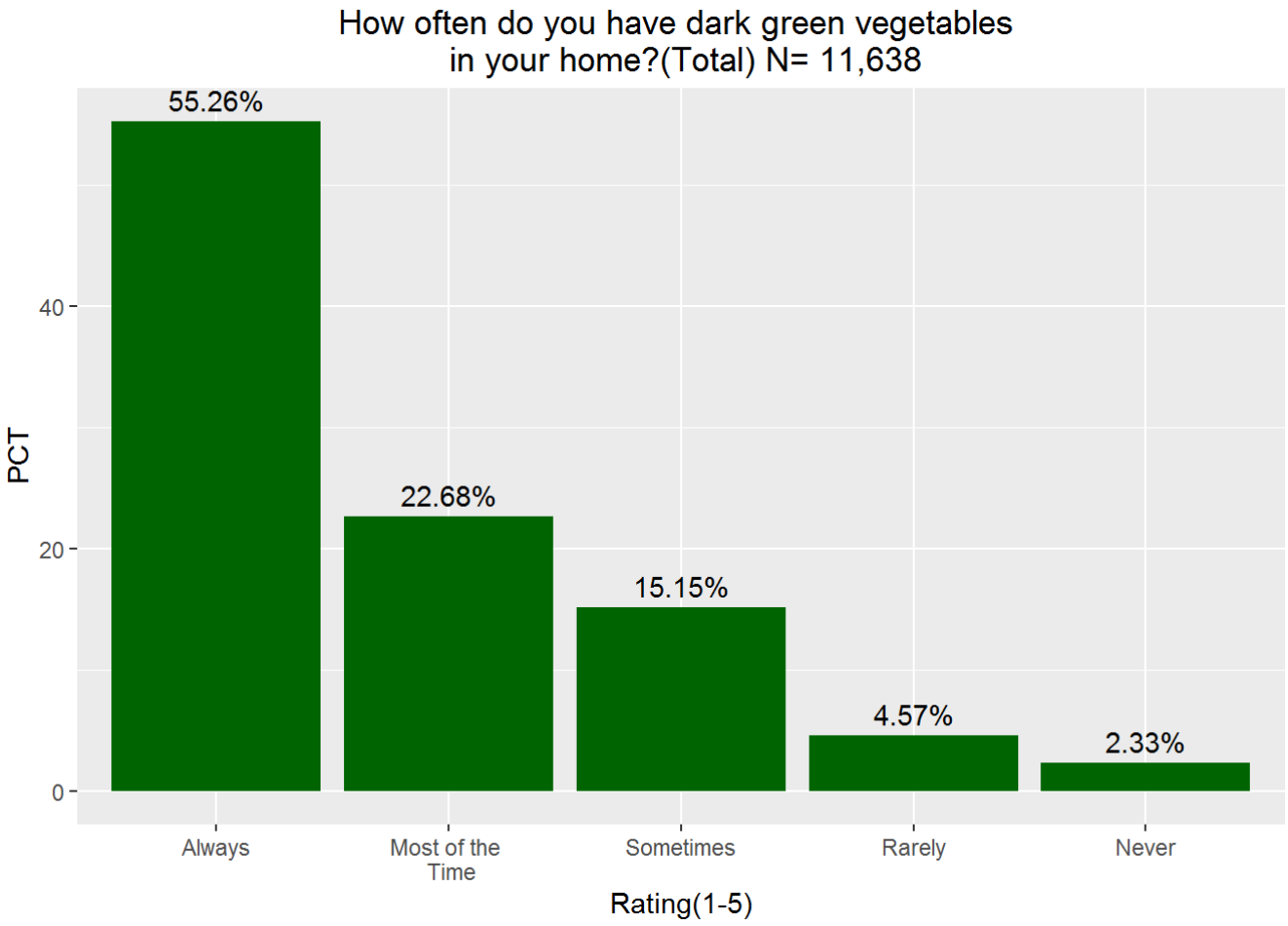
```
fruitQ$CBQ020 <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
positions <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
ggplot(fruitQ, aes(x=CBQ020, y = FREQ)) + geom_bar(stat="identity" , fill = "red")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ), vjust = -.5))+ labs(title = "How often do you have fruits
 \n in your home?(Total) N= 11,638", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits = positions)
```



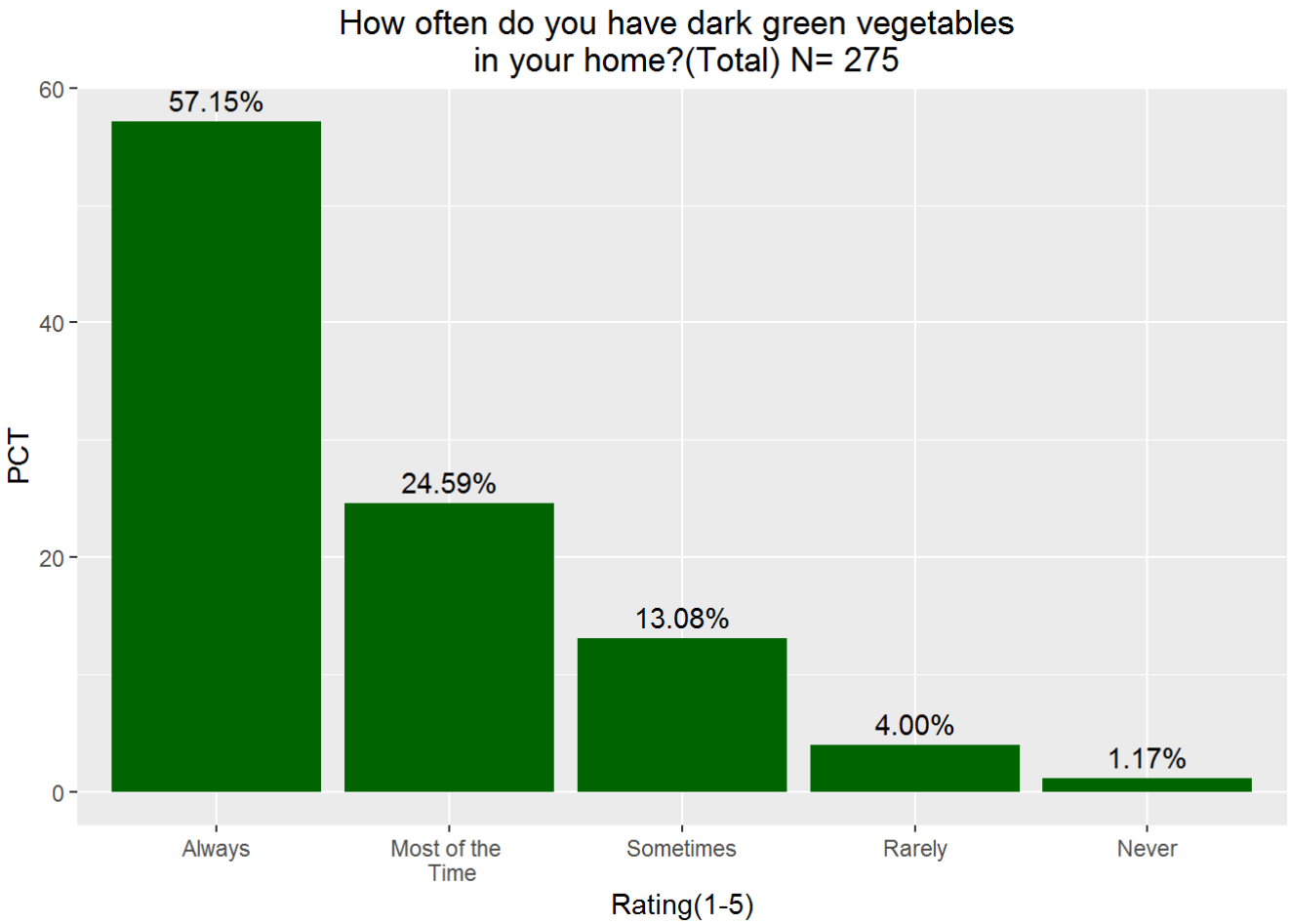

```
teachfruitQ$CBQ020 <- c("Always","Most of the \n Time", "Sometimes", "Rarely")
positions <- c("Always","Most of the \n Time", "Sometimes", "Rarely")
ggplot(teachfruitQ, aes(x=CBQ020, y = FREQ)) + geom_bar(stat="identity" , fill = "red")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ), vjust = -.5))+ labs(title = "How often do you have fruits
 \n in your home?(Total) N= 275", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits = positions)
```



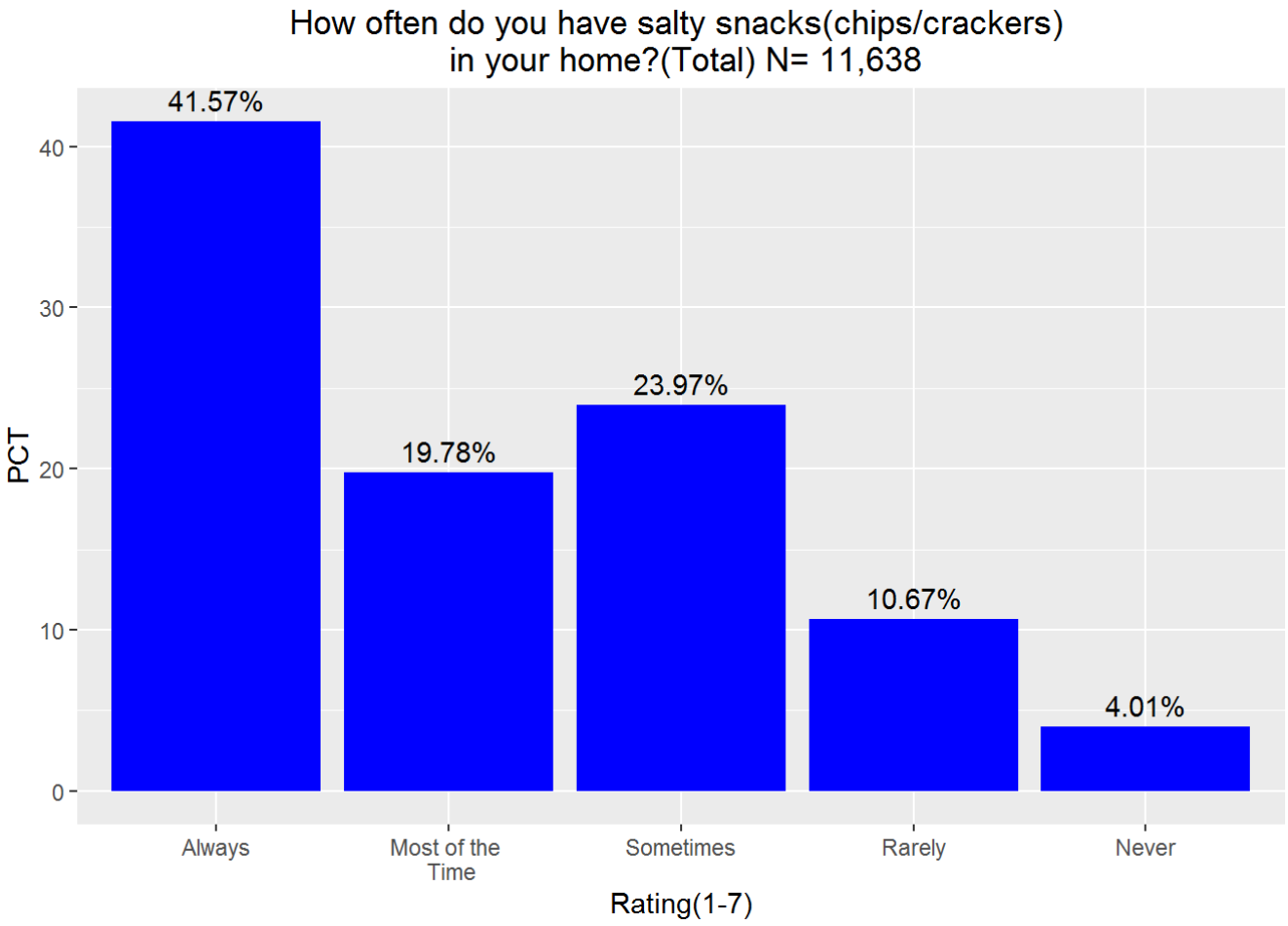

```
vegQ$CBQ030 <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
positions <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
ggplot(vegQ, aes(x=CBQ030, y = FREQ)) + geom_bar(stat="identity" , fill = "dark green")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How often do you have dark gre
en vegetables \n in your home?(Total) N= 11,638", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits = p
ositions)
```



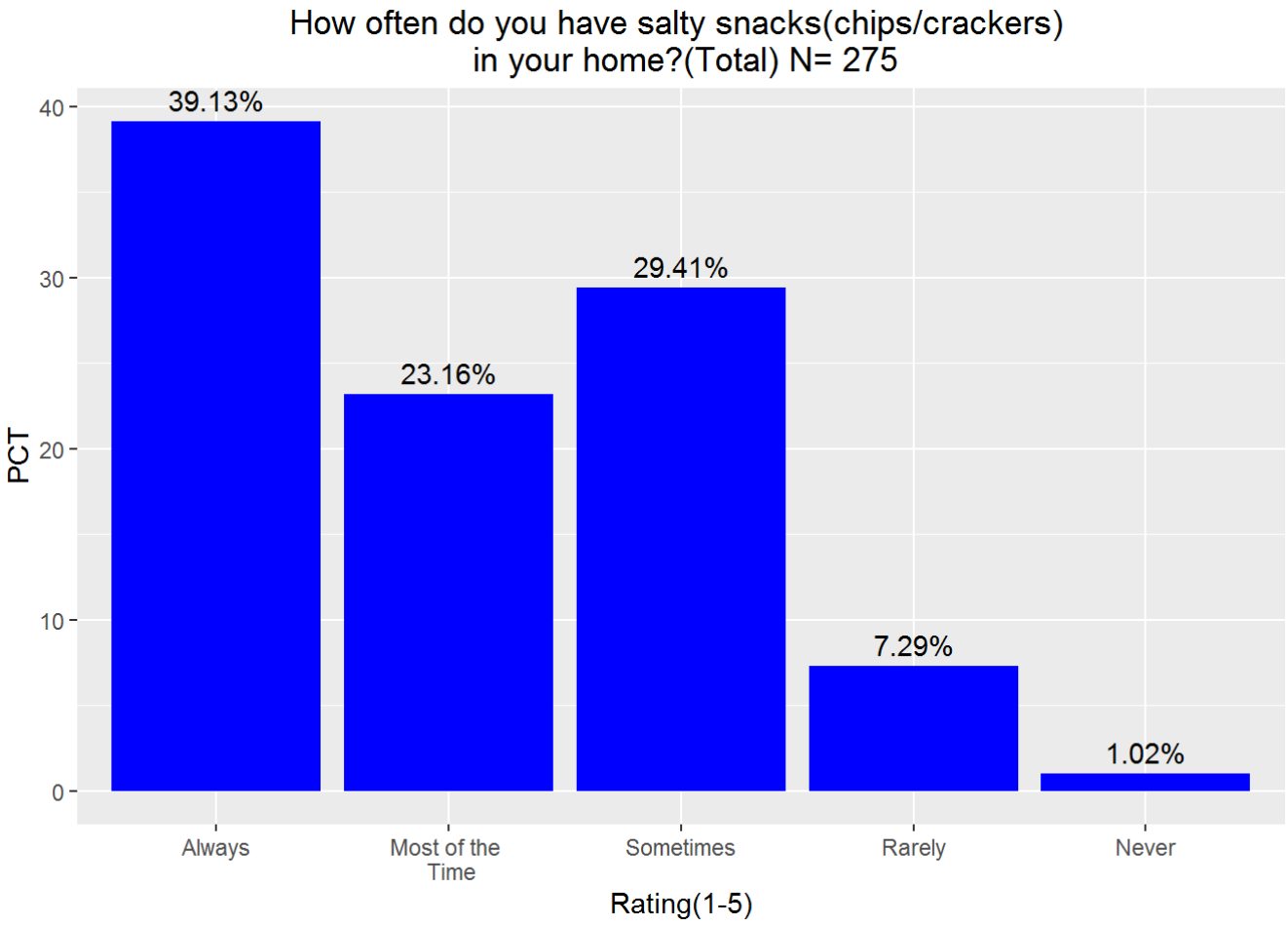

```
teachvegQ$CBQ030 <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
positions <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
ggplot(teachvegQ, aes(x=CBQ030, y = FREQ)) + geom_bar(stat="identity" , fill = "dark green")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How often do you have dark gre
en vegetables \n in your home?(Total) N= 275", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits = posi
tions)
```



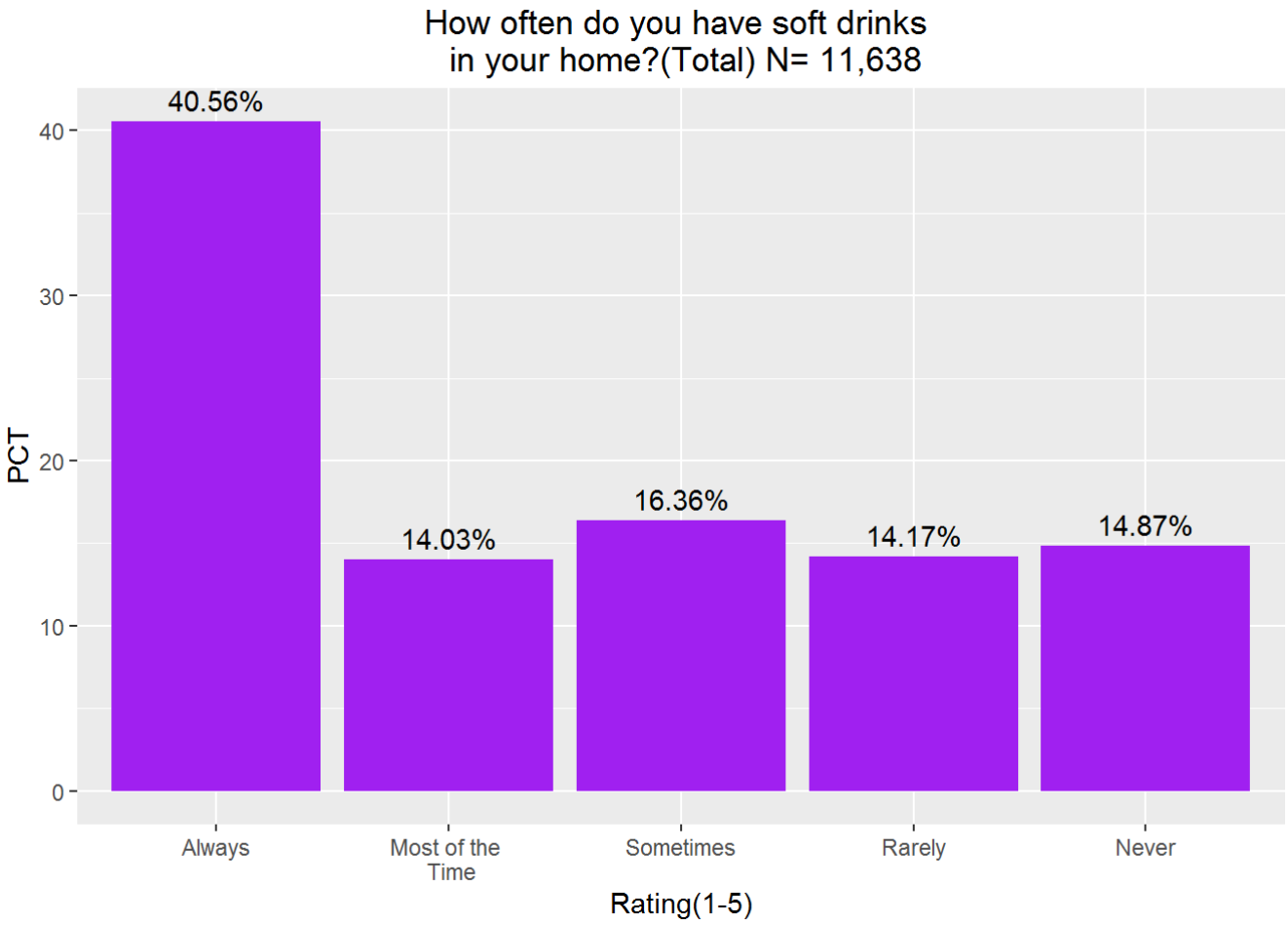

```
saltQ$CBQ040<- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
positions <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
ggplot(saltQ, aes(x=CBQ040, y = FREQ)) + geom_bar(stat="identity" , fill = "blue")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How often do you have salty sn
acks(chips/crackers) \n in your home?(Total) N= 11,638", x = "Rating(1-7)", y = "PCT")+ scale_x_discrete(lim
its = positions)
```



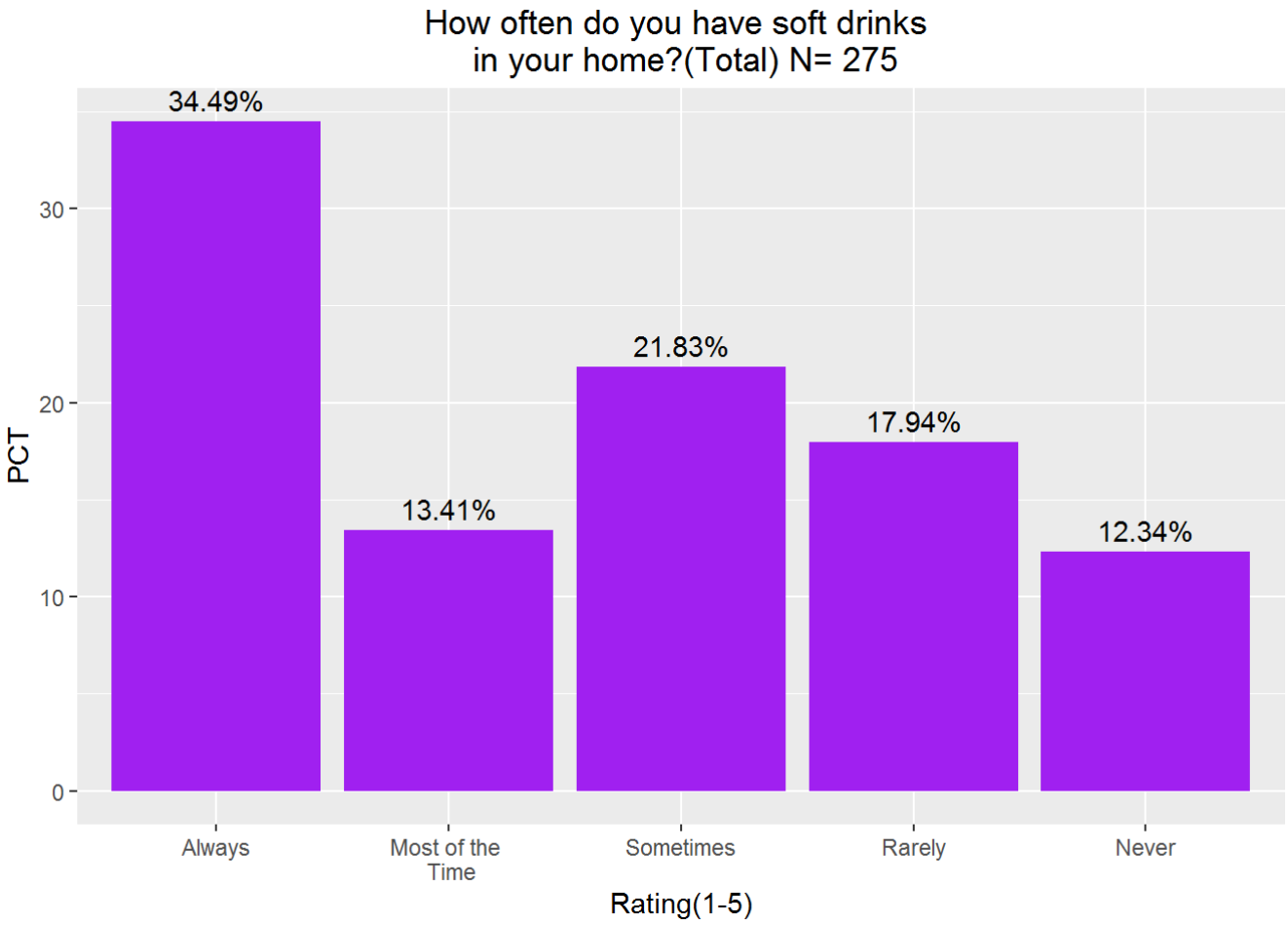

```
teachsaltQ$CBQ040<- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
positions <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
ggplot(teachsaltQ, aes(x=CBQ040, y = FREQ)) + geom_bar(stat="identity" , fill = "blue")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How often do you have salty sn
acks(chips/crackers) \n in your home?(Total) N= 275", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits
 = positions)
```



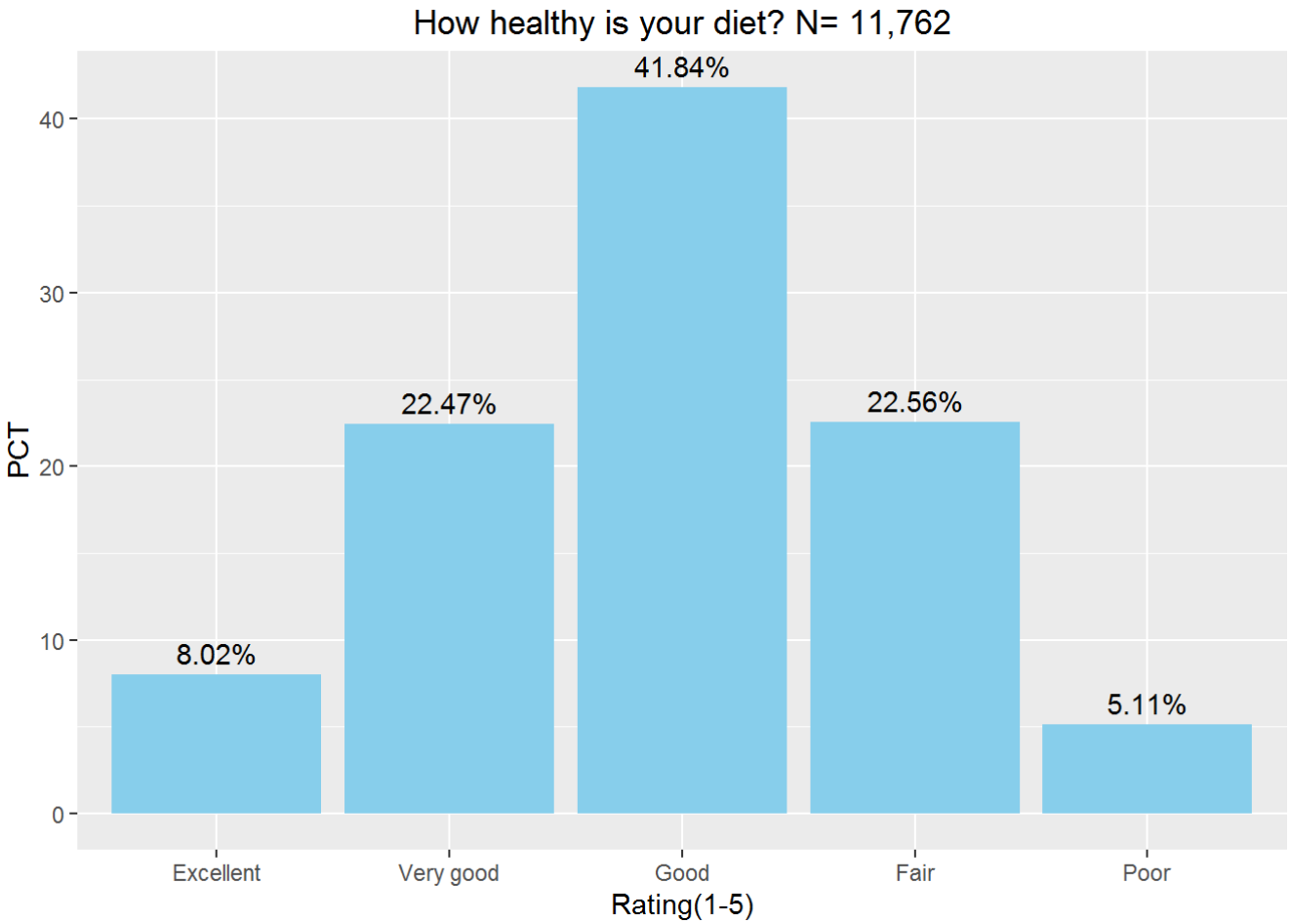

```
softQ$CBQ060<- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
positions <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
ggplot(softQ, aes(x=CBQ060, y = FREQ)) + geom_bar(stat="identity" , fill = "purple")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How often do you have soft dri
nks \n in your home?(Total) N= 11,638", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits = positions)
```



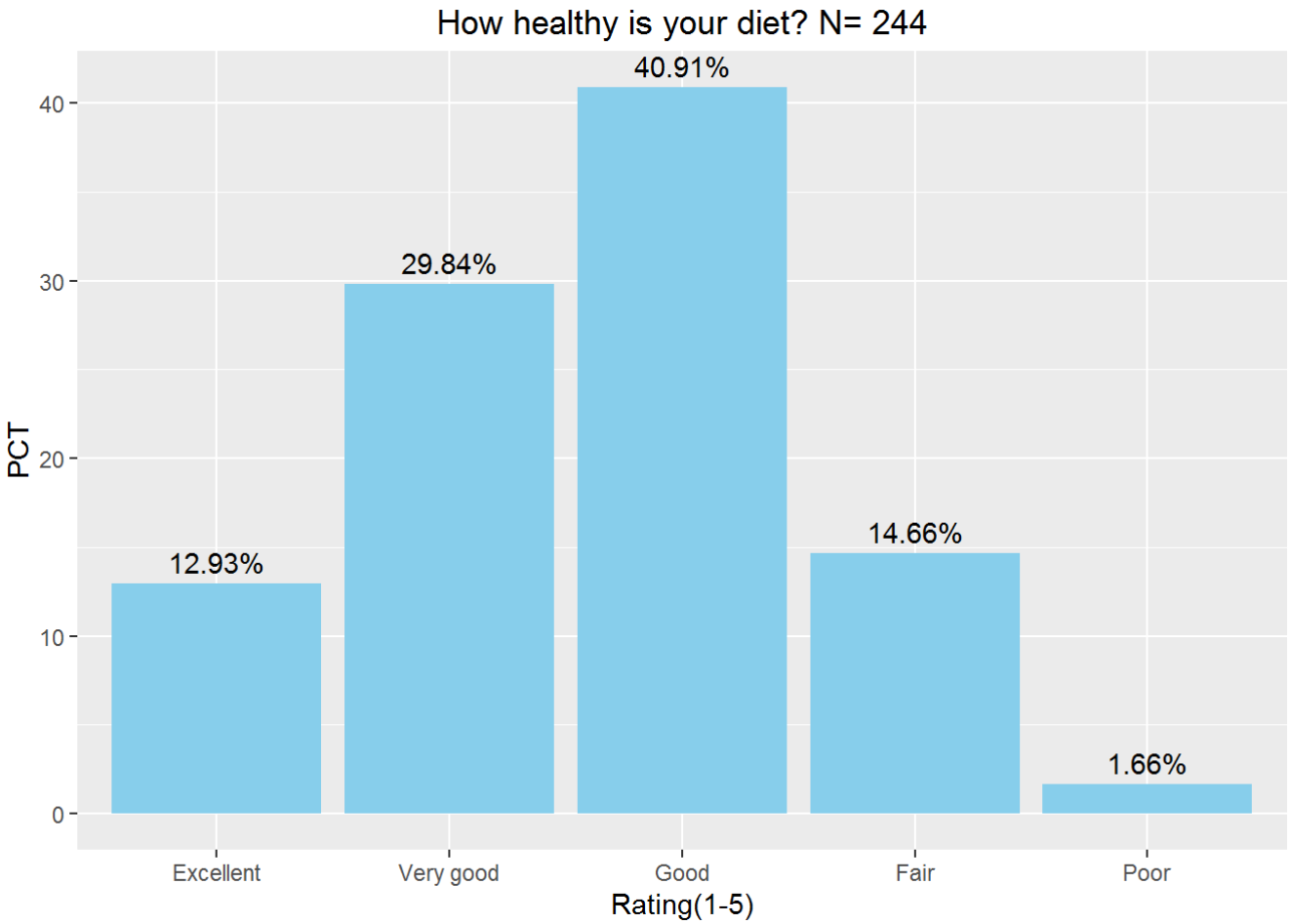

```
teachsoftQ$CBQ060<- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
positions <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
ggplot(teachsoftQ, aes(x=CBQ060, y = FREQ)) + geom_bar(stat="identity" , fill = "purple")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How often do you have soft dri
nks \n in your home?(Total) N= 275", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits = positions)
```



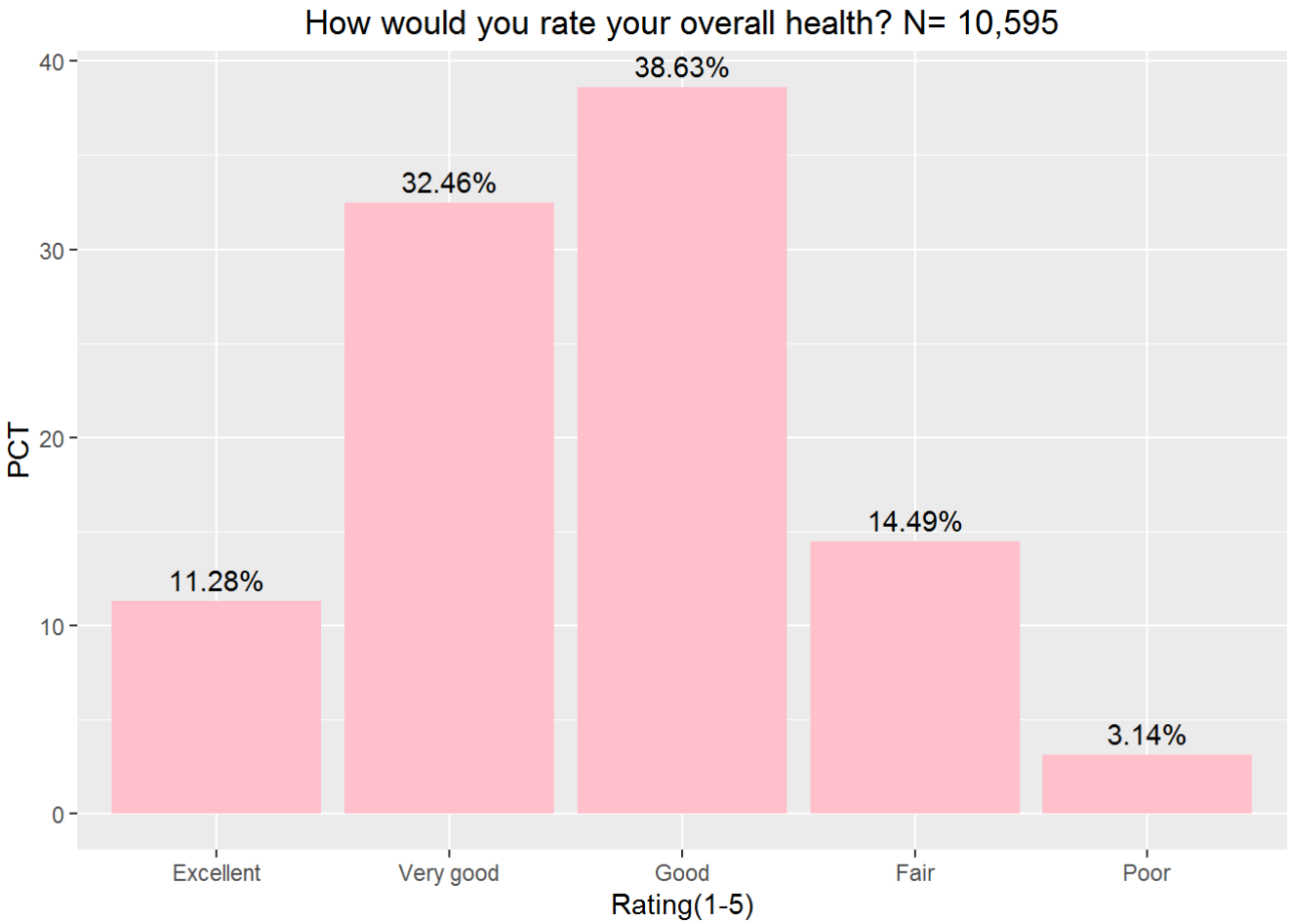

```
DBQ$DBQ700<- c("Excellent","Very good", "Good", "Fair", "Poor")
positions <- c("Excellent","Very good", "Good", "Fair", "Poor")
ggplot(DBQ, aes(x=DBQ700, y = FREQ)) + geom_bar(stat="identity" , fill = "sky blue")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How healthy is your diet? N= 1
1,762", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits = positions)
```



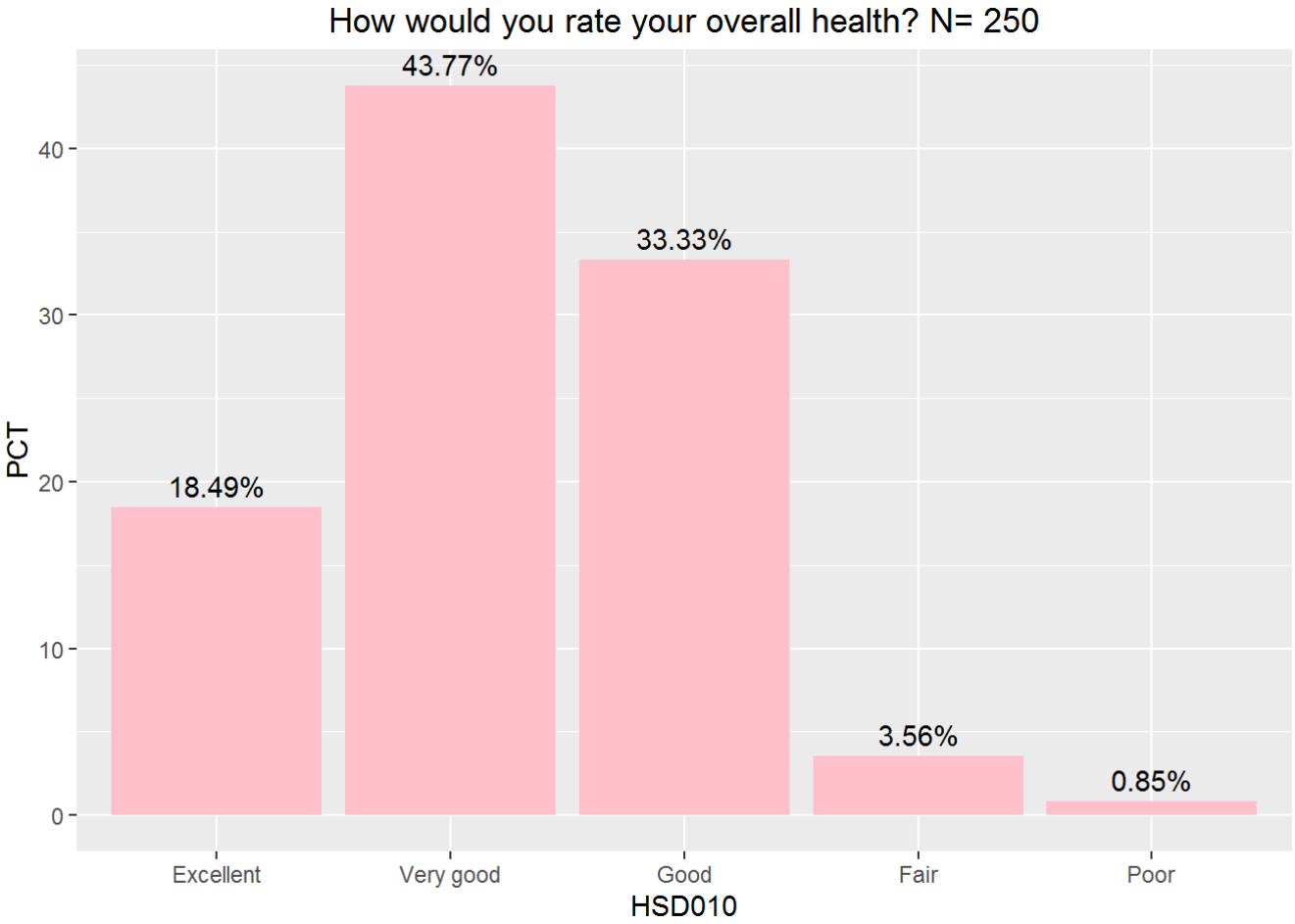

```
teachDBQ$DBQ700<- c("Excellent","Very good", "Good", "Fair", "Poor")
positions <- c("Excellent","Very good", "Good", "Fair", "Poor")
ggplot(teachDBQ, aes(x=DBQ700, y = FREQ)) + geom_bar(stat="identity" , fill = "sky blue")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How healthy is your diet? N= 2
44", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits = positions)
```




```
genHealth$HSD010<- c("Excellent","Very good", "Good", "Fair", "Poor")
positions <- c("Excellent","Very good", "Good", "Fair", "Poor")
ggplot(genHealth, aes(x=HSD010, y = FREQ)) + geom_bar(stat="identity" , fill = "pink")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How would you rate your overal
l health? N= 10,595", x = "Rating(1-5)", y = "PCT")+ scale_x_discrete(limits = positions)
```




```
teachGenHealth$HSD010<- c("Excellent","Very good", "Good", "Fair", "Poor")
positions <- c("Excellent","Very good", "Good", "Fair", "Poor")
ggplot(teachGenHealth, aes(x=HSD010, y = FREQ)) + geom_bar(stat="identity" , fill = "pink")+
  geom_text(aes(label = sprintf("%.2f%%", FREQ)), vjust = -.5)+ labs(title = "How would you rate your overal
l health? N= 250", y = "PCT")+ scale_x_discrete(limits = positions)
```


How would you rate your overall health? N= 250
43.77%
33.33%
18.49%
3.56%
0.85%
40
30
20
10
0
PCT
Excellent
Very good
Good
Fair
Poor
HSD010