# Teacher Descriptive Stats

***Evan Yacek***

***Tuesday, July 12, 2016***

```
#UPLOAD NHANES DATASETS
OCP_E2007 <- nhanes('OCQ_E')
```

```
## Processing SAS dataset OCQ_E      ..
```

```
OCP_F2009 <- nhanes('OCQ_F')
```

```
## Processing SAS dataset OCQ_F      ..
```

```
DEMO2007 <- nhanes('DEMO_E')
```

```
## Processing SAS dataset DEMO_E     ..
```

```
DEMO2009 <- nhanes('DEMO_F')
```

```
## Processing SAS dataset DEMO_F     ..
```

```
Diet2007 <- nhanes('DBQ_E')
```

```
## Processing SAS dataset DBQ_E      ..
```

```
Diet2009 <- nhanes('DBQ_F')
```

```
## Processing SAS dataset DBQ_F      ..
```

```
cons2007 <- nhanes('CBQ_E')
```

```
## Processing SAS dataset CBQ_E      ..
```

```
cons2009 <- nhanes('CBQ_F')
```

```
## Processing SAS dataset CBQ_F      ..
```

```
bodyMeasure2007 <- nhanes('BMX_E')
```

```
## Processing SAS dataset BMX_E      ..
```

```
bodyMeasure2009 <- nhanes('BMX_F')
```

```
## Processing SAS dataset BMX_F      ..
```

```
bp2007 <- nhanes('BPX_E')
```

```
## Processing SAS dataset BPX_E      ..
```

```
bp2009 <- nhanes('BPX_F')
```

```
## Processing SAS dataset BPX_F      ..
```

```
chol2007 <- nhanes('HDL_E')
```

```
## Processing SAS dataset HDL_E      ..
```

```
chol2009 <- nhanes('HDL_F')
```

```
## Processing SAS dataset HDL_F      ..
```

```
tri2007 <- nhanes('TRIGLY_E')
```

```
## Processing SAS dataset TRIGLY_E   ..
```

```
tri2009 <- nhanes('TRIGLY_F')
```

```
## Processing SAS dataset TRIGLY_F   ..
```

```
apob2007 <-nhanes('APOB_E')
```

```
## Processing SAS dataset APOB_E     ..
```

```
apob2009 <-nhanes('APOB_F')
```

```
## Processing SAS dataset APOB_F     ..
```

```
#FILTER BASED ON TEACHER OCCUPATION CODES
TEACHERS2007 <- filter(OCP_E2007, OCD231 == 15 & OCD241 == 8 )
TEACHERS2009 <- filter(OCP_F2009, OCD231 == 15 & OCD241 == 8 )

#MERGE INTO SINGLE TEACHER DATASET
TEACHERDEMO2007 <- merge(TEACHERS2007 , DEMO2007)
td <- merge(TEACHERDEMO2007 , Diet2007)
td <- merge(td , cons2007)
td <- merge(td, bp2007)
td <- merge(td, bodyMeasure2007)
td <- merge(td, chol2007)







TEACHERDEMO2009 <- merge(TEACHERS2009 , DEMO2009)
td2 <- merge(TEACHERDEMO2009 , Diet2009)
td2 <- merge(td2, cons2009)
td2 <- merge(td2, bp2009)
td2 <- merge(td2, bodyMeasure2009)
td2 <- merge(td2, chol2009)









common_cols <- intersect(colnames(td), colnames(td2))
newCombine<-
  rbind(
  subset(td, select = common_cols),
  subset(td2, select = common_cols)
)

#CREATE FREQUENCY TABLE FOR GENDER
MF_Frequency <- table(newCombine$RIAGENDR)
res <- cbind(MF_Frequency,round(prop.table(MF_Frequency)*100, 2))
rownames(res) <- c('Male','Female')
colnames(res) <- c("Freq", "Pct")
res
```

```
##        Freq   Pct
## Male     79 28.32
## Female  200 71.68
```

```
#INTERVALED AGE DISTRIBUTION FOR TEACHERS

x <- newCombine$RIDAGEYR
factorx <- factor(cut(x, breaks=8))
#Tabulate and turn into data.frame
xout <- as.data.frame(table(factorx))
#Add cumFreq and proportions
xout <- transform(xout, cumFreq = cumsum(Freq), relative = prop.table(Freq)*100)
colnames(xout) <- c("Age-Range" , "Freq" , "Total", "PCT")
xout
```

```
##   Age-Range Freq Total        PCT
## 1 (15.9,24]   12    12  4.3010753
## 2   (24,32]   51    63 18.2795699
## 3   (32,40]   51   114 18.2795699
## 4   (40,48]   50   164 17.9211470
## 5   (48,56]   49   213 17.5627240
## 6   (56,64]   47   260 16.8458781
## 7   (64,72]   17   277  6.0931900
## 8 (72,80.1]    2   279  0.7168459
```

```
ggplot(newCombine, aes(x = RIDAGEYR, fill=..count.. )) + geom_histogram(bins = 8, col="black" ) +
labs(title="Distribution for Age of Teachers") + labs(x="Age", y="Count") + scale_fill_gradient("Count", low
 = "green", high = "red")
```

```
## Don't know how to automatically pick scale for object of type labelled/integer. Defaulting to continuous
```


```
#Ethnicity Distribution
Ethnic_Frequency <- table(newCombine$RIDRETH1)
rest <- cbind(Ethnic_Frequency,round(prop.table(Ethnic_Frequency)*100, 2))
colnames(rest) <- c("Freq", "Pct")
rownames(rest) <- c("Mexican American", "Other Hispanic", "Non-Hispanic White", "Non-Hispanic Black", "Othe
r")
rest
```

```
##                   Freq   Pct
## Mexican American    43 15.41
## Other Hispanic      34 12.19
## Non-Hispanic White 148 53.05
## Non-Hispanic Black  38 13.62
## Other               16  5.73
```

```
#Calculate mean Waist Cirucumference
#Seperate Men and Women
maleTeachers <- filter(newCombine,RIAGENDR == 1 )
femTeachers <- filter(newCombine,RIAGENDR == 2 )
maleTeacherWaistMean <- mean(maleTeachers$BMXWAIST, na.rm = TRUE)

#Average
maleTeacherWaistMean
```

```
## [1] 100.6077
```

```
femTeacherWaistMean <- mean(femTeachers$BMXWAIST, na.rm = TRUE)
femTeacherWaistMean
```

```
## [1] 93.79581
```

```
totalPop <- rbind(DEMO2007, DEMO2009)
totalPop2 <- rbind(bodyMeasure2007, bodyMeasure2009)
totalPopset <- merge(totalPop, totalPop2)
totalPopMale <- filter(totalPopset, RIAGENDR == 1)
totalPopFemale <- filter(totalPopset, RIAGENDR == 2)
femPopMeanWaist <- mean(totalPopFemale$BMXWAIST, na.rm = TRUE)
malePopMeanWaist <- mean(totalPopMale$BMXWAIST, na.rm = TRUE)
femPopMeanWaist
```

```
## [1] 86.76449
```

```
malePopMeanWaist
```

```
## [1] 88.55798
```

Compare Triglcerides

```
teacherTri2007 <- filter(OCP_E2007, OCD231 == 15 & OCD241 == 8 )
teacherTri2007 <- merge(teacherTri2007, tri2007 )
teacherTri2007 <- merge(teacherTri2007, DEMO2007 )
teacherTri2009 <- filter(OCP_F2009, OCD231 == 15 & OCD241 == 8 )
teacherTri2009 <- merge(teacherTri2009, tri2009 )
teacherTri2009 <- merge(teacherTri2009, DEMO2009 )

common_cols2 <- intersect(colnames(teacherTri2007), colnames(teacherTri2009))
newCombine2<-
  rbind(
  subset(teacherTri2007, select = common_cols2),
  subset(teacherTri2009, select = common_cols2)
)


maleTeachTri <- filter(newCombine2 ,RIAGENDR == 1 )
maleTeachTriMean <- mean(maleTeachTri$LBXTR, na.rm = TRUE)
maleTeachTriMean
```

```
## [1] 135.9487
```

```
femaleTeachTri <- filter(newCombine2 ,RIAGENDR == 2 )
femaleTeachTriMean <- mean(femaleTeachTri$LBXTR, na.rm = TRUE)
femaleTeachTriMean
```

```
## [1] 93.34524
```

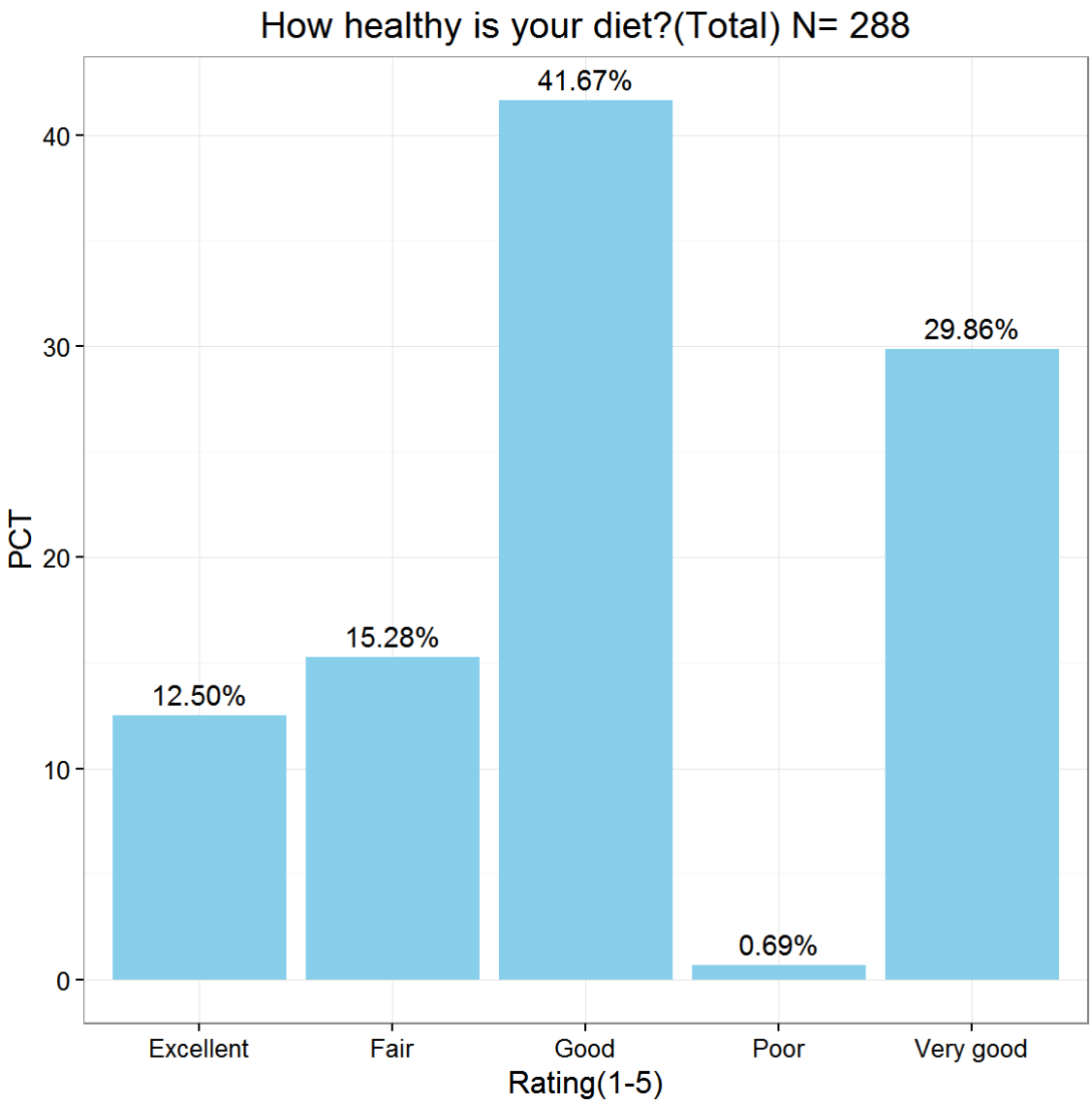
```
totalPop3 <- rbind(DEMO2007, DEMO2009)
totalPop4 <- rbind(tri2007, tri2009)
totalPopset2 <- merge(totalPop3, totalPop4)
totalPopMale2 <- filter(totalPopset2, RIAGENDR == 1)
totalPopFemale2 <- filter(totalPopset2, RIAGENDR == 2)

femPopMeanTri <- mean(totalPopFemale2$LBXTR, na.rm = TRUE)
femPopMeanTri
```

```
## [1] 118.8999
```

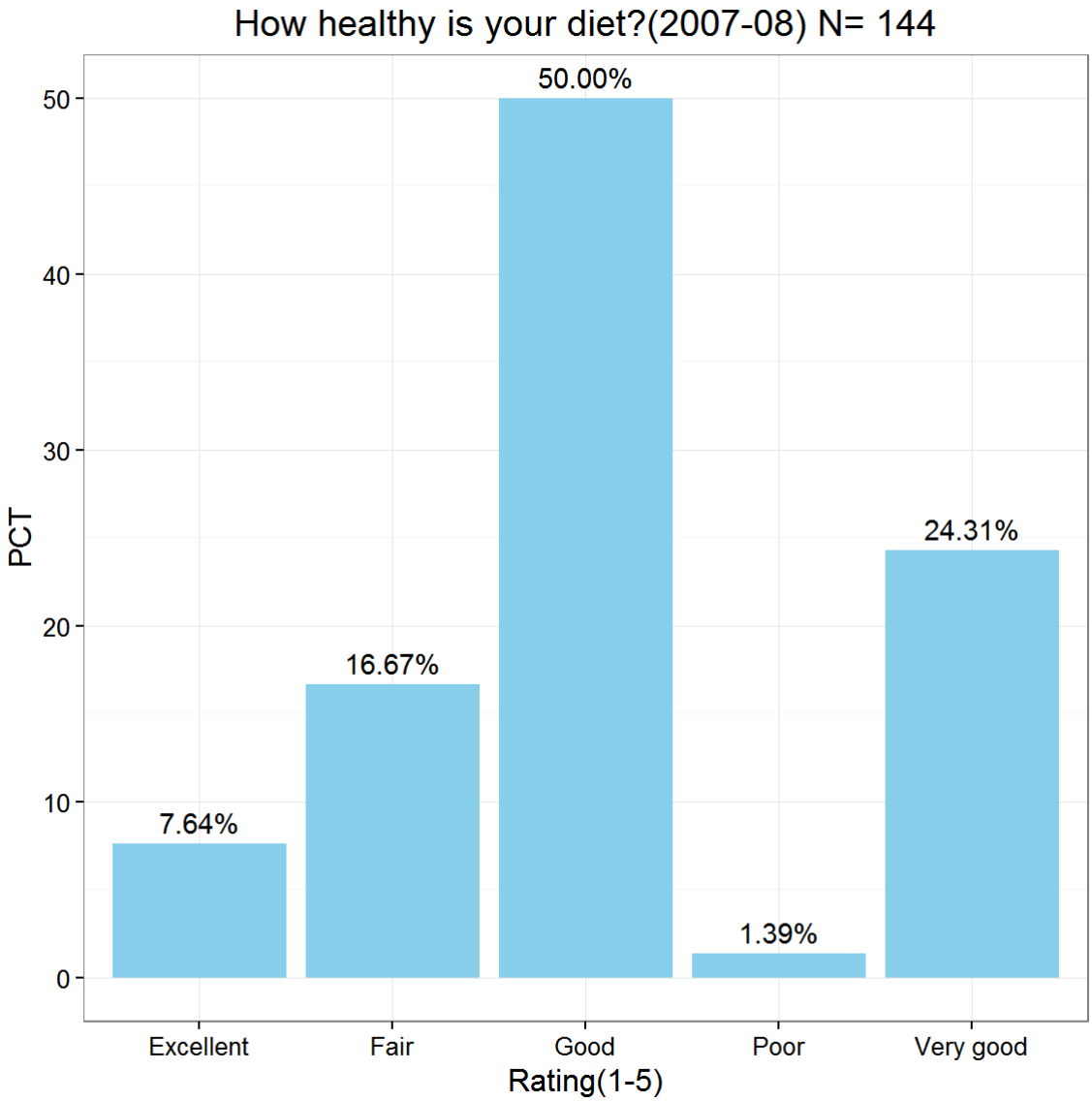
```
malePopMeanTri <- mean(totalPopMale2$LBXTR, na.rm = TRUE)
malePopMeanTri
```

```
## [1] 135.7783
```

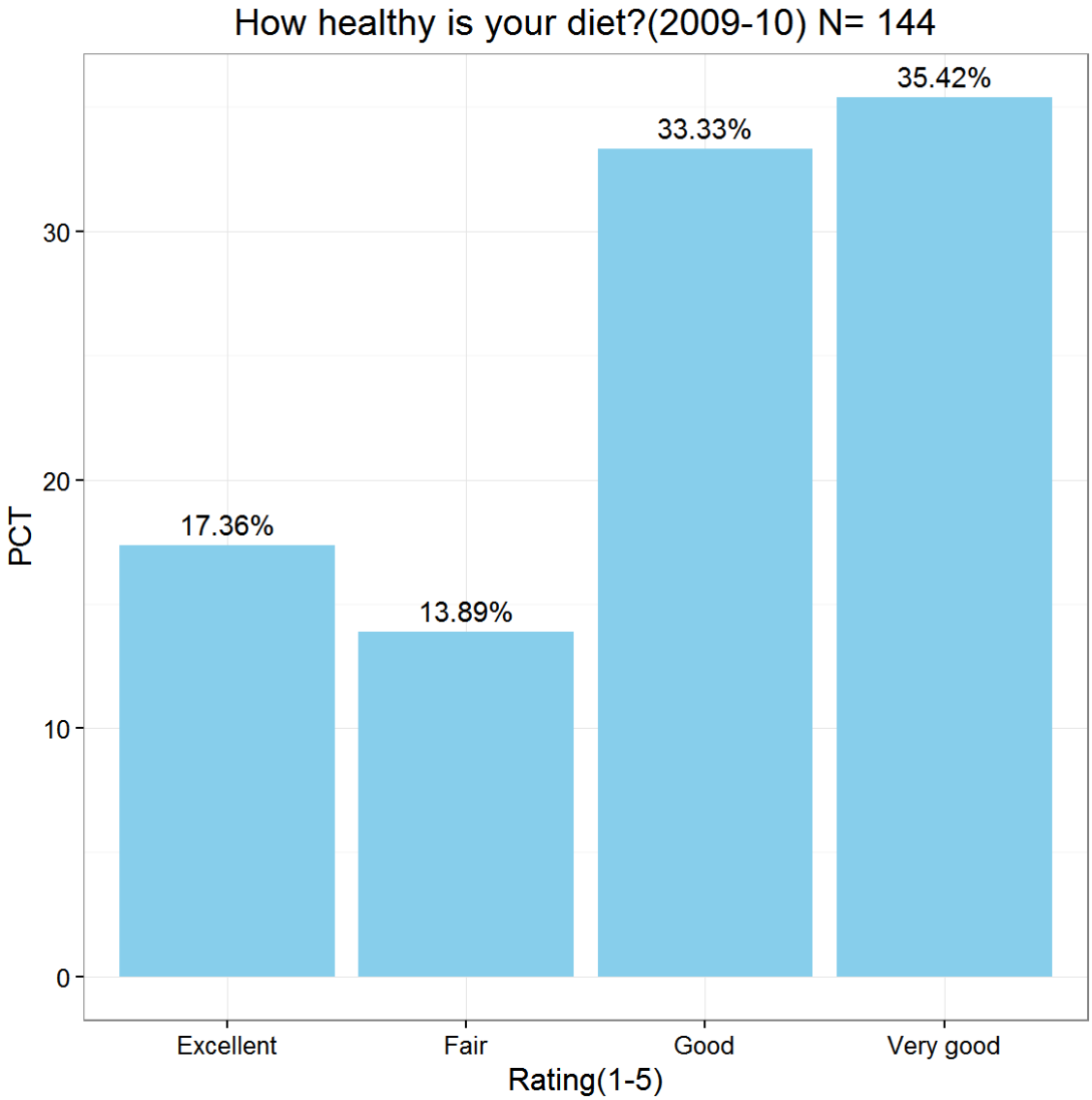
```
TEACHERS2007 <- merge(TEACHERS2007, Diet2007)
TEACHERS2009 <- merge(TEACHERS2009, Diet2009)
commoncols <- intersect(colnames(TEACHERS2007), colnames(TEACHERS2009))
teacherCombine<-
  rbind(
    subset(TEACHERS2007, select = commoncols),
    subset(TEACHERS2009, select = commoncols)
  )

healthyDietQ <- as.data.frame(table(teacherCombine$DBQ700))
healthyDietQ$Var1 <- c("Excellent","Very good", "Good", "Fair", "Poor")
ggplot(healthyDietQ, aes(x=Var1, y = Freq*100/288)) + geom_bar(stat="identity" , fill = "sky blue")+
  geom_text(aes(label = sprintf("%.2f%%", Freq/sum(Freq) * 100)), vjust = -.5)+ labs(title = "How healthy is
 your diet?(Total) N= 288", x = "Rating(1-5)", y = "PCT")
```



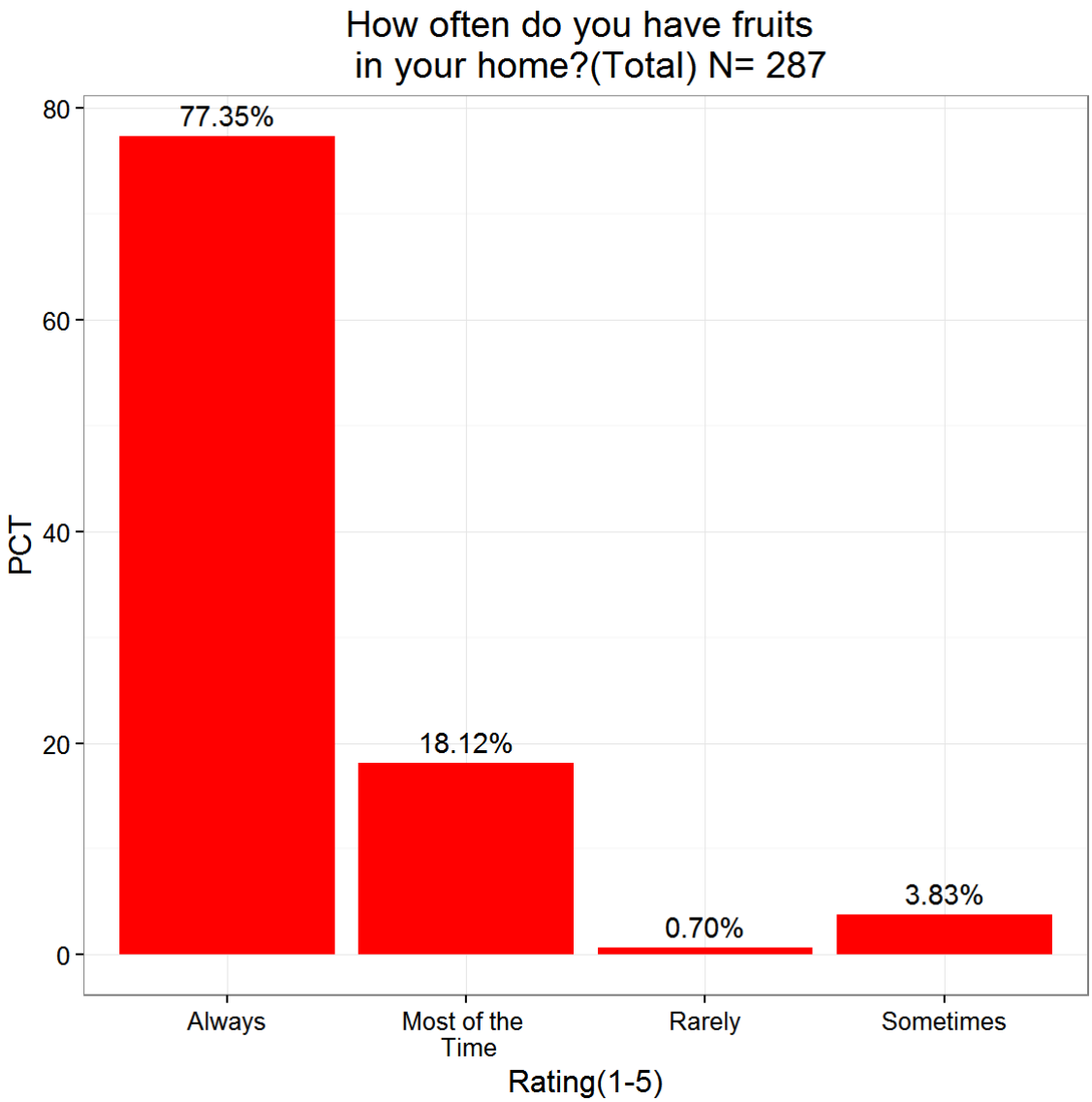

```
healthyDietQ2007 <- as.data.frame(table(TEACHERS2007$DBQ700))
healthyDietQ2007$Var1 <- c("Excellent","Very good", "Good", "Fair", "Poor")
ggplot(healthyDietQ2007, aes(x=Var1, y = Freq*100/144)) + geom_bar(stat="identity" , fill = "sky blue")+
  geom_text(aes(label = sprintf("%.2f%%", Freq/sum(Freq) * 100)), vjust = -.5)+ labs(title = "How healthy is
 your diet?(2007-08) N= 144", x = "Rating(1-5)", y = "PCT")
```



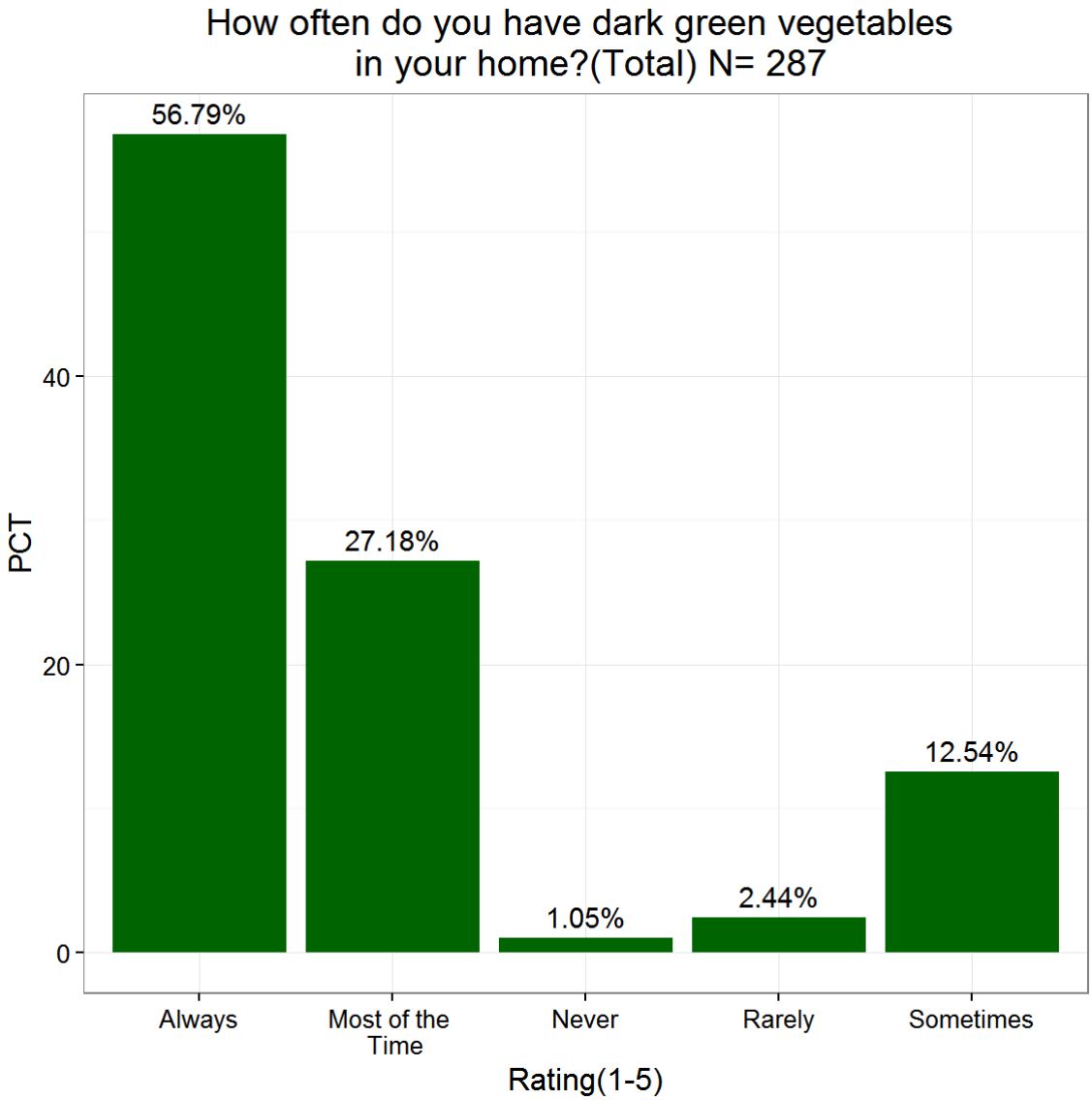

```
healthyDietQ2009 <- as.data.frame(table(TEACHERS2009$DBQ700))
healthyDietQ2009$Var1 <- c("Excellent","Very good", "Good", "Fair" )
ggplot(healthyDietQ2009, aes(x=Var1, y = Freq*100/144)) + geom_bar(stat="identity" , fill = "sky blue")+
  geom_text(aes(label = sprintf("%.2f%%", Freq/sum(Freq) * 100)), vjust = -.5)+ labs(title = "How healthy is
 your diet?(2009-10) N= 144", x = "Rating(1-5)", y = "PCT")
```



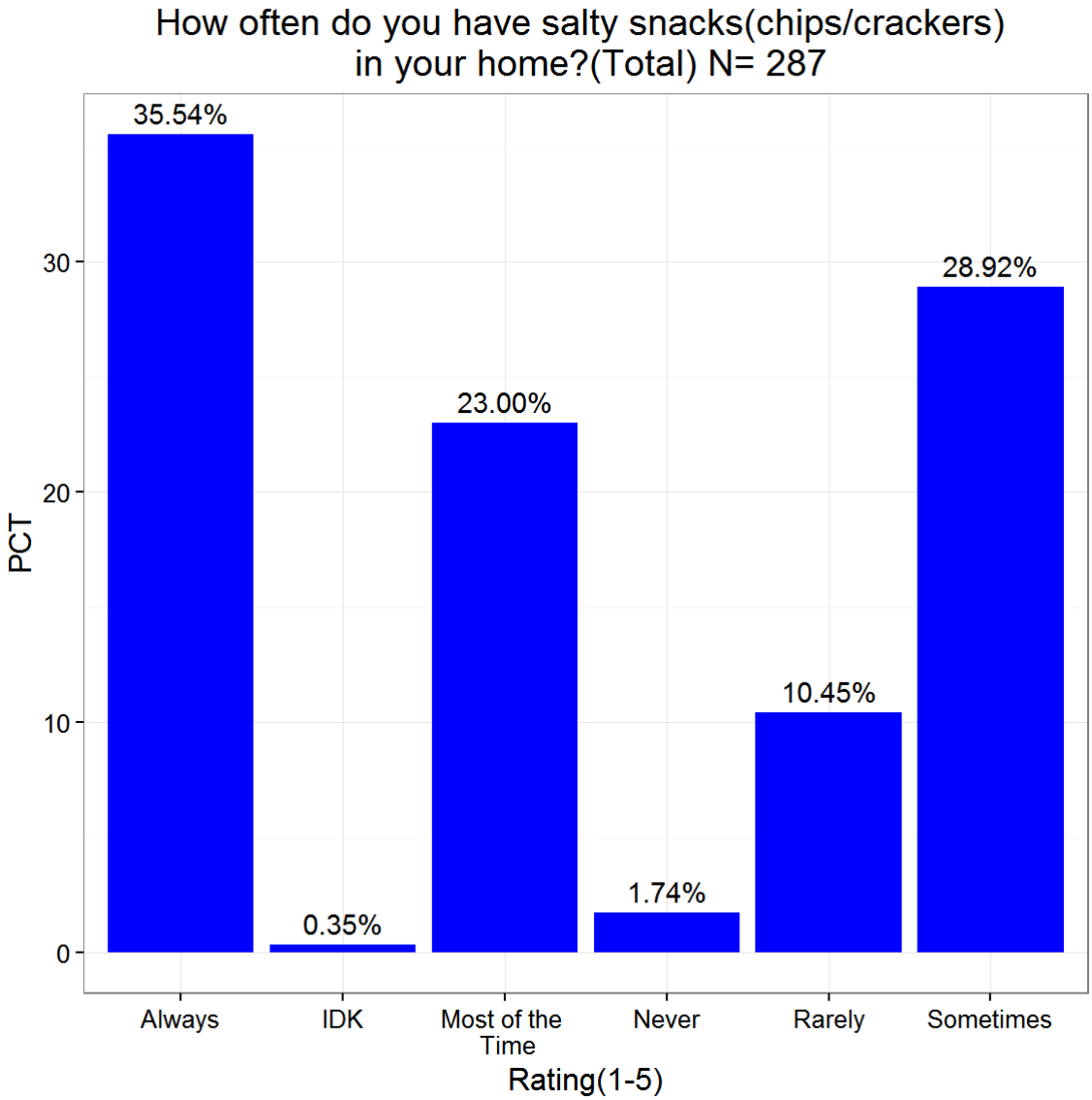

```
TEACHERS2007 <- merge(TEACHERS2007, cons2007)
TEACHERS2009 <- merge(TEACHERS2009, cons2009)
commoncols <- intersect(colnames(TEACHERS2007), colnames(TEACHERS2009))
teacherCombine<-
  rbind(
    subset(TEACHERS2007, select = commoncols),
    subset(TEACHERS2009, select = commoncols)
  )

fruitsQ <- as.data.frame(table(teacherCombine$CBQ020))
fruitsQ$Var1 <- c("Always","Most of the \n Time", "Sometimes", "Rarely")
ggplot(fruitsQ, aes(x=Var1, y = Freq*100/287)) + geom_bar(stat="identity" , fill = "red")+
  geom_text(aes(label = sprintf("%.2f%%", Freq/sum(Freq) * 100)), vjust = -.5)+ labs(title = "How often do y
ou have fruits \n in your home?(Total) N= 287", x = "Rating(1-5)", y = "PCT")
```



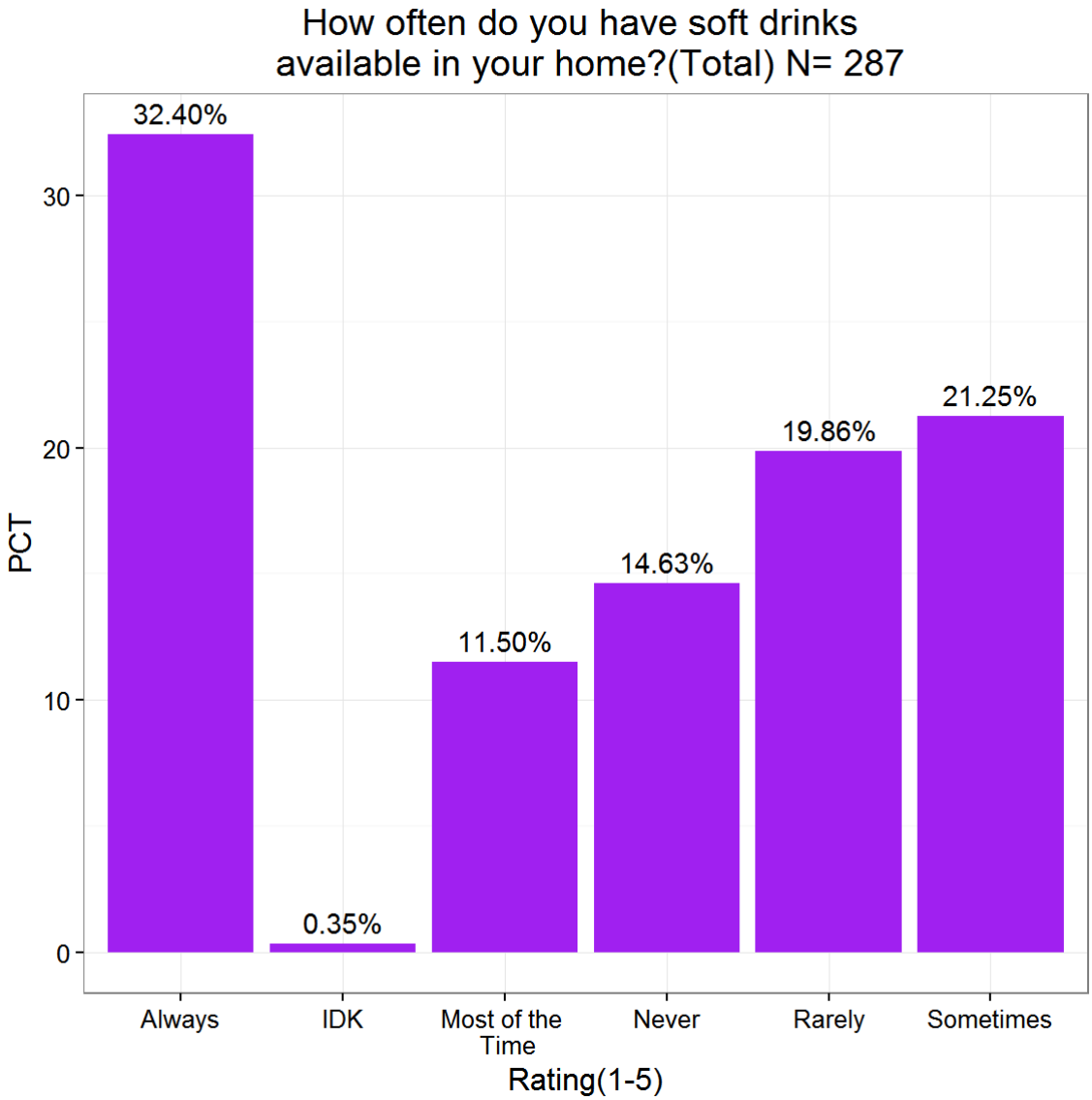

```
vegQ <- as.data.frame(table(teacherCombine$CBQ030))
vegQ$Var1 <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never")
ggplot(vegQ, aes(x=Var1, y = Freq*100/287)) + geom_bar(stat="identity" , fill = "dark green")+
  geom_text(aes(label = sprintf("%.2f%%", Freq/sum(Freq) * 100)), vjust = -.5)+ labs(title = "How often do y
ou have dark green vegetables \n in your home?(Total) N= 287", x = "Rating(1-5)", y = "PCT")
```




```
saltQ <- as.data.frame(table(teacherCombine$CBQ040))
saltQ$Var1 <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never", "IDK")
ggplot(saltQ, aes(x=Var1, y = Freq*100/287)) + geom_bar(stat="identity" , fill = "blue")+
  geom_text(aes(label = sprintf("%.2f%%", Freq/sum(Freq) * 100)), vjust = -.5)+ labs(title = "How often do y
ou have salty snacks(chips/crackers) \n in your home?(Total) N= 287", x = "Rating(1-5)", y = "PCT")
```




```
softQ <- as.data.frame(table(teacherCombine$CBQ060))
softQ$Var1 <- c("Always","Most of the \n Time", "Sometimes", "Rarely" , "Never", "IDK")
ggplot(softQ, aes(x=Var1, y = Freq*100/287)) + geom_bar(stat="identity" , fill = "purple")+
  geom_text(aes(label = sprintf("%.2f%%", Freq/sum(Freq) * 100)), vjust = -.5)+ labs(title = "How often do y
ou have soft drinks \n available in your home?(Total) N= 287", x = "Rating(1-5)", y = "PCT")
```


How often do you have soft drinks available in your home?(Total) N= 287
32.40%
0.35%
11.50%
14.63%
19.86%
21.25%
30
20
10
0
PCT
Always
IDK
Most of the Time
Never
Rarely
Sometimes
Rating(1-5)