Classwork

Darya Nemirich

10 May 2019

```r
load("D:/Bioinformatics and System Biology/2nd term/R/R_classwork/Case/CHIS2009_reduced_2.Rdata")
```

```r
head(adult)
```

```
##   RBMI BMI_P RACEHPR2 SRSEX SRAGE_P MARIT2 AB1 ASTCUR AB51 POVLL
## 1    3 28.89        6     1      32      1   1      2   -1     4
## 2    3 26.15        6     2      80      3   1      2   -1     4
## 3    3 25.06        6     1      71      1   2      1   -1     4
## 4    2 24.99        6     1      39      4   1      2   -1     4
## 5    3 25.09        6     1      75      1   2      2   -1     4
## 6    4 32.21        6     2      53      1   3      1   -1     4
```

```r
str(adult)
```

```
## 'data.frame':    44346 obs. of  10 variables:
##  $ RBMI    : num  3 3 3 2 3 4 3 2 3 3 ...
##  $ BMI_P   : num  28.9 26.1 25.1 25 25.1 ...
##  $ RACEHPR2: num  6 6 6 6 6 6 6 6 6 6 ...
##  $ SRSEX   : num  1 2 1 1 1 2 1 2 1 2 ...
##  $ SRAGE_P : num  32 80 71 39 75 53 42 33 67 52 ...
##  $ MARIT2  : num  1 3 1 4 1 1 1 1 3 3 ...
##  $ AB1     : num  1 1 2 1 2 3 2 2 1 5 ...
##  $ ASTCUR  : num  2 2 1 2 2 1 2 2 2 2 ...
##  $ AB51    : num  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
##  $ POVLL   : num  4 4 4 4 4 4 4 3 4 4 ...
```

```r
adult$RBMI <- as.factor(adult$RBMI)
```

```r
str(adult)
```

```
## 'data.frame':    44346 obs. of  10 variables:
##  $ RBMI    : Factor w/ 4 levels "1","2","3","4": 3 3 3 2 3 4 3 2 3 3 ...
##  $ BMI_P   : num  28.9 26.1 25.1 25 25.1 ...
##  $ RACEHPR2: num  6 6 6 6 6 6 6 6 6 6 ...
##  $ SRSEX   : num  1 2 1 1 1 2 1 2 1 2 ...
##  $ SRAGE_P : num  32 80 71 39 75 53 42 33 67 52 ...
##  $ MARIT2  : num  1 3 1 4 1 1 1 1 3 3 ...
##  $ AB1     : num  1 1 2 1 2 3 2 2 1 5 ...
##  $ ASTCUR  : num  2 2 1 2 2 1 2 2 2 2 ...
##  $ AB51    : num  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
##  $ POVLL   : num  4 4 4 4 4 4 4 3 4 4 ...
```

```r
library(ggplot2)
```

```r
plot1 <- ggplot(adult, aes(x = SRAGE_P, fill = RBMI)) +
  geom_histogram(binwidth = 1) +
  xlab("age")

plot1
```
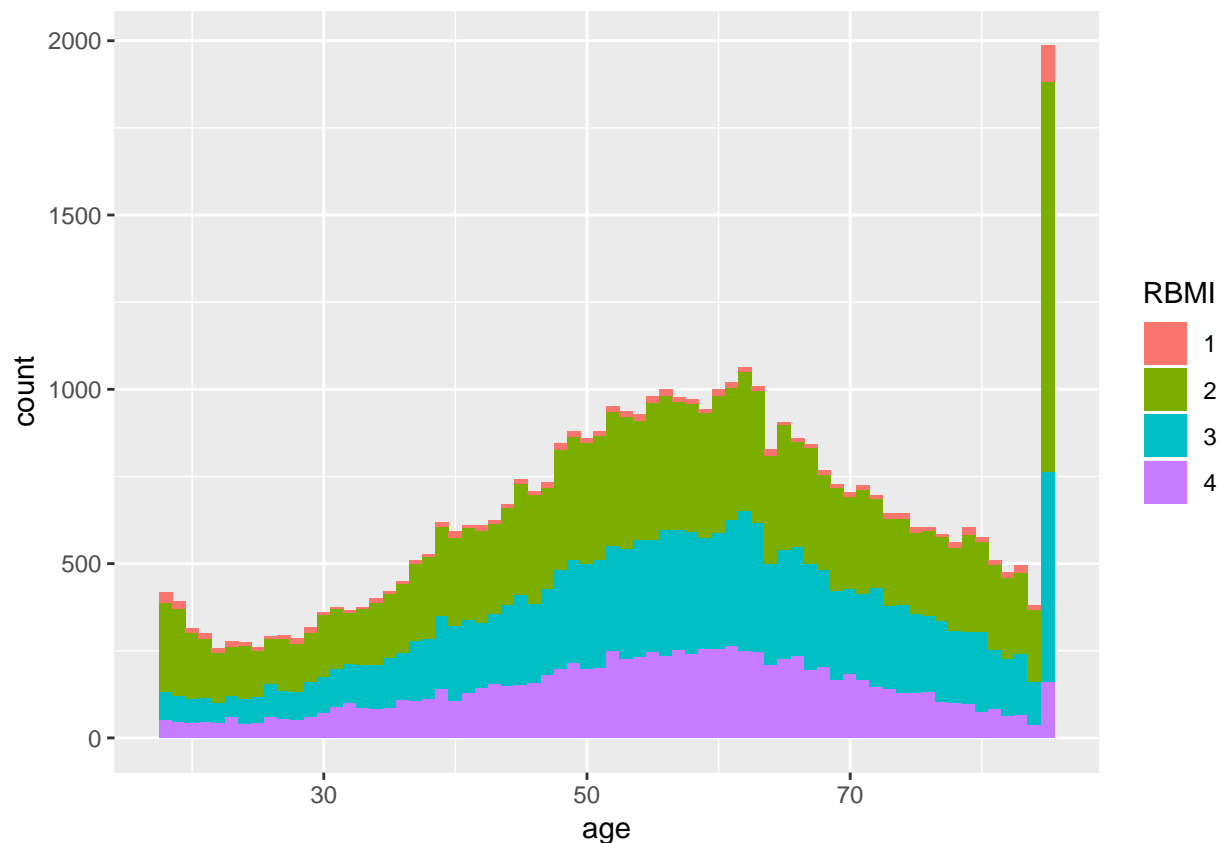
```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
adult_tidy <- adult %>%
  filter(SRAGE_P < 85) %>%
  filter(BMI_P < 52 & BMI_P >= 16) %>%
  mutate(RACEHPR2 = factor(RACEHPR2,
                    levels = c(1, 4, 5, 6),
                    labels = c("Latino", "Asian", "African American", "White"))) %>%
  mutate(RBMI = factor(RBMI,
                    levels = c(1, 2, 3, 4),
                    labels = c("Under", "Normal", "Over", "Obese")))
```

```r
str(adult_tidy)
```

```
## 'data.frame':    42167 obs. of  10 variables:
##  $ RBMI    : Factor w/ 4 levels "Under","Normal",..: 3 3 3 2 3 4 3 2 3 3 ...
```

```
## $ BMI_P   : num  28.9 26.1 25.1 25 25.1 ...
## $ RACEHPR2: Factor w/ 4 levels "Latino","Asian",..: 4 4 4 4 4 4 4 4 4 4 ...
## $ SRSEX   : num  1 2 1 1 1 2 1 2 1 2 ...
## $ SRAGE_P : num  32 80 71 39 75 53 42 33 67 52 ...
## $ MARIT2  : num  1 3 1 4 1 1 1 1 3 3 ...
## $ AB1     : num  1 1 2 1 2 3 2 2 1 5 ...
## $ ASTCUR  : num  2 2 1 2 2 1 2 2 2 2 ...
## $ AB51    : num  -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 ...
## $ POVLL   : num  4 4 4 4 4 4 4 3 4 4 ...
```

```r
plot2 <- ggplot(adult_tidy, aes(x = SRAGE_P, fill = RBMI)) +
  geom_histogram(binwidth = 1, position = "stack") +
  xlab("age") +
  facet_grid(RBMI ~ .) +
  scale_fill_brewer("BMI group", palette = "Reds") +
  theme_classic() +
  theme(strip.text.y = element_blank())
```
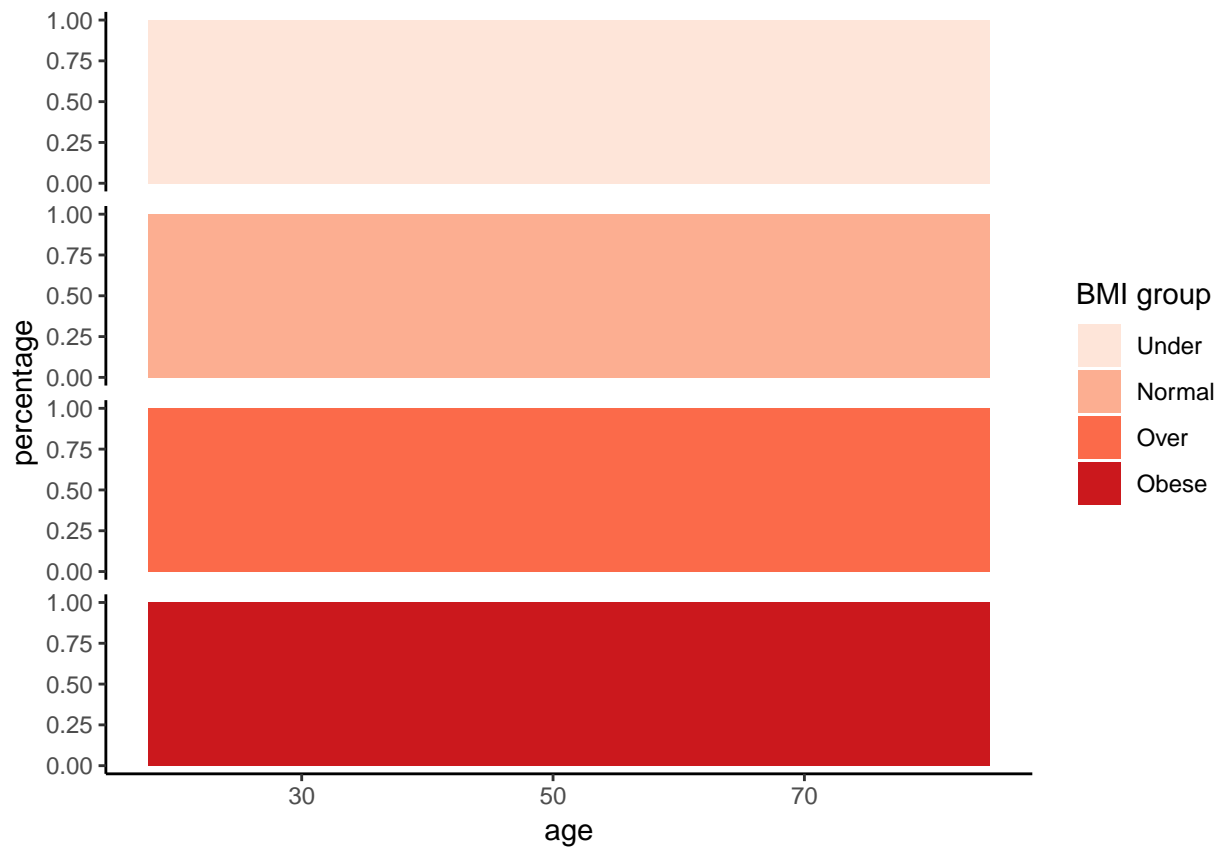
```r
plot2
```
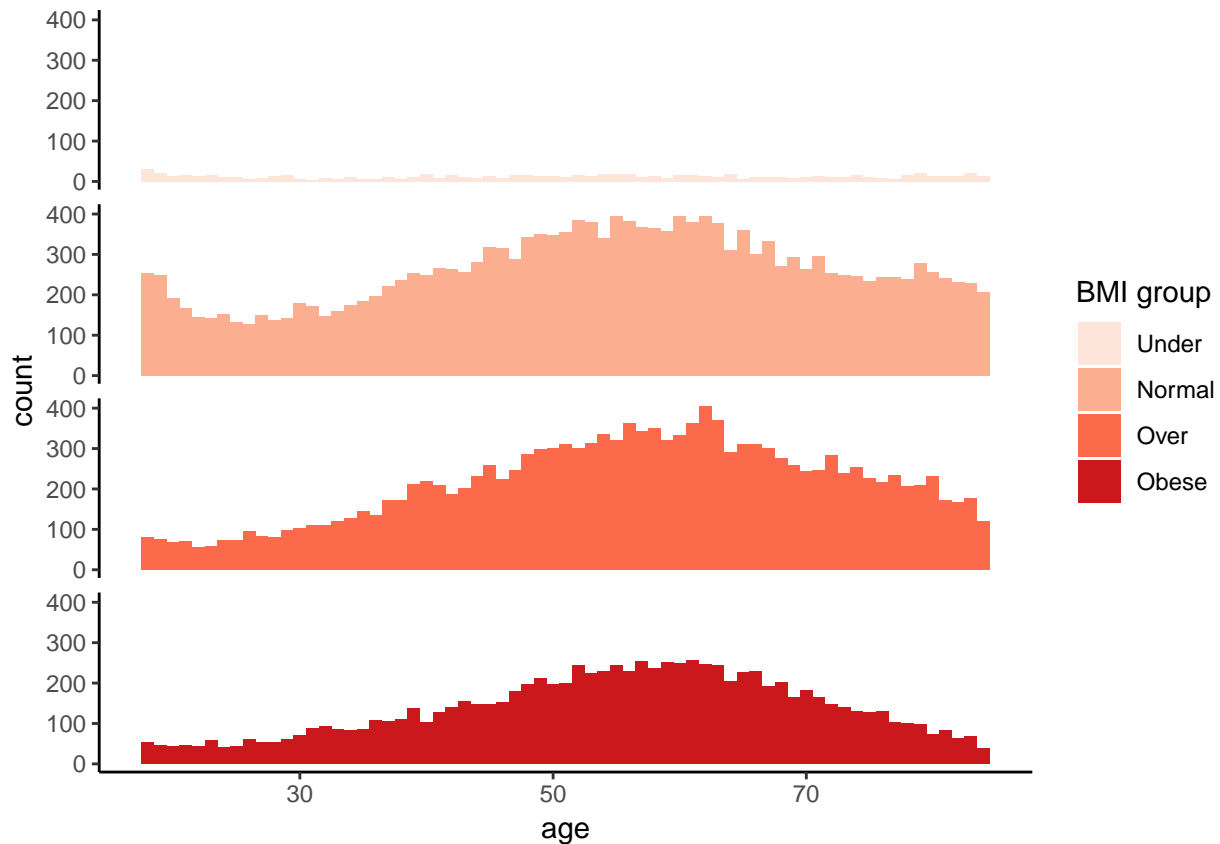


```r
plot3 <- ggplot(adult_tidy, aes(x = SRAGE_P, fill = RBMI)) +
  geom_histogram(position = "fill", bins = 66) +
  xlab("age") +
  ylab("percentage") +
  facet_grid(RBMI ~ .) +
  scale_fill_brewer("BMI group", palette = "Reds") +
  theme_classic() +
```

```
    theme(strip.text.y = element_blank())
```

plot3


```
library(reshape2)
```

```
frequencies_table <- dcast(adult_tidy, SRAGE_P ~ RBMI, length)
```

```
## Using POVLL as value column: use value.var to override.
```
```
head(frequencies_table)
```

```
##   SRAGE_P Under Normal Over Obese
## 1      18    30    254   80    52
## 2      19    22    248   76    45
## 3      20    14    191   68    43
## 4      21    15    168   70    45
## 5      22    13    145   56    44
## 6      23    15    142   59    58
```

```
melted_freq <- reshape2::melt(frequencies_table, id.vars="SRAGE_P", measured.vars=c("Under", "Normal",
```

```
head(melted_freq)
```

```
##   SRAGE_P variable value
## 1      18    Under    30
## 2      19    Under    22
## 3      20    Under    14
```

```
## 4        21     Under        15
## 5        22     Under        13
## 6        23     Under        15
```

```r
plot4 <- ggplot(melted_freq, aes(x = SRAGE_P, y = value, fill = variable)) +
  geom_col() +
  xlab('age') +
  ylab('count') +
  facet_grid(. ~ variable) +
  scale_fill_brewer("BMI group", palette = "Reds") +
  theme_classic() +
  theme(strip.text.y = element_blank())

plot4
```