HM08_final

Sedreh

5/16/2019

```
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
```

```
library(tidyr)
library(ggplot2)
```

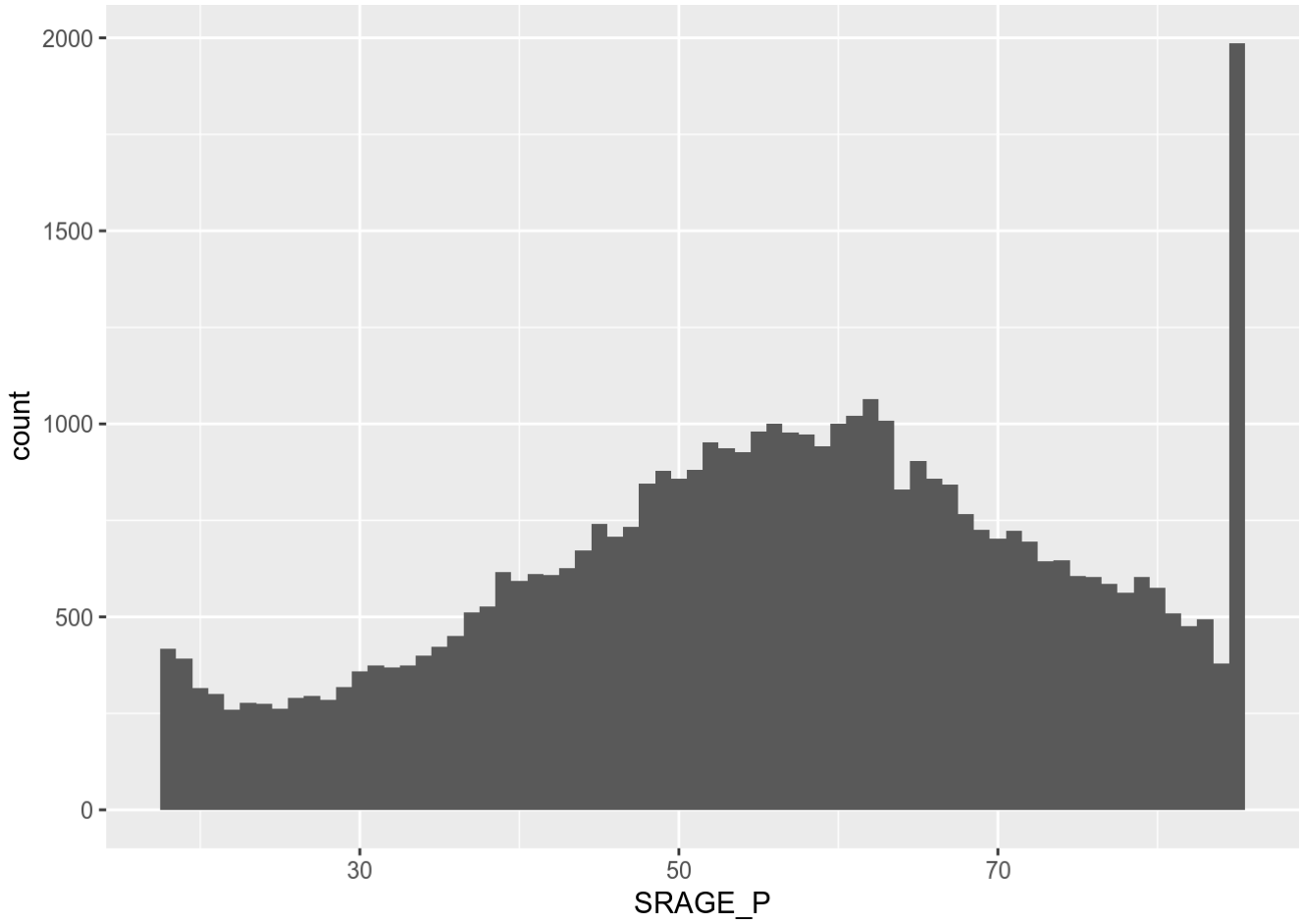
```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

```
data("chis2009")
```

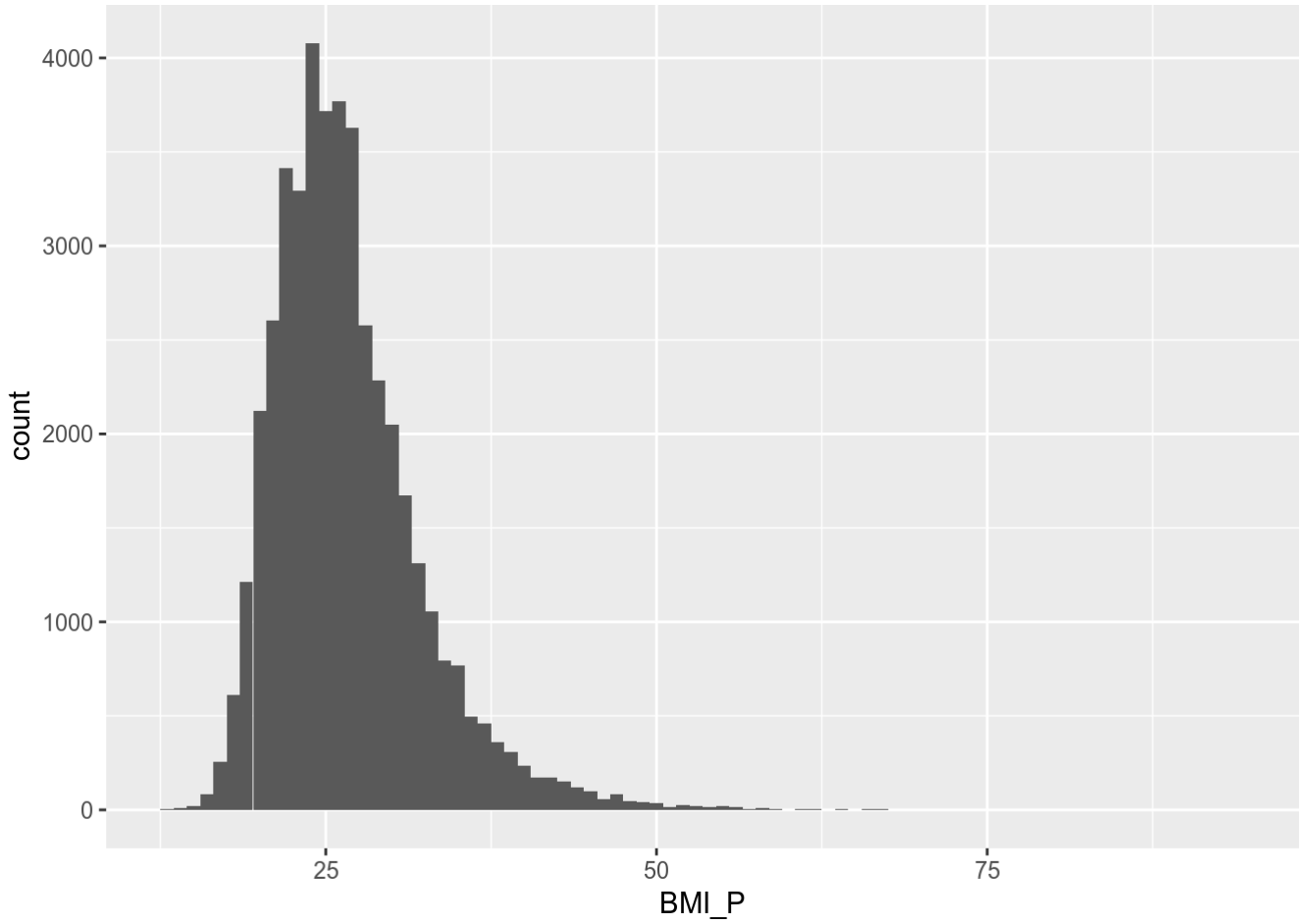
```
## Warning in data("chis2009"): data set 'chis2009' not found
```

```
data = load("/home/sedreh/ITMO/semester2/Statistic-R/8/CHIS2009_reduced_2.Rdata")
data <- adult
```

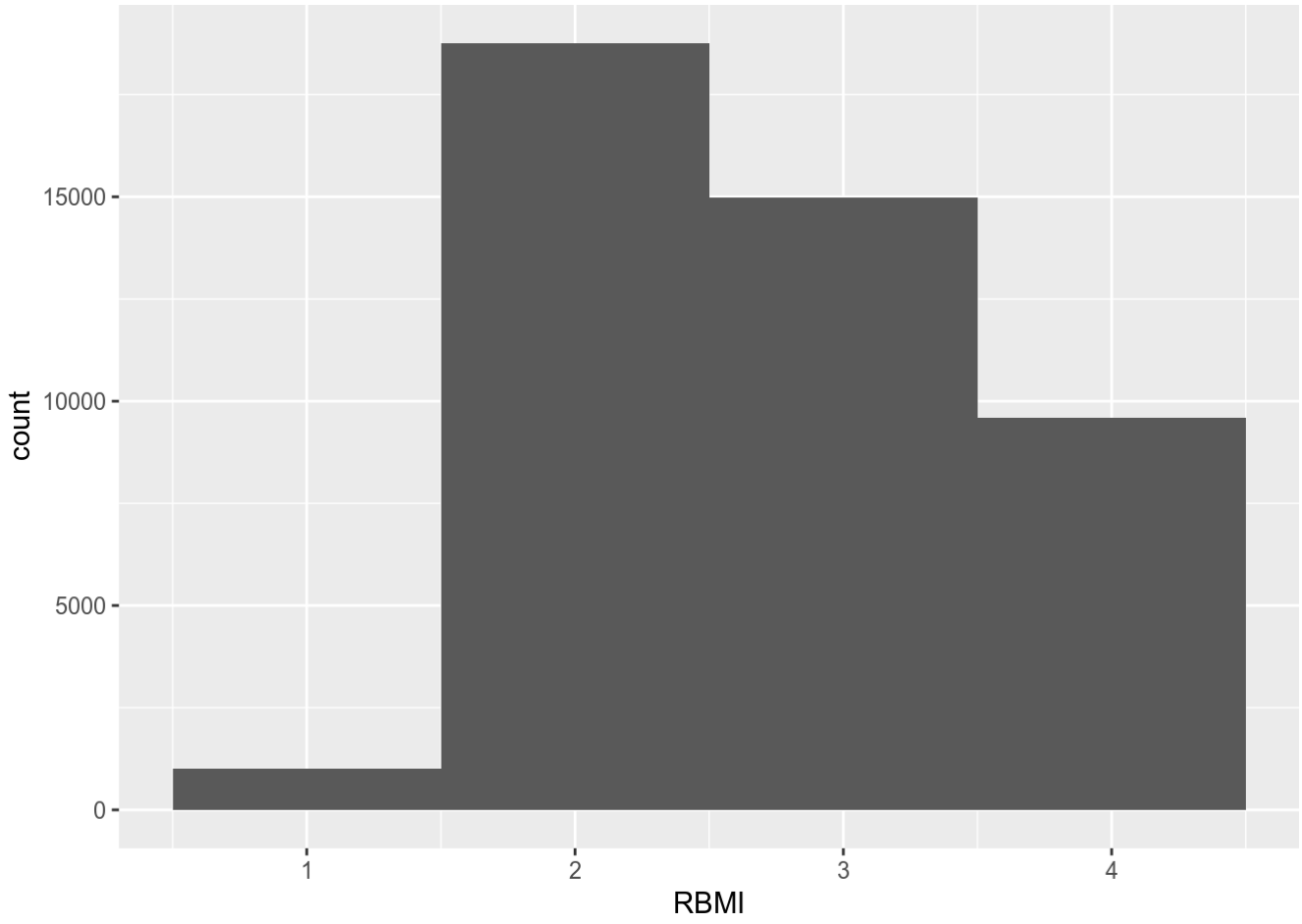
```
#Explore age (SRAGE_P)
p <- ggplot(adult, aes(SRAGE_P) )+
  geom_histogram(binwidth = 1)
p
```



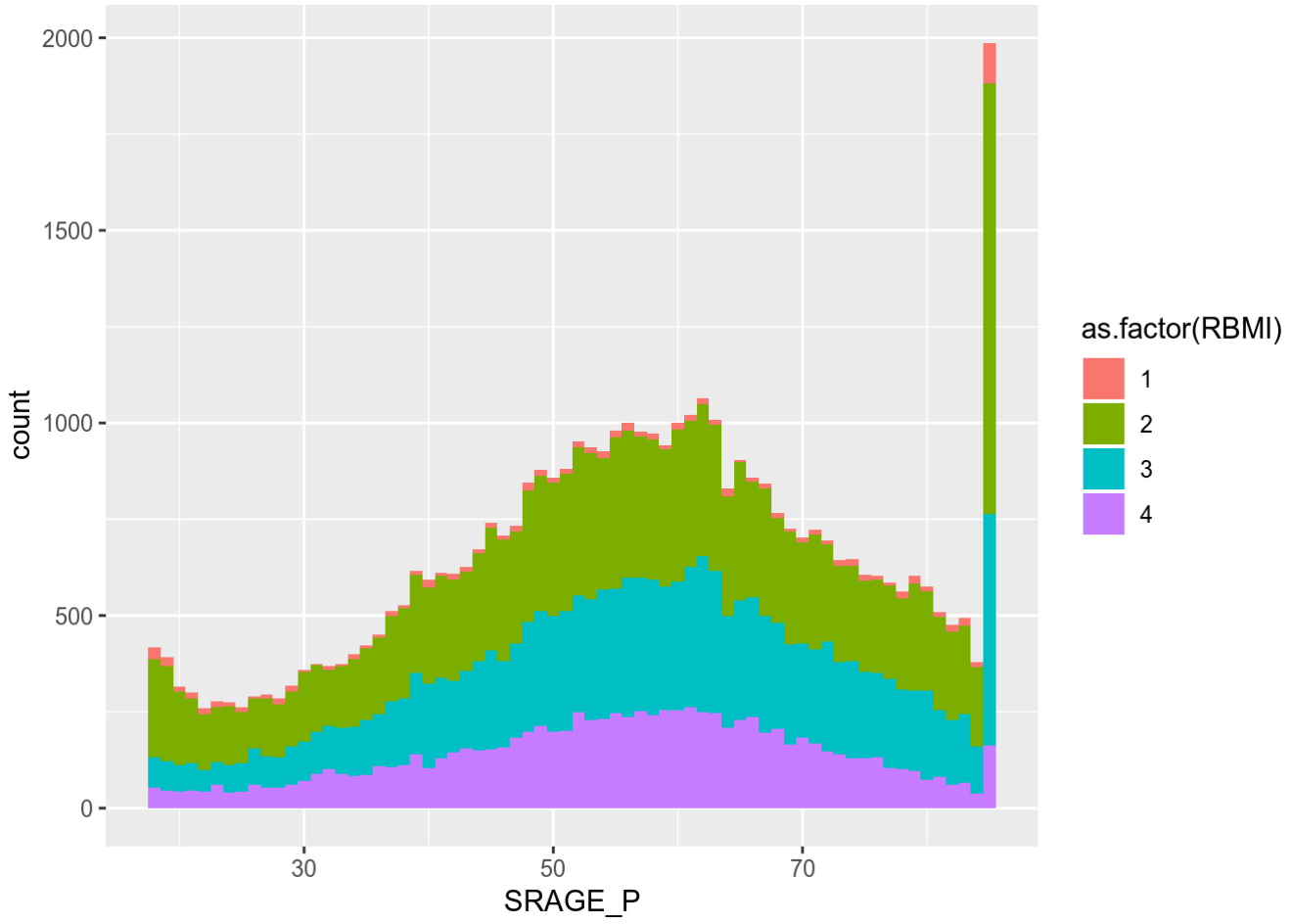

```
#Explore BMI (BMI_P)
p <- ggplot(adult, aes(BMI_P)) +
  geom_histogram(binwidth = 1)
p
```




```
#Explore BMI groups (RBMI)
p <- ggplot(adult, aes(RBMI)) +
  geom_histogram(binwidth = 1)
p
```




```
#histogram of ages colored by BMI groups
p <- ggplot(adult, aes(SRAGE_P, fill = as.factor(RBMI))) + geom_histogram(binwidth =
1)
p
```



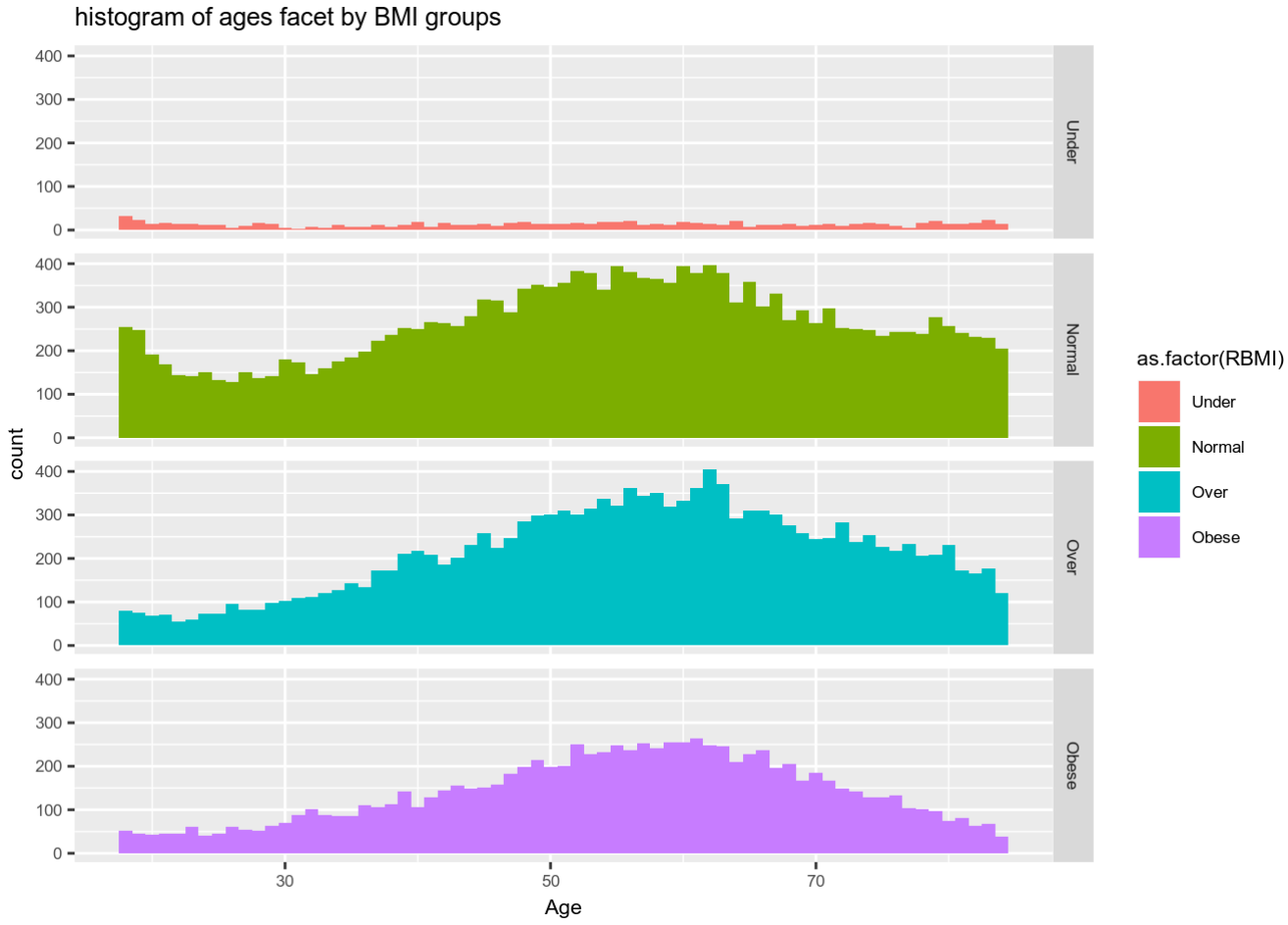

```
#Data cleaning: age under the value
# Age under the value
# Keep BMI between 16 (incl) and 52 (excl)

clean_data <- filter(data, BMI_P >= 16 & BMI_P <= 52)
clean_data <- filter(adult, SRAGE_P < 85)
```

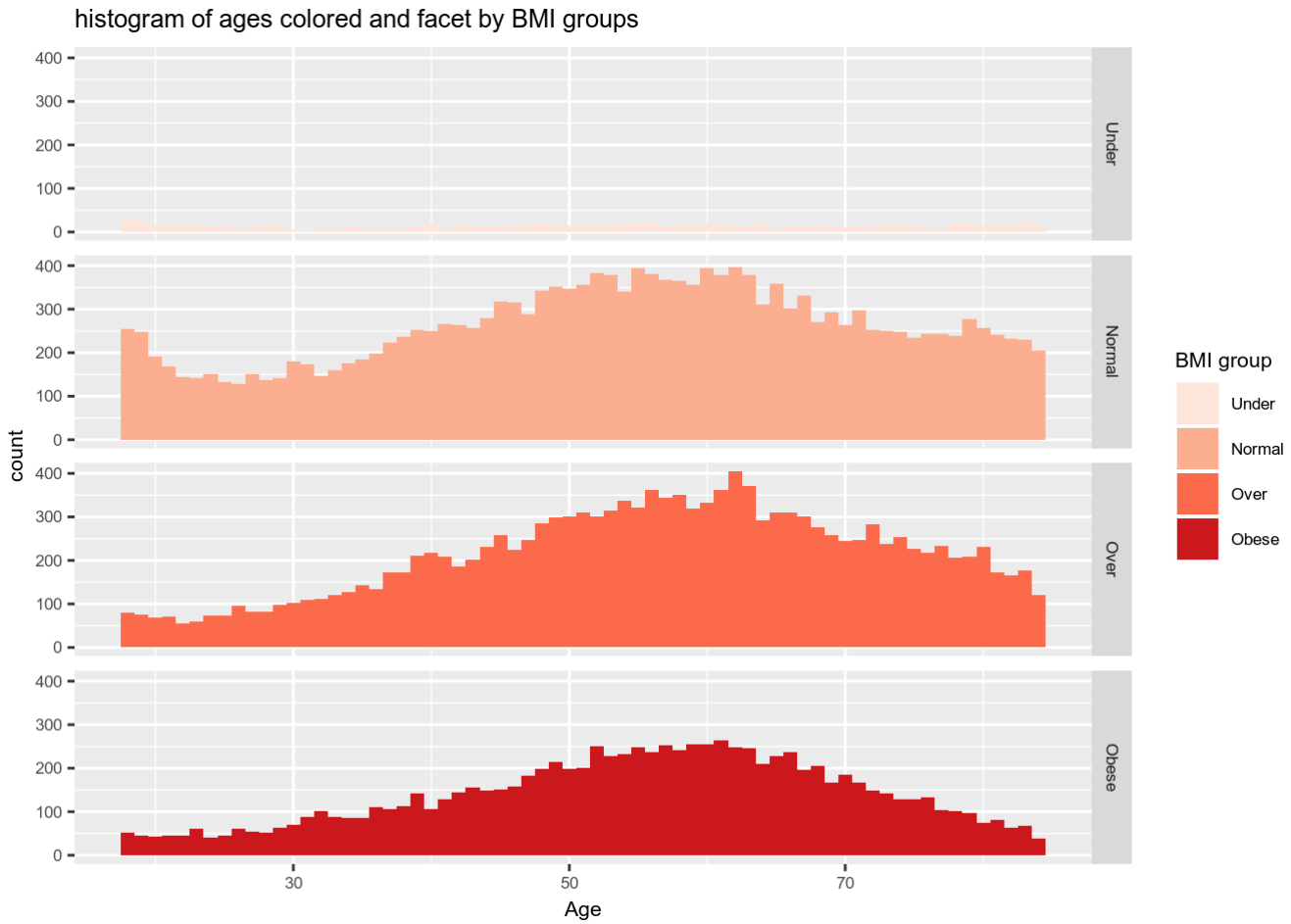
```
#Relabel race (RACEHPR2)
clean_data$RACEHPR2 <- factor(clean_data$RACEHPR2, labels = c("Latino", "Asian", "Afr
ican American", "White"))
```

```
#Relabel BMI groups
clean_data$RBMI <- factor(clean_data$RBMI, labels = c("Under", "Normal", "Over", "Obe
se"))
```

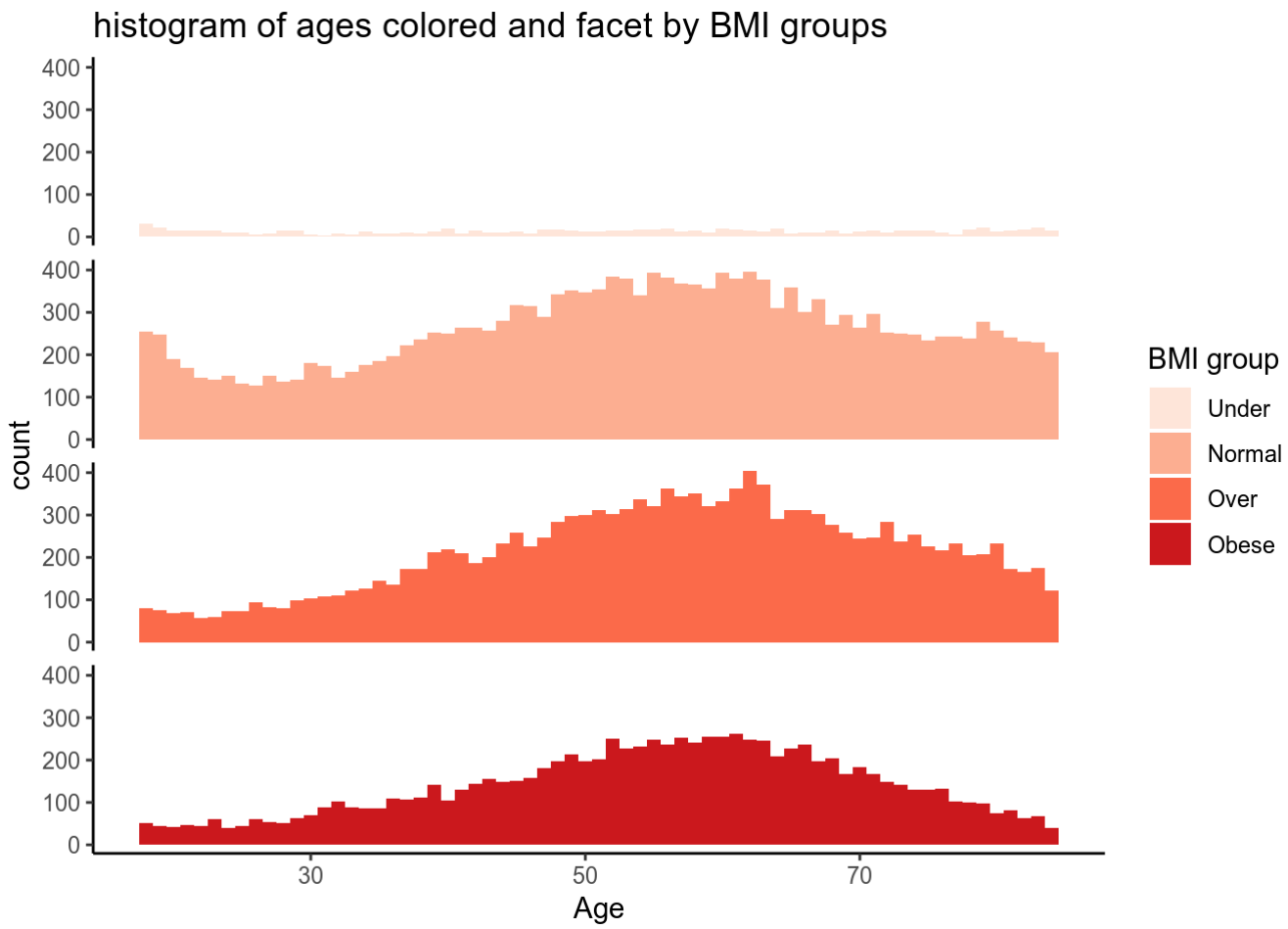
```
#Build a histogram of ages colored and facet by BMI groups
p <- ggplot(clean_data, aes(SRAGE_P, fill = as.factor(RBMI))) +
  geom_histogram(binwidth = 1)+
  facet_grid(RBMI~.)+
  labs(x = "Age")+
  ggtitle("histogram of ages facet by BMI groups")+
  theme(text = element_text(size = 8))
p
```



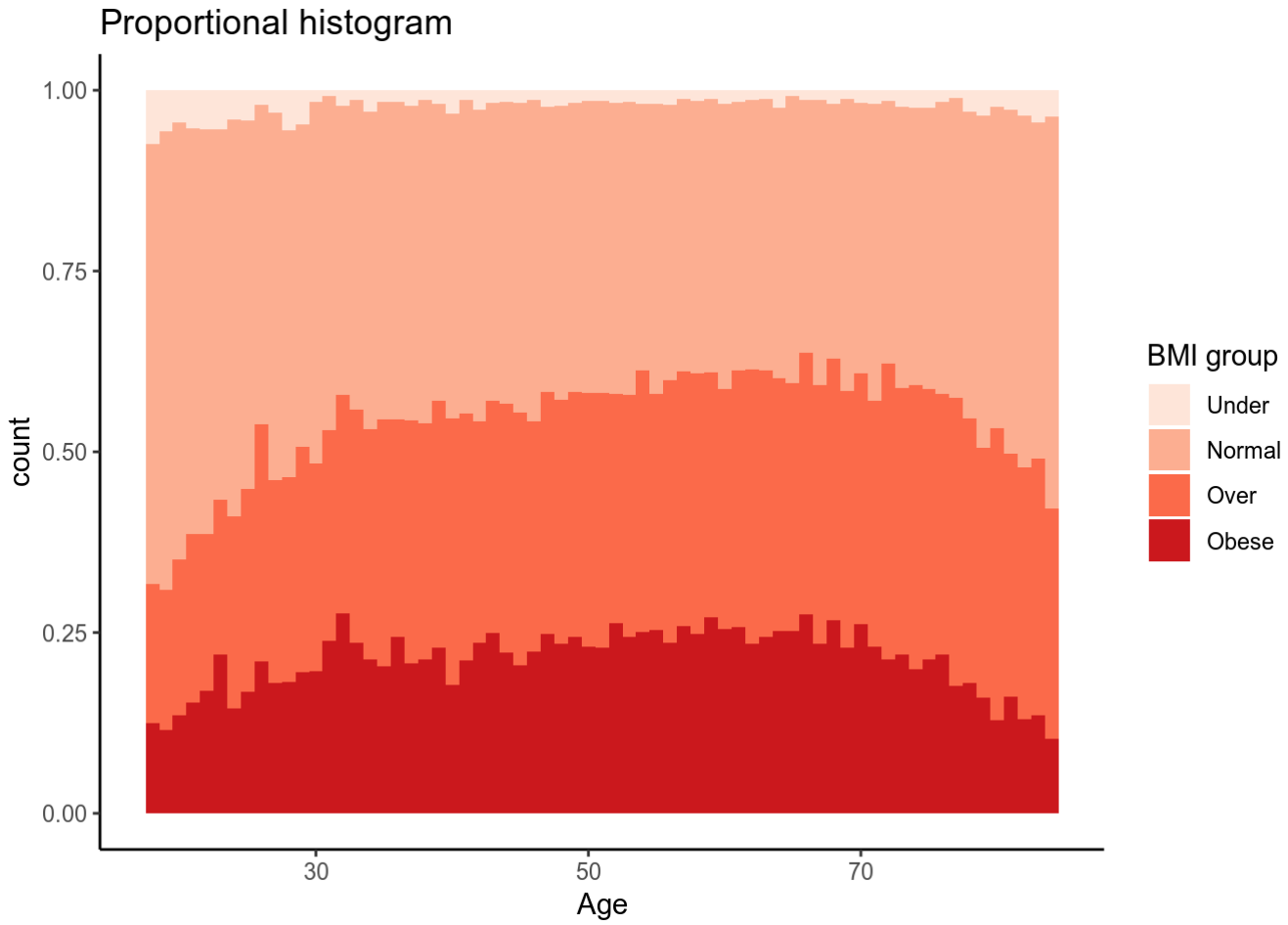

```
#Color with another palette
p <- ggplot(clean_data, aes(SRAGE_P, fill = as.factor(RBMI))) +
  geom_histogram(binwidth = 1)+
  facet_grid(RBMI~.)+
  labs(x = "Age")+
  ggtitle("histogram of ages colored and facet by BMI groups")+
  theme(text = element_text(size = 8))+
  scale_fill_brewer('BMI group', palette = 'Reds')
p
```




```
#Use theme_classic() and theme(strip.text.y = element_blank())
p <- ggplot(clean_data, aes(SRAGE_P, fill = as.factor(RBMI))) +
  geom_histogram(binwidth = 1)+
  facet_grid(RBMI~.)+
  labs(x = "Age")+
  ggtitle("histogram of ages colored and facet by BMI groups")+
  theme(text = element_text(size = 8))+
  scale_fill_brewer('BMI group', palette = 'Reds')+
  theme_classic()+
   theme(strip.text.y = element_blank())
p
```

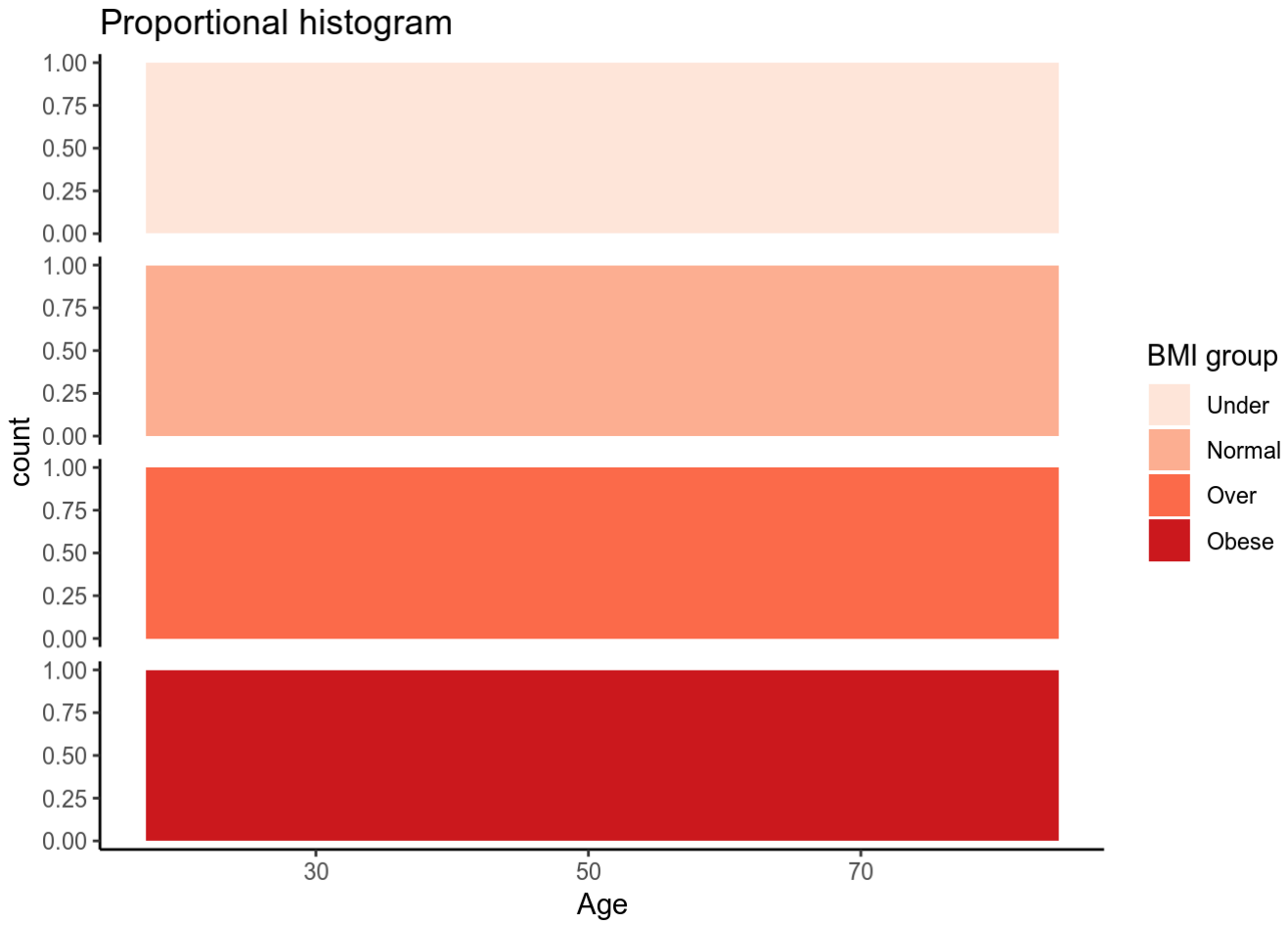



```
#proportional histogram
p <- ggplot(clean_data, aes(SRAGE_P, fill = as.factor(RBMI))) +
  geom_histogram(binwidth = 1, position = "fill")+
  labs(x = "Age")+
  ggtitle("Proportional histogram")+
  theme(text = element_text(size = 8))+
  scale_fill_brewer('BMI group', palette = 'Reds')+
  theme_classic()+
  theme(strip.text.y = element_blank())
p
```




```
#use both facet and proportion
p <- ggplot(clean_data, aes(SRAGE_P, fill = as.factor(RBMI))) +
  geom_histogram(binwidth = 1, position = "fill")+
  facet_grid(RBMI~.)+
  labs(x = "Age")+
  ggtitle("Proportional histogram")+
  theme(text = element_text(size = 8))+
  scale_fill_brewer('BMI group', palette = 'Reds')+
  theme_classic()+
  theme(strip.text.y = element_blank())

p
```




```
#we can't understand any useful information from this plot!!
```

```
#Make frequency table with ages(columns) and proportions of each BMI group relative t
o each age(rows)

# A frequency table is a table that represents the number of occurrences of every uni
que value in the variable.
#You can generate frequency tables using the table( ) function, tables of proportions
using the prop.table( ) function, and marginal frequencies using margin.table( ).

Two_way_table <- table(clean_data$RBMI, clean_data$SRAGE_P )
#Two_way_table

library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
prop_df <- melt(Two_way_table)
names(prop_df) <- c("mesure", "Age", "value")
head(prop_df)
```

```
##   mesure Age value
## 1  Under  18    31
## 2 Normal  18   254
## 3   Over  18    80
## 4  Obese  18    52
## 5  Under  19    22
## 6 Normal  19   248
```

```
#second method
#Transform this frequency table using reshape2::melt() function
# melting this new_df_prop (new matrix of proportions)
# new_melted <- data.frame(reshape2::melt(prop_df))
# new_melted
```

```
#using geom_col() instead of histogram
# Build a histogram of ages colored by BMI groups
# Add facet by BMI group
# Color with another palette: scale_fill_brewer("BMI group", palette = "Reds")
# Use theme_classic() and theme(strip.text.y = element_blank())
p <- ggplot(prop_df, aes(x=Age, y = value, fill = mesure)) +
  geom_col(position = "stack") +
  facet_grid(mesure~.)+
  scale_fill_brewer("BMI group", palette = "Reds")+
  theme_classic() + theme(strip.text.y = element_blank())
p
```