STA 478 Assingment 1

Jared Brotamonte

8/29/2023

Exercise 1

```r
# (a)
setwd("C:/Users/jkbro/OneDrive/Desktop/STA 478/Homework/HWK 1")
load("compstats1.Rdata")
ls()
```

```
## [1] "ages"
```

```r
print(ages)
```

```
## [1] 17 18 16 20 22 23
```

```r
# (b)
ageLevels <- function(age){
  if(age < 18){
    return("Minor")
  }
  else{
    return("Adult")
  }
}
ageLevels <- factor(sapply(ages, ageLevels), levels = c("Minor","Adult"))
print(ageLevels)
```

```
## [1] Minor Adult Minor Adult Adult Adult
## Levels: Minor Adult
```

```r
agesOrdered <- ages[order(ageLevels)]
print(agesOrdered)
```

```
## [1] 17 16 18 20 22 23
```

Exercise 2

```
# (a)
data("ChickWeight")

# (b)
ChickFiltered <- ChickWeight[ChickWeight$Time %in% c(10, 20), ]
head(ChickFiltered)
```

```
##    weight Time Chick Diet
## 6      93   10     1    1
## 11    199   20     1    1
## 18    103   10     2    1
## 23    209   20     2    1
## 30     99   10     3    1
## 35    198   20     3    1
```

```
# (c)
FilteredMeanAndSD <- ChickFiltered %>%
  group_by(Diet, Time) %>%
  summarise(mean_weight = mean(weight), sd_weight = sd(weight))
```

```
## 'summarise()' has grouped output by 'Diet'. You can override using the
## '.groups' argument.
```

```
print(FilteredMeanAndSD)
```

```
## # A tibble: 8 x 4
## # Groups:   Diet [4]
##   Diet   Time mean_weight sd_weight
##   <fct> <dbl>       <dbl>     <dbl>
## 1 1        10        93.1      22.5
## 2 1        20       170.       55.4
## 3 2        10       108.       24.3
## 4 2        20       206.       70.3
## 5 3        10       117.       20.2
## 6 3        20       259.       65.2
## 7 4        10       126        11.4
## 8 4        20       234.       37.6
```

Exercise 3

```
# (a)
Survey <- read.csv("https://www.lock5stat.com/datasets3e/StudentSurvey.csv",
                   na.strings=c(""," ")) %>%
  mutate(Year = factor(Year,
                       levels=c("FirstYear","Sophomore","Junior","Senior"))) %>%
  mutate(Sex = factor(Sex))

# (b)
```

```
SurveyCounts <- Survey %>%
  filter(!is.na(Year)) %>%
  group_by(Year, Sex) %>%
  summarise(response_count = n())
```

```
## 'summarise()' has grouped output by 'Year'. You can override using the
## '.groups' argument.
```

```
print(SurveyCounts)
```

```
## # A tibble: 8 x 3
## # Groups:   Year [4]
##   Year      Sex   response_count
##   <fct>     <fct>          <int>
## 1 FirstYear F                 43
## 2 FirstYear M                 51
## 3 Sophomore F                 96
## 4 Sophomore M                 99
## 5 Junior    F                 18
## 6 Junior    M                 17
## 7 Senior    F                 10
## 8 Senior    M                 26
```

```
# (c)
SurveyTable <- SurveyCounts %>%
  pivot_wider(names_from = Year, values_from = response_count)
print(SurveyTable)
```

```
## # A tibble: 2 x 5
##   Sex   FirstYear Sophomore Junior Senior
##   <fct>     <int>     <int>  <int>  <int>
## 1 F            43        96     18     10
## 2 M            51        99     17     26
```
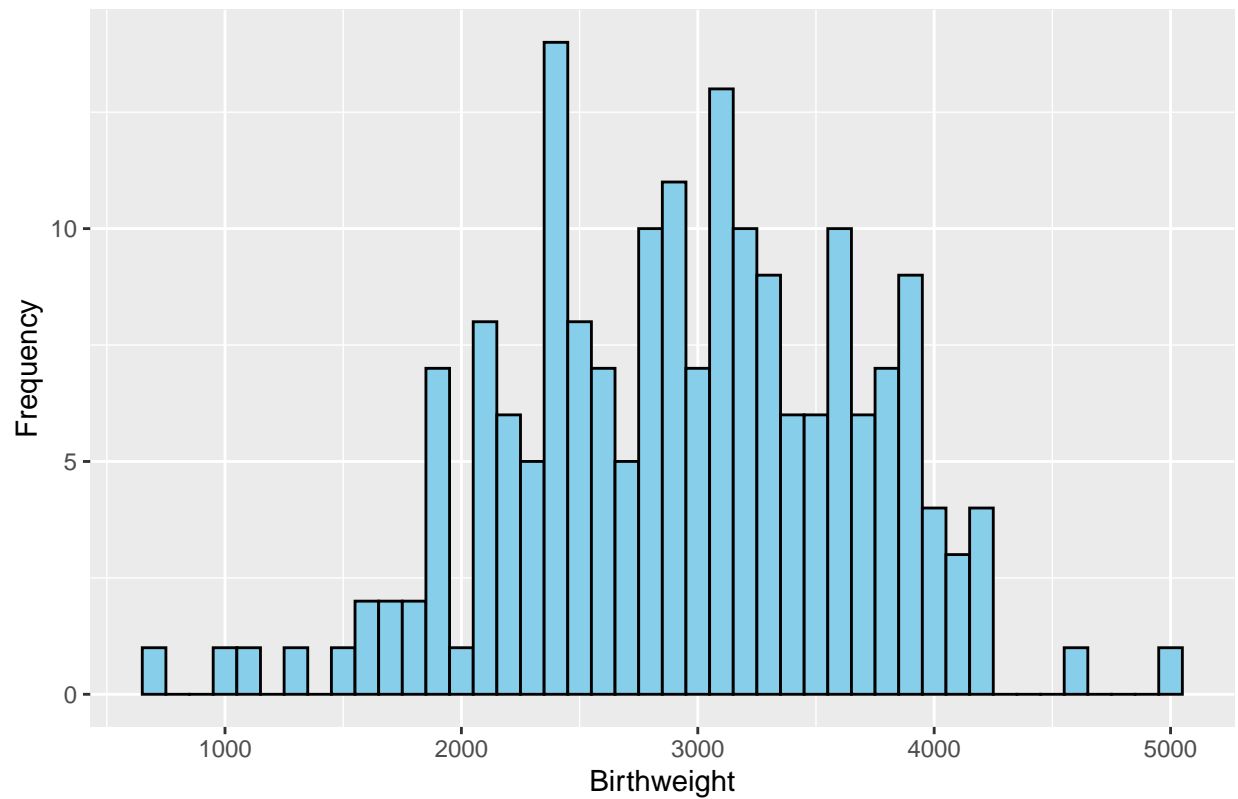
Exercise 4

```
# (a)
# Loaded MASS at the beginning of session

# (b)
birthwt <- birthwt %>%
  mutate(race = factor(race, labels=c("White","Black","Other")),
         smoke = factor(smoke, labels=c("No Smoke", "Smoke")))

# (c)
ggplot(data = birthwt, aes(x = bwt)) +
  geom_histogram(binwidth = 100, fill = "skyblue", color = "black")+
  labs(title = "Histogram of Birthweights",
       x = "Birthweight",
       y = "Frequency")
```
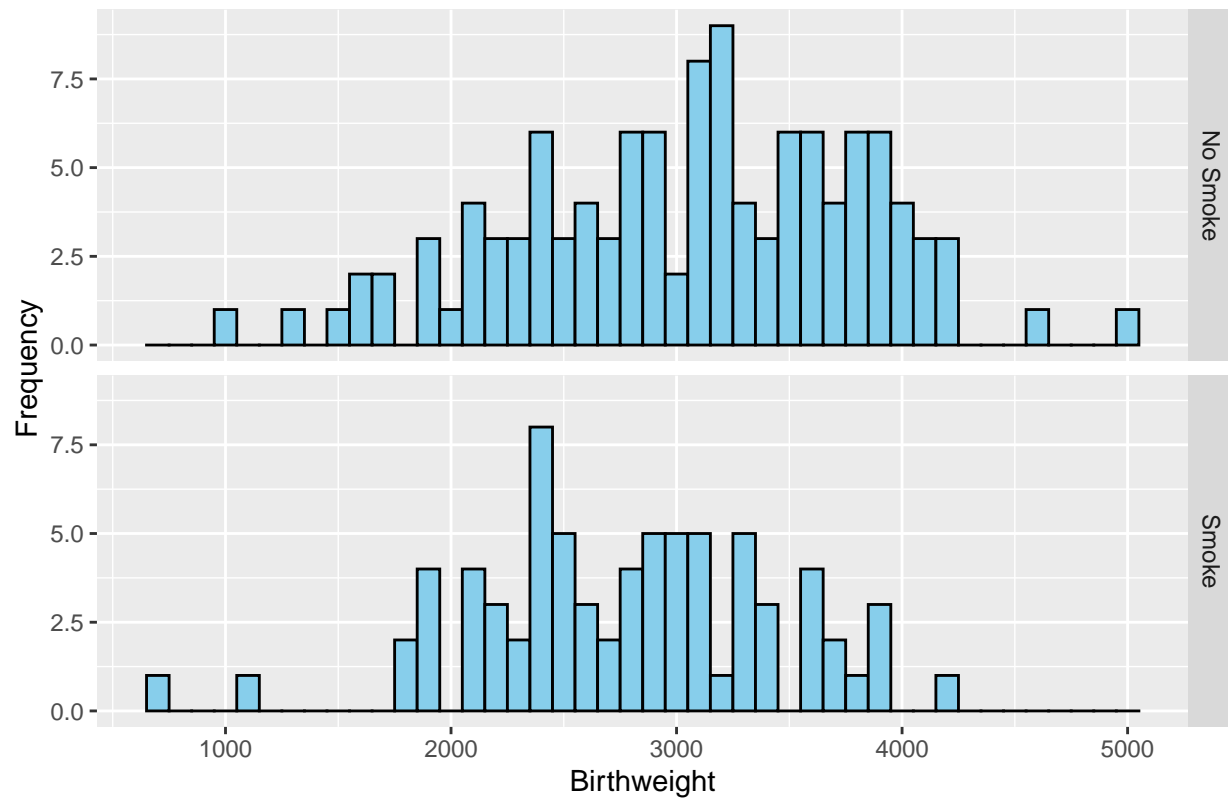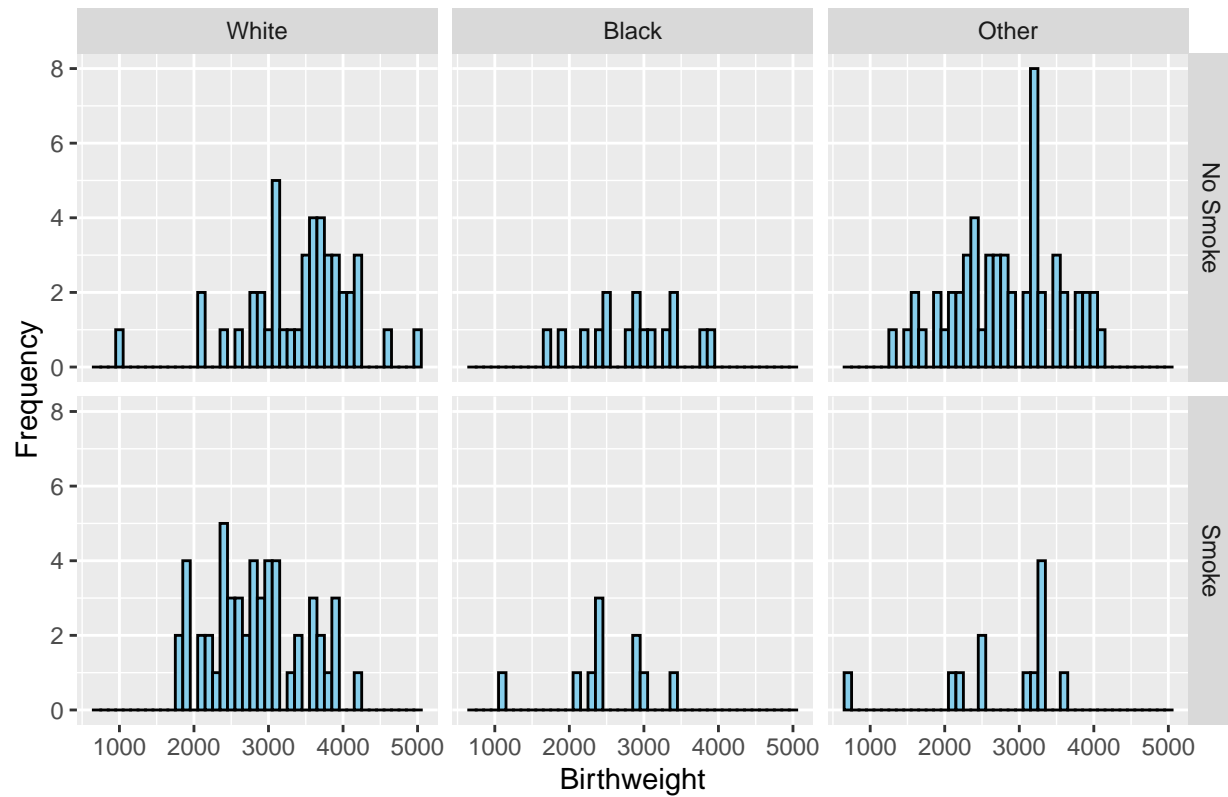
Histogram of Birthweights


```
# (d)
ggplot(data = birthwt, aes(x = bwt)) +
  geom_histogram(binwidth = 100, fill = "skyblue", color = "black")+
  labs(title = "Histogram of Birthweights",
       x = "Birthweight",
       y = "Frequency")+
  facet_grid(smoke ~.)
```
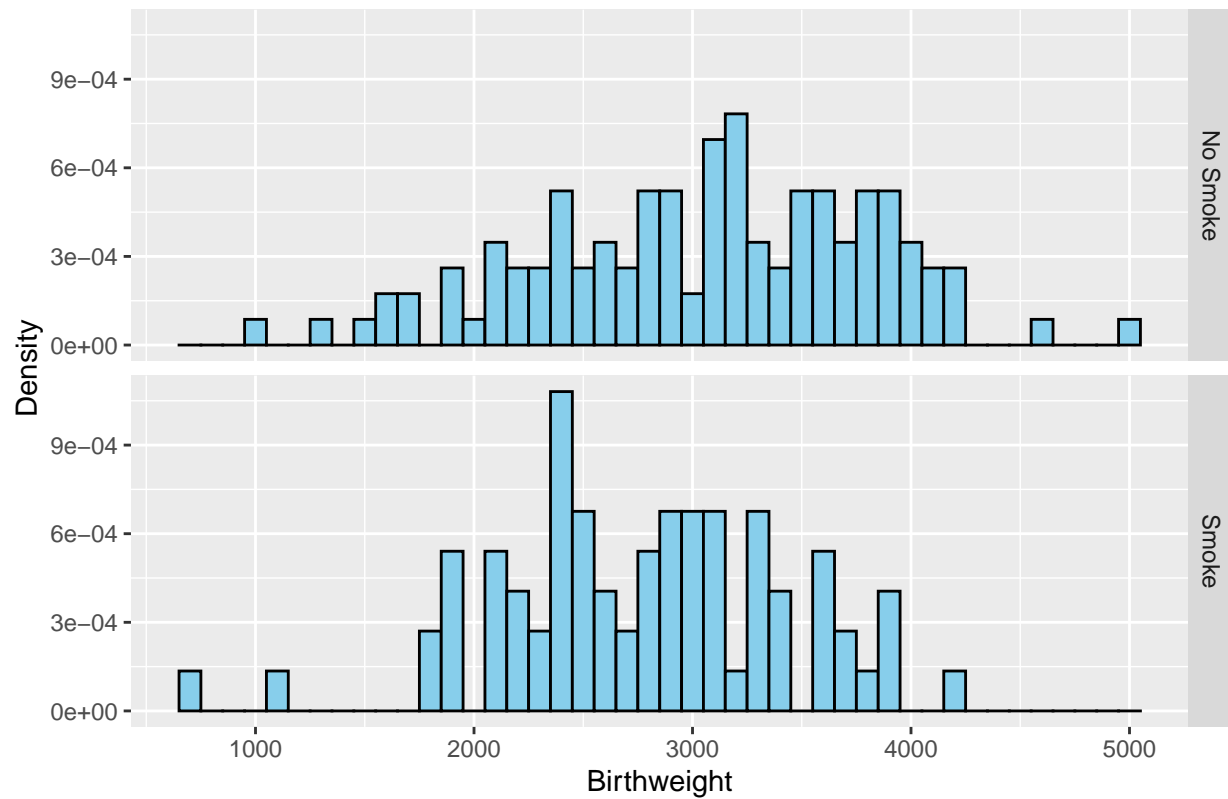
Histogram of Birthweights


```
# (e)
ggplot(data = birthwt, aes(x = bwt)) +
  geom_histogram(binwidth = 100, fill = "skyblue", color = "black")+
  labs(title = "Histogram of Birthweights",
       x = "Birthweight",
       y = "Frequency")+
  facet_grid(smoke ~ race)
```
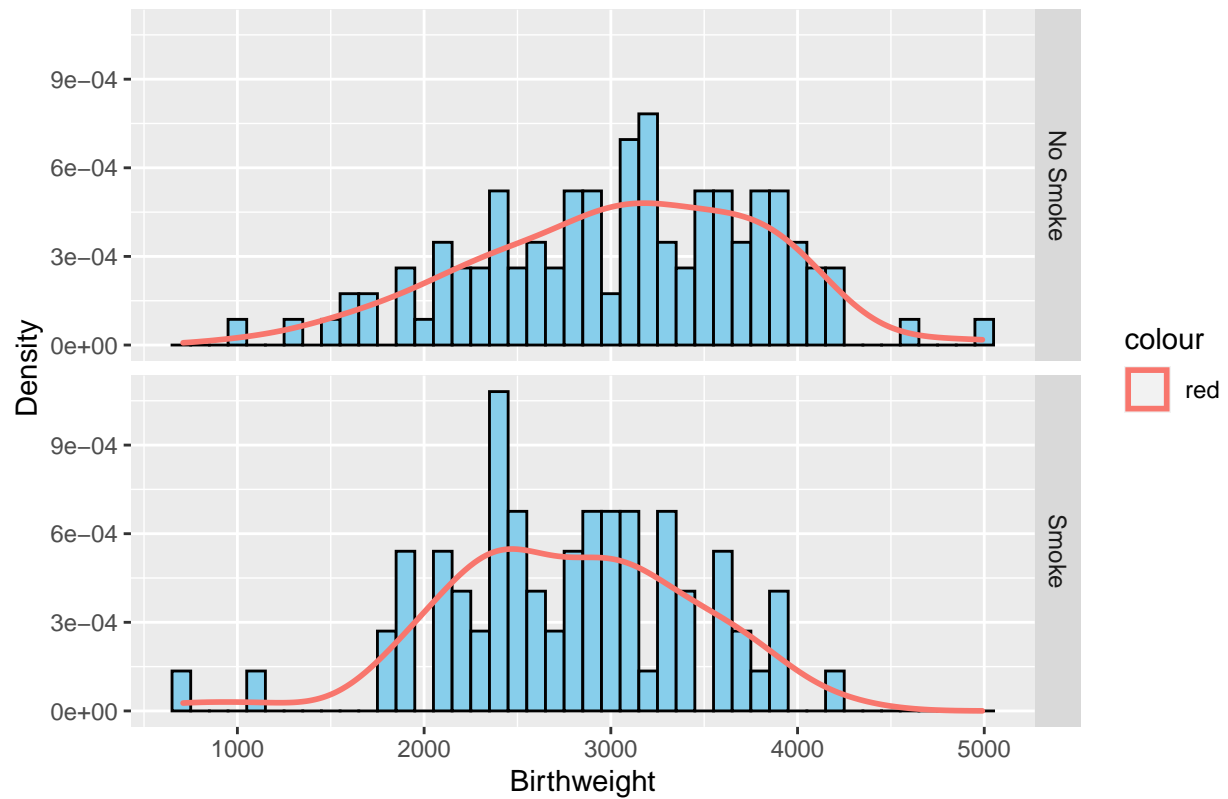
Histogram of Birthweights


```
# (f)
ggplot(data = birthwt, aes(x = bwt, y = ..density..)) +
  geom_histogram(binwidth = 100, fill = "skyblue", color = "black") +
  labs(title = "Histogram of Birthweights",
       x = "Birthweight",
       y = "Density")+
  facet_grid(smoke ~.)
```

Histogram of Birthweights


```
# (g)
ggplot(data = birthwt, aes(x = bwt, y = ..density..)) +
  geom_histogram(binwidth = 100, fill = "skyblue", color = "black") +
  geom_density(aes(color = "red"), size = 1) +
  labs(title = "Histogram of Birthweights",
       x = "Birthweight",
       y = "Density")+
  facet_grid(smoke ~.)
```
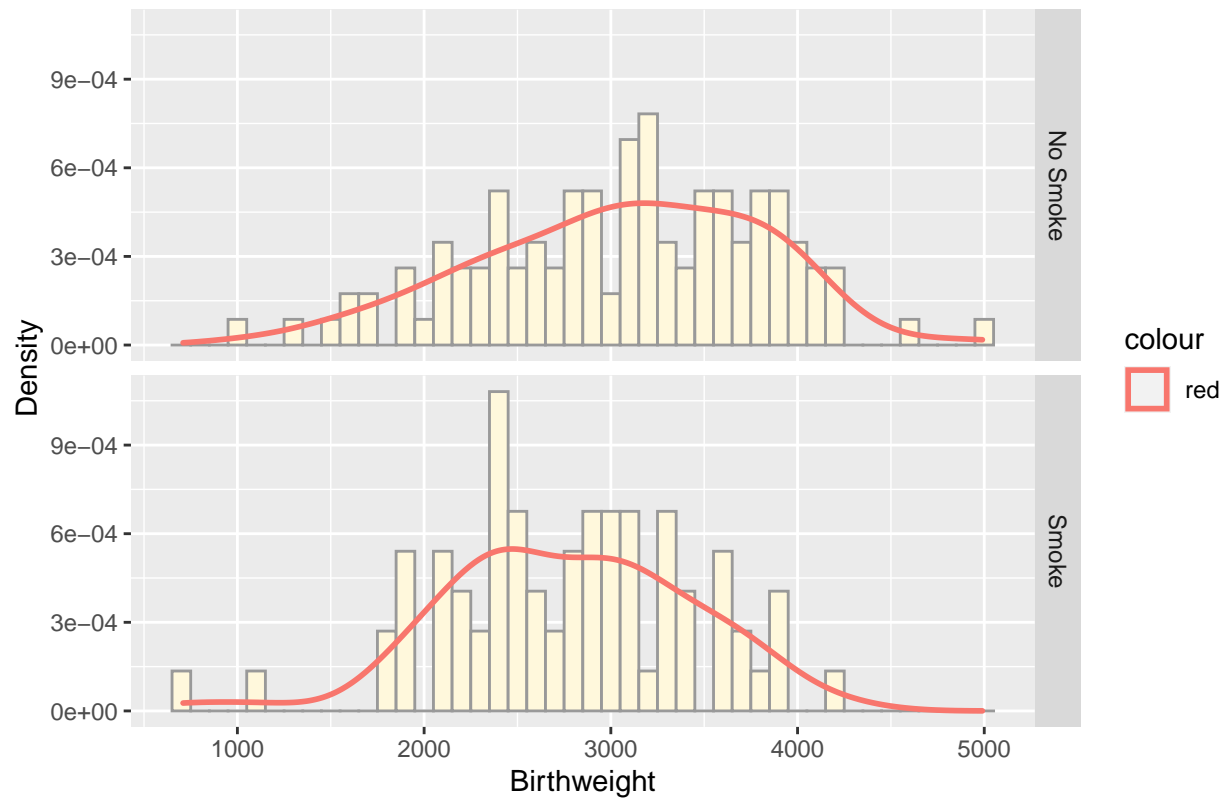
Histogram of Birthweights


```
# (h)
ggplot(data = birthwt, aes(x = bwt, y = ..density..)) +
  geom_histogram(binwidth = 100, fill = "cornsilk", color = "grey60") +
  geom_density(aes(color = "red"), size = 1) +
  labs(title = "Histogram of Birthweights",
       x = "Birthweight",
       y = "Density")+
  facet_grid(smoke ~.)
```
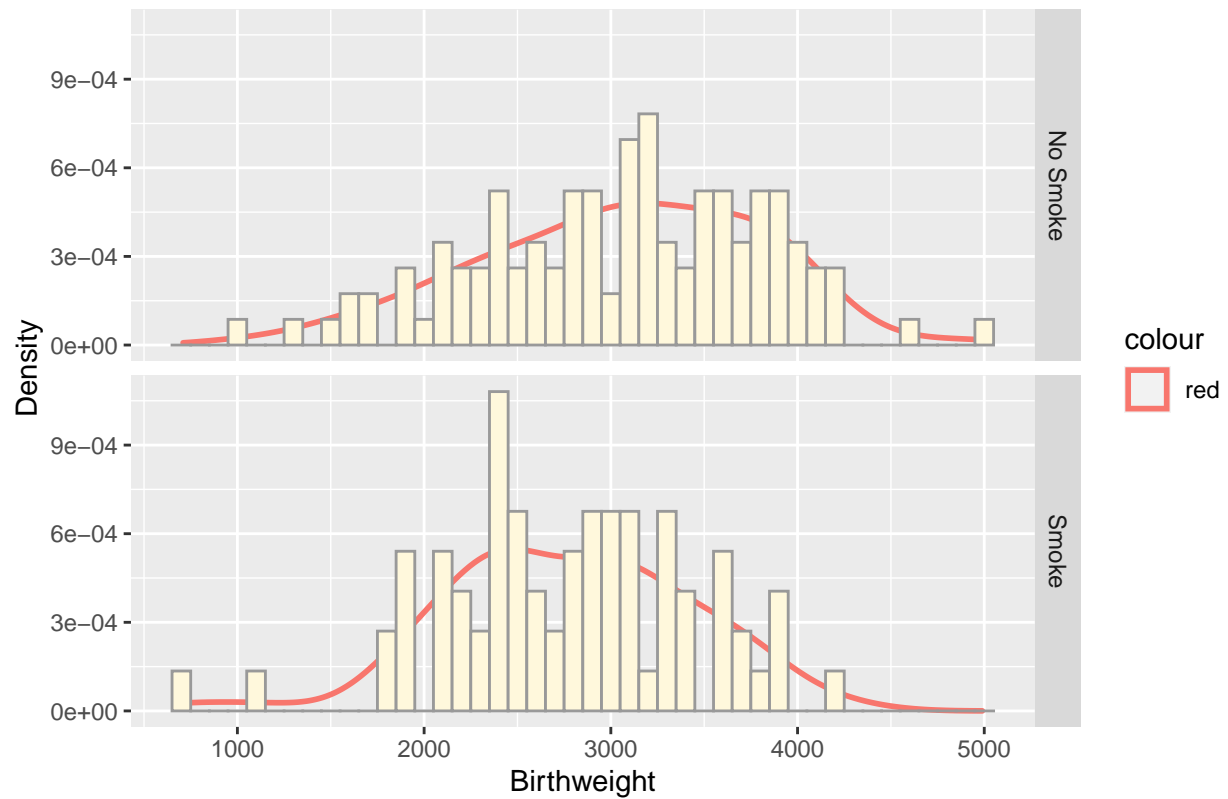
Histogram of Birthweights

```r
# (i)
ggplot(data = birthwt, aes(x = bwt, y = ..density..)) +
  geom_density(aes(color = "red"), size = 1) +
  geom_histogram(binwidth = 100, fill = "cornsilk", color = "grey60") +
  labs(title = "Histogram of Birthweights",
       x = "Birthweight",
       y = "Density")+
  facet_grid(smoke ~.)
```

Histogram of Birthweights

For part(i) of question 4, the order does matter. Personally I prefer doing the histogram first then the density line because it looks cleaner and is easier to see the smooth density throughout the data.