Unit 7 - Visualizing Attributes of Parole Violators

```
library(ggplot2)
parole <- read.csv("parole.csv")
parole$male = as.factor(parole$male)
parole$state = as.factor(parole$state)
parole$crime = as.factor(parole$crime)
```

Problem 1.1

```
table(parole$male, parole$violator)
```

```
##
##       0   1
##   0 116  14
##   1 481  64
```

```
14 / (14 + 64)
```

```
## [1] 0.1794872
```

Problem 1.2

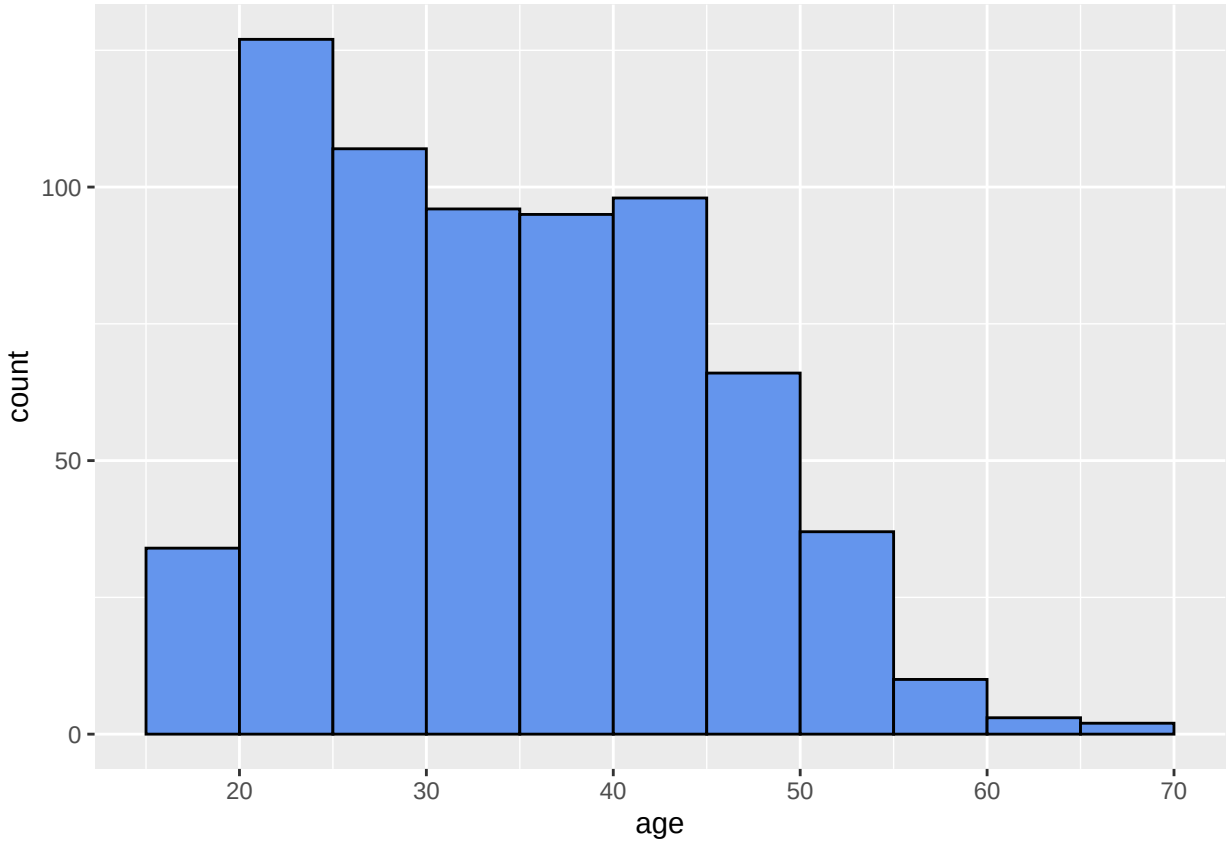
```
table(parole$state, parole$crime)
```

```
##
##       1   2   3   4
##   1  66   9  34  34
##   2  42  10  64   4
##   3  42  15  20   5
##   4 165  72  35  58
```

Drug-related crime

Problem 2.1

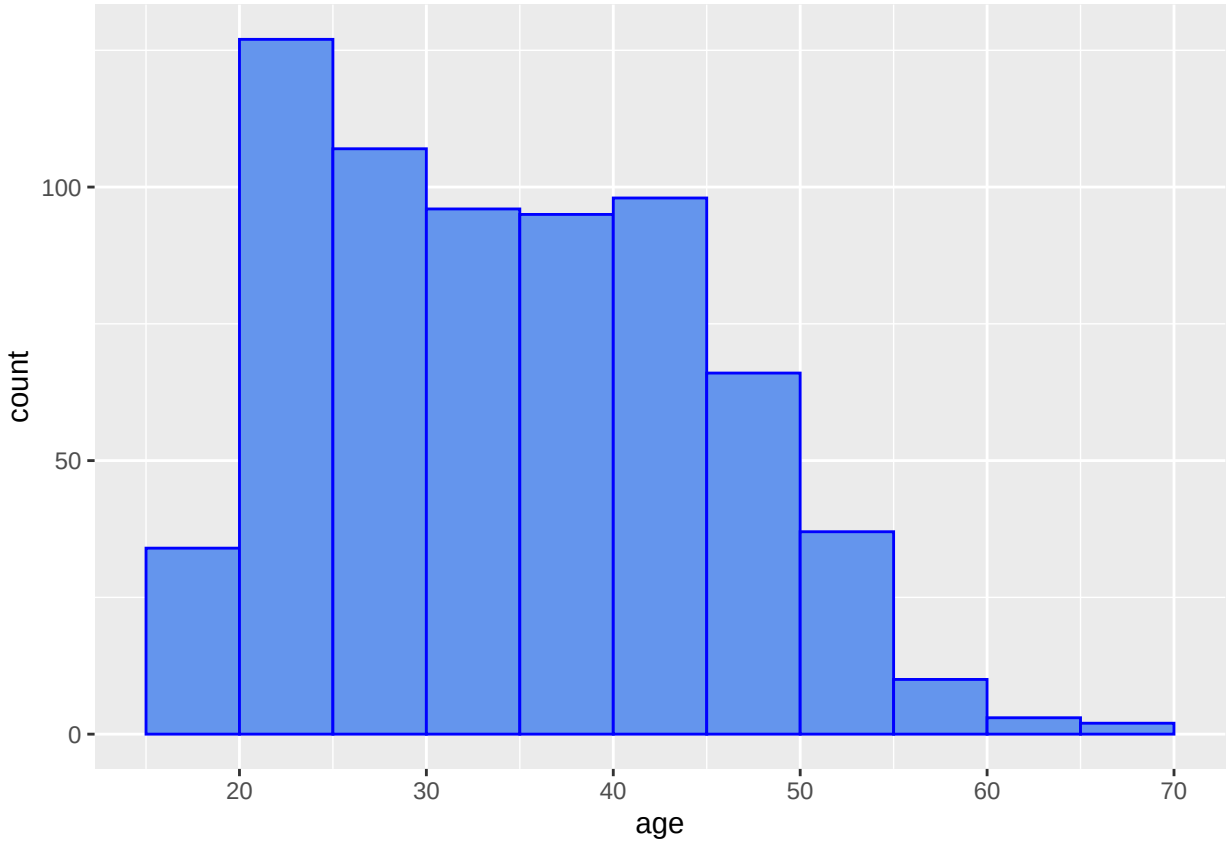
```
ggplot(data = parole, aes(x = age)) +
  geom_histogram(binwidth = 5, boundary = 0, color = 'black', fill = 'cornflowerblue')
```



20-24

Problem 2.2

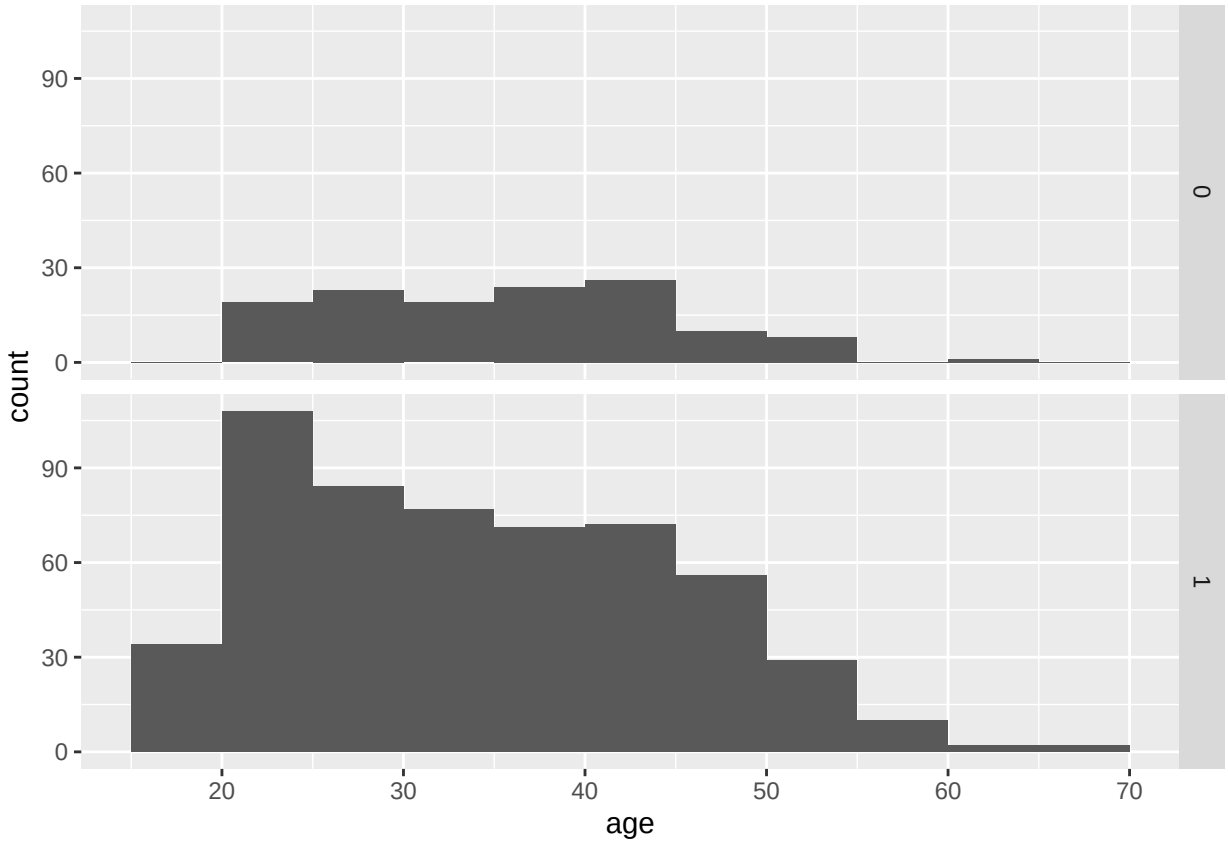
```
ggplot(data = parole, aes(x = age)) +
  geom_histogram(binwidth = 5, boundary = 0, color = 'blue', fill = 'cornflowerblue')
```



Changes the outline color of the bars

Problem 3.1

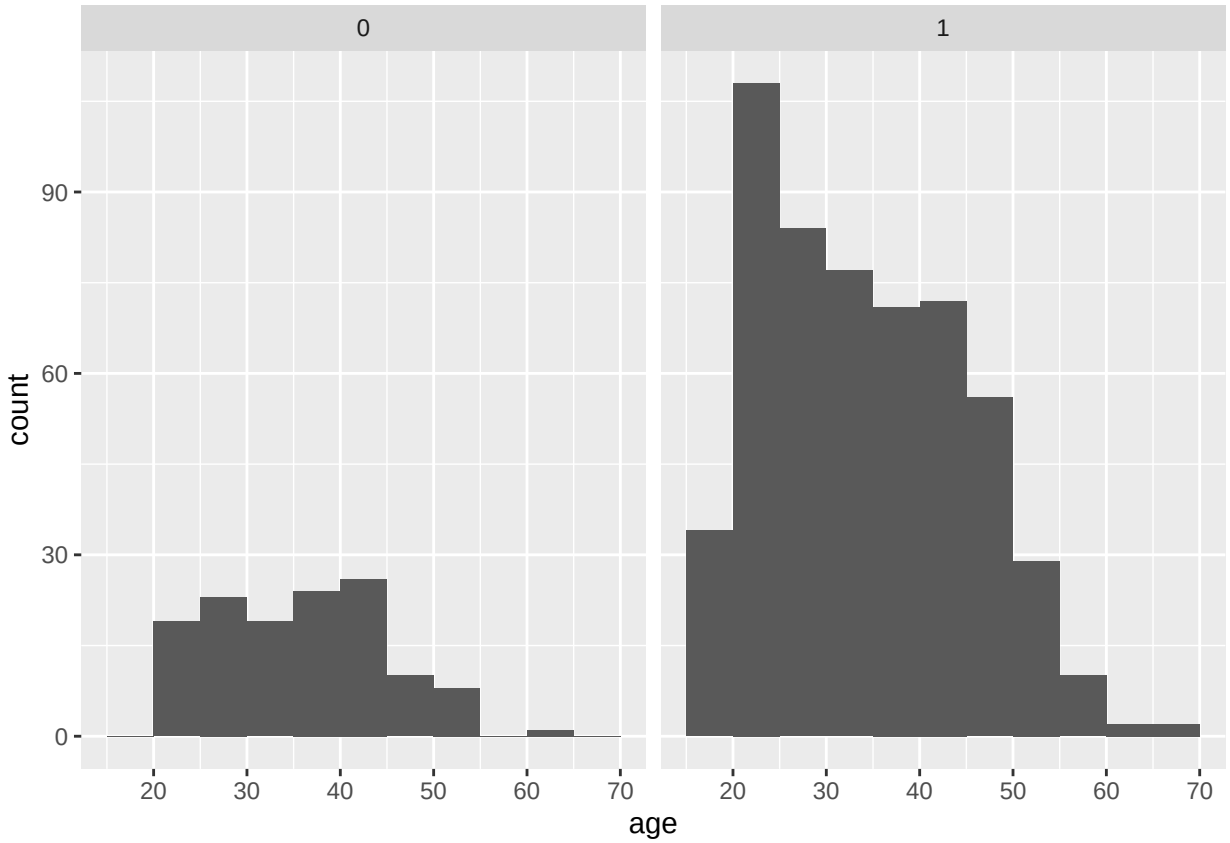
```
ggplot(data = parole, aes(x = age)) +
  geom_histogram(binwidth = 5, boundary = 0) + facet_grid(male ~ .)
```



35-39

Problem 3.2

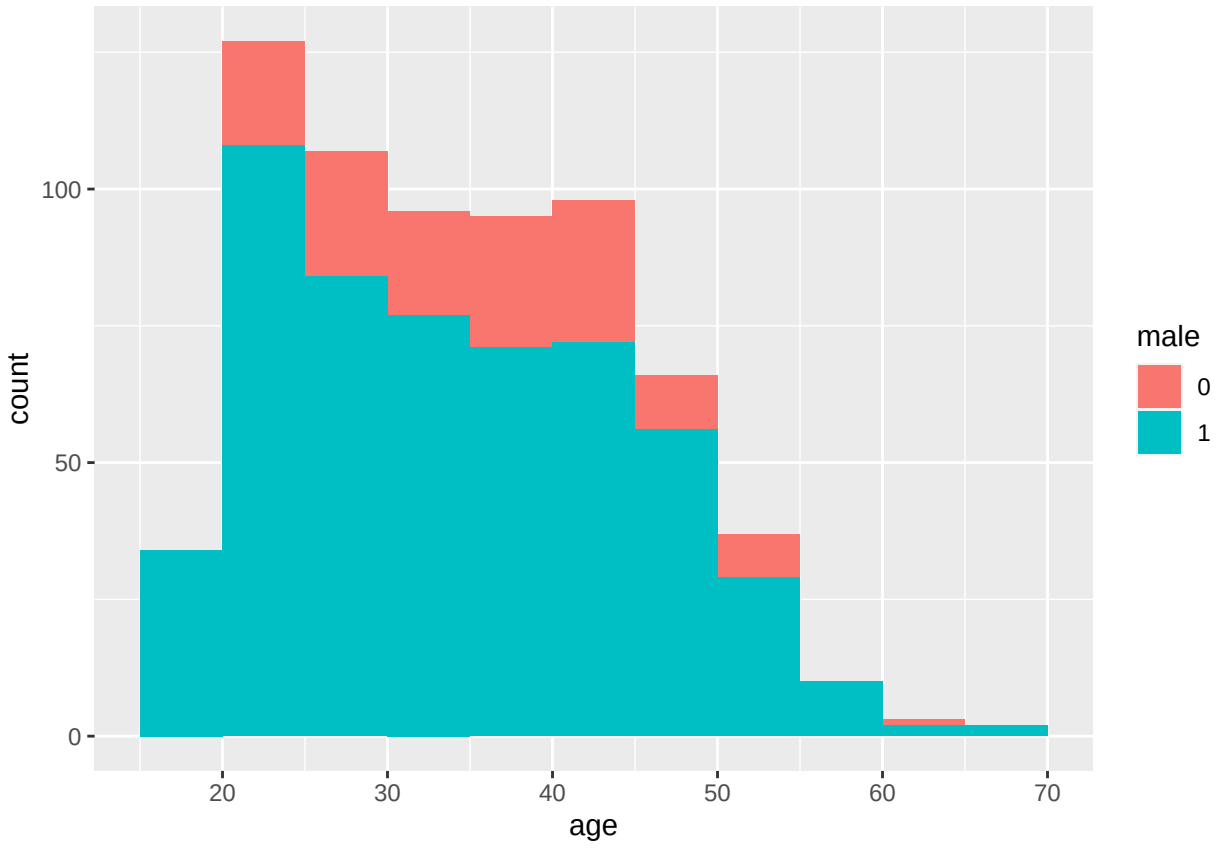
```
ggplot(data = parole, aes(x = age)) +
  geom_histogram(binwidth = 5, boundary = 0) +
  facet_grid(. ~ male)
```



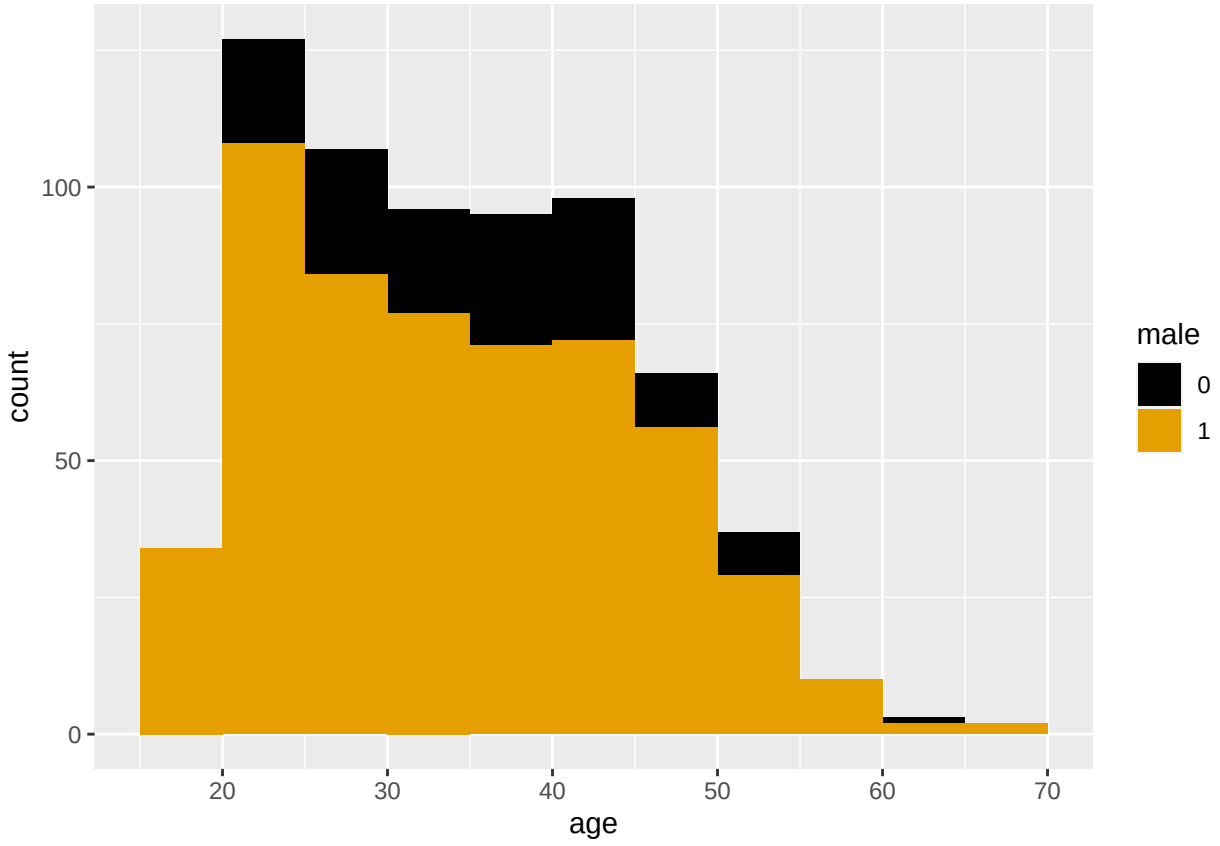
Puts the histograms side-by-side instead of on top of each other.

Problem 3.3

```
ggplot(data = parole, aes(x = age, fill = male)) +
  geom_histogram(binwidth = 5, boundary = 0)
```



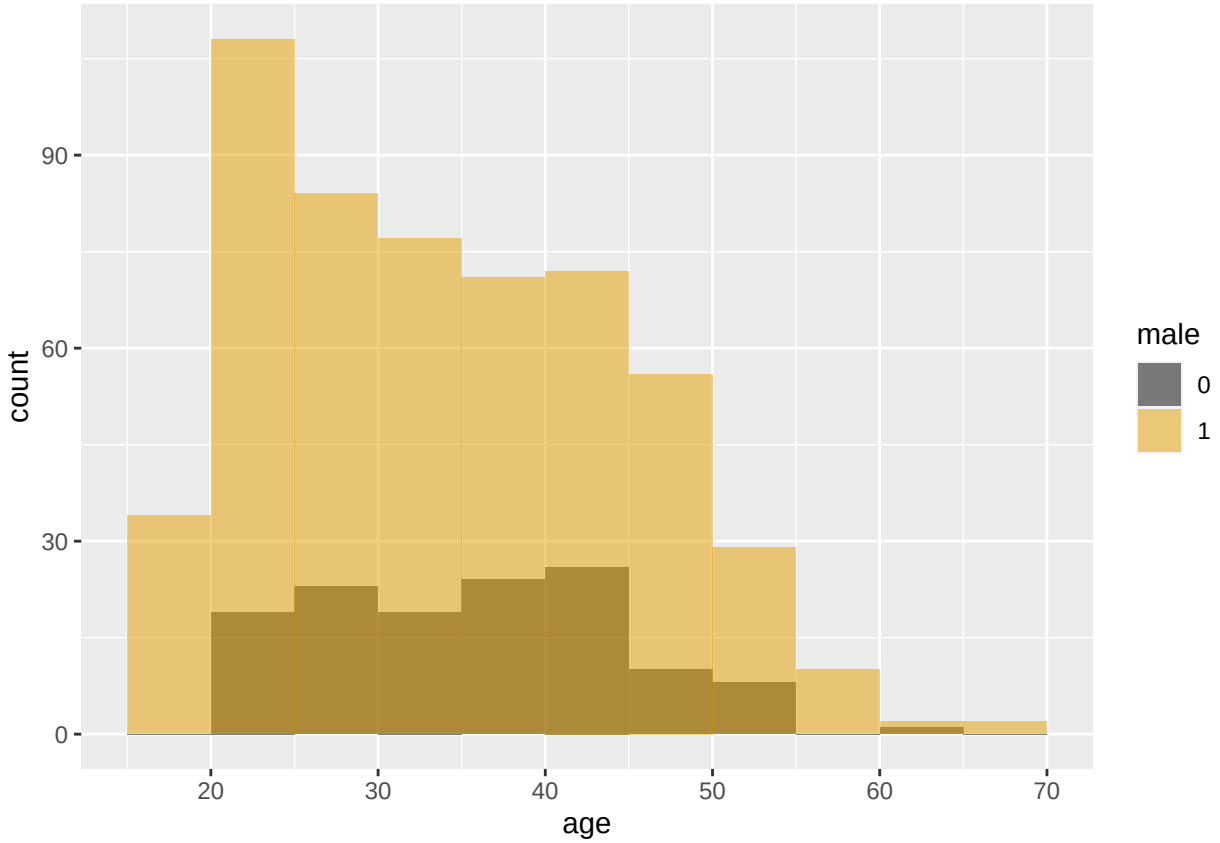

```
colorPalette = c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7"
ggplot(data = parole, aes(x = age, fill = male)) +
  geom_histogram(binwidth = 5, boundary = 0) +
  scale_fill_manual(values=colorPalette)
```



Black

Problem 3.4

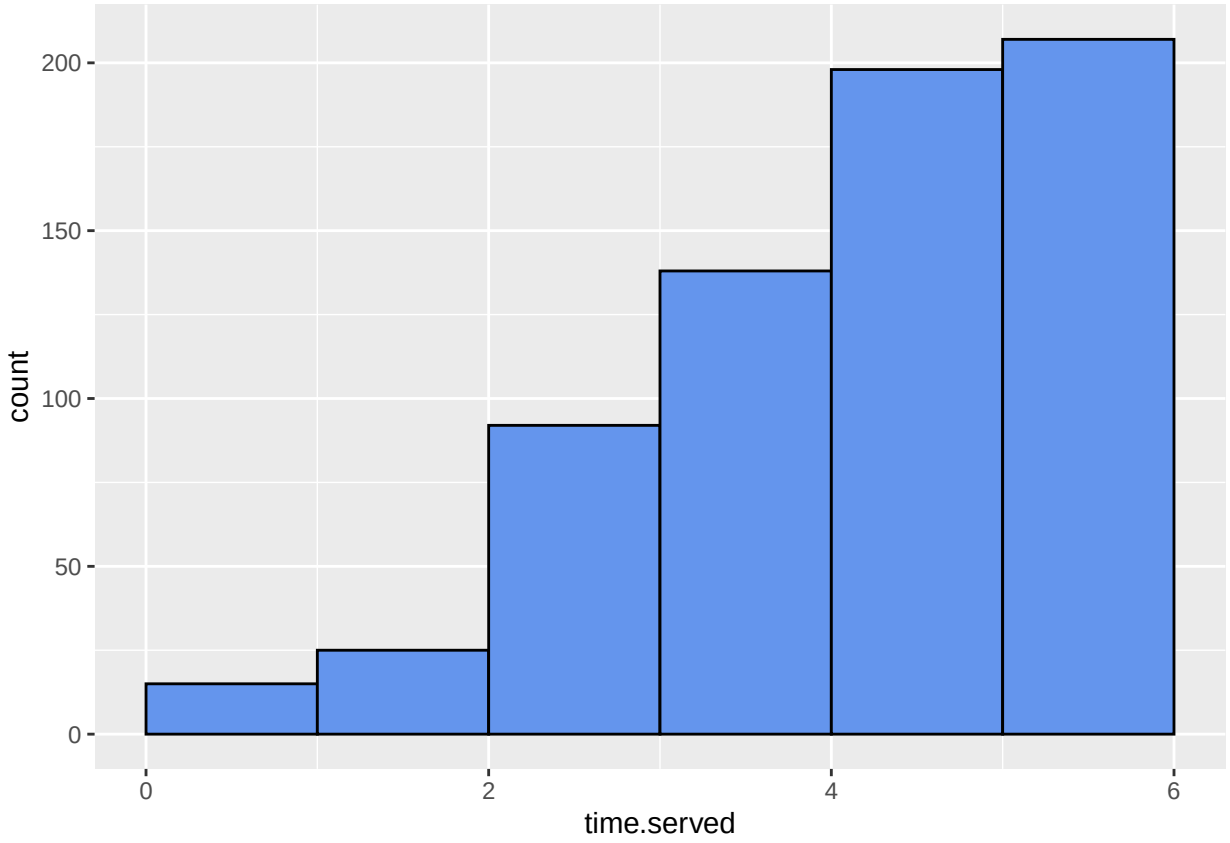
```
ggplot(data = parole, aes(x = age, fill = male)) +
  geom_histogram(binwidth = 5, boundary = 0, position = "identity", alpha = 0.5) +
  scale_fill_manual(values=colorPalette)
```



15-19, 55-59, 65-69

Problem 4.1

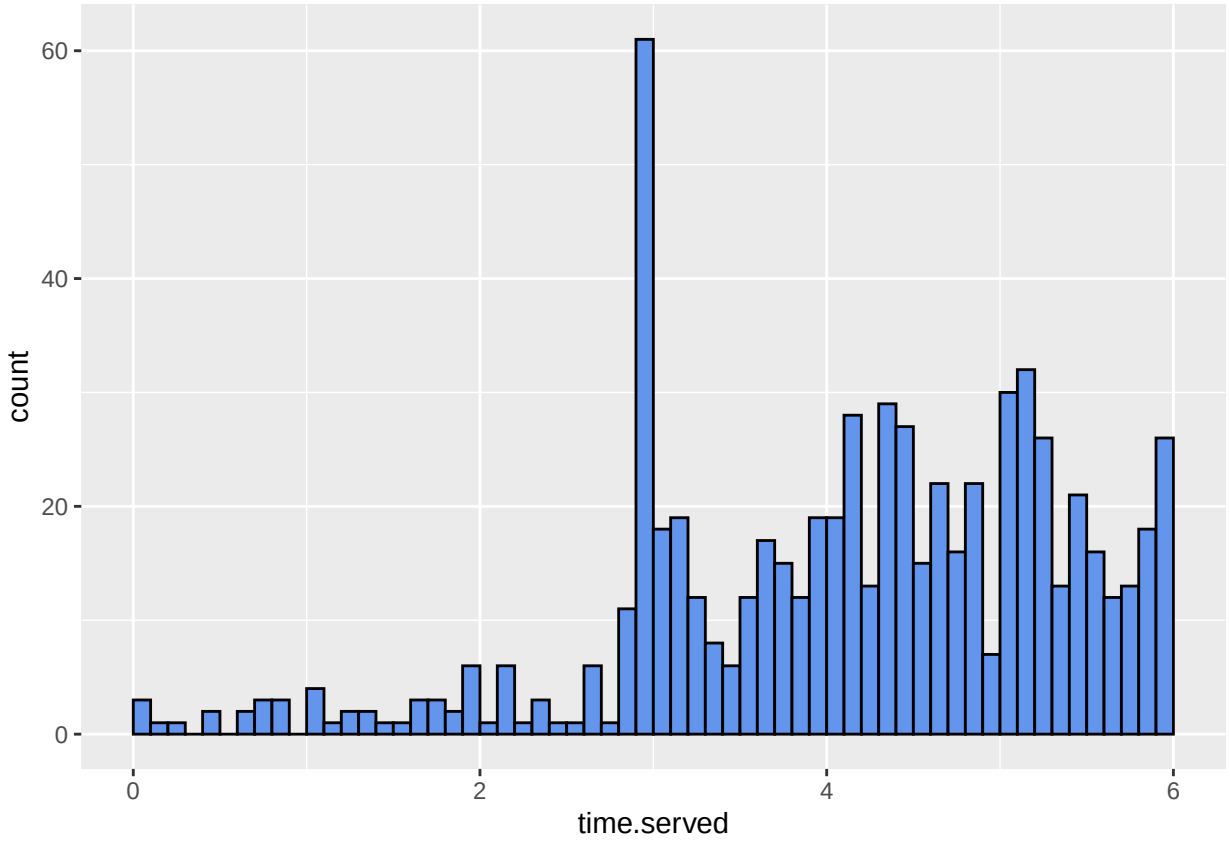
```
ggplot(data = parole, aes(x = time.served)) +
  geom_histogram(binwidth = 1, boundary = 0, color = 'black', fill = 'cornflowerblue')
```



Between 5 and 6 months

Problem 4.2

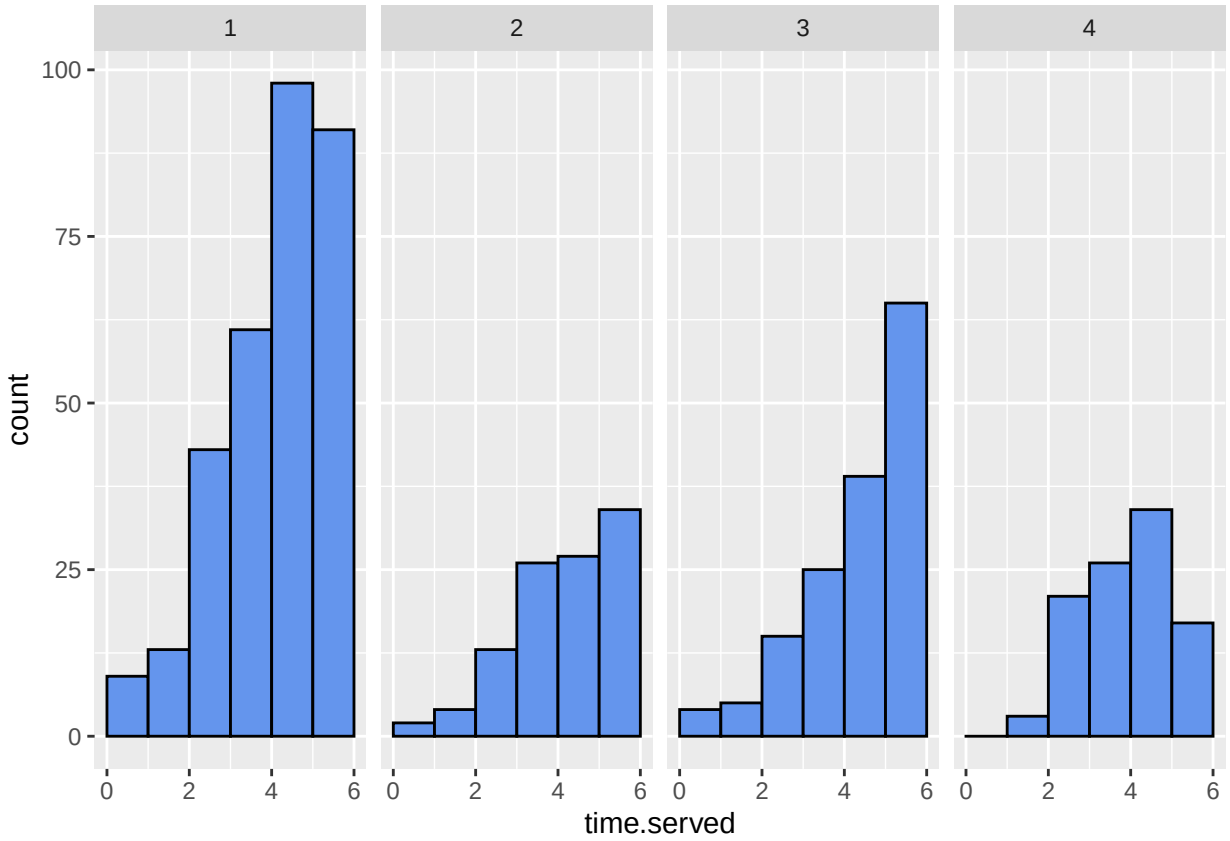
```
ggplot(data = parole, aes(x = time.served)) +
  geom_histogram(binwidth = 0.1, boundary = 0, color = 'black', fill = 'cornflowerblue')
```



Between 2.9 and 3.0 months
Be careful when choosing the binwidth - it can significantly affect the interpretation of a histogram! When visualizing histograms, it is always a good idea to vary the bin size in order to understand the data at various granularities.

Problem 4.3

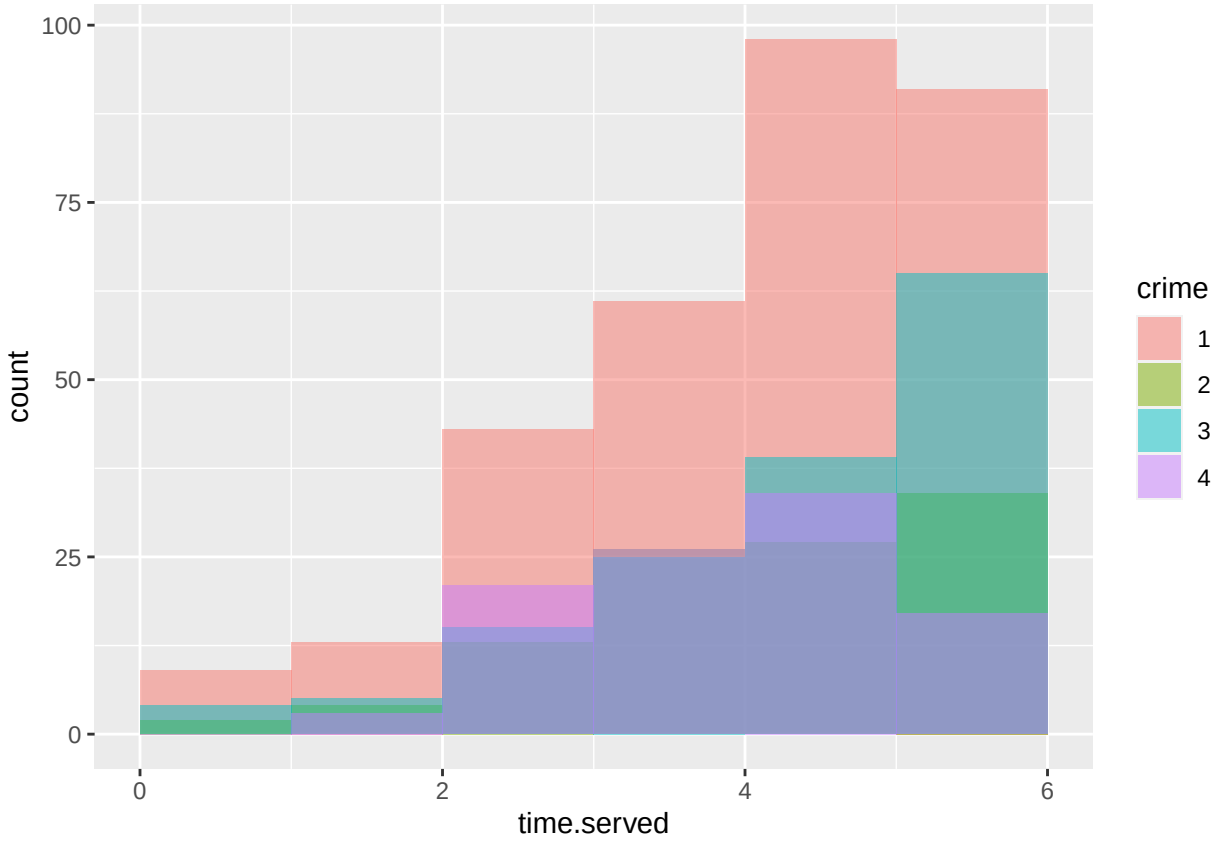
```
ggplot(data = parole, aes(x = time.served)) +
  geom_histogram(binwidth = 1, boundary = 0, color = 'black', fill = 'cornflowerblue') +
  facet_grid(. ~ crime)
```



Driving-related, Drug-related correct

Problem 4.4

```
ggplot(data = parole, aes(x = time.served, fill = crime)) +
  geom_histogram(binwidth = 1, boundary = 0, position = "identity", alpha = 0.5)
```



With four different groups, it can be hard to tell them apart when they are overlayed.