Homework 2

Benjamin Anderson II

```
library(tidyverse)
```

Tasks that require an answer are bolded (inside ** in the .Rmd file). For any task that includes a question (i.e. it ends with "?"), you should also answer the question in sentence form.

Looking at Data

The following tasks all relate to the dataset `starwars` that comes with the dplyr package. Since, dplyr is included in the tidyverse, you don't need to import this data, just type its name to see it:

```
starwars
```

```
# A tibble: 87 x 14
   name     height  mass hair_color skin_color eye_color birth_year sex   gender
   <chr>     <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
 1 Luke Sk~    172    77 blond      fair       blue            19   male  mascu~
 2 C-3PO       167    75 <NA>       gold       yellow         112   none  mascu~
 3 R2-D2        96    32 <NA>       white, bl~ red             33   none  mascu~
 4 Darth V~    202   136 none       white      yellow          41.9 male  mascu~
 5 Leia Or~    150    49 brown      light      brown           19   fema~ femin~
 6 Owen La~    178   120 brown, gr~ light      blue            52   male  mascu~
 7 Beru Wh~    165    75 brown      light      blue            47   fema~ femin~
 8 R5-D4        97    32 <NA>       white, red red             NA   none  mascu~
 9 Biggs D~    183    84 black      light      brown           24   male  mascu~
10 Obi-Wan~    182    77 auburn, w~ fair       blue-gray       57   male  mascu~
# i 77 more rows
# i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#   vehicles <list>, starships <list>
```

It contains attributes of the characters in *some* of the Star Wars films. You can find out more about the variables by examining the help page:

```
?starwars
```

1.

How many rows and columns does the dataset `starwars` have? *(Use the output from the above chunk to answer this question)* (1pt)

```
nrow(starwars)
```

```
[1] 87
```

```
ncol(starwars)
```

```
[1] 14
```

There are 87 rows and 14 columns.

2.

The following code extracts the characters from the starwars data are not human and saves the result to a variable called `not_humans`.

```
not_humans <- filter(starwars, species != "Human")
```

How many non-human characters are in the dataset? (2pts) (*Your answer should include code, relevant output, and a complete sentence answer to the question.*)

```
paste("There are", nrow(not_humans), "non-humans in the data set.", sep = " ")
```

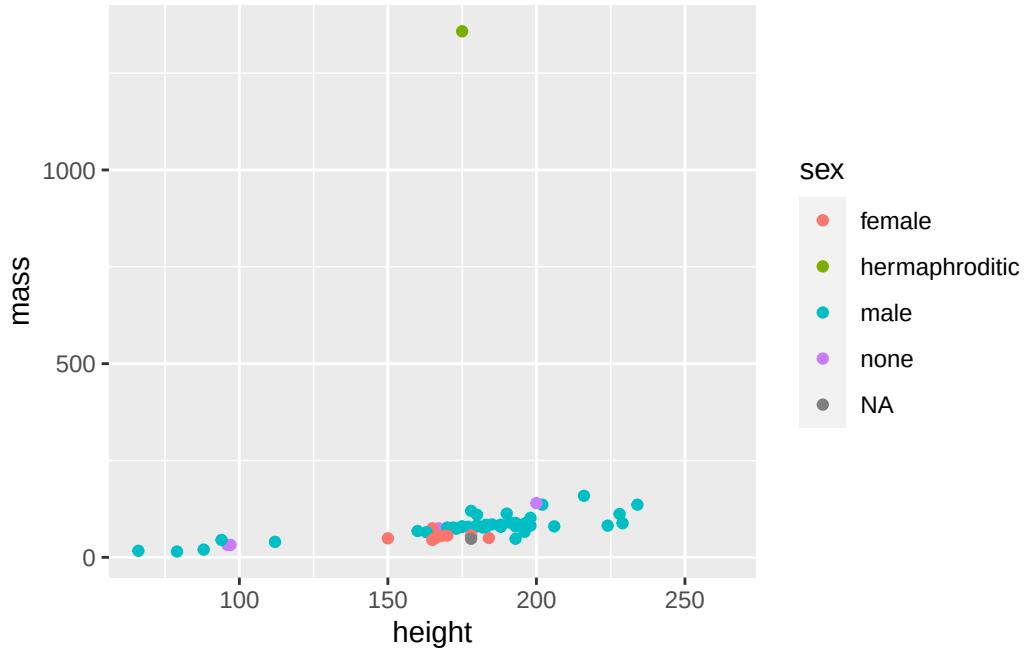
```
[1] "There are 48 non-humans in the data set."
```

3.

Create a scatterplot of all (human and non-human) characters' mass against their height, using color to represent sex. (2pts)

```
ggplot(data = starwars, mapping = aes(x = height, y = mass)) +
  geom_point(mapping = aes(color = sex))
```

```
Warning: Removed 28 rows containing missing values (`geom_point()`).
```

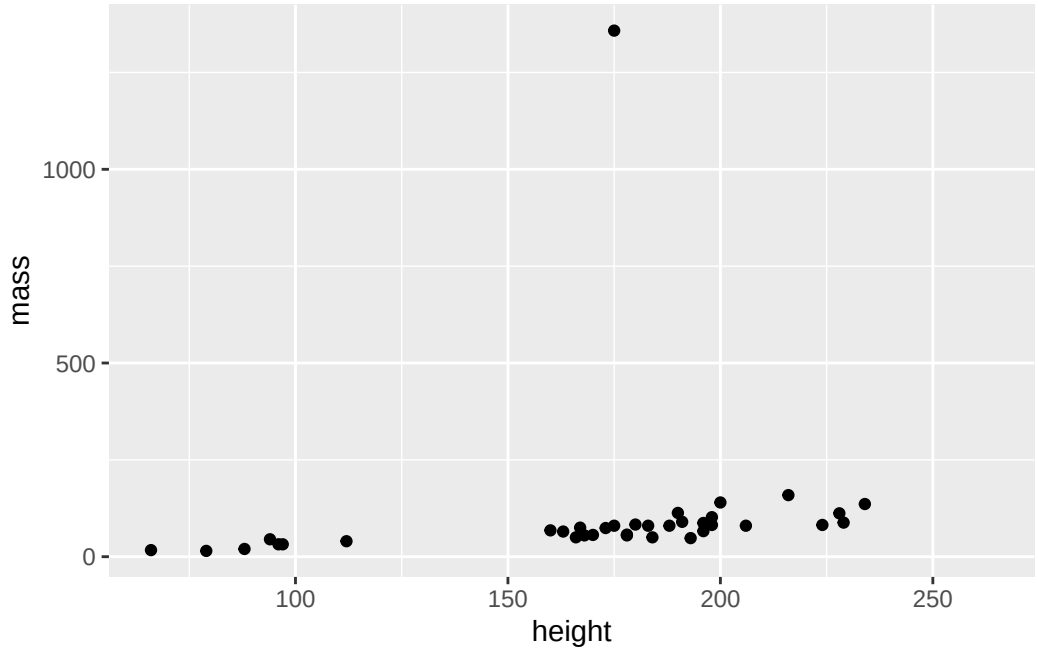


4.

Make a scatter plot of the non-human characters mass against their height. (1pt)

```
ggplot(data = not_humans, mapping = aes(x = height, y = mass)) +
  geom_point()
```

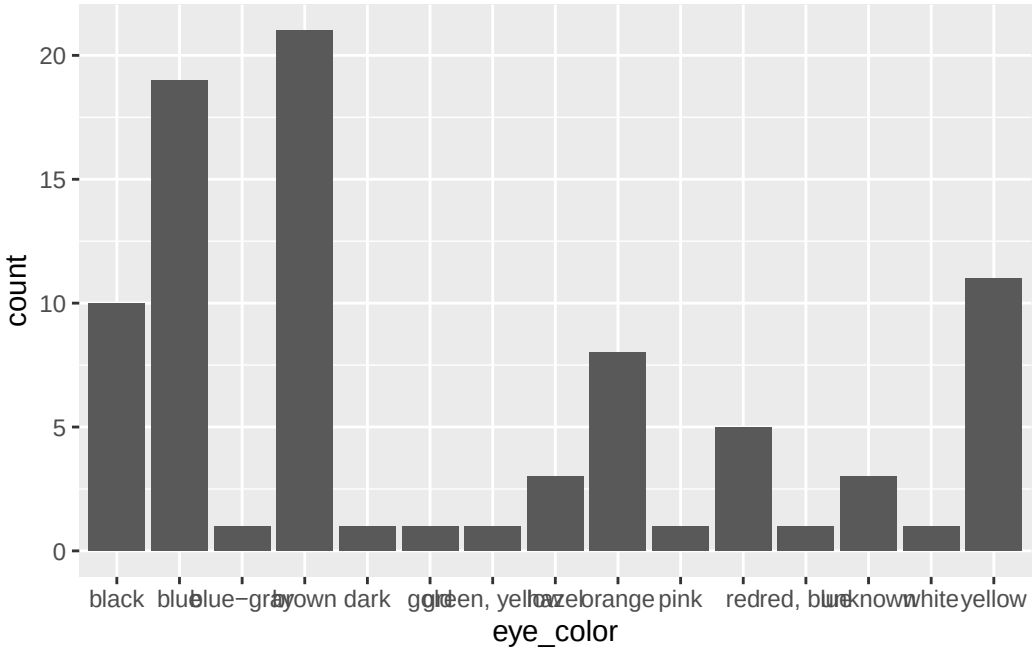
```
Warning: Removed 12 rows containing missing values (`geom_point()`).
```



5.

Make a barchart of all (human and non-human) characters' eye colors. (1pt)

```
ggplot(data = starwars, mapping = aes(x = eye_color)) +
  geom_bar()
```



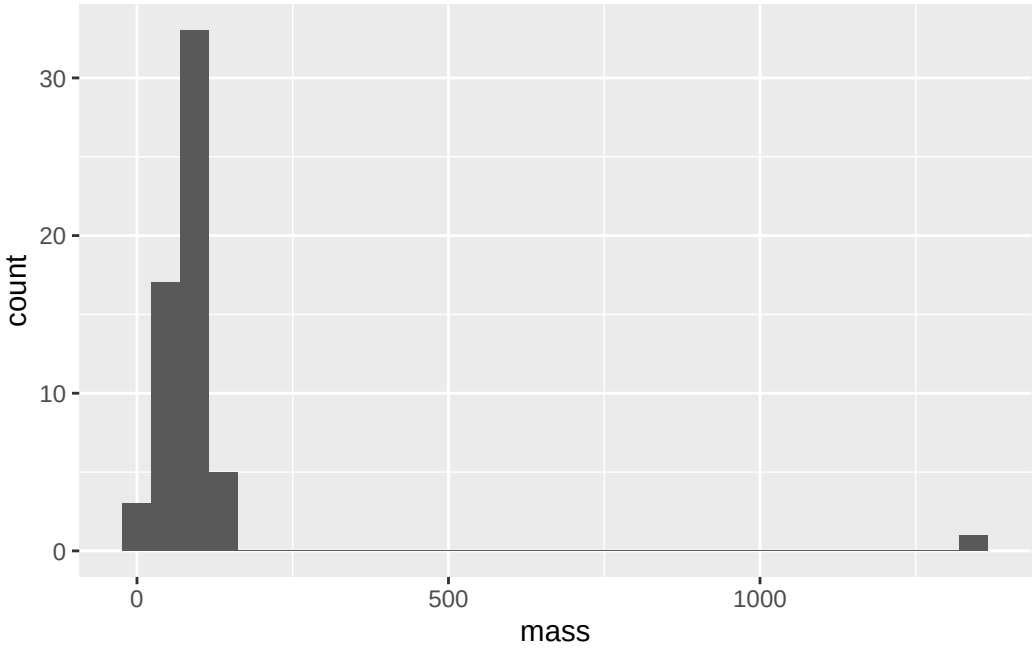
6.

Make a histogram of all (human and non-human) character's mass. (1pt)

```
ggplot(data = starwars, mapping = aes(x = mass)) +
  geom_histogram()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
Warning: Removed 28 rows containing non-finite values (`stat_bin()`).
```



(2pts) For correct `author` in header, and submitting both PDF and Quarto files.