STA_445_Assignment 7

Paige Hawkinson

Sys.Date

Load your packages here:

Problem 1:

The `infmort` data set from the package `faraway` gives the infant mortality rate for a variety of countries. The information is relatively out of date, but will be fun to graph. Visualize the data using by creating scatter plots of mortality vs income while faceting using `region` and setting color by `oil` export status. Utilize a $\log_{10}$ transformation for both `mortality` and `income` axes. This can be done either by doing the transformation inside the `aes()` command or by utilizing the `scale_x_log10()` or `scale_y_log10()` layers. The critical difference is if the scales are on the original vs log transformed scale. Experiment with both and see which you prefer.

a. The `rownames()` of the table gives the country names and you should create a new column that contains the country names. *`rownames`

```
rownames(infmort)
```

```
##   [1] "Australia          " "Austria            " "Belgium            "
##   [4] "Canada             " "Denmark            " "Finland            "
##   [7] "France             " "West_Germany       " "Ireland            "
##  [10] "Italy              " "Japan              " "Netherlands        "
##  [13] "New_Zealand        " "Norway             " "Portugal           "
##  [16] "South_Africa       " "Sweden             " "Switzerland        "
##  [19] "Britain            " "United_States      " "Algeria            "
##  [22] "Ecuador            " "Indonesia          " "Iran               "
##  [25] "Iraq               " "Libya              " "Nigeria            "
##  [28] "Saudi_Arabia       " "Venezuela          " "Argentina          "
##  [31] "Brazil             " "Chile             " "Colombia           "
##  [34] "Costa_Rica         " "Dominican_Republic " "Greece             "
##  [37] "Guatemala          " "Israel            " "Jamaica            "
##  [40] "Lebanon            " "Malaysia          " "Mexico             "
##  [43] "Nicaragua          " "Panama            " "Peru               "
##  [46] "Singapore          " "Spain             " "Taiwan             "
##  [49] "Trinidad_and_Tobago" "Tunisia           " "Uruguay            "
##  [52] "Yugoslavia         " "Zambia            " "Bolivia            "
##  [55] "Cameroon           " "Congo             " "Egypt              "
##  [58] "El_Salvador        " "Ghana             " "Honduras           "
##  [61] "Ivory_Coast        " "Jordan            " "South_Korea        "
##  [64] "Liberia            " "Moroco            " "Papua_New_Guinea   "
##  [67] "Paraguay           " "Philippines       " "Syria              "
```

```
##  [70] "Thailand           " "Turkey             " "South_Vietnam      "
##  [73] "Afganistan         " "Bangladesh        " "Burma              "
##  [76] "Burundi            " "Cambodia          " "Central_African_Rep"
##  [79] "Chad               " "Dahomey           " "Ethiopia           "
##  [82] "Guinea             " "Haiti             " "India              "
##  [85] "Kenya              " "Laos              " "Madagascar         "
##  [88] "Malawi             " "Mali              " "Mauritania         "
##  [91] "Nepal              " "Niger             " "Pakistan           "
##  [94] "Rwanda             " "Sierra_Leone      " "Somalia            "
##  [97] "Sri_Lanka          " "Sudan             " "Tanzania           "
## [100] "Togo               " "Uganda            " "Upper_Volta        "
## [103] "Southern_Yemen     " "Yemen              " "Zaire              "
```

```
infmort.cont <- infmort %>%
  mutate(Country = rownames(infmort))
infmort.cont
```

```
##                       region income mortality           oil
## Australia               Asia   3426      26.7 no oil exports
## Austria               Europe   3350      23.7 no oil exports
## Belgium               Europe   3346      17.0 no oil exports
## Canada              Americas   4751      16.8 no oil exports
## Denmark               Europe   5029      13.5 no oil exports
## Finland               Europe   3312      10.1 no oil exports
## France                Europe   3403      12.9 no oil exports
## West_Germany          Europe   5040      20.4 no oil exports
## Ireland               Europe   2009      17.8 no oil exports
## Italy                 Europe   2298      25.7 no oil exports
## Japan                 Europe   3292      11.7 no oil exports
## Netherlands           Europe   4103      11.6 no oil exports
## New_Zealand             Asia   3723      16.2 no oil exports
## Norway                Europe   4102      11.3 no oil exports
## Portugal              Europe    956      44.8 no oil exports
## South_Africa          Africa   1000      71.5 no oil exports
## Sweden                Europe   5596       9.6 no oil exports
## Switzerland           Europe   2963      12.8 no oil exports
## Britain               Europe   2503      17.5 no oil exports
## United_States       Americas   5523      17.6 no oil exports
## Algeria               Africa    400      86.3    oil exports
## Ecuador             Americas    250      78.5    oil exports
## Indonesia               Asia    110     125.0    oil exports
## Iran                    Asia   1280        NA    oil exports
## Iraq                    Asia    560      28.1    oil exports
## Libya                 Africa   3010     300.0    oil exports
## Nigeria               Africa    220      58.0    oil exports
## Saudi_Arabia            Asia   1530     650.0    oil exports
## Venezuela           Americas   1240      51.7    oil exports
## Argentina           Americas   1191      59.6 no oil exports
## Brazil              Americas    425     170.0 no oil exports
## Chile               Americas    590      78.0 no oil exports
## Colombia            Americas    426      62.8 no oil exports
## Costa_Rica          Americas    725      54.4 no oil exports
## Dominican_Republic  Americas    406      48.8 no oil exports
## Greece                Europe   1760      27.8 no oil exports
```

```
## Guatemala           Americas    302      79.1 no oil exports
## Israel                 Asia    2526      22.1 no oil exports
## Jamaica             Americas    727      26.2 no oil exports
## Lebanon                Asia    631      13.6 no oil exports
## Malaysia               Asia    295      32.0 no oil exports
## Mexico              Americas    684      60.9 no oil exports
## Nicaragua           Americas    507      46.0 no oil exports
## Panama              Americas    754      34.1 no oil exports
## Peru                Americas    335      65.1 no oil exports
## Singapore              Asia    1268      20.4 no oil exports
## Spain                Europe    1256      15.1 no oil exports
## Taiwan                 Asia    261      19.1 no oil exports
## Trinidad_and_Tobago Americas    732      26.2 no oil exports
## Tunisia              Africa    434      76.3 no oil exports
## Uruguay             Americas    799      40.4 no oil exports
## Yugoslavia           Europe    406      43.3 no oil exports
## Zambia               Africa    310     259.0 no oil exports
## Bolivia             Americas    200      60.4 no oil exports
## Cameroon             Africa    100     137.0 no oil exports
## Congo                Africa    281     180.0 no oil exports
## Egypt                Africa    210     114.0 no oil exports
## El_Salvador         Americas    319      58.2 no oil exports
## Ghana                Africa    217      63.7 no oil exports
## Honduras            Americas    284      39.3 no oil exports
## Ivory_Coast          Africa    387     138.0 no oil exports
## Jordan                 Asia    334      21.3 no oil exports
## South_Korea            Asia    344      58.0 no oil exports
## Liberia              Africa    197     159.2 no oil exports
## Moroco               Africa    279     149.0 no oil exports
## Papua_New_Guinea       Asia    477      10.2 no oil exports
## Paraguay            Americas    347      38.6 no oil exports
## Philippines            Asia    230      67.9 no oil exports
## Syria                  Asia    334      21.7 no oil exports
## Thailand               Asia    210      27.0 no oil exports
## Turkey                 Asia    435     153.0 no oil exports
## South_Vietnam          Asia    130     100.0 no oil exports
## Afganistan             Asia     75     400.0 no oil exports
## Bangladesh             Asia    100     124.3 no oil exports
## Burma                  Asia     73     200.0 no oil exports
## Burundi              Africa     68     150.0 no oil exports
## Cambodia               Asia    123     100.0 no oil exports
## Central_African_Rep  Africa    122     190.0 no oil exports
## Chad                 Africa     70     160.0 no oil exports
## Dahomey              Africa     81     109.6 no oil exports
## Ethiopia             Africa     79      84.2 no oil exports
## Guinea               Africa     79     216.0 no oil exports
## Haiti               Americas    100        NA no oil exports
## India                  Asia     93      60.6 no oil exports
## Kenya                Africa    169      55.0 no oil exports
## Laos                   Asia     71        NA no oil exports
## Madagascar           Africa    120     102.0 no oil exports
## Malawi               Africa    130     148.3 no oil exports
## Mali                 Africa     50     120.0 no oil exports
## Mauritania           Africa    174     187.0 no oil exports
```

```
## Nepal                   Asia     90        NA no oil exports
## Niger                  Africa     70     200.0 no oil exports
## Pakistan                Asia    102     124.3 no oil exports
## Rwanda                 Africa     61     132.9 no oil exports
## Sierra_Leone           Africa    148     170.0 no oil exports
## Somalia                Africa     85     158.0 no oil exports
## Sri_Lanka               Asia    162      45.1 no oil exports
## Sudan                  Africa    125     129.4 no oil exports
## Tanzania               Africa    120     162.5 no oil exports
## Togo                   Africa    160     127.0 no oil exports
## Uganda                 Africa    134     160.0 no oil exports
## Upper_Volta            Africa     82     180.0 no oil exports
## Southern_Yemen          Asia     96      80.0 no oil exports
## Yemen                   Asia     77      50.0 no oil exports
## Zaire                  Africa    118     104.0 no oil exports
##                                 Country
## Australia           Australia
## Austria             Austria
## Belgium             Belgium
## Canada              Canada
## Denmark             Denmark
## Finland             Finland
## France              France
## West_Germany        West_Germany
## Ireland             Ireland
## Italy               Italy
## Japan               Japan
## Netherlands         Netherlands
## New_Zealand         New_Zealand
## Norway              Norway
## Portugal            Portugal
## South_Africa        South_Africa
## Sweden              Sweden
## Switzerland         Switzerland
## Britain             Britain
## United_States       United_States
## Algeria             Algeria
## Ecuador             Ecuador
## Indonesia           Indonesia
## Iran                Iran
## Iraq                Iraq
## Libya               Libya
## Nigeria             Nigeria
## Saudi_Arabia        Saudi_Arabia
## Venezuela           Venezuela
## Argentina           Argentina
## Brazil              Brazil
## Chile               Chile
## Colombia            Colombia
## Costa_Rica          Costa_Rica
## Dominican_Republic  Dominican_Republic
## Greece              Greece
## Guatemala           Guatemala
## Israel              Israel
```

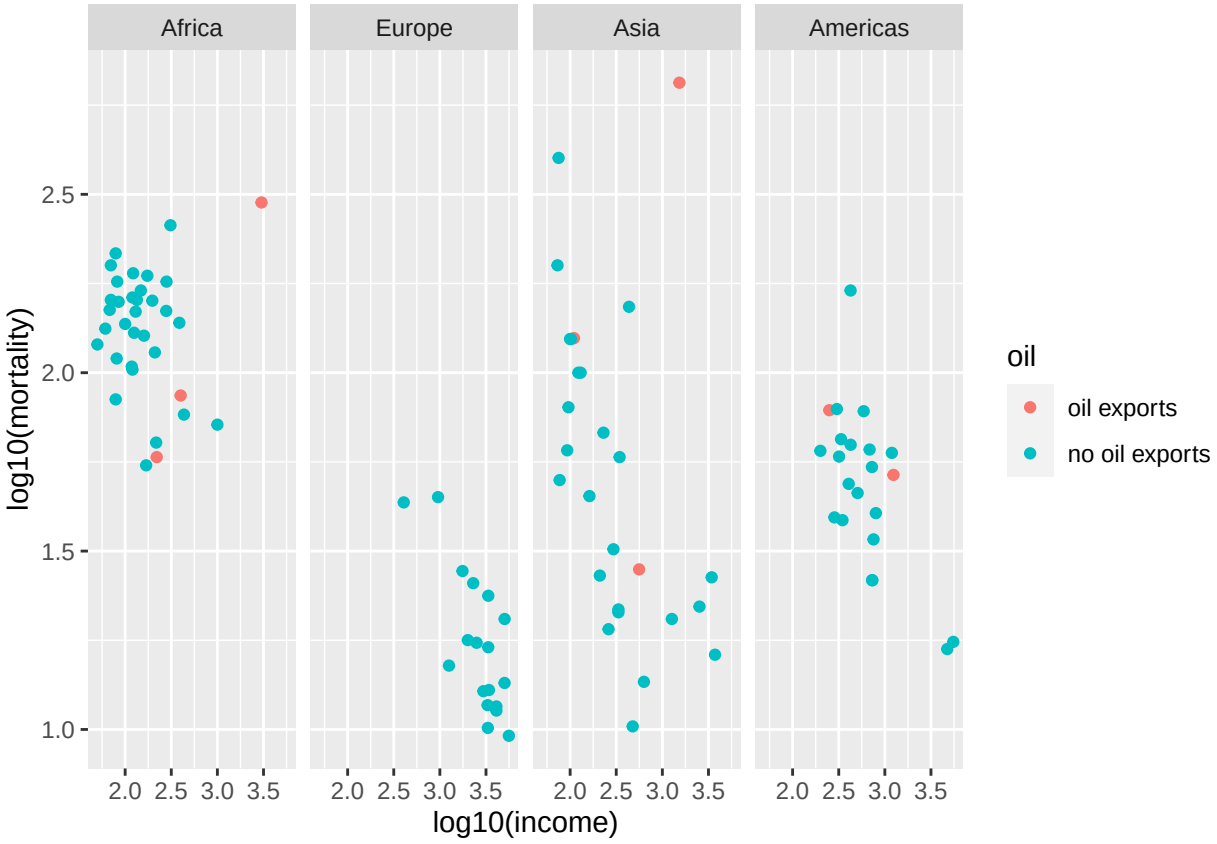
```
## Jamaica              Jamaica
## Lebanon              Lebanon
## Malaysia             Malaysia
## Mexico               Mexico
## Nicaragua            Nicaragua
## Panama               Panama
## Peru                 Peru
## Singapore            Singapore
## Spain                Spain
## Taiwan               Taiwan
## Trinidad_and_Tobago Trinidad_and_Tobago
## Tunisia              Tunisia
## Uruguay              Uruguay
## Yugoslavia           Yugoslavia
## Zambia               Zambia
## Bolivia              Bolivia
## Cameroon             Cameroon
## Congo                Congo
## Egypt                Egypt
## El_Salvador          El_Salvador
## Ghana                Ghana
## Honduras             Honduras
## Ivory_Coast          Ivory_Coast
## Jordan               Jordan
## South_Korea          South_Korea
## Liberia              Liberia
## Moroco               Moroco
## Papua_New_Guinea     Papua_New_Guinea
## Paraguay             Paraguay
## Philippines          Philippines
## Syria                Syria
## Thailand             Thailand
## Turkey               Turkey
## South_Vietnam        South_Vietnam
## Afganistan           Afganistan
## Bangladesh           Bangladesh
## Burma                Burma
## Burundi              Burundi
## Cambodia             Cambodia
## Central_African_Rep Central_African_Rep
## Chad                 Chad
## Dahomey              Dahomey
## Ethiopia             Ethiopia
## Guinea               Guinea
## Haiti                Haiti
## India                India
## Kenya                Kenya
## Laos                 Laos
## Madagascar           Madagascar
## Malawi               Malawi
## Mali                 Mali
## Mauritania           Mauritania
## Nepal                Nepal
## Niger                Niger
```

```
## Pakistan            Pakistan
## Rwanda              Rwanda
## Sierra_Leone        Sierra_Leone
## Somalia             Somalia
## Sri_Lanka           Sri_Lanka
## Sudan               Sudan
## Tanzania            Tanzania
## Togo                Togo
## Uganda              Uganda
## Upper_Volta         Upper_Volta
## Southern_Yemen      Southern_Yemen
## Yemen               Yemen
## Zaire               Zaire
```

b. Create scatter plots with the `log10()` transformation inside the `aes()`command.

```
ggplot(data=infmort.cont, aes( x = log10(income), y = log10(mortality), color=oil)) +
  geom_point() +
  facet_grid(.~region)
```

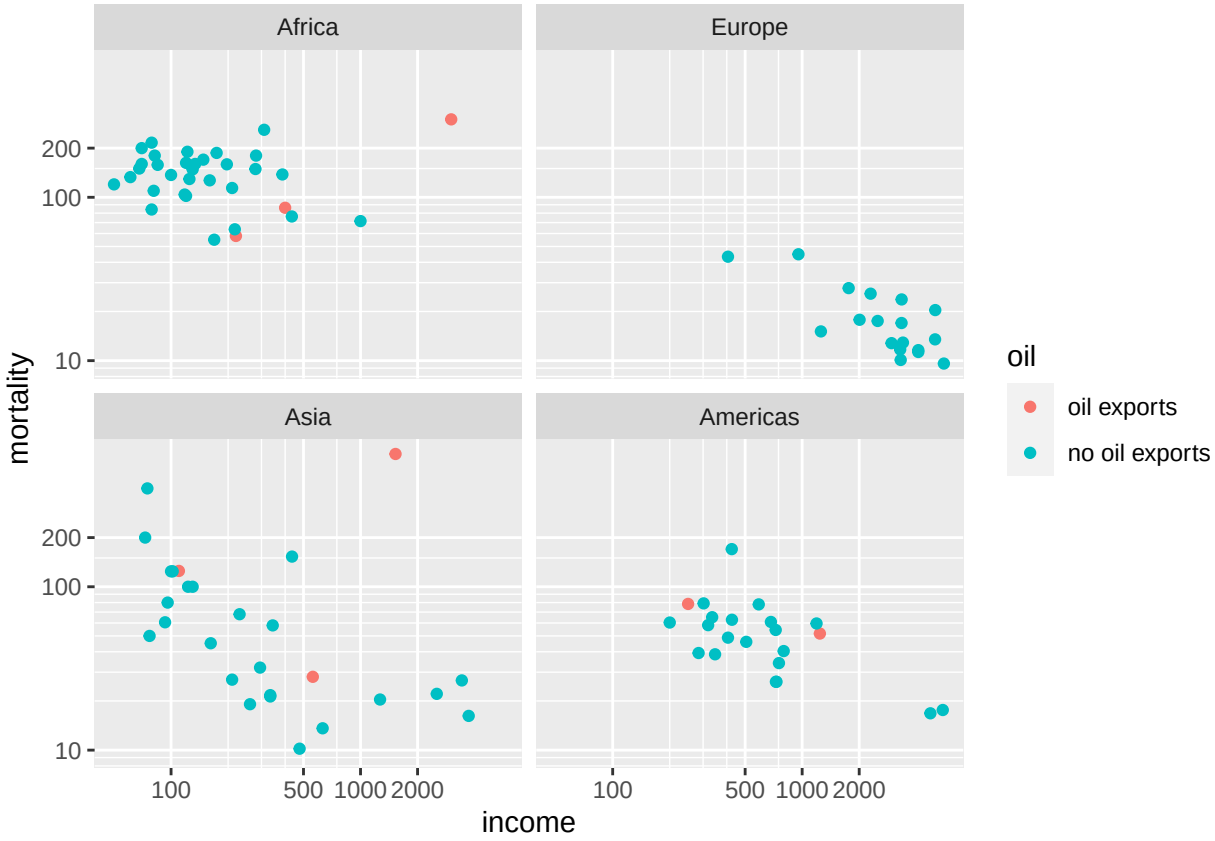
```
## Warning: Removed 4 rows containing missing values (‘geom_point()‘).
```



c. Create the scatter plots using the `scale_x_log10()` and `scale_y_log10()`.Set the major and minor breaks to be useful and aesthetically pleasing. Comment on which version you find easier to read.

```
ggplot(data=infmort.cont, aes(x=income, y=mortality, color=oil)) +
    geom_point() +
    facet_wrap(.~region) +
    scale_x_log10(breaks=c(100,500,1000,2000),
                  minor=c(1:10,
                        seq( 100,500,by=100 ),
                        seq(500,1000,by=250),
                        seq(1000,2000,by=1000))) +
    scale_y_log10(breaks=c(1,10,100,200),
                  minor=c(1:10,
                        seq( 10, 100,by=10 ),
                        seq(100,200,by=50)))
```

```
## Warning: Removed 4 rows containing missing values (`geom_point()`).
```



I find the plot from part c easier to read and interpret because the x and y axis are measured the same as the origal variables.

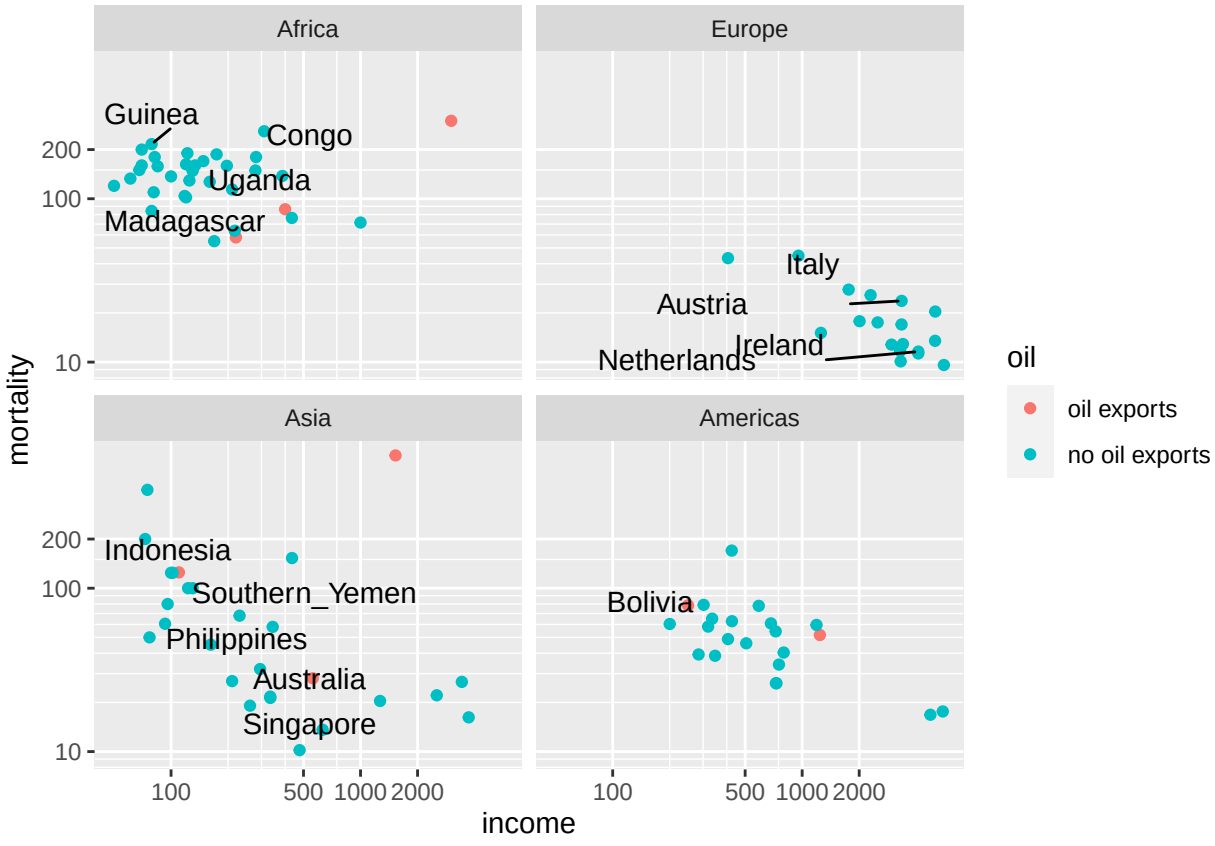
d. The package `ggrepel` contains functions `geom_text_repel()` and `geom_label_repel()` that mimic the basic `geom_text()` and `geom_label()`functions in `ggplot2`, but work to make sure the labels don't overlap. Select 10-15 countries to label and do so using the `geom_text_repel()` function.

```
country.new <- slice_sample(infmort.cont, n=15)
```

```
ggplot(data=infmort.cont) +
    geom_point(aes(x=income, y=mortality, color=oil)) +
    facet_wrap(.~region) +
    scale_x_log10(breaks=c(100,500,1000,2000),
                  minor=c(1:10,
                        seq( 100,500,by=100 ),
                        seq(500,1000,by=250),
                        seq(1000,2000,by=1000))) +
    scale_y_log10(breaks=c(1,10,100,200),
                  minor=c(1:10,
                        seq( 10, 100,by=10 ),
                        seq(100,200,by=50))) +
  geom_text_repel(data=country.new, aes(x = income, y = mortality,
                                label = Country))
```

```
## Warning: Removed 4 rows containing missing values (‘geom_point()‘).
```

```
## Warning: Removed 1 rows containing missing values (‘geom_text_repel()‘).
```



Problem 2

Using the `datasets::trees` data, complete the following:

a. Create a regression model for $y =$ `Volume` as a function of $x =$ `Height`.

```
trees.mod <- lm(data=trees, Volume ~ Height )
```

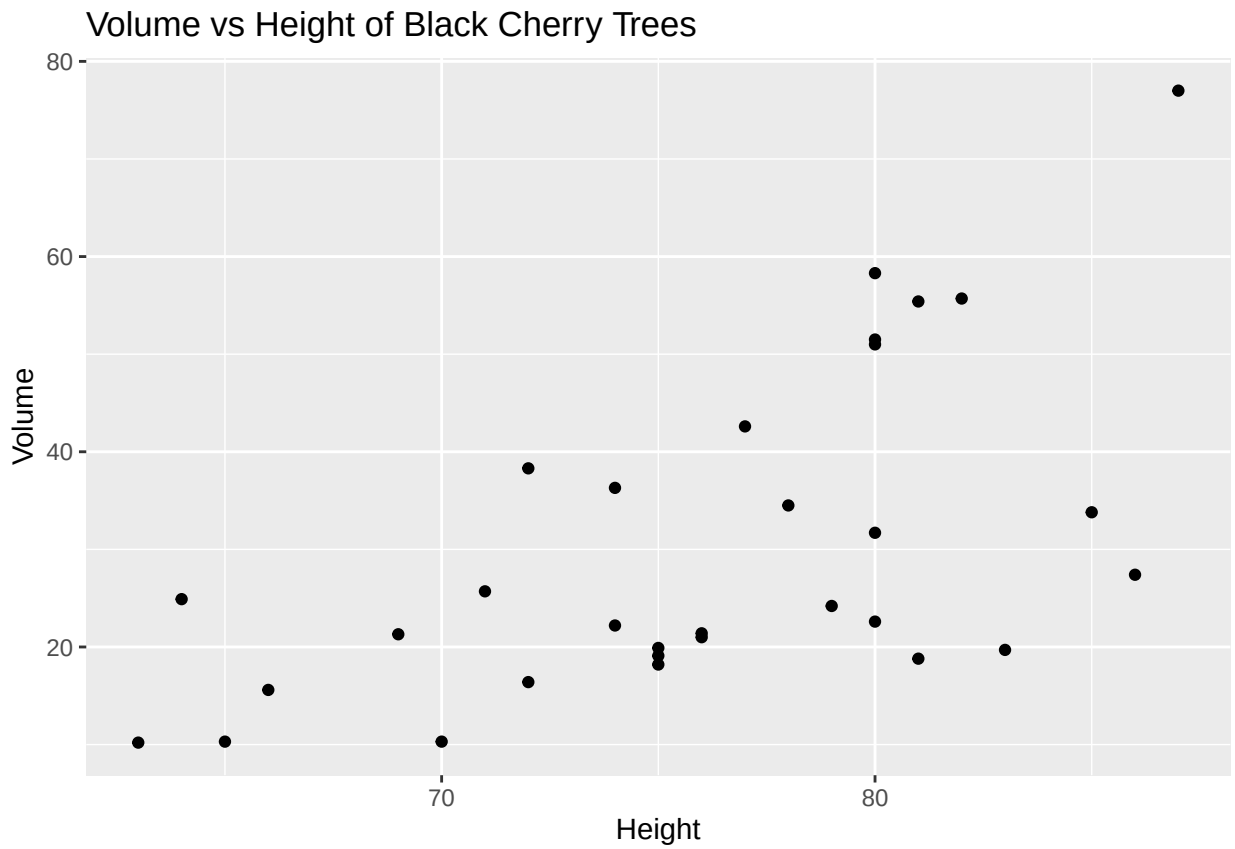
b. Using the str(your model's name) command, to get a list of all the information stored in the linear model object. Use $ to extract the slope and intercept of the regression line (the coefficients).

```
str(trees.mod$coefficients)
```

```
##  Named num [1:2] -87.12 1.54
##  - attr(*, "names")= chr [1:2] "(Intercept)" "Height"
```

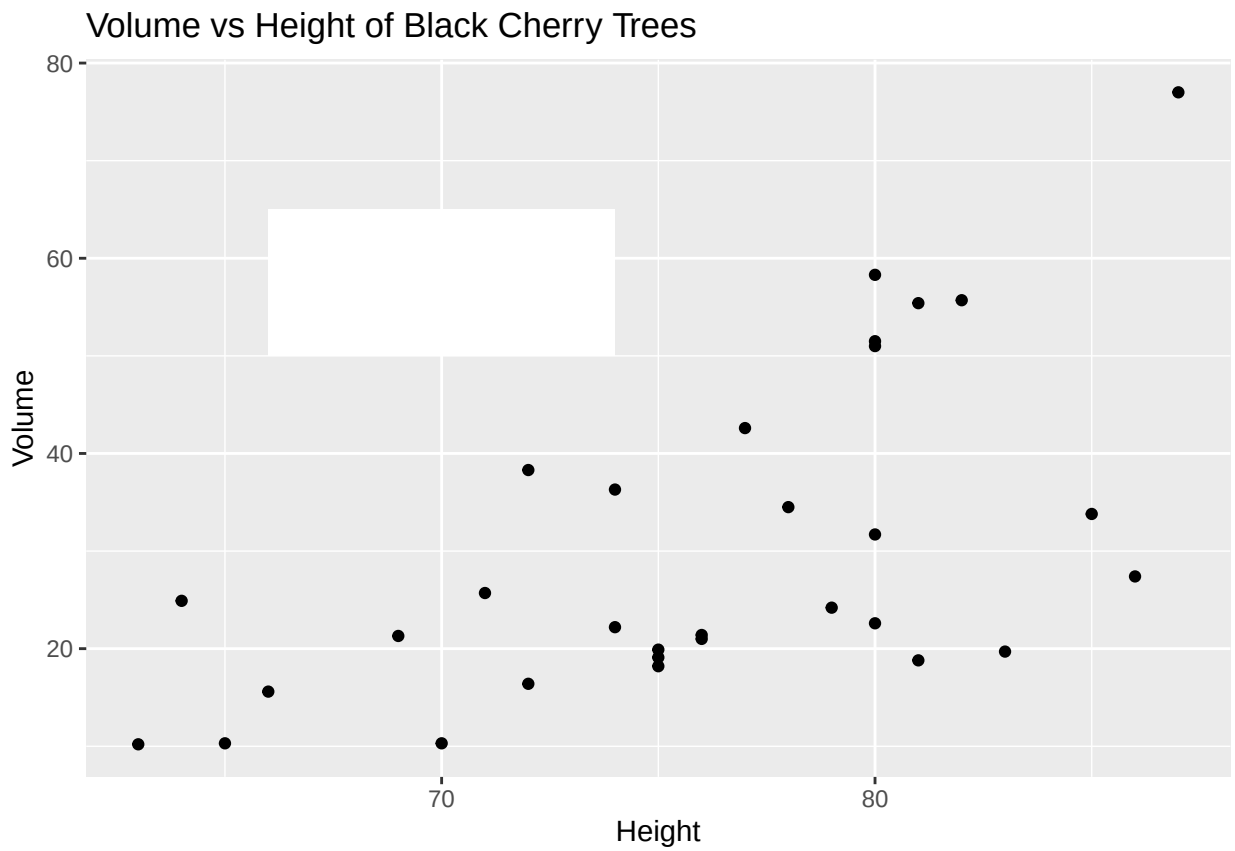
c. Using ggplot2, create a scatter plot of Volume vs Height.

```
ggplot(data=trees, aes(x = Height, y = Volume)) +
  geom_point() +
  labs( title = "Volume vs Height of Black Cherry Trees" )
```



d. Create a nice white filled rectangle to add text information to using by adding the following annotation layer.

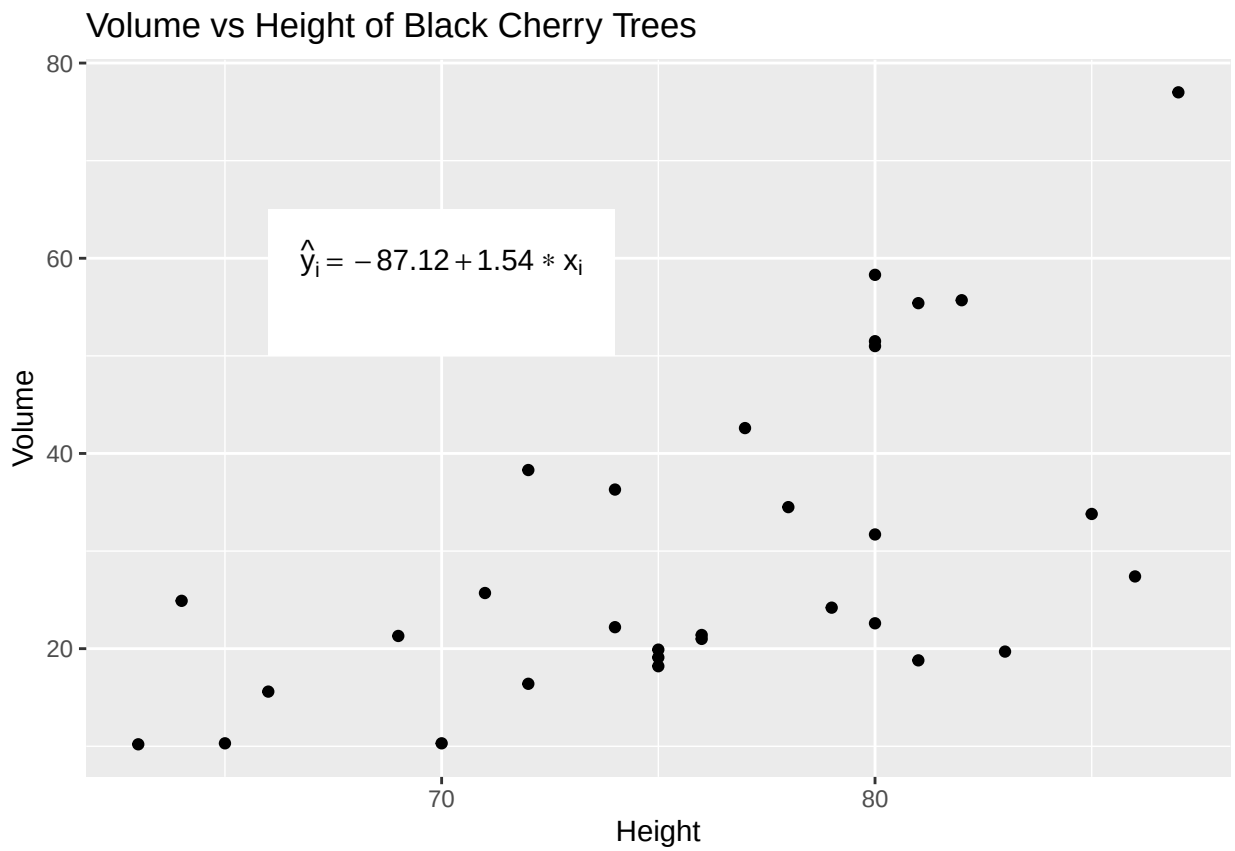
```
ggplot(data=trees, aes(x = Height, y = Volume)) +
  geom_point() +
  labs( title = "Volume vs Height of Black Cherry Trees" ) +
   annotate('rect', xmin = 66, xmax = 74, ymin = 50, ymax = 65, fill = "white")
```



e. Add some annotation text to write the equation of the line $\hat{y}_i = -87.12 + 1.54 * x_i$ in the text area.

```
ggplot(data=trees, aes(x = Height, y = Volume)) +
  geom_point() +
  labs( title = "Volume vs Height of Black Cherry Trees" ) +
  annotate('rect', xmin = 66, xmax = 74, ymin = 50, ymax = 65, fill = "white") +
  annotate('text', x=70.0, y=60,
           label = latex2exp::TeX('$\\hat{y}_i = -87.12 + 1.54 * x_i$'))
```

```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```

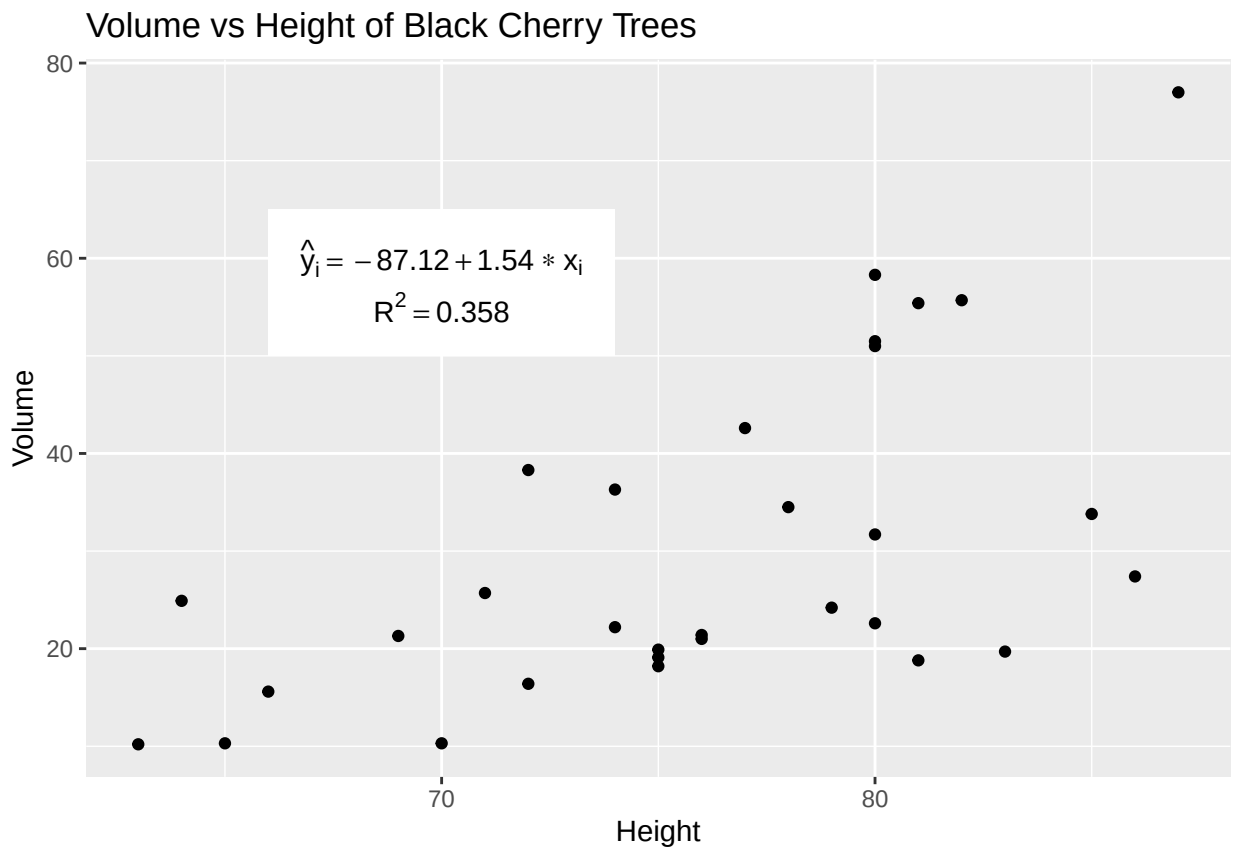


f. Add annotation to add $R^2 = 0.358$

```
ggplot(data=trees, aes(x = Height, y = Volume)) +
  geom_point() +
  labs( title = "Volume vs Height of Black Cherry Trees" ) +
  annotate('rect', xmin = 66, xmax = 74, ymin = 50, ymax = 65, fill = "white") +
   annotate('text', x=70.0, y=60, size=4,
            label = latex2exp::TeX('$\\hat{y}_i = -87.12 + 1.54 * x_i$')) +
  annotate('text', x=70.0, y=55,
           label = latex2exp::TeX('$R^2 = 0.358$'))
```

```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```

```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```

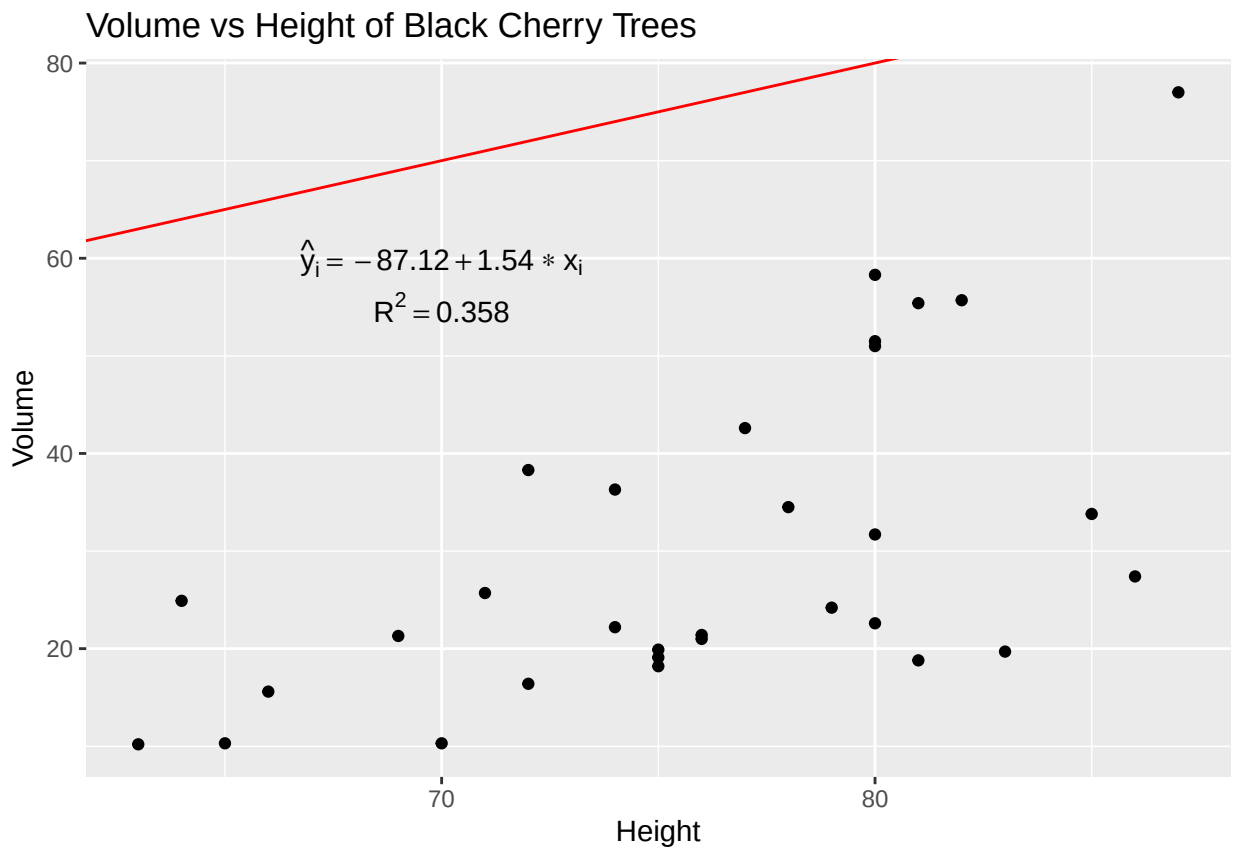


g. Add the regression line in red. The most convenient layer function to use is `geom_abline()`.

```
ggplot(data=trees, aes(x = Height, y = Volume)) +
  geom_point() +
  labs( title = "Volume vs Height of Black Cherry Trees" ) +
   annotate('text', x=70.0, y=60, size=4,
            label = latex2exp::TeX('$\\hat{y}_i = -87.12 + 1.54 * x_i$')) +
  annotate('text', x=70.0, y=55,
           label = latex2exp::TeX('$R^2 = 0.358$')) +
  geom_abline(color='red')
```

```
## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'

## Warning in is.na(x): is.na() applied to non-(list or vector) of type
## 'expression'
```



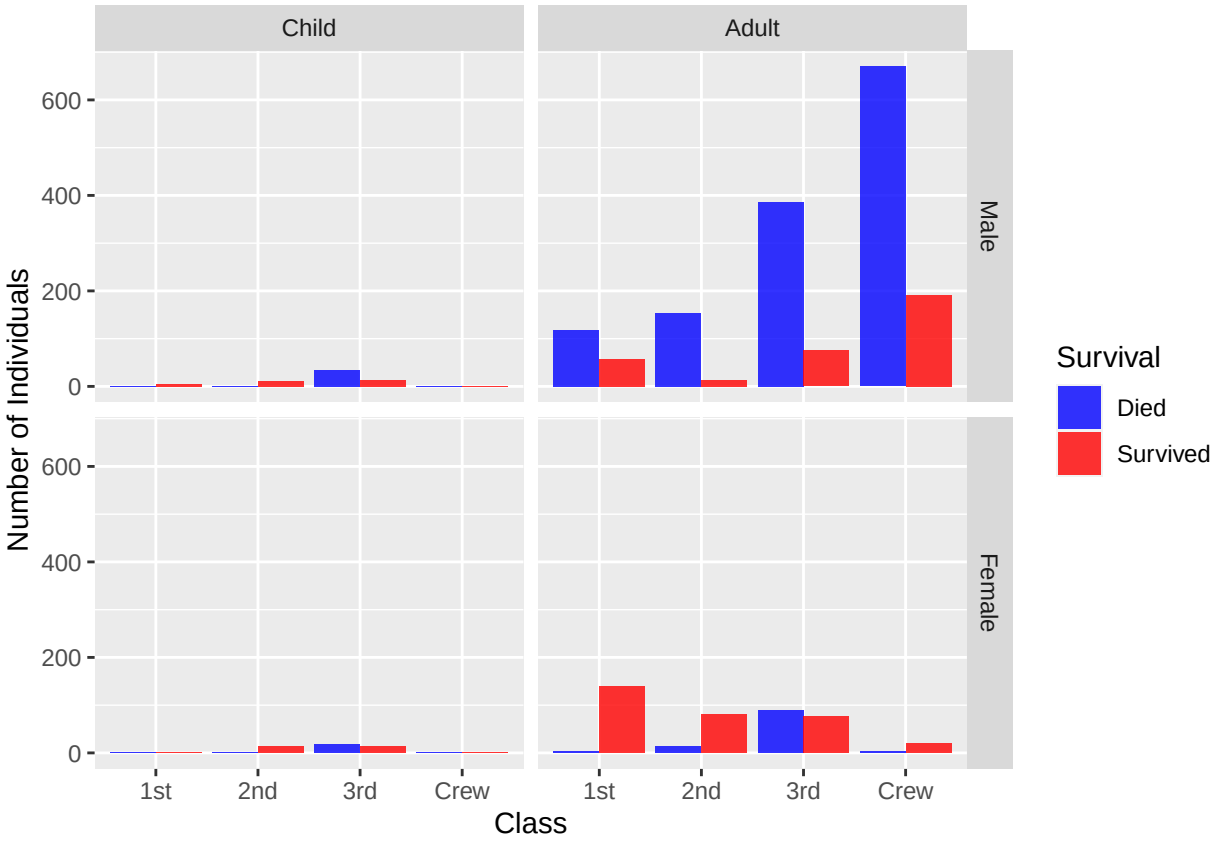
Problem 3

In `datasets::Titanic` table summarizes the survival of passengers aboard the ocean liner *Titanic*. It includes information about passenger class, sex, and age (adult or child). Create a bar graph showing the number of individuals that survived based on the passenger `Class`, `Sex`, and `Age` variable information. You'll need to use faceting and/or color to get all four variables on the same graph. Make sure that differences in survival among different classes of children are perceivable. *Unfortunately, the data is stored as a* `table` *and to expand it to a data frame, the following code can be used.*

```
Titanic <- Titanic %>% as.data.frame()
```

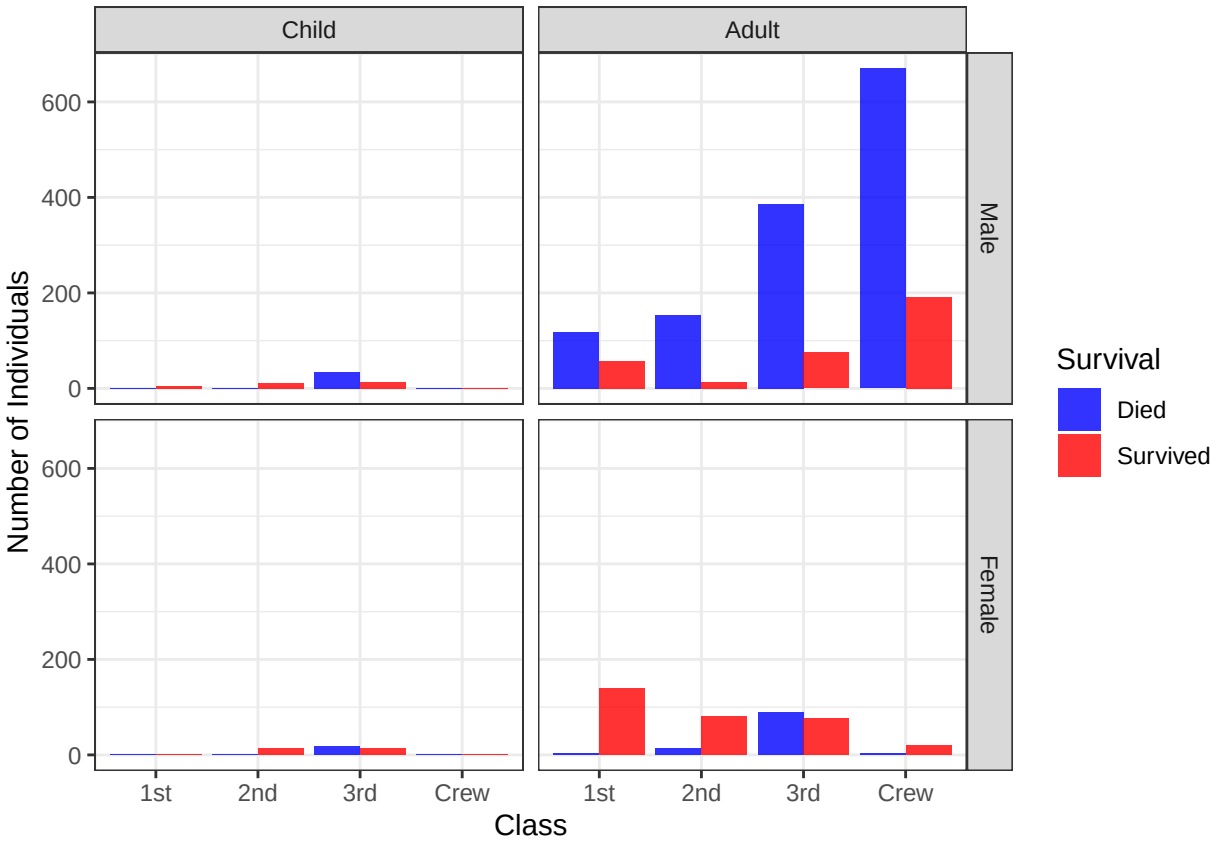
a. Make this graph using the default theme. *If you use color to denote survivorship, modify the color scheme so that a cold color denotes death.*

```
plot.1 <- ggplot(data=Titanic, aes(x = Class)) +
  geom_bar(aes(y = Freq, fill = Survived),
           position = "dodge", alpha = 0.8, stat = "identity") +
  facet_grid(Sex ~ Age) +
  scale_fill_manual(values = c("blue", "red"), labels = c("Died", "Survived")) +
  scale_color_manual(values = c("blue", "red"), labels = c("Died", "Survived")) +
  labs(x = "Class", y = "Number of Individuals", fill = "Survival", color = "Survival")
plot.1
```



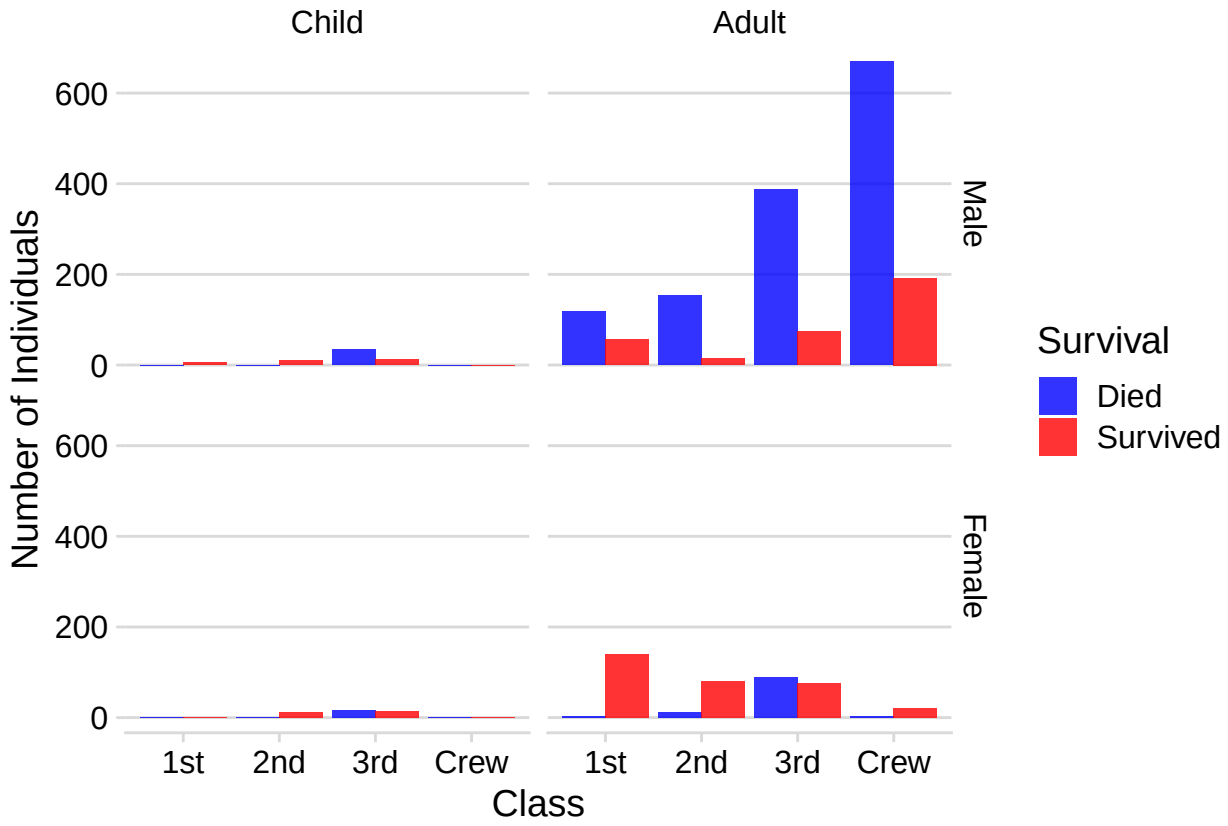
b. Make this graph using the `theme_bw()` theme.

```
plot.1 + theme_bw()
```



c. Make this graph using the `cowplot::theme_minimal_hgrid()` theme.

```
plot.1 + cowplot::theme_minimal_hgrid()
```



d. Why would it be beneficial to drop the vertical grid lines? For some graphs, it may be more beneficial to drop the vertical grid lines for the purpose of making the graph more appealing visually. After making a few different plots with different themes, I personally like the layout of the graph in part c. This is a personal preference but the graph feels less cluttered and is easier on the eyes.