Module14

Joe Vargovich

10/27/2020

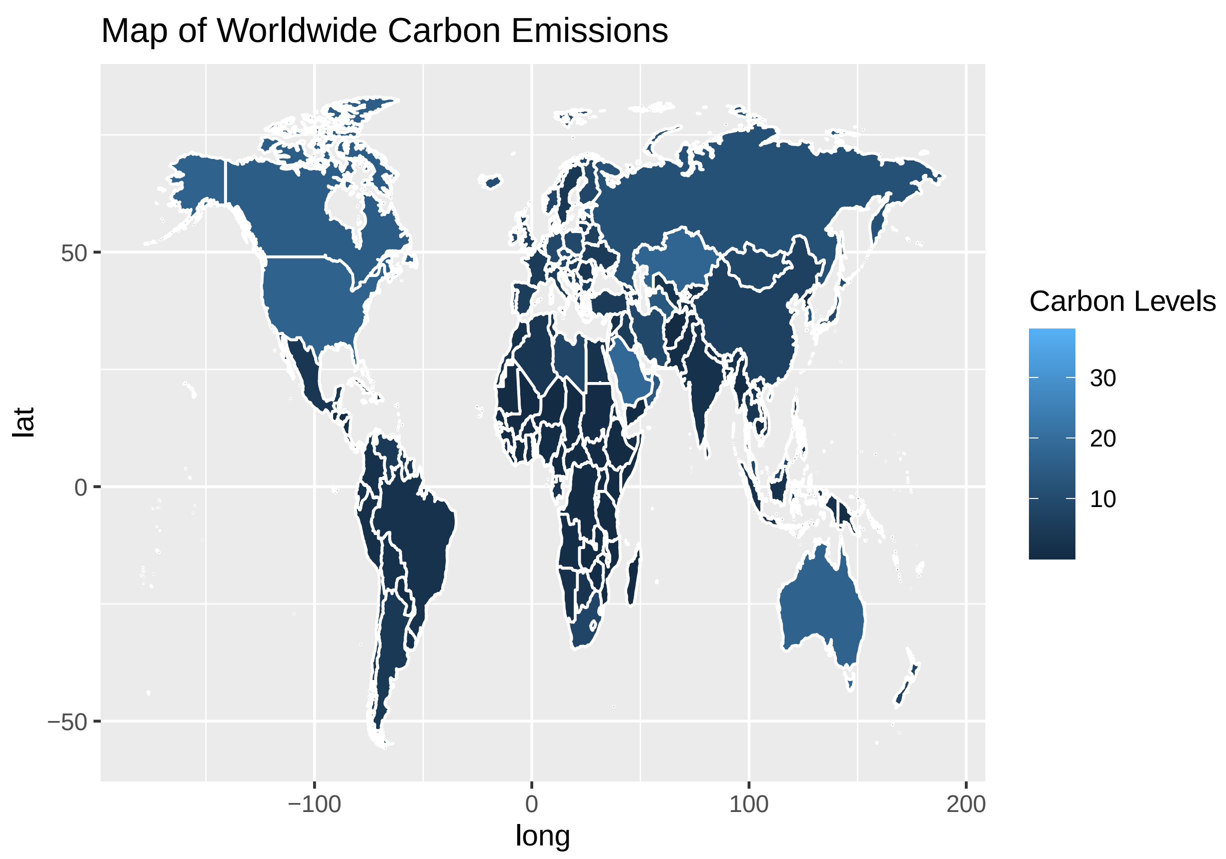
Exercise 1 - CO2 emissions worldwide

```r
#a. Load the data and filter to the current year.
carbonDf = read_csv("co2_emissions_tonnes_per_person.csv")
#Select only 2018
carbonDf = carbonDf %>%
  select("country","2018")

#Load the geoDataSet
geoData = ggplot2::map_data('world')

#Here I tried the standardization method, but it wouldn't join right for me.
# standardCountries = tribble(
#   ~raw, ~standardized,
#  # 'Antigua' = 'Antigua and Barbuda',
#   #'Barbuda' = 'Antigua and Barbuda',
#   'Congo, Rep.' = 'Democratic Republic of the Congo',
#   'Congo, Dem. Rep.' = 'Democratic Republic of the Congo',
#   'Congo, Rep.' = 'Democratic Republic of the Congo',
#   'Cote d\'Ivoire' = 'Ivory Coast',
#   'Swaziland' = 'Estwatini',
#   'Kyrgystan' = 'Kyrgyz Republic',
#   'Laos' = 'Laos',
#   'Micronesia' = 'Micronesia Fed States',
#   'Macedonia' = 'North Macedonia',
#   'Slovakia' = 'Slovak Republic',
#   'Saint Kitts' = 'St. Kitts and Nevis',
#   'Saint. Lucia' = 'St. Lucia',
#   'Grenadines' = 'St. Vincent and the Grenadines',
#   'Trinidad' = 'Trinidad and Tobago',
#   'Tobago' = 'Trinidad and Tobago',
#   'Tuvalu' = 'Tuvalu',
#   'UK', 'United Kingdom',
#   'USA', 'United States'
# )
#
# standardCountries
#
# carbonDf = inner_join(standardCountries, carbonDf, by=c("raw"="country"))
# carbonDf
```

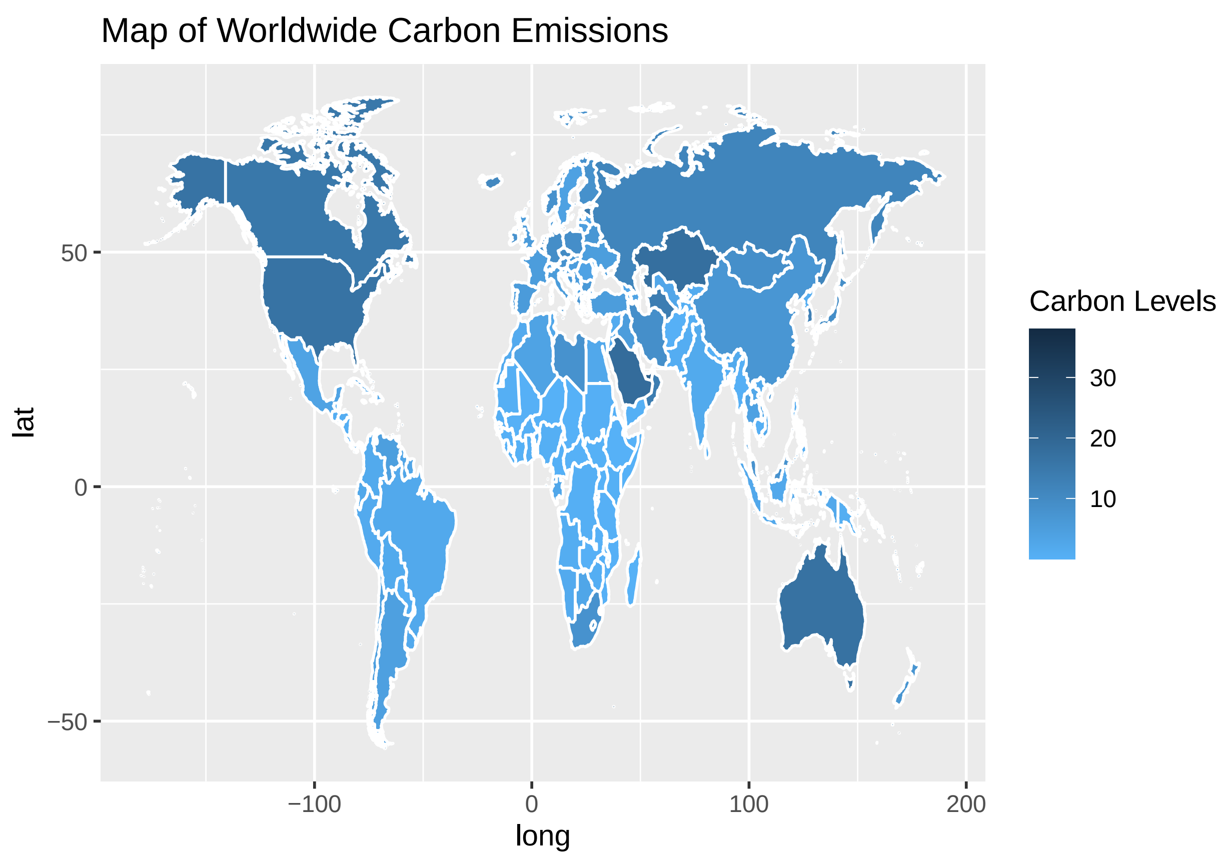
```r
#Thus I did some recodings to match the geoData to the carbonDf.
 geoData = geoData %>%
  mutate(region = factor(region),
         region = fct_recode(region, 'United States' = 'USA'),
         region = fct_recode(region, 'United Kingdom' = 'UK'),
         region = fct_recode(region, 'Cote d\'Ivoire' = 'Ivory Coast'),
         region = fct_recode(region, 'Congo, Rep.' = 'Democratic Republic of the Congo'),
         region = fct_recode(region, 'St. Lucia' = 'Saint Lucia'),
         region = fct_recode(region, 'Slovak Republic' = 'Slovakia'),
         region = fct_recode(region, 'Lao' = 'Laos'),
         region = fct_recode(region, 'Eswatini' = 'Swaziland'),
         region = fct_recode(region, 'Micronesia, Fed. Sts.' = 'Micronesia'),
         region = fct_recode(region, 'Kyrgyz Republic' = 'Kyrgyzstan'),
         region = fct_recode(region, 'St. Kitts and Nevis' = 'Saint Kitts'),
         region = fct_recode(region, 'St. Vincent and the Grenadines' = 'Grenadines'),
         region = fct_recode(region, 'Antigua and Barbuda' = 'Antigua'),
         region = fct_recode(region, 'Antigua and Barbuda' = 'Barbuda'),
         region = fct_recode(region, 'North Macedonia' = 'Macedonia'),
         region = fct_recode(region, 'Trinidad and Tobago' = 'Trinidad'),
  )
#Join with our new matching labels
joinedDf = inner_join(carbonDf, geoData, by=c("country"="region"))

#Rename the 2018 variable to something more descriptive.
joinedDf = rename(joinedDf, carbon = '2018')
head(joinedDf)
```

```
## # A tibble: 6 x 7
##   country     carbon  long   lat group order subregion
##   <chr>        <dbl> <dbl> <dbl> <dbl> <int> <chr>
## 1 Afghanistan  0.254  74.9  37.2     2    12 <NA>
## 2 Afghanistan  0.254  74.8  37.2     2    13 <NA>
## 3 Afghanistan  0.254  74.8  37.2     2    14 <NA>
## 4 Afghanistan  0.254  74.7  37.3     2    15 <NA>
## 5 Afghanistan  0.254  74.7  37.3     2    16 <NA>
## 6 Afghanistan  0.254  74.7  37.3     2    17 <NA>
```

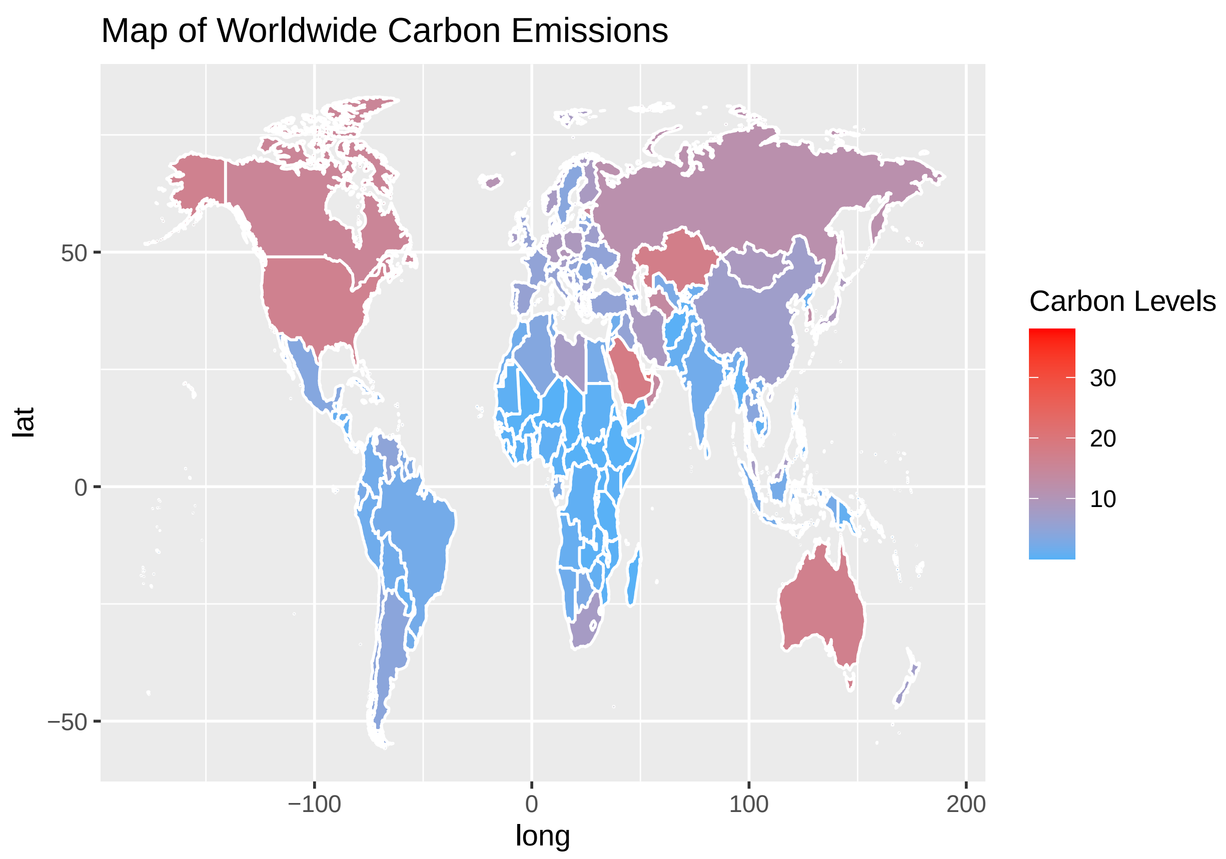
```r
#Create our carbon map with fill color based on carbon levels.
plot = ggplot(joinedDf, aes(x = long, y = lat, group = group, fill=carbon)) +
  geom_polygon( colour = "white") +
  labs(
    fill = "Carbon Levels",
    title = "Map of Worldwide Carbon Emissions"
  )
plot #Default color scheme and gradient
```



```r
#Inverted plot (I think this communicates high emissions better.)
plotInverted = plot +
  scale_fill_gradient(
  low = "#56B1F7",
  high = "#132B43",
)
plotInverted
```



```
#Another cool color idea, even clearer!
plotRed = plot +
 scale_fill_gradient(
  low = "#56B1F7",
  high = "red",
)
plotRed
```

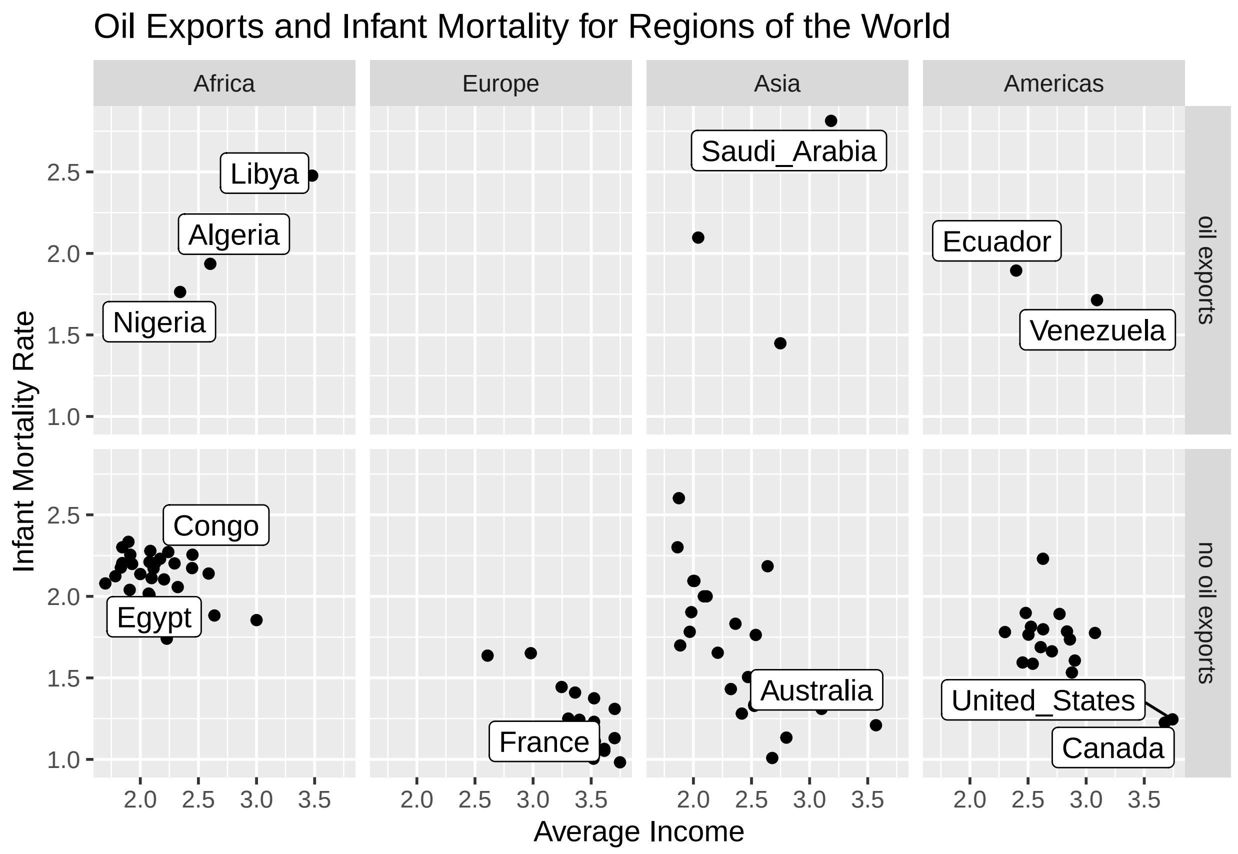


Exercise 3 - Infmort dataset scatterplots

```r
#Load the infmort dataset
data(infmort, package='faraway')

#infmort can be visualized using region and oil as facets.

#Rownames are something we want to access for our labels, let's make them a col.
infmortNew = rownames_to_column(infmort, var="country")

#Filter the infmort data down into some countries we want to label
target = c("Australia", "France", "Saudi_Arabia", "United_States",
           "Canada", "Australia", "Congo", "China", "Egypt", "Algeria",
            "Nigeria", "Libya", "Ecuador", "Venezuela")
labeledCountries = infmortNew %>%
  mutate(
    country = str_trim(country)) %>%
  filter(country %in% target)
head(labeledCountries)
```

```
##         country   region income mortality           oil
## 1     Australia     Asia   3426      26.7 no oil exports
## 2        Canada Americas   4751      16.8 no oil exports
## 3        France   Europe   3403      12.9 no oil exports
## 4 United_States Americas   5523      17.6 no oil exports
## 5       Algeria   Africa    400      86.3     oil exports
## 6       Ecuador Americas    250      78.5     oil exports
```

```
#a. Create scatter plots of countries income and infant mortality using a
# log10 transformation for both axes.
plot = ggplot(infmort, aes(x=log10(income), y= log10(mortality))) +
  geom_point() +
  facet_grid(oil ~ region) +
  #b. Label 10-15 countries across the regions.
  geom_label_repel(data=labeledCountries, aes(label=country)) +
  labs(
    title = "Oil Exports and Infant Mortality for Regions of the World",
    x = "Average Income",
    y = "Infant Mortality Rate"
  )
plot
```

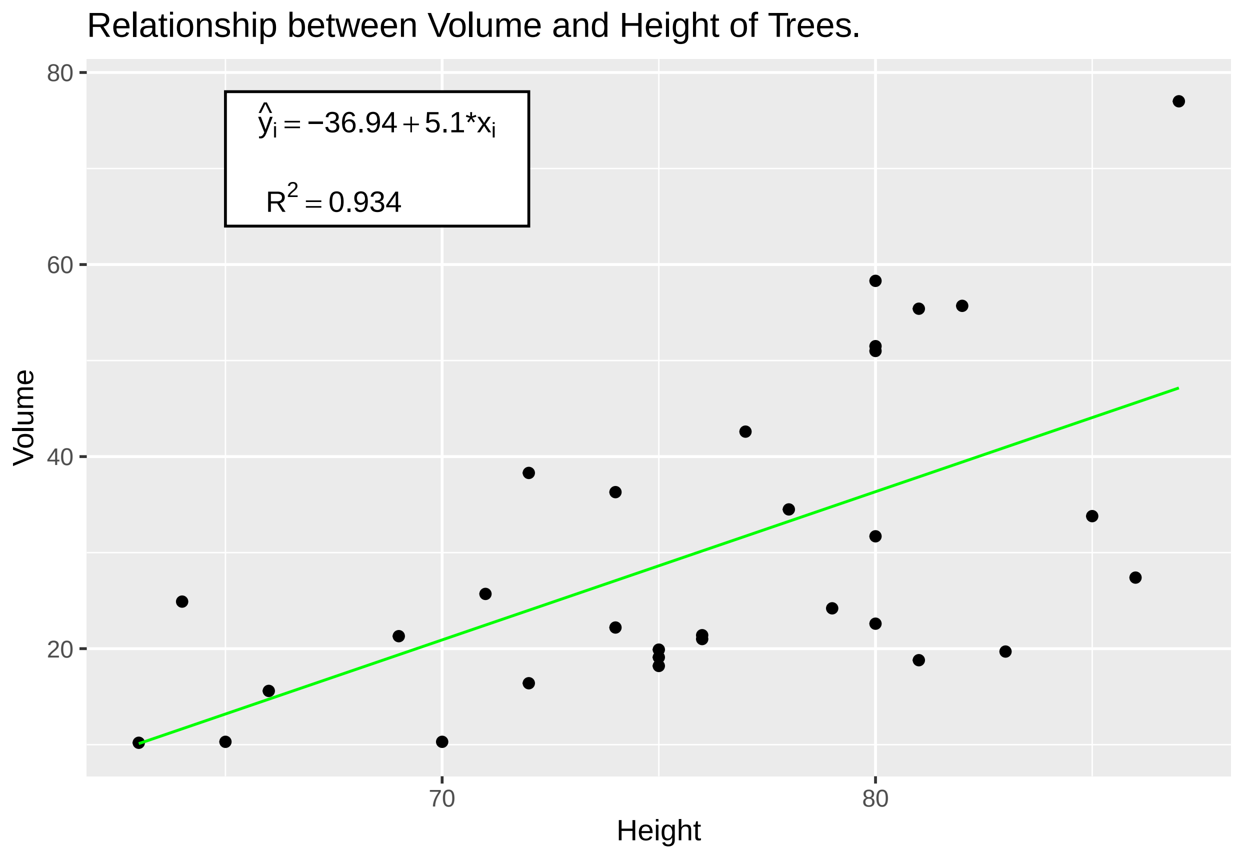


Exercise 4 - Trees dataset regression model and graphing

```
trees = datasets::trees
#a. Create regression model for Volume as a function of Height
model <- lm( Volume ~ Height, data=trees)      # Fit a regression model
trees <- trees %>%                         # save the regression line yhat points
  mutate(fit=fitted(model))
head(trees)
```

```
##   Girth Height Volume      fit
## 1   8.3     70   10.3 20.91087
## 2   8.6     65   10.3 13.19412
## 3   8.8     63   10.2 10.10742
## 4  10.5     72   16.4 23.99757
## 5  10.7     81   18.8 37.88772
## 6  10.8     83   19.7 40.97442
```

```
#b. Get the y-intercept and slope of regression line
#This is given by the linear model object
coefs = model$coefficients

#c. Plot the data
plot = ggplot(trees, aes(x=Height)) +
  geom_point(aes(y=Volume)) +
  geom_line(aes(y=fit), color='green') +
  annotate('rect', xmin=65, xmax=72, ymin=64, ymax=78,
           fill='white', color='black') +
  annotate('text',x=68.5, y=75, label = TeX('$\\hat{y}_i = -36.94 + 5.1*x_i$')) +
  annotate('text', x=67.5, y = 67, label=TeX('$R^2 = 0.934$')) +
  labs(
    title = "Relationship between Volume and Height of Trees."
  )
plot
```

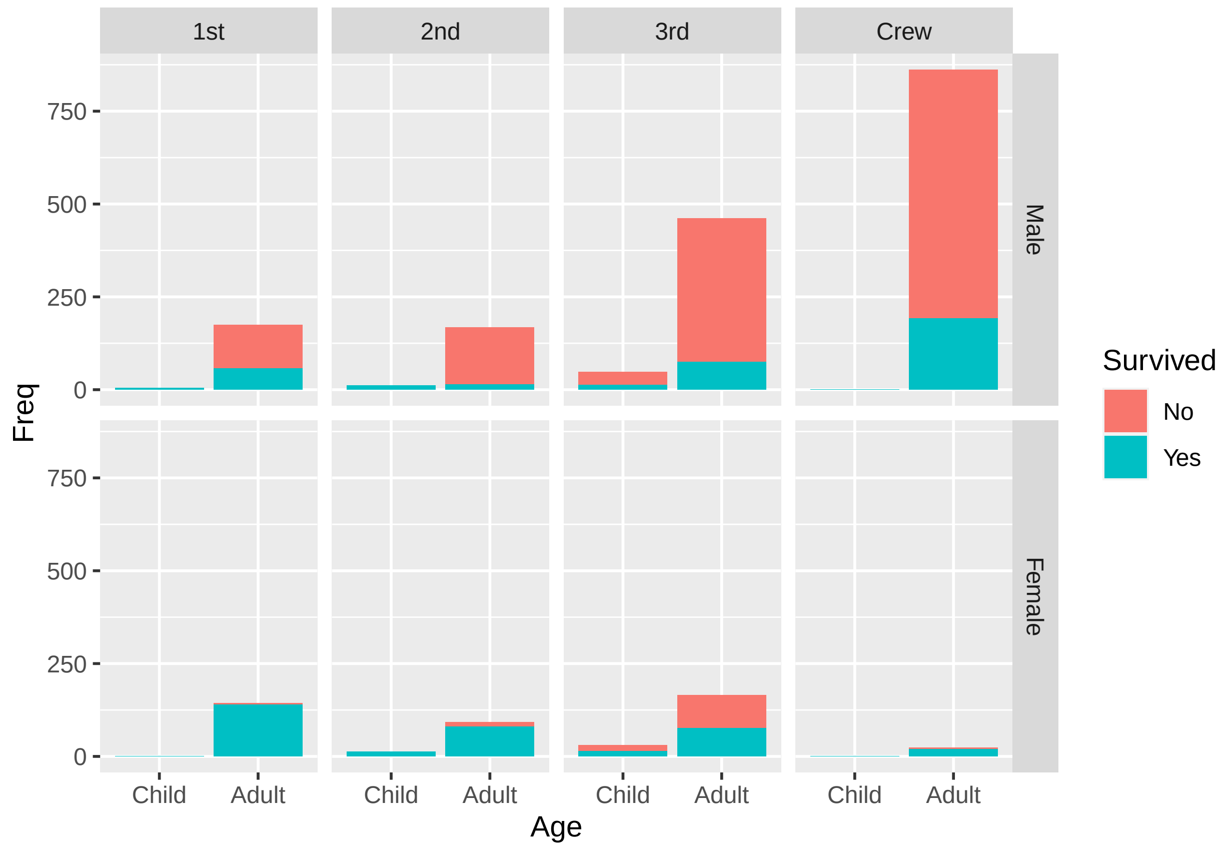


Exercise 5 - Titanic Survival Statistics

```r
#Load the data
Titanic = datasets::Titanic
Titanic = Titanic %>% as.data.frame() %>%
  group_by(Age, Class)
head(Titanic)
```

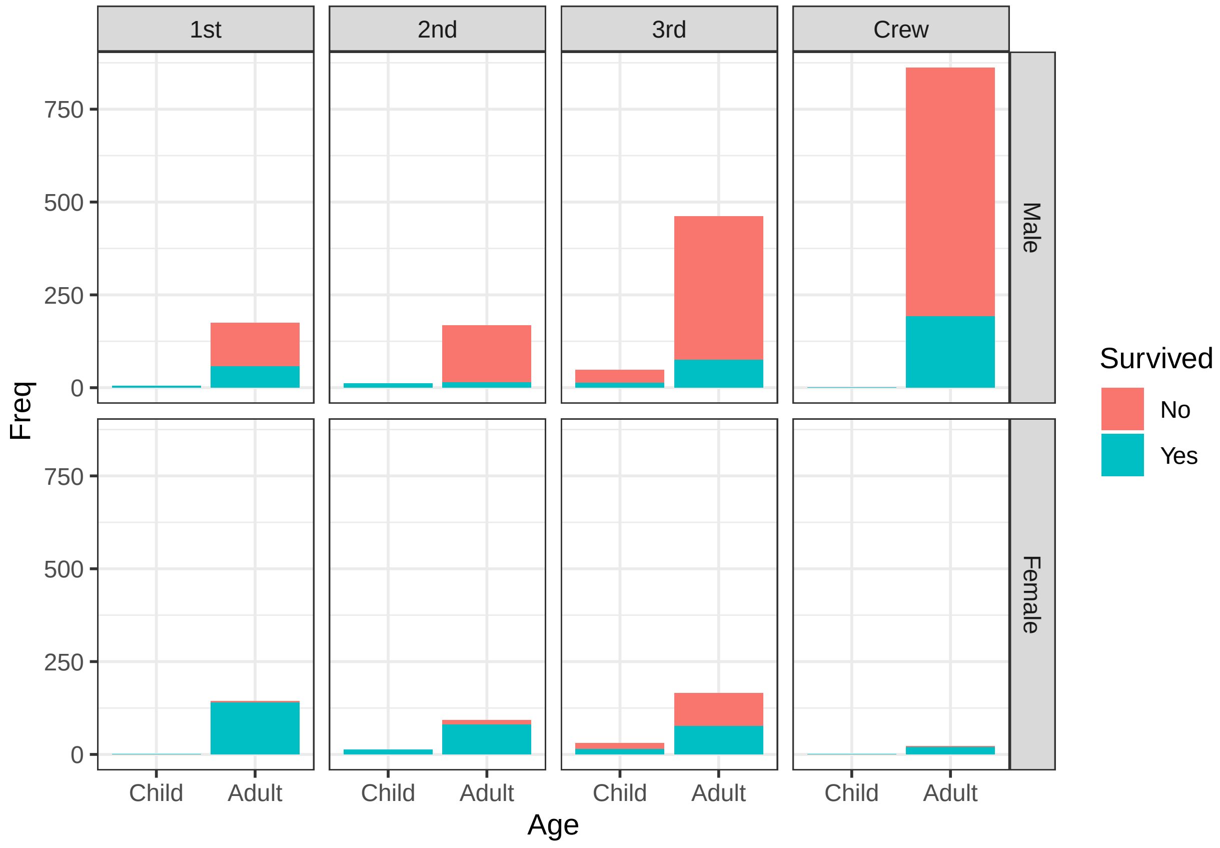
```
## # A tibble: 6 x 5
## # Groups:   Age, Class [4]
##   Class Sex    Age   Survived  Freq
##   <fct> <fct>  <fct> <fct>    <dbl>
## 1 1st   Male   Child No           0
## 2 2nd   Male   Child No           0
## 3 3rd   Male   Child No          35
## 4 Crew  Male   Child No           0
## 5 1st   Female Child No           0
## 6 2nd   Female Child No           0
```

```r
#a. Create the plot showing the number of survivors of each class and gender.
defaultPlot = ggplot(Titanic, aes(x=Age, y=Freq), group=Sex, color=Class) +
 geom_col(aes(x=Age,fill=Survived)) +
  facet_grid(Sex ~ Class)
defaultPlot
```

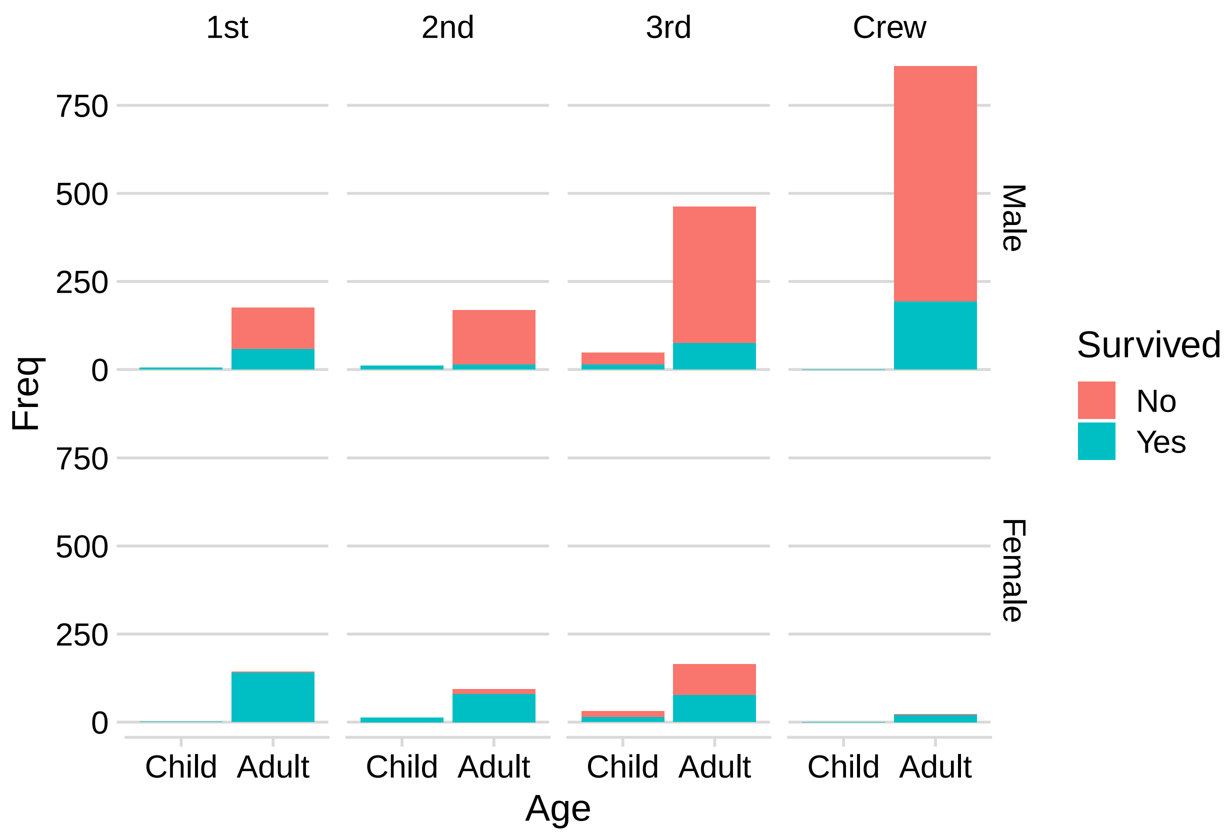


```r
#b. Create the plot using the theme_bw
bwPlot = defaultPlot + theme_bw()
```

```
bwPlot
```



```
#c. Create the plot using a minimial hgrid theme
minimalPlot = defaultPlot + cowplot::theme_minimal_hgrid()
minimalPlot
```



```
#d. Is removing the vertical grid lines good or not?
#Removing the vertical grid lines declutters the noise of the default
#grid theme. This allows us to see the bars more easily while still retaining
# the useful horizontal grid lines used to measure the values of the bars.
```