# Module 2 Solutions

Make sure ggplot2 library is installed and then loaded correctly!

Hopefully you have saved the data and RMD into the same folder, and created a project file for that folder. If so then the Pulse dataset can be read in without modifying the code below.

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.3

Pulse<-read.csv("Pulse.csv")

## Exercises

**Exercise 1. Fix the error(s) in this code to produce a boxplot of attractiveness by age group**

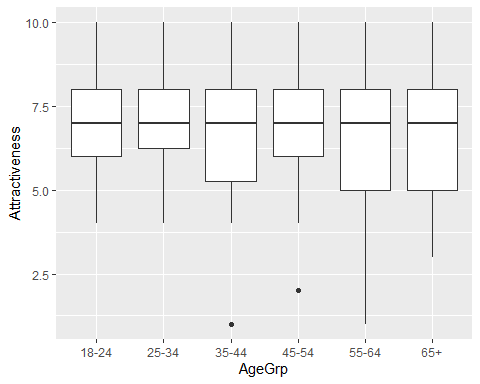
ggplot2(data = Pulse, aes(x = agegroup, y = attractiveness) geom\_boxplot()

## Answer 1

There are a few errors in this code: 1. The function is ggplot(). The library is ggplot2 2. The variable names are both incorrect. x should be “AgeGrp” and y should be “Attractiveness”. Remember case is important! 3. Only one bracket is being closed in the first line of code. Both ggplot() and aes() need to have brackets closed 4. There is no + at the end of the line.

So it should look like:

ggplot(data = Pulse, aes(x = AgeGrp, y = Attractiveness))+   
 geom\_boxplot()

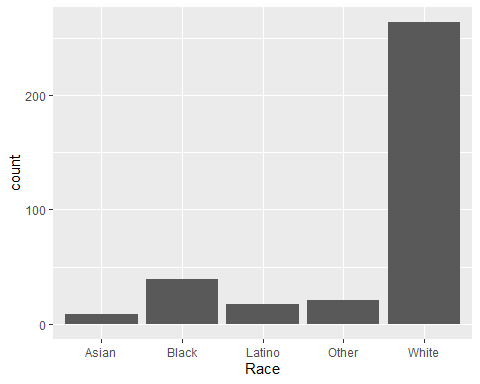


**Exercise 2. Produce a bar chart showing the frequencies of the Race variable**

\*\*Answer 2.

I need to use the Pulse data, map Race to either the x or y aesthetic, and use geom\_bar geom\_bar should only have x or y aesthetic specified. The frequencies will appear by default on the unmapped aesthetic

ggplot(data=Pulse,aes(x=Race))+  
 geom\_bar()

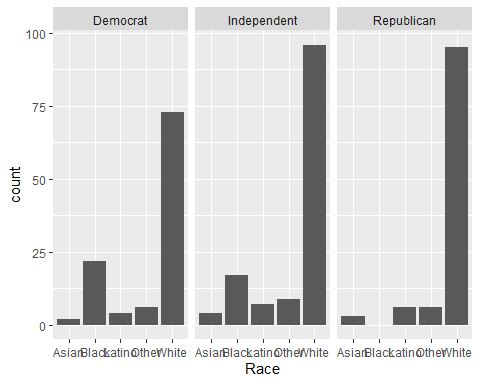


**Exercise 3. Continuing the previous example, use facets to make separate bar charts of race by political party**

\*\*Answer 3

I can copy and paste my previous response, and then use facet\_wrap() with the PoliticalParty variable. Remember to use the tilde sign

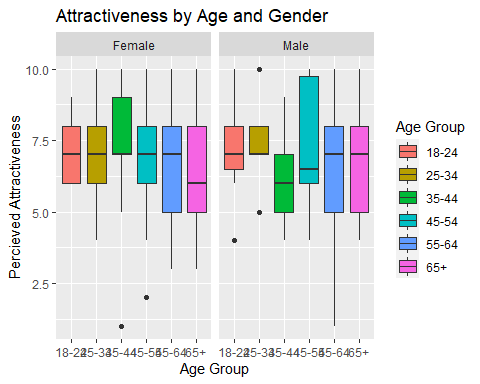
ggplot(data=Pulse,aes(x=Race))+  
 geom\_bar()+  
 facet\_wrap(~PoliticalParty)



**Exercise 4. Produce a set of boxplots of attractiveness by age group, with separate panels for each gender. Use some colours and add sensible titles and labels**

This is a continuation of what we started with exercise 1; but we need to use facet\_wrap() with gender. I will also map the fill axis to the AgeGrp and use lab() to provide sensible labels

ggplot(data = Pulse, aes(x = AgeGrp, y = Attractiveness,fill=AgeGrp))+   
 geom\_boxplot() +  
 facet\_wrap(~Gender)+  
 labs(title = "Attractiveness by Age and Gender",x="Age Group",y="Percieved Attractiveness",fill="Age Group")

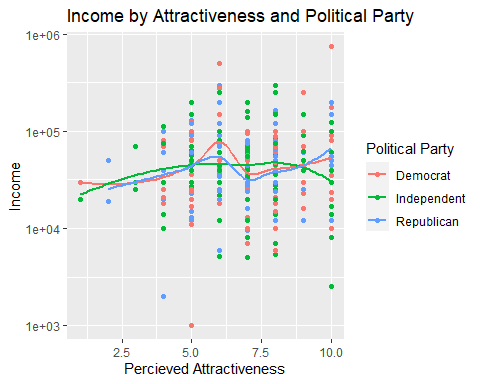


**Exercise 5. Produce a graph to investigate whether there is a relationship between income, on the y axis, and attractiveness, on the x axis. Use different colours for each political party, a log axis for income and add in some sensible titles and labels. You also might want to consider using more than one layer.**

Here I am going to use a scatter plot with geom\_point() and also add a trend line using geom\_smooth(). To set different colours I will need to map the color aesthetic, not the fill aesthetic, as this affects points and lines. I’ve also turned off the standard error bars around the trendline since they were a bit in the way for me!

ggplot(data = Pulse, aes(x = Attractiveness, y = Income,color=PoliticalParty))+   
 geom\_point() +  
 geom\_smooth(se=FALSE)+  
 labs(title = "Income by Attractiveness and Political Party",y="Income",x="Percieved Attractiveness",col="Political Party")+  
 scale\_y\_log10()

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



My general conclusion is there is not much relationship. The people rating themselves with very low attractiveness scores do maybe have lower incomes on average, but there are very few of these people to be drawing conclusions from. Equally there doesnt seem to be any differences in this relationship by political party.