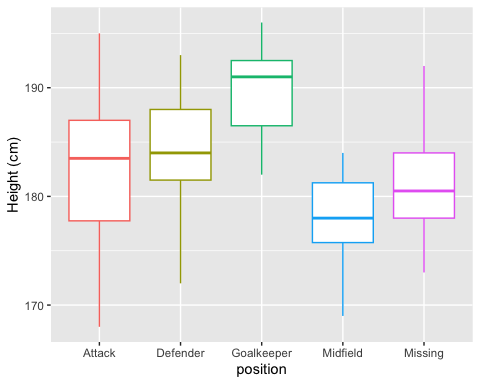
SoccerModule

A researcher wants to investigate whether there are differences among the mean height for different positions (Goalkeeper, Defender, Midfield, or Attack) of soccer players. Goalkeepers are typically thought to be taller than other positions to make up in wingspan to cover the goal better. The ‘players\_df’ dataset contains information on each player including their name, current club, position, and height. For this analysis, only the position and height of the player will be used. A sample of 20 players was taken from each position and will be used for this analysis.



#### Question 1.

1. Calculate the mean height for each position.
2. Are there any issues with the data that could affect the results?
3. Write of the null and alternative hypotheses that would be consistent with the research question.
4. Complete the ANOVA F-test.

Df Sum Sq Mean Sq F value Pr(>F)   
position 4 1540 385.1 13.26 1.27e-08 \*\*\*  
Residuals 95 2758 29.0   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

1. Appropriately interpret results of analysis of variance F-test.

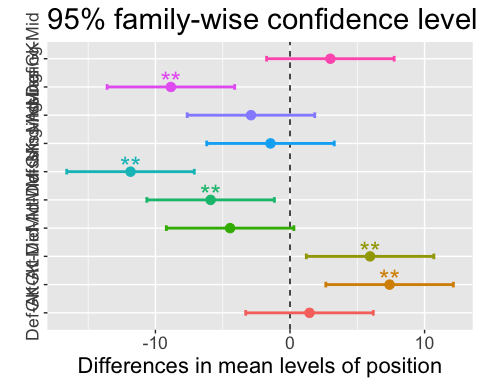
#### Question 2.

We will now perform a follow-up analysis to determine which pairs of positions have significantly different height means.

Tukey multiple comparisons of means  
 95% family-wise confidence level  
  
Fit: aov(formula = height\_in\_cm ~ position, data = height\_sample\_df)  
  
$position  
 diff lwr upr p adj  
Def-Att 1.45 -3.288606 6.188606 0.9137637  
GK-Att 7.40 2.661394 12.138606 0.0003340  
Mid-Att -4.45 -9.188606 0.288606 0.0763027  
Missing-Att -1.45 -6.188606 3.288606 0.9137637  
GK-Def 5.95 1.211394 10.688606 0.0064112  
Mid-Def -5.90 -10.638606 -1.161394 0.0070357  
Missing-Def -2.90 -7.638606 1.838606 0.4377532  
Mid-GK -11.85 -16.588606 -7.111394 0.0000000  
Missing-GK -8.85 -13.588606 -4.111394 0.0000115  
Missing-Mid 3.00 -1.738606 7.738606 0.4026003

#### Tukey Pair-wise comparison Plot

Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as  
of ggplot2 3.3.4.  
ℹ The deprecated feature was likely used in the ggiraphExtra package.  
 Please report the issue to the authors.



1. Write a few sentences summarizing the interesting findings between the groups