

# 5240 Workshop 09

Digvijay Jondhale 0862899

## Loading the Data ...

```
library(tidyr)
library(ggplot2)
football <- read.csv(file = 'footballweights.csv')
```

```
str(football)
```

```
'data.frame':  17 obs. of  5 variables:
 $ cowboys      : int  258 263 263 272 258 273 253 260 274 254 ...
 $ packers      : int  260 271 258 263 267 254 255 250 248 240 ...
 $ broncos      : int  270 250 281 273 257 264 233 254 268 252 ...
 $ dolphins     : int  260 255 265 257 268 263 247 253 251 252 ...
 $ fortyniners : int  247 249 255 247 244 245 249 260 217 208 ...
```

## a) Converting to longer format

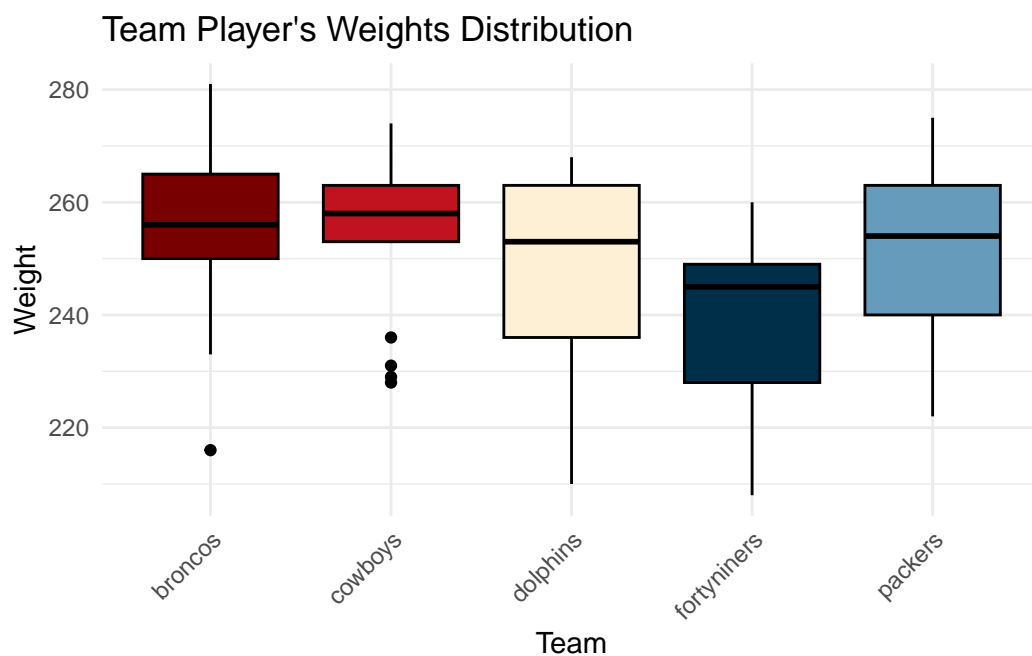
```
long_football <- pivot_longer(
  football,
  cols = everything(),
  names_to = "team",
  values_to = "weights"
)
```

```
str(long_football)
```

```
tibble [85 x 2] (S3: tbl_df/tbl/data.frame)
 $ team      : chr [1:85] "cowboys" "packers" "broncos" "dolphins" ...
 $ weights: int [1:85] 258 260 270 260 247 263 271 250 255 249 ...
```

## b) Box Plot

```
ggplot(long_football, aes(x = team, y = weights)) +  
  geom_boxplot(fill = c("#780000", "#c1121f", "#fdf0d5", "#003049", "#669bbc"), col="black") +  
  labs(  
    title = "Team Player's Weights Distribution",  
    x = "Team",  
    y = "Weight"  
  ) +  
  theme_minimal() +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



## c) ANOVA

$H_0$  : All Teams have same mean weights of players.  $H_a$  : At least one Team has different mean weight of players.

```
anova_result <- aov(weights~team, data = long_football)  
summary(anova_result)
```

```

      Df Sum Sq Mean Sq F value Pr(>F)
team      4   2851    712.7    2.742 0.0341 *
Residuals 80  20795    259.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

As the p-values(0.03) from the anova is  $< 0.05$ , we have to reject the null hypothesis and conclude that at least one team has a different mean weight. To find the teams who has different mean weight, we can perform Tukey's Honestly Significant Difference Test.

#### d) Tukey's Honestly Significant Difference Test

```
TukeyHSD(anova_result)
```

```

Tukey multiple comparisons of means
 95% family-wise confidence level

```

```
Fit: aov(formula = weights ~ team, data = long_football)
```

```

$team
      diff      lwr      upr    p adj
cowboys-broncos  0.8823529 -14.551681 16.3163866 0.9998513
dolphins-broncos -5.2352941 -20.669328 10.1987396 0.8776395
fortyniners-broncos -15.2941176 -30.728151  0.1399160 0.0533123
packers-broncos    -3.2941176 -18.728151 12.1399160 0.9753589
dolphins-cowboys   -6.1176471 -21.551681  9.3163866 0.8027693
fortyniners-cowboys -16.1764706 -31.610504 -0.7424369 0.0352373
packers-cowboys     -4.1764706 -19.610504 11.2575631 0.9424607
fortyniners-dolphins -10.0588235 -25.492857  5.3752102 0.3700418
packers-dolphins     1.9411765 -13.492857 17.3752102 0.9966796
packers-fortyniners 12.0000000  -3.434034 27.4340337 0.2017663

```

#### Results

Fortyniners and cowboys are the teams , who has significant difference in their mean weights with fortyniners mean score significantly lower than cowboys by 16.17 units.