IS-4300 Final Project

Milciades Arauz

4/30/2021

## Introduction

This project is about the comparison between passes in a league match and in a tournament match. The variables that we will use are **CompT.Pass** for completed passes in a tournament and **CompL.Pass** for completed passes in the league. Even though the analysis is comparing the amount of passes, it also be in usage the variables **Player** and **Pos** to get further detail in who made the passes and in which position they play.

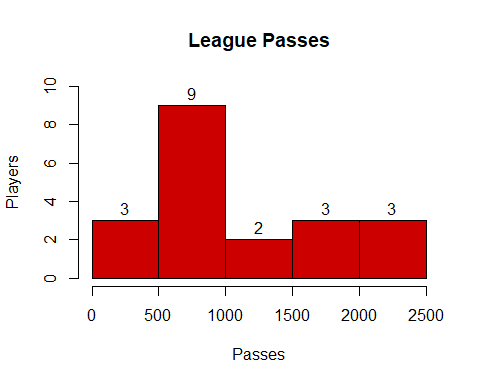
Player Pos CompT.Pass CompL.Pass  
1 Manuel Neuer GK 430 1272  
2 Joshua Kimmich DF,MF 732 2440  
3 Robert Lewandowski FW 196 592  
4 Thiago Alcántara MF 760 1649  
5 Serge Gnabry FW 318 883  
6 Alphonso Davies DF 465 1585

## Heads-Up

Before we take a look to the histograms is that there is a huge difference because of number if games each “category” has, in the german league or Bundesliga there are thirty-four games while in the tournament also known as Champions League there are six for group stage or thirteen if the team go all the way to the final, which Bayern Munich did.

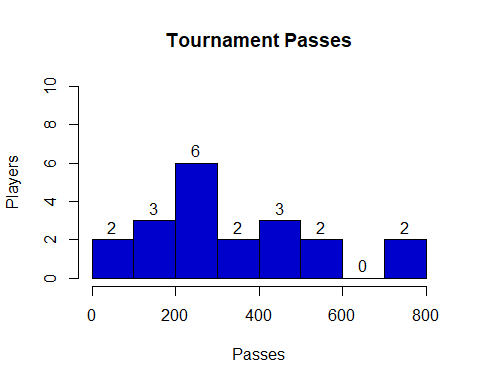
### League’s Histogram

The first histogram that were going to check is the league’s histogram, it has the number of players on the “y” and the amount of passes in the axis x. One way to read this histogram is tha nine players made more than 500 passes, other example would be only two players made more than 1000 passes, but they are six players that achieved more than 1500 passes.



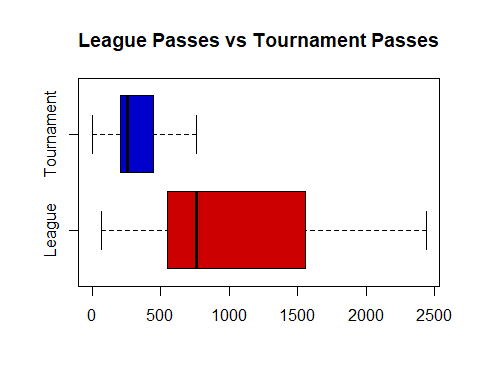
### Tournament’s Histogram

The next histogram is more unique way, instead of a ascending rate of passes, in this histogram it suddenly stops at 600 passes, and skips the 600-700 range to keep going in the 700-800 range, only *two* players were able to break the constant rate.



## Boxplot

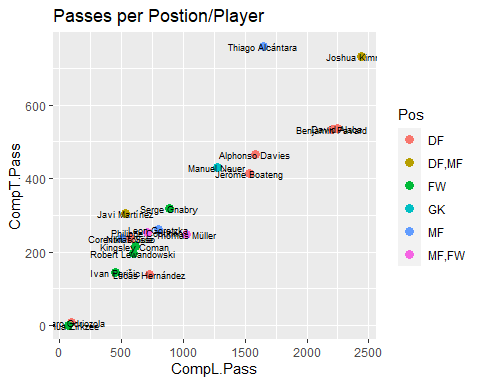
The next visual is comparing between the two quantities of passes.



As you can see in the box plot, both data have huge difference between them,the max number in the Tournament’s boxplot is the median of the League’s boxplot and the reason behind is the amount of games each competion has, just as it was briefly explained in the **Heads-Up**

## Scatter Plot

Another data was unseen, it’s about the variable *Player*, which it comes with two questions: Which position provide the most completed passes? and Who is player?



The scatter plot not only output the position but the names as well. Apparently the player who most provided completed passes is Joshua Kimmich, the other player very close is Thiago Alcantara. The reason why Joshua a Kimmich is high it’s because he plays two different position DF as a Right Back or *RB* and MF as Central Defense Midfielder or *CDM*. Because of these two positions Kimmich has more chances to pass the ball.

### Type of Play

You wonder why these two players have the most amount of passes. In the 2019-2020 the team used these lineup Thiago in the middle as a *CDM* and Joshua as *RB*, not only them but most of the players that have a decent number of passes are defenders. In the scatter plot you can “see” their way of play is passing the nack retaining possession to use this as an counter attack.



Bayern Munich 2019-2020

## 

## Two Sample t-test

Now that we have the visual out of the way. Lets have a test on the sample data, *CompL.Pass* would be the independent variable and *CompT.Pass* is the dependent variable. The decision of what variable is independent or dependent were based the amount of games each competition has.

Welch Two Sample t-test  
  
data: Passes$CompL.Pass and Passes$CompT.Pass  
t = 4.289, df = 22.269, p-value = 0.0002911  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 363.1865 1042.3135  
sample estimates:  
mean of x mean of y   
 1024.50 321.75

We can conclude that with at the .05 level of significance, there is very significant difference between League passes and Tournament passes with the difference of 702.75 passes.