

STEP 1 CREATING DATASET

```
FIFAdata=read.csv(file.choose(),header =  
TRUE)
```

```
Impluslf= FIFAdata %>%
```

```
  filter(FIFAdata$Position == "LM" |  
FIFAdata$Position == 'LF' )
```

```
Impluslf=data.frame(Impluslf)
```

STEP 2 CREATING GGPAIR()

THERE ARE VERY LITTLE "LF"
POSITION IN THE LEAGUE
COMPARED TO "LM"

"LF" HAS ON AVERAGE HIGHER
AGILITY WHILE "LM" HAS HIGHER
ACCELERATION ON AVERAGE.

GENERALLY SPEAKING, THE
HIGHER THE AGILITY, THE HIGHER
THE ACCELERATION



T - TEST

Step 1 : Create Null Hypothesis

Ho: $\mu(\text{LM}) - \mu(\text{LF}) = 0$, Ha: $\mu(\text{LM}) - \mu(\text{LF}) \neq 0$

Step 2: Find Critical Level (Skipped)

Step 3: Find Test Statistics

"t = -10.095, df = 57.404, p-value = 2.452e-14

95 percent confidence interval:

-4.804353 -3.214072

sample estimates:

mean of x mean of y

```
xbar_holder1=numeric(l=30)
n1=30
for( i in 1:30 ){
  sample1=rnorm(30,mean = mean(LMA$Agility), sd= sd(LMA$Agility))
  xbar1=mean(sample1)
  xbar_holder1[i]=xbar1
}
xbar_holder1

xbar_holder2=numeric(l=30)
for( i in 1:30 ){
  sample2=rnorm(30,mean = mean(LFA$Agility), sd= sd(LFA$Agility))
  xbar2=mean(sample2)
  xbar_holder2[i]=xbar2
}
xbar_holder2

t.test(x=xbar_holder1,y=xbar_holder2)
```

Step 4: Find P Value

P= 2.452e⁽⁻¹⁴⁾

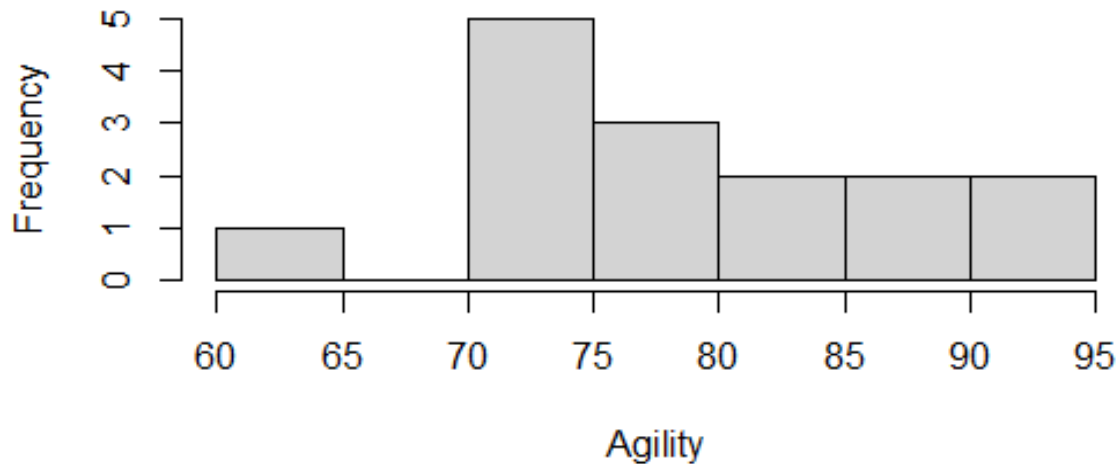
Step 5: Conclusion

Reject Ho

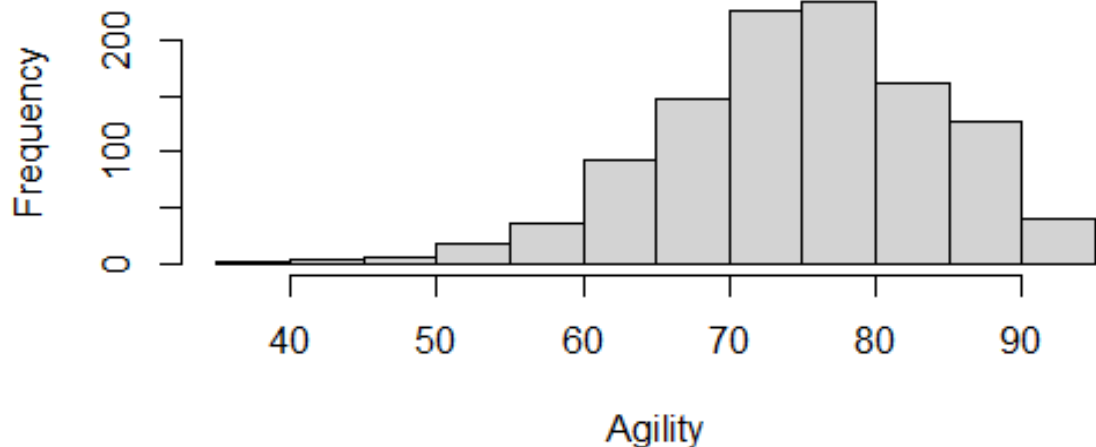
Step 6: Explanation

There is strong evidence suggesting that $\mu(\text{LM}) \neq \mu(\text{LF})$

Histogram of LF's Agility



Histogram of LM's Agility

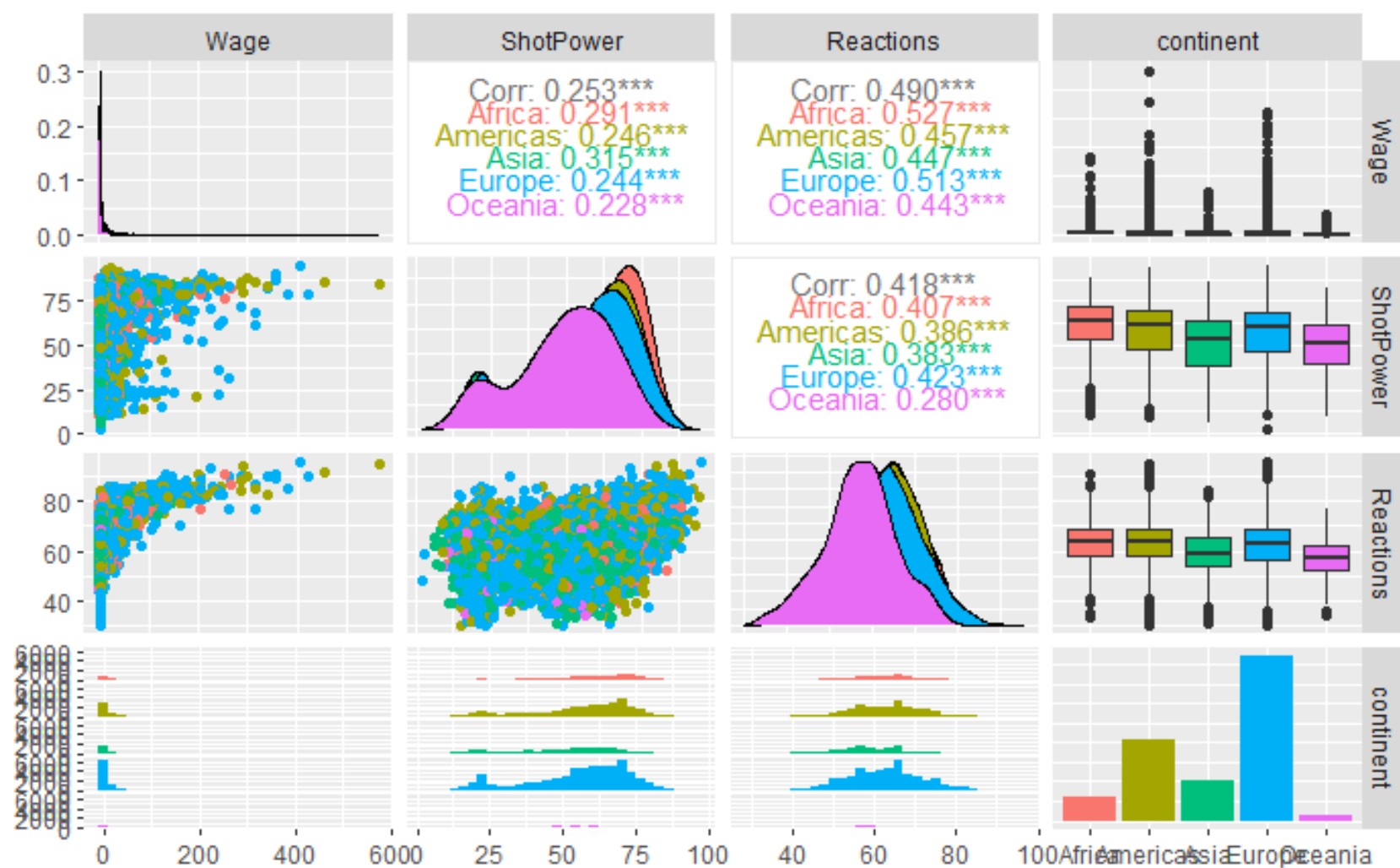


COMMENT (HAVE NOT HAD STAT1)

- By looking at the two histograms on the left, we can visually see that $\mu(\text{LM}) \neq \mu(\text{LF})$ which match our T-test in the previous slide
- I selected sample size of 30 and according to CLT, the sample distribution of the sample means should be reasonably normally distributed.
- Two histogram present two different skewness one to the right and one to the left, therefore visually I would say that there is a difference between two standard deviations.

THE RELATIONSHIP BETWEEN

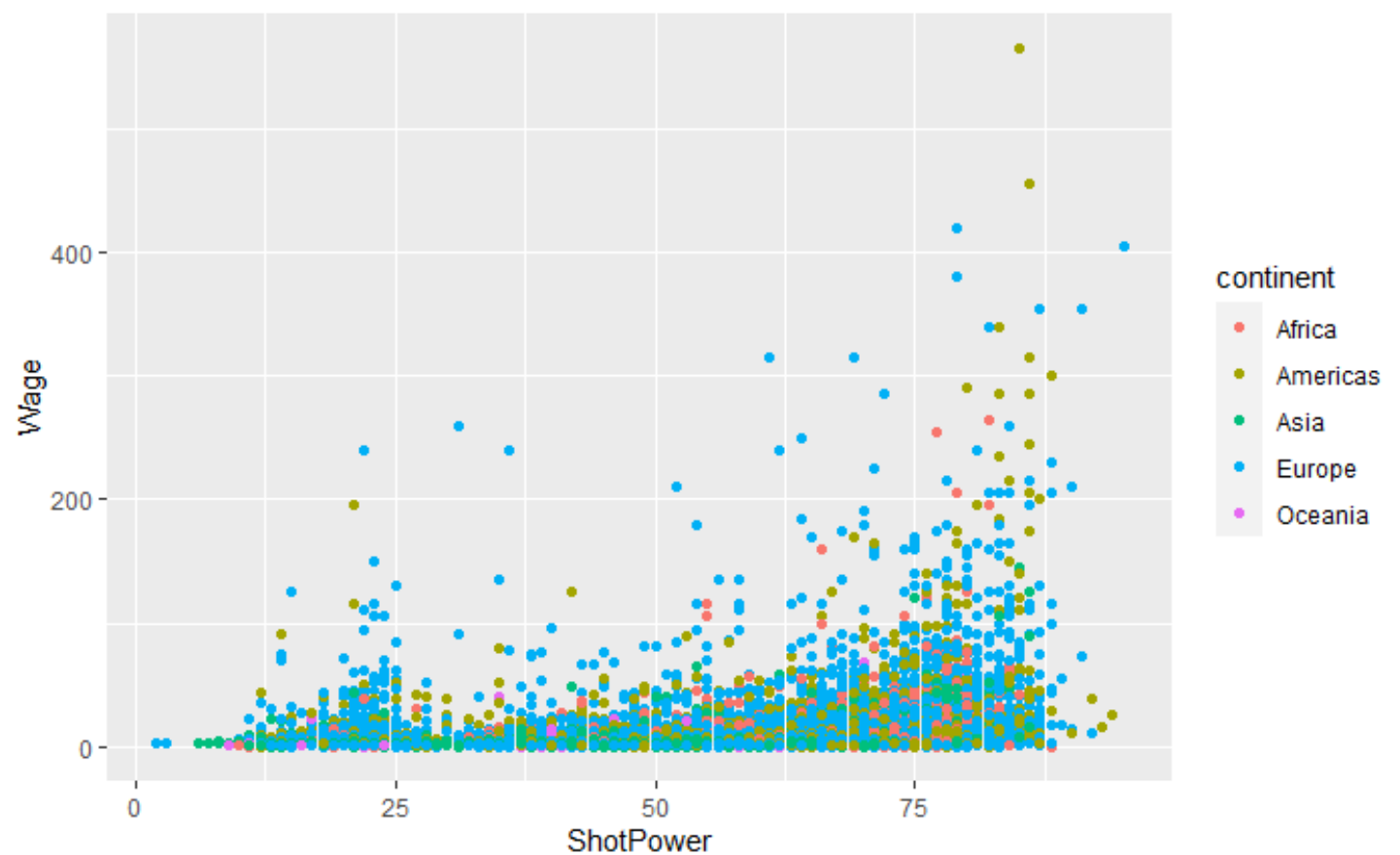
-WAGE
-SHOT POWER
-REACTION
-CONTIENT
(CREATED BY
NATIONALITY)



CREATED CATEGORICAL VARIABLE:

-AFRICA
-AMERICAS
-ASIA
-EUROPE
-OCEANIA

```
##{r}
#creating new variable by contient
install.packages("countrycode")
library(countrycode)
fifa %>%
  add_column(add_column = "continent")
fifa$continent=countrycode(fifa$Nationality, origin = "country.name",destination = "continent")
##
```



QUESTION:

IS SHOT POWER AND CREATION TIME POSITIVELY CORRELATED WITH WAGE?

WHAT'S THE RELATIONSHIP BETWEEN WAGE AND CONTINENT?

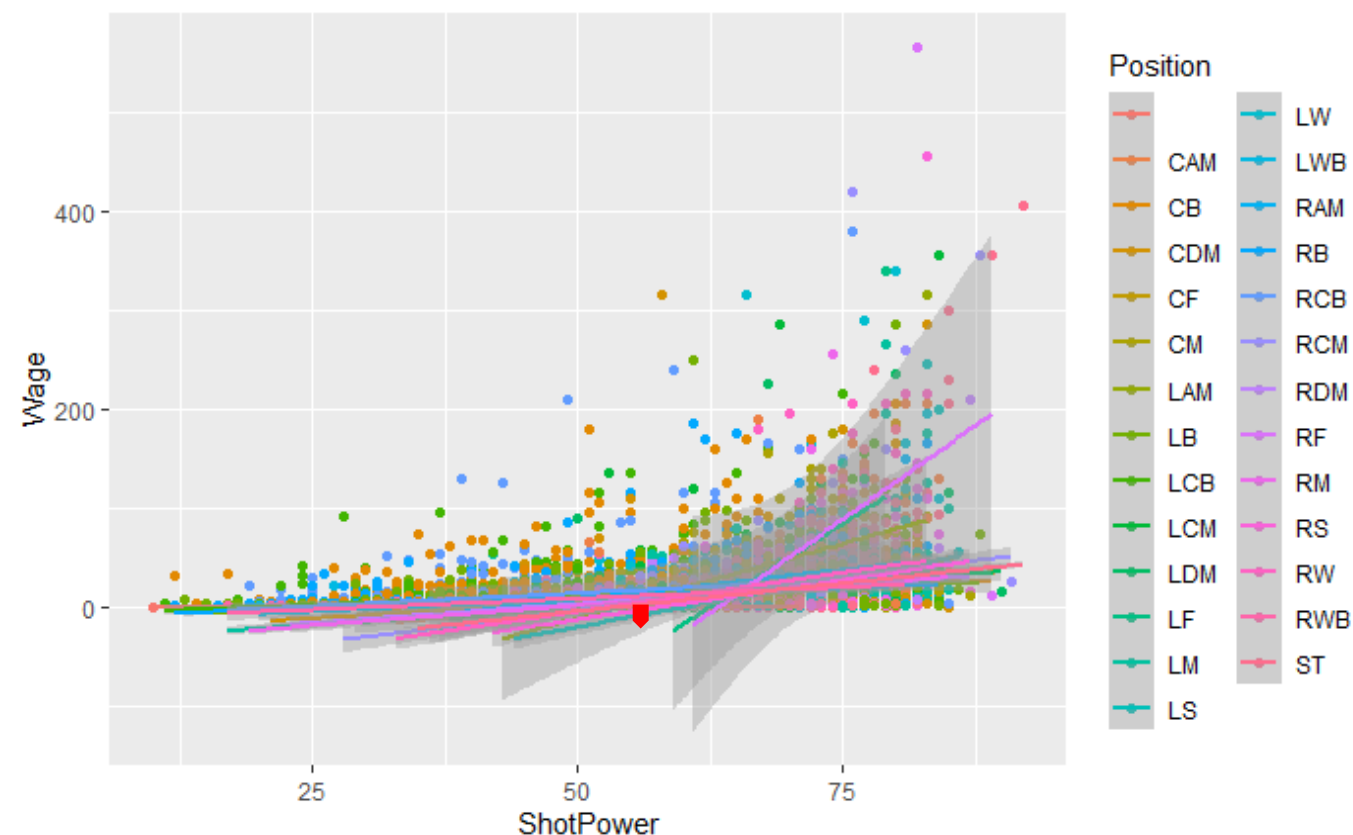
-There is no clear visual evidence suggesting that higher shot power = higher wages.

-However, Most high wage players have nationality in America or Europe



IF WE BRING "POSITION" INTO THE EQUATION

- CAN'T FIGURE OUT WHY THERE IS AN EMPTY CATEGORY IN POSITION DID `FILTER(!IS.NA)`
- FILTER OUT GOALKEEPER POSITION
- WE CAN SEE THAT, FOR POSITION "RF" THERE SEEMS TO BE A POSITIVE RELATIONSHIP.



```
fifa %>%  
  filter(!is.na(fifa$Position)) %>%  
  filter(Position != "GK") %>%  
  ggplot(mapping=aes(x = ShotPower , y = Wage, color=Position)) +  
  geom_point()+  
  geom_smooth(method=lm)
```

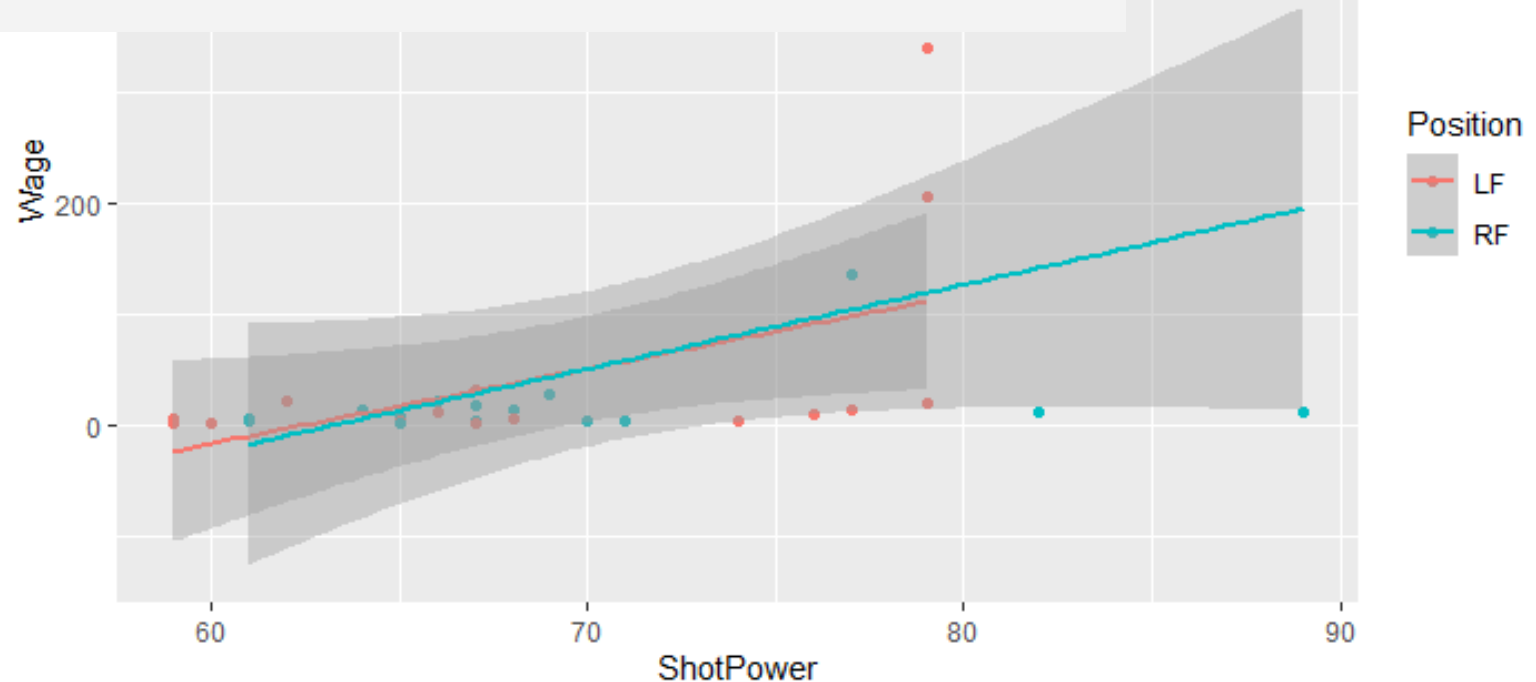
SHOTPOWER VS WAGE (POSITION = LF,RF)

If we look deeper,

It seems that shot power is visually positively correlated with players that play position LF and Rf.

Intuitively, you would think that people playing forward position will tend to have higher Shot power, which is positively correlated with wage, that matches our assumption

```
fifa %>%  
  filter(!is.na(fifa$Position)) %>%  
  filter(Position == "LF" | Position == "RF") %>%  
  ggplot(mapping=aes(x = ShotPower , y = Wage, color=Position)) +  
  geom_point()+  
  geom_smooth(method=lm)
```





TAKE AWAY

There were many questions that came up during the EDA, however theorizing and testing these observations was time consuming and tedious.

I was a little taken away while doing my EDA and went off track a little bit, there are many things that I feel like are correlated, but I fell short on proving them

For example, I wanted to test wages based on locations using nationality, I have built my continent variable but was stuck on getting a meaningful result.

I think things will gradually improve over time however I do very much enjoy the process of solving and attacking the problems.