HW3\_pnegandh

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.3

library(readr)

## Warning: package 'readr' was built under R version 3.5.3

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(plyr)

## Warning: package 'plyr' was built under R version 3.5.3

## -------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

## -------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

nba <- read\_csv("nba2017-18.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_double(),  
## Player = col\_character(),  
## Season = col\_character(),  
## Tm = col\_character(),  
## Lg = col\_character(),  
## Ht = col\_character()  
## )

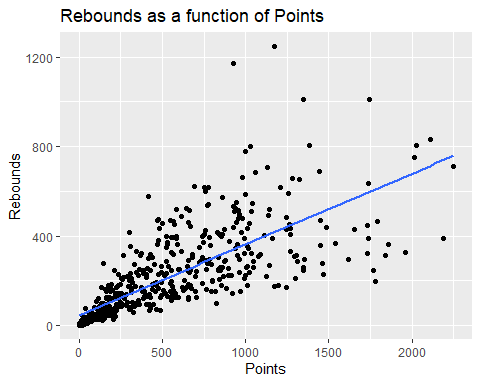
## See spec(...) for full column specifications.

nba = data.frame(nba)  
nba\_tr = select(nba, PTS, TRB, MP, Age)  
head(nba\_tr)

## PTS TRB MP Age  
## 1 235 144 337 29  
## 2 1088 316 1553 22  
## 3 214 246 729 31  
## 4 531 361 2069 20  
## 5 300 418 1151 35  
## 6 1258 592 2408 33

# Q1

gg = ggplot(nba\_tr, aes(x = PTS, y = TRB)) + geom\_point() + geom\_smooth(method = "lm", se = FALSE) + labs(x = "Points", y = "Rebounds", title = "Rebounds as a function of Points")  
gg



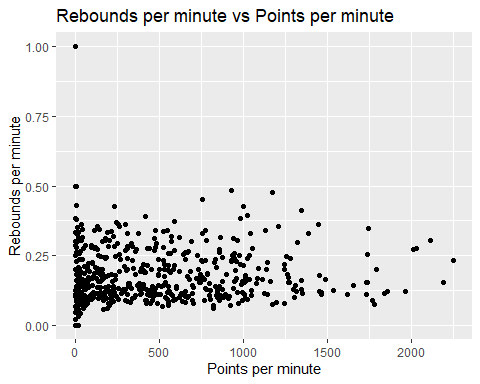
There seems to be strong relationship between rebounds and points. As the line is steadily increasing.

# Q2)

nba\_tr$RPM = nba\_tr$TRB / nba\_tr$MP  
nba\_tr$PPM = nba\_tr$PTS / nba\_tr$MP  
head(nba\_tr)

## PTS TRB MP Age RPM PPM  
## 1 235 144 337 29 0.4272997 0.6973294  
## 2 1088 316 1553 22 0.2034771 0.7005795  
## 3 214 246 729 31 0.3374486 0.2935528  
## 4 531 361 2069 20 0.1744804 0.2566457  
## 5 300 418 1151 35 0.3631625 0.2606429  
## 6 1258 592 2408 33 0.2458472 0.5224252

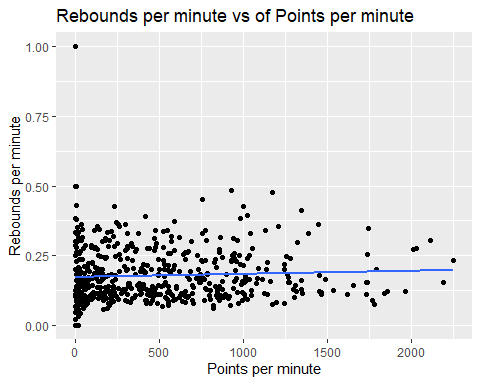
gg = ggplot(nba\_tr, aes(x = PTS, y = RPM)) + geom\_point() + labs(x = "Points per minute", y = "Rebounds per minute", title = "Rebounds per minute vs Points per minute")  
gg



As we can see the rebounds per minute is almost scattered horizontally with a very slight tendency to increase as points per minute increases.

# Q3

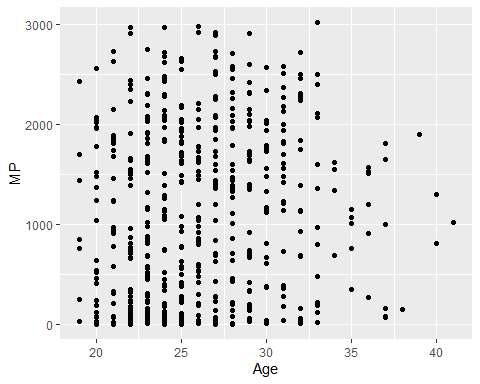
gg = ggplot(nba\_tr, aes(x = PTS, y = RPM)) + geom\_point() + geom\_smooth(method = "lm", se = FALSE) + labs(x = "Points per minute", y = "Rebounds per minute", title = "Rebounds per minute vs of Points per minute")  
gg



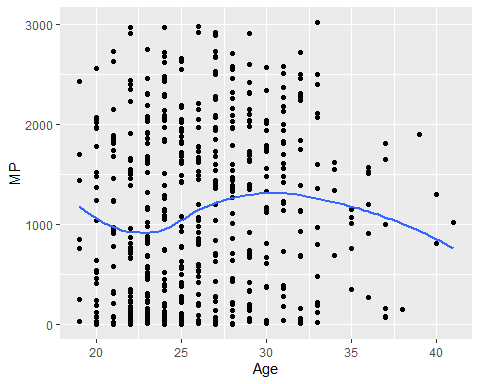
As we can see graphically that there is a somewhat weak relationship between them as the line is not completely horizontal but tilts upwards as points per minute increases. Thus the relationship is positive.

# Q4

gg = ggplot(nba\_tr, aes(x = Age, y = MP)) + geom\_point()  
gg

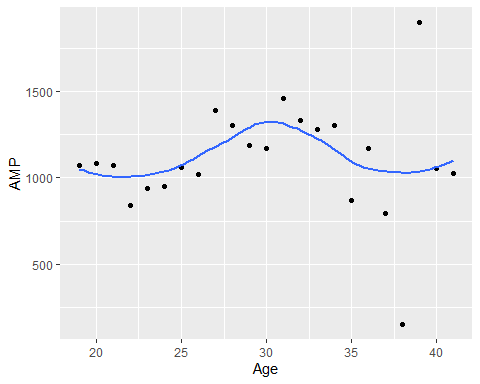


gg = ggplot(nba\_tr, aes(x = Age, y = MP)) + geom\_point() + geom\_smooth(method = "loess", se = FALSE)  
gg



nba\_4 = ddply(nba\_tr, ~Age, summarize, AMP = mean(MP))

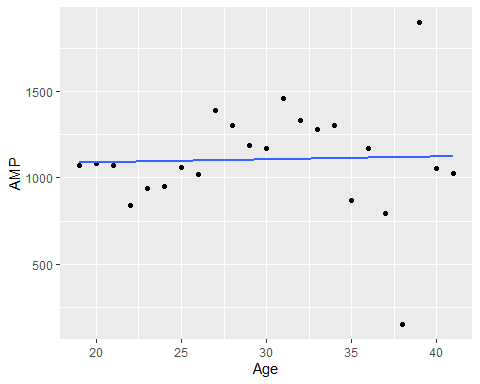
gg = ggplot(nba\_4, aes(x = Age, y = AMP)) + geom\_point() + geom\_smooth(method = "loess", se = FALSE)  
gg



As seen we can see that young players get less time on court but as they gain experience they get more time on court. But as they age more their time on court decreases again.

# Q5

gg = ggplot(nba\_4, aes(x = Age, y = AMP)) + geom\_point() + geom\_smooth(method = "lm", se = FALSE)  
gg



This line tells us that as players age, their time on court increases.