HomeWork#5

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.4  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(ggplot2)  
library(dplyr)  
library(Lahman)  
view(Batting)

#Ques1  
is\_tibble(Batting)

## [1] FALSE

#Since a tibble is characterized by not more than 10 rows, by running  
#r code above, we conclude that database "Battling" is not a tibble  
#since it clearly consists of > 10 rows. This is confirmed with the   
#output result: FALSE

#Ques2:  
?Batting

## starting httpd help server ... done

#Rows/Observations: 105861   
#Columns/Variables: 22

#Ques3:  
as\_tibble(Batting)

## # A tibble: 105,861 x 22  
## playerID yearID stint teamID lgID G AB R H X2B X3B HR  
## <chr> <int> <int> <fct> <fct> <int> <int> <int> <int> <int> <int> <int>  
## 1 abercda~ 1871 1 TRO NA 1 4 0 0 0 0 0  
## 2 addybo01 1871 1 RC1 NA 25 118 30 32 6 0 0  
## 3 allisar~ 1871 1 CL1 NA 29 137 28 40 4 5 0  
## 4 allisdo~ 1871 1 WS3 NA 27 133 28 44 10 2 2  
## 5 ansonca~ 1871 1 RC1 NA 25 120 29 39 11 3 0  
## 6 armstbo~ 1871 1 FW1 NA 12 49 9 11 2 1 0  
## 7 barkeal~ 1871 1 RC1 NA 1 4 0 1 0 0 0  
## 8 barnero~ 1871 1 BS1 NA 31 157 66 63 10 9 0  
## 9 barrebi~ 1871 1 FW1 NA 1 5 1 1 1 0 0  
## 10 barrofr~ 1871 1 BS1 NA 18 86 13 13 2 1 0  
## # ... with 105,851 more rows, and 10 more variables: RBI <int>, SB <int>,  
## # CS <int>, BB <int>, SO <int>, IBB <int>, HBP <int>, SH <int>, SF <int>,  
## # GIDP <int>

#Ques4:  
Batting%>%  
 select(playerID,teamID,yearID,HR)%>%  
 filter(yearID == 1991, HR>30) -> new1991  
new1991

## playerID teamID yearID HR  
## 1 cansejo01 OAK 1991 44  
## 2 cartejo01 TOR 1991 33  
## 3 dawsoan01 CHN 1991 31  
## 4 fieldce01 DET 1991 44  
## 5 gantro01 ATL 1991 32  
## 6 johnsho01 NYN 1991 38  
## 7 mcgrifr01 SDN 1991 31  
## 8 ripkeca01 BAL 1991 34  
## 9 tartada01 KCA 1991 31  
## 10 tettlmi01 DET 1991 31  
## 11 thomafr04 CHA 1991 32  
## 12 willima04 SFN 1991 34

#Ques5:  
mean(new1991$HR)

## [1] 34.58333

#Ques6:  
Batting%>%  
 select(playerID,yearID,HR)%>%  
 filter(yearID == 1990, HR>30) -> new1990  
new1990

## playerID yearID HR  
## 1 bondsba01 1990 33  
## 2 bonilbo01 1990 32  
## 3 cansejo01 1990 37  
## 4 fieldce01 1990 51  
## 5 gantro01 1990 32  
## 6 grubeke01 1990 31  
## 7 mcgrifr01 1990 35  
## 8 mcgwima01 1990 39  
## 9 mitchke01 1990 35  
## 10 sandbry01 1990 40  
## 11 strawda01 1990 37  
## 12 willima04 1990 33

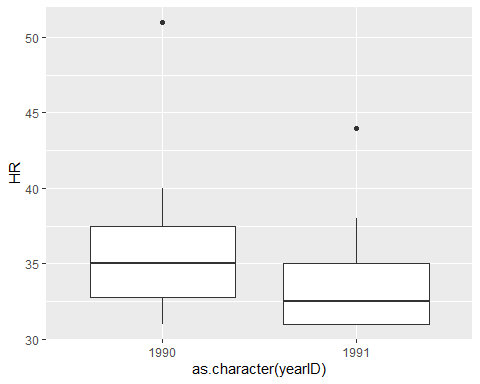
#Ques7:  
mean(new1990[[3]])

## [1] 36.25

#Ques8:  
Batting%>%  
 select(playerID,teamID,yearID,HR)%>%  
 filter(yearID==1990|yearID==1991, HR>30) ->N  
N

## playerID teamID yearID HR  
## 1 bondsba01 PIT 1990 33  
## 2 bonilbo01 PIT 1990 32  
## 3 cansejo01 OAK 1990 37  
## 4 fieldce01 DET 1990 51  
## 5 gantro01 ATL 1990 32  
## 6 grubeke01 TOR 1990 31  
## 7 mcgrifr01 TOR 1990 35  
## 8 mcgwima01 OAK 1990 39  
## 9 mitchke01 SFN 1990 35  
## 10 sandbry01 CHN 1990 40  
## 11 strawda01 NYN 1990 37  
## 12 willima04 SFN 1990 33  
## 13 cansejo01 OAK 1991 44  
## 14 cartejo01 TOR 1991 33  
## 15 dawsoan01 CHN 1991 31  
## 16 fieldce01 DET 1991 44  
## 17 gantro01 ATL 1991 32  
## 18 johnsho01 NYN 1991 38  
## 19 mcgrifr01 SDN 1991 31  
## 20 ripkeca01 BAL 1991 34  
## 21 tartada01 KCA 1991 31  
## 22 tettlmi01 DET 1991 31  
## 23 thomafr04 CHA 1991 32  
## 24 willima04 SFN 1991 34

ggplot(data=N)+  
 geom\_boxplot(mapping=aes(x=as.character(yearID), y=HR))



#Ques9:  
#Using DATA "N"  
ggplot(data = N)+  
 geom\_boxplot(mapping=aes(x=teamID, y=HR))+  
 xlab("teamID")+  
 ylab("HR")

