Problem Set 1

Pat Ausman

Total Rcode is at the bottom. Snippets included in each question

1) *Include in your Problem Set a script of R code (it can be within the same file as the script from class) that obtains the answers to the following questions using the Lahman database:*
   a) *Answer the following questions (and show your R code):*
      i) *How many players with the last name "Saltalamacchia" have played MLB?*

```
> who_to_look_for = "Saltalamacchia"
> Salty <- Master %>% select(nameLast) %>% filter(nameLast == who_to_look_for)
> Salty
          nameLast
         1 Saltalamacchia
> # output the number
> cat("The number of players with a last name of",who_to_look_for,"is",count(Salty)[[1]])
The number of players with a last name of Saltalamacchia is 1
```

      ii) *How many players with the first name "Jarrod" have played MLB?*

```
> who_to_look_for = "Jarrod"
> Salty <- Master %>% select(nameFirst) %>% filter(nameFirst == who_to_look_for)
> Salty
  nameFirst
1    Jarrod
2    Jarrod
3    Jarrod
4    Jarrod
5    Jarrod
> # output the number
> cat("The number of players with a first name of",who_to_look_for,"is",count(Salty)[[1]])
The number of players with a first name of Jarrod is 5
```

      iii) *What is the name of the player with the playerID "parkeja01" (you may have seen a different playerID, "parkeja02," in a previous result)?*

```
> who_to_look_for = "parkeja01"
> results <- Master %>% select(playerID, nameFirst, nameLast) %>% filter(playerID ==
who_to_look_for)
> results
  playerID nameFirst nameLast
1 parkeja01       Jay   Parker
> # output the number - assuming there is only one record
>   cat("The name of the player with a playerID
of",who_to_look_for,"is",results[[2]],results[[3]])
The name of the player with a playerID of parkeja01 is Jay Parker
```

Problem Set 1
Pat Ausman

b) Go to retrosheet and download game logs from any year in baseball (the yearly links appear below the section on that page called "Game Logs for Individual Seasons.") and unzip the file, and move the ".TXT" file in your RStudio Working Directory. Read that file in RStudio and re-run the appropriate commands from the R script from class on your chosen year. Include the R script snippet in your Problem Set.

Downloaded 2015 Game logs. Looked at Patriots Day 2015 Bos vs Bal. Bos won 7-1

```
> # downloaded the game log from 2015. Use the headers for input since they are much clearer
>   GL.headers = read_csv('game_log_header.csv')
Parsed with column specification:
cols(
  .default = col_character()
)
See spec(...) for full column specifications.
> GL.2015 = read_csv('GL2015.TXT',
+                    col_names = names(GL.headers))
Parsed with column specification:
cols(
  .default = col_double(),
  DayOfWeek = col_character(),
  VisitingTeam = col_character(),
  VisitingTeamLeague = col_character(),
  HomeTeam = col_character(),
  HomeTeamLeague = col_character(),
  DayNight = col_character(),
  CompletionInfo = col_number(),
  ForfeitInfo = col_logical(),
  ProtestInfo = col_character(),
  ParkID = col_character(),
  VisitorLineScore = col_character(),
  HomeLineScore = col_character(),
  UmpireHID = col_character(),
  UmpireHName = col_character(),
  Umpire1BID = col_character(),
  Umpire1BName = col_character(),
  Umpire2BID = col_character(),
  Umpire2BName = col_character(),
  Umpire3BID = col_character(),
  Umpire3BName = col_character()
  # ... with 58 more columns
)
See spec(...) for full column specifications.
> names(GL.2015)
  [1] "Date"                  "DoubleHeader"          "DayOfWeek"
  [4] "VisitingTeam"          "VisitingTeamLeague"    "VisitingTeamGameNumber"
  [7] "HomeTeam"              "HomeTeamLeague"        "HomeTeamGameNumber"
 [10] "VisitorRunsScored"     "HomeRunsScore"         "LengthInOuts"
 [13] "DayNight"              "CompletionInfo"        "ForfeitInfo"
 [16] "ProtestInfo"           "ParkID"                "Attendance"
 [19] "Duration"              "VisitorLineScore"      "HomeLineScore"
 [22] "VisitorAB"             "VisitorH"              "VisitorD"
 [25] "VisitorT"              "VisitorHR"             "VisitorRBI"
 [28] "VisitorSH"             "VisitorSF"             "VisitorHBP"
```

Problem Set 1
Pat Ausman

```
 [31] "VisitorBB"              "VisitorIBB"              "VisitorK"
 [34] "VisitorSB"              "VisitorCS"               "VisitorGDP"
 [37] "VisitorCI"              "VisitorLOB"              "VisitorPitchers"
 [40] "VisitorER"              "VisitorTER"              "VisitorWP"
 [43] "VisitorBalks"           "VisitorPO"               "VisitorA"
 [46] "VisitorE"               "VisitorPassed"           "VisitorDB"
 [49] "VisitorTP"              "HomeAB"                  "HomeH"
 [52] "HomeD"                  "HomeT"                   "HomeHR"
 [55] "HomeRBI"                "HomeSH"                  "HomeSF"
 [58] "HomeHBP"                "HomeBB"                  "HomeIBB"
 [61] "HomeK"                  "HomeSB"                  "HomeCS"
 [64] "HomeGDP"                "HomeCI"                  "HomeLOB"
 [67] "HomePitchers"           "HomeER"                  "HomeTER"
 [70] "HomeWP"                 "HomeBalks"               "HomePO"
 [73] "HomeA"                  "HomeE"                   "HomePassed"
 [76] "HomeDB"                 "HomeTP"                  "UmpireHID"
 [79] "UmpireHName"            "Umpire1BID"              "Umpire1BName"
 [82] "Umpire2BID"             "Umpire2BName"            "Umpire3BID"
 [85] "Umpire3BName"           "UmpireLFID"              "UmpireLFName"
 [88] "UmpireRFID"             "UmpireRFName"            "VisitorManagerID"
 [91] "VisitorManagerName"     "HomeManagerID"           "HomeManagerName"
 [94] "WinningPitcherID"       "WinningPitcherName"      "LosingPitcherID"
 [97] "LosingPitcherNAme"      "SavingPitcherID"         "SavingPitcherName"
[100] "GameWinningRBIID"       "GameWinningRBIName"      "VisitorStartingPitcherID"
[103] "VisitorStartingPitcherName" "HomeStartingPitcherID"  "HomeStartingPitcherName"
[106] "VisitorBatting1PlayerID"  "VisitorBatting1Name"   "VisitorBatting1Position"
[109] "VisitorBatting2PlayerID"  "VisitorBatting2Name"   "VisitorBatting2Position"
[112] "VisitorBatting3PlayerID"  "VisitorBatting3Name"   "VisitorBatting3Position"
[115] "VisitorBatting4PlayerID"  "VisitorBatting4Name"   "VisitorBatting4Position"
[118] "VisitorBatting5PlayerID"  "VisitorBatting5Name"   "VisitorBatting5Position"
[121] "VisitorBatting6PlayerID"  "VisitorBatting6Name"   "VisitorBatting6Position"
[124] "VisitorBatting7PlayerID"  "VisitorBatting7Name"   "VisitorBatting7Position"
[127] "VisitorBatting8PlayerID"  "VisitorBatting8Name"   "VisitorBatting8Position"
[130] "VisitorBatting9PlayerID"  "VisitorBatting9Name"   "VisitorBatting9Position"
[133] "HomeBatting1PlayerID"     "HomeBatting1Name"      "HomeBatting1Position"
[136] "HomeBatting2PlayerID"     "HomeBatting2Name"      "HomeBatting2Position"
[139] "HomeBatting3PlayerID"     "HomeBatting3Name"      "HomeBatting3Position"
[142] "HomeBatting4PlayerID"     "HomeBatting4Name"      "HomeBatting4Position"
[145] "HomeBatting5PlayerID"     "HomeBatting5Name"      "HomeBatting5Position"
[148] "HomeBatting6PlayerID"     "HomeBatting6Name"      "HomeBatting6Position"
[151] "HomeBatting7PlayerID"     "HomeBatting7Name"      "HomeBatting7Position"
[154] "HomeBatting8PlayerID"     "HomeBatting8Name"      "HomeBatting8Position"
[157] "HomeBatting9PlayerID"     "HomeBatting9Name"      "HomeBatting9Position"
[160] "AdditionalInfo"           "AcquisitionInfo"
> head(GL.2015)
# A tibble: 6 x 161
     Date DoubleHeader DayOfWeek VisitingTeam VisitingTeamLea… VisitingTeamGam… HomeTeam
    <dbl>        <dbl> <chr>     <chr>        <chr>                      <dbl> <chr>
1 2.02e7            0 Sun       SLN          NL                             1 CHN
2 2.02e7            0 Mon       MIN          AL                             1 DET
3 2.02e7            0 Mon       CLE          AL                             1 HOU
```

Problem Set 1
Pat Ausman

```
4 2.02e7          0 Mon       CHA         AL                      1 KCA
5 2.02e7          0 Mon       TOR         AL                      1 NYA
6 2.02e7          0 Mon       TEX         AL                      1 OAK
# … with 154 more variables: HomeTeamLeague <chr>, HomeTeamGameNumber <dbl>,
#   VisitorRunsScored <dbl>, HomeRunsScore <dbl>, LengthInOuts <dbl>, DayNight <chr>,
#   CompletionInfo <dbl>, ForfeitInfo <lgl>, ProtestInfo <chr>, ParkID <chr>, Attendance <dbl>,
#   Duration <dbl>, VisitorLineScore <chr>, HomeLineScore <chr>, VisitorAB <dbl>, VisitorH <dbl>,
#   VisitorD <dbl>, VisitorT <dbl>, VisitorHR <dbl>, VisitorRBI <dbl>, VisitorSH <dbl>,
#   VisitorSF <dbl>, VisitorHBP <dbl>, VisitorBB <dbl>, VisitorIBB <dbl>, VisitorK <dbl>,
#   VisitorSB <dbl>, VisitorCS <dbl>, VisitorGDP <dbl>, VisitorCI <dbl>, VisitorLOB <dbl>,
#   VisitorPitchers <dbl>, VisitorER <dbl>, VisitorTER <dbl>, VisitorWP <dbl>, VisitorBalks <dbl>,
#   VisitorPO <dbl>, VisitorA <dbl>, VisitorE <dbl>, VisitorPassed <dbl>, VisitorDB <dbl>,
#   VisitorTP <dbl>, HomeAB <dbl>, HomeH <dbl>, HomeD <dbl>, HomeT <dbl>, HomeHR <dbl>,
#   HomeRBI <dbl>, HomeSH <dbl>, HomeSF <dbl>, HomeHBP <dbl>, HomeBB <dbl>, HomeIBB <dbl>,
#   HomeK <dbl>, HomeSB <dbl>, HomeCS <dbl>, HomeGDP <dbl>, HomeCI <dbl>, HomeLOB <dbl>,
#   HomePitchers <dbl>, HomeER <dbl>, HomeTER <dbl>, HomeWP <dbl>, HomeBalks <dbl>, HomePO
<dbl>,
#   HomeA <dbl>, HomeE <dbl>, HomePassed <dbl>, HomeDB <dbl>, HomeTP <dbl>, UmpireHID <chr>,
#   UmpireHName <chr>, Umpire1BID <chr>, Umpire1BName <chr>, Umpire2BID <chr>, Umpire2BName
<chr>,
#   Umpire3BID <chr>, Umpire3BName <chr>, UmpireLFID <lgl>, UmpireLFName <chr>, UmpireRFID <lgl>,
#   UmpireRFName <chr>, VisitorManagerID <chr>, VisitorManagerName <chr>, HomeManagerID <chr>,
#   HomeManagerName <chr>, WinningPitcherID <chr>, WinningPitcherName <chr>, LosingPitcherID <chr>,
#   LosingPitcherNAme <chr>, SavingPitcherID <chr>, SavingPitcherName <chr>, GameWinningRBIID
<chr>,
#   GameWinningRBIName <chr>, VisitorStartingPitcherID <chr>, VisitorStartingPitcherName <chr>,
#   HomeStartingPitcherID <chr>, HomeStartingPitcherName <chr>, VisitorBatting1PlayerID <chr>,
#   VisitorBatting1Name <chr>, …
> tail(GL.2015)
# A tibble: 6 x 161
     Date DoubleHeader DayOfWeek VisitingTeam VisitingTeamLea… VisitingTeamGam… HomeTeam
    <dbl>        <dbl> <chr>     <chr>        <chr>                      <dbl> <chr>
1 2.02e7            0 Sun       SDN          NL                           162 LAN
2 2.02e7            0 Sun       CHN          NL                           162 MIL
3 2.02e7            0 Sun       WAS          NL                           162 NYN
4 2.02e7            0 Sun       MIA          NL                           162 PHI
5 2.02e7            0 Sun       CIN          NL                           162 PIT
6 2.02e7            0 Sun       COL          NL                           162 SFN
# … with 154 more variables: HomeTeamLeague <chr>, HomeTeamGameNumber <dbl>,
#   VisitorRunsScored <dbl>, HomeRunsScore <dbl>, LengthInOuts <dbl>, DayNight <chr>,
#   CompletionInfo <dbl>, ForfeitInfo <lgl>, ProtestInfo <chr>, ParkID <chr>, Attendance <dbl>,
#   Duration <dbl>, VisitorLineScore <chr>, HomeLineScore <chr>, VisitorAB <dbl>, VisitorH <dbl>,
#   VisitorD <dbl>, VisitorT <dbl>, VisitorHR <dbl>, VisitorRBI <dbl>, VisitorSH <dbl>,
#   VisitorSF <dbl>, VisitorHBP <dbl>, VisitorBB <dbl>, VisitorIBB <dbl>, VisitorK <dbl>,
#   VisitorSB <dbl>, VisitorCS <dbl>, VisitorGDP <dbl>, VisitorCI <dbl>, VisitorLOB <dbl>,
#   VisitorPitchers <dbl>, VisitorER <dbl>, VisitorTER <dbl>, VisitorWP <dbl>, VisitorBalks <dbl>,
#   VisitorPO <dbl>, VisitorA <dbl>, VisitorE <dbl>, VisitorPassed <dbl>, VisitorDB <dbl>,
#   VisitorTP <dbl>, HomeAB <dbl>, HomeH <dbl>, HomeD <dbl>, HomeT <dbl>, HomeHR <dbl>,
#   HomeRBI <dbl>, HomeSH <dbl>, HomeSF <dbl>, HomeHBP <dbl>, HomeBB <dbl>, HomeIBB <dbl>,
#   HomeK <dbl>, HomeSB <dbl>, HomeCS <dbl>, HomeGDP <dbl>, HomeCI <dbl>, HomeLOB <dbl>,
```

Problem Set 1
Pat Ausman

```
#   HomePitchers <dbl>, HomeER <dbl>, HomeTER <dbl>, HomeWP <dbl>, HomeBalks <dbl>, HomePO
<dbl>,
#   HomeA <dbl>, HomeE <dbl>, HomePassed <dbl>, HomeDB <dbl>, HomeTP <dbl>, UmpireHID <chr>,
#   UmpireHName <chr>, Umpire1BID <chr>, Umpire1BName <chr>, Umpire2BID <chr>, Umpire2BName
<chr>,
#   Umpire3BID <chr>, Umpire3BName <chr>, UmpireLFID <lgl>, UmpireLFName <chr>, UmpireRFID <lgl>,
#   UmpireRFName <chr>, VisitorManagerID <chr>, VisitorManagerName <chr>, HomeManagerID <chr>,
#   HomeManagerName <chr>, WinningPitcherID <chr>, WinningPitcherName <chr>, LosingPitcherID <chr>,
#   LosingPitcherNAme <chr>, SavingPitcherID <chr>, SavingPitcherName <chr>, GameWinningRBIID
<chr>,
#   GameWinningRBIName <chr>, VisitorStartingPitcherID <chr>, VisitorStartingPitcherName <chr>,
#   HomeStartingPitcherID <chr>, HomeStartingPitcherName <chr>, VisitorBatting1PlayerID <chr>,
#   VisitorBatting1Name <chr>, …
> summary(GL.2015)
      Date          DoubleHeader      DayOfWeek         VisitingTeam       VisitingTeamLeague
 Min.   :20150405   Min.   :0.00000   Length:2429        Length:2429        Length:2429
 1st Qu.:20150521   1st Qu.:0.00000   Class :character   Class :character   Class :character
 Median :20150704   Median :0.00000   Mode  :character   Mode  :character   Mode  :character
 Mean   :20150682   Mean   :0.03582
 3rd Qu.:20150822   3rd Qu.:0.00000
 Max.   :20151004   Max.   :2.00000
 VisitingTeamGameNumber   HomeTeam          HomeTeamLeague     HomeTeamGameNumber VisitorRunsScored
 Min.   :  1.00         Length:2429        Length:2429         Min.   :  1.00     Min.   : 0.000
 1st Qu.: 41.00         Class :character   Class :character    1st Qu.: 41.00     1st Qu.: 2.000
 Median : 82.00         Mode  :character   Mode  :character    Median : 81.00     Median : 4.000
 Mean   : 81.47                                                Mean   : 81.46     Mean   : 4.154
 3rd Qu.:122.00                                                3rd Qu.:122.00     3rd Qu.: 6.000
 Max.   :162.00                                                Max.   :162.00     Max.   :21.000
 HomeRunsScore     LengthInOuts      DayNight          CompletionInfo      ForfeitInfo
 Min.   : 0.000   Min.   : 31.00   Length:2429        Min.   :2.015e+11   Mode:logical
 1st Qu.: 2.000   1st Qu.: 51.00   Class :character   1st Qu.:2.015e+11   NA's:2429
 Median : 4.000   Median : 54.00   Mode  :character   Median :2.015e+11
 Mean   : 4.346   Mean   : 53.61                      Mean   :2.015e+11
 3rd Qu.: 6.000   3rd Qu.: 54.00                      3rd Qu.:2.015e+11
 Max.   :22.000   Max.   :114.00                      Max.   :2.015e+11
 ProtestInfo          ParkID           Attendance       Duration      VisitorLineScore
 Length:2429        Length:2429        Min.   :    0   Min.   :100.0   Length:2429
 Class :character   Class :character   1st Qu.:22584   1st Qu.:163.0   Class :character
 Mode  :character   Mode  :character   Median :30792   Median :177.0   Mode  :character
                                       Mean   :30379   Mean   :180.8
                                       3rd Qu.:38098   3rd Qu.:194.0
                                       Max.   :53518   Max.   :409.0
 HomeLineScore        VisitorAB        VisitorH         VisitorD         VisitorT
 Length:2429        Min.   :17.00   Min.   : 0.000   Min.   :0.000   Min.   :0.0000
 Class :character   1st Qu.:32.00   1st Qu.: 6.000   1st Qu.:1.000   1st Qu.:0.0000
 Mode  :character   Median :34.00   Median : 8.000   Median :1.000   Median :0.0000
                    Mean   :34.82   Mean   : 8.669   Mean   :1.684   Mean   :0.1816
                    3rd Qu.:37.00   3rd Qu.:11.000   3rd Qu.:2.000   3rd Qu.:0.0000
                    Max.   :71.00   Max.   :23.000   Max.   :9.000   Max.   :4.0000
  VisitorHR        VisitorRBI        VisitorSH        VisitorSF        VisitorHBP
 Min.   :0.000   Min.   : 0.000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
 1st Qu.:0.000   1st Qu.: 2.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
 Median :1.000   Median : 3.000   Median :0.0000   Median :0.0000   Median :0.0000
 Mean   :1.003   Mean   : 3.942   Mean   :0.2528   Mean   :0.2417   Mean   :0.3413
 3rd Qu.:2.000   3rd Qu.: 6.000   3rd Qu.:0.0000   3rd Qu.:0.0000   3rd Qu.:1.0000
 Max.   :8.000   Max.   :20.000   Max.   :4.0000   Max.   :4.0000   Max.   :4.0000
  VisitorBB        VisitorIBB        VisitorK         VisitorSB        VisitorCS
 Min.   : 0.000   Min.   :0.0000   Min.   : 1.00   Min.   :0.0000   Min.   :0.0000
```

Problem Set 1

Pat Ausman

```
1st Qu.: 1.000   1st Qu.:0.0000   1st Qu.: 6.00   1st Qu.:0.0000   1st Qu.:0.0000
Median : 3.000   Median :0.0000   Median : 8.00   Median :0.0000   Median :0.0000
Mean   : 2.832   Mean   :0.1853   Mean   : 8.03   Mean   :0.5228   Mean   :0.2301
3rd Qu.: 4.000   3rd Qu.:0.0000   3rd Qu.:10.00   3rd Qu.:1.0000   3rd Qu.:0.0000
Max.   :13.000   Max.   :4.0000   Max.   :19.00   Max.   :7.0000   Max.   :3.0000
  VisitorGDP       VisitorCI         VisitorLOB     VisitorPitchers   VisitorER
Min.   :0.0000   Min.   :0.000000   Min.   : 0.000   Min.   : 1.00   Min.   : 0.000
1st Qu.:0.0000   1st Qu.:0.000000   1st Qu.: 5.000   1st Qu.: 3.00   1st Qu.: 2.000
Median :1.0000   Median :0.000000   Median : 7.000   Median : 4.00   Median : 4.000
Mean   :0.7843   Mean   :0.008646   Mean   : 6.772   Mean   : 4.01   Mean   : 4.022
3rd Qu.:1.0000   3rd Qu.:0.000000   3rd Qu.: 8.000   3rd Qu.: 5.00   3rd Qu.: 6.000
Max.   :5.0000   Max.   :2.000000   Max.   :25.000   Max.   :13.00   Max.   :20.000
  VisitorTER       VisitorWP       VisitorBalks       VisitorPO        VisitorA
Min.   : 0.000   Min.   :0.0000   Min.   :0.00000   Min.   :15.00   Min.   : 1.000
1st Qu.: 2.000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:24.00   1st Qu.: 8.000
Median : 4.000   Median :0.0000   Median :0.00000   Median :27.00   Median :10.000
Mean   : 4.016   Mean   :0.3697   Mean   :0.02635   Mean   :26.04   Mean   : 9.828
3rd Qu.: 6.000   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:27.00   3rd Qu.:12.000
Max.   :20.000   Max.   :4.0000   Max.   :2.00000   Max.   :57.00   Max.   :23.000
   VisitorE      VisitorPassed       VisitorDB        VisitorTP            HomeAB
Min.   :0.0000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000000   Min.   :18.00
1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.0000000   1st Qu.:31.00
Median :0.0000   Median :0.00000   Median :1.0000   Median :0.0000000   Median :33.00
Mean   :0.5932   Mean   :0.05846   Mean   :0.8872   Mean   :0.0004117   Mean   :33.31
3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.0000000   3rd Qu.:35.00
Max.   :4.0000   Max.   :3.00000   Max.   :5.0000   Max.   :1.0000000   Max.   :67.00
    HomeH           HomeD           HomeT          HomeHR         HomeRBI          HomeSH
Min.   : 0.000   Min.   :0.000   Min.   :0.000   Min.   :0.000   Min.   : 0.000   Min.   :0.0000
1st Qu.: 6.000   1st Qu.:1.000   1st Qu.:0.000   1st Qu.:0.000   1st Qu.: 2.000   1st Qu.:0.0000
Median : 8.000   Median :1.000   Median :0.000   Median :1.000   Median : 4.000   Median :0.0000
Mean   : 8.666   Mean   :1.709   Mean   :0.205   Mean   :1.018   Mean   : 4.148   Mean   :0.2413
3rd Qu.:11.000   3rd Qu.:2.000   3rd Qu.:0.000   3rd Qu.:2.000   3rd Qu.: 6.000   3rd Qu.:0.0000
Max.   :26.000   Max.   :8.000   Max.   :4.000   Max.   :8.000   Max.   :22.000   Max.   :3.0000
    HomeSF          HomeHBP          HomeBB          HomeIBB          HomeK
Min.   :0.0000   Min.   :0.0000   Min.   : 0.000   Min.   :0.0000   Min.   : 0.000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 2.000   1st Qu.:0.0000   1st Qu.: 5.000
Median :0.0000   Median :0.0000   Median : 3.000   Median :0.0000   Median : 7.000
Mean   :0.2655   Mean   :0.3182   Mean   : 2.962   Mean   :0.2063   Mean   : 7.387
3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.: 4.000   3rd Qu.:0.0000   3rd Qu.: 9.000
Max.   :3.0000   Max.   :4.0000   Max.   :12.000   Max.   :3.0000   Max.   :19.000
    HomeSB          HomeCS          HomeGDP          HomeCI          HomeLOB
Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.00000   Min.   : 0.000
1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.00000   1st Qu.: 5.000
Median :0.0000   Median :0.0000   Median :1.000   Median :0.00000   Median : 7.000
Mean   :0.5084   Mean   :0.2079   Mean   :0.755   Mean   :0.00494   Mean   : 6.715
3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:1.000   3rd Qu.:0.00000   3rd Qu.: 8.000
Max.   :6.0000   Max.   :3.0000   Max.   :5.000   Max.   :1.00000   Max.   :19.000
 HomePitchers       HomeER          HomeTER          HomeWP          HomeBalks          HomePO
Min.   : 1.00   Min.   : 0.000   Min.   : 0.000   Min.   :0.0000   Min.   :0.0000   Min.   :16.00
1st Qu.: 3.00   1st Qu.: 2.000   1st Qu.: 2.000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:27.00
Median : 4.00   Median : 3.000   Median : 3.000   Median :0.0000   Median :0.0000   Median :27.00
Mean   : 4.21   Mean   : 3.845   Mean   : 3.837   Mean   :0.3541   Mean   :0.0317   Mean   :27.57
3rd Qu.: 5.00   3rd Qu.: 5.000   3rd Qu.: 5.000   3rd Qu.:1.0000   3rd Qu.:0.0000   3rd Qu.:27.00
Max.   :11.00   Max.   :20.000   Max.   :20.000   Max.   :4.0000   Max.   :2.0000   Max.   :57.00
    HomeA           HomeE         HomePassed         HomeDB          HomeTP
Min.   : 2.00   Min.   :0.000   Min.   :0.00000   Min.   :0.0000   Min.   :0.000000
1st Qu.: 8.00   1st Qu.:0.000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:0.000000
Median :10.00   Median :0.000   Median :0.00000   Median :1.0000   Median :0.000000
Mean   :10.24   Mean   :0.571   Mean   :0.07369   Mean   :0.9152   Mean   :0.001235
3rd Qu.:12.00   3rd Qu.:1.000   3rd Qu.:0.00000   3rd Qu.:1.0000   3rd Qu.:0.000000
Max.   :26.00   Max.   :4.000   Max.   :4.00000   Max.   :5.0000   Max.   :1.000000
```

Problem Set 1

Pat Ausman

```
 UmpireHID        UmpireHName        Umpire1BID       Umpire1BName       Umpire2BID
Length:2429       Length:2429       Length:2429       Length:2429       Length:2429
Class :character  Class :character  Class :character  Class :character  Class :character
Mode  :character  Mode  :character  Mode  :character  Mode  :character  Mode  :character
 Umpire2BName       Umpire3BID       Umpire3BName       UmpireLFID     UmpireLFName
Length:2429       Length:2429       Length:2429       Mode:logical   Length:2429
Class :character  Class :character  Class :character  NA's:2429      Class :character
Mode  :character  Mode  :character  Mode  :character                 Mode  :character
UmpireRFID     UmpireRFName       VisitorManagerID   VisitorManagerName HomeManagerID
Mode:logical   Length:2429        Length:2429        Length:2429        Length:2429
NA's:2429      Class :character   Class :character   Class :character   Class :character
               Mode  :character   Mode  :character   Mode  :character   Mode  :character
HomeManagerName    WinningPitcherID   WinningPitcherName LosingPitcherID    LosingPitcherNAme
Length:2429        Length:2429        Length:2429        Length:2429        Length:2429
Class :character   Class :character   Class :character   Class :character   Class :character
Mode  :character   Mode  :character   Mode  :character   Mode  :character   Mode  :character
SavingPitcherID    SavingPitcherName  GameWinningRBIID   GameWinningRBIName
Length:2429        Length:2429        Length:2429        Length:2429
Class :character   Class :character   Class :character   Class :character
Mode  :character   Mode  :character   Mode  :character   Mode  :character
VisitorStartingPitcherID VisitorStartingPitcherName HomeStartingPitcherID HomeStartingPitcherName
Length:2429              Length:2429                Length:2429            Length:2429
Class :character         Class :character           Class :character       Class :character
Mode  :character         Mode  :character           Mode  :character       Mode  :character
VisitorBatting1PlayerID VisitorBatting1Name VisitorBatting1Position VisitorBatting2PlayerID
Length:2429             Length:2429         Min.   : 2.000          Length:2429
Class :character        Class :character    1st Qu.: 5.000          Class :character
Mode  :character        Mode  :character    Median : 8.000          Mode  :character
                                            Mean   : 6.814
                                            3rd Qu.: 8.000
                                            Max.   :10.000
VisitorBatting2Name VisitorBatting2Position VisitorBatting3PlayerID VisitorBatting3Name
Length:2429         Min.   : 2.000          Length:2429             Length:2429
Class :character    1st Qu.: 4.000          Class :character        Class :character
Mode  :character    Median : 6.000          Mode  :character        Mode  :character
                    Mean   : 5.949
                    3rd Qu.: 8.000
                    Max.   :10.000
VisitorBatting3Position VisitorBatting4PlayerID VisitorBatting4Name VisitorBatting4Position
Min.   : 2.000          Length:2429             Length:2429         Min.   : 2.000
1st Qu.: 4.000          Class :character        Class :character    1st Qu.: 3.000
Median : 7.000          Mode  :character        Mode  :character    Median : 5.000
Mean   : 6.483                                                      Mean   : 5.867
3rd Qu.: 9.000                                                      3rd Qu.: 9.000
Max.   :10.000                                                      Max.   :10.000
VisitorBatting5PlayerID VisitorBatting5Name VisitorBatting5Position VisitorBatting6PlayerID
Length:2429             Length:2429         Min.   : 2.000          Length:2429
Class :character        Class :character    1st Qu.: 3.000          Class :character
Mode  :character        Mode  :character    Median : 5.000          Mode  :character
                                            Mean   : 5.705
                                            3rd Qu.: 8.000
                                            Max.   :10.000
VisitorBatting6Name VisitorBatting6Position VisitorBatting7PlayerID VisitorBatting7Name
Length:2429         Min.   : 2.000          Length:2429             Length:2429
Class :character    1st Qu.: 3.000          Class :character        Class :character
Mode  :character    Median : 6.000          Mode  :character        Mode  :character
                    Mean   : 5.545
                    3rd Qu.: 7.000
                    Max.   :10.000
VisitorBatting7Position VisitorBatting8PlayerID VisitorBatting8Name VisitorBatting8Position
Min.   : 2.000          Length:2429             Length:2429         Min.   : 1.000
```

Problem Set 1

Pat Ausman

```
1st Qu.: 3.000         Class :character       Class :character    1st Qu.: 2.000
Median : 5.000         Mode  :character       Mode  :character    Median : 5.000
Mean   : 5.087                                                    Mean   : 4.788
3rd Qu.: 7.000                                                    3rd Qu.: 7.000
Max.   :10.000                                                    Max.   :10.000
VisitorBatting9PlayerID VisitorBatting9Name VisitorBatting9Position HomeBatting1PlayerID
Length:2429             Length:2429         Min.   : 1.000          Length:2429
Class :character        Class :character    1st Qu.: 1.000          Class :character
Mode  :character        Mode  :character    Median : 2.000          Mode  :character
                                            Mean   : 3.259
                                            3rd Qu.: 6.000
                                            Max.   :10.000
HomeBatting1Name   HomeBatting1Position HomeBatting2PlayerID HomeBatting2Name
Length:2429        Min.   : 2.00        Length:2429         Length:2429
Class :character   1st Qu.: 5.00        Class :character    Class :character
Mode  :character   Median : 8.00        Mode  :character    Mode  :character
                   Mean   : 6.81
                   3rd Qu.: 8.00
                   Max.   :10.00
HomeBatting2Position HomeBatting3PlayerID HomeBatting3Name   HomeBatting3Position
Min.   : 2.000       Length:2429          Length:2429        Min.   : 2.000
1st Qu.: 4.000       Class :character     Class :character   1st Qu.: 4.000
Median : 6.000       Mode  :character     Mode  :character   Median : 7.000
Mean   : 5.995                                               Mean   : 6.406
3rd Qu.: 8.000                                               3rd Qu.: 9.000
Max.   :10.000                                               Max.   :10.000
HomeBatting4PlayerID HomeBatting4Name   HomeBatting4Position HomeBatting5PlayerID
Length:2429          Length:2429        Min.   : 2.000       Length:2429
Class :character     Class :character   1st Qu.: 3.000       Class :character
Mode  :character     Mode  :character   Median : 5.000       Mode  :character
                                        Mean   : 5.996
                                        3rd Qu.: 9.000
                                        Max.   :10.000
HomeBatting5Name   HomeBatting5Position HomeBatting6PlayerID HomeBatting6Name
Length:2429        Min.   : 2.000       Length:2429         Length:2429
Class :character   1st Qu.: 3.000       Class :character    Class :character
Mode  :character   Median : 5.000       Mode  :character    Mode  :character
                   Mean   : 5.678
                   3rd Qu.: 8.000
                   Max.   :10.000
HomeBatting6Position HomeBatting7PlayerID HomeBatting7Name   HomeBatting7Position
Min.   : 2.000       Length:2429          Length:2429        Min.   : 2.00
1st Qu.: 3.000       Class :character     Class :character   1st Qu.: 3.00
Median : 5.000       Mode  :character     Mode  :character   Median : 5.00
Mean   : 5.518                                               Mean   : 5.14
3rd Qu.: 7.000                                               3rd Qu.: 7.00
Max.   :10.000                                               Max.   :10.00
HomeBatting8PlayerID HomeBatting8Name   HomeBatting8Position HomeBatting9PlayerID
Length:2429          Length:2429        Min.   : 1.00        Length:2429
Class :character     Class :character   1st Qu.: 2.00        Class :character
Mode  :character     Mode  :character   Median : 5.00        Mode  :character
                                        Mean   : 4.73
                                        3rd Qu.: 7.00
                                        Max.   :10.00
HomeBatting9Name   HomeBatting9Position AdditionalInfo     AcquisitionInfo
Length:2429        Min.   : 1.000       Length:2429         Length:2429
Class :character   1st Qu.: 1.000       Class :character    Class :character
Mode  :character   Median : 2.000       Mode  :character    Mode  :character
                   Mean   : 3.225
                   3rd Qu.: 6.000
                   Max.   :10.000
```

Problem Set 1

Pat Ausman

```
 [ reached getOption("max.print") -- omitted 1 row ]
> GL.2015 %>% filter(HomeTeam == 'BOS')
# A tibble: 81 x 161
    Date DoubleHeader DayOfWeek VisitingTeam VisitingTeamLea… VisitingTeamGam… HomeTeam
   <dbl>        <dbl> <chr>     <chr>        <chr>                      <dbl> <chr>
 1 2.02e7           0 Mon       WAS          NL                             7 BOS
 2 2.02e7           0 Tue       WAS          NL                             8 BOS
 3 2.02e7           0 Wed       WAS          NL                             9 BOS
 4 2.02e7           0 Fri       BAL          AL                            10 BOS
 5 2.02e7           0 Sat       BAL          AL                            11 BOS
 6 2.02e7           0 Sun       BAL          AL                            12 BOS
 7 2.02e7           0 Mon       BAL          AL                            13 BOS
 8 2.02e7           0 Mon       TOR          AL                            20 BOS
 9 2.02e7           0 Tue       TOR          AL                            21 BOS
10 2.02e7           0 Wed       TOR          AL                            22 BOS
# … with 71 more rows, and 154 more variables: HomeTeamLeague <chr>, HomeTeamGameNumber
<dbl>,
#   VisitorRunsScored <dbl>, HomeRunsScore <dbl>, LengthInOuts <dbl>, DayNight <chr>,
#   CompletionInfo <dbl>, ForfeitInfo <lgl>, ProtestInfo <chr>, ParkID <chr>, Attendance <dbl>,
#   Duration <dbl>, VisitorLineScore <chr>, HomeLineScore <chr>, VisitorAB <dbl>, VisitorH <dbl>,
#   VisitorD <dbl>, VisitorT <dbl>, VisitorHR <dbl>, VisitorRBI <dbl>, VisitorSH <dbl>,
#   VisitorSF <dbl>, VisitorHBP <dbl>, VisitorBB <dbl>, VisitorIBB <dbl>, VisitorK <dbl>,
#   VisitorSB <dbl>, VisitorCS <dbl>, VisitorGDP <dbl>, VisitorCI <dbl>, VisitorLOB <dbl>,
#   VisitorPitchers <dbl>, VisitorER <dbl>, VisitorTER <dbl>, VisitorWP <dbl>, VisitorBalks <dbl>,
#   VisitorPO <dbl>, VisitorA <dbl>, VisitorE <dbl>, VisitorPassed <dbl>, VisitorDB <dbl>,
#   VisitorTP <dbl>, HomeAB <dbl>, HomeH <dbl>, HomeD <dbl>, HomeT <dbl>, HomeHR <dbl>,
#   HomeRBI <dbl>, HomeSH <dbl>, HomeSF <dbl>, HomeHBP <dbl>, HomeBB <dbl>, HomeIBB <dbl>,
#   HomeK <dbl>, HomeSB <dbl>, HomeCS <dbl>, HomeGDP <dbl>, HomeCI <dbl>, HomeLOB <dbl>,
#   HomePitchers <dbl>, HomeER <dbl>, HomeTER <dbl>, HomeWP <dbl>, HomeBalks <dbl>, HomePO
<dbl>,
#   HomeA <dbl>, HomeE <dbl>, HomePassed <dbl>, HomeDB <dbl>, HomeTP <dbl>, UmpireHID <chr>,
#   UmpireHName <chr>, Umpire1BID <chr>, Umpire1BName <chr>, Umpire2BID <chr>, Umpire2BName
<chr>,
#   Umpire3BID <chr>, Umpire3BName <chr>, UmpireLFID <lgl>, UmpireLFName <chr>, UmpireRFID <lgl>,
#   UmpireRFName <chr>, VisitorManagerID <chr>, VisitorManagerName <chr>, HomeManagerID <chr>,
#   HomeManagerName <chr>, WinningPitcherID <chr>, WinningPitcherName <chr>, LosingPitcherID <chr>,
#   LosingPitcherNAme <chr>, SavingPitcherID <chr>, SavingPitcherName <chr>, GameWinningRBIID
<chr>,
#   GameWinningRBIName <chr>, VisitorStartingPitcherID <chr>, VisitorStartingPitcherName <chr>,
#   HomeStartingPitcherID <chr>, HomeStartingPitcherName <chr>, VisitorBatting1PlayerID <chr>,
#   VisitorBatting1Name <chr>, …
> GL.2015 %>%
+     filter(HomeTeam == 'BOS') %>%
+     # look at patriots day game in 2015
+     filter(Date == 20150420)
# A tibble: 1 x 161
    Date DoubleHeader DayOfWeek VisitingTeam VisitingTeamLea… VisitingTeamGam… HomeTeam
   <dbl>        <dbl> <chr>     <chr>        <chr>                      <dbl> <chr>
1 2.02e7           0 Mon       BAL          AL                            13 BOS
# … with 154 more variables: HomeTeamLeague <chr>, HomeTeamGameNumber <dbl>,
#   VisitorRunsScored <dbl>, HomeRunsScore <dbl>, LengthInOuts <dbl>, DayNight <chr>,
```

Problem Set 1
Pat Ausman

```
#   CompletionInfo <dbl>, ForfeitInfo <lgl>, ProtestInfo <chr>, ParkID <chr>, Attendance <dbl>,
#   Duration <dbl>, VisitorLineScore <chr>, HomeLineScore <chr>, VisitorAB <dbl>, VisitorH <dbl>,
#   VisitorD <dbl>, VisitorT <dbl>, VisitorHR <dbl>, VisitorRBI <dbl>, VisitorSH <dbl>,
#   VisitorSF <dbl>, VisitorHBP <dbl>, VisitorBB <dbl>, VisitorIBB <dbl>, VisitorK <dbl>,
#   VisitorSB <dbl>, VisitorCS <dbl>, VisitorGDP <dbl>, VisitorCI <dbl>, VisitorLOB <dbl>,
#   VisitorPitchers <dbl>, VisitorER <dbl>, VisitorTER <dbl>, VisitorWP <dbl>, VisitorBalks <dbl>,
#   VisitorPO <dbl>, VisitorA <dbl>, VisitorE <dbl>, VisitorPassed <dbl>, VisitorDB <dbl>,
#   VisitorTP <dbl>, HomeAB <dbl>, HomeH <dbl>, HomeD <dbl>, HomeT <dbl>, HomeHR <dbl>,
#   HomeRBI <dbl>, HomeSH <dbl>, HomeSF <dbl>, HomeHBP <dbl>, HomeBB <dbl>, HomeIBB <dbl>,
#   HomeK <dbl>, HomeSB <dbl>, HomeCS <dbl>, HomeGDP <dbl>, HomeCI <dbl>, HomeLOB <dbl>,
#   HomePitchers <dbl>, HomeER <dbl>, HomeTER <dbl>, HomeWP <dbl>, HomeBalks <dbl>, HomePO
<dbl>,
#   HomeA <dbl>, HomeE <dbl>, HomePassed <dbl>, HomeDB <dbl>, HomeTP <dbl>, UmpireHID <chr>,
#   UmpireHName <chr>, Umpire1BID <chr>, Umpire1BName <chr>, Umpire2BID <chr>, Umpire2BName
<chr>,
#   Umpire3BID <chr>, Umpire3BName <chr>, UmpireLFID <lgl>, UmpireLFName <chr>, UmpireRFID <lgl>,
#   UmpireRFName <chr>, VisitorManagerID <chr>, VisitorManagerName <chr>, HomeManagerID <chr>,
#   HomeManagerName <chr>, WinningPitcherID <chr>, WinningPitcherName <chr>, LosingPitcherID <chr>,
#   LosingPitcherNAme <chr>, SavingPitcherID <chr>, SavingPitcherName <chr>, GameWinningRBIID
<chr>,
#   GameWinningRBIName <chr>, VisitorStartingPitcherID <chr>, VisitorStartingPitcherName <chr>,
#   HomeStartingPitcherID <chr>, HomeStartingPitcherName <chr>, VisitorBatting1PlayerID <chr>,
#   VisitorBatting1Name <chr>, …
> glimpse(GL.2015 %>%
+           filter(HomeTeam == 'BOS') %>%
+           filter(Date == 20150420))
Observations: 1
Variables: 161
$ Date                    <dbl> 20150420
$ DoubleHeader            <dbl> 0
$ DayOfWeek               <chr> "Mon"
$ VisitingTeam            <chr> "BAL"
$ VisitingTeamLeague      <chr> "AL"
$ VisitingTeamGameNumber  <dbl> 13
$ HomeTeam                <chr> "BOS"
$ HomeTeamLeague          <chr> "AL"
$ HomeTeamGameNumber      <dbl> 13
$ VisitorRunsScored       <dbl> 1
$ HomeRunsScore           <dbl> 7
$ LengthInOuts            <dbl> 39
$ DayNight                <chr> "D"
$ CompletionInfo          <dbl> NA
$ ForfeitInfo             <lgl> NA
$ ProtestInfo             <chr> NA
$ ParkID                  <chr> "BOS07"
$ Attendance              <dbl> 36829
$ Duration                <dbl> 141
$ VisitorLineScore        <chr> "0100000"
$ HomeLineScore           <chr> "104002x"
$ VisitorAB               <dbl> 24
$ VisitorH                <dbl> 4
$ VisitorD                <dbl> 2
$ VisitorT                <dbl> 1
$ VisitorHR               <dbl> 0
$ VisitorRBI              <dbl> 1
$ VisitorSH               <dbl> 0
```

Problem Set 1

Pat Ausman

```
$ VisitorSF          <dbl> 0
$ VisitorHBP         <dbl> 0
$ VisitorBB          <dbl> 3
$ VisitorIBB         <dbl> 0
$ VisitorK           <dbl> 7
$ VisitorSB          <dbl> 1
$ VisitorCS          <dbl> 0
$ VisitorGDP         <dbl> 1
$ VisitorCI          <dbl> 0
$ VisitorLOB         <dbl> 5
$ VisitorPitchers    <dbl> 2
$ VisitorER          <dbl> 2
$ VisitorTER         <dbl> 2
$ VisitorWP          <dbl> 0
$ VisitorBalks       <dbl> 0
$ VisitorPO          <dbl> 18
$ VisitorA           <dbl> 3
$ VisitorE           <dbl> 3
$ VisitorPassed      <dbl> 0
$ VisitorDB          <dbl> 1
$ VisitorTP          <dbl> 0
$ HomeAB             <dbl> 22
$ HomeH              <dbl> 5
$ HomeD              <dbl> 1
$ HomeT              <dbl> 0
$ HomeHR             <dbl> 0
$ HomeRBI            <dbl> 4
$ HomeSH             <dbl> 1
$ HomeSF             <dbl> 1
$ HomeHBP            <dbl> 2
$ HomeBB             <dbl> 6
$ HomeIBB            <dbl> 1
$ HomeK              <dbl> 3
$ HomeSB             <dbl> 1
$ HomeCS             <dbl> 0
$ HomeGDP            <dbl> 1
$ HomeCI             <dbl> 0
$ HomeLOB            <dbl> 7
$ HomePitchers       <dbl> 3
$ HomeER             <dbl> 1
$ HomeTER            <dbl> 1
$ HomeWP             <dbl> 0
$ HomeBalks          <dbl> 0
$ HomePO             <dbl> 21
$ HomeA              <dbl> 6
$ HomeE              <dbl> 0
$ HomePassed         <dbl> 0
$ HomeDB             <dbl> 1
$ HomeTP             <dbl> 0
$ UmpireHID          <chr> "fleta901"
$ UmpireHName        <chr> "Andy Fletcher"
$ Umpire1BID         <chr> "hobep901"
$ Umpire1BName       <chr> "Pat Hoberg"
$ Umpire2BID         <chr> "mealj901"
$ Umpire2BName       <chr> "Jerry Meals"
$ Umpire3BID         <chr> "woodt901"
$ Umpire3BName       <chr> "Tom Woodring"
$ UmpireLFID         <lgl> NA
$ UmpireLFName       <chr> "(none)"
$ UmpireRFID         <lgl> NA
$ UmpireRFName       <chr> "(none)"
```

Problem Set 1

Pat Ausman

```
$ VisitorManagerID          <chr> "showb801"
$ VisitorManagerName        <chr> "Buck Showalter"
$ HomeManagerID             <chr> "farrj001"
$ HomeManagerName           <chr> "John Farrell"
$ WinningPitcherID          <chr> "mastj001"
$ WinningPitcherName        <chr> "Justin Masterson"
$ LosingPitcherID           <chr> "chenw001"
$ LosingPitcherNAme         <chr> "Wei-Yin Chen"
$ SavingPitcherID           <chr> NA
$ SavingPitcherName         <chr> "(none)"
$ GameWinningRBIID          <chr> NA
$ GameWinningRBIName        <chr> "(none)"
$ VisitorStartingPitcherID  <chr> "chenw001"
$ VisitorStartingPitcherName <chr> "Wei-Yin Chen"
$ HomeStartingPitcherID     <chr> "mastj001"
$ HomeStartingPitcherName   <chr> "Justin Masterson"
$ VisitorBatting1PlayerID   <chr> "deaza001"
$ VisitorBatting1Name       <chr> "Alejandro de Aza"
$ VisitorBatting1Position   <dbl> 7
$ VisitorBatting2PlayerID   <chr> "parej002"
$ VisitorBatting2Name       <chr> "Jimmy Paredes"
$ VisitorBatting2Position   <dbl> 10
$ VisitorBatting3PlayerID   <chr> "jonea003"
$ VisitorBatting3Name       <chr> "Adam Jones"
$ VisitorBatting3Position   <dbl> 8
$ VisitorBatting4PlayerID   <chr> "davic003"
$ VisitorBatting4Name       <chr> "Chris Davis"
$ VisitorBatting4Position   <dbl> 3
$ VisitorBatting5PlayerID   <chr> "snidt001"
$ VisitorBatting5Name       <chr> "Travis Snider"
$ VisitorBatting5Position   <dbl> 9
$ VisitorBatting6PlayerID   <chr> "machm001"
$ VisitorBatting6Name       <chr> "Manny Machado"
$ VisitorBatting6Position   <dbl> 5
$ VisitorBatting7PlayerID   <chr> "flahr001"
$ VisitorBatting7Name       <chr> "Ryan Flaherty"
$ VisitorBatting7Position   <dbl> 4
$ VisitorBatting8PlayerID   <chr> "lavar001"
$ VisitorBatting8Name       <chr> "Ryan Lavarnway"
$ VisitorBatting8Position   <dbl> 2
$ VisitorBatting9PlayerID   <chr> "cabre001"
$ VisitorBatting9Name       <chr> "Everth Cabrera"
$ VisitorBatting9Position   <dbl> 6
$ HomeBatting1PlayerID      <chr> "bettm001"
$ HomeBatting1Name          <chr> "Mookie Betts"
$ HomeBatting1Position      <dbl> 8
$ HomeBatting2PlayerID      <chr> "pedrd001"
$ HomeBatting2Name          <chr> "Dustin Pedroia"
$ HomeBatting2Position      <dbl> 4
$ HomeBatting3PlayerID      <chr> "ortid001"
$ HomeBatting3Name          <chr> "David Ortiz"
$ HomeBatting3Position      <dbl> 10
$ HomeBatting4PlayerID      <chr> "ramih003"
$ HomeBatting4Name          <chr> "Hanley Ramirez"
$ HomeBatting4Position      <dbl> 7
$ HomeBatting5PlayerID      <chr> "napom001"
$ HomeBatting5Name          <chr> "Mike Napoli"
$ HomeBatting5Position      <dbl> 3
$ HomeBatting6PlayerID      <chr> "victs001"
$ HomeBatting6Name          <chr> "Shane Victorino"
$ HomeBatting6Position      <dbl> 9
```

Problem Set 1
Pat Ausman

```
$ HomeBatting7PlayerID      <chr> "holtb002"
$ HomeBatting7Name          <chr> "Brock Holt"
$ HomeBatting7Position      <dbl> 5
$ HomeBatting8PlayerID      <chr> "bogax001"
$ HomeBatting8Name          <chr> "Xander Bogaerts"
$ HomeBatting8Position      <dbl> 6
$ HomeBatting9PlayerID      <chr> "hanir001"
$ HomeBatting9Name          <chr> "Ryan Hanigan"
$ HomeBatting9Position      <dbl> 2
$ AdditionalInfo            <chr> NA
$ AcquisitionInfo           <chr> "Y"
```

c) ***Cut and paste this R code into your script in RStudio:***
Master %>% inner_join(Batting) %>% filter(playerID=="aaronha01") %>% select(playerID, yearID, stint, nameFirst, nameLast, HR)

Master %>% inner_join(Batting) %>% filter(playerID=="aaronha01") %>% select(playerID, yearID, stint, nameFirst, nameLast, HR)

```
Joining, by = "playerID"
   playerID yearID stint nameFirst nameLast HR
1  aaronha01   1954     1     Hank    Aaron 13
2  aaronha01   1955     1     Hank    Aaron 27
3  aaronha01   1956     1     Hank    Aaron 26
4  aaronha01   1957     1     Hank    Aaron 44
5  aaronha01   1958     1     Hank    Aaron 30
6  aaronha01   1959     1     Hank    Aaron 39
7  aaronha01   1960     1     Hank    Aaron 40
8  aaronha01   1961     1     Hank    Aaron 34
9  aaronha01   1962     1     Hank    Aaron 45
10 aaronha01   1963     1     Hank    Aaron 44
11 aaronha01   1964     1     Hank    Aaron 24
12 aaronha01   1965     1     Hank    Aaron 32
13 aaronha01   1966     1     Hank    Aaron 44
14 aaronha01   1967     1     Hank    Aaron 39
15 aaronha01   1968     1     Hank    Aaron 29
16 aaronha01   1969     1     Hank    Aaron 44
17 aaronha01   1970     1     Hank    Aaron 38
18 aaronha01   1971     1     Hank    Aaron 47
19 aaronha01   1972     1     Hank    Aaron 34
20 aaronha01   1973     1     Hank    Aaron 40
21 aaronha01   1974     1     Hank    Aaron 20
22 aaronha01   1975     1     Hank    Aaron 12
23 aaronha01   1976     1     Hank    Aaron 10
```

d) ***Run the line of code in c). Answer the following: Based on the output in the console, what does the `inner_join` function do? (HINT: first look at the columns in the Batting and Master tables)***

Inner_join combines the two data_files based on a common key between the two data_files. Since no key was specified files are joined by any/all common fields in this case PlayerID. Only data that exists in the Batting file is selected from the Masters.

e) ***Add a line of code to yoru script that joins together the Batting and Pitching tables with*** *Batting %>% inner_join(Pitching, by=c("playerID", "yearID"))* ***and adds filters to find a player who recorded more than 100 outs as a pitcher and had over 100 at-bats as a hitter in a season. In how many player-seasons has this occurred (i.e., how many observation does this script yield)? In how many player-seasons has this occurred since the year 1990?***

> **Combo <- Batting %>% inner_join(Pitching, by=c("playerID", "yearID")) %>%**

Problem Set 1
Pat Ausman

```
+     filter(lPouts > 100 & AB > 100)
>   cat ("The number of players that have had 100 outs as a pitcher and 100 AB as a hitter
is",count(Combo)[[1]])
The number of players that have had 100 outs as a pitcher and 100 AB as a hitter is 1550
> #  How many players have done this since the year 1990?
>   Combo2 <- Combo %>%
+     filter(yearID >= 1900)
> cat ("The number of players that have had 100 outs as a pitcher and 100 AB as a hitter since 1900
is",
+       count(Combo2)[[1]])
The number of players that have had 100 outs as a pitcher and 100 AB as a hitter since 1900 is 735
```

f) BONUS: Join the Master table to your answer from e) and select the first and last names of the players returned.

```
  # BONUS: Join the Master table to your answer from e)
>   # and select the first and last names of the players returned.
>   # only records in Combo2 are matched in Master by playerID
>   whoarethey <-inner_join(Combo2,Master) %>% select(nameFirst, nameLast)
Joining, by = "playerID"
> head(whoarethey)
  nameFirst     nameLast
1       Ted Breitenstein
2     Jimmy     Callahan
3      Bill      Carrick
4      Bill      Dinneen
5   Noodles         Hahn
6      Pink       Hawley
```

Problem Set 1

Pat Ausman

2) ***What kinds of questions might the Lahman database help you answer? Give an example of a hypothesis one could test using data from the Lahman database. If you are struggling with an answer, consider reviewing all of the tables Lahman has to offer.***

Who is in the master table but was not a player?

First try is to find anyone in the master table but not in the Batting Table.

```
> a <- anti_join(Master,Batting)
Joining, by = "playerID"
> cat("The number of people in the master table that are not players",count(a)[[1]])
The number of people in the master table that are not players 190
```

But browsing through the data a few records looked odd. It looked like most of the records had playerID's ending in 99. How many didn't and what were they?

```
> # lets try to find all the records in Masters that have 99 in the id
>   f <- filter(Master,grepl('99', playerID))
> cat("The number of people in the master table have 99 in their player ID is",count(f)[[1]])
The number of people in the master table have 99 in their player ID is 187
> # okay what records are in a but not in f and who are they
>   f <- filter(a,!grepl('99', playerID)) %>% select (playerID,nameFirst,nameLast)
> f
   playerID nameFirst nameLast
1 crossjo01       Joe    Cross
2 kigerma01      Mark    Kiger
3 sulliwi01   William Sullivan
```

Researched the records. Also checked that they didnt appear in the Pitching file. Two of three are just missing from Batting though in baseball reference their data appears. One only appeared in post season. This show that as much as we would like to just write some code to analyse data you must still check the results.

#1 crossjo01 Joe Crossk - batter in 1 game in 1888

#2 kigerma01 Mark Kiger - only batted in the post season

#3 sulliwi01 William Sullivan - picher 1 game in 1988 - not in pitching

Problem Set 1
Pat Ausman

All the code in the file.

```
#####################################
#        Baseball Analytics         #
#           Problem Set 1           #
#          Pat Ausman               #
#           Feb 2019              #
#####################################

# clean up, set the environment
rm(list=ls())
setwd("/Users/pausman/Desktop/sabermetrics/problem1")

# check working directory just to make sure
getwd()
list.files()

# make sure tidyverse is loaded
if (!require(tidyverse)) install.packages('tidyverse')
library(tidyverse)

# setup the Lahman database which is built into R
if (!require(Lahman)) install.packages('Lahman')
library(Lahman)

#
# part a
#

# question 1 How many players with the last name “Saltalamacchia” have played MLB?
who_to_look_for = "Saltalamacchia"
Salty <- Master %>% select(nameLast) %>% filter(nameLast == who_to_look_for)
Salty
# output the number
cat("The number of players with a last name of",who_to_look_for,"is",count(Salty)[[1]])

# question 2 How many players with the first name “Jarrod” have played MLB?
who_to_look_for = "Jarrod"
Salty <- Master %>% select(nameFirst) %>% filter(nameFirst == who_to_look_for)
Salty
# output the number
cat("The number of players with a first name of",who_to_look_for,"is",count(Salty)[[1]])

# question 3 What is the name of the player with the playerID “parkeja01”
who_to_look_for = "parkeja01"
results <- Master %>% select(playerID, nameFirst, nameLast) %>% filter(playerID == who_to_look_for)
results
# output the number - assuming there is only one record
  cat("The name of the player with a playerID of",who_to_look_for,"is",results[[2]],results[[3]])


#
# part b
#

# cleanup and reinitialize
  rm(list=ls())
  setwd("/Users/pausman/Desktop/sabermetrics/problem1")

  # check working directory just to make sure
  getwd()
```

Problem Set 1

Pat Ausman

```
list.files()

# make sure tidyverse is loaded - this is probably still loaded
if (!require(tidyverse)) install.packages('tidyverse')
library(tidyverse)

# downloaded the game log from 2015. Use the headers for input since they are much clearer
GL.headers = read_csv('game_log_header.csv')
GL.2015 = read_csv('GL2015.TXT',
                   col_names = names(GL.headers))
names(GL.2015)
head(GL.2015)
tail(GL.2015)
summary(GL.2015)

GL.2015 %>% filter(HomeTeam == 'BOS')
GL.2015 %>%
  filter(HomeTeam == 'BOS') %>%
  # look at patriots day game in 2015
  filter(Date == 20150420)

glimpse(GL.2015 %>%
          filter(HomeTeam == 'BOS') %>%
          filter(Date == 20150420))

#
# part c - back to Lahman
#Cut and paste this R code into your script in RStudio:
Master %>% inner_join(Batting) %>% filter(playerID=="aaronha01") %>% select(playerID, yearID, stint, nameFirst, nameLast, HR)

#
# part d - what does inner_join do?
# part of dplyr library which is in rstudio
#
cat ("Inner_join combines the two data_files based on a common key between the two data_files.")
cat ("Since no key was specified files are joined by any/all common fields in this case PlayerID")
cat ("Only data that exists in the Batting file is selected from the Masters")


#
# part e
#  Add a line of code to yoru script that joins together the Batting and Pitching tables
#  with Batting %>% inner_join(Pitching, by=c("playerID", "yearID")) and adds filters to
#  find a player who recorded more than 100 outs as a pitcher and had over 100 at-bats
#  as a hitter in a season.
#  How many players have done this?


Combo <- Batting %>% inner_join(Pitching, by=c("playerID", "yearID")) %>%
  filter(IPouts > 100 & AB > 100)

#  How many players have done this?
cat ("The number of players that have had 100 outs as a pitcher and 100 AB as a hitter is",count(Combo)[[1]])

#  How many players have done this since the year 1990?
Combo2 <- Combo %>%
  filter(yearID >= 1900)
cat ("The number of players that have had 100 outs as a pitcher and 100 AB as a hitter since 1900 is",
     count(Combo2)[[1]])

# part f
```

Problem Set 1

Pat Ausman

```
# BONUS: Join the Master table to your answer from e)
# and select the first and last names of the players returned.
# only records in Combo2 are matched in Master by playerID
whoarethey <-inner_join(Combo2,Master) %>% select(nameFirst, nameLast)
head(whoarethey)
#
# What kinds of questions might the Lahman database help you answer?
# Give an example of a hypothesis one could test using data from the Lahman database
# I started with one question but it lead to more questions
# question 1: who is in the master table but was not a player

a <- anti_join(Master,Batting)
cat("The number of people in the master table that are not players",count(a)[[1]])
cat("Looking at the data there is at least one player that has no batting but did appear in the post season Matt Kiger")

# lets try to find all the records in Masters that have 99 in the id
f <- filter(Master,grepl('99', playerID))
cat("The number of people in the master table have 99 in their player ID is",count(f)[[1]])

# okay what records are in a but not in f and who are they
f <- filter(a,!grepl('99', playerID)) %>% select (playerID,nameFirst,nameLast)
f
```