

# 597A HW4

*Anya Conti*

*February 28, 2017*

1.

```
library(DBI)
library(RSQLite)
library(fImport)

## Loading required package: timeDate
## Loading required package: timeSeries

library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

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

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

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

setwd("/Users/Anya/Documents/JuniorYear/Spring/Stat 597A")
```

2.a.

```
drv = dbDriver("SQLite")
con = dbConnect(drv, dbname="baseball.db")

dbListTables(con)

## [1] "AllstarFull"      "Appearances"      "AwardsManagers"
## [4] "AwardsPlayers"    "AwardsShareManagers" "AwardsSharePlayers"
## [7] "Batting"          "BattingPost"      "Fielding"
## [10] "FieldingOF"       "FieldingPost"     "HallOfFame"
## [13] "Managers"         "ManagersHalf"     "Master"
## [16] "Pitching"         "PitchingPost"     "Salaries"
## [19] "Schools"         "SchoolsPlayers"   "SeriesPost"
## [22] "Teams"           "TeamsFranchises"  "TeamsHalf"
## [25] "sqlite_sequence" "xref_stats"
```

2.b.

```
Salaries <- dbReadTable(con, "Salaries")

Payroll <- aggregate(x = Salaries$salary, by = list(Salaries$teamID, Salaries$yearID), FUN = "sum")
```

```
Payroll2010 <- (Payroll[which(Payroll$Group.2 == 2010),])
Payroll2010
```

```
##      Group.1 Group.2      x
## 716      ARI      2010 60718166
## 717      ATL      2010 84423666
## 718      BAL      2010 81612500
## 719      BOS      2010 162447333
## 720      CHA      2010 105530000
## 721      CHN      2010 146609000
## 722      CIN      2010 71761542
## 723      CLE      2010 61203966
## 724      COL      2010 84227000
## 725      DET      2010 122864928
## 726      FLO      2010 57029719
## 727      HOU      2010 92355500
## 728      KCA      2010 71405210
## 729      LAA      2010 104963866
## 730      LAN      2010 95358016
## 731      MIL      2010 81108278
## 732      MIN      2010 97559166
## 733      NYA      2010 206333389
## 734      NYN      2010 134422942
## 735      OAK      2010 55254900
## 736      PHI      2010 141928379
## 737      PIT      2010 34943000
## 738      SDN      2010 37799300
## 739      SEA      2010 86510000
## 740      SFN      2010 98641333
## 741      SLN      2010 93540751
## 742      TBA      2010 71923471
## 743      TEX      2010 55250544
## 744      TOR      2010 62234000
## 745      WAS      2010 61400000
```

```
Payroll2010[which(Payroll2010$x == max(Payroll2010$x)),]
```

```
##      Group.1 Group.2      x
## 733      NYA      2010 206333389
```

2.c.

```
dbGetQuery(con,paste("SELECT teamID, SUM(salary) AS payroll",
                      "FROM Salaries",
                      "WHERE yearID = 2010",
                      "GROUP BY teamID"))
```

```
##      teamID  payroll
## 1      ARI  60718166
## 2      ATL  84423666
## 3      BAL  81612500
## 4      BOS 162447333
## 5      CHA 105530000
## 6      CHN 146609000
## 7      CIN  71761542
## 8      CLE  61203966
```

```
## 9    COL  84227000
## 10   DET 122864928
## 11   FLO  57029719
## 12   HOU  92355500
## 13   KCA  71405210
## 14   LAA 104963866
## 15   LAN  95358016
## 16   MIL  81108278
## 17   MIN  97559166
## 18   NYA 206333389
## 19   NYN 134422942
## 20   OAK  55254900
## 21   PHI 141928379
## 22   PIT  34943000
## 23   SDN  37799300
## 24   SEA  86510000
## 25   SFN  98641333
## 26   SLN  93540751
## 27   TBA  71923471
## 28   TEX  55250544
## 29   TOR  62234000
## 30   WAS  61400000
```

```
dbGetQuery(con,paste("SELECT teamID, SUM(salary) AS payroll",
                      "FROM Salaries",
                      "WHERE yearID = 2010",
                      "GROUP BY teamID",
                      "ORDER BY payroll desc" ,
                      "LIMIT 1"))
```

```
##   teamID  payroll
## 1    NYA 206333389
```

**2.d.**

```
YrlyPR <- dbGetQuery(con,paste("SELECT yearID, teamID, SUM(salary) AS payroll",
                                "FROM Salaries",
                                "WHERE yearID >= 1985 AND yearID <= 2010",
                                "GROUP BY teamID, yearID"))
```