DaigleInClassLab_Wk5D1.R

daiglechris

Tue Feb 13 22:28:17 2018

```
## Chris Daigle
## Week 5 Day 1 In Class Lab
getwd()
```

```
## [1] "/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/R/W
```

```
setwd("/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/R/We
dir()
```

```
## [1] "complete.csv"                  "district_rev_exp_readtable.txt"
## [3] "district_rev_exp.csv"          "district_rev_exp.txt"
## [5] "district_rev_exp.xlsx"         "school15doc.pdf"
```

```
fifa <- read.csv("complete.csv")
attach(fifa)

# 1
# Segment the data
league1 <- fifa[ , c("league", "eur_wage", "overall", "club", "age")]
# Find the mean for the variables of interest grouped by league
meanLeague1 <- aggregate(fifa[ , c("eur_wage", "overall")], list(league), mean, na.rm=TRUE)
# Decompose the sgmented data into just the mean wages and the leagues
meanWage <- meanLeague1[,1:2]
# Sort the mean wages from highest to lowest
meanWage <- meanWage[order(meanWage$eur_wage, decreasing = TRUE),]
# Select the first element of each column in the data.frame to see what the name and highest mean wage
bestWage <- data.frame(meanWage[[1]][1],meanWage[[2]][1])
# Pretty it up by making the data frame have legible and reasonable column names
colnames(bestWage) <- c("League with Highest Expected Wage", "Expected Wage")

# Decompose the segmented data into the leagues and the mean overall ranking (then do all the procedure
meanOverall <- meanLeague1[,c(1,3)]
meanOverall <- meanOverall[order(meanOverall$overall, decreasing = TRUE),]
bestOverall <- data.frame(meanOverall[[1]][1], meanOverall[[2]][1])
colnames(bestOverall) <- c("League with Highest Average Overall", "Average Overall")

bestWage
```

```
##   League with Highest Expected Wage Expected Wage
## 1            English Premier League      57840.98
```

```
bestOverall
```

```
##   League with Highest Average Overall Average Overall
## 1            Ukrainian Premier League        74.70833
```

```
#2
# I interpret this as "of the top ranked teams by eur_value, what team has the most players", and I bel
# "club"
is.factor(club)
```

```
## [1] TRUE
```

```
# Find the mean for the variables of interest grouped by teams
meanClubWage <- aggregate(fifa[ , c("eur_wage")], list(club), mean, na.rm=TRUE)
meanClubWage <- meanClubWage[order(meanClubWage$x, decreasing = TRUE),]
topClubWage <- meanClubWage[1:100,]
playerCount <- aggregate(full_name ~ club, fifa, FUN = function(x) length(unique(x)))
topClubWage <- merge(topClubWage, playerCount, by=0)
topClubWage <- data.frame(topClubWage[4], topClubWage[5], topClubWage[3])
colnames(topClubWage) <- c("Team", "numPlayers", "meanEurWage")
topClubWage <- topClubWage[order(topClubWage$numPlayers, decreasing = TRUE), ]
# This seems incorrect:
mostPlayers <- topClubWage$Team[which("numPlayers"==33)]
mostPlayers
```

```
## factor(0)
## 648 Levels:   SSV Jahn Regensburg 1. FC Heidenheim ... Zenit St. Petersburg
```

```
# Select the first element of each column in the data.frame to see what the name and highest mean wage

detach(fifa)
```