

```

1  # Exercise 1.#####
2  # How many days are between Jan 13, 1903 and Feb 27, 2019?
3  a <- c("1903-01-13", "2019-02-27")
4  dates <- as.Date(a)
5  class(dates)
6  dates[2]-dates[1]
7  difftime(dates[2],dates[1],units="weeks")
8
9  # Construct a sequence of dates starting from Jan 1, 2019 to Feb 27, 2019.
10 a <- as.Date(c("01-01-2019","02-27-2019"), format="%m-%d-%Y")
11 seq(a[1],a[2],by="day")
12
13 # Exercise 2.#####
14
15 # Download results.csv from HustkyCT (lecture11-data) in your computer.
16 # Import this dataset to R.
17
18 sc.game <- read.csv("results.csv", stringsAsFactors = FALSE, na.strings="")
19 head(sc.game)
20 str(sc.game)
21
22 sc.game$date <- as.Date(sc.game$date)
23 str(sc.game)
24
25 # Compare the average number of goals in 1930-1939 and in 2005-2014 and see
26 # whether it increased or decreased.
27
28 sc.game$year <- format(sc.game$date, format="%Y")# Review of Exercise
#####
29
30 # Download complete.csv from HustkyCT (lecture10-data) in your computer.
31 # Import this dataset to R.
32
33
34 rm(list=ls())
35
36 # 1. Which league is the best in terms of wage (eur_wage) and overall?
37
38 sc <- read.csv("complete.csv", stringsAsFactors = FALSE, na.strings="")
39 #sc <- read.csv("complete.csv", stringsAsFactors = FALSE)
40 sc <- sc[,c("name", "club", "age", "league", "eur_value", "eur_wage", "overall")]
41 sum(is.na(sc))
42 sc <- na.omit(sc)
43
44 nrow(sc)
45 head(sc)
46
47 str(sc)
48 summary(sc)
49
50 league_wage <- aggregate(sc[, "eur_wage"], list(sc$league), mean)
51 league_wage
52 names(league_wage) <- c("league", "ave_wage")
53 head(league_wage)
54 tail(league_wage)
55
56 # This is the way to sort several variables by one variable.
57 rank_wage <- order(league_wage[, "ave_wage"], decreasing = TRUE)
58 league_wage <- league_wage[rank_wage,]
59 head(league_wage)
60
61 league_overall <- aggregate(sc[, "overall"], list(sc$league), mean)
62 names(league_overall) <- c("league", "ave_overall")
63
64 rank_overall <- order(league_overall[, "ave_overall"], decreasing = TRUE)
65 league_overall <- league_overall[rank_overall,]
66 head(league_overall)
67
68 cbind(1:nrow(league_wage), league_wage, league_overall)

```

```

69
70 # 2. Based on "eur_value", which team has the most players in top 100?
71 # hint: 1. Sort based on eur_value, 2. Make sure that team is a factor
72
73 rank_value <- order(sc[, "eur_value"], decreasing = TRUE)
74 league_value <- sc[rank_value,]
75 league_value100 <- league_value[1:100,]
76 league_value100$rankvalue <- 1:100
77
78 league_value100
79 str(league_value100)
80 league_value100$club = factor(league_value100$club)
81 team_stat <- summary(league_value100$club) # or you can also use table()
82 team_stat
83 sort(team_stat, decreasing=TRUE)
84
85 # 3. Present the change of average wage (eur_wage) over age.
86 # hint: 1. Use aggregate() function to obtain average wages for each age,
87 # 2. Use plot() function
88
89 age_wage <- aggregate(sc[, "eur_wage"], list(sc$age), mean)
90 names(age_wage) <- c("age", "ave_wage")
91 age_wage
92 plot(age_wage$age, age_wage$ave_wage, xlab = "age", ylab="average wage", type="s")
93 boxplot(eur_wage~age, data=sc, xlab = "age", ylab="wage", range=5)
94
95 # 4. Which team has the most players under 23?
96 # hint: Make sure that team is a factor
97
98 team23 <- factor(sc[sc$age<23, "club"])
99 team_age <- summary(team23)
100 team_age_rank <- sort(team_age, decreasing = TRUE)
101 team_age_rank[1:10]
102
103 #####
104 # Time and date data
105
106 today <- Sys.Date()
107 today
108 class(today)
109
110 now <- date()
111 now
112 class(now)
113
114 # Use as.Date() to convert string to dates
115 a <- c("2007-06-22", "2004-02-13")
116 class(a)
117 mydates <- as.Date(a)
118 mydates
119 class(mydates)
120
121 a1 <- c("01/13/2013", "Feb-03-2013", "12/11/14", "02/26/2019")
122 as.Date(a1)
123 mydates1 <- as.Date(a1, format=c("%m/%d/%Y", "%b-%d-%Y", "%m/%d/%y") )
124 mydates1
125
126 # The following symbols can be used with the format() function to print dates
127
128 # Symbol      Meaning      Example
129 # %d          day as a number (0-31)  25
130 # %a          abbreviated weekday    Mon
131 # %A          unabbreviated weekday   Monday
132 # %m          month (00-12)           08
133 # %b          abbreviated month       Jan
134 # %B          unabbreviated month     January
135 # %y          2-digit year            19
136 # %Y          4-digit year            2019

```

```

137
138 format(mydates1, format = "%B %d %Y")
139 format(mydates1, format = "%m-%d-%y")
140 format(mydates1, format = "%Y")
141 format(mydates1, format = "%B %d (%a)")
142
143 # We can compare dates
144 mydates[1] > mydates[2]
145 days <- mydates[1] - mydates[2]
146 days
147 class(days)
148 difftime(mydates[1],mydates[2])
149
150 mydates[1]
151 mydates[1] + 1
152 mydates[1] + 30
153 mydates[1] + 365
154
155 # Exercise 1.#####
156 # How many days are between Jan 13, 1903 and Feb 27, 2019?
157
158 # Construct a sequence of dates starting from Jan 1, 2019 to Feb 27, 2019.
159 # Hint: check seq() function.
160
161 # Exercise 2.#####
162
163 # Download results.csv from HustkyCT (lecture11-data) in your computer.
164 # Import this dataset to R.
165
166 # Compare the average number of goals in 1930-1939 and in 2005-2014 and see
167 # whether it increased or decreased.
168
169 # What is the ratio of home winning games to the total number of games in 1950s?
170
171 # plot the number of games over each year from 1900 to 1960 and see whether
172 # there was decrease in game in world war I (1914-1919) and world war II (1939-1945)
173
174
175 sc.game$score <- sc.game$home_score + sc.game$away_score
176
177 ave.score.1 <- colMeans(sc.game[sc.game$year>="1930" & sc.game$year<="1939",
178                           c("home_score", "away_score", "score")])
179 ave.score.2 <- colMeans(sc.game[sc.game$year>="2005" & sc.game$year<="2014",
180                           c("home_score", "away_score", "score")])
181
182 ave.score.1
183 ave.score.2
184
185 # What is the ratio of home winning games to the total number of games in 1950s?
186 game.1950s <- sc.game[sc.game$date >="1950-01-01" & sc.game$date <= "1959-12-31",]
187 home_win <- game.1950s$home_score > game.1950s$away_score
188 sum(home_win) / nrow(game.1950s)
189
190 # plot the number of games over each year from 1900 to 1960 and see whether
191 # there was decrease in game in world war I (1914-1919) and world war II (1939-1945)
192
193 games <- summary(factor(sc.game$year[sc.game$year>1900 & sc.game$year<1960]))
194 plot(games, type="s")
195

```