```
library(XML)
library(httr)

tables <- GET ("http://en.wikipedia.org/wiki/List_of_countries_by_population")

poptable <- readHTMLTable (rawToChar (tables$content), which=2 )

head(poptable)

options(digits = 2)

# we have a dataframe with the desired information
# however it is not in the desired format

# all of the veriables seem to be factor, but some should be numeric or character
str(poptable) #every column is a factor

# before converting them to numeric or character values, there are some things to get rid of
# like commas, or "%" sign, etc.  which should be deleted

# delete commas in the "% of worl population" column, which is the 5th column

poptable[,5] <- gsub ("%", "", poptable[,5])

# now, it is ready to convert character to numeric
poptable[,5] <- as.numeric (poptable[,5])

# try to get mean, but it fails since it is a factor, not a numeric.
mean (poptable$Population) #fails becuse it is a factor
str(poptable$Population) #see?

# remove commas in the "Population" column
# there are 2 ways to write it, choose one:

poptable[,3] <- gsub(",", "", poptable[,3])
# This is the other way:
# poptable$Population <- gsub(",","",poptable$Population)

# it is character now (check it!),  go on:
# convert character to numeric
poptable[,3] <- as.numeric(poptable[,3])
# OR: poptable$Population <- as.numeric(poptable$Population)

# converting  factor to numeric
# is somewhat different than converting char to numeric
# we must use levels of that factor here!

# convert rank column (factor) into numeric
poptable[,1] <- as.numeric ( levels ( poptable[,1]))[ poptable[,1]]
```

```
# HOMEWORK 3

# 1. convert "Date" column (factor) into date
# Please find date formats in r here: https://www.r-bloggers.com/date-formats-in-r/

# 2. CHANGE THE COLUMN NAMES AS: "Rank","Country", "Population", "Date",   "% of world",
"Source"

# 3.There are some strings btw [] in country names column. get rid of those:

# 4. Show that you can get the difference btw two dates. Choose two dates by yourself

# 5. Get the difference btw the population of some countries you choose: ex.China-USA,  Turkey-USA:

# 6. Find the most crowded country

# 7. Extract the countries whose population is greater than 100000000 and assign it to
#    a new data frame called"pop_crowded"

# 8. Draw the barplot of these countries (pop_crowded) such that:
#    Y axis is the population, and X axis indicates the countries
#    You can google search or go to http://www.theanalysisfactor.com/r-11-bar-charts/  for help
```