

# 1 Introducere in R

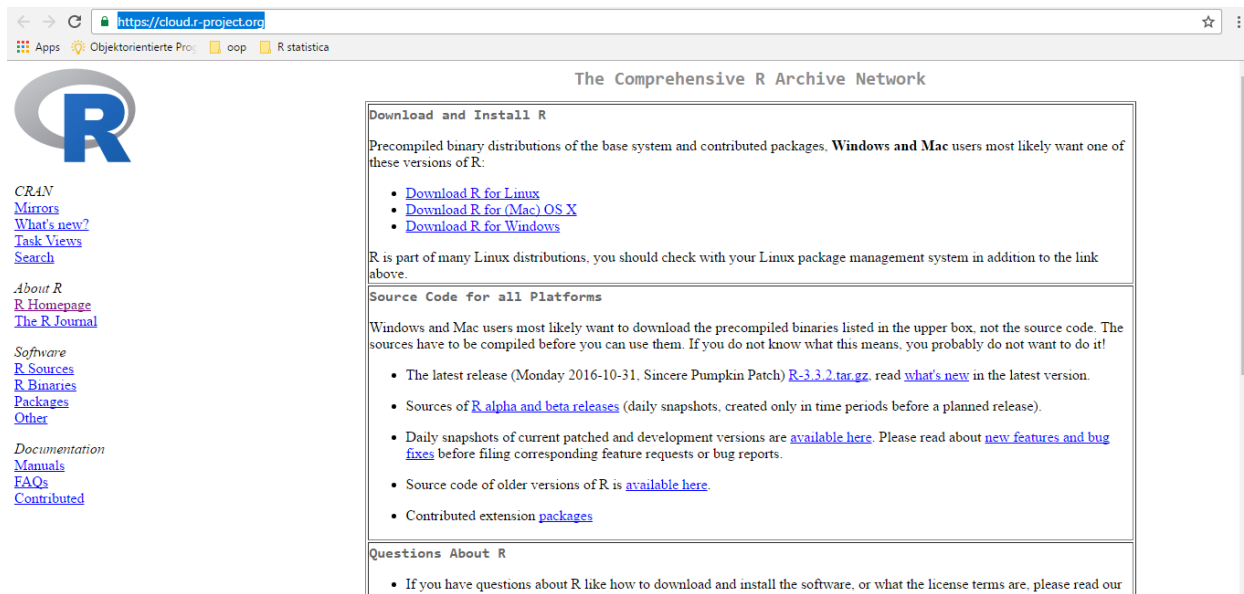
R este un soft pentru prelucrări statistice a datelor foarte performant. Este asemănător ca și performanțe cu SPSS doar că este un mediu care seamănă mai mult cu mediul linux lucrându-se mult din linie de comandă. Limbajul R este un limbaj orientat obiect. De fapt toate variabilele cu care vom lucra sunt obiecte.

## 1.1 Instalarea

Informații referitoare la proiectul R pentru statistică computațională se pot accesa la adresa de web <https://www.r-project.org/>. Pentru a downloada softul trebuie accesat un server CRAN (Comprehensive R Archive Network) de la adresa <https://cran.r-project.org/mirrors.html>.

Ultima versiune a limbajului R este versiunea R-3.3.2 pentru 32-bit și 64-bit. În mod implicit se instalează versiunea pe 32 de biți pentru Windows, dar se poate alege versiunea pe 64 de biți. Evident ca fanii sistemelor de operare Linux sau MacOX au la dispoziție variantele corespunzătoare.

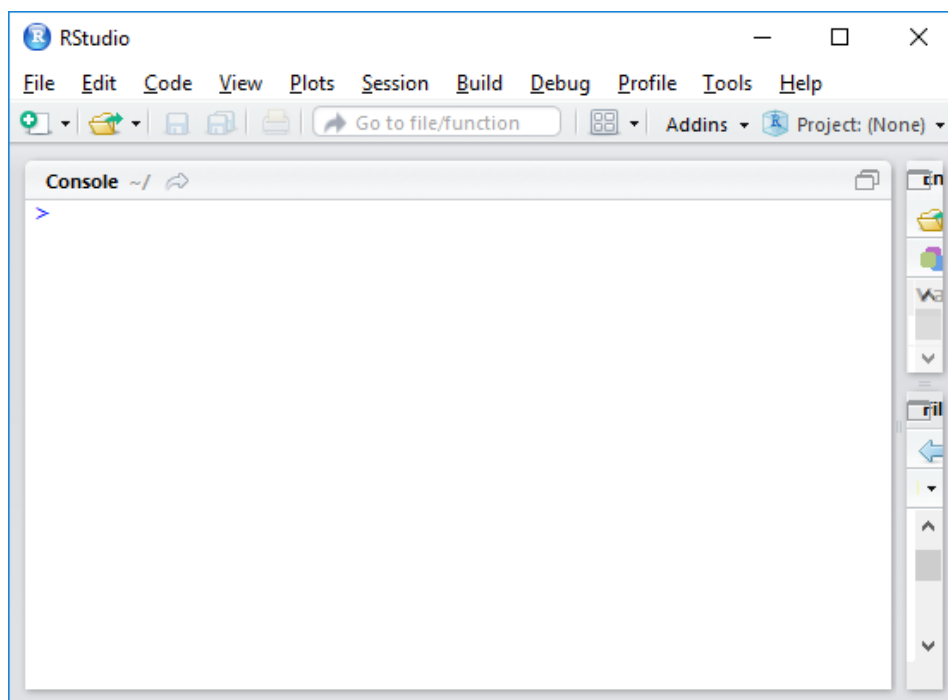
Adresa de download de la care am instalat softul este <https://cloud.r-project.org/>



După instalarea unei versiuni de R, pentru a simplifica modul de lucru vom instala o interfață grafică (tot gratuită) oferită de R-Studio de la adresa de web: <https://www.rstudio.com/products/rstudio/download/>

## 1.2 Lucrul în mediul R

La pornirea RStudio interfața pornește afișând un prompter „>” semn care indică faptul că mediul așteaptă comenzi:



Comenzile se lansează apăsând tasta ENTER. În cazul în care în loc de < apare + înseamnă că instrucțiunea se completează pe rândul următor:

### ***Exercițiu 1.1***

Ștergeți conținutul consolei.

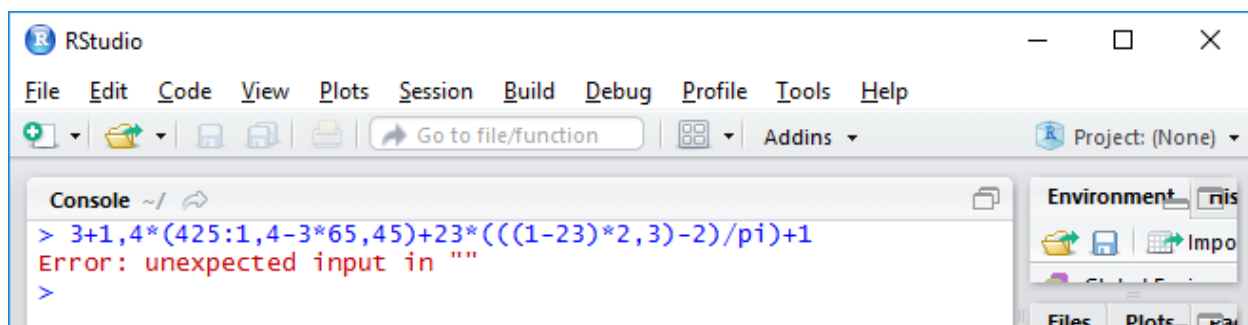
**Indicație:** *Ctrl+I*

### ***Ex: 1.2***

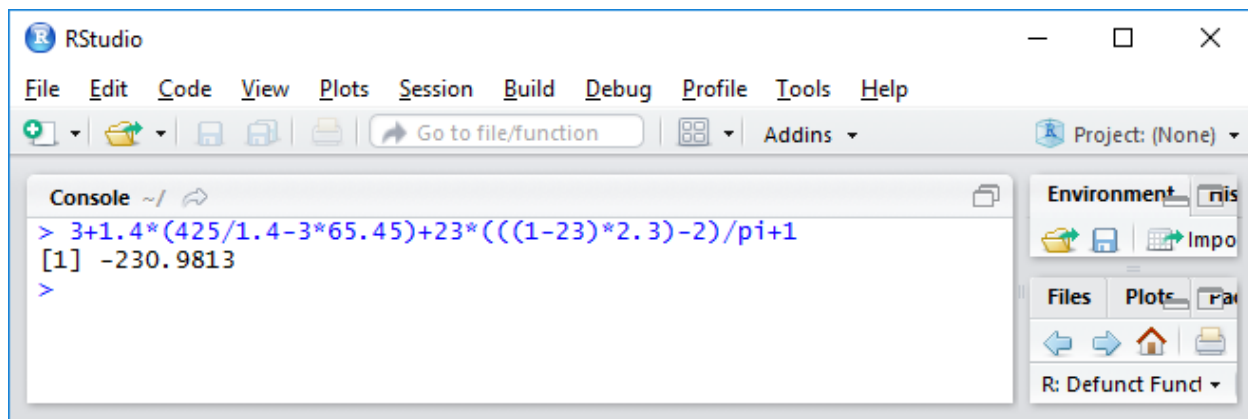
Calculați expresia  $3+1,4*(425:1,4-3*65,45)+23*(((1-23)*2,3)-2)/\pi)+1$

**Indicație:** *Ctrl+I*

Atenție la utilizarea separatorului zecimal și la semnul împărțirii altfel veți obține un mesaj de eroare:



Corect este:



### Exercițiu 1.3

Testați până la câte paranteze imbricate poate lucra limbajul R?

În linia de comandă se pot scrie mai multe instrucțiuni care se vor separa prin ”;”

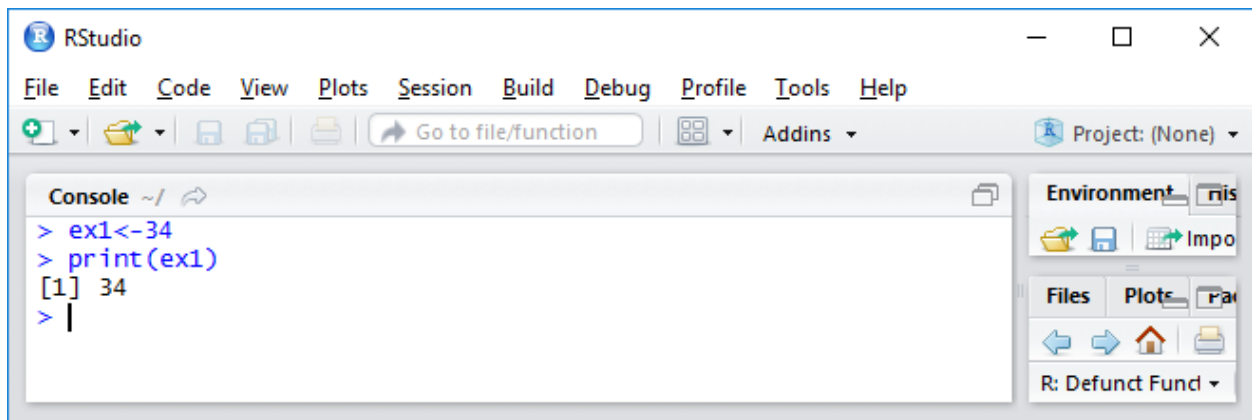
Pentru a vizualiza comenzi anterioare se folosește tasta ”săgeată sus”. Pentru a ieși din mediul de dezvoltare R se folosește funcția `q()`. La ieșire avem opțiunea de a salva istoricul instrucțiunilor scrise în consolă.

### 1.3 Variabile

În limbajul R variabilele sunt definite cu ajutorul caracterelor literale. Atenție la numele variabilelor deoarece în R numele este ”case sensitive”. În R toate variabilele sunt de fapt obiecte.

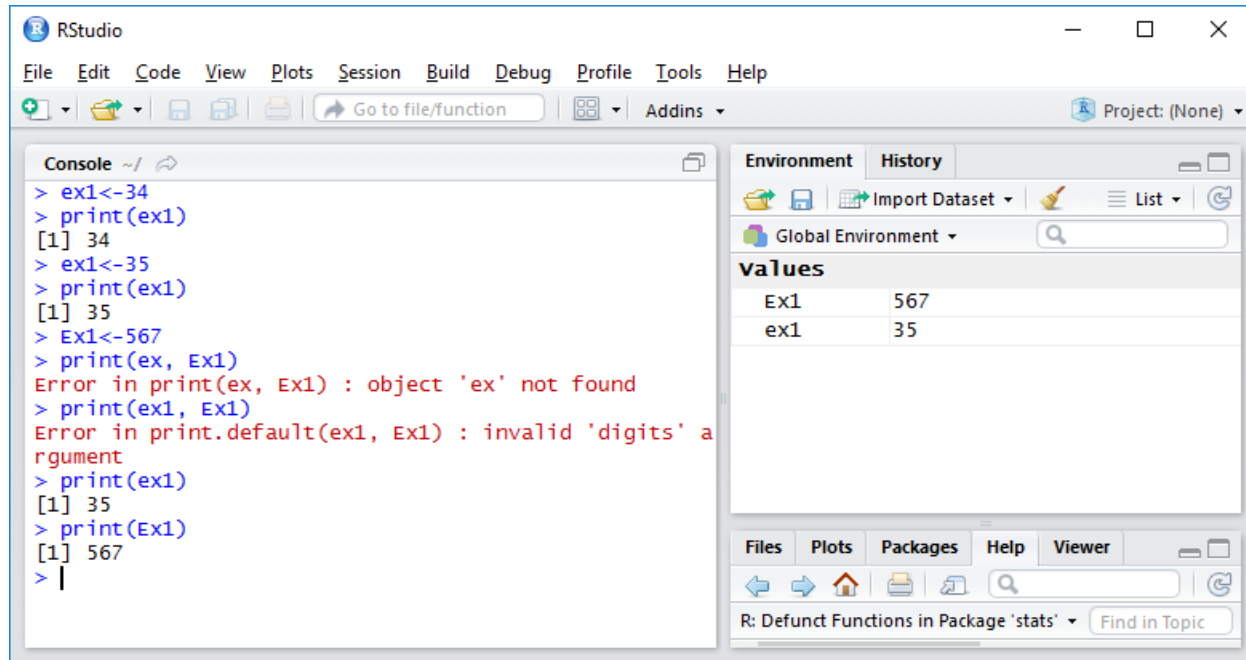
Atribuirea unei valori unui obiect se face cu ”<-”.

#### Exemplu:

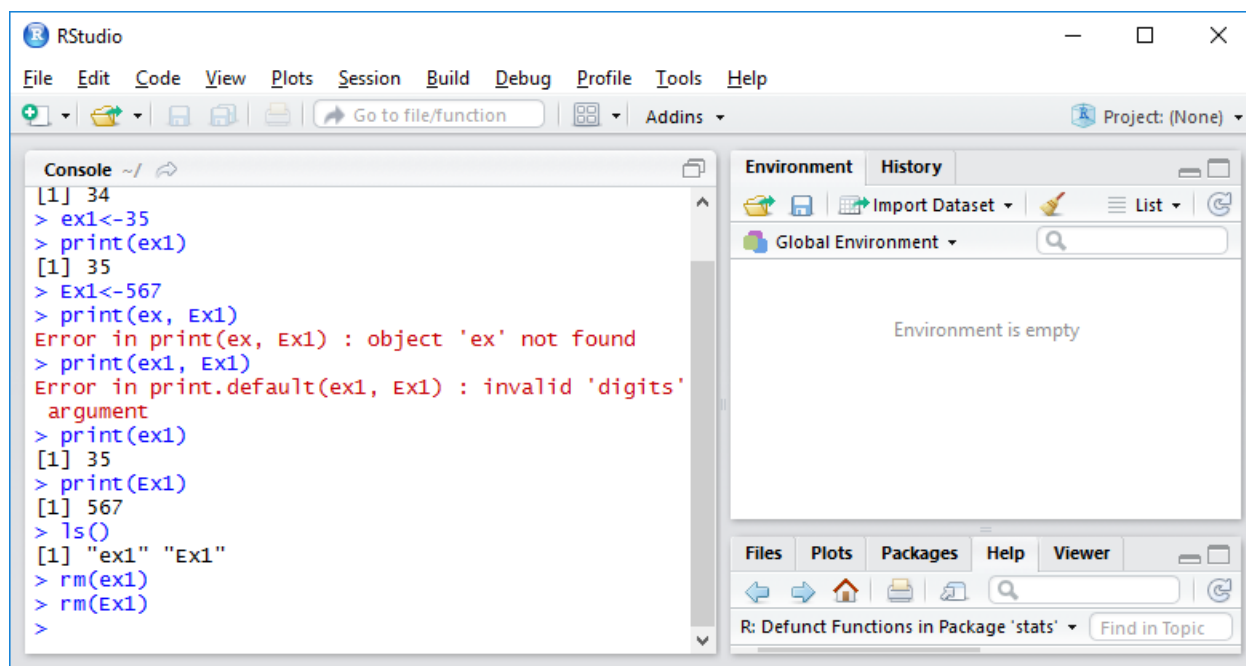


În exemplul de mai sus variabilei `ex1` i s-a atribuit valoarea 34. Apelând funcția `print` cu numele obiectului ca parametru se afișează valoarea variabilei.

## Exemplu:



Observăm că în exemplul de mai sus valoarea inițială a variabilei `ex1` a fost suprascrisă și este alta ca cea a variabilei `Ex1`. În panelul de "Environment" se afișează valorile curente ale variabilelor utilizate. Listarea variabilelor utilizate se face apelând funcția `ls()`. Pentru a distruge variabilele utilizate se folosește funcția `rm()` care primește ca parametru numele variabilei de șters.



## 1.4 Tipuri de date în R

În continuare vom prezenta cele mai utilizate tipuri de date în limbajul R

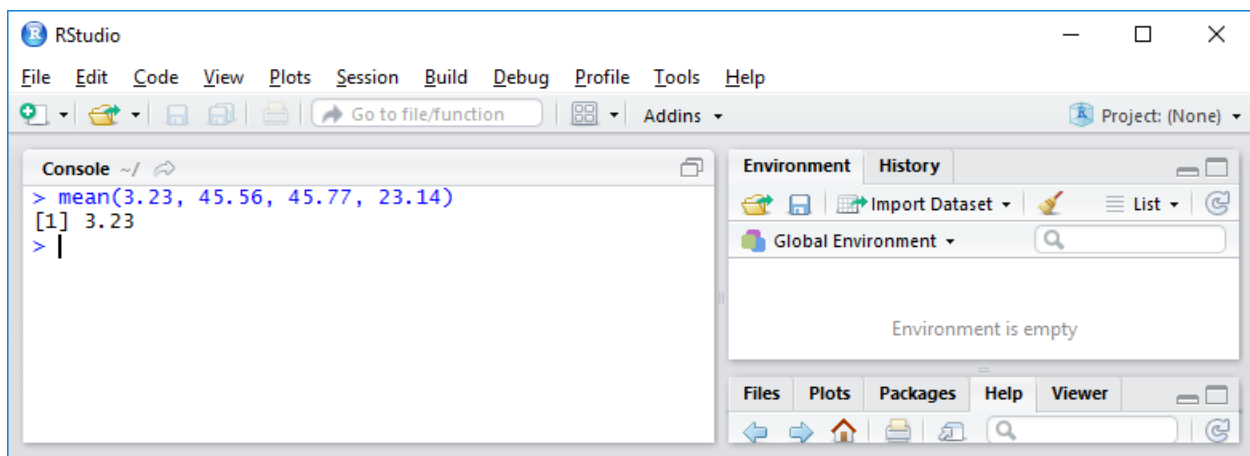
### 1.4.1 Vectori

#### *Exercițiu 1.4*

Să se calculeze media aritmetică a valorilor 3.23 45.56 45.77 23.14.

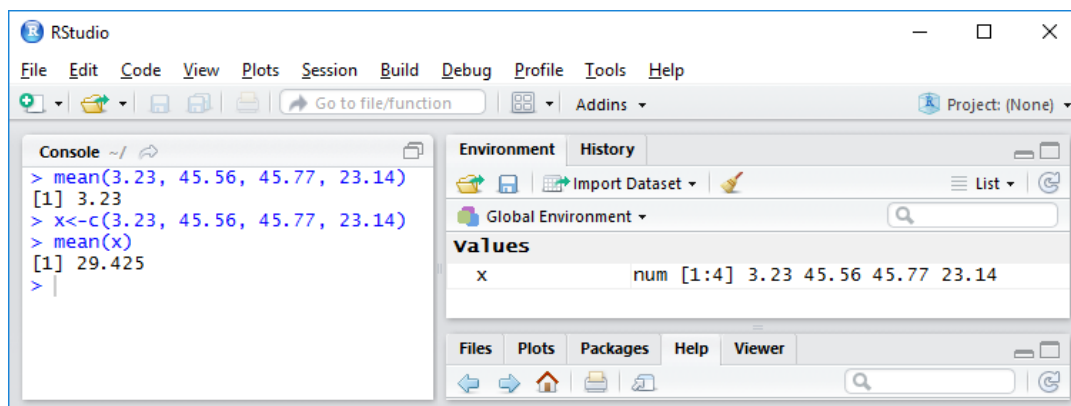
Soluția 1: utilizarea formulei pentru medie.

Soluția 2: utilizarea funcției `mean()`.



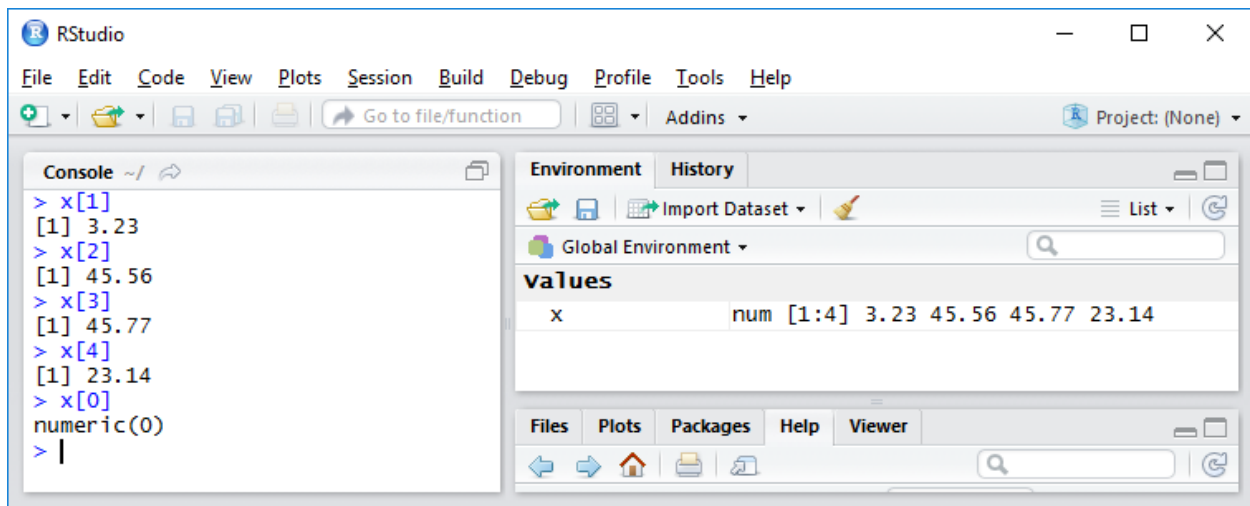
Observăm că funcția `mean()` afișează ca rezultat prima valoare care am trimis-o ca parametru. Celelalte valori sunt ignorate. Pentru a putea calcula media corect trebuie să atribuim valorile prezentate mai sus unui vector și apoi să trimitem acest vector ca parametru funcției `mean()`.

Practic trebuie să "concatenăm" valorile utilizând funcția `c()`. (`c` este prescurtarea de la "concatenate")



Să înțelegem ce s-a întâmplat:

Făcând atribuirea `x<-c(3.23, 45.56, 45.77, 23.14)` variabila x a devenit vector cu 4 valori. Pentru a le accesa evident că putem folosi notația clasică `x[i]`.

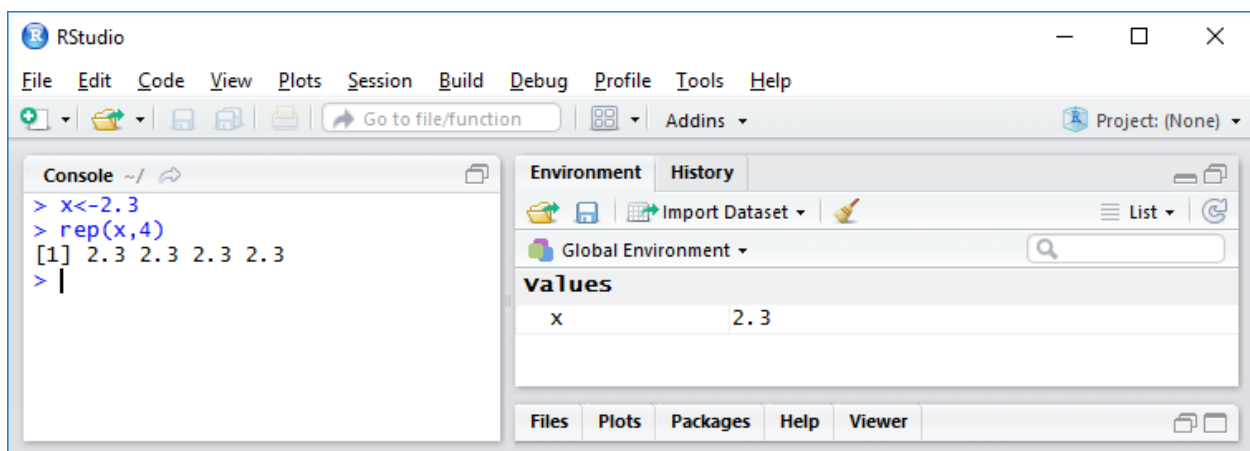


Observăm că în R indecșii uni vector încep de la 1. Acest lucru se poate observa și în panelul de valori în care variabila x conține valori numerice de la indexul 1 la 4.

## 1.4.2 Generarea de vectori

### 1.4.2.1 Funcția `rep()` :

Sintaxa: `rep(x, times=n)` repetă obiectul x de n ori



Observație:

În exemplul de mai sus observăm că în consolă valoarea 2.3 a fost repetată de 4 ori dar valoarea obiectului x nu s-a modificat dacă verificăm valoarea lui x afișată în "global environment".

## Exercițiu 1.5

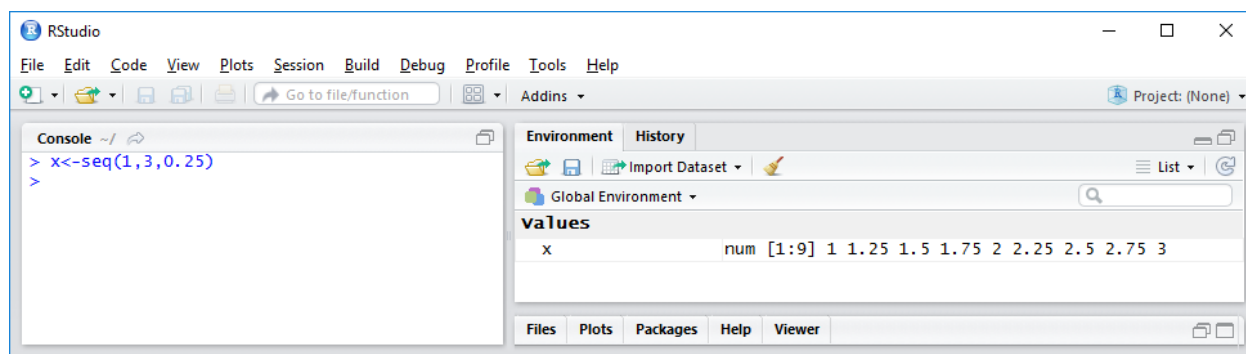
Atribuiți obiectului `x` de 5 ori valoarea 3.24 utilizând doar funcția `rep()`.

Observație

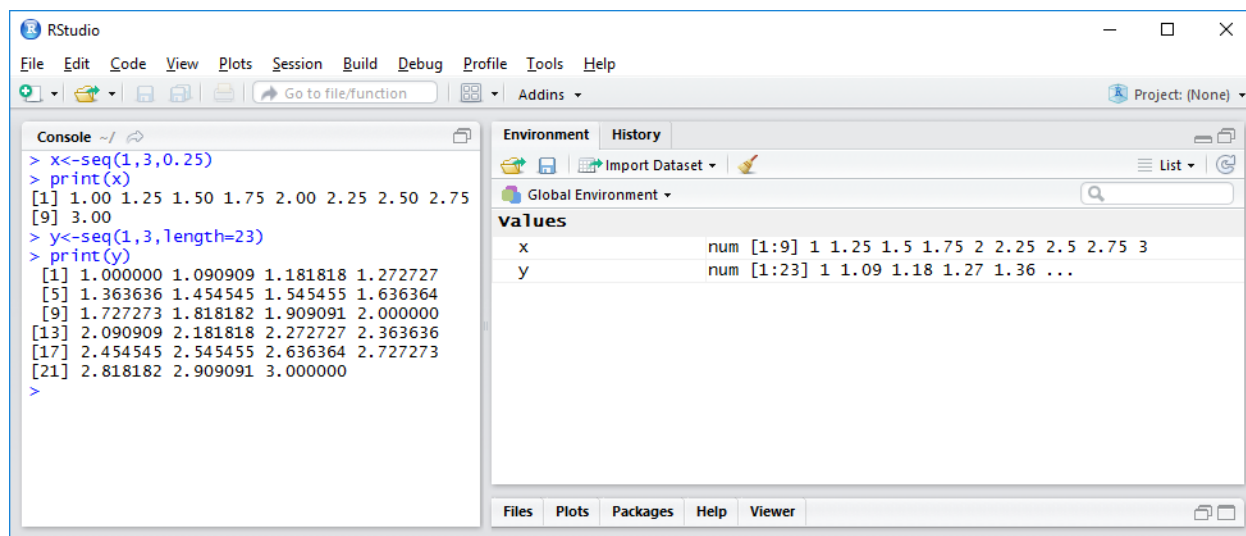
`rep(1, 4)` este identic cu `c(1, 1, 1, 1)`

### 1.4.2.2 Funcția `seq()`:

Sintaxa: `seq(from=, to=, by=, length=)`



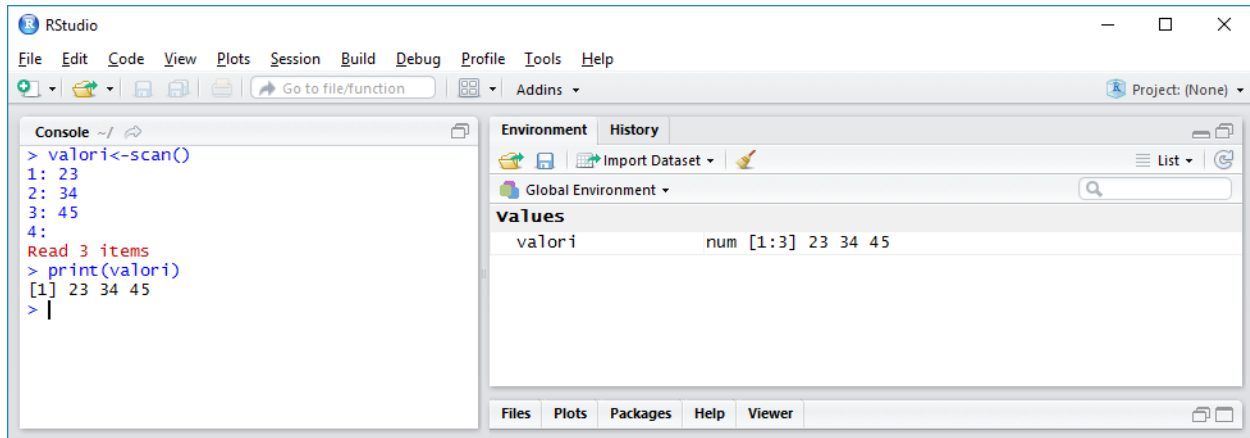
În exemplul de mai sus am omis parametrul `length` dacă îl utilizăm atunci vom obține un vector de valori de lungime `length` și lăsăm mediul R să calculeze pasul.



### 1.4.2.3 Funcția scan()

Citește de la tastatură sau dintr-un fișier câte un rând.

Exemplu citire de la tastatură:



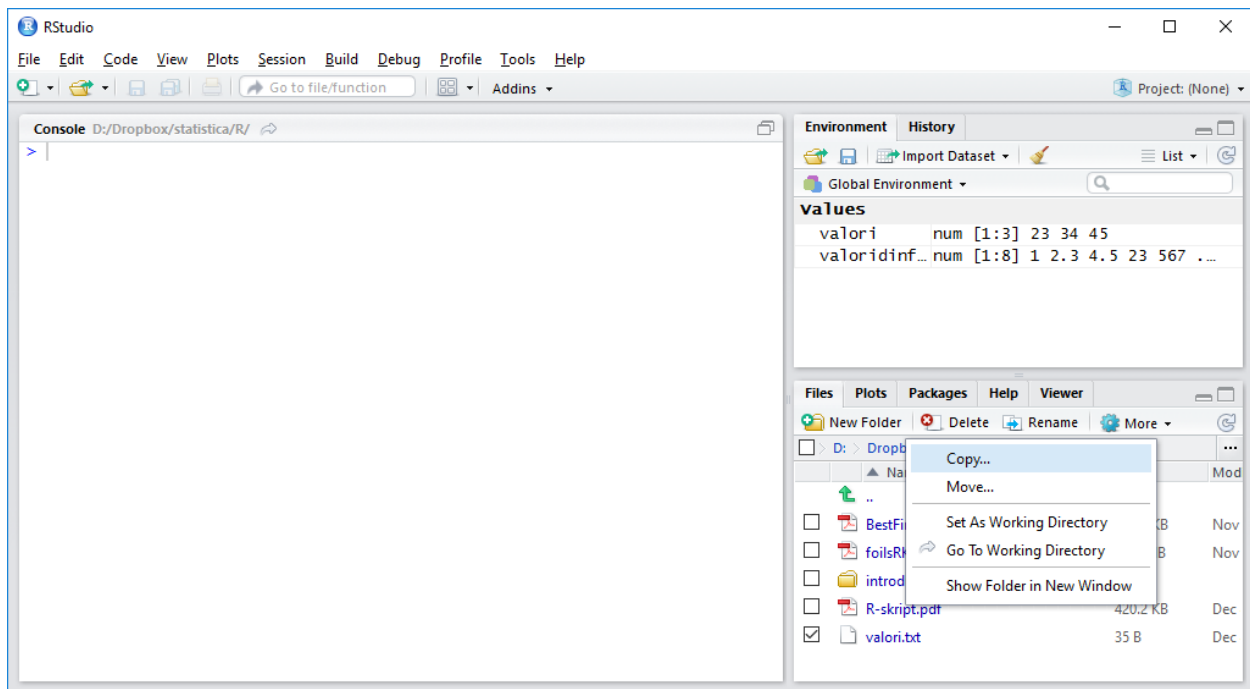
Observăm că citirea se încheie dacă la un pas apăsăm doar ENTER fără să mai adăugăm o valoare.

### Exercițiu 1.6

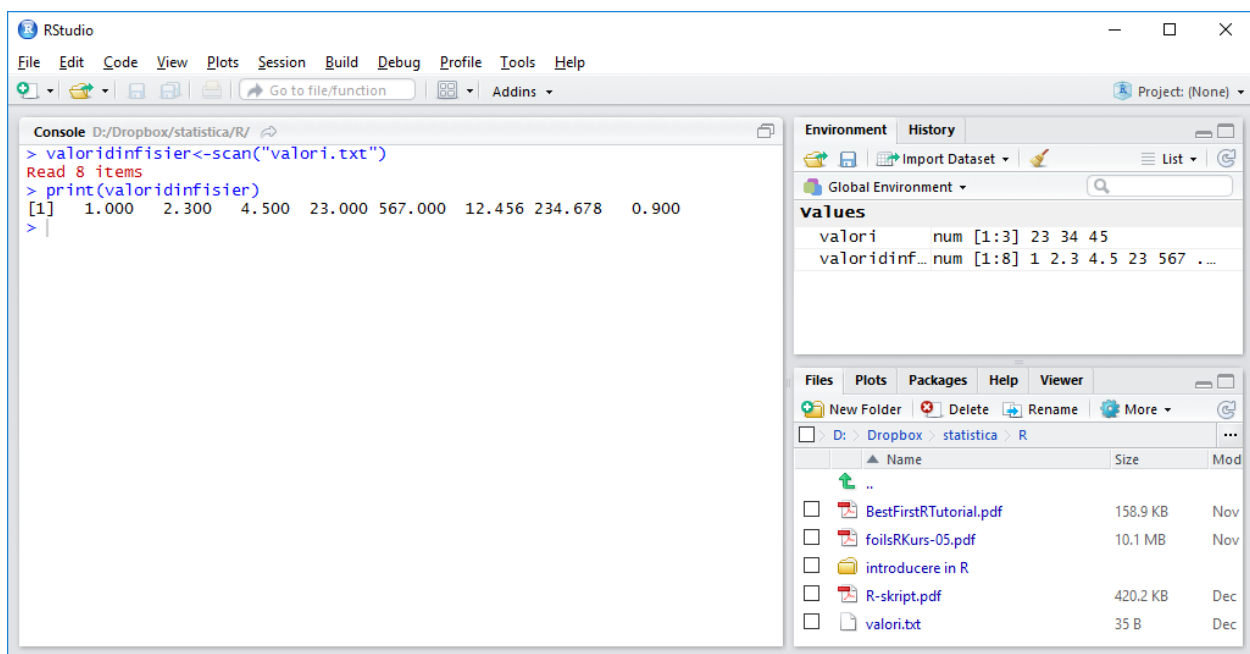
Care este o altă modalitate de a obține vectorul "valori"?

Citirea dintr-un fișier cu ajutorul funcției scan()

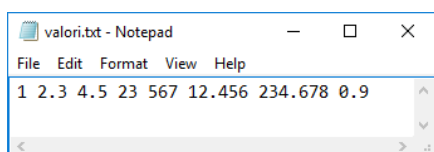
Exemplu citire dintr-un fișier „valori.txt”. Numele fișierului trebuie dat ca parametru între ghilimele duble funcției scan("valori.txt"). Aici apare problema locației fișierului. Prima varianta ar fi să folosim calea absoluta spre fișier sau să stabilim un director de lucru și să copiem fișierul în acel director. Stabilirea directorului de lucru se poate face dând click pe meniul MORE din panelul de FILES și alegând opțiunea "Set as working directory" directorul curent.



După stabilirea directorului de lucru (a se vedea bara consolei):



Datele din fisier



### ***Exercițiu 1.7***

În obiectele `alfa`, `beta`, `gama1` și `gama2` atribuiți 20 de valori utilizând funcțiile `c()`, `seq()`, `scan()` și `scan("fisier.txt")` și apoi calculați media aritmetică pentru fiecare obiect.