The Impatient Guide to Getting Started With R and RStudio

This is a very quick guide. We assume you are insanely impatient and want to get going as quick as possible. More details can be found in our longer Introductory Guide. This is not a recommended guide and truly only for the impatient.

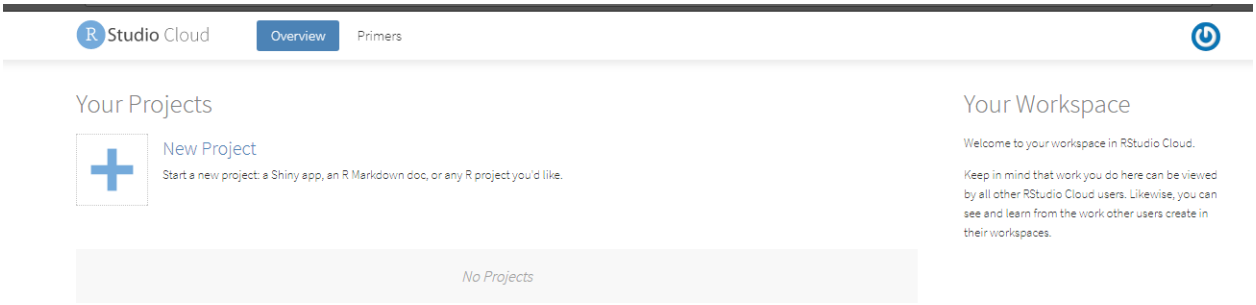
Overview

R is an environment for analysing and visualising data, as well as being a programming language, but it is much more than just that. The following is intended to give a brief flavour of R and to introduce some of its key features. Note that R is the program and Rstudio is the interface-in this impatient guide we will use the terms back and forth.

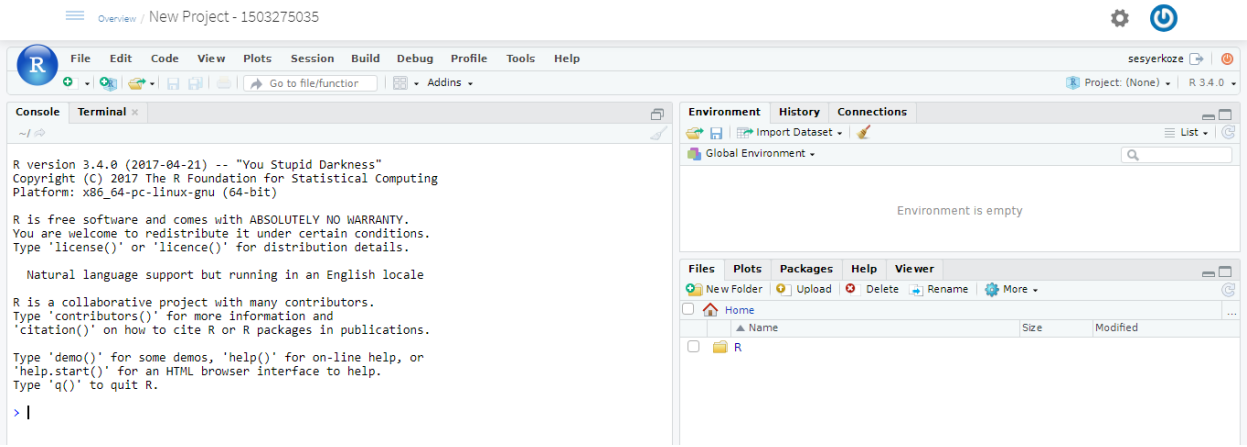
Do First

Create an account on rstudio.cloud.

Once logged in fully you will see the following screen.



Click on the big "plus" sign, give it a few moments and you're ready to go. Your screen should look something as follows:



The left side window is called the Console and the > sign is where we enter commands into R. Our output will either show up in the Console window or as a plot in the lower right window.

Some Data

We have national data on high school graduation rate and percent of children living in poverty. Some of the data looks as follows:

state	hsgradrate	childpov
ALABAMA	65.9	24.6
ALASKA	70.6	13.9
ARIZONA	68.5	22.1
ARKANSAS	76.4	24.9
CALIFORNIA	63.1	21.7
COLORADO	73.0	17.1
CONNECTICUT	78.5	10.2

Read the Data into R

For this class, all our data sets will be on the web and easily read into R. This data set is read in as follows. Enter the following command in the Console

```
> mydata=read.csv("https://tinyurl.com/teendata1")
```

You can view the data by entering the following command

```
> View(mydata)
```

You can check the dimenions of the data set as follows.

```
> dim(mydata)
[1] 51  3
```

This says we have 51 rows (District of Columbia!) and 3 columns of data.

The names of the columns can be found as follows

```
> names(mydata)
[1] "state"      "hsgradrate" "childpov"
```

Summarise the Data

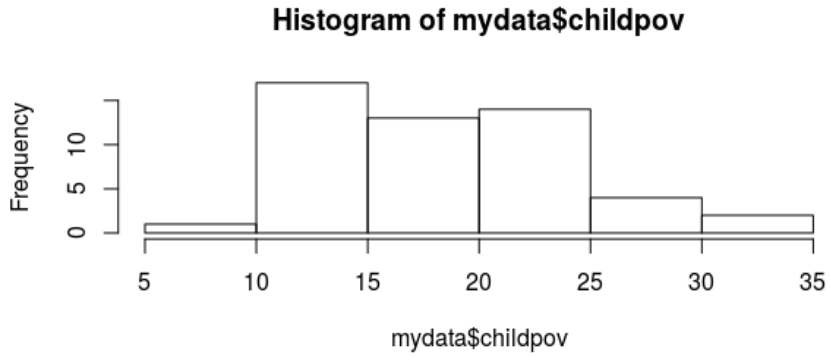
We can calculate the min,max, mean and median of the high school graduation rate as follows:

```
> mean(mydata$hsgradrate)
[1] 72.20196
> median(mydata$hsgradrate)
[1] 72.2
> min(mydata$hsgradrate)
[1] 56.2
> max(mydata$hsgradrate)
[1] 89.4
```

Visualize the Data

We can make a histogram as follows. This will put a graph in the plot window on the lower right of the screen.

```
> hist(mydata$childpov)
```



A scatter plot is also easily made.

```
> plot(x=mydata$childpov,y=mydata$hsgradrate)
```