

R болон RStudio-г ашиглаж эхлэх заавар

Энэбишийн Тэмүүлэн

22/09/2020

Эхлэл

Та бүгдтэй энэхүү хичээлийн үеэр уулзахдаа тун баяртай байна!

Бидэнд хичээлийн үеэр цаг бага байх учир доорх алхмуудыг хичээл эхлэхээс өмнөх өдрүүдэд хийж ирнэ үү.

- i) R-ыг <http://cran.r-project.org> вэбсайтаас татан авна. Татан авах буюу download хийх линкийг нь дараад өөрийн компьютерийн үйлдлийн системд (Mac эсвэл Windows) тохирохыг нь суулгана. Хэрвээ компьютертээ өмнө нь R суулгасан байсан бол хамгийн сүүлийн үеийн хувилбар мөн эсэхийг нь шалгаад биш бол дахин татаж авч суулгана уу.
- ii) RStudio-гийн үнэгүй хувилбарыг доорх линкээс татан авч суулгана.

<https://www.rstudio.com/products/rstudio/download/>.

- iii) Суулгасныхаа дараа RStudio-г нээгээд консол (зүүн доод цонх) руу дараах кодыг бичнэ:

```
install.packages(c("tidyverse", "remotes"))
```

Энэ код tidyverse болон remotes гэсэн хоёр R багцыг суулгана. tidyverse гэдэг багц нь багцуудын багц юм. Учир нь тэр дотроо бидний нарийвчлан үзэх dplyr, ggplot2 гэх мэт олон багцуудыг агуулсан байдаг. Багц бүр нь дата шинжлэх ухаанд хэрэглэгддэг функцуудын сан юм.

remotes багц нь Tidyverse-ийн хэсэг биш боловч бидэнд энэхүү сургалтын материалуудыг ашиглахад тус болох юм.

R гэж юу вэ?

- R бол програмчлалын хэл. Бид датаг цэвэрлэх, зураглах болон шинжлэхдээ үүнийг ашиглах болно.
- Та өмнө нь хэзээ ч код бичиж үзээгүй байсан ч зүгээр. Эдгээр зааврууд танд үндэс суурийг нь өгөх бөгөөд сургалтанд суухын тулд таны мэдэх ёстой бүхнийг заах болно.

Яагаад R ашиглана гэж?

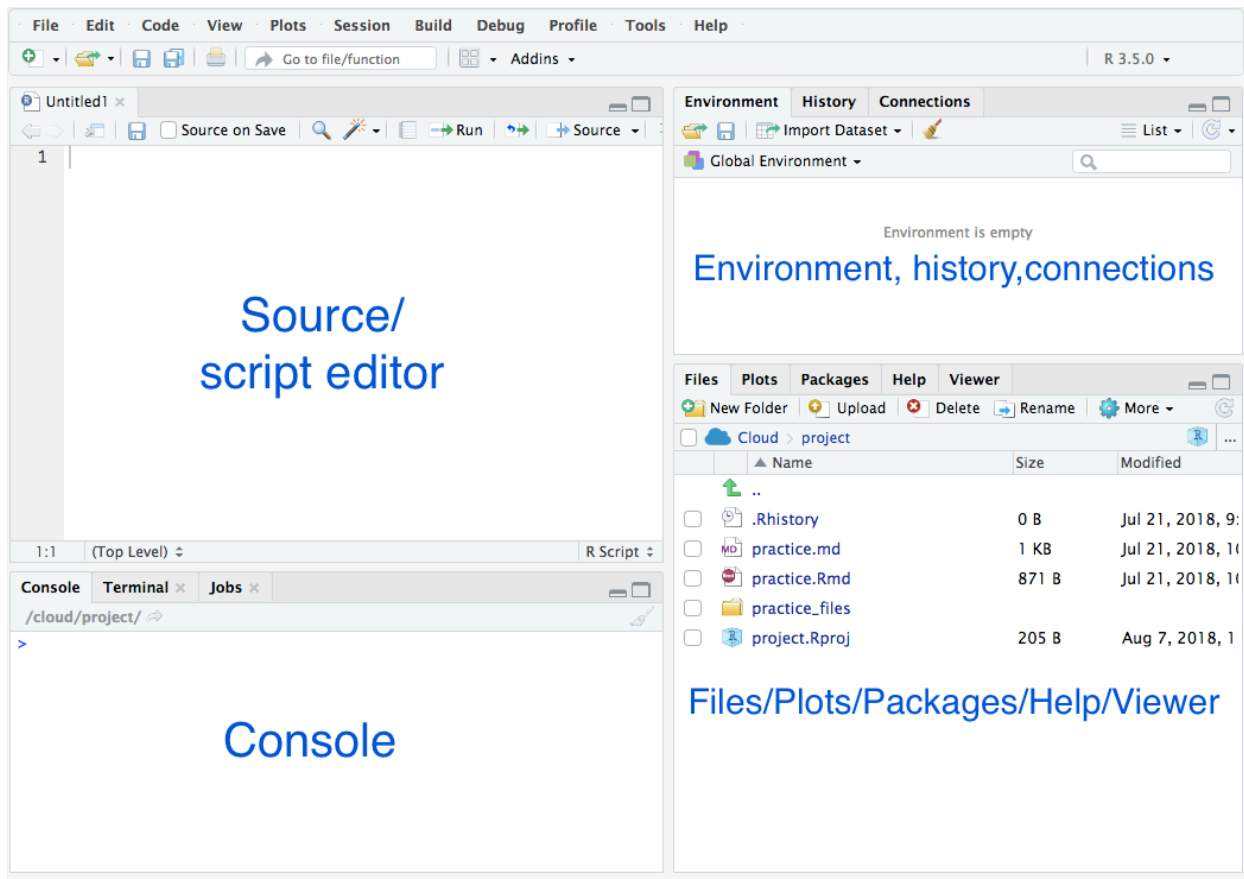
- Яагаад гэвэл R нь програмчлалын хэл тул та өөрийн хийсэн бүх алхмаа хадгалж болно. Тэгснээр:
 - Дахин кодоо ашиглах, хувилахад амар.
 - Өргөжүүлэхэд хялбар
- R нь үнэгүй бөгөөд нээлттэй эхийн програм. Тиймээс R-ыг хүссэн хүн бүр суулган ашиглаж болох бөгөөд SPSS, SAS, STATA зэрэг үнэтэй програмуудаас илүү олон нийтэд хүртээмжтэй юм.
- R нь маш уян хатан бөгөөд үргэлж цаг үетэйгээ нийлэн хувьсан өөрчлөгдөж байдаг.

RStudio гэж юу вэ?

RStudio нь код хөгжүүлэлтийн цогц орчин (integrated development environment, IDE) юм. RStudio-ийн ачаар R дээр код бичих нь илүү амар болсон. RStudio байхгүй байсан бол жирийн R нээхэд бидэнд зүгээр кодоо оруулах консол л харагдах болно.

RStudio-г ашиглах нь

Компьютер дээрээ RStudio-г нээнэ. Анх нээхэд буюу өгөгдмөл (default) хэлбэрээрээ RStudio доорх дөрвөн цонхтойгоор харагддаг:



Өнөөдөр та бүхэн эдгээр дөрвөн цонхтой танилцахын тулд дасгалууд гүйцэтгэх болно.

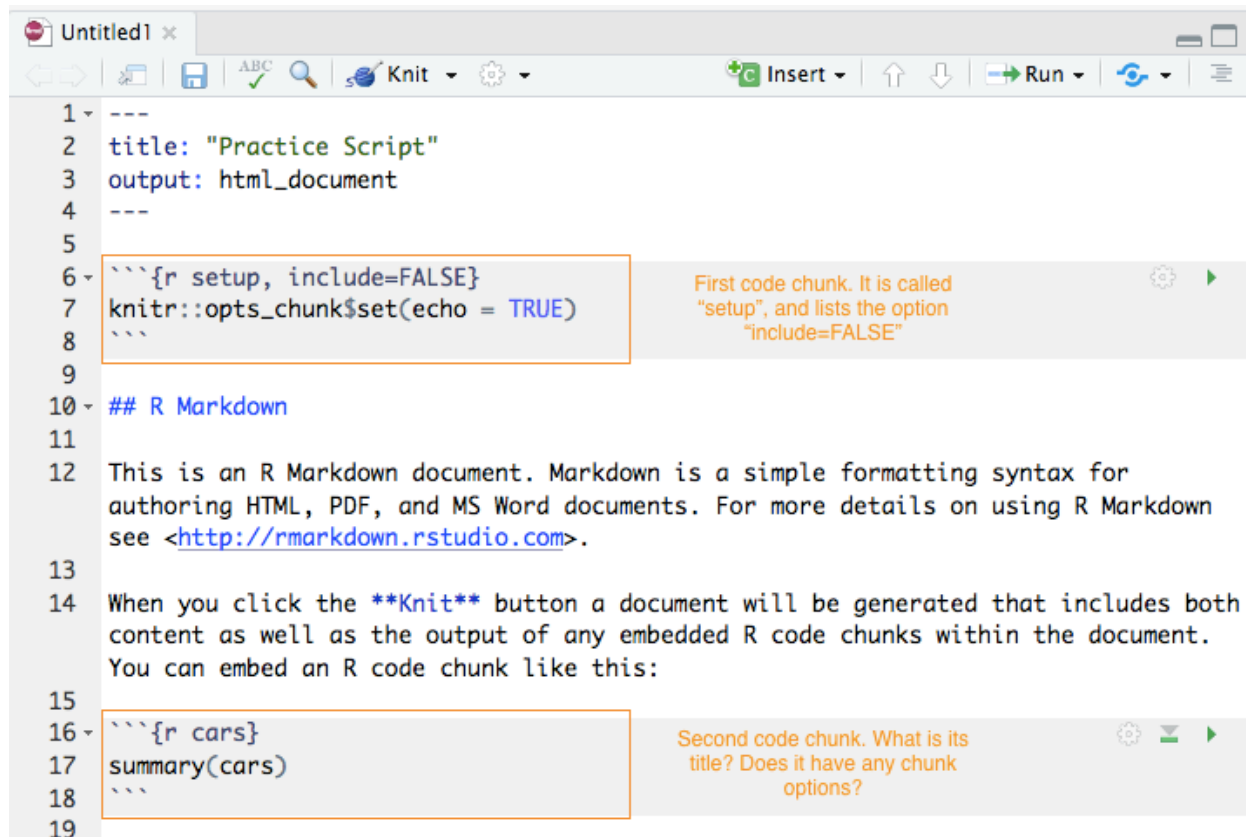
Цонх 1. Консол (зүүн доод цонх) Энэ бол таны кодоо оруулж уншуулдаг хэсэг. Тоон үр дүнгүүд Консол дотор харагдах болно.

- **Дасгал 1:** Консолыг тооны машин байдлаар хэрэглэ. Консол дотор дараад 2+2 гэж бичээд Enter дар. Дараа нь нэмэх (+), хасах (-), үржүүлэх (*), хуваах (/), болон зэрэг дэвшүүлэх (^) үйлдлүүдийг хийж үзээрэй. Олон үйлдлүүд зэрэг хийхдээ бөөрөнхий хаалтыг ашиглаарай.
- **Дасгал 2:** Та шинэ ажилдаа сард хэдэн төгрөг авах вэ гэдгээ тооцмоор байна гэж бодъё. Цалин нь цагт 50 000₮ бөгөөд та 7 хоногт 15 цаг ажиллана, гэвч таны цалингаас 20%-ийн татвар авагдах болно. Та 1 сар ажилласны дараа таны татварын дараах цалин хэд байх вэ? Консолыг тооны машин байдлаар ашиглан хариугаа бодож гаргана уу.
- **Дасгал 3:** Оноох операторыг (<-) ашиглан өмнөх дасгалын хариултаа сарын_цалин хэмээх обьектод хадгална. Оноох операторыг хэрэглэхдээ шинэ обьектоо зүүн гар талд, утгыг нь баруун гар талд

байрлуулна. Жишээ нь, консол дотор `a <- 2` гэсэн кодыг бичихэд 2 гэсэн утга `a` хэмээх объектод хадгална.

Цонх 2. Үүсвэр (зүүн дээд цонх) - мөн код бичигч эсвэл скрипт бичигч хэмээдэг. **Консол** нь шууд харьцан ажиллах (интерактив) зориулалттай байхад **үүсвэр** нь бүрэн хэмжээний анализ хийхэд зориулсан олон мөр код хадгалдаг хэсэг юм. Та компьютерийн гарны товчлолууд (shortcuts) болон айконууд ашиглан үүсвэрээс консол руу кодоо илгээж болно. Мөн та энэхүү “кодны ном”-оо хадгалаад дараа нь нээх боломжтой. Эсвэл өөр хүний бичсэн кодыг ч энэхүү цонхонд нээж болно.

- **Дасгал 4:** Анхны скриптээ үүсгэнэ үү. File > New File > R Markdown гэсэн цэсээр орно. Файлдаа “Сургалтын скрипт” гэсэн нэр өгнө. **R Markdown** загвартай шинэ файл нээгдэнэ. Эхний мөр дээрх “title” хэсэгт таны файлдаа өгсөн нэр байгааг анхаарна уу. **R Markdown** бидэнд жирийн текст (хүн унших өгүүлбэрүүд) болон кодыг (компьютер унших өгүүлбэрүүд) хослуулах боломж олгодог юм. Код нь **R кодын хэсгүүдэд (R code chunks)** бичигдэх бөгөөд гурван хойшоо харсан тэмдэгээр (“”) эхэлдэг, дуусдаг байна. Эхлэл тэмдгийн бүрдэл нь R код бичиж байгааг илтгэх “r” агуулах бөгөөд мөн энэ хэсэгт кодын агуулгыг илэрхийлэх нэр өгч болно. Таслалын дараа **хэсгийн тохируулгууд** орж ирж болдог бөгөөд үүний талаар бид дараа ярилцах болно. Доорх зураг нь өгөгдмөл (default) байдлаар үүсгэсэн R Markdown файлын эхний 19 мөр кодыг харуулан кодын хэсгүүдийг тайлбарласан байна:



```
1 ---
2 title: "Practice Script"
3 output: html_document
4 ---
5
6 ```{r setup, include=FALSE}
7 knitr::opts_chunk$set(echo = TRUE)
8 ```
9
10 ## R Markdown
11
12 This is an R Markdown document. Markdown is a simple formatting syntax for
13 authoring HTML, PDF, and MS Word documents. For more details on using R Markdown
14 see <http://rmarkdown.rstudio.com>.
15
16 When you click the **Knit** button a document will be generated that includes both
17 content as well as the output of any embedded R code chunks within the document.
18 You can embed an R code chunk like this:
19
20 ```{r cars}
21 summary(cars)
22 ```
```

- **Дасгал 5:** Та өөрийн үүсгэсэн Сургалтын скриптэнд байгаа текстийг унш. **Knit** гэсэн товчийг дар. Файлаа “Сургалтын скрипт” гэж нэрлээд хадгал. “Knit” дарсны дараа үүссэн файлыг (шинэ цонхонд нээгдэх ёстой) код бичигч дотор буюу скрипттэй харьцуул.
 - Энэхүү вэбсайтын зураг 2.1-ийг хар: <http://socviz.co/gettingstarted.html>. Зураг 2.1-ээс markdown-д бичигдсэн зүйлс гаралтан (output) дээр хэрхэн харагдаж байгааг анхаарна уу.

- **Дасгал 6:** Сургалтын скриптийн markdown файлын 5-28 дахь мөрүүдийг устга. Ингэснээр зөвхөн толгой хэсгийн мэдээлэл үлдэх ёстой (дээр, доороо “—” багтаасан дээд талын дөрвөн мөр). Файлаа knit хийгээд Viewer цонхонд юу гарч ирж байгааг хар.
- **Дасгал 7:** Файлынхаа нэрийг (1-р мөр) “Цалингийн тайлан” болгон шинэчил. Тэгээд 6-р мөрнөөс эхлээд дараах өгүүлбэрийг бич: “Сарынхаа цалинг тооцоолохын тулд би цагийнхаа цалинг ажилласан цагтаа үржүүлээд үр дүнг нь нэг сард байдаг долоо хоногийн тоогоор үржүүлсэн. Эцэст нь татварыг оруулахын тулд үржвэрийг нь би 80%-иар үржүүлсэн.” Одоо файлаа дахин Knit хийнэ үү.
- **Дасгал 8:** 10-р мөр дээр кодын хэсэг нэм. Ингэхийн тулд загвар файл дотроо код нэмэхийг хүссэн хэсэгтээ очоод **Insert** товчийг дарж гарч ирсэн цэснээс R сонгоно. Энэхүү үйлдлийн гарны товчлол нь Windows үйлдлийн систем дээр Ctrl + Alt + i Mac үйлдлийн систем дээр option + cmd + i байдаг.
- **Дасгал 9:** Шинээр үүсгэсэн кодын хэсэг дотроо дараах кодыг бич: сарын_цалин <- 20 * 15 * 4 * 0.8. Кодын хэсгээ “цалин тооцоолох” гэж нэрлэ (2 дахь зургийг харж хэрхэн кодын хэсгийг нэрлэхээ өөртөө сануулж болно).
- **Дасгал 10:** Дараагийн мөрөнд нь (кодын хэсгийнхээ дотор) сарын_цалин гэж бич. Бид энэ мөрийг уншуулах үед R-аас бид сарын цалингийн хувьсагчид оноогдсон утгыг бидэнд үзүүлэхийг асууж байна гэсэн үг. Кодын хэсгийг уншуулахдаа түүний ногоон сумыг нь дарж эсвэл курсороо тухайн мөрөн дээр авчраад PC дээр Ctrl + enter Mac дээр cmd + enter дарна.

Цонх 3. Environment/History/Connections: Яг одоо environment таб дотор юу байна вэ? Энэ таб нь бидэнд **Environment** дотроо хадгалсан **объектүүдийг** харуулж байгаа юм. Шинээр дата ачааллах үед түүний объектыг бид мөн энд харж болно. Одоо History таб дээр дар. Энд юу байна? Кодны мөрөн дээр дараад “To Console” эсвэл “To Source” гэсэн товчны аль нэгийг дар. Юу болж байна? History таб нь консол дотроо бичсэн боловч код бичигч дотроо хадгалаагүй байсан кодыг дахин хэрэглэх шаардлагатай үед дахин дуудах боломжийг олгодог юм. Connections табын талаар бид ярилцахгүй.

Цонх 4. Files/Plots/Packages/Help/Viewer Эдгээр табууд бүгд их хэрэгтэй. Files таб нь файлуудаа харах хэсэг. Таны фолдер дотор ямар файлууд байна? Plots таб нь консол дотроо уншуулсан кодыг ашиглан бүтээсэн бүх плотыг харуулна. Packages таб нь ямар R багцууд суусан байгааг жагсаалт хэлбэрээр харуулдаг бөгөөд яг одоо ачаалсан байгаа нь тэмдэглэгдсэн байдаг. Бид дараагийн хэсэгтээ R багцуудын талаар дэлгэрүүлэн ярих болно. Help таб тусламжийн файлуудыг хайхад хэрэглэгдэнэ. Viewer табыг бусад төрлийн гаралтыг (ихэнхдээ html гаралт) харахад ашигладаг. Жишээ нь, бид дээрх R markdown файлыг knit хийхэд html файл үүссэн.

R багц (package) гэж юу вэ?

- R багц гэдэг нь хэн нэгний бүтээсэн **функцуудын** бүрдэл юм. Жишээ нь, ggplot2 багц нь плотууд хийдэг функцуудын бүрдэл байхад dplyr багц нь шинэ хувьсагч үүсгэх, нэрийг нь өөрчлөх гэх мэт дата цэвэрлэх, өөрчлөх функцуудын бүрдэл юм.
- Та R анх ачааллах үед үндсэн суурь багцууд хамтдаа ачаалдаг. Одоогоор R хөгжүүлэгчид нэмэлтээр 10,000 илүүтэй багц бүтээгээд байна. Эдгээрийг R багцуудын үндсэн агуулах болох **CRAN** (The Comprehensive R Archive Network) сүлжээнээс татан авч болно.

Дасгалуудын эхэнд бид хоёр багцыг суулгасан. Тэгвэл тэдгээрийн доторх функцуудыг ашиглахын тулд бид library хэмээх функцийг ашиглан тэдний кодын санг ачаалдаг. Доорх кодыг консол руугаа бичээд эдгээр багцыг ачаална уу.

```
library(dplyr)
library(ggplot2)
```

Tidyverse багц нь tidyverse системийн цөм болсон 8 багцыг ачаалдаг мета-багц юм. Та аль хэдийн хоёрыг нь буюу dplyr, ggplot2-ыг ачаалсан. Доор буй кодыг уншуулан гаралт нь ямар байгааг хар.

```
library(tidyverse)
```

Tidyverse багцуудыг ачааллахад танд халхлагдаж (masked) буй функцуудын талаар “conflicts” буюу “зөрчлүүд” хэмээх хэсэг харагдаж магад. Энэ нь алдаа биш бөгөөд тэд таны ачаалсан багцад багтсан функцуудтай ижил нэртэй функцууд өөр газар мөн байгааг харуулж буй юм. Бид үүний талаар тун удахгүй ярилцана.

Функц (function) гэж юу вэ?

Бид функцыг амархан таньж болно. Учир нь тэд үргэлж хоёр бөөрөнхий хаалтаар () төгсөж байдаг. **Функцын** өөр нэр нь **команд**. Жишээ нь, `library()` бол функц юм. Бөөрөнхий хаалтан дотор байгаа зүйлүүдийг тухайн **функцын аргумент** гэдэг. Ихэнх функцууд хэд хэдэн аргументтэй байдаг. Бидний дээр бичсэн код нь эхний мөрөн дээр `dplyr` багцыг, хоёр дахь мөрөн дээр `ggplot2` багцыг ачаалахыг тулд `library()` гэсэн функцыг ашиглаж байна.

Багцыг ачаалахад юу болдог вэ?

Packages табыг (баруун доод цонхонд буй) дараад та өөрийн ачаалсан багцынхаа нэрийг хайж ол. Тэрхүү багц нь одоо ачаалагдсан байгаа учир тэмдэгтэй байх ёстой. Заримдаа багц ачаалах үед мессеж гарч ирдэг. Эдгээр мессежүүдийг одоохондоо учрыг нь ойлгохгүй байгаадаа санаа зовох хэрэггүй.

Тусламж!

Функц болгон тухайн функц юу хийдэг болон ямар аргуудыг оруулах ёстой талаар туслах хуудастай байдаг. Хоёр замаар функцын туслах хуудсыг харж болно:

```
help("library")
?library
```

Бид дээрх хоёр кодын аль нэгийг уншуулахад Help таб онгойж `library()` функцын туслах хуудас руу аваачих болно. Ийм байдлаар туслах хуудас дуудах нь хэрэглэмээр байгаа функцынхаа нэрийг мэддэг боловч яг яаж ашиглахаа мартсан тохиолдолд хамгийн тустай байдаг. Туслах хуудсанд буй бүх зүйлсийг хараахан ойлгохгүй байх нь зүгээр ээ, та яваандаа ойлгодог болно. Хэрвээ туслах хуудаснууд хэрхэн бэлтгэгддэг талаар илүү сайн ойлгомоор байвал энэхүү вэбсайтыг хараарай <http://socviz.co/appendix.html#a-little-more-about-r>

Хэрвээ танд тусламж хэрэгтэй боловч тусалцаа хэрэгтэй байгаа функцынхаа нэрийг санахгүй байгаа бол тухайн функцтайгаа холбоотой үгийн урд давхар асуултын тэмдэг бичин хайж болно:

```
??histogram
```

Энэ нь танд хистограмын талаарх бүх багцууд, тэдгээрийн функцуудыг харуулна. Жишээ нь, `ggplot2::geom_freqpoly()` гэдэг нь `ggplot2` багцад `geom_freqpoly()` гэдэг хистограм зурдаг функц байна гэсэн үг. Та өөрийн хайж буй функцаа олтоо энд буй функцуудыг хайж болно.

Эцэст нь хэлэхэд, Google таны хамгийн сайн найз. Ерөнхийд нь хэлэхэд, таны ихэнх асуултын хариу Stack Overflow хэмээх вэбсайтад байгаа. Өөр тус болох вэбсайт нь RStudio Community.