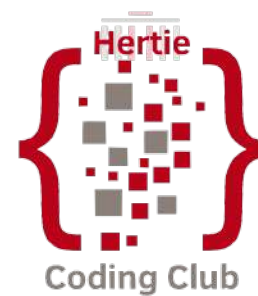


Hertie Coding Club



(Re) Introduction to R

Jorge Roa

Welcome







Agenda for today

- Objective of the club. Why this club?
- Who we are?
- Motivation
- Dinamic of the sessions
- Slack
- GitHub and GitHub Desktop
- R and RStudio
- Customize our RStudio appearance
- Create our first project
- RStudio elements
- Working directory
- Packages
- Objects
- Summary
- Next session

Objective of the club.
Why this club?

Objective of the club. Why this club?

-  Data is a relevant trend and us, as future policy makers, must have those skills.
-  **Programming:** You are developing your problem-solving skills. The time you spend learning how to code and resolving code problems makes us think more logically.
-  **Gap:** A lot of people want to learn code, but you must take Stats I and Stats II to have the right to apply to the Introduction to Data Science Course. Why should we not start learning from the first semester?
-  **Democratize learning code:** No matter if you are from the MPP, MIA, or another program, everyone should learn how to code and possess those abilities; we must exploit our potential.

Who are we?

Who are we?

Instructors



Gabriel Zech

MDS 2023

Project Manager

(Data Science Lab)
Bertelsmann Stiftung



Carmen Garro

MDS

2024



Jorge Roa

MDS

2024



Lukas Warode

MDS 2023

Data Researcher

POLITICO



Rodrigo Dornelles

MDS

2024

Who are we?

Communications



Ileana Marroquin (MPP 2023)
Coordinator alligator



Ridhima Singh (MPP 2024)
Coordinator alligator



Mary Nguyen (MPP 2023)
Celebrity promoter (Social Media)



Chhengpor Aun (MIA 2024)
Celebrity promoter (Social Media)



Lisa Bewermeier (MPP 2024)
Celebrity promoter (Social Media)



Natascha Schoepl (MPP 2023)
Central office emissary



Abigail Pena Alejos (MPP 2023)
Central office emissary



Kaja Buxrud (MPP 2024)
Promotion emotion



Issac John (MPP 2024)
Promotion emotion



Fernando Segovia (MIA 2024)
Promotion emotion

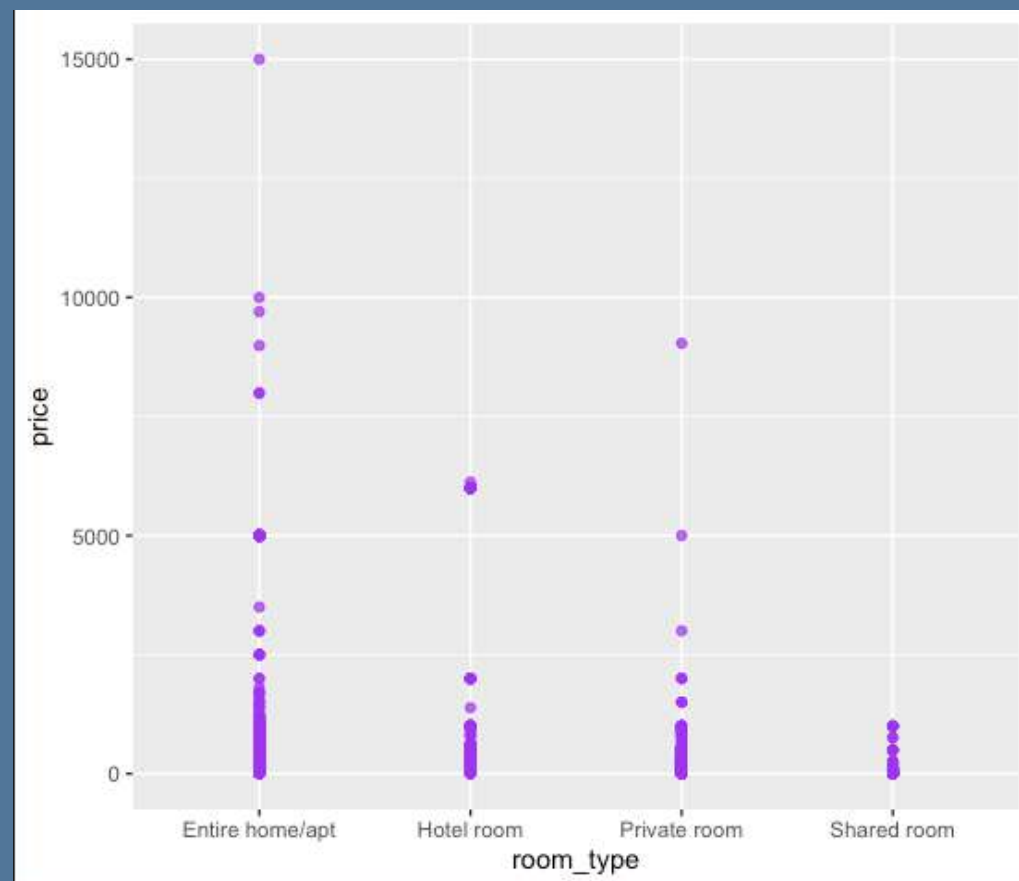


Aayran Salman (MPP 2024)
Canva Wizard

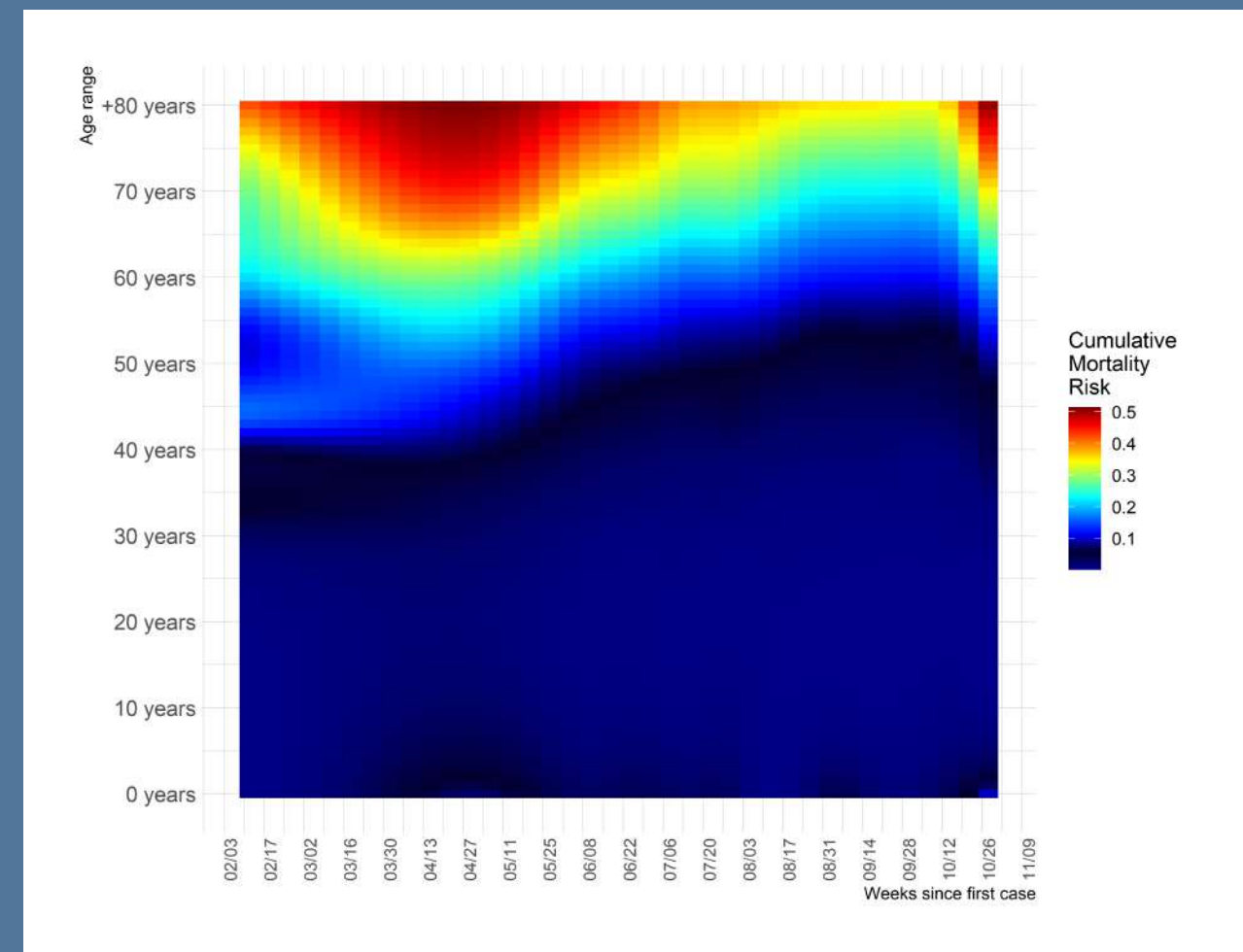
Motivation

Motivation

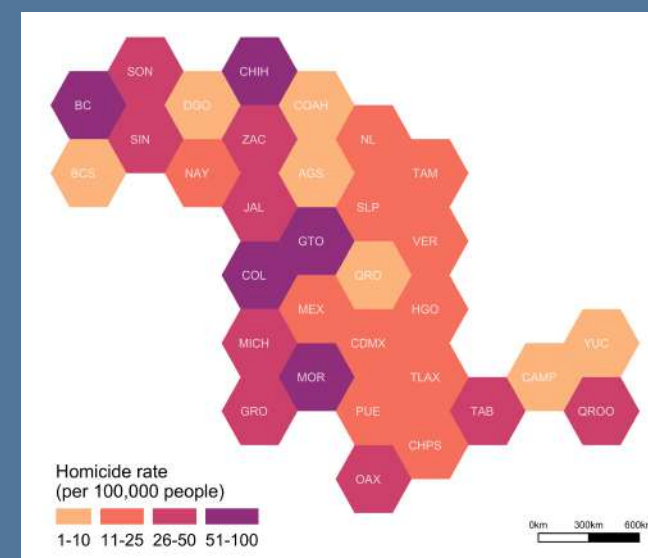
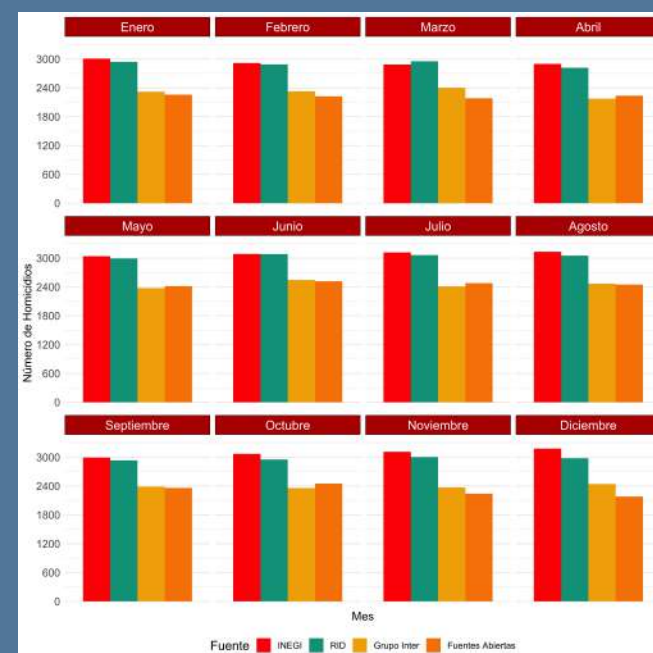
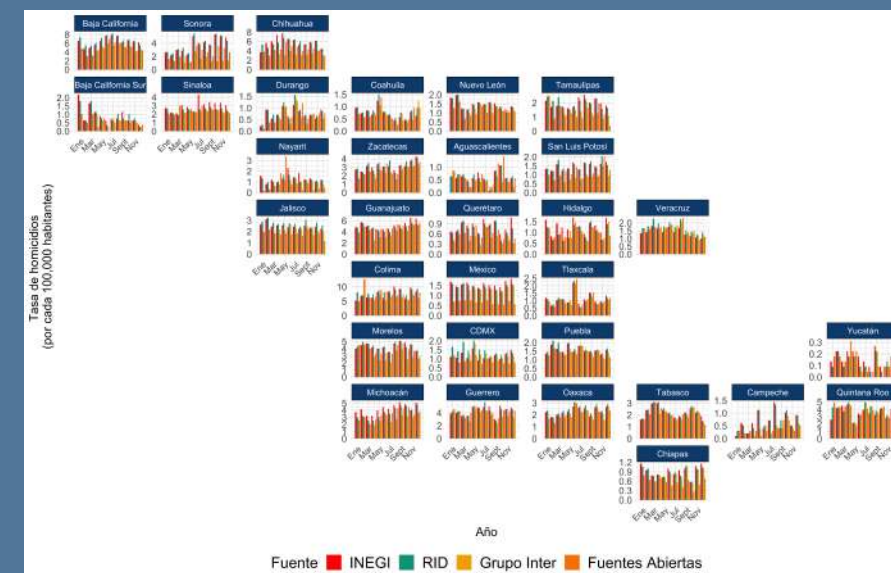
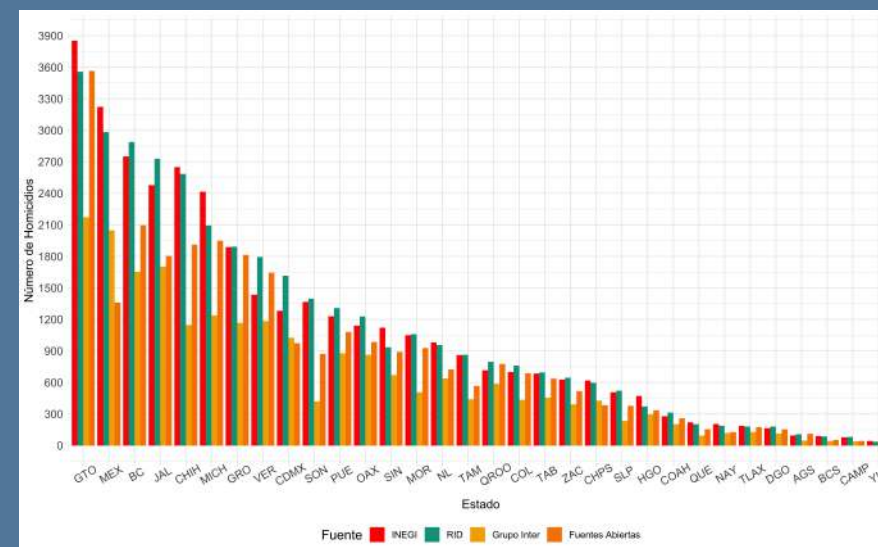
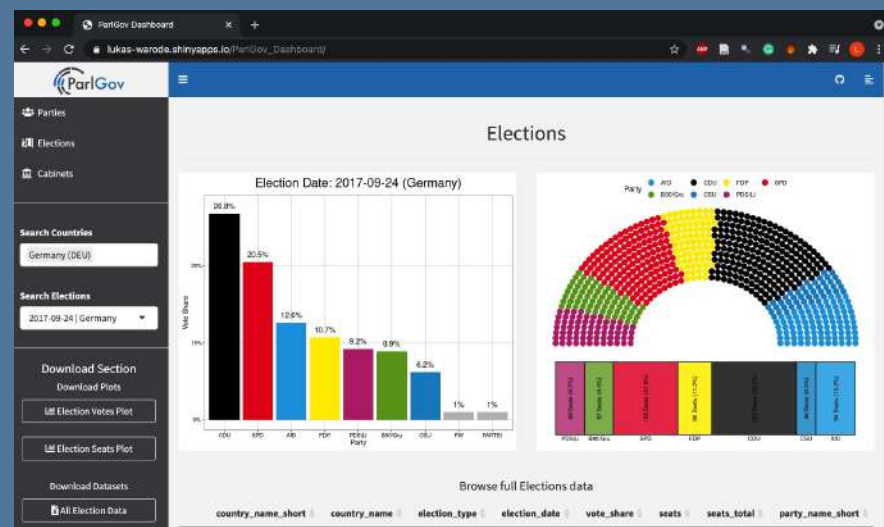
This was my first graph in R



How I improved



Motivation



Motivation

Víctimas en Ciudad de México por tipo de delito y año				
Delito	2019	2020	2021	Promedio de % Crecimiento
Encubrimiento Por Favorecimiento	1	25	357	1864.0
Delitos Electorales	9	22	533	2233.6
Robo De Vehículo En Pensión, Taller Y Agencias S/V	1	18	25	869.4
Daño En Propiedad Ajena Intencional A Vías De Comunicación	4	45	333	832.5
Contra La Intimidad Sexual	56	808	1534	716.4
Retención De Menores	12	132	575	667.8
Tala	7	69	148	500.1
Usurpación De Identidad	1	5	30	450.0
Enriquecimiento Ilícito	1	6	15	325.0
Daño En Propiedad Ajena Culposa Por Tránsito Vehicular A Vías De Comunicación	9	51	56	238.2
Ddh Anónimas	35	143	188	170.0
Daño Suelo (Actividad, Invasión O Extracción)	13	30	89	163.7
Robo A Pasajero En Trolebus Sin Violencia	39	158	171	156.7
Ataque A Las Vías Generales De Comunicación	116	233	644	136.6
Robo A Pasajero En Ecobus Sin Violencia	1	2	5	125.0
Robo A Transeunte En Parques Y Mercados Con Violencia	347	629	1470	107.5
Contaminación O Residuos	15	42	49	98.3
Cambio De Uso De Suelo	17	31	62	91.2
Ataque A Las Vías De Comunicación (Daño A Vías O Medios De Transporte)	33	57	111	83.7
Inhumación, Exhumación Y Respeto A Los Cadáveres O Restos Humanos	8	12	26	83.3
Ddh Sin Datos	60	114	145	58.6
Ddh Cerso	27	54	61	56.5
Violación De Correspondencia	21	36	51	56.5
Violación Equiparada	291	454	688	53.8
Tráfico De Influencia	14	17	28	43.1
Tentativa De Violación	17	19	33	42.7
Robo A Transeunte En Negocio Con Violencia	440	516	858	41.8
Robo De Placa De Automovil	796	1206	1591	41.7
Robo A Pasajero A Bordo De Transporte Público Sin Violencia	413	468	792	41.3
Tortura	223	348	414	37.5
Robo A Pasajero En Rtp Sin Violencia	11	12	18	29.5
Ddh Relacionadas	213	313	326	25.6
Uso Indevido De Atribuciones Y Facultades	34	49	52	25.1
Pérdida De La Vida Por Enfermedad	126	128	186	23.4
Delitos De Abogados, Patronos, Litigantes Y Asesores Jurídicos	106	109	156	23.0
Tentativo De Fraude	27	34	39	20.3
Pérdida De La Vida Por Otras Causas	1241	1347	1686	16.9
Violencia Familiar	27443	30064	37068	16.4
Homicidio Culposo Por Tránsito Vehicular (Colisión)	282	369	370	15.6
Abuso De Confianza	3782	3837	4932	15.0
Daño En Propiedad Ajena Intencional A Casa Habitación	528	617	687	14.1
Contra El Cumplimiento De La Obligación Alimentaria	561	581	700	12.0
Robo De Fluidos	75	83	94	12.0
Robo De Vehículo De Pedales	1455	1545	1807	11.6
Amenazas	14094	14952	17354	11.1
Violación Equiparada Por Conocido	23	24	28	10.5
Fraude	14290	14478	17072	9.6
Robo A Pasajero En Trolebus Con Violencia	20	21	23	7.3
La Administración De Justicia	991	1033	1128	6.7
Despajo	3808	4118	4225	5.4

Elaboración con datos de la Fiscalía General de Justicia de la Ciudad de México

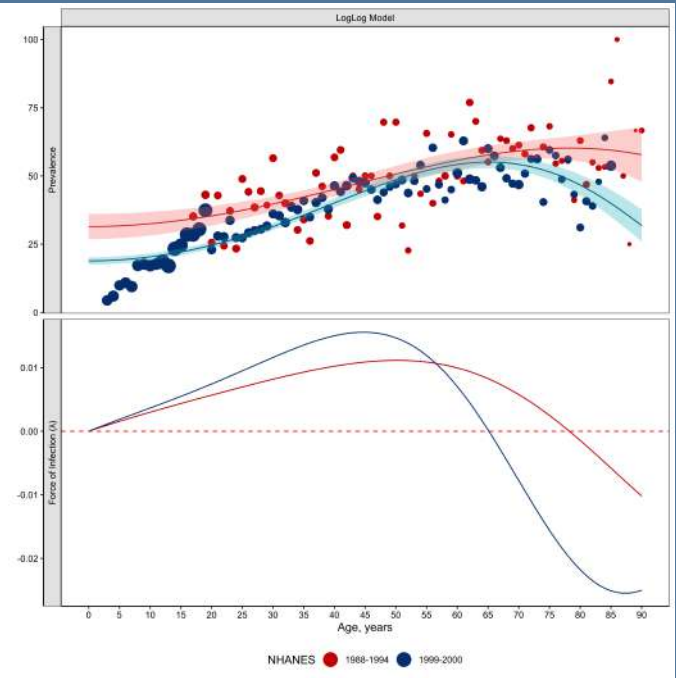
Pruebas de series de tiempo por tipo de delito en 2021			
Delitos	T-Test ¹	ADF-Test ²	Trend
Abuso Sexual	NS ³	NS	
Acoso Sexual	NS	***	
Amenazas	NS	NS	
Cobranza Ilegítima	NS	***	
Cohecho	NS	NS	
Daño En Propiedad Ajena Intencional A Automovil	NS	NS	
Ddh Otras Materias	NS	NS	
Ddh Relacionadas	NS	NS	
Falsedad Ante Autoridades	NS	NS	
Falsificación O Alteración Y Uso Indevido De Documentos	NS	NS	
Lesiones Culposas	NS	NS	
Lesiones Culposas Por Caída	NS	NS	
Lesiones Culposas Por Tránsito Vehicular	NS	NS	
Lesiones Intencionales Por Arma Blanca	NS	*	
Lesiones Intencionales Por Golpes	NS	NS	
Retención De Menores	NS	NS	
Robo A Pasajero A Bordo De Metro Sin Violencia	***	NS	
Robo A Transeunte A Bordo De Taxi Público Y Privado Con Violencia	NS	NS	
Robo De Objetos	NS	NS	
Tortura	NS	NS	
Violación Equiparada	***	NS	
Violencia Familiar	NS	NS	

¹ Prueba t de Student asumiendo distribución normal.
² Prueba estacionaria de Dickey-Fuller Aumentada.
³ *** Valor-p:significancia al 99%, ** Valor-p:significancia al 95%, * Valor-p:significancia al 90%
Elaboración con datos de la Fiscalía General de Justicia de la Ciudad de México

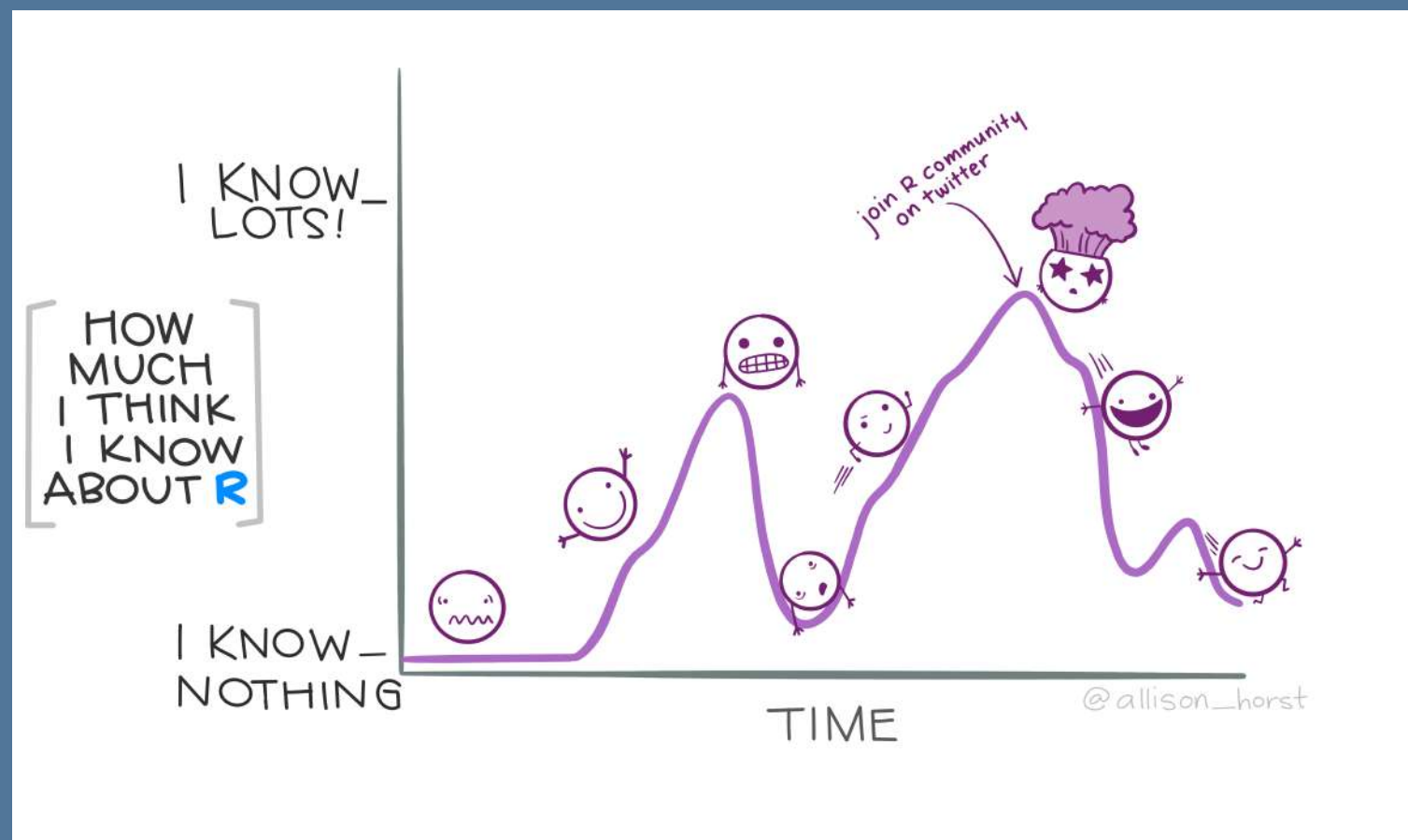
Víctimas de delitos en Ciudad de México durante 2021							
Hora	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
00:00	874	993	992	1023	992	1060	862
01:00	992	733	628	603	673	764	940
02:00	871	731	556	598	621	731	834
03:00	905	620	509	570	644	713	829
04:00	731	543	485	435	541	583	635
05:00	680	655	617	623	630	793	758
06:00	682	856	781	780	790	843	737
07:00	819	1145	967	975	964	1101	863
08:00	1000	1500	1261	1324	1239	1422	1083
09:00	1147	1930	1765	1876	1766	1762	1328
10:00	1656	2671	2490	2456	2533	2615	1885
11:00	1462	1997	2042	2051	2082	2038	1697
12:00	2519	4485	4354	4399	4352	4445	2950
13:00	1160	1743	1701	1772	1622	1678	1409
14:00	1377	1863	1907	1847	1838	1845	1562
15:00	1359	1843	1831	1895	1797	1849	1543
16:00	1359	1693	1707	1709	1621	1701	1453
17:00	1376	1574	1568	1677	1529	1605	1351
18:00	1351	1588	1633	1564	1554	1594	1351
19:00	1337	1525	1570	1565	1469	1491	1273
20:00	1424	1522	1509	1584	1495	1603	1475
21:00	1314	1236	1400	1337	1274	1516	1404
22:00	1190	1035	999	1040	1077	1214	1320
23:00	1015	763	796	808	936	1129	1167

Elaboración con datos de la Fiscalía General de Justicia de la Ciudad de México

Resultados de modelo lineal			
Estado	Intercepto	Pendiente	Significancia
Aguascalientes	0.5172	-0.0002	-
Baja California	6.9525	0.0023	-
Baja California Sur	0.7883	-0.0003	-
Campeche	0.4543	0.0024	-
Chiapas	0.6753	0.0047	**
Chihuahua	7.0099	-0.0076	**
Ciudad de México	1.4195	-0.0009	-
Coahuila	0.5992	0.0003	-
Colima	7.1268	-0.0075	-
Durango	0.9703	-0.0003	-
Estado de México	1.4550	-0.0006	-
Guanajuato	6.1729	-0.0005	-
Guerrero	3.3311	-0.0016	-
Hidalgo	1.0219	-0.0008	-
Jalisco	2.3481	0.0050	-
Michoacán	4.0650	0.0000	-
Morelos	4.5759	-0.0135	-
Nayarit	1.3919	-0.0021	-
Nuevo León	1.4900	-0.0008	-
Oaxaca	2.1603	-0.0038	-
Puebla	1.3892	-0.0013	-
Querétaro	0.7127	0.0007	-
Quintana Roo	2.8956	0.0013	-
San Luis Potosí	2.1926	0.0015	-
Sinaloa	2.0328	0.0027	-
Sonora	4.2197	0.0023	**
Tabasco	1.4306	0.0027	***
Tamaulipas	1.5665	0.0000	-
Tlaxcala	0.5776	0.0023	-
Veracruz	1.4574	0.0013	-
Yucatán	0.1908	0.0000	-
Zacatecas	5.2941	0.0037	-



What is the key to learn how to code



- Learning a new thing can be difficult sometimes.
- Getting errors can be scary: don't worry, we are here for you.
- In my opinion: the more errors you get, the better. Why? Because you learn from those errors. (It doesn't apply to real life :p)
- From my own experience: patience.
- The goal of the club is help us each other learn how to code.

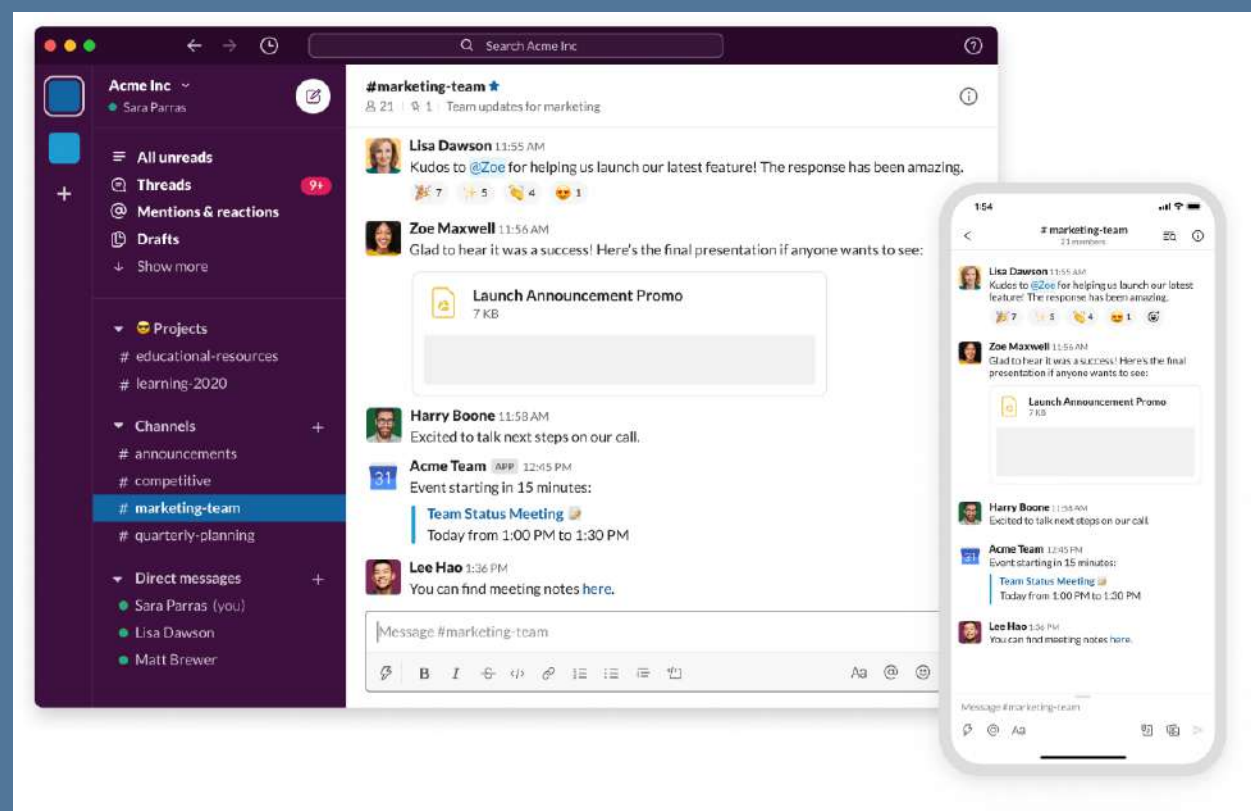
Dynamic of the sessions

Dynamic of the sessions



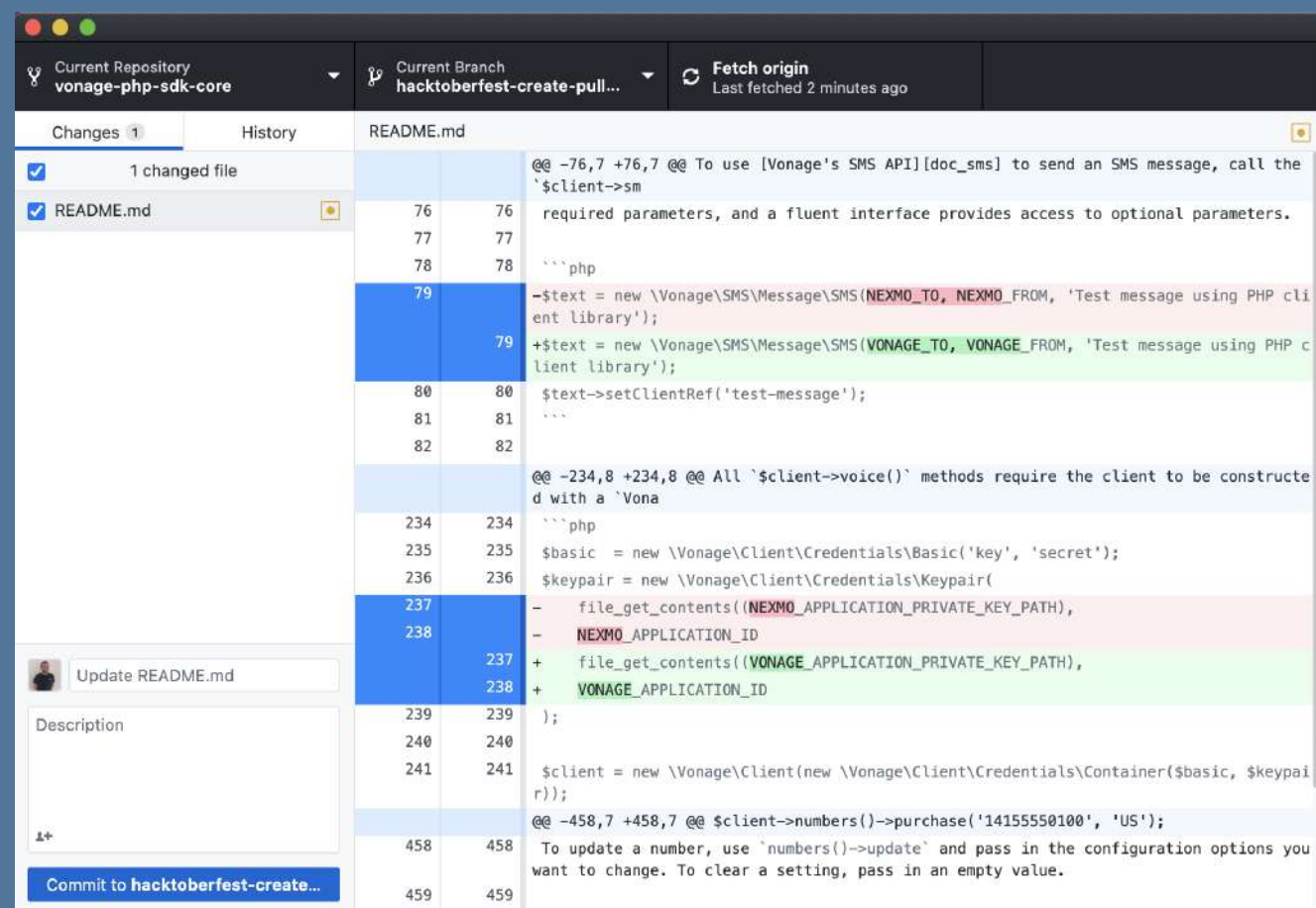
- In the short run, we will have sessions for introduction to R (weekly)
- We hope this semester, we can have specific sessions for advanced R users and other languages (Python, SQL).
- We will record the sessions for future reference.

Slack



- Permanent platform for Hertie Community
- Like a WhatsApp for code.
- Share code, share files, and ask questions.
- One of the most popular platforms in the data science world.

GitHub and GitHub Desktop



- Website and cloud-based service to store and manage code
- Git IDE: used in the programming world. It is used for tracking changes in the source code during software development.
- It makes it easier for individuals and teams to use Git for version control and collaboration.

R and RStudio

R and RStudio

Why R?

- Open source
- One of the most used language programs in the world
- Works for every type of data (shape and sizes)
- Easy to learn it (We will help you with that)
- More than 18,000 packages

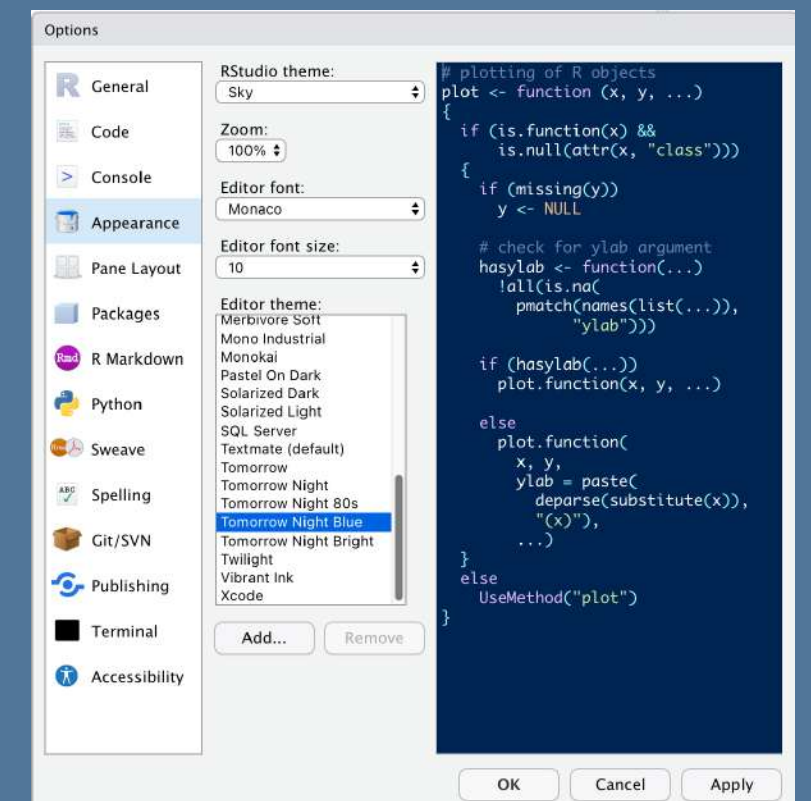
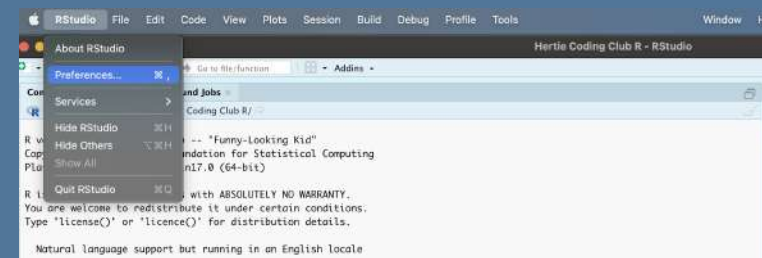
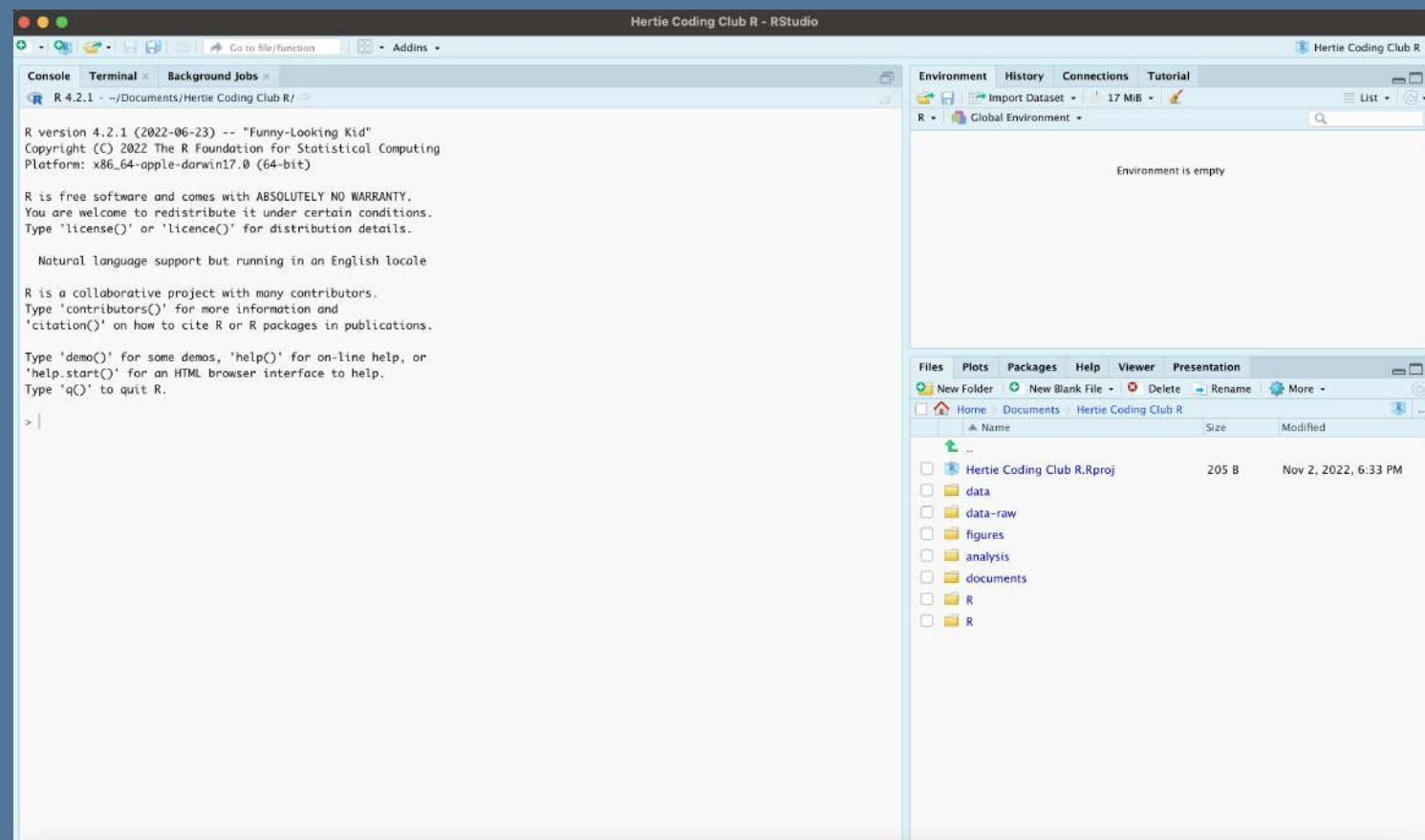


Customize our RStudio appearance

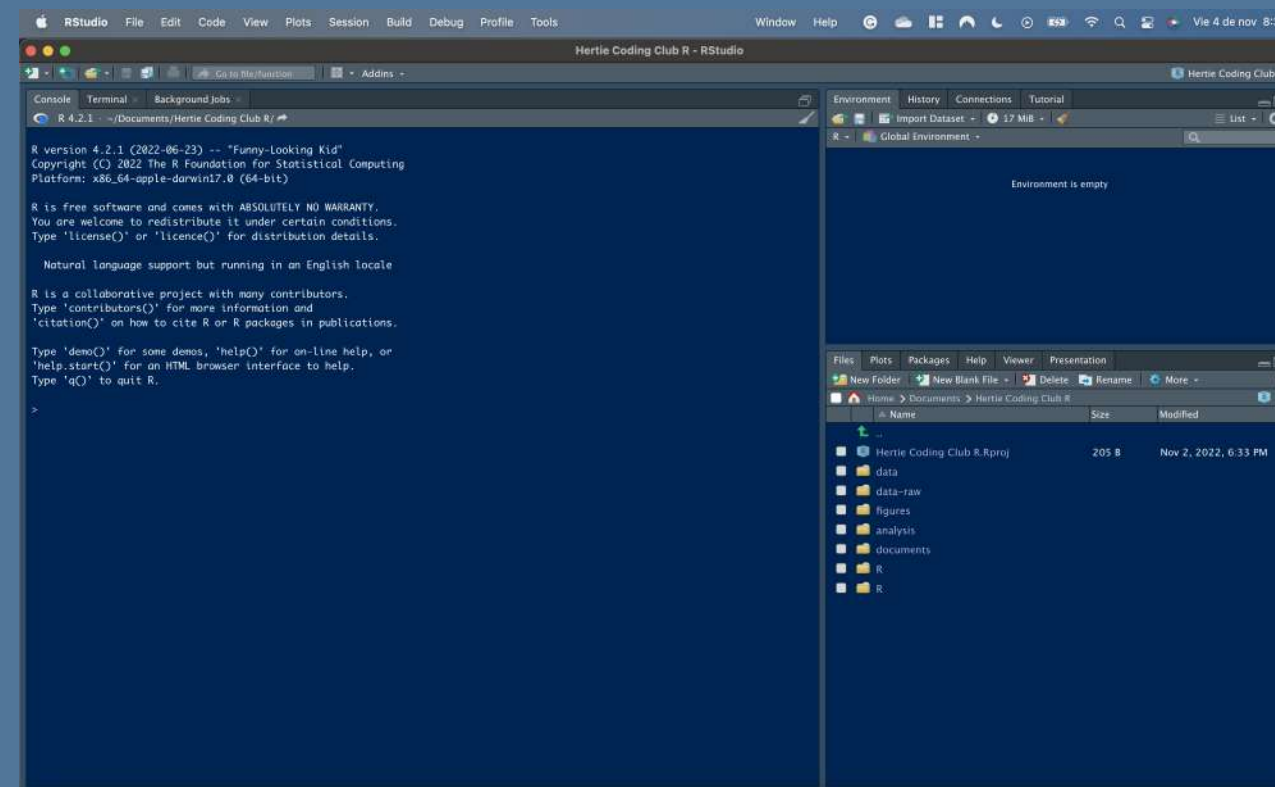
Customize our RStudio appearance

Welcome to RStudio

- First window that appears when you open RStudio
- Let's change the appearance of RStudio



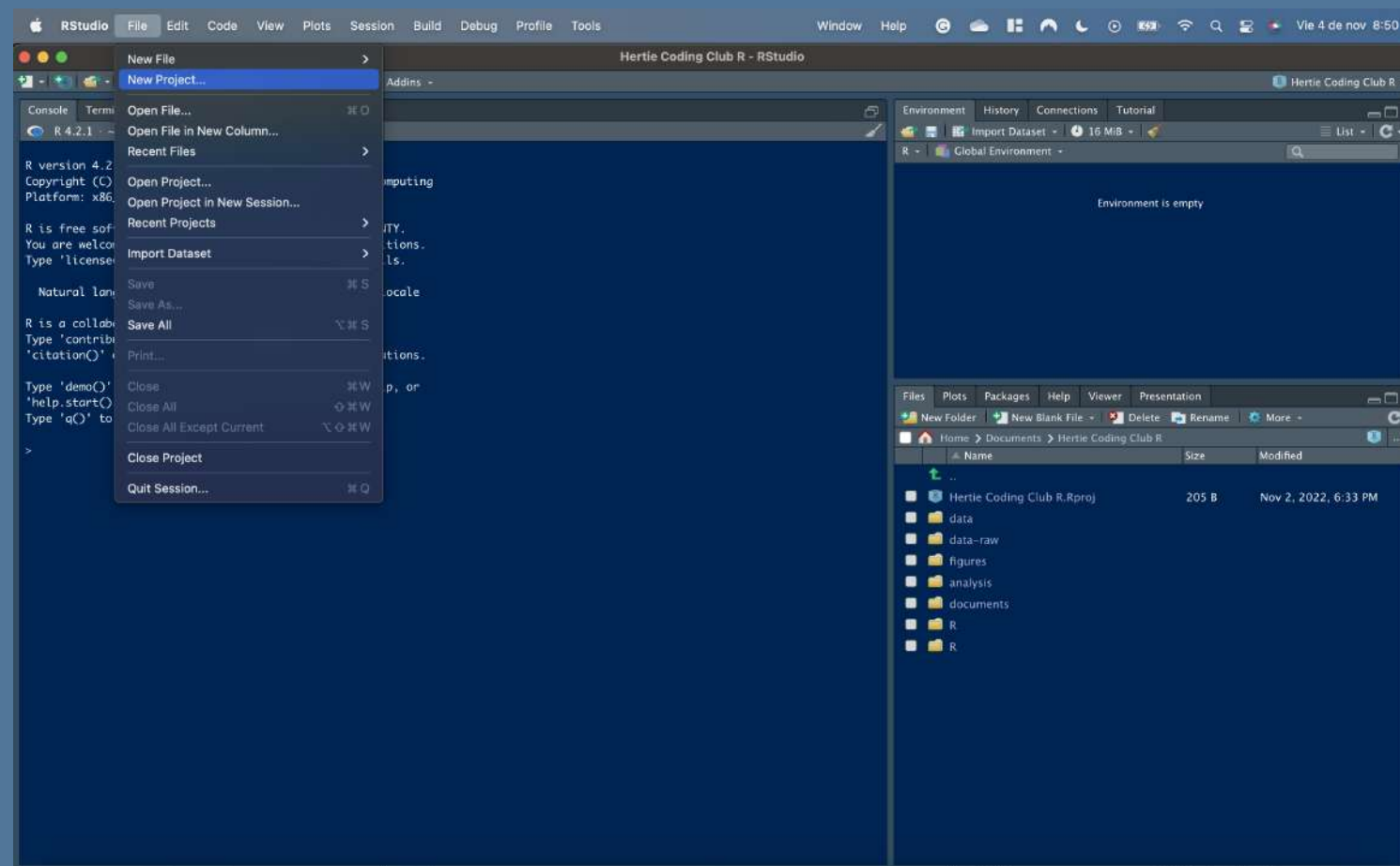
Customize our RStudio appearance



- Blue is my favorite color. For me looks awesome! What about you?

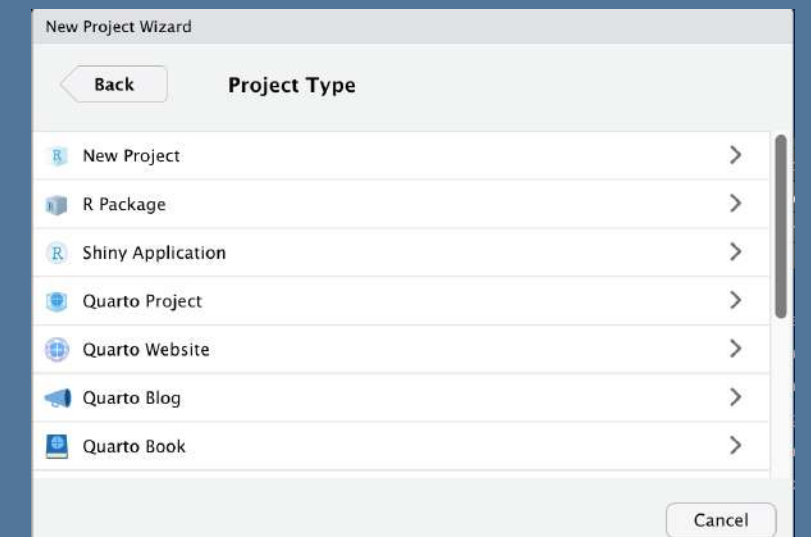
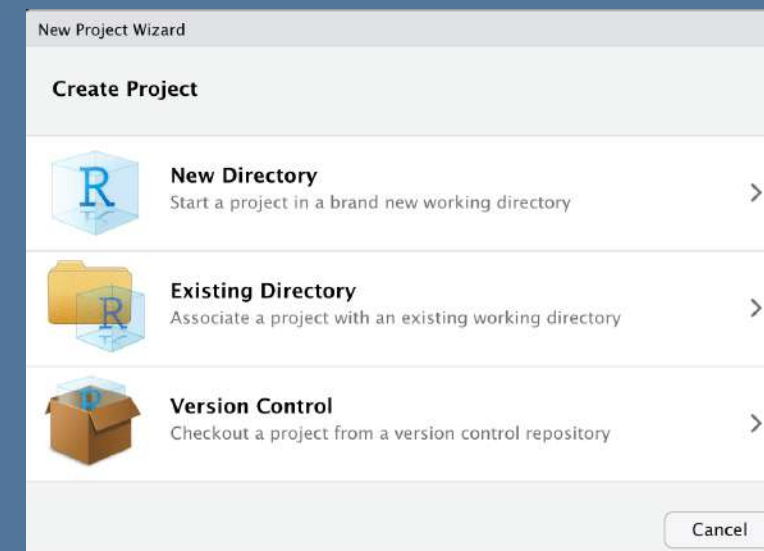
Create our first project

Create our first project (Step 1)

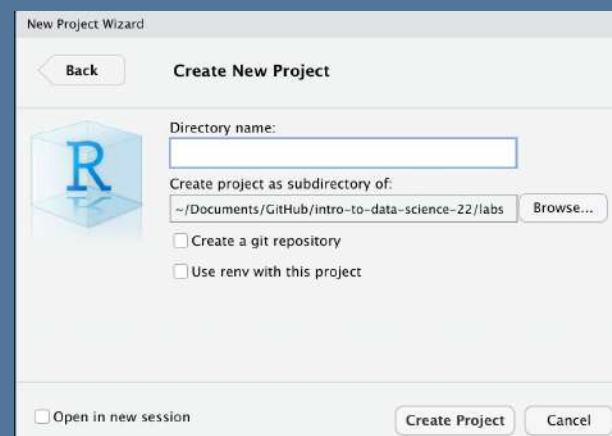
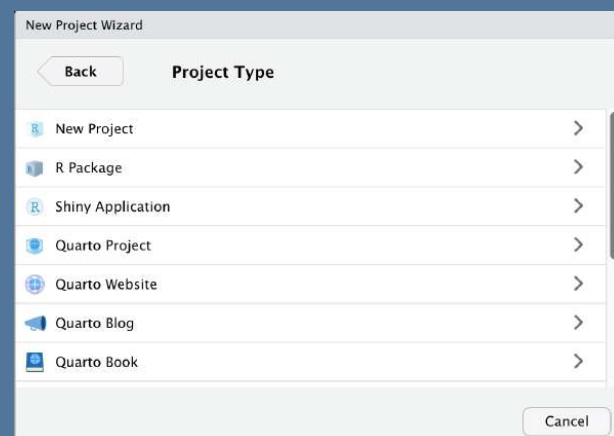


Go to the next steps:

- Create a new directory (Where we will store our project)
- Select “New Project”

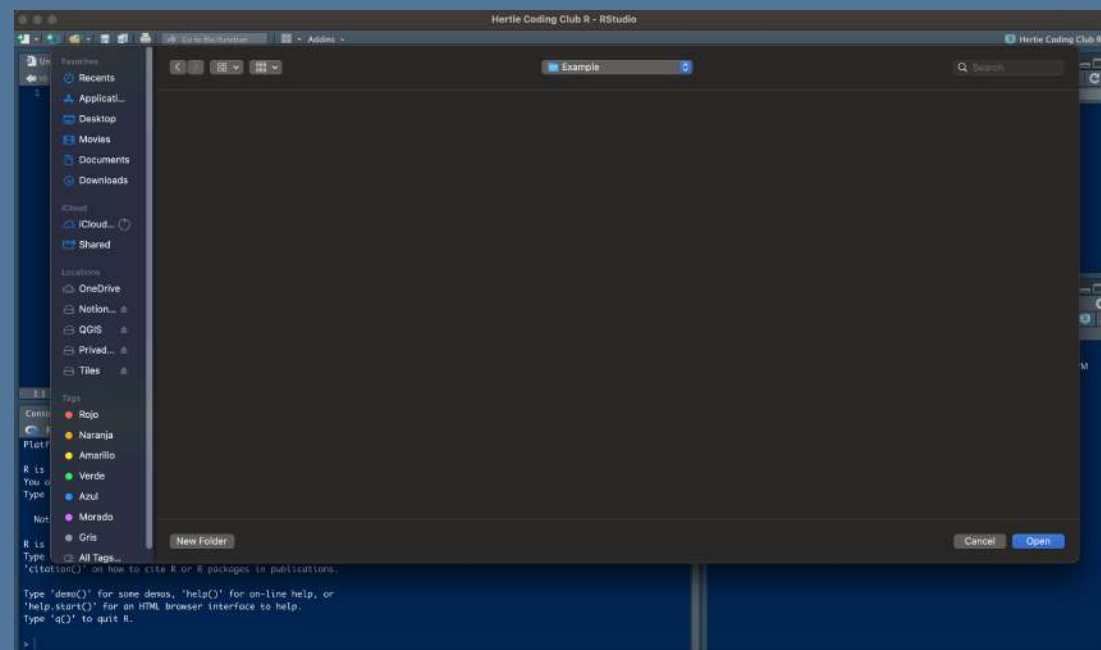


Create our first project (Step 2)

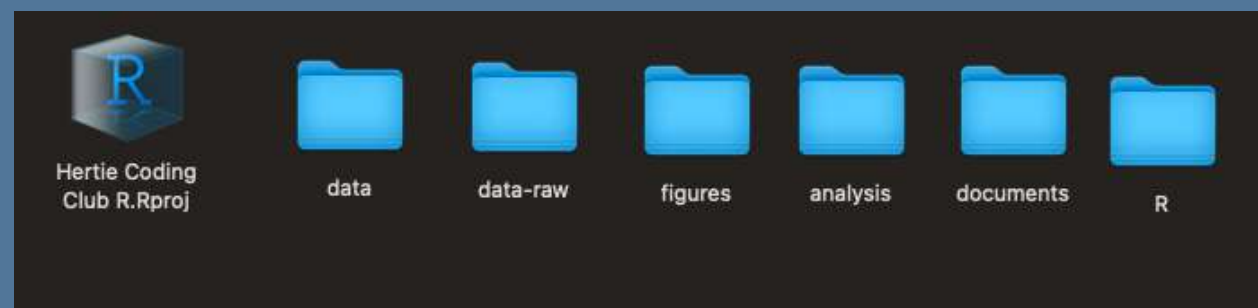


Go to the next steps:

- Select again New Project
- Choose the path where you want to create your project (It should be a folder)



Working directory (How we load and save things in R)



Go to the next steps:

- data/: for processed data
- data-raw/: for raw data and intermediate datasets (Internet, f.e.)
- figures/: for save your plots
- analysis/: for save your scripts
- documents/: for outlines, drafts, other text
- R/: for functions

Alarid et. al. (2019)

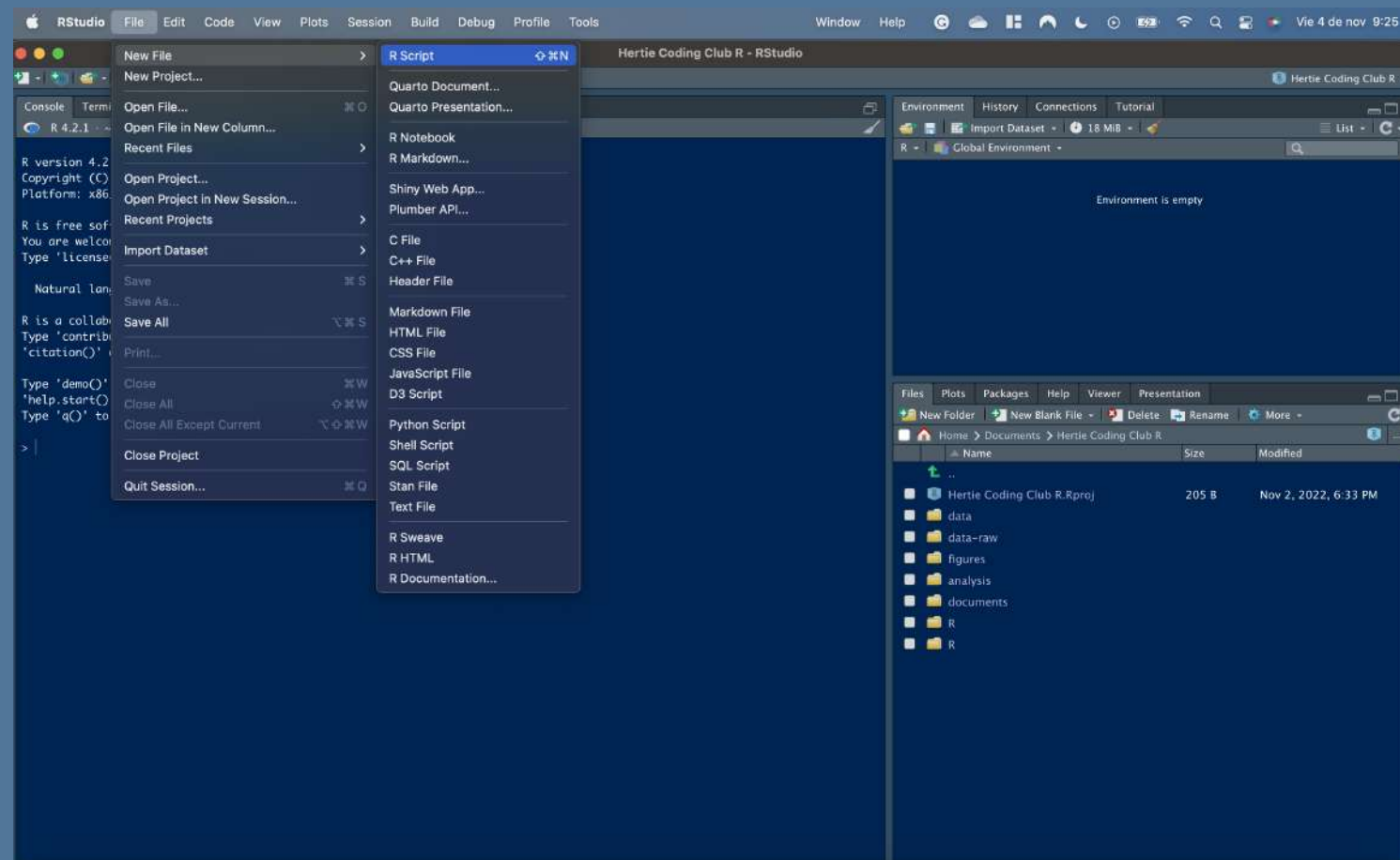
Folder name	Folder function
data-raw	This is where raw data is stored alongside ".R" scripts that read in raw data, process these data, and calls use_this::use_data(<processed data>) to save .rda formatted data files in the "data" folder. These data could include a ".csv" file with input parameters derived from the published literature, as well as internal R data files (with .RData, .rds, or .rda extensions) containing primary data from which model input values will be estimated through statistical models embedded into the analysis.
data	This is where input data is stored to be used in the different components of the CEA. These data could be generated from raw data stored in the "data-raw" folder. Essentially, this folder stores the cleaned or processed versions of raw data that has been gathered from elsewhere
R	This is where ".R" files that define functions to be used as part of the analysis are stored. These are functions that are specific to the analysis. The model will be one such function; however, other functions will likely be used, such as computing the fit of the model output to the specific calibration targets of the analysis. This folder also stores ".R" scripts that document the datasets in the "data" folder.
analysis	This is where interactive scripts of the analysis would be stored. These scripts control the overall flow of the analysis. This is also where many operations that ultimately become functions will be developed and debugged.
output	This is where output files of the analysis should be stored. These files may be internal R data files (.RData, .rds, .rda) or external data files (such as .csv). Examples of files stored here would be the output of the model calibration component or the PSA dataset generated in the uncertainty analysis component. These data files can then be loaded by other components without having to first rerun previous components (e.g. the calibrated model values can be loaded for a base case analysis without re-running the calibration).
figs	For analyses that will include figures, we generally create a separate figures folder. Though these could be stored in the output folder, it can be helpful to have a separate folder so that the images of the figure files can be easily previewed. This is particularly important for analyses that generate a large number of figures.
tables	This folder includes tables to be included in a publication or report, such as the table of intervention costs and effects and ICERs.
report	A report folder could be used to store R Markdown files to describe in detail the model-based CEA by using all the functions and data of the framework, run analyses and display figures. The R Markdown files can be compiled into .html, .doc or .pdf files to generate a report of the CEA. This report could be the document submitted to HTA agencies accompanying the R code of the model-based CEA.
vignettes	A vignettes folder could be used to describe the usage of the functions and data of each of some or all components of the framework through accompanying R Markdown files as documentation. The R Markdown file can use all the functions, outputs, and figures to integrate the R code into the Markdown text.
tests	A tests folder includes ".R" scripts that runs all the unit tests of the functions in the framework. A good practice is to have one file of tests for each complicated function or for each of the components of the framework.

YOU ARE READY



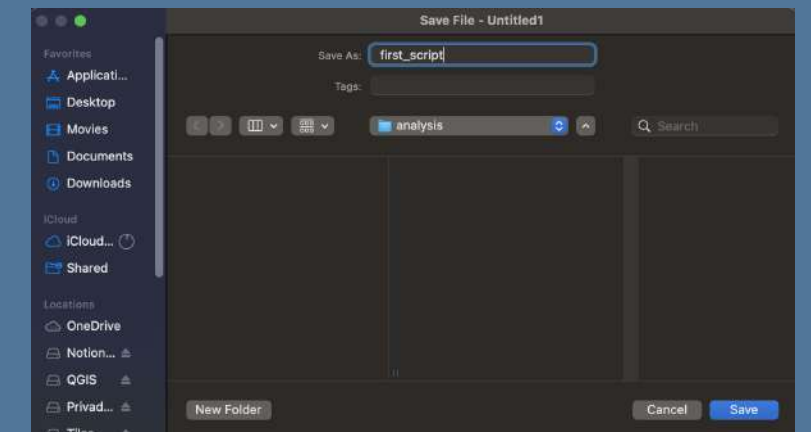
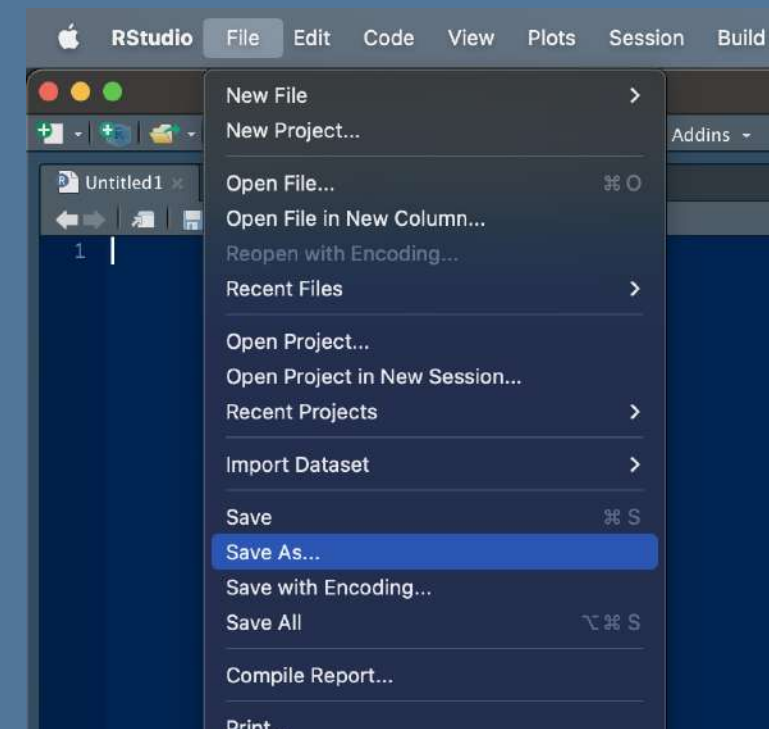
RStudio

Create a script (This is the place where we write code)



Go to the next steps:

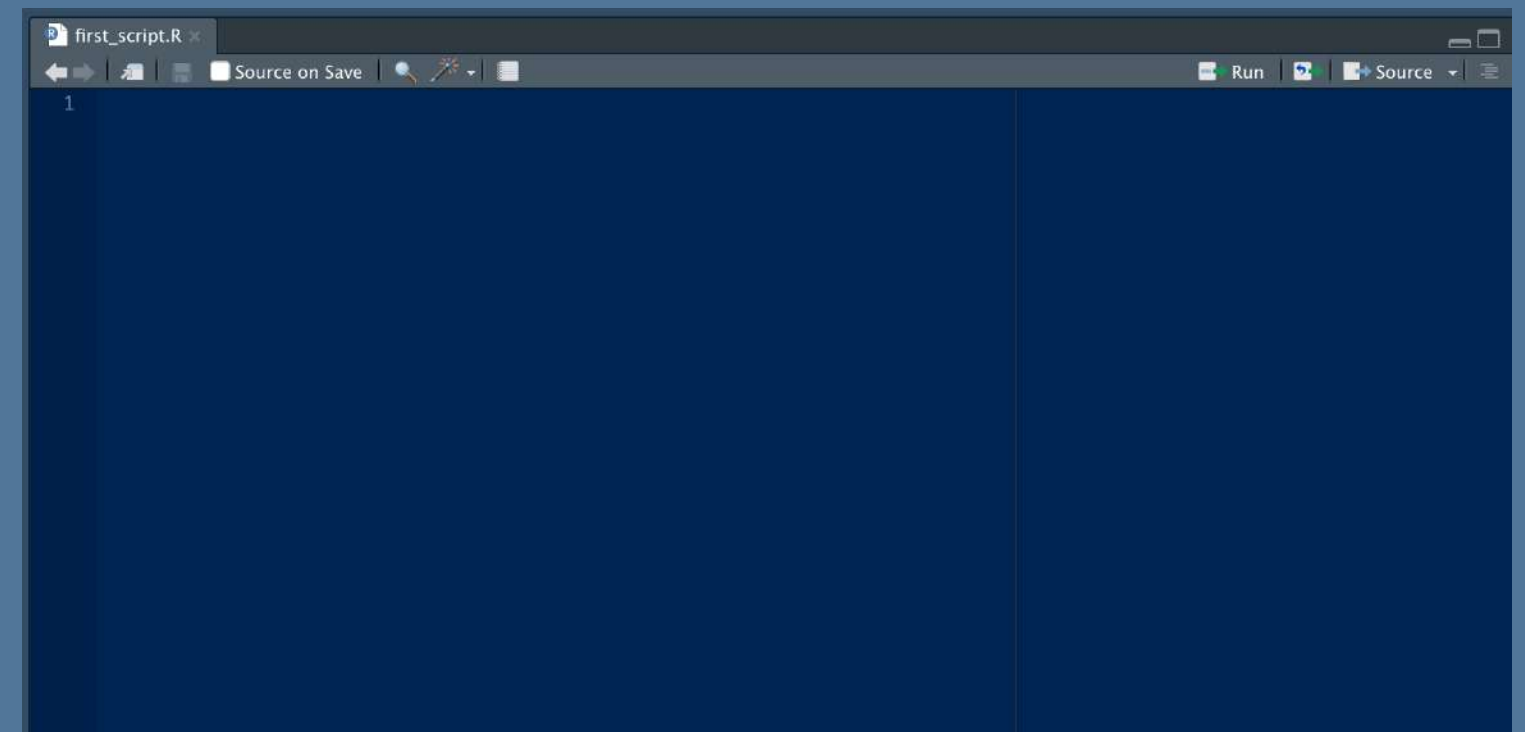
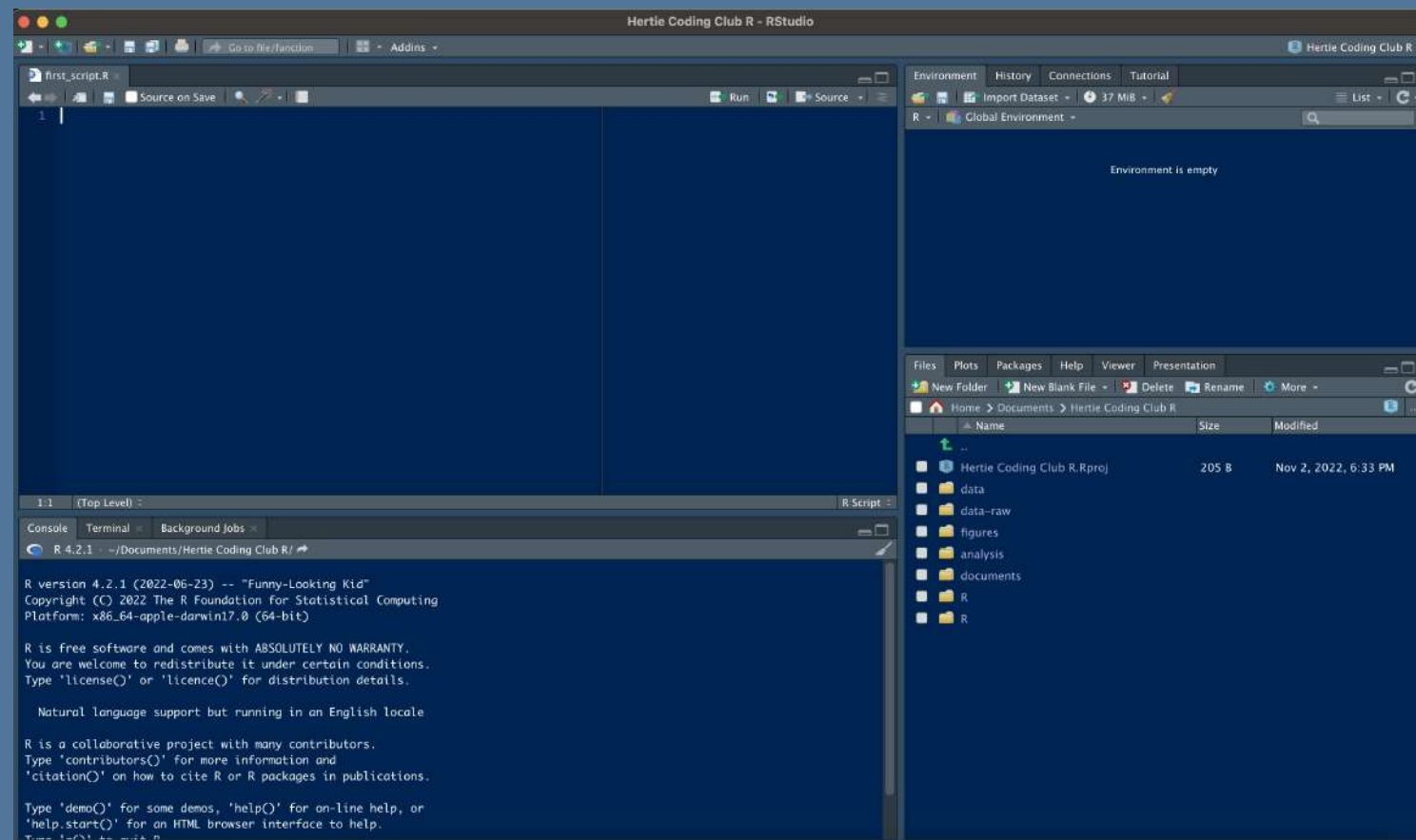
- In the file section, create an R script
- Save inside your analysis folder



Source

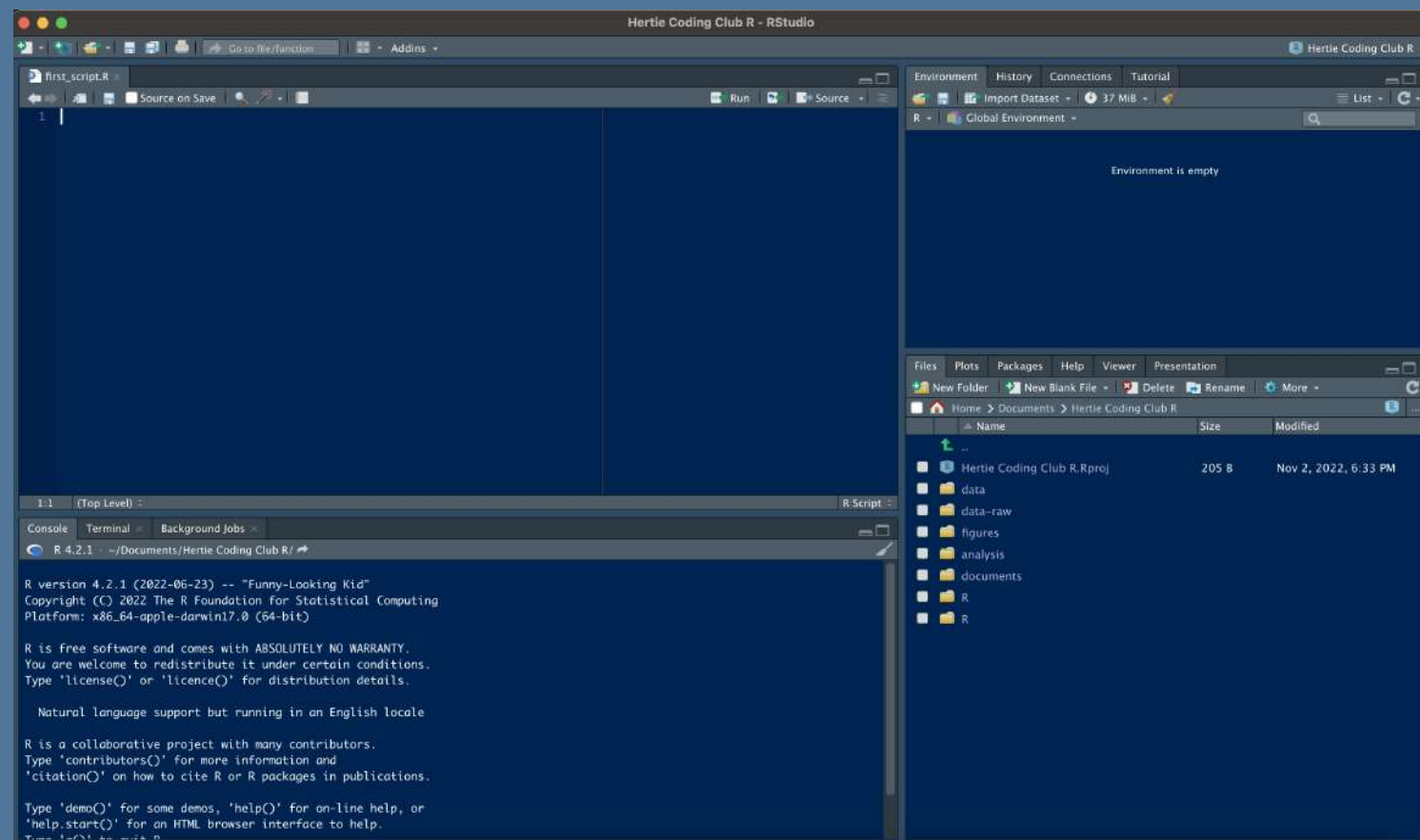
The source pane is where you create and edit R Scripts

- “Command + Return” on Mac, or “Control + Enter” on PC to send all highlighted code to the console.
- Save it inside your analysis folder



Console

The console is the heart of R.



- Here R actually evaluates your code
- Try to write most of your code in a document in the Source. Only type directly into the Console to de-bug or do quick analyses.
- When ready: > and If waiting: +
- Cancel commands by pressing Esc.

```

Console Terminal Background Jobs
R 4.2.1 - ~/Documents/Hertie Coding Club R/
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 2+2

```

```

Console Terminal Background Jobs
R 4.2.1 - ~/Documents/Hertie Coding Club R/
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

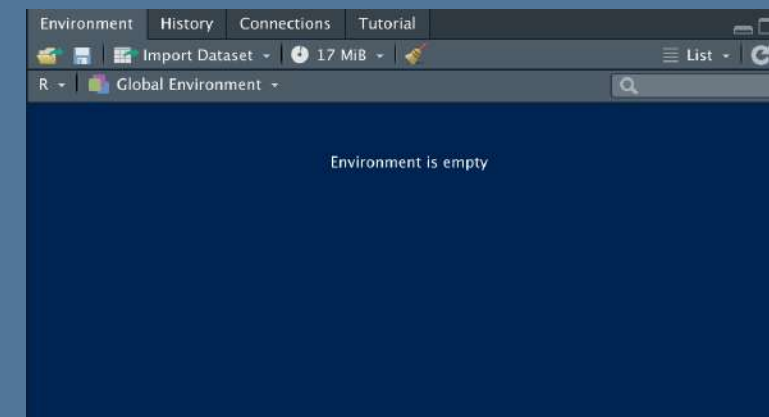
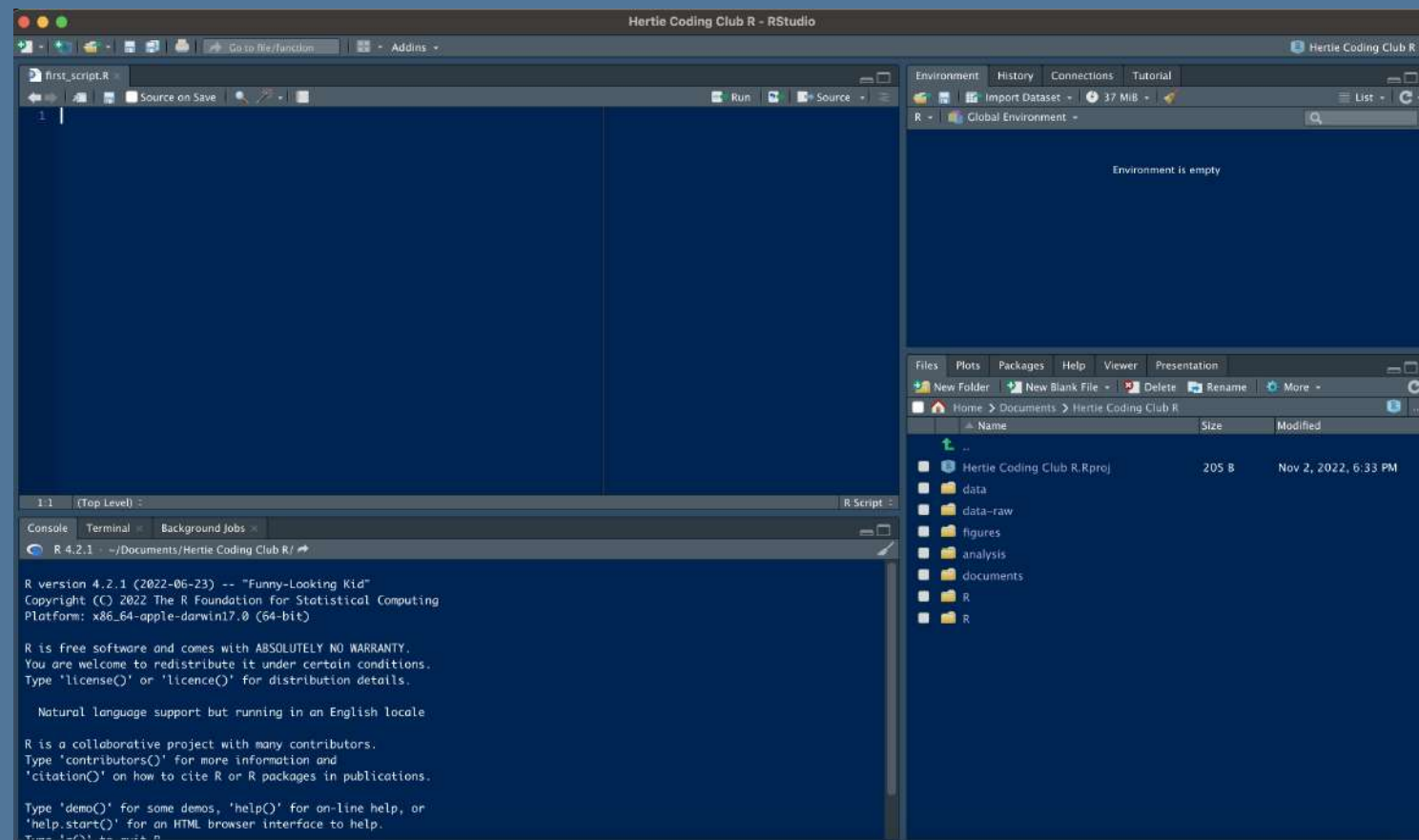
> 2+2
[1] 4
> |

```

Environment/History

The Environment tab of this panel shows you the names of all the data objects

- You can also see information like the number of observations and rows in data objects

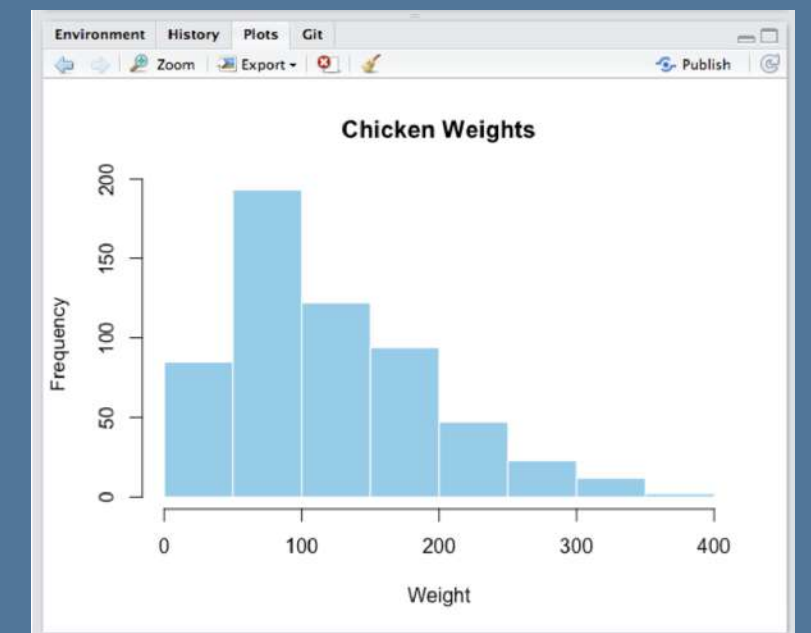
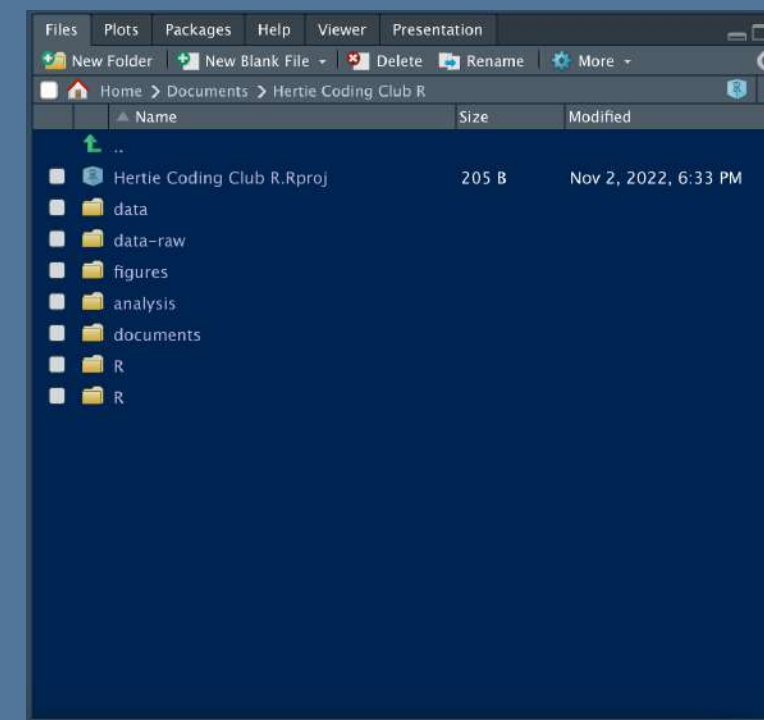
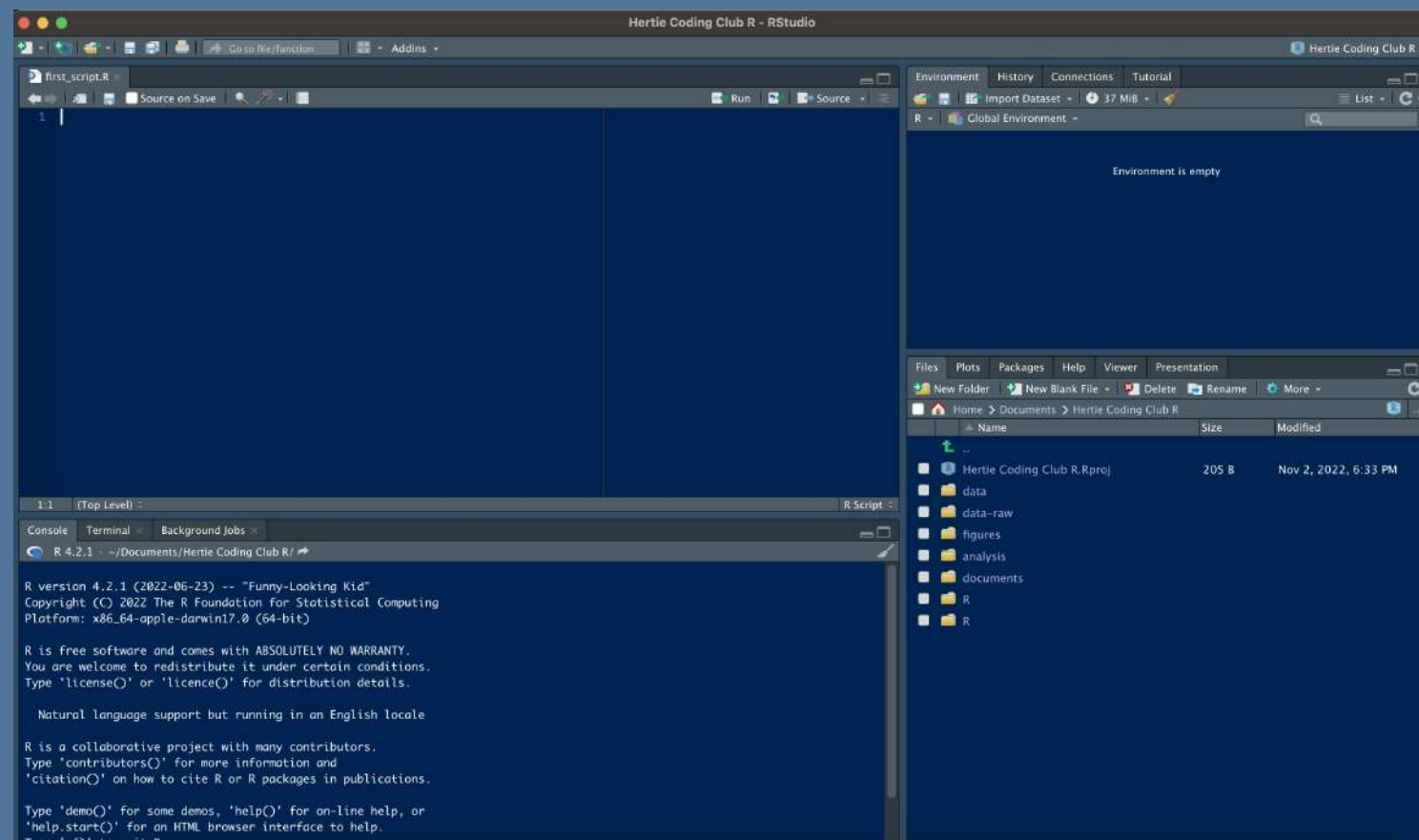


Environment	History	Connections
<div> <div>Import Dataset</div> <div>17 MiB</div> <div>List</div> </div>		
Global Environment		
RAWOccWye	List of 9	
RelevantTrusts	9 obs. of 8 variables	
RelevantTrusts2	2 obs. of 8 variables	
SPECIALITY2014_Q1	228 obs. of 81 variables	
SPECIALITY2014_Q2	228 obs. of 81 variables	
SPECIALITY2014_Q3	225 obs. of 81 variables	
SPECIALITY2014_Q4	224 obs. of 81 variables	
SPECIALITY2015_Q1	219 obs. of 81 variables	
SPECIALITY2015_Q2	218 obs. of 81 variables	
SPECIALITY2015_Q3	218 obs. of 81 variables	
SPECIALITY2015_Q4	217 obs. of 81 variables	

Files and more

The Files / Plots / Packages / Help panel shows you lots of helpful information

- Files - The files panel gives you access to the file directory on your hard drive.
- Plots - The Plots panel shows all your plots.



Packages

Packages

Why R?

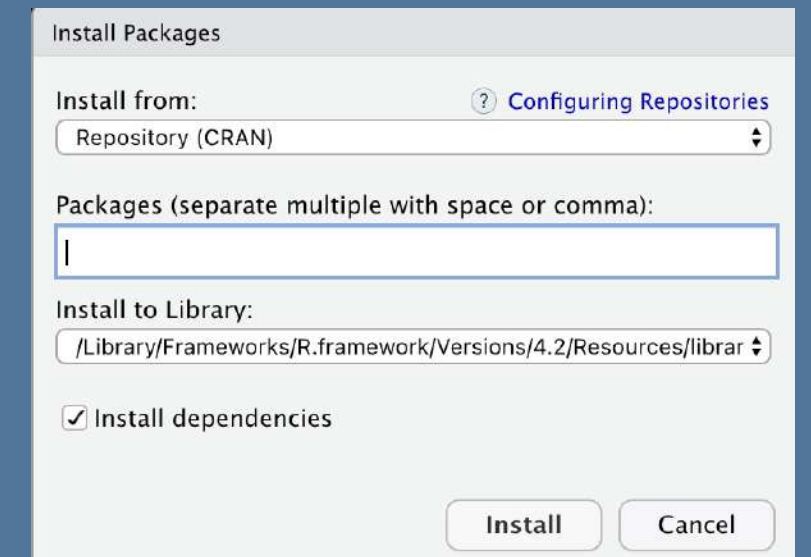
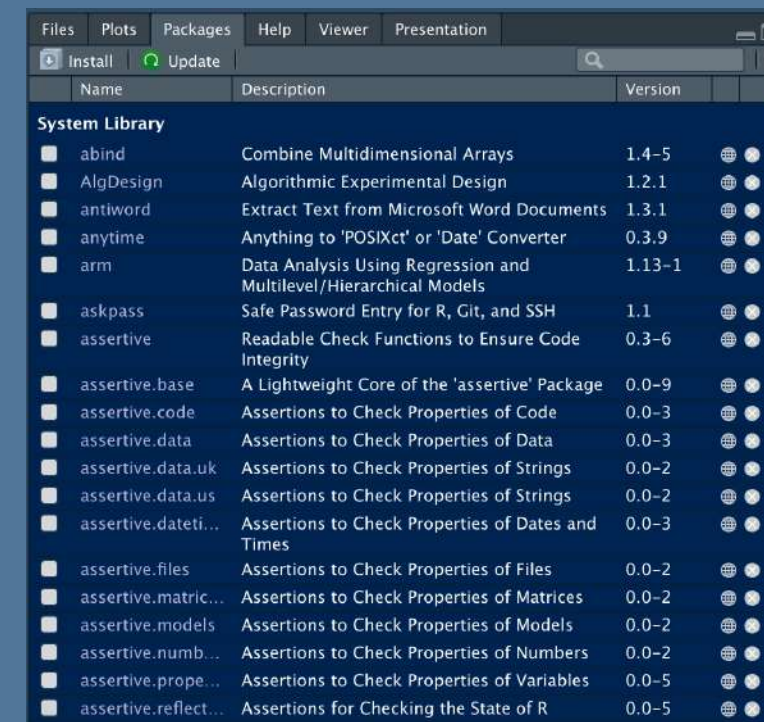
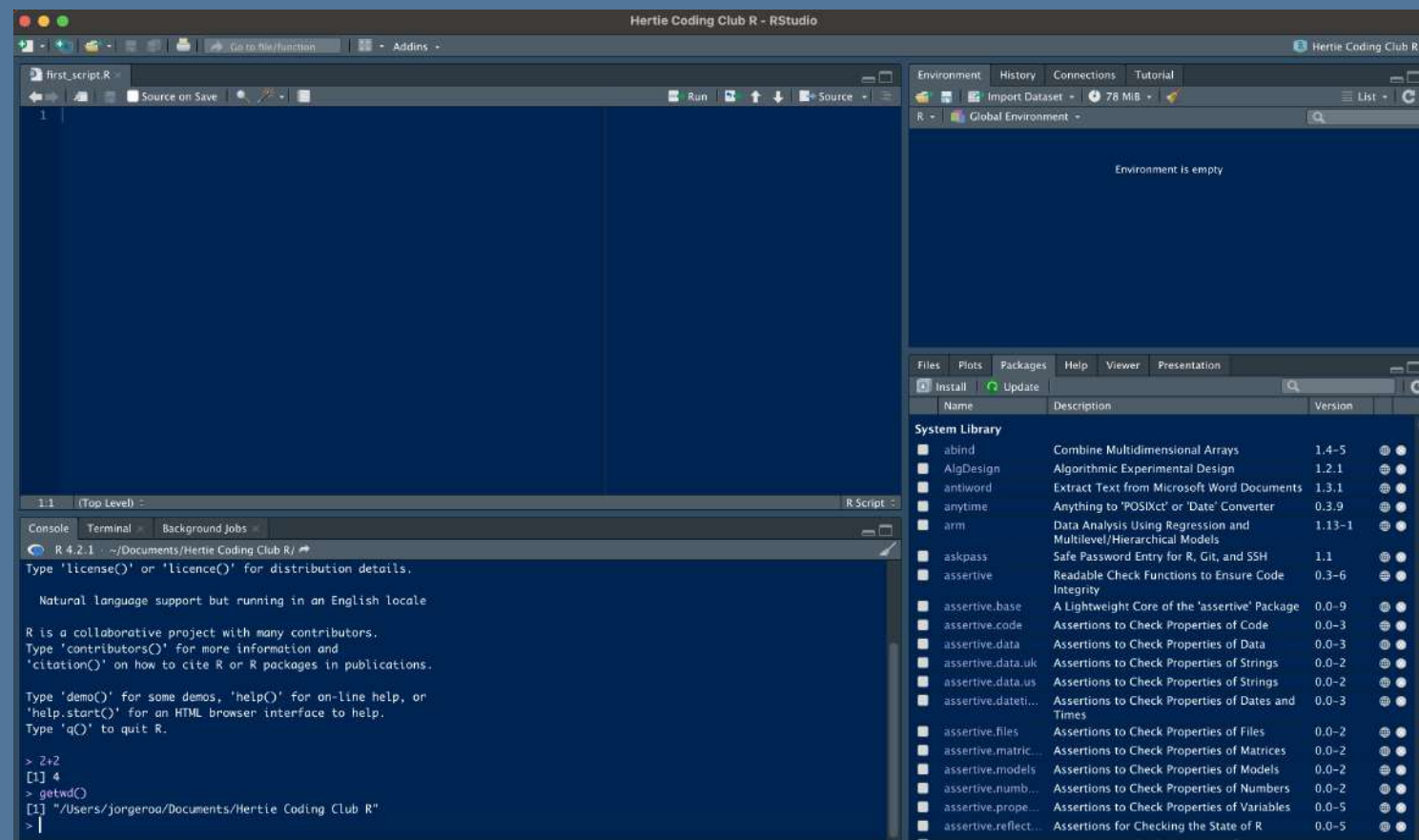
- Packages in R Programming language are a set of R functions, compiled code, and sample data. These are stored under a directory called “library” within the R environment.
- R packages provide a simple way to distribute R code and documentation.
- Packages are the fundamental units of reproducible R code. They include reusable R functions, the documentation that describes how to use them, and sample data
- More than 18,000 packages
- Extends the capacity of R and allows us to do much more



How we install packages

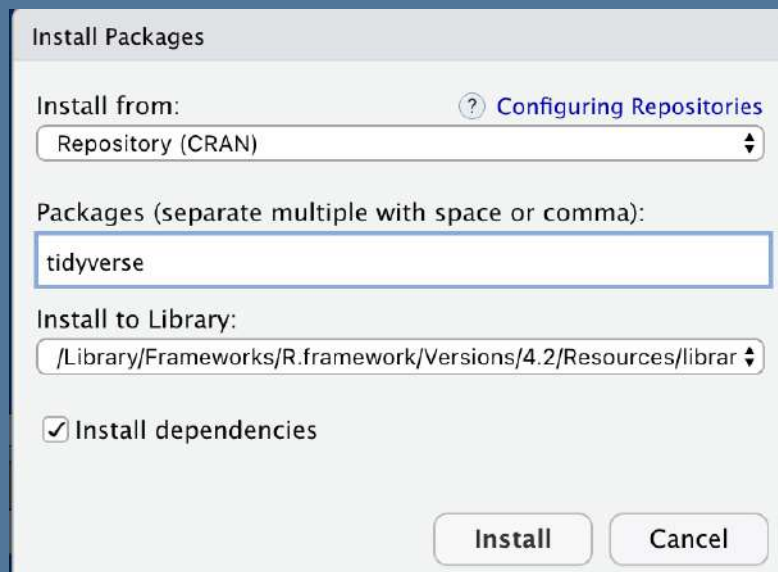
The easy way

- Go to the “Packages” tab
- Press the “Install” button



How we install packages

The easy way



Install Packages

Install from: [? Configuring Repositories](#)

Repository (CRAN)

Packages (separate multiple with space or comma):

tidyverse

Install to Library:

/Library/Frameworks/R.framework/Versions/4.2/Resources/librar

☒ Install dependencies

Install Cancel

- Other way is type in the console
- `install.packages("tidyverse")`

```
> install.packages("tidyverse")
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/tidyverse_1.3.2.tgz'
Content type 'application/x-gzip' length 420896 bytes (411 KB)
=====
downloaded 411 KB

The downloaded binary packages are in
  /var/folders/q3/ztffc4r10tv5h4pftjr2qw0r0000gn/T//RtmpKwSpEV/downloaded_packages
>
```

Objects

Objects

Why objects?

- R works with objects (vectors, dataframes, lists, numbers, etc)
- To do useful and interesting things in R, we need to assign values to objects.
- To create an object, give it a name followed by the assignment operator, followed by the value.
- Assignment operator `<-`
- Can also use `=` but not recommended
- Shortcut: “Alt + -” on PC, “Option + -” on Mac

```
1 x <- 2 + 2
2
3 x
```

```
[1] 4
```



Type of objects

There are 5 basic types of objects in the R language:

- **Atomic vectors** are one of the basic types of objects in R programming. Atomic vectors can store homogeneous data types such as character, doubles, integers, raw, logical, and complex.
- **List** is another type of object in R programming. List can contain heterogeneous data types such as vectors or another lists.

```
1 #Numeric vector
2 numbers <- c(1, 2, 3, 4)
3
4 #String vector
5 characters <- c("a", "b", "c", "d")
6
7 #Numeric value
8 value <- 5
9
10 #List
11 my_list <- list(c(1, 2, 3, 4), list("a", "b", "c"))
```

```
1 print(numbers)
[1] 1 2 3 4

1 print(characters)
[1] "a" "b" "c" "d"

1 print(value)
[1] 5

1 print(my_list)

[[1]]
[1] 1 2 3 4

[[2]]
[[2]][[1]]
[1] "a"

[[2]][[2]]
[1] "b"

[[2]][[3]]
[1] "c"
```


More type of objects

- **Matrices:** To store values as 2-Dimensional array, matrices are used in R. Data, number of rows and columns are defined in the `matrix()` function.
- **Factors:** Factor object encodes a vector of unique elements (levels) from the given data vector.
- **Arrays:** `array()` function is used to create n-dimensional array. This function takes `dim` attribute as an argument and creates required length of each dimension as specified in the attribute.

```
1 x <- c(1, 2, 3, 4, 5, 6)
2
3 # Matrix
4 mat <- matrix(x, nrow = 2)
5
6 # Matrix
7 s <- c("spring", "autumn", "winter", "summer",
8 "spring", "autumn")
9
10 arr <- array(c(1, 2), dim = c(3, 3))
```

```
1 print(mat)
```

	[,1]	[,2]	[,3]
[1,]	1	3	5
[2,]	2	4	6

```
1 print(factor(s))
```

```
[1] spring autumn winter summer spring autumn
Levels: autumn spring summer winter
```

```
1 print(nlevels(factor(s)))
```

```
[1] 4
```

```
1 print(arr)
```

	[,1]	[,2]	[,3]
[1,]	1	2	1
[2,]	2	1	2
[3,]	1	2	1

Finally: dataframes

- Data frames are 2-dimensional tabular data object in R programming.
- Data frames consists of multiple columns and each column represents a vector.
- Columns in data frame can have different modes of data unlike matrices.

```
1 # Create vectors
2 who <- c("Mom", "Sister", "Myself", "Dad", "Brother", "Brother", "Our dog (:)")
3 age <- c(58, 17, 25, 60, 29, 27, 5)
4 names <- c("Carmen", "Fernanda", "Jorge", "Arturo", "Ale", "Eduardo", "Rocky")
5
6 # Create data frame of vectors
7 df_my_family <- data.frame(who, age, names)
```

```
1 print(df_my_family)
```

	who	age	names
1	Mom	58	Carmen
2	Sister	17	Fernanda
3	Myself	25	Jorge
4	Dad	60	Arturo
5	Brother	29	Ale
6	Brother	27	Eduardo
7	Our dog (:)	5	Rocky

Now it's your turn

Exercise

- Create a dataframe of your family with multiple vectors.
- Include age, position (mom, dad, etc), their names and their birthdays.
- Hint:

```
1 # Create vectors
2 who <- c()
3 age <- c()
4 names <- c()
5 birthday <- "?"
6
7 # Create data frame of vectors
8 df_my_family <- data.frame("Here you put your vectors")
```

REMEMBER: They must have the same length.



Summary

- Customize our RStudio appearance
- Create our first project
- RStudio elements
- Working directory
- Packages
- Objects
- Summary
- Next session



Next session

- Base R
 - Load and save objects
 - Pipelines
 - Tidyverse
 - And more...
- Same channel (Forum)
 - Same day and hour Friday, November 11 (14:00 hrs - 16:00 hrs)
 - Join Slack if you haven't.
 - Use Slack for questions.
 - Follow us in our social media:
 - 📷 Instagram: hertiecodingclub
 - 🐦 Twitter: @HertieCodingCLB

Thanks for your time and welcome



Remember that everybody can learn how to code!!

1200 lines of code were created for this presentation.