Data Science for Everyone – Intro to Coding with R

Dr. Ab Mosca (they/them)

Slides based off slides courtesy of Jordan Crouser (https://jcrouser.github.io/)

Plan for Today

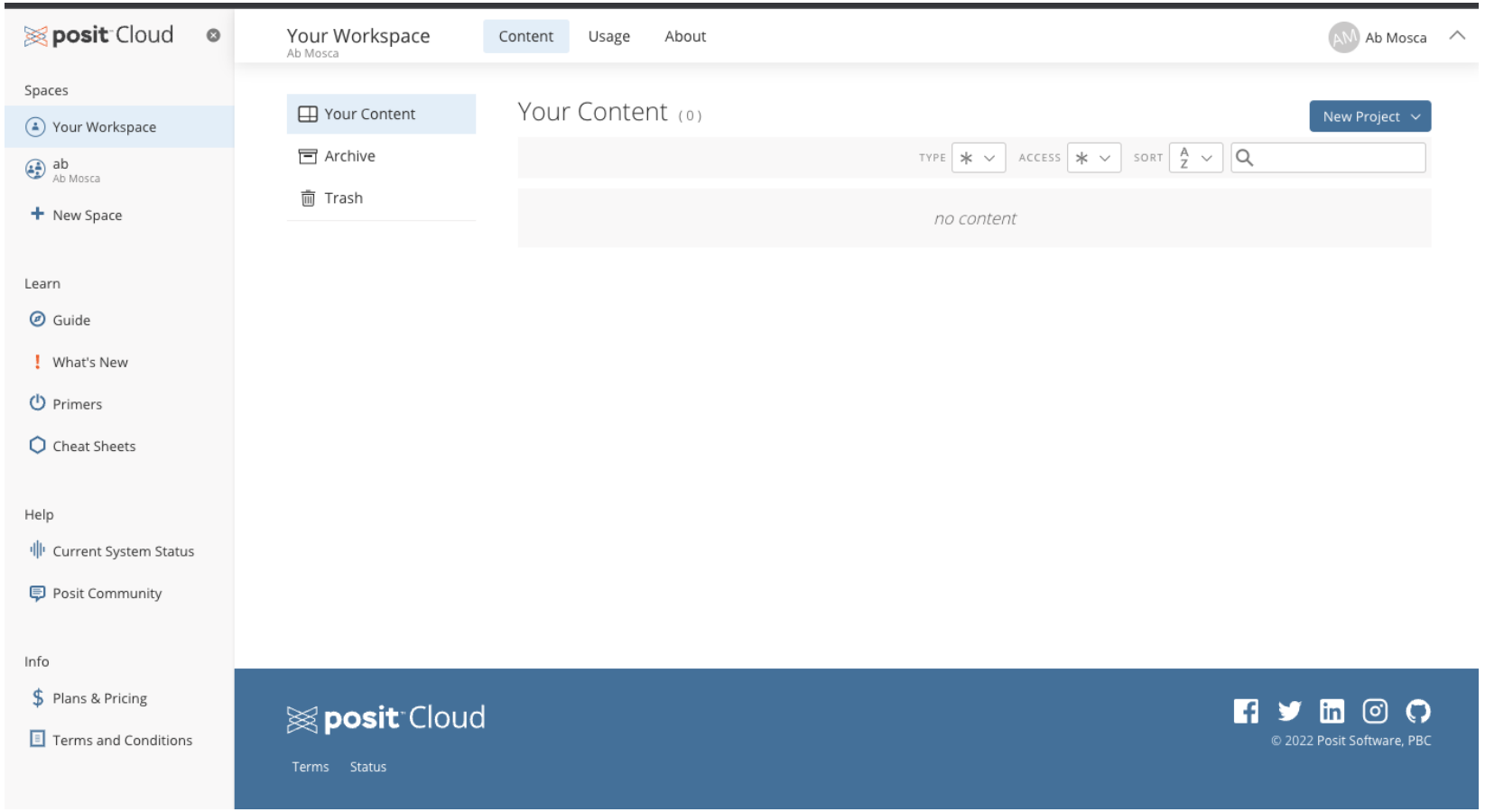
- Practice coding with the R programming language

R
R Studio
R Markdown

- R is a programming language
- R Studio is an interface for coding
- R Markdown is a type of file that allows you to intersperse writing, code, and output

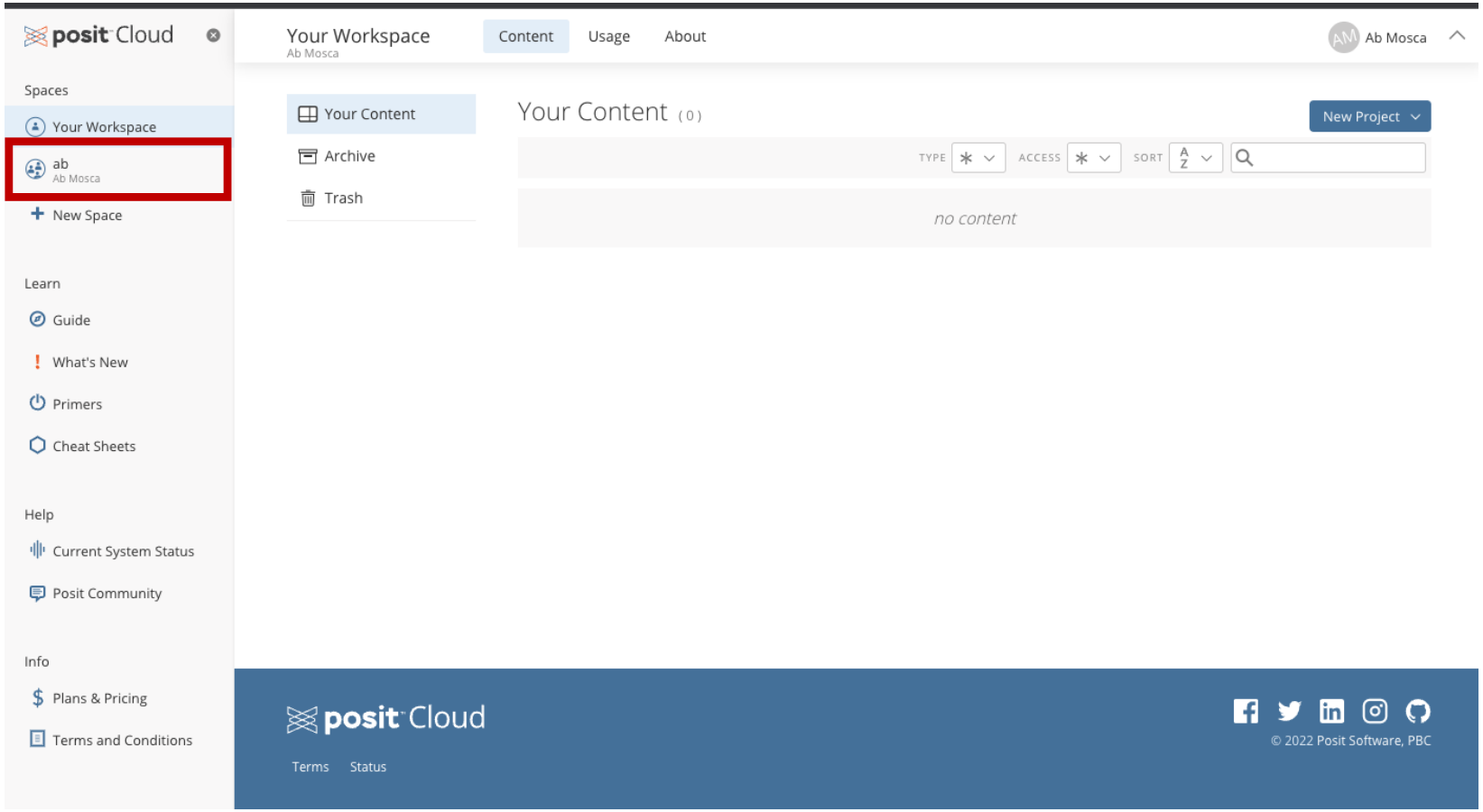
R Studio

- Launch R Studio:
- https://posit.cloud/



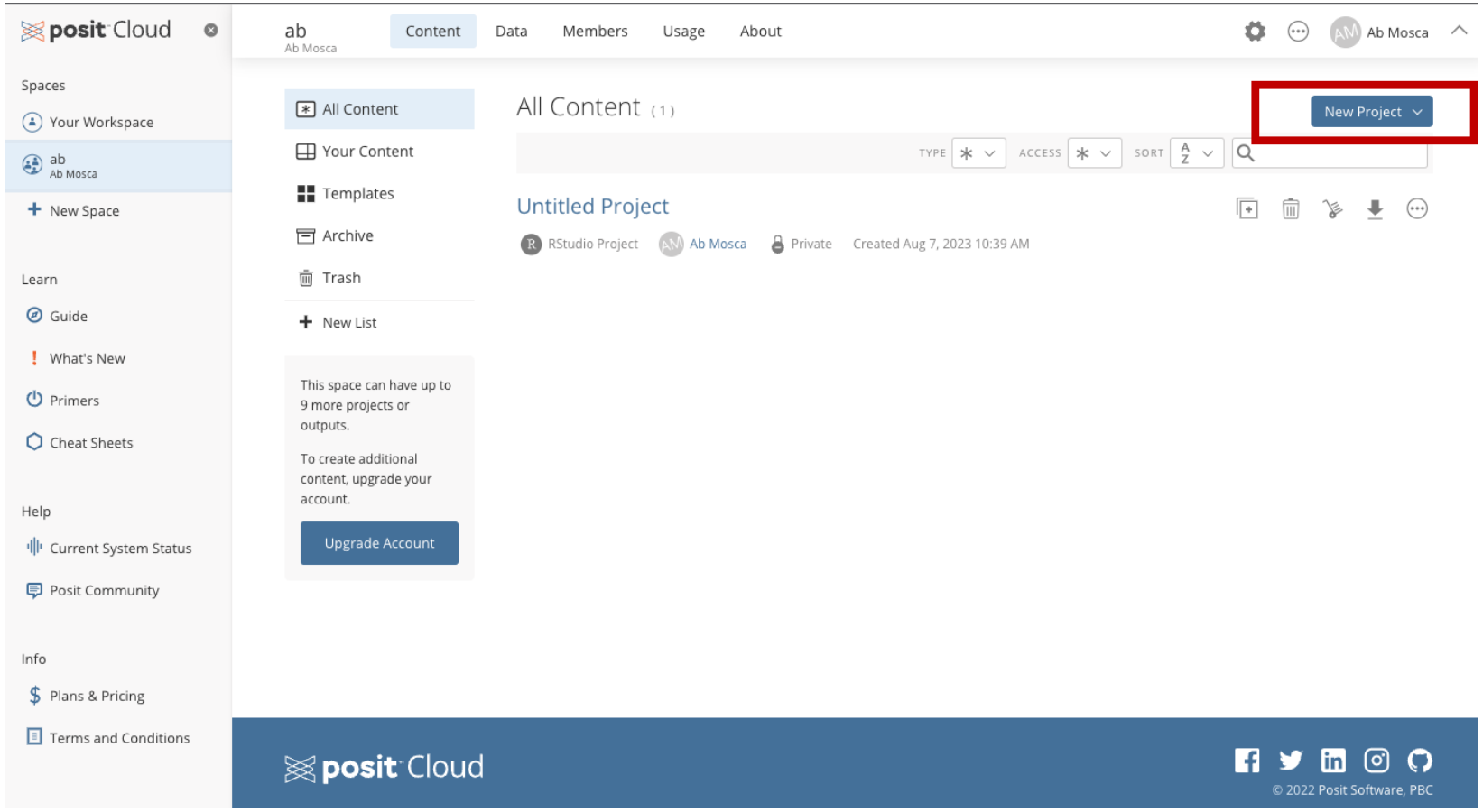
R Studio

- Move to your personal space



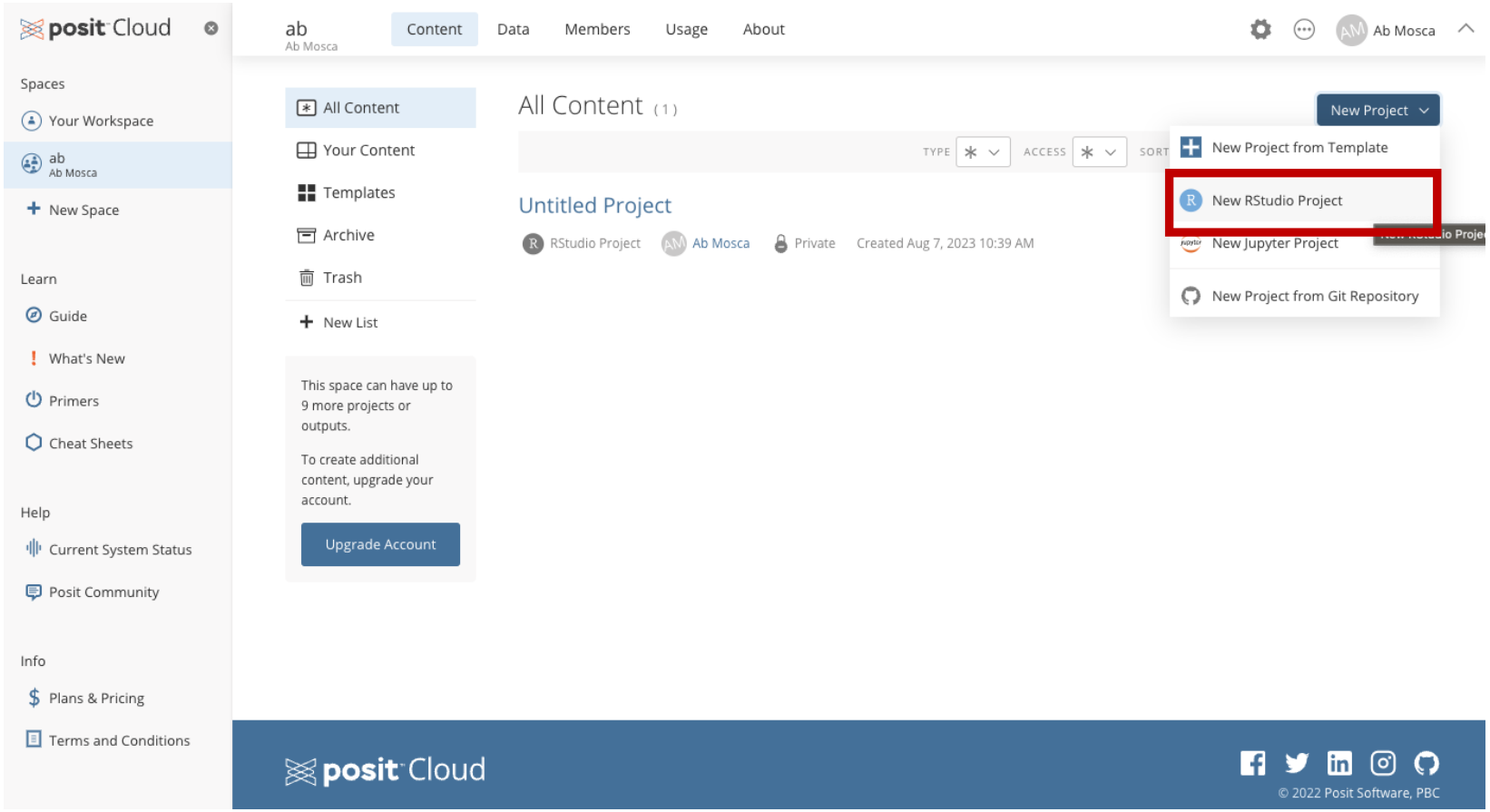
R Studio

- Start a new project



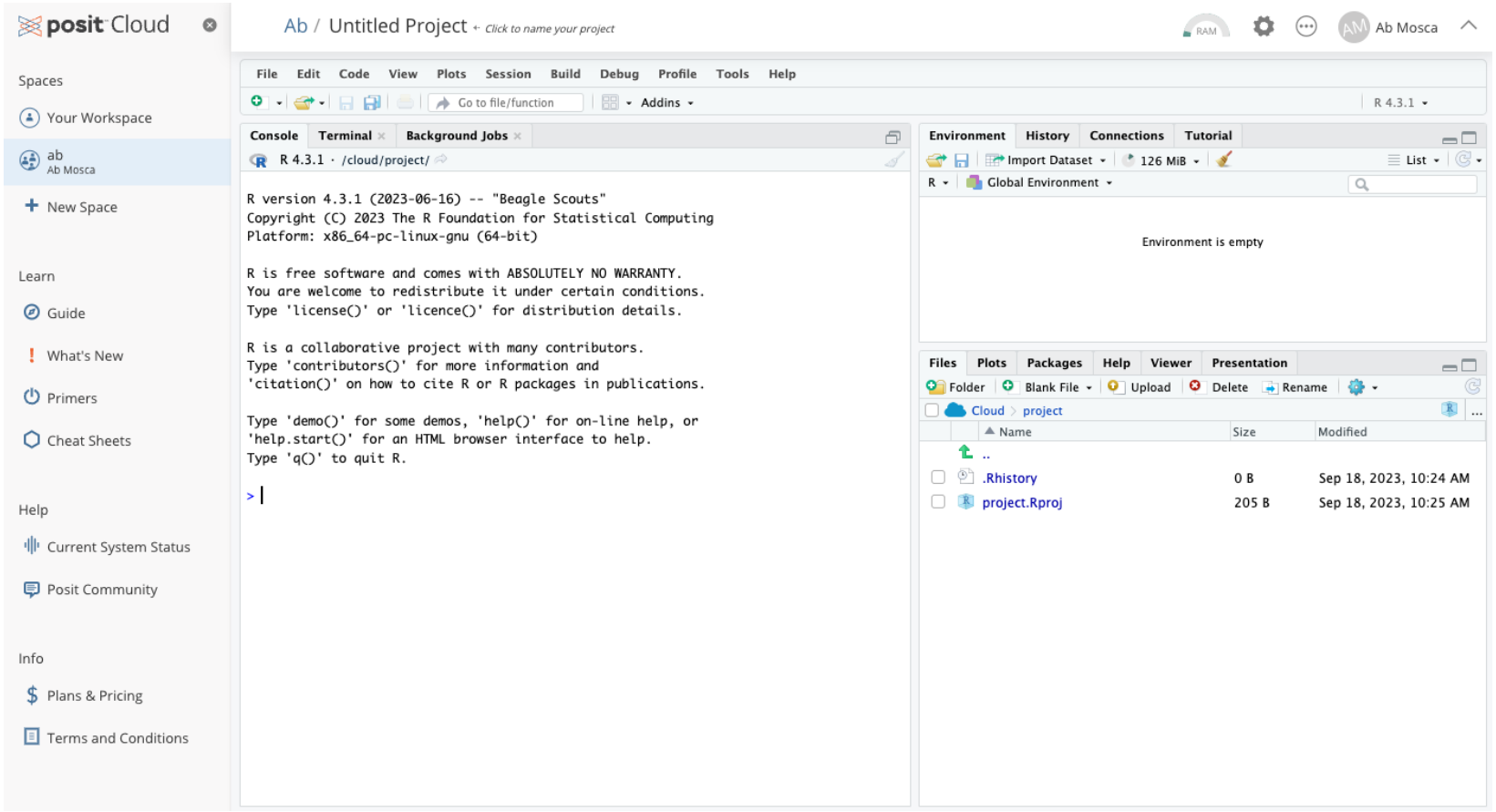
R Studio

- Start a new project



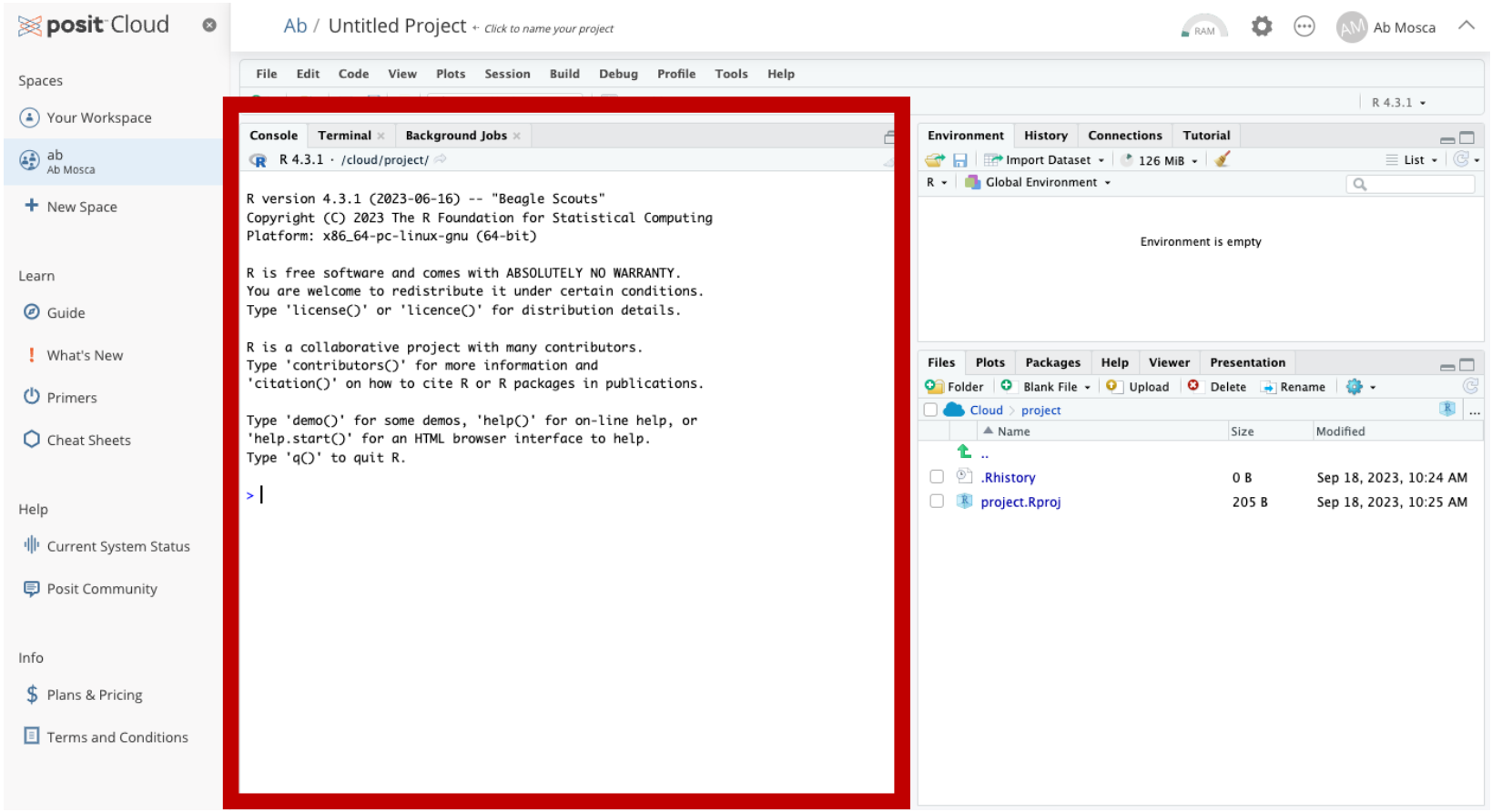
R Studio

- New project



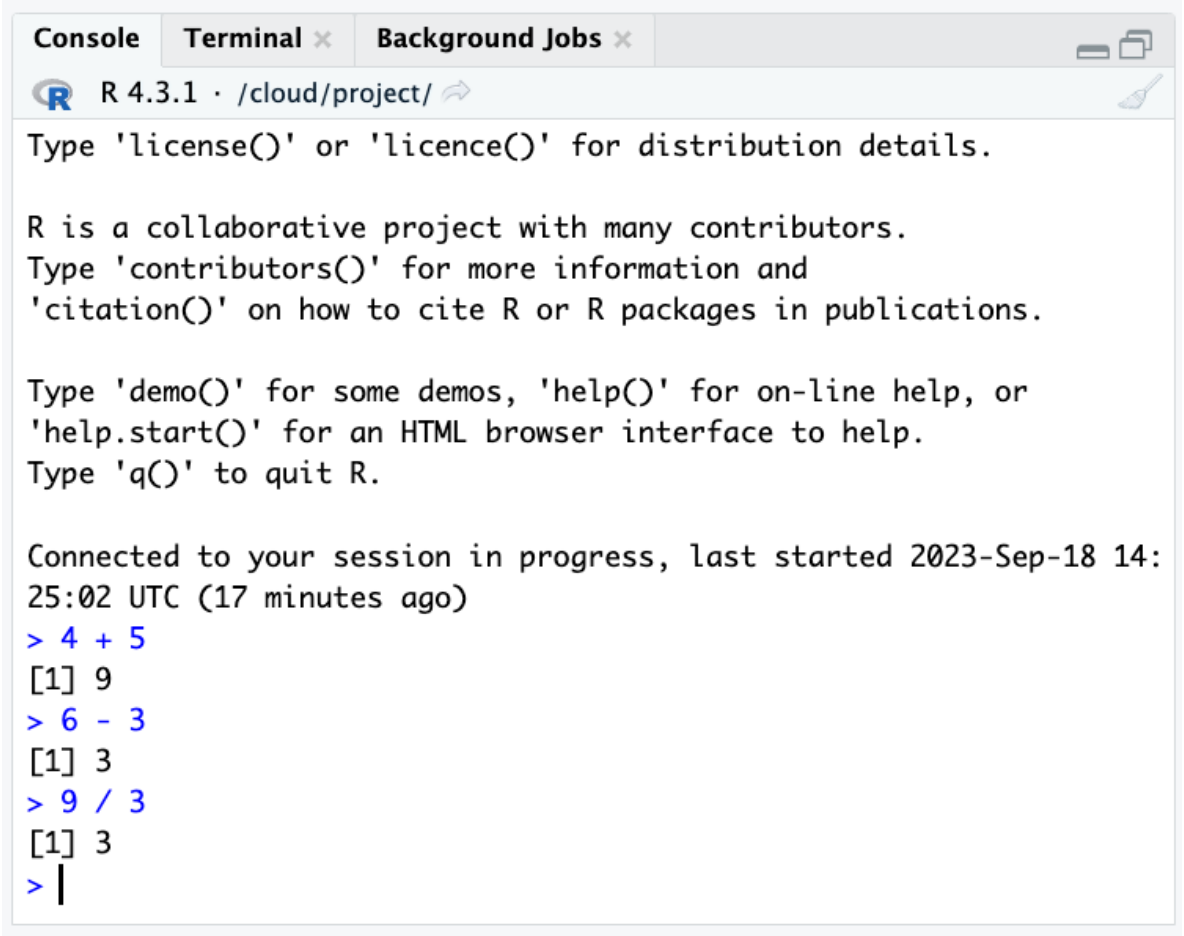
R Studio

- Console



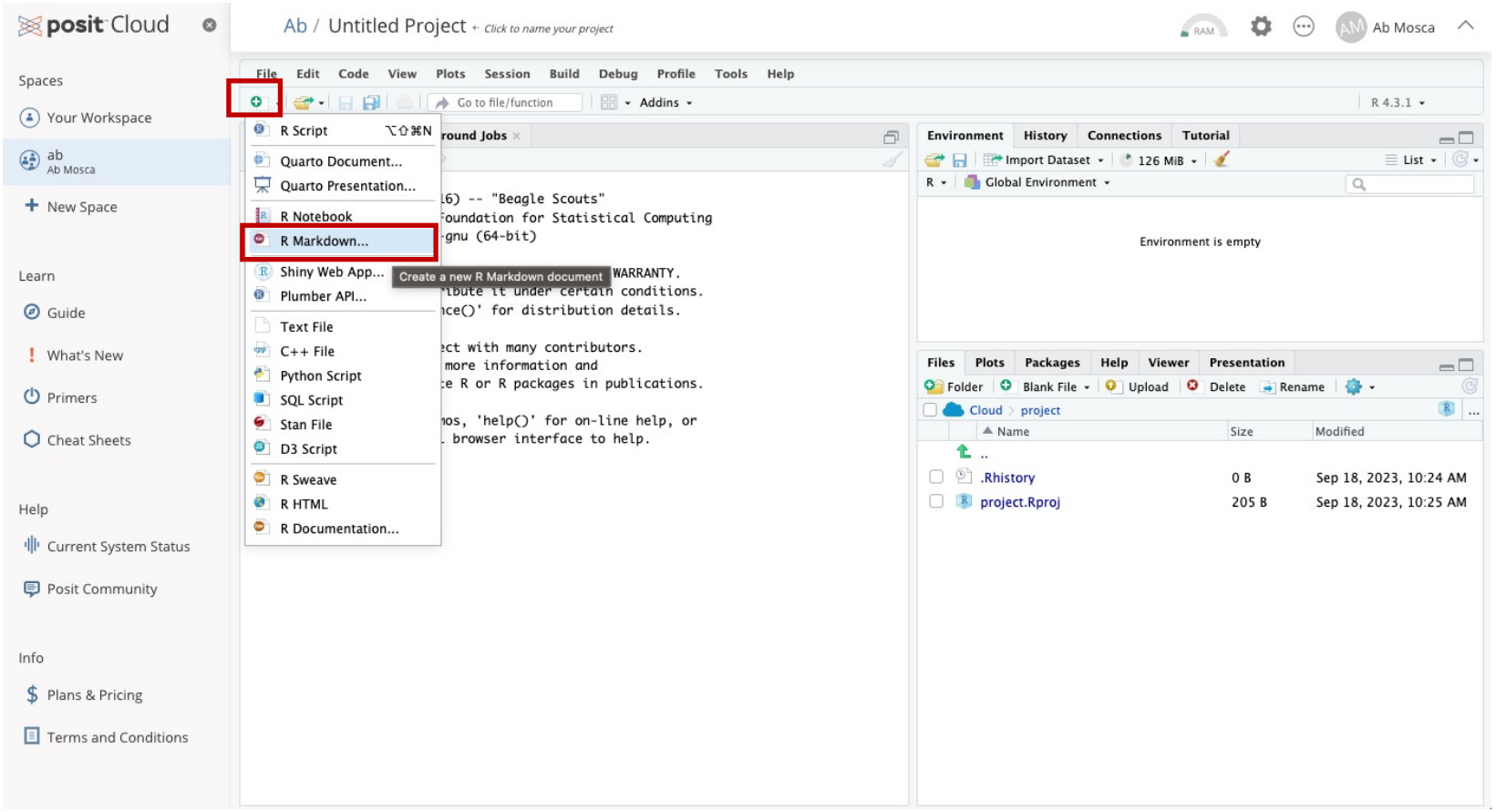
R Studio

- Console



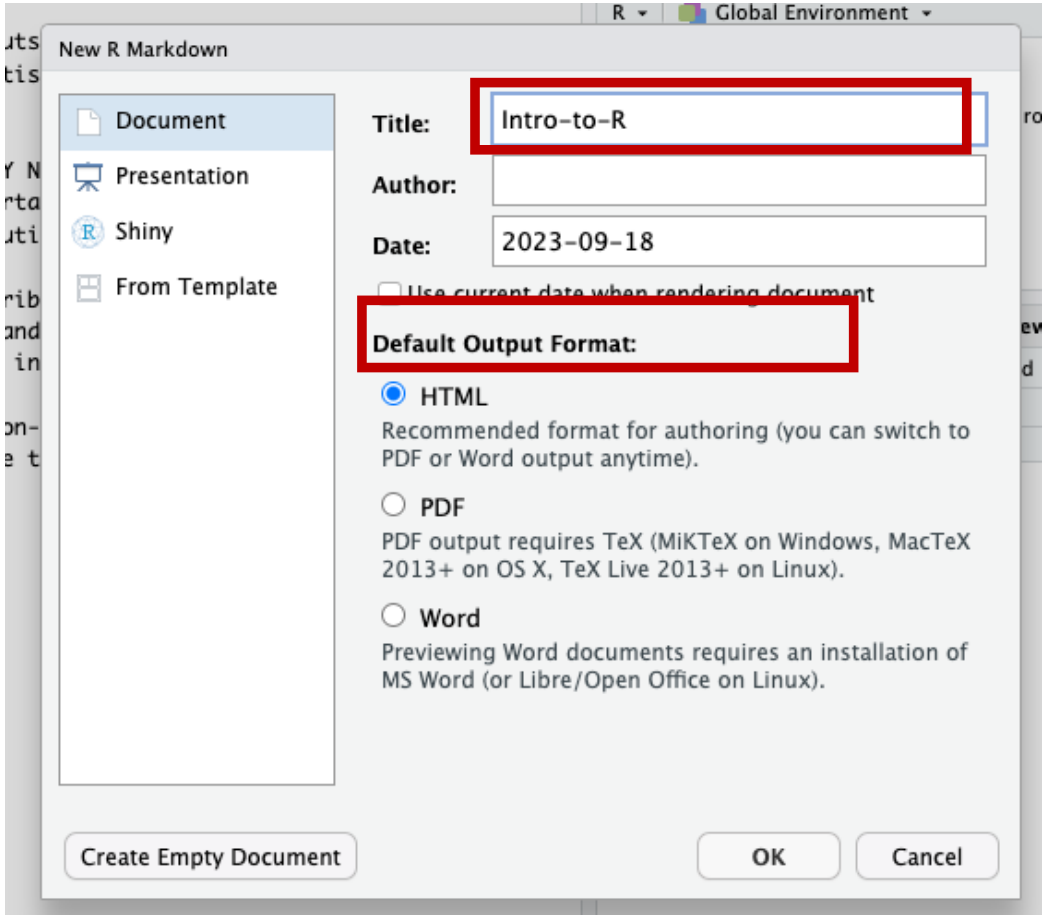
R Studio

- New file



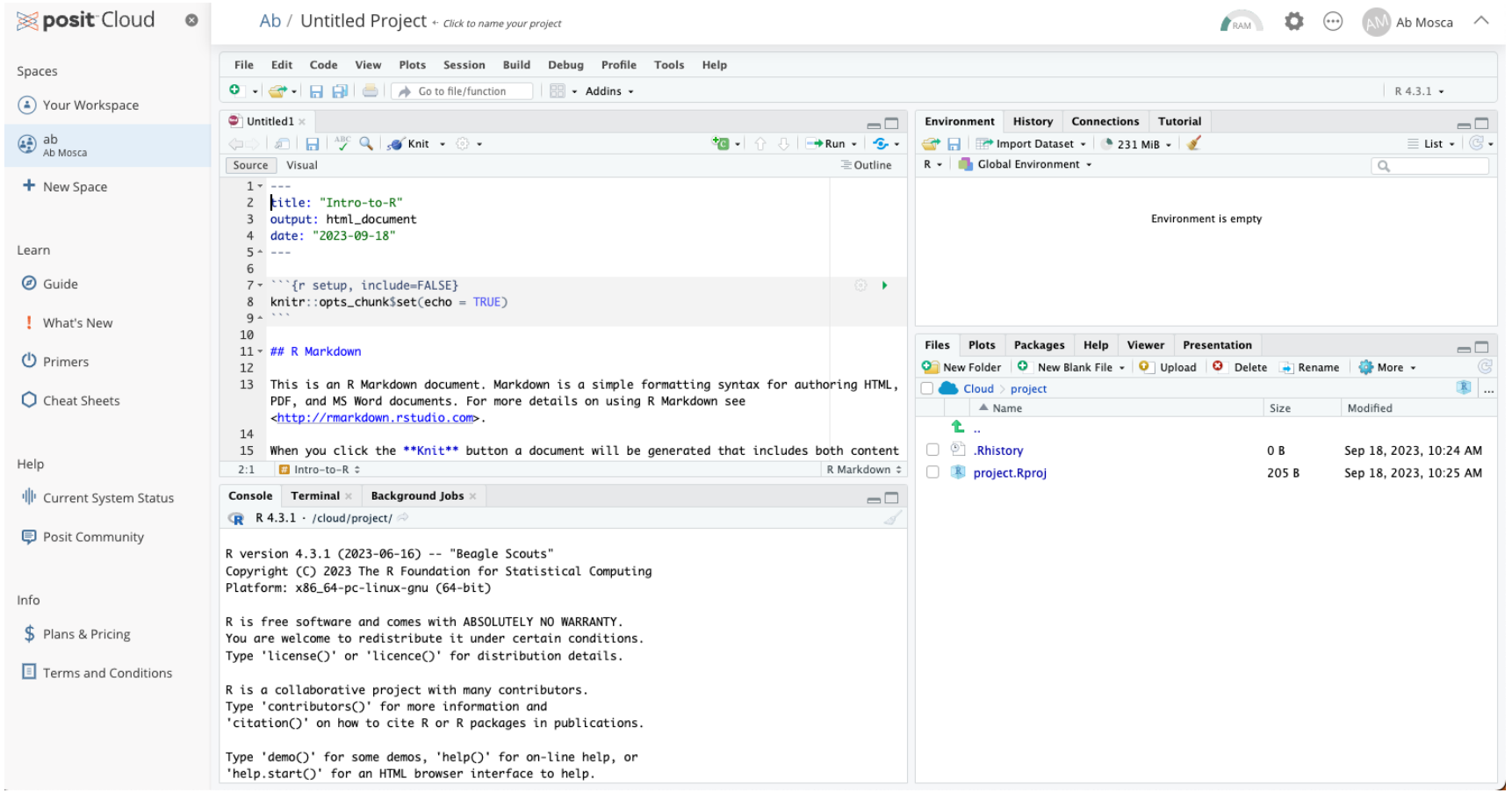
R Studio

- New file



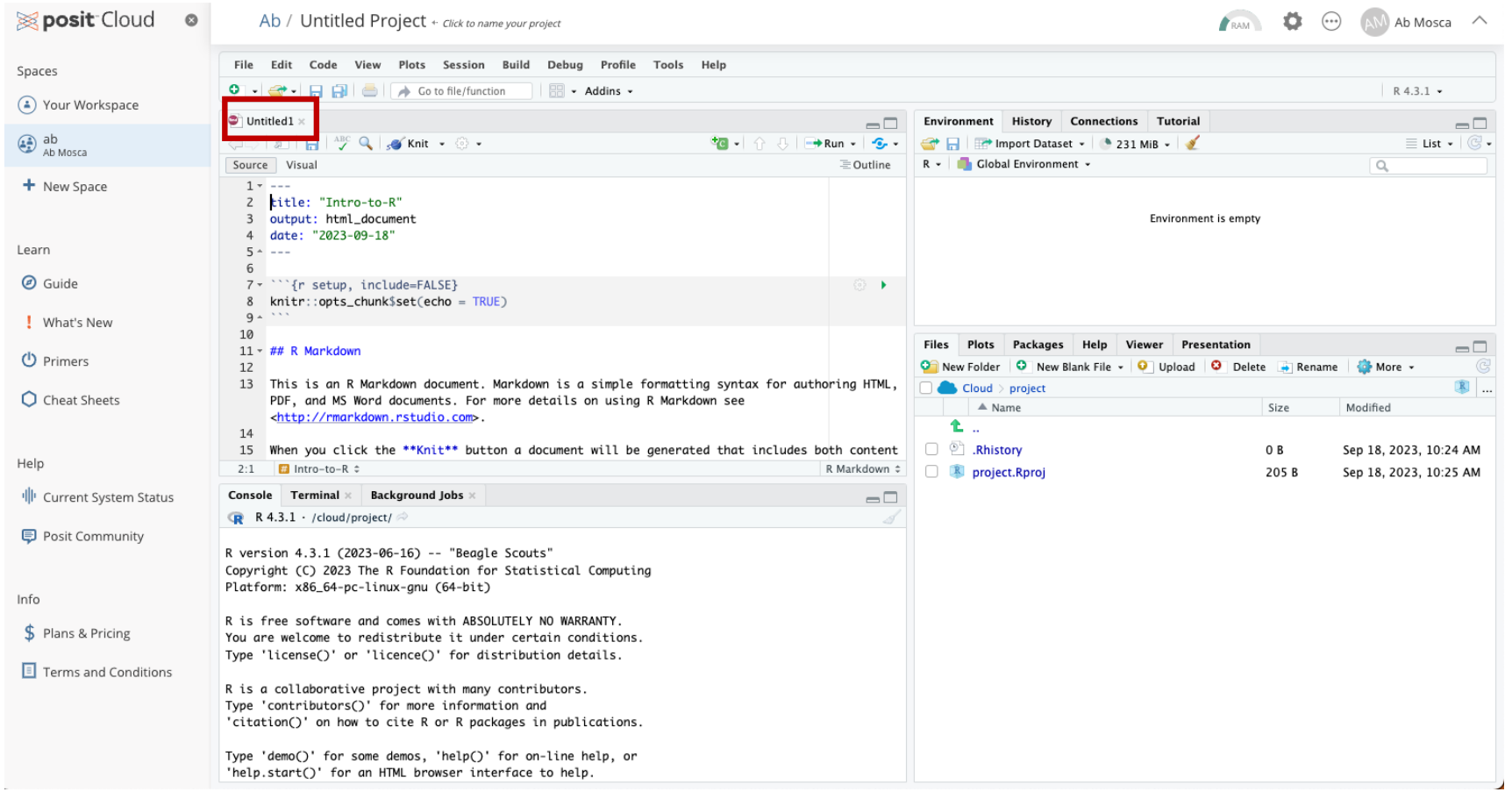
R Studio

- New file



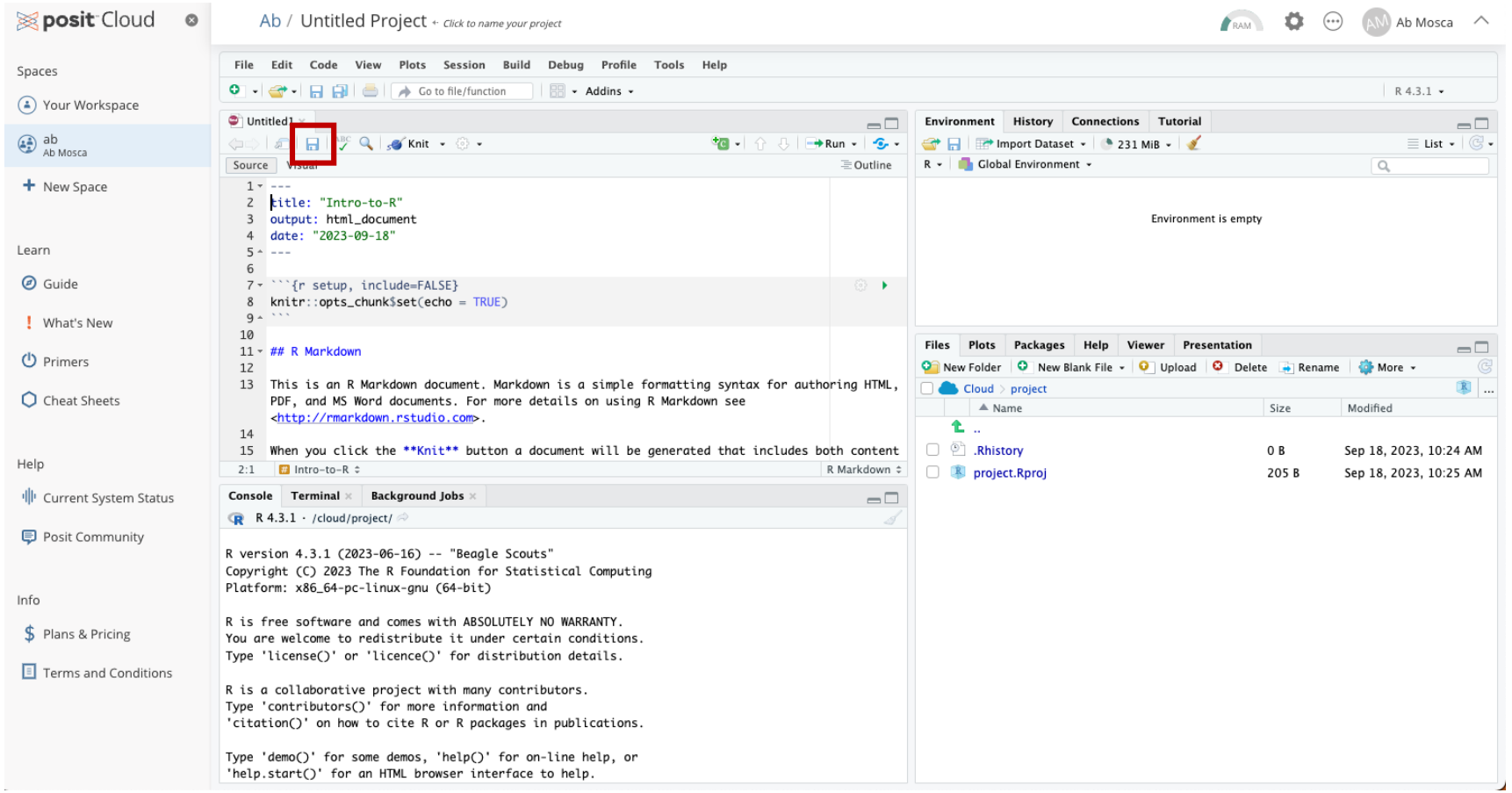
R Studio

- New file



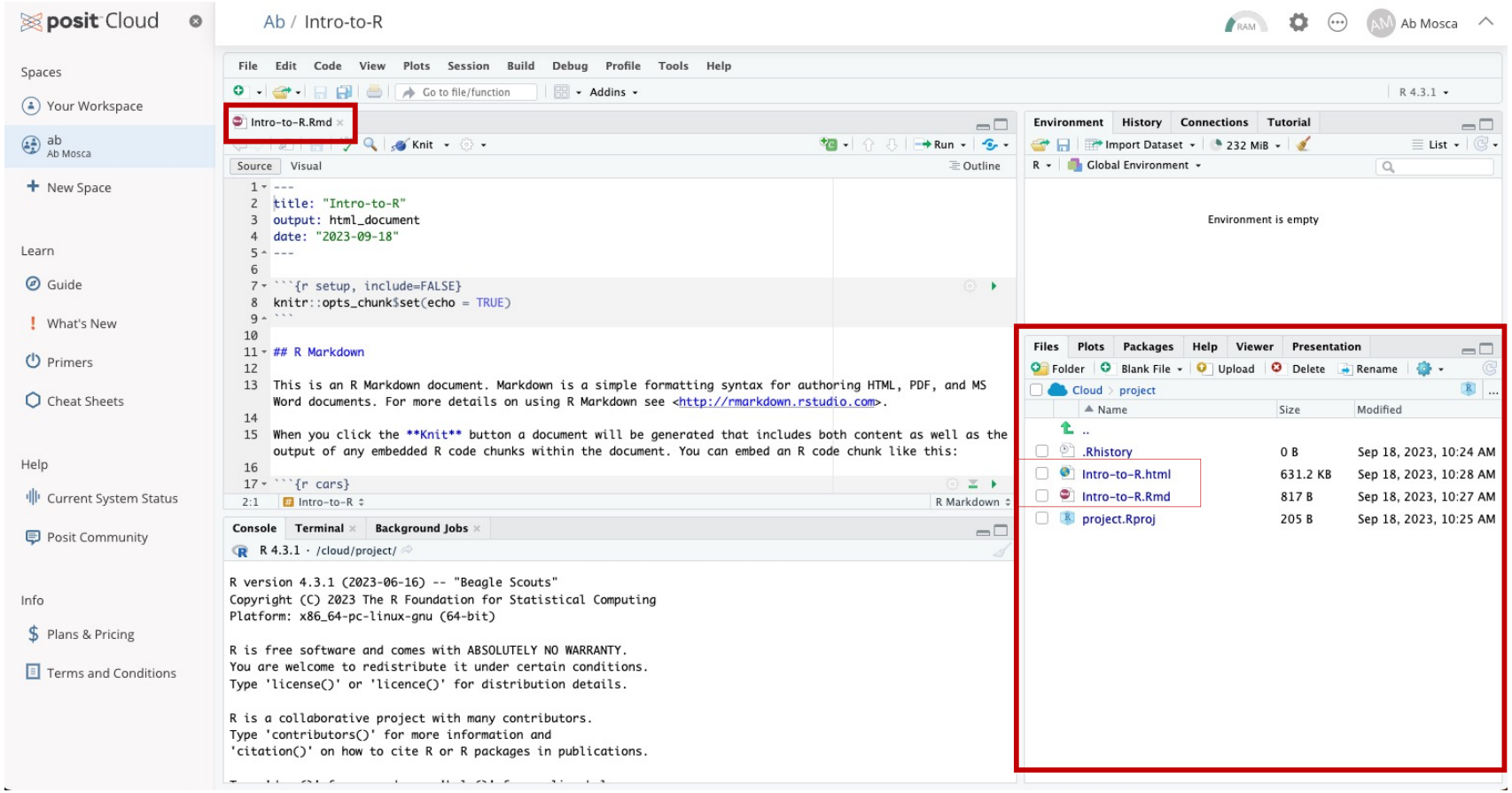
R Studio

- New file - SAVE



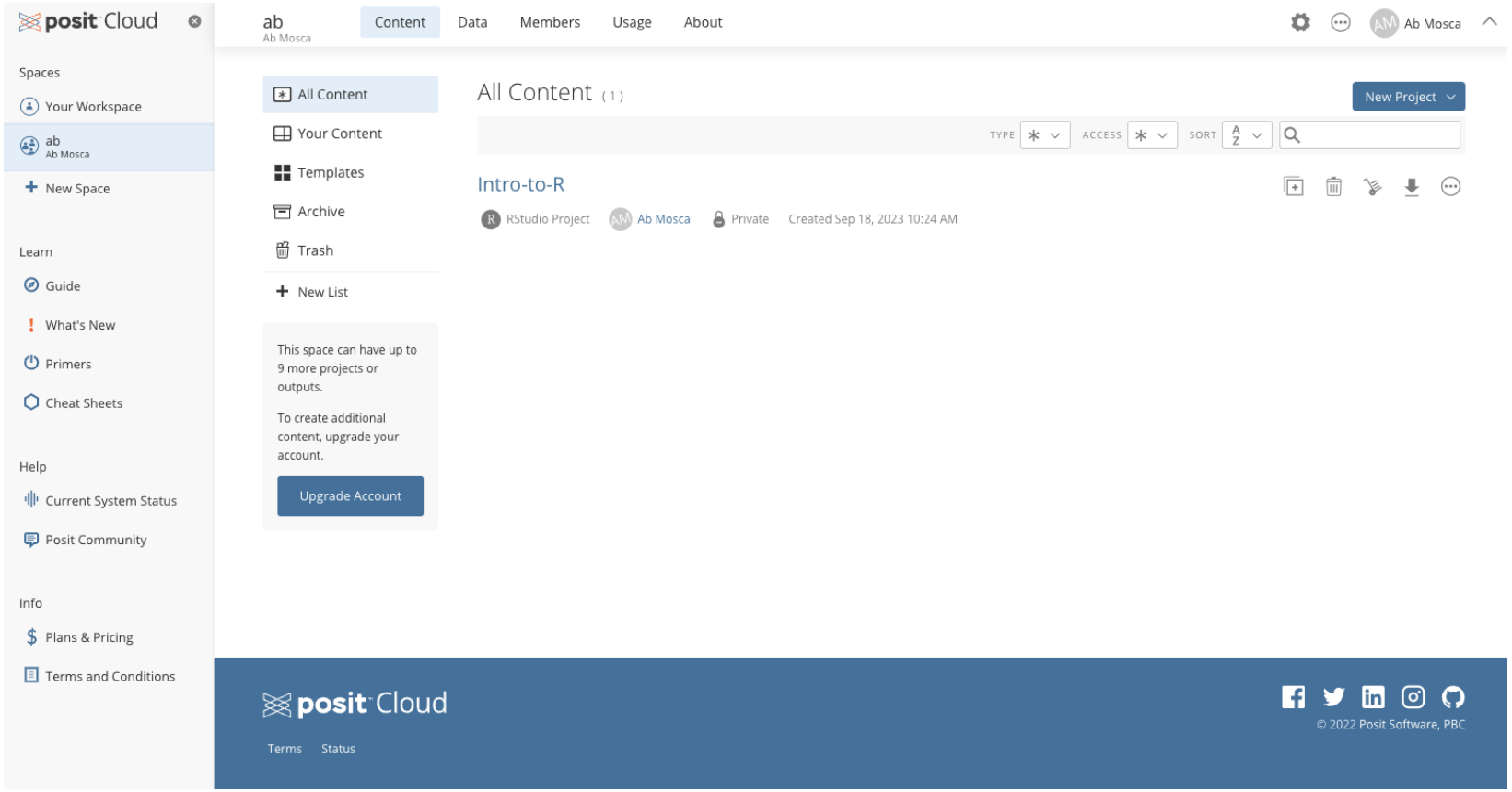
R Studio

- New file - saved



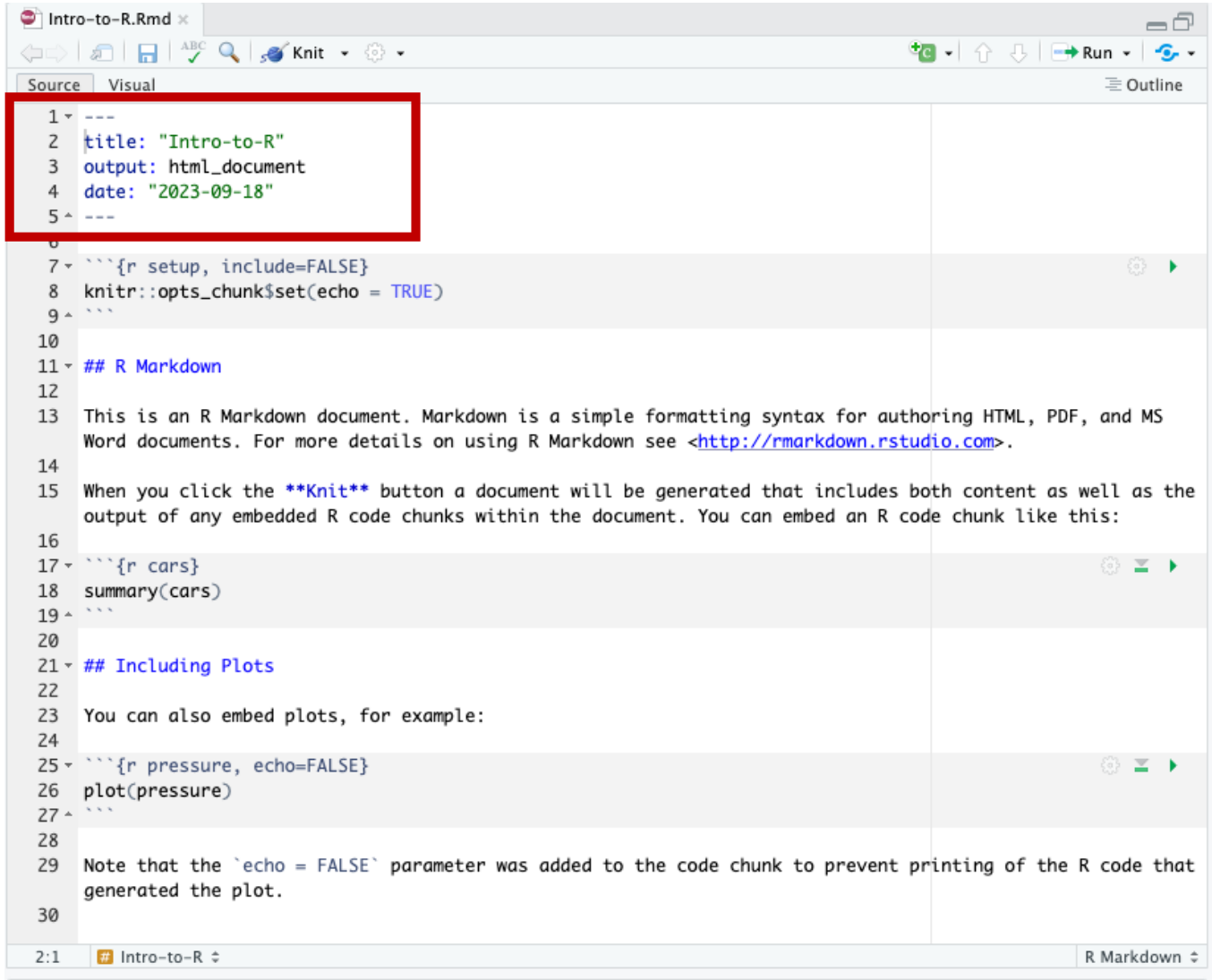
R Studio

- New file - saved



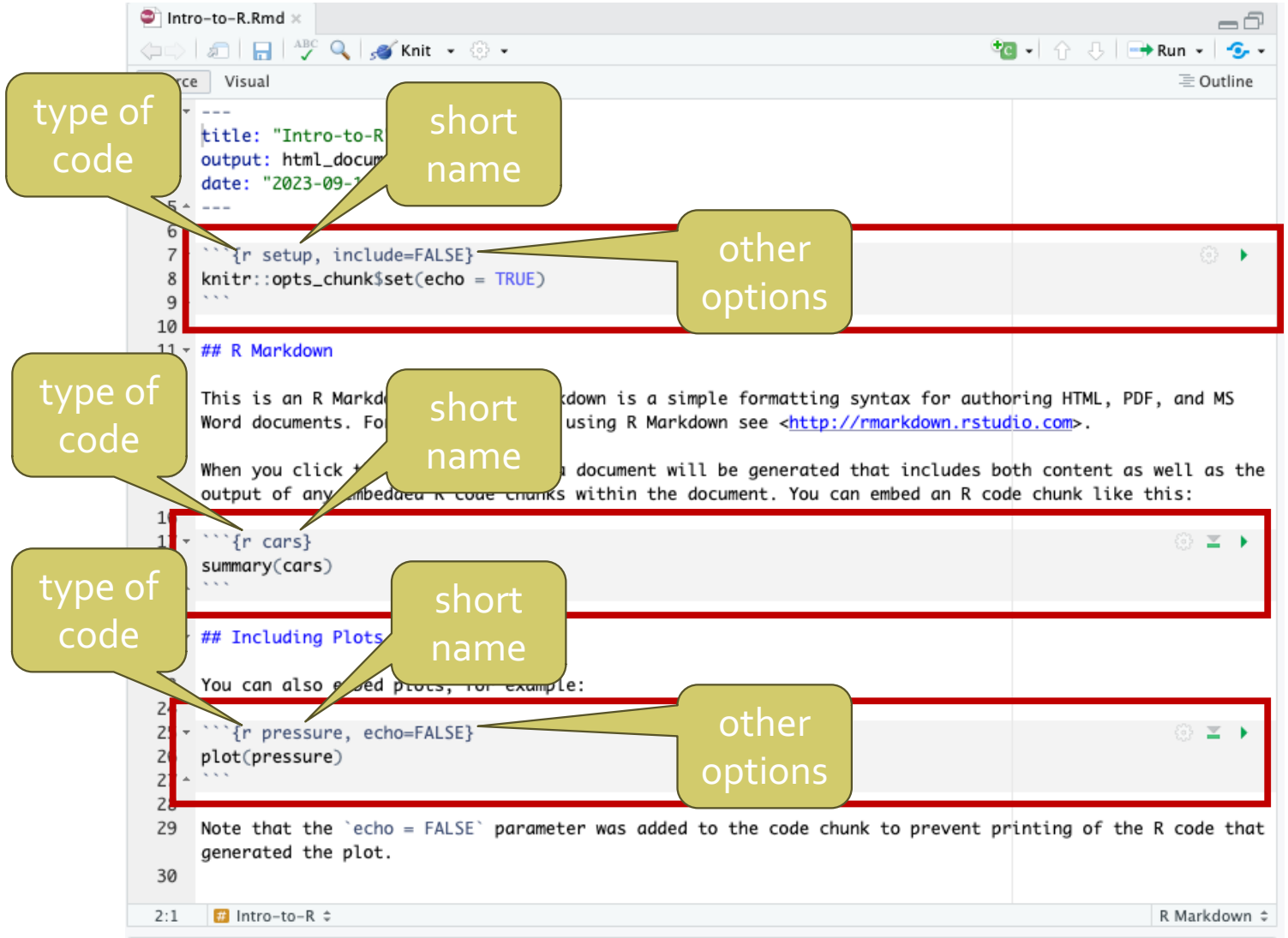
R Markdown

- Meta data



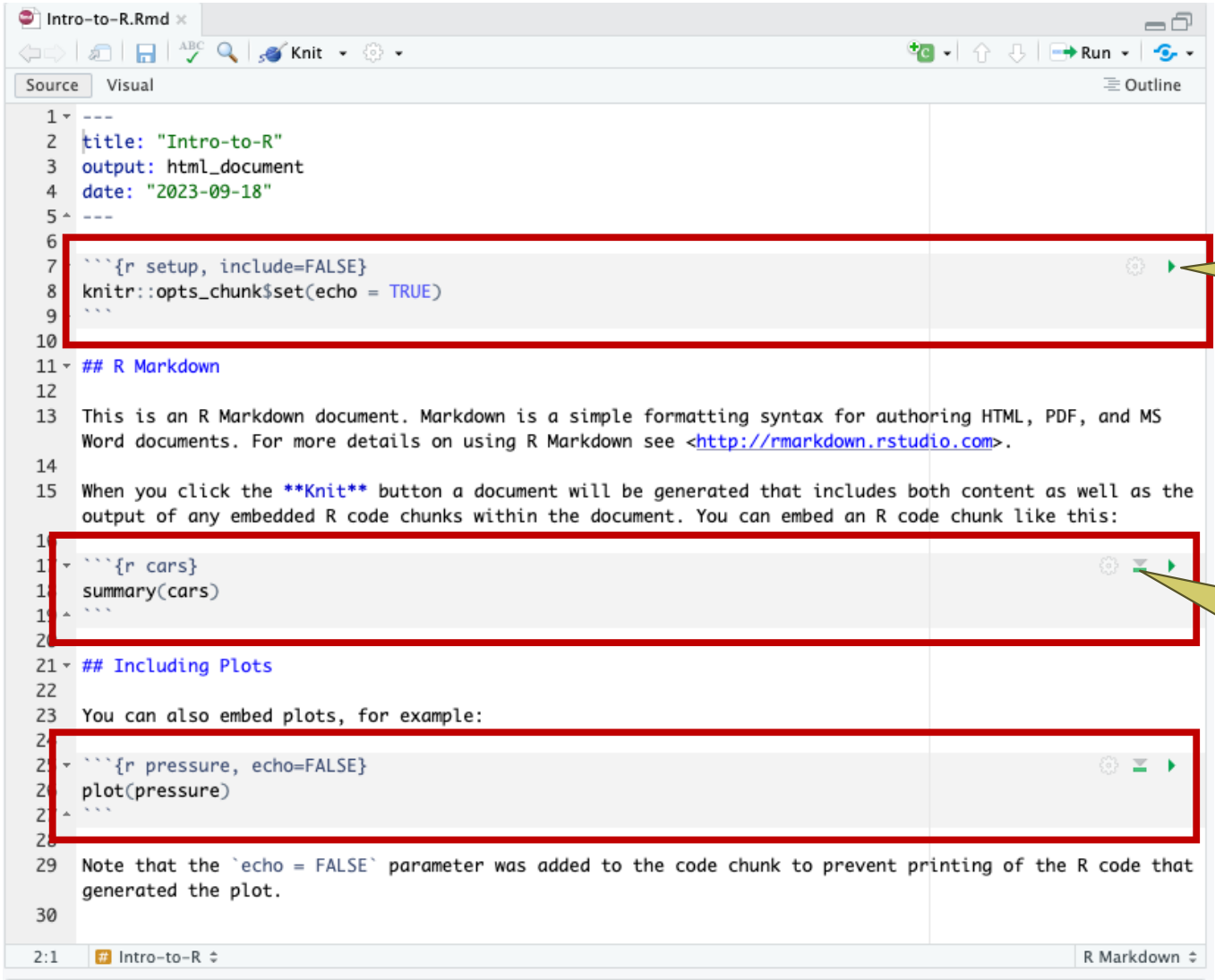
R Markdown

- Code chunks



R Markdown

- Code chunks

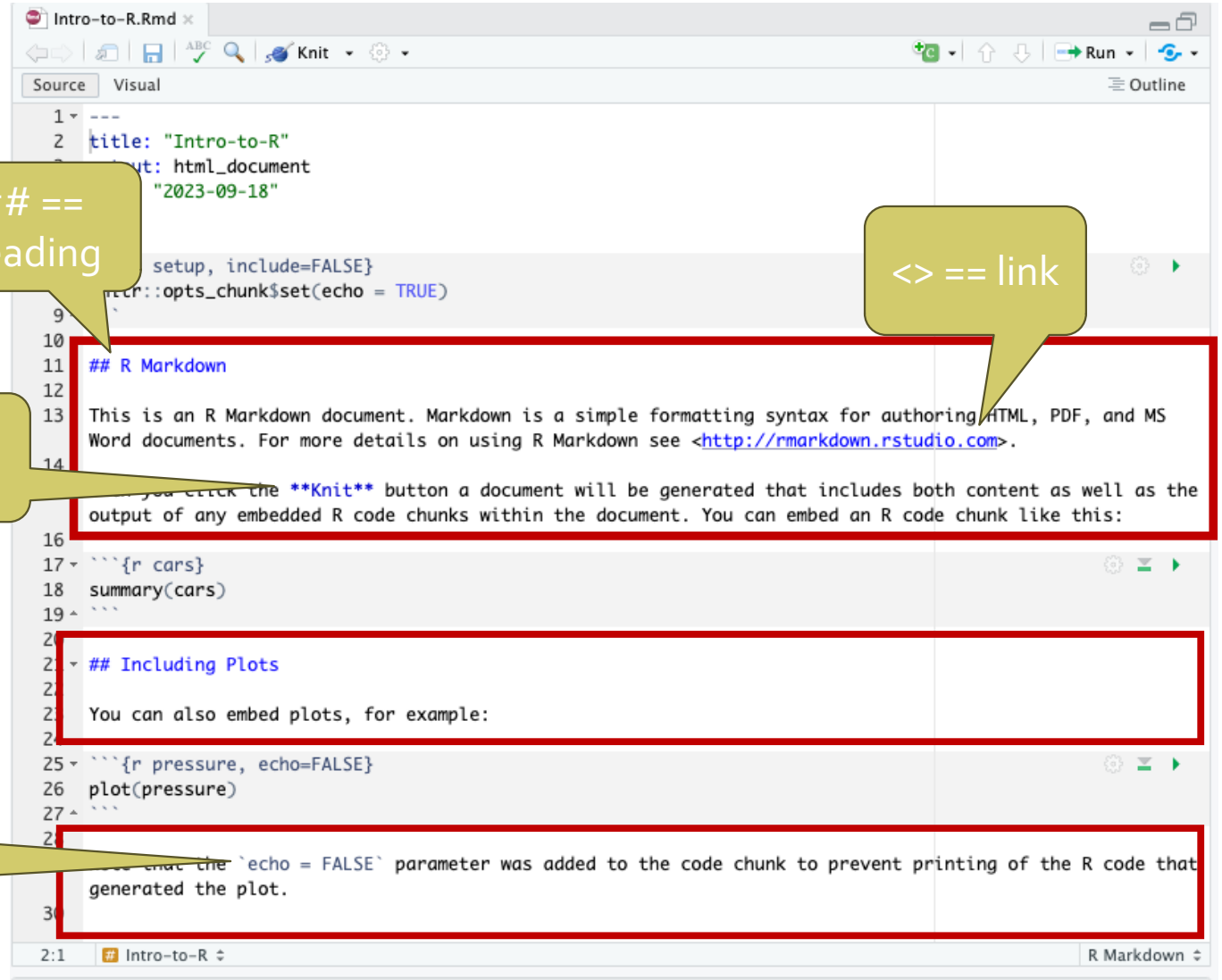


run chunk

run all preceding chunks

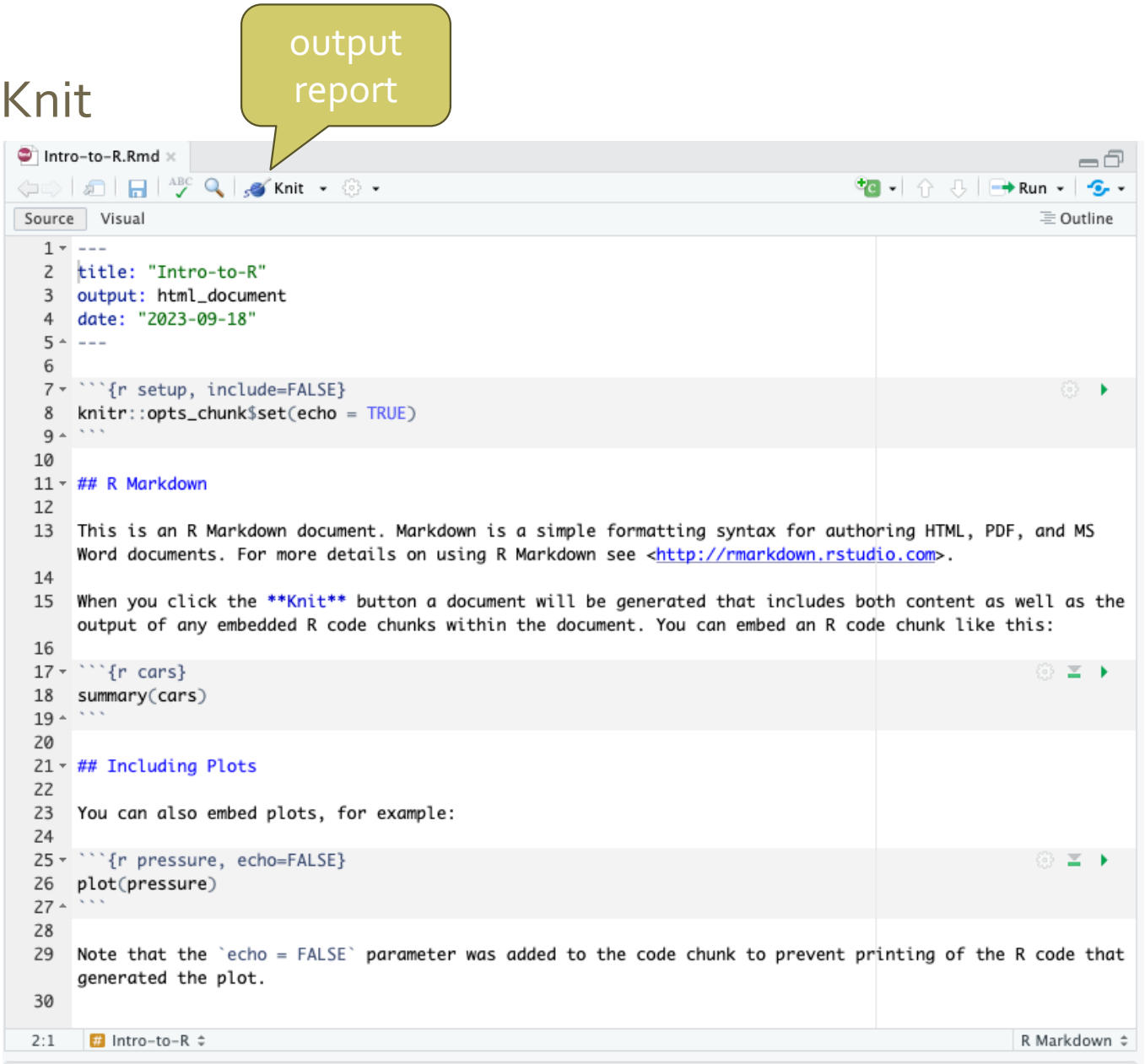
R Markdown

- Text



R Markdown

- Knit



R Markdown

- Knit

Intro-to-R.Rmd

Knit | Run

Source | Visual

```
1  ---
2  title: "Intro-to-R"
3  output: html_document
4  date: "2023-09-18"
5  ---
6
7  ```{r setup, include=FALSE}
8  knitr::opts_chunk$set(echo = TRUE)
9  ```
10
11 ## R Markdown
12
13 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS
   Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
14
15 When you click the **Knit** button a document will be generated that includes both content as well as the
   output of any embedded R code chunks within the document. You can embed an R code chunk like this:
16
17 ```{r cars}
18 summary(cars)
19 ```
20
21 ## Including Plots
22
23 You can also embed plots, for example:
24
25 ```{r pressure, echo=FALSE}
26 plot(pressure)
27 ```
28
29 Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that
   generated the plot.
30
```

2:1 | Intro-to-R | R Markdown

a2725c493c49447e91580193bcead4fa.app.posit.cloud/?view=rmarkdown

Intro-to-R.html | Open in Browser | Find

Intro-to-R

2023-09-18

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

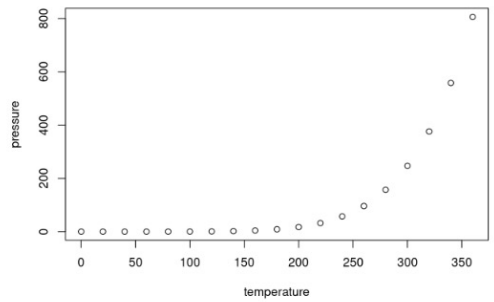
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

R Markdown

- Knit

Intro-to-R.Rmd

Knit | Run

Source | Visual | Outline

```
---
title: "Intro-to-R"
output: html_document
date: "2023-09-18"
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

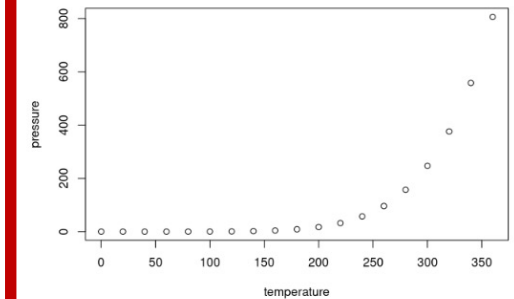
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```{r cars}
summary(cars)
```

Including Plots

You can also embed plots, for example:

```{r pressure, echo=FALSE}
plot(pressure)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```

2:1 | Intro-to-R | R Markdown

a2725c493c49447e91580193bcead4fa.app.posit.cloud/?view=rmarkdown

Intro-to-R.html | Open in Browser | Find

Intro-to-R

2023-09-18

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

R Markdown

- Knit

Intro-to-R.Rmd

Knit | Run | Source | Visual | Outline

```
---
title: "Intro-to-R"
output: html_document
date: "2023-09-18"
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

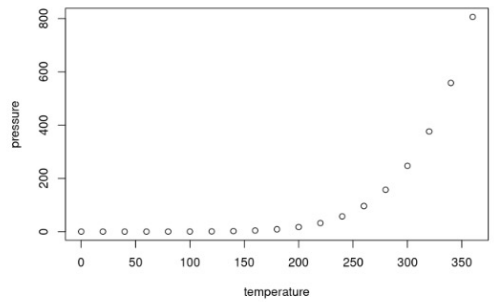
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```{r cars}
summary(cars)
```

Including Plots

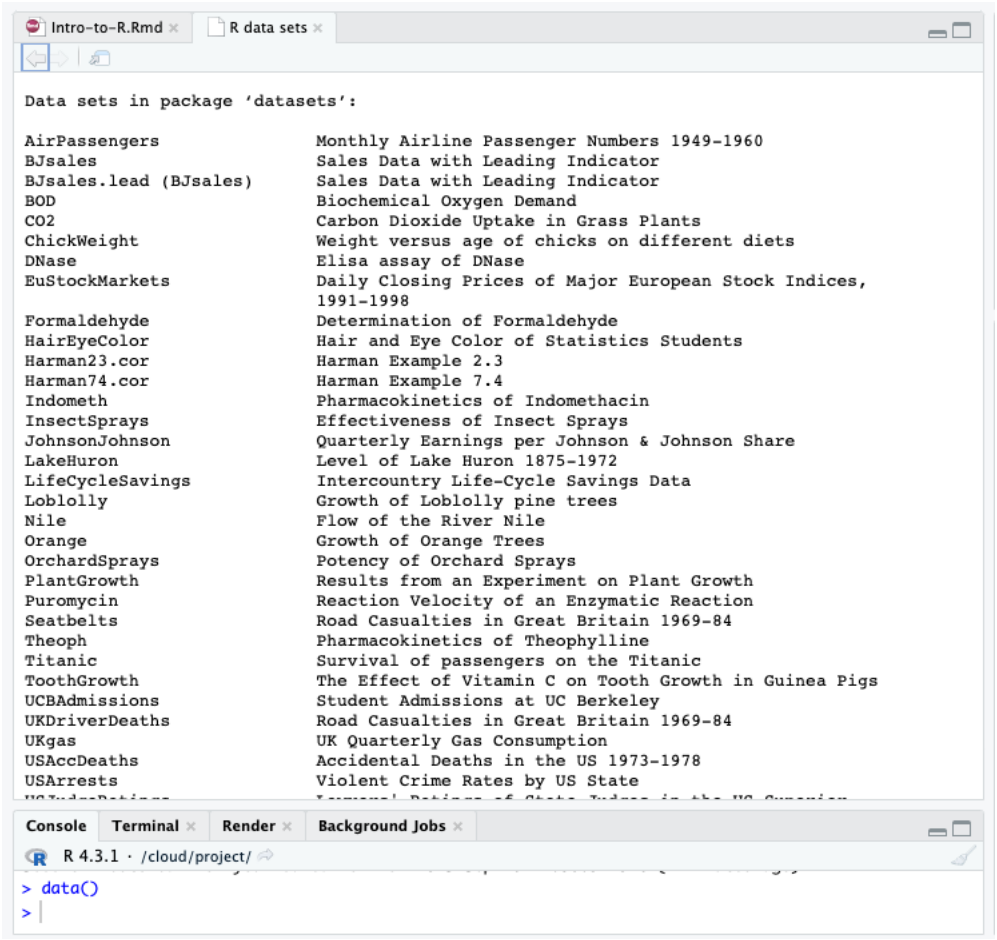
You can also embed plots, for example:

```{r pressure, echo=FALSE}
plot(pressure)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```

2:1 Intro-to-R — R Markdown

a2725c493c49447e91580193bcead4fa.app.posit.cloud/?view=rmarkdown

Intro-to-R.html | Open in Browser | Find

Intro-to-R

2023-09-18

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

R

- Packages

```
```{r load-packages, message = FALSE, eval=TRUE}
library(tidyverse)
```
```

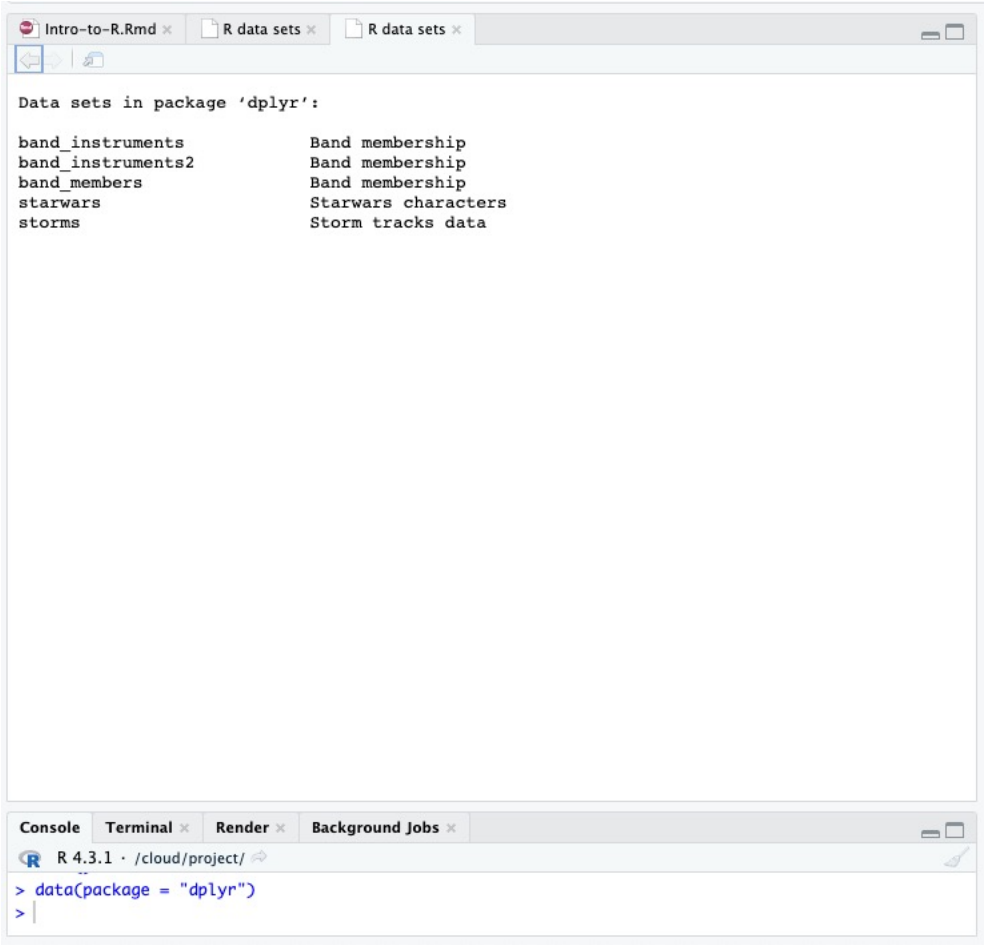
R

- Data


Intro-to-R.Rmd
R data sets
Data sets in package 'datasets':
AirPassengers Monthly Airline Passenger Numbers 1949-1960
BJsales Sales Data with Leading Indicator
BJsales.lead (BJsales) Sales Data with Leading Indicator
BOD Biochemical Oxygen Demand
CO2 Carbon Dioxide Uptake in Grass Plants
ChickWeight Weight versus age of chicks on different diets
DNase Elisa assay of DNase
EuStockMarkets Daily Closing Prices of Major European Stock Indices, 1991-1998
Formaldehyde Determination of Formaldehyde
HairEyeColor Hair and Eye Color of Statistics Students
Harman23.cor Harman Example 2.3
Harman74.cor Harman Example 7.4
Indometh Pharmacokinetics of Indomethacin
InsectSprays Effectiveness of Insect Sprays
JohnsonJohnson Quarterly Earnings per Johnson & Johnson Share
LakeHuron Level of Lake Huron 1875-1972
LifeCycleSavings Intercountry Life-Cycle Savings Data
Loblolly Growth of Loblolly pine trees
Nile Flow of the River Nile
Orange Growth of Orange Trees
OrchardSprays Potency of Orchard Sprays
PlantGrowth Results from an Experiment on Plant Growth
Puromycin Reaction Velocity of an Enzymatic Reaction
Seatbelts Road Casualties in Great Britain 1969-84
Theoph Pharmacokinetics of Theophylline
Titanic Survival of passengers on the Titanic
ToothGrowth The Effect of Vitamin C on Tooth Growth in Guinea Pigs
UCBAdmissions Student Admissions at UC Berkeley
UKDriverDeaths Road Casualties in Great Britain 1969-84
UKgas UK Quarterly Gas Consumption
USAccDeaths Accidental Deaths in the US 1973-1978
USArrests Violent Crime Rates by US State
Console
Terminal
Render
Background Jobs
R 4.3.1 · /cloud/project/
> data()
>

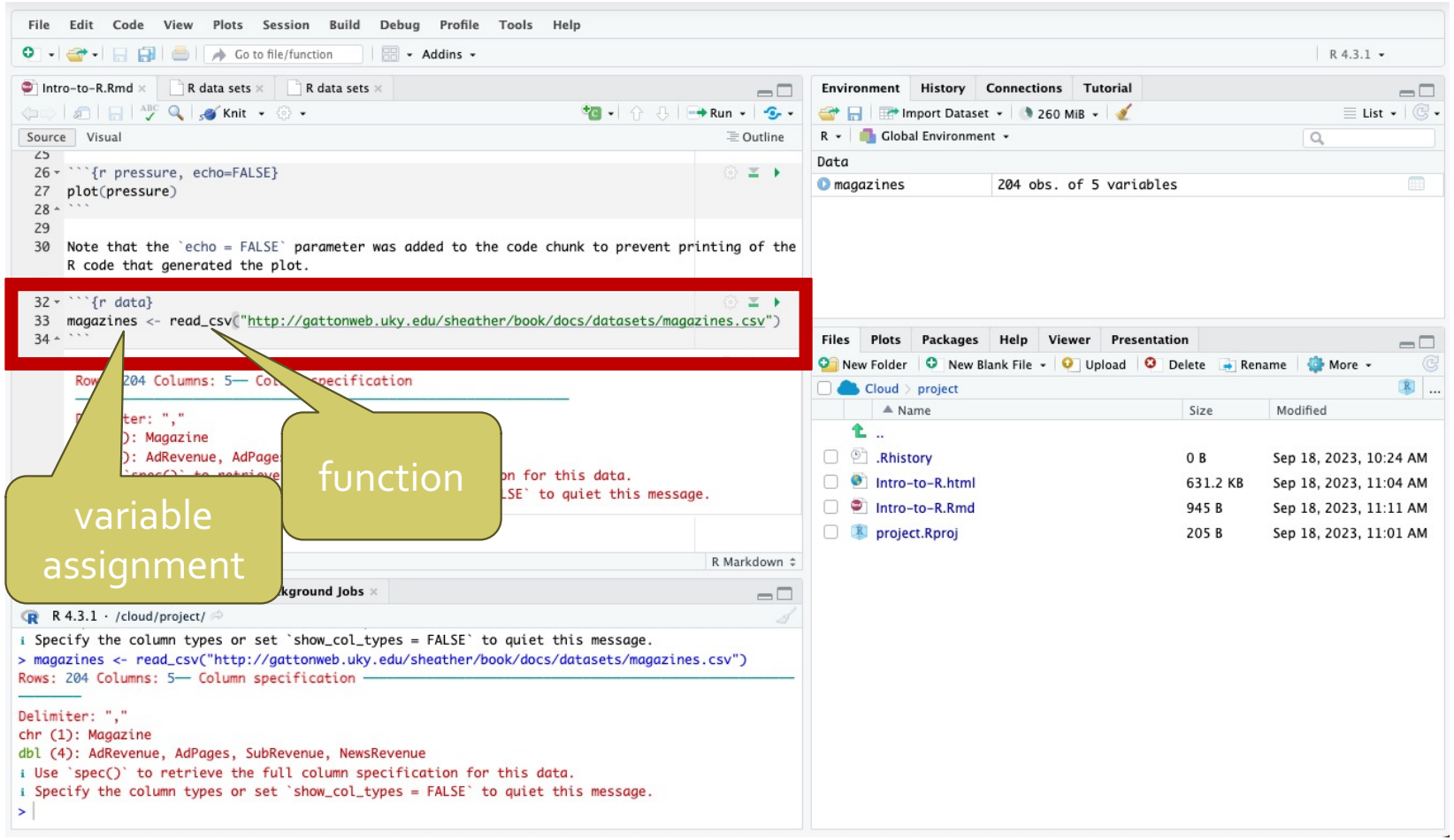
R

- Data


Intro-to-R.Rmd
R data sets
R data sets
Data sets in package 'dplyr':
band_instruments Band membership
band_instruments2 Band membership
band_members Band membership
starwars Starwars characters
storms Storm tracks data
Console
Terminal
Render
Background Jobs
R 4.3.1 · /cloud/project/
> data(package = "dplyr")
>

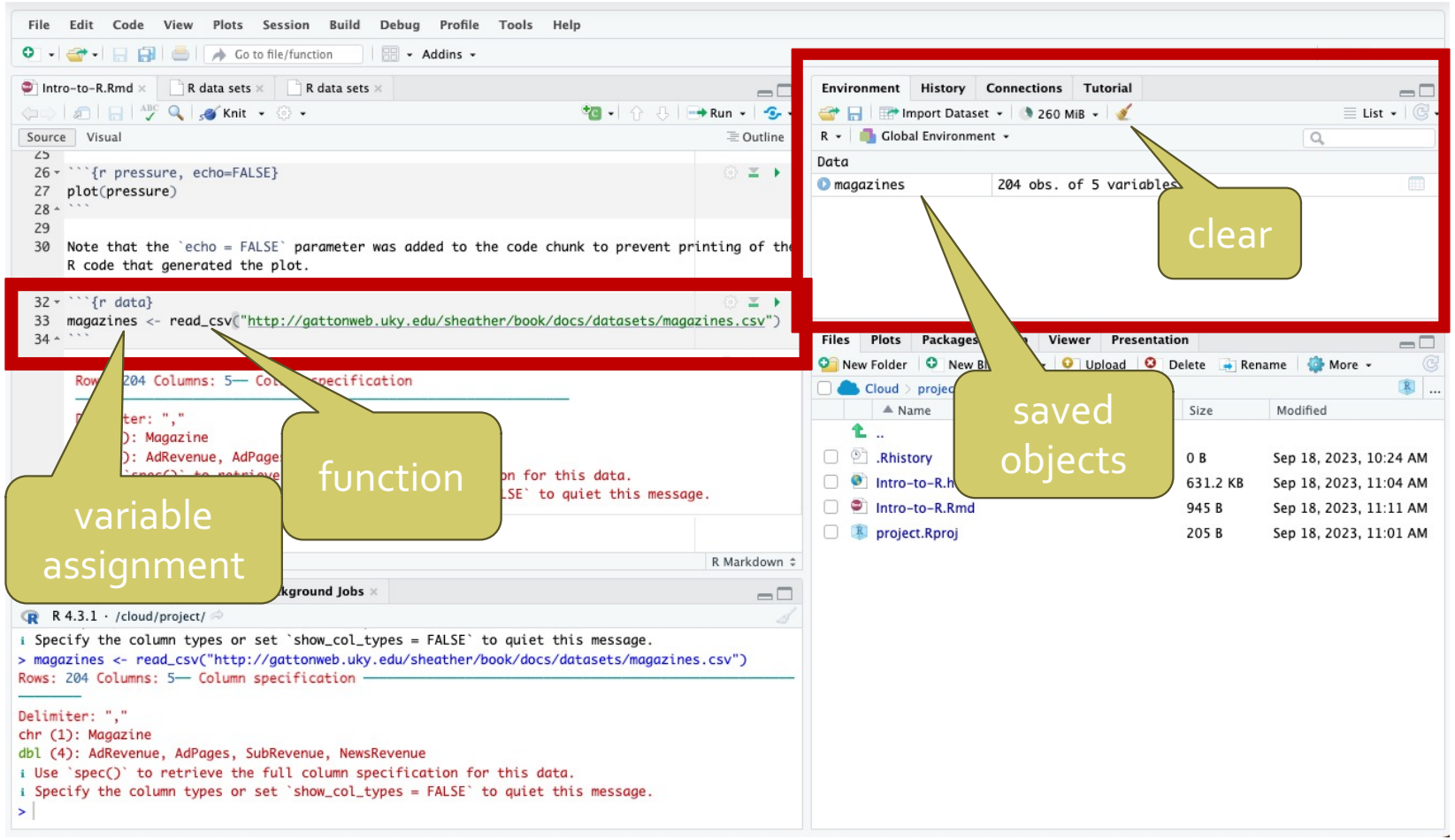
R

- Data – reading a CSV



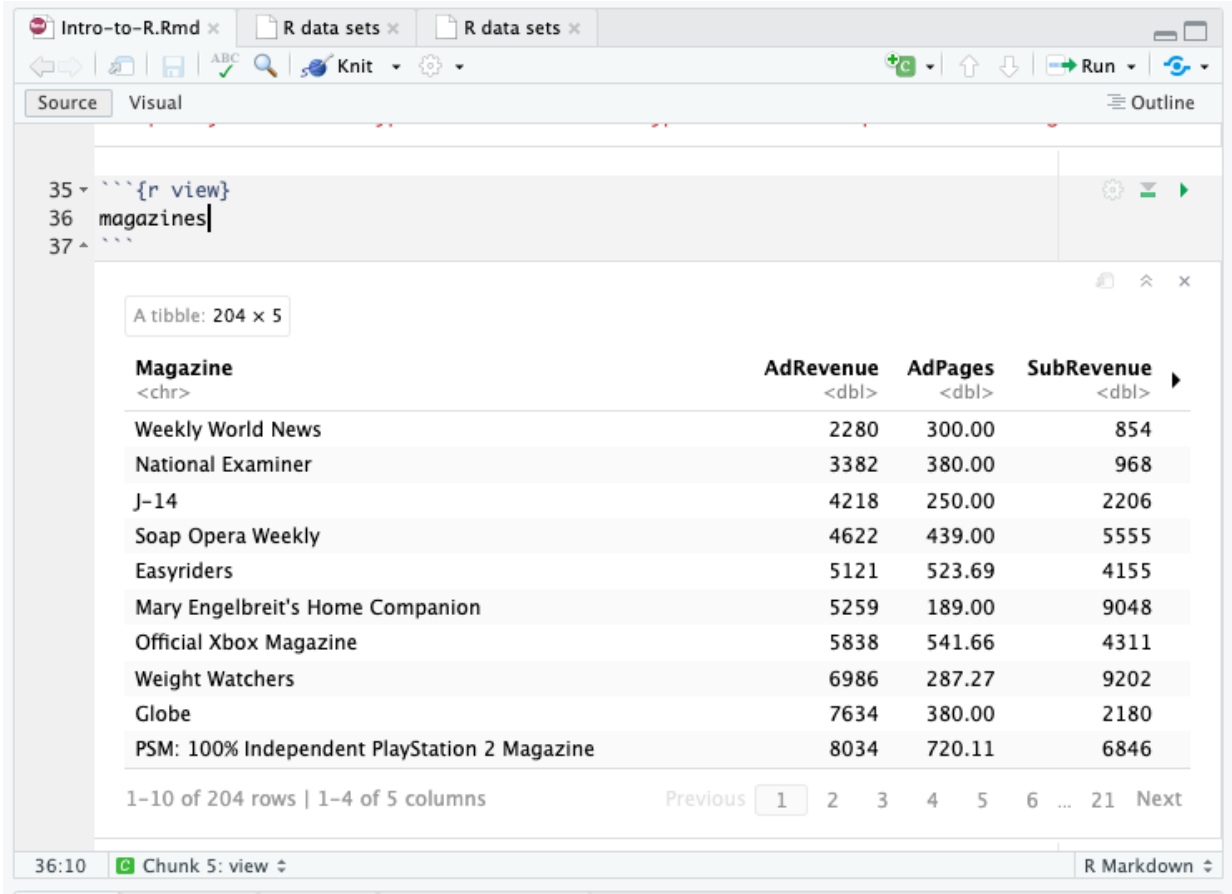
R

- Data – reading a CSV



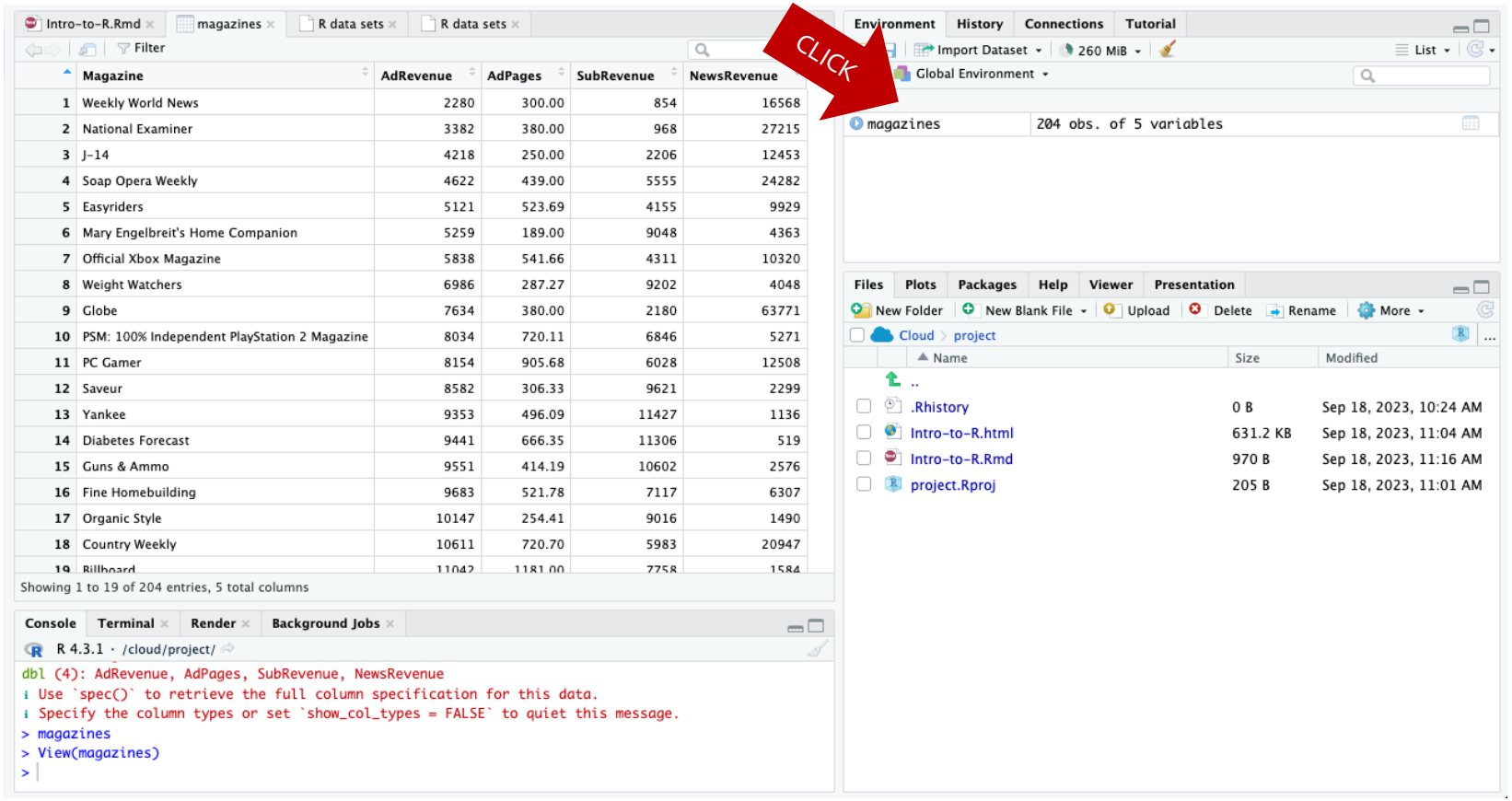
R

- Data – viewing



R

- Data – viewing



R

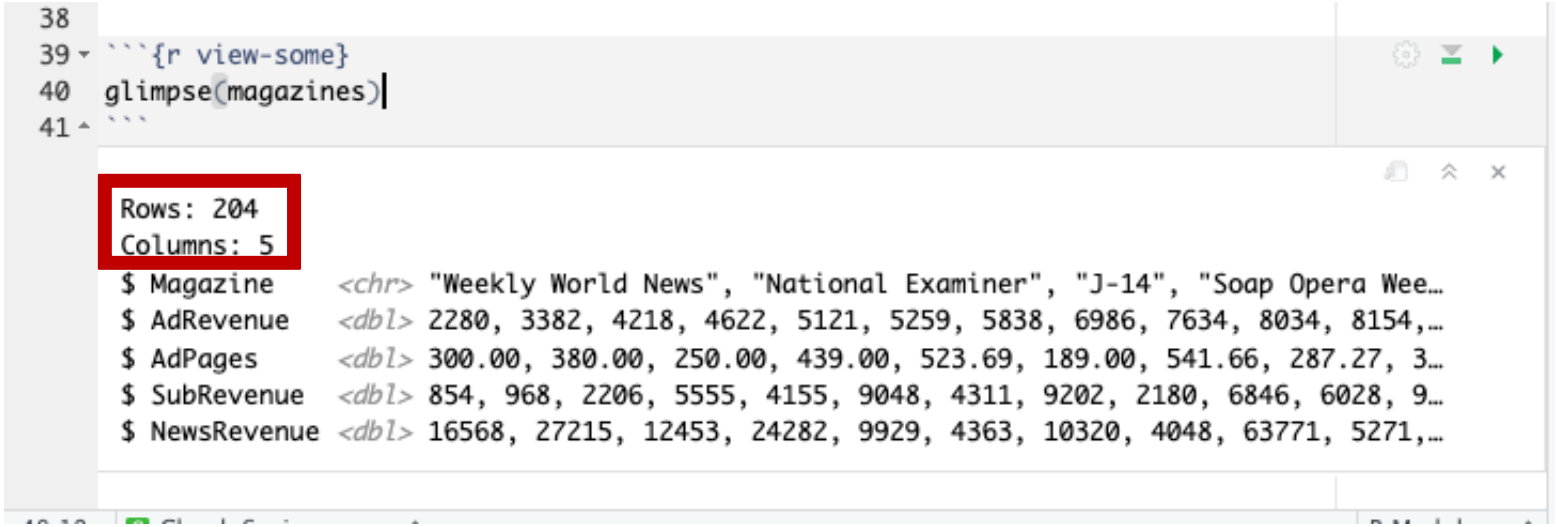
- Data – viewing

```
38
39 ```{r view-some}
40 glimpse(magazines)
41 ```
```

```
Rows: 204
Columns: 5
$ Magazine    <chr> "Weekly World News", "National Examiner", "J-14", "Soap Opera Wee…
$ AdRevenue   <dbl> 2280, 3382, 4218, 4622, 5121, 5259, 5838, 6986, 7634, 8034, 8154,…
$ AdPages     <dbl> 300.00, 380.00, 250.00, 439.00, 523.69, 189.00, 541.66, 287.27, 3…
$ SubRevenue  <dbl> 854, 968, 2206, 5555, 4155, 9048, 4311, 9202, 2180, 6846, 6028, 9…
$ NewsRevenue <dbl> 16568, 27215, 12453, 24282, 9929, 4363, 10320, 4048, 63771, 5271,…
```

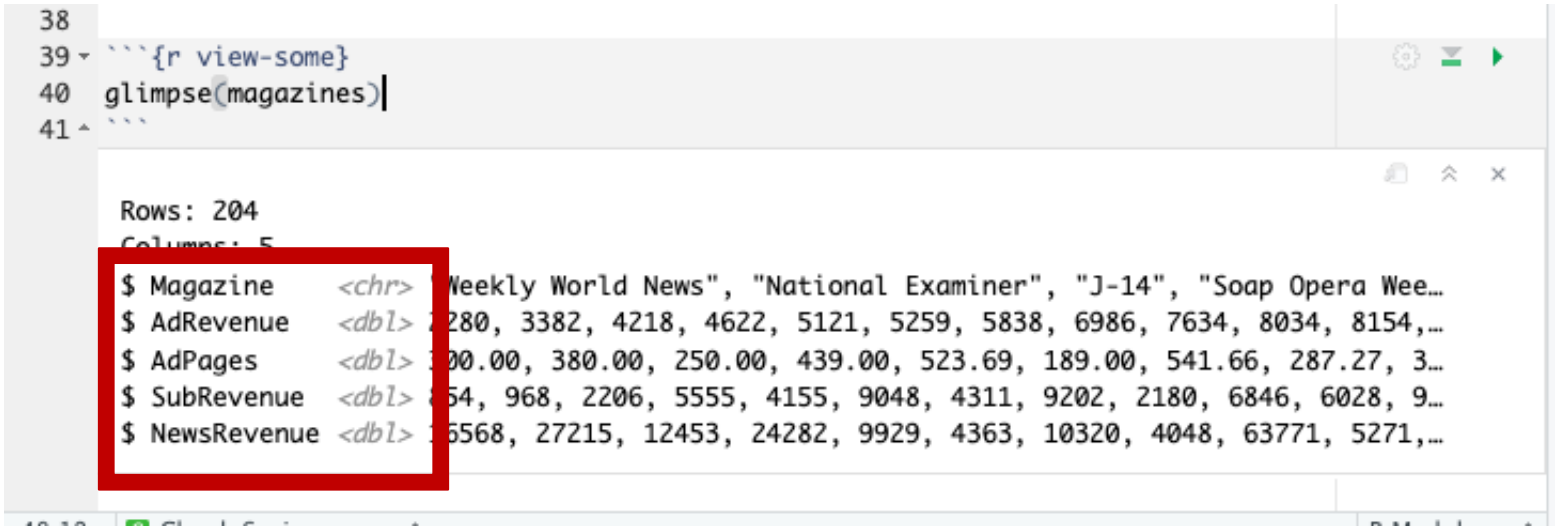
R

- Data – viewing



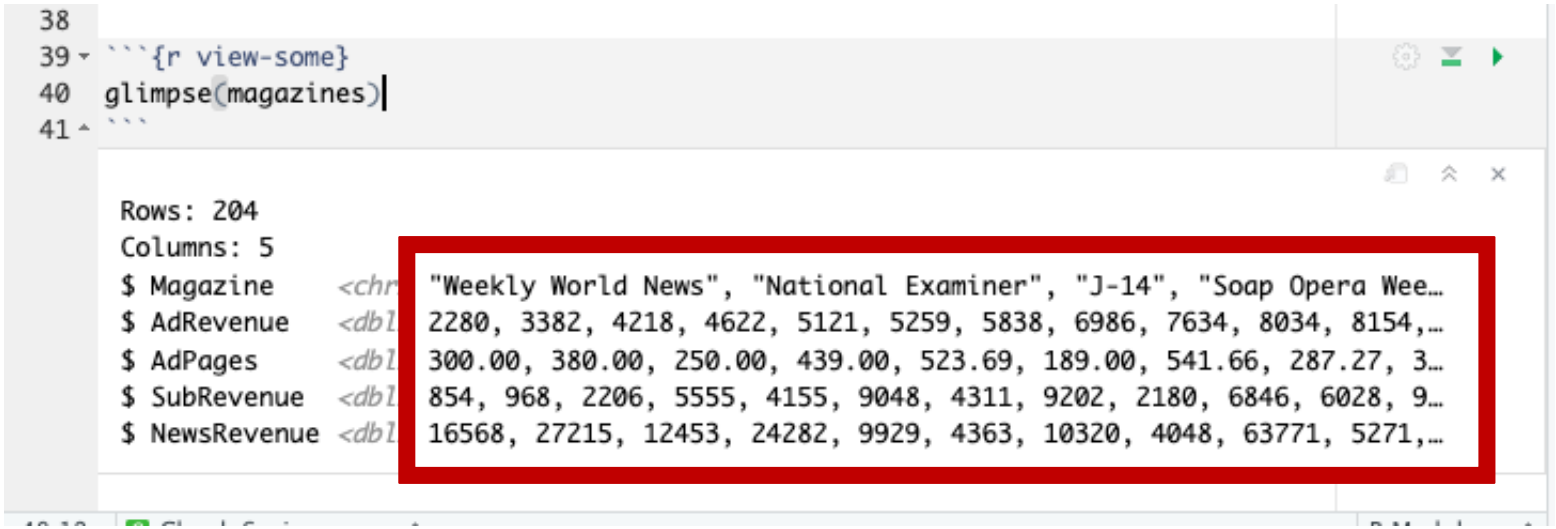
R

- Data – viewing



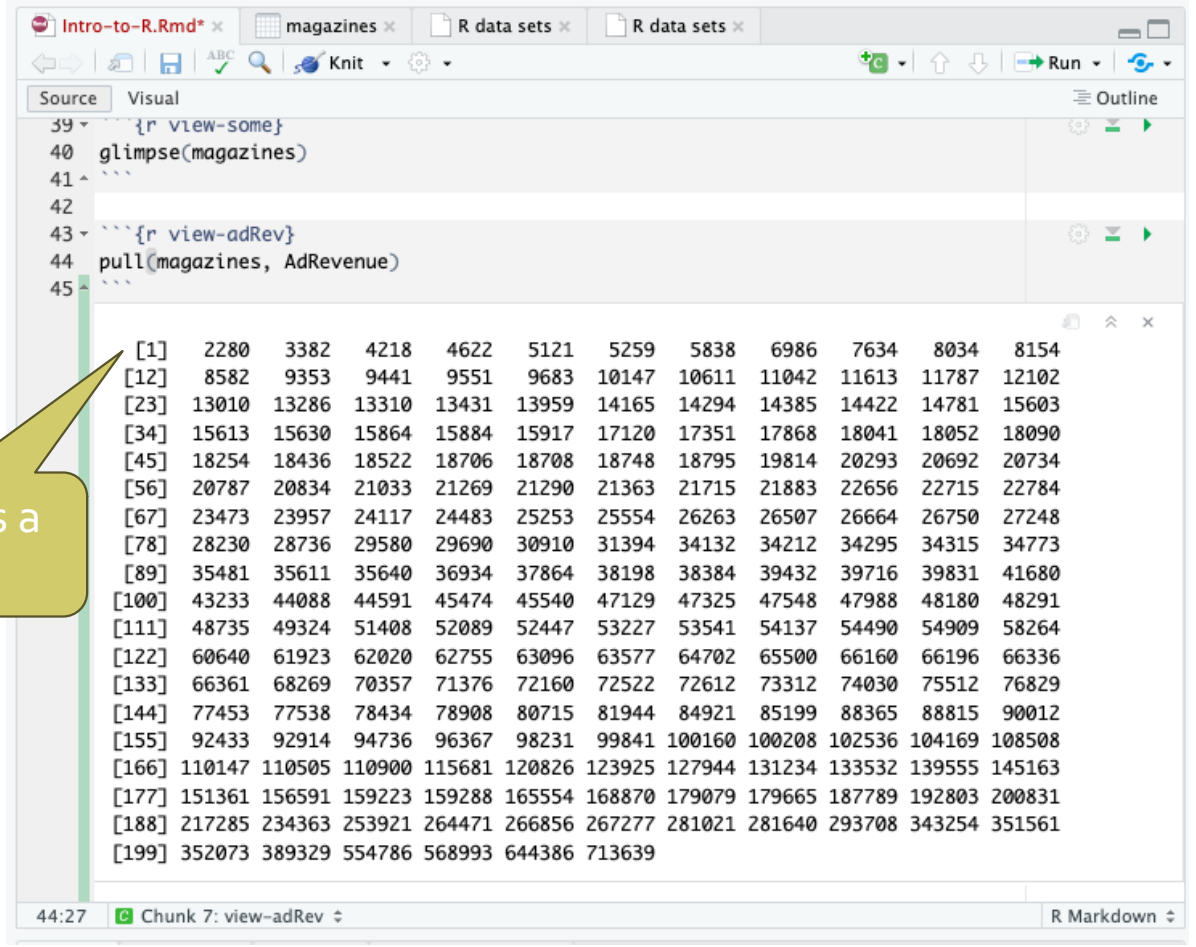
R

- Data – viewing



R

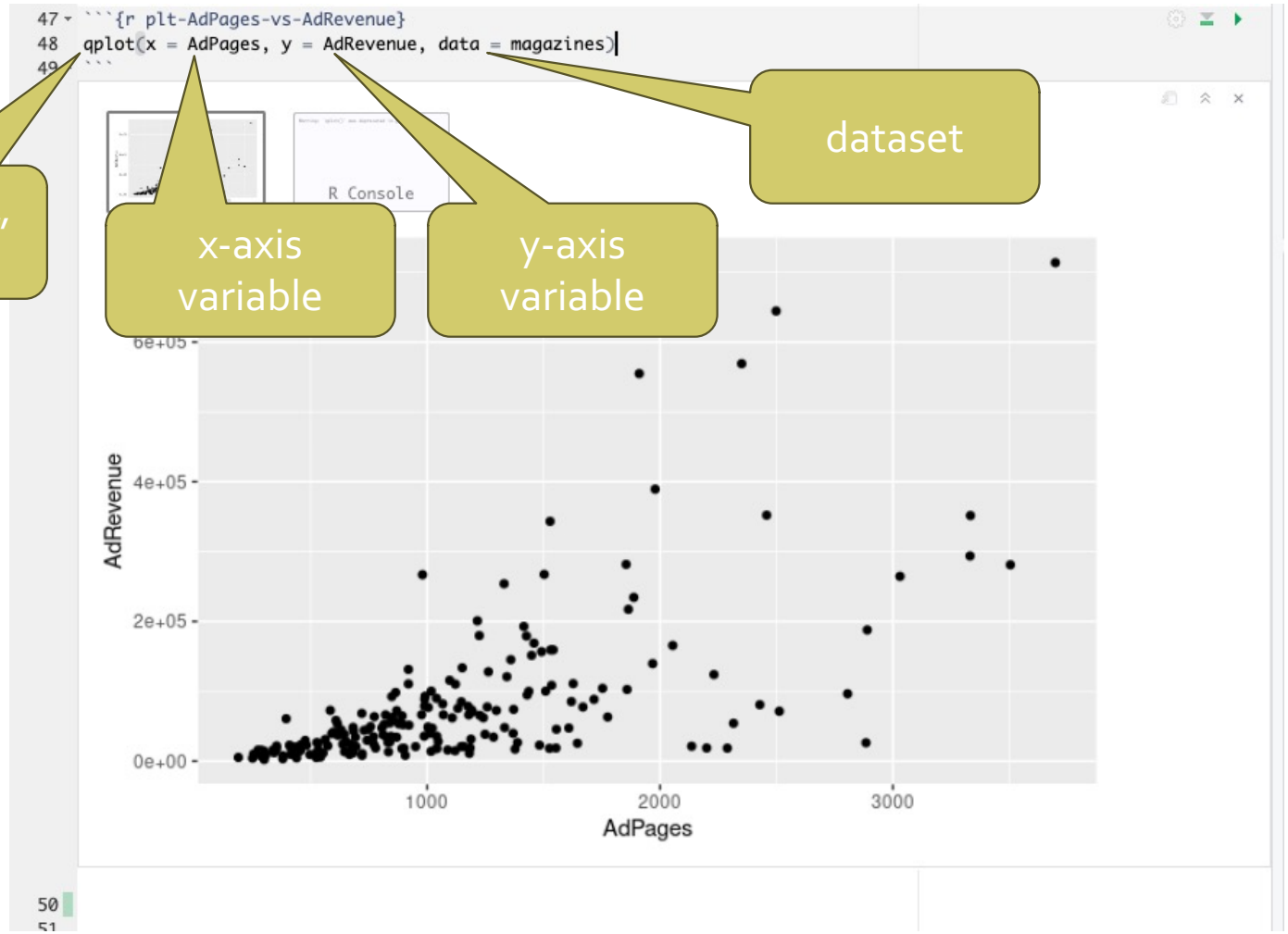
- Data – single column



column is a vector

R

- Visualization

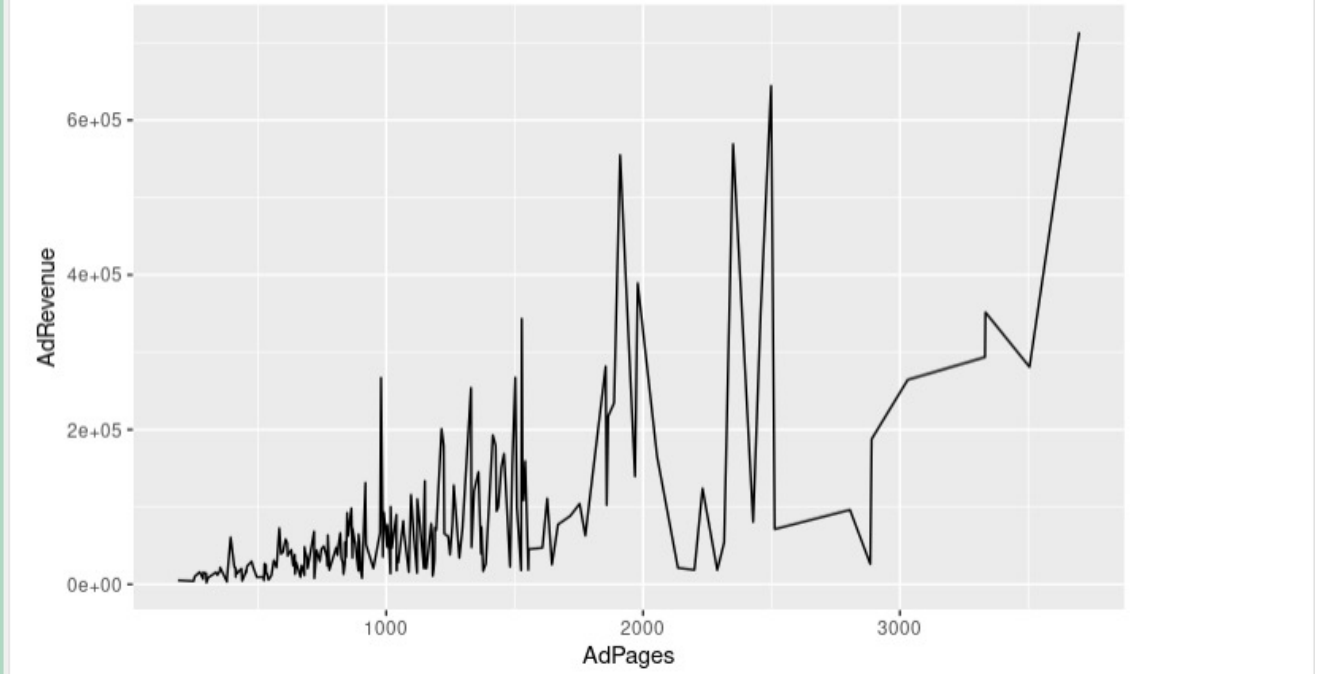


R

- Visualization

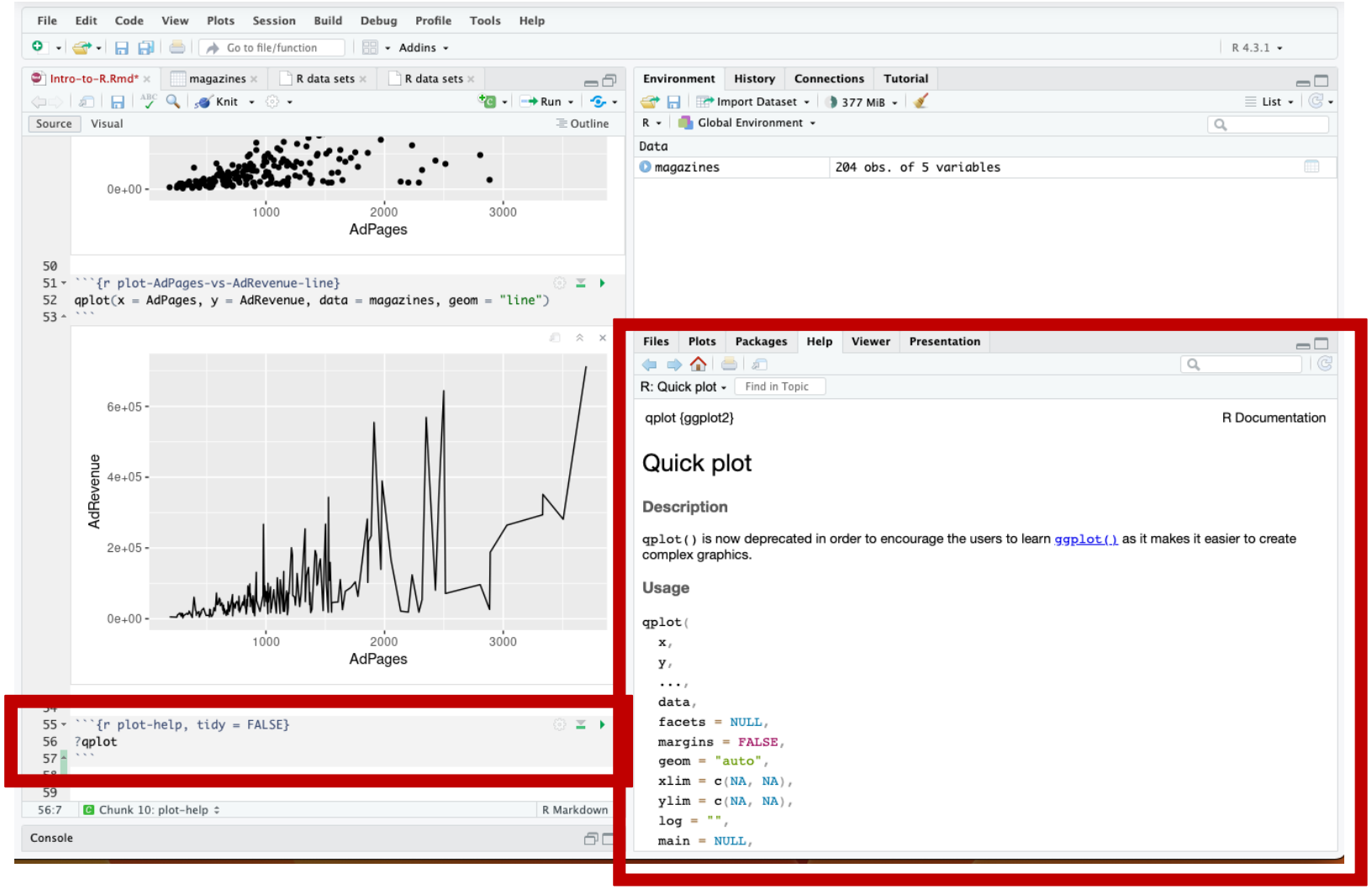
```
50
51 ```{r plot-AdPages-vs-AdRevenue-line}
52 qplot(x = AdPages, y = AdRevenue, data = magazines, geom = "line")
53 ```
54
```

"geometry"



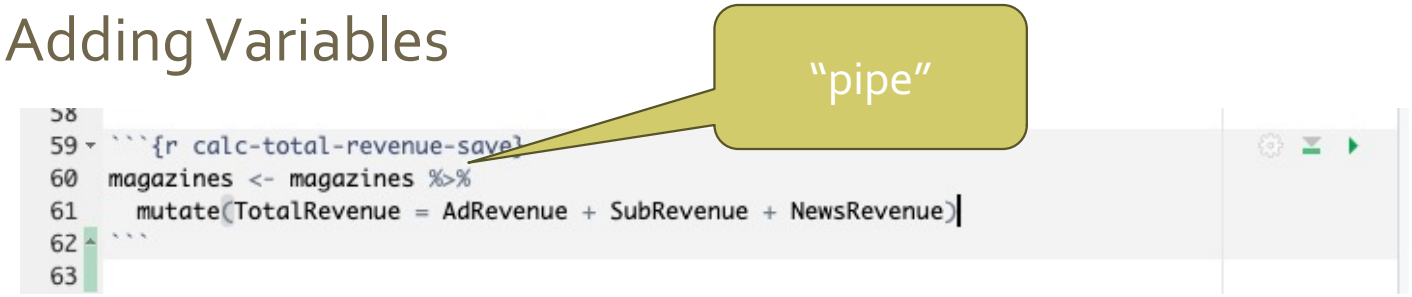
R

- Documentation



R

- Adding Variables

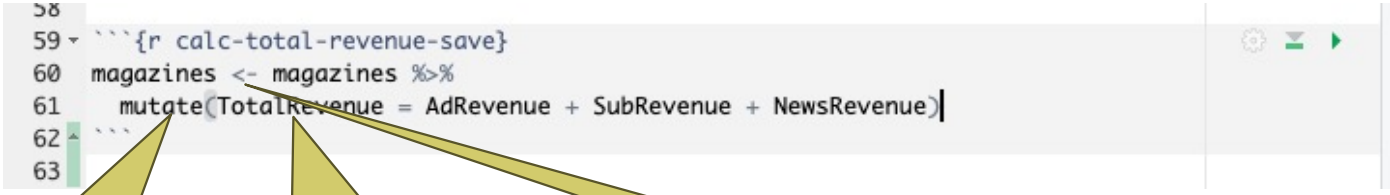


- %>% strings functions together
- Takes output of pervious expression and sends it as first parameter to next function in the chain

- Ex. `x %>% f(y)` is equivalent to `f(x, y)`

R

- Adding Variables



change magazines by …

adding new var

set magazines to new dataset

R

- Adding Variables

```
58
59 ```{r calc-total-revenue-save}
60 magazines <- magazines %>%
61   mutate(TotalRevenue = AdRevenue + SubRevenue + NewsRevenue)
62 ```
63
```

Intro-to-R.Rmd | magazines | R data sets

Filter

	Magazine	AdRevenue	AdPages	SubRevenue	NewsRevenue	TotalRevenue
1	Weekly World News	2280	300.00	854	16568	19702
2	National Examiner	3382	380.00	968	27215	31565
3	J-14	4218	250.00	2206	12453	18877
4	Soap Opera Weekly	4622	439.00	5555	24282	34459
5	Easyriders	5121	523.69	4155	9929	19205
6	Mary Engelbreit's Home Companion	5259	189.00	9048	4363	18670
7	Official Xbox Magazine	5838	541.66	4311	10320	20469
8	Weight Watchers	6986	287.27	9202	4048	20236
9	Globe	7634	380.00	2180	63771	73585
10	PSM: 100% Independent PlayStation 2 Magazine	8034	720.11	6846	5271	20151
11	PC Gamer	8154	905.68	6028	12508	26690
12	Saveur	8582	306.33	9621	2299	20502
13	Yankee	9353	496.09	11427	1136	21916
14	Diabetes Forecast	9441	666.35	11306	519	21266
15	Guns & Ammo	9551	414.19	10602	2576	22729
16	Fine Homebuilding	9683	521.78	7117	6307	23107
17	Organic Style	10147	254.41	9016	1490	20653
18	Country Weekly	10611	720.70	5983	20947	37541
19	Billboard	11042	1181.00	7758	1584	20384
20	Honey	11613	553.00	5467	1454	18534
21	Computer Gaming World	11787	681.26	4376	4413	20576
22	Wood	12102	342.30	14345	3082	29529
23	Fit Pregnancy	13010	536.75	5352	3311	21673
24	Windows & .NET Magazine	13286	833.75	4225	1571	19082

R

- Adding Variables

```
67
68 ```{r calc-adrev-per-page}
69 magazines <- magazines %>%
70   mutate(PerPageRevenue = AdRevenue / AdPages)
71 ```
72
73
```

What does this do?

R

- Adding Variables

```
67
68 ```{r calc-adrev-per-page}
69 magazines <- magazines %>%
70   mutate(PerPageRevenue = AdRevenue / AdPages)
71 ```
72
73
```

What does this do?

R

- Adding Variables

What did this do?

```
73 ```{r newsstand-vs-subscription}
74 magazines <- magazines %>%
75   mutate(OneOff = NewsRevenue > SubRevenue)
76
77 pull(magazines, OneOff)
78 ```
```

```
  [1]  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
 [18]  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
 [35] FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE
 [52] FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE
 [69] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
 [86] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE
[103] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE
[120] FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE
[137] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
[154] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
[171] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE
[188] FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
```

```
79
```

R

- Try on your own!
- Lab is on the course website under "Labs" tab (https://amosca01.github.io/Intro-DS-F23/)
- Go to the "More Practice" section of the lab for today and complete as much as you can
- Ask questions when they come up!