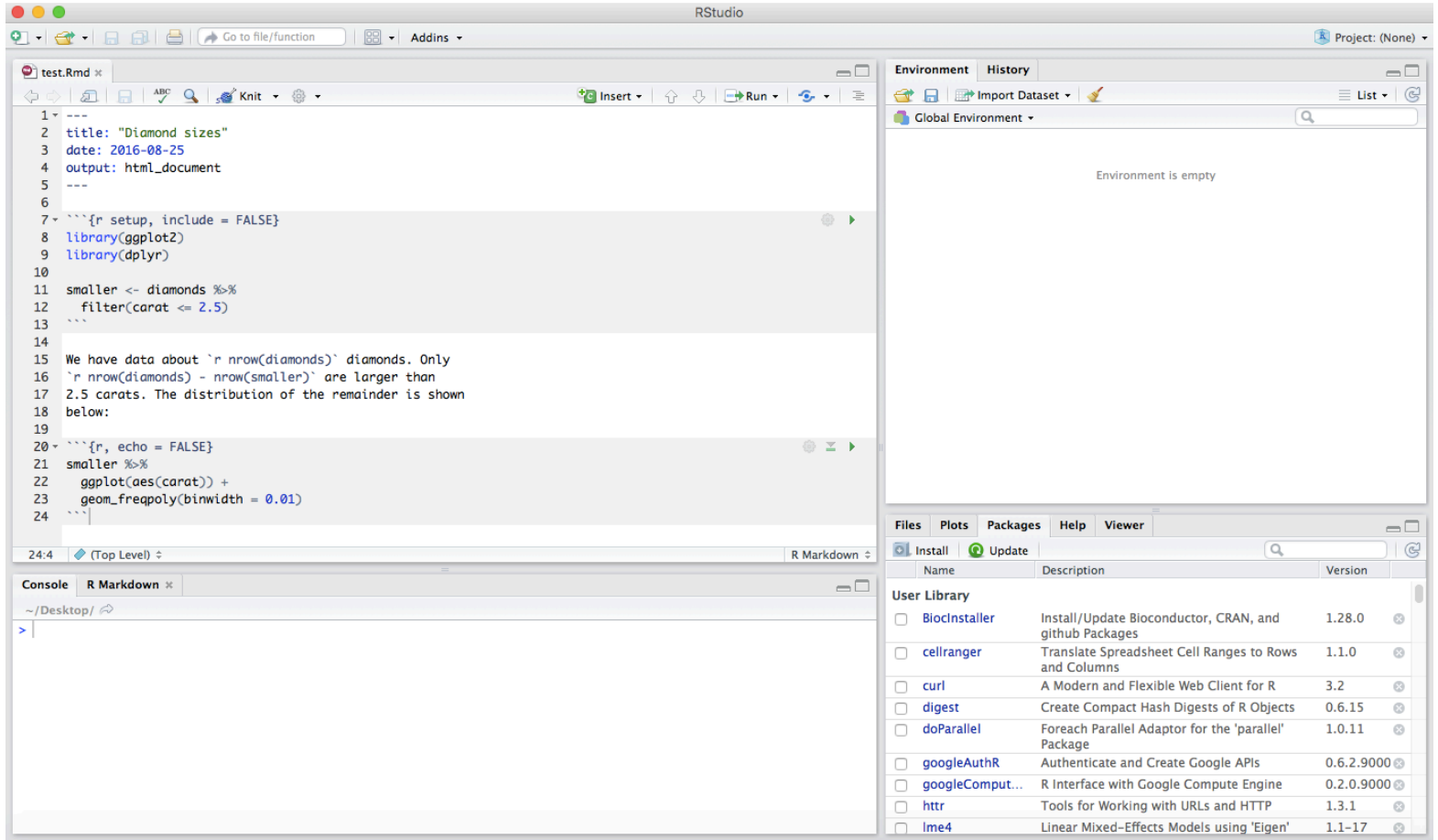
R Markdown & Import

Carolina A. de Lima Salge
Assistant Professor
Terry College of Business
University of Georgia

Business Intelligence
Spring 2021



Terry College of Business
UNIVERSITY OF GEORGIA

R Markdown

Provides a unified authoring framework for data science, combining your code, its results, and your comments.

- Reproducible
- Shareable
- Many output formats (PDF, Word, Slides)

R Markdown

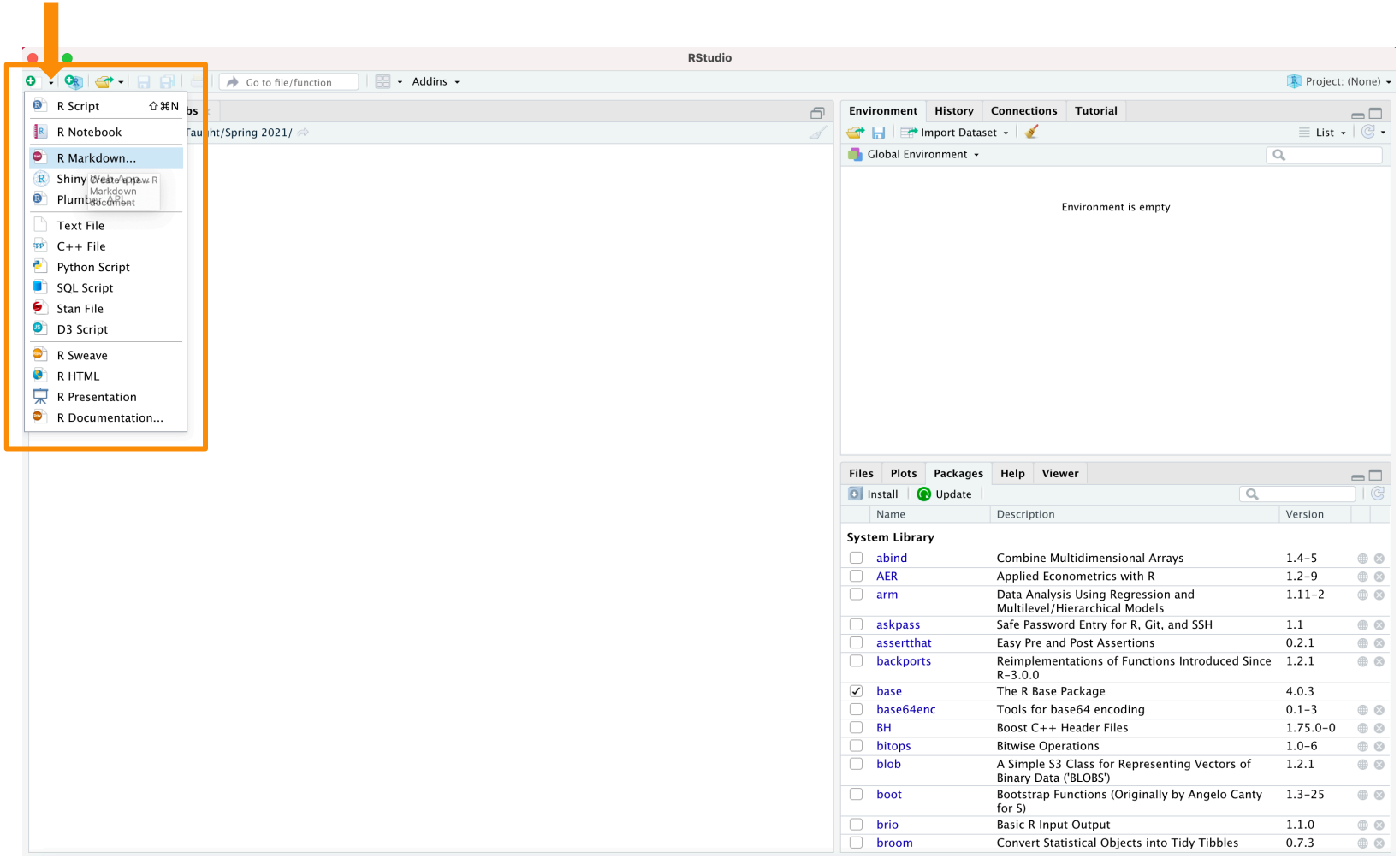
Why R Markdown?

1. For communicating to decision makers, who want to focus on the conclusions, not the code behind the analysis.

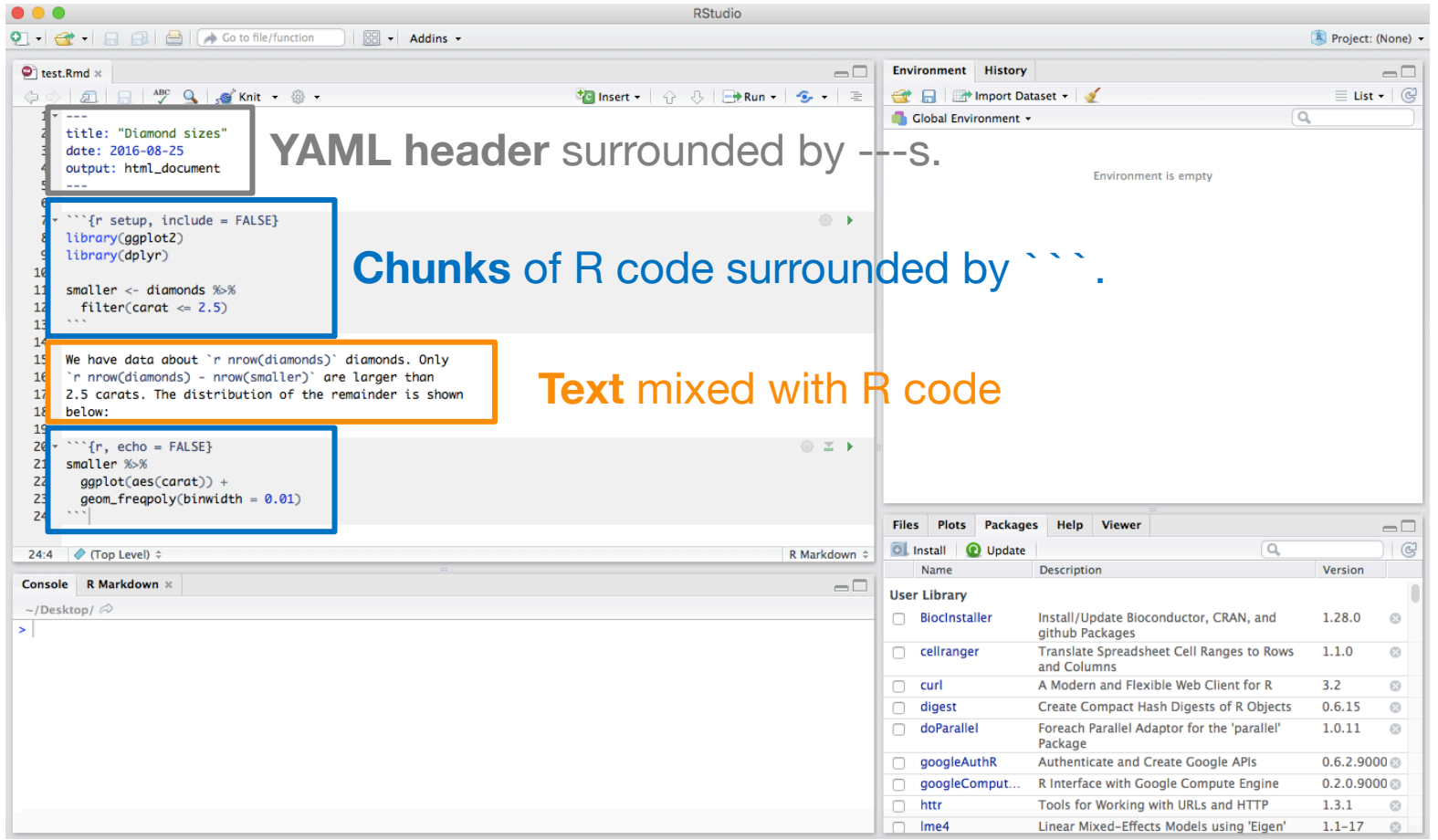
2. For collaborating with other data scientists (including future you!), who are interested in both your conclusions, and how you reached them (i.e. the code).

3. As an environment in which to *do* data science, as a modern day lab notebook where you can capture not only what you did, but also what you were thinking.

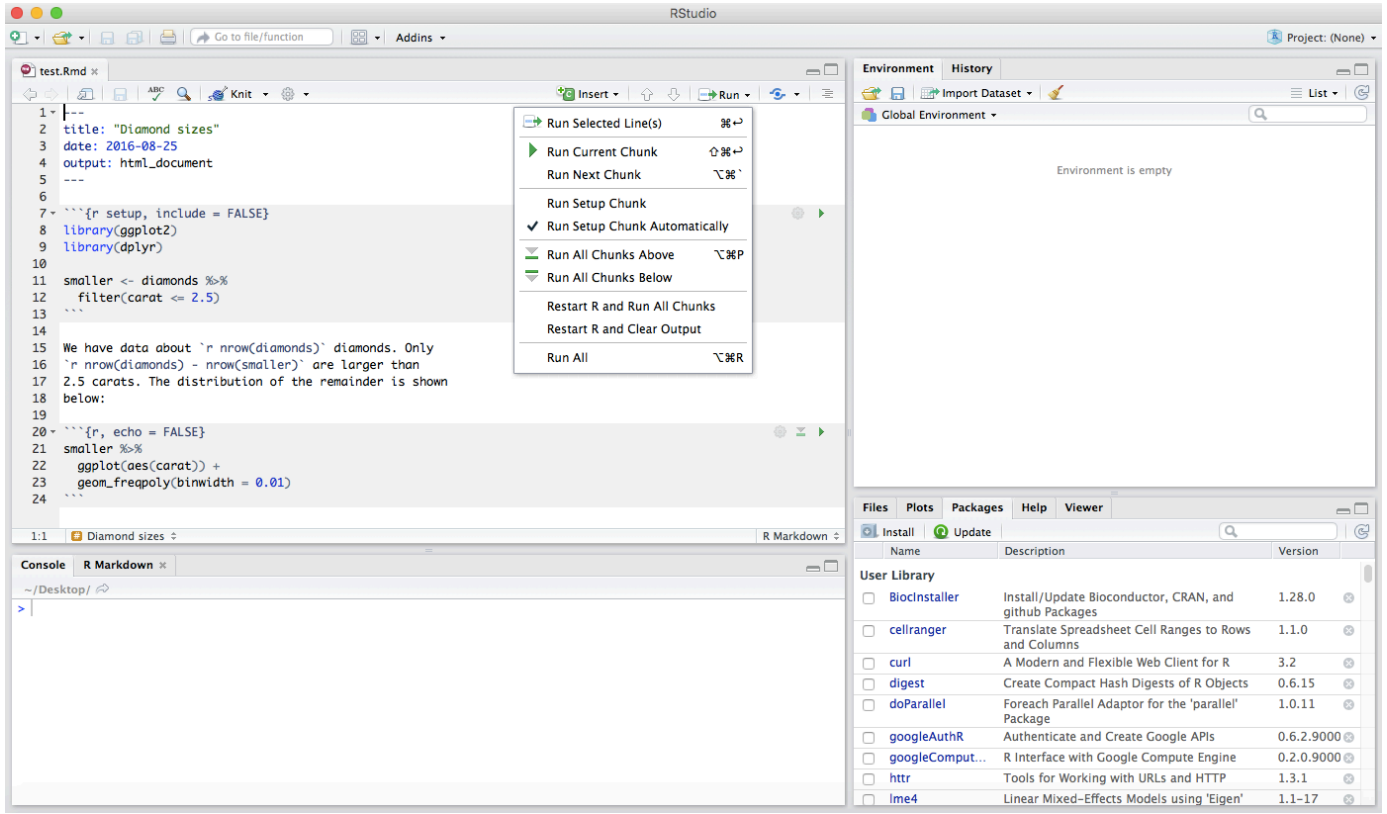
R Markdown in RStudio



R Markdown in Rstudio (Rmd file)



Run Code


RStudio
Addins
Project: (None)
test.Rmd
Knit
Insert
Run
---
title: "Diamond sizes"
date: 2016-08-25
output: html_document

```{r setup, include = FALSE}
library(ggplot2)
library(dplyr)
smaller <- diamonds %>%
filter(carat <= 2.5)
```
We have data about `r nrow(diamonds)` diamonds. Only
`r nrow(diamonds) - nrow(smaller)` are larger than
2.5 carats. The distribution of the remainder is shown
below:
```{r, echo = FALSE}
smaller %>%
ggplot(aes(carat)) +
geom_freqpoly(binwidth = 0.01)
```
Run Selected Line(s) ⌘↩
Run Current Chunk ⇧⌘↩
Run Next Chunk ⌥⌘`
Run Setup Chunk
Run Setup Chunk Automatically
Run All Chunks Above ⌥⌘P
Run All Chunks Below
Restart R and Run All Chunks
Restart R and Clear Output
Run All ⌥⌘R
Diamond sizes
R Markdown
Console
R Markdown
~/Desktop/
Environment
History
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Install
Update
Name
Description
Version
User Library
BiocInstaller
Install/Update Bioconductor, CRAN, and github Packages
1.28.0
cellranger
Translate Spreadsheet Cell Ranges to Rows and Columns
1.1.0
curl
A Modern and Flexible Web Client for R
3.2
digest
Create Compact Hash Digests of R Objects
0.6.15
doParallel
Foreach Parallel Adaptor for the 'parallel' Package
1.0.11
googleAuthR
Authenticate and Create Google APIs
0.6.2.9000
googleComput...
R Interface with Google Compute Engine
0.2.0.9000
httr
Tools for Working with URLs and HTTP
1.3.1
lme4
Linear Mixed-Effects Models using 'Eigen'
1.1-17

Knit

To create a report, click “Knit” or press Cmd/Ctrl + Shift + K

- Rmd file is sent to **knitr**, which executes the code and creates a new markdown (.md) document
- The markdown file is then processed by **pandoc**, which creates the finished file

HTML

```
---
title: "R Markdown"
author: "Carolina Alves de Lima Salge"
date: "1/25/2021"
output: html_document
---

```{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.
```

~/Desktop/Untitled.html

Untitled.html | Open in Browser | Find | Publish

R Markdown

Carolina Alves de Lima Salge

1/25/2021

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.

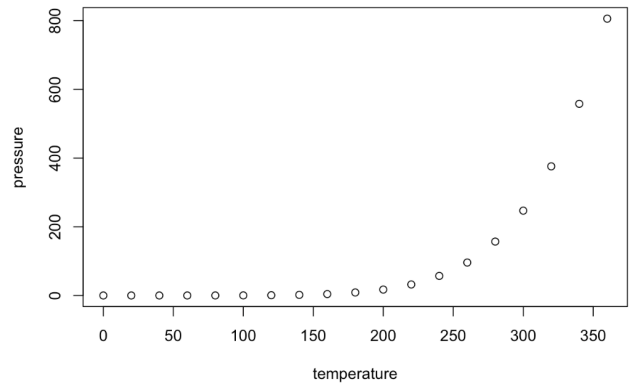
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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.

PDF

```
---
title: "R Markdown"
author: "Carolina Alves de Lima Salge"
date: "1/25/2021"
output: pdf_document
---

```{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.
```

R Markdown

Carolina Alves de Lima Salge

1/25/2021

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.

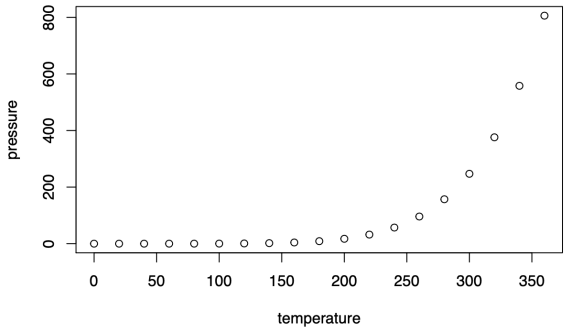
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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.

Word

```
---
title: "R Markdown"
author: "Carolina Alves de Lima Salge"
date: "1/25/2021"
output: word_document
---

```{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.
```

R Markdown

Carolina Alves de Lima Salge

1/25/2021

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.

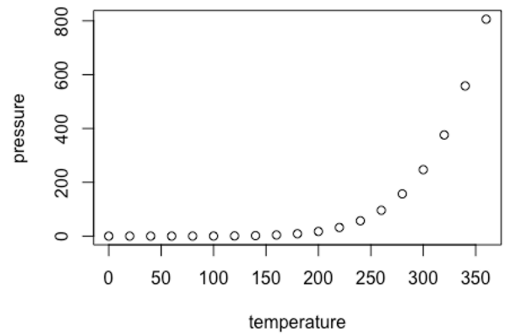
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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.

Data Import

Load flat files (e.g., csv) in R with the **readr** package, which is part of the core tidyverse

```
install.packages("tidyverse") # install package

library(tidyverse) # load already installed package
```

Getting Started

Turn flat files into tables (e.g., data frames) by using `read_csv()`

```
read_csv() # reads comma delimited files

read_csv2() # reads semicolon separated files
# common in countries where , is used as a decimal place

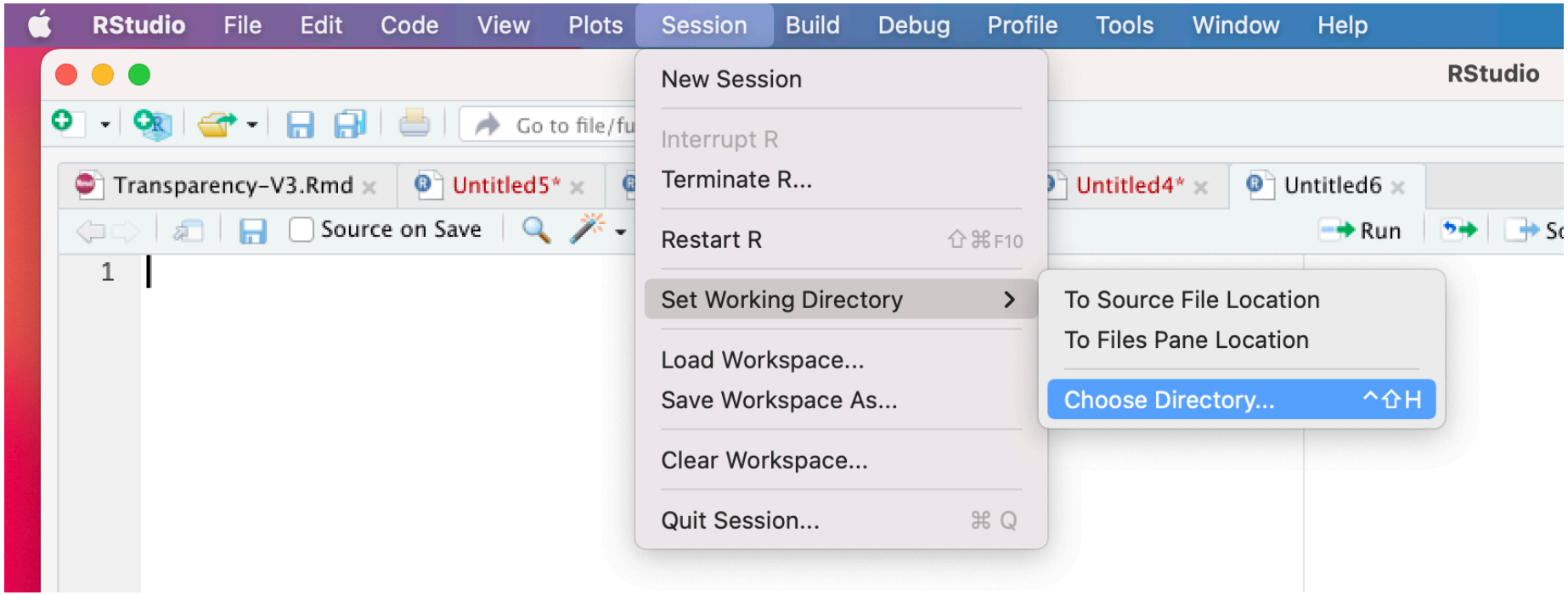
read_tsv() # reads tab delimited files

read_delim() # reads in files with any delimiter

# see http://r4ds.had.co.nz/data-import.html for more functions
```

Getting Started

Set a working directory, and dump your files in there



Import CSV files

Then use `read_csv()` to import the file of interest

```
CoffeeChain <- read_csv("CoffeeChain.csv")

── Column specification ────────────────────────────────────────────────────────
cols(
  Profit = col_double(),
  Margin = col_double(),
  Sales = col_double(),
  COGS = col_double(),
  TotExp = col_double(),
  Marketing = col_double(),
  Inventory = col_double(),
  BudgProf = col_double(),
  BudgMarg = col_double(),
  BudgSales = col_double(),
```

Save CSV files

Use `write_csv()` to save a file of interest. This function increases the chances of the output file being read back in correctly by:

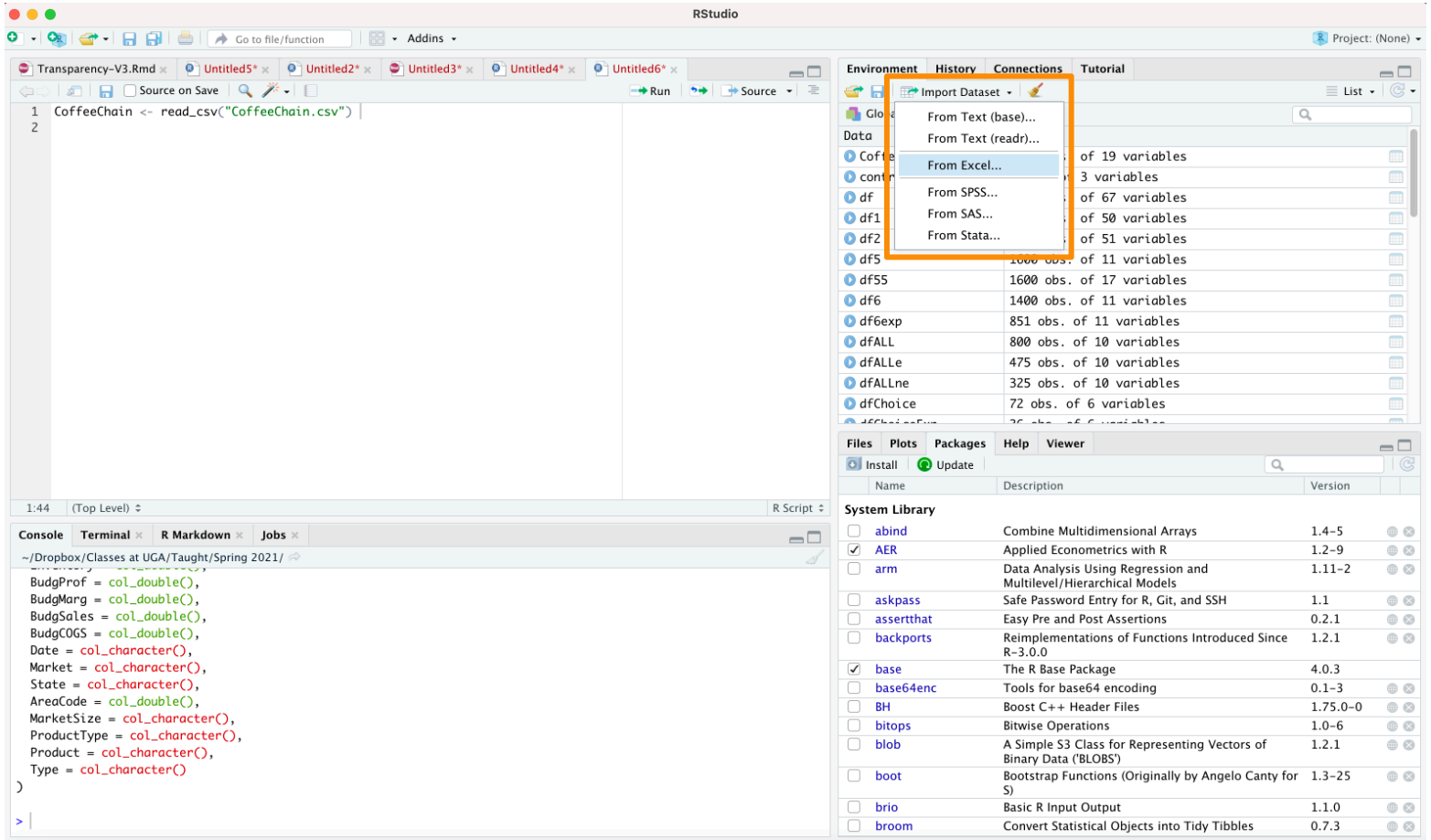
- Always encoding strings in UTF-8
- Saving dates and date-times in ISO8601 format

```
write_csv(CoffeeChain, "CoffeeChain.csv") # save csv file
```

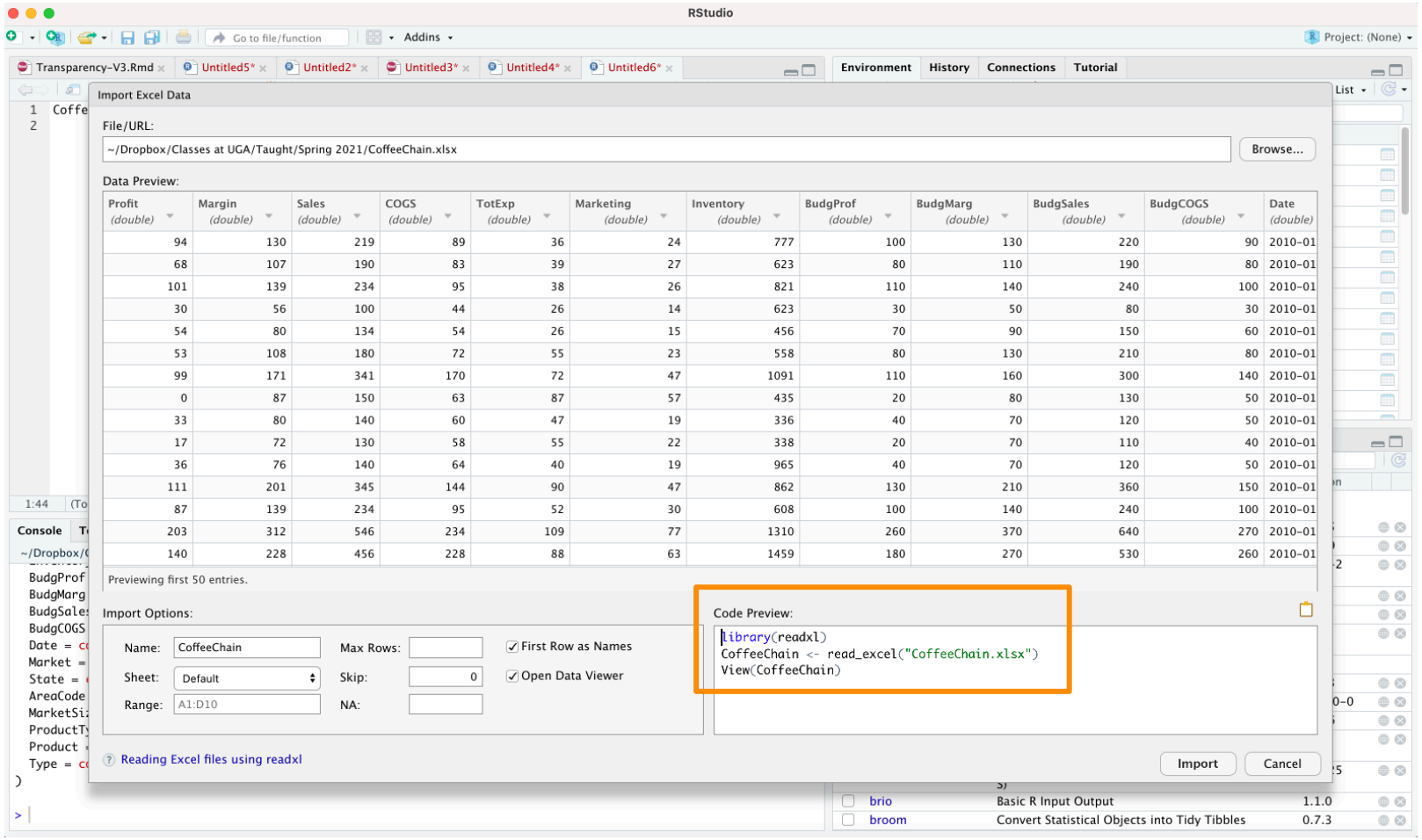
“This is all you need to know to read ~75% of CSV files”

Wickham & Grolemund in *R for data science*
http://r4ds.had.co.nz/data-import.html

Import Excel files



Import Excel files



Import Excel files

Then use `read_excel()` to import the file of interest

```
library(readxl)
CoffeeChain <- read_excel("CoffeeChain.xlsx")
View(CoffeeChain)

> library(readxl)
> CoffeeChain <- read_excel("CoffeeChain.xlsx")

> View(CoffeeChain)
```

Import from Database

Use **DBI**, along with a database specific package (e.g. **RMySQL**, **RSQLite**, **RPostgreSQL**)

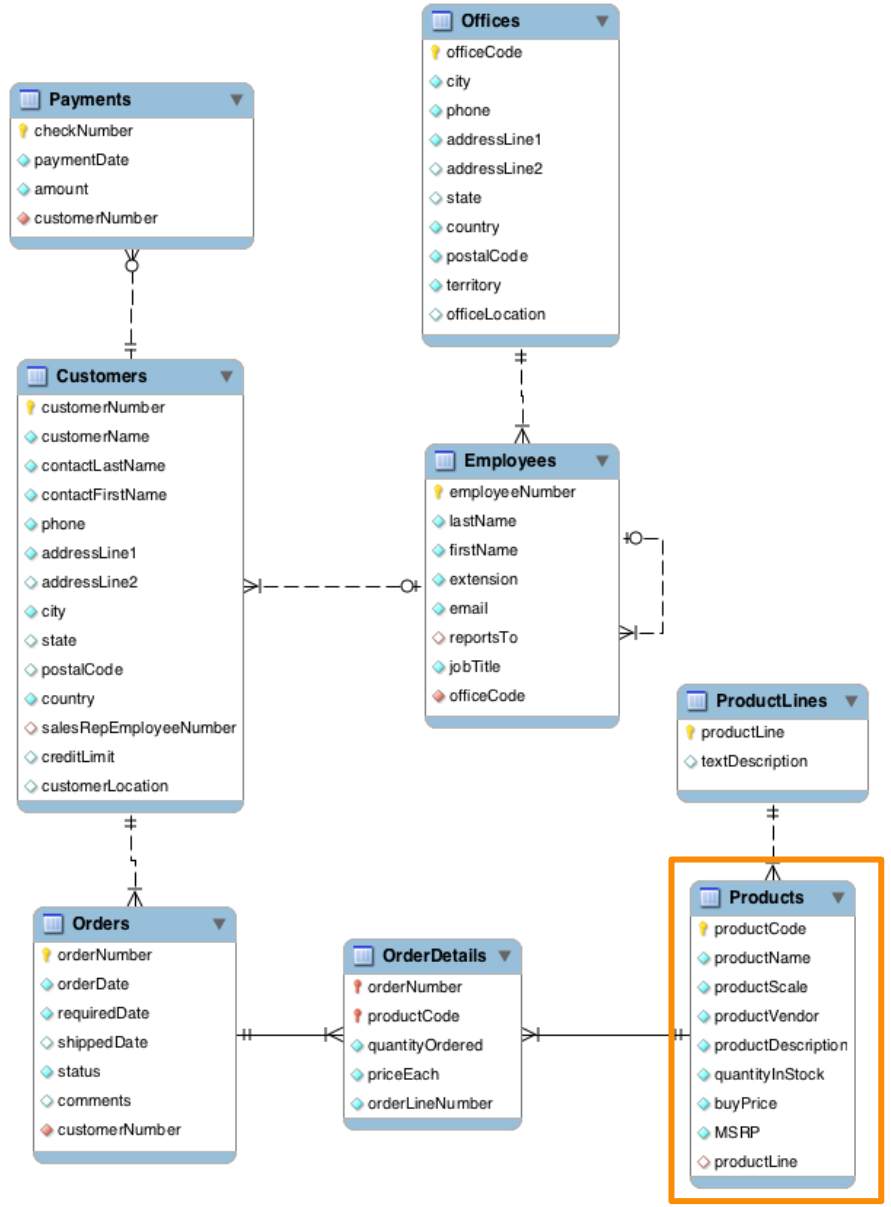
```
library(RMySQL)
library(DBI)
# Connect to Classic Models database
conn1 <- dbConnect(MySQL(),
       host= "",
       dbname= "",
       user= "",
       password= "")

# pull data from the database
products <- dbGetQuery(conn1, "select * from Products;")
```

Import from Database

```
library(RMySQL)
library(DBI)
# Connect to Classic Models database
conn1 <- dbConnect(MySQL(),
       host= "",
       dbname= "",
       user= "",
       password= "")

# Pull data from the database with SQL
products <- dbGetQuery(conn1, "select *
from Products;")
```



Other Types of Data

Start with the tidyverse list of packages below:

- **haven** reads SPSS, Stata, and SAS files
- **jsonlite** for json and **xml2** for XML

Exercise

Open RStudio and do the following:

- Install the **readxl**, **RMySQL**, and **DBI** packages
- Create a R Markdown file (see slide 5)
- Set a working directory (see slide 14)
- Download the CoffeeChain files from eLc and save them in your working directory folder

Exercise (cont.)

Open RStudio and do the following:

- Copy and paste the following code into one or many chunks of R code in R Markdown

```
library(tidyverse)
CoffeeChain <- read_csv("CoffeeChain.csv")
CoffeeChain # Print first 10 rows of the data

library(readxl)
CoffeeChain2 <- read_excel("CoffeeChain.xlsx")
CoffeeChain2

library(RMySQL)
library(DBI)
# Connect to Classic Models database
conn1 <- dbConnect(MySQL(),
        host= "",
```

Exercise (cont.)

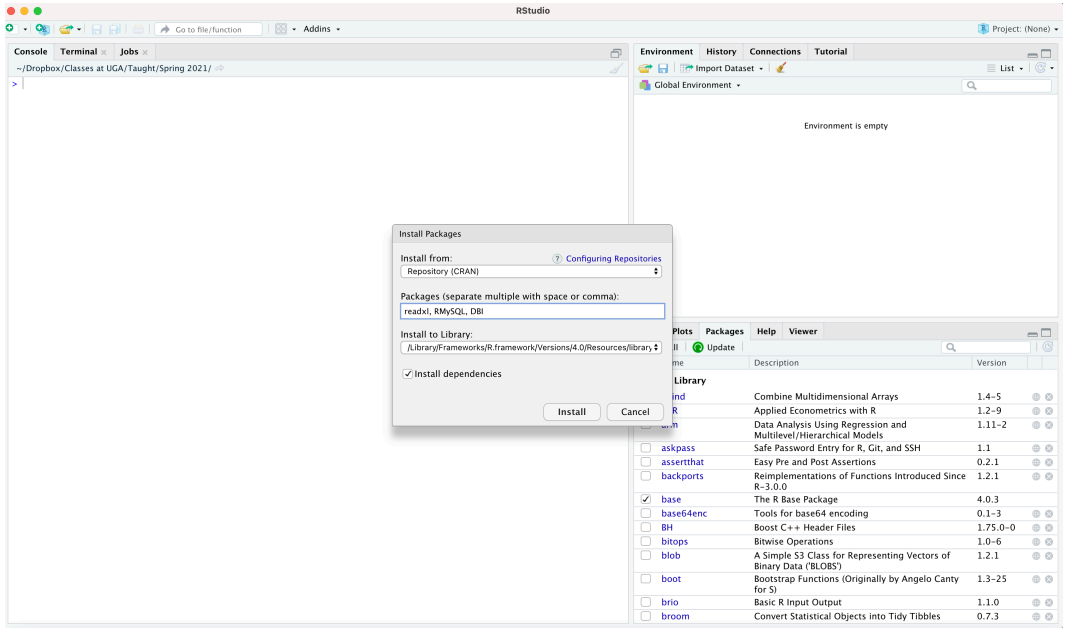
Open RStudio and do the following:

- Change {r setup, include=FALSE} to {r setup, include=TRUE} as well as Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot. to Note that the `echo = TRUE` parameter was added to the code chunk to print the R code that generated the results.
- Knit the file in your saved directory to PDF or Word

Exercise (visuals)

Open RStudio and do the following:

- Install the **readxl**, **RMySQL**, and **DBI** packages



Exercise (visuals cont.)

Open RStudio and do the following:

- Copy and paste the following code into one or many chunks of R code in R Markdown

Exercise (visuals cont.)

Open RStudio and do the following:

- Knit the file in your saved directory to PDF or Word

Thank You!