



Introduction to Tidyverse/R

Data Science Institute
January 7, 2020

Giang Rudderham
Data Science Consultant, ITS Research Services
The University of Iowa



Welcome to Introduction to Tidyverse/R

Data Science Institute
January 7, 2020

Please sign in with your HawkID or ID card using the iPad.

Giang Rudderham
Data Science Consultant, ITS Research Services
The University of Iowa

Before the Workshop

- ▶ Please make sure that you installed R, RStudio, the `tidyverse` package, and the `dslabs` package on your laptop.
- ▶ Please sign in with your HawkID or ID card using the iPad.
- ▶ Download materials from <https://github.com/grudderham/dsi-2020-01>

About This Workshop

- ▶ Intended for complete beginners in R
- ▶ Tidyverse and base R
- ▶ Introduction to data transformation, data visualization, exploratory data analysis
- ▶ Get you started in R so that you can continue learning after this workshop

About Me

- BA, Mathematics, College of Wooster (Ohio)
- MS, Statistics, University of Iowa
 - Started learning R
 - Found it really hard
- Currently support data science in ITS Research Services
 - Support the new [Interactive Data Analytics Service \(IDAS\)](#)



Workshop Outline

- ▶ Basic workflow in R
- ▶ Dataset 1: NYC Regents Exams Scores 2010
- ▶ Dataset 2: Gapminder Data

Please follow along using RStudio on your laptop

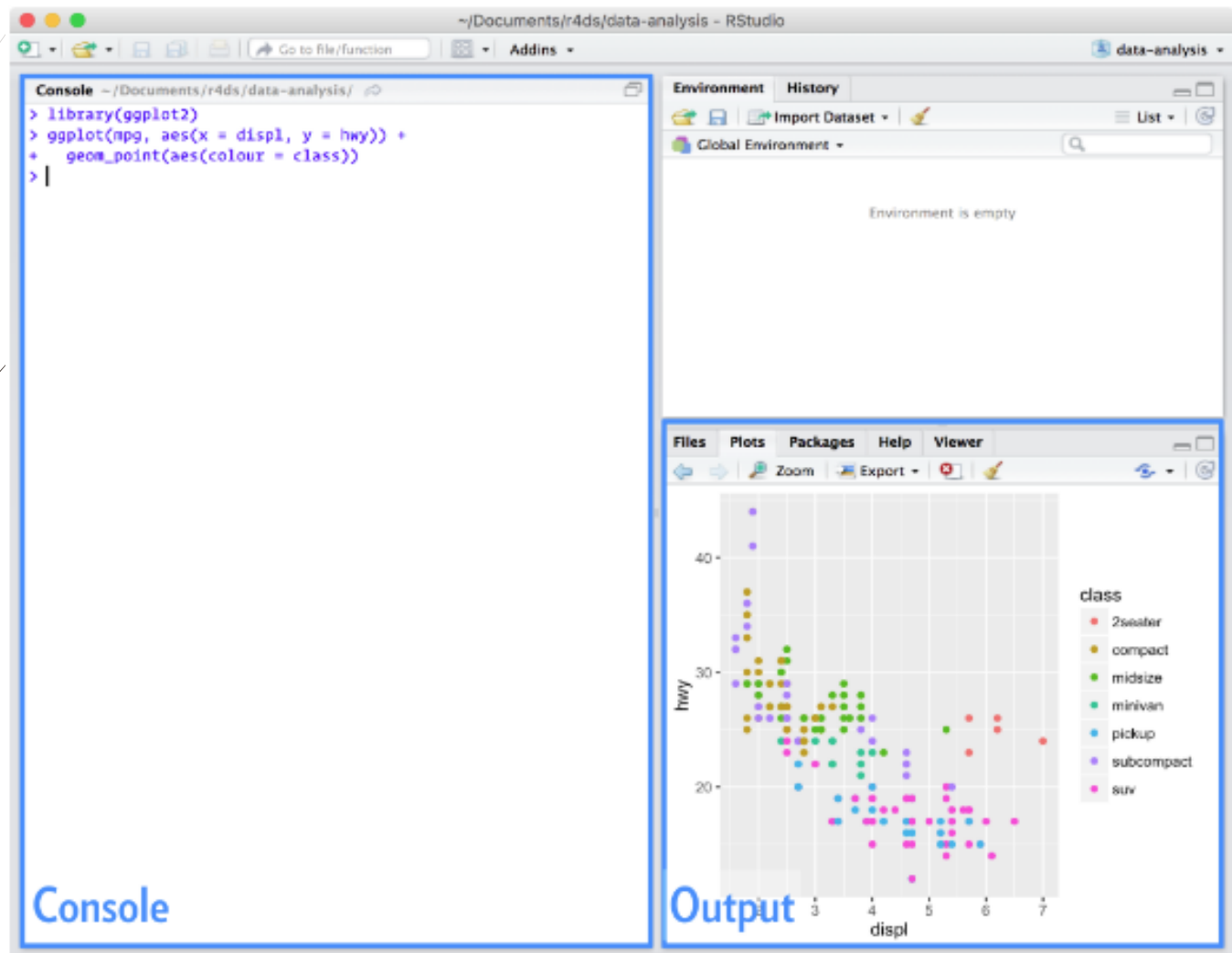
The Tools

- ▶ What is R?
 - ▶ Open source!
 - ▶ For statistical computing and graphics
- ▶ What is RStudio?
 - ▶ Integrated development environment (IDE) for R programming
- ▶ Base R and the tidyverse

Workflow Basics in R

Let's launch RStudio

Layout of RStudio



Let's try typing in the Console

- ▶ R can be used as a calculator. What do the following return at the Console?

```
20 * 4  
(100 + 50) / 2  
2 ^ 2  
pi
```

- ▶ To create an object, we use assignment statements.

General form: `object_name <- value`

```
x <- 3  
y <- "Hi! How are you?"
```

- ▶ To inspect the objects we just created:

```
x  
y
```

- ▶ What's the difference between `(x <- 3)` and `x <- 3`?

Please follow along using RStudio on your laptop

Object Names in R

- ▶ Must start with a letter
- ▶ Can only contain letters, numbers, `_` and `.`

```
i_use_snake_case  
otherPeopleUseCamelCase  
some.people.use.periods  
And_aFew.People_RENOUNCEconvention
```

Source: R for Data Science

- ▶ Let's try at the Console:

```
my_very_very_long_name <- 18
```

1. Then, start typing `my_` at the Console. What happens?
2. How do I change `my_very_very_long_name` to 20?
3. What if I type `my_very_long_name`?

Calling Functions in R

- ▶ Call built-in functions in R:

```
function_name(arg1 = val1, arg2 = val2, ...)
```

- ▶ Let's try at the Console:

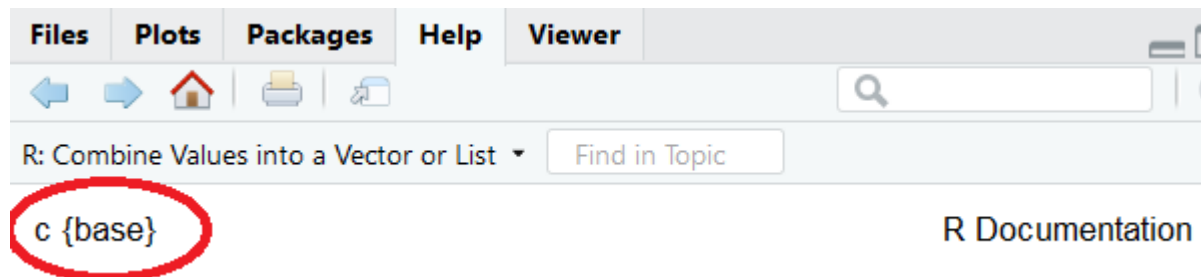
1. What does `rep(1, times = 3)` return?
2. What do you think the `rep` function do?

- ▶ We can use a function to create an object:

3. What does this return? `(z <- c(1, 3))`
4. What does “c” do? Try `?c` at the Console

Help

- Help is in the lower right pane in RStudio.
- It tells you the name of the function and the package.



Combine Values into a Vector or List

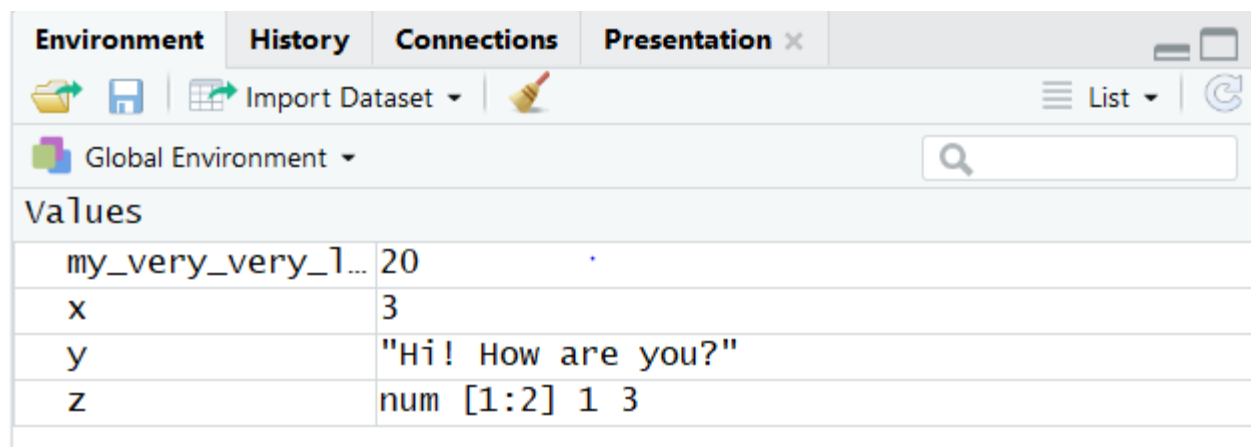
Description

This is a generic function which combines its arguments.

- Scroll down to the Examples section. It's usually helpful.
- I also Google a lot!

Environment

- Environment is in the upper right pane.
- By this time yours probably look like this:



The screenshot shows the R Studio Environment pane. At the top, there are tabs for 'Environment', 'History', 'Connections', and 'Presentation'. Below the tabs is a toolbar with icons for file operations and a search bar. The main area is titled 'Global Environment' and contains a table of variables.

Values	
my_very_very_1...	20
x	3
y	"Hi! How are you?"
z	num [1:2] 1 3

1. What does `rm(x)` do? Try `?rm` at the Console.
2. How do I remove both `y` and `z`?
3. Type `another_var <- 37` and then `rm(list = ls())`. What does the second command do?
Be careful! There was no warning from R!

One more thing before we work with datasets

- ▶ Coding standards are useful.

- ▶ These are super hard to read:

```
x<-1/200*30^5
```

```
z<-runif(3,0,1)
```

- ▶ Instead, do these:

```
x <- 1 / 200 * 30 ^ 5
```

```
z <- runif(3, 0, 1)
```

- ▶ By the way, what does `runif` do? Try `?runif` at the Console.

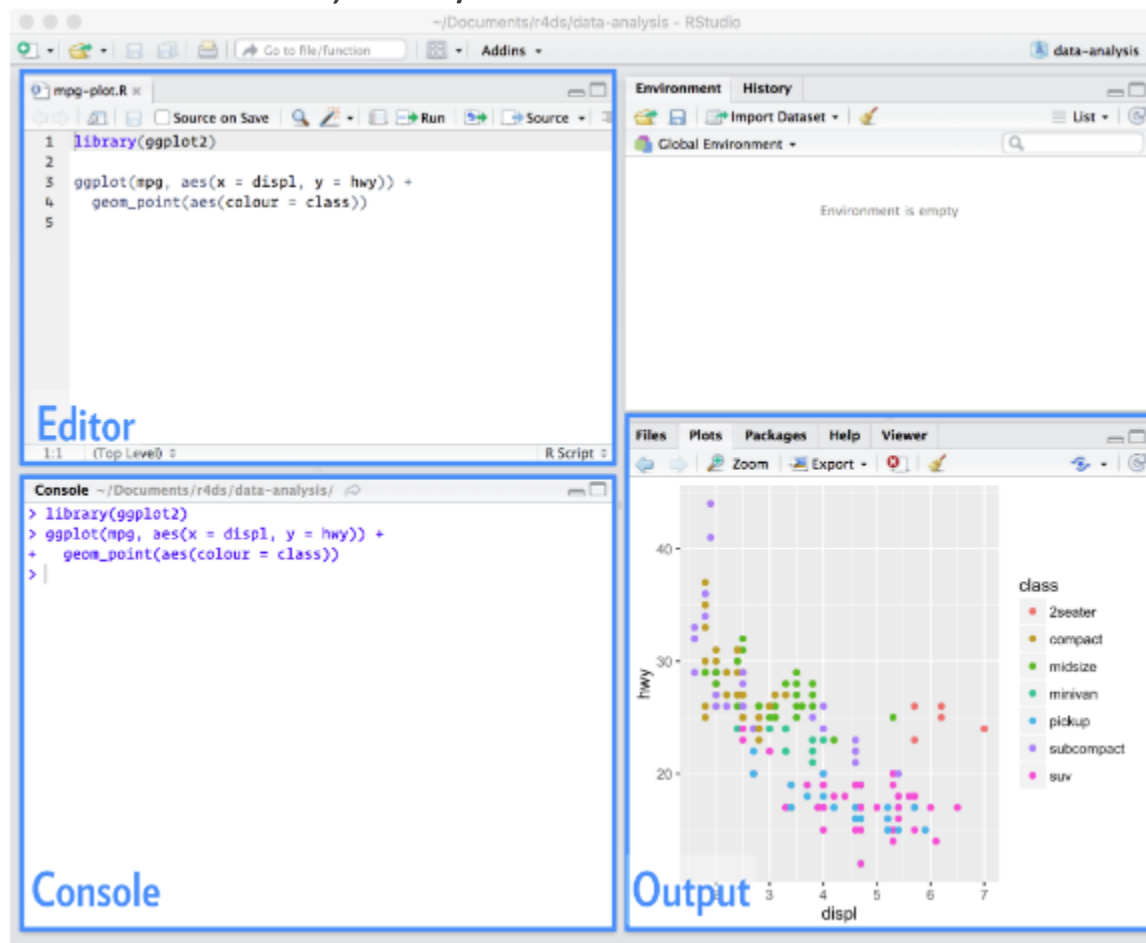
- ▶ Its cousins: `rbinom`, `rnorm`, `rchisq`, ...

Dataset 1: NYC Regents Exams Scores 2010

Let's open the R script file # 1

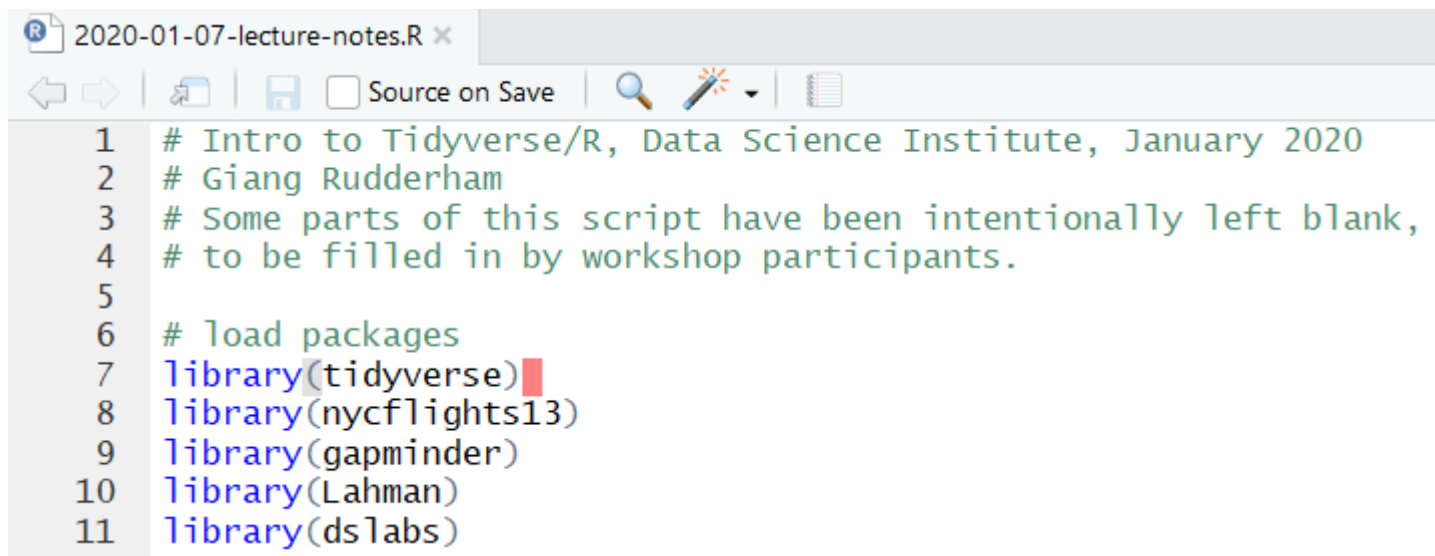
The Script Editor

- To open a script file, double click on the .R file
- Or in RStudio, Cmd/Ctrl + O



The Script Editor

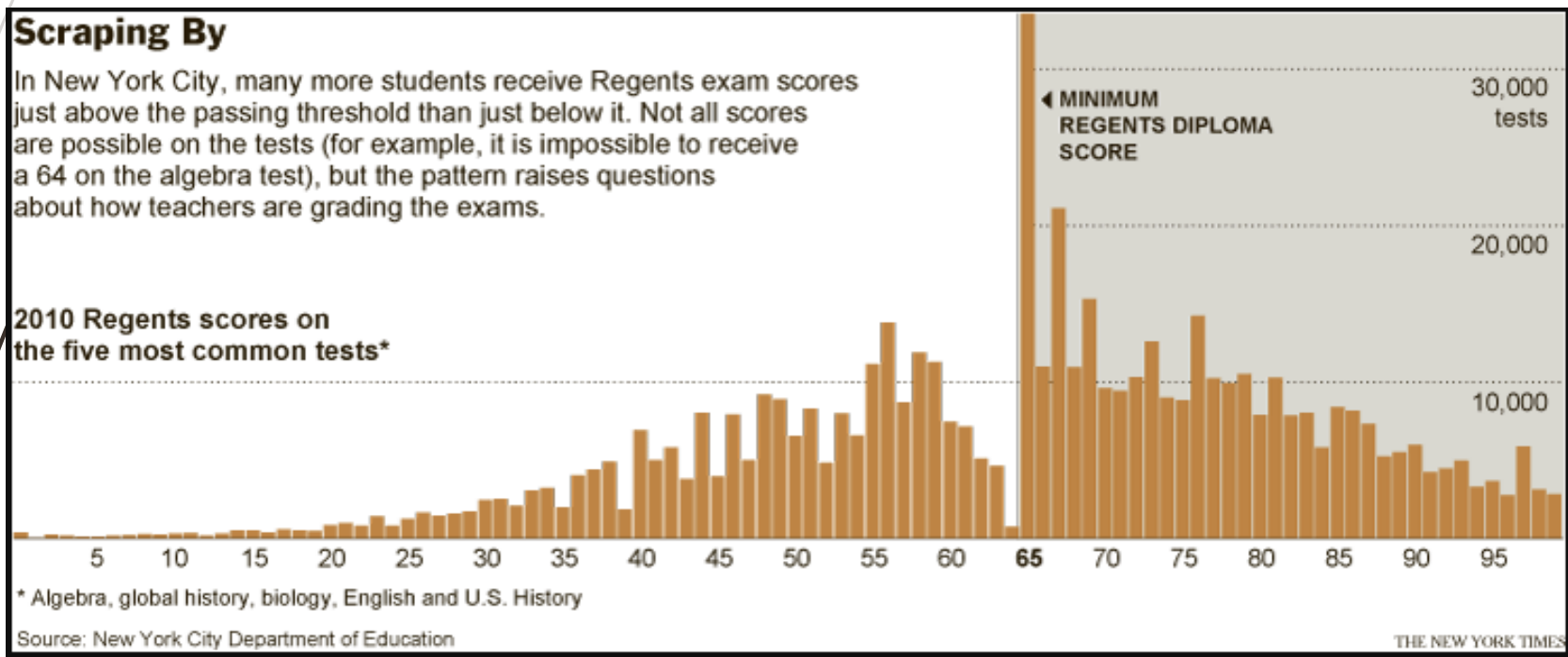
- ▶ Always use an R script file to organize code.
- ▶ To execute a line: Cmd / Ctrl + Enter
 - ▶ R will run the line where my red cursor is.
- ▶ Load packages at the top of the script file
- ▶ Use a lot of comments.
 - ▶ Comments start with # and are in green.



```
2020-01-07-lecture-notes.R x
Source on Save
1 # Intro to Tidyverse/R, Data Science Institute, January 2020
2 # Giang Rudderham
3 # Some parts of this script have been intentionally left blank,
4 # to be filled in by workshop participants.
5
6 # load packages
7 library(tidyverse)
8 library(nycflights13)
9 library(gapminder)
10 library(Lahman)
11 library(dslabs)
```

The NYC Regents exams scores 2010 dataset

- ▶ The dataset was used to make the plot below.
- ▶ What do you think is the story here?



Source: Figure is from The New York Times

The NYC Regents exams scores 2010 dataset

- ▶ We'll continue by working with the R script file.

Dataset 2: Gapminder Data

Let's open the R script file # 2

Gapminder Data

- ▶ We'll continue by working with the R script file.

Conclusions

- R is fun and very powerful!
- Use script files (.R) to organize your code.
- Use comments in your .R files
- Follow coding standards as best as we can
- We can learn a lot from reading code written by others.

Thank you for your time!

- ▶ Questions specific to this workshop:
 - ▶ giang-rudderham@uiowa.edu
- ▶ Questions about using computing in research, including
 - ▶ [Storage](#)
 - ▶ [High Performance Computing](#)
 - ▶ [Interactive Data Analytics Service \(IDAS\)](#)
 - ▶ Please write to: research-computing@uiowa.edu

Resources

Base R

- ▶ Book: “Cookbook for R”
 - ▶ <http://www.cookbook-r.com/>
- ▶ Learn R at the Console (interactively): the `swirl` package
 - ▶ <https://swirlstats.com/>

Tidyverse

- ▶ Book: “R for Data Science”
 - ▶ <https://r4ds.had.co.nz/>
- ▶ Article: “The Layered Grammar of Graphics”
 - ▶ <http://vita.had.co.nz/papers/layered-grammar.pdf>

Resources

- ▶ `dplyr` reference
 - ▶ <https://dplyr.tidyverse.org/reference/index.html>
- ▶ `ggplot2` reference
 - ▶ <https://ggplot2.tidyverse.org/reference/index.html>
- ▶ Cheat sheets from RStudio
 - ▶ <https://rstudio.com/resources/cheatsheets/>
- ▶ Color brewer, for plotting
 - ▶ <http://colorbrewer2.org/#type=sequential&scheme=YlGnBu&n=4>

Resources: learn R on campus

- ▶ Workshops in Python, HPC, and R hosted by ITS Research Services
 - ▶ <https://hpc.uiowa.edu/events>
- ▶ Workshops hosted by the Iowa Social Science Research Center (ISRC), including R and other topics
 - ▶ <http://ppc.uiowa.edu/isrc/workshops>