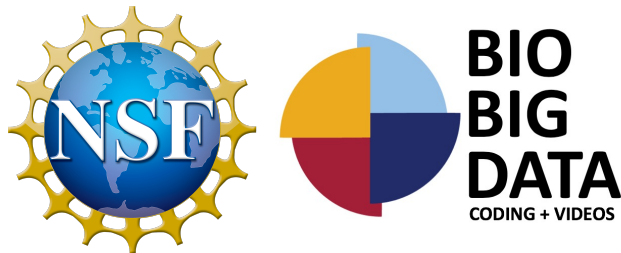


# R Data Science Coding Bootcamp



# Before we start

- Anonymous Pre-survey



# What is R?

- It is a popular language for data analysis.
- [https://en.wikipedia.org/wiki/R\\_\(programming\\_language\)](https://en.wikipedia.org/wiki/R_(programming_language))
- Open source

# Install R

<https://cran.r-project.org/>



CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

## The Comprehensive R Archive Network

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

# Install RStudio

 [rstudio.com/products/rstudio/download/](https://rstudio.com/products/rstudio/download/)



Products ▾ Solutions ▾ Customers Resources

DOWNLOAD SUP

Download the RStudio IDE

RStudio Desktop

Open Source License

**Free**

**DOWNLOAD**

[Learn more](#)

Integrated Tools for R



# Rstudio interface

The screenshot displays the RStudio interface with three main panels. The left panel shows R code for loading and analyzing data. The top-right panel shows the Environment pane with a table of variables. The bottom-right panel shows the Files pane with a list of files in the project directory. The bottom-left panel shows the Console with the output of the R code.

**CODE**

```
1 ---
2 title: "Analysis of survey data for metrics, scientific literacy and attitude"
3 author: "H Qin"
4 date: "5/30/2019"
5 output:
6   html_document: default
7   pdf_document: default
8   layout: topic
9 ---
10
11 -----
12
13 > ## Learning Objectives
14 >
15 > * Load external tabular data from a .csv file into R.
16 > * Describe what an R data frame is.
17 > * Summarize the contents of a data frame in R.
18 > * Manipulate categorical data in R using factors.
19
20 -----
21
22 # Check files in the working directory
23
24 ```{r list, echo=FALSE}
25 rm(list=ls()) #clean up workspace
26
```

**Environment**

Variable	Class
m1	List of 12
m2	List of 12
m3	List of 13
m4	List of 12
tb	316 obs. of 7 variables
tb3	269 obs. of 7 variables

**Files**

Name	Size	Modified
..		
.RData	17.3 KB	May 13, 2020, 11:25 PM
.Rhistory	13.7 KB	May 13, 2020, 11:25 PM
Learning_R_by_metricExample.ppt	16 MB	May 13, 2020, 11:33 PM
learningR_through_metric_example.ht...	1.9 MB	May 13, 2020, 11:25 PM
learningR_through_metric_example.pdf	344.4 KB	May 13, 2020, 11:25 PM
learningR_through_metric_example.R...	5.9 KB	May 13, 2020, 11:25 PM
metric-attitude-literacy.csv	17.1 KB	May 13, 2020, 11:25 PM
metric_survey_form.pdf	370.5 KB	May 13, 2020, 11:25 PM
simpleR.html	1.1 MB	May 13, 2020, 11:25 PM
simpleR.Rmd	1.9 KB	May 13, 2020, 11:25 PM
~\$Learning_R_by_metricExample.ppt	165 B	May 13, 2020, 11:33 PM

**Running**

```
tb3$age                2.985  0.00311 **
tb3$genderFemale       -0.760  0.44791
tb3$genderMale         -0.333  0.73966
tb3$country             0.706  0.48094
tb3$degreeBachelor Degree in Science or equivalent  0.464  0.64277
tb3$degreeHigh School or equivalent -0.817  0.41447
tb3$degreeM.D. or equivalent  1.090  0.27674
tb3$degreeMaster Degree or equivalent  0.738  0.46099
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9451 on 259 degrees of freedom
Multiple R-squared:  0.2354,    Adjusted R-squared:  0.2089
F-statistic: 8.861 on 9 and 259 DF,  p-value: 1.249e-11

> #age is still signicant after PhD are removed from the sample
>
> |
```

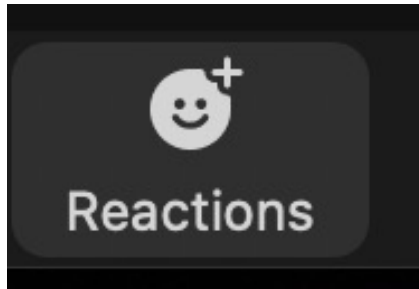
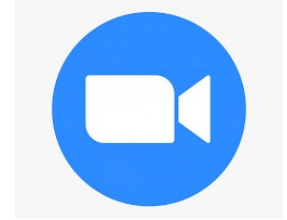
**Variables**

**Files, Plot, Help**

# Participatory Live Coding



# Asking for helps on ZOOM



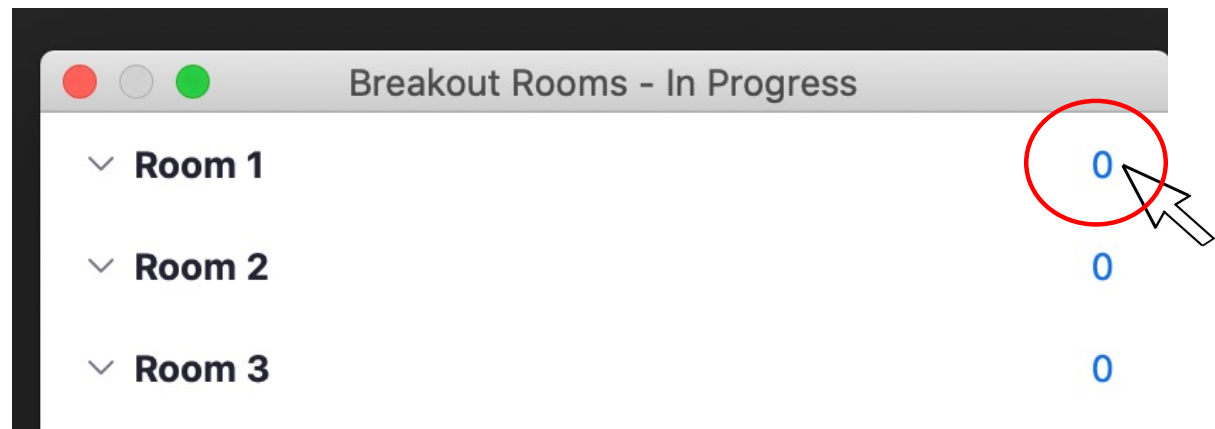
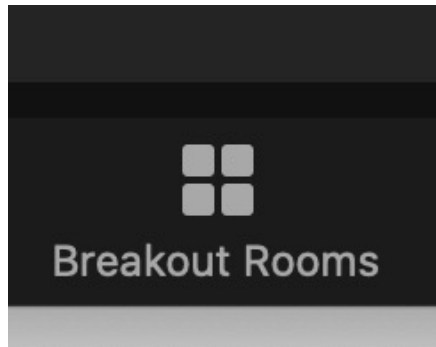
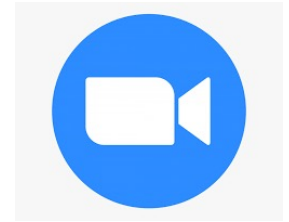
Ziwei Ma <ziweinmsu@gmail.com>



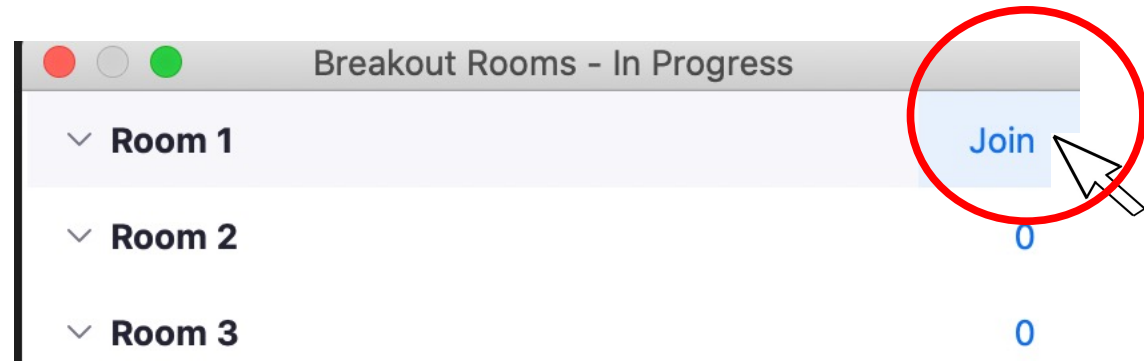
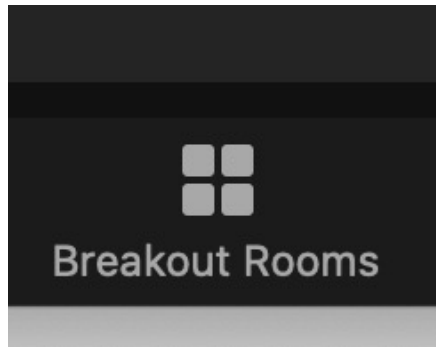
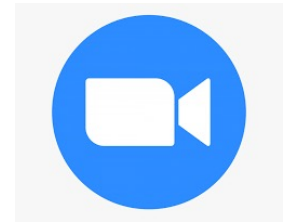
**Raise Hand**



# Joining Breakout Rooms in ZOOM



# Joining Breakout Rooms in ZOOM



# R markdown file

```
▼ # Check files in the working directory
```

```
▼ ```{r list, echo=FALSE}  
rm( list=ls()) #clean up worksapce  
list.files()  
```
```



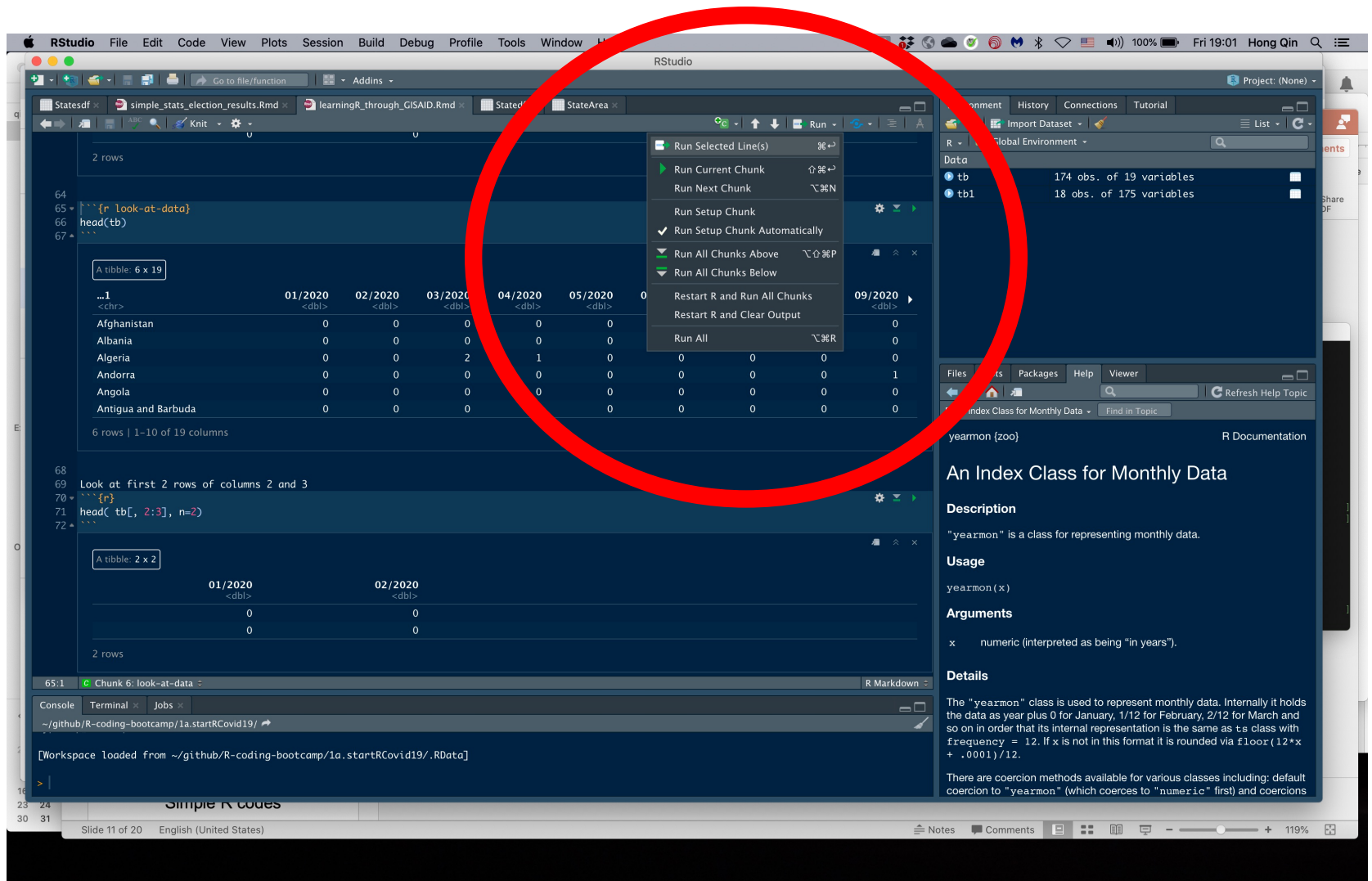
```
[1] "~$Learning_R_by_metricExample.ppt"  
[2] "Learning_R_by_metricExample.ppt"  
[3] "learningR_through_metric_example.html"  
[4] "learningR_through_metric_example.pdf"  
[5] "learningR_through_metric_example.Rmd"  
[6] "metric_survey_form.pdf"  
[7] "metric-attitude-literacy.csv"  
[8] "simpleR.html"
```

**Comment  
Chunks**

**Code  
Chunks**

**Results**

# Run selected lines



The screenshot shows the RStudio interface with a red circle highlighting the 'Run' menu. The menu options are:

- Run Selected Line(s)
- Run Current Chunk
- Run Next Chunk
- Run Setup Chunk
- Run Setup Chunk Automatically
- Run All Chunks Above
- Run All Chunks Below
- Restart R and Run All Chunks
- Restart R and Clear Output
- Run All

The background shows two R code chunks and their outputs.

**Chunk 6: look-at-data**

```
64 {r look-at-data}
65 head(tb)
67
```

A tibble: 6 x 19

|                     | 01/2020 | 02/2020 | 03/2020 | 04/2020 | 05/2020 | 06/2020 | 07/2020 | 08/2020 | 09/2020 |
|---------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Afghanistan         | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       |
| Albania             | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       |
| Algeria             | 0       | 0       | 2       | 1       | 0       | 0       | 0       | 0       | 0       |
| Andorra             | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 1       |
| Angola              | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       |
| Antigua and Barbuda | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       |

6 rows | 1-10 of 19 columns

**Chunk 7: Look at first 2 rows of columns 2 and 3**

```
68 Look at first 2 rows of columns 2 and 3
69 {r}
70 head( tb[, 2:3], n=2)
71
72
```

A tibble: 2 x 2

|  | 01/2020 | 02/2020 |
|--|---------|---------|
|  | 0       | 0       |
|  | 0       | 0       |

2 rows

**Console**

```
~/github/R-coding-bootcamp/1a.startRCovid19/
[Workspace loaded from ~/github/R-coding-bootcamp/1a.startRCovid19/.RData]
```

**R Documentation**

## An Index Class for Monthly Data

**Description**

"yearmon" is a class for representing monthly data.

**Usage**

```
yearmon(x)
```

**Arguments**

x numeric (interpreted as being "in years").

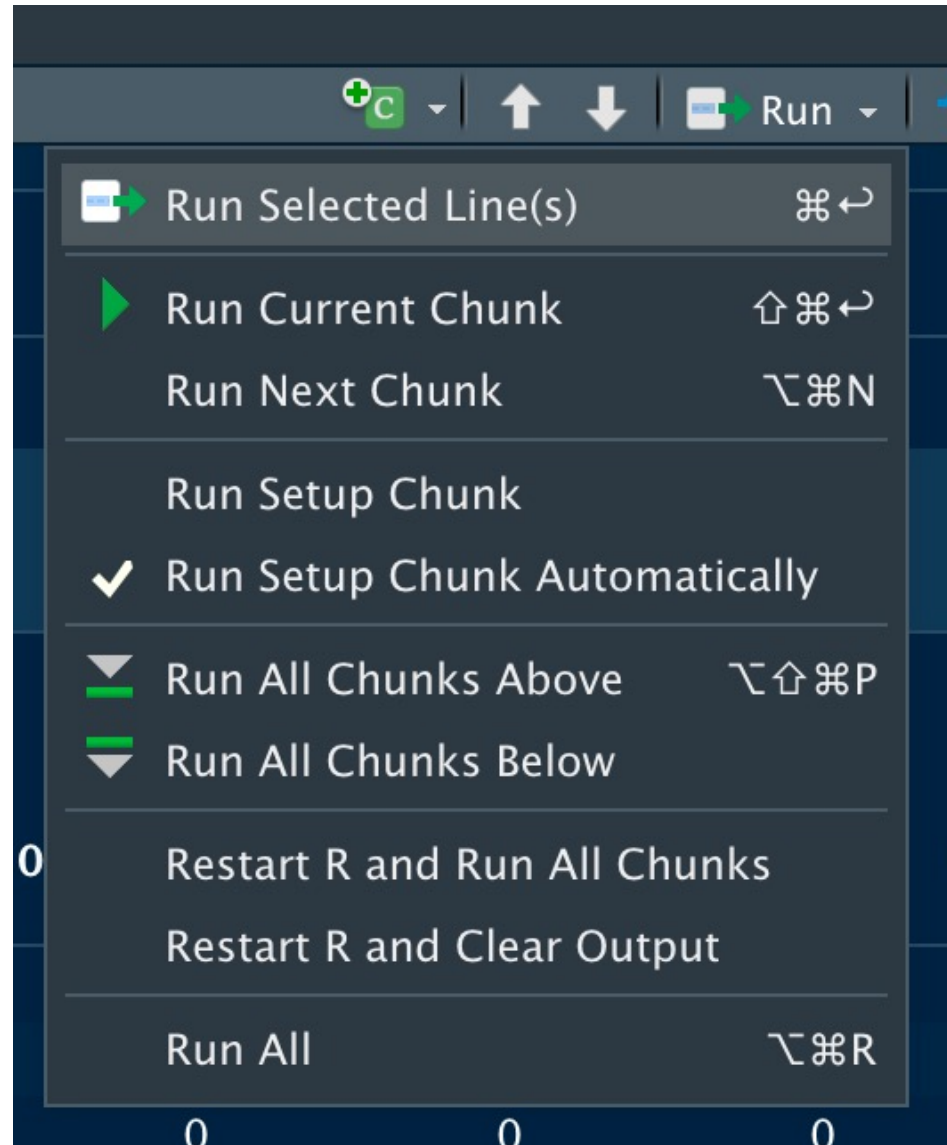
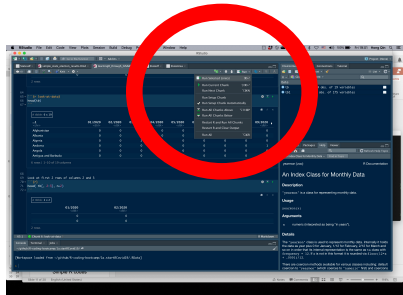
**Details**

The "yearmon" class is used to represent monthly data. Internally it holds the data as year plus 0 for January, 1/12 for February, 2/12 for March and so on in order that its internal representation is the same as ts class with frequency = 12. If x is not in this format it is rounded via floor(12\*x + .0001)/12.

There are coercion methods available for various classes including: default coercion to "yearmon" (which coerces to "numeric" first) and coercions

Slide 11 of 20 English (United States) 119%

# Run selected lines/ chunks



# Editing: Inset a code chunk

