



Flexible R code for Assessment Practitioners

Laura Lambert
James Madison University

Overview

- Who are you?
- What will we be doing for the next hour?
- Will I become an R expert?
- Is this going to be scary?

Overview

What are

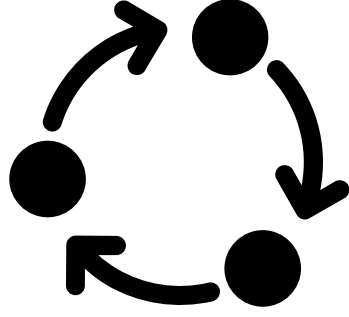
Who I am not

- An Expert Coder
- A computer science major
- Someone who has received formal training in how to ‘properly’ code

Who I am

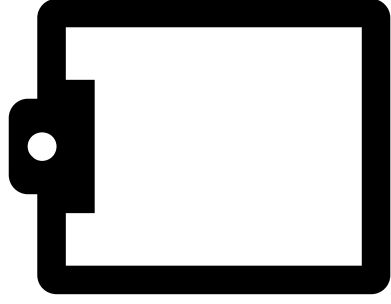
- Doctoral candidate in James Madison University's Assessment & Measurement program
- Full-time A&P faculty in Integrated Science and Technology (also at JMU)
- Assessment (and measurement) enthusiast
- Really good at crafting Google search terms
- Someone who enjoys using available tools to make my life easier!

Overview



Contextualize this in the Assessment Cycle

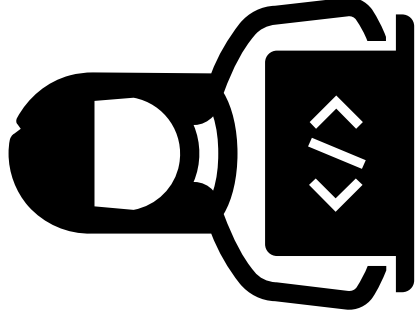
Overview



Gather some data from you



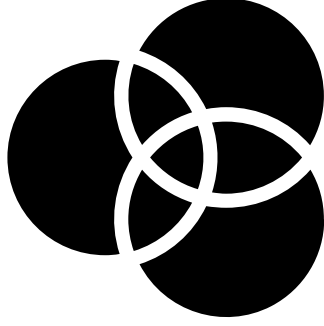
Overview



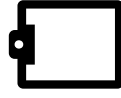
Overview of R and R Studio



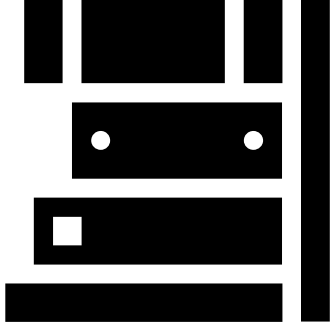
Overview



Comparison of R to other programs



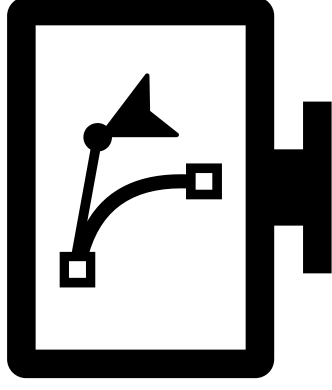
Overview



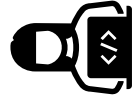
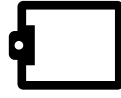
Share some resources



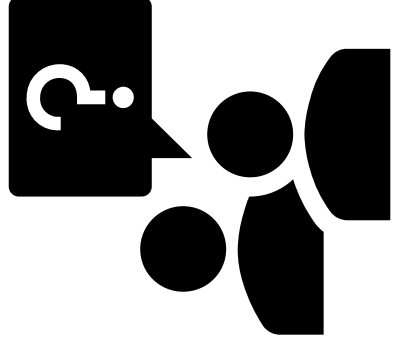
Overview



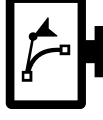
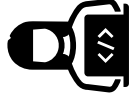
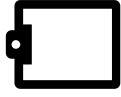
Live walk-through



Overview

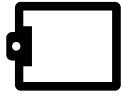


Time for questions



Overview

Let's get started!



Assessment Cycle

Analyzing Data,
Reporting Results, and
Maintaining
Information

Information

Fidelity

Assessment Cycle: Analyze Data

FileEditViewDataTransformAnalyzeGraphsUtilitiesExtensionsWindowHelp																
FontFontid																
ClipboardFormat Painter																
CutCopyFormat Painter																
B I U Merge & Center																
Alignment																
General																
Number																
A B C D E F G H I J K																
1	Id	gender	age	anx1	anx2	anx3	anx4	anx5	anx6	anx7	anx8					
2	1 M		45	3	3	2	2	1	3	2	3					
3	2 M		57	3	3	1	3	2	4	3	3					
4	3 F		50	4	5	2	4	3	4	2	2					
5	4 F		54	5	4	3	2	4	2	4	2					
6	5 F		37	4	5	5	5	5	5	5	5					
7	6 F		48	4	5	5	5	4	5	5	5					
8	7 M		43	3	4	3	4	2	1	3	3					
9	8 F		47	1	2	2	2	2	3	1	1					
10	9 NB		52	4	1	1	3	2	3	3	3					
11	10 F		52	5	5	3	3	3	4	3	3					
12	11 M		61	2	2	2	1	3	4	3	3					
13	12 M		51	4	5	2	4	4	3	4	2					
14	13 NB		52	4	4	3	2	2	3	2	2					
15	14 M		39	5	5	5	4	5	5	5	5					
16	15 NB		56	1	1	1	1	1	1	1	1					
17	16 NB		54	3	4	3	3	3	3	4	3					
18	17 M		53	1	2	2	2	2	2	1	2					
19	18 F		54	2	2	2	2	2	2	1	2					
20	19 F		50	3	2	3	2	4	1	3	3					
21	20 M		57	1	2	1	2	1	2	2	1					
22	21 F		47	3	1	2	1	3	3	2	2					
23	22 F		47	4	4	5	5	5	5	5	5					
24	23 F		44	5	4	5	5	4	4	4	4					
25	24 F		42	4	4	3	4	4	4	2	1					
26	25 M		58	3	2	1	1	2	2	2	1					
27	26 M		46	2	1	2	2	4	3	1	1					
28	27 M		44	4	3	5	3	3	4	4	4					
29	28 M		45	1	1	2	1	2	1	2	1					
30	29 M		46	3	5	3	5	3	3	4	5					
31	30 F		40	2	1	2	1	3	4	2	2					
32	31 M		40	1	1	2	3	2	2	2	2					
33	32 M		47	3	4	3	1	5	4	4	4					
34	33 F		50	4	3	4	2	2	2	2	2					
35	34 F		46	4	5	5	2	4	3	4	4					
36	35 M		58	5	5	5	4	5	5	5	5					
37	36 F		54	2	3	1	2	2	1	2	3					

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help																
Power Analysis																
Meta Analysis																
Reports																
Descriptive Statistics																
Bayesian Statistics																
Tables																
Compare Means																
General Linear Model																
Generalized Linear Models																
Mixed Models																
Correlate																
Regression																
Loglinear																
Neural Networks																
Classify																
Dimension Reduction																
Scale																
Nonparametric Tests																
Forecasting																
Survival																
Multiple Response																
Missing Value Analysis...																
Multiple Imputation																
Complex Samples																
Simulation...																
Quality Control																
Spatial and Temporal Modeling...																
Direct Marketing																
1	age	14														
2		14														
3		14														
4		15														
5		15														
6		14														
7		15														
8		15														
9		14														
10		14														
11		14														
12		14														
13		14														
14		15														
15		15														
16		15														
17		14														
Data View Variable View																

First Survey Question

These are just for fun - we'll be illustrating some code later
with this data



Introduction to R and R Studio

- R is a programming language
- R Studio is an IDE - a user-friendly interface
- Both R and R Studio are free and open source

Where to find them

<https://posit.co/download/rstudio-desktop/>

1: Install R

RStudio requires R 3.6.0+. Choose a version of R that matches your computer's operating system.

R is not a Posit product. By clicking on the link below to download and install R, you are leaving the Posit website. Posit disclaims any obligations and all liability with respect to R and the R website.

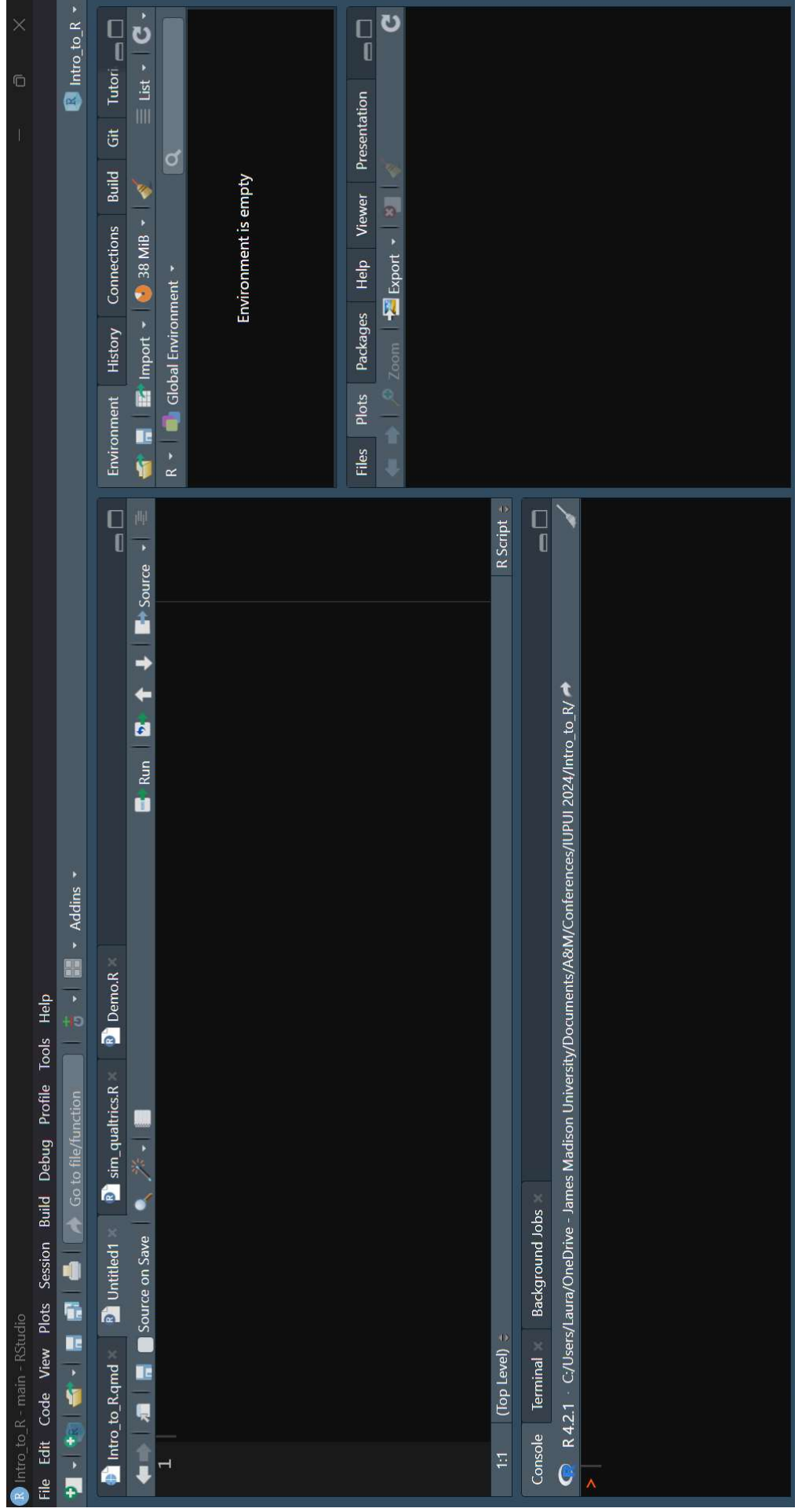
DOWNLOAD AND INSTALL R

2: Install RStudio

DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Size: 265.55 MB | SHA-256: 513216FE | Version:
2024.09.0+375 | Released: 2024-09-23

R Studio window



Script vs. Project

- A script file is a set of instructions

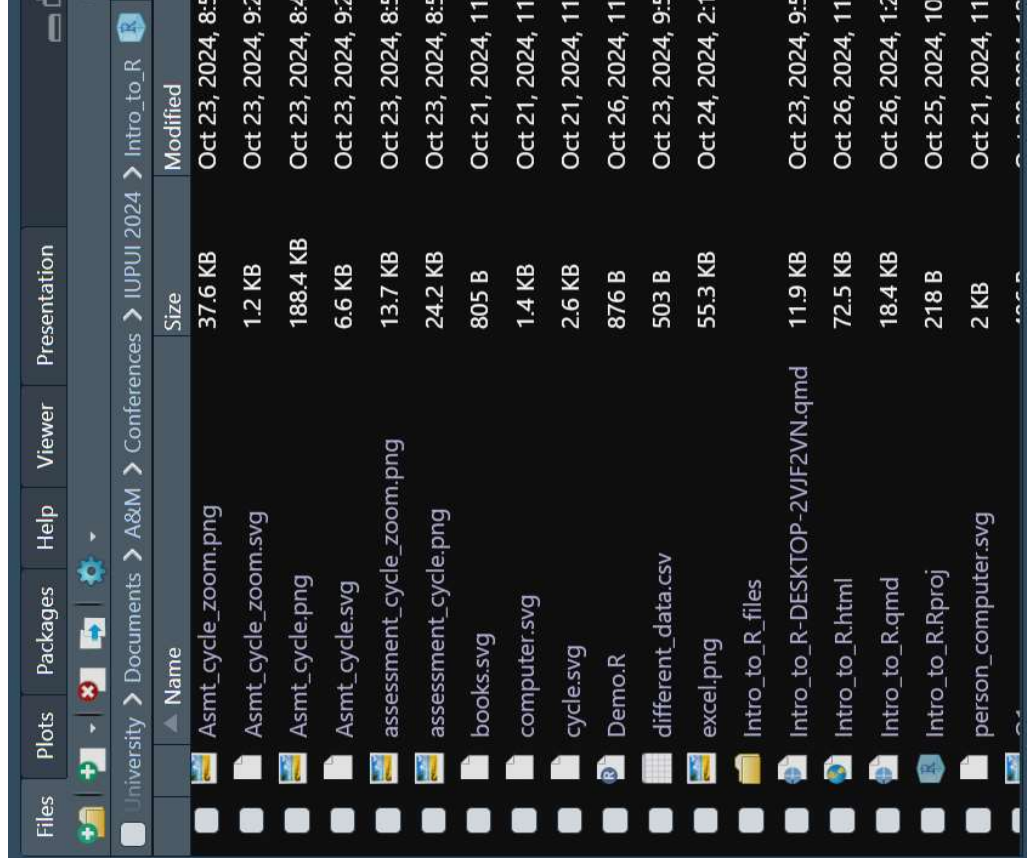
Script vs. Project

```
1 #Set seed
2 set.seed(123456)
3
4 #First survey
5 programs <- c("Excel", "SPSS dropdowns", "SPSS syntax",
6               "SAS", "R", "Python", "Something not listed",
7               "What is this 'analyze' you speak of?")
8
9
10 Q1 <- sample(programs, size = 50, replace = TRUE, prob = c(0.35, 0.35, 0.05,
11                                                         0.05, 0.15, 0.01,
12                                                         0.02, 0.02))
13
14 survey1 <- data.frame(Q1 = Q1)
15
16 write.csv(survey1, "survey1.csv")
17
18 #Second survey
19 age <- round(rnorm(50, mean = 35, sd = 5.1), 0)
20 articles <- round(rnorm(50, mean = 12, sd = 3))
21
22 #Combine into dataframe
23 #Make sure vars are named correctly
24 #Need to match what I'm doing with Qualtrics
25 survey_age <- data.frame(Q1 = age,
26                           Q2 = articles)
27
28
29 46:1 (Top Level)
R Script
Console
```

Script vs. Project

- A script file is a set of instructions
- A project is a directory (folder)

Script vs. Project



The screenshot shows a file explorer window with a dark theme. The top navigation bar includes tabs for Files, Plots, Packages, Help, Viewer, and Presentation. The breadcrumb path is University > Documents > A&M > Conferences > IUPUI 2024 > Intro_to_R. The main area displays a list of files and folders with columns for Name, Size, and Modified date.

	Name	Size	Modified
	Asmt_cycle_zoom.png	37.6 KB	Oct 23, 2024, 8:5
	Asmt_cycle_zoom.svg	1.2 KB	Oct 23, 2024, 9:2
	Asmt_cycle.png	188.4 KB	Oct 23, 2024, 8:4
	Asmt_cycle.svg	6.6 KB	Oct 23, 2024, 9:2
	assessment_cycle_zoom.png	13.7 KB	Oct 23, 2024, 8:5
	assessment_cycle.png	24.2 KB	Oct 23, 2024, 8:5
	books.svg	805 B	Oct 21, 2024, 11
	computer.svg	1.4 KB	Oct 21, 2024, 11
	cycle.svg	2.6 KB	Oct 21, 2024, 11
	Demo.R	876 B	Oct 26, 2024, 11
	different_data.csv	503 B	Oct 23, 2024, 9:5
	excel.png	55.3 KB	Oct 24, 2024, 2:1
	Intro_to_R_files		
	Intro_to_R-DESKTOP-2VJF2VN.qmd	11.9 KB	Oct 23, 2024, 9:5
	Intro_to_R.html	72.5 KB	Oct 26, 2024, 11
	Intro_to_R.qmd	18.4 KB	Oct 26, 2024, 1:2
	Intro_to_R.Rproj	218 B	Oct 25, 2024, 10
	person_computer.svg	2 KB	Oct 21, 2024, 11

Script vs. Project

- A script file is a set of instructions
- A project is a directory (folder)
 - Script(s)
 - Data
 - Output
 - Anything, really

Script vs. Project

Main difference: calling files

Script file

```
df <- read.csv("path/to/data/here.csv")
```

Project

```
df <- read.csv("file.csv")
```

Base R and Packages

- ‘base’ R has quite a bit of functionality on its own
- ‘packages’ extend what base R can do
 - Specialized collections of functions (e.g. [psych](#), [lavaan](#))
 - Make data wrangling and visualization easier (e.g. [tidyverse](#), [ggplot2](#))
 - Advanced analysis (e.g. [lme4](#), [brms](#))

Installing and Loading Packages

```
1 #Install a package
2 #Only need to do this once on any given machine
3 #If you switch machines (e.g. office to home computer) you'll need to re-in
4 install.packages("ggplot2")
5
6 #Call the package
7 #You will need to call your packages for each R session
8 library(ggplot2)
```

Installing and Loading Packages

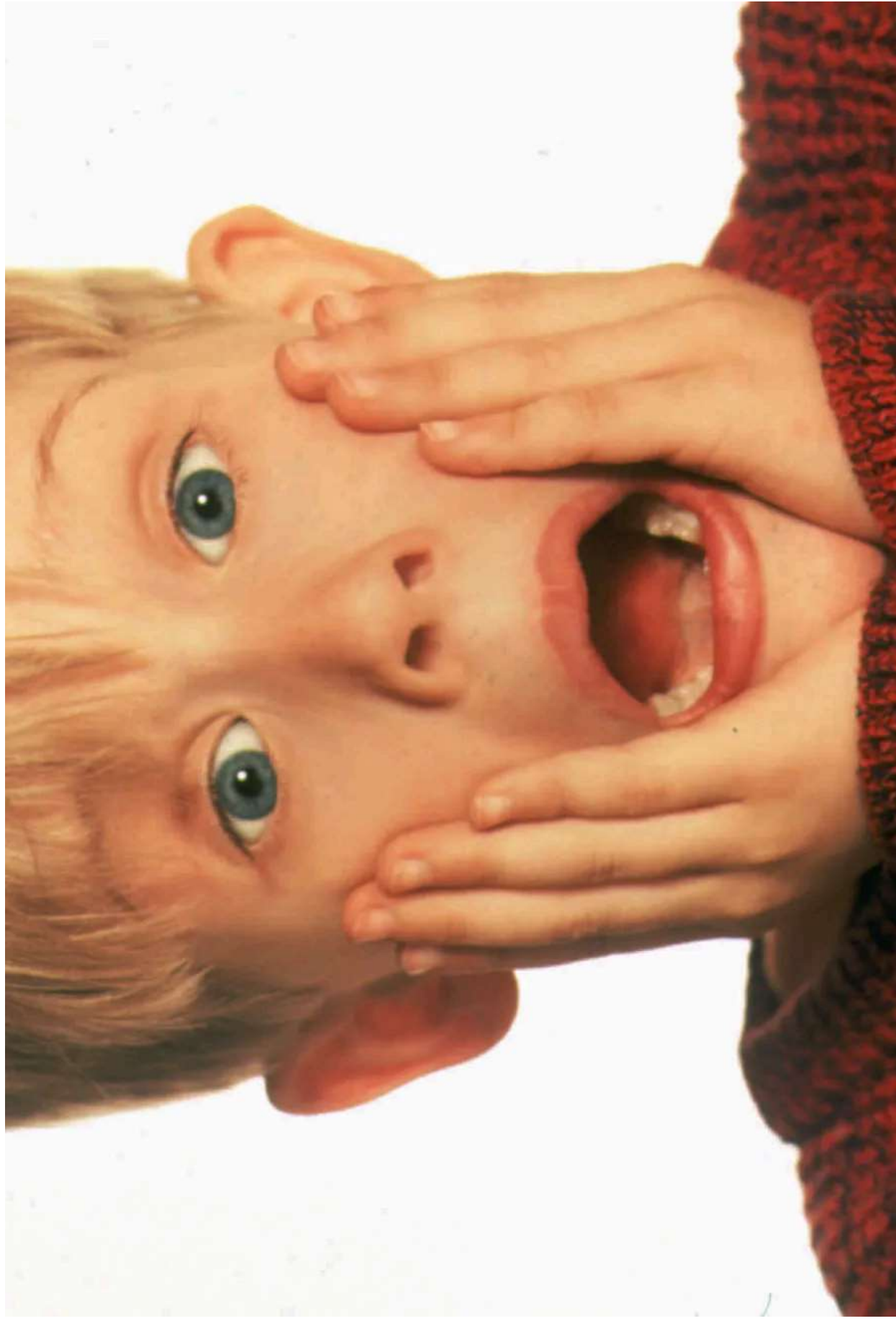
```
1 #Handy to have a 'set-up' area at the beginning of your code
2 #Load necessary packages
3 library(tidyverse)
4 library(psych)
5 library(car)
6
7 #Then proceed with the rest of your code....
```

Second Survey Question

Feel free to be as honest or dishonest as you'd like here...



Syntax



(Source: "Columbus, C. (1990). Home Alone. Twentieth Century

Syntax

- Like many things, an up-front investment of time can save a lot of time long-term
- MANY available resources for code
- Take and modify what you need

The Argument for Syntax

Let's imagine two scenarios where you're trying to describe your work:

Scenario 1

Scenario 2

So, you have this Excel file and you want to make a formula in one of the cells to sum up the 'pre' and 'post' columns. Then, you need to reverse score one of the items, so you have to make that, then go ahead and delete the old column and put the new one in its place. Then, to do a t-test, you use this formula across these columns....

The Argument for Syntax

1. Documentation

2. Reproducibil

Documentation

```
1 #Comments can tell you what you did
2 #Or, what dataset you read in
3
4 #Read in dataset A
5 data <- read.csv("groupA.csv")
```

Documentation

```
1 #Read in data
2 test_data <- read.csv("some/file/path.csv")
3
4 #What does our data look like?
5 head(test_data)
6
7 #Take a peek at a variable
8 #Check if any missing data
9 table(test_data$sp_2024, useNA = "always")
10
11 #Two missing values
12 #Remove them from the data - can only have complete cases here
13 test_data_clean <- na.omit(test_data)
```

Documentation

But, what if you had a different criteria? What if you only removed them if some other variable was above/below a certain value?

[illegible]

Reproducibility

```
1 #Read in data
2 data1 <- read.csv("some_data.csv")
3
4 #Examine distributions
5 #Pre first
6 ggplot(data = data1) +
7   geom_histogram(aes(pre))
8
9 #Then post
10 ggplot(data = data1) +
11   geom_histogram(aes(post))
12
13 #Do a t-test
14 t.test(data1$pre, data1$post, paired = TRUE)
```

Reproducibility

```
1 #Read in data
2 data1 <- read.csv("different_data.csv")
3
4 #Examine distributions
5 #Pre first
6 ggplot(data = data1) +
7   geom_histogram(aes(pre))
8
9 #Then post
10 ggplot(data = data1) +
11   geom_histogram(aes(post))
12
13 #Do a t-test
14 t.test(data1$pre, data1$post, paired = TRUE)
```

Excel vs. R

- Excel can hold a decent amount of data
- Functions aid in data analysis
- Pivot tables
- Built-in data visualization

Excel vs. R

But...

How many times have you messed up an Excel sheet when trying to do an analysis and didn't realize it until too late?

Excel vs. R

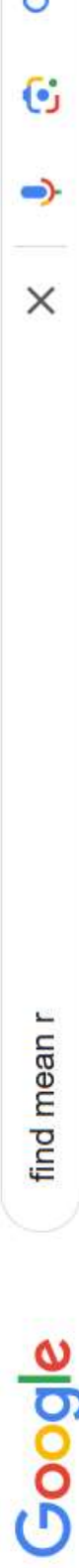
- In Excel, you perform calculations and manipulations on your original dataset
 - Unless you save a copy!
- With R, manipulations don't affect the original dataset
 - This was mind-blowing to me

Available R Resources

- A major benefit to R being open source: near infinite Google-ability
- Many other free resources if you wanted to learn more
- Can borrow code from other folks who have done it before
 - No need to reinvent the wheel

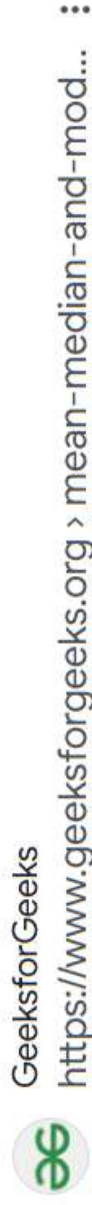
Available R Resources

Example: Google



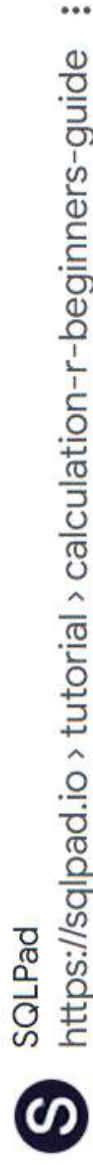
Learn R: Mean, Median, and Mode Cheatsheet

In R, the mean of a vector is calculated using the `mean()` function. The function accepts a vector as input, and returns the average as a numeric.



Mean, Median and Mode in R Programming

Jul 8, 2024 – In R calculating the **mean** and median is straightforward using the built-in functions `mean()` and `median()`. Calculating the mode requires a ...



Mean Calculation in R: A Beginner's Guide

May 5, 2024 – In R, the simple **mean** is calculated using the `mean()` function, and the weighted **mean** can be calculated using the weighted `mean(x, w)` function, ...

Introduction : Basic Syntax and Functions for ... Advanced Techniques for ...

Available R Resources

Example: YouTube

<https://www.youtube.com/>

Available R Resources

Example: Course materials

https://bookdown.org/laura_lambert_99/intermediate_stats/

Available R Resources

Example: Giant repository of R books

<https://www.bigbookofr.com/>

Available R Resources

Example: R for Data Science book

<https://r4ds.hadley.nz/>

Available R Resources

Example: GitHub for this presentation!

https://github.com/lauralambert99/Intro_to_R

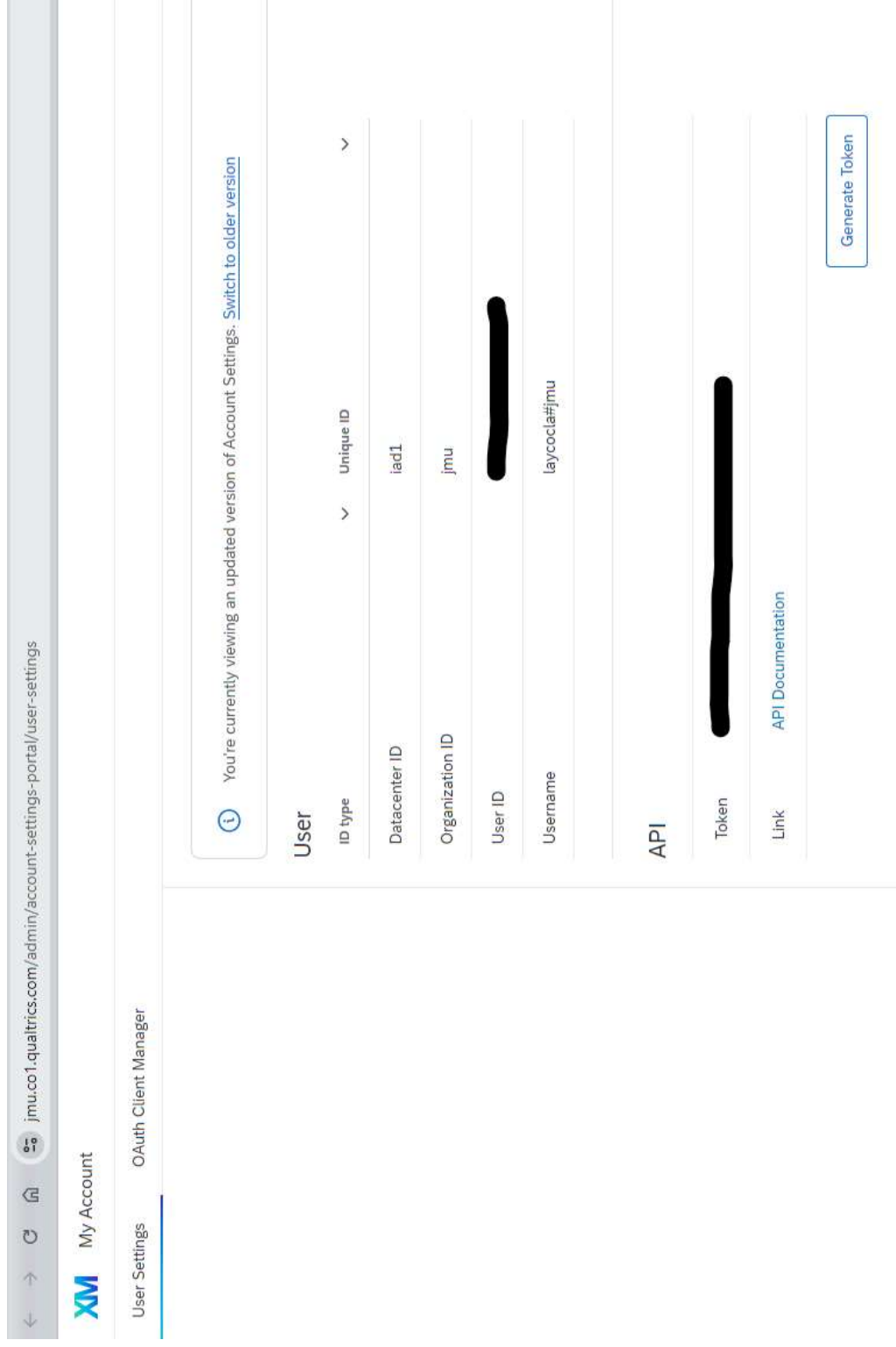
Example Use Case

- Quick dive into some code, then we'll come back up
- Newer package: [qualtRics](#)
- Allows you to pull survey data straight into R for analysis

Example Use Case

```
1 #Call the package
2 library(qualtRics)
3
4 #This is an initial set-up - only need to do once per machine!
5 qualtrics_api_credentials(api_key = "YOUR_KEY_HERE",
6   base_url = "URL.GOES.HERE",
7   install = TRUE)
```

Example Use Case



Example Use Case

```
1 #See what surveys are associated with your account
2 #Here, saving to an object
3 surveys <- all_surveys()
4
5 #We can see what the column names are
6 names(surveys)
7
8 #And then see what surveys we have
9 surveys$name
```

Example Use Case

```
1 #See what surveys are associated with your account
2 #Here, saving to an object
3 surveys <- all_surveys()
4
5 #We can see what the column names are
6 names(surveys)
7
8 #And then see what surveys we have
9 surveys$name
```

Example Use Case

```
1 #See what surveys are associated with your account
2 #Here, saving to an object
3 surveys <- all_surveys()
4
5 #We can see what the column names are
6 names(surveys)
7
8 #And then see what surveys we have
9 surveys$name

[1] "id"      "name"    "ownerId" "lastModified" "creationDate"
[6] "isActive"
```

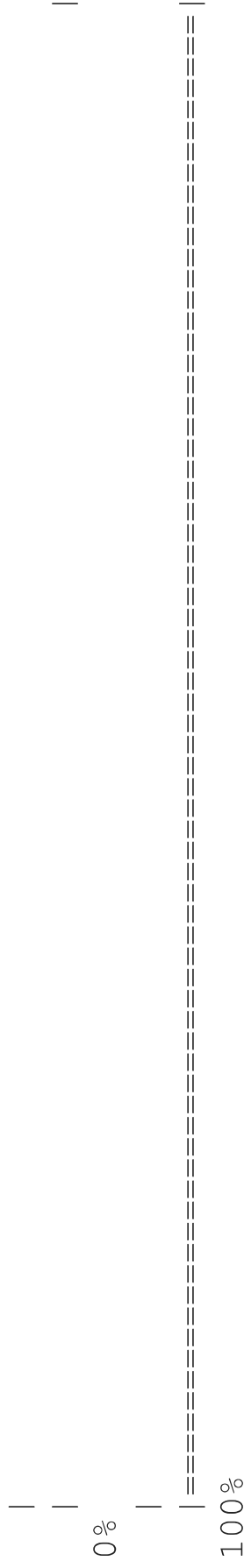
Example Use Case

```
1 #See what surveys are associated with your account
2 #Here, saving to an object
3 surveys <- all_surveys()
4
5 #We can see what the column names are
6 names(surveys)
7
8 #And then see what surveys we have
9 surveys$name

[1] "2023 ISAT Senior Survey"      "MS ISAT/MSc EMS Orientation Survey"
[3] "AI24_Q3"                     "2024 ISAT Senior Survey"
[5] "30th Anniversary Alumni Survey" "STEM Belonging Survey"
[7] "AI24_Q2"                     "AI24_Q1"
[9] "STEM Focus Group Availability"
```

Example Use Case

```
1 #Get survey data  
2 survey_data <- fetch_survey(surveyID = surveys$id[8])
```



Example Use Case

```
1 #Save question names to an object
2 questions <- survey_questions(surveyID = surveys$id[8])
3
4 #See what the questions are
5 print(questions)

# A tibble: 1 × 4
  qid  qname question
force_resp
<chr> <chr> <chr>
1 QID1  Q1    What software do you currently use to analyze your ass... FALSE
<lg1>
```

Example Use Case

```
1 #Now, we can do things!  
2 table(survey_data$Q1)
```

```
Excel      SPSS dropdowns  
16         2  
SPSS syntax SAS  
4          1  
R          Python  
8          0  
Something not listed What is this 'analyze' you speak of?  
4          6
```

Parting Thoughts Before Demo

- You don't need to be an expert in R to use it
- Use it however works for **you** - there isn't a "right" way
- Projects make things easier
- Don't reinvent the wheel!
- You'll learn real fast how well you spell

Last Survey Question

Last bit of data!



You Promised Flexibility!

```
creative_name <- read.csv("some/data/file.csv")
```

You Promised Flexibility!

```
creative_name <- read.csv("some/data/file.csv")
```

You Promised Flexibility!

creative_name <- read.csv("some/data/file.csv")

You Promised Flexibility!

```
creative_name <- read.csv("some/data/file.csv")
```

You Promised Flexibility!

```
creative_name <- read.csv("some/data/file.csv")
```

You Promised Flexibility!

creative_name\$some_variable

You Promised Flexibility!

creative_namesome_variable

You Promised Flexibility!

creative_name\$**some_variable**

Live Demo time

Now, let's run some code, using the data you have so kindly provided

If you want to run it on your computer, access some fake data here: <https://bit.ly/3NINVha>

**What questions can I answer for you
now?**

My Contact Info

Laura Lambert laycocla@jmu.edu