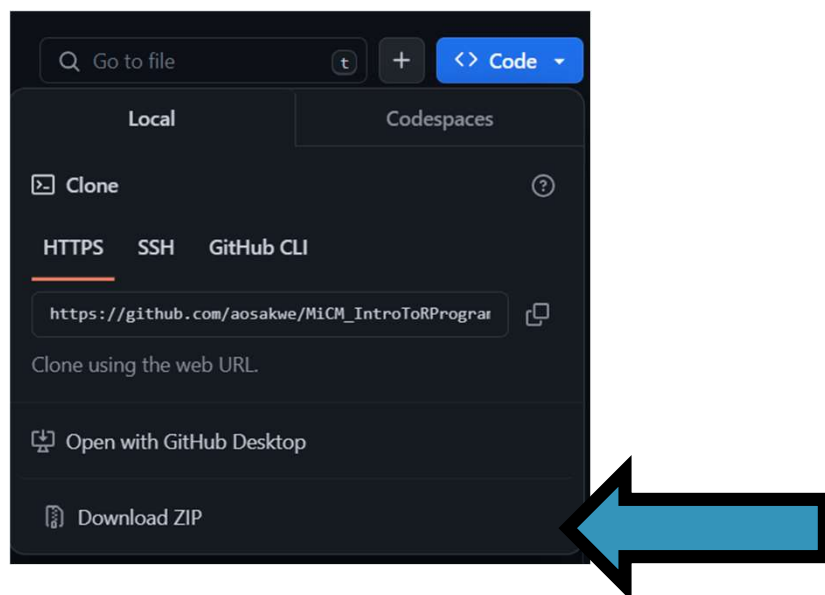
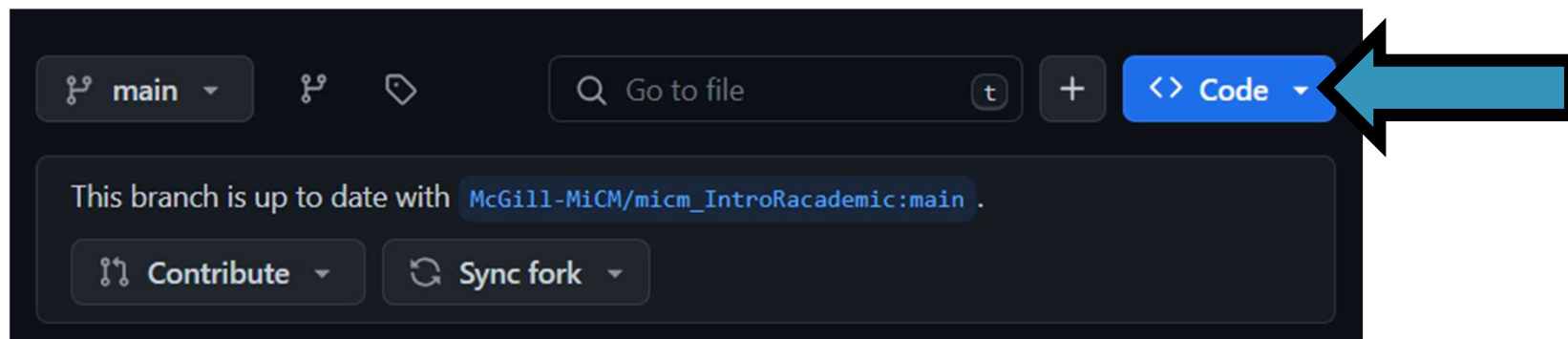


Download Workshop Materials

1. Go to https://github.com/aosakwe/MiCM_IntroToRProgramming

2.



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Intro to programming in R

Lead: Adrien Osakwe

Facilitator: Bangli Cao

February 13, 2025

Materials adapted from Larisa M. Soto and Xiaoqi Xie

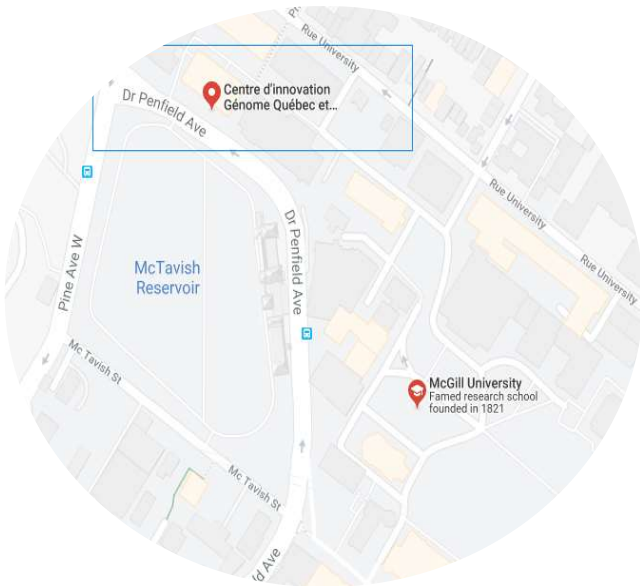


McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Mission statement: deliver quality workshops designed to help biomedical researchers develop the skills they need to succeed.



Location: 550 Sherbrooke West,
Montreal, Quebec



Scan the QR code to sign up
for our **mailing list**

Contact: workshop-micm@mcgill.ca



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Winter Series

Workshop	Date	Location	Registration
How to think in Code	Jan. 28 1PM-3PM	EDUC 133	Closed
Intro to Git & GitHub	Jan. 30 1PM-5PM	EDUC 133	Closed
Intro to Unix	Feb. 6 1PM-5PM	EDUC 133	Closed
Intro to Python (Part 1)	Feb. 11 1PM-5PM	EDUC 133	Open
Intro to R (Part 1)	Feb. 13 1PM-5PM	EDUC 133	Open
Exploring MATLAB	Feb. 18 1PM-5PM	EDUC 133	Open
Statistics in R (Part 2)	Feb. 20 1PM-5PM	EDUC 133	Open
Data Processing in Python	Feb. 25 1PM-5PM	EDUC 133	Open
Intro to Machine Learning	Mar. 13 1PM-5PM	EDUC 133	TBA
Intro to R (Part 1)	TBA	EDUC 133	TBA
Intro to Python (Part 1)	TBA	EDUC 133	TBA

<https://www.mcgill.ca/micm/training/workshops-series>



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Workshop outline Part 1

1

The language

History
Foundation
Syntax
Logical ops
Help
Packages

2

Data types

Vectors
Factors
Lists
Data Frames
Arrays
Hands on

3

Control Structures

Functions
If statement
for loop
Hands on



Workshop outline Part 2

4

Basic data manipulation

Read & Write
Subset
Split
Join

Hands on

5

Advanced data Manipulation

dplyr
tidyr
plyr
DataTable

Hands on

6

Generating Outputs

Graphics
ggplot2
RMarkdown
Templates

7

Software development

Good coding practices
Documentation standards
Debugging



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Workshop Components

- Theory
- Code Examples
- Hands-on Activities



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

1. The R programming language

Learning objectives

- Why Excel is not enough
- What is R
- What is an IDE
- Basic Operations



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Why not Excel?

- Easy at first glance
- **Issues**
 1. Hard to automate
 2. Hard to reproduce
 3. Inflexible
 4. Slow!



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant



- **Statistical** Programming Language
- Integrated suite for **data manipulation, analysis, and graphical visualization**
- Environment where **statistical tests can be performed**
- Its functionality can be easily extended with ***packages***
- GNU project of free software
- Users have the freedom to:
 - Run the program
 - View and modify the source code
 - Redistribute copies and
 - Distribute their modifications

R facts

- Interpreted language
- Object-oriented
- No spaces allowed in variable names
- Case sensitive
- 1-based indexing
- Allows user-defined functions
- Works with environments



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

R Files

- Many types of files can contain R code
 - **.R 'Script'**
 - **.Rmd 'R Notebook'**
 - **.qmd 'Quarto Notebook'**
 - **.ipynb 'Jupyter Notebook'**
- Scripts
 - Automation & Portability
- Notebooks
 - Documentation
 - Accessibility



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

R & RStudio



- R & RStudio are **different entities**
- R is the programming language
 - The actual code we execute
 - Developed at the University of Auckland
- RStudio is an **Integrated Development Environment (IDE)**
 - A GUI software to develop and execute R code
 - Developed by Posit



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

```
PS C:\Program Files\R\R-4.2.2\bin> .\R.exe
```

```
R version 4.2.2 (2022-10-31 ucrt) -- "Innocent and Trusting"  
Copyright (C) 2022 The R Foundation for Statistical Computing  
Platform: x86_64-w64-mingw32/x64 (64-bit)
```

```
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.
```

```
  Natural language support but running in an English locale
```

```
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.
```

```
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.
```

```
> print('hello world')  
[1] "hello world"  
> |
```

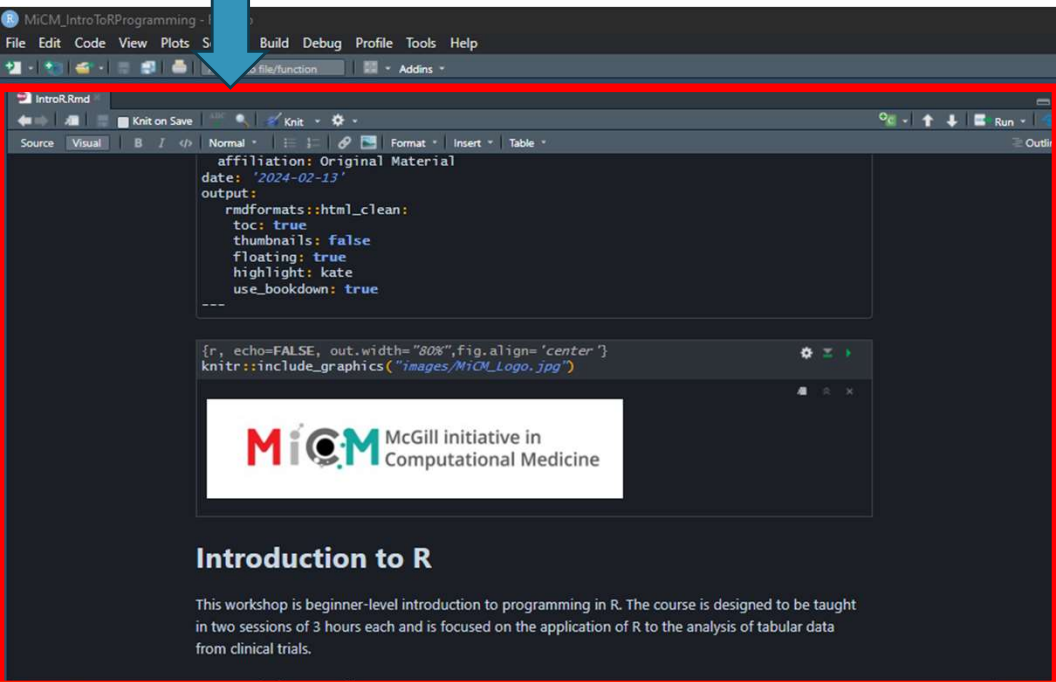


McGill

Quantitative Life
Sciences

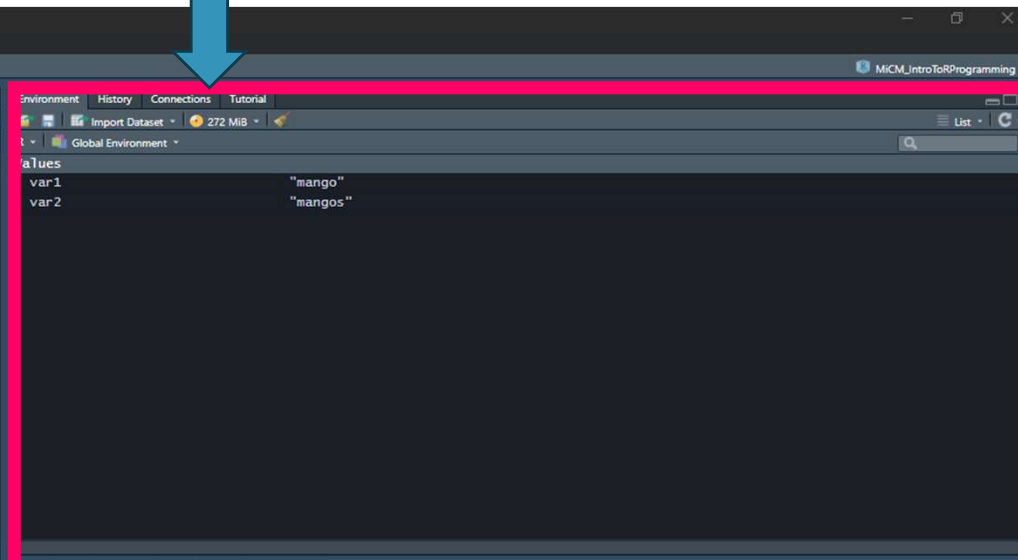
Sciences quantitatives
du vivant

Source Pane

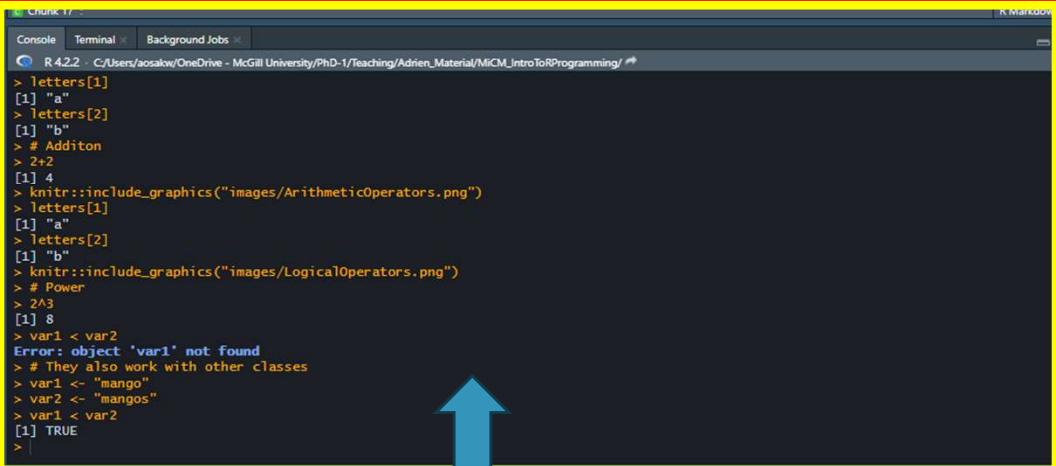


The Source Pane displays the R Markdown source code for a file named 'IntroR.Rmd'. The code includes a YAML header with metadata like 'affiliation: Original Material' and 'date: '2024-02-13'', followed by an 'output' section with options like 'rmdformats::html_clean' and 'knitr::include_graphics'. Below the code, the rendered HTML output is shown, featuring the 'MiCM' logo (McGill initiative in Computational Medicine) and a title 'Introduction to R'. A paragraph of introductory text follows, stating the workshop is a beginner-level introduction to R programming.

Environment Pane

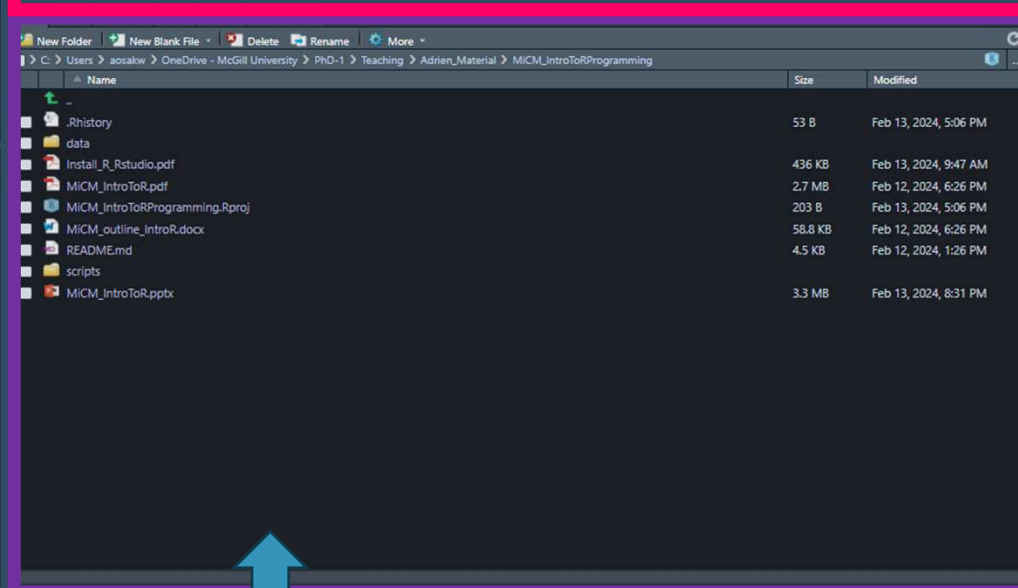


The Environment Pane shows the 'Global Environment' with two variables: 'var1' with the value 'mango' and 'var2' with the value 'mangos'.



The Console Pane shows the R session output. It displays the execution of several R commands, including creating vectors 'letters[1]' and 'letters[2]', performing arithmetic operations like '2+2' and '2^3', and using 'knitr::include_graphics' to load images. The output shows the results of these operations, such as '[1] 4' and '[1] 8'. An error message 'Error: object 'var1' not found' is also visible, followed by a comment '# They also work with other classes' and the successful execution of 'var1 <- "mango"' and 'var2 <- "mangos"'. The final output is '[1] TRUE'.

Console Pane



The Files/Plot/Help Pane shows a file explorer view of the directory 'C:\Users\aosakw\OneDrive - McGill University\PhD-1\Teaching\Adrien_Material\MiCM_IntroToRProgramming'. The directory contains several files and folders, including '.Rhistory', 'data', 'Install_R_Studio.pdf', 'MiCM_IntroToR.pdf', 'MiCM_IntroToRProgramming.Rproj', 'MiCM_outline_IntroR.docx', 'README.md', 'scripts', and 'MiCM_IntroToR.pptx'. The files are listed with their names, sizes, and modification dates.

Files/Plot/Help Pane



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Data types and data structures

Learning objectives

- Understand the differences between classes, objects and data types in R
- Create objects of different types
- Subset and index objects
-
- Learn and use vectorized operations



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Atomic Classes

Also called data types

Character	A,b,c,d,e,..
Numeric (real numbers)	1.00,2.00,... Inf, NaN
Integer	1L,2L,3L,4L,....
Complex	2i
Logical (True/False)	TRUE,FALSE
Missing Value	NA



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Arithmetic operators

Addition	+
Subtraction	-
Division	/
Power	\wedge
Scalar multiplication	*
Matrix multiplication	%*%



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Syntax operators

Comment line	#
Assignment	<-
Access content	\$
Equal	=



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Logical operators

Equal	==
Not equal	!=
Greater than	>
Greater than or equal to	>=
Less than	<
Less than or equal to	<=
contains	%in%
x AND y	x & y
x OR y	x y
NOT x	!x



Objects

Also called data structures

Vector	Only elements of the same class
List	Elements of any class
Factor	Categorical data
Matrix	Elements of the same class in 2D
Data frame	Elements of multiple classes in 2D
NULL	Empty object



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

One dimension

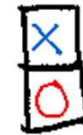
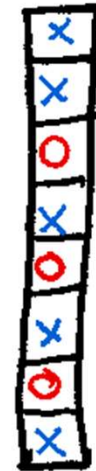
Vector



List



Factor



↑
levels

↑
elements



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Vectors

- Can only contain objects of the **same class**
- Most basic type of R object
- Variables are vectors

`var1 ← 23`
→ 23 vector of length 1

`var2 ← "abc"`
→ abc vector of length 1



Vectorized operations

Diagram illustrating vectorized addition of two 8x1 matrices:

1	2	3
2	5	7
1	3	4
4	9	13
5	7	12
7	5	12
8	6	14
2	7	9

Diagram illustrating vectorized addition of two 3x3 matrices:

3	6	8	8	7	15
2	1	5	5	5	11
7	4	9	15	13	11

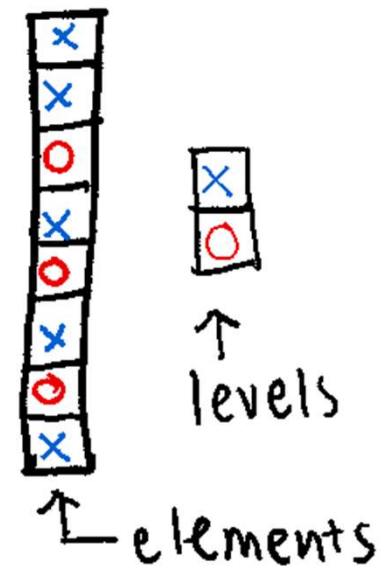
Lists

- Can contain objects of multiple classes
- Very important data type in R
- Extremely powerful when combined with some built-in functions



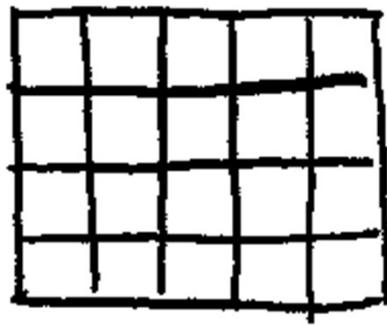
Factors

- Useful when for categorical data
- Can have implicit order, if needed
- Each **element** has a label or **level**
- They are important in statistical modelling and plotting with ggplot
- Some operations behave differently on factors



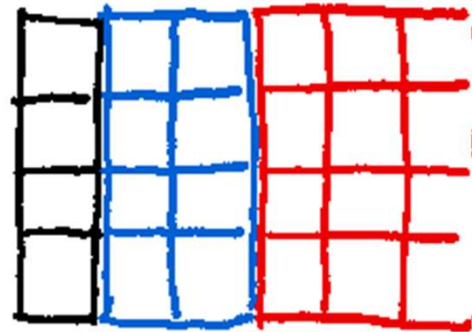
Multiple dimensions

Matrix



4x5

Data Frame



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Break



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Control structures and functions

Learning objectives:

- Understand the concept of environments in R
- Create new functions
- Implement conditional statements
- Implement a for loop to iterate over a list of files



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Conditional statements

- When we want a set of actions to be executed only if certain conditions are met

```
# if
if (condition is true) {
  perform action
}

# if ... else
if (condition is true) {
  perform action
} else { # that is, if the condition is false,
  perform alternative action
}
```



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

For loop

- Repeat a set of operations a certain number of times

```
for (iterator in set of values) {  
  do a thing  
}
```



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

While loop

- Repeat a set of operations until a condition is no longer met

```
while(condition_is_true){  
    do a thing  
}
```


What if base R is not enough?

- Sometimes your analysis requires tools that are not available in base R
- Two options:
 1. **Create new functions**
 2. **Packages** provide a way to incorporate methods and functions from

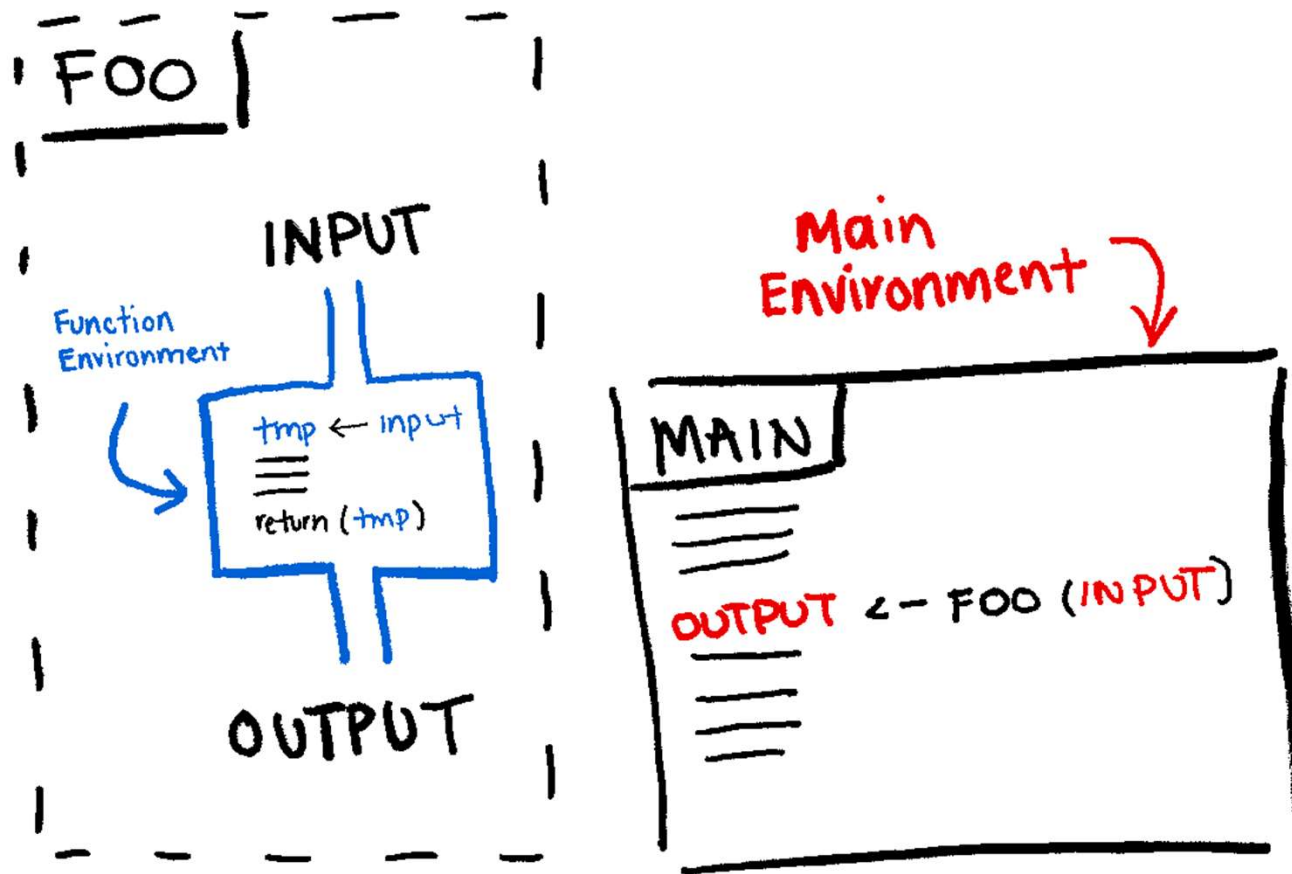


McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Functions and environments



Pass by value and scope

- When we pass an object to a function, a copy of it is created internally
- The changes made inside the function won't modify the original object we passed to it
- Any variables created inside the function will only exist during the function's execution time



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Packages

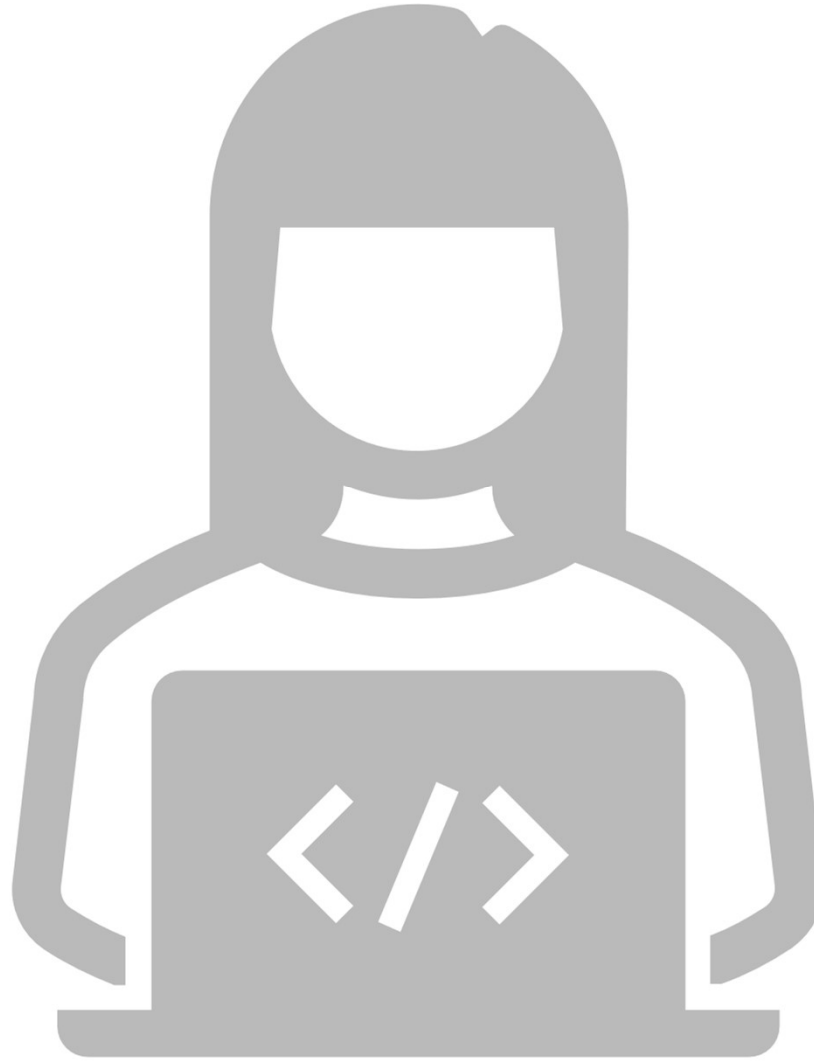
- Packages are a way for users to share methods they have developed
- Incorporate novel methods, datasets, or visualization tools
- Downloaded from many places:
 - **Comprehensive R Archive Network (CRAN)**
 - **Bioconductor**
 - GitHub, Bitbucket etc.



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant



Basic data manipulation

Learning objectives:

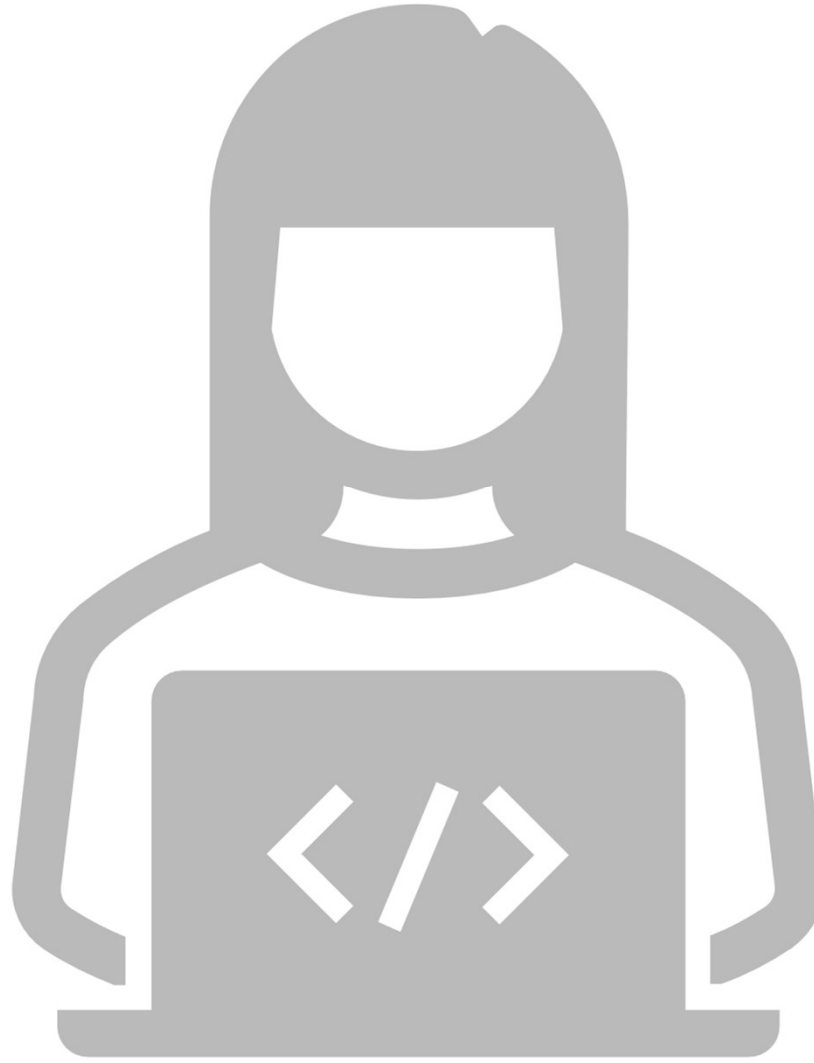
- Learn how to read/write data to/from files with different formats (.tsv, .csv)
- Familiarize with basic operations of data frames
- Index and subset data frames using base R functions
- Manipulate specific data frame columns
- Joining by columns and rows



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant



Advanced data manipulation

Learning objectives:

- Become familiar with the dplyr syntax
- Create pipes with the operator %>%
- Perform operations on data frames using dplyr and tidyr functions
- Implement functions from other external packages



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Split-Apply-Combine problem

INPUT

x	y
a	2
a	4
b	0
b	5

SPLIT

x	y
a	2
a	4
b	0
b	5

APPLY

x	y
a	3
b	2.5

COMBINE

x	y
a	3
b	2.5

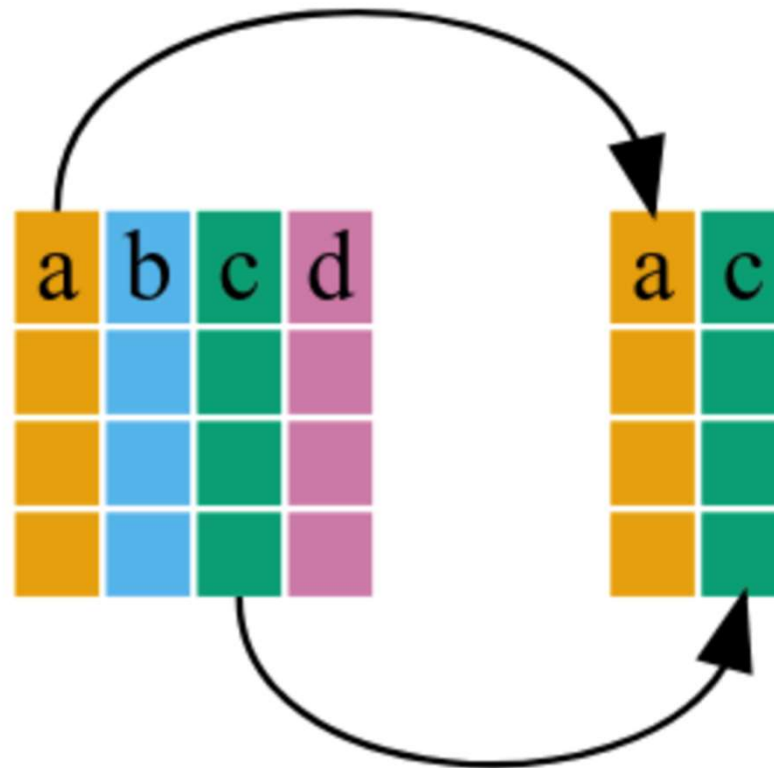


McGill

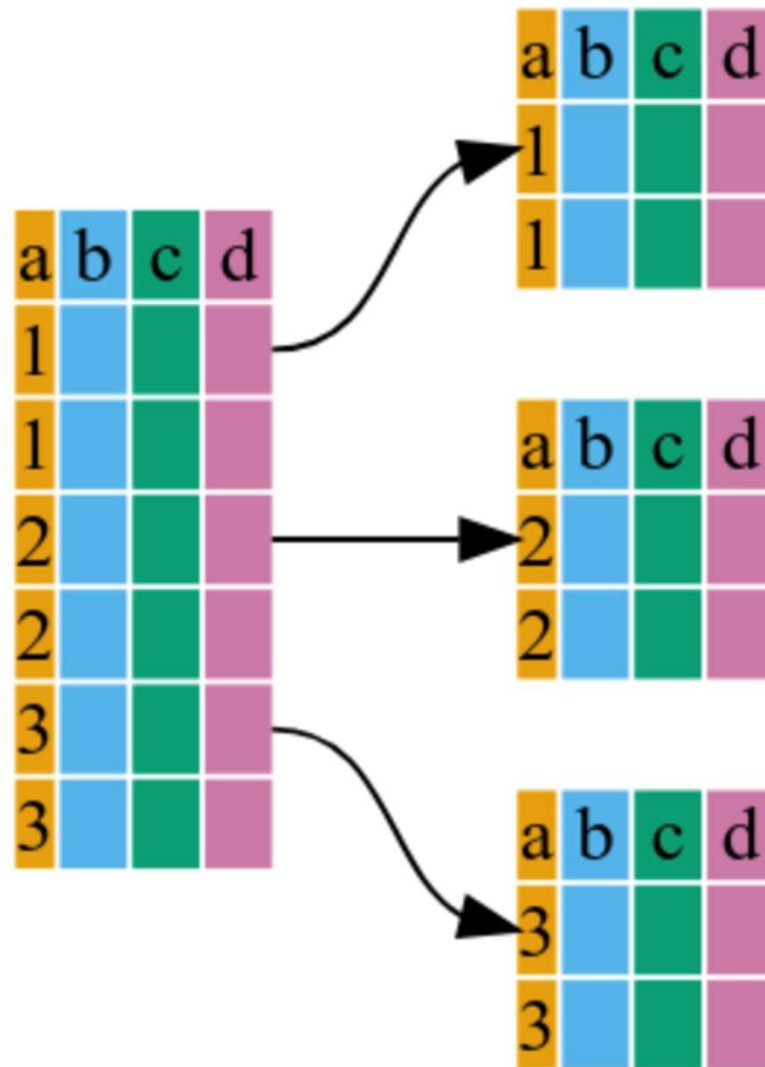
Quantitative Life
Sciences

Sciences quantitatives
du vivant

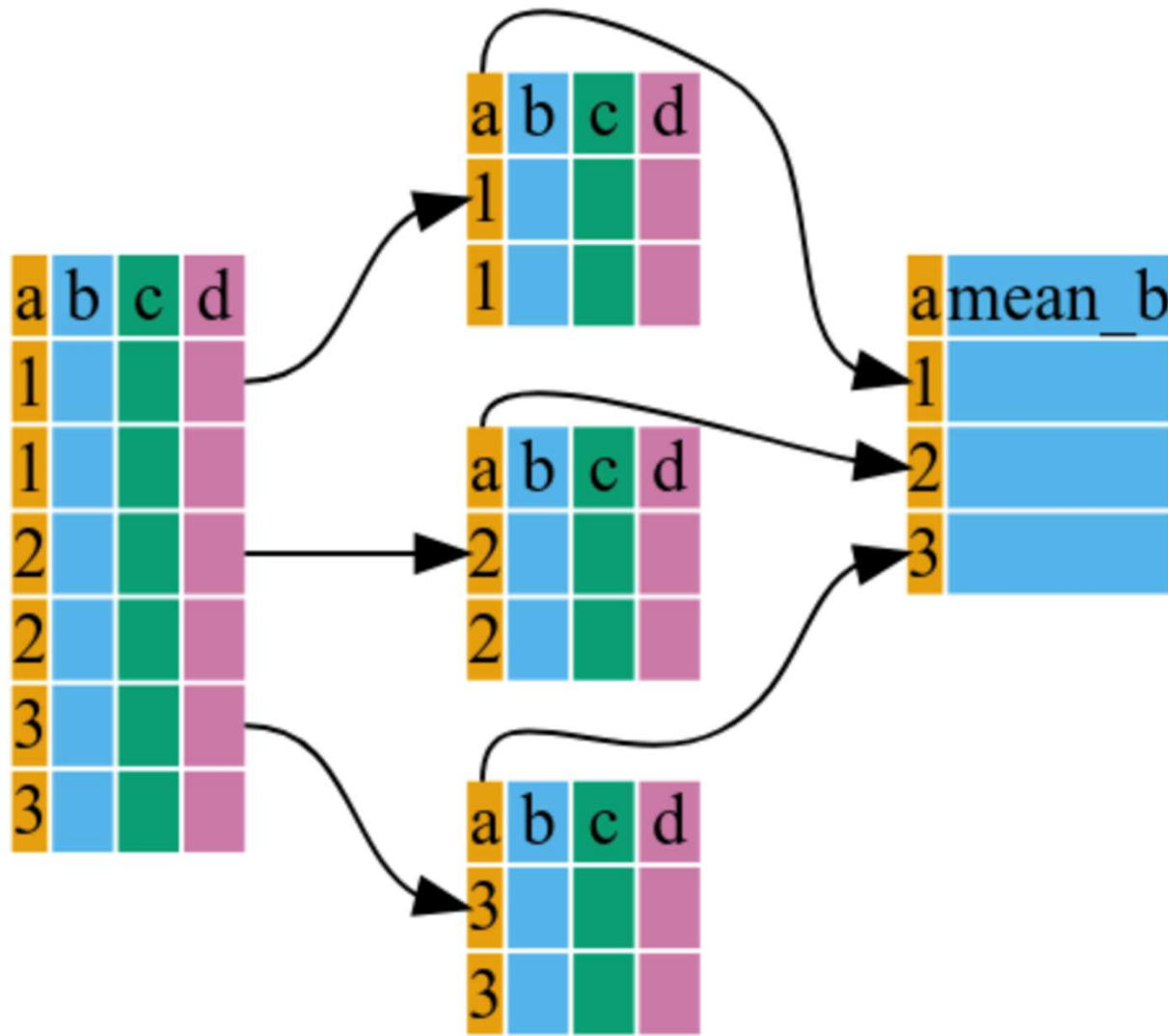
Select

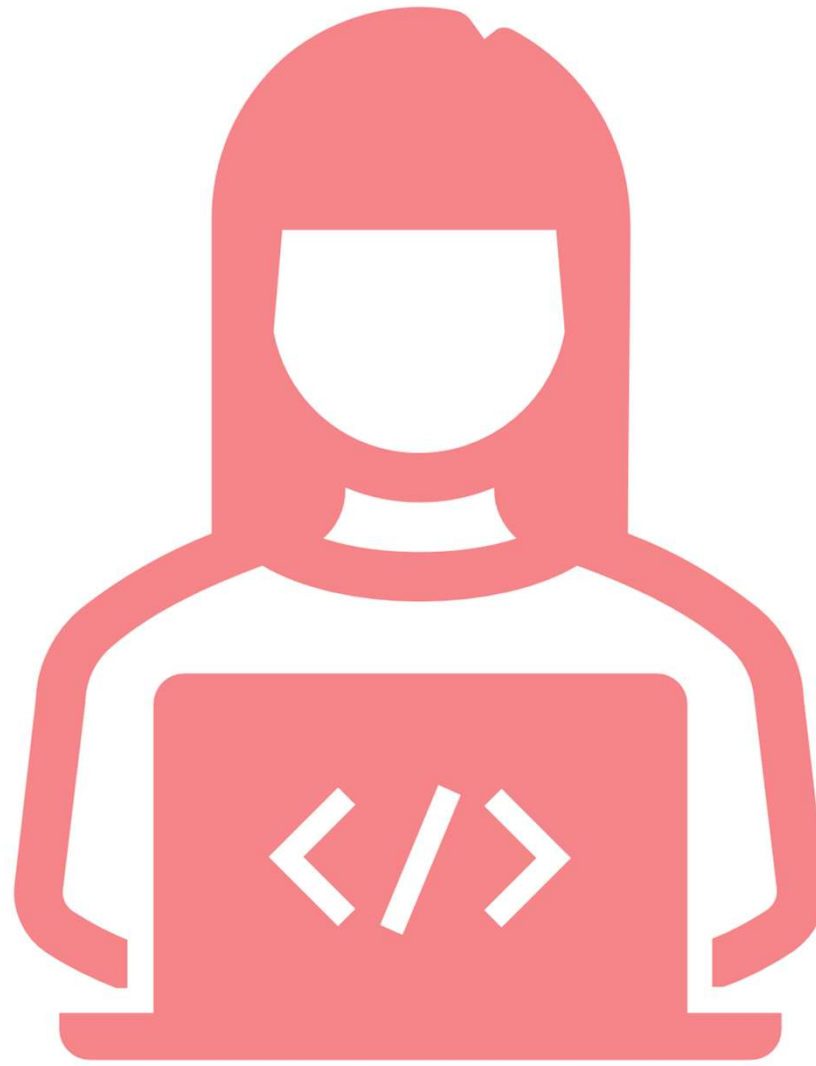


Group by



Summarize





Break



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Generating visual outputs

Learning objectives:

- Create basic plots using base R functions
- Understand the connection between data frames and ggplot2
- Create basic graphs with ggplot2
- Use factors to customize graphics in ggplot2
- Learn about RMarkdown syntax to create reports
- Get familiar with existing RMarkdown templates



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Formatting data for ggplot

WIDE

	a	b	c
1	5	10	
2	7	11	
1	4	9	

Values

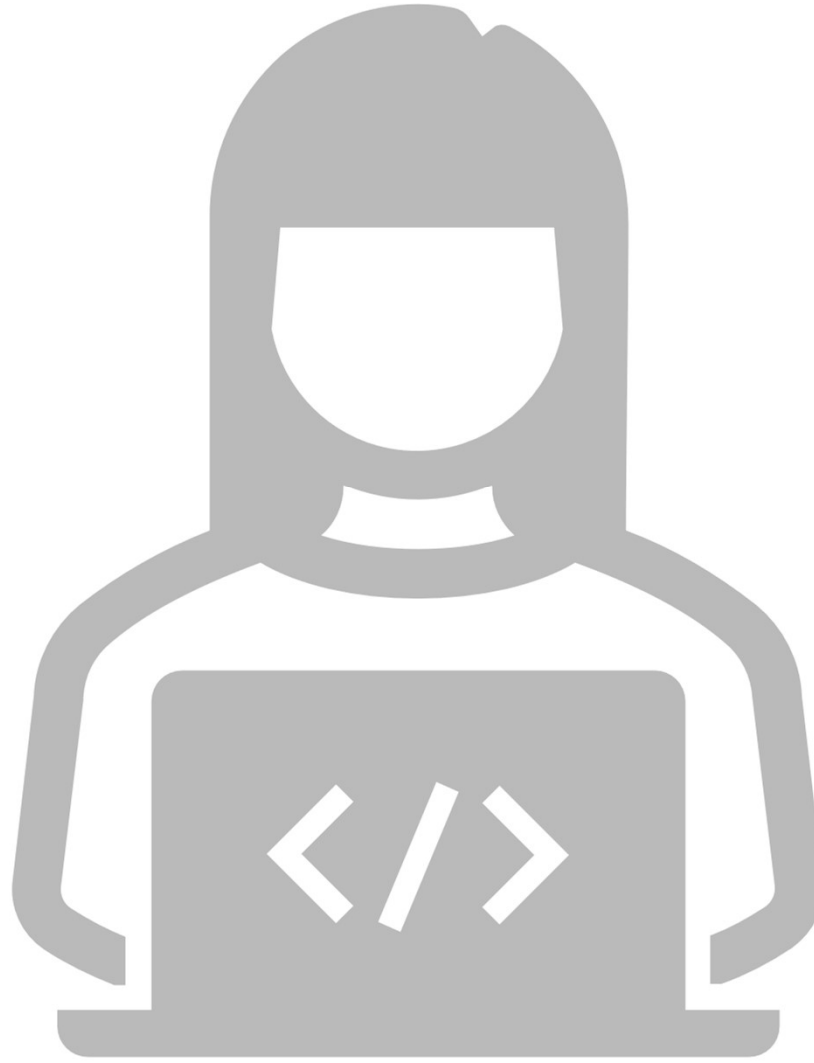
Variables

LONG

Values	variables
1	a
2	q
1	a
5	b
7	b
4	b
10	c
11	c
9	c

grouping factor





Activity: Analyzing a medical data set

Learning objectives:

- Familiarize with a real-life use case of R
- Apply the knowledge from previous modules to create an analysis pipeline



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

COVID testing dataset

Details

Data on testing for SARS-CoV2 from days 4-107 of the COVID pandemic in **2020**. CHOP is a pediatric hospital in Philadelphia, Pennsylvania, USA. These data have been anonymized, time- shifted, and permuted.



The dataset

Documentation

- Part of the medicaldata package
- https://htmlpreview.github.io/?https://github.com/higgi13425/medicaldata/blob/master/man/description_docs/covid_desc.html
- https://htmlpreview.github.io/?https://github.com/higgi13425/medicaldata/blob/master/man/codebooks/covid_testing_codebook.html



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Format

A data frame with 15524 observations and 17 variables

subject_id id number for each subject; type: numeric

fake_first_name an auto-generated fake first name; type: character

fake_last_name an auto-generated fake last name; character

gender anonymized Gender, levels: female, male; type: character

pan_day day after start of pandemic; type: numeric

test_id test that was performed, levels: covid, xcvd1; type: character

clinic_name Clinic or ward where the specimen was collected, 88 levels; type: character

result result of test, levels: positive, negative, invalid; type: character

demo_group patient group, levels: patient, misc_adult, client, other adult, unidentified; type: character

age Age of subject at time of specimen collection (Anonymized), units = years; type: numeric

drive_thru_ind Whether the specimen was collected via a drive-thru site, levels: 1: Collected at drive-thru site; 0: Not collected at drive-thru site; type: numeric

ct_result Cycle at which threshold reached during PCR, range: 14.05-45; type: numeric

orderset Whether an order set was used for test order, levels: 1: Collected via orderset; 0: Not collected via orderset; numeric

payor_group Payor associated with order, levels: commercial, government, unassigned, medical assistance, self pay, charity care, other; type: character

patient_class Disposition of subject at time of collection, levels: inpatient, emergency, observation, recurring outpatient, outpatient, not applicable, day surgery, admit after surgery-obs, admit after surgery-ip; type: character

col_rec_tat Time elapsed between collect time and receive time, range: 0 - 61370.2, units = hours; type: numeric

rec_ver_tat Time elapsed between receive time and verification time, range: -18.6 - 218.2, units = hours; type: numeric ...



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Software development concepts

Learning objectives:

- Familiarize with general good coding practices
- Learn about documentation standards
- Things to avoid when programming in R
- Learn how to debug and troubleshoot



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

What we learned today:

- What is R
- Basic syntax, data types
- Data Manipulation and Visualization
- Package Installation



What's next?

Statistics in R Workshop (Part 2)

- Data Wrangling
- Regression
- Statistical Analysis

Statistics in R (Part 2)	Feb. 20 1PM-5PM	EDUC 133	Open
--------------------------	-----------------	----------	----------------------



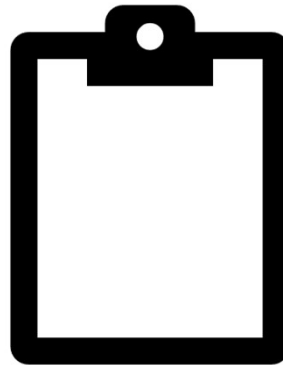
Thank you for attending!

1



Scan the QR code to confirm you attended today's workshop.

2



Fill out the feedback survey in the next 72h.

3



Get recognition for this workshop on your co-curricular record.



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant

Useful links

- [R software project](#)
- [RStudio Cheatsheet](#)
- [R ggplot2 Cheatsheet](#)
- [R dplyr Cheatsheet](#)
- [More resources](#)



McGill

Quantitative Life
Sciences

Sciences quantitatives
du vivant