

# DSA2101

## Essential Data Analytics Tools: Data Visualization

Yuting Huang

AY23/24 Semester 2

Course Policies

# The teaching team

## **Instructor:**

- ▶ Dr. Huang Yuting ([yhuang@nus.edu.sg](mailto:yhuang@nus.edu.sg))
- ▶ Office: S16 04-01
- ▶ Office hour: In-person and by appointment

## **Teaching assistants (TAs):**

- ▶ Consultation: By appointment via Zoom/MS Teams.
- ▶ Contacts of the TAs will be announced before Week 3.

# Topics covered

Here is a list of topics we plan to cover this semester:

1. R programming
2. Importing data into R
3. Data manipulation with R
4. Principles of data visualization
5. The grammar of graphics
6. Exploring data through visualization

# Tentative teaching plan

- ▶ Weeks 1–2: Basics in R programming
- ▶ Weeks 3–4: Importing data into R
- ▶ Week 5: Data manipulation
- ▶ Week 6: Tidy data
- ▶ Week 7: Relational data
- ▶ **Week 8: In-class midterm test**
- ▶ Week 9: Principles of data visualization
- ▶ Weeks 10–11: The grammar of graphics
- ▶ Weeks 12–13: Exploring data through visualization

## Lectures

- ▶ Mondays and Wednesdays **from 8 to 9:30 am** at LT32.
- ▶ Although lectures will be recorded and web-cast, we encourage you to attend the live lectures in person.

**Tutorials** begin in Week 3.

- ▶ Attendance is compulsory.
- ▶ You will also need to bring your own laptop.
- ▶ Tutorial worksheets will be released one week in advance. You are expected to **work on the worksheet beforehand**.

# Evaluation components

▶ Tutorial	5%
▶ DataCamp assignments	10%
▶ Group project	15%
▶ Midterm test	30%
▶ Final exam	40%

# Tutorial

**Tutorial attendance is compulsory** - worth 5% of your final grade.

- ▶ Please adhere to the tutorial session you selected.
- ▶ In case of illness, send your medical certificate (MC) to your TA. You will be excused from attending the tutorial on that day.

# DataCamp assignments

As part of the class, you are required to complete DataCamp assignments (10% of final grade).

- ▶ The activation link is available on Canvas. Use it to activate your DataCamp Classroom account.
  - ▶ Sign in with your NUS email (with domain @u.nus.edu).
- ▶ After activation, you will have access to courses there for 6 months. Do make full use of it!



For each DataCamp assignment, you will receive a binary grade – either full or none.

- ▶ They are graded based on completion by the due date, not by XP points.

The screenshot displays the DataCamp interface. On the left is a dark sidebar with navigation options: 'Dashboard', 'Members', 'Teams', 'Settings', and a 'LEARN' section containing 'Custom Tracks', 'Assignments' (highlighted), 'Leaderboard', 'Insights & Analytics', 'Reporting', 'Custom Reports', and 'Skill Matrix'. The main area is titled 'Assignments / Everyone' and has tabs for 'ACTIVE', 'PAST DUE', and 'ARCHIVED'. Below the tabs is a search bar labeled 'Search assignments...'. A table lists active assignments with the following columns: 'TITLE', 'ASSIGNED TO', 'STATUS', 'DUE DATE TYPE', 'CREATED AT', and 'EXPIRES AT'. The table contains seven rows of assignment data.

TITLE	ASSIGNED TO	STATUS	DUE DATE TYPE	CREATED AT	EXPIRES AT
Introduction to R Course	Organization	Active	Fixed	Jan 9, 11:55 +08	Feb 2, 23:59 +08
Reporting with R Markdown Getting Started with R Markdown Chapter	Organization	Active	Fixed	Jan 9, 11:56 +08	Feb 9, 23:59 +08
Introduction to Importing Data in R Course	Organization	Active	Fixed	Jan 9, 11:56 +08	Feb 23, 23:59 +08
Data Manipulation with dplyr Course	Organization	Active	Fixed	Jan 9, 11:57 +08	Feb 23, 23:59 +08
Joining Data with dplyr Course	Organization	Active	Fixed	Jan 9, 11:57 +08	Mar 8, 23:59 +08
Reporting with R Markdown Course	Organization	Active	Fixed	Jan 9, 11:58 +08	Mar 15, 23:59 +08
Introduction to Data Visualization with ggplot2 Course	Organization	Active	Fixed	Jan 9, 11:58 +08	Mar 22, 23:59 +08

# Group project

Project guideline will be released after the midterm exam.

- ▶ You will work as a team of **3 ~ 5 members**.
- ▶ More details will be announced in due course.

# Late submission policy

- ▶ Late submissions will not be graded.
- ▶ Email submissions will not be graded.
- ▶ Only submissions through Canvas will be graded.

# Software

For this class, we will be using the following software.

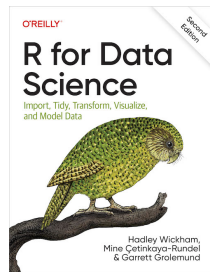
1. R
2. RStudio
3. Exemplify

Make sure that you have a laptop (Windows or Mac), and the latest versions of software above.

# Main references

## R for Data Science

- ▶ Online textbook:  
<https://r4ds.hadley.nz/preface-2e>
- ▶ Great resource for R using the `tidyverse` syntax.
- ▶ Lots of examples and exercises.



Due to time constraint, we won't be able to cover all the chapters. If you are serious about the field of data science, we encourage you to read through the book as carefully as you can.

# How to reach us?

You can talk to us after lectures and tutorials.

- ▶ Or email us using your NUS email account.

## How to ask a coding question?

- ▶ Copy and paste your code and the entire error message.
- ▶ A code that allows us to reproduce the error (i.e., a minimal working example) is ideal.
- ▶ **Do not send screenshots.**

R-Bloggers: How to ask good questions that prompt useful answers?

## BAD ways of asking questions.

```
0- ***[r setup, include=FALSE]
7-
8-
9- library(readxl)
10- data_read_excel("../data/t1-9.xls")
11-
12-
13- ***

Error in inDL(x, as.logical(local), as.logical(now), ...) :
  unable to load shared object 'C:/Users/Spencer/Documents/R/win-library/4.1/vctr/Libs/x64/v
  LoadLibrary failure: The specified procedure could not be found.
```



```
Error in 'mutate()':
! In argument: 'BMI = mass/(height_m^2)'.
Caused by error:
! object 'height_m^2' not found
Backtrace:
1. dplyr::mutate(starwars_small, height_m = height/100, BMI = mass/(height_m^2))
2. dplyr::mutate.data.frame(starwars_small, height_m = height/100, BMI = mass/(height_m^2))
3. dplyr::mutate.cols<data, dplyr_quosures(...), by>
4. dplyr::mutate_col(dots[[1]], data, mask, new_columns)
5. mask$eval_all_mutate(quo)
6. dplyr (local) eval()

Error in mutate(starwars_small, height_m = height/100, BMI = mass/(height_m^2)) :
  Caused by error:
! object 'height_m^2' not found
```

## GOOD way of asking questions.

I have an RScript file (let's call it `main.R`) which has a reference to another file, using the below code:

```
source("functions.R")
```

But, when I run the RScript file, it complains with the below error:

```
Error in file(filename, "r", encoding = encoding) :
  cannot open the connection
In addition: Warning message:
In file(filename, "r", encoding = encoding) :
  cannot open file 'functions.R': No such file or directory
```

I am sure, my `main.R` file is next to `functions.R` in the same directory. I can call the `functions.R` in the `Rmd` (RMarkdown) file which exist in the same directory

# Good ways of asking questions

1. Be clear and concise.
  - ▶ What are you trying to achieve? What is the issue you are encountering?
2. Provide context, show the code
  - ▶ Copy and paste the actual line of code that is causing the problem.
3. Explain what you have tried.
  - ▶ Mention any attempts to solve the problem.
4. Be specific about the error.
  - ▶ Include error message as text rather than screenshots.