

# Business Analytics Practicum (MGT 4803)

Zhaohu (Jonathan) Fan

Information Technology Management

Scheller College of Business

Georgia Institute of Technology

January 11, 2024

# Team formation goals

- Creating teams mixing business, engineering, and science majors
  - We aim for diverse skill sets in each group.
  - Groups will consist of 3-5 people.
    - Please click on the link provided here (Google form)

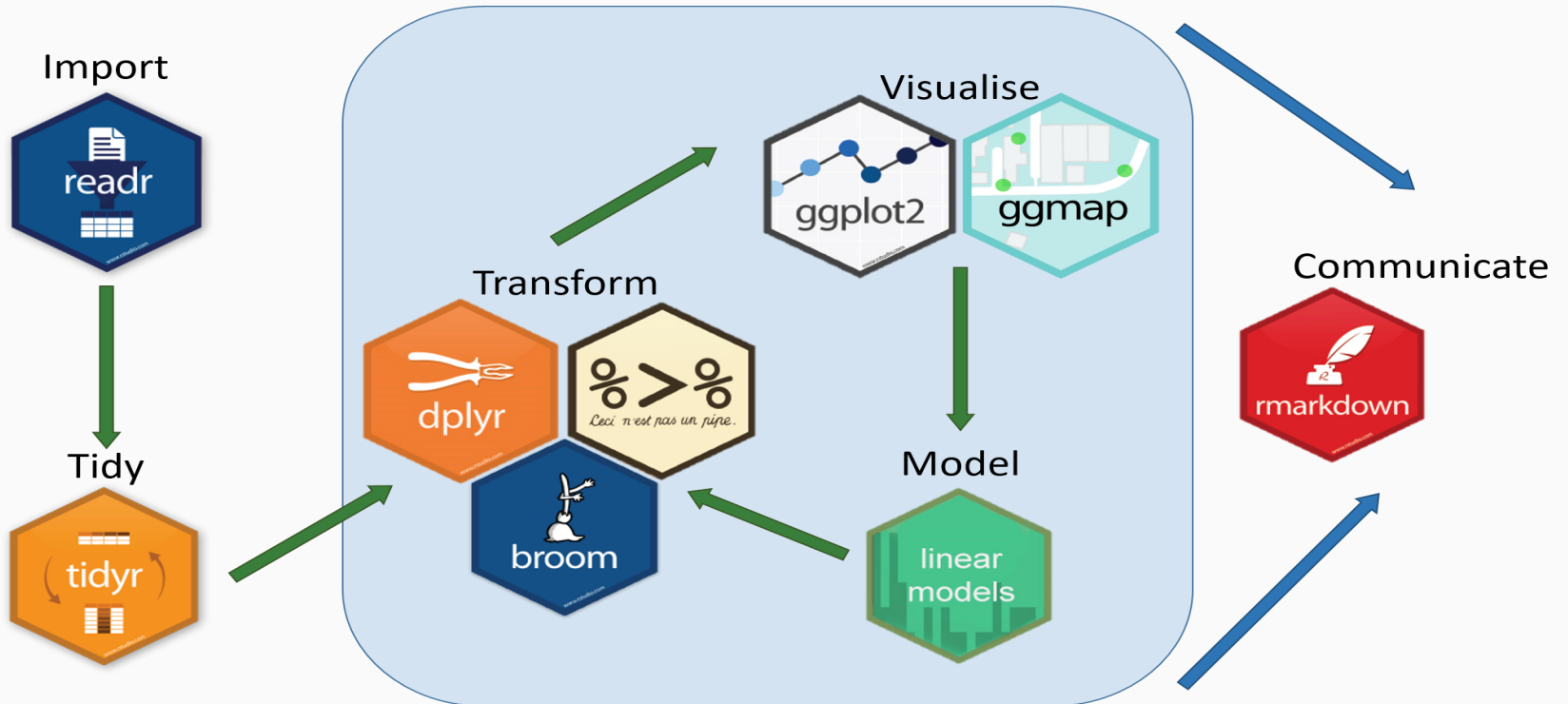
# What's happening next Tuesday?

- Projects leverage data engineering, data exploration and analysis, statistical modeling (**Python/R**), data visualization and dashboard (**Tableau**), and automation techniques.
- Projects presentation by Sponsors (Next Tuesday, Jan.16th)
  - **The Home Depot** (5:00 pm-5:35pm)
  - **Sparck Technologies** (5:40 pm-6:15pm)
- Post-presentations: Teams are required to submit their rank-ordered list of project preferences.
  - The instructor will assign each student team to a project.

# How it works?

- **Small team collaboration (3-5)**
- **Support and Guidance:**
  - Provided by the instructor, a TA (1st-year MBA), and a group of 2nd-year MBA coaches.
    - Each team will have a designated MBA coach.
    - **Tuesdays** are allocated for meetings with student teams to review weekly project progress.
    - **Thursdays** are dedicated to hands-on tutorials, and practical demonstrations.

# Data science pipeline



Step 1: Reading data sets using Python and R

# Reading data sets using Python



- **Pandas** is a library for data manipulation **and data analysis**.
- **Numpy** is an **array manipulation** library, used for linear algebra, Fourier transform, and random number capabilities.
- **Matplotlib** is a library which generates **figures** and provides graphical user interface toolkit.



# Car sales example

Diagram illustrating a car sales dataset with annotations for indexing and data access.

**Column (axis = 1)**

**Index number (starts at 0 by default)**

**Row (axis = 0)**

**Data**

**Column name**

	<b>Make</b>	<b>Colour</b>	<b>Odometer</b>	<b>Doors</b>	<b>Price</b>
0	Toyota	White	150043	4	\$4,000
1	Honda	Red	87899	4	\$5,000
2	Toyota	Blue	32549	3	\$7,000
3	BMW	Black	11179	5	\$22,000
4	Nissan	White	213095	4	\$3,500

The diagram shows a table with 5 rows and 6 columns. The first column is labeled 'Index number (starts at 0 by default)' and contains values 0, 1, 2, 3, 4. The second column is labeled 'Column (axis = 1)' and contains the column names: 'Make', 'Colour', 'Odometer', 'Doors', 'Price'. The first row is highlighted with a green border, and its cells (0, Toyota, White, 150043, 4, \$4,000) are also highlighted with a green border. An arrow labeled 'Data' points from the 'Data' label to the 'Price' cell in the first row. The label 'Column name' is positioned to the right of the 'Price' column header.

# Car sales example

- Please click on the link provided below.
  - [Hands-on tutorials \(Car sales example\)](#)

# Reading data sets using R

- Read the CSV File: Use the `read.csv()` function to load the CSV file into a data frame.
- View the Data: Use functions like `head()` to view the first few rows of the data frame.

## R code

```
# Load the necessary library  
library(readr)  
# Read the CSV file  
car_sales_data ← read.csv("car-sales.csv")  
# Display the first few rows of the data frame  
head(car_sales_data)
```