# Data Importation

## Using R and RStudio

Dr. Austin Brown

Kennesaw State University

# Data Importation

- **Objective**: Learn how to import data from CSV and XLSX files into RStudio
- **Importance**: Importing data is a fundamental step in data analysis and data science workflows.

# Importing CSV Files in RStudio

- In the Module 1 folder in D2L, there is a CSV file called "Cars.csv".
- Download the file to your computer and then upload to your Posit Cloud project folder. If you need help understanding how to upload the file, please refer to the video tutorial at this link.
- After uploading the file, we can use the `read.csv` function to help us import the data into RStudio.

# Importing CSV Files in RStudio

- We can think of functions as tools that help us perform specific tasks. The `read.csv` function is a tool that helps us import data from a CSV file.
- Each function has specific arguments that we must specify in order to the function to execute the way we expect it to.
- For the `read.csv` function, we must specify the file path to the CSV file we want to import.

# Importing CSV Files in RStudio

```
## Importing Data from CSV ##
cars <- read.csv("Cars.csv")
```

- Now in the upper right hand corner of our RStudio window, we should see a data object called `cars` which contains 32 observations (or rows) and 12 variables (or columns).

# Importing XLSX Files in RStudio

- We can also import Excel files into RStudio. For this example, please download the Penguins.xlsx file from D2L and upload to your Posit Cloud project folder.
- This time, we will make use of a function called `read_xlsx`. However, this function is not part of base R, but instead is part of an external package called `readxl`.
- To use the `readxl` package, we must first install it using the `install.packages` function.

```
## Installing readxl ##
install.packages('readxl')
```

# Importing XLSX Files in RStudio

- Once installed, we can load the `readxl` package using the `library` function.
  - Loading a package is similar to opening a toolbox. It gives us access to all the tools (or functions) that are part of the package.

```
library(readxl)
```

# Importing XLSX Files in RStudio

- Once the package is loaded, we can use the `read_xlsx` function in a similar manner to how we used the `read.csv` function.

```
## Importing Data from XLSX ##
penguins <- read_xlsx("Penguins.xlsx")
```

# Exploring Imported Data

- Once we have imported data, it is important to explore it to ensure that it was imported correctly.
- There are many ways to go about this using R, but one of the simplest ways is to use the `head` function. The `head` function displays the first few rows of a data frame.
- Let's use the `head` function on both the `cars` and `penguins` dataframes:

# Exploring Imported Data

```
## Use your head function: ##
head(cars)
```

```
                  X  mpg cyl disp  hp drat    wt  qsec vs am gear carb
1         Mazda RX4 21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
2     Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
3        Datsun 710 22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
4    Hornet 4 Drive 21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
5 Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
6           Valiant 18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

# Exploring Imported Data

```
head(penguins)
```

```
# A tibble: 6 x 8
  species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g
  <chr>   <chr>              <dbl>         <dbl>             <dbl>       <dbl>
1 Adelie  Torgersen           39.1          18.7               181        3750
2 Adelie  Torgersen           39.5          17.4               186        3800
3 Adelie  Torgersen           40.3          18                 195        3250
4 Adelie  Torgersen           NA            NA                  NA          NA
5 Adelie  Torgersen           36.7          19.3               193        3450
6 Adelie  Torgersen           39.3          20.6               190        3650
# i 2 more variables: sex <chr>, year <dbl>
```