

# DATA 106 - Lab 2

*Jillian Morrison*

*September 16, 2019*

## General rules

- For some questions, the needed methods may not have been covered in class. For them, please do some research to solve them.
- You must show your work in order to get points. Providing correct answers without supporting codes or intermediate steps does not receive full credit.
- You must submit both the R file as a .R file and the Assignment file as a PDF. For the Assignment file include the code, the output and explanations (if necessary).

## Questions

1. Using the `Cars93` dataset in the `MASS` package, do the following:
  - a. Create two new data frames for USA and non-USA cars. Name the new datasets `USA` and `Non` respectively. You can use the filter function in `{dplyr}` package.
  - b. Find the cheapest US car and non-US car. Here, consider using the `filter()` function along with `min()` and `max()` functions to find minimum and maximum values.
  - c. find the most expensive USA and non-USA Car
  - d. The `Type` variable classifies the type of market the car is aimed at. Find the cheapest car in each type, and the car with the greatest fuel efficiency. (Hint: In part a, you separated by a specific variable and b and c, you filtered to find the cheapest car in each group. You will need to combine both in this part. However, instead of using filter to separate by a specific variable (part a), consider using `group_by()` in `{dplyr}`. You will also want to use piping (`%>%`) to make this easier.)
  - e. Compute the mean horsepower for each type. (Hint: Still using piping (`%>%`), try using `group_by()` and `summarize()`. See: <https://datacarpentry.org/R-genomics/04-dplyr.html> for more info on summarize. Note: `na.rm=TRUE` removes missing values from the dataset before making calculations.).
  - f. Save the resulting table in part e to a .csv file called `Summary.csv`. You will upload this file to moodle along with your R script and pdf. (Hint: Remember to set your working directory so you know where your file is saved. Also make sure that you save your table as an object in R so you can save it to a csv.)
2. Using the `gapminder` dataset from the `{gapminder}` package, do the following:
  - a. Save the dataset to an object called "gap" and convert it to a dataframe
  - b. How many different countries are covered by the data. List them.
  - c. Extract all the 2002 life expectancies for African countries

(Note: the `select()` function is available in both `{dplyr}` and `{MASS}` packages. To specify you want to use the `{dplyr}` package, use instead `dplyr::select()`)

(Other Note: here you have 2 conditions - Africa and 2002. As well, you are to produce a table with only the necessary variables - country and life expectancy)

- d. Extract the 2005 population for African countries
- e. Extract the country with the highest gdp value for 2007 for each continent.