

Module 1 - MATH 241

Contents

I. Data Wrangling (Homework)	3
II. BibTeX References and Citations in R Markdown (Instructional)	4
III. Block Quotes in R Markdown (Instructional)	5
References	6

Posted: February 5, 2022

Due: February 11, 2022

Instructions:

- **Please knit your R Markdown file as PDF file.** Don't put your name in any part of the document. Your document upload will correspond to your name automatically in Gradescope.
- **Please provide complete solutions for each problem.** If it involves mathematical computations, explanations, or analysis, please provide your reasoning or detailed solutions.
- **It is recommended that you work on the module incrementally.** The modules are designed for you to work on it at anytime until the designated deadline. This method will let you transfer what you have learned from lectures and mini-assignments in small parts rather than stressing on completing the entire assignment at the last minute.
- **Note that some problems have multiple solutions or ways to solve it.** Make sure that your solutions are clear enough to showcase your work and understanding of the material.
- **Creativity and collaborations are encouraged.** Use all of the resources you have and what you need to complete the module. Each student must take personal responsibility and submit their work individually. Please abide by the Reed College Honor Principle.
- If you can't figure out why a code chunk is preventing you from knitting the document, replace "r" at the top of the code chunk with "r eval = FALSE, echo = TRUE." The code will not be executed, but it will be printed in your pdf, earning you some partial credit.
- Please follow the general module guidelines written in the course website. If you have any questions, please send them to the instructor as a direct message on Slack or through email.

R Packages:

Below are pre-loaded general packages required for this module assignment. You can load more packages here or throughout the module if necessary. Note that you need to install R packages before you can use them. You can use the `install.packages()` in the R console, or go to the “Tools” tab and click “Install Packages...” in R Studio.

```
# pre-load packages here  
library(dplyr)  
library(bookdown)  
library(tidyverse)
```

I. Data Wrangling (Homework)

Materials

This part of the module uses the following packages and/or data files.

- The `nycflights13` package: This package contains information about all flights that departed from NYC (e.g. EWR, JFK and LGA) to destinations in the United States, Puerto Rico, and the American Virgin Islands) in 2013: 336,776 flights in total. This package provides the following data tables. Note that you can use the `?` or `help()` functions to understand the variables involved in these tables.
 - `flights`: all flights that departed from NYC in 2013
 - `weather`: hourly meteorological data for each airport
 - `planes`: construction information about each plane
 - `airports`: airport names and locations
 - `airlines`: translation between two letter carrier codes and names

```
# load the package
library(nycflights13)
```

Exercises

The exercise problems shown below was taken and slightly modified from your textbook Modern Data Science with R [Ch. 4.4](#) and [Ch. 5.5](#).

1. The `nycflights13` package includes a table (`weather`) that describes the weather during 2013. Use that table to answer the following questions:
 - a. What is the distribution of temperature in July, 2013? Identify any important outliers in terms of the `wind_speed` variable.
 - b. What is the relationship between `dewp` and `humid`?
 - c. What is the relationship between `precip` and `visib`?
2. Use the `nycflights13` package and the `flights` and `planes` tables to answer the following questions:
 - a. How many planes have a missing date of manufacture?
 - b. What are the five most common manufacturers?
 - c. How many airplanes that flew from New York City are included in the `planes` table?
3. Use the `weather` table in the `nycflights13` package to convert the `temp` variable from Fahrenheit to Celsius. Create a user-defined function that converts Fahrenheit to Celsius and then apply the function. Below is the equation of the conversion formula. Using the converted values, either create a new column within the table or make a new separate variable. Provide an output showing the results.

$$f_C(x_F) = \frac{5}{9}(x_F - 32)$$

II. BibTeX References and Citations in R Markdown (Instructional)

Materials

This part of the module uses the following files.

- The file named `references.bib` contains an entry of a BibTeX formatted reference shown below. The bib key to this entry is `R-base`.

```
@Manual{R-base,  
  title = {R: A Language and Environment for Statistical Computing},  
  author = {{R Core Team}},  
  organization = {R Foundation for Statistical Computing},  
  address = {Vienna, Austria},  
  year = {2016},  
  url = {https://www.R-project.org/},  
}
```

- The file named `apa.csl` is a CSL (Citation Style Language) file which was downloaded from [Zotero Style Repository](#). It contains code to generate citations and references within your document using the American Psychological Association (APA) 7th edition style.

Exercises

1. Visit [Google Scholar](#) and search for “data science.” Using the first item in the results, click the “Cite” button. Look for the BibTeX link, where it leads to another page showing the BibTeX formatted citation of the reference. Copy and paste the citation into a new line in the `references.bib` file. Show what you copied and pasted by filling in the verbatim code block below.

put your BibTeX formatted citation here and in the `.bib` file

2. We can cite references listed in our BibTeX file by using the bib keys of the references. We can start using those bib keys, using the following syntax:
 - `[@key]` for single citation or `[@key1; @key2]` for multiple citations
 - `@key` is also a valid syntax when using it in a sentence

For citing the `R-base` bib key, we use `[@R-base]`, which is rendered as (R Core Team, 2016). Notice the `@` key character before the bib key to indicate that we are citing an entry from the BibTeX file which will show in the [References](#) section.

Identify the bib key of the BibTeX reference you just added and cite it in two different ways: (a) as a single citation, and (b) use it in a sentence. You can use a made-up sentence for the purpose of this exercise.

III. Block Quotes in R Markdown (Instructional)

Materials

We can add quotes from sources by using the block quotes feature in R Markdown. This is similar to a code block but block quotes are used to show text. Below is an example.

“Sometimes the best way to solve your own problems is to help someone else.”
— Iroh, (*Avatar the Last Airbender*, 2005-2008)

Exercises

1. Use the block quotes feature to quote something from your favorite movie or series.

References

Avatar the last airbender. (2005-2008). Premiered on Nickelodeon.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>