An Educator’s Guide to the Open Case Studies

February, 2022

Table of Contents

# About this guide

The [Open Case Studies project](https://www.opencasestudies.org), developed at the [Johns Hopkins Data Science Lab](https://jhudatascience.org/), is an education platform that provides self-contained, multimodal, peer-reviewed, and open-source guides for real-world examples for active experiences of complete data analyses.

Case study guides can be used 1) in the classroom (either onsite or online) by engaging students to actively participate in a hands-on problem-solving experience, and 2) outside of the classroom by providing an archive of stand-alone examples of best practices.

To help guide educators on how to most effectively use the Open Case Studies (either in or outside of the classroom), this guide documents examples of how to use the case studies, various entry points to using the case studies (including an R package to enable modular use of the case studies), how to modify and adapt components of the case studies for the classroom, and how to contribute new case studies.

# 1 Introduction

## 1.1 Motivation

The [Open Case Studies project](https://www.opencasestudies.org), developed at the [Johns Hopkins Data Science Lab](https://jhudatascience.org/), is an education platform that provides self-contained, multimodal, peer-reviewed, and open-source guides for real-world examples for active experiences of complete data analyses. The intention of this guide is to provide instructors more information about how to make the most of our case studies.

## 1.2 Target Audience

This guide is intended for educators who are interested in using Open Case Studies for instruction.

Elements of the case studies can be helpful for instructors who teach:

* High school students
* Undergraduate students
* Graduate students

Elements of the case studies can assist with teaching courses about the following topics:

* Data science
* Statistics
* Public health
* Programming
* Technical writing

… and more

## 1.3 Curriculum

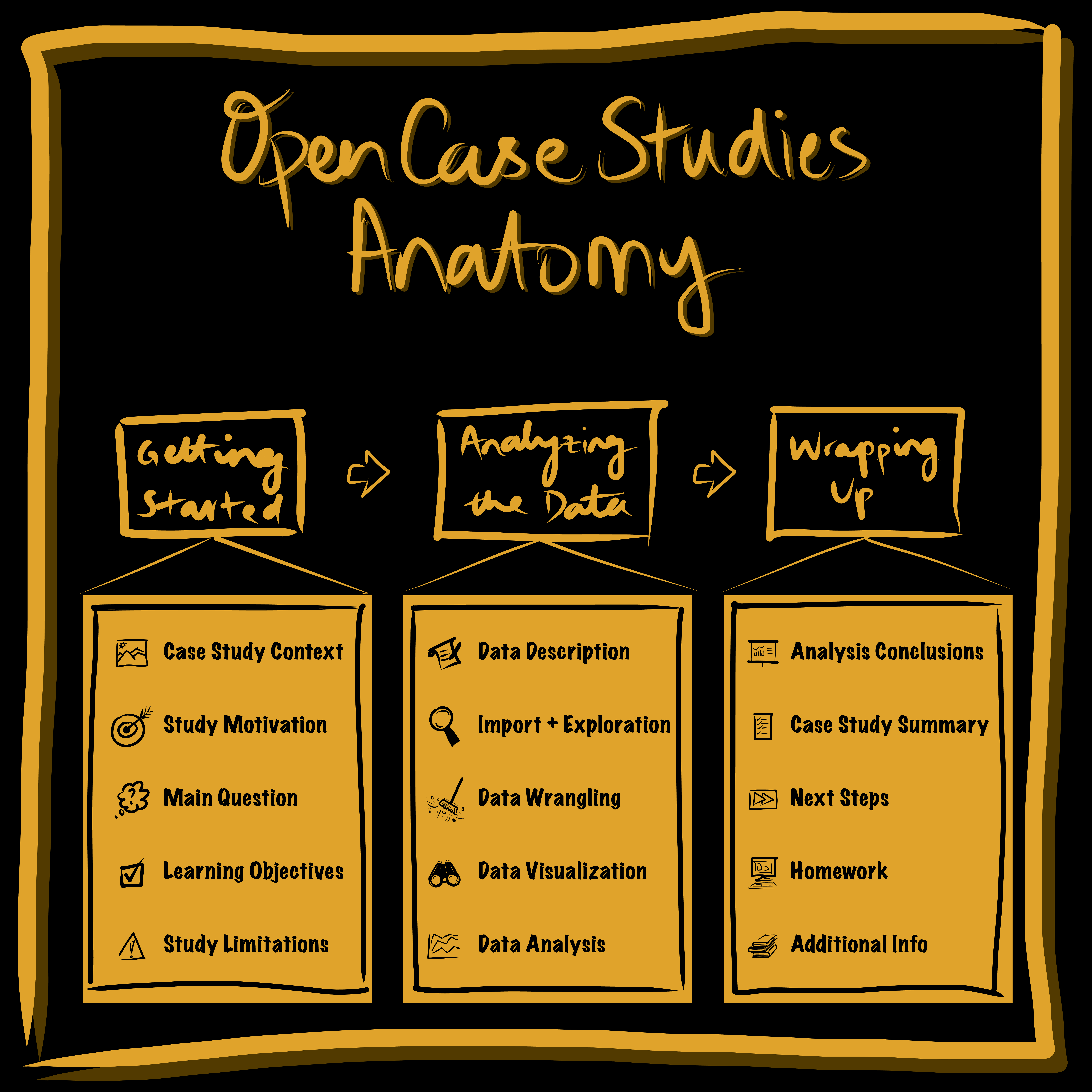
This guide documents:

* The Open Case Studies philosophy
* The general structure of case studies
* Various entry points to using the case studies (including an R package to enable modular use of the case studies)
* Examples of how to use the case studies
* How to modify and adapt components of the case studies for the classroom
* How to contribute new case studies.

## 1.4 Open Case Studies Philosophy



## 1.5 Open Case Studies Anatomy



Case studies are designed to be self-contained examples of complete data analyses. They start with a subject area of interest and a problem definition.

Each case study is composed of three main stages:

* Stage 1: Getting Started

In the first stage, students learn about the subject area they will be investigating in their data analysis. This can include historical perspectives, previous literature, and current overarching questions in the field. Ideally, this will provide motivation as to why the data analysis they will be performing is interesting and useful. Students will then further refine questions of interest and define a specific question to be investigated by the data at hand. This question will define the statistical and data science learning objectives for the case study. Finally, students will explore the limitations of their investigation and discuss why their data analysis may or may not be able to fully answer the main question of interest.

- Case study context  
- Study motivation  
- Main question  
- Learning objectives   
- Study limitations

* Stage 2: Analyzing the Data

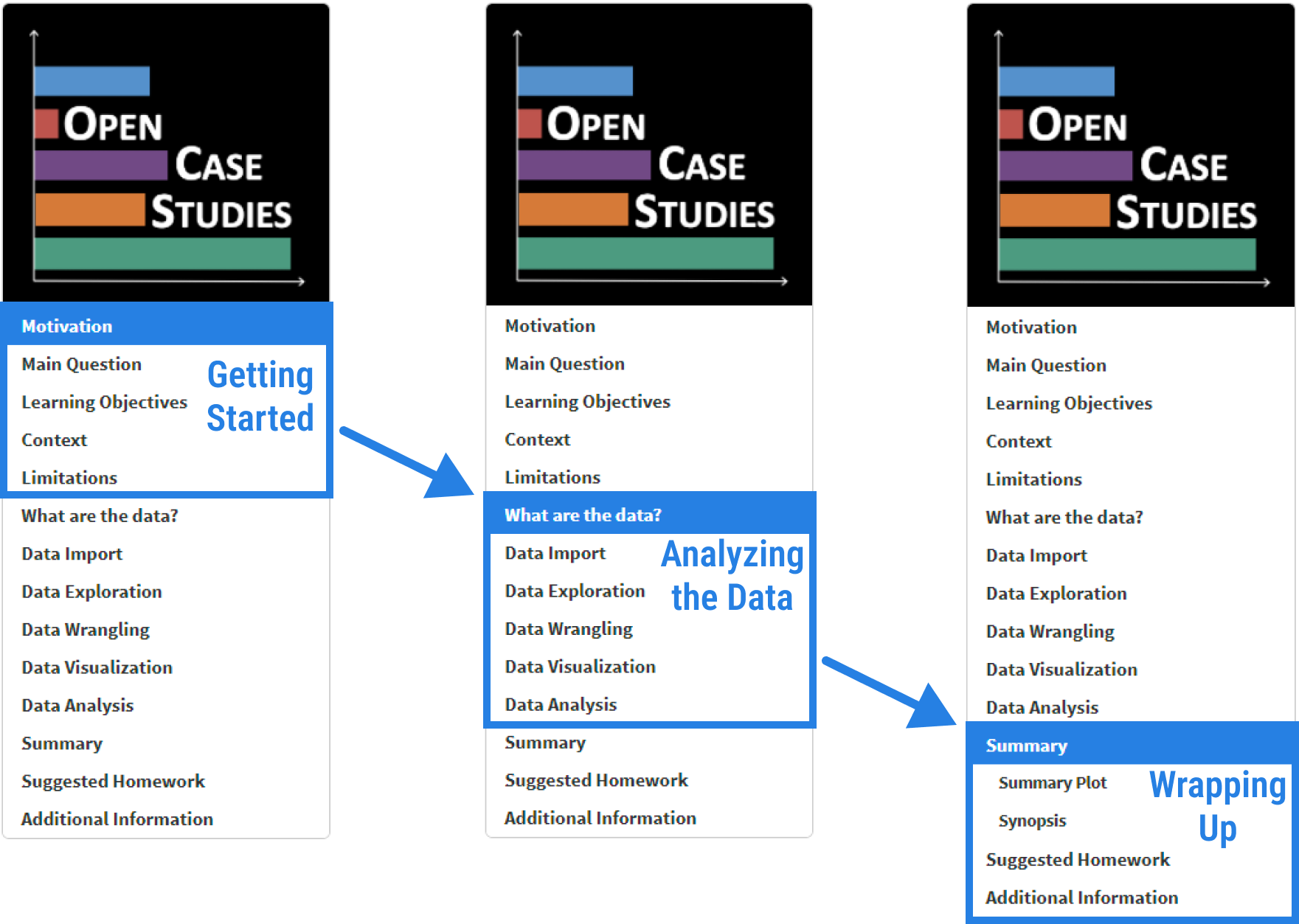
This stage includes the bulk of the data analysis. It begins with a detailed description of the data used in the case study, how it was generated, and where is was obtained. This sets the stage for a walk-through of the data import procedure and exploration of the available data. From there, the specific subsets of the data relevant to answering the main question of interest can be identified and the data can be cleaned and wrangled to give the data in the required format. The cleaned data can then be visualized and analyzed to answer the study question.

- Data description  
- Import and exploration   
- Wrangling   
- Visualization   
- Analysis

* Stage 3: Wrapping-Up

In the final stage of the case study, students synthesize conclusions to the main study question based on the results of their data analysis. They summarize the data exploration and analysis steps they took and discuss potential next steps they could take based on their results. Finally, additional analyses are suggested as homework and additional information relevant to the case study is provided.

- Analysis conclusions   
- Case study summary  
- Next steps   
- Homework

This anatomy is reflected in the table of contents: 

## 1.6 Starter Kit

Open Case Studies are designed to be beginner friendly. Users can work through an entire case study with no prerequisite knowledge in programming, public health, or statistics. However, the appropriate technology, software, and a basic familiarity with R Studio is required. This section will detail what students will need to be able to jump into a case study.

### 1.6.1 Technical Requirements

(avocado note from slack: what physical resources are needed - software installed - what type of computers - think about interactive case studies)

All case studies use the R statistical programming language for data analysis. R is available for Windows, MacOS, and Linux and can be downloaded from the [The Comprehensive R Archive Network (CRAN)](https://cran.r-project.org/). While there is no specific R version requirement for the case studies, the OCSdata package, which can be used to get and load the data, does require R 3.5 (avocado). Furthermore, R packages used to run specific analyses in each case study may have their own R version requirements. R version requirements may be checked in the sessionInfo() section in each case study.

Cheatsheets detailing how to use common packages and functions are available at the [RStudio website](https://www.rstudio.com/resources/cheatsheets/).

### 1.6.2 RStudio

To work with R, we recommend the RStudio Integrated Development Environment. RStudio includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. RStudio is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or in a browser connected to RStudio Server. More information and program installation instruction are available at the [RStudio website](https://www.rstudio.com/products/rstudio/).

For reproducible data analyses and easy publishing of reports and presentations, we recommend using RMarkdown. More information about RMarkdown is also available at the [RStudio website](https://rmarkdown.rstudio.com/index.html).

### 1.6.3 GitHub

Open Case Studies is hosted on GitHub. While GitHub is not needed for users to access the case studies, users can use GitHub to streamline download and modification of case studies. Check out the GitHub Organization and OCSdata sections in Chapter 2 for more details on how to use GitHub to access case studies and modify content for personalized use. An introduction on how to get started with GitHub is available at [Happy Git](https://happygitwithr.com/index.html).

## 1.7 Feedback

We are continually working to improve the Open Case Studies project to address learner and educator needs. Feedback is essential for this goal. If you use Open Case Studies material, we would love to hear from you! Please submit your feedback in the [user experience survey](https://www.opencasestudies.org/). We appreciate feedback about all aspects of the Open Case Studies user experience including but not limited to technical challenges, potential improvements, or new ideas on how to use the case studies in the classroom.

## 1.8 Contact Us

[Email:] [opencasestudies@gmail.com](mailto:opencasestudies@gmail.com)

[Twitter:](https://twitter.com/opencasestudies)

# 2 Open Case Study Infrastructure

## 2.1 Learning Objectives

In this chapter we will discuss the overall infrastructure of the Open Case Studies platform, which includes:

* our website
* an Open Case Studies search tool
* our GitHub organization
* an R package called OCSdata

## 2.2 OCS Website

### 2.2.1 Open Case Studies mission

The [Open Case Studies website](https://www.opencasestudies.org/) describes the mission of the Open Case Studies project, the history of its inception, and current and previous members of the OCS team, among other information.

### 2.2.2 Open Case Studies feedback

The website also includes a link to a [user experience survey](https://www.opencasestudies.org/) (takes up to 10 minutes to complete), which helps us learn more about how to improve the data science education experience. Part of this includes getting a better understanding of who is using our case studies and how so that we can better design our case studies. We would greatly appreciate you filling it out if you have the time!

### 2.2.3 Open Case Studies search tool

Finally, the website includes the OCS case study search tool to aid instructors in finding appropriate case studies for their learning objectives. The search tool is described in more detail below.

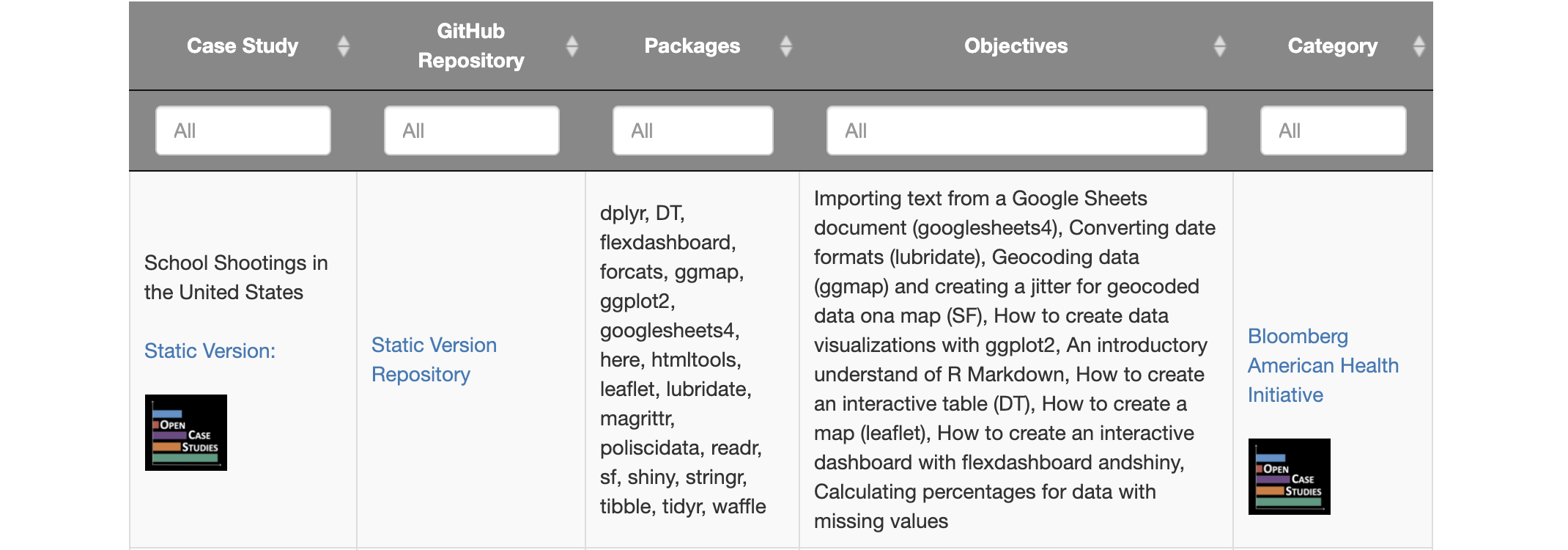
## 2.3 OCS Case Study Search Tool

A brief description of the subject areas of each case study is included at the [Bloomberg American Health Initiative website](https://americanhealth.jhu.edu/open-case-studies).

To aid with case study selection, we created the [OCS case study search tool](https://www.opencasestudies.org/#searchtab). This search tool is designed to aid instructors in identifying appropriate case studies for their learning objectives. The tool consists of a table with searchable columns, where each row is an individual case study.

The main two columns likely to be helpful in identifying appropriate case studies are the **“Packages”** and **“Objectives”** columns. The **Packages** column details all the R packages used in the case study, and can help identify if a case study teaches a specific data import, wrangling, analysis, or visualization skill. The **Objectives** column details the learning objectives of each case study (e.g. importing data from PDF files, reshaping data, specific statistical analysis, etc.).

The table also includes links to both the static and interactive versions of the case studies and their GitHub repositories. Below is an example of the information included in the search tool results for the case study on [School Shootings in the United States](https://www.opencasestudies.org/ocs-bp-school-shootings-dashboard/).



## 2.4 GitHub Organization

The GitHub repository contains all the materials needed for the case study. This includes the case study text to be distributed to students, the data used in the case study (discussed below), additional documents and references, and brief guidelines on case study use.

Data included in the GitHub repository is available in multiple formats to facilitate modularization of the case studies as described below. To use the case study data, you can download the GitHub repository directly or use the OCSdata R package described below.

(avocado Add figure 2.4 outlining repo structure)

## 2.5 OCSdata

To simplify the process of accessing the data required for each case study, we have created the OCSdata R package. Briefly, the OCSdata package creates a new folder called “OCSdata” where it downloads the data needed for a specific case study. Users can download the data in its original raw format or in various processed formats that correspond to different stages of data wrangling and cleaning. This allows users to perform the data exploration and wrangling or the data visualization and analysis sections of the case study without having to process the data from the raw files. For some of the case studies, the OCSdata package also downloads extra source data that is not used in the case study.

(Add table 2.1 summarizing data formats available for download via OCSdata)

### 2.5.1 Getting Started with OCSdata

The OCSdata package is available on the package repository CRAN and can be installed in R as follows:

[avocado R version requirement?]

install.packages("OCSdata") #only run once to install package  
library(OCSdata) #run every new R session to load package

### 2.5.2 Downloading raw data

The raw\_data function will download the raw data files that can be imported into R.

The first argument is the name of the case study. A list of case study names can be found in the package documentation [online](https://cran.r-project.org/web/packages/OCSdata/vignettes/instructions.html#casestudy) or by typing ?raw\_data in R.

The outpath argument is a string specifying the folder where the data should be downloaded. To download the data to a folder named “OCS\_data” in the current working directory, you can supply getwd() to the output argument. If nothing is provided for the argument, you will be prompted to enter 1, 2, or 3 to download the data into the current director, to specify the download path, or to cancel, respectively.

In the following example, we download the raw data for the “Opioids in the United States” case study to the current directory.

raw\_data("ocs-bp-opioid-rural-urban", outpath = getwd())

### 2.5.3 Downloading data in other formats

The OCSdata package can be used to download the data in various processed formats that may be helpful in skipping certain case study sections and focusing on data wrangling and/or analysis and visualization. All of the functions take the same arguments described above.

#### 2.5.3.1 Simpler import

The simpler\_import\_data function will download raw data files that have been converted to file formats that are easier to import into R, typically .csv. Some case studies offer this option when the original raw files require a more complicated import step.

simpler\_import\_data("ocs-bp-opioid-rural-urban", outpath = getwd())

#### 2.5.3.2 Importing data as R objects

The imported\_data function will download raw data files in .rda format. This means the data have already been imported into R objects. This can be used to skip the data import section and start directly with data wrangling. The R objects files can be imported into R by either double clicking on the files in Rstudio or using the load() function as follows.

imported\_data("ocs-bp-opioid-rural-urban", outpath = getwd()) #download data in .rda format   
file\_path = "~/OCS\_data/data/imported/land\_area.rda" #specify download directory   
load(file\_path) #load R objects

#### 2.5.3.3 Importing wrangled data

The following functions will download the data files that have already been wrangled and are ready to be analyzed. These come in both .csv and .rda formats.

Download as csv files:

wrangled\_csv("ocs-bp-opioid-rural-urban", outpath = getwd())

Downloading as R objects:

wrangled\_rda("ocs-bp-opioid-rural-urban", outpath = getwd())

### 2.5.4 Downloading extra data

Some case studies have extra data are not used in the case study but can be used to explore the case study subject from different perspectives. These data but can This data can be downloaded using the extra\_data function.

extra\_data("ocs-bp-opioid-rural-urban", outpath = getwd())

### 2.5.5 Downloading all case study data

The zip\_ocs function will download the all of the repository files in a .zip folder and unzip them into a specified directory. This includes the case study data in all the formats detailed above (raw, simpler\_import, imported, wrangled, and extra). It also includes the case study .Rmd file, which can be modified by instructors as needed.

zip\_ocs("ocs-bp-opioid-rural-urban", outpath = getwd())

### 2.5.6 Forking the case study repository

## R version 4.0.2 (2020-06-22)  
## Platform: x86\_64-pc-linux-gnu (64-bit)  
## Running under: Ubuntu 20.04.3 LTS  
##   
## Matrix products: default  
## BLAS/LAPACK: /usr/lib/x86\_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.8.so  
##   
## locale:  
## [1] LC\_CTYPE=en\_US.UTF-8 LC\_NUMERIC=C   
## [3] LC\_TIME=en\_US.UTF-8 LC\_COLLATE=en\_US.UTF-8   
## [5] LC\_MONETARY=en\_US.UTF-8 LC\_MESSAGES=C   
## [7] LC\_PAPER=en\_US.UTF-8 LC\_NAME=C   
## [9] LC\_ADDRESS=C LC\_TELEPHONE=C   
## [11] LC\_MEASUREMENT=en\_US.UTF-8 LC\_IDENTIFICATION=C   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## loaded via a namespace (and not attached):  
## [1] bookdown\_0.24 crayon\_1.4.2 digest\_0.6.25 R6\_2.5.1   
## [5] lifecycle\_1.0.1 magrittr\_2.0.2 evaluate\_0.14 pillar\_1.4.6   
## [9] rlang\_0.4.10 stringi\_1.5.3 fs\_1.5.2 ellipsis\_0.3.1   
## [13] vctrs\_0.3.4 rmarkdown\_2.10 ottr\_0.1.2 tools\_4.0.2   
## [17] stringr\_1.4.0 readr\_1.4.0 hms\_0.5.3 xfun\_0.26   
## [21] yaml\_2.2.2 compiler\_4.0.2 pkgconfig\_2.0.3 htmltools\_0.5.0  
## [25] knitr\_1.33 tibble\_3.0.3

# 3 Use of Open Case Studies

## 3.1 Learning Objectives

This chapter will provide guidance on how to use Open Case Studies in classroom settings. We will give a coarse guide on which case studies include material appropriate for beginner, intermediate, or advanced learners. We will present some examples of assignments that can accompany the case studies and the rubrics that they can be evaluated with. Furthermore, we provide examples of extensions to case studies that can serve as a template for homework assignments or independent student exploration.

The examples presented in this chapter are merely suggestions - modifications to the case studies to fit student needs are expected and encouraged! If you come up with a different way to use the case studies, please let us know what you come up with so that other educators may be inspired by your creativity.

### 3.1.1 Prerequisites

**Public Health Subject Matter**

The case studies in Open Case Studies generally deal with topics in public health. We do not require any prior knowledge on the public health subjects examined in the case studies. The Getting Started section in each case study (specifically, the case study context) will present the subject material relevant to understanding the data and the case study implications.

**Statistics**

The case studies cover a range of statistical approaches and skillsets. Some case studies will build on more foundational statistical concepts. The Experience Level categorization for each case study (see Section 3.3) will indicate the expected prior skills that the case study will expect. Furthermore, the exact skills that the case study will use are listed in the Case Study Search Tool under the Objectives column.

**Coding/Data Science**

All case studies use the R statistical programming language for data analysis. Some familiarity with R basics is expected for effective use of the case studies. However, depending on the data used and the extent to which it needs to be cleaned and processed before analysis, each case study may require experience with additional programming and data wrangling skills. This will be indicated by the Experience Level designation for the case study (see Section 3.3).

**Software**

All case studies use the R statistical programming language for data analysis. While there is no specific R version requirement for the case studies, the OCSdata package, which can be used to get and load the data, does require R 3.5 (avocado). Furthermore, R packages used to run specific analyses in each case study may have their own R version requirements. R version requirements may be checked in the sessionInfo() section in each case study.

### 3.1.2 Experience Level Descriptions

### 3.1.3 Experience Level Descriptions

The table below explains how we define the beginner, intermediate, and advanced experience levels.

| Experience Level | Description |
| --- | --- |
| Beginner | Little to no previous experience with coding and/or statistical analysis. |
| Intermediate | Familiar with at least one programming language and has experience working with data and statistics. |
| Advanced | Fluent in at least one programming language and likely familiar with more. Has a depth of experience working with data science projects. |

Typically, most middle/high school and first year undergraduate students will fit in the beginner category. Upperclassmen undergraduates and some graduate students are often at the intermediate level. Most advanced level students will be at the graduate level. However, this is a generalization, and a student may be considered beginner, intermediate, or advanced at any academic level depending on their independent studies and experiences.

## 3.2 Open Case Studies in the Classroom

The case studies are structured to support both partial and full use of a case study. Educators are also free to use case study data by itself.

### 3.2.1 Teaching a Full Case Study

The case studies are written to provide a cohesive story that simulates data science in the real world. Reading through an entire case study is an excellent exercise for students to experience a standard data science workflow and learn the best practices of today from start to finish. The following list provides a few examples of how educators have used a full case study in the past:

* Converted the case study into a slide deck (add link, source and example image of slide)
* Assigned students to read the case study and write a report as homework
* Suggested reading and supplemental material
* more?

(avocado add discussion of interactive case studies)

### 3.2.2 Teaching Part of a Case Study

Some educators may find that only certain sections of a case study are relevant to their course learning objectives. For these educators the case studies are structured to allow for modular use. The case study [GitHub repositories](https://github.com/opencasestudies) provide the appropriate data files to be used at the start of each section. These data files can be downloaded directly from GitHub or with the [OCSdata](https://github.com/opencasestudies/OCSdata) package (see chapter 2.5). The following table describes which data folder contains the corresponding data for each case study section. An example use for each data type is provided as well.

| Data Folder | Case Study Section | Example Use |
| --- | --- | --- |
| raw | Data Import | Assigning public health students to read through a case study starting from the beginning |
| imported | Data Exploration, Data Wrangling | Data science students practicing wrangling methods with in class exercises |
| wrangled | Data Visualization, Data Analysis | Statistics course practicing analysis methods with real data as a live lesson |
| simpler\_import | Data Import | Introductory data science instructor wants students to practice data import without over-complication |
| extra | Not Used in Case Study | Public health course assigns homework to practice case study methods with similar but new data |

(avocado add discussion of interactive case studies)

### 3.2.3 Teaching With Case Study Data Only

* Data available on GitHub
* Can use OCSdata package for streamlined data retrieval process
* Suggestions on how to use data?
* Look at case studies for exercise ideas
* more?

## 3.3 Case Study Recommendations

The following table lists a few example case studies that would be suitable for each experience level.

| Experience Level | Case Studies |
| --- | --- |
| Beginner | Exploring CO2 emissions across time: [Static](https://www.opencasestudies.org/ocs-bp-co2-emissions/), [Interactive](https://rsconnect.biostat.jhsph.edu/ocs-bp-co2-emissions-interactive/). Vaping Behaviors in American Youth: [Static](https://www.opencasestudies.org/ocs-bp-vaping-case-study/). |
| Intermediate | Mental Health of American Youth: [Static](https://www.opencasestudies.org/ocs-bp-youth-mental-health/), [Interactive](https://rsconnect.biostat.jhsph.edu/ocs-bp-youth-mental-health-interactive/). Opioids in United States: [Static](https://www.opencasestudies.org/ocs-bp-opioid-rural-urban/), [Interactive](https://rsconnect.biostat.jhsph.edu/ocs-bp-opioid-rural-urban-interactive). |
| Advanced | Influence of Multicollinearity on Measured Impact of Right-to-Carry Gun Laws Part 1: [Static](https://www.opencasestudies.org/ocs-bp-RTC-wrangling/). Influence of Multicollinearity on Measured Impact of Right-to-Carry Gun Laws Part 2: [Static](https://www.opencasestudies.org/ocs-bp-RTC-analysis/), [Interactive](https://rsconnect.biostat.jhsph.edu/ocs-bp-RTC-analysis-interactive). |

## 3.4 Example Use Cases

Because the case studies were developed to be modular and stand-alone, they can be used in a variety of ways that cater to the learner’s goals, experience, and interests. Below, we provide a few examples of how case studies have been used previously. If you use Open Case Studies in a new way, we would love to hear about it!

**Beginner Undergraduate Data Science Course**

[Practical Data Science in R](https://cogs137.github.io/website/) is a 10-week intermediate undergraduate course taught be Dr. Shannon Ellis at University of California Santa Barbara. It includes three Open Case Studies and uses them to illustrate how foundational data science skills and statistical concepts taught throughout the course can be applied to real data.

The course incorporates labs and homeworks, which have guided coding and analysis exercises related to the concepts discussed in lecture. The course also assigns written reports where students present the analysis they conduct related to the case study in the format of a scientific article (see example assignment below).

**Advanced Graduate Data Science Course**

[Advanced Data Science](http://jtleek.com/ads2020/) is a semester-long course taught by Dr. Jeff Leek and Dr. Roger Peng in 2020 at Johns Hopkins Bloomberg School of Public Health, primarily for PhD students. This course is designed for students to gain experience in designing and communicating data analyses effectively and critically analyzing analyses. Assignments included [writing scientific journal sections](https://github.com/advdatasci/homework9) (e.g. Introduction, Methods, Results, Discussion) based on the case studies and [extending analyses](https://github.com/advdatasci/homework11) based on results presented in the case study.

**Independent Study**

Case studies can be used for learners to gain experience in statistics and data science independently. We strongly recommend that independent learners aim to actively engage with the case study by running the analyses independently, exploring the data beyond what is presented in the case study, and extending the analyses by to investigate their own hypotheses. Furthermore, creating a finished product, such as a blog post or a presentation, can be an excellent demonstration of the skills learned.

**Interactive Case Studies**

Some of the case studies are also included in interactive versions. These versions allow students to write and run code in the browser interactively, with hints and answers available for students to check their progress as they go through the case study. Interactive case studies could be appropriate for independent learning or for in class labs, as they provide real time feedback and can reduce demands on the educator to provide intensive personalized feedback.

## 3.5 Examples of assignments

Educators are not limited to having the students go through the case study in their current format. Case studies can be a spring board for further exploration and additional assignments. For example, in addition to helping students develop data analysis skills, we hope that the case studies can also help students develop their curiosity, technical writing and communication skills. Additional assignments can include but is not limited to data visualization and presentation, written reports, and oral presentations. Below we provide a few examples of potential assignments that educators can use to tailor instruction to the desired learning objectives.

**Further Exploration**

Several case studies have additional data that is not discussed. This can be used for further exploration of the subject area that was discussed in the case study. This can be guided by the questions included in the Homework section of the case studies. Case studies that have additional data include the [Opioid Use Case Study](https://www.opencasestudies.org/ocs-bp-opioid-rural-urban/#Main_Question), the [Right to Carry Case Study](https://www.opencasestudies.org/ocs-bp-RTC-analysis/), and the [CO2 Emissions Case Study](https://www.opencasestudies.org/ocs-bp-co2-emissions/).

**Written Report**

Below is an example of a scientific-style paper written based on the [Opioid Use Case Study](https://www.opencasestudies.org/ocs-bp-opioid-rural-urban/#Main_Question). We also include an example rubric by which this paper can be evaluated adapted from [here](https://ocw.mit.edu/courses/biological-engineering/20-109-laboratory-fundamentals-in-biological-engineering-spring-2010/assignments/guidelines-for-writing-up-your-research/#Evaluation).

[OCS Example Report Assignment Guide](https://cogs137.github.io/website/project/cs01.html)

[OCS Opioid Use in the US Example Report](https://raw.githubusercontent.com/opencasestudies/OCS_Guide/main/assets/OCS_Opioids_Example_Report.pdf)

[OCS Opioid Use in the US Example Report Rubric](https://raw.githubusercontent.com/opencasestudies/OCS_Guide/main/assets/OCS_Opioids_Example_Report_Rubric.pdf)

**Oral Presentation**

Below is an example of a presentation based on the [Vaping Behaviors in American Youth Case Study](https://www.opencasestudies.org/ocs-bp-vaping-case-study/). This presentation focuses on the context of the study and the methods used in the analysis. Assignments for presentations can be modified to emphasize presenting results and conclusions or to emphasize communication to different audiences (e.g. policy makers, other researchers, the public, etc. ).

(avocado insert presentation assignment guide, example, and grading rubric)

**Data Visualization**

The data visualizations included in the case studies are not the only way to present the results of the analyses. While the principles of effective data visualization are not a focus of the case studies, the data included as well as the study questions can be used to guide students through the design choices that are commonly considered when determining how to best present data. As examples, we link a [lecture on the principles of data visualization](https://paldhous.github.io/ucb/2019/dataviz/week2.html) from Dr. Peter Aldhous’ Introduction to Data Visualization Course and a [data visualization assignment and accompanying grading rubric](http://vis.mit.edu/classes/6.894/A1) from the Interactive Data Visualization course at MIT’s Department of Computer Science.

*Acknowledgements*

We would like to thank the following people for generously sharing the ways in which they used Open Case Studies materials in their teaching:

* Shannon Ellis PhD., \*University of California, Santa Barbara

## R version 4.0.2 (2020-06-22)  
## Platform: x86\_64-pc-linux-gnu (64-bit)  
## Running under: Ubuntu 20.04.3 LTS  
##   
## Matrix products: default  
## BLAS/LAPACK: /usr/lib/x86\_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.8.so  
##   
## locale:  
## [1] LC\_CTYPE=en\_US.UTF-8 LC\_NUMERIC=C   
## [3] LC\_TIME=en\_US.UTF-8 LC\_COLLATE=en\_US.UTF-8   
## [5] LC\_MONETARY=en\_US.UTF-8 LC\_MESSAGES=C   
## [7] LC\_PAPER=en\_US.UTF-8 LC\_NAME=C   
## [9] LC\_ADDRESS=C LC\_TELEPHONE=C   
## [11] LC\_MEASUREMENT=en\_US.UTF-8 LC\_IDENTIFICATION=C   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## loaded via a namespace (and not attached):  
## [1] bookdown\_0.24 crayon\_1.4.2 digest\_0.6.25 R6\_2.5.1   
## [5] lifecycle\_1.0.1 magrittr\_2.0.2 evaluate\_0.14 pillar\_1.4.6   
## [9] rlang\_0.4.10 stringi\_1.5.3 fs\_1.5.2 ellipsis\_0.3.1   
## [13] vctrs\_0.3.4 rmarkdown\_2.10 ottr\_0.1.2 tools\_4.0.2   
## [17] stringr\_1.4.0 readr\_1.4.0 hms\_0.5.3 xfun\_0.26   
## [21] yaml\_2.2.2 compiler\_4.0.2 pkgconfig\_2.0.3 htmltools\_0.5.0  
## [25] knitr\_1.33 tibble\_3.0.3

# 4 Modifying and creating open case studies

## 4.1 Learning Objectives

This chapter will cover how to modify the case studies to your own needs using the following methods:

* Modular case study use with the help of the [OCSdata package](https://github.com/opencasestudies/OCSdata)
* Modifying a case study with [GitHub](https://github.com/opencasestudies) and [RStudio](https://www.rstudio.com/)
* Creating your own case study with our [template](https://github.com/opencasestudies/ocs-bp-template) and [MakeCaseStudies](https://rsconnect.biostat.jhsph.edu/MakeCaseStudies/)

## 4.2 Libraries

For this chapter, we’ll need the following packages attached:

library(OCSdata)

## 4.3 Modular use

*short explanation of how case studies can be used modularly for those who want to break the case study up or are only interested in part of the case study* table about which package functions to use for each section in a case study you could start from \*step by step example of how to jump into the middle of a case study (do data analysis section as an example?)

## 4.4 Modify a case study

*short explanation of case study structure: RStudio projects written in Rmd, knitted to HTML and hosted on GitHub* tutorial on .Rmd and knitting in RStudio *tutorial on GitHub: forking* step by step example of forking a case study repository, editing Rmd content, knitting, and sharing the output

## 4.5 Create a case study

*tutorial on how to create a case study using GitHub template* tutorial on how to create a case study using [MakeCaseStudies](https://rsconnect.biostat.jhsph.edu/MakeCaseStudies/) app

## 4.6 Session info

## R version 4.0.2 (2020-06-22)  
## Platform: x86\_64-pc-linux-gnu (64-bit)  
## Running under: Ubuntu 20.04.3 LTS  
##   
## Matrix products: default  
## BLAS/LAPACK: /usr/lib/x86\_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.8.so  
##   
## locale:  
## [1] LC\_CTYPE=en\_US.UTF-8 LC\_NUMERIC=C   
## [3] LC\_TIME=en\_US.UTF-8 LC\_COLLATE=en\_US.UTF-8   
## [5] LC\_MONETARY=en\_US.UTF-8 LC\_MESSAGES=C   
## [7] LC\_PAPER=en\_US.UTF-8 LC\_NAME=C   
## [9] LC\_ADDRESS=C LC\_TELEPHONE=C   
## [11] LC\_MEASUREMENT=en\_US.UTF-8 LC\_IDENTIFICATION=C   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] OCSdata\_1.0.2  
##   
## loaded via a namespace (and not attached):  
## [1] knitr\_1.33 ottr\_0.1.2 magrittr\_2.0.2 usethis\_2.1.5.9000  
## [5] hms\_0.5.3 R6\_2.5.1 rlang\_0.4.10 stringr\_1.4.0   
## [9] httr\_1.4.2 tools\_4.0.2 xfun\_0.26 htmltools\_0.5.0   
## [13] ellipsis\_0.3.1 yaml\_2.2.2 digest\_0.6.25 tibble\_3.0.3   
## [17] lifecycle\_1.0.1 crayon\_1.4.2 bookdown\_0.24 purrr\_0.3.4   
## [21] readr\_1.4.0 vctrs\_0.3.4 fs\_1.5.2 glue\_1.6.1   
## [25] evaluate\_0.14 rmarkdown\_2.10 stringi\_1.5.3 compiler\_4.0.2   
## [29] pillar\_1.4.6 pkgconfig\_2.0.3

# 5 New Case Studies - Building and Contributing

In this chapter we will discuss the guidelines for creating new case studies as well as how to publish your own new case study as part of our project. To preserve the integrity of the core ideas of our resource we provide the following guidelines for case studies to be included in the Open Case Studies collection:

### 5.0.1 Guidelines:

* **Programming language:** Case studies should be written in open source programming languages (preferably R).
* **Data:**
  + Case studies should use data that is publicly available or can be made publicly available. Please ensure that you are allowed to make the data public if it is not already.
  + Transparent descriptions of data sources and how data was generated should be included when possible.
* **Core sections**: all case studies should include the following sections consistent with the published case studies. See Chapter 1, Case Study Anatomy for a detailed description on what each section should include:
  + Case study context
  + Study motivation
  + Main question
  + Learning objectives
  + Study limitations
  + Data description
  + Import and exploration
  + Wrangling
  + Visualization
  + Analysis
  + Analysis conclusions
  + Case study summary
  + Next steps
  + Homework
  + Additional information
* **Content:** Despite often being motivated by articles, case studies are not intended to demonstrate the methods of a paper - they are intended as an educational resource where users are guided through the data science process.
  + Links to literature or other sources to motivate the scientific topic of the case study should be included where possible.
  + Case studies should aim to describe the decision making process involved in performing data science related tasks.
* **Attribution:** all outside resources used in the case study should be referenced appropriately.
  + Case studies should include disclaimers and appropriate license agreements.
  + All included images (that are not original to the case study) should include relevant sources.

### 5.0.2 Submission:

sessionInfo()

## R version 4.0.2 (2020-06-22)  
## Platform: x86\_64-pc-linux-gnu (64-bit)  
## Running under: Ubuntu 20.04.3 LTS  
##   
## Matrix products: default  
## BLAS/LAPACK: /usr/lib/x86\_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.8.so  
##   
## locale:  
## [1] LC\_CTYPE=en\_US.UTF-8 LC\_NUMERIC=C   
## [3] LC\_TIME=en\_US.UTF-8 LC\_COLLATE=en\_US.UTF-8   
## [5] LC\_MONETARY=en\_US.UTF-8 LC\_MESSAGES=C   
## [7] LC\_PAPER=en\_US.UTF-8 LC\_NAME=C   
## [9] LC\_ADDRESS=C LC\_TELEPHONE=C   
## [11] LC\_MEASUREMENT=en\_US.UTF-8 LC\_IDENTIFICATION=C   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## loaded via a namespace (and not attached):  
## [1] bookdown\_0.24 crayon\_1.4.2 digest\_0.6.25 R6\_2.5.1   
## [5] lifecycle\_1.0.1 magrittr\_2.0.2 evaluate\_0.14 pillar\_1.4.6   
## [9] rlang\_0.4.10 stringi\_1.5.3 fs\_1.5.2 ellipsis\_0.3.1   
## [13] vctrs\_0.3.4 rmarkdown\_2.10 ottr\_0.1.2 tools\_4.0.2   
## [17] stringr\_1.4.0 readr\_1.4.0 hms\_0.5.3 xfun\_0.26   
## [21] yaml\_2.2.2 compiler\_4.0.2 pkgconfig\_2.0.3 htmltools\_0.5.0  
## [25] knitr\_1.33 tibble\_3.0.3

# About the Authors

These credits are based on our [course contributors table guidelines](https://github.com/jhudsl/OTTR_Template/wiki/How-to-give-credits).

| Credits | Names |
| --- | --- |
| **Pedagogy** |  |
| Lead Content Instructor(s) | [FirstName LastName](link%20to%20personal%20website) |
| Lecturer(s) (include chapter name/link in parentheses if only for specific chapters) - make new line if more than one chapter involved | Delivered the course in some way - video or audio |
| Content Author(s) (include chapter name/link in parentheses if only for specific chapters) - make new line if more than one chapter involved | If any other authors besides lead instructor |
| Content Author(s) | Michael Breshock and Lyla Atta |
| Content Contributor(s) (include section name/link in parentheses) - make new line if more than one section involved | Wrote less than a chapter |
| Content Editor(s)/Reviewer(s) | Checked your content |
| Content Director(s) | Helped guide the content direction |
| Content Consultants (include chapter name/link in parentheses or word “General”) - make new line if more than one chapter involved | Gave high level advice on content |
| Acknowledgments | Gave small assistance to content but not to the level of consulting |
| **Production** |  |
| Content Publisher(s) | Helped with publishing platform |
| Content Publishing Reviewer(s) | Reviewed overall content and aesthetics on publishing platform |
| **Technical** |  |
| Course Publishing Engineer(s) | Helped with the code for the technical aspects related to the specific course generation |
| Template Publishing Engineers | [Candace Savonen](https://www.cansavvy.com/), [Carrie Wright](https://carriewright11.github.io/) |
| Publishing Maintenance Engineer | [Candace Savonen](https://www.cansavvy.com/) |
| Technical Publishing Stylists | [Carrie Wright](https://carriewright11.github.io/), [Candace Savonen](https://www.cansavvy.com/) |
| Package Developers ([ottr](https://github.com/jhudsl/ottr)) | [John Muschelli](https://johnmuschelli.com/), [Candace Savonen](https://www.cansavvy.com/), [Carrie Wright](https://carriewright11.github.io/) |
| **Art and Design** |  |
| Illustrator(s) | Created graphics for the course |
| Figure Artist(s) | Created figures/plots for course |
| Videographer(s) | Filmed videos |
| Videography Editor(s) | Edited film |
| Audiographer(s) | Recorded audio |
| Audiography Editor(s) | Edited audio recordings |
| **Funding** |  |
| Funder(s) | Institution/individual who funded course including grant number |
| Funding Staff | Staff members who help with funding |

## ─ Session info ───────────────────────────────────────────────────────────────  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Ubuntu 20.04.3 LTS   
## system x86\_64, linux-gnu   
## ui X11   
## language (EN)   
## collate en\_US.UTF-8   
## ctype en\_US.UTF-8   
## tz Etc/UTC   
## date 2022-02-09   
##   
## ─ Packages ───────────────────────────────────────────────────────────────────  
## package \* version date lib source   
## bookdown 0.24 2022-02-08 [1] Github (rstudio/bookdown@88bc4ea)   
## callr 3.4.4 2020-09-07 [1] RSPM (R 4.0.2)   
## cli 3.1.1 2022-01-20 [1] CRAN (R 4.0.2)   
## crayon 1.4.2 2021-10-29 [1] CRAN (R 4.0.2)   
## desc 1.4.0 2021-09-28 [1] CRAN (R 4.0.2)   
## devtools 2.3.2 2020-09-18 [1] RSPM (R 4.0.3)   
## digest 0.6.25 2020-02-23 [1] RSPM (R 4.0.0)   
## ellipsis 0.3.1 2020-05-15 [1] RSPM (R 4.0.3)   
## evaluate 0.14 2019-05-28 [1] RSPM (R 4.0.3)   
## fs 1.5.2 2021-12-08 [1] CRAN (R 4.0.2)   
## glue 1.6.1 2022-01-22 [1] CRAN (R 4.0.2)   
## htmltools 0.5.0 2020-06-16 [1] RSPM (R 4.0.1)   
## knitr 1.33 2022-02-08 [1] Github (yihui/knitr@a1052d1)   
## lifecycle 1.0.1 2021-09-24 [1] CRAN (R 4.0.2)   
## magrittr 2.0.2 2022-01-26 [1] CRAN (R 4.0.2)   
## memoise 1.1.0 2017-04-21 [1] RSPM (R 4.0.0)   
## pkgbuild 1.1.0 2020-07-13 [1] RSPM (R 4.0.2)   
## pkgload 1.1.0 2020-05-29 [1] RSPM (R 4.0.3)   
## prettyunits 1.1.1 2020-01-24 [1] RSPM (R 4.0.3)   
## processx 3.4.4 2020-09-03 [1] RSPM (R 4.0.2)   
## ps 1.3.4 2020-08-11 [1] RSPM (R 4.0.2)   
## purrr 0.3.4 2020-04-17 [1] RSPM (R 4.0.3)   
## R6 2.5.1 2021-08-19 [1] CRAN (R 4.0.2)   
## remotes 2.2.0 2020-07-21 [1] RSPM (R 4.0.3)   
## rlang 0.4.10 2022-02-08 [1] Github (r-lib/rlang@f0c9be5)   
## rmarkdown 2.10 2022-02-08 [1] Github (rstudio/rmarkdown@02d3c25)  
## rprojroot 2.0.2 2020-11-15 [1] CRAN (R 4.0.2)   
## sessioninfo 1.1.1 2018-11-05 [1] RSPM (R 4.0.3)   
## stringi 1.5.3 2020-09-09 [1] RSPM (R 4.0.3)   
## stringr 1.4.0 2019-02-10 [1] RSPM (R 4.0.3)   
## testthat 3.0.1 2022-02-08 [1] Github (R-lib/testthat@e99155a)   
## usethis 2.1.5.9000 2022-02-08 [1] Github (r-lib/usethis@57b109a)   
## withr 2.4.3 2021-11-30 [1] CRAN (R 4.0.2)   
## xfun 0.26 2022-02-08 [1] Github (yihui/xfun@74c2a66)   
## yaml 2.2.2 2022-01-25 [1] CRAN (R 4.0.2)   
##   
## [1] /usr/local/lib/R/site-library  
## [2] /usr/local/lib/R/library

# References