[Your School’s] Lesson Plan

**Teacher:** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Class**: ­­­­­­­­­­­­­­­­­­­­ Introduction to Data Science

**Week**: ­­­­­­­­­­­­­­­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Duration:** 2 weeks -

**Lesson:**  Cleaning Data in R

|  |
| --- |
| Materials and Resources: |
| Textbook: R for Data Science by Hadley Wickham**Free downloads for students:** [**https://r4ds.had.co.nz/**](https://r4ds.had.co.nz/)**Video for the Students:****Mutate:**[**https://www.youtube.com/watch?v=2dFpblO7MB8&t=11s**](https://www.youtube.com/watch?v=2dFpblO7MB8&t=11s)**Pivot Longer:**[**https://www.youtube.com/watch?v=UR-4vBEN3Fw&t=3s**](https://www.youtube.com/watch?v=UR-4vBEN3Fw&t=3s)**Pivot Wider:**[**https://www.youtube.com/watch?v=YpAdZ4079qs&t=3s**](https://www.youtube.com/watch?v=YpAdZ4079qs&t=3s)**Separate:**[**https://www.youtube.com/watch?v=yDZljicdLAk&t=3s**](https://www.youtube.com/watch?v=yDZljicdLAk&t=3s) |
| Learning Objectives: |
| **Topics:**Data cleansing and pre-processing                 A.           Data preparation methods and tools               B.           Data extraction and mining               C.           Data wrangling                              1.            Blending data from multiple sources and formats                              2.            Transforming the data for analysis                              3.            Querying and summarizing data                              4.            Data munging        CLO- 2 -Collect, clean and prepare the data for analysis and identify problems that might arise from assumptions made during this process. |
| Instructional Strategies: |
| Students should learn to use data science technology to clean data, understand different data types and the difference between long and wide data formats. Students should learn to create customized and new data fields from existing data, pivot data, sperate out fields and join multiple data frames using a primary key. |
| Tools: |
| This lesson plan was written with the intention of using R. |
| Content: |
| Assessment: |
| Students will be assessed through projects, and quizzes. |
| Homework/Assignments: |
| R for Data Science by Hadley Wickham Chapter 12 Tidy Data (pivots, separate, missing values), Chapter 5 Data Transformations and Chapter 13 Relational Data (joins).* [12.1 Introduction](https://r4ds.had.co.nz/tidy-data.html#introduction-6)
	+ [12.1.1 Prerequisites](https://r4ds.had.co.nz/tidy-data.html#prerequisites-6)
* [12.2 Tidy data](https://r4ds.had.co.nz/tidy-data.html#tidy-data-1)
	+ [12.2.1 Exercises](https://r4ds.had.co.nz/tidy-data.html#exercises-23)
* [12.3 Pivoting](https://r4ds.had.co.nz/tidy-data.html#pivoting)
	+ [12.3.1 Longer](https://r4ds.had.co.nz/tidy-data.html#longer)
	+ [12.3.2 Wider](https://r4ds.had.co.nz/tidy-data.html#wider)
	+ [12.3.3 Exercises](https://r4ds.had.co.nz/tidy-data.html#exercises-24)
* [12.4 Separating and uniting](https://r4ds.had.co.nz/tidy-data.html#separating-and-uniting)
	+ [12.4.1 Separate](https://r4ds.had.co.nz/tidy-data.html#separate)
	+ [12.4.2 Unite](https://r4ds.had.co.nz/tidy-data.html#unite)
	+ [12.4.3 Exercises](https://r4ds.had.co.nz/tidy-data.html#exercises-25)
* [12.5 Missing values](https://r4ds.had.co.nz/tidy-data.html#missing-values-3)
	+ [12.5.1 Exercises](https://r4ds.had.co.nz/tidy-data.html#exercises-26)
* [12.6 Case Study](https://r4ds.had.co.nz/tidy-data.html#case-study)
	+ [12.6.1 Exercises](https://r4ds.had.co.nz/tidy-data.html#exercises-27)
* [12.7 Non-tidy data](https://r4ds.had.co.nz/tidy-data.html#non-tidy-data)
* [5 Data transformation](https://r4ds.had.co.nz/transform.html#transform)
* [5.1 Introduction](https://r4ds.had.co.nz/transform.html#introduction-2)
	+ [5.1.1 Prerequisites](https://r4ds.had.co.nz/transform.html#prerequisites-2)
	+ [5.1.2 nycflights13](https://r4ds.had.co.nz/transform.html#nycflights13)
	+ [5.1.3 dplyr basics](https://r4ds.had.co.nz/transform.html#dplyr-basics)
* [5.2 Filter rows with filter()](https://r4ds.had.co.nz/transform.html#filter-rows-with-filter)
	+ [5.2.1 Comparisons](https://r4ds.had.co.nz/transform.html#comparisons)
	+ [5.2.2 Logical operators](https://r4ds.had.co.nz/transform.html#logical-operators)
	+ [5.2.3 Missing values](https://r4ds.had.co.nz/transform.html#missing-values)
	+ [5.2.4 Exercises](https://r4ds.had.co.nz/transform.html#exercises-8)
* [5.3 Arrange rows with arrange()](https://r4ds.had.co.nz/transform.html#arrange-rows-with-arrange)
	+ [5.3.1 Exercises](https://r4ds.had.co.nz/transform.html#exercises-9)
* [5.4 Select columns with select()](https://r4ds.had.co.nz/transform.html#select)
	+ [5.4.1 Exercises](https://r4ds.had.co.nz/transform.html#exercises-10)
* [5.5 Add new variables with mutate()](https://r4ds.had.co.nz/transform.html#add-new-variables-with-mutate)
	+ [5.5.1 Useful creation functions](https://r4ds.had.co.nz/transform.html#mutate-funs)
	+ [5.5.2 Exercises](https://r4ds.had.co.nz/transform.html#exercises-11)
* [5.6 Grouped summaries with summarise()](https://r4ds.had.co.nz/transform.html#grouped-summaries-with-summarise)
	+ [5.6.1 Combining multiple operations with the pipe](https://r4ds.had.co.nz/transform.html#combining-multiple-operations-with-the-pipe)
	+ [5.6.2 Missing values](https://r4ds.had.co.nz/transform.html#missing-values-1)
	+ [5.6.3 Counts](https://r4ds.had.co.nz/transform.html#counts)
	+ [5.6.4 Useful summary functions](https://r4ds.had.co.nz/transform.html#summarise-funs)
	+ [5.6.5 Grouping by multiple variables](https://r4ds.had.co.nz/transform.html#grouping-by-multiple-variables)
	+ [5.6.6 Ungrouping](https://r4ds.had.co.nz/transform.html#ungrouping)
	+ [5.6.7 Exercises](https://r4ds.had.co.nz/transform.html#exercises-12)
* [5.7 Grouped mutates (and filters)](https://r4ds.had.co.nz/transform.html#grouped-mutates-and-filters)
	+ [5.7.1 Exercises](https://r4ds.had.co.nz/transform.html#exercises-13)
* [13 Relational data](https://r4ds.had.co.nz/relational-data.html#relational-data)
* [13.1 Introduction](https://r4ds.had.co.nz/relational-data.html#introduction-7)
	+ [13.1.1 Prerequisites](https://r4ds.had.co.nz/relational-data.html#prerequisites-7)
* [13.2 nycflights13](https://r4ds.had.co.nz/relational-data.html#nycflights13-relational)
	+ [13.2.1 Exercises](https://r4ds.had.co.nz/relational-data.html#exercises-28)
* [13.3 Keys](https://r4ds.had.co.nz/relational-data.html#keys)
	+ [13.3.1 Exercises](https://r4ds.had.co.nz/relational-data.html#exercises-29)
* [13.4 Mutating joins](https://r4ds.had.co.nz/relational-data.html#mutating-joins)
	+ [13.4.1 Understanding joins](https://r4ds.had.co.nz/relational-data.html#understanding-joins)
	+ [13.4.2 Inner join](https://r4ds.had.co.nz/relational-data.html#inner-join)
	+ [13.4.3 Outer joins](https://r4ds.had.co.nz/relational-data.html#outer-join)
	+ [13.4.4 Duplicate keys](https://r4ds.had.co.nz/relational-data.html#join-matches)
	+ [13.4.5 Defining the key columns](https://r4ds.had.co.nz/relational-data.html#join-by)
	+ [13.4.6 Exercises](https://r4ds.had.co.nz/relational-data.html#exercises-30)
	+ [13.4.7 Other implementations](https://r4ds.had.co.nz/relational-data.html#other-implementations)
* [13.5 Filtering joins](https://r4ds.had.co.nz/relational-data.html#filtering-joins)
	+ [13.5.1 Exercises](https://r4ds.had.co.nz/relational-data.html#exercises-31)
* [13.6 Join problems](https://r4ds.had.co.nz/relational-data.html#join-problems)
* [13.7 Set operations](https://r4ds.had.co.nz/relational-data.html#set-operations)
 |
|  |
|   |
| Notes: |
| For Chapter 5 and 13 students will access the nycflights13 package and use the build in data from R, airlines, airports, planes and weather.For Chapter 12 students will access the building dataframes in R; table1, table2, table 3, table4a and table 4b.Please become familiar with these built in dataframes in R.  |
| Standards/Alignment: |
|   |

Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_