

Lab 4: 2019 Novel Coronavirus

Name: Sal Figueroa

2025-04-17

Contents

Github Repository	1
Required data sets	1
Project Objectives	1
Objective 1	1
Objective 2	1
Objective 3	2
Objective 4	2
Objective 5	2

Github Repository

Holds all related files

github: Figgs0bit-Lab4 (<https://github.com/Figgs0bit/CSIT165-Lab4.git>)

Required data sets

2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by John Hopkins CSSE

The following data set is operated by the John Hopkins University Center for Systems Science and Engineering (JHU CSSE). Data sets include daily time series CSV summary tables for confirmations and deaths associated with COVID-19. Lat and Long refer to coordinate references for the data field. Date fields are stored in MM/DD/YYYY format. For this laboratory, we will be use confirmed and deaths data sets for the US.

Project Objectives

Before beginning your objectives in your final document, please state which day you downloaded the data sets on for analysis. For this laboratory, use the packages described in Module 10 and 11 to accomplish each objective where appropriate. The surgeon general for the United States recently created a new data science initiative, CSIT-165, that uses data science to characterize pandemic diseases. CSIT-165 disseminates data driven analyses to state governors. You are a data scientist for CSIT-165 and it is up to you and you alone to manipulate and visualize COVID-19 data for disease control.

Objective 1

Which state has the most confirmed cases of COVID-19? Which state has the most deaths?

Objective 2

Create a new table called `current_tally` by merging confirmed and deaths containing only the columns `iso2`, `Province_State`, `Admin2`, `Lat`, `Long_`, and today's count for confirmed and deaths. Rename `current_tally`'s

columns as *Country*, *State*, *City*, *Lat*, *Long*, *Confirmed*, and *Deaths* respectively. Filter *current_tally* for data points where the country is the US. Please print *current_tally* as a tibble.

Objective 3

In *current_tally*, create a new variable, *unaccounted*, as the absolute difference between confirmed and deaths for each city. Determine the number of unaccounted for each state using grouped operation.

Objective 4

Using the package *kable*, create a table showing the top 5 unaccounted states.

Objective 5

Create any visualization you like using the data sets used in this laboratory. Provide your visualization with a title and custom axis labels. Describe how this visualization conveys information that is not readily apparent from looking at the data alone.