

Project 1

Jason Ballinger, Jason Perez, Megan Juza

2024-04-02

Contents

Background	1
Data	1
Project Objectives	1
Objective 1	1
Objective 2	2
Objective 3	2
Objective 4	2
GitHub Log	2

Background

The World Health Organization has recently employed a new data science initiative, *CSIT-165*, that uses data science to characterize pandemic diseases. *CSIT-165* disseminates data driven analyses to global decision makers.

CSIT-165 is a conglomerate comprised of two fabricated entities: *Global Health Union (GHU)* and *Private Diagnostic Laboratories (PDL)*. Your and your partner's role is to play a data scientist from one of these two entities.

Data

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

Data for 2019 Novel Coronavirus is operated by the John Hopkins University Center for Systems Science and Engineering (JHU CSSE). Data includes daily time series CSV summary tables, including confirmations, recoveries, and deaths. Country/region are countries/regions that conform to World Health Organization (WHO). Lat and Long refer to coordinates references for the user. Date fields are stored in MM/DD/YYYY format.

Project Objectives

Objective 1

```
# Objective 1: Determine where COVID-19 originated from

library(readr)

# Import data sets
cases <- data.frame(read_delim(file="time_series_covid19_confirmed_global.csv", delim=","))

## Rows: 289 Columns: 1147
## -- Column specification -----
```

```
## Delimiter: ","
## chr      (2): Province/State, Country/Region
## dbl (1145): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
deaths <- data.frame(read_delim(file="time_series_covid19_deaths_global.csv", delim=","))

## Rows: 289 Columns: 1147
## -- Column specification -----
## Delimiter: ","
## chr      (2): Province/State, Country/Region
## dbl (1145): Lat, Long, 1/22/20, 1/23/20, 1/24/20, 1/25/20, 1/26/20, 1/27/20,...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
# Highest confirmed cases & deaths for first day (1/22/2020)
highestCasesIndex <- which.max(cases$X1.22.20)
highestDeathsIndex <- which.max(deaths$X1.22.20)

# Get location given the index
highestCasesLocation <- paste(cases$Province.State[highestCasesIndex], cases$Country.Region[highestCasesIndex])
highestDeathsLocation <- paste(deaths$Province.State[highestDeathsIndex], cases$Country.Region[highestDeathsIndex])

# Check if the locations are the same
if (highestCasesLocation == highestDeathsLocation) {
  x <- highestCasesLocation
  x
} else {
  x <- "Unable to find origin location"
  x
}

## [1] "Hubei China"
```

Objective 2

Objective 3

Objective 4

Objective 4.1

Objective 4.2

GitHub Log

```
git log --pretty=format:"%nSubject: %s%nAuthor: %aN%nDate: %aD%nBody: %b"

##
## Subject: Initial commit
## Author: Jason Ballinger
## Date: Fri, 29 Mar 2024 10:39:40 -0700
## Body:
```