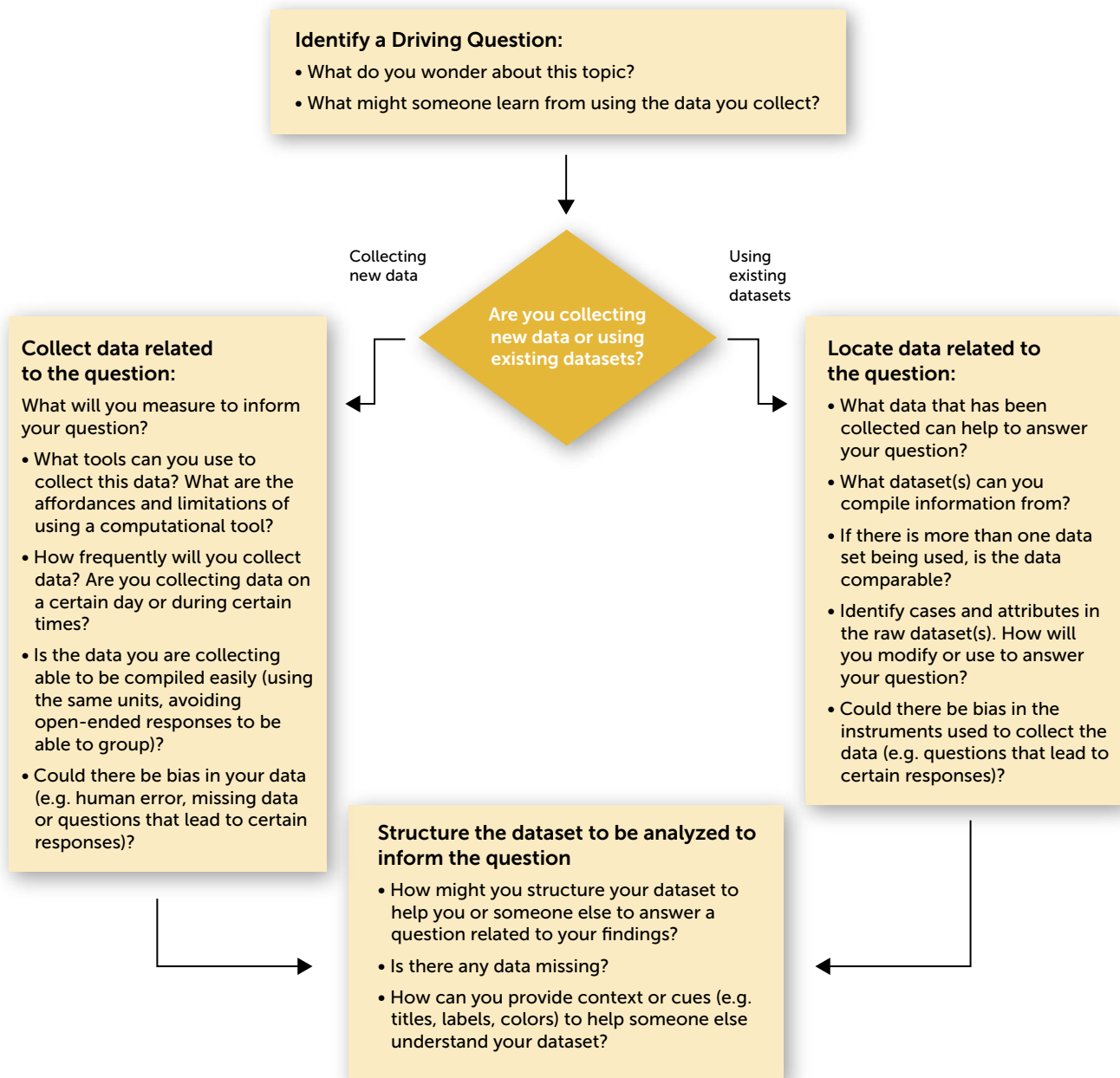


# Data Practices: Collecting, Analyzing and Evaluating Data

## 1

### Part 1: Collecting and Structuring Data to answer a driving question

Use the flowchart below to **1) identify a driving question**, **2) collect or locate data related to the question**, and **3) structure the dataset to be analyzed to inform the question**.



In this activity, a student explored existing data about the novel coronavirus (COVID-19). The student followed the algorithm above to:

**1. Identify a driving question:**

The student began by brainstorming questions they had about the coronavirus, such as:

- Where in the world has the highest number of cases?
- How many cases have been diagnosed in the US?

**2. Locate data related to the question:**

The student identified a dataset related to their COVID-19 questions from [Our World in Data](#). The raw data was downloaded from the Our World in Data website. Then, it was uploaded into a Google Sheet (see [Raw Data Sheet 1](#)).

**3. Structure the dataset to be analyzed to inform the question**

The data needed to be structured in a way that allowed it to be analyzed to answer the driving questions. First, the raw data was formatted into columns/rows. (See [Raw Data Formatted Sheet 2](#).) Then, the rows and columns were flipped to make it easier to filter data for individual countries. (See [Select Countries Formatted Sheet 4](#).)

**2**

**Part 2: Analyzing Data using data moves**

Then, the student used data moves (Erickson et al., 2019) to answer the questions they had about the dataset. Examples of data moves are:

- **Filtering** is selecting only a subset of the data that is available to explore. For example, you may view only a portion of the data because only some of it is relevant to the question you are exploring.
- **Grouping** is used to make comparisons between different subgroups of a data set.
- **Summarizing** is computing a value (e.g. mean, median, mode) to summarize a group.
- **Calculating** is creating a new attribute based on values of one or more existing attributes. For example, if an attribute is distance in kilometers, you may want to convert that value to distance in miles.
- **Merging/joining** is combining more than one dataset together. You may add more cases, or add more attributes to existing cases.
- **Making a hierarchy** is grouping data using multiple levels.

In order to complete data moves, you should become familiar with a program that allows you to manipulate data. Google Sheets, Microsoft Excel, CODAP (Concord Consortium) are all great resources for using data moves to analyze your data.

## Question: How many cases have been confirmed in the US between Jan and Feb?

### Data Move(s)

Dive deeper into the data set by combining moves

Sketch or describe how you organized or illustrated the data:

What does this tell you about the relationships/patterns within your dataset?

What new questions do you have?

Filtering and summarizing

- Filter by country to view only confirmed cases from the US
- Find total number of cases at the end of February
- Find total number of cases at the end of January

There have been approximately 60 cases diagnosed in the past month.

How long will the cases increase?

Country	2020-01-31	2020-02-29
United States	6	66
New Cases in Feb: (Total # of confirmed cases in Jan) - (Total # of confirmed cases in Feb) $66 - 6 = 60$		

## Question: What country has the highest number of confirmed cases?

### Data Move(s)

Dive deeper into the data set by combining moves

Sketch or describe how you organized or illustrated the data:

What does this tell you about the relationships/patterns within your dataset?

What new questions do you have?

Sorting

Sorted from most recent date by highest to lowest

Country	2020-05-03
World	3,388,665
United States	1,133,069
Spain	216,582
Italy	209,328
United Kingdom	182,260
Germany	162,496

The United States has the highest number of cases, followed by Spain, Italy, the United Kingdom, and Germany.

What continent has the highest number of cases? (See analysis below)

## Question: What continent has the highest number of confirmed cases?

### Data Move(s)

Dive deeper into the data set by combining data moves.

### Sketch or describe how you organized or illustrated the data:

### What does this tell you about the relationships/patterns within your dataset?

### What new questions do you have?

### Sorting

- Identified the 25 countries with the highest total number of confirmed cases
- Created a new column to show which continent each country is on
- Filtered each column by continent and add total confirmed cases

Country	Continent	2020-05-03
World		3,388,665
United States	North America	1,133,069
Spain	Europe/Africa	216,582
Italy	Europe	209,328
United Kingdom	Europe	182,260
Germany	Europe	162,496
France	Europe	130,979
Turkey	Europe/Asia	124,375
Russia	Europe/Asia	124,054
Brazil	South America	96,559
Iran	Asia	96,448
China	Asia	83,961
Canada	North America	56,714
Belgium	Europe	49,517
Peru	South America	42,534
Netherlands	Europe	40,236
India	Asia	39,980
Switzerland	Europe	29,734
Ecuador	South America	27,464
Saudi Arabia	Asia	25,459
Portugal	Europe/Africa	25,190
Mexico	North America	22,088
Sweden	Europe	22,082
Ireland	Europe	21,176
Pakistan	Asia	19,103
Chile	South America	18,435
Asia		264,951
North America		1,211,871
Europe		847,808
South America		184,992
Europe/Africa		241,772
Europe/Asia		248,429

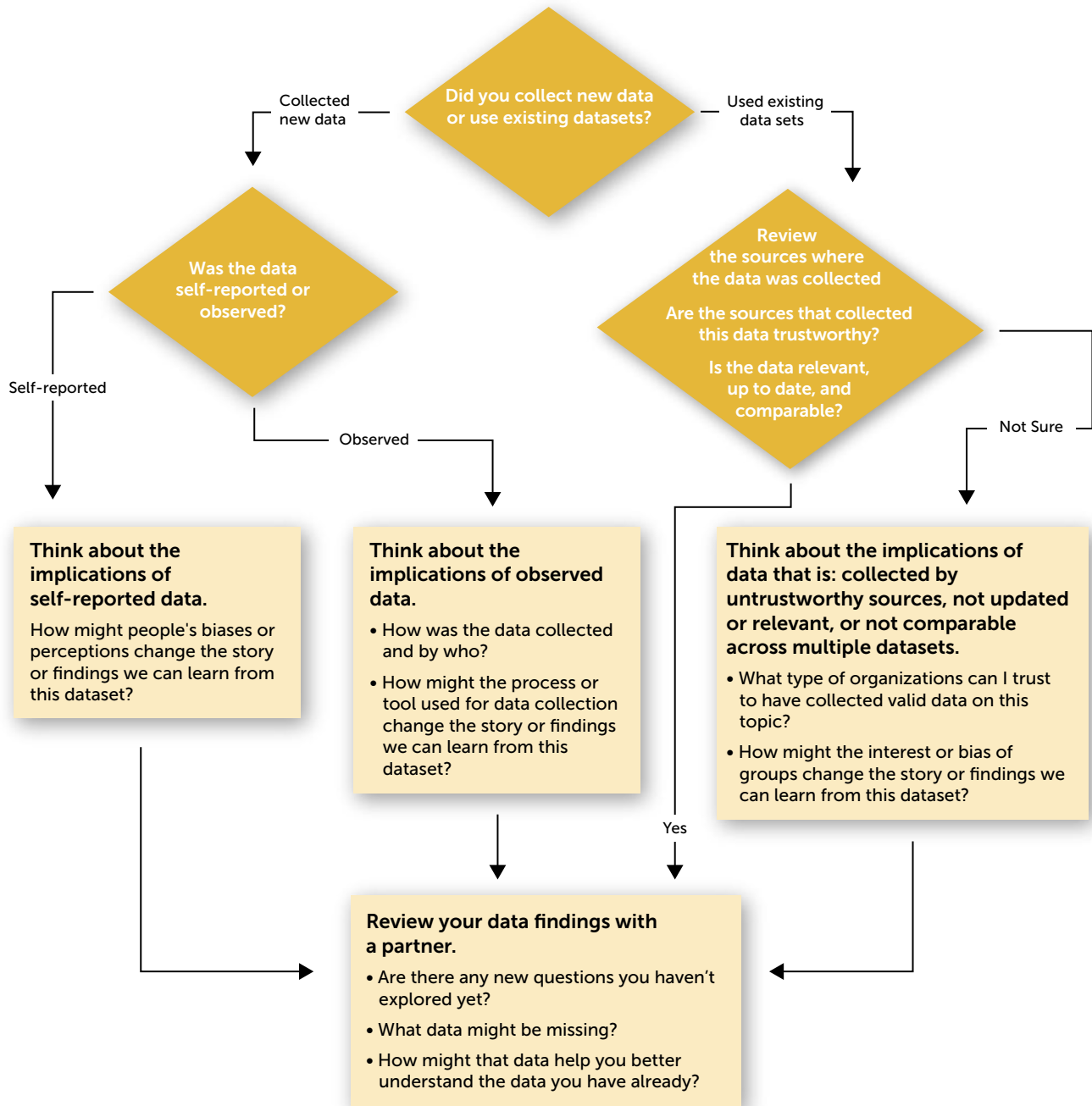
North America has the most cases diagnosed followed by Europe. There are fewer cases in South American and Asia.

What measures would be effective to isolate the disease to a specific location or continent?

## 3

## Part 3: Evaluating Data

Explore the sources that reported the data. Use the flowchart below to consider error and/or bias that could be present.



This student found that [Our World in Data](#) is sourced from the European CDC. They stopped relying on data from the WHO in mid-March 2020 after finding [multiple reporting errors](#). However, the total number of COVID-19 cases is likely higher than the number of confirmed cases reported on the infographic. This is because some individuals/countries have more access to testing than others.