Exercises

Complete the following exercises using what we have learned so far in this book and the data in the exercises/ directory:

1. With faang.csv file from previous exercise: Right now, the data is somewhere between long and wide format. Use melt() to make it completely long format.

   Hint: date and ticker are our ID variables (they uniquely identify each row). We need to melt the rest so that we don't have separate columns for open, high, low, close, and volume.

2. The European Centre for Disease Prevention and Control (ECDC) provides an open dataset on COVID-19 cases called daily number of new reported cases of COVID-19 by country worldwide (https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographicdistribution-covid-19-cases-worldwide). This dataset is updated daily, but we will use a snapshot that contains data from January 1, 2020 through September 18, 2020. Clean and pivot the data so that it is in wide format:

   a) Read in the covid19_cases.csv file.

   b) Create a date column using the data in the dateRep column and the pd.to_datetime() function.

   c) Set the date column as the index and sort the index.

   d) Replace all occurrences of United_States_of_America and United_Kingdom with USA and UK, respectively.

   Hint: the replace() method can be run on the dataframe as a whole.

   e) Using the countriesAndTerritories column, filter the cleaned COVID-19 cases data down to Argentina, Brazil, China, Colombia, India, Italy, Mexico, Peru, Russia, Spain, Turkey, the UK, and the USA.

   f) Pivot the data so that the index contains the dates, the columns contain the country names, and the values are the case counts (the cases column). Be sure to fill in NaN values with 0.