

CIC Student Working Group COVID-19 Data Visualization Project



Build your data analysis skills!

Learn Time Series Analysis,
practice Geospatial Analysis,
& create a Data Visualization
for your portfolio!



COVID INFORMATION COMMONS (CIC) STUDENT WORKING GROUP

COVID-19 Data Visualization Project

The [COVID Information Commons \(CIC\) Student Working Group](#) is excited to support this COVID-19 data analysis and visualization project!

Project Overview:

With this exercise, you will embark on a data-driven journey to analyze and visualize COVID-19 data. This project will not only enhance your data analysis and visualization skills but also offer valuable insights into the ongoing pandemic. This project was designed with beginner to intermediate data science learners in mind. We recommend students to spend 5-8 hours on each Milestone (each student can spend more or less time depending on their interest and level of experience).

Data visualization is an essential skill for all STEM researchers and students - this is an opportunity for you to learn this skill in a friendly and supportive environment!

Key Highlights:

- You'll gain hands-on experience in data cleaning, exploration, and preparation.
- We'll delve into time series analysis, geospatial mapping, and advanced data visualization techniques.
- You will develop: a COVID-19 research question, a comprehensive set of visualizations, and a written description of the insights you found in the given dataset.

When your project is complete, email your final work to Lauren Close at CICStudentWorkingGroup@columbia.edu. Participants who finish this project will receive a certificate of completion. Please reach out to our team at any time for support and questions!

Tools You May Need:

- Microsoft Excel/Google Sheets
- A free [Kaggle Account](#)
- A [Tableau Public](#) Account (optional)
- A [Google Colab Notebook](#) or [Jupyter Notebook](#) (optional)

Important Links:

- [Slack Channel for Discussion](#) (#student wg-data-visualization channel)

- CICStudentWorkingGroup@columbia.edu (email us with any questions)
 - Join the [Working Group Listserv](#) to receive updates
-

COVID-19 Data Visualization Project

Milestone 1: Data Exploration and Preparation

Objective: Understand the dataset, clean it, and prepare it for visualization.

Steps:

1. What is Data Visualization and why it is important?:
 - Learn [what a data story is \(and is not\)](#). A data story has three elements: *data*, *visuals*, and *narrative*. We will explore *data* in Milestones 1-3, then discuss *visuals* in Milestone 4, and wrap up with a *narrative* description of your research in Milestone 5.
2. [Get Started with the Dataset!](#)

We have the following different types of data available via the link above. Download one or more of these data sets. In Excel, explore the data and consider the type of insights you want to discuss in a data visualization.

 - full_grouped.csv - Day to day country wise number of cases (Has County/State/Province level data)
 - covid_19_clean_complete.csv - Day to day country wise number of cases (Doesn't have County/State/Province level data)
 - country_wise_latest.csv - Latest country level number of cases
 - day_wise.csv - Day wise number of cases (Doesn't have country level data)
 - usa_county_wise.csv - Day to day county level number of cases
 - worldometer_data.csv - Latest data from <https://www.worldometers.info/>
3. Dataset Overview:
 - Explore the dataset's structure, including columns like date, location, cases, deaths, etc.
 - As you explore, [consider the research question you want to answer!](#) What interesting ideas emerge as you crawl through these spreadsheets?
 - Also, [decide who you are going to tell your data story to](#). Will your final audience be a group of scientists and peers? Is it the general public? How would you frame your analysis differently for each of these groups? Write down your initial ideas in an informal document. You will explain your rationale in Milestone 5.

4. Data Cleaning and Integration:

- [Learn best practices for deduplicating your data](#), fixing structures, and removing irrelevant data.
- [Begin the data cleaning process](#) in Excel.

5. Data Transformation:

- [Learn about data transformation and how it differs from data cleaning](#)
 - Convert date columns to a standard date format.
 - Aggregate data wherever needed (e.g., daily, weekly or monthly).
- Begin the data transformation process on your spreadsheet, remembering your initial research question. Keep the data you need to answer this question.

6. Exploratory Data Analysis (EDA):

- [Learn about EDA and initial investigations](#)
 - Generate summary statistics for key variables (cases, deaths, etc.). You may wish to [learn about pivot tables and create a few in your excel workbook](#).
- [Create basic visualizations](#) like line plots or bar charts to understand the data's distribution in Excel or Google Sheets. This will help you visualize key concepts as you refine your ideas.

Milestone 2: Time Series Analysis

Objective: Dive deeper into time series analysis and trends.

Steps:

1. Time Series Decomposition:

- Now, let's do a bit of math! We want to explore this dataset using statistics. We will deconstruct the data into components that represent categories and patterns. We'll explore Time Series Analysis by reviewing trends and seasonality. We will consider forecasting techniques, then wrap up by writing down our initial findings from this exercise.
- Learn how to [decompose the time series into trend, seasonality, and residual components](#) and why it's important for data visualization.

2. Trend Analysis:

- Let's start with trend analysis! [Learn what trends are in time series analysis](#).
- [Plot the trend component](#) in Excel (or a [Google Colab](#) notebook if you're [comfortable in Python](#)) to identify long-term patterns or trends.

3. Seasonal Analysis:

- Learn what [seasonality is in a time series](#) and why it matters to your analysis.
- [Visualize seasonal effects using appropriate plots](#) in Excel or a Google Colab notebook. Identify any recurring patterns (e.g. weekly or monthly effects). ([An advanced tutorial for seasonal analysis visualization can be found here.](#))

4. Time Series Forecasting (Optional):

- Learn about [forecasting techniques like ARIMA](#) or [exponential smoothing](#).

5. Your Summary Analysis:

- As you complete this Milestone, take some time to jot down notes on what you have learned. What key insights about COVID-19 have you learned from this exploration? What pieces of analysis do you want to bring into your final visualization and data story? How does this compare with the EDA you identified in Milestone 1?
- Put any questions you have into the [Slack channel](#) - we'll check in regularly to see if you're experiencing problems. Feel free to respond to other members' questions or even share a link to your work in progress!

Milestone 3: Geospatial Analysis

Objective: Explore geographical aspects of COVID-19 data.

Steps:

1. Geospatial Visualization:

- In the previous Milestone, we analyzed data by examining change over time. Now, let's consider how [data represented spatially](#) can also give us useful insights. We'll observe data when plotted at a regional level, then learn how to compare that data. We'll wrap up by writing down our initial findings from this exercise.
- Learn about [geospatial visualizations](#) and [how they can help with data analysis](#). Consider how this [visualization format is \(or is not\) helpful for analyzing COVID cases](#).

2. Regional Analysis:

- Plot the data on maps to visualize the geographic spread of COVID-19 cases using either [Excel](#) or a [Google Colab notebook](#) if you're [comfortable in Python](#).

- Consider using libraries like 'folium' or 'geopandas' for this. You can also create [heatmaps](#), [choropleth](#) maps, or other relevant plots.
3. Comparative Analysis:
 - Try a few different visualizations with your data. Compare the impact of COVID-19 across different locations or regions.
 4. Your Summary Analysis:
 - As you complete this Milestone, take some time to jot down notes on what you have learned. What key insights about COVID-19 have you learned from this exploration? What pieces of analysis do you want to bring into your final visualization and data story? How does this compare with the EDA you identified in Milestone 1?
 - Put any questions you have into the [Slack channel](#) - we'll check in regularly to see if you're experiencing problems. Feel free to respond to other members' questions or even share a link to your work in progress!

Milestone 4: Advanced Visualization and Storytelling

Objective: Create advanced visualizations and communicate insights effectively.

Steps:

1. Exploring Advanced Visualizations:
 - Now that we've considered some ways to analyze COVID-19 data, we're ready to visualize our ideas. Refer back to Milestone 1 and the research question you developed. Consider your target audience and what kind of story we'll need to tell to persuade them that our insights are relevant and informative. In this Milestone, we'll brainstorm and then build a visualization.
 - Explore [advanced chart types](#) (e.g., stacked area charts, bubble charts, etc.).
 - [Decide which visualization type is most appropriate for your research question](#). Compare types and how they communicate specific ideas or trends.
2. Build your Visualization:
 - Now we come to the fun part! Give yourself plenty of time to build your visualization. You can either use your Google Colab notebook (again, if you're comfortable in Python) or learn Tableau.
 - If you're interested in Tableau, [begin by creating a Tableau Public account](#). There are many good Tableau tutorials out there. We recommend the following to get you started:
 - [Learn Tableau in 15 minutes](#)

- [Tableau for Data Science and Data Visualization](#)
 - [Create Covid-19 in India Dashboard](#)
- Put any questions you have into the [Slack channel](#) - we'll check in regularly to see if you're experiencing problems. Feel free to respond to other members' questions or even share a link to your work in progress!
- 3. Data Storytelling:
 - Write 2-3 paragraphs summarizing your research question, your reasoning for selecting your data visualization choices, and key insights from your data analysis. You may also wish to include your outstanding research questions that could not be answered by the dataset.
 - Learn to [tell a compelling data-driven story](#).
- 4. Accessible Visualizations:
 - Consider [how your visualization might appear to people with disabilities](#) who may not be able to distinguish between muted colors or see your chart at all. Are there any changes you can make so that differently-abled scientists can also learn about your research?
 - Other tools:
 - [Color Contrast Checker](#)
 - [SAS Graphics Accelerator](#)
 - [TwoTone Data Sonification Tool](#)
 - [Making Visual Studio Accessible](#)

Milestone 5: Finalizing Visualizations

Steps:

1. Submit your Final Visualization:
 - You're nearly done! In previous Milestones, we've learned about the importance of data visualization, the different ways we can analyze data, and then found new ways to plot data through data storytelling. In this Milestone, we'll refine our visualization and get ready to share our visualization with others. We'll prepare how we discuss our visualization and request feedback from our peers.
 - Submit a final visualization that you feel best answers your research question from Milestone 1 to CICStudentWorkingGroup@columbia.edu. In your packet (any format you choose), include the 2-3 paragraph summary you developed in Milestone 4. Students who create and share a research submission may receive a certificate of completion for this project.
 - If you are using a Google Colab notebook, share your final visualizations with lc3460@columbia.edu.

- If you are using Tableau, publish your Tableau Public visualization so others can view your final workbook. Alternative: [Instructions for Desktop to Public publishing](#)

2. Future Practice (Optional):

- The best way to improve your data visualization skills is through continued practice. Keep your skills sharp by taking on additional visualization challenges, like these ones offered through [Makeover Monday](#).

Congratulations! You've generated a fantastic set of data visualizations for your educational and professional portfolios!