Data Analysis in Astronomy and Physics (SoSe22)

Instructor: PD Dr. Markus Röllig
TA: Dr. Christof Buchbender
TA: Craig Yanitski

1. Global CO_2 emissions [100 points]

Open the data in `co2-data.csv` to answer the following questions. It contains historical CO_2 emissions of each country per capita.

a. Compute the mean CO_2 per capita emission in 2017. What is the standard deviation and median? **20 points**

- Compute and display the statistics

b. This compares the data of countries in numerous stages of development. Try separating the data by continent, then calculate the mean, standard deviation, and median in 2017. Show this data in a convenient plot (maybe in a *box plot*). What are the limitations to this data reduction? **30 points**

- Determine a way to group the countries by continent
- Compute and show the statistics for the world and for each continent
- Compare the data in with a useful representation (some type of plot)
- Ensure the axes are properly labeled

c. To fully assess the contribution of each country to global emissions, we should look at the cumulative emission. What are the mean, standard deviation, and median of the dataset? Show this information for both the world and separated by continent. **20 points**

- Compute and show the statistics for the *cumulative data* (either calculated or use the `cumulative_co2` column) for the world and each continent
- Ensure you correct for the *time-dependance* of the data
- Compare the data in with a useful representation (some type of plot)
- Ensure the axes are properly labeled

d. We can also look at this history of CO_2 emissions by each country. Calculate the mean and standard deviation of the annual emission for the U.S., U.K., Germany, and China. Is this metric useful? In which year was the peak emission from these countries? Are they starting to gain control of their emissions? **30 points**

- Compute and display the statistics for the required countries over time
- Plots are very useful and important for understanding the data
- Infer and describe what the data shows