Programming in Science (420-SN1-RE)
Tiago Bortoletto Vaz
Lindsay V. Joseph 2466129 and Angeliki Stathopoulos 2373277
Due on October 31th, 2025


Lab 06 Assessment - Exploratory Data Analysis with Python

GitHub repository link:
https://github.com/lindsayjoseph-blip/Lab06-Exploratory-Data-Analysis-with-with-Python


Part 4 - Visualizing statistical relationships

1. Using the relplot() function from seaborn, generate a plot that helps you answer the following question: "Is there any association between GNI per capita and life expectancy?" (one plot for each gender).
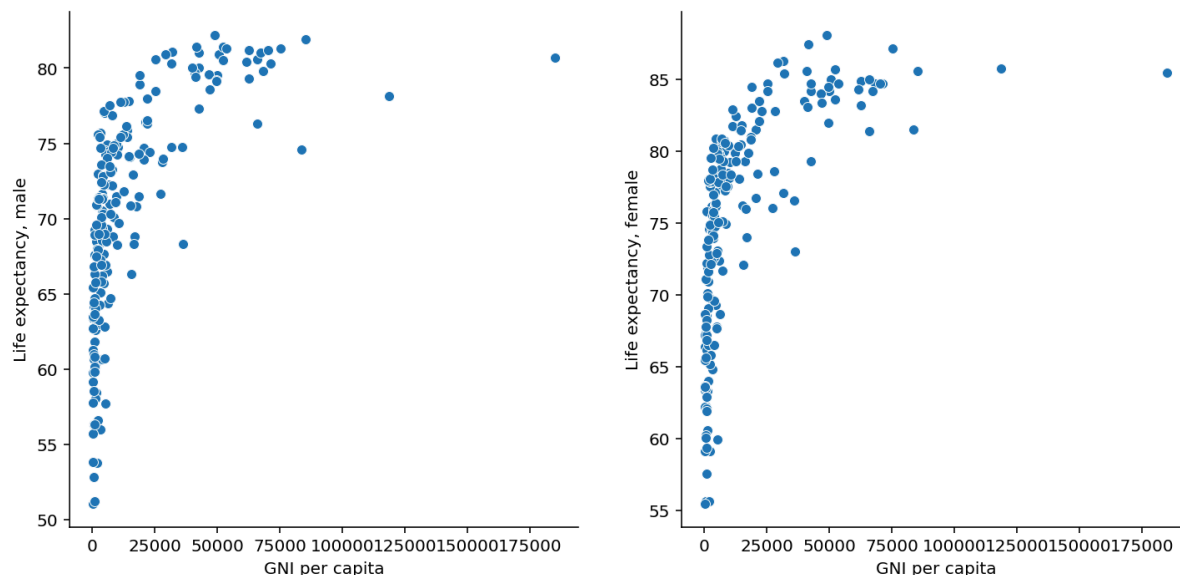
These plots showcase that there is indeed a significant relationship between the GNI per capita and the life expectancy for males and females. The life expectancy seems to increase as the GNI increases as

well, meaning that life expectancy is longer in higher income countries.

2. By adding a third "feature" to your plot using colors to represent it in order to answer the following question: "Does the association between GNI per capita and life expectancy vary by region?" (one plot for each gender)
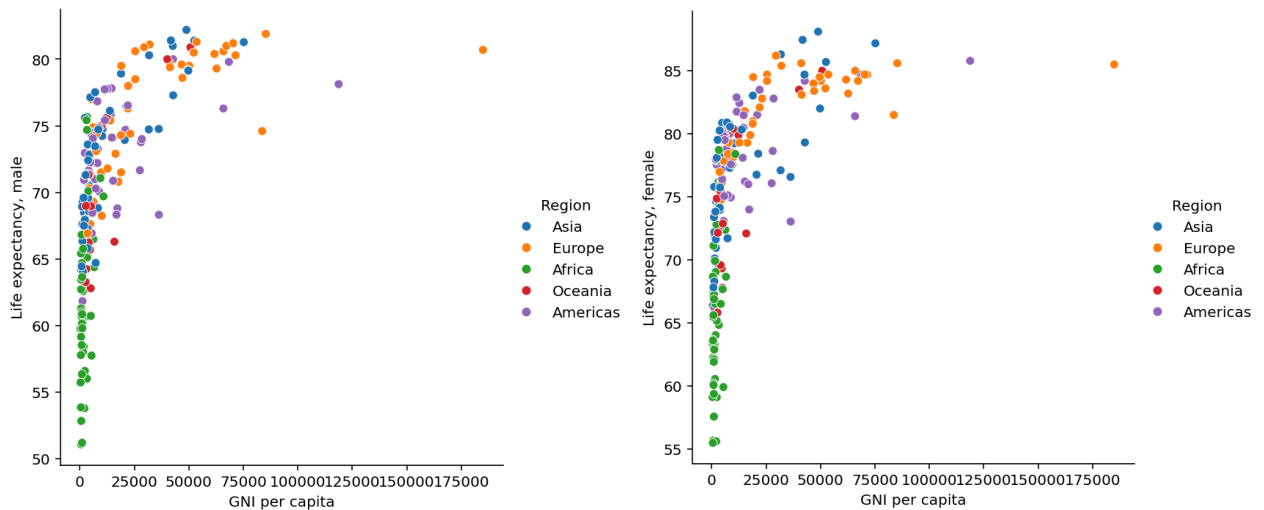
These graphs shows that countries in the region of Africa have lower life expectancy then other regions. This confirms our observations in the previous question, stating that life expectancy seems to increase as the GNI increases. Regions like Europe and the Americas have a higher GNI per capita and a longer life expectancy for both males and females.

3. Generate the plot from item 2, now using lines along with standard deviation.
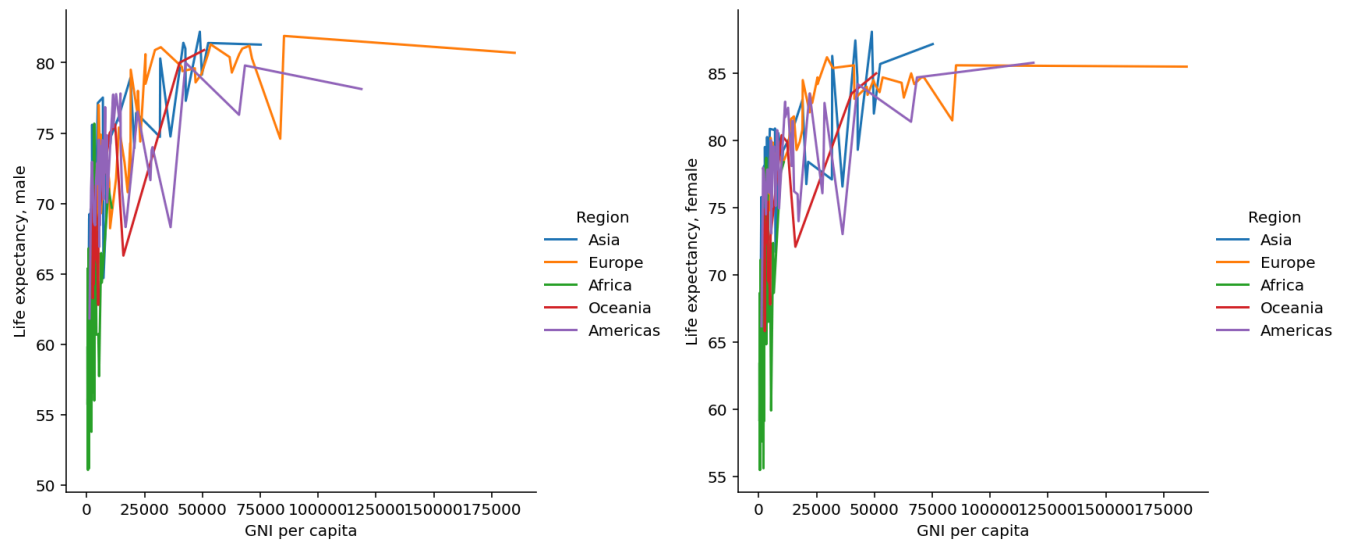
These plots represent the life expectancy of both male and female depending on GNI per capita in different region. This form of representation can be a clearer picture of how the life expectancy increases as the GNI increases (with a few fluctuations here and there).

#The reason why we can not see the standard deviation as a line is that it cannot be represented through only having one value per item category.

**4. Use the lmplot() function to generate a linear regression for the previous plot.**
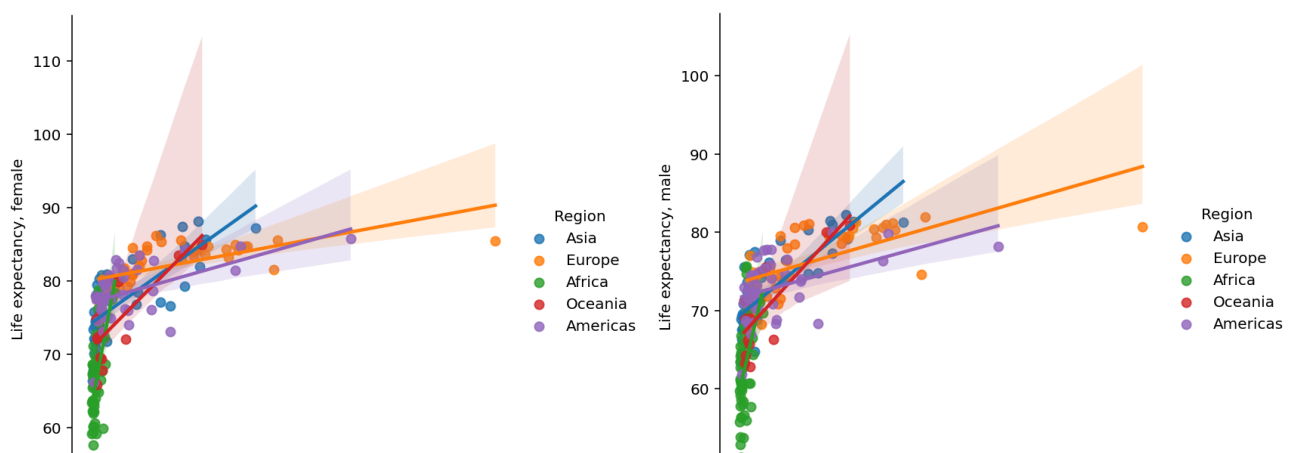
5.

    a) Use relplot() to explore relationships between female life expectancy and some of the other numerical features. Are these relationships similar for male life expectancy?
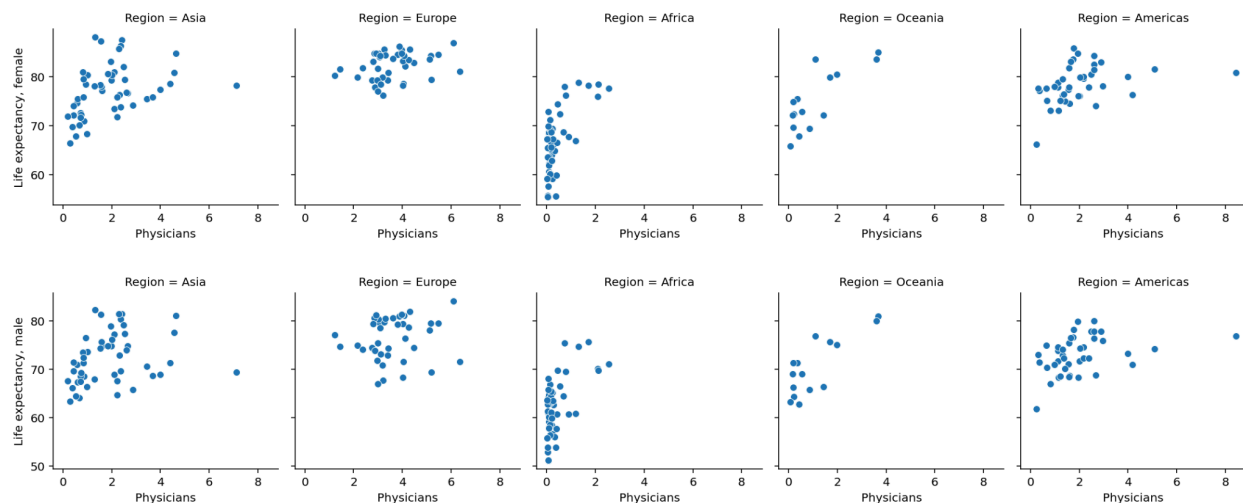
The relationships between the life expectancy of males and females in their respective regions as a function of the chosen numerical feature (Physician proportion) are indeed similar for each gender. Indeed, the more the proportion of physicians in each region, the greater the life expectancy for most. However, the region of Africa seems to be an exception, as life expectancy increases when the number of physicians stays low.

    b) Elaborate at least 5 more questions and generate one plot for each to help you answering them. Use "FacetGrid" feature from seaborn to visualize side by side results for each Region.

1. How does the relationship of woman in national parliament and females in tertiary education vary from a region to another?
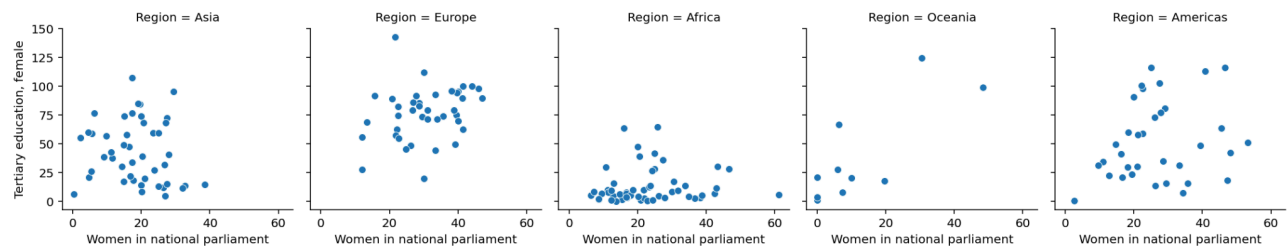
There does not seem to be a great correlation between the number of women in tertiary education and the proportion of females in national parliament for the region of Africa. Although there are not many women in tertiary education, they still manage to fill up space in the parliament. For the regions of Oceania and America, there seems to be a trend hinting that the more women in tertiary education, the more women in national parliament. The region of Europe has a very random distribution, while the region of Asia simply has a lower proportion of women in parliament than the others, regardless of the amount of women in tertiary education.

2. How does the relationship between the GNI and the International tourism vary from a region to another?
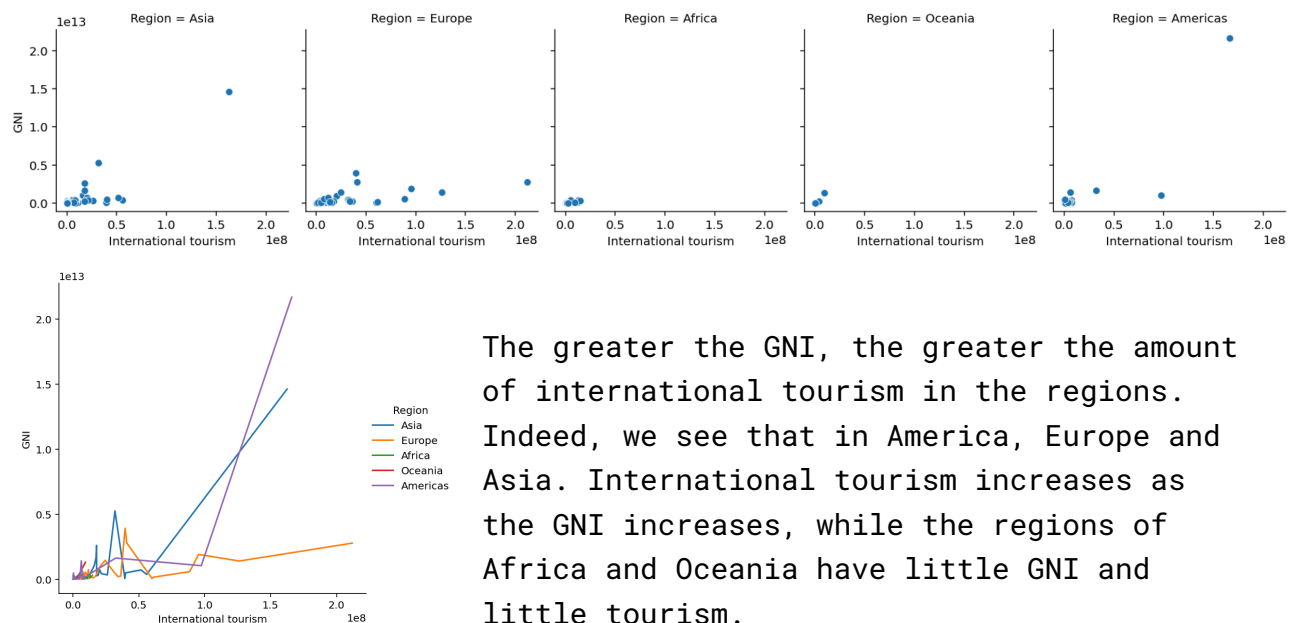

The greater the GNI, the greater the amount of international tourism in the regions. Indeed, we see that in America, Europe and Asia. International tourism increases as the GNI increases, while the regions of Africa and Oceania have little GNI and little tourism.

3. How does the relationship between the proportion of physicians and the women in tertiary education vary from a region to another?
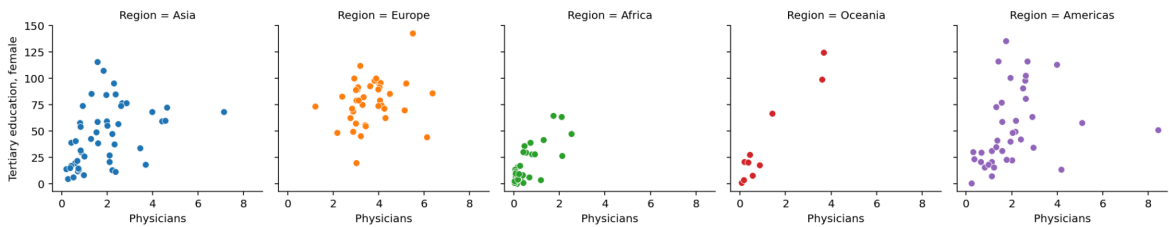

For all regions, the more women in tertiary education, the greater the proportion of physicians. However, it is notable that in the region of Africa, there are significantly fewer women in tertiary education than in other regions.

4. How does the relationship between the proportion of physicians and the men in tertiary education vary from a region to another?
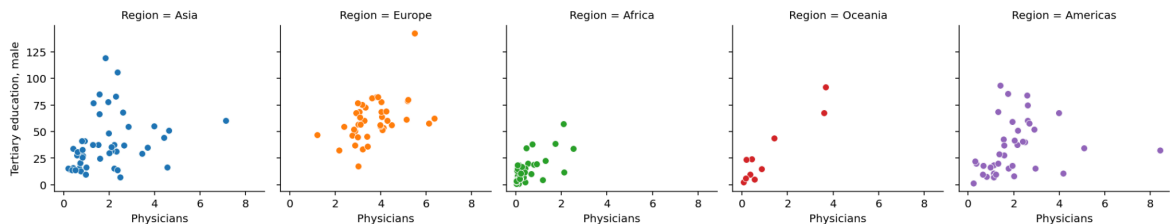
For all regions, the more men in tertiary education, the greater the proportion of physicians. However, just like previously, it is notable that in the region of Africa, there are significantly fewer men in tertiary education than in other regions.

5. Do regions with higher GNI per capita have higher internet use (access to internet)?
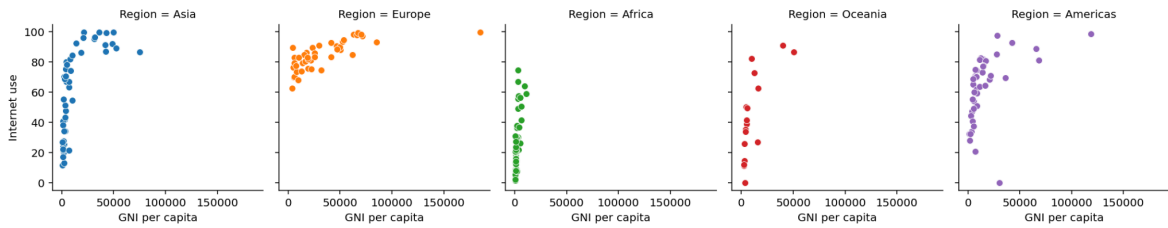
Yes, the graph shows that with greater GNI per capita, the countries tend to have higher internet usage. The richer the region, the more access to internet.

6. Using your new programming skills, answer the following questions:

**a) Is there any association between Internet use and emissions per capita?**
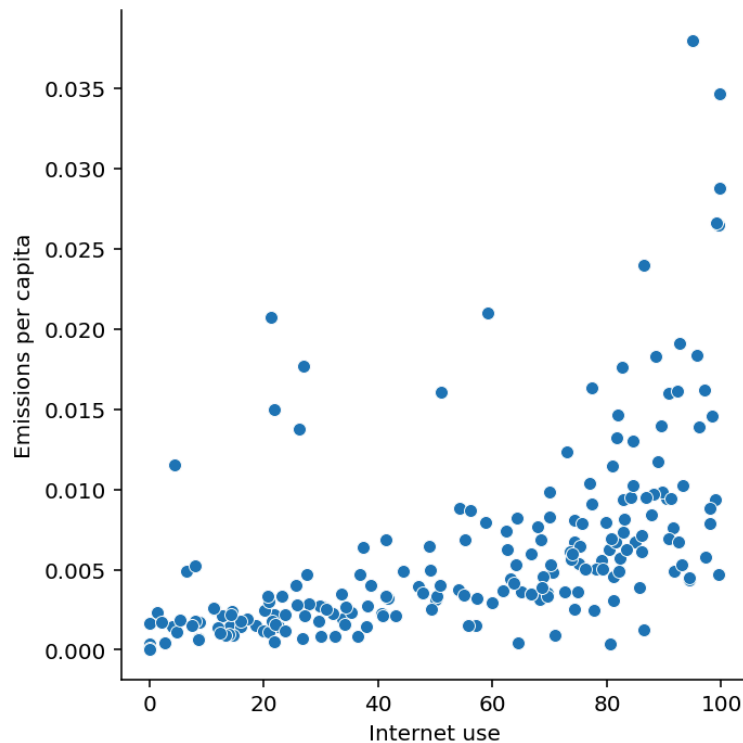
Yes, the graph shows a relationship between Internet use and emissions per capita however, it's not a very strong one. The scatter of the

data is forming an exponential curve but there are still a lot of points that don't follow the trend. Overall though, the data proves that as internet use goes up, so does emissions per capita and vice-versa.

b) Which are the countries with high emissions? (> 0.03)

The countries with high emissions are 'Brunei Darussalam'and 'Qatar'.

c) Is there much variation by region (with respect to high emissions vs Internet use)?
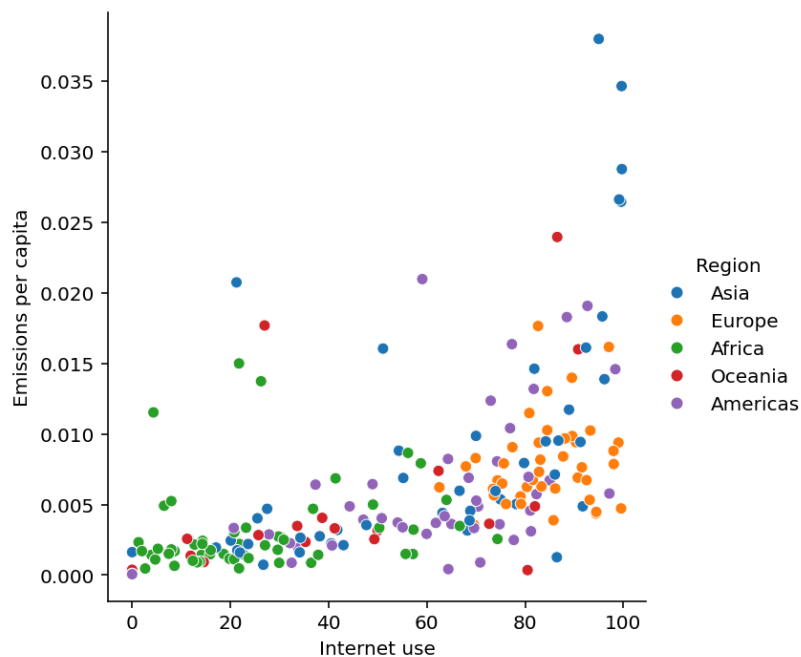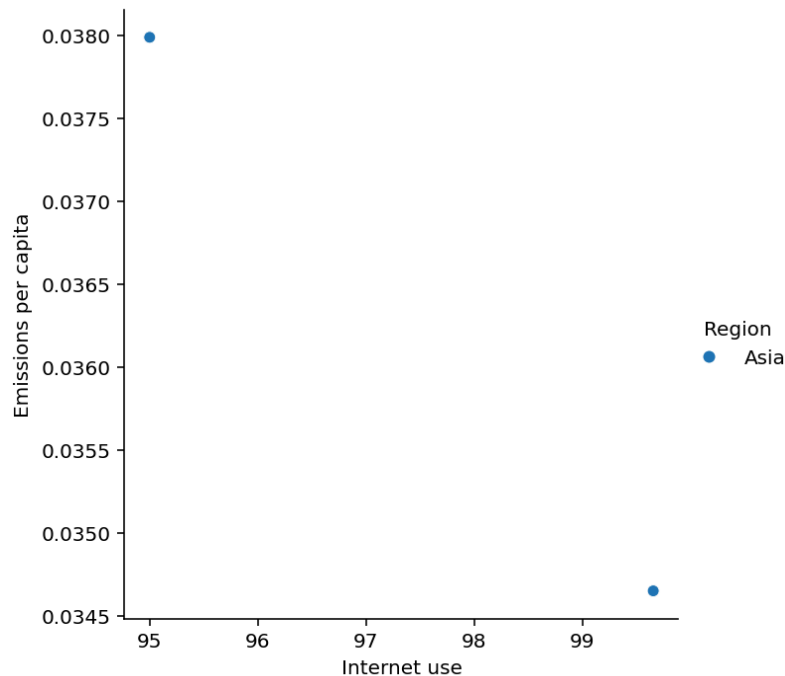
There is a slight variation between countries with high emissions
'Brunei Darussalam'and 'Qatar', in the same region (Asia), when it
comes to internet use. Indeed, we see that these 2 countries have an
internet usage between the range of 95 and 100, very closely related
but still varying slightly.

However, there is variation by region because when we color coordinate
the point by region, we see that Asia and Europe cluster towards a
higher internet use and emissions per capita while Africa clusters
lower for both. However, Americas and Oceania are pretty evenly
distributed and are towards the middle so they have medium internet
use and medium emissions per capita.

d) Do all high-income economies have high emissions?
No, not all high incomes have high emissions because we have learned
that 66 countries have high income economies while only 2 countries
have high emissions.

Findings and Difficulties

Through this lab, we learned how to use seaborn plotting. We encountered some difficulties because the seaborn functions all work slightly differently. However, we navigated the seaborn website which was very well organized and helpful. We also learned the fundamentals of analyzing data through python (.info(), .describe(), creating a new column) to comparing multiple plots (Faceting). The main difficulties we faced in this lab involved finding meaningful relationships in the data to build our questions and choosing representations that clearly showed our results. At times, it was challenging to decide which variables to compare, interpret trends correctly, and make sure our graphs supported our conclusions. However, we used our critical thinking to find questions we wanted answered about the data. We also developed the skill of a new way to analyze data and plot graphs. Of course, we encountered minor coding errors (missing parentheses, typos) which was frustrating, as always. It was often difficult to interpret the plots and find the trends and determine if the relation we found is significant.