

# Introduction to Stat for DS Groupwork-Based Assignment 2

## Reflection on Teamwork Experience

Durham Student ID: **001102553**, IT account: **bjsn39**

December 7, 2023

**Q1: Identify an aspect of your groupwork. What do you consider the primary objective or objectives of that aspect? Did you achieve this objective or objectives?**

I would like to discuss the data analysis segment in our group report, which stands at the core of our recent series of tasks. In my view, the most crucial aspect of this phase is the proficient application of mathematical knowledge and R programming commands acquired during the ISDS course to analyze the "Happy.csv" dataset. Even though the subsequent slides may be aesthetically pleasing, without careful consideration and extensive discussions among team members during this phase, they would merely be empty rhetoric. Despite discovering that this data analysis project was more complex than anticipated as our understanding of statistical concepts grew, we are grateful for successfully completing this task.

**Q2: Why did the team achieve/not achieve your objective or objectives? What challenges did you face, and how did you attempt to overcome them? Were you successful in overcoming them fully?**

I believe we successfully achieved our objectives because our data analysis in the group presentation was comprehensive. Starting from the data preprocessing phase, we addressed missing values, corrected errors in continent data, and later efficiently segmented data for different continents, we also developed functions for multiple linear regression analysis. In addition to statistical analysis visuals generated in R Studio, our slides feature beautifully crafted flowcharts, aiding both our team members and the audience in understanding the work of our team.

One of the major challenges we encountered was the absence of some team members from classes or their undergraduate backgrounds not being in STEM fields, making it somewhat challenging for them to grasp statistical concepts. This resulted in a lack of presence in later group discussions, making it difficult for them to integrate into our conversations and provide effective suggestions for our project. I empathize with the situation of those team members and understand that hands-on experience is the best way to master knowledge. Therefore, in the work of the data analysis module, I initially assigned them tasks related to drawing

flowcharts, helping them understand which statistical tools our team would be using. This significantly boosted their confidence. Simultaneously, in the code section of data analysis, each time I updated the R code for this project, I shared it on GitHub to assist everyone in understanding. Moreover, I provided detailed comments for each code segment. In the end, when each person recorded their video segment for the presentation, I am confident that everyone comprehended each step of our data analysis.

**Q3: What have you, as a member of this team, learned from attempting to meet this objective or these objectives?**

As written in Group-based Individual Assignment 1, I consider myself a proficient organizer within the team. Even though assisting team members in learning statistical knowledge consumed a significant amount of our team discussion time, as these concepts and techniques could have been acquired through ISDS workshops and video resources on Ultra, in the long run, not giving up on each team member and enhancing their engagement in the project is the optimal way to improve the quality of our group presentation.

At the same time, thorough consideration of issues remains crucial. The following example illustrates this: the first instance was when we were addressing the issue of regions in the dataset, around the seventh week. Initially, our idea was to map continents from strings to integers. However, through discussions, we realized that assigning a value like 1 or 6 to a continent in linear regression was not logically sound. Therefore, after comprehensive deliberation, we decided to split the data for each continent into independent datasets. Additionally, we analyzed the six datasets without including the continent information in the overall analysis. The advantage of this approach is that it allows for a global analysis and provides customized solutions for each continent.

**Q4: Imagine this module lasted for another term, and I asked you to prepare another mini-report and presentation in the same team. What changes would you wish to make in the way you work as a team?**

If there is a new project, I hope our team can achieve a roughly similar level of proficiency in statistical knowledge, rather than having a lag of 1-2 weeks in progress. Otherwise, it may lead to limited cognitive engagement for some team members, making it challenging for them to integrate into group discussions, and it could adversely affect the overall team morale. Through this group report project, I have gained insights into each person's expertise in video production, R programming, slide creation, and chart design. For the new project, we plan to directly assign each task to the individual's strongest area, avoiding the extensive time required for team coordination as seen in this assignment.

In addition, we also need to ensure that the technologies we use are the most recently learned. In the process of learning our statistics course, especially in the R language-related parts, during theoretical learning, we often implement specific functionalities by defining our own

functions. However, in practical exercises, particularly in workshop files, we discovered that certain functionalities are already implemented internally in R language, or can be directly achieved by installing external packages. Therefore, I believe keeping ourselves updated with the latest knowledge is essential, significantly enhancing the readability of our project code.