PACC – Data In Action
Part 1 Tasks

Description of Data

The University of Washington Athletics provided data for this challenge, which we have shared with you.

Athletes on the women's volleyball team and men's soccer team submitted daily self-reports of their fatigue, sleep quality, general soreness, stress, and mood on a scale of 1 to 5 and how many hours of sleep they got. For each of these metrics, the scale is as follows:

(1)	→	(5)
less fatigued	→	more fatigued
worse sleep quality	→	better sleep quality
less stressed	→	more stressed
worse mood	→	better mood

This information is contained in files called `player_data_soccer.csv` and `player_data_volleyball.csv`. We've also provided you with the dates, locations, and results of the teams' games, in separate files called `team_results_soccer.csv` and `team_results_volleyball.csv`.

These files should have been emailed to you along with this task list, but please contact us if you believe you did not receive them. `.csv` files can be opened and processed with Excel or Google Drive Spreadsheets. You should even be able to open them as plaintext to get a general idea of what they contain. Although it is not required, you may want to write code to analyze this data. There are many examples online

Instructions

Please complete each of the tasks below using the data provided. Your answers will range from one-line sentences to full paragraphs, and may or may not include a visualization such as a timeline or graph. Feel free to include a visualization if you think it is useful, even if it is not required for a certain question.

The questions were designed so that they can be answered without needing to write code, though writing code may be easier in some cases, or might allow you to explore the data more thoroughly, which could be particularly useful for answering some of the more theoretical questions.

In addition to simply answering the question, let us know how you approached the problem and tell us if any of the answers surprised you. Did you sort the Excel spreadsheet by a specific column? Did you write some code? Did you generate a graph and then examine it? If

you write code or use graphs to help answer the questions, please include these at the end of your answer.

If possible, work together on each of these tasks (don't just divide them up and go your separate ways). Your teammates' varied perspectives will help you think more deeply about the questions. If you disagree (or find discrepancies in the data), feel free to provide multiple answers and an explanation of why you disagreed.

Tasks

General Data Analysis:
1. Which soccer player had the most sleep per night throughout the season?
2. Is there a significant difference in the average amount of sleep for athletes on weekends vs. weekdays?

General Charts & Graphs:
3. Create a histogram of the hours of sleep players on the volleyball team got throughout the season. What is the range of the number of hours of sleep? Which player(s) slept the most/least in a night? Make one observation of your own about the data in this chart.
4. Plot stress vs. mood for volleyball players (scatter plot). Do you think this plot is useful for showing how (or if) mood and stress are related? Why or why not? Suggest an alternate plot or a few changes that would improve this visualization. You may actually generate this plot or simply describe and sketch it.

Game Day Preparation:
5. Do athletes sleep more the night before a game?

Game Day Impacts:
6. Do athletes feel more stress on game days or on days without games?
7. Do athletes feel more stress before away games? If so, does the distance travelled to the game seem to impact stress levels?
8. Which team's players (soccer or volleyball) report being more sore on the day after games?
9. Are athletes in better (or worse) moods the day after a game? Does it matter whether their team won or lost?

Annotated Timelines:
10. Choose any soccer player or volleyball player (let us know who you chose). Create a set of plots (ex. timelines) that show all of the data for this player throughout the season. Annotate the timeline with information that you think might explain any patterns or spikes in the data. For example, if fatigue appears to fluctuate regularly with the day of the week, you might draw tick marks in one-week increments on the

time axis; if there is a huge spike in soreness right after a big game, you could add an arrow and text to highlight this. Here are a few examples found online based on different data: fitness trackers after an earthquake; history of oil prices; GM recalls. You are encouraged to produce something very different from these examples -- get creative!

Open-ended questions:
11. How might a coach, player, or athletic trainer use this data? For example, could the coach use this data to improve win/loss records for the team; could the player improve his/her own performance; could the trainer use it to help prevent injuries?
12. How might the fact that this data is self-reported impact how useful it is?
13. What additional types/sources of data might be useful for teams?

What's Next?

An answer key for evaluation will be available here (sharing link will only be open after the challenge has finished).

Your answers will be evaluated based on completeness and correctness. Partial credit will be awarded for teams that show their work. Extra points (and kudos) will be awarded for especially creative methods, thoroughness, or original thought in the open-ended questions section. Don't worry too much about your score in Part 1 -- all teams that complete Part 1 will be invited to participate in Part 2, which will count significantly more in the overall competition.