Student	Chris Hampton	86

Part 1

	Available	Marks
Q1 The file *roller_coasters.csv* contains data on a number of roller coasters from all around the world. This shows the name, type, material, manufacturer, location and current status for each roller coaster and may also show data on the height, length, speed, and number of inversions achieved by each roller coaster. Load this data into a data frame *coasterData*, ensuring that the headers are imported and that the type, material, manufacturer, and current status are imported as factors. Your script should assume that the data file is stored in the same folder as the script file. Display the resulting data frame.	5	5
Q2: Create a new data frame *highestData* that shows the name, height, park and status of the highest operating roller coaster of each type of roller coaster. Display this data frame.	10	10
Q3: Include a function *displayComparison()* that will take the seating_type as a parameter and will produce a bar chart showing the speed of the top ten fastest operating rollercoasters of that seating type. Display this using inverted roller coasters.	15	14
Q4: Include a function *correlateLengthSpeed()* that generates a data frame *lengthSpeed* that shows the mean height, length, and coefficient of correlation of length and speed for Sit Down roller coasters These calculations should be based on records in *coasterData* that contain no NA fields.	15	10
Q5: Include a function *compareWoodenSteel()* that performs a t-test to see if there is a significant difference at the p=0.05 level between the speeds of wooden rollercoasters and steel rollercoasters. Include your hypotheses and your conclusions in your comments. These calculations should be based on records in *coasterData* that contain no NA fields.	10	5
Commenting, white space and code presentation	5	4
TOTAL		48

Comment

Q1: Excellent error handling.

Q2: Good use of order()

Q3: The grey background is a bit confusing, but otherwise excellent. The selection of axes to allow the names to be clearly visible is an excellent idea.

Q4: Your function is elegant and works well, but you don't actually create the required data frame *lengthSpeed*.

Q5: Provides the wrong answer – there is a significant difference between the values for Steel and Wooden. Elegant code, but what are your null and alternative hypotheses? These should be included in the comments.

File should be rollercoaster.r Otherwise very good.

Part 2

	Available	Marks
Explain how your code works for each section (including flow charts) and how you have extracted the required data from the data frame where you have stored the imported data. TASK 1	30	30
A reflection on what you found challenging, what you learned, and what you would do differently next time.	10	8
TOTAL		38

Comments

Starting with an overall flow diagram is a sensible approach, as is the inclusion of the prerequisites. Excellent level of detail. The comparison with the AI explanation is interesting; I can certainly see it becoming standard practice to have Ais do the documentation of code.
The reason for the disparity between the Excel and R calculations is that there is a flaw in the R code.

An excellent piece of work. Well done.