Question 1

(a) Does mobile phone usage affect young adult brain's memory capabilities?

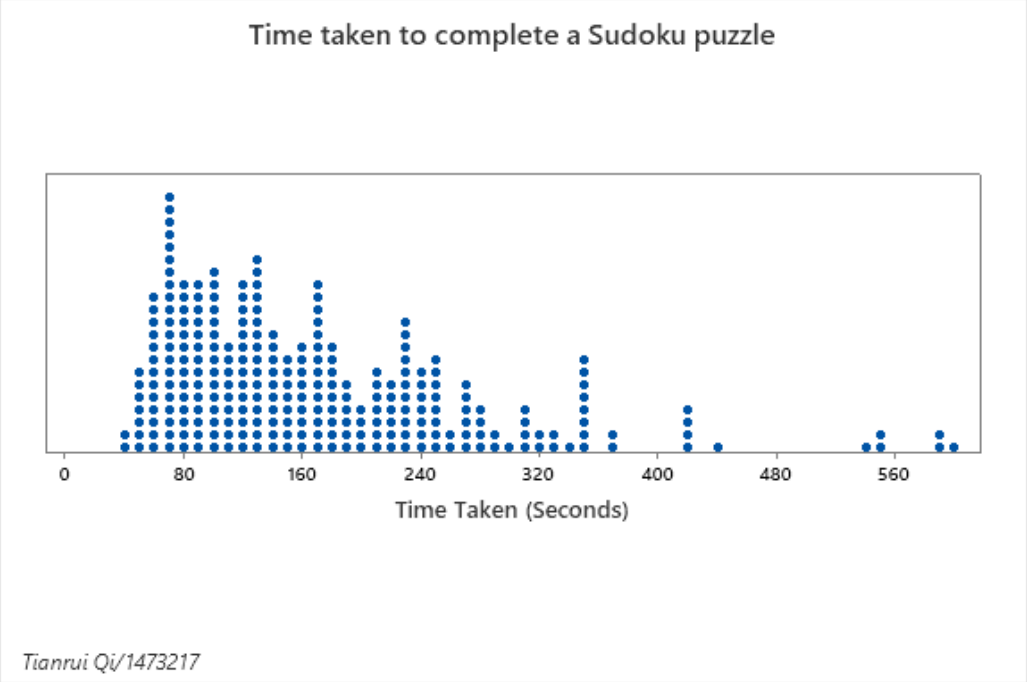
(b) The response variable is the score (out of 100) on a 30 minute memory test taken by the subject. It is a discrete numeric data. The primary explanatory variable is the grouped average hours of non-essential mobile phone usage per day across a month, between low (less than 1 hour), medium (1 hour to 2 hours), and high (more than 2 hours). It is an ordered categorical data.

(c)

- The study would aim to randomly sample young adults both through online methods and at concentrated young adults areas. Randomizing to assign subjects to the study groups (of low, medium and high), may not be desirable as there is likely a confounding variable between existing device use and brain functions — blocking between existing average device uses may bring less biased results.
- The study should enact control by restricting, based on age (say from 18 to 30), the subjects sampled as to reduce the confounding variable underlying age and brain memory functions. A sufficient protocol to reduce bias could be to conduct the memory tests in the morning starting at the same time for the same duration. Blinding is not possible for the test would be objective and the subjects clearly know their experimental groups.
- To implement replication, the study aims to sample around 500 to 1000 subjects to ensure sufficient subjects as to drown out the unaccounted confounding variables. It should also aim to sample at least 100 of subjects each for the 3 blocks used, based on their current device usage hours, to increase replications within the blocking groups. It is also possible to perform another memory test in conjunction with the existing morning test, but in the afternoon, to increase precision through repetition in measurements.

Question 2

(a) Figure



Statistics

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	IQR
Time (Seconds)	269	171.90	106.25	40.00	90.50	142.00	228.37	600.00	137.87

(b)

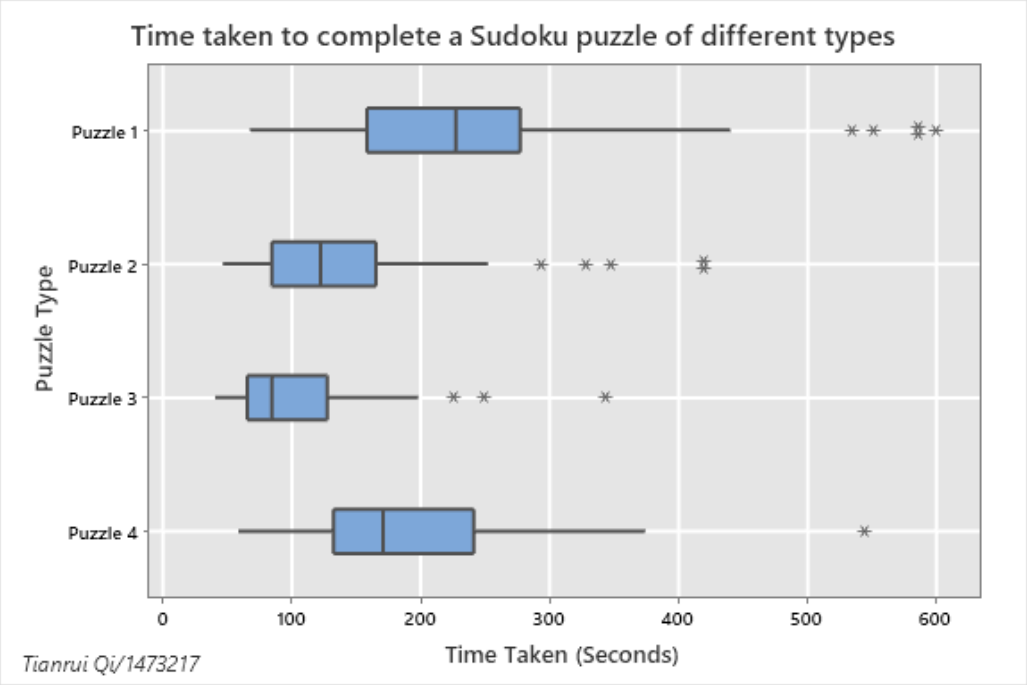
- The distribution is positively skewed with a tail to the right. Its shape resembles a skewed bell curve.
- The median of the distribution is 142 seconds, with the mean larger at 171.9 seconds.
- The spread, using the IQR, is quite large at 137.87 seconds. When measuring in standard deviations, the data points are on average 106.25 seconds away from the mean.

(c) The lower scoring times are all completed on the puzzle types of 2 and 3, namely the alphabetical and numeric Sudoku puzzles.

Question 3

(a) I think that, barring some minuscule differences in the coloring of the papers, this was a fair comparison between Sudoku puzzles of different symbol types. The reasoning behind is the identical placements of the 6 symbols on the starting tiles between each puzzle types, and a one-to-one mapping of symbols between the puzzles: the beta on puzzle 1 is equivalent to the number 2 in puzzle 3. This ensures that the puzzles are only different by the symbols presented, not in the difficulty inherent of the puzzle.

(b) Figure



Statistics

Variable	Puzzle number	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	IQR
Time (Seconds)	Puzzle 1	67	244.5	125.5	67.0	158.0	227.0	277.0	600.0	119.0
	Puzzle 2	75	139.30	80.14	46.00	84.00	122.00	165.00	419.00	81.00
	Puzzle 3	59	104.93	57.88	40.00	65.00	84.00	127.00	343.00	62.00
	Puzzle 4	68	194.5	93.4	58.0	132.0	170.5	241.0	545.0	109.0

(c) From the boxplots, we can see various outliers to the right as well as some skewness for each puzzle (puzzle 3 and 4 seems positively skewed). This indicates that we should use the medians for a less-biased comparison of the distributions.

Statistically, we can see that puzzle 3 had the least median time taken of (84 seconds), and puzzle 1 took the longest median time to complete (227 seconds), with puzzle 2 being second place (122 seconds) and puzzle 4 being third place (170.5 seconds). This order is also exhibited throughout the entire 5 number summaries.

Since the data originated from a designed study and the puzzles only differed in symbol types, the data supports the idea that the symbols in a Sudoku puzzle affects the time taken to successfully complete it for students in this class. Numbers lead to the fastest times and Greek letters to the slowest.

Question 4

(a) Judging by the small differences between the observed and expected counts (computed through row and column percentages) in most the table cells, there seems to be no significant association between the puzzle types assigned and the Sudoku experiences of the subjects.

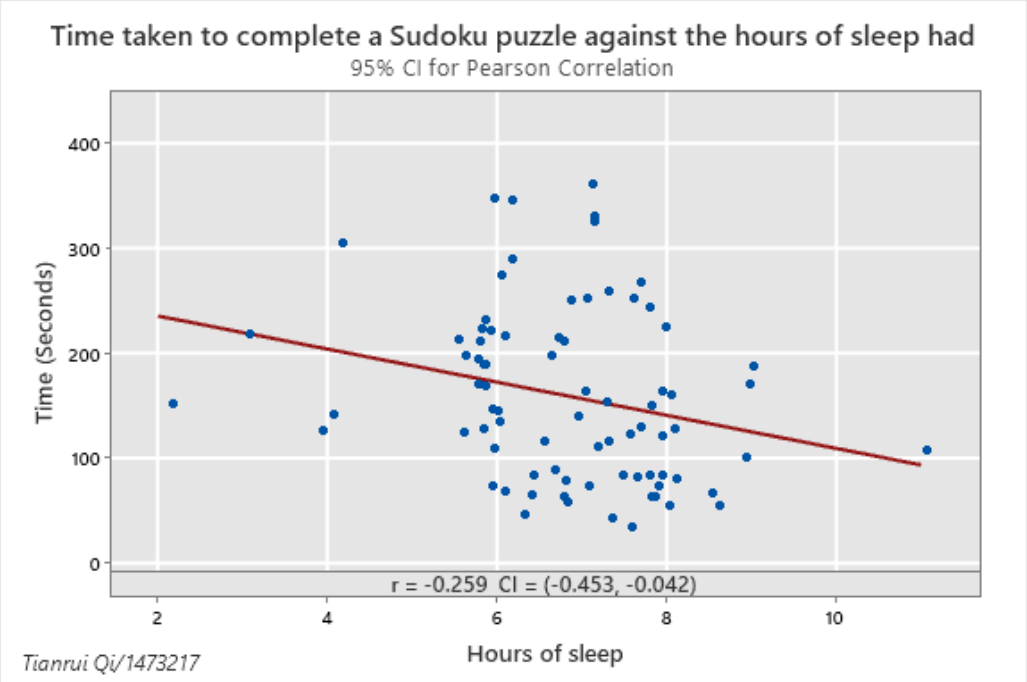
The randomization was adequate for the group sizes of all four puzzles are relatively equal at around 59 to 68. The lack of significant associations also indicates that a representative proportion of subjects with no Sudoku experiences were assigned to each puzzle type due to randomization.

Rows: Have you played Sudoku before? Columns: Puzzle number

	Puzzle 1	Puzzle 2	Puzzle 3	Puzzle 4	All
No	19	17	9	17	62
	15.44	17.29	13.60	15.67	
Yes	48	58	50	51	207
	51.56	57.71	45.40	52.33	
All	67	75	59	68	269

Cell Contents
Count
Expected count

(b) From the figure, there is a weak relationship between the hours of sleep the subjects had last night and their time taken to complete a Sudoku puzzle. The respective r value is negative but small at -0.259, indicating that higher hours of sleep is weakly associated with faster puzzle completions.



Question 5

(a)

- The graph has a clear title that includes both the primary explanatory variable (surgery types) and the response variable (regret rates).
- The pie chart is wrongly used to display unrelated regret percentages between the studies of different populations; the sum of the percentages in the slices adds to 103%, a pie chart should have its slices add to 100%.
- There are no units under the labels of each slice. For instance, it should read "Bowel 32%" instead.

(b) Figure