

Figure That!

EVENT CATEGORY: Inquiry and Problem Solving

EVENT DIVISION: Division C

1. **DESCRIPTION:** This event aims to assess student mastery of scientific figure interpretation, creation, and critique. (For any clarifications/questions, competitors should reach out to kevinbao1444@gmail.com.)

A TEAM OF UP TO: 3

EYE PROTECTION: None

IMPOUND: No

APPROXIMATE TIME/EVENT TIME: 50 minutes

2. **EVENT PARAMETERS:**

- a. Each team will be provided with 1 computer that has access to Google Colab set to the R runtime. There will also be a local folder on the computer with relevant files for the event. No other files, directories, websites or applications may be accessed during the event. The Google Colab file will contain lines of code at the top to install the “ggplot2” and “dplyr” package; these lines may be run during the tournament, and no other packages may be installed by the student. A Google Form URL will also be included in the Google Colab file for final submission. Event supervisors will ensure that all clocks on all computers are synced.
- b. Teams may not use any other electronic device other than the provided computer. The computer is to only be used for the figure creation task.
- c. Students’ electronic files will be submitted or collected electronically. Event supervisor will provide details at the start of the event.

3. **THE COMPETITION:**

- a. Students will be given a test booklet containing the questions they are to answer and a full page, separate description of their figure creation task.
- b. Teams will begin the event simultaneously at the indication of the Event Supervisor. At this point, the timer also begins (see 4. SCORING)
- c. Students will be asked to be familiar with coding in R and graphics in both base R and ggplot2.
- d. The competition will consist of two parts: (i) a Test and (ii) a Figure Creation Task
 - i. The Test Component will consist of questions asking students to interpret a scientific figure or point out issues with its construction/implementation. All figures presented will include a legend that details all abbreviations, scientific jargon, and key information required to interpret the results. Competitors should not be required to have extensive specific subject knowledge to answer a question correctly but instead be tested on their understanding of statistical tests, data visualization, and general data science principles. Whenever possible figures should be drawn from primary literature and sources included. Topics and tasks that can be assigned are as follows
 - (1) Code analysis: students will be asked to be familiar with **Loops (for, while, for each), “if” structures, Piping, Boolean logic and operation, “plot” function, Variables, Manipulating data structures (vectors, arrays, data frames), the “apply” family of functions (e.g. sapply, capply), arithmetic and math.**
 - (a) Students will not be asked to assess the appropriateness of statistics or methods in code. Students will not be asked to correct mistakes or identify the efficiency of code/improve it.
 - (2) Figure analysis: students should be familiar with **types of graphs, how data is presented in a figure, graphical abstracts and other similar figures, best practices for figures, statistical tests, common figure fallacies, how to modify a graph to display other information, reverse engineering code from figures.**
 - (3) General knowledge about **statistics, statistical tests, data science, data types, and data manipulation.**

ii. The Figure Creation Task will require the students to generate a figure from a given dataset (csv, txt, or other file type openable with a simple text editor). Students will be provided with a readme file that details the type(s) of data in the data file and how large the data file should be. Students are tasked with reading the data into Google Colab and creating a figure given certain parameters. The type of figure to be created should be left to student discretion; however, event supervisors may choose to restrict, promote, or otherwise limit what kinds of figures students may choose. Once students are done, they should export their Notebook file as a “.ipynb” and submit to a Google Form with a link located at the top of provided code file. If more than one file is submitted, students will receive points only for the earliest time-stamped file submitted. Students may be tasked to do the following:

- (1) Import, wrangle and rearrange data
- (2) Adhere to specific aesthetic considerations
- (3) Screen data for outliers, junk, and specific categories
- (4) Create and modify axes, error bars, legends, jitter, and other common plot elements
- (5) Compute statistical metrics including but not limited to mean, standard deviation, confidence intervals, coefficient of variation, and error.
- (6) Resample the data (e.g. bootstrapping, jackknifing) (NATIONALS ONLY)

4. **SAMPLE TASKS/STATIONS/QUESTIONS:**

- a. If the above figure was made in Base R using the code “plot(data\$time~data\$height)”, what would you add to that code to make the y axis labels horizontal?
- b. Based on the figure above which of the following conclusions can you draw? What about this figure could be misleading?

5. **SCORING:** Describe how scores will be awarded in your event. Topics that you need to address are:

- a. Participants will be given points based on how well they do on the Test and Figure Creation Task.
Highest Score wins
 - i. Final Score = Test Score + Figure Score + Timing Bonus
- b. Test Score = Raw Score/Max Possible Score * 1000
- c. The Figure Score = Raw Score/20 * 1000
- d. Timing Bonus = Time Remaining/Total Time * 200 (rounded up to nearest minute)

6. **RECOMMENDED RESOURCES:** Students and event supervisors are encouraged to read *The R Graphics Cookbook* by Winston Chang. It is available for free online at r-graphics.org. Google Colab is available for free online.

7. **Sample Figure Rubric:** Figures are to be scored based on the following Rubric. Event supervisors are to score students based on the following and sum up their score for all categories

- a. Figure Basics (3 max): Has all basic aspects of a figure like title, axis labels, axis titles, etc
 - i. 3: has all basic elements
 - ii. 2: is missing 1 element
 - iii. 1: is missing 2-3 elements
 - iv. 0: is missing critical parts of the figure/many elements
- b. Data (5 max): The data is presented in a way that does not change its meaning and is easy to interpret
 - i. 5: All relevant data is present and clear and all data to be filtered is not present.
 - ii. 4: All relevant data is present and clear but some data to be filtered is present even if not visible
 - iii. 3: All relevant is present but some of it is unclear or muddled no data has been filtered
 - iv. 2: Not all relevant data is present but all present data is clear to interpret
 - v. 1: Not all data is present and what is present is somewhat unclear OR all data is present but it is presented in a way that is impossible to interpret
 - vi. 0: data not at all present or not presented in a way that can be interpreted.
- c. Statistics (5 max): Requested statistical values are present and shown in a sensible easily understandable way (e.g. brackets with stars)
 - i. 5: Relevant statistical tests are performed and the results presented clearly in the figure
 - ii. 4: Relevant statistical test are performed but the results are not presented clearly in the figure
 - iii. 3: Statistical Tests are not performed correctly but results are presented clearly in the figure

- iv. 2: Statistical tests are not performed correctly, and results are not presented clearly in the figure
 - v. 1: Statistical tests are performed correctly but results are in the code only.
 - vi. 0: No statistical tests are performed
 - d. Caption (3 max): Figures should be accompanied by a caption
 - i. 3: A figure caption accompanies the figure and describes the type of plot, what the data represents and accounts for all elements not otherwise described in a legend or key.
 - ii. 2: A figure caption accompanies the figure but it fails to describe one of the above
 - iii. 1: A figure caption accompanies the figure but it fails to describe more than one of the above
 - iv. 0: Figure caption is absent or nonsensical
 - e. Aesthetics (2 max): Figure should be pleasing to the eye and easy to read for all people
 - i. 2: Figure is clear and there is color or patterning clearly separating different data groups. Patterns and colors are used efficiently and clearly. If colors are not needlessly garish or inaccessible to the colorblind (e.g. using green and red when other options are available)
 - ii. 1: Figure is unclear and/or color and pattern choices are wanting
 - iii. 0: Figure lacks all color or patterning and/or is entirely illegible
 - f. Figure Specifics (1 max): All other figure specifications not covered in other sections were adhered to.
 - i. 1: all parameters provided by the event supervisor was adhered to.
 - ii. 0: not all parameters provided by the event supervisor were adhered to.
 - g. Reproducibility (1 max): Figure should be able to be recreated by running the code submitted
 - i. 1: figure can be recreated by running all the code submitted
- 0: figure cannot be recreated by running the code submitted