



DataJam Guidebook - 2024

DataJam Guidebook

Welcome to the Pittsburgh DataJam! We hope this will be a rewarding and enriching experience for you. This guidebook will help guide your journey through the process. The guidebook is laid out as follows:

Timeline	2-3
Proposal Guidelines.....	4
Proposal Template.....	5
Parent Permission Slip.....	6-7
Poster & Presentation Guidelines.....	8
Rubric for Assessment of Final Submission.....	9-13

Timeline

There will be several important dates to keep track of, with the proposal due date the first deadline. If your team feels you can not meet a deadline, please contact us at datajam@thedatajam.org and we will try to accommodate your needs if possible. However, as time gets closer to the DataJam finale we have less flexibility.

September – December 2023

Sign up to be a DataJam team; send an email to pghdataworks.org with (1) the name of your school, (2) the name, email and phone number of the teacher advisor, (3) the names and email addresses of each student on the team. We strongly encourage all students to have their parents sign a permission slip and scan it and email it to us to allow participation on the DataJam 2024 Edtera workspace: an online workspace for all DataJam teams where they can easily communicate with each other and with DataJam mentors.

December 8, 2023

Deadline for teams to turn in DataJam proposals. Use the proposal template on page 6 of this guidebook and email it to datajam@pghdataworks.org. Your team will receive feedback from DataWorks and further guidance on your proposed project by mid-December.

January – Late March 2024

Teams work on DataJam projects. DataJam mentors are available to meet with your team by zoom or a videoconferencing service that works best for your team. You can arrange for a team mentor by contacting us at datajam@thedatajam.org. Give us some options when you contact us about the days/times that would work best for your team to meet. Once you have a team mentor you can communicate with them regularly on the DataJam 2024 Edtera workspace.

March 29, 2023

Poster submission deadline. Follow the instructions on page 8. Send your final poster as a .pdf file to datajam@thedatajam.org

Week of April 15, 2024

Final slide presentations will be scheduled for each team, at a time that works for both the team and the 3 judge panel who will be listening to each team. Dr. Cameron at The Datajam will be scheduling the final presentation times. Your presentation will be evaluated using the criteria outlined on page 9.

April 25, 2024

DataJam 2024 Finale! The finale will be held on Thursday, April 25, 2024 from 5:30-7:00 PM, EST on Zoom at link: <https://pitt.zoom.us/j/99862628516>
DataJam teams, parents, teachers, mentors, and all who are interested are invited to attend. DataJam team projects will be presented, and a variety of awards will be given!

Proposal Guidelines

Create a one-page proposal describing the question your team is asking or the problem your team is addressing. The proposal should follow the proposal template on page 5 of this guide and should describe:

- A problem or question your team's project will address,
- A statement of why the problem or question is interesting,
- The team's hypothesis,
- What dataset(s) you will use to address your question (provide links),
- How do you plan to go about analyzing the data to address the problem? What types of data visualizations do you expect to need?

Data: It is the responsibility of each team to find and download the dataset(s) needed for their project. The goal is not to have a really big dataset to work with (terabytes of data are not needed), but rather to have enough data to gain insights into the problem. Some good websites to use to find data are listed on our *DataSets page*. Finding datasets can be difficult; DataJam mentors are available to help.

Visualization Tools: Simple visualizations can be made with programs such as Google Sheets and Excel. More complex visualizations can be made with Tableau Public, which is a freely available program. See videos for how to make visualizations on our *Resources page*.

Statistical Analysis: Each team will decide the appropriate statistical method to apply to their problem. We have lots of resources available to help. See the *videos at the top of the DataJam page, the DataJam manual and many videos and tip sheets on the Resources page*. DataJam mentors are available to help think about what analyses to use and to teach teams how do various analyses.

Judging Criteria for the Final Project: Participants are encouraged, but not required, to consider tackling a problem in their community. The following criteria will be used by judges in evaluating the final poster and presentation that each team prepares:

- Understanding the Problem: Each team should clearly present and show how they are solving a specific civic issue or challenge with their project.
- Feasibility: The project selected should demonstrate the clear possibility of solving the problem with data that is or could be available.
- Critical Thinking: Teams should demonstrate the ability to ask insightful questions and relate different concepts to solve a problem. The approach the team takes is as important as being able to conduct a particular analysis.
- Artistic Merit: Final presentations should have a consistent art design and be aesthetically pleasing.
- Presentation Skills: Focus on making an interesting, concise and engaging presentation.
- Commitment & Engagement: Each member of the team should have played a role in the project.



DataJam Project Proposal Template

School Name (& Team Number if there is more than one team/school):

Project Title:

Team Member Names & Emails:

Problem:

IN THE FORM OF A QUESTION STATE WHAT PROBLEM YOUR DATAJAM PROJECT WILL ADDRESS?

Why is it Important?:

IN A PARAGRAPH EXPLAIN WHY THIS IS AN IMPORTANT QUESTION AND WHO WOULD BE INTERESTED IN THE ANSWER.

Hypothesis:

A STATEMENT SHOULD BE MADE ABOUT WHAT THE TEAM THINKS THE ANSWER TO THEIR QUESTION WILL BE.

Data Required:

PROVIDE A LIST OF DATASETS YOU HAVE FOUND THAT YOU WILL ANALYZE TO ANSWER YOUR QUESTION. PLEASE INCLUDE THE FULL LINKS SO WE CAN LOOK AT THE DATASETS WHILE WE ARE LOOKING AT YOUR PROPOSAL.

Analysis Plan:

IN ONE OR TWO PARAGRAPHS DESCRIBE HOW YOU PLAN TO ANALYZE YOUR DATA, WHAT ANALYTICAL STRATEGIES YOU WILL USE AND WHAT DATA VISUALIZATIONS YOU THINK YOU WILL USE? (No work needs to have been done at the time the proposal is submitted)



DataJam 2024

Parent Permission Slip, Waiver of Liability, Assumption of Risks, Indemnity Agreement

The DataJam (thedatajam.org) is the educational nonprofit that runs the DataJam each year to introduce high school students to learning about data science, how critical it is for understanding everything in our current world, and how to work with big data sets from a data visualization and statistics perspective.

We encourage all high schools to put together a DataJam team (2 or more students constitute a team; generally teams have 3-8 students). Each high school team is asked to have a teacher at their high school to serve as a DataJam Advisor. Teams are mentored by university students who are trained as DataJam Mentors in how to teach data science, best practices in mentoring, working with diverse communities, and how to give scientific presentations. Mentors are advised to use Zoom and Edtera for mentoring and NOT share personal contact information, “friend” youth on social media, or have direct contact with youth.

The DataJam will make available an online Edtera workspace so all teams can easily get assistance from DataJam mentors for helping them think about the research question they want to study, how to find data sets, how to analyze data, how to make graphs and how to put together a presentation. They will also be able to ask to have a zoom meeting with a mentor so they can get assistance at any time that works for them. All interactions between mentors and students will be online and by videoconference. Mentors will be advised to NOT hold Zoom conferences with single students.

The timeline for this year is (1) Fall 2023 through December 8 2023 – teams sign up to participate in DataJam 2024, (2) December 8, 2023 – team proposals should be submitted, (3) Jan-Mar 2024 – teams work on their projects with help of mentors, (4) March 29, 2024 - posters due (5) April 25, 2024 – DataJam finale on zoom – teams give presentations about their projects and many awards are given.

Waiver: In turn for being permitted to participate in DataJam 2024 I do hereby release, waive, discharge, and promise not to sue The DataJam or any of the colleges or universities training DataJam mentors from any and all claims, including negligence and sexual misconduct in connection with my child's participation in DataJam 2024.

Assumption of Risks: Participation in DataJam 2024 carries with it certain inherent risks that can not be eliminated regardless of the care taken to avoid risks. Risks vary from hurt feelings to exposure to inappropriate language or behaviors.

Indemnification and Hold Harmless: I also agree to indemnify and hold The DataJam and all universities training DataJam mentors harmless from any and all claims, actions, suits, procedures, costs, expenses, damages and liabilities, including attorney’s fees, arising out of my child’s involvement in DataJam 2024.

Severability: I further agree that this Waiver of Liability, Assumption of Risk, and Indemnity Agreement is intended to be as broad and inclusive as permitted by law, and that if any portion is held invalid the remaining portions will continue to have full legal force and effect.

I am the parent of _____, who is a current high school student at _____ High School. I give my permission for my child to participate in online videoconferences and use the Edtera website to participate in the DataJam 2024. I also give permission for The DataJam to post pictures from videoconferences of my child participating in DataJam 2024 activities.

Acknowledgment of Understanding: I have read this Waiver of Liability, Assumption of Risk, and Indemnity Agreement, fully understand its terms, and **understand that I am giving up substantial rights, including my right to sue.** I confirm that I am signing the agreement freely and voluntarily, and **intend my signature to be a complete and unconditional release of all liability** to the greatest extent allowed by law.

Parent Signature

Date

Printed Parent Name

Student contact email address (a publicly available email address, like a gmail address works best)

[Scan and EMAIL TO datajam@thedatajam.org](mailto:datajam@thedatajam.org)

Poster and Final Presentation Guidelines

Data analytics consists of two broad tasks: (a) extracting meaningful information from raw data and (b) explaining the impact and value of that information. Your poster and final presentation for the DataJam should show how your DataJam project accomplished these two tasks.

Poster Guidelines:

You should follow the guidelines below for the development of your poster. However, we encourage you to exercise your creativity in designing your poster and how you would like to communicate the details of your project.

These three guidelines are essential:

- Poster size: 24" x 36". The poster orientation can be vertical or horizontal.
- Your poster should clearly display: (1) Title of your project, (2) The full names of the all team members, (3) the name of your school.
- Submit your poster as a .pdf file and send to datajam@thedatajam.org by March 29, 2024.

In your poster you will want to have information addressing the questions below. Remember, pictures, graphs, tables and diagrams are very helpful ways of presenting your results.

- What is your research question?
- What was your hypothesis when you started the project?
- What dataset(s) did you use?
- Briefly describe the analyses you did to answer your question. Show graphs and diagrams if relevant.
- What challenges did you face in finding or using the data?
- Summarize your results.
- Can you make a recommendation based on your results?

Except for the essential guidelines, which each team must follow, you can tailor your poster according to your project. Remember judges will view your poster without you being there to explain it.

Presentation Guidelines:

The week of April 15, 2024, your team will be scheduled to give a 10-minute oral presentation, using slides to a panel of three judges. This presentation will be scheduled at a time that is mutually possible for both your team and the judges. The presentation will be conducted on Zoom.

The oral presentation should address the same questions listed above for the poster. Remember to not have too many slides; for a 10-minute talk, 10-13 slides is appropriate.

It is best if each member of your team can have some speaking role in the presentation and question answering. A 10-minute period will be allotted for questions.

Poster						
Scoring Category	5=Highly Competent	4	3	2	1=Not Competent	Weight
Uncover and use a variety of data sources.	<i>Sources are varied, high quality, pertinent, sufficient.</i>		<i>Sources are pertinent and sufficient, but more were required of a higher quality.</i>		<i>Too few sources are used, some or all are not pertinent.</i>	12%
Form a research question based upon the relevant data.	<i>Research question is specific, clear, original, and thoughtful. Data are relevant, important, and balanced.</i>		<i>Research question is clear, but lacks originality. Data sources are relevant, but lack depth.</i>		<i>Research question is posed with so little clarity as to be confusing, or it is absent. Data lack relevance, quality, depth and/or balance.</i>	16%
Analyze data/information.	<i>Data analysis is clear, thorough, and appropriate. Findings are presented clearly.</i>		<i>Data analysis is clear, but needs further elaboration. Findings are relevant, if simplistic.</i>		<i>Data analysis is missing, unclear, simplistic, or biased. Analysis is unrepresentatively selective or unrelated to the research question.</i>	14%
Data Visualization	<i>Research question is further explored through the use of relevant visualizations. The group shows the growth of their understanding in the role of visualization in exploring data questions.</i>		<i>Visualization is somewhat relevant. The group can articulate why it was chosen.</i>		<i>No visualization was used or the group cannot explain why the particular visualization was chosen.</i>	9%

Draw logical and defensible conclusions.	<i>Conclusions are present, logical, related to the research question, supported by evidence</i>	<i>Conclusions relate to the research question, but are only weakly supported by the evidence.</i>	<i>Conclusions are missing, unclear, illogical, irrelevant to the research question, or unsupported by data.</i>	16%
Communicate clearly and effectively findings and conclusions.	<i>Writing is free of grammatical, syntax and typographical errors, is well organized, and enhances the reader's ability to understand the findings and conclusions. Use of relevant visualization is included.</i>	<i>Writing is mostly free of grammatical, syntax and typographical errors, and is organized well enough for the readers to understand the findings and conclusions. Use of simple visualization is included.</i>	<i>Writing contains significant grammatical, syntax, and typographical errors and is poorly organized. Errors significantly impair the reader's understanding of the findings and conclusions. Visualizations are poorly used or not used at all.</i>	8%
				75%

Presentation

Scoring Category	5=Highly Competent	4	3	2	1=Not Competent	Weight
<p>Effective Communication</p> <p><i>We will not be penalizing people for nervousness or speaking skills. Focus on the content.</i></p>	<p>Presentation clearly articulates the work done and the conclusions reached.</p> <p>Presentation is delivered effectively with out a single person dominating.</p> <p>All team members contributed in a meaningful way.</p>		<p>Presentation articulates the work and conclusions, but there are gaps in logic or how they reached their conclusion.</p> <p>Presentation is delivered effectively with all people contributing, but one person being dominant.</p>		<p>Presentation DOES NOT articulates the work done or present clear conclusions .</p> <p>Presentation is delivered By one or two people with one person dominating time.</p>	10.0%
<p>Slides</p>	<p>Presentation slides are presented in an orderly and engaging format.</p> <p>Includes sections for Problem Statement, approach, analysis, conclusions and recommendations</p>		<p>Presentation is well organized, but not all slides are effective in communicating the project.</p> <p>Some sections are missing from presentation</p>		<p>Presentation is dis-organized, lacks several sections, lack graphical appeal.</p>	8.0%

<p>Answering Questions</p>	<p>Answers always address question asked.</p> <p>Answers were thoughtful and thorough.</p> <p>Multiple people answer questions or it is clear the team as a whole could answer the question.</p>	<p>Answers generally address the question asked.</p> <p>Some responses are guesses or they are winging it.</p> <p>It is not clear that everyone could answer the question.</p>	<p>Answers do not address the question.</p> <p>It is clear team does not fully understand the work they did from their responses.</p>	<p>7.0%</p>
				<p>25.0%</p>