Electron Collision Visualization

Brian Gilmore, Rebecca Moore and Chiamaka Aghaizu

Main Project Link: https://github.com/rmoore2738/Electron_Collision_Visualization

Brian's Github Link:https://github.com/B-swagG
Rebecca's Github Link: https://github.com/rmoore2738
Chiamaka's Github Link: https://github.com/optimak

Introduction

Electron collisions are a fundamental building block of our world. The ability to combine and break apart molecules allows for the creation of almost everything. Since human eyes aren't able to see at such a small scale, visualizations of the collisions have to be generated. By visualizing these events, we are able to better understand what is physically happening and the effects different parameters have on the collisions. For this project, we will focus on highlighting the process of the collisions visually and introducing the idea of electron collisions in a digestible way.

Domain

For this dashboard, we wanted to create a domain targeted at scientists and novel physics enthusiasts with a basic understanding of statistics, closer to the general public. We found that there exists different visualizations and data for both domains, but not something that can be interesting and helpful for both. To do this, we used detailed scientific data that includes many of the features involved in electron collisions and can provide the details we need for the scientists. For the general public, we are focusing on high-level, interactive visualizations that ensure they get the main idea from the data, and travel deeper to understand further, if they desire. Since we are combining these two domains, we will focus on one question from each. For the scientists, their main question seems to be the transfer of energy that occurs during collisions. For the public, their main question seems to ask what an electron collision is.

To address these two questions, we created layers to our dashboard. The top layer addresses the public and their question about what an electron collision is. The middle layer provides more visualizations in higher dimensions that begin the transition deeper into the data. Finally, the bottom's dynamic visualization is focused on the scientists and provides a concise way for them to analyze the energy of each electron during collisions, the overall transfer, or whatever their expertise tells them to look at by selecting the attributes. By breaking the dashboard down like this, we are able to create a custom domain and address the main questions from each of our target audiences.

Workflow

In order to effectively complete the project, we all worked concurrently on specific aspects. By doing these, we were able to manage our time more effectively and create a more well rounded project. Once we were each happy with our contribution, we came together as a team and combined each of the parts. Below is a detailed timeline and description of our workflow throughout this project.

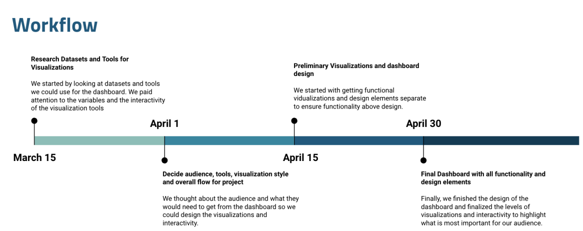


Figure 1: Workflow

Significance

The field of electron collision physics is currently lacking in visualizations. There are several robust datasets, but most do not have sufficient dashboards accompanying them. A new focus in the field has been on the visualization of the data, specifically in the subatomic fields [5]. Some of the current visualization works include electron wavefunctions [6], electron-neutral collisions [7] and Proton visualizations [8], but very little on dielectron visualizations, which is what we will focus on. These current visualizations also focus on line graphs and do not have any interactive elements, as they are meant for published papers, not the public. Our dashboard will allow physicists and the general public to understand and enjoy this niche field of electron collisions, unlike some of the confusing visualizations currently available. By presenting the information in an interactive and still meaningful way, we can engage with more people than the high-level academic papers have been able to and spark interests into the field that many may have never been introduced to.

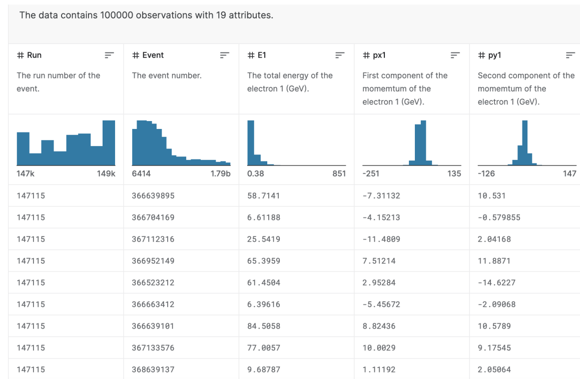
Background

A lot of work in electron collision visualizations centers around simplified 3D renderings of electrons, such as NASA's Scientific Visualization Studio, which shows a very simple representation of two electrons colliding. This work is great for reaching a wider audience due to the simple nature, but lacks the details that make it meaningful for scientists and physicists. Some work has been done on representing electron collisions through computer generated art, which shows the beautiful and intuitive way electrons collide, but again lacks the detailed breakdown of what happens during the collision. Through researching the way electron collisions have been visualized and presented thus far, we found that there is a need to combine the visually appealing images that come from the sources above and the tables/charts from scientific papers that have the detailed breakdown of the collisions.

Data Abstraction

Dataset Attributes and types

Our dielectron data is a tabular dataset composed completely of numerical data. This includes four categorical attributes: Run, Event, and Charge (electron 1 and 2) [2]. The dataset also contains 15 quantitative attributes: x, y, and z components of momentum (for each electron), transverse momentum (for each electron), pseudorapidity (for each electron), phi angle (for each electron), and the mass of both electrons [2]. Each item in the dataset represents a single collision of two electrons. Below is a summary of the kaggle dataset we used:



The data contains 100000 observations with 19 attributes.

# Run	# Event	# E1	# px1	# py1
The run number of the event.	The event number.	The total energy of the electron 1 (GeV).	First component of the momemtum of the electron 1 (GeV).	Second component of the momemtum of the electron 1 (GeV).
147k – 149k	6414 – 1.79b	0.38 – 851	-251 – 135	-126 – 147
147115	366639895	58.7141	-7.31132	10.531
147115	366704169	6.61188	-4.15213	-0.579855
147115	367112316	25.5419	-11.4809	2.04168
147115	366952149	65.3959	7.51214	11.8871
147115	366523212	61.4504	2.95284	-14.6227
147115	366663412	6.39616	-5.45672	-2.09068
147115	366639101	84.5058	8.82436	10.5789
147115	367133576	77.0057	10.0029	9.17545
147115	368639137	9.68787	1.11192	2.05064

Figure 2: Data summary

Design of Features

Our dashboard is a multi-level, dynamic interface that displays high level analysis of the data, such as a breakdown and comparison of the attributes of electron 1 and 2 during the collision. It allows visitors to dynamically control a graph to visualize their choice of attributes. To do this, we have created an interactive web page using a Python backend that will export user selections to HTML and allows users to select attributes from the dataset and view the graph of the selected data. Our dashboard will highlight the high level aspects of the data, such as the average energy of the electrons, offer mid level comparisons between the two electrons in the collision, and finally have a fully interactive interface that allows for custom visualizations and lets users dig deeper into their chosen attributes. The dashboard highlights the flow of information from high level and in lower dimension, to more detail as the user scrolls down the page, ending with the interactive visualization at the bottom of the page. Below is a diagram that describes the hierarchical design of features we developed:

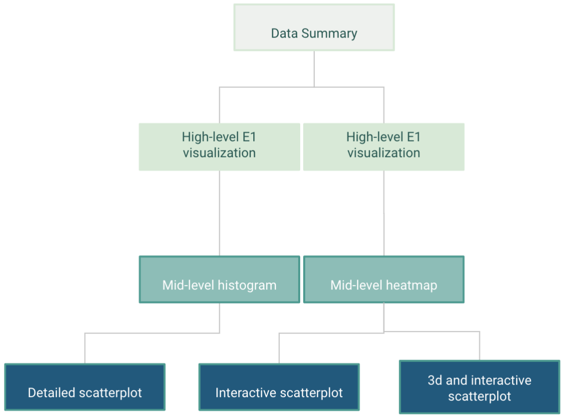


Figure 3: Feature Diagram

As this diagram shows, the deeper a user scrolls into the dashboard, the more details they are presented. At each level they can also click to view more or adjust for their preferences/needs. In order to avoid overwhelming the user, we made more interactive idioms that have simple starting information that is easy to digest.

Data Transformation

The dataset came as a CSV format, which we were able to read directly into our dash app. Once we had the csv loaded, we transformed it to a pandas dataframe for the creation of the visualizations. We checked to ensure there were no missing values for the features. For the table summary at the top of the dashboard, we needed to convert the dataframe into a dictionary. We did this by using the python "OrderedDict" library. This had to be done in order to correctly display the columns and values using the dash_table library.

Task Abstraction

Our dashboard allows users to analyze increasingly complex targets as they progress from top to bottom. We begin by targeting single attributes by visualizing distributions using histograms and their linear correlation using a heatmap. We then target multi-attribute distributions using pie charts. Finally, further correlation, trends, features, and outliers are targeted using interactive scatterplots.

The main action for both target groups to analyze; the visualizations are intended to support consumption and understanding of data. We expect the users with less domain expertise to discover and enjoy the data through the visualizations. Another expected action of our audience is to query the data; they can compare attributes of the electron collision and their relationships with each other.

The tasks we allow on our dashboard each have a specific purpose for the user. We first start by presenting simple summaries of the data and two high level visualizations.At this point, users are able to adjust what values and features they look at, but not deeply visualize. Once they understand the data, they can derive personal visualizations using the containers we set up for histograms and heatmaps. This allows them to explore deeper and become even more familiar with certain attributes. Once they reach the bottom, they can begin comparing attributes. We wanted to make sure people got a chance to understand and appreciate the data before they jump into comparisons and conclusions. At this point, the user can travel as deep as they wish into the dataset and create many custom plots. Below is a diagram showing the flow of our actions:

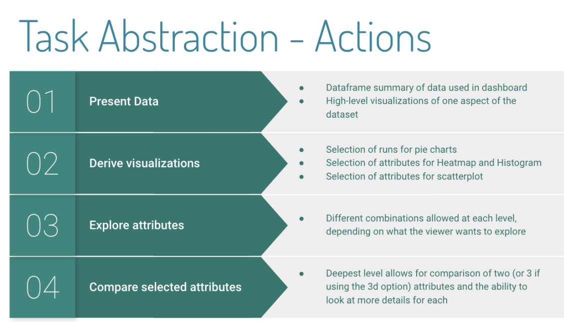


Figure 4: Task abstraction Diagram

Implementation Using Tools

For this dashboard, we used several different tools. The languages we used include Python and HTML. From python, we used Dash, Plotly and Pandas. These libraries allow us to use python as the backend for our webpage and to run the entire program in a python script, rather than using another tool as an API. Dash is specifically designed to create data visualization dashboards by being able to host the backend, API, and front end all in one script. To make the plots, we used python Plotly. Plotly is used by the graph component of Dash to create each of the visualizations seen on the dashboard. Dash also creates the HTML components needed to work with the HTML code and utilizes callback functions to make interactive visualizations.

Final Results for Analysis

Each aspect of our dashboard is shown below. We carefully chose the level and idiom type to use to create a cohesive dive into the dataset. By doing these, we are able to reach a further audience and have more meaningful conclusions drawn from our visualizations.

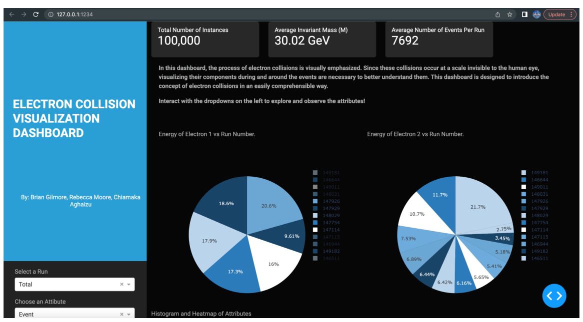


Figure 5: data summary & Pie charts

Figure 6: Histogram

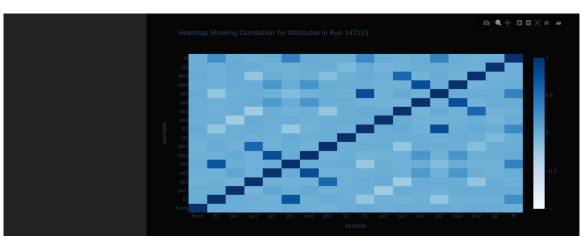


Figure 7: Heatmap

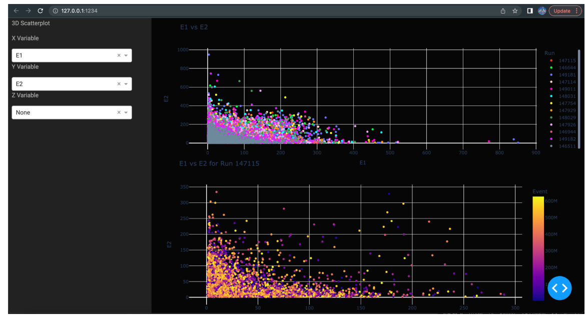


Figure 8: 2d scatterplots

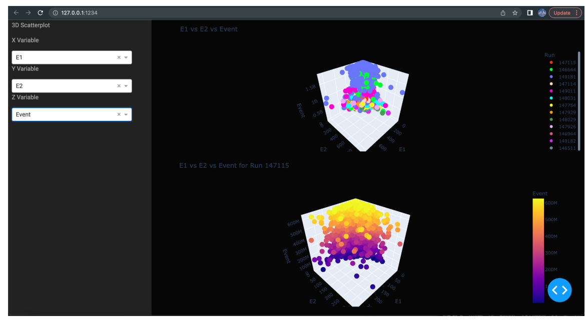


Figure 9: 3d scatterplots

Storytelling

The necessity and impact of this project and its application can be fully understood by looking into its target audience and domain.

Life

• ***Who*** *are the people or communities in need of help?*

Scientists and Physics aficionados are the communities in need of help.

• ***What*** *problem happened to them?*

They have no way of visualizing and understanding the properties of an electron during electron collision, a popular event studied in atomic and molecular physics.

• ***When*** *did the problem take place?*

It can be estimated that this problem began when electron collision was discovered in 1914 [10].

• ***Where*** *1) The environment and settings that the people or the community is living in, and 2) the place/location where the problem takes place.*

1) This problem is most likely to take place in a learning environment.
2) These places include college and online or in-person science forums.

• ***Why***? *-the possible causes and/or origin of the problem?*

The problem exists because the actual animation of electrons colliding has been the primary method of visualizing this occurrence.

Data

• ***Who*** *is the data set about?*

The data set is about electrons, and was obtained as a subset of the total experiment run by Thomas McCauley in 2010 [3].

Who were sampled in this data set?

A subset of elections undergoing collisions with one another was sampled.These data were selected for use in education and outreach and contain a subset of the total event information, where the two electrons being observed have invariant mass between 2-110 GeV .

Who were over sampled or under sampled? Are they representative of the main characters?

The data doesn't represent electron collisions with either or both electrons having invariant mass outside the 2-110 GeV range.

Is there any identifiable information or is there any risk of disclose identifiable information?

Since this is electron data, there was no need to anonymize the data set.

• ***What*** *properties are recorded by the data set?*

The actual properties recorded by the data set are The Run number of the experiment, the event number within that run, the total energy of the electron, the components of the electron's momentum, the transverse momentum, the pseudorapidity, phi angle and charge of each electron and the invariance mass of the two electrons [3].

Does the data set record the targeted events, activities, behaviors, etc.?

The targeted event, the behavior of electron's attributes, are recorded in the dataset.

• ***When*** *did the observation take place? When were the data collected?*

These properties were observed in 2010, and the data was collected at that time[3].

Is it longitudinal or cross-sectional?

The data is cross-sectional in nature.

Are they real time data? How old or fresh are the data?

It is also collected in real-time.

To what extent generalization can be made across time?

Generalizations can be made with similar particles, subjected to the same conditions and having the same properties.

• ***Where*** *did the event, activity, behavior, and observation, etc. take place? Where was the data collected, if the information is available?*

The event took place inside the Large Hadron Collider located in Switzerland and was collected there .

What does the geographical coverage of the data set look like? Does the data set contain geographical information (GIS)? Is this a local, regional, national, or global data set? To what extent generalization can be made across settings

There is no geographical coverage related to the measures collected in this data, so no generalization based on geography can be made.

• ***Why*** *did the experimentation/observation etc. take place?*

The experiment where the properties were recorded took place as part of the CMS Experiment to ultimately investigate completely new and unpredicted physics phenomena [9].

Why was the data collected?

The reason for collecting this particular subset of data was because the two electrons being observed have invariant mass between 2-110 GeV for outreach and education [3].

Users

• ***Who***: *the main character is the targeted user or audience?*

The main target users of this application are also the community in need of help in this story; scientists and Physics enthusiasts. \

• ***What*** *can the application do? What does the visualization show?*

The application allows the target users to observe how the properties of the electrons undergoing electron collision change in relation to each other, in 1-3 dimensions.

• ***When*** *can the user use the application/visualization?*

The user can utilize the application when they want to completely understand the attributes of electron collision, and perhaps confirm already determined physics principles.

• ***Where*** *will the visualization and applications be deployed?*

This application is deployed on the web and allows full utilization on a web-page.

• ***Why*** *is the visualization or application useful to the user?*

The application can be used to choose aspects of the experiments that can be further explored by interested physicists, particularly when in a research setting.

Project Management

Implementation Status Report

To finalize the project, we took all of our functional visualizations from before and combined them in an appealing way. We also added some more layers to the visualizations, such as a high level data summary and a deeper layer to the scatterplots for more in-depth analysis.

Works Completed

Description

We have completed the main visualizations that include dynamic and interactive elements and the outline for the final dashboard design.

Responsibility

- Rebecca
  - dashboard design & integration
  - Electron 1 & 2 pie charts
  - Supplemental second level visualizations
- Brian
  - Interactive Scatterplot
  - Sub-scatterplot for deeper visualization
  - 3d scatterplot
- Chiamaka
  - Dynamic histogram and summary cards
  - Dynamic heatmap
  - Live site and Github repo updates

Contribution

- Rebecca - 34%
- Brian - 33%
- Chiamaka - 33%

References

1. D. Gupta, Applied Science and Convergence Technology, 30-Nov-2020. [Online]. Available: https://www.e-asct.org/journal/view.html?doi=10.5757%2FASCT.2020.29.6.125#:~:text=Electron%20collision%20data%20are%20important,the%20realization%20of%20plasma%20reactors [Accessed: 20-Mar-2023].
2. McCauley, Thomas; (2014). https://opendata.cern.ch/record/304. Available: https://www.kaggle.com/datasets/fedesoriano/cern-electron-collision-data
3. McCauley, Thomas; (2014). Events with two electrons from 2010. CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.PCSW.AH

VG Available:
https://opendata.cern.ch/record/304

4. J. Nam, H. Yong, J. Hwang, and J. Choi, “Training an artificial neural network for recognizing electron collision patterns,” Physics Letters A, vol. 387, p. 127005, 2021.
5. *Direct Observation of Spin-Polarized Surface States in the Parent ...* https://iopscience.iop.org/article/10.1088/1367-2630/12/12/125001
6. X, Science. “Visualizing a Complex Electron Wavefunction Using High-Resolution Attosecond Technology.” *Phys.org*, Phys.org, 11 Jan. 2023, https://phys.org/news/2023-01-visualizing-complex-electron-wavefunction-high-resolution.html
7. Brask, S. M., et al. ``Electron–Neutral Collisions Effects on Langmuir Probe in the Lower e-Region Ionosphere.” *AIP Publishing*, AIP Publishing LLCAIP Publishing, 1 Jan. 1970, https://aip.scitation.org/doi/10.1063/5.0079761
8. Costello, Sarah. “‘Visualizing the Proton’ – Physicists' Innovative Animation Depicts the Subatomic World in a New Way.” *SciTechDaily*, 11 May 2022, https://scitechdaily.com/visualizing-the-proton-physicists-innovative-animation-depicts-the-subatomic-world-in-a-new-way/
9. “CERN accelerating science,” What is CMS? | CMS Experiment. [Online]. Available: https://cmsexperiment.web.cern.ch/news/what-cms. [Accessed: 30-Apr-2023].
10. R. Williams, “This Month in physics history,” *American Physical Society*. [Online]. Available: https://www.aps.org/publications/apsnews/201704/physicshistory.cfm#:~:text=April%201914%3A%20James%20Franck%20Studies%20Atoms%20with%20Electron%20Collisions . [Accessed: 30-Apr-2023].