

Knowledge Discovery for Biomechanical Data

Sift is designed to be a tool that helps you, the user, discover useful knowledge from your data. This process of knowledge discovery is iterative, requiring users to collect, clean, and shape their data before performing analysis and then communicating their results. Each of these steps requires experience to be done well, the aim of this article is to outline the goal of each step, how Sift lets you accomplish these goals, and to point you on to additional resources.

Gather Data

The first step in learning from your data is bring it all together for analysis. Challenges can arise here if the data you are interested in has been collected by different researchers in different locations over many years. These challenges can be legal in nature - whether adequate permission was given by participants during data collection - or technical - how to transfer that many 1s and 0s. No matter how complex your study is, questions about how data is collected, how meta-data and data are linked, how data is shared or centralized, and how data is protected should always be answered before you start collecting.

So you have established a data collection plan, worked hard to collect you data, and now you have an initial or complete data set to analyze. At this point you want to bring your data set under one roof and explore what you've collected to see what patterns and connections you can find. Sift lets you do this by loading [CMZ files](#) which contain all of the [.c3d files](#) as a library. Since the .c3d format is standard across biomechanics research, this lets you easily group your data according to collection session and trials and to import this data with only a few clicks. You can also include arbitrary metadata alongside your .c3d files by using the [Build CMZs](#) feature.

Read about the [Directory Structure for Input Data](#) to learn more about how to organize your data most efficiently.

Complete the [tutorial](#) for loading and viewing your data to see how you can get your data into Sift and ready for analysis.

Cleaning Data

The second step in learning from your data is to clean your data. If we're honest with ourselves, no dataset from the real-world is perfect. Sometimes sensors fail, gait events are incorrectly identified, or something else just goes wrong. That's why it's important to clean your data and confirm that every piece of data that you put into analysis is a piece of data that you trust. This is also a chance for you to make sure that you have all of the data you expect. If you're missing something, then it's time to go back and make sure it gets collected.

Sift lets you visualize your data as individual traces, workspace means, or group means so that you can assess it in whichever way makes sense for you. You can click on a specific trace in the [explore page](#) in order to determine exactly which file and frames it comes from and you can animate specific traces to help your quality control process. It is also possible to automate some of the data cleaning process by using statistical methods such as [dynamic time warping](#).

Although we are ideally able to rectify any data quality issues, at the end of the day, you can choose to exclude specific traces from your dataset so that they are not included in your analysis. Consistent with our philosophy, exclusion does not imply deletion - it simply means that the trace or metric is flagged to be excluded. It's also important to note that excluding data is not the same as “cherry-picking” since we only exclude data from our analysis when we believe that it does represent a movement that we recorded from the real-world.

Complete the [tutorial](#) for cleaning your data to learn more about these processes.

Shaping Data

Now that you have a clean dataset in front of you, it's time to start analyzing the data! But wait, because a single study can contain multiple questions and each question might be concerned with a different portion of your dataset. Before we can jump into analysis, we have to shape our data to make sure that we are getting the right “view” into our data to answer the question we have in mind.

Sift lets you shape your data by [querying](#) the library. A query defines which signals and traces you want to extract from the library and executing a query produces a group in the queried group widget. Sift can automatically define queries for you or you can save these as [.q3d files](#) for easy reuse and to help you track exactly how you performed your analysis. If you decide that you can't quite get the data you want from your queries or for some reason you can't work with it, try going back to the collection and cleaning stages of the process. Read about Sift's [Query Builder](#) dialog to learn more about how you can define and refine groups of signals and read about how to query [multi-subject](#) data.

Beyond deciding which traces should be assigned to which group, we can also shape our data by deciding if and how these traces should be transformed. Sift's default behaviour is to time-normalize queried traces to 101 points, but users can also register their traces to intermediate events and points of interest using Sift's [curve registration](#) feature.

Analysing Data

Having queried and shaped your clean dataset to get exactly the traces and metrics that you wanted, now you're finally ready to start analyzing. We learn a lot about analytical techniques in our courses and throughout our formal training, but it's only one of the five steps here. Even though this is what we often think of as the difficult work of research, you've already put in a lot of effort to get through the first three steps and get to this point!

The type of analysis you perform is obviously going to depend on the question you're trying to answer and the dataset that you have. Sift implements a range of common data analysis techniques such as summary statistics calculation, [Principal Component Analysis \(PCA\)](#), [Statistical Parametric Mapping](#), [Dynamic Time Warping \(DTW\)](#), [Global Gait Asymmetry \(GGA\)](#), [Gait Profile Score \(GPS\)](#), and clustering algorithms, with new techniques being added. Sometimes your analysis will prompt new questions or new ways of looking at your data. Don't be afraid to go back to collecting or shaping your data to see what else you might find.

Complete the [tutorial](#) for performing PCA and the [tutorial](#) for performing SPM to see some of the different ways you can analyze your data.

Communicating Results

Once your analysis has produced results, your last step is to communicate your findings to the wider world. Whether you're presenting to a collaborator, talking at a conference, or producing publication-ready figures, there is a lot to think about when communicating your results.

Sift's different visualization tools all let you control over the colours, line styles, and axis labels used to allow you to produce the figures you want. If more work is required, then you can export your analysis results to a number of different text formats including Visual3D ASCII, P2D, and SPSS.

Complete the [tutorial](#) on customizing and shaping your data visualizations to best communicate your results.

Complete the [tutorial](#) on exporting results to see the variety of options available.

Read about [The misuse of colour in science communication](#) to learn how properly chosen colour palettes help you report data variation, reduce complexity, and increase accessibility.

Visit [Color Brewer](#) for advice on choosing accessible colour palettes.

Visit [The Data Visualization Catalogue](#) to learn about the different ways you can visualize data, along with the pros and cons of each.

From:
<https://has-motion.com/wiki/> - Wiki Documentation Site

Permanent link:
https://has-motion.com/wiki/doku.php?id=sift:documentation:knowledge_discovery_for_biomechanical_data&rev=1724870426

Last update: 2024/08/28 18:40

