

Literature Review

Joe Zemmels

A scientific method must be effective, accessible, and understandable for it to be accepted by the wider scientific community. As the need for computationally-intensive methods to analyze forensic data has grown, so too has the need for user-friendly tools to experiment with and improve upon these methods.

In this work, we discuss developments to a particular class of algorithms used to compare cartridge case evidence known as the Congruent Matching Cells (CMC) method. Chapter [#] discusses a modularization of the algorithm into a “pipeline” that enables reproducibility, experimentation, and comprehension. Chapter [#] introduces novel pieces of this pipeline that we demonstrate provide improvements to the current state-of-the-art. Chapter [#] details a suite of diagnostic tools that illuminate the inner-workings of the algorithm and help determine when and why the algorithm does or does not “work” correctly.

Forensic Examinations

A primary goal of a forensic examination is to determine the source of a piece of evidence. This is commonly referred to as the *source identification* problem [Ommen and Saunders, 2018]. A common setting for source identification problems involves obtaining evidence of unknown source from a crime scene and either evidence from a known source or other evidence of unknown source. For example, a bullet found at a crime scene may be compared to a suspect’s firearm. Such evidence is sent to a forensic lab where a trained Firearms and Toolmarks examiner compares the “questioned” bullet to bullets fired from the suspect’s firearm to determine whether the suspect’s firearm was the original source. In this work, we develop a method to supplement such an examination by providing an objective measure of similarity between the two pieces of evidence.

Firearms and Toolmarks Examination

Firearms and toolmarks (F&T) examination involves studying markings or impressions left by a firearm or other tool (e.g., a screwdriver) on a surface [Thompson, 2017]. The focus of this work is on a subset of F&T problems that deal with the comparison of cartridge case evidence. A *cartridge case* is the portion of firearm ammunition that encases a projectile (e.g., bullet, shots, or slug) along with the explosive used to propel the projectile through the firearm. When a firearm is discharged, the projectile is propelled down the barrel of the firearm, while the cartridge case is forced towards the back of the barrel. It strikes the back wall, known as the *breech face*, of the barrel with considerable force, thereby imprinting

any markings on the breech face onto the cartridge case, creating the so-called *breech face impressions*. These markings have been suggested to be unique to a firearm and are used in forensic examinations to determine whether two cartridge cases have been fired by the same firearm.

During a forensic examination, two pieces of ballistic evidence are placed under a comparison microscope. Comparison microscopes allow for a side-by-side comparison of two objects within the same viewfinder, as seen in [Figure 1](#). A pair of breech face images is aligned along the thin black line in the middle of the images. The degree to which these breech face markings can be aligned is used to determine whether the two cartridge cases came from the same source; i.e., were fired from the same firearm. These breech face impressions are considered to be a firearm’s unique “fingerprint” left on a cartridge case [[Thompson, 2017](#)].

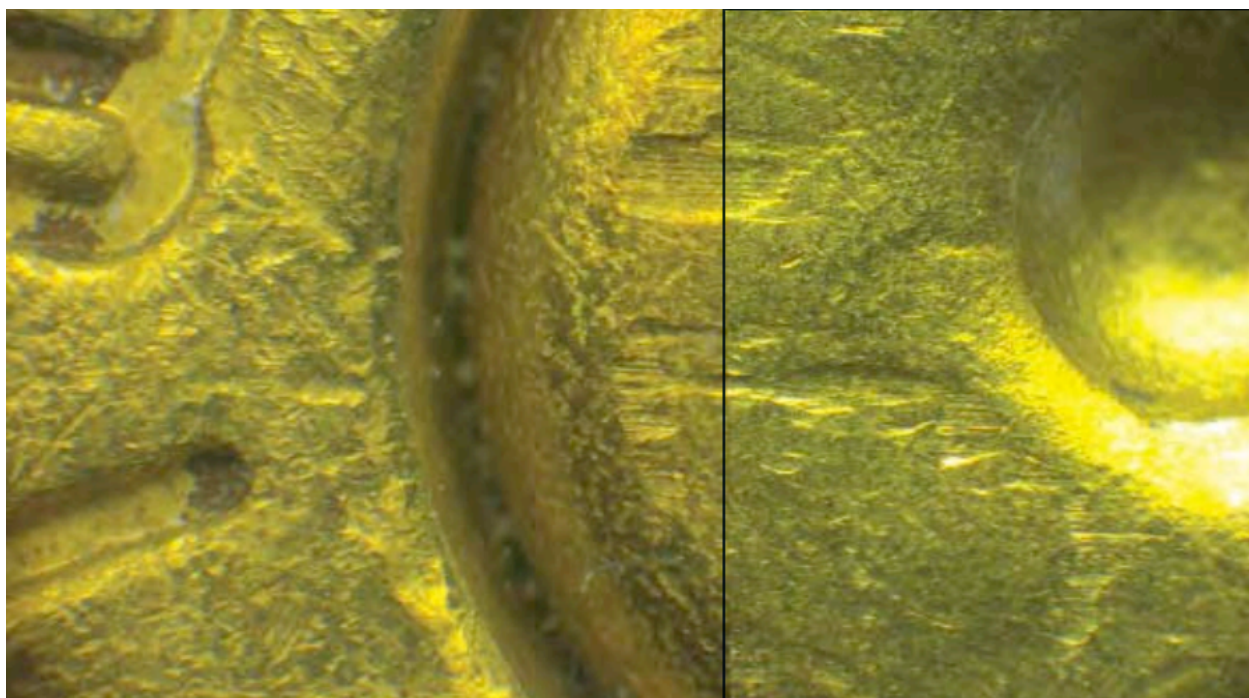


Figure 1: A cartridge case pair with visible breech face impressions under a microscope. A thin line can be seen separating the two views. The degree to which the markings coincide is used to conclude whether the pair comes from the same source.

A F&T examination typically ends in one of three conclusions: identification, meaning the evidence originated from the same source, exclusion, meaning the evidence did not originate from the same source, or inconclusive, meaning there is insufficient information to conclude identification or exclusion [[AFTE Criteria for Identification Committee, 1992](#)]. Examiners rarely need to provide quantitative justification for their conclusion. Even for qualitative justifications, it can be difficult to determine what the examiner is actually “looking at” to arrive to their conclusion [[Ulery et al., 2014](#)]. For cartridge case comparisons, similarity between the breech face impressions of two cartridge cases, among other factors, is used by the examiner to make their determination.

Due to the opacity in the decision-making process, examiners have been referred to as “black boxes” in a similar sense to black box algorithms [OSAC Human Factors Committee, 2020]. Their evidentiary conclusions are fundamentally subjective, and there is empirical evidence to suggest that conclusions differ across examiners when presented with the same evidence and even within a single examiner when presented with the same evidence on two different occasions [Ulery et al., 2011, 2012]. This suggests the need to supplement these black box decisions with transparent, objective techniques that quantitatively measure the similarity between pieces of evidence [National Research Council, 2009, President’s Council of Advisors on Sci. & Tech., 2016]. In this work, we focus on a specific set of techniques used to compare cartridge case evidence.

Forensic Comparison Pipelines

Recent work in many forensic disciplines has focused on the development of algorithms to measure the similarity between pieces of evidence including glass [Curran et al., 2000, Park and Tyner, 2019, Tyner et al., 2019], handwriting [Crawford, 2020], shoe prints [Park and Carriquiry, 2020], ballistics [Hare et al., 2017, Tai and Eddy, 2018], and toolmarks [Hadler and Morris, 2017, Krishnan and Hofmann, 2018]. These algorithms often result in a numerical, non-binary (dis)similarity score for two pieces of evidence. A non-binary score adds additional nuance to an evidentiary conclusion beyond simply stating whether the evidence did or did not originate from the same source as would be the case in binary classification. For example, the larger the similarity score, the “more similar” the evidence. However, a binary (or ternary, if admitting inconclusives) conclusion must ultimately be reached by an examiner. Whether a decision should be reached based solely on results of a comparison algorithm (e.g., defining a score-based decision boundary) or if an examiner should incorporate the similarity score into their own decision-making process is still up for debate [Swofford and Champod, 2021]. In this work, we view forensic comparison algorithms as a supplement to, rather than a replacement of, the forensic examination.

We conceptualize forensic comparison algorithms as evidence-to-classification “pipelines.” Broadly, the steps of the pipeline include:

1. capturing a digital representation of the evidence,
2. pre-processing this representation to isolate or emphasize a region of interest of the evidence,
3. comparing regions of interest from two different pieces of evidence to obtain a (perhaps high-dimensional) set of similarity features,
4. combining these features into a low-dimensional set of similarity scores, and
5. defining a classification rule based on these similarity features.

This is similar to the structure discussed in Rice [2020]. We add to this structure the emphasis that each step of the pipeline can be further broken-down into modularized pieces. For example, the pre-processing step may include multiple sub-procedures to isolate a region of interest of the evidence. Figure 2 shows three possible variations of the cartridge case

comparison pipeline as well as the parameters requiring manual specification and alternative modules. The benefits of this modularization include easing the process of experimenting with different parameters/sub-procedures and improving the comprehensibility of the overall pipeline.

In the following sections, we detail recent advances to each of the five steps in the pipeline outlined above. We narrow our focus to advances made in comparing F&T evidence.

Digital Representations of Cartridge Case Evidence

Digital representations of cartridge case evidence commonly come in one of two modes: 2D optical images or 3D topographic scans. A common way to take 2D optical images is to take a picture of the cartridge case under a microscope. This implies that the digital representation of the cartridge case surface is dependent on the lighting conditions under which the picture was taken. Some recent work has focused on comparing 2D optical images [Tai and Eddy, 2018, Tong et al., 2014], although the use of 3D microscopes has become more prevalent to capture the surface of ballistics evidence.

Using a 3D microscope, we can obtain scan at the micron (or micrometer) level that is more light-agnostic than a 2D image [Weller et al., 2012]. One common 3D scanning procedure is disc scanning confocal microscopy. This procedure works by shining a focused beam of light on the cartridge case surface. This light is reflected back onto a pinhole allowing a limited height range to pass through. The microscope scans through different height range “slices” and compiles all of these slices into a single 3D topography of the cartridge case primer surface. Figure 3 shows a 2D image and 3D topography of the same cartridge case primer.

More recently, Cadre Forensics™ introduced the TopMatch-3D High-Capacity Scanner [Weller et al., 2015]. This scanner collects images under various lighting conditions of a gel pad into which the cartridge case surface is impressed and combines these images into a regular 2D array called a *surface matrix*. The physical dimensions of these objects are about 5.5 mm^2 captured at a resolution of 1.84 microns per pixel (1000 microns equals 1 mm).

When applied to ballistics evidence, these 3D scans are commonly stored in the ISO standard x3p file format [ISO 25178-72, 2017]. x3p is a container consisting of a single surface matrix representing the height value of the surface and metadata concerning the parameters under which the scan was taken (including size, resolution, creator, etc.). It has been empirically demonstrated that comparing 3D topographic scans of cartridge case evidence leads to more accurate conclusions compared to comparing 2D optical images of the same evidence [Tai, 2019, Tong et al., 2014, Song et al., 2014].

Pre-processing Procedures for Cartridge Case Data

When capturing the surface of a cartridge case, the result is bound to contain extraneous regions due to the incongruity between the circular primer and the rectangular array in which the surface data are stored. Figure 3 shows an example of a 2D image and 3D scan of the same cartridge case. We can see, for example, that the corners of these arrays include non-primer regions of the cartridge case surface. Additionally, the center of the cartridge

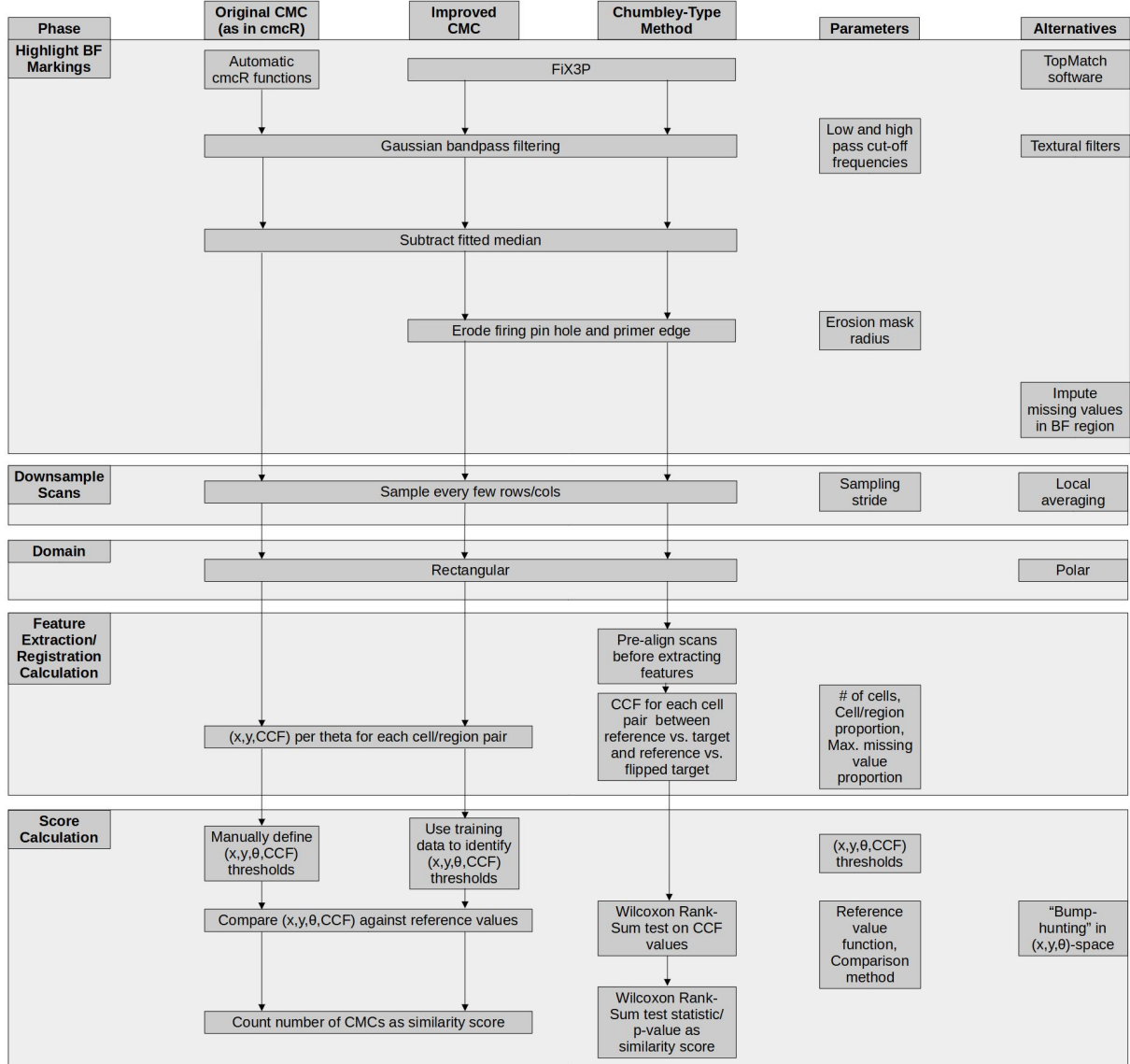


Figure 2: Variations upon the cartridge case comparison pipeline. The first three columns detail the pipeline with different sub-procedures. The fourth column shows the parameters that require manual specification at each step. The fifth column shows alternative processing steps that could replace steps in the existing pipeline.

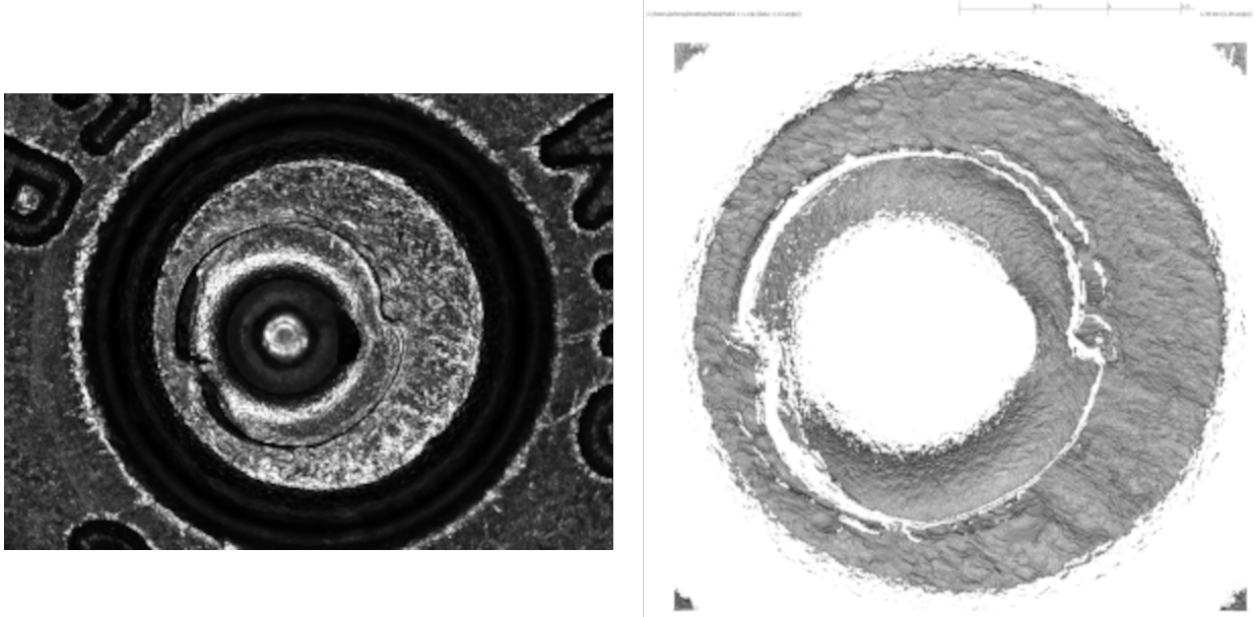


Figure 3: A cartridge case from [Fadul et al. \[2011\]](#) captured using 2D confocal reflectance microscopy (left) and 3D disc scanning confocal microscopy (right).

case primer features an impression left by the firing pin during the firing process. In most applications, impressions left by the firing pin are compared separately from the breech face impressions [\[Zhang et al., 2016\]](#) [cite others?]. As the focus of this work is on the comparison of breech face impressions between two cartridge cases, only the annular region surrounding the firing pin impression is of interest. The annular breech face impression region must be segmented away from the rest of the captured surface.

Both the 2D optical and 3D topographic representations of cartridge case surfaces are fundamentally pictorial in nature. As such, many image processing and computer vision techniques are used to automatically isolate the breech face impression region. [Tai and Eddy \[2018\]](#) uses a combination of histogram equalization, Canny edge detection, and morphological operations to isolate the breech face impressions in 2D images. Various types of Gaussian filters are commonly employed to remove unwanted structure. [Tong et al. \[2014\]](#) uses a low-pass Gaussian filter that removes noise via a Gaussian-weighted moving average operation. [Chu et al. \[2013\]](#), [Song et al. \[2018\]](#) use a bandpass Gaussian filter, which simultaneously performs the function of a low-pass filter along with a high-pass filter to remove global structure from the scan. Other versions of the bandpass filter are used in [Song et al. \[2014\]](#), [Chen et al. \[2017\]](#), [Ott et al. \[2017\]](#) that accomplish tasks such as omitting outlier surface values or addressing boundary effects [\[ISO 16610-71, 2014, Brinkman and Bodschinna, 2003\]](#).

Instead of using automatic procedures, others have used subjective human intervention to isolate the breech face impressions. For example, [\[Song et al., 2018\]](#) indicate that cartridge cases are “manually trimming to extract the breech face impression of interest.” In [Roth et al. \[2015\]](#), examiners manually identify the borders of the breech face impression region by

placing points around an image of the cartridge case primer.

Figure 4 illustrates a sequence of pre-processing procedures applied to a cartridge case surface. By breaking the broader pre-processing step into modularized pieces, we can easily devise other arrangements of these pre-processing procedures that may improve the segmenting or emphasizing of the breech face impressions. Experimentation is enabled by the modularity of the pipeline.

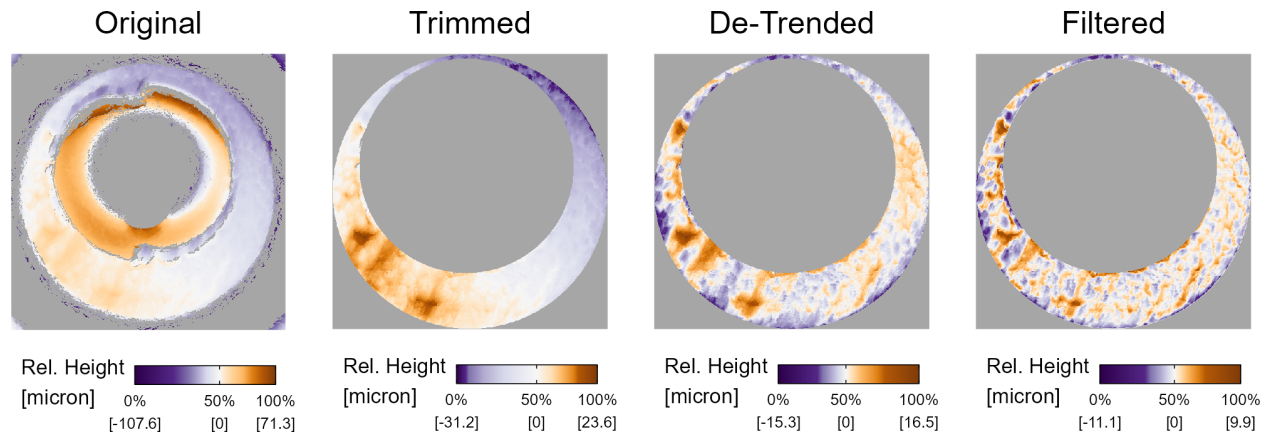


Figure 4: A cartridge case undergoing various pre-processing steps. Note that the distribution of height values changes from left to right as more extraneous noise is removed from the scan.

Cartridge Case Evidence Feature Extraction

After applying the pre-processing procedures to two cartridge case scans, their breech face impressions are compared and similarity features are extracted. Given that the cartridge cases at this point are represented as high-dimensional matrices, this can be thought of as a dimensionality reduction of the high-dimensional surface arrays to a set of similarity statistics.

A variety of features have been proposed to quantify the similarity between two cartridge case surface arrays. [Tai and Eddy \[2018\]](#) propose calculating the cross-correlation function (CCF) value between two cartridge cases across a grid of rotations. It is assumed that the CCF will be larger around the “true” rotation for matching cartridge case pairs than for non-matching pairs. [Riva and Champod \[2014\]](#) proposed combining the CCF between the two aligned scans with the element-wise median Euclidean distance and median difference between the normal vectors at each point of the surface. Later, [Riva et al. \[2016, 2020\]](#) applied Principal Component Analysis to reduce these three features down to two principal components onto which a 2D Kernel Density Estimator could be fit.

Pertinent to this work is the cell-based comparison procedure originally outlined in [Song \[2013\]](#). The underlying assumption of [Song \[2013\]](#) is similar to that of [Tai and Eddy \[2018\]](#): that two matching cartridge cases will exhibit higher similarity when they are “close” to being correctly aligned. While [Tai and Eddy \[2018\]](#) measured similarity using the CCF between the two full scans, [Song \[2013\]](#) proposes partitioning the scans into a grid of “correlation cells” and counting the number of similar cells between the two scans. The rationale behind this

procedure is that many cartridge case scans have regions that do not contain discriminatory markings. As such, comparing full scans may result in a lower correlation than if one were to focus on the highly-discriminatory regions. Dividing the scans into cells allows for the identification of these regions.

After breaking a scan into a grid of cells, each cell is compared to the other scan to identify the rotation and translation, known together as the *registration*, at which the cross-correlation is maximized. Song [2013] assume that the cells from a truly matching pair of cartridge cases will “agree” on their registration in the other scan. Details of this procedure are provided in [Chapter].

Similarity Scores for Cartridge Case Evidence

Following feature extraction, the dimensionality of these features is further reduced to a low-dimensional, usually univariate, similarity score.

After calculating the CCF across various possible registrations, Tai and Eddy [2018] propose using the maximum observed CCF value as the univariate similarity score. In this case, a binary classification can be achieved by setting a CCF threshold above which pairs are classified as “matches” and below which as “non-matches.” Tai [2019] proposes setting a CCF cut-off that maximizes the precision and recall in a training set of pairwise comparisons.

Riva et al. [2016, 2020] use a training set to fit two 2D kernel density estimates to a set of features from matching and non-matching comparisons. Using these estimates, they are able to estimate the score-based likelihood ratio (SLR) for a new set of features. This SLR can be viewed as a similarity score [Garton et al., 2020].

In the case of the cell-based comparison procedure discussed above, the total number of cells that are deemed “congruent matching” is used as a similarity score. The criteria used to define “congruent matching” has changed across papers [Song et al., 2014, Tong et al., 2014, 2015, Chen et al., 2017] and will be discussed in greater detail in [Chapter]. The authors of these papers have consistently used a decision boundary of six “Congruent Matching Cells” to distinguish matches from non-matches.

Zhang et al. [2020] applies the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to the features from the cell-based comparison procedure to determine if any clusters form amongst the per-cell estimated registration values. This is based on the assumption that any cells that come to a consensus on their registration should form a cluster in translation (x, y) and rotation θ space. Zhang et al. [2020] proposes a binary classifier based on whether any clusters are identified by the DBSCAN algorithm. If a cluster is found for a particular pairwise comparison, then that pair is classified as a “match” and otherwise as a “non-match.”

Implementation Considerations

This cartridge case comparison pipeline is similar to other data analysis pipelines. Much like other data analysis pipelines, the procedural details can be obscured as the goals of the analysis become more sophisticated. This is helpful neither for the individual performing the

analysis nor for any consumer of the results. As such, it is worthwhile to design tools that make the data analysis procedure easier to implement and understand [Wickham, 2014].

Beyond simply conceptualizing the cartridge case comparison procedure as a pipeline, we also implement the procedure as a sequence of algorithms that can programatically be connected together. In particular, we utilize the pipe operator `%>%` available from the `magrittr` R package [Bache and Wickham, 2022]. This operator allows the output of one function to be passed as input to another without assigning a new variable. Data can be incrementally transformed as they move from one function to another.

Implementing a data analysis procedure using the pipe operator allows the user to think intuitively in terms of verbs applied to the data. Table 1 illustrates two examples of pipelines that utilize the pipe operator. The left-hand example shows how an R data frame can be manipulated using functions from the `dplyr` package. Functions like `group_by`, `summarize`, and `filter` are simple building blocks that can be strung together to create complicated workflows. The right-hand example similarly illustrates a cartridge case object passing through the comparison pipeline. While the full comparison procedure is complex, the modularization to the `preProcess_`, `comparison_`, and `decision_` steps, which can further be broken-down into simple building blocks, renders the process more understandable to, and flexible for, the user.

Data Frame Manipulation Example	Cartridge Case Comparison Example
<code>dataFrame %>%</code> <code> group_by(category) %>%</code> <code> summarize(x = summary(var)) %>%</code> <code> filter(x > 0) ...</code>	<code>cartridgeCase1 %>%</code> <code> preProcess_func(params1) %>%</code> <code> comparison_func(cartridgeCase2,</code> <code> params2) %>%</code> <code> decision_func(params3) ...</code>

Table 1: Two examples of data analysis workflows that utilize the pipe operator. The left side shows a data frame manipulation while the right side shows a comparison of two cartridge cases.

Our implementation is structured to adhere to the “tidy” principles of design [Wickham et al., 2019]. The `tidyverse` is a collection of R packages that share an underlying design philosophy and structure. The four principles of a tidy API are:

1. Reuse existing data structures.
2. Compose simple functions with the pipe.
3. Embrace functional programming.
4. Design for humans.

Adherence to these principles makes it easier to engage with and understand the overall data analysis pipeline. In our application it also enables experimentation by making it easy to change one step of the pipeline and measure the downstream effects [Zimmerman et al., 2019]. Each step of the cartridge case comparison pipeline requires the user to define parameters. These can range from minor, such as the standard deviation used in a Gaussian filter, to

substantial, such as choosing the algorithm used to calculate the similarity score. So far, there is no consensus on the “best” parameter settings. A large amount of experimentation is yet required to establish these parameters. A tidy implementation of the cartridge case comparison pipeline allows more people to engage in the validation and improvement of the procedure.

Visual Diagnostics

Forensic examiners often provide expert testimony in court cases. As part of this testimony, an examiner is allowed to not only provide facts about the outcome of a forensic examination, but also their opinion about what the results mean. A party to a court may challenge the examiner on the validity of the underlying scientific method or whether they interpreted the results correctly [AAFS, 2021]. In these situations, examiners need to be able to explain the process by which they reached an evidentiary conclusion to the fact finders of the case; namely, the judge or jury. As algorithms are more often used in forensic examinations, the technical knowledge required to understand and explain an algorithm to lay-people has increased. While in some cases the authors of the algorithm have been willing to provide testimony to establish the validity of the algorithm [Indiana County Court of Common Pleas, 2009], this will become less viable as algorithms become more prevalent. Indeed, even the most elegant improvements to an algorithm may be moot if an examiner can’t explain the improvements in their testimony.

The resources required to educate examiners on the use of highly technical algorithms makes additional training seem currently implausible. An alternative is to develop algorithms from the ground-up to be intuitive for examiners to understand and explain to others. *Explainability* refers to the ability to identify the factors that contributed to the results of an algorithm [Belle and Papantonis, 2021]. For example, understanding “why” a classifier predicted one class over another.

Visual Diagnostic Features

We use the visual diagnostic tools discussed in Chapter [5] to develop a set of features. By definition, these features are human-interpretable unlike, for example, features that are calculated in the convolution layer of a convolutional neural network. The interpretability of these features imply that they can be explained to forensic examiners or lay-people. This will make it easier to introduce such methods into forensic labs and court rooms.

Reproducibility, Comprehensibility, and Approachability of Algorithms

[Note that in the case of human intervention in the pipeline, it is difficult to quantify what we mean by “parameters.” For example, how would we measure the consistency, attentiveness, or skill of an individual performing manual pre-processing on a piece of evidence? This directly impacts the reproducibility of the pipeline as it is highly unlikely that the conditions under which a manual intervention occurred can be exactly reproduced. As we will see in [Chapter],

ensuring reproducibility of the pipeline in this case all but requires providing a copy of the manually-processed data.]

When evidence derived from a scientific method is presented to a judge or jury during court proceedings, there is an ethical imperative that the underlying scientific method be proven effective in its intended usage. Established in U.S. [1993], the Daubert standard legally codifies this notion by requiring a scientific method to satisfy a set of criteria before being considered admissible. These criteria include that the method be generally accepted in the scientific community, that it can be and has been tested, and that it has a known error rate.

An impediment to achieving general acceptance of a scientific method is if the method is difficult to test by the wider scientific community, and therefore difficult to estimate its error rate [cite a paper on difficult to reproduce studies]. This prompts the question: how can we make scientific methods easier to test? Generally, “easier” implies that the method requires few resources or resources that the scientific community has access to. In the case of computational algorithms, virtually everyone has access to a computer and therefore has the means to run (reasonably-sized) algorithms. Additional requirements include code and input data which, if not available, must be produced before an algorithm can be executed. This can be expensive and time-consuming [citation here].

Comprehensibility = understanding what the algorithm is doing. Approachability = being able to use or change the algorithm yourself.

Baggerly and Coombes (2009) *Deriving Chemosensitivity from Cell Lines: Forensic Bioinformatics and Reproducible Research in High-Throughput Biology*

Donoho (2017) *50 Years of Data Science*

tidyverse functionality

National Academies of Sciences, Engineering, and Medicine (2019) *Reproducibility and Replicability in Science*

In cartridge case evidence, the Congruent Matching Cells (CMC) methods are one class of algorithms used to measure the similarity between two cartridge cases. Numerous authors have demonstrated the ability of the CMC methods to effectively distinguish between matching and non-matching pairs of cartridge cases. However, to-date only conceptual descriptions of the CMC methods, along with results derived from an internal implementation of the described algorithm, have been published for the wider scientific community. These published descriptions and results demonstrate that the authors’ implementation of the CMC methods work as intended, yet fail to ensure others can reproduce or develop upon the published work without having to create their own implementation. By “reproduce,” we mean *computational reproducibility* as defined by the National Academy of Science, Engineering, and Medicine:

Definition here

For a method to be widely accepted by the scientific community, results must be reproducible (by others).

By definition, algorithms are repeatable assuming the same data are provided as input and

a seed is set for any internal random number generation. That is to say, if the exact same procedure is performed on the same data on two separate occasions, then the results will be the same. However, reproducibility is still in-question for many forensic comparison algorithms. By reproducibility, we mean *computational reproducibility* as defined by the [National Academy of Sciences, Engineering, and Medicine]: “obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis.” As we argue in [Chapter], many published forensic comparison algorithms fail to provide sufficient detail or resources to satisfy one or more of these criteria and, therefore, are not reproducible. In-short: if the code and data exist, then you should share it. Using a specific cartridge case comparison algorithm as an example, we detail a development process by which algorithms can not only satisfy reproducibility, but can also be more easily understood and accessed by the wider scientific community.

In Chapter 6, we introduce a taxonomy to classify various levels of computational reproducibility. We argue that currently-published versions of the CMC method are provided in the form of *conceptual descriptions* while our implementation, paired with data that are open-source on the National Ballistics Toolmark Database [cite NBTRD], satisfies the definition of computational reproducibility provided by the NASEM [cite NASEM].

Beyond considerations of reproducibility, which is a core tenet of science, are considerations of comprehensibility and approachability of an algorithm. [Discuss how pipeline aids in comprehension, experimentation].

For data discussion: cite Vorberger et al. (2007), NBTRD, and TopMatch

[Carpentry R package, gapminder reproducibility seminar] <https://swcarpentry.github.io/r-novice-gapminder/>

References

- AAFS. What is forensic science?, 2021. URL <https://www.aafs.org/careers-forensic-science/what-forensic-science>.
- AFTE Criteria for Identification Committee. Theory of identification, range striae comparison reports and modified glossary definitions. *AFTE Journal*, 24(3):336–340, 1992.
- Stefan Milton Bache and Hadley Wickham. *magrittr: A Forward-Pipe Operator for R*, 2022. URL <https://CRAN.R-project.org/package=magrittr>. R package version 2.0.2.
- Vaishak Belle and Ioannis Papantonis. Principles and practice of explainable machine learning. *Frontiers in Big Data*, 4, 2021.
- S. Brinkman and H. Bodschwinna. Advanced Gaussian filters. In L. Blunt and X. Jiang, editors, *Advanced Techniques for Assessment Surface Topography: Development of a Basis for 3D Surface Texture Standards "SURFSTAND"*. Elsevier Inc., United States, 2003. URL <https://doi.org/10.1016/B978-1-903996-11-9.X5000-2>.
- Zhe Chen, John Song, Wei Chu, Johannes A. Soons, and Xuezheng Zhao. A convergence algorithm for correlation of breech face images based on the congruent matching cells (CMC)

- method. *Forensic Science International*, 280:213–223, November 2017. ISSN 03790738. URL <https://doi.org/10.1016/j.forsciint.2017.08.033>.
- Wei Chu, Mingsi Tong, and John Song. Validation Tests for the Congruent Matching Cells (CMC) Method Using Cartridge Cases Fired with Consecutively Manufactured Pistol Slides. *Journal of the Association of Firearms and Toolmarks Examiners*, 45(4):6, 2013. URL <https://www.nist.gov/publications/validation-tests-congruent-matching-cells-cmc-method-using-cartridge-cases-fired>.
- Amy Crawford. *Bayesian hierarchical modeling for the forensic evaluation of handwritten documents*. Ph.D thesis, Iowa State University, 2020. URL <https://doi.org/10.31274/etd-20200624-257>.
- James Michael Curran, Tacha Natalie Hicks Champod, and John S Buckleton, editors. *Forensic interpretation of glass evidence*. CRC Press, Boca Raton, FL, June 2000.
- T. Fadul, G. Hernandez, S. Stoiloff, and Gulati Sneha. An Empirical Study to Improve the Scientific Foundation of Forensic Firearm and Tool Mark Identification Utilizing 10 Consecutively Manufactured Slides, 2011. URL <https://www.ojp.gov/ncjrs/virtual-library/abstracts/empirical-study-improve-scientific-foundation-forensic-firearm-and>.
- Nathaniel Garton, Danica Ommen, Jarad Niemi, and Alicia Carriquiry. Score-based likelihood ratios to evaluate forensic pattern evidence. 2020. doi: 10.48550/ARXIV.2002.09470. URL <https://arxiv.org/abs/2002.09470>.
- Jeremy R. Hadler and Max D. Morris. An improved version of a tool mark comparison algorithm. *Journal of Forensic Sciences*, 63(3):849–855, September 2017. doi: 10.1111/1556-4029.13640. URL <https://doi.org/10.1111/1556-4029.13640>.
- Eric Hare, Heike Hofmann, and Alicia Carriquiry. Automatic Matching of Bullet Land Impressions. *The Annals of Applied Statistics*, 11(4):2332–2356, 2017. URL <http://arxiv.org/abs/1601.05788>.
- Indiana County Court of Common Pleas. *Commonwealth of Pennsylvania vs. Kevin J. Foley*. 03 2009.
- ISO 16610-71(2014). Geometrical product specifications (GPS) - Filtration - Part 71: Robust areal filters: Gaussian regression filters. Standard, International Organization for Standardization, Geneva, CH, 2014. URL <https://www.iso.org/standard/60159.html>.
- ISO 25178-72(2017). Geometrical product specifications (GPS) — Surface texture: Areal — Part 72: XML file format x3p. Standard, International Organization for Standardization, Geneva, CH, 2017. URL <https://www.iso.org/standard/62310.html>.
- Ganesh Krishnan and Heike Hofmann. Adapting the chumbley score to match striae on land engraved areas (leas) of bullets,. *Journal of Forensic Sciences*, 64(3):728–740, November 2018. doi: 10.1111/1556-4029.13950. URL <https://doi.org/10.1111/1556-4029.13950>.
- National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, Washington, DC, 2009. ISBN 978-0-309-13130-8. URL <https://doi.org/10.17226/12589>.

- Danica M Ommen and Christopher P Saunders. Building a unified statistical framework for the forensic identification of source problems. *Law, Probability and Risk*, 17(2):179–197, May 2018. doi: 10.1093/lpr/mgy008. URL <https://doi.org/10.1093/lpr/mgy008>.
- OSAC Human Factors Committee. Human factors in validation and performance testing of forensic science. Technical report, May 2020. URL <https://doi.org/10.29325/osac.ts.0004>.
- Daniel Ott, Robert Thompson, and Junfeng Song. Applying 3D measurements and computer matching algorithms to two firearm examination proficiency tests. *Forensic Science International*, 271:98–106, February 2017. ISSN 03790738. URL <https://doi.org/10.1016/j.forsciint.2016.12.014>.
- Soyoung Park and Alicia Carriquiry. An algorithm to compare two-dimensional footwear outsole images using maximum cliques and speeded-up robust feature. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 13(2):188–199, 2020. ISSN 1932-1872. URL <https://doi.org/10.1002/sam.11449>.
- Soyoung Park and Sam Tyner. Evaluation and comparison of methods for forensic glass source conclusions. *Forensic Science International*, 305:110003, December 2019. doi: 10.1016/j.forsciint.2019.110003. URL <https://doi.org/10.1016/j.forsciint.2019.110003>.
- President’s Council of Advisors on Sci. & Tech. Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods. 2016. URL https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.
- Kiegan E. Rice. *A Framework for Statistical and Computational Reproducibility in Large-Scale Data Analysis Projects with a Focus on Automated Forensic Bullet Evidence Comparison*. PhD thesis, 2020. URL <https://www.proquest.com/dissertations-theses/framework-statistical-computational/docview/2446975934/se-2?accountid=10906>. Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2021-05-25.
- Fabiano Riva and Christophe Champod. Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases. *Journal of Forensic Sciences*, 59(3):637–647, February 2014. doi: 10.1111/1556-4029.12382. URL <https://doi.org/10.1111/1556-4029.12382>.
- Fabiano Riva, Rob Hermsen, Erwin Mattijssen, Pascal Pieper, and Christophe Champod. Objective evaluation of subclass characteristics on breech face marks. *Journal of Forensic Sciences*, 62(2):417–422, December 2016. doi: 10.1111/1556-4029.13274. URL <https://doi.org/10.1111/1556-4029.13274>.
- Fabiano Riva, Erwin J.A.T. Mattijssen, Rob Hermsen, Pascal Pieper, W. Kerkhoff, and Christophe Champod. Comparison and interpretation of impressed marks left by a firearm on cartridge cases – towards an operational implementation of a likelihood ratio based technique. *Forensic Science International*, 313:110363, August 2020. doi: 10.1016/j.forsciint.2020.110363. URL <https://doi.org/10.1016/j.forsciint.2020.110363>.

- Joseph Roth, Andrew Carriveau, Xiaoming Liu, and Anil K. Jain. Learning-based ballistic breech face impression image matching. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8, 2015. doi: 10.1109/BTAS.2015.7358774.
- John Song. Proposed “NIST Ballistics Identification System (NBIS)” Based on 3D Topography Measurements on Correlation Cells. *American Firearm and Tool Mark Examiners Journal*, 45(2):11, 2013. URL https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=910868.
- John Song, Wei Chu, Mingsi Tong, and Johannes Soons. 3D topography measurements on correlation cells—a new approach to forensic ballistics identifications. *Measurement Science and Technology*, 25(6):064005, June 2014. ISSN 0957-0233, 1361-6501. URL <https://doi.org/10.1088/0957-0233/25/6/064005>.
- John Song, Theodore V. Vorburger, Wei Chu, James Yen, Johannes A. Soons, Daniel B. Ott, and Nien Fan Zhang. Estimating error rates for firearm evidence identifications in forensic science. *Forensic Science International*, 284:15–32, March 2018. ISSN 03790738. URL <https://doi.org/10.1016/j.forsciint.2017.12.013>.
- H. Swofford and C. Champod. Implementation of algorithms in pattern & impression evidence: A responsible and practical roadmap. *Forensic Science International: Synergy*, 3:100142, 2021. doi: 10.1016/j.fsisyn.2021.100142. URL <https://doi.org/10.1016/j.fsisyn.2021.100142>.
- Xiao Hui Tai. *Matching Problems in Forensics*. PhD thesis, 2019. URL https://kilthub.cmu.edu/articles/Matching_Problems_in_Forensics/9963596/1.
- Xiao Hui Tai and William F. Eddy. A Fully Automatic Method for Comparing Cartridge Case Images,. *Journal of Forensic Sciences*, 63(2):440–448, March 2018. ISSN 00221198. URL <http://doi.wiley.com/10.1111/1556-4029.13577>.
- Robert Thompson. *Firearm Identification in the Forensic Science Laboratory*. National District Attorneys Association, 08 2017. URL <https://doi.org/10.13140/RG.2.2.16250.59846>.
- Mingsi Tong, John Song, Wei Chu, and Robert M. Thompson. Fired Cartridge Case Identification Using Optical Images and the Congruent Matching Cells (CMC) Method. *Journal of Research of the National Institute of Standards and Technology*, 119:575, November 2014. ISSN 2165-7254. URL <https://doi.org/10.6028/jres.119.023>.
- Mingsi Tong, John Song, and Wei Chu. An Improved Algorithm of Congruent Matching Cells (CMC) Method for Firearm Evidence Identifications. *Journal of Research of the National Institute of Standards and Technology*, 120:102, April 2015. ISSN 2165-7254. URL <https://doi.org/10.6028/jres.120.008>.
- Sam Tyner, Soyoung Park, Ganesh Krishnan, Karen Pan, Eric Hare, Amanda Luby, Xiao Hui Tai, Heike Hofmann, and Guillermo Basulto-Elias. sctyner/openforscir: Create doi for open forensic science in r, 2019. URL <https://zenodo.org/record/3418141>.
- Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts. Accuracy and reliability of forensic latent fingerprint decisions. *Proceedings of the National Academy*

- of Sciences*, 108(19):7733–7738, April 2011. doi: 10.1073/pnas.1018707108. URL <https://doi.org/10.1073/pnas.1018707108>.
- Bradford T. Ulery, R. Austin Hicklin, JoAnn Buscaglia, and Maria Antonia Roberts. Repeatability and reproducibility of decisions by latent fingerprint examiners. *PLoS ONE*, 7(3):e32800, March 2012. doi: 10.1371/journal.pone.0032800. URL <https://doi.org/10.1371/journal.pone.0032800>.
- Bradford T. Ulery, R. Austin Hicklin, Maria Antonia Roberts, and JoAnn Buscaglia. Measuring what latent fingerprint examiners consider sufficient information for individualization determinations. *PLoS ONE*, 9(11):e110179, November 2014. doi: 10.1371/journal.pone.0110179. URL <https://doi.org/10.1371/journal.pone.0110179>.
- U.S. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, volume 509. 1993.
- Todd Weller, Marcus Brubaker, Pierre Duez, and Ryan Lilien. Introduction and initial evaluation of a novel three-dimensional imaging and analysis system for firearm forensics. *AFTE Journal*, 47:198, 01 2015.
- Todd J. Weller, Alan Zheng, Robert Thompson, and Fred Tulleners. Confocal microscopy analysis of breech face marks on fired cartridge cases from 10 consecutively manufactured pistol slides. 57(4), mar 2012. doi: 10.1111/j.1556-4029.2012.02072.x. URL <https://doi.org/10.1111/j.1556-4029.2012.02072.x>.
- Hadley Wickham. Tidy data. *The Journal of Statistical Software*, 59, 2014. URL <http://www.jstatsoft.org/v59/i10/>.
- Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019. doi: 10.21105/joss.01686.
- Hao Zhang, John Song, Mingsi Tong, and Wei Chu. Correlation of firing pin impressions based on congruent matching cross-sections (CMX) method. *Forensic Science International*, 263:186–193, June 2016. doi: 10.1016/j.forsciint.2016.04.015. URL <https://doi.org/10.1016/j.forsciint.2016.04.015>.
- Hao Zhang, Jialing Zhu, Rongjing Hong, Hua Wang, Fuzhong Sun, and Anup Malik. Convergence-improved congruent matching cells (CMC) method for firing pin impression comparison. *Journal of Forensic Sciences*, 66(2):571–582, November 2020. doi: 10.1111/1556-4029.14634. URL <https://doi.org/10.1111/1556-4029.14634>.
- Naupaka Zimmerman, Greg Wilson, Raniere Silva, Scott Ritchie, François Michonneau, Jeffrey Oliver, Harriet Dashnow, Andrew Boughton, Andy Teucher, David Mawdsley, Andrew MacDonald, Timothy Rice, Rémi Emonet, Remi Daigle, Bill Mills, Ben Bolker, Sam Penrose, Clare Sloggett, John Blischak, Timothy Eoin Moore, David Mawdsley, Jeffrey Arnold, Dave Bridges, Erin Alison Becker, Giulio Valentino Dalla Riva, Liz Ing-Simmons, Research Bazaar, Trevor Bekolay, Julia Piaskowski, Marc Sze, Martin John Hadley, Nima Hejazi,

Fernando Mayer, Katrin Leinweber, Lachlan Deer, Nicholas Lesniak, Olivia Rata Burge, Paula Andrea Martinez, Ana Costa Conrado, Andris Jankevics, Jaime Ashander, Jonah Duckles, Luke Zappia, Marie-Helene Burle, Nora Mitchell, Phil Bouchet, Rayna Michelle Harris, Sebastien Renaut, Adam H. Sparks, , Daniel, Dean Attali, Drew Tyre, Elise Morrison, Gordon McDonald, Ido Bar, James Mickley, Jamie McDevitt-Irwin, Katherine Koziar, Kieran Samuk, Kunal Marwaha, Kyriakos Chatzidimitriou, Lucy Chang, Melissa Kardish, Nicholas Potter, Philipp Boersch-Supan, Scott Allen Funkhouser, Tobin Magle, Waiteb5, Ahsan Ali Khoja, Amy Lee, Antonio Berlanga-Taylor, Ashwin Srinath, Bippuspm, Bret Beheim, Butterflyskip, David J. Harris, Diego Rabatone Oliveira, James Balamuta, Josh Quan, Kara Woo, Kate Hertweck, Kellie Ottoboni, Kevin Weitemier, Lex Nederbragt, Andrew Lonsdale, Luke W Johnston, Marieke Frassl, Mark Dunning, Mary Donovan, Matt Clark, Mike Jackson, Murray Cadzow, Narayanan Raghupathy, Nelly Sélem, Pete Bachant, Piotr Banaszkiewicz, Richard Barnes, Robert Bagchi, Sandra Brosda, Sarah Munro, Sasha Lavrentovich, Thea Van Rossum, Tyler Crawford Kelly, Vicken Hillis, Kiana Ashley West, and Yuka Takemon. swcarpentry/r-novice-gapminder: Software carpentry: R for reproducible scientific analysis, june 2019, 2019. URL <https://zenodo.org/record/3265164>.