Journal of Data Science 0 (0), 1–15
??? 2025

DOI: 10.6339/25-JDS1189
Education in Data Science

# Quantifying the Alignment of a Data Analysis Between Analyst and Audience

Lucy D'Agostino McGowan[1], Roger D. Peng[2,*], and Stephanie C. Hicks[3,4,*]

[1]Department of Statistical Sciences, Wake Forest University, USA
[2]Department of Statistics and Data Sciences, University of Texas at Austin, USA
[3]Department of Biostatistics, Johns Hopkins University, USA
[4]Department of Biomedical Engineering, Johns Hopkins University, USA

**Abstract**

A challenge that data scientists face is building an analytic product that is useful and trustworthy for a given audience. Previously, a set of principles for describing data analyses were defined that can be used to create a data analysis and to characterize the variation between analyses. Here, we introduce a concept called the *alignment* of a data analysis, which is between the data analyst and an audience. We define an aligned data analysis as the matching of principles between the analyst and the audience for whom the analysis is developed. In this paper, we propose a model for evaluating the alignment of a data analysis and describe some of its properties. We argue that more generally, this framework provides a language for characterizing alignment and can be used as a guide for practicing data scientists to building better data products.

**Keywords** *analytic design theory; data science; evaluation*

## 1 Introduction

In the practice of data science, a data scientist builds a data analysis to extract knowledge and insights from examining data (Tukey, 1962; Tukey and Wilk, 1966). More recently, the knowledge developed from data analyses is often encoded into machine learning algorithms or related data products to facilitate use by large numbers of users. Yet, the discussion of how to build a data analysis that is trusted by its users often proceeds without explicit reference to an audience or consumer for whom the data analysis is being developed. Indeed, there is much for a data analyst to consider on their own with respect to statistical techniques, visualization methods, data processing approaches, and computational algorithms that do not involve the needs or requirements of an audience member in particular. However, a critical goal for many data analyses is to be useful or persuasive to another person (Kimball, 1957). The audience could range from simply the person doing the analysis to a much larger external group.

A general goal for the data scientist is to build data analyses that are trustworthy and for others to have trust in the work that they produce. The extent to which results from data analyses are used for key policy decisions enhances the need for trust between analyst and audience. Broderick et al. (2023) note that complex data analyses with long data pipelines present numerous opportunities for trust to break down. In particular, they note that trust can break down if the evidence generated by an analysis is not useful for decision-making. An

*Corresponding author. Email: roger.peng@austin.utexas.edu or shicks19@jhu.edu.


Received November 26, 2024; Accepted May 22, 2025

alternative framework is proposed by Yu and Kumbier (2020), who argue that trustworthiness in the data science life cycle can be achieved through building analyses that have predictability, computability, and stability (see also Yu and Barter, 2024). A principle that ties both frameworks together is the characterization of the development of trust as primarily being in the hands of the analyst, via decisions made about study design, data collection, model choice, and other aspects of the data science process. Although the human element of data analysis is clearly acknowledged, it is primarily in reference to the analyst. The identity and characteristics of the audience, i.e. the person designated to receive the analysis results, are not specified in any detail.

A significant challenge for the data analyst then is to build a data analysis that is useful for the intended audience while also being trustworthy and adhering to generally accepted statistical principles for high quality analyses. One extreme would be for the data analyst to ignore the audience and build the data analysis as they see fit, hoping that the audience simply accepts the analysis upon presentation. Another extreme would involve the data analyst simply giving the audience whatever they asked for (within the confines of statistical theory, of course) without the need for additional discussion. Neither extreme seems ideal, in general. Analysts who ignore the specific interests of the audience risk being ignored in turn, rendering their analysis ultimately useless. Analysts who meet the audience's specifications exactly miss an opportunity to educate the audience on techniques or approaches that might better serve their needs. In either case, the analysis conducted is not as useful as it perhaps could have been (Mira and Wit, 2021).

One approach to addressing the challenge of building a useful analysis is to consider the audience as part of the design of the analysis itself. The role of the audience here is to contribute the definition of usefulness for a given analysis and specify their requirements for meeting that definition, similar to the way an architect designs a house for someone to live in or the way an engineer designs equipment for a specific purpose. Initially, the analyst may not completely agree with the definition of usefulness nor have a desire or ability to meet the requirements. There may be a lack of *alignment* between the analyst and the audience over what constitutes utility in the context of the analysis. Subsequently, a negotiation may occur between the analyst and audience in order to come to agreement over what would make the analysis most useful and how the analyst might shape the analysis to achieve that goal.

This paper seeks to define the concept of alignment between a data analyst and an audience and to discuss its role in the design of data analyses and in ensuring the usefulness of analytic outputs. Usefulness can mean a variety of things here, but typically, a useful analysis will influence a decision-making process in a scientific, business, or policy context. Analyses designed with a broad range of audiences in mind may serve a rhetorical purpose, aiming to convince an audience of a specific point or to inform them about an important issue. From the analyst's perspective, alignment can be thought of as a heuristic for guiding the organization of the elements of a data analysis and its presentation.

We start by leveraging a set of previously introduced principles for data analyses that can be used to guide the creation of a data analysis and to characterize the variation between data analyses (D'Agostino McGowan et al., 2022). These principles of data analysis are prioritized qualities or characteristics that are relevant to the analysis, as a whole or individual components. For a given data analysis, a data analyst can assign allocations to these principles to increase or decrease the amount of resources dedicated to these characteristics in a given data analysis. These allocations can be highly influenced by outside constraints such as time or budget. In this way, different allocations of the principles by the analyst can lead to different data analyses, all addressing the same underlying question (Silberzahn et al., 2018).

Next, we use this set of principles for data analysis to propose a framework for modeling the alignment of a data analysis that relies critically on the audience for whom the analysis is developed. In particular, as every data analysis has an audience that views the analysis with their own preconceived notions, characteristics, and biases, we consider the allocations of the principles by both the analyst and the audiences, who may have a different perspectives on how these various principles should be allocated for a given data analysis. Neither set of principles allocated by the analyst or the audience is necessarily correct or incorrect. However, we previously hypothesized that how successfully aligned a data analysis is may depend on how well-matched the analyst's allocations are to the audience's allocations for a given analysis (Hicks and Peng, 2019). In this paper, we make these ideas more concrete and introduce an evaluation metric that we call the *alignment* of a data analysis between the analyst and the audience. We define an aligned data analysis as the matching of allocated principles between the analyst and the audience on which the analysis is developed. In the following sections, we formalize those ideas by proposing a general framework for evaluating the alignment of a data analysis (Sections 2–3) and describe an example of how the concept of alignment can be used in the classroom to train students in the analysis design process (Section 4).

## 2 Components of Variation in Principle Allocation

As described above and in previous work, we consider data analyses to be constructed in a manner guided by a set of design principles (D'Agostino McGowan et al., 2022) characterizing the data analysis. Specifically, D'Agostino McGowan et al. (2022) defined six principles of data analysis as

1. *Data matching*: The extent to which the analysis has data readily measured or available to the analyst that directly matches the data needed to investigate a question.
2. *Exhaustive*: The extent to which specific questions are addressed using multiple, complementary methods, tooling, or workflows.
3. *Skeptical*: The extent to which multiple, related questions or alternative hypotheses are considered using the same data.
4. *Second-order*: The extent to which methods, tooling, or workflows that do not directly address the primary question, but give important context or supporting information to the analysis, are included.
5. *Clarity*: The extent to which key pieces of evidence in the data that explain the most "variation" or are most influential to understanding the results or conclusions are summarized or visualized.
6. *Reproducible*: The extent to which someone who is not the original analyst can take the published code and data and compute the same results as the original analyst. For the purposes of this paper, we assume the analyst can *always* reproduce the analysis for themselves.

In this paper, we assume that the data analyst allocates a set of resources to each principle. We allow for the possibility that there will be variation in the allocation of the principles from analysis to analysis, for both analyst and audience.

### 2.1 Analyst Characteristics

From the analyst's perspective, some of the determinants for how a given principle may be allocated are:

1. *Analysis-specific Resources.* Considerations about computing resources, time, budget, personnel, and other such resources and analysis characteristics can often require that an analyst allocate more or less to certain principles for analysis. For example, analyses that must be conducted in a short amount of time may be limited in their ability to explore multiple competing hypotheses and exhibit low skepticism.
2. *Question Significance.* The significance of the question being addressed with the data may play a role in determining principle allocations. Questions of high significance, for example, may require a high degree of transparency or reproducibility for the audience. Questions of lower significance may be done in a "quick-and-dirty" fashion; should the question's significance change in the future the analysis may need to be re-done with a different set of principle allocations.
3. *Field-specific Conventions.* Analysts are often members of a field from which they may have received their training (e.g. statistics, economics, computer science, bioinformatics). Each field develops conventions regarding how analyses in their field should be conducted. Tukey (1962) emphasized that in data analysis, there is a heavy emphasis on "judgment", one particular form of which is based upon the experience of members of a given field.
4. *Analytic Product.* Depending on the analytic product that will ultimately be presented to the audience (e.g. PDF document, web-based dashboard, executable R Markdown document), the analyst may determine that certain principles should receive a greater or lesser allocation.

## 2.2 Audience Characteristics

Similarly, the audience for whom the analysis is being developed will determine their principle allocations based on a variety factors, including their *perception* of resources available to the analyst, their judgment of the significance of the question, their own field-specific conventions (assuming the audience and the analyst are not members of the same field), and their perception of what the analytic product should contain.

The enumerated list in Section 2.1 describes some of the fixed factors that may drive variation in how various data analytic principles are allocated. However, there may be variation that is more random in nature. In particular, we consider the randomness as arising from sampling from a population of analysts or potential audiences trained in specific fields. Different analysts, presented with the exact same question and data, will likely allocate principles differently and hence produce different analyses based on their own personal characteristics. Similarly, different audiences, considering the same analytic product, will allocate principles differently and evaluate the alignment of the analysis differently.

Data analyses are built to be viewed by an audience and the nature of that audience can affect how analyses are designed and planned. For the purposes of this paper, we describe audiences as falling into three categories:

1. Analyst only: Here, the analyst is building an analysis for themselves and is not anticipating that other people will be viewing it. Because the analyst is serving as the audience, the analyst is able to allocate principles for the analysis that will be well-aligned.
2. Single or Small Audience: Many analyses are built for either a single person, such as collaborator or client, or a small group of people. The assumption in this scenario is that the analyst would have access to the audience and would be able to discuss with them what principles should be emphasized in the analysis.
3. Multiple or Large Audience: Analyses built for very large audiences, such as in a research paper or a large conference presentation, generally cannot be designed in a manner that

assumes access to the audience. In this scenario, the analyst must to some degree speculate about how the typical audience member would allocate different principles for the analysis. In general, the audience will have their own preferences for how resources should be allocated for the analyst building the data analysis and those preferences could be in conflict with how the analyst would allocate those resources.

### 2.3 Example: Initially Misaligned Analysis

Consider two principles that may play a role in a data analysis: The reproducibility of the analysis by people who are not the analyst (Peng, 2011) and how well-matched the data are to the analysis (D'Agostino McGowan et al., 2022). Suppose an analyst is in a organization where there is a request to do a quick analysis for a presentation in an internal meeting one week from now. The analyst, as a general matter, may feel it worthwhile to dedicate time and energy to ensure that the analysis is reproducible by others, even if it makes the analysis take longer to prepare and execute. In addition, the analyst notices that while the exact dataset for the analysis is not yet available, a dataset containing a similar surrogate measure is available now. Given that the analysis is needed relatively quickly, the analyst figures that the surrogate measure is sufficient. Meanwhile the person requesting the analysis (the audience) would greatly prefer if more effort were taken to obtain the exact dataset needed for the analysis, and that less time be spent on making it reproducible, given that the audience at this internal meeting has no interest in re-analyzing the data for themselves. For this analyst and audience pair, we have a mismatch on these two principles in that the analyst would prefer to devote more time to making the analysis reproducible relative to matching the data while the audience would prefer that more effort be spent on getting the better data relative to making the analysis reproducible by others.

As we will describe in greater detail below in Section 3, the mismatch in priorities between the analyst and the audience in this example indicate a lack of alignment. If, for example, we could quantify the extent to which the analyst and the audience prioritize either reproducibility or the matching of the datasets, we could then quantify the extent to which the analyst and audience are misaligned (what we might call baseline alignment). This quantitative difference in priorities will serve as the basis for defining alignment in an data analysis. Furthermore, it is possible that the analyst an audience could engage in negotiation to adjust their respective priorities for these principles (reproducibility and data matching), thereby improving their overall analysis alignment.

## 3 Stages of Analytic Design

We can imagine that there are a fixed amount of resources (whether it be time, money, etc), that can be allocated to each of the principles. Both the analyst and audience may have different expectations of how these should be allocated. We conceive of the analytic design process as broadly occurring in a sequence of stages. At the first or *baseline* stage the analyst and audience independently allocate various principles based on their field-specific conventions and personal views of the analysis. Following this stage is the *analytic negotiation* stage, where the analyst and audience discuss the proposed analysis and negotiate over how various principles will be allocated. We do not view the analytic negotiation stage as necessarily contentious, but rather an opportunity for analyst and audience to understand the tradeoffs that each party is considering.

Finally, in the *resolution* stage, both analyst and audience adjust their principle allocations based on the results of the analytic negotiation.

### 3.1 Defining the Alignment of a Data Analysis

With the analyst allocations to specific principles and the audience allocations, we can proceed to define the alignment of a data analysis.

**Definition 1** (Baseline Alignment)**.** The baseline alignment of a data analysis between analyst $i = 1, \ldots, N$ and audience $j = 1, \ldots, M$ is defined in terms of the principle-specific allocation difference for principle $k = 1, \ldots, K$,

$$B_{ij}^{(k)} \quad = \quad V_i^{(k)} - W_j^{(k)}, \tag{1}$$

where $V_i^{(k)}$ is the analyst's baseline allocation for principle $k$ and $W_j^{(k)}$ is the audience's baseline allocation for principle $k$. The overall baseline alignment for an analysis is defined via the collection of principle-specific allocation differences across all principles, $\mathbf{B}_{ij} = \left(B_{ij}^{(1)}, \ldots, B_{ij}^{(K)}\right)$.

The baseline alignment for an analysis represents the alignment that exists before the analyst and and audience meet and negotiate any possible adjustments (Figure 1). After the negotiation stage, we define the analyst's principle adjustments in the resolution stage, i.e. the change in principle allocation from the baseline stage, as $\phi_i^{(k)}$ for analyst $i$ and principle $k$. Similarly, the audience's principle adjustments are indicated as as $\theta_j^{(k)}$ for audience $j$ and principle $k$.

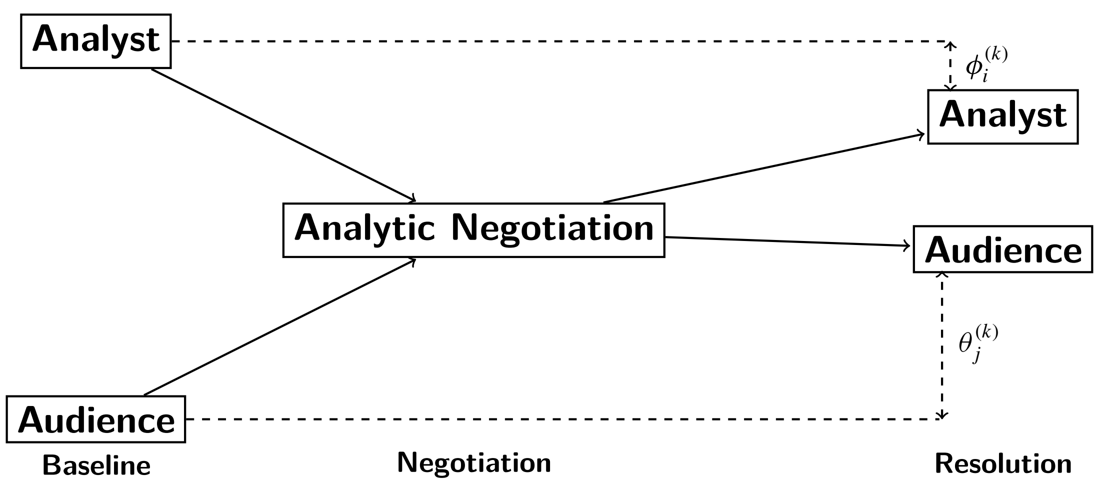


Figure 1: Three stages of analytic design. At the baseline stage, the analyst and audience independently allocate various principles based on their field-specific conventions and personal views of the analysis. Next, the analyst and audience engage in stage of analytic negotiation, where they discuss the proposed analysis and negotiate over how various principles will be allocated. Finally, in the resolution stage, both analyst $i$ and audience $j$ adjust their $k$th principle allocations ($\phi_i^{(k)}$ and $\theta_j^{(k)}$, respectively) based on the results of the analytic negotiation.

In some cases, the analyst may not have an opportunity to interact directly with the audience to negotiate the principle allocations. In those cases, the analyst may obtain some indirect understanding of the audience's expectations for an analysis by doing some background research.

Such background research may lead to an adjustment of the analyst's principle allocations via $\phi_i^{(k)}$. However, the audience's allocations will remain unchanged in the resolution stage (i.e. $\theta_j^{(k)} = 0$) because there was never any direct interaction with the analyst.

Next, given a $B_{ij}^{(k)}$, $\phi_i^{(k)}$, and $\theta_j^{(k)}$, we define the overall analysis alignment for principle $k$.

**Definition 2** (Overall Analysis Alignment)**.** For a given principle $k$, let $B_{ij}^{(k)}$ represent the baseline alignment between analyst $i$ and audience $j$ for principle $k$. Let $\phi_i^{(k)}$ and $\theta_j^{(k)}$ represent allocation adjustments made by analyst $i$ and audience $j$, respectively, upon reaching the resolution stage. Then the overall analysis alignment for principle $k$ is defined as

$$\begin{aligned} D_{ij}^{(k)} &= B_{ij}^{(k)} + \left(\phi_i^{(k)} - \theta_j^{(k)}\right) \\ &= B_{ij}^{(k)} + R_{ij}^{(k)}. \end{aligned} \tag{2}$$

The overall analyst-audience alignment for a given data analysis is then characterized by the collection $D_{ij}^{(k)}$s for the entire set of $K$ principles $\mathbf{D}_{ij} = \left(D_{ij}^{(1)}, \ldots, D_{ij}^{(K)}\right)$.

In Equation (2) above, $R_{ij}^{(k)} = \phi_i^{(k)} - \theta_j^{(k)}$ represents the "residual alignment adjustment" for principle $k$ made by the analyst and the audience after considering their baseline difference.

### 3.2 Measures of the Strength of Alignment of a Data Analysis

In this section we propose ways to measure the alignment of a data analysis pairwise between the analyst $i$ and audience $j$. For example, you could set some maximum distance, $\varepsilon$, and require that all distances, $D_{ij}^{(k)}$, are less than that. Formally, we could refer to this as *Strong Pairwise Alignment* (Definition 3). In constrast, you could consider whether the average distance across all principles is less than some determined distance, $\varepsilon$. This would be a weaker form of analysis alignment that allows for some differences in how the principles are allocated, but places a limit on the total variation of those differences; we refer to this as *Weak Pairwise Alignment* (Definition 4).

**Definition 3** (Strong Pairwise $\varepsilon$-Alignment)**.** A data analysis is strongly $\varepsilon$-aligned between the pairing of analyst $i$ with audience $j$ if for some small $\varepsilon > 0$,

$$\left\|\mathbf{D}_{ij}\right\|_\infty = \max_{k=1,\ldots,K} \left|D_{ij}^{(k)}\right| < \varepsilon.$$

The definition of strong pairwise alignment requires that the differences are never too large for any given principle.

**Definition 4** (Weak Pairwise $\varepsilon$-Alignment)**.** A data analysis is weakly $\varepsilon$-aligned between the pairing of analyst $i$ with audience $j$ if for some $p \geqslant 1$ and small $\varepsilon > 0$,

$$\left\|\mathbf{D}_{ij}\right\|_p = \left(\frac{1}{K}\sum_{k=1}^{K}\left|D_{ij}^{(k)}\right|^p\right)^{1/p} < \varepsilon. \tag{3}$$

With this definition, the analyst and audience may differ slightly with respect to how each principle is allocated, but the overall differences between analyst and audience must be small. The choice of $p$ here (and hence, the norm) will have an impact on how much deviation is allowed between analyst and audience and how much any single principle may differ.

### 3.3 Analyst–Audience Scenarios

In this section we present some examples of the implications of aspects of our model for analytic alignment on the design of data analyses. We present a series of scenarios to demonstrate the models discussed above. First, we will demonstrate three profiles of analyst/audience pairs: (1) accommodating analyst—intransigent audience, (2) intransigent analyst—accommodating audience, (3) design-focused analyst—design-focused audience. Assume for the moment that $N = M = 1$ so that we have a single analyst and single audience member. In all three scenarios, we reach alignment in expectation.

In the first scenario, the driving force is the intransigent audience, meaning they are inflexible with respect to updating their baseline expectations, in other words, $\theta_j^{(k)} = 0$ for $k = 1, \ldots, K$. If the analyst is completely accommodating, meaning alignment is always achieved, $\phi_i^{(k)} = -B_{ij}^{(k)}$ for $k = 1, \ldots, K$. Similiarly, in the second scenario, the analyst is intransigent and therefore $\phi_i^{(k)} = 0$ for $k = 1, \ldots, K-1$ and $\theta_j^{(k)} = -B_{ij}^{(k)}$ for $k = 1, \ldots, K$.

In the third scenario, we describe both the analyst and audience as *design-focused*, by which we mean they are both motivated to improve the alignment of the data analysis and are willing to negotiate over principle allocations. In this scenario, we can describe bounds for $\phi_i^{(k)}$ and $\theta_j^{(k)}$. In absolute value, the analyst's post-resolution adjustment coefficient, $\phi_i^{(k)}$, is bounded by $\left|-B_{ij}^{(k)}\right|$ and 0 with the qualification that the strict equality with 0 only occurs when alignment is achieved at baseline. Likewise, in absolute value the audience's post-resolution adjustment coefficient, $\theta_j^{(k)}$, is bounded by $\left|-B_{ij}^{(k)}\right|$ and 0, again with the qualification that the strict equality with 0 only occurs when alignment is achieved at baseline.

## 4 Case Study

In this section we describe a case study of analytic alignment. The data were collected from 26 students enrolled in a undergraduate capstone course at Wake Forest University titled *Seminar in Mathematical Business Analysis*. This course is designed for seniors who have majored in a joint program between the Department of Statisitcal Sciences and the Business School. The study was approved by the Wake Forest University Institutional Review Board (IRB00025295).

Students in the class were assigned to one of seven groups, each tasked with completing a data analysis for an external client. The analysis questions that each of the groups addressed, as well as information about the client, are shown in Table 1. The students were taught six principles for designing a data analysis (D'Agostino McGowan et al., 2022) (clarity, exhaustive, data-matching, reproducibility, second order, and skeptical) and given an initial problem statement from their assigned client. The concept of alignment between the students (analysts) and the audience (clients) was discussed in class. Each of the 26 students were first asked to allocate each of the six design principles in the context of their assigned data analysis task to describe how much relative time they would allocate to each one, resulting in baseline analyst allocations of $\hat{V}_i^{(k)}$ for all $i = 1, \ldots, 26$ and $k = 1, \ldots, 6$. Subsequently, in class, students discussed these baseline allocations with their group, estimating the *group-specific* allocation for each principle. These two steps fall in the *baseline* stage of Figure 1.

The student groups were then asked to hold a meeting with their client where they discussed the proposed principle allocations and received input on the client's expectations (the *analytic negotiation*). These discussions were held outside of class. Finally, the students reported a post-negotiation set of allocations, taking into account both their initial allocation and the clients

Table 1: Group analysis questions and clients.

| Group | Analysis Question | Client |
|---|---|---|
| 1 | Which county-level, socioeconomic variables are most highly associated with higher rates of alcohol-related deaths? | Researcher |
| 2 | How much of an impact does a broken-down ride have on the wait times of other rides in a Disney World park? | Data scientist |
| 3 | How accurately do first-year students predict the major they ultimately graduate with? | University administrator |
| 4 | Create a dashboard detailing grant support for an institute in the medical school | Grant manager |
| 5 | What are the factors that contribute to a university's peer ranking score in the US News yearly report? | University administrator |
| 6 | Which variables are most influential for predicting Space Mountain wait time at Disney World? | Data scientist |
| 7 | What demographic information and clinical values predict whether a child has primary or secondary hypertension? | Clinician |

expectation, in the *resolution* stage, i.e. $\hat{V}_i^{(k)} + \hat{\phi}_i^{(k)}$. We have included the resources used for this class activity in the Supplementary Materials.

Figure 2 displays the principle allocation results for all students across all groups and principles. Each row of panels represents a student group and each column of panels represents a specific principle. Within each panel, there are two sets of points representing the students' allocations for each principle. The left group of points represents the students' individual allocations before meeting the client and the right point (connected via the lines) represents the group's collective allocation after meeting the client. In the left group of points, the blue circle represents the group's average pre-negotiation principle allocation.

The data in Figure 2 show variability both within and between groups, indicating differences in how each group and its members perceived the importance of the design principles before and after client negotiations. Assuming each group represents its member's "field", the differences within group at baseline are captured by the left group of points in each panel. The differences between groups, as indicated by the differences in the group means, may represent the differences in the analyses themselves, as each group was assigned a different analysis task.

Table 2 shows the change in each group's average principle allocations between the baseline stage and the resolution stage. We see that most groups updated their allocations post-negotiation by either increasing or decreasing them, potentially improving pairwise alignment. For example, Group 5 decreased its allocation to the matching principle ($-0.048$) after the negotiation phase but increased its allocation to the skeptical principle (0.038).

Due to the nature of this consulting class project, we assumed $\theta_j^{(k)} = 0$ for $j = 1, \ldots, 7$ and for each of the principles. That is, only the students would update their allocations in the resolution stage; the seven clients' baseline and resolution allocations would remain the same. This is an example of the "accomodating analyst-intransigent audience" scenario described in Section 3.3. Upon examining the data, it seems possible that groups 3 and 6 are a reflection of the "intrasigent analyst-accomodating audience" scenario, as there were no changes between

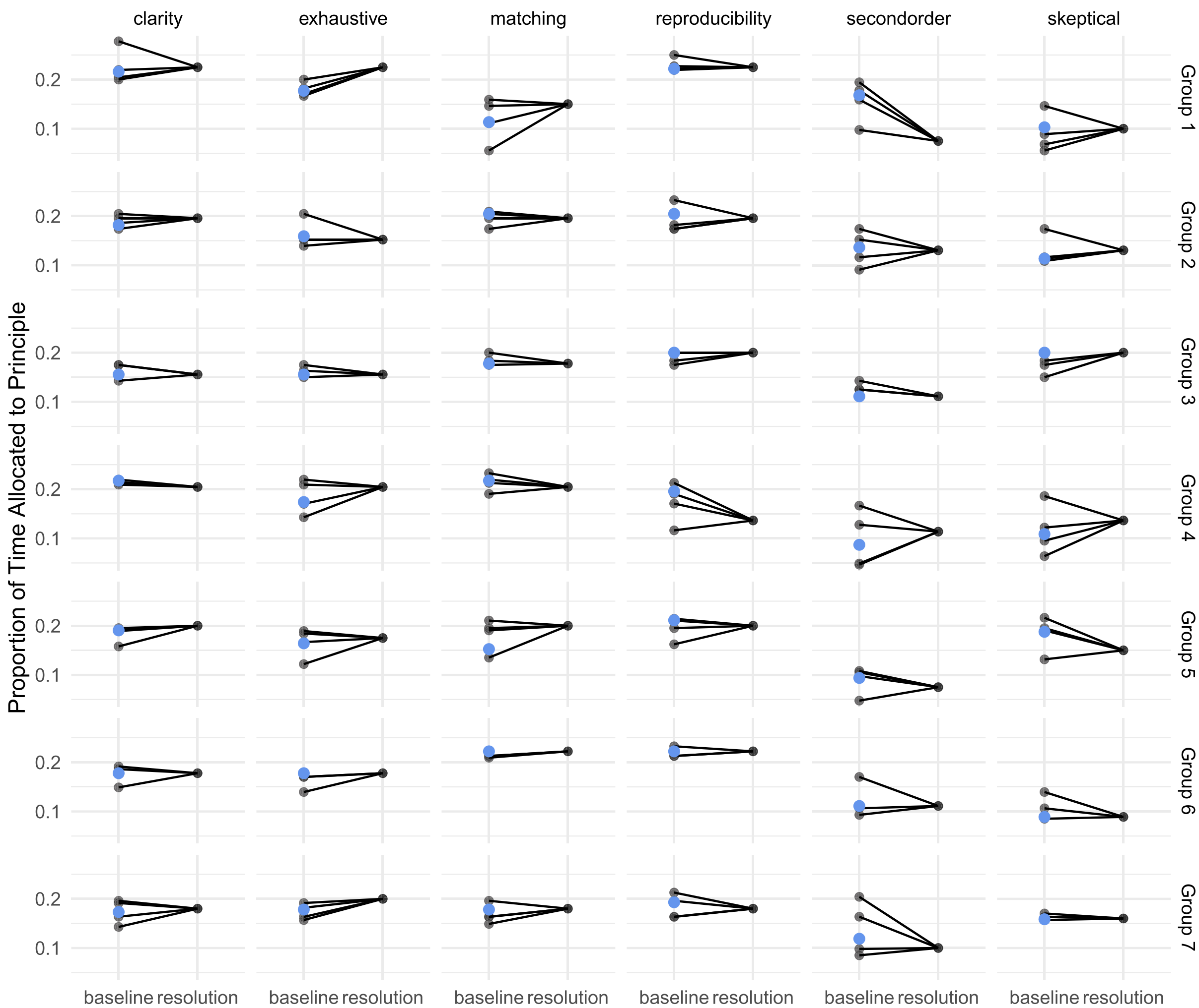


Figure 2: Principle allocations before and after analytic negotiation. Each point represents the allocation assigned to the given principle by the analyst pre-negotiation ($V_i^{(k)}$) and post-analytic negotiation ($V_i^{(k)} + \phi_i^{(k)}$). The units on the $y$-axis represent the proportion of time allocated by the analyst to that particular principle. Each line connects an individual's pre- and post-analytic negotiation principle allocation (i.e. each line represents a single analyst). The blue point represents the agreed upon group-specific allocation, as described in Section 2.

their baseline and post-negotiation principle allocation values. It is worth noting that these two groups worked with the same client for their project.

# 5 Discussion

In this paper, we propose a framework for characterizing the alignment of an analyst and an audience in the development of a data analysis and demonstrated its use in a teaching case study. This framework consists of a set of principles introduced earlier (D'Agostino McGowan et al., 2022) and a model that explains potential variation in the allocation of these principles for a given analysis. We define the alignment of an analysis as the matching of allocations

Table 2: Change in average principle allocation between baseline stage and resolution stage, by group.

| Group | Matching | Exhaustive | Skeptical | 2nd Order | Clarity | Reproducibility |
|---|---|---|---|---|---|---|
| Group 1 | −0.036 | −0.048 | 0.003 | 0.093 | −0.009 | −0.003 |
| Group 2 | 0.009 | 0.007 | −0.017 | 0.006 | −0.014 | 0.009 |
| Group 3 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Group 4 | 0.013 | −0.031 | −0.028 | −0.027 | 0.013 | 0.059 |
| Group 5 | −0.048 | −0.011 | 0.038 | 0.019 | −0.009 | 0.011 |
| Group 6 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Group 7 | −0.002 | −0.022 | −0.002 | 0.019 | −0.007 | 0.013 |

across data analytic principles between the analyst and audience. Alignment can be obtained via negotiation between the analyst and audience between the baseline and resolution stages of analytic planning. In the development of machine learning models based on information learned from data analysis, building in alignment with the appropriate audience can serve as an element in getting users to adopt new technologies.

The case study described in Section 4 illustrates how elements of our model could be mapped to numerical quantities that can be used to track analysts' relative allocations to different principles. The case study describes a classroom project where students developed baseline principle allocations, conducted a negotiation with the client/audience, and then updated their principle allocations in the post-negotiation phase. The data from this case study indicate a few interesting aspects of our framework. First, from Figure 2, the wide variation in the individual principle allocations at the baseline stage indicates that individual students were not necessarily aligned with the client's interests before meeting with them. For example, Group 4's baseline allocations for the second order principle ranged from 0.05 to nearly 0.2. We can therefore infer that discussion with the client and amongst the group members had some impact on the members' principle allocations. Second, Table 2 shows that some groups, on average, had to make much bigger changes in their principle allocations than other groups, suggesting that some groups began with much greater baseline misalignment. For example, Group 1 had to make some large changes to their allocations for data matching, exhaustive, and second order. The data in Figure 2 and Table 2 present numerical evidence that initially, analysts can be far apart from each other regarding which aspects of an analysis deserve more emphasis than others. An area for future work might be to develop a better understanding of this baseline misalignment and identifying what predictors might explain this variation. Section 2 proposes some potential factors for both the analyst and the audience that may be worth considering.

Considerable literature to date has documented the lack of replication of significant findings in the medical and social sciences (Wen et al., 2018; Coiera et al., 2018; Coiera and Tong, 2021; Valentine et al., 2011; Open Science Collaboration, 2015; Dreber and Johannesson, 2019). The proposed reasons for this replication crisis include publication bias, perverse incentives, questionable research practices, and insufficient education (Edwards and Roy, 2017; Nosek et al., 2012; Van Aert et al., 2019; Franco et al., 2014; Moonesinghe et al., 2007; Artino Jr et al., 2019; Simmons et al., 2011; Gigerenzer, 2018). Recommendations for addressing this problem range from increasing training to constraining research degrees of freedom (Gelman and Loken, 2014; Nosek et al., 2018) to outright banning of specific analysis practices (Schirm et al., 2019;

Hand, 2022). While some of the proposed recommendations address the human element to doing data analysis, none of them recognizes the role of the audience in the design of data analyses. The literature on statistical consulting stresses the importance of communicating statistics to clients (Cabrera and McDougall, 2013) and the need to develop an understanding of the underlying nature of the scientific question (Hand et al., 2007). While the concept of alignment seems clearly embedded in these recommendations for consulting, no explicit framework for alignment is proposed.

A key point that we raise in this paper is the need for analyses to be designed properly in order to characterize their usefulness. Because the usefulness of an analysis varies between potential audiences of an analysis, there is a need to involve the audience in the design of the analysis itself. Seasoned practitioners of data analysis will likely recognize this aspect of data analysis, as often there is an initial consultation that occurs where the specific requirements of an analysis are negotiated (Kimball, 1957). We propose that such requirements can be driven by an underlying set of design principles and that negotiating agreement on how to allocate those principles to a given analysis is important for making an analysis useful. The data presented in Section 4 suggest that the alignment framework can serve as a teaching tool for students learning about the data analysis process. A broader point that we raise here is the value of externalizing aspects of a data analysis that might otherwise be implicit in the analysis process. Having a concrete representation of the alignment between analyst and audience provides a reference for future iterations of the analysis or for future analysts that might address the same problem.

The presentation we have taken is formal, mathematically speaking, because we believe the formalism can lead to valuable insights. For example, it allows for a precise statement of what it means for an analysis to be aligned between audience and analyst. Looking to the future, this mathematical formalism could provide a road map for developing approaches to estimating parameters we have specified from data, such as data analytic reports or papers. For example, the *tidycode* package (D'Agostino McGowan, 2019) already allows for the automatic processing and classification of R code into data analysis activities using crowd-sourcing or pre-specified lexicons. Additionally, formalizing the mathematical framework for negotiation between analyst and audience allows us to leverage tools such as artificial intelligence, both for wide-spread pedagogical use as well as the study of how analysts and audiences interact in various environments. Our hope is that the mathematical model presented here serves as a solid foundation on which to build future knowledge about data analysis.

This framework suggests a potential mechanism to investigate the ways in which the development of a data analysis could fail. Specifically, a lack of alignment between analyst and audience could be an important failure mode for a data analysis and that consideration of alignment in the early stages could help to clarify the requirements of an analysis. Although there are numerous descriptions of failed data analyses (Baggerly and Coombes, 2009), specific definitions of how data analyses can fail, with detailed discussions of potential root causes, are lacking in the literature. Learning from failed data analyses is an important aspect of the training of any data analyst and the first step in that process is identifying when failure has occurred. Dialog between the analyst and the audience about why a data analysis has failed can improve the quality of future analyses, as well as improve the quality of the relationship between analyst and audience. Critical to such "post-mortem" discussions is that it be conducted in a blameless manner (Parker, 2017) so that analyst and audience can quickly come to a resolution over how problems should be fixed.

Our approach to characterizing data analysis failures shares many elements with the field of design thinking in its approach to building a solution matched to a specific audience (Cross, 2011). In some ways, one could think of a data analysis as a kind of "product", in the sense that it is not a naturally occurring object in nature. As such, someone—the analyst—must design the analysis in a manner that makes it useful to the audience, or is aligned with the audience's expectations and needs, much like any designed product. While the audience could be one individual or a group of individuals, each individual audience member plays a critical role in evaluating the quality of a given data analysis. Each audience member evaluates the quality with their own preconceived notions, characteristics, and biases towards valuing what makes a good or bad analysis (Wild and Pfannkuch, 1999).

Our work overlaps with the expanding literature on best practices in statistical consulting. For example, Maimone et al. (2024) discuss the skills needed for data science consulting teams, including making tacit knowledge explicit, having difficult conversations (for example, with collaborators), and learning for diverse experiences. Successfully negotiating alignment between analyst and audience likely requires each of these skills. Rubio et al. (2011) discuss metrics used by the Biostatistics, Epidemiology, and Research Design cores of the National Institutes of Health's Clinical and Translational Science Awards. The metrics cover a range of domains including "development and maintenance of collaborations with clinical and translational science investigators." The framework of alignment described in this paper could be adapted as a collaboration metric to monitor the progress of statistical consultations.

Our definition of alignment in data analysis depends solely on the participants—the analyst and the audience—and the outputs of the data analysis. In theory, one could calculate the pairwise alignment of an analysis with just those elements. Critically, we do not consider events or information that occur outside the analysis or perhaps in the future. For example, an analysis may make certain conclusions based on the evidence available in the data that are later invalidated by more in-depth analysis (perhaps with better data). We do not therefore conclude that the original analysis was by definition a failure. At any given moment, an analysis can only draw on the data and evidence that are available. It therefore seems inappropriate to judge the alignment of a data analysis based on information that were not accessible at the time.

## Supplementary Material

In the supplementary materials we provide the lecture slides used for the case study and the code and data used for the analysis in Section 4.

## Funding

The authors do not have any funding to acknowledge.

## References

Artino Jr AR, Driessen EW, Maggio LA (2019). Ethical shades of gray: International frequency of scientific misconduct and questionable research practices in health professions education. *Academic Medicine*, 94(1): 76–84. https://doi.org/10.1097/ACM.0000000000002412

Baggerly KA, Coombes KR (2009). Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *Annals of Applied Statistics*, 3(4): 1309–1334.

Broderick T, Gelman A, Meager R, Smith AL, Zheng T (2023). Toward a taxonomy of trust for probabilistic machine learning. *Science Advances*, 9(7): eabn3999. https://doi.org/10.1126/sciadv.abn3999

Cabrera J, McDougall A (2013). *Statistical Consulting*. Springer Science & Business Media.

Coiera E, Ammenwerth E, Georgiou A, Magrabi F (2018). Does health informatics have a replication crisis? *Journal of the American Medical Informatics Association*, 25(8): 963–968. https://doi.org/10.1093/jamia/ocy028

Coiera E, Tong HL (2021). Replication studies in the clinical decision support literature–frequency, fidelity, and impact. *Journal of the American Medical Informatics Association*, 28(9): 1815–1825. https://doi.org/10.1093/jamia/ocab049

Cross N (2011). *Design Thinking: Understanding How Designers Think and Work*. Berg.

D'Agostino McGowan L (2019). *tidycode: Analyze Lines of R Code the Tidy Way*. R package version 0.1.1.

D'Agostino McGowan L, Peng RD, Hicks SC (2022). Design principles for data analysis. *Journal of Computational and Graphical Statistics*, 32(2): 754–761.

Dreber A, Johannesson M (2019). Statistical significance and the replication crisis in the social sciences. In: *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Edwards MA, Roy S (2017). Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science*, 34(1): 51–61. https://doi.org/10.1089/ees.2016.0223

Franco A, Malhotra N, Simonovits G (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203): 1502–1505. https://doi.org/10.1126/science.1255484

Gelman A, Loken E (2014). The statistical crisis in science. *American Scientist*, 102(6): 460–465. https://doi.org/10.1511/2014.111.460

Gigerenzer G (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, 1(2): 198–218. https://doi.org/10.1177/2515245918771329

Hand DJ (2022). Trustworthiness of statistical inference. *Journal of the Royal Statistical Society. Series A. Statistics in Society*, 185(1): 329–347. https://doi.org/10.1111/rssa.12752

Hand DJ, Everitt BS, Everitt B (2007). *The Statistical Consultant in Action*. Cambridge University Press.

Hicks SC, Peng RD (2019). Elements and principles of data analysis. arXiv preprint, 1–13.

Kimball A (1957). Errors of the third kind in statistical consulting. *Journal of the American Statistical Association*, 52(278): 133–142. https://doi.org/10.1080/01621459.1957.10501374

Maimone C, Sharp JL, Schwartz-Soicher O, Oliver JC, Beltran L (2024). Do good: Strategies for leading an inclusive data science or statistics consulting team. *Statistica*, 13(2): e687. https://doi.org/10.1002/sta4.687

Mira A, Wit E (2021). The capstone in everyone's delivery room: Placing 'practice'at the center of data science education. *Harvard Data Science Review*, 3(1). https://doi.org/10.1162/99608f92.539432b5

Moonesinghe R, Khoury MJ, Janssens ACJW (2007). Most published research findings are false—but a little replication goes a long way. *PLoS Medicine*, 4(2): e28. https://doi.org/10.1371/journal.pmed.0040028

Nosek BA, Ebersole CR, DeHaven AC, Mellor DT (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11):

2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek BA, Spies JR, Motyl M (2012). Scientific utopia: Ii. restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7(6): 615–631. https://doi.org/10.1177/1745691612459058

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). aac4716.

Parker H (2017). Opinionated analysis development. *PeerJ Preprints*, 5: e3210v1.

Peng RD (2011). Reproducible research in computational science. *Science*, 334(6060): 1226–1227. https://doi.org/10.1126/science.1213847

Rubio DM, Del Junco DJ, Bhore R, Lindsell CJ, Oster RA, Wittkowski KM, et al. (2011). Evaluation metrics for biostatistical and epidemiological collaborations. *Statistics in Medicine*, 30(23): 2767–2777. https://doi.org/10.1002/sim.4184

Schirm A, Lazar N, et al. (2019). Moving to a world beyond "$p < 0.05$". *American Statistician*, 73(sup1): 1–19. https://doi.org/10.1080/00031305.2019.1583913

Silberzahn R, Uhlmann EL, Martin DP, Anselmi P, Aust F, Awtrey E, et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1(3): 337–356. https://doi.org/10.1177/2515245917747646

Simmons JP, Nelson LD, Simonsohn U (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11): 1359–1366. https://doi.org/10.1177/0956797611417632

Tukey JW (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1): 1–67. https://doi.org/10.1214/aoms/1177704711

Tukey W, Wilk MB (1966). Data analysis and statistics: An expository overview. In: *Proceedings of the November 7–10, 1966, Fall Joint Computer Conference*, 695–709.

Valentine JC, Biglan A, Boruch RF, Castro FG, Collins LM, Flay BR, et al. (2011). Replication in prevention science. *Prevention Science*, 12: 103–117. https://doi.org/10.1007/s11121-011-0217-6

Van Aert RC, Wicherts JM, Van Assen MA (2019). Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PLoS ONE*, 14(4): e0215052. https://doi.org/10.1371/journal.pone.0215052

Wen H, Wang HY, He X, Wu CI (2018). On the low reproducibility of cancer studies. *National Science Review*, 5(5): 619–624. https://doi.org/10.1093/nsr/nwy021

Wild CJ, Pfannkuch M (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3): 223–248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Yu B, Barter RL (2024). *Veridical Data Science: The Practice of Responsible Data Analysis and Decision Making*. MIT Press.

Yu B, Kumbier K (2020). Veridical data science. *Proceedings of the National Academy of Sciences of the United States of America*, 117(8): 3920–3929. https://doi.org/10.1073/pnas.1901326117