man

Z. Palmore

2022-11-29

About

This manual focuses on reading and making sense of the times data set. It is intended to help guide the reader through the process of analysis to gain insights from the data. It is a living document updated as changes occur. A breakdown of times contents and research plans are provided.

The data dictionary contains a definition and description of every field recorded. The timeline contains a chronological series of events that influenced or directly changed the layout of the fields or data set's purpose. The research section details, in chronological order, what studies were conducted a brief description the research goals. When applicable, the outcomes or intended outcomes are also described.

Data Dictionary

This section contains definitions and descriptions of every field recorded in the times. It does not track changes over time although there may be special notes about how a field was defined before or after a significant event. All significant events are recorded in the timeline.

Date

A date object that represents the day that the project and takes the form of a single digit month, day, and four digit year, for example m/dd/yyyy. This field begins at 3/21/2022.

Start

The start time of the task for the project in the format of h:mm (single digit hour : two digit minutes) followed by an AM or PM symbol to indicate morning or night. Although working hours generally consist of the hours between 7:00 AM and 4:00 PM, there are deviations outside of these boundaries.

End

The end time of the task for the project in the format of h:mm (single digit hour : two digit minutes) followed by an AM or PM symbol to indicate morning or night. Although working hours generally consist of the hours between 7:00 AM and 4:00 PM, there are deviations outside of these boundaries.

Project

The name of the project as provided at the start of the project. It is important to note that the name is most often given by others who may not be performing any work on the task.

This field is subject to change if other individuals start to refer to the project as something new. This change happens when others refer to the project often enough in a new name that it would be confusing to recording. Otherwise, the project is given an new name because it constitutes a new project.

The values in this field take the form of any characters or numbers.

Subproject

In some cases there are projects that have multiple phases or components that are large enough to be considered a project in-and-of-themselves. This field provides the project name of those phases or components and each fits within a larger project.

The values in this field take the form of any characters or numbers.

Notes

Description of what work was performed during the date between the start and end times. This does not have to be specific. In general, descriptions are kept as brief as possible while also trying to effectively describe the bulk of the work performed.

Timeline

Origins

The origin of times tracking is traced to the start of employment for the data scientist position. As written in Times: A Reproducible Guide to Measure and Quantify Time Estimates of Projects the current methods for tracking and estimating time spent on projects and other staff activities are inadequate. The goal of the times tracking is to know what people spend their time on and to be able to bill grants appropriately. Currently, this is not possible to do accurately under any method.

In short, there are 3 main methods for tracking times including a method developed by human resources (HR), another by the information technology (IT) department known as Dailies, and the In-Out Health Board. The HR method is the most commonly utilized since, it is how staff get paid. Everyone at some point fills out a timesheet. However, it is provides no details on the work the staff member performs. The Dailies method was developed to fix that but seems to consistently fail to deliver. The main limitation is that staff are unaware of how to bill to each grant. Lastly, the In-Out Health Board was developed as a tracker to know when staff are in the office or out in the field. This had no intention of being used to track time but largely functions the same as Dailies. These each have a multitude of limitations not listed here and that influence the quality of the information greatly due to method setup.

The report provided the reasoning for development of a reproducible measure to quantify and estimate time spent on various projects. There are no method capable of collecting reasonably accurate data and no standards. The definition, for example, of what counts as staff time is not agreed upon and the individuals who openly disagree on what constitutes staff time each track their times individually. Attempts at gaining clarity on these limitations have not been successful. Entry is performed by hand and is intended to track time in 15 minute intervals but can often be delayed for months or not tracked by staff. For further details and justification for why this times data set was needed, see the full report.

From his own words on the presentation of the yrbs data. He said coming up with data and the visuals for the data then the nurses take the data and tackle how we're using this information. I don't want to go out and volunteer to present to the board. Given how things are going with the board, I don't want [to]. He was having a conversation with Jen Wedge and another nurse or public health strategist. He said Katrina had said she wanted this presented to the board at some point but he again stated he did not want to. Someone likely asked about what if they ask directly for this to be presented to the board or the board asked, to which he said, we'll cross that bridge when we get there but he is not going to go out and volunteer with the way some people are treating members of the department whether they are there or not there. Curious what would be most important to talk about beyond the substance use and data that is related to absenteeism. We could look at general topics and he listed a bunch, to say, "given all of that, I want to pull out 4-5 topics that we could share with the school health collaborative. But, he does not know what that group is interested in. The nurses (middle-aged women) gave the response that driving was important. To which they are trying to emphasize things like seat belt wearing and vehicle safety to prevent accidental deaths and other forms of vehicle accidents among young people. Someone mentioned the idea of sexual behavior and there was a single question on the use of"profelaxus." Someone was trying to link food insecurity with disparities and probably

sex riskiness. There is a massive focus on the data visualization with no regard to the underlying process of how the data got to that point. He tried to use a visualization to show that the pandemic has led to much higher screen time. This was one thing he thought was a big change from the previous cycle to this cycle. There was a 25% increase in the rate of non-school work screen time. He thinks the lack of being able to send kids to school led to the increase in screen time and that maybe led to their decline in performance? It was not clear. He did not know whether tiktok was around in 2018-2019. That this could be telling in terms of those apps and being "extremely addictive." Decrease in use of marry Jane. from 1/3 to about 1/4 but he doesn't know what that's about really. They have a confidence interval that they provide but basically he did not know how to read or use this he just kept repeating I don't know. Looks like it indicates sexual activity is decreasing overall so that's good. He wondered if there was a missing question that he forgot to include because he put this together rather quickly. The rest of the conversation just spiraled downhill.

On March 14, 2022 the assistant director of the department requested that the data scientist track the time spent on projects and tasks throughout the day. The data scientist is the subordinate of the assistant director and it was expressed that the time tracking would be used to present to the director how time was being spent. A new system was needed because the assistant director clearly had no way of accessing or tracking the time of the data scientist and at best, current systems are inaccurate and costly measures for estimating times which do no get down to position-specific time tracking. Thus, the report was developed and a spreadsheet started.

Initial Setup

On March 21, the first version of times (a tracking spreadsheet) for the data scientist position included the fields: Date, Start, End, Project, Subprojects, and Notes. In discussions with the assistant director after March 14 2022, it was confirmed that these fields captured the parameters of interest. There was no disagreement with this initial setup. The assistant director encouraged this form of tracking in addition to entering times in Dailies, the In-Out Health Board, and for HR. Although the data scientist started on February 7, 2022 recording did not begin until March 21, 2022. Anything prior to March 21, 2022 was ignored. There was simply no way to know what had happened since the data scientist's start date without guessing at times. With that, it is important to note several prominent projects that were ignored from times but took a substantial amount of time, here defined as several days to weeks worth of work.

The first major project was originally, not thought of as a project at all. Within the first week of joined the Health Department staff, the data scientist was asked to ask other staff what they had for me to do and find out about the projects available for the data scientist. Taking initiative, the data scientist added responses from the staff based on an example within the orientation workbook that was provided by the assistant director and director. This was intended to foster an interest in the work of the data scientist, provide transparency, and had been discussed and approved of as a nice idea from the assistant director. After speaking to every supervisor, various other staff selected by the data scientist based on the work they performed, and other members of the staff deemed important enough by the assistant director to put more people in front of the data scientist, a list of over 41 projects was developed.

This changed the strategy for tracking time by changing the reasoning behind why time should be tracked in the first place. It changed from a method to gain an understanding of something to a method to simply take up time. The project list was later called upon by the assistant director throughout the summer to reference while meeting with the epidemiologist and during regular check-ins with the data scientist. It was used to provide details on what each of the epidemiologist and data scientist roles were going to do on any project that was a 'true' project and identify gap areas. What a 'true' project was remained undocumented. It was at the discretion of the staff member who chose to work on it to define whether it was a true project or "pie in the sky idea" as defined by the assistant director, except in cases of disagreement, where no decisions meant putting any project ideas on hold until further notice.

Tracking time on projects is difficult with no definition of what constitutes time spent on a project or what to define the project as. For the assistant director, the list of projects was a tool used to push the data scientist to do projects even when the assistant director had no idea what the project entailed or who to reach out to ensure that the project would be deemed important enough to continue. Instead, we spent months worth of

time discussing what would be worth working on only to defer the decision the staff epidemiologist.

These discussions were not recorded but made up the bulk of time spent from February 7, 2022 until early March, 2022, prior to a decision to move forward on any project. The assistant director stated to the data scientist without the director or epidemiologist around on multiple occasions that it was not clear what other people might want from a data scientist nor was it clear what a data scientist can or should do for the department in general. Additionally, the assistant director added, that it was important to the director that the project come from staff because they might not be ready for a data scientist to perform projects independently.

A contradiction that led to a reformatting of how the times data was to be recorded and to the addition of a sub project field came from the assistant director shortly after hearing from the assistant director that the director preferred projects come from staff. The assistant director made clear, that rather than ask about which projects take priority, strategically planning for the project, or having a discussion of the barriers faced when taking initiative and working on projects, it would be best for the data scientist to work with others independently without input from the assistant director until such input was necessary. Later, we would determine that following this process frequently led to holds being placed on projects by the assistant director after which the assistant director recommended the data scientist reference back to the list but also to remove things from the list that might not constitute a true project using the best judgement of the data scientist. Soon to be recorded times were reformatted again to capture all time related to projects even in cases where the project had been put on hold or was just an idea. However, it was still unclear when the idea became a project.

While the project list was by far the most important and influential component to how times began it was certainly not the only project. There was a project which had approval from the assistant director for an office health study. This took a few weeks and was the assistant director's idea. It was also the assistant director who encouraged the data scientist to reach out to a public health strategist who specialized in mental health to get a study going. After completing a survey, the assistant director informed the director, who expressed disapproval without ever knowing how far along the project had gone. This, and other projects like it, led to a cyclic approval-disapproval process that wasted time from February 7, 2022 until recording began.

May Changes

Management was interested learning what the data scientist was doing, when, and gaining an overall better understanding of how much time the data scientist spends on various projects. However, there were no expectations of how this would be completed other than that the data scientist complete this task alone.

By April 1, 2022 the data scientist recognized management had no intention to review or provide feedback on time. The assistant director expressed interest in having While management claimed to be curious they expressed minimal interest in the setup for

Research Plans

In this section we have research plans for collection and analysis. It is structured such t

Original

This is the first plan as discussed with the assistant director. It is the most basic plan and only considers the variables Date, Start, End, Project, and Notes. However, this plan was quickly made obsolete when subsequent discussions with management where there was no intention to review the data they requested the data scientist collect on the projects performed while in the position.

Objective In this

Time

Conda Version

This is the plan that used the Anaconda State of Data Science and led to the creation of the fields: Primary_Venn_Class, Venn_Class_Heirarchy, Language DataScience_Anaconda_Survey_TimeSpentOn_Area, Job_Category.

Progress Tracking

In this plan we retroactively classify the project as a project as negative, neutral, or positive toward the completion of the project.