Use of Data Journals as Data Investigative Tools

Atri Basu

University of Illinois, Urabana Champagne

ABSTRACT

As a product owner I often find myself needing to express the reasons behind my decisions or the current state of my project/product using data not intuitive to anyone who isn't a core team member. And for the past 12 years that I have been leading projects, I've been trying to figure out the best way to express the narrative of each of my projects. Ship captains and pilots in the olden days kept closely guarded log books that recorded every minutiae of a journey so as to allow easier subsequent travel. I think it would be tremendously helpful for product/project managers/owners to be able to review a similar progression of a successful product/project launch.

Product/project managers are also required to steer their teams and products towards the envisioned destination and the easiest way to help keep everyone on the team aligned is to ensure everyone understands the relevant data they need to be paying attention to and what it indicates about the product/project. Thus the data journal. Project/Product Data Journals should use data visualizations techniques to be:

1. Informative – somebody with no knowledge of the project upon reviewing the journal feels they have an idea of the state of the project.
2. effortlessly expressive – The reader should not need any explanatory guide or training to intuit the information from the data.

In the past I've found annotated graphs to serve really well as a *data explainability tool*, a tool that allows one to explain the inferences that can be reasonably drawn from the data presented to the viewer.

However as a *data investigative tool,* a tool that allows a lay person to draw their own inferences and through a reasonable understanding of the data presented to them, *data explainability tools* are too static, more interactive content is required.

Over the past couple of semesters I've taken two very engaging courses on data visualization as part of my Master's degree in Data Science and been exposed to the inspirational work of Data Investigator extraordinaire Hans Rosling, so I thought I would try to make a *data journal,* a data investigative tool that allows the author to assist the reader in their investigative journey through the data presented in the journal.

The data I've presented in the data journal, while modeled on the data from one of my projects, is merely that – a representation of what a real life project report might look like in the form of a data journal.

Keywords: data investigative tool, data explainability tool, data visualization, data reporting, product management, product ownership, data journal, data journalism.

Index Terms: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 INTRODUCTION

Over the last 12 years I've been a thought leader on a number of projects of varying nature, e.g. I've helped launch, curate and grow a social knowledge management platform, helped establish a new support center in Mexico City, been a significant contributor to a crowd-sourced automation platform for log analysis and more recently I successfully launched my orgs first AI-powered workflow augmentation application. Throughout my journey from ideator to product owner, the single underlying lynchpin of every project/product has been finding the right mix between intuition and data to make the right decision at the right time. However, it doesn't just stop there. For those of us closest to the data, much that seems intuitive and self-explanatory requires deeper exploration and introspection. I have learnt this the hard way both in terms of having misread the data as well as from having to explain my thought process to my project/product stakeholders. E.g., during one particular turbulent period of a product launch, I was finding it difficult to explain the rationale behind the current state of the project and what should be our primary focus, when I drew up the following chart to explain the data as I saw it:

Figure 1. A Data Narrative

Without meaning to rabbit hole this discourse in an *exploration* of this other project, I present this as an example of the kind of visualisations that were met with a lot of alacrity. They created a data narrative that allowed lay individuals, individuals who do not deal with the data minutiae of the project/product, to get their hooks into the stream of data being hurled at them and follow along.

2 THEORY

My theory was that if a simple Data Narrative was so powerful then for people seeking to make decisions on a project/product, a data journal - one that told of the confluence of all the narratives associated with a project/product; whilst also allowing these decisions makers to explore and validate or identify alternate narratives; would be an extremely powerful tool. To prove this

atrib3@illinois.edu

theory I have harnessed anonymized data modeled to mimic real life data from one of my projects in a tableau dashboard titled:

Investigating the impact and adoption of NLP on search @ a multinational

2.1 Is the data journal of any use with obfuscated data?

Yes! The obfuscation of the data was necessary to for data protection and privacy issues, but it also serves the function of focusing the discourse in this paper to the manner of data visualization and presentation rather than the content of it. The data is of placeholder-value only and should be viewed as a means of displaying the various characteristics of the data journal. This is also why the data has not been published anywhere, as it wouldn't serve any purpose to anyone, being fictitious.

2.2 The structure of the data journal

Since I visualised the data jounral as a medium of data investigation to help enable a particular vision for the project/product, I figured it should take a form that seeks to provide answers to pertinent questions. Each page of the data journal should address the most essential needs, the ones most important to fulfil to ensure the success of the project/product. Thus I set about coming up with some broad-strokes qualitative questions; the answers to which I thought helped shaped the narrative I wanted to protray of the project. Each qualitative question should reflect a point of view that provides enough scope for the team to start thinking about solutions which go beyond the staus quo.

Figure 4. "A question well framed is a problem half solved"

So on each page of the data journal, the reader would be presented with one such qualitative question to ponder along with data points to answer more granular and specific quantitative questions that agglomeratively could help answer the qualitative question. Thus each quantitative question needs to express some insight developed through the synthesis of the gathered project/product information that can be leveraged in designing the solution to the qualitative question. Tldr; Metrics! Each quantiative question needed to draw a line in the sand, it needed to be a measurable target that would determine whether or not the some particular intent was succesful. What the quantitative question should ***NOT*** be is a mere "Fun Fact" – a contextless number that obfuscates actual success measurment, e.g. "Over 4000 people have used this tool already!" sure but how frequently. You don't want one of these insights to be quoted by an executive and result in misguided product strategy. These quantitative questions should have the potential to serve as metrics with specific targets for each release – *Key Performance Indicators* for the project/product. They could be things like:

1. Discovery - %age of new unique visitors to a website
2. Acquisition – number of users signing up for a service
3. Engagement - % of Daily Active Users(DAU) that perform a particular action
4. Retention – 7-day, 14-day retention rates
5. ROI – resulting value generation

They could also be ones that validate design decisions. Missing these metrics should mean some adjustment is necessary to the project.

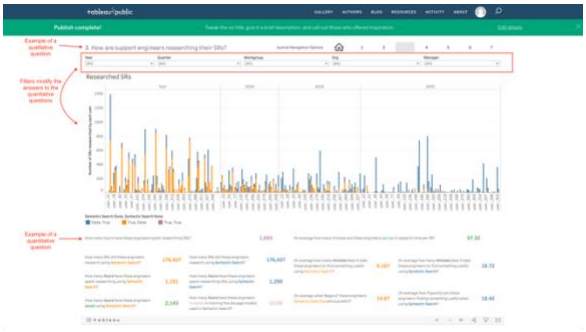

Figure 2. Qualitative vs Quantitative; Filters ftw

However each of these qualitative questions/pages of the data journal must also be woven together into a composite and expressive narrative about the project. To allow the reader to get a clearer idea about my train of thought, I've provided a table of content(see figure 2) and navigation buttons(see figure 4) on each page of the magazine which allows the reader to tab to the relevant section easily and intuitively.

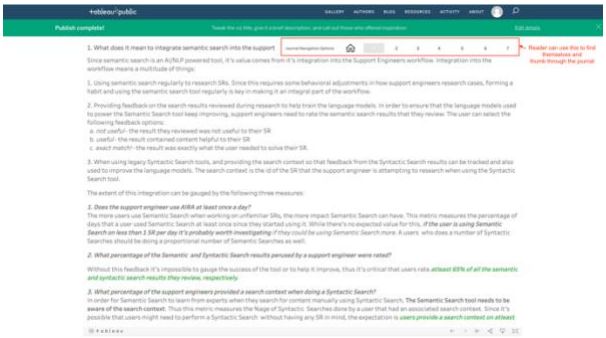

Figure 3. Thumbing through the journal

2.3 The use of filters for self-exploration

Identifying the qualitative questions allows the author of the data journal to initially control the narrative, but the real power of data investigative tool is in it's ability to allow the reader to explore the data more thoroughly and come to their own data-driven answers to the qualitative questions. To do this we have to identify generic and yet pertinent quantitative questions whose answers could help articulate the answer to the overarching qualitative question. For example of filters, refer to Figure 3.

atrib3@illinois.edu

2.4 The dynamic annotations hack – tableau feature request?

Keeping in mind that the data journal is meant for lay users and needs to replace the kind of visualisation I depicted in Figure 1, the jounral also needs to allow me to present my annotations on the readers demand. Interestingly, Tableau doesn't have any straightforward implementation of such a feature. However using some *marks & parameters* magic I was able to add annotations tied to marks on the jounral which would only show up when the reader set a "show annotations" parameter in the data jounral:

Figure 6. Dynamic annotations using markes and parameters

3 Discussion

3.1 How does the data journal help the team?

By visually representing targets for KPIs in the context of the grander vision for a product/project, the data journal does the following:

1. Aligns the team – allowing team members to easily see and understand the metrics tied to a particular release builds focus. It can help resolve design disputes that come from a misalignment of goals. Quantifying what success looks like within tangible time windows(one sprint cycle), gives folks the right context.
2. Builds autonomy – knowing what the goal is and how success is being measured allows the team to independently come up with solutions to the problem and not be reliant on a single source for direction.
3. Introduces accountability – the entire team understands what the product/project is trying to address and this empowers them to think ahead to the next iteration and thus become more accountable towards the overall success of the project.

3.2 The use of tree bar charts

When toying around with the bar charts in Figure 5 above, I discovered that I could create a third dimension to my bars. While this dimension was a low fidelity dimension, by which I mean a dimension of the visualization that could not be used to easily discern between values of the same variable but could be used to visualize trends. It successfully employs a *focus+context approach* - the user's attention is directed toward the focus, but the context is still available in a less emphasized and unobtrusive way. And so instead of using simple stacked bar charts I decided to a create tree bar charts instead. You can see an example of how the additional dimension adds value in Figure 6, below:

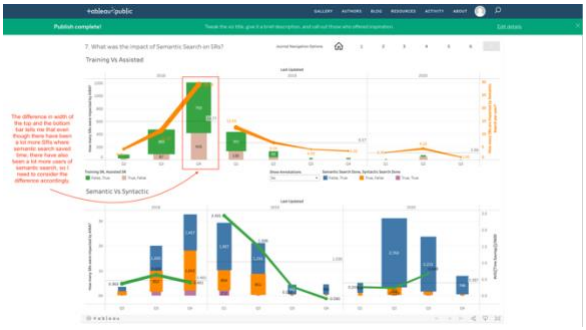

Figure 5. Tree bar charts lending *focus+context*

3.3 Future Considerations

The current state of the data journal is by no means it's final. When I first started on this journey I found it very hard to think from the perspective of a lay individual rather the the most intimately knowledgable person on the project/product. I think that the journal will continue to evolve iteratively with each project/product experience. Here are some of the things I hope to explore in the near future.

3.3.1 Adding dynamic colour commentary to every data point

I want to identify a structured way in which I can come up with generic colour commentary for the answer to each quantitative question. This commentary would serve as tips the reader can use to interpret the data, e.g., the following text could be presented as a tooltip to a reader when they hover over the data point for *Total Research Time* in the data journal:

"If you see the number of total research hours increase without any increase in semantic search hours, this indicates poor adoption of the semantic search tool."

3.3.2 Explore the use of AI to create narratives

When you have a hammer… everything looks like a nail. Well, as someone who has been fully immersed in the business of AI for the past 4 years, it would seem natural that I would want to explore ways that AI, specifically the conjunction of NLP and statistics could be used to automatically generate this kind of narrative. I believe that if the trend of maintaining data journals for project/product reporting catches on then being able to recommend elegant suggestions, similar to the MS Power Point wizard, it would be a much sought after application.

4 Conclusion

I hope in tandem with this paper, the tableau data journal help readers find more powerful and less painstaking ways to create powerful narratives using data investigative tools and, like Prof. Rosling's, Prof Eric Shaffer's and Prof. Josh C. Hart's lectures have done for me, this data journal inspires new trends and techniques in data reporting related to projects/products.

atrib3@illinois.edu

REFERENCES

[1] Rikke Friis Dam and Teo Yu Siang, Define and Frame Your Design Challenge by Creating Your Point Of View and Ask "How Might We", *Interaction Design Foundation,* August 2020

[2] Paul Koullick, How to Set Metrics for Product Launches, *Amplitude Blog,* 9, January 2018

[3] Atri Basu, Investigating the impact and adoption of NLP on search @ a multinational, *Tableau Public,* 26 November, 2020

[4] Chris Love. Adding a Guided Tutorial to your Tableau Visualisation. *The Information Lab*, 2, July 2019.

[5] Hans Rosling. Let my dataset change your mindset. TED, June 2009.

[6] Eric G. Chaffer. Lectures from CS 519 - Scientific Visualization, *UIUC* 2020.

[7] Josh C. Hart. Lectures from Data Visualisation, *Coursera* 2020