No Free Hunch (http://blog.kaggle.com/)



★ (HTTP://BLOG.KAGGLE.COM) > COMMUNICATING DATA SCIENCE: A GUIDE TO PRESENTING YOUR WORK
★ (HTTP://BLOG.KAGGLE.COM/2016/07/06/JUNE-2016-SCRIPTS-OF-THE-WEEK/)
★ (HTTP://BLOG.KAGGLE.COM/2016/06/17/COMPETITION-SCRIPTS-TECHNIQUES-FOR-TACKLING-IMAGE-PROCESSING/)

Communicating data science: A guide to presenting your work

Megan Risdal (http://blog.kaggle.com/author/mrisdal/) | 06.29.2016



See the forest, see the trees. Here lies the challenge in both performing and presenting an analysis. As data scientists, analysts, and machine learning engineers faced with fulfilling business objectives, we find ourselves bridging the gap between <u>The Two Cultures</u>

(https://en.wikipedia.org/wiki/The_Two_Cultures): sciences and humanities. After spending countless hours at the terminal devising a creative and elegant solution to a difficult problem, the insights and business applications are obvious in our minds. But how do you distill them into something you can communicate?

- Team player, willing to do what it takes to make others on team and in other functional areas successful
- Good communicator; able to talk at all levels about deep data science and high level value propositions
- Visionary; able to excite & participate in roadmap discussions regarding how we innovate to excellence
- Hard working/dedicated

(http://blog.kaggle.com/wp-content/uploads/2016/06/job-listing.png)

QUALIFICATIONS AND REQUIREMENTS FOR A SENIOR DATA SCIENTIST POSITION.

Presenting my work is one of the surprising challenges I faced in my recent transition from academia to life as a data analyst at a <u>market research and strategy firm (http://campos.com/)</u>. When I was a linguistics PhD student at UCLA studying learnability theory in a classroom or measuring effects of an oral constriction on glottal vibration in a sound booth, my colleagues and I were comfortable speaking the same language. Now that I work with a much more diverse crowd of co-workers and clients with varied backgrounds and types of expertise, I need to work harder to ensure that the insights of my analyses are communicated effectively.

In this second entry in the communicating data science series

(http://blog.kaggle.com/tag/communicating-data-science/), I cover some essentials when it comes to presenting a thorough, comprehensible analysis for readers who want (or need) to know how to get their work noticed and read.

Get your head in the game

Imagine you've just completed the so-called heavy lifting, whatever it may be, and you're ready to present your results and conclusions in a report. Well, step away from the word processor! There are two things you must first consider: your audience and your goals. This is your forest.

Who is your audience?

The matter of who you're speaking to will influence every detail of how you choose to present your analysis from whether you use technical jargon or spend time carefully defining your terms. The formality of the context may determine whether a short, fun tangent or personal anecdote will keep your audience happily engaged or elicit eye rolls ... and worse.

This is all important to consider because once you've envisioned your audience, you take stock of what may and may not be shared knowledge and how to manage their expectations. In your writing (and in everyday life), it's useful to be cognizant of Grice's principles of cooperative communication:

- 1. Maximum of quantity: be informative, without giving overwhelming amounts of extraneous detail.
- 2. **Maximum of quality**: be truthful. Enough said.
- 3. **Maximum of relation**: be relevant. I'll give you some tips on staying topical shortly!
- 4. **Maximum of manner**: be clear. Don't be ambiguous, be orderly.

So be cooperative! Know your audience and do what you can to anticipate their expectations. This will ensure that you cover all ground in exactly as much detail as necessary in your report.

What is the goal?

Also before you put pen to paper, it's helpful remind yourself again (and again) of what your goal is. If you're working in a professional environment, you're aware that it's important to be continually mindful of the goal or business problem and why you're tasked with solving it.

Or perhaps it's a strategic initiative you're after: Did you set out to learn something new about some data (and the world)? Or have you been diligently working on a new skill you'd like to showcase? Do you want to test out some ideas and get feedback? It's okay to make it your goal to find out "Can I do this?" Maybe you want to share some of your expertise with the community on Kaggle Scripts. In that case, it's even more imperative that you have a buttoned-up analysis!

"If we can really understand the problem, the answer will come out of it, because the answer is not separate from the problem."

— Jiddu Krishnamurti

If you've reached the point of having an analysis to report, you've more than likely familiarized yourself with the goals of the initiative, but you must also keep them at the forefront of your thoughts when presenting your results as well. Your work should be contextualized in terms of your understanding of the research objectives. Often in my own day job this means synthesizing many analyses I've performed into a few key pieces of evidence which support a story; this can't be done well except by accident without keeping in mind the ultimate objective at hand.

The preamble

Now that you've got yourself in the right frame of mind—you can see the forest and you know the trees—you're ready to start thinking about the content of your report. However, before you start furiously spilling ink, first remind yourself of the three elements required to ask an askable question in science:

- 1. The question itself along with some justification of how it addresses your objectives
- 2. A hypothesis
- 3. A feasible methodology for addressing your question

Much as I implore you to consider who your audience is and what your objectives are in order to get your mind in the right place, I'm recommending that you have the answers to these three things ready because they will dictate the content of your report. You don't want to throw everything and the kitchen sink into a report!

What's the question?

On Kaggle, the competition hosts very generously provide their burning questions to the community. Outside of this environment, the challenge is to come up with one on your own or work within the business objectives of your employer. At this point, you make sure that you can appropriately state the question and how it relates to your objective(s).

As an aside, if you need some exercise in the area of asking insightful questions (a skill unto its own), I hereby challenge you to scroll through some of Kagglers' most recent scripts, find and read one, and think of one new question you could ask the author. If you find that this is a stumbling block preventing you from proceeding with your analysis, many dataset publishers include a number of questions they'd like to see addressed. Or read the Script of the Week blogs (http://blog.kaggle.com/2016/06/07/may-2016-scripts-of-the-week/) and see what other ideas script authors would like to see explored in the same dataset.

What's the answer?

Now that you have your question, what do you think the answer will be? It's good practice, of course, to consider what the possible answers may be before you dig into the data, so hopefully you've already done that! Clearly delimiting the hypothesis space at this point will guide the evidence and arguments you use in the body of your report. It will be easier to evaluate what constitutes weak and strong support of your theory and what analyses may be absolutely irrelevant. Ultimately you will prevent yourself from attacking straw men (https://en.wikipedia.org/wiki/Straw man) in faux support of your theory.



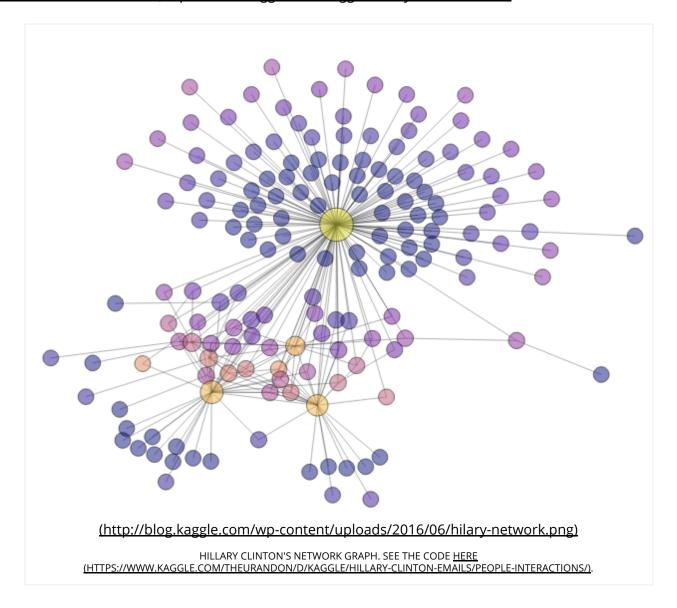
(http://blog.kaggle.com/wp-content/uploads/2016/06/straw-800x.png)
DON'T BUILD STRAW MEN.

What's your methodology?

Let's say you're asking whether Twitter users with dense social networks in the <u>How ISIS Uses Twitter</u> <u>dataset (https://www.kaggle.com/kzaman/how-isis-uses-twitter)</u> express greater negative sentiment than users with less dense networks. Your first step is to confirm that the data available is sufficient to address your research question. If there's major missing information, you may want to rethink your question, revise your methodology, or even collect new data.

If you're unsure of how to put language to a particular methodology, this is a good opportunity to flex your Googling skills. Search for "social network analysis in r" or "sentiment analysis in python." Dive into some academic papers if it's appropriate and see how it's presented. Peruse the <u>natural language</u> <u>processing tags (http://blog.kaggle.com/tag/natural-language-processing/)</u> on No Free Hunch and read the <u>winners' interviews (http://blog.kaggle.com/category/winners-interviews/)</u>. Get inspiration from

scripts on similar datasets on Kaggle. For example, <u>a similar analysis</u> (https://www.kaggle.com/theurandon) was performed by Kaggle user Khomutov Nikita (https://www.kaggle.com/theurandon) using the Hillary Clinton's Emails dataset (https://www.kaggle.com/kaggle/hillary-clinton-emails).



Even if you don't end up needing to share every nuance of your methodology with your given audience, you should always document your work thoroughly to the extent possible. Once you're ready to present your analysis, you'll be capable of determining how much is the right amount to share when discussing the nitty gritty mechanics of your model. Similarly, I've been able to pleasantly surprise my boss many times because I have an answer ready at-hand for immediate questions thanks to keeping my exploratory analyses well-documented.

By the way, if you've felt overwhelmed by the task of putting together a solid methodology for tackling a question, it can't hurt to lob an idea and some code to the community for feedback. Especially once you have solid presentation of analysis skills! Be honest about where you feel you could use extra input and maybe a fellow Kaggler will come forth with different angle on the problem.

Putting the pieces together

Finally, you're ready to write. Keep in mind that a good analysis should facilitate its own interpretation as much as possible. Again, this requires anticipating what information your likely audience will be seeking and what knowledge they're coming in with already. One method which is both tried-and-true and friendly to the academic nature of the discipline is following a template for your analysis. With that, this section covers the structure which when fleshed out will help you tell the story in the data.

Keep in mind that a good analysis should facilitate its own interpretation as much as possible.

Not so abstract

Make it easy for your audience to quickly determine what they're about to digest. Use an abstract or introduction to recall your objectives and clearly state them for your readers. What is the problem that you've set out to solve? If you have a desired outcome or any expectations of your audience, say it, as this is the entire reason you're presenting them with your analysis.

You then cover everything from your preamble in this section: the question you've been on a mission to answer, your hypothesis, and the methodology you've used. Finally, you will often provide a high level summary of your results and key findings. Don't worry about spoiler alerts or boring your readers to death with the content that's about to follow. Trust that if they pay attention past the introduction that they are interested in how you achieve what you claim you have.

Many people I've talked to have said that they often find it easier to write the abstract *after* having already completely documented the detailed findings of the analysis. I think that this is at least in part because you've familiarized yourself with your own work through the lens of your readership by doing so. Slowly but surely you're extracting yourself from the trees and the bigger picture becomes apparent.

The content: Break off what you can chew

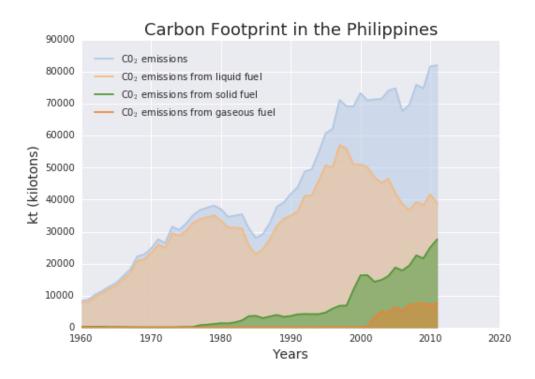
This is where the good stuff lives. You've laid the foundation for your analysis such that your audience is prepared to read or listen intently to your story. I can't tell you the specifics of what goes here, but I can tell you how to structure it.

Take your analysis in small bits by breaking your question into subparts. For a data-driven analysis, it can make sense to tackle each piece of evidence one-by-one. You may have a dissertation's worth of data to report on, but more likely than not you must pick and choose what will best support your analysis succinctly and effectively. Again, having the objectives and audience in mind will help you decide what's critical. Lay it all out before you and pair sub-questions with evidence until you have a story.

Once you've presented the evidence, explain why it supports (or doesn't support) your hypothesis or your objectives. A good analysis also considers alternative hypotheses or interpretations as well. You've already surveyed the hypothesis space, so you should be ready-armed to handle contrary evidence. Doing so is also a way of anticipating the expectations of your audience and the skepticism they may harbor. It's at this point that it's most critical to keep in mind your objectives and the question you're addressing with your analysis. Ask how every piece of evidence you offer takes you one step closer to confirming or disproving your hypothesis.

Other tips and tricks

Visualize the problem. Seeing is believing. It's cliched to say in any statement asserting the value of data visualization, but it's so incredibly true. This "trick" is so effective that I'm going to spend more time talking about it in a future post. If you can plainly "state" something with a graph or chart, go for it!



Variety is the spice of life. And it can liven up your writing (and speaking) as well. For example, use a mix of short and sweet sentences interspersed among longer, more elaborate ones. Find where you accidentally used the word "didactic" four times on one page and change it up! Related to my first point, use effective variety in types of visualizations you employ. Small things like this will keep your readers awake and interested.

Check your work. I don't like to emphasize this too much because I'm a <u>descriptivist</u> (https://en.wikipedia.org/wiki/Linguistic description), but make sure your writing is grammatical, fluent, and free of typos. For better or worse, trivial mistakes can discredit you in the eyes of many. I find that it helps to read my writing aloud to catch disfluencies.

Gain muscle memory. If you really struggle with transforming your analysis into a form that can be shared more broadly, begin by writing *anything* until writing prose feels as natural as writing code. For example, I actually suggest sitting down and copying a report word-for-word. Or even any instance of

persuasive writing. Not to be used as your own in any way (i.e., plagiarism), but to remove one more unknown from the equation: what it literally *feels like* to go through the motions of stringing words and sentences and paragraphs together to tell a story.

Conclusions & Next Steps

A good analysis is repetitive. You know the intricacies of your work in and out, but your audience does not. You've told your readers in your abstract (or introduction, if you prefer) what you had ventured to do and even what you end up finding and the content lays this all out for them. In the conclusions section you hit them with it again. At this point, they've seen the relevant data you've carefully chosen to support your theory so it's time to formally draw your conclusions. Your readers can decide if they agree or not.

Speaking of being repetitive, after making your conclusions, you again remind your readers of the objective(s) of this report. Restate them again and help your readers help you—what do you expect now? What feedback would you like? What decision-making can happen now that your report is presented and the insights have been shared? In my work, I often collaborate with strategists to develop a set of recommendations for our clients. Typically I'll take a stab at it based on the expertise I've gained in working with the data and a strategist will refine using their business insights.

fin

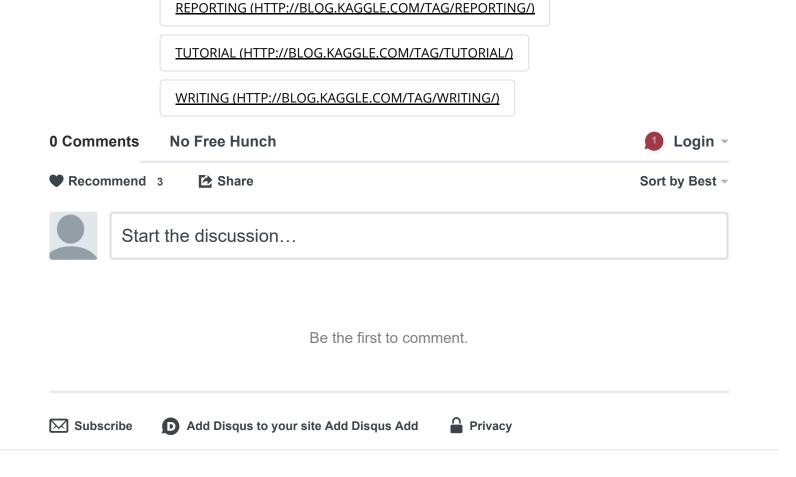
And this is exactly where the beauty of the analysis and your skillful presentation thereof meet. Because you've managed to package your approach in a fashion digestible to your audience, your readers, collaborators, and clients have comprehended and learned from your analysis and what its implications are without getting lost in the trees. They are equipped to react to the value in your work and participate in the next step of realizing its objectives.

Thanks for reading the second entry in this series on <u>communicating data science</u> (http://blog.kaggle.com/tag/communicating-data-science/). I covered the basics of presenting an analysis at a very high level. I'd love to learn what your approach is, how you realize the value in your work, and how you collaborate with others to achieve business goals. Leave a comment or <u>send me a note</u> (mailto:mrisdal@kaggle.com)!

If you missed my interview with Tyler Byers, a data scientist and storytelling expert, <u>check it out here</u> (http://blog.kaggle.com/2016/06/13/communicating-data-science-an-interview-with-a-storytelling-expert-tyler-byers/). Stay tuned to learn some data visualization fundamentals.

COMMUNICATING DATA SCIENCE
(HTTP://BLOG.KAGGLE.COM/TAG/COMMUNICATING-DATA-SCIENCE/)

DATA ANALYSIS (HTTP://BLOG.KAGGLE.COM/TAG/DATA-ANALYSIS/)



> The Official Blog of Kaggle.com

Q Search

■ Categories

DATA SCIENCE NEWS (HTTP://BLOG.KAGGLE.COM/CATEGORY/DATA-SCIENCE-NEWS/) (36)

KAGGLE NEWS (HTTP://BLOG.KAGGLE.COM/CATEGORY/KAGGLE-NEWS/) (118)

SCRIPTS (HTTP://BLOG.KAGGLE.COM/CATEGORY/SCRIPTS/) (22)

TUTORIALS (HTTP://BLOG.KAGGLE.COM/CATEGORY/TUTORIALS/) (28)

WINNERS' INTERVIEWS (HTTP://BLOG.KAGGLE.COM/CATEGORY/WINNERS-INTERVIEWS/) (174

Popular Tags

ALGO TRADING CHALLENGE (HTTP://BLOG.KAGGLE.COM/TAG/ALGO-TRADING-CHALLENGE/)

ANNUAL SANTA COMPETITION (HTTP://BLOG.KAGGLE.COM/TAG/ANNUAL-SANTA-COMPETITION/)

BINARY CLASSIFICATION (HTTP://BLOG.KAGGLE.COM/TAG/BINARY-CLASSIFICATION/)

COMMUNITY (HTTP://BLOG.KAGGLE.COM/TAG/COMMUNITY/)

COMPUTER VISION (HTTP://BLOG.KAGGLE.COM/TAG/COMPUTER-VISION/)

CROWDFLOWER SEARCH RESULTS RELEVANCE (HTTP://BLOG.KAGGLE.COM/TAG/CROWDFLOWER-SEARCH-RESULTS-RELEVANCE/)

DARK MATTER (HTTP://BLOG.KAGGLE.COM/TAG/DARK-MATTER/)

DEEP NEURAL NETWORKS (HTTP://BLOG.KAGGLE.COM/TAG/DEEP-NEURAL-NETWORKS/)

DELOITTE (HTTP://BLOG.KAGGLE.COM/TAG/DELOITTE/)

DIABETES (HTTP://BLOG.KAGGLE.COM/TAG/DIABETES/)

DIABETIC RETINOPATHY (HTTP://BLOG.KAGGLE.COM/TAG/DIABETIC-RETINOPATHY/)

EEG DATA (HTTP://BLOG.KAGGLE.COM/TAG/EEG-DATA/)

ELO CHESS RATINGS COMPETITION (HTTP://BLOG.KAGGLE.COM/TAG/ELO-CHESS-RATINGS-COMPETITION/)

EUROVISION CHALLENGE (HTTP://BLOG.KAGGLE.COM/TAG/EUROVISION-CHALLENGE/)

FACEBOOK RECRUITING (HTTP://BLOG.KAGGLE.COM/TAG/FACEBOOK-RECRUITING/)

FLAVOURS OF PHYSICS: FINDING T → MMM (HTTP://BLOG.KAGGLE.COM/TAG/FLAVOURS-OF-PHYSICS-FINDING-%CF%84-%E2%86%92-%CE%BC%CE%BC%CE%BC/)

FLIGHT QUEST (HTTP://BLOG.KAGGLE.COM/TAG/FLIGHT-QUEST/)

GRASP-AND-LIFT EEG DETECTION (HTTP://BLOG.KAGGLE.COM/TAG/GRASP-AND-LIFT-EEG-DETECTION/)

HERITAGE HEALTH PRIZE (HTTP://BLOG.KAGGLE.COM/TAG/HERITAGE-HEALTH-PRIZE/)

HOW MUCH DID IT RAIN? (HTTP://BLOG.KAGGLE.COM/TAG/HOW-MUCH-DID-IT-RAIN/)

IMAGE CLASSIFICATION (HTTP://BLOG.KAGGLE.COM/TAG/IMAGE-CLASSIFICATION/)

INTEL (HTTP://BLOG.KAGGLE.COM/TAG/INTEL/)

KAGGLE INCLASS (HTTP://BLOG.KAGGLE.COM/TAG/KAGGLE-INCLASS/)

LOGISTIC REGRESSION (HTTP://BLOG.KAGGLE.COM/TAG/LOGISTIC-REGRESSION/)

MARCH MANIA (HTTP://BLOG.KAGGLE.COM/TAG/MARCH-MANIA/)

MERCK (HTTP://BLOG.KAGGLE.COM/TAG/MERCK/)

MULTICLASS CLASSIFICATION (HTTP://BLOG.KAGGLE.COM/TAG/MULTICLASS-CLASSIFICATION/)

NATURAL LANGUAGE PROCESSING (HTTP://BLOG.KAGGLE.COM/TAG/NATURAL-LANGUAGE-PROCESSING/)

OPTIMIZATION PROBLEM (HTTP://BLOG.KAGGLE.COM/TAG/OPTIMIZATION-PROBLEM/) OTTO PRODUCT CLASSIFICATION (HTTP://BLOG.KAGGLE.COM/TAG/OTTO-PRODUCT-CLASSIFICATION/) OWEN ZHANG (HTTP://BLOG.KAGGLE.COM/TAG/OWEN-ZHANG/) PRACTICE FUSION (HTTP://BLOG.KAGGLE.COM/TAG/PRACTICE-FUSION/) PRODUCT NEWS (HTTP://BLOG.KAGGLE.COM/TAG/PRODUCT-NEWS/) PROFILING TOP KAGGLERS (HTTP://BLOG.KAGGLE.COM/TAG/PROFILING-TOP-KAGGLERS/) RECRUITING (HTTP://BLOG.KAGGLE.COM/TAG/RECRUITING/) REGRESSION PROBLEM (HTTP://BLOG.KAGGLE.COM/TAG/REGRESSION-PROBLEM/) SCIKIT-LEARN (HTTP://BLOG.KAGGLE.COM/TAG/SCIKIT-LEARN/) SCRIPTS (HTTP://BLOG.KAGGLE.COM/TAG/SCRIPTS/) SCRIPTS OF THE WEEK (HTTP://BLOG.KAGGLE.COM/TAG/SCRIPTS-OF-THE-WEEK/) THE HUNT FOR PROHIBITED CONTENT (HTTP://BLOG.KAGGLE.COM/TAG/THE-HUNT-FOR-PROHIBITED-CONTENT/) TOURISM FORECASTING (HTTP://BLOG.KAGGLE.COM/TAG/TOURISM-FORECASTING/) TUTORIAL (HTTP://BLOG.KAGGLE.COM/TAG/TUTORIAL/) VIDEO SERIES (HTTP://BLOG.KAGGLE.COM/TAG/VIDEO-SERIES/) WIKIPEDIA CHALLENGE (HTTP://BLOG.KAGGLE.COM/TAG/WIKIPEDIA-CHALLENGE/) XGBOOST (HTTP://BLOG.KAGGLE.COM/TAG/XGBOOST/)

Archives

Select Month

(https://www.facebook.com/kaggle) (https://twitter.com/kaggle)