Communicating Results and Creating a Data Product

Therri Usher

Tuesday, April 21, 2015

Discussion Time

As a millenial, what are some of the ways that we can use new areas of technology, such as social media, to disseminate scientific results out to researchers? The general public?

Think about it as a group before we discuss as a class.

Communicating Results

What is the point of completing an entire data analysis if no one else ever knows of the results?

It is our responsibility as medical and/or public health researchers to communicate effectively so that we may help others.

Research Papers

We have used academic papers to communicate scientific results for decades, if not centuries. It is the backbone of scientific knowledge.

The peer-review process is meant to ensure that our research is vetted properly and able to withstand scientfic inquiry.

Popular software to create statistical papers:

1. Microsoft Word
2. LaTeX

While it continues to be the main form of communicating new information in academia, academic papers are no longer the only way to communicate results.

Research Posters

Research posters are widely used in the scientific community, especially at conferences. They are a quick and attractive way to present the main points of an analysis to an audience.

Software to create statistical research posters:

1. Microsoft PowerPoint
2. Adobe Illustrator
3. LaTeX

Research Presentations

Research presentations are also widely used in the scientific community, particularly at conferences. It is a way to communicate your results from your data analysis orally to an audience.

Software to create statistical research presentations:

1. Microsoft PowerPoint
2. Markdown
3. Beamer
4. Slidify

Poster or Presentation?

- Presentations tend to reach a broader audience than posters.
- Posters allow you to converse with your audience.
- Presentations can have a bit more prestige, according to some.
- Posters allow you to present smaller findings or practice for a larger presentation.

My Opinion: It's great to communicate your results to others in either way.

New Software For Communicating Results

While papers, posters, and presentations might not be very new to the scientific community, some of the tools we can use to create these products are newer.

NOTE: This discussion is not meant to teach you how to use these tools but to inform you of their existence and capabilities.

Beamer

LaTeXuses document classes to create different text products. For instance, the article class is used to create papers.

Beamer is the LaTeXdocument class for creating presentations. The slides for this course are created using R Markdown and Beamer. The same rules of LaTeXstill apply for Beamer.

An alternative class for making presentations in LaTeXis Powerdot.

Slidify

Slidify is an R package that allows users to generate interactive, web-ready HTML slides. It also allows for quick publishing through GitHub or Dropbox.

Once Slidify is set up, it uses the same syntax as R Markdown to create the slides. The interactive component is a bit more complicated.

Slidify has a lot of customization for the advanced user that might be confusing to the beginner. However, the creator's website and GitHub repository is helpful.

Rpubs

Rpubs is a publishing service that allows RStudio users to publish reports made using R Markdown online.

Once you register, you can create HTML documents and publish them directly to Rpubs by pressing a couple of buttons.

Data Products

Data products are an effective method of disseminating your results to as many people as possible.

Data products are useful when your analysis creates something new to offer, such as:

- New data sets
- Data visualizations
- Statistical methods
  - Data cleaning
  - Statistical modeling

Researchers are more likely to use your new data or methods if they are able to obtain them or easily use them.

Shiny

Shiny is a package created by RStudio. It allows users to create web applications without having to explicitly code in HTML, CSS, or Java. In particular, it is a great tool for data visualization.

Their website includes a gallery of apps as well as a beginner tutorial and articles for those who want more information.

R Packages

R gives users the capabilities to create their own packages. It can be very helpful to have your own personal R package if you have functions you commonly use. However, it is also great for distributing data products, such as data and methods, that you might have created during your data analysis.

To build R packages, you must have Rtools downloaded on your computer. (See `http://cran.r-project.org/bin/windows/Rtools/`)

Hilary Parker, an alumnus of the Biostatistics department at JHSPH, has a blog post (`http://hilaryparker.com/2014/04/29/writing-an-r-package-from-scratch/`) that explains how to create a basic R package simply and cleanly. We will now go through the post.

R Packages

The Comprehensive R Archive Network, or CRAN, houses many if not all of the packages we have used in this course. It is possibly the best way to ensure that your package is used by others.

However, not all R packages are on CRAN. You can house your packages on GitHub as well as other repositories. The R package devtools has an install_github command that allows you to download R packages from GitHub if you know the package name and the creator's name.

Think Outside The Box

We live in a new age of technology. More and more statisticians and researchers are making their presence known on the internet, through blogs and social media.

Don't forget about the classic ways of communicating results, such as papers, but use all the available forms of communication to your advantage. Don't be afraid to try something new.

Bottom Line

While we are researchers interested in answering a research question, we should think of ourselves as storytellers. Our data analyses should tell a story that also answers our research questions. The use of data products can help us tell the story so that it reaches as many people as possible and can help us all become great storytellers.

Active Learning Exercise

Given what we have discussed, how would you disseminate the results of your data analysis project to an audience? Devise a plan of doing so. Be sure to address the classical methods of communication, the kinds of software and products you would use, as well as the use of social media, blogs, and other forms of technology in communicating your results.

Does it agree with what you would have done before this discussion? If so, how? If not, why?

References

1. How to Create a Research Poster
2. Designing Effective Research Posters
3. LaTeX
4. LaTeXWikibook
5. Slidify
6. Slidify's GitHub Repository
7. Rpubs
8. Shiny
9. Rtools
10. Writing an R package from scratch
11. Coursera: Developing Data Products