Origin of Data Science Studio Blogs

"A peer-reviewed blogging journal about reproducible analytics, functional graphing, statistical modeling, and data visualization"

Dr. Craig Hutton, PhD

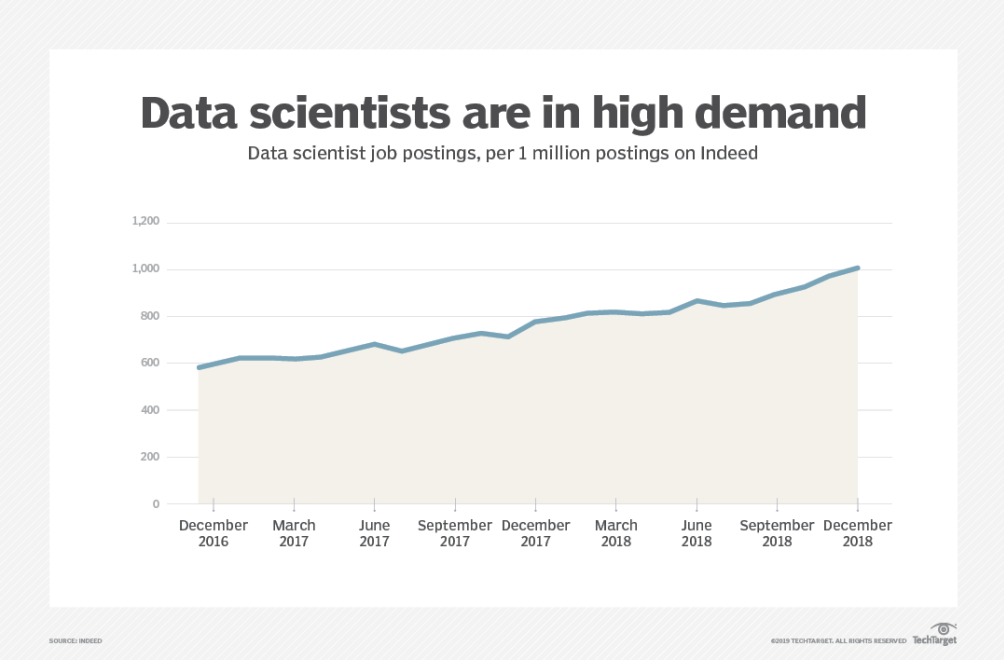
Data Science Fellow

Office of the Chief Information Officer

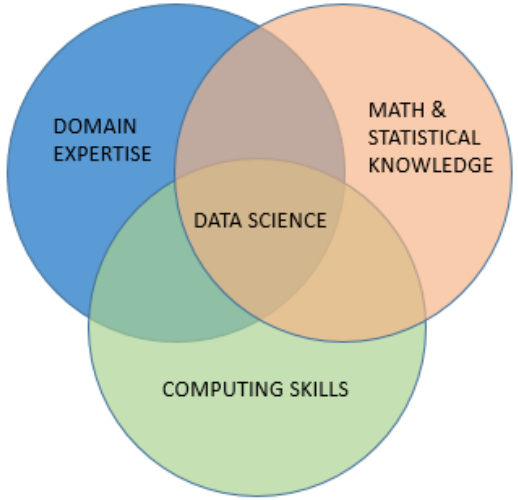
Province of British Columbia

Why a blogging platform?

- The demand people who can use data effectively is increasing.
- The field is new: many aspiring professionals are supplementing their education by learning online.
- Too much unregulated information – tough to know what to trust or where to find it
  - Most of it is scattered amongst numerous personal blogs or disorganized
- This is of particular concern for health data science



https://searchbusinessanalytics.techtarget.com/feature/Demand-for-data-scientists-is-booming-and-will-increase



https://medium.com/@anandr42/the-data-science-delusion-7759f4eaac8e

DSS

Origin of the DSS blogging platform

- Promoting quality reproducible analytics is a core goal of the DSS
- Many highly skilled analysts are too busy to update a blog frequently
- A way to help those passionate about teaching to reach a wider audience
  - DSS blogging community can help get your content online
- What if there was a way to contribute infrequently on a personal blog (e.g. monthly) but have your work featured on a frequently-updated site?
- Peer-review enables a degree of quality control
- Institutional affiliation is also nice.
- Through the DSS we can create a high-quality repository of useful content for readers of all skill levels.

DSS

The DSS blogging platform

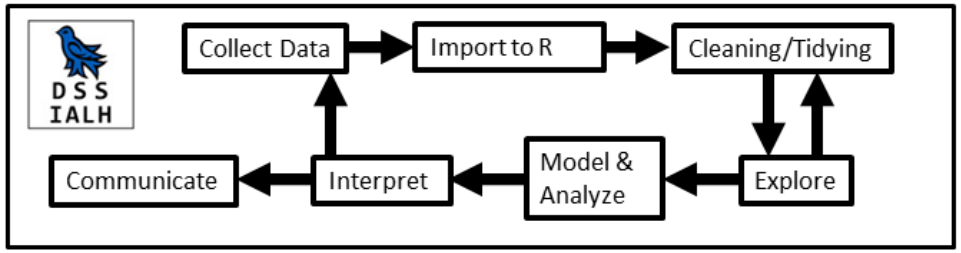
- A platform to promote reproducible analytics, functional graphing, statistical modeling, and data visualization.
  - Emphasis on health research applications
  - https://dss-ialh.rbind.io/post/welcome-to-dss-blogs/index.html

- Me: A Scientist's guide to R
  - https://craig.rbind.io/
  - Tutorials for science trainees on reproducible data analysis using R
    - Will cover examples and applications in mental health, neuroscience, psychology

- Matthew: Statistics, (health) data visualization, blogging tutorials
  - https://matthew-parker.rbind.io

- Andriy: reproducible research and (health) data visualization
  - https://andriy.rbind.io/

Example



Data analysis workflow outline

- What are the main steps of analyzing data? My typical analytical workflow follows these 7 steps:
- **0. Install and load necessary packages**
- **1. Import the data.**
- **2. Clean/Transform the data**
- **3. Explore the data**
- **4. Resolve structural issues**
- **5. Model**
- **6. Analyze and Interpret**
- **7. Communicate the results**

https://craig.rbind.io/post/a-scientist-s-guide-to-r-intro-basic-workflow/