Demystifying Big Data: Value of Data Analysis Skills for Research Librarians

Tammy Ann Syrek-Marshall, MLS
Tami.sky.mars@gmail.com
This Presentation will briefly cover…

- The relationship of Library Science to the Data, Information and Knowledge Sciences
- Understanding what is Data, Data Science and Data Analysis
- Career Paths in Data Science for Librarians
- Valuation of Data Analysis skills
- Overview of pathways to obtaining skills and knowledge
- Bad Data and the hidden dangers of Data Analysis
DIKUW Pyramid and the Hierarchical Relationships Between the Sister Sciences
The DIKUW Cloverleaf, a Feedback Loop on the Path to Wisdom
What is Data?
Big Data’s Seven V’s

• **Volume**: Sample Size
• **Velocity**: Speed of Data Creation, how Fresh it is
• **Variety**: How many additional Variables are in the Set
• **Veracity**: Accuracy of the Data
• **Variability**: Consistency of the Data over time
• **Visualization**: Creating Visual Representations of Data
• **Value**: How Relevant and how Useful the Data is
Descriptive Statistics to Predictive Analysis

• **Data Creation:** Surveys, Research, Behavior Analysis, and so on

• **Data Warehousing:** Specialized systems for data storage, archiving and retrieval

• **Data Retrieval:** Locating, identifying and extracting relevant data.

• **Data Mining:** Using algorithms and machine learning to identify and study data

• **Data Analysis:** Using statistics, programming language, and custom software to turn data into usable information.

• **Data Visualization:** Taking processed data and translating it into graphics like charts or histograms

• **Storytelling:** Interpreting the graphic representation of analyzed data and presenting it in a format that conveys its ‘story’ or meaning.
Data Analysis is a Team Effort

• Data Science is one of the hottest fields of the 21st Century
• Current predictions see an increase in demand of as much as 28% by the year 2020
• A search of online job sites identified at least 28 different job titles in the field.
• Of those 28, 8 job titles made it into Glassdoor’s top 50 ranking
• Topping the ranking in positions 1 and 2 are Data Scientist and DevOps Engineer
• The current demand has not yet met the current supply of qualified candidates.
Some Sample Roles for Librarians in Data Science

• **Data Librarian**

• **Data Warehouse Specialist**: Uses recommended practices to create effective storage and access to data

• **Data Quality Analyst**: Reviews and audits the health and quality of data

• **Analytics Manager**: Team leader for the creation of reports and presentations of post analysis for use by clients

• **Data Storyteller**: Transforming post-analysis Big Data into a text and graphics ‘story’ that conveys the meaning within the data
The Valuation of Data Skills

• **Personal Value**: Knowing where your own interests lie

• **Professional Value**: Knowing what your career goals are and Exploring other career options

• **Organizational Value**: Knowing the needs of your current employer and taking advantage of opportunities when they arise

• **Shared Value**: Expanding the roles of librarians to create new pathways for the future.
Taking the Path Forward

• Online/Video Courses
  • Coursera
  • DataCamp
  • EdX
  • Khan Academy
  • Lynda.com
  • Udacity

• Skill Building Books
  • “The Accidental Data Scientist” by Amy Affelt
  • O’Reilly Data Science Series
Taking the Path Forward

- SLA Certificate and Conference Programs
- SLA Data Caucus
- Data Science Professional Organizations
  - The Data Science Association
  - American Statistical Association
  - The International Institute for Analytics
- Professional Websites and Online Communities
  - Quora – Data Science
  - Data Science Central
  - Kdnuggets
  - Data Mining Research Blog
- College/University Certificate and Degree Programs
Bad Data’s Seven I’s

• Incomplete Data: 1,2,3,5,6,7,9,11
• Inaccurate Analysis: 2+2=
• Ill-conceived Algorithms: If X=1 then Y=
• Implicit Bias: Men are better with computers than women
• Inappropriate Sourcing: Using data on heart arrhythmia to predict outcomes of treatment for asthma.
• Invasion of Privacy: Unauthorized access to SSNs and PINs
• Illegal Access: see CA & Facebook
Case Study: Puerto Rico and Hurricane Maria
Are There Any Questions?