

Data Science for the Public Good Young Scholars Program Curriculum

DSPG trainings focus on learning data science methods and tools. The students engage in 47 classroom training modules covering the topics listed in Table 1. By the end of the program, the students have learned about and used the skills learned in their projects. All students have learned or increased their proficiency in using data science software and tools.

Table 1. DSPG Training Classes, 2020

Getting Ready 1 SDAD Data Science Process 2 Data Security Requirements 3 Introduction to the SDAD Computing Platforms Code Management with Git 4 Naming Things 5 Project Setup, Git Setup, Git on your own 6 Git with branches for collaboration R Basics Boot Camp 7 R Notebooks 8 Strings, Factors, Numerics, Date, Times 9 Data Objects 10 Iteration (Loops and Vectorization), Functions 11 Regex Data Discovery 12 FIELD TRIP: Library of Congress LABS 13 Literature Review 14 Data Inventory & Screening Research Ethics & Behavioral Experiments 15 Research Ethics 16 Brown Bag: Data Privacy & Stakeholder Expectations 17 Behavioral Experiments Data Ingestion 18 From Files & APIs 19 From Databases & Spatial Scraping Data 20 Web-Scraping: HTML 21 Web-Scraping: Javascript Transforming (& Creating) Data 22 Restructuring on Ingest 23 Creating Data from Satellite Imagery	Data Management 24 Storage, Security, Destruction 25 Metadata Data Profiling 26 Structure 27 Quality 28 Metadata & Provenance Data Preparation 29 Restructuring 30 Cleaning 31 Transformation 32 Joining 33 Deduplication Data Exploration & Analysis 34 Exploration & Visualization 35 Mapping 36 Text Mining Media Interaction 37 Media Writing 38 Media Speaking Data Modeling 39 Modeling: Social Network Analysis 40 Modeling: Regression / GLM 41 Modeling: Classification & Clustering 42 Modeling: Decision Trees Data Presentation 43 APIs 44 Graphics Cookbook 45 Latex 46 Shiny Documents & Dashboards, part 1 47 Shiny Documents & Dashboards, part 2
---	--

As we switched to a virtual program last summer, we are benefiting from online recorded trainings from DSPG 2020.

The recorded training material from last year's virtual program is posted on:

<https://www.youtube.com/playlist?list=PL5pmbS-xLPOFeiLGKYwunU5EBpFzLd-bJ>

Our current training modules for DSPG 2021 program and associated materials are listed on:

<https://github.com/DSPG-Young-Scholars-Program/DSPGTraining2021>