⌂ (/bin/view/Main/) ▾ / DSPG (/bin/view/DSPG/) ▾ / DSPG Program Schedule (/bin/view/DSPG/Schedule/) ▾

DSPG Program Schedule

Last modified by Aaron Schroeder (/bin/view/XWiki/aschroed) on 2019/05/31 15:58

Skillz Types	Professional	Technical	Research

Getting Ready

Welcome to the SDAD research family!

Date	Activity	Instructor
06/03/2019	**The SDAD Data Science Process Overview** People, Processes & Platforms	Aaron Schroeder
06/03/2019	**Data Security Requirements** SDAD Data Management Plan, SDAD Data Use Management Agreement	Aaron Schroeder
06/03/2019	**Introduction to the SDAD Computing Platforms** Components, Accessing Servers (RStudio, PostgreSQL, Files)	Aaron Schroeder

Code Management with Git

A free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency

Date	Activity	Instructor

| 06/04/2019 | **Naming Things** | Aaron Schroeder |
| | Machine Readable, Human Readable, Plays well with Default Ordering | |

| 06/04/2019 | **Project Setup, Git Setup, Git on your own** | Devika Nair |
| | Learning the SDAD Way! Git with it! | |

| 06/04/2019 | **Git with branches for collaboration** | Devika Nair |
| | The why and how of source control branching | |

Date	**Activity**	**Trip Leader**
06/05/2019 1:00PM	**FIELD TRIP: Library of Congress LABS**	Vicki Lancaster
	A place to encourage innovation with Library of Congress digital collections	

R Basics Boot Camp

Getting started with some critical basics

Date	**Activity**	**Instructor**
06/06/2019	**R Notebooks**	Aaron Schroeder
	An R Notebook is an R Markdown document with chunks that can be executed independently and interactively, with output visible immediately beneath the input. R Notebooks are an implementation of Literate Programming (https://en.wikipedia.org/wiki/Literate_programming) that allows for direct interaction with R while producing a reproducible document with publication-quality output.	
06/06/2019	**Strings, Factors, Numerics, DateTimes**	Bayoán Santiago
	R has a wide variety of data types including scalars, vectors (numerical, character, logical), matrices, data frames, and lists.	

| 06/06/2019 | **Data Objects**

Base VS Data.Table VS The Acolytes of Wickham (The Tidy Way): data.frames, data.tables & tibls | Aaron Schroeder |

| 06/07/2019 | **Iteration (Loops and Vectorization), Functions**

"Looping", "cycling", "iterating" or just replicating instructions is an old practice that originated well before the invention of computers. It is nothing more than automating a multi-step process by organizing sequences of actions or 'batch' processes and by grouping the parts that need to be repeated.

A function is a set of statements organized together to perform a specific task. R has a large number of in-built functions and the user can create their own functions. | Bayoán Santiago |

| 06/10/2019 | **Regex**

A regular expression (regex or regexp for short) is a special text string for describing a search pattern. You can think of regular expressions as wildcards on steroids. But with regular expressions you can do much more. | Teja Pristavec |

Date	Activity	Speaker
06/11/2019 LUNCH	**BROWNBAG: Data Privacy & Stakeholder Expectations** Data and privacy is more than law and regulation!	Joy Tobin

Data Discovery

Identifying potential data sources that could be related to the specific topic of interest

Date	Activity	Instructor

| 06/11/2019 | **Literature Reviews**

A scientific literature review is a focused, systematic, and critical synthesis of key research conducted on a given topic. We'll discuss designing your review, selecting databases and journals, devising search queries, integrating your findings, using citations, and composing bibliographies, and we'll cover tools that'll help you do it all efficiently. | Teja Pristavec

Vicki Lancaster |
| 06/12/2019 | **Data Inventory & Screening**

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor | Devika Nair |

Research Ethics & Behavioral Experiments

Date	Activity	Instructor
06/13/2019	**Research Ethics** Research that involves human subjects or participants raises unique and complex ethical, legal, social and political issues. Research ethics is specifically interested in the analysis of ethical issues that are raised when people are involved as participants in research.	Sallie Keller Stephanie Shipp
06/13/2019	**Behavioral Experiments (Hands-On and !FUN! Exercise)** **Human subject research** is systematic, scientific investigation that can be either interventional (a "trial") or observational (no "test article") and involves human beings (https://en.wikipedia.org/wiki/Human_being) as research subjects.	Gizem Korkmaz

Data Ingestion

Data Ingestion is the process of bringing data into the Data Management Platform. The variety of ingestion processes will be as diverse as all the potential data sources and will account for a broad and heterogeneous spectrum of required access requirements, protocols, data formats, schemas, and update frequencies.

Importing Data

Date	Activity	Instructor

06/14/2019	**From Files & APIs** ACS, LODES, GTFS?	Aaron Schroeder

06/14/2019	**From Databases & Spatial** Indexes	Aaron Schroeder

Scraping Data

Date	Activity	Instructor
06/17/2019	**Web-Scraping: HTML** The rvest package provides an easy to use, out of the box solution to fetch the html code that generates a webpage	Bayoán Santiago
06/17/2019	**Web-Scraping: Javascript** When the website or webpage makes use of JavaScript to display the data you're interested in, the rvest package misses the required functionality and we must instead to look to approaches using packages like Selenium or Phantomjs (headless chrome)	Bayoán Santiago

Transforming (& Creating) Data

Date	Activity	Instructor
06/18/2019	Restructuring on Ingest Eviction Data	Aaron Schroeder
06/18/2019	Creating Data from Satellite Imagery Land Cover	Aaron Schroeder

Data Management

Date	Activity	Instructor

| 06/19/2019 | Storage, Security, Destruction

ggg | Joel Thurston |

| 06/19/2019 | Metadata

ggg | Joel Thurston |

Data Profiling

sdfsdf sdfsdf sdsdfsd

Date	Activity	Instructor
06/20/2019	**Structure** Is the data appropriately structured for purposes of analysis? Because datasets are often created for reasons of administration and reporting, and further that there are generally few constraints on their organization, administrative datasets are often constructed in manners not conducive to statistical analyses.	Aaron Schroeder
06/20/2019	**Quality** Does the data correctly represent the real-world construct to which it refers? Investigating and describing various categories of desirable attributes (or dimensions) of data.	Aaron Schroeder
06/20/2019	**Metadata & Provenance** When dealing with data that is to be used for research purposes, it is of vital importance to discover if the datasets (tables), their observational observation units (records/rows), and their attributes (fields/columns) are consistently named, sufficiently described, and appropriately formatted for analysis and for combination with other project datasets.	Aaron Schroeder

Data Preparation

sdsdsdsd sdsdssd

Date	Activity	Instructor

| 06/21/2019 | **Restructuring** | Aaron Schroeder |

To address issues of structure discovered during Data Profiling, it often necessary to restructure the data source (dataset) into multiple new datasets that are more easily analyzed. This activity can be thought of as being akin to the process of database normalization, the process of organizing the columns (attributes) and tables (relations) of a relational database to minimize data redundancy.

| 06/24/2019 | **Cleaning** | Nathaniel Ratcliff |

Data cleaning refers to the process of either fixing or removing data in a data source that is incorrect, incomplete, improperly formatted, or duplicated.

| 06/24/2019 | **Transformation** | Nathaniel Ratcliff |

Data transformation refers to the mapping of dataset field values from a provided format into an expected or more useful format.

- Aggregation
- Imputation
- Smoothing

| 06/25/2019 | **Joining** | Aaron Schroeder |

- Deterministic
- Probabilistic
- Geographic

| 06/26/2019 | **Deduplication** | Aaron Schroeder |

Record Deduplication, or more generally, Record Linkage is the task of finding which records refer to the same entity, like a person or a company. It's used mainly when there isn't a unique identifier in records like Social Security Number. This means one can't trivially find duplicate records in a single dataset, nor easily link records from different datasets.

Data Exploration & Analysis

Data Exploration refers to the analysis of the datasets by summarizing main characteristics,often with visual methods. Data exploration is used throughout the data framework.

Date	Activity	Instructor
06/27/2019	**Exploration & Visualization** Data Exploration refers to the analysis of the datasets by summarizing main characteristics, often with visual methods. Data exploration is used throughout the data framework. Descriptive statistics play a principal role in data profiling, from identifying valid attribute values to checking for semantic consistency. The use of visual techniques like boxplots support iterations between data cleaning and transformation during the data preparation. Distributional characterizations of the data help identify needs and opportunities for data linkage.	
06/28/2019	**Mapping** An introduction to analyzing spatial data in R	Aaron Schroeder
07/01/2019	**Text Mining** This tutorial will cover common steps in processing text for meaningful analysis in R. We'll practice tokenization and removing stopwords, examine word frequency, conduct sentiment analysis, create word clouds, find n-grams, and visualize word networks and word correlations.	Teja Pristavec

Media Interaction

Date	Activity	Instructor
6/27/2019	**Media Writing** Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor	Michael Todd
6/28/2019	**Media Speaking** Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor	Grace Dusseau?

Data Modeling

| 07/02/2019 | **Modeling: Social Network Analysis** | Gizem Korkmaz |
| | Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor | |

| 07/08/2019 | **Modeling: Regression / GLM** | Brandon Kramer |
| | Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor | |

| 07/09/2019 | **Modeling: Classification & Clustering** | Josh Goldstein |
| | Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor | |

| 07/10/2019 | **Modeling: Decision Trees** | Josh Goldstein |
| | Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor | |

Data Presentation

Date	Activity	Instructor
07/11/2019	**APIs** Making new data sets available programmatically. Creating RESTful APIs using R is straightforward using the plumber package. Even if you have never written an API, plumber makes it easy to turn existing R functions into API endpoints.	Aaron Schroeder
07/12/2019	**Graphics Cookbook** Adopting sound standards for graphical output	Vicki Lancaster
07/15/2019	**Latex** Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor	Gizem Korkmaz

07/16/2019	### Shiny Documents & Dashboards Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor	Aaron Schroeder
07/17/2019	### Shiny Documents & Dashboards Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor	Aaron Schroeder

Tags:

Created by Aaron Schroeder (/bin/view/XWiki/aschroed) on 2019/04/16 22:16

No comments for this page