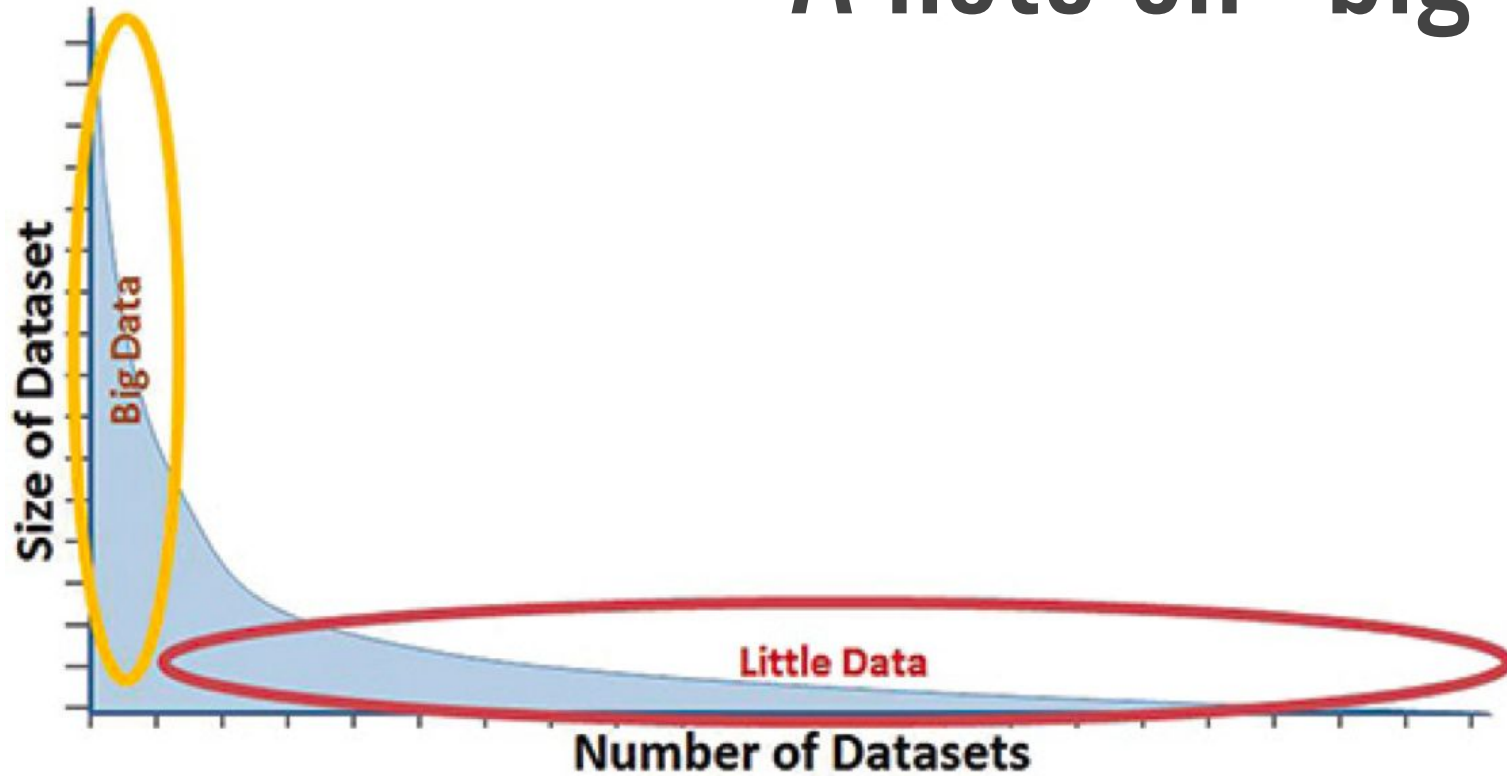# Course Reminders

- A1 - due Sunday 4/14 (11:59 PM)
- Project Proposal due *next* Sunday

Today's slides are both a PDF and a Jupyter notebook. You'll have to pull the whole directory and unzip to have access to everything locally (or upload everything to datahub).

A note on "big data"

Types of data: Big vs. Little

- There are difficulties in working with large data sets.
  - The analysis cycle time slows as data size grows (slow to iterate)
  - Large data sets are complex to visualize

- Simple models do not require massive data to fit or evaluate

Big Data Approach? Small Data Approach?

What are current voter preferences about the democratic presidential campaign pool?

Which approach is more accurate?

<u>Take away</u>: The right data set is the one most directly relevant to the tasks at hand, not necessarily the biggest one.

COGS108 Final Projects

The **COGS108 Final Project** will give you the chance to explore a topic of your choice and to expand your analytical skills. By working with real data of your choosing you can examine questions of particular interest to you.

Final Project: Objectives

- Identify the problems and goals of a *real* situation and dataset.
- Choose an appropriate approach for formalizing and testing the problems and goals, and be able to articulate the reasoning for that selection.
- Implement your analysis choices on the dataset(s).
- Interpret the results of the analyses.
- Contextualize those results within a greater scientific and social context, acknowledging and addressing any potential issues related to privacy and ethics.
- Work effectively to manage a project as part of a team.

Deadlines & Action Items

Project Proposal: Due Sunday, April 21st @ 11:59 PM (Week 3)

Final Project: Due Wednesday, June 12th @ 11:59 PM (Finals Week)

Note: there was a typo in the original date in the slide for proposal due date. Date here has been edited.

Project Proposal (10%)

All Parts are required for project completion:

1. Questions(s) + Dataset(s)
2. Team Expectations
3. Schedule for Completion

- Make copy of proposal & share with group
    - Work on it as a team
    - Save as PDF
    - Submit on gradescope
- Be sure to read COGS108 Team Policies

Full project guidelines are here: https://github.com/COGS108/Projects/blob/master/FinalProject_Guidelines.pdf

The best projects start with a question NOT the dataset. The most boring projects are dataset-first.

Where to look for and get data for your projects?

Available Datasets

- US Census Data
- data.gov
- Awesome Public Datasets
- Data Is Plural
- Datasets | Deep Learning
- Stanford | Social Science Data Collection
- Open Climate Data
- Eviction Lab (email required)
- Data and Story Library

There are more dataset sources listed on the FinalProject_Guidelines document on GitHub

Finding Data & Statistics: Home

Home | **What is Data?** | **Frequently Used Statistics** | **Frequently Used Data** | **Find Data by Topic** ▾ | **Data APIs** | **Text Data** ▾

Statistical Analysis Software ▾ | **Data Visualization** ⧉

Library Data Services

- Data Services
- Data & GIS Lab
- GIS @ UCSD
- Finding Data & Statistics
- Research Data Curation

Featured UC San Diego Collections

- UC San Diego Dataverse 🔒
  Miscellaneous datasets
  purchased for the UC San Diego
  community. Includes:
  Data on Terrorist Suspects
  (DOTS)
  Field (California) Poll, 1956 -
  [Latest Release]
  International Country Risk Guide:
  Table 3B: Political Risk Points by
  Component
  International Terrorism: Attributes
  of Terrorist Events (ITERATE),
  1968 - [most recent]
  Latin American Public Opinion
  Project (LAPOP) 1978-2003

Finding Data & Statistics

Welcome to the UC San Diego Library's guide

Data repositories and datasets linked here are ou

specific guides, many of which include sections

are also available. If you need help finding particu

out to a librarian for assistance.

Data Spotlight

UN data
A world of information

A one-stop-shop interface for
accessing statistics collected by
United Nations agencies. Search
across 32 databases (60 million
records!) by topic, country, or
region.

Not finding what you need? Check
the website of the specific UN
agency for additional statistics.

◄

Off-Campus Access & Wireless

Many of the resources listed on this guide hav

Diego Library. Off-campus access, as well as

Find Data by Topic (dropdown)

- Art and Culture
- Country Statistics & Data
- Crime
- Data Science
- Economic & Financial Data
- Economics: Datastream software
- Education
- Environment & Energy
- Food & Beverage
- Government Spending & Infrastructure
- Health/Health Care & Mortality
- Labor, Employment, Wages
- Latin American Public Opinion Project (LAPOP)
- Latinobarómetro
- Marketing
- Migration & Immigration
- People: Census guide ⧉
- People: Children, Families, Aging
- People: Demographics & Population (general)
- People: Gender Studies & Women
- People: Race & Ethnicity
- People: Religion
- Political Science
- Political Science: Worldwide Elections Guide
- Public Opinion, Social Attitudes and Values
- San Diego & California
- Social Media

(partially obscured right column)

...nsultations

...nter Quarter 2019

...n consultation hours

...esdays, 11am-12pm

...Data & GIS Lab

...who can provide

...nie Labou
...cience Librarian

...Barsh
...an for Economics and
...ss

...se Sklar
...an for Political Science,
...Society, and International
...nment Information

...Smith
...an for US Government
...ation, Urban Studies &
...ng, Environmental Policy,
...an Diego Government

stics/artculture

https://ucsd.libguides.com/data-statistics/home

When the data aren't ready and waiting for you

- APIs
- Web Scraping
- Collecting your own data

Final Project Grading (35%)

Category	Percentage of Project Grade
Introduction and Background	10%
Data Description	10%
Data Cleaning/Pre-processing	10%
Data Visualization	15%
Data Analysis and Results	25%
Privacy/Ethics Considerations	15%
Conclusions and Discussion	15%

Data Wrangling

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu