Plan:
1. Discuss the benefits/limitations to "big data"
2. Review where to find datasets
3. What to do if the data aren't readily available?

# Finding Data

Shannon E. Ellis, Ph.D
UC San Diego

Department of Cognitive Science
sellis@ucsd.edu

Finding Data

Once you formulate a question, you'll need data. Data are *everywhere*. It's up to you to find the data that can best answer your question of interest. Often, this will involve multiple datasets. Typically, no one dataset will be perfect. And, often, even the combination datasets will not be perfect, but they will be good enough.

A note on "big data"

Size of Dataset (y-axis), Number of Datasets (x-axis). Big Data, Little Data.
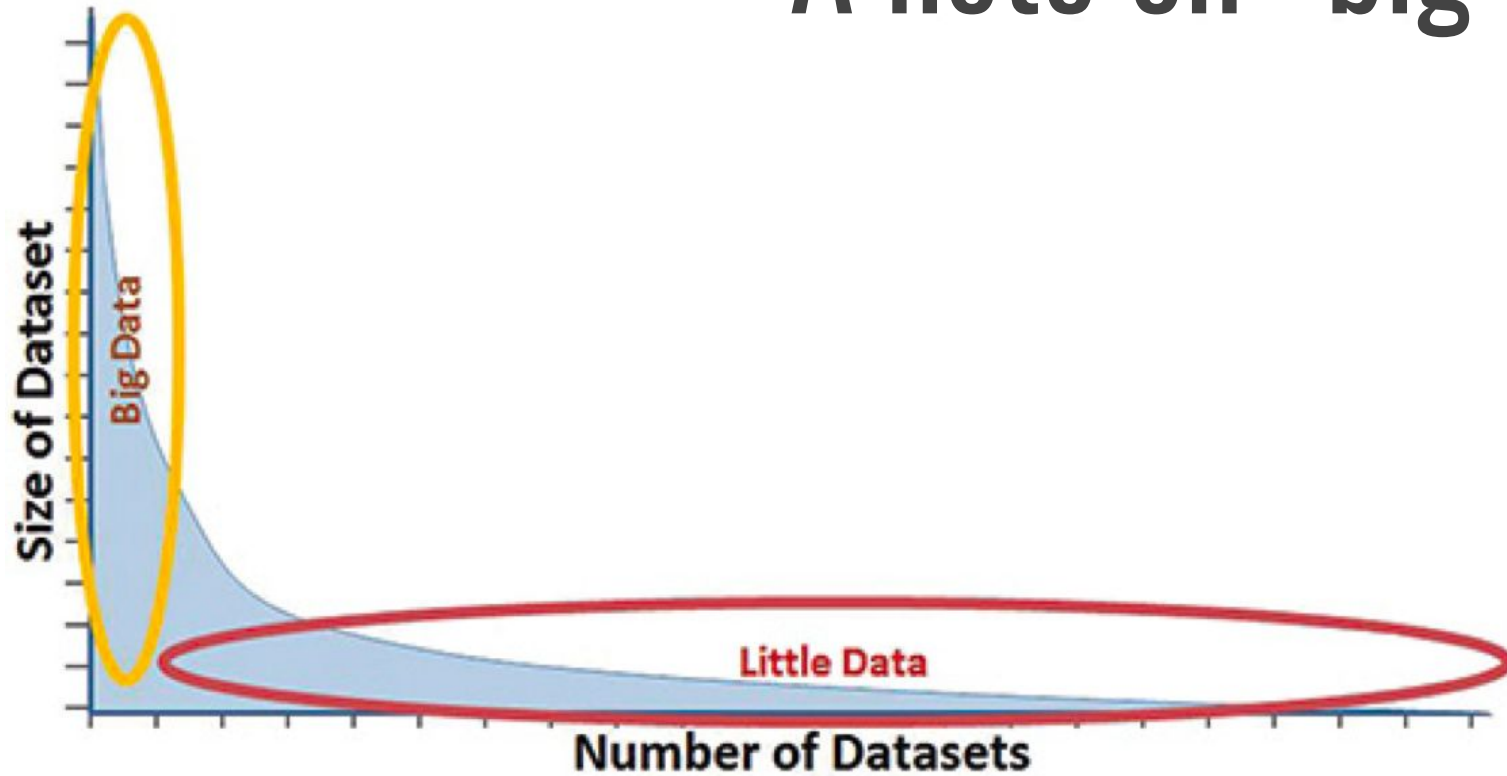
Types of data: Big vs. Little

- 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?

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

The best projects start with a question NOT the dataset.

The most boring projects are dataset-first.

Once you figure out the question, the links on GitHub may be helpful:
https://github.com/COGS108/Projects/blob/master/Option1_Group/FinalProject_Guidelines.md#dataset-resource-list

Where to look for and get data for your projects?

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

UNdata
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

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

Consultations

nter Quarter 2019

n consultation hours

esdays, 11am-12pm

Data & GIS Lab

who can provide

nie Labou
cience Librarian

Barsh
an for Economics and
s

se Sklar
an for Political Science,
Society, and International
ment Information

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

stics/artculture

When the data aren't ready and waiting for you

- APIs
- Web Scraping
- Collecting your own data

You can see previous COGS 108 Projects

https://github.com/COGS108/FinalProjects-Sp17

https://github.com/COGS108/FinalProjects-Wi18

https://github.com/COGS108/FinalProjects-Sp19

https://github.com/COGS108/FinalProjects-Fa19

https://github.com/COGS108/FinalProjects-Wi20