

Creating data

PSYC 11: Laboratory in Psychological Science

Jeremy Manning
jeremy@dartmouth.edu

Data sleuthing lab

- You'll play two roles: **data creator** and **data sleuth**
- First your group will **create** a dataset to be passed off to another group
- Next, your group will **explore** a dataset given to you by another group

Major objective: learn about “results” sections

- What goes into creating a dataset?
- What can vs. can't you conclude from a given dataset?
- Wrangle/analyze data, create figures, run statistical tests
- Interpret results

What's the “point” of data?

- Tells us about something we’re interested in
- Compare conditions and test hypotheses

Where do data come from?

- Measurements from sensors and devices
- Self reports or observations
- Simulations and equations

How can you “create” a dataset?

- Download an existing one!
- Random number generators
- Data spoofing libraries

Existing datasets



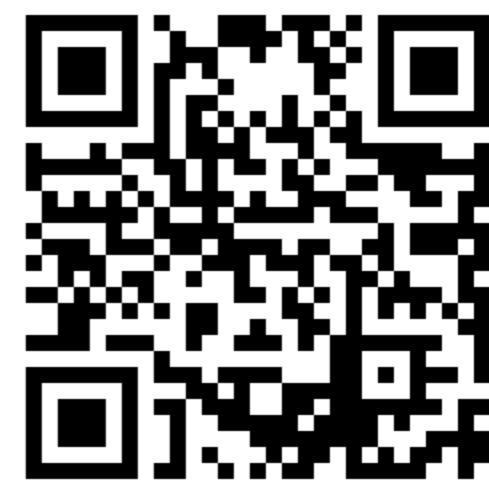
Awesome public data



Fivethirtyeight



Amazon Open Data



Kaggle



Google Data



World Health Organization



Worldbank

Random number generators



NumPy Random

Data spoofing



Faker



Memesis

This week's lab

- Today: find or generate a dataset + come up with 5 questions about the dataset
 - Make the dataset (and questions) interesting!
 - At least 1 question must be **possible** to answer with the dataset
 - At least 1 question must be **impossible** to answer with the dataset
 - Organize into a spreadsheet for easy analysis
 - At least 5 “features” per observation
 - At least 500 observations

This week's lab

- Wednesday: hand off datasets to another group, explore your received dataset:
 $(A \rightarrow B, B \rightarrow C, C \rightarrow A)$
- Friday: wrap up data analysis, discussion

This week's lab

