Data Scavenger Hunts

Learning about datasets together

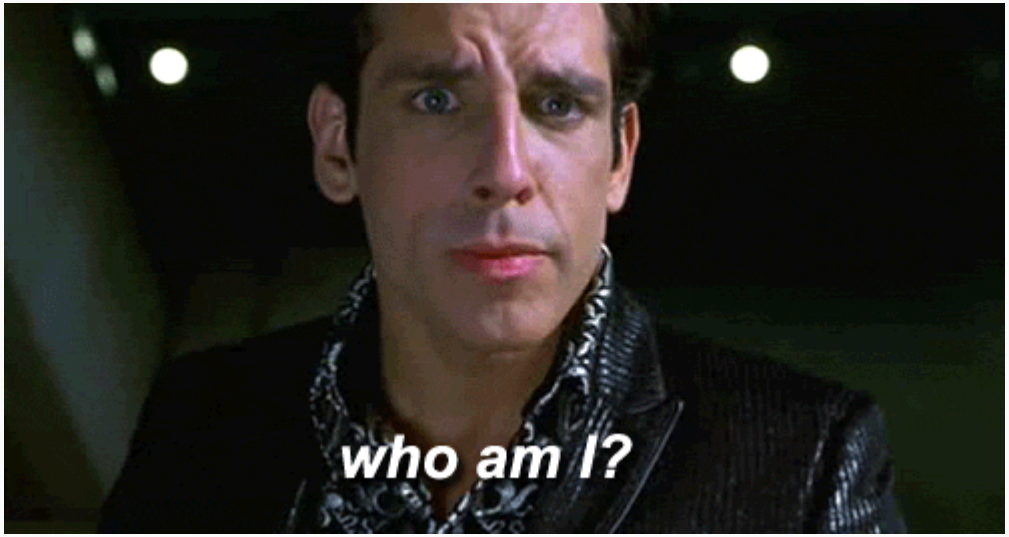
Ted Laderas

2019-05-09

Overview: Data Scavenger Hunts

- Introduction
- Why?
- What are they?
- Who are they For?
- Who are they for? (reprise)
- Call to Action

Who am I?



- Assistant Professor at Oregon Health & Science University
- Many years working with bioinformatics data
- I want people to understand the data they generate
- I love teaching visualization

Why? Looking at Data Together

- Visualization is empowering
- Visualization is accessible
- Visualization lets us have a conversation
- Visualization/Interpreting is a valuable skill

What are Data Scavenger Hunts?

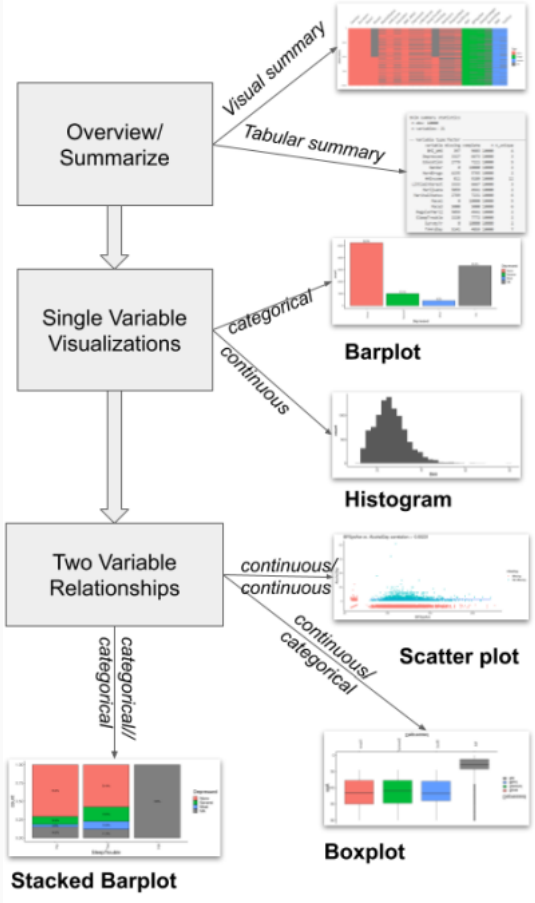
- Social learning activity about a dataset
- Inspired by Exploratory Data Analysis
  - 'Find patterns, reveal structure (Behrens)'
- Understand associations between variables
  - How is Body Mass Index associated with Diabetes?
- Reflect/present discoveries to the group

What? Burro

- Burro[w] into the data!
- R Package/Shiny App for exploring many kinds of data
- Spin up a web app for any dataset
- Maps variable types to the appropriate visualizations
- Allow people to explore the data quickly

What? Burro Overview

- Organized by variable types
- Overview Data
- Continuous vs Categorical Data
- Assess single variables first, then look at interactions by variable type

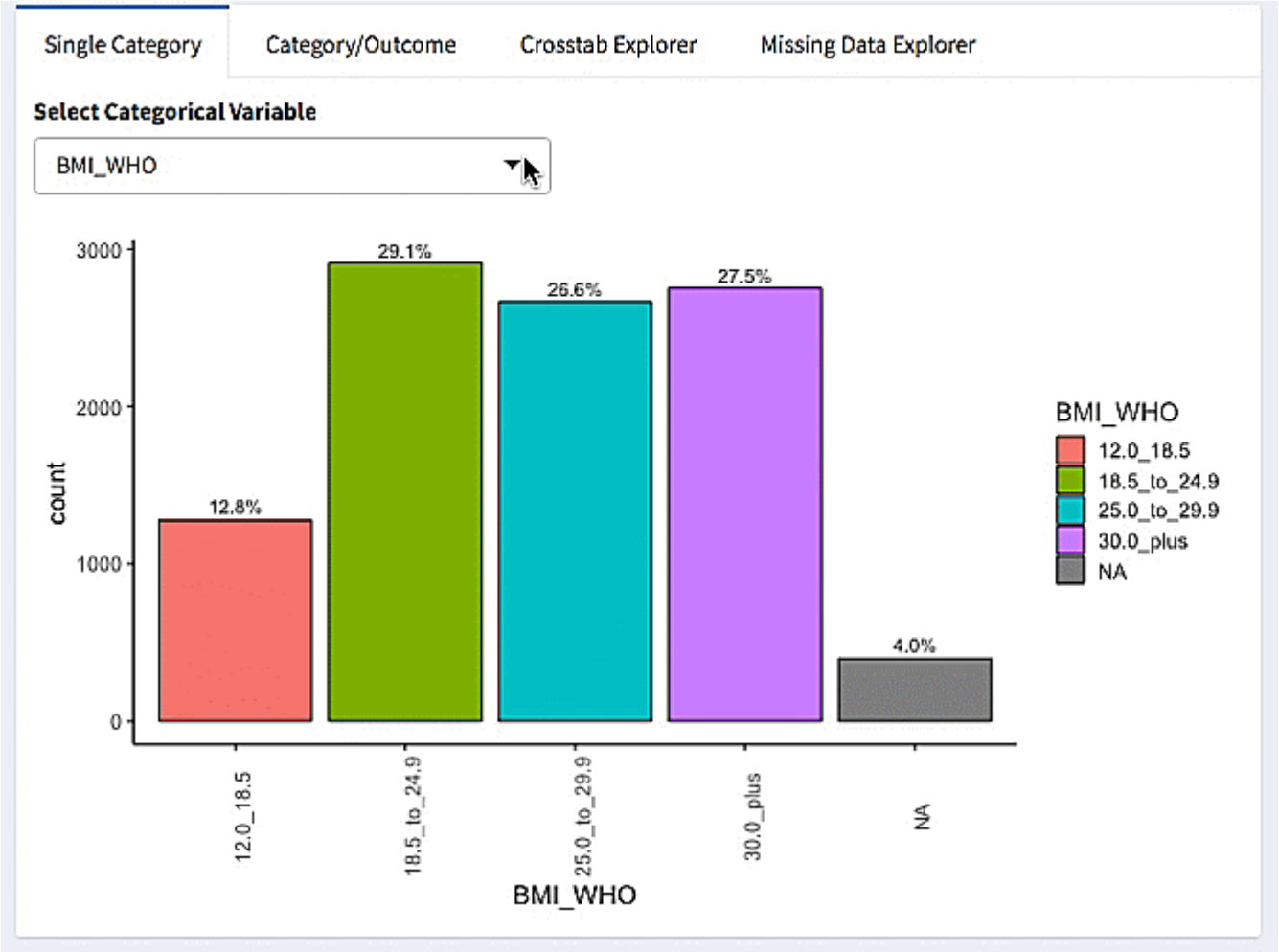


What? NHANES Dataset

- National Health and Nutrition Examination Survey
- Survey to answer nutrition habits
- Lots of outcomes, and is of broad interest
- Looked at Depression as outcomes

What? Burro Example

Looking for association of episodes of little interest with depressive episodes:



What? The Questions

- Have pre-prepared questions that guide exploration
- Questions range from simple to complex
- Map question types into visualization types

What? Try it out

http://bit.ly/csv_burro

Try one of these questions:

- Why is age capped at 80 years in this dataset?
- Is marijuana use associated with depressive episodes?
- Are hours of sleep associated with depressive episodes?

What: Reflection/Discussion

- Bring back our observations to the group
- Simple Template:
  - What was our question?
  - What was our evidence? How did we find it?
  - What other variables should we look at?
- Group reflection: what did we learn as a group?

Who has participated?

- Graduate Students
  - Portland State Students
  - Cardiovascular Risk Data (synthetic data sets)
- Graduate Students (clinical data wrangling course)
  - Clinical and Bioinformatics incoming students
  - Sleep Heart Health Study Data
- BioData Club Members (NHANES Scavenger Hunt)
  - NHANES Data
- Undergraduate Students (Data Literacy)
  - Public Health Education
  - NHANES Data

What do they say (BioData Club)?

- "The scavenger hunt. It is great place to apply this knowledge."
- "I enjoyed the collaborative and self-paced approach."
- "I liked the dataset and the open curiosity element of the workshop."

Who: Useful for Citizen Data Efforts?

- What would it take to make `burro` useful for citizen data efforts?
  - Better "just in time" documentation
- Can it be used to teach graph literacy and data curiosity?
- Can we use this framework to democratize data science?

Who: Spatial Data Exploration

- New example: Social Determinants of Health
- Prototyping with spatial data frames (shapefiles, sf)
  - Example: https://tladeras.shinyapps.io/oregon_county/

Call to action

We need beginners/novices!

- Need help with usability testing/documentation
- Need help with workflows for different data types
- Need help with standardizing template for scavenger hunts

Funding Acknowledgement

- T15 NLM Training Supplement Grant: 5T15LM007088
- NCATS Center for Data to Health (CD2H) Grant: U24TR002306

Stay in Touch

These slides: http://bit.ly/csv_hunt

- email me: laderast@ohsu.edu
- @tladeras
- More about data scavenger hunts and `burro`:
  - http://laderast.github.io/burro/
- Burro email list:
  - http://tinyletter.com/tladeras
- Contributing Info:
  - https://laderast.github.io/burro/CONTRIBUTING
- BioData Club:
  - Local portland club that focuses on learning about data
  - http://biodata-club.github.io