Exploratory Data Analysis

Professor Anna Neufeld

Tuesday, September 10

Announcements

General Weekly Schedule: Subject to change; check GLOW for the most updated version.

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday
		Section 1 Lecture 8:30am-9:45am Wachenheim 113		Section 1 Lecture 8:30am-9:45am Wachenheim 113	
		Section 2 Lecture 9:55am-11:10am Wachenheim 113		Section 2 Lecture 9:55am-11:10am Wachenheim 113	
			TA Session Elizabeth 12pm-3pm Math/stat library		HW and/or project milestone due on GLOW at 1pm (most weeks)
		Prof. N office hours 2:30pm-4pm Wachenheim 239		Prof. N office hours 2:30pm-4pm Wachenheim 239	
TA session Jina 7-8:30pm Griffin 7	TA session Ali + Rachel 7-8:30pm Griffin 7		TA session Ali 7pm-9pm Griffin 7	TA session Rachel + Jina 7pm-9pm Griffin 7	

- Additional office hours: I will often be in my office from 1pm-2:30pm on Thursdays, and from 2pm-3pm on Wednesdays.

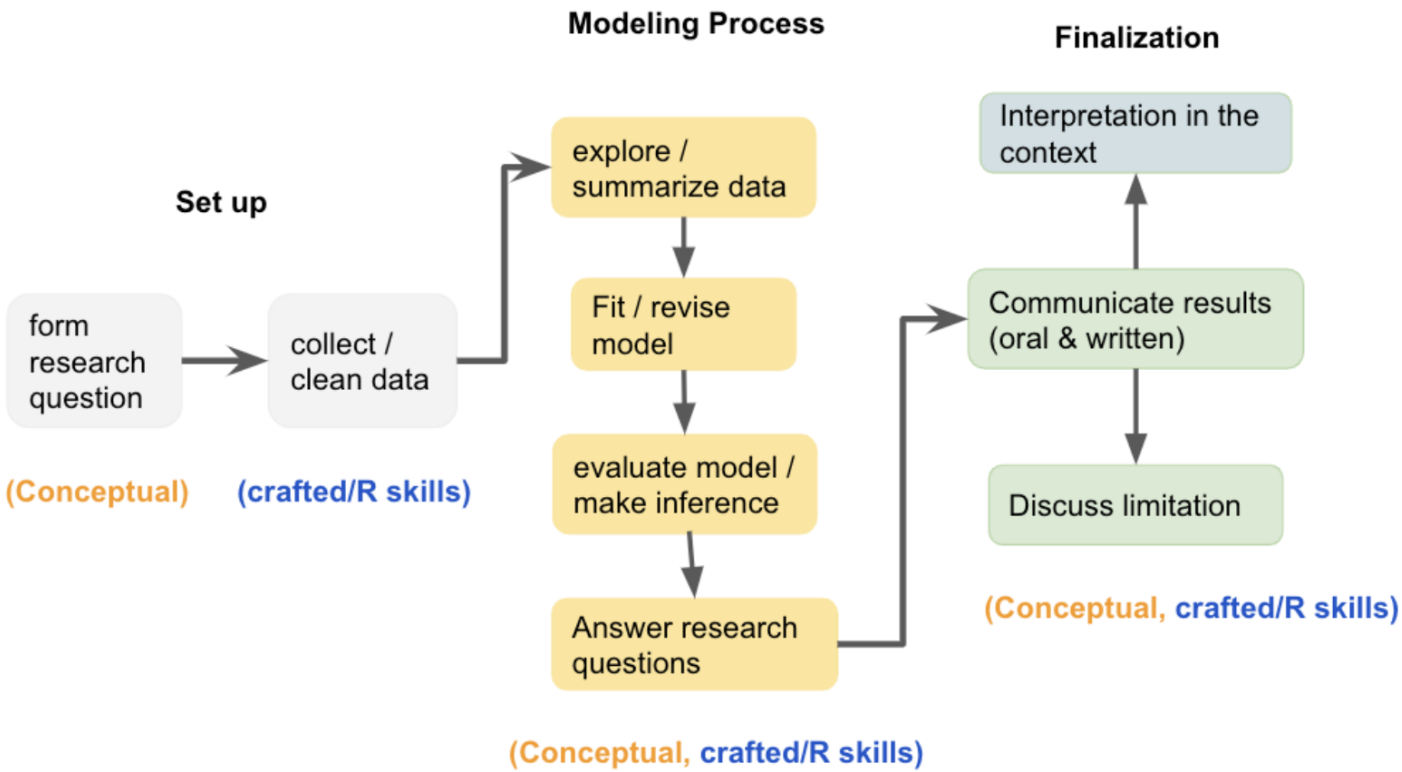
Announcements

- HW1 due Friday on GLOW at 1pm.
- Keep thinking about datasets and groupmates for the project:
  - First milestone is due Friday, 9/27.
  - You can start working on this milestone alone or with groupmates.
  - Example final projects that won national prizes posted on GLOW.
- Stat kick-off colloquium:
  - Wednesday 9/18 at 1pm, NSB 015
  - Ville Satopaa '11

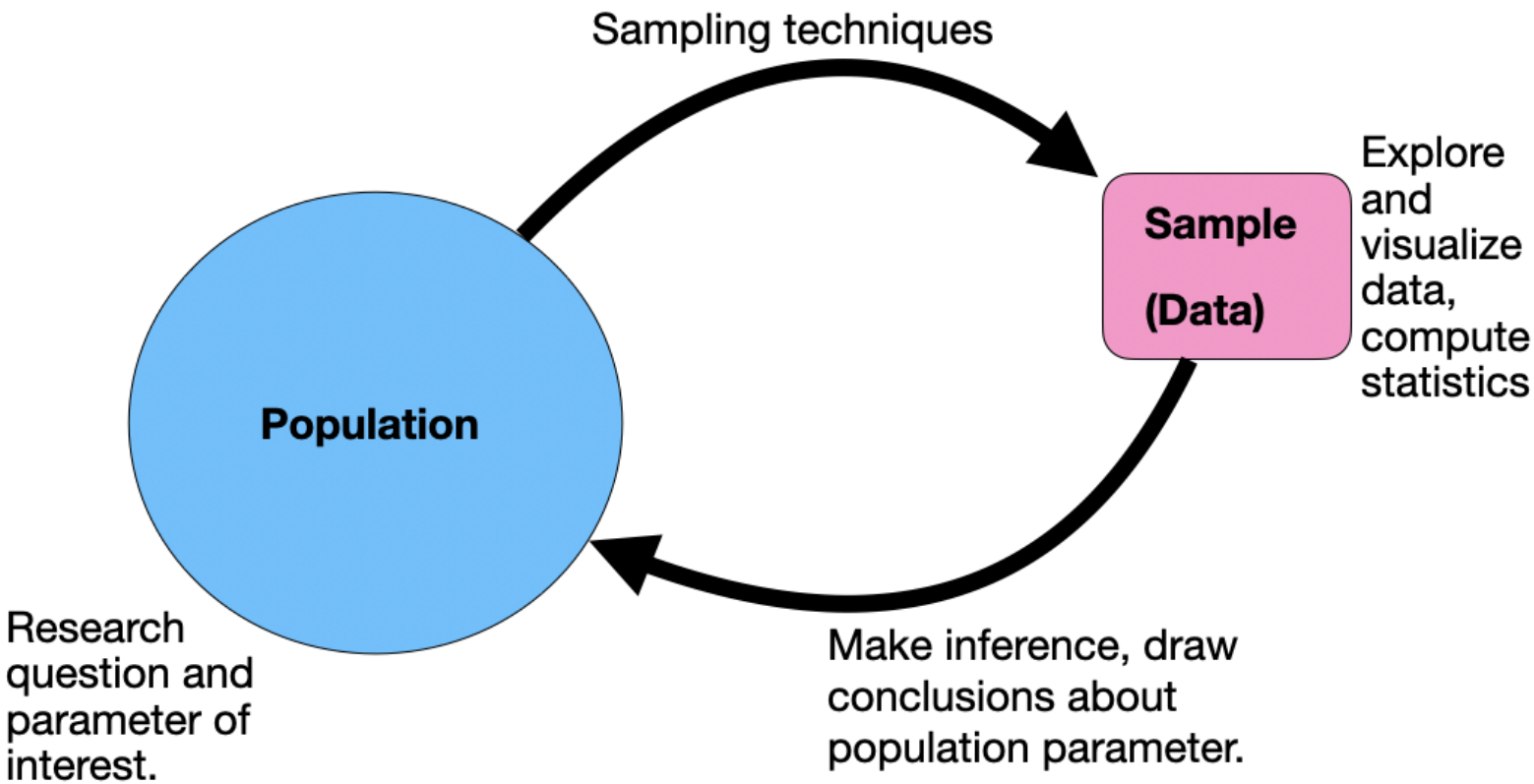
Overview

- **Today:** Exploratory data analysis in R!
- **Thursday:** More of the same
- **Next week:** Sampling and random variables!

Where are we?


Set up
Modeling Process
Finalization
form research question
collect / clean data
explore / summarize data
Fit / revise model
evaluate model / make inference
Answer research questions
Interpretation in the context
Communicate results (oral & written)
Discuss limitation
(Conceptual)
(crafted/R skills)
(Conceptual, crafted/R skills)
(Conceptual, crafted/R skills)

Where are we?


Sampling techniques
Sample
(Data)
Explore and visualize data, compute statistics
Population
Research question and parameter of interest.
Make inference, draw conclusions about population parameter.

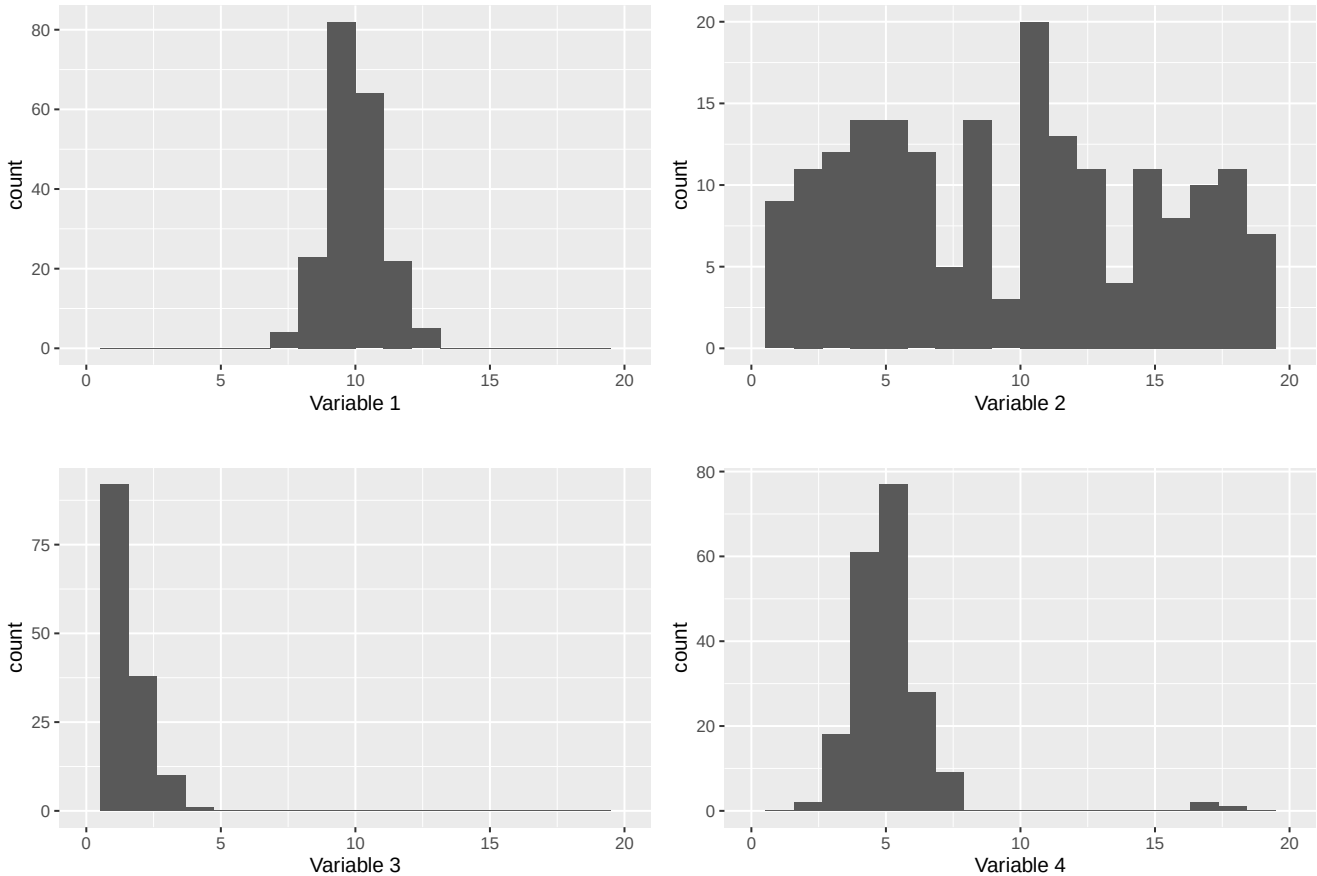
Key words

- Observational unit
- Variable
  - Quantitative
    - Discrete, continuous.
  - Categorical:
    - Ordinal
  - Other types?
- Univariate? Multivariate?
- Response vs. predictor

A single quantitiative variable

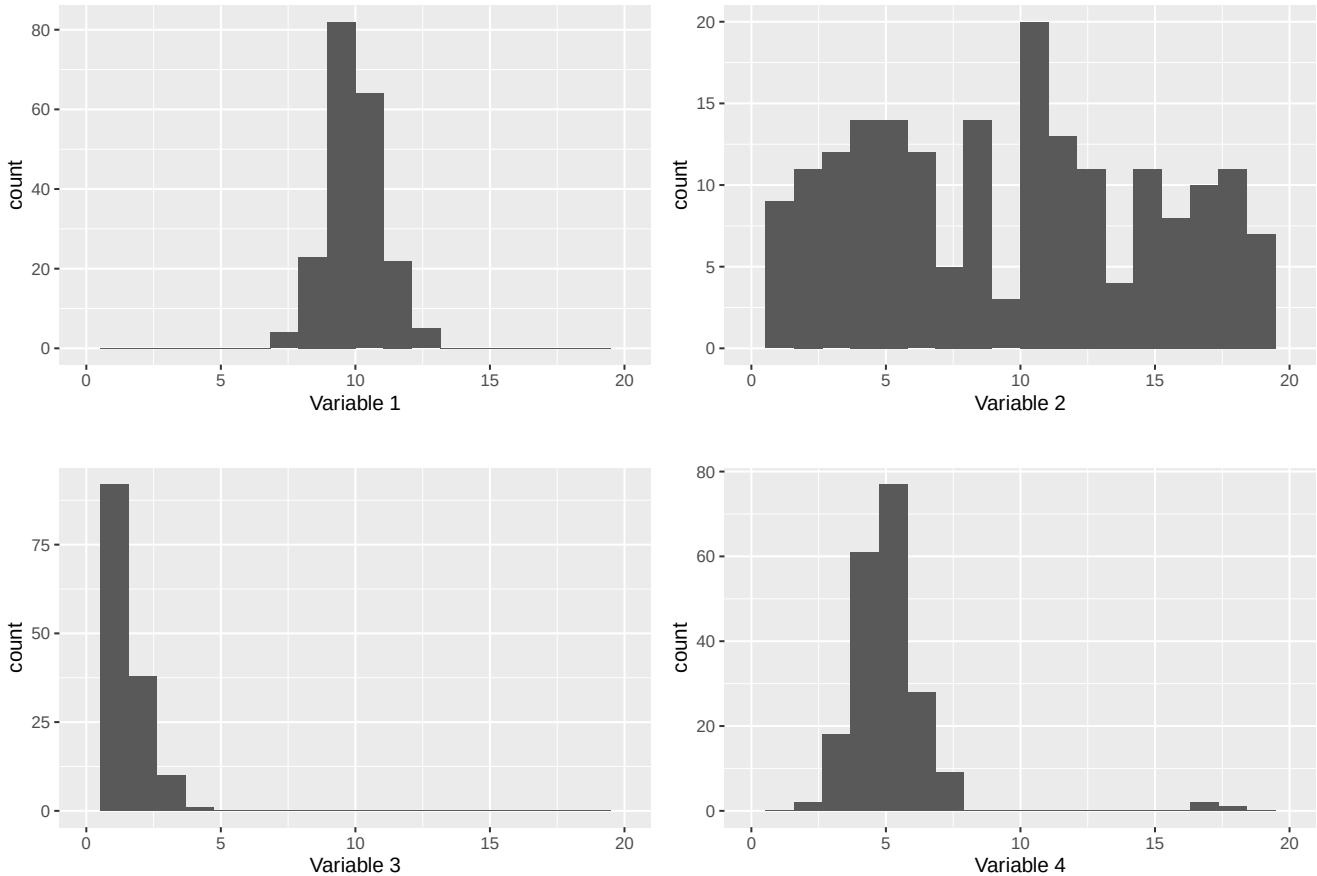
- How do we explore visually?
- How do we summarize numerically?
- What should we be thinking about?

A single quantitiative variable



- We think about SOCS (shape, outliers, center, spread).

A single quantitiative variable



- We think about SOCS (shape, outliers, center, spread).
- Numerical summaries: mean, median, mode, sd, IQR.

Quantitative response, categorical predictor

- How do we explore visually?
- How do we summarize numerically?
- What should we be thinking about?

Quantitative response, categorical predictor

Anna to add: side by side boxplots. Maybe two types. SOCS still important. All the same summaries, just by group.

Quantitative response, quantitative predictor

- How do we explore visually?
- How do we summarize numerically?
- What should we be thinking about?

Quantitative response, quantitative predictor

Anna to add: scatterplots. side by side boxplots. Maybe two types. SOCS still important. All the same summaries, just by group.

Single categorical variable

- How do we explore visually?
- How do we summarize numerically?
- What should we be thinking about?

Single categorical variable

- **Insert table and bar plot.**

Categorical response, categorical predictor

Categorical response, quantitative predictor