Navigating the Frontier:
Statistics, Data Science, and AI in the First Two Years

Patti Frazer Lock

St. Lawrence University

Canton, New York

Joint Mathematics Meetings

January 2025

GAISE

Guidelines for Assessment and Instruction in Statistics Education

Originally published in 2005
Updated in 2016
(Now being updated again)

From the American Statistical Association

Considered the Gold Standard for Introductory Statistics courses

Current GAISE: 2005/2016

1. Teach statistical thinking. Include multivariable thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

GAISE 2016 is far more than just these recommendations!

Contents

GAISE 2016:
What can be omitted?

Contents

- Formal probability
- Constructing plots by hand
- Computation of basic statistics (mean, median, …)
- Drills with z-, t-, χ^2, and F-tables
- Advanced training on a statistical software package

Revising GAISE: GAISE 2025

Incorporate Data Science

Incorporate issues of Diversity, Equity, and Inclusion

Make recommendations more clear

Web-based Document

Steering Committee:

Patti Frazer Lock, St. Lawrence University (co-chair)
Jamie Perrett, Brigham Young University (co-chair)
David Hunter, Penn State University
Lisa Kay, Eastern Kentucky University
Kate Kozak, Coconino Community College
Chris Malone, Winona State University
Maria Tackett, Duke University
Donna Lalonde, ASA Liaison

GAISE 2025: Current Status

Recommendations for Statistics and Data Science

Ten of them. Finalized.

Student Learning Outcomes for Intro Stat

Ten of them? First draft done.

Stuff on Data Science

Not done. But started!

More Chapters: History, Assessment, Current Status, and so on.

Not done. Also started!

Finding the current College GAISE Report.

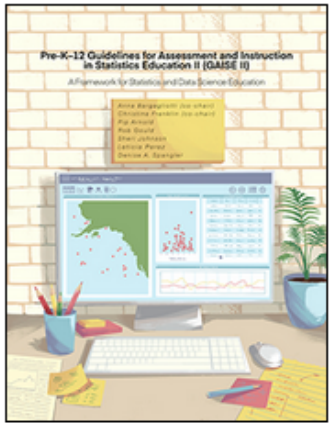
On the ASA site.

Google "GAISE"

Guidelines for Assessment and Instruction in Statistics Education (GAISE) Reports

Participants in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) project have created two reports of recommendations for introductory statistics courses (college level) and statistics education in Pre-K–12 years.

Pre-K–12 Report



Full Report (PDF format)

Purchase a copy of the Pre-K–12 GAISE II Report on Amazon for $15 plus shipping.

Pre-K–12 GAISE II writing team: Anna Bargagliotti, co-chair (Loyola Marymount University); Christine Franklin, co-chair (American Statistical Association); Pip Arnold (Karekare Education New Zealand); Rob Gould (University of California at Los Angeles); Sheri Johnson (The Mount Vernon School); Leticia Perez (University of California at Los Angeles Center X); Denise A. Spangler (University of Georgia)

College Report

LINK TO REVISION!

2016 GAISE College Report

Full Report (PDF Format)

GAISE College Report writing team: Robert Carver (Stonehill College), Michelle Everson, co-chair (The Ohio State University), John Gabrosek (Grand Valley State University), Nicholas Horton (Amherst College), Robin Lock (St. Lawrence University), Megan Mocko, co-chair (University of Florida), Allan Rossman (Cal Poly – San Luis Obispo), Ginger Holmes Rowell (Middle Tennessee State University), Paul Velleman (Cornell University), Jeffrey Witmer (Oberlin College), and Beverly Wood (Embry-Riddle Aeronautical University)

GAISE 2025

Recommendations for Statistics and Data Science

1. Teach statistics and data science as iterative processes of gleaning insights from data to inform evidence-based decisions.
2. Emphasize effective written and oral communication of results from data, with attention to the scope and limitations of conclusions.
3. Focus on conceptual understanding rather than algebraic manipulation and formulas.
4. Integrate real data with a context and purpose throughout the course. Select data that are meaningful and engaging to the students.
5. Encourage multivariable thinking.

GAISE 2025

Recommendations for Statistics and Data Science

6. Incorporate software/apps to explore concepts and work with data.
7. Emphasize responsible and ethical conduct in the collection and use of data and in their analysis.
8. Employ evidence-based pedagogies that actively engage students in the learning process.
9. Use a variety of formative and summative assessments to improve teaching and learning.
10. Implement a course design that uses inclusive strategies to foster a sense of belonging.

Student Learning Outcomes for **Intro Stats** DRAFT

After completing an Introductory Statistics course, a student should be able to:

1. Identify cases and variables in a data set, including multivariable data sets. Recognize whether variables are categorical or quantitative.
2. Assess how the data were collected, and recognize how data collection affects what conclusions can be drawn from the data. Explain the importance of random sampling and random assignment; know the distinction between the two.
3. Identify appropriate graphs and summary statistics for variables and relationships between variables, and correctly interpret information from graphs and summary statistics.
4. Use software or apps to create visualizations, produce summary statistics, and carry out the computational aspects of statistical analysis.
5. Explain the importance of random variation in statistical inference. Recognize that statistical inference involves generalizing from a sample to a population.

Student Learning Outcomes for **Intro Stats** DRAFT

After completing an Introductory Statistics course, a student should be able to:

6. Identify appropriate methods to use in statistical inference, and interpret results from those methods.
7. Understand that conclusions drawn from statistical inference might be wrong, due to sampling variation. Recognize the problem of publication bias and the key importance of replication.
8. Demonstrate effective use of predictive models, such as regression lines.
9. Evaluate ethical issues in statistical practice.
10. Effectively communicate results obtained from data.

DATA SCIENCE

Two important points to consider:

1. How does the impact and growth of data science affect the intro stat course?
2. What might an intro DS course look like?

Impact of Data Science on Intro Stats

More “big” datasets!

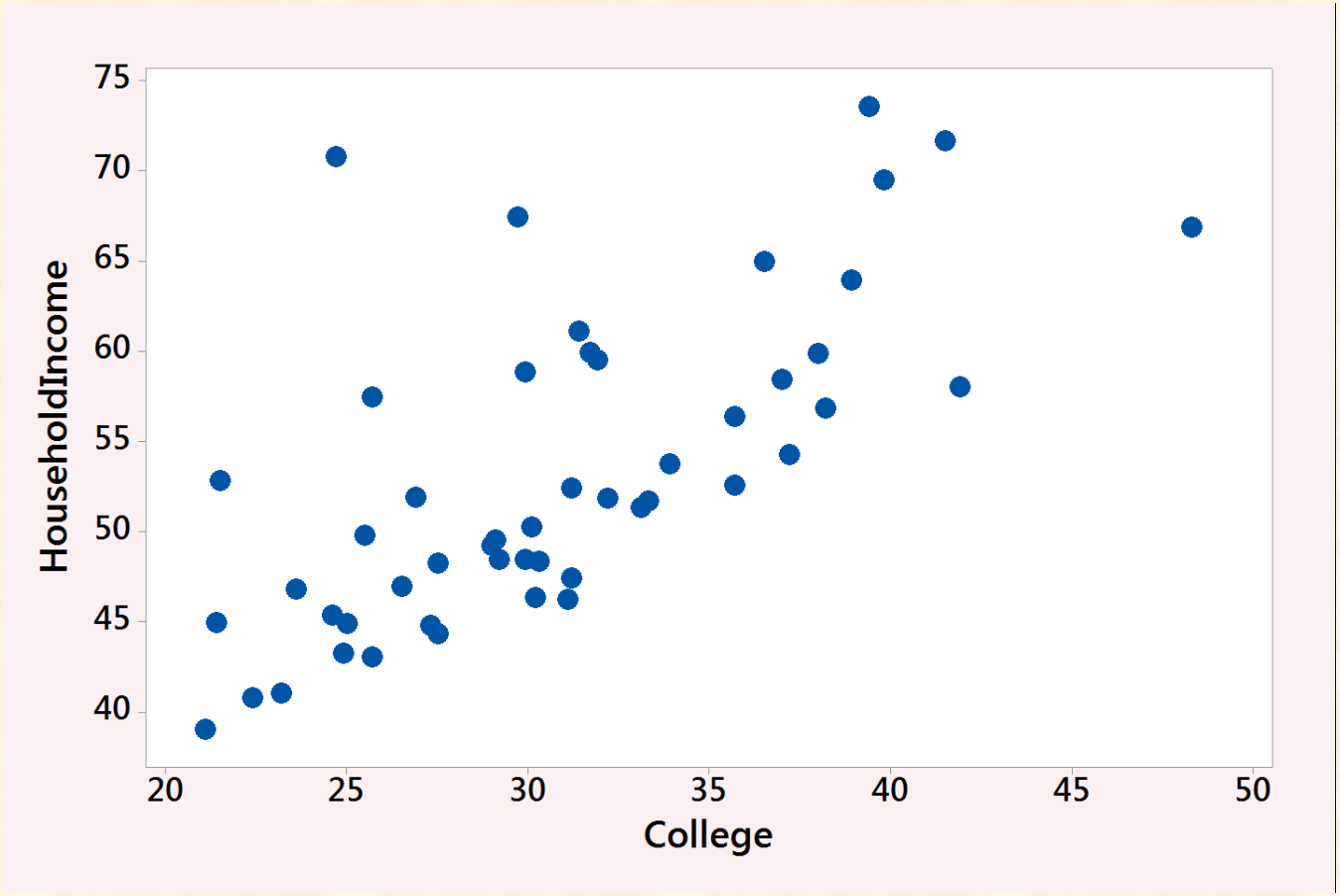
Multiple variables!

Data visualization!

Technology!

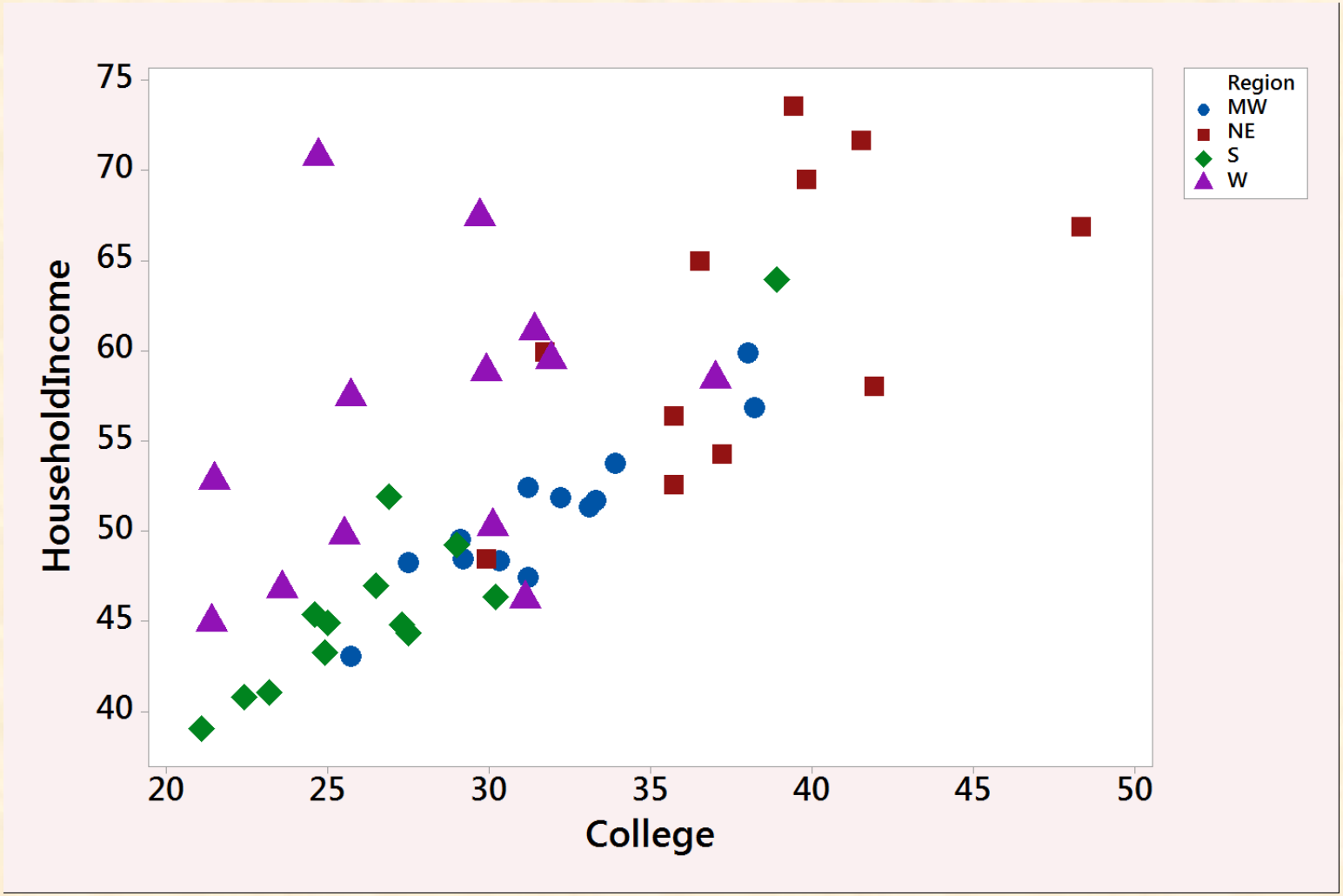
Scatterplot: Two Quantitative Variables

Percent to graduate college. Median Household Income.
For 50 US States.



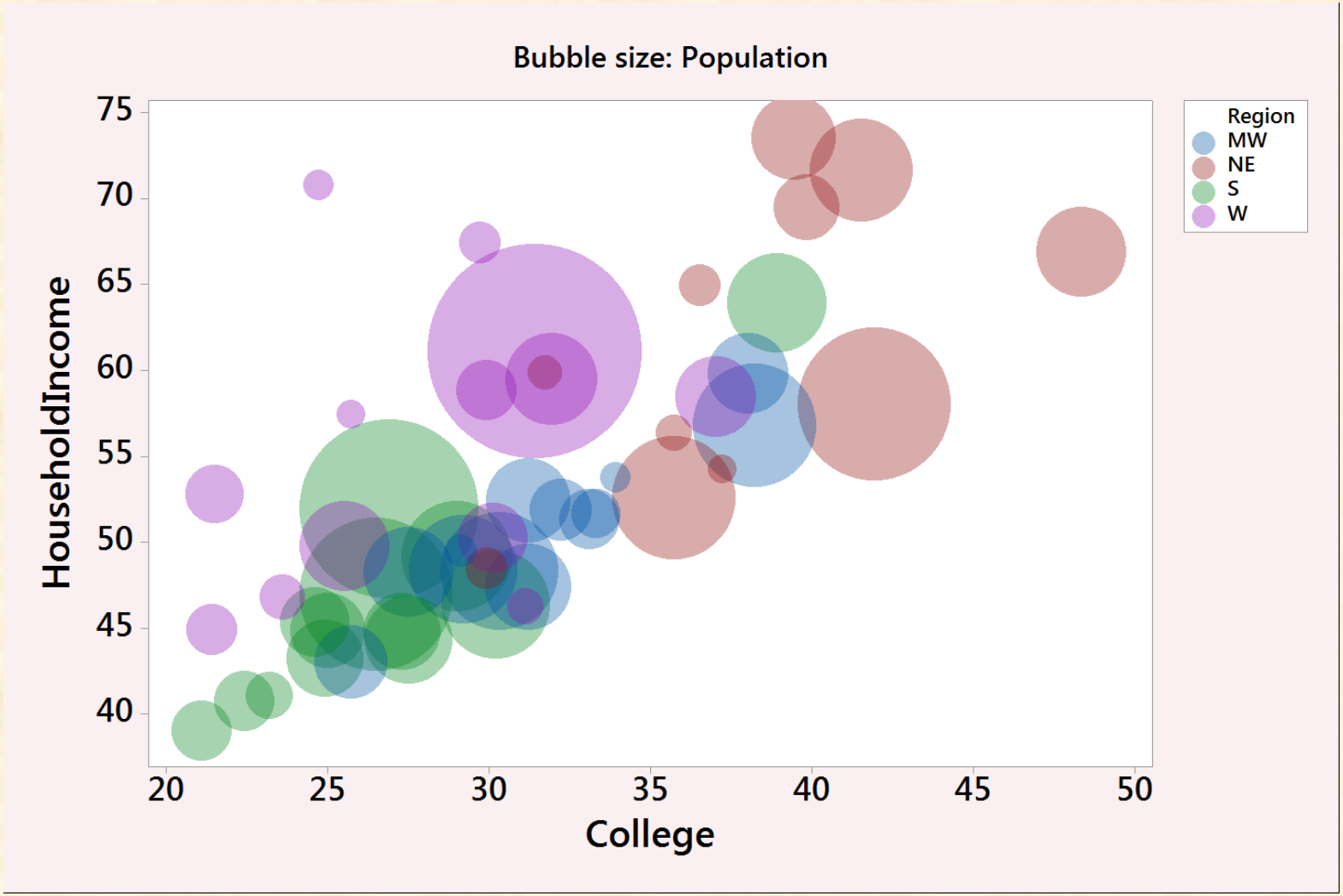
Scatterplot: Add a Third Variable

Categorical variable: Region of the country.



Scatterplot: Add a Fourth Variable!

Quantitative variable: Population.



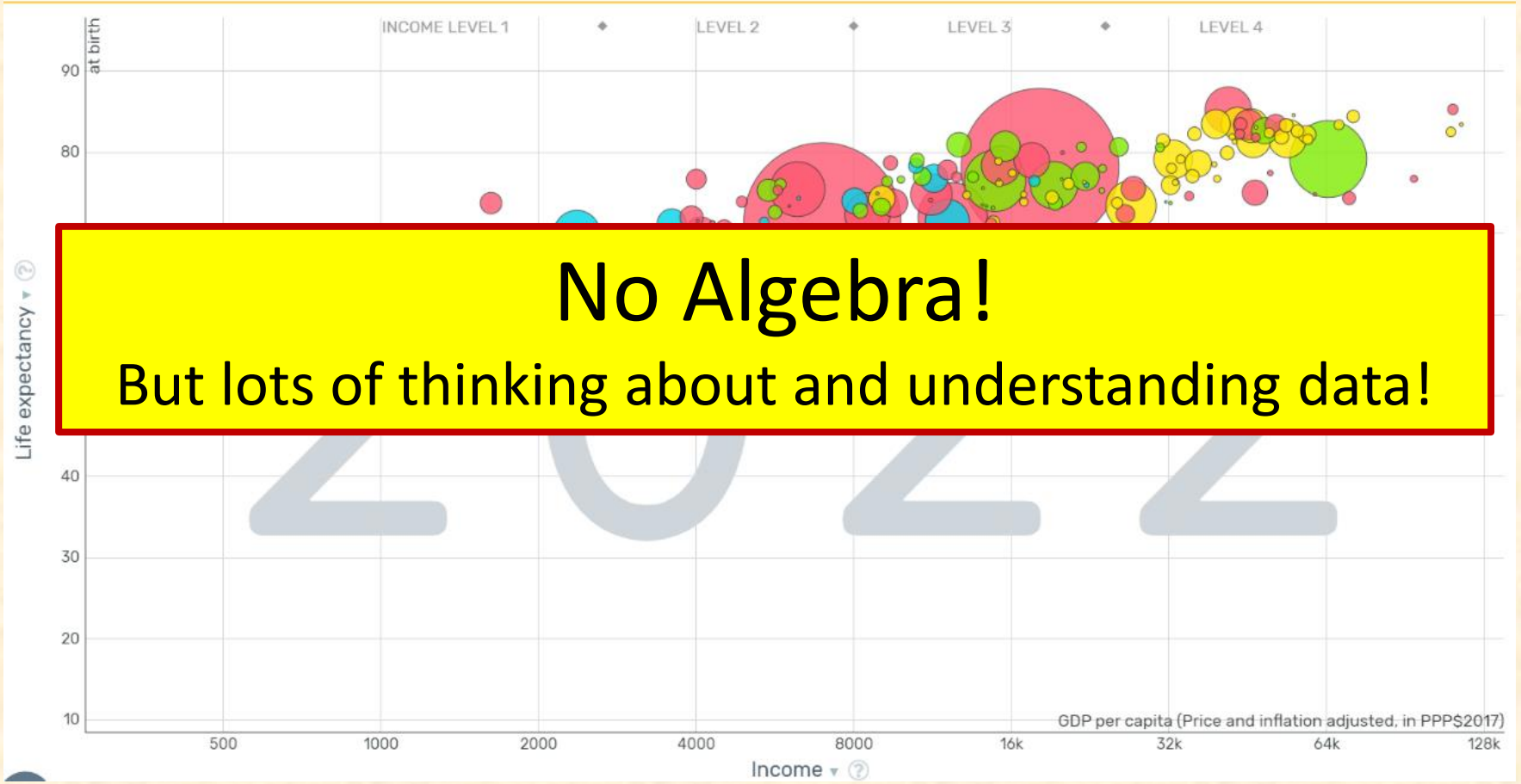
But we want more!

Make it INTERACTIVE!

Make it DYNAMIC!

www.gapminder.com/tools

Gapminder



No Algebra!

But lots of thinking about and understanding data!

www.gapminder.org/tools

Reducing Algebra: Why It Matters

From the *Common Vision* report of the Mathematical Association of America:

Algebra is "**the most significant barrier**" to students finishing a degree in both STEM and non-STEM fields.

This barrier is most acute for low-income students, first generation college students, and underrepresented groups.

<u>And</u>: It's 2024! We have computers!
Algebra is decreasing in importance.
Statistics/Data Science is increasing in importance.
We can use technology to decrease the algebraic load.

A Common Vision for Undergraduate Mathematical Sciences Programs in 2025, Mathematical Association of America, 2015

What about Technology?

Having students use online resources, apps, menu-based statistical software is almost essential in a modern Intro Stats class.

What about coding?

My thoughts:

Coding is also a barrier for many of our students.

Adding "data science" on top of a full Intro Stat course is TOO MUCH.

Coding is hard. Inference is hard. Doing both in one semester is too much!

We need two semesters! (Let's think aspirationally!)

And data is so pervasive in our society, and evidence-based decision-making is so important, that we should have two semesters!

Maybe: Introduce students to (easy) statistical software in Intro Stats.
Introduce them to (advanced) statistical software in Intro DS.

Intro Data Science

So many versions!

Intro DS after Intro Stat as a prerequisite.

Intro DS as a completely separate course. Can be taken in any order.

Intro DS as an "everyone" course; perhaps a precursor to Intro Stat.

ALL GOOD!

More versions!

Intro DS combined with Intro Stat in one course.

Intro DS replacing Intro Stat for students.

NOT SO GOOD?

GAISE and Data Science

Stay tuned as we work through this!

We welcome your insights!

Thanks for listening!

Thoughts?
Comments?
Questions?

plock@stlawu.edu