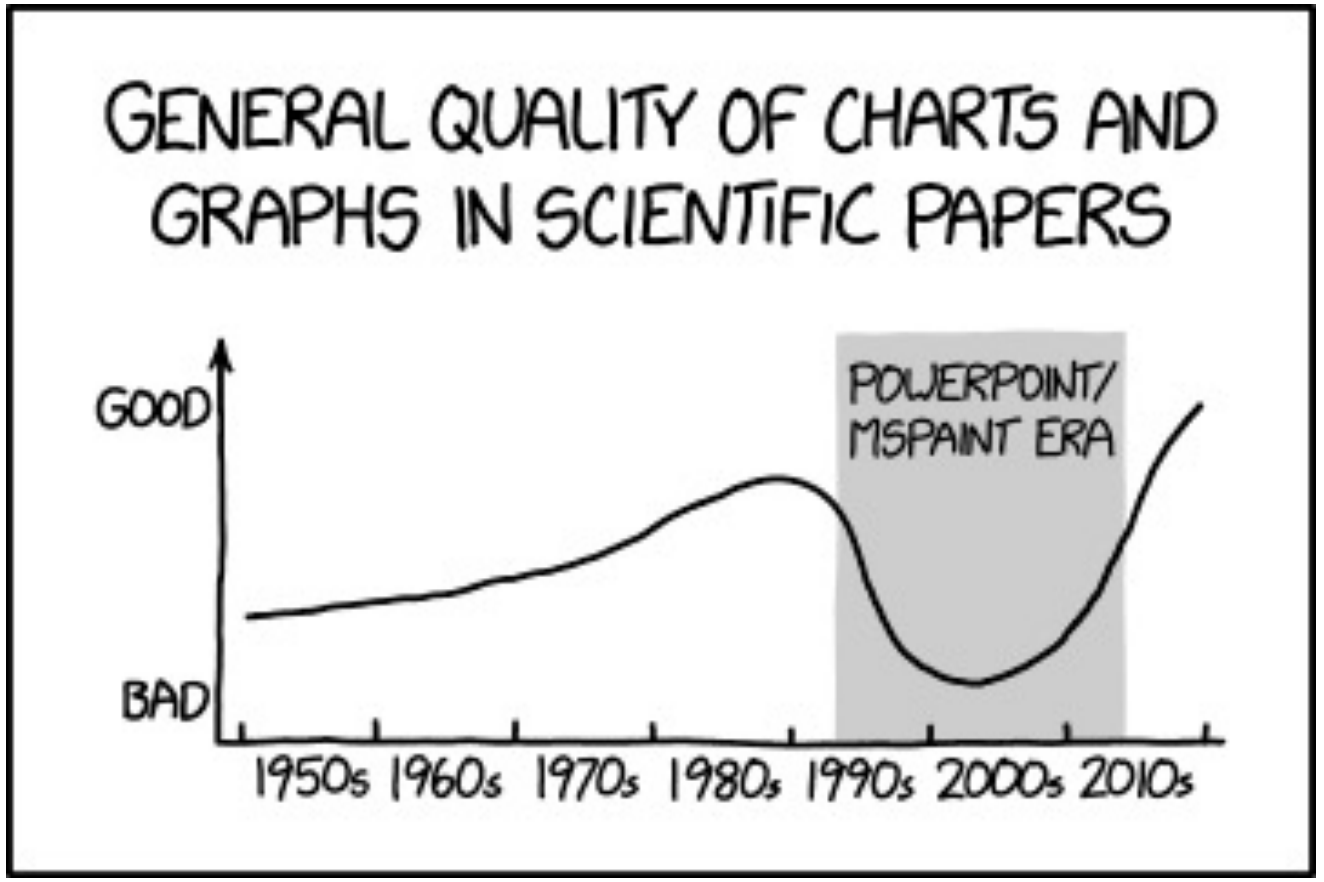
Fidelity, integrity, and sophistication: Edward Tufte's principles of data visualization

so like, how to not make crappy plots

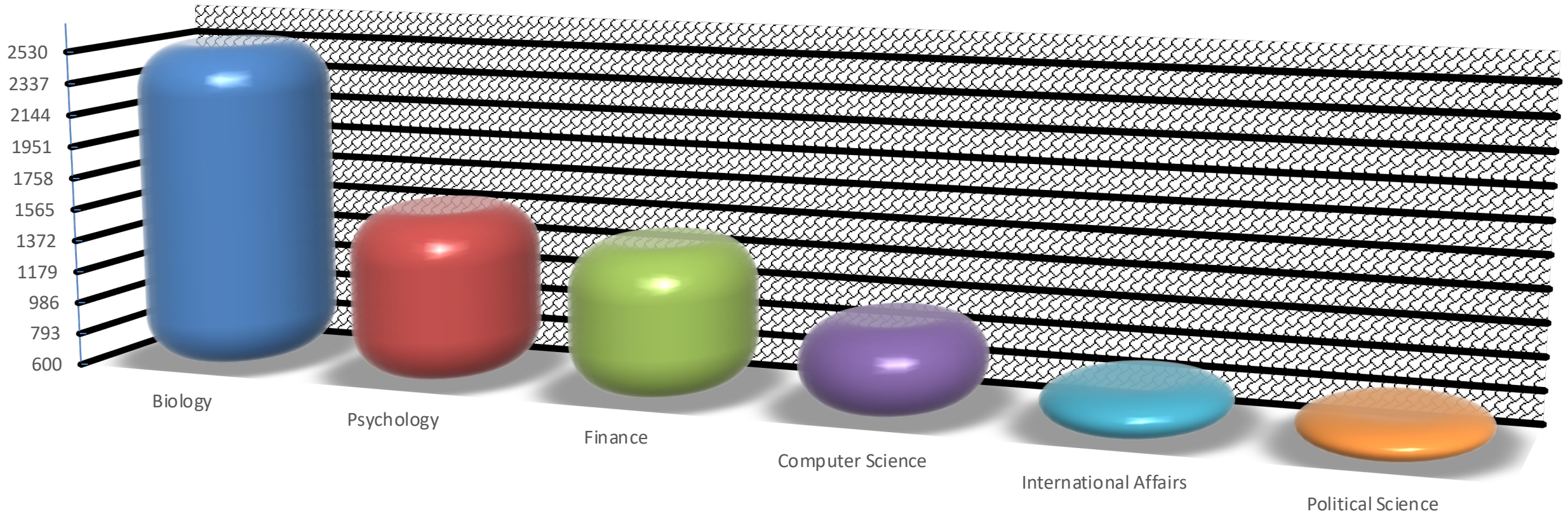
Joey Stanley
DigiLab Research Assistant
@joey_stan joeystanley.com

October 16, 2019

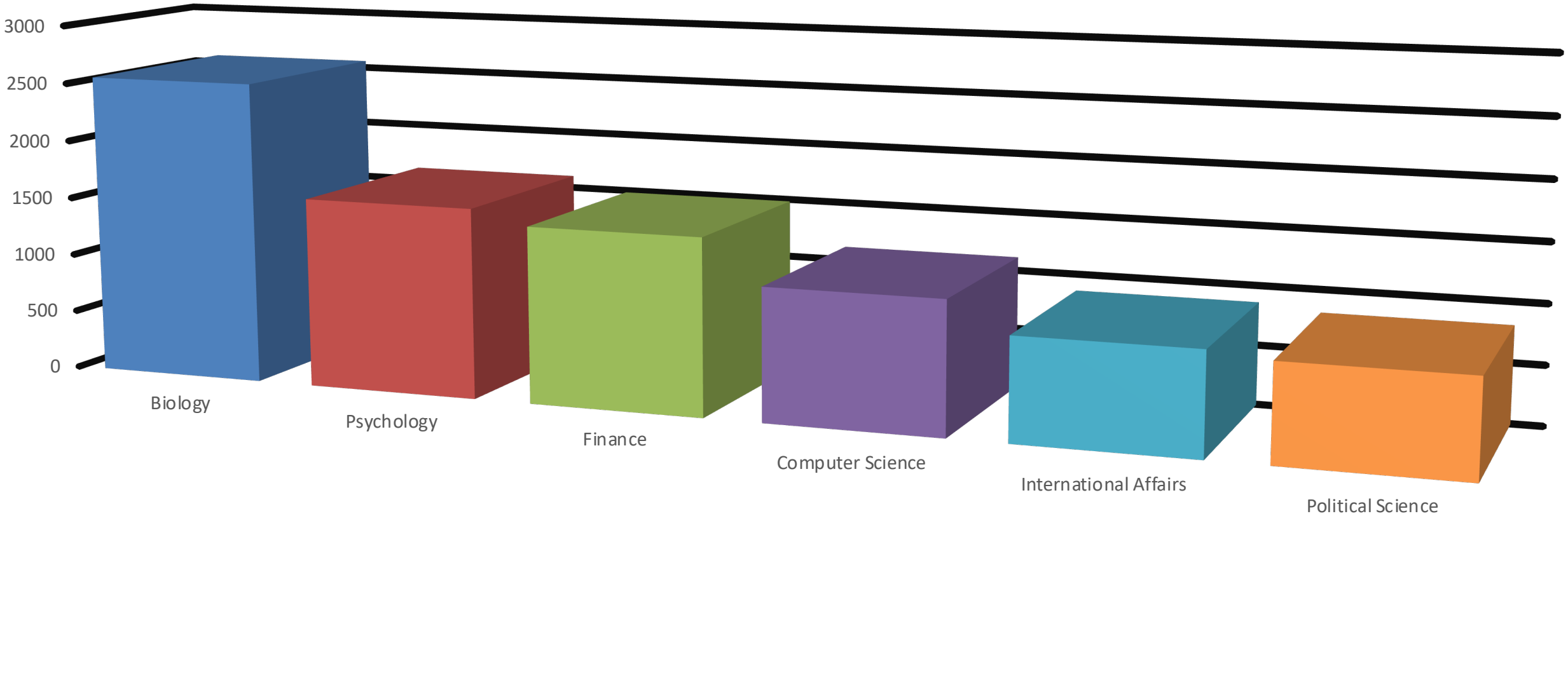


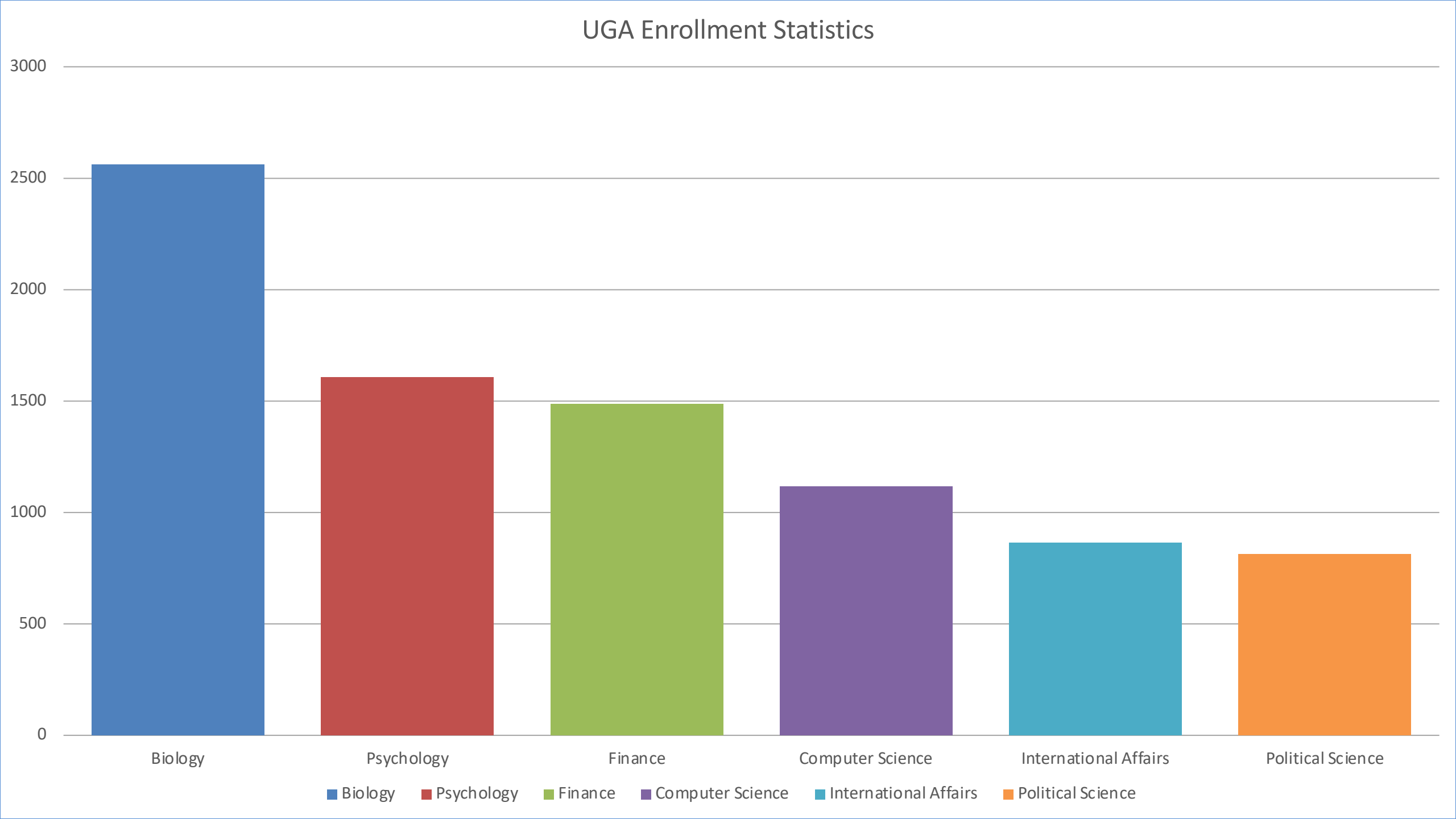
https://xkcd.com/1945/

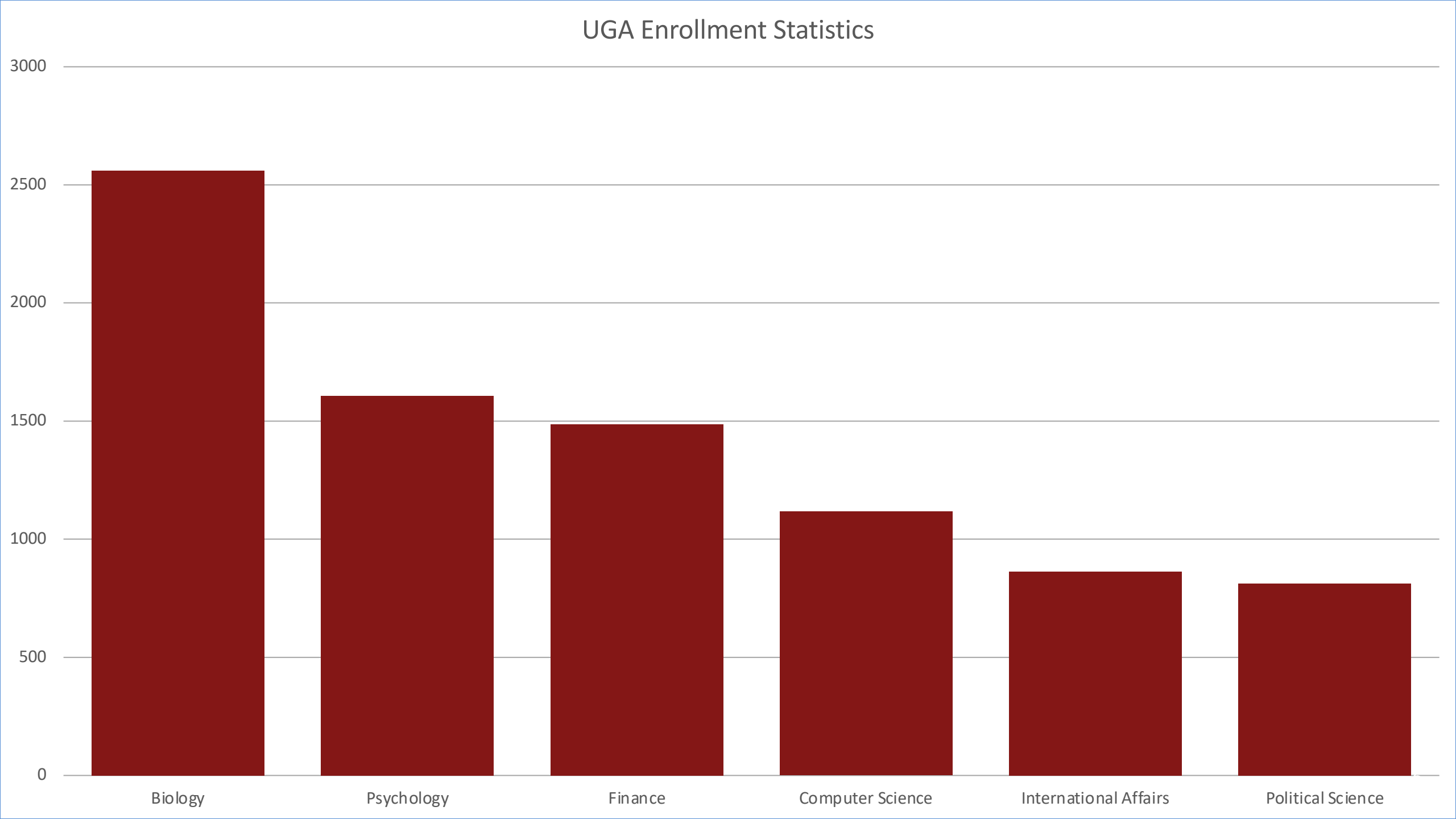
UGA Enrollment Statistics

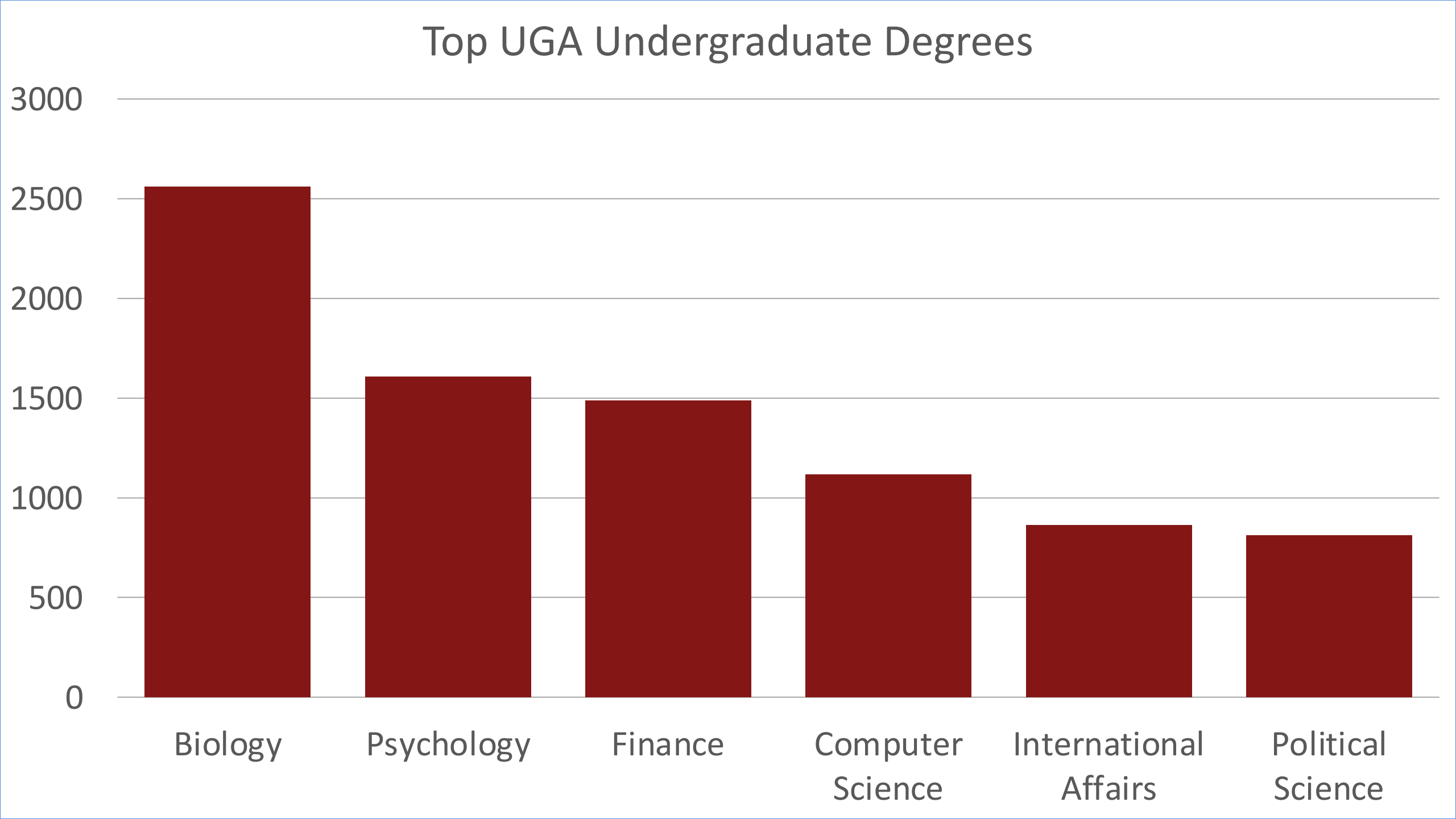


UGA Enrollment Statistics

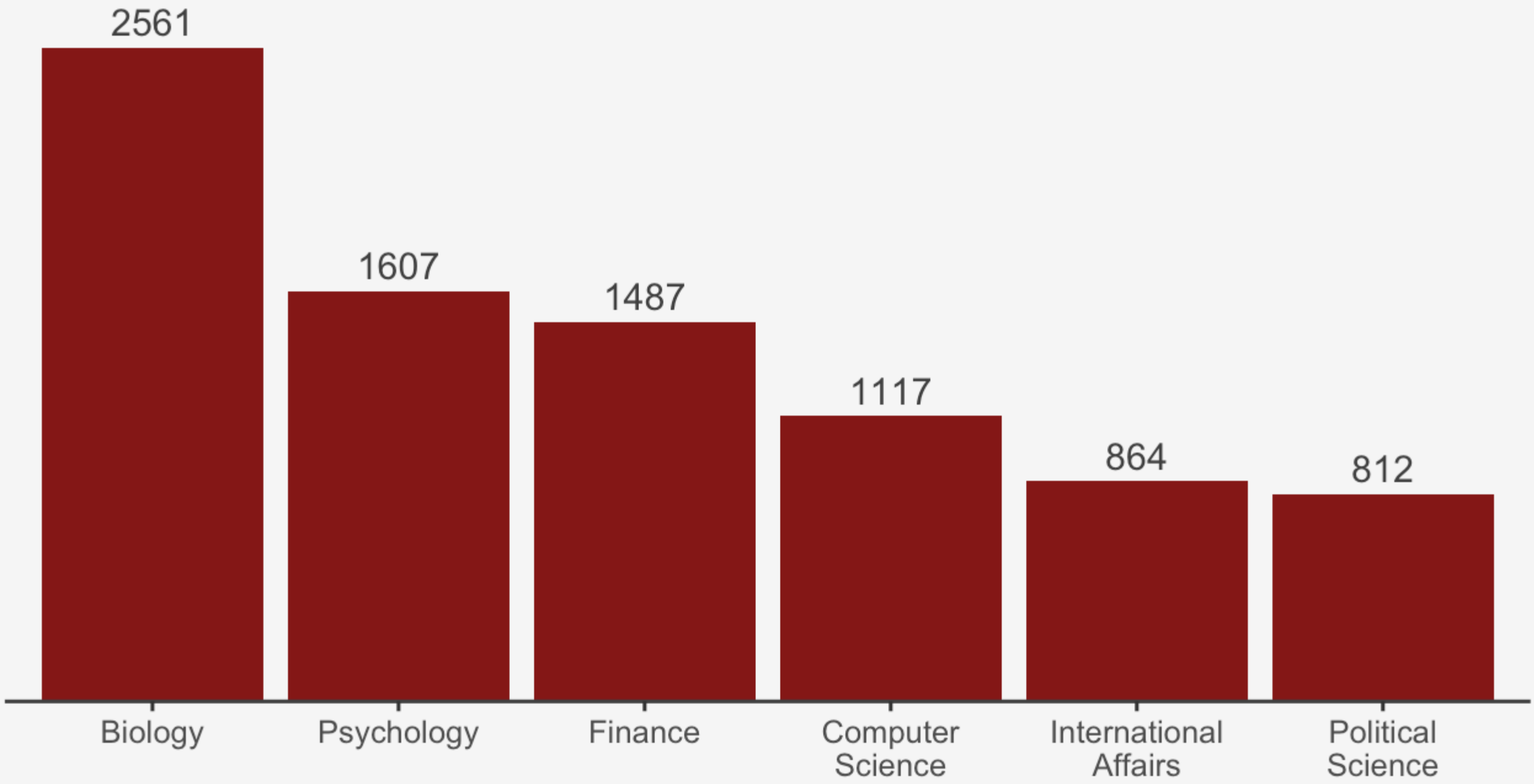



UGA Enrollment Statistics
3000
2500
2000
1500
1000
500
0
Biology
Psychology
Finance
Computer Science
International Affairs
Political Science
Biology
Psychology
Finance
Computer Science
International Affairs
Political Science


UGA Enrollment Statistics
3000
2500
2000
1500
1000
500
0
Biology
Psychology
Finance
Computer Science
International Affairs
Political Science


Top UGA Undergraduate Degrees
3000
2500
2000
1500
1000
500
0
Biology
Psychology
Finance
Computer Science
International Affairs
Political Science

Top UGA Undergraduate Degrees



Today's Presentation

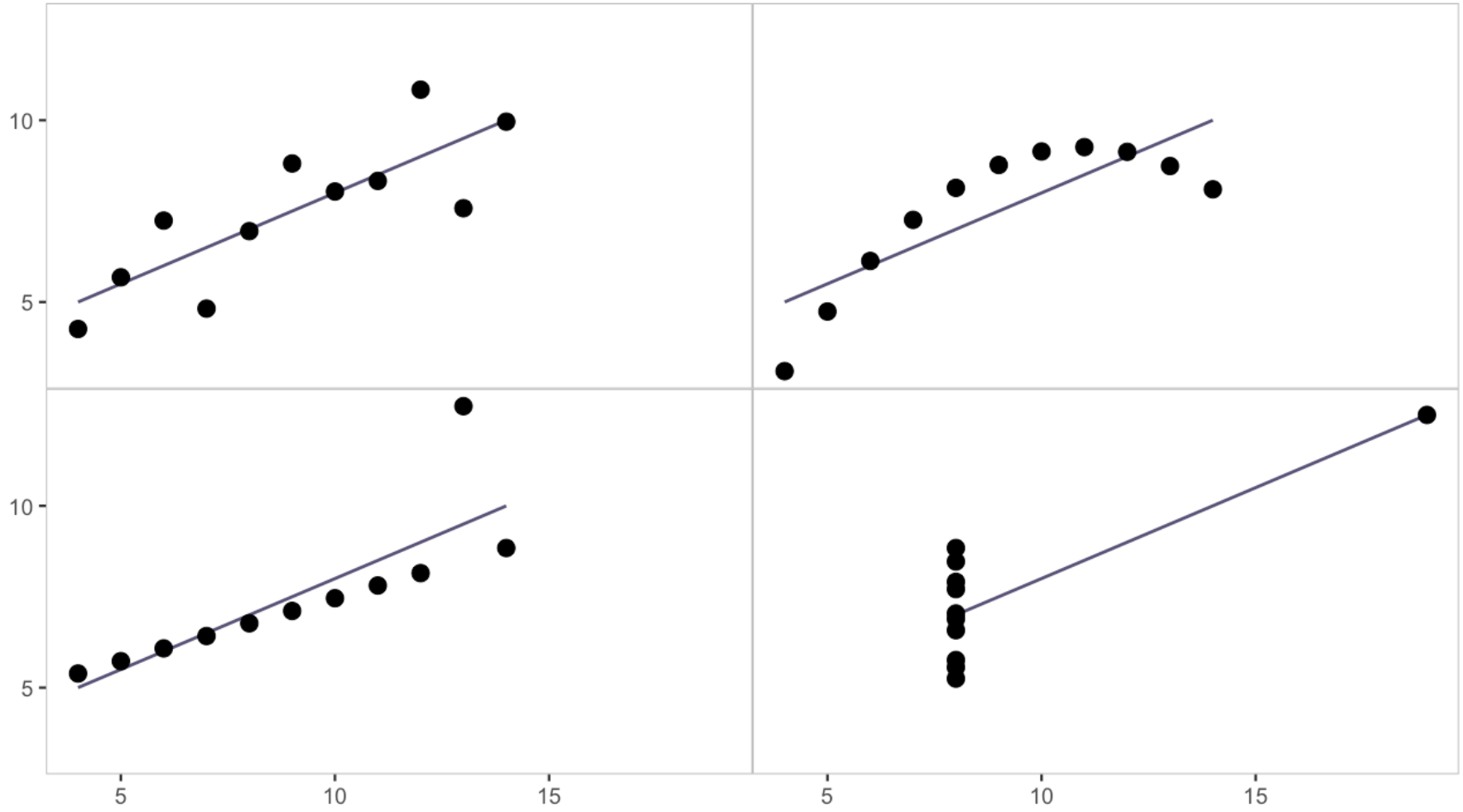
1. Purposes of data visualization
   - Why visualize data?
   - Who are data visualizations for?

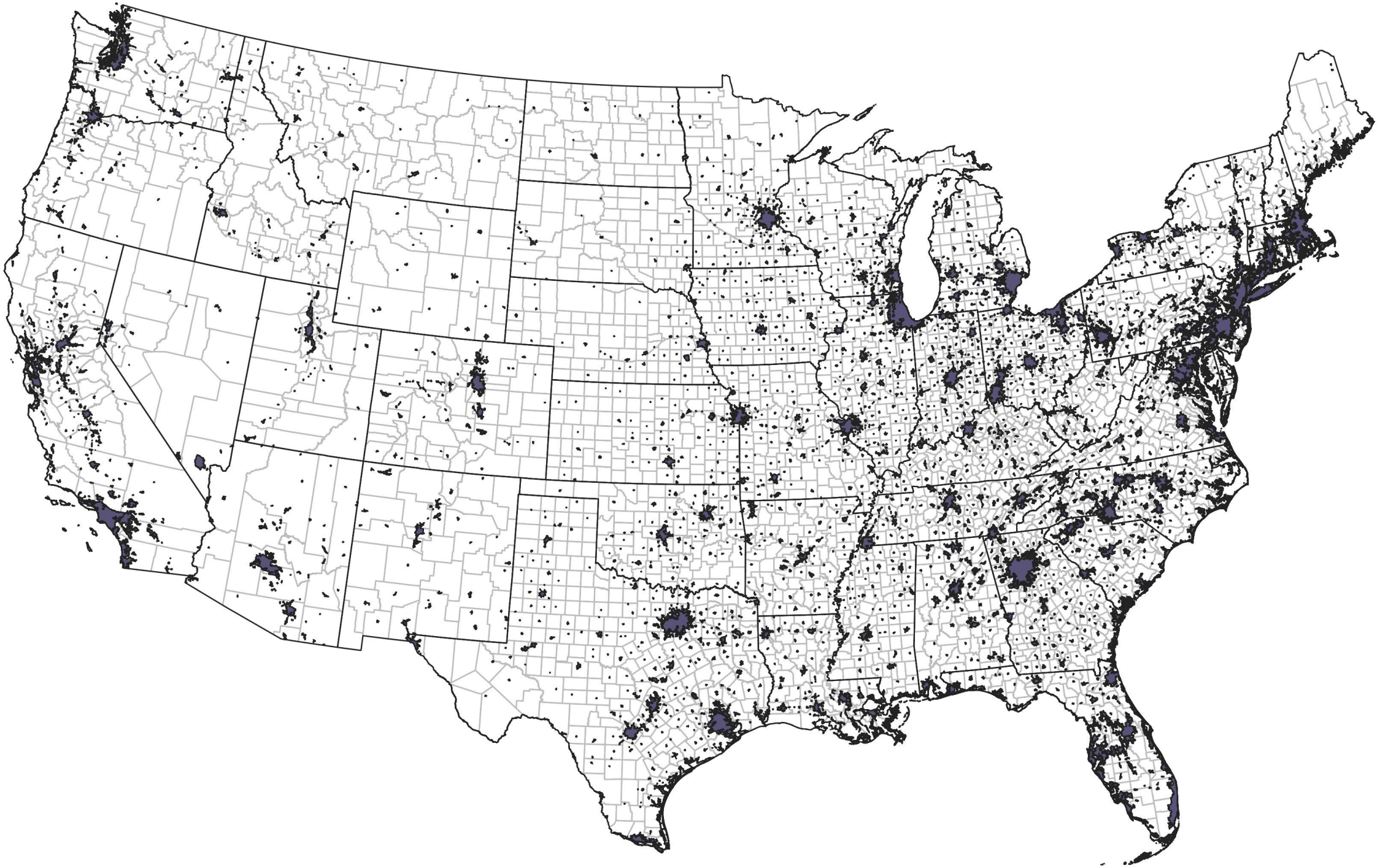
2. Graphical integrity & fidelity
   - (How to not lie and how to spot liars.)

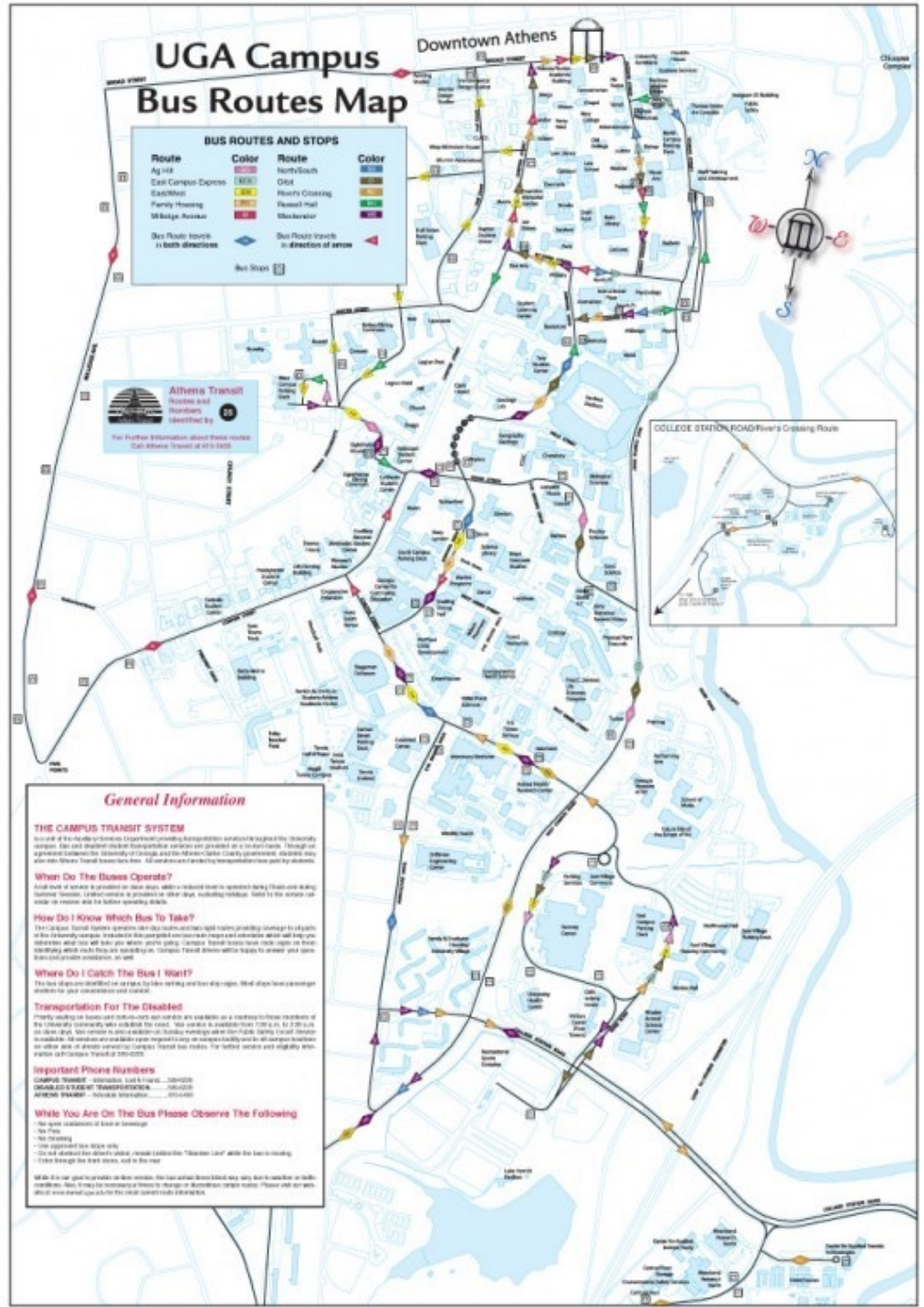
3. How to make plots look better
   - Remove chartjunk
   - Data-to-ink ratio and maximumization

Purposes of Data Visualization

Anscombe's Quartet






UGA Campus
Bus Routes Map
Downtown Athens
BUS ROUTES AND STOPS
Route
Color
Ag Hill
East Campus Express
East/West
Family Housing
Milledge Avenue
North/South
Orbit
River's Crossing
Russell Hall
Woodworker
Bus Route travels in both directions
Bus Route travels in direction of arrow
Bus Stops
Athens Transit
Routes and Numbers identified by
For Further Information about these routes Call Athens Transit at 613-3400
COLLEGE STATION ROAD/River's Crossing Route
General Information
THE CAMPUS TRANSIT SYSTEM
When Do The Buses Operate?
How Do I Know Which Bus To Take?
Where Do I Catch The Bus I Want?
Transportation For The Disabled
Important Phone Numbers
While You Are On The Bus Please Observe The Following

Who is Data Visualization for?

Audience 1: You

Reasons

- Familiarize yourself with your data
- Analyzing model output

Properties

- Quick and dirty; unpolished
- Easy to switch from plot to plot

Audience 2: Not You

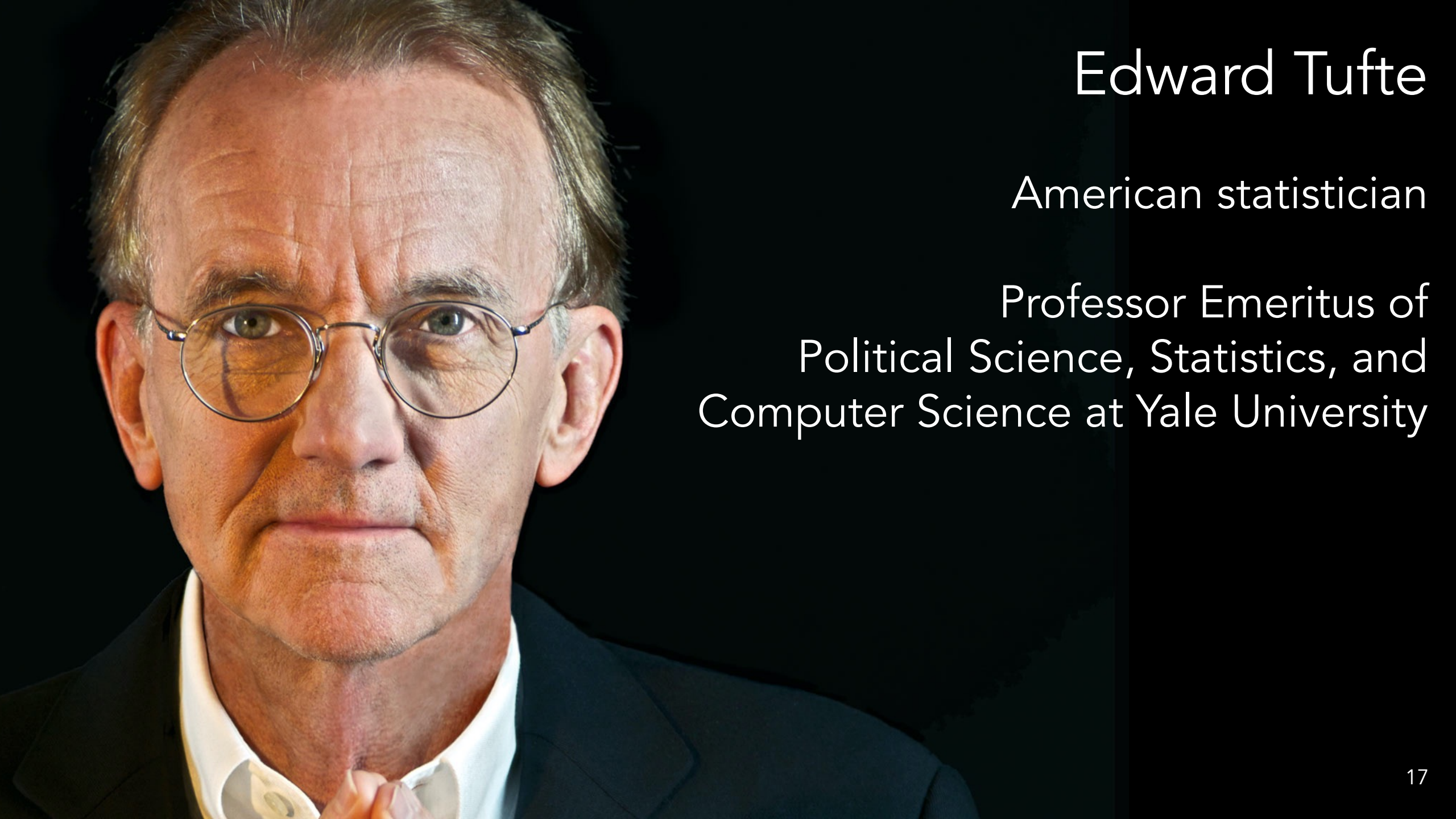
Reasons

- Evidence for your hypothesis
- Tell a story

Properties

- Sophisticated, clean, polished
- Easily interpretable

Graphical Integrity and Fidelity

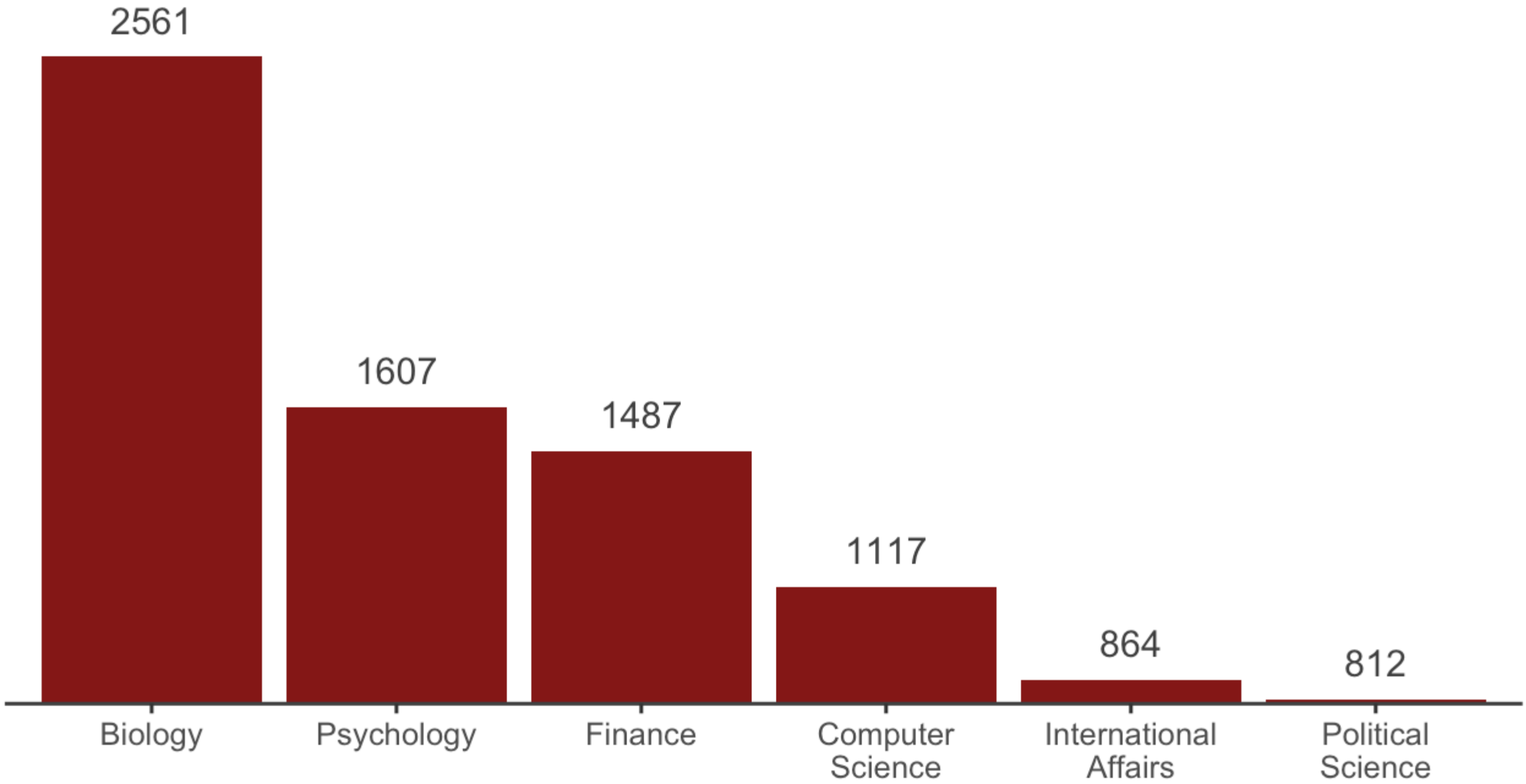


Edward Tufte

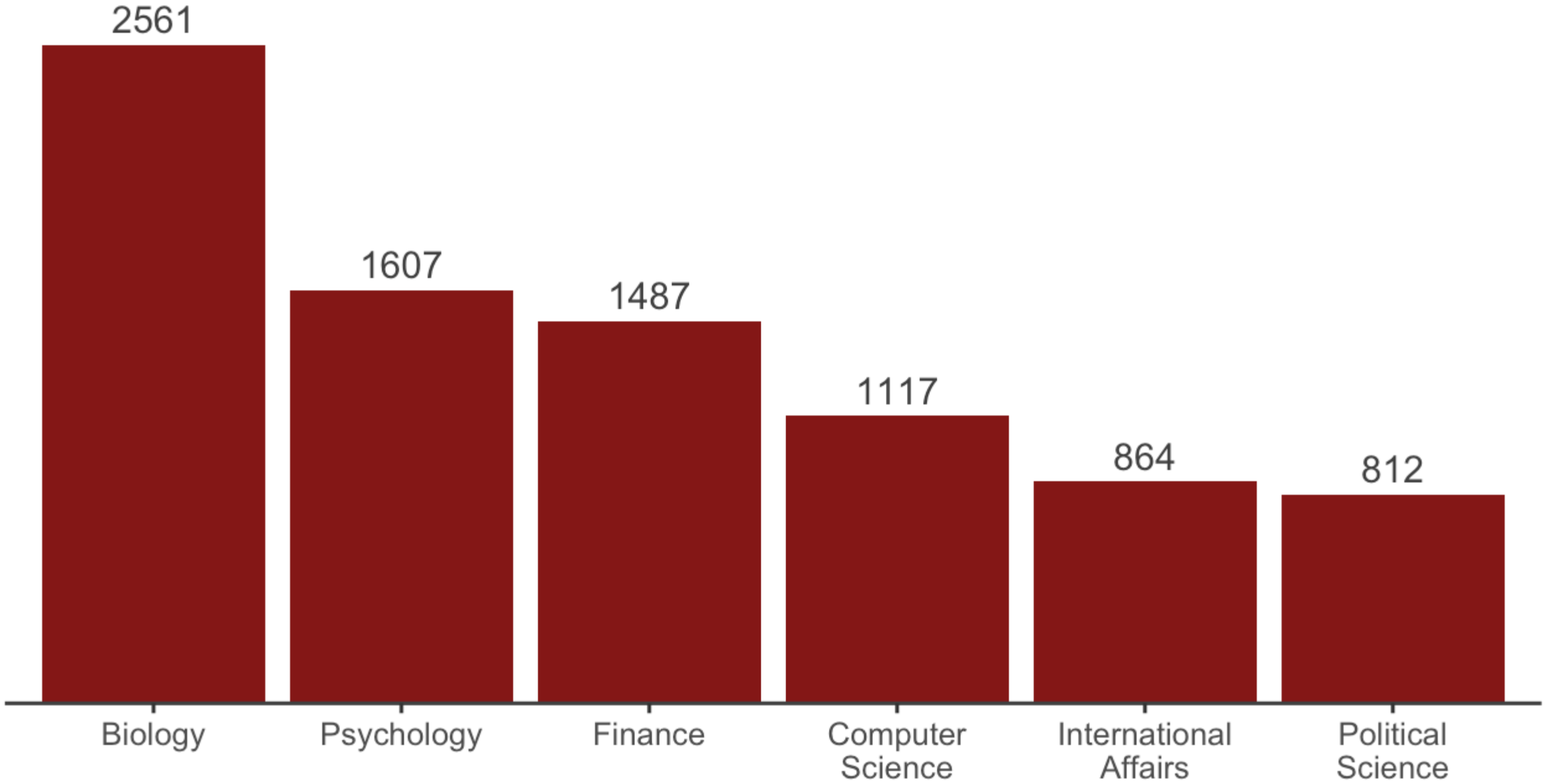
American statistician

Professor Emeritus of
Political Science, Statistics, and
Computer Science at Yale University

Top UGA Undergraduate Degrees



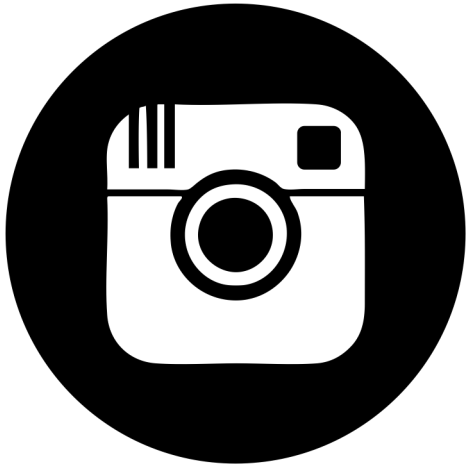
Top UGA Undergraduate Degrees



Facebook's Dominance over Social Media

Facebook
2.23b

YouTube
1.9b



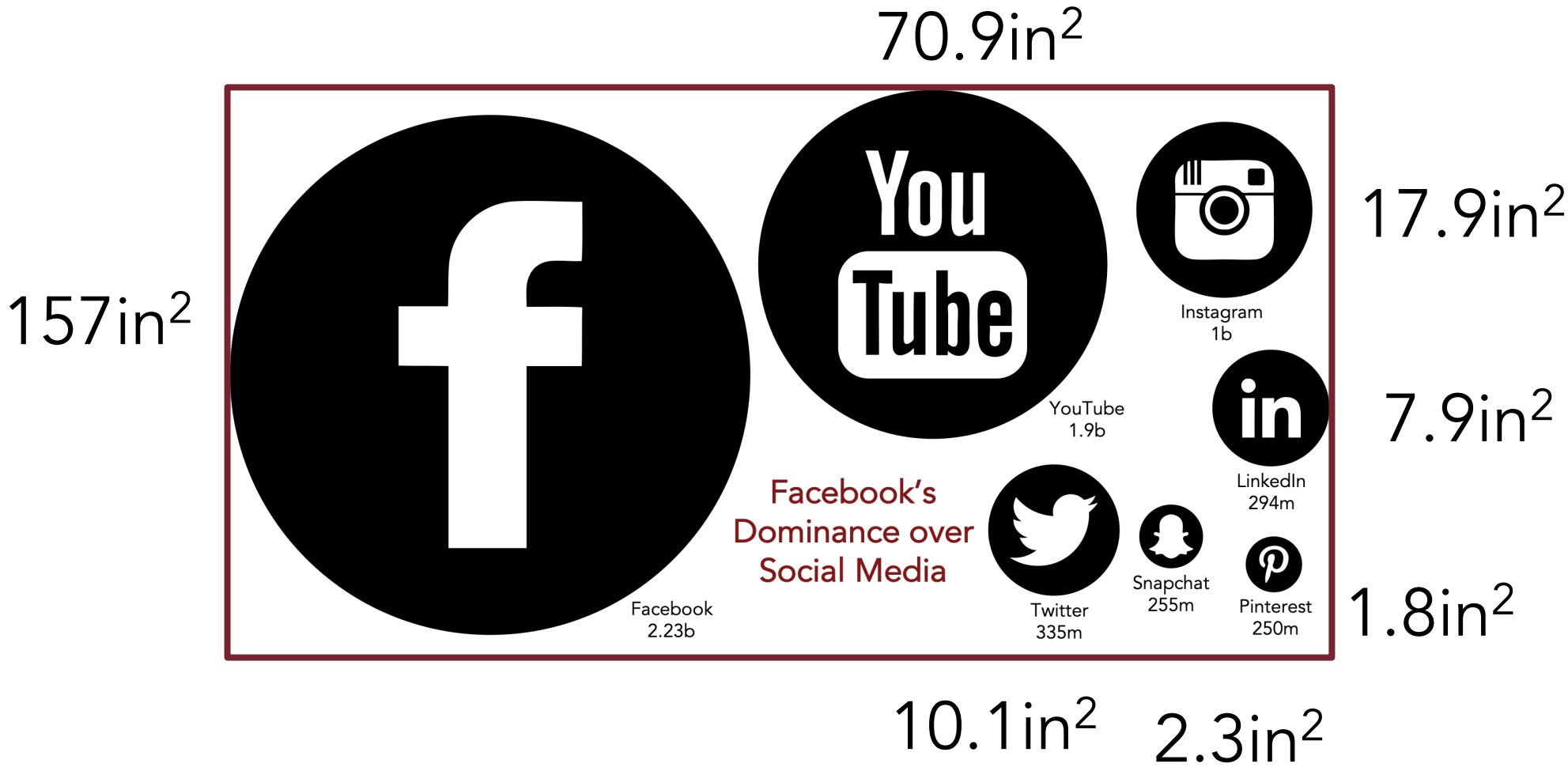
Instagram
1b

LinkedIn
294m

Twitter
335m

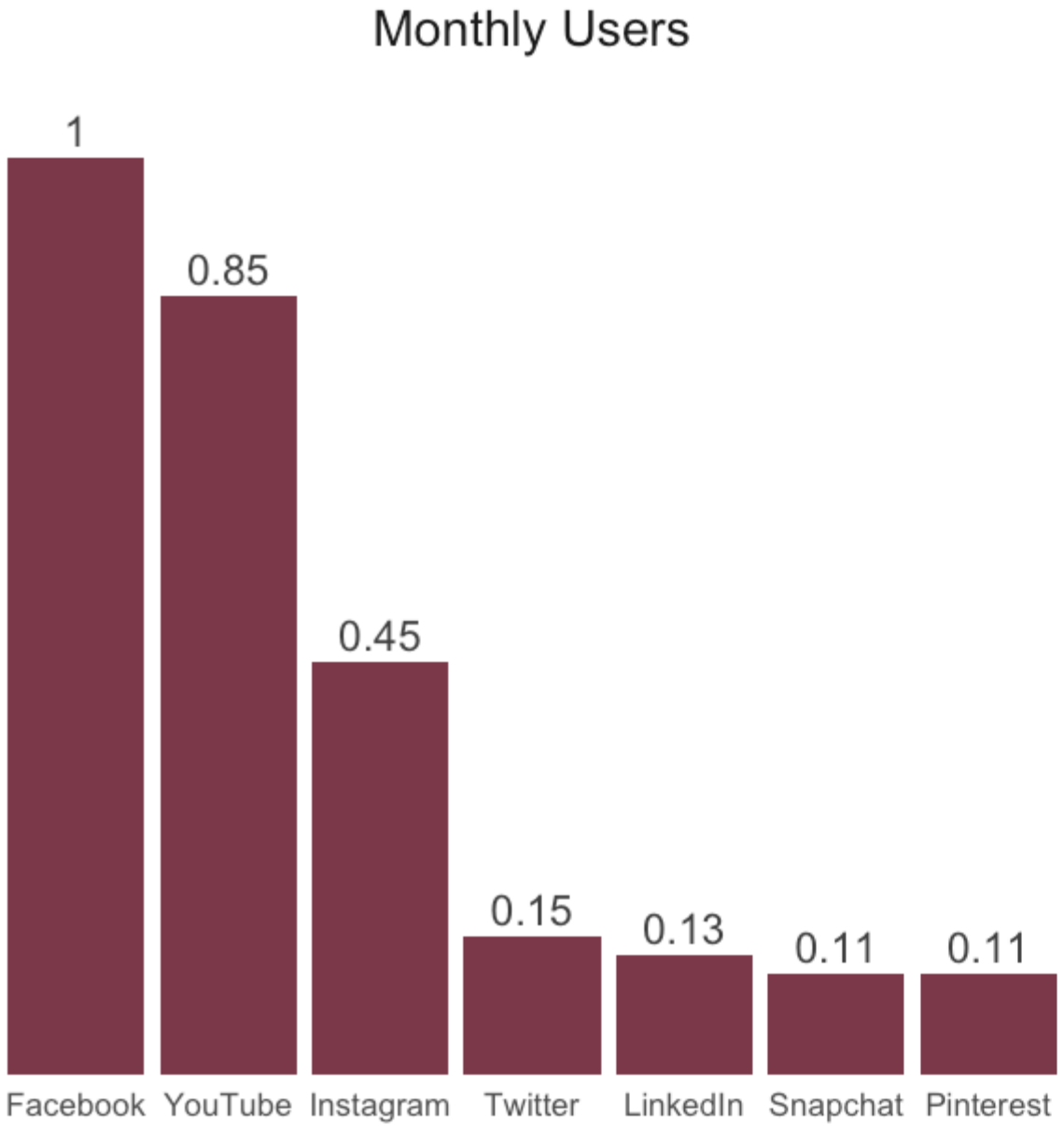
Snapchat
255m

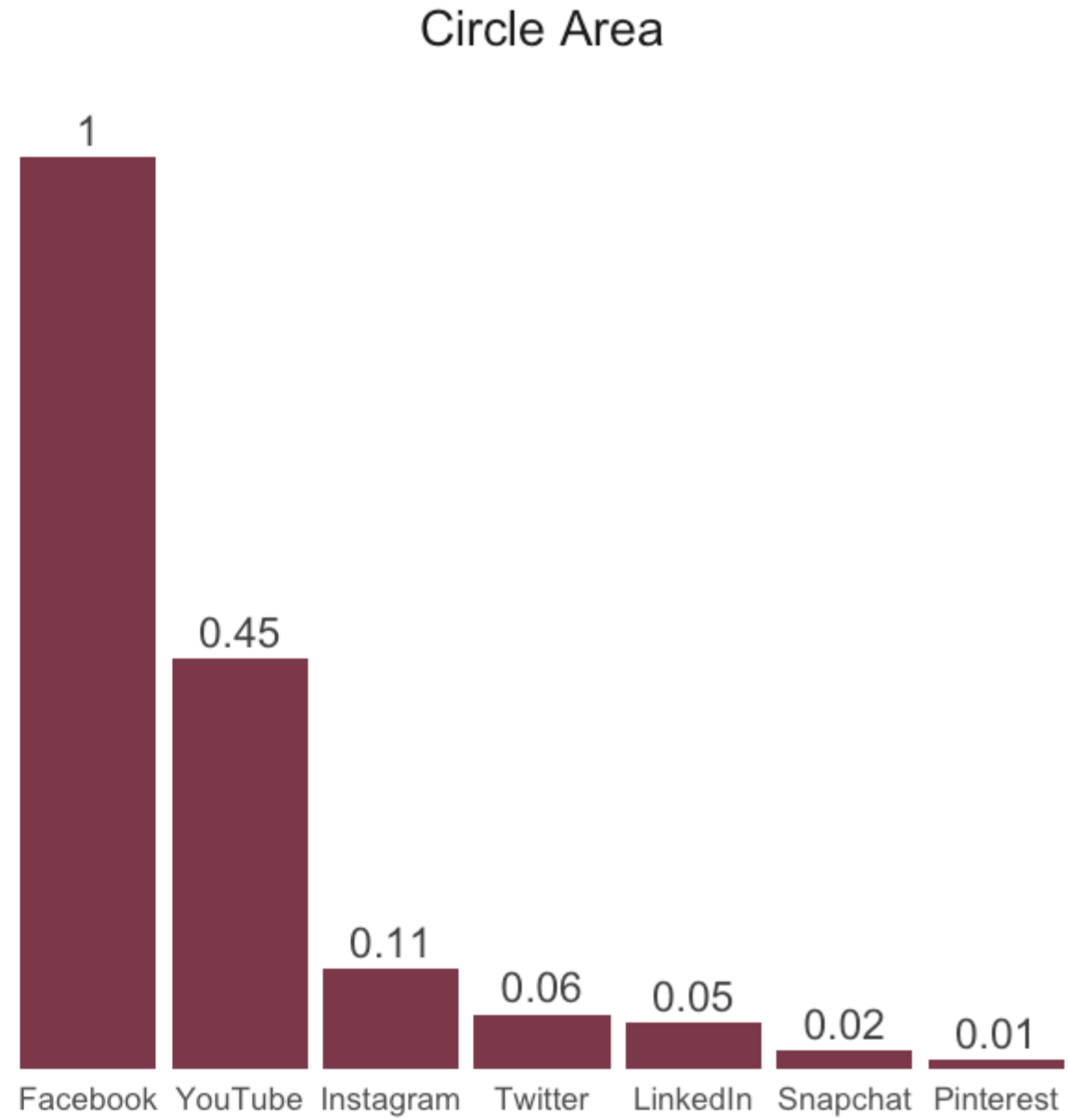
Pinterest
250m


70.9in²
157in²
17.9in²
7.9in²
1.8in²
10.1in²
2.3in²
Facebook's Dominance over Social Media
Facebook 2.23b
YouTube 1.9b
Instagram 1b
LinkedIn 294m
Twitter 335m
Snapchat 255m
Pinterest 250m

Proportion of Users compared to Facebook

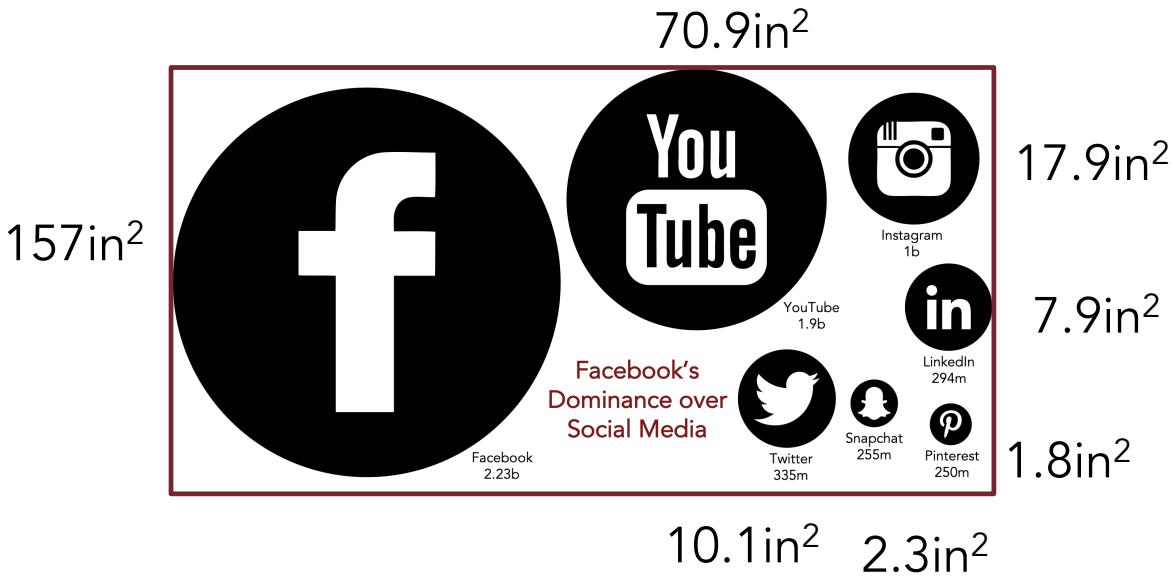
Comparing actual data to the circle size in the previous graphic

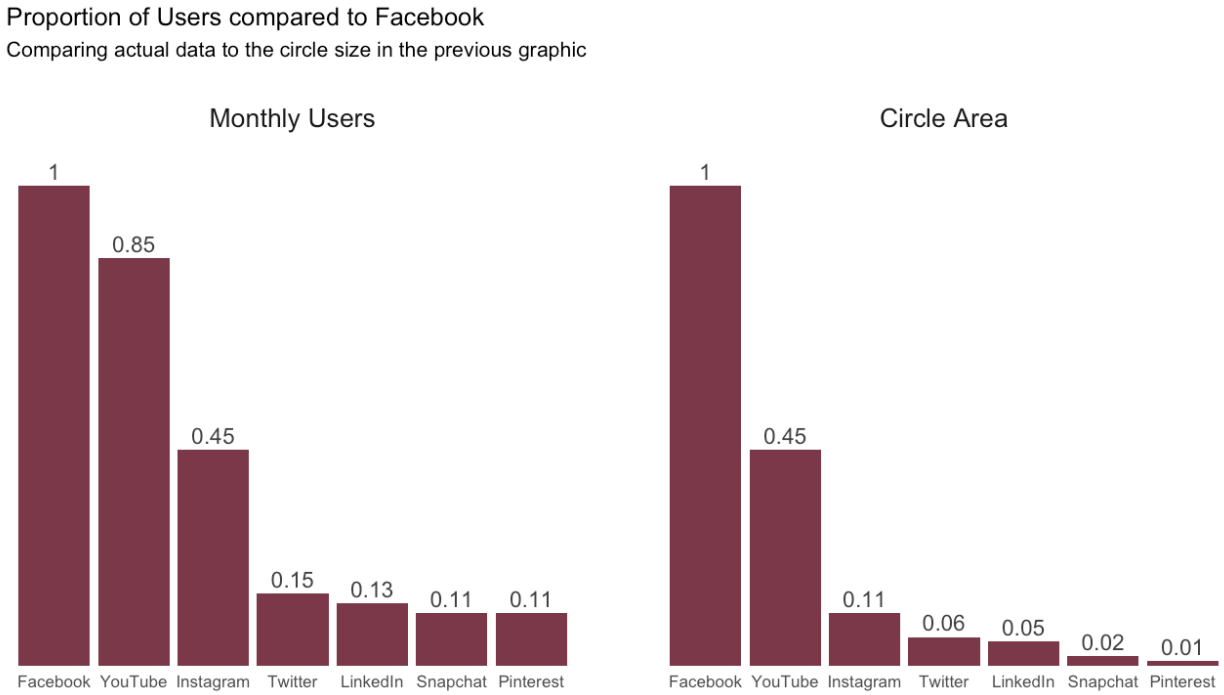




$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

Lie Factor $_{\text{Facebook to YouTube}} = 0.52$ Lie Factor $_{\text{Facebook to Pinterest}} = 0.09$

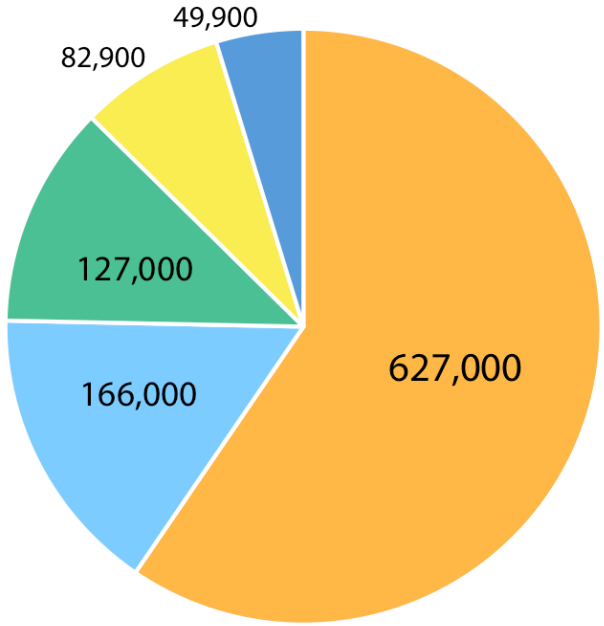


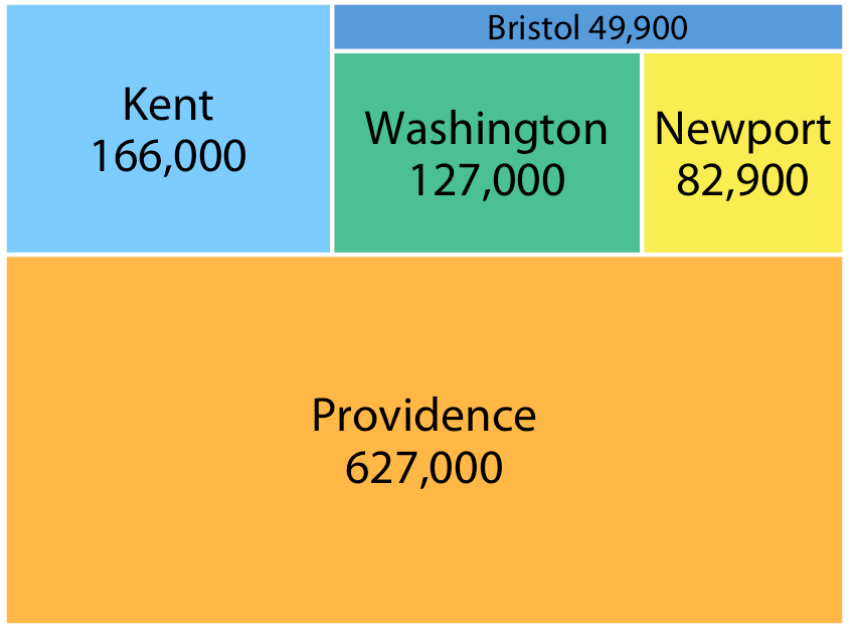


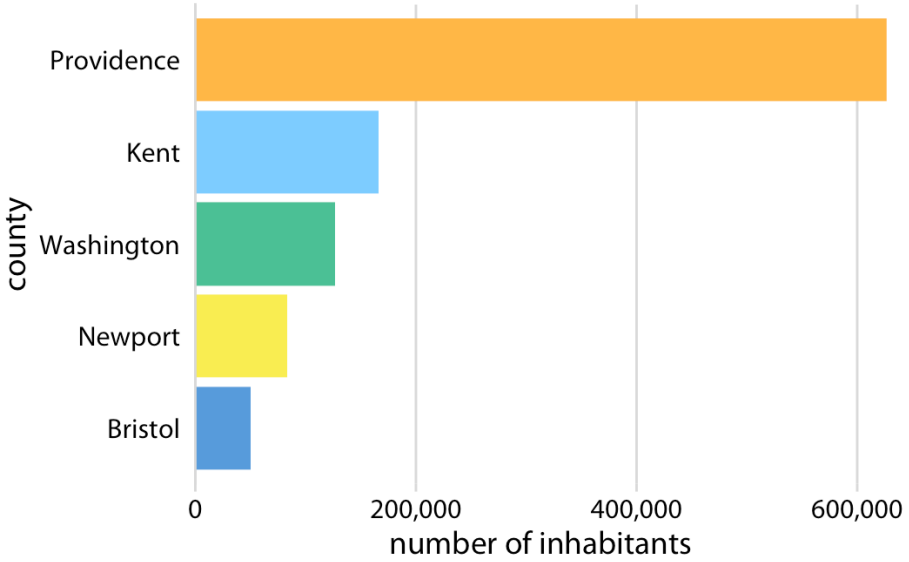
"The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented."

Edward Tufte
The Visual Display of Quantitative Information p. 56

Number of Inhabitants in Rhode Island counties



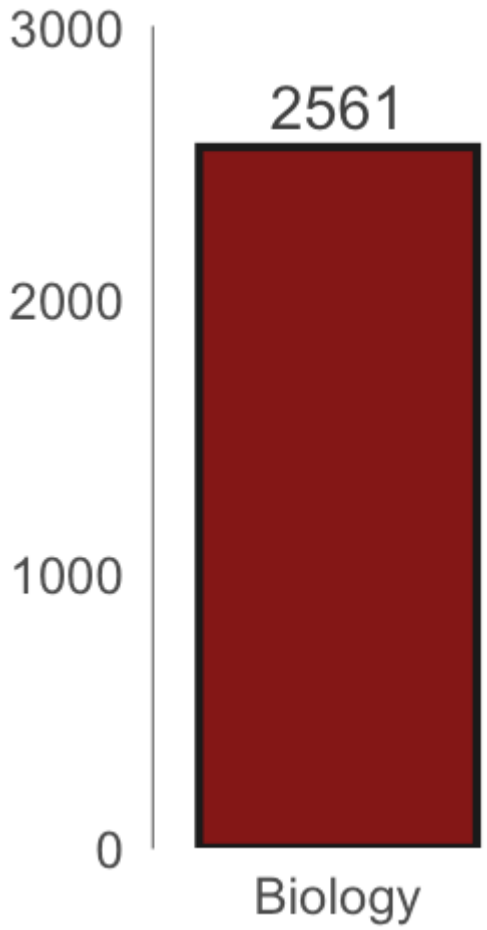




Claus O. Wilke, *Fundamentals of Data Visualization* p 216–217

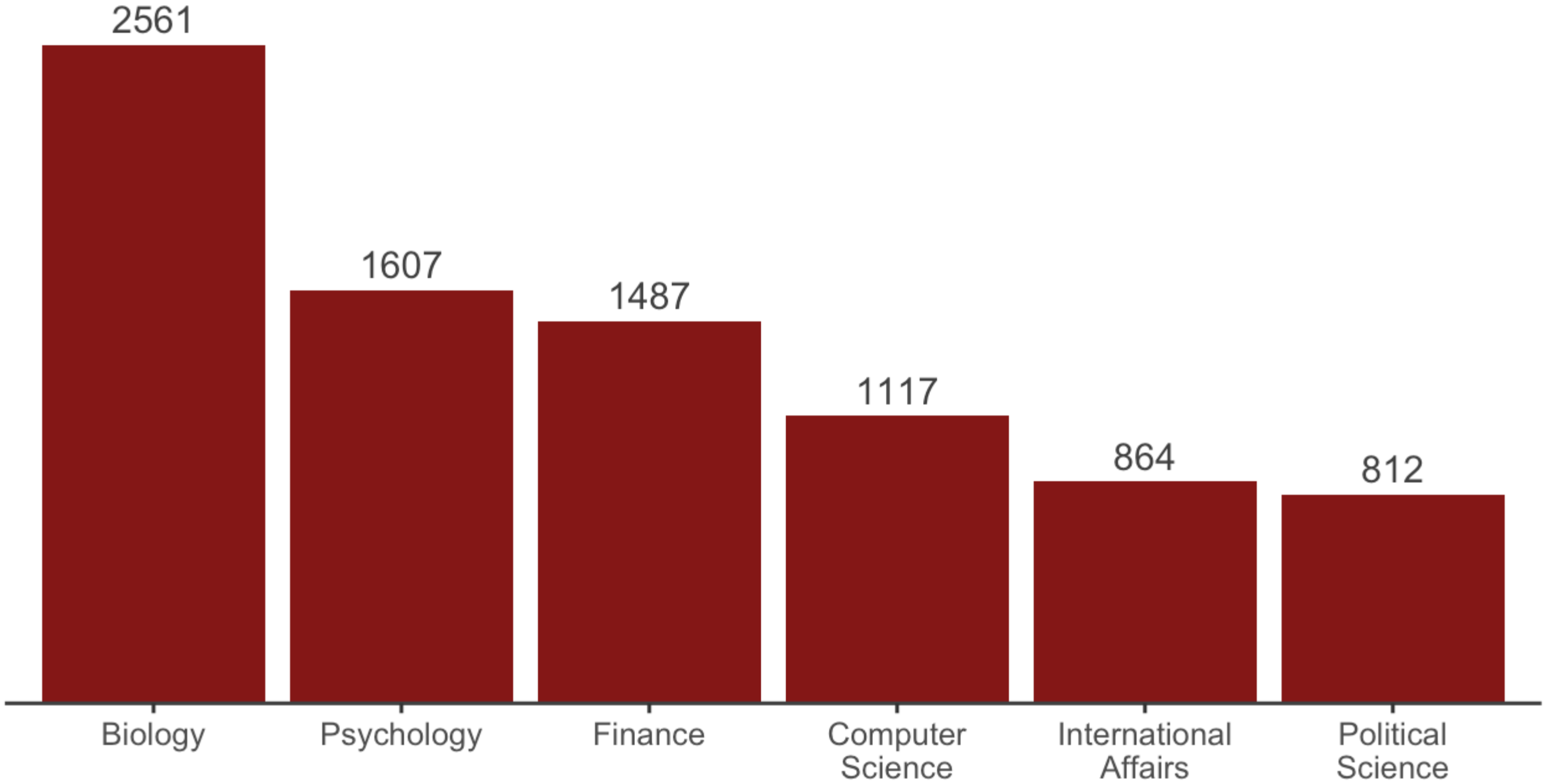
Making Plots Look Better

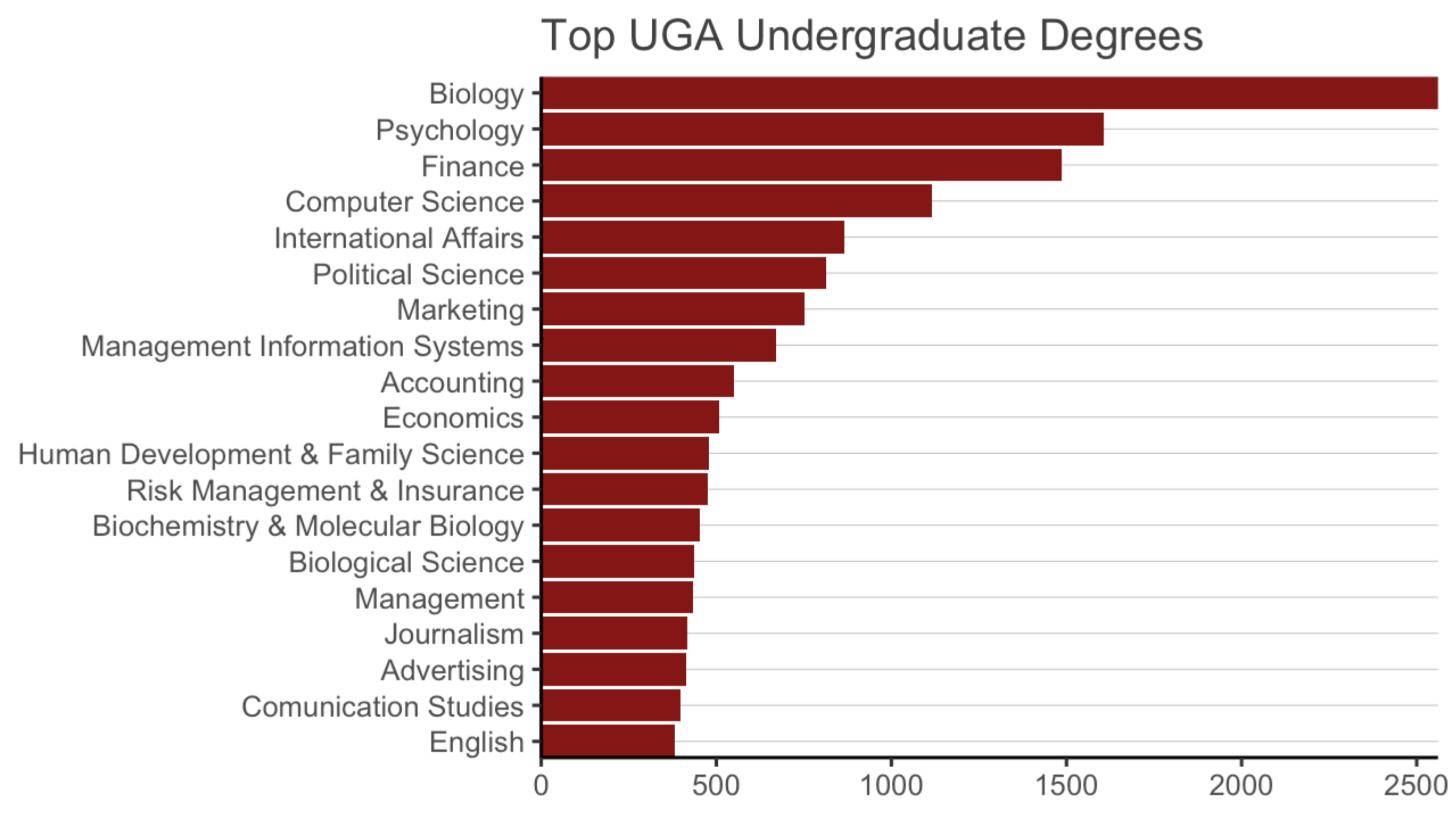
How is quantity represented?

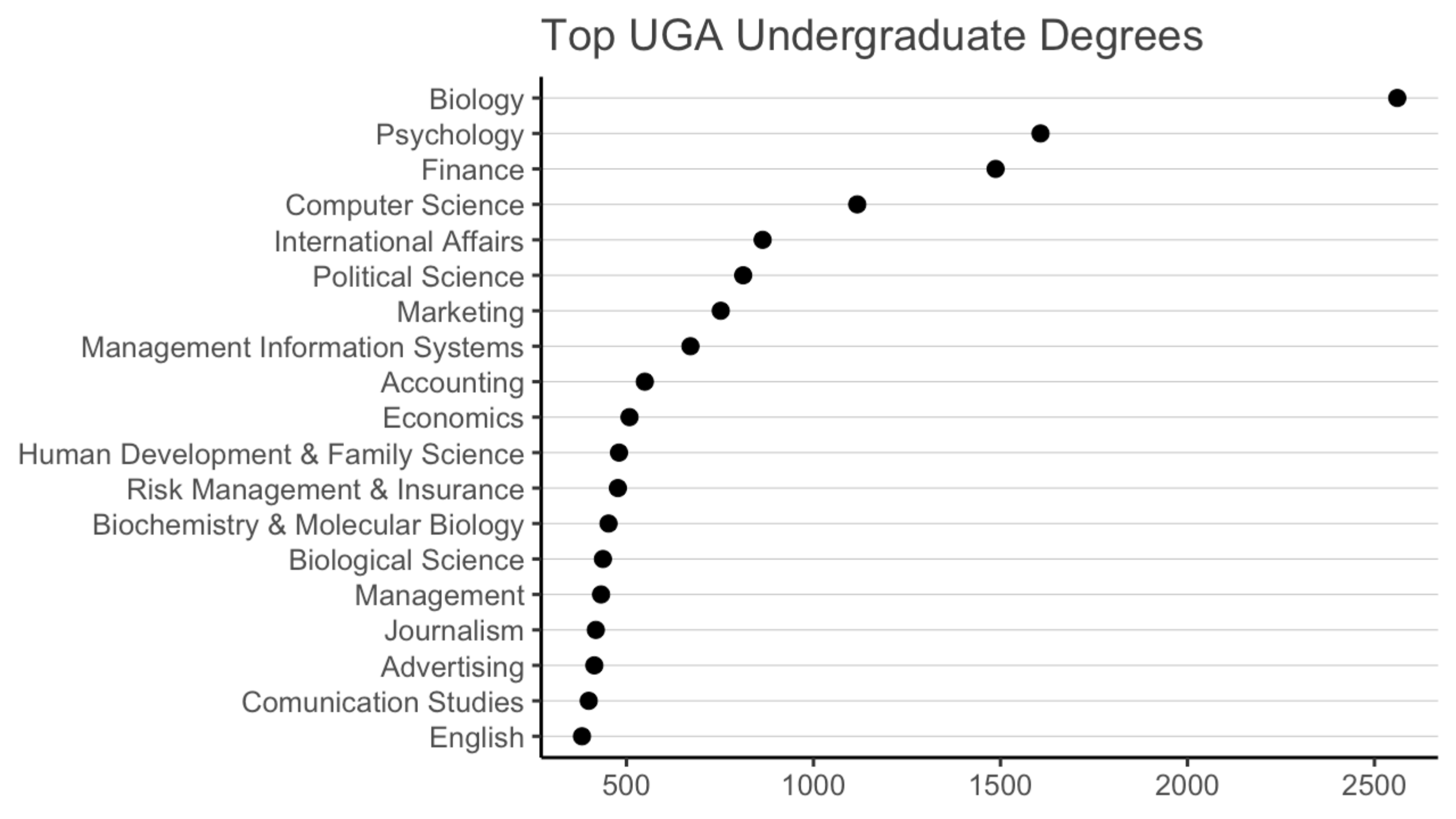


How is quantity represented?

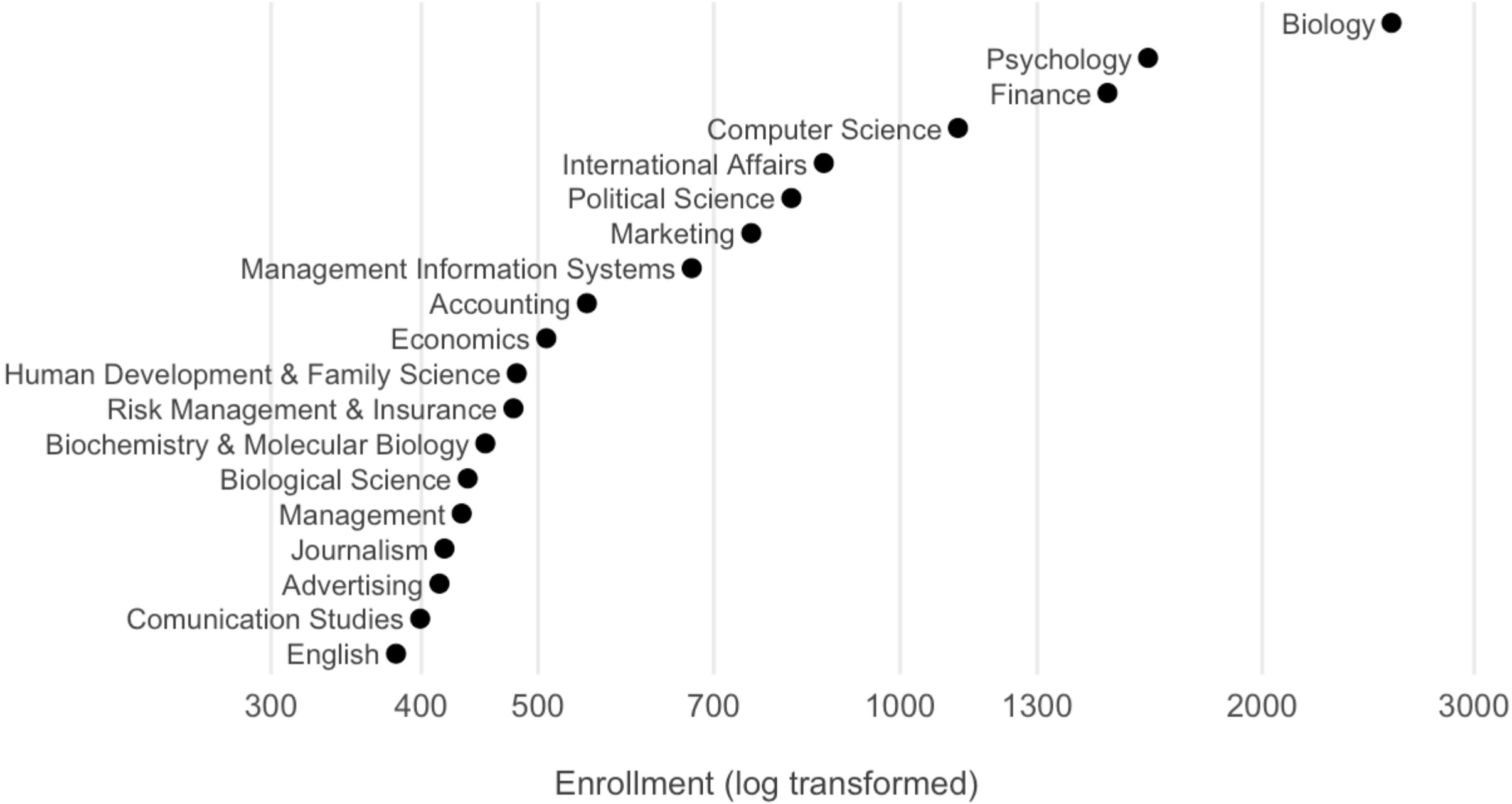
Top UGA Undergraduate Degrees




Top UGA Undergraduate Degrees
Biology
Psychology
Finance
Computer Science
International Affairs
Political Science
Marketing
Management Information Systems
Accounting
Economics
Human Development & Family Science
Risk Management & Insurance
Biochemistry & Molecular Biology
Biological Science
Management
Journalism
Advertising
Comunication Studies
English
0
500
1000
1500
2000
2500


Top UGA Undergraduate Degrees
Biology
Psychology
Finance
Computer Science
International Affairs
Political Science
Marketing
Management Information Systems
Accounting
Economics
Human Development & Family Science
Risk Management & Insurance
Biochemistry & Molecular Biology
Biological Science
Management
Journalism
Advertising
Comunication Studies
English
500
1000
1500
2000
2500

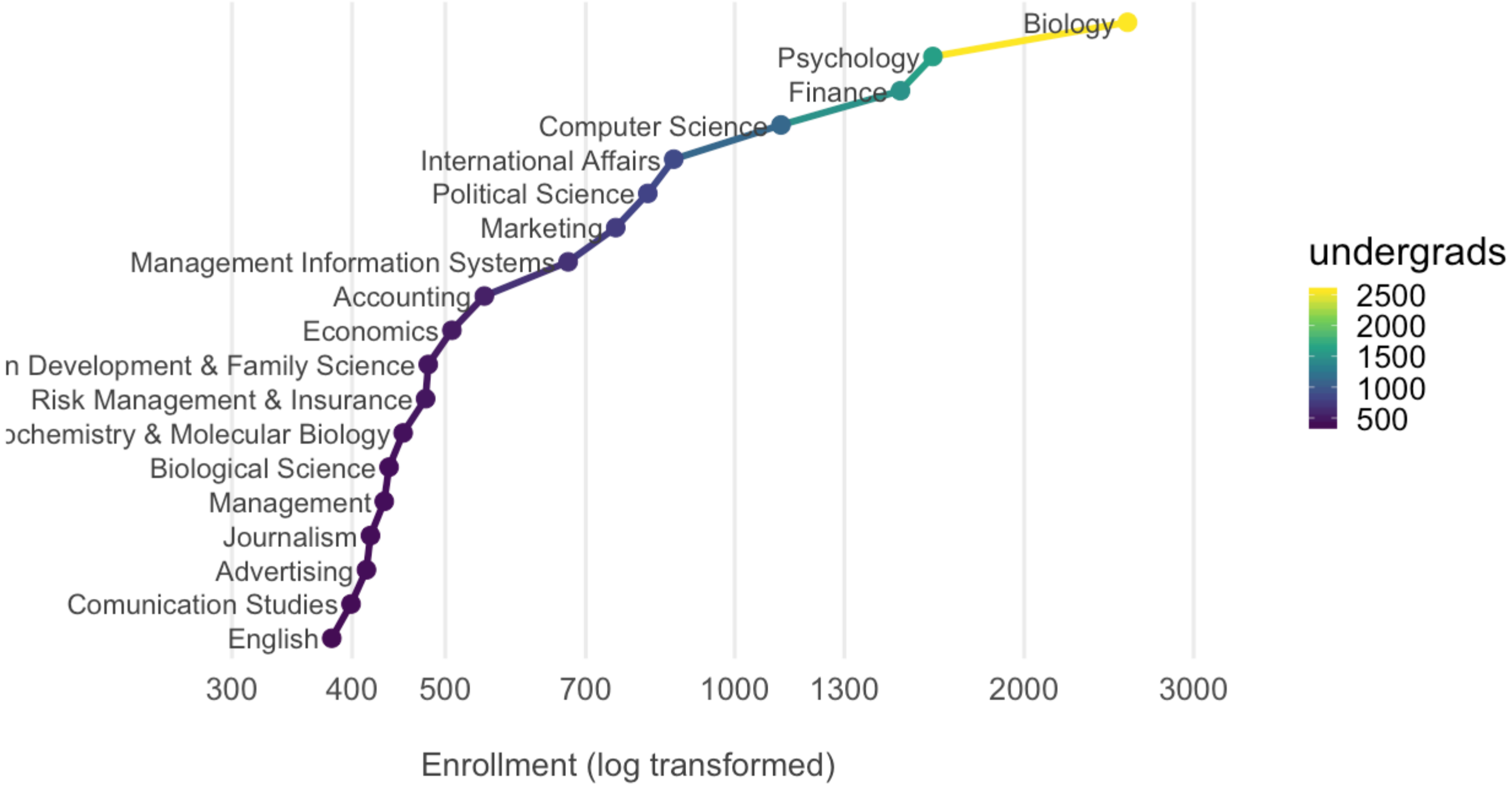
Top UGA Undergraduate Degrees



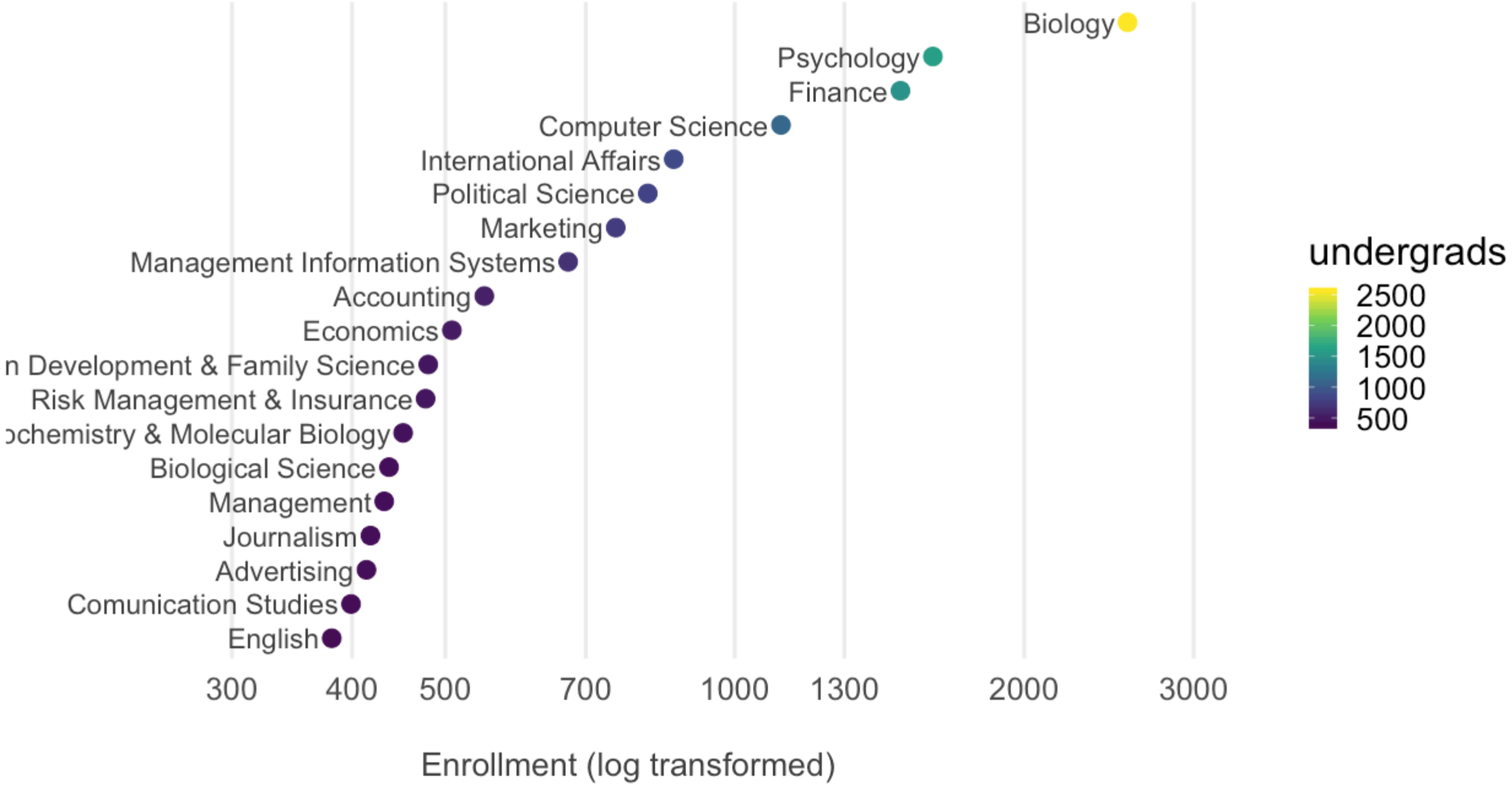
Above all else show the data.
Maximize the data-ink ratio.
Erase non-data-ink.
Erase redundant data-ink.
Revise and edit.

Edward Tufte
The Visual Display of Quantitative Information p. 105

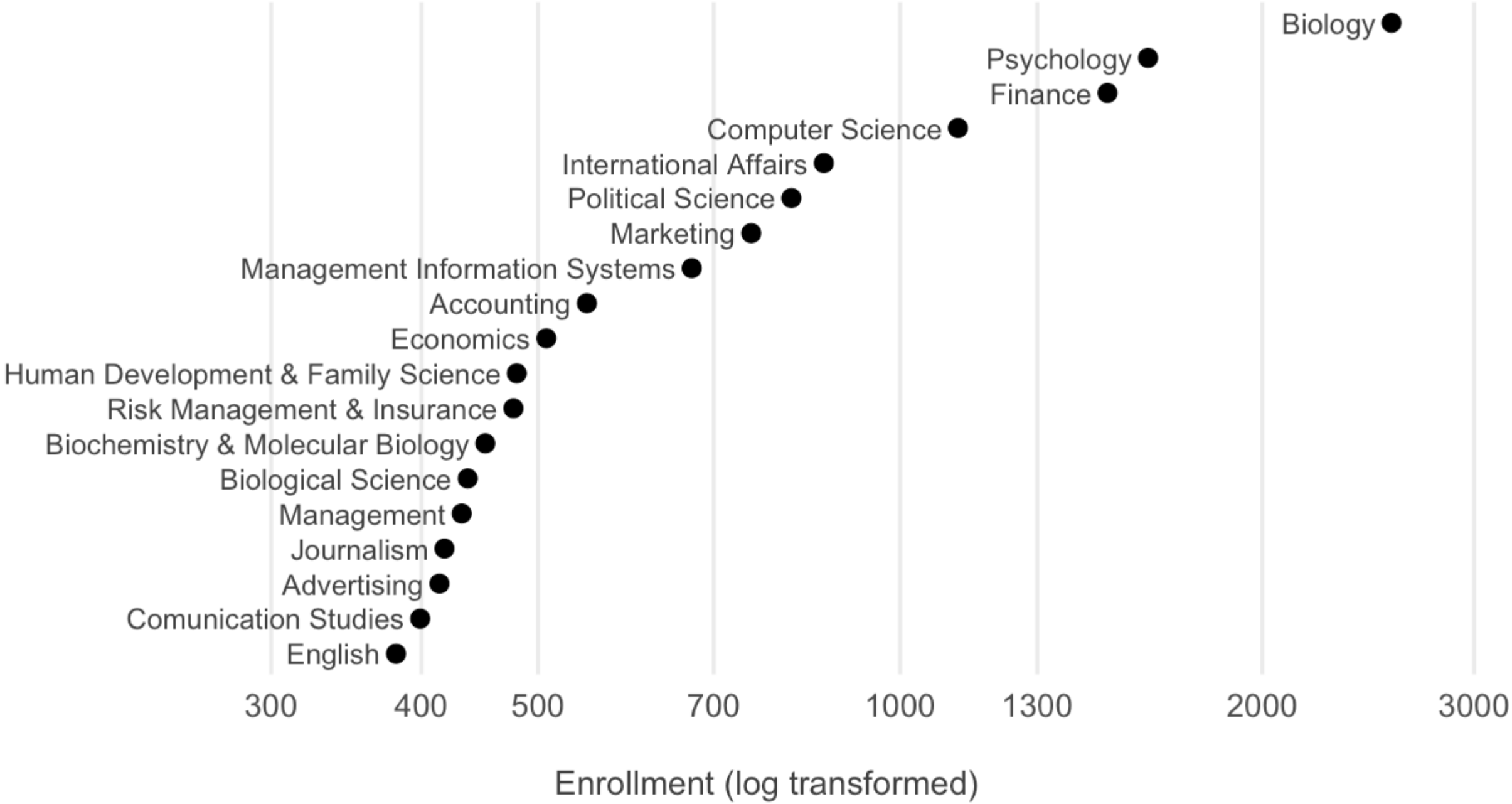
Top UGA Undergraduate Degrees



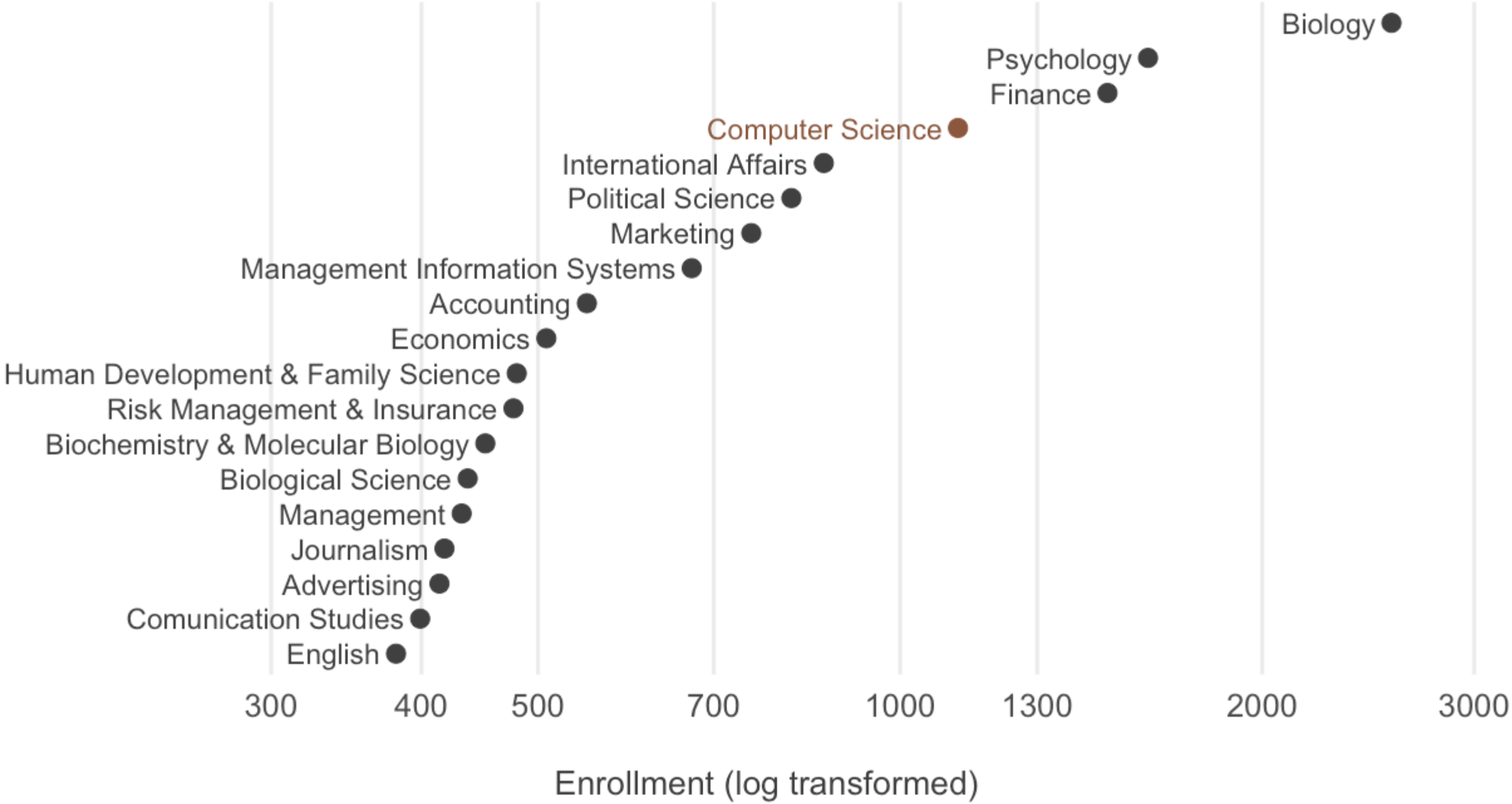
Top UGA Undergraduate Degrees



Top UGA Undergraduate Degrees



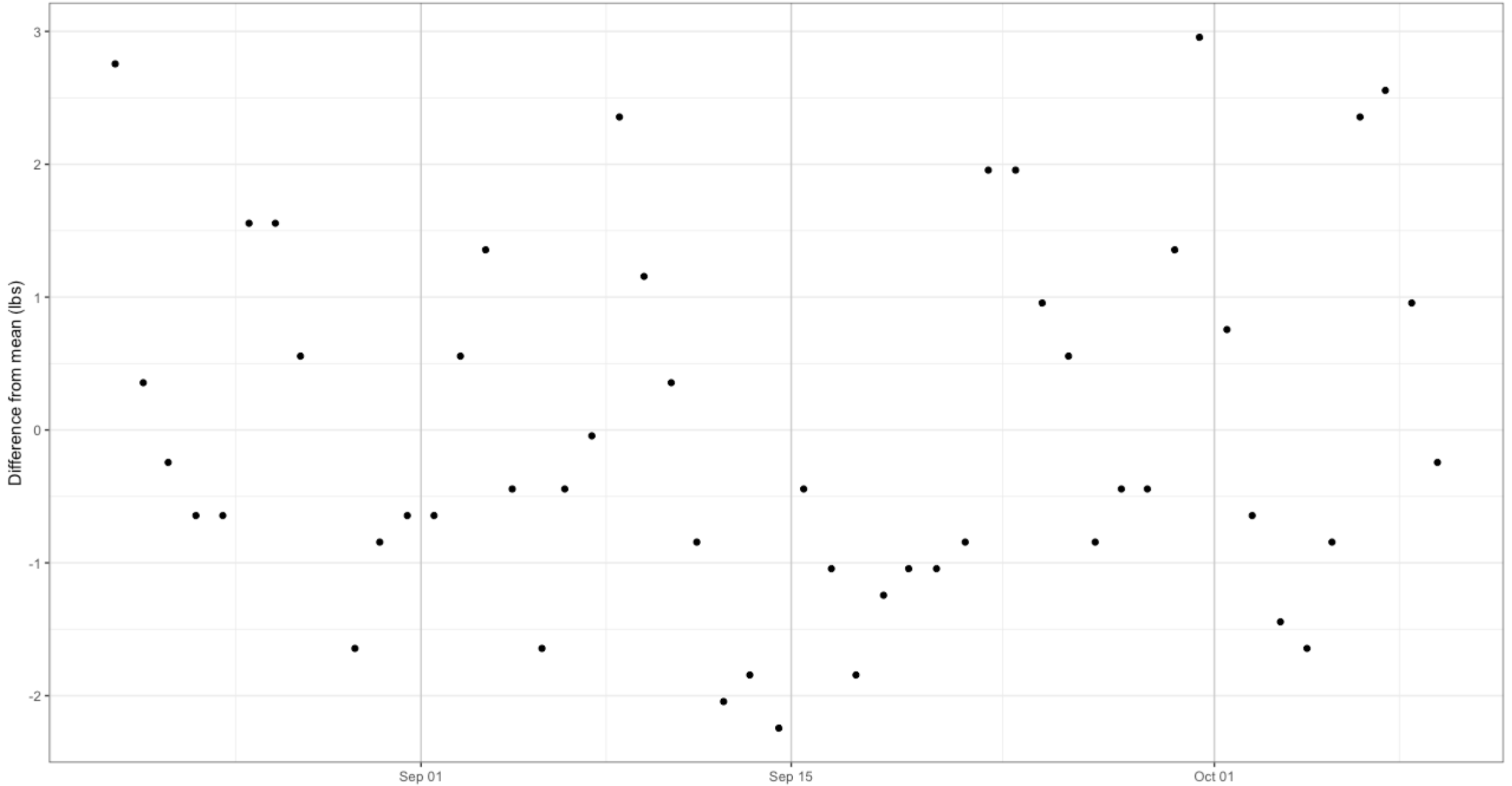
Top UGA Undergraduate Degrees

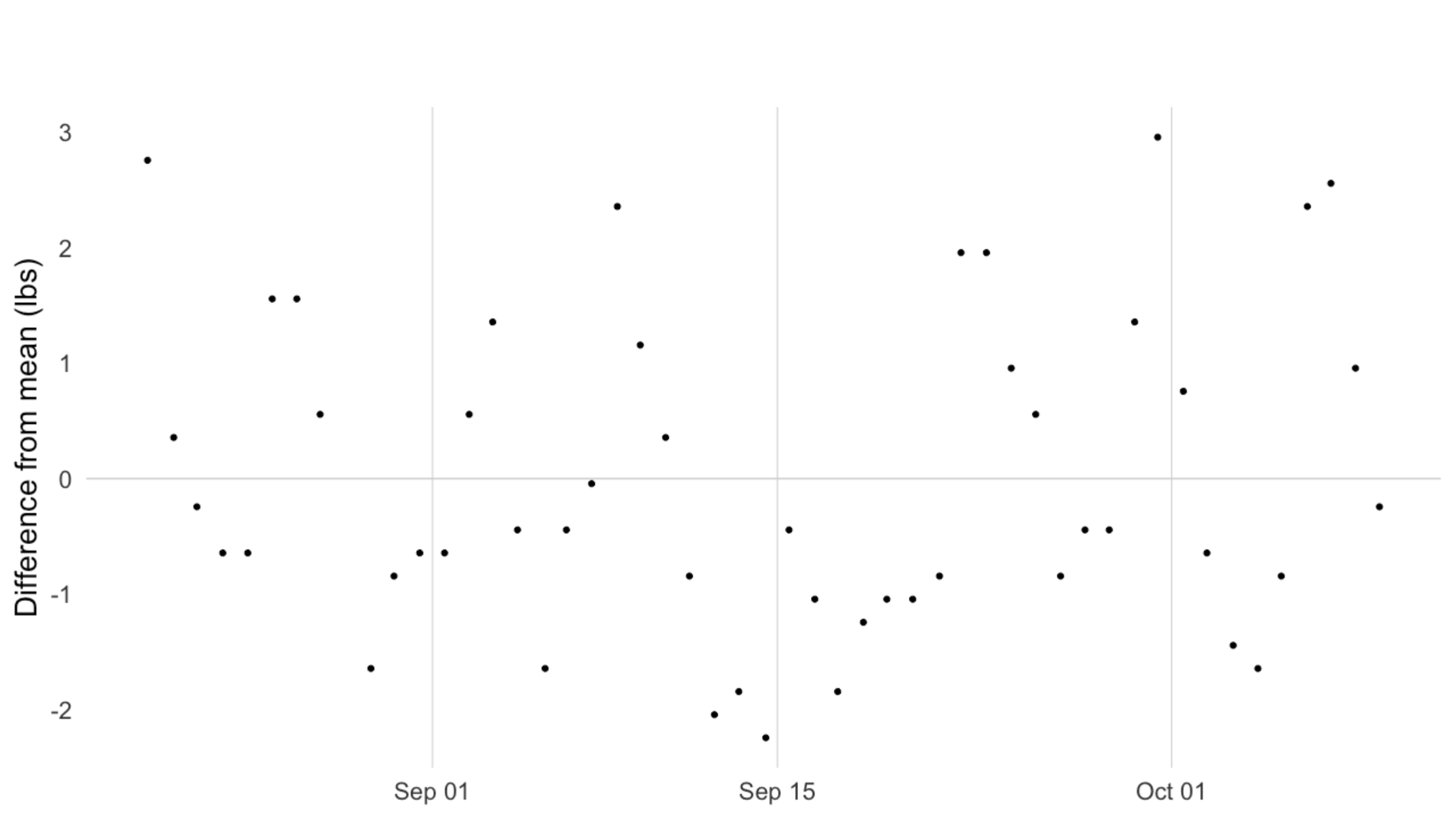


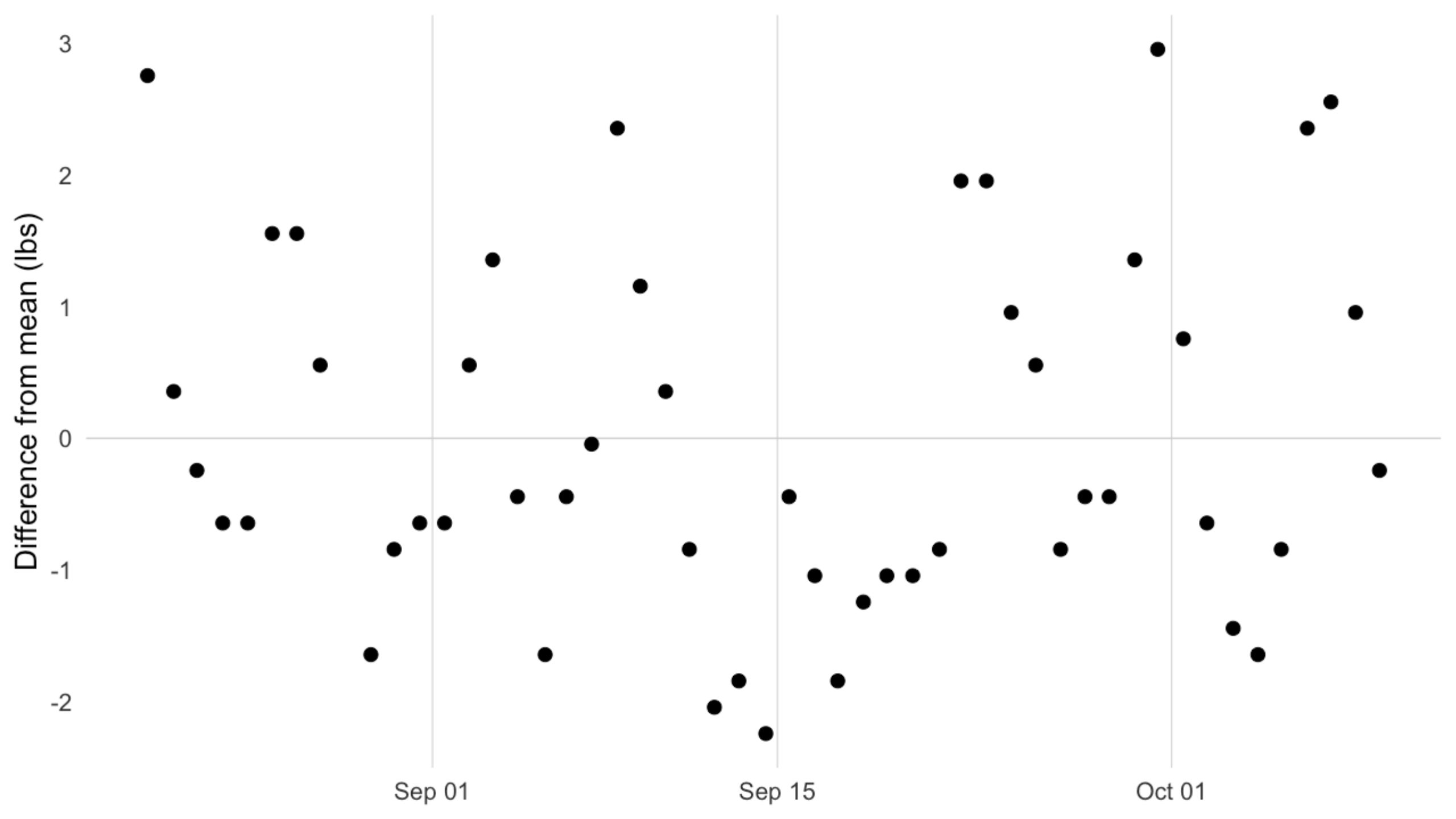
Remove unnecessary legends.

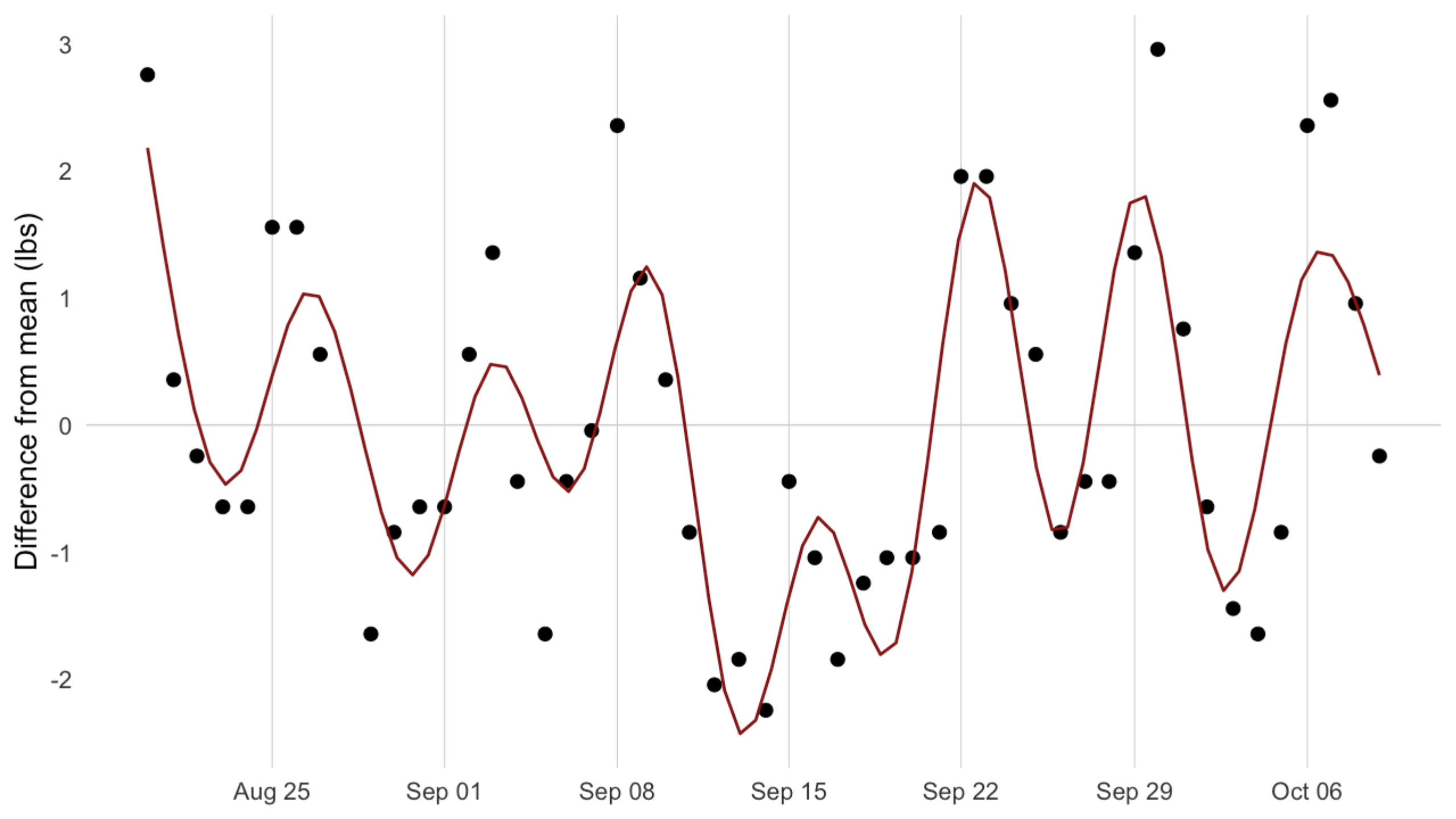
Use color judiciously.

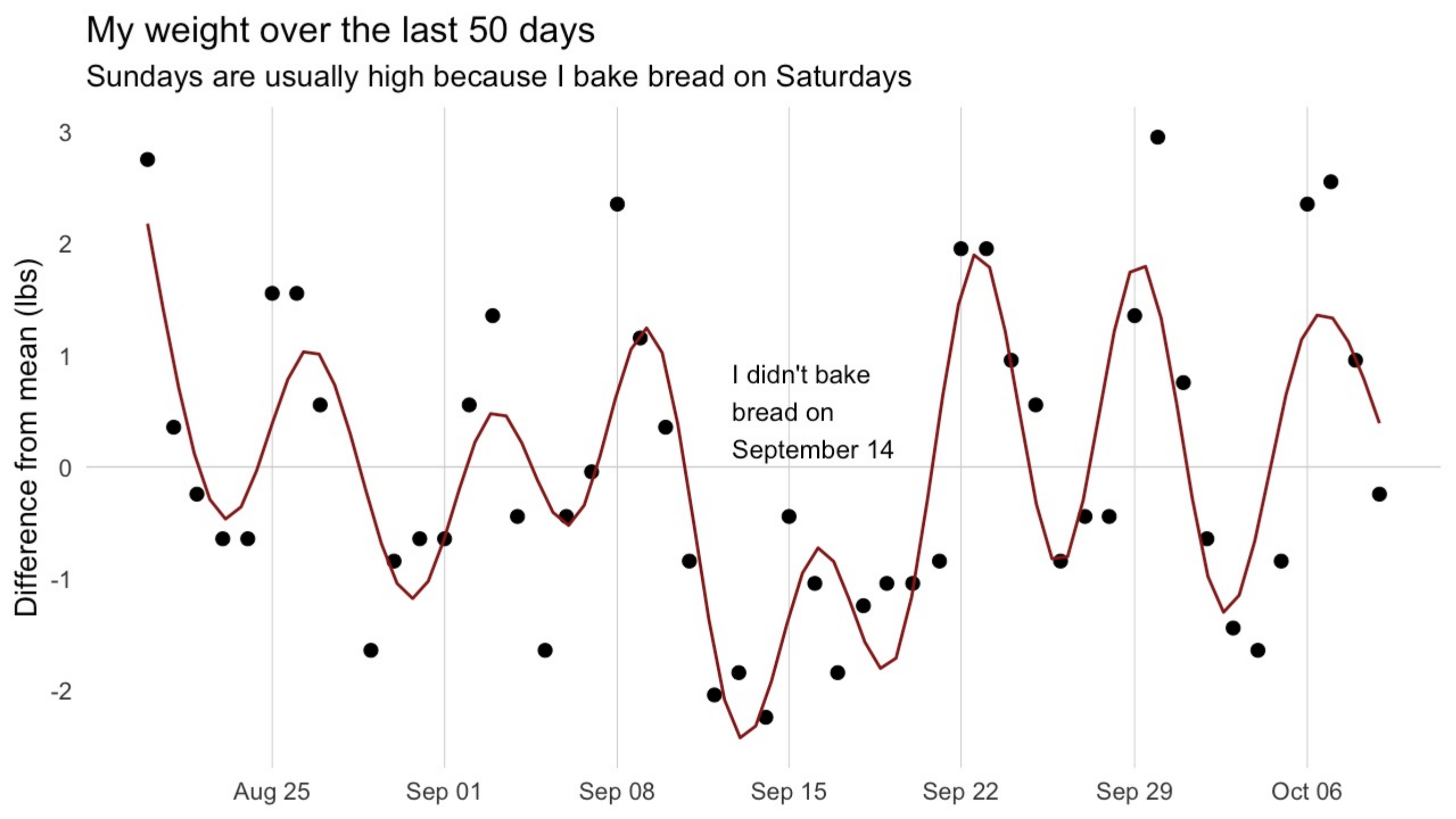
(See next week's workshop on color!)


Difference from mean (lbs)
3
2
1
0
-1
-2
Sep 01
Sep 15
Oct 01

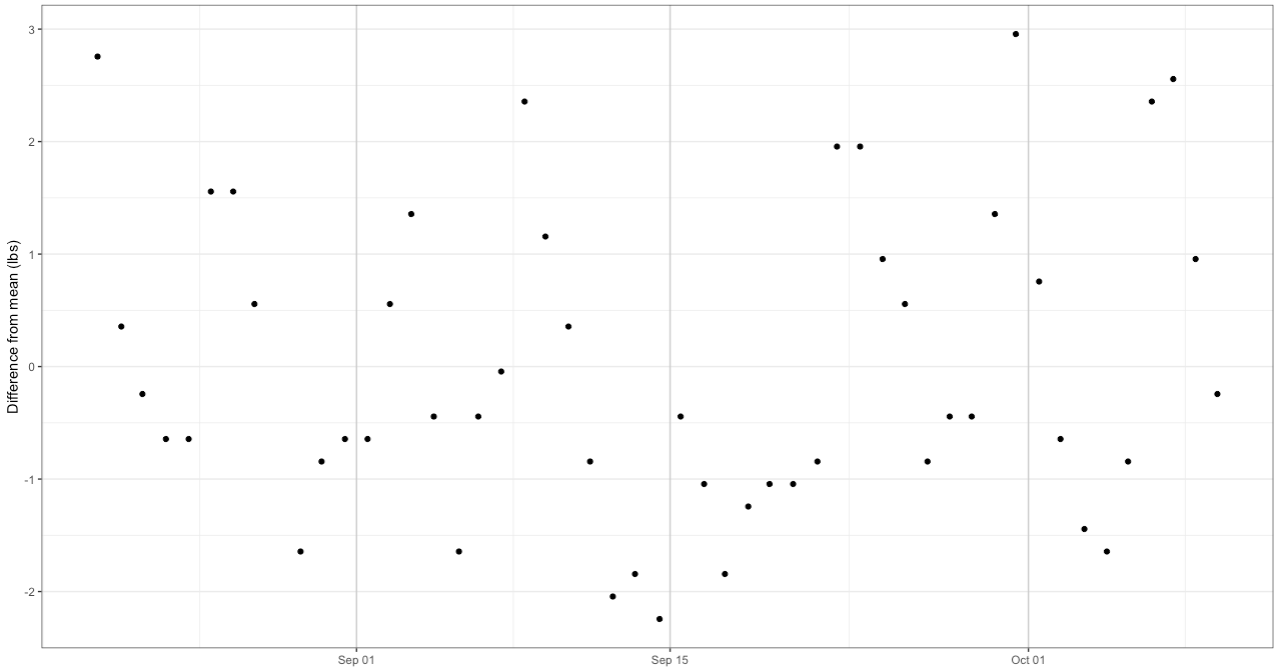

Difference from mean (lbs)
3
2
1
0
-1
-2
Sep 01
Sep 15
Oct 01

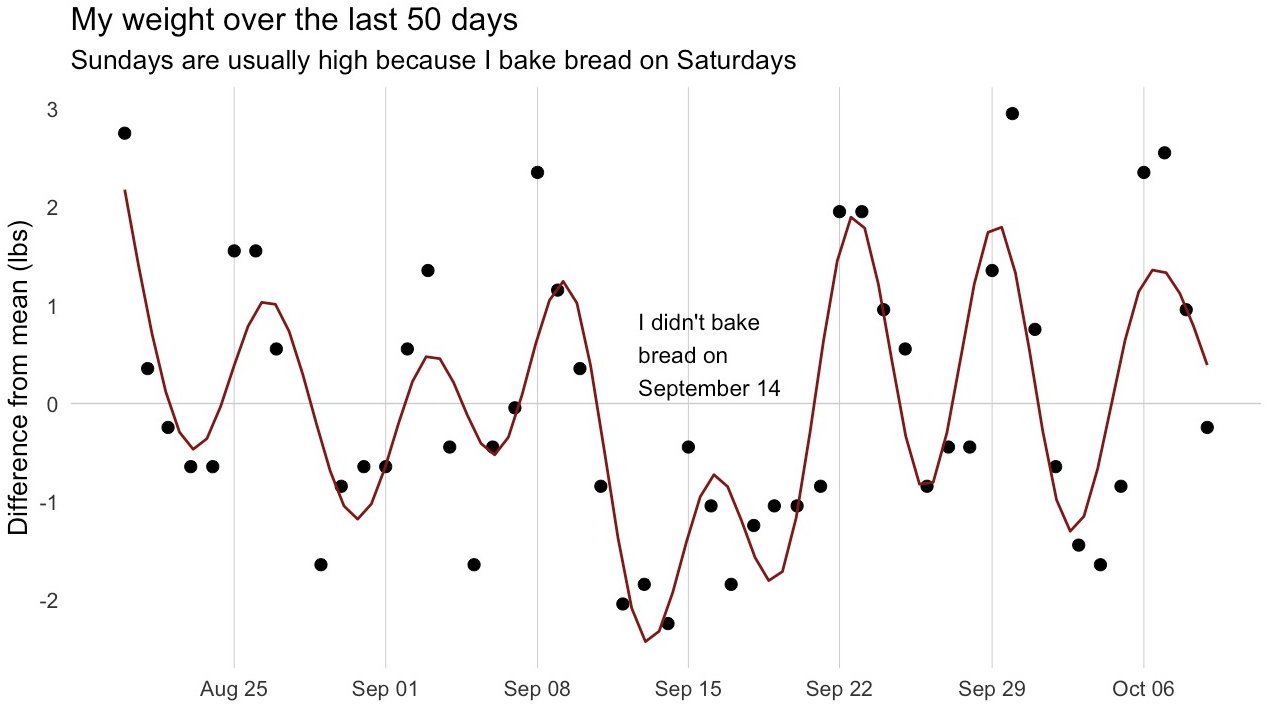

Difference from mean (lbs)
3
2
1
0
-1
-2
Sep 01
Sep 15
Oct 01


Difference from mean (lbs)
3
2
1
0
-1
-2
Aug 25
Sep 01
Sep 08
Sep 15
Sep 22
Sep 29
Oct 06


My weight over the last 50 days
Sundays are usually high because I bake bread on Saturdays
Difference from mean (lbs)
3
2
1
0
-1
-2
I didn't bake
bread on
September 14
Aug 25
Sep 01
Sep 08
Sep 15
Sep 22
Sep 29
Oct 06

Tell the Story





Remove

to improve

the **pie chart** edition

Created by Darkhorse Analytics

www.darkhorseanalytics.com

Remove

to improve

(the **data-ink** ratio)

Remove

to improve

the **data tables** edition

Remove

to improve

the **map** edition

Created by Darkhorse Analytics

Today's Presentation

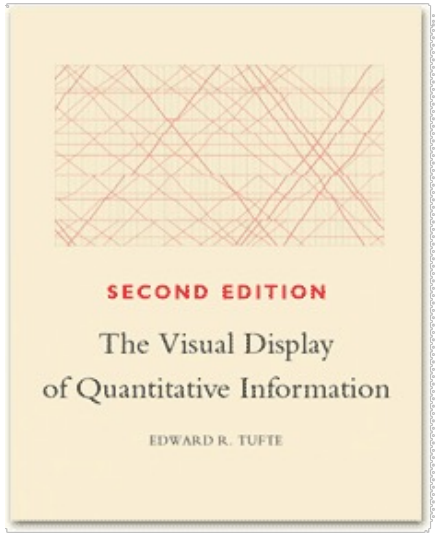
1. Purposes of data visualization
   - Why visualize data?
   - Who are data visualizations for?

2. Graphical integrity & fidelity
   - (How to not lie and how to spot liars.)

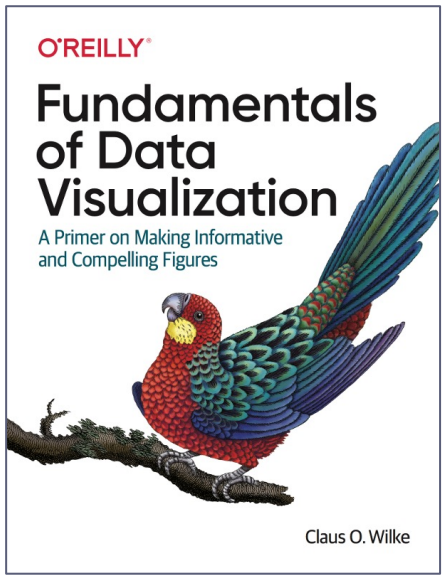
3. How to make plots look better
   - Remove chartjunk
   - Data-to-ink ratio and maximumization

CREDITS

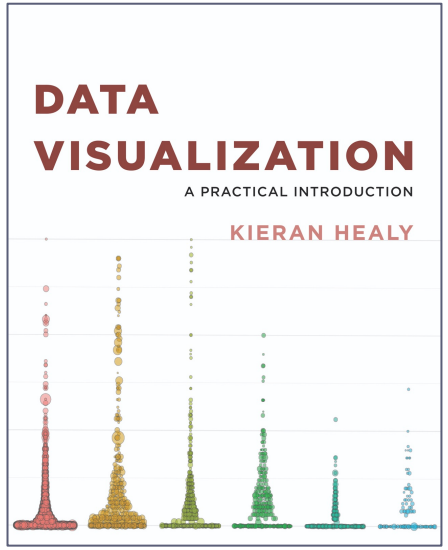
Edward R. Tufte (2001). *The Visual Display of Quantitative Information.* 2nd ed.



Claus O. Wilke (2019). *Fundamentals of Data Visualization.* serialmentor.com/dataviz/



Kearan Healy (2018). *Fundamentals of Data Visualization.*



Dark Horse Analytics for the Less is More Animations (darkhorseanalytics.com)