

Sharing Stata Knowledge Online: Existing Examples & How to Do It More Effectively

JOHN V. KANE

ASSOCIATE PROF, NEW YORK UNIVERSITY

JOHN.KANE@NYU.EDU

Introduction

- ▶ Personally, a strong appeal of Stata has always been: **ease of use + powerful capability**
- ▶ But finding resources online can be tricky, especially for data visualization
 - ▶ Important because data viz. is increasingly popular (vis-à-vis tabular output)
 - ▶ Users don't always know where to go to learn about new techniques & pick up new skills from Stata users
- ▶ And ways of sharing knowledge (w/ students, broader Stata community, etc.) are also limited
 - ▶ E.g., easily sharing ideas, best practices, and copyable code

Introduction



- ▶ **How can we both learn and share Stata knowledge more effectively?**
 - ▶ Will cover some common resources and their limitations
 - ▶ Will discuss my experience with **Medium.com** and the advantages it offers for learning & sharing Stata knowledge

What do we want?

A specific fix or general guidance?

- ▶ Sometimes we have very **specific** things we want to fix
 - ▶ For example, the syntax for a particular type of graph, or how to adjust specific features of our graphs
- ▶ Where can we get online help for specific fixes?
 - ▶ help files and online manuals
 - ▶ Search/ask Statalist
 - ▶ Emailing Stata
 - ▶ ChatGPT

What do we want?


A specific fix or general guidance?

- ▶ But other times we want more **general** guidance
 - ▶ For example...
 - ▶ best practices for graphing regression output
 - ▶ how to use Stata with LaTeX
 - ▶ how and why to use particular types of bar graphs
 - ▶ how to make publication-quality graphs
 - ▶ how to make maps with Stata
 - ▶ how to use Stata's "Graph Editor" more effectively
 - ▶ how to use Stata to make specialized graphs that are not native to Stata & don't (yet) have user-written packages
- ▶ The aforementioned resources won't be much help
 - ▶ E.g., you can't really ask "how do I use Graph Editor?" on Statalist


What do we want?

On the hunt for general guidance

- ▶ When we want general guidance, where should we go online?



A screenshot of a Reddit post from the subreddit r/stata. The post is titled "Looking for some good graphs" and is categorized as a "Question". The user "kingbreadmess" posted it 3 years ago. The text of the post asks for recommendations on websites or codes for creating high-quality graphs, noting that the pre-installed Stata ones are not satisfactory.

←  r/stata • 3 yr. ago
kingbreadmess

Looking for some good graphs

Question

Does anyone know any websites or codes for great looking graphs? The preinstalled stata ones aren't quite hitting the spot for me. I've found a couple on some forums but I'm just wondering where everyone gets theirs from.

What do we want?

On the hunt for general guidance

- ▶ When we want general guidance, where should we go online?
 - ▶ Stata's YouTube channel and online tutorials
 - ▶ Limitations:
 - ▶ More emphasis on statistics than visualizations
 - ▶ Usually a bit more basic visualizations
 - ▶ Don't typically feature user-written commands or "tricks"
 - ▶ YouTube
 - ▶ Limitations:
 - ▶ Sometimes outdated
 - ▶ Sometimes not great quality (not easy to see/hear clearly)
 - ▶ Don't include code

What do we want?

On the hunt for general guidance

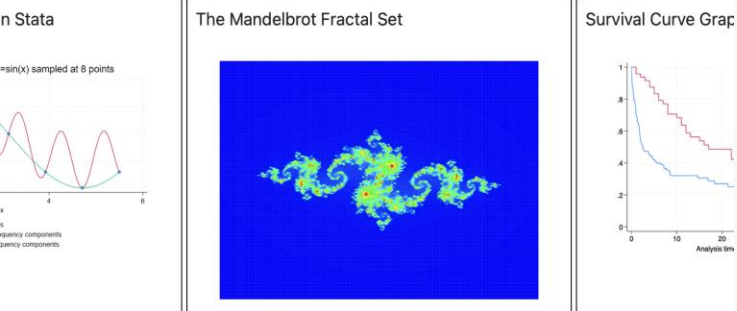
- ▶ When we want general guidance, where should we go online?
 - ▶ Various Stata groups' /users' own sites that are great

STATA[®]

SDAS Stata Graph Gallery

Learn to create these graphs with step-by-step guidance from SDAS TechTip

Show graphs by complexity



Looking for how to make the best graphs in Stata? Data visualizations and predictive margins plots? How to use coefplot in Stata? Let's do this.



Getting started – How to use coefplot

- Basic usage
- Keep and drop
- Plotting multiple models
 - Models as separate series
 - Changing the offsets
 - Using multiple axes
 - Appending models
 - Models as subgraphs
 - Multiple models per subgraph
 - Different plot styles per subgraph
 - How subgraphs are combined
 - Subgraphs by coefficients
 - Using wildcards in model names
- How coefficients are matched
- Ordering and sorting
 - Automatic order of coefficients
 - Explicit order of coefficients
 - Sort coefficients by size
- Plotting results from matrices
- Changing the plot type (recast)
- Using a continuous axis (at)
- Types and placement of options

Basic usage

The basic procedure is to compute one or more sets of estimates (e.g. regress) to draw a plot displaying the point estimates and their confidence intervals. The `coefplot` command returns (type `ereturn list` after running an estimation command) the point estimates from (the first equation in) vector `e(b)` estimates found in matrix `e(V)`. See the [Estimates](#) and [Confidence intervals](#) defaults. Furthermore, `coefplot` can also read results from matrices that are read from matrices below.

The syntax to produce a plot of the coefficients of a single model is

```
coefplot [name] [, options]
```

where `name` is the name of a stored model (see [help estimates store](#)), and `options` are used to plot the point estimates and 95% confidence intervals for the most recent model.

```
. sysuse auto, clear  
      (1978 automobile data)  
  
. regress price mpg trunk length turn
```


What do we want?

On the hunt for general guidance

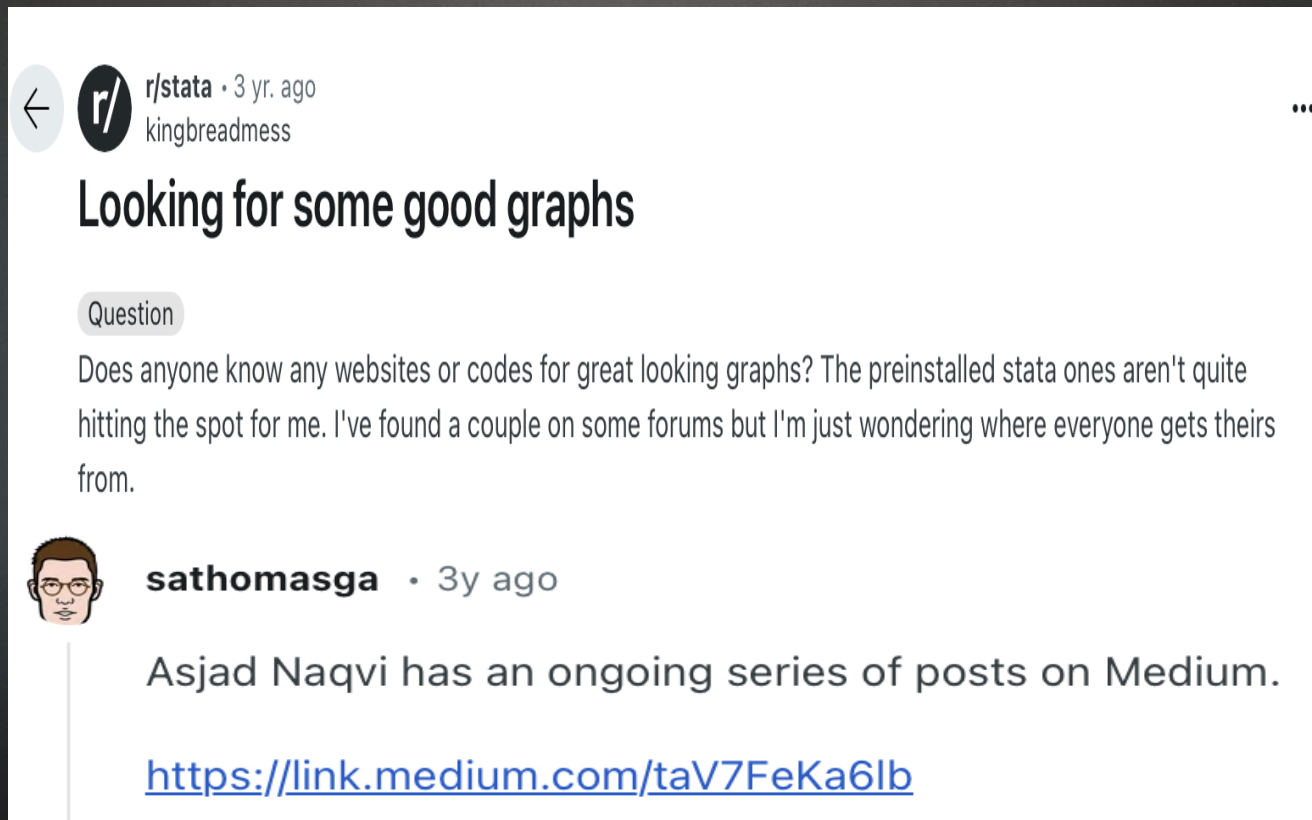
- ▶ When we want general guidance, where should we go online?
 - ▶ Various Stata users' own sites
 - ▶ Limitations:
 - ▶ Can't necessarily ask the authors questions
 - ▶ Inefficient searching / have to continually check for new content
 - ▶ Limited content/examples
 - ▶ Don't always have detailed explanations
 - ▶ As authors: We need our own sites to post our own content to share
 - ▶ Social media isn't great for providing code, reaching large audiences


What do we want?

On the hunt for general guidance

► When we want general guidance, where should we go online?

► **Medium.com**




◀  **r/stata** · 3 yr. ago
kingbreadmess

Looking for some good graphs

Question

Does anyone know any websites or codes for great looking graphs? The preinstalled stata ones aren't quite hitting the spot for me. I've found a couple on some forums but I'm just wondering where everyone gets theirs from.

 **sathomasga** · 3y ago

Asjad Naqvi has an ongoing series of posts on Medium.

<https://link.medium.com/taV7FeKa6lb>

Using Medium for General Stata Knowledge Sharing

- ▶ **Medium.com** has quickly become a valuable resource for Stata users
- ▶ Longer-form articles (5-25 mins)
 - ▶ for sharing detailed guidance
- ▶ Features “The Stata Guide” & “The Stata Gallery”



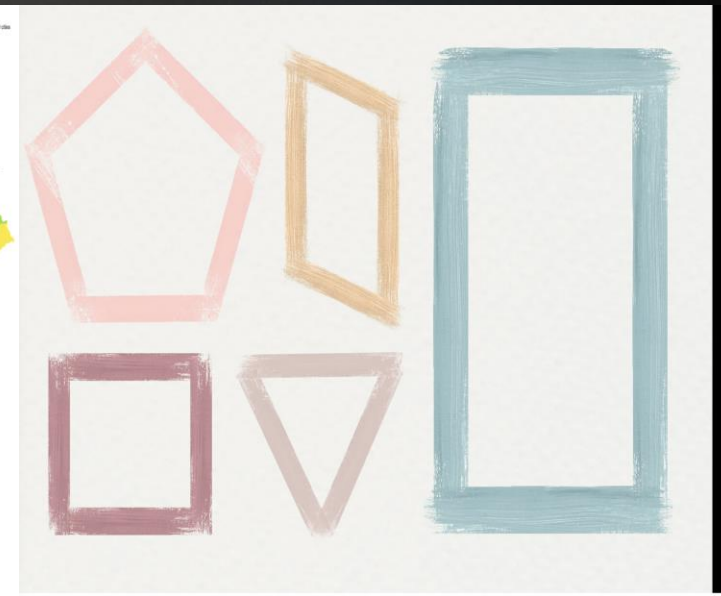
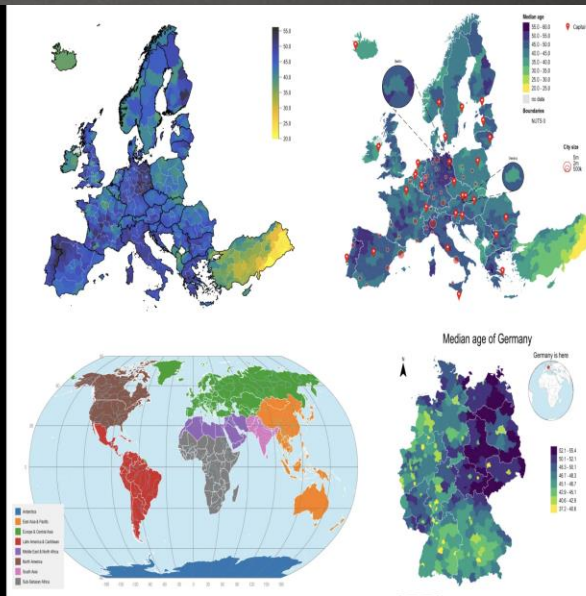
“The Stata Guide” on Medium

Amazing Resources from Asjad Naqvi



MAPS IN STATA PART III: "GEOPLOT"

JULY 2024 UPDATE



THE STATA FRAMES GUIDE

- Creator of 25 Stata packages (and is the one who nudged me to use Medium for my guides)

“The Stata Gallery”

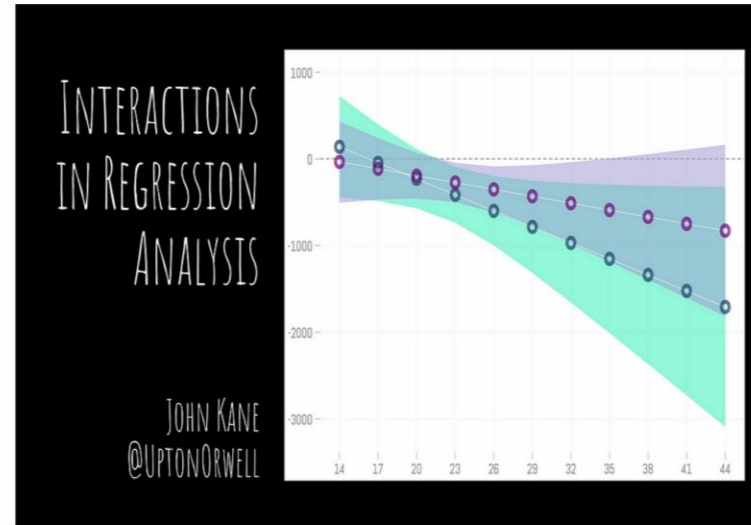
- ▶ Hosted by Asjad Naqvi, but allows submissions from anyone (guidelines available online)
- ▶ Hundreds of followers
- ▶ Diverse array of medium/long-form guides
 - ▶ Some on general statistical techniques
 - ▶ Some on user-written packages
 - ▶ Some on specific types of graphs or tasks
 - ▶ Some on general principles for better data visualization
 - ▶ Some on data cleaning & management
 - ▶ Some on various tricks and hacks

Medium



The Stata Gallery ABOUT

Writer



Interactions in Regression Models: What Are They & How Should We Visualize Them?

Want to use interactions in regression but find them challenging? This guide covers the key concepts & how to visualize them effectively!



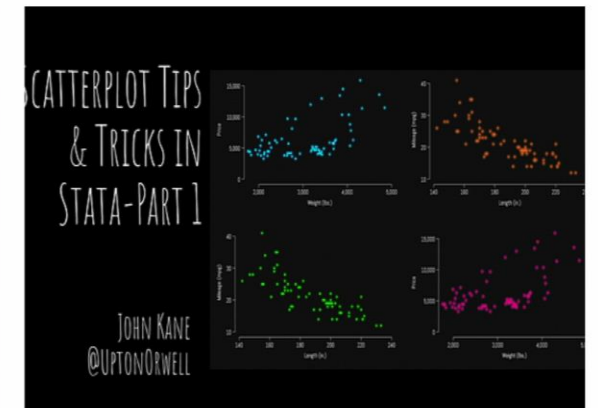
John V. Kane

Dec 14, 2024 · 25 min read

Latest



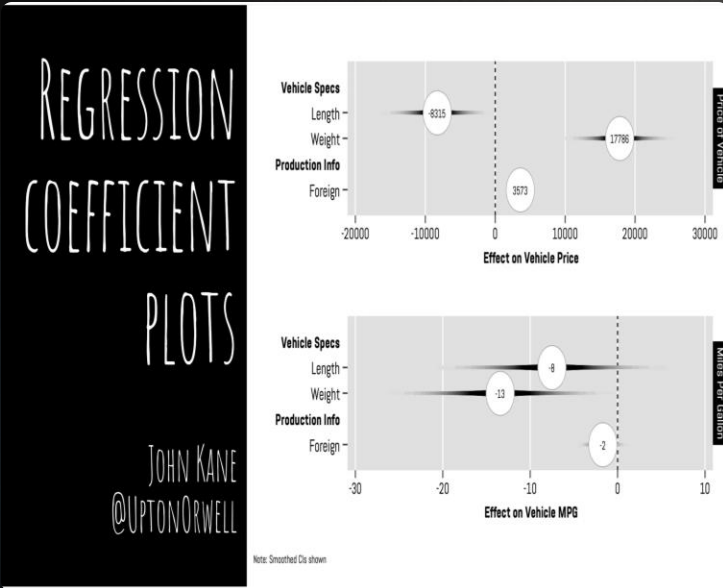
Stata to LaTeX: Converting SVG Graphs to TikZ



Scatterplot Tips & Tricks in Stata (#1)

Stata Gallery Guide Examples

Key Options for Regression Coefficient Plots



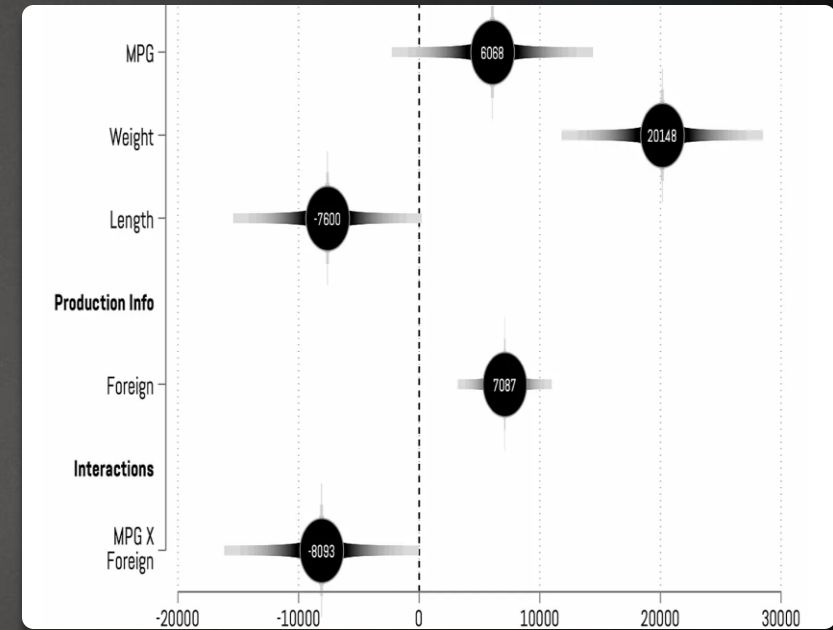
Example article title

And here is a second example, with code showing the models, individual coefplots, and process for combining the coefplots:

```
*Example #2: Style with marker labels
*First model:
reg price length_01 weight_01 i.foreign

coefplot, drop(_cons) xline(0, lcolor(black) lwidth(2)) scheme(gg_jet) ///
xtitle("{bf: Effect on Vehicle Price}") /// bolded text
graphregion(margin(2)) ///
xsize(6.5) ysize(4.5) ///
grid(none) ///
ylab(, labsize(*1.1)) /// enlarge size of y-axis labels
coeflabels(length_01="Length" weight_01="Weight" 1.foreign="Foreign") ///
headings(length_01= "{bf: Vehicle Specs}") ///
1.foreign = "{bf: Production Info}", gap(0)) ///
mlabel mlabcolor(black) mlabsize(vsmall) format(%4.0f) ///
mlabposition(center) msize(ehuge) mcolor(white) mcolor(gs10) ///
csmooth(color(black) lwidth(2 20)) /// default lwidth( ) range is 2 to 15
subtitle("Price of Vehicle", size(small) color(white) box fcolor(black) position
orientation(rvertical) margin(vsmall)) ///
saving(modi_labels.gph, replace)
```

Example section of code



Example figure from article

Stata Gallery Guide Examples:

Three-part Series on Bar Graphs



ADVANCED BAR GRAPHS IN STATA PART I

JOHN KANE @UPTONORWELL

Average Life Expectancy in the U.S. 1900 - 1999

Note: 83% & 95% CIs Shown

ADVANCED BAR GRAPHS IN STATA PART II

JOHN KANE @UPTONORWELL

Another Example of Using a 'Total' Category
Smoking Status & Doctor Visits in First Trimester

ADVANCED BAR GRAPHS IN STATA PART III

JOHN KANE @UPTONORWELL

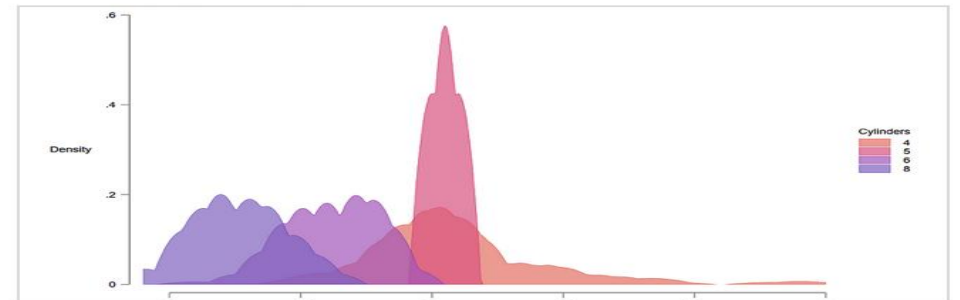
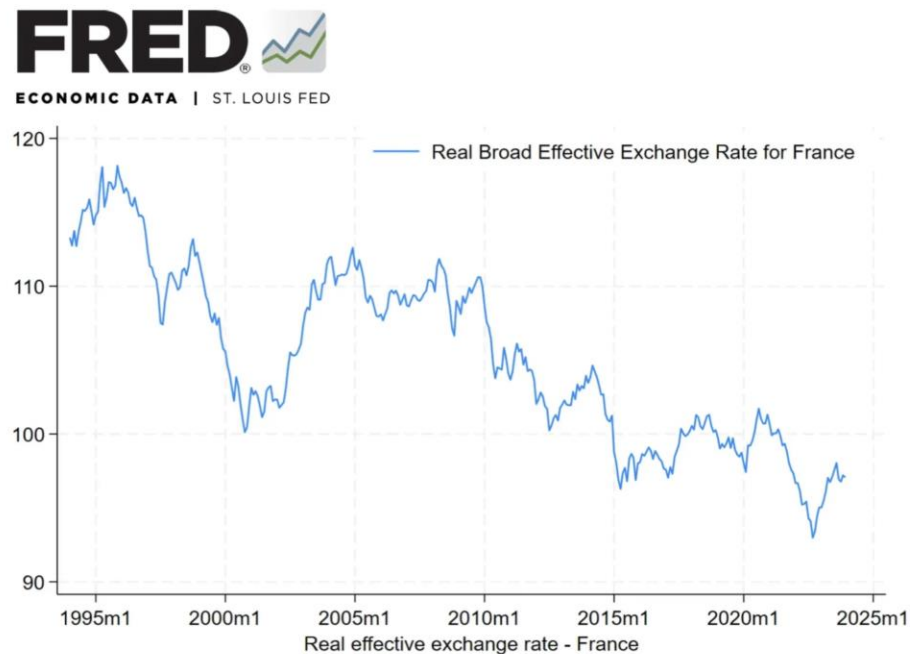
% Low Birthweight by Mothers' Age Group and Smoking Status
Example Using -grc1leg- Package

- Each article features links to the other two

Stata Gallery Guide Examples: Guides on FRED Data & Specialized Graphs

USING FRED
DATA WITH
STATA

JAMEL SAADAoui
WWW.JAMELSAADAoui.COM



Top 25 Stata Visualizations — With Full Code

Having trouble deciding what visualization to use for the type of problem at hand in Stata? This tutorial can help you decide!



Fahad Mirza

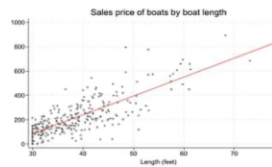
Nov 19, 2022 · 25 min read

Additional Stata Guide Examples: Guides on Statistical Concepts, Teaching & Techniques

Multicollinearity non-technically explained and visualized: regression modeling of the price of...

The multicollinearity problem is a recurrent topic in statistical research and data science, and textbooks abound with...

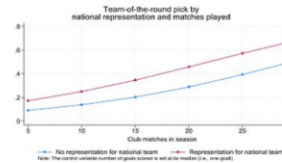
Apr 11, 2024 6



Tips for teaching MULTIPLE logistic regression more effectively ...

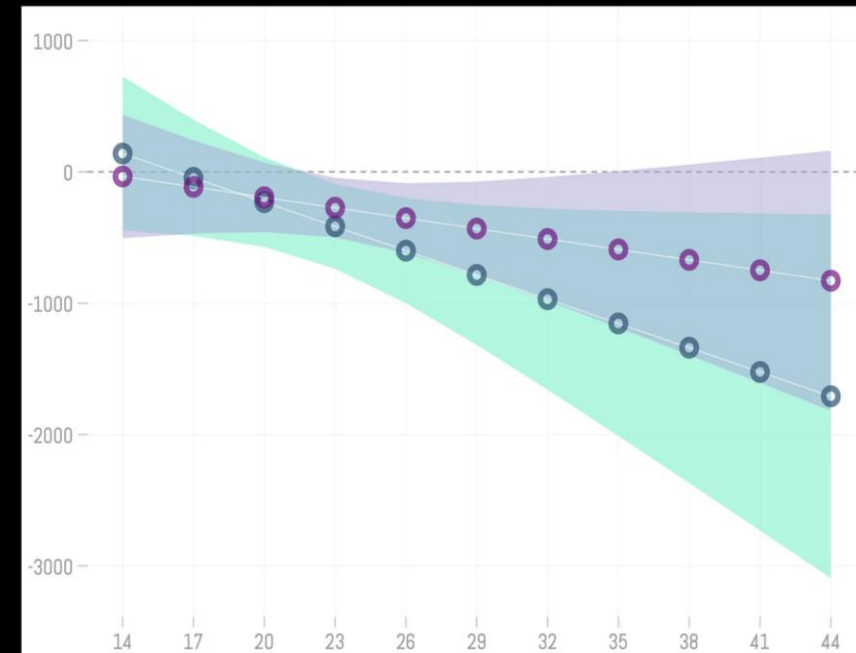
As promised given a certain amount of feedback, here is the sequel to my post "Tips for teaching logistic regression analysis..."

May 27, 2024 3



INTERACTIONS IN REGRESSION ANALYSIS

JOHN KANE
@UPTONORWELL

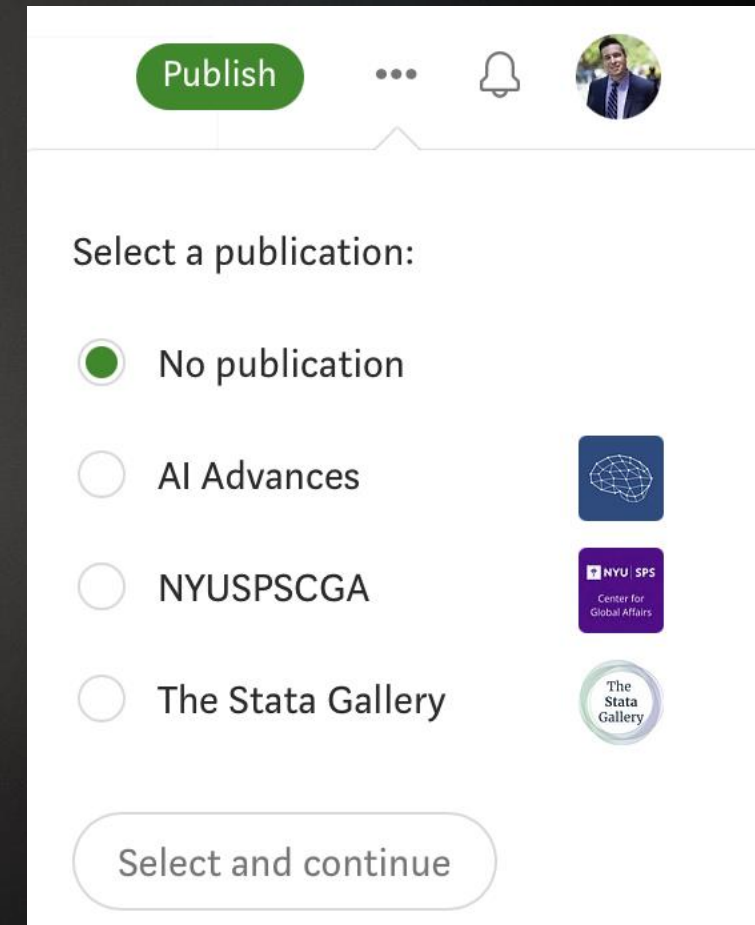


Advantages of Medium for Finding and Sharing Guides

- ▶ **Large audience**
 - ▶ Over 100 million active readers per month
 - ▶ Social Media: 2 million Twitter/X followers; 192k LinkedIn Followers
- ▶ **Easy to Search for and Share Articles**
 - ▶ Share article links with students, colleagues, co-authors*
 - ▶ Post article links on social media
- ▶ **Helpful Engagement Options**
 - ▶ You can follow authors and/or subscribe to receive emails when they write new articles
 - ▶ Ask authors questions via comment section
 - ▶ Can bookmark articles for future reading
 - ▶ Can highlight key sections of stories

Advantages of Medium for Authoring Guides

- ▶ Can publish stories on your own or via other groups
 - ▶ e.g., The Stata Gallery, Towards Data Science (TDS)
- ▶ No limit on length of articles
 - ▶ Users can see how long it takes to read
 - ▶ Option available to listen to any story (AI-generated voice)
- ▶ Gain followers / email subscribers
- ▶ *Articles can be edited / updated anytime*



Advantages of Medium for Authoring Guides

▶ Ease of use

- ▶ Typing and inserting images, hyperlinks, etc. is very simple
 - ▶ I often link to Statalist and other Medium guides
- ▶ Can easily add code blocks and paste in Stata code from a .do file
- ▶ Readers can copy code and paste directly into their own .do files
 - ▶ Useful for sharing installation code for user-written packages with a broader audience



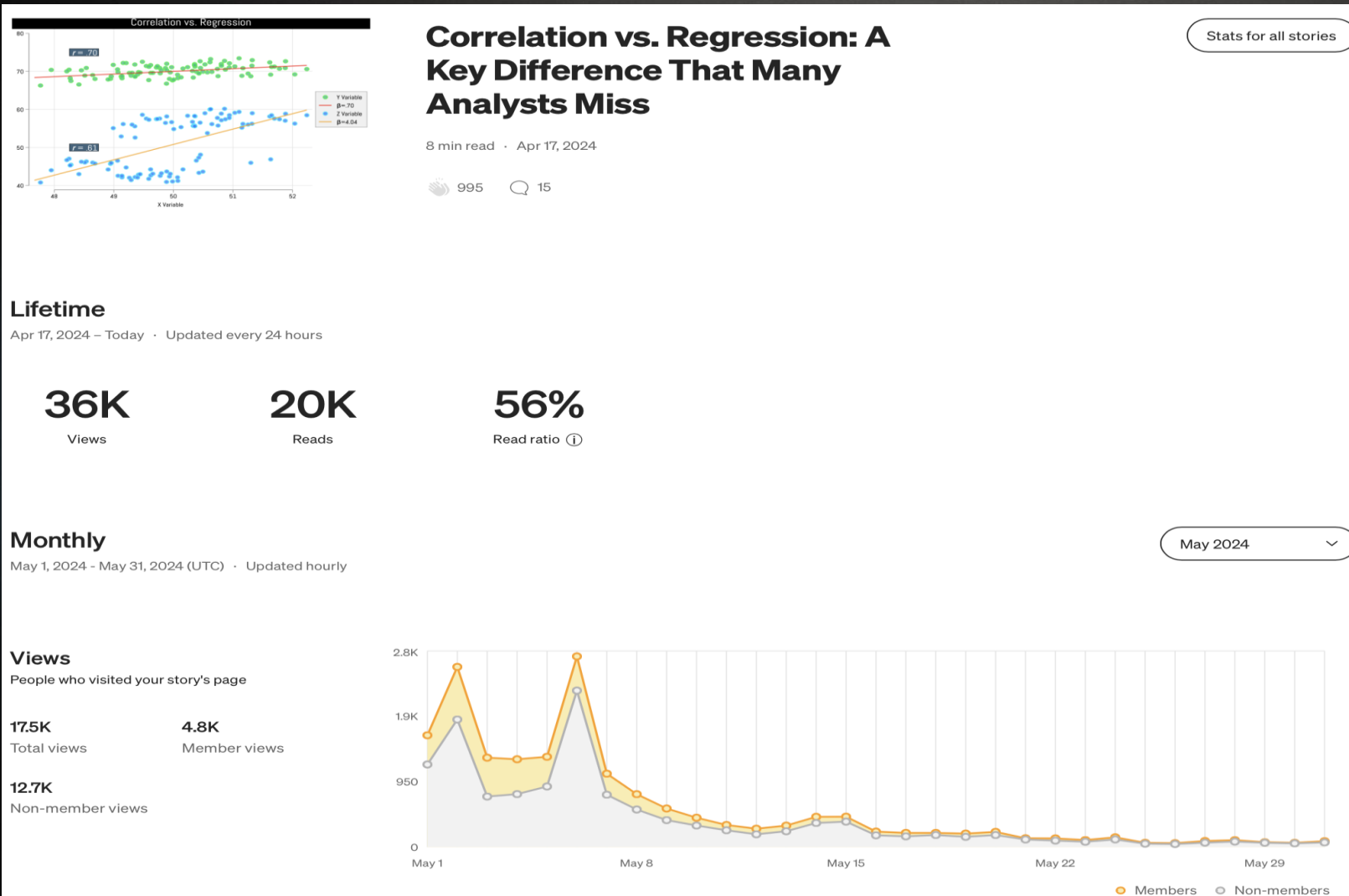
Other Advantages for Authors

- ▶ High visibility / search engine optimization (SEO)
 - ▶ Example: Googling Stata's Graph Editor
 - ▶ We see Stata's website, suggested YouTube videos, Statalist, then my Medium guide



Other Advantages for Authors

► Detailed stats for stories and info regarding how readers find your article



Views by traffic source

Internal ⓘ	52%
External referrals	48%
email, IM, and direct	14.8K
twitter.com	1.8K
mail.google.com	1.5K
google.com	397
towardsdatascience.com	369
Android device (not Medium app)	331
facebook.com	135
linkedin.com	92
bsky.app	49
news.ycombinator.com	38
All other external referrals	392

Costs and Benefits

- ▶ No paid membership required to write articles
 - ▶ Can access a specific number of articles per month
- ▶ Paid membership allows reading unlimited articles
 - ▶ \$5 per month (student discounts available)
- ▶ Partner program pays authors (but articles are limited access for free accounts)

Conclusions



- ▶ If you are looking for general guidance on how to get more out of Stata, and/or
- ▶ You have knowledge about Stata features, packages, tricks, tips, etc. that you would be willing to share...
- ▶ Consider using [Medium.com](https://medium.com)!
 - ▶ Easy to use, low cost, high visibility, growing Stata presence
 - ▶ I've really enjoyed my experience with it and hope to make more guides in the coming year!

Thank you!

▶ Questions? 🤔

▶ Feel free to connect!

▶ **Email:** john.kane@nyu.edu

▶ **Medium:**
www.medium.com/@jvk221

▶ **X/Twitter:** @UptonOrwell

▶ **Bluesky:**
@UptonOrwell.bsky.social

▶ **LinkedIn:** John V. Kane

Search

ults in County (Controlli

50
40
30

John V. Kane

700 Followers

John V. Kane is an Associate Professor at NYU's Center for Global Affairs. He researches political attitudes & experimental methods. Website: www.johnvkane.com

[Edit profile](#)

Following

- Arieda Muço
- The Medium Blog
- Cai Parry-Jones
- Shritha Sampath
- Mark Peffley

See all (21)

Help Status About Careers Press Blog Privacy Terms
Text to speech Teams