

# Training Computational Social Science PhD Students for Academic and Non-Academic Careers

Jae Yeon Kim

Johns Hopkins, ex-Code for America

July 14, 2024

# About me

- ▶ Political science PhD in UC Berkeley (2021).

# About me

- ▶ Political science PhD in UC Berkeley (2021).
- ▶ Previously, data scientist at the Safety Net Innovations Lab at Code for America.

# About me

- ▶ Political science PhD in UC Berkeley (2021).
- ▶ Previously, data scientist at the Safety Net Innovations Lab at Code for America.
- ▶ Currently, research scientist at the SNF Agora Institute at Johns Hopkins and fellow at Harvard Kennedy School's Center for Public Leadership.

# Plan

- 1 What's CSS portfolio?
- 2 How to build CSS portfolio

- ▶ CSS portfolio: a *purposeful* collection of CSS projects + outputs (Nolan and Stoudt, 2021; Robinson and Nolis, 2020).

- ▶ CSS portfolio: a *purposeful* collection of CSS projects + outputs (Nolan and Stoudt, 2021; Robinson and Nolis, 2020).
  - ▶ It's about communication and presentation of yourself as a computational social scientist.

- ▶ CSS portfolio: a *purposeful* collection of CSS projects + outputs (Nolan and Stoudt, 2021; Robinson and Nolis, 2020).
  - ▶ It's about communication and presentation of yourself as a computational social scientist.
  - ▶ It's a **CSS** portfolio: There must be a (social sciency) argument (and research design) + skill demonstration that convinces your deliverables are desirable and credible (Nolan and Stoudt, 2021).



- ▶ CSS portfolio: a *purposeful* collection of CSS projects + outputs (Nolan and Stoudt, 2021; Robinson and Nolis, 2020).
  - ▶ It's about communication and presentation of yourself as a computational social scientist.
  - ▶ It's a **CSS** portfolio: There must be a (social sciency) argument (and research design) + skill demonstration that convinces your deliverables are desirable and credible (Nolan and Stoudt, 2021).
  - ▶ As we shall see, these distinct components' proportions vary by audience and deliverable types.

- ▶ Portfolio is entrepreneurial: if you CREATE values, you must CAPTURE them.

- ▶ Portfolio is entrepreneurial: if you CREATE values, you must CAPTURE them.
  - ▶ In life and also in career: You Don't Get What You Deserve, You Get What You Negotiate.

- ▶ Portfolio is entrepreneurial: if you CREATE values, you must CAPTURE them.
  - ▶ In life and also in career: You Don't Get What You Deserve, You Get What You Negotiate.
  - ▶ If you want to get something, you must communicate.

- ▶ Portfolio is entrepreneurial: if you CREATE values, you must CAPTURE them.
  - ▶ In life and also in career: You Don't Get What You Deserve, You Get What You Negotiate.
  - ▶ If you want to get something, you must communicate.
  - ▶ From time to time, we need to reflect on, reframe, and reorient our research projects (3r).

- ▶ Portfolio is entrepreneurial: if you **CREATE** values, you must **CAPTURE** them.
  - ▶ In life and also in career: You Don't Get What You Deserve, You Get What You Negotiate.
  - ▶ If you want to get something, you must communicate.
  - ▶ From time to time, we need to **reflect** on, **reframe**, and **reorient** our research projects (3r).
- ▶ tl;dr: Academics spend so much time on **R&D** (knowledge production). In contrast, we spend too little time on **marketing** (packaging and selling ideas) ([Moynihan, 2022](#)).

- ▶ Portfolio is entrepreneurial: if you **CREATE** values, you must **CAPTURE** them.
  - ▶ In life and also in career: You Don't Get What You Deserve, You Get What You Negotiate.
  - ▶ If you want to get something, you must communicate.
  - ▶ From time to time, we need to **reflect** on, **reframe**, and **reorient** our research projects (3r).
- ▶ tl;dr: Academics spend so much time on **R&D** (knowledge production). In contrast, we spend too little time on **marketing** (packaging and selling ideas) ([Moynihan, 2022](#)).
  - ▶ Or we are not well trained to be good communicators and presenters (going beyond job talks).

- ▶ CSS portfolio is unique.



- ▶ CSS portfolio is unique.
- ▶ CSS portfolio is not limited to academic papers/conference proceedings because computational social scientists (not all but many of us are scholar-practitioners) have dual audiences.

- ▶ CSS portfolio is unique.
- ▶ CSS portfolio is not limited to academic papers/conference proceedings because computational social scientists (not all but many of us are scholar-practitioners) have dual audiences.
  - ▶ Paper (for academic, non-academic).

- ▶ CSS portfolio is unique.
- ▶ CSS portfolio is not limited to academic papers/conference proceedings because computational social scientists (not all but many of us are scholar-practitioners) have dual audiences.
  - ▶ Paper (for academic, non-academic).
  - ▶ Teaching materials (for academic, non-academic).

- ▶ CSS portfolio is unique.
- ▶ CSS portfolio is not limited to academic papers/conference proceedings because computational social scientists (not all but many of us are scholar-practitioners) have dual audiences.
  - ▶ Paper (for academic, non-academic).
  - ▶ Teaching materials (for academic, non-academic).
  - ▶ GitHub repo (for non-academic).

- ▶ CSS portfolio is unique.
- ▶ CSS portfolio is not limited to academic papers/conference proceedings because computational social scientists (not all but many of us are scholar-practitioners) have dual audiences.
  - ▶ Paper (for academic, non-academic).
  - ▶ Teaching materials (for academic, non-academic).
  - ▶ GitHub repo (for non-academic).
  - ▶ R package, Python library (for non-academic).

- ▶ CSS portfolio is unique.
- ▶ CSS portfolio is not limited to academic papers/conference proceedings because computational social scientists (not all but many of us are scholar-practitioners) have dual audiences.
  - ▶ Paper (for academic, non-academic).
  - ▶ Teaching materials (for academic, non-academic).
  - ▶ GitHub repo (for non-academic).
  - ▶ R package, Python library (for non-academic).
  - ▶ Interactive maps, dashboards (for non-academic).

- ▶ CSS portfolio is unique.
- ▶ CSS portfolio is not limited to academic papers/conference proceedings because computational social scientists (not all but many of us are scholar-practitioners) have dual audiences.
  - ▶ Paper (for academic, non-academic).
  - ▶ Teaching materials (for academic, non-academic).
  - ▶ GitHub repo (for non-academic).
  - ▶ R package, Python library (for non-academic).
  - ▶ Interactive maps, dashboards (for non-academic).
  - ▶ Blog posts (for non-academic).

# Plan

- 1 What's CSS portfolio?
- 2 How to build CSS portfolio



# Portfolio Development Tips

- ▶ "One way to think about building a successful portfolio is to imagine it as a series of 'deliverables' that demonstrate that one understands the CSS pipeline."

# Portfolio Development Tips

- ▶ "One way to think about building a successful portfolio is to imagine it as a series of 'deliverables' that demonstrate that one understands the CSS pipeline."
- ▶ Start early.

# Portfolio Development Tips

- ▶ "One way to think about building a successful portfolio is to imagine it as a series of 'deliverables' that demonstrate that one understands the CSS pipeline."
- ▶ Start early.
- ▶ Use GitHub (there's a student developer pack!) to manage your research processes as well as outputs and document them (using README, etc).

# Portfolio Development Tips

- ▶ "One way to think about building a successful portfolio is to imagine it as a series of 'deliverables' that demonstrate that one understands the CSS pipeline."
- ▶ Start early.
- ▶ Use GitHub (there's a student developer pack!) to manage your research processes as well as outputs and document them (using README, etc).
  - ▶ Bonus: it helps you prepare replication code and data for journal publications.

# Additional Steps

- ▶ Take an extra step.

# Additional Steps

- ▶ Take an extra step.
  - ▶ If you wrote code, develop a package.

# Additional Steps

- ▶ Take an extra step.
  - ▶ If you wrote code, develop a package.
  - ▶ If you taught a course, create a website.

# Additional Steps

- ▶ Take an extra step.
  - ▶ If you wrote code, develop a package.
  - ▶ If you taught a course, create a website.
  - ▶ If you published a paper, write a public-facing blog post.



# Additional Steps

- ▶ Take an extra step.
  - ▶ If you wrote code, develop a package.
  - ▶ If you taught a course, create a website.
  - ▶ If you published a paper, write a public-facing blog post.
- ▶ Document feedback: start writing a **brag document**

# Additional Steps

- ▶ Take an extra step.
  - ▶ If you wrote code, develop a package.
  - ▶ If you taught a course, create a website.
  - ▶ If you published a paper, write a public-facing blog post.
- ▶ Document feedback: start writing a **brag document**
  - ▶ Publication credits are easy to capture; others are not.

# Kudos Example 1



**Drew Engelhardt** @amengel.bsky.social · 15h

Huge thanks to [@jaeyeonkim.bsky.social](https://jaeyeonkim.bsky.social) for his {tidytweetjson} package and helpful tutorial. Went from crashing my RStudio attempting to read in a JSON file with 1.1 million tweets running overnight to loading and reformatting everything in just 10 minutes.

[jaeyk.github.io/tidytweetjson/](https://jaeyk.github.io/tidytweetjson/)

jaeyk.github.io

## Tidying Tweet JSON files

Twitter data is an important resource for social science research. However, parsing a great deal of Twitter JSON data is not an easy task for researchers with little programming experience. This packa...



1



1



4



**Jae Yeon Kim** @jaeyeonkim.bsky.social · 1h

I'm glad to know that the code still works after 4 years!



1



**Drew Engelhardt**

@amengel.bsky.social

Works great! The package and the tip to use gsplint have made dealing with >200GB of tweets much easier. Appreciate it!

# Kudos Example 2



Figure 2: Kudos example 2

# Takeaways

1. If you create values, you also need to capture them.
2. You capture values by purposefully creating a CSS portfolio (!= just papers).
3. You create a portfolio by taking one extra step.
4. Practice makes you more efficient. So, start early!
5. Finally, balance between efforts and opportunities.

# Thank You

Comments or questions?  
E-mail: [jkim638@jhu.edu](mailto:jkim638@jhu.edu)

Moynihhan, Donald P. 2022. “How Can Scholars Help to Embed Institutions of Public-Sector Change?(Or Things I Wish I’d Known When I Was a Grad Student).” *Perspectives on Public Management and Governance* 5(4):276–287.

Nolan, Deborah and Sara Stoudt. 2021. “The promise of portfolios: Training modern data scientists.” *Harvard Data Science Review* 3(3).

Robinson, Emily and Jacqueline Nolis. 2020. *Build a career in data science*. Manning Publications.