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

Jae Yeon Kim[1]

Code for America, Johns Hopkins

April 19, 2024

[1] Joint work with Aniket Kesari (Fordham), Sono Shah (Pew), Taylor Brown (Meta), Tiago Ventura (Georgetown), and Tina Law (UC Davis)

About me

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

About me

- **Political science** PhD in UC Berkeley (2021).
- I am a **data scientist** at Code for America, where we work with the U.S. federal, state, and local governments to make safety net programs (e.g., Medicaid, SNAP, WIC, etc) more accessible.

- My role (discipline) is **quantitative research**.

- My role (discipline) is **quantitative research**.
- I design and implement **field experiments (RCTs in the field)** and surveys with these agencies and evaluation partners (e.g., the federal Office of Evaluation Sciences and Georgetown's Better Government Lab).

- My role (discipline) is **quantitative research**.
- I design and implement **field experiments (RCTs in the field)** and surveys with these agencies and evaluation partners (e.g., the federal Office of Evaluation Sciences and Georgetown's Better Government Lab).
- **I didn't leave academia!**: I still do academic research as a fellow at the SNF Agora Institute at Johns Hopkins and the Center for Public Leadership at Harvard Kennedy School.

- My role (discipline) is **quantitative research**.
- I design and implement **field experiments (RCTs in the field)** and surveys with these agencies and evaluation partners (e.g., the federal Office of Evaluation Sciences and Georgetown's Better Government Lab).
- **I didn't leave academia!**: I still do academic research as a fellow at the SNF Agora Institute at Johns Hopkins and the Center for Public Leadership at Harvard Kennedy School.
- I prefer the "**building another bridge $>$ leaving academia**" frame.

- Data science skills and my academic training (**domain knowledge** + **research design**) as a social scientist were crucial in my professional journey.

- Data science skills and my academic training (**domain knowledge** + **research design**) as a social scientist were crucial in my professional journey.
- A typical workflow (in my opinion) of solving problems as a quant researcher / data scientist:

- Data science skills and my academic training (**domain knowledge** + **research design**) as a social scientist were crucial in my professional journey.
- A typical workflow (in my opinion) of solving problems as a quant researcher / data scientist:
    1. mapping problems (questions) to research design

- Data science skills and my academic training (**domain knowledge** + **research design**) as a social scientist were crucial in my professional journey.
- A typical workflow (in my opinion) of solving problems as a quant researcher / data scientist:
  1. mapping problems (questions) to research design
  2. mapping research design to data analysis

- Data science skills and my academic training (**domain knowledge** + **research design**) as a social scientist were crucial in my professional journey.
- A typical workflow (in my opinion) of solving problems as a quant researcher / data scientist:
  1. mapping problems (questions) to research design
  2. mapping research design to data analysis
  3. mapping data analysis to deliverables

Why did we write this paper?

- Aniket and I discussed how we trained ourselves to be a computational social scientist at Berkeley.

Why did we write this paper?

- Aniket and I discussed how we trained ourselves to be a computational social scientist at Berkeley.
- We contacted colleagues (smarter than us / at least me) to combine diverse experiences with the goal of describing and exposing the hidden script for professionalization in the field of computational social science (social sciences + data science).

Why did we write this paper?

- Our core belief is that CSS (computational social science) is exciting and empowering, but its career pathways remain hidden.

Why did we write this paper?

- Our core belief is that CSS (computational social science) is exciting and empowering, but its career pathways remain hidden.
  - CSS provides many *exciting* research opportunities for almost any empirical problem.

Why did we write this paper?

- Our core belief is that CSS (computational social science) is exciting and empowering, but its career pathways remain hidden.
  - CSS provides many *exciting* research opportunities for almost any empirical problem.
  - We would like to *empower* students to define their career path(s) and success metrics in their own terms.

Why did we write this paper?

- Hidden opportunities (I wish I could have known them earlier)!

Why did we write this paper?

- Hidden opportunities (I wish I could have known them earlier)!
  - Many organizations are hiring computational social scientists (sometimes not using data science titles): academic departments, professional schools, nonprofits, tech companies, international organizations, and government agencies.

Why did we write this paper?

- Hidden opportunities (I wish I could have known them earlier)!
  - Many organizations are hiring computational social scientists (sometimes not using data science titles): academic departments, professional schools, nonprofits, tech companies, international organizations, and government agencies.
  - Even if you ultimately aim to take an academic position, a summer internship at an applied research organization is not a bad idea as it provides you with perspectives, skills, and networks.

Outline

1. Three-step framework
   - Learning data science skills as a social scientist
   - Building CSS portfolio
   - Networking in CSS
2. Conclusion and Discussions

Plan

1. Three-step framework
   - Learning data science skills as a social scientist
   - Building CSS portfolio
   - Networking in CSS
2. Conclusion and Discussions

1. Learning data science skills (step 1)

1. Learning data science skills (step 1)
2. Building a data science portfolio (step 2)

1. Learning data science skills (step 1)
2. Building a data science portfolio (step 2)
3. Connecting with computational social scientists (step 3)

Learning Data Science Skills	
Core Competencies	• Ability to design and execute research projects from end to end (data to report) • Domain expertise • Programming fluency in `R` and/or `Python` • Experience with data management, particularly with managing large, messy, and unstructured data • Effective communication and collaborative research skills with both technical and nontechnical colleagues (e.g., version control and documentation) • Practiced knowledge of machine learning and traditional quantitative social science paradigms • Engagement with ethical concerns about digital and digitized data and computational methods (e.g., privacy protection and algorithmic bias)
Additional Market-Specific Skills	• Ability to apply theory, methods, and findings to the practical aims of a product and/or organization (*non-academic*) • Proficiency with relational database languages (e.g., SQL) and cloud-based databases *(non-academic especially)*
Building a CSS Portfolio	
Core Competencies	• Publicly available research projects documented from end to end demonstrating engagement with social science and applied aspects of a research project via problem definition, hypothesis generation, data and outcome selection, and measurement and method application • Reproducible, efficient, and communicable code via GitHub • Publish and serve as reviewer for journal publications/conference proceedings
Additional Market-Specific Skills	• Sharing learnings through research notes *(non-academic)* and tutorials *(academic)*
Connecting with Computational Social Scientists	
Core Competencies	• Attend and know how to navigate cross-disciplinary computational social science conferences
Additional Market-Specific Skills	• Work with computational social scientists through internships and work with civic, social, and nonprofit organizations *(non-academic)* • Connect with computational social scientists working on similar topics in different sectors via online platforms (e.g., LinkedIn and Slack) *(non-academic)*

Figure 1: Computational Social Science professionalization process

- Core competencies: research design + domain knowledge (CSS is still social science)

- Core competencies: research design + domain knowledge (CSS is still social science)
- "We argue that effective CSS training begins—first and foremost—with strong training in two areas that social science PhD programs already focus on: research design and domain expertise."

- Research design (e.g., statistical and causal inference, experiment and survey design):

- Research design (e.g., statistical and causal inference, experiment and survey design):
  - Research design (writing for the clarity of thought) is your superpower as a computational ***social*** scientist

- Research design (e.g., statistical and causal inference, experiment and survey design):
  - Research design (writing for the clarity of thought) is your superpower as a computational ***social*** scientist
  - For instance, you can inform your team whether current efforts will gain insights based on ***design*** alone (without data)

- Domain knowledge (e.g., policy knowledge, behavioral science, opinion research, game theory)

- Domain knowledge (e.g., policy knowledge, behavioral science, opinion research, game theory)
  - You can tell what your team should aim to learn (=learning goals).

- Domain knowledge (e.g., policy knowledge, behavioral science, opinion research, game theory)
  - You can tell what your team should aim to learn (=learning goals).
  - You can help your team to interpret and communicate findings and support decision-making (=insights).

- Domain knowledge (e.g., policy knowledge, behavioral science, opinion research, game theory)
  - You can tell what your team should aim to learn (=learning goals).
  - You can help your team to interpret and communicate findings and support decision-making (=insights).
  - In social science research, improving efficiency is rarely a satisfying goal (our goal is not to develop a faster algorithm).

- Domain knowledge (e.g., policy knowledge, behavioral science, opinion research, game theory)
  - You can tell what your team should aim to learn (=learning goals).
  - You can help your team to interpret and communicate findings and support decision-making (=insights).
  - In social science research, improving efficiency is rarely a satisfying goal (our goal is not to develop a faster algorithm).
  - Computation is crucial but only one part of a large empirical research process.

- Programming fluency: At least Python/R fluency (pick one and try to be a multi-lingual). SQL is a plus (database query is usually a first step in the workflow in applied settings).

- Programming fluency: At least Python/R fluency (pick one and try to be a multi-lingual). SQL is a plus (database query is usually a first step in the workflow in applied settings).
- Data management (e.g., workflow, documentation, and version control)

- Programming fluency: At least Python/R fluency (pick one and try to be a multi-lingual). SQL is a plus (database query is usually a first step in the workflow in applied settings).
- Data management (e.g., workflow, documentation, and version control)
- Collaborative research skills (make other peoples' lives easier; you create impacts by being helpful)

- Programming fluency: At least Python/R fluency (pick one and try to be a multi-lingual). SQL is a plus (database query is usually a first step in the workflow in applied settings).
- Data management (e.g., workflow, documentation, and version control)
- Collaborative research skills (make other peoples' lives easier; you create impacts by being helpful)
- Machine learning paradigms (use cases: unstructured data, automation, etc)

- Programming fluency: At least Python/R fluency (pick one and try to be a multi-lingual). SQL is a plus (database query is usually a first step in the workflow in applied settings).
- Data management (e.g., workflow, documentation, and version control)
- Collaborative research skills (make other peoples' lives easier; you create impacts by being helpful)
- Machine learning paradigms (use cases: unstructured data, automation, etc)
- Research ethics (e.g., differential privacy, synthetic data)

- If your home dept doesn't teach these skills ...

- If your home dept doesn't teach these skills ...
- Look for other department courses: CS, Stat, I-School, etc.

- If your home dept doesn't teach these skills ...
- Look for other department courses: CS, Stat, I-School, etc.
- External learning opportunities:

- If your home dept doesn't teach these skills ...
- Look for other department courses: CS, Stat, I-School, etc.
- External learning opportunities:
  - Summer Institute in Computational Social Science (SICSS) (full disclosure: I'm an alum, former organizer, and current advisor)

- If your home dept doesn't teach these skills ...
- Look for other department courses: CS, Stat, I-School, etc.
- External learning opportunities:
  - Summer Institute in Computational Social Science (SICSS) (full disclosure: I'm an alum, former organizer, and current advisor)
  - Data Science for Social Good (DSSG)

- If your home dept doesn't teach these skills ...
- Look for other department courses: CS, Stat, I-School, etc.
- External learning opportunities:
  - Summer Institute in Computational Social Science (SICSS) (full disclosure: I'm an alum, former organizer, and current advisor)
  - Data Science for Social Good (DSSG)
  - The Inter-University Consortium for Political and Social Research (ICPSR)

- If your home dept doesn't teach these skills ...
- Look for other department courses: CS, Stat, I-School, etc.
- External learning opportunities:
  - Summer Institute in Computational Social Science (SICSS) (full disclosure: I'm an alum, former organizer, and current advisor)
  - Data Science for Social Good (DSSG)
  - The Inter-University Consortium for Political and Social Research (ICPSR)
- Online tutorials and resources: Data Carpentry and R-Ladies

- Portfolio: projects + outputs

- Portfolio: projects + outputs
- CSS portfolio is not limited to academic papers / conference proceedings

- Portfolio: projects + outputs
- CSS portfolio is not limited to academic papers / conference proceedings
  - Paper (for academic, non-academic)

- Portfolio: projects + outputs
- CSS portfolio is not limited to academic papers / conference proceedings
  - Paper (for academic, non-academic)
  - Teaching materials (for academic, non-academic)

- Portfolio: projects + outputs
- CSS portfolio is not limited to academic papers / conference proceedings
  - Paper (for academic, non-academic)
  - Teaching materials (for academic, non-academic)
  - GitHub repo (for non-academic)

- Portfolio: projects + outputs
- CSS portfolio is not limited to academic papers / conference proceedings
  - Paper (for academic, non-academic)
  - Teaching materials (for academic, non-academic)
  - GitHub repo (for non-academic)
  - R package, Python library (for non-academic)

- Portfolio: projects + outputs
- CSS portfolio is not limited to academic papers / conference proceedings
  - 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)

- Portfolio: projects + outputs
- CSS portfolio is not limited to academic papers / conference proceedings
  - 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)

- "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."

- "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.

- "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.
- If you were a graduate student, use GitHub (there's a student developer pack!) and manage your research processes as well as outputs and document them (using README, etc).

- "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.
- If you were a graduate student, use GitHub (there's a student developer pack!) and 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.

- Take an extra step.

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

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

- Take an extra step.
- If you wrote code, develop a package too.
- If you taught a course, develop a website too.
- Start writing a brag document where you document your kudos.

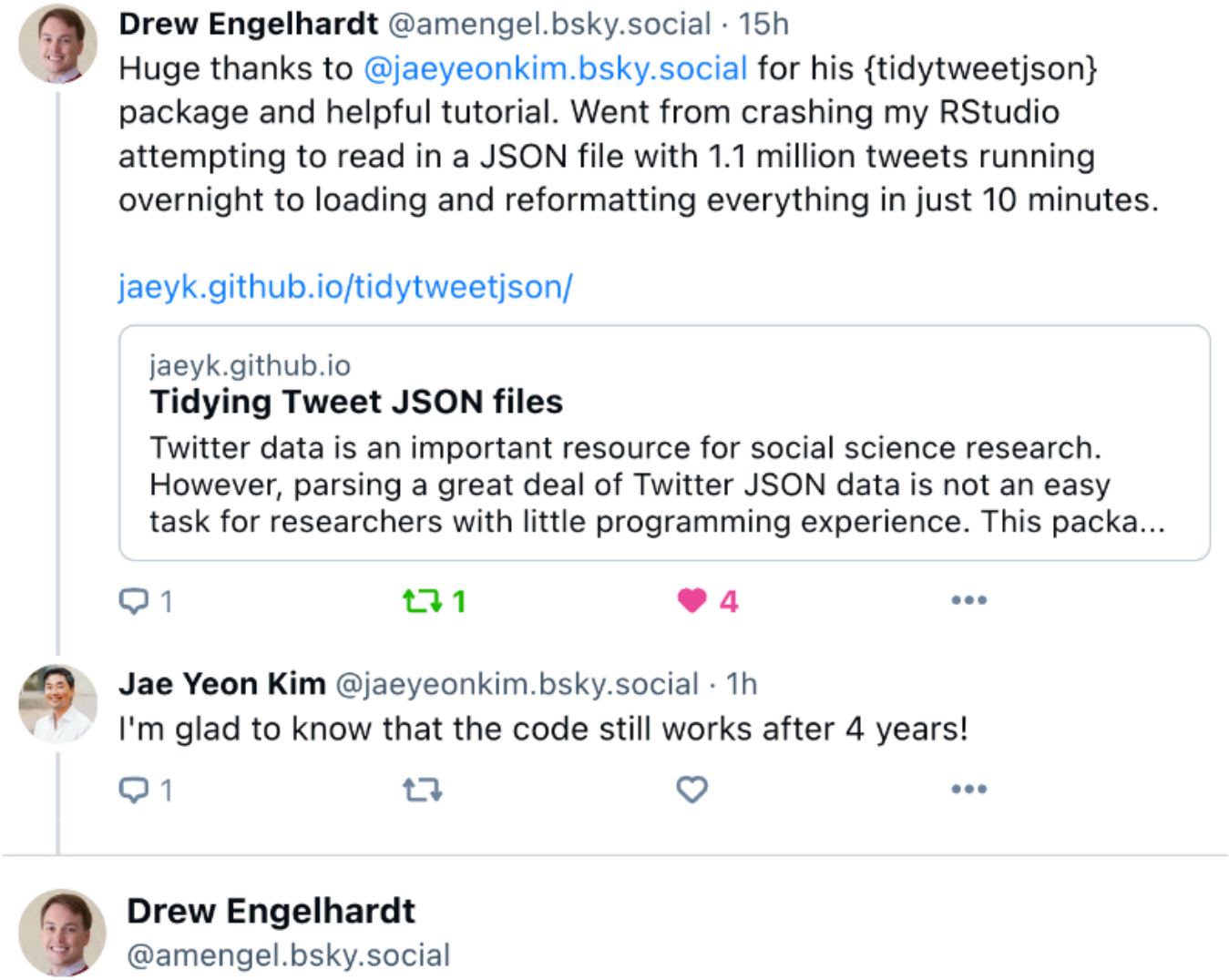

Figure 2: Kudos example 1

Figure 3: Kudos example 2

- "Networking is as valuable to computational social scientists in terms of finding collaborators and jobs; however, it operates slightly differently in CSS because the opportunities to connect span more spaces across disciplines and sectors."

- Tips on conferences.
- The ACM Conference on Human Factors in Computing Systems
- The International Conference on Web and Social Media
- The Text as Data Conference; the Network Science Society Conference
- The International Social Networks Conference
- The Politics and Computational Social Science Conference
- The ACM Conference on Fairness, Accountability, and Transparency
- Many others

- Many of these conferences are international.

- Many of these conferences are international.
- Open to researchers as well as practitioners in the industry, government, and nonprofits.

- Many of these conferences are international.
- Open to researchers as well as practitioners in the industry, government, and nonprofits.
- Conference proceedings matter (highly selective) and get published.

- Many of these conferences are international.
- Open to researchers as well as practitioners in the industry, government, and nonprofits.
- Conference proceedings matter (highly selective) and get published.
- Poster sessions are well-attended.

- Many of these conferences are international.
- Open to researchers as well as practitioners in the industry, government, and nonprofits.
- Conference proceedings matter (highly selective) and get published.
- Poster sessions are well-attended.
- Many disciplinary conferences have added preconferences focused on CSS topics (e.g., APSA's PolNet, PolMeth).

- Tips on internships. If possible, do (summer) internships.

- Tips on internships. If possible, do (summer) internships.
- Places:

- Tips on internships. If possible, do (summer) internships.
- Places:
  - Tech companies (e.g., Meta, Google, Amazon, Microsoft Research)

- Tips on internships. If possible, do (summer) internships.
- Places:
  - Tech companies (e.g., Meta, Google, Amazon, Microsoft Research)
  - Nonprofits (e.g., Urban, Pew, Mathematica, RAND)

- Tips on internships. If possible, do (summer) internships.
- Places:
  - Tech companies (e.g., Meta, Google, Amazon, Microsoft Research)
  - Nonprofits (e.g., Urban, Pew, Mathematica, RAND)
  - Government agencies (e.g., Coding It Forward)

- Tips on internships. If possible, do (summer) internships.
- Places:
  - Tech companies (e.g., Meta, Google, Amazon, Microsoft Research)
  - Nonprofits (e.g., Urban, Pew, Mathematica, RAND)
  - Government agencies (e.g., Coding It Forward)
  - International organizations (e.g., World Bank)

- Tips on internships. If possible, do (summer) internships.
- Places:
  - Tech companies (e.g., Meta, Google, Amazon, Microsoft Research)
  - Nonprofits (e.g., Urban, Pew, Mathematica, RAND)
  - Government agencies (e.g., Coding It Forward)
  - International organizations (e.g., World Bank)
- In general, paid well (compared to the academic standards)

- Tips on internships. If possible, do (summer) internships.
- Places:
  - Tech companies (e.g., Meta, Google, Amazon, Microsoft Research)
  - Nonprofits (e.g., Urban, Pew, Mathematica, RAND)
  - Government agencies (e.g., Coding It Forward)
  - International organizations (e.g., World Bank)
- In general, paid well (compared to the academic standards)
- Other benefits:

- Tips on internships. If possible, do (summer) internships.
- Places:
  - Tech companies (e.g., Meta, Google, Amazon, Microsoft Research)
  - Nonprofits (e.g., Urban, Pew, Mathematica, RAND)
  - Government agencies (e.g., Coding It Forward)
  - International organizations (e.g., World Bank)
- In general, paid well (compared to the academic standards)
- Other benefits:
  - Experience, skills, and networking

- Tips on internships. If possible, do (summer) internships.
- Places:
  - Tech companies (e.g., Meta, Google, Amazon, Microsoft Research)
  - Nonprofits (e.g., Urban, Pew, Mathematica, RAND)
  - Government agencies (e.g., Coding It Forward)
  - International organizations (e.g., World Bank)
- In general, paid well (compared to the academic standards)
- Other benefits:
  - Experience, skills, and networking
  - Useful to decide their career paths

- Tips on internships. If possible, do (summer) internships.
- Places:
  - Tech companies (e.g., Meta, Google, Amazon, Microsoft Research)
  - Nonprofits (e.g., Urban, Pew, Mathematica, RAND)
  - Government agencies (e.g., Coding It Forward)
  - International organizations (e.g., World Bank)
- In general, paid well (compared to the academic standards)
- Other benefits:
  - Experience, skills, and networking
  - Useful to decide their career paths
- Highly selective (prepare early)

Plan

1. Three-step framework
   - Learning data science skills as a social scientist
   - Building CSS portfolio
   - Networking in CSS
2. Conclusion and Discussions

Recommendations

- Some recommendations for departments

Recommendations

- Some recommendations for departments
- Provide information on non-academic career opportunities, including internships, to students at the beginning of PhD training

Recommendations

- Some recommendations for departments
- Provide information on non-academic career opportunities, including internships, to students at the beginning of PhD training
- Integrate data science skills building into existing curriculum (e.g., integrating R or Python in introductory statistics courses)

Recommendations

- Some recommendations for departments
- Provide information on non-academic career opportunities, including internships, to students at the beginning of PhD training
- Integrate data science skills building into existing curriculum (e.g., integrating R or Python in introductory statistics courses)
- Offer new courses on computational methods and data management

Recommendations

- Some recommendations for departments
- Provide information on non-academic career opportunities, including internships, to students at the beginning of PhD training
- Integrate data science skills building into existing curriculum (e.g., integrating R or Python in introductory statistics courses)
- Offer new courses on computational methods and data management
- Identify relevant data science coursework in other departments and recognize earned credits

- Identify relevant data science faculty in other departments who can serve on dissertation committees

- Identify relevant data science faculty in other departments who can serve on dissertation committees
- Offer options for students to substitute a program requirement (e.g., one field exam) for an internship or advanced CSS training

- Provide support for current faculty to pursue CSS training

- Provide support for current faculty to pursue CSS training
- Hire more CSS faculty and recruit computational social scientists from industry and nonprofit organizations for faculty and visiting scholar positions

- Provide support for current faculty to pursue CSS training
- Hire more CSS faculty and recruit computational social scientists from industry and nonprofit organizations for faculty and visiting scholar positions
- Evolve publication standards to increasingly value CSS conference proceedings, journals, and the value of collaborative CSS project work

Thank you

Comments or questions?
E-mail: jkim638@jhu.edu