GEORGETOWN UNIVERSITY

The Secure Query System

Providing Privacy-Preserving Statistics on Administrative Tax Data

Joshua Snoke

Massive Data Institute

What is the Secure Query System (SQS)?

- Collaborative project between MDI, Yale, and IRS Statistics of Income (SOI)
- SQS provides clients with aggregate statistics from IRS earnings data
  - Securely linking client data with tax data
  - Applying disclosure avoidance to ensure privacy is preserved in the statistics

<u>MDI Team:</u> *Amy O'Hara, Ella Blue, Jake Pasner, Kangheng Liu, Stephanie Straus, Vince Dorie*

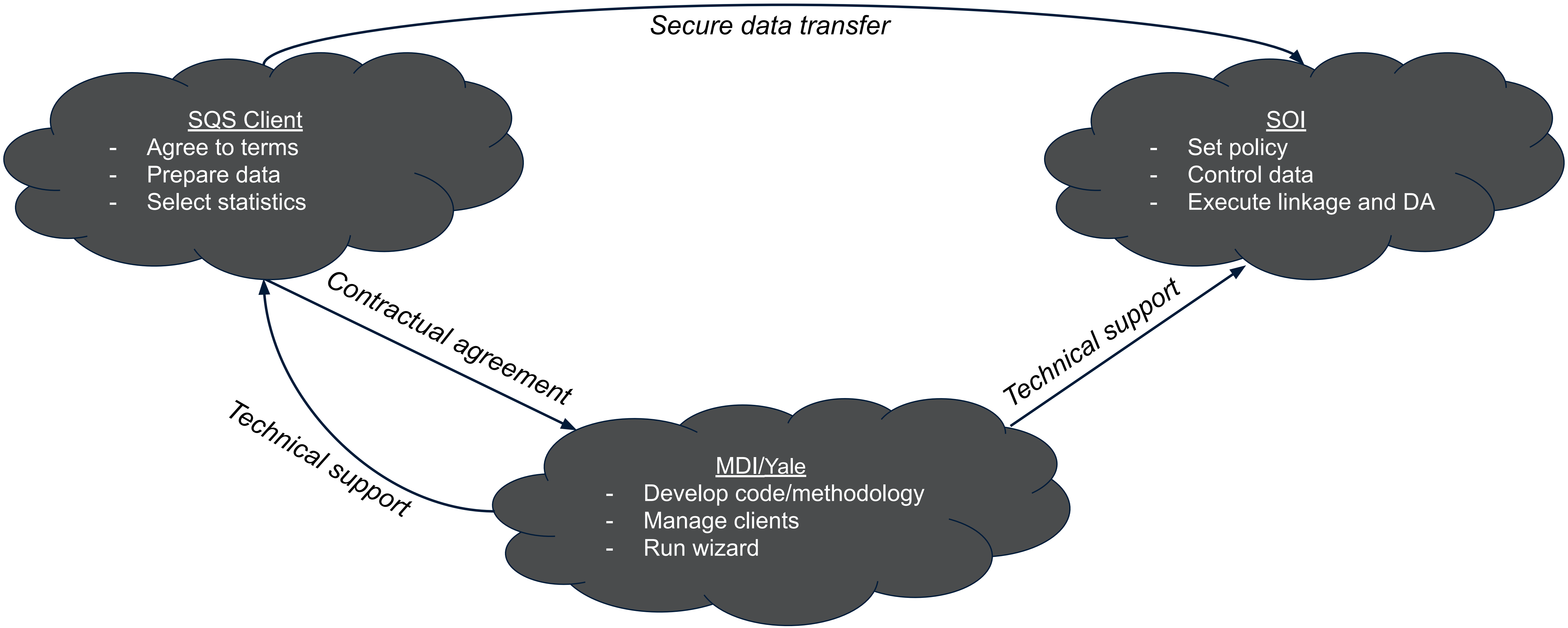
<u>Yale Team:</u> *Eliza McKenny, Kathy Stack, Ron Borzekowski*

<u>SOI Team:</u> *Brian Foltz, Caiyi Lang, Corbin Miller, Kelly Dauberman, Mark Xu*

SQS is a New Tier of Access

- Historically, access to earnings data has been restricted
  - Joint Statistical Research Program → **highly selective**
  - Published tables or the Public Use File (PUF)* microdata → **limited utility**
- For many, **linking earnings** to a specific cohort is crucial for evaluation
  - Training programs
  - Policy initiatives
  - Post-secondary outcomes

Overview of SQS Entities



The Key Advantages of SQS

- Streamlined contractual agreements
  - Remove administrative burden from SOI
- Data standardization and validation
  - No need for customized inputs or data cleaning
- Automated disclosure avoidance
  - Fast review by SOI staff
- ***SQS is designed from start to finish to be efficient and sustainable***

Pathway to SQS - Define Client Use Cases

- MDI/Yale conducted extensive outreach
  - State or local agencies
  - Non-profit service providers
- Use cases abound
  - Education (K-12, Post-Secondary Education Institutions or Systems)
  - Workforce (Training and Employment Programs)
  - Health and Human Services (Crossover Youth, SNAP, TANF, Child Welfare)
  - Justice (Training, Reentry)
  - Housing (Housing Subsidy Recipients, Homeless/Transitional Housing Services)

Pathway to SQS - Standardize Statistics

- Develop set of measures that answer common user requests
  - Value comes from data coverage and quality
- Users provide grouping information (covariates)
  - Enables extensions to hypothesis testing, multivariate analyses
- Focus on common sources of earnings, common tax credits

SQS Statistics: 1040 Product

Available statistics for matched individuals with an IRS Form 1040

Measure	N	Mean	Variance	Quartiles
Filed 1040s	yes	–	–	–
Non-zero CTC on 1040	yes	yes	yes	yes
Non-zero EITC on 1040	yes	yes	yes	yes
Non-zero Wages on 1040	yes	yes	yes	yes
Non-zero AGI on 1040	yes	yes	yes	yes
Schedule C on 1040	yes	–	–	–

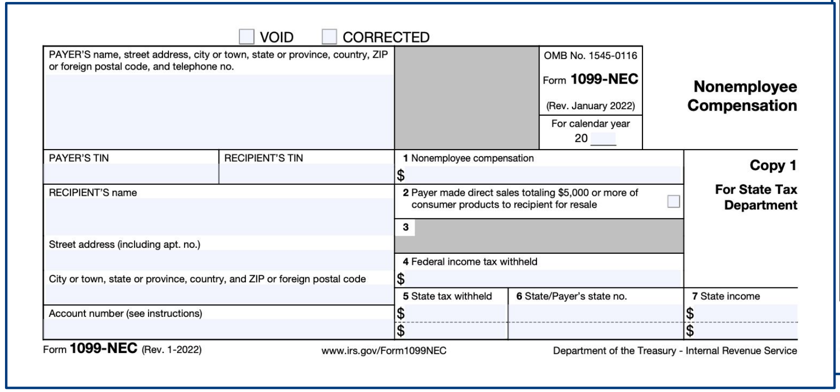
SQS Statistics: Info Product

Available statistics for matched individuals with 1 or more Form W-2 or Form 1099-NEC

Measure	N	Mean	Variance	Quartiles
Total Non-zero W2(s) Income	yes	yes	yes	yes
Total Non-zero 1099NEC(s) Compensation	yes	yes	yes	yes
Combined W2(s) Income and 1099NEC(s) Compensation	yes	yes	yes	yes
Number of W2(s) Forms with Non-Zero Income	–	yes	yes	yes
Number of 1099NEC(s) Forms with Non-Zero Compensation	–	yes	yes	yes
All W2(s) In-State of Choice	yes	–	–	–
Some W2(s) In-State of Choice	yes	–	–	–
All W2(s) Out-of-State of Choice	yes	–	–	–

Pathway to SQS - Link Records

- Need to link individuals with Form 1040, Form W-2, and Form 1099-NEC
  - Varying PII on forms
- Conducted tests to determine optimal matching strategies
- No retention of PII by SOI → no connection to clients



VOID CORRECTED

PAYER'S name, street address, city or town, state or province, country, ZIP or foreign postal code, and telephone no.

OMB No. 1545-0116

Form 1099-NEC

(Rev. January 2022)

For calendar year 20__

Nonemployee Compensation

PAYER'S TIN | RECIPIENT'S TIN

1 Nonemployee compensation $

Copy 1 For State Tax Department

RECIPIENT'S name

2 Payer made direct sales totaling $5,000 or more of consumer products to recipient for resale

3

Street address (including apt. no.)

4 Federal income tax withheld $

City or town, state or province, country, and ZIP or foreign postal code

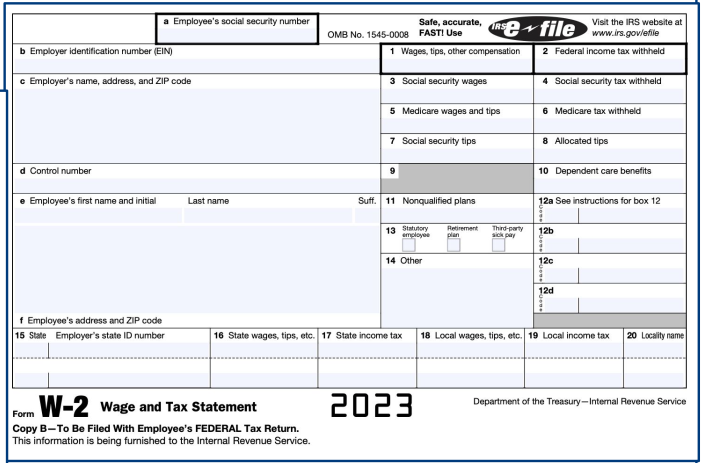
5 State tax withheld $ $

6 State/Payer's state no.

7 State income $ $

Account number (see instructions)

Form 1099-NEC (Rev. 1-2022) www.irs.gov/Form1099NEC Department of the Treasury - Internal Revenue Service



a Employee's social security number

OMB No. 1545-0008

Safe, accurate, FAST! Use IRS e-file

Visit the IRS website at www.irs.gov/efile

b Employer identification number (EIN)

1 Wages, tips, other compensation

2 Federal income tax withheld

c Employer's name, address, and ZIP code

3 Social security wages

4 Social security tax withheld

5 Medicare wages and tips

6 Medicare tax withheld

7 Social security tips

8 Allocated tips

d Control number

9

10 Dependent care benefits

e Employee's first name and initial | Last name | Suff.

11 Nonqualified plans

12a See instructions for box 12

13 Statutory employee | Retirement plan | Third-party sick pay

12b

14 Other

12c

12d

f Employee's address and ZIP code

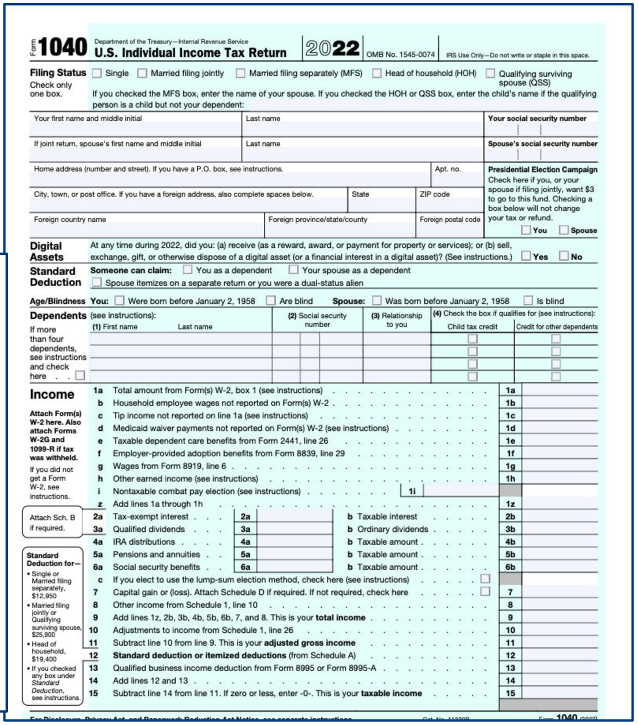
15 State | Employer's state ID number | 16 State wages, tips, etc. | 17 State income tax | 18 Local wages, tips, etc. | 19 Local income tax | 20 Locality name

Form W-2 Wage and Tax Statement 2023

Department of the Treasury—Internal Revenue Service

Copy B—To Be Filed With Employee's FEDERAL Tax Return.

This information is being furnished to the Internal Revenue Service.



Form 1040 Department of the Treasury—Internal Revenue Service U.S. Individual Income Tax Return 2022 OMB No. 1545-0074 IRS Use Only—Do not write or staple in this space.

Filing Status Check only one box. Single Married filing jointly Married filing separately (MFS) Head of household (HOH) Qualifying surviving spouse (QSS)

If you checked the MFS box, enter the name of your spouse. If you checked the HOH or QSS box, enter the child's name if the qualifying person is a child but not your dependent:

Your first name and middle initial | Last name | Your social security number

If joint return, spouse's first name and middle initial | Last name | Spouse's social security number

Home address (number and street). If you have a P.O. box, see instructions. | Apt. no.

Presidential Election Campaign Check here if you, or your spouse if filing jointly, want $3 to go to this fund. Checking a box below will not change your tax or refund. You Spouse

City, town, or post office. If you have a foreign address, also complete spaces below. | State | ZIP code

Foreign country name | Foreign province/state/county | Foreign postal code

Digital Assets At any time during 2022, did you: (a) receive (as a reward, award, or payment for property or services); or (b) sell, exchange, gift, or otherwise dispose of a digital asset (or a financial interest in a digital asset)? (See instructions.) Yes No

Standard Deduction Someone can claim: You as a dependent Your spouse as a dependent Spouse itemizes on a separate return or you were a dual-status alien

Age/Blindness You: Were born before January 2, 1958 Are blind Spouse: Was born before January 2, 1958 Is blind

Dependents (see instructions): (1) First name Last name (2) Social security number (3) Relationship to you (4) Check the box if qualifies for (see instructions): Child tax credit Credit for other dependents

If more than four dependents, see instructions and check here

Income

1a Total amount from Form(s) W-2, box 1 (see instructions)

b Household employee wages not reported on Form(s) W-2

c Tip income not reported on line 1a (see instructions)

d Medicaid waiver payments not reported on Form(s) W-2 (see instructions)

e Taxable dependent care benefits from Form 2441, line 26

f Employer-provided adoption benefits from Form 8839, line 29

g Wages from Form 8919, line 6

h Other earned income (see instructions)

i Nontaxable combat pay election (see instructions)

z Add lines 1a through 1h

Attach Form(s) W-2 here. Also attach Forms W-2G and 1099-R if tax was withheld. If you did not get a Form W-2, see instructions.

Attach Sch. B if required.

2a Tax-exempt interest | 2a | b Taxable interest | 2b

3a Qualified dividends | 3a | b Ordinary dividends | 3b

4a IRA distributions | 4a | b Taxable amount | 4b

5a Pensions and annuities | 5a | b Taxable amount | 5b

6a Social security benefits | 6a | b Taxable amount | 6b

c If you elect to use the lump-sum election method, check here (see instructions)

7 Capital gain or (loss). Attach Schedule D if required. If not required, check here

8 Other income from Schedule 1, line 10

9 Add lines 1z, 2b, 3b, 4b, 5b, 6b, 7, and 8. This is your total income

10 Adjustments to income from Schedule 1, line 26

11 Subtract line 10 from line 9. This is your adjusted gross income

12 Standard deduction or itemized deductions (from Schedule A)

13 Qualified business income deduction from Form 8995 or Form 8995-A

14 Add lines 12 and 13

15 Subtract line 14 from line 11. If zero or less, enter -0-. This is your taxable income

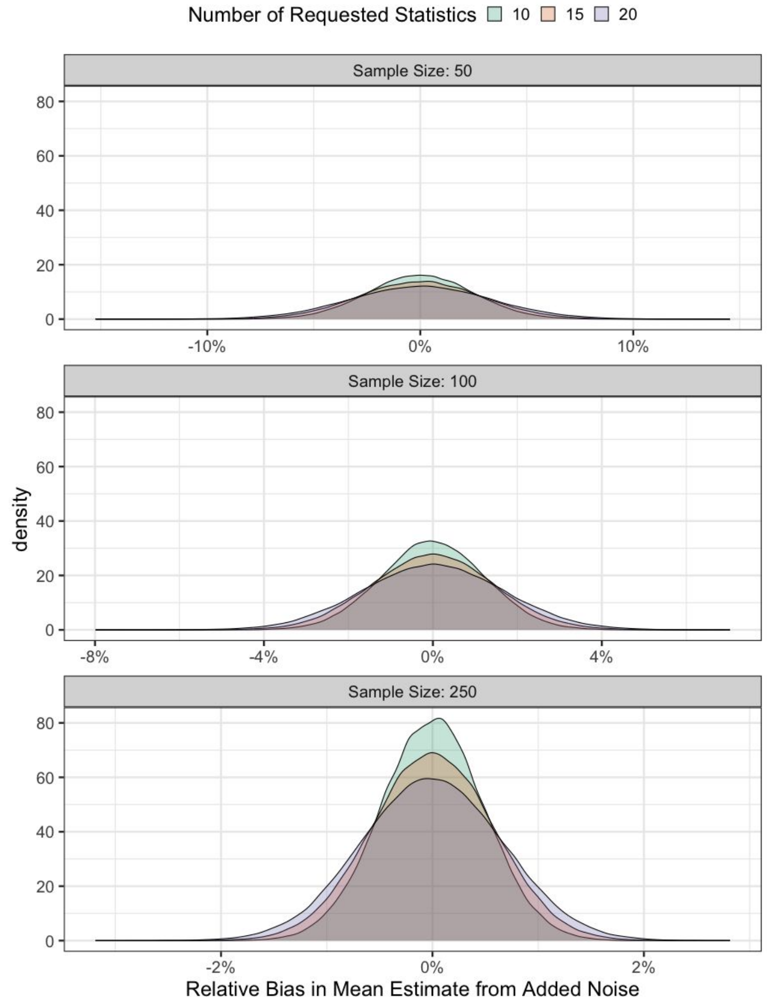
Standard Deduction for— • Single or Married filing separately, $12,950 • Married filing jointly or Qualifying surviving spouse, $25,900 • Head of household, $19,400 • If you checked any box under Standard Deduction, see instructions.

Pathway to SQS - Preserve Privacy

- Data minimization
  - No data retention from linkage
  - Only provide users with statistics they request
- Satisfying IRS Publication 1075
  - No statistics based on fewer than 20 records
  - Noise addition to protect against disclosing fact of filing or return information
- Follow ethical and technical best practices
  - Using gaussian differential privacy with per-record privacy loss accounting framework
  - Scaling protections to number of requested statistics, magnitude of measures
  - Enables tracking and managing cumulative risk

Pathway to SQS - Provide Valid Outputs

- Clients receive confidence intervals (CIs) that account for both sampling and noise addition
- Clients receive expected amount of noise added, based on:
  - Number of statistics requested
  - Number of input records
  - Expected match/filing rates



Takeaways for Applied Data Privacy Work

- Add value the easy way
  - Create simple products with rich untapped data
- Walk before you run
  - Improve the DA but be willing to relax theoretical best practices
- Always read the rule book
  - Know how to satisfy the regulations
- Cost is king
  - Make the process fast and efficient for responsible agency staff

Next Steps

- SQS is going live within the next couple weeks!
- Planned future improvements:
  - Additional tooling for technical support for clients
    - Impact of disclosure avoidance
    - Guidance for creating grouping variables
  - Record-level privacy loss tracking infrastructure
    - Enable longitudinal studies
    - Improved privacy accounting

Next Steps

- SQS is going live within the next couple weeks!
- Planned future improvements:
  - Additional tooling for technical support for clients
    - Impact of disclosure avoidance
    - Guidance for creating grouping variables
  - Record-level privacy loss tracking infrastructure
    - Enable longitudinal studies
    - Improved privacy accounting

Thank you! Questions?

joshua.snoke@georgetown.edu