

## 1. Introduction

This poster describes research that seeks to establish a baseline of Excel proficiency among undergraduates in a Principles of Economics course at a large U.S. public university. Data literacy is an important skill for the economics curriculum (Halliday 2019, Hoyt and McGoldrick 2017), and is in demand by employers of college graduates (Marshall and Underwood 2020). According to Conerly 2017 and Barreto 2015, Microsoft Excel is the most widely used program for data analysis in industry and is a recommended skill to be taught in undergraduate economics programs.

We administered an IRB approved survey in large lecture classes of Principles of Microeconomics to measure proficiency in several fundamental Excel commands. The respondents are asked to analyze MLB data using Excel functions and respond to the questions while reporting their degree of confidence for each correct answer. An image of the data and questions are shown below.

PLAYER DATA							
First Name	Last Name	Full Name	Fantasy	Home Runs	Batting Avg	Dummy	Previous
Aaron	Hicks	AaronHicks	133	10	0.228		
Adeniy	Hechavarría	AdeniyHechavarría	189	8	0.238		
Albert	Almonor	AlbertAlmonor	158	5	0.246		
Austin	Meadows	AustinMeadows	183	12	0.263		
Austin	Riley	AustinRiley	100	9	0.27		
Austin	Romine	AustinRomine	161	7	0.275		
Austin	Slater	AustinSlater	177	11	0.221		
Boog	Powell	BoogPowell	112	1	0.279		
Brandon	Crawford	BrandonCrawford	189	6	0.236		
Brandon	Drury	BrandonDrury	112	7	0.269		
Brandon	Lowe	BrandonLowe	172	15	0.231		
Brett	Gardner	BrettGardner	156	4	0.241		
Brett	Phillips	BrettPhillips	174	14	0.233		
Brook	Holt	BrookHolt	86	2	0.24		
Geoff	Mullins	GeoffMullins	125	7	0.298		
Chad	Pinder	ChadPinder	159	12	0.245		
Chane	Stiso	ChaneStiso	140	10	0.237		
Christian	Vazquez	ChristianVazquez	181	11	0.243		
Corey	Dickerson	CoreyDickerson	202	11	0.243		
Cristian	Pache	CristianPache	149	6	0.238		
Danny	Jansen	DannyJansen	183	11	0.258		
Dansby	Swanson	DansbySwanson	83	6	0.243		
Daniel	Lugo	DanielLugo	108	4	0.261		
Derek	Dietrich	DerekDietrich	126	10	0.235		
Deven	Marrero	DevenMarrero	157	8	0.237		
Domingo	Santana	DomingoSantana	164	18	0.238		
Donovan	Selano	DonovanSelano	120	8	0.246		
Eddie	Rosario	EddieRosario	181	8	0.247		
Eric	Segund	EricSegund	200	9	0.246		
Eugenio	Suarez	EugenioSuarez	192	24	0.189		
Evan	Longoria	EvanLongoria	182	8	0.243		
Evan	White	EvanWhite	148	9	0.281		
Francisco	Cervelli	FranciscoCervelli	97	6	0.218		
Francisco	Mejia	FranciscoMejia	176	7	0.275		
Franklin	Barreto	FranklinBarreto	123	13	0.222		
Grady	Gschler	GradyGschler	106	6	0.153		

Questions: Enter your response, and your self repo completion of all questions, return to the Qualtrics

1) You are asked to SORT the first name column in ascending order. Report the 5th player from the top and the last player:

YOUR ANSWER:

2-4) Your next task is to calculate statistics for the Home Runs variable. You are asked to calculate the TOTAL (SUM) of Home Runs, the AVERAGE, and the MEDIAN. YOUR ANSWERS:

5) Create an XY scatter plot of Fantasy Points and Home Runs. Describe the general relation that you see between the two variables. Choose "positive", "negative" or "no relation" YOUR DESCRIPTION:

6) In the column labeled Dummy, use the IF THEN function to have the rows in Dummy variable column equal 1 when Home Runs > 18. Report the sum of dummy variables.

YOUR ANSWER:

7) Run a regression of Fantasy Points on Home Runs. Report the Intercept. (Two decimal points, do not round.)

YOUR ANSWER:

8) Sort the data by Last Name. In the column named Previous Year Points, use the VLOOKUP function to copy each player's Previous Year Points from worksheet Previous Player Points. Report the value for Evan Longoria.

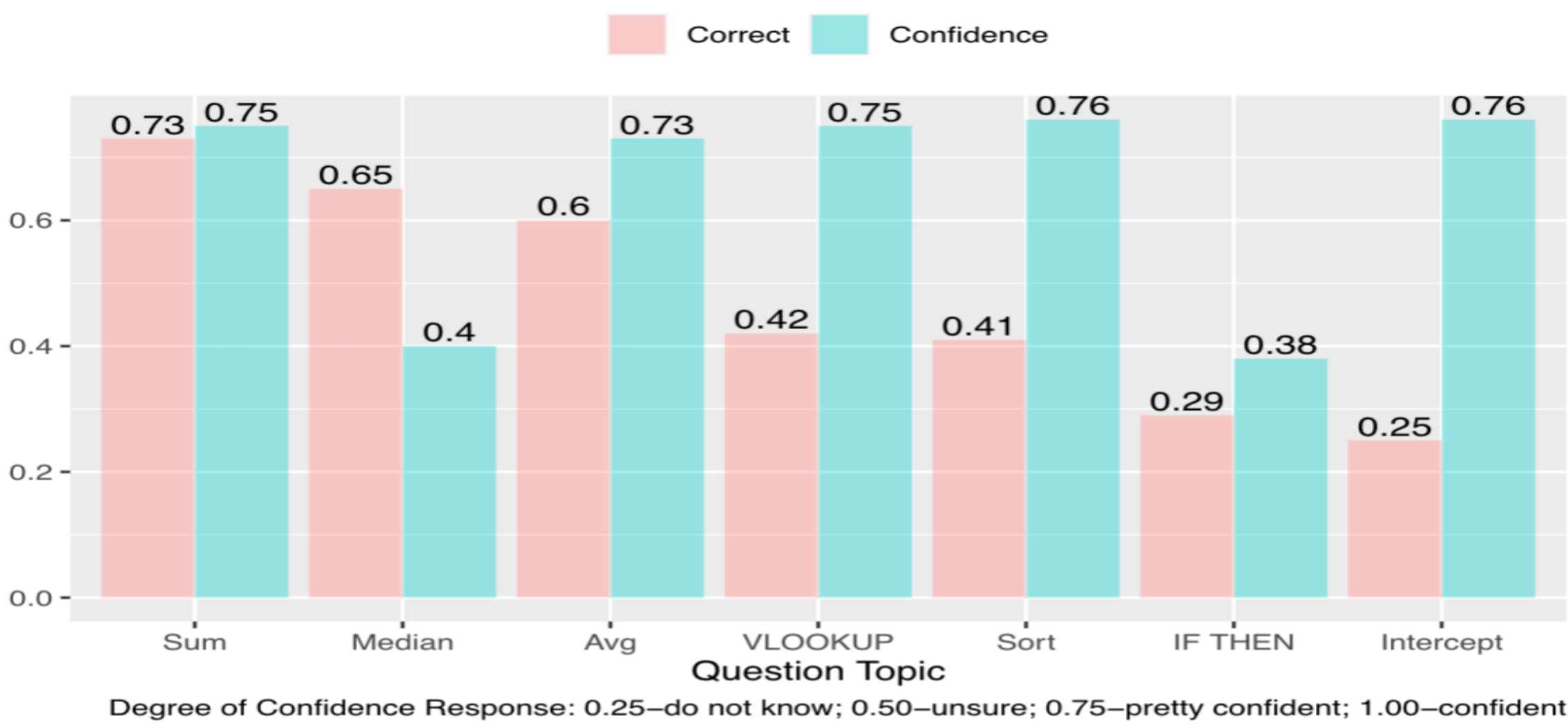
YOUR ANSWER \_\_\_\_\_

## 2. Overall Results

Table 1: Descriptive Statistics of Variables

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Percent Correct	468	0.476	0.287	0.000	0.286	0.714	1.000
Percent Correct Log	468	0.370	0.204	0.000	0.251	0.539	0.693
Average Confidence	468	2.459	0.683	1.000	2.143	2.857	4.000
Cumulative GPA	468	3.587	0.411	0.817	3.400	3.862	4.000
Male	468	0.421	0.494	0	0	1	1
Female	468	0.579	0.494	0	0	1	1
Asian	468	0.122	0.327	0	0	0	1
Black	468	0.038	0.193	0	0	0	1
Hispanic	468	0.083	0.277	0	0	0	1
Multiracial	468	0.049	0.216	0	0	0	1
Non-Resident Alien	468	0.030	0.171	0	0	0	1
White	468	0.677	0.468	0	0	1	1

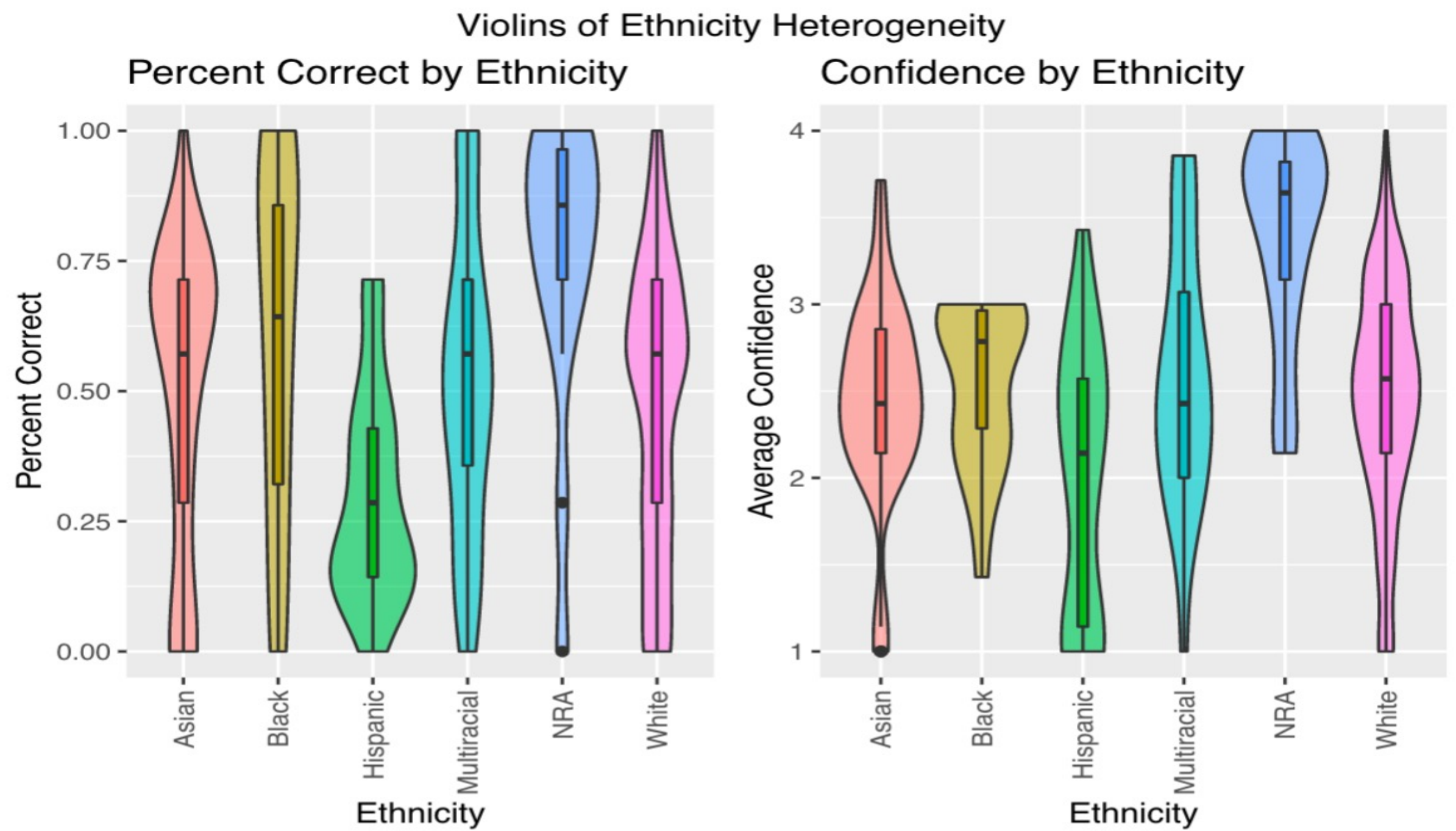
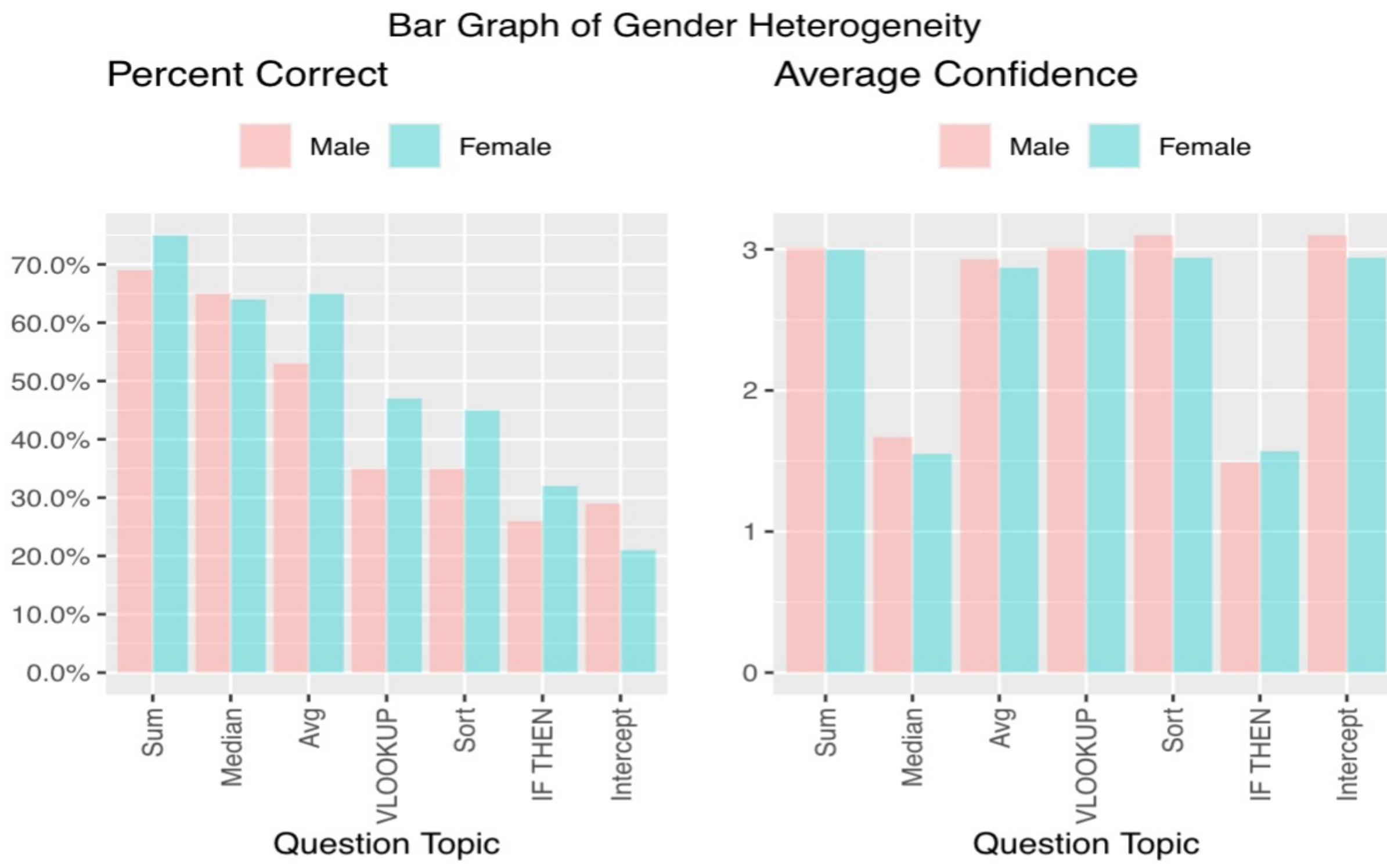
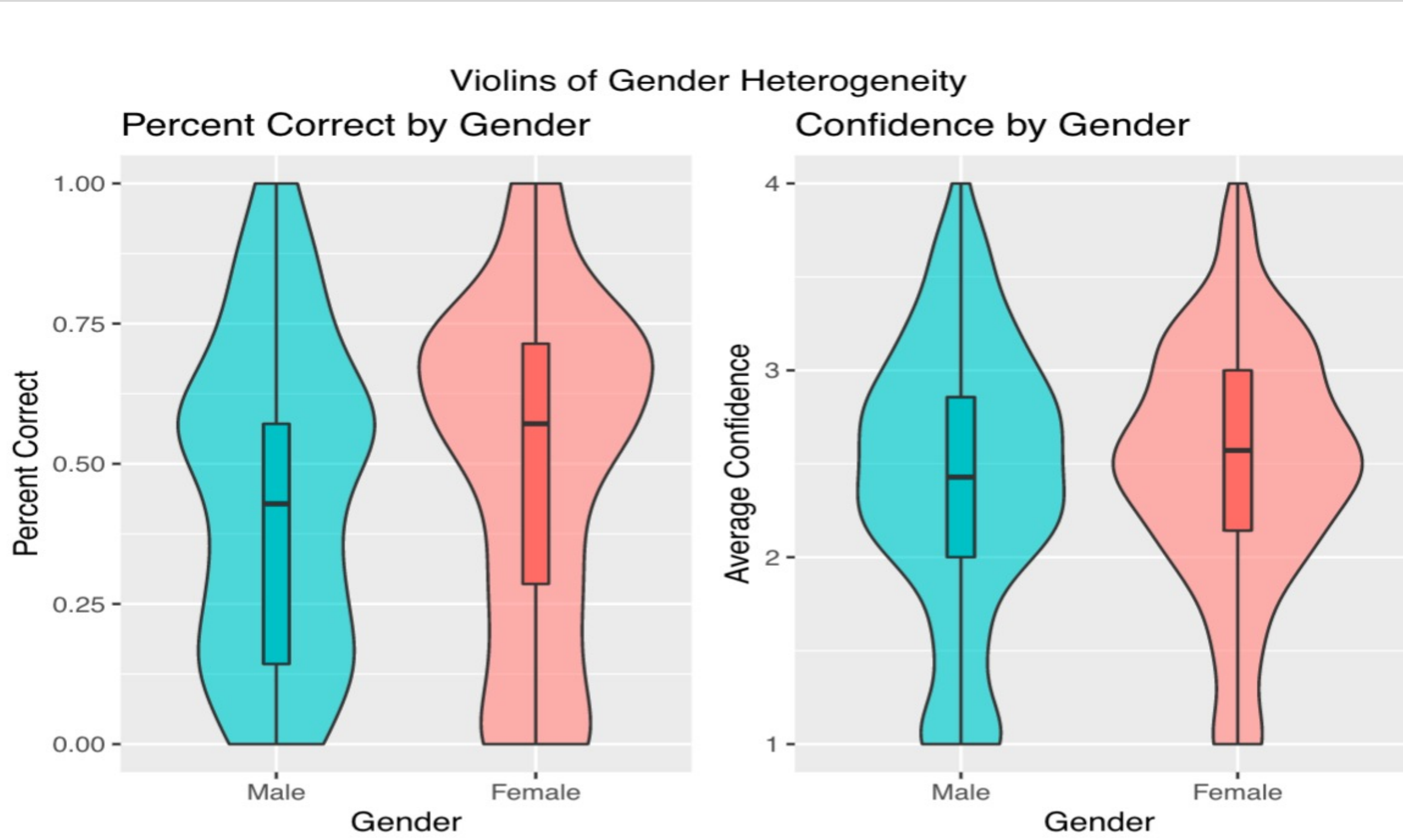
Percent Correct and Degree of Confidence per Question



## 3. Heterogeneity by Gender and Ethnicity

### Preliminary Results

- Findings indicate that, as expected, the percentage of correct responses decline significantly from the relatively simple Excel function of SUM to the more complex REGRESSION function. The results suggest significant gender differences in percentage correct and degree of confidence. The output also suggests significant heterogeneity among ethnic groups.



## 4. Regression Results

- All Ethnic groups are compared against the Asian group, which was dropped to avoid multicollinearity. Similarly, students reporting self confidence of one are the baseline for comparison of confidence factors.

Table 2: Regression of Correct Percent Log on Covariates

	Dependent variable:
	Percent Correct Log
Cumulative GPA	0.075*** (0.018)
Male	−0.013 (0.015)
Black	0.090** (0.043)
Latino	−0.043 (0.033)
NRA	0.028 (0.050)
Multiracial	0.0004 (0.038)
White	−0.015 (0.022)
Confidence 2	0.212*** (0.025)
Confidence 3	0.361*** (0.025)
Confidence 4	0.491*** (0.041)
Intercept	−0.155** (0.073)
Observations	468
R <sup>2</sup>	0.445
Adjusted R <sup>2</sup>	0.433
Residual Std. Error	0.154 (df = 457)
F Statistic	36.651*** (df = 10; 457)
Note:	*p<0.1; **p<0.05; ***p<0.01

### Future Research

- Future research plans include developing homework exercises for Principles of Microeconomics where students reinforce learning the fundamental economic concepts by analyzing data with Excel. Then using a randomized experimental design, and administering a pre-/post-survey, investigate their effectiveness of promoting Excel and data literacy. We will also explore a multinomial logit model to better understand the determinants of factors influencing each degree of confidence selected. An LPM model will also be examined to evaluate covariate probabilities of a correct response.

## 5. References

Barreto. 2015. Why Excel?. *Journal of Economic Education* 46 (3): 300–309.

Conerly. 2017. *Forbes*, May 4, 2017.

Halliday, S. D. 2019. Data literacy in economic development. *Journal of Economic Education* 50 (3): 284–98.

Hoyt, G. M., and K. McGoldrick. 2017. Promoting undergraduate research in economics. *American Economic Review* 107 (5): 655–59.

Marshall and Underwood. 2020. Is Economics STEM? : Trends in the discipline from 1997 to 2018. *Journal of Economic Education* 51(2): 167-174.