Supplemental Online Material

Community-Engaged "Data Fest" Events: Applying Econometric Skills to Build Confidence

By Caroline Krafft, Kristine West, Meghan Mason, Louise Ba, and Joy Moua

Table 4. Factor analysis

	Scoring coefficient	Loading	Uniqueness
Relevance			
Data analysis skills will make me more employable	0.044	0.348	0.879
I will have some application for data analysis and statistical software skills in my profession	0.052	0.409	0.832
Data analysis gives me a useful framework for thinking about issues that are important to my life	0.101	0.502	0.748
Data analysis is relevant for social justice work	0.055	0.392	0.846
Belonging			
I feel different than the typical student in data science or economics	-0.023	-0.188	0.965
I feel comfortable asking questions when I am in a group working on a data project	0.046	0.510	0.740
Growth mindset			
I can learn how to use statistical software	0.115	0.590	0.652
I find it easy to understand data analysis because of how I think	0.072	0.566	0.679
I can learn how to analyze large data sets	0.086	0.592	0.649
Data analysis skills are quickly learned by most people	0.032	0.239	0.943
Learning statistical software requires a great deal of discipline	0.005	-0.018	1.000
I can learn the skills I need to become a data scientist or economist	0.081	0.515	0.735
I believe I can learn how to do data analysis	0.150	0.632	0.600
Confidence			
I like using statistical software to analyze data	0.080	0.689	0.525
I feel confident using statistical software to analyze data	0.273	0.756	0.428
Statistical software code is easy to understand	0.066	0.438	0.808
I understand the basics of how to analyze large data sets using statistical software	0.132	0.698	0.513
Data analysis is complicated	-0.031	-0.294	0.914
I get frustrated using statistical software to analyzing data	-0.083	-0.361	0.870
I am under stress while using statistical software to analyze data	-0.091	-0.526	0.723
How comfortable you are running R	0.041	0.481	0.768
How comfortable you are running Stata	-0.011	0.215	0.954
Eigenvalue	5.230		

Source: Authors' calculations based on all observations

Table 5. Paired t-tests from pre-first to post-second Data Fest

	Pre-first Data Fest	Post-second Data Fest	P-value paired t-test
Relevance			
Data analysis skills will make me more employable	1.88	1.98	0.10*
I will have some application for data analysis and statistical software skills in my profession	1.67	1.81	0.11
Data analysis gives me a useful framework for thinking about issues that are important to my life	1.81	1.92	0.13
Data analysis is relevant for social justice work	1.92	1.94	0.71
Belonging			
I feel different than the typical student in data science or economics	0.13	0.58	0.02**
I feel comfortable asking questions when I am in a group working on a data project	1.42	1.71	0.02**
Growth mindset			
I can learn how to use statistical software	1.94	1.96	0.57
I find it easy to understand data analysis because of how I think	0.58	0.88	0.06*
I can learn how to analyze large data sets	1.81	1.79	0.81
Data analysis skills are quickly learned by most people	-0.31	-0.21	0.50
Learning statistical software requires a great deal of discipline	1.04	1.10	0.66
I can learn the skills I need to become a data scientist or economist	1.77	1.79	0.85
I believe I can learn how to do data analysis	1.85	1.92	0.26
Confidence			
I like using statistical software to analyze data	1.19	1.69	0.00***
I feel confident using statistical software to analyze data	0.04	1.06	0.00***
Statistical software code is easy to understand	-0.21	0.06	0.15
I understand the basics of how to analyze large data sets using statistical software	0.75	1.63	0.00***
Data analysis is complicated	0.90	1.00	0.42
I get frustrated using statistical software to analyzing data	0.23	0.44	0.18
I am under stress while using statistical software to analyze data	0.02	-0.10	0.53
How comfortable you are running R	2.27	2.50	0.30
How comfortable you are running Stata	1.90	3.17	0.00***
Factor variable			
Factor	-0.19	0.31	0.00***

Source: Authors' calculations based on pre-first and post-second Data Fest.
Notes: *p<0.1; **p<0.05; ***p<0.01

Table 6. Fixed effects ordered logit regression models for relevance, belonging, and growth mindset outcomes

	Relevance		Belonging		Growth mindset					
	Profession	Framework	Feel different	Group questions	Understand data	Learn analysis	Quick learning	Software req. discipline	Data sci. or econ.	Belief learn
Pre vs. post (pre omit.)										
Post Data Fest	0.876	1.852**	0.672**	0.422	0.696	1.471**	0.343	0.499	0.260	2.256**
	(0.683)	(0.738)	(0.313)	(0.428)	(0.430)	(0.577)	(0.383)	(0.336)	(0.636)	(1.026)
Calendar year (2021 omit.)										
2022	17.400***	18.134***	-1.446*	0.725	1.065***	1.071	1.248	0.593	17.380***	1.450
	(0.799)	(0.617)	(0.869)	(0.739)	(0.409)	(1.071)	(0.892)	(0.545)	(0.590)	(1.209)
2023	18.387***	19.169***	0.361	1.869**	0.467	-0.743	1.328	1.582**	17.673***	0.242
	(1.197)	(1.535)	(1.109)	(0.894)	(0.592)	(0.928)	(0.984)	(0.721)	(1.085)	(1.539)
First or second Data Fest of the year (first omit.)										
Second Data Fest of the year	0.422	-0.928	0.332	1.376***	0.094	-1.530**	0.007	-0.015	0.064	-0.984
	(0.676)	(0.798)	(0.344)	(0.496)	(0.316)	(0.741)	(0.362)	(0.406)	(0.584)	(0.973)
N (Obs.)	77	65	157	116	146	74	150	133	72	42
N (individuals)	23	18	44	33	43	21	43	37	20	12

Notes: *p<0.1; **p<0.05; ***p<0.01. Cells are coefficients. Clustered standard errors in parentheses. Note that N (number of observations) varies as only clusters (students) with variation in outcomes are included.

Table 6 continued. Fixed effects ordered logit regression models for confidence outcomes

	Confidence								
	Like software	**Confident software**	**Code easy**	**Software understanding**	**Complicated**	**Frustrated software**	**Stress software**	**R comfort**	**Stata comfort**
Pre vs. post (pre omit.)									
Post Data Fest	1.275***	1.813***	0.767**	1.675***	-0.118	0.303	-0.017	0.222	0.436
	(0.463)	(0.497)	(0.345)	(0.492)	(0.466)	(0.325)	(0.294)	(0.414)	(0.339)
Calendar year (2021 omit.)									
2022	0.573	1.450**	0.576	1.402	1.107	-1.059	-0.243	1.307**	-1.107
	(1.115)	(0.730)	(0.512)	(0.894)	(0.756)	(0.743)	(0.969)	(0.658)	(1.043)
2023	1.632	4.282***	0.267	4.619***	1.064	-0.963	-0.680	3.481***	-0.434
	(1.630)	(1.263)	(0.559)	(1.538)	(0.930)	(0.836)	(1.093)	(0.784)	(0.906)
First or second Data Fest of the year (first omit.)									
Second Data Fest of the year	1.173***	1.883***	0.302	1.975***	0.510	0.223	-0.410	0.322	1.629***
	(0.444)	(0.353)	(0.350)	(0.452)	(0.435)	(0.302)	(0.307)	(0.334)	(0.317)
N (Obs.)	121	165	154	154	127	173	168	143	172
N (individuals)	37	47	45	45	39	50	48	40	52

Notes: *p<0.1; **p<0.05; ***p<0.01. Cells are coefficients. Clustered standard errors in parentheses. Note that N (number of observations) varies as only clusters (students) with variation in outcomes are included.

Table 7. Fixed effects OLS regression models for relevance, belonging, and growth mindset outcomes, with controls

	Relevance				Belonging		Growth mindset						
	Employable	**Profession**	**Framework**	**Social justice**	**Feel different**	**Group questions**	**Learn software**	**Understand data**	**Learn analysis**	**Quick learning**	**Software req. discipline**	**Data sci. or econ.**	**Belief learn**
Pre vs. post (pre omit.)													
Post Data Fest	0.035	0.102	0.127	0.027	0.256*	-0.009	0.016	0.110	0.065	0.143	0.253***	0.021	0.088
	(0.034)	(0.116)	(0.078)	(0.037)	(0.152)	(0.132)	(0.024)	(0.201)	(0.086)	(0.166)	(0.095)	(0.065)	(0.079)
Calendar year (2021 omit.)													
2022	-0.091*	0.106	0.016	0.100	-0.911**	1.791***	-0.013	1.952***	0.550**	0.776***	0.044	0.204	0.265
	(0.053)	(0.329)	(0.201)	(0.120)	(0.434)	(0.299)	(0.107)	(0.282)	(0.253)	(0.288)	(0.263)	(0.167)	(0.210)
2023	-0.143	-0.022	-0.072	0.087	-0.652	3.625***	-0.013	3.293***	0.593**	1.188***	-0.049	0.125	0.221
	(0.088)	(0.329)	(0.211)	(0.146)	(0.462)	(0.345)	(0.114)	(0.316)	(0.270)	(0.348)	(0.317)	(0.184)	(0.224)
First or second Data Fest of the year (first omit.)													
Second Data Fest of the year	0.070	0.153	0.050	-0.002	0.225	0.342*	-0.016	0.209	-0.151	0.034	-0.067	0.136	-0.000
	(0.047)	(0.141)	(0.094)	(0.031)	(0.178)	(0.182)	(0.023)	(0.168)	(0.105)	(0.146)	(0.134)	(0.098)	(0.100)
Year in college (first-year student omit.)													
Second-year student	0.052	0.627***	0.088**	0.012	0.241**	-1.834***	0.000	-2.341***	-0.543***	1.088***	1.093***	0.079**	0.044
	(0.038)	(0.057)	(0.040)	(0.033)	(0.108)	(0.069)	(0.020)	(0.086)	(0.051)	(0.094)	(0.078)	(0.035)	(0.031)
Third-year student	0.166*	0.919***	0.382*	-0.039	0.885**	-3.306***	0.149	-3.675***	-1.046***	0.857**	1.409***	0.437***	-0.099
	(0.099)	(0.190)	(0.206)	(0.141)	(0.338)	(0.242)	(0.122)	(0.227)	(0.258)	(0.398)	(0.364)	(0.153)	(0.202)
Fourth-year student	0.252	1.007**	0.703**	-0.147	1.081*	-5.170***	0.194	-5.203***	-1.198***	-0.225	1.642***	0.380	-0.097
	(0.156)	(0.462)	(0.318)	(0.246)	(0.621)	(0.547)	(0.154)	(0.483)	(0.364)	(0.585)	(0.540)	(0.282)	(0.276)
Majors (multiple possible)													
Economics major	-0.248**	-0.558**	-0.754***	0.620**	2.288*	0.037	-0.452*	-0.810***	-0.124	0.610*	0.929***	-0.703***	0.566**

	Relevance				Belonging		Growth mindset						
	Employable	**Profession**	**Framework**	**Social justice**	**Feel different**	**Group questions**	**Learn software**	**Understand data**	**Learn analysis**	**Quick learning**	**Software req. discipline**	**Data sci. or econ.**	**Belief learn**
	(0.113)	(0.220)	(0.088)	(0.311)	(1.306)	(0.160)	(0.260)	(0.219)	(0.693)	(0.306)	(0.306)	(0.132)	(0.229)
Data/math major	0.044	0.929**	0.245	0.078	0.201	-0.606***	0.050	0.022	-0.087	1.980***	1.698***	0.024	0.466***
	(0.039)	(0.364)	(0.213)	(0.056)	(0.446)	(0.221)	(0.083)	(0.246)	(0.173)	(0.176)	(0.142)	(0.124)	(0.104)
Poli. sci./public policy major	0.025	0.509**	0.550***	-0.038	-2.449**	-0.300**	0.484**	-0.094	-0.320	-0.415*	-0.708***	0.500***	-0.021
	(0.017)	(0.209)	(0.075)	(0.047)	(1.228)	(0.137)	(0.236)	(0.194)	(0.211)	(0.209)	(0.159)	(0.120)	(0.074)
Constant													
Constant	1.913***	0.763***	1.547***	1.658***	-0.349	2.730***	1.859***	2.639***	2.393***	-1.905***	-0.675***	1.488***	1.404***
	(0.068)	(0.231)	(0.137)	(0.138)	(0.473)	(0.158)	(0.095)	(0.169)	(0.300)	(0.239)	(0.219)	(0.088)	(0.118)
N (Obs.)	160	160	160	160	160	160	160	160	160	160	160	160	160
N (individuals)	76	76	76	76	76	76	76	76	76	76	76	76	76
R-sq.	.064	.103	.172	.15	.283	.2	.235	.162	.141	.154	.127	.245	.116

Notes: *p<0.1; **p<0.05; ***p<0.01. Cells are coefficients. Clustered standard errors in parentheses. Ns are smaller than main model due to missing data on year in college and major (only collected in pre-surveys). Business, other, and international majors' controls dropped because there was not within-person variation over time.

Table 7 continued. Fixed effects OLS regression models for confidence and factor outcomes, with controls

	Confidence									Factor
	Like software	Confident software	Code easy	Software understanding	Complicated	Frustrated software	Stress software	R comfort	Stata comfort	Factor
Pre vs. post (pre omit.)										
Post Data Fest	0.196*	0.567***	0.440**	0.233	-0.024	0.301*	-0.032	0.047	0.400**	0.284***
	(0.104)	(0.159)	(0.188)	(0.147)	(0.145)	(0.174)	(0.163)	(0.225)	(0.187)	(0.082)
Calendar year (2021 omit.)										
2022	-0.012	0.924**	0.972***	-0.248	-0.238	-2.140***	-0.818**	-1.537**	0.096	0.901***
	(0.373)	(0.421)	(0.312)	(0.452)	(0.195)	(0.463)	(0.406)	(0.631)	(0.608)	(0.334)
2023	-0.269	1.941***	1.312***	-0.098	-0.294	-3.790***	-1.780***	-1.603**	0.504	1.458***
	(0.413)	(0.463)	(0.352)	(0.478)	(0.235)	(0.464)	(0.452)	(0.625)	(0.628)	(0.361)
First or second Data Fest of the year (first omit.)										
Second Data Fest of the year	0.318***	0.398**	-0.120	0.468***	0.136	-0.001	-0.043	0.084	0.783***	0.215**
	(0.119)	(0.187)	(0.181)	(0.160)	(0.135)	(0.206)	(0.159)	(0.167)	(0.217)	(0.084)
Year in college (first-year student omit.)										
Second-year student	0.757***	0.983***	-0.340***	1.350***	0.056	2.650***	0.463***	1.566***	-0.908***	0.047
	(0.071)	(0.081)	(0.106)	(0.080)	(0.081)	(0.076)	(0.107)	(0.152)	(0.163)	(0.054)
Third-year student	1.047**	0.483	-0.972***	1.721***	0.387*	4.465***	1.584***	3.333***	-0.945**	-0.310
	(0.437)	(0.466)	(0.290)	(0.364)	(0.217)	(0.318)	(0.380)	(0.510)	(0.437)	(0.326)
Fourth-year student	1.654***	0.118	-1.652***	2.757***	0.559	6.307***	2.124***	5.388***	-1.127	-0.545
	(0.579)	(0.761)	(0.516)	(0.758)	(0.412)	(0.508)	(0.577)	(0.846)	(0.935)	(0.508)
Majors (multiple possible)										
Economics major	-0.374	-0.309	0.906	-1.088***	-0.713***	-1.140***	0.074	0.094	-0.090	-0.383
	(0.564)	(0.575)	(0.935)	(0.374)	(0.197)	(0.419)	(0.632)	(1.001)	(0.881)	(0.278)
Data/math major	1.250***	0.964***	0.146	0.713**	-0.193	1.091*	-1.308***	-0.558	-0.789	0.810***
	(0.223)	(0.297)	(0.306)	(0.345)	(0.214)	(0.566)	(0.389)	(0.540)	(0.515)	(0.225)
Poli. sci./public policy major	0.414***	-0.060	0.395	1.158***	0.319*	0.513**	0.181	-0.797	1.171***	0.477*
	(0.092)	(0.202)	(0.932)	(0.177)	(0.162)	(0.239)	(0.586)	(0.802)	(0.210)	(0.257)
Constant										
Constant	0.133	-1.322***	-0.429	-0.790***	1.025***	-1.862***	-0.355	0.497	2.500***	-0.822***

	Confidence									Factor
	Like software	**Confident software**	**Code easy**	**Software understanding**	**Complicated**	**Frustrated software**	**Stress software**	**R comfort**	**Stata comfort**	**Factor**
	(0.298)	(0.272)	(0.326)	(0.274)	(0.170)	(0.419)	(0.365)	(0.470)	(0.505)	(0.181)
N (Obs.)	160	160	160	160	160	160	160	160	160	160
N (individuals)	76	76	76	76	76	76	76	76	76	76
R-sq.	.188	.385	.149	.327	.0336	.14	.0726	.338	.277	.318

Notes: *p<0.1; **p<0.05; ***p<0.01. Cells are coefficients. Clustered standard errors in parentheses. Ns are smaller than main model due to missing data on year in college and major (only collected in pre-surveys). Business, other, and international majors' controls dropped because there was not within-person variation over time.

Table 8. Fixed effects OLS regression models for relevance, belonging, and growth mindset outcomes, only first calendar year of data

	Relevance				Belonging		Growth mindset						
	Employable	**Profession**	**Framework**	**Social justice**	**Feel different**	**Group questions**	**Learn software**	**Understand data**	**Learn analysis**	**Quick learning**	**Software req. discipline**	**Data sci. or econ.**	**Belief learn**
Pre vs. post (pre omit.)													
Post Data Fest	0.060*	0.149*	0.167**	0.018	0.295**	0.115	0.037	0.300**	0.139*	0.178	0.210*	0.029	0.112*
	(0.035)	(0.089)	(0.065)	(0.036)	(0.132)	(0.090)	(0.032)	(0.148)	(0.072)	(0.148)	(0.118)	(0.086)	(0.060)
First or second Data Fest of the year (first omit.)													
Second Data Fest of the year	0.003	0.034	-0.092	-0.005	0.132	0.265**	-0.011	0.152	-0.126	0.031	0.105	0.012	-0.039
	(0.037)	(0.098)	(0.077)	(0.037)	(0.139)	(0.112)	(0.024)	(0.117)	(0.082)	(0.138)	(0.147)	(0.090)	(0.071)
Constant	1.902***	1.595***	1.800***	1.925***	0.154*	1.321***	1.889***	0.549***	1.771***	-0.374***	0.897***	1.710***	1.819***
	(0.023)	(0.046)	(0.034)	(0.024)	(0.087)	(0.057)	(0.017)	(0.065)	(0.040)	(0.073)	(0.071)	(0.057)	(0.030)
N (Obs.)	182	182	182	182	182	182	182	182	182	182	182	182	182
N (individuals)	93	93	93	93	93	93	93	93	93	93	93	93	93
R-sq.	.0279	.042	.0607	.00229	.0586	.0999	.0157	.0915	.0511	.0237	.0449	.00174	.0359

Notes: *p<0.1; **p<0.05; ***p<0.01. Cells are coefficients. Clustered standard errors in parentheses. No calendar year coefficients due to including only first calendar year of data and individual fixed effects.

Table 8 continued. Fixed effects OLS regression models for confidence and factor outcomes, only first calendar year of data

	Confidence									Factor
	Like software	**Confident software**	**Code easy**	**Software understanding**	**Complicated**	**Frustrated software**	**Stress software**	**R comfort**	**Stata comfort**	**Factor**
Pre vs. post (pre omit.)										
Post Data Fest	0.370***	0.589***	0.452***	0.512***	0.024	0.063	-0.063	0.060	0.382***	0.423***
	(0.100)	(0.146)	(0.145)	(0.119)	(0.128)	(0.149)	(0.148)	(0.135)	(0.128)	(0.083)
First or second Data Fest of the year (first omit.)										
Second Data Fest of the year	0.457***	0.813***	0.249	0.725***	0.135	0.255*	-0.202	0.177	1.033***	0.278***
	(0.099)	(0.141)	(0.158)	(0.131)	(0.110)	(0.130)	(0.157)	(0.118)	(0.149)	(0.081)
Constant	0.986***	-0.287***	-0.267***	0.473***	0.865***	0.063	0.082	1.910***	1.862***	-0.377***
	(0.063)	(0.078)	(0.087)	(0.077)	(0.060)	(0.071)	(0.098)	(0.081)	(0.093)	(0.052)
N (Obs.)	182	182	182	182	182	182	182	182	182	182
N (individuals)	93	93	93	93	93	93	93	93	93	93
R-sq.	.297	.429	.142	.406	.0181	.0487	.0235	.028	.445	.314

Notes: *p<0.1; **p<0.05; ***p<0.01. Cells are coefficients. Clustered standard errors in parentheses. No calendar year coefficients due to including only first calendar year of data and individual fixed effects.

Table 9. Means and standard deviations of outcome variables

	Relevance				Belonging		Growth mindset						
	Employable	Profession	Framework	Social justice	Feel different	Group questions	Learn software	Understand data	Learn analysis	Quick learning	Software req. discipline	Data sci. or econ.	Belief learn
Mean	1.93	1.7	1.84	1.93	0.28	1.52	1.92	0.77	1.76	-0.24	1.04	1.77	1.85
Standard deviation	0.29	0.61	0.42	0.26	1.29	0.8	0.29	1.02	0.54	1.09	0.91	0.58	0.42

	Confidence									Factor
	Like software	Confident software	Code easy	Software understanding	Complicated	Frustrated software	Stress software	R comfort	Stata comfort	Factor
Mean	1.39	0.46	0	1.11	0.95	0.16	-0.09	2.29	2.48	0
Standard deviation	0.83	1.28	1.1	1.11	1.09	1.22	1.17	1.43	1.33	0.96

Surveys

Unless noted otherwise, questions were included in both the pre- and post-surveys

1. Email [text box]
2. What's your name? [text box]
3. Who is your professor (can check multiple) [List of faculty names with participating classes, along with other specify]

Data analytics and statistical software

4. How comfortable you are running the following statistical software program: R
   Not 1 2 3 4 5 Extremely comfortable

5. How comfortable you are running the following statistical software program: Stata
   Not 1 2 3 4 5 Extremely comfortable

6. Rate the following statements

	Agree	Somewhat agree	Neutral	Somewhat disagree	Disagree
I like using statistical software to analyze data					
I feel confident using statistical software to analyze data					
I get frustrated using statistical software to analyzing data					
I am under stress while using statistical software to analyze data					

7. Rate the following statements

	Agree	Somewhat agree	Neutral	Somewhat disagree	Disagree
Data analysis skills will make me more employable					

I will have some application for data analysis and statistical software skills in my profession					
Data analysis gives me a useful framework for thinking about issues that are important to my life					
Data analysis is relevant for social justice work					

8. Rate the following statements

	Agree	Somewhat agree	Neutral	Somewhat disagree	Disagree
I can learn how to use statistical software					
I find it easy to understand data analysis because of how I think					
I can learn how to analyze large data sets					
I understand the basics of how to analyze large data sets using statistical software					

9. Rate the following statements

	Agree	Somewhat agree	Neutral	Somewhat disagree	Disagree
Data analysis is complicated					
Statistical software code is easy to understand					
Data analysis skills are quickly learned by most people					
Learning statistical software requires a great deal of discipline					

10. Rate the following statements

	Agree	Somewhat agree	Neutral	Somewhat disagree	Disagree
I feel comfortable asking questions when I am in a group working on a data project					
I can learn the skills I need to become a data scientist or economist					
I feel different than the typical student in data science or economics					
I believe I can learn how to do data analysis					

11. What is something you liked about Data Fest? [post-survey only, paragraph response]
12. What is something you would like to see changed for the next Data Fest? [post-survey only, paragraph response]
13. What year are you in college? [pre-survey only]
    a. First year
    b. Sophomore
    c. Junior
    d. Senior
    e. Graduate student
    f. Other
14. What is your major? [text box, pre-survey only]
15. What is your gender identity?
    a. Female
    b. Male
    c. Non-binary
    d. Prefer not to say
    e. Other (specify)
16. What is your racial identity? (you may select more than one checkbox) [pre-survey only]
    a. White
    b. Black or African American
    c. American Indian/Alaskan Native

    d. Asian
    e. Hawaiian Native & Pacific Islander
    f. Prefer not to say
    g. Other (specify)
17. What is your ethnic identity? [pre-survey only]
    a. Hispanic/Latinx
    b. Not Hispanic/Latinx
    c. Prefer not to say
    d. Other
18. Do you have any dietary restrictions or allergies we should be aware of for the dinner/snacks?