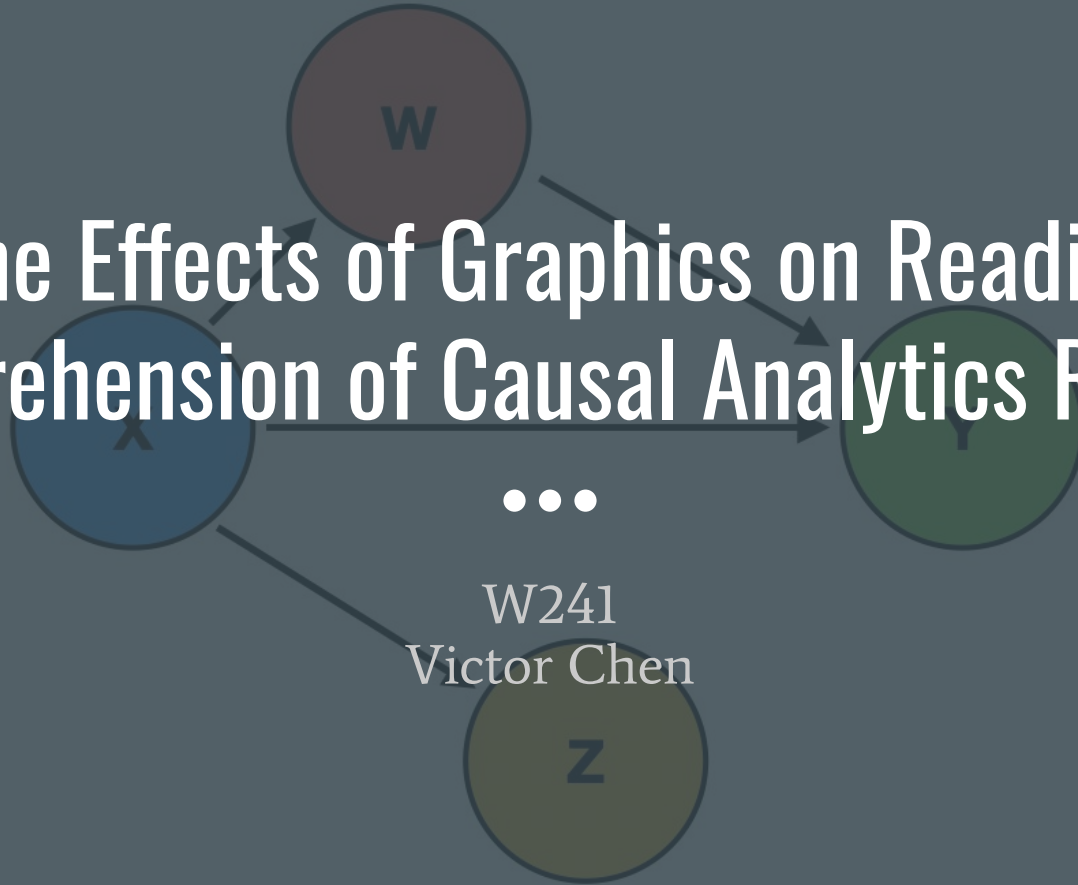


The Effects of Graphics on Reading Comprehension of Causal Analytics Reports



Background

Pressed-for-time,
story-focused
executives



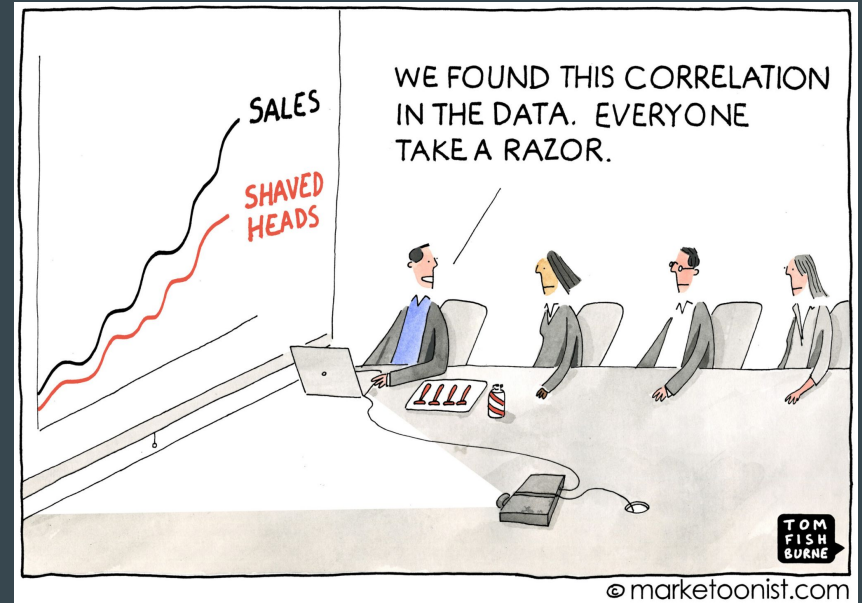
Data-savvy
analysts

Background

Data analytics



Prescriptive analytics



Hypotheses

A graphic/visual representation of causal knowledge will lead to more effective translation of data insights into actionable prescriptions

- More accurate reading comprehension of complex analytics reports
- Higher confidence in the results
- Greater actionability of recommendations

Example

Data analytics



Prescriptive analytics

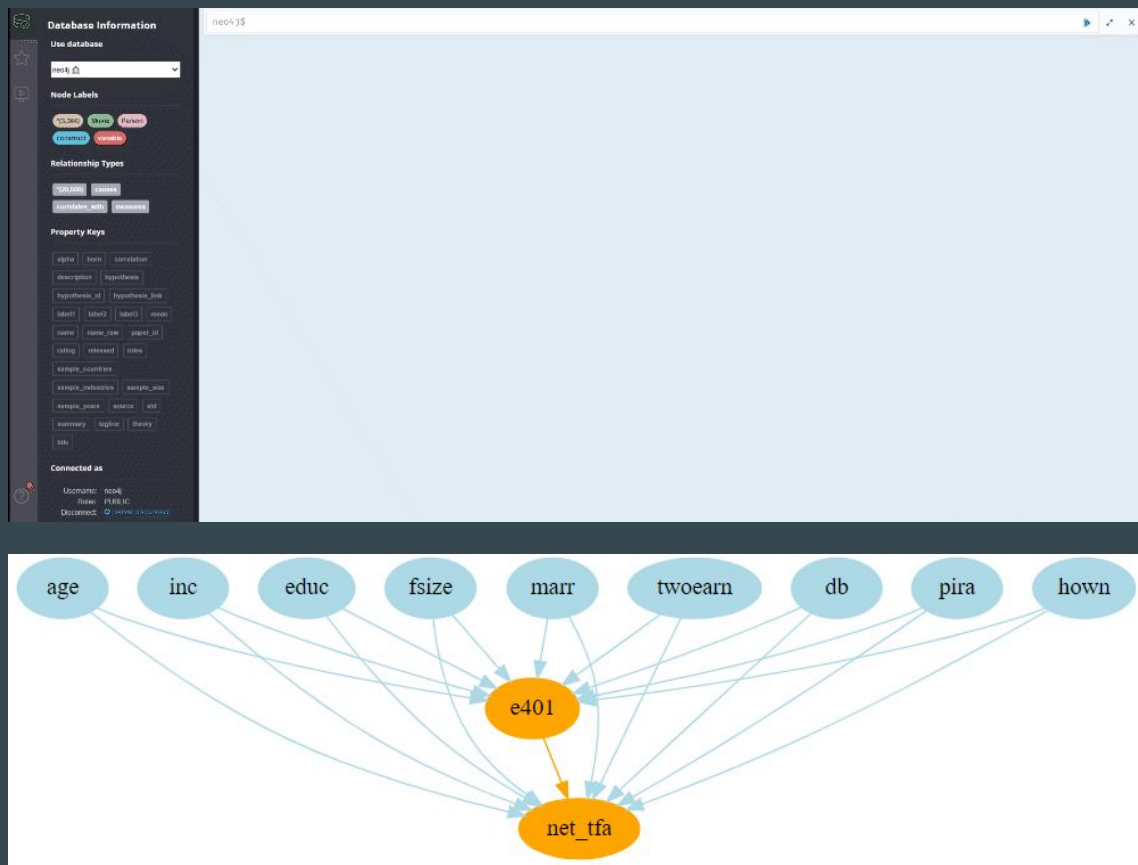


Example

Data analytics



Prescriptive analytics

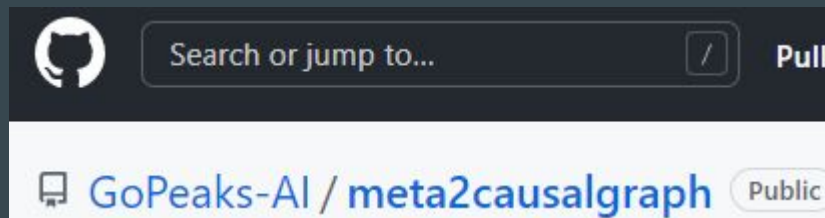


Example

Data analytics

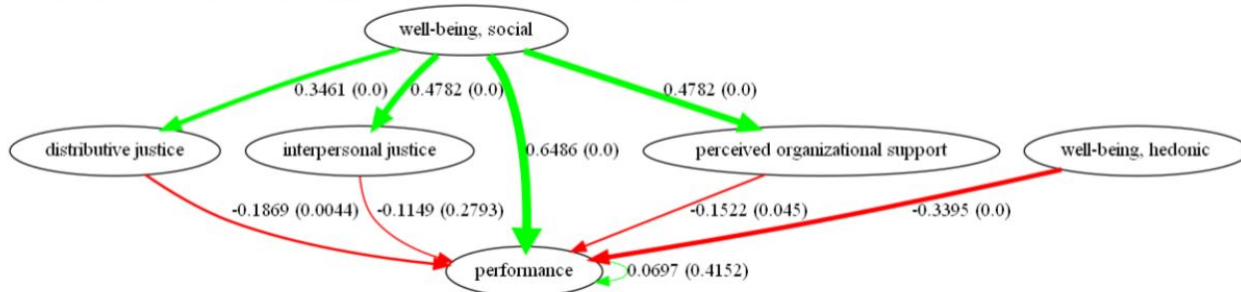


Prescriptive analytics



```
causal_graph(n1_level='label1',  
             n2_level='label1',  
             n1_searchterm='well-being',  
             n2_searchterm='performance',  
             search_method='contain',  
             draw_graph=True  
)
```

Query has returned 5 findings on the direct causal link between 'well-being' and 'performance'.
Query has returned 3 findings on mediation between 'well-being' and 'performance'.
Query has returned 16 findings on the confounder between 'well-being' and 'performance'.
Query has returned 0 findings on the collider between 'well-being' and 'performance'.



Notes:

- [1] link label shows the mean effect size (p-value in the parenthesis);
- [2] link arrow shows the direction of causality;
- [3] link color suggests the sign of the effect size: green for positive, and red for negative;
- [4] link thickness suggests the magnitude of the effect size.

Hypothesis

A graphic/visual representation of causal knowledge will lead to a more accurate reading comprehension of analytics reports.



Control



Does 401(k) Eligibility affect Financial Wealth?

A recent study on the effects of 401(k) eligibility on financial wealth was published in American Economic Review in 2017. This study examined a sample of households from wave 4 of the 1990 U.S. Survey of Income and Program Participation (SIPP), where the observations are limited to households in which the reference person is 25-64 years old, at least one person is employed, and no one is self-employed. The sample consists of 9915 households, and all dollar amounts are in 1991 dollars. The sample shows a \$19,559 mean difference between households ineligible for 401(k) and those eligible for 401(k). That is, on average, households eligible for 401(k) had almost \$20,000 more in net financial assets than those ineligible for 401(k). But this simple mean difference has not controlled for covariates like age, income, family size, marriage status, two-earner status, defined benefit (DB) pension status, IRA participation status, and homeownership status. Because these covariates may directly affect a household's 401(k) eligibility, the study further used machine learning techniques to reduce the confounding biases from the data. After debiasing the data, there is an approximately \$8000 - \$9000 effect of the debiased measure of 401(k) eligibility on the debiased measure of a household's net financial assets. Additionally, a linear regression estimated a coefficient of \$5896 of 401(k) eligibility on a household's net financial assets, after controlling for all the covariates, suggesting that everything being equal, a household eligible for 401(k) on average tends to earn almost \$6000 more in net financial assets. These findings suggest a clear positive relationship between 401(k) eligibility and a household's financial wealth. But the exact estimates vary depending on the research methods.

Based on this study, how many additional 1991 dollars in a household's net financial assets would be caused by 401(k) eligibility, everything else being equal?

References

- [1] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/Neyman machine learning of treatment effects. *American Economic Review*, 107(5), 261-65.
- [2] DoubleML (2021). Python: Impact of 401(k) on financial wealth. Accessible at https://docs.doubleml.org/stable/examples/py_double_ml_pension.html.

\$19,559

\$8000-\$9000

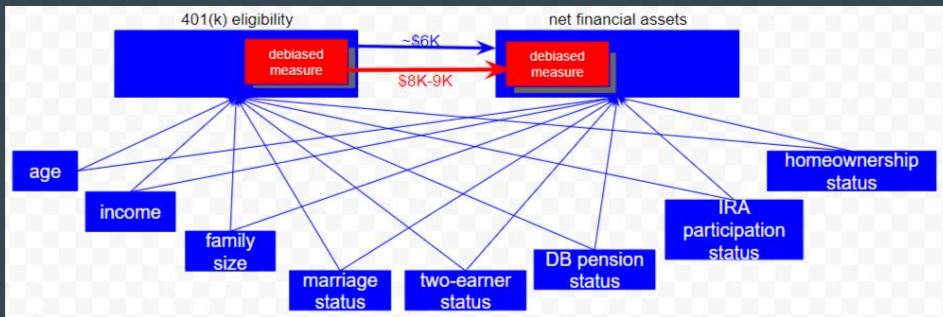
Almost \$6000

Treatment



Does 401(k) Eligibility affect Financial Wealth?

A recent study on the effects of 401(k) eligibility on financial wealth was published in *American Economic Review* in 2017. This study examined a sample of households from wave 4 of the 1990 U.S. Survey of Income and Program Participation (SIPP), where the observations are limited to households in which the reference person is 25-64 years old, at least one person is employed, and no one is self-employed. The sample consists of 9915 households, and all dollar amounts are in 1991 dollars. The sample shows a \$19,559 mean difference between households ineligible for 401(k) and those eligible for 401(k). That is, on average, households eligible for 401(k) had almost \$20,000 more in net financial assets than those ineligible for 401(k). But this simple mean difference has not controlled for covariates like age, income, family size, marriage status, two-earner status, defined benefit (DB) pension status, IRA participation status, and homeownership status. Because these covariates may directly affect a household's 401(k) eligibility, the study further used machine learning techniques to reduce the confounding biases from the data. After debiasing the data, there is an approximately \$8000 - \$9000 effect of the debiased measure of 401(k) eligibility on the debiased measure of a household's net financial assets. Additionally, a linear regression estimated a coefficient of \$5896 of 401(k) eligibility on a household's net financial assets, after controlling for all the covariates, suggesting that everything being equal, a household eligible for 401(k) on average tends to earn almost \$6000 more in net financial assets. These findings suggest a clear positive relationship between 401(k) eligibility and a household's financial wealth. But the exact estimates vary depending on the research methods.



Based on this study, how many additional 1991 dollars in a household's net financial assets would be caused by 401(k) eligibility, everything else being equal?

References

- [1] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/Neyman machine learning of treatment effects. *American Economic Review*, 107(5), 261-65.
- [2] DoubleML (2021). Python: Impact of 401(k) on financial wealth. Accessible at https://docs.doubleml.org/stable/examples/py_double_ml_pension.html.

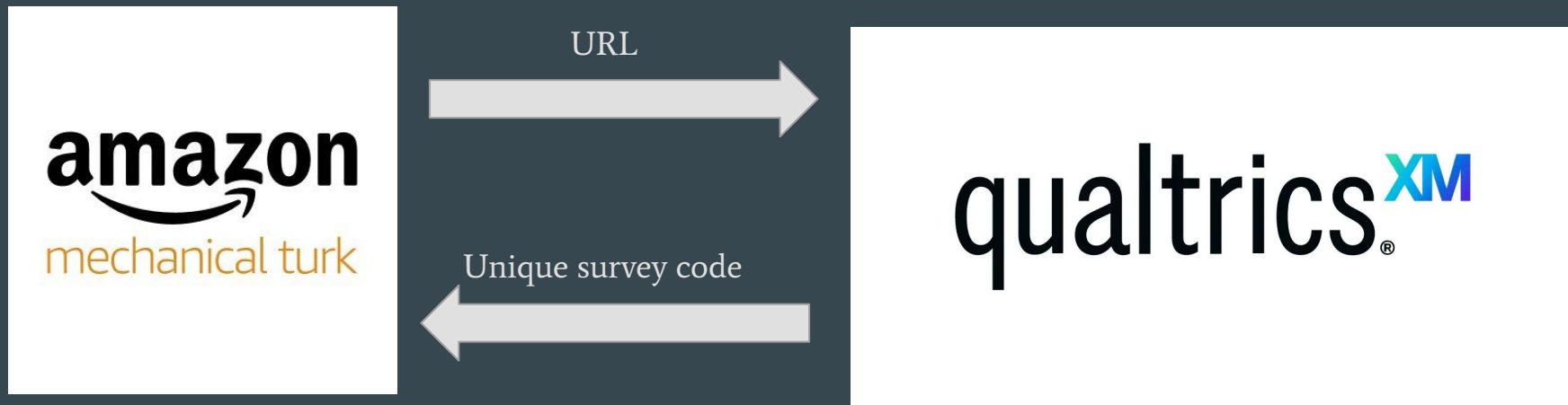
\$19,559

\$8000-\$9000

Almost \$6000

Measurement Unit

Unique individual Amazon M-Turks workers whose first language is EN, and are living in the United States



Randomization

XM Study of 401(k) on financial wealth

Survey Workflows Distributions Data & Analysis Results Reports

Survey flow Published

Show Block: Background (6 Questions) [Add Below](#) [Move](#) [Duplicate](#) [Delete](#)

Randomizer
Randomly present of the following elements ☒ Evenly Present Elements [Edit Count](#)
[Add Below](#) [Move](#) [Duplicate](#) [Collapse](#) [Delete](#)

Show Block: Reading Comprehension 0 (1 Question) [Add Below](#) [Move](#) [Duplicate](#) [Delete](#)

Show Block: Reading Comprehension 0 (1 Question) [Add Below](#) [Move](#) [Duplicate](#) [Delete](#)

[+ Add a New Element Here](#)

Show Block: References (1 Question) [Add Below](#) [Move](#) [Duplicate](#) [Delete](#)

[+ Add a New Element Here](#)

Workflow and Causation



Outcome measures

Main Y:

- Correct answer (1 correct, 0 wrong) on the expected effects of enrollment in 401k on net financial assets in 1991 dollars, on average

Secondary Ys:

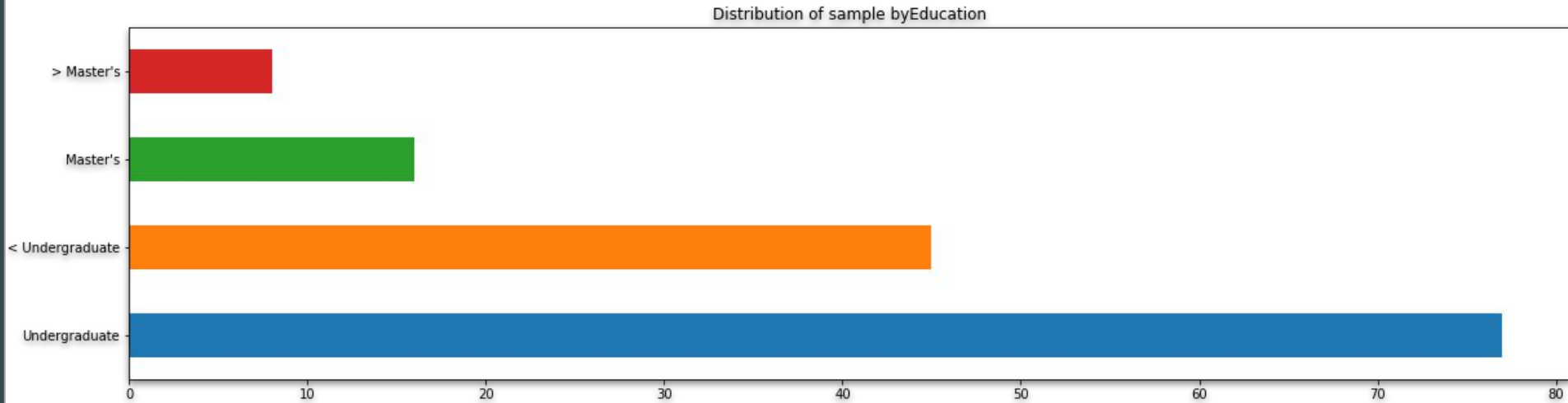
- Reading speed (in seconds)
- Whether to check external sources (1 yes, 0 no)

Control group
N = 75

Treatment group
N = 72



Data Analysis - Summary



Data Analysis - Summary

Group	Education	
Control	Undergraduate	37
	< Undergraduate	29
	Master's	7
	> Master's	1
Treatment	Undergraduate	40
	< Undergraduate	16
	Master's	9
	> Master's	7

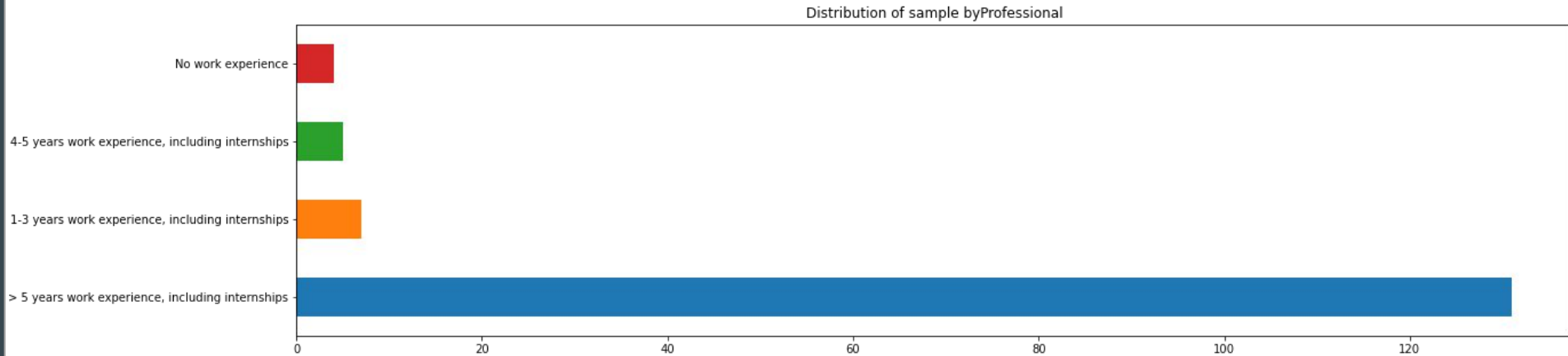
Name: Education, dtype: int64

Difference in education score: 0.11555555555555554

The variances are 0.09528888888888885 and 0.17283950617283966 for the control and treatment groups, respectively.

The t-test of the mean education difference has a p-value of 0.058709107659612805.

Data Analysis - Summary (cont'd)



Data Analysis - Summary (cont'd)

Group	Professional	
Control	> 5 years work experience, including internships	67
	No work experience	4
	1-3 years work experience, including internships	3
	4-5 years work experience, including internships	1
Treatment	> 5 years work experience, including internships	64
	1-3 years work experience, including internships	4
	4-5 years work experience, including internships	4

Name: Professional, dtype: int64

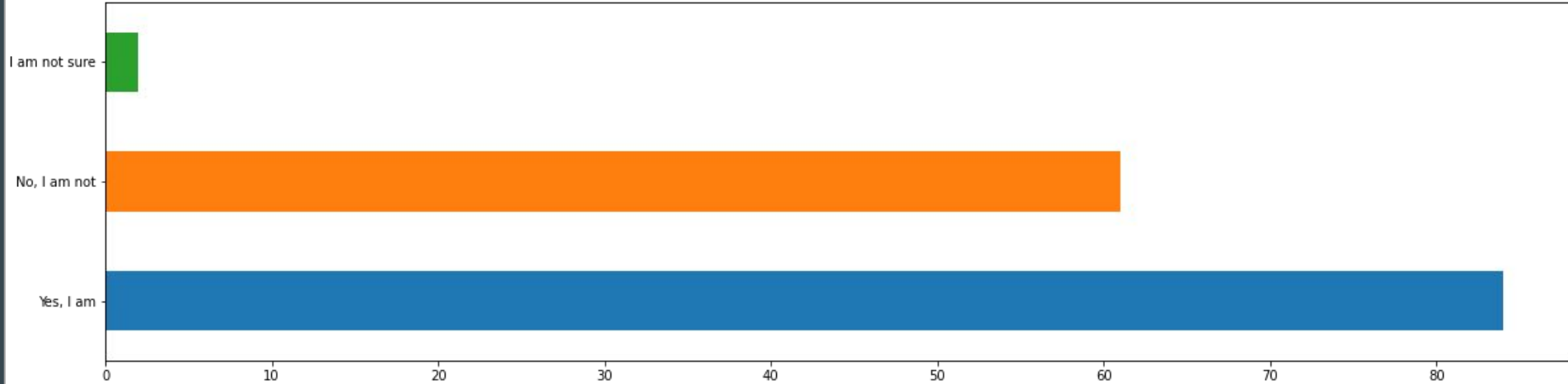
Difference in experience score: 0.086666666666666689

The variances are 0.58915555555555563 and 0.25000000000000001 for the control and treatment groups, respectively.

The t-test of the mean experience score difference has a p-value of 0.4238107594329512.

Data Analysis - Summary (cont'd)

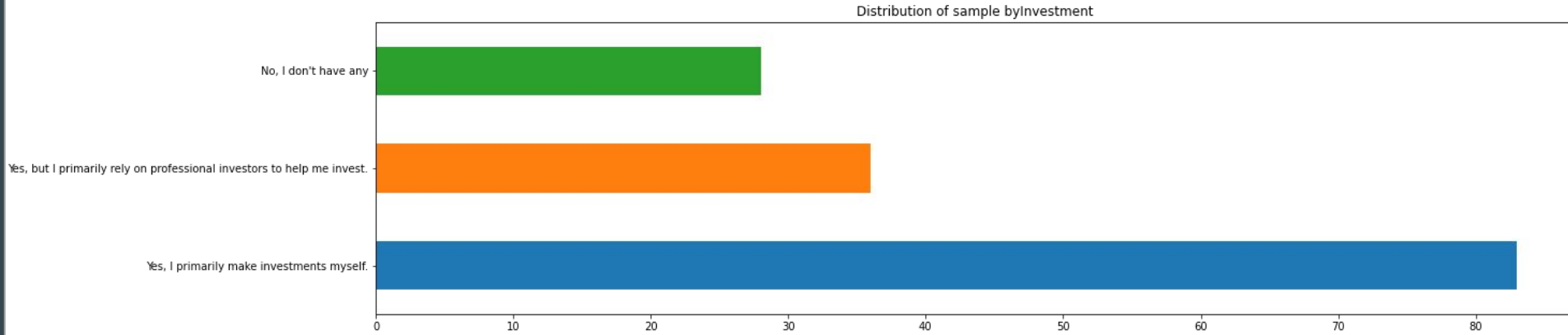
Distribution of sample by ln401k



Data Analysis - Summary (cont'd)

```
Group      In401k
Control    Yes, I am      39
           No, I am not   36
Treatment  Yes, I am      45
           No, I am not   25
           I am not sure   2
Name: In401k, dtype: int64
Difference in 401k status: 0.10499999999999998
The variances are 0.2496 and 0.234375 for the control and treatment groups, respectively.
The t-test of the mean 401k status difference has a p-value of 0.20103306368696297.
```

Data Analysis - Summary (cont'd)



Data Analysis - Summary (cont'd)

Group	Investment	
Control	Yes, I primarily make investments myself.	39
	No, I don't have any	19
	Yes, but I primarily rely on professional investors to help me invest.	17
Treatment	Yes, I primarily make investments myself.	44
	Yes, but I primarily rely on professional investors to help me invest.	19
	No, I don't have any	9

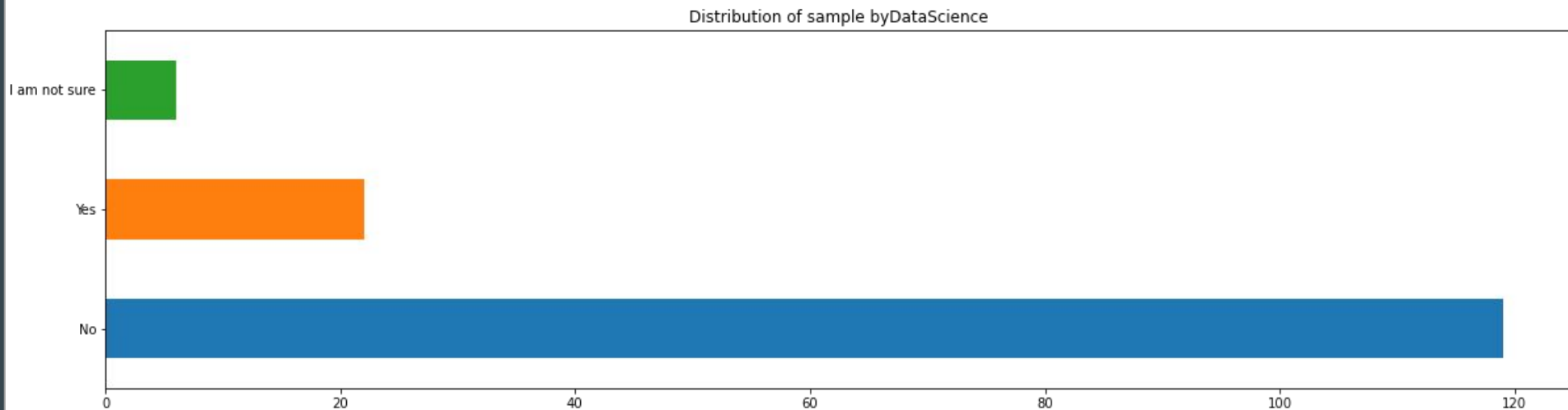
Name: Investment, dtype: int64

Difference in investment background: 0.10972222222222228

The variances are 0.175555555555555546 and 0.12495177469135793 for the control and treatment groups, respectively.

The t-test of the mean difference in investment background has a p-value of 0.09108924246338256.

Data Analysis - Summary (cont'd)



Data Analysis - Summary (cont'd)

```
Group      DataScience
Control    No          62
           Yes          10
           I am not sure 3
Treatment  No          57
           Yes          12
           I am not sure 3
```

Name: DataScience, dtype: int64

Difference in Data science background: 0.033333333333333326

The variances are 0.11555555555555558 and 0.13888888888888895 for the control and treatment groups, respectively.

The t-test of the mean difference in Data science background has a p-value of 0.5742522857891523.

Data Analysis - Correctness of the answer

Control group answers:

Almost \$6000 41

\$19,559 20

\$8000-\$9000 14

Name: Control, dtype: int64

Treatment group answers:

Almost \$6000 32

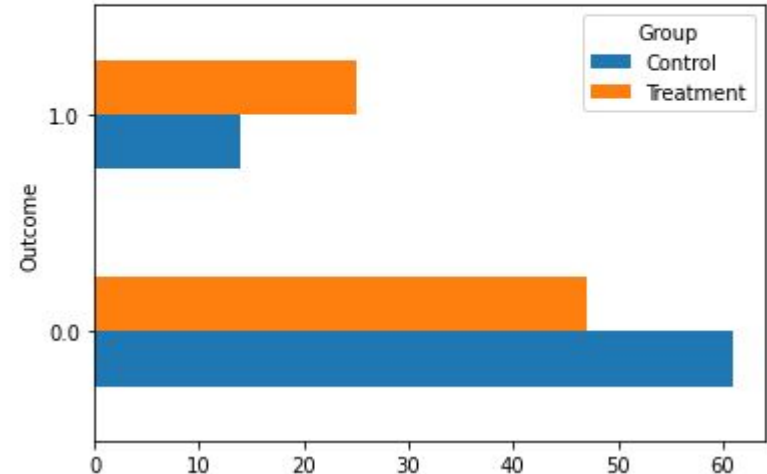
\$8000-\$9000 25

\$19,559 15

Name: Treatment, dtype: int64

Distribution of outcome by group

<AxesSubplot:ylabel='Outcome'>



Data Analysis - Correctness of the answer (Cont'd)

```
print("-" * 50)
print("Mean outcome by group")
print("-" * 50)
print(df.groupby('Group')['Outcome'].mean())
```

```
-----
Mean outcome by group
-----
Group
Control    0.1867
Treatment   0.3472
Name: Outcome, dtype: float64
```

```
Y1 = df[df['Group']=='Treatment']['Outcome']
Y0 = df[df['Group']=='Control']['Outcome']
print("ATE:", Y1.mean() - Y0.mean())
print("The variances are {} and {} for the control and treatment groups, respectively.".format(np.var(Y0), np.var(Y1)))
print("The t-test of the mean outcome difference has a p-value of {}".format(stats.ttest_ind(Y1, Y0, equal_var=True)[1]))
```

ATE: 0.16055555555555553

The variances are 0.1518222222222222 and 0.22665895061728392 for the control and treatment groups, respectively.
The t-test of the mean outcome difference has a p-value of 0.027544900175178758.

Data Analysis - Correctness of the answer (Cont'd)

Logit Regression Results

Dep. Variable:	Outcome	No. Observations:	147
Model:	Logit	Df Residuals:	140
Method:	MLE	Df Model:	6
Date:	Tue, 12 Apr 2022	Pseudo R-squ.:	0.04912
Time:	23:44:23	Log-Likelihood:	-80.867
converged:	True	LL-Null:	-85.045
Covariance Type:	HC3	LLR p-value:	0.2133

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-1.4978	0.766	-1.955	0.051	-2.999	0.004
C(Group)[T.Treatment]	0.8923	0.400	2.231	0.026	0.109	1.676
education_score	-0.3790	0.421	-0.901	0.367	-1.203	0.445
experience_score	0.0663	0.630	0.105	0.916	-1.169	1.302
e401k_status	0.2912	0.411	0.708	0.479	-0.515	1.098
investment_background	-0.1354	0.519	-0.261	0.794	-1.152	0.881
data_science_background	0.6818	0.522	1.307	0.191	-0.341	1.704

Data Analysis - Subsample test 1: Education

0 for < UG

1 for UG +

```
1    101
0     46
Name: education_score, dtype: int64
```

education score: 0

ATE: 0.14166666666666666

The variances are 0.17888888888888888 and 0.234375 for the control and treatment groups, respectively.

The t-test of the mean outcome difference has a p-value of 0.3202249474781813.

education score: 1

ATE: 0.18373015873015874

The variances are 0.1313580246913581 and 0.2241709183673471 for the control and treatment groups, respectively.

The t-test of the mean outcome difference has a p-value of 0.03607277577286542.

Data Analysis - Subsample test 2: Work experience

0 for ≤ 5 years work experience

1 for 5+ years work experience

```
1    131
0     16
Name: experience_score, dtype: int64
```

Work experience score: 1

ATE: 0.19519589552238806

The variances are 0.1372243261305412 and 0.230224609375 for the control and treatment groups, respectively.

The t-test of the mean outcome difference has a p-value of 0.010613096924951873.

Work experience score: 0

ATE: -0.125

The variances are 0.234375 and 0.1875 for the control and treatment groups, respectively.

The t-test of the mean outcome difference has a p-value of 0.6185599512335765.

Data Analysis - Subsample test 3: 401k status

Yes, I'm currently enrolled in 401k

No, I'm not

```
Yes, I am      84
No, I am not   63
Name: In401k, dtype: int64
```

Enrolled in 401k: No, I am not

ATE: 0.12962962962962962

The variances are 0.13888888888888887 and 0.20850480109739364 for the control and treatment groups, respectively.

The t-test of the mean outcome difference has a p-value of 0.22725246004317023.

Enrolled in 401k: Yes, I am

ATE: 0.17264957264957265

The variances are 0.1630506245890861 and 0.23506172839506173 for the control and treatment groups, respectively.

The t-test of the mean outcome difference has a p-value of 0.08624412598748671.

Data Analysis - Subsample test 4: Investment background

0 if not involved in personal investments

1 if involved in personal investments

```
1    83
```

```
0    64
```

```
Name: investment_background, dtype: int64
```

```
investment background: 1
```

```
ATE: 0.2097902097902098
```

```
The variances are 0.13017751479289943 and 0.23140495867768582 for the control and treatment groups, respectively.
```

```
The t-test of the mean outcome difference has a p-value of 0.030818486949751207.
```

```
investment background: 0
```

```
ATE: 0.09920634920634924
```

```
The variances are 0.17283950617283966 and 0.21811224489795905 for the control and treatment groups, respectively.
```

```
The t-test of the mean outcome difference has a p-value of 0.38070771208171583.
```


Data Analysis - Subsample test 5: Data science training

Yes, I did

No, I did not

```
No      125
Yes      22
Name: DataScience, dtype: int64
```

Data science: No

ATE: 0.1474358974358974

The variances are 0.14059171597633133 and 0.2163888888888889 for the control and treatment groups, respectively.

The t-test of the mean outcome difference has a p-value of 0.05443764110466267.

Data science: Yes

ATE: 0.2

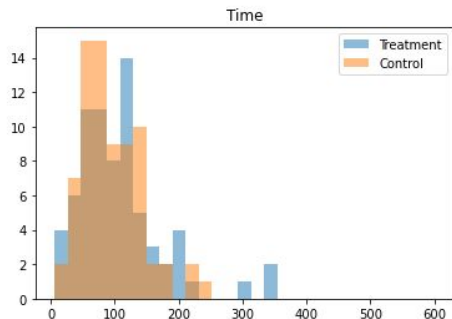
The variances are 0.21000000000000002 and 0.25 for the control and treatment groups, respectively.

The t-test of the mean outcome difference has a p-value of 0.3659958339806202.

Data Analysis - Other findings

Reading speed (time taken in seconds to complete the survey)

Histogram of Time (seconds)



```
Y1 = df[df['Group']=='Treatment']['Time']
Y0 = df[df['Group']=='Control']['Time']

print("ATE:", Y1.mean() - Y0.mean())
print("The variances are {} and {} for the control and treatment groups, respectively.".format(np.var(Y0), np.var(Y1)))
print("The t-test of the mean time difference has a p-value of {}".format(stats.ttest_ind(Y1, Y0, equal_var=True)[1]))
```

ATE: -0.3072222222222223

The variances are 16839.43395555556 and 4579.161844135802 for the control and treatment groups, respectively.

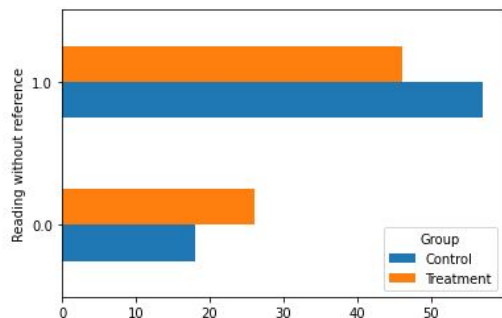
The t-test of the mean time difference has a p-value of 0.9858491683781828.

Data Analysis - Other findings

Checking external sources (1 yes, 0 no)

Distribution of Reading without reference by group

<AxesSubplot:ylabel='Reading without reference'>



```
Y1 = df[df['Group']=='Treatment']['Reading without reference']
Y0 = df[df['Group']=='Control']['Reading without reference']

print("ATE:", Y1.mean() - Y0.mean())
print("The variances are {} and {} for the control and treatment groups, respectively.".format(np.var(Y0), np.var(Y1)))
print("The t-test of the mean difference in Reading without reference has a p-value of {}".format(stats.ttest_ind(Y1, Y0, eq
```

ATE: -0.1211111111111117

The variances are 0.1824000000000003 and 0.23070987654321 for the control and treatment groups, respectively.

The t-test of the mean difference in Reading without reference has a p-value of 0.11044704614405117.

Conclusion and Discussion

- Causal graph/visualization has
 - Increased the accuracy of reading comprehension
 - Without changes in the reading speed or whether or not to seek external sources
- Subsample tests are generally consistent, with some subsamples too small to have enough statistical power (e.g., p value).
- Results remain consistent after controlling for covariates in a regression model

Conclusion and Discussion

- If I had more budget and time, what would I have done differently?
 - Blocking and randomization within each block by key predictors: data science training, etc.
 - More examples in multiple disciplines and fields, from daily topics like healthcare to specialized topics such as neuroscience;
 - More varieties of the underlying analytics reports, e.g., DML to synthetic control, instrumental variables, etc.
 - Effects of causal graphs on the entire workflow between data science and decision-making

Q&As

Project website:

https://github.com/victorchenberkeley/w241_finalproject