

DataLit Week 3 Homework Assignment

Question

The “wage gap” is widely reported in the news. Do attending certain schools (Ivy League, Public/Private), produce different wage gap or is the gap education-independent?

Data Sources

A data set was synthesized from a School-Men-Women-Gap dataset from a Plotly and publicly available attributes about the schools such as City, State, Region and Endowment.

Application

“...you are free to use any notebook or application that can give interactive visualization.”

Due to time constraints, PowerBI was used as that was the most familiar tool for creating interactive visualizations.

Data Assumptions

This is a small data set

It is summary data, not broken out by industry

Data Preparation

Columns

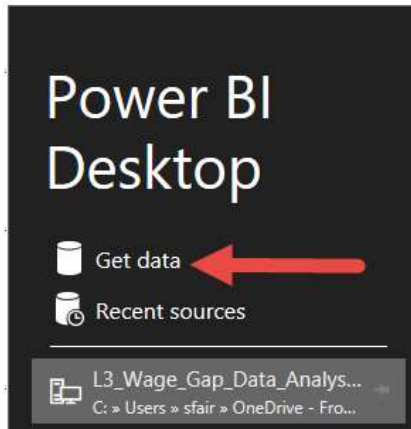
- School
- Women
- Men
- Gap
- Type
- Region
- State
- City
- Academic_staff
- Students
- Endowment(Billion)
- Endowment
- Campus
- Mascot

Dimensionality Reduction

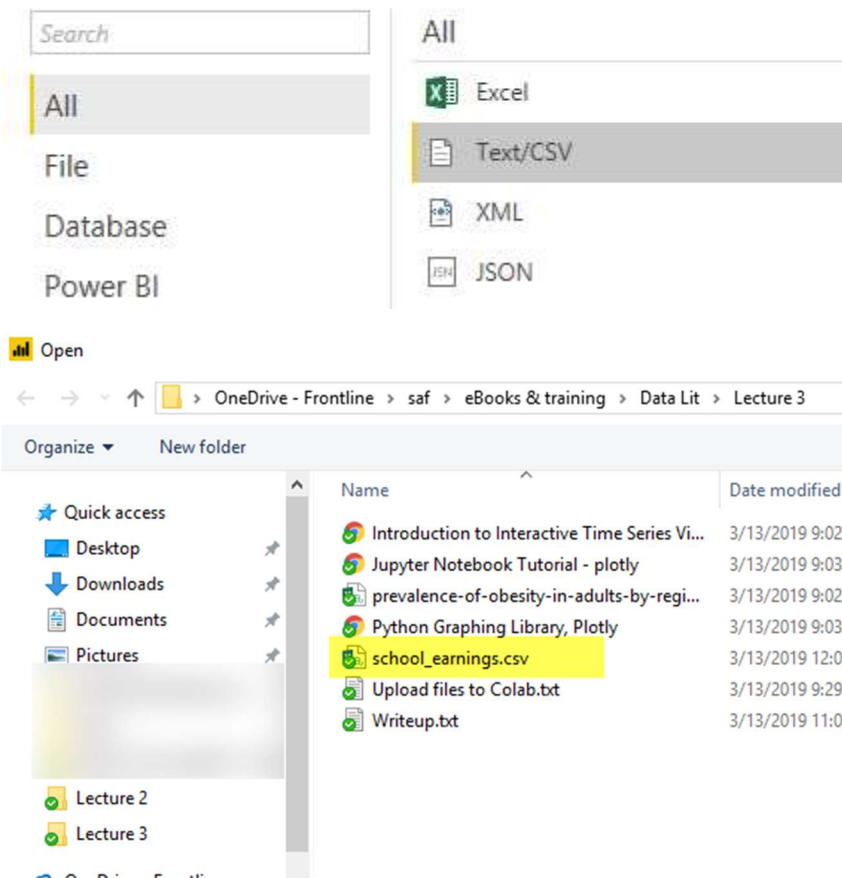
Extraneous data, like Mascot, was removed by not selecting it in the model.

Visualization Steps

Import school_earnings.csv file into PowerBI



Get Data



school_earnings.csv

File Origin: 1252: Western European (Windows) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

School	Women	Men	Gap	Type	Region	State	City	Academic_staff	Students	Endowment(Billion)	Endowm
MIT	94	152	58	Private	Northeast	MA	Amherst	1074	11574	16.4	15-20
Stanford	96	151	55	Private	West	CA	Stanford	2219	16430	24.8	25+
Harvard	112	165	53	Private	Northeast	MA	Cambridge	4671	22000	39.2	25+
U.Penn	92	141	49	Private	Northeast	PA	Philadelphia	4638	21599	13.8	10-15
Princeton	90	137	47	Private	Northeast	NJ	Princeton	1238	8273	25.9	25+
Chicago	78	118	40	Private	Midwest	IL	Chicago	2860	16446	8.2	5-10
Georgetown	94	131	37	Private	South	DC	Washington	2173	17849	1.661	0-5
Tufts	76	112	36	Private	Northeast	MA	Somerville	1706	11489	1.85	0-5
Yale	79	114	35	Private	Northeast	CT	New Haven	4410	12312	29.4	25+
Columbia	86	119	33	Private	Northeast	NY	New York	3999	27942	10.9	10-15
Duke	93	124	31	Private	South	NC	Durham	3774	15892	8.5	5-10
Dartmouth	84	114	30	Private	Northeast	NH	Hanover	750	6409	5.5	5-10
NYU	67	94	27	Private	Northeast	NY	New York	2242	51848	4.1	0-5
Notre Dame	73	100	27	Private	Midwest	IN	Notre Dame	1309	12292	13.1	10-15
Cornell	80	107	27	Private	Northeast	NY	Ithaca	2908	23600	7.2	5-10
Michigan	62	84	22	Public	Midwest	MI	Ann Arbor	6771	46002	11.9	10-15
Brown	72	92	20	Private	Northeast	RI	Providence	758	9731	3.8	0-5
Berkeley	71	88	17	Public	West	CA	Berkeley	null	42519	4.6	0-5
Emory	68	82	14	Private	South	GA	Atlanta	null	15451	7.31	5-10
UCLA	64	78	14	Public	West	CA	Westwood	4016	45428	5	5-10

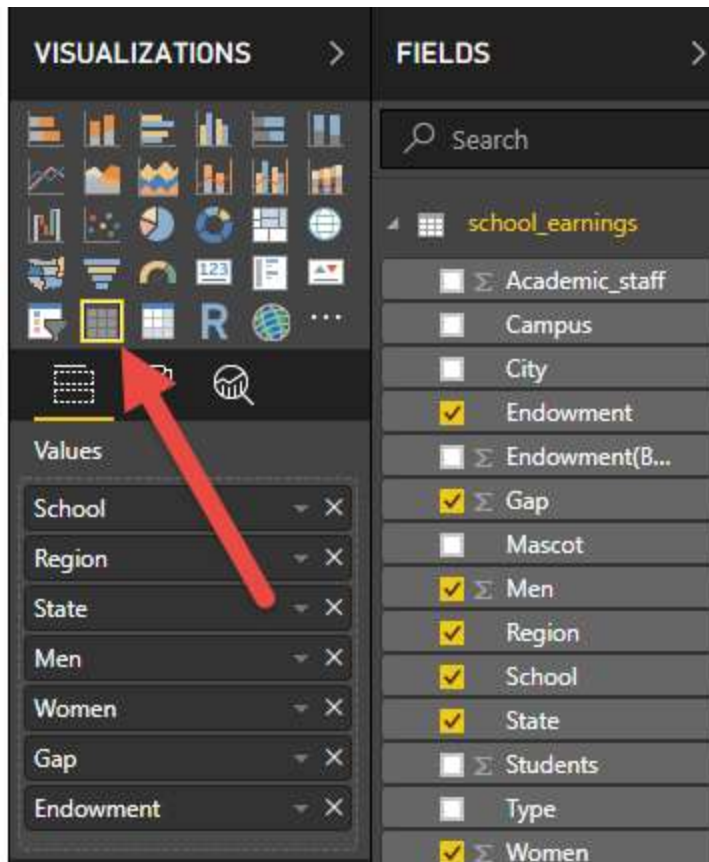
Load Edit Cancel

The school_earnings table now has the data columns added to a model

school_earnings

<input type="checkbox"/>	Academic_staff
<input type="checkbox"/>	Campus
<input type="checkbox"/>	City
<input checked="" type="checkbox"/>	Endowment
<input type="checkbox"/>	Endowment(B...
<input checked="" type="checkbox"/>	Gap
<input type="checkbox"/>	Mascot
<input checked="" type="checkbox"/>	Men
<input checked="" type="checkbox"/>	Region
<input checked="" type="checkbox"/>	School
<input checked="" type="checkbox"/>	State
<input type="checkbox"/>	Students
<input type="checkbox"/>	Type
<input checked="" type="checkbox"/>	Women

Select the Table visualization and select School, Region, State, Men, Women, Gap, Endowment



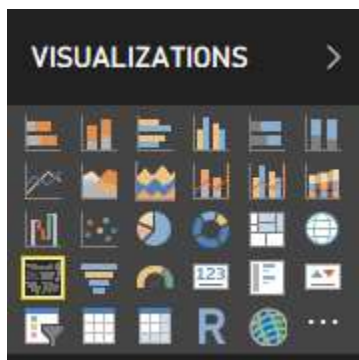
Add a Clustered Column chart for Gap by School (select School, Gap)



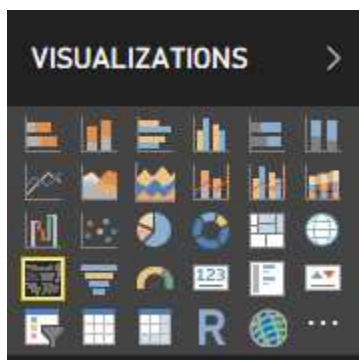
Add a Funnel Chart for Average of Gap by Endowment (select Gap, Endowment), change Gap to Average



Add a Filled Map chart (select State, Gap)

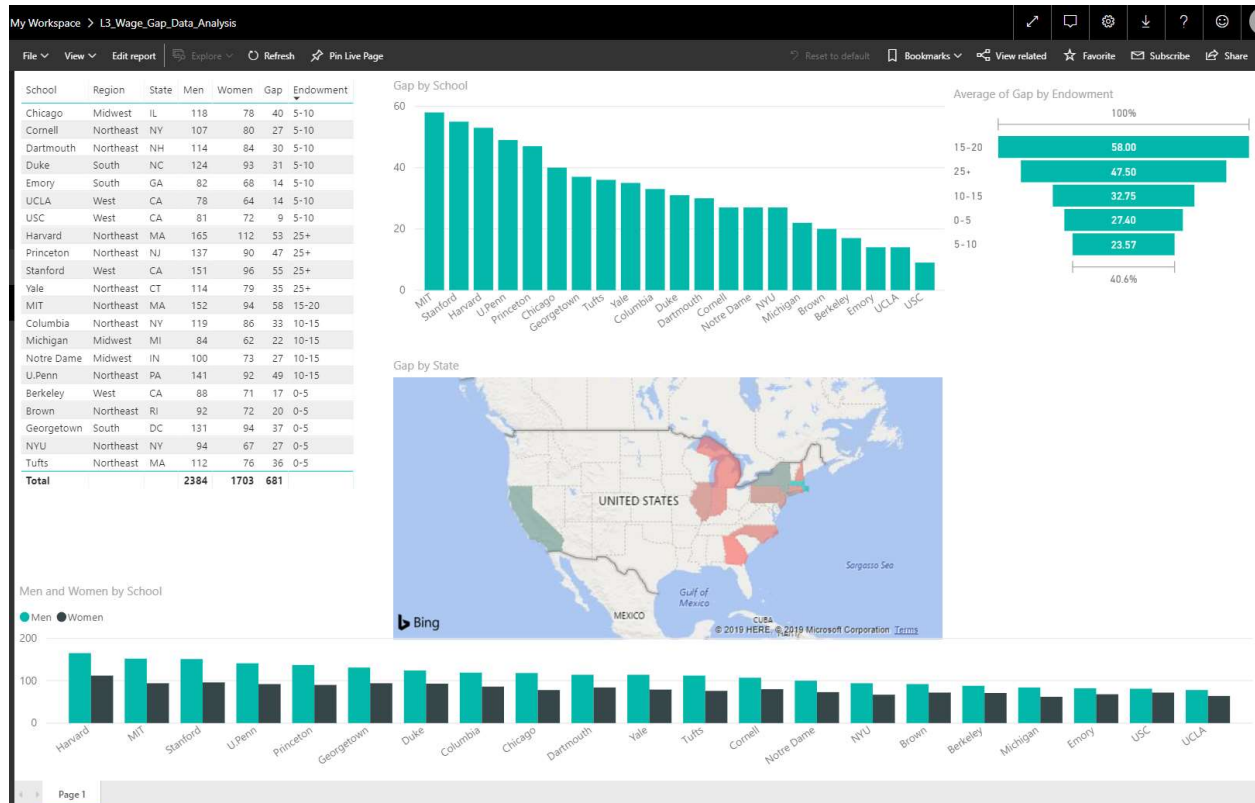


Add a Line and Clustered Bar Chart (select School, Men, Women)



PowerBI Report

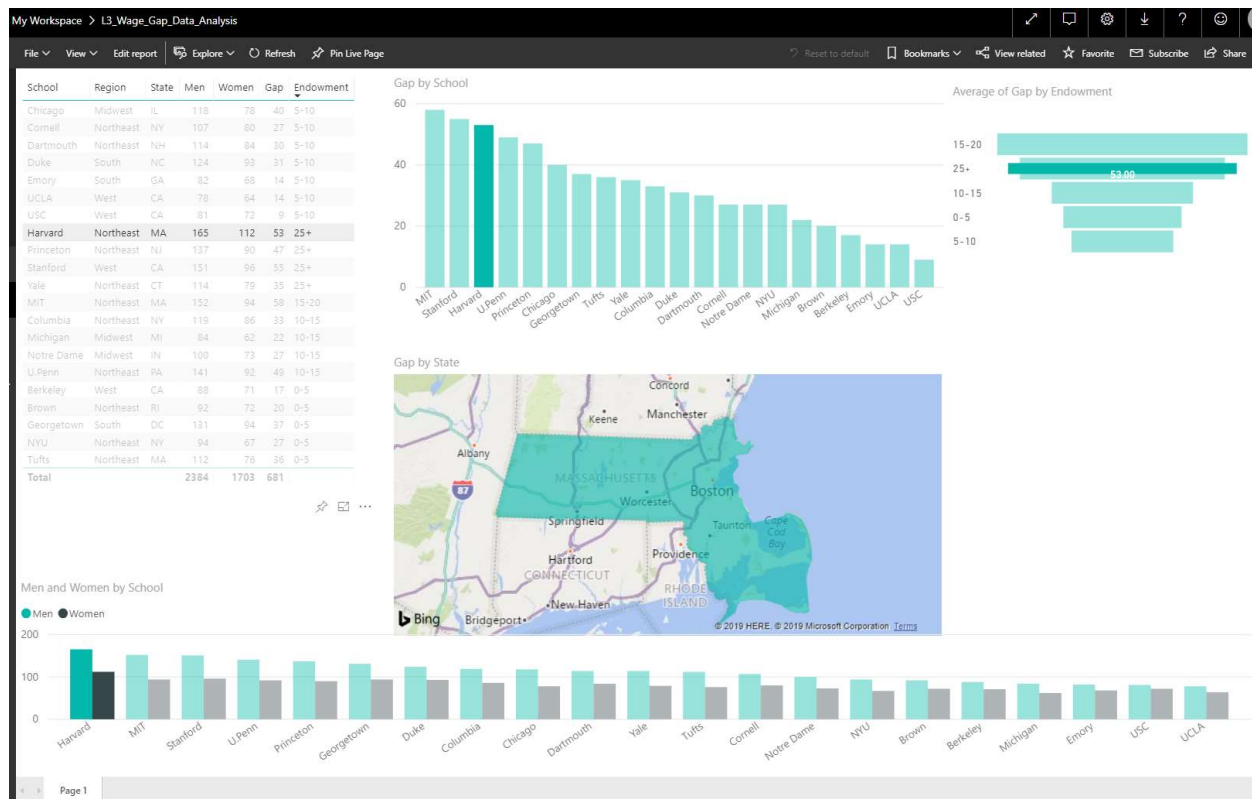
The report was deployed to a free PowerBI account (<http://app.powerbi.com>):



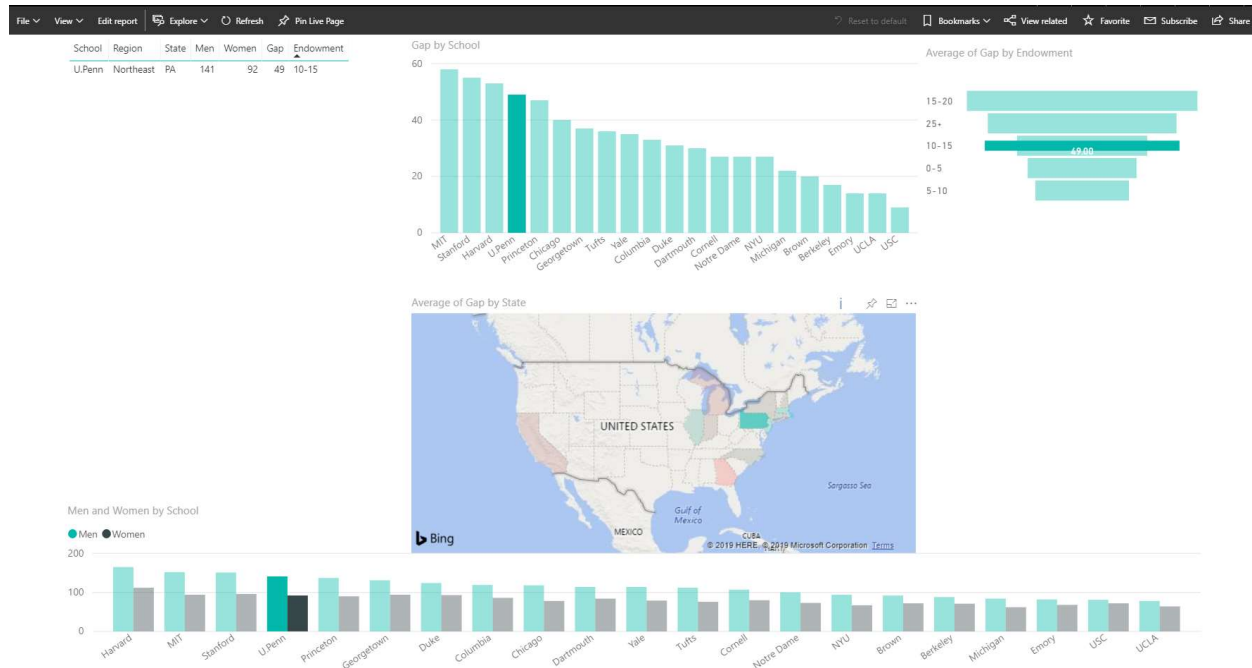
Interactivity

The great thing about PowerBI is that all the visualizations are interconnected by default. Clicking on the School in the table, highlights the bar in the bar chart, the school's section in the Mand and Women by School and the State in the Filled Map. Clicking on a state in the filled map, selects all of the corresponding data in the other data visualizations.

For example, select Harvard in the table:



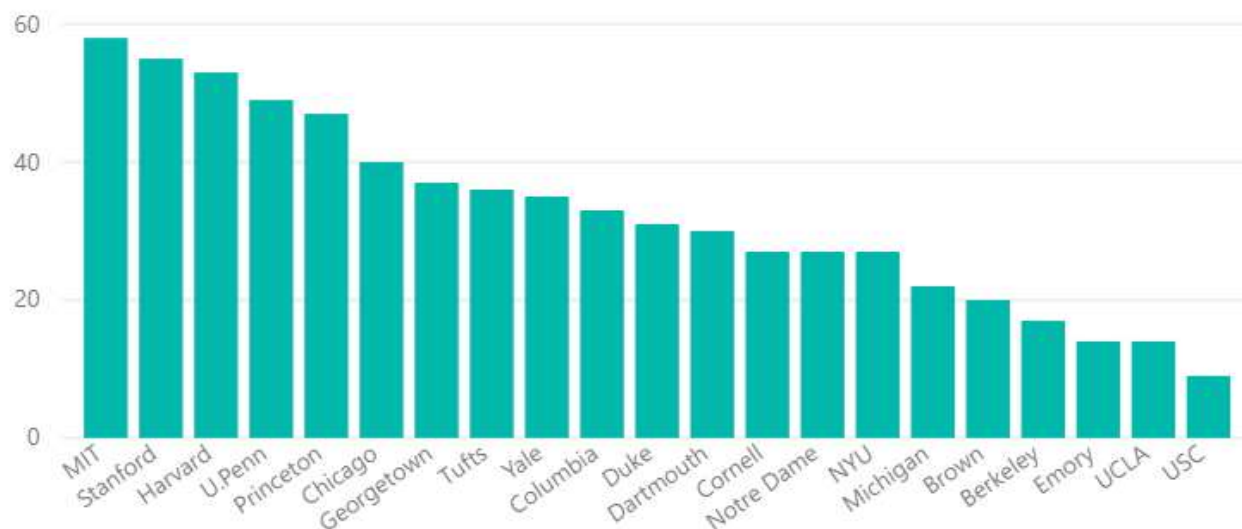
In another example, on the map select New York state:



Investigation

Is the wage gap constant, regardless of school?

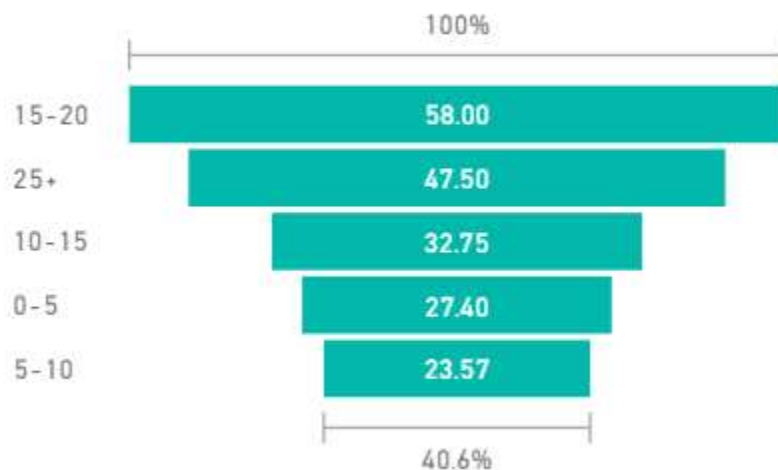
Gap by School



No, the Wage Gap by School chart shows that there are different values

Do attending certain schools (Ivy League, Public/Private), produce different wage?

Average of Gap by Endowment



There appears to be a bigger gap for schools that with an Endowment over \$15 Billion

Conclusion

Based on the sample data, there appears to be a wage gap between men and women. This gap is not constant across sampled institutions. The wage gap appears to be more pronounced when attending the “better” schools.

Future Work

The data needs to be improved to compare men and women by years’ of experience, job level, industry. The difference could be accounted for by higher level executives skewing the results.