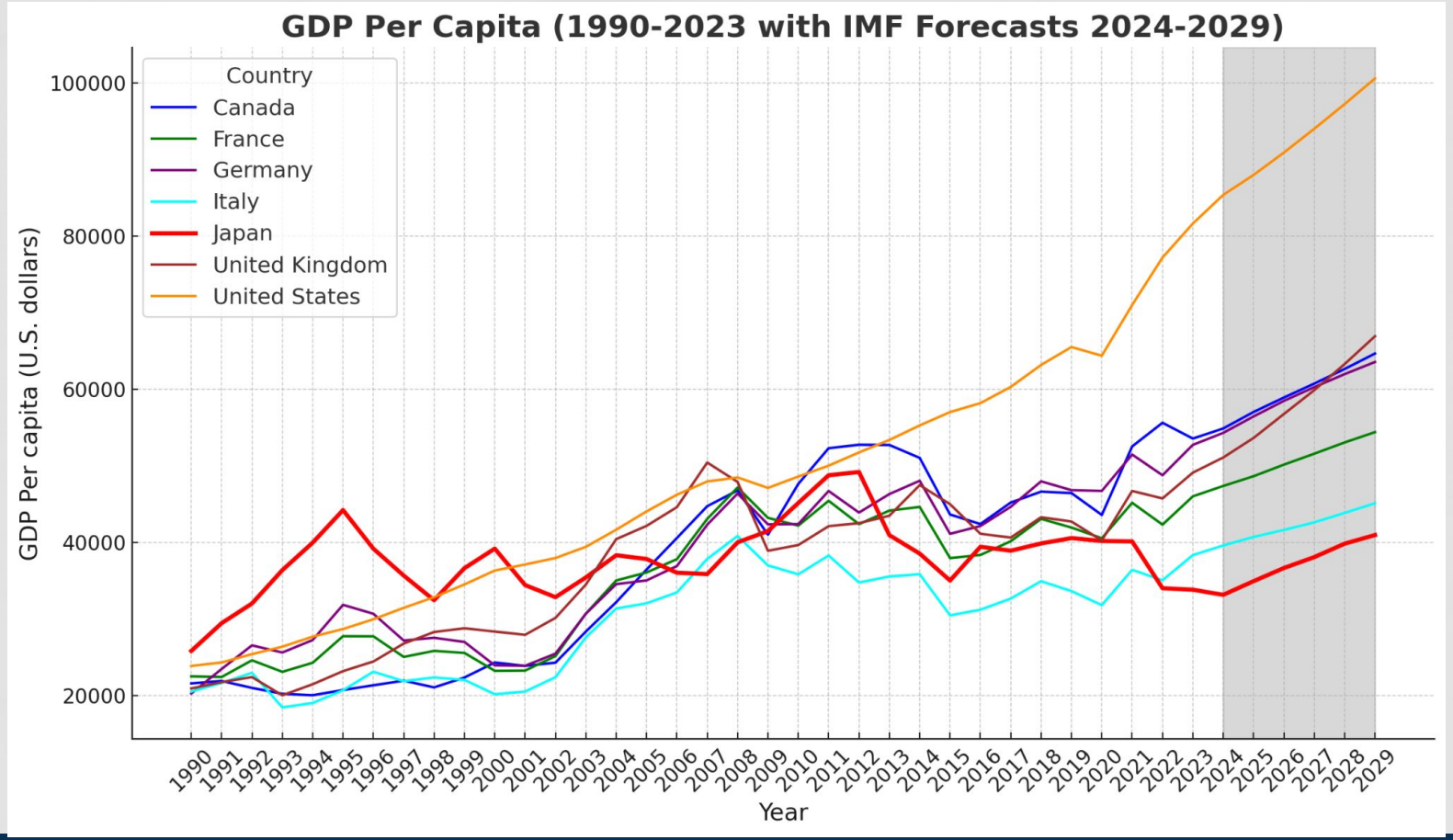
Data Analytics Using Python

Final Project

John Wu

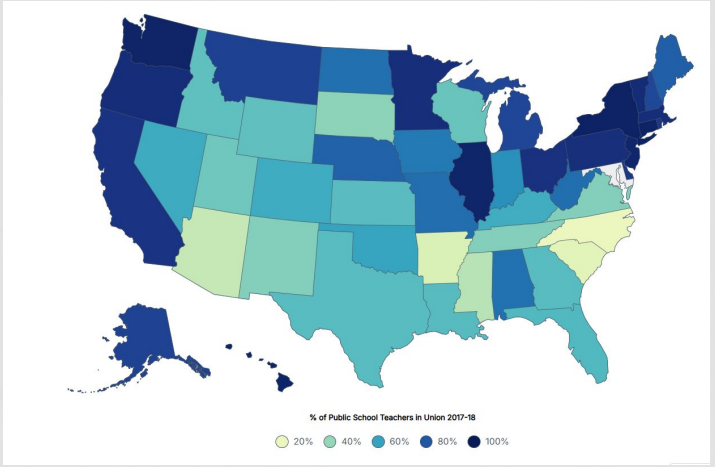
01

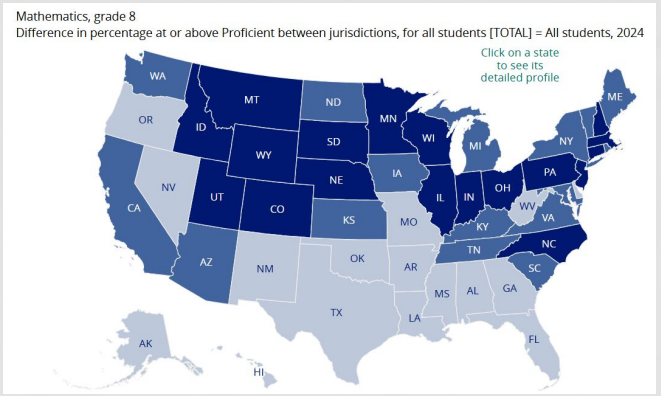
The Problem


GDP Per Capita (1990-2023 with IMF Forecasts 2024-2029)
Country
Canada
France
Germany
Italy
Japan
United Kingdom
United States
GDP Per capita (U.S. dollars)
100000
80000
60000
40000
20000
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
Year

02

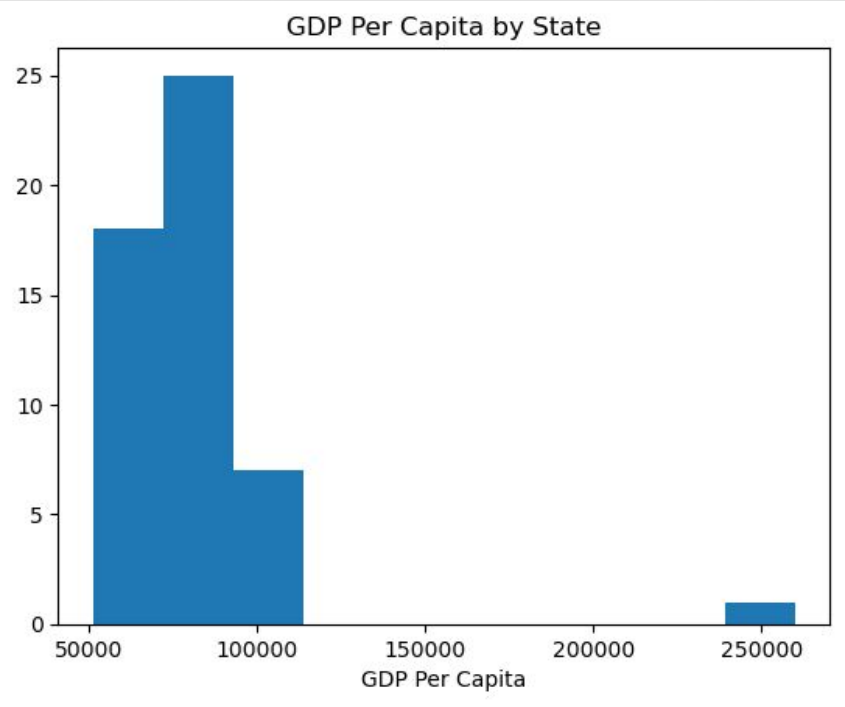
Investigation

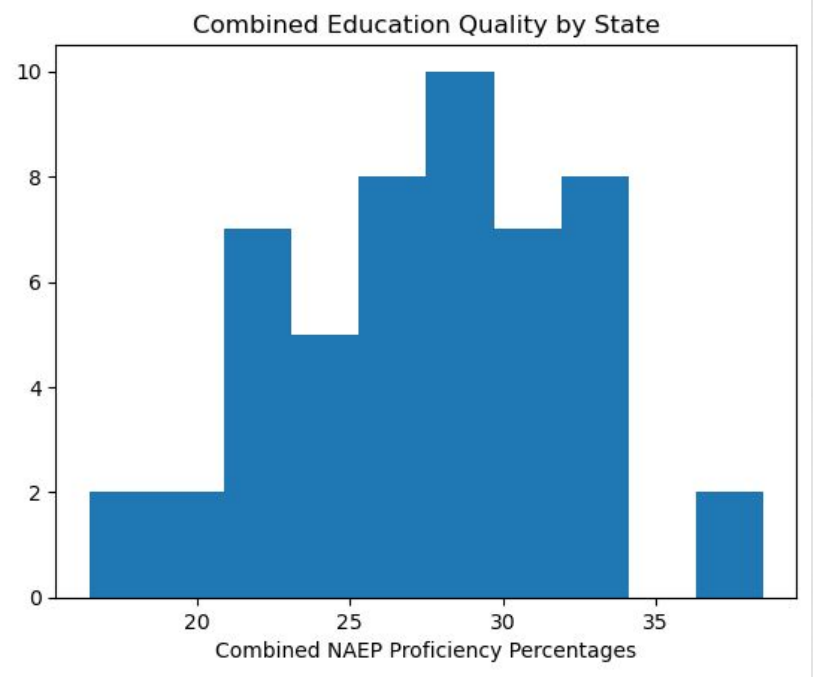

% of Public School Teachers in Union 2017-18
20%
40%
60%
80%
100%

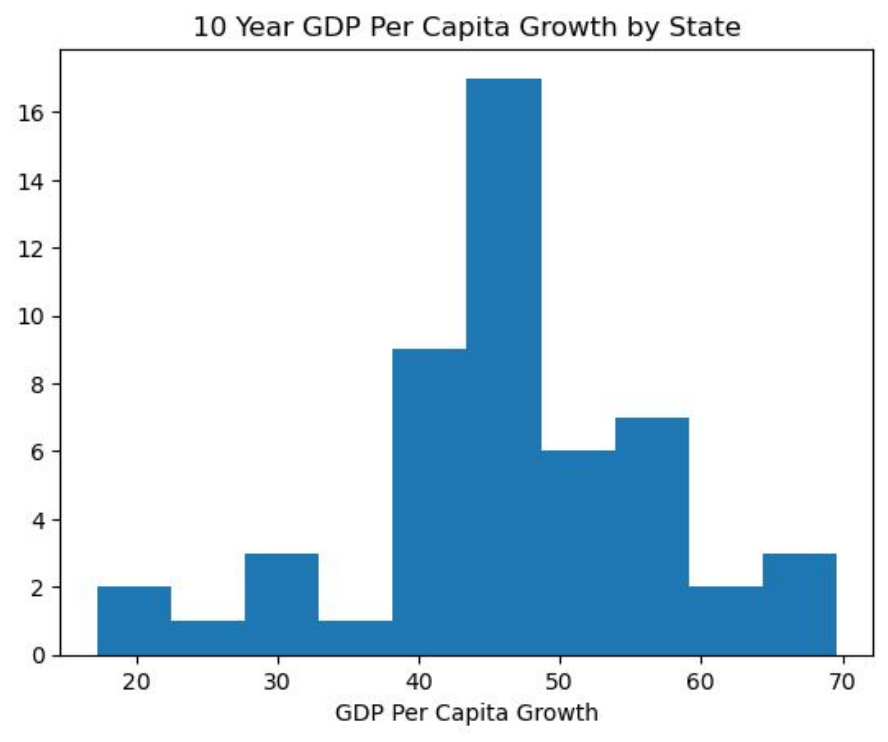

Mathematics, grade 8
Difference in percentage at or above Proficient between jurisdictions, for all students [TOTAL] = All students, 2024
Click on a state to see its detailed profile
WA
OR
CA
NV
ID
MT
WY
UT
AZ
CO
NM
ND
SD
NE
KS
OK
TX
MN
IA
MO
AR
LA
WI
IL
MS
MI
IN
KY
TN
AL
OH
WV
GA
FL
SC
NC
VA
PA
NY
ME
AK
HI

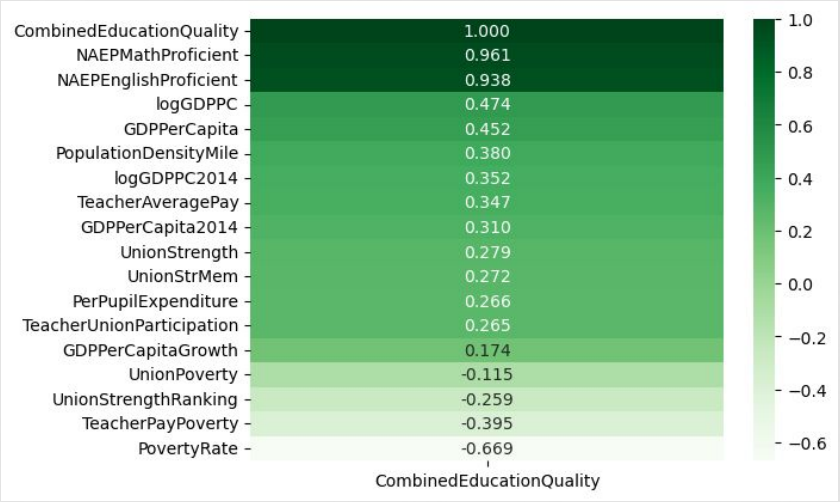
03

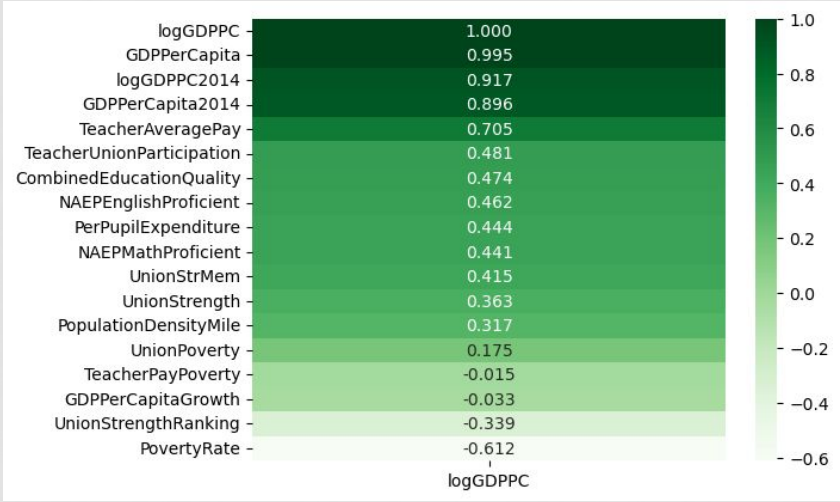
Exploration

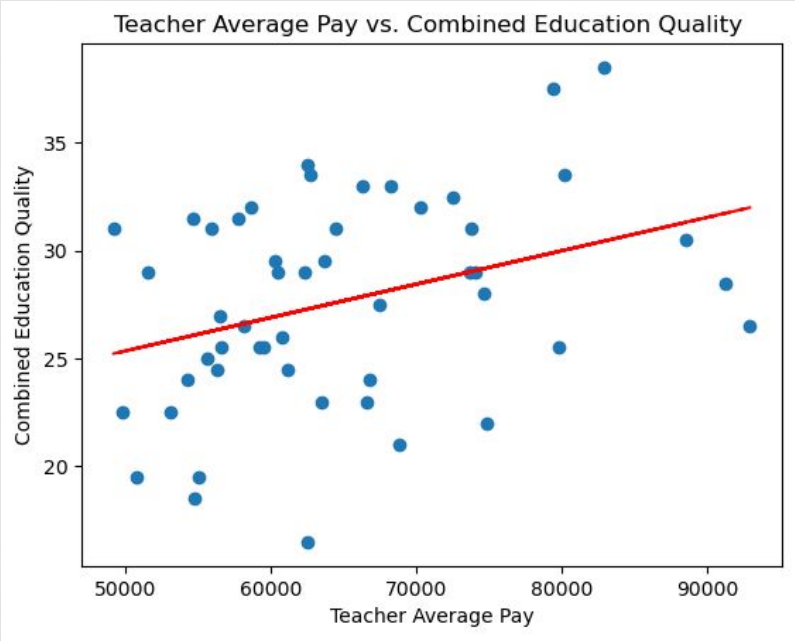

GDP Per Capita by State
25
20
15
10
5
0
50000
100000
150000
200000
250000
GDP Per Capita

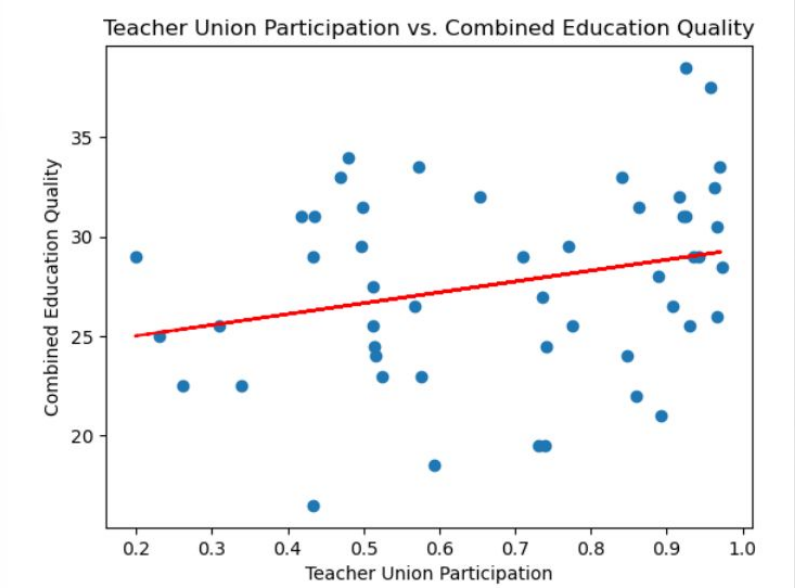

Combined Education Quality by State
10
8
6
4
2
0
20
25
30
35
Combined NAEP Proficiency Percentages

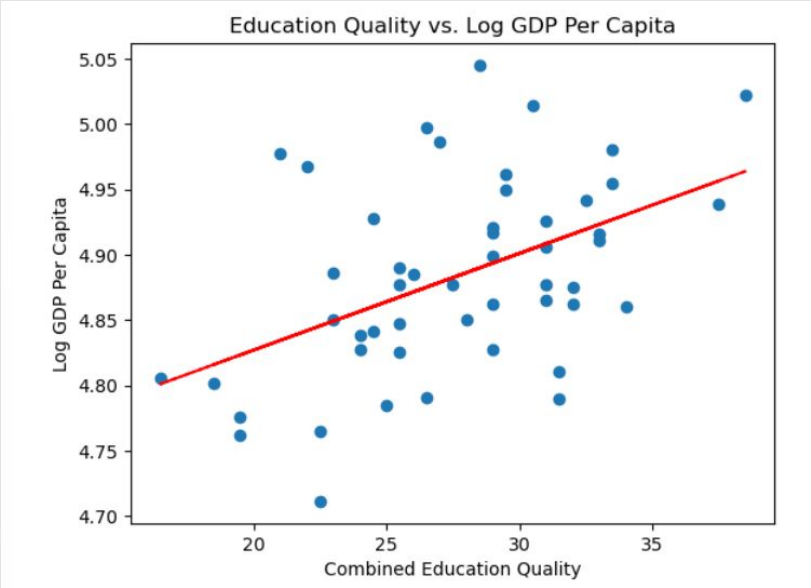

10 Year GDP Per Capita Growth by State
16
14
12
10
8
6
4
2
0
20
30
40
50
60
70
GDP Per Capita Growth

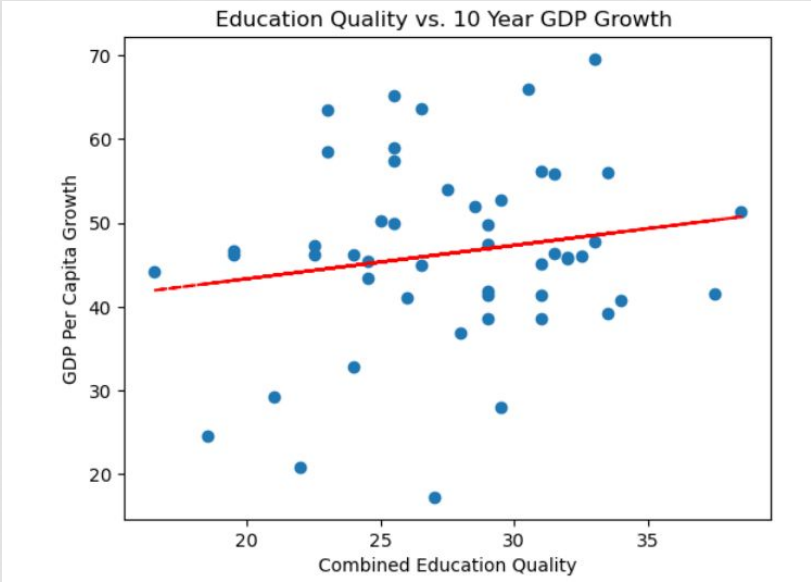

CombinedEducationQuality 1.000
NAEPMathProficient 0.961
NAEPEnglishProficient 0.938
logGDPPC 0.474
GDPPerCapita 0.452
PopulationDensityMile 0.380
logGDPPC2014 0.352
TeacherAveragePay 0.347
GDPPerCapita2014 0.310
UnionStrength 0.279
UnionStrMem 0.272
PerPupilExpenditure 0.266
TeacherUnionParticipation 0.265
GDPPerCapitaGrowth 0.174
UnionPoverty -0.115
UnionStrengthRanking -0.259
TeacherPayPoverty -0.395
PovertyRate -0.669
CombinedEducationQuality
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

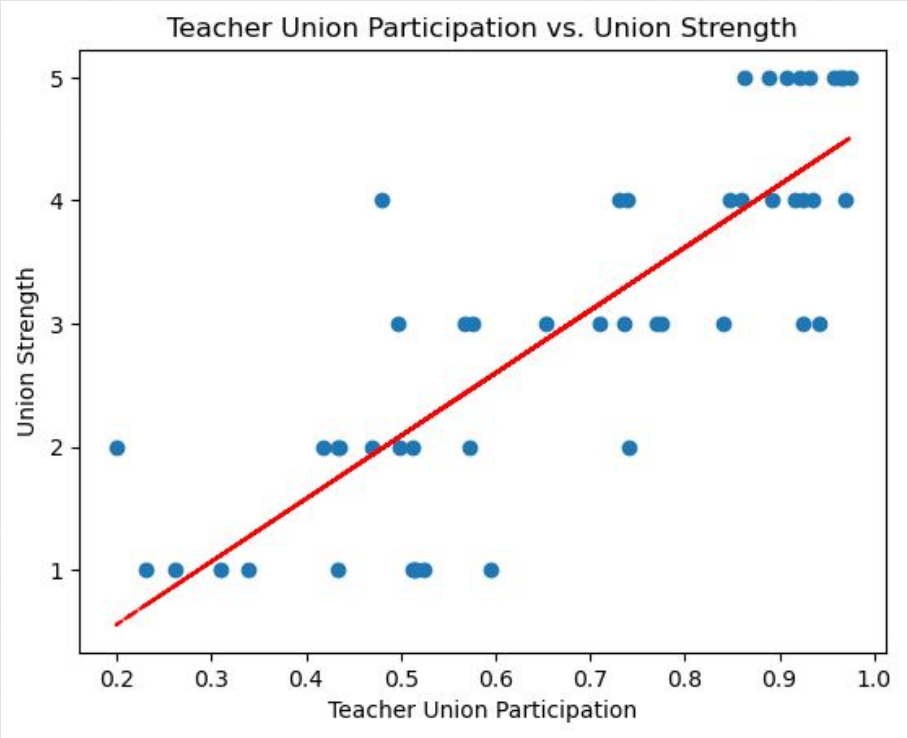

logGDPPC 1.000
GDPPerCapita 0.995
logGDPPC2014 0.917
GDPPerCapita2014 0.896
TeacherAveragePay 0.705
TeacherUnionParticipation 0.481
CombinedEducationQuality 0.474
NAEPEnglishProficient 0.462
PerPupilExpenditure 0.444
NAEPMathProficient 0.441
UnionStrMem 0.415
UnionStrength 0.363
PopulationDensityMile 0.317
UnionPoverty 0.175
TeacherPayPoverty -0.015
GDPPerCapitaGrowth -0.033
UnionStrengthRanking -0.339
PovertyRate -0.612
logGDPPC
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


Teacher Average Pay vs. Combined Education Quality
Combined Education Quality
Teacher Average Pay
50000
60000
70000
80000
90000
20
25
30
35


Teacher Union Participation vs. Combined Education Quality
Combined Education Quality
Teacher Union Participation
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
20
25
30
35

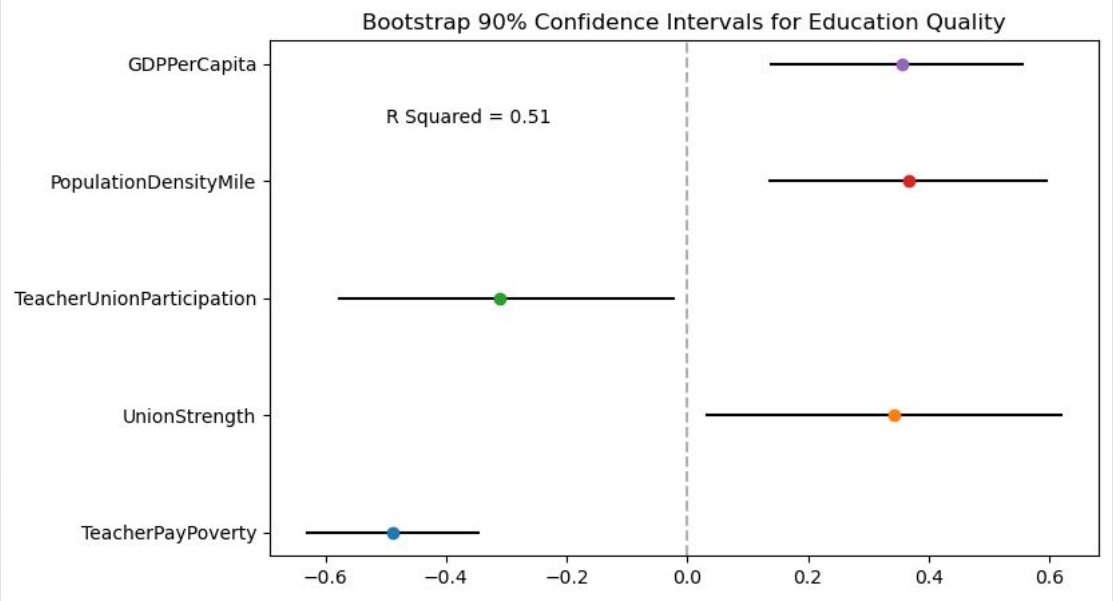

Education Quality vs. Log GDP Per Capita
Log GDP Per Capita
Combined Education Quality
5.05
5.00
4.95
4.90
4.85
4.80
4.75
4.70
20
25
30
35

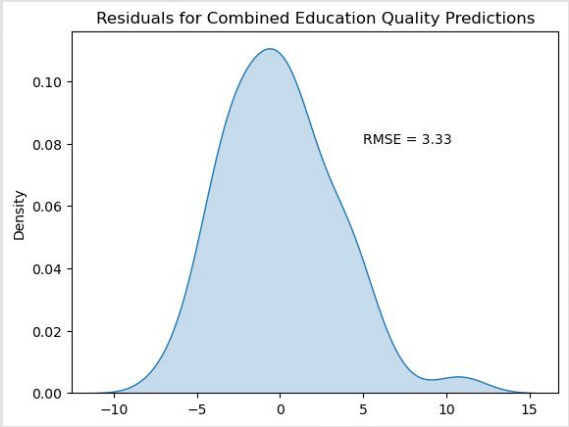

Education Quality vs. 10 Year GDP Growth
GDP Per Capita Growth
Combined Education Quality
70
60
50
40
30
20
20
25
30
35

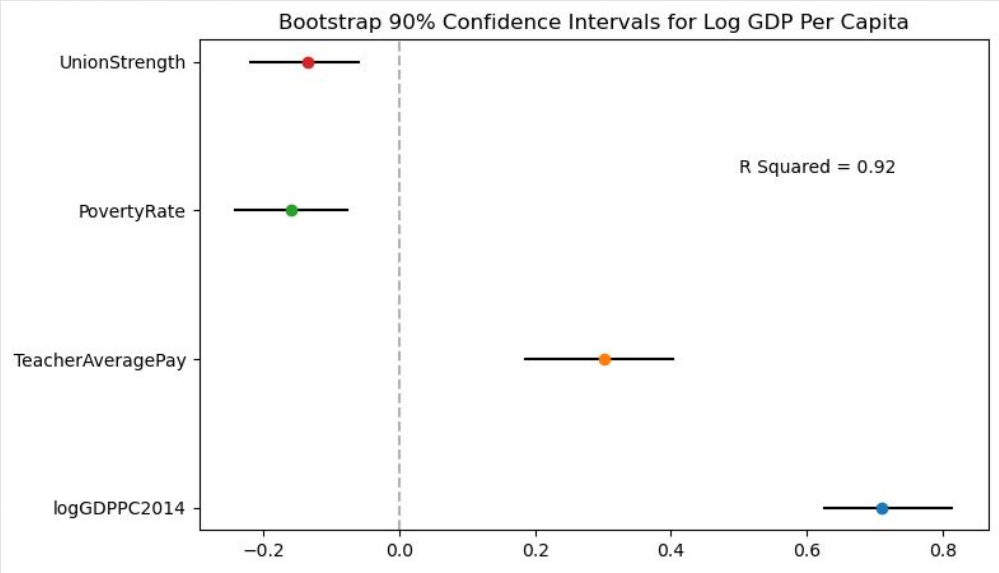

Teacher Union Participation vs. Union Strength
Union Strength
Teacher Union Participation
1
2
3
4
5
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

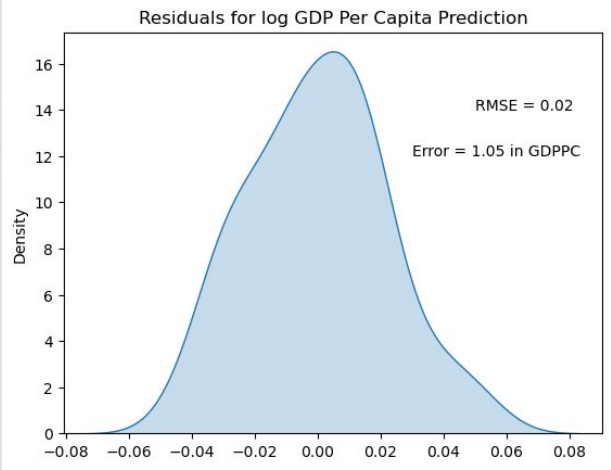
04

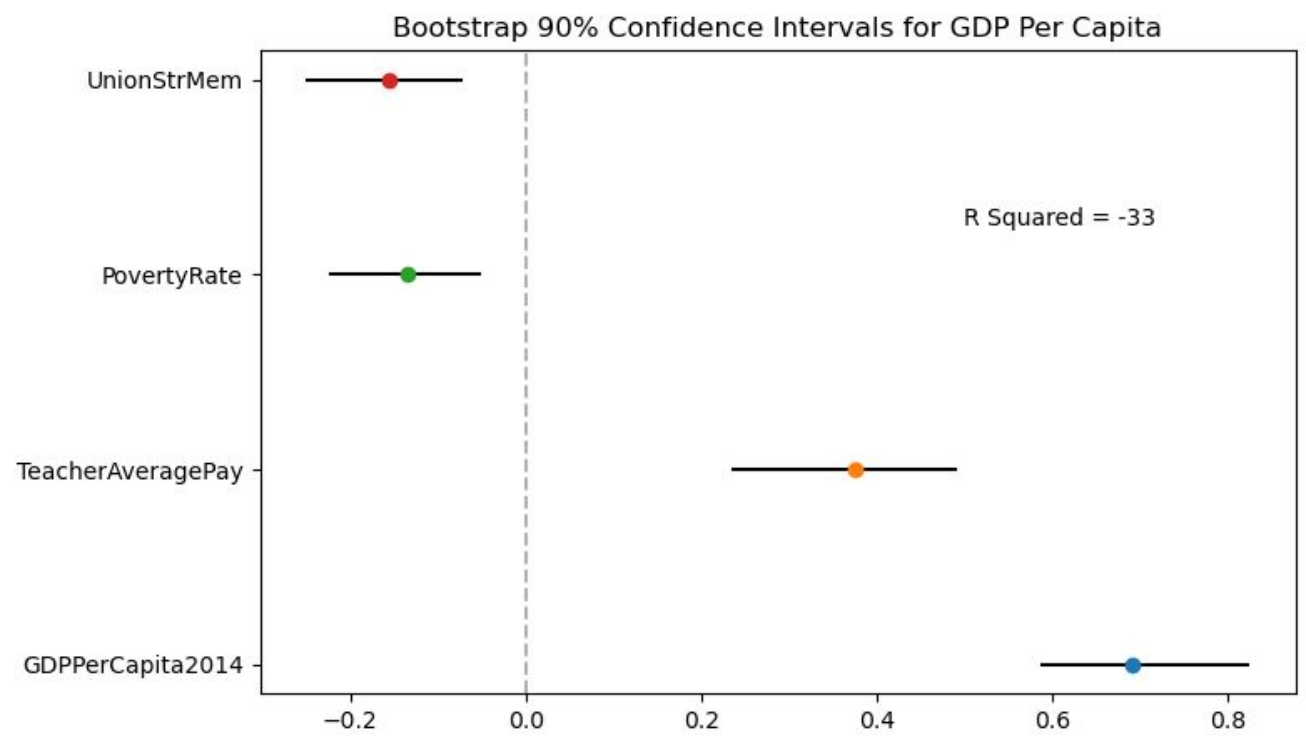
Models

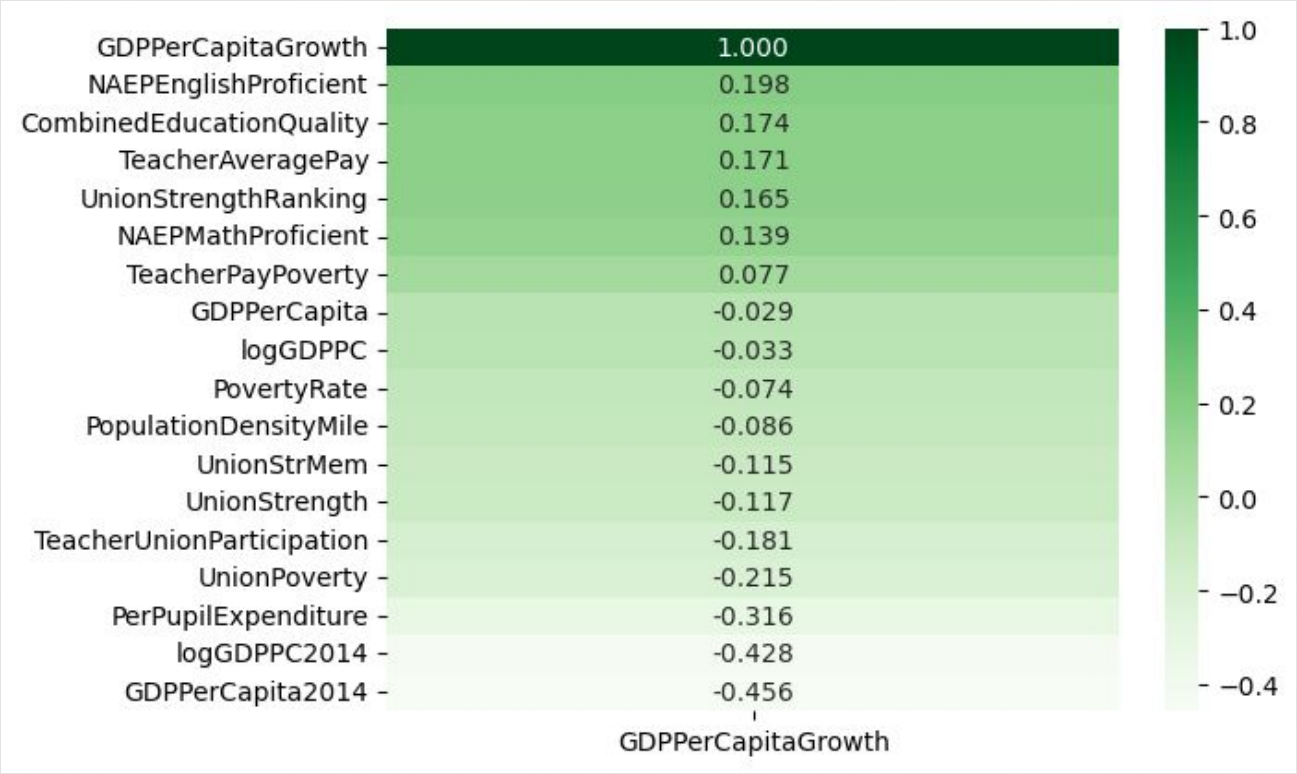

Bootstrap 90% Confidence Intervals for Education Quality
GDPPerCapita
PopulationDensityMile
TeacherUnionParticipation
UnionStrength
TeacherPayPoverty
R Squared = 0.51
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6

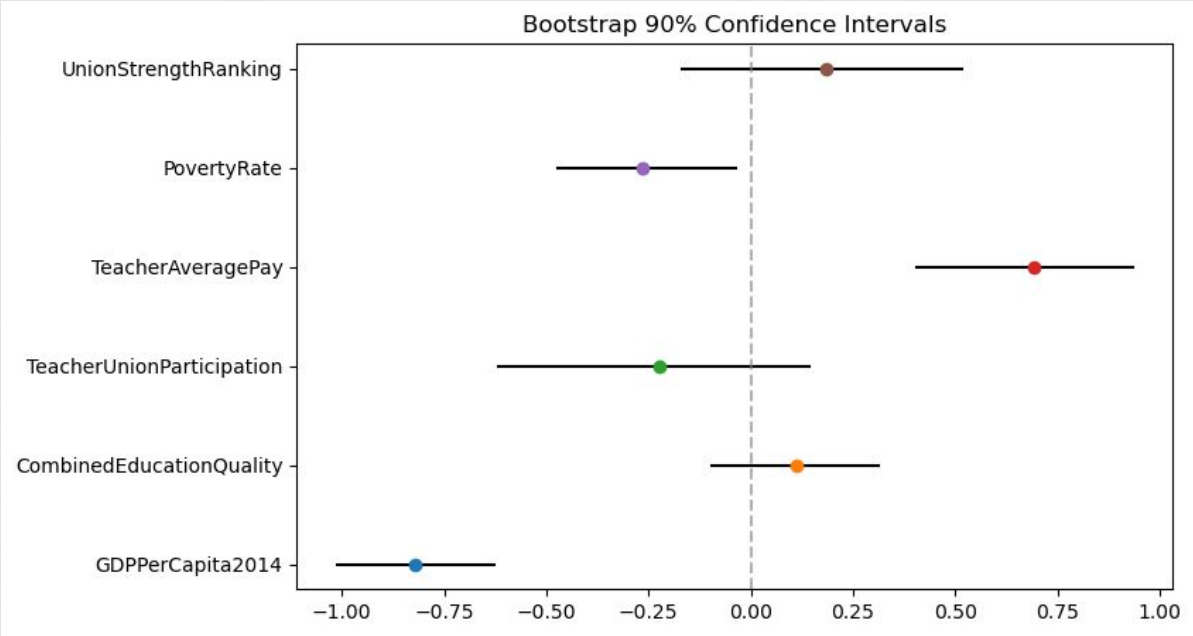

Residuals for Combined Education Quality Predictions
Density
RMSE = 3.33
0.00
0.02
0.04
0.06
0.08
0.10
−10
−5
0
5
10
15

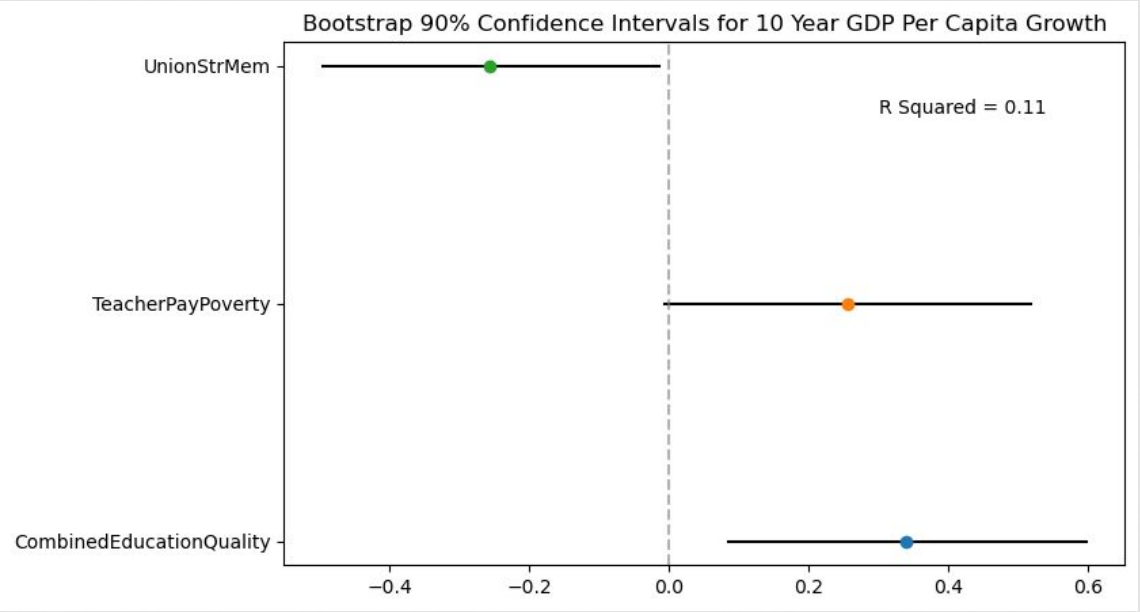

Bootstrap 90% Confidence Intervals for Log GDP Per Capita
UnionStrength
PovertyRate
TeacherAveragePay
logGDPPC2014
R Squared = 0.92
−0.2
0.0
0.2
0.4
0.6
0.8

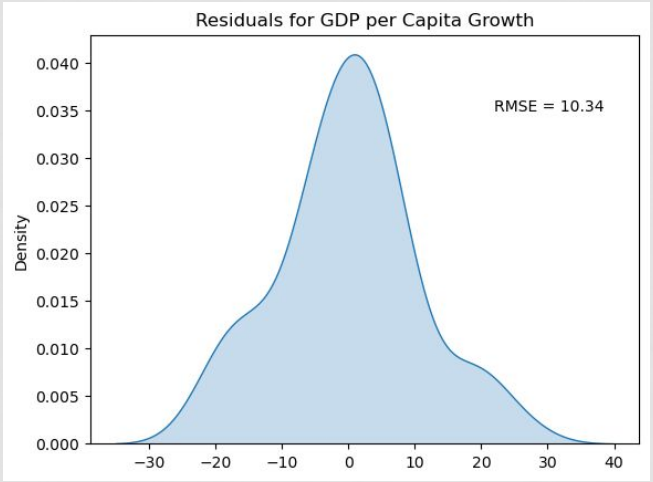

Residuals for log GDP Per Capita Prediction
Density
RMSE = 0.02
Error = 1.05 in GDPPC
0
2
4
6
8
10
12
14
16
−0.08
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06
0.08


Bootstrap 90% Confidence Intervals for GDP Per Capita
UnionStrMem
PovertyRate
TeacherAveragePay
GDPPerCapita2014
R Squared = -33
−0.2
0.0
0.2
0.4
0.6
0.8


GDPPerCapitaGrowth 1.000
NAEPEnglishProficient 0.198
CombinedEducationQuality 0.174
TeacherAveragePay 0.171
UnionStrengthRanking 0.165
NAEPMathProficient 0.139
TeacherPayPoverty 0.077
GDPPerCapita -0.029
logGDPPC -0.033
PovertyRate -0.074
PopulationDensityMile -0.086
UnionStrMem -0.115
UnionStrength -0.117
TeacherUnionParticipation -0.181
UnionPoverty -0.215
PerPupilExpenditure -0.316
logGDPPC2014 -0.428
GDPPerCapita2014 -0.456
GDPPerCapitaGrowth
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

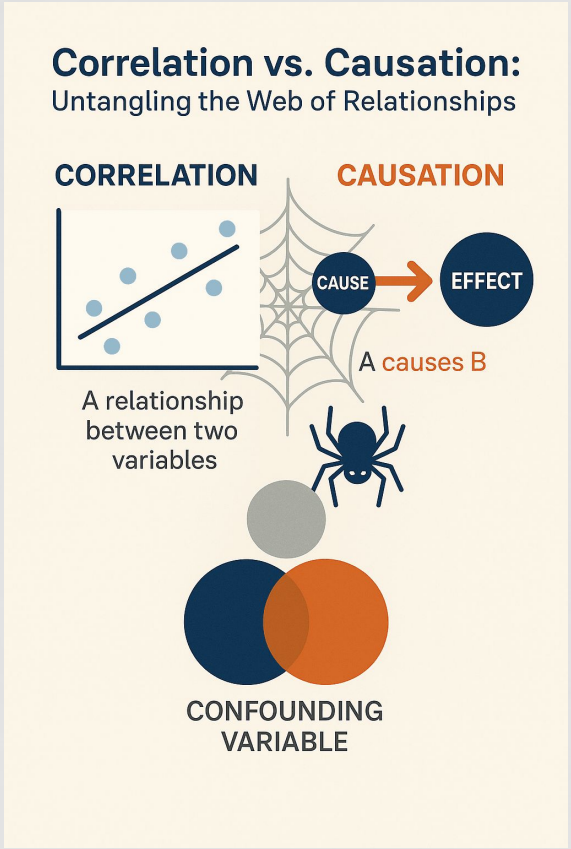

Bootstrap 90% Confidence Intervals
UnionStrengthRanking
PovertyRate
TeacherAveragePay
TeacherUnionParticipation
CombinedEducationQuality
GDPPerCapita2014
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

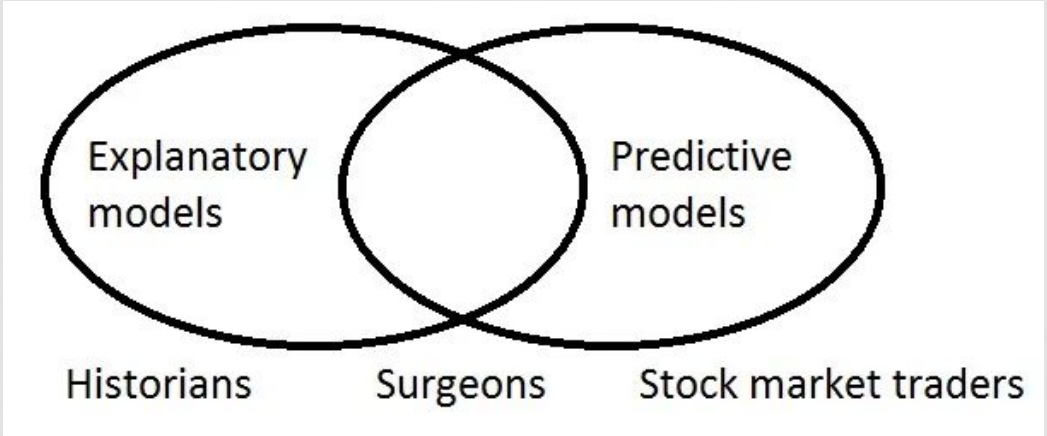

Bootstrap 90% Confidence Intervals for 10 Year GDP Per Capita Growth
UnionStrMem
TeacherPayPoverty
CombinedEducationQuality
R Squared = 0.11
−0.4
−0.2
0.0
0.2
0.4
0.6

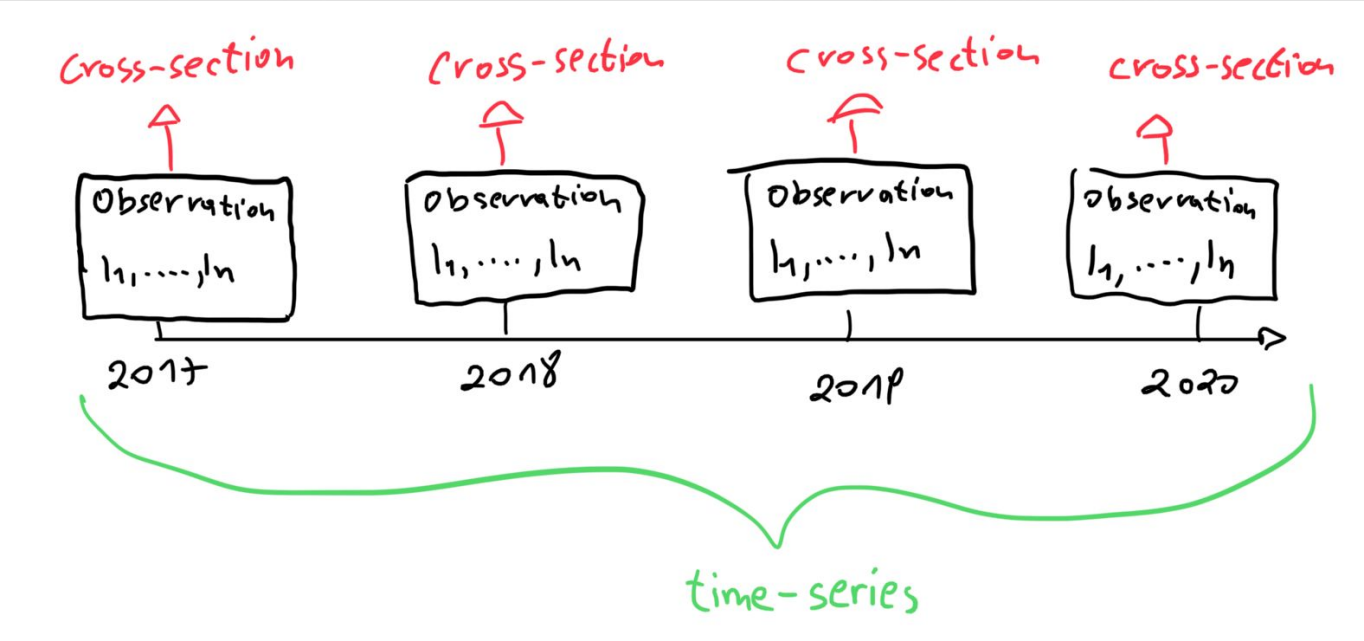

Residuals for GDP per Capita Growth
RMSE = 10.34
Density
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
−30
−20
−10
0
10
20
30
40

05

Summary & Future


Correlation vs. Causation:
Untangling the Web of Relationships
CORRELATION
CAUSATION
CAUSE
EFFECT
A causes B
A relationship between two variables
CONFOUNDING VARIABLE


Explanatory models
Predictive models
Historians
Surgeons
Stock market traders


Cross-section
Cross-section
Cross-section
Cross-section
Observation
$l_1, \ldots, l_n$
Observation
$l_1, \ldots, l_n$
Observation
$l_1, \ldots, l_n$
Observation
$l_1, \ldots, l_n$
2017
2018
2019
2020
time-series