The STEM Pipeline

www.businessandeducation.org

Sean Connin
Initial EDA and Models
5/21

Questions:

1. Has the percentage of women completing STEM degrees changed during the past decade?

2. Do women STEM completion rates vary in relation to extrinsic factors such as campus location, size, instructional expenditures, etc?

IPEDS Integrated Postsecondary Education Data System

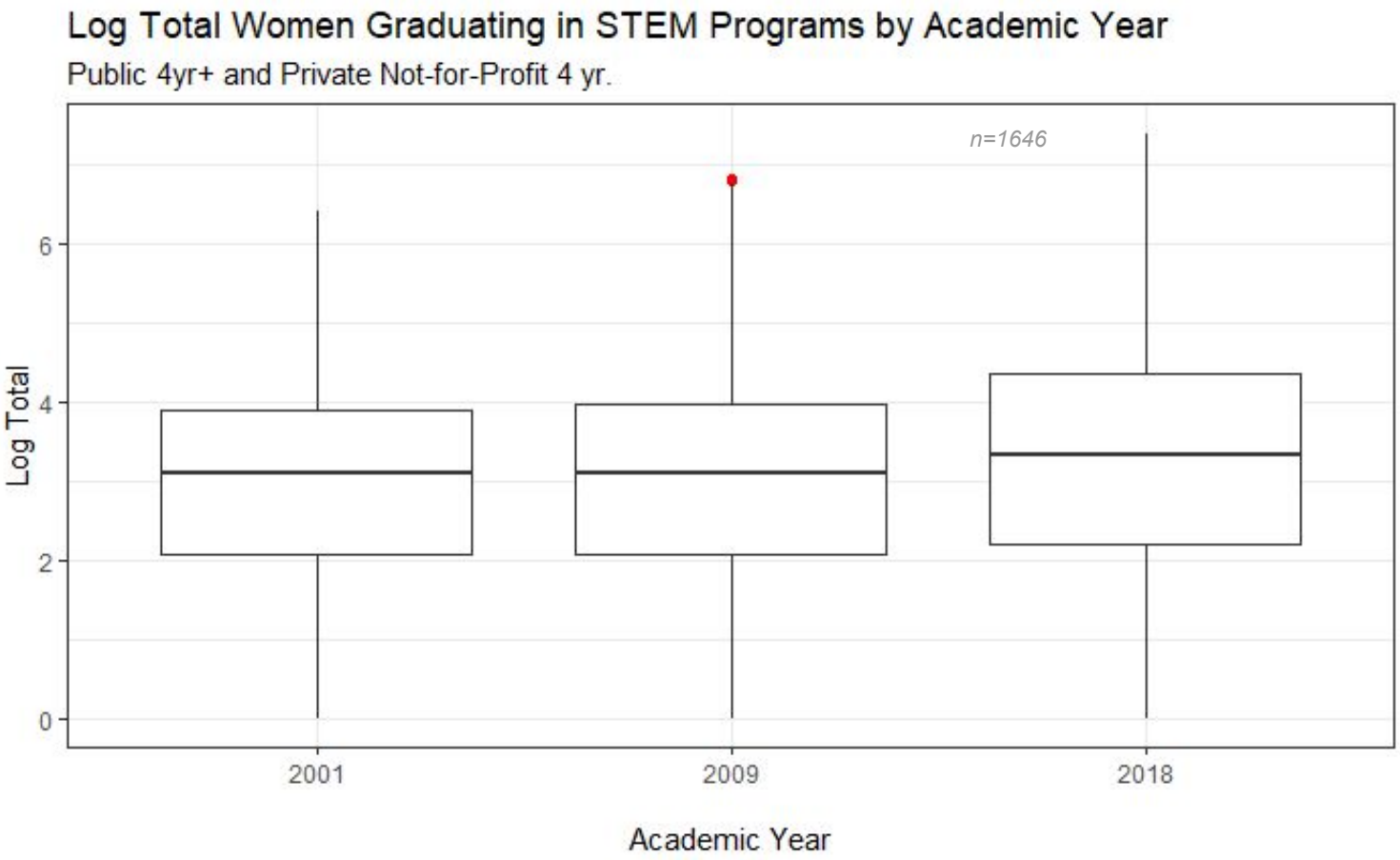
https://nces.ed.gov

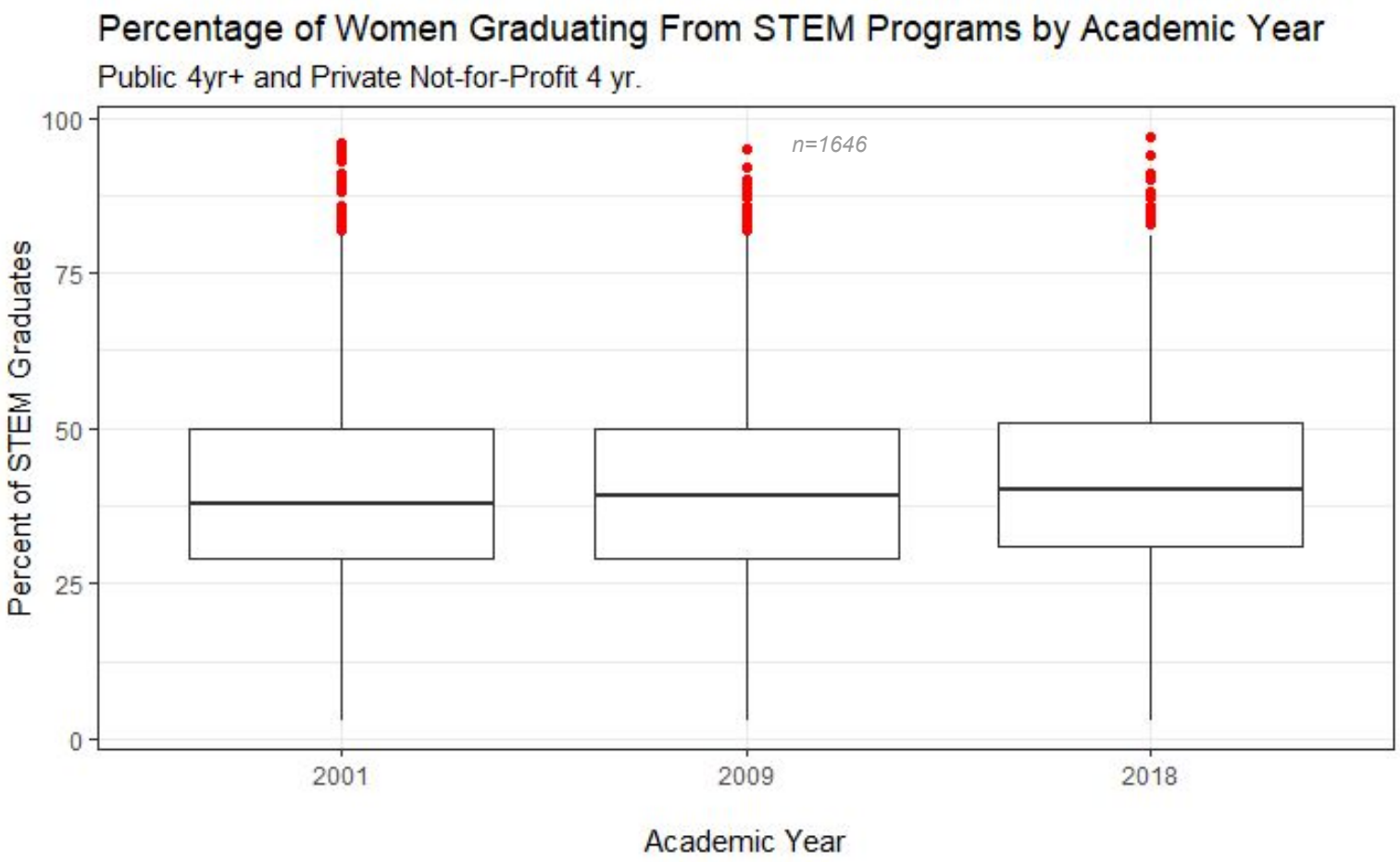
- Institutional Characteristics
- Institutional Prices
- Admissions
- Enrollment
- Student Financial Aid
- Degrees and Certificates Conferred (Completions)
- Student Persistence and Success
- Institutional Resources

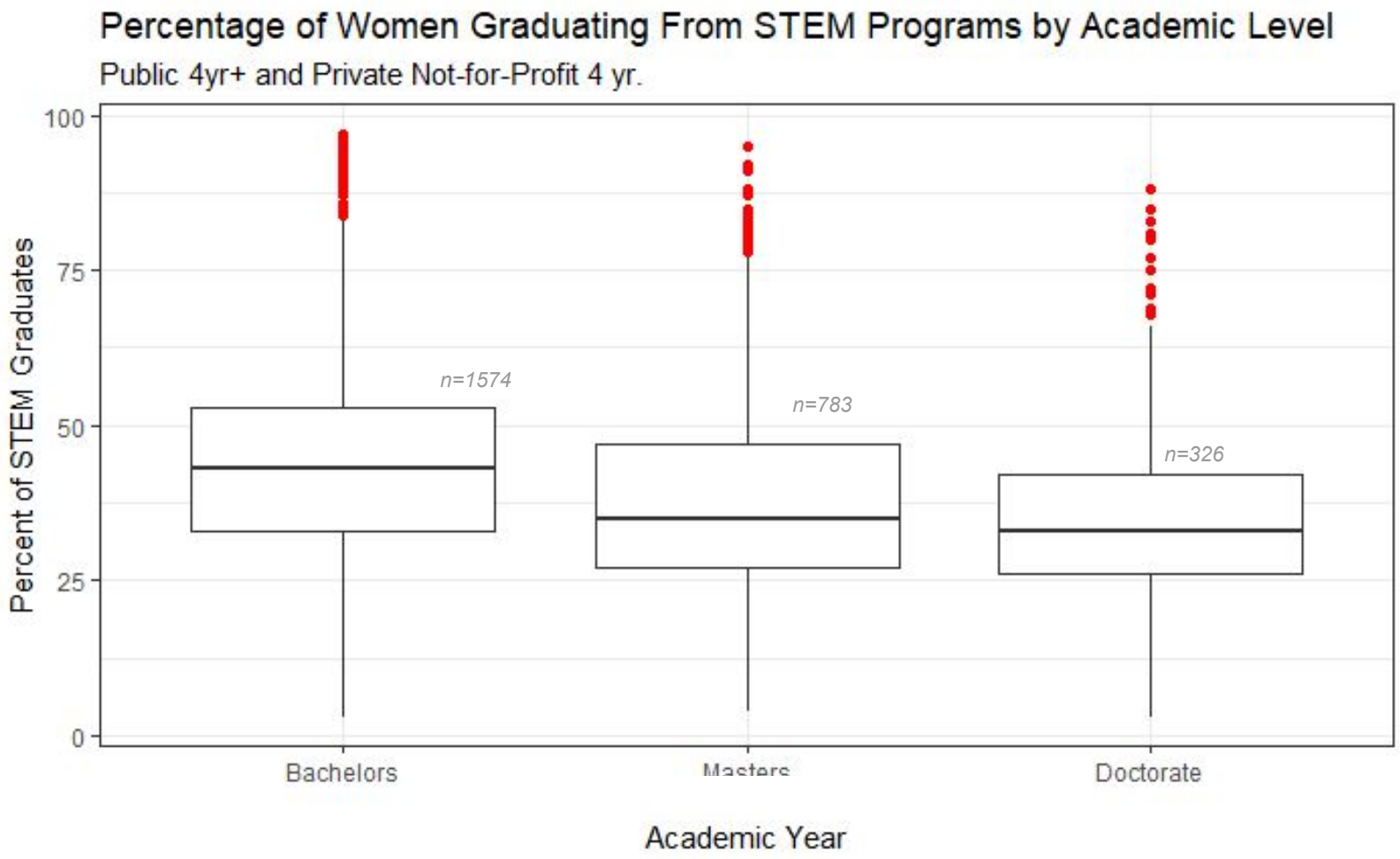
Have STEM completion rates among women changed during the past decade?

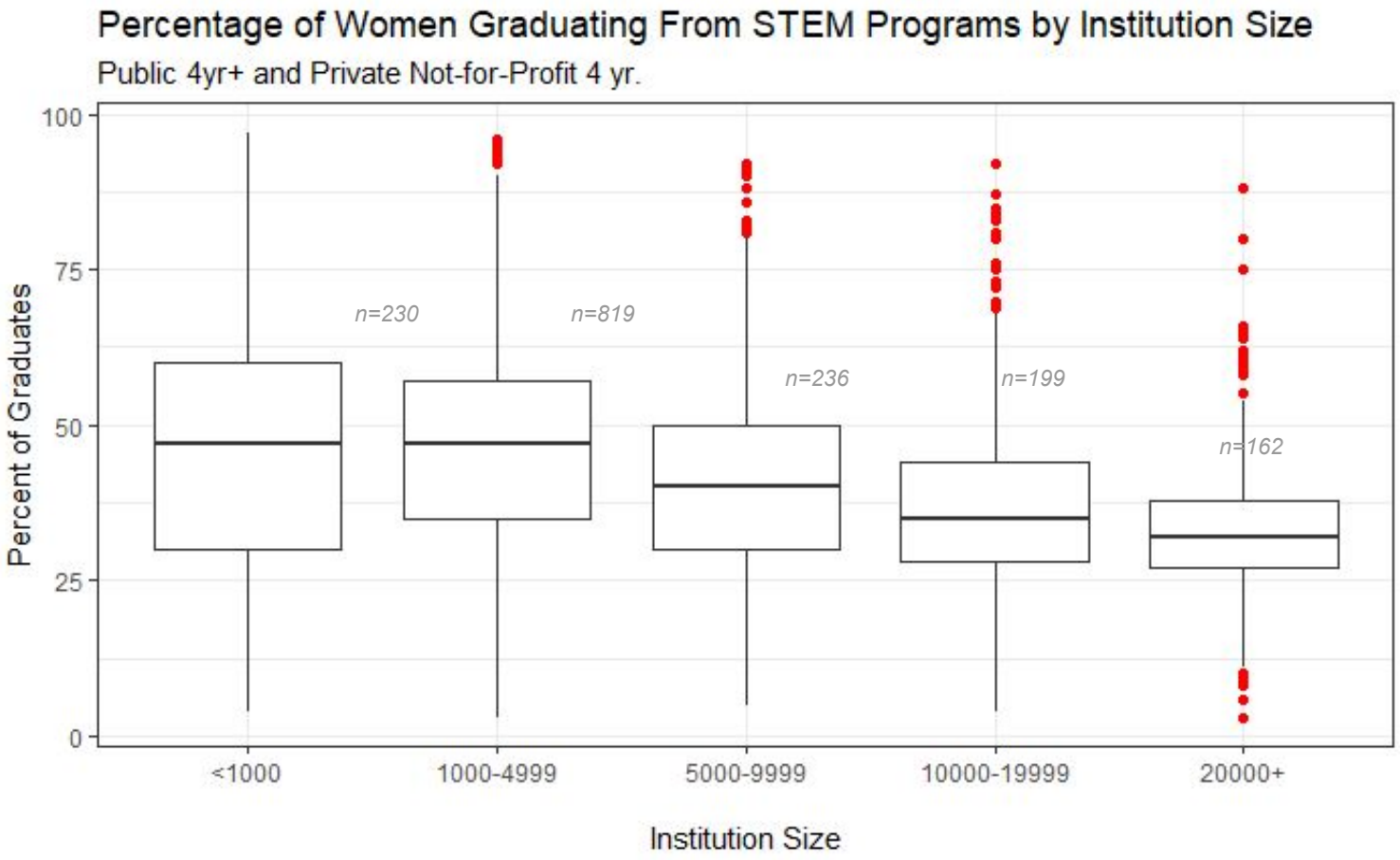
:

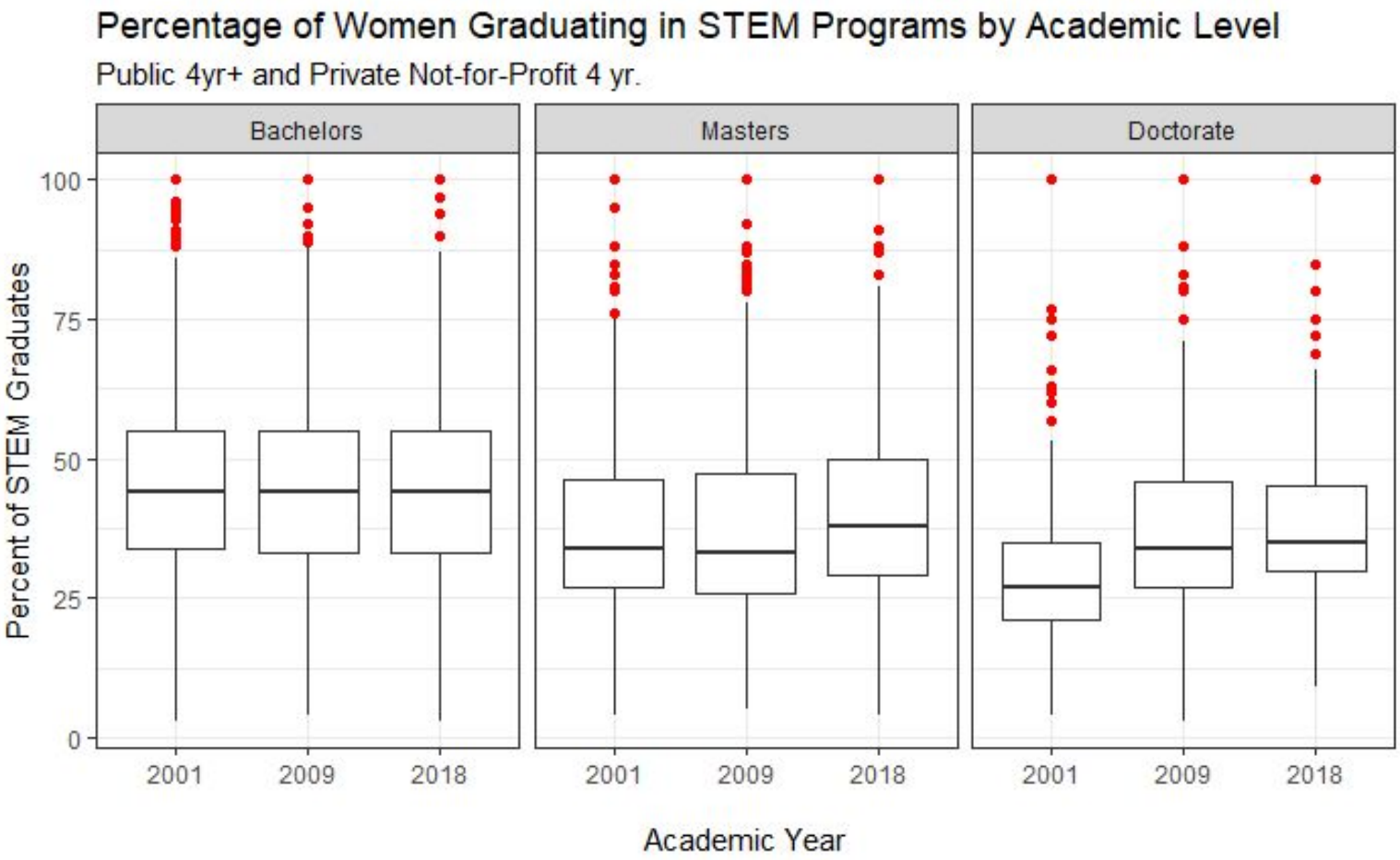
- **Year**
  - 2000-01
  - 2009-10
  - 2018-19
- **Degree**
  - Bachelors
  - Masters
  - Doctorate
- **Size**
  - < 4999
  - 5000-9999
  - 10000-1999
  - 20000+

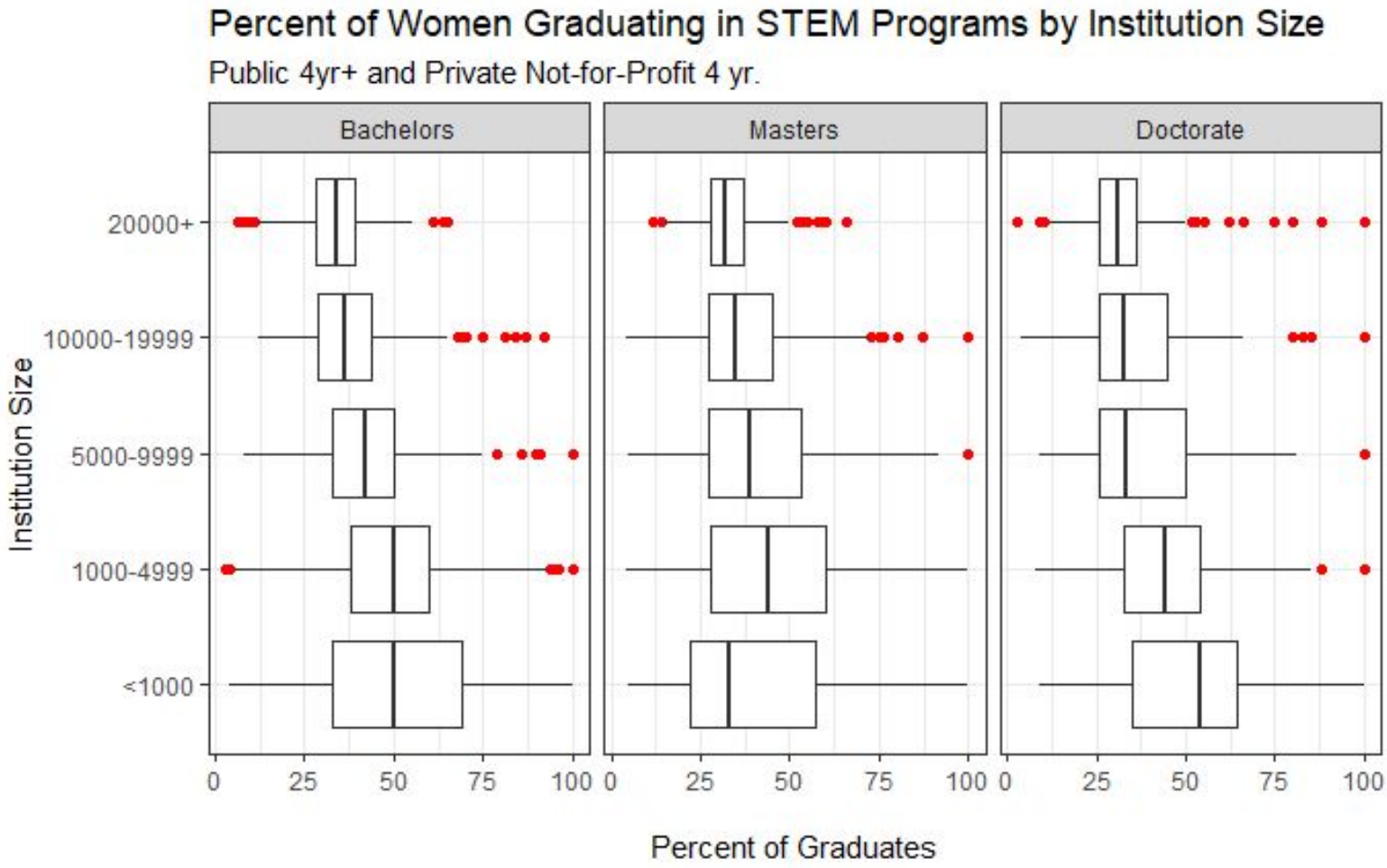

Log Total Women Graduating in STEM Programs by Academic Year
Public 4yr+ and Private Not-for-Profit 4 yr.
n=1646
Log Total
0
2
4
6
2001
2009
2018
Academic Year


Percentage of Women Graduating From STEM Programs by Academic Year
Public 4yr+ and Private Not-for-Profit 4 yr.
n=1646
Percent of STEM Graduates
100
75
50
25
0
2001
2009
2018
Academic Year


Percentage of Women Graduating From STEM Programs by Academic Level
Public 4yr+ and Private Not-for-Profit 4 yr.
Percent of STEM Graduates
100
75
50
25
0
n=1574
n=783
n=326
Bachelors
Masters
Doctorate
Academic Year

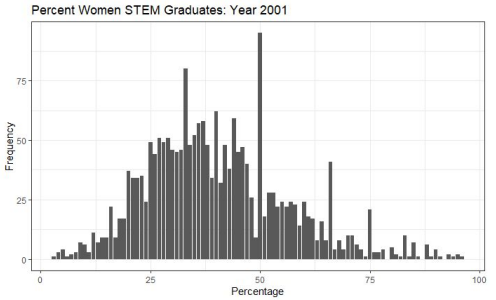

Percentage of Women Graduating From STEM Programs by Institution Size
Public 4yr+ and Private Not-for-Profit 4 yr.
Percent of Graduates
100
75
50
25
0
n=230
n=819
n=236
n=199
n=162
<1000
1000-4999
5000-9999
10000-19999
20000+
Institution Size

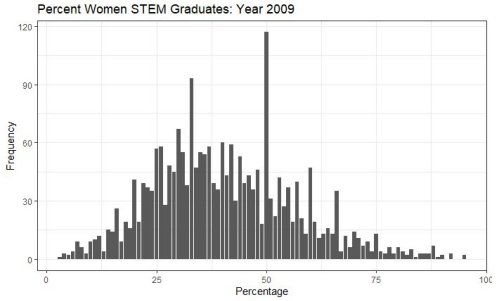

Percentage of Women Graduating in STEM Programs by Academic Level
Public 4yr+ and Private Not-for-Profit 4 yr.
Bachelors
Masters
Doctorate
Percent of STEM Graduates
100
75
50
25
0
2001
2009
2018
Academic Year

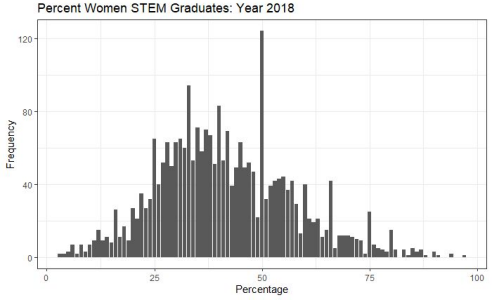

Percent of Women Graduating in STEM Programs by Institution Size
Public 4yr+ and Private Not-for-Profit 4 yr.
Bachelors
Masters
Doctorate
Institution Size
20000+
10000-19999
5000-9999
1000-4999
<1000
0
25
50
75
100
Percent of Graduates

Assumptions for ANOVA Tests

1. The responses for each factor level have a normal population distribution.
2. These distributions have the same variance.
3. The data are independent.







Three Way ANOVA

Main Effects

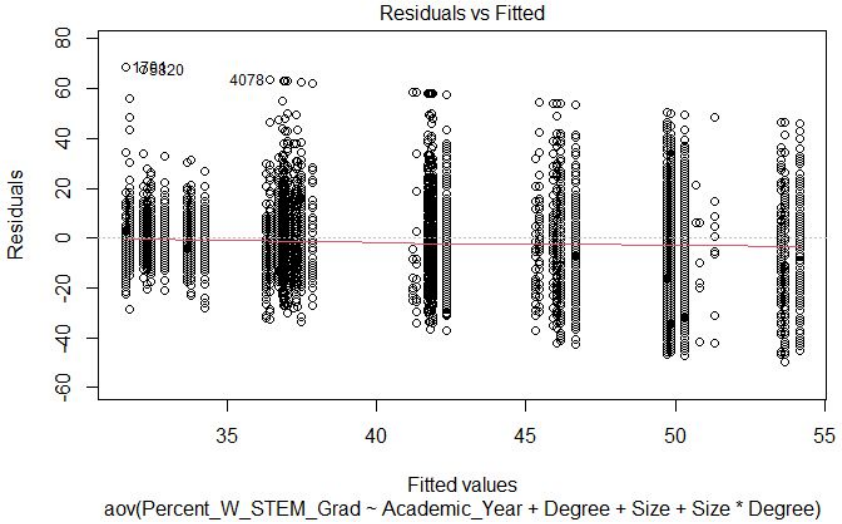
```
                Df  Sum Sq Mean Sq F value Pr(>F)
Academic_Year    2    1807     903   4.013 0.0181 *
Degree           2   70581   35290 156.765 <2e-16 ***
Size             4  130867   32717 145.332 <2e-16 ***
Residuals     6715 1511661     225
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Main Effects with Interaction

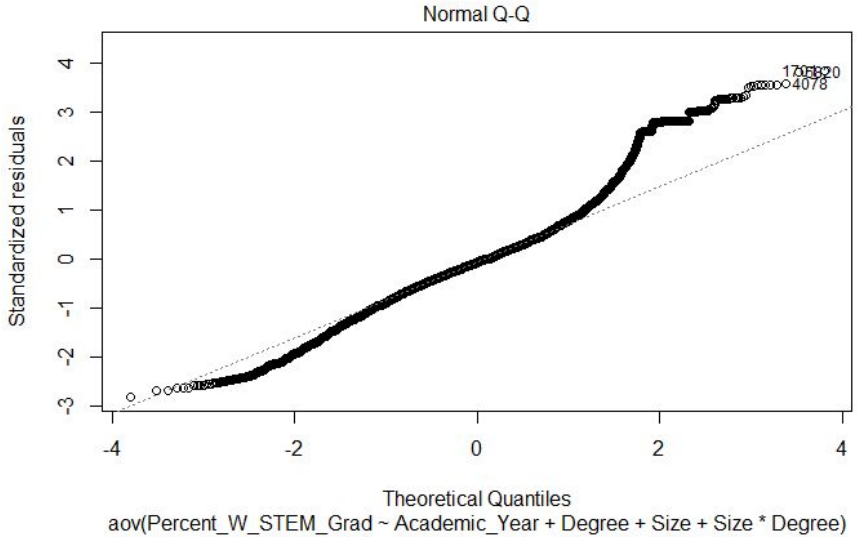
```
                Df  Sum Sq Mean Sq F value   Pr(>F)
Academic_Year    2    1807     903   4.025 0.017900 *
Degree           2   70581   35290 157.252  < 2e-16 ***
Size             4  130867   32717 145.785  < 2e-16 ***
Degree:Size      8    6483     810   3.611 0.000339 ***
Residuals     6707 1505178     224
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Fit: aov(formula = Percent_W_STEM_Grad ~ Academic_Year + Degree + Size + Size *
Degree, data = wstm)

$Degree
                          diff        lwr        upr     p adj
Masters-Bachelors    -6.221245  -7.377690 -5.064799 0.00e+00
Doctorate-Bachelors -9.476410 -11.048605 -7.904216 0.00e+00
Doctorate-Masters    -3.255166  -4.986769 -1.523562 3.17e-05


$Academic_Year
                diff        lwr      upr     p adj
2009-2001 0.3403896 -0.9197382 1.600517 0.8018176
2018-2001 0.9383620 -0.2849710 2.161695 0.1702536
2018-2009 0.5979724 -0.5957514 1.791696 0.4686077


$Size
                             diff        lwr         upr     p adj
1000-4999-<1000          -2.716281  -5.022173  -0.4103899 0.0114950
5000-9999-<1000          -8.631014 -11.161237  -6.1007910 0.0000000
10000-19999-<1000       -12.262252 -14.781947  -9.7425560 0.0000000
20000+-<1000            -15.624807 -18.140617 -13.1089965 0.0000000
5000-9999-1000-4999      -5.914733  -7.587799  -4.2416660 0.0000000
10000-19999-1000-4999    -9.545970 -11.203073  -7.8888676 0.0000000
20000+-1000-4999        -12.908525 -14.559713 -11.2573367 0.0000000
10000-19999-5000-9999    -3.631237  -5.588501  -1.6739735 0.0000042
20000+-5000-9999         -6.993792  -8.946052  -5.0415330 0.0000000
20000+-10000-19999       -3.362555  -5.301151  -1.4239592 0.0000223
```

Homoscedasticity



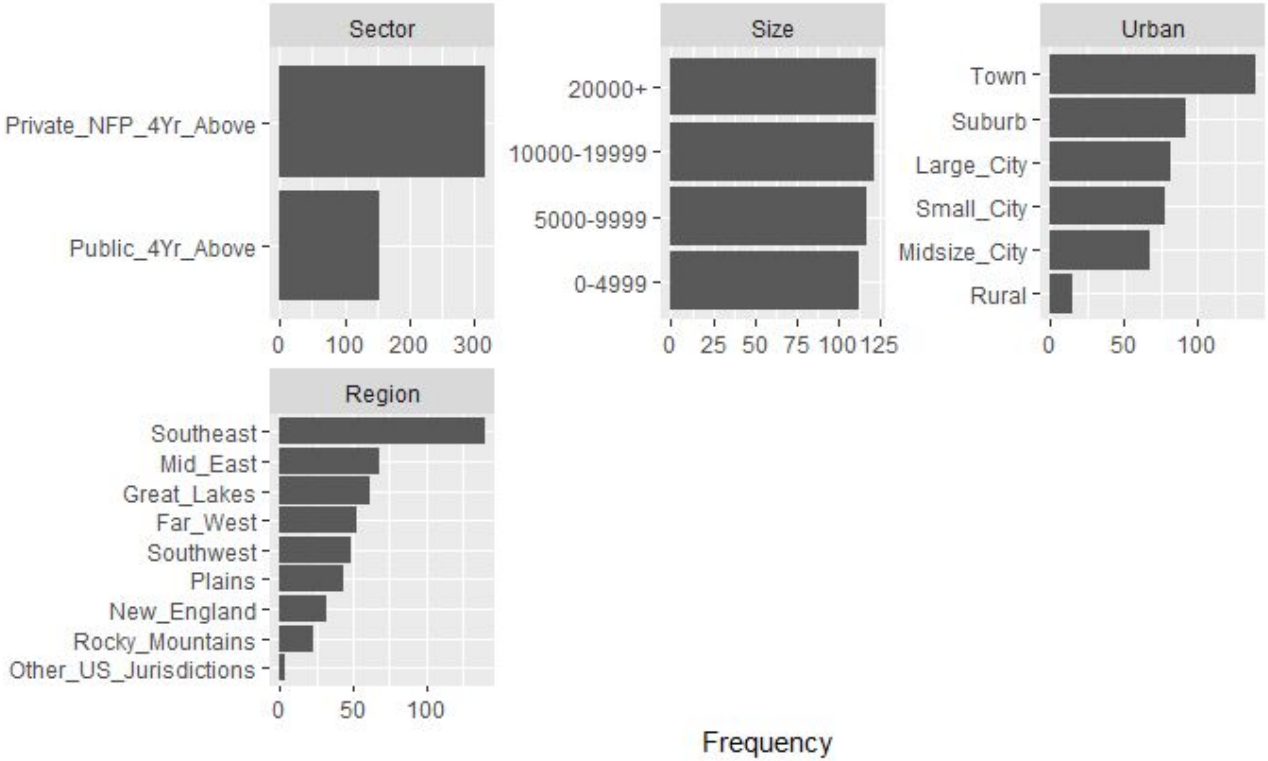
Normal Model



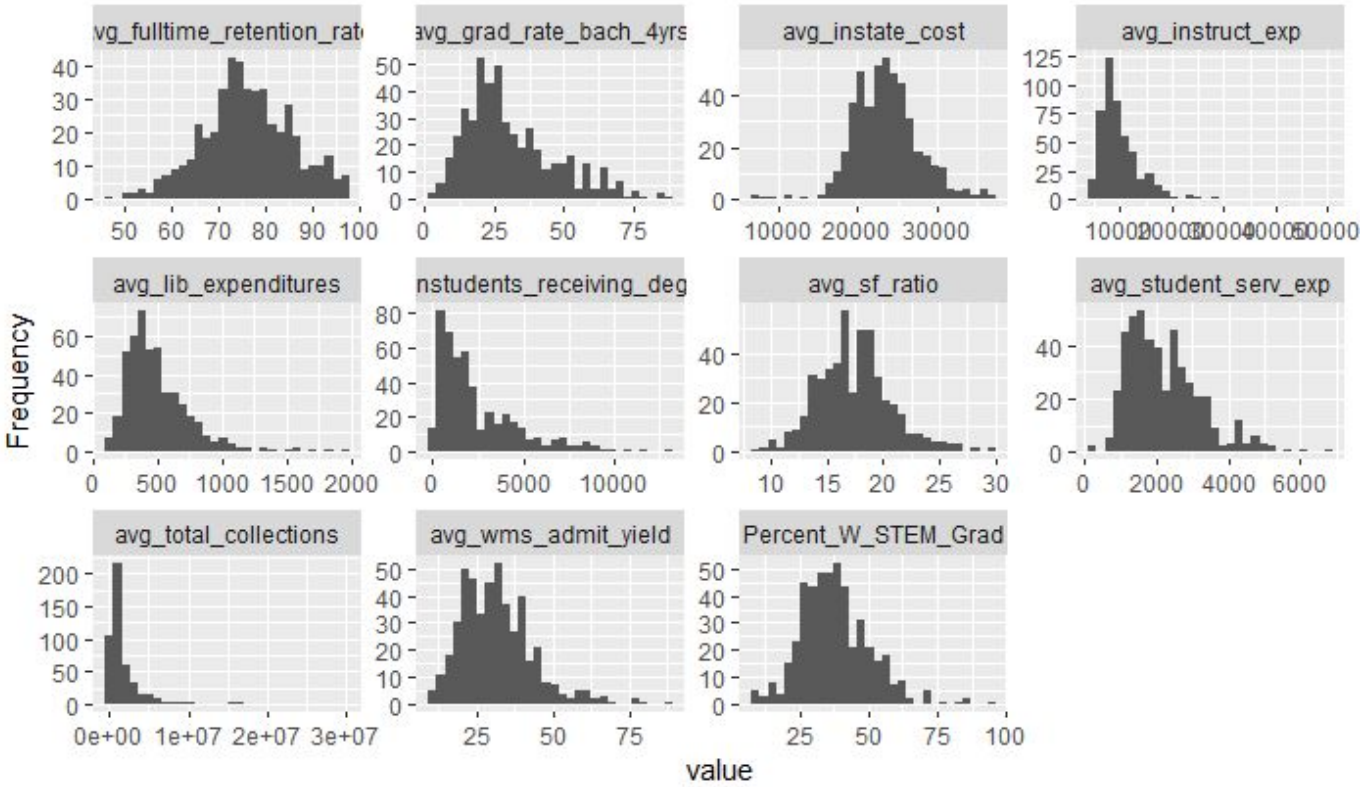
Do STEM completion rates among women vary in relation to extrinsic factors such:

- Campus Location
- Size
- Instructional Expenditures
- Student : Faculty Ratios
- Retention Rates
- Etc…….

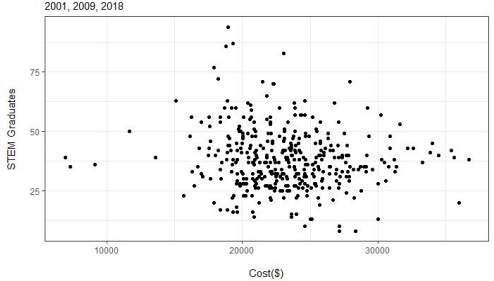
Categorical: Independent Variables



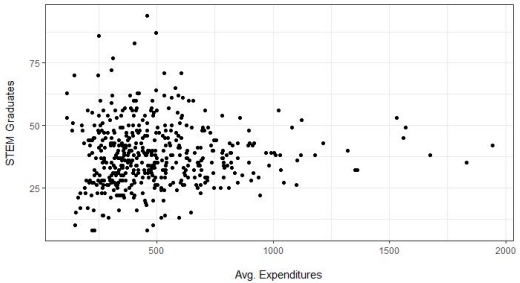
Continuous: Independent Variables



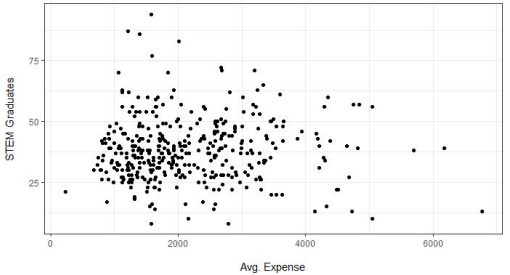
Percent of Women STEM Graduates vs. Instate Cost
2001, 2009, 2018



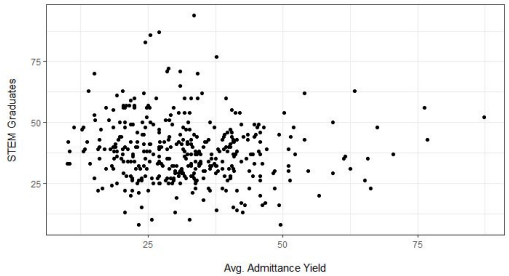
Percent of Women STEM Graduates vs. Avg. Library Expenditures
2001, 2009, 2018

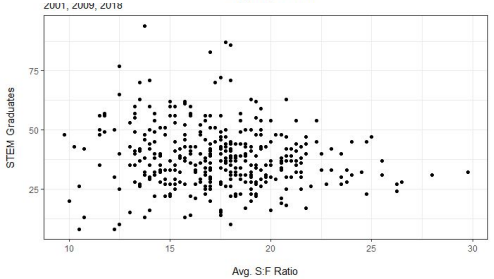


Percent of Women STEM Graduates vs. Avg. Student Service Expense
2001, 2009, 2018

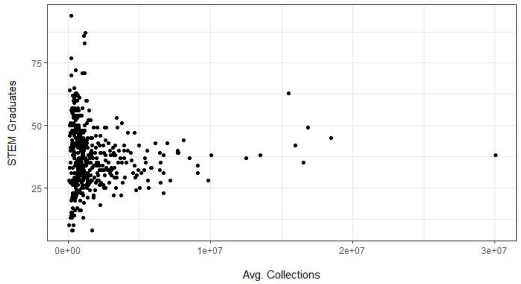


Percent of Women STEM Graduates vs. Avg. Womens Admittance Yield
2001, 2009, 2018

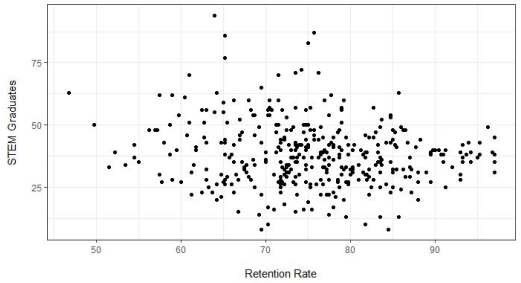




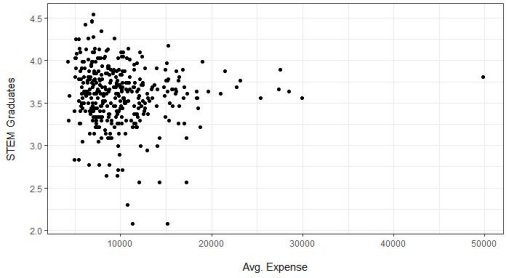
Percent of Women STEM Graduates vs. Avg. Total Library Collections
2001, 2009, 2018



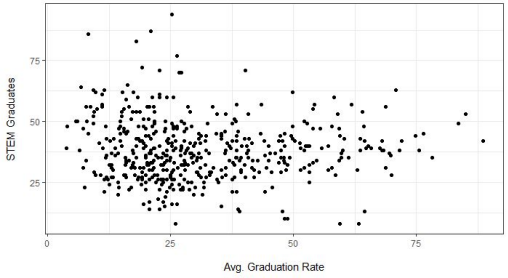
Percent of Women STEM Graduates vs. Avg. Full Time Retention Rate
2001, 2009, 2018



Percent of Women STEM Graduates vs. Avg. Instructional Expense
2001, 2009, 2018



Percent of Women STEM Graduates vs. Avg. 4yr Graduation Rate
2001, 2009, 2018



Variance Inflation Factor $\text{VIF} = \frac{1}{1 - R_i^2}$

```
vif(full)
```

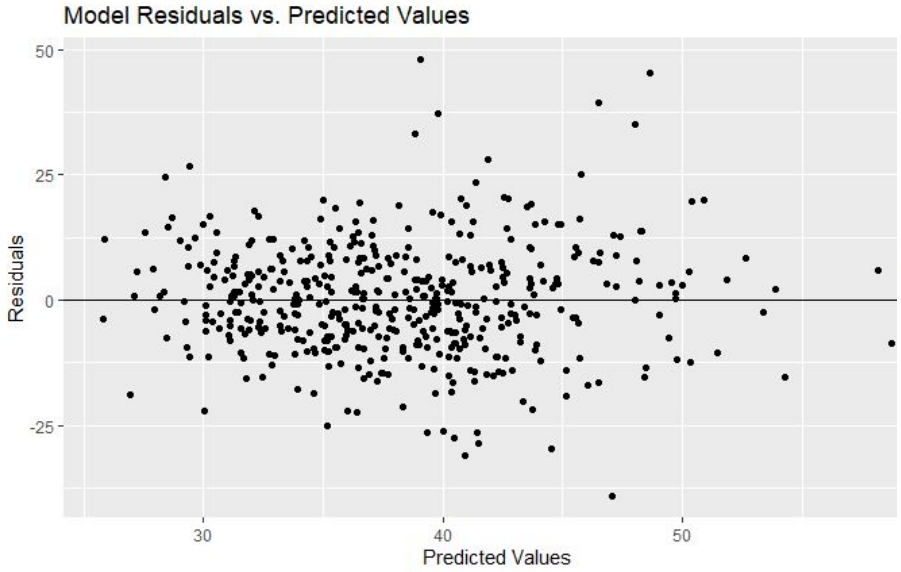
SectorPublic_4Yr_Above	Size0-4999	Size10000-19999
1.130668	6.914135	3.557593
avg_instruct_exp	avg_student_serv_exp	avg_total_collections
2.803585	1.519561	2.387525
avg_fulltime_retention_rate	avg_sf_ratio	UrbanLarge_City
4.565872	2.405782	2.044875
UrbanSuburb	RegionFar_West	RegionGreat_Lakes
1.674227	1.760709	1.461010
RegionPlains	RegionRocky_Mountains	RegionSouthwest
1.302097	1.202760	1.395287

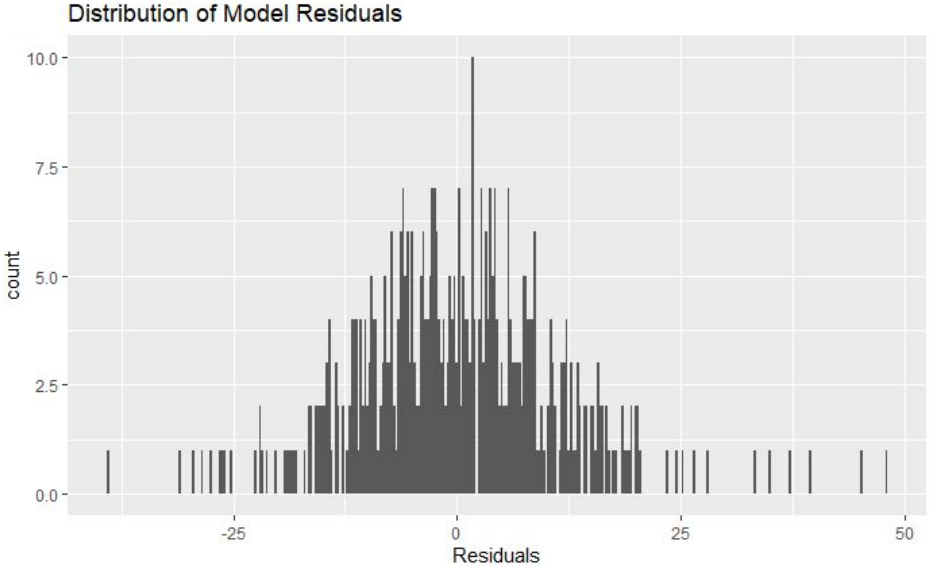
Size5000-9999	avg_grad_rate_bach_4yrs	avg_instate_cost
5.209813	3.951906	2.078146
avg_lib_expenditures	avg_nstudents_receiving_degrees	avg_wms_admit_yield
2.908216	6.244225	1.458504
UrbanMidsize_City	UrbanRural	UrbanSmall_City
1.548694	1.201342	1.578419
RegionMid_East	RegionNew_England	RegionOther_US_Jurisdictions
1.851075	1.472864	1.198965

```
lm(formula = Percent_W_STEM_Grad ~ Size + avg_instate_cost + 
    avg_lib_expenditures + avg_wms_admit_yield + avg_sf_ratio + 
    Region, data = mod2)

Residuals:
    Min      1Q  Median      3Q     Max 
-37.888  -6.968   0.051   6.489  48.537 

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                 4.400e+01  6.791e+00   6.480 2.38e-10 ***
Size0-4999                  9.840e+00  2.000e+00   4.920 1.21e-06 ***
Size10000-19999             2.549e+00  1.636e+00   1.558 0.119931    
Size5000-9999               7.217e+00  1.757e+00   4.107 4.75e-05 ***
avg_instate_cost           -4.112e-04  1.624e-04  -2.532 0.011667 *  
avg_lib_expenditures        1.021e-02  2.727e-03   3.745 0.000204 ***
avg_wms_admit_yield        -2.120e-01  5.171e-02  -4.101 4.88e-05 ***
avg_sf_ratio                3.211e-01  2.181e-01   1.472 0.141655    
RegionFar_West             -6.333e+00  2.021e+00  -3.133 0.001840 ** 
RegionGreat_Lakes          -8.711e+00  1.782e+00  -4.889 1.41e-06 ***
RegionMid_East             -4.935e+00  1.856e+00  -2.659 0.008115 ** 
RegionNew_England          -8.917e+00  2.350e+00  -3.795 0.000168 ***
RegionOther_US_Jurisdictions 1.140e+01  6.984e+00   1.633 0.103240    
RegionPlains               -8.921e+00  2.014e+00  -4.430 1.18e-05 ***
RegionRocky_Mountains      -1.050e+01  2.621e+00  -4.004 7.26e-05 ***
RegionSouthwest            -5.182e+00  1.934e+00  -2.679 0.007645 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.43 on 456 degrees of freedom
Multiple R-squared:  0.1961,    Adjusted R-squared:  0.1696 
F-statistic: 7.415 on 15 and 456 DF,  p-value: 8.735e-15
```

Percent_W_STEM_Grads

+ 9.84 * $size_{0\text{-}4999}$
+ 7.22 * $size_{5000\text{-}9999}$
+ 1.02 * avg_lib_expenditures
+ 2.21 * avg_wmns_admit_yield
- 4.11 * avg_instate_cost
- 8.71 * Great_Lakes
- 4.94 * Mid_East (Atlantic)
- 9.92 * New_England
- 9.92 * Plains
- 1.05 * Rocky_Mts
- 5.18 * Southwest


Model Residuals vs. Predicted Values
Residuals
50
25
0
-25
30
40
50
Predicted Values


Distribution of Model Residuals
count
10.0
7.5
5.0
2.5
0.0
-25
0
25
50
Residuals

Percent of Women Completing STEM Degrees

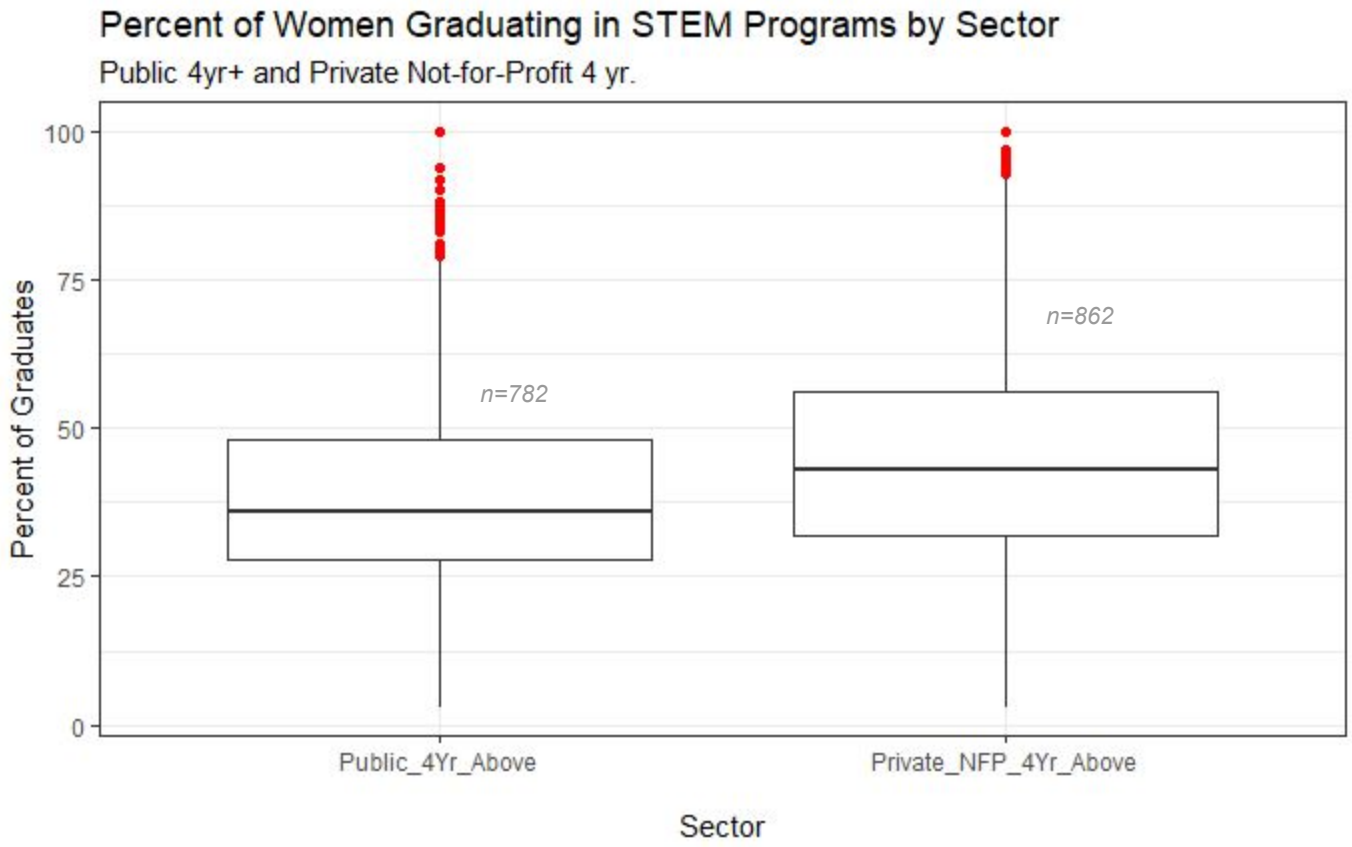
Increased ~ 9% in the past decade.

Decreased ~ 9.5% from Bachelor -to- Doctorate

Decreased ~ 16% from enrollments of 5,000 or less -to- 20000+

Smaller Institutions
Higher Enrollment Yields (W)
Southeastern location

Lower cost of study
Higher Library Investments


Percent of Women Graduating in STEM Programs by Sector
Public 4yr+ and Private Not-for-Profit 4 yr.
Percent of Graduates
100
75
50
25
0
n=782
n=862
Public_4Yr_Above
Private_NFP_4Yr_Above
Sector