

A decorative graphic on the left side of the slide, consisting of a network of white lines and small circles on a dark blue background, resembling a circuit board or a neural network.

EXPLORING COLLEGE.CSV

BY DANIEL GILLINGER & BRIAN FREISS

DATABASE: COLLEGE.CSV

Name	Private	Apps	Accept	Enroll	Top10perc	Top25perc	F.Undergrad	P.Undergrad	Outstate	Room.Board	Books	Personal	PhD	Terminal	S.F.Ratio	perc.alumni	Expend	Grad.Rate
Abilene Christian University	Yes	1660	1232	721	23	52	2885	537	7440	3300	450	2200	70	78	18.1	12	7041	60
Adelphi University	Yes	2186	1924	512	16	29	2683	1227	12280	6450	750	1500	29	30	12.2	16	10527	56
Adrian College	Yes	1428	1097	336	22	50	1036	99	11250	3750	400	1165	53	66	12.9	30	8735	54
Agnes Scott College	Yes	417	349	137	60	89	510	63	12960	5450	450	875	92	97	7.7	37	19016	59
Alaska Pacific University	Yes	193	146	55	16	44	249	869	7560	4120	800	1500	76	72	11.9	2	10922	15
Albertson College	Yes	587	479	158	38	62	678	41	13500	3335	500	675	67	73	9.4	11	9727	55
Albertus Magnus College	Yes	353	340	103	17	45	416	230	13290	5720	500	1500	90	93	11.5	26	8861	63
Albion College	Yes	1899	1720	489	37	68	1594	32	13868	4826	450	850	89	100	13.7	37	11487	73
Albright College	Yes	1038	839	227	30	63	973	306	15595	4400	300	500	79	84	11.3	23	11644	80
Alderson-Broaddus College	Yes	582	498	172	21	44	799	78	10468	3380	660	1800	40	41	11.5	15	8991	52
Alfred University	Yes	1732	1425	472	37	75	1830	110	16548	5406	500	600	82	88	11.3	31	10932	73

- 18 Variables
 - Index: Name
- 777 United States Colleges

FIRST QUESTION & HYPOTHESIS

- Q: How do college's out-of-state tuition affect graduation rates?
 - We created a sub-table to find out.
- H: The higher the out-of-state tuition, the higher the graduation rate.

```
Outstate_vs_Gradrate=College[['Outstate', 'Grad.Rate']]  
Outstate_vs_Gradrate
```

	Outstate	Grad.Rate
0	7440	60
1	12280	56
2	11250	54
3	12960	59
4	7560	15
...
772	6797	40
773	11520	83
774	6900	49
775	19840	99
776	4990	99

777 rows × 2 columns

INITIAL DATA EXPLORATION

- Just some data we thought would be interesting
- Grad rate max is a problem

```
grad_rate=College['Grad.Rate']  
grad_rate.mean()
```

65.46332046332046

```
Student_Faculty=College['S.F.Ratio']  
Student_Faculty.mean()
```

14.089703989703986

```
Student_Faculty.min()
```

2.5

```
Student_Faculty.max()
```

39.8

```
grad_rate.min()
```

10

```
grad_rate.max()
```

118

FIXING GRAD_RATE

```
def grad_cutoff_at_100(rate):  
    return min(rate, 100)  
College['Grad.Rate']=College['Grad.Rate'].apply(grad_cutoff_at_100)
```

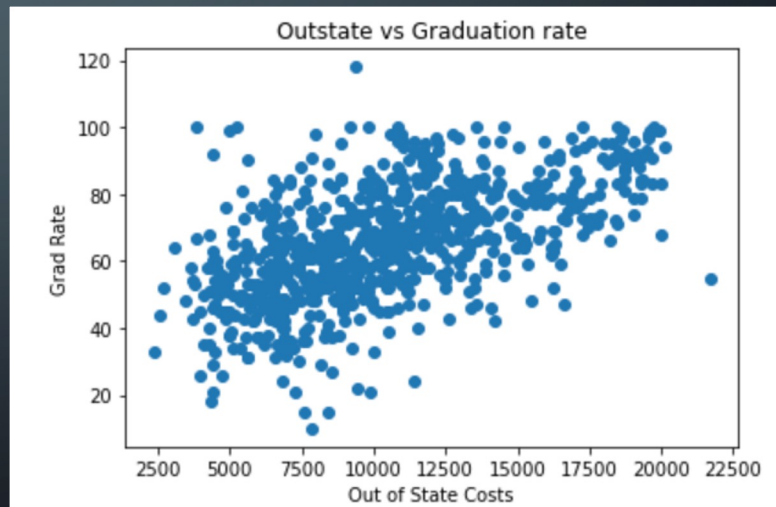
```
grad_rate.max()
```

```
100
```

It's all good now

FIRST QUESTION ANSWERED

- Out-of-State Tuition & Graduation cost
 - Positive Correlation
 - Correlation Coefficient = .57 (Moderate Positive)



EXPANDED QUESTION & HYPOTHESIS

- Will all cost variables give us a stronger correlation with Grad Rate?
- New row for the total costs

```
College['Total Costs']=College['Outstate']+College['Room.Board']+College['Books']+College['Personal']  
College['Total Costs']
```

```
0      13390  
1      20980  
2      16565  
3      19735  
4      13980  
...  
772    12397  
773    18330  
774    12498  
775    29095  
776    10300
```


EXPANDED QUESTIONS

- This was not the case
- Slightly weaker correlation

```
np.corrcoef(College['Total Costs'],College['Grad.Rate'])  
  
array([[1.          , 0.55229142],  
       [0.55229142, 1.          ]])
```


EXPANDED QUESTIONS

- Possibly one negative correlation
- A smaller S.F. Ratio is one of DeSales' selling points after all

```
np.corrcoef(College['S.F.Ratio'],College['Grad.Rate'])  
  
array([[ 1.          , -0.30671041],  
       [-0.30671041,  1.          ]])
```

EXPANDED QUESTIONS

- Is there a difference between public and private schools in terms of graduation rates?

```
where_warriors = (College["Private"].str.contains("Yes"))  
private=College[where_warriors ]  
#all privates are rounded up
```

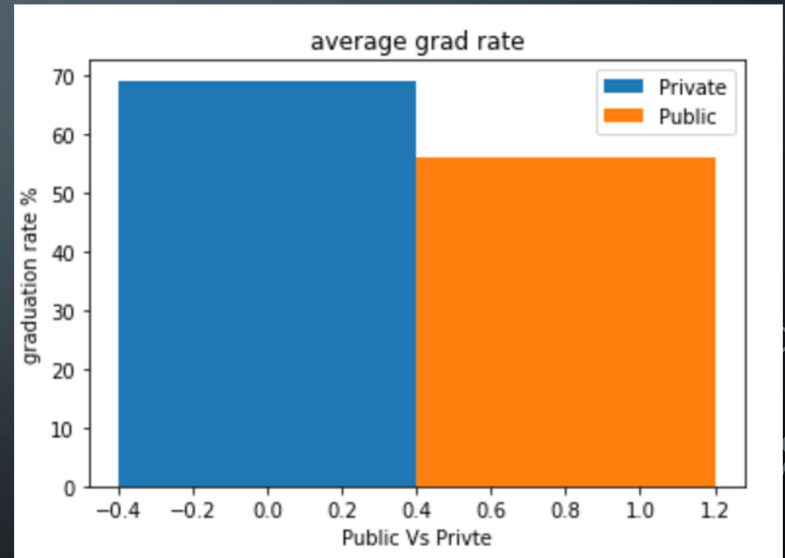
```
where_warrior = (College["Private"].str.contains("No"))  
public=College[where_warrior ]  
#all publics are rounded up
```

```
a=public['Grad.Rate'].sum()/212  
#getting average
```

```
b=private['Grad.Rate'].sum()/565  
#getting average
```

EXPANDED QUESTIONS

- As you can see private schools outperform public by around 10%



BIGGEST CHALLENGE/DREAM DATA SET

- Biggest Challenge
 - Column name meaning & measurement
- Dream Data:
 - Instate costs
 - Success after graduation
 - i.e. mean income, careers, etc.
 - Continued education
 - MBA, PhD, etc.



THANKS FOR LISTENING!

Questions, comments, concerns?

