WORLD UNIVERSITY RANKING

BEHIND THE SCENES

DIEGO A. PETTOROSSI (MSDA '22)

OBJECTIVE

- Ranking system methodology

- Top ranked universities overview

- Myths and stereotypes validation

- What features really matter?

- UTSA vs the WORLD

METHODOLOGY

| RANKING INDICATOR	WEIGHT
Global research reputation	12.5%
Regional research reputation	12.5%
Publications	10%
Books	2.5%
Conferences	2.5%
Normalized citation impact	10%
Total citations	7.5%
Number of publications among top 10% most cited	12.5%
% of publications among top 10% most cited	10%
International collaboration- relative to country	5%
International collaboration	5%

HOW THE DATASET LOOKS LIKE?

	year	rank_display	score	country	city	region	type	research_output	student_faculty_ratio	international_students	size	faculty_count
	<int>	<chr>	<dbl>	<chr>	<chr>	<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<chr>
1	2017	1	100.0	United States	Cambridge	North America	Private	Very High	4	3,730	M	3,065
2	2017	2	98.7	United States	Stanford	North America	Private	Very High	3	3,879	L	4,725
3	2017	3	98.3	United States	Cambridge	North America	Private	Very High	5	5,877	L	4,646
4	2017	4	97.2	United Kingdom	Cambridge	Europe	Public	Very high	4	7,925	L	5,800
5	2017	5	96.9	United States	Pasadena	North America	Private	Very High	2	692	S	968
6	2017	6	96.8	United Kingdom	Oxford	Europe	Public	Very High	3	8,442	L	6,708
7	2017	7	95.6	United Kingdom	London	Europe	Public	Very High	5	21,824	XL	7,195
8	2017	8	94.2	Switzerland	Zürich	Europe	Public	Very High	7	7,733	L	2,719
9	2017	9	94.1	United Kingdom	London	Europe	Public	Very High	5	11,143	L	8,000
10	2017	10	93.0	United States	Chicago	North America	Private	Very High	6	4,696	L	2,703

PUBLIC OR PRIVATE? ASIA OR AMERICA?
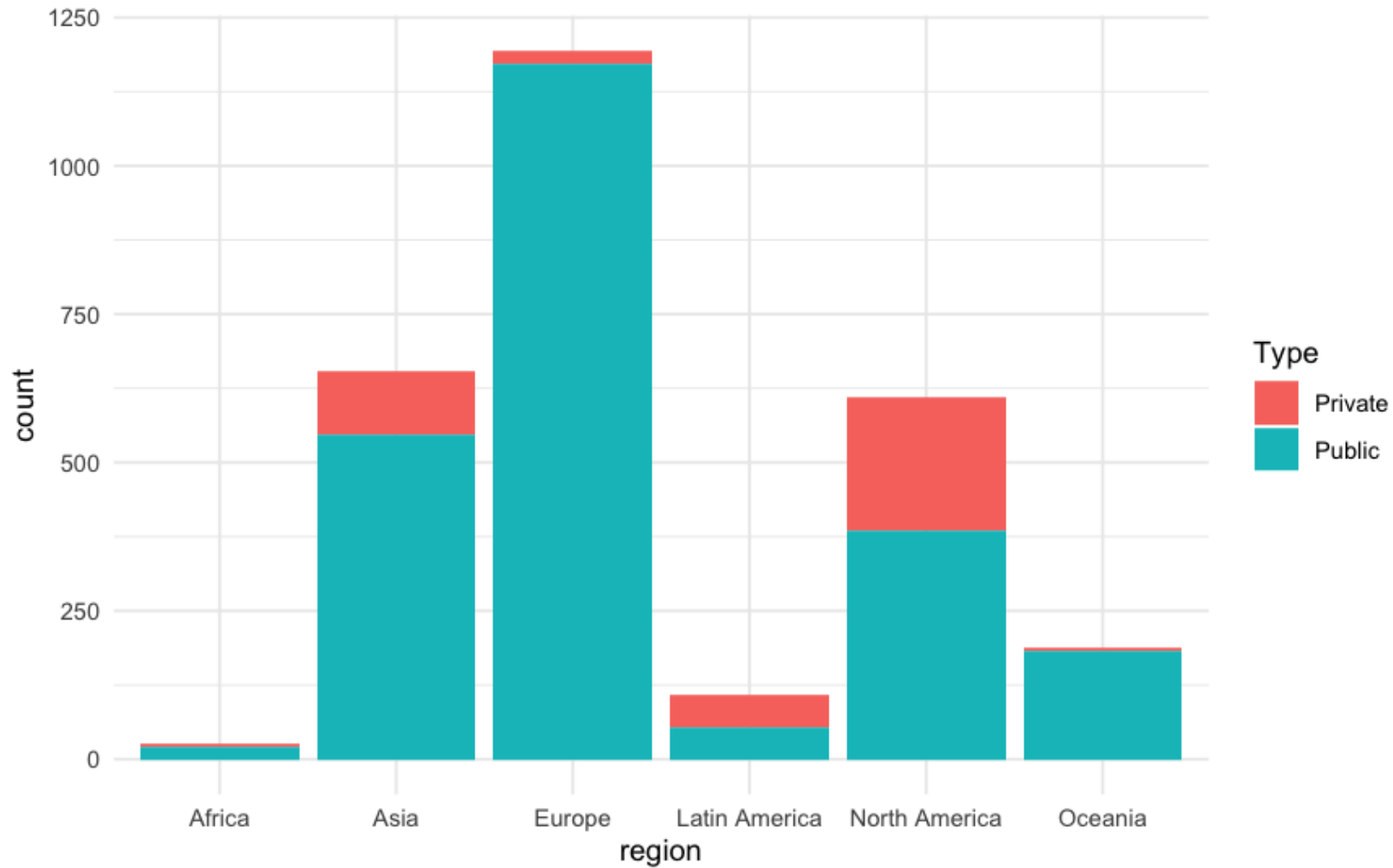
DOES SIZE MATTER?
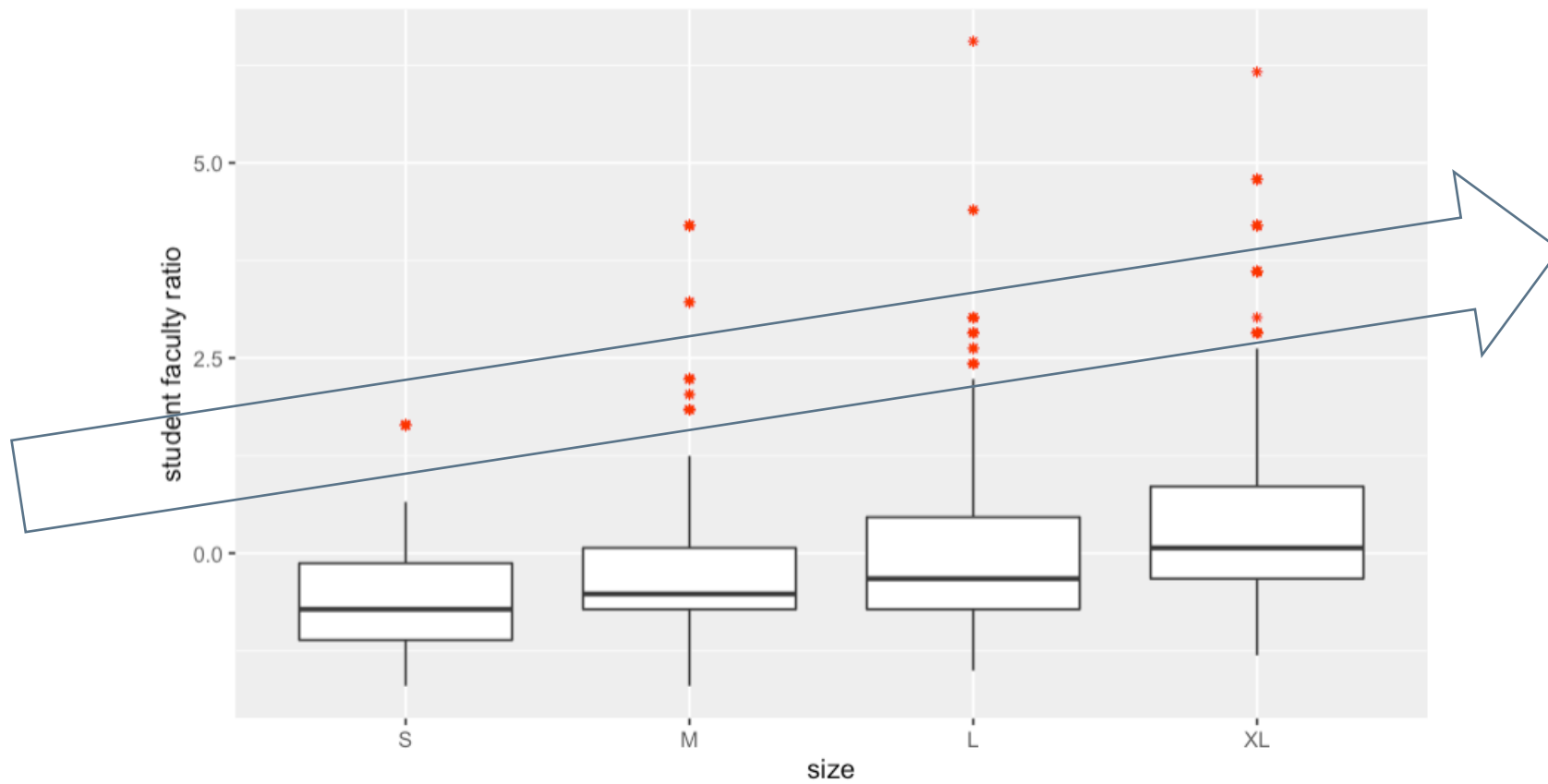
WHAT VARIABLES HAVE AN IMPACT ON THE RANKING SCORE?

|	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
region	5	35573	7115	35.226	< 2e-16	***
type	1	4636	4636	22.952	1.75e-06	***
research_output	3	50324	16775	83.055	< 2e-16	***
student_faculty_ratio	1	109039	109039	539.873	< 2e-16	***
international_students	1	213790	213790	1058.519	< 2e-16	***
size	3	2533	844	4.181	0.00581	**
faculty_count	1	6763	6763	33.483	8.00e-09	***
Residuals	2764	558248	202			

WHAT'S IMPORTANT?
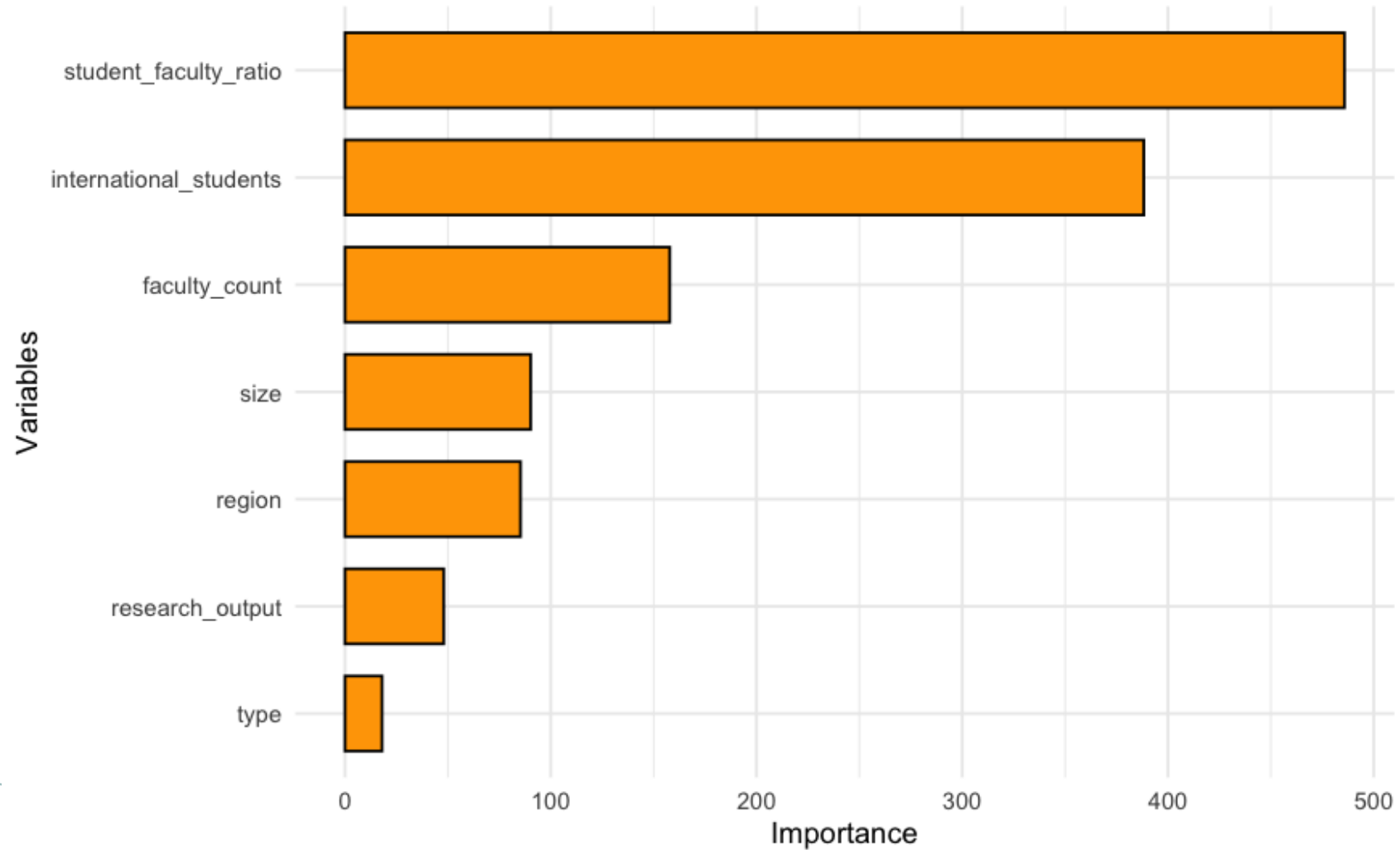
MODEL SELECTION

	ORDINAL LOGISTIC REGRESSION	LINEAR REGRESSION
INTERPRETABILITY (1-5)	3	5
COMPLEXITY	COMPLEX	EASY
ASSUMPTIONS	PROPORTIONAL ODDS, MULTICOLLINEARITY	NORMALITY, HOMOSCEDASTICITY, LINEARITY
HYPERPARAMETERS	NOT REQUIRED	NOT REQUIRED

ASSUMPTIONS CHECK
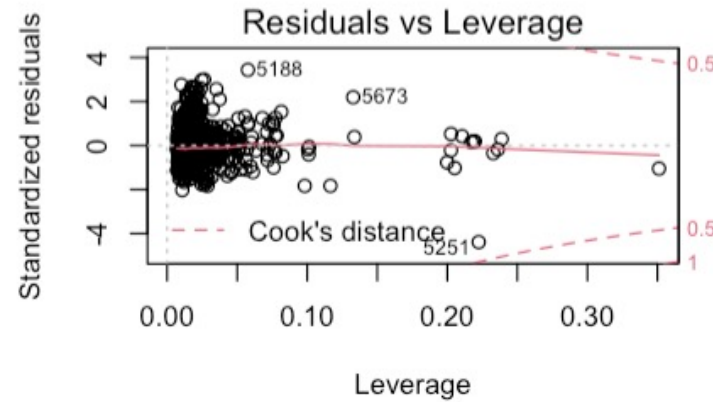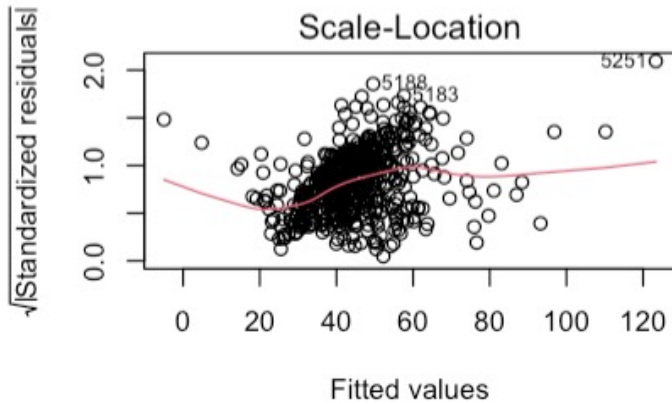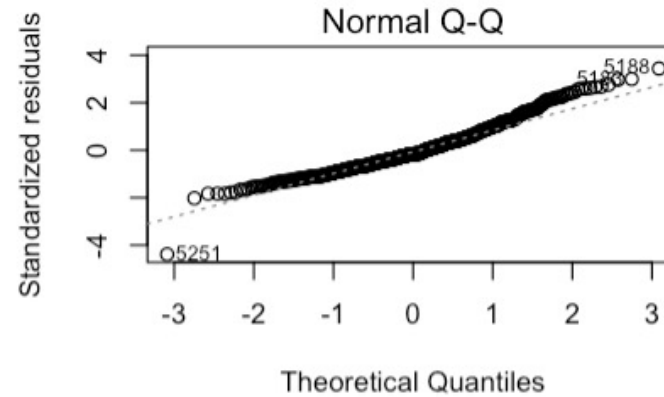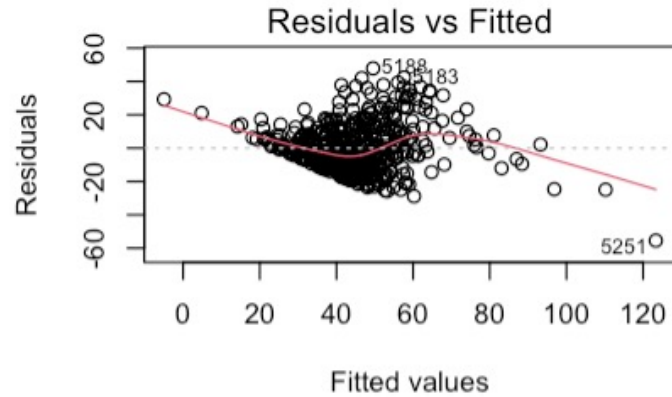
	GVIF	Df	GVIF^(1/(2*Df))
research_output	1.087825	3	1.014129
student_faculty_ratio	1.505190	1	1.226862
size	1.951502	3	1.117879
faculty_count	1.864635	1	1.365516

LINEAR MODEL

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                309.926     26.429  11.727  < 2e-16 ***
research_outputlow          95.293    145.688   0.654 0.513438
research_outputmedium       57.588     68.289   0.843 0.399579
research_outputvery high   -60.941     27.203  -2.240 0.025641 *
student_faculty_ratio       23.531      7.904   2.977 0.003090 **
sizeM                       57.056     21.010   2.716 0.006909 **
sizeS                       55.294     36.390   1.519 0.129455
sizeXL                     -47.296     19.010  -2.488 0.013264 *
faculty_count               34.475     10.015   3.442 0.000639 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 139.8 on 390 degrees of freedom
Multiple R-squared:  0.06194,   Adjusted R-squared:  0.04269
F-statistic: 3.219 on 8 and 390 DF,  p-value: 0.00148
```

UTSA VS TOP 500 UNIVERSITIES

```
      year        rank_display    score                    country        city                    region
 Min.   :2022    494    : 8   Min.   : 24.10   United States   : 87   Length:499          Africa          : 5
 1st Qu.:2022    414    : 7   1st Qu.: 29.60   United Kingdom  : 49   Class :character    Asia            :123
 Median :2022    334    : 6   Median : 38.70   Germany         : 31   Mode  :character    Europe          :212
 Mean   :2022    290    : 5   Mean   : 44.83   Australia       : 26                       Latin America: 21
 3rd Qu.:2022    112    : 4   3rd Qu.: 55.60   China (Mainland): 26                       North America:104
 Max.   :2022    226    : 4   Max.   :100.00   Canada          : 17                       Oceania         : 34
                 (Other):465                   (Other)         :263
      type         research_output student_faculty_ratio.V1 international_students.V1 size
 Private: 77    high     : 38   Min.   :-1.703193      Min.   :-1.087421        L :242
 Public :422    low      :  1   1st Qu.:-0.719577      1st Qu.:-0.683258        M : 93
                medium   :  5   Median :-0.129408      Median :-0.273067        S : 25
                very high:455   Mean   : 0.044449      Mean   :-0.036193        XL:139
                                3rd Qu.: 0.460761      3rd Qu.: 0.318466
                                Max.   : 6.559176      Max.   : 6.398749

  faculty_count.V1        rankUS
 Min.   :-1.205747   Min.   :  1.0
 1st Qu.:-0.684749   1st Qu.:125.5
 Median :-0.233919   Median :249.5
 Mean   :-0.030244   Mean   :250.0
 3rd Qu.: 0.375040   3rd Qu.:374.5
 Max.   : 8.883923   Max.   :495.5

> UTSA
      year rank_display score      country       city        region   type research_output
6154  2022          801     0 United States San Antonio North America Public       very high
      student_faculty_ratio international_students size faculty_count
6154               2.231269            -0.9017377    L    -0.7035442
```

STRENGTHS	OPPORTUNITY TO GROW
RESEARCH OUTPUT	STUDENT - FACULTY RATIO
LOCATION	INTERNATIONAL STUDENTS
TYPE	FACULTY COUNT
	SIZE

CONCLUSIONS

- The ranks cannot be accurately be predicted using the data provided by the QS

- Linear regressions overperforms Ordinal Logistic Regression in accuracy and interpretability

- 91% of the top 500 university worldwide has a "very high" research output

- Large campuses with low student-faculty ratio AND faculty count appears to be associated with better ranks.

RECOMMENDATIONS

- Collect more data about the universities, especially regarding:
  - Publications (text mining)
  - Research reputation (sentiment analysis)

- Try different models to predict ranking score, which is highly related to ranking