

RateMyProfessor.com: Students' Paradise, Professors' Nightmare?

Colleen Su, Derick Yang



Hypothesis

1. Students use Rate My Professor to either rave or rant about their professors.
2. Students' ratings reach a higher consensus when the professor's teaching quality is very high or very low; ratings have more variance when the professor is mediocre.
3. The level of difficulty of the class impacts the overall quality rating of the professor.

Data Collection

- Use the “Developer Tools” to access the hidden URL that is used when reload the page.

The image shows a web browser window with a course page on the left and the Chrome Developer Tools network tab on the right.

Course Page (Left):

- Header: RMP (Rate My Professors)
- Search bar: "Search ..."
- Advertisement: "FREE TO JOIN WHEELS WHEN YOU WANT THEM" with a Zipcar logo and a "Sign up" button.
- Text: "Start typing your comment..." and "350 characters left"
- Button: "CONTINUE YOUR RATING"
- Date: "04/07/2019"
- Rating: "AWESOME" (5 stars)
- Course ID: "PSYC110"
- Details: "For Credit: Yes", "Attendance: Not Mandatory", "Textbook Used: Yes", "Would Take Again: Yes", "Grade Received: Not sure yet"
- Tags: "GRADED BY FEW THINGS", "AMAZING LECTURES", "HILARIOUS"
- Description: "Kazama is a phenomenal lecturer whose PowerPoints add a lot to the lecture experience visually. He's great at explaining things in a way that will help the information stick in your mind, and he can be very funny. His tests can be difficult and considering they're your only grades you definitely need to study, but Kazama is definitely worth taking."
- Feedback: "0 people found this useful" (thumbs up) and "1 person did not find this useful" (thumbs down)
- Button: "report this rating"
- Date: "03/18/2019"

Developer Tools Network Tab (Right):

- Filter: "Hide data URLs All XHR JS CSS Img Media Font Doc WS Manifest Other"
- Timeline: A horizontal timeline showing network activity from 20000 ms to 180000 ms.
- Selected Request: "http://www.ratemyp Professors.com/paginate/professors/ratings?tid=1976448&filter=&courseCode=6page=2"
- Request Method: GET
- Status Code: 200 OK
- Remote Address: 199.232.34.217:80
- Referrer Policy: no-referrer-when-downgrade
- Response Headers: "Accept-Ranges: bytes", "Age: 0", "Connection: keep-alive", "Content-Encoding: gzip", "Content-Length: 3602", "Content-Type: application/json; charset=UTF-8", "Date: Wed, 24 Apr 2019 04:18:48 GMT", "Edge-control: dca=esi, lno-store, max-age=0m", "Server: Apache-Coyote/1.1", "Vary: Accept-Encoding", "Via: 1.1 varnish", "X-Cache: MISS, MISS", "X-Cache-Hits: 0, 0", "X-Served-By: cache-lad2133-IAD, cache-fty21329-FTY", "X-Timer: S1556079529.657566, V50, VE76"
- Request Headers: "Accept: application/json, text/javascript, */*; q=0.01"

Data Collection

```
{
  "ratings": [
    {
      "attendance": "Not Mandatory",
      "clarityColor": "good",
      "easyColor": "average",
      "helpColor": "good",
      "helpCount": 0,
      "id": 30896209,
      "notHelpCount": 0,
      "onlineClass": "",
      "quality": "awesome",
      "rClarity": 5,
      "rClass": "PSY C110",
      "rComments": "Andy Kazama is an inspirational professor and person. His personal touch to the class and vibrant personality help this class stay at one of Emory's finest and most in-demand",
      "rDate": "12/06/2018",
      "rEasy": 3.0,
      "rEasyString": "3.0",
      "rErrorMsg": null,
      "rHelpful": 5,
      "rInterest": "N/A",
      "rOverall": 5.0,
      "rOverallString": "5.0",
      "rStatus": 1,
      "rTextBookUse": "Yes",
      "rWouldTakeAgain": "Yes",
      "sId": 340,
      "takenForCredit": "Yes",
      "teacher": null,
      "teacherGrade": "A",
      "teacherRatingTags": [
        "Amazing lectures",
        "Inspirational",
        "Caring"
      ],
      "unUsefulGrouping": "people",
      "usefulGrouping": "people"
    },
    {
      "attendance": "Mandatory",
      "clarityColor": "good",
      "easyColor": "good",
      "helpColor": "good",
      "helpCount": 0,
      "id": 30875706,
      "notHelpCount": 0,
      "onlineClass": "",
      "quality": "awesome",
      "rClarity": 5,
      "rClass": "PSYC385",
      "rComments": "I meant to give this class a 4 in difficulty in the previous review not a 1. Lots of time put into it but it isnt super difficult especially as you get used to it",
      "rDate": "12/05/2018",
      "rEasy": 4.0,
      "rEasyString": "4.0",
      "rErrorMsg": null,
      "rHelpful": 5,
      "rInterest": "N/A",
      "rOverall": 5.0,
      "rOverallString": "5.0",
      "rStatus": 1,
      "rTextBookUse": "No",
      "rWouldTakeAgain": "Yes",
      "sId": 340,
      "takenForCredit": "Yes",
      "teacher": null,
      "teacherGrade": "A",
      "teacherRatingTags": [],
      "unUsefulGrouping": "people",
      "usefulGrouping": "people"
    },
    {
      "attendance": "Mandatory",
      "clarityColor": "good",
      "easyColor": "poor",
      "helpColor": "good",
      "helpCount": 0,
      "id": 30873538,
      "notHelpCount": 0,
      "onlineClass": "",
      "quality": "awesome",
      "rClarity": 5,
      "rClass": "PSYC385",
      "rComments": "Fantastic course. I learned so much about PTSD as well as how to read and review scientific papers. This course is super good prep for graduate school, but be ready to commit a decent chunk of time to the course. There are two papers a week where you review scientific articles, but he rewards with easier exams so you will likely get an A or high B",
      "rDate": "12/05/2018",
      "rEasy": 1.0,
      "rEasyString": "1.0",
      "rErrorMsg": null,
      "rHelpful": 5,
      "rInterest": "N/A",
      "rOverall": 5.0,
      "rOverallString": "5.0",
      "rStatus": 1,
      "rTextBookUse": "No",
      "rWouldTakeAgain": "Yes",
      "sId": 340,
      "takenForCredit": "Yes",
      "teacher": null,
      "teacherGrade": "A+",
      "teacherRatingTags": [
        "Caring",
        "Participation matters",
        "So many papers"
      ],
      "unUsefulGrouping": "people",
      "usefulGrouping": "people"
    },
    {
      "attendance": "Not Mandatory",
      "clarityColor": "good",
      "easyColor": "average",
      "helpColor": "good",
      "helpCount": 1,
      "id": 30873005,
      "notHelpCount": 0,
      "onlineClass": "",
      "quality": "awesome",
      "rClarity": 5,
      "rClass": "PSY C110",
      "rComments": "Professor Kazama is absolutely amazing! He manages to make tough, dry subject matter fun, interesting and easy to learn and is extremely knowledgeable about his subject. He genuinely cares for his students and makes every effort to help them. Do take his course even if it isn't for a major or GER. You won't regret it!",
      "rDate": "12/05/2018",
      "rEasy": 3.0,
      "rEasyString": "3.0",
      "rErrorMsg": null,
      "rHelpful": 5,
      "rInterest": "N/A",
      "rOverall": 5.0,
      "rOverallString": "5.0",
      "rStatus": 1,
      "rTextBookUse": "Yes",
      "rWouldTakeAgain": "Yes",
      "sId": 340,
      "takenForCredit": "Yes",
      "teacher": null,
      "teacherGrade": "N/A",
      "teacherRatingTags": [
        "Accessible outside class",
        "Inspirational",
        "Amazing lectures"
      ],
      "unUsefulGrouping": "people",
      "usefulGrouping": "person"
    },
    {
      "attendance": "Not Mandatory",
      "clarityColor": "good",
      "easyColor": "average",
      "helpColor": "good",
      "helpCount": 0,
      "id": 30755992,
      "notHelpCount": 0,
      "onlineClass": "",
      "quality": "awesome",
      "rClarity": 5,
      "rClass": "PSY C110",
      "rComments": "Dr. Kazama is the most amazing professor I have ever had the pleasure to learn from. I guarantee that you will walk out of his class with a whole new perspective about what it means to be human, as well as a new appreciation for the world around you. If you keep up with readings, go to class, and stay engaged, you will do well in the class.",
      "rDate": "11/19/2018",
      "rEasy": 3.0,
      "rEasyString": "3.0",
      "rErrorMsg": null,
      "rHelpful": 5,
      "rInterest": "N/A",
      "rOverall": 5.0,
      "rOverallString": "5.0",
      "rStatus": 1,
      "rTextBookUse": "Yes",
      "rWouldTakeAgain": "Yes",
      "sId": 340,
      "takenForCredit": "Yes",
      "teacher": null,
      "teacherGrade": "A+",
      "teacherRatingTags": [
        "Respected",
        "Skip class? You won't pass."
      ],
      "unUsefulGrouping": "people",
      "usefulGrouping": "people"
    },
    {
      "attendance": "Not Mandatory",
      "clarityColor": "good",
      "easyColor": "good",
      "helpColor": "good",
      "helpCount": 0,
      "id": 30665933,
      "notHelpCount": 0,
      "onlineClass": "",
      "quality": "awesome",
      "rClarity": 5,
      "rClass": "PSYC110",
      "rComments": "Sweet guy, but definitely makes you work for the grade. Well worth taking him and getting to know him due to the fact that he cares so much for his students. 10/10 would recommend.",
      "rDate": "11/06/2018",
      "rEasy": 4.0,
      "rEasyString": "4.0",
      "rErrorMsg": null,
      "rHelpful": 5,
      "rInterest": "N/A",
      "rOverall": 5.0,
      "rOverallString": "5.0",
      "rStatus": 1,
      "rTextBookUse": "Yes",
      "rWouldTakeAgain": "Yes",
      "sId": 340,
      "takenForCredit": "Yes",
      "teacher": null,
      "teacherGrade": "N/A",
      "teacherRatingTags": [
        "Test heavy",
        "Accessible outside class",
        "Graded by few things"
      ],
      "unUsefulGrouping": "people",
      "usefulGrouping": "people"
    },
    {
      "attendance": "Not Mandatory",
      "clarityColor": "good",
      "easyColor": "average",
      "helpColor": "good",
      "helpCount": 0,
      "id": 30896209,
      "notHelpCount": 0,
      "onlineClass": "",
      "quality": "awesome",
      "rClarity": 5,
      "rClass": "PSY C110",
      "rComments": "Andy Kazama is an inspirational professor and person. His personal touch to the class and vibrant personality help this class stay at one of Emory's finest and most in-demand",
      "rDate": "12/06/2018",
      "rEasy": 3.0,
      "rEasyString": "3.0",
      "rErrorMsg": null,
      "rHelpful": 5,
      "rInterest": "N/A",
      "rOverall": 5.0,
      "rOverallString": "5.0",
      "rStatus": 1,
      "rTextBookUse": "Yes",
      "rWouldTakeAgain": "Yes",
      "sId": 340,
      "takenForCredit": "Yes",
      "teacher": null,
      "teacherGrade": "A",
      "teacherRatingTags": [
        "Amazing lectures",
        "Inspirational",
        "Caring"
      ],
      "unUsefulGrouping": "people",
      "usefulGrouping": "people"
    }
  ]
}
```

Data Collection

- Extract variables and comments from the source page into dataframe.]

```
def gender(comments):
    m = 0
    f = 0
    for c in comments:
        if 'He' in c or 'he' in c or 'His' in c or 'his' in c:
            m += 1
        if 'She' in c or 'she' in c or 'Her' in c or 'her' in c:
            f += 1

    gen = []
    if m > f:
        gen = ['M']
    elif m < f:
        gen = ['F']
    else:
        gen = ['N/A']

    return gen

def instructorinfo(ID):
    url = 'https://www.ratemyprofessors.com/ShowRatings.jsp?tid='+ID
    r = requests.get(url)
    html = BeautifulSoup(r.text, 'html.parser')

    name = html.find('h1')
    firstlast = []
    professorName=[]
    university=[]

    college=html.find('a', {"class": "school"}).string.strip()
    university.append(college)

    for span in name.find_all('span'):
        del span['class']
        firstlast.append(span)

    professorName.append(firstlast[0].string.strip()+' '+firstlast[2].string.strip())

    return(professorName, university)

def parse(ID):
    l=[]
    for page in range(15):
        r_sanfran = requests.get("http://www.ratemyprofessors.com/paginate/professors/ratings?tid="+ID+"&",
                                q=r_sanfran["ratings"])
        l.append(q)
```

```
classID=[]
overall=[]
helpfulness=[]
easiness=[]
clarity=[]
quality=[]
TakeAgain=[]
teacherGrade=[]
comment=[]
tag=[]
commentNum=[]

for i in range(1,100):
    for k in range(100):
        try:
            date.append(a[i][k]['rDate'])
            classID.append(a[i][k]['rClass'])
            overall.append(a[i][k]['rOverall'])
            helpfulness.append(a[i][k]['rHelpful'])
            easiness.append(a[i][k]['rEasy'])
            clarity.append(a[i][k]['rClarity'])
            quality.append(a[i][k]['rQuality'])
            TakeAgain.append(a[i][k]['rWouldTakeAgain'])
            teacherGrade.append(a[i][k]['rTeacherGrade'])
            comment.append(a[i][k]['rComments'])
            tag.append(a[i][k]['rTeacherRatingTags'])

        except:
            pass

namelist=instructorinfo(ID)[0]
namelist=namelist*len(date)

genderlist=gender(comment)*len(date)

collegelist=instructorinfo(ID)[1]
collegelist=collegelist*len(date)

Num=int(len(date))
commentNum.append(Num)
commentNum=commentNum*len(date)

pol = []
subj = []
for i in comment:
    sentence = TextBlob(i)
    pol.append(round(sentence.sentiment.polarity,2))
    subj.append(round(sentence.sentiment.subjectivity,2))
```

Data Overview

Total Number of Reviews	3383
Max Number of Reviews	104
Number of Professors	332
Male	276
Female	56

Selection criteria:

- School size
- US News National Ranking
- US News Math Ranking

College	Reviews	Professors	Average Number of Reviews	US News. Math Ranking
Emory University	601	69	9	55
Yale University	42	10	4	9
Stanford University	153	34	5	2
University of Chicago	153	33	5	6
Carnegie Mellon University	665	52	13	32
University of Notre Dame	68	20	3	39
Washington University of St.Louis	924	54	17	34
Tufts University	777	60	13	74

Regression

- Overall quality (1.0 - 5.0)
- Level of difficulty(1.0 -5.0)
- Gender (Dummy)
- Number of Comments
- Colleges (Dummy)

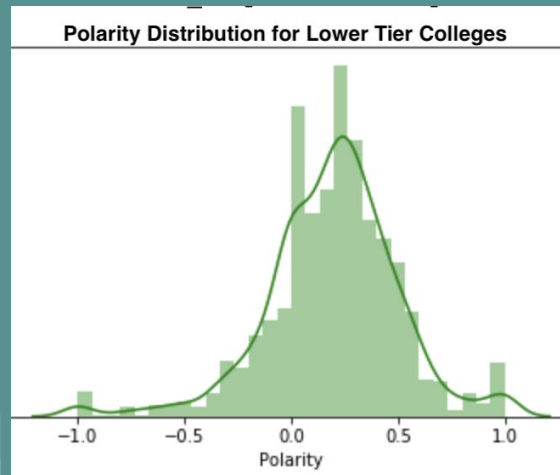
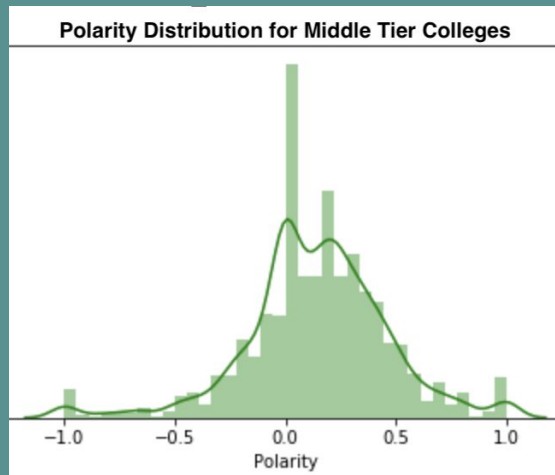
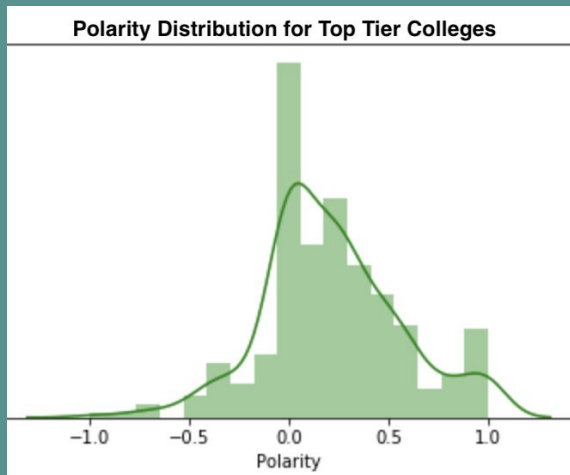
$$\begin{aligned} \text{OverallQualityScore} = & \beta_0 + \beta_1 * \text{LevelofDiff} + \beta_2 * \text{Gender} + \beta_3 * \text{LoD} : \text{Gender} + \beta_4 \\ & * \log(\text{Comment}) + \sigma_1 * \text{Carnegie} + \sigma_2 * \text{Stanford} + \sigma_3 * \text{Tufts} + \sigma_4 * \text{Chicago} + \sigma_5 \\ & * \text{NotreDame} + \sigma_6 * \text{Washington} + \sigma_7 * \text{Yale} \end{aligned}$$

Regression Result

	Estimate	Error	t	P
(Intercept)	4.71920	0.50336	9.375	< 2e-16
Level of Diff	-0.49102	0.14825	-3.312	0.00103
Gender	-1.04415	0.50649	-2.062	0.04006
log(Comment)	0.01312	0.04743	0.277	0.78222
Polarity	3.24916	0.23095	14.069	< 2e-16
Diff: Gender	0.27939	0.15636	1.787	0.07491 .
Carnegie	0.02553	0.13400	0.191	0.84904
Stanford	-0.08421	0.15535	-0.542	0.58812
Tufts	0.11821	0.13001	0.909	0.36392
Chicago	-0.03792	0.15582	-0.243	0.80787
Notre Dame	0.20293	0.18783	1.080	0.28076
Washington	-0.23108	0.13561	-1.704	0.08936 .
Yale	-0.34573	0.24815	-1.393	0.16452

Sentiment Analysis

- TextBlob: a Python Library for processing textual data
 - Polarity (-1,1)
 - Subjectivity (0,1)
 - Accuracy: 84%



One-way ANOVA and Tukey Multiple Pairwise Comparison

```
> summary(ANV_Pol)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
RMP\$Tier_cat	2	2.9	1.4414	13.77	1.1e-06 ***
Residuals	3384	354.2	0.1047		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
4 observations deleted due to missingness

```
> TukeyHSD(ANV_Pol)
```

Tukey multiple comparisons of means
95% family-wise confidence level

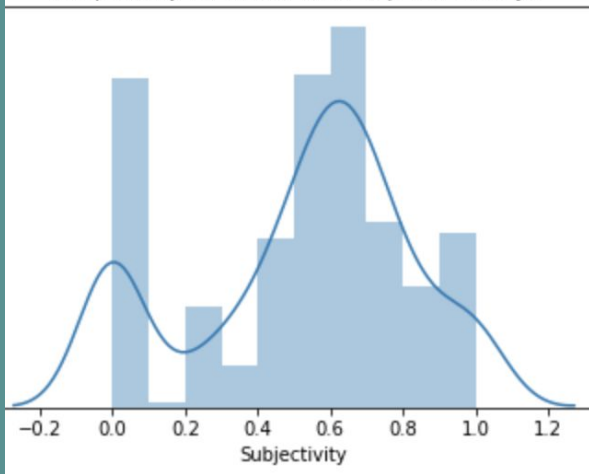
Fit: aov(formula = RMP\$Polarity ~ RMP\$Tier_cat)

\$`RMP\$Tier_cat`

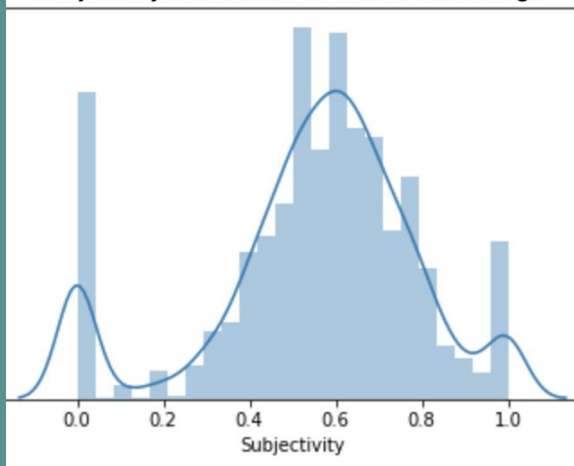
	diff	lwr	upr	p adj
mid-top	-0.07368820	-0.11841022	-0.02896619	0.0003352
low-top	-0.02072554	-0.06623038	0.02477930	0.5341047
low-mid	0.05296266	0.02532430	0.08060103	0.0000216

Sentiment Analysis

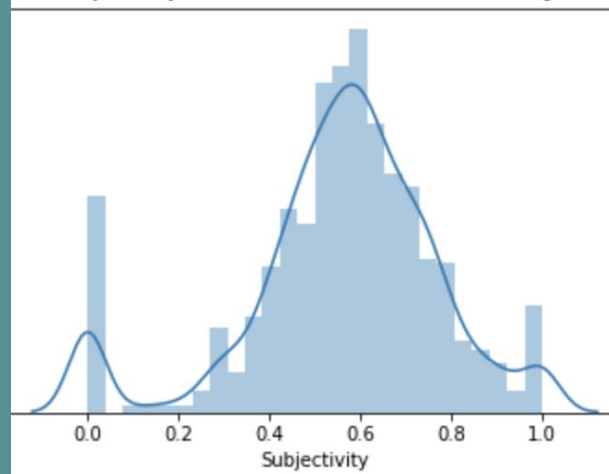
Subjectivity Distribution for Top Tier College



Subjectivity Distribution for Middle Tier College



Subjectivity Distribution for Lower Tier College



Sentiment Analysis

```
> summary(ANV_Sub)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
RMP\$Tier_cat	2	0.38	0.19160	3.374	0.0344 *
Residuals	3384	192.19	0.05679		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> TukeyHSD(ANV_Sub)
```

Tukey multiple comparisons of means
95% family-wise confidence level

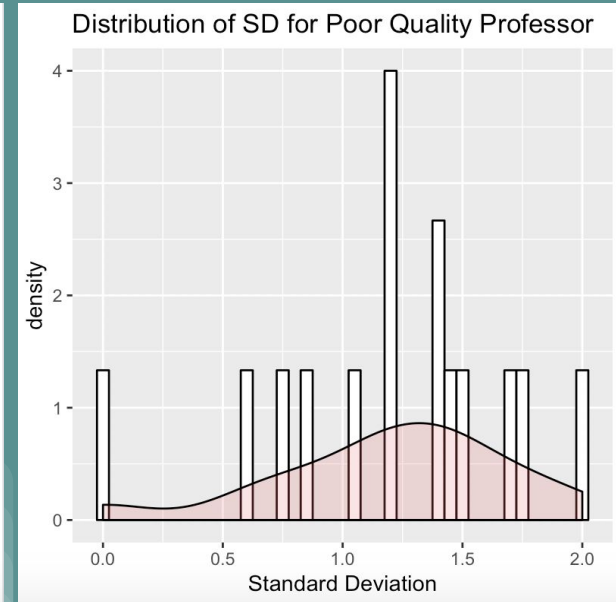
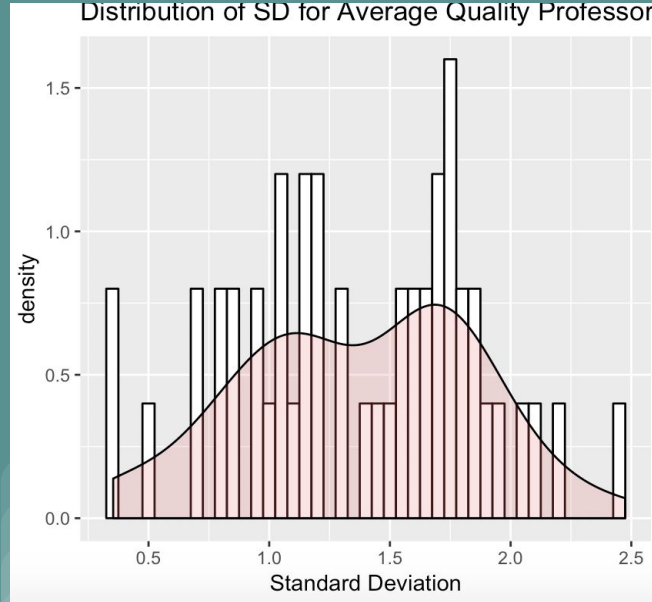
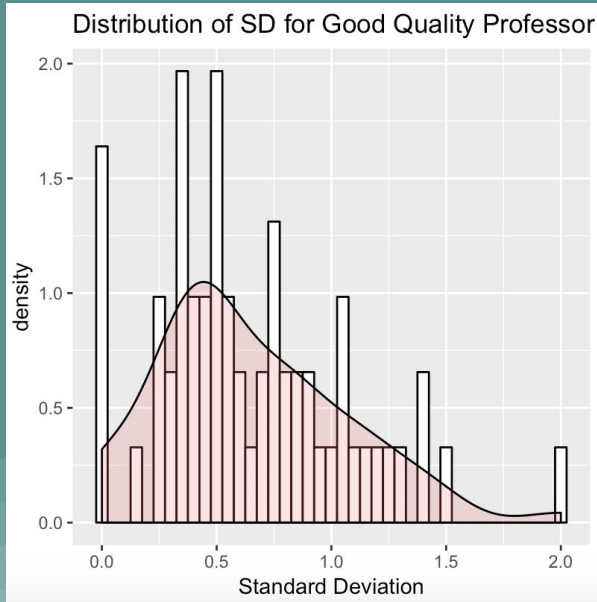
Fit: aov(formula = RMP\$Subjectivity ~ RMP\$Tier_cat)

\$`RMP\$Tier_cat`

	diff	lwr	upr	p adj
mid-top	0.029427365	-0.003516919	0.06237165	0.0911158
low-top	0.037133272	0.003612323	0.07065422	0.0255472
low-mid	0.007705907	-0.012653778	0.02806559	0.6482737

Students' consensus

- Overall Quality Score: (1,5)
- Good: Average Overall Quality Score ≥ 4
- Average: ≥ 2.5 , < 4
- Poor: < 2.5
- High consensus should present low standard deviation of overall quality score



One-way ANOVA and Tukey Multiple Pairwise Comparison

```
> summary(ANOVA)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
D\$Quality	2	17.70	8.850	48.66	2.9e-16 ***
Residuals	122	22.19	0.182		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
4 observations deleted due to missingness

```
> TukeyHSD(ANOVA)
```

Tukey multiple comparisons of means
95% family-wise confidence level

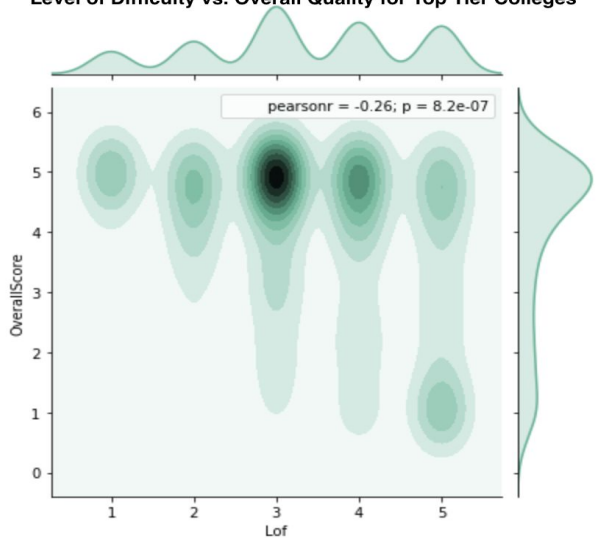
Fit: aov(formula = D\$OverallStd ~ D\$Quality)

\$`D\$Quality`

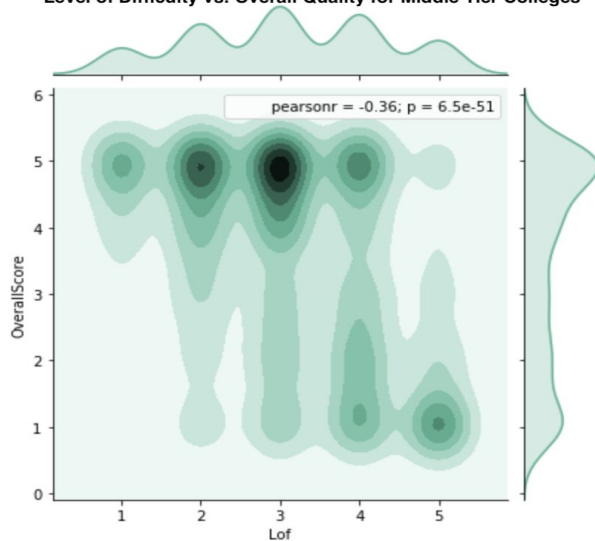
	diff	lwr	upr	p adj
Avrg-Poor	0.06511081	-0.2367120	0.3669336	0.8657254
Good-Poor	-0.70637397	-1.0098454	-0.4029025	0.0000006
Good-Avrg	-0.77148478	-0.9636397	-0.5793299	0.0000000

Overall Quality vs. Level of Difficulty

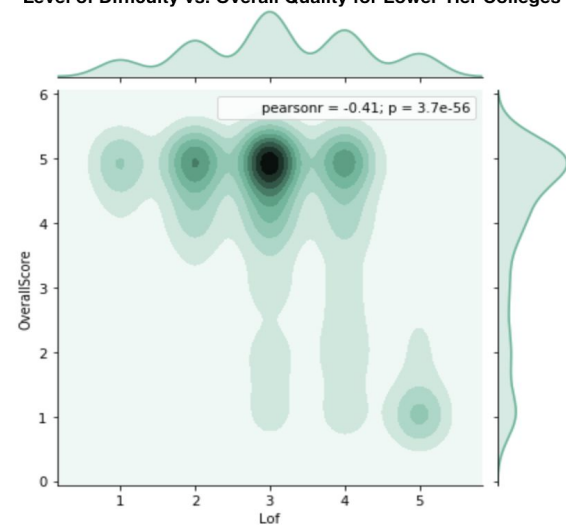
Level of Difficulty vs. Overall Quality for Top Tier Colleges



Level of Difficulty vs. Overall Quality for Middle Tier Colleges



Level of Difficulty vs. Overall Quality for Lower Tier Colleges




```
Call:
lm(formula = top$OverallScore ~ top$Lof)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.5133	-0.9187	0.4867	1.0813	1.6759

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.81059	0.20962	22.949	< 2e-16 ***
top\$Lof	-0.29730	0.05919	-5.022	8.19e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Call:
lm(formula = low$OverallScore ~ low$Lof)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.8990	-0.8759	0.1357	1.1241	2.1473

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.41056	0.10272	52.67	<2e-16 ***
low\$Lof	-0.51156	0.03094	-16.54	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Call:
lm(formula = mid$OverallScore ~ mid$Lof)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.3258	-1.3780	0.1742	1.1220	2.4654

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.77364	0.09595	49.75	<2e-16 ***
mid\$Lof	-0.44780	0.02882	-15.54	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Into the comments: College Correlation

TF-IDF

Number of Reviews	665	601	153	777	153	68	924	42
	CMU	EU	SU	TU	Uchi	UND	WashU	YU
	Greggo(19.0)	QTM (28.0)	Meches(5.0)	Kye(17.0)	Gerry(9.0)	Huy(3.0)	Blank(25.0)	Frame (4.0)
	MacKey(14.0)	Rodl(14.0)	Bray(3.0)	Dyer(13.0)	Brady(5.0)	Huynh(2.0)	Shapiro(18.0)	120(2.0)
	Walker(12.0)	Gong(13.0)	Batteh(3.0)	Nitecki(12.0)	Algorithms(4.0)	Dee(2.0)	webwork(14.0)	Jae jeong(2.0)
	Flaherty(8.0)	Jin(9.0)	White(3.0)	Adler(11.0)	distill(2.0)	Inspired(2.0)	Hasting(12.0)	Expositur(1.0)
	Irina(6.0)	Louie(8.0)	Passion(3.0)	Glaser(10.0)	quarter(2.0)	Polini(2.0)	Shareshian(12.0)	
	Hrusa(6.0)	Jacobson(8.0)	Caltech(2.0)	Garmirian(8.0)	Boller(2.0)	Walsh(2.0)	McCarthy(10.0)	
	Schaeffer(5.0)	Nance(8.0)	Brubaker(2.0)		Sinan(2.0)	Native(1.0)	Spitznagel(9.0)	
		Roth(7.0)	Siegmund(2.0)		Sally(2.0)			
		HATE(7.0)			Babai(2.0)			

Cosine Similarity

US News Ranking		CMU	EU	SU	TU	Uchi	UND	WashU	YU
32	CMU	1	0.98	0.98	0.99	0.97	0.98	0.99	0.96
55	EU	0.98	1	0.96	0.99	0.94	0.97	0.99	0.94
2	SU	0.98	0.96	1	0.97	0.98	0.97	0.97	0.96
74	TU	0.99	0.99	0.97	1	0.96	0.98	0.99	0.95
6	Uchi	0.97	0.94	0.98	0.96	1	0.96	0.96	0.95
39	UND	0.98	0.97	0.97	0.98	0.96	1	0.97	0.95
34	WashU	0.99	0.99	0.97	0.99	0.96	0.97	1	0.95
9	YU	0.96	0.94	0.96	0.95	0.95	0.95	0.95	1

Further Improvement:

- School and sample size selection
- Fixed effect elimination through demean regression
- Eliminate “No Comment” from the dataset to improve the sentiment analysis
- Improve TF-IDF.
- Build vector space model, apply dimension reduction and TSNE plotting to delve into college similarity

For Future Research:

- Students academic experience and school ranking
- Overall quality, polarity, time series and comment bias

Thank You