Flatiron Project 3

Grouping Grade Distributions For Better Class Selection

-By Max and Georgiy

How do you choose your classes?

- Choosing classes that fit your interests
- Choosing classes that fit your schedule
- Choosing classes that won’t hurt your GPA with unfriendly grading curves

We Can Help! (so long as you go to UW-Madison)

- We've compiled a master list of all courses taught 2006-2017!
- We've matched them to their grade distributions!
- We've broken down each course by section (even though UW-Madison's sections table was poorly organized and required cleaning)!
- Our well-designed python formulas allow you to effortlessly compare any two classes or sets of classes to see which has the friendlier grade distribution!
- Let's walk through an example to see how it works:

Subject Comparison - with A ratio

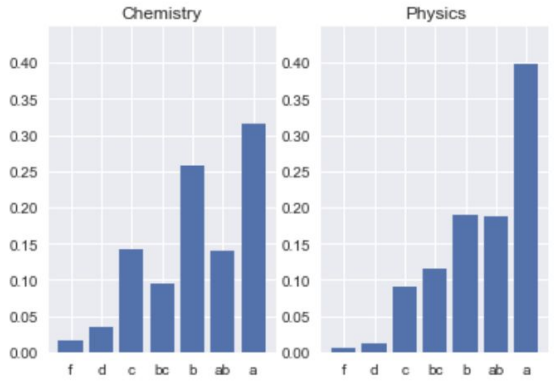
- Let's say you're a liberal arts major who needs to choose a science course to fulfill a requirement
- Do you choose physics or chemistry?
- Luckily, we can help you make that choice with one of two statistical tests!
- Is the ratio of A's different between the two departments: using Fisher's Exact Test
  - Null- Hypothesis - $H_{0:}\ \mu_1 = \mu_2$
  - Alternative Hypothesis $\mu_1 \neq \mu_2$

```
#Build lists of all classes in the relevant departments
chem = list(set(courses[courses['subject']=='Chemistry']['name']))
physics = list(set(courses[courses['subject']=='Physics']['name']))


#Insert into comparison function
subject_graph_compare('Chemistry',chem, 'Physics', physics)
```

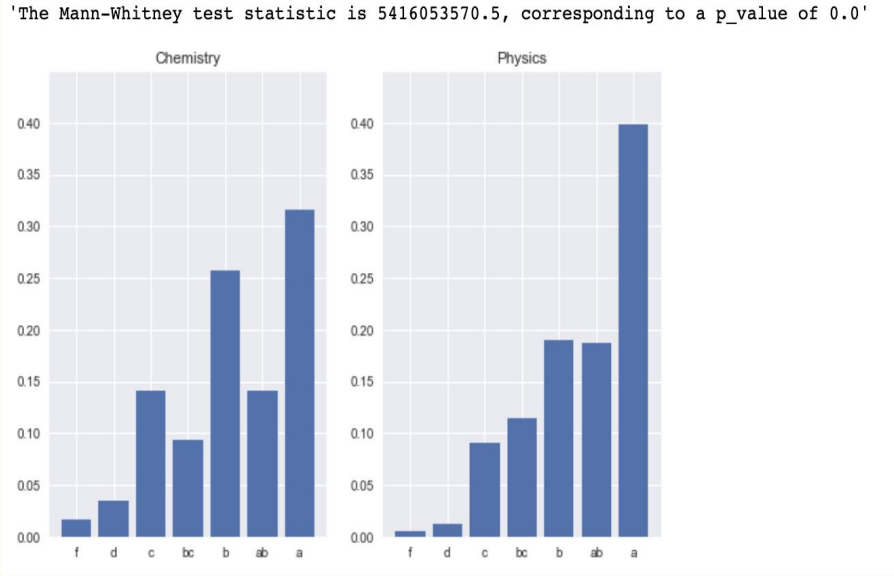
```
31.6% of people of in Chemistry get As
39.9% of people of in Physics get As
The p-value associated with this 8.299999999999997% difference is 0.0

(31.6, 39.9, 0.0)
```



Subject Comparison - with Mann-Whitney Test

- You can also compare classes and subjects with the Mann-Whitney U Test
- This is a rank sum test which compares the overall distributions
  - Null Hypothesis: distribution 1 is the same as distribution 2
- Do the distributions overall differ substantially?
- Still looks like Physics gives out higher grades!

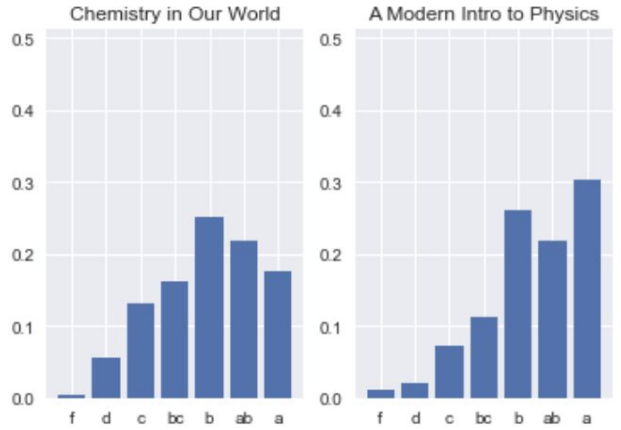


Individual Class Comparison

- These course wide comparisons include many advanced courses, maybe that's skewing the results
- After all it shouldn't be surprising if physics majors are getting good grades in the advanced classes!
- Luckily, we can easily compare individual classes!
- It looks like the grade distributions do differ significantly for the intro-overview courses

```
class_graph_mw_comp('Chemistry in Our World',
                    'A Modern Intro to Physics')
```

```
'The Mann-Whitney test statistic is 419625.0, corresponding to a
p_value of 0.0'
```

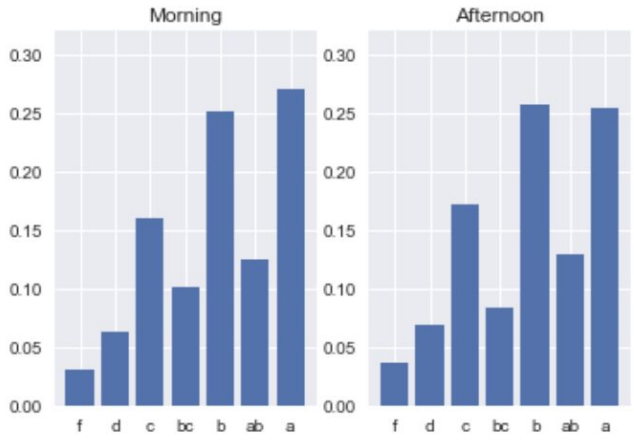


Does Class Section Matter?

- Of course, you also plan to take statistics, the most important class
- That gets you wondering - does it matter which section I'm in? Do grades differ between the morning and the afternoon sections?
- Luckily, we can test that too!
- The distributions are a little different... but not enough to be considered statistically significant

```
morn_aft_comp('Intro-Theory of Probability')
```

```
'The Mann-Whitney test statistic is 2183716.0, corresponding to
a p_value of 0.11473'
```



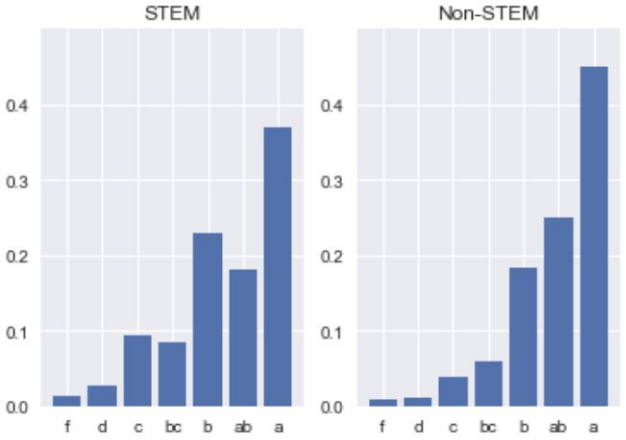
The Great STEM vs. non-STEM debate

- We can also once and for all, scientifically answer which subjects give out the easiest As
- We compiled all the relevant subjects - around 3 million grades overall
- Given the size of these samples we can say with a high degree of certainty that the nearly 8% more A's seen in non-STEM courses is not due to chance

```
subject_graph_compare('STEM',stem_summary,
                      'Non-STEM', non_stem_summary)
```

```
37.02% of people of in STEM get As
44.97% of people of in Non-STEM get As
The p-value associated with this 7.949999999999996% difference
is 0.0

(37.02, 44.97, 0.0)
```



STEM and non-STEM graphs