QQQ Report

Lecture 20

Report format using QQQ

1. Qualitative:
   - Question, problem, hypothesis, claim, context, motivation
   - Definitions, data, methods to be used
   - Rationale, assumptions, biases

1. Quantitative:
   - Data processing, analysis, visualization
   - Documented code and results
   - Summary visuals

1. Qualitative:
   - Answer, update question/claim, summary, re-contextualization, story, relate to domain knowledge
   - Uncertainty, limitations, caveats
   - New problems, next steps

1. Repeat. QQQ-QQQ-QQQ-...
   - Break down a large problem into parts
   - Alternative approaches to a problem
   - Sequence of related problems, "vignettes"
   - Follow-up problems

References:

- Good example of a Jupyter Notebook report: https://nbviewer.jupyter.org/gist/nealcaren/5105037 (https://nbviewer.jupyter.org/gist/nealcaren/5105037)
- QQQ: https://www.bava.stat.vt.edu/wp-content/uploads/2017/08/Developing-a-New-Interdisciplinary-Computational-Analytics-Undergraduate-Program-A-Qualitative-Quantitative-Qualitative-Approach.pdf (https://www.bava.stat.vt.edu/wp-content/uploads/2017/08/Developing-a-New-Interdisciplinary-Computational-Analytics-Undergraduate-Program-A-Qualitative-Quantitative-Qualitative-Approach.pdf)
- Using visuals to support claims: https://www.cbre.com/research-and-reports/Scoring-Tech-Talent-in-North-America-2018 (https://www.cbre.com/research-and-reports/Scoring-Tech-Talent-in-North-America-2018)
- Typical industry spam: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf)

Example...

Why ACC is almost the best (and rankings are wrong)?

Which NCAA basketball conference is best? I think ACC did really well last year. Every year my buddies argue about which conf is the best, so lets settle the issue.

How can we define "best'? A good conf is defined by containing good schools. A the good-ness of a school is defined by its ranking. We will assume rankings from NCAAA https://www.ncaa.com/rankings/basketball-men/d1/ncaa-mens-basketball-net-rankings (https://www.ncaa.com/rankings/basketball-men/d1/ncaa-mens-basketball-net-rankings)

To aggregate school rankings into conf rankings, we will use the mean school ranking of the conf. There lots of ways we could do this, ... We are choosing mean ranking because ...

detail on what we do below...

In step 1 we scrape the data from NCAA rankings and capture the Ranking School name and COnf name. We output the number of schools ranked.

In [3]:
```
###
import requests, bs4, pandas

url = 'https://www.ncaa.com/rankings/basketball-men/d1/ncaa-mens-basketball-net-rankings'

schools = pandas.DataFrame(
    [(int(r.contents[1].text), r.contents[5].text, r.contents[7].text)
        for r in
            bs4.BeautifulSoup(requests.get(url).text, 'html5lib')
                .find_all('tr')[1:]],
    columns=['Ranking','School','Conference'])

len(schools)
```

Out[3]: 353

In [5]:
```
schools.set_index('Ranking')
```

Out[5]:

Ranking	School	Conference
1	Virginia	ACC
2	Gonzaga	WCC
3	Duke	ACC
4	Kentucky	SEC
5	Michigan St.	Big Ten
...	...	...
349	Mississippi Val.	SWAC
350	UNC Asheville	Big South
351	Delaware St.	MEAC
352	UMES	MEAC
353	Chicago St.	WAC

353 rows × 2 columns

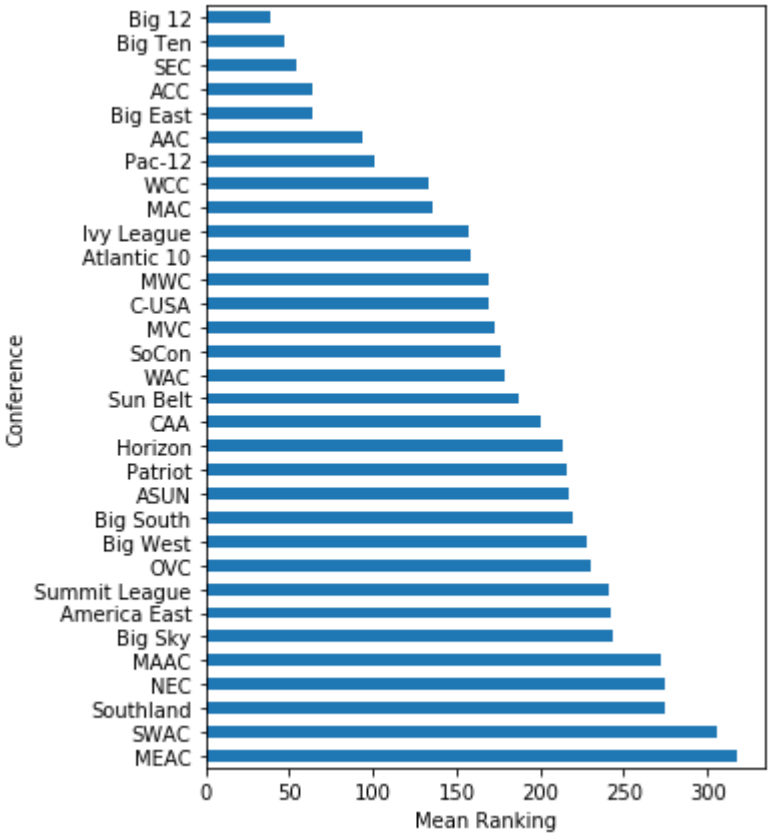
In step 2, we aggregate the schools by conf, using mean ranking and output the mean ranmk for each conf.

In [7]:
```
###
conf = schools.groupby('Conference').mean().sort_values('Ranking')
conf['Mean Ranking'] = conf.Ranking.map(round)
conf[['Mean Ranking']]
```

Out[7]:

Conference	Mean Ranking
Big 12	39
Big Ten	47
SEC	54
ACC	64
Big East	64
AAC	94
Pac-12	101
WCC	133
MAC	136
Ivy League	157
Atlantic 10	158
MWC	169
C-USA	169
MVC	173
SoCon	177
WAC	179
Sun Belt	187
CAA	200
Horizon	213
Patriot	216
ASUN	217
Big South	220
Big West	228
OVC	230
Summit League	241
America East	242
Big Sky	244
MAAC	272
NEC	274
Southland	275
SWAC	306
MEAC	318

In [25]:
```
%matplotlib inline
chart = conf['Mean Ranking'].sort_values(ascending=False).plot.barh(f
igsize=(5,7), sharex=True)
chart.set_xlabel('Mean Ranking');
#chart.grid()
```



Alas, ACC is 4th in the conf mean rankings. Big 12 is the 'best'. Interestingly, the Power 5 is indeed distinct from all the remaining conferences. This should settle my buddys' arguments once and for all.

Admittedly, this is only one way to compute the 'best' conf. It would be interesting to compare the results to some other ways to compute best, like .... counting # teams in top 25.

In future work, Maybe looking at Median as an aggregation measure would overcome some issues of skew in the conf rankings...