

1. Enter:

https://github.com/rsonka/TQFRanalysis



Features Business Explore Marketplace Pricing

This repository Search

Sign in or Sign up

rsonka / TQFRanalysis

Watch 1

Star 0

Fork 0

Code

Issues 0

Pull requests 0

Projects 0

Insights

Join GitHub today

GitHub is home to over 20 million developers working together to host and review code, manage projects, and build software together.

Sign up

Dismiss

A project that helps with course planning by pulling data from the TQFRs website, and analyzing it.

10 commits

1 branch

0 releases

2. Click:

Branch: master

New pull request

Find file

Clone or download

rsonka More fixes for 80 char monitors

TQFRdata Adding a good collection of physics electives for fall 2017.

README.md README update: -42 className bug was fixed, no longer noted in REA

TQFRanalyzer.py Some things really only allow 80 characters. Fixed some menus for tha

TQFRpage.py More fixes for 80 char monitors

TQFRscraper.py Some things really only allow 80 characters. Fixed some menus for tha

Clone with HTTPS

Use Git or checkout with SVN using the web URL.

https://github.com/rsonka/TQFRanalysis

3. Click:

Open in Desktop

Download ZIP

1. Enter:

https://github.com/rsonka/TQFRanalysis

Sign in or Sign up

rsonka / TQFRanalysis

Watch 1

Star 0

Fork 0

Code

Issues 0

Pull requests 0

Projects 0

Insights

Opening TQFRanalysis-master.zip

You have chosen to open:

TQFRanalysis-master.zip

which is: Compressed (zipped) Folder (26.6 MB)

from: https://code.load.github.com

What should Firefox do with this file?

☐ Open with Windows Explorer (default)

☒ Save File

☐ Do this automatically for files like this from now on.

OK

Cancel

4. Save the file

2. Click:

Clone or download

Clone with HTTPS

Use Git or checkout with SVN using the web URL.

https://github.com/rsonka/TQFRanalysis

3. Click:

Open in Desktop

Download ZIP

5. I suggest creating the MyPython directory and saving to C:\Python27\MyPython.

C:\Python27\MyPython

Organize Open Burn New folder

★ Favorites

Autodesk 360
Desktop
Downloads
Recent Places

Libraries

Documents
Music
Pictures
Videos

Computer

Windows (C:)
Removable Disk (E:)
Autodesk 360

Network

Name

Date modified

Type

Size

TQFRanalysis-master.zip

11/5/2017 6:16 PM

Compressed (zippe...

27,236 KB

Open

Open in new window

Extract All...

7-Zip

CRC SHA

Scan with Sophos Anti-Virus

Edit with Notepad++

Open with...

Restore previous versions

Send to

Cut

Copy

Create shortcut

Delete

Rename

Properties

6. Right-click zip file. Extract All.

This window will appear:

Extract Compressed (Zipped) Folders

Extract Compressed (Zipped) Folders

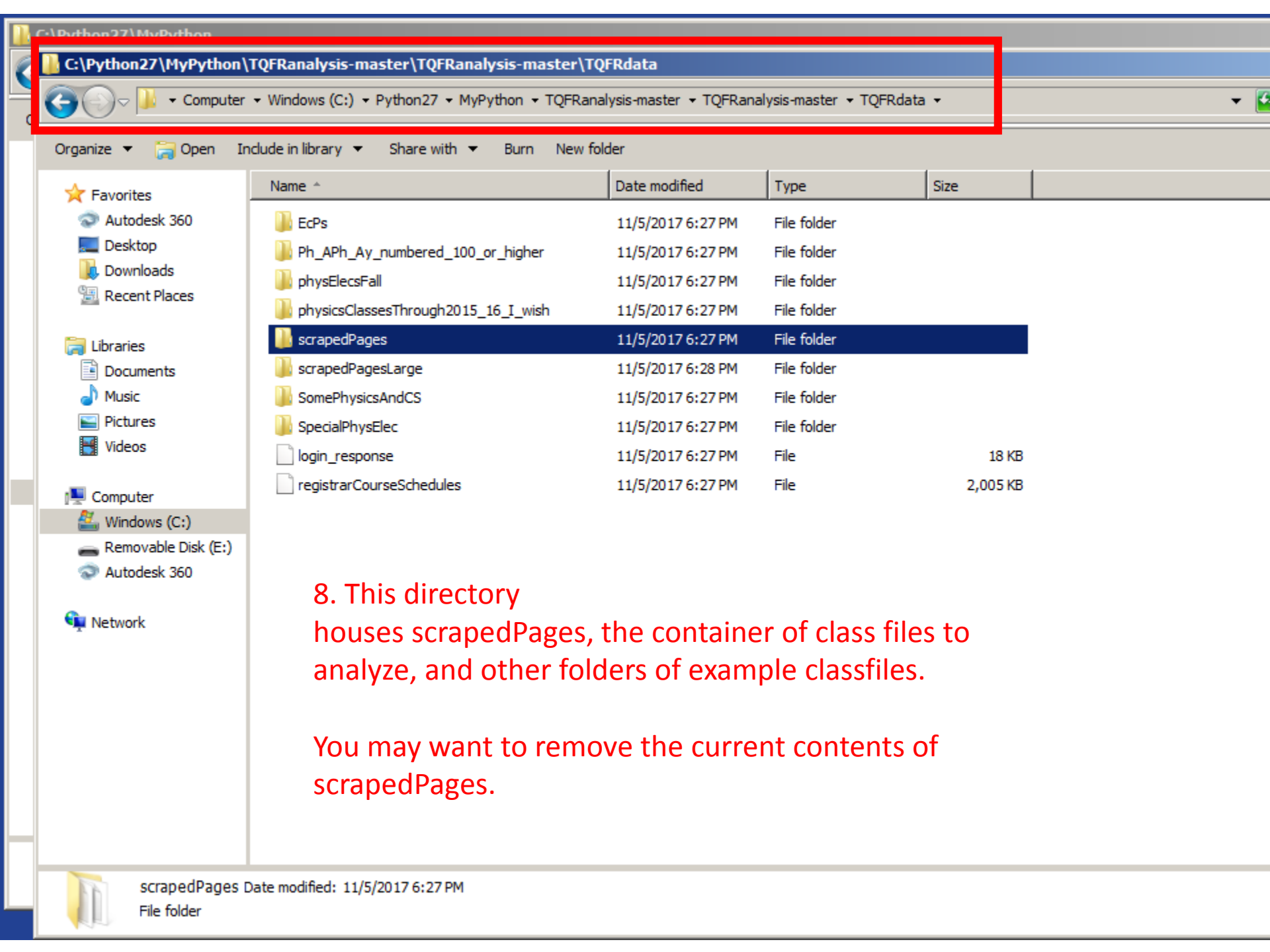
Select a Destination and Extract Files

Files will be extracted to this folder:

C:\Python27\MyPython\TQFRanalysis-master

☒ Show extracted files when complete

7. Extract!



8. This directory houses scrapedPages, the container of class files to analyze, and other folders of example classfiles.

You may want to remove the current contents of scrapedPages.

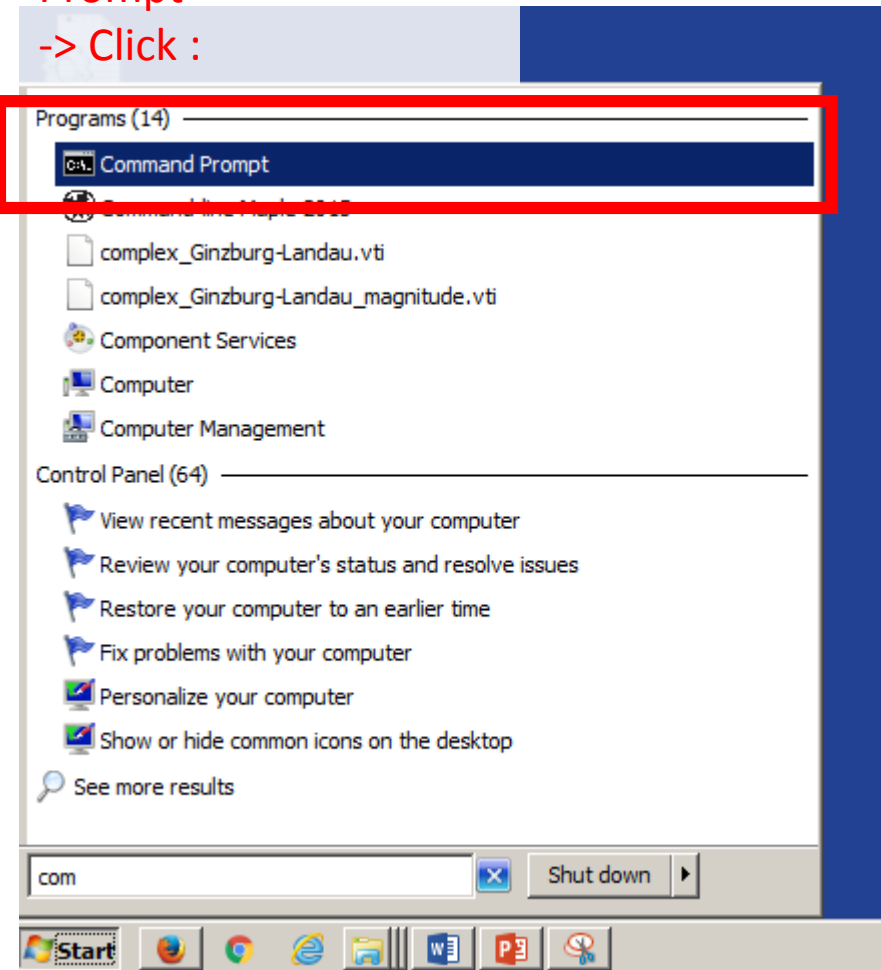


scrapedPages Date modified: 11/5/2017 6:27 PM
File folder

9. Windows Start menu

-> type "Command Prompt"

-> Click :



10. Navigate to location of our files by entering

```
cd ..\..\Python27\MyPython\TQFRanalysis-master\TQFRanalysis-master
```

```
Microsoft Windows [Version 6.1.7601]  
Copyright (c) 2009 Microsoft Corporation. All rights reserved.
```

```
C:\Users\rsonka>cd ..\..\Python27\MyPython\TQFRanalysis-master\TQFRanalysis-master
```

```
C:\Python27\MyPython\TQFRanalysis-master\TQFRanalysis-master>
```

11. Install needed python package 'bs4' by entering

`py -2 -m pip install bs4`

```
Command Prompt
```

```
Microsoft Windows [Version 6.1.7601]  
Copyright (c) 2009 Microsoft Corporation. All rights reserved.
```

```
C:\Users\rsonka>cd ..\..\Python27\MyPython\TQFRanalysis-master\TQFRanalysis-master
```

```
C:\Python27\MyPython\TQFRanalysis-master\TQFRanalysis-master>py -2 -m pip install bs4
```

```
You are using pip version 7.0.1, however version 7.0.2 is available.  
You should consider upgrading via the 'pip install --upgrade pip' command.
```

```
Collecting bs4
```

```
Using cached bs4-0.0.1.tar.gz
```

```
Requirement already satisfied (use --upgrade to upgrade): BeautifulSoup4 in c:\python27\lib\site-packages (from bs4)
```

```
Installing collected packages: bs4
```

```
Running setup.py install for bs4
```

```
Successfully installed bs4-0.0.1
```

```
C:\Python27\MyPython\TQFRanalysis-master\TQFRanalysis-master>
```

12. Then install needed python packages 'requests' and 'numpy' by entering

`py -2 -m pip install requests`

And then

`py -2 -m pip install numpy`

Your installation is now complete! I was able to log out and back in and it was still there, so you should only have to do this once.

1. Now we can run our program. Enter
py -2 runThisOne.py

```
Command Prompt: py -2 runThisOne.py
C:\Python27\MyPython\TQFRanalysis-master\TQFRanalysis-master>py -2 runThisOne.py

using: C:\Python27\MyPython\TQFRanalysis-master\TQFRanalysis-master
found existing TQFRdata folder
login
    Login to Caltech access to enable scraping. Do this once after starting the
    program if you intend to scrape new data.
scrape
    Interactively scrape class information from TQFRs for analysis. Does not
    load the scraped pages base info into the system-there's an option for that
    under analyze.
analyze
    Load and analyze scraped data.
instructions ! help ! info ! information
    Repeat these instructions.
commands
    Prints JUST the names and arguments of the valid commands.
done
    End the program.

[Main] Command: login

    Login to Caltech access to enable scraping. Do this once after starting the
    program if you intend to scrape new data.
scrape
    Interactively scrape class information from TQFRs for analysis. Does not
    load the scraped pages base info into the system-there's an option for that
    under analyze.
analyze
    Load and analyze scraped data.
instructions ! help ! info ! information
    Repeat these instructions.
commands
    Prints JUST the names and arguments of the valid commands.
done
    End the program.

[Main] Command: login
Enter access.caltech.edu username: rsonka
Enter access.caltech.edu password:
Getting login page
Logging in
Login response will be written to 'login_response', under the TQFRdata folder. If you have weird problems, possibly che
k this file-you may not have logged in correctly.

[Main] Command: scrape
```

2. We have to login to Caltech Access to scrape new pages.

3. Let's go scrape some data!

```
Ctrl Command Prompt - py -2 runThisOne.py
scrapeClass <department1> <number> [termChar] ['prac', 'anal', or nothing]
Usage example: scrapeClass Ma 1 A
Copies every html file you do not yet have that have that number and that
department. If the class is cross-listed, <department1> should be the
PRIMARY department-the one the class is listed under on the registrar.
scrapeProfessor <ProfessorFirstname> <ProfessorLastname>
Copies every html file you do not yet have taught by that professor. Names
should be capitalized and spelled correctly.
scrapeAdvanced
Construct a specific template to match classes to for scraping.
Interactively set any of year, term, division, professor, a range of
numbers, department, termChar, whether it's practical or analytical,
actual classname. Can be SIGNIFICANTLY faster than scrapeClass or
scrapeProfessor depending on how much information you specify. Just giving
a division can speed things up dramatically, though make sure you're
actually writing the division the way it is listed on TQFRs!
instructions
Prints these instructions again.
done
Indicates you have finished scraping and want to stop.

[Scraping] Command: scrapeAdvanced
```

(You might want to actually scroll up and read this menu. I'm going to skip to doing stuff because this is a 10 minute presentation.)

4. I'm going to pick out the best Winter-term economics classes to fulfill my Advanced Social Science requirement with.

```
Ctrl Command Prompt - py -2 runThisOne.py
WARNING! ODD CLASS NAME: "Psy015". Probably fine if program does not immediately crash.
[DEBUG] Psy--15 ANY 2008-09 WI HSS
WARNING! ODD CLASS NAME: "Psy101". Probably fine if program does not immediately crash.
[DEBUG] Psy--101 ANY 2008-09 WI HSS
[DEBUG] SS--201B ANY 2008-09 WI HSS
[DEBUG] SS--222B ANY 2008-09 WI HSS
[DEBUG] SS--223B ANY 2008-09 WI HSS

[Scraping] Command: scrapeAdvanced
NOTE: For any of the following, input 'ANY' to match anything. DON'T INCLUDE THE QUOTES ('') FOR ANY OF THIS.
Year, e.g. '2015-16': ANY
Term, e.g. 'PA', 'WI', 'SP': WI
Division, e.g. BBE, CHCHE, EAS, GPS, HSS, Institute, PMA, Performing and Visual Arts, or Physical Education: HSS
Enter Professor(s), separated by '_' e.g. 'Firstname McLastname' or 'Firstname McLastname_Othername Lastname': ANY
The lowest classnumber you want to match: 100
The highest classnumber you want to match: 1000000000
Enter departments, separated by spaces (e.g. CS Ph): Ec
If you want to match any class that SHARES a department with one of the ones you input, enter Y. Otherwise, enter anything, and it will ONLY match those that are listed under all those departments (and not a single one more). E.g.: Y if you want to include cross-listings.Y
Enter termChar ('A', 'B', 'C', or '') ['ANY' is also an option, as always]: ANY
Enter 'prac' for practical, 'anal' for analytical, '' for a class that isn't split like that:
Enter classname. You should put 'ANY' here unless you're dealing with one of the RARE cases where the previous stuff doesn't fully identify the class; I'm not aware of any. If using this, will need the exact classname used in TQFRS; this is mostly here for if they do something stupid in the future, or if I missed something in the present. ANY
```

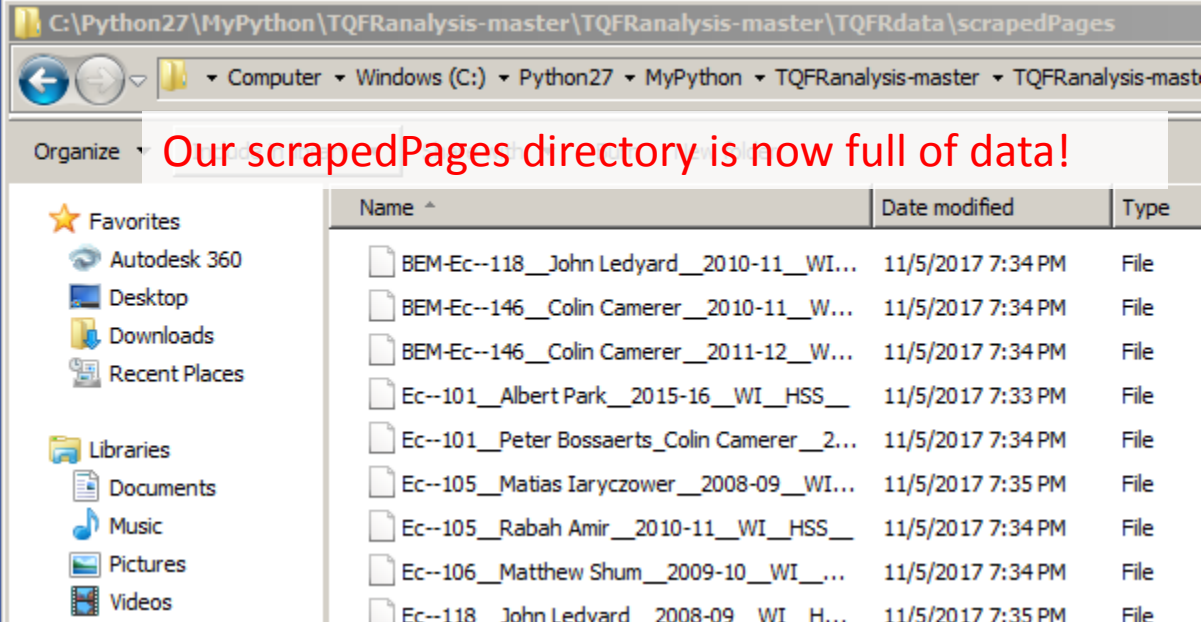
The program then downloads TQFR reports for every winter Economics class #100 or above.


```

[DEBUG] SS--201B ANY 2008-09 WI HSS
[DEBUG] SS--222B ANY 2008-09 WI HSS
[DEBUG] SS--223B ANY 2008-09 WI HSS
[Scraping] Command: done
[Main] Command: analyze

```

5. Time to examine our options.



```

C:\Python27\MyPython\TQFRanalysis-master\TQFRanalysis-master\TQFRdata\scrapedPages
Command Prompt - py -2 runThisOne.py
analyzeClagg <className, exactly as listed on one of the clagg displays>
Does what analyzeClass does, except by picking an already constructed clagg
from the list. Unlike analyzeClass, analyzeClagg will affect which pages are
included in the stats displayed for the clagg in
claggStats/sortClaggs/importantNumbers etc. Try looking at/excluding the
individual pages with this option if you see weird things in the claggStats,
or just to look at a class in more detail.
importantNumbers
Prints the numbers I look at first for hums for all classAggs constructed.
constructProfessorAggregates
# TODO!!!!:
Assigns every loaded page to a ProfessorAggregate, if it is not already
assigned. Necessary if you want to search for a professor by their aggregate
data qualities (e.g., 'overall teaching' score).
instructions ! help ! info ! information
Prints these instructions again.
commands
Prints JUST the names and arguments of the valid commands.
done
Indicates you have finished analyzing and return to main menu.

Typical usage:
fullLoad -> constructClaggs -> claggStats or sortClaggs or analyzeClagg

[Analyzer] Command: fullLoad
PS-Ec--172__Omer Tamuz__2016-17__WI__HSS__
PS-Ec--173__Christopher Chambers__2009-10__WI__HSS__
[Analyzer] Command: constructClaggs

```

(Again, you might want to actually scroll up and read this menu.)

6. A couple setup commands to load our data and calculate the stats for the classes.

7. If you're like me, you probably don't really care about economics. Thus, your first instinct is to ask "which class is the least work?"

```
Command Prompt - py -2 runThisOne.py

[Analyzer] Command: sortClaggs
Legit statNames:
['expectedGrade', 'homeworkCompletion', 'hoursOutsideClass', 'lectureAttendance', 'passFail', 'co
untPerception']
Enter statname to sort by exactly as above: hoursOutsideClass
['expectedGrade', 'homeworkCompletion', 'hoursOutsideClass', 'lectureAttendance', 'passFail', 'co
untPerception']

|className|hoursOutsideClass|
|Ec 140|12.75|
|Ec 121B|13.24|
|BEM/Ec 146|13.31|
|Ec 101|13.71|
|Ec/SS 124|14.47|
|BEM/Ec 118|14.7|
|Ec 118|15.0|
|Ec/Psy 109B|15.0|
|Ec/SS 129|15.0|
|PS/Ec 173|15.0|
|Ec 105|16.2|
|PS/Ec 172|16.48|
|Ec 106|16.5|

[Analyzer] Command:
```

This list takes about 1.5 numbingly boring hours to compile by hand due to how TQFR report is laid out. I guarantee you can install TQFRscraper on an SFL computer and get this list in way less than 1.5 hours.

Cross-reference with registrar list of classes available next term (<http://schedules.caltech.edu/WI2017-18.html>) ... 121B, 124, and 172 are offered next term, plus the completely-new or never-surveyed classes 101, 149, and 181.

So Ec 121B is our top contender, followed by Ec/SS 124. We don't really want to take the underunited PS/Ec 172.

8. Let's quickly check overall class stats to make sure we're not missing anything nasty, like an arbitrary grading scheme or horrific content quality.

```
[Analyzer] Command: claggStats
Legit statNames:
['expectedGrade', 'homeworkCompletion', 'hoursOutsideClass', 'lectureAttendance', 'passFail', 'contentQuality']
Now prompting for stats. Enter statNames EXACTLY as above, or enter 'all' to include all of them.
Otherwise, enter 'done' when finished, 'undo' to remove last statName entered.
StatName, all, done, or undo: all
My stats: ['expectedGrade', 'homeworkCompletion', 'hoursOutsideClass', 'lectureAttendance', 'passFail', 'contentQuality']
|className|expectedGrade|homeworkCompletion|hoursOutsideClass|lectureAttendance|passFail|contentQuality|
|Ec 140|14.0|1.0|12.75|10.88|1.0|14.5|
|Ec 121B|13.82|1.0|13.24|10.83|1.0|13.31|
|BEM/Ec 146|13.86|1.0|13.31|10.9|1.0|13.88|
|Ec 101|14.0|1.0|13.71|10.97|1.0|14.71|
|Ec/SS 124|13.75|10.99|14.47|10.66|1.0|14.0|
|BEM/Ec 118|13.0|10.98|14.7|10.76|1.0|13.6|
|Ec 118|1.0|10.97|15.0|10.65|1.0|13.67|
|Ec/Psy 109B|14.0|1.0|15.0|1.0|1.0|15.0|
|Ec/SS 129|13.67|10.91|15.0|10.84|1.0|14.5|
|PS/Ec 173|14.0|10.97|15.0|10.97|1.0|14.0|
|Ec 105|13.64|10.96|16.2|10.8|10.75|13.6|
|PS/Ec 172|13.37|10.98|16.48|10.56|1.0|13.88|
|Ec 106|13.5|1.0|16.5|10.95|1.0|14.0|

[Analyzer] Command: analyzeClagg Ec/SS 124
```

9. Hmm, Ec/SS 124 gets a much higher contentQuality rating. Is it enough to justify the extra hour of work per week? Let's look at the class's consolidated TQFRdata in depth.

```

Homework Completed: 10.99 +/- 0.03 [1.0, 1.0, 1.0] 17
Course Content Quality [0 (bad) to 5 (good)]: 13.98 +/- 0.86 [3.5, 4.0, 5.0] 47
PROFESSOR DATA: Robert Sherman
NOTE: if responders for individual item < responders for whole (under Percent Response), remainder selected
Statistical Quantity: Average +/- stdev Quartiles Responder
-----
Percent Response: 29.0 32
---Overall Ratings---
The instructor's overall teaching 4.18 +/- 0.91 [4.0, 4.0, 5.0] 44
---Organization/Clarity---
Set out and met clear objectives announced for the course 4.11 +/- 0.88 [3.0, 4.0, 5.0] 44
Displayed thorough knowledge of course material 4.66 +/- 0.71 [5.0, 5.0, 5.0] 44
Explained concepts clearly 4.18 +/- 0.83 [3.0, 4.0, 5.0] 44
Distinguished between more important and less important topics 3.98 +/- 0.89 [3.0, 4.0, 5.0] 44
Presented material at an appropriate pace 4.11 +/- 0.8 [3.0, 4.0, 5.0] 44
---Ability to Engage and Challenge Students Intellectually---
Emphasized conceptual understanding and/or critical thinking 4.11 +/- 0.71 [4.0, 4.0, 5.0] 44
Related course topics to one another 4.05 +/- 0.93 [3.0, 4.0, 5.0] 44
---Interaction with Students---
Demonstrated concern about whether students were learning 4.39 +/- 0.71 [4.0, 5.0, 5.0] 44
Inspired and motivated student interest in the course content 4.11 +/- 1.09 [4.0, 4.0, 5.0] 44
Was available for consultation outside of class 4.49 +/- 0.74 [4.0, 5.0, 5.0] 41
---Course Organization, Content, and Evaluation---
Selected course content that was valuable and worth learning 4.18 +/- 0.83 [4.0, 4.0, 5.0] 44
Organized course topics in a coherent fashion 3.98 +/- 0.89 [3.0, 4.0, 5.0] 44
Chose assignments that solidified understanding 4.18 +/- 0.91 [3.0, 5.0, 5.0] 44
Explained clearly how students would be evaluated 4.59 +/- 0.61 [4.0, 5.0, 5.0] 44
Designed and used fair grading procedures 4.59 +/- 0.61 [4.0, 5.0, 5.0] 44
Gave tests and quizzes that accurately reflected material taught 4.53 +/- 0.66 [4.0, 5.0, 5.0] 32
Comments:
-----
-----BELOW FROM CLASS: Ec-SS--124_Robert Sherman_2012-13_WI_HSS_ Introduction to Empirical Process Met
!
!No comments were entered for this subject.
!
!-----BELOW FROM CLASS: Ec-SS--124_Robert Sherman_2014-15_WI_HSS_ Identification Problems in the Social
!
!This is a fairly easy course. The content is somewhat difficult to get excited
about, as bounds analyses almost always end with the conclusion that you can't
say anything about the quantity of interest. Having said that, this is extremely
important to learn! Scientists need to be more aware of the assumptions that
they make in performing data analysis, and going to the trouble of analyzing
data with a variety of explicitly-stated assumptions, both on the homework and
on the final project, serves this purpose well. I would be in favor of making
this a generally required course (for undergraduates, graduate students,
faculty...).!
!This class is a good follow-up to Ec 122 (econometrics), but can be done without
122 as well. The textbook is necessary as homeworks and notes were based on it,
but it was super useful to have. The material, just like 122, was interesting to
me as an ACM major and should be a good class to fulfill HSS requirements for
anyone looking for a SS class without much writing. Also, Prof. Sherman cares a
ton about the material and his students which is a definite plus.

```

!The course is united appropriately and if you start the sets early enough there is usually no issue. I would recommend that students be comfortable with R/Python/Mathematica etc before taking this class.

!Textbook necessary for homework3 sets and either a final or final project.

!-----BELOW FROM CLASS: Ec-SS--124__Robert Sherman__2016-17__WI__HSS__ Identification Problems in the

!Interesting course with a light workload. Three sets, a project, and a final. Weeks that have an assignment are appropriately united, weeks that don't aren't. Good class if you take about this sort of thing.

!This class is not a lot of work. It's pretty boring at times, but the concepts are not too challenging if you pay attention in class or read the book. There are 3 homework assignments, a project, and a final.

COMMANDS: =====

selectIncluded

Select which of the aggregate's files should be included in the analysis.

calculate

Calculates and displays data, as calculated from all files in current included list.

displayComments <Y or N>

Sets whether to display comments or not.

instructions

Prints these instructions again.

done

Return to main analyzer menu.

[Aggregate] Command: **selectIncluded**

#	In?	Identifier
0	Y	Ec-SS--124__Robert Sherman__2012-13__WI__HSS__
1	Y	Ec-SS--124__Robert Sherman__2014-15__WI__HSS__
2	Y	Ec-SS--124__Robert Sherman__2015-16__WI__HSS__
3	Y	Ec-SS--124__Robert Sherman__2016-17__WI__HSS__

Enter which numbers to include, separated by spaces: 0 1 2 3

New included:

#	In?	Identifier
0	Y	Ec-SS--124__Robert Sherman__2012-13__WI__HSS__
1	Y	Ec-SS--124__Robert Sherman__2014-15__WI__HSS__
2	Y	Ec-SS--124__Robert Sherman__2015-16__WI__HSS__
3	Y	Ec-SS--124__Robert Sherman__2016-17__WI__HSS__

[Aggregate] Command: **calculate**

```

C:\> Command Prompt - py -2 runThisOne.py

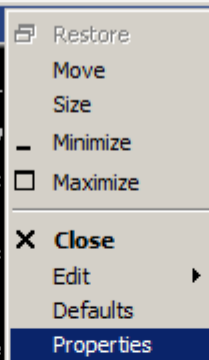
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation

C:\Users\rsonka>cd ..\..\Python27\MyPython\TQFRanalysis-master
This is the main program.

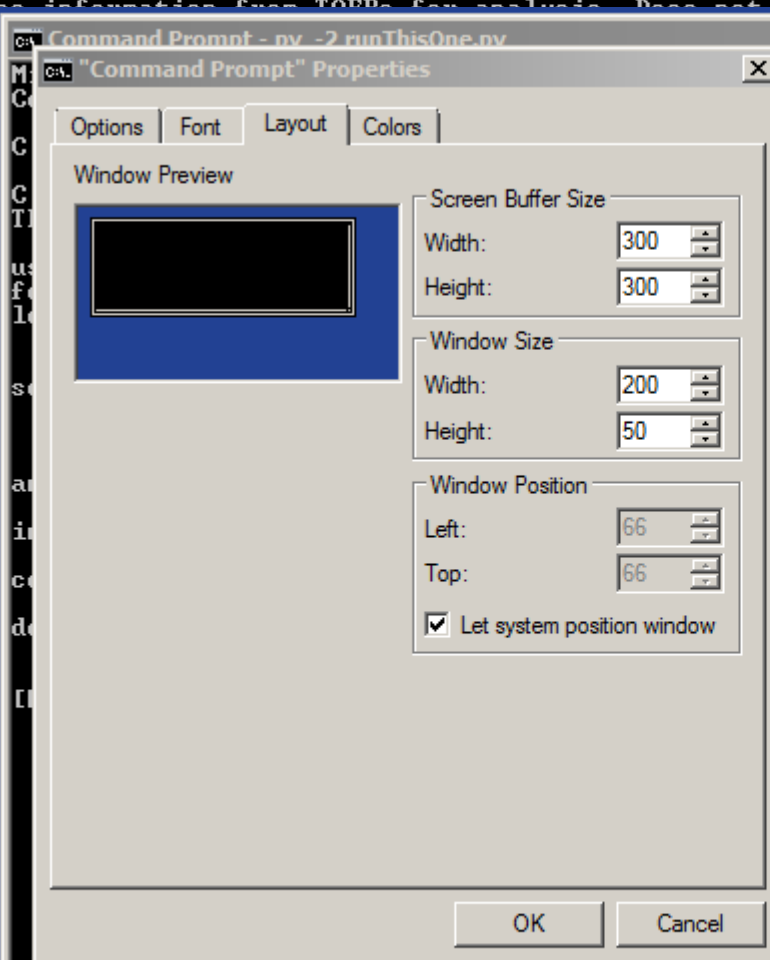
using: C:\Python27\MyPython\TQFRanalysis-master\TQFRdata folder
login
  Login to Caltech access to enable
  program if you intend to scrape ne
scrape
  Interactively scrape class information from TQFRs for analysis. Does not
  load the scraped pages b
  under analyze.
analyze
  Load and analyze scraped
instructions ! help ! info !
  Repeat these instruction
commands
  Prints JUST the names an
done
  End the program.

[Main] Command: _

```



One last trick: right-click
the Command Prompt bar
and open Properties.



Then you can resize the
Command Prompt
Window and buffer to
actually view everything
on one screen.