**My Process**

1. Wrote a script ([OTRSite\_Log\_Scrape](file://localhost/Users/bridget/Desktop/Pratt%20Junk/664/pfch-2015/OTRSite_Log_Scrape.py)) to scrape the individual show URL’s from [www.otrsite.com/radiolog/index.html](http://www.otrsite.com/radiolog/index.html) and save to a CSV file (Log\_Links\_2)
2. Manually cleaned up Log\_Links CSV (removed duplicates/extra links) and saved as Log\_Links\_4
3. Wrote another script ([OTRSite\_Log\_Scrape\_Pt\_2](file://localhost/Users/bridget/Desktop/Pratt%20Junk/664/pfch-2015/OTRSite_Log_Scrape_Pt_2.py)) that would loop through the new CSV, open each link, scrape any text preceded by font size 3 tag or a font size 4 tag and write that to a txt file (OTRSiteScrape.txt)
4. Wrote a script ([OTRSite\_Parse\_Log3](file://localhost/Users/bridget/Desktop/Pratt%20Junk/664/pfch-2015/OTRSite_Parse_Log3.py)) to loop through the text file, pulling out the show titles and names of stars, announcers, and hosts with RegEx and writing each to a separate CSV file.
5. Created a CSV file for Tableau entitled All\_Personalities.csv with Names, Types, and Shows. Reconciled any misspellings or multiple spellings of names. Loaded into Tableau and pulled out the personalities that appeared on the most shows.
6. Wrote those names in another CSV file along with links to their DBPedia page entitled Bio\_Links.csv
7. Wrote a script ([bio\_scrape](file://localhost/Users/bridget/Desktop/Pratt%20Junk/664/pfch-2015/bio_scrape.py)) to loop through each link, connect to the JSON file of the dbpedia page, and pull out the biographies and write them to a new CSV file (Bios.csv)
8. Filled in any bios for personalities w/o a dbpedia page from IMDB.com
9. Put both Bios.csv and All\_Personalities.csv into Tableau and created visualizations.
10. Created an edge file and loaded into Gephi. Weighted nodes by degree.