GROUP 4 PROPOSAL:  
Participants: @Radhika @Jen  
**DATA SETS:**

1. We’re using this wildlife data from Wilson and Reeder’s Mammal Species of the World:

<https://www.departments.bucknell.edu/biology/resources/msw3/>

1. US Fish and Wildlife Service Environmental Conservation Online System(ECOS) shows the endangered status of species in the US

<https://ecos.fws.gov/ecp/report/species-listings-by-tax-group?statusCategory=Listed&groupName=Mammals>

**EXTRACT:**

* Both files will be downloaded as CSVs (Both: 10 min)
* CHALLENGE: Radhika will load one of the CSVs into Postgres using PG Admin’s import tool (Radhika: 15 min)
* One CSV will be imported to pandas using from\_csv, the other using from\_sql with pandas (Radhika: 25 min)

**TRANSFORM**

* Break out the ECOS scientific name into Genus, Species columns (Radhika: 30 min)
* Once imported, we will join the tables on Genus, Species (Radhika: 15 min)
* Remove any rows not related to US mammals (Radhika: 15 min)

**LOAD**

* While Radhika is extracting/loading, Jen will read through each dataset and create a Postgres table to load into. (Jen: 1 hr)
* Jen will work with Radhika to rename columns appropriately (Jen: 15 min)
* Jen will write the portion of the python script that loads the data into the Postgres database. (20 min)

**Notes:**  
This is slated for about 2 hours of work for each of us; we plan to use extra time wednesday to explore any bugs and verify data integrity. Project to be complete Wednesday night.