Project 3

Author: Chase Pashkowich, Chunjie Nan, Jiho Kim

Assignment:

Use the data to answer the question, “Which are the most valued data science skills?”

Collaboration Tools:

We will mainly use our own slack rooms to communicate and shared GitHub will be used for code sharing and documentation.

Data Source:

Main data:

1. “Data Scientist Jobs” from Kaggle (url: <https://www.kaggle.com/andrewmvd/data-scientist-jobs>) that contains information on Job title, Salary and Description which would explain what skills are highly desired.
2. “The most in Demand Skills for Data Scientists” from Kaggle (url: https://www.kaggle.com/discdiver/the-most-in-demand-skills-for-data-scientists/data)

Other possible data: we can web-scrape and pull data “data scientist” job posts from websites like LinkedIn or Indeed to see skills that are required.

Data will be uploaded to Github so that works can be reproduced as a csv file. We will lean on the dplyr() and other R functions to extract keywords and information to create normalized database for analysis and visualization