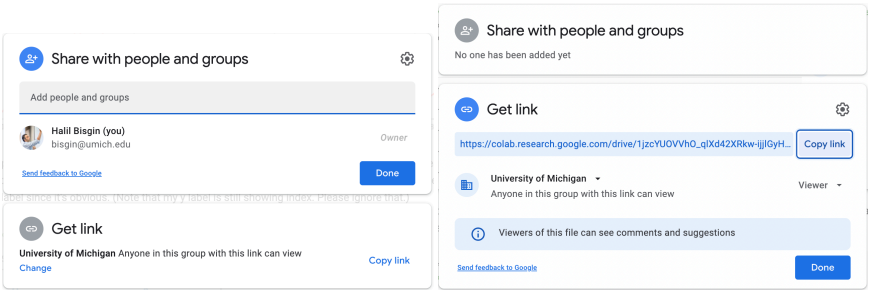
# Homework 6

Introduction to Data Visualization - Winter 2022

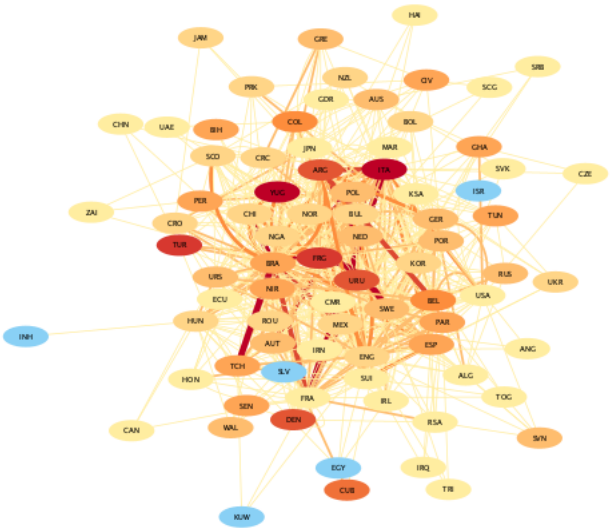
**Due on April 6, 11.59PM**

1. Create a repository called, HW6, on your GitHub to upload your work.
2. Load `house_prices.rda` in R and perform the followings:
   (a) Use `facet_wrap` to draw line plots to represent the trend for `house_price_index` over years for each state. . Please have three ticks on the x axis which are for years 1980, 2000, and 2020. Since you have limited space your labels can be '80, '00, and '20 for these ticks.
   (b) Use `gather()` function from `tidyr` to reshape your data in a way that it will have measure and value columns for `house_price_index` and `unemploy_perc`, respectively, in two columns. You should leave out remaining attributes by using "-" in front them inside the function. Then, direct the output to a new data frame called `house_reshaped` Please refer to examples we did in class.
   (c) Use `house_reshaped` to replicate the graph in 2a with two lines where each represents `house_price_index` and `unemploy_perc` over the years for each state. Once you have the chart, please comment on it as to whether it's a good graph to present those two pieces of information.
   (d) When you achieve the goals above, please create a R markdown file and create an HTML page which has your codes and charts including your comments.
3. Please see the Python notebook **here** and make a copy to yourself.




4. In the same Python notebook, read `WorldCupMatches.csv` by using Pandas.
   (a) Then using Pandas groupby feature on 'Home Team Initials' and 'Away Team Initials' find how many times these countries' national teams played a match. You can use the `count()` after grouping. These counts will be the weight attribute when you create your final data frame explained below.
   (b) Find the total goals scored by each home team for the matches played above by using groupby as earlier and sum on the 'Home Team Goals' column. These summations will be your HomeGoalTotal' attribute below.
   (c) Use the values you found above, come up with a new data frame called `team_pairs` which has 'Home Team Initials', 'Away Team Initials', 'weight', 'HomeGoalTotal' columns where latter two columns should have the values from your earlier steps.
   (d) Write your `team_pairs` to a csv file by using Pandas to_csv feature. (Please ignore the fact that there may be symmetric pairs such as ARG-BRA and BRA-ARG in your data frame. Please proceed what you have for the sake of simplicity. If you can consolidate them, you can get 5 points bonus.)
   (e) When you finish coding, please upload this Python notebook to your GitHub and share the link. Please also share the link to your google drive for me to be able to run if needed. When you want to share your Google Colab, please don't put my email address. Instead, click on **Change** and make it viewable to anyone from University of Michigan as below:
5. Perform the following steps in Cytoscape:
   (a) Import your csv file as a network.
   (b) Fill your node colors based on the HomeGoalTotal' values and select a color scale which has red on the higher end of the spectrum.
   (c) Color your edges based on the weight column of the network file so that we can see how often national teams played together. Please use the same colorscale above.
   (d) Use the weight column to change the width of edges in your network in a way that width should go from 1 to 10 depending on the weight. (Sliding bar may not let you hit exact numbers. As long as you're close to 1 and 10, that is fine)

(e) Use the yFiles Organic Layout to change the layout of your network. Then, use the Annotation tab on the control panel, insert a text box in which you should put a title for your network. At the end, your network should like similar to the one below with an annotation on it.




(f) Please export your network as an image and upload it to the submission page.

---

- Please follow the submission instructions in each question.
- No late submissions will be accepted **unless you have an excuse**.