**DATA 608 – Knowledge and Visual Analytics**

**Final Project Proposal by Jagdish Chhabria**

**Data**

I intend to use bank branch data from the website of the Federal Deposit Insurance Corporation (FDIC) to show the distribution of physical bank branches across the USA. I plan to focus on the branch distributions of the 5 largest national banks and pull in historical data to show how their branch network has been changing in the past 10-15 years.

The link to the data source can be found at:

<https://www7.fdic.gov/sod/dynaDownload.asp?barItem=6>

**Objective and Approach**

The reason I chose this data to visualize is because I think it would be very relevant as a first step in Exploratory Data Analysis (EDA) into current economic policy topics such as:

1. linkage between bank branches and small business credit.

or

1. business impact of newer digital banking channels on physical branch networks. Given the increasing penetration of electronic banking as well as growth of fintechs, the cost-benefit analysis of brick-and-mortar branches has high strategic importance to both incumbent and new challenger banks.

While I will not have time to conduct a full-fledged study, I will instead focus on the visualization aspect of bank branch and deposit data, since this course is about data visualization.

I will explore using graphics libraries in either R or Python for this. If time permits, I may try to create a R Shiny interactive application instead of static graphs.

The FDIC website has a lot of data. I will need to spend time on understanding the data dictionary, select datasets, clean/remove unwanted data attributes and potentially merge data from different files. Given that I intend to use historical data, I will likely have to deal with missing attributes and imputing their values if required.