The attached excel file has 11 quarters (each quarter is its own tab) of selected data from the Y9C, a financial regulatory report for bank holding companies.

By quarter, we don’t mean 1st, 2nd, 3rd, 4th quarters of the year; we mean unique date – so 12/31/2012 is the first quarter observed in the data and 6/30/2015 is the 11th quarter in the data.

Each bank is identified by its RSSD, which is a unique identifier that does not change over time (note, a bank can go away if it fails/closes or if it is purchased (after being purchased, the bank/RSSD won’t be observed in the dataset but the data for the purchasing bank/RSSD will reflect the addition of the purchased bank in subsequent quarters).

Please answer/complete the following questions. You should include the code that you have used to complete each question.

In addition, your graphs and charts should be self-contained. That is, someone should be able to look at the chart and understand what you’ve done. This will include an informative title, labeled axes and a description of the variables on each axis (including the units), a legend (if appropriate) and any explanatory notes (particularly helpful if the chart is hard to understand even after looking at the above items)

1. Put the excel file into one dataset, where each observation is a date/RSSD combination.
   1. Be clear about your variable names and labeling conventions. Feel free to drop variables that you won’t need as part of the questions below. In addition to any variables that you keep (which should include the variables that you use to identify your observations), create and/or name the following variables:

Assets = **BHC2170**

Return on assets (ROA) = **BHCK4301 /** Assets

Return on equity (ROE) = **BHCK4301 / BHCKG105**

Asset growth (AGR) = (Current Assets – Prior quarter’s assets) / Prior quarter’s assets

1. You will have RSSDs that show up less than 11 times due to mergers and acquisitions as well as firms that fall out of, or enter, a pre-set lower asset threshold.
   1. Figure out what the asset threshold referred to above is?
   2. How many unique RSSDs are in your dataset. Compute the frequency (i.e., number of quarters) in which each RSSD is present in your data, and produce a histogram showing that frequency across the RSSDs.
   3. Determine whether there are any cases where the “FIRM NAME” for a particular RSSD changes over the period.
2. Diagnose and explain cases where your calculations in question 1a resulted in missing values.
3. For the *most recent quarter* in the data, plot a graph that is appropriate to look for a possible relationship between:
   1. ROA vs. Assets
   2. ROA vs. the log of Assets (does the x-axis look better now?)

Insert a note on the graph that provides the correlation between the two variables that you have plotted.

1. Do both parts of question 4, but for AGR vs. Assets and ROE vs. Assets.
2. Plot the Sum of Assets for the largest 10 banks (the sum, not the assets of each firm) for each quarter of your data. So, Y axis=sum of assets of 10 largest firms, x axis=quarter.
3. Plot the average AGR, ROA, and ROE of the 10 largest banks (largest is measured by assets) each quarter.
   1. Include the following notes at the bottom of your graph

“The assets of the largest bank in the U.S. were $XX billion as of December 31, 2012 and $XX billion as of June 30, 2015” and,

“The assets of the 10th largest bank in the U.S. were $XX billion as of December 31, 2012 and $XX billion as of June 30, 2015”

1. Plot the AGR over time (on one graph) for the following firms: JPMorgan, Morgan Stanley, Goldman Sachs, Wells Fargo, and Bank of New York Mellon
   1. Include the following note for each bank at the bottom of your graph:

“BANK X’s assets were $XX billion as of December 31, 2012 and $XX billion as of June 30, 2015”

1. Plot the average AGR over time (on one graph) for each quartile (by assets) of the data – 1st quartile = average AGR among banks with assets at the 25th percentile or below, …
   1. Include the following note for each quartile at the bottom of your graph:

“The median amount of assets for banks in the Xth asset quartile was $XX billion as of December 31, 2012 and $XX billion as of June 30, 2015”

Extra credit for anyone who can work the notes in questions 7-9 into your legends. More extra credit for approaches that are not hard-coded.

1. Create a dataset where an observation corresponds to an RSSD and your variables are the values of Assets, ROA, ROE and AGR at each quarter. Thus, you should have 11 variables describing Assets, 11 variables describing ROA, 11 variables describing ROE and 11 variables describing AGR.
2. For each observation, i.e., RSSD, create variables for the mean, median, min, max of the following: ROA, ROE, and Assets. Also include a variable with the number of observations that the RSSD shows up in.
3. For all of the RSSDs that are observed in at least 7 quarters, compute the compound annual growth rate (CAGR) of assets – defined as

CAGR = (Assets in last quarter observed / Assets in 1st quarter observed)**1/# years in calculation** – 1

# of years in calculation is the # of quarters (including first and last) between the last quarter observed and the first quarter observed divided by 4

So, # of years is 2.75 (i.e., 11/4) if a particular RSSD is observed in both the first and last quarter of the data.

Produce a histogram of the CAGRs for the banks you included.