

The following is an edited version of the questions and answers given during the first Q&A session for the FDIC Academic Challenge on October 7, 2020. Thanks to all participants in the conference call for their engagement and interest.

**Q: Can we use data other than the base dataset that FDIC provides?**

A: Any data that are public data are fine to use as part of an Academic Challenge submission. Public data are data that are equally accessible to all people in the public. If a university has paid for access to a dataset, these data should not be used as part of an Academic Challenge submission. On the other hand, data downloaded from the website of a U.S. government agency, such as the U.S. Census Bureau, and merged with the FDIC-provided dataset are freely available for anyone to download and fine to use as part of a submission with proper attribution of the data source. A conversation with a loan officer to gather insight on how community banks or larger banks make loan decisions is also a fine source of information. If there is any uncertainty about whether a dataset is acceptable to use as part of a submission, please send a question to [AcademicChallenge@fdic.gov](mailto:AcademicChallenge@fdic.gov) along with some information about the dataset for a determination. A link to the webpage where the data reside or documentation that describes the data and its original sources will help produce a timely answer.

**Q: What specific things are FDIC most looking for in the written submission?**

A: The Grading Rubric on the Academic Challenge website shows the broad categories on which submissions will be evaluated. Is the submission well-organized? Is the theory reasonable and thought-out? Does the submission present interesting ideas about how community banks might affect economic development? Does it lay those ideas out clearly? Does it use the data to support the theory?

Specifically, the Rubric says that the judges will evaluate whether conclusions are supported by data, and if any extra data that a team might bring in are taken from a trusted source. They will evaluate if the information presented is accurate, whether the team uses data correctly in arriving at conclusions, and whether the method of analysis is appropriate for the question. They will look for teams to address the prompts with clear and organized writing. Arguments should be reasonable and compelling, and visualizations should support the text. When assembling effective graphs and tables, make sure visualizations are self-contained. A reader should not have to refer back to the text to understand what is shown in a graph or table.

The last part of the Challenge concerns the FDIC and policy recommendations. This requires an understanding of the FDIC's mission and approaches that the FDIC can use to accomplish

goals within the banking sector. Overall, the judges will be looking for something well-written and something where the data can be used to illustrate the conceptual points made in the paper.

**Q: Is FDIC expecting teams to look at time periods outside of the time period of the data provided or examine the effects of COVID-19 on banks?**

A: The decline in community banking in the U.S. is a long-run phenomenon. This decline may be very relevant to the response of communities today to COVID-19 and related challenges. The data that FDIC provides run from 1994 to 2018. The time period selected for the question was determined, in part, because several datasets are combined, and recent data that might show impacts from COVID-19 are not yet available for all the variables included in the dataset. If a team wanted to use a reliable source of more recent data to analyze the effects of the pandemic, that is certainly a possibility. The data that FDIC provides do not cover recent months, so a team would need to consider the best way to align time periods or obtain additional banking data from more recent months.

**Q: How should teams define “local” when considering local economic development? Is local limited to a single county, or can teams choose to expand the region?**

A: Teams may choose whatever definition of local is most appropriate for the outcomes they are examining. The FDIC-provided data are at the county level, but this reflects the smallest geographic unit available in many common public datasets rather than any preferred definition of “local.”

Using counties as the main geographic unit under study is certainly a reasonable choice, though teams should be aware that counties can have greatly varying populations and geographic sizes. Additionally, county governments may have different policy-making powers in different states. For some outcomes of interest, using counties may be appropriate, but for others, aggregating observations up to the state level may be a more reasonable choice. We anticipate that different teams will make different choices.

The provided data can be aggregated up to state or other regions as needed, but see the question on data structure below for issues that teams should consider before doing so.

**Q: Is there a preferred approach that teams should use when answering the question? Statistical, graphical, qualitative?**

A: Teams should use the approach that best allows them to illustrate the ideas contained in their conceptual framework. A well-thought-out framework with some graphs that provide strong suggestive evidence in favor of the theory could form a very strong answer. Teams should not feel like regression techniques are a necessary component for a strong answer. An approach is successful if it presents evidence that highlights a team's thinking on the question.

**Q: How does FDIC measure banks' market presence?**

A: There is no single best metric of market presence, but there are some common choices. Often, this is done using relative deposit shares, as this provides an indication of which bank or set of banks is holding the most money in a community. Alternatively, market presence could be measured using the number of physical bank offices in an area. Each of these measures has some strengths and weaknesses. Deposit shares are dollar-weighted, while bank offices might capture service provision or the place of a bank in the consciousness of the community. Teams should consider which measure best fits their conceptual framework and their outcomes of interest if they wish to compare market presence across banks or types of banks.

**Q: Is it necessary to address all of the example subquestions in each part of the Challenge question? Should teams focus on these subquestions in their answers?**

A: The subquestions are examples of questions teams might answer while addressing the main question posed in each part of the Challenge question. Complete answers to all of the subquestions would be very difficult to provide within the page limits of the Challenge, so teams are not expected to provide detailed answers to every subquestion.

Instead, teams should focus their efforts on addressing the issues in the main prompt for each part of the Challenge question. For example, Part I asks teams to “develop a concise conceptual framework” and “use this framework to motivate hypotheses to illustrate with data.” Accomplishing these goals is more important than specifically addressing any of the subquestions raised below this prompt.

The subquestions provide a way to test the usefulness of a framework. For example, Part I says that the framework should discuss “potential ways in which the decline in community banking could affect outcome measures of your choosing.” One of the subquestions poses a more specific issue: “what should we expect to see in the data in areas where larger banking institutions have a large market presence relative to community banks?” This subquestion prompts teams to consider how to test the hypotheses generated by their framework. As such, answering this question is a natural part of setting up a framework that will be useful in

addressing the rest of the Challenge. Many of the subquestions are best thought of as issues to consider when answering the main prompt, but they do not need to be specifically addressed on their own.

**Q: What are self-contained figures and tables? Should text be included with figures and tables?**

A: Figures and tables should have enough text included that a reader looking at a visualization does not need to return to the main text to understand what a visualization is showing. This should include an informative title, clearly labeled axes and rows/columns, and a footnote underneath that gives more details on the data sources and variables appearing in the visualization. As a check, teams may wish to show only the figure/table to their faculty advisor, and ask the faculty advisor to describe what is in the figure/table. For a self-contained visualization, it should be possible for a person to describe what is in the visualization while looking only at the visualization. Providing appropriate text to describe the visualization is an important part of making it self-contained.

**Q: Are there any features of the data that teams should be aware of when preparing their answers to the Challenge?**

A: The data that FDIC provides for the Challenge have one observation for each bank-county-year. That is, an observation contains information about a particular bank, in a particular county, in a particular year. Teams should be aware, however, that not all the data provided actually vary at this level.

Specifically, data from the Summary of Deposits vary at this level—different banks have different deposit amounts and different numbers of bank branches in each county and each year. However, data from the Call Report vary only at the bank and year levels, not at the county level. For example, a given bank in a given year will have the same dollar amount of commercial and industrial loans listed for every county. The data do not distinguish which loans were made in which counties. Similarly, data from the Bureau of Economic Analysis vary only at the county and year levels, not at the bank level. For example, data on the number of jobs in the farming sector are identical for all banks in a given county in a given year. Depending on the specific outcomes that teams are interested in studying and the approach that the teams elect to use to study those outcomes, teams will want to consider whether this data structure might affect their conclusions and how they might need to adjust the data to make sure that their conclusions are sound.

**Q: Are savings and loan institutions included in the data provided by FDIC? Are credit unions?**

A: The provided dataset includes institutions that have deposit insurance with the FDIC and submit their Call Report information to the FDIC. As a result, credit unions are not included in the provided data, but savings and loan institutions are. Teams interested in adding data on credit unions should refer to the National Credit Union Administration (<https://www.ncua.gov/>) for information very similar to that on the Call Report collected by FDIC.