**Exercise ideas:**

**--Visualize some data and have the group come up with questions**

**--go through some recent stories or items from the paper and explain how to approach them from a data state of mind**

**--use a dataset to walk through the process of thinking of your questions (in English) and then figuring out how to answer them in software**

**--show the difference between summary data and raw data**

**CAR editing road map:**

**Beat development:**

* Encourage reporters to find their data “sources,” add them to their Rolodex, take them out for “coffee”, etc.
* Give them time to explore a dataset that is crucial to their beat; perhaps generate a minimum story from it

**Idea generation:**

* Questions you should be asking/Things to keep in mind:
  + Story idea should be framed as a question, not a noun. For example: “I want to do a story about how many bridges are deficient” is better than “I want to do a story about unsafe bridges” because the question helps point you to the data that will help tell your story. It’s okay if there’s more than one question, especially in the beginning.
  + Tipoffs for you that maybe the reporter should consider their own data analysis:
    - The reporter has some summary numbers but there are many questions that can’t be answered with the summary numbers. For example: let’s say the idea is to look at whether some neighborhoods have had an uptick in crime recently, but the police department is only able to provide the change over time for the city as a whole. Or they can only show the change from last month to this month and you think there’s a bigger trend (a better story would be to compare a recent time period to the same months in previous years)
    - The reporter is unable to find summary data that answers their question and/or confirms that nobody has done that sort of analysis.
    - “Myth” stories are good contenders for data analysis. For example, there’s an on-going myth that the holidays in December are the “busiest” days at the MSP airport. But if you look at the data about how many planes are coming in and out, that’s wrong – it’s actually the first week of June, right after school lets out.
    - Putting breaking news events into context. When the 35W bridge collapsed, we used inspection data to find out how many other deficient bridges were out there and where they were. When there’s a bad car accident, it’s possible to get crash data to find out if that intersection is problematic, etc.
  + There’s a balancing act you need to consider in deciding whether to pursue data analysis for a story.
    - One factor is time – how quickly do you have to turn this around
    - Another is whether somebody else has already done the analysis that you want (i.e. a government agency or non-profit).
    - Another is how complicated the question is. For example, if the question is whether the percentage of students on free or reduced price lunch has gone up in recent years and where it’s gone up the most – that data is very easy to obtain and answering that question would only take a couple hours, at most. But if you want to prove that county workers are “spiking” their pensions by working excessive overtime in their last five years on the job, that data is harder to obtain and this question would require a more complicated analysis.
    - Another is how important the story is.
  + If you’re going to launch into an analysis that is even a little bit complex, always, always make sure the reporter spends time looking for examples of similar analyses that could serve as a template – i.e. other news organizations, non-profits, academics, etc. There’s a very good chance somebody has done something in the same vein and you could learn from their work, rather than reinventing the wheel. A good place to start is IRE’s Extra! Extra! Website (publicly accessible) or the story database in the Resource Center (requires IRE membership). Also check the IRE tipsheets for guidance on the topic (also requires IRE membership)

**Finding/Requesting data:**

* Number one rule: Not all data is on the Internet. In fact, usually the good stuff is not.
* The existence of summary data or reports that lay out analysis on the topic tend to be good tipoffs for where you might find the underlying (“raw”) data.
* Ask how readily accessible the data is. If the data will require a public records request, be prepared for it to take longer than you might expect.
* Reporters should do some “reporting” in advance of making a request. They should have a good understanding of what’s in the data (what’s public, what’s not) and also what might not be in the data.
* It’s best to get as detailed data as you can (not summarized) and as many fields of information as are publicly available. Whether or not to get more than one year worth of data is dependent on the topic. My rule of thumb is that 5 to 10 years’ worth of data is the minimum for doing anything that looks at trends over time. A common rookie mistake is to request a subset of data because that makes the request process easier, but then later they discover that it limits what they can say from the analysis.
* If it’s necessary to make a formal Data Practices Act or FOIA request, see my handout on “Using the Minnesota Data Practices Act” for guidance on the wording of the request letter and other tips, including how to deal with cost and denials.
* It is likely, although not a given, that you may have to pay for the data you request -- regardless of the level of government you’re dealing with. Rule of thumb: if they are saying it will cost more than $50, something might not be right. In Minnesota, they are allowed to charge the hourly rate of the lowest-paid person qualified to do the data export (regardless of who actually does the work) and can only charge for the time the person spends actually working on it. They can’t charge for the time that the computer is running the export, can’t charge for separating private from public and can’t charge for any time they spend verifying that the data is correct. As a result, it shouldn’t take more than 1 hour for them to export your data (and most IT people are paid roughly $50 per hour)
* Other questions you should ask: Who collects/maintains this data? Do we have a name and contact info for the actual data person? What is the process for requesting the data?
* What if the data doesn’t exist? There may be times when there is not an existing database of what you want – but maybe there are paper records – or you’re having trouble getting a request fulfilled, but the data is sitting out on a web page. Number one rule: don’t just start typing it in by hand. There are ways to scrape web pages, pull things out of PDFs or get paper records turned into data.

**Reporting/Data analysis:**

* Ask about the quality of the data. Does it need to be “cleaned.” If so, build in extra time for that cleaning. Did we get all of the data we expected? Are all the fields there?
* As the reporter starts the analysis, ask some questions about the data to get a feel for how well the reporter understands the data. If he/she can’t answer your questions, encourage them to seek out someone who knows the data and can answer those questions. Some potential questions to ask: what does each record represent? What’s the time frame of the data – is it just one year or multiple years? Are there any fields that you don’t understand? Are there fields we can’t use because they are incomplete or inconsistent?
* Ask the reporter if they’ve done some data integrity checks. If possible, pull a few paper records to match to the database to see if the information is consistent. Run summaries and try to match them up with summaries that the government may have generated.
* Encourage the reporter to come up with some questions and working theories. Doing pre-reporting (with human sources and other paper trails) before analyzing data is something that is often overlooked by rookies. It’s kind of like prepping for an important interview. You need to have some general questions in mind and an idea of what direction you want to go with the “interview.”
* Encourage the reporter to keep a “trail” of their data cleaning and analysis. This trail should be a form of documentation that you would be able to understand and would enable an outsider to recreate the analysis on their own
* Sit down with the reporter and have them show you the data and the analysis they’ve done. You may want to repeat this multiple times during the process (if it’s a larger project)
* Don’t wait until the last minute to show key sources the findings; in fact, that should be the first step in the reporting process. Often the sources know the data better than you do and they might point out something that you missed or didn’t understand.
* Get digital and graphics people into the loop to discuss data visualization and/or other digital options as soon as you have the analysis defined. It doesn’t need to be perfected, but just so that you don’t have any major lingering questions or problems.
* After the reporter has done the first round of analysis, have them write a 2-3 graph summary. That will help both of you see where the holes are. If the reporter can crank out this summary in a few minutes, then you’ve got good reason to believe the analysis is sound. If they struggle and find it difficult to phrase their findings, the analysis is probably a little too much like Swiss Cheese.
* If something seems too good to be true, it probably is
* Do the findings make sense? A newspaper published a story about the Lhasa Apso being the most popular dog breed (according to the city’s pet licensing database). But it turns out the reporter erred in sorting the Excel file.
* Be sure the reporter repeats the analysis, start to finish, to ensure accuracy
* The reporter should attempt to prove the key finding wrong before going ahead with it. Doing data analysis is kind of like being a police detective. You can’t go into interviews with preconceived notions that are set in stone. Once you find the direction you want to go, you also need to see if you can prove it wrong to ensure you have a solid case.
* Expect that the data analysis might need to be revised or repeated during the reporting and writing process. As the reporter gathers new information from human sources, that will put the data into a new light.

**Writing/Editing:**

* Numeracy—“star number”; use graphics to share the details; avoid too many numbers in same sentence or paragraph; you truly don’t need decimal points (almost never). Fight the urge to throw every number into the story.
* What’s the strategy for putting “life” into the numbers? Your goal should be to have a story that is supported by data but is not solely focused on the data (with some very rare exceptions). Data will either be the “spine” of the story (if that’s what’s driving the content) or it will be used in smaller bits to support key parts of the story.
* Do the math
* Do we have the necessary context? Do we have the “compared to what” answered?
* Are we portraying the results honestly and not making too much out of it?
* Do our results contradict what our sources are saying? Or are there plausible explanations for our results (which essentially mean you no longer have a story)?
* Don’t be afraid to consult outside experts, especially on advanced statistical analyses
* As assignment editor, you should review any digital content, just like you’d review the story. Make sure that it makes sense, that it complements the story and consider how friendly it is for the reader (usability is particularly important if it’s an interactive element)
* Make sure that it also gets reviewed by a copy editor just before publication.
* See bulletproofing guides for more