Daily guide for projects

Summary

This project plan explicitly encourages the iterative nature of research as a series of questions and answers that gradually refine your hypotheses. You will start with the D1-D2 course content, which teaches you about modelling, neural data and the scientific process. You will then apply the steps taught there to a vaguely formed hypothesis or topic of interest to you. This will not be perfect, because you still don't know much about the data or about a phenomenon of interest at this point. In D3-D9 you will develop your knowledge about your topic, and frequently go back to the cheat sheet of research steps taught in D2. During D10, you will know enough to be able to write a short abstract about your project, which may or may not include results, but it should at least include a testable hypothesis. For the rest of the project on D11-D15 you will focus on getting evidence for/against your hypothesis. Finally, in D15 you will meet with other groups in your pod and superpod, and tell them the story of your project. This is a low-key presentation that may include some of the plots you made along the way, but it is not meant as a real research presentation with high "production values". See some of the examples from last year to get a sense of what the presentation will look like.

Project templates

Project templates are research ideas developed by the NMA team that can be used in conjunction with the datasets we provided. Project templates can be used in a variety of ways.

- For starters, you can use the project templates just to get familiarized with some of our datasets or one of the provided models. They can provide keywords for you to use in your proposal on D2, or python libraries you can reuse to answer your own questions.
- You should use the project templates extensively if you are new to neuroscience and/or you don't have a lot of research experience. They have been designed to give you enough structure to get started, and enough options to keep you going if you stick with the template. Or you may start with a template, use it the first week and then in the second week diverge from it as your group develops their own new idea or question to test.
- Templates have a natural flow of questions, but don't hesitate to skip or completely change some of these. They are meant to be used very flexibly!

Project TAs

Project TAs are a new role at NMA this year, and they are your friendly dataset experts to consult with on all issues related to datasets. They can also help with other aspects of a project, including brainstorming, literature searches and coding. You will have a one-hour meeting with one of them on your first two project days. During this time, they will help you refine your question and hypothesis into something that can be answered with our datasets. Since they can arrive unannounced at any time (busy schedules!), please stop what you were doing to have the meeting, and then resume your work when the project TA leaves.

In later days, project TAs will be assigned meetings with junior groups, but can also be recruited to senior groups for meetings when you need them. For the assigned meetings, project TAs will generally come only during project times, but sometimes they might need to schedule meetings slightly earlier or later. We encourage you to reach out to them for extra meetings whenever you need them, and to post questions on discord in the #dataset-X channels. All project TAs have time set aside specifically to answer discord questions and to provide additional meetings when necessary.

Project Block 1 (3 h)

This time block happens either after coursework on Monday (timeslots 1,3,5) or before coursework Tuesday (timeslots 2,4).

Goal is to decide what dataset/topic to work on and to form two separate groups (4-6 students/group). TAs make a Miro concept board for brainstorming at the beginning of this session and they invite the students to join during the brainstorming times below.

- Introductions (30 min = 2 min/student): say a few things about yourself, then about your research area or research interests. What are you really curious about, that you might explore in your NMA project?
- Listen carefully as others talk about their interests. If you are curious about something, ask them. If their interests match yours, you can try to recruit them to your group later!
- Individual reading time (30 min): browse the projects booklet which includes dataset details and project template details. Watch some of the videos that are of interest to you.
- Brainstorm at the pod level (60 min): do you have an idea or a rough direction? Bring it out to the group. Start recruiting others to your group!
- The TA will pay close attention to what you all are interested in, and will produce an initial set of two topics / datasets / ideas, written down on the Miro board. join one of these by writing your name down underneath the topic.
- Tentatively separate into groups (not final). Now brainstorm within your group (60 min, breakout rooms). Definitely choose a data type and choose a specific dataset from that data type if you can. Make sure the dataset you choose is well suited for answering the broad range of questions you are interested in. Try to come up with one or a few topics of interest, either by yourselves or directly from the booklet (i.e. project templates).

Tips:

- Use the project booklet like you would a conference booklet: skim through it, read a paragraph here and there, try to decide which of the videos to watch. No one is expected to read through all the materials.
- No need to have a very concrete project after this day. You will determine the feasibility of your questions in the next few days, and you will likely change your question completely. That's how research works!

W1D2 coursework (Tuesday) + project block 2 (Tuesday for slot 1,3,5 / Wednesday for slot 2,4)

The W1D2 tutorials are set up to launch your projects. Once you're done with them, you will complete your literature review. During the next available project block, you will finalize your project groups, and create your project proposal: a short, unstructured proposal about what you'd like to do. We will use the project proposal to match your group to a mentor based on their research interests (think of the matching as done by "keywords", but in an automated manner).

(3h) Complete the intro/tutorial/outro for this day

- You will need to use your own project for some of this content. If you don't yet have concrete ideas, or you haven't done a research project before, use one of the provided project templates to walk through the four steps.
- If you are using a project template, your goal is to translate the information from the slide and colab notebook into a 4-step format. Some information might not be readily available in the slide or notebook, and you might have to find it in your literature review later this day.
- Try to write down a few sentences for each of the four steps applied to your project. You will re-use these in your proposal later today.

(2.5h) Literature review: identify interesting papers The goal of this literature review is to situate your question in context and help you acquire some keywords that you will use in your proposal today.

- (30min) on your own, start doing a literature review using google searches and only look at abstracts to select 2-3 promising ones.
- (10min) report to the whole group what papers you found and pool them together. Assign one paper per person to read/skim in the next 1h.
- (1h) on your own, read the paper that was assigned to you. Make notes in a common google doc shared with your group, and especially write down important keywords or concepts which you might use in your proposal later today. If you are not connected to an .edu domain or a VPN, try to find full versions of papers on preprint servers like arXiv / bioRxiv. You could also ask your TA to get it for you (and they might in turn ask someone who has access to a university VPN). There might be other options too...
- (1h) report back to the group, and try to tell them as much as you understood about the paper. Get into details, but don't just read to them whole sections from the paper. Ask the other students questions about the papers they are presenting to understand them better.

Project block task: (3h) Project proposal

- Try to write a proposal for this project based on the way you understand it now. This should re-use some of the text you wrote down for the four steps, and should include keywords and concepts that you identified in your literature review. Don't worry too much about the structure of this paragraph! The goal is to get as many words (200-300) on paper as possible. You have the entire day 10 to learn how to write a properly structured scientific abstract.

- It is important to include the concepts which you identified as relevant, and the keywords that go with them. This will allow us to identify mentors whose research covers your interests.
- You should submit your proposal into the Airtable form at the end of this day. Remember that only an algorithm will read it, so it definitely doesn't need to be perfect!

Block 3 (2-3 h)

You should now get your hands dirty by running some of the notebooks provided for your dataset. There should be a loading notebook, and there should be some further analysis notebooks interspersed among the project templates. Even if you are not using a project template, they are likely to contain useful code to get you started, especially if you are doing a theory project.

*...but first, check your inbox! You will receive mentor assignments at some point late in the day, with a first meeting scheduled for Friday. Reach out to your mentor to confirm the first meeting or reschedule if it's necessary for them. Also try to arrange a meeting for day 10, ideally the second half of the day, when their feedback on your abstract could be useful.

- Most of this day you should be getting familiar with the data/model, and making exploratory plots.
- If all you do is run the provided notebooks that's fine! Try to understand what is being plotted and how. Pay attention to the code libraries being used, and the way the data is accessed / binned / aligned. You will reuse some of these code elements to start doing your own analyses in later days.

Blocks 4-5

Over the next two weeks, you will iterate between refining your question and trying to answer it. Be on the lookout for interesting hypotheses. You might notice something weird in the data, and if you dig deeper it might lead you directly to a result. For this to work, you must keep an open mind about what your questions are. If you feel like your question is starting to change, go back to Steps1-5 and see if it's easier to formulate those steps with the new question. A good question/hypothesis makes the 5 steps really easy to think through. Here are some generally useful tips & tricks:

- The hardest part will be wrestling with the data to try to answer your question. You can rely on your TAs, the dedicated project TAs and the Discord channels to make this process easier.
- For theory projects, wrestling with your model can be equally challenging. If your model generates data, for example a neural network simulation, then you can still use some of the tricks below to analyze that data.
- If your model makes a hypothesis that needs to be tested, then your theory project might become a data project. The opposite may also happen: you may find something interesting in the data, and realize that you need a model to understand it better.
- Always be on the lookout for bugs in your code, or "bugs" in your analysis plan. If a plot/result looks too good to be true, it might be! Make sure you always split your data train/test, even for simple analyses where you think it might not matter (i.e. for making tuning curves).
- If your question does change, remember to always do a quick literature survey (i.e. google search) to see if others thought about your question in the past. You don't need to come up with a completely original question! Do however situate your research within the relevant literature, and try to get hints/suggestions from other papers.
- Depending how complex your question is, there could be several data analysis steps:
    - data wrangling: some questions can be answered simply by plotting the right variable from the data! Some generally useful strategies: make PSTHs and tuning curves; try scatter plots of different variables; plot across neurons or across trials; select the most tuned neurons and look just at those; if there are multiple sessions pick a good one and dig deep into that one.
    - simple, linear analyses: most questions can be answered at this stage. This is often needed if you are doing a "population analysis", i.e. trying to determine if a set of neurons or voxels collectively encode a certain variable. By far, the most used linear analyses are linear regression, PCA and k-means clustering.
    - Linear regression is often a good first step, even if your variables are binary/categorical. Once you have a pipeline, it will be easy to switch to logistic regression or other predictors from the scikit-learn library.
    - For visualizations, you might want to reduce a population of neurons to just a few components using PCA, then go back to the "data wrangling" steps and make the same kinds of plots for PCs that you made for neurons.
    - Another way to reduce the size of data is to cluster neurons (or trials!) into a few subsets, then average within that cluster. The simplest clustering model is k-means, which is a "linear" clustering model.
- complicated, nonlinear analyses: if the simple analyses fail, you might think you have no choice but to try something fancy, like deep learning or ISOMAP. This is often a dead end where projects go to die! You probably will make a lot more progress by slightly (or greatly) changing your question, or refining your

hypothesis. The reason complicated analysis are so hard to do and interpret is that they often function as black boxes that are hard to look into. What do the parameters of a deep neural network mean? That is a hard research question in its own right. This is not to say that your hypothesis cannot be a nonlinear model, just that you can often test nonlinear hypotheses with simple, even linear analyses. If you must, however, use complicated analyses, then deep learning "replaces" linear regression, t-SNE / ISOMAP replaces PCA, and hdbscan replaces k-means.

- deep learning as a prediction tool. This is unlikely to do better than linear/logistic regression, because neural data is noisy, and you need a lot of training data to really train a deep network well. This is because deep networks have a lot of parameters.
- There are many "nonlinear dimensionality reduction" methods like t-SNE / ISOMAP, but these are often not meant as replacements for PCA, but instead as visualization tools to try to see a clustering structure in your data. They can be useful for making hypotheses based on interesting-looking plots. You still need to validate those hypotheses using simpler methods, like clustering and PCA.
- There are many nonlinear clustering models like hdbscan and spectral clustering, but those are fickle for high-dimensional data and difficult to interpret. You will have to carefully try different parameters, and think through what the clusters mean.

Block 6 (2-3h)

You should now have a sense of the data, and you have probably refined your hypothesis a little. You might have a vague idea of what it would take for your project to work, what tools you might use, and what the answer could look like. Let's make these things explicit, by continuing with steps 6-9 of the modelling practice in this steps 5-10 notebook.

- (0.5h) Go through the first five steps again with your own refined question. Try to write down what the steps looks like for your data.
- (1-2h) Go through steps 6-10 for the example project in the new notebook and watch the videos.
- (1-2h) The rest of the day, start thinking what these steps would look like for your project, without actually doing the steps. Do you need to select a toolkit and where can you find some options? Do you need to implement a model? Don't actually implement the model on this day! Try to complete the low-hanging fruit first, because you'll have the rest of this week and next for the actual full implementation of your project.

Blocks 7-9 (2-3h / day)

Leading up to day 10, you will implement Steps 5-9 in your project. If you are already experienced with research projects, this might just look like a continuation of last week and you don't need to stick to the steps too closely. If you are not so experienced, you could benefit from implementing the steps one after the other, perhaps at a rate of 1 step / day.

Day 10 (8h, Friday for everyone)

Abstract writing day! One of the best ways to understand your own research is to try to write about it. You should write early and often, not just at the end when you're trying to write a paper or your thesis. Science conferences are a great way to present your intermediate work, and they give you a chance to write an abstract. For example, the Neuromatch Conferences are a great venue for this. However, we don't have to wait so long to write our project abstract, we'll do it today. If you already have an answer to your question that's great, but it's not necessary. Most of the components of an abstract do not in fact require results. There are several materials associated with this day, to be used loosely and when needed. The goal is to workshop your abstract with your group, and then present this to your entire pod.

If you have been using a project template, this is a good time to branch out and pursue your own questions. The template was only meant to get you started on some concrete analyses, so that you become familiar with the data, but now that you have more knowledge, you should be able to come up with your own question. Practice the 4-steps again if necessary, they should be easier once you have a good question.

With your group

- (20 min) discuss situations in which you've had to write an abstract before. This "short format" is not only used in academia, but also in industry. If you feel confident, show this abstract to your group.
- (30 min) by yourself, read the "10-steps to writing a paper". Those are also the 10 steps to writing an abstract.
- (30 min) workshop together one of the provided abstracts from some of the papers of our mentors. Say what you like and what you don't like about it. Does it follow the 10-steps recipe closely and if not, can you re-write it in such a way that it does?

- (1h) As a group try to write a first draft of your abstract in a google doc. 30min break
- (1h) Edit the abstract individually in your own google doc. At this stage, it is also important to control the flow of the abstract, in addition to keeping the structure from the 10 steps-paper. The flow relates to the "writing style", which is generally no different for scientists than for other writers. Most importantly, make sure each sentence picks up where the previous one left, and do not use jargon without defining it first. Use this quick reference about writing, and make a note to check some of these great books about writing later (book1, book2, book3).
- (30 min) You should now have as many copies of your abstract as there are students in your group. Put them all into the same google doc, and try to see what you all did the same / differently. What sounds better? Pick and choose different sentences from different abstracts.

With your mentor (timing is not precise!)

- (30-60min) Try to schedule a meeting with your mentor to be about now (or any time in the second half of this day). Show them your abstract. Try to get explicit feedback, and ask them to edit their own version of the abstract.

With the pod

- (30min / group = 1.5h) It is always revealing to present your research to someone who has never heard about it. Take turns in your pod to read each other's abstracts (out loud and shared google doc). Tell others what you understand and what you don't from their research project.

Back in your group

- (1-2h) Has the abstract refined or changed your question? Use the rest of this day to make a concrete plan for the final week of your project. If you already answered your question, then you will need to plan control analyses, maybe including some simulated data that you need to also generate yourself.

Blocks 11-14

Abstract writing day should have helped you narrow down what results (positive or negative) you would actually need to answer your question. You will use the rest of this time to try to get a result, or make progress towards an answer. This might not work out in such a short time, but don't get discouraged: this part normally takes months if not years of work.

- If you know what analysis you need, but don't know how to do it, the TAs are there to help you. They can point you to useful toolkits that may be difficult to find otherwise.
- Try not to implement complicated analyses from scratch. Use existing toolkits, and learn how to use them well. This kind of knowledge is very helpful long-term.
- If you find a negative answer to your question, that is absolutely ok! Please do report that. Then go back and think about how this affects your initial hypothesis. Does it rule it out, or could there be limitations in this particular data that lead to the negative result? What other data would you collect that would be better suited for answering this question. Try to design a new experiment in very specific detail and tell us about it. Who knows, somebody might run that experiment someday!
- If you find a positive result to your question, then you should spend the rest of your time validating it to make absolutely sure it is really true. You will need to design controls using the data (shuffling controls), or using simulated data, and you need to check the logic of your pipeline from start to end. Did you accidentally select only neurons that were tuned to a behavior, and then showed that they respond to particular aspects of that behavior? Did you sort neurons by their peak response time and then found sequences in your data? There are some obvious and some not-so-obvious circular analyses that can catch even experienced researchers off-guard. This is what the controls are especially useful at catching.

Day 15 (Friday tutorial block for everyone)

- This is the day where you tell your superpod about your project. The groups will take turns to share their screens. You can use figures and other graphics, but this is meant to be told as a story, and everyone from your group should take a turn telling a part of the story. Tell us about the different hypotheses you've had at different points and how you refined them using some of the tools we taught. If the tools were not useful, tell us why, and tell us what you did instead to make progress, or to get unstuck when you didn't know what to do.
- Most groups won't have a result and this is absolutely normal. However, the main goal anyway is to communicate the logic of your project proposal. Did you design a smart way to test the neural binding hypothesis, but then didn't find the data to get answers? That can also be very interesting for others to hear

about! Furthermore it will make it clear that research never stops. It continues as a series of questions and answers, not just within your own project, but at the level of the entire research field. Tell us what got you excited about this particular project, and try to dream big. One day, models like yours could be used to ... ?
