

Biology 125 Lab Manual

2021-12-01

Contents

Welcome	5
Labs	5
Assignments	6
Absences	6
Copyright	6
Conventions	7
 Lab 1	 11
Welcome	11
Your first project	11
But first...	12
 Research Project: Part 1	 13
Meet Elliot	13
Finding property	14
Your mission	16
 Research Question	 17
A good research question	17
Background	18
Phrasing	18
An Example	19

Research Data Management	23
Assignment: Lab 1	25
Lab 2	29
Open Science	29
Lab 3	33
Research Project: Part 2	33
Overview: Lab 3	33
Assignment: Lab 3	35
Lab 4	39
Research Project: Part 3	39
Lab 5	43
Research Project: Part 4	43
Recommendation Report	45
Science writing	45
Preparing to write	46
A good report	46
RMarkdown and RStudio	50
Rubric: Lab 5	55
Assignment: Lab 5	59

Welcome

To your Biology 125 Labs!

First, a few important and relevant links...

- Canvas course shell
- Syllabus

Second, some key pieces of information for how labs will be run this term...

The following pages detail how labs, assignments, and absences will be handled. In addition to the content here, the lab manual, assignments, quizzes, and additional resources are all supplied on Canvas so take some time to look over the information.

If in doubt about course policies, refer to your syllabus or talk to your TA. To get started read through your course syllabus and the lab manual.

Above all, have fun this term and enjoy the process of science!

Sincerely,

Dr. Tristyn Hay

Labs

Similar to term 1, labs will be alternating between on-campus and online, with some online labs being conducted in Zoom during your regular scheduled lab section. In order to be successful this term please pay attention to your syllabus where you can find your schedule, mark breakdown, missed lab policy, and all lab requirements for your BIOL125 lab. You will need the information found in this syllabus to be able to answer the Introductory Quiz on canvas. Students will not be excused of any missed or late assignments due to being "unaware" of these policies so be sure you reference this information often.

Your first lab, Lab 1, is synchronous - meaning that you complete it on-campus - with the relevant material being posted on Canvas. You will have the opportunity to meet your TA and the other students registered in your lab section. This week will be a very important week as you will have your partner assigned to you and will be setting up for an experiment. For in-person labs, it is vital that you attend the lab for which you are registered - check your registration online to be sure of where you are supposed to be!

Your second lab, Lab 2, will be asynchronous - meaning you will work on it in your own time and online. Your TA will be available online during your regular scheduled lab sections if you have any questions or concerns.

During the synchronous labs you may bring a lab coat but this is not required. Closed toed shoes and pants are required for this lab and your TA will turn you away if you are not in proper attire. Any goggles and gloves will be provided in lab.

Assignments

See Canvas for due dates.

All assignments, even those worked on as a group, will need to be submitted individually unless directed to do so by your TA.

Failure to submit an assignment will result in a mark of 0 for that assignment. No exceptions!

Please be sure you have not only submitted assignments on time but that you have double-checked they have been uploaded properly.

Absences

Pay attention to your absence scores - see syllabus for details. Students receiving an absence score over 5 will fail the lab. This applies to both on campus and online assignments. If an assignment is not submitted it counts as an absence. Excused absences get a score of 1 and unexcused a score of 2.5.

Absences can only be excused by myself (Dr. Hay) and not your TA.

Copyright

This work is licenced under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Please use the following for citing this document

Hay, T., Vis-Dunbar, M. (2021). *Biology 125 Lab Manual*. <https://ubco-biology.github.io/BIOL-125-Lab-Manual/>

All source files are available <https://github.com/ubco-biology/BIOL-125-Lab-Manual>.

Conventions

Information relevant to lab logistics and grading.

Further insights or notes on presented materials.

Highlights and key take aways.

Optional material that dives deeper into a presented concept.

Lab 1

Welcome

Last updated 2021-12-01

Welcome to your first Biology 125 Lab for Biology for Science Majors II!

For your first week your TA will spend a bit of time going over the lab schedule and mark breakdown, along with some safety information you will need to work in this space safely and professionally. Your TA will then get you going on your first project!

Your TA will provide you with their contact information along with office hours and location. It is your responsibility to ensure you have this information readily available.

Your first project

After this brief introduction you will get into pairs - with one group of 3 in uneven class sizes - and your TA will assign you and your partner(s) an environmental problem in which you both will be spending the better part of the term trying to work through.

But first...



Did you know this little guy is actually real??? Anyone guess what these guys are called?

If you have a guess send me an email at trisyn.hay@ubc.ca or feel free to just say hello.

Research Project: Part 1

Study protocol

This research project will have you looking at specific environmental features - climate, exposure, chemistry - to answer a pretty practical real world problem: will a piece of arable land support the cultivation of a specific crop.

You will be expected to design a study, collect data and report on your findings. You will be recording your data and authoring your report following Open Science best practices, namely, ensuring that your work is transparent, reproducible and well documented.

In this lab, we'll be introduced to the problem, explore how to propose a research question to address the issue and scope out a study design that includes what data we'll need to collect, how we'll organize that data, and how we propose to analyze that data. You'll also get things set up to run your first experiment in Lab 3.

Your assignment will be a protocol, detailing the above.

Meet Elliot

Farmer Elliot is looking to acquire a 20-acre farm to grow mung beans. Farmer Elliot has some basic knowledge of farming as he did some residential gardening in the city and thus understands the basic requirements of plants but has limited knowledge of growing mung beans.

You might be asking why Farmer Elliot is pursuing this when without the needed knowledge ahead of time? Let's just leave it as Farmer Elliot is a "act now think later" kind of person.



Figure 1: Elliot. The character behind the farmer. Find more pics of Elliot in your syllabus!

Finding property

In order to start their search Farmer Elliot has enlisted the help of a real-estate company - "llamaste Realty". Their real-estate company has shown them three properties thus far.

Property 1



Figure 2: Property 1.

This property has 19-acres of flat usable land and is situated next to the base of a large mountain. This land comes with fully automated irrigation and a small one-bedroom home. Though this land is start up ready, Farmer Elliot is concerned that the property may not have enough hours of sunlight due to its proximity to the base of the mountain.

In order to help determine if this is a viable option for their mung bean farm, Farmer Elliot will need to know how many hours of sunlight is required for optimum mung bean germination and growth.

Property 2



Figure 3: Property 2.

This property is 21 acres large with 3 acres designated to residential space and the remaining land is usable agricultural acreage. Running parallel to the longest section of this property is a small alkaline lake. Farmer Elliot has been informed that the adjacent alkali lake has, on average, a salinity concentration ranging from 0-5%.

Not knowing if salinity has an impact on germination and growth of mung beans Farmer Elliot is not sure if this is the best property for their endeavor.

Property 3

This property is located at a much higher elevation than the other properties and located in a very different biogeoclimatic zone than what Farmer Elliot is familiar with and the differences in mineral content and soil texture in this area results in soil pH levels ranging between 6.2-7.2. It comes with a 2-bedroom home and 20 acres of usable ready to grow acreage.

This property appears ideal but Farmer Elliot is unsure if this pH range is tolerable for mung bean germination and growth.



Figure 4: Property 3.

Your mission

In order to better inform Farmer Elliot's decision they've hired you and your partner(s) as consultants.

Your TA will assign a specific property listing to you and your partner(s) and you will be tasked with designing an experiment to test the specified variable on both mung bean germination and growth.

Following this you will need to develop a recommendation report for your client Farmer Elliot.

To help you in this endeavour, you may wish to review the content from BIOL116 on experimental design – Designing the Experiment.

Research Question

First things first, we need a good research question.

In BIOL116 you were introduced to experimental design and hypothesis testing. One of the things we didn't touch on in great detail was how to develop a research question.

How you formulate your research question will impact what you study and how you conduct that study.

When we think about transparency and reproducibility in research design and implementation, every step we take and every decision we make is predicated on earlier decisions; and things begin with a research question. Well, to be fair, developing a research question is an iterative process, but it underpins so many future decisions - it will inform your hypothesis – after all, your hypothesis is the testable statement that addresses your research question – which will then inform your study design and so on. So, we shouldn't gloss over it's importance!

A good research question

A good research question will help to limit many biases that Open Science is trying to combat in the conduct of research, including HARKing and making decisions after having looked at one's data.

A good research question is primarily informed by two things:

- Research done to date that has addressed this problem.
- The problem at hand

Background

Consulting the literature

Consulting research done to date will allow you to see how this or similar questions have been addressed by other researchers. Novel ways of addressing the same question are important to move science forward; consulting previous research will help to identify gaps that are opportunities for these novel approaches.

At the same time, consistency in methodology underpins reproducibility, and it's consequently just as important to test the same the questions with the same methods in both similar and novel populations as previous research has done, helping to build a body of evidence and identify if earlier findings are generalizable to other populations.

The problem at hand

The problem at hand will come with certain known and unknown elements. In this assignment, depending on the plot of land that you're looking at, you already know certain things about the soil chemistry, geological features or water sources of a given plot.

It is the unknown elements – or a portion of the unknown elements – that your research question will try and address. In this instance, how these factors will impact mung bean production.

A research question that asks

What is the impact of fertilizer x on the growth of mung beans?

would seem a reasonable first attempt at addressing one potential issue at hand. However, it doesn't give a clear definition of what we're measuring as either a dependent or independent variable.

Since the research question informs the hypothesis, which then guides your design, you're leaving yourself with a lot of wriggle room here further down the line.

Phrasing

A testable question

Using the word **what** re-enforces this less than concise formulation of the research question. In fact, predicating your question with **what** or **why** doesn't allow your question to ask exactly what you need it to ask.

In experimental design, we're testing for relationships - asking "is there a relationship?". In fact, we're asking a question that allows for the proposal of a hypothesis; a prediction of what that relationship might be. So, we should think about how we can ask a question that reflects the test or experiment we're planning.

A succinct question

In addition to re-framing our question to one which is phrased as a testable question, we want to clearly articulate our population of interest and our main variables of interest. When phrased as

What is the impact of fertilizer x on the growth of mung beans?

the variables ostensibly include fertilizer and plant growth. But plant growth is more nuanced than this, and our study might be too. In fact, arguably, plant growth is not a variable, but a composite of variables; so, we should ask ourselves, "what do we mean by growth? What about growth are we interested in? Germination rate, germination survival, biomass, height, flower set, fruit set?"

Defining the scope of your population and variables is a key consideration when developing a research question; defining these early means that you won't be asking these questions later, once you've already started to collect, or work with, your data.

So, ultimately, we want a question that:

- Is testable;
- Clearly identifies our population;
- Clearly identifies our primary variables of interest; and
- Is concise

An Example

Let's say our farmer is concerned primarily about fruit set. It seems reasonable then to test for fruit set. Again, fruit set could be defined in many ways - average biomass per fruit, average count per plant etc. And we may or may not be interested in each of these outcomes. In either case, a more concise, testable research question might then look like

Will the application of fertilizer x increase the quantity of fruit set of *Vigna radiata*?

Compare this with what we had before

What is the impact of fertilizer x on the growth of mung beans?

By specifying a proposed relationship that articulates our variables, not only can we now test this question directly, it identifies exactly what we're interested in testing, and it's concise, which means that we can then readily propose a hypothesis and null hypothesis to address it:

- Ho: fertilizer x will increase the quantity of fruit set of *Vigna radiata*.
- Ha: fertilizer x will have no impact on the quantity of fruit set of *Vigna radiata*.

Reproducibility, meta-analyses, and the evidence base

When we reproduce a study, we always know that there is a chance of error or bias resulting from our sample not being truly representative of its population, for any number of reasons including sampling error, lack of power etc. This is why we should never rely on the findings of just one study.

A meta-analysis is a study of already conducted studies to try and determine if across a series of studies addressing the same research question there is enough agreement in the findings to accept one conclusion, even though this conclusion may be contradicted by individual studies.

Replication enables this aggregation of findings, helping to sift through studies that have suffered from systematic error. To do this well, meta-analyses rely on documentation and homogeneity; studies that use similar methods, instruments, and techniques to address the same question and describe in detail how this was done. This is because comparing two studies of the same phenomenon with two different research questions and two different methodological approaches and data collection tools is extremely confounding and limiting.

Meta-analyses are based on extremely comprehensive literature reviews, reviews that attempt to uncover all literature – published and unpublished – addressing a given research question. Your research question not only informs your hypothesis and study design, it also frames your title and abstract, whether for a lab report, poster, or one day a manuscript. Expressing your research question in a way that clearly and succinctly outlines the variables you plan to test makes the inclusion of your results in a meta-analysis more likely, as your work will be more easily discovered and identified.

In fact, with this in mind, if you were conducting your mung bean research for a particular plot of land in a particular region, this might impact the variables you choose to work with, and you might end up with a still more concise research question that would allow for identification of potential homogeneity and then for comparing your data against other similar studies in a meaningful way. So, for example, in the Okanagan, your research question might be adjusted to

Will the application of fertilizer x increase the quantity of fruit set of *Vigna radiata* in a sandy loam soil of the BC Okanagan Valley?

Research Data Management

Once we know what we want to ask, we need to consider how we're going to organize our project and its data. How we do this - Research Data Management or RDM for short - is a critical component of reproducibility and transparency in the sciences.

In BIOL116, you were introduced to best practices in file naming. You may wish to review that content. In this lab, you'll be looking at best practices in directory structure management; that is, how we organize our individual files.

Just like with file naming conventions, it is extremely important that our files are organized in a way that logically reflects the structure of our project and can be easily navigated with computational tools, allowing for, at a minimum, computational reproducibility. There is also an increased need to provide documentation that describes the chosen structure; in fact, the more complex a project becomes, the more this documentation is important.

So, please review the content on directory structure management in the BIOL Procedures and Guidelines.

Assignment: Lab 1

Please use the following template for this assignment:

20220101_Lab01_125_Protocol-Assignment_V2.docx (17 KB)

Putting it into practice

Drawing on what you learned in BIOL116 and after reviewing the content for this lab, this assignment asks you to articulate the key components of a protocol: a research question, hypothesis, and proposed study time line, as well as to describe the kind of data (variables as well as data types) you'll be collecting, how you'll be collecting it, and what you'll be doing with these data.

Lab 2

Open Science

Last updated 2021-12-01

This week's lab content is the second half of Open Science: An Introduction.

You are asked to cover Parts 2 and 3, Open Science in Action: Benefits, and Open Science in Action: Challenges, respectfully.

The accompanying quiz can be found in Canvas.

Lab 3

Research Project: Part 2

Last updated 2021-12-01

Data collection

In Lab 1 we established our protocol. In this lab, we'll put that protocol into practice and start collecting some data.

Your assignment this week will help prepare you for the report that you will ultimately be preparing.

Overview: Lab 3

During this lab, you will...

1. Run your first trial and collect data (this was set up in Lab 1)
2. Prepare your set up for your second trial

Collecting data

You will need to set up your data recording tool - you sketched this out in your protocol and now you'll need to implement it. Keep in mind that at some point you'll need to submit your data as a `.csv` file. How you record your data now might save you time later on.

When collecting data be sure to take notes on any and all observations. These will be very important in helping you interpret your data later on. Take your time when recording and organizing your data.

Preparing for the next trial

Once you have finished collecting your data you will need to set up your experiment again in order to collect a second round of data during your next synchronous week.

If you are wanting to increase your sample size and thus are doing more trials ensure that you follow the same process as the previous synchronous week. Alternatively, you may be setting up to ask an alternative question. In this case, be sure you have your TA take a look and approve the design / modifications before you leave your lab today.

Assignment: Lab 3

Writing a report can be extremely daunting and confusing especially if you have not written many of them. Many students struggle to know what information goes in what part of a report and in particular what should be included in a results section and what should not.

In order to help you better understand how to write a results section you are asked to edit the results section of the following document:

- Lab-03_Results_Assignment.pdf (306 KB)

This results example section contains at least 10 errors. Using the rubric for writing a good results section - found on Canvas, in Lab 5, and pasted below - and the information provided, you must edit this results section by clearly identifying and explaining each error. Take your time as some of them may not jump out at you immediately.

Though there may be more than one type of error each error only counts as one. For example, if you see that multiple graphs are missing a figure caption that still counts as one error, the error being no caption provided.

Regardless of the fact that you are working in partners for your experimental project this assignment is an individual assignment.

If you decide to work together please be cautious as plagiarism is a serious issue that comes up when students work together and as per your academic misconduct unit from last term has serious academic consequences.

Criteria	Description
Results	Includes graphs / figures. No raw data is provided outside of supplemental. Clearly outlines the findings from the study. Flow is sensible with figures present immediately following paragraphs describing the results of figure.
Figures	Are present. Figure selected is best for this type of data. All axes are labelled with units present where applicable and legends found. All figure present have been discussed in write up. Only averages are being shown. Appropriate statistical measures are present. Figures are clear and easy to interpret. No figures present without being discussed.

Lab 4

Research Project: Part 3

Last updated 2021-12-01

Data analysis

In Lab 3 we started collecting data according what we had mapped out in our protocol. In this lab we'll be working through analysing the data we've collected to date.

Lab 5

Research Project: Part 4

Last updated 2021-12-01

Bringing things together

At this stage, you've drafted a protocol, set up your experiment, collected some data, and analyzed that data. Now it's time to communicate your findings.

Recommendation Report

In lab 1 you were introduced to Markdown in the context of documenting your project with `readmes` and `data-dictionaries`.

The basic syntax used in Markdown can be found in the BIOL Procedures and Guidelines. In the Procedures and Guidelines you are introduced to generic text editors for writing Markdown.

Markdown is a powerful authoring tool. Part of what makes it powerful is it's integration with other tools, such as `R`. In BIOL202 you will be introduced to statistical analyses using `R`. You will also be asked to author reports using `R` and `RMarkdown` - `RMarkdown` is one flavour of Markdown. Remember, the content that you're reading right now is all authored using `R` - when there's analyses being presented - and `RMarkdown`.

For this assignment, you will be learning `RMarkdown` and the environment in which we author `R` and `RMarkdown` documents - `RStudio`. You will not be expected to do your analysis in `R`.

Your Recommendation Report Draft will be submitted as both an `RMarkdown` document and a pdf. You will also need to include a copy of your data in long format saved as `.csv` and a `_DATA-DICTIONARY.md` file. More details on this in the following sections.

While your Recommendation Report Draft should be approximately 5 double-spaced pages, Times New Roman and font size 12, if you're using `RStudio` and the templates provided in this class, you should only have to concern yourself with the length of your report; the rest of the formatting will be handled when you export from `RMarkdown` to pdf.

Science writing

Technical science writing is an art. Unlike English style writing, technical science is clear-cut and lacking in artistic enhancements.

Do not quote your sources but rather read through the information and write it in your own words and cite it. It is a good idea to read an article once all

the way through without making any notes. Then come back and read it again this time making notes in the margins or on some scrap paper. This will help ensure you not only understand the material you are reading but that you are able to describe it in your own words and avoid issues of plagiarism which so often become in issue for students.

You may wish to review the BIOL Procedures and Guidelines content on APA Citations and Academic Integrity.

Preparing to write

Read a lot! It is important that you have a thorough understanding of the topic. At the very least you should have at least 3 primary source papers you are referring too throughout your report to provide further credibility to your recommendation.

Start writing early! Students often make the mistake of starting the night before the lab report is due. This more often than not results in poor submissions and thus lower grades. You should expect that you will have at least 3 rounds of revisions before you submit.

Someone reading your report should be able to tell what question(s) you addressed, why the topic is important, how you tackled the problem, the types of data you will collect, and how your research helps to inform your client.

Need help?

Book an appointment with the the Student Learning Hub's writing consultants!

A good report

A good report includes the following headings / sections

- Abstract
- Data availability statement
- Introduction
- Methods
- Results
- Discussion / Conclusion / Recommendations
- References

The goal is to clearly describe to Farmer Elliot what your question was, how you went about answering it, what your results told you and what recommendations you have to help Farmer Elliot make his decision.

Abstract

An abstract is a brief summary of what the report is all about.

Abstracts in the sciences are approached in a couple of different ways, depending on the sub-discipline and journal preferences. For BIOL125, your abstract should be a single paragraph and no more than 250 words. It should clearly outline the question or problem your research is investigating, describe how the question or problem was addressed and identify the key results and recommendations.

In less than 250 words, the reader should be able to attain the most crucial aspects of each segment of the report within this one paragraph.

Data availability statement

As we learned in BIOL116, when appropriate and feasible, the data underlying our analyses should be made available. You will be asked to submit a `.csv` file of your data in long format. You may wish to review the content from BIOL116 on Preparing Your Data and the content on Tidy Data from the BIOL Procedures and Guidelines. You will also be asked to submit a `_DATA-Dictionary.md` file describing your data. Refer back to the Data Dictionary section of the BIOL Procedures and Guidelines for guidance and an example.

This is a short statement that indicates if data is available and if it is, how it can be acquired.

Introduction

~ 1 page

The introduction should begin with the general topic and then narrow the focus of the details pertinent to the research.

Your introduction should discuss what is currently understood about the topic and how this ties into the study. This is where you want to get across the interesting points of the field that led you to develop your hypothesis and your experimental design. You want to use many sources, particularly primary sources such as journal articles. Ensure your information is cited appropriately (see guidance in the BIOL Procedures and Guidelines). You should have a clear hypothesis stated at the end of this section. This section will be the lengthiest section of your report. Ensure your reader has no doubt where the source of your information comes from.

Your introduction should situate, explain, and identify your research project. It should do this by providing relevant background information that frames the current project and is directly relevant. It should then identify the importance

of this particular project. And finally it should clearly articulate the research question and hypothesis being addressed.

Methods

~ 1/2 - 1 page

This section of your report involves producing a written description of the materials used and the methods involved in performing your experiment. Under no circumstances should you provide bullet points or list one by one the materials used. Rather you need to describe each step clearly enough that someone else could replicate your experiment exactly. You should also include a section outlining what statistical measure(s) you used and how you transformed your data if need be.

It is highly recommended you show this to someone not in your class and see if they can follow along. If they can't you need to ask them where they get stuck and re-write to make sure it's clear. Think of this like following a recipe while cooking. Don't leave anything out that isn't obvious or the recipe will fail for the next person trying to cook.

Remember, for transparency and reproducibility, your methods are key to your audience understanding how you did exactly what you did. And if you wrote a protocol, it is the methods section against which that protocol will be screened to identify bias.

So, it should be clear, concise, and contain sufficient information for someone else to reproduce the experiment. This means it should include things like, how specimens were procured, how data was collected (tools, measurements etc), and how the data was analysed.

The steps should flow logically, and, while being concise, you should not use bullet points.

Results

~ 1/2 - 1 page

The results section is where you will describe what you saw. That is, what the response was to your variable. This should be the driest and easiest section to write as you are just stating what you found and nothing more. There should be no mention of what you did to attain this data or how you went about doing it - that's for your methods section. This is not where you describe why you saw what you saw - that's for your discussion and recommendations section. Nor is it where you try and tie in other research to your research - that's for your introduction and discussion sections.

The results section should include all averaged data from observations during your experiment. This includes charts, tables, graphs, and any other illustrations of data you feel best represents the information you would like to convey. It should not include any raw data. Raw data should be attached as a separate file.

Depending on the information you wish to convey you may feel that a box plot, bar graph or line graph is most descriptive. Whichever way you decide think about what message you are trying to convey and ask yourself if an audience was to quickly look at your graph would they get that messaging easily. If not, you should look at an alternative way to display your graph. Your TA will be able to help you sort this out as well.

Be sure to provide all labels, legends and axes where necessary and a caption which informs the reader of what they are looking at. Remember anyone who is not familiar with your research should be able to quickly look at your figure and understand what message you are trying to show. Please review the BIOL Procedures and Guidelines section on Figures & Tables.

Your results section should clearly outline the relevant findings from your study and should flow directly from your research question and hypothesis.

This section should include graphs or figures to highlight key findings. Graphs and figures should be present immediately following paragraphs describing the results described by these graphs and figures.

While summary data should be provided, raw data should be not; raw data should be included as supplementary content.

Discussion, Conclusions & Recommendations

~ 1 page

This section is where you will discuss what you saw. Were you able to answer the question you set out to answer? Why or why not? In either case try and explain and interpret your results.

This is where you will want to go back to the journals you found and see what they found. Is it similar or not? Why or why not? Did they do something different from you? You can often explain results you may not have anticipated seeing by looking at what others in the area have found. Think about the why?

Is this the right property for Farmer Elliot or should he keep looking?

Your job here is to try and explain what you found and how it relates to what others have found. From here you will make your recommendation to your client.

References

All references used should be included at the end of your report on a separate page. That includes any books, articles, lab manuals, etc. that you used when writing your report. APA citations are required. Ensure you provide a properly formatted list with sufficient references. At least 3 primary source papers should be listed.

For formatting guidance, refer to the APA section of the BIOL Procedures and Guidelines.

RMarkdown and RStudio

When you're writing **readme** files and **data-dictionaries** - or even taking notes in class - a text editor like Atom is extremely convenient and versatile. When it comes to authoring reports, however, we're going to move you into RStudio.

RStudio is an IDE - an Integrated Development Environment - for R. This is just a fancy way of saying that it's an application that helps you write R code. In BIOL202 and BIOL228, you'll start using RStudio to do analyses in R. Right now, we're just using RStudio to write in RMarkdown and to get used to using the RStudio environment. Along the way, we'll see some R code as we get things set up.

Since RStudio is designed for working with R, we need to install both R and Rstudio. So let's do this.

If you're running a Chromebook, using a tablet, or don't want to install anything new on your computer, all of the Windows computers in the library have R and RStudio installed on them.

While the **tinytex** package is installed on these machines, it's not loaded out of the box. So, you will need to run the following code in the console

```
tinytex::install_tinytex()
```

If prompted to update the **rmarkdown** package, do so. There are more details on **tinytex** and **rmarkdown** in the 'Getting set up' section below.

Installing R

R is available from CRAN - the Comprehensive R Archive Network - and is available for all operating systems. Find and download the installer for your operating system at <https://cran.r-project.org/>. At the time of writing, the latest version is 4.1.2 Any version that is 4.x.x should be fine for what we'll be doing.

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [Ubuntu](#))
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Installing RStudio

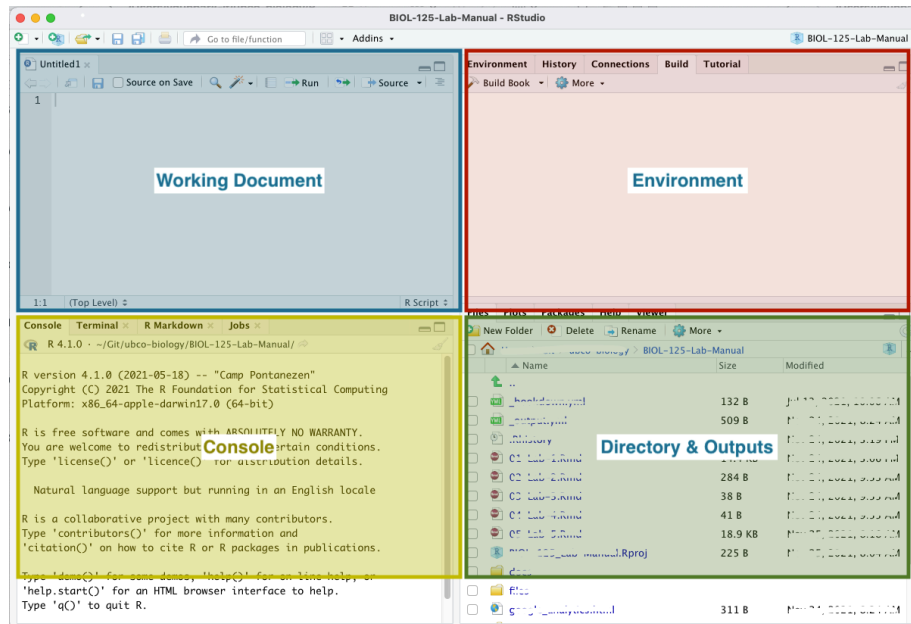
Once you have R installed, you can go ahead and install RStudio, also available for all operating systems. Find and download the installer for your operating system at <https://www.rstudio.com/products/rstudio/download/#download>. At the time of writing, the latest version is 2021.09.0+351 This or any later version that is published should be fine for what we'll be doing.

OS	Download	Size	SHA-256
Windows 10	 RStudio-2021.09.1-372.exe	156.89 MB	1c3d27f5
macOS 10.14+	 RStudio-2021.09.1-372.dmg	203.00 MB	daec6a40
Ubuntu 18/Debian 10	 rstudio-2021.09.1-372-amd64.deb	117.89 MB	921b4f23
Fedora 19/Red Hat 7	 rstudio-2021.09.1-372-x86_64.rpm	133.83 MB	f1be5848
Fedora 28/Red Hat 8	 rstudio-2021.09.1-372-x86_64.rpm	133.85 MB	ba36870d
Debian 9	 rstudio-2021.09.1-372-amd64.deb	118.10 MB	637cd465
OpenSUSE 15	 rstudio-2021.09.1-372-x86_64.rpm	119.78 MB	678d020e

A quick intro to RStudio

Your RStudio window is comprised of 4 panes.

- The upper left is where you'll find your working documents.
- The lower left is your console. It is in the console that we can run R code directly if needed.
- The upper right displays information related to your working environment.
- The lower right is where you'll see any output generated by your R code, like figures, help pages etc. It's also where you'll see a file manager so that you can interact with your files directly from within RStudio.



For report authoring using RMarkdown, we'll mostly be concerned with the upper left pane. We'll occasionally use the console to run a bit of R. And we'll occasionally use the file manager to access files. We won't worry at all about the upper right pane - this will have more relevance when you start computing statistics in R using RStudio.

When you launch RStudio, unless you're opening an existing document, you will only see 3 panes, your console will be on the left, and your environment and output panes will be on the right.

Getting set up

Being able to convert from markdown to pdf is not something we can do with the default install of R. We need to get two add-ons to be able to do this. Add-ons in R are called **packages**. The first package we need to install is **rmarkdown**. The second is **tinytex**. **rmarkdown** handles the general process of reading through your report and getting it ready to be output to a different format. **tinytex** contains the necessary information to produce a pdf, so **rmarkdown** will use **tinytex** for that one part of the conversion process.

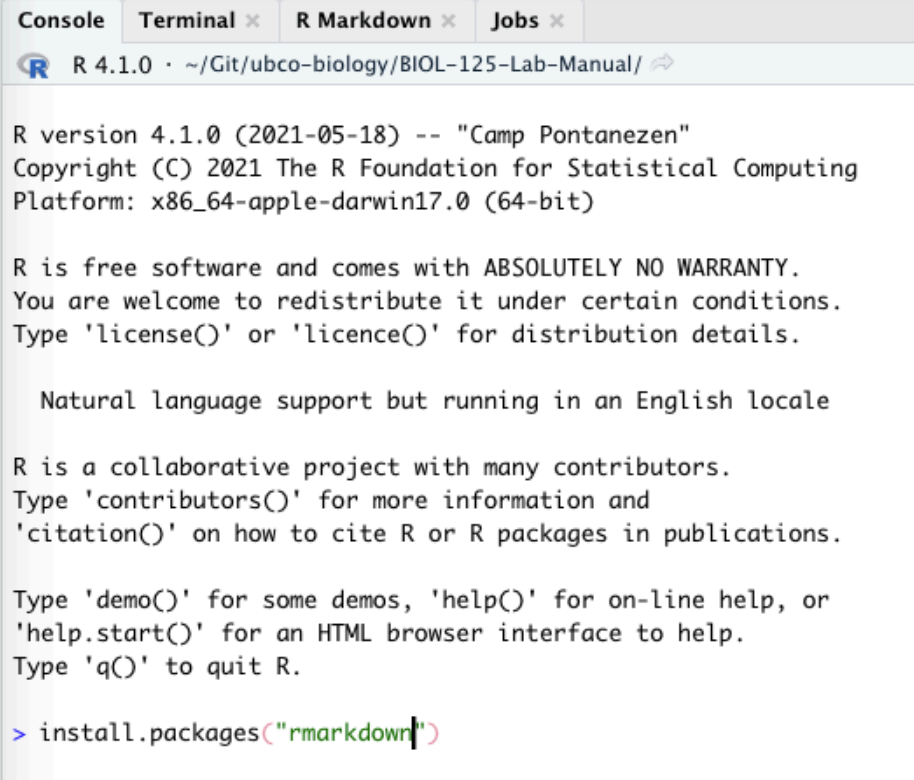
The **x** in **tinytex** is pronounced like a **k**, so should read more like **tinytek** or **tinytech**.

Installing Rmarkdown

Open RStudio, in the console type the following and hit 'Enter'.

```
install.packages("rmarkdown")
```

You'll see a bunch of stuff written to the console. When it's all done, you'll see your prompt - > - return.



The screenshot shows the RStudio interface with the 'Console' tab selected. The title bar indicates the path is ~/Git/ubco-biology/BIOL-125-Lab-Manual/. The console output displays the R version 4.1.0 (2021-05-18) -- "Camp Pontanezen" and various copyright and platform information. It also includes a disclaimer about the warranty and instructions on how to use R, including commands like 'license()', 'contributors()', 'citation()', 'demo()', 'help()', 'help.start()', and 'q()'. At the bottom, the command > install.packages("rmarkdown") is entered, with the closing quote highlighted in green and a cursor at the end.

```
R version 4.1.0 (2021-05-18) -- "Camp Pontanezen"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> install.packages("rmarkdown")
```

Installing tinytex

Now, in the console type the following and hit 'Enter' again.

```
install.packages("tinytex")
```

When that's all done, type the following and hit 'Enter'

```
tinytex::install_tinytex()
```

That's it. You should be good to go to open the .Rmd template available in the Assignment tab for this lab.

Rubric: Lab 5

Criteria	Description	Pts
Abstract	Brief, no more than 250 words. Clearly outlines the question / problem. Clearly describes how the question/problem was addressed. Results and recommendations are provided.	4
Data Availability	Data availability statement is present.	1
Introduction	Relevant background information provided. Clearly articulates how the background information is connected to the current project. Importance of this project has been described. Written well and easy to follow. Flows from more general and broad background information to the focus of the project. Hypothesis and questions posed are outlined at the end of this section. No factual errors are present.	7

Criteria	Description	Pts
Methods	No bullet points. All methods and materials are clearly described. Easy to follow. Enough information has been provided for others to be able to reproduce the experiment. Data analysis procedure is also included.	4
Experimental Design	Procedure is specific and addresses the question / problem. Data collection is clearly defined. Appropriate control (where applicable) has been used. Independent and dependent variables are identified. Student testing only one variable.	5
Results	Includes graphs / figures. No raw data is provided outside of supplemental. Clearly outlines the findings from the study. Flow is sensible with figures present immediately following paragraphs describing the results of figure.	4

Criteria	Description	Pts
Figures	Are present. Figure selected is best for this type of data. All axes are labelled with units present where applicable and legends found. All figure present have been discussed in write up. Only averages are being shown. Appropriate statistical measures are present. Figures are clear and easy to interpret. No figures present without being discussed.	7
Discussion & Recommendation	Student displays clear understanding of results. Student displays a clear understanding of the meaning of these results. Interpretation of results is founded in the data, observations and/or other studies. Recommendations are sound and based on the current study and / or other studies.	4
Spelling & Grammar	No spelling errors. No grammar errors. No awkward sentence structures.	3
References & in-text citations	APA format used properly and consistently. Minimum of 3 primary source papers used in the report. In-text citations are used when required. Citations and references match up.	4

Criteria	Description	Pts
Plagiarism & Quotations	No plagiarism of any kind has been found. No quotations present. Information attained from outside resources are properly cited.	3
File Uploads & Format	A total of 4 files have been submitted. Report has been submitted as pdf and RMarkdown. Report is no more than 5 pages (excluding references). Data has been submitted in Tidy format as csv . Data dictionary has been submitted as .md .	4
Total		50

Assignment: Lab 5

Please use the following template for this assignment:

20220101_Lab05_125_Assignment_V1.Rmd (3 KB)

You will need to submit 4 files for this assignment:

- Recommendation report as `.Rmd`
- Recommendation report as `.pdf`
- Data in long, tidy, format as `.csv`
- Data dictionary as `.md`

You will receive your marked lab report one week from the time it is submitted.

You can decide to resubmit the same lab report draft without making any changes or you will have the opportunity to review the edits and make the needed changes in order to increase your mark.

If you have any questions regarding your mark and / or the comments from your TA please ensure you take the opportunity to chat with your TA to go over these. This will ensure that you are in the best position to attain the highest marks possible for this assignment.

Using the template

All the markdown syntax that you need for RMarkdown can be found in the Markdown section of the BIOL Procedures and Guidelines.

Directory structure & file naming

It is expected that you will have a root project folder for your work associated with this lab. And that at the minimum you will have a folder for your report, your data, and your figures. And that you will download this template into your **report/** directory. And that lastly, you will rename the template in accordance with the file naming convention you outlined in your first assignment.

This structure and hierarchy will be important when it comes time to include figures and images in your report.

YAML

The top of the template contains some front matter called YAML. YAML provides instructions to all the pieces of software involved in converting your RMarkdown document to its outputs, in this case, **pdf**. YAML is very specific to spacing, so don't add any extra spaces!

What you need to do.

1. Provide a title within the quotations after **title**.
2. Provide your name within the quotations after **author**.
3. Provide your abstract within the quotations after **abstract**.

What might be nice to know.

1. `r Sys.Date()` pulls the date from your computer and auto populates this for you.
2. The **output** tag defines the output format. Other options include **html_document** and **word_document**.

What exactly is YAML?

YAML™ (rhymes with “camel”) is a human-friendly, cross language, Unicode based data serialization language designed around the common native data types of dynamic programming languages. It is broadly useful for programming needs ranging from configuration files to internet messaging to object persistence to data auditing and visualization.

Read more at the Official YAML Web Site

Document body

The template is then pre-populated with first level headers for each section you're expected to include in your report. Each heading re-iterates the key elements the content of these headings should address. This is just place holder text, so replace it with your own.

Images & graphs

There is one sample graph included. Note how it references the figure to be included `../figures/image-name.png` The `../` means 'go one level up in the directory' which, if you have your project set up in the following way and your `.Rmd` file is in your `report/` directory it means 'look in the `root/` directory for a folder called `figures/`.

```
root/  
  report/20220101_Lab05_125_Assignment_V1.Rmd  
  data/  
  figures/MVD_BIOL125-Lab5_Fig-1-Boxplot_V1.png
```

If you make a mistake in setting this path, you'll get the following error in RStudio

```
(No image at path ...)
```

You'll also note the following directly after the image path: `{width=50%}`. This reduces the image size by 50%. This works well for the images produced by the ShinyApp used in this course.

As noted in the template, you do not need to write **Figure 1:** before your figures; this small piece of text is handled during the conversion from RMarkdown to pdf. Any other information that you would like to include in the caption should go in the `[]` before the `()` that contain the path to the image.

Figure placement

The engine behind the conversion from RMarkdown to pdf is a typesetting application, one with pretty strict rules about how content should be formatted - much more strict than something like Microsoft Word.

What this means is that if the placement of your images will disrupt your prose - by creating large amounts of empty white space for example - this typesetting application will *push* your figure to somewhere lower in your report where it won't create this white space.

Your figures should be adjacent to the relevant text in your RMarkdown file. How this manifests to your pdf might look a little different; that's ok.

References

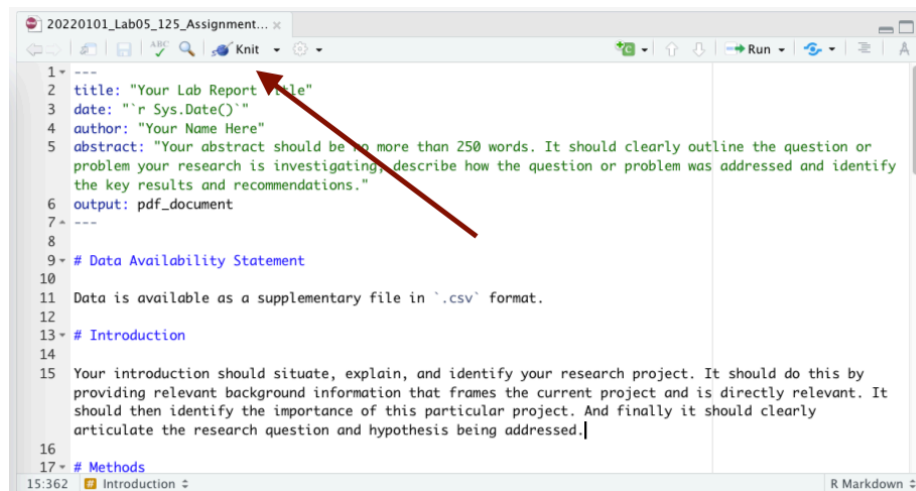
Just before the heading for references you'll see the following

```
\clearpage
```

This creates a page break between your references section and the rest of your report.

Building the pdf

If you've installed R, RStudio, and the `markdown` and `tinytex` packages successfully, when you open the template `.Rmd` file you should see an option to Knit.



Click this button or select the drop down arrow and select **Knit to pdf**. This will generate a pdf in the same directory as your `.Rmd` file.

To test this with the template, ensure the template `.Rmd` file is in your `report/` directory and download the following image into your `figures/` directory

- MVD_BIOL125-Lab5_Fig-1-Boxplot_V1.png (4 KB)

You should get something that looks like this after **Knitting** the `.Rmd` file

- 20220101_Lab05_125_Assignment_V1.pdf (180 KB)