

BIOL 59000: DATA SCIENCE PROJECT FOR LIFE SCIENTISTS

WRITTEN PROJECT REPORT

The individual capstone research project must focus on a topic that applies data analytics in the context of bioinformatics or computational biology. This work must make a scholarly contribution to existing knowledge and practice in the field. Although the topic of each student's project will vary, all should draw upon concepts and skills learned in all other courses in the MSDS program, with focus potentially falling on one or two areas more than others. Projects may:

- Focus on analyzing a large-scale, real world data set
- Develop a software tool for Data Science related work (identifying a data set that could be used to test the functionality of this software)
- Implement machine learning or other data science-related techniques in real-time systems

I. GENERAL INFORMATION

The written report of this research project will ultimately be unique for each student, but needs to be reported in a consistent format. The final report should include the following components:

- 1.) An abstract which addresses the following:
 - a. Statement of the research problem (including the objective(s) of the study)
 - b. Methods (brief overview)
 - c. Results (brief summary)
 - d. Conclusions (including major findings of the project, interpretation of outcomes, basic implications)
- 2.) An introduction – This will provide background and reviews the current literature. Also state the purpose of the study and research objectives.
- 3.) Methods/Procedure – Explain the methodology used in the research project, indicating novel approaches as well as use of procedures initially proposed by others
- 4.) Results – Report outcomes of method/procedure, providing data in tables/figures as appropriate.
- 5.) Discussion/Conclusions – Discuss interpretation of results in the context of the larger literature, summarize the outcome(s) of the project, propose future work that could be performed as an extension of this project
- 6.) Literature cited – Provide a comprehensive list of all sources cited throughout the body of the paper

II. ORGANIZATION OF RESEARCH PAPER

A. Arrangement of Sections:

1. Title Page
2. Abstract
3. Acknowledgment Page (Optional)
4. Table of Contents
5. List of Tables
6. List of Figures
7. List of Abbreviations
8. Main Text (divided into chapters)
9. Literature Cited
10. Appendix (if necessary)

B. Description of Sections:

1. Title Page: Refer to enclosed example in appendix.
2. Abstract: See description in GENERAL INFORMATION section. Typically ~200-250 words in length. Because it is an overview of the entire document, a good rule of thumb is to write this section last, as a summary of content contained throughout the document.
3. Acknowledgement Page: This section is at the discretion of the author. It is not required, but any information provided should be consistent with the professionalism applied to the rest of the document.
4. Table of Contents: It is recommended that this section is generated using the automatic function in Word (or similar software). Pages included in the document prior to the Table of Contents (title page/abstract/acknowledgement page) should NOT be included. Refer to enclosed example in appendix.
5. List of Tables/List of Figures: It is recommended that these are generated using the formatting option in Word. Refer to enclosed example in appendix
6. List of Abbreviations: This is a place to report all abbreviations that may be used in the body of the text. When an abbreviation is first introduced in the paper, write out the full term, followed by the abbreviation in parentheses. For example, you may want to use the abbreviation for Human Immunodeficiency Virus (HIV) rather than writing out the full length term repeatedly. For all subsequent usage of the term in the body of the text, strictly use the abbreviation (do not revert back to the full length). On the abbreviation list, alphabetize the abbreviations in one column, and the full length definition of the abbreviation in a corresponding second column. Refer to enclosed example in appendix.
7. Main Text: The body of the paper should be divided into chapters (following the outline proposed in the submitted prospectus), including the introduction, method/procedure, results and discussion/conclusion portion of your work.

8. Literature Cited: This section should report full references for all works cited within the text. If additional sources were consulted in the process of completing this project, but were not directly cited, they should be listed separately in a table of Secondary References. For format, refer to enclosed example in the appendix.
9. Appendix: The appendix can be used as a way to organize supplementary materials not placed within the body of the paper. This section is not required; include as needed.

III. MATTERS OF STYLE

- A. Including Tables/Figures. Any “data panel” (either table or figure reporting data) should be included in the document according to the following formatting rules:
 1. Each table/figure should be placed on its own separate page in the document, immediately following the page on which it is referenced in the text.
 2. Each table/figure needs to be labeled with a number and a descriptive title. If necessary, the title can be followed by a descriptive legend. Table numbers/titles/legends should be placed **immediately above** the table. Figure numbers/titles/legends should be placed **immediately below** the figure. Table/Figure numbers and titles should be reported, along with the page number, on the “Lists...” pages at the beginning of the document. Examples:
 - i. Table 1. Summary of cancer samples studied.
 - ii. Figure 1. Correlation between age at time of disease outset and length of remission.
 3. Broadside tables and figures: Certain oversized tables or figures often require their placement on the page in “landscape” – this is appropriate, but please orient such that the top of the table/figure aligns with the equivalent of the left-hand margin when the page is printed in “portrait.”
- B. In-text Literature Citations. Typically footnotes are not used in the context of this type of scientific literature. Literature citations giving the author(s) and year of publication are used in the following manner:
 1. Indirect quotation (paraphrased material, the preferred method):
 - i. Seyle (2011) states that stress does have a direct impact on physical well-being.
 - ii. Stress has been shown to have a direct impact on physical well-being (Seyle, 2011).
 2. Direct quotations (not typical in this type of writing – whenever possible, use indirect quotations/paraphrasing):
 - i. (Short quotation) Noble (1999) states, “The pelvis of the Salienta is characterized by its long ilia.”
 - ii. (Long quotation) Noble (1999) states,

The pelvis of the Salienta is especially characterized by its long ilia which makes a ligamentous connection with the diapophyses of the sacral vertebrae. There may be [...]

3. Report of author(s) and year of publication should be adapted depending on the number of authors for the document.
 - i. One author: (Cummings, 2011)
 - ii. Two authors: (Powers & Pifer, 2012)
 - iii. Three to five authors:
First citation: (Clark, Tanaka, Powers & Veselits, 2011)
Second and subsequent citations: (Clark et al., 2011)
 - iv. Six or more authors:
First citation: (Mandal et al., 2011)
Second and subsequent citations: (Mandal et al., 2011)
- C. General Formatting: Use Times New Roman, 12 point font that is double-spaced. All paragraphs should begin with tab indentation.
- D. Margins: All margins should be 1" with the following exceptions:
 1. The TOP margin of ALL first pages of new chapters should be 2"
 2. If present, the TOP margin of the Acknowledgement Page should be 2"
- E. Page Numbers:
 1. Preliminary pages (Title Page, Abstract, Acknowledgement Page, etc.) should be numbered with lowercase Roman Numerals (i, ii, iii, etc.) placed at the BOTTOM CENTER of the page. However, the title page should not show the number (it is implied as "i"), and the numbers should first appear on the second page of the document (on the Abstract page, numbered "ii").
 2. Beginning with the Main Text, page numbers should switch to Arabic Numerals (1, 2, 3, etc.). The first page of all chapters should have the number placed at the BOTTOM CENTER of the page. All following pages within the chapter should then have the page number placed in the TOP RIGHT corner. Use continuous numbering from the first page of the first chapter through the last page of the document.
- F. Person and Tense: This paper, like any scientific paper, should be constructed in accordance with objective principles and is not personal in nature. Therefore, it should be written in third person, avoiding use of any/all personal pronouns. Regarding verb tense:
 1. Background factual information that continues to hold true should be written in present tense. Example: "DNA is a double helix."
 2. Methods and results that report the outcome of research conducted (steps that have been completed in the process of completing this project) should be written in past tense.
 3. Future tense may be used when future work is proposed.
- G. Formatting Sections and Subsections (per Turabian): The written report will be divided into sections, and in the body of the Main Text, often chapters will need to be further divided into additional sections and subsections. To denote importance (rank) of each heading as well as provide the greatest clarification in style for the reader, a standard

method of placing section/subsection headers has been devised. For this report, enact this following scheme (includes out to six ranks of subtitles, if needed):

1. Centered heading, ALL CAPITALS
 2. Centered heading, underlined
 3. Centered heading, not underlined
 4. Side heading (flush with left margin), underlined
 5. Side heading (flush with left margin), not underlined
 6. Paragraph heading (one tab-indent from the left margin, followed by a period or dash), underlined
- H. Widows and Orphans. Be aware of situations in typesetting where new units of information (heading, paragraph) starts or ends a page as a single line, as this can lead to a break in thought and, for presentation purposes, is considered unacceptable. When a new section heading or a paragraph-opening line appears at the bottom of a page (orphan), adjust the text such that the line is moved to the top of the next page, and remains associated with the corresponding content. When a paragraph-ending line appears at the top of a page (widow), adjust the text so that at least two lines (or more) appear at the top of that page. Do not change the margins or spacing of the paragraph to achieve this end – rather, widows and orphans should be accommodated by adjusting space (lines) between the lines of text in question.
- I. Use of Numbers. Numbers should always be spelled out if the first word in a sentence, EXCEPT when reporting values associated with units of measure (volume, mass, length, time), which should be reported as Arabic Numerals. Spell out numbers one to nine, EXCEPT when accompanied by units of measure, but report 10+ as Arabic Numerals in all instances. If reporting a series and the series contains numbers of 10 or higher, use numerals to report all values (such as 3 mice, 4 rats and 11 hamsters). When reporting values in terms of decimals, always report using numerals (use 0.1, not .1 always; use 6.0, not 6 only if significant). Decimals should always be used for precise measurements (rather than fractions; 0.5 second rather than one-half second). Always spell out a fraction unless it is part of a larger number (such as three-fourths versus 25.75). Percentages should always be given in Arabic Numerals (98%); if the words for “%” are used, should be written as “per cent” not “percent.”

IV. APPENDIX

The appendix contains the following examples:

- Title page
- Table of Contents
- List of Tables
- List of Figures
- List of Abbreviations
- Literature Cited

LEWIS UNIVERSITY

<TITLE>

RESEARCH PROJECT SUBMITTED
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE, DATA SCIENCE

BY

<NAME>

DIRECTOR: <FACULTY MEMBER'S NAME>

ROMEDEVILLE, IL

<DATE: MONTH, YEAR>

TABLE OF CONTENTS

LIST OF TABLES.....	v
LIST OF FIGURES.....	vi
LIST OF ABBREVIATIONS.....	vii
CHAPTER I – INTRODUCTION.....	1
B CELL DEVELOPMENT.....	2
PRE-B CELL RECEPTOR SIGNALING.....	4
IL-7 RECEPTOR SIGNALING.....	7
CELL CYCLE.....	10
CHAPTER II – METHODS.....	12
CHAPTER III – RESULTS.....	15
CHAPTER IV – CONCLUSIONS.....	25
LITERATURE CITED.....	30

LIST OF TABLES

Table 1. Summary of Cancer Samples Studied.....	16
Table 2. Public Databases Used.....	18

LIST OF FIGURES

Figure 1. Stages during B cell development.....	3
Figure 2. Overview of signaling pathways.....	6
Figure 3. Cyclin D2 and cyclin D3 protein stability in pro-B cells.....	20
Figure 4. Cell cycle analysis of B cell progenitors upon PI3K inhibition.....	22

LIST OF ABBREVIATIONS

BCR	B Cell Receptor
FACS	Fluorescence Activated Cell Sorting
HSC	Hematopoietic Stem Cell
ITAM	Immunoreceptor Tyrosine-based Activation Motif
PTM	Post-translational Modification
Rb	Retinoblastoma protein
Syk	Spleen Tyrosine Kinase
TSLP	Thymic Stromal Lymphopoietin

LITERATURE CITED

- Dias, S., Xu, W., McGregor, S. & Kee, B. (2008) Transcriptional regulation of lymphocyte development. *Current Opinion in Genetics & Development*, 18, 441-448.
- Gong, S. & Nussenzweig, M.C. (1996) Regulation of Early Developmental Checkpoint in the B Cell Pathway by Igbeta. *Science*, 272, 411-414.
- Milne, C.D. & Paige, C.J. (2006) IL-7: A key regulator of B lymphopoiesis. *Seminars in Immunology*, 18, 20-30.