

# Assessment Rubric for Senior Thesis in Data Science

Student Name: \_\_\_\_\_  
 Reader Name: \_\_\_\_\_  
 Term: \_\_\_\_\_

## Learning Outcomes

Students with a major in Data Science will learn the following topics.

1. Relevant programming abilities.
2. Understanding computational efficiency and data structures.
3. Proficiency with statistical analysis of data from one or more disciplinary areas.
4. Working with remote databases (often referred to as big data.)
5. Understanding of probability and theory of statistics.
6. Communication skills including the use of professional level statistical software for analysis, visualization, and reporting.
7. Making decisions based on statistical analysis of data together with an ethical framework.
8. Familiarity with common tools of data science.

Learning Outcomes	Advanced 4	Proficient 3	Developing 2	Emerging 1	Score
<b>I. Programming</b>	Students can build complex algorithms using advanced data structures in multiple languages.	Students can combine algorithms and functions to achieve complex effects in at least two high level languages appropriate for work in data science.	Students can combine algorithms and function appropriately to achieve complex methods in at least one high level language appropriate for work in data science.	Given a simple algorithm or method, student can code them in a high-level programming language.	
<b>II. Computational efficiency/Data structures</b>	Designing advanced data structures appropriate to a specific problem.	Implementing advanced data structures to improve speed of algorithms.	Understand how the use of more advanced data structures can change the running time of operations.	Understand the runtime of operations needed to perform basic tasks as they scale with problem input.	
<b>III. Statistical analysis of data from different disciplines</b>	Hierarchical Bayesian models and networks. Regression using variable selection and regularization techniques such as LASSO. Deep understanding of the statistical tools used for a particular discipline or domain.	Advanced one-dimensional analysis using non-parametric and conjugate Bayesian approaches. Multivariable data set analysis, including ANOVA.	Building confidence intervals for different applications.	Working with simple one-dimensional data sets and estimators from one or more application areas.	
<b>IV. Remote databases</b>	Understand advanced computing technologies for working with remote data sets, including knowledge of speed and reliability tradeoffs.	Being able to undertake data analyses using remote databases in one or more domains.	Learning to access large data sets through tools such as SQL.	Understand the issues involved with data that can not be contained in local memory.	

<b>V. Probability and theory of statistics</b>	Understanding limits of information contained in random variables leading to best possible estimators in certain classes.	Building estimators with properties such as consistency and confidence intervals using pivots, the central limit theorem, and distributional properties.	Making connections between properties of random variables and qualities of estimators.	Working with distributions, understanding expectation and variance.	
<b>VI. Decision making</b>	Combining ethics and statistical knowledge to make informed decisions.	Knowledge of ethical frameworks for evaluation of decisions.	Interpreting statistical models to make optimal decisions.	Understanding how models can be used to inform decisions.	
<b>VII. Communication</b>	Complete fluency with typesetting, visualization, and reporting for written documents. Ability to create dynamic presentation and web applications to illustrate results.	Use of advanced library functions for advanced statistical analyses, typesetting, and reporting.	Use of visualization using prepackaged graphical functions and methods.	Being able to write simple reports including basic estimation tasks.	
<b>VIII. Data science tools</b>	Linked together multiple steps for a large complex analysis of untidy data.	Building models for tidied data, and using them for prediction and statistical inference.	Learning to clean and tidy data.	Understanding how to import tidy data.	