CSC 530 Data Analysis Topics:

1. Careers after the college degree
2. Surviving on minimum wage
3. Amount of snow to cancel school
4. Body maps show where we feel emotion
5. Global forest change
6. Personal space per person in various countries
7. Woman as academic authors over the years
8. Longer life expectancy, more years of disease
9. Emotions of countries mapped
10. Solar energy potential, mapped in Cambridge
11. Soundscapes of city pollution
12. Life expectancy and child poverty as a tube map
13. High school drop-outs (857/per hour)
14. How recruiters look at your resume
15. Cell phones and cancer
16. Exploratory tool for school admissions
17. Music listening preferences by gender/age
18. Driving habit and gas prices shift into reverse
19. Visualize this: class size and quality of education
20. Class size and SAT scores by state
21. Caffeine vs. Calories – Buzz vs. Bulge
22. Visualize this: Obesity rates by state
23. How long will the world’s natural resources left?
24. Google users search terms to predict flu activity
25. Decline of U.S. Men’s tennis

**For the first half of the data analysis project, we focus on the statistical analysis part, which applies the CI, hypothesis tests, and linear/nonlinear regression techniques to approximate the model. It should include all the details of your implementation, including:**

* Topic selection or problem statement
  + What’s your problem?
  + Why do you choose this topic?
* Data collection
  + Existing public data, real or artificial?
* Background review
  + State-of-the-art techniques for this problem or similar problems
* Implementation approach
  + What’s your proposed approach to solve the problem?
  + How do you implement your approach?
  + What kind of specific technologies or tools are needed?
* Results analysis
  + What are your results?
  + How do you verify your results are correct, e.g., statistical tests, subjective and objective assessments?
  + Do the results verify your solution, i.e., does your approach solve the problem?
* Conclusion
  + What are your conclusions?
  + Lessons learned and future work.

Based on the above statistics analysis, we will conduct data mining/learning on the same (or similar) topic in the end of this semester. Here is the final report template, which can be used for your reference at this point.

1. Title
2. Abstract
3. Introduction
   * Covers the overview of your work
   * Problem statement and why it is important
   * Data collection or available public data
   * Your implementation plan (comparison study or a specific technique)
4. Background
   * Related works and applications
   * History and state-of-the-art works
5. Implementation
   * A detailed description of the algorithm should be included. The related mathematic proof can be added to the appendix.
   * Data collection and description
   * Data analysis.
6. Experiments
   * How many experiments? What are the purposes of the experiments?
   * You should include the hypotheses tests to verify the statistical significance of your results.
   * Use figures and tables to present your results.
7. Conclusion
   * Strength and shortage of the approach
8. Reference
9. Appendix
   * Algorithm details
   * Source codes