The following list consists of past final projects for BIOS 669. Hopefully these examples will inspire new ideas for your final project topics. **Please note: these are *examples*, not a list of topics to choose from.**

- Download and analyze (in relation to weather) some months of data from the City of New York's Citi Bike program.
- Scrape and analyze data related to Major League Baseball's Beat the Streak online contest.
- Download data from the National Immunization Survey for Teens and investigate possible determinants of HPV and meningitis vaccination rates.
- Try to reproduce tables and figures in a paper on health insurance, alcohol, and tobacco use among pregnant and non-pregnant women, which was based on data from the National Study on Drug Use and Health (available from the University of Michigan's ICPSR site)
- Graphically look at NC State Fair attendance over the years, especially in relation to weather during the week of the fair. Both kinds of data were obtained by web scraping.
- Look at player salary compared with productivity on the Pittsburgh Penguins NHL team (data obtained by web scraping).
- Investigate possible relationships among country-specific variables obtained from the World Health Organization's Global Health Observatory data repository.
- Predicting NBA end-of-season player awards based on player performance.
- Take a longitudinal look at data from the National Youth Tobacco Survey – do findings lend support to any particular methods of tobacco use control among young people?
- Look at how the introduction of the 3-point shot in the NBA in 1980 has affected aspects of professional basketball
- Download and analyze mobile food (AKA food truck) data for San Francisco – look at permits by space and time to see where and when food trucks are most common (data source: San Francisco's open data portal)
- Using MEPS data (Medical Expenditure Panel Survey), look at demographic and health-related factors associated with having private insurance and government-provided insurance as well as health care utilization and expenditures
- Analyze NBA draft picks (pick order 1-60) to see if draft order is related to minutes played, points scored, win share of the team, etc. – data obtained by web scraping
- Look at customer complaints registered with the Consumer Financial Protection Bureau (CFPB).
- Download and analyze fire incidents data from the City of San Francisco database.
- Download and analyze storm data from the NOAA website. What kinds of storms (characterized by rain, wind, flooding, etc.) cause the most fatalities and property damage?
- NHANES data has been relatively popular:
  - look at self-reported physical activity in relation to Hispanic ethnicity
  - look at prevalence of dietary supplement use in different demographic groups
  - look at soda (and other beverage) consumption among college students
- Download and analyze data from the CCC-ACS (Cardiovascular Disease in China project, looking at acute coronary syndrome)
- Answer questions about refugee movements based on United Nations data

- Using data from Yelp (via kaggle), analyze restaurant ratings in relation to business attributes for restaurants in selected cities
- Analyzed crowd-sourced data on police violence referenced on FiveThirtyEight.com
- Using data from MediaCloud.org, try to reproduce FiveThirtyEight.com analyses comparing media coverage of Hurricanes Irma, Harvey, and Maria (2017) and states/territories Florida, Texas, and Puerto Rico
- Using data from www.basketball-reference.com and focusing specifically on the Philadelphia 76ers, try to answer the question "Is it worth tanking in the NBA?" (that is, lose deliberately so that you are awarded higher draft picks)
- Use data from the National Ambulatory Medical Care Survey (NAMCS) to describe trends in antibiotic prescriptions over time and geographic regions
- After downloading data using open data APIs maintained by each city, compare crime rates for Cary, Chapel Hill, and Raleigh across several years
- Reproduce the results in a PLOS Medicine paper on a trial of social dancing and incidence of falls in older adults
- Analyze changing characteristics of adlibs in rap music over time using web scraped song lyrics and billboard rankings of top songs on selected dates
- Analyze how prominent topics in statistics and probability have changed over time using web scraped thesis keywords
- Build a stock portfolio using Yahoo Finance data and the methodology suggested in a certain paper
- Download state-level data on educational funding and educational outcomes and look for associations
- Write a recommender system for NC state parks based on activities available, trail characteristics, distance from a specific zip code, and so forth
- Build a "neighborhood selector" application for New York City based on data pulled from various sources (NYC OpenData, Zillow, NYC health department)
- Using Yelp data on Chapel Hill restaurants, build and compare two types of recommender systems
- Download COVID-19 case rate data and analyze it over time or for different geographic areas or in relation to socio-economic factors
- Download CSV files from a regional anti-poverty organization website and try to reproduce figures from a report put out by the organization
- Reproduce tables and figures from the paper *The Paradox of Music-Evoked Sadness*
- A BIOS MS student in his final semester wrote complicated macros for preparation and analysis of genetic data used for his Masters thesis – turned in both R and SAS code
- Download the IMDb database and analyze it for rating patterns by genre
- Download and analyze concussion data from NEISS (National Electronic Injury Surveillance System)
- Download airline flight performance data from TranStats and look for factors related to on-time arrival and departure

- Looked at air quality data from Beijing and tried to relate it to measures of industrial activity over time
- Used data from Kaggle to look at how published coronavirus research topics changed from late 2019 into spring 2020
- Used the Tweeter Search API to obtain about 25,000 COVID-related tweets in China in April 2020 – analyzed sentiment and popularity in relation to tweet source (government or media)