Intro

* Population: Summer 2021 STA302H1 cohort, roughly n = 227 students at the beginning of the course, n = 198 students as of August 13, 2021.
* Professor surveyed students (all short answer) on Quercus for the first 4 weeks of STA302H1:
* Study the relationship between interim STA302H1 quiz scores, study time, and COVID contemplation time, vs. final STA302H1 quiz score
* Many factors can affect quiz 4 score:
  + Internal factors within STA302H1
    - Studying
    - Attending lectures
    - Attending office hours
    - Completing assignments
    - Completing readings
  + External factors outside of STA302H1
    - Prerequisite knowledge
    - Current work ethic, discipline, diligence
    - TODO: Do grade adjustments count?
* ~~Develop a model to learn which of the following -- interim STA302H1 quiz scores, study time, and COVID contemplation time – are most important for predicting quiz 4 score.~~
* Developing this model will serve the following purposes:
  + Helps current professors identify possible weak topics, reflect on what things they tried to do that helped students and things that didn’t help, and tailor their future lessons that way
  + Guides future professors establish course expectations, and common pitfalls so they have a chance to prepare extra well for those formative lectures
  + Helps current students focus on key components to getting good Quiz 4 grades
  + Helps future students establish reasonable expectations and develop strategies to maximize their time and success in STA302H1
  + Helps UofT admin decide who deserves a raise/promotion?

Exploratory Data Analysis

* Quiz scores are out of 10, STA302H1 study time and COVID contemplation time are measured in hours.
* Professor collected survey data for the first 4 weeks of STA302H1:
  + End of Week 1 (July 5 – July 9)
  + End of Week 2 (July 12 – July 16)
  + End of Week 3 (July 19 – July 23)
  + End of Week 4 (July 26 – July 30)
* Response variable:
  + quiz 4 score
* Predictor variables:
  + quiz 1 – 3 scores (out of 10)
  + weeks 1 – 4 sta302h1 study time (hours)
  + weeks 1 – 4 covid19 time (hours)
* How to calculate values for predictor variables
  + Median of 3 quiz scores
  + Median of weeks 1 – 4 COVID hours
  + Median of weeks 1 – 4 STA302H1 study hours
  + Country – use factors
    - When I tried to filter by country, it turns out all countries but Canada, China, and Unknown contain only 1 – 3 entries. When I tried to construct the correlation matrix, I obtained a bunch of NAs or r = +/- 1.
    - Therefore, it doesn’t make sense to analyze each country separately. Instead, I’ve decided not to group data by country.
  + 1 – 2 NAs for quiz scores is OK, 3 – 4 NAs indicates dropped course ~~Missing both quiz 2 and quiz 3 marks indicates dropped course~~
  + ~~1 – 2 NAs for COVID and STA302 hours is OK, 3 – 4 NAs indicates dropped course~~
  + Missing the country is ok, just mark as unknown
  + Missing any number of COVID and STA302H1 hours is OK, those students either forgot or abstained. These variables have negligible effect on quiz 4 grades anyways
* Look at descriptive statistics (i.e., histograms, boxplots, scatterplots)
  + TODO: insert scatterplots of strong relationships between quiz 4 score and one of the predictor variables
  + TODO: insert histograms of noteworthy relationships between quiz 4 score and one of the predictor variables
  + ~~TODO: Take mean of 3 quiz scores, weeks 1 – 4 COVID, weeks 1 – 4 STA302H1 study~~
  + TODO: Take median of 3 quiz scores, weeks 1 – 4 COVID, weeks 1 – 4 STA302H1 study
  + TODO: Analyze similar countries separately or add country as a predictor variable and group similar countries together to reduce number of categories.
* TODO: Consult 3 – 4 external sources to confirm your findings.

Model Development Selection

* Model selection
  + Tentative models:
    - Quiz 4 ~ Quizzes 1 – 3
    - Quiz 4 ~ Weeks 1 – 4 COVID-19
    - Quiz 4 ~ Weeks 1 – 4 STA302H1
  + Model selection criterion
    - For each of the three models, there should be about 8 (2^3) – 16 (2^4) possible models to choose from
    - Use R^2, adjusted R^2, C, AIC, forward selection, backwards selection criterion
    - Majority model(s) win. TODO: Is this true?
* First order model?
  + Linear model?
  + Quadratic model?
  + Polynomial model?
* Second-order model?
  + Think about interaction terms & multicollinearity
  + Correct multicollinearity with recentering, and perform analysis based on recentered model
  + Correct skewness with variable transformation, and perform analysis based on transformed model
* TODO: Add justification for your model
  + Checking Gauss-Markov assumptions
  + Showing residual plots
  + Showing skewed histograms
  + Showing well fitted qqplots
  + Using prior knowledge (e.g., quiz grades tend to be left skewed because few STA302H1 students fail quizzes)

Limitations

* Online offerings are different from equivalent in-person offerings of STA302H1
  + More discipline necessary to succeed
  + More technological literacy required to succeed, privileging students who have access to computers or internet – some families are too poor to afford computers and internet, and must share or use public computers
  + Performance has more room to fluctuate – some people may prefer online courses over in person lectures due to long commute times, and vice versa
  + Harder to create boundaries with work/rest, and work environments vs. play/sleep environment
* Events outside of one’s time in STA302H1 (studying for quizzes, attending lectures, and doing assignments)
  + Mental health
    - Better mental health = more resilient individuals, more positive outlook on life, more altruistic
  + Commute times
    - Increased commute times increases students stress levels because it leaves less time available for STA302H1 and other courses, and affects one’s disposition towards class format (in-person vs. online)
  + Social connections
    - Staying connected helps stave off negative COVID-19 thoughts and promote community
  + Number of hours slept, quality of sleep
    - Well-rested brains tend to respond quicker, can better regulate their mood, have an easier time focusing during lecture, and make fewer mistakes on quizzes and assignment (or notice them more easily and readily)
  + Anxiety levels?
    - Anxiety may influence quiz scores and assignment performance because it affects cognitive performance – brain may hyperfocus (fixation) or hypofocus (distraction) – fight or flight, focus on survival (not dying of COVID-19) rather than thriving (succeeding in STA302H1)
  + Number of extra-curricular activities, professional (e.g., preparing for job interviews) or recreational (e.g., going outside, biking, etc.)
    - Alternatives for STA302H1 studying, effects are debatable depending on how related they are to STA302H1 or statistics in general
  + Level of physical activity (exercise helps brain learn, which may improve performance)
    - Exercise and physical activity make you more alert, less prone to illness and injury, promote a positive attitude, and improves relationships with other people
  + Caring for family
    - Family responsibilities may distract or interfere with one’s progress on STA302H1 assignments and consume STA302H1 study time
    - Family members may also provide a nurturing environment for your studies through quality family time, strong family values, and family work connections
  + Discipline
    - Exercising discipline allows one to consistency interact with STA302H1 material and increase their chances of retaining the material and performing well in STA302H1 assignments and quizzes.
  + Prior background – direct (past stats/ML courses) or indirect (quantitative background, programming background)
    - Those who are more familiar with prerequisite statistics material and do well in prior courses have an easier time studying for STA302H1 and are more likely to succeed – ceteris paribus.
  + Attitude towards school (keen vs. disengaged)
    - Those who are keen in school tend to perform well, regardless of class format, and vice versa
  + Thoughts about COVID
    - Positive thoughts can help tune out some negative COVID-19 thoughts and focus on what you have control over. Persisting COVID-19 thoughts can cause one to feel despair and pessimism
  + COVID case counts and current COVID restrictions in country of origin? 🡨 can be inferred from Country of origin
    - COVID-heavy countries tend to require more social distancing and public has greater anxiety over contracting COVID-19
  + Time zone 🡨 can be inferred from Country of origin
    - Time zone may influence sleep schedule, may be trickier to coordinate group projects and multi-person assignments
    - One upside with working with teammates in opposite time zones is that someone is always working on the final project
  + Chronotypes
    - Students in other time zones may alter their chronotypes to accommodate STA302H1, which may offset adverse effects of sleep loss on performance.
* Time spent in STA302H1 during weeks 3 – 4 includes completing the mini assignments
  + This explains why Weeks 3 – 4 STA302H1 study times may be slightly inflated
* Sample size is roughly n = 200 people
  + Excludes students who dropped STA302H1
  + Blank entries and imputations for missing survey responses and missing quiz scores due to some students skipping quizzes