There are many ideas for the thesis that will be done in the summer of 2023. While most of them are related to interests of mine, I do want to do any of them for the thesis. What I want to do is a project that might lead to a publication in a journal. I have never done any sort of academic research/writing and would welcome any feedback on how to proceed.

To answer the prompt, I have broken my ideas down into several categories:

1. **Sports**
   1. Hockey – determine the most valuable offensive player by season
   2. Tennis – test for a correlation of success due to height, weight, country
   3. Tennis – track the ball movement to come up with discernible patterns
   4. NHL/NBA – travelling salesman problem
   5. Major Sports – criminal background per league
   6. Physical performances based upon age (run times for person A from age 20 to age 60)
2. **New R Language**
   1. Create a text dictionary (AFINN, NRC, BING) for slang (Urban Dictionary)
   2. Create a sparklines package that allows for multiple columns
3. **Geospatial**
   1. Determine the true square footage per state - flatten out the state
   2. Railways (dependence, multi-variate layer with airports and streets)
   3. Modelling invasive species
4. **Cryptography**
   1. Identifying deep fake video
   2. Decoding secret/hidden messages
5. **Finance**
   1. Modeling uncertainty
   2. Crypto stock price compared to normal stock price with respect to social media sentiment
   3. Peak-Load Pricing
   4. Kuhn-Tucker for rationing
   5. Weekly Economic report
      1. Interest rates, population change, housing prices, GDP, etc.
   6. Cobweb model for agriculture
6. **Work**
   1. Labor analysis
   2. Predictions
   3. Dashboards

Most of the academic ideas listed below came from:

* <https://math.williams.edu/majors/colloquium-advisors-and-topics/>
* <https://population-europe.eu/events/calls-papers/call-papers-international-journal-data-science-and-analytics>
* <https://www.middlebury.edu/academics/math/requirements/seniorthesis/topics>

**Academic**

* Analyzing variable selection strategies
* Accounting for spare feature matrices
* Graph theory
* Game theory
* Mediation analysis
* Statistical distribution of eigenvalues for random matrices
* Point process models
* Distance between clusters
* Human migration
  + Environmental impact
  + Historical changes
  + Country -> city
  + Covid affect
* Clustering residuals to identify missing variables
* Mersennes Prime for odd numbers
* Mathematical models of conventional warfare
* Galois Theory
* Twin Primes
* Digital coin price compared to bank stability ratings (random panel regression)
* Analysis of “call for papers” to see text-sentiment/time-series clustering of topics