Kin 575: Principles of Applied Sports Analytics  
(replaced Kin 530A starting September 2021)

Date: September – December 2021  
Time: Tuesdays 9am and Thursdays 5pm

Instructors:  
Ming-Chang Tsai, Ph.D.  
Johann Windt, Ph.D.

Course Description  
This experiential course will enable students to develop interdisciplinary skills and gain an understanding of technical and applied knowledge in data science and analytics related to high performance sports. Students will develop techniques in data cleaning and processing, filtering, extraction of subject- / sport-specific metrics, and data interpretation with an emphasis on the application of technical skills and theoretical concepts on real-life competition, performance testing, and wellness monitoring data. Students will develop basic competencies in data manipulation using the programming language R and a basic conceptual understanding of statistical methods and data visualization methods.

Learning Objectives  
Students will;

- Understand data collection protocols, their advantages/limitations, various data types, cleaning, privacy and storage considerations.  
- Learn and use basic programming skills necessary for data collection and analysis.  
- Understand different types of data, descriptive and inferential analysis, and their use or mis-use in High Performance sport related contexts.  
- Apply data analysis techniques to establish evidence-based and data informed decisions related to high performance sport and or health outcomes  
- Students have basic competency to do data processing and statistical analysis using R software.  
- Learn how to visualize data and present key findings and analyses to synthesize critical decisions that are made in HP sport of health-related outcomes.

Expectations  
Students will  
- Be required to attend two (2) online sessions per week. Each session will be 1.5 hours in duration
● Be required to perform weekly data processing assignments which may require additional
time (~60-90 minutes) outside of weekly sessions
● Complete individual projects that relate to their specific context and will be able to
interpret results and deliver accurate recommendations from the analyses used.
● Do a final presentation on improving their current reporting templates using the
knowledge learned in class.

Course Evaluation
● Weekly Assignment 30%
  ○ 10 assignments - due weekly before lecture starts
● Sport Data Processing project 30%
  ○ Data cleaning
  ○ Data processing (ie. filtering…)
  ○ Metrics extraction
  ○ Statistical analysis
  ○ Reporting (interpretation and visuals)
● Final Project 40%
  ○ Project proposal (5%)
  ○ Presentation (20%)
    ■ Peer evaluation (10% - done by 2 students, 5% each)
    ■ Instructor Evaluation (10%)
      ● Identify areas for improvement
      ● Describe processing steps
        ○ Demonstrate along with the analysis code
      ● Integrate inferential statistics into reporting template
      ● Interpretation of results
  ○ Report (15%)
    ■ Background
      ● Gaps the report is addressing
      ● Metrics used in report template
    ■ Methods
      ● Technique used to improve template
    ■ Results and interpretation
      ● Improvements made
      ● Interpretation on results

Course Schedule
Lectures will be delivered as synchronous (ie virtual live) or asynchronous (ie pre-recording)
Lab sessions will be delivered as synchronous
<table>
<thead>
<tr>
<th>Week</th>
<th>Dates</th>
<th>Class</th>
<th>Topics</th>
</tr>
</thead>
</table>
| 1    | Sept 9, 2021 Thurs 5pm | Lecture | Course introduction
Adopting a Data Informed Approach
- Course overview
- Why data analytics?
  - Definition of data science
- History of data science in sport
- Importance of data driven culture
- Data science framework
- Fit into Canadian high performance system |
| 2    | Sept 14, 2021 Tues 9am  | Lab    | Introduction to software
- Installation of R and RStudio
- Installing libraries
- Basic data structure
Numbers, Strings, Variables
- Data type concepts
- Numeric operations
- Character operations
- Conversion of data type |
| 2    | Sept 16, 2021 Thurs 5pm | Lecture | Data Governance
- Definition
- Ownership
- Research ethics and informed consent
- Safeguarding information
- Secondary uses of data
Developing Research Plan
- Identifying investigative question
- Assessing metrics/equipment/tools used
- Analysis to be used
- Data Collection
- Data structure/entry/formatting
- Identify different types of sensors they could use. |
- Present the basis of data acquisition. Analog vs digital signals. Sampling frequency, dynamic range.
Specific utilization of different sensors.

| 3  | Sept 21, 2021 Tues 9am | Lab | Comparison and logic  
- Comparison operators  
- Logic operators  
- Filtering data frame with operators  
Assignment #1 due (before class) |
|----|-----------------------|-----|--------------------------------------------------------------------|
| 3  | Sept 23, 2021 Thurs 5pm | Lecture | Validating Research Methods and Tools  
- Sensitivity of methods and tools  
- Variance in data  
- Reliability of methods and tools  
  - internal/external reliability  
  - ICC  
  - Bland-Altman plots  
- Validity of methods and tools  
- Accuracy  
Precision |
| 4  | Sept 28, 2021 Tues 9am | Lab | Syncing data sources  
Descriptive statistics  
- Mean  
- Standard deviation  
- Interquartile range  
- outliers  
Assignment #2 due (before class) |
| 4  | Sept 30, 2021 Thurs 5pm | NO CLASS |  |
| 5  | Oct 5, 2021 Tues 9am | Lab | If statements  
For loops  
Assignment #3 due (before class) |
| 5  | Oct 7, 2021 Thurs 5pm | Lecture | Formatting and structuring data  
• Examples of proper data structure conventions  
  o Date  |
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Time</th>
<th>Location</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Oct 12, 2021</td>
<td>Tues 9am</td>
<td>Lab</td>
<td>Filtering&lt;br&gt;Basic plotting&lt;br&gt;- Line&lt;br&gt;- Bar&lt;br&gt;Assignment #4 due (before class)</td>
</tr>
<tr>
<td>6</td>
<td>Oct 14, 2021</td>
<td>Thurs 5pm</td>
<td>Lecture</td>
<td>Introduction to sources of data&lt;br&gt;Descriptive statistics&lt;br&gt;- Distributions&lt;br&gt;- Mean&lt;br&gt;- Standard deviation/variance/noise&lt;br&gt;- Coefficient of variance&lt;br&gt;- Effect size&lt;br&gt;- Smallest worthwhile change</td>
</tr>
<tr>
<td>7</td>
<td>Oct 19, 2021</td>
<td>Tues 9am</td>
<td>Lab</td>
<td>More descriptive metrics&lt;br&gt;Assignment #5 due (before class)</td>
</tr>
<tr>
<td>7</td>
<td>Oct 21, 2021</td>
<td>Thurs 5pm</td>
<td>Lecture</td>
<td>Interpreting and reporting descriptive data&lt;br&gt;- Reporting requirements&lt;br&gt;- Appropriate statement to communicate results&lt;br&gt;Filtering&lt;br&gt;- Moving averages</td>
</tr>
<tr>
<td>8</td>
<td>Oct 26, 2021</td>
<td>Tues 9am</td>
<td>Lab</td>
<td>Rearranging data frame for analysis&lt;br&gt;T-Test&lt;br&gt;Assignment #6 due (before class)</td>
</tr>
<tr>
<td>Week</td>
<td>Date</td>
<td>Time</td>
<td>Location</td>
<td>Topic</td>
</tr>
<tr>
<td>------</td>
<td>--------------</td>
<td>-------</td>
<td>----------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>8</td>
<td>Oct 28, 2021</td>
<td>Thurs 5pm</td>
<td>Lecture</td>
<td>Inferential Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Purpose of inferential analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Null hypothesis significance testing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Level of significance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Type 1 error</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Type 2 error</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Comparing means</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- T-test (one sample, independent, paired)</td>
</tr>
<tr>
<td>9</td>
<td>Nov 2, 2021</td>
<td>Tues 9am</td>
<td>Lab</td>
<td>ANOVA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Assignment #7 due (before class)</td>
</tr>
<tr>
<td>9</td>
<td>Nov 4, 2021</td>
<td>Thurs 5pm</td>
<td>Lecture</td>
<td>ANOVA</td>
</tr>
<tr>
<td>10</td>
<td>Nov 9, 2021</td>
<td>Thurs 5pm</td>
<td>Lecture</td>
<td>Investigating patterns/trends and relationships</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Simple Linear regression</td>
</tr>
<tr>
<td>10</td>
<td>Nov 11, 2021</td>
<td></td>
<td>NO CLASS</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Nov 16, 2021</td>
<td>Tues 9am</td>
<td>Lecture</td>
<td>Investigating patterns/trends and relationships</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Multiple Linear regression</td>
</tr>
<tr>
<td>11</td>
<td>Nov 18, 2021</td>
<td>Thurs 5pm</td>
<td>Lab</td>
<td>Linear Regression</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Simple</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- multiple</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Assignment #8 due (before class)</td>
</tr>
<tr>
<td>12</td>
<td>Nov 23, 2021</td>
<td>Tues 9am</td>
<td>Lecture</td>
<td>Data Visualization</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Univariate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Bivariate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Categorical vs Categorical</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Categorical vs Quantitative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Quantitative vs Quantitative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Multivariate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>- Grouping</td>
</tr>
<tr>
<td>12</td>
<td>Nov 25, 2021</td>
<td>Thurs 5pm</td>
<td>Lab</td>
<td>Advanced plotting (GGPLOT2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Assignment #9 due (before class)</td>
</tr>
<tr>
<td>Week</td>
<td>Date</td>
<td>Time</td>
<td>Type</td>
<td>Topic</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>--------</td>
<td>-------</td>
<td>----------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| 13   | Nov 30, 2021 | Tues 9am | Lecture | **Data Visualization**  
- Time Series  
- Customizing graphs  
Axes, colors, points & lines, fonts. Legends, labels, annotations   |
| 13   | Dec 2, 2021  | Thurs 5pm | Lab    | More advanced plotting (GGPLOT2)  
Reporting  
Assignment #10 due (before class)   |
| 14   | Dec 7, 2021  | Tues 9am | Lecture | **Integrating Technology into your sporting environment**   |