Topic: Principles of Sport Data Analytics

Date: September – December
Time: Tuesdays 9:00-10:30am and Thursdays 5:00-6:30pm, PT

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

Course Description
This experiential course will empower students to 1) critically think about how data can help them in their practice, 2) to feel more comfortable and confident to collect, handle, and report data, and 3) to make more data-informed decisions. Students will develop interdisciplinary skills and better understand the data pipeline in high performance sports. They will develop techniques in data collection, data cleaning and processing, filtering, extracting subject-/sport-specific metrics, and interpreting data related to their own sporting context. Students will develop their skills in Microsoft Excel and learn basic competencies in the programming language R so that they feel more comfortable with sport related data in their applied practice. to be able to better describe and implement statistical and data visualization methods.

Learning Objectives
By the end of course, students will be able to:

- Collect several data types in different ways, compare their advantages/limitations, describe various data types, and consider how to clean data appropriately, store it properly, and ensure data privacy is adhered to.
- Efficiently use Microsoft Excel for data collection and storage, perform basic programming skills in R, and use both platforms to visualize and analyze data. These tools will increase students' proficiency to handle and analyze data in their own coaching practice.
- Appraise and implement different types of data analysis (e.g. descriptive and inferential analysis), to establish evidence-based and data-informed decisions related to high performance sport and or health outcomes.
- Visualize data effectively and communicate analytical findings. These skills will facilitate students' abilities to inform key decisions in high performance sport or health-related environments.

Expectations

Students will

- Be required to attend one synchronous online session per week, which will be 1.5 hours in duration (Thursday 5:00-6:30pm). Thirty minutes of this lecture will be devoted to discussing the pre-recorded lecture content and reading material for the week. Sixty minutes will be considered 'lab' time where students synchronously work through practical assignments.
- The second class of the week (in lieu of the Tuesday AM time allotment) will be asynchronous. Students will be expected to:
    - Listen to the pre-recorded lecture.
    - Read the assigned articles.
    - Engage in critical, online discussion boards regarding assigned course content.
- Be required to complete four analytics assignments which will require additional time (~60-90 minutes) outside of weekly sessions.
- Complete an overarching final project using all the skills learned in class. This project will relate to students' own applied environments/practice that either answers a specific research question or improves an existing reporting template that coaches use in their everyday practice. This may also relate to their planned KIN 596 project. Students will interpret results and deliver accurate recommendations from the analyses used.

Course Requirements

Students will be required to attend one synchronous class each week, participate in class discussions, and engage in online discussion forums. Additionally, there will be a requirement to complete pre-class activities and all course related assignments within the timeframes provided.

Academic Accommodation for Students with Disabilities

The University's goal is to ensure fair and consistent treatment of all students, including students with a disability, in accordance with their distinct needs and in a manner consistent with academic principles. Students with a disability who wish to have academic accommodation should contact Access and Diversity without delay.

Academic Integrity

All UBC students are expected to behave as honest and responsible members of an academic community. Breach of those expectations or failure to follow the appropriate policies, principles, rules, and guidelines of the University with respect to academic honesty may result in disciplinary action. It is your responsibility to become familiar with the University of British Columbia's Academic Honesty and Plagiarism Policies, as well as the Student Declaration and the consequences of violating these policies.

Readings and Resources

There are no required texts for this course. Assigned and recommended reading materials will be provided in Canvas. All lecture and specific assignment content will also be available on Canvas.

Course Evaluation

Participation 20%:

- In class participation (10%): Students will be expected to attend lectures and labs and participate in class discussions when topics are presented to debate/discuss.
- Online participation (10%): Students will receive one paper to read every ~2 weeks to facilitate their critical thought and appraisal of key course concepts. To facilitate engagement with the material, a discussion prompt on Canvas will be provided for each required reading. Students will be expected to critically respond to each reflection question and interact with their peers' responses, with evaluation based on their consistency in responding to each discussion prompt and interacting with at least 2 of their peers, as well as the quality of their responses.

Category	Poor	Below Average	Above Average	Good	Total
In Class Participation	Below 50% attendance at synchronous classes, very little engagement when present. 1-2.5	50-75% Attendance at synchronous classes Sometimes engages in discussions 2.5-5	>75% attendance at synchronous classes. Regularly engages in discussions and asks relevant questions. 5-7.5	>75% attendance at synchronous classes. Regular, critical, thoughtful engagement. 7.5-10	_____ 10
Online Participation	Student answers <50% of online discussion prompts and/or fails to interact with classmate posts. 1-2.5	Student answers 50-75% of online discussion prompts and peer posts with some critical reflection shown. 2.5-5	Student answers >75-95% of online discussion prompts and peer post requirements with critical thought and engagement. 5-7.5	Student answers >95% of online discussion prompts and peer response requirements with critical reflection evident. 7.5-10	_____ 10
					/20

Lab Assignments 30%

- Four lab assignments that focus on different components of the data pipeline are to be completed by each student individually:
    - Data collection, organization, and cleaning (7.5%)
    - Descriptive analysis (7.5%)
    - Inferential analysis (7.5%)
    - Data visualization and reporting. (7.5%)
- Each lab will consist of 1-5 questions from the course material and be graded out of 10 total marks.

Final Project 50%

The final project is an original data analysis project where students synthesize and implement the knowledge they've learned over the semester. In consultation with the course professors, students are to develop/identify a relevant dataset, visualize the data, develop a specific research question, apply a statistical analysis, interpret the findings appropriately, and communicate them effectively. Students will write their findings in a report and communicate them in a brief presentation – covering all key components of communicating data analysis in sport.

Interim Check-Ins Completed on Time (10%)

Each main 'component' of the final project will be due at interim time points, to ensure that students progress efficiently on their project throughout the term. These check-ins will coincide with lab projects - due 1 week after each corresponding lab is due.

Presentation (20%: 10% instructor; 10% peer evaluation, done by 2 students, 5% each)

Presentations should be a minimum of 10 minutes, and a maximum of 15 minutes. Although students have freedom in presentation flow, they must include the following information.

Introduction:

- Subject matter background and theory behind the research question.
- A clearly stated hypothesis / research question.

Dataset Description and Analysis Methods:

- Describe the dataset you had (fictitious or real), which variables were in the set, how you structured the dataset.
- Describe what analysis you chose to implement, and why.
- Include a slide on the code you used to run your test, and the output.

Results:

- Include a data visualization that represents the data in a way that reflects your research question and statistical analysis.
- Report the test results – which test you ran (e.g. t-test, ANOVA, regression), the effect estimate (e.g. group difference), the 95% confidence interval, and the p-value, along with any other relevant values (e.g. R^2 of the regression model).

Discussion:

- What do your results/findings mean for your research question?
- Strengths/limitations of your approach?
- Recommended/desired next steps.

Presentation Rubric:

Group #					Total
Organization	Audience cannot understand presentation because there is no sequence of information. (1)	Audience has difficulty following presentation because student jumps around. (2)	Student presents information in logical sequence which audience can follow. (3)	Student presents information in logical, interesting sequence which audience can follow. (4)	_____ 4
Subject Knowledge	Student does not have grasp of information and content presented. No objectives/agenda, intro and analysis of content very poor & no references. (1)	Student is uncomfortable with information but is able to show some content knowledge. No objectives/agenda, intro and analysis of content poor & no references. (2)	Student is at ease with the content and presents it well. Objectives/agenda present, intro and analysis of content good but no references. (2)	Student demonstrates full knowledge (more than required) and is confident with material. Clear objectives/agenda, intro and analysis of content great w references. (4)	_____ 4
Graphics/Engagement	Student uses superfluous graphics or no graphics. (0.5)	Student occasionally uses graphics that rarely support text and presentation. (1)	Student's graphics relate to text and presentation. (1.5)	Student's graphics explain and reinforce screen text and presentation. (2)	_____ 2
Mechanics	Four or more spelling errors and/or grammatical errors. (1)	Three misspellings and/or grammatical errors. (2)	No more than two misspellings and/or grammatical errors. (3)	No misspellings or grammatical errors. (4)	_____ 4
Elocution/Clarity And Engagement	Student mumbles, incorrectly pronounces terms, speaks too quietly for students in the back of class to hear. Presenter is not clear. (0.5)	Student's voice is low. Student incorrectly pronounces terms. Audience members have difficulty hearing presentation. Presenter isn't always clear. (1)	Student's voice is clear. Student pronounces most words correctly. Most audience members can hear presentation. Presenter is clear. (1.5)	Student uses a clear voice and correct, precise pronunciation of terms so that all audience members can hear presentation. Presenter is always clear. (2)	_____ 2
Q & A	Does not answer questions posed by the class. (1)	Answers some/or parts of questions posed by the class. (2)	Answers to all questions, but fails to elaborate. (3)	Answered the questions very well and was able to pose a thought-provoking statement/rebuttal. (4)	_____ 4
					/20

Report (20%) (6-10 Pages)

Background/Introduction:

- Subject matter background and theory behind the research question.
- A clearly stated hypothesis / research question.

Dataset Description and Analysis Methods:

- Describe the dataset you had (fictitious or real), which variables were in the set, how you structured the dataset.
- Describe what analysis you chose to implement, and why.
- Include a slide on the code you used to run your test, and the output.

Results:

- Include a data visualization that represents the data in a way that reflects your research question and statistical analysis.
- Report the test results – which test you ran (e.g. t-test, ANOVA, regression), the effect estimate (e.g. group difference), the 95% confidence interval, and the p-value, along with any other relevant values (e.g. R^2 of the regression model).

Discussion:

- What do your results/findings mean for your research question?
- Strengths/limitations of your approach? Recommended/desired next steps.

Course Schedule

Lectures will be delivered as asynchronous (i.e. virtual, pre-recorded), with 30 minutes of weekly synchronous time devoted to discussion of the pre-recorded content.

Lab sessions will be delivered as synchronous, tutorial-type sessions. Some introductory lab-material will also be pre-recorded and expected learning for students (e.g. Installing and getting familiar with R & RStudio).

Online Learning Resources for Excel & R:

Working in R:

- R For Data Science:
    - https://r4ds.had.co.nz/
- R Cheat Sheets
    - https://www.rstudio.com/resources/cheatsheets/

Working in Excel:

- Adam Virgile Excel & Google Sheets YouTube Videos:
    - https://www.youtube.com/c/AdamVirgile34/playlists
- Excel Tricks for Sports YouTube Channel
    - https://www.youtube.com/channel/UCagflprv_C-UPPdzSJ0bMCA

Course Schedule

Week	Lecture	Lab	Reading / Assignment
1	**Overview of sport analytics** - Define data science/analytics - Explain why data science can benefit coaches - List the steps of the data pipeline - Consider how to implement analytics in your practice.	**Project workflows and Excel 'Top up'** - Understanding and setting up analytical workflows in Excel and R. - Get on the 'fast track' to being an Excel power user.	Gamble, Paul, Lionel Chia, and Sian Allen. "The Illogic of Being Data-Driven: Reasserting Control and Restoring Balance in Our Relationship with Data and Technology in Football." *Science and Medicine in Football* 4, no. 4 (October 1, 2020): 338–41.
2	**Applied examples of how analytics have impacted sports.** - Athlete monitoring systems - Baseball - 'Moneyball' - NFL - 4th down conversion and the NFL big data bowl - NBA – 3-point shooting - Soccer – Expected goals - Hawkeye 'tennis'	**Setting up your data source/spreadsheet** - How do you structure a data source so that it can be analyzed and visualized appropriately? - Understand basic data types/formats - How to identify and prevent common errors.	Broman, Karl W., and Kara H. Woo. "Data Organization in Spreadsheets." *The American Statistician* 72, no. 1 (January 2, 2018): 2–10. https://doi.org/10.1080/00031305.2017.1375989.
3	**Data Governance** - Definition & Ownership - Research ethics and informed consent - Safeguarding information **Developing Research Plan** - Identifying investigative question - Assessing metrics/equipment/tools used - Analysis to be used - Data Collection - Data structure/entry/formatting	**Exploring your data 1 - data visualization** - Data visualization for exploratory data analysis.	Midway, Stephen R. "Principles of Effective Data Visualization." *Patterns* 1, no. 9 (December 11, 2020): 100141. https://doi.org/10.1016/j.patter.2020.100141.

4	**Data Collection & Processing**
	- Identify different types of sensors.
	  - Present the basis of data acquisition. Analog vs digital signals. Sampling frequency, etc.
	- Why collection & processing are important
	- Data cleaning
	  - Outliers (causes, finding, cleaning)
	  - Formatting types
	- Data filtering approaches
		Exploring your data 2 - descriptive statistics
		- Basic descriptive statistics for exploratory data analysis
		- Applied examples of descriptive statistics
			Lab Assignment 1 Due: Setting up your data

5	**Validity** - Define validity and reliability - List the different types of validity and reliability - Explain why validity and reliability are important - Describe how unified validity theory informs critical thinking about data for coaches	**Data cleaning and manipulation.** - Joining data sets	**Final Project Check-In 1: Dataset Prep** Milligan, Gemma S., Tara J. Reilly, Bruno D. Zumbo, and Michael J. Tipton. "Validity and Reliability of Physical Employment Standards." *Applied Physiology, Nutrition, and Metabolism* 41, no. 6 (Suppl. 2) (June 2016): S83–91. https://doi.org/10.1139/apnm-2015-0669.
6	**Introduction to statistical analysis:** - Overview types of approaches (e.g. descriptive, inferential, predictive). - Provide examples of how each type of analysis is used in practice.	**Inferential statistics 1.** **T-test & ANOVA**	**Lab 2 Due: Exploring your data (visualization and descriptive statistics)** Assignment 2 due and initial dataset prepped for final project.
7	**Descriptive statistics** - Distributions - Mean, standard deviation/variance/noise - Coefficient of variance - Effect size, smallest worthwhile change	**Inferential statistics 2.** **Regression** - Understand and perform a linear regression model. - Overview the principles of logistic regression for binary outcomes.	**Final Project Check-In (exploring your data)** Robertson, Samuel, Jonathan D. Bartlett, and Paul B. Gastin. "Red, Amber, or Green? Athlete Monitoring in Team Sport: The Need for Decision-Support Systems." *International Journal of Sports Physiology and Performance* 12, no. Suppl 2 (December 14, 2016): S2-73. https://doi.org/10.1123/ijspp.2016-0541.

8	**Basic probability theory and inferential statistics**
	- Purpose of inferential analysis
	- Null hypothesis significance testing
	  - Level of significance
	  - Type 1 error
	  - Type 2 error
	- Comparing means
		Advanced data visualization and reporting	**Inferential statistics lab assignment due.**

9	**Inferential statistics 2** - Identifying relationships between variables - Multiple regression analysis	**Creating a session report** - Create a report summarizing data from a training session.	**Final project interim check-in (inferential stats)** Rijn, Marieke H.C. van, Marieke H.c, Anneke Bech, Jean Bouyer, Van Den Brand, and Jan A.j.g. "Statistical Significance versus Clinical Relevance." *Nephrology Dialysis Transplantation* 32, no. suppl_2 (April 1, 2017): ii6–12. https://doi.org/10.1093/ndt/gfw385.

10	**Data communication and visualization** - Describe the basic principles of data visualization and how humans perceive images/information - List the main types of data visualizations and the general use cases for each type. - Explain best practices for building a chart from a dataset, and common ways that images are used to deceive viewers.	**Building a monitoring dashboard** - Using data on athlete load and wellness, build a dashboard that visualizes the team and individual athlete data over time.	**Data visualization/reporting assignment & final project check-in due**

11	**Integrating technology into your practice**
	- Understand the key questions you should ask when you consider potential technologies for your practice, in particular those around:
	  - Potential uses
	  - Validity related evidence
	  - Data management bandwidth
	  - Implementation burden
		Project tutorial - from raw data to reporting
		- This lab is for any applied Q&A related to the final project, and we will work through an example research project comparing the means of two groups.
		- Q&A
			Windt, Johann, Kerry MacDonald, David Taylor, Bruno D. Zumbo, Ben C. Sporer, and David T Martin. "'To Tech or Not to Tech?' A Critical Decision-Making Framework for Implementing Technology in Sport." *Journal of Athletic Training* 55, no. 9 (September 1, 2020): 902–10. https://doi.org/10.4085/1062-6050-0540.19.

12	**Exploring the future of analytics in sport - 'big data', 'machine learning', 'computer vision' etc.**	**Project tutorial - from raw data to reporting** - This lab is for any applied Q&A related to the final project, and we will work through an example research project using regression modeling.	**Final Project Due**

Additional Recommended Readings/Resources:

Visualization:

- Franconeri, Steven L., Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman. "The Science of Visual Data Communication: What Works." *Psychological Science in the Public Interest* 22, no. 3 (December 1, 2021): 110–61. https://doi.org/10.1177/15291006211051956.
- Nordmann, Emily, Phil McAleer, Wilhelmiina Toivo, Helena Paterson, and Lisa DeBruine. "Data Visualisation Using R, for Researchers Who Don't Use R." PsyArXiv, June 21, 2021. https://doi.org/10.31234/osf.io/4huvw.
- From Data to Viz: https://www.data-to-viz.com/
- Hands on Data Viz (Open Access Book): https://handsondataviz.org/

Statistics:

- Riemann, Bryan L., and Monica Lininger. "Statistical Primer for Athletic Trainers: The Difference Between Statistical and Clinical Meaningfulness." *Journal of Athletic Training* 50, no. 12 (December 2015): 1223–25. https://doi.org/10.4085/1062-6050-51.1.04.
- Mark, Daniel B., Kerry L. Lee, and Frank E. Harrell. "Understanding the Role of P Values and Hypothesis Tests in Clinical Research." *JAMA Cardiology* 1, no. 9 (December 1, 2016): 1048–54. https://doi.org/10.1001/jamacardio.2016.3312.
- Interactive visualizations of statistical concepts: https://rpsychologist.com/viz
- **Seeing Theory:** Interactive visualizations of probability theory and statistics. https://seeing-theory.brown.edu/
- Using ggplot to visualize statistical comparisons: https://indrajeetpatil.github.io/ggstatsplot/articles/web_only/ggbetweenstats.html

Technology/Integration/Data Science Systems:

- Windt, Johann, David Taylor, Dean Little, and Ben C. Sporer. "Making Everyone's Job Easier. How Do Data Scientists Fit as a Critical Member of Integrated Support Teams?" *British Journal of Sports Medicine*, August 14, 2020. https://doi.org/10.1136/bjsports-2020-102938.
- Robertson, Samuel, Jonathan D. Bartlett, and Paul B. Gastin. "Red, Amber, or Green? Athlete Monitoring in Team Sport: The Need for Decision-Support Systems." *International Journal of Sports Physiology and Performance* 12, no. Suppl 2 (December 14, 2016): S2-73. https://doi.org/10.1123/ijspp.2016-0541.