School of Business Administration

SBA Mission Statement:
"We develop socially responsible business leaders with a global mindset through academically rigorous, relevant, and values-based education and research."

BUAN 370 (01 & 02): DESCRIPTIVE ANALYTICS & DATA MANAGEMENT

LOCATION: KIPJ H

TIMES: MONDAYS & WEDNESDAYS, 2:30PM-3:50PM (01) & 4pm-5:20pm (02)

FALL SEMESTER 2021
SYLLABUS

Instructor(s): Dr. Steve Levkoff, PhD, CAP® & Dr. Carl Rebman Jr, PhD
E-mail: slevkoff@sandiego.edu, carlr@sandiego.edu
Instructor Webpage(s): http://stevelevkoff.com(not recently updated);
http://carl.sandiego.edu/buan370/
Course Webpage: http://ole.sandiego.edu
Office: https://sandiego.zoom.us/j/98929383879
Office Hours: Tuesdays & Thursdays, 4pm-5pm on Zoom (virtual location TBD). Please let me know in advance via email if you plan on attending so I can have the Zoom session set appropriately.

Course Description
Advances in our capability to generate and collect information coupled with decreasing disk-space prices are pushing us toward a world centered around data management. Data preparation and storage are the foundation of today's business analytics. They ensure data are properly processed for later meaningful analysis. Data preparation includes data cleansing and data transformation. The objective of data preparation is to collect the data from various sources into a single location and transform it into a form that is ready for later analysis. Databases are at the heart of modern commercial application development for data storage. Once data is prepared and properly stored, the first step of analysis usually involves summarizing basic facts about what has happened in the past. This preliminary examination of data falls in the category of descriptive analytics (exploratory data analysis). The purpose of this course is to provide a comprehensive introduction of the data management process - from data preparation, storage, to descriptive analytics applications.

Course Outcomes

After taking this course you will be able to:

- Describe and identify different basic data structure and their characteristics
- Access large datasets and doing the "wrangling" needed to prepare them for analysis
- Describe basic relational database and data warehouse concepts
- Write SQL queries for a variety of data definition and data manipulation scenarios
- Identify appropriate visualizations for different types of data and best practices for creating visualizations use interactive data visualization to understand and analyze data

Prerequisites

In order to be enrolled in this course, you should have satisfactorily completed ITMG 100, ECON 216 (OR ECON 217). Some experience scripting/programming will be useful, but not required.

Workload & Expectations

Analytics is time-intensive, but it pays off! Average "Business Analyst" salary is ≈$72,000 and "Analytics Manager" salary is ≈ $105,000 (Indeed.com, 2018). Salaries of seasoned "Data Scientists" exceed $120,000 annually. Analytics is naturally a time-intensive field, which is why salaries for business analytics professionals are higher than in other fields. Therefore, you should expect to spend an average of 6 – 8 hours OUTSIDE of class per week working on and learning the material. Some weeks you may spend more than 6 – 8 hours and some weeks you may spend less. It is vitally important that you be open to "hacking" by taking advantage of open access resources on the internet (ie: Stack Overflow, DataCamp, etc.) to *independently* supplement the course materials assigned. Being able to experiment *on your own* with trial and error is an important part of studying analytics. It is also critically important that you *follow directions carefully* to streamline efficiency and to avoid redundancies in questions and processes. An example of how you may allocate your 6 – 8 hours study time each week is as follows (everyone is unique so you may personally have a different allocation of your time; this is just an example):

- 2 hours: reading the relevant assigned readings and taking your own notes on the readings
- 1 hour: reviewing examples and notes from lecture
- 1 – 2 hours: working on assignments (worked examples, code chunks, problem sets, projects)
- 1 – 2 hours: going to office hours and/or reviewing with peers (group study)
- 1 hour: troubleshooting and researching

Readings

Required :

[1] R Programming for Data Science, Roger D. Peng, The Johns Hopkins University, Leanpub. This book will act as a reference guide for programming in R. It contains useful examples including the code, output, and hyperlinks to video lessons associated with the book made by the authors. It can be found for free online at https://leanpub.com/.

[2] Database Design – 2nd Edition by Adrienne Warr and Nelson Eng is used under a CC BY 4.0 International License. Download for free from the B.C. Open Textbook Collection.

Recommended:

[3] A First Course in Database Systems, 3rd Edition, by Jeffrey Ullman and Jennifer Widom. ISBN- 13: 978-0136006374 ISBN-10: 9780136006374

[4] ggplot2: Elegant Graphics for Data Analysis, Hadley Wickham, Springer, 2nd Edition. Several of the worked examples will be referenced as case studies.

[5] Exploratory Data Analysis with R, Roger D. Peng, The Johns Hopkins University, Leanpub. This book is also written in R and contains some useful worked examples with code, output, and hyperlinks to video content made by the author. This will also act as a reference guide for a few of the worked case studies we will investigate. It can be found for free online at https://leanpub.com/.

[6] Statistical Inference for Data Science, Brian Caffo, The Johns Hopkins University, Leanpub. This book is a reference guide intended to review concepts introduced in ECON 216 but introduces the material with coded examples in R. We will not work directly from this resource, but it should provide for a good review of material if you are rusty. It can be found for free online at https://leanpub.com/.

[7] Regression Models for Data Science in R, Brian Caffo, The Johns Hopkins University, Leanpub. This book is a bit more advanced in terms of mathematical background, but will be a useful reference guide for reviewing material introduced in ECON 216 and

examples implementing regression models in R. It can be found for free online at https://leanpub.com/.

[8] Resource for relational database and SQL: http://ovid.cs.depaul.edu/Classes/CSC355-S14/CSC355-links.htm

Software Tools:

Recommended:

[1] R, a statistical computing package that can be downloaded for free here: https://www.r-project.org/.

[2] RStudio, An integrated development environment (IDE) for R, an open source statistical computing software package. You can download the most recent (desktop) version (compatible for Mac and Windows) for free at www.rstudio.com. All other packages and extensions can be loaded seamlessly from within the IDE. (It is also possible to run RStudio remotely using a virtual machine (VM) by setting up a free account and carefully poking around here: https://dataplatform.cloud.ibm.com/.)

[3] Tableau. The course project will be done using this tool to generate visualizations. You can download a free student license here: https://www.tableau.com/academic/students. Be sure to retain your access code for your records (they will email it to you when you start your account).

 [4] Microsoft Excel. All computers in the labs have access to this software.

[5] MS SQL Server. Access will be provided via EPIC.

[6] Microsoft Access. All computers in the labs have access to this software.

Professionalism (attendance, participation, etc.)
Attendance is critical for success in this class. It is in the student's best interest to attend class every day. It is understood that events may cause one to have to miss class to which advance notice is preferred. A portion of the course grade is allocated towards professionalism (includes attendance, class participation and professionalism). Failure to be in class in the event attendance is taken can result in a reduction in this grade component. This can also include being late after attendance has been taken or leaving class early. Furthermore, failure to attend more than 50% of all class lecture (and/or lab sessions) will result in failing the course. Lastly keep in mind that attendance,

participation, and professionalism are three separate items, that is, you need to more than just show up; you are expected to be prepared and contribute to class activities and discussion. Also engaging in Internet surfing, IMs, games, or sending emails during class will not be tolerated and will result in a reduction in points. Plus, you will find it much easier to keep up and master the material if you attend regularly. You are responsible for all material covered in class or assigned during a class even if you were absent. Please contact me as soon as possible if you know you must miss a class.

SWIRL Submissions: Swirl, a series of open source (free) R packages, is a self-instructional programming tutorial that is run through the RStudio IDE. The SWIRL package exercises function to quickly train you in RStudio and in the R programming syntax. Completion of SWIRL assignments for the R Programming series are sent via email notification to the instructor. You can find the instructions for installing the swirl packages here: http://swirlstats.com/students.html. The five packages that will be used in this course are R Programming, Getting and Cleaning Data, Exploratory Data Analysis, Statistical Inference, and Regression Models. The swirl course repository can be found here: https://github.com/swirldev/swirl_courses#swirl-courses. There may also be several other benchmark submissions that will be required to be posted to your GitHub repository and shared with the instructor for grading aside from the swirl exercises (see projects below). The R Programming E SWIRL modules (15 of them) are due for completion on 9/19/21 by 11:59pm. It shouldn't take more than a few hours to complete these modules to ensure you have a (very) basic understanding of how to use R. I will post instructions for installation as well as instructions for submission that must be followed *extremely carefully*. Completion of these tasks will guarantee at some very fundamental level that everyone is on the same page in terms of utilizing the software package. I will obviously provide additional instruction to supplement when appropriate as the expectation for this course is that nobody has any coding/programming experience.

Exams
There will be two noncumulative midterm exams administered at approximately weeks 5 and weeks 10 of the course. The final exam takes the form of the final project for the course (Monday 12/13/21 from 2-4pm for section 01 and Wednesday 12/15/21 from 5-7pm for section 02).

Course Project
The goals of the team project are (1) to integrate and apply what you have learned in the class; (2) to give you experience working in team situations; and (3) to enhance your communication and presentation skills. Throughout the course, you will complete two major analytics tasks: a descriptive task and a database task. Details for these two tasks will be made available after the appropriate material has been covered. In general, the projects will provide you the opportunity to express yourself creatively within the constraints of the task using the statistical software package to showcase your analysis and demonstrate your competency with coding and debugging your work.

Grading:

Professionalism (attendance, participation, etc.)	10%
SWIRL Programming Assignments (15 of them)	15%
Exam 1	15%
Database Assignments (5 of them)	15%
Exam 2	15%
Project	30%
Total	100%

The course is graded on a relative curve (as is any college course). In particular, students will all be ranked from highest to lowest course score according to your final course grade calculated from the raw exam score weighting above. Letter grade assignments will depend on your percentile ranking in the class and a subjective assessment by the instructor in borderline cases (say, if there was marked improvement).

Classroom Decorum and Email

To avoid distracting others in the classroom, please arrive on time. In extreme cases the instructor reserves the right to decrease the letter grade by an entire letter for a student due to inappropriate behavior. Please restrict the use of email to the minimally necessary volume, put the course number (BUAN 370) in the subject of the email
– I teach other classes and this is the fastest way for me to know who I'm responding to. Put your *full name* at the end of email messages as well. Your email decorum should be professional. You should make sure you address the email formally and properly and use college level grammar.

Academic Integrity

You, your colleagues, faculty, staff, and alumni are the University of San Diego. These and many other persons have worked very hard since the founding of USD in 1949 to build a quality university. The philosophy and mission of USD http://www.sandiego.edu/about/mission-vision-values.php emphasizes the idea of personal and academic integrity. The following is a synopsis of the academic integrity policy. For more information click on this hyperlink http://www.sandiego.edu/associated-students/branches/vice-president/honor-council/integrity-policy.php or download this pdf http://www.sandiego.edu/conduct/documents/HonorCode.pdf

"All members of the University community share the responsibility for maintaining an environment of academic integrity since academic dishonesty is a threat to the University. Acts of academic dishonesty include a) unauthorized assistance on an examination; b) falsification or invention of data; c) unauthorized collaboration on an academic exercise; d) plagiarism; e) misappropriation of resource materials; f) any

unauthorized access of an instructor's files or computer account; or g) any other serious violation of academic integrity as established by the instructor."

Academic Dishonesty will not be tolerated in any form. Helping each other study is anticipated. However, only original work will be accepted. There will be no sharing of materials, wearing hats, or using cell phones/PDAs during tests. All assignments unless otherwise noted are individual assignments. If an incident of academic dishonesty occurs in this course the student could receive a grade of "F" for the semester and could possibly face further disciplinary action.

Student Disability Policy

The University of San Diego complies with the American with Disabilities Act and Section 504 of the Rehabilitation Act It is a University of San Diego Disability Services policy that when students are scheduling exams (midterm exams or final exams) in the DLDRC, they must submit an "Authorization to Administer Exam" at least one week prior to the exam date. If a student does not give DLDRC this notice, the DLDRC can deny them the right to the accommodation, as the DLDRC is not given adequate time to prepare (and set up office space during established exam scheduling times.) Please provide me (your instructor) with a Letter of Accommodation drafted by the DLDRC as soon as possible. There are no retroactive accommodations for Disabled students. As your instructor I am not obligated to provide accommodations until I receive the Letter of Accommodation drafted by the DLDRC. If a student does not give the Letter of Accommodation to me (your instructor), within adequate time to make exam arrangements, I am not obligated to fulfill any such request. If you have any questions or concerns about the process please contact, the USD Disability Services office (Serra Hall, Rm 300), phone at 619-260-4655 or via email at disabilityservices@sandiego.edu as soon as possible.

Student Athletes

You are responsible for providing me advanced written notice, (email), any time you will be unable to attend class. You are also responsible for determining alternate dates/times to make up missed work in class. Without written notice, no assignments, quizzes or tests, either due or taken in class, can be made up.

Tentative Course Schedule of Select Topics (Subject to Change)

Week	New
1	Intro to Analytics / Case Studies / Branches of Problem Types
2	The Learning Problem / Getting Started / Importing Data in R
3	Descriptive Statistics in R / An Example
4	Preprocessing & Cleaning Data Fundamentals / Data Structures
5	Panel Data Example / Cross Sectional Cleaning Example w / NHIS Survey Data

6	Automating Processes / For & While Loops / Boolean Logic / Automating Cleaning Example
7	Analytics Programming Exam / Stochastic Simulation / Monte Carlo
8	Intro to Database / Overview / Relational Databases / Entity-Relationship Diagram ERD
9	ERD / Database Design / Normalization
10	SQL / Connection to SQL Server / Writing SQL CRUD Queries
11	SQL Aggregates/ Joins Two Tables / Nested Queries / Views
12	SQL IIS Interactions / SQL Injections / Data Warehousing / Other OOB DB
13	SQL Skills Tests
14	Visualization Best Practices / GGPLOT2
15	Final Project Work Week
	FINAL EXAM - PROJECT PRESENTATION