Official Calendar: Techniques for computation, analysis, and visualization of data using software. Manipulation of small and large data sets. Automation using scripting. Real-world applications from life sciences, physical sciences, economics, engineering, or psychology. No prior computing background is required. Credit will be granted for only one of COSC 301, DATA 301 or DATA 501. [3-2-0] Prerequisite: Either (a) third-year standing, or (b) one of COSC 111 or COSC 122. Specific description: This course provides an introduction to data analytics to train students with practical industrial techniques for data manipulation, analysis, reporting, and visualization. This is not an introduction to programming using Python. Programming techniques will be taught to automate data analysis. Introduction to programming courses are COSC 111 or COSC 123. Prior computing experience is not required, but is helpful including COSC 122 or COSC 111. See course descriptions here. Topics include introduction to course. What is data analytics? Data Representation: Data and metadata; file formats and encoding; text/binary files 2 Excel I: Intro to Excel – formulas, formatting, aggregate functions Excel II: Data Analysis in Excel – sorting, filtering, charts, what-if scenarios, pivot tables 3 Excel III: Excel scripting – macros, VBA Databases I: Introduction to Relational Databases; Creating a Database 4 Databases II: Querying using SQL Databases III: Advanced Querying using SQL 5 Command Line: Linux/Unix Command Line Introduction Python I: Introduction to Python 6 Midterm 1: Wednesday October 9 Python II: Decisions and Loops 7 Python III: Reading and Writing Files Python IV: Data Analysis with Python 8 Python V: Data Analysis with Map-Reduce R I: Data Analysis with R – Brief statistics intro/review 9 R II: Data Analysis with R (cont.) Midterm 2: Monday November 4 10 R III: Data Analysis with R (cont.) GIS⇤ 11 Data Visualization I: Reporting Data Visualization II: Tableau 12+ Open science (guest lecture by Dr. Jason Pither Nov 20th) Review for final exam Data 501 final project presentations. Learning Outcomes: Ability to manipulate, extract, convert, and integrate data from different sources. Ability to perform advanced Excel analysis including what-if scenarios, pivot tables, and VBA scripting. Ability to use relational databases including creating tables and querying using SQL. Ability to use scripting programs to automate repetitive and large tasks and improve efficiency. Course Objectives: Understand data representation formats and techniques and how to use them. Experience using a wide-range of data analytics tools including Excel, SQL databases, GIS, and visualization and reporting software. Develop a computational thinking approach to problem solving and use programs to solve data tasks.