## Geography 30323: Data Analysis & Visualization

**Fall 2015**

**Texas Christian University**

**Tuesday/Thursday 9:30-10:50**

**Classroom: Tucker 353**

**Instructor: Dr. Kyle Walker** ([kyle.walker@tcu.edu](mailto:kyle.walker@tcu.edu))

### Course overview

Data literacy has emerged as an essential skill of the 21st century workforce across many different professional fields and contexts. Says Hal Varian, Chief Economist at Google:

“The ability to take data – to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it – is going to be a hugely important skill in the next decades, not only at the professional level but even at the elementary level for elementary school kids, for high school kids, for college kids. Because now we really do have essentially free and ubiquitous data.”

In this course, you will gain experience with several aspects of the data analysis process. Course topics will include:

* Obtaining data from a variety of sources, including databases and the Web;
* Cleaning messy datasets, and converting data between different formats and types;
* Using exploratory data analysis to summarize and generate insights from your data;
* Producing both static and interactive visualizations to both explore your data and communicate your findings with a wider audience.

In order to perform these tasks, you will also learn the basics of the Python programming language, which you will use for most of the data analysis activities in this course. Python has rapidly become one of the most popular languages for working with data and has numerous extensions for data wrangling, data analysis, and data visualization.

Writing code for data analysis is an essential skill, as it allows you to document your workflow and structure your analyses in a coherent, logical way. That said, this is not a course in software development, nor is it a programming course *per se*. Instead, think of Python as a tool you’ll be using to work in more flexible and powerful ways with data; in turn, you’ll be gaining skills in technical computing along the way.

There are no prerequisites for this course; however, you should bring a curiosity about data and an eagerness to learn new technologies!

### Course format

Mondays will be devoted to lecture and discussion, and Wednesdays will be used for applied tutorials and assignment work, with a couple exceptions. The class will be managed through its corresponding Learning Studio website, accessible through the portals at <http://my.tcu.edu> or <http://tcuglobal.edu>, and course materials will be available in the course GitHub repository at <https://github.com/walkerke/geog30323>. You will submit your assignments and view your grades on the Learning Studio website.

### Readings

There are two required texts for this course:

Downey, Allen (2014). *Think Python: How to Think Like a Computer Scientist*. Needham, MA: Green Tea Press. This book is available for free at <http://www.greenteapress.com/thinkpython/thinkpython.pdf>.

Yau, Nathan (2013). *Data Points: Visualization that Means Something*. Indianapolis, IN: Wiley. This book is available for free as an e-book through the TCU Library. If you would like a hard copy, you can purchase it from Amazon relatively inexpensively: <http://www.amazon.com/Data-Points-Visualization-Means-Something/dp/111846219X>.

### Evaluation

You will complete **three take-home examinations**, which collectively comprise **45 percent of your grade**. Each examination will be comprised of a structured set of homework problems that test you on the computing and analytical skills you've learned to that point in the course. Deliverables will be in the form of Jupyter Notebooks that show and describe the work you've done. Take-home exams will be due on **Sunday, October 18**; **Sunday, November 22**; and **Tuesday, December 15**, which is the day of our scheduled final exam. You'll receive the prompt for the take-home exam 2 weeks prior to the due date.

You will also have 14 **weekly assignments** which in total make up **42 percent of your grade**. Weekly assignments are assigned each Thursday, and due by the end of the day the following Thursday. You'll have devoted class time each week to work on the assignments.

In-class tutorials and participation exercises will constitute the remaining **13 percent** of your grade.

**The minimum percentages** to achieve a specific grade are as follows:

|  |  |
| --- | --- |
| Percentage | Grade |
| 94 | A |
| 90 | A- |
| 87 | B+ |
| 83 | B |
| 80 | B- |
| 77 | C+ |
| 73 | C |
| 70 | C- or Pass |
| 67 | D+ |
| 63 | D |
| 60 | D- |
| Below 60 | F |

I will award an incomplete (I) only in the most extreme and exceptional circumstances. Please notify me as soon as possible if you are in a situation where you feel you require an I.

The attendance policy for this course corresponds to the official TCU attendance policy, which reads:

Regular and punctual class attendance is essential, and no assigned work is summarily excused because of absence, no matter what the cause.

Extensions to deadlines and make-up participation credit will only be permitted in the instance of a **documented illness or emergency** or a **documented TCU-sanctioned activity**.

### Course Schedule

|  |  |  |  |
| --- | --- | --- | --- |
| **Week** | **Topic** | **Reading** | **Assignments** |
| Week 1: Aug 25-27 | What is data? | *Data Points*, Ch.1 | Data journalism |
| Week 2: Sept 1-3 | Programming I | *Think Python*, Chs. 1-2 | Python basics; intro to the Jupyter Notebook |
| Week 3: Sept 8-10 | Programming II | *Think Python*, Ch. 3 | Functions and modules in Python |
| Week 4: Sept 15-17 | Programming III | *Think Python*, Chs. 7-8 | Loops/conditionals/object-oriented programming in Python |
| Week 5: Sept 22-24 | Exploratory data analysis I | *Exploratory Data Analysis*, Chs. 2-3 (on Learning Studio) | Introduction to pandas |
| Week 6: Sept 29-Oct 1 | Exploratory data analysis II | *Data Points*, Ch. 4 | Exploratory visualization with pandas and seaborn |
| Week 7: Oct 6-8 | Data wrangling I | Wickham (2014), "Tidy Data" | pandas: subsetting, indexing, merging |
| Week 8: Oct 15 | Data wrangling II |  | pandas: groupby operations, reshaping data |
| Week 9: Oct 20-22 | Advanced I/O: databases, web, APIs | TBD | Web scraping and APIs in Python |
| Week 10: Oct 27-29 | Visualization best practices I | *Data Points*, Ch. 2 | Visualization with seaborn I |
| Week 11: Nov 3-5 | Visualization best practices II | *Data Points*, Ch. 5 | Visualization with seaborn II |
| Week 12: Nov 10-12 | Interactive data visualization | *Data Points*, Ch. 6 | Interactive plots with plotly and bokeh |
| Week 13-14: Nov 17-23 | Geographic data & visualization | *The Nature of Geographic Information*, Ch. 1 | Mapping with CartoDB |
| Week 15: Dec 1-3 | Communicating with data | TBD | Data dashboards with Tableau Public |
| Week 16: Dec 8 | Where to go from here? | TBD | N/A |

### Software

The main software used in this course is the Anaconda distribution of the Python programming language (version 3.4) from Continuum Analytics, which includes over 330 popular Python packages for data analysis, visualization, and scientific computing. Anaconda is completely free and can be installed from this link: <http://continuum.io/downloads>, and works on the most popular operating systems (Windows, Mac OS, Linux). I’ll show you how to install Anaconda on your own computer in class. If you prefer to do your work on campus, Anaconda will also be available in our classroom (Tucker 353) and in the Center for Urban Studies computer lab, SCHAR 2015A. Late in the course, we will also be using CartoDB and Tableau Public for interactive visualization.

All assignments and tutorials in this course will be completed in the Jupyter Notebook, which is an extension to Python that facilitates interactive computing. The Jupyter Notebook exemplifies literate programming, which refers to a combination of textual description, code, and images/graphics in the same document. In turn, you’ll use the Jupyter notebook to document your workflow and make your work reproducible.

### Other issues

**Academic conduct:**

Although attendance is not an explicit part of your course grade, it is necessary to successfully complete the in-class tutorials. If you have to miss class for a TCU-related event and you know when these events will take place (i.e. athletic competition, musical performance, etc.), give me advance notice so that we can work together to ensure that you remain on track. Additionally, if you need to miss class for illness or family emergency, please give me documentation and I’ll count the absence as excused.

This course will comply with TCU policies on academic conduct and plagiarism. The TCU statement on academic misconduct from the Student Handbook (Section 3.4) is below:

Academic Misconduct (Sec. 3.4 from the Student Handbook) –Any act that violates the academic integrity of the institution is considered academic misconduct. The procedures used to resolve suspected acts of academic misconduct are available in the offices of Academic Deans and the Office of Campus Life and are listed in detail in the Undergraduate Catalog (Student Policies>Academic Conduct Policy Details; <http://www.catalog.tcu.edu/current_year/undergraduate/>). Specific examples include, but are not limited to:

* Cheating: Copying from another student’s test paper, laboratory report, other report, or computer files and listings; using, during any academic exercise, material and/or devices not authorized by the person in charge of the test; collaborating with or seeking aid from another student during a test or laboratory without permission; knowingly using, buying, selling, stealing, transporting, or soliciting in its entirety or in part, the contents of a test or other assignment unauthorized for release; substituting for another student or permitting another student to substitute for oneself.
* Plagiarism: The appropriation, theft, purchase or obtaining by any means another’s work, and the unacknowledged submission or incorporation of that work as one’s own offered for credit. Appropriation includes the quoting or paraphrasing of another’s work without giving credit therefore.
* Collusion: The unauthorized collaboration with another in preparing work offered for credit.

In short: please don’t cheat, as it is a very serious offense and you will get caught. If you are in any way struggling in the course and tempted to cheat, please come talk to me so we can address your issues face to face.

Finally, the classroom is a place where diversity of opinions and perspectives is not only welcomed, but highly encouraged. I ask you to always be mindful and respectful of the diversity (broadly defined) of your classmates.

**Disability statement:**

TCU’s statement on disabilities is as follows:

Texas Christian University complies with the Americans with Disabilities Act and Section 504 of the Rehabilitation Act of 1973 regarding students with disabilities. Eligible students seeking accommodations should contact the Coordinator of Student Disabilities Services in the Center for Academic Services located in Sadler Hall, 1010. Accommodations are not retroactive, therefore, students should contact the Coordinator as soon as possible in the term for which they are seeking accommodations. Further information can be obtained from the Center for Academic Services, TCU Box 297710, Fort Worth, TX 76129, or at (817) 257-6567.

### Statement on use of the syllabus

This syllabus is intended for your use as a guide to assist in your planning for the semester. I reserve the right to make changes to the syllabus and schedule if necessary. However, rest assured that if I do make any changes to the syllabus, I will give you plenty of advance notice.