

PROGRAMA ABIERTO DE COMPLEMENTACIÓN Y AMPLIACIÓN DE LA CURRÍCULA DE LA MAESTRÍA 2020

La Maestría en Generación y Análisis de Información Estadística, en el marco de su Programa de Actualización Permanente de las Orientaciones de Estadísticas Económicas, Estadísticas Sociodemográficas y Estadísticas de Opinión y Mercado, en coordinación con la Secretaría de Extensión Universitaria y la Dirección de Posgrado, presenta el Ciclo de Seminarios y Cursos extracurriculares en Estadística para este año lectivo.

Coordinador: Jorge Fernández Bussy

Curso extracurricular: “Análisis espacial con Python”

Profesores: Felipe Gonzalez

Licenciado en Sociología (Universidad de Buenos Aires), maestrando en Generación y Análisis de Información Estadística (Universidad Nacional de Tres de Febrero) y MS en Urban Science and Informatics (New York University - Tandon School of Engineering).

Presentación y objetivo del curso

El interés por el análisis de datos ha crecido exponencialmente en los últimos tiempos con el aumento de la cantidad de datos producidos y las nuevas técnicas para procesarlos, analizarlos y modelarlos. El análisis espacial es uno de los campos específicos en esta nueva era, con sus técnicas particulares. Estos datos y técnicas comienzan a exigirnos interactuar con las computadoras y los programas de un modo que hasta ahora no era necesario. Nos exigen empezar “saber a programar” o mejor dicho “a pensar computacionalmente”. Este curso utilizará el lenguaje Python, un lenguaje de programación de licencia libre con una curva de aprendizaje relativamente sencilla y de crecimiento exponencial en el mundo académico y profesional ligado a la ciencia de datos. Este lenguaje es uno más de los muchos que hay y de los que vendrán. Por eso se priorizará aprender a pensar computacionalmente antes que saber programar. El mismo enfoque se dará a las herramientas de análisis espacial, se presentarán como formas de resolver problemas, intentando no adaptar la formulación del problema a la herramienta limitando el rango de acción (cuando uno solo tiene un martillo todos los problemas son clavos). En este curso uno podrá desarrollar herramientas de análisis espacial customizadas a su propio problema.

El objetivo del seminario es brindar conocimientos para utilizar el lenguaje de programación Python de modo práctico a través del uso de herramientas de análisis y visualización espacial.

El análisis espacial ofrece la facilidad de abordar procesos de pensamiento abstracto como el computacional, matemático y estadístico de un modo práctico, concreto y palpable para aquellos que nunca han tenido contacto con algún lenguaje de programación.

Todos los ejercicios y ejemplos serán sobre casos de estudio espaciales, con la posibilidad de visualizar en mapas el producto de nuestro trabajo. Esto permite que la introducción al pensamiento computacional y

programático sea menos ardua para los/las estudiantes que se encuentran haciendo sus primeras armas en el tema.

El abordaje programático y computacional presenta numerosas ventajas cuando trabajamos con grandes conjuntos o sets de datos, tareas repetitivas o demasiado particulares que requieren de procesos personalizados que no se encuentran predefinidos en ningún software. También tiene la ventaja de dejar rastro de todos nuestros procesos, garantizando que nuestro trabajo sea reproducible y verificable por nuestros pares.

Objetivos particulares:

- Obtener un manejo introductorio y práctico del lenguaje de programación Python
- Obtener un manejo introductorio y práctico de las librerías para manejo de datos de Python, en particular de las dedicadas a datos de tipo espacial
- Desarrollar estrategias y herramientas de visualización de datos, particularmente en mapas
- Obtener un manejo de las herramientas fundamentales de análisis de datos espaciales

Destinatarios

El seminario está destinado a estudiantes, técnicos y profesionales que quieran introducirse a algún lenguaje de programación, a la metodología programática y a la forma computacional de resolver problemas. También a aquellos que deseen profundizar su conocimiento en técnicas de análisis espacial. No se requiere conocimientos de programación previos. El seminario pretende introducir en el mundo de la programación a quienes no lo han hecho nunca, mediante actividades aplicadas al manejo y visualización de datos, particularmente en torno a temas espaciales. Sin embargo, sí se recomienda un conocimiento previo en programas de procesamiento de datos como SPSS, STATA, QGIS, ARCGIS, etc. También, por la naturaleza del seminario, aquellos/as con un gusto y aptitud por el pensamiento lógico-matemático les resultará más atractivo y productivo para sus trayectorias académico-profesionales. Una recomendación: es más importante contar con ganas y voluntad de hacer un pequeño esfuerzo en aprender algo nuevo antes que traer muchos conocimientos previos.

Temario de clases

Clase 0 – Presentación e Introducción al entorno Google Colab y Python – 14 de octubre

Esta clase procurará ofrecer una introducción al análisis espacial abordado desde una perspectiva programática. Se ofrecerá un ejemplo a modo de introducción del potencial de las herramientas de Python en general y de las utilizadas para el análisis espacial en particular. Luego se ahondará en los elementos fundamentales de Python como el lenguaje de programación y las librerías adicionales.

Clase 1 – Introducción a los objetos fundamentales de Python – 21 de octubre

En esta clase se introducirán los objetos fundamentales del lenguaje de programación Python (listas, diccionarios, etc) como los controles de flujo y la definición de funciones. Se abordará el proceso de iteración y aplicación de funciones como así también los objetos básicos de librerías como Numpy y Pandas.

Clase 2 – Librería de Python para manejar tablas de datos tabulares: Pandas – 28 de octubre

Habiendo recorrido los componentes fundamentales de Python en la primera clase, esta clase profundizará en las librerías que nos permiten analizar datos en formatos de tablas, en filas y columnas. En esta clase veremos la librería Pandas.

Clase 3 - Librerías de Python para manejar datos espaciales (Geopandas y Shapely) y visualizarlos en mapas (Matplotlib). Puntos y líneas – 4 de Noviembre

En esta clase, veremos cómo trabajar con las librerías de Python que manejan objetos espaciales: Geopandas y Spatially. Dentro del lenguaje Python, estas son las herramientas que permitirán llevar adelante operaciones espaciales. Se observará cómo tratar con Puntos, Polígonos y Líneas de modo individual como así también en una tabla de datos. También se profundizará un aspecto indispensable para el análisis de datos y en particular para el análisis de datos espaciales: la visualización de los datos en un modo que comuniquen con eficiencia nuestro trabajo. En el caso del análisis espacial esto implica la confección de mapas utilizando las librerías de visualización de gráficos en general que trae Python: Matplotlib.

Clase 4 – Manipulación de datos entre diferentes tablas y unidades espaciales: Joins, Group By y Dissolve. Profundización en visualizaciones y paletas de colores. Polígonos y coropletas - 11 de Noviembre

En esta clase se presentarán en las operaciones que permiten vincular set de datos entre si ya sea por atributos en común o por una relación particular en el espacio. Al mismo tiempo se profundizará en la obtención de estadísticos agrupados en función de atributos comunes o de unidades espaciales de diferente escala. En este proceso se profundizará en las estrategias de visualización y paleta de colores.

Clase 5 – Otros operadores geográficos (overlays) y formas de particionar el espacio (buffers, mallas, voronoi, delauney) – 18 de noviembre

En esta clase se profundizarán en los operadores geográficos y el tipo de relaciones espaciales entre polígonos y entre todos los tipos de unidades espaciales (puntos, líneas, polígonos). Estos procesos permiten una manipulación ad hoc de los objetos espaciales con los que se trabaja, para crear otras geometrías producto de su interrelación.

Clase 6 – Geolocalización y uso de APIs – 25 de noviembre

En esta clase se introducirá a algunas librerías abiertas y consultas de interfaces de programación de aplicaciones (APIs, del inglés application programming interface) para poder obtener datos abiertos de gran utilidad. Se realizará foco en dos casos particulares extensamente usados en el mundo del análisis espacial y los estudios urbanos: Open Street Map y Google Maps. Entre las ventajas del enfoque programático es que lenguajes como Python permiten una gran interacción con estos servicios para obtener datos.

Clase 7 – Estructurando el espacio – 2 de diciembre

Esta clase presentará algunas librerías para estructurar el espacio urbano en modelos de datos que nos permitan manipular las relaciones espaciales de otro modo. Por un lado, utilizaremos un criterio de grafos o redes por un lado para modelar la infraestructura urbana utilizando un modelo conocido y los estadísticos correspondientes a través de la librería Osmnx. También veremos la librería de teselación h3 que utiliza un esquema anidado de hexágonos para estructurar el espacio.

Metodología de trabajo

Las clases se dividirán en dos partes. La primera será de carácter más expositivo donde se presentarán los temas de cada clase a nivel conceptual. La segunda parte será de carácter práctico donde repasaremos cómo resolver diferentes problemas desde una perspectiva programática. También servirá como espacio de preguntas sobre temas de la clase anterior. Todos los trabajos, presentaciones, materiales y ejercicios estarán disponibles online para que cada uno/a de los estudiantes pueda reforzar y repasar a su ritmo fuera de la clase.

Bibliografía principal

- Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49, 45–53. <http://doi.org/http://dx.doi.org/10.1016/j.apgeog.2013.09.012> The New Urban World .
- Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211–221.
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
- Lazer, D., & Radford, J. (2017). Data ex Machina: Introduction to Big Data. *Annual Review of Sociology*, (0).
- Rey, S. (2015). Geovisualization. In GPH471: Geographic Information Analysis. Lecture slides from course taught at Arizona State University.
- Tufte, E. R. (1983). *The visual display of quantitative information*. Graphics press Cheshire, CT.
- Openshaw, S., & Openshaw, S. (1984). The modifiable areal unit problem. *Geo Abstracts* University of East Anglia.
- Lloyd, Christopher (2014), *Spatial Data Analysis: An Introduction for GIS users*, Oxford Press

Requisitos para la cursada y promoción

Para la aprobación del seminario es necesario cumplimentar el 80% de asistencia presencial y presentar un trabajo práctico que aplique las técnicas aprendidas en el curso. El mismo deberá ser reproducible, citando las fuentes y haciendo explícitas todas las decisiones y procesos.