|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  **POSGRADO EN LINGÚÍSTICA**  **MAESTRÍA EN LINGÜÍSTICA APLICADA**  **Programa de la actividad académica** | | | | | | | | | |
| Denominación: Programación de R para datos lingüísticos  Tutor/a: Max Carey | | | | | | | | | |
| Clave: | | Semestre:  **2018-2** | Campo de conocimiento:  Lingüística (Aplicada o Hispánica) Ambos | | | | | No. Créditos: | |
| Carácter: Obligatoria/Optativa | | | | Horas | | Horas por semana | | Horas al semestre | |
| Tipo:Teórica | | | | Teoría: | Práctica: |  | | 60 | |
| 6 | 36 |
| Modalidad: Curso/Seminario | | | | Duración del programa: Semestral | | | | | |
| **Actividad académica con seriación subsecuente:** na  **Actividad académica con seriación antecedente:** na | | | | | | | | | |
| **Objetivo(s) del programa:**  Aprender a utilizar R para analizar datos lingüísticos y fomentar | | | | | | | | | |
| **Objetivo general**: Familiarizar el alumno, de manera muy práctica, con la programación en R, la estadística en R, y otras herramientas y conceptos básicos de la informática para que puedan aplicarlos a sus futuros y actuales proyectos de investigación. | | | | | | | | | |
| **Objetivos específicos:**   * Manejar las herramientas y conceptos básicos para la programación con R * Utilizar R para la visualización y transformación de datos * Llevar a cabo análisis estadísticos con R * Incorporar R en los actuales programas de investigación | | | | | | | | | |
| **Índice Temático** | | | | | | | | | |
| Unidad | Tema | | | | | | Horas | | |
|  | | | | | | | Teóricas | | Prácticas |
| 1 | Herramientas y conceptos básicos de la informática | | | | | | 0 | | 10 |
| 2 | Programación básica con R | | | | | | 0 | | 14 |
| 3 | R Studio y El Tidyverse | | | | | | 0 | | 6 |
| 4 | Estadística descriptiva e inferencial con R | | | | | | 0 | | 20 |
| 5 | Lingüística de corpus con R y otras herramientas | | | | | | 0 | | 10 |
| 6 |  | | | | | |  | |  |
| 7 |  | | | | | |  | |  |
| 8 |  | | | | | |  | |  |
| 9 |  | | | | | |  | |  |
| 10 |  | | | | | |  | |  |
| **Total de horas** | | | | | | |  | | 60 |
| **Suma total de horas:** | | | | | | |  | | |

**Contenido Temático desglosado**

|  |  |
| --- | --- |
| **Unidad** | **Tema/subtemas** |
|  | 1. Cómo interactuar con el sistema a través del *command line* para poder tener una buena organización de los archivos en el disco duro (4 horas datacamp.com y 2 horas de taller) 2. Herramientas básicas: (cómo instalarlas o usarlas): editores de texto (*sublime text, text wrangler*), *syntax highligting*, *stack overflow* para preguntas y dudas, codificación de los archivos (p. ej., utf-8 vs. ASCII), control de versión básico *(Git y Github)*, R Studio (2 horas de taller) |
|  | 1. Cómo mostrar *output* en R, (programas de tipo “*hello world*”) las estructuras de control (condicionales, iteraciones), en R variables, funciones, funciones de la familia *apply()* de R (10 horas de datacamp.com en casa y 4 horas de taller) |
|  | 1. *Dplyr* para manipulación de datos, *ggplot2* para la visualización de datos (4 horas de datacamp.com en casa y 2 horas de taller) |
|  | 1. Distribuciones, histogramas, *boxplots*, *scatterplots*, wordclouds, correlación, regresión simple y múltiple, regresión logística, modelos de efectos mixtos (20 horas de datacamp.com en casa y 8 horas de taller ) |
|  | 1. Anotación de partes de la oración, medidas de frecuencia: *frecuencia de token vs. frecuencia de tipo*, generación de concordancias, gestores de corpus (*GECO*, *AntConc*, *SketchEngine*) (5 horas de taller) 2. Trabajar con los propios datos de los hablantes (5 en casa y 10 horas de taller) |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
| **Bibliografía básica:**  **Unidad 1** **(**Herramientas y conceptos básicos de la informática**)**  1. Introduction to Shell for Data Science, Chapter 1 (Obligatoria), Chapter 2-6 (Opcional) (datacamp.com) <https://www.datacamp.com/courses/introduction-to-shell-for-data-science> (4 horas)  **Unidad 2 (**Programación básica con R)  2. Introduction to R (datacamp.com) <https://www.datacamp.com/courses/free-introduction-to-r> (4 horas)  **Unidad 3 (**R Studio y El Tidyverse)  4. Introduction to the Tidyverse <https://www.datacamp.com/courses/introduction-to-the-tidyverse> (4 horas)  **Unidad 4 (**Estadística descriptiva y inferencial con R)  5. Introduction to Data (datacamp.com) <https://www.datacamp.com/courses/introduction-to-data> (4 horas)  6. Exploratory Data Analysis (datacamp.com) <https://www.datacamp.com/courses/exploratory-data-analysis> (4 horas)  5. Correlation and Regression (datacamp.com) <https://www.datacamp.com/courses/correlation-and-regression> (4 horas)  6. Multiple and Logistic Regression (datacamp.com) (4 horas)  7. Foundations of Inference (datacamp.com) <https://www.datacamp.com/courses/foundations-of-inference>  (4 horas)  **Unidad 5:**  8. Adolphs, S. (2011). Corpus Linguistics. En J. Simpson (Ed.), The Routledge handbook of applied linguistics (1st ed). Milton Park, Abingdon, [UK] ; New York: Routledge. | |
| **Bibliografía complementaria:**   1. Gries, S. T. (2017). Quantitative corpus linguistics with R: a practical introduction (Segunda Edición). New York: Routledge, Taylor & Francis Group. 2. Adolphs, S. (2011). Corpus Linguistics. En J. Simpson (Ed.), The Routledge handbook of applied linguistics (1st ed). Milton Park, Abingdon, [UK] ; New York: Routledge. 3. Weisser, M. (2016). Practical corpus linguistics: an introduction to corpus-based language analysis. Chichester, West Sussex [England] ; Malden, MA: Wiley Blackwell. 4. Intro to Computer Science (udacity.com) https://www.udacity.com/course/intro-to-computer-science--cs101 5. Importing Data in R (Part 1) (datacamp.com) <https://www.datacamp.com/courses/importing-data-in-r-part-1> 6. Importing Data in R (Part 2) (datacamp.com) <https://www.datacamp.com/courses/importing-data-in-r-part-2> 7. Otros cursos en datacamp.com para los propósitos específicos del alumno. \*Datacamp.com tiene muchos cursos y se esperaría que los alumnos se animaran para seguir aprendiendo en esta plataforma u otra. | |
| **Estrategias didácticas:**  Exposición oral (x)  Exposición audiovisual ( )  Ejercicios dentro de clase (x)  Ejercicios fuera del aula (x)  Lecturas obligatorias ( )  Trabajo de Investigación ( )  Otro: | **Mecanismos de evaluación del aprendizaje de los alumnos:**  Exámenes Parciales ( )  Examen final escrito ( )  Trabajos y tareas fuera del aula (x)  Exposición de seminarios por los alumnos ( )  Participación en clase ( )  Asistencia (x)  Seminario ( )  Otro: Ejercicios fuera de aula |
| **Perfil profesiográfico:**  Licenciado en lingüística romance, experiencia profesional enseñando a programar, publicaciones y tesis que utilizaron R, Diplomado en Programación para aplicaciones web, experiencia profesional programación para el análisis de datos | |

Fecha: 5 de enero de 2018